

# A Cross-Scene Few-Shot Learning Based on Intra–Inter Domain Contrastive Alignment for Hyperspectral Image Change Detection

Wenhui Hou, Jiangtao Peng<sup>✉</sup>, Senior Member, IEEE, Bing Yang<sup>✉</sup>, Lanxin Wu,  
and Weiwei Sun<sup>✉</sup>, Senior Member, IEEE

**Abstract**—Recently, deep neural networks have demonstrated outstanding performance in hyperspectral image (HSI) change detection (CD), especially when there are sufficient labeled samples. However, the labels of HSI are difficult to obtain, and acquiring enough labels to train deep network is a great challenge in practice. Therefore, to mitigate the effect of insufficient labels in detection results, this article proposes a cross-scene few-shot learning (FSL) network based on intra–inter domain contrastive alignment (CAFSL) for HSI-CD, which combines contrastive learning (CL) and FSL into a unified framework, aiming to achieve better detection results using only a few labeled samples. Specifically, we perform cross-scene FSL using a pair of dual-phase images from a very high-resolution image (VHRI) as the source domain and a pair of HSI as the target domain. Then, an intra-domain supervised CL (INSCL) module is designed to enhance the compactness within classes and widen the discrimination between classes by maximizing the feature similarity of intra-class and minimizing the feature similarity of inter-class. Finally, a cross-domain contrastive alignment (CRCA) module is proposed to align the features of source and target domains, which mitigates the effect of domain migration problems caused by different data types. Experiments on three HSI benchmark datasets reveal that the CAFSL algorithm outperforms current advanced algorithms based on deep learning (DL) and FSL while with limited labeled samples.

**Index Terms**—Contrastive learning (CL), domain alignment, few-shot learning (FSL), hyperspectral image (HSI) change detection (CD).

## I. INTRODUCTION

**H**YPERSPECTRAL imaging is an important remote sensing technology. Unlike conventional three-channel

images, hyperspectral image (HSI) can reach dozens or even hundreds of spectral dimensions, in which each pixel is represented as a high-dimensional vector, which is commonly used in diverse fields because of the richer spectral information it can provide [1], [2], [3]. As technology evolves, HSI change detection (HSI-CD) has become an essential tool to monitor the process of change in ground cover [4]. HSI-CD aims to identify the process of change occurring in the same geographic region at various observation times [5], [6], [7]. By analyzing dual time-phase images of the same region, HSI-CD can provide detailed information about changes in the region. Over the past few decades, HSI-CD techniques are widely and successfully applied in many remote sensing fields such as environmental monitoring, ecological planning, disaster monitoring, etc. [8], [9], [10], [11], [12], [13], [14].

According to the research of previous experts and scholars, HSI-CD methods include algebra-based methods, transformation-based methods, classification-based methods, etc. [15]. Earlier algebra-based methods included image differencing, change vector analysis (CVA) [16], etc. CVA is a classical algebraic approach that utilizes spectral change vectors for two different dates for calculations, where the change is shown by the magnitude of vectors between two temporal phases and the changed/unchanged regions between dual-phases are determined based on thresholds. The common point of the algebra-based method is a threshold that needs to be determined to judge whether the region has changed or not, so the limitation of these methods is the difficulty in choosing a suitable threshold. Transformation-based methods project HSIs into a new feature space, and the representative methods include principal components analysis (PCA) [17], [18], [19], Kauth–Thomas transformation [20], and other methods [21]. Nielsen [20] proposed an iteratively reweighted multivariate alteration detection approach, assigning different weights to observations and performing weighted iterations based on the chi-square distance. Deng et al. [18] used the PCA to improve the change information in multisensor data after geometrically correcting and radiometrically normalizing the images. Classification-based methods contain strategies such as post-classification comparison and direct classification. Direct classification strategies are generally based on a classifier such as support vector machine (SVM) for classifying the areas as changed or unchanged [22]. The traditional

Received 18 February 2025; revised 2 April 2025; accepted 28 April 2025. Date of publication 5 May 2025; date of current version 19 May 2025. This work was supported in part by Zhejiang Province “Pioneering Soldier” and “Leading Goose” Research and Development Project under Grant 2023C01027, in part by the National Natural Science Foundation of China under Grant 42471417 and Grant 42471404, and in part by Zhejiang Provincial Natural Science Foundation of China under Grant LQ23F020012. (Corresponding authors: Jiangtao Peng; Bing Yang; Weiwei Sun.)

Wenhui Hou, Jiangtao Peng, and Lanxin Wu are with Hubei Key Laboratory of Applied Mathematics, Faculty of Mathematics and Statistics, and the Key Laboratory of Intelligent Sensing System and Security, Ministry of Education, Hubei University, Wuhan 430062, China (e-mail: pengjt1982@hubei.edu.cn).

Bing Yang is with the College of Sciences, China Jiliang University, Hangzhou 310018, China (e-mail: bingyang0517@163.com).

Weiwei Sun is with Ningbo Key Laboratory of Remote Sensing and Ecological Security of Coastal Zone, Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo 315211, China (e-mail: sunweiwei@nbu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2025.3566701

HSI-CD methods have the advantages of simple calculation and easy implementation, but their feature extraction capability is limited and it is hard to take full advantage of the rich spatial-spectral information, which limits the model's performance to some extent.

In the past few years, deep learning (DL) is gradually used in remote sensing and has shown excellent performance with sufficient samples [23], [24], [25], [26], [27]. Many DL methods have obtained outstanding results in HSI-CD. For example, Luo et al. [28] proposed a multiscale diff-changed feature fusion network (MSDFFN), utilizing the proposed bidirectional diff-changed feature representation module to learn the fine-changed features of dual-phase images at different scales and to improve the fine-variation discriminative performance. Hu et al. [29] proposed a pixel-level self-supervised hyperspectral spatial-spectral feature understanding network for HSI-CD, which designs a spatial-spectral attention module to explore the spatial correlation and discriminative spectral features of HSIs, respectively, and uses a focal cosine loss function to address the imbalance between easy and hard positive samples. Yang et al. [30] proposed a sharpening-aware graph network (SAGN) to achieve HSI-CD. SAGN includes a graph convolution module based on Laplacian sharpening to emphasize the change information between dual-phase HSI. In addition, an untied "difference map" was constructed for dual-phase HSI to simulate the difference between the changed pixels and their neighbors while dynamically updating the graph structure to further maximize the divergence. Ding et al. [31] proposed a multilevel feature fused and change information-enhanced neural network. This method utilizes hierarchical features of dual-temporal images to achieve accurate detection through multilevel feature extraction, change enhancement, and feature fusion. Yu et al. [32] proposed a gated spectral-spatial-temporal attention network with spectral similarity filtering, which extracted more information without sacrificing the performance of the model through three mechanisms. All of the above DL methods achieved better performance in HSI-CD. Nevertheless, DL approaches generally require lots of labels to train the model and tune built-in parameters. In reality, collecting labels needs manual involvement, which takes a lot of human, material, and financial resources. Therefore, a major challenge in DL is the scarcity of labels. In order to address this issue, scholars have proposed many approaches for small-sample situations, such as few-shot learning (FSL), lightweight network, etc.

Inspired by the human's ability to learn fast, researchers hope machine learning models can adapt to new categories with minimal samples quickly after learning from abundant data [33], [34], [35]. This is what FSL needs to solve. FSL applies meta-learning in the field of supervised learning. There are three main meta-learning strategies for FSL: model-based methods, metric-based methods, and optimization-based methods [36]. Specifically, model-based methods [37] adapt to new tasks fast on few samples by designing model structures; the metric-based method [38], [39] measures the distance between samples in each batch and samples in support set, and utilizes the nearest neighbor (NN) idea to complete classification; and the optimization-based method [40] completes the FSL task

by adjusting the optimization approach. In existing studies, the metric-based method has become the most commonly used scheme in meta-learning because of its simplicity, efficiency, stable performance, and ability to quickly adapt to FSL by learning similarity metrics in feature space.

During the meta-training phase of FSL, different meta-tasks are obtained for each training. Thus, the overall training contains different categories combinations. When facing a new category, it can also categorize better by this learning mechanism. Based on this advantage, cross-scene FSL has received attention from a wide range of experts and scholars. For instance, in HSI classification field, Li et al. [41] proposed a deep cross-domain FSL (DCFSL) method, which uses a conditional adversarial domain adaptation (DA) strategy to overcome the domain bias, and solves the FSL and DA problems within an integrated framework. Liu et al. [42] proposed a refined prototype contrastive learning FSL (RPCL-FSL) network to optimize the support prototype by triple constraints based on contrastive learning (CL), self-calibration, and cross-calibration. Cao et al. [43] proposed a spatial-spectral-semantic method based on FSL, which introduces a semantic-aware strategy to enhance the association between cross-scene data and semantic information, and utilizes feature denoising to improve the meta-learning effect. FSL-based methods use the idea of prototype to enhance the intra-class compactness, avoids the instability problem of feature distribution when there is insufficient data, and can improve the generalization performance of the model. Therefore, FSL can still be used to solve the HSI-CD problem when there are few labels. With the continuous advancement of FSL research, a small number of studies have applied FSL in HSI-CD. For example, Wang et al. [44] designed a bidirectional generation-based cross-domain FSL framework (BiG-FSLF) for HSI-CD, and proposed a bidirectional generative DA method based on generative adversarial strategies to implement adaptive alignment of cross-domain feature distributions. Song et al. [45] proposed an FSL method based on 3-D convolutional neural network and normalized flow for HSI-CD, which analyzes the spectral-spatial distribution of hyperspectral variations by normalizing flow and generates hyperspectral tensor samples approximating the distribution. However, the existing FSL-based methods in HSI-CD do not fully consider the distinction between categories. Therefore cross-scene FSL still has broad research potential in HSI-CD.

The two datasets for the cross-scene FSL problem mentioned above are from different domains, which leads to one of the main problems, that is, the different data structure between the source and target domains, which with significant domain differences. For the cross-domain problem, DA is a valid method to solve this issue [46], [47], [48], [49]. Huang et al. [47] proposed a two-branch attention adversarial DA network, which utilizes a two-branch feature extraction sub-network to extract features and improves the cross-domain classification ability through adversarial learning of generators and discriminators. Feng et al. [50] proposed a DA method for class alignment and class balance generation for HSI classification, which extracts features through a hierarchical capsule network and generates pseudo-labels for target

domain. Meanwhile, the sampler adaptively adjusts the sample distribution according to the pseudo-labels and credibility. The above methods show that the DA approach can improve the network's discriminative and generalization abilities under cross-scene conditions.

Meanwhile, CL has a wide range of applications in the HSI-CD field. Ou et al. [51] proposed an HSI-CD framework with a self-supervised CL pretrained model, which utilizes the contrastive loss function based on Pearson correlation coefficient and negative cosine correlation to make the features extracted by the network close to each other. Wu et al. [52] proposed a differential band reweighting and unbalanced CL framework for HSI-CD. This method takes into account the sample distribution imbalance and intra-class differences, and designs an unbalanced CL method based on multiple prototypes to adjust the contribution of samples in different categories to the loss by estimating multiple prototypes.

In summary, to solve the problem of insufficient labels in DL and domain bias across domains, we propose a cross-scene FSL network based on intra-inter domain contrastive alignment (CAFSL) for HSI-CD, which integrates supervised CL and FSL into a unified framework. After differentiating the images in source and target domains, respectively, all samples are subjected to feature extraction by designed feature extractor, and the similarity between the prototypes in support set and the samples in query set is calculated using a metric function. At the same time, intra-class similarity and inter-class distinction of features are emphasized through intra-domain CL. Since the source and target domains are from different scenes, we compute the similarity of positive and negative sample pairs by cross-domain contrastive alignment (CRCA), aiming to align features of same class in different domains. The main contributions of the CAFSL algorithm are as follows.

- 1) This article proposes a new CAFSL framework for HSI-CD, combining supervised CL and cross-scene FSL under the same framework. Good results are achieved on three HSI-CD benchmark datasets using just a few labeled samples.
- 2) A CRCA strategy is designed to alleviate the domain transfer problem. The cross-domain contrastive loss is calculated using the source and target domain samples as anchor points, respectively, and the features of source and target domains are aligned to the same feature space. This module effectively handles the distribution differences between source and target domains, improving the performance of cross-domain tasks.
- 3) We construct an intra-domain contrastive support set to improve intra-class separation and inter-class dispersion, allowing the model to better distinguish feature representations of different categories.

The rest of this article is assigned as follows. Section II describes the CAFSL algorithm proposed in this article detailedly. Section III performs a series of experiments on three hyperspectral datasets of HSI-CD, and Section IV shows parameter analysis and visualization results of the article. Finally, Section V summarizes this article and discusses potential directions for future research.

## II. PROPOSED METHOD

This section describes the overall framework and detailed design of CAFSL algorithm. The CAFSL network integrates supervised CL and cross-scene FSL into a unified framework. The framework of CAFSL is shown in Fig. 1.

For the CAFSL algorithm, we use an alternating training method for overall training, that is, training the source domain in odd rounds and the target domain in even rounds. That is, our CAFSL framework contains two meta-training processes. After differentiating the dual-phase very high-resolution image (VHRI) in source domain and the dual-phase HSI in target domain, the source domain dataset  $D_s$  and the target domain dataset  $D_t$  are obtained. The training process of the algorithm is described using  $D_s$  as an example. After dividing  $D_s$  and  $D_t$  into contrastive support set and query set, feature extractor  $\mathcal{F}$  is utilized to unify spectral dimensions and feature extraction. Subsequently, the contrastive loss of sample pairs in contrastive support set and the similarity between prototypes in support set and the samples in query set are computed. Finally, the labeled samples of  $D_s$  are used as an anchor point, and the CRCA loss of  $D_s$  and  $D_t$  is computed. The algorithm has completed one round of iterations at this point. The training process on  $D_t$  is similar. After completing all the training, the model is tested by the NN classifier [53], which is shown in Fig. 2.

### A. Data Preprocessing

In FSL problems, the training set is usually divided into support set and query set. Assuming that the data has  $N$  classes,  $K$  labeled samples are randomly selected from each class as the support set, meanwhile a small number of samples are picked from remaining samples as the query set, that is, the  $N$ -way,  $K$ -shot problem. In our algorithm, the division of support and query sets is the same for both source and target domains. Taking the source domain  $D_s$  with  $N$  classes as an example, we select  $k_s$  samples from each class randomly as the support set  $S_s = (x_s^i, y_s^i)_{i=1}^{n_s}$ , and then select  $k_q$  samples randomly as the query set  $Q_s = (x_s^j, y_s^j)_{j=1}^{n_q}$ , where  $n_s = k_s \times N$  and  $n_q = k_q \times N$ .  $S_t$  and  $Q_t$  are constructed as above.

In order to adapt to the construction of the contrastive group of the subsequent CL training set, we assign the two samples of the same category into separate groups in  $S_s$  and  $S_t$ , thereby forming two contrastive groups, in which the two samples of the same class are positive samples to each other, the samples of different classes are negative samples to each other.

### B. Feature Mapping Extraction Network

Considering that both the VHRI of  $D_s$  and the HSI of  $D_t$  have rich spatial information, we design a 2-D convolution-based DL network named feature extraction mapping network ( $\mathcal{F}$ ) as the feature extractor. The detailed architecture of  $\mathcal{F}$  is shown in Fig. 3.

Since different images have different feature dimensions (e.g., the spectral dimension  $B_s$  of the LEVIR-CD in  $D_s$  is 3, and the spectral dimension  $B_t$  of the Bay Area in  $D_t$  is 224). To unify the input dimensions of  $D_s$  and  $D_t$ , the feature dimensions of  $D_s$  and  $D_t$  are unified to  $d$  using two mapping

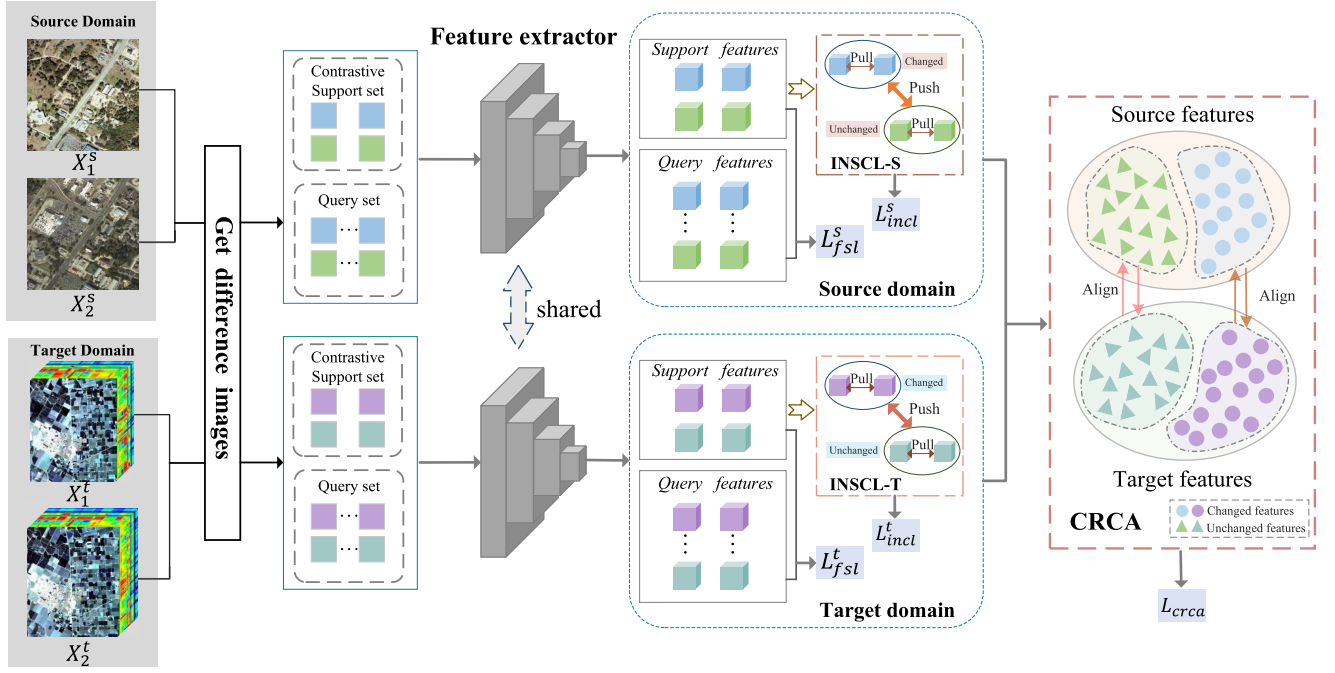


Fig. 1. Framework of CAFSL. The labeled samples in source and target domains are divided into contrastive support set and query set, respectively. After extracting features from two domains, the INSCl module is utilized to enhance the differentiation between changed and unchanged categories. Then, CRCA strategy is used to align features to mitigate the domain shift.

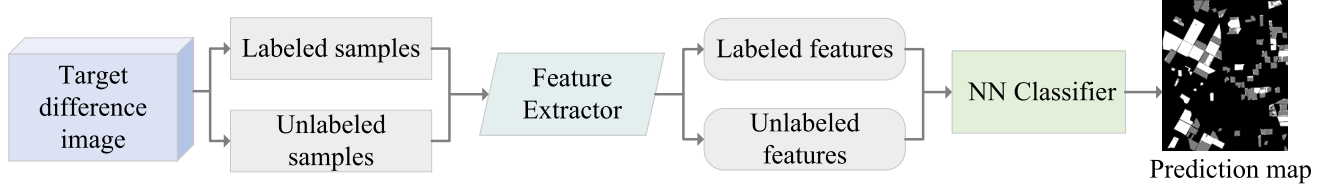


Fig. 2. Testing process of CAFSL.

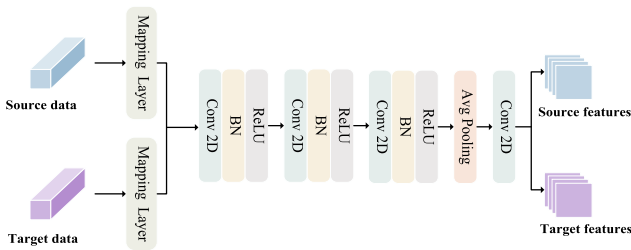


Fig. 3. Architecture of  $\mathcal{F}$ .

layers  $M_s$  and  $M_t$  in  $\mathcal{F}$  initially. The output of mapping layers is as follows:

$$f_{\text{map}_s} = M_s(f_s) \quad (1)$$

$$f_{\text{map}_t} = M_t(f_t) \quad (2)$$

where  $f_{\text{map}_s}, f_{\text{map}_t} \in \mathbb{R}^{p \times p \times d}$ , and  $f_s \in \mathbb{R}^{p \times p \times B_s}$  and  $f_t \in \mathbb{R}^{p \times p \times B_t}$  are the input features of  $D_s$  and  $D_t$ , respectively.  $p$  is the patch size and  $d$  is set to 100 in our algorithm.

Subsequently,  $f_{\text{map}_s}$  and  $f_{\text{map}_t}$  are input into  $\mathcal{F}$  to further extract the spatial-spectral features, respectively. Specifically,  $\mathcal{F}$  is composed of three 2-D convolutional layers and a

pooling layer, where each 2-D convolution is followed by a normalization (BN) layer and a rectified linear unit (ReLU) activation function. The process can be expressed as follows:

$$\mathcal{F}(f_{\text{map}_i}) = \text{AvgPooling}(\Phi_3(\Phi_2(\Phi_1(f_{\text{map}_i})))) \quad (3)$$

where  $i = s, t$  and  $\Phi_j$  is as follows:

$$\Phi_j(H) = \text{Relu}(\text{BN}_j(\text{Conv}_j(H))), \quad j = 1, 2, 3. \quad (4)$$

Considering the need to utilize the features output from the network for subsequent procedures, we connect a 2-D convolution after the pooling layer for FSL- and CL-related computations.

### C. Cross-Scene FSL

FSL can learn the feature differences of different categories with only a little label and improve the generalization performance of the model. The training process of FSL requires two datasets: a training set with sufficient labels and a test set with just a few labels, and the two datasets can be the same or different types. In our CAFSL method, after the feature extraction by  $\mathcal{F}$ , the samples of support set in  $D_s$  and  $D_t$  obtain a set of features  $F_s = \{f_s^1, \dots, f_s^{n_s}\}$  and  $F_t = \{f_t^1, \dots, f_t^{n_t}\}$ ,



respectively. Then, the  $c$ th support set prototype of  $D_s$  is calculated as follows:

$$\text{Pro}_s^c = \frac{1}{k_s} \sum_{f_s^i \in F_s^c} f_s^i. \quad (5)$$

Similarly, the prototype of the  $c$ th support set in  $D_t$  is as follows:

$$\text{Pro}_t^c = \frac{1}{k_s} \sum_{f_t^i \in F_t^c} f_t^i. \quad (6)$$

Likewise, the samples of the query sets in  $D_s$  and  $D_t$  also have a set of features after  $\mathcal{F}$ . Taking  $D_s$  as an example, for any  $x_s^j \in Q_s$ , which belongs to the  $c$ th class, the probability distribution of  $x_s^j$  for class  $c_j$  can be expressed as follows:

$$p(y_j = c_j | x_s^j) = \frac{\exp(-d(f_s^j, \text{Pro}_s^{c_j}))}{\sum_{c=1}^N \exp(-d(f_s^j, \text{Pro}_s^c))} \quad (7)$$

where  $f_s^j$  represents the features of  $x_s^j$  after extracting by  $\mathcal{F}$ , and  $d(\cdot)$  represents the Euclidean distance between features of the prototype of the support set and query set.

Ultimately, the loss function  $L_{\text{fsl}}^s$  for the source domain is obtained by minimizing the negative value of the class probability distribution of the query set samples, which can be expressed as follows:

$$L_{\text{fsl}}^s = -\frac{1}{n_q} \sum_{j=1}^{n_q} \log p(y_j = c_j | x_s^j). \quad (8)$$

The FSL process of  $D_t$  is the same as  $D_s$ , and its loss function can be expressed as follows:

$$L_{\text{fsl}}^t = -\frac{1}{n_q} \sum_{j=1}^{n_q} \log p(y_j = c_j | x_t^j). \quad (9)$$

In summary, the overall FSL loss function of the model can be denoted as follows:

$$L_{\text{fsl}} = L_{\text{fsl}}^s + L_{\text{fsl}}^t. \quad (10)$$

#### D. Intra-Domain Supervised CL

The FSL in Section II-C utilizes the idea of prototype-based learning. However, such an approach focuses only on intra-class similarity, and does not consider the similarity of the changed and unchanged samples in the feature spaces, which may lead to inter-class confusion. In order to address this problem, we design an intra-domain supervised CL (INSCL) module.

The purpose of CL is to promote the closeness of intra-class samples and the discreteness of inter-class samples, and thus to learn common and divergent features simultaneously [54]. The optimization of CL is to minimize positive pair distances while maximizing negative pair distances in the feature space, that is,

$$\text{Sim}(\mathcal{F}(x), \mathcal{F}(x^+)) \gg \text{Sim}(\mathcal{F}(x), \mathcal{F}(x^-)) \quad (11)$$

where  $\text{Sim}(\cdot)$  represents the similarity between features,  $x$  and  $x^+$  indicate the positive sample pair, and  $x$  and  $x^-$  indicate the negative sample pair.

In INSCL module, we consider the information of the real labels of samples, and regard the samples belonging to the same class as positive samples, and samples of different classes as negative samples. Taking the source domain support set  $S_s$  as an example, we select two samples for each category as support set separately, and subsequently set them into two contrastive groups:  $C_s^1 = \{x_s^1, x_s^3, \dots, x_s^{2N-1}\}$  and  $C_s^2 = \{x_s^2, x_s^4, \dots, x_s^{2N}\}$ , where  $x_s^{2k-1}$  and  $x_s^{2k}$  belong to the same class, that is, mutually positive samples. At this point, the contrastive loss between  $x_s^m$  and  $x_s^n$  can be calculated as follows:

$$l_{m,n}^s = -\log \frac{\exp(\phi(f_s^m, f_s^n)/\tau)}{\sum_{k=1}^{2N} \mathbf{1}_{k \neq m} \exp(\phi(f_s^m, f_s^k)/\tau)} \quad (12)$$

where  $f_s^m$  and  $f_s^n$  are the features of samples  $x_s^m$  and  $x_s^n$  belonging to the same category in  $D_s$ ;  $\phi$  represents the Euclidean distance, which is used to measure the similarity;  $\mathbf{1}$  is the indicator function; and  $\tau$  is the temperature parameter of CL.

To compute the intra-domain contrastive loss in an episode in source domain, for each sample pair  $(x_s^{2k-1}, x_s^{2k})$  and its inverse pair  $(x_s^{2k}, x_s^{2k-1})$ , the mean is taken after calculating the contrastive loss  $l_{2k-1,2k}^s$  and  $l_{2k,2k-1}^s$ , respectively. The final supervised contrastive loss of  $D_s$  can be obtained as follows:

$$L_{\text{incl}}^s = \frac{1}{2N} \sum_{k=1}^N (l_{2k-1,2k}^s + l_{2k,2k-1}^s). \quad (13)$$

Similarly, the supervised contrastive loss in  $D_t$  is obtained as follows:

$$L_{\text{incl}}^t = \frac{1}{2N} \sum_{k=1}^N (l_{2k-1,2k}^t + l_{2k,2k-1}^t). \quad (14)$$

Therefore, the overall loss of the INSCL module can be denoted as follows:

$$L_{\text{incl}} = L_{\text{incl}}^s + L_{\text{incl}}^t. \quad (15)$$

#### E. Cross-Domain Contrastive Alignment

For traditional cross-scene FSL problems, the source and target domains have different classes and data distributions, which lead to limitations when using domain alignment. However, in CD task, although the image and land cover types in the source and target domains are different, all the samples can be categorized as “changed” or “unchanged” after differencing. Even with a significant difference between data in source and target domains, the changed/unchanged samples in source domain are always more close to the changed/ unchanged samples in target domain in the feature space. Based on this situation, we hope that using certain strategies can make the feature distributions of source and target domains consistent, decrease the domain bias, and improve the generalization performance of the target domain. It also makes the features under similar classes (e.g., the changed class of source domain and the changed class of target domain) as close as possible, and the features of different classes (e.g., the changed class of source domain and the unchanged class of target domain) as far as possible. Therefore, we propose a CRCA

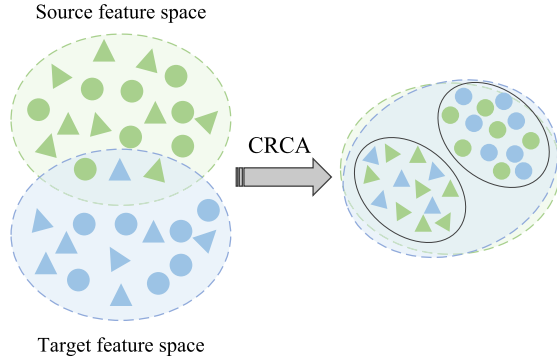


Fig. 4. Schematic of the feature space of CRCA.

module, which aims to unify the data of source and target domains under the same feature space for feature matching. The CRCA representation is shown in Fig. 4.

As there are only a very small number of labeled samples in  $D_t$ , in each batch we treat the labeled samples in  $D_s$  (i.e.,  $CL_s = S_s \cup Q_s$ ) and the labeled samples in  $D_t$  ( $CL_t = S_t \cup Q_t$ ) as the sample sets for domain alignment, and perform the CRCA operations from source to target domain and from target to source domain in different episodes, respectively.

Here, we show the details of aligning the source to target domain. After feature extraction by  $\mathcal{F}$ , each sample in  $CL_s$  and  $CL_t$  is given a feature, and the feature set can be denoted as  $F_{cl_s} = \{f_s^1, \dots, f_s^{n_s+n_q}\}$  and  $F_{cl_t} = \{f_t^1, \dots, f_t^{n_t+n_q}\}$ . Regarding sample  $x_s^i$  of the  $c$ th class in source domain as an anchor point, all the samples in the  $c$ th class that belong to  $CL_t$  are positive pairs with each other, and the rest of the samples, that are not in the  $c$ th class, is negative samples. Then the similarity matrix of positive samples in  $D_s$  and  $D_t$  can be denoted as follows:

$$\mathbf{sim}_{s \rightarrow t}^+(x_s^i, x_t^j) = \exp\left(\frac{\varphi(x_s^i, x_t^j)}{\tau}\right) \quad (16)$$

where  $x_{t+}^j$  represents the samples which are of the same kind as  $x_s^i$  in target domain (i.e., positive samples), and  $\varphi$  is the similarity measure, and here we choose cosine similarity for calculation.  $\tau$  is the temperature hyperparameter. Similarly, the similarity matrix of negative sample pairs can be written as follows:

$$\mathbf{sim}_{s \rightarrow t}^-(x_s^i, x_t^j) = \exp\left(\frac{\varphi(x_s^i, x_{t-}^j)}{\tau}\right). \quad (17)$$

Inspired by [55], the denominator in our CRCA loss only computes the similarity between the anchor and all negative samples  $\mathbf{sim}_{s \rightarrow t}^-(x_s^i, x_t^j)$  in the sum (hereafter denoted as  $\mathbf{sim}_{s \rightarrow t}^+$ ,  $\mathbf{sim}_{s \rightarrow t}^-$ ). Then the CRCA loss from the source to target domain is as follows:

$$L_{crca}^{s \rightarrow t} = -\frac{1}{|CL_t^+|} \sum \log \frac{\mathbf{sim}_{s \rightarrow t}^+}{\mathbf{sim}_{s \rightarrow t}^+ + \sum_{|CL_t^-|} \mathbf{sim}_{s \rightarrow t}^-} \quad (18)$$

where  $CL_t^+/CL_t^-$  represents the set of positive/negative samples in the target domain set, and  $|CL_t^+|/|CL_t^-|$  denotes the number of positive/negative samples in the set.

#### Algorithm 1 CAFSL

**Input:** Source-domain VHR dual time-phase images  $D_s^1$  and  $D_s^2$ , labels  $y_s$ ; Target domain HS dual-phase images  $D_t^1$  and  $D_t^2$ , a small number of labels  $y_t$ ; Number of iterations  $R$ .

**Output:** Target domain predicted label  $Y$ .

**Begin**

For episode in range  $[1, R]$ :

Feature extraction of the difference images  $D_s$  and  $D_t$  in source and target domains by  $\mathcal{F}$

**If episode%2 == 0:**

(1) Calculating source domain FSL loss  $L_{fsl}^s$  by (8)

(2) Calculating source domain INSCL loss  $L_{incl}^s$  by (13)

(3) Calculating the CRCA loss  $L_{crca}^{s \rightarrow t}$  from source to target domain by (18)

(4) Get the overall loss  $L^s$  of source domain and return the parameters

**else:**

(1) Calculating target domain FSL loss  $L_{fsl}^t$  by (9)

(2) Calculating target domain INSCL loss  $L_{incl}^t$  by (14)

(3) Calculating the CRCA loss  $L_{crca}^{t \rightarrow s}$  from target to source domain by (19)

(4) Get the overall loss  $L^t$  of target domain and return the parameters

**End if**

**End**

In the same way, it can be obtained that the contrastive loss from the target to source domain in an episode is as follows:

$$L_{crca}^{t \rightarrow s} = -\frac{1}{|CL_s^+|} \sum \log \frac{\mathbf{sim}_{t \rightarrow s}^+}{\mathbf{sim}_{t \rightarrow s}^+ + \sum_{|CL_s^-|} \mathbf{sim}_{t \rightarrow s}^-}. \quad (19)$$

The final contrastive loss needs to traverse each sample in  $CL_s$  and  $CL_t$ , respectively; then the overall CRCA loss can be expressed as follows:

$$L_{crca} = \frac{1}{|CL_s|} L_{crca}^{s \rightarrow t} + \frac{1}{|CL_t|} L_{crca}^{t \rightarrow s}. \quad (20)$$

#### F. Total Loss Function

In the overall training process of our proposed CAFSL algorithm, three main processes are included, namely cross-scene FSL, INSCL, and CRCA. We use three loss functions  $L_{fsl}$ ,  $L_{incl}$ , and  $L_{crca}$  to update the relevant parameters of feature extractor  $\mathcal{F}$ . Therefore, the overall loss function of CAFSL can be denoted as follows:

$$L_{total} = L_{fsl} + \alpha L_{incl} + \beta L_{crca} \quad (21)$$

where  $\alpha$  and  $\beta$  are both internal hyperparameters of the algorithm, and the specific values of both will be explained in detail in Section IV. The overall procedure of the CAFSL algorithm is summarized in Algorithm 1.

TABLE I  
DETAILS OF FOUR DATASETS

Dataset	Size	T1 (Year)	T2 (Year)	Unchanged Samples	Changed Samples
LEVIR-CD	1024×1024	2002	2013	926627	121949
Bay Area	600×500	2013	2015	34211	39270
Santa Barbara	984×740	2013	2014	80418	52134
Farmland	450×140	2006	2007	44723	18277

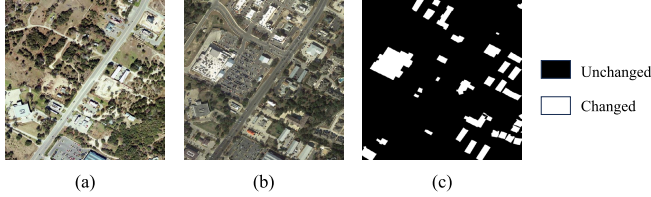


Fig. 5. LEVIR-CD dataset. (a) Pseudo-color composite image on T1. (b) Pseudo-color composite image on T2. (c) Ground truth.

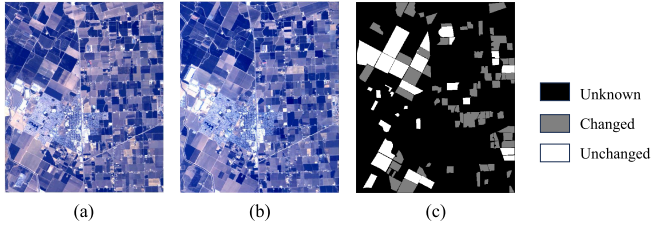


Fig. 6. Bay Area dataset. (a) Pseudo-color composite image on T1. (b) Pseudo-color composite image on T2. (c) Ground truth.

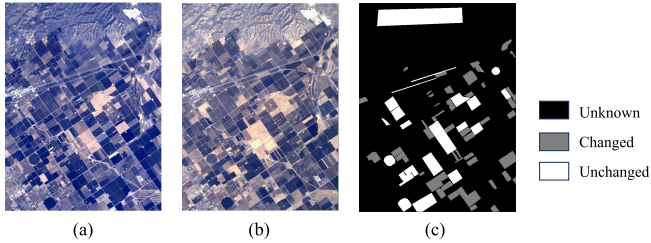


Fig. 7. Santa Barbara dataset. (a) Pseudo-color composite image on T1. (b) Pseudo-color composite image on T2. (c) Ground truth.

### III. EXPERIMENTS

#### A. Datasets

In our algorithm, four datasets are used totally, including a VHRI dataset as source domain and three HSI datasets as target domain. The detailed information and schematic diagrams of four datasets are shown in Table I and Figs. 5–8, respectively. The introduction is as follows.

**LEVIR-CD:** LEVIR-CD is a large-scale remote sensing building CD dataset, which consists of 637 very high-resolution Google Earth image patch pairs with a size of  $1024 \times 1024$  pixels. We randomly select a pair of bi-temporal images as the source domain data.

**Bay Area:** The Bay Area dataset was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor in the San Francisco Bay Area in Paterson, CA, USA, in

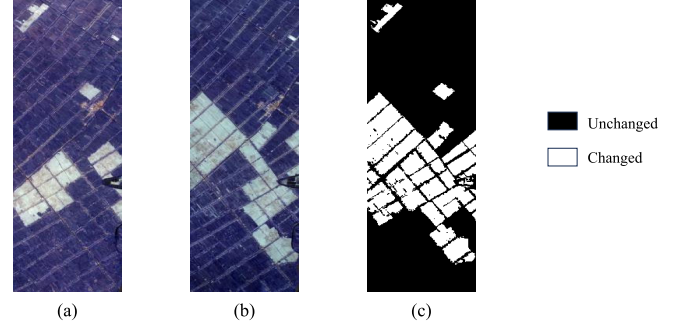


Fig. 8. Farmland dataset. (a) Pseudo-color composite image on T1. (b) Pseudo-color composite image on T2. (c) Ground truth.

2013 and 2015. The HSI contained in this dataset shows the distribution of buildings and crops in the area. Each HSI contains 224 spectral bands and the image size is  $600 \times 500$ .

**Santa Barbara:** The Santa Barbara dataset was collected by the AVIRIS sensor over the Santa Barbara area in California, USA, and contains two HSIs which reflect the land cover in 2013 and 2014. The spatial size of each HSI is  $984 \times 740$  pixels and contains 224 spectral bands.

**Farmland:** The Farmland dataset was collected by the Hyperion sensor of the Earth Observation (EO-1) in the wetland agricultural area of Yancheng city, Jiangsu, China, in 2006 and 2007. After noise processing and removal of water absorption bands, the dataset contains  $420 \times 140$  pixels and 155 spectral bands totally. The main changes in the dataset are related to changes in crop planting.

#### B. Evaluation Criteria

In this article, five commonly used metrics in CD task are applied as evaluation criteria for the performance of different models, which are overall accuracy (OA), kappa coefficient ( $\kappa$ ), precision ( $P$ ), recall ( $R$ ), and  $F1$  score ( $F1$ ), and each metric is defined as follows:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (22)$$

$$\kappa = \frac{OA - p_e}{1 - p_e} \quad (23)$$

$$P = \frac{TP}{TP + FP} \quad (24)$$

$$R = \frac{TP}{TP + FN} \quad (25)$$

$$F1 = \frac{2P \times R}{P + R} \quad (26)$$

TABLE II  
EXPERIMENT RESULTS ON BAY AREA DATASET

Method	SVM	ML-EDAN	MSDFFN	CSANet	GTMSiam	DIEFEN	DCFSL	BiG-FSLF	CAFSL
OA(%)	61.78	64.34	73.23	67.37	70.00	81.32	87.81	87.71	<b>92.49</b>
$\kappa \times 100$	25.95	28.26	47.15	36.86	41.02	61.97	75.37	75.31	<b>85.00</b>
P(%)	78.71	74.91	83.06	90.97	82.15	78.03	<b>92.48</b>	86.92	89.96
R(%)	39.03	63.50	62.83	43.37	57.01	90.91	86.00	90.58	<b>95.89</b>
F1(%)	52.19	64.22	71.53	58.72	66.88	83.96	89.01	88.42	<b>92.67</b>

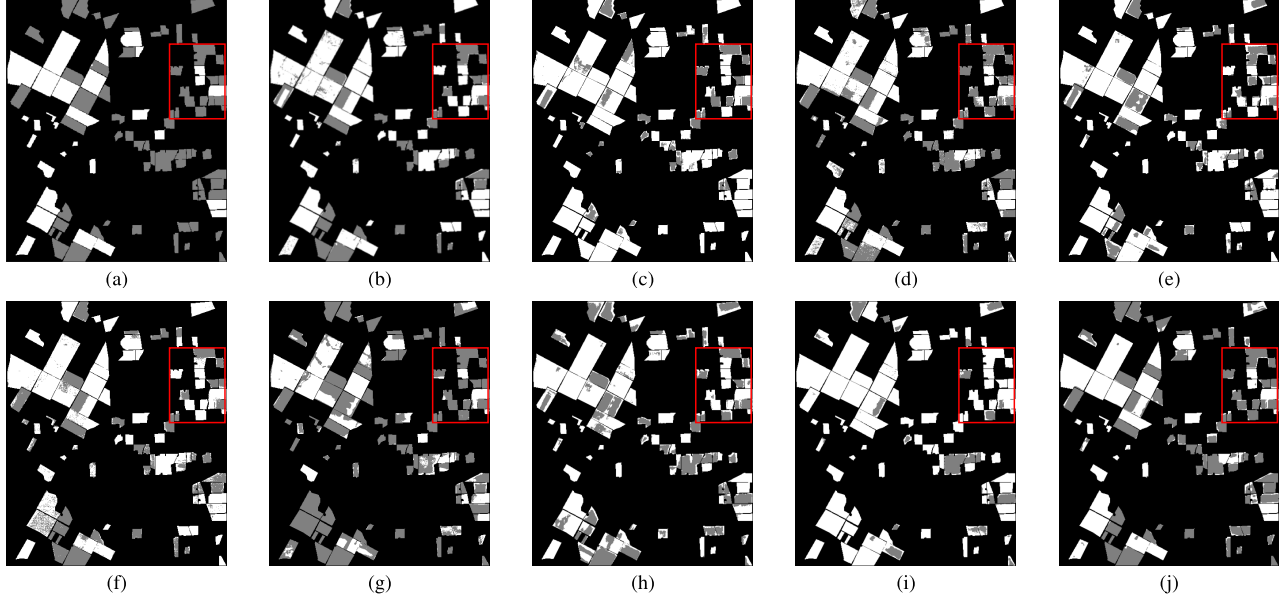


Fig. 9. Detection map on Bay Area. (a) Ground truth. (b) SVM (61.78%). (c) ML-EDAN (64.34%). (d) MSDFFN (73.23%). (e) CSANet (67.37%). (f) GTMSiam (70.00%). (g) DIEFEN (81.32%). (h) DCFSL (87.81%). (i) BiG-FSLF (87.71%). (j) CAFSL (92.49%).

$$p_e = \frac{(TP \cdot FN) + (TP \cdot FP) + (TN \cdot FN) + (TN \cdot FP)}{N^2} \quad (27)$$

where true positive (TP) refers to the number of positive examples correctly predicted by model; true negative (TN) refers to the number of negative examples correctly predicted by model; false positive (FP) refers to the number of positive examples incorrectly predicted by model, and false negative (FN) refers to the number of negative examples incorrectly predicted by model. The larger the value of the above five metrics, the better the result of the model.

### C. Experiment Settings and Comparison Algorithms

The CAFSL algorithm proposed in this article and all comparison methods use a device equipped with a CPU of 2.70 GHz, 128GB of RAM, and NVIDIA GeForce RTX 2080Ti GPU, and the operating environment is Python 3.6. In CAFSL algorithm, the patch size is set at  $9 \times 9$  and the learning rate is set at 0.01.

We selected eight algorithms for comparison, namely SVM [56], multilevel encoder-decoder attention network (ML-EDAN) [57], MSDFFN [28], cross-temporal interaction symmetric attention network (CSANet) [58], differential

information-enhanced feature exchange network (DIEFEN) [59], gated transmitting-based multiscale siamese network (GTMSiam) [60], DCFSL [41], and BiG-FSLF [44]. Among them, SVM is a traditional algorithm; ML-EDAN, MSDFFN, CSANet, DIEFEN, and GTMSiam are DL-based methods for HSI-CD; and DCFSL and BiG-FSLF are FSL-based algorithms. All comparison algorithms use the original parameters of the article.

In the sample selection of target domain, we randomly selected five samples from the changed and unchanged categories for experiment, that is, our algorithm used a total of 10 labeled samples. To maintain the fairness and stability of experiments, we selected the same number of labeled samples for training in all comparison algorithms, and each algorithm is run 10 times to take the average value as the final results.

### D. Results on Three Datasets

1) *Bay Area Dataset*: The experiment results and visualizations of eight comparison algorithms and the proposed CAFSL algorithm on Bay Area are shown in Table II and Fig. 9, respectively. As shown in Table II, our CAFSL achieves the best results on OA, which is at least 4% higher compared to



TABLE III  
EXPERIMENT RESULTS ON SANTA BARBARA DATASET

Method	SVM	ML-EDAN	MSDFFN	CSANet	GTMSiam	DIEFEN	DCFSL	BiG-FSLF	CAFSL
OA(%)	64.66	74.88	81.72	85.12	81.88	88.02	91.03	92.39	<b>93.14</b>
$\kappa \times 100$	25.70	47.00	59.80	66.61	59.87	75.89	81.54	84.00	<b>85.58</b>
P(%)	55.17	70.26	85.97	<b>98.56</b>	89.01	78.09	92.71	89.60	90.75
R(%)	54.12	67.71	63.78	63.02	61.64	<b>96.64</b>	86.36	90.98	92.00
F1(%)	54.64	66.75	73.14	76.69	72.77	86.37	89.15	90.23	<b>91.20</b>

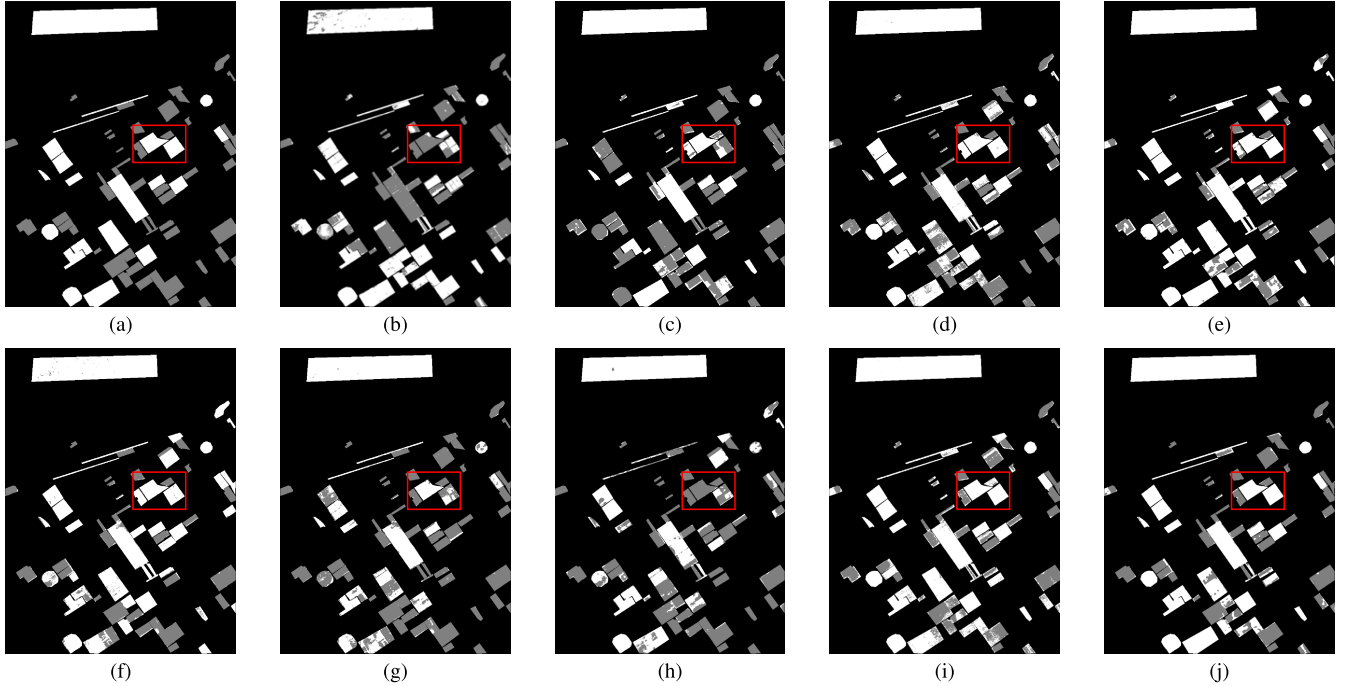


Fig. 10. Detection map on Santa Barbara. (a) Ground truth. (b) SVM (64.66%). (c) ML-EDAN (74.88%). (d) MSDFFN (81.72%). (e) CSANet (85.12%). (f) GTMSiam (81.88%). (g) DIEFEN (88.02%). (h) DCFSL (91.03%). (i) BiG-FSLF (92.39%). (j) CAFSL (93.14%).

other algorithms. It reflects the superiority and applicability of our algorithm on the HSI-CD task.

In addition, it is obvious from Table II that the FSL-based method performs better in all five metrics, and our algorithm is greatly improved compared to the current representative FSL-based algorithms, indicating that the INSCL module can enhance feature consistency among samples in the same domain effectively, while the CRCA module can achieve feature alignment in different domains. The synergy of two modules significantly enhances the model's performance in HSI-CD task.

Fig. 9 shows a visualization of results of nine algorithms on Bay Area dataset and the ground truth map, where gray represents areas predicted to be changed and white represents areas predicted to be unchanged. The red box indicates where our algorithm is significantly better than other algorithms. As shown in Fig. 9, due to the insufficient number of labels, the traditional algorithm and DL-based algorithm [i.e., Fig. 9(b)–(g)] are quite different from the ground truth and detect more noise points. In addition, it can be seen from the red box area that our algorithm is basically consistent with

Fig. 9(a), which indicates that the CAFSL algorithm has a higher accuracy in detail areas which are difficult to detect, and also fully proves the excellent performance of our algorithm.

2) *Santa Barbara Dataset*: Experimental results of nine algorithms on Santa Barbara are shown in Table III. From Table III, the three FSL-based methods achieve good performances. However, our method improves on all metrics over BiG-FSLF, the second best algorithm on OA. Especially on  $\kappa$ , our algorithm has a 1.58% improvement. For algorithms based on DL, high accuracy is hard to obtain because of the scarcity of labels. Among them, CSANet has the highest  $P$  but lower  $R$ , which indicates that the algorithm has poor detection ability for changed samples and high false positive rate. In contrast, DIEFEN has high  $R$  and low  $P$ , which means that the network can detect most of changed areas, but at the same time it will also mark many unchanged areas as changed mistakenly, resulting in more false detections. The CAFSL algorithm we proposed achieved relatively balanced results in all five metrics, indicating that our algorithm is accurate in detecting both changed and unchanged samples, and also reflects the superior performance of our algorithm.

TABLE IV  
EXPERIMENT RESULTS (%) ON FARMLAND DATASET

Method	SVM	ML-EDAN	MSDFFN	CSANet	GTMSiam	DIEFEN	DCFSL	BiG-FSLF	CAFSL
OA(%)	82.32	88.45	88.93	85.92	85.71	86.36	92.86	92.99	<b>94.72</b>
$\kappa \times 100$	55.65	70.25	72.28	62.65	67.07	67.89	83.12	83.39	<b>87.28</b>
P(%)	71.90	86.21	83.12	86.05	73.99	73.59	91.92	91.75	<b>92.46</b>
R(%)	64.10	73.56	77.73	61.44	84.98	84.21	85.05	85.72	<b>89.83</b>
F1(%)	67.78	77.83	79.85	71.68	77.17	77.56	88.22	88.38	<b>91.01</b>

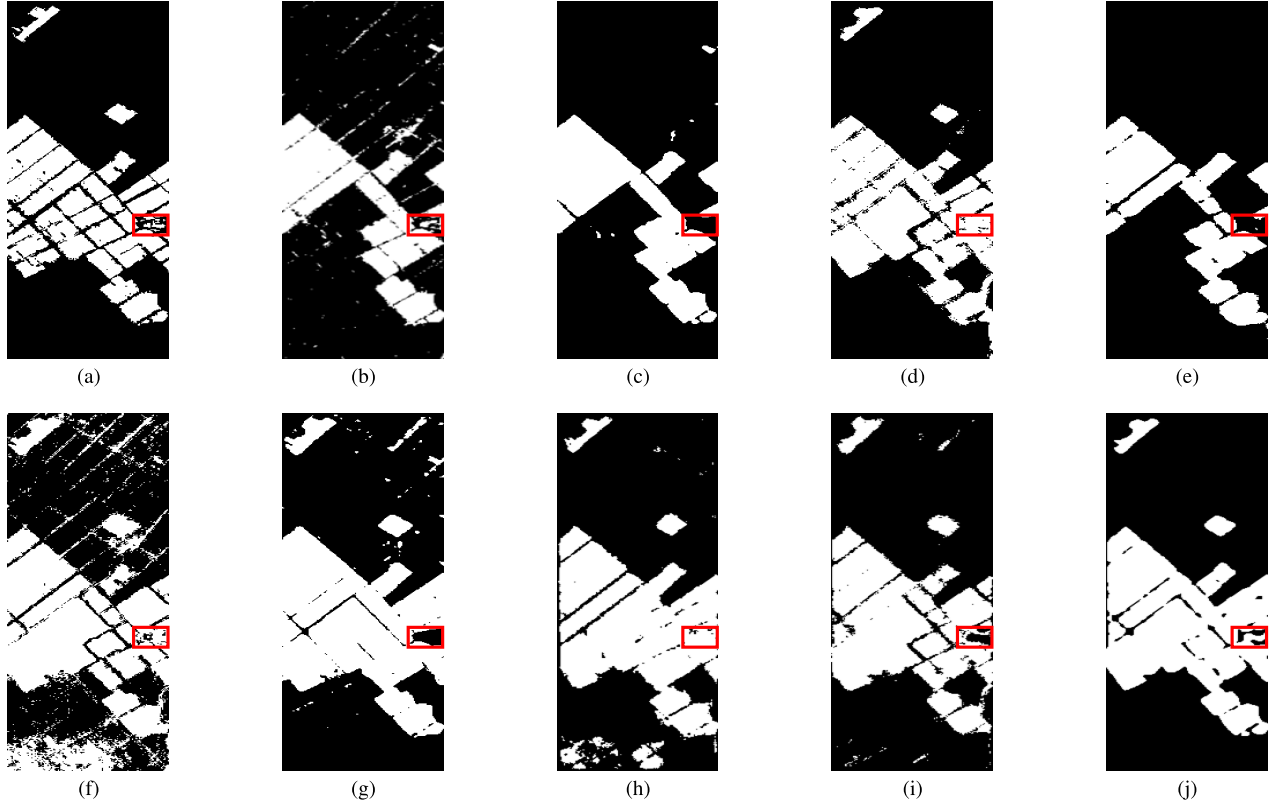


Fig. 11. Detection map on Farmland. (a) Ground truth. (b) SVM (82.32%). (c) ML-EDAN (88.45%). (d) MSDFFN (88.93%). (e) CSANet (85.92%). (f) GTMSiam (85.71%). (g) DIEFEN (86.36%). (h) DCFSL (92.86%). (i) BiG-FSLF (92.99%). (j) CAFSL (94.72%).

Fig. 10 shows the detection results of nine algorithms on Santa Barbara, where Fig. 10(a) shows the ground truth and the red box area is the best detection area for our algorithm. From Fig. 10, we can see that our experiment results [Fig. 10(j)] are closer to ground truth and with fewer detection noise points, which also verifies that the CAFSL we proposed performs better compared to other algorithms.

3) *Farmland Dataset*: Table IV and Fig. 11 show the detection results of nine algorithms on Farmland and the visual results, respectively. From Table IV, it is evident that our CAFSL gets the best results on all five metrics. Table IV shows that the OA of both traditional and DL-based algorithms is lower, while the OA of FSL-based method is relatively high, suggesting that the FSL strategy has a positive effect on HSI-CD task in the case of small samples. Compared with the second-best method BiG-FSLF, CAFSL improved by nearly 3% in *F1*, indicating that CAFSL has a good balance between

*P* and *R*, which can capture the changed areas comprehensively and adapt to CD tasks.

Fig. 11 shows the feature map of the Farmland dataset and the detection results of nine algorithms, where white is the area predicted to be changed and black is the area predicted to be unchanged. As shown in Fig. 11, due to the scarcity of labeled samples, all algorithms on Farmland dataset are more confusing for distinguishing the boundaries of changed and unchanged regions. However, compared with the FSL-based algorithms, the DL-based methods have higher false detection and missed detection rates for changed regions which results in larger errors. Our algorithm is more close to the ground truth than the other two FSL-based algorithms, especially in the red boxed region, our method obtains better detection relatively in the area where the changed and unchanged categories are obviously confused.

4) *Ablation Experiments*: The main contributions of the proposed CAFSL are INSCL module and CRCA module.

TABLE V  
ABLATION EXPERIMENT RESULTS ON THREE DATASETS

Datasets	FSL	INSCL	CRCA	OA(%)	$\kappa \times 100$	P(%)	R(%)	F1(%)
Bay Area	✓			89.81	79.62	87.02	93.68	90.11
	✓	✓		91.77	83.49	<b>91.12</b>	93.48	92.24
	✓	✓	✓	<b>92.49</b>	<b>85.00</b>	89.96	<b>95.89</b>	<b>92.67</b>
Santa Barbara	✓			90.86	81.16	<b>92.70</b>	85.48	88.88
	✓	✓		92.03	83.44	92.25	88.43	90.07
	✓	✓	✓	<b>93.14</b>	<b>85.58</b>	90.75	<b>92.00</b>	<b>91.20</b>
Farmland	✓			94.13	85.64	89.87	<b>90.30</b>	89.73
	✓	✓		94.36	86.37	91.60	89.47	90.34
	✓	✓	✓	<b>94.72</b>	<b>87.28</b>	<b>92.46</b>	89.83	<b>91.01</b>

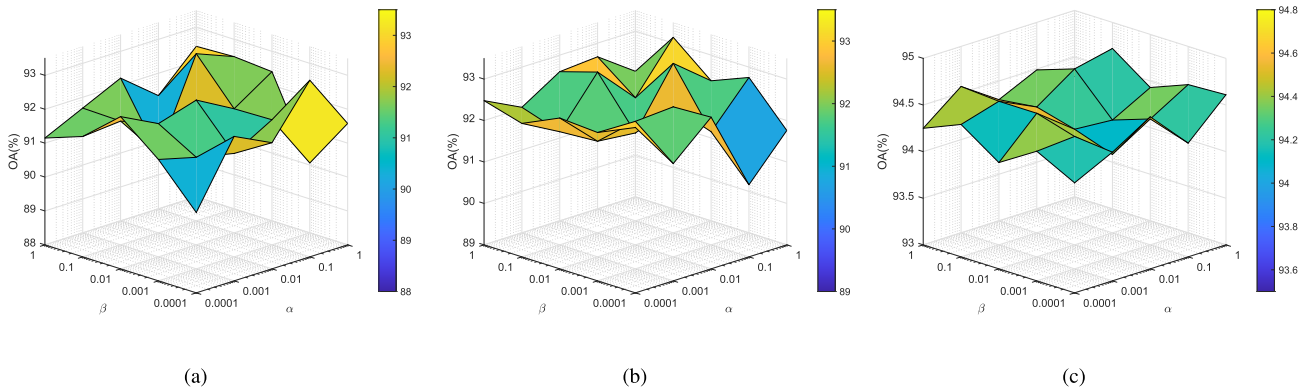


Fig. 12. OA (%) versus hyperparameter  $\alpha$  and  $\beta$  on three datasets. (a) Bay Area. (b) Santa Barbara. (c) Farmland.

To assess the effectiveness of each module, we performed ablation experiments on three datasets, respectively, where FSL is the baseline for the whole algorithm using only feature extractor and FSL strategy. After that, the INSCL module and CRCA module are added to baseline for intra-domain feature contrastive and inter-domain feature alignment sequentially. The overall experiment results are shown in Table V.

As can be seen from Table V, after joining INSCL to three datasets, OA has been improved to a certain extent, indicating that the INSCL module we designed is able to learn both common and divergent features of samples simultaneously, thereby promoting the compactness of intra-class samples and the discreteness of inter-class samples. On this basis, by adding the CRCA module, the OA and  $\kappa$  of our algorithm on all three datasets are significantly improved, indicating that the CRCA module can keep the feature distributions of source and target domains consistent, which reduces the domain bias effectively, and achieves the purpose of improving the generalization performance of target domain. In addition, in the final CAFSL model, although the  $P$  is slightly decreased on some datasets compared to without CRCA, the  $P$  and  $R$  values reach a balance, indicating that the model shows a good comprehensive ability to reduce FP and FN. On the other hand, it also shows that the CAFSL neither misses too many changed samples nor misjudges unchanged samples as changed too much, and the overall result is more satisfactory.

In conclusion, on three datasets, the proposed INSCL and CRCA modules have a positive effect on the original basis and achieve expected experiment results.

#### IV. DISCUSSION

##### A. Hyperparameter Analysis

The overall loss function of our proposed CAFSL contains two hyperparameters  $\alpha$  and  $\beta$ , which are intended to control the contribution of intra-domain CL and inter-domain feature alignment, respectively. We select  $\alpha$  and  $\beta$  in the set  $\{0.0001, 0.001, 0.01, 0.1, 1\}$  for experiments to analyze the impact of two parameters. The parameter analysis results on the three datasets are shown in Fig. 12.

As shown in Fig. 12(b) and (c), our CAFSL algorithm shows the best performance on Santa Barbara and Farmland datasets when  $\alpha = 1$  and  $\beta = 0.1$ . From (a) in Fig. 12, we can see that, when  $\alpha = 1$  and  $\beta = 0.1$ , although the performance of the CAFSL algorithm on Bay Area dataset is not the best, it still has a better performance. To maintain parameter consistency, we also set  $\alpha$  to 1 and  $\beta$  to 0.1 on all three datasets.

##### B. Effect of Different Number of Labeled Samples

In order to verify the effect of different labeled samples on the performance of algorithms, we take the Bay Area dataset as an example, select 3, 4, 5, 6, and 7 as the number

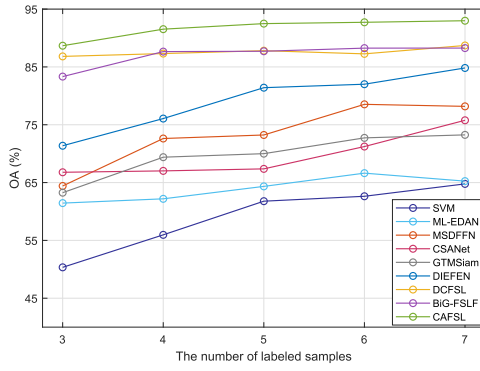


Fig. 13. OA varies with the number of labeled samples on Bay Area dataset.

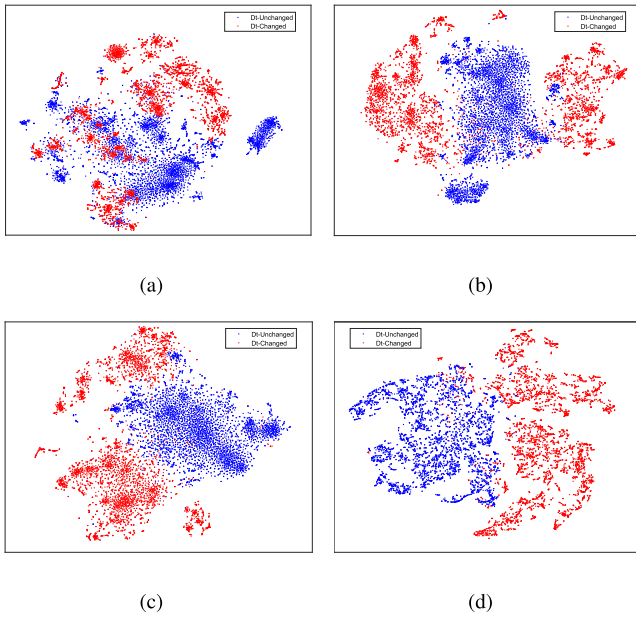


Fig. 14. Feature visualization of four algorithms on Bay Area dataset. (a) MSDFFN. (b) DCFSL. (c) BiG-FSLF. (d) CAFSL.

of labeled samples in target domain, respectively, to test the performance of nine algorithms. The experiment results are shown in Fig. 13. According to Fig. 13, the OA of nine algorithms increases with the number of samples basically. The CAFSL algorithm we proposed achieves the best performance under different labels, which also shows that our algorithm can adapt to a variety of small-sample situations and can better perform the HSI-CD task.

### C. Feature Visualization

The proposed CAFSL algorithm makes the features tighter between the same class and more dispersed between different classes through INSCL module and CRCA module. In order to illustrate the advantages of the CAFSL algorithm more intuitively, this section takes the Bay Area dataset as an example and use the t-distributed stochastic neighbor embedding (t-SNE) diagram to display the features of one representative DL-based method MSDFFN and three FSL-based algorithms (i.e., DCFSL, BiG-FSLF, and CAFSL). The results are shown in Fig. 14.

The blue dots in Fig. 14 indicate the feature distribution of the unchanged category and the red dots indicate the changed category. Since the features of unchanged pixels are easier to detect, it can be seen from Fig. 14 that the four algorithms are basically effective in aggregating the features of unchanged pixels. However, as shown in Fig. 14(a), the DL-based algorithm MSDFFN cannot separate the changed and unchanged samples well due to the limited labels. At the same time, from Fig. 14(b) and (c), we can see that the red dots are distributed in a rather chaotic manner, which indicates that the other three methods cannot aggregate the changed features well, while the CAFSL can effectively aggregate the changed features and separate them from the unchanged features, indicating that the module we designed can improve the inter-class separability of samples in the feature space.

### V. CONCLUSION

To solve the problem of insufficient number of HSI labels, this article proposes an algorithm based on FSL named CAFSL for small-sample HSI-CD tasks. The CAFSL model utilizes the VHRI of source domain to assist the CD task of the HSI of target domain and integrates CL into the FSL framework. After feature extraction for source and target domains, respectively, the distance between the support set prototype and query set is calculated to obtain the corresponding FSL loss. Subsequently, an INSCL module based on CL is designed to make the intra-class samples in the same domain closer and the inter-class samples more dispersed. At the same time, to mitigate the domain offset caused by the different data types of the source and target domains, a CRCA module is used to align the features of the source and target domains. Experiment results on three datasets show that our algorithm obtains better detection results while using fewer labeled samples and achieves the desired goals. In future research, we can consider expanding the source domain to multimodal data and using CL to improve the cross-modal feature fusion capability.

### REFERENCES

- [1] A. Plaza et al., "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, Apr. 2009.
- [2] L. Zhang and L. Zhang, "Artificial intelligence for remote sensing data analysis: A review of challenges and opportunities," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 2, pp. 270–294, Jun. 2022.
- [3] B. Tu, X. Yang, B. He, Y. Chen, J. Li, and A. Plaza, "Anomaly detection in hyperspectral images using adaptive graph frequency location," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–15, 2024.
- [4] G. Cheng et al., "Change detection methods for remote sensing in the last decade: A comprehensive review," *Remote Sens.*, vol. 16, no. 13, p. 2355, Jun. 2024.
- [5] S. Liu, L. Bruzzone, F. Bovolo, M. Zanetti, and P. Du, "Sequential spectral change vector analysis for iteratively discovering and detecting multiple changes in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4363–4378, Aug. 2015.
- [6] A. Schaum and A. Stocker, "Hyperspectral change detection and supervised matched filtering based on covariance equalization," *Proc. SPIE*, vol. 5425, pp. 77–90, Aug. 2004.
- [7] W. Zhao, L. Mou, J. Chen, Y. Bo, and W. J. Emery, "Incorporating metric learning and adversarial network for seasonal invariant change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2720–2731, Apr. 2020.
- [8] Z. Lv, M. Zhang, W. Sun, J. A. Benediktsson, T. Lei, and N. Falco, "Spatial-contextual information utilization framework for land cover change detection with hyperspectral remote sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4411911.



- [9] F. Ye, Z. Wu, X. Jia, J. Chanussot, Y. Xu, and Z. Wei, "Bayesian nonlocal patch tensor factorization for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 32, pp. 5877–5892, 2023.
- [10] D. Wen et al., "Change detection from very-high-spatial-resolution optical remote sensing images: Methods, applications, and future directions," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 68–101, Dec. 2021.
- [11] G. Tagliabue et al., "Hybrid retrieval of crop traits from multi-temporal PRISMA hyperspectral imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 187, pp. 362–377, May 2022.
- [12] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised deep change vector analysis for multiple-change detection in VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3677–3693, Jun. 2019.
- [13] Y. Huang, X. Li, Z. Du, and H. Shen, "Spatiotemporal enhancement and interlevel fusion network for remote sensing images change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5609414.
- [14] L. Zhang, L. Song, B. Du, and Y. Zhang, "Nonlocal low-rank tensor completion for visual data," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 673–685, Feb. 2021.
- [15] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Change detection techniques," *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, 2004.
- [16] W. A. Malila, "Change vector analysis: An approach for detecting forest changes with Landsat," in *Proc. LARS Symposia*, Jan. 1980, p. 385.
- [17] G. Licciardi, P. R. Marpu, J. Chanussot, and J. A. Benediktsson, "Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 447–451, May 2012.
- [18] J. S. Deng, K. Wang, Y. H. Deng, and G. J. Qi, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *Int. J. Remote Sens.*, vol. 29, no. 16, pp. 4823–4838, Aug. 2008.
- [19] V. Ortiz-Rivera, M. Vélez-Reyes, and B. Roysam, "Change detection in hyperspectral imagery using temporal principal components," *Proc. SPIE*, vol. 6233, pp. 368–377, May 2006.
- [20] A. A. Nielsen, "The regularized iteratively reweighted mad method for change detection in multi-and hyperspectral data," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 463–478, Feb. 2007.
- [21] K. C. Seto, C. E. Woodcock, C. Song, X. Huang, J. Lu, and R. K. Kaufmann, "Monitoring land-use change in the pearl river delta using Landsat TM," *Int. J. Remote Sens.*, vol. 23, no. 10, pp. 1985–2004, Jan. 2002.
- [22] F. Bovolo, L. Bruzzone, and M. Marconcini, "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2070–2082, Jul. 2008.
- [23] A. Voulodimos et al., "Deep learning for computer vision: A brief review," *Comput. Intel. Neurosci.*, vol. 2018, no. 1, 2018, Art. no. 7068349.
- [24] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.
- [25] B. Yang, S. Pan, W. Sun, Z. Guo, Z. Ye, and J. Peng, "FCFDA: Fine-coarse-fine progressive graph framework with distribution alignment for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5538714.
- [26] L. Wu, J. Peng, B. Yang, W. Sun, and Z. Ye, "AIWSEN: Adaptive information weighting and synchronized enhancement network for hyperspectral change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5503412.
- [27] X. Ou, L. Liu, B. Tu, G. Zhang, and Z. Xu, "A CNN framework with slow-fast band selection and feature fusion grouping for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5524716.
- [28] F. Luo, T. Zhou, J. Liu, T. Guo, X. Gong, and J. Ren, "Multiscale diff-changed feature fusion network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5502713.
- [29] M. Hu, C. Wu, and L. Zhang, "HyperNet: Self-supervised hyperspectral spatial-spectral feature understanding network for hyperspectral change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5543017.
- [30] B. Yang, W. Sun, and J. Peng, "SAGN: Sharpening-aware graph network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5518812.
- [31] J. Ding, X. Li, S. Xiang, and S. Chen, "Multilevel features fused and change information enhanced neural network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2023, Art. no. 5502413.
- [32] H. Yu, H. Yang, L. Gao, J. Hu, A. Plaza, and B. Zhang, "Hyperspectral image change detection based on gated spectral-spatial-temporal attention network with spectral similarity filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5511313.
- [33] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 29, Dec. 2016, pp. 3637–3645.
- [34] V. G. Satorras and J. B. Estrach, "Few-shot learning with graph neural networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Feb. 2018, pp. 1–13.
- [35] Q. Sun, Y. Liu, T.-S. Chua, and B. Schiele, "Meta-transfer learning for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 403–412.
- [36] S. Ravi and H. Larochelle, "Optimization as a model for few-shot learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Mar. 2017, pp. 1–11.
- [37] L. Yang, Y. Li, J. Wang, and N. N. Xiong, "FSLM: An intelligent few-shot learning model based on Siamese networks for IoT technology," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 9717–9729, Jun. 2021.
- [38] P. Zhang, C. Liu, X. Chang, Y. Li, and M. Li, "Metric-based meta-learning model for few-shot PolSAR image terrain classification," in *Proc. CIE Int. Conf. Radar (Radar)*, Dec. 2021, pp. 2529–2533.
- [39] R. Yang, X. Xu, X. Li, L. Wang, and F. Pu, "Learning relation by graph neural network for SAR image few-shot learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 1743–1746.
- [40] M. Andrychowicz et al., "Learning to learn by gradient descent by gradient descent," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Jan. 2016, pp. 1–9.
- [41] Z. Li, M. Liu, Y. Chen, Y. Xu, W. Li, and Q. Du, "Deep cross-domain few-shot learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5501618.
- [42] Q. Liu et al., "Refined prototypical contrastive learning for few-shot hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5506214.
- [43] M. Cao, X. Zhang, J. Cheng, G. Zhao, W. Li, and X. Dong, "Spatial-spectral-semantic cross-domain few-shot learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5525315.
- [44] X. Wang, S. Li, X. Zhao, and K. Zhao, "BiG-FSLF: A cross heterogeneous domain few-shot learning framework based on bidirectional generation for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5516213.
- [45] B. Song et al., "3DCNN-NF: Few-shot hyperspectral image change detection based on 3-D convolution neural network and normalizing flow," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5530615.
- [46] J. Peng, Y. Huang, W. Sun, N. Chen, Y. Ning, and Q. Du, "Domain adaptation in remote sensing image classification: A survey," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9842–9859, 2022.
- [47] Y. Huang et al., "Two-branch attention adversarial domain adaptation network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5540813.
- [48] Y. Zhang, M. Zhang, W. Li, S. Wang, and R. Tao, "Language-aware domain generalization network for cross-scene hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5501312.
- [49] Z. Fang et al., "Confident learning-based domain adaptation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5527116.
- [50] J. Feng et al., "Class-aligned and class-balancing generative domain adaptation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5509617.
- [51] X. Ou, L. Liu, S. Tan, G. Zhang, W. Li, and B. Tu, "A hyperspectral image change detection framework with self-supervised contrastive learning pretrained model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7724–7740, 2022.
- [52] X. Wu, P. Gamba, J. Feng, R. Shang, X. Zhang, and L. Jiao, "A multitask framework for hyperspectral change detection and band reweighting with unbalanced contrastive learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5530213.

- [53] L. Samaniego, A. Bardossy, and K. Schulz, "Supervised classification of remotely sensed imagery using a modified  $k$ -NN technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2112–2125, Jul. 2008.
- [54] P. Khosla et al., "Supervised contrastive learning," in *Proc. NIPS*, 2020, pp. 18661–18673.
- [55] Y. Ning, J. Peng, Q. Liu, Y. Huang, W. Sun, and Q. Du, "Contrastive learning based on category matching for domain adaptation in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5301814.
- [56] T. Habib, J. Inglada, G. Mercier, and J. Chanussot, "Support vector reduction in SVM algorithm for abrupt change detection in remote sensing," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 3, pp. 606–610, Jul. 2009.
- [57] J. Qu, S. Hou, W. Dong, Y. Li, and W. Xie, "A multilevel encoder-decoder attention network for change detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518113.
- [58] R. Song, W. Ni, W. Cheng, and X. Wang, "CSANet: Cross-temporal interaction symmetric attention network for hyperspectral image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6010105.
- [59] L. Wu et al., "DIEFEN: Differential information-enhanced feature exchange network for hyperspectral change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5533312.
- [60] X. Wang, K. Zhao, X. Zhao, and S. Li, "GTMSiam: Gated transmitting-based multiscale Siamese network for hyperspectral image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5511805.



**Bing Yang** received the B.S. degree in information and computing science from Henan University, Kaifeng, China, in 2015, and the Ph.D. degree in statistics from the Huazhong University of Science and Technology, Wuhan, China, in 2021.

He is currently a Lecturer at the College of Sciences, China Jiliang University, Hangzhou, China. His research interests include hyperspectral image processing, graph neural network, and deep learning.



**Lanxin Wu** received the B.S. degree in mathematics and applied mathematics from the Faculty of Mathematics and Statistics, Hubei University, Wuhan, China, in 2023, where he is currently pursuing the M.S. degree in applied mathematics.

His research interests focus on deep learning and remote sensing change detection.



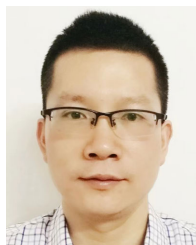
**Wenhui Hou** received the B.S. degree in mathematics and applied mathematics and the M.S. degree in applied mathematics from the Faculty of Mathematics and Statistics, Hubei University, Wuhan, China, in 2021 and 2024, respectively, where she is currently pursuing the Ph.D. degree in applied mathematics.

Her research interests include machine learning and hyperspectral image processing.



**Jiangtao Peng** (Senior Member, IEEE) received the B.S. degree in information and computing science and the M.S. degree in applied mathematics from Hubei University, Wuhan, China, in 2005 and 2008, respectively, and the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2011.

He is currently a Professor at the Faculty of Mathematics and Statistics, Hubei University. His research interests include machine learning and hyperspectral image processing.



**Weiwei Sun** (Senior Member, IEEE) received the B.S. degree in surveying and mapping and the Ph.D. degree in cartography and geographic information engineering from Tongji University, Shanghai, China, in 2007 and 2013, respectively.

From 2011 to 2012, he studied at the Department of Applied Mathematics, University of Maryland, College Park, MD, USA, working as a Visiting Scholar with the famous professor John Benedetto to study on the dimensionality reduction of hyperspectral image. From 2014 to 2016, he studied at the

State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China, working as a Post-Doctoral Researcher to study intelligent processing in hyperspectral imagery. From 2017 to 2018, he worked as a Visiting Scholar at the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. He is currently a Full Professor at Ningbo University, Ningbo, Zhejiang, China. He has published more than 70 journal articles. His research interests include hyperspectral image processing with manifold learning, anomaly detection, and target recognition of remote sensing imagery using compressive sensing.