

Easy-to-Hard Domain Adaptation With Human Interaction for Hyperspectral Image Classification

Cheng Zhang¹, Shengwei Zhong¹, Sheng Wan¹, and Chen Gong¹, *Senior Member, IEEE*

Abstract—In real-world hyperspectral image classification (HSIC), the limited annotated examples usually lead to an insufficiently trained classifier, which further generates low classification accuracy. To overcome this challenge, domain adaptation (DA) methods have been developed to transfer learnable knowledge from the external HSIs with sufficient labeled examples (i.e., the source domain) to the interested HSIs with scarce labeled examples (i.e., the target domain). Conventional DA approaches often pseudo-label the examples with high classification confidence and then incorporate them into the training process. However, due to the significant domain gap, relying solely on confident examples may not be adequate to achieve satisfactory performance. Therefore, this article proposes an interactive easy-to-hard DA method (IEH-DA) to arrange the adaptation process so that the “easy” examples are adapted ahead of the “hard” ones. In an early stage, the easy examples with high pseudo-labeling confidence are selected for the adversarial learning-based DA. In a later stage, the “hard” examples with high informativity are further selected, and they are interactively labeled by human experts to provide accurate supervision information for adaptation. As a result, the examples in the target domain are used in an easy-to-hard way, which forms a curriculum sequence for orderly model training. Extensive experiments conducted on typical public datasets demonstrate that IEH-DA outperforms other state-of-the-art DA methods for HSIC.

Index Terms—Curriculum learning (CL), domain adaptation (DA), example selection, hyperspectral image classification (HSIC).

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) are widely used in many remote-sensing applications such as smart agriculture [1], [2] target detection [3], [4], and environmental

Manuscript received 27 September 2023; revised 21 December 2023; accepted 22 January 2024. Date of publication 26 January 2024; date of current version 7 February 2024. This work was supported in part by the National Science Foundation (NSF) of China under Grant 12371510 and Grant 62101261; in part by the NSF for Distinguished Young Scholar of Jiangsu Province under Grant BK20220080; in part by the NSF of Jiangsu Province under Grant BK20210332; in part by the Fundamental Research Funds for the Central Universities under Grant 30920032202, Grant 30921013114, and Grant 30923011027; in part by the China Postdoctoral Science Foundation (CPSF) under Grant 2023M741708 and Grant 2023TQ0159; and in part by the Postdoctoral Fellowship Program of CPSF under Grant GZC20233503. (Corresponding authors: Shengwei Zhong; Sheng Wan.)

The authors are with the PCA Laboratory, the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, the Jiangsu Key Laboratory of Image and Video Understanding for Social Security, and the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: 121106010702@njust.edu.cn; zhongsw_91@foxmail.com; wansheng315@hotmail.com; chen.gong@njust.edu.cn).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TGRS.2024.3358869>, provided by the authors.

Digital Object Identifier 10.1109/TGRS.2024.3358869

monitoring [5]. An HSI consists of hundreds of spectral bands [6], providing abundant spectral and spatial information for accurate target classification. However, the training of classifiers usually needs massive labeled examples (i.e., pixels), and manually labeling the examples in each HSI can be expensive and time-consuming. To overcome the challenge of requiring a large number of labeled examples, DA methods [7], [8] are proposed to adapt the classifier trained in the historical image to the currently investigated image. To achieve this goal, it is crucial to align the data distributions between the source domain (i.e., the historical image) and the target domain (i.e., the image under investigation). The objective is to transfer the knowledge gained from the labeled examples in the historical images to the current image, thereby reducing the need for extensive manual labeling. As one of the most common and challenging misalignments between HSIs, spectral shift often occurs in HSIs due to various factors such as atmospheric effects and variations in illumination conditions [9]. Therefore, we primarily focus on addressing the issue of spectral shift, which is similar to most existing DA methods.

Pseudo-labeling is a widely applied technique in DA, which involves assigning pseudo-labels to the unlabeled data in the target domain using a model initially trained on the source domain. It leverages the unlabeled data to improve the model performance on the target domain so that the model can be generalized across different domains. For example, Zhong and Zhang [10] proposed an iterative training method by using source data and pseudo-labeled target data, where the labels are generated by a trained classifier. Besides, pseudo-labeling techniques have also been investigated in adversarial DA approaches. Fang et al. [11] proposed confident learning-based domain adaptation (CLDA) to train deep neural networks by using high-confidence pseudo-labeled target examples.

Although pseudo-labeling examples in the target domain can provide additional supervised information to enhance the adaptation capability of the model [10], [11], [12], existing pseudo-labeling-based DA methods still face two challenges. The first challenge lies in the way to utilize the obtained pseudo-labels for domain distribution alignment, as most existing DA methods apply all the obtained pseudo-labels uniformly for domain distribution alignment, which overlooks their different certainty levels of correctness. Specifically, the pseudo-labels associated with the examples near the class boundary (i.e., the hard examples) tend to be less certain, whereas those at the class center (i.e., the easy examples) are usually more reliable. Therefore, equally treating all the

pseudo-labels for classification may result in substantial performance degradation. The second challenge is the difficulty in ensuring the label correctness of the hard examples due to the high uncertainty of their corresponding pseudo-labels. This issue will ultimately hamper the generalization capability of the model.

To address these challenges, we proposed a novel method called interactive easy-to-hard DA (IEH-DA), which forms an easy-to-hard learning paradigm for DA. Concretely, we first utilize easy examples with confident pseudo-labels for adversarial learning-based DA. In the following, we select the hard examples for model learning, since they can offer informative supervision signals that are essential for improving the model performance. Furthermore, to ensure reliable labels for the hard examples, IEH-DA obtains labels by incorporating the interaction with human experts. By this means, a curriculum sequence enhanced by reliable labels is formed for orderly model training. Considering the limitations of relying solely on a single type of adversarial training for DA, our IEH-DA proposes a hybrid adversarial learning objective, which simultaneously considers the feature discriminability and domain invariance. This can ensure a balanced and effective approach to the adaptation process. The main contributions of IEH-DA are listed as follows.

- 1) To the best of our knowledge, this is the first application of curriculum learning (CL) to DA problems in HSIs. By progressively leveraging pixels with increasing difficulty levels, the proposed method ensures a smooth and effective learning trajectory for the model, avoiding excessive difficulties during the initial training stage. Learning from these orderly curricula can enhance the capabilities of the DA model to generalize to unseen domains, which facilitates effective knowledge transfer across diverse HSIs.
- 2) To achieve robust DA, we devise a dual cross-domain adversarial learning approach that combines discriminator- and classifier-based adversarial learning techniques for domain alignment. Here, an additional classifier is designed to pseudo-label the target examples, ensuring domain alignment without compromising feature discriminability.
- 3) We have conducted comprehensive empirical studies using multiple public benchmark datasets under the commonly-used evaluation protocol. The results demonstrate that the proposed IEH-DA consistently outperforms the state-of-the-art methods.

II. RELATED WORK

A. Hyperspectral Image Classification

HSI classification (HSIC) emerges as a significant branch of remote sensing. The early-staged methods for HSIC are mostly based on conventional pattern recognition approaches, such as support vector machines (SVMs) [13] and K -nearest neighbor classifiers. However, these methods independently treat every pixel and fail to capture the spatial context of HSIs. To address this issue, edge-preserving filtering (EPF)-based approaches are widely used to take care of both spectral and spatial correlation. For example, Zhong et al. [14] introduced iterative EPF

to generate new image cubes by combining the currently processed image cube with soft probability maps obtained through EPF. Considering the challenges in the classification of small-sized categories, class feature weighted HSIC (CFW-HSIC) [15] introduced three types of class features to calculate the probability for each class to represent the corresponding class significance and enhance HSIC by extracting features from the classes of interest. With the advances in deep learning, convolutional neural networks (CNN) have gained significant attention and have been extensively employed in HSIC. For instance, Jia et al. [16] employed CNNs to extract spectral features and achieved good performance. To effectively exploit the contextual relations in HSIs, a multiscale dynamic graph convolutional network (MDGCN) [17] is devised to dynamically exploit the multiscale spatial information in HSIs, resulting in competitive performance, especially in boundary regions of class. Moreover, Wan et al. [18] captured long-range contextual relations in HSIs by performing successive graph convolution on a learnable region-induced graph, which contributes to accurate classification. In addition, Yu et al. [20] proposed a GCN model with contrastive learning to explore the supervision signals contained in both spectral information and spatial relations. Recently, transformer-based methods have attracted a lot of attention. Hong et al. [20] proposed SpectralFormer to learn spectrally local sequence information from neighboring bands of HSIs and generate groupwise spectral embeddings, to effectively mine and represent the sequence attributes of spectral signatures. Besides, He et al. [21] combined HSIC with the bidirectional encoder representation from transformers (BERT) [22], which effectively addresses the challenges of limited receptive field, inflexibility, and poor generalizability in HSIC.

B. Domain Adaptation

To transfer knowledge from the source domain with rich supervision information to the target domain with insufficient supervision information, DA [7] is proposed. Recent mainstream DA methods can be broadly categorized into classifier-based methods [23] and feature-based methods [24]. In classifier-based DA methods, the objective is to train a suitable classifier using the data of the source domain, enabling the model to make predictions on the target domain. For example, Sun et al. [25] proposed a semisupervised approach based on multikernel SVMs to overcome the distribution bias between domains. Besides, based on a generalization bound derived from prior knowledge, Mansour and Schain [26] trained a robust SVM for binary classification and regression DA algorithms. Meanwhile, Ma et al. [27] used adaptive ranking SVMs to learn distance models. By relaxing the constraints on the necessary conditions, a discriminative model with high confidence in target positive mean and low confidence in target negative image pairs is trained. A key challenge in classifier-based DA methods is the difficulty in identifying parameters that are independent of the domain, which is important for the model to generalize effectively across various domains.

Unlike the classifier-based methods, feature-based DA approaches have been closely integrated with deep learning

and gained significant popularity. These algorithms focus on extracting domain-invariant feature representations and achieve DA through feature matching or adversarial learning techniques. Feature-matching methods aim to minimize the discrepancy between feature distributions by utilizing various measures of domain differences. For example, Long et al. [28] introduced the deep adaptation network (DAN) to extract the features from the convolutional layer by minimizing a multiple kernel maximum mean discrepancy (MMD). In addition, Zellinger et al. [29] employed the minimization of central moment discrepancy (CMD) to align the high-order moments in the distribution of the two domains, therefore facilitating knowledge transfer between domains. In recent years, generative adversarial networks (GANs) [30] have emerged as a powerful tool to boost the performance of feature-based DA approaches. By leveraging the adversarial training paradigm, GAN-based approaches aim to learn a mapping between the source domain and target domain which improves the transferability of learned representations. Ganin and Lempitsky [31] proposed the domain adversarial neural network (DANN) that incorporated a gradient reversal layer to facilitate adversarial learning. Meanwhile, Pei et al. [32] proposed multiadversarial DA (MADA) that employs multiple class domain discriminators to align the data distribution of the two domains. In [33], classifiers were introduced as a replacement for the domain discriminator, addressing the issue of deteriorated feature discriminability commonly observed in standard domain discriminator-based adversarial learning. Although feature-based DA methods are effective in extracting domain-invariant features, they could struggle with the complexity of feature space alignment and may not ensure the transferability of these features to new unseen domains.

C. Curriculum Learning

In 2009, Bengio et al. [34] proposed CL, a learning approach that aims to imitate the human learning process, which can be viewed as a special kind of continuation method. CL focuses on gradually transitioning from simpler to more challenging problems during the training process, mimicking the way humans learn. By initially optimizing smoother problems and gradually increasing the difficulty, CL assists the model to find the global minimum and improve generalization performance [35]. For instance, multimodal CL (MMCL) [36] determined the optimal curriculum sequence by evaluating the reliability and the discriminability of images from the viewpoints of multiple modalities. Besides, by comprehensively investigating both the individuality and commonality of different modalities, Gong [37] used a row-sparse matrix and a sparse noise matrix to explore the common preferences and the distinct opinions of different modalities on curriculum difficulty, respectively, and performed the “soft” fusion of multiple curricula from different modalities. With the rise of graph machine learning, a graph-based CL method [38] has been proposed, assuming that different unlabeled examples have different levels of difficulty for propagation, and the label learning should adhere to the easy-to-hard sequence.

In addition, there have been successful combination ways of CL with HSIC. For example, Luo et al. [39] proposed

a small data learning framework (SDLF) that introduces a dynamic training sample amplification (TSA) under a unified CL strategy to generate reliable labels for training. Furthermore, Wei et al. [40] designed a binary neural network (BNN) and utilized a CL-based progressive binarization strategy to improve the classification performance. Besides, Yang et al. [41] proposed self-paced learning-based probability subspace projection (SL-PSP) to develop two regularizers from a self-paced maximum margin and a probability label graph, respectively. This approach accurately revealed the affinity between mixed pixels and achieved good performance with very few labeled examples.

Recently, CL has also been leveraged in DA. For example, Choi et al. [42] employed CL to gradually incorporate local and global information for cross-domain remote sensing image segmentation. Analogously, Zhang et al. [43] proposed curriculum-style cross-DA (CCDA) that uses an entropy-based ranking strategy to measure example difficulty and performed the adaptation process by utilizing easy-to-hard examples and their corresponding pseudo-labels. Despite the advancements achieved by current CL methods in the field of remote sensing, the neglect of the label correctness of the examples in the target domain could potentially degrade the model training process.

D. Active Learning

Active learning is a classical interactive learning algorithm that acquires informative examples through various example selection strategies and then confirms the corresponding labels manually. It can provide rich information to gradually improve the generalization ability of the model [44]. The characteristics of active learning make it particularly effective for the design of hard curricula in CL, inspiring the difficulty measurer in CL. However, it should be noted that active learning primarily focuses on the information richness of each example, while CL emphasizes a sequential progression from easy to hard curricula. Recently, active learning has been applied to HSIC tasks. For example, Ye et al. [45] proposed a graph convolutional network (GCN)-based active learning method, which extracts both global as well as local graph-based features to gauge the discriminative information in unlabeled examples. In addition, Li et al. [46] proposed an active-learning-based prototypical network (ALPN) that utilizes the prototypical network to extract representative features and integrates semisupervised clustering with active learning methods to actively select valuable examples. Analogously, Li et al. [47] proposed a novel subpixel-pixel-superpixel-based multiview active learning method to reflect various characteristics of HSIs and contribute to good identification abilities. Besides, active learning has been proven to be effective when combined with DA techniques [48]. For instance, Lin et al. [49] introduced an active learning process to initialize the salient examples in HSI data and utilized higher-level features to overcome the cross-domain disparities.

III. METHOD

In this section, detailed information on our proposed IEH-DA method is provided. Our method focuses on

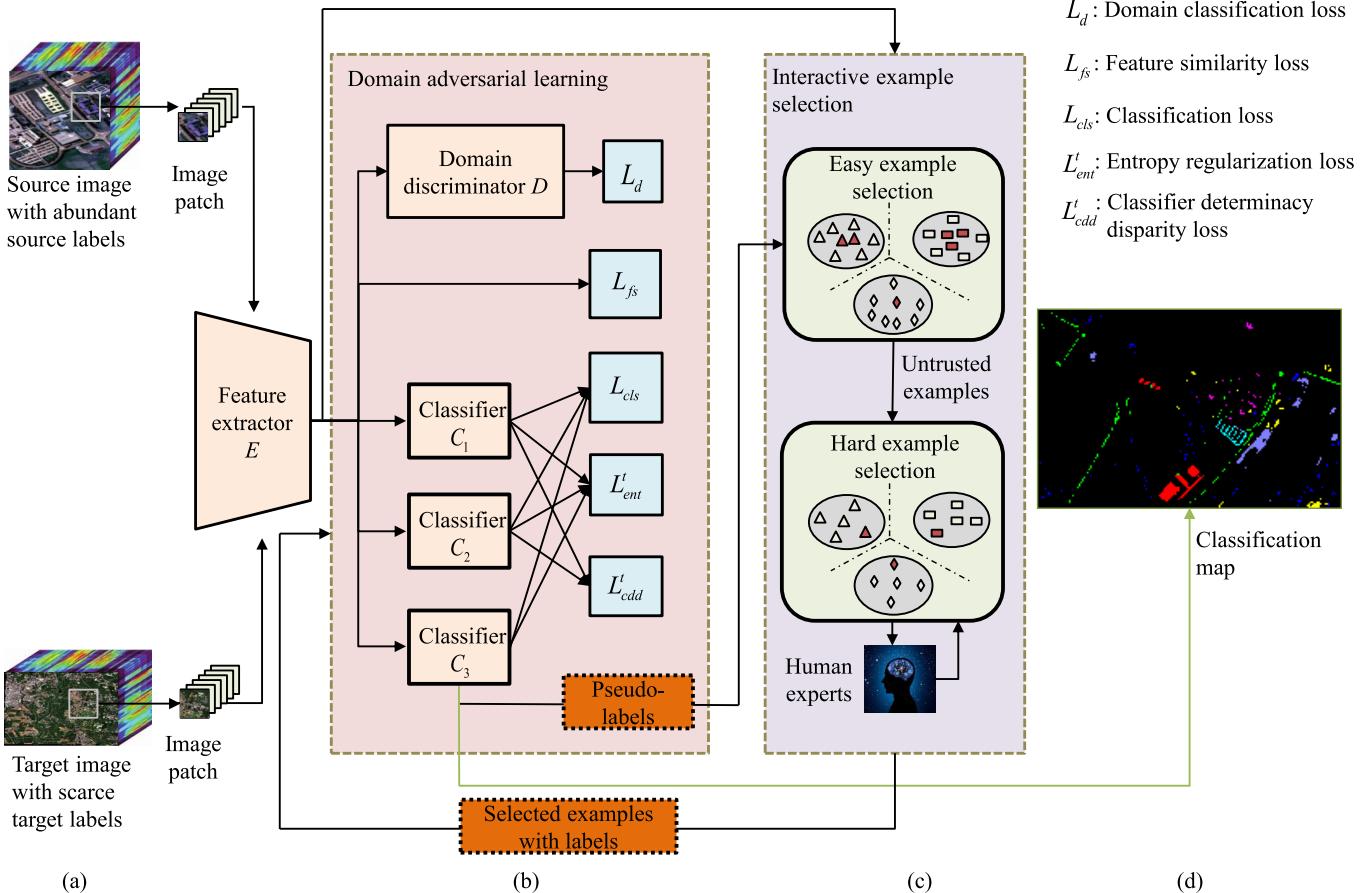


Fig. 1. Framework of our proposed IEH-DA. (a) Source image and the target image. (b) Dual cross-domain adversarial learning is performed by jointly optimizing the feature extractor, the domain discriminator, and the three classifiers. The adversarial learning between the feature extractor and domain discriminator is performed to obtain domain-invariance of the extracted features. Besides, the adversarial learning between the feature extractor and the first two classifiers (namely, C_1 and C_2) aims to enhance the feature discriminability. The third classifier (i.e., C_3) updates the parameters synchronously with the feature extractor to overcome the negative impacts on the classifiers caused by the biclassifier adversarial learning. The well-trained classifier C_3 is applied for pseudo-label generation and target-domain classification. (c) Interactive example selection module consists of easy example selection and hard example selection. (d) Classification map is obtained from classifier C_3 .

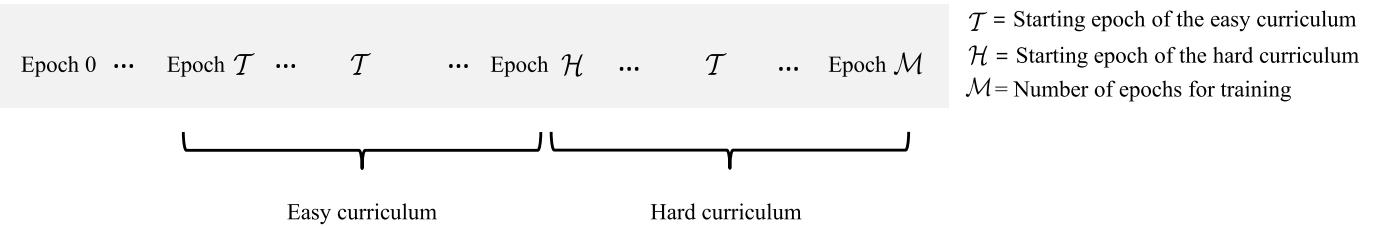


Fig. 2. Model training with an easy-to-hard curriculum sequence. \mathcal{T} is the epoch number to start the easy curriculum, and our model updates easy examples every \mathcal{T} epochs. \mathcal{H} indicates the epoch number to start the hard curriculum, and the selected hard examples will not be updated. \mathcal{M} represents the total number of epochs for the training cycle.

addressing the semisupervised DA problems. As shown in Figs. 1 and 2, IEH-DA consists of three modules: domain adversarial learning, interactive example selection, and model training with an easy-to-hard curriculum sequence. Here, domain adversarial learning aims to generate features that exhibit both class-discriminativeness and domain-invariance, which is achieved by jointly optimizing the feature extractor, the domain discriminator, and three classifiers. In the initial training stage, abundant labeled examples from the source image and scarce labeled examples from the target image are exploited for model training. Afterward, only the “easy”

examples with confident pseudo-labels from the target image are utilized for model training, which is considered the “easy curriculum” for the domain adversarial learning module. After learning with the easy curriculum, a few informative examples, referred to as “hard” examples, are selected with the labels given by experts. On this basis, the domain adversarial learning is finally performed with both easy and hard examples, forming the “hard curriculum.” By implementing the easy-to-hard curriculum training, the classification performance of our proposed method can be improved steadily in a progressive manner.

A. Problem Formulation

In the semisupervised cross-domain classification problem, the set of examples from the source domain is defined as $\mathbf{X}_s = \{\mathbf{x}_i^s\}_{i=1}^{n_s}$, where \mathbf{x}_i^s represents the i th example from the source domain and n_s is the number of labeled source examples. The corresponding labels for the source-domain examples are denoted as $\mathbf{Y}_s = \{\mathbf{y}_i^s\}_{i=1}^{n_s}$, where \mathbf{y}_i^s represents the label for the i th example. Similarly, the labeled target domain example set is defined as $\mathbf{X}_t = \{\mathbf{x}_i^t\}_{i=1}^{n_t}$, where \mathbf{x}_i^t represents the i th example from the target domain and n_t is the number of labeled target examples. The corresponding labels for the target-domain examples are denoted as $\mathbf{Y}_t = \{\mathbf{y}_i^t\}_{i=1}^{n_t}$, where \mathbf{y}_i^t represents the label for the i th example. The complete set of target domain examples including both labeled and unlabeled target examples, is denoted as \mathbf{X}_{ut} . It is important to note that the number of labeled source examples, that is, n_s , is significantly greater than the number of labeled target examples, that is, n_t . In our method, the number of shared land-cover categories between the source and target domains is assumed to be K . The objective of our method is to leverage the labeled source examples, labeled target examples, and unlabeled target examples, to improve the classification performance for the target domain.

B. Dual-Cross-Domain Adversarial Learning

This module consists of a feature extractor E , three classifiers C_1 , C_2 , and C_3 , and a domain discriminator D , as depicted in Fig. 1. Within this module, HSI patches from both the source domain and the target domain are individually fed into the feature extractor E . The extracted features are then fed into the three classifiers C_1 , C_2 , C_3 , and the domain discriminator D , respectively. In this module, adversarial learning is conducted from two perspectives, namely classifier-based and discriminator-based adversarial learning. To be specific, classifiers C_1 and C_2 fight against E to generate transferable and discriminative features. Meanwhile, the domain discriminator D aims at confusing E by obfuscating domain information, to align the distributions of the source domain and the target domain. Classifier C_3 is employed to overcome the defects caused by classifiers C_1 and C_2 in adversarial learning. Subsequently, we will provide a detailed introduction to classifier-based adversarial learning and discriminator-based adversarial learning, respectively.

1) *Classifier-Based Adversarial Learning*: Inspired by biclassifier determinacy maximization [33], the objective of classifier-based adversarial learning is to encourage C_1 and C_2 to produce different predictions on shared examples, thus exploring a diverse output space. To measure the divergence of classifiers, classifier determinacy disparity (CDD) [33] is employed, which can be defined as follows:

$$L_{cdd}^t(\mathbf{X}_t) = \frac{1}{n_{ut}} \sum_{i=1}^{n_{ut}} \left(\sum_{m,n=1}^K \mathbf{A}_{mn} - \sum_{m=1}^K \mathbf{A}_{mm} \right) \quad (1)$$

$$\mathbf{A} = \mathbf{p}_1 \mathbf{p}_2^T = C_1(E(\mathbf{x}_i^t)) C_2(E(\mathbf{x}_i^t))^T \quad (2)$$

where n_{ut} represents the number of all target-domain examples, the output of $C_j(\cdot)$ represents the category prediction of

the j th classifier, the output of $E(\cdot)$ corresponds to the features extracted by E , and the prediction correlation matrix \mathbf{A} is calculated by the prediction results of C_1 and C_2 , that is, \mathbf{p}_1 and \mathbf{p}_2 , after applying the softmax operation. Here, the prediction results satisfy the condition that $\sum_{k=1}^K \mathbf{p}_i^k = 1$, so $\mathbf{A}_{mn} = \mathbf{p}_1^m \mathbf{p}_2^n$ represents the probability that classifiers C_1 and C_2 assign an example to the m th and the n th categories, respectively.

2) *Discriminator-Based Adversarial Learning*: In addition to classifier-based adversarial learning, adversarial learning is also performed between the feature extractor E and the domain discriminator D . The objective is to blur the domain information present in the features. Specifically, the feature extractor E is trained to produce domain-invariant features that confuse the domain discriminator D , while D is simultaneously trained to distinguish the categories of features as accurately as possible. Here, the domain classification loss is used to compute the cost associated with misclassifying the feature domains, which can be expressed as

$$L_d(\mathbf{X}_s, \mathbf{X}_t) = \sum_{\mathbf{x}_i \in \mathbf{X}_s \cup \mathbf{X}_t} L_{cr}(D(E(\mathbf{x}_i)), \mathbf{y}_i^d) \quad (3)$$

where $D(\cdot)$ is the domain prediction result of the domain discriminator and \mathbf{y}_i^d is the domain label of \mathbf{x}_i . The parameters of the aforementioned feature extractor E and classifiers C_1 , C_2 , and C_3 can be updated by minimizing the source domain classification loss L_{cls}^s and the target domain classification loss L_{cls}^t , simultaneously

$$L_{cls}^s(\mathbf{X}_s, \mathbf{Y}_s) = \frac{1}{3n_s} \sum_{i=1}^{n_s} \sum_{j=1}^3 L_{cr}(C_j(E(\mathbf{x}_i^s)), \mathbf{y}_i^s) \quad (4)$$

$$L_{cls}^t(\mathbf{X}_t, \mathbf{Y}_t) = \frac{1}{3n_t} \sum_{i=1}^{n_t} \sum_{j=1}^3 L_{cr}(C_j(E(\mathbf{x}_i^t)), \mathbf{y}_i^t). \quad (5)$$

In addition to this, we utilize the feature similarity loss to minimize the discrepancy between the source features and target features from the same category

$$L_{fs}(\mathbf{X}_s, \mathbf{Y}_s, \mathbf{X}_t, \mathbf{Y}_t) = \sum_{y_i=y_j} L_f(\mathbf{x}_i, \mathbf{x}_j) - \sum_{y_i \neq y_j} L_f(\mathbf{x}_i, \mathbf{x}_j) \quad (6)$$

where $\sigma(\cdot)$ represents the sigmoid function, $L_{cr}(\cdot, \cdot)$ is the cross-entropy loss function, $E(\cdot)$ is the extracted features of examples, and $L_f(\mathbf{x}_i, \mathbf{x}_j) = L_{cr}(\sigma(E(\mathbf{x}_i)), \sigma(E(\mathbf{x}_j)))$. By using (6), the discriminability among the features belonging to different categories can be enhanced.

C. Interactive Example Selection for the Curriculum

It is noteworthy that some pixels in HSIs exhibit a mixture of spectral information from multiple categories [50], which correspond to hard examples near the class boundary. Incorporating the hard examples directly into the model training process can cause problems. Therefore, in the example selection stage, we employ the CL strategy by gradually introducing increasingly challenging examples for model training based on spectral characteristics of HSIs. This strategy ensures a smooth and effective learning trajectory for the model, avoiding overwhelming difficulties during the initial stages of training.

Meanwhile, it gradually enhances the model's capacity to handle sophisticated spectral characteristics. To be specific, examples with high classification confidence are first selected as easy examples, and the prediction results of classifier C_3 are taken as pseudo-labels. After a period of training, the remaining examples with the highest uncertainties are selected as hard examples and their labels are actively acquired from the domain experts. By incorporating hard examples into the training process, the model can benefit from the highly confident examples that provide meaningful supervision for model learning.

1) *Selection and Pseudo-Labeling of Easy Examples*: The easy examples are defined as the target examples with high classification confidence. Here, the classification confidence is calculated by estimating the joint probability distribution of the classification results obtained by two different classifiers. To be specific, an example is regarded as a confident example when the classification results produced by the two classifiers are consistent. In our method, classifier C_3 , together with an SVM classifier that is trained using the source features and the labeled target features produced by E , is employed for the adversarial learning module [11].

Suppose that \mathbf{p}_3^i is the predicted probability vector produced by classifier C_3 of the i th unlabeled example, \mathbf{x}_{ut}^i , and the according label is denoted as $y_{C_3}^i = \arg \max_m \mathbf{p}_{3,m}^i$, where $\mathbf{p}_{3,m}^i$ is the m th element in the vector \mathbf{p}_3^i and $i \in 1, \dots, n_{ut}$. Then $y_s^i \in \{1, \dots, K\}$ is the predicted label of \mathbf{x}_{ut}^i generated by the SVM, the confident joint matrix $\mathbf{C} \in \mathbb{N}^{K \times K}$ is a counting matrix, where the element in the m th row and the n th column of \mathbf{C} is defined as follows:

$$\mathbf{C}_{mn} := |\{\mathbf{x}_{ut}^i | y_{C_3}^i = m, y_s^i = n, \mathbf{p}_{3,m}^i \geq t_m\}| \quad (7)$$

where $|\cdot|$ is the number of elements of a set, and the threshold t_m is adaptively calculated as follows:

$$t_m = \frac{1}{|\{\mathbf{x}_{ut}^i | y_{C_3}^i = m\}|} \sum_{\{\mathbf{x}_{ut}^i | y_{C_3}^i = m\}} \mathbf{p}_{3,m}^i. \quad (8)$$

Afterward, the joint probability distribution matrix $\mathbf{Q} \in \mathbb{R}^{K \times K}$ can be obtained with weighted normalization based on the confident joint matrix \mathbf{C}

$$\mathbf{Q}_{mn} = \frac{\frac{\mathbf{C}_{mn}}{\sum_{n=1}^K \mathbf{C}_{mn}} |\{\mathbf{x}_{ut}^i | y_{C_3}^i = m\}|}{\sum_{m,n=1}^K \frac{\mathbf{C}_{mn}}{\sum_{n=1}^K \mathbf{C}_{mn}} |\{\mathbf{x}_{ut}^i | y_{C_3}^i = m\}|} \quad (9)$$

where \mathbf{Q}_{mn} is the element in the m th row and the n th column of \mathbf{Q} .

Based on the joint probability distribution matrix \mathbf{Q} , the prune by class (PBC) [51] method is used to prune untrusted examples in the target domain, to clean the target pseudo-labels. To be specific, in the m th category, all the unlabeled examples with the lowest values of $\mathbf{p}_{3,m}^i$ are selected as untrusted examples and pruned. The number of pruned examples is calculated as follows:

$$N_{\text{untrusted}}^m = n_{ut} \cdot \sum_{n=1, m \neq n}^K \mathbf{Q}_{mn}. \quad (10)$$

2) *Selection of Hard Examples*: In our proposed method, the hard examples are defined as those with the greatest amount of uncertainty in their classification results that are produced by classifier C_3 . These examples can be relatively informative for model learning if their labels can be determined by domain experts. Inspired by Rangwani et al. [52], the proposed method first finds perturbations that maximize the Kullback–Leibler (KL) divergence between the softmax outputs of the original example and its perturbed counterpart. The perturbations are then used to calculate the uncertainty score of each example according to the sensitivity to adversarial perturbations. To avoid the issue that most uncertain examples can be too similar to each other, a diversity score is further calculated to regulate the uncertainty score.

Specifically, for each pruned examples \mathbf{x} in the easy example selection procedure, the corresponding perturbation \mathbf{r}_i is created by

$$\mathbf{r}_i^* = \max_{\|\mathbf{r}_i\| \leq \epsilon} D_{KL}(C_3(E(\mathbf{x})) || C_3(E(\mathbf{x}_{r_i}))) \quad (11)$$

where $D_{KL}(\cdot || \cdot)$ is the KL divergence and $\mathbf{x}_{r_i} = \mathbf{x} + \mathbf{r}_i$ is the perturbed example. The optimal solution \mathbf{r}_i^* can be found using the Power Method [53] with a random initialization. In the multiclass classification scenario, R different types of perturbations are created, and then the pairwise mean KL divergence is calculated to measure the uncertainty of each example

$$\begin{aligned} \text{Uncertainty}(\mathbf{x}) &= \frac{1}{R^2} \left(\sum_{i=1}^R D_{KL}(C_3(E(\mathbf{x})) || C_3(E(\mathbf{x}_{r_i}))) \right. \\ &\quad \left. + \sum_{i=1}^R \sum_{j=1, i \neq j}^R D_{KL}(C_3(E(\mathbf{x}_{r_i})) || C_3(E(\mathbf{x}_{r_j}))) \right). \end{aligned} \quad (12)$$

Nevertheless, relying solely on the uncertainty score could potentially diminish the diversity of the examples. To ensure that the selected examples exhibit sufficient diversity and enhance the representation ability of the model, examples are iteratively scored and incorporated into the diversity dataset \mathbf{S} based on their similarity to the examples already present in \mathbf{S} . The following diversity measurement of a new example \mathbf{x}_n is used:

$$d(\mathbf{S}, \mathbf{x}_n) = \max_{\mathbf{x} \in \mathbf{S}} D_{KL}(C_3(E(\mathbf{x})) || C_3(E(\mathbf{x}_n))). \quad (13)$$

A higher score indicates lower similarity between the examples inside and outside \mathbf{S} , thus ensuring diversity across the categories of the selected examples. Taking into account both uncertainty and category diversity, the total score of the example \mathbf{x}_n is calculated as

$$\text{TotSco}(\mathbf{x}_n) = e \cdot \text{Uncertainty}(\mathbf{x}_n) + (1 - e) \cdot d(\mathbf{S}, \mathbf{x}_n) \quad (14)$$

where $e \in (0, 1)$ is the weight assigned to the uncertain score and set to 0.6 in the experiments. The top five examples with the highest TotSco are selected as hard examples, which will subsequently form the hard curriculum for model training along with easy examples.

3) *Easy-to-Hard Curriculum*: Training with an easy-to-hard curriculum sequence is an effective strategy that gradually increases the difficulty of model learning. Concretely, by initially exposing the model to easily comprehensible examples, CL enables the model to establish a strong foundational understanding. As the model becomes increasingly proficient and robust, the curriculum progressively introduces more challenging examples, gradually improving the model's performance. As illustrated in Fig. 2, the process mimics the gradual learning trajectory of humans, which begins with simpler questions and progressively advances to more complex concepts.

During the initial training phase, the performance of the classifiers could be suboptimal. Hence, we intend to enhance the capabilities of the classifiers via training them for T epochs, to generate high-quality pseudo-labels for the subsequent CL. After T epochs, easy examples \mathbf{X}_{ct} are selected from the target domain, which are pseudo-labeled as $\hat{\mathbf{Y}}_{ct}$. The labeled example set \mathbf{X}_t and \mathbf{Y}_t are updated by merging the original labeled target examples with the easy examples. The process is iteratively conducted every T epoch. Meanwhile, the loss of batch nuclear-norm maximization (BNM) [54] is used to enhance the diversity of classification results of each batch.

After \mathcal{H} epochs, the model is expected to be well-trained on the easy examples. On this basis, we begin to perform the hard example selection and label these examples through interaction with human experts. Here, the hard examples and their corresponding labels are denoted as \mathbf{X}_{at} and \mathbf{Y}_{at} , respectively. It is noted that the aforementioned selection and update of the easy examples will be performed during the hard curriculum sequence. Afterward, the example set \mathbf{X}_t and the label set \mathbf{Y}_t can be updated with both the easy and hard examples. In the hard CL, the normal entropy regularization loss [11] is employed, considering the hard examples are relatively challenging and uncertain, which may potentially have a negative impact on the model's ability to make accurate predictions. By adhering to this easy-to-hard curriculum, the model is trained to initially handle the simple concepts and gradually progress to the challenging ones. This approach significantly enhances the model's generalization ability and thus can produce satisfactory results on unseen examples.

D. Optimization of IEH-DA

The optimization of our approach follows a three-step process, and the specific training process is outlined in Algorithm 1. First, the parameters of the feature extractor and three classifiers are updated based on all the source examples and labeled target examples, to narrow the gap between the source features and the target features, namely,

$$\begin{aligned} \min_{\theta_E, \theta_{C1}, \theta_{C2}, \theta_{C3}} & L_{cls}^s(\mathbf{X}_s, \mathbf{Y}_s) + \beta L_{ent}^t(\mathbf{X}_{ut}) \\ & + \mu L_{fs}(\mathbf{X}_s, \mathbf{Y}_s, \mathbf{X}_t, \mathbf{Y}_t) + \alpha L_{cls}^t(\mathbf{X}_t, \mathbf{Y}_t). \end{aligned} \quad (15)$$

In the second step, the domain classification loss L_d is used to guide the domain discriminator in correctly identifying the domains of the examples. Meanwhile, the CDD loss L_{cdd}^t is used to enhance the differences among the classifiers. The

Algorithm 1 Training Algorithm of IEH-DA

Input: Source domain examples $\mathbf{X}_s, \mathbf{Y}_s$, target domain examples \mathbf{X}_{ut} , target domain labeled examples $\mathbf{X}_t, \mathbf{Y}_t$, the epoch number to start easy example selection T , the epoch number to start hard example selection \mathcal{H} , the epoch number to end the cycle M , trade-off parameters $\alpha, \beta, \gamma, \delta, \mu$.

Output: Feature extractor E and three classifiers C_1, C_2, C_3 , and domain discriminator D .

- 1: Randomly initialize E, C_1, C_2, C_3, D and set $\alpha = \beta = 0$.
- 2: Set Indx-hard=FALSE.
- 3: **While** epoch $\leq M$ or Indx-hard==FALSE:
 - // **Domain adversarial learning**:
 - 4: **If** epoch $\geq T$, set the values of α, β .
 - 5: Update E, C_1, C_2, C_3 through Eq. (15).
 - 6: Update C_1, C_2, D through Eq. (16).
 - 7: Update E, C_3 through Eq. (17).
 - // **Easy example selection**:
 - 8: **If** epoch % $T == 0$:
 - 9: Initialize \mathbf{X}_t as the existing target labeled examples.
 - 10: Obtain the joint distribution matrices \mathbf{C} and \mathbf{Q} according to Eq. (7) and Eq. (9), respectively.
 - 11: Find and prune untrusted examples to obtain the easy examples $\mathbf{X}_{ct}, \hat{\mathbf{Y}}_{ct}$.
 - 12: Update $\mathbf{X}_t \leftarrow \mathbf{X}_t \cup \mathbf{X}_{ct}, \mathbf{Y}_t \leftarrow \mathbf{Y}_t \cup \hat{\mathbf{Y}}_{ct}$.
 - // **Hard example selection**:
 - 13: **If** epoch == \mathcal{H} and Indx-hard == False:
 - 14: Calculate the score of untrusted target examples by Eq. (14).
 - 15: Select hard examples \mathbf{X}_{at} and interactively label them and set Indx-hard = True.
 - 16: Update $\mathbf{X}_t \leftarrow \mathbf{X}_t \cup \mathbf{X}_{at}, \mathbf{Y}_t \leftarrow \mathbf{Y}_t \cup \mathbf{Y}_{at}$
 - 17: Update the parameters of E, C_1, C_2, C_3 , and D with $\mathbf{X}_t, \mathbf{Y}_t, \mathbf{X}_s$ and \mathbf{Y}_s .

classification losses L_{cls}^s and L_{cls}^t are used to ensure basic classification ability of the classifiers

$$\begin{aligned} \min_{\theta_{C1}, \theta_{C2}, \theta_D} & L_{cls}^s(\mathbf{X}_s, \mathbf{Y}_s) + \alpha L_{cls}^t(\mathbf{X}_t, \mathbf{Y}_t) \\ & + \delta L_d(\mathbf{X}_s, \mathbf{X}_t) - \gamma L_{cdd}^t(\mathbf{X}_t). \end{aligned} \quad (16)$$

In the third step, the feature extractor can fight against the domain discriminator and biclassifiers by using the domain classification loss L_d and the CDD loss L_{cdd}^t . In addition, the feature similarity loss L_{fs} and the entropy regularization loss L_{ent}^t are employed to further enhance the feature extractor E and classifier C_3

$$\begin{aligned} \min_{\theta_E, \theta_{C3}} & \gamma L_{cdd}^t(\mathbf{X}_t) + \mu L_{fs}(\mathbf{X}_s, \mathbf{Y}_s, \mathbf{X}_t, \mathbf{Y}_t) \\ & - \delta L_d(\mathbf{X}_s, \mathbf{X}_t) + \beta L_{ent}^t(\mathbf{X}_{ut}). \end{aligned} \quad (17)$$

IV. EXPERIMENTAL RESULTS

In this section, extensive experiments will be conducted to evaluate the effectiveness of the proposed IEH-DA method. Specifically, we first compare IEH-DA with other state-of-the-art approaches on three datasets, where four metrics including

TABLE I
CLASSIFICATION ACCURACIES FOR EACH CLASS, OA, AA, AND KAPPA COEFFICIENT ON THE PC USING TEN METHODS

Class	SVM [13]	Random Forest [55]	3D-CNN [56]	ResNet [57]	DCFSL [58]	CDADA [59]	UJADA [60]	CLDA [11]	S3VAADA [52]	IEH-DA
1	79.04	82.19	81.53	87.16	92.43	95.97	97.15	96.17	92.92	96.12
2	93.86	89.01	80.30	29.45	98.47	81.78	97.06	98.68	97.96	99.29
3	74.47	52.00	100.00	100.00	98.62	60.63	72.10	78.22	77.09	86.00
4	46.86	79.55	89.03	86.75	84.14	19.04	72.22	82.71	87.35	90.05
5	100.00	100.00	100.00	97.20	99.97	99.21	99.37	100.00	100.00	97.37
6	87.42	90.82	88.20	98.64	95.36	78.76	91.46	92.07	89.61	93.51
7	66.69	96.58	61.90	50.40	86.34	58.59	83.00	91.21	92.66	94.09
OA	76.37	85.63	82.47	69.87	92.49	68.61	88.15	92.17	92.20	94.60
AA	78.34	84.31	82.58	80.20	93.62	70.57	87.48	91.29	91.08	93.78
Kappa	71.89	82.76	79.21	64.74	91.00	62.28	85.71	90.57	90.61	93.49

TABLE II
CLASSIFICATION ACCURACIES FOR EACH CLASS, OA, AA, AND KAPPA COEFFICIENT ON HOUSTON 2018 USING TEN METHODS

Class	SVM [13]	Random Forest [55]	3D-CNN [56]	ResNet [57]	DCFSL [58]	CDADA [59]	UJADA [60]	CLDA [11]	S3VAADA [52]	IEH-DA
1	81.29	73.65	96.51	62.66	69.12	49.77	70.20	58.63	65.41	67.08
2	61.25	92.48	41.11	55.12	90.25	90.01	86.84	89.89	85.38	62.95
3	86.30	61.81	80.14	70.87	84.35	65.72	68.78	69.22	56.61	67.93
4	100.00	100.00		87.5	87.5	75.00	97.27	89.09	78.18	79.07
5	78.09	76.02	92.72	77.35	71.09	88.28	77.33	95.71	93.92	80.61
6	48.97	42.67	63.45	39.61	73.08	48.65	35.60	54.63	65.70	86.92
7	80.85	72.20	40.92	67.76	51.55	78.06	72.20	76.85	71.18	69.87
OA	59.61	55.93	63.36	53.64	71.59	60.89	51.51	65.53	70.53	80.56
AA	66.68	64.12	54.26	47.75	72.29	73.97	71.43	74.73	73.90	75.96
Kappa	45.64	41.89	47.67	40.71	57.12	47.92	39.91	52.82	57.17	68.13

per-class accuracy, overall accuracy (OA), average accuracy (AA), and Kappa coefficient are utilized for the evaluation of model performance. Then, we investigate the parametric sensitivity of hyperparameters. Afterward, we demonstrate that both the hard examples and the curriculum sequence in our IEH-DA are beneficial to obtaining promising performance.

A. Experimental Settings

To evaluate the performance of the proposed method, three publicly available real-world HSI datasets are adopted in our experiments, including Pavia, Houston, and Indiana, which are introduced in the Supplementary Material. For all three real-world datasets, we randomly selected 180 labeled examples in the source domain from each category and five labeled target examples per class for network training. In the meanwhile, five additional labeled examples in the target domain will be provided by human experts during the training process. The details of our network architectures will also be introduced in the Supplementary Material.

In our experiments, to evaluate the classification ability of our proposed IEH-DA, nine approaches are adopted for comparison with the proposed method, including two traditional classification algorithms (i.e., SVM [13] and Random Forest [55]), two deep-learning-based methods (i.e., 3-D-CNN [56] and ResNet [57]), one semisupervised DA method [deep cross-domain few-shot learning (DCFSL) [58]], three unsupervised DA methods [cross-dataset hyperspectral image classification based on adversarial DA (CDADA) [59], unsupervised joint adversarial DA for cross-scene hyperspectral image classification (UJADA) [60], and CLDA [11]], and one informative example selection method [submodular subset selection for virtual adversarial active DA (S3VAADA) [52]]. Note that the hyperparameters of these methods have been

carefully tuned, which are presented in the Supplementary Material. Meanwhile, all experiments are implemented five times to mitigate the effects of random sampling, and the mean values are reported. Besides, since the informative example selection in our method introduced five additional labeled target examples provided by human experts, we added one labeled target example for each category for other comparison experiments to ensure fairness (i.e., six labeled target examples in total for each category). The detailed numbers of the labeled training examples for different methods are also presented in the Supplementary Material.

B. Experimental Results

As shown in Tables I–III, we can draw several observations as follows.

- For the Pavia dataset, our IEH-DA method gains a margin of 2.11 percentage points, 0.16 percentile points, and 2.49 percentile points improvement over the second-best method in terms of OA, AA, and Kappa coefficient, respectively. For the Houston dataset, our method achieves an advantage of 8.97 percentile points OA, 1.23 percentile points AA, and 10.96 percentile points Kappa coefficient compared with the second-best method. Analogously, for the Indiana dataset, IEH-DA outperforms the second-best method by a margin of 6.77 percentile points in terms of OA, 8.13 percentile points in terms of AA, and 7.96 percentile points in terms of Kappa coefficient, respectively. Besides, the proposed IEH-DA achieves the best classification performance in the first, second, fourth, and seventh categories, respectively, in the Pavia dataset. In the Houston dataset, IEH-DA achieves the best classification performance in the sixth category, while other methods obtain poor

TABLE III

CLASSIFICATION ACCURACIES FOR EACH CLASS, OA, AA, AND KAPPA COEFFICIENT ON THE TARGET SCENE OF INDIANA USING TEN METHODS

Class	SVM [13]	Random Forest [55]	3D-CNN [56]	ResNet [57]	DCFSL [58]	CDADA [59]	UJADA [60]	CLDA [11]	S3VAADA [52]	IEH-DA
1	30.21	52.11	63.28	86.48	33.07	37.31	23.64	17.68	40.34	47.88
2	3.10	33.62	13.68	20.72	40.38	5.45	8.51	1.26	3.87	58.97
3	32.39	19.90	70.39	8.93	41.01	40.15	37.45	51.49	49.63	43.59
4	97.04	93.64	99.02	85.99	75.71	68.78	89.83	92.94	81.09	89.45
5	54.74	34.62	16.95	39.93	56.60	12.38	1.82	0.34	39.94	53.31
6	21.01	56.00	22.12	86.21	50.74	0.99	3.61	0.63	2.38	58.51
7	83.83	89.01	98.99	99.50	92.74	92.57	92.63	98.01	97.86	97.12
OA	48.05	52.52	60.35	55.63	59.87	45.55	43.98	46.83	53.55	66.64
AA	46.06	54.13	53.02	55.99	55.75	36.78	36.81	37.48	45.02	64.12
Kappa	37.96	43.02	49.93	47.07	50.82	33.40	31.83	34.17	42.13	58.78

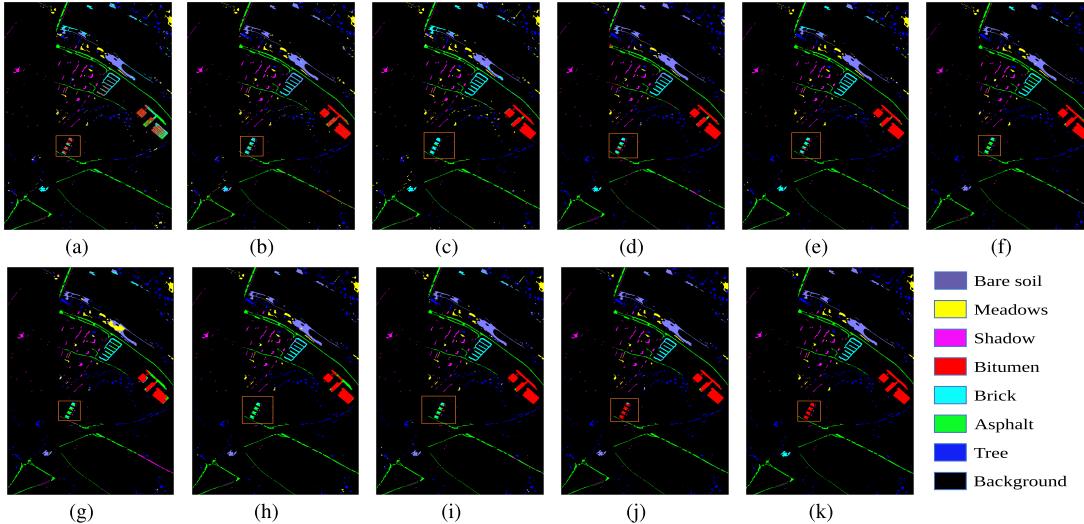


Fig. 3. Classification maps for PC with different methods. (a) SVM. (b) Random Forest. (c) 3-D-CNN. (d) ResNet. (e) DCFSL. (f) CDADA. (g) UJADA. (h) CLDA. (i) S3VAADA. (j) IEH-DA. (k) Ground truth.

classification performance in the sixth category. In the Indiana dataset, our method achieves the best classification performance in the second category, while there is room for improvement in the classification performance of the third category. This is probably attributed to the high similarity between classes of the Indiana dataset, which makes it difficult for classifiers to accurately distinguish the categories.

- 2) With the selection of “easy” examples and the use of pseudo-labels, the unsupervised DA algorithms gain improvements over both the traditional semisupervised algorithms and the deep-learning-based semisupervised traditional methods. Taking Pavia as an example, the CLDA algorithm outperforms Random Forest by a margin of 6.54 percentage points in terms of OA and achieves an advantage of 9.7 percentile points OA compared with a 3-D-CNN algorithm, respectively. The results highlight that the use of high confident examples and pseudo-labels can effectively enhance the classification ability of the model. In addition, the performance of the model trained by “hard” examples is better than that of the model trained by “easy” examples, which is revealed by the performance gap between S3VAADA and CLDA. Moreover, we also find that the DCFSL algorithm gains a margin of 10.02 percentile points improvement over the 3-D-CNN in terms of OA.

This improvement can be attributed to the ability of DA algorithms to align the distributions of the two domains by leveraging adversarial training. Additionally, the utilization of both “easy” and “hard” examples, along with the easy-to-hard curriculum sequence, contributes to the excellent classification ability of our model. In the Houston dataset, our IEH-DA achieves an advantage of 8.97 and 15.03 percentage points in terms of OA compared with DCFSL and CLDA, respectively. Figs. 3–5 also show the classification prediction maps for the target domain.

C. Parametric Sensitivity

Three critical hyperparameters should be pretuned manually in our proposed method, that is, β , δ , and μ . Here, β is the weight of the entropy regularization loss, μ is the weight of the feature similarity loss, and δ represents the weight of the domain classification loss. Here, the Houston dataset is adopted for evaluation. Fig. 6(a)–(c), respectively, shows the OAs of IEH-DA with different values of different hyperparameters. In Fig. 6(a), we find that using a large β usually leads to unsatisfactory performances, while promising results can be obtained with a small β . Hence, it is reasonable to select a relatively small value for β . In addition, an appropriate δ is critical for achieving satisfactory performance. In Fig. 6(b),

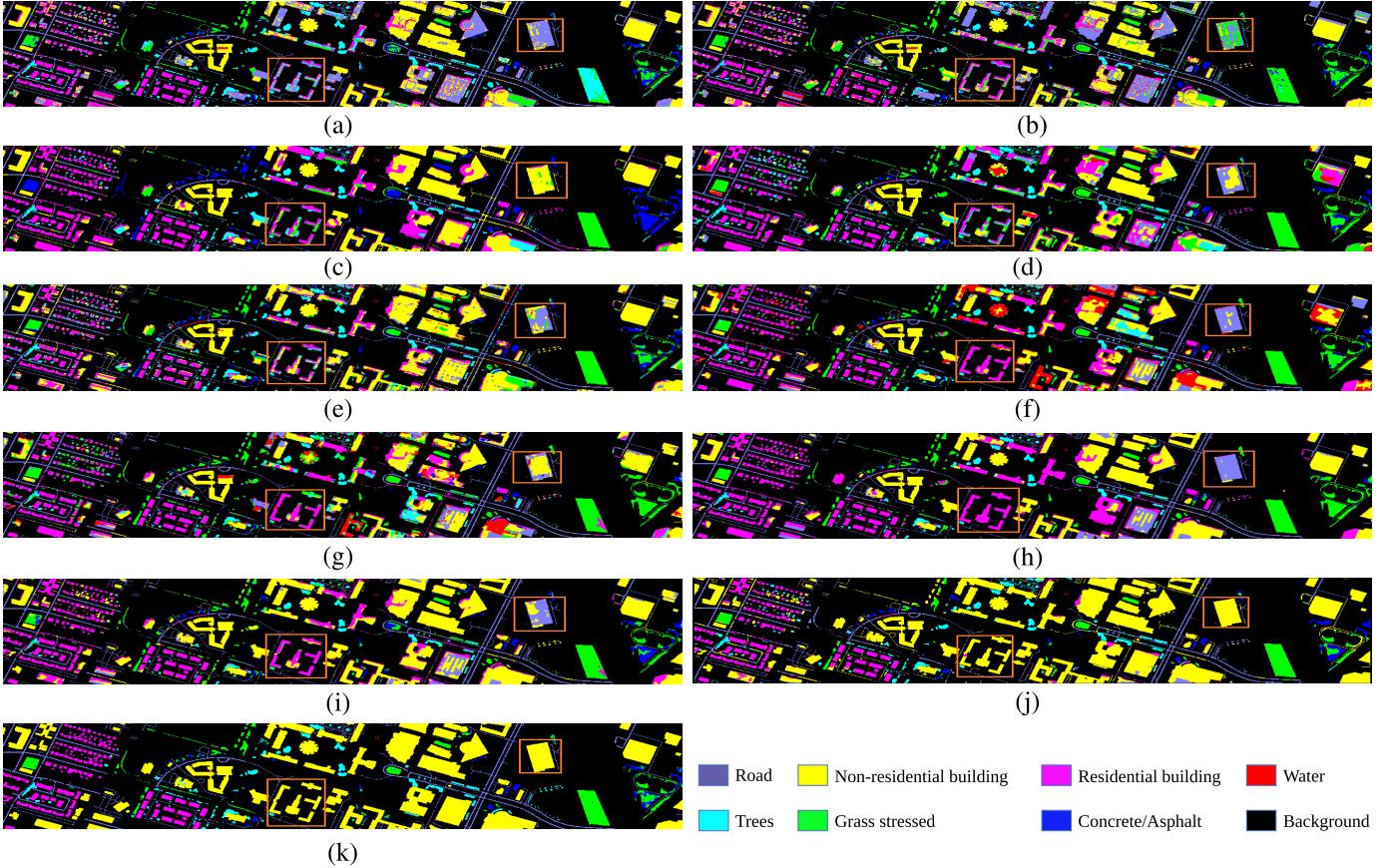


Fig. 4. Classification maps for Houston 2018 with different methods. (a) SVM. (b) Random Forest. (c) 3-D-CNN. (d) ResNet. (e) DCFSL. (f) CDADA. (g) UJADA. (h) CLDA. (i) S3VAADA. (j) IEH-DA. (k) Ground truth.

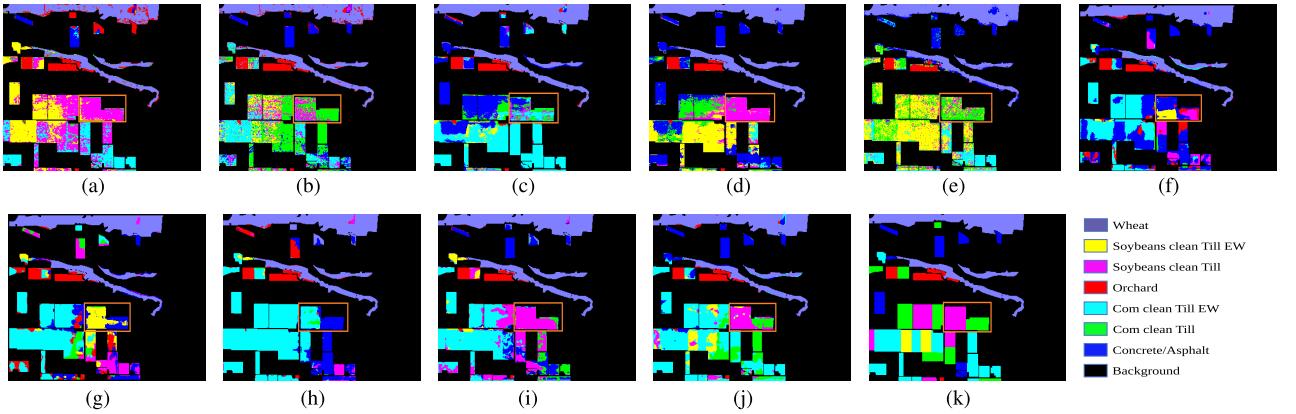


Fig. 5. Classification maps for the target scene of Indiana with different methods. (a) SVM. (b) Random Forest. (c) 3-D-CNN. (d) ResNet. (e) DCFSL. (f) CDADA. (g) UJADA. (h) CLDA. (i) S3VAADA. (j) IEH-DA. (k) Ground truth.

we find that the best result is usually reached with a relatively large δ , since a small δ may lead to a residual domain gap for the adversarial learning. For μ , although increasing the value of μ will enhance the representation ability of the network, the risk of overfitting also gets higher. Therefore, we should carefully select a reasonable μ .

To investigate the impact of different input patch sizes on different classification methods, we conducted experiments with patch sizes of 3×3 , 5×5 , 7×7 , and 9×9 using three DA methods, respectively (i.e., CLDA [11], S3VAADA [52], and our proposed IEH-DA). Fig. 7(a)–(c) displays the OAs

of different methods with different patch sizes on three datasets. The optimal performance is typically achieved with a medium-sized patch size. Therefore, we have chosen the patch size 5×5 for the above three DA methods.

D. Ablation Study

Our proposed IEH-DA contains two key components, that is, the hard example selection and the easy-to-hard curriculum. Therefore, we use three datasets to demonstrate the usefulness of them. Here, we use two reduced models for comparison:

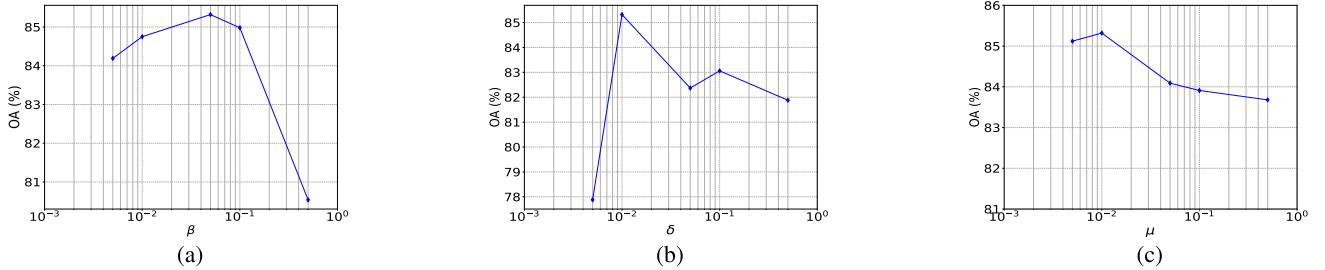
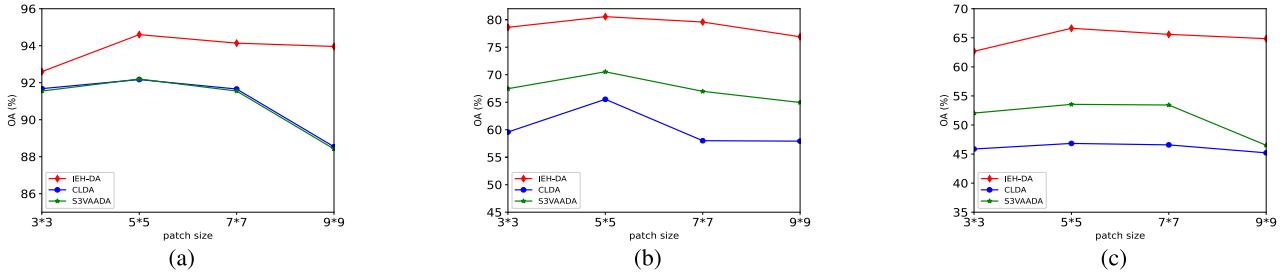
Fig. 6. Sensitivity analysis of the hyperparameter on the Houston dataset. (a) β . (b) δ . (c) μ .

Fig. 7. Sensitivity analysis of patch sizes on different methods. (a) Pavia. (b) Houston. (c) Indiana.

TABLE IV

OAs (%) OF THE PROPOSED IEH-DA WITH DIFFERENT ADVERSARIAL LEARNING MODELS AND CURRICULA ON THE INDIANA, HOUSTON, AND PAVIA DATASETS

Methods	Pavia	Houston	Indiana
IEH-DA w/o hard examples	93.20	75.42	59.21
IEH-DA w/o curriculum	94.25	77.01	63.68
IEH-DA	94.60	80.56	66.64

1) IEH-DA without hard example selection and active label acquisition (IEH-DA w/o hard examples) and 2) IEH-DA without the progressive easy-to-hard curriculum (IEH-DA w/o curriculum). Note that, in 2), the adversarial learning model is trained by easy examples and hard examples simultaneously. As shown in Table IV, IEH-DA without using hard examples or the easy-to-hard curriculum performs worse than the original IEH-DA, which proves the effectiveness of the two components.

V. CONCLUSION

In this article, we proposed a semisupervised DA framework combining dual-cross-domain adversarial learning with an easy-to-hard curriculum. The key idea of IEH-DA involves the use of two distinct strategies to select easy and hard examples, respectively, which are then formed into a curriculum sequence. Accordingly, IEH-DA initially identifies the examples at the cluster center as simple curricula for training. As the model gains proficiency with the initial simple curricula, the examples near the cluster edge are then selected as hard curricula for training. By this means, the model can initially handle the simple curriculum with high confidence, before addressing the hard curriculum with rich information, which can help improve the generalization ability of the model. In addition, considering the high uncertainty in the pseudo-labels of the hard examples, the interaction with human experts is incorporated to provide accurate supervision information for subsequent adaptation. The superiority and

stability of the proposed method are demonstrated by the experimental results of various methods on three real-world HSI datasets. Future work could explore the combination of curriculum-based DA methods that incorporate different difficulty measurers or training schedulers to further enhance the model performance.

REFERENCES

- [1] J. Liu et al., "Estimating the forage neutral detergent fiber content of Alpine grassland in the Tibetan Plateau using hyperspectral data and machine learning algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4405017.
- [2] B. Luo, C. Yang, J. Chanussot, and L. Zhang, "Crop yield estimation based on unsupervised linear unmixing of multivariate hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 162–173, Jan. 2013.
- [3] P. W. Yuen and M. Richardson, "An introduction to hyperspectral imaging and its application for security, surveillance and target acquisition," *Imag. Sci. J.*, vol. 58, no. 5, pp. 241–253, Oct. 2010.
- [4] C. Gong, H. Shi, J. Yang, and J. Yang, "Multi-manifold positive and unlabeled learning for visual analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 5, pp. 1396–1409, May 2020.
- [5] S. Zhong, T. Zhou, S. Wan, J. Yang, and C. Gong, "Dynamic spectral-spatial Poisson learning for hyperspectral image classification with extremely scarce labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5517615.
- [6] S.-E. Qian, "Hyperspectral satellites, evolution, and development history," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7032–7056, 2021.
- [7] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [8] J. Lin, C. He, Z. J. Wang, and S. Li, "Structure preserving transfer learning for unsupervised hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1656–1660, Oct. 2017.
- [9] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 41–57, Jun. 2016.
- [10] S. Zhong and Y. Zhang, "An iterative training sample updating approach for domain adaptation in hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 10, pp. 1821–1825, Oct. 2021.
- [11] Z. Fang et al., "Confident learning-based domain adaptation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5527116.

- [12] H. Zhang, L. Liu, Y. Long, and L. Shao, "Unsupervised deep hashing with pseudo labels for scalable image retrieval," *IEEE Trans. Image Process.*, vol. 27, pp. 1626–1638, 2018.
- [13] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [14] S. Zhong, C.-L. Chang, and Y. Zhang, "Iterative edge preserving filtering approach to hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 90–94, Jan. 2019.
- [15] S. Zhong et al., "Class feature weighted hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4728–4745, Dec. 2019.
- [16] P. Jia, M. Zhang, W. Yu, F. Shen, and Y. Shen, "Convolutional neural network based classification for hyperspectral data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 5075–5078.
- [17] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang, and J. Yang, "Multiscale dynamic graph convolutional network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3162–3177, May 2019.
- [18] S. Wan, C. Gong, P. Zhong, S. Pan, G. Li, and J. Yang, "Hyperspectral image classification with context-aware dynamic graph convolutional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 597–612, Jan. 2021.
- [19] W. Yu, S. Wan, G. Li, J. Yang, and C. Gong, "Hyperspectral image classification with contrastive graph convolutional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5503015.
- [20] D. Hong et al., "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518615.
- [21] J. He, L. Zhao, H. Yang, M. Zhang, and W. Li, "HSI-BERT: Hyperspectral image classification using the bidirectional encoder representation from transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 165–178, Jan. 2020.
- [22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
- [23] L. Zhang and D. Zhang, "Robust visual knowledge transfer via extreme learning machine-based domain adaptation," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4959–4973, Oct. 2016.
- [24] B. Sun, J. Feng, and K. Saenko, "Correlation alignment for unsupervised domain adaptation," in *Domain Adaptation in Computer Vision Applications*. Berlin, Germany: Springer, 2017, pp. 153–171.
- [25] Z. Sun, C. Wang, H. Wang, and J. Li, "Learn multiple-kernel SVMs for domain adaptation in hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1224–1228, Sep. 2013.
- [26] Y. Mansour and M. Schain, "Robust domain adaptation," *Ann. Math. Artif. Intell.*, vol. 71, no. 4, pp. 365–380, 2014.
- [27] A. J. Ma, J. Li, P. C. Yuen, and P. Li, "Cross-domain person reidentification using domain adaptation ranking SVMs," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1599–1613, May 2015.
- [28] M. Long, Y. Cao, and J. Wang, "Learning transferable features with deep adaptation networks," in *Proc. ICML*, 2015, pp. 97–105.
- [29] W. Zellinger, T. Grubinger, E. Lughofer, T. Natschläger, and S. Saminger-Platz, "Central moment discrepancy (CMD) for domain-invariant representation learning," 2017, *arXiv:1702.08811*.
- [30] I. Goodfellow, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [31] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. ICML*, 2015, pp. 1180–1189.
- [32] Z. Pei, Z. Cao, M. Long, and J. Wang, "Multi-adversarial domain adaptation," in *Proc. AAAI*, vol. 32, 2018.
- [33] S. Li, F. Lv, B. Xie, C. H. Liu, J. Liang, and C. Qin, "Bi-classifier determinacy maximization for unsupervised domain adaptation," in *Proc. AAAI*, vol. 35, 2021, pp. 8455–8464.
- [34] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proc. ICML*, 2009, pp. 41–48.
- [35] X. Wang, Y. Chen, and W. Zhu, "A survey on curriculum learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4555–4576, Mar. 2021.
- [36] C. Gong, D. Tao, S. J. Maybank, W. Liu, G. Kang, and J. Yang, "Multi-modal curriculum learning for semi-supervised image classification," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3249–3260, Jul. 2016.
- [37] C. Gong, "Exploring commonality and individuality for multi-modal curriculum learning," in *Proc. AAAI*, vol. 31, 2017.
- [38] C. Gong, J. Yang, and D. Tao, "Multi-modal curriculum learning over graphs," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 4, pp. 1–25, Jul. 2019.
- [39] X. Luo, S. Li, X. Shi, and J. Yin, "Learning from small data for hyperspectral image classification," *Signal Process.*, vol. 213, Dec. 2023, Art. no. 109183.
- [40] W. Wei, C. Song, L. Zhang, and Y. Zhang, "Lightweighted hyperspectral image classification network by progressive bi-quantization," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5501914.
- [41] S. Yang, Z. Feng, M. Wang, and K. Zhang, "Self-paced learning-based probability subspace projection for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 2, pp. 630–635, Feb. 2019.
- [42] J. Choi, M. Jeong, T. Kim, and C. Kim, "Pseudo-labeling curriculum for unsupervised domain adaptation," 2019, *arXiv:1908.00262*.
- [43] B. Zhang, T. Chen, and B. Wang, "Curriculum-style local-to-global adaptation for cross-domain remote sensing image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5611412.
- [44] R. Thoreau, V. Achard, L. Risser, B. Berthelot, and X. Briottet, "Active learning for hyperspectral image classification: A comparative review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 3, pp. 256–278, Sep. 2022.
- [45] Z. Ye, T. Sun, S. Shi, L. Bai, and J. E. Fowler, "Local–global active learning based on a graph convolutional network for semi-supervised classification of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5501805.
- [46] X. Li, Z. Cao, L. Zhao, and J. Jiang, "ALPN: Active-learning-based prototypical network for few-shot hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5508305.
- [47] Y. Li, T. Lu, and S. Li, "Subpixel-pixel-superpixel-based multiview active learning for hyperspectral images classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4976–4988, Jul. 2020.
- [48] M. Ahmad et al., "A disjoint samples-based 3D-CNN with active transfer learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5539616.
- [49] J. Lin, L. Zhao, S. Li, R. Ward, and Z. J. Wang, "Active-learning-incorporated deep transfer learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4048–4062, Nov. 2018.
- [50] J. M. Bioucas-Dias et al., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
- [51] C. Northcutt, L. Jiang, and I. Chuang, "Confident learning: Estimating uncertainty in dataset labels," *J. Artif. Intell. Res.*, vol. 70, pp. 1373–1411, Apr. 2021.
- [52] H. Rangwani, A. Jain, S. K. Aithal, and R. V. Babu, "S3VAADA: Submodular subset selection for virtual adversarial active domain adaptation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 7516–7525.
- [53] G. H. Golub and H. A. van der Vorst, "Eigenvalue computation in the 20th century," *J. Comput. Appl. Math.*, vol. 123, nos. 1–2, pp. 35–65, 2000.
- [54] S. Cui, S. Wang, J. Zhuo, L. Li, Q. Huang, and Q. Tian, "Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations," in *Proc. CVPR*, Jun. 2020, pp. 3941–3950.
- [55] B. Liu et al., "Morphological attribute profile cube and deep random forest for small sample classification of hyperspectral image," *IEEE Access*, vol. 8, pp. 117096–117108, 2020.
- [56] C. Yu, R. Han, M. Song, C. Liu, and C.-I. Chang, "A simplified 2D-3D CNN architecture for hyperspectral image classification based on spatial-spectral fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2485–2501, 2020.
- [57] Y. Jiang, Y. Li, and H. Zhang, "Hyperspectral image classification based on 3-D separable ResNet and transfer learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1949–1953, Dec. 2019.
- [58] 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.
- [59] X. Ma, X. Mou, J. Wang, X. Liu, J. Geng, and H. Wang, "Cross-dataset hyperspectral image classification based on adversarial domain adaptation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4179–4190, May 2021.
- [60] X. Tang, C. Li, and Y. Peng, "Unsupervised joint adversarial domain adaptation for cross-scene hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5536415.



Cheng Zhang received the B.S. degree from the Nanjing University of Technology, Nanjing, China, in 2021, where he is currently pursuing the M.S. degree in computer science and technology with the School of Computer Science and Engineering.

His research interests include domain adaptation and classification for hyperspectral images.



Shengwei Zhong received the B.E. degree in information countermeasure technology and the M.S. and Ph.D. degrees in electronics and communication engineering from the Harbin Institute of Technology, Harbin, China, in 2013, 2015, and 2020, respectively.

She was an Exchange Ph.D. Student visiting the Remote Sensing Signal and Image Processing Laboratory (RSSIPL) as a Faculty Research Assistant at the University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA. She is currently an Associate Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. Her research interests include hyperspectral image processing, remote-sensing image fusion, and applications.



Sheng Wan received the Ph.D. degree in computer science and technology from the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China, in 2023.

He is currently a Post-Doctoral Researcher with the School of Computer Science and Engineering, Nanjing University of Science and Technology. His research interests include graph machine learning, weakly supervised learning, and hyperspectral image processing.



Chen Gong (Senior Member, IEEE) received the dual Ph.D. degrees from Shanghai Jiao Tong University (SJTU), Shanghai, China, and the University of Technology Sydney (UTS), Ultimo, NSW, Australia, in 2016 and 2017, respectively.

He is currently a Full Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. He has authored more than 100 technical papers at prominent journals and conferences, such as *Journal of Machine Learning Research* (JMLR), IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (TNNLS), IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), IEEE TRANSACTIONS ON CYBERNETICS (TCYB), IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (TCSVT), IEEE TRANSACTIONS ON MULTIMEDIA (TMM), IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS (TITS), ACM Transactions on Intelligent Systems and Technology (TIST), International Conference on Machine Learning (ICML), Conference on Neural Information Processing Systems (NeurIPS), International Conference on Learning Representations (ICLR), Conference on Computer Vision and Pattern Recognition (CVPR), Association for the Advancement of Artificial Intelligence (AAAI), International Joint Conference on Artificial Intelligence (IJCAI), and IEEE International Conference on Data Mining (ICDM). His research interests mainly include machine learning, data mining, and learning-based vision problems.

Dr. Gong received the “Excellent Doctoral Dissertation Award” of the Chinese Association for Artificial Intelligence, “Young Elite Scientists Sponsorship Program” of the China Association for Science and Technology, “Wu Wen-Jun AI Excellent Youth Scholar Award,” and the Science Fund for Distinguished Young Scholars of Jiangsu Province. He was also selected as the “Global Top Chinese Young Scholars in AI” released by Baidu. He serves as the Area Chair or a Senior PC Member for several top-tier conferences, such as AAAI, IJCAI, ACM International Conference on Multimedia (ACM MM), European Conference on Machine Learning (ECML), and International Conference on Artificial Intelligence and Statistics (AISTATS). He serves as an Associate Editor for the IEEE TCSVT and *Neural Processing Letters* (NePL).