Combine Meta-Learning with Feature Alignment for Cross-Domain Heterogeneous Hyperspectral Image Classification

Minchao Ye, *Member, IEEE,* Yuheng Jin, Jianwei Zhao, Weiqi Yan, Yuntao Qian, *Senior Member, IEEE*

***Abstract*—The scarcity of labeled samples results in the challenge of small-sample-size in hyperspectral image (HSI) classification. Transfer learning offers hope for solving this problem. In cross-domain transfer learning, the source domain boasts abundant labeled training samples, whereas the target domain comprises only limited labeled training samples. Leveraging the information from the source domain can benefit the classification of the target domain. However, inconsistencies in land-cover classes between source and target domains may hinder knowledge transfer between domains. Fortunately, few-shot learning (FSL) provides an effective solution to this challenge. In recent years, meta-learning has gained widespread attention as a mainstream approach within FSL. This paper proposes a novel method for cross-domain heterogeneous HSI classification, called cross-domain meta-learning with feature alignment (CD-MFA). CD-MFA enhances the generalization performance of the inner-loop optimization by incorporating task-adaptive loss function. The adaptive weighting strategy is used in the outer-loop optimization to balance the classification losses of the source and target domains to learn more discriminative features. Additionally, by aligning the features of the source and target domains under the guidance of the Gaussian prior, the impact of domain shift can be mitigated. It is worth noting that CD-MFA is trained concurrently on both the source and target domains so that the two domains are will bound, thereby enhancing the effectiveness of knowledge transfer. Experimental results on four publicly available HSI datasets validate the effectiveness of CD-MFA.**

***Index Terms*—Cross-domain heterogeneous hyperspectral image classification, meta-learning, task-adaptive loss function, adaptive weighting strategy**

I. Introduction

HYPERSPECTRAL images (HSIs) are three-dimensional (3-D) data cubes that encapsulate both spatial and spectral information. Due to the wealth of spatial and spectral information, HSI has become widely employed in classification tasks [1]–[5]. In past years, traditional machine learning methods have been extensively applied to HSI classification [6], [7]. In comparison to traditional machine learning methods, deep learning approaches can automatically extract deep abstract features conducive to classification tasks from input data, thereby achieving heightened classification accuracy. Consequently, deep learning is introduced into HSI classification, yielding superior results [8]–[11]. Nonetheless, a notable drawback is the reliance of most deep learning-based approaches on a large number of training samples. Simultaneously, accurately labeling the newly collected HSI proves to be a considerable difficulty, leading to a small-sample-size problem in HSI classification. Effectively addressing the utilization of limited labeled samples for HSI classification has become a primary research focus.

Transfer learning enables the transference of knowledge learned from the source domain rich in labeled samples to the target domain with limited labeled samples [12]–[16]. This process serves to enhance the classification accuracy of the target domain. This is referred to as cross-domain HSI classification. Therefore, transfer learning emerges as an effective strategy for mitigating the small-sample-size problem in HSI classification [17]. Liu *et al.* [18] proposed an unsupervised domain adaptation network based on class-wise adversarial learning. This approach produces domain-invariant features by employing adversarial learning between the feature extractor and multiple domain discriminators. Furthermore, to enhance feature alignment performance, a probability-prediction-based maximum mean discrepancy (MMD) method is incorporated into the adversarial adaptation network. Ning *et al.* [19] proposed an instance-to-instance contrastive learning framework based on category matching (CLCM). This approach draws upon category information within the feature space and employs a sample from the source domain as an anchor to find its cross-domain positive and negative matching samples. Simultaneously, contrastive learning is implemented by bringing positive matching pairs closer together and pushing negative matching pairs farther apart in the two domains, thereby mitigating domain shift. Wang *et al.* [20] proposed a cross-scene classification model based on feature learning (CSCFL) for HSI classification. In this approach, unsupervised domain adaptation technology is applied to mitigate domain shift. Additionally, it incorporates depth separable convolution to enhance the model's ability to capture fine spatial features.

The effectiveness of these transfer learning approaches

Manuscript received Xxx xx, xxxx; revised Xxxxx xx, xxxx. Corresponding author: Yuntao Qian. This work was supported in part by Zhejiang Provincial Natural Science Foundation of China under Grant No. ZCLMS26F0104, and in part by the National Natural Science Foundation of China under Grant No. 62571475.

Minchao Ye, Yuheng Jin and Jianwei Zhao are with the Zhejiang-New Zealand Joint Laboratory on Vision-Based Intelligent Metrology, College of Information Engineering, China Jiliang University, Hangzhou 310018, China.

Weiqi Yan is with the Zhejiang-New Zealand Joint Laboratory on Vision-Based Intelligent Metrology, School of Engineering, Computer & Mathematical Sciences, Auckland University of Technology, Auckland 1142, New Zealand.

Yuntao Qian is with the College of Computer Science, Zhejiang University, Hangzhou 310027, China.

relies on the premise that the source and target domains are imaged by the same sensor, meaning both domains have identical feature dimensions. This is referred to as homogeneous transfer learning [21], [22]. However, in most cases, the source and target domains are imaged by different sensors, resulting in different feature dimensions, which need to be processed by heterogeneous transfer learning [23], [24]. Zhong *et al.* [25] proposed a cross-domain deep transfer learning approach based on spectral feature adaptation (SFA) for HSI classification. This approach employs the joint probability distribution adaptation technique to alleviate the domain shift between the source and target domains. Additionally, to comprehensively extract discriminant features from HSI, a multiscale spectral-spatial unified network is designed. Wang *et al.* [26] proposed an HSI classification approach based on an unsupervised heterogeneous domain adaptation CycleGan (UHDAC). In this method, the features of the source and target domains are constrained through two-way adversarial transfer, allowing for the alignment of the distributions of the extracted features. Simultaneously, the Coral loss function is applied to address the challenge of insufficient mapping relationship constraints caused by the low consistency in data structure between the source and target domains. Zhong *et al.* [27] proposed a cross-domain HSI classification approach. The approach leverages multiscale convolutional sparse decomposition (MCSD) to acquire transferable features for classification. Additionally, it incorporates a structure preserved distribution alignment (SPDA) model to effectively transfer discriminant knowledge.

While these methods have demonstrated effectiveness in cross-domain HSI classification, most existing transfer learning methods' success relies on the assumption that the source and target domains share the same classes. These approaches encounter challenges when classes present in the target domain are absent in the source domain. Fortunately, few-shot learning can leverage a limited number of training samples to identify unseen classes in the target domain. As a mainstream method in few-shot learning, meta-learning holds promise in addressing this challenge. Unlike traditional deep learning methods, meta-learning utilizes tasks comprised of support sets and query sets as the foundational units of training. This enables the model to acquire the capacity to learn how to learn. In the cross-domain HSI classification task, meta-learning distills the general knowledge of few-shot classification by simulating scenarios of the few-shot classification. Consequently, the problem of inconsistent classes between the source and target domains can be mitigated through the application of meta-learning. Cheng *et al.* [28] proposed a cross-domain HSI classification approach with causal meta-transfer learning (CMTL). CMTL employs a mask transformer to identify noncausal factors unrelated to classes, achieving the separation of causal and noncausal factors through the imposition of a causal constraint. Concurrently, to ensure a robust causal association between causal factors and classification tasks, the method introduces a causal association module (CAM). Li *et al.* [29] integrated model-agnostic meta-learning (MAML) with cutout data augmentation, introducing an approach termed MAML and regularization fine-tuning few-shot learning (MRFSL). This approach is applied to address the challenge posed by the difficulty in utilizing transfer learning for cross-domain HSI classification, stemming from the disparity in ground object categories between two datasets. Hu *et al.* [30] proposed a dual-adjustment mode-based cross-domain meta-learning (DMCM) approach. DMCM employs a dual-adjustment model incorporating intracorrection (IC) and interalignment (IA) learning strategies. IC is dedicated to the precise generation of class prototypes, while the adoption of the IA learning strategy facilitates mitigating the impact of domain shift.

The optimization-based meta-learning methods can quickly learn new tasks through the optimization process [31]. Therefore, this paper proposes a novel optimization-based meta-learning method for cross-domain HSI classification, called cross-domain meta-learning with feature alignment (CD-MFA). CD-MFA incorporates the task-adaptive loss function during inner-loop optimization to improve the model's generalization performance. Then, to alleviate the impact of domain shift, we align the features of both the source and target domains, guided by the Gaussian prior. Additionally, the adaptive weighting strategy (AWS) is employed to effectively balance source and target domain classification losses during the outer-loop optimization process. CD-MFA can be characterized as bi-level optimization: 1) Inner-loop optimization is dedicated to optimizing the base learner, enabling it to adapt to new tasks with limited samples. 2) Outer-loop optimization is to balance source and target domain classification losses via the AWS. The balanced classification loss and feature alignment (FA) loss are then used to train the model to achieve better generalization performance. Notably, CD-MFA is trained simultaneously on both the source and target domains, effectively binding the two domains together and thereby improving the effectiveness of knowledge transfer. The main contributions of this work are as follows.

1) Unlike many existing cross-domain HSI classification algorithms, CD-MFA allows the source and target domains to have different land-cover classes. This is a more common scenario in practical applications.
2) The task-adaptive loss function is employed in inner-loop optimization to enhance the model's generalization performance. Concurrently, the AWS effectively balances source and target domain classification losses in the outer-loop optimization, making it convenient for learning more discriminative features.
3) By aligning the features of the source and target domains under the guidance of the Gaussian prior, the impact of domain shift is mitigated.

The subsequent sections of this paper are structured as follows. Section II provides an introduction to the fundamentals of domain adaptation and meta-learning. Following that, Section III introduces the CD-MFA algorithm. Section IV shows and analyzes the experimental results from four publicly available HSI datasets. Lastly, Section V serves as the conclusion of this paper.

II. Related works

This work integrates domain adaptation and meta-learning, both of which are briefly elucidated in this section.

TABLE I
SYMBOLS AND DESCRIPTIONS OF PROPOSED ALGORITHM IN THIS PAPER

Symbol	Description
$p_s(T)$ and $p_t(T)$	Source and target task distributions.
α	Learning rate of the inner-loop optimization.
β	Learning rate of the outer-loop optimization.
$\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$	Affine transformation parameters in the inner-loop optimization.
$\boldsymbol{\theta}$	Base learner parameters.
$\boldsymbol{\phi}$	Meta-learnable loss function parameters.
$\boldsymbol{\psi}$	Meta-learner parameters.
$\mathcal{DS}_i^S$ and $\mathcal{DT}_i^S$	Source and target domain support sets.
$\mathcal{DS}_i^Q$ and $\mathcal{DT}_i^Q$	Source and target domain query sets.
$\mathbf{xs}_i^s$ and $\mathbf{xt}_i^s$	Source and target domain support set samples.
$\mathbf{xs}_i^q$ and $\mathbf{xt}_i^q$	Source and target domain query set samples.
ξ	Hyperparameter for balancing classification and FA losses.
$\mathcal{L}_{cls}^s$ and $\mathcal{L}_{cls}^t$	Cross-entropy loss.
$\mathcal{L}_{da}^s$ and $\mathcal{L}_{da}^t$	KL divergence.

A. Domain Adaptation

The goal of domain adaptation technology is to address the issue of disparate data distributions between the source and target domains. Traditional domain adaptation methods can be categorized into three main groups: instance-based, feature-based, and classifier-based domain adaptation methods [32]. Through the adjustment of the marginal distribution of either the source or target samples, instance-based domain adaptation methods can effectively align the distributions of both the source and target domains. Zhou *et al.* [33] proposed a transfer learning algorithm for remote sensing image classification based on extreme learning machines with weighted least squares. This method transfers prior knowledge from historical images to the target model by adjusting historical and target samples using different weighting strategies. The feature-based domain adaptation methods transform data from source and target domains into a new feature space. Its objective is to minimize the distribution disparity between the data from these domains within the new feature space during the process of optimization. Zhu *et al.* [34] aligned the two domains by shifting the target domain samples towards the source domain to align the class centroids of the two domains. Due to their powerful modeling capabilities [35], graphs are also employed to model correlations between source and target domains in feature-based adaptation. Ye *et al.* [22] introduced a graph loss to enforce feature alignment across the two domains. The classifier-based domain adaptation methods adapt the classifier trained in the source domain to better suit the target domain. Zhong *et al.* [36] introduced an iterative training sample updating (ITSU) method based on posterior spatial feature extraction. The method iteratively generates training samples from the target domain, ultimately yielding a classifier with enhanced performance in the target domain. Deep learning techniques excel at automatically extracting deep abstract features from data. In recent years, deep domain adaptation has emerged as a prominent research focus and has found extensive application. Deep domain adaptation methods can be categorized into two main types: discrepancy-based and adversarial-based domain adaptation methods [32]. By integrating adaptation layers into the network, the discrepancy-based domain adaptation methods ensure consistency in the feature distribution between the source and target domains. Zhu *et al.* [37] proposed a local maximum mean discrepancy (LMMD) to quantify the discrepancy between related subdomains within the same category across various domains. Adversarial-based domain adaptation methods aim to minimize the distribution discrepancy between domains through adversarial learning, thereby acquiring transferable and domain-invariant features. Wang *et al.* [38] aligned the distributions between the two domains by employing domain adversarial learning techniques simultaneously in both the source and target domains.

Although the existing domain adaptation methods have achieved remarkable results, most of them assume that the source and target domains share the same land-cover classes. When the source and target domains contain different classes, directly aligning their distributions becomes challenging. To overcome this challenge, we improve classification accuracy in the target domain by aligning the features of the source and target domains under the guidance of the Gaussian prior.

B. Meta-learning

The objective of meta-learning is to equip models with the capability to learn how to learn. Unlike traditional deep learning methods, meta-learning methods are trained on tasks consisting of support sets and query sets. Meta-learning methods can be categorized into three main groups: metric-based, model-based, and optimization-based meta-learning methods [31]. Metric-based meta-learning methods predict the labels of query set samples by considering the distances between query set samples and support set samples in an embedded space. Snell *et al.* [39] proposed employing prototypical networks to address few-shot classification challenges. This method conducts classification by calculating the distance of a sample to the prototype representations of each class within a metric space. Model-based meta-learning methods rely upon an adaptive internal state. Munkhdalai *et al.* [40] proposed a meta-learning method called meta networks. This method utilizes loss gradients as meta information, achieving better results. Most of the optimization-based meta-learning methods treat it as a bi-level optimization problem. Inner-loop optimization allows the base learner to adapt to new tasks with a small number of samples, while outer-loop optimization aims to enhance the model's performance. Finn *et al.* [41] proposed an optimization-based meta-learning method called MAML. In this method, a weight initialization is learned, serving as an effective starting point for executing a few update steps on a new task with a limited number of training samples to achieve good generalization performance.

In this paper, we employ the task-adaptive loss function in inner-loop optimization to improve the generalization performance of the model. Simultaneously, the AWS is utilized to balance source and target domain classification losses in outer-loop optimization.

III. PROPOSED CD-MFA ALGORITHM

Meta-learning is the predominant method in few-shot learning, aimed at solving the problem of less-than-ideal clas-

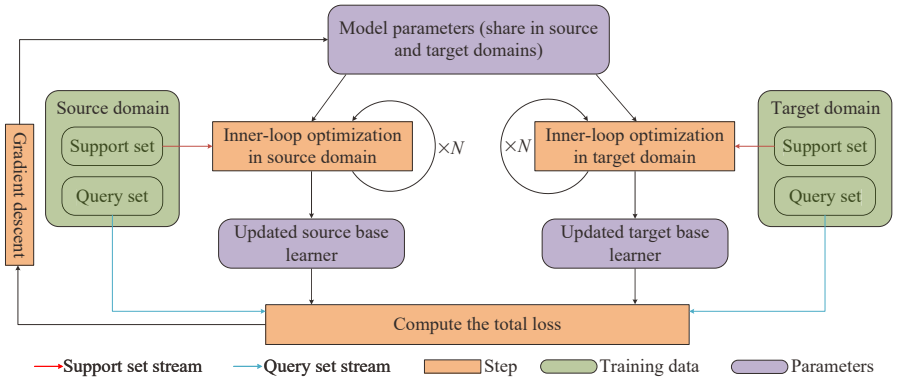

Fig. 1. The flowchart of our proposed CD-MFA.

sification when the source and target domain classes are inconsistent. The CD-MFA method proposed in this paper is an optimization-based meta-learning method. CD-MFA is trained through bi-level optimization. 1) Inner-loop optimization is used to adjust the base learner, enabling it to adapt to new tasks with limited samples. 2) Outer-loop optimization is utilized to train the model. The flowchart and algorithm framework proposed for CD-MFA are depicted in Fig. 1 and Fig. 2. As shown in Fig. 1, the training process of CD-MFA is as follows:

1) The inner-loop optimizations are performed in the source and target domains, respectively, and the base learners are updated.
2) The total loss is computed based on the updated base learners in the source and target domains.
3) Gradient descent is applied to train the model.

By utilizing the updated base learners from the source and target domains to compute the total loss and perform gradient descent for updating model parameters, the two domains are bound for facilitating effective knowledge transfer.

By integrating the task-adaptive loss function into inner-loop optimization, CD-MFA enhances the adaptive ability of inner-loop optimization, thereby enhancing model generalization performance. In the outer-loop optimization, the features of the source and the target domains are aligned with the guidance of the Gaussian prior. Additionally, the AWS is employed to balance source and target domain classification losses, facilitating the learning of more discriminant features.

As illustrated in Fig. 2, our approach aims to leverage both spectral and spatial information present in HSIs effectively. Our base learner is structured into two branches. The first branch employs one-dimensional (1-D) convolutional operations to extract spectral features, while the second branch employs two-dimensional (2-D) convolutional operations to extract spatial features. A fully connected (FC) layer is then employed to fuse the extracted spectral and spatial features.

A. Meta-learning with Task-Adaptive Loss Function

In optimization-based meta-learning methods, the objective of inner-loop optimization is to optimize the base learner, allowing it to adapt to new tasks. Similar to MAML, the inner-loop optimization at n-th step is represented as:

$$\boldsymbol{\theta}_{i,n+1} = \boldsymbol{\theta}_{i,n} - \alpha \nabla_{\boldsymbol{\theta}_{i,n}} \mathcal{L}(\mathcal{D}_i^S; \boldsymbol{\theta}_{i,n}) \tag{1}$$

where $\mathcal{D}_i^S$ denotes the support set, α denotes the learning rate of the inner-loop optimization, and $\boldsymbol{\theta}$ denotes the parameters of the base learner.

Inspired by [42], CD-MFA improves the inner-loop optimization process by integrating the task-adaptive loss function to enhance the model's generalization performance. The detailed computation process of task-adaptive loss is shown in Fig. 3a. This loss function is generated from a small neural network with parameters $\boldsymbol{\phi}$. For convenience, we refer to it as a meta-learnable loss function. The inner-loop optimization at step n is formulated as follows:

$$\boldsymbol{\theta}_{i,n+1} = \boldsymbol{\theta}_{i,n} - \alpha \nabla_{\boldsymbol{\theta}_{i,n}} \mathcal{L}_{\boldsymbol{\phi}_{i,n}}(\boldsymbol{\tau}_{i,n}) \tag{2}$$

where $\boldsymbol{\tau}_{i,n}$ represents the task state of inner-loop optimization at step n. To ensure the task-adaptive ability of the meta-learnable loss function, it is common practice to utilize gradient descent to update the parameters of the meta-learnable loss function. However, the implementation of gradient descent often leads to a substantial computational burden [42]. In order to reduce the computational burden and ensure the adaptive ability of the meta-learnable loss function, we use affine transformation to dynamically change the parameters $\boldsymbol{\phi}$ of the meta-learnable loss function:

$$\boldsymbol{\phi}^{'} = \boldsymbol{\gamma}\boldsymbol{\phi} + \boldsymbol{\eta} \tag{3}$$

where $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ are generated by the meta-learner, as illustrated in Fig. 4. With the affine transformation, the detailed computation process of task-adaptive loss is shown in Fig. 3b.

As both the meta-learnable loss function and the meta-learner are constructed using neural networks, during the n-th step of inner-loop optimization for a given task T_i (comprising the support set $\mathcal{D}_i^S$ and the query set $\mathcal{D}_i^Q$), its input may encompass not only the loss information $\mathcal{L}(\mathcal{D}_i^S)$ of the support set but also task-specific details, like the parameters $\boldsymbol{\theta}_{i,n}$ of the base learner and the output value $f(\mathbf{x}_i^s)$ of the support set samples $\mathbf{x}_i^s$ after passing through the base learner. The task state $\boldsymbol{\tau}_{i,n}$ is represented as follows:

$$\boldsymbol{\tau}_{i,n} = [\mathcal{L}(\mathcal{D}_i^S), \boldsymbol{\theta}_{i,n}, f(\mathbf{x}_i^s)] \tag{4}$$

It aims to enhance the inner-loop optimization process by incorporating the task-adaptive loss function, thereby bolstering the adaptive abilities of the inner-loop optimization and enhancing the model's generalization performance.

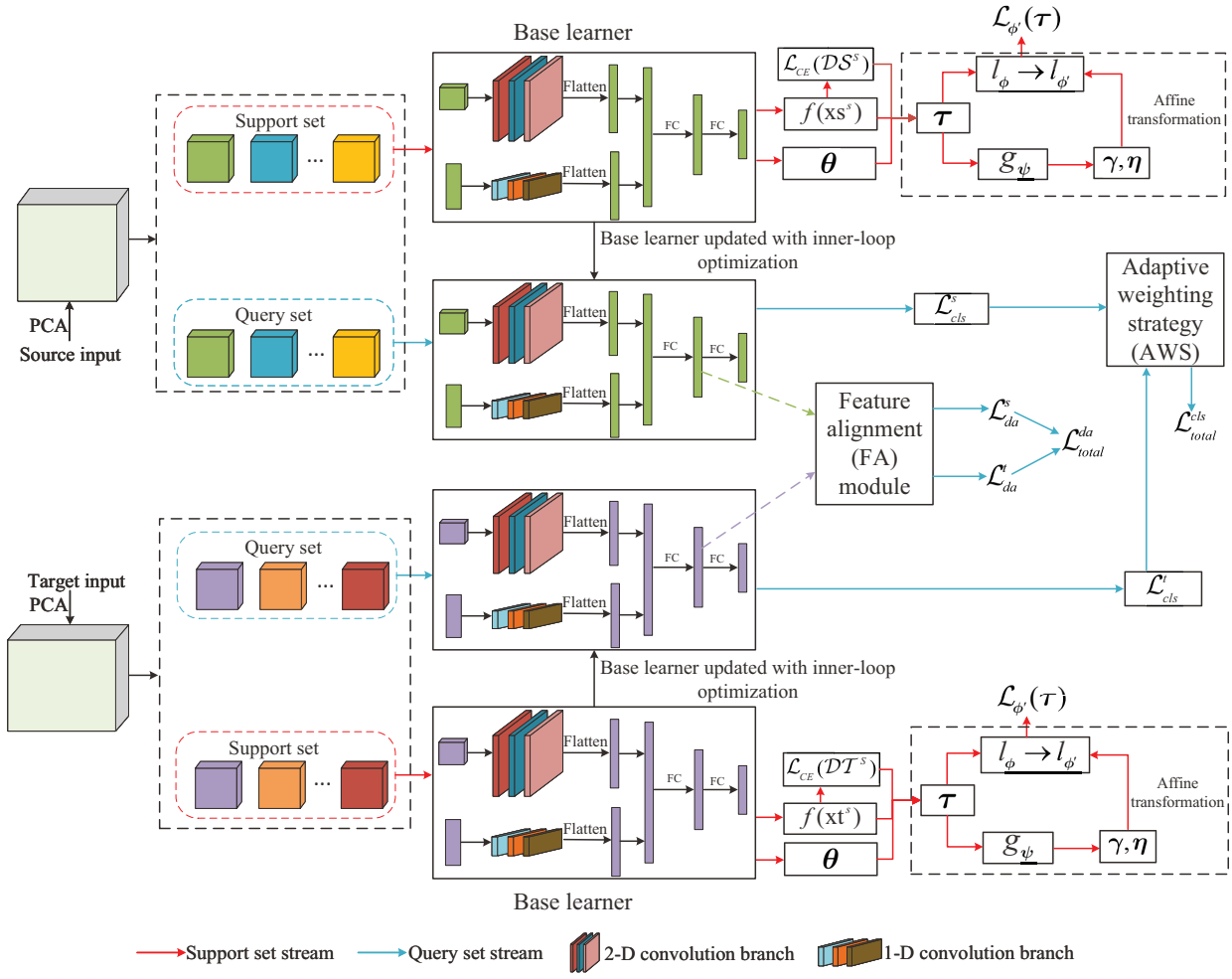

Fig. 2. The framework of our proposed CD-MFA.

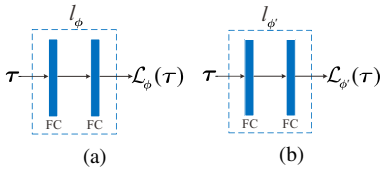

Fig. 3. Detailed computation process of task-adaptive loss. (a) without affine transformation. (b) with affine transformation.

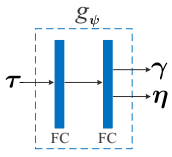

Fig. 4. Detailed computation process of affine transformation parameters γ and η.

B. Source and Target Meta-Learning

After completing the inner-loop optimization, the query set samples need to be employed to assess the optimized base learner. The source domain classification loss $\mathcal{L}_{cls}^{s}(\mathcal{DS}_i^Q)$ can be determined using the source domain query set $\mathcal{DS}_i^Q$. Likewise, the target domain classification loss $\mathcal{L}_{cls}^{t}(\mathcal{DT}_i^Q)$ can be computed from the target domain query set $\mathcal{DT}_i^Q$. In cross-domain HSI classification, the land-cover classes of the source and target domains are different. Although some land-cover classes may be shared between them, the class conditional distributions of the two domains vary, resulting in variations in the optimization objectives for the classification losses of each domain. The total classification loss in this method is composed of the classification losses from both the source and target domains. If the classification loss of the source domain is significantly larger than that of the target domain, the classification loss from the target domain may be overlooked during training, leading to poor model performance in the target domain. Conversely, if the classification loss of the target domain is much higher than that of the source domain, knowledge transfer from the source domain to the target domain becomes extremely difficult. Thus, maintaining an appropriate balance between these objectives is essential.

A conventional method involves linearly weighting source and target classification losses.

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{cls}^{s} + \lambda_2 \mathcal{L}_{cls}^{t} \tag{5}$$

However, manually adjusting these weight hyperparameters demands considerable manpower and time. Inspired by [43],

CD-MFA utilize an adaptive multi-task loss function based on Gaussian likelihood maximization to automatically determine the weights of source and target classification losses, which is named to as the AWS.

Given input $\mathbf{x}$, let the output of a neural network with parameter $\boldsymbol{W}$ be denoted as $f^{\boldsymbol{W}}(\mathbf{x})$. We compress a scaled version of the model output using a softmax function as follows:

$$p(\mathbf{y}|f^{\boldsymbol{W}}(\mathbf{x}),\sigma) = \mathrm{Softmax}(\frac{1}{\sigma^2}f^{\boldsymbol{W}}(\mathbf{x})) \tag{6}$$

where $\mathbf{y}$ denotes the ground-truth, σ is a positive number, $f^{\boldsymbol{W}}(\mathbf{x})$ is scaled using σ^2. The number σ is adjustable through learning. The log likelihood of this output can be formulated as:

$$\log p(\mathbf{y}|f^{\boldsymbol{W}}(\mathbf{x}),\sigma) = \frac{1}{\sigma^2}f_c^{\boldsymbol{W}}(\mathbf{x}) - \log\sum_{c'}\exp(\frac{1}{\sigma^2}f_{c'}^{\boldsymbol{W}}(\mathbf{x})) \tag{7}$$

where $f_c^{\boldsymbol{W}}(\mathbf{x})$ represents the c-th element of $f^{\boldsymbol{W}}(\mathbf{x})$. There are two discrete outputs, denoted as $\mathbf{y}_1$, and $\mathbf{y}_2$, modeled with softmax likelihoods, respectively. The joint loss $\mathcal{L}(\boldsymbol{W},\sigma_1,\sigma_2)$ is [43]:

$$\begin{aligned}&\mathcal{L}(\boldsymbol{W},\sigma_1,\sigma_2)\\ &= -\log p(\mathbf{y}_1|f^{\boldsymbol{W}}(\mathbf{x}),\sigma_1) - \log p(\mathbf{y}_2|f^{\boldsymbol{W}}(\mathbf{x}),\sigma_2)\\ &\approx \frac{1}{\sigma_1^2}\mathcal{L}_1(\boldsymbol{W}) + \frac{1}{\sigma_2^2}\mathcal{L}_2(\boldsymbol{W}) + \log\sigma_1 + \log\sigma_2\end{aligned} \tag{8}$$

As same as [43], we define $\lambda_j = \log\sigma_j^2 (j=1,2)$. Let $\mathcal{L}_1 = \mathcal{L}_{cls}^s$, $\mathcal{L}_2 = \mathcal{L}_{cls}^t$. The total classification loss in the outer-loop optimization can be expressed as follows:

$$\mathcal{L}_{total}^{cls} = \exp(-\lambda_1)\mathcal{L}_{cls}^s + \exp(-\lambda_2)\mathcal{L}_{cls}^t + \lambda_1/2 + \lambda_2/2 \tag{9}$$

By incorporating the AWS into the outer-loop optimization process, source and target classification losses can be effectively balanced and significantly reduce the search time for weight hyperparameters.

C. Feature Alignment (FA)

In cross-domain HSI classification, the source and target domains are collected by different sensors. Although applying principal component analysis (PCA) can equalize the feature dimensions between the source and target domains, it cannot ensure consistency in their feature distributions. Therefore, it is necessary to design a feature alignment module to align the features between the source and target domains, to reduce the impact of distribution discrepancies on classification performance. Inspired by [44], [45], Kullback-Leibler (KL) divergence is used to measure the discrepancy between two distributions. Due to the inconsistency in land-cover classes between the source and target domains, directly aligning the features is challenging. Therefore, we introduce the Gaussian prior in the feature space. By aligning the features from both domains to the Gaussian prior, the source and target domain features are indirectly aligned, effectively mitigating the effects of domain shift.

In the FA module, KL divergence is utilized to align the features of the source and target domains with the constructed Gaussian prior, respectively

$$\mathcal{L}_{da}^s(\mathcal{DS}_i^Q) = \frac{1}{CM}\sum_{q=1}^{CM}\mathrm{Softmax}(\mathbf{z})\log\frac{\mathrm{Softmax}(\mathbf{z})}{\mathrm{Softmax}(f^s(\mathbf{xs}_i^q))} \tag{10}$$

$$\mathcal{L}_{da}^t(\mathcal{DT}_i^Q) = \frac{1}{CM}\sum_{q=1}^{CM}\mathrm{Softmax}(\mathbf{z})\log\frac{\mathrm{Softmax}(\mathbf{z})}{\mathrm{Softmax}(f^t(\mathbf{xt}_i^q))} \tag{11}$$

$$f^s(\mathbf{xs}_i^q) = FC(Concat(f_{2D}^s(\mathbf{xs}_i^q), f_{1D}^s(\mathbf{xs}_i^q)) \tag{12}$$

$$f^t(\mathbf{xt}_i^q) = FC(Concat(f_{2D}^t(\mathbf{xt}_i^q), f_{1D}^t(\mathbf{xt}_i^q)) \tag{13}$$

where $\mathbf{xs}_i^q$ and $\mathbf{xt}_i^q$ represent samples in the source domain query set $\mathcal{DS}_i^Q$ and target domain query set $\mathcal{DT}_i^Q$, respectively. $f_{2D}^s(\cdot)$ and $f_{2D}^t(\cdot)$ represent the 2-D convolutional branch of the base learner updated after inner-loop optimization in the source and target domains, respectively. $f_{1D}^s(\cdot)$ and $f_{1D}^t(\cdot)$ represent the 1-D convolutional branch of the base learner updated after inner-loop optimization in the source and target domains, respectively. $\mathbf{z}\sim\mathcal{N}(0,1)$ represents the constructed Gaussian prior used for feature alignment. It is worth noting that the feature alignment module uses query set samples, hence there are rich samples.

In the outer-loop optimization, the FA loss is expressed as:

$$\mathcal{L}_{total}^{da} = \mathcal{L}_{da}^s + \mathcal{L}_{da}^t \tag{14}$$

Then, the total loss for the CD-MFA is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_{total}^{cls} + \xi\mathcal{L}_{total}^{da} \tag{15}$$

where ξ denotes the hyperparameter used to balance the classification and FA losses.

Lastly, the gradient descent method is applied to update the model parameters in the outer-loop optimization

$$(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\psi},\lambda) \leftarrow (\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\psi},\lambda) - \beta\nabla_{\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\psi},\lambda}\sum\nolimits_{T_i^s,T_i^t}\mathcal{L}_{total} \tag{16}$$

The detailed training process for CD-MFA and the inner-loop optimization process are described in Algorithm 1 and Algorithm 2, respectively.

D. Implementation Details

According to Eq. (3), we apply affine transformation to dynamically change the meta-learnable loss function's parameters. Nevertheless, as $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ are produced by a neural network, their value range is unpredictable, potentially impacting the stability of model training. To enhance model stability during training, we introduce two learnable parameters, $\boldsymbol{\gamma}_0$ and $\boldsymbol{\eta}_0$, initialized to 0

$$\boldsymbol{\phi}^{'} = (1+\boldsymbol{\gamma}_0\boldsymbol{\gamma})\boldsymbol{\phi} + \boldsymbol{\eta}_0\boldsymbol{\eta} \tag{17}$$

Algorithm 1 CD-MFA algorithm

Input: Task distribution $p_s(T), p_t(T)$, hyperparameters ξ, Learning rates α, β
Output: The parameters of the optimized model

Initialize $\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\psi}, \lambda$
while not done **do**
  Sample batch of tasks $T_i^s \sim p_s(T), T_i^t \sim p_t(T)$
  for each task T_i^s **do**
    Data set $\mathcal{DS}_i^S, \mathcal{DS}_i^Q$ from T_i^s
    for n in number of inner-loop updates N **do**
      Adapt $\boldsymbol{\theta}_{i,n+1}^s \leftarrow \boldsymbol{\theta}_{i,n}$ using Algorithm 2
    end for
    Compute the $\mathcal{L}_{cls}^s(\mathcal{DS}_i^Q)$ on the $\mathcal{DS}_i^Q$
    Compute the $\mathcal{L}_{da}^s(\mathcal{DS}_i^Q)$using (10)
  end for
  for each task T_i^t **do**
    Data set $\mathcal{DT}_i^S, \mathcal{DT}_i^Q$ from T_i^t
    for n in number of inner-loop updates N **do**
      Adapt $\boldsymbol{\theta}_{i,n+1}^t \leftarrow \boldsymbol{\theta}_{i,n}$ using Algorithm 2
    end for
    Compute the $\mathcal{L}_{cls}^t(\mathcal{DT}_i^Q)$ on the $\mathcal{DT}_i^Q$
    Compute the $\mathcal{L}_{da}^t(\mathcal{DT}_i^Q)$ using (11)
  end for
  Compute the $\mathcal{L}_{total}^{cls}$ and $\mathcal{L}_{total}^{da}$ using (9) and (14)
  Compute the total loss:$\mathcal{L}_{total} = \mathcal{L}_{total}^{cls} + \xi\mathcal{L}_{total}^{da}$
  Gradient descent is performed to update the weights: $(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\psi}, \lambda) \leftarrow (\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\psi}, \lambda) - \beta\nabla_{\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\psi},\lambda} \sum_{T_i^s,T_i^t} \mathcal{L}_{total}$
end while

Algorithm 2 Inner-loop optimization algorithm

Input: Parameters $\boldsymbol{\theta}_{i,n}, \boldsymbol{\phi}_{i,n}, \boldsymbol{\psi}_{i,n}$ of the f, l, g, Data set $\mathcal{D}_i^S$
Output: The parameters of the optimized base learner

Compute the $\mathcal{L}_{CE}(\mathcal{D}_i^S)$ on the $\mathcal{D}_i^S$
Compute the task state:$\boldsymbol{\tau}_{i,n} = [\mathcal{L}_{CE}(\mathcal{D}_i^S), \boldsymbol{\theta}_{i,n}, f(\mathbf{x}_i^s)]$
Compute the affine transformation parameters: $\boldsymbol{\gamma}_{i,n}, \boldsymbol{\eta}_{i,n} = g(\boldsymbol{\tau}_{i,n})$
Adapt the loss function parameters: $\boldsymbol{\phi}_{i,n}^{'} = \boldsymbol{\gamma}_{i,n}\boldsymbol{\phi}_{i,n} + \boldsymbol{\eta}_{i,n}$
Compute task-adaptive loss:$\mathcal{L}_{\boldsymbol{\phi}_{i,n}^{'}}(\boldsymbol{\tau}_{i,n})$
Gradient descent is performed to adapt f: $\boldsymbol{\theta}_{i,n+1} = \boldsymbol{\theta}_{i,n} - \alpha\nabla_{\boldsymbol{\theta}_{i,n}}\mathcal{L}_{\boldsymbol{\phi}_{i,n}^{'}}(\boldsymbol{\tau}_{i,n})$

TABLE II
LAND-COVER CLASSES AND NUMBER OF SAMPLES IN CHIKUSEI DATASET

Class	Name	Samples
1	Water	2845
2	Bare soil (school)	2859
3	Bare soil (park)	286
4	Bare soil (farmland)	4852
5	Natural plants	4297
6	Weeds in farmland	1108
7	Forest	20516
8	Grass	6515
9	Rice field (grown)	13369
10	Rice field (first stage)	1268
11	Row crops	5961
12	Plastic house	2193
13	Manmade (non-dark)	1220
14	Manmade (dark)	7664
15	Manmade (blue)	431
16	Manmade (red)	222
17	Manmade grass	1040
18	Asphalt	801
19	Paved ground	145
	Total	77592

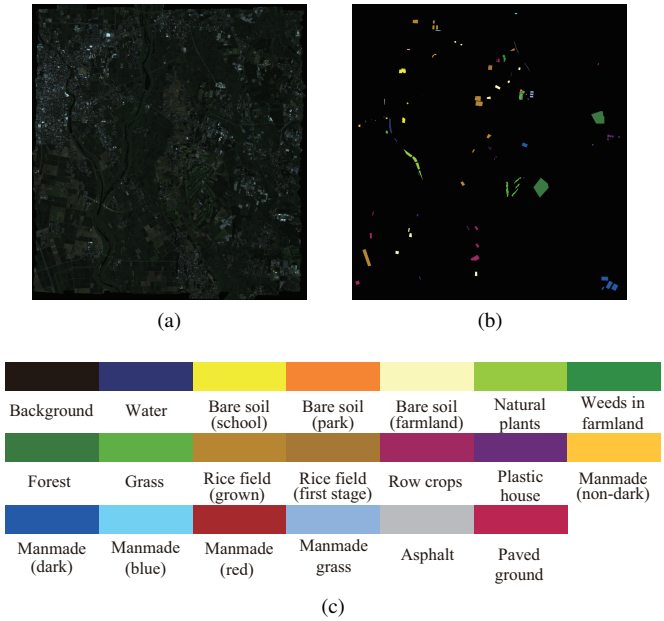

Fig. 5. Chikusei dataset. (a) False-color image with bands 29, 39, 49. (b) Ground truth map. (c) Color coding.

IV. EXPERIMENTS

A. Experimental Datasets

To validate the performance of CD-MFA, we conducted experiments utilizing four public HSI datasets: Chikusei [46], Indian Pines (IP), University of Pavia (UP), and Salinas (SA) datasets. The Chikusei dataset is chosen as the source domain dataset, while the IP, UP, and SA datasets are selected as the target domain datasets.

1) Source Domain: The Chikusei dataset was captured in Chikusei, Ibaraki, Japan by a Headwall Hyperspec-VNIR-C imaging sensor. The Chikusei dataset comprises 128 spectral bands, ranging from 363 to 1018nm, with a total of 2517×2335 pixels, each with a spatial resolution of 2.5m. It contains nineteen unique land-cover classes. Table II displays the land-cover classes and their respective number of samples in the Chikusei dataset, while Fig. 5 shows the false-color images and ground truth maps.

2) Target Domain: The IP dataset was collected by an AVIRIS sensor in 1992 from Northwest Indiana. It comprises 200 spectral bands ranging from 400 to 2500nm. It consists 145×145 pixels, offering a spatial resolution of 20m, and encompasses sixteen unique land-cover classes. Table III describes the land-cover classes with their corresponding sample numbers in the IP dataset. The false-color image and ground truth map are presented in Fig. 6.

The UP dataset was acquired using a ROSIS sensor. It comprises 103 spectral bands, ranging from 430 to 860 nm. It

TABLE III
LAND-COVER CLASSES AND NUMBER OF SAMPLES IN INDIAN PINES DATASET

Class	Name	Samples
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-mintill	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pasture-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-mintill	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings-Grass-Trees-Drives	386
16	Stone-Steel-Towers	93
	Total	10249

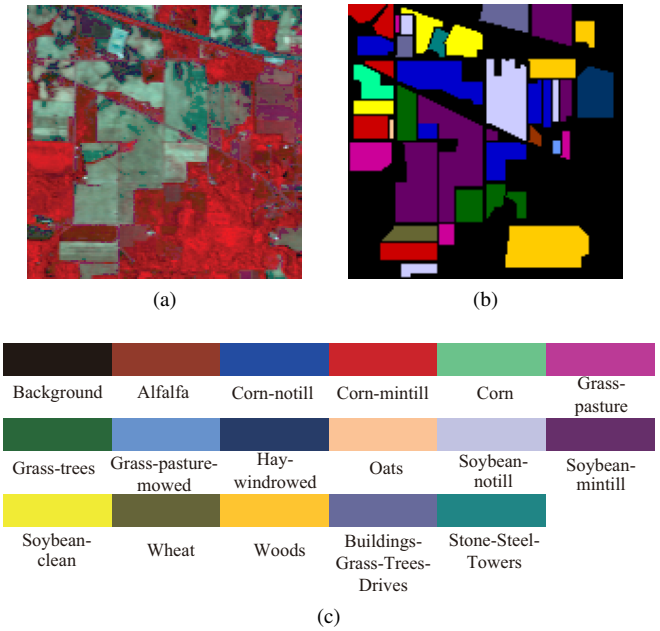

Fig. 6. Indian Pines dataset. (a) False-color image with bands 19, 29, 39. (b) Ground truth map. (c) Color coding.

consists of 610×340 pixels with a spatial resolution of 1.3 m. There are nine unique land-cover classes within this dataset. Table IV shows the land-cover classes of the UP dataset and their corresponding sample numbers. The false-color image and ground truth map are shown in Fig. 7.

The SA dataset was captured by the AVIRIS sensor in Salinas Valley, CA, USA. The SA dataset comprises 204 spectral bands ranging from 400 to 2500 nm. It consists of 512×217 pixels, each with a spatial resolution of 3.7 m. It includes sixteen unique land-cover classes. Table V shows the land-cover classes of the SA dataset and their corresponding number of samples. Fig. 8 shows the false-color image and ground truth map.

TABLE IV
LAND-COVER CLASSES AND NUMBER OF SAMPLES IN UNIVERSITY OF PAVIA DATASET

Class	Name	Samples
1	Asphalt	6631
2	Meadows	18649
3	Gravel	2099
4	Trees	3064
5	Painted metal sheets	1345
6	Bare Soil	5029
7	Bitumen	1330
8	Self-Blocking Bricks	3682
9	Shadows	947
	Total	42776

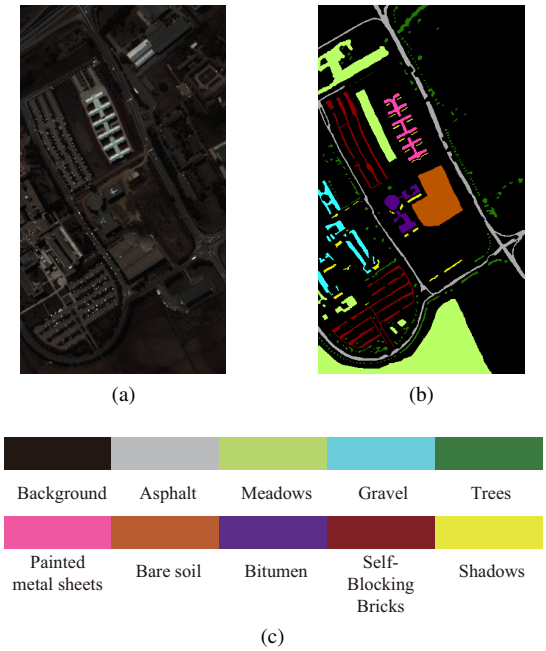

Fig. 7. University of Pavia dataset. (a) False-color image with bands 10, 20, 30. (b) Ground truth map. (c) Color coding.

B. Experimental Settings

To validate the efficacy of our proposed method, we compared CD-MFA with the following seven methods.

1) Support vector machine (SVM): SVM utilizes raw spectral features to classify HSI. Within SVM, we designate the penalty parameters as $C \in \{0.001, 0.01 \cdots, 100\}$ and the parameters of the Radial Basis Function (RBF) kernel as $R \in \{1, 2, 4, 8, 16, 32, 64\}$.

2) Hybrid Spectral Convolutional Neural Network (HybridSN) [47]: HybridSN is a hybrid model combining 3-D and 2-D Convolutional Neural Networks (CNNs) for single-scene HSI classification. The method fully utilizes the spatial and spectral information by combining 3-D and 2-D CNNs. Even with limited training data, this method shows good performance.

3) Deep Cross-Domain Few-Shot Learning (DCFSL) [48]: DCFSL is a method aimed at addressing cross-domain FSL challenges in HSI classification. It employs a conditional adversarial domain adaptation strategy to mitigate domain shift and conducts FSL in both the source and target domains. This

TABLE V
LAND-COVER CLASSES AND NUMBER OF SAMPLES IN SALINAS DATASET

Class	Name	Samples
1	Brocoli_green_weeds_1	2009
2	Brocoli_green_weeds_2	3726
3	Fallow	1976
4	Fallow_rough_plow	1394
5	Fallow_smooth	2678
6	Stubble	3959
7	Celery	3579
8	Grapes_untrained	11271
9	Soil_vinyard_develop	6203
10	Corn_senesced_green_weeds	3278
11	Lettuce_romaine_4wk	1068
12	Lettuce_romaine_5wk	1927
13	Lettuce_romaine_6wk	916
14	Lettuce_romaine_7wk	1070
15	Vinyard_untrained	7268
16	Vinyard_vertical_trellis	1807
	Total	54129

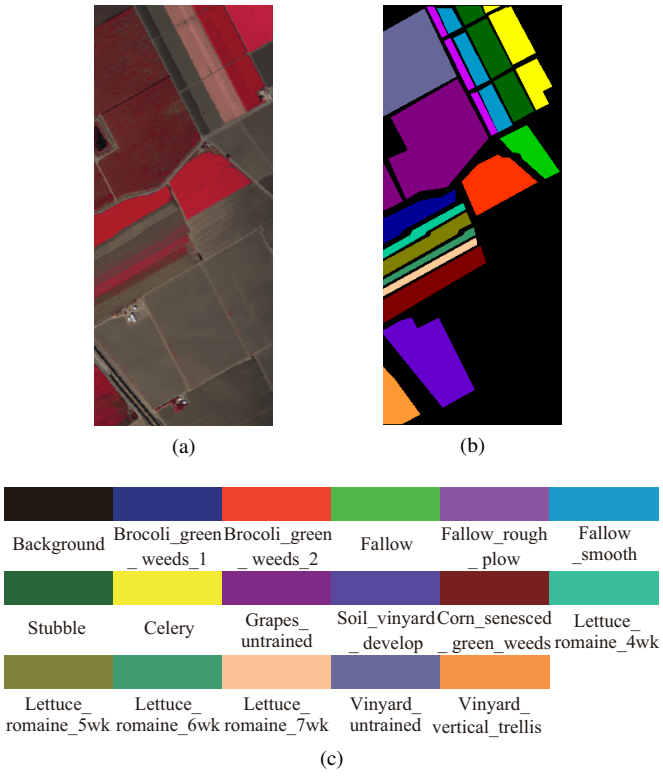

Fig. 8. Salinas dataset. (a) False-color image with bands 19, 29, 39. (b) Ground truth map. (c) Color coding.

approach enables the acquisition of transferable knowledge from the source domain while simultaneously learning a discriminative embedding space in the target domain. Compared to traditional FSL methods, DCFSL has a better performance.

4) Graph information aggregation cross-domain few-shot learning (Gia-CFSL) [49]: Gia-CFSL addresses cross-domain HSI classification challenges by fusing FSL and domain adaptation. To achieve domain alignment, Gia-CFSL designs an intradomain distribution extraction block (IDE-block) and a cross-domain similarity aware block (CSA-block).

5) Refined prototypical contrastive learning network for few-shot learning (RPCL-FSL) [50]: RPCL-FSL addresses the issue of prototype instability by applying triple constraints on the prototypes of support set samples and introduces a fusion training strategy to mitigate domain shift.

6) Feature disentanglement few-shot learning (FDFSL) [51]: By decoupling diverse feature components for few-shot learning, FDFSL effectively suppresses metaknowledge biased toward the source domain and facilitates the learning of target domain specific information. Furthermore, this method proposes a self-distillation-based feature constraint strategy to prevent the model from overfitting to domain alignment tasks.

7) Graph convolution contrast few-shot learning (GCC-FSL) [52]: GCC-FSL proposes a positive and negative pairs module to align the data distributions of the source and target domains. Additionally, this method introduces a graph convolution contrast module, which not only enhances the ability of feature expression but also further mitigates the domain shift problem.

For the single-domain algorithms (SVM and HybridSN methods), only training samples from the target domain are utilized because the algorithm's training class must be the same as that of the test. In cross-domain algorithms (DCFSL, Gia-CFSL, RPCL-FSL, FDFSL, GCC-FSL, and CD-MFA methods), transferable knowledge is learned from source domain samples to aid in target domain classification. From the source domain, we randomly select 200 labeled samples from each class for training. To ensure fair comparison, 5 labeled samples are randomly chosen from each class in the target domain as training samples. The remaining samples in the target domain are used as testing samples. Subsequently, in deep learning methods, Gaussian random noise is uniformly applied as the data augmentation method for the target domain training samples [48], [49].

In cross-domain HSI classification, the source and target domains exhibit heterogeneous features. To address this issue, CD-MFA employs PCA to preprocess the HSI data. We retain the first 100 principal components of the spectral dimension to ensure consistency between the source and target domains.

Unlike the training methods of traditional deep learning, CD-MFA employs episodic training. Each episode during the training phase is a C-way K-shot task. C denotes the number of classes in the target domain. $K = 1$ signifies the number of samples per class in the support sets $\mathcal{DS}_i^S$ and $\mathcal{DT}_i^S$. Furthermore, the number of samples for each class in the query sets $\mathcal{DS}_i^Q$ and $\mathcal{DT}_i^Q$ is $M = 19$. We perform $N = 5$ updates for the inner-loop optimization, and the number of training iterations is 1000. Additionally, the patch size is set to 9×9. The learning rate of the inner-loop optimization α is set to 0.05. The initial learning rate of the outer-loop optimization β_1 for the base learner, the meta-learnable loss function, and the meta-learner is set to 0.005, while the initial learning rate of the outer-loop optimization β_2 for the AWS is set to 0.01.

To evaluate the classification performance of all methods, we employed three quality indicators: the overall accuracy (OA), the average accuracy (AA), and the kappa coefficient (κ). The experimental results are derived from the average of 10 independent runs to eliminate the influence of random sampling.

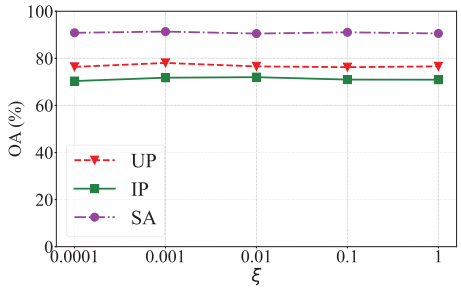

Fig. 9. The impact of hyperparameter ξ on the overall accuracy (OA) of the UP, IP, and SA datasets.

C. Experiments of Hyperparameter Settings

In CD-MFA, the hyperparameter ξ is crucial for balancing classification and FA losses, thus requiring an in-depth discussion. We set $\xi \in \{0.0001, 0.001, 0.01, 0.1, 1\}$. To evaluate the effect of the hyperparameter ξ on model performance, we use OA to assess the performance of CD-MFA on the UP, IP, and SA datasets. The experimental results are displayed in Fig. 9. As shown in Fig. 9, the experimental results on the three datasets are insensitive to changes in the hyperparameter ξ, demonstrating the robustness of CD-MFA to varying hyperparameter ξ settings. On the UP dataset, the best performance is obtained when the hyperparameter is set to $\xi = 0.001$. On the IP dataset, the best performance is obtained when the hyperparameter is set to $\xi = 0.01$. On the SA dataset, the best performance is obtained when the hyperparameter is set to $\xi = 0.001$.

D. Comparison With Different Algorithms

After delving into the significant hyperparameters within CD-MFA, a comparison of the performance of different algorithms ensues. Table VI illustrates the results of all algorithms on the IP dataset. Based on the analysis of the classification results depicted in Table VI, the following summary is drawn.

1) The experimental results in Table VI demonstrate that the performance of the cross-domain classification methods (DCFSL, Gia-CFSL, RPCL-FSL, FDFSL, GCC-FSL, and CD-MFA) exceeds that of the single-domain classification methods (SVM and HybridSN). This indicates that transferring knowledge from the source domain to the target domain effectively enhances the classification performance in the target domain.
2) The performance of CD-MFA surpasses that of DCFSL, Gia-CFSL, RPCL-FSL, FDFSL, and GCC-FSL, highlighting the advantages of CD-MFA. The advantages of CD-MFA can be summarized in two aspects: on the one hand, the introduction of the task-adaptive loss function into the inner-loop optimization improves the optimization process and subsequently enhances the model's generalization performance. On the other hand, in the outer-loop optimization, the AWS is used to balance the classification losses of the source and target domains, while the FA module is employed to reduce domain shift, thereby improving the model's classification performance.
3) Of all the algorithms, CD-MFA achieves the best classification performance, demonstrating its effectiveness.

TABLE VI
OA, AA AND κ OF EACH ALGORITHM ON IP DATASET.

Method	OA (%)	AA (%)	κ ($\times$100)
SVM	41.90	50.44	35.48
HybridSN	53.25	71.91	48.71
DCFSL	63.75	75.82	59.34
Gia-CFSL	64.21	74.74	59.78
RPCL-FSL	64.23	76.50	59.86
FDFSL	64.55	76.83	60.42
GCC-FSL	64.02	75.02	59.55
CD-MFA	**72.00**	**83.30**	**68.64**

TABLE VII
OA, AA AND κ OF EACH ALGORITHM ON UP DATASET.

Method	OA (%)	AA (%)	κ ($\times$100)
SVM	56.29	56.16	45.95
HybridSN	50.74	59.64	40.80
DCFSL	76.19	79.73	69.71
Gia-CFSL	77.40	80.75	71.23
RPCL-FSL	75.09	80.59	68.59
FDFSL	76.98	**84.26**	71.12
GCC-FSL	76.50	81.23	70.13
CD-MFA	**78.04**	82.58	**71.99**

After analyzing the classification results of the IP dataset, the classification results of the UP dataset are analyzed, with classification results shown in Table VII. The summary is as follows.

1) Observing the experimental results in Table VII, we find that the classification accuracy of the cross-domain classification methods (DCFSL, Gia-CFSL, RPCL-FSL, FDFSL, GCC-FSL, and CD-MFA) remains superior to that of the single-domain classification methods (SVM and HybridSN), demonstrating the effectiveness of transfer learning in addressing small-size-sample problem.
2) Compared to DCFSL, Gia-CFSL, FDFSL, GCC-FSL, and RPCL-FSL, CD-MFA achieves higher OA, underscoring that CD-MFA enhances the model's generalization performance through a task-adaptive loss function. In the outer-loop optimization, AWS effectively balances the classification loss between the two domains, while the FA module efficiently aligns the features of the source and target domains.
3) Compared to other methods, CD-MFA demonstrates superior OA, further highlighting its advantages.

Finally, the experimental results on the SA dataset are analyzed. The classification results of different algorithms on the SA dataset are shown in Table VIII. The summary is as follows.

1) According to Table VIII, similar to the experimental results from the IP and UP datasets, the cross-domain classification methods (DCFSL, Gia-CFSL, RPCL-FSL, FDFSL, GCC-FSL, and CD-MFA) continue to outperform the single-domain classification methods (SVM and HybridSN) on the SA dataset. This further emphasizes the effectiveness of transfer learning.

TABLE VIII
OA, AA AND κ OF EACH ALGORITHM ON SA DATASET.

Method	OA (%)	AA (%)	κ ($\times$100)
SVM	76.02	84.21	73.58
HybridSN	80.89	90.02	79.03
DCFSL	89.30	93.84	88.12
Gia-CFSL	88.46	92.31	87.18
RPCL-FSL	88.16	92.89	86.84
FDFSL	90.04	94.02	88.94
GCC-FSL	88.11	92.32	86.79
CD-MFA	**91.36**	**95.62**	**90.39**

TABLE IX
ABLATION STUDY

#	Baseline	Task-adaptive loss function	AWS	FA module	OA (%)		
					UP	IP	SA
1	✓				73.57	68.06	89.50
2	✓	✓			76.45	70.81	90.73
3	✓	✓	✓		77.70	71.72	91.01
4	✓	✓	✓	✓	**78.04**	**72.00**	**91.36**

2) As shown in Table VIII, CD-MFA exhibits superior classification performance compared to DCFSL, Gia-CFSL, RPCL-FSL, FDFSL, and GCC-FSL, further demonstrating the effectiveness of the task-adaptive loss function, AWS, and FA module in CD-MFA.
3) The classification performance of the CD-MFA proposed in this paper surpasses that of other comparison algorithms, demonstrating the effectiveness of combining meta-learning and transfer learning for cross-domain heterogeneous HSI classification.

E. Ablation Study

We conducted an ablation study on the three key components of CD-MFA: task-adaptive loss function, AWS, and FA module. The experiment is performed on the UP, IP, and SA datasets, with OA as the evaluation metric. We used a model with the cross-entropy loss in the inner-loop optimization as the baseline. This baseline is a cross-domain meta-learning model for HSI classification. Additionally, in this baseline model, the task-adaptive loss function is not comprised of inner-loop optimization, while the FA module and AWS are not comprised of the outer-loop optimization. Table IX presents the experimental results from four ablation studies. Based on the experimental results, the specific analysis is as follows.

1) *Effects on task-adaptive loss function:* Table IX shows that incorporating the task-adaptive loss function results in OA increases of 2.88%, 2.75%, and 1.23% on the UP, IP, and SA datasets, respectively. These results indicate that the task-adaptive loss function improves the inner-loop optimization process, thereby enhancing the model's generalization performance.
2) *Effects on AWS:* Observing the experimental results in Table IX shows that AWS improves performance on the UP, IP, and SA datasets by 1.05%, 0.91%, and 0.28%, respectively, compared to the approach without AWS. It is demonstrated that the AWS effectively balances the classification losses between the source and target domains, thereby facilitating the learning of more discriminative features and improving the model's classification performance.
3) *Effects on FA module:* As shown in Table IX, the FA module performs well on the three datasets, with OA increasing by 0.34%, 0.28%, and 0.35% on the UP, IP, and SA datasets, respectively. This indicates that the FA module can effectively align the features of the source and target domains, thereby reducing the impact of domain shift on classification performance.

F. Training Stability Analysis

To evaluate the training stability of the proposed CD-MFA, we plotted the training loss curves under four different random seeds, each corresponding to a distinct initialization. As illustrated in Fig. 10, the total loss fluctuates during training due to differences in task difficulties, yet it demonstrates a consistent overall decreasing trend across all four initializations on various datasets. This behavior indicates that the proposed CD-MFA exhibits robust training stability.

G. Comparison on Computational Complexity

We record the computational complexity of the comparison methods, including SVM, HybridSN, DCFSL, Gia-CFSL, RPCL-FSL, FDFSL, GCC-FSL, and our proposed CD-MFA. The training time, the inference time, and the number of parameters for all compared methods are listed in Table X. The results indicate that the training time of the cross-domain classification methods (DCFSL, Gia-CFSL, RPCL-FSL, FDFSL, GCC-FSL, and CD-MFA) is longer than that of the single-domain classification methods (SVM and HybridSN). This is because the cross-domain method requires training samples from the source domain. Compared with RPCL-FSL and FDFSL, CD-MFA requires more training time and involves more parameters due to the need for multiple inner-loop optimizations, but it remains within an acceptable range. Besides, it is worth noting that the inference time of our proposed method is faster than other methods, proving the feasibility of CD-MFA in practical application.

V. CONCLUSION AND FUTURE WORK

This paper proposes a novel heterogeneous cross-domain HSI classification algorithm called CD-MFA. CD-MFA comprises two main components: 1) Inner-loop optimization aims to enable the base learner to adapt effectively to new tasks with limited samples. 2) Outer-loop optimization focuses on training the model. The advantages of CD-MFA are as follows: 1) CD-MFA employs the task-adaptive loss function to enhance the model's generalization performance in the inner-loop optimization. 2) During the outer-loop optimization process, the AWS is employed to balance the classification losses between the source and target domains, which facilitates knowledge transfer and learning more discriminative features. 3) Additionally, the FA module is utilized to align the features of the source and target domains, thereby mitigating the impact of domain shift on classification performance. Our experiments

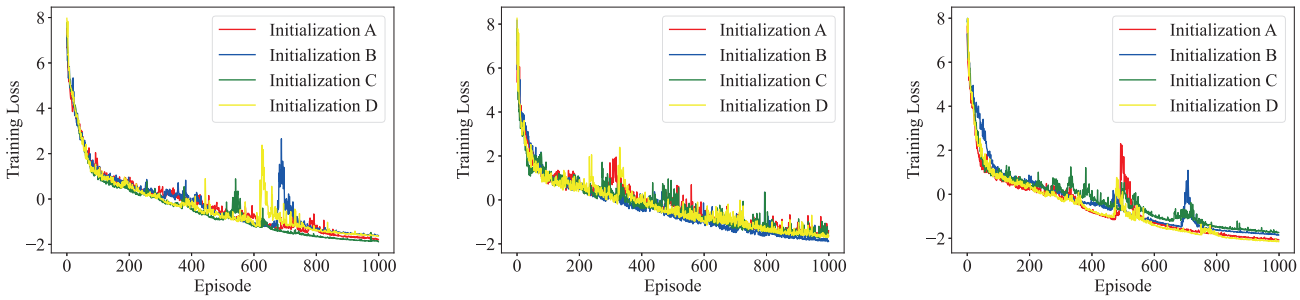

Fig. 10. Training loss curves across different random seeds. (a) IP dataset. (b) UP dataset. (c) SA dataset.

TABLE X
COMPUTATIONAL TIME (SECONDS) AND PARAMETERS (MILLION) OF DIFFERENT METHODS

Method	IP Dataset			UP Dataset			SA Dataset		
	Training Time (s)	Inference Time (s)	Parameters (M)	Training Time (s)	Inference Time (s)	Parameters (M)	Training Time (s)	Inference Time (s)	Parameters (M)
SVM	0.09	0.01	-	0.15	0.02	-	0.47	0.03	-
HybridSN	407.27	0.43	1.69	237.86	2.56	1.69	407.77	2.04	1.69
DCFSL	1450.85	0.89	4.36	1027.72	3.12	4.35	1501.32	4.76	4.36
Gia-CFSL	3094.98	0.91	1.46	2174.36	3.30	1.44	3249.91	5.11	1.46
RPCL-FSL	414.34	0.90	0.32	298.80	3.15	0.31	475.34	4.76	0.32
FDFSL	896.18	1.84	0.56	715.94	9.34	0.55	1064.66	17.52	0.55
GCC-FSL	4032.23	0.98	0.91	2003.28	3.44	0.89	4163.26	5.45	0.91
CD-MFA	1101.36	0.32	2.33	884.44	1.44	2.32	1108.38	1.88	2.33

conducted on four publicly available HSI datasets demonstrate that CD-MFA exhibits better performance in heterogeneous cross-domain HSI classification.

Although CD-MFA exhibits strong cross-domain transfer performance, there is nevertheless potential for further enhancement. In its current inner-loop design, there is no alignment between the two domains. Future research could focus on integrating domain adaptation techniques into the inner-loop optimization process to promote better domain alignment and enhance overall classification performance.

REFERENCES

[1] F. Cao and W. Guo, "Cascaded dual-scale crossover network for hyperspectral image classification," *Knowl-based. Syst.*, vol. 189, pp. 1–11, 2020, Art. no. 105122.

[2] Q. Qian, X. Fan, and M. Ye, "Improving hyperspectral image classification using graph wavelets," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 842–845.

[3] A. Sellami and S. Tabbone, "Deep neural networks-based relevant latent representation learning for hyperspectral image classification," *Pattern Recogn.*, vol. 121, pp. 1–13, 2022, Art. no. 108224.

[4] C. Zhao, B. Qin, S. Feng, W. Zhu, W. Sun, W. Li, and X. Jia, "Hyperspectral image classification with multi-attention transformer and adaptive superpixel segmentation-based active learning," *IEEE Trans. Image Process.*, pp. 3606–3621, 2023.

[5] Z. Zhang, Y. Ding, X. Zhao, L. Siye, N. Yang, Y. Cai, and Y. Zhan, "Multireceptive field: An adaptive path aggregation graph neural framework for hyperspectral image classification," *Expert Systems with Applications*, vol. 217, pp. 1–15, 2023, Art. no. 119508.

[6] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, 2011.

[7] J. Li, X. Huang, P. Gamba, J. M. Bioucas-Dias, L. Zhang, J. A. Benediktsson, and A. Plaza, "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, 2015.

[8] X. Yang, Y. Ye, X. Li, R. Y. K. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5408–5423, 2018.

[9] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, 2020.

[10] F. Cao and W. Guo, "Deep hybrid dilated residual networks for hyperspectral image classification," *Neurocomputing*, vol. 384, pp. 170–181, 2020.

[11] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, "Spectral-spatial feature tokenization transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022, Art. no. 5522214.

[12] M. Ye, W. Zheng, H. Lu, X. Zeng, and Y. Qian, "Cross-scene hyperspectral image classification based on DWT and manifold-constrained subspace learning," *Int. J. Wavelets Multiresolution Inf. Process.*, vol. 15, no. 06, p. 1750062, 2017.

[13] B. Yang, S. Hu, Q. Guo, and D. Hong, "Multisource domain transfer learning based on spectral projections for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3730–3739, 2022.

[14] M. Zhang, H. Liu, M. Gong, H. Li, Y. Wu, and X. Jiang, "Cross-domain self-taught network for few-shot hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–19, 2023, Art. no. 4501719.

[15] Y. Yang, Y. Xu, Z. Wu, B. Wang, and Z. Wei, "Cross-scene classification of hyperspectral images via generative adversarial network in latent space," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–17, 2023, Art. no. 5526217.

[16] M. Ye, C. Wang, Z. Meng, F. Xiong, and Y. Qian, "Domain-invariant attention network for transfer learning between cross-scene hyperspectral images," *IET Comput. Vision*, vol. 17, no. 7, pp. 739–749, 2023.

[17] Y. Qu, R. K. Baghbaderani, W. Li, L. Gao, Y. Zhang, and H. Qi, "Physically constrained transfer learning through shared abundance space for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10 455–10 472, 2021.

[18] Z. Liu, L. Ma, and Q. Du, "Class-wise distribution adaptation for unsupervised classification of hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 508–521, 2020.

[19] 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, pp. 508–521, 2023, Art. no. 5301814.

[20] A. Wang, C. Liu, H. Zhou, Y. Song, H. Wu, and Y. Iwahori, "Cross-scene hyperspectral image classification based on feature learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.* IEEE, 2022, pp. 3568–3571.

[21] M. Ye, Y. Xu, C. Ji, H. Chen, H. Lu, and Y. Qian, "Feature selection for cross-scene hyperspectral image classification using cross-domain ReliefF," *Int. J. Wavelets Multiresolution Inf. Process.*, vol. 17, no. 05, p. 1950039, 2019.

[22] M. Ye, J. Chen, F. Xiong, and Y. Qian, "Learning a deep structural subspace across hyperspectral scenes with cross-domain VAE," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, Art. no. 5522613.

[23] L. Lei, B. Huang, M. Ye, F. Yao, and Y. Qian, "Cross-domain residual deep NMF for transfer learning between different hyperspectral image scenes," *Int. J. Wavelets Multiresolution Inf. Process.*, vol. 21, no. 02, p. 2250046, 2023.

[24] M. Ye, J. Chen, F. Xiong, and Y. Qian, "Adaptive graph modeling with self-training for heterogeneous cross-scene hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–15, 2024, Art. no. 5503815.

[25] C. Zhong, J. Zhang, S. Wu, and Y. Zhang, "Cross-scene deep transfer learning with spectral feature adaptation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2861–2873, 2020.

[26] X. Wang, Y. Li, and Y. Cheng, "Hyperspectral image classification based on unsupervised heterogeneous domain adaptation CycleGan," *Chinese J. Electron.*, vol. 29, no. 4, pp. 608–614, 2020.

[27] C. Zhong, J. Zhang, Q. Guo, and Y. Zhang, "Heterogeneous spectral-spatial feature transfer with structure preserved distribution alignment for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5545–5558, 2022.

[28] Y. Cheng, W. Zhang, H. Wang, and X. Wang, "Causal meta-transfer learning for cross-domain few-shot hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–14, 2023, Art. no. 5521014.

[29] W. Li, Q. Liu, Y. Zhang, Y. Wang, Y. Yuan, Y. Jia, and Y. He, "Few-shot hyperspectral image classification using meta learning and regularized finetuning," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–14, 2023, Art. no. 5529514.

[30] L. Hu, W. He, L. Zhang, and H. Zhang, "Cross-domain meta-learning under dual-adjustment mode for few-shot hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–16, 2023, Art. no. 5526416.

[31] M. Huisman, J. N. Van Rijn, and A. Plaat, "A survey of deep meta-learning," *Artif. Intell. Rev.*, vol. 54, no. 6, pp. 4483–4541, 2021.

[32] 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.

[33] Y. Zhou, J. Lian, and M. Han, "Remote sensing image transfer classification based on weighted extreme learning machine," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 10, pp. 1405–1409, 2016.

[34] L. Zhu and L. Ma, "Class centroid alignment based domain adaptation for classification of remote sensing images," *Pattern Recogn. Lett.*, vol. 83, pp. 124–132, 2016.

[35] Q. Qian and T. Mallick, "Wavelet-inspired multiscale graph convolutional recurrent network for traffic forecasting," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2024, pp. 5680–5684.

[36] 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, 2021.

[37] Y. Zhu, F. Zhuang, J. Wang, G. Ke, J. Chen, J. Bian, H. Xiong, and Q. He, "Deep subdomain adaptation network for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1713–1722, 2021.

[38] H. Wang, Y. Cheng, C. L. P. Chen, and X. Wang, "Hyperspectral image classification based on domain adversarial broad adaptation network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, Art. no. 5517813.

[39] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[40] T. Munkhdalai and H. Yu, "Meta networks," in *Proc. Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 2554–2563.

[41] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1126–1135.

[42] S. Baik, J. Choi, H. Kim, D. Cho, J. Min, and K. M. Lee, "Meta-learning with task-adaptive loss function for few-shot learning," in *Proc. IEEE int. Conf. Comput. Vis.*, 2021, pp. 9465–9474.

[43] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proc. IEEE conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7482–7491.

[44] J. Lei, S. Fang, W. Xie, Y. Li, and C.-I. Chang, "Discriminative reconstruction for hyperspectral anomaly detection with spectral learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7406–7417, 2020.

[45] Z. Leng, M. Wang, Q. Wan, Y. Xu, B. Yan, and S. Sun, "Meta-learning of feature distribution alignment for enhanced feature sharing," *Knowledge-based Systems*, pp. 1–11, 2024, Art. no. 111875.

[46] N. Yokoya and A. Iwasaki, "Airborne hyperspectral data over Chikusei," Space Appl. Lab., Univ. Tokyo, Tokyo, Japan, Tech. Rep., May 2016.

[47] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 277–281, 2020.

[48] 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, no. 99, pp. 1–18, 2022, Art. no. 5501618.

[49] Y. Zhang, W. Li, M. Zhang, S. Wang, R. Tao, and Q. Du, "Graph information aggregation cross-domain few-shot learning for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 1912–1925, 2024.

[50] Q. Liu, J. Peng, Y. Ning, N. Chen, W. Sun, Q. Du, and Y. Zhou, "Refined prototypical contrastive learning for few-shot hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–14, 2023, Art. no. 5506214.

[51] B. Qin, S. Feng, C. Zhao, W. Li, R. Tao, and W. Xiang, "Cross-domain few-shot learning based on feature disentanglement for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–15, 2024, Art. no. 5514215.

[52] Z. Ye, J. Wang, T. Sun, J. Zhang, and W. Li, "Cross-domain few-shot learning based on graph convolution contrast for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–14, 2024, Art. no. 5504614.

Minchao Ye (Member, IEEE) received the B.E. degree in computer science and technology from Sichuan University, Chengdu, China, in 2010, and the Ph.D. degree in computer science and technology from Zhejiang University, Hangzhou, China, in 2016. Since 2016, he has been with the College of Information Engineering, China Jiliang University, Hangzhou, China, where he is currently an Associate Professor of computer science and technology.

His research interests encompass hyperspectral image processing, deep learning, and pattern recognition.

Yuheng Jin received the B.E. in Communication Engineering from Henan Polytechnic University, Jiaozuo, China, in 2021 and the M.E. in Electronic and Information Engineering from China Jiliang University, Hangzhou, China, in 2025. He joined Hangzhou Jingsheng Hangxing Technology Company, Hangzhou, China, in 2025, as a Technical Position.

His research interests include hyperspectral image processing and deep learning.

Jianwei Zhao received the Ph.D. degree from the Chinese Academy of Sciences, on the subject of mathematics, in 2006. She was a postdoctoral researcher with the Department of Electronics and Information Engineering, Chonbuk National University, South Korea, in 2011. She is currently a professor with the College of Information Engineering, China Jiliang University. Her research interests include machine learning, image processing, and computer vision. She has published more than 100 scientific papers and these papers have been SCI-cited more than 660 times in the Web of Sciences. She is a member of Professional Committee of Machine Learning of Chinese Association for Artificial Intelligence, and also a member of Professional Committee of Computer Vision of China Computer Federation.

Wei Qi Yan 's expertise covers intelligent robotics, deep learning, computer vision, intelligent surveillance, and multimedia computations. Dr. Yan is an Associate Editor of IEEE Transactions on Circuits and Systems for Video Technology, an Associate Editor of ACM Transactions on Multimedia Computing, Communications and Applications, a Senior Area Editor of IEEE Signal Processing Letters, a Section Editor of Springer Journal Discover Artificial Intelligence, he has worked as an exchange computer scientist between the Royal Society Te Apārangi (RSNZ) and the Chinese Academy of Sciences (CAS) in China. Dr. Yan is recognised as one of the "Top Two Percent of Scientists in the World" by Stanford University, he currently holds the position of Chair of ACM Multimedia Chapter of New Zealand, a Full Board Member of AI Researchers Association (NZ), he is a Fellow of Engineering New Zealand Te Ao Rangahau (FEngNZ).

Yuntao Qian (Senior Member, IEEE) received the B.E. and M.E. degrees in automatic control from Xi'an Jiaotong University, Xi'an, China, in 1989 and 1992, respectively, and the Ph.D. degree in signal processing from Xidian University, Xi'an, China, in 1996.

From 1996 to 1998, he was a Post-Doctoral Fellow with Northwestern Polytechnical University, Xi'an. Since 1998, he has been with the College of Computer Science, Zhejiang University, Hangzhou, China, where he became a Professor in 2002. He was a Visiting Professor with Concordia University, Montreal, QC, Canada, from 1999 to 2001; Hong Kong Baptist University, Hong Kong, China, in 2006; Carnegie Mellon University, Pittsburgh, PA, USA, in 2010; the Canberra Research Laboratory, NICTA, Canberra, ACT, Australia, in 2013; the University of Macau, Macau, China, from 2015 to 2016; and Griffith University, Nathan, QLD, Australia, in 2018. His research interests include machine learning, signal and image processing, pattern recognition, and hyperspectral imaging.

Prof. Qian is currently an Associate Editor of IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.