Few-shot Learning with Class-Covariance Metric for Hyperspectral Image Classification

Bobo Xi, *Member, IEEE,* Jiaojiao Li, *Member, IEEE,* Yunsong Li, *Member, IEEE,* Rui Song, *Member, IEEE,* Danfeng Hong, *Senior Member, IEEE* and Jocelyn Chanussot, *Fellow, IEEE*

***Abstract*—Recently, embedding and metric-based few-shot learning (FSL) has been introduced into hyperspectral image classification (HSIC) and achieved impressive progress. To further enhance the performance with few labeled samples, we in this paper propose a novel FSL framework for HSIC with a class-covariance metric (CMFSL). Overall, the CMFSL learns global class representations for each training episode by interactively using training samples from the base and novel classes, and a synthesis strategy is employed on the novel classes to avoid overfitting. During the meta-training and meta-testing, the class labels are determined directly using the Mahalanobis distance measurement rather than an extra classifier. Benefiting from the task-adapted class-covariance estimations, the CMFSL can construct more flexible decision boundaries than the commonly used Euclidean metric. Additionally, a lightweight cross-scale convolutional network (LXConvNet) consisting of 3-D and 2-D convolutions is designed to thoroughly exploit the spectral-spatial information in the high-frequency and low-frequency scales with low computational complexity. Furthermore, we devise a spectral-prior-based refinement module (SPRM) in the initial stage of feature extraction, which can not only force the network to emphasize the most informative bands while suppressing the useless ones, but also alleviate the effects of the domain shift between the base and novel categories to learn a collaborative embedding mapping. Extensive experiment results on four benchmark data sets demonstrate that the proposed CMFSL can outperform the state-of-the-art methods with few-shot annotated samples.**

***Index Terms*—Few-shot learning, class-covariance estimations, cross-scale, spectral prior, HSI classification.**

I. Introduction

Hyperspectral image (HSI) is captured through dozens to hundreds of narrow and contiguous bands. Benefited from the refined spectral information, it has been investigated in a variety of tasks, such as target/abnormal detection, fusion, etc [1]. For instance, Shi et al. [2] proposed an unconstrained linear mixture model to extract the objects of interest in the HSIs accurately. Sun et al. [3] presented a band divide-and-conquer multispectral and hyperspectral image fusion method, which enhances the spatial details of the HSI and preserves the spectral fidelity with high computational efficiency.

In HSI, each pixel is represented by a spectral curve containing wealthy discriminative information, which can be utilized to identify its owned category precisely [4]. Thus, HSI classification (HSIC) has been applied to various applications, such as medical care, environmental protection, and urban planning, to name a few [5], [6]. However, due to the high-dimensional spectral signal versus limited annotated training samples, HSIC with high accuracy is still a challenging problem and deserves further study [7], [8].

To settle this issue, researchers have devoted their efforts to various learning strategies and feature engineering. The former explores different learning schemes to maximize the prior information embodied in the limited labeled samples, and even involve the unlabeled data (e.g., active learning [9], self-learning [10]) and the data from other fields (e.g., transfer learning [11]). The latter seeks to reduce the feature dimension of the hyperspectral data while preserving and enhancing the contained discriminative information. Particularly, few-shot learning (FSL) is one of the most effective learning paradigms, which can rapidly generalize to new tasks with prior knowledge and limited supervisory experience [12].

Leveraging meta-learning (learning to learn) [13], many FSL approaches based on embedding and metric learning methods are developed and introduced to HSIC tasks [14]–[17]. Generally, these methods aim to learn a transformation function by reorganized meta-tasks (i.e., episodes including support and query samples), such that when projected into the embedding space, novel-class samples are easy to distinguish through using a linear classifier based on the distance measure. For instance, Snell *et al.* proposed a prototypical network (PNet) [18], which learns prototypes of each class by averaging the underlying features and using the Euclidean distance. Then, Liu *et al.* applied the PNet into HSIC for deep few-shot learning (DFSL) [14], which promotes the classification results with a handful of training samples. However, the labeled examples of the target data are not harnessed during training, which hinders the limited but significant utility of prior labeling information. To overcome this drawback, Zhang *et al.* presented a global prototypical network (GPN) to learn global prototypical representations of the base and novel classes, thus acquiring better performance [15].

This work was supported in part by the National Nature Science Foundation of China (no.61901343), the State Key Laboratory of Geo-Information Engineering (no.SKLGIE2020-M-3-1), Science and Technology on Space Intelligent Control Laboratory (no.ZDSYS-2019-03), and the China Postdoctoral Science Special Foundation (no.2018T111019). The project was also partially supported by the National Nature Science Foundation of China (no.61671383, 91538101), the 111 project (no.B08038), and Wuhu and Xidian University special fund for industry- university- research cooperation (no.XWYCXY-012021002). (Corresponding authors: Jiaojiao Li, Yunsong Li.)

B. Xi, J. Li, Y. Li, and R. Song are with the State Key Laboratory of Integrated Service Networks, School of Telecommunications Engineering, Xidian University, Xi'an 710071, China. (e-mail: xibobo1301@foxmail.com; jjli@xidian.edu.cn; ysli@mail.xidian.edu.cn; ruiScientific@gmail.com).

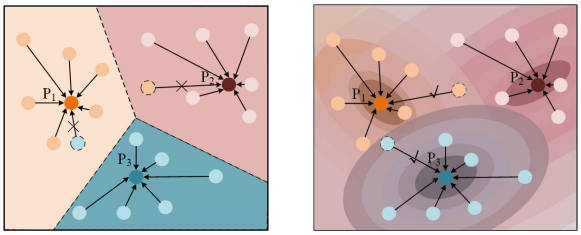
D. Hong is with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: hongdf@aircas.ac.cn).

J. Chanussot is with the Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, 38000 Grenoble, France, also with the Aerospace Information Research Institute, Chinese Academy of Sciences, 100094 Beijing, China. (e-mail: jocelyn@hi.is).

According to the theory of Bregman divergence, all the abovementioned FSL frameworks employ Euclidean metric to measure the distance in the latent space, which generally performs better than the Cosine metric [18], [19]. Nevertheless, the choice of Euclidean measure involves a flawed assumption, namely that the feature dimensions are uncorrelated with uniform variances. Furthermore, the Euclidean metric is insensitive to the distribution of within-class samples associated with their prototype, which is a problem suggested by the study in Ref. [20]. To solve these issues, we propose a novel FSL framework for HSIC with a class-covariance metric (CMFSL) through using the Mahalanobis distance with the task-adapted class-covariance estimations. For conceptual clarity, Fig. 1 illustrates the advantages of exploiting the estimated class covariance during the classification. In Fig. 1, $\{\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3\}$ refers to the prototypes (i.e., centers) representing each class. It is observed that the class-covariance-based metric can contribute to improving the decision boundaries of the non-linear classifier compared to the Euclidean metric by considering the distribution of each class in the latent space. Specifically, the new query samples, i.e., the dotted circles, that are misclassified in the Euclidean metric can be correctly recognized profited by the class-covariance estimations.

Additionally, the feature engineering can be divided into two categories: band selection (BS) [21] and feature extraction (FE) [22]. The BS directly chooses a most informative subset of all spectral bands, which remains the physical meaning of the reflectance/radiance records. In contrast, the FE aims to create a series of new representative features through combinations of the existing features. With convolutional neural networks (CNNs) flourishing in the computer vision field, 1-D-CNN [23], 2-D-CNN [24], and 3-D-CNN [25] are successively employed to explore the spectral, spatial, and spectral-spatial information contained in the HSI, which are verified to be more powerful than the traditional methods.

Specifically, since the 3-D convolution (3DCONV) incorporates the 3-D patches as input, 3-D-CNN can simultaneously investigate the spectral and spatial features and obtain satisfactory performance. However, the 3DCONV consumes more computing resources than the 2-D convolution (2DCONV) due to the involved extra spectral kernel dimension, and the overfitting occurs as more sophisticated networks go deeper, especially under limited training samples [26], [27]. To tackle this issue, Roy *et al.* proposed a hybrid spectral convolutional neural network (HybridSN) [28] comprising spectral-spatial 3DCONVs and spatial 2DCONVs, which reduces the computational complexity than the model with 3DCONV alone. Additionally, Tang *et al.* extended the Octave convolution [29] into a 3-D Octave convolution model (3-D-OCM) [30] by using the 3DCONVs, which separates the feature cubes into the high-frequency and low-frequency portions along the channel dimension and then realize information exchange between the two spatial scales. Since the spatial dimension of the feature maps is shrunk for the low-frequency component, the calculation amount is decreased in some degree, but the purely used 3DCONVs still cause a heavy computational burden. To conquer the deficiency, we present a lightweight cross-scale convolutional network (LXConvNet) composed of both 3DCONV and 2DCONVs, simultaneously exploiting the distinguishing features involved in the high-frequency and low-frequency scales with less computational complexity.



(a) Euclidean metric (b) Class-covariance metric

Fig. 1: Schematic diagram of different measurement metrics. The circles with different colors represent different classes.

From another perspective, to enhance the nontrivial spectral-spatial features and constrain the unimportant ones, attention mechanism [31]–[34] has been extensively studied in the FE process. In this study, we present a novel spectral-prior-based refinement module (SPRM) that adopts the spectral-prior of the patches to gather the global spatial correlations. Here, the spectral-prior includes the central pixel, mean and standard deviation vectors along the band dimension. Because the central pixel is the most fundamental information of the neighborhood patches, while the mean and standard deviation vectors are low-cost and effective band-wise statistics. By applying the SPRM to patch-wise samples, the crucial bands can be emphasized whereas the useless ones can be suppressed. Moreover, since the SPRM is interactively trained by the base and novel classes, it is expected to narrow the domain gaps between the source and target samples in the initial FE stage.

In summary, the main contributions of the proposed CMFSL are as follows:

1) A novel FSL framework for HSIC is proposed, which can achieve meta-task adapted classification and obtain state-of-the-art performance with few-shot labeled samples by learning a class-covariance-based metric space.

2) Inspired by the efficient Octave convolutions, we propose an LXConvNet, which is able to exploit the spectral-spatial features across high-frequency and low-frequency scales by joint 3DCONV and 2DCONVs, and transform the 3-D patch samples into the discriminative embedding space with comparatively low computational complexity.

3) We propose a new spectral-prior-based refinement module (SPRM). Apart from adaptively refining the band-wise information of the 3-D patch samples, SPRM is supposed to reduce the domain shift between the source and target data sets in the initial stage of the feature extraction.

The remains of this paper are organized as follows. Section II briefly introduces the related works. Section III describes our proposed method in detail. Section IV carries out abundant experiments and analyzes the classification results. Section V draws the conclusion.

II. Related works

As described above, existing methods of HSIC to cope with the high dimensional data against limited labeled samples

can be categorized into various learning strategies and feature engineering (mainly feature extraction). In this section, some typical HSIC methods are introduced sequentially.

A. Different Learning Strategies for HSIC

Generally, different learning strategies include supervised learning, semi-supervised learning, and unsupervised learning [35]–[37]. Remarkably, semi-supervised learning can fully exploit the limited supervision information and take a large amount of unlabeled data or the data from other sources as an auxiliary, which commonly achieves more excellent classification performance than the other two learning paradigms. Thus, it attracts more researches. For instance, Haut *et al.* presented an active learning approach associated with a Bayesian CNN [38], which was reinforced by the acquisition of new hard unlabeled samples to obtain robust identification results. Wu and Prasad proposed a semi-supervised learning framework with pseudo labels [39]. The deep networks are pre-trained by utilizing all data and their pseudo labels obtained by non-parametric Bayesian clustering, and then fine-tuned by the limited labeled samples. Additionally, to transfer the knowledge from the natural image with plenty of labeled samples, e.g., ImageNet, Chen *et al.* introduced a heterogeneous transfer learning framework to the HSIC tasks, achieving notable performance [40].

Unlike the typical transfer learning paradigm that initializes the network parameters using the pre-trained model, FSL encourages the model to learn fast-learning abilities from previous experience and rapidly generalize to the new concepts, which has demonstrated impressive performance on HSIC with a handful of labeled samples. Except for the early mentioned representative DFSL [14] and GPN [15], Chen *et al.* recently proposed a deep cross-domain few-shot learning (DCFSL) [41] framework, taking the domain shift between the base and novel classes into account, which obtains state-of-the-art classification results. Furthermore, considering the defects of the commonly used Euclidean distance in the above networks, we propose a novel FSL network for HSIC with a class-covariance metric in the embedding space, which can improve the non-linear classifier decision boundaries and acquire more excellent performance.

B. Various Feature Extraction for HSIC

Traditional FE approaches include the canonical machine learning methods, such as principle component analysis (PCA) [42], local linear embedding (LLE) [43], and linear discriminant analysis (LDA) [44], etc. Additionally, the morphological profiles (MPs) [45], extinction profiles (EPs) [46], attribute profiles (APs) [47] are also demonstrated to be effective FE strategies for HSI. However, the above methods are criticized due to that the extracted shallow features lack representativeness, and the design of the hyperparameters is complicated and time-consuming [22]. Besides, the FE process is independent of the subsequent classifier, thus the satisfied performance cannot be achieved. During the last decade, the deep learning (DL) based methods have flourished in HSIC, which can achieve deep FE and classification in an end-to-end manner [48]–[50]. Notably, since the 3DCONV can simultaneously investigate spectral and spatial correlations, more and more sophisticated DL frameworks based on 3DCONV are developed. For instance, Zhong *et al.* presented a spectral-spatial residual network (SSRN) [51], which investigates the 3-D features by skip-connections to ensure the generalization ability of the model. Furthermore, Wang *et al.* proposed a fast dense spectral-spatial convolution network (FDSSC) [52] by using dense connections between the low- and high-level features.

More significantly, inspired by the human perception mechanism, the attention scheme has been verified as a practical paradigm in HSIC. For instance, Mou *et al.* proposed a spectral attention module (SAM) [31] to attach different importance to the spectral bands of the 3-D patch samples, which achieves improved classification performance. However, the attention factors are produced through conducting global convolution (GCONV) on the patch, which may be cursory to obtain the desired weights, and restrict the capacity of the attention mechanism. To model the dependence of the spectral bands, Zheng et al. [32] employed the squeeze-and-excitation block (SE-Block) [53] in which the comprehensive spectral descriptors are acquired by the global average pooling (GAP), and the attention weights are generated by two successive fully connected (FC) layers and a sigmoid function.

Nonetheless, it is believed that the central target pixel contains the essential information in the pixel-centered neighborhoods, thus it should be treated with difference [54]. That indicates, the GAP misses this prior information and tends to obtain sub-optimal results. Besides, the FC layers in the SE-Block may bring redundant trainable parameters. To solve the above issues, this paper proposes an SPRM. The spectral-prior comprises the central pixel, mean vector, and standard deviation vector along the band dimension, which are utilized to describe the local and global information comprehensively. Furthermore, the attention weights are calculated using the 2DCONV with shareable weights rather than the FC layers. It is worth noting that the SPRM is supposed to not only highlight the informative bands but also narrow the domain gap between the base and novel classes in the initial stage of the FE process.

III. Proposed CMFSL for HSIC

The architecture of the proposed CMFSL is shown in Fig. 2. It can be observed that the proposed framework mainly consists of three parts: source and target data sets reorganization in the episode format, embedding feature extractor, and the class-covariance metric-based meta-training and meta-test. In the following, we will elaborate on the three portions in detail.

A. Source and Target Data Sets Reorganization

Suppose that we have a set of data sets in the source domain with C_b categories in total, which are called "base classes" having sufficient labeled samples. Meanwhile, in the target domain data set, we only have K annotated samples for each category of the N "novel classes". Due to the number of the spectral bands being various in different data sets, we conduct band selection on them to reduce the spectral bands to the

This article has been accepted for publication in IEEE Transactions on Image Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIP.2022.3192712

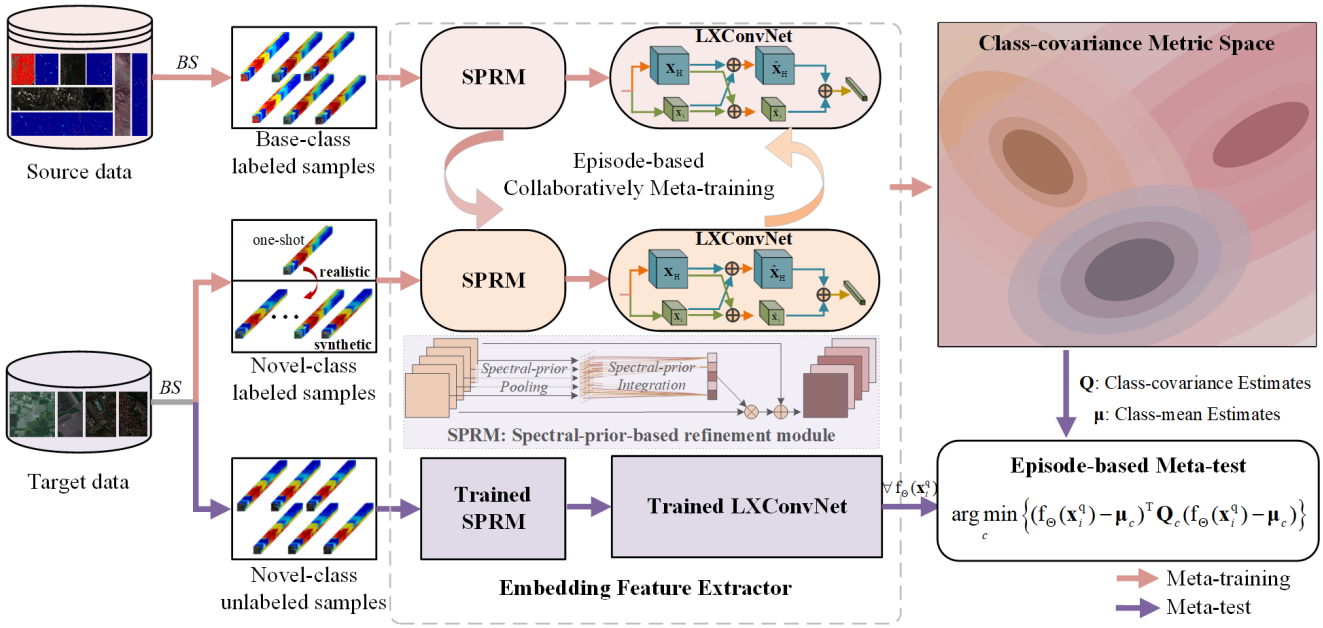


Fig. 2: The architecture of the proposed CMFSL for HSIC. Based on the class-covariance metric, the classification process is completed by the episode-based collaboratively meta-training of the source and target data sets, and the episode-based meta-test of the target data set. Notably, the embedding feature extractor comprises a new SPRM and a novel LXConvNet.

same number, which facilitates training the networks. Here, the unsupervised optimal clustering band selection (OCBS) method [55] is employed, which has been demonstrated effective in Ref. [56]. Then, based on the assumption that adjacent pixels tend to belong to the same class, we adopt the 3-D neighborhood patches as the input of the networks. Note that both base and novel classes are employed for training the networks. Thus, we reorganize the source and target data from the sample level into the meta-tasks (i.e., episodes) using a sampling strategy.

In detail, for the base classes, we first randomly sample N classes from the C_b classes, and then randomly sample N_S instances from each class as the support set. After that, N_Q instances are sampled from the rest of each class to form the query set. Finally, each episode is comprised of $N \times (N_S + N_Q)$ samples. For the novel classes, the sampling scheme is similar to that of the base classes. The difference is that we conduct the sample synthesis strategy to generate more annotated samples by adding random Gaussian noise to the few-shot realistic labeled samples, which is widely used in Refs. [57] and [58]. In this way, the scarcity of labeled samples in the novel classes is alleviated to construct more diverse episodes. After forming the meta-training tasks, we aim to learn a robust embedding feature extractor to map the source and target data into a shareable discriminative metric space, where the intra-class samples are compact and the inter-class ones are separated.

B. Embedding Feature Extractor

For the FSL based on embedding and metric learning, many studies have verified the significance of powerful feature representations of the embedding network [59]. Therefore, considering the characteristics of the HSI, i.e., image-spectrum merged structure, we design a novel embedding feature extractor, which mainly comprises an SPRM and an LXConvNet.

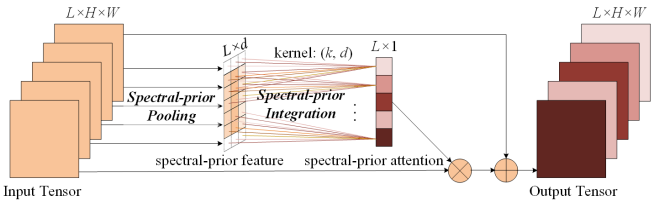


Fig. 3: The detailed framework of the SPRM.

1) Spectral-prior-based refinement module: The framework of the presented SPRM is illustrated in Fig. 3. It is mainly composed of two components: Spectral-prior Pooling (SPP) and Spectral-prior Integration (SPI). We will elaborate on the design of them in the following.

Spectral-prior Pooling: In the initial feature extraction stage, the input patch sample can be represented as $\mathbf{X} \in \mathbb{R}^{L\times H\times W}$, where $H\times W$ means the spatial dimension and L refers to the number of the selected bands. The SPP sub-module is designed to extract the spectral-prior representation $\mathbf{T} \in \mathbb{R}^{L\times d}$ from $\mathbf{X}$, where d is the number of spectral-prior features. Here, we take the central pixel of the patch, mean and standard deviation vectors along the band dimension into account, thus $d = 3$. In the formula, the feature vector $\mathbf{t}_l \in \mathbb{R}^3$ that summaries the spectral-prior information in the lth band of $\mathbf{X}$ is obtained by:

$$\mu_l = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} s_{lhw} \tag{1}$$

$$\sigma_l = \sqrt{\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}(s_{lhw}-\mu_l)^2} \tag{2}$$

$$\mathbf{t}_l = [\mu_l, \sigma_l, x_{lc}] \tag{3}$$

where $l \in [1, 2, ..., L]$. In Section IV-C, we demonstrate the practical benefits of the proposed SPP compared with other approaches for gathering the global information, i.e., using GCONV and GAP as in SAM [31] and SE-Block [32], respectively.

Spectral-prior Integration: Next, the spectral-prior feature is transformed to the spectral-prior attention weights for the bands by an SPI sub-module. Most importantly, inspired by the efficient channel attention network (ECANet) [60], which utilizes 1-D convolution instead of the FC layer to exploit the adjacent channel correlations with low computing cost, we adopt the 2DCONV operation to integrate the spectral-prior statistics. In specific, the calculation to obtain the spectral-prior attention can be expressed as:

$$\mathbf{A} = \sigma(\text{CONV2D}(\mathbf{T}, \mathbf{W})) \tag{4}$$

where σ refers to the sigmoid function to scale the attention weights to [0, 1]. $\mathbf{W} \in \mathbb{R}^{k\times d}$ is the trainable kernel of the 2DCONV. Note that the second dimension of $\mathbf{W}$ is assigned to d so that the second dimension of $\mathbf{T}$ can be squeezed to 1. Inspired by Ref. [60], the first dimension of $\mathbf{T}$ controlling the interaction range of the bands is determined by:

$$k = \left|\frac{\log_2(L)}{\gamma} + \frac{b}{\gamma}\right|_{odd} \tag{5}$$

in which $b = 1$ and $\gamma = 2$. Furthermore, to avoid the undesired information loss, we employ a skip connection to add the original $\mathbf{X}$ with the feature refined by the attention weight $\mathbf{A} \in \mathbb{R}^L$. Finally, the recalibrated output of the SPRM can be represented as:

$$\mathbf{F} = \mathbf{X} \odot (1+\mathbf{A}) \tag{6}$$

where $\odot$ represents the product with broadcasting. The attention weights can model the importance of the spectral-prior belonging to individual channels so as to emphasize or suppress them accordingly. Moreover, the adaptive weights are interactively optimized by the base and novel classes, thus they are supposed to narrow the domain gaps between the source and target domain at the beginning of the feature extraction process.

2) Lightweight cross-scale convolutional network: Inspired by the Octave convolution, we design an LXConvNet comprising the successive 3DCONV and 2DCONV. Note that each convolution is followed by the batch normalization and ReLU activation function to accelerate the training process. Specifically, the 3DCONVs are conducted in parallel and interact across the two feature scales, where the high spatial frequency component can capture the delicate details, and the low spatial frequency component can describe the smooth structure. For a brief, we first formulate the 2DCONV and 3DCONV as follows:

$$v_{ij}^{xy} = b_{ij} + \sum_{c}\sum_{\tau=0}^{h_i-1}\sum_{\sigma=0}^{w_i-1} k_{ijc}^{\tau\sigma} \times v_{(i-1)c}^{(x+\tau)(y+\sigma)} \tag{7}$$

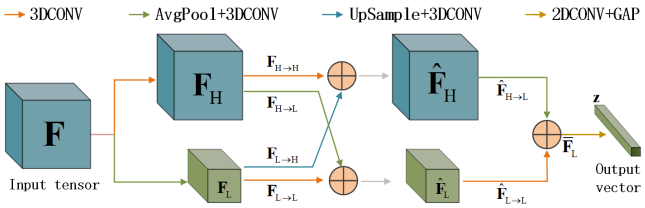


Fig. 4: The diagram of the proposed LXConvNet.

$$v_{ij}^{xyz} = b_{ij} + \sum_{c}\sum_{\tau=0}^{h_i-1}\sum_{\sigma=0}^{w_i-1}\sum_{\delta=0}^{d_i-1} k_{ijc}^{\tau\sigma\delta} \times v_{(i-1)c}^{(x+\tau)(y+\sigma)(z+\delta)} \tag{8}$$

in which v represents the value of the feature map, (x, y, z) are the position indexes of the *j*th feature map in the *i*th layer. (τ, σ, δ) are indexes of the kernels, c indexes over the series of feature maps in the $(l-1)$th layer. b means the bias. (h, w, c_{in}, c_{out}) and $(h, w, d, c_{in}, c_{out})$ describe the 2DCONV and 3DCONV kernels, where (h, w, d) are the kernel size along the height, width, and depth dimension, respectively. c_{in} and c_{out} represent the numbers of the input and output channels.

Particularly, the diagram of the proposed LXConvNet is delineated in Fig. 4. Given the input tensor $\mathbf{F} \in \mathbb{R}^{L\times H\times W}$, two 3DCONVs are firstly performed to generate different spatial frequency components, which can be formulated as:

$$\mathbf{F}_{\mathrm{H}} = \text{3DCONV}(\mathbf{F}, \mathbf{W}_{\mathrm{H}\to\mathrm{H}}) \tag{9}$$

$$\mathbf{F}_{\mathrm{L}} = \text{3DCONV}(\text{AvgPool}(\mathbf{F}), \mathbf{W}_{\mathrm{H}\to\mathrm{L}}) \tag{10}$$

where AvgPool means the average pooling operation with 2×2 size and stride 2. $\mathbf{W}$ represents the trainable convolutional kernel parameters. After that, we perform an extra average pooling on $\mathbf{F}_{\mathrm{H}}$ and $\mathbf{F}_{\mathrm{L}}$ along the depth dimension with 1×3 kernel and stride 3 to reduce the calculation burden. Next, for inter-frequency communication to explore the complementary information, the features across the two scales are added through using 3DCONV, AvgPool, and UnSample as bridges:

$$\begin{aligned}\hat{\mathbf{F}}_{\mathrm{H}} = \mathbf{F}_{\mathrm{H}\to\mathrm{H}} + \mathbf{F}_{\mathrm{L}\to\mathrm{H}} &= \text{3DCONV}(\mathbf{F}_{\mathrm{H}}, \mathbf{W}_{\mathrm{H}\to\mathrm{H}}) \\ &+\text{3DCONV}(\text{UpSample}(\mathbf{F}_{\mathrm{L}}), \mathbf{W}_{\mathrm{L}\to\mathrm{H}})\end{aligned} \tag{11}$$

$$\begin{aligned}\hat{\mathbf{F}}_{\mathrm{L}} = \mathbf{F}_{\mathrm{L}\to\mathrm{L}} + \mathbf{F}_{\mathrm{H}\to\mathrm{L}} &= \text{3DCONV}(\mathbf{F}_{\mathrm{L}}, \mathbf{W}_{\mathrm{L}\to\mathrm{L}}) \\ &+\text{3DCONV}(\text{AvgPool}(\mathbf{F}_{\mathrm{H}}), \mathbf{W}_{\mathrm{H}\to\mathrm{L}})\end{aligned} \tag{12}$$

where the UpSample is completed by the interpolate operation with factor 2 in the trilinear mode. Then, similar to the generation of $\hat{\mathbf{F}}_{\mathrm{L}}$, the feature maps in high and low frequencies are summarized into a comprehensive feature cube $\bar{\mathbf{F}}_{\mathrm{L}}$ in the low frequency:

$$\begin{aligned}\bar{\mathbf{F}}_{\mathrm{L}} = \hat{\mathbf{F}}_{\mathrm{L}\to\mathrm{L}} + \hat{\mathbf{F}}_{\mathrm{H}\to\mathrm{L}} &= \text{3DCONV}(\hat{\mathbf{F}}_{\mathrm{L}}, \mathbf{W}_{\mathrm{L}\to\mathrm{L}}) \\ &+\text{3DCONV}(\text{AvgPool}(\hat{\mathbf{X}}_{\mathrm{H}}), \mathbf{W}_{\mathrm{H}\to\mathrm{L}})\end{aligned} \tag{13}$$

Finally, in order to further exploit the deep spatial information, as well as to reduce the feature dimensionality, we conduct

2DCONV on $\bar{\mathbf{F}}_{\mathrm{L}}$ and perform the GAP to obtain the final embedding feature vector $\mathbf{z} \in \mathbb{R}^D$:

$$\mathbf{z} = \mathrm{GAP}(\mathrm{CONV2D}(\bar{\mathbf{F}}_{\mathrm{L}}), \mathbf{W}) \tag{14}$$

It is worth noting that the proposed LXConvNet is comparatively lightweight compared to the network with full 3DCONVs, because the used 2DCONV omits one spectral dimension and focuses on the spatial calculation. Besides, the cross-scale architecture also saves the computational burden compared to the network that performs on the high frequency all through.

C. Details of the Meta-training and Meta-test Processes

Taking the SPRM and the LXConvNet as a unified embedding feature extractor, it can be represented by a mapping function $\mathbf{z}_i = \mathrm{f}_\Theta(\mathbf{x}_i)$. Then the objective of the meta learning is to optimize the parameter Θ to maximize $\mathbb{E}_\tau[\prod_{Q^\tau} p(y_i^q | \mathbf{z}_i^q, S^\tau)]$, where Q^τ and S^τ represent the query and support set in each meta-task, respectively. y_i^q and $\mathbf{z}_i^q$ denote the true label and embedding feature of the query sample $\mathbf{x}_i$, respectively. $\mathbb{E}_\tau$ means the expectation over all meta-tasks. To allow the meta-training and meta-testing to be better adaptive to each specific task, we calculate the prediction probabilities as:

$$p(y_i^q = c | \mathbf{z}_i^q, S^\tau) = \mathrm{softmax}(-d_c(\mathbf{z}_i^q, \mu_c)) \tag{15}$$

where $d_c(x, y) = (x-y)^{\mathrm{T}} \mathbf{Q}_c (x-y)$ and $\mathbf{Q}_c$ is a covariance matrix corresponding to the cth class of current task. μ_c is the mean vector of the cth class of current task.

Due to that the value of $\mathbf{Q}_c$ cannot be known in advance, we estimate it through using the feature embeddings of the support set in each meta-task. To be more precise, we form $\mathbf{Q}_c$ as a convex combination of the intra-class covariance matrices Σ_c and inter-class covariance matrices Σ as:

$$\mathbf{Q}_c = \lambda_c \Sigma_c + (1-\lambda_c)\Sigma + \mathbf{I} \tag{16}$$

where the Σ_c is estimated by the embedding features of the S^τ of the c-class as:

$$\Sigma_c = \frac{1}{|S_c^\tau| - 1} \sum_{x_i \in S^\tau} (\mathbf{z}_i - \mu_c)(\mathbf{z}_i - \mu_c)^{\mathrm{T}} \tag{17}$$

in which $|S_c^\tau|$ refers to the number of the support samples of the cth class in current task. And if $|S_c^\tau| = 1$, the Σ_c is set to be the zero matrix with a suitable size. Similar to (17), the Σ is estimated through using all support samples in S_c^τ ignoring their particular classes. Besides, the weight of Σ_c is assigned in a deterministic mode as $\lambda_c = |S^\tau| / (|S^\tau| + 1)$, which intuitively increases as the shot number growing and the $\mathbf{Q}_c$ will pay more attention on the better estimated intra-class covariance Σ_c.

In the meta-testing phase, different from the common embedding and metric-based FSL frameworks that using an extra classifier, we directly determine the categories of the query samples by using the Mahalanobis distance measurement, which can comprehensively exploit the feature distribution of the entire support set. It can be formulated as:

$$y_i^{q*} = \arg\min_c \left\{ (\mathrm{f}_\Theta(\mathbf{x}_i^{\mathrm{q}}) - \mu_c)^{\mathrm{T}} \mathbf{Q}_c (\mathrm{f}_\Theta(\mathbf{x}_i^{\mathrm{q}}) - \mu_c) \right\} \tag{18}$$

where y_i^{q*} is the predicted label of the query sample $\mathbf{x}_i^{\mathrm{q}}$.

IV. Experimental results and analysis

A. Data sets and Evaluation Measures

The experimental data sets used in this work are divided into two groups: the source data sets of the base classes with sufficient labeled samples, and the target data sets of the novel classes with few-shot annotated instances. As the research in Refs. [14] and [15], we select the Houston University 2013 (UH2013)[1], Chikusei (CKS)[2], Kennedy Space Center (KSC)[3], and Botswana (BW)[3] scenes to serve as the source data. The brief descriptions of them are summarized in Table I, including the senor that collects the dataset, number of the bands, wavelength in the spectral dimension, spatial resolution, spatial size, number of labeled samples and classes. From Table I, we can observe that the four data sets are captured with different sensors and their characteristics from other aspects are also diverse, which offers rich knowledge to enhance the generalization capacity of the embedding mappings.

In terms of the target data sets, four benchmarks are chosen to evaluate the designed frameworks, i.e., Indian Pines (IP), Salinas (SA), Pavia University (UP), and Pavia Center (PC). The first IP data set is acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor with 224 bands from 0.4 to 2.5 μm, and 200 bands are preserved after removing the spectrum covering the scope of water absorption. The image comprises 145×145 pixels with 20 m/pixel spatial resolution. The second SA data set is also obtained by the AVIRIS in the wavelength of 0.4-2.5 μm. Similar to the IP scene, 204 bands are retained after discarding the water-absorbed bands. The spatial size is 512×217 with a higher geometric resolution of 3.7 m/pixel. The third UP data set is collected by the Reflective Optics Spectrographic Imaging System (ROSIS) airborne instrument. It is composed of 610×340 pixels with a 1.3 m/pixel spatial resolution and 103 spectral bands ranging from 0.43 to 0.86 μm after abandoning 12 noisy channels. The last PC data set is also captured by the ROSIS sensor with 102 spectral bands and a larger spatial size as 1096×715 with 1.3 m/pixel geometric resolution. The groundtruth, labeled land-cover types, and detailed sample numbers of the four data sets are depicted in Table II. For each data set, different land-cover classes are represented by different colors.

With respect to the performance evaluation, we adopt the class-specified classification accuracy (CA), average classification accuracy (AA), overall classification accuracy (OA), and Kappa coefficient (Kappa) as criteria, which are widely used in the HSIC tasks.

B. Experimental Settings

To maximize the effectiveness of the proposed networks, we conduct intensive experimental settings. In the reorganization of the source and target data sets, we firstly screen 55 categories with more than 200 labeled samples from the total 61 base classes to ensure enough annotated samples

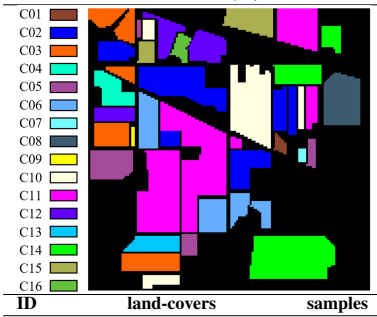
[1][Online]. Available: https://hyperspectral.ee.uh.edu/?page_id=459
[2][Online]. Available: https://naotoyokoya.com/Download.html
[3][Online]. Available: https://rslab.ut.ac.ir/data

TABLE I: DESCRIPTIONS OF THE SOURCE DATA SETS OF THE BASE CLASSES WITH SUFFICIENT LABELED SAMPLES

Data	Sensor	Band Num.	Wavelength (μm)	Spatial Resolution (m)	Spatial Size	Sample Num.	Class Num.
UH2013	ITRES-CASI 1500	144	0.364-1.046	2.5	349×1,905	5,211	13
CKS	Headwall Hyperspec-VNIR-C	128	0.363-1.018	2.5	2,517×2,335	77,592	19
KSC	AVIRIS	176	0.4-2.5	18	614×512	5,211	15
BW	EO-1	145	0.4-2.5	30	1,476×256	3,248	14

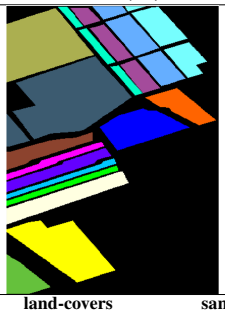
TABLE II: GROUNDTRUTH AND THE DETAILED NUMBERS OF THE SAMPLES FOR EACH CLASS IN THE TARGET DATA SETS

Indian Pines (IP)



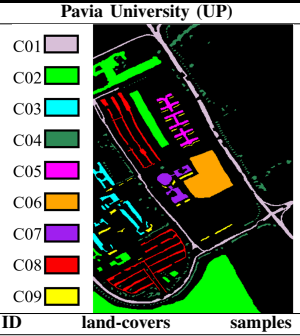
ID	land-covers	samples
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-mintill	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pasture-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-mintill	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings-Grass-Trees-Drives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

Salinas (SA)



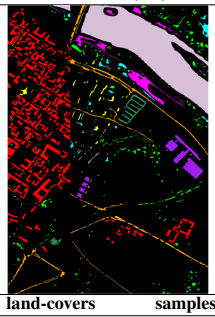
land-covers	samples
Brocoli_green_weeds_1	2009
Brocoli_green_weeds_2	3726
Fallow	1976
Fallow_rough_plow	1394
Fallow_smooth	2678
Stubble	3959
Celery	3579
Grapes_untrained	11271
Soil_vinyard_develop	6203
Corn_senesced_green_weeds	3278
Lettuce_romaine_4wk	1068
Lettuce_romaine_5wk	1927
Lettuce_romaine_6wk	916
Lettuce_romaine_7wk	1070
Vinyard_untrained	7268
Vinyard_vertical_trellis	1807
Total	54129

Pavia University (UP)



ID	land-covers	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

Pavia Center (PC)



land-covers	samples
Water	65971
Trees	7598
Asphalt	3090
Self-Blocking Bricks	2685
Bitumen	6584
Tiles	9248
Shadows	7287
Meadows	42826
Bare Soil	2863
Total	148152

for each base class. Then, for all data sets, the OCBS[4] [55] method is adopted to select 100 informative bands to ensure the consistent dimension of the samples fed into the feature extractor. In each episode, the number of the query samples N_Q is set to 19 and the patch size is set to 9×9 as the previous few-shot HSIC works [14], [15], [41]. In addition, the numbers of the sampled classes are set to 16 when classifying IP and SA and 9 for UP and PC data sets, respectively. This setting can facilitate collaboratively learning a shareable discriminant embedding space for both the source and target classes.

With regard to the embedding feature extractor, the detailed network configuration is shown in Table III. In particular, since the number of the 3-D convolutional kernels (M) highly influences the performance of the feature extractor, we investigate this parameter on the four test data sets, and the experiment results are depicted in Fig. 5. From Fig. 5, for the IP, SA, and UP data sets, it can be seen that the OA first rises and then falls as M increasing from 2 to 16, and it reaches the highest value when M equals to 8. On the other hand, the OA is promoted a little when M gains to 16 for the PC data set, but the computational cost increases a lot in this condition. Thus, we set it to 8 in our experiments.

TABLE III: CONFIGURATIONS OF THE PROPOSED EMBEDDING FEATURE EXTRACTOR

No.	Layer type	Kernel	Output shape	Connected to
1	Input	-	(100,9,9,1)	-
2	SPRM	(3,5)	(100,9,9,1)	1
3	3DCONV	(M,3,3,3)	(100,9,9,M)	2
4	AvgPool	(1,2,2)	(100,4,4,M)	2
5	AvgPool	(3,1,1)	(33,9,9,M)	3
6	AvgPool	(3,1,1)	(33,4,4,M)	4
7	3DCONV	(M,3,3,3)	(33,9,9,M)	5
8	AvgPool	(1,2,2)	(33,4,4,M)	5
9	3DCONV	(M,3,3,3)	(33,4,4,M)	8
10	UpSample	(1,2,2)	(33,9,9,M)	6
11	3DCONV	(M,3,3,3)	(33,9,9,M)	10
12	3DCONV	(M,3,3,3)	(33,4,4,M)	6
12	Add	-	(33,9,9,M)	7,11
13	Add	-	(33,4,4,M)	8,12
14	AvgPool	(1,2,2)	(33,4,4,M)	12
15	3DCONV	(2M,3,3,3)	(33,4,4,2M)	14
16	3DCONV	(2M,3,3,3)	(33,4,4,2M)	13
17	Add	-	(33,4,4,2M)	15,16
18	Reshape	-	(4,4,33*2M)	17
19	2DCONV	(128,3,3)	(4,4,128)	18
20	Flatten	-	(1,128)	19

[4][Online]. Available: https://github.com/tanmlh

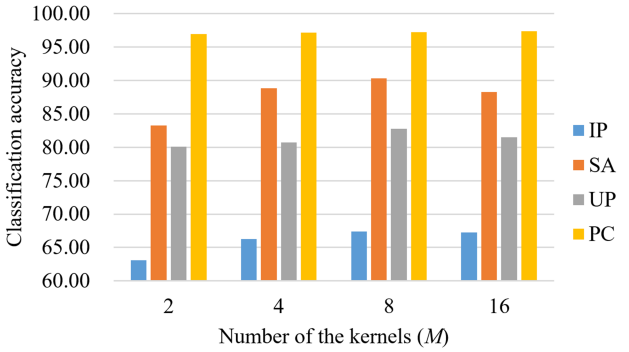


Fig. 5: Classification accuracy (%) with different numbers of the 3-D convolutional kernels (M).

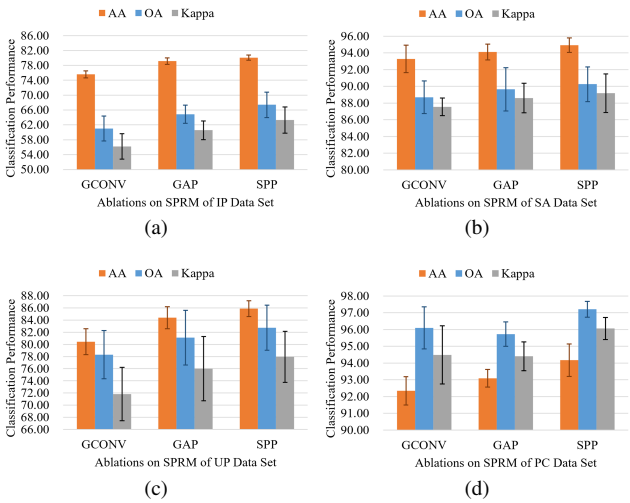


Fig. 6: Classification performance (%) of the ablations on the proposed SPRM. (a) IP, (b) SA, (c) UP, (d) PC.

For the meta-training process, the mini-batch i.e., episodes, constructed from the source and target data sets are interactively fed into the networks, and the total number of the training episodes is experimentally set to 10000. Additionally, we adopt the Adam algorithm to optimize the trainable parameters. Meanwhile, the learning rate is experimentally set to 0.001, which is appropriate for the networks to achieve a stable convergence.

C. Ablations on the SPRM

As described in Section III-B, the proposed SPRM is mainly comprised of the SPP to gather the global spatial information of each band, and the SPI to generate the attention weights. Specifically, the SPI is achieved by the convolutional operation, which has been demonstrated to be more efficient and effective in exploiting the correlations of the adjacent channels than the FC layers in the previous work [60]. Thus, we conduct the ablation studies primarily on the SPP component in this subsection.

In particular, keeping other settings the same, we replace the SPP in the SPRM with the GCONV [31] and the GAP [32], respectively. We randomly select five labeled samples in each novel class (i.e., five shots) and the experiments are repeated ten times independently. The average classification performance along with the standard deviation on the four experiment data sets is depicted in Fig. 6. From Fig. 6, it can be observed that the proposed SPP can achieve the best classification performance on all data sets. Additionally, the GAP obtains the sub-optimal results except for the OA and Kappa on PC data set. It indicates that learning the attentions weights directly by the GCONV under the few-shot conditions is hard for the networks. By contrast, the proposed SPP in the SPRM can better exploit the prior spectral feature to gather the global information contained in the 3-D patch samples, facilitating the learning process to pay more attention to the influential bands and suppress the useless ones.

D. Comparing with Other Methods

In order to validate the performance of the proposed CMFSL, we compare its classification results with other state-of-the-art methods. Specifically, 3-D-CNN [25], HybridSN [28], SSRN [51], and FDSSC [52] are four prevalent DL-based spectral-spatial classifiers. In addition, the DFSL [14] is the first work to introduce the FSL paradigm into the HSIC tasks. Besides, the GPN [15] is similar to the proposed CMFSL that utilizes the labeled samples of the novel classes to train the networks. Furthermore, the DCFSL [41] is recently proposed to consider the domain-shift problem in few-shot HSIC, which achieves advanced performance. Thus, DFSL, GPN, and DCFSL are also employed as comparisons. The configuration of the hyperparameters in the compared methods is assigned as the original papers. Five samples are randomly selected in the target classes to involve in the training process. To ensure the reliability of the experiment results, ten independent runs are executed for each method and data set. The mean accuracies and the standard deviations are shown from Table IV to VII. For comparison purposes, the highest and the suboptimum accuracies of each row are highlighted in bold and underline, respectively.

Table IV shows the classification performance of various methods on the IP data set. We can find that the proposed CMFSL outperforms other comparisons regarding AA, OA and Kappa coefficient. In concrete, the AA, OA, and Kappa exceed the second-best DCFSL algorithm as 0.87%, 2.07%, and 1.45%, respectively. Regarding the CA, the FDSSC obtains the optimal results on seven classes benefited to the dense connection structure, which is the best than other methods. Particularly, the presented CMFSL ranks first place up to five categories and second place in other five classes. It indicates that the feature extractor of the CMFSL can transform the original data into a more discriminative embedding space so that different classes are more clearly separated and more likely to be distinguished.

Table V provides the classification results obtained by the considered methods on the SA data set. It can be observed that the presented CMFSL achieves the most excellent classification performance compared to other methods. In terms of AA, OA, and Kappa, it surmounts the second-best values

TABLE IV: CLASSIFICATION PERFORMANCE (%) OF VARIOUS METHODS ON THE IP DATA SET

No.	3-D-CNN [25]	HybridSN [28]	SSRN [51]	FDSSC [52]	DFSL [14]	GPN [15]	DCFSL [41]	CMFSL
1	96.04±6.77	91.16±12.97	98.54±2.49	**99.27±1.56**	90.24±12.11	94.63±6.34	98.29±3.09	**99.27±1.12**
2	30.38±21.81	27.04±11.44	45.51±23.12	44.81±19.45	30.73±4.01	42.51±6.88	46.59±11.79	**48.76±8.13**
3	35.99±13.22	41.03±9.74	50.10±18.90	**71.27±13.04**	56.52±1.34	52.86±7.08	58.75±8.86	59.21±7.12
4	55.07±16.04	52.96±18.94	76.33±17.78	**92.37±13.39**	75.29±12.07	78.02±13.01	84.22±19.96	83.97±13.39
5	41.08±19.38	62.40±8.81	66.03±22.25	72.59±19.79	**78.45±1.79**	75.73±9.47	**78.45±8.76**	76.57±8.30
6	62.31±20.74	89.50±4.91	91.82±6.74	**93.05±4.58**	90.30±2.95	86.11±5.40	89.77±5.17	90.83±4.85
7	**100.00±0.00**	99.46±1.54	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	99.57±1.30	99.57±1.30	**100.00±0.00**
8	**90.86±7.48**	88.79±12.29	81.01±17.92	84.02±15.92	89.15±6.20	86.26±11.84	84.19±13.75	88.08±12.42
9	95.83±6.11	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	99.33±2.00	**100.00±0.00**
10	35.48±23.34	47.52±19.63	58.21±15.14	50.87±17.37	**60.29±6.12**	59.39±5.83	58.04±9.12	59.45±9.04
11	51.40±20.02	35.78±13.86	48.97±19.28	41.44±19.20	46.53±15.55	**61.98±12.18**	55.55±15.11	57.92±14.97
12	37.93±13.67	37.07±5.25	52.55±14.35	45.70±23.60	50.00±13.01	45.26±7.83	52.67±14.45	**53.10±10.46**
13	87.50±13.83	96.38±5.27	**99.60±0.70**	98.25±2.36	95.33±6.25	99.45±0.65	97.60±5.41	99.30±1.14
14	79.51±7.79	69.25±15.57	**87.94±7.36**	85.99±11.41	81.85±4.28	81.29±10.84	83.38±8.02	83.76±9.36
15	65.26±12.28	59.12±11.06	72.81±15.91	**86.93±11.41**	83.90±6.16	60.58±9.89	80.76±12.19	81.05±16.01
16	88.92±9.60	94.75±4.54	98.64±2.77	97.96±3.08	98.48±2.14	97.84±3.50	**99.66±0.52**	99.55±0.75
AA	65.85±2.28	68.26±1.73	76.75±2.12	79.03±2.17	76.69±0.68	76.34±2.27	79.18±1.46	**80.05±0.69**
OA	52.52±4.25	51.30±2.23	63.17±2.42	63.05±5.93	61.46±1.88	64.82±3.23	66.07±3.07	**67.40±3.43**
Kappa	46.68±4.40	46.52±2.28	58.70±2.28	59.02±6.09	56.95±1.72	60.39±3.37	61.84±3.13	**63.29±3.51**

TABLE V: CLASSIFICATION PERFORMANCE (%) OF VARIOUS METHODS ON THE SA DATA SET

No.	3-D-CNN [25]	HybridSN [28]	SSRN [51]	FDSSC [52]	DFSL [14]	GPN [15]	DCFSL [41]	CMFSL
1	63.74±9.47	**99.56±0.79**	99.50±0.60	86.14±28.22	96.75±1.21	99.33±0.71	97.83±1.24	98.54±1.80
2	75.07±12.24	99.27±1.23	**100.00±0.00**	99.89±0.31	96.01±1.75	99.63±0.41	97.64±0.39	99.75±0.33
3	67.83±15.60	82.21±29.18	95.25±7.54	87.64±12.31	97.17±0.76	89.49±10.59	96.44±1.44	**99.23±0.52**
4	88.59±6.12	89.93±11.53	**99.72±0.36**	99.23±1.72	97.59±0.22	99.21±1.18	95.74±2.12	99.16±0.90
5	75.97±11.91	83.03±17.75	**96.03±2.21**	93.72±11.46	91.95±0.91	92.45±2.20	92.79±3.90	94.20±3.46
6	93.44±8.46	99.27±1.22	**99.95±0.08**	99.77±0.32	97.86±0.20	99.35±0.77	97.33±1.95	98.56±0.96
7	83.37±5.93	81.53±20.49	**99.91±0.10**	99.81±0.26	97.83±0.02	99.72±0.33	97.24±0.52	99.47±0.49
8	58.40±10.39	55.84±15.64	64.01±17.77	61.05±32.23	75.52±2.38	71.85±11.04	75.60±9.42	**75.87±11.11**
9	88.66±3.92	98.51±2.97	**99.32±1.04**	98.73±2.32	96.53±1.81	99.11±0.98	97.39±1.26	98.38±2.15
10	61.24±17.96	82.01±15.76	88.93±6.06	89.14±4.77	86.83±5.14	84.35±7.89	89.25±4.37	**89.73±4.50**
11	89.78±13.05	95.76±4.04	97.59±2.67	94.66±7.22	96.96±0.67	**98.33±1.70**	97.19±1.16	97.99±2.37
12	86.91±12.10	83.45±14.93	99.61±0.89	**99.81±0.39**	98.12±1.80	99.53±0.39	97.74±1.33	99.26±1.93
13	79.56±12.67	98.48±1.90	**99.44±0.98**	99.20±2.05	96.87±1.23	99.18±1.03	97.52±1.36	99.26±0.77
14	83.63±19.34	93.81±7.57	97.57±2.79	97.92±2.86	96.00±0.34	98.66±1.35	96.46±0.78	**98.98±0.85**
15	40.56±14.54	**78.21±16.03**	72.80±22.40	61.94±32.57	63.46±7.49	77.01±7.15	72.35±5.58	77.54±8.75
16	65.28±16.00	**95.09±9.26**	91.73±7.60	94.99±5.50	92.62±4.43	90.64±5.46	91.99±3.66	92.90±6.30
AA	75.13±3.51	88.50±3.88	93.84±1.42	91.48±1.97	92.37±0.18	93.61±0.95	93.28±1.66	**94.93±0.87**
OA	69.91±3.57	82.53±4.15	87.30±2.45	84.30±3.17	87.03±1.25	88.70±1.91	89.57±1.77	**90.26±2.08**
Kappa	66.66±3.91	80.69±4.56	85.90±2.69	82.59±3.39	85.79±1.40	87.46±2.09	88.51±1.87	**89.18±2.30**

TABLE VI: CLASSIFICATION PERFORMANCE (%) OF VARIOUS METHODS ON THE UP DATA SET

No.	3-D-CNN [25]	HybridSN [28]	SSRN [51]	FDSSC [52]	DFSL [14]	GPN [15]	DCFSL [41]	CMFSL
1	56.31±16.91	45.06±14.71	**87.36±5.74**	77.20±9.27	73.11±8.96	74.11±8.96	80.37±10.02	76.35±8.44
2	80.18±12.07	76.23±9.30	65.48±16.77	68.92±14.75	75.35±11.56	76.35±11.56	**85.49±8.41**	82.28±9.82
3	38.91±12.21	**80.19±13.54**	79.59±16.88	70.06±19.68	64.43±10.23	65.43±10.23	63.74±11.59	71.63±11.28
4	84.17±12.28	71.20±15.07	**95.75±3.12**	91.75±8.03	88.71±3.30	89.71±3.30	94.27±2.44	93.54±3.32
5	82.04±12.24	96.99±5.03	99.66±0.58	90.73±23.41	95.72±0.60	96.72±0.60	99.51±0.44	**99.86±0.27**
6	47.27±15.36	78.10±18.98	81.16±11.52	85.41±10.74	84.71±8.71	**85.71±8.71**	76.30±8.46	85.00±9.11
7	48.61±14.16	**97.02±4.04**	87.11±14.35	96.29±4.12	80.24±7.20	81.24±7.20	79.42±7.09	85.59±7.20
8	43.01±17.75	43.92±10.01	70.39±26.34	**82.99±22.41**	72.56±8.67	73.56±8.67	59.94±12.94	79.38±9.73
9	88.25±9.83	54.30±13.66	**99.67±0.57**	98.58±2.00	95.17±0.84	96.17±0.84	98.67±2.32	99.46±0.80
AA	63.19±3.32	71.45±3.31	84.13±1.81	84.66±4.24	81.11±1.29	82.11±1.29	81.97±1.76	**85.90±1.30**
OA	66.93±4.50	69.48±4.43	76.50±5.67	77.23±5.45	77.51±4.86	80.51±4.86	81.52±3.50	**82.75±3.71**
Kappa	56.96±5.12	61.32±5.33	70.88±6.01	71.63±5.96	72.52±5.59	75.52±5.59	76.10±4.06	**77.96±4.21**

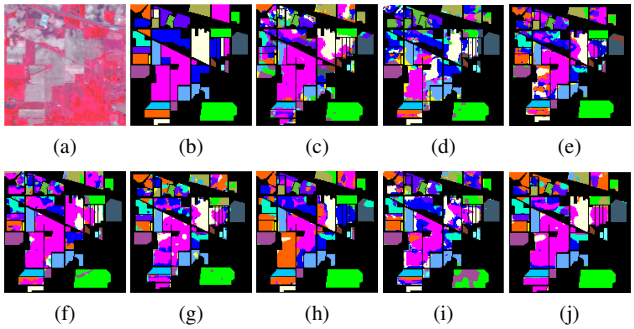


Fig. 7: Classification maps obtained via various algorithms for IP data set. (a) False-color image, (b) Groundtruth, (c) 3-D-CNN, (d) HybridSN, (e) SSRN, (f) FDSSC, (g) DFSL, (h) GPN, (i) DCFSL, (j) CMFSL. Best zoomed-in view.

by 1.09%, 0.69%, and 0.67%, respectively. Additionally, the GPN and DCFSL obtain more promising OA and Kappa than the methods without the FSL paradigm, which indicates the potential of the FSL in HSIC. Moreover, the DFSL method acquires unsatisfactory results mainly because of the lost utilization of the prior labeled information from the target classes. Remarkably, for the third class, "Fallow", the CMFSL outperforms other methods by a substantial margin, surpassing 2.06% over the second-best CA. It implies that the learned class-covariance metric can facilitate the networks to make more accurate decisions.

Table VI and Table VII summarize the classification accuracies on the UP and PC data sets, respectively. It can be observed that the proposed CMFSL achieves the top-level classification performance, which again demonstrates the capacity of the presented frameworks. Particularly, the CMFSL can reach the OA of 97.21% with only five labeled samples on the PC data set, which is inconceivable for the traditional CNN-based supervised HSIC methods. This is mainly due to the fact that, the limited prior knowledge embodied in the few-shot labeled samples and the spectral-prior of the 3-D patches can be fully exploited by the proposed CMFSL.

For visual comparison, we exhibit the best classification maps of various methods on the four data sets from Fig. 7 to Fig. 10, which are in consistent with Table IV to VII. The corresponding composite false-color image and groundtruth are also depicted together for convenience to compare. In Fig. 7, we can find that the classification maps obtained by different methods contain many undesired misclassifications. Because the IP data set has much noise and the spectral uncertainty problem is more intractable than other data sets. Thus, the limited few-shot labeled samples cannot meet the methods to produce a satisfactory classification map. By contrast, in Fig. 8 of SA data set, it can be observed that the classification map generated by the CMFSL is closest to the groundtruth and the edge structure is better preserved than the other methods. This phenomenon demonstrates that CMFSL can extract the distinguishing inter-class features, even for the border pixels with similar spectral-spatial information. Moreover, Fig. 9 and Fig. 10 can also verify this point. For instance, it can be seen that the "Bitumen" class of UP data set in red (Fig. 9 (j)) is

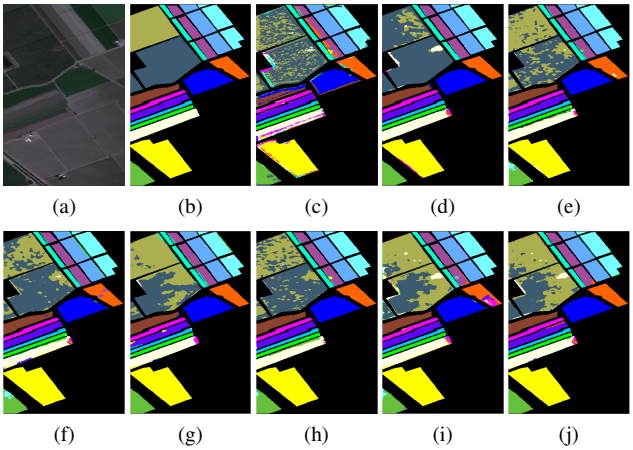


Fig. 8: Classification maps obtained via various algorithms for SA data set. (a) False-color image, (b) Groundtruth, (c) 3-D-CNN, (d) HybridSN, (e) SSRN, (f) FDSSC, (g) DFSL, (h) GPN, (i) DCFSL, (j) CMFSL. Best zoomed-in view.

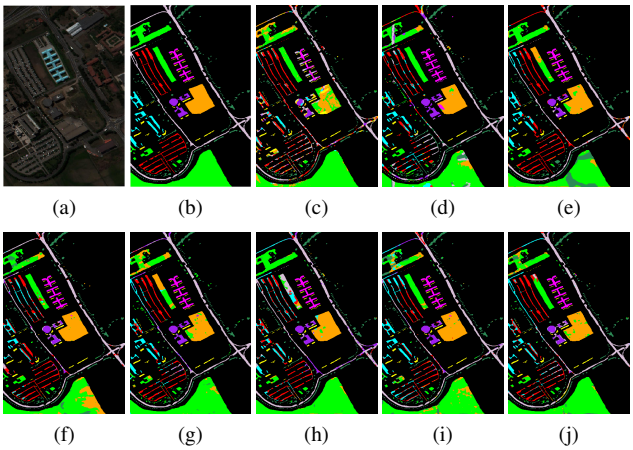


Fig. 9: Classification maps obtained via various algorithms for UP data set. (a) False-color image, (b) Groundtruth, (c) 3-D-CNN, (d) HybridSN, (e) SSRN, (f) FDSSC, (g) DFSL, (h) GPN, (i) DCFSL, (j) CMFSL. Best zoomed-in view.

clearer and more compact than other methods.

Next, we reduce the number of the labeled samples in the target classes to further validate the effectiveness of the proposed CMFSL. Specifically, from 1 to 5 annotated samples are randomly selected and the experiments are repeated ten times. The mean OAs of various methods on the four data sets are shown in Fig. 11. As expected, the classification performance is promoted with the increase of labeled information for all methods. Most significantly, our proposed CMFSL is always in the first place, verifying the robustness of the proposed networks.

E. Evaluation on the class-covariance estimations

In this work, we employ the Mahalanobis distance with task-adapted class-covariance estimations to collaboratively learn a discriminative metric space for both the base and novel classes. In this manner, the category of the query samples is determined by the distribution of the support data, but not

TABLE VII: CLASSIFICATION PERFORMANCE (%) OF VARIOUS METHODS ON THE PC DATA SET

No.	3-D-CNN [25]	HybridSN [28]	SSRN [51]	FDSSC [52]	DFSL [14]	GPN [15]	DCFSL [41]	CMFSL
1	99.73±0.35	99.79±0.38	98.50±0.27	98.72±0.20	98.51±0.20	98.71±0.31	**99.83±0.16**	99.78±0.18
2	81.34±13.08	75.46±20.52	87.50±7.29	87.58±3.45	87.11±3.88	91.16±3.15	91.06±3.73	**91.77±4.11**
3	82.35±7.75	84.32±11.43	89.72±10.20	91.90±9.47	**95.32±2.84**	93.03±4.59	93.89±2.60	93.75±3.41
4	77.66±12.35	67.94±18.55	88.91±12.47	93.93±8.24	**97.51±1.33**	93.24±6.31	93.49±7.93	95.80±5.52
5	71.38±15.33	65.91±21.74	88.16±10.22	86.18±5.65	85.70±3.47	89.92±5.64	**90.26±4.80**	88.69±6.64
6	56.20±15.28	48.73±19.81	**98.33±0.43**	98.04±0.84	92.79±5.07	94.49±3.38	94.95±3.70	94.45±4.13
7	78.60±6.34	**90.93±4.79**	83.71±5.29	85.53±5.85	89.68±2.59	84.07±4.11	85.31±2.43	85.59±3.95
8	70.69±17.47	94.80±3.25	97.46±1.26	96.99±1.65	96.35±0.41	97.36±0.79	**98.42±0.84**	98.28±0.84
9	83.68±8.48	66.17±27.73	98.95±0.13	98.97±0.06	98.79±0.27	98.43±0.71	98.73±0.84	**99.45±0.75**
AA	77.96±3.24	77.12±5.95	92.36±1.55	93.09±1.24	93.53±0.87	93.38±0.81	94.00±0.78	**94.17±0.97**
OA	84.31±5.21	90.42±1.77	96.09±0.63	95.18±0.46	95.86±0.36	96.33±0.51	97.07±0.41	**97.21±0.47**
Kappa	78.48±6.72	86.43±2.51	94.89±0.89	95.01±0.64	94.57±0.51	95.24±0.71	95.87±0.58	**96.06±0.66**

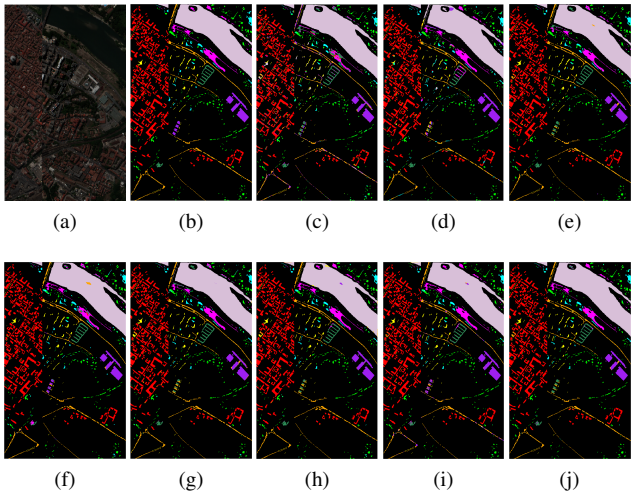


Fig. 10: Classification maps obtained via various algorithms for PC data set. (a) False-color image, (b) Groundtruth, (c) 3-D-CNN, (d) HybridSN, (e) SSRN, (f) FDSSC, (g) DFSL, (h) GPN, (i) DCFSL, (j) CMFSL. Best zoomed-in view.

TABLE VIII: CLASSIFICATION PERFORMANCE (%) WITH DIFFERENT DISTANCE METRICS ON IP AND UP DATA SETS

		AA	OA	Kappa
IP	CMFSL	**80.05±0.69**	**67.40±3.43**	**63.29±3.51**
	Var.1 (Euclidean)	77.82±1.38	65.04±2.60	60.73±2.62
	Var.2 (Cosine)	77.97±1.44	65.53±2.72	61.17±2.70
UP	CMFSL	**85.90±1.30**	**82.75±3.71**	**77.96±4.21**
	Var.1 (Euclidean)	84.23±1.44	81.61±3.87	76.50±4.48
	Var.2 (Cosine)	84.22±1.22	80.58±5.02	75.27±5.63

the prototypes (mean vectors of the embedding feature of the support set) as in the DFSL [14] and GPN [15]. To demonstrate the effectiveness of the estimated class covariance, we compare the Mahalanobis distance with the commonly used Euclidean distance and Cosine distance. In specific, keeping other experiment settings unchanged, the classification performance of IP and UP data sets with the three distance metrics are listed in Table VIII. From Table VIII, we can observe that the best performance is acquired by the Mahalanobis distance metric

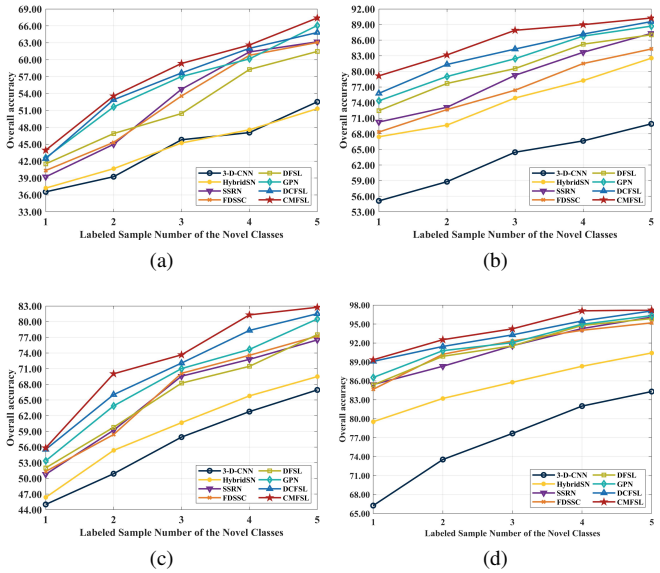


Fig. 11: Classification performance (%) of various methods on the four data sets with different numbers of labeled samples. (a) IP, (b) SA, (c) UP, (d) PC.

and the Euclidean and Cosine distance obtain the suboptimal results. In concrete, the Mahalanobis distance surpasses the second-best Cosine metric by 2.08%, 1.87%, 2.12% of AA, OA, and Kappa on IP data set, respectively; surmounts the second-best Euclidean metric by 1.67%, 1.14%, 1.46% of AA, OA, and Kappa on UP data set, respectively. Moreover, we visualize the embeddings in Fig. 12 by using t-SNE [61] to enhance the interpretability of the proposed model. From Fig. 12, we can observe that the inter-class features generated by the Mahalanobis distance are more separable than that of the other two distance metrics, especially for the category in yellow, leading to a more promising classification result.

F. Analysis of the Computational Complexity

In this section, we analyze the computational complexity of the proposed CMFSL. Specifically, we compare the number of parameters and inference time of the CMFSL with the FSL-related HSIC methods, i.e., DFSL, GPN, and DCFSL. All

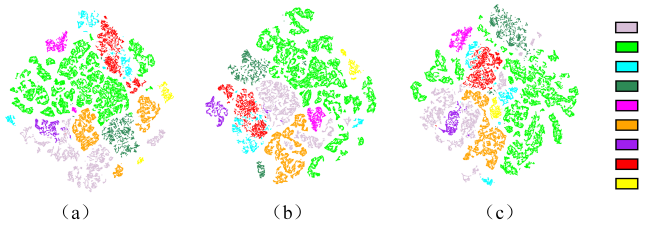


Fig. 12: Visualization of the embedding by the t-SNE method. The features are obtained by different distance metrics from UP data set. (a) Cosine distance, (b) Euclidean distance, (c) Mahalanobis distance with task-adapted class-covariance estimations. Here different colors represent different categories.

TABLE IX: PARAMETERS AND INFERENCE TIME OF THE PROPOSED CMFSL AND THE RELATED FEW-SHOT LEARNING-BASED METHODS.

		IP	SA	UP	PC
DFSL	*para.*		38,140		
	time	2.19	10.75	8.13	28.8
GPN	*para.*		137,033		
	time	2.37	12.57	10.01	34.90
DCFSL	para.	4,268,994	4,269,394	4,259,294	4,259,194
	time	2.28	12.46	8.48	29.51
CMFSL	*para.*		622,943		
	time	1.14	5.81	4.58	16.12

methods are performed on the NVIDIA GeForce GTX 1080Ti GPU, and the detailed statistics are summarized in Table IX.

From Table IX, it can be observed that the DFSL, GPN, and CMFSL have the same number of parameters for different data sets. It is because that the input data of the networks has the same number of bands after the band selection method. In contrast, DCFSL has different parameters due to the mapping layer, which transforms the source and various target data sets into the same band dimension. Particularly, CMFSL has a moderate number of parameters compared to other methods, which can facilitate the model to transfer sufficient knowledge from the source to the target classes and alleviate the overfitting problem. Notably, the CMFSL consumes the shortest period for predicting the classification results for the four data sets. This is mainly benefited from the proposed LXConvNet from two perspectives. Firstly, compared with the networks performed on the high-frequency features all through, LXConvNet decreases the computational complexity by exploiting the high-frequency and low-frequency crossed features. Secondly, the 2DCONV operation reduces the computational burden along the spectral dimension.

V. Conclusion

In this paper, we propose a novel FSL framework with a class-covariance metric (CMFSL) for accurate HSI classification. First of all, to exploit the information of the few-shot annotated labels, the network is interactively trained between the base and novel classes, aiming to learn a collaborative discriminative embedding space for them. Secondly, to emphasize the most informative bands and narrow the domain shift between the source and target classes, an SPRM is designed in the initial phase of the feature extraction, which reasonably utilizes the spectral prior information. Thirdly, we propose an LXConvNet consisting of 3-D and 2-D convolutions to investigate the spectral-spatial features across the high-frequency and low-frequency scales with comparatively low computational complexity. The experiments demonstrate that the CMFSL can achieve superior results than other methods with few-shot labeled samples. Moreover, the used Mahalanobis distance with task-adapted class-covariance estimations is more competitive in modeling the decision boundaries than the Euclidean and Cosine metrics in the proposed frameworks.

References

[1] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 37–78, Dec 2017.

[2] Y. Li, Y. Shi, K. Wang, B. Xi, J. Li, and P. Gamba, "Target detection with unconstrained linear mixture model and hierarchical denoising autoencoder in hyperspectral imagery," *IEEE Transactions on Image Processing*, vol. 31, pp. 1418–1432, 2022.

[3] W. Sun, K. Ren, X. Meng, C. Xiao, G. Yang, and J. Peng, "A band divide-and-conquer multispectral and hyperspectral image fusion method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[4] B. Xi, J. Li, Y. Li, R. Song, W. Sun, and Q. Du, "Multiscale context-aware ensemble deep kelm for efficient hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 6, pp. 5114–5130, 2021.

[5] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, pp. 279 – 317, 2019.

[6] Y. Liu, D. Sun, X. Hu, X. Ye, Y. Li, S. Liu, K. Cao, M. Chai, W. Zhou, J. Zhang, Y. Zhang, W. Sun, and L. Jiao, "The advanced hyperspectral imager: Aboard china's gaofen-5 satellite," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 4, pp. 23–32, 2019.

[7] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral–spatial hyperspectral image classification: An overview and new guidelines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 3, pp. 1579–1597, March 2018.

[8] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, "A survey: Deep learning for hyperspectral image classification with few labeled samples," *Neurocomputing*, vol. 448, pp. 179–204, 2021.

[9] Q. Shi, B. Du, and L. Zhang, "Spatial coherence-based batch-mode active learning for remote sensing image classification," *IEEE Transactions on Image Processing*, vol. 24, no. 7, pp. 2037–2050, 2015.

[10] J. Yue, L. Fang, H. Rahmani, and P. Ghamisi, "Self-supervised learning with adaptive distillation for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[11] X. He and Y. Chen, "Transferring cnn ensemble for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 5, pp. 876–880, 2021.

[12] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–34, 2020.

[13] M. Huisman, J. Rijn, and A. Plaat, "A survey of deep meta-learning," *Artificial Intelligence Review*, 2021.

[14] B. Liu, X. Yu, A. Yu, P. Zhang, G. Wan, and R. Wang, "Deep few-shot learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2290–2304, 2019.

[15] C. Zhang, J. Yue, and Q. Qin, "Global prototypical network for few-shot hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 4748–4759, 2020.

[16] Y. Wang, M. Liu, Y. Yang, Z. Li, Q. Du, Y. Chen, F. Li, and H. Yang, "Heterogeneous few-shot learning for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[17] X. Li, Z. Cao, L. Zhao, and J. Jiang, "ALPN: Active-learning-based prototypical network for few-shot hyperspectral imagery classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[18] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4077–4087.

[19] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra *et al.*, "Matching networks for one shot learning," *Advances in neural information processing systems*, vol. 29, pp. 3630–3638, 2016.

[20] P. Bateni, R. Goyal, V. Masrani, F. Wood, and L. Sigal, "Improved few-shot visual classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 493–14 502.

[21] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 118–139, 2019.

[22] B. Rasti, D. Hong, R. Hang, P. Ghamisi, X. Kang, J. Chanussot, and J. A. Benediktsson, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geoscience and Remote Sensing Magazine*, vol. 8, no. 4, pp. 60–88, 2020.

[23] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors*, vol. 2015, 2015.

[24] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2015, pp. 4959–4962.

[25] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sensing*, vol. 9, no. 1, 2017.

[26] B. Xi, J. Li, Y. Li, R. Song, Y. Shi, S. Liu, and Q. Du, "Deep prototypical networks with hybrid residual attention for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 3683–3700, 2020.

[27] S. K. Roy, P. Kar, D. Hong, X. Wu, A. Plaza, and J. Chanussot, "Revisiting deep hyperspectral feature extraction networks via gradient centralized convolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–19, 2022.

[28] 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 Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 277–281, 2020.

[29] Y. Chen, H. Fan, B. Xu, Z. Yan, Y. Kalantidis, M. Rohrbach, S. Yan, and J. Feng, "Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 3435–3444.

[30] X. Tang, F. Meng, X. Zhang, Y.-M. Cheung, J. Ma, F. Liu, and L. Jiao, "Hyperspectral image classification based on 3-D octave convolution with spatial–spectral attention network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2430–2447, 2021.

[31] L. Mou and X. X. Zhu, "Learning to pay attention on spectral domain: A spectral attention module-based convolutional network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 110–122, 2020.

[32] Z. Zheng, Y. Zhong, A. Ma, and L. Zhang, "FPGA: Fast patch-free global learning framework for fully end-to-end hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5612–5626, 2020.

[33] R. Hang, Z. Li, Q. Liu, P. Ghamisi, and S. S. Bhattacharyya, "Hyperspectral image classification with attention-aided CNNs," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2281–2293, 2021.

[34] M. Zhu, L. Jiao, F. Liu, S. Yang, and J. Wang, "Residual spectral–spatial attention network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 449–462, 2021.

[35] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2020.

[36] B. Xi, J. Li, Y. Li, R. Song, Y. Xiao, Q. Du, and J. Chanussot, "Semisupervised cross-scale graph prototypical network for hyperspectral image classification," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2022.

[37] Y. Li, B. Xi, J. Li, R. Song, Y. Xiao, and J. Chanussot, "SGML: A symmetric graph metric learning framework for efficient hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 609–622, 2022.

[38] J. M. Haut, M. E. Paoletti, J. Plaza, J. Li, and A. Plaza, "Active learning with convolutional neural networks for hyperspectral image classification using a new bayesian approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 11, pp. 6440–6461, 2018.

[39] H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1259–1270, 2018.

[40] X. He, Y. Chen, and P. Ghamisi, "Heterogeneous transfer learning for hyperspectral image classification based on convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3246–3263, 2020.

[41] Z. Li, M. Liu, Y. Chen, Y. Xu, W. Li, and Q. Du, "Deep cross-domain few-shot learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2022.

[42] X. Kang, X. Xiang, S. Li, and J. A. Benediktsson, "PCA-based edge-preserving features for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 12, pp. 7140–7151, Dec 2017.

[43] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[44] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 862–873, March 2009.

[45] J. Benediktsson, J. Palmason, and J. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480–491, 2005.

[46] P. Ghamisi, R. Souza, J. A. Benediktsson, L. Rittner, R. Lotufo, and X. X. Zhu, "Hyperspectral data classification using extended extinction profiles," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 11, pp. 1641–1645, 2016.

[47] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 3791–3808, 2020.

[48] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[49] N. Audebert, B. Le Saux, and S. Lefevre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 159–173, June 2019.

[50] H. Sun, X. Zheng, and X. Lu, "A supervised segmentation network for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 30, pp. 2810–2825, 2021.

[51] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 847–858, Feb 2018.

[52] W. Wang, S. Dou, Z. Jiang, and L. Sun, "A fast dense spectral–spatial convolution network framework for hyperspectral images classification," *Remote Sensing*, vol. 10, no. 7, 2018.

[53] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 7132–7141.

[54] B. Xi, J. Li, Y. Li, R. Song, Y. Xiao, Y. Shi, and Q. Du, "Multi-direction networks with attentional spectral prior for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[55] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 5910–5922, 2018.

[56] Y. Shi, J. Li, Y. Li, and Q. Du, "Sensor-independent hyperspectral target detection with semisupervised domain adaptive few-shot learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 8, pp. 6894–6906, 2021.

[57] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.

[58] H. Lee and H. Kwon, "Going deeper with contextual CNN for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4843–4855, 2017.
[59] M. N. Rizve, S. Khan, F. S. Khan, and M. Shah, "Exploring complementary strengths of invariant and equivariant representations for few-shot learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 836–10 846.
[60] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11 531–11 539.
[61] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

Bobo Xi (Member, IEEE) received the B.E. degree in information engineering and Ph.D. degree in information and communication engineering from Xidian University, Xi'an, China, in 2017 and 2022, respectively.

He is currently a Lecturer with the State Key Laboratory of Integrated Services Networks, School of Telecommunications, Xidian University. He has published over ten papers in refereed journals, including the IEEE Transactions on Image Processing, the IEEE Transactions on Neural Networks and Learning Systems, and the IEEE Transactions on Geoscience and Remote Sensing. His research interests include hyperspectral image processing, machine learning, and deep learning.

Jiaojiao Li (Member, IEEE) received the B.E. degree in computer science and technology, the M.S. degree in software engineering, and the Ph.D. degree in communication and information systems from Xidian University, Xi'an, China, in 2009, 2012, and 2016, respectively.

She was an exchange Ph.D. Student of Mississippi State University, supervised by Dr. Qian Du. She is currently an Associate Professor and doctoral supervisor with the State Key Laboratory of Integrated Service Networks, School of Telecommunications, Xidian University. Her research interests include hyperspectral remote sensing image analysis and processing, pattern recognition, and data compression.

Yunsong Li (Member, IEEE) received the M.S. degree in telecommunication and information systems and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 1999 and 2002, respectively.

In 1999, he joined the School of Telecommunications Engineering, Xidian University, where he is currently a Professor. He is also the Director of the State Key Laboratory of Integrated Service Networks, Image Coding and Processing Center. His research interests include image and video processing, hyperspectral image (HSI) processing, and high-performance computing.

Rui Song (Member, IEEE) received the M.S. and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 2006 and 2009, respectively.

He is currently a Professor with the State Key Laboratory of Integrated Service Networks, School of Telecommunications, Xidian University. His research interests include video coding algorithms, VLSI architecture design, and 3-D reconstruction.

Danfeng Hong (Senior Member, IEEE) received the M.Sc. degree (summa cum laude) in computer vision from the College of Information Engineering, Qingdao University, Qingdao, China, in 2015, the Dr. -Ing degree (summa cum laude) from the Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), Munich, Germany, in 2019.

He is currently a Professor with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences (CAS). Before joining CAS, he has been a Research Scientist and led a Spectral Vision Working Group at the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. He was also an Adjunct Scientist at GIPSA-lab, Grenoble INP, CNRS, Univ. Grenoble Alpes, Grenoble, France. His research interests include signal / image processing, hyperspectral remote sensing, machine / deep learning, artificial intelligence, and their applications in Earth Vision.

Dr. Hong is currently serving as the Associate Editor of the IEEE Transactions on Geoscience and Remote Sensing (TGRS), an Editorial Board Member of Remote Sensing, and an Editorial Advisory Board Member of ISPRS Journal of Photogrammetry and Remote Sensing. He has been serving as Technical Committee Member of IEEE Workshop on Hyperspectral Image and Signal Processing (WHISPERS) since 2022. He was a recipient of the Best Reviewer Award of the IEEE TGRS in 2021 and 2022, and the Best Reviewer Award of the IEEE JSTARS in 2022, the Jose Bioucas Dias Award for recognizing the outstanding paper at WHISPERS in 2021, the Remote Sensing Young Investigator Award in 2022, and the IEEE GRSS Early Career Award in 2022.

Jocelyn Chanussot (Fellow, IEEE) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree in electrical engineering from the Université de Savoie, Annecy, France, in 1998.

Since 1999, he has been with Grenoble INP, Grenoble, France, where he is currently a Professor of signal and image processing. He was a Visiting Scholar at Stanford University, Stanford, CA, USA, KTH Royal Institute of Technology, Stockholm, Sweden, and National University of Singapore, Singapore. Since 2013, he has been an Adjunct Professor of the University of Iceland, Reykjavk, Iceland, and the Chinese Academy of Sciences, Aerospace Information research Institute, Beijing, China. In 2015-2017, he was a Visiting Professor at the University of California, Los Angeles (UCLA), Los Angeles, CA, USA. His research interests include image analysis, hyperspectral remote sensing, data fusion, machine learning, and artificial intelligence.

Prof. Chanussot holds the AXA Chair in remote sensing with the Chinese Academy of Sciences, Aerospace Information research Institute. He is the founding President of IEEE Geoscience and Remote Sensing French chapter (2007-2010), which received the 2010 IEEE GRSS Chapter Excellence Award. He has received multiple outstanding paper awards. He was the Vice-President of the IEEE Geoscience and Remote Sensing Society, in charge of meetings and symposia (2017-2019). He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing (WHISPERS). He was the Chair (2009-2011) and Cochair of the GRS Data Fusion Technical Committee (2005-2008). He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006-2008) and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing (2009). He is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE TRANSACTIONS ON IMAGE PROCESSING, and the PROCEEDINGS OF THE IEEE. He was the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (2011-2015). In 2014 he served as a Guest Editor for the IEEE Signal Processing Magazine. He is a member of the Institut Universitaire de France (2012-2017) and a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters).