Unsupervised Meta Learning With Multiview Constraints for Hyperspectral Image Small Sample set Classification

Kuiliang Gao, Bing Liu, Xuchu Yu, and Anzhu Yu

***Abstract*— The difficulties of obtaining sufficient labeled samples have always been one of the factors hindering deep learning models from obtaining high accuracy in hyperspectral image (HSI) classification. To reduce the dependence of deep learning models on training samples, meta learning methods have been introduced, effectively improving the classification accuracy in small sample set scenarios. However, the existing methods based on meta learning still need to construct a labeled source data set with several pre-collected HSIs, and must utilize a large number of labeled samples for meta-training, which is actually time-consuming and labor-intensive. To solve this problem, this paper proposes a novel unsupervised meta learning method with multiview constraints for HSI small sample set classification. Specifically, the proposed method first builds an unlabeled source data set using unlabeled HSIs. Then, multiple spatial-spectral multiview features of each unlabeled sample are generated to construct tasks for unsupervised meta learning. Finally, the designed residual relation network is used for meta-training and small sample set classification based on the voting strategy. Compared with existing supervised meta learning methods for HSI classification, our method can only utilize HSIs without any label for unsupervised meta learning, which significantly reduces the number of requisite labeled samples in the whole classification process. To verify the effectiveness of the proposed method, extensive experiments are carried out on 8 public HSIs in the cross-domain and in-domain classification scenarios. The statistical results demonstrate that, compared with existing supervised meta learning methods and other advanced classification models, the proposed method can achieve competitive or better classification performance in small sample set scenarios.**

***Index Terms*— Hyperspectral image, small sample set classification, unsupervised meta learning, multiview learning, deep learning.**

I. Introduction

SINCE the 1970s, remote sensing technology has become one of the most important technical means to obtain earth observation information. Compared with panchromatic and multispectral images, HSIs contain both rich spectral details and spatial structure information, providing the possibility for accurate identification and classification [1].

Manuscript received June 5, 2021; revised February 19, 2022 and March 24, 2022; accepted April 10, 2022. Date of publication May 5, 2022; date of current version May 12, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 41801388, Grant 42101458, and Grant 42130112; and in part by the Natural Science Foundation of Henan Province under Grant 222102210035 and Grant 222300420387. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Chun-Shien Lu. *(Corresponding author: Bing Liu.)*

The authors are with PLA Strategic Support Force Information Engineering University, Zhengzhou 450001, China (e-mail: liubing220524@126.com).

Digital Object Identifier 10.1109/TIP.2022.3169689

The main task of HSI classsification is to assign a unique label to each pixel in the images, so as to produce classification maps reflecting the distribution information of ground objects. Classical machine learning classifiers such as support vector machine (SVM) [2] and random forest (RF) [3] can directly classify HSIs, but they cannot obtain satisfactory classification results due to the high-dimensional and non-linear characteristics of hyperspectral data. To make use of the potential features in HSIs, principal component analysis (PCA) [4], extended morphological profile (EMP) [5], local binary patterns (LBP) [6], Gabor [7] and other feature extraction methods are combined with the above classifiers, effectively improving classification performance. However, these traditional methods based on artificial features rely heavily on expert knowledge and lack universality, so fail to obtain stable results for multi-source and multi-resolution HSIs.

With the continuous development of remote sensing imaging technology, high-performance computing unit and computer vision theory, deep learning methods have been applied to HSI classification and constantly refreshing the classification accuracy in theoretical research [8]. Compared with traditional methods, deep learning methods can automatically mine the deep abstract features conducive to classification task from input data, so as to obtain higher classification and recognition accuracy. Stacked autoencoder (SAE) [9], recurrent neural network (RNN) [10], deep belief networks (DBN) [11] and convolutional neural networks (CNN) [12]–[16] are first introduced, achieving better classification performance than traditional methods with sufficient training samples. Compared with the three deep learning models, CNN can directly operate across high-dimensional data with grid structure using 2D or 3D convolution, thus effectively utilizing the spatial-spectral information in HSIs. For example, Chen et al. explored the feature extraction performance of 1D, 2D and 3D convolution in HSIs, and made a detailed analysis on the hyperparameters including the number of network layer, learning rate and so on [17]. Liu et al. built a deep feature extraction network using 3D convolution layers, further improving the accuracy of HSI classification [16]. In addition, residual structure [18], dense connection [19], capsule neuron [20], attention mechanism [21] and Network In Network (NIN) [22] structure are also combined with CNN model, to extract the more abundant features and further improve classification accuracy. To sum up, a large number of existing literatures show that CNN has become the mainstream model in HSI classification.

As we all know, collecting a large number of high-quality HSI labeled samples is very laborious and time-consuming in practice. However, most supervised deep learning models require sufficient labeled samples for parameter optimization. This contradiction shadows the classification performance of supervised deep learning models when the number of training samples are limited. To this end, advanced learning methods such as semi-supervised learning [23], unsupervised feature extraction and transfer learning have been introduced successively, to improve the classification accuracy with limited training samples. Two semi-supervised models, Generative adversarial networks (GANs) [24], [25] and graph convolutional network (GCNs) [26], [27] have been widely used in HSI classification. The former can effectively enhance training process by generating synthetic samples, while the latter can make full use of the potential features in unlabeled samples by building graph models. Both can significantly improve the classification performance with limited training samples. Unsupervised feature extraction can learn deep abstract representations by minimizing reconstruction errors, and is always combined with machine learning classifiers such as multilayer perceptron (MLP) and SVM for classification [28]. For example, Mei et al. designed an unsupervised spatial-spectral features learning method by combining 3D CNN and SAE, obtaining better results with insufficient training samples [29]. In transfer learning methods, the relevant data sets are used to pre-train the model, to improve the classification performance on target HSIs [30], [31]. In addition, active learning [32], contrastive learning [33], metric learning [34], [35], data augmentation [36] and deep forest [37] have also been introduced, improving classification performance to a certain extent.

Recently, with the rapid increase in the volume of hyperspectral data, how to use a few labeled samples to achieve rapid and accurate classification on target HSIs, that is, HSI small sample set classification, has attracted a lot of attention. In HSI small sample set classification, most deep learning models fail to obtain satisfactory results, because in the target HSIs, only 1-10 labeled samples per class are available, which makes the deep learning model suffer from overfitting in the training process. To improve the accuracy of HSI small sample set classification, meta learning methods are introduced and has achieved encouraging results. Different from conventional deep learning methods, meta learning methods take tasks containing support sets and query sets as the basic units for training, enabling the model to acquire the ability of learning how to learn [38]–[40]. Typically, meta learning methods first learn more general-purpose knowledge on the source data sets, and then fine-tune the model using a few labeled samples in the target HSIs, to quickly adapt to new classification tasks [41], [42]. Liu et al. take the lead in exploring the performance of meta learning methods in HSI classification [43]. Specifically, Liu et al. proposed a novel classification method based on prototype network, and constructed the source data sets using different HSIs, effectively improving the accuracy of HSI small sample set classification. Then, Gao et al. and Ma et al. introduced the relation network into HSI small sample set classification, further improving classification performance [44], [45]. Furthermore, Li et al. fully considered the domain shift between data sets, and designed a novel classification framework by combining a domain adaptation method and the relation network, obtaining more excellent results [46]. In addition, induction network [47], model-agnostic meta learning algorithm [48] and temporal convolution [49] are also applied in HSIs small sample set classification, obtaining better results than most existing deep learning models.

The above meta learning-based methods can effectively improve the accuracy of HSI small sample set classification, but they all carry out supervised learning on a labeled source data set, which means that a large number of labeled samples are still required in advance. For example, Liu et al. constructed a source data set containing 55 classes and 11000 labeled samples using four HSIs [43]. In fact, collecting a large number of high-quality labeled samples from different HSIs and constructing source data sets is still time-consuming and labor-intensive. In contrast, we can quickly collect a large number of unlabeled HSIs. Therefore, how to utilize unlabeled HSIs for unsupervised meta learning, to further reduce the number of requisite labeled samples and improve the performance of HSI small sample set classification, is more meaningful in practical application. Two typical unsupervised meta learning methods, CACTUs (Clustering to Automatically Construct Tasks for Unsupervised meta learning) [50] and UMTRA (Unsupervised Meta learning with Tasks constructed by Random sampling and Augmentation) [51], have achieved promising results on the Omniglot and Mini-Imagenet few-shot learning benchmarks. Briefly, the two methods allocate pseudo labels to unlabeled samples through clustering and data augmentation respectively, to construct the support sets and query sets for unsupervised meta learning. However, according to our experiments, simply applying CACTUs and UMTRA to HSIs cannot obtain satisfactory classification results, because the two methods designed for natural image cannot fully consider the characteristics of high-dimensional, nonlinearity, and spatial-spectral information fusion in hyperspectral data.

Multiview learning aims to model the features of the same object under different views and conduct joint optimization, so as to improve the classification performance and generalization ability [52]. Obviously, different spectral bands in HSIs can be regarded as different views. Our previous studies have shown that using the spatial-spectral features of the same sample in different bands for multiview learning can effectively improve the accuracy of HSI classification [33]. This also indirectly indicates that, constructing tasks with the multiview features of the same sample has great potential for unsupervised meta-learning on HSIs. Therefore, based on the idea of meta learning and multiview learning, this paper proposes a novel unsupervised meta learning method with multiview constraints, UM^{2}L, to further improve the performance of HSI small sample set classification and reduce the dependence on requisite labeled samples. Specifically, the proposed method first generates multiple spatial-spectral multiview features by combining different bands of the same sample and data augmentation, to constract the tasks for unsupervised meta learning. Then, the designed residual relation network is adopted for meta-training and small sample set classification based on the voting strategy. Extensive experiments are carried on four pre-collected HSIs and four target HSIs. The statistical results show that, compared with existing supervised meta

learning methods and other advanced classification models, the proposed method can achieve competitive or better results, and can significantly reduce the number of requisite labeled samples in training process.

The main contributions of this paper are as follows.

1) We propose a novel unsupervised meta learning method with multiview constraints for HSI small sample set classification. Experimental results show that the proposed method can significantly reduce the number of requisite labeled samples and effectively improve the accuracy of HSI small sample set classification. To the best of our knowledge, our method is the first one based on unsupervised meta learning in HSI classification.
2) Based on the idea of multiview learning, a task construction method for unsupervised meta learning is proposed. In this method, multiview features along the spectral dimension are obtained by combining different bands of the same sample, and multiview features along the spatial dimension are obtained by data augmentation such as rotation and flipping, to generate multiple spatial-spectral multiview features belonging to the same sample together.
3) To make full use of the spatial-spectral information in HSIs, we design an deep residual relation network for metric learning. The designed network can extract deep features with short inner-class distance and long inter-class distance, effectively enhancing separability.
4) In the experiments, two classification scenarios, cross-domain and in-domain HSI small sample set classification, are used to fully verify the effectiveness of the proposed method. The statistical results show that, compared with existing supervised meta learning methods and other advanced classification models, the proposed method can achieve competitive or better results.

The remainder of this paper is structured as follows. Section II introduces the general process of meta learning-based methods for HSI small sample set classification. Section III describes the proposed method in detail. IV presents the experimental results and detailed analysis. Finally, conclusions are provided in Section V.

II. Meta Learning-Based HSI Small Sample Set Classification

A. Supervised Meta Learning

Fig. 1 illustrates the general process of supervised meta learning methods classifying HSIs, including three phases: meta-training, fine-tuning and classification. To illustrate the whole process more clearly, let's first explain several concepts. The labeled source data set is constructed with pre-collected HSIs, including a large number of labeled samples belonging to different classes. The pre-collected HSIs possess different spatial resolutions and spectral ranges, which can effectively increase the diversity and richness of samples in the source data set, thus improving the effectivesess of meta-training. The target HSI are completely different from the pre-collected HSIs, and are used to evaluate classification performance. Specifically, only a few labeled samples are randomly selected

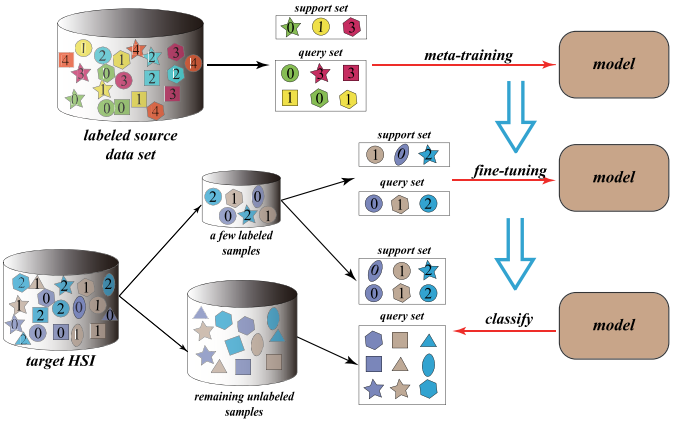


Fig. 1. HSI small sample set classification based on supervised meta learning. Different colors distinguish different classes.

from the target HSI to fine-tune the model, and then the remaining unlabeled samples are used to evaluate the classification results. In conclusion, in the supervised meta-learning methods for HSI small sample set classification, the model first utilizes the labeled source data set for meta-training, developing the ability of learning how to learn, and then utilizes only a few labeled samples from the target HSI for fine-tuning, to quickly adapt to the new and unseen classes, and finally classifies the target HSI.

B. Tasks-Based Learning Strategy

Indeed, meta learning transfers the knowledge learned from the source data set to the target data set. However, different from the general transfer learning, meta learning takes tasks as basic units for training. A task contains a support set and a query set. In a task, the support set and the query set contain the same classes, and the number of samples in the support set is usually smaller than that in the query set, to simulate the situation of small sample set classification. A task can be described using three keywords: way, $shot$, and $query$. The number of classes in a task is denoted by way, and the number of samples per class in the support set and query set are denoted by $shot$ and $query$, respectively. For example, in Fig. 1, the task in the meta-training phase can be denoted as 3-way 1-$shot$ 2-$query$. Each task is generated by a random combination of classes in the same data set, and the number of classes in a task is less than that in the corresponding data set, to increase the diversity of tasks. In the learning process, the samples in the support set are clearly labeled, while the samples in the query set are regarded as unknown. The model updates its parameters by predicting the labels of the query samples and calculating the loss.

As shown in Fig. 1, in the meta-training phase, tasks are constructed with the samples in the labeled source data set. In the fine-tuning phase, tasks are constructed with only a few labeled samples from the target HSI. In the classification phase, tasks are constructed with the selected labeled samples in the previous phase as support sets and the remaining labeled samples as query sets.

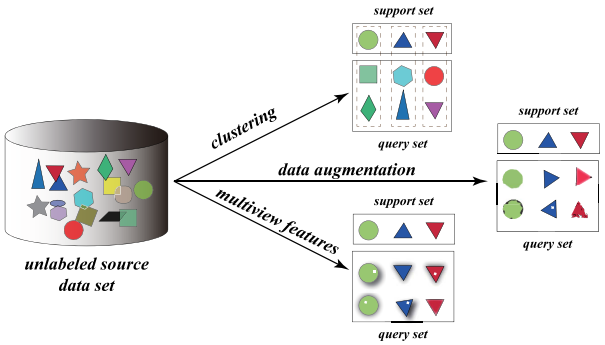


Fig. 2. Schematic of different task construction methods in unsupervised meta learning. Different colors distinguish different samples.

C. Unsupervised Meta Learning

Unsupervised meta learning attempts to utilize unlabeled HSI source data sets for meta-training, enabling the model to learn more general-purpose knowledge from tasks constructed with unlabeled samples. What distinguishes unsupervised meta learning from supervised meta learning is the unlabeled source data set. Therefore, how to utilize unlabeled source data set to construct tasks for meta-training is the key problem that must be solved in unsupervised meta learning. As far as we know, there are few studies related to unsupervised meta learning. CACTUs and UMTRA are two representative unsupervised meta learning methods designed for natural image. CACTUs first conducts clustering on the unlabeled samples and utilizes the obtained results with pseudo labels to construct tasks for meta-training. UMTRA performs data augmentation, such as random shift and flipping, on each unlabeled sample, and treats each sample and its augmentation samples as a class. The proposed method constructs tasks by generating spatial-spectral multiview features of the same sample. Fig. 2 briefly depicts the three task construction methods.

Supervised meta learning requires a large number of labeled samples for meta-training. Therefore, in the theoretical research on HSI small sample set classification, most of the existing methods utilize several labeled HSIs to build source data sets [43], [44], [46]. This means that the model performs meta-training on the pre-collected HSIs, and performs cross-domain classification on another completely different HSI. However, in the unsupervised meta learning methods, only one target HSI can be used to complete meta-training and evaluation. Similar to most supervised deep learning methods, the target HSI is first treated as unlabeled data set for meta-training, and then the original labeled samples in the target HSI are used for fine-tuning and evaluation. This means that the methods based on unsupervised meta learning can perform cross-domain and in-domain small sample set classification.

III. The Proposed Method

A. Workflow

To further improve the performance of HSI small sample set classification and alleviate the dependence of the model on requisite labeled samples, we propose a novel method, UM^{2}L, based on unsupervised meta learning and multiview learning. Similar to the existing meta learning methods for HSI classification, the proposed method includes three phases: unsupervised meta learning (meta-training), fine-tuning and classification. Fig. 3 shows the workflow of the proposed method performing unsupervised meta learning. First, spatial-spectral multiview features are generated for each sample in the unlabeled source data set. Then, each sample and its multiview features are treated as a class, and the tasks are constructed by random selection. Next, the designed residual relation network performs meta-training on the tasks constructed with multiview features, to learn more general-purpose feature knowledge. It should be noted that, HSIs without any label information are used to build the unlabeled source data set, and multiview features of unlabeled samples are used to construct the tasks for meta-training. Therefore, the proposed method is actually an unsupervised meta learning method with multiview constraints. After meta-training, the designed model is fine-tuned with only a few labeled samples in the target HSIs, to quickly adapt to the new unseen classes. Finally, based on the voting strategy in Fig. 6, the classification performance of the model is evaluated with the remaining labeled samples in the target HSIs.

B. Spatial-Spectral Multiview Features

In real world applications, multiview data are very common. Data describing the same object are often collected from different measuring methods as particular singleview data cannot comprehensively describe the information of all examples [52]. It is significant to make good use of the information from different views. Obviously, different bands in HSIs can be regarded as different views because they can reflect different properties. Therefore, based on the idea of multiview learning, we propose a novel task construction method, i.e., generating the spatial-spectral multiview features of the same sample.

Fig. 4 describes the process of generating the spatial-spectral multiview features. For a cube sample with $m \times m$ pixels and c bands, band random selection and data augmentation are successively adopted. HSIs possess dozens to hundreds of bands, and there is a strong correlation between adjacent bands. Taking each band as a view will undoubtedly increase the complexity of meta learning. Therefore, multiview features along the spectral dimension are obtained by random selection. Specifically, three bands are randomly selected from the original spectral bands and stacked, which is repeated several times to generate multiple different spectral multiview features. After that, the spatial information is transformed using several data augmentation methods, such as crop, flip, rotation and cutout, to generate the multiview features along the spatial dimensions.

In brief, multiview features along the spectral dimension are obtained by band random selection, and multiview features along the spatial dimension are obtained by data augmentation, to generate the spatial-spectral multiview features together. After generating multiple multiview features, each sample could be treated as a particular class. This effectively increases the size of the source data set and provides pseudo label information for task construction.

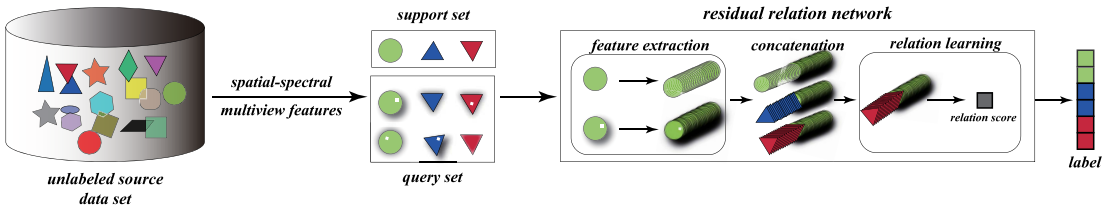


Fig. 3. Workflow of the proposed method performing unsupervised meta learning.

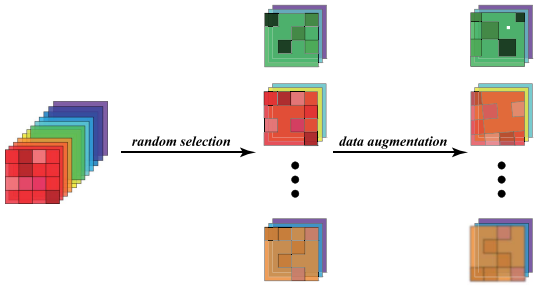


Fig. 4. Schematic of generating the spatial-spectral multiview features.

C. Residual Relation Network

To make full use of the spatial-spectral information in HSIs, a deep residual relation network is designed. As shown in Fig. 3, the designed model consists of three parts: feature extraction module, concatenation operation and relation learning module. The feature extraction module and relation learning module are explained more detailedly in Fig. 5. The designed network possess an end-to-end structure, which takes tasks as input and directly outputs the predictive class labels of the query sample.

Formally, a task containing C classes is denoted as $\mathcal{T} = \{\mathcal{S}, \mathcal{Q}\}$ where $\mathcal{S} = \{(x_i, y_i)\}_{i=1}^{C\times K}$ represents the support set, $\mathcal{Q} = \{(x_j, y_j)\}_{j=1}^{C\times N}$ represents the query set, K and N denotes the number of samples per class in support and query set respectively. Firstly, the feature extraction module is used to map each sample in the input task into a deep metric space. Specifically, the feature extraction module is mainly composed of convolution blocks, residual connections and max pooling layers. As shown in Fig. 5(a), a convolution block including a convolution layer, a batch normalization layer and a ReLU (Rectified Linear Unit) activation layer. Convolutional layer is the most important part in convolutional block, which is responsible for extracting abstract features from the input data. Batch normalization layer can effectively alleviate the problem of vanishing gradient and improve the training speed. ReLU activation function can increase the nonlinearity of the model and accelerate convergence. Max pooling layers are added between convolution blocks, to gradually reduce the spatial size of feature maps and ensure that the convolution kernels at deep level have a larger receptive field. Furthermore, residual connection is introduced to establish shortcuts, so as to improve the reuse rate of feature maps at different levels and the training effect of the designed deep network. Fig. 5(b) shows the overall structure of the designed feature extraction module. Actually, The feature extraction module is equivalent to a nonlinear embedding function f that maps input samples x_i and x_j to deep features $f(x_i)$ and $f(x_j)$.

The feature extraction module maps each sample in the input task to a feature vector. For subsequent relation learning and class prediction, the vectors obtained from support set are averaged according to class, to generate the class vectors. This means that, no matter how many support samples are included in a task, the support set will always produce C feature vectors. After that, each feature vector obtained from the query set is concatenated with the class vectors, to generate $C \times C \times K$ concatenations $\mathcal{C}(f(x_i), f(x_j))$.

The relation learning module, responsible for calculating the similarity between two feature vectors in the concatenations, could also be regard as a nonlinear transformation function g. Compared with feature extraction module, the relation learning module is simple and shallow. As shown in Fig. 5(c), a convolution block is first used to process the input concatenations, so as to reduce the dimensions of the vectors while learning the cross-channel features. Then, two fully connected layers are added to enhance the fitting ability of the module and output the relation score. The relation score denoted as $r_{i,j} = g[\mathcal{C}(f(x_i), f(x_j))]$ represents the similarity between the feature vectors $f(x_i)$ and $f(x_j)$. If the similarity between two feature vectors is high, the relation score should be high; otherwise, the relation score should be low. Finally, the class of the query sample will be determined by the support sample corresponding to the max relation score.

The designed residual relation network is trained with mean square error (MSE) as loss function (Equation 1). If the support sample and query sample in the concatenations belong to the same class, i.e. $y_i == y_j$, the relation score is optimized to 1, otherwise 0. Under the constraints of MSE loss function, the model can effectively learn the relationship between features and predict classes by comparison.

$$L_{MSE} = \sum_{i=1}^{C\times K}\sum_{j=1}^{C\times N}(r_{i,j} - 1\cdot(y_i == y_j))^2. \tag{1}$$

It should be further explained that there are two main considerations for using MSE as the loss function: firstly, most existing classification frameworks based on relation network adopt MSE for model optimization, to achieve the purpose of similarity comparison and relation learning [39], [44], [45], [47]; Secondly, by comparing the influence of cross entropy loss and MSE on classification results, it is found that MSE can enable the proposed method to obtain the more excellent

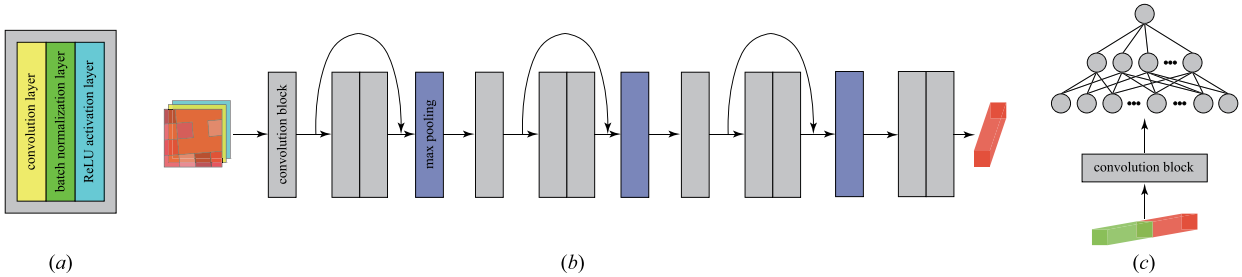


Fig. 5. The designed residual relation network. (*a*) A convolution block including a convolution layer, a batch normalization layer and a ReLU activation layer. (*b*) Deep feature extraction module combined with residual structure. (*c*) Relation learning module.

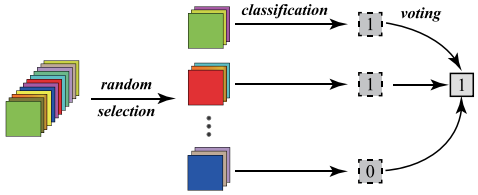


Fig. 6. Schematic of classification with voting strategy.

classification performance, and both its classification accuracy and robustness are improved.

D. Classification With Voting Strategy

In the meta-training stage, the model learns deep features based on three randomly selected bands. To maintain consistency, the voting strategy is used to classify the target HSIs by referring to the joint classification method in [53]. As shown in Fig. 6, for each sample, 3 bands were randomly selected for 10 times. In consequence, 10 different multiview features are generated. Then, the trained model classifies the 10 multiview features respectively, and the final class of the sample is determined according to the voting result. On the one hand, the voting strategy keeps the consistency of the dimension of input samples in the process of meta-training, fine-tuning and classification, and skillfully deals with the large difference in the number of bands between different HSIs. On the other hand, the voting classification process can be regarded as an ensemble learning method, which can further improve the accuracy and robustness of the classification results.

IV. Experiments and Discussion

A. Experimental Data Sets

To fully verify the effectiveness of the proposed method, 8 widely used public HSIs are selected for the experiments. The 8 HSIs are Houston 2013 (HS13), Botswana (BO), Kennedy Space Center (KSC), Chikusei (CH), University of Pavia (UP), Pavia Center (PC), Salinas (SA) and Indian Pines (IP). Referring to [43], [44], these HSIs are artificially divided into two groups. Specifically, the first four HSIs are used to build unlabeled source data set, and the last four HSIs are used to evaluate the classification performance of the model.

The details of HS13, BO, KSC, CH are listed in Table I. The four HSIs are acquired by different sensors respectively, with completely different classes, ground sample distances and spectral ranges. Building the source data set with different HSIs can effectively increase the diversity and richness of the samples, so as to improve the effect of meta-training. Specifically, all the samples in the four HSIs are treated as unlabeled, and 40,000 samples are randomly selected to build the unlabeled source data set. To make full use of the spatial-spectral information in HSIs, 28×28 patches within the neighborhood of center pixels are selected as samples. For each sample, 20 spatial-spectral multiview features, each of which contains 3 randomly selected bands, are generated using the method in Section III-B, to construct the tasks for meta-training. Therefore, the dimension of the built unlabeled source data set is (40000, 20, 28, 28, 3).

Table I also lists the details of four target HSIs including UP, PC, SA, IP. It should be noted that the four target HSIs and the pre-collected HSIs have completely different classes, that is, there is no intersection in their label space. After meta-training, the model is used to classify the four target HSIs respectively. Only 5 labeled samples per class are randomly selected from each target HSI to fine-tune the model, and the remaining samples are used to evaluate the classification performance. The 28×28 patches around pixels are still selected as input. Therefore, the dimension of input sample is $28 \times 28 \times 3$ in the phase of fine-tuning and classification according to the voting strategy in Fig. 6. The four target HSIs have different classes, sample distributions, spatial resolutions and spectral ranges, which can evaluate the classification performance of the model more comprehensively.

B. Experimental Settings

All algorithms are developed and implemented by Python and libraries such as Pytorch, sklearn and numpy. All results are generated on a computer equipped with an Intel(R) Xeon(R) Gold 6152 CPU and an Nvidia A100 PCIE GPU.

In the residual relation network, the size of convolution kernel in each residual convolution block is set to 3×3, and the number of convolution kernel gradually increases by 128,256,512,1024. The kernel size in max pooling layer is set to 2×2. In the relation learning module, the 1×1 convolution kernel is applied to reduce the dimension of the concatenations. The number of neurons in two fully connected layers is set to 128 and 1 respectively, and Dropout is adopted to further increase the generalization ability of the model.

TABLE I

DETAILS OF THE PRE-COLLECTED HSIS AND TARGET HSIS. HOUSTON 2013 (HS13), BOTSWANA(BO), KENNEDY SPACE CENTER (KSC), CHIKUSEI(CH), UNIVERSITY OF PAVIA (UP), PAVIA CENTER (PC), SALINAS (SA), INDIAN PINES (IP), GROUND SAMPLE DISTANCE (GSD)(M), SPATIAL SIZE (PIXEL), SPECTRAL RANGE (NM), AIRBORNE VISIBLE INFRARED IMAGING SPECTROMETER (AVIRIS), REFLECTIVE OPTICS SYSTEM IMAGING SPECTROMETER (ROSIS)

	Pre-collected HSIs				Target HSIs			
	HS13	BO	KSC	CH	UP	PC	SA	IP
Spatial size	349×1905	1476×256	512×614	2517×2335	610×340	1096×715	512×217	145×145
Spectral range	380-1050	400-2500	400-2500	363-1018	430-860	430-860	400-2500	400-2500
No. of bands	144	145	176	128	103	102	204	200
GSD	2.5	30	18	2.5	1.3	1.3	3.7	20
Sensor type	ITRES-CASI 1500	EO-1	AVIRIS	Hyperspec -VNIR-C	ROSIS	ROSIS	AVIRIS	AVIRIS
Areas	Houston	Botswana	Florida	Chikusei	Pavia	Pavia	California	Indiana
No. of classes	15	14	13	19	9	9	16	16
Labeled samples	15029	3248	5211	77592	42776	148152	54129	10249

In the process of unsupervised meta learning, 20-*way* 5-*shot* 15-*query* tasks are randomly constructed. The learning rate is set as 0.0001 and the number of iterations is set as 40,000, to ensure that the model is adequately trained. In the phase of fine-tuning, the number of iterations is set to 1000 and C_{target}-*way* 2-*shot* 3-*query* tasks are constructed (C_{target} is 9 for UP and PC, C_{target} is 16 for SA and IP). In addition, the Adam algorithm is selected for model optimization.

To quantitatively evaluate the classification results of the model, the overall accuracy (OA), average accuracy (AA) and kappa coefficient are selected as evaluation criteria. In addition, to alleviate the instability of results caused by random sample selection, the average value of 10 experiments is regarded as the final result.

C. Cross-Domain HSI Small Sample set Classification

In the experiments of cross-domain small sample set classification, the four pre-collected HSIs used to build the unlabeled source data set are completely different from the target HSIs. The proposed method, UM^{2}L, first performs meta-training on the source data set, and then transfers the learned knowledge to the classification tasks in target HSIs. To demonstrate the effectiveness of the proposed method, 6 different classification methods are selected for comparison, including the classical machine learning classifier RBF-SVM, a supervised deep learning model 3D-CNN [54], an semi-supervised method EMP+GCN [55], and three supervised meta learning methods DFSL+SVM [43], RN-FSC [44] and DCFSL [46].

Compared with other classifiers, RBF-SVM is better at processing high-dimensional data. In the experiments, grid search strategy is adopted to determine the parameters from a wide range $(2^{-3}, 2^{-2}, \ldots, 2^{8})$. 3D-CNN can directly process hyperspectral data using 3D convolution, and effectively make use of the spatial-spectral information. EMP+GCN is a semi-supervised method based on EMP features and GCN model, which can effectively utilize the potential features in unlabeled samples. For the above three models, only 5 labeled samples per class from the target HSIs are randomly selected for supervised training. DFSL+SVM, RN-FSC and DCFSL are all supervised meta learning methods. DFSL+SVM first learns a feature extraction network using the labeled source data set, then extracts features from the target HSI, and finally performs classification with RBF-SVM. RN-FSC constructs an end-to-end framework for small sample set classification, and performs classification by comparing the similarity between the extracted deep features. DCFSL combines small sample set classification with a domain adaptation method, and attempts to extract the domain-invariant features hy introducing conditional adversarial domain adaptation strategy. Referring to relevant literatures, for DFSL+SVM and RN-FSC, four pre-collected HSIs including HS13, BO, KSC and CH, are used to construct the labeled source data set. For DCFSL, only the CH data set is used for meta-training.

Table II reports the classification results of different methods on the four target HSIs, from which we can obtain the following observations.

(1) On the whole, the classification performance of RBF-SVM is inferior to that of deep learning models. The deep learning models can extract more informative and discriminative features layer by layer, to obtain higher classification accuracy.
(2) The classification accuracy of EMP+GCN is significantly improved. Compared with 3D-CNN, the OA of EMP+GCN is improved by 11.00%, 14.11%, 2.93% and 7.32% respectively on four target HSIs.
(3) Compared with RBF-SVM, 3D-CNN and EMP+GCN, the classification accuracy of the three supervised meta learning methods is further improved. In the small sample set scenario, the deep learning models trained by a few labeled samples often has the problem of overfitting. In contrast, the supervised meta learning method can utilize the knowledge learned from the source data set to guide the classification on the target HSIs, so as to obtain better classification performance.
(4) Comparing the three supervised meta learning methods (DFSL+SVM, RN-FSC and DCFSL) and UM^{2}L, it can be found that on the PC data sets, the classification accuracy of UM^{2}L is slightly lower than that of the supervised meta learning methods. On the UP, SA and IP data sets, the classification accuracy of UM^{2}L is higher than that of the supervised meta learning methods. Specifically, on the PC data set, the OA of UM^{2}L is

TABLE II

THE CLASSIFICATION RESULTS OF DIFFERENT METHODS IN THE CROSS-DOMAIN SCENARIO. SD DENOTES THE STANDARD DEVIATION OF 10 EXPERIMENTAL RESULTS

HSI	Criteria	RBF-SVM *Mean±SD*	3D-CNN *Mean±SD*	EMP+GCN *Mean±SD*	DFSL+SVM *Mean±SD*	RN-FSC *Mean±SD*	DCFSL *Mean±SD*	UM^{2}L *Mean±SD*
UP	OA	62.12±4.47	66.88±2.82	77.88±1.15	75.61±4.56	80.16±2.40	80.65±4.68	80.92±3.55
	AA	64.12±2.36	61.89±3.18	76.48±0.73	75.91±2.92	78.36±3.38	81.37±1.68	77.77±3.19
	kappa	0.5328±0.0448	0.5726±0.0327	0.7037±0.0141	0.6909±0.0506	0.7423±0.0280	0.7499±0.0527	0.7556±0.0406
PC	OA	81.38±3.66	80.91±5.09	95.02±0.94	94.90±1.97	95.93±1.16	97.45±0.55	94.27±1.43
	AA	77.70±2.82	64.06±5.22	88.66±1.24	86.41±2.33	87.94±2.41	94.25±0.94	84.85±2.93
	kappa	0.7476±0.0452	0.7398±0.0657	0.9301±0.0129	0.9289±0.0262	0.9427±0.0160	0.9640±0.0077	0.9194±0.0199
SA	OA	78.34±3.03	80.12±3.08	83.05±1.47	85.31±3.47	84.54±1.77	89.13±1.69	90.34±2.54
	AA	83.91±1.63	82.62±2.19	83.61±1.63	91.23±1.11	89.47±1.36	94.02±0.72	93.04±1.30
	kappa	0.7601±0.0329	0.7802±0.0342	0.8594±0.0126	0.8371±0.0378	0.8288±0.0194	0.8794±0.0184	0.8929±0.0280
IP	OA	45.75±4.03	54.46±2.84	61.78±2.77	63.11±3.13	62.47±1.78	64.58±1.89	72.09±2.20
	AA	46.90±3.05	51.99±3.16	56.64±2.06	62.23±4.09	62.70±1.03	76.36±1.76	64.84±3.17
	kappa	0.3958±0.0430	0.4954±0.0280	0.5778±0.0292	0.5853±0.0320	0.5825±0.0187	0.6010±0.0204	0.6883±0.0237

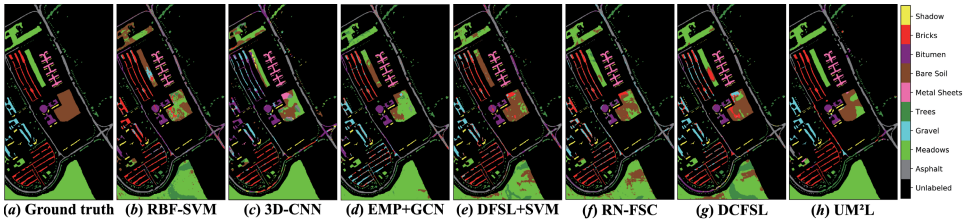


Fig. 7. The classification maps resulting from different methods on UP in the cross-domain scenario.

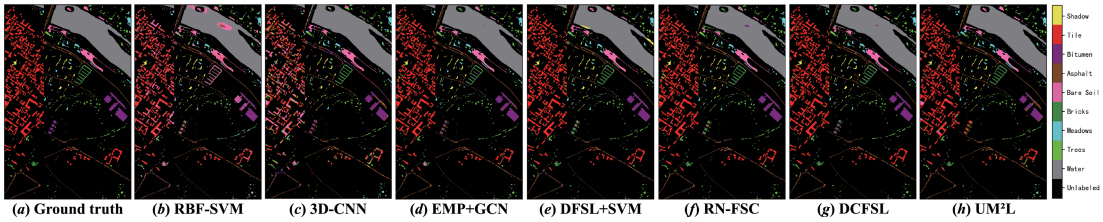


Fig. 8. The classification maps resulting from different methods on PC in the cross-domain scenario.

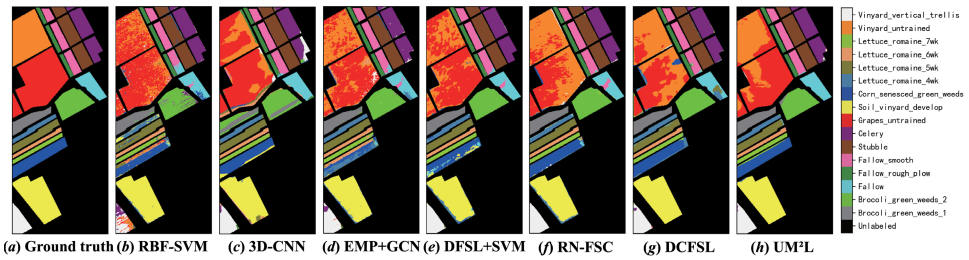


Fig. 9. The classification maps resulting from different methods on SA in the cross-domain scenario.

about 0.6%-3.2% lower than that of the supervised meta learning methods. On the UP, SA and IP data sets, the OA of UM^{2}L is higher than that of the three supervised meta learning methods. Especially on the IP data set, the OA of UM^{2}L increases by about 7.5%-9.6%.

It should be emphasized that UM^{2}L obtains the above results under the condition that only 5 labeled samples per class in the target HSIs are used for fine-tunning. Compared with the three supervised meta learning methods, UM^{2}L does not need to construct a labeled source data set, greatly reducing the number of requisite labeled samples. The excellent performance of UM^{2}L can be attributed to two points: unsupervised meta learning and deep residual relation network. On the one hand, the spatial-spectral multiview features of each sample are generated to construct tasks, enabling the model to perform sufficient meta-training. On the other hand, the deep residual relation network can make full use of the deep spatial-spectral features in the input samples, to obtain better learning effect and classification accuracy.

To directly observe the classification results of different methods and perform qualitative evaluation, Figs. 7-10 show the classification maps of different methods. It can be seen that methods with higher classification accuracy can produce more accurate classification maps. On the UP and PC data

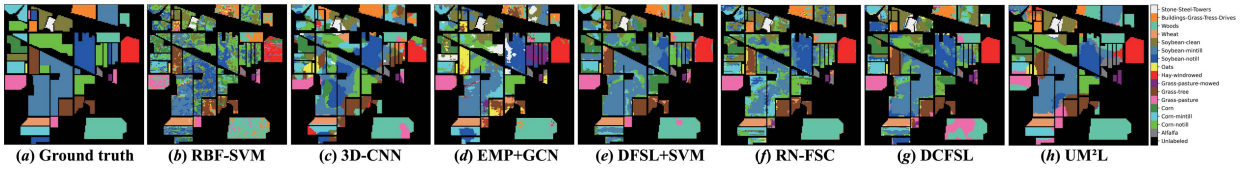


Fig. 10. The classification maps resulting from different methods on IP in the cross-domain scenario.

sets, the classification maps of the proposed method have certain misclassifications. However, from the perspective of class consistency, the classification maps of the proposed method are smoother compared with other methods. It is believed that this phenomenon benefits from the samples with large neighborhood and the data augmentation methods along spatial dimensions during the process of constructing multiview features. The former can make full use of the spatial information around the center pixels, while the latter can effectively improve the robustness of the learned spatial features. On SA and IP data sets, the proposed method has the most excellent classification results, and its classification maps are also the closest to the ground truths.

D. In-Domain HSI Small Sample set Classification

Supervised meta learning methods must utilize the labeled samples from other HSIs for meta-training. In contrast, unsupervised meta learning methods are able to perform meta-training, fine-tuning and classification on the same target HSI. Therefore, unsupervised meta learning methods are more flexible. This section explores the classification performance of UM^{2}L in the in-domain scenario. Four different methods including an unsupervised classification method 3D-CAE [29], a self-supervised learning method SimCLR and two unsupervised meta learning methods CACTUs [50] and UMTRA [51] are used for comparison.

3D-CAE combines 3D convolutional layer with stack autoencoder and utilizes reconstruction loss for training. SimCLR contains a siamese feature extractor and a MLP predictor, and can perform self-supervised learning based on contrastive loss. Referring to [33], PCA and data augmentation are used to generate input samples for contrastive learning. CACTUs and UMTRA are two typical unsupervised meta learning methods designed for natural image. CACTUs assigns pseudo labels to the unlabeled samples by clustering, to construct tasks for meta-training. UMTRA performs data augmentation on each unlabeled sample, and treats each sample and its augmentation samples as a class. In the experiments, PCA is firstly used to reduce the dimensionality of HSIs and the first three principal components are retained. Then, CACTUs and UMTRA methods are used for unsupervised meta learning and classification. The above four methods and the proposed method all first treat the target HSIs as unlabeled data for unsupervised or self-supervised learning, then carry out classification and evaluation on the same HSIs. It should be noted that, for the three unsupervised meta learning methods (CACTUs, UMTRA and UM^{2}L), the target HSIs are both the source data sets and the target data sets, which is different from the cross-domain classification scenario. In addition, 5 labeled samples per class are randomly selected for fine-tunning.

TABLE III
THE CLASSIFICATION RESULTS OF DIFFERENT METHODS IN THE IN-DOMAIN SCENARIO. SD DENOTES THE STANDARD DEVIATION OF 10 EXPERIMENTAL RESULTS

HSI	Criteria	3D-CAE *Mean* ±*SD*	PCA-SimCLR *Mean* ±*SD*	PCA-CACTUs *Mean* ±*SD*	PCA-UMTRA *Mean* ±*SD*	UM^{2}L *Mean* ±*SD*
UP	OA	67.11 ±1.98	78.18 ±4.77	63.00 ±3.11	72.76 ±4.35	81.71 ±3.14
	AA	64.93 ±1.31	75.31 ±2.64	66.59 ±3.60	72.04 ±3.40	78.48 ±2.83
	kappa	0.5821 ±0.0240	0.7242 ±0.0532	0.5392 ±0.0367	0.6579 ±0.0479	0.7631 ±0.0365
PC	OA	90.03 ±2.39	88.78 ±2.25	91.69 ±3.17	91.81 ±1.91	94.95 ±1.19
	AA	78.41 ±3.52	78.18 ±3.99	81.30 ±4.49	80.04 ±3.81	86.08 ±2.94
	kappa	0.8603 ±0.0335	0.8452 ±0.0298	0.8849 ±0.0423	0.8852 ±0.0260	0.9289 ±0.0167
SA	OA	78.65 ±2.59	87.78 ±3.39	78.58 ±1.75	87.06 ±2.02	90.30 ±2.64
	AA	81.36 ±1.72	88.16 ±3.06	80.47 ±2.54	86.97 ±1.70	93.34 ±1.38
	kappa	0.7631 ±0.0287	0.8647 ±0.0370	0.7629 ±0.0189	0.8565 ±0.0225	0.8923 ±0.0292
IP	OA	46.76 ±5.37	68.82 ±3.94	49.76 ±1.95	48.17 ±1.84	71.36 ±2.56
	AA	49.13 ±2.85	60.28 ±5.16	49.28 ±4.67	50.36 ±1.77	66.62 ±1.73
	kappa	0.4077 ±0.0550	0.6528 ±0.0412	0.4378 ±0.0200	0.4281 ±0.0171	0.6785 ±0.0280

The classification results of different methods in the in-domain scenario are listed in Table III. Again, the mean and standard deviation of 10 experiments are listed to enhance the persuasiveness of results. Three observations can be made.

(1) On the whole, the classification performance of the three unsupervised meta learning methods is better than that of 3D-CAE. In most cases, the OA of 3D-CAE is lower than that of the three meta learning methods, except that on the UP data set, the OA of 3D-CAE is significantly higher than that of PCA-CACTUs.
(2) On the UP, SA and IP data sets, the classification accuracy of PCA-SimCLR is significantly improved compared with those of 3D-CAE, PCA-CACTUs and PCA-UMTRA, indicating the potential of contrast learning in HSI small sample set classification.
(3) CACTUs and UMTRA construct the tasks by clustering and data augmentation respectively, which cannot fully adapt to the characteristics of high-dimensional, nonlinearity, and spatial-spectral information fusion in hyperspectral data. Therefore, the classification accuracy of PCA-CACTUs and PCA-UMTRA are not satisfactory.

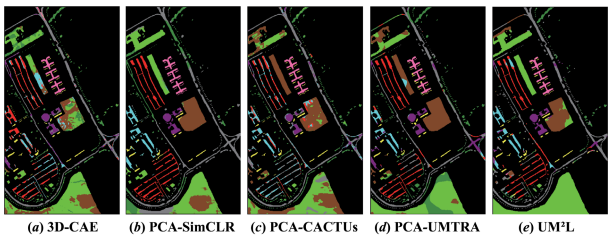


Fig. 11. The classification maps resulting from different methods on UP in the in-domain scenario.

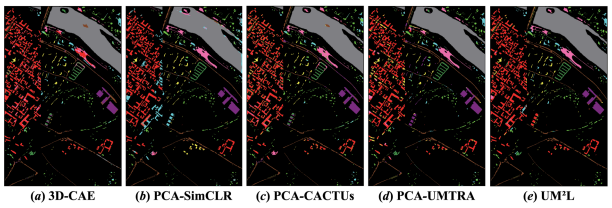


Fig. 12. The classification maps resulting from different methods on PC in the in-domain scenario.

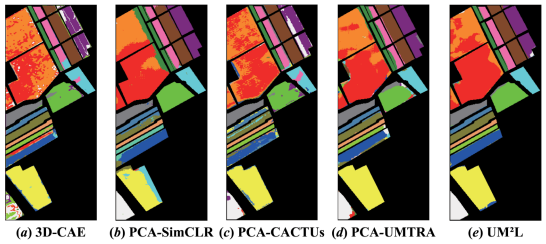


Fig. 13. The classification maps resulting from different methods on SA in the in-domain scenario.

(4) The proposed method achieves the highest OA, AA and kappa coefficient on four target HSIs. Specifically, the OA of the proposed method is 3.53%, 3.14%, 2.52% and 2.54% higher than that of the second place in the four target HSIs, respectively.

The unsupervised learning method 3D-CAE only utilizes reconstruction loss for model training, and cannot make full use of the deep spatial-spectral features in HSIs. PCA-SimCLR and the proposed method both utilize the same unlabeled data set for pre-training. The proposed method achieves better classification performance on four target HSIs, indicating the effectiveness of meta-training. The general-purpose representations learned in the meta-training phase can effectively guide the classification on the target HSIs. Therefore, compared with 3D-CAE and PCA-SimCLR, the classification performance of the proposed method is further improved. Compared with PCA-CACTUs and PCA-UMTRA, the proposed method constructs tasks by generating spatial-spectral multiview features, and utilizes the designed deep residual relation network for meta-training and small sample set classification based on the voting strategy, obtaining better classification results. This directly shows the effectiveness of the task construction method, model design, and voting strategy in the proposed method.

Fig. 11-14 show the classification maps of the different methods in the in-domain scenario. It can be seen that, there are a large number of misclassifications in the classification maps generated by 3D-CAE, PCA-SimCLR, PCA-CACTUs and PCA-UMTRA, which cannot accurately display the spatial distribution information of ground objects. The classification maps of UM^{2}L are more accurate and closest to the ground truths, which once again verifies the effectiveness of the proposed method from a visual perspective.

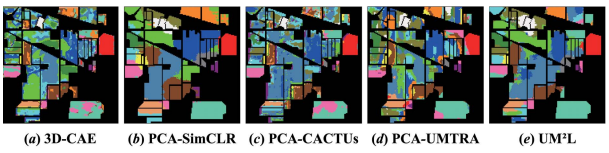


Fig. 14. The classification maps resulting from different methods on IP in the in-domain scenario.

TABLE IV

OVERALL ACCURACY(%) OF THE PROPOSED METHOD ON THE TARGET HSIS WITH DIFFERENT NUMBERS OF SUPPORT SAMPLES AND QUERY SAMPLES

HSI	1-*shot* 19-*query*	5-*shot* 15-*query*	10-*shot* 10-*query*
UP	80.10	80.92	79.83
PC	93.05	94.27	93.17
SA	89.66	90.34	88.52
IP	70.67	72.09	70.83

TABLE V

OVERALL ACCURACY(OA, %) OF THE PROPOSED METHOD ON THE TARGET HSIS WITH DIFFERENT FEATURE EXTRACTION MODULE

No.	**1**	**2**	**3**	**4**
feature extraction module	conv_1 res_1 maxpool_1 conv_2 res_2 maxpool_2 conv_3 conv_4 conv_5	conv_1 res_1 maxpool_1 conv_2 res_2 maxpool_2 conv_3 res_3 maxpool_3 conv_4 conv_5	conv_1 res_1 maxpool_1 conv_2 res_2 maxpool_2 conv_3 res_3 maxpool_3 conv_4 res_4 maxpool_4 conv_5	conv_1 res_1 maxpool_1 conv_2 res_2 maxpool_2 conv_3 res_3 maxpool_3 conv_4 res_4 maxpool_4 conv_5 res_5 maxpool_5
UP	78.43	80.92	79.76	78.93
PC	89.45	94.27	92.27	91.36
SA	85.96	90.34	89.43	89.94
IP	70.33	72.09	70.58	70.85

E. Hyperparameters Analysis

Meta learning methods take tasks as basic units for training, and the form of tasks directly affects the learning effectiveness of the model. Referring to [43], [44], [47], the number of classes in a task is set to 20, and the two hyperparameters *shot* and *query* are explored. Table IV lists OA of the proposed method with different numbers of support and query samples in a task. It can be seen that, when *shot* is 5 and *query* is 15, the proposed method can achieve higher classification accuracy. However, when the value of *shot* is too large or too small, the accuracy will decline to different degrees.

To make full use of the deep features in HSIs, a deep residual relation network is designed as the model for meta

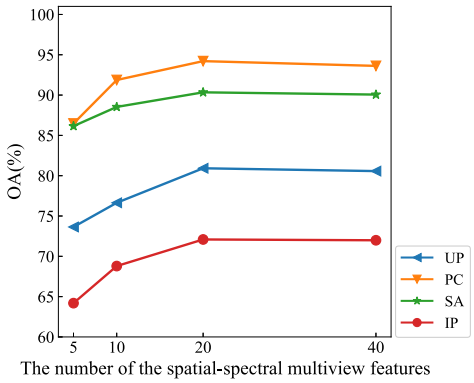


Fig. 15. The influence of the number of multiview features on classification results.

TABLE VI

OVERALL ACCURACY(%) OF THE PROPOSED METHOD WITH DIFFERENT TASK CONSTRUCTION MODES. SDA STANDS FOR SPATIAL DATA AUGMENTATION AND BRS STANDS FOR BANDS RANDOM SELECTION

HSI	PCA+SDA	BRS	BRS+SDA				
			crop	cutout	rotation	flip	all
UP	73.38	70.83	76.87	73.19	74.25	76.06	80.92
PC	91.64	88.76	92.18	90.39	91.57	91.80	94.27
SA	88.83	84.65	89.01	87.29	88.04	87.82	90.34
IP	68.89	61.60	70.35	69.41	67.78	68.37	72.09

learning. The influence of network structure on classification accuracy is explored by increasing or decreasing the residual convolution blocks in the feature extraction module. In the designed model, a residual convolution block consists of a convolution block, a residual block and a pooling layer. Table V shows the relationship between network structure and the classification accuracy. In Table V, the network structure numbered 1, 2, 3, 4 possesses 2, 3, 4 and 5 residual convolution block, respectively, and the network structure numbered 2 corresponds to Fig. 5(b). As we can see, the second network structure enables the model to obtain more excellent classification performance, while too deep or too shallow network will lead to a decline in classification accuracy.

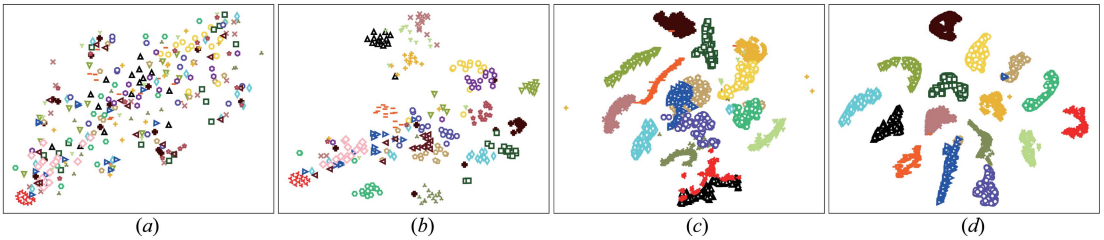
In the proposed method, it is a key step to construct tasks for unsupervised meta learning using the spatial-spectral multiview features. To verify its effectiveness, the influence of the number of multiview features and different task construction modes are further analyzed. Theoretically within a certain range, as the number of multiview features increases, the model can learn richer features and possesses more excellent classification ability. In Fig. 15, all curves show a trend of rising at first and then tending to be stable, indicating that excessive multiview features can no longer improve the classification performance. Table VI lists the classification accuracy under different task construction modes, which is actually an ablation study on task construction methods. PCA+SDA means that PCA replaces bands random selection and combines with spatial data augmentation to construct tasks; BRS means that only bands random selection is used to construct tasks without data augmentation; different modes in BRS+SDA explore the influence of different data augmentation techniques on classification results. Obviously, both PCA+SDA and BRS lead to significant decrease in accuracy, indicating the importance of combining the multiview features along spectral and spatial dimension. In spatial dimension, data augmentation using crop, cutout, rotation or flip methods alone also lead to varying degrees of decline in classification accuracy, and the highest accuracy can be achieved by the comprehensive application of multiple techniques.

F. Feature Visualization Analysis

To show the effect of unsupervised meta learning more intuitively, feature visualization analysis is carried out. Specifically, the t-SNE (t-distributed Stochastic Neighbor Embedding) algorithm [56] is utilized to reduce the dimensionality of the input samples and the features extracted by the designed residual relation model, and the separability among the obtained data is observed. Fig. 16 shows the visualization results. Fig. 16(*a*) and Fig. 16(*b*) show the original input samples and features generated by the feature extraction module in the phase of unsupervised meta learning, respectively. The separability among the input samples is very poor as the tasks are randomly constructed with unlabeled samples, while the separability of extracted features is significantly improved. This indicates that the model can extract the features with small inner-class distance and large inter-class distance after unsupervised meta learning. In addition, the SA data set is taken as an example to visualize the input samples and the extracted features in the classification phase. It can be seen that, the model after meta-training can effectively improve the separability among the samples, and make the samples belonging to same class more aggregated and the samples belonging to different classes separated from each other. It can be seen from the visualization results that, the process of unsupervised meta learning can endow the model with the ability of extracting more discriminant features, thus improving the accuracy in the subsequent classification tasks.

G. Comprehensive Comparison

Existing literatures and experimental results have shown that, the meta learning-based methods can effectively improve the classification accuracy of HSIs in small sample set scenario. Table VII shows a comprehensive comparison of four meta learning-based classification methods in terms of the number of required labeled samples, execution time and classification performance. The required labeled samples includes the labeled samples used for meta-training and fine-tuning. The execution time of each method can be artificially divided into three parts: meta-training, fine-tuning, and classification. The classification performance of each method on different data sets is measured by OA. In terms of the number of required labeled samples, DFSL+SVM and RN-FSC both need to build source data sets containing a large number of labeled samples in the meta-training phase, so the number of labeled samples they need is the largest. DCFSL only utilizes the CH data set



(a) (b) (c) (d)

Fig. 16. Visual presentation of the input samples and extracted features. (*a*) Input samples in meta-training phase. (*b*) Features generated by feature extraction module in meta-training phase. (*c*) Input samples in classification phase. (*d*) Features generated by feature extraction module in classification phase.

TABLE VII

A COMPREHENSIVE COMPARISON OF META LEARNING-BASED CLASSIFICATION METHODS

Target HSI	items	DFSL+SVM	RN-FSC	DCFSL	UM^{2}L (cross-domain)	UM^{2}L (in-domain)
UP	labeled samples	11000+45	11000+45	3600+45	0+45	0+45
	meta-training	118.23 *min*	319.84 *min*	24.32 *min*	316.87 *min*	315.34 *min*
	fine-tuning	9.15 *s*	80.47 *s*	0.10 *s*	363.32 *s*	363.32 *s*
	classification	1.88 *s*	19.12 *s*	4.56 *s*	141.83 *s*	141.83 *s*
	OA	75.61	80.16	80.65	80.92	81.71
PC	labeled samples	11000+45	11000+45	3600+45	0+45	0+45
	meta-training	118.23 *min*	319.84 *min*	26.31 *min*	316.87 *min*	316.57 *min*
	fine-tuning	25.47 *s*	79.13 *s*	0.09 *s*	361.02 *s*	361.02 *s*
	classification	3.43 *s*	73.79 *s*	15.44 *s*	483.91 *s*	483.91 *s*
	OA	94.90	95.93	97.45	94.27	94.95
SA	labeled samples	11000+80	11000+80	3600+80	0+80	0+80
	meta-training	118.23 *min*	319.84 *min*	39.93 *min*	316.87 *min*	317.04 *min*
	fine-tuning	9.63 *s*	186.33 *s*	0.20 *s*	381.13 *s*	381.13 *s*
	classification	1.82 *s*	44.63 *s*	7.03 *s*	204.30 *s*	204.30 *s*
	OA	85.31	84.54	89.13	90.34	90.30
IP	labeled samples	11000+80	11000+80	3600+80	0+80	0+80
	meta-training	118.23 *min*	319.84 *min*	38.17 *min*	316.87 *min*	316.53 *min*
	fine-tuning	2.92 *s*	185.65 *s*	0.17 *s*	375.08 *s*	375.08 *s*
	classification	1.65 *s*	9.18 *s*	1.33 *s*	42.89 *s*	42.89 *s*
	OA	63.11	62.47	64.58	72.09	71.36

for meta-training, effectively reducing the number of required labeled samples, but nearly 3700 labeled samples are still needed in the whole process. In both the cross-domain and in-domain scenarios, the proposed method only utilizes the source data set without any label information for unsupervised meta learning. Therefore, the number of required labeled samples in the meta-training phase is 0, greatly alleviating the dependence of the model on labeled samples. In terms of execution time, the scale of source data set in DCFSL is the smallest, so the required meta-training time is the shortest. DFSL+SVM and DCFSL use RBF-SVM and KNN classifiers for fine-tuning and classification respectively, so their fine-tuning and classification time is significantly shorter than other methods. UM^{2}L utilizes the voting strategy to classify target HSIs, and actually runs the fine-tuning and classification process for 10 times. Therefore, the proposed method requires longer fine-tuning and classification time. In terms of classification performance, the OA of the proposed method is lower than that of RN-FSC and DCFSL on the PC data set. However, the proposed method can achieve higher classification accuracy than other meta learning methods on the other three data sets, which directly indicates the effectiveness of the proposed method in HSI small sample set classification.

In summary, the proposed method greatly reduces the number of requisite labeled samples in the whole classification process, and can achieve competitive or better classification results within acceptable execution time.

V. CONCLUSION

To further improve the accuracy of HSI small sample set classification and reduce the dependence of meta learning models on the labeled samples, a novel unsupervised meta learning method with multiview constraints, UM^{2}L, is proposed in this paper. The proposed method first builds unlabeled source data set using HSIs without any label, and then generates the spatial-spectral multiview features for each sample based on the idea of multiview learning. Specifically, spectral multiview features are obtained by band random selection and spatial multiview features are obtained by data augmentation, to generate multiple spatial-spectral multiview features together. The tasks for unsupervised meta learning is constructed with the generated multiview features. Finally, a deep residual relation network is designed to perform meta-training and small sample set classification based on the voting strategy. The experimental results show that, compared with existing supervised meta learning methods and other advanced classification models, the proposed method can achieve competitive or better classification performance in small sample set scenarios, and significantly reduce the

number of requisite labeled samples in the whole classification process.

It should be pointed out that the source data sets and the target HSIs have different data distributions in the cross-domain classification scenario. In the future, the domain adaptation methods will be introduced to narrow the distribution difference between the two data sets in meta-training and the performance of the combination of domain adaptation methods and the proposed method will be explored in details, to further improve the accuracy and robustness of classification results.

REFERENCES

[1] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, Jun. 2019.

[2] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.

[3] Y. Ren, Y. Zhang, W. Wei, and L. Li, "A spectral-spatial hyperspectral data classification approach using random forest with label constraints," in *Proc. IEEE Workshop Electron., Comput. Appl.*, May 2014, pp. 344–347.

[4] P. Deepa and K. Thilagavathi, "Data reduction techniques of hyperspectral images: A comparative study," in *Proc. 3rd Int. Conf. Signal Process., Commun. Netw. (ICSCN)*, Mar. 2015, pp. 1–6.

[5] Y. Gu, T. Liu, X. Jia, and J. A. Benediktsson, "Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3235–3247, Jun. 2016.

[6] S. Jia, J. Hu, J. Zhu, X. Jia, and Q. Li, "Three-dimensional local binary patterns for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2399–2413, Apr. 2017.

[7] S. Jia, B. Deng, H. Xie, and L. Deng, "A Gabor feature fusion framework for hyperspectral imagery classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2394–2397.

[8] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 279–317, Dec. 2019.

[9] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[10] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Apr. 2017.

[11] A. Mughees and L. Tao, "Multiple deep-belief-network-based spectral-spatial classification of hyperspectral images," *Tsinghua Sci. Technol.*, vol. 24, no. 2, pp. 183–194, Apr. 2019.

[12] M. Zhang, W. Li, and Q. Du, "Diverse region-based CNN for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2623–2634, Jun. 2018.

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

[14] X. He and Y. Chen, "Transferring CNN ensemble for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 1–5, May 2020.

[15] C. Yu, R. Han, M. Song, C. Liu, and C.-I. Chang, "A simplified 2D-3D CNN architecture for hyperspectral image classification based on spatial–spectral fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2485–2501, 2020.

[16] B. Liu, X. Yu, P. Zhang, A. Yu, Q. Fu, and X. Wei, "Supervised deep feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 1909–1921, Nov. 2017.

[17] 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 Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Jul. 2016.

[18] Z. Xue, X. Yu, B. Liu, X. Tan, and X. Wei, "HResNetAM: Hierarchical residual network with attention mechanism for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3566–3580, 2021.

[19] L. Zhi, X. Yu, B. Liu, and X. Wei, "A dense convolutional neural network for hyperspectral image classification," *Remote Sens. Lett.*, vol. 10, no. 1, pp. 59–66, Jan. 2019.

[20] M. E. Paoletti *et al.*, "Capsule networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2145–2160, Apr. 2019.

[21] C. Yu, R. Han, M. Song, C. Liu, and C.-I. Chang, "Feedback attention-based dense CNN for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022.

[22] M. Lin, Q. Chen, and S. Yan, "Network in network," in *Proc. Int. Conf. Learn. Represent.*, 2014.

[23] B. Liu, X. Yu, P. Zhang, X. Tan, A. Yu, and Z. Xue, "A semi-supervised convolutional neural network for hyperspectral image classification," *Remote Sens. Lett.*, vol. 8, no. 9, pp. 839–848, Sep. 2017.

[24] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5046–5063, Sep. 2018.

[25] X. Wang, K. Tan, Q. Du, Y. Chen, and P. Du, "Caps-TripleGAN: GAN-assisted CapsNet for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7232–7245, Sep. 2019.

[26] A. Qin, Z. Shang, J. Tian, Y. Wang, T. Zhang, and Y. Y. Tang, "Spectral-spatial graph convolutional networks for semisupervised hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 241–245, Sep. 2019.

[27] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang, and J. Yang, "Multiscale dynamic graph convolutional network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3162–3177, May 2020.

[28] L. Mou, P. Ghamisi, and X. X. Zhu, "Unsupervised spectral–spatial feature learning via deep residual Conv–Deconv network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 391–406, Jan. 2018.

[29] S. Mei, J. Ji, Y. Geng, Z. Zhang, X. Li, and Q. Du, "Unsupervised spatial–spectral feature learning by 3D convolutional autoencoder for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6808–6820, Apr. 2019.

[30] X. He, Y. Chen, and P. Ghamisi, "Heterogeneous transfer learning for hyperspectral image classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3246–3263, May 2019.

[31] C. Deng, Y. Xue, X. Liu, C. Li, and D. Tao, "Active transfer learning network: A unified deep joint spectral–spatial feature learning model for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1741–1754, Mar. 2019.

[32] 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 Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6440–6461, Nov. 2018.

[33] B. Liu, A. Yu, X. Yu, R. Wang, K. Gao, and W. Guo, "Deep multiview learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 1–15, Sep. 2020.

[34] Y. Dong, C. Yang, and Y. Zhang, "Deep metric learning with online hard mining for hyperspectral classification," *Remote Sens.*, vol. 13, no. 7, p. 1368, Apr. 2021.

[35] Y. Dong, W. Shi, B. Du, X. Hu, and L. Zhang, "Asymmetric weighted logistic metric learning for hyperspectral target detection," *IEEE Trans. Cybern.*, early access, May 26, 2021, doi: 10.1109/TCYB.2021.3070909.

[36] J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and J. Li, "Hyperspectral image classification using random occlusion data augmentation," *EEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1751–1755, Nov. 2019.

[37] B. Liu *et al.*, "Morphological attribute profile cube and deep random forest for small sample classification of hyperspectral image," *IEEE Access*, vol. 8, pp. 117096–117108, 2020.

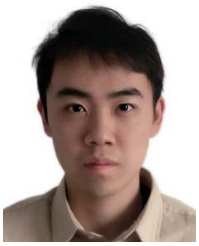
[38] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," in *Proc. Conf. Neural Inform. Process. Syst.*, 2017, pp. 4080–4090. [Online]. Available: http://papers.nips.cc/paper/6996-prototypicalnetworks-for-few-shot-learning

[39] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 1199–1208. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2018/html/Sung_Learning_to_Compare_CVPR_2018_paper.html

[40] W. Jiang, K. Huang, J. Geng, and X. Deng, "Multi-scale metric learning for few-shot learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 3, pp. 1091–1102, Mar. 2021.

[41] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 1126–1135.

[42] N. Mishra, M. Rohaninejad, X. Chen, and P. Abbeel, "A simple neural attentive meta-learner," in *Proc. Int. Conf. Learn. Represent.*, 2018.
[43] B. Liu, X. Yu, A. Yu, P. Zhang, G. Wan, and R. Wang, "Deep few-shot learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2290–2304, Apr. 2019.
[44] K. Gao, B. Liu, X. Yu, J. Qin, P. Zhang, and X. Tan, "Deep relation network for hyperspectral image few-shot classification," *Remote Sens.*, vol. 12, no. 6, p. 923, Mar. 2020.
[45] X. Ma, S. Ji, J. Wang, J. Geng, and H. Wang, "Hyperspectral image classification based on two-phase relation learning network," *IEEE Trans. Geoence Remote Sens.*, vol. 57, no. 12, pp. 10398–10409, Dec. 2019.
[46] 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, pp. 1–18, 2022.
[47] K. Gao, W. Guo, X. Yu, B. Liu, A. Yu, and X. Wei, "Deep induction network for small samples classification of hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3462–3477, 2020.
[48] K. Gao, B. Liu, X. Yu, P. Zhang, X. Tan, and Y. Sun, "Small sample classification of hyperspectral image using model-agnostic meta-learning algorithm and convolutional neural network," *Int. J. Remote Sens.*, vol. 42, no. 8, pp. 3090–3122, Apr. 2021.
[49] K. Gao, X. Yu, X. Tan, B. Liu, and Y. Sun, "Small sample classification for hyperspectral imagery using temporal convolution and attention mechanism," *Remote Sens. Lett.*, vol. 12, no. 5, pp. 510–519, May 2021.
[50] K. Hsu, S. Levine, and C. Finn, "Unsupervised learning via meta-learning," in *Proc. Int. Conf. Learn. Represent.*, 2019.
[51] S. Khodadadeh, L. Bölöni, and M. Shah, "Unsupervised meta-learning for few-shot image classification," in *Proc. 33rd Conf. Neural Inf. Process. Syst. (NeurIPS)*, 2019.
[52] J. Zhao, X. Xie, X. Xu, and S. Sun, "Multi-view learning overview: Recent progress and new challenges," *Inf. Fusion*, vol. 38, pp. 43–54, Nov. 2017.
[53] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, Feb. 2017.
[54] B. Liu, X. Yu, P. Zhang, X. Tan, R. Wang, and L. Zhi, "Spectral-spatial classification of hyperspectral image using three-dimensional convolution network," *J. Appl. Remote Sens.*, vol. 12, no. 1, pp. 1–18, 2018.
[55] B. Liu, K. Gao, A. Yu, W. Guo, R. Wang, and X. Zuo, "Semisupervised graph convolutional network for hyperspectral image classification," *J. Appl. Remote Sens.*, vol. 14, no. 2, pp. 1–14, 2020.
[56] L. van der Maaten and G. Hinton, "Viualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.



Kuiliang Gao received the B.S. degree in remote sensing science and technology from PLA Strategic Support Force Information Engineering University, Zhengzhou, China, in 2019, where he is currently pursuing the M.S. degree.

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

Bing Liu received the B.S. degree in measurement and control engineering, the M.S. degree in pattern recognition and intelligent system, and the Ph.D. degree in surveying and mapping science and technology from Information Engineering University, Zhengzhou, China, in 2013, 2016, and 2019, respectively.

He is currently working as an Associate Professor at PLA Strategic Support Force Information Engineering University. His research interests include machine learning, pattern recognition, and signal processing in Earth observation. He is an Active Reviewer of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE TRANSACTIONS ON CYBERNETICS, IEEE ACCESS, the *International Journal of Remote Sensing*, the *Remote Sensing Letter*, and the *Journal of Applied Remote Sensing*.

Xuchu Yu received the Ph.D. degree from the Institute of Surveying and Mapping, PLA Strategic Support Force Information Engineering University, Zhengzhou, China, in 1997.

He is currently working as a Professor at the Institute of Surveying and Mapping, PLA Strategic Support Force Information Engineering University, and a Ph.D. Supervisor. His research interests include photogrammetry, remote sensing, and pattern recognition.

Anzhu Yu received the B.S. degree in remote sensing science and technology, the M.S. degree in photogrammetry and remote sensing, and the Ph.D. degree from the Institute of Surveying and Mapping, PLA Strategic Support Force Information Engineering University, Zhengzhou, China, in 2011, 2014, and 2017, respectively.

He is currently working as an Associate Professor at PLA Strategic Support Force Information Engineering University. His research interest includes signal processing in Earth observation.