

Spatial-Spectral Deep Transfer Learning Network for Few-Shot Hyperspectral Image Classification

Jay Kishor Sah*, Dipak Kumar Ghosh, Usha Chauhan, Ravi Anand
Department of Electrical, Electronics and Communication Engineering,
Galgotias University, India

**Corresponding author E-mail: Jaykishor.sah@galgotiasuniversity.edu.in*

Received 12 January 2026; Revised 4 August 2026; Accepted 4 September 2026

Abstract

Background and Objectives: In recent decades, classification of hyperspectral images (HSIs) has made significant progress with the development of deep learning (DL). However, learning with a few samples remains a difficulty due to the high cost of annotating samples and potential errors in manual evaluation. Most DL-related methods require a sufficient number of annotated samples to learn the intricate spatial-spectral structure and spectrum of HSI data, making them unsuitable for practical situations with limited samples. To overcome the difficulty in learning with a few samples, we propose a deep transfer contrastive learning approach known as Spatial-Spectral Deep Transfer Learning Network (SSDTLN) for HSI classification. The proposed approach utilizes a spectral augmentation process to increase the diversity of pairs of samples with diverse characteristics. In addition, a spatial feature extraction component is incorporated to leverage complementary information by supporting the spatial component with pretrained representations from the source domain and the spectral residual component, focusing on precise spectral feature extraction. Moreover, contrastive learning can increase the discriminative ability of the representations with the goal of improving the ability to deal with tasks concerning high similarities between classes and large variances within classes with few samples.

Methodology: The proposed approach involved a spatial-spectral few-shot deep transfer learning approach for the classification of hyperspectral images, which was broken down into three broad phases: spatial feature extraction via a CNN-based spatial network pre-trained on the ImageNet dataset; spectral feature extraction via a residual convolutional network coupled with spectral enhancement techniques such as random spectral shift and Gaussian noise injection;

spatial-spectral feature fusion via convolutional layers; and a spatial attention module (SAM). Hyperspectral image patches were then created and arranged in a C-way K-shot episode, where the labeled images comprised the support set, while the remaining comprised the query set. A mapping layer was applied for projecting the high-dimensional hyperspectral image data into a three-channel format, compatible with the spatial network. Finally, the classification was conducted via a nearest-neighbor classifier, while the optimization of the approach was conducted via a joint loss function that combined cross-entropy loss and a supervised contrastive loss.

Main Results: The quantitative and qualitative analysis of the proposed and other state-of-the-art techniques like SVM, SSRN, DFSL, S3Net, HTLN, DCLN, and CPPM are validated on four benchmark hyperspectral image (HSI) datasets known as Pavia University (PU) dataset, Salinas (SA) dataset, Indian Pines (IP) dataset, and LongKou (LK) dataset. Experimental results indicate that the proposed model outperforms other techniques in terms of classification accuracy, boundary identification, and generalization ability, using minimal samples, while also having acceptable complexity.

Conclusions: This work successfully addresses its goal of designing a strong few-shot classification framework for hyperspectral images (HSIs) leveraging deep transfer learning and contrastive learning methods. The proposed model effectively boosts feature learning by leveraging the strengths of ImageNet-pretrained spatial networks, spectral residual learning, and spectral data augmentation methods. Additionally, the Spatial Attention Module (SAM) helps better capture the context at multiple scales, especially in boundary regions. Finally, the hybrid loss function further promotes inter-class distinction and decreases intra-class variations. Experiments conducted on four popular HSI datasets validate the superior accuracy and generalization performance of the proposed approach in few-shot learning scenarios. Even though the approach is effective, there are some issues related to computational complexity and adaptability for low-resolution and cross-sensor learning that show promising research directions.

Practical Application: The findings of this study can be effectively used in real-world applications of hyperspectral image analysis when there is a lack of annotated data, including precision farming, environmental studies, urban planning, mineral exploration, and disaster response. The developed few-shot learning framework can be effectively used for land cover mapping with only a few training samples, thereby reducing the cost of manual annotation of data. The framework's noise-robust properties make it useful for real-world applications, including UAV- or satellite-based remote sensing systems.

Keywords: Contrastive Learning, Few-Shot Learning, Hyperspectral Image Classification, Spatial–Spectral Features, Transfer Learning

Introduction

The data acquired by airborne and spaceborne hyperspectral imagery (HSI) consists of a large amount of closely distributed narrow bands, allowing for the collection of precise spectrographic signatures and spatial data. Due to their high data capacity, great interest has been raised for HSI in many areas over the past few years, especially precision agriculture, meteorology, environmental surveillance, and mineral exploration. Although great achievements have been made, HSI classification remains a challenging task, especially when limited training data are available, thus triggering intense interest for few-shot HSI classification [1].

One of the key difficulties in hyperspectral research is related to the issue of acquiring high-quality labeled training examples. The simple fact that each pixel in a hyperspectral image provides a strong response to numerous spectral bands generates a huge amount of data. This requires a lot of high-quality labeled training examples. However, labeling in the field of hyperspectral images is very expensive and time-consuming. This may require domain knowledge, and in several cases, even analysis of a scene in a distant or impossible-to-reach region.

While some advancements have been made in the classification of few-shot hyperspectral images, the existing systems still suffer from several limitations. The use of data augmentation can bring about noisy data, while the meta-learning framework relies on episodic training, which is complex. Also, transfer learning algorithms are limited due to differences in the domains between natural and hyperspectral images. Besides, most of the existing systems rely solely on the spatial features or the spectral features without considering the advantages of combining them. These challenges lead to the need for an SSDTLN system. Deep learning (DL) has made substantial contributions to HSI classification tasks by providing the capability to self-learn hierarchical spatial-spectral representations. Techniques like deep belief networks, convolutional neural networks (CNNs), and stacked autoencoders have achieved appreciable results even for difficult few-shot settings. These include light-weight 3D-CNN designs that handle direct HSI cubes and spatial-spectral residual networks (SSRN) that employ techniques like residual learning to avoid vanishing gradient issues and transfer complex information depicted in HSI images. The major drawback of DL techniques is their need for large amounts of labeled training samples and their tendency to degrade severely when the training samples are limited to cause overfitting [2].

The few-shot hyperspectral image classification (HSI) has gained widespread research interest in the last decade due to the challenge of achieving good classification accuracy using few labeled samples. Existing solutions can be classified as either data augmentation methods or deep feature learning approaches. The former attempts to diversify the training sample pool. Specifically, they increased the sample size by randomly obscuring pixels within different rectangular spatial blocks and incorporated GAN-generated samples into the true data in order to help train the CNN [3]. Though both methods could effectively increase the sample size, they are likely to add some noise to the training set and cannot well maintain the abundant spatial-spectral features of the hyperspectral images. On the contrary, the latter methods attempt to extract more discriminative features from limited samples. For example, the adaptive graph convolution is used to capture multi-scale spatial information, while the BiLSTM learns the cross-scale spatial-spectral dependency. Likewise, the proposal is a Siamese network based on depth-wise separable convolutions to explore the relationship between feature samples. Despite their excellent performance, these deep models usually involve many trainable parameters, which make optimization hard and are prone to overfitting in few training samples [4].

Various paradigms of learning, which have shown good promise for addressing the problem, include meta-learning, active learning, self-supervised learning, transfer learning, and contrastive learning. All of these paradigms use different approaches to help the classifier better generalize to unseen data due to restricted training data. The goal of meta-learning is to facilitate fast adaptation through the optimization of initialization and learning processes and thus is well suited for few-shot scenarios. Self-supervised learning uses the inherent structure of data to create pseudo-labels and therefore requires little annotation effort. Transfer learning allows transferring the knowledge learned in a data-rich domain to the task with limited examples. Active learning selects the most informative examples for labeling and thus makes the most of limited annotations [5]. However, in spite of these promising results obtained by existing few-shot and transfer learning-based methods for HSIC, several issues still exist. Namely, many algorithms use data augmentation strategies that are usually based on GAN, and this causes generation of noise in the form of samples which do not correspond to the intrinsic characteristics of hyperspectral images. Moreover, deep learning models like CNNs, GCNs, and Siamese networks have a huge amount of parameters, and therefore, they are highly likely to suffer from overfitting if there is a small

number of labeled samples. Also, the problem of domain gap remains critical in existing transfer learning-based algorithms because the knowledge transfer occurs directly from natural images whose feature distributions are different from the feature distributions of hyperspectral images [6]. Besides that, existing few-shot learning methods consider spectral or spatial information separately without learning complementary representations thereof. In order to solve these problems, this research proposes the Spatial–Spectral Deep Transfer Learning Network (SSDTLN) for few-shot HSIC. SSDTLN combines transfer learning and joint extraction of spatial and spectral features in order to reduce the domain gap and to benefit from the complementary nature of hyperspectral images at the same time. Besides that, the integration of SAM and SpeM helps to achieve higher classification accuracy using few-shot HSIC due to learning of more robust and discriminative feature representations.

The major contributions of the work are as follows:

1. In order to perform few-shot hyperspectral image classification, a novel Spatial–Spectral Deep Transfer Learning Network (SSDTLN) is presented, which helps transfer knowledge effectively while simultaneously learning discriminative features in both spatial and spectral domains.
2. In the proposed SSDTLN, an auxiliary technique for learning discriminative features is provided by combining Spatial Attention Module (SAM) and Spectral Residual Module (SpeM), which allows the model to learn discriminative regions and spectral features at the same time, which helps to improve the performance of the proposed method compared to existing techniques.
3. An effective domain adaptation method is designed to bridge the gap between source and target domains and alleviate overfitting due to a small number of labeled data.
4. The performance of the proposed SSDTLN is extensively tested using benchmarking hyperspectral images, and the experimental results confirm the superiority of the proposed approach compared to various other existing techniques such as few-shot [12, 15], transfer learning [36], and deep learning algorithms [20, 22, 24] in terms of OA, AA, and Kappa.

Methodology

Description of Proposed Model

The proposed SSDTLN model is shown in Figure 1 for the classification of hyperspectral images. The proposed approach involved a spatial-spectral few-shot deep transfer learning approach, which was broken down into three broad phases: spatial feature extraction via a CNN-based spatial network pre-trained on the ImageNet dataset, spectral feature extraction

via a residual convolutional network coupled with spectral enhancement techniques such as random spectral shift and Gaussian noise injection, and spatial-spectral feature fusion via convolutional layers and a Spatial Attention Module (SAM). The proposed model operates in two major stages involving: (i) few-shot pre-training based on data samples from both the source and target domains and subsequently (ii) evaluation of remaining data samples in the target domain. Before initiating training, there is learning of transferable features based upon classes of images belonging to the source domain. Despite ImageNet offering considerable spatial information, there could be an incomplete resemblance of hyperspectral characteristics through features learned therefrom. The training parameters learned in the source domain are thus used for forming a spatial and spectral feature extraction module for the target domain.

In the testing stage, both the labeled and the unlabeled samples from the target HSI are mapped through the spatial-spectral extractor and the mapping-layer feature extractor, and the embedding features are obtained. Then, a nearest-neighbor classifier is trained on the labeled samples and used to predict the label of the unlabeled samples. The spatial-spectral feature extractor helps in getting compact clustering in the embedding space for samples belonging to the same class and pushing samples from other classes away from each other in the embedding space, thus facilitating the classification of samples in the embedding space based on similarity in the NN classifier. Then, the classification maps are obtained for the corresponding unlabeled samples to test the performance of the proposed method. The details of different execution phases for the proposed SSDTLN are explained as follows.

1. Few-Shot for Pre-Training Phase

The training of a feature extraction module that can effectively work in a few-shot learning environment is at the heart of the presented approach. The pre-training process of the spatial-spectral feature extraction module, shown in the top part of Figure 1, consists of two stages. The initial stage includes learning spatial features from the ImageNet dataset, and for the second stage, the learning process is fine-tuned on the hyperspectral dataset to focus more on the domain-specific spatial-spectral information.

2. Pre-training on the source domain

This architecture offers the basic structure for scaling the model to the target domain. It is highly effective for fine-tuning the model on HSI data without undermining the robustness and capacity of the pre-trained model. The final layers are therefore retrained, and the transfer learning component on spatial features is able to efficiently tap into the multi-scale

hierarchical features transferred from the source domain. It is able to capture the significant feature extraction in the lower levels and focus on the domain-specific hyperspectral features.

3. Pre-Training Phase on the Target Domain

The learned transfer-based spatial feature extractor from the former stage is incorporated into the developed spatial-spectral feature extraction module. Usually, after the leveling is done, the hyperspectral image is divided into patches around every pixel, and each patch is assigned the name X_k . In order to increase the diversity of the samples and cause the model to focus on the spectral aspects extensively, the spectral enrichment technique is used. In this technique, the random spectral shift operation is applied by picking a set of spectral bands for which half are shifted by five pixels horizontally and the other half are shifted by five pixels vertically. Gaussian noise is also considered for the introduction of interferences from the acquisition system that occur in the actual world. The augmented sample $\tilde{x}_k$ is obtained as

$$\tilde{x}_k = f(x_k) + \gamma n, \quad (1)$$

where spectral augmentation function $f(x)$, γ is set to 0.01 and defines the level of the introduced noise, where $n \sim N(0, 1)$ is the Gaussian noise with zero mean and standard deviation equal to one.

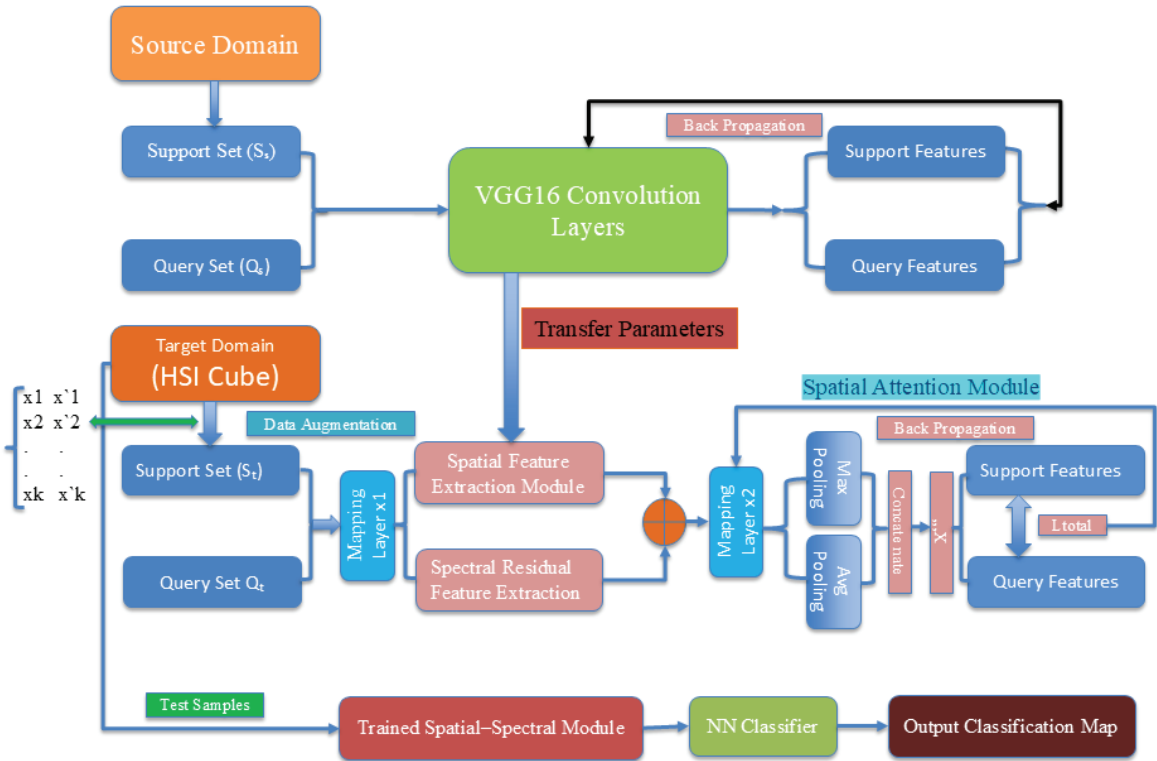


Figure 1 Architecture of the Proposed Model

In few-shot learning, just a limited number of labeled samples, the “support set” are provided to describe a huge pool of unidentified data, the “query set.” To adapt the model to this learning approach, data are separated into a support set and a “query set Q.” Each training data randomly takes C-classes, giving a C-way K-shot job. For each chosen class, K labeled instances compose the support set. support set “S” and query set “Q” represented as,

$$S = \{(\tilde{x}_i, y_i)\}_{i=1}^{C \times K},$$

While N unlabeled samples per class from the query set.

$$Q = \{(\tilde{x}_j, y_j)\}_{j=1}^{C \times N}.$$

However, the direct application of models that have been pre-trained on RGB image datasets to hyperspectral images faces difficulties owing to the disparity of the spectra. Therefore, a spectral mapping layer is proposed to map the hyperspectral image cube into

three-channel representations that can be handled by the pretrained backbone network. Given a hyperspectral image cube $\tilde{\mathbf{x}} \in \mathbb{R}^{h \times w \times b}$, where h is the height of the image, and w refers to the width of the image, while b refers to the total number of spectral bands. the hyperspectral image cube is projected into the three-channel representations through the learnable spectral mapping matrix is denoted by $\mathbf{K} \in \mathbb{R}^{b \times 3}$, where the matrix $\mathbf{K}$ maps the hyperspectral data of b -bands into three channels compatible with pretrained CNNs.

$$\mathbf{x}' = \tilde{\mathbf{x}} \times \mathbf{K}. \quad (2)$$

The settings of the mapping layer are tuned using backpropagation utilizing the existing training data.

The spatial feature extraction module uses a transfer learning backbone trained on ImageNet for accurate extraction of two-dimensional spatial information. In order to address the continuous and high-dimensional spectral information of hyperspectral images (HSIs), a specially designed spectral residual feature extraction network [7] is used in the proposed model. As shown in Figure 2.

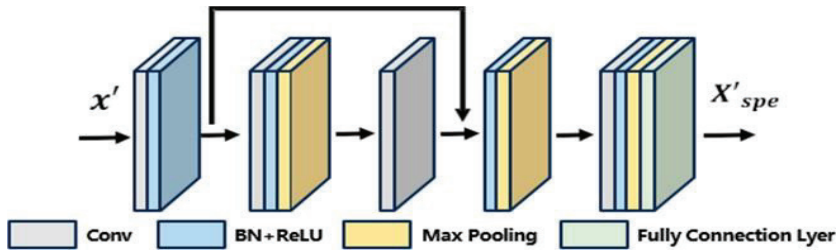


Figure 2 Framework of the Spectral Residual Feature Module

It can be seen that the Spatial- Spectral network consists of all convolutional layers, spectral residual layers, and a fully connected (FC) layer. The first convolutional layer is responsible for transforming an input spectral vector $\mathbf{x}' \in \mathbb{R}^{B \times 1}$ into a series of spectral feature vectors, which takes part in residual learning [7]. The learned feature vectors are then passed through the spectral residual block. To retain the dimensionality of the input features, padding is introduced in all convolutional layers, thus maintaining an identical input and output dimension. After completion of all residual operations, an extra convolutional layer is introduced to inject extra information into the spectral representation. Each convolutional step is accompanied by batch normalization and ReLU activation functions,

thus helping in stabilizing training and speeding up convergence. Finally, a fully connected layer consisting of 100 neurons is introduced to extract X_{spe}' features.

To combine spatial and spectral information learnt from separate training stages, a uniform spatial- spectral feature extraction module is developed. The spatial feature vector X_{spa}' and the spectral feature vector X_{spe}' are combined along the channel dimension to create a composite image and the 2-D convolutional layer is then utilized to combine these complementary features and reduce variation coming from different data sources. Because spatial traits in the target area may change greatly, the fused feature map X'' is then sent via a spatial attention module (SAM). To successfully record multi-scale spatial context under few-shot settings, SAM blends two concurrent pooling operations: max-pooling highlights locally prominent structures, and average-pooling saves global contextual information. The outputs from both branches are combined and fed into a 2D convolutional layer with a 3×3 kernel, lowering the channel dimension to one-fourth of its original size for speed. A ReLU activation adds non-linearity, and a dropout layer is utilized to reduce overfitting given the restricted data supply. Finally, a completely linked layer yields the final embedded representation X'' . This multi-stage fusion method boosts the model's capacity to incorporate different and complementary spatial-spectral information, boosting flexibility in few-shot categorization circumstances [8].

To further address class imbalance and the existence of difficult-to-discriminate data, a combined loss function is applied to improve the network. This method uses the stability of cross-entropy (CE) loss while including supervised contrastive loss as a data-driven regularizer, hence improving the discriminative powers of the learnt embeddings.

The classification problem is optimized via the cross-entropy (CE) criterion initially, where the CE measures how far the predicted probabilities deviate from ground-truth labels. The CE loss function is calculated for the query data set, after the model was learned on the support data set and here, L_{ce} denotes the cross-entropy loss, defined as

$$L_{ce} = \mathbb{E}_{S,Q} \left[- \sum_{(x,y) \in Q} 1 \log p_{\phi}(y = k | x) \right], \quad (4)$$

where S and Q denote the support sets and query sets, respectively, and $x \in Q$ is a query sample with true label k . Prior studies suggest that CE loss alone is restricted by its susceptibility to label noise and inadequate margin expansion, which decreases its generalization potential.

To mitigate these difficulties and increase feature transferability, supervised contrastive learning is added into the optimization. For a batch index $i \in I = \{1, 2, \dots, N\}$, the supervised contrastive loss is defined as

$$L_{\text{con}} = \sum_{i \in I} \frac{1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(\text{sim}_g(Z_i, Z_p)/\tau)}{\sum_{t \in O(i)} \exp(\text{sim}_g(Z_i, Z_t)/\tau)}, \quad (5)$$

where Z_i , Z_p , and Z_t denotes the feature embedding of the i -th sample generated after going through the feature extractor, and sim_g denotes the cosine similarity between the two normalized projection vectors. The set $O(i)$ consists of all the sample indexes in the current mini-batch except for the anchor one, and $P(i) = \{p \in O(i) | y_p = y_i\}$ denotes the positive samples with the same class label as the anchor. The temperature scaling coefficient τ is set to 0.5 for all the experiments in the paper.

Through aggregating information of features in the same category and distinguishing features in different categories, the supervised contrastive loss is capable of directly alleviating this serious problem in HSI classification, in which high intra-class variability and high inter-class similarity coexist.

Results and Discussion

Experimental Data and Analysis:

The efficiency of the proposed model was tested using four well-known HSI standard datasets: (a) Pavia University (PU) dataset, (b) Salinas (SA) dataset, (c) Indian Pines (IP) dataset, and (d) LongKou (LK) dataset. This approach has been developed using the Python programming language along with the use of open-source deep learning libraries TensorFlow and Keras. TensorFlow is utilized for enabling efficient computation through the use of data-flow graphs, whereas Keras provides a high-level abstraction API for the creation, training, and evaluation of models. In order to study the implementation and deployment efficiency of the proposed SSDTLN framework, computational complexity analysis and runtime efficiency analysis have been performed. The proposed framework requires around 2.84 GFLOPs of computation per inference along with the requirement of around 1.32 GB GPU memory per inference. Furthermore, the framework achieves around 215 image inferences per second on an NVIDIA RTX 3060 GPU, which makes it suitable for efficient hyperspectral image classification.

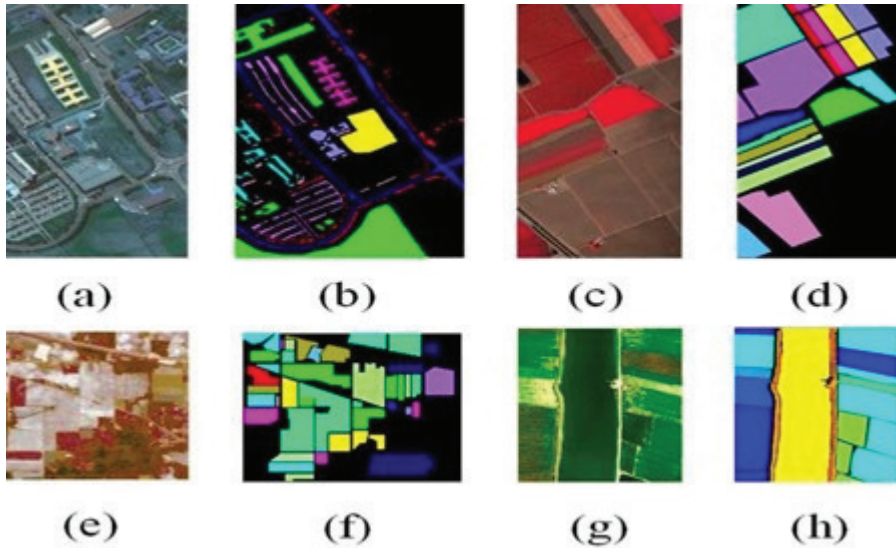


Figure 3 The False-Color Pictures and Ground-Truth Maps. (a, b): Pavia University Dataset. (c, d): Salinas Dataset. (e, f): Indian Pines Dataset. (g, h): LongKou Dataset

These datasets span multiple regional areas and include various land-cover categories, providing a thorough basis for testing the model's generalization skills [9]. False-color (FC) pictures and the linked ground-truth (GT) maps for all datasets are presented in Figure 3, while detailed dataset specs are explained in Tables 1-4.

Table 1 Data Samples of Pavia University Dataset

Pavia University Dataset			
S. No.	Lands - Cover Types	Training	Test
(i)	Asphalt	V	6,626
(ii)	Bitumen	V	1,325
(iii)	Gravel	V	2,094
(iv)	Trees	V	3,059
(v)	Metal -sheets	V	1,340
(vi)	Bis-soil	V	5,024
(vii)	Bricks	V	3,677
(viii)	Medows	V	18,644
(ix)	Shadows	V	942
-	Total	XLV	42,731

Table 2 Data Samples of the Salinas University Dataset

Salinas D			
	Lands-Cover Types	Training	Test
(i)	Broccoli -1	V	2,004
(ii)	Broccoli -2	V	3,721
(iii)	Soil	V	6,198
(iv)	Corn	V	3,273
(v)	Fallow -smooth	V	2,673
(vi)	Stubble	V	3,954
(vii)	Celery	V	3,574
(viii)	Grapes -untrained	V	
(ix)	Fallow	V	1,971
(x)	Fallow -plow	V	1,389
(xi)	Lettuce_4_wk	V	1,063
(xii)	Lettuce_5_wk	V	1,922
(xiii)	Lettuce_6_wk	V	911
(xiv)	Lettuce_7_wk	V	1,065
(xv)	Vinyard -untrained	V	7,263
(xvi)	Vinyard -vertical	V	1,802
-	Total	LXXX	54,049

Table 3 Data Samples of the Indian Pines Dataset

Indian Pines (IP) Dataset			
S. No.	Lands-Cover Types	Training	Test
(i)	Alfalfa	V	41
(ii)	Corn-notill	V	1,432
(iii)	Corn-mintill	V	825
(iv)	Grass - pasture mowed	V	23
(v)	Grass-pasture	V	478
(vi)	Grass-trees	V	725
(vii)	Corn	V	232

Table 3 Data Samples of the Indian Pines Dataset (continued)

Indian Pines (IP) Dataset			
S. No.	Lands-Cover Types	Training	Test
(viii)	Hay-windrowed	V	473
(ix)	Woods	V	1,260
(x)	Soybean -notill	V	967
(xi)	Soybean -mintill	V	2,450
(xii)	Soybean -clean	V	588
(xiii)	Wheat	V	200
(xiv)	Oats	V	15
(xv)	Buildings -grass	V	381
(xvi)	Stone -steel	V	88
-	Total	LXXX	10,169

Table 4 Data Samples of the LongKou Dataset

LongKou Dataset			
S. No.	Lands-Cover Types	Training	Test
1	Corn	V	34,506
2	Cotton	V	8369
3	Sesame	V	3026
4	Broad -leaf soy -bean	V	63,207
5	Narrow -leaf soybean	V	4146
6	Rice	V	11,849
7	Water	V	67,051
8	Roads and Houses	V	7119
9	Mixed weed	V	5224
-	Total	XLV	204,497

Table 5 Overall Accuracy (OA) of Different Methods for the Four Datasets

S. No.	Datasets Type	Scheme	OA%
1	Pavia University	SpaM (pre -trained)	82.25 ± 1.47
		SpaM (Scratch)	78.14 ± 2.03
		SpeM	81.73 ± 1.21
		SpaM (Pre -trained) + SpeM	85.49 ± 1.52
2	Salinas	SpaM (Pre-trained)	89.45 ± 0.81
		SpaM (Scratch)	86.73 ± 1.34
		SpeM	87.93 ± 0.97
		SpaM (Pre -trained) + SpeM	91.11 ± 2.31
3	Indian Pines	SpaM (pre -trained)	74.61 ± 2.08
		SpaM (Scratch)	71.09 ± 2.41
		SpeM	73.89 ± 1.76
		SpaM (Pre -trained) + SpeM	75.46 ± 2.15
4	LongKou	SpaM (Pre -trained)	91.55 ± 1.78
		SpaM (Scratch)	89.23 ± 1.65
		SpeM	90.84 ± 1.05
		SpaM (pre -trained) + SpeM	94.11 ± 1.31

Indian Pines (IP), Pavia University (PU), Salinas (SA), and LongKou (LK) datasets are the hyperspectral image (HSI) datasets used to test the performance of the developed model. IP is a hyperspectral dataset taken by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor in northwestern India. It has 200 spectral bands, a spatial size of 145x145 pixels, and 16 classes of land cover. Pavia University dataset is also a hyperspectral dataset taken by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor in northern Italy. It has 103 spectral bands, a spatial size of 610x340 pixels, and nine classes of land cover. Salinas dataset is a hyperspectral dataset taken by the AVIRIS sensor, and it has 204 spectral bands, a spatial size of 512x217 pixels, and 16 classes of land cover. It also includes 54,129 labeled pixels, excluding background.

This WHU-Hi LongKou data is obtained on the 17th of July 2018 for the LongKou Town in the province of Hubei in China using the Headwall Nano-Hyperspectral camera that is mounted on an unmanned aerial vehicle (UAV). This data has a spatial resolution of about 0.463 m with an image size of 550 x 400 pixels. This data can be described with the help of 270 spectral bands with wavelengths ranging from 400 nm to 1000 nm.

In order to conduct a fair and complete evaluation of the algorithm, three different training protocols were conducted according to the experimental goals. (1) Few-shot episodic learning: SSDTLN was trained using an episodic learning paradigm where there was one labeled support sample per class and 19 query samples per episode in order to conduct the evaluation in a situation with an extremely low number of labels. (2) Practical few-shot evaluation: The setup involving 1-shot ($K = 1$, $N = 19$) is used for meta-training and meta-testing for the few-shot learning setup. The 5-shot setup is used to test the robustness of the proposed architecture under a slightly increased number of labels per class. (3) Supervised learning: In order to benchmark the results against other fully supervised deep learning architectures, 200 labeled samples per class were randomly picked for training, and the rest were left for testing purposes. To lessen the effect of chance on group selection, all tests were repeated 10 times, and the best overall accuracy. Model training was achieved using the Adam optimizer with a learning rate of 0.001. Classification ability was studied using three standard metrics: overall accuracy (OA), average accuracy (AA), and the kappa coefficient (κ).

Table 6 Influence of the Spatial Attention Module (SAM) on Overall Accuracy (OA%)

Datasets	With_SAM	Without_SAM	OA (%)
Pavia University	85.49 $\pm$ 1.44	82.12 $\pm$ 0.3	+3.37
Salinas	91.11 $\pm$ 2.31	88.98 $\pm$ 0.1	+2.13
Indian Pines	75.46 $\pm$ 2.15	71.83 $\pm$ 0.5	+3.63
LongKou	94.11 $\pm$ 1.31	91.75 $\pm$ 0.2	+2.35

1. Comparison of Different Models

In order to determine the effectiveness of the transfer-based spatial feature extraction module (SpaM) and the spectral residual feature extraction module (SpeM) in the few-shot hyperspectral image (HSI) classification, four experimental settings were considered on every one of the four datasets. The overall accuracies (OAs) achieved by the different settings are shown in Table 5.

Within the baseline scenario, known as the SpaM-Scratch experiment, training was performed on HSI solely without using the ImageNet framework; this was used as the reference point for determining the effect of transfer learning [10]. When compared to the reference point, the relative accuracy improvement of the pre-trained SpaM on ImageNet was shown to be 3.3% to 4.11%, which clearly shows the pre-trained spatial knowledge was quite beneficial.

When a comparative analysis of SpaM and SpeM is conducted individually, it has been noticed that SpaM performs better, implying that spatial features obtained from a pre-trained model are superior to spectral features captured by a small dataset. However, the best overall accuracy is reached by the combination of the pre-trained model and SpeM across all datasets [11]. These results validate the assumption that by incorporating spatial and spectral features, optimal results can be achieved through few-shot learning of HSI classification.

2. Spatial Attention Module (SAM) Contribution

To evaluate the role of the Spatial Attention Module (SAM), the new approach was trained with and without the SAM in the standard five-shot setting across the five categories. The results for the remaining components and parameters remained unchanged for fair comparison reasons, as asserted in [12]. Table 6 shows the combined results of 10 different experiments for each case.

As revealed from the findings presented in Table 6, the addition of SAM achieves greater accuracy improvements for the four traditional datasets, with the largest accuracy improvement of 3.6% for the Indian Pines imaging dataset, characterized by very complex spatial patterns.

This can be explained by the dual-path structure of the SAM network, which merges the results of the average and max pooling layers. In fact, the average pooling pathway focuses on the disjoint patches in the farmland areas, while the max pathway maintains the spectral continuity in the ‘crop’ areas. It should also be noted that the contribution made by the SAM method in variance reduction is limited. However, it is interesting to see that the method can increase the accuracy of boundary areas. For instance, the F1 measure for the “Gravel” category is improved by 9.2% using the proposed method.

(1) Loss Function

The loss function also acts as an essential component in improving the results of the classification of the proposed method. The proposed loss function combines cross-entropy loss functions with supervised contrastive learning in order to attain domain-invariant and highly discriminative feature spaces that can efficiently discriminate among various classes [13]. In this combination, both methods can effectively benefit from each other since cross-entropy has domain-invariant properties that can efficiently discriminate among various classes, while contrastive learning can efficiently discriminate among clean features. To test the effectiveness of the proposed loss function strategy, the other components in the model remain fixed with only changes in the loss function [14]. From the results shown in Table 7, there is a distinct trend in the performance of the model for the four datasets. The model trained only with a cross-entropy loss function has a lower accuracy rate. The supervised contrastive learning helps to achieve a significant boost in performance. The combined loss function that aims to optimize both cross-entropy and supervised contrastive objectives performs the best. These results validate the effectiveness of the proposed combined loss strategy and its applicability for few-shot HSI classification tasks [15].

(2) Comparison With State-of-the-Art Algorithms:

In this experiment, the proposed method was compared against several state-of-the-art approaches, including the different classification supervised machine learning classifier SVM [14], the 3-D convolution-based model spatial-spectral residual network SSRN [22], the random occlusion data augmentation technique RODA [24], the heterogeneous transfer learning network HTLN [36], the deep few-shot learning methods DFSL [12] and S3Net [11], the deep cross-domain few-shot network DCLN [15], and the contrastive learning enhanced model CPPM [20]. These similarities were made to completely examine the effectiveness of our method.

Table 7 Ablation Research on Loss Function

S. No.	Datasets	Loss Function	OA%
1	Pavia University	LCE	84.75 $\pm$ 2.45
		Lcon	85.24 $\pm$ 1.09
		LCE + Lcon	85.49 $\pm$ 1.53
2	Salinas	LCE	90.75 $\pm$ 1.47
		Lcon	90.14 $\pm$ 1.75
		LCE + Lcon	91.11 $\pm$ 2.31
3	Indian Pines	LCE	73.81 $\pm$ 0.95
		Lcon	74.36 $\pm$ 1.31
		LCE + Lcon	75.46 $\pm$ 2.15
4	LongKou	LCE	89.98 $\pm$ 1.71
		Lcon	91.41 $\pm$ 1.60
		LCE + Lcon	94.11 $\pm$ 1.31

For the Support Vector Machines algorithm, a rigorous evaluation procedure was followed. Five labeled examples of each class were randomly picked as training examples. For all methods both the proposed approaches and the baselines validation and training were conducted under the same conditions.

(3) The Quantitative Evaluations

The results of the classification ability shown in Tables 8-11 indicate a thorough comparison between the effectiveness of the classifiers on the four benchmark datasets. A close analysis reveals that the poorest performance is produced by SVM, indicating the weakness of shallow models in efficiently learning high-dimensional spectral data. Although the data augmentation method RODA outperforms SVM, its results are extremely oscillatory owing to the noisy nature of the artificially generated data, a property contrasting with strict spectral consistency in hyperspectral data [16]. The 3D CNN-based SSRN model works well under a scenario with copious amounts of labeled training data, but a considerable drop in accuracy occurs under the five-shot learning scenario, indicating the dependence of deep CNNs on a considerably larger training dataset.

Few-shot-oriented models like DFSL, S3Net, and CPPM rely on meta-learning and self-supervised learning strategies to improve embedding features and reach competitive results on both the PU and SA datasets. By contrast, their performance drops sharply on the IP dataset, which means that learning complex spatial-spectral correlations from various agricultural environments still faces difficulties [17]. The heterogeneous transfer learning network HTLN obtains competitive results on various datasets and shows that transfer learning is a promising solution for handling data scarcity.

Table 8 Results of Classification for Pavia University Dataset Using Multiple Models

S. No	SVM	SSRN	RODA	DFSL	HTLN	S3Net	CPPM	DCLN	Proposed Model
1	71.43	75.98	36.52	99.27	88.99	82.76	80.02	74.86	84.31
2	83.68	64.96	52.81	99.46	73.36	74.51	84.38	60.77	90.24
3	41.72	54.97	83.68	49.05	63.88	71.08	64.23	40.89	81.46
4	66.91	92.83	85.87	48.92	95.68	84.83	93.17	93.34	91.44
5	99.98	99.51	99.09	98.91	98.87	98.93	98.62	99.79	99.58
6	28.82	62.28	48.68	78.97	56.78	65.17	70.17	85.05	71.44
7	40.88	54.68	95.62	72.56	94.16	94.14	80.27	91.76	87.50
8	55.85	46.83	21.34	85.23	80.36	73.45	68.87	62.91	75.33
9	80.76	98.99	95.92	76.69	99.16	95.90	98.50	99.97	99.28
OA	60.15 $\pm$ 2.28	68.09 $\pm$ 1.21	55.40 $\pm$ 1.04	79.59 $\pm$ 2.67	77.21 $\pm$ 0.99	78.25 $\pm$ 1.21	82.38 $\pm$ 2.94	81.94 $\pm$ 1.32	86.56 $\pm$ 1.66
AA	62.99 $\pm$ 3.75	73.51 $\pm$ 2.88	67.95 $\pm$ 0.13	79.57 $\pm$ 1.75	82.43 $\pm$ 0.55	81.29 $\pm$ 0.67	82.92 $\pm$ 1.96	69.82 $\pm$ 2.41	85.60 $\pm$ 1.22
K	49.88 $\pm$ 2.61	58.99 $\pm$ 0.57	45.89 $\pm$ 2.24	79.51 $\pm$ 2.11	73.09 $\pm$ 1.78	72.22 $\pm$ 2.46	75.36 $\pm$ 3.68	76.20 $\pm$ 1.86	82.18 $\pm$ 2.02

Table 9 Results of Classification for Salinas Dataset Using Multiple Models

S.No	SVM	SSRN	RODA	DFSL	HTLN	S3Net	CPPM	DCLN	Proposed Model
1	88.84	95.27	82.36	98.60	99.10	98.81	98.11	98.37	99.34
2	98.46	97.47	86.82	99.70	99.70	95.44	98.51	98.31	97.69
3	83.43	91.66	76.36	99.93	96.30	95.84	97.44	92.70	98.89
4	98.21	97.83	74.84	98.10	84.17	99.70	95.64	98.57	98.44
5	94.03	96.33	88.31	72.29	88.32	96.42	97.37	93.30	98.48
6	96.23	98.15	89.12	98.50	98.87	98.38	99.08	97.63	96.86
7	97.68	99.92	86.47	98.32	97.48	99.69	98.04	98.23	99.55
8	65.58	77.40	82.73	81.14	92.38	82.86	64.38	74.38	77.12
9	98.51	95.21	94.43	99.10	98.35	98.81	99.98	98.85	99.71
10	75.53	84.46	83.94	75.73	91.15	93.47	93.91	92.53	92.89
11	77.29	91.73	63.84	94.40	96.16	98.43	97.41	97.36	98.30
12	87.51	98.38	78.84	82.22	98.31	89.77	84.18	99.70	98.24
13	92.30	21.18	63.05	99.10	96.19	98.86	98.04	97.85	97.20
14	78.23	65.19	64.73	33.73	82.86	92.47	97.98	98.87	99.88

Table 9 Results of Classification for Salinas Dataset Using Multiple Models (continued)

S.No	SVM	SSRN	RODA	DFSL	HTLN	S3Net	CPPM	DCLN	Proposed Model
15	47.37	60.24	70.87	56.82	61.29	72.29	89.14	86.86	82.23
16	90.23	65.86	71.57	96.52	98.58	97.20	85.51	90.36	96.19
OA	80.40	84.10	81.40	83.77	89.09	89.95	89.36	91.47	92.20
	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$ 2.22
	2.19	1.41	1.29	2.23	1.37	1.64	0.74	1.20	
AA	84.53	84.47	77.74	87.51	91.48	95.19	92.31	93.63	94.63
									$\pm$ 0.61
	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	
	2.18	2.68	2.55	0.77	1.78	2.19	0.55	0.32	
K	78.16	82.09	79.32	82.08	87.81	90.92	88.13	90.41	91.01
	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$ 1.42
	3.41	2.71	1.83	1.48	2.73	0.44	0.91	0.56	

Across all datasets, the suggested technique consistently demonstrates good performance. On the PU dataset, our model exhibits competitive class-level accuracy, notably in classes C1, C2, and C8. For the Salinas dataset, practically all classes are categorized properly, which may be ascribed to the spectral residual network's capacity to capture narrow-band spectral signatures and the SAM's multi-scale spatial fusion capabilities [18]. On the Indian Pines dataset, our approach attains the maximum accuracy in seven out of the available classifications, demonstrating its resilience in complicated agriculture contexts. Similarly, performance increases for all models on the LongKou dataset because to its increased spatial resolution; yet, our technique continues to outperform others, attaining the best OA, AA, and K across all four datasets [19].

These findings together show that the proposed architecture offers a robust and effective solution for few-shot HSI classification, enabling better generalization and increased discrimination across various situations.

Table 10 Results of Classification for Indian Pines Dataset Using Multiple Models

S.No	SVM	SSRN	RODA	DFSL	HTLN	S3Net	CPPM	DCLN	Proposed Model
1	29.18	76.45	93.64	81.53	37.55	98.77	99.02	81.42	90.53
2	38.75	19.42	54.35	28.23	67.15	48.37	39.43	58.11	71.67
3	38.04	55.76	52.51	78.03	66.62	57.81	55.71	61.45	75.11
4	25.07	31.20	36.51	88.23	48.03	90.62	75.71	90.26	83.21
5	43.92	10.83	86.35	84.21	87.02	77.96	84.32	64.23	91.49
6	79.91	84.43	87.15	89.79	98.54	96.46	92.25	71.16	93.64
7	24.71	11.37	99.01	99.01	20.04	99.01	99.01	96.43	96.20
8	91.35	77.40	62.30	99.04	46.05	84.31	86.91	92.41	96.74
9	13.93	13.22	92.32	99.04	24.87	99.01	99.01	92.24	86.10
10	37.64	61.38	16.32	57.08	63.70	58.20	69.60	65.72	77.87
11	57.26	67.96	31.58	43.74	75.97	57.12	73.42	61.15	60.61
12	21.21	27.15	10.17	36.71	57.75	61.42	44.93	66.06	81.86
13	84.21	98.88	94.01	95.65	75.95	98.04	98.72	85.26	98.44
14	84.41	86.28	69.16	75.69	93.30	84.17	90.33	85.29	78.18
15	28.95	14.62	41.67	59.14	50.45	77.26	78.72	93.84	10.45
16	84.21	86.98	99.01	99.10	58.15	98.44	96.02	83.58	98.64
OA	48.99 $\pm$ 3.71	59.02 $\pm$ 1.14	49.73 $\pm$ 2.91	60.71 $\pm$ 2.96	70.07 $\pm$ 3.02	68.53 $\pm$ 3.18	69.81 $\pm$ 1.31	71.26 $\pm$ 2.51	76.41 $\pm$ 2.23
AA	47.78 $\pm$ 2.90	51.57 $\pm$ 2.27	65.09 $\pm$ 2.49	74.87 $\pm$ 3.26	61.67 $\pm$ 2.67	81.42 $\pm$ 2.39	81.18 $\pm$ 1.60	76.87 $\pm$ 1.71	85.06 $\pm$ 1.31
K	41.81 $\pm$ 2.55	50.08 $\pm$ 1.68	41.34 $\pm$ 4.72	54.27 $\pm$ 4.23	67.22 $\pm$ 2.52	61.58 $\pm$ 3.61	67.84 $\pm$ 1.37	69.38 $\pm$ 1.14	71.08 $\pm$ 0.96

Table 11 Results of Classification for LongKou Dataset Using Multiple Models

S.No	SVM	SSRN	RODA	DFSL	HTLN	S3Net	CPPM	DCLN	Proposed Model
1	88.01	91.49	82.31	92.09	78.63	96.27	88.28	89.27	93.31
2	63.60	79.13	77.43	79.78	91.92	86.29	77.08	96.09	71.21
3	58.39	79.84	81.34	92.67	88.12	98.57	96.91	76.28	98.62
4	94.56	93.64	93.14	93.36	76.61	95.94	78.48	96.87	89.16
5	98.81	98.77	99.76	98.86	89.12	98.25	81.11	99.99	93.25
6	59.83	70.78	74.97	92.28	95.11	93.24	90.03	92.98	84.63
7	79.76	98.93	92.43	81.67	90.26	95.93	99.68	76.36	87.64

Table 11 Results of Classification for LongKou Dataset Using Multiple Models

S.No	SVM	SSRN	RODA	DFSL	HTLN	S3Net	CPPM	DCLN	Proposed Model
8	67.11	85.53	78.25	91.74	82.95	96.51	52.26	78.89	82.72
9	98.27	87.05	98.92	98.89	57.41	94.12	52.04	98.26	88.68
OA	72.71	82.99	82.29	87.38	86.71	92.44	84.98	91.28	93.15
	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$ 1.47
	4.43	2.54	1.82	2.47	2.53	1.14	2.21	1.09	
AA	79.58	86.93	87.62	90.38	84.25	94.88	78.44	90.42	94.94
	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$ 2.44
	3.66	1.21	3.13	1.97	2.43	1.32	3.09	1.78	
K	63.43	79.33	76.71	83.67	82.64	90.84	83.12	88.74	93.89
	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$ 1.89
	4.12	1.87	2.54	2.87	3.89	2.78	4.03	2.02	

(4) The Qualitative Evaluations

The visual categorization maps created by different approaches for the four datasets are displayed in Figures 4-7. A comparative evaluation demonstrates that few-shot learning methods such as DFSL, S3Net, DCLN, and the suggested method generate classification maps that are considerably cleaner and more spatially consistent than those generated by previous techniques. In contrast, SVM findings reveal severe speckling and unclear borders across many areas, highlighting the difficulties of shallow models in handling tiny training data. Although RODA yields considerably cleaner maps than SVM, the results remain inferior [20]. This implies that although data augmentation alleviates the small-sample problem to some extent, the limited quantity of labeled data still leaves the model vulnerable to fluctuations in the sampling procedure. SSRN also delivers inferior visual results, possibly owing to its dependence on greater training volumes to capture important spatial-spectral patterns [21].

Although few-shot learning models are not fully free from mistakes, their performance is often adequate given the constrained amount of labeled examples. For the Pavia University dataset, our proposed SSDTLN technique properly classifies almost all pixels in the “Gravel,” “Metal-sheets,” “Bare-soil,” and “Bitumen” classes (C3, C5, C6, and C7). The borders between these classes are significantly clearer than those generated by SVM and SSRN, illustrating the usefulness of the transferred spatial characteristics in modeling heterogeneous surface materials

[22]. CPPM, however, misclassifies numerous samples of the “Bare-soil” class (C6), illustrating the limits of simply contrastive learning when dealing with spectrally similar categories. In contrast, SSDTLN corrects these mistakes via the use of spectral residual representations [23-24].

In the Salinas dataset, substantial uncertainty is evident in S3Net’s predictions for the “Lettuce” classes (C11-C14), but SSDTLN maintains an accuracy above of 97.20%. This improvement is related to the spectral enhancement module, where noise injection enhances resilience against spectral distortions and band shifts.

On the Indian Pines dataset, despite having just 200 labeled samples for the “Wheat” class (C13), SSDTLN attains 98.44% accuracy [25]. This result highlights the usefulness of the hybrid loss function in avoiding overfitting to minority classes. Collectively, the visual findings across all datasets indicate the great discriminative power and resilience of the proposed technique, particularly under severe few-shot situations [26].

Overall, compared with prior few-shot learning techniques, the proposed model provides classification maps that closely match the underlying ground-truth pictures across all four datasets [27]. These gains emerge from the model’s increased capacity to extract discriminative spatial-spectral characteristics even when only a limited count of labeled examples is available. This efficacy is primarily owing to the integration of a spatial feature extractor ported from large-scale pretrained models, along with a customized spectral residual network that boosts spectral representation learning [28].

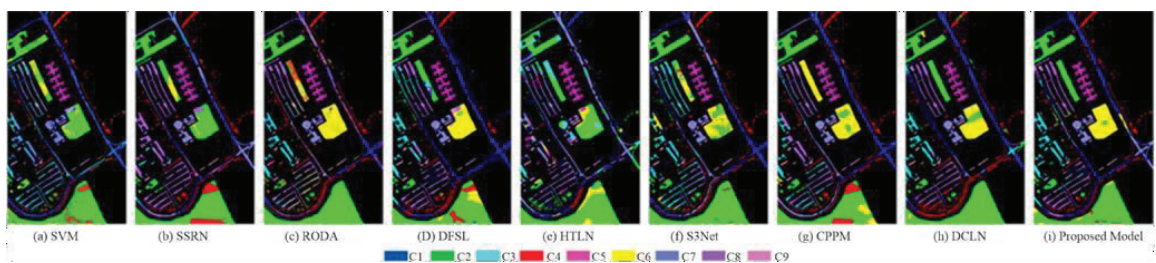


Figure 4 Comparison Between Classification Results on the Pavia University Dataset Using the Following Techniques: (a) SVM, (b) SSRN, (c) RODA, (d) DFSL, (e) HTLN, (f) S3Net, (g) CPPM, (h) DCLN, and (i) Proposed Model

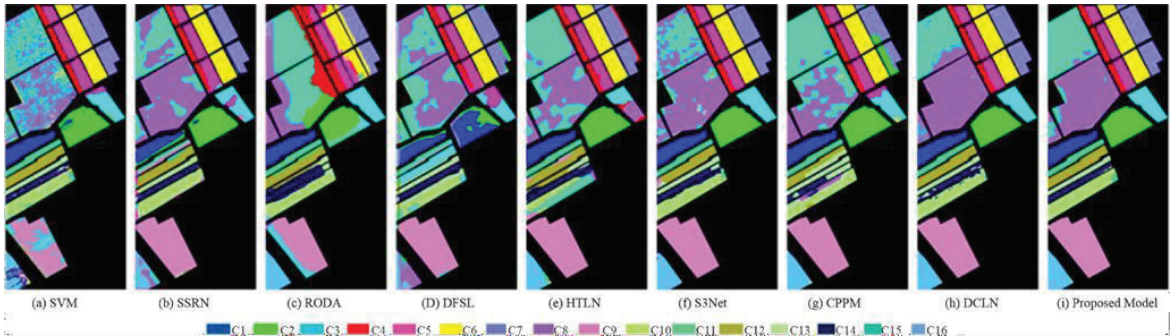


Figure 5 Comparison Between Classification Results on the Salinas Dataset Using the Following Techniques: (a) SVM, (b) SSRN, (c) RODA, (d) DFSL, (e) HTLN, (f) S3Net, (g) CPPM, (h) DCLN, and (i) Proposed Model

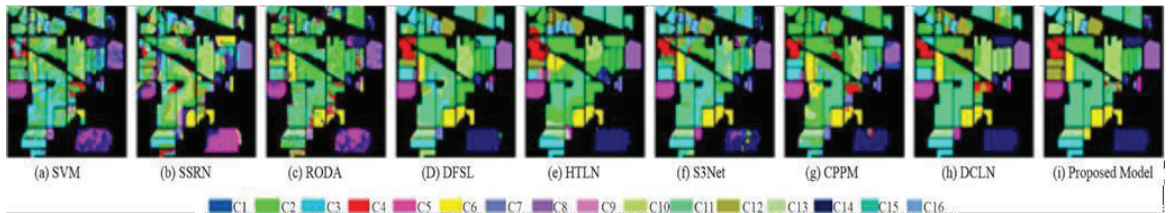


Figure 6 Comparison Between Classification Results on the Indian Pines Dataset Using the Following Techniques: (a) SVM, (b) SSRN, (c) RODA, (d) DFSL, (e) HTLN, (f) S3Net, (g) CPPM, (h) DCLN, and (i) Proposed Model

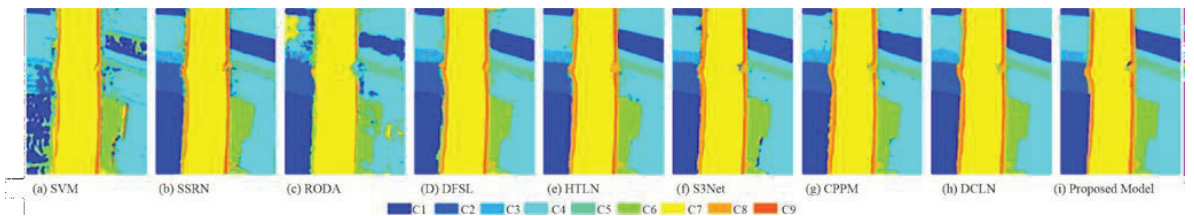


Figure 7 Comparison Between Classification Results on the LongKou Dataset Using the Following Techniques: (a) SVM, (b) SSRN, (c) RODA, (d) DFSL, (e) HTLN, (f) S3Net, (g) CPPM, (h) DCLN, and (i) Proposed Model

(5) Effect of the Labeled Samples:

To further investigate the model's generalization capabilities under very low supervision, further trials were conducted utilizing 3-7 labeled samples per class. As seen in Figures 8-11, the overall accuracies (OAs) of all techniques progressively rise when more labeled samples become available, which accords with the well-acknowledged

idea that extra training data typically improves performance [29]. Across this range, the suggested technique consistently obtains reasonably high OAs, demonstrating excellent adaptability to various sample sizes. Moreover, even in the most constrained situation with just three labeled examples per class, the model retains competitive performance, showing its potential usefulness in ultra-low-sample learning settings [30-31].

Table 12 Computational Times (Seconds) & Parameters (M) of Various Approaches for Each Dataset

Datasets	Metric	SVM	SSRN	RODA	DFSL	HTLN	S3Net	CPPM	DCLN	Proposed Model
PU	Train	4.26	721.43	644.23	808.58	1472.3	1214.34	1837.30	1447.82	1547.32
	Test	0.58	3.88	3.37	4.56	8.41	6.22	7.97	7.31	8.61
	Params	-	0.36	0.21	0.04	0.69	0.06	0.84	0.64	0.48
SA	Train	7.31	1019.4	873.87	1199.3	1614.1	1362.33	1619.04	1576.95	1421.61
	Test	0.81	6.24	5.28	7.31	8.12	7.38	8.33	7.92	8.95
	Params	-	0.41	0.21	0.02	0.69	0.08	0.86	0.48	0.48
IP	Train	3.98	473.21	422.25	917.43	1422.6	973.46	1573.97	1370.32	1533.96
	Test	0.48	3.68	3.11	5.42	7.28	5.89	7.93	8.14	7.35
	Params	-	0.36	0.21	0.02	0.69	0.08	0.86	0.64	0.48
LK	Train	9.43	50.32	1312.3	1618.9	1947.3	1798.21	2374.22	1989.29	1813.99
	Test	2.36	13.96	20.22	19.38	22.33	22.15	21.32	24.91	20.23
	Params	-	0.37	0.21	0.02	0.67	0.08	0.86	0.66	0.48

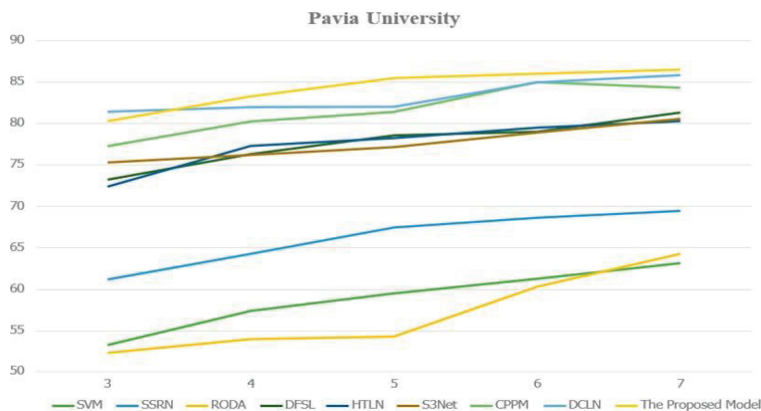


Figure 8 Classification Accuracies (%) Achieved by Different Numbers of Labeled Samples from the PU Dataset

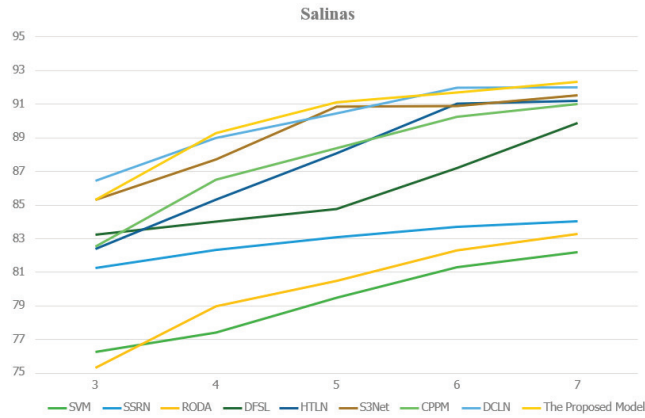


Figure 9 Classification Accuracies (%) Obtained by Changing Numbers of Labeled Samples from SA Dataset

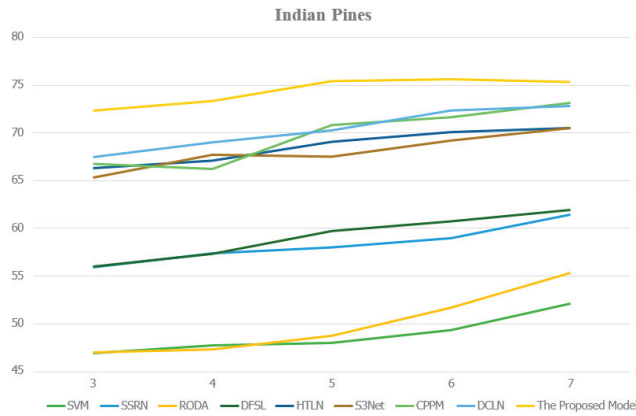


Figure 10 Classification Accuracies (%) Obtained by Different Numbers of Labeled Samples from IP Dataset

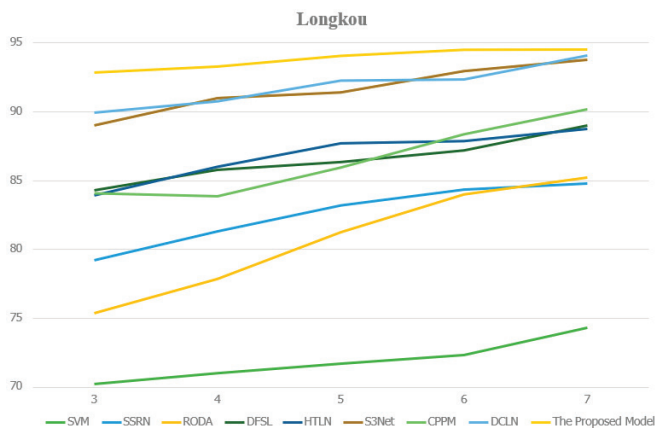


Figure 11 Classification Accuracies (%) Achieved by Different Amounts of Labeled Samples from LK Dataset

(6) Analysis of Computational Complexity

The experimental results shown in Table 12 above highlight the variances between the approaches considered for training time, testing time, and the number of trainable parameters. Not surprisingly, the fastest training time, which ranges from seconds, is taken by traditional approaches using a small number of parameters [32]. On the other hand, approaches based on deep learning have significantly higher training and testing costs owing to the complex network architecture and back-propagation [33].

In the group of deep learning models, SSRN and RODA have a moderate training time range of 400-1300 seconds for the four datasets with efficient feature selection and sample augmentation techniques. Their number of parameters remains well under 0.4 million. However, DFSL and S3Net minimize the number of parameters to 0.03-0.09 million with testing times under 7 seconds, showcasing the effectiveness of using efficient model architecture design [34].

Area-based methods include HTLN, CPPM, and DCLN, whose designs involve further modules for inter-domain matching and/or mining of positive pairs, thus increasing their number of parameters to 0.63-0.85 million. This increase is reflected in significantly increased training times, which take about 1400 seconds, as well as testing times that approach 25 seconds, indicating serious computational constraints [35].

The technique presented in this study provides better trade-off benefits because it needs fewer trainable parameters when compared to the HTLN, DCLN, and CPPM networks and still keeps the training and testing times within acceptable limits. This is because the computational efficiency and classification accuracy are the key benefits of the presented design technique according to Wei et al. [36].

Conclusions

This paper proposes a novel spatial-spectral deep transfer learning network (SSDTLN) for HSI classification. The proposed model integrates a spatial-spectral few-shot deep transfer learning approach, which includes spatial feature extraction via a CNN-based spatial network pre-trained on the ImageNet dataset, spectral feature extraction via a residual convolutional network coupled with spectral enhancement techniques such as random spectral shift and Gaussian noise injection, and spatial-spectral feature fusion via convolutional layers and a Spatial Attention Module (SAM). Here, a spectral data augmentation strategy, stacking random spectral translation and noise addition, strengthens the robustness against real-world spectrum

perturbations and may overcome challenges posed by a few training samples. Additionally, the network architecture employs transfer learning for the spatial-spectral feature extraction module. In this work, a pre-trained spatial network (pre-trained using ImageNet) is combined with a spectral residual network to effectively improve feature extraction. The Spatial Attention Module (SAM), with dual-path pooling (max pooling and average pooling), helps multi-scale environmental information acquisition, ensuring higher accuracy for boundary-sensitive areas in classification tasks. The combination of several loss functions further promotes higher inter-class discriminability and lower intra-class variability, overcoming challenges of high inter-class similarity and large intra-class diversity, which lie in HSI data. Experimental results based on four publicly available HSI datasets confirm the effective and competitive performance of the proposed approach. Experimental results also indicate that the proposed approach gives better performance compared to other state-of-the-art techniques such as SVM [12], SSRN [22], RODA [24], DFSL [12], HTLN [36], S3Net [11], CPPM [20], and DCLN [15]. There are some challenges, especially those posed by higher computational complexity and adaptability for tough environments, such as low-resolution cases or cross-sensor HSI analysis, which may be further studied in the future.

References

1. He, X., Chen, Y. and Ghamisi, P. 2019. Heterogeneous transfer learning for hyperspectral image classification based on convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 58, 3246-3263.
2. Zhao, X., Liang, Y., Guo, A.J. and Zhu, F. 2020. Classification of small-scale hyperspectral images with multi-source deep transfer learning. *Remote Sensing Letters*, 11, 303-312.
3. Zhang, H., Li, Y., Jiang, Y., Wang, P., Shen, Q. and Shen, C. 2019. Hyperspectral classification based on lightweight 3-D-CNN with transfer learning. *IEEE Transactions on Geoscience and Remote Sensing*, 57, 5813-5828.
4. Qu, Y., Baghbaderani, R.K. and Qi, H. 2019. Few-Shot hyperspectral image classification through multitask transfer learning. *10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Amsterdam, The Netherlands, 1-5.
5. Zhou, H., Wang, X., Xia, K., Ma, Y. and Yuan, G. 2024. Transfer learning-based hyperspectral image classification using residual dense connection networks. *Sensors*, 24, 2664.

6. Yang, B., Hu, S., Guo, Q. and Hong, D. 2022. Multisource domain transfer learning based on spectral projections for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 3730-3739.
7. Lin, J., Zhao, L., Li, S., Ward, R. and Wang, Z.J. 2018. Active-learning-incorporated deep transfer learning for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11, 4048-4062.
8. Xue, Z., Zhou, Y. and Du, P. 2022. S3Net: Spectral-spatial Siamese network for few-shot hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-19.
9. Masarczyk, W., Głomb, P., Grabowski, B. and Ostaszewski, M. 2020. Effective training of deep convolutional neural networks for hyperspectral image classification through artificial labeling. *Remote Sensing*, 12, 2653.
10. Masarczyk, W., Głomb, P., Grabowski, B. and Ostaszewski, M. 2019. Effective transfer learning for hyperspectral image classification with deep convolutional neural networks. *arXiv Preprint arXiv*, 1909.05507.
11. Xie, F., Gao, Q., Jin, C. and Zhao, F. 2021. Hyperspectral image classification based on superpixel pooling convolutional neural network with transfer learning. *Remote Sensing*, 13, 930.
12. Cheng, Y., Zhang, W., Wang, H. and Wang, X. 2023. Causal meta-transfer learning for cross-domain few-shot hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1-14.
13. Li, W., Wang, Z., Wang, Y., Wu, J., Wang, J., Jia, Y. and Gui, G. 2020. Classification of high-spatial- resolution remote sensing scenes method using transfer learning and deep convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 1986-1995.
14. Ahmad, M., Ghous, U., Hong, D., Khan, A.M., Yao, J., Wang, S. and Chanussot, J. 2022. A disjoint samples-based 3D-CNN with active transfer learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-16.
15. Sun, J., Shen, X. and Sun, Q. 2022. Hyperspectral image few-shot classification network based on the earth mover's distance. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-14.

16. Cao, M., Zhang, X., Cheng, J., Zhao, G., Li, W. and Dong, X. 2024. Spatial-spectral-semantic cross-domain few-shot learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-15.
17. Ullah, F., Ullah, I., Khan, K., Khan, S. and Amin, F. 2025. Advances in deep neural network-based hyperspectral image classification and feature learning with limited samples: a survey. *Applied Intelligence*, 55, 370.
18. Xi, B., Zhang, Y., Li, J., Huang, Y., Li, Y., Li, Z. and Chanussot, J. 2025. Transductive few-shot learning with enhanced spectral-spatial embedding for hyperspectral image classification. *IEEE Transactions on Image Processing*, 34, 854-868.
19. Deng, F., Pu, S., Chen, X., Shi, Y., Yuan, T. and Pu, S. 2018. Hyperspectral image classification with capsule network using limited training samples. *Sensors*, 18, 3153.
20. Singh, M.K. and Kumar, B. 2023. Fine tuning the pre-trained convolutional neural network models for hyperspectral image classification using transfer learning. *Computer Vision and Robotics: Proceedings of CVR 2022*, Springer Nature Singapore, Singapore, 271-283.
21. Feng, S., Zhang, H., Xi, B., Zhao, C., Li, Y. and Chanussot, J. 2024. Cross-domain few-shot learning based on decoupled knowledge distillation for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-14.
22. Guo, B., Liu, T., Zhang, X. and Gu, Y. 2024. Few-shot multispectral-hyperspectral image collaborative classification with feature distribution enhancement and subdomain alignment. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-17.
23. Gong, Z., Duan, L., Xiao, F. and Tong, L. 2025. Hyperspectral image classification via cross-domain few-shot learning with multi-level style and semantic enhancement. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 1-18.
24. Li, Z., Xue, Z., Xu, Q., Zhang, L., Zhu, T. and Zhang, M. 2023. SPFormer: Self-pooling transformer for few-shot hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1- 19.
25. Zhang, C., Yue, J. and Qin, Q. 2020. Global prototypical network for few-shot hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 4748- 4759.

26. Rise, B., Uney, M. and Huang, X. 2024. Two-stage transfer learning for fusion and classification of airborne hyperspectral imagery. *ICASSP 2024–2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, South Korea, 6555–6559.
27. Xia, M., Yuan, G., Yang, L., Xia, K., Ren, Y., Shi, Z. and Zhou, H. 2023. Few-Shot hyperspectral image classification based on convolutional residuals and SAM Siamese networks. *Electronics*, 12, 3415.
28. He, X., Chen, Y. and Lin, Z. 2021. Spatial-spectral transformer for hyperspectral image classification. *Remote Sensing*, 13, 498.
29. He, X. and Chen, Y. 2020. Transferring CNN ensemble for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 18, 876-880.
30. Wang, T., Liu, H. and Li, J. 2022. Spectral–spatial classification of few shot hyperspectral image with deep 3-D convolutional random Fourier features network. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-18.
31. Tang, H., Yang, X., Tang, D., Dong, Y., Zhang, L. and Xie, W. 2024. Tensor-based few-shot learning for cross-domain hyperspectral image classification. *Remote Sensing*, 16, 4149.
32. Jia, S., Jiang, S., Lin, Z., Xu, M., Sun, W., Huang, Q., Zhu, J. and Jia, X. 2021. A semisupervised Siamese network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–17.
33. Li, X., Ma, H., Guo, S., Zhang, D. and Li, Z. 2025. Graph knowledge distillation-guided few-shot learning for hyperspectral image classification: X. *The Journal of Supercomputing*, 81, 1181.
34. Gao, K., Yu, A., You, X., Qiu, C., Liu, B. and Guo, W. 2023. Learning general-purpose representations for cross-domain hyperspectral images classification with small samples. *Remote Sensing*, 15, 1080.
35. Mdrafi, R., Du, Q., Gurbuz, A.C., Tang, B., Ma, L. and Younan, N.H. 2020. Attention-based domain adaptation using residual network for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 6424-6433.
36. Wei, X., Li, W., Zhang, M. and Li, Q. 2019. Medical hyperspectral image classification based on end-to-end fusion deep neural network. *IEEE Transactions on Instrumentation and Measurement*, 68, 4481-4492.