Fully Convolutional Spectral–Spatial Fusion Network Integrating Supervised Contrastive Learning for Hyperspectral Image Classification

Yifan Shen, Ling Shi, Ji Zhao, *Member, IEEE*, Yuting Dong, *Member, IEEE*, and Lizhe Wang, *Fellow, IEEE*

***Abstract*—Hyperspectral image classification using deep learning techniques has received great attention in recent years, considering the powerful spatial feature mining ability of deep learning. Fully convolutional network is an effective deep learning architecture that exploits spatial contextual information through a hierarchical convolutional structure. However, it often ignores the relationships between samples of the same category and different categories within the global context. Therefore, a fully convolutional spectral–spatial fusion network based on supervised contrastive learning (FCSCL) is proposed for hyperspectral image classification to enhance the separability between different categories and class aggregation among the same category. In the FCSCL framework, the spectral–spatial fusion classification network is developed to capture subtle spectral variations and spatial patterns by adaptively fusing the features extracted by the spectral branch and spatial branch. To improve intraclass compactness and interclass separability, the SCL module is integrated into the FCSCL framework. The positive and negative sample pairs are constructed by the designed hard example pairs sampling strategy. These constructed sample pairs are used to guide the network to learn more discriminative feature representations that pixels of the same category are closer to each other and pixels of different categories are pushed further apart in the feature space. The experiments using three public hyperspectral datasets verify the effectiveness of the FCSCL algorithm, and the FCSCL method achieves better classification performance.**

***Index Terms*—Contrastive learning, fully convolutional network (FCN), hyperspectral image, image classification.**

I. Introduction

HYPERSPECTRAL imagery (HSI) has rich spectral information and often has hundreds of nearly continuous spectral bands, which can be used to identify different materials with small spectral differences. Therefore, hyperspectral images play an important role in precision agriculture, mineral exploration, environmental monitoring, and other fields [1], [2], [3]. HSI classification is a fundamental task for various applications by labeling each pixel with a semantic category [4].

The early methods of HSI classification were mainly based on the spectral information of hyperspectral images to identify the category of materials, such as classification based on artificial neural network [5], extreme learning machines [6], multilayer perceptron [7], and support vector machine (SVM) [8]. These methods achieve good performance on HSI classification. However, due to the spectral variability in HSI images, the same land cover category may have significantly different spectra and different land cover categories may have similar spectra, so it is difficult to effectively and accurately represent the characteristics of land covers using only single spectral information.

To alleviate the confusion of land cover categories caused by spectral variability and improve classification accuracy, as another aspect of HSI image, the spatial information is fully exploited, and it gradually transitions from manually designed spatial features based on expert knowledge to automatic spatial feature representation based on deep learning. The handcrafted spatial features design some artificial rules to mine the spatial pattern of pixels and neighboring pixels. For example, the morphological profile feature uses the dilation and erosion operations to extract spatial structure features [9], [10]. In addition, the random field model can also mine the spatial information of HSI images by modeling the relationship between pixels based on the graph structure [11]. The model mainly includes two types of conditional random fields (CRF) [12], [13], [14] and Markov random fields [15], [16]. These methods significantly improve classification accuracy compared to the method that only uses spectral information. However, these methods rely more on handcrafted features, and the generalization ability of the models is limited, which makes them difficult to achieve satisfactory classification performance.

To automatically mine spatial structure information of the image, deep learning classification methods have received great attention [17], [18]. Based on the different feature information used in the classification process, the deep learning models for HSI classification can be mainly divided into the following three categories. One kind of deep learning algorithm for HSI classification is spectral information feature extraction network. The model takes 1-D spectral vectors as input, and learns the correlation and complex patterns between these spectral features to achieve classification, such as deep belief networks [19],

Manuscript received 8 August 2023; revised 13 September 2023; accepted 19 September 2023. Date of publication 27 September 2023; date of current version 13 October 2023. This work was supported by the National Natural Science Foundation of China under Grant 42171384. *(Corresponding author: Ji Zhao.)*

Yifan Shen, Ling Shi, Ji Zhao, and Lizhe Wang are with the School of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: shenyf@cug.edu.cn; lingshi0265@gmail.com; zhaoji@cug.edu.cn; lizhe.wang@foxmail.com).

Yuting Dong is with School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China (e-mail: dongyt@cug.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3319587

one-dimension convolutional neural networks (1-D CNNs) [20], and stacked autoencoders [21]. The described network type sacrifices a certain level of spatial information and experiences exponential growth in network parameters with the addition of fully connected layers. The second kind of deep learning algorithm for HSI classification is spatial feature-based extraction network. These models take spatial patches as input to focus on spatial information. Since hyperspectral images often have a large number of spectral bands, these method often combined with dimensionality reduction algorithms, such as principal component analysis (PCA) [22], [23]. The spatial patch-based mechanism divides the hyperspectral image into multiple small patches for processing, which may lose important information due to inappropriate segmentation, thus affecting classification accuracy. To effectively utilize the spectral and spatial information of hyperspectral images simultaneously, another method is the spatial–spectral fusion extraction network. The input of this type of model is a spatial patch-based approach that combines spectral and spatial information. By this way, deep learning models can fully utilize all spectral information and neighborhood spatial information to improve classification accuracy. This type of model includes 2D-CNN [24], [25] for extracting spectral information, and 3D-CNN [26], [27], [28] for performing convolution operations in both spectral and spatial dimensions, and further derived 3D-GAN [25], [29], spectral–spatial 3-D fully convolutional network [30], spectral–spatial residual network [31], oriented self-similarity network [32], and other high-performance network architectures. In addition, some network structures construct two branches to extract features from spatial and spectral information separately, and then integrate these features into fully connected layers to achieve spectral–spatial information fusion. These network architectures include the two-stream deep architecture [33], spectral-spatial unified networks [34], and dual-branch multiscale spectral attention network [35].

Although the HSI classification task using deep learning is equivalent to the semantic segmentation of computer vision, there are some differences in hyperspectral image classification, such as the sparsity of training data and lots of unlabeled images. To enable deep models, such as CNN to effectively train hyperspectral data, a lot of previous work is patch-based learning strategy, which will lead to high computational complexity [36]. To solve this problem, some researchers have started using the whole image as input to achieve end-to-end spectral–spatial fusion classification. For example, the spectral–spatial fully convolutional network (SSFCN) [37] was proposed and achieved the aggregation of global long-range contextual information by introducing dense CRF (DenseCRF). Zheng et al. [36] proposed a fast patch-free global learning approach for HSI classification. Its speed and accuracy are greatly improved compared to deep learning models with spatial patch-fetching mechanism. But it only considers the relationship between pixels in a local region and ignored the relationships between samples of the same category and different categories within the global context.

In this study, to effectively utilize the global relationship between samples, a fully convolutional spectral–spatial fusion classification network based on supervised contrastive learning (FCSCL) is proposed. Contrastive learning aims to learn similar/dissimilar distribution properties by comparing pairs of samples [38], [39], [40], and achieves a tighter feature space, which means that the features of similar samples are brought closer together and the features of different categories of sample are pushed apart. Contrastive learning has also recently been a popular research topic for HSI classification. For example, Zhao et al. [41] proposed an innovative contrastive learning approach that aims to extract spectral–spatial features from readily available unlabeled samples and subsequently enhance classification performance by fine-tuning the parameters of the model using labeled samples. Hou et al. [42] adopt a self-supervised training strategy, where differentiating positive and negative sample pairs are used as pretext tasks during training. Subsequently, the pretrained model is employed in downstream classification tasks. To ensure that features with similar semantic meaning are brought closer together, a spectral–spatial semantic feature learning network is proposed by Xu et al. [43] to align the projected spectral spatial features through contrastive loss. However, contrastive learning in the above methods is often mainly used in patch-based hyperspectral classification frameworks. The anchor is the original image patch, its augmented version serves as the positive sample, and all other image patches in the batch function as negative samples [44]. This will not take advantage of the FCN. Therefore, this study integrates an SCL module into a fully convolutional spectral–spatial fusion classification network to take the global labeled samples into consideration, which helps to enhance the distinctiveness between different categories and the cohesion within the same category. The main contributions are summarized as follows.

1) An FCSCL framework is proposed for hyperspectral image classification. The FCSCL framework integrates FCN and contrastive learning by sharing the feature extraction network and designing an SCL head and a semantic segmentation head. Therefore, the advantages of the two technologies can be combined, and the aggregate information of the same category and the difference information of different categories can be considered in the full convolutional network through contrastive learning.

2) To effectively utilize the spectral and spatial information of hyperspectral images, a spectral–spatial fusion classification architecture is developed. We design the architecture as a concatenated structure of two branches, one for extracting spectral features and the other for extracting spatial features. To maximize the utilization of the spectral–spatial features, an adaptive fusion approach is introduced to integrate the spectral–spatial features to improve the classification accuracy and robustness of the architecture.

3) An SCL module with an effective sample pair construction strategy is designed to better learn the similarity and dissimilarity of data. Sample pair construction strategy is a key part of contrastive learning. We select the sample pair with the lowest intra-class similarity as the positive sample pair, maximizing the difference between the positive sample pairs to improve the discriminability of the learned features. Meanwhile, new

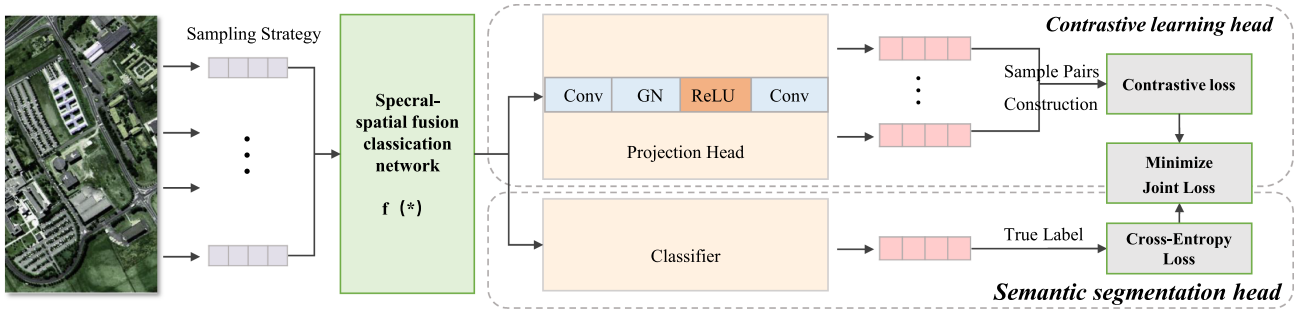


Fig. 1. Flowchart of the FCSCL framework.

negative sample pairs are constructed through pairwise combinations of positive sample pairs to increase the number of negative sample pairs, so that the model can better learn the differences of interclass.

The rest of this article is organized as follows. The FCSCL framework is described in Section II. The relevant experimental results and discussions are presented in Section III. Finally, Section IV concludes this article.

II. FCSCL for HSI Classification

In this article, to mine the relationship between different samples within the global context in hyperspectral classification, an FCSCL framework is proposed (Fig. 1). The framework mainly includes a feature extraction network, an SCL head, and a semantic segmentation head. The feature extraction network enhances class separability from spectral and spatial perspectives, and a attention mechanism is incorporated at the end of the network to achieve adaptive fusion to mine useful spectral and spatial feature information. After that, the contrastive loss is obtained by the SCL head and the cross-entropy loss is obtained by the semantic segmentation head. Joint loss consists of the above two losses, and finally constrains the feature extraction network.

A. Spectral–Spatial Fusion Classification Network

To better utilize the spectral relationships between spectral bands and the spatial relationships between pixels, we developed a spectral–spatial fusion classification network (Fig. 2), and the network consists of feature extraction modules based on spatial branch and spectral branch, and a spectral–spatial adaptive fusion module based on the attention mechanism.

The spatial feature extraction branch mainly improves the discrimination ability of different categories by mining multiscale spatial information. Since hyperspectral datasets have different spectral dimensions, we first compress the spectral dimensions with 1×1 convolution to fit different datasets. To capture the global context information, inspired by U-Net [45], the branch contains a top-down contraction path, and a bottom-up expansion path. The contraction path is used to compress the size of the input hyperspectral image layer by layer while adjusting the spectral dimension. The path consists of four repeated convolutional modules, each containing two cascaded convolutional groups. The convolution group consists of a convolution layer with 3×3 kernel, a group normalization layer (GN), and a linear activation layer. Each module of the extended path contains a

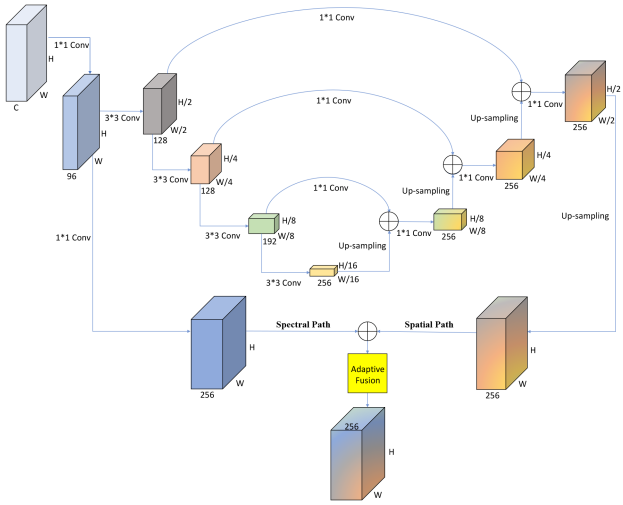


Fig. 2. Schematic diagram of spectral–spatial fusion classification network.

convolutional layer with 1×1 kernel to adjust the dimension of the corresponding layer of the contraction path to be consistent with the dimension of that layer. The higher level semantic features are spliced with more complete spatial information by upsampling, and then the number of channels is adjusted by the convolution with 1×1 kernel to utilize the global and local spatial information at the same time.

For the spectral feature extraction branch, we extract spectral features by two consecutive convolutions with 1×1 kernel acting only on the spectral dimension, where the first convolution shares weights with the first convolution in the spatial feature extraction branch. In addition, after each convolutional layer, a GN layer is added. To better fuse spectral and spatial extracted features, we design an adaptive fusion mechanism. The spectral and spatial information are first spliced in dimension. The adaptive fusion based on squeeze-and-excitation block [46] is used to assign different weights to the spectral and spatial features extracted by the spectral and spatial feature extraction branches to improve the classification performance. The specific structure of the proposed adaptive fusion module is shown in Fig. 3.

B. Supervised Contrastive Learning Module

The SCL is integrated into the FCSCL framework, and its main role is to learn the similarity and dissimilarity of data, which mainly includes projection head, construction of sample

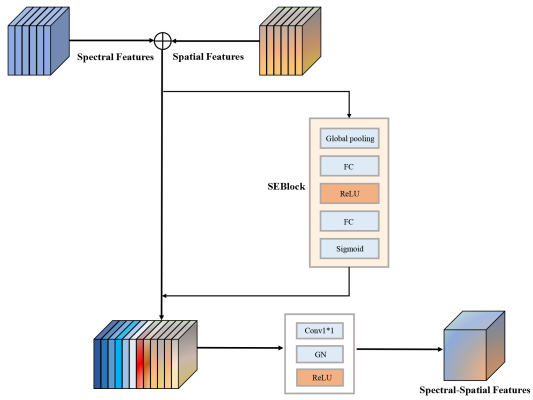


Fig. 3. Specific structure of the proposed adaptive fusion module.

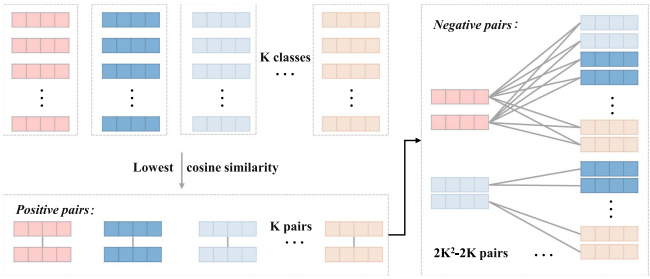


Fig. 4. Schematic diagram of sample pairs construction.

pairs, and pixel-to-pixel contrastive loss. The function of projection head is to map the high-dimensional feature representation to a more compact and discriminative low-dimensional feature space, and then the similarity and difference between data pairs can be learned through contrastive loss. By minimizing the distance between feature vectors of positive pairs and maximizing the distance between feature vectors of negative pairs, the projection head learns a representation where similar samples' feature vectors are closer, and dissimilar samples' feature vectors are more distant. Therefore, the construction of sample pairs is very critical in contrastive learning. Considering that the discriminative ability of samples is crucial for contrastive learning, a hard example pairs sampling strategy is designed, which is shown in Fig. 4. For each class (total K classes), the cosine similarity operation is performed on any two samples selected within the class. The two sample pairs with the lowest similarity are chosen as positive sample pairs for the class. These same-class sample pairs with lower similarity can motivate the model to pay closer attention to intraclass differences, leading to better discrimination of different samples. In the end, a total of $2K$ samples were selected. Except the anchor point x_i and its corresponding positive sample x_i^+, the remaining $2K-2$ samples constitute negative sample pairs with x_i. Therefore, the designed hard example pairs sampling strategy will form K positive pairs and $2\,K^2-2\,K$ negative pairs.

Based on the constructed sample pairs, the contrastive loss function can maximize the similarity of positive sample pairs and the dissimilarity of negative sample pairs to encourage the model to learn feature representations with stronger discriminative capabilities. This leads to similar samples being closer to each other in the feature space, while dissimilar samples are pushed further apart. The pixel contrastive loss function is InfoNCE which was first proposed by Oord et al. [47]. It is defined as

$$\mathcal{L}^{\text{InfoNCE}} = -E_x\left[\log\frac{\exp\left(s\left(q,h^+\right)/\tau\right)}{\sum_{x_i\in X, i\neq q}\exp\left(s\left(q,h_i\right)/\tau\right)}\right] \tag{1}$$

where $X=\{x_1,\ldots,x_N\}$ represent the sampled N samples. $q=f(x_q)$ is the feature value of query samples, where f is the feature extraction network. $h^+=f(x_i)$ is the feature vector of positive samples corresponding to x_q, and τ denotes a temperature parameter. s is the similarity measure function, which is the cosine similarity function here.

C. FCSCL Framework for HSI Classification

The FCSCL framework mainly composes of three parts: 1) spectral–spatial fusion classification network, 2) SCL module, and 3) a fully convolutional semantic segmentation module. By jointly considering both spectral and spatial cues, the spectral–spatial fusion network can capture subtle spectral variations and spatial patterns, leading to more discriminative feature representations. The SCL learns the similarity and dissimilarity between data pairs. The semantic segmentation head is responsible for transforming the high-dimensional features into semantic-level outputs. The cross entropy loss is often used as its loss function, which can compare the classification result of each pixel with the ground-truth to get a loss value. Specifically, assuming a hyperspectral image has K categories, the one-hot representation of the true label of each pixel i is $\boldsymbol{y_i}=(y_{i1},y_{i2},\ldots,y_{iK})$, and $y_{ij}=1$ when pixel i belongs to the category j. The model's prediction probability for that pixel is $\hat{\boldsymbol{y}}_{\boldsymbol{i}}=(p_{i1},p_{i2},\ldots,p_{iK})$, then the cross entropy loss of the pixel is

$$\mathcal{L}^{CE} = -E_x\left[\sum_{j=1}^{K} y_{ij}\log\left(p_{ij}\right)\right]. \tag{2}$$

It is noticed that the cross-entropy loss focuses on the construction of the decision boundary, and does not consider the structure of the feature space used for classification. SCL can accurately construct a tighter feature space, which will be more conducive to the construction of decision boundaries. Therefore, the weighted sum of cross-entropy loss and contrastive loss is used to constrain the whole network at the same time. To integrate SCL module in the FCSCL framework, a joint loss function is used, which combines the pixel cross-entropy loss from the semantic segmentation head and the contrastive loss from the SCL head. These two losses are complementary to some extent. The former helps to improve intraclass compactness and interclass separability by exploring the semantic relationship between pixel samples, and the latter expects the network to learn discriminative features for HSI classification. Therefore, the joint loss can be expressed as follows:

$$\mathcal{L}^{\text{Joint}} = \mathcal{L}^{CE} + \lambda\,\mathcal{L}^{\text{InfoNCE}} \tag{3}$$

where $\lambda>0$ is the coefficient. $\mathcal{L}^{CE}$ in (3) is the pixel cross-entropy loss shown in (2), and $\mathcal{L}^{\text{InfoNCE}}$ is the aforementioned

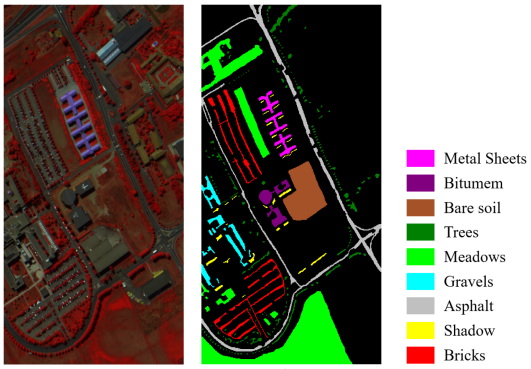


Fig. 5. Pavia University dataset. (a) Three-band false color image. (b) Ground-truth image.

contrastive loss shown in (1). This joint loss guides the optimization of the spectral–spatial fusion classification network, ensuring that the final feature representations are both semantically meaningful and discriminative for accurate HSI classification.

III. Experimental Results

We evaluate the proposed approach on three typical datasets (WHU_Hi_HongHu, WHU_Hi_HanChuan [48], [49], and Pavia University datasets) and compare it with other typical hyperspectral image classification algorithms. The overall accuracy (OA), Kappa, average accuracy (AA), and classification accuracy of each class are used to evaluate the performance of various algorithms. In the following we will describe our experiments in several sections.

A. Data Description

1) Pavia University Dataset: This image was collected by the ROSIS sensor over the Pavia University, Italy. The image contains 103 spectral bands ranging from 430–860 nm, and the spatial size is 610×340 with a spatial resolution of 1.3 m. The dataset mainly contains nine categories, including trees, asphalt, bricks, meadows, etc. Fig. 5 shows the false color image and the ground-truth image.

2) WHU_Hi_HongHu Dataset: This dataset is a UAV image collected in HongHu, Hubei Province, China in 2017. The spatial resolution of the data is 0.043 m when the UAV is flying at an altitude of 100 m. The spatial size of the dataset is 940×475, and it has 270 bands ranging from 400–1000 nm. This dataset contains a complex agricultural scene with a wide variety of crops, and different varieties of the same crop are grown in the area, including Chinese cabbage and cabbage, and small bok choy and baby cabbage. The dataset contains a total of 22 categories. Fig. 6 shows the false color image and the corresponding ground-truth image.

3) WHU_Hi_HanChuan Dataset: This dataset is also a UAV image collected in Hanchuan, Hubei Province, China in 2017. The spatial resolution of the data is 0.109 m when the UAV is flying at an altitude of 250 m. The dataset size is 1217×303 and contains 274 bands ranging from 400–1000 nm. This dataset contains a total of 16 categories. Fig. 7 shows the false color image and the ground truth.

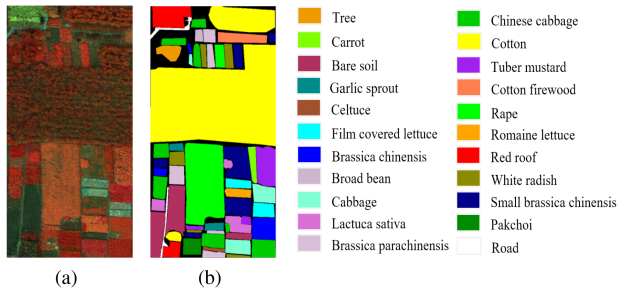


Fig. 6. WHU_Hi_HongHu dataset. (a) Three-band false color image. (b) Ground-truth image.

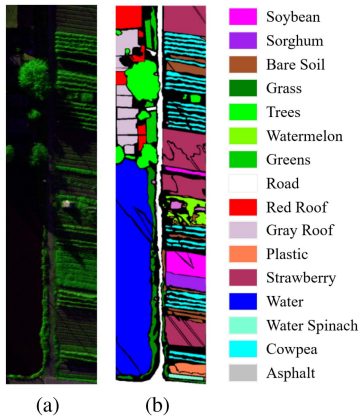


Fig. 7. WHU_Hi_HanChuan data. (a) Three-band false color image. (b) Ground-truth image.

B. Experimental Setup

In the experiments, the FCSCL method was compared with several HSI classification approaches, including traditional classification algorithm and five recently proposed deep learning classification algorithms. They are SVM [50], CVSSN [32], RESN [51], DFFN [52], SSFCN-CRF [37], and FreeNet [36]. The SVM [50] algorithm uses the radial basis function (RBF) as the kernel function for classification. The CVSSN [32] method utilizes the neighborhood adaptive fusion module of spectral information and the feature-oriented neighborhood similarity metric module to better fuse spatial information. The RESN [51] is a self-ensembling network based on teacher–student model, which takes full advantage of unlabeled spectral information through consistency regularization and a consistency filter. The DFFN [52] is a deep residual network, which incorporates multilevel features. The SSFCN-CRF [37] is a spectral–spatial full convolutional network to combine spectral and spatial features. To balance global and local information, CRF is applied as a post-processing. The FreeNet [36] is a fully convolutional end-to-end classification framework, which designs a sampling strategy to allow the network to converge better and faster.

To verify the classification performance of our algorithm using limited samples, 25 samples are selected for each category on the three experimental datasets as training samples. Tables I–III show the number of training samples and the corresponding testing samples. All experiments were performed with a single NVIDIA GTX 1080ti GPU. To ensure the training convergence of the proposed FCSCL model, the global stochastic stratified

TABLE I
CLASS INFORMATION FOR THE PAVIA UNIVERSITY DATASET

Class	Name	Training	Testing	Sum
C1	Asphalt	25	6606	6631
C2	Meadows	25	18624	18649
C3	Gravel	25	2074	2099
C4	Trees	25	3039	3064
C5	Painted metal sheets	25	1320	1345
C6	Bare Soil	25	5004	5029
C7	Bitumen	25	1305	1330
C8	Self-Blocking Bricks	25	3657	3682
C9	Shadows	25	922	947

TABLE II
CLASS INFORMATION FOR THE WHU_HI_HONGHU DATASET

Class	Name	Training	Testing	Sum
C1	Red roof	25	14016	14041
C2	Road	25	3487	3512
C3	Bare soil	25	21796	21821
C4	Cotton	25	163260	163285
C5	Cotton firewood	25	6193	6218
C6	Rape	25	44532	44557
C7	Chinese cabbage	25	24078	24103
C8	Pakchoi	25	4029	4054
C9	Cabbage	25	10794	10819
C10	Tuber mustard	25	12369	12394
C11	Brassica parachinensis	25	10990	11015
C12	Brassica chinensis	25	8929	8954
C13	Small Brassica chinensis	25	22482	22507
C14	Lactuca sativa	25	7331	7356
C15	Celtuce	25	977	1002
C16	Film covered lettuce	25	7237	7262
C17	Romaine lettuce	25	2985	3010
C18	Carrot	25	3192	3217
C19	White radish	25	8687	8712
C20	Garlic sprout	25	3461	3486
C21	Broad bean	25	1303	1328
C22	Tree	25	4015	4040

TABLE III
CLASS INFORMATION FOR THE WHU_HI_HANCHUAN DATASET

Class	Name	Training	Testing	Sum
C1	Strawberry	25	44710	44735
C2	Cowpea	25	22728	22753
C3	Soybean	25	10262	10287
C4	Sorghum	25	5328	5353
C5	Water spinach	25	1175	1200
C6	Watermelon	25	4508	4533
C7	Greens	25	5878	5903
C8	Trees	25	17953	17978
C9	Grass	25	9444	9469
C10	Red roof	25	10491	10516
C11	Gray roof	25	16886	16911
C12	Plastic	25	3654	3679
C13	Bare soil	25	9091	9116
C14	Road	25	18535	18560
C15	Bright object	25	1111	1136
C16	Water	25	75376	75401

sampling strategy proposed in FreeNet is used and the parameter α is set to 5. We set the temperature coefficient in InfoNCE to 0.3, λ set to 1. The SGD optimizer with a learning rate of 0.01 is used.

C. Experimental Results and Analysis

The classification accuracies of different algorithms on the three datasets are shown in Tables IV–VI. The highest accuracy for each category is shown in bold. The classification accuracy of the HSI classification methods based on deep learning are better than the SVM algorithm on all datasets. For the classification performance of the Pavia University dataset (Table IV), RESN achieves the best results in the second category, because the number of samples in the second category is obviously more than other categories, and the integrated network of RESN can fully extract information from unlabeled samples. The proposed algorithm achieves second and close accuracy on the second category. Overall, our proposed method achieves OA (98.57%), AA (99.03%), and Kappa (0.9811), which is the best result among all algorithms. Out of nine categories, the proposed FCSCL algorithm achieves top results in five categories, including 100% accuracy on one category. Compared with other algorithms on this dataset, the proposed FCSCL algorithm improves at least 1.99%, 1.71%, and 0.026 on the OA, AA, and Kappa evaluation metrics, respectively.

Table V reports the classification accuracy of various methods for the WHU_Hi_HongHu dataset. Since this dataset has more categories and many indistinguishable crop categories, the accuracy obtained on this dataset is lower than that obtained on the Pavia University dataset using the same algorithm. Among the deep learning-based HSI classification approaches, the FCSCL algorithm achieved the best results on OA, AA, and Kappa evaluation metrics, which are 95.47%, 95.58%, and 0.9429, respectively, significantly outperforming the patch-based deep learning networks. Compared with FreeNet and SSFCN-CRF, which are also FCN, the FCSCL algorithm can achieve better classification performances since the contrastive learning module is considered to better extract global features of images. In addition, deep learning-based HSI classification methods generally achieve high accuracy, but some deep learning models may not perform well in terms of classification accuracy in some categories, since they cannot be fully trained using limited samples, such as the accuracies of category 19 and category 20 for the SSFCN-CRF model. As reported in Table VI, similar conclusions can be obtained for the WHU_Hi_HanChuan dataset. Overall, the FCSCL method achieves comparable accuracies on all three HSI classification datasets and has improved the accuracy by an average of 2%, compared with the best result (FreeNet) in the comparison algorithm.

To qualitatively analyze the experimental results, we visualized the classification results of various approaches on three HSI classification datasets, which is exhibited in Figs. 8–10. In all three classification maps of SVM, there is lots of salt–pepper classification noise because it only utilizes spectral features for classification and may have the problem of insufficient feature extraction. Compared with other algorithms, the DFFN algorithm produces obvious oversmooth classification results due to the weak feature discrimination ability of spatial patches, especially on the Pavia University dataset, and also produces a large number of boundary misclassifications in other two datasets. Overall, the classification approaches based on FCN (the proposed FCSCL algorithm and FreeNet) are closest to the ground-truth maps in the three datasets, and the FCSCL method has better classification results at the boundary, since it constructs a tighter feature space through contrastive learning, making the outline clearer and more continuous, which is closest to the real boundary.

TABLE IV
CLASSIFICATION RESULTS OF SVM, CVSSN, RESN, DFFN, SSFCN, FREENET, AND FCSCL ON THE PAVIA UNIVERSITY DATASET

	class	SVM	CVSSN	RESN	DFFN	SSFCN-CRF	FreeNet	FCSCL
Accuracies (%)	C1	70.50	93.55	87.89	83.59	91.13	94.94	**96.94**
	C2	70.50	94.40	**99.84**	82.56	88.17	96.58	98.53
	C3	76.08	72.54	90.07	95.31	78.50	95.61	**99.08**
	C4	93.58	82.63	**98.88**	80.36	97.20	97.24	**98.88**
	C5	99.32	99.55	99.24	99.31	99.47	**100.00**	**100.00**
	C6	87.09	60.46	83.79	85.72	96.88	98.30	**99.30**
	C7	88.12	82.59	88.74	98.83	80.54	**100.00**	99.69
	C8	77.22	91.96	88.46	86.65	67.51	94.00	**98.91**
	C9	99.89	94.70	99.46	83.84	99.02	99.24	**99.89**
OA(%)		77.02	86.22	93.67	84.94	88.39	96.58	**98.57**
AA(%)		84.72	85.82	92.93	88.46	88.69	97.32	**99.03**
Kappa		0.7113	0.8213	0.9159	0.8059	0.8494	0.9547	**0.9811**

TABLE V
CLASSIFICATION RESULTS OF SVM, CVSSN, RESN, DFFN, SSFCN, FREENET, AND FCSCL ON THE WHU_HI_HONGHU DATASET

	class	SVM	CVSSN	RESN	DFFN	SSFCN-CRF	FreeNet	FCSCL
Accuracies (%)	C1	82.80	**98.90**	94.78	88.47	77.09	96.45	96.63
	C2	71.41	74.17	75.53	80.36	42.41	**97.93**	95.73
	C3	69.14	96.62	79.03	86.02	87.75	90.99	**95.24**
	C4	65.24	**99.59**	93.95	93.91	98.90	98.19	98.98
	C5	59.33	26.96	84.72	96.11	94.14	96.66	**97.98**
	C6	72.97	91.06	86.58	88.75	91.01	89.91	**91.41**
	C7	48.51	77.99	61.03	50.86	58.80	77.38	**82.70**
	C8	26.23	33.14	60.14	40.76	88.06	89.68	**96.05**
	C9	90.30	98.94	95.47	92.06	85.94	96.67	**98.94**
	C10	45.79	84.67	63.98	65.04	46.07	94.18	**95.93**
	C11	37.22	67.15	64.91	46.50	47.29	89.06	**92.68**
	C12	52.45	68.94	80.85	94.86	56.42	90.26	**95.23**
	C13	45.86	89.63	70.2	58.98	79.37	81.22	**89.62**
	C14	64.30	88.98	86.88	86.33	62.38	92.57	**95.35**
	C15	77.07	71.63	93.35	90.34	93.45	**100.00**	**100.00**
	C16	68.80	**94.09**	85.27	91.24	54.43	93.72	93.26
	C17	68.24	84.32	**98.76**	93.28	92.33	97.55	97.66
	C18	47.34	56.83	92.36	60.56	65.70	98.72	**99.72**
	C19	59.10	91.59	76.74	97.39	11.87	**93.58**	91.23
	C20	63.77	83.46	87.20	70.81	11.01	95.09	**98.96**
	C21	64.47	46.82	95.63	93.04	99.62	99.85	**100.00**
	C22	62.12	50.90	76.16	81.40	94.00	96.71	**99.45**
OA(%)		63.12	85.90	85.24	84.27	83.07	93.42	**95.47**
AA(%)		61.02	76.20	81.98	79.41	69.91	93.47	**95.58**
Kappa		0.5666	0.8267	0.8152	0.8043	0.7865	0.9173	**0.9429**

TABLE VI
CLASSIFICATION RESULTS OF SVM, CVSSN, RESN, DFFN, SSFCN, FREENET, AND FCSCL ON THE WHU_HI_HANCHUAN DATASET

	class	SVM	CVSSN	RESN	DFFN	SSFCN-CRF	FreeNet	FCSCL
Accuracies (%)	C1	59.03	**95.61**	90.01	76.76	94.87	94.52	93.97
	C2	40.02	87.52	64.51	83.21	66.17	81.37	**88.01**
	C3	64.66	84.67	**94.82**	81.86	72.19	93.80	91.56
	C4	85.15	92.84	**98.09**	93.81	84.68	93.77	97.52
	C5	52.51	61.05	97.87	**100.00**	94.3	98.04	99.74
	C6	39.37	34.08	63.00	59.11	49.38	93.57	**99.36**
	C7	76.13	40.52	97.77	95.68	81.63	95.65	**98.77**
	C8	45.67	75.97	63.91	49.86	55.95	79.23	**86.63**
	C9	36.11	51.29	68.04	86.72	39.22	84.54	**88.36**
	C10	83.06	75.63	**96.76**	92.92	80.88	96.02	95.19
	C11	78.99	88.23	87.39	64.45	**97.25**	88.81	89.36
	C12	44.44	44.07	92.97	76.99	93.46	**100.00**	99.75
	C13	39.47	44.45	69.64	50.51	67.90	83.94	**85.34**
	C14	62.54	87.75	85.8	72.27	70.20	92.86	**97.73**
	C15	66.70	34.24	87.76	91.99	80.74	92.07	**94.15**
	C16	76.79	98.58	99.18	71.93	89.73	98.67	**99.22**
OA(%)		63.25	81.59	87.03	73.85	76.16	92.38	**94.27**
AA(%)		59.42	68.53	84.84	78.00	80.62	91.67	**94.04**
Kappa		0.5837	0.7867	0.8490	0.7037	0.7758	0.9111	**0.9331**

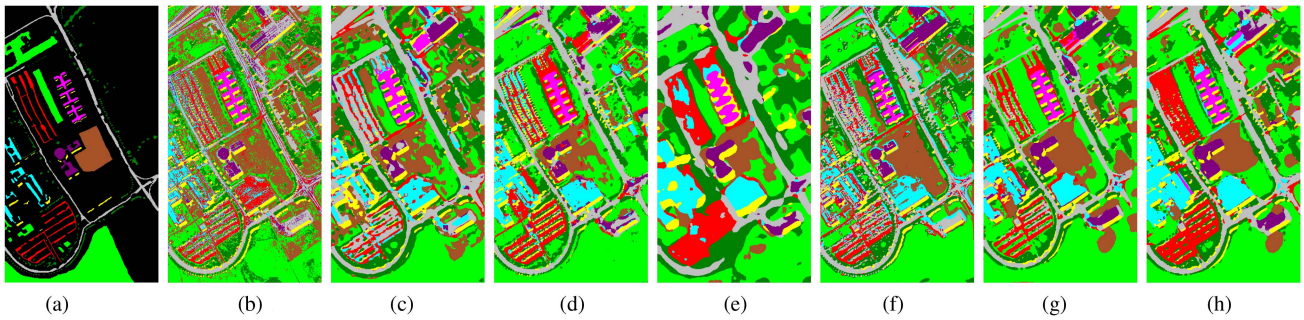


Fig. 8. Classification maps of various algorithms for the Pavia University dataset. (a) Groundtruth. (b) SVM. (c) CVSSN. (d) RESN. (e) DFFN. (f) SSFCN. (g) FreeNet. (h) FCSCL.

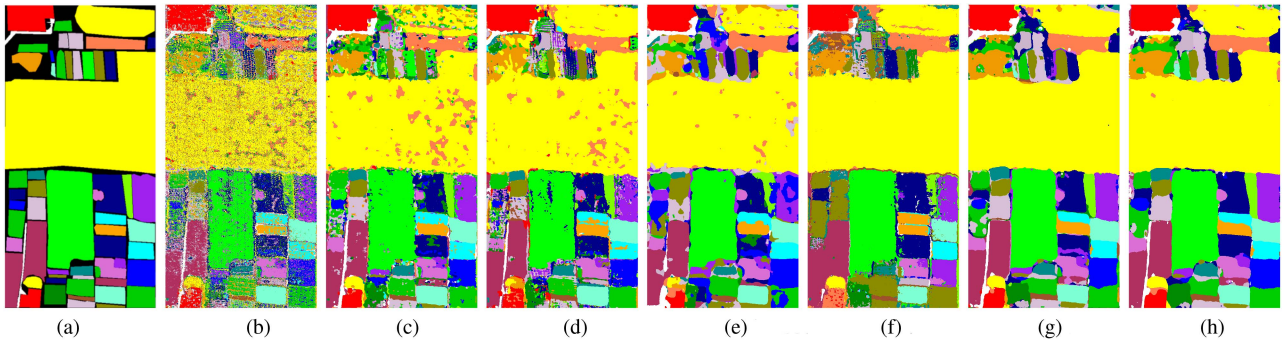


Fig. 9. Classification maps of various algorithms for the WHU_Hi_HongHu dataset. (a) Groundtruth. (b) SVM. (c) CVSSN. (d) RESN. (e) DFFN. (f) SSFCN. (g) FreeNet. (h) FCSCL.

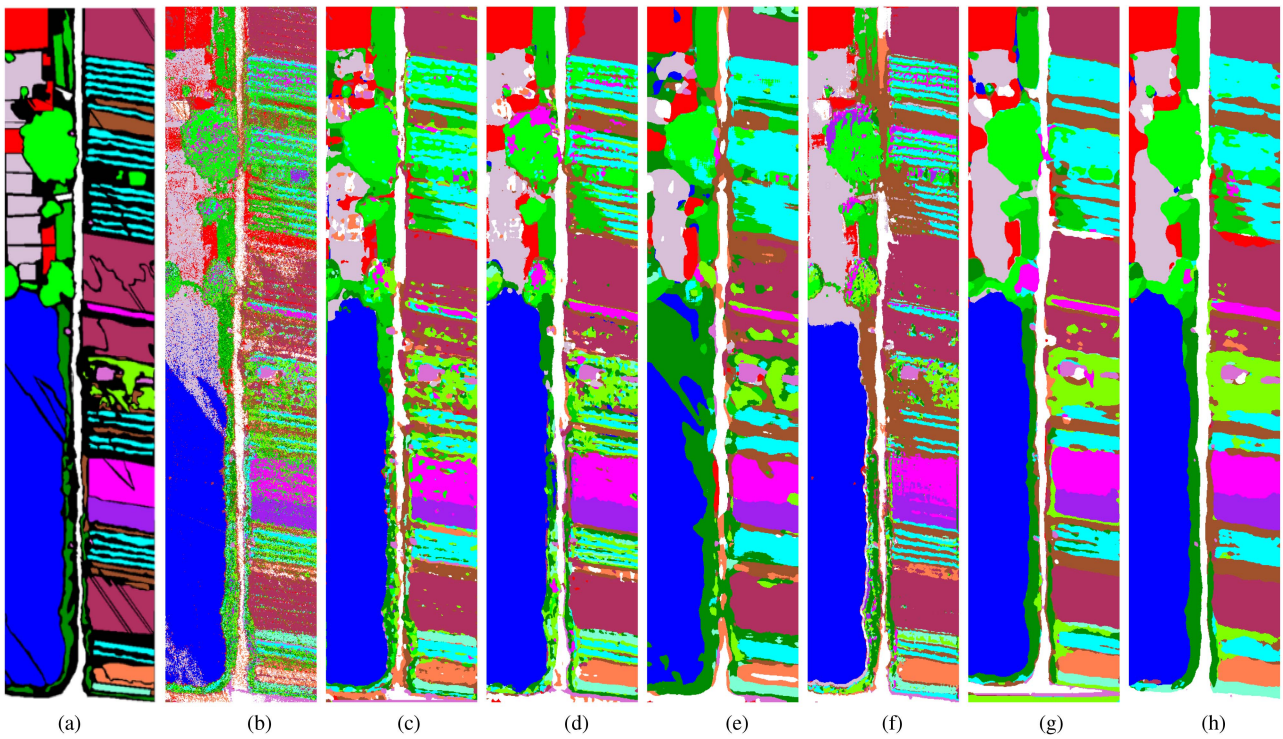


Fig. 10. Classification maps of various algorithms for the WHU_Hi_HanChuan dataset. (a) Groundtruth. (b) SVM. (c) CVSSN. (d) RESN. (e) DFFN. (f) SSFCN. (g) FreeNet. (h) FCSCL.

TABLE VII
EFFECTS OF ADAPTIVE FUSION MECHANISM (AF) AND CONTRASTIVE LEARNING MODULE (CL) ON THREE DATASETS

	Pavia University			WHU_Hi_HongHu			WHU_Hi_HanChuan		
	w/o CL	w/o AF	FCSCL	w/o CL	w/o AF	FCSCL	w/o CL	w/o AF	FCSCL
OA(%)	96.22	97.11	98.57	94.45	95.03	95.47	92.05	92.63	94.27
AA (%)	98.14	98.62	99.03	94.59	95.27	95.58	91.52	91.83	94.04
Kappa	0.9504	0.9619	0.9811	0.9301	0.9374	0.9429	0.9074	0.9141	0.9331

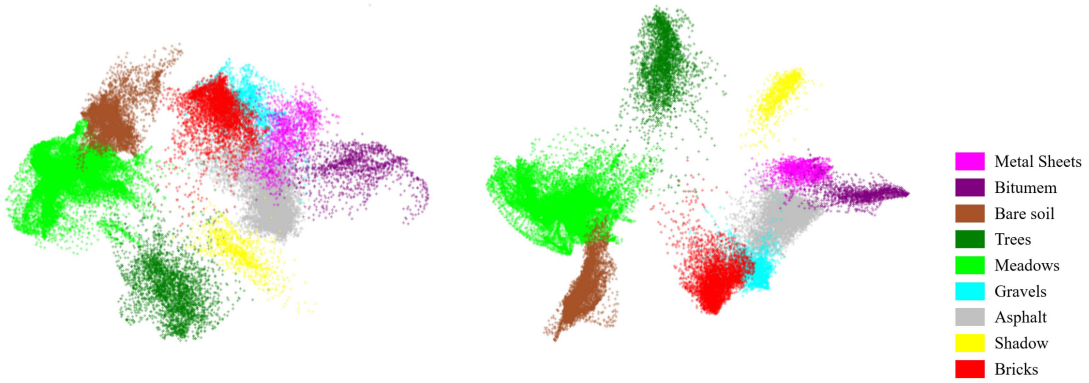


Fig. 11. Data distributions of the labeled samples of the Pavia University dataset in (a) the original spectral feature space and (b) the learned feature space.

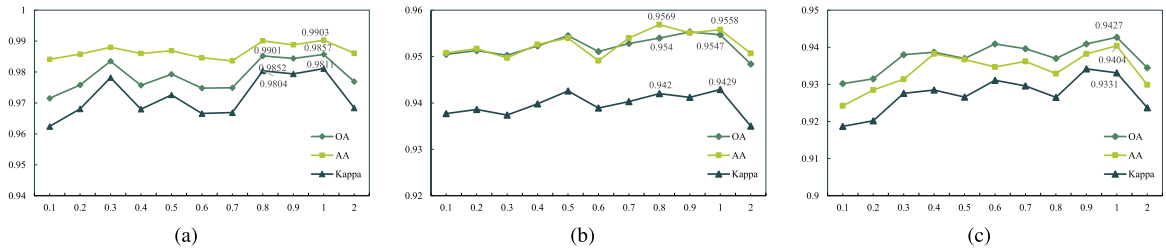


Fig. 12. Sensitivity analysis of parameter λ on (a) Pavia University dataset, (b) WHU_Hi_HongHu dataset, and (c) WHU_Hi_HanChuan dataset.

D. Ablation Experiments

To fully validate the impact of our proposed adaptive fusion mechanism and contrastive learning module on the overall framework, we conduct ablation experiments on these three datasets in this section. For the scene without contrastive learning module, the InfoNCE loss function is removed and only the cross-entropy loss is used to optimize the model. For the scene without adaptive fusion mechanism, the spectral branch is removed from network to observe the results of experiments.

The results of the ablation experiments on the three HSI classification datasets are shown in Table VII. For the effect of adaptive fusion mechanism, the proposed FCSCL algorithm extracts discriminative features from spectral and spatial dimensions for adaptive fusion, resulting in an average accuracy improvement of about 1% on the three datasets. For the effect of contrastive learning module, taking the Pavia University dataset as an example, the proposed SCL module can improve OA by 2.35%, Kappa by 0.0307, and AA by 0.89%. This is because the contrastive learning module using the InfoNCE loss function constructs an efficient classification feature space to provide the ability of intraclass compactness and interclass separability, and the FCN using cross-entropy loss can find the optimal decision boundary in this classification feature space. Using an FCN alone may only focus on the decision boundary, but not on the feature distribution in the classification space. To show the role of contrastive learning more intuitively, the classification feature space is visualized, as shown in Fig. 11. Each category has significantly tighter feature distributions in the feature space using the contrastive learning module. In the ablation experiments of the remaining two datasets, the conclusions that are basically consistent with the Pavia University dataset can be obtained.

E. Sensitivity Analysis

The parameter λ is used to balance the impact of FCN classification and SCL in the loss function, which will directly affect the final classification performance. To explore the effect of λ on the network performance, we use the grid search method on the Pavia University, WHU_Hi_HongHu, and WHU_Hi_HanChuan datasets to explore the optimal value of

λ, $\lambda \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 2.0\}$. The experimental results are shown in Fig. 12.

As can be seen from Fig. 12, as the parameter λ increases, the impact of SCL in the network will also become greater, so the classification accuracy of three datasets (Pavia University, WHU_Hi_HongHu, and WHU_Hi_HanChuan datasets) have an increasing trend. However, when the parameter λ reaches a certain value, the classification accuracy has an obvious downward trend. This is mainly because it reduces the relative role of cross-entropy loss in the network, which will affect the learning of discriminative features. Therefore, the parameter λ was set to 1 in the experiments to better balance the effects of semantic segmentation and SCL.

IV. Conclusion

In this study, an FCSCL framework integrating FCN and contrastive learning is developed for HSI classification. To exploit the spectral and spatial information of hyperspectral images, the spectral–spatial fusion classification network is designed in the FCSCL framework, which uses the spectral–spatial adaptive fusion module to merge the spectral and spatial feature extraction branches. Based on the shared spectral–spatial fusion classification network, SCL head and a fully convolutional semantic segmentation head are jointly applied in the framework, which can complement each other to some extent. For SCL, an effective sample pair construction strategy is developed by sampling hard example pairs, and the contrastive loss is used to optimize the similarity of these constructed positive and negative sample pairs to improve intraclass compactness and interclass separability. The fully convolutional semantic segmentation module expects the network to learn discriminative features for classification. Therefore, the SCL is integrated in fully convolutional semantic segmentation to learn semantically meaningful and discriminative feature representation for accurate hyperspectral image classification. In the future, we will construct a hyperspectral classification dataset with spatial separation of training data and test data based on existing works [53], [54] to test the generalization of the algorithm.

Acknowledgment

The authors would like to thank the editor, associate editor and anonymous reviewers for their helpful comments and suggestions, and also thank the RSIDEA group supervised by Prof. Yanfei Zhong, Wuhan University for providing the free downloads of the WHU_Hi_HongHu and WHU_Hi_HanChuan datasets.

References

[1] A. Shukla and R. Kot, "An overview of hyperspectral remote sensing and its applications in various disciplines," *IRA-Int. J. Appl. Sci.*, vol. 5, pp. 85–90, 2016.

[2] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118–14129, 2018.

[3] T. A. Carrino, A. P. Crósta, C. L. B. Toledo, and A. M. Silva, "Hyperspectral remote sensing applied to mineral exploration in southern Peru: A multiple data integration approach in the Chapi Chiara gold prospect," *Int. J. Appl. Earth Observation Geoinf.*, vol. 64, pp. 287–300, 2018.

[4] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.

[5] Y. Zhong and L. Zhang, "An adaptive artificial immune network for supervised classification of multi-/hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 894–909, Mar. 2012.

[6] A. Samat, P. Du, S. Liu, J. Li, and L. Cheng, "E^2LMs : Ensemble extreme learning machines for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1060–1069, Apr. 2014.

[7] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.

[8] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[9] P. Ghamisi, J. A. Benediktsson, and J. R. Sveinsson, "Automatic spectral–spatial classification framework based on attribute profiles and supervised feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5771–5782, Sep. 2014.

[10] A. Samat, E. Li, W. Wang, S. Liu, C. Lin, and J. Abuduwaili, "Meta-xgboost for hyperspectral image classification using extended MSER-guided morphological profiles," *Remote Sens.*, vol. 12, no. 12, 2020, Art. no. 1973.

[11] B. Liang, C. Liu, J. Li, A. Plaza, and J. M. Bioucas-Dias, "Semisupervised discriminative random field for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 12403–12414, 2021.

[12] Y. Zhong, X. Lin, and L. Zhang, "A support vector conditional random fields classifier with a Mahalanobis distance boundary constraint for high spatial resolution remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1314–1330, Apr. 2014.

[13] J. Zhao, Y. Zhong, X. Hu, L. Wei, and L. Zhang, "A robust spectral-spatial approach to identifying heterogeneous crops using remote sensing imagery with high spectral and spatial resolutions," *Remote Sens. Environ.*, vol. 239, 2020, Art. no. 111605.

[14] J. Zhao, S. Tian, C. Geiß, L. Wang, Y. Zhong, and H. Taubenböck, "Spectral-spatial classification integrating band selection for hyperspectral imagery with severe noise bands," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1597–1609, 2020.

[15] W. Li, S. Prasad, and J. E. Fowler, "Hyperspectral image classification using Gaussian mixture models and Markov random fields," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 153–157, Jan. 2014.

[16] H. Yu, L. Gao, J. Li, S. S. Li, B. Zhang, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification using subspace-based support vector machines and adaptive Markov random fields," *Remote Sens.*, vol. 8, no. 4, 2016, Art. no. 355.

[17] M. Ahmad et al., "Hyperspectral image classification—traditional to deep models: A survey for future prospects," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 968–999, 2022.

[18] K. Jiang et al., "Multi-scale hybrid fusion network for single image deraining," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 7, pp. 3594–3608, Jul. 2023.

[19] Y. Chen, X. Zhao, and X. Jia, "Spectral–spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.

[20] 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, Oct. 2016.

[21] 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.

[22] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 4959–4962.

[23] 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.

[24] L. Fang, G. Liu, S. Li, P. Ghamisi, and J. A. Benediktsson, "Hyperspectral image classification with squeeze multibias network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1291–1301, Mar. 2019.
[25] 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.
[26] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, 2017, Art. no. 67.
[27] 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, Sep. 2019.
[28] M. Ahmad et al., "A disjoint samples-based 3D-CNN with active transfer learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5539616.
[29] 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.
[30] L. Zou, X. Zhu, C. Wu, Y. Liu, and L. Qu, "Spectral—spatial exploration for hyperspectral image classification via the fusion of fully convolutional networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 659–674, 2020.
[31] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.
[32] M. Li, Y. Liu, G. Xue, Y. Huang, and G. Yang, "Exploring the relationship between center and neighborhoods: Central vector oriented self-similarity network for hyperspectral image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 4, pp. 1979–1993, Apr. 2023.
[33] S. Hao, W. Wang, Y. Ye, T. Nie, and L. Bruzzone, "Two-stream deep architecture for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2349–2361, Apr. 2018.
[34] Y. Xu, L. Zhang, B. Du, and F. Zhang, "Spectral–spatial unified networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5893–5909, Oct. 2018.
[35] C. Shi, D. Liao, Y. Xiong, T. Zhang, and L. Wang, "Hyperspectral image classification based on dual-branch spectral multiscale attention network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10450–10467, 2021.
[36] Z. Zheng, Y. Zhong, A. Ma, and L. Zhang, "FPGA: Fast patch-free global learning framework for fully end-to-end hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5612–5626, Aug. 2020.
[37] Y. Xu, B. Du, and L. Zhang, "Beyond the patchwise classification: Spectral-spatial fully convolutional networks for hyperspectral image classification," *IEEE Trans. Big Data*, vol. 6, no. 3, pp. 492–506, Sep. 2020.
[38] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9729–9738.
[39] X. Chen et al., "Unpaired deep image deraining using dual contrastive learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 2017–2026.
[40] Y. Xiao, Q. Yuan, K. Jiang, J. He, Y. Wang, and L. Zhang, "From degrade to upgrade: Learning a self-supervised degradation guided adaptive network for blind remote sensing image super-resolution," *Inf. Fusion*, vol. 96, pp. 297–311, 2023.
[41] L. Zhao, W. Luo, Q. Liao, S. Chen, and J. Wu, "Hyperspectral image classification with contrastive self-supervised learning under limited labeled samples," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6008205.
[42] S. Hou, H. Shi, X. Cao, X. Zhang, and L. Jiao, "Hyperspectral imagery classification based on contrastive learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5521213.
[43] H. Xu, W. He, L. Zhang, and H. Zhang, "Unsupervised spectral–spatial semantic feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5526714.
[44] D. Ivanitsa and W. Wei, "Efficient hyperspectral imagery classification method with lightweight structure and image transformation-based data augmentation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 3560–3563.
[45] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.
[46] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
[47] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2019, *arXiv:1807.03748*.
[48] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, and L. Zhang, "WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF," *Remote Sens. Environ.*, vol. 250, 2020, Art. no. 112012.
[49] X. Hu et al., "SPNet: Spectral patching end-to-end classification network for UAV-borne hyperspectral imagery with high spatial and spectral resolutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5503417.
[50] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, pp. 1–27, 2011.
[51] Y. Xu, B. Du, and L. Zhang, "Robust self-ensembling network for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 19, 2022, doi: 10.1109/TNNLS.2022.3198142.
[52] W. Song, S. Li, L. Fang, and T. Lu, "Hyperspectral image classification with deep feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3173–3184, Jun. 2018.
[53] D. He, Q. Shi, X. Liu, Y. Zhong, G. Xia, and L. Zhang, "Generating annual high resolution land cover products for 28 metropolises in China based on a deep super-resolution mapping network using Landsat imagery," *GIScience Remote Sens.*, vol. 59, no. 1, pp. 2036–2067, 2022.
[54] Q. Shi, M. Liu, A. Marinoni, and X. Liu, "UGS-1 m: Fine-grained urban green space mapping of 31 major cities in China based on the deep learning framework," *Earth System Sci. Data*, vol. 15, no. 2, pp. 555–577, 2023.

Yifan Shen is currently working toward the undergraduate degree in computer science and technology from the China University of Geosciences, Wuhan, China.

His research interests include computer vision algorithms and image processing.

Ling Shi is currently working toward the undergraduate degree in intelligent science and technology from the School of Computer Science, China University of Geosciences, Wuhan, China.

Her research interests include hyperspectral image classification and natural language processing.

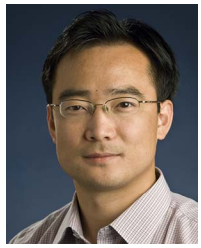
Ji Zhao (Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China, in 2017.

He is currently an Associate Professor with the School of Computer Science, China University of Geosciences, Wuhan. His research interests include remote sensing image classification, scene analysis, remote sensing applications, and machine learning algorithms.

Yuting Dong (Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China, in 2018.

From 2019 to 2020, she was a Postdoc with the German Aerospace Center, Cologne, Germany. She is currently a Professor with the China University of Geosciences, Wuhan. Her research interests include remote sensing data processing and InSAR topographic mapping.

Dr. Dong was the recipient of the Alexander von Humboldt Research Fellowship in 2022.



Lizhe Wang (Fellow, IEEE) received the B.S. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1998 and 2001, respectively, and the D.E. degree (magna cum laude) in applied computing from the University of Karlsruhe, Karlsruhe, Germany, in 2007.

He is currently the "ChuTian" Chair Professor with the School of Computer Science, China University of Geosciences, Wuhan, China. His research interests include remote sensing data processing, digital earth, and big data computing.

Dr. Wang is a Fellow of the Institution of Engineering and Technology and the British Computer Society. He was selected for Distinguished Young Scholars of the National Natural Science Foundation of China, the National Leading Talents of Science and Technology Innovation, and the 100-Talents Program of Chinese Academy of Sciences. He is an Associate Editor for Remote Sensing, International Journal of Digital Earth, ACM Computing Surveys, IEEE Transactions on Parallel and Distributed Systems, and IEEE Transactions on Sustainable Computing.