



## Full length article

## Deep learning for hyperspectral image classification: A comprehensive review and future predictions

Yongchao Song<sup>a</sup>, Junhao Zhang<sup>a,\*</sup>, Zhaowei Liu<sup>a</sup>, Yang Xu<sup>b</sup>, Siwen Quan<sup>c</sup>, Lijun Sun<sup>a</sup>, Jiping Bi<sup>a</sup>, Xuan Wang<sup>a,\*</sup>

<sup>a</sup> School of Computer and Control Engineering, Yantai University, Yantai, 264005, China

<sup>b</sup> Science and Technology Office, Yantai University, Yantai, 264005, China

<sup>c</sup> School of Electronics and Control Engineering, Chang'an University, Xi'an, 710064, China

## ARTICLE INFO

**Keywords:**

Deep learning  
Hyperspectral image  
Target classification  
Spectral-spatial features

## ABSTRACT

Hyperspectral image classification (HSIC) is an important research direction in the field of remote sensing image analysis and computer vision, which is of great practical significance. Hyperspectral imaging (HSI) is widely used in a variety of scenarios with its rich spectral and spatial information, but problems such as high-dimensional data characteristics and scarcity of labeled samples challenge the classification accuracy. Deep learning (DL), with its powerful feature extraction and modeling capabilities, provides an effective means to solve the nonlinear problems in HSIC. In this survey, we systematically review the research progress and applications of DL in HSIC. Firstly, we outline the importance of accurate classification, analyze the features of HSI and the challenges faced by DL in this area. Secondly, we introduce different feature representations of HSI and provide a comprehensive describe of the application of various DL models in HSIC. Meanwhile, we also explore DL methods that can effectively improve the classification performance in the case of insufficient training samples. Finally, we summarize the current research situation, and put forward the future development direction and suggestions.

### 1. Introduction

Hyperspectral imaging (HSI) can acquire reflectance data for hundreds of continuous spectral bands in the wavelength range of 0.4 to 2.4 μm [1]. These spectral bands cover the visible and short-wave infrared regions, capturing detailed spectral characteristics of different materials. This helps distinguish objects that may appear similar. With high spectral resolution and a broad wavelength range, HSI is an effective tool for material detection and analysis. It can also identify emission signatures, including those in the mid- to long-infrared range [2].

While HSI provides a wealth of information, it also presents several significant challenges. HSI data consist of both spatial resolution and spectral dimensions, forming a three-dimensional data cube. Consequently, traditional analysis methods designed for monochrome, RGB, and multispectral images are not directly applicable for effectively extracting meaningful insights from HSI data. To address this complexity, specialized techniques and methodologies are essential.

The main advantage of HSI is the ability to distinguish materials that are similar in visual features such as color, texture, shape, and structure, but differ in spectral properties. This capability is essential across

a multitude of application domains, such as atmospheric monitoring [3, 4], urban planning [5], agricultural production [6], marine and forest health monitoring [7,8], healthcare and food processing [9,10]. HSI also has a wide range of military applications, including camouflage identification, mine detection and coastal area mapping [11–13]. In addition, HSI is also widely used in aerospace and underwater equipment, through the acquisition of detailed spectral information to complete a variety of mission requirements.

HSI research consists of the following six main areas: HSI unmixing [14], HSI change detection [15,16], HSI classification [17], HSI reconstruction and restoration [18–20], HSI data fusion [21], HSI dimensionality reduction [22], and HSI segmentation [23]. According to statistics obtained from “app.dimension.ai”, Fig. 1 illustrates the cumulative number of publications and patents in the HSI field over the past decade. This trend suggests an imminent increase in the use and significance of HSI applications.

In this survey, we focus on hyperspectral image classification (HSIC), a field that has seen significant progress. With the continuous evolution of machine learning, learning-based algorithms have been introduced into HSIC and achieved good results. Due to the rich spectral

\* Corresponding authors.

E-mail addresses: [zhangjunhao@s.ytu.edu.cn](mailto:zhangjunhao@s.ytu.edu.cn) (J. Zhang), [xuanwang91@ytu.edu.cn](mailto:xuanwang91@ytu.edu.cn) (X. Wang).

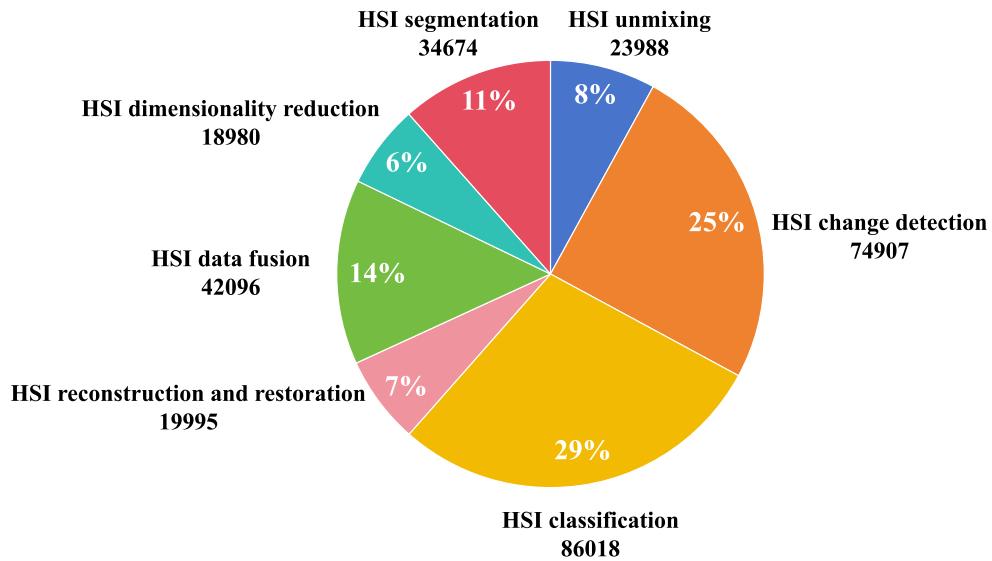


Fig. 1. 2014–2024 Number of HSI-related publications and patents published.

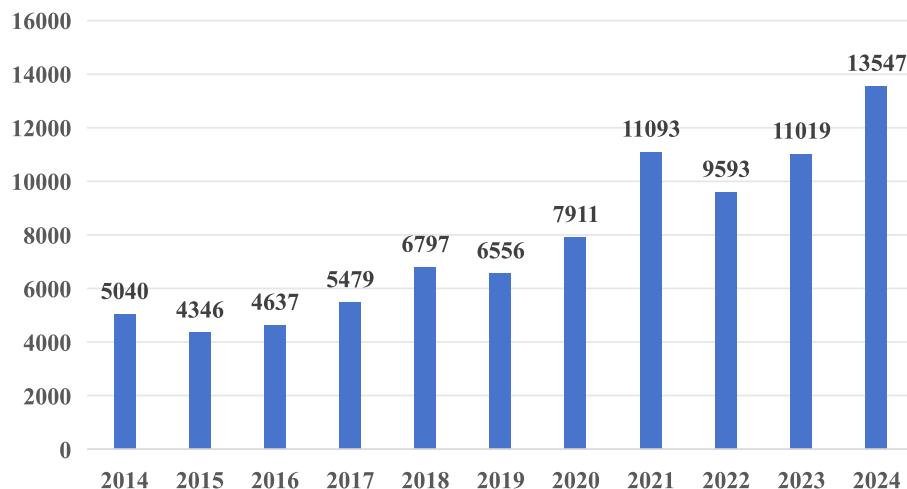


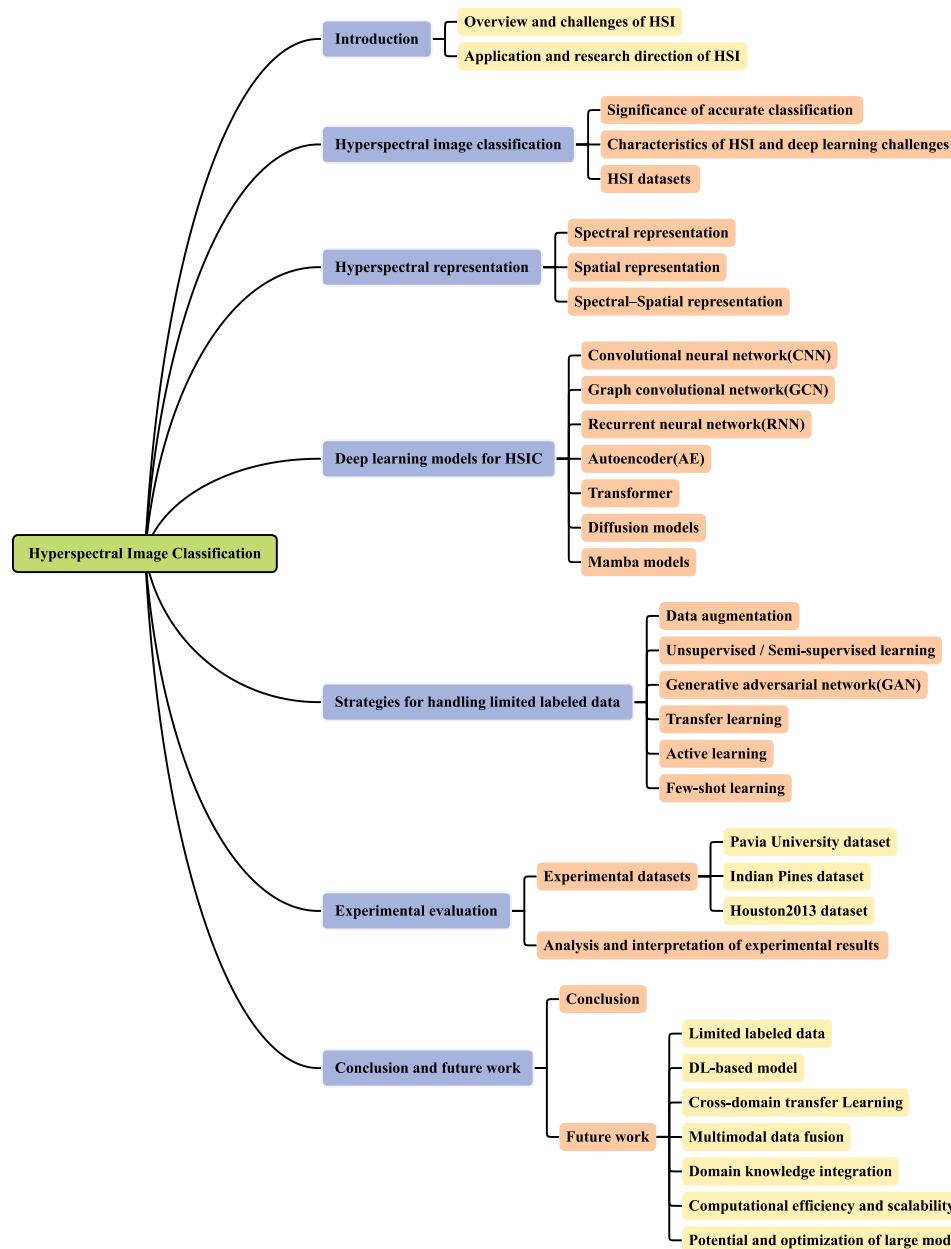
Fig. 2. 2014–2024 Number of publications and patents on DL-based HSIC.

information of HSI, early studies mainly focused on extracting high-dimensional spectral information, such as support vector machines (SVM) [24]. On this basis, Li et al. [25] proposed a multi-core based SVM model. Liu et al. [26] proposed an SVM model based on non-parallelism. In addition, some machine learning related algorithms such as Random Forest [27], Polynomial Logistic Regression [28], and Random Subspace Methods [29] have also been introduced into the field of HSIC. However, there is still room for optimization of traditional machine learning for spatial feature extraction.

In recent years, with the improvement of GPU computational performance, Deep learning (DL) has emerged as a popular direction for HSIC. Chen et al. [30] first applied DL to the HSIC task and combined it with principal component analysis (PCA) [31]. Early deep networks were predominantly shallow, focusing on extracting either spectral or spatial features. Subsequent studies have progressively introduced deeper, multi-layered architectures that effectively integrate spectral-spatial information. For instance, numerous studies have employed

multi-scale, multi-view convolutional networks to capture subtle variations in HSI data [32,33]. Moreover, models leveraging attention mechanisms and Transformer architectures have demonstrated superior ability to capture long-range dependencies, thereby enhancing both classification accuracy and generalization [34,35]. According to the statistics of “app.dimension.ai”, Fig. 2 shows a significant increase in the number of publications and patents for DL-based HSIC methods in the last few years. This trend shows that researchers in the field have a strong interest in DL methods. It also suggests that DL technology will play an increasingly crucial role in HSIC.

The rapid development of DL in HSIC has driven the diversification of model architectures and improved performance. It has also introduced new approaches to address the challenges of high dimensionality, redundancy, and labeling difficulties in HSI data. In the future, as innovative algorithms emerge and cross-domain technologies integrate, HSIC research will reach new heights and yield further breakthroughs in practical applications.



**Fig. 3.** Structural framework of this survey.

The contributions of this survey are as follows:

- This survey summarizes the typical application scenarios of HSIC and analyzes the key challenges faced by HSIC approaches based on DL.
- We systematically introduce the spectral features, spatial features and spectral-spatial features of HSI, and comprehensively review the spectral-spatial HSIC research methods based on DL.
- We propose strategies to address the problem of limited training samples and explore possible research directions in the future.

The remaining investigations are organized as follows: Section 2 provides an overview of HSIC applications and describes the characteristics of HSI and the challenges faced by DL in this area. Section 3 describes the various representations of HSI. Section 4 systematically reviews various DL-based classification methods. Section 5 discusses solution strategies to cope with the problem of limited labeled data.

Section 6 selected different models for comparative experiments to further analyze the characteristics between the different models. Section 7 concludes and gives suggestions for future research. The structural framework of this survey is shown in Fig. 3.

## 2. Hyperspectral image classification

### 2.1. Significance of accurate classification

The main objective of HSIC is to accurately identify and classify different types of characteristics, such as forests, agricultural land, water bodies, and urban areas. By accurately classifying HSI, confusion caused by similarity of spectral features between different feature classes can be avoided, thus effectively preventing erroneous decisions. For example, in areas such as land management administration, environmental monitoring or urban planning, the accuracy of classification is crucial [36–38]. In addition, classification results can show spatial

and temporal trends of different features. This provides valuable insights into ecosystems, land-use changes, and environmental shifts. It also helps support scientific and informed decision-making.

In agricultural monitoring, by analyzing the spectral characteristics of different crops in different bands, HSIC can be used to classify crop types, such as precise fertilizer application, irrigation, and pest control [39]. The higher the classification accuracy, the better the effect of measures such as fertilizer application, watering and pest control, etc. HSIC can also be used to monitor the type, moisture, pH and nutrient content of soil to optimize the management of agricultural production and resource allocation [40,41].

In the field of environmental monitoring, HSIC can accurately classify different features and is widely used in land use, land cover change monitoring and urban expansion analysis. With the help of HSI data, HSIC can effectively identify and monitor environmental problems such as water pollution and air pollution sources, and can also be used to detect potential environmental hazards such as oil spills and industrial wastewater discharges, thus providing important support for environmental protection and governance [42–44].

In mineral exploration, HSI can accurately distinguish between different minerals, rocks and soil types on the surface, so accurate classification can help locate mineral deposits, evaluate mineral resources and formulate exploration plans in geological exploration [45,46]. By analyzing the hyperspectral features of minerals, the distribution of different minerals can be accurately identified, reducing exploration costs.

In addition, HSIC offers significant advantages in disaster monitoring and military security [47]. Accurate classification of HSI is crucial for better understanding the Earth's surface. It plays a vital role in addressing complex challenges and promoting sustainable development.

## 2.2. Characteristics of HSI and deep learning challenges

The main task of HSIC is to assign a unique label to each pixel vector based on the spectral or spectral-spatial properties of the HSI cube. Mathematically the HSI cube can be represented as:

$$\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3, \dots, \mathbf{t}_K]^T \in \mathbb{R}^{K \times (I \times J)} \quad (1)$$

where  $K$  is the total number of spectral bands and each band consists of  $I \times J$  samples of category  $L$ . Each sample  $\mathbf{t}_i = [t_{1,i}, t_{2,i}, t_{3,i}, \dots, t_{K,i}]^T$  has a category label  $l_i \in \mathbb{R}^L$ . HSIC problem can be considered as an optimization problem, which we can represent as:

$$\mathbf{L} = g_\alpha(\mathbf{T}, \theta) \quad (2)$$

where the mapping function  $g_\alpha(\cdot)$  is responsible for converting the input data  $T$  into predicted category labels  $L$ . The adjustable parameter  $\theta$  is used to transform the input data  $T$ , allowing  $g$  to map the spectral data to the category label space  $L$  [48]. By adjusting  $\theta$ , the mapping process can be optimized to minimize the gap between the predicted results and the actual labels.

DL has revolutionized HSIC by surpassing traditional methods, integrating neural networks with attention mechanisms [49,50]. For instance, convolutional neural networks (CNNs) excel at learning hierarchical representations, effectively capturing both spatial and spectral dependencies in HSI [51]. Recurrent neural networks (RNNs), designed for sequence data processing, offer distinct advantages for temporal analysis in HSIC [52]. RNNs efficiently capture contextual information across time, which aids in tracking land-cover changes and monitoring dynamic environmental processes. Autoencoders (AEs), another type of neural network, have also been utilized in HSIC, particularly when trained on unlabeled HSI data to capture underlying data structures [53]. Recently, Transformers have gained significant attention in fields like natural language processing and computer vision [54,55]. By leveraging a self-attention mechanism, Transformers can capture dependencies between various positions in a sequence. This ability

to model interactions between spectral bands allows Transformers to exploit the rich information in HSI, thereby enhancing classification performance and capturing complex spectral patterns. Additionally, diffusion models [56] and Mamba models [57] have been incorporated into HSIC due to their unique advantages.

While these DL models have found broad potential in HSIC tasks, they still face significant challenges in practical applications. These challenges stem primarily from the fact that HSI data contains a large number of continuous and narrow spectral channels with very high spectral resolution but low spatial resolution. Although the rich spectral information is very useful for classification tasks, processing such complex data requires a lot of time and computational resources. When processing HSI data, with the increase of feature dimension, the computational complexity of the data also rises sharply, which is called the curse of dimensionality [58]. Especially in supervised learning, dimensionality catastrophe can significantly affect classification accuracy. With a limited number of training samples, an increase in the feature dimension often leads to a decrease in the performance of the classification algorithm, and the classification accuracy may be significantly reduced, which results in the Hughes phenomenon [59]. To address this challenge, several strategies can be adopted, such as feature selection, dimensionality reduction (DR), and utilizing unsupervised or semi-supervised learning methods to extend the training dataset.

DL models face multiple challenges in HSIC. The training and parameter optimization of deep neural networks (DNN) are NP-complete problems, and the optimization process is complex and the convergence is unstable. The HSI task requires an adjustment of a large number of parameters and the training difficulty is high. Although optimization algorithms significantly improve convergence efficiency and classification performance, the “black box” nature of DNNs limits model interpretability, which has not been solved. The high dimensionality of HSI data and the limited training samples easily lead to overfitting, especially the Hughes phenomenon, which needs to be fine-tuned to alleviate. The computational complexity of DNN requires high memory, computing power and storage, but parallel and distributed computing technologies have been partially alleviated. Increasing the depth of the network may reduce the training accuracy due to the explosion or disappearance of the gradient, which puts forward higher requirements for deep network design and optimization.

## 2.3. HSI datasets

**Table 1** offers a comprehensive overview of several widely used HSI datasets [47,60–62]. It features key datasets including Pavia University (PU), Pavia Center (PC), Indiana Pines (IP), Houston2013, Houston2018, Salinas Valley (SA), Kennedy Space Center (KSC), Washington DC Mall, AeroRIT, Botswana (BS), WHU-OHS, WHU-Hi-LongKou, WHU-Hi-HongHu, WHU-Hi-HanChuan, QUH-Tangdaowan, QUH-Qingyun, QUH-pingan, Matiwan Village. The table enumerates essential attributes for each dataset, such as Wavelength, Spatial size, the number of spectral bands, ground sampling distance (GSD) in meters per pixel, the number of distinct classes, and the data acquisition mode. The detailed information provides a valuable reference that offers critical insights into the characteristics of these datasets and supports informed decision-making in HSIC research.

## 3. Hyperspectral representation

HSI data is represented as a 3D hypercube, denoted as  $X \in \mathbb{R}^{B \times (N \times M)}$ , which contains both 1D spectral samples and 2D spatial details [63]. Here,  $B$  denotes the number of spectral bands, while  $N$  and  $M$  represent the spatial dimensions (width and height), respectively. **Fig. 4** illustrates an HSI cube from the Pavia University dataset.

**Table 1**

The table lists publicly available HSI datasets used for classification methods.

Dataset	Sensor	Mode	Wavelength	Spatial size	Bands	GSD	Classes
Pavia University (PU)	ROSIS	Aerial	0.43–0.86 $\mu\text{m}$	610 $\times$ 340	103	1.3 m	9
Pavia Center (PC)	ROSIS	Aerial	0.43–0.86 $\mu\text{m}$	1096 $\times$ 715	102	1.3 m	9
Indiana Pines (IP)	AVIRIS	Aerial	0.4–2.5 $\mu\text{m}$	145 $\times$ 145	224	20 m	16
Houston2013	NCALM	Aerial	0.38–1.05 $\mu\text{m}$	349 $\times$ 1905	144	2.5 m	15
Houston2018	NCALM	Aerial	0.38–1.05 $\mu\text{m}$	601 $\times$ 2384	48	1 m	20
Salinas Valley (SA)	AVIRIS	Aerial	0.4–2.5 $\mu\text{m}$	512 $\times$ 217	204	3.7 m	16
Kennedy Space Center (KSC)	AVIRIS	Aerial	0.4–2.5 $\mu\text{m}$	512 $\times$ 614	176	18 m	13
Washington DC Mall	HYDICE	Aerial	0.4–2.4 $\mu\text{m}$	1208 $\times$ 307	191	1.5–3 m	7
AeroRIT	Headwall Micro E	Aerial	0.397–1.003 $\mu\text{m}$	1973 $\times$ 3975	372	0.4 m	5
Botswana (BS)	Hyperion	Satellite	0.4–2.5 $\mu\text{m}$	512 $\times$ 614	145	30 m	14
WHU-OHS	Orbita hyperspectral micro	Satellite	0.4–1 $\mu\text{m}$	512 $\times$ 512	32	10 m	24
WHU-Hi-LongKou	Headwall Nano-Hyperspec	UAV aerial	0.4–1 $\mu\text{m}$	550 $\times$ 400	270	0.463 m	9
WHU-Hi-HongHu	Headwall Nano-Hyperspec	UAV aerial	0.4–1 $\mu\text{m}$	940 $\times$ 475	270	0.043 m	22
WHU-Hi-HanChuan	Headwall Nano-Hyperspec	UAV aerial	0.4–1 $\mu\text{m}$	1217 $\times$ 303	274	0.109 m	16
QUH-Tangdaowan	Gaiasky mini2-VN	UAV aerial	0.4–1 $\mu\text{m}$	1740 $\times$ 860	176	0.15 m	18
QUH-Qingyun	Gaiasky mini2-VN	UAV aerial	0.4–1 $\mu\text{m}$	880 $\times$ 1360	176	0.15 m	6
QUH-pingan	Gaiasky mini2-VN	UAV aerial	0.4–1 $\mu\text{m}$	1230 $\times$ 1000	176	0.15 m	10
Matiwan Village	V-NIR imaging	UAV aerial	0.4–1 $\mu\text{m}$	3750 $\times$ 1580	250	0.5 m	20

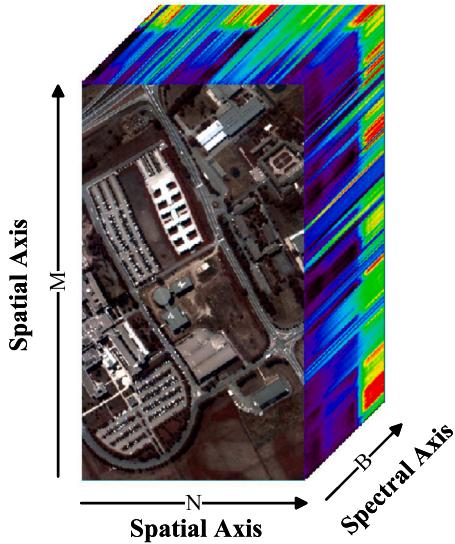


Fig. 4. Hyperspectral spatial and spectral cube.

### 3.1. Spectral representation

In spectral representation, the spectral information of each pixel can be represented as a spectral vector, denoted as  $x_i \in \mathbb{R}^B$ . Here,  $B$  denotes the spectral channel, which can be the actual number of channels or a simplified set obtained by DR techniques. Usually, to avoid redundancy and improve the class separability, data processing prefers to use a low-dimensional HSI representation.

The application of DR Method in HSI can be divided into two categories: supervised and unsupervised. Supervisory methods rely on label information to guide dimension reduction processes, such as linear discriminant analysis (LDA) [64], local fisher discriminant analysis (LFDA) [65], local discriminant embedding (LDE) [66], and non-parametric weighted feature extraction (NWFE) [67]. Unsupervised dimensionality reduction methods do not rely on label information. They reduce the dimensionality based on the structural characteristics of the data. Common examples include PCA, local linear embedding (LLE) [68], and non-negative matrix factorization (NMF) [69]. However, spectral mixing makes it hard to distinguish categories using only spectral reflectance. Therefore, advanced techniques are needed to improve classification accuracy in HSI.

### 3.2. Spatial representation

To overcome the limitations of spectral representation in HSI, spatial information can be incorporated. Each band can be represented as a matrix, denoted as  $x_i \in \mathbb{R}^{N \times M}$ , which captures the spatial structure of pixels. In most cases, neighboring pixels share similar characteristics. Therefore, it is important to include the information of adjacent pixels during spatial representation. This can be done using core-based or pixel-centric window techniques [70].

There are several methods to extract spatial information from HSI cubes. In some cases that do not rely on DL, traditional image processing techniques can be used to extract spatial information. For example, morphological contours (MPs) and region growth algorithms can effectively extract spatial structure and connectivity features of images. In addition, grey level co-occurrence matrix (GLCM), gabor filter, and local binary mode (LBP) are used to analyze spatial structure features such as texture and shape of images through mathematical models to further enhance the extraction effect of spatial information. Zhang et al. [71] combined gabor filters with differential morphological profiles and applied them to HSIC. This shows that these technologies can be fully integrated for extracting spatial information. In addition, spatial features can be extracted based on multi-scale analysis. For example, the pyramid method captures spatial features at different scales by downsampling images at multiple scales to generate images with different resolutions. The wavelet transform can gradually extract spatial information from low frequency to high frequency, which is especially suitable for capturing local structural changes of images.

### 3.3. Spectral-spatial representation

HSI are essentially composed of multiple continuous band images, and the spectral value of each pixel in each band reflects the composition and properties of the object. However, spectral information only provides the radiation characteristics of each pixel, and cannot fully capture the spatial relationship in the image. Spatial information describes the spatial distribution and interrelationship of pixels, which can reveal the shape, boundary and surrounding environment of objects in the image. The combination of spectral information and spatial information makes the analysis of HSI more comprehensive and can overcome the limitations of single spectral or spatial analysis methods.

There are two common spectral-spatial fusion methods. One method is based on feature level fusion, which is usually processed by combining spectral features with spatial features into new feature vectors. For example, pixel-level fusion directly combines the spectral information of each pixel with the spatial information of its neighborhood (such as texture, edge, etc.) to form a high-dimensional feature vector. Local neighborhood feature extraction extracts spatial features (such as

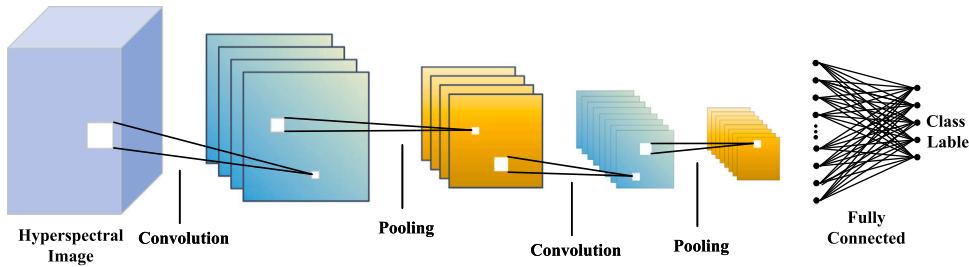


Fig. 5. General structure of CNN.

mean, variance, gradient, etc.) by analyzing the neighborhood around each pixel, and combines them with spectral information. In order to improve the computational efficiency, the feature dimension is often reduced by PCA or LDA. Another approach is learning-based fusion. Recently, many spectral-spatial fusion methods have used DL models such as CNNs and Transformers. These models automatically learn how to combine features from the data, avoiding manual feature design.

In the subsequent sections, we focus on the models for extracting spectral-spatial features.

#### 4. Deep learning models for HSIC

##### 4.1. Convolutional neural network (CNN)

The CNN architecture is inspired by Hubel and Wiesel's bio-visual system [72] and Fukushima's neocognitron model [73]. It consists of two main parts: a feature extraction (FE) network and a classification stage for the extracted feature maps. As shown in Fig. 5, the general model of CNN includes convolutional layer, pooling layer, and fully connected(FC) layer. The FE network extracts features from the input data step by step using convolution and pooling layers. The classification stage then uses these features for prediction. Due to this structure, CNNs are widely used in image classification and object detection tasks.

Convolutional layers are characterized by the use of a set of convolutional kernels and biases to process the input data. These convolutional kernels have small receptive fields and are used to extract specific features. The feature map of a convolutional layer can be defined as follows:

$$f_j^l = \sum_{i=1}^T \alpha \left( f_i^{l-1} * w_{ij}^l + b_i^l \right) \quad (3)$$

where  $f_j^l$  denotes the  $j$ th feature map of the  $l$ th layer,  $f_i^{l-1}$  represents the  $i$ th feature map of the previous layer ( $l-1$ ),  $T$  is the number of channels,  $w_{ij}^l$  and  $b_i^l$  are the weights and biases, respectively. The operator  $*$  denotes the convolution operation, and  $\alpha$  represents the activation function. In each convolutional layer, the convolutional kernel is shared across the entire input data. A spatial sharing mechanism helps reduce model complexity, which in turn improves training efficiency. After the convolution operation, the results are fed into an activation layer, allowing the network to learn nonlinear patterns through the application of nonlinear activation functions. Pooling is then used to downsample the feature maps, decreasing their resolution while retaining essential discriminative information. This operation also contributes to the network's translation invariance. Following several convolution and pooling layers, the feature maps are flattened into one-dimensional vectors and passed to FC layers. In the classification stage, FC layers and a softmax function are used to compute the probability distribution over the classes based on the extracted features.

Recently, multi-scale feature extraction has gained significant attention in research. However, a common problem with traditional DL models is that the learned model may not be effective in images with high intra-class variance and low inter-class variance. Gong et al. [74]

have proposed a multiscale CNN (MS-CNN) to effectively extract multiscale deep features from HSI, aiming to address the challenge of capturing complex spectral-spatial patterns. To improve the limitations of traditional metric learning, where metric parameters tend to become overly similar and lead to redundancy, they incorporated determinantal point process (DPP) priors into the model. These priors encourage diversity among the learned factors, enhancing the network's ability to extract more discriminative features for HSIC. Zhang et al. [75] have proposed the SPRN model, which divides the input spectral bands into multiple non-overlapping contiguous sub-bands. Spectral-spatial features from each sub-band are independently extracted using improved residual cascading parallel blocks. The multi-scale features are subsequently fused and passed to the classifier for final prediction. To further enhance performance, the model incorporates a spatial attention module called the homogeneous pixel detection module (HPDM), designed to capture uniform pixels within the input image block. The module effectively mitigates the performance degradation typically caused by increasing block size and can be flexibly integrated into any CNN-based HSIC framework.

Some researchers have proposed innovative solutions by combining CNNs to address the challenges of small sample training and unsupervised learning. Dong et al. [76] have proposed a pixel cluster CNN and spectral-spatial fusion (SSF) algorithm for HSIC. First, spatial information is extracted using a grey scale covariance matrix. Then, the spatial information is fused with spectral data through band superposition to create spectral-spatial features. To increase the number of training samples, the pixels processed by SSF are grouped into pixel clusters based on specific rules. Finally, effective features are extracted from these pixel clusters using the CNN framework. Due to the limited adaptability of this method across different types of HSI or datasets, more tuning and optimization may be required for specific datasets. In contrast, Ye et al. [77] have introduced self-supervised learning with the multiscale densely connected network (SS-MSDCNet), a novel unsupervised method designed to fully leverage unlabeled samples for HSIC. The network employs a spectral segmentation-based data enhancement technique to synchronize the dual-stream structure of SS-MSDCNet, improving spectral representation. Additionally, multi-scale spectral-spatial features are extracted using a 3D dense connectivity module and a 3D spatial attention module.

Although these models perform well on specific tasks or datasets, they still face many challenges. For example, techniques like multi-scale convolution, 3D convolution, and spatial attention modules can lead to high computational costs, making them difficult to apply in resource-limited environments. Additionally, small sample learning methods often rely on more refined feature selection and data augmentation techniques, while CNNs may not fully capture the complex patterns in the data.

##### 4.2. Graph convolutional network (GCN)

GCNs demonstrate notable strengths in modeling complex spatial dependencies, making them particularly effective for HSIC. HSI typically contains hundreds to thousands of consecutive spectral bands,

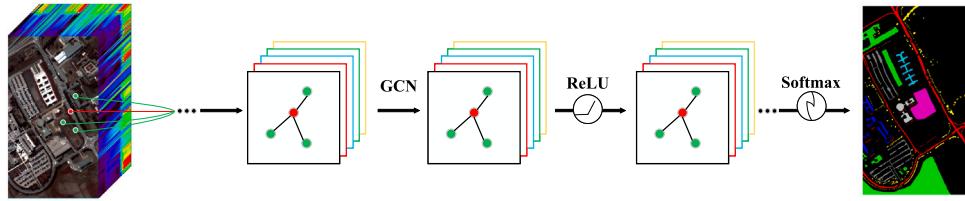


Fig. 6. HSIC general process based on GCN.

each representing a spectral feature. This high-dimensional spectral data makes traditional CNNs difficult to capture the spatial associations between pixels. To address this problem, GCNs model the relationships between pixels by constructing graph structures. In GCNs, the graph structure is the basis of the convolution operation, and the accuracy of its design directly affects the effectiveness of the feature representation. Therefore, it is crucial to customize the appropriate graph structure for a specific task, while selecting the appropriate graph nodes is a key step in constructing the graph. Fig. 6 illustrates the general process of HSIC based on GCN. In GCN-based HSIC methods, pixels or superpixels are typically used as nodes to construct the graph structure.

Shahraki and Prasad [78] have first proposed the use of GCN model for HSIC. Due to the lack of inherent graph structure in hyperspectral data, it is necessary to construct an adjacency matrix based on pixel relationships to enable graph convolution operations. In this context, the Spectral-Spatial GCN (S<sup>2</sup>GCN) [79] leverages both spectral features and spatial distances to define graph connections, enhancing the model's ability to capture local and global context. However, constructing such adjacency matrices can be computationally intensive, especially for large-scale datasets. To overcome this limitation, the miniGCN [80] introduces a small-batch training strategy, which significantly reduces the computational burden and enables scalable learning for HSIC tasks.

Gao et al. [81] have introduced an interaction-enhanced network that integrates multiple self-attention (MSA) mechanisms with graph convolution for hyperspectral and LiDAR data fusion. Initially, spectral and spatial features are independently extracted from the two modalities, followed by the construction of interaction-enhanced fusion features (IEFFs) based on the rank information of central pixels. To further capture spatial dependencies, graph structures are built for four distinct feature sets: interaction-enhanced HSI, interaction-enhanced LiDAR, HSI-only, and LiDAR-only features. Each pixel is treated as a node in the graph, enabling the model to represent complex spatial relationships. A multimodal gated fusion module (MGFM) is then employed to adaptively fuse the extracted features using a learnable weighting mechanism, where the weights are dynamically determined according to feature characteristics. This process ensures an effective and balanced integration of multimodal information.

Traditional GCN methods usually use hyperpixel-based nodes to reduce the computational complexity, but this approach cannot effectively capture pixel-level spectral-spatial features. In addition, these methods usually only focus on the consistency of predicted labels with real labels, ignoring the relationship between intra-class and inter-class distances, resulting in low feature differentiation. To address these issues, Chu et al. [82] have proposed a feature fusion fuzzy graph generalization network (F<sup>3</sup>GBN) for HSIC. They have extracted pixel-level attribute profile features using attribute filters and have fused these features with superpixel features through typical correlation analysis. In addition, a Breadth Learning System (BLS) is employed as a classifier to fully exploit the spectral-spatial information through nonlinear transformations, resulting in significant classification results. Unlike existing hypergraph models, Duan et al. [83] have proposed a structure-preserved hyper GCN (SPHGCN) that combines regular convolution and irregular hypergraph convolution to break the spatial constraints of traditional convolution, aggregate information between different image blocks, and acquire deep spatial-spectral features.

Graph attention network (GAT) [84] can accurately depict the relationship between samples through automatic learning and optimization of node connections, which has attracted the attention of researchers. To address the depth limitations of standard GAT architectures, the Spectral-Spatial Residual GAT (S<sup>2</sup>RGANet) [85] integrates residual connections with graph attention mechanisms, thereby enhancing feature learning capabilities. In addition, the Multi-Relational GAT (MR-GAT) [86] incorporates multi-scale receptive fields to extract both local and global features from neighboring nodes and edges, enriching the model's representation capacity. For multimodal fusion tasks, the Graph-Attention based Multimodal Fusion Network (GAMF) [87] leverages heterogeneous graphs to capture deep semantic relationships and effectively resolve long-range dependency challenges. Further extending the capabilities of attention mechanisms, the Extended Context GAT (ExGAT) [88] improves the modeling of long-distance dependencies through a broader contextual aggregation strategy.

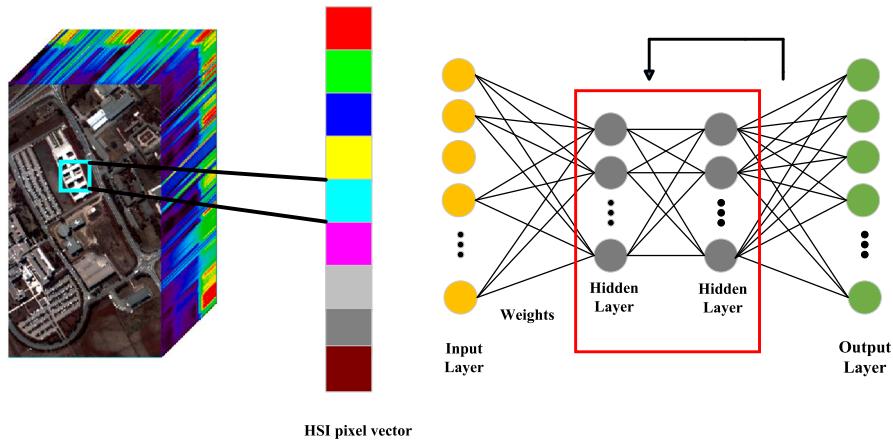
While GCNs capture spatial dependencies by constructing a graph structure between pixels, their effectiveness heavily depends on the design of this structure. The construction of the adjacency matrix or the selection of the graph structure is critical to the network's performance. An inadequate graph structure can hinder information transfer, ultimately diminishing model performance. Designing and constructing an appropriate graph structure becomes even more challenging when dealing with complex or noisy datasets.

#### 4.3. Recurrent neural network (RNN)

RNNs have demonstrated strong capabilities in modeling sequential data, making them particularly suitable for capturing the temporal and semantic characteristics inherent in HSI. Their effectiveness in learning spectral sequences and spatial dependencies enables enhanced representation of complex spectral-spatial information. The fundamental RNN architectures include the Vanilla RNN, Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). These models address the challenge of long-term dependencies through distinct mechanisms. In particular, LSTM and GRU are effective in alleviating the vanishing gradient problem commonly encountered in long sequences, thereby improving the network's ability to learn long-range contextual information. The overall RNN-based architecture is shown in Fig. 7, including cyclic connectivity, where the activation of the next node depends on the previous step [89].

Vanilla RNN is the simplest RNN model, which leads to information degradation when dealing with high-dimensional data. LSTM is able to better learn long-term dependencies in sequences by introducing three gates: input, forget, and output gates and a cell state, which utilizes a gating mechanism to continually discard extraneous information, but this also increases the computational burden. GRU functions similarly to LSTM but has a simpler structure. It changes the input and forget gates to update and reset gates and removes the output gates, thus making training easier.

Gündüz et al. [90] have proposed a model that uses CNNs for spatial feature extraction and GRUs to model spectral dependencies. A double focus mechanism is introduced in the spectral and spatial modules to highlight important spectral bands and key spatial regions, improving model discrimination. However, in complex image data,



**Fig. 7.** The general structure of the HSIC based on RNN.

over-relying on attention mechanisms may overlook some potential key information. To address this, Ramkumar et al. [91] have proposed the Multi-Dimensional RNN (MDRNN) for HSIC. It has applied min-max normalization as a preprocessing technique to reduce image noise and enhance the quality of the HSI. Features were then extracted using a spatial attention-based technique. Finally, MDRNN has learned middle-level visual patterns and has enhanced the spatial relationships between these patterns to improve HSIC.

To address the challenge of small sample learning, Dasu et al. [92] proposed a two-stage framework consisting of feature extraction and classification. HSIs from the RIT-18 dataset are first divided into  $16 \times 16$  patches. An enhanced CNN, optimized by the Modified Velocity-based Colliding Bodies Optimization (MV-CBO), is used to extract spectral-spatial features from each patch. These features are then aggregated into a unified feature vector. In the classification stage, an MV-CBO-optimized RNN is employed to perform classification, leading to improved accuracy. Additionally, Kang et al. [93] have proposed the Spectral-Spatial Double-Branch Network (SSDBN). SSDBN consists of two independent branches for extracting spectral and spatial features, respectively. The framework incorporates multi-scale 2D convolution modules, LSTM, and attention mechanisms. The spectral-spatial double-branch feature extraction method of SSDBN performs well in handling small sample problems.

The 3D CNN model uses a feed-forward structure that reduces the processing burden for 3D data. However, due to its vector-based design, it struggles to fully capture the complex information in HSI, leading to less discriminative feature extraction. In contrast, convolutional long short-term memory (CLSTM) networks can recursively process 3D data and extract more abstract and distinctive features, but at the cost of higher computational complexity. To this end, Seydgar et al. [94] have proposed a two-stage approach combining CNN and CLSTM. The first stage utilizes a 3D CNN to extract low-dimensional shallow spectral-spatial features from HSI with relatively little spatial information, while the second stage applies CLSTM for the first time to recursively analyze the spatial information while taking into account the spectral information. This recursive analysis enhances the robustness of the model in extracting patches at different spatial sizes, and the strategy of using 3D CNN before applying CLSTM effectively reduces the computational burden. Ranjan et al. [95] have also proposed a novel CLSTM classifier for dealing with the complex spatial and spectral correlations inherent in HSI data. The classifier performs well in modeling the multidimensional relationships of the data and is effective in capturing long-term dependencies, thus improving the accuracy of feature extraction and classification.

Although LSTM and GRU are highly effective for time-series data, their application in HSIC still faces notable challenges. Many existing methods treat each pixel in an HSI as a sequence, with the spectral

bands forming the sequence elements. This results in significantly long input sequences, which increases the risk of overfitting. Moreover, handling such lengthy sequences leads to greater computational demands and slower training processes. As a result, there is a pressing need to explore parallel processing techniques to enhance the generalization capability and efficiency of RNN-based models in HSIC.

#### 4.4. Autoencoder (AE)

AE is a symmetric neural network widely used in HSIC, mainly for unsupervised feature learning. Its core goal is to generate a compressed feature representation of high-dimensional HSI data, rather than directly performing classification. As shown in Fig. 8, the AE consists of an input layer, an encoding layer (hidden layer), a decoding layer (reconstruction layer), and an output layer, which is learned by minimizing the difference between the input and the reconstructed output. In addition, AE is also used as an unsupervised DR technique, which is widely used in image preprocessing, feature extraction and image classification. Stacked auto-encoders (SAEs) learn more abstract features by stacking multiple AE layers, while denoising auto-encoders (DAEs) enhance the model's ability to handle noisy data by adding noise to the input. These methods effectively deal with high-dimensional data and noise and are suitable for a wide range of HSI data analysis tasks.

Zhao et al. [96] have proposed a method that combines SAE with a three-dimensional deep residual network (3DDRN). The key idea is to first reduce spectral redundancy using SAE through unsupervised dimensionality reduction. Then, 3DDRN is used to extract joint spectral-spatial features by leveraging its residual learning capability. The hierarchical design helps reduce overfitting while preserving important information. In addition, to address parameter instability caused by noise and limited training data, Yuvaraj et al. [97] have introduced an improved SAE combined with a restricted boltzmann machine (RBM). It has applied data augmentation with different noise levels to enhance category separation. The approach aims to utilize the SAE for robust feature extraction while incorporating the RBM to perform effective classification and denoising. This combination enhances the model's generalization performance, especially in the presence of noise.

To enhance the compatibility between learned features and classification objectives, the Deep Margin Cosine Autoencoder (DMCA) [98] introduces a margin-based constraint during training. This constraint encourages the autoencoder to form a more discriminative feature space suited for softmax-based classification. The framework adopts a two-stage training process, where general feature representations are initially learned and subsequently fine-tuned for classification tasks. Building upon this approach, the Two-stage Multidimensional Convolutional SAE (TMC-SAE) [99] separates spectral and spatial processing

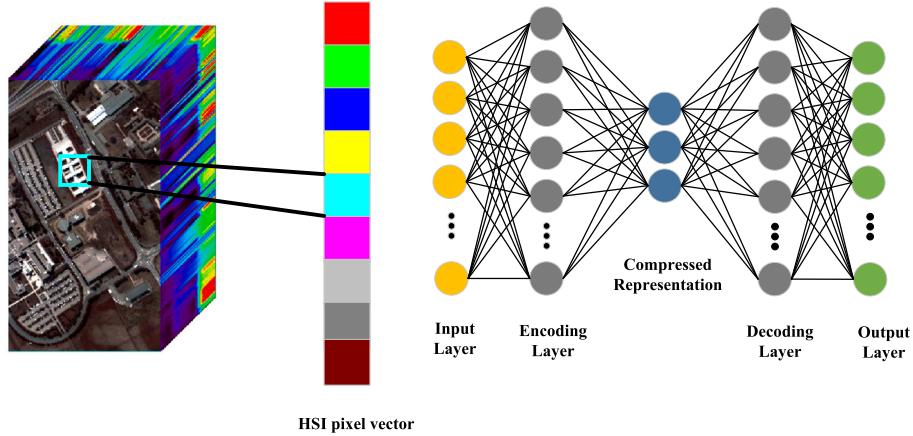


Fig. 8. The general structure of the HSIC based on AE.

into distinct phases. The first stage, SAE-1, focuses on reducing spectral redundancy through 1D convolutions and fully connected layers. The second stage, SAE-2, leverages a combination of 2D and 3D convolutional operations to extract joint spectral-spatial features. This modular design not only enhances model interpretability but also contributes to more efficient training.

Addressing the scarcity of labeled data and the need for improved feature generalization, Ranjan et al. [100] have proposed a semi-supervised framework. It combines 3D convolutional autoencoders (3D-CAE), twin networks, and attention mechanisms. This model uses unlabeled data to pretrain the 3D-CAE. The twin network encourages better separation between classes. An attention module helps the model focus on more discriminative features.

Beyond traditional AE frameworks, recent studies have adapted the masked autoencoder (MAE) approach for HSIC. The Diffusion-Enhanced Masked Autoencoder (DEMAE) [101] is designed to capture long-range spectral dependencies. It improves robustness by incorporating auxiliary tasks such as denoising and reconstruction. By leveraging noise-aware training alongside a customized loss function designed to improve the signal-to-noise ratio, the model facilitates more effective feature learning. In a parallel effort to disentangle spectral and spatial information under a self-supervised paradigm, the Spatial-Spectral MAE (SS-MAE) [102] employs separate masking and reconstruction branches for spatial patches and spectral bands. This architecture enables the model to more effectively exploit both spatial structures and spectral signatures in HSI data. Wang et al. [103] have followed a similar approach. They have introduced a unified MAE framework (HSIMAE) with spatial-spectral encoders and a two-branch fine-tuning strategy. This allows the model to effectively adapt to unlabeled data in the target domain.

AE shows great potential in HSI processing but still faces certain challenges. The extremely high dimensionality of HSI data increases the risk of overfitting. Therefore, how to effectively perform feature selection and DR remains a key issue. In addition, due to the problems of high intra-class variability and high inter-class similarity, there is an urgent need to further explore methods like pre-training, co-training, and adaptive neural networks to enhance the framework of self-coder-based HSIC.

#### 4.5. Transformer

The Transformer architecture, introduced by Vaswani et al. [104] in 2017, has since emerged as a powerful deep learning model widely applied across various tasks, including image processing, machine translation, and text generation. The core encoder-decoder architecture of Transformer is shown in Fig. 9. The encoder and decoder are

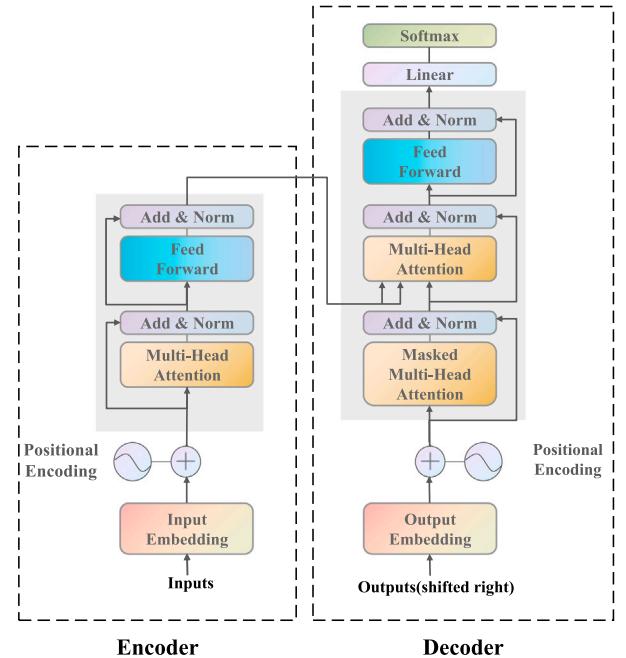


Fig. 9. Encoder-decoder architecture in Transformer.

responsible for processing input sequences and generating output sequences, respectively. Each layer in the encoder and decoder contains the same sub-layers, including the self-attention mechanism and the feed-forward network. This architecture not only helps to fully understand the input sequences, but also generates context-rich output sequences.

The Transformer architecture has several advantages in HSIC. The Transformer architecture can effectively capture the dependency between spectral bands and spatial locations in HSI. By focusing on both spectral and spatial dimensions, the model is able to learn to extract discriminative features and fully utilize the rich information in HSI data [105]. HSI often contains long-range dependencies. The spectral features of one pixel may depend not only on nearby pixels, but also on those that are far apart in both space and spectrum. The self-attention mechanism in the Transformer architecture helps capture these complex relationships. It allows the model to learn from distant pixel interactions, which leads to more accurate predictions [106]. Additionally, the Transformer can create contextualized representations

for each pixel. By analyzing the entire image, it learns to encode information about surrounding pixels and their categories. This improves the model's ability to classify hyperspectral data more accurately.

In order to capture the long-range relationships between sequence spectra in HSI, He et al. [107] have proposed a new classification framework called Spatial-Spectral Transformer (SST). The model first utilizes CNNs to extract spatial features and introduces an improved Transformer structure, called DenseTransformer, to efficiently capture the relationships between sequence spectra. Finally, a multilayer perceptron is used for the classification task. Sun et al. [108] have proposed a spectral-spatial feature tokenization transformer (SSFTT) method for capturing spectral-spatial features as well as high-level semantic features. The method extracts shallow spectral and spatial features through a spectral-spatial feature extraction module consisting of 3D convolutional layers and 2D convolutional layers, then transforms the features using a Gaussian-weighted feature tokenizer, then learns the high-level semantic representations of the features through a Transformer coding module, and finally outputs the classification results through a linear classifier.

Qin et al. [109] have presented a contrastive learning framework in a self-supervised setting, where separate Transformer modules are used to capture spatial and spectral representations. By incorporating spatial-spectral enhancement and entropy-based pseudo-labeling, the framework effectively reduces dependence on labeled data while improving domain-specific feature learning. Similarly, Dang et al. [110] have employed a dual-branch structure, in which one branch focuses on spectral information and the other on spatial features. Attention-based modules are then used to fuse the outputs, reflecting a broader trend in Transformer-based HSI classification that emphasizes the separation and adaptive integration of spectral and spatial cues. The inclusion of dual attention mechanisms further underscores the significance of maintaining a balance between these complementary feature types.

The Vision Transformer has achieved better classification performance than CNNs by using self-attention to capture global relationships in the data. The morphological Transformer model, morphFormer [111], integrates learnable spectral and spatial morphology networks to enhance HSIC. By combining morphological convolutions with attention mechanisms, the architecture facilitates stronger interactions between hyperspectral tokens and the classification token. The design enables the network to more effectively capture structural and shape-related information within the data, ultimately improving classification performance. In an effort to enhance spectral-spatial interaction, the Double-Attention Transformer Encoder (DATE) [112] was introduced. The model incorporates two self-attention modules: the spectral attention module (SPE) and the spatial attention module (SPA). The SPE is designed to capture global spectral relationships, while the SPA focuses on local spatial features. The combination of these modules enables effective fusion of spatial and spectral information for improved classification. Liu et al. [113] have used Swin Transformer for HSIC. Swin applies hierarchical windows and shifted attention. This allows it to capture both local details and global context. To further enrich feature diversity, Tan et al. [114] have proposed ELViT to further enrich features. ELViT leverages CNN-based token generators to produce multiscale semantic tokens, strengthening local feature representation. Additionally, a linear-complexity attention mechanism is employed to efficiently model long-range dependencies.

A common limitation in many Transformer-based HSIC methods is the reliance on average pooling, which often leads to the loss of fine-grained information. To overcome this issue, Ahmad et al. [55] have introduced WaveFormer. The model incorporates wavelet transform for reversible downsampling, preserving data integrity while enabling effective attention learning. By combining downsampling with wavelet transform, WaveFormer decompresses feature maps without loss, achieving a balance between accuracy and efficiency. Additionally, wavelet decomposition enhances the interaction between structural and shape information across image blocks and channels. In

addressing another challenge of fixed positional encodings, which limit flexibility for variable-length inputs, Ahmad et al. [115] have introduced SSFormer. The model utilizes implicit conditional positional encodings (CPEs) that adjust according to the local neighborhood structure, improving context capture and increasing classification robustness. Furthermore, cross-attention and learnable embeddings are integrated to enhance feature discrimination.

Transformer offers significant benefits in HSIC, including a deep understanding of the global context, adaptive feature learning, and the ability to efficiently process spectral-spatial information. However, in order to continue to realize Transformer's potential in HSIC, concerns such as computational requirements, limited data availability, and model interpretability must be addressed.

#### 4.6. Diffusion models

Diffusion modeling is a DL method based on probabilistic generation that gradually transforms data into noise by simulating the physical diffusion process. The core of the model lies in learning the inverse process to gradually recover the original data from the noise, thus achieving high-quality generation results [116,117]. The denoising process is formulated as an iterative optimization procedure, optimized using Langevin dynamics [118]. The diffusion model introduces many degrees of freedom, with noise predicting additional information at each time step. This allows the model to implicitly capture both high-level and low-level visual features, enhancing generalization and modeling complex spectral-spatial relationships [119,120]. Currently, research on diffusion models focuses on three main forms: denoising diffusion probabilistic models (DDPMs) [121], score-based generative models [119], and stochastic differential equations [122]. Fig. 10 illustrates the workflow of a manually selected, single-timestep, single-stage diffusion feature extraction method for HSIC. The horizontal arrows indicate the sequential denoising process from  $x_T$  to  $x_0$ . At a specific timestep, a feature extractor is employed to obtain diffusion features, which are then directly fed into the classifier for final categorization.

Zhou et al. [123] have proposed Diff-HSI, an unsupervised spectral-spatial feature learning framework based on a diffusion model. The framework uses unlabeled HSI image blocks to pre-train the diffusion model. Intermediate-level features are then extracted from different time steps. To make better use of these time-step features, they have designed a time-step feature library and a dynamic feature fusion module. This approach allows the model to adaptively learn multi-level information representations from different time steps. Sigger et al. [124] have proposed DiffSpectralNet, a novel method combining diffusion modeling and transformer techniques for HSIC. The method extracts both high-level and low-level spectral-spatial features using diffusion modeling. A pre-trained denoising U-Net classifies intermediate-level features, with a Transformer used for final supervised classification. Zhang et al. [125] have introduced DKDMN, a data- and knowledge-driven deep multiview fusion network. DKDMN uses a diffusion model-based knowledge learning framework (DMKL) to extract knowledge from unlabeled data. This knowledge combines with a deep multi-view network architecture (DMNA) to drive feature extraction through unsupervised diffusion knowledge.

The SpectralDiff [126] leverages iterative denoising and construction to generate data, enhancing the capture of relationships between samples. SpectralDiff integrates a spectral-spatial diffusion module and an attention-based classification module. These modules adaptively construct sample relationships, capture spectral-spatial distributions, and utilize contextual information to improve classification performance. However, previous approaches extract diffusion features from only a single time step and a single phase of the denoised U-Net, with their selection manually determined through extensive dataset-specific experimentation. The reliance on limited temporal and structural information often results in the loss of critical spectral-spatial features,

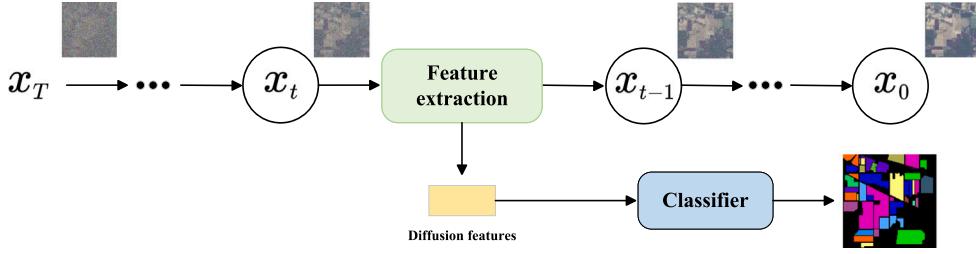


Fig. 10. Single-timestep, single-stage diffusion feature extraction.

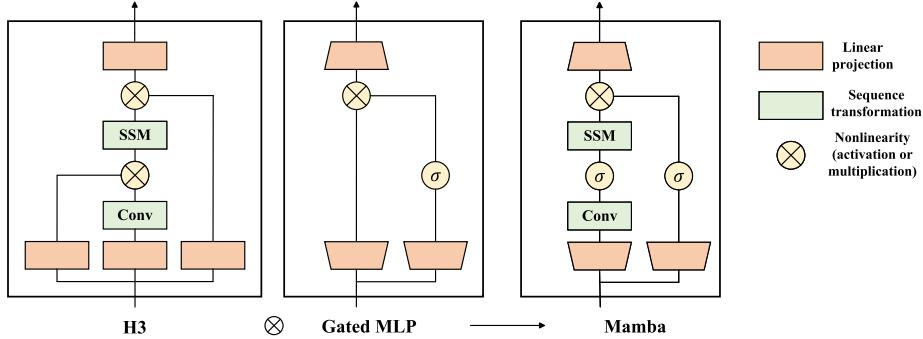


Fig. 11. Mamba module - Simplified block that integrates H3 and MLP modules.

hindering the model's capacity to effectively capture complex relationships. Addressing this limitation, the Multi-Timestep Multi-Stage Diffusion (MTMSD) framework [127] is the first to explore the integration of multi-timestep and multi-stage diffusion features for HSIC. By progressively propagating features through different temporal steps and structural stages, the framework effectively captures complex spectral-spatial dependencies, thereby enhancing classification performance. Specifically, MTMSD first uses a diffusion model through unlabeled HSI blocks for pre-training to learn diffusion features and uncover the spectral-spatial dependencies in the data. Then, multi-time-step and multi-stage diffusion features are extracted from the pre-trained denoised U-Net decoder, and time-step-level central and global feature libraries are constructed using center extraction and average pooling methods.

Effectively utilizing the rich multi-timestep and multi-stage features extracted from various stages of denoising U-Net is a key challenge. Multi-stage features, compared to single-stage ones, provide richer semantic and reconstruction information, which helps model spectral-spatial relationships more accurately. However, the enhancement in classification performance varies across datasets. Different sensors capture HSI with distinct spectral-spatial features, leading to variations in their representations. Moreover, different regions of the same HSI may emphasize different textural and semantic details, affecting the preference for time-step selection. Additionally, the large number of multi-stage features can lead to high memory consumption and computational load, especially when training large-scale datasets. Thus, managing and utilizing these diverse features efficiently remains a critical challenge to improve algorithm performance.

#### 4.7. Mamba models

The Mamba model, introduced by Gu et al. [128], is a selectively structured state-space model that excels in long sequence modeling tasks. Unlike CNNs, Mamba overcomes their modeling limitations by leveraging global receptive fields and dynamic weighting, offering advanced modeling capabilities similar to Transformers. However, it

avoids the secondary computational complexity typically associated with Transformer models, making it a more efficient alternative.

Mamba is a simplified state-space model (SSM) architecture, as illustrated in Fig. 11. Unlike the typical SSM architecture, which stacks linear attention-like blocks and multilayer perceptron (MLP) blocks as seen in Transformers, Mamba combines and evenly stacks these two components. In contrast to the H3 block, Mamba replaces the first multiplication gate with an activation function. Additionally, compared to the MLP block, Mamba introduces an SSM into the main path. The overall structure of Mamba consists of multiple repeated Mamba blocks, interspersed with standard normalization layers (such as batch normalization or layer normalization) and residual connections. Mamba retains the linear scalability of state-space models for sequence lengths, while also incorporating the powerful modeling capabilities of Transformers.

The spectral-spatial Mamba (SS-Mamba) model [129] introduces a structured approach to HSIC by integrating a token generation module with stacked spectral-spatial Mamba blocks (SS-MBs). The architecture begins by converting HSI cubes into spatial and spectral tokens. These tokens are then passed through SS-MBs, with each block containing two Mamba modules and a feature enhancement component. SS-MBs fuse information by modulating spatial and spectral tokens with data from the central region of the HSI sample. Building on a similar foundation, SpectralMamba [130] operates in a two-stage manner. Initially, spatial-spectral dynamic masks are learned via efficient convolutions to suppress spectral variability and inter-class confusion. In the subsequent stage, hidden-state representations are constructed by learning parameterized selective responses, thus eliminating the need for redundant attention mechanisms. A segmented scanning strategy is introduced to compress continuous spectral sequences, maintaining both short- and long-range contextual information while reducing computational overhead. MambaHSI [131] adopts a dual-branch design to enhance both spatial and spectral representations. The spatial Mamba block (SpaMB) enables the extraction of long-range dependencies across the entire image, while the spectral Mamba block (SpeMB) partitions spectral vectors into groups to capture intra-group relationships and derive refined spectral features. This coordinated design

facilitates the simultaneous modeling of global spatial interactions and detailed spectral dependencies.

He et al. [132] have proposed a novel 3D-spectral-spatial mamba (3DSS-Mamba) framework for HSIC to model spectral-spatial relationships with improved computational efficiency. A spectral-spatial token generation (SSTG) module first transforms the HSI cube into 3D spectral-spatial tokens. To process these tokens, a 3D spectral-spatial selective scanning (3DSS) mechanism is employed, enabling pixel-wise scanning across both spectral and spatial dimensions. By integrating this scanning mechanism with conventional mapping operations, the model constructs 3D spectral-spatial Mamba blocks (3DMBs), enhancing representation capacity while maintaining computational efficiency. Ahmad et al. [133] have proposed the Spatial-Spectral Morphological Mamba (MorpMamba) model. Initially, a token generation module transforms HSI blocks into spatial-spectral tokens. These tokens are processed through morphological operations and depth-wise separable convolutions to capture structural characteristics. A feature enhancement module further refines the representations. To improve discriminative power, the model incorporates a multi-head self-attention mechanism in the feature space. Finally, the aggregated features are passed through a state-space block to perform classification and generate the corresponding truth maps.

The Mamba model overcomes the limitations of traditional CNNs by using a global sensing field and dynamic weighting mechanism. This gives it advanced modeling capabilities, similar to Transformers. Mamba reduces computational costs and speeds up training and inference. This makes it ideal for real-time monitoring and large-scale image processing, such as in environmental monitoring, agricultural analysis, and urban planning. Mamba-based HSIC is a cutting-edge technology. It can be enhanced by combining it with other DL models to better extract spectral-spatial features in HSI.

## 5. Strategies for handling limited labeled data

### 5.1. Data augmentation

Data augmentation helps create more training data by applying transformations like rotation, translation, and scaling. This approach addresses the issues of limited data and overfitting. The transformed sample  $y_i$  corresponding to the input sample  $x_i$  is expressed as follows:

$$y = f(x_i) \quad (4)$$

where  $f$  is the transformation function. In HSIC research, many frameworks use data augmentation methods. These methods help improve classification accuracy and reduce overfitting caused by limited training data.

For example, Yu et al. [134] have improved the training data by applying three data enhancement operations (flip, rotate, and pan) and have utilized this enhanced dataset to train CNNs for HSIC. Inspired by the recently proposed principal component analysis-based data augmentation (PCA-DA) and Superpixelwise PCA (SuperPCA), Gao et al. [135] have introduced SuperPCA-DA in the model. Specifically, the HSI is first divided into blocks of hyperpixels, and then local reconstruction is performed using image segmentation techniques for denoising. Then, SuperPCA-DA is applied in each hyperpixel for data enhancement, where low-dimensional features are extracted by local PCA and projected back to the high-dimensional spectral space, and random noise is added to the projection matrix. Tan et al. [136] have proposed a data enhancement method based on improved deep convolutional generative adversarial network (DCGAN) to extend the spectral data of cantaloupe under different pesticide residues. Zhu et al. [137] have also proposed a conditional deep convolutional generative adversarial network (cDCGAN) model. This model generates labeled HSI samples to enhance the training data.

### 5.2. Unsupervised/semi-supervised learning

While supervised DL models have made significant progress on HSIC, the current demand favors unsupervised or semi-supervised DL models. Since HSI data lacks labeled training samples, unsupervised/semi-supervised methods can utilize unlabeled data samples to aid training.

Unsupervised methods utilize an encoder-decoder structure to train the network without labeling the samples. Ghasroodashti et al. [138] have proposed a method for HSIC by improving the traditional autoencoder and combining it with Majorization Minimization (MM)-based technology. They construct the weights of the autoencoder using the similarity angle map (SAM) criterion as a regularization term. To improve classification accuracy, they use a fuzzy mode to estimate the parameters. Sellars et al. [139] have proposed a novel graph-based semi-supervised framework. The framework uses a hyperpixel approach to define meaningful local regions in the HSI that are likely to share the same classification labels. Spectral and spatial features are extracted from these regions, and a contracted weighted graph representation is constructed, where each node represents a region rather than a single pixel. This graph is then fed into a graph-based semi-supervised classifier to obtain the final classification results. Chen et al. [140] have proposed an unsupervised multivariate feature fusion network ( $M^3FuNet$ ) for HSIC. The method extracts spectral and spatial features using multiscale supervector matrix correction (MSMC) and multiscale random convolution dispersion (MRCD), which significantly improves feature retention and spatial-spectral dependence. In addition, unsupervised HSIC based on data clustering attracts increasing attention. Liu et al. [141] have combined entropy rate superpixel segmentation (ERS), superpixel-based PCA, and PCA-domain 2D singular spectral analysis (SSA) to improve the efficiency and effectiveness of feature extraction, and use anchor point-based graph clustering (AGC) for efficient classification.

### 5.3. Generative adversarial network(GAN)

The GAN model is proposed by Goodfellow et al. [142], which contains two sub-models: generator and discriminator. Fig. 12 illustrates how the GAN model works. The generator  $G$  captures the underlying distribution  $p_d$  of real data and learns a new distribution  $p_g$  for generating fake data, with the goal of making  $p_g$  similar to  $p_d$ . The discriminator  $D$  computes the distance between  $p_g$  and  $p_d$  to determine whether the input sample comes from real data.  $G$  and  $D$  are trained synchronously, with their parameters updated alternately.

Zhu et al. [143] have first explored the effectiveness and usefulness of GAN in HSIC. They design two CNNs, one CNN for distinguishing input data and the other for generating pseudo-inputs. To address the challenge of limited labeled samples, the study introduces two complementary classification strategies: a one-dimensional GAN (1D-GAN) focused on spectral feature learning, and a three-dimensional GAN (3D-GAN) designed to capture spectral-spatial features more robustly. These approaches demonstrated the feasibility of using GANs to enhance feature representation and classification performance in HSIC tasks. Xue et al. [144] have proposed a semi-supervised convolutional GAN classification model, which includes a 3D-CNN-based generator network and a 3D deep residual-based discriminator network. The generated samples, labeled samples, and unlabeled samples are fed into the discriminator together for joint training. The trained discriminator is able to recognize both the authenticity of the samples and their category labels. Wang et al. [145] have proposed a new discriminator that combines CapsNet and CLSTM. This structure extracts low-level features and combines them with local spatial sequence information to form high-level context features.

Yu et al. [146] have utilized a novel GAN, Sill-RGAN, to enhance the classification performance of HSI under different lighting conditions. Sill-RGAN enhances the robustness of classification by handling varying

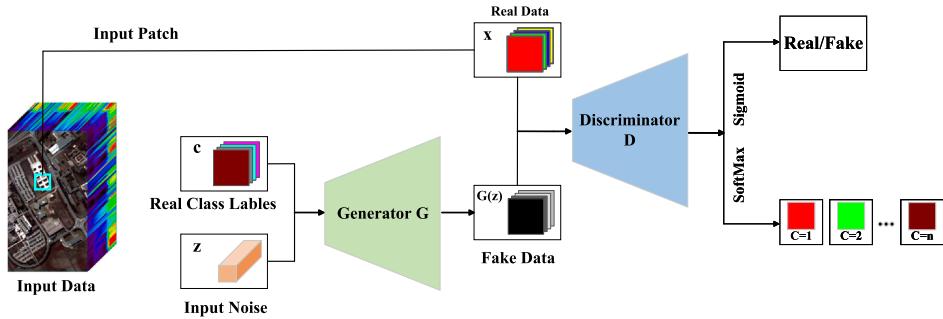


Fig. 12. HSIC based on GAN.

lighting conditions through a unique mapping method and generates new virtual samples to support model training. To address the scarcity of labeled samples, a combination of kernel PCA for dimensionality reduction, CycleGAN for sample augmentation, and EfficientNet for classification has been explored in [147]. Hao et al. [148] have proposed the Transformer with Residual Upscale GAN (TRUG) model, which addresses the challenge of limited labeled samples and demonstrates strong performance in spectral information processing. TRUG consists of a generator  $G$  and a discriminator  $D$ . Residual Upscale (RU) is used in  $G$  to improve the resolution of generated features, extract texture features, and capture context.

#### 5.4. Transfer learning

Transfer learning typically involves a source domain  $D_s$  and a target domain  $D_t$ . The source domain  $D_s = \{x_i, y_i\}_i^{N_s}$  and the target domain  $D_t = \{x_i, y_i\}_i^{N_t}$ , where  $x_i$  and  $y_i$  represent the data samples and their corresponding class labels. The goal of transfer learning is to use the knowledge from the source domain  $D_s$  and source task  $T_s$  to enhance the learning of the prediction function  $f_t(\cdot)$  for the target domain  $D_t$  and target task  $T_t$ . In HSIC, transfer learning uses weights from a pre-trained network to initialize deep learning models. This approach is particularly effective when dealing with tasks with a limited number of samples.

In [149], transfer learning is integrated with Wasserstein GAN (WGAN-GP), where the fully trained discriminator is subsequently repurposed as a feature extractor. This approach enables the model to capture spectral-spatial features without the need for labeled data. To address scenarios with extremely limited supervision, such as having only one labeled sample per class, [150] segments the image into homogeneous regions to augment training data and employs a deep Siamese CNN to minimize the domain gap. For heterogeneous transfer learning in crop classification, [151] adapts both 2D-CNN and 3D-CNN architectures using isomorphic transfer strategies on HSI datasets.

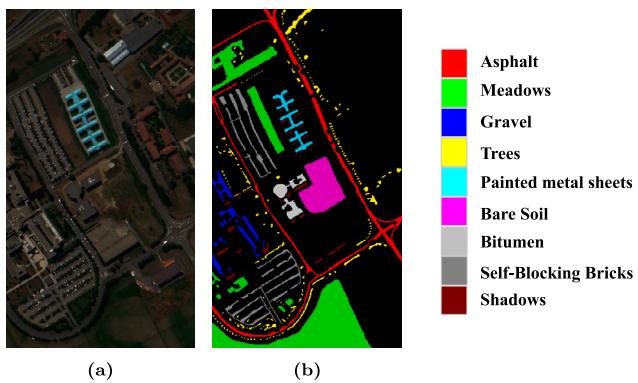
Zhou et al. [152] have combined spectral-feature extraction with meta-transfer learning. The model is first trained on a richly labeled source-domain dataset and then transferred to a sparsely labeled target domain. The feature extraction capability is enhanced using dense connection blocks and three-dimensional convolutional residual connections to maximize spatial and spectral information. Su et al. [153] have proposed the adaptive cuckoo refinement-based graph transfer network (ACGT-Net) method. A GCN is first pre-trained to learn transferable weight parameters. Then, a cuckoo search strategy (CSS) is integrated into the trained GCN to adaptively refine the graph structure. In addition, Hong et al. [154] have introduced SpectralGPT, an innovative generalized remote sensing base model specialized for HSI processing with an advanced 3D generation pre-trained Transformer.

#### 5.5. Active learning

Active learning (AL) progressively reduces the training sample size by selecting the most representative instances from unlabeled samples using uncertainty measures and structure. The goal is to increase the amount of information while reducing redundancy and accelerating the learning process with fewer samples [155]. In each cycle, AL selects the most valuable samples from the unlabeled data, which are then assigned labels by an “oracle” (either human or machine). These valuable samples are then added to the training set and the model is retrained on the updated dataset. This process is repeated until a certain termination condition is met, which may be the size of the training dataset, the number of iterations, or the desired classification result.

As shown in [156], DL has been combined with AL for HSIC. First, a CNN model is trained using a small set of labeled pixels. Then, AL selects the most representative samples from the unlabeled pixels for annotation. Finally, the CNN is fine-tuned using the newly labeled samples to improve performance. In [157], training sample selection is addressed through two strategies. Active learning (AL) is employed to iteratively expand the labeled dataset by incorporating informative unlabeled samples. In parallel, the iterative training sampling (ITS) method enriches the data cube by embedding spatial categorization information into the training set, thereby enhancing the representativeness of the samples. In [158], an active deep learning framework is introduced, incorporating an auxiliary network to estimate the uncertainty of unlabeled samples. It fuses features from both the original data and the base learner’s hidden layers, which are then processed by a fully connected network. A customized loss function guides the uncertainty prediction for more effective sample selection.

In [159], an AL-based spectral-spatial classification (ALSSC) model has been proposed for tree species classification. To reduce the demand for training samples, after each round of classification, AL is used to select the most informative samples from the unlabeled test set to enrich the training set. However, statistical methods have limited effectiveness, and their performance depends on specific scenarios. Therefore, Patel et al. [160] have proposed a reinforced pool-based deep AL (RPDAL) method. This method designs and trains a reinforcement learning (RL)-based agent for selecting informative samples for labeling. After training, this RL-based agent is able to migrate to other HSI datasets and efficiently select samples to be labeled. Some AL algorithms usually tend to select unlabeled samples near category boundaries, which are difficult to distinguish. However, during the learning process, samples in the center of the category are often hard to identify accurately. This issue may lead to inaccurate category metrics and affect the effectiveness of sample selection. To address this problem, Ding et al. [161] have proposed a new AL method called PLGCN, which combines prototype learning (PL) and GCN. In this method, the prototype of each category is iteratively updated at each sampling stage to ensure its optimality.



**Fig. 13.** Pavia University dataset. (a) False color. (b) Ground truth.

**Table 2**

The number of training, validation and test samples of the Pavia University dataset.

Class No.	Class name	Train Num	Valid Num	Test Num
1	Asphalt	66	66	6499
2	Meadows	186	187	18276
3	Gravel	21	21	2057
4	Trees	30	31	3003
5	Painted metal sheets	14	13	1318
6	Bare soil	50	50	4929
7	Bitumen	13	14	1303
8	Self-Blocking Bricks	37	37	3608
9	Shadows	10	9	928
Total		427	428	41921

## 5.6. Few-shot learning

Few-shot learning (FSL) addresses the challenge of insufficient training samples by being able to recognize and classify new categories using only a small number of samples when training the model. HSIC is performed by analyzing spectral information to assign each pixel to a predefined category. In the HSIC task, due to the variety and complexity of categories, models trained using FSL are able to capture latent patterns in the spectral data, making them better adapted to new categories.

There have been many studies incorporating FSL in HSIC. Liu et al. [162] have proposed a deep sample-less learning method to solve the problem of sample shortage in HSIC. The method first extracts spectral-spatial features using a deep residual 3D-CNN to reduce labeling uncertainty. Subsequently, a metric space is constructed through staged training, which clusters similar samples and separates dissimilar ones. Finally, a nearest neighbor classifier is used to classify the test samples in this metric space. Wang et al. [163] have proposed a heterogeneous FSL (HFSL) method, which is designed to use only a small number of labeled samples per category for classification tasks.

However, HFSL also has some problems, such as poor generalization from natural to HSI and unstable prototype due to limited labeled samples. To solve these problems, Zhang et al. [164] have proposed a mutual information enhancement FSL (MIEFSL) method. The method contains three main modules: mutual information enhancement (MIE), intradomain prototype rectification (IPR), and interdomain distribution alignment (IDA). Xi et al. [165] have proposed an FSL framework for HSIC with a class-covariance metric (CMFSL). CMFSL learns the global class representation by alternating training samples of the base class and the new class in each training round and applies a synthesis strategy to the new class to prevent overfitting.

## 6. Experimental evaluation

In this study, in order to ensure the fairness and comparability of the experimental results, a standardized experimental setup was used to

Table 3

The number of training, validation and test samples of the Indian Pines dataset.

The number of training, validation and test samples of the Indian Pine dataset				
Class No.	Class name	Train Num	Valid Num	Test Num
1	Alfalfa	3	2	41
2	Corn-notill	71	72	1285
3	Corn-mintill	42	41	747
4	Corn	12	12	213
5	Grass-pasture	24	24	435
6	Grass-trees	37	36	657
7	Grass-pasture-mowed	2	1	25
8	Hay-windrowed	24	24	430
9	Oats	1	1	18
10	Soybean-notill	49	48	875
11	Soybean-mintill	122	123	2210
12	Soybean-clean	30	29	534
13	Wheat	10	10	185
14	Woods	63	63	1139
15	Buildings-Grass-Trees-Drives	19	20	347
16	Stone-Steel-Towers	4	5	84
Total		513	511	9225

Table 4

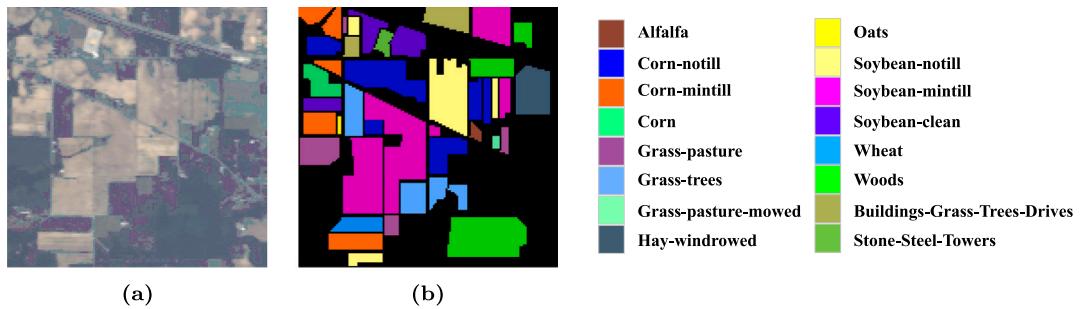
The number of training, validation, and test samples of the Houston2013 dataset.

The number of training, validation, and test samples of the Houston2010 dataset.				
Class No.	Class name	Train Num	Valid Num	Test Num
1	Healthy grass	63	62	1126
2	Stressed grass	63	62	1129
3	Synthetic grass	35	35	627
4	Trees	62	62	1120
5	Soil	62	62	1118
6	Water	16	17	292
7	Residential	63	64	1141
8	Commercial	62	62	1120
9	Road	63	62	1127
10	Highway	61	62	1104
11	Railway	61	62	1112
12	Parking Lot 1	61	62	1110
13	Parking Lot 2	24	23	422
14	Tennis Court	22	21	385
15	Running Track	33	33	594
Total		751	751	13527

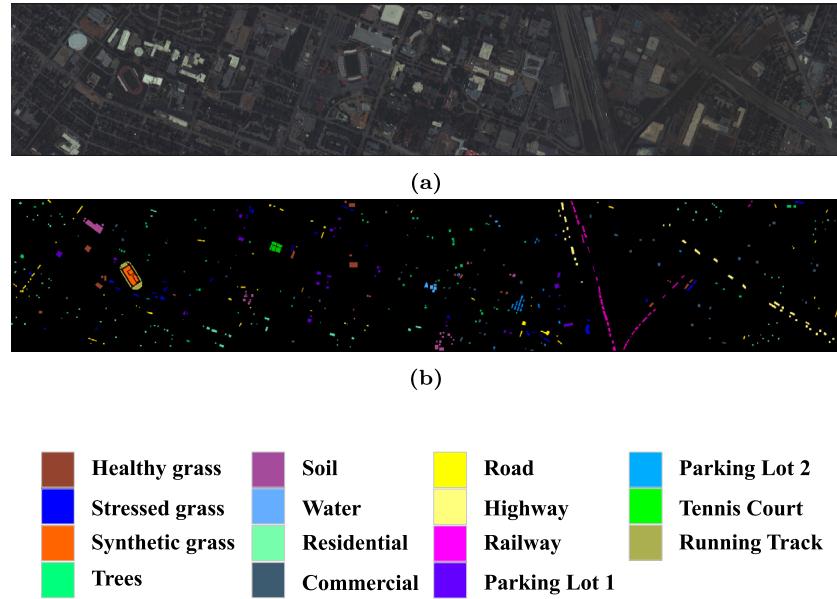
unify the way of selecting the training, validation, and testing samples. Overall classification accuracy (OA), average classification accuracy (AA) and kappa coefficient (Kappa) were utilized as evaluation metrics. We ensured that all models were experimented on the same sample set, guaranteeing consistency in geographic location and the number of training samples in each round of cross-validation. In addition, the overlap between the training and test samples was specifically considered. A stratified sampling delineation was used in this study to ensure that there was no overlap between the training and test samples, thus eliminating any possible bias due to sample overlap and ensuring the integrity of the evaluation process.

### 6.1. Experimental datasets

In this paper, we employed the Pavia University (PU), Indiana Pines (IP), and Houston2013 datasets for training and testing. Fig. 13(a) and (b) respectively display the false-color composite images and corresponding ground truth maps of the PU dataset. The dataset comprises 42,776 labeled samples and is partitioned into a training set 1%, a validation set 1%, and a test set 98%, as presented in Table 2. Fig. 14(a) and (b) respectively present the false-color composite image and the corresponding ground truth map of the IP dataset. The dataset contains 10,249 labeled samples and the data is divided into 5% training set, 5% validation set and 90% test set as shown in Table 3. Fig. 15(a) and (b) present the false-color composite images and their corresponding ground truth maps of the Houston2013 dataset. This dataset contains 15,029 labeled samples, which are divided into 5% for the training set, 5% for the validation set, and 90% for the test set, as shown in Table 4.



**Fig. 14.** Indian Pines dataset. (a) False color. (b) Ground truth.



**Fig. 15.** Houston2013 dataset. (a) False color. (b) Ground truth.

**Table 5**  
Experimental results of different methods on Pavia University dataset.

	3D CNN	SPRN	AMGCFN	CEGCN	SSFTT	MorphFormer	SS-Mamba	3DSS-Mamba	CL-MGNet
1	95.69	98.36	97.78	98.79	98.55	98.02	96.60	99.78	<b>100</b>
2	98.75	99.93	99.89	<b>99.95</b>	99.82	99.77	94.34	99.86	99.83
3	77.37	95.62	95.40	<b>100.00</b>	98.38	94.85	99.65	98.09	99.75
4	98.43	95.23	79.36	94.49	98.43	96.54	98.81	98.04	<b>99</b>
5	99.55	98.44	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	98.84	<b>100</b>
6	84.81	99.84	98.94	99.67	<b>99.68</b>	99.54	99.03	98.83	<b>99.92</b>
7	70.53	85.26	80.83	<b>100</b>	95.49	92.48	99.96	97.74	99.84
8	90.39	93.69	97.93	99.86	97.23	97.31	98.69	98.97	<b>100</b>
9	88.38	92.92	73.20	<b>100.00</b>	95.14	95.46	99.98	95.77	97.89
OA	93.76	97.93	96.41	99.34	99.02	98.48	98.94	99.24	<b>99.78</b>
AA	89.32	95.48	91.48	99.19	98.11	97.11	99.35	98.44	<b>99.57</b>
Kappa	91.67	97.26	95.23	99.13	98.7	97.98	98.60	99.00	<b>99.71</b>

## *6.2. Analysis and interpretation of experimental results*

To further analyze the characteristics of different models, we select several representative methods for comparison. These include 3D CNN [166] and SPRN [75], which are based on CNN; AMGCFN [167] and CEGCN [168], which are based on GCN; SSFTT [108] and MorphFormer [111], which are based on Transformer; SS-Mamba [129] and 3DSS-Mamba [132], which are based on Mamba; and CL-MGNet [169], based on class-level band learning. The models are evaluated on three benchmark HSI datasets — Pavia University, Indian Pines, and Houston2013. The classification results are shown in Tables 5, 6, and 7, as well as Figs. 16, 17, and 18.

CL-MGNet outperforms other methods in Overall Accuracy (OA), Average Accuracy (AA), and Kappa coefficient across all three datasets. It shows a particularly strong performance when the training samples are limited. In the Pavia University dataset, CL-MGNet significantly exceeds other methods in classification accuracy. It also successfully handles the challenge of distinguishing between similar categories, such as grassland and bare soil. Additionally, as shown in [Table 5](#), the classification accuracy for category 9 fluctuates greatly with other methods, ranging from 73.20% to 100%. However, CL-MGNet consistently maintains high performance. This is due to its class-level spectral band learning strategy, which uses multi-scale convolution and attention

**Table 6**

Experimental results of different methods on Indian Pines dataset.

	3D CNN	SPRN	AMGCFN	CEGCN	SSFTT	MorphFormer	SS-Mamba	3DSS-Mamba	CL-MGNet
1	78.26	84.78	90.91	97.56	95.65	95.65	99.61	97.56	<b>100</b>
2	89.29	98.39	92.84	<b>99.29</b>	98.67	95.66	78.46	92.92	98.52
3	88.92	96.39	96.86	99.72	99.28	98.79	89.08	96.79	<b>100</b>
4	87.76	98.73	98.68	98.57	99.16	98.73	<b>99.4</b>	95.77	97.52
5	92.96	98.14	71.89	98.36	<b>99.59</b>	90.89	93.32	97.24	99.27
6	<b>99.86</b>	98.63	92.89	99.84	96.03	99.04	98.14	98.33	99.84
7	<b>100</b>	78.57	<b>100</b>	<b>100</b>	<b>100</b>	82.14	<b>100</b>	68.00	<b>100</b>
8	98.54	<b>100</b>	<b>100</b>	<b>100</b>	99.79	<b>100</b>	<b>100</b>	99.77	<b>100</b>
9	55.00	95.00	73.68	77.78	65.00	40.00	<b>100</b>	77.78	<b>100</b>
10	90.02	92.39	90.62	97.23	95.88	94.24	88.23	95.77	<b>99.52</b>
11	95.64	97.11	97.20	<b>99.86</b>	99.47	99.06	90.08	96.83	99.76
12	88.53	91.74	85.79	97.18	94.09	93.42	92.68	88.95	<b>99.8</b>
13	<b>100</b>	97.56	<b>100</b>	97.28	98.05	99.51	99.94	98.92	98.86
14	99.76	99.68	99.18	99.91	99.92	98.49	98.35	98.6	<b>100</b>
15	91.71	98.70	<b>100</b>	95.86	93.26	99.48	97.54	96.83	<b>100</b>
16	97.85	67.74	<b>100</b>	98.78	<b>100</b>	79.57	99.72	94.05	97.5
OA	93.72	96.81	94.32	99.01	98.18	97.02	97.89	98.09	<b>99.53</b>
AA	90.88	93.35	93.16	97.32	95.88	91.53	96.86	95.38	<b>99.41</b>
Kappa	92.82	96.36	93.51	98.87	97.93	96.60	97.04	97.54	<b>99.46</b>

**Table 7**

Experimental results of different methods on Houston2013 dataset.

	3D CNN	SPRN	AMGCFN	CEGCN	SSFTT	MorphFormer	SS-Mamba	3DSS-Mamba	CL-MGNet
1	95.60	97.44	93.78	<b>100</b>	97.68	99.92	93.55	99.47	<b>100.00</b>
2	97.76	98.01	99.75	97.45	99.68	99.60	95.61	99.65	<b>100.00</b>
3	<b>100</b>	98.85	99.70	97.19	99.28	99.71	<b>100.00</b>	99.84	<b>100.00</b>
4	99.12	92.92	99.08	98.34	99.44	98.47	99.11	98.21	<b>99.72</b>
5	99.84	99.76	99.92	<b>100</b>	<b>100</b>	<b>100</b>	98.24	99.46	<b>100</b>
6	88.61	90.77	96.81	96.20	<b>98.15</b>	93.23	95.9	97.26	97.83
7	94.32	97.00	98.35	98.30	97.24	98.58	90.98	93.08	<b>100</b>
8	88.99	93.09	85.79	84.20	97.11	97.27	83.99	97.50	<b>98.87</b>
9	92.33	95.45	93.69	<b>99.59</b>	98.08	96.96	92.85	96.72	99.25
10	88.75	<b>100</b>	<b>100</b>	<b>100</b>	99.18	98.78	97.46	99.91	<b>100</b>
11	94.49	91.98	93.22	94.73	99.51	99.76	92.9	97.57	<b>100</b>
12	97.49	97.32	98.56	92.90	96.92	97.65	90.98	<b>99.28</b>	98.95
13	92.11	92.75	91.57	91.43	99.57	98.93	90.75	98.10	<b>100</b>
14	97.89	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
15	99.54	97.19	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
OA	95.18	96.45	96.51	96.68	98.66	98.73	95.47	98.28	<b>99.69</b>
AA	95.12	96.36	96.68	96.69	98.79	98.59	95.82	98.40	<b>99.64</b>
Kappa	94.79	96.17	96.22	96.41	98.55	98.62	95.1	98.14	<b>99.66</b>

mechanisms to capture subtle feature differences in high-dimensional HSI data.

In Table 6, due to the severe sample imbalance in the Indian Pines dataset, the comparison of categories 1, 7, and 9 reveals that most methods tend to overfit in these categories. Notably, only CL-MGNet effectively alleviates this issue. This is especially evident in Category 9, where the performance of other models fluctuates drastically, ranging from 40% to 99.52%. However, CL-MGNet consistently maintains relatively stable high performance.

3DSS-Mamba successfully captures the complex dependencies of spectral-spatial data by adopting an innovative state-space approach. It also demonstrates relatively stable performance across the three datasets. In contrast, traditional CNN models such as 3D CNN and SPRN perform relatively weaker. Its symmetric convolutional structure leads to a rapid increase in parameters, which adds to the complexity of model training. Methods based on Transformer and GCN show moderate performance, with relatively stable results on some datasets, but lack the excellent generalization ability seen in CL-MGNet and 3DSS-Mamba.

The strong performance of CL-MGNet and 3DSS-Mamba reflects progress in algorithm performance. It also points to a key trend in HSIC research. This trend focuses on using innovative network architectures and learning strategies. These approaches help better understand and utilize the complex features of the high-dimensional spectral data. It offers valuable technical guidance and theoretical insights for future HSIC research.

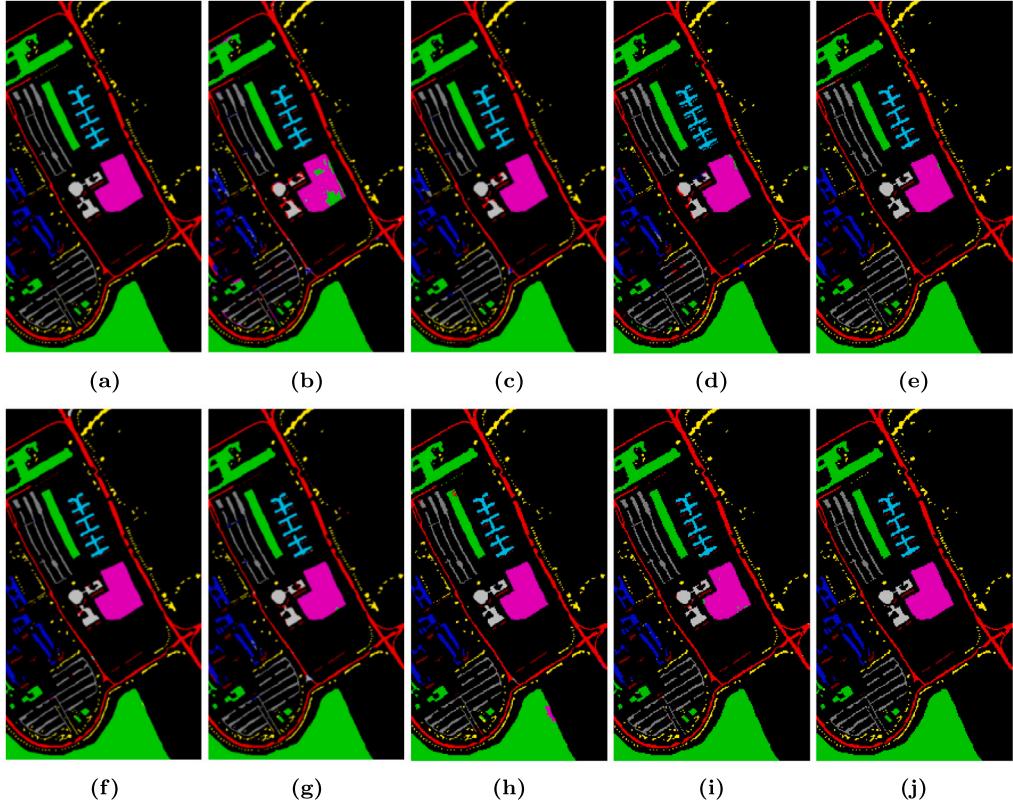
## 7. Conclusion and future work

### 7.1. Conclusion

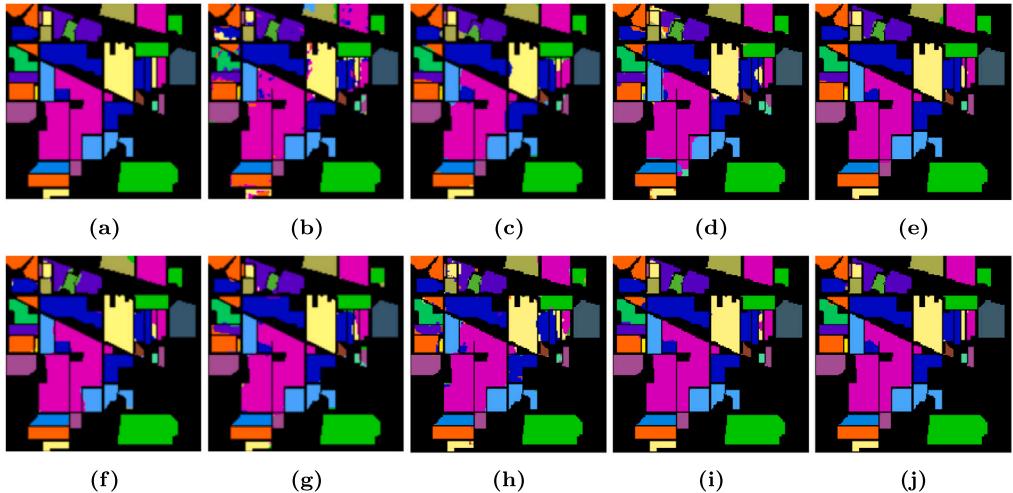
Compared with traditional manual feature extraction methods, DL-based HSIC models have attracted wide attention due to their high efficiency in many fields. In this review, we first introduce the importance of HSIC, the characteristics of HSI data, and the challenges faced by DL in this area. Next, we explore different spectral, spatial, and spectral-spatial representations for efficient processing of HSI data. We then review various DL classification methods, including CNN, GCN, RNN, AE, Transformer, diffusion model, and Mamba model, and analyze their advantages and limitations. In addition, we also discuss the strategy to solve the problem of limited labeled data. Finally, we conducted experiments on multiple DL models on three widely used benchmark datasets aimed at evaluating their performance in HSIC tasks, thus providing a valuable reference and guidance for future research.

### 7.2. Future work

Although these DL methods have made remarkable progress in classification accuracy, they still face difficulties in improving the computational efficiency of the model, processing high-dimensional data ability, and coping with the challenges of complex scenarios, so further research is needed to improve the generalization ability of the model.



**Fig. 16.** Pavia University dataset classification maps by various methods. (a) ground truth, (b) 3D CNN, (c) SPRN, (d) AMGCFN, (e) CEGCN, (f) SSFTT, (g) MorphFormer, (h) SS-Mamba, (i) 3DSS-Mamba, (j) CL-MGNet.

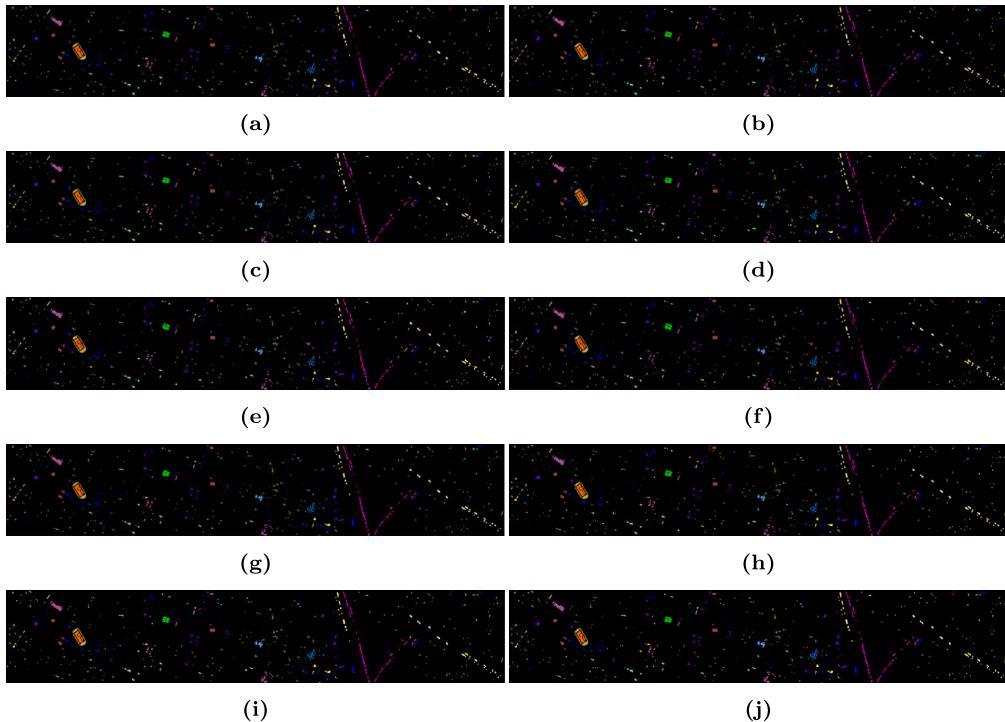


**Fig. 17.** Indian Pines dataset classification maps by various methods. (a) ground truth, (b) 3D CNN, (c) SPRN, (d) AMGCFN, (e) CEGCN, (f) SSFTT, (g) MorphFormer, (h) SS-Mamba, (i) 3DSS-Mamba, (j) CL-MGNet.

- Limited labeled data: Due to the scarcity and limited availability of labeled data in HSI, Section 5 explores a variety of solutions, and we need to combine these methods to fully utilize their combined advantages. In addition, unsupervised learning and FSL have become a hot topic in current research because they can accurately predict category labels with only a small number of labeled samples.
- DL-based model: In terms of model architecture, DL-based diffusion model and Mamba model will become the main direction of future development. These models have significant advantages in feature expression and can effectively mine the complex spatial

and spectral information in HSI. The diffusion model can better deal with the nonlinear relationship and global dependence in the image, so as to improve the robustness of the model in complex scenes. The Mamba model is expected to improve the recognition ability of HSI fine-grained features through its unique network structure.

- Cross-domain transfer Learning: The study of cross-domain transfer learning in HSIC will focus on optimizing domain adaptation techniques to cope with distribution differences between different hyperspectral datasets. By designing more efficient adaptive methods, we can reduce the need for labeling in the target domain and improve the generalization ability of the model in



**Fig. 18.** Houston2013 dataset classification maps by various methods. (a) ground truth, (b) 3D CNN , (c) SPRN, (d) AMGCFN, (e) CEGCN, (f) SSFTT, (g) MorphFormer, (h) SS-Mamba, (i) 3DSS-Mamba, (j) CL-MGNet.

different environments. At the same time, research on how to combine domain specific knowledge with transfer learning and develop customized transfer learning framework to enhance the performance of models in complex scenarios will be an important direction in the future.

- Multimodal data fusion: In HSIC, combining HSI data with other remote sensing data, such as LiDAR, radar imagery or visible imagery, can provide richer feature information and improve the accuracy and robustness of classification to compensate for the inadequacy of a single data source. Future research will focus on developing advanced data fusion techniques to effectively deal with the heterogeneity of different data sources, especially in large-scale and real-time applications, while optimizing the computational efficiency to meet the practical needs.
- Domain knowledge integration: In the future, effective integration of domain knowledge will play an important role in HSIC. Through cooperation with experts in agriculture, environmental monitoring, and other domains, model architectures and feature extraction methods can be customized to improve classification accuracy. Meanwhile, the development of model architectures that effectively integrate domain knowledge and utilize specific sensor data and physical properties will help to improve model performance. In addition, feature engineering based on domain knowledge can also reduce the dependence on a large amount of labeled data and further improve the efficiency and accuracy of HSIC, thus promoting the application of HSIC technology in multiple fields.
- Computational efficiency and scalability: HSIC still faces high computational costs and long training times. Future research can explore lightweight model architectures, such as sparse neural networks, quantization techniques, and model pruning, to reduce resource consumption. Integrating distributed and cloud computing may enhance the practicality of HSIC for large-scale data processing. Additionally, hardware acceleration and specialized chip optimization could offer new possibilities for efficiently processing HSI data.

- Potential and optimization of large models: In the future, HSIC will benefit from the development of large models like SpectralGPT. These models have strong feature learning and generalization capabilities, allowing them to capture complex patterns in HSI data more effectively. Meanwhile, optimizing their computational efficiency, reducing inference time, and maintaining high performance with limited resources will be key research directions. Additionally, the previously mentioned technological advances in unsupervised learning and pre-training of HSI data will further unlock the potential of large models for HSIC tasks.

#### CRediT authorship contribution statement

**Yongchao Song:** Writing – review & editing, Writing – original draft, Methodology, Funding acquisition, Conceptualization. **Junhao Zhang:** Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Zhaowei Liu:** Supervision, Investigation, Funding acquisition. **Yang Xu:** Software, Resources, Investigation, Formal analysis. **Siwen Quan:** Writing – review & editing, Software, Formal analysis. **Lijun Sun:** Visualization, Methodology, Data curation. **Jiping Bi:** Writing – original draft, Software, Data curation. **Xuan Wang:** Writing – review & editing, Writing – original draft, Resources, Project administration, Conceptualization.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

This research was funded by the Natural Science Foundation of Shandong Province, China (ZR2022QF037), in part by the Shandong Province Science and Technology-based Small and Medium-sized Enterprises Innovation Capacity Enhancement project (2024TSGC0016), and in part by the School and Locality Integration Development Project of Yantai City (2022).

## Data availability

Data will be made available on request.

## References

- [1] D. Landgrebe, Hyperspectral image data analysis, *IEEE Signal Process. Mag.* 19 (1) (2002) 17–28.
- [2] D. Hong, W. He, N. Yokoya, J. Yao, L. Gao, L. Zhang, J. Chanussot, X. Zhu, Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing, *IEEE Geosci. Remote. Sens. Mag.* 9 (2) (2021) 52–87.
- [3] H. Ye, H. Shi, X. Wang, E. Sun, C. Li, Y. An, S. Wu, W. Xiong, Z. Li, J. Landgraf, Improving atmospheric CO<sub>2</sub> retrieval based on the collaborative use of greenhouse gases monitoring instrument and directional polarimetric camera sensors on Chinese hyperspectral satellite GF5-02, *Geo-Spat. Inf. Sci.* 27 (3) (2024) 572–584.
- [4] C. Xiong, H. Tao, S. Liu, G. Liu, Z. Wen, Y. Shang, Q. Wang, C. Fang, S. Li, K. Song, Using satellite imagery to estimate CO<sub>2</sub> partial pressure and exchange with the atmosphere in the Songhua river, *J. Hydrol.* 634 (2024) 131074.
- [5] A. Nisha, A. Anitha, Current advances in hyperspectral remote sensing in urban planning, in: 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies, ICICICT, IEEE, 2022, pp. 94–98.
- [6] L. Shuai, Z. Li, Z. Chen, D. Luo, J. Mu, A research review on deep learning combined with hyperspectral imaging in multiscale agricultural sensing, *Comput. Electron. Agric.* 217 (2024) 108577.
- [7] A. Faltynkova, C.E. Deschênes, A. Zolich, M. Wagner, T.A. Johansen, G. Johnsen, Use of an uncrewed surface vehicle and near infrared hyperspectral imaging for sampling and analysis of aquatic microplastics, *Marine Poll. Bull.* 201 (2024) 116214.
- [8] F.-C. Lin, Y.-S. Shiu, P.-J. Wang, U.-H. Wang, J.-S. Lai, Y.-C. Chuang, A model for forest type identification and forest regeneration monitoring based on deep learning and hyperspectral imagery, *Ecol. Inform.* 80 (2024) 102507.
- [9] Y. Pan, H. Zhang, Y. Chen, X. Gong, J. Yan, H. Zhang, Applications of hyperspectral imaging technology combined with machine learning in quality control of traditional Chinese medicine from the perspective of artificial intelligence: a review, *Crit. Rev. Anal. Chem.* 54 (8) (2024) 2850–2864.
- [10] D.-W. Sun, H. Pu, J. Yu, Applications of hyperspectral imaging technology in the food industry, *Nat. Rev. Electr. Eng.* 1 (4) (2024) 251–263.
- [11] B. Liu, Z. Liu, S. Men, Y. Li, Z. Ding, J. He, Z. Zhao, Underwater hyperspectral imaging technology and its applications for detecting and mapping the seafloor: A review, *Sensors* 20 (17) (2020) 4962.
- [12] J. Cai, S. Ding, Q. Zhang, R. Liu, D. Zeng, L. Zhou, Broken ice circumferential crack estimation via image techniques, *Ocean Eng.* 259 (2022) 111735.
- [13] L. Zhou, J. Cai, S. Ding, The identification of ice floes and calculation of sea ice concentration based on a deep learning method, *Remote. Sens.* 15 (10) (2023) 2663.
- [14] Y. Su, Z. Zhu, L. Gao, A. Plaza, P. Li, X. Sun, X. Xu, DAAN: A deep autoencoder-based augmented network for blind multilinear hyperspectral unmixing, *IEEE Trans. Geosci. Remote Sens.* (2024).
- [15] S. Liu, D. Marinelli, L. Bruzzone, F. Bovolo, A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges, *IEEE Geosci. Remote. Sens. Mag.* 7 (2) (2019) 140–158.
- [16] J. Qu, W. Dong, Y. Yang, T. Zhang, Y. Li, Q. Du, Cycle-refined multidecision joint alignment network for unsupervised domain adaptive hyperspectral change detection, *IEEE Trans. Neural Netw. Learn. Syst.* (2024).
- [17] F. Ullah, I. Ullah, R.U. Khan, S. Khan, K. Khan, G. Pau, Conventional to deep ensemble methods for hyperspectral image classification: A comprehensive survey, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* (2024).
- [18] X. Wang, Q. Hu, Y. Cheng, J. Ma, Hyperspectral image super-resolution meets deep learning: A survey and perspective, *IEEE/CAA J. Autom. Sin.* 10 (8) (2023) 1668–1691.
- [19] H. Fu, Z. Ling, G. Sun, J. Ren, A. Zhang, L. Zhang, X. Jia, HyperDehazing: A hyperspectral image dehazing benchmark dataset and a deep learning model for haze removal, *ISPRS J. Photogramm. Remote Sens.* 218 (2024) 663–677.
- [20] Q. Zhang, Y. Zheng, Q. Yuan, M. Song, H. Yu, Y. Xiao, Hyperspectral image denoising: From model-driven, data-driven, to model-data-driven, *IEEE Trans. Neural Netw. Learn. Syst.* (2023).
- [21] S. Kahraman, R. Bacher, A comprehensive review of hyperspectral data fusion with lidar and sar data, *Annu. Rev. Control.* 51 (2021) 236–253.
- [22] S. Hongjun, Dimensionality reduction for hyperspectral remote sensing: Advances, challenges, and prospects, *Natl. Remote. Sens. Bull.* 26 (8) (2022) 1504–1529.
- [23] R. Grewal, S.S. Kasana, G. Kasana, Hyperspectral image segmentation: a comprehensive survey, *Multimedia Tools Appl.* 82 (14) (2023) 20819–20872.
- [24] F. Melgani, L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, *IEEE Trans. Geosci. Remote Sens.* 42 (8) (2004) 1778–1790.
- [25] F. Li, H. Lu, P. Zhang, An innovative multi-kernel learning algorithm for hyperspectral classification, *Comput. Electr. Eng.* 79 (2019) 106456.
- [26] G. Liu, L. Wang, D. Liu, L. Fei, J. Yang, Hyperspectral image classification based on non-parallel support vector machine, *Remote. Sens.* 14 (10) (2022) 2447.
- [27] J. Ham, Y. Chen, M.M. Crawford, J. Ghosh, Investigation of the random forest framework for classification of hyperspectral data, *IEEE Trans. Geosci. Remote Sens.* 43 (3) (2005) 492–501.
- [28] W. Liu, J.E. Fowler, C. Zhao, Spatial logistic regression for support-vector classification of hyperspectral imagery, *IEEE Geosci. Remote. Sens. Lett.* 14 (3) (2017) 439–443.
- [29] H. Yu, L. Gao, J. Li, S.S. Li, B. Zhang, J.A. Benediktsson, Spectral-spatial hyperspectral image classification using subspace-based support vector machines and adaptive Markov random fields, *Remote. Sens.* 8 (4) (2016) 355.
- [30] Y. Chen, Z. Lin, X. Zhao, G. Wang, Y. Gu, Deep learning-based classification of hyperspectral data, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* 7 (6) (2014) 2094–2107.
- [31] H. Abdi, L.J. Williams, Principal component analysis, *Wiley Interdiscip. Rev.: Comput. Stat.* 2 (4) (2010) 433–459.
- [32] M. Zhang, Z. Lei, L. Liu, K. Ma, R. Shang, L. Jiao, Efficient evolutionary multi-scale spectral-spatial attention fusion network for hyperspectral image classification, *Expert Syst. Appl.* 262 (2025) 125672.
- [33] R. Shang, J. Yang, J. Feng, Y. Li, S. Xu, Hyperspectral image classification based on ConvGRU and spectral-spatial joint attention, *Appl. Soft Comput.* (2025) 112949.
- [34] K. Li, Y. Wan, A. Ma, Y. Zhong, A lightweight multi-scale and multi-attention hyperspectral image classification network based on multi-stage search, *IEEE Trans. Geosci. Remote Sens.* (2025).
- [35] Q. Wu, M. He, Q. Chen, L. Sun, C. Ma, Integrating multi-scale spatial-spectral shuffling convolution with 3D lightweight transformer for hyperspectral image classification, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* (2025).
- [36] P. Sevugan, V. Rudhrakoti, T.-h. Kim, M. Gunasekaran, S. Purushotham, R. Chinthaginjala, I. Ahmad, A. Kumar, Class-aware feature attention-based semantic segmentation on hyperspectral images, *PLoS One* 20 (2) (2025) e0309997.
- [37] H. Mahdipour, A. Sharifi, M. Soohak, C.R. Medrano, Ultrafusion: Optimal fuzzy fusion in land-cover segmentation using multiple panchromatic satellite images, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* 17 (2024) 5721–5733.
- [38] A. Sharifi, M.M. Safari, Enhancing the spatial resolution of sentinel-2 images through super-resolution using transformer-based deep learning models, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* (2025).
- [39] C. Wang, B. Liu, L. Liu, Y. Zhu, J. Hou, P. Liu, X. Li, A review of deep learning used in the hyperspectral image analysis for agriculture, *Artif. Intell. Rev.* 54 (7) (2021) 5205–5253.
- [40] S.M. MirhoseiniNejad, D. Abbasi-Moghadam, A. Sharifi, ConvLSTM-ViT: A deep neural network for crop yield prediction using earth observations and remotely sensed data, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* (2024).
- [41] A. Vafaeinejad, N. Alimohammadi, A. Sharifi, M.M. Safari, Super-resolution AI-based approach for extracting agricultural cadastral maps: Form and content validation, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* (2025).
- [42] R. Rajabi, A. Zehtabian, K.D. Singh, A. Tabatabaeenejad, P. Ghamisi, S. Homayouni, Hyperspectral imaging in environmental monitoring and analysis, 2024.
- [43] M.M. Safari, A. Sharifi, J. Mahmood, D. Abbasi-Moghadam, Mesoscale eddy detection and classification from sea surface temperature maps with deep neural networks, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* (2024).
- [44] S. Marzvan, K. Moravej, S. Felegari, A. Sharifi, M.S. Askari, Risk assessment of alien Azolla filiculoides lam in Anzali lagoon using remote sensing imagery, *J. Indian Soc. Remote. Sens.* 49 (2021) 1801–1809.
- [45] S. Hajaj, A. El Harti, A.B. Pour, A. Jellouli, Z. Adiri, M. Hashim, A review on hyperspectral imagery application for lithological mapping and mineral prospecting: Machine learning techniques and future prospects, *Remote. Sens. Appl.: Soc. Environ.* (2024) 101218.
- [46] G. Cai, X. Zheng, J. Guo, W. Gao, Real-time identification of borehole rescue environment situation in underground disaster areas based on multi-source heterogeneous data fusion, *Saf. Sci.* 181 (2025) 106690.
- [47] Z. Zhang, L. Huang, Q. Wang, L. Jiang, Y. Qi, S. Wang, T. Shen, B.-H. Tang, Y. Gu, UAV hyperspectral remote sensing image classification: A systematic review, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* (2024).
- [48] M.E. Paoletti, J.M. Haut, J. Plaza, A. Plaza, Deep learning classifiers for hyperspectral imaging: A review, *ISPRS J. Photogramm. Remote Sens.* 158 (2019) 279–317.
- [49] M. Ahmad, A.M. Khan, M. Mazzara, S. Distefano, S.K. Roy, X. Wu, Hybrid dense network with attention mechanism for hyperspectral image classification, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* 15 (2022) 3948–3957.
- [50] F. Ji, W. Zhao, Q. Wang, W.J. Emery, R. Peng, Y. Man, G. Wang, K. Jia, Spectral-spatial evidential learning network for open-set hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* (2024).
- [51] M. Ahmad, U. Ghous, D. Hong, A.M. Khan, J. Yao, S. Wang, J. Chanussot, A disjoint samples-based 3D-CNN with active transfer learning for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 60 (2022) 1–16.

- [52] L. Liang, S. Zhang, J. Li, Multiscale DenseNet meets with bi-RNN for hyperspectral image classification, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* 15 (2022) 5401–5415.
- [53] G. Jaiswal, R. Rani, H. Mangotra, A. Sharma, Integration of hyperspectral imaging and autoencoders: Benefits, applications, hyperparameter tuning and challenges, *Comput. Sci. Rev.* 50 (2023) 100584.
- [54] X. Cao, Y. Lian, K. Wang, C. Ma, X. Xu, Unsupervised hybrid network of transformer and CNN for blind hyperspectral and multispectral image fusion, *IEEE Trans. Geosci. Remote Sens.* (2024).
- [55] M. Ahmad, U. Ghous, M. Usama, M. Mazzara, WaveFormer: Spectral-spatial wavelet transformer for hyperspectral image classification, *IEEE Geosci. Remote. Sens. Lett.* (2024).
- [56] Y. Liu, J. Yue, S. Xia, P. Ghamisi, W. Xie, L. Fang, Diffusion models meet remote sensing: Principles, methods, and perspectives, 2024, arXiv preprint arXiv:2404.08926.
- [57] W. Zhou, S.-i. Kamata, H. Wang, M.S. Wong, H.C. Hou, Mamba-in-Mamba: Centralized Mamba-cross-scan in tokenized Mamba model for hyperspectral image classification, *Neurocomputing* 613 (2025) 128751.
- [58] P. Hammer, Adaptive control processes: a guided tour (R. Bellman), 1962.
- [59] G. Hughes, On the mean accuracy of statistical pattern recognizers, *IEEE Trans. Inform. Theory* 14 (1) (1968) 55–63.
- [60] V. Kumar, R.S. Singh, M. Rambabu, Y. Dua, Deep learning for hyperspectral image classification: A survey, *Comput. Sci. Rev.* 53 (2024) 100658.
- [61] H. Fu, G. Sun, L. Zhang, A. Zhang, J. Ren, X. Jia, F. Li, Three-dimensional singular spectrum analysis for precise land cover classification from UAV-borne hyperspectral benchmark datasets, *ISPRS J. Photogramm. Remote Sens.* 203 (2023) 115–134.
- [62] J. Li, X. Huang, L. Tu, WHU-OHS: A benchmark dataset for large-scale herseptical image classification, *Int. J. Appl. Earth Obs. Geoinf.* 113 (2022) 103022.
- [63] Y. Zhang, L. Liang, J. Li, A. Plaza, X. Kang, J. Mao, Y. Wang, Structural and textural-aware feature extraction for hyperspectral image classification, *IEEE Geosci. Remote. Sens. Lett.* (2024).
- [64] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, L. Zhao, Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey, *Multimedia Tools Appl.* 78 (2019) 15169–15211.
- [65] M. Sugiyama, Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis, *J. Mach. Learn. Res.* 8 (5) (2007).
- [66] H.-T. Chen, H.-W. Chang, T.-L. Liu, Local discriminant embedding and its variants, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'05, Vol. 2, IEEE, 2005, pp. 846–853.
- [67] B.-C. Kuo, D.A. Landgebre, Nonparametric weighted feature extraction for classification, *IEEE Trans. Geosci. Remote Sens.* 42 (5) (2004) 1096–1105.
- [68] Y. Fang, H. Li, Y. Ma, K. Liang, Y. Hu, S. Zhang, H. Wang, Dimensionality reduction of hyperspectral images based on robust spatial information using locally linear embedding, *IEEE Geosci. Remote. Sens. Lett.* 11 (10) (2014) 1712–1716.
- [69] D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, *Adv. Neural Inf. Process. Syst.* 13 (2000).
- [70] K. Chhapariya, K.M. Buddhiraju, A. Kumar, A deep spectral-spatial residual attention network for hyperspectral image classification, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* (2024).
- [71] X. Zhang, Y. Sun, K. Jiang, C. Li, L. Jiao, H. Zhou, Spatial sequential recurrent neural network for hyperspectral image classification, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* 11 (11) (2018) 4141–4155.
- [72] D.H. Hubel, T.N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *J. Physiol.* 160 (1) (1962) 106.
- [73] K. Fukushima, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, *Biol. Cybernet.* 36 (4) (1980) 193–202.
- [74] Z. Gong, P. Zhong, Y. Yu, W. Hu, S. Li, A CNN with multiscale convolution and diversified metric for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 57 (6) (2019) 3599–3618.
- [75] X. Zhang, S. Shang, X. Tang, J. Feng, L. Jiao, Spectral partitioning residual network with spatial attention mechanism for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 60 (2021) 1–14.
- [76] S. Dong, Y. Quan, W. Feng, G. Dauphin, L. Gao, M. Xing, A pixel cluster CNN and spectral-spatial fusion algorithm for hyperspectral image classification with small-size training samples, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* 14 (2021) 4101–4114.
- [77] Z. Ye, Z. Cao, H. Liu, H. Liu, W. Li, L. Bai, X. Li, Self-supervised learning with multiscale densely-connected network for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* (2024).
- [78] F.F. Shahrai, S. Prasad, Graph convolutional neural networks for hyperspectral data classification, in: 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, 2018, pp. 968–972.
- [79] A. Qin, Z. Shang, J. Tian, Y. Wang, T. Zhang, Y.Y. Tang, Spectral-spatial graph convolutional networks for semisupervised hyperspectral image classification, *IEEE Geosci. Remote. Sens. Lett.* 16 (2) (2018) 241–245.
- [80] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, J. Chanussot, Graph convolutional networks for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 59 (7) (2020) 5966–5978.
- [81] H. Gao, H. Feng, Y. Zhang, S. Fei, R. Cheng, S. Xu, B. Zhang, Interactive enhanced network based on multihead self-attention and graph convolution for classification of hyperspectral and LiDAR data, *IEEE Trans. Geosci. Remote Sens.* (2024).
- [82] Y. Chu, J. Cao, W. Ding, J. Huang, H. Ju, H. Cao, G. Liu, Hyperspectral image classification using feature fusion fuzzy graph broad network, *Inform. Sci.* 689 (2025) 121504.
- [83] Y. Duan, F. Luo, M. Fu, Y. Niu, X. Gong, Classification via structure-preserved hypergraph convolution network for hyperspectral image, *IEEE Trans. Geosci. Remote Sens.* 61 (2023) 1–13.
- [84] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, 2017, arXiv preprint arXiv:1710.10903.
- [85] K. Xu, Y. Zhao, L. Zhang, C. Gao, H. Huang, Spectral-spatial residual graph attention network for hyperspectral image classification, *IEEE Geosci. Remote. Sens. Lett.* 19 (2021) 1–5.
- [86] Y. Ding, Z. Zhang, X. Zhao, D. Hong, W. Cai, N. Yang, B. Wang, Multi-scale receptive fields: Graph attention neural network for hyperspectral image classification, *Expert Syst. Appl.* 223 (2023) 119858.
- [87] J. Cai, M. Zhang, H. Yang, Y. He, Y. Yang, C. Shi, X. Zhao, Y. Xun, A novel graph-attention based multimodal fusion network for joint classification of hyperspectral image and LiDAR data, *Expert Syst. Appl.* 249 (2024) 123587.
- [88] P. Quan, L. Zheng, W. Zhang, Y. Xiao, L. Niu, Y. Shi, ExGAT: Context extended graph attention neural network, *Neural Netw.* 181 (2025) 106784.
- [89] R.J. Williams, D. Zipser, A learning algorithm for continually running fully recurrent neural networks, *Neural Comput.* 1 (2) (1989) 270–280.
- [90] A. Gündüz, Z. Orman, Hyperspectral image classification using a hybrid RNN-CNN with enhanced attention mechanisms, *J. Indian Soc. Remote. Sens.* (2024) 1–17.
- [91] K. Ramkumar, H.M. Al-Jawahry, N. Shilpa, G. Shalini, M. Arun, Multi-dimensional recurrent neural networks hyperspectral image classification, in: 2024 International Conference on Data Science and Network Security, ICDSNS, IEEE, 2024, pp. 1–4.
- [92] M.V. Dasu, P.V.N. Reddy, S.C.M. Reddy, Deep concatenated features with improved heuristic-based recurrent neural network for hyperspectral image classification, *Multimedia Tools Appl.* 83 (17) (2024) 49875–49904.
- [93] J. Kang, Y. Zhang, X. Liu, Z. Cheng, Hyperspectral image classification using spectral-spatial double-branch attention mechanism, *Remote. Sens.* 16 (1) (2024) 193.
- [94] M. Seygar, A. Alizadeh Naeini, M. Zhang, W. Li, M. Satari, 3-D convolution-recurrent networks for spectral-spatial classification of hyperspectral images, *Remote. Sens.* 11 (7) (2019) 883.
- [95] P. Ranjan, A. Girdhar, Ankur, R. Kumar, A novel spectral-spatial 3D auxiliary conditional GAN integrated convolutional LSTM for hyperspectral image classification, *Earth Sci. Inform.* (2024) 1–21.
- [96] J. Zhao, L. Hu, Y. Dong, L. Huang, S. Weng, D. Zhang, A combination method of stacked autoencoder and 3D deep residual network for hyperspectral image classification, *Int. J. Appl. Earth Obs. Geoinf.* 102 (2021) 102459.
- [97] N. Yuvaraj, K. Praghosh, R. Arshath Raja, S. Chidambaram, D. Shreecharan, Hyperspectral image classification using denoised stacked auto encoder-based restricted Boltzmann machine classifier, in: International Conference on Hybrid Intelligent Systems, Springer, 2022, pp. 213–221.
- [98] M. Wang, Y. Xu, Z. Wang, C. Xing, Deep margin cosine autoencoder based medical hyperspectral image classification for tumor diagnosis, *IEEE Trans. Instrum. Meas.* (2023).
- [99] Y. Bai, X. Sun, Y. Ji, W. Fu, J. Zhang, Two-stage multi-dimensional convolutional stacked autoencoder network model for hyperspectral images classification, *Multimedia Tools Appl.* 83 (8) (2024) 23489–23508.
- [100] P. Ranjan, R. Kumar, A. Girdhar, A 3D-convolutional-autoencoder embedded siamese-attention-network for classification of hyperspectral images, *Neural Comput. Appl.* 36 (15) (2024) 8335–8354.
- [101] Z. Li, Z. Xue, M. Jia, X. Nie, H. Wu, M. Zhang, H. Su, DEMA: Diffusion enhanced masked autoencoder for hyperspectral image classification with few labeled samples, *IEEE Trans. Geosci. Remote Sens.* (2024).
- [102] J. Lin, F. Gao, X. Shi, J. Dong, Q. Du, SS-MAE: Spatial-spectral masked autoencoder for multisource remote sensing image classification, *IEEE Trans. Geosci. Remote Sens.* 61 (2023) 1–14.
- [103] Y. Wang, M. Wen, H. Zhang, J. Sun, Q. Yang, Z. Zhang, H. Lu, HSIMAE: A unified masked autoencoder with large-scale pre-training for hyperspectral image classification, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* (2024).
- [104] A. Vaswani, Attention is all you need, *Adv. Neural Inf. Process. Syst.* (2017).
- [105] N. Chen, L. Fang, Y. Xia, S. Xia, H. Liu, J. Yue, Spectral query spatial: Revisiting the role of center pixel in transformer for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* (2024).
- [106] J. Zhao, J. Wang, C. Ruan, Y. Dong, L. Huang, Dual-branch spectral-spatial attention network for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* (2024).

- [107] X. He, Y. Chen, Z. Lin, Spatial-spectral transformer for hyperspectral image classification, *Remote. Sens.* 13 (3) (2021) 498.
- [108] L. Sun, G. Zhao, Y. Zheng, Z. Wu, Spectral-spatial feature tokenization transformer for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 60 (2022) 1–14.
- [109] J. Qin, H. Zhao, Spatial-spectral-associative contrastive learning for satellite hyperspectral image classification with transformers, *Remote. Sens.* 15 (6) (2023) 1612.
- [110] L. Dang, L. Weng, Y. Hou, X. Zuo, Y. Liu, Double-branch feature fusion transformer for hyperspectral image classification, *Sci. Rep.* 13 (1) (2023) 272.
- [111] S.K. Roy, A. Deria, C. Shah, J.M. Haut, Q. Du, A. Plaza, Spectral-spatial morphological attention transformer for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 61 (2023) 1–15.
- [112] P. Tang, M. Zhang, Z. Liu, R. Song, Double attention transformer for hyperspectral image classification, *IEEE Geosci. Remote. Sens. Lett.* 20 (2023) 1–5.
- [113] B. Liu, Y. Liu, W. Zhang, Y. Tian, W. Kong, Spectral swin transformer network for hyperspectral image classification, *Remote. Sens.* 15 (15) (2023) 3721.
- [114] Y. Tan, M. Li, L. Yuan, C. Shi, Y. Luo, G. Wen, Hyperspectral image classification with embedded linear vision transformer, *Earth Sci. Inform.* 18 (1) (2025) 69.
- [115] M. Ahmad, M. Usama, A.M. Khan, S. Distefano, H.A. Altuwaijri, M. Mazzara, Spatial spectral transformer with conditional position encoding for hyperspectral image classification, *IEEE Geosci. Remote. Sens. Lett.* (2024).
- [116] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, Deep unsupervised learning using nonequilibrium thermodynamics, in: International Conference on Machine Learning, PMLR, 2015, pp. 2256–2265.
- [117] A.Q. Nichol, P. Dhariwal, Improved denoising diffusion probabilistic models, in: International Conference on Machine Learning, PMLR, 2021, pp. 8162–8171.
- [118] M. Welling, Y.W. Teh, Bayesian learning via stochastic gradient langevin dynamics, in: Proceedings of the 28th International Conference on Machine Learning, ICML-11, Citeseer, 2011, pp. 681–688.
- [119] Y. Song, S. Ermon, Generative modeling by estimating gradients of the data distribution, *Adv. Neural Inf. Process. Syst.* 32 (2019).
- [120] W. Zhao, Y. Rao, Z. Liu, B. Liu, J. Zhou, J. Lu, Unleashing text-to-image diffusion models for visual perception, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 5729–5739.
- [121] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, *Adv. Neural Inf. Process. Syst.* 33 (2020) 6840–6851.
- [122] Y. Song, C. Durkan, I. Murray, S. Ermon, Maximum likelihood training of score-based diffusion models, *Adv. Neural Inf. Process. Syst.* 34 (2021) 1415–1428.
- [123] J. Zhou, J. Sheng, J. Fan, P. Ye, T. He, B. Wang, T. Chen, When hyperspectral image classification meets diffusion models: An unsupervised feature learning framework, 2023, arXiv preprint [arXiv:2306.08964](https://arxiv.org/abs/2306.08964).
- [124] N. Sigger, Q.-T. Vien, S.V. Nguyen, G. Tozzi, T.T. Nguyen, Unveiling the potential of diffusion model-based framework with transformer for hyperspectral image classification, *Sci. Rep.* 14 (1) (2024) 8438.
- [125] J. Zhang, F. Zhao, H. Liu, J. Yu, Data and knowledge-driven deep multiview fusion network based on diffusion model for hyperspectral image classification, *Expert Syst. Appl.* 249 (2024) 123796.
- [126] N. Chen, J. Yue, L. Fang, S. Xia, SpectralDiff: A generative framework for hyperspectral image classification with diffusion models, *IEEE Trans. Geosci. Remote Sens.* (2023).
- [127] J. Zhou, J. Sheng, P. Ye, J. Fan, T. He, B. Wang, T. Chen, Exploring multistep multi-stage diffusion features for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* (2024).
- [128] A. Gu, T. Dao, Mamba: Linear-time sequence modeling with selective state spaces, 2023, arXiv preprint [arXiv:2312.00752](https://arxiv.org/abs/2312.00752).
- [129] L. Huang, Y. Chen, X. He, Spectral-spatial mamba for hyperspectral image classification, 2024, arXiv preprint [arXiv:2404.18401](https://arxiv.org/abs/2404.18401).
- [130] J. Yao, D. Hong, C. Li, J. Chanussot, Spectralmamba: Efficient mamba for hyperspectral image classification, 2024, arXiv preprint [arXiv:2404.08489](https://arxiv.org/abs/2404.08489).
- [131] Y. Li, Y. Luo, L. Zhang, Z. Wang, B. Du, MambahSI: Spatial-spectral mamba for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* (2024).
- [132] Y. He, B. Tu, B. Liu, J. Li, A. Plaza, 3DSS-Mamba: 3D-spectral-spatial Mamba for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 62 (2024) 1–16.
- [133] M. Ahmad, M.H.F. Butt, M. Usama, A.M. Khan, M. Mazzara, S. Distefano, H.A. Altuwaijri, S.K. Roy, J. Chanussot, D. Hong, Spatial-spectral morphological Mamba for hyperspectral image classification, 2024, arXiv preprint [arXiv:2408.01372](https://arxiv.org/abs/2408.01372).
- [134] X. Yu, X. Wu, C. Luo, P. Ren, Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework, *GIScience Remote. Sens.* 54 (5) (2017) 741–758.
- [135] S. Gao, X. Jiang, Y. Zhang, X. Liu, Q. Xiong, Z. Cai, Superpixelwise PCA based data augmentation for hyperspectral image classification, *Multimedia Tools Appl.* (2024) 1–21.
- [136] H. Tan, Y. Hu, B. Ma, G. Yu, Y. Li, An improved DCGAN model: Data augmentation of hyperspectral image for identification pesticide residues of Hami melon, *Food Control* 157 (2024) 110168.
- [137] F. Zhu, J. Wang, P. Lv, X. Qiao, M. He, Y. He, Z. Zhao, Generating labeled samples based on improved cDCGAN for hyperspectral data augmentation: A case study of drought stress identification of strawberry leaves, *Comput. Electron. Agric.* 225 (2024) 109250.
- [138] E.K. Ghasrodashti, N. Sharma, Hyperspectral image classification using an extended auto-encoder method, *Signal Process., Image Commun.* 92 (2021) 116111.
- [139] P. Sellars, A.I. Aviles-Rivero, C.-B. Schönlieb, Superpixel contracted graph-based learning for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 58 (6) (2020) 4180–4193.
- [140] H. Chen, H. Long, T. Chen, Y. Song, H. Chen, X. Zhou, W. Deng, M<sup>3</sup>FuNet: An unsupervised multivariate feature fusion network for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* (2024).
- [141] Q. Liu, D. Xue, Y. Tang, Y. Zhao, J. Ren, H. Sun, PSSA: PCA-domain superpixelwise singular spectral analysis for unsupervised hyperspectral image classification, *Remote. Sens.* 15 (4) (2023) 890.
- [142] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, *Adv. Neural Inf. Process. Syst.* 27 (2014).
- [143] L. Zhu, Y. Chen, P. Ghamisi, J.A. Benediktsson, Generative adversarial networks for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 56 (9) (2018) 5046–5063.
- [144] Z. Xue, Semi-supervised convolutional generative adversarial network for hyperspectral image classification, *IET Image Process.* 14 (4) (2020) 709–719.
- [145] W.-Y. Wang, H.-C. Li, Y.-J. Deng, L.-Y. Shao, X.-Q. Lu, Q. Du, Generative adversarial capsule network with ConvLSTM for hyperspectral image classification, *IEEE Geosci. Remote. Sens. Lett.* 18 (3) (2020) 523–527.
- [146] Z. Yu, W. Cui, Robust hyperspectral image classification using generative adversarial networks, *Inform. Sci.* 666 (2024) 120452.
- [147] T.S. Reddy, S. Anusha, E.V.V. Reddy, V.V. Shivamani, V. Mrudula, Hyperspectral image classification based on cycle GAN and EfficientNet, in: 2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies, ICAECT, IEEE, 2024, pp. 1–6.
- [148] S. Hao, Y. Xia, Y. Ye, Generative adversarial network with transformer for hyperspectral image classification, *IEEE Geosci. Remote. Sens. Lett.* (2023).
- [149] Q. Sun, S. Bourennane, Hyperspectral image classification with unsupervised feature extraction, *Remote. Sens. Lett.* 11 (5) (2020) 475–484.
- [150] M. Shabaz, M. Soni, Cognitive digital modelling for hyperspectral image classification using transfer learning model, *Turk. J. Electr. Eng. Comput. Sci.* 31 (6) (2023) 1039–1060.
- [151] U. Patel, M. Pathan, P. Kathiria, V. Patel, Crop type classification with hyperspectral images using deep learning: a transfer learning approach, *Model. Earth Syst. Environ.* 9 (2) (2023) 1977–1987.
- [152] H. Zhou, X. Wang, K. Xia, Y. Ma, G. Yuan, Transfer learning-based hyperspectral image classification using residual dense connection networks, *Sensors* 24 (9) (2024) 2664.
- [153] Y. Su, J. Chen, L. Gao, A. Plaza, M. Jiang, X. Xu, X. Sun, P. Li, ACGT-Net: Adaptive cuckoo refinement-based graph transfer network for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* (2023).
- [154] D. Hong, B. Zhang, X. Li, Y. Li, C. Li, J. Yao, N. Yokoya, H. Li, P. Ghamisi, X. Jia, et al., SpectralGPT: Spectral remote sensing foundation model, *IEEE Trans. Pattern Anal. Mach. Intell.* (2024).
- [155] E. Yanik, T.M. Sezgin, Active scene learning, 2019, arXiv preprint [arXiv:1903.02832](https://arxiv.org/abs/1903.02832).
- [156] X. Cao, J. Yao, Z. Xu, D. Meng, Hyperspectral image classification with convolutional neural network and active learning, *IEEE Trans. Geosci. Remote Sens.* 58 (7) (2020) 4604–4616.
- [157] K.Y. Ma, C.-I. Chang, Iterative training sampling coupled with active learning for semisupervised spectral-spatial hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 59 (10) (2021) 8672–8692.
- [158] Z. Lei, Y. Zeng, P. Liu, X. Su, Active deep learning for hyperspectral image classification with uncertainty learning, *IEEE Geosci. Remote. Sens. Lett.* 19 (2021) 1–5.
- [159] F. Tong, Y. Zhang, Active learning-based spectral-spatial classification for discriminating tree species in hyperspectral images, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* (2024).
- [160] U. Patel, V. Patel, Active learning-based hyperspectral image classification: A reinforcement learning approach, *J. Supercomput.* 80 (2) (2024) 2461–2486.
- [161] C. Ding, M. Zheng, S. Zheng, Y. Xu, L. Zhang, W. Wei, Y. Zhang, Integrating prototype learning with graph convolution network for effective active hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* (2024).
- [162] B. Liu, X. Yu, A. Yu, P. Zhang, G. Wan, R. Wang, Deep few-shot learning for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 57 (4) (2018) 2290–2304.
- [163] Y. Wang, M. Liu, Y. Yang, Z. Li, Q. Du, Y. Chen, F. Li, H. Yang, Heterogeneous few-shot learning for hyperspectral image classification, *IEEE Geosci. Remote. Sens. Lett.* 19 (2021) 1–5.
- [164] Q. Zhang, J. Peng, W. Sun, Q. Liu, Few-shot learning with mutual information enhancement for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 62 (2024) 1–14.

- [165] B. Xi, J. Li, Y. Li, R. Song, D. Hong, J. Chanussot, Few-shot learning with class-covariance metric for hyperspectral image classification, *IEEE Trans. Image Process.* 31 (2022) 5079–5092.
- [166] A.B. Hamida, A. Benoit, P. Lambert, C.B. Amar, 3-D deep learning approach for remote sensing image classification, *IEEE Trans. Geosci. Remote Sens.* 56 (8) (2018) 4420–4434.
- [167] H. Zhou, F. Luo, H. Zhuang, Z. Weng, X. Gong, Z. Lin, Attention multihop graph and multiscale convolutional fusion network for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 61 (2023) 1–14.
- [168] Q. Liu, L. Xiao, J. Yang, Z. Wei, CNN-enhanced graph convolutional network with pixel-and superpixel-level feature fusion for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* 59 (10) (2020) 8657–8671.
- [169] Y. Fu, H. Liu, Y. Zou, S. Wang, Z. Li, D. Zheng, Category-level band learning based feature extraction for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* (2023).