Dual Domain Feature Extraction and Adaptive Spectral-Spatial Feature Fusion Network for Hyperspectral Image Classification

Ziqi Sun[1], Xiaoqing Wan[1,2,*], Yupeng He[1] and Feng Chen[1]

[1]College of Computer Science and Technology, Hengyang Normal University, Hengyang, 421002, China

[2]Hunan Provincial Key Laboratory of Intelligent Information Processing and Application, Hengyang, 421002, China

E-mail: wanxiaoqingkaixin@163.com

Abstract. Although some progress has been made in hyperspectral image (HSI) classification, it still faces many challenges due to limited training samples, insufficient fusion of spectral and spatial information, and consumption of computing resources. In order to effectively address the above problems, this paper proposes a novel combination of dual domain feature extraction and adaptive spectral-spatial feature fusion (DDFE-ASFS), which fully extracts global and local spectral-spatial features and deep high-level semantic features. Firstly, a dual domain feature extraction (DDFE) module is proposed by integrating deep CNNs, fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT), which can fully characterize local and global spectral-spatial and frequency features. Secondly, an efficient adaptive spectral-spatial fusion (EASSF) module is designed to capture the dependency between cross-views by using the attention mechanism while maintaining the consistency of spectral and spatial features. Then, two convolution layers are used to further optimize the features, and pixel-attention and residual path are combined to achieve dynamic fusion of spectral and spatial features. Finally, the spectral graph context optimizer (SGCO) is used to model the long-range dependency relationship, and improve the classification efficiency and accuracy. Extensive evaluations on four popular HSIs show that, with 10% of the training samples, the proposed method reaches 99.57% average accuracy on the Houston2013 dataset, 99.80% on the Pavia University dataset, 99.85% on the WHU-Hi-HanChuan dataset, and 99.70% on the WHU-Hi-HongHu dataset, superior to some existing advanced technologies.

Keywords: Hyperspectral image (HSI) classification, convolutional neural networks (CNNs), attention mechanism, Fourier transform, deep semantic features.

1. Introduction

Hyperspectral image classification has undergone significant technological evolution over recent decades, emerging as a critical analytical tool across diverse scientific and

industrial domains, including environmental monitoring [1–3], precision agriculture [4], urban planning [5, 6], mineral exploration [7–9], and military reconnaissance [10].

During the early development of HSI analysis, conventional machine learning techniques, including random forest (RF) [11–13], K-nearest neighbors (KNN) [14–16], support vector machines (SVM) [17–19], and decision tree (DT) [20, 21] classifiers, have been extensively studied. However, these early methods showed inherent limitations in dealing with high-dimensional features of HSIs, which often contained hundreds of continuous spectral bands, and often led to suboptimal classification performance due to dimensional limitations and inadequate feature representation capabilities. Subsequent studies have shown that this traditional approach is difficult to effectively model complex nonlinear relationships between spectral features and land cover classes, especially when dealing with subtle spectral variations inherent in mixed pixels and hyperspectral datasets.

The advancement of deep learning (DL) architectures [22–24] has been widely applied to image processing tasks, particularly image classification, leading to substantial improvements in feature representation and classification accuracy. The initial exploration of HSI classification based on deep learning focused on dimensionality reduction techniques using stacked autoencoders (SAE) [25] and recursive autoencoders (RAE) [26]. While these architectures demonstrate improved spectral feature extraction compared to traditional machine learning methods, their inherent vectorization processing destroys critical spatial context information in local HSIs. This limitation has prompted the development of two-dimensional convolutional neural networks (CNNs) [27] to maintain spatial relationships in HSIs. However, 2D CNN exhibits suboptimal performance due to its inability to simulate spectral correlations across hundreds of continuous bands. To address the dimensional limitations of conventional 2D convolutional architectures, researchers developed three-dimensional convolutional neural network variants. Lee et al. [28] proposed depth-aware architectures capable of modeling inter-band correlations through localized spectral-spatial operators. Roy et al. [29] proposed exploring a 3D–2D CNN feature hierarchy (HybridSN) for HSI classification, which effectively combines 3D CNN and 2D CNN and achieves good performance. As the number of network layers increases, the Hughes phenomenon may appear; in order to alleviate this phenomenon, Zhong et al. [30] proposed an end-to-end spectral spatial residual network (SSRN), which takes raw 3D cubes as input data and can classify HSIs without feature engineering. Paoletti et al. [31] proposed a pyramid residual network (PyResNet), which enhances a CNN by incorporating extra residual connections and progressively expanding the feature map dimensions across all convolutional layers. Hu et al. [32] proposed MDRDNet, an integrated neural network that utilizes depthwise separable convolutions to efficiently capture spectral-spatial features, significantly reducing the computational burden compared to 3D CNNs. In addition, Zhang et al. [33] proposed a lightweight 3D asymmetric initial network (AINet) that improves classification performance by emphasizing spectral features. Although CNN-based HSI classification methods have achieved significant progress

in performance due to their remarkable ability to learn local feature representations, they face limitations in capturing global contextual information and deeper semantic features. This limitation arises primarily from the convolution operation being restricted to localized receptive fields. Such constraints can hinder the improvement of classification accuracy, particularly when dealing with complex tasks and diverse datasets. Furthermore, as the network depth increases, the computational cost grows substantially.

The successful application of CNNs in HSI classification has been further enhanced in recent years by the emergence of transformer [34–36] architecture. The transformer architecture has demonstrated an exceptional ability to model long-distance dependencies and capture global context information, making it an effective tool to address the spectral-spatial complexity inherent in HSI data. The evolution of this architecture effectively overcomes the inherent limitations of CNN-based methods, especially when dealing with high-dimensional hyperspectral data cubes, and solves the problems of limited receptive fields and low computational efficiency in deep networks. Sun et al. [37] proposed a spectral–spatial feature tokenization transformer (SSFTT) that utilizes a hybrid 3D-2D convolution system followed by Gauss-weighted feature tokenization. A hierarchical transformer encoder is used to capture the local-to-global spectrum spatial relationship step by step. To reduce the quadratic complexity of standard self-attention, Ma et al. [38] proposed a variant of the vision transformer (ViT) with lightweight self-Gaussian attention (LSGA) that reduces computational overhead while maintaining classification accuracy through adaptive spectral attention graphs. Tu et al. [39] proposed LSFAT, a converter based on local semantic feature aggregation, which achieves multi-scale feature fusion and effectively improves classification accuracy. In addition, Roy et al. [40] proposed MASSFormer, which utilizes spectral and spatial morphology convolution operations to enhance the interaction between structure and shape information among various tokens. Meng et al. [41] introduced the multi-scale super-token transformer (MSSTT), which employs a divide-and-conquer approach to extract multi-scale local features and global dependencies, thereby enhancing the modeling of remote dependencies.

Contemporary computer vision research has witnessed transformative developments through attention-based architectures, which computationally emulate the biological selective perception mechanism observed in human visual cognition. These neural attention paradigms, inspired by the psychovisual prioritization of salient stimuli while suppressing irrelevant sensory inputs, have demonstrated remarkable efficacy across vision tasks, including object detection and semantic segmentation [42–44]. Cui et al. [45] proposed a novel dual triple attention network (DTAN), which effectively improves the performance of HSI classification by capturing interdimensional interactive information. Similarly, Li et al. [46] proposed a dual attention network (DANet) and achieved excellent results in HSI classification. In order to further improve the classification performance of HSI, Li et al. [47] proposed a two-branch two-attention network (DBDA). Although the transformer architecture has demonstrated strong

performance in HSI classification due to its ability to capture long-range dependencies within spectral sequences, several challenges persist. Despite the success of these sequential models, inherent inefficiencies—arising from difficulties in parallelization and the computationally intensive nature of attention mechanisms—continue to hinder their practical application, particularly in large-scale remote sensing tasks. Furthermore, transformer-based HSI classification methods still struggle to effectively capture local, global, and multi-scale deep semantic features simultaneously.

Although current HSI classification models are proficient at feature extraction, challenges such as limited training samples, inadequate spectral-spatial information utilization, and high computational resource consumption still persist in HSI classification. To address these issues, this paper proposes an innovative dual domain feature extraction and adaptive spectral-spatial feature fusion network for HSI classification. The main contributions of this work are as follows:

(1) A dual domain feature extraction (DDFE) module is proposed. The model first utilizes point convolution and deep convolution to extract multi-scale local features, generating spatial domain features. These features are then transformed into frequency domain features using fast Fourier transform (FFT). The frequency domain features are convolved to enhance global information and subsequently transformed back to the spatial domain through inverse FFT (IFFT). Finally, the spatial and frequency domain features are fused using residual connections to capture both local and global information.

(2) An efficient adaptive spectral-spatial fusion (EASSF) module is proposed. A hybrid operator combining 1×1 convolution and 3×3 deep convolution preserves the spectral-spatial relationship. An attention mechanism captures feature dependencies between views, with stepwise enhancement and refinement through two 1×1 convolutions. Finally, the pixel-attention mechanism, integrated through a residual path, fuses spatial and spectral features, resulting in a more comprehensive and refined feature representation. (3) Finally, the spectral graph context optimizer (SGCO) module models long-range dependencies to enhance contextual information, improving both classification accuracy and efficiency for land cover classes. Extensive experiments on four benchmark HSIs, including Houston2013, Pavia University, WHU-Hi-HanChuan, and WHU-Hi-HongHu, demonstrate that the proposed method outperforms several existing methods.

2. Methodology

The workflow of DDFE-ASFS, illustrated in Fig.1, is organized into four distinct stages. First, principal component analysis (PCA) is applied to reduce the dimensionality of the data, thereby simplifying its structure. In the second stage, the DDFE module is used to extract spectral and frequency domain features. In the third stage, the EASSF module is used to realize the dynamic spectral and spatial features. In the fourth stage, SGCO modules are used to model long-range dependencies to optimize context information.

Finally, a linear layer is used to classify the samples and generate corresponding labels.

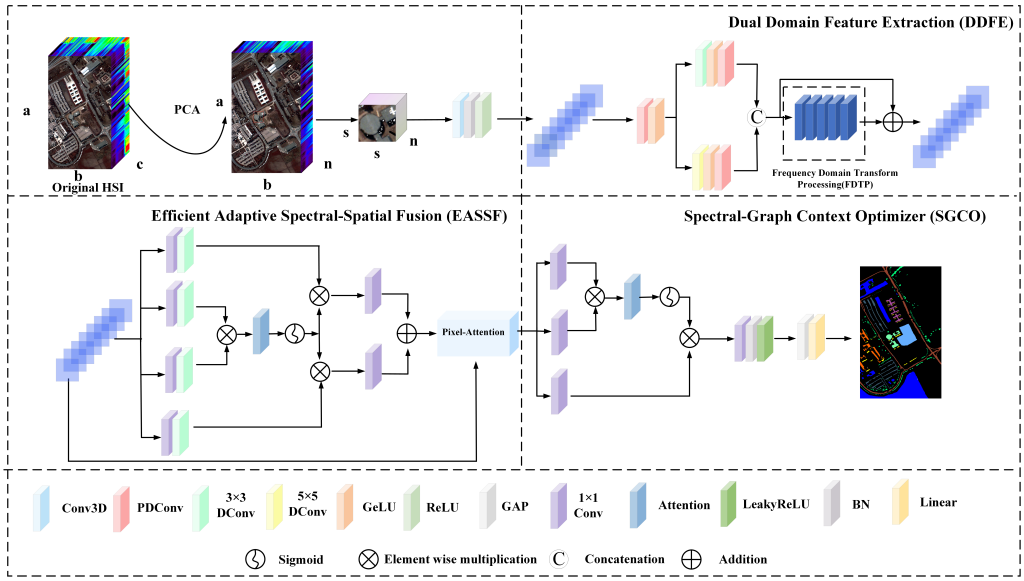


Figure 1. Outline of the proposed DDFE-ASFS architecture.

2.1. DDFE module

1) Dimensionality reduction and convolution operations: Given raw HSIs $I \in R^{a\times b\times c}$. Where $a \times b$ represents the spatial dimension, c is the number of spectral bands. Each pixel contains c spectral dimensions in I, and that corresponds to a unimodal class vector $Y = (y_1, y_2, \cdots, y_C) \in R^{1\times 1\times C}$, where C is the number of land cover categories. HSIs are composed of l bands. Although these bands contain rich spectral information, they also lead to excessive data dimension, which increases the computational burden. Therefore, in order to reduce the computation and spectral dimension, PCA [48] is used to reduce the dimensionality of HSIs.

Next, the 3D patch extraction of HSIs I_{pca} after dimension reduction is carried out. Each 3D adjacent patch $P \in Rs \times s \times n$ is extracted from $Ipca$, where $s \times s$ represents the spatial size of the window and n is the spectral dimension. The center pixel position for each patch is set to (x_i, x_j), where $0 \leq i < a$, $0 \leq j < b$. The true label of each patch is determined by the label of the center pixel. When extracting patches around individual pixels, data on edge pixels cannot be obtained. Therefore, a fill operation is performed on these pixels with a fill width of $\frac{s-1}{2}$. The resulting number of final 3D patches is $a \times b$, and the range covered by each patch is from $x_i - \frac{s-1}{2}$ to $x_i + \frac{s-2}{2}$, the width of the $x_j - \frac{s-1}{2}$ to $x_j + \frac{s-1}{2}$ height, and all n spectral bands. After removing the pixel patch labeled zero, all remaining sample patches are divided into the training sample set and the test sample set.

Then, 3D convolution is used for initial feature extraction to generate richer feature maps for subsequent network layers. In 3D convolution, the spatial position $(\alpha\ ,\ \beta\ ,\ \gamma)$

in the j feature cube of layer i is denoted as $\omega_{i,j}^{\alpha,\beta,\gamma}$, and its calculated value is given by the following formula:

$$\omega_{i,j}^{\alpha,\beta,\gamma} = \nu\left(\sum_{\delta=1}\sum_{\varepsilon=-x}^{x}\sum_{\eta=-y}^{y}\sum_{\theta=-z}^{z}\lambda_{i,j,\delta}^{\varepsilon,\eta,\theta} \times \mu_{i-1,\delta}^{\alpha+\varepsilon,\beta+\eta,\gamma+\theta} + b_{i,j}\right) \tag{1}$$

where ν represents the activation function, $b_{i,j}$ represents the bias parameter of the j feature graph in the i layer, and x, y and z represent the width, height, and number of channels of the three-dimensional convolution kernel, respectively. z represents the spectral dimension. $\lambda_{i,j,\delta}^{\varepsilon,\eta,\theta}$ is the weight parameter attached to the δ^{th} feature cube position $(\varepsilon, \eta, \theta)$.

The DDFE module first expands the channel dimension by point convolution: if the input feature is $X \in R^{C\times H\times W}$

$$X' = GELU(PWConv(x)), x' \in R^{2C\times H\times W} \tag{2}$$

where is the point-wise convolution. X' is divided into two parts, and deep convolution is performed separately.

$$\begin{cases} X_1 = DWConv_{3\times3}(X_1') \\ X_2 = DWConv_{5\times5}(X_2') \end{cases} \tag{3}$$

where $X_1', X_2' \in R^{C\times H\times W}$, $DWConv_{k\times k}$ represents a depth-separable convolution with kernel size k. Splice the features of X_1 and X_2 to form a new feature.

$$X_{concat} = [X_1, X_2] \in R^{2C\times H\times W} \tag{4}$$

The new features after concatenation are processed by frequency domain transform processing (FDTP) to capture global features effectively.

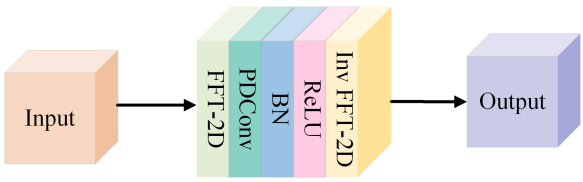


Figure 2. Outline of the proposed FDTP architecture.

2) Component FDTP: The structure of FDTP is shown in Fig.2. Fast Fourier transform on X_{concat} is performed as:

$$F(x) = FFT(X_{concat}) \in R^{2C\times H\times(W/2+1)} \tag{5}$$

where FFT [49] represents the Fourier transform, the formula is:

$$F(u,v) = \frac{1}{\sqrt{HW}}\sum_{h=0}^{H-1}\sum_{w=0}^{W-1} x[h,w]\cdot e^{-j2\pi(\frac{hu}{H}+\frac{wv}{W})} \tag{6}$$

where $F(u, v)$ is a complex component built on fast Fourier transform spatial, where u and v represent coordinates in Fourier spatial, respectively.

The transform $F(x)$ is enhanced in the frequency domain to capture its global features effectively.

$$F'(x) = ReLU(BN(PDConv(F(x)))) \in R^{2C\times H\times(W/2+1)} \tag{7}$$

By inverse fast Fourier transform, the characteristic information in the frequency domain is restored to the spatial domain.

$$X_{freq} = IFFT(F'(x)) \in R^{2C\times H\times W} \tag{8}$$

where IFFT represents the inverse fast Fourier transform.

The X_{freq} obtained after inverse transformation carries global structural information and is complemented and enhanced by residual connections with local features X_{concat} in the spatial domain.

$$X_{fused} = \varrho X_{freq} + (1 - \varrho)X_{concat} \tag{9}$$

where ϱ is a learnable parameter that adaptively adjusts the frequency spatial contribution.

2.2. Structure of EASSF

Split the global and local X_{fused} extracted in the previous step into X_l and X_r and input them into the EASSF module for effective feature fusion. Firstly, depth-separable convolution is used for multi-scale feature enhancement.

$$\begin{cases} Q_l = DWConv_{3\times3}(PWConv(X_l)) \in R^{C\times H\times W} \\ Q_r = DWConv_{3\times3}(PWConv(X_r)) \in R^{C\times H\times W} \\ V_l = DWConv_{3\times3}(PWConv(X_l)) \in R^{C\times H\times W} \\ V_r = DWConv_{3\times3}(PWConv(X_r)) \in R^{C\times H\times W} \end{cases} \tag{10}$$

Then, using the attention matrix, the global correlation between spectral and empty spectral features is established.

$$A = \frac{Q_l \times Q_r^T}{\sqrt{C}} \tag{11}$$

where $\sqrt{C}$ is the scaling factor to prevent gradient explosion. Feature interaction is realized through softmax normalization.

$$\begin{cases} F_{l\to r} = Softmax(A) \cdot V_r \\ F_{r\to l} = Softmax(A^T) \cdot V_l \end{cases} \tag{12}$$

Next, the feature dimensions are adjusted using 1×1 convolution.

$$\begin{cases} F'_{l\to r} = Conv_{1\times1}(F_{l\to r}) \\ F'_{r\to l} = Conv_{1\times1}(F_{r\to l}) \end{cases} \tag{13}$$

By introducing a pixel-attention gating mechanism, spectral-spatial sensitive gating is constructed.

$$\upsilon = \phi(PA(X_l + X_r, F'_{l\to r} + F'_{r\to l})) \tag{14}$$

where ϕ is the sigmoid function and PA is the pixel-attention module. Finally, through adaptive feature fusion, the empty spectral-spatial features are effectively integrated.

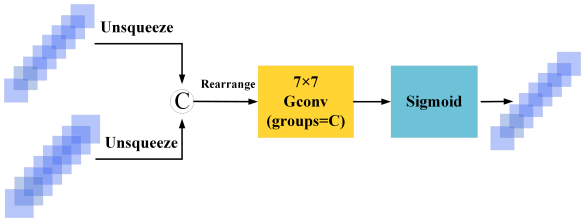


Figure 3. Outline of the proposed pixel-attention architecture.

$$X_{fused1} = \upsilon \odot X_l + (1 - \upsilon) \odot X_r \tag{15}$$

where $\odot$ is the channel-by-channel product. The structure of the pixel-attention module is depicted in Fig.3, and its corresponding formula is as follows:

$$PA(F_1, F_2) = Sigmoid(Conv_{7\times7}^{group=C})([F_1]||[F_2]) \tag{16}$$

where $||$ represents channel splicing, grouping convolution to maintain parametric efficiency.

2.3. Structure of SGCO

In this part, we use the SGCO module for feature optimization to improve the model's ability to recognize and classify complex ground objects so as to significantly improve the classification accuracy. Input features X_{fused} are entered into each of the three 1×1 convolution layers, ensuring that the output dimensions remain unchanged.

$$\begin{cases} Q = Conv_{1\times1}(X_{fused1}) \in R^{c\times H\times W} \\ K = Conv_{1\times1}(X_{fused1}) \in R^{c\times H\times W} \\ V = Conv_{1\times1}(X_{fused1}) \in R^{c\times H\times W} \end{cases} \tag{17}$$

where c is $C/3$. Then, the correlation degree between spectral features and empty spectral features is calculated using the attention mechanism.

$$A = softmax(\frac{Q \cdot K}{\sqrt{d}} \odot \hbar) \in R^{2c\times H\times W} \tag{18}$$

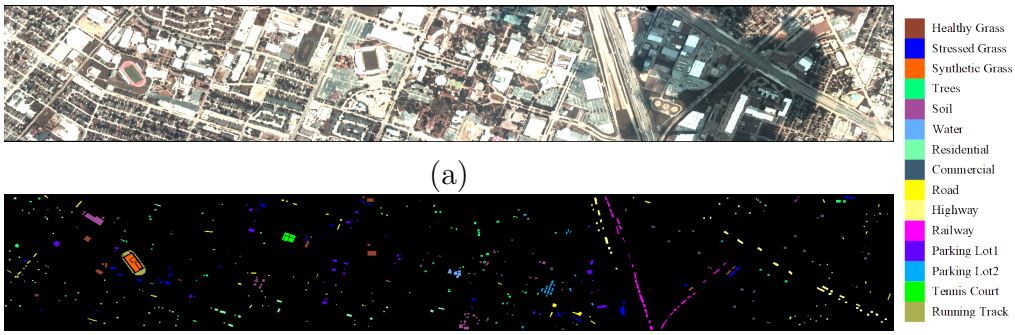
where $\odot$ represents the Hadamard product, $\hbar$ is the normalized adjacency matrix. Then, feature aggregation is carried out to optimize the context information.

$$Z = A \cdot V \in R^{C\times H\times W} \tag{19}$$

Finally, the LeakyReLU activation function [50] is used to effectively alleviate the problem of gradient disappearance in the attention matrix, thus achieving adaptive feature calibration.

3. EXPERIMENT AND ANALYSIS

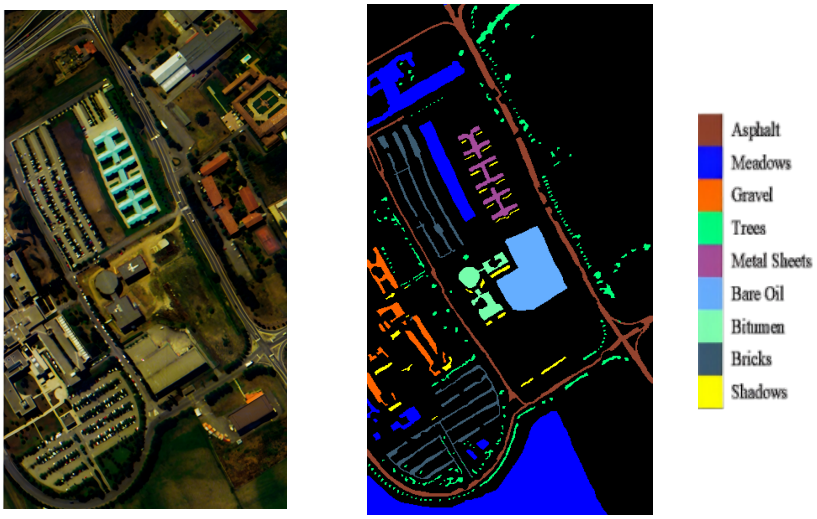
3.1. Dataset Description



(a)

(b)

Figure 4. The Houston2013 dataset. (a) Three-band color composite image. (b) Ground truth map.

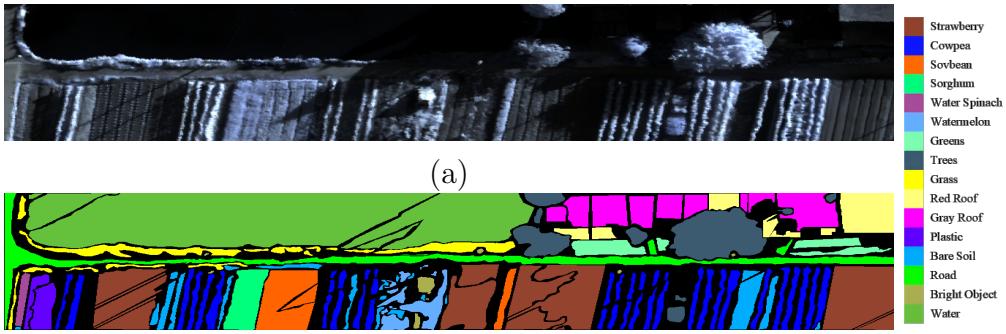


(a) (b)

Figure 5. The Pavia University dataset. (a) Three-band color composite image. (b) Ground truth map.

Experiments are conducted using four publicly available HSI datasets: the Houston2013 , the Pavia University , the WHU-Hi-HanChuan , and the WHU-Hi-HongHu datasets.

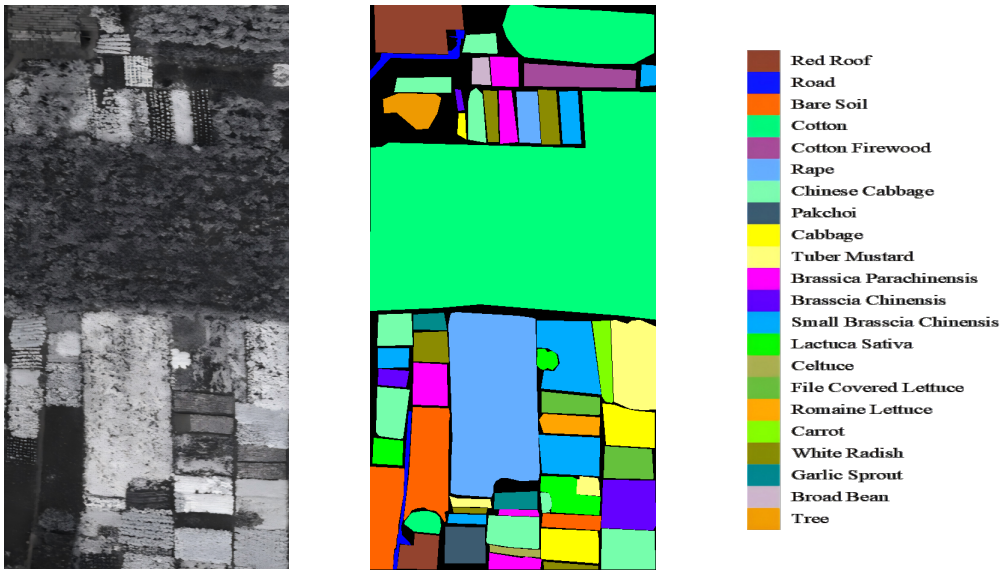
The Houston2013 dataset was acquired by a hyperspectral sensor in Houston, Texas, USA, and has a high spatial resolution of about 1.5 meters. The dataset covers 144



(a)

(b)

Figure 6. The WHU-Hi-HanChuan dataset. (a) Three-band color composite image. (b) Ground truth map.



(a) (b)

Figure 7. The WHU-Hi-HongHu dataset. (a) Three-band color composite image. (b) Ground truth map.

bands, covering the wavelength range from 0.38 to 1.05 μm, providing detailed spectral information. The dataset contains 15 different categories of ground features.

The Pavia University dataset was acquired in 2001 using the Reflective Optics System Imaging Spectrometer (ROSIS) sensor, encompassing an urban-academic environment in northern Italy's Lombardy region. This hyperspectral cube contains 115 calibrated spectral channels spanning the visible to near-infrared spectrum ($430 \sim 860$ nm wavelength range). With a spatial dimension of 610×340 pixels and 1.3 *meter* ground sampling distance, the dataset provides nine annotated land-cover categories representing typical urban terrain features.

The WHU-Hi-HanChuan dataset was collected via airborne hyperspectral survey conducted on *June*17, 2016 over transitional urban-agricultural landscapes in Hubei Province's HanChuan district. Acquisition utilized a Headwall Nano-Hyperspec imaging

Table 1. The number of training and test samples for each category in the Houston2013 and University of Pavia datasets.

	Houston2013				University of Pavia			
No	Class	Train	Test	Total Samples	Class	Train	Test	Total Samples
1	Healthy Grass	125	1126	1251	Asphalt	663	5968	6631
2	Stressed Grass	125	1129	1254	Meadows	1865	16784	18649
3	Synthetic Grass	70	627	697	Gravel	210	1889	2099
4	Tress	124	1120	1244	Trees	306	2758	3064
5	Soil	124	1118	1242	MetalSheets	134	1211	1345
6	Water	33	292	325	BareOil	503	4526	5029
7	Residential	127	1141	1268	Bitumen	133	1197	1330
8	Commercial	124	1120	1244	Bricks	368	3314	3682
9	Road	125	1127	1252	Shadows	95	852	947
10	Highway	123	1104	1227				
11	Railway	123	1112	1235				
12	Parking Lotl	123	1110	1233				
13	Parking Lot2	47	422	469				
14	Tennis Court	43	385	428				
15	Running Track	66	594	660				
-	Total	1502	13527	15029	Total	4277	38499	42776

Table 2. The number of training and test samples for each category in the WHU-Hi-HanChuan and WHU-Hi-HongHu datasets.

	WHU-Hi-HanChuan				WHU-Hi-HongHu			
No	Class	Train	Test	Total Samples	Class	Train	Test	Total Samples
1	Strawberry	4473	40262	44735	RedRoof	1404	12637	14041
2	Cowpea	2275	20478	22753	Road	351	3161	3512
3	Soybean	1029	9258	10287	BareSoil	2182	19639	21821
4	Sorghum	535	4818	5353	Cotton	16328	146957	163285
5	WaterSpinach	120	1080	1200	CottonFirewood	622	5596	6218
6	Watermelon	453	4080	4533	Rape	4456	40101	44557
7	Greens	590	5313	5903	ChineseCabbage	2410	21693	24103
8	Trees	1798	16180	17978	Pakchoi	405	3649	4054
9	Grass	947	8522	9469	Cabbage	1082	9737	10819
10	RedRoof	1052	9464	10516	TuberMustard	1239	11155	12394
11	GrayRoof	1691	15220	16911	Brassicaparachinensis	1102	9913	11015
12	Plastic	368	3311	3679	BrassicaChinensis	895	8059	8954
13	Baresoil	912	8204	9116	SmallBrassicaChinensis	2251	20256	22507
14	Road	1856	16704	18560	LactucaSativa	736	6620	7356
15	BrightObject	114	1022	1136	Celtuce	100	902	1002
16	Water	7540	67861	75401	FilmCoveredlettuce	726	6536	7262
					RomaineLettuce	301	2709	3010
					Carrot	322	2895	3217
					WhiteRadish	871	7841	8712
					GarlicSprout	349	3137	3486
					BroadBean	133	1195	1328
					Tree	404	3636	4040
-	Total	25753	231777	257530	Total	38669	348024	386693

spectrometer integrated with the Leica Aibot ×6 UAV platform, achieving 0.109 m ground sampling distance from 250 m flight altitude. This 274 *channel* spectral cube with 1217×303 pixel dimensions documents seven key cash crops within heterogeneous land landcover comprising built-up areas, aquatic systems, and cultivated fields.

The WHU-Hi-HongHu hyperspectral dataset employed a DJI Matrice 600 Pro UAV platform equipped with a Headwall Nano-Hyperspec sensor. Under stable meteorological conditions (overcast sky, ambient temperature maintained at 8°C with 55% relative humidity), the system recorded 270 discrete spectral channels over

diversified agricultural terrain featuring leafy vegetable cultivars including Chinese cabbage, bok choy, and kale variants. The nadir-oriented imaging geometry at 100 m altitude yielded 940×475 pixel scenes with 0.043 *meter* ground resolution.

Our experimental protocol employed a stratified random sampling strategy, allocating 10% of each dataset's total instances (Houston2013, Pavia University, WHU-Hi-HanChuan, and WHU-Hi-HongHu) for training subsets, with the remaining 90% reserved for model testing. Detailed categorical taxonomy with corresponding sample partitions is systematically organized in Tables 1-2. Complementary visual representations in Figs. 4-7 provide geo-referenced pseudo-color composites (generated from optimized spectral band combinations) juxtaposed with reference cartography, employing standardized chromatic coding to differentiate land-cover categories across the multi-sensor datasets.

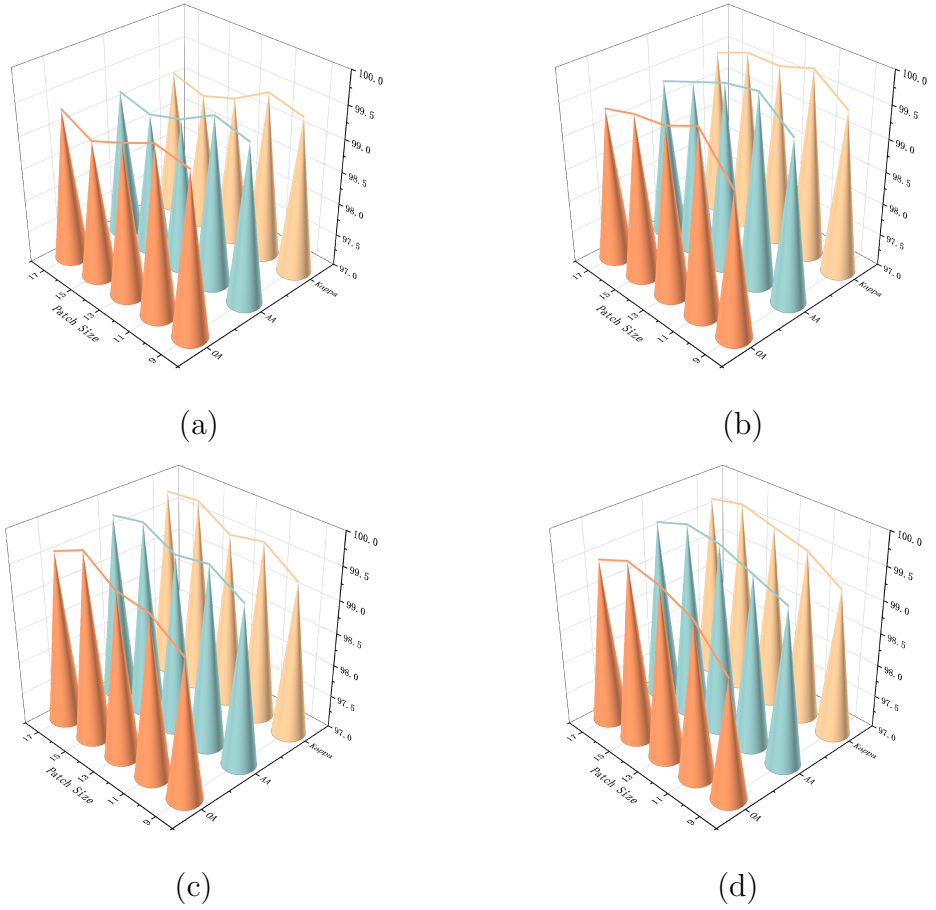


Figure 8. Impact of varied patch sizes on quantitative evaluation metrics. (a) Houston2013, (b) Pavia University, (c) WHU-Hi-HanChuan, (d)WHU-Hi-HongHu.

3.2. Experimental Setting

1) Evaluation Metrics: To quantitatively evaluate the performance of the proposed methods and compare them, we introduce three key evaluation metrics: Overall

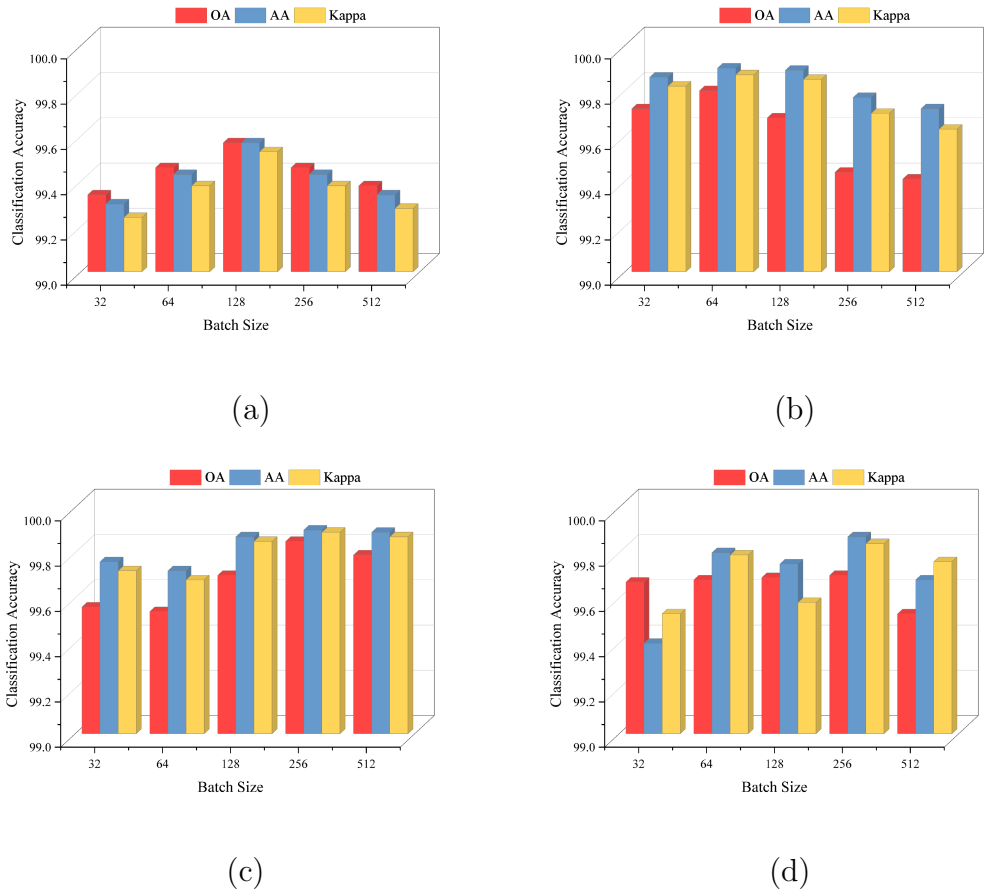


Figure 9. Impact of varied batch sizes on quantitative evaluation metrics. (a) Houston2013, (b) Pavia University, (c) WHU-Hi-HanChuan, (d)WHU-Hi-HongHu.

Accuracy (OA), Average Accuracy (AA), and Kappa coefficient (κ). 2) Configuration: All validation experiments for the classification methods are performed within the PyTorch framework, leveraging an Intel Core i5-12700KF processor, an NVIDIA RTX 4060Ti GPU, and 16 GB of RAM. 3) Parameter Analysis: In the parameter analysis, we examined various factors that influence the classification accuracy, specifically the patch size $p \times p$, batch size $s \times s$, and learning rate r.

3.2.1. Patch size impact on classification accuracy The choice of patch size significantly influences the model's performance. If the patch size is too small, the learning process may lack sufficient information, thereby limiting the model's effectiveness. On the other hand, selecting an excessively large patch size may lead to information redundancy, potentially causing overfitting and reducing accuracy. Furthermore, larger patch sizes increase computational complexity. Fig. 8 illustrates the impact of various patch sizes. Each experimental result represents the average results of five experimental runs. From the Figure, it is evident that the optimal classification accuracy for the four datasets in this model corresponds to patch sizes of 11 ×11, 11×11, 15×15, and 11×11, respectively.

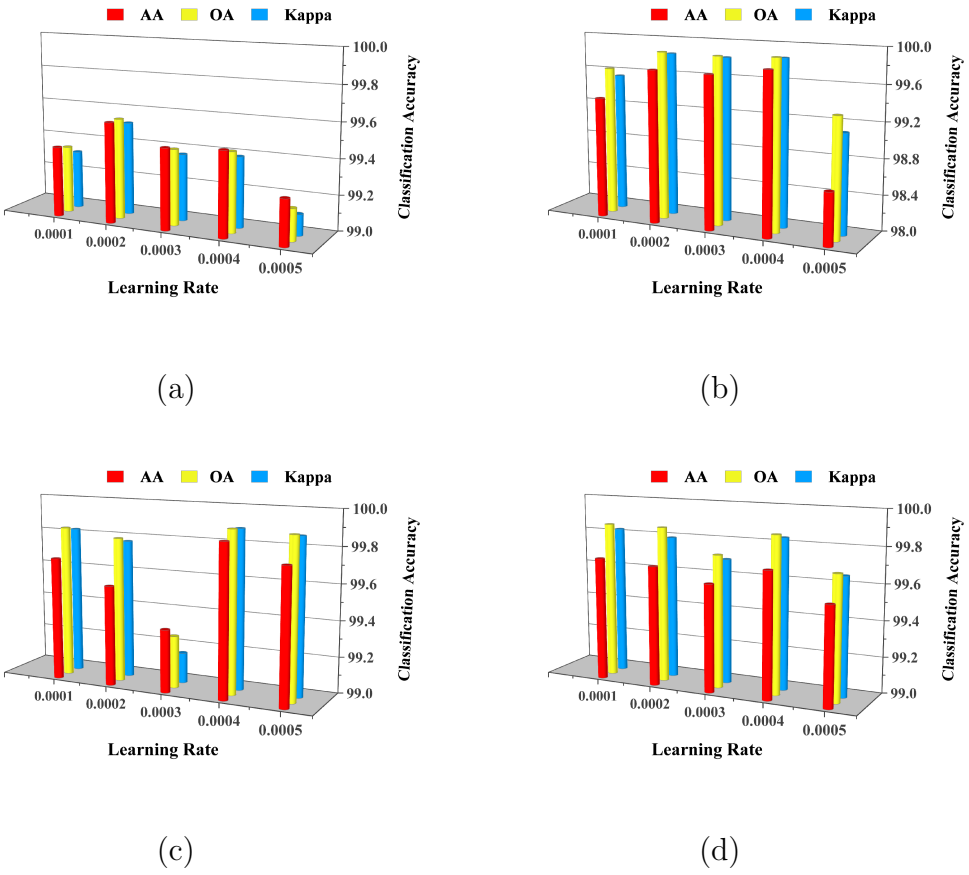


Figure 10. Impact of varied learning rates on quantitative evaluation metrics. (a) Houston2013, (b) Pavia University, (c) WHU-Hi-HanChuan, (d)WHU-Hi-HongHu.

3.2.2. Impact of batch size on classification accuracy Generally, employing a smaller batch size can improve a model's generalization ability, albeit at the cost of slower convergence during training. In contrast, utilizing a larger batch size accelerates the training process but may adversely affect the model's stability and its ability to generalize effectively. Fig. 9 shows how classification accuracy varies with batch size. Each data point represents the average result of five experiment runs. It is obvious from the figure that the optimal classification accuracy of the four datasets in the model is 128, 64, 256, and 256, respectively.

3.2.3. Effect of learning rate on classification performance The learning rate controls the size of the updated parameters during each update, and each iteration of training plays a crucial role in the convergence speed and stability of the model. Too high a learning rate may cause the model to oscillate during training or fail to converge. Too low a learning rate may cause the model to converge too slowly, thus prolonging the training process. Fig. 10 shows the accuracy differences for the four datasets at the different learning rate. As can be seen from the Figure, the accuracy of the four datasets

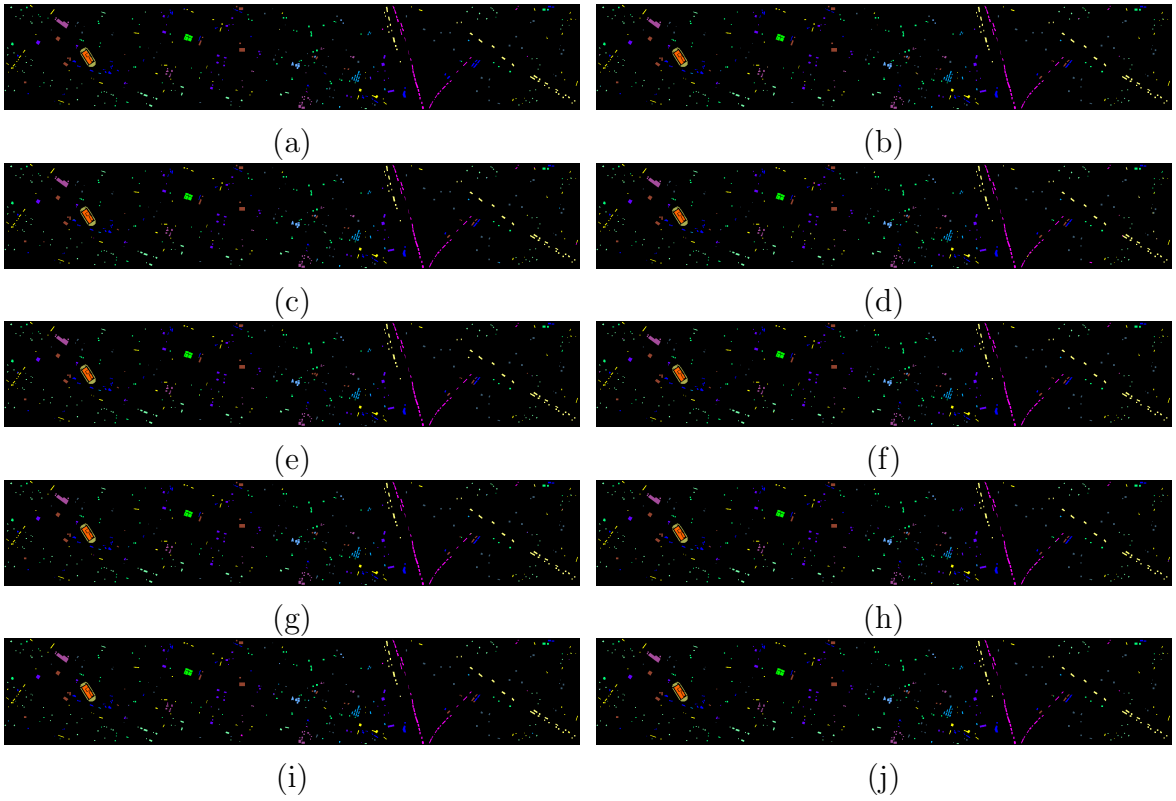


Figure 11. Visualization of classification outcomes using various classification methods applied to the Houston2013 dataset with 10% of the training samples. (a) SSFTT (OA=98.63%), (b) morphFormer (OA=99.12%), (c) LSGA (OA=98.86%), (d) DCTN (OA=98.99%), (e) MSSTT (OA=97.80%), (f) RDTN (OA=99.17%), (g) MASSFormer (OA=99.05%), (h) LSFAT (OA=97.31%), (i) DBCT (OA=99.30%) and (j) DDFE-ASFS (OA=99.57%).

is optimal when the learning rates are 0.0002, 0.0002, 0.0004, and 0.0002, respectively.

4. Comparison of Classification Performance

4.1. Quantitative Accuracy Analysis

To comprehensively assess the classification performance of the proposed DDFE-ASFS framework, we perform an extensive comparison with nine advanced spectral-spatial methods: SSFTT [37], morphFormer [51], LSGA [38], DCTN [52], MSSTT [41], RDTN [53], MASSFormer [54], LSFAT [39], and DBCT [55]. This comparison aims to provide a thorough evaluation of the proposed model against leading techniques in the field, with all contrastive algorithms being assessed according to the parameter configurations specified in the original publications.

Table 3 presents the classification results of various methods on the Houston2013 dataset, including the accuracy for each class, while the corresponding classification maps for each algorithm are shown in Fig. 11. Table 3 compares the performance of ten algorithms, with DDFE-ASFS achieving outstanding results across all metrics.

Table 3. Accuracy for every class (%), OA (%), AA (%), and κ (%) of various techniques for the Houston2013 dataset

NO.	SSFTT	morphFormer	LSGA	DCTN	MSSTT	RDTN	MASSFormer	LSFAT	DBCT	DDFE-ASFS
1	98.49	99.20	99.82	99.82	95.63	98.59	**99.91**	98.90	98.58	**99.91**
2	99.38	98.94	99.91	98.77	95.26	99.91	99.21	95.52	99.91	**100.0**
3	99.68	**100.0**	99.68	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
4	97.97	99.56	99.30	98.41	97.27	99.19	98.59	96.18	98.92	**99.64**
5	**100.0**	99.20	**100.0**	**100.0**	**100.0**	**100.0**	100.0	**100.0**	**100.0**	**100.0**
6	93.89	**100.0**	**100.0**	**100.0**	**100.0**	97.99	**100.0**	**100.0**	**100.0**	**100.0**
7	95.43	97.50	96.57	97.46	95.97	97.36	97.69	93.52	**98.15**	97.82
8	**100.0**	99.27	96.15	**100.0**	99.64	**100.0**	99.27	99.71	**100.0**	**100.0**
9	97.57	99.28	98.09	98.64	94.97	98.07	99.18	96.98	98.25	**99.21**
10	**100.0**	98.57	99.91	99.28	98.48	99.37	96.50	99.91	99.55	99.91
11	98.84	98.14	99.55	98.76	99.73	99.20	98.22	99.91	**100.0**	99.11
12	**99.91**	99.46	97.97	97.02	98.22	99.28	99.10	89.05	99.37	99.28
13	99.52	97.66	97.91	99.52	99.74	99.04	97.44	**100.0**	97.42	**100.0**
14	**100.0**	98.21	**100.0**	**100.0**	99.48	**100.0**	99.22	**100.0**	99.74	**100.0**
15	96.74	99.83	**100.0**	99.83	98.02	**100.0**	**100.0**	99.83	**100.0**	**100.0**
OA(%)	98.63	99.12	98.86	98.99	97.80	99.17	99.05	97.31	99.31	**99.57**
AA(%)	98.84	98.88	98.99	99.17	97.24	99.20	98.77	96.38	99.33	**99.57**
κ(%)	98.52	98.96	98.77	98.90	97.62	99.10	98.86	97.09	99.24	**99.53**

Table 4. Accuracy for every class (%), OA (%), AA (%), and κ (%) of various techniques for the Pavia University dataset

NO.	SSFTT	morphFormer	LSGA	DCTN	MSSTT	RDTN	MASSFormer	LSFAT	DBCT	DDFE-ASFS
1	99.48	99.76	98.86	99.01	99.22	97.36	99.12	99.26	98.83	**100.0**
2	99.94	99.72	99.86	99.90	99.92	98.78	99.82	99.99	99.07	**100.0**
3	**99.90**	99.27	97.52	97.38	98.54	94.81	94.52	98.58	93.92	99.73
4	98.28	99.35	90.10	99.78	97.22	98.04	99.78	98.39	99.42	**99.96**
5	99.21	**100.0**	99.62	**100.0**	98.21	**100.0**	**100.0**	99.68	**100.0**	**100.0**
6	99.90	99.95	**100.0**	99.82	**100.0**	99.22	99.38	99.61	99.03	99.98
7	98.14	98.66	**100.0**	99.15	98.93	99.34	99.66	99.67	99.81	**100.0**
8	97.89	84.57	98.03	94.53	98.73	98.84	95.79	93.14	97.92	**99.16**
9	95.79	99.77	96.98	99.77	99.23	100.00	**100.0**	97.43	**100.0**	**100.0**
OA(%)	99.39	96.03	98.68	99.13	99.35	98.45	98.41	98.95	98.77	**99.90**
AA(%)	98.45	96.35	97.88	98.82	98.29	98.49	98.75	97.80	98.67	**99.80**
κ(%)	99.20	97.25	98.26	98.85	99.14	97.95	99.05	98.61	98.38	**99.87**

Table 5. Accuracy for every class (%), OA (%), AA (%), and κ (%) of various techniques for the WHU-Hi-HanChuan dataset

NO.	SSFTT	morphFormer	LSGA	DCTN	MSSTT	RDTN	MASSFormer	LSFAT	DBCT	DDFE-ASFS
1	98.52	99.14	99.25	98.94	99.11	98.80	99.29	98.64	99.02	**99.93**
2	99.40	99.58	98.94	98.99	99.37	98.92	99.06	98.65	99.32	**99.91**
3	99.66	99.09	99.62	99.77	99.33	98.32	99.55	98.85	99.16	**99.80**
4	99.79	99.63	99.64	99.92	99.82	99.69	99.96	99.73	99.28	**100.0**
5	98.33	97.64	98.11	98.83	99.54	85.83	99.35	95.59	97.91	**99.82**
6	98.58	91.49	95.76	96.55	99.18	97.65	98.45	95.27	96.64	**99.75**
7	98.45	92.62	95.99	97.67	98.83	94.82	98.54	97.90	96.01	**99.92**
8	99.11	99.41	97.21	98.34	98.35	97.91	99.12	98.55	98.50	**99.90**
9	98.82	98.97	97.94	94.46	98.67	98.10	98.46	98.07	97.58	**99.91**
10	99.69	98.80	99.58	99.57	99.45	99.80	99.82	99.04	99.94	**99.85**
11	99.38	99.19	99.21	98.56	99.15	99.24	99.43	98.96	98.61	**99.96**
12	99.16	97.83	96.05	99.08	99.40	97.9	97.94	97.26	98.92	**100.0**
13	**99.50**	97.04	95.42	95.10	95.59	98.18	97.31	96.53	97.41	99.17
14	99.60	97.46	99.30	99.08	99.52	98.92	99.65	98.74	98.79	**99.99**
15	98.40	99.50	99.20	99.28	99.12	98.06	99.50	96.74	99.49	**100.0**
16	99.94	**100.0**	99.97	99.97	99.95	99.94	99.98	99.94	99.84	99.97
OA(%)	99.35	98.65	98.95	98.89	99.25	98.91	98.85	98.88	99.03	**99.90**
AA(%)	98.33	98.72	98.20	98.38	98.20	97.63	99.28	97.48	98.53	**99.85**
κ(%)	99.24	98.91	98.78	98.70	99.12	98.72	99.39	98.69	98.86	**99.89**

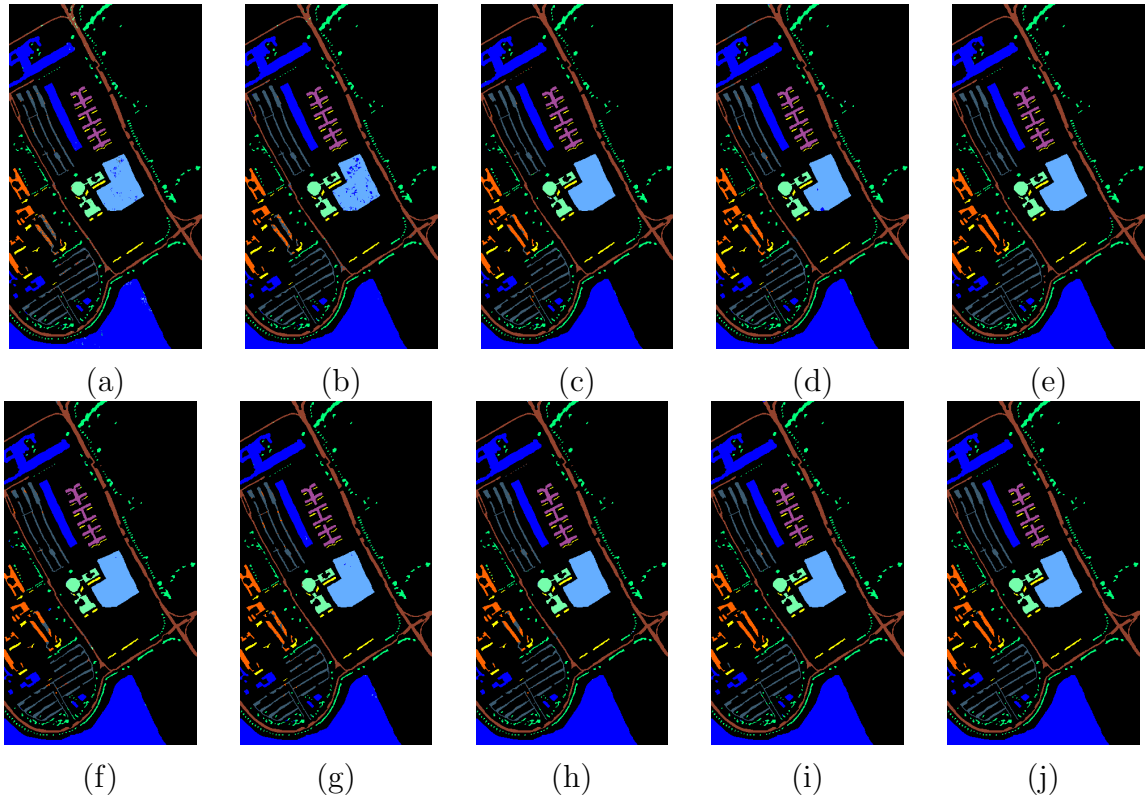


Figure 12. Visualization of classification outcomes using various classification methods applied to the Pavia University dataset with 10% of the training samples. (a) SSFTT (OA=99.39%), (b) morphFormer (OA=96.03%), (c) LSGA (OA=98.68%), (d) DCTN (OA=98.13%), (e) MSSTT (OA=99.35%), (f) RDTN (OA=98.45%), (g) MASSFormer (OA=98.41%), (h) LSFAT (OA=98.95%), (i) DBCT (OA=98.77%) and (j) DDFE-ASFS (OA=99.90%).

According to Table 3, the OA of DDFE-ASFS reaches 99.57%, outperforming all other algorithms. This high OA value indicates a strong consistency between the classification results of DDFE-ASFS and the ground truth, thereby demonstrating its reliability and stability in classifying ground objects. When comparing the performance of DDFE-ASFS with other methods, it is evident that DDFE-ASFS consistently maintains high classification accuracy. For instance, DDFE-ASFS achieves 100% accuracy in most individual classes, highlighting its advanced data learning and feature extraction capabilities. Other algorithms, such as MASSFormer and DBCT, also perform well; however, DDFE-ASFS consistently delivers high performance across the entire dataset.

Table 4 presents the classification results of ten algorithms, including DDFE-ASFS, on the Pavia University dataset. Fig. 12 presents a visual comparison of the ground truth maps and the classification maps produced by each algorithm. As shown in Table 4, the DDFE-ASFS method achieved the highest classification accuracy in 7 out of the 9 ground object categories, further confirming its superior performance in HSI classification. In contrast, the OA of the MorphFormer and RDTN methods is 96.03% and 98.45%, respectively, indicating minor classification errors in certain categories. In

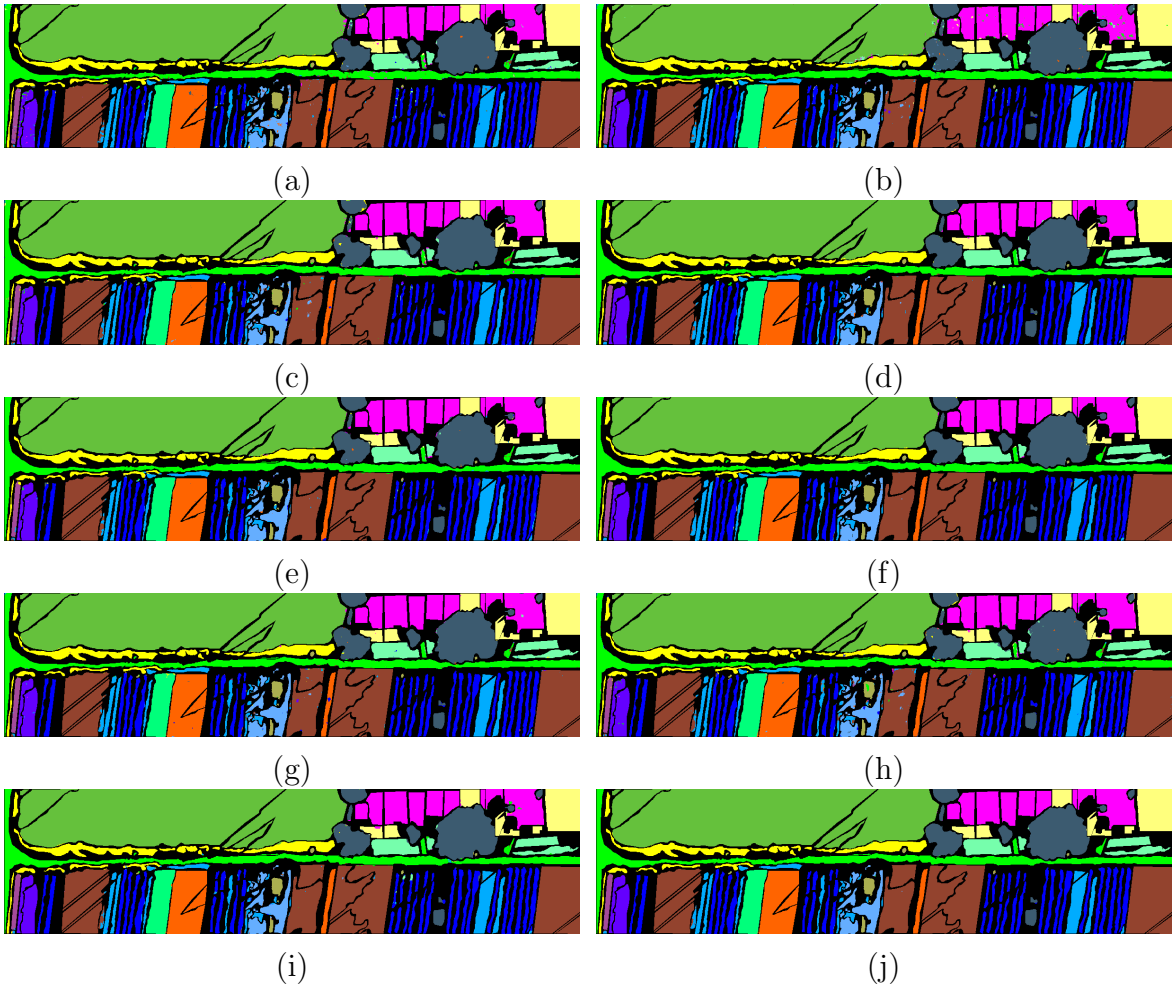


Figure 13. Visualization of classification outcomes using various classification methods applied to the WHU-Hi-HanChuan dataset with 10% of the training samples. (a) SSFTT (OA=99.35%), (b) morphFormer (OA=98.65%), (c) LSGA (OA=98.95%), (d) DCTN (OA=98.89%), (e) MSSTT (OA=99.25%), (f) RDTN (OA=98.91%), (g) MASSFormer (OA=98.85%), (h) LSFAT (OA=98.88%), (i) DBCT (OA=99.03%) and (j) DDFE-ASFS (OA=99.90%).

addition, the LSFAT method has a classification accuracy of 93.14% in category eight, indicating a decline in accuracy in some categories compared to other methods.

Table 5 presents the classification accuracy of ten algorithms on the WHU-Hi-HanChuan dataset, while Fig.13 illustrates the classification results of each algorithm. From the table, it is evident that the DDFE-ASFS algorithm demonstrates exceptional performance, achieving an OA of 99.90%, significantly outperforming other algorithms. In contrast, the OA for the SSFTT and MSSTT algorithms are 98.63% and 97.80%, respectively. Despite their overall strong performance, these algorithms exhibit certain limitations in the classification accuracy of specific feature categories. For instance, the MSSTT algorithm shows a noticeable decline in accuracy in complex background scenarios, which may be attributed to its insufficient capability in capturing subtle spectral differences during feature extraction. Similarly, although SSFTT achieves

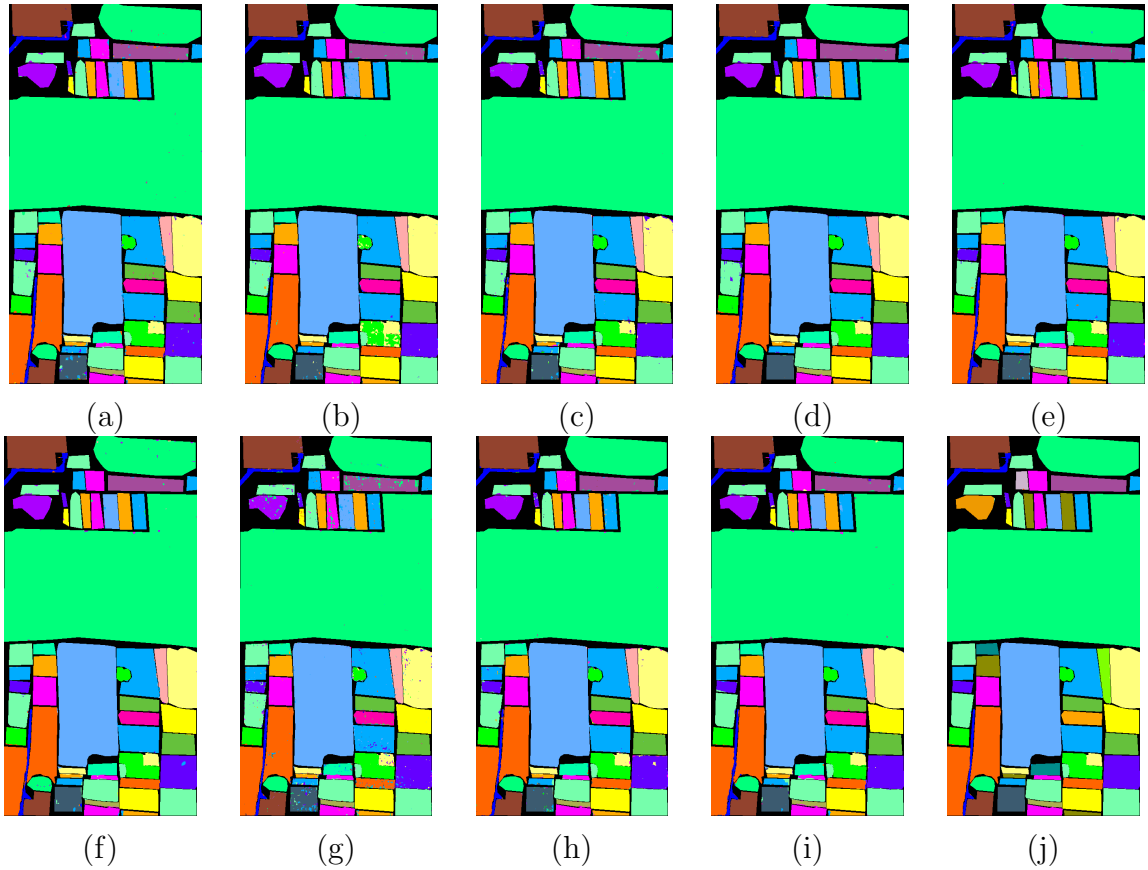


Figure 14. Visualization of classification outcomes using various classification methods applied to the WHU-Hi-HongHu dataset with 10% of the training samples. (a) SSFTT (OA=98.94%), (b) morphFormer (OA=98.16%), (c) LSGA (OA=99.39%), (d) DCTN (OA=99.08%), (e) MSSTT (OA=99.39%), (f) RDTN (OA=98.97%), (g) MASSFormer (OA=95.28%), (h) LSFAT (OA=99.12%), (i) DBCT (OA=99.10%) and (j) DDFE-ASFS (OA=99.87%).

an overall accuracy of 98.63%, its classification accuracy significantly decreases in certain ground object classification tasks, particularly for categories with similar spectra. This limitation may stem from the relatively restricted feature fusion capacity of morphological operators in hyperspectral data, which leads to suboptimal classification performance in dealing with hyperspectral-related objects.

Table 6 presents the classification accuracy of ten different methods on the WHU-Hi-HongHu dataset, while Fig.14 displays the classification results for each algorithm. As shown in Table 6, the DDFE-ASFS algorithm achieves the best performance, with an OA of 99.87%, an AA of 99.70%, and a κ of 99.84%, significantly outperforming the other methods. The LSGA algorithm yields an OA of 99.39%, while DCTN achieves an OA of 99.08%, both performing well but still lagging behind DDFE-ASFS. In contrast, the MASSFormer algorithm demonstrates a lower OA of 95.28%, highlighting the challenges it faces in achieving high classification accuracy on this dataset. Nonetheless, it still exhibits strong performance in certain categories. Similarly, other methods, such as LSFAT, DBCT, and MSSTT, achieve OAs of 99.12%, 99.10%,

Table 6. Accuracy for every class (%), OA (%), AA (%), and κ (%) of various techniques for the WHU-Hi-HongHu dataset

NO.	SSFTT	morphFormer	LSGA	DCTN	MSSTT	RDTN	MASSFormer	LSFAT	DBCT	DDFE-ASFS
1	99.16	99.84	99.41	99.26	99.30	99.58	98.71	99.24	99.38	**99.62**
2	92.91	97.26	96.63	92.33	95.48	91.52	97.04	96.59	92.37	**99.68**
3	98.57	99.60	99.34	99.54	99.22	98.66	99.01	99.21	99.27	**99.83**
4	99.88	99.82	99.70	99.78	99.92	99.66	98.93	99.89	99.82	**99.97**
5	98.18	99.25	99.26	97.97	99.15	98.9	95.33	99.07	98.63	**99.98**
6	99.63	99.77	99.78	99.51	99.72	99.66	99.05	99.46	99.44	**99.97**
7	98.26	99.07	98.33	99.16	98.68	97.89	97.61	98.50	98.27	**99.67**
8	98.95	99.46	99.14	94.08	98.14	98.37	95.10	97.00	96.96	**99.78**
9	99.30	99.79	99.75	99.43	99.61	99.62	99.82	99.40	99.34	**99.98**
10	99.28	91.14	99.45	97.38	99.08	98.56	98.12	98.20	98.58	**99.77**
11	97.36	99.27	98.34	98.08	97.73	97.23	95.70	95.76	98.42	**99.93**
12	96.69	97.97	99.05	97.85	98.47	96.59	92.79	97.83	98.03	**99.85**
13	95.65	97.79	98.86	98.29	98.40	98.62	94.44	99.15	98.77	**99.84**
14	99.08	**99.98**	99.66	98.56	98.99	99.07	99.30	98.85	98.76	99.58
15	**99.66**	97.16	99.52	97.43	97.69	97.39	98.61	97.14	97.14	99.34
16	99.62	99.71	99.41	98.25	99.30	98.63	97.65	98.75	99.56	**99.95**
17	94.48	99.55	98.92	98.76	98.76	98.48	99.28	100.00	96.61	**99.89**
18	94.71	98.79	98.44	96.41	95.95	94.02	96.76	95.87	96.86	**98.87**
19	99.47	98.49	99.44	98.77	98.76	98.95	96.91	97.38	98.67	**99.82**
20	98.48	94.21	98.51	96.9	98.68	96.14	96.10	95.42	93.99	**99.49**
21	95.83	95.52	98.17	95.82	97.35	**99.74**	87.76	95.05	91.17	99.09
22	97.81	99.72	99.21	98.55	97.82	95.64	96.76	97.69	97.94	**99.86**
OA(%)	98.94	98.16	99.39	99.08	99.39	98.97	95.28	99.12	99.10	**99.87**
AA(%)	97.37	98.89	99.01	97.82	98.41	97.86	97.56	97.91	97.64	**99.70**
κ(%)	98.66	99.12	99.23	98.83	99.14	98.69	98.07	98.89	98.86	**99.84**

Table 7. Ablation experiment results on Houston2013, Pavia University, WHU-Hi-HanChuan and WHU-Hi-HongHu datasets

Cases	Components			Indicators							
				Houston2013		Pavia University		WHU-Hi-HanChuan		WHU-Hi-HongHu	
	DDFE	EASSF	SGCO	OA (%)	AA (%)	OA (%)	AA (%)	OA (%)	AA (%)	OA (%)	AA (%)
1	✓	×	×	99.29	99.30	99.56	99.34	99.61	99.10	99.10	99.21
2	✓	✓	×	99.50	99.51	99.88	99.71	99.71	99.41	99.57	99.62
3	✓	×	✓	99.37	99.41	99.77	99.59	99.77	99.58	99.51	99.52
4	×	✓	✓	99.31	99.32	99.62	99.43	99.69	99.47	99.20	99.24
5	✓	✓	✓	**99.57**	**99.57**	**99.90**	**99.80**	**99.90**	**99.85**	**99.87**	**99.70**

and 99.39%, respectively. While these algorithms are relatively effective, they do not match the exceptional performance demonstrated by DDFE-ASFS. The robustness of the DDFE-ASFS algorithm in handling various spectral data, particularly in more complex scenarios, sets it apart from the others.

4.2. Ablation Study

To validate the effectiveness of each module in the proposed model, we conduct ablation experiments on the Houston2013, Pavia University, WHU-Hi-HanChuan, and WHU-Hi-HongHu datasets to evaluate the contributions of each module in the proposed DDFE-ASFS framework, which includes the DDFE, EASSF, and SGCO modules. The experimental results are summarized in Table 7. The lowest classification accuracy across all four datasets is observed when using only the DDFE module. In the second

Table 8. The running time in seconds (s) between the contrast methods and the proposed method on four datasets (optimal results are bolded)

Methods	Houston2013	Pavia University	WHU-Hi-HanChuan	WHU-Hi-HongHu
	Running time(s)	Running time(s)	Running time(s)	Running time(s)
SSFTT	50.78	114.30	664.30	544.30
morphFormer	41.06	107.82	494.13	709.35
LSGA	102.24	**49.80**	367.36	512.72
DCTN	49.66	68.35	423.12	504.68
MSSTT	56.04	216.60	401.10	746.45
RDTN	108.17	248.31	513.56	656.87
MASSFormer	60.62	87.63	505.77	732.37
LSFAT	107.76	61.52	382.24	532.72
DBCT	101.17	71.78	448.37	612.89
DDFE-ASFS	**38.56**	200.05	**339.21**	**488.73**

scenario, adding the EASSF module results in improved classification performance compared to the first scenario. Specifically, on the Houston2013 dataset, OA increased by 0.21%, and AA increased by 0.21%. On the WHU-Hi-HongHu dataset, OA increased by 0.47%, and AA increased by 0.41%. These results show that the inclusion of the EASSF module enhances the classifier's recognition performance. This indicates that integrating high-level spectral-spatial information and deep semantic information helps to more accurately classify different land cover types. In the third scenario, the addition of the SGCO module leads to a more significant improvement in classification accuracy compared to the first scenario. For instance, on the Pavia University dataset, OA increased by 0.36%, and AA increased by 0.25%. On the WHU-Hi-HongHu dataset, OA increased by 0.13%, and AA increased by 0.48%. These results demonstrate that the SGCO module also contributes to the classifier's improved recognition performance. Finally, comparing the fifth and fourth scenarios shows that the classification accuracy reached its optimal level when all modules (DDFE, EASSF, and SGCO) were included. Specifically, on the Houston2013 dataset, OA increased by 0.26%, and AA increased by 0.25%. On the WHU-Hi-HongHu dataset, OA increased by 0.67%, and AA increased by 0.46%. These findings highlight that the excellent overall classification performance of the proposed framework is primarily attributed to the integration of the DDFE, EASSF, and SGCO modules.

4.3. Time Cost Comparison

A comparison of running time for SSFTT, morphFormer, LSGA, DCTN, MSSTT, RDTN, MASSFormer, LSFAT, DBCT, and the proposed DDFE-ASFS is presented in Table 8. From Table 8, it can be seen that the DDFE-ASFS model has the shortest execution time among the three datasets. In general, the main indicators for evaluating the robustness of classification algorithms are classification accuracy and execution

time. By comparing Tables 3, 4, 5, and 6, it can be found that the DDFE-ASFS model performs the best in OA, AA, and κ values, indicating that the model has strong robustness.

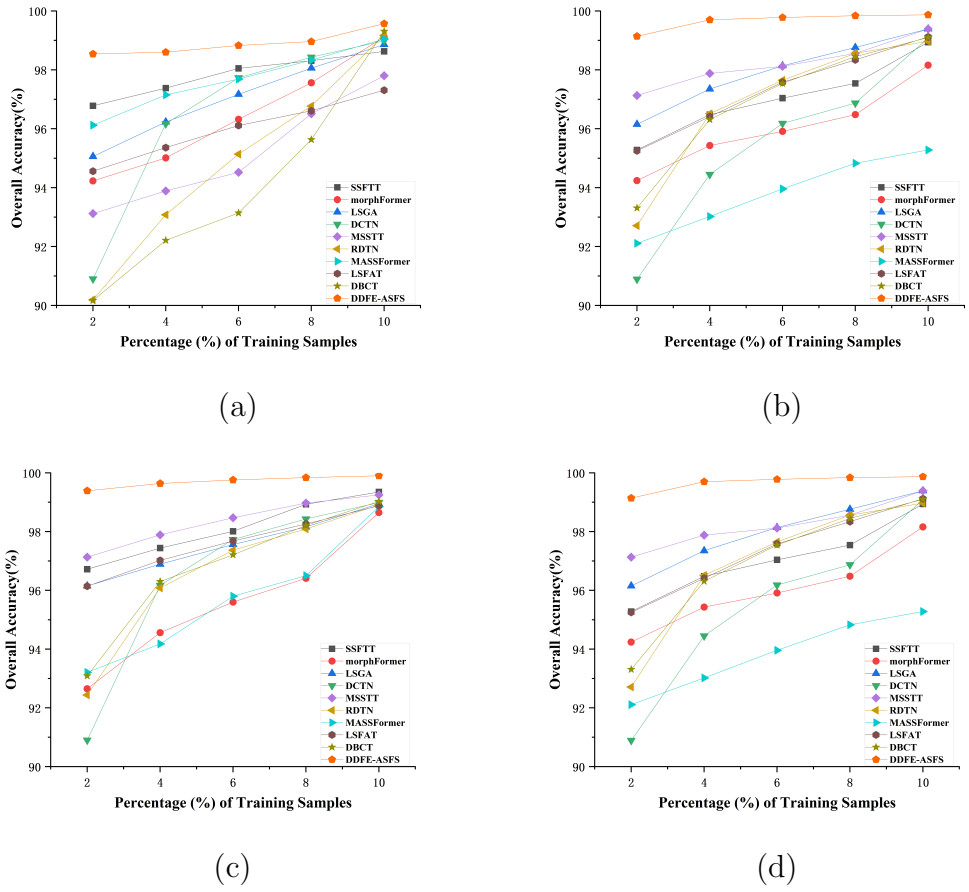


Figure 15. Impact of varied training sizes on overall accuracy. (a) Houston2013, (b) Pavia University, (c) WHU-Hi-HanChuan, (d) WHU-Hi-HongHu.

4.4. Impact of training sample proportion on overall accuracy

Fig. 15 illustrates the variations in classification accuracy across four datasets when different algorithms and training sample ratios are used. The proposed DDFE-ASFS algorithm demonstrates outstanding performance on all four datasets. Specifically, when the training sample ratio ranges from 2% to 10%, DDFE-ASFS consistently achieves high OA values, highlighting its exceptional performance, generalization ability, and stability. First, on the Houston2013 dataset, DDFE-ASFS exhibits superior performance across various training sample ratios. Even with a low training sample ratio, the algorithm achieves high classification accuracy, and as the training sample size increases, the classification accuracy steadily improves. In contrast, algorithms such as DBCT, RDTN, and DCTN perform well at higher sample ratios but exhibit instability at lower sample ratios. This phenomenon can be attributed to the relatively narrow inter-class

separations in the Houston2013 and Pavia University datasets, where the distribution of pixels within the same class tends to be more dispersed. The combination of narrow class boundaries and wide pixel distributions within each class can lead to a large number of mixed pixels, especially in boundary regions. These mixed pixels make accurate land cover classification more challenging, particularly when the available sample size is limited. Next, on the Pavia University dataset, while the accuracies of different algorithms vary, all methods exhibit a marked improvement in performance as the training sample ratio increases. For instance, DDFE-ASFS achieves 98.1% OA with a 2% training sample ratio and 99.90% OA with a 10% training sample ratio, demonstrating its exceptional ability to handle limited labeled data. Other algorithms, such as MorphFormer and RDTN, generally perform poorly at lower sample ratios, but their accuracy improves considerably as the training sample size increases. For the WHU-HI-HanChuan dataset, MSSTT and DDFE-ASFS stand out, achieving OA values of 99.25% and 99.90%, respectively, at a 10% training sample ratio, confirming their advantages in handling high-dimensional data. The MASSFormer algorithm shows subpar performance at lower training sample ratios, particularly at 2% and 4%, with OA of 93.21% and 94.18%, respectively. In the MASSFormer framework, shallow CNNs are used for spectral-spatial feature extraction but are limited in capturing high-level, complex features. This inability to extract deeper, abstract features from hyperspectral data hampers the model's performance, leading to suboptimal feature representations and reduced classification accuracy. The shallow depth of the feature extraction network is a key factor in the accuracy degradation observed in MASSFormer. Finally, on the WHU-Hi-HongHu dataset, although LSGA and SSFTT exhibit some stability, their initial accuracy at lower sample ratios is relatively low. This could be due to their need for larger datasets to effectively learn complex features. When the sample size is small, this limitation negatively affects their performance. This may be because their model designs are not optimized for HSIs, leading to challenges in extracting effective features when data is scarce. Notably, DDFE-ASFS outperforms all other algorithms at every training sample ratio, achieving an exceptional accuracy of 99.87% at a 10% training sample ratio, underscoring its superior feature extraction and learning capability. This comprehensive analysis highlights the robustness and efficiency of the DDFE-ASFS algorithm in various scenarios, particularly when dealing with limited training data and high-dimensional HSIs.

5. Conclusion

In this study, we introduce a novel DDFE-ASFS model aimed at enhancing the accuracy and efficiency of HSI classification. By introducing three core modules, DDFE, EASSF, and SGCO, the proposed classification architecture not only effectively obtains rich local-global spatial-spectral and frequency domain features of hyperspectral datasets but also consumes relatively short execution time. First, DDFE helps capture and integrate global and local feature information simultaneously, enabling models to recognize

fine-grained spectral and spatial details. Then, EASSF dynamically fuses spectral features with spatial features by using two attention mechanisms, thus enhancing the representation ability of the model in spectral and spatial dimensions. Finally, SGCO characterizes long-range dependencies, thereby enhancing both the classification efficiency and accuracy. Through comprehensive experimental evaluation on four HSIs, we demonstrate that the DDFE-ASFS model achieves classification performance on par with several current state-of-the-art methods while offering certain extent improvements in computational efficiency and resource utilization.

Acknowledgements

The authors would like to thank the HSI Analysis group and the NSF Funded Center for Airborne Laser Mapping (NCALM) at the University of Houston for providing the HU dataset used in this work.

Funding

This study was funded by the Scientific Research Fund of Hunan Provincial Education Department (No.23B0666), the Science and Technology Innovation Program of Hunan Province (2016TP1020), the Science Foundation of Hengyang Normal University (2022QD07), the 14th Five-Year Plan Key Disciplines and Application-oriented Special Disciplines of Hunan Province (Xiangjiaotong [2022] 351).

Conflicts of interest

No potential conflict of interest was reported by the authors.

References

[1] Xia Yue, Anfeng Liu, Ning Chen, Shaobo Xia, Jun Yue, and Leyuan Fang. Hypermll: Toward robust hyperspectral image classification with multi-source label learning. *IEEE Transactions on Geoscience and Remote Sensing*, 62:5526515, 2024.

[2] Yimin Zhu, Kexin Yuan, Wenlong Zhong, and Linlin Xu. Spatial–spectral convnext for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:5453–5463, 2023.

[3] Hao Zhou, Fulin Luo, Huiping Zhuang, Zhenyu Weng, Xiuwen Gong, and Zhiping Lin. Attention multihop graph and multiscale convolutional fusion network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–14, 2023.

[4] Nizom Farmonov, Khilola Amankulova, József Szatmári, Alireza Sharifi, Dariush Abbasi-Moghadam, Seyed Mahdi Mirhoseini Nejad, and László Mucsi. Crop type classification by desis hyperspectral imagery and machine learning algorithms. *IEEE Journal of selected topics in applied earth observations and remote sensing*, 16:1576–1588, 2023.

[5] Leyuan Fang, Yifan Jiang, Yinglong Yan, Jun Yue, and Yue Deng. Hyperspectral image instance segmentation using spectral–spatial feature pyramid network. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–13, 2023.

[6] Zhiyong Lv, Pengfei Zhang, Weiwei Sun, Jón Atli Benediktsson, Junhuai Li, and Wei Wang. Novel adaptive region spectral–spatial features for land cover classification with high spatial resolution remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–12, 2023.

[7] Ping Ma, Jinchang Ren, Genyun Sun, Huimin Zhao, Xiuping Jia, Yijun Yan, and Jaime Zabalza. Multiscale superpixelwise prophet model for noise-robust feature extraction in hyperspectral images. *IEEE transactions on geoscience and remote sensing*, 61:1–12, 2023.

[8] Hojat Shirmard, Ehsan Farahbakhsh, R Dietmar Müller, and Rohitash Chandra. A review of machine learning in processing remote sensing data for mineral exploration. *Remote Sensing of Environment*, 268:112750, 2022.

[9] Sen Jia, Shuangzhao Zhu, Zhihao Wang, Meng Xu, Weixi Wang, and Yujuan Guo. Diffused convolutional neural network for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023.

[10] Jean-Pierre Ardouin, Josée Lévesque, and Terry A Rea. A demonstration of hyperspectral image exploitation for military applications. In *2007 10th International Conference on Information Fusion*, pages 1–8. IEEE, 2007.

[11] Ke Wu, Jiayuan Fan, Peng Ye, and Mingzhen Zhu. Hyperspectral image classification using spectral–spatial token enhanced transformer with hash-based positional embedding. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–16, 2023.

[12] Heng-Chao Li, Zhi-Xin Lin, Tian-Yu Ma, Xi-Le Zhao, Antonio Plaza, and William J Emery. Hybrid fully connected tensorized compression network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–16, 2023.

[13] Ting Lu, Mengkai Liu, Wei Fu, and Xudong Kang. Grouped multi-attention network for hyperspectral image spectral-spatial classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–12, 2023.

[14] Yishu Peng, Yaru Liu, Bing Tu, and Yuwen Zhang. Convolutional transformer-based few-shot learning for cross-domain hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:1335–1349, 2023.

[15] Qingwang Wang, Jiangbo Huang, Tao Shen, and Yanfeng Gu. Ehgnn: Enhanced hypergraph neural network for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5, 2024.

[16] Jiansheng Wang, Xintian Mao, Yan Wang, Xiang Tao, Junhao Chu, and Qingli Li. Automatic generation of pathological benchmark dataset from hyperspectral images of double stained tissues. *Optics & Laser Technology*, 163:109331, 2023.

[17] Artur Miroszewski, Jakub Mielczarek, Grzegorz Czelusta, Filip Szczepanek, Bartosz Grabowski, Bertrand Le Saux, and Jakub Nalepa. Detecting clouds in multispectral satellite images using quantum-kernel support vector machines. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:7601–7613, 2023.

[18] Jiaxin Lu, Ling Han, Lei Liu, Junfeng Wang, Zhaode Xia, Dingjian Jin, and Xinlin Zha. Lithology classification in semi-arid area combining multi-source remote sensing images using support vector machine optimized by improved particle swarm algorithm. *International Journal of Applied Earth Observation and Geoinformation*, 119:103318, 2023.

[19] Joy Sim, Yash Dixit, Cushla Mcgoverin, Indrawati Oey, Russell Frew, Marlon M Reis, and Biniam Kebede. Support vector regression for prediction of stable isotopes and trace elements using hyperspectral imaging on coffee for origin verification. *Food research international*, 174:113518, 2023.

[20] Hongyuan Zha, Yujun Guo, Yicen Liu, Xueqin Zhang, Song Xiao, Guoqiang Gao, and Guangning Wu. The characteristic analysis of esdd and nsdd detection of composite insulators based on hyperspectral technology. *IEEE Transactions on Instrumentation and Measurement*, 72:1–8, 2023.

[21] Md Hafiz Ahamed, Md Ali Hossain, and Yeahia Sarker. Dynamic kernel network for hyperspectral image classification. *International Journal of Remote Sensing*, 44(9):2847–2866, 2023.

[22] Weiwei Liu, Kai Liu, Weiwei Sun, Gang Yang, Kai Ren, Xiangchao Meng, and Jiangtao Peng. Self-supervised feature learning based on spectral masking for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023.

[23] Chang Liu, Yinpeng Dong, Wenzhao Xiang, Xiao Yang, Hang Su, Jun Zhu, Yuefeng Chen, Yuan He, Hui Xue, and Shibao Zheng. A comprehensive study on robustness of image classification models: Benchmarking and rethinking. *International Journal of Computer Vision*, 133(2):567–589, 2025.

[24] Jun-Jie Pei, Li-Hua Gong, Li-Guo Qin, and Nan-Run Zhou. One-to-many image generation model based on parameterized quantum circuits. *Digital Signal Processing*, 168:105340, 2025.

[25] Behnood Rasti, Danfeng Hong, Renlong Hang, Pedram Ghamisi, Xudong Kang, Jocelyn Chanussot, and Jon Atli Benediktsson. Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox. *IEEE Geoscience and Remote Sensing Magazine*, 8(4):60–88, 2020.

[26] Xiangrong Zhang, Yanjie Liang, Chen Li, Ning Huyan, Licheng Jiao, and Huiyu Zhou. Recursive autoencoders-based unsupervised feature learning for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 14(11):1928–1932, 2017.

[27] Random Fields. Hyperspectral image classification with markov random fields and a convolutional neural network. *Learning*, 42:1109.

[28] Hyungtae Lee and Heesung Kwon. Going deeper with contextual cnn for hyperspectral image classification. *IEEE Transactions on Image Processing*, 26(10):4843–4855, 2017.

[29] Swalpa Kumar Roy, Gopal Krishna, Shiv Ram Dubey, and Bidyut B Chaudhuri. Hybridsn: Exploring 3-d–2-d cnn feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 17(2):277–281, 2019.

[30] Zilong Zhong, Jonathan Li, Zhiming Luo, and Michael Chapman. Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2):847–858, 2017.

[31] Mercedes E Paoletti, Juan Mario Haut, Ruben Fernandez-Beltran, Javier Plaza, Antonio J Plaza, and Filiberto Pla. Deep pyramidal residual networks for spectral–spatial hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2):740–754, 2018.

[32] Yicheng Hu, Shufang Tian, and Jia Ge. Hybrid convolutional network combining multiscale 3d depthwise separable convolution and cbam residual dilated convolution for hyperspectral image classification. *Remote Sensing*, 15(19):4796, 2023.

[33] Haokui Zhang, Yu Liu, Bei Fang, Ying Li, Lingqiao Liu, and Ian Reid. Hyperspectral classification based on 3d asymmetric inception network with data fusion transfer learning. *arXiv preprint arXiv:2002.04227*, 2020.

[34] Zhuoyi Zhao, Xiang Xu, Shutao Li, and Antonio Plaza. Hyperspectral image classification using groupwise separable convolutional vision transformer network. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–17, 2024.

[35] Zhenqiu Shu, Yuyang Wang, and Zhengtao Yu. Dual attention transformer network for hyperspectral image classification. *Engineering Applications of Artificial Intelligence*, 127:107351, 2024.

[36] Fulin Xu, Shaohui Mei, Ge Zhang, Nan Wang, and Qian Du. Bridging cnn and transformer with cross attention fusion network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024.

[37] Le Sun, Guangrui Zhao, Yuhui Zheng, and Zebin Wu. Spectral–spatial feature tokenization transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.

[38] Chao Ma, Minjie Wan, Jian Wu, Xiaofang Kong, Ajun Shao, Fan Wang, Qian Chen, and Guohua Gu. Light self-gaussian-attention vision transformer for hyperspectral image classification. *IEEE transactions on instrumentation and measurement*, 72:1–12, 2023.

[39] Bing Tu, Xiaolong Liao, Qianming Li, Yishu Peng, and Antonio Plaza. Local semantic feature

aggregation-based transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022.

[40] Le Sun, Hang Zhang, Yuhui Zheng, Zebin Wu, Zhonglin Ye, and Haixing Zhao. Massformer: Memory-augmented spectral-spatial transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024.

[41] Zhe Meng, Taizheng Zhang, Feng Zhao, Gaige Chen, and Miaomiao Liang. Multi-scale super token transformer for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 21:5508105, 2024.

[42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[43] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[44] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11534–11542, 2020.

[45] Ying Cui, Zikun Yu, Jiacheng Han, Shan Gao, and Liguo Wang. Dual-triple attention network for hyperspectral image classification using limited training samples. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2021.

[46] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3146–3154, 2019.

[47] Rui Li, Shunyi Zheng, Chenxi Duan, Yang Yang, and Xiqi Wang. Classification of hyperspectral image based on double-branch dual-attention mechanism network. *Remote Sensing*, 12(3):582, 2020.

[48] Md Palash Uddin, Md Al Mamun, and Md Ali Hossain. Pca-based feature reduction for hyperspectral remote sensing image classification. *IETE Technical Review*, 38(4):377–396, 2021.

[49] Hao Shi, Guo Cao, Youqiang Zhang, Zixian Ge, Yanbo Liu, and Di Yang. F 3 net: Fast fourier filter network for hyperspectral image classification. *IEEE Transactions on Instrumentation and Measurement*, 72:1–18, 2023.

[50] Anish Sarkar, Utpal Nandi, Moirangthem Marjit Singh, Bachchu Paul, and Rajasekaran Selvaraju. Finding efficient activation functions for deep learning based hyperspectral image classification. In *International Conference on Data Analytics and Insights*, volume 1233, pages 283–295. Springer, 2024.

[51] Swalpa Kumar Roy, Ankur Deria, Chiranjibi Shah, Juan M Haut, Qian Du, and Antonio Plaza. Spectral–spatial morphological attention transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023.

[52] Yunfei Zhou, Xiaohui Huang, Xiaofei Yang, Jiangtao Peng, and Yifang Ban. Dctn: Dual-branch convolutional transformer network with efficient interactive self-attention for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–16, 2024.

[53] Yan Li, Xiaofei Yang, Dong Tang, and Zheng Zhou. Rdtn: Residual densely transformer network for hyperspectral image classification. *Expert Systems with Applications*, 250:123939, 2024.

[54] Le Sun, Hang Zhang, Yuhui Zheng, Zebin Wu, Zhonglin Ye, and Haixing Zhao. Massformer: Memory-augmented spectral-spatial transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024.

[55] Rui Xu, Xue-Mei Dong, Weijie Li, Jiangtao Peng, Weiwei Sun, and Yi Xu. Dbctnet: Double branch convolution-transformer network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024.