Expert Systems With Applications 296 (2026) 129198

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

HorD^{2}CN: High-order deformable differential convolution network for hyperspectral image classification

Zitong Zhang[a,b], Fujie Jiang[a,b,*], Chengcheng Zhong[c], Qiaoyu Ma[c]

[a] State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), 102249, Beijing, China
[b] College of Geoscience, China University of Petroleum (Beijing), 102249, Beijing, China
[c] College of Information and Electrical Engineering, China Agricultural University, 100083, Beijing, China

ARTICLE INFO

Keywords:
Hyperspectral image classification
High-order feature interaction
Differential convolution
Deformable operation
Deep learning

ABSTRACT

Hyperspectral image (HSI) classification confronts significant challenges in modeling high-order spectral-spatial interactions and adaptively capturing fine-grained structural details, as existing deep learning methods typically suffer from inherent limitations in modeling nonlinear high-order features and reliance on fixed spatial sampling that fails to adapt to complex geometric variations of ground objects. To address these limitations, we propose a novel high-order deformable differential convolution network (HorD^{2}CN) that enables explicit modeling of high-order feature interactions and adaptive capture of spatial structural details. The core innovation lies in the design of a high-order multi-scale differential convolution (HorMSDC) block, which is engineered to enhance the extraction of complex high-order patterns from HSI data and facilitate the representation of discriminative spectral-spatial information. This block integrates two synergistic modules: the high-order spectral differential convolution (HSEDC) module, which performs 1D high-order differential convolution via an adaptive spectral shift (ASES) operation to capture subtle spectral band variations between distinct land cover types and enhance fine-grained feature discriminability, and the high-order spatial differential convolution (HSADC) module, which employs 2D differential convolution with a deformable spatial shift (DSAS) operation to strengthen the modeling of multi-scale spatial structural details. By integrating these modules, HorD^{2}CN enables adaptive extraction of high-order spectral-spatial features. Comprehensive experiments on five benchmark HSI datasets demonstrate that HorD^{2}CN outperforms ten state-of-the-art deep learning methods, validating its effectiveness in HSI classification tasks.

1. Introduction

Hyperspectral image (HSI) data provide rich spectral and spatial information, enabling a comprehensive characterization of the optical properties in the physical world (Bian et al., 2024). With hundreds to thousands of contiguous spectral bands and superior spatial mapping capabilities, HSI classification involves assigning labels to individual pixels based on their high-dimensional spectral-spatial features, allowing for precise discrimination of materials with similar visual characteristics. This capability has established HSI classification as a powerful tool in diverse applications, including environmental monitoring (Camps-Valls et al., 2013), agriculture (Wang et al., 2021), and land use management (Qin et al., 2024a).

Over the past few years, scholars have made great efforts to extract discriminative spectral-spatial features and enhance HSI classification performance. Early HSI classification primarily relied on traditional machine learning methods, such as support vector machine (SVM) and random forest (RF) (Hasan et al., 2019; Zhang et al., 2018). However, due to their dependence on manually crafted features and limitations of shallow learning capabilities, HSI classification has gradually transitioned into the era of deep learning (DL) (Feng et al., 2024). The success of DL in computer vision has been demonstrated across diverse classification tasks (Kaur et al., 2023; Verma & Yadav, 2025), prompting widespread interest in using DL-based methods to automatically extract rich spectral-spatial features from raw HSI data.

In particular, the past three years have witnessed rapid advancements in transformer-based, deep MLP-based, and state-space models for HSI classification. Currently, the most popular and state-of-the-art DL methods used in HSI classification can be categorized into the following branches: convolutional neural network (CNN) (Torun et al., 2024), vision transformer (ViT) (Zhou et al., 2023), MLP-Mixer (Tolstikhin et al., 2021), and Mamba (Gu & Dao, 2023).

[*] Corresponding author.
E-mail addresses: 3001200116@email.cugb.edu.cn (Z. Zhang), jiangfj@cup.edu.cn (F. Jiang), b20233080710@cau.edu.cn (C. Zhong), b20223080660@cau.edu.cn (Q. Ma).

https://doi.org/10.1016/j.eswa.2025.129198
Received 19 February 2025; Received in revised form 17 June 2025; Accepted 27 July 2025
Available online 31 July 2025

Among these methods, CNNs have emerged as pioneering tools, effectively capturing spatial patterns in HSI data through local receptive fields and shared weights. Traditional 2D-CNNs focus on spatial patterns, while the spectral domain of HSI data is equally crucial for effective analysis. Consequently, 1D-CNNs and 3D-CNNs have been developed to separately extract spectral features and simultaneously capture spectral-spatial features, respectively (Jijón-Palma et al., 2021; Li et al., 2022). More sophisticated network architectures have since been proposed to further enhance HSI classification performance. For instance, Zhong et al. (2023) introduced the lightweight criss-cross large kernel convolutional neural network (LiteCCLKNet) for HSI classification, aggregating long-range contextual features efficiently. Zhang et al. (2024c) designed a spectral-spatial difference convolution network (S^2DCN), integrating the difference principle into the deep learning framework and utilizing a learnable gradient encoding pattern to extract detailed features in spectral and spatial domains. Meanwhile, Zhang et al. (2024b) developed a tree-shaped multiobjective evolutionary CNN (TMOE-CNN), organizing the convolution operation hierarchically to extract features of different types and scales.

The advent of attention mechanisms has significantly advanced HSI classification by enabling models to focus on the most informative regions across both spectral and spatial domains (Qin et al., 2024b; Waswani et al., 2017). Building upon the success of CNNs in capturing local patterns, attention mechanisms offer a complementary approach by emphasizing global dependencies and long-range interactions. Inspired by human visual attention, ViT employs a self-attention mechanism that allows each image patch to interact with others, learning the relative importance of various parts of the image and effectively capturing global information (Dosovitskiy et al., 2021). ViT-based methods have shown promise in learning long-range dependencies and capturing non-local relationships for HSI classification (Ahmad et al., 2024a). Zhao et al. (2024) proposed a groupwise separable convolutional ViT (GSC-ViT), combining groupwise separable convolution for local spectral-spatial feature extraction with groupwise separable multi-head self-attention for both local and global spatial feature extraction. This architecture reduced model complexity while maintaining strong classification performance. Ahmad et al. (2024b) employed a spectral-spatial wavelet transformer (WaveFormer), utilizing wavelet transforms for invertible downsampling to preserve data integrity and enable enhanced attention learning.

The MLP-Mixer model, a novel architecture based on multilayer perceptrons (MLPs), has recently emerged as a competitive approach for HSI classification. Unlike traditional CNNs or ViTs, MLP-Mixer leverages a simple yet powerful structure that does not rely on convolutions or self-attention mechanisms. Instead, it processes spatial and channel-wise features using MLPs in separate stages (Tolstikhin et al., 2021). The key innovation of MLP-Mixer is the decoupling of spatial and channel interactions, where the model first mixes features within each spatial location and then mixes features across channels, enabling it to capture both local and global dependencies effectively. Zhang et al. (2024a) proposed a learnable dilated spectral-spatial MLP (LDS^{2}MLP), introducing a learnable dilated receptive field and grouped MLP to extract discriminative spectral-spatial features at different scales and promote feature interaction. Shao et al. (2022) designed a spatial-spectral involution MLP network (SSIN), combining an involution MLP in the image path to improve spatial interaction and a coordinate path to incorporate global spatial distribution, thereby enhancing long-distance dependencies.

Recently, Mamba (Gu & Dao, 2023) has developed into a promising alternative to transformers by further improving structured state-space sequence models (S4) (Gu et al., 2021) with a selective mechanism. Benefiting from its strong long-distance modeling capabilities while maintaining linear computational complexity, Mamba has received substantial research across many fields, such as image segmentation (Li et al., 2025; Zhang et al., 2025b), image extrapolation (Zhang et al., 2025a), and emotion recognition (Yang et al., 2024). Taking cues from the success of Mamba, various Mamba variants have been proposed for HSI classification, such as S^2Mamba (Wang et al., 2025), MambaHSI (Li et al., 2024), and HyperMamba (Liu et al., 2024), demonstrating the superiority and prospect of Mamba.

However, the nature of HSI data presents unique challenges that demand specialized approaches for effective feature extraction and classification. High spectral resolution means that the satellite is capable of capturing hundreds of bands of the same spatial area, and the wavelength for each band is typically narrow (Aburaed et al., 2023). The first challenge that follows is that the rich spectral-spatial information in HSI contains subtle variations, which are crucial for distinguishing similar materials, posing the requirement to enhance sensitivity to subtle spectral and spatial variations. Second, HSI data exhibit significant high-order statistical structures, which go beyond simple first-order (e.g., pixel values or mean values) and second-order (e.g., variance or covariance) representations and require advanced modeling techniques to capture (Chang et al., 2014). Third, the spatial variability of HSI motivates the need for deformable or adaptive spatial modeling, as fixed-kernel operations are inadequate for capturing non-uniform spatial dependencies in complex HSI data (Fang et al., 2025).

While ViT-, deep MLP-, and Mamba-based models have improved global feature representation, they still suffer from insufficient modeling of high-order spectral-spatial interactions beyond second-order statistics and fixed sampling kernels that fail to adapt to complex geometric deformations of ground objects. For example, existing ViT variants rely on fixed patch partitioning, while Mamba-based models often overlook fine-grained spectral-spatial differences.

To address these challenges, we propose a novel high-order deformable differential convolution network (HorD^{2}CN) that integrates high-order differential convolution and deformable operations to enhance the feature extraction, ensuring robust and accurate classification even with limited labeled data. The main contributions of this paper are as follows.

- A HorD^{2}CN is proposed for HSI classification, which significantly enhances high-order feature interactions in HSI data.
- A high-order multi-scale differential convolution (HorMSDC) block is introduced to explicitly model the intricate high-order features inherent in HSI data and enhances the sensitivity of the model to capture subtle spectral-spatial variations by emphasizing pixel-wise differences.
- The deformable operations are integrated into the HorMSDC block to adaptively adjust sampling positions, extracting complex spectral patterns and spatial geometric structural features.
- Comprehensive experiments are conducted on five benchmark HSI datasets, and the experimental results demonstrate that the proposed HorD^{2}CN achieves competitive performance compared to ten state-of-the-art methods in the field.

The structure of this paper is organized as follows. Section 2 reviews related work on high-order feature interaction, differential convolution, and deformable operations. Section 3 details the proposed HorD^{2}CN. Section 4 presents a comprehensive analysis of experimental results, comparing the performance of HorD^{2}CN with ten deep learning methods across five HSI datasets. Finally, Section 5 concludes the paper.

2. Related works

2.1. High-order feature interaction

High-order feature interactions in images refer to the complex and nonlinear relationships between different features within an image, which result in higher-level semantic information or more refined feature representations. Unlike first-order (linear) or second-order (simple nonlinear) interactions, high-order feature interactions involve intricate dependencies among multiple features, capturing deeper contextual information and inter-feature relationships. While there exist complex and often high-order interactions between two spatial locations in a deep model due to the non-linearity, the success of self-attention and other dynamic networks highlights the benefits of explicitly incorporating high-order spatial interactions, thereby enhancing the modeling capacity of vision models (Rao et al., 2022).

Currently, approaches to realizing high-order feature interactions can be broadly classified into two categories. One approach involves the derivation of local pattern descriptors. For example, Liu et al. (2019) proposed an improved derived mean complete local binary pattern (DM CLBP) based on high-order derivatives, capturing imaging information from the concave-convex regions between the high-order pixel and the neighboring sampling points. Zhang et al. (2010) introduced a local derivative pattern (LDP) to extract high-order local information by encoding various distinctive spatial relationships contained in a given local region. Similarly, Fan and Hung (2014) presented a local vector pattern (LVP) in the high-order derivative space for face recognition, effectively extracting 1D and 2D texture features through constructing vectors based on pixel differences and encoding vector pairwise directions via comparative space transform.

The second approach leverages deep learning methods to model high-order feature interactions. For instance, Xie et al. (2021) employed a broad attentive graph fusion network (BaGFN) to better model high-order feature interactions in a flexible and explicit manner. Rao et al. (2022) designed a HorNet to perform high-order spatial interactions through gated convolutions and recursive mechanisms. Additionally, Jiang et al. (2023) introduced the idea of high-order feature extraction and interaction into the HSI classification task and proposed a spectral-spatial multi-order interaction network (S^2MoINet), capturing high-order and generalized features to improve classification accuracy.

2.2. Differential convolution

Differential convolution (Yu et al., 2020a,b), inspired by the differential principles of the traditional Local Binary Pattern (LBP) operator (Ojala et al., 2002), addresses the limitations of standard convolutions in effectively capturing fine-grained image details. Unlike conventional convolution, which directly applies fixed kernels to extract features, differential convolution emphasizes the differences between neighboring pixels and the central pixel within a local patch. Specifically, it computes the difference between each surrounding element and the center pixel, followed by applying a convolution kernel to highlight subtle variations or structural details. This approach has gained significant traction in computer vision due to its ability to enhance sensitivity to high-frequency information, as illustrated in Fig. 1.

By focusing on pixel-wise differences, differential convolution excels at capturing intricate local patterns, making it particularly effective for tasks requiring fine detail extraction. Its variants have demonstrated remarkable success across diverse applications, including face anti-spoofing (Yu et al., 2020b), edge detection (Su et al., 2021), and video gesture/action recognition (Yu et al., 2021).

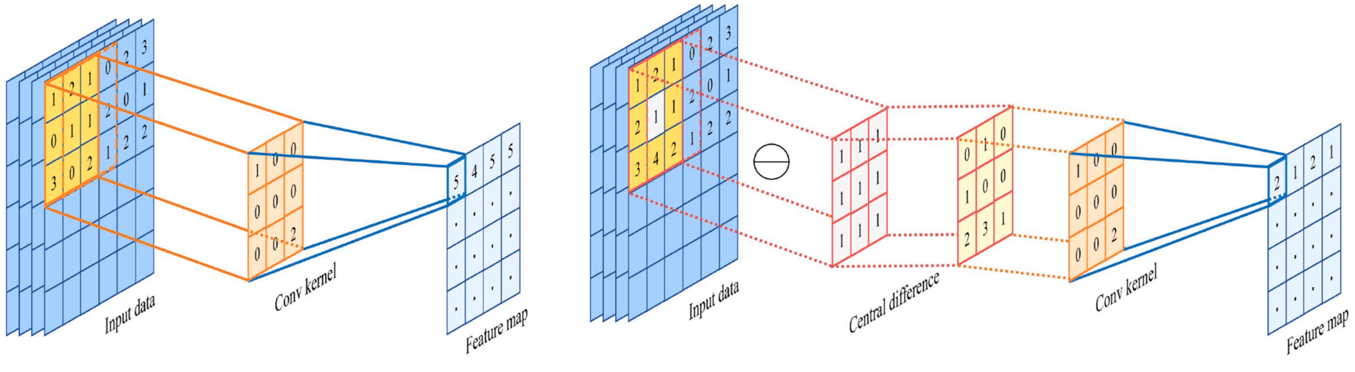
In the context of HSI classification, differential convolution has proven advantageous due to its ability to model subtle spectral variations and spatial structural details across different land cover types. For instance, Zhang et al. (2024c) introduced the spectral-spatial difference convolution network (S^2DCN) for HSI classification, leveraging a learnable gradient encoding mechanism to extract discriminative features in both spectral and spatial domains. This underscores the potential of differential convolution as a powerful tool for processing high-dimensional HSI data, warranting further exploration in this field.

2.3. Deformable operation

Deformable operation marks a significant advancement in deep learning by enabling dynamic adjustment of spatial sampling positions, thereby improving the model's ability to capture complex geometric shapes and structures (Zhang et al., 2024a). Among these techniques, deformable convolution has become the most widely adopted method (Dai et al., 2017; Zhu et al., 2019). By incorporating learnable offsets into convolution kernels, deformable convolution can adaptively adjust to geometric variations in the input data, significantly enhancing the model's feature extraction capabilities and robustness. This adaptability is particularly essential when dealing with data containing intricate or irregular patterns, where traditional convolutions with fixed kernels often fail to perform effectively.

In HSI classification, the inherently high-dimensional and complex spectral-spatial characteristics present unique challenges for feature extraction. Ground objects in HSI data commonly exhibit diverse shapes, scale variations, and local deformations, making it difficult for conventional models to capture informative features. Deformable operations provide a promising solution by enabling adaptive changes in the receptive field, allowing the model to focus on critical regions and reduce information loss (Zhu et al., 2018). Recent progress in deformable operations has further expanded their applicability in HSI classification. Techniques such as multiscale deformable convolution (Fang et al., 2025; Yang et al., 2023), deformable 3D convolution (Tang et al., 2022), superpixel guided deformable convolution (Zhao et al., 2022), and learnable dilated operation (Zhang et al., 2024a) have significantly improved the modeling of intricate spectral-spatial relationships in HSI data by explicitly addressing its geometric complexity.

Nevertheless, existing studies on deformable operations for HSI classification have primarily emphasized spatial flexibility based on traditional convolution kernels with learnable offsets. Although such methods improve the network's ability to handle irregular object boundaries, they generally operate in low-order feature spaces, which limits their capacity to model more abstract spatial relationships in complex land cover structures. To overcome this limitation, HorD^{2}CN integrates de-



(a) (b)

Fig. 1. Schematic diagrams of the two convolutions. (a) Conventional convolution. (b) Differential convolution.

formable operations into high-order computational flows. This integration enables the model not only to adjust sampling positions adaptively but also to capture rich spectral-spatial dependencies beyond the pixel level.

Building on these advancements, the unified framework HorD^{2}CN combines high-order differential convolution with deformable operations to address the unique challenges of HSI classification. Our approach ensures robust and precise classification, even in scenarios with limited labeled data.

3. Methodology

3.1. Overall

In this section, we propose a network structure named HorD^{2}CN, which aims to improve classification accuracy by effectively extracting and fusing HSI features through an adaptive spatial-spectral differential structure. Fig. 2 illustrates the general classification framework. Its first component is a convolution layer that adjusts the number of channels in the input features, the subsequent feature extraction process consists of a high-order multi-scale differential convolution (HorMSDC) block. Each HorMSDC block includes three parallel branches: a 1 × 1 convolution (conv) branch that captures pixel-level features and achieves dimensionality reduction, a 3 × 3 deformable differential convolution (diffconv) branch that captures medium-scale spatial structures, and a 5 × 5 deformable diffconv branch that captures large-scale spatial structures. Each diffconv branch consists of two modules: the high-order spectral differential convolution (HSEDC) module and the high-order spatial differential convolution (HSADC) module.

The HSEDC module improves the extraction of subtle spectral differences between various ground object types through deformable adaptive spectral shift (ASES). It enables the module to dynamically adapt to different spectral characteristics, thus enhancing the ability to distinguish between different ground objects in the spectral dimension. Meanwhile, the HSADC module leverages differential convolution and deformable spatial shift (DSAS) to enhance spatial context awareness and optimize structural detail modeling. DSAS allows the module to better capture the spatial relationships and structural details of the input features, contributing to a more accurate representation of the spatial information. These modules enable adaptive extraction of geometric deformations and complex patterns in the spatial and spectral domains of the input features. Finally, the multi-scale features obtained from the multiple HorMSDC blocks are concatenated along the channel dimension and passed through a fully connected layer for the final classification.

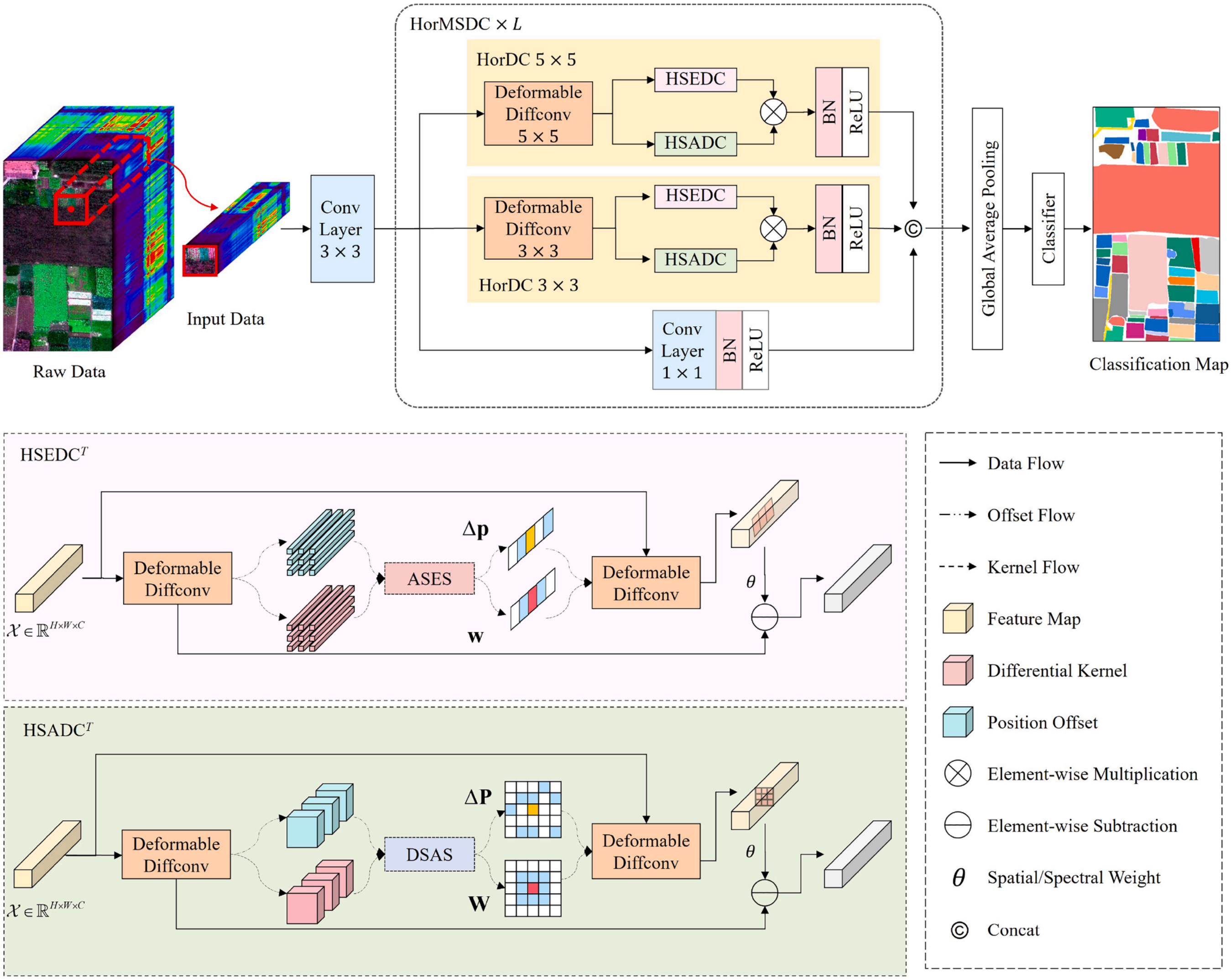


Fig. 2. The architecture of the proposed HorD^{2}CN.

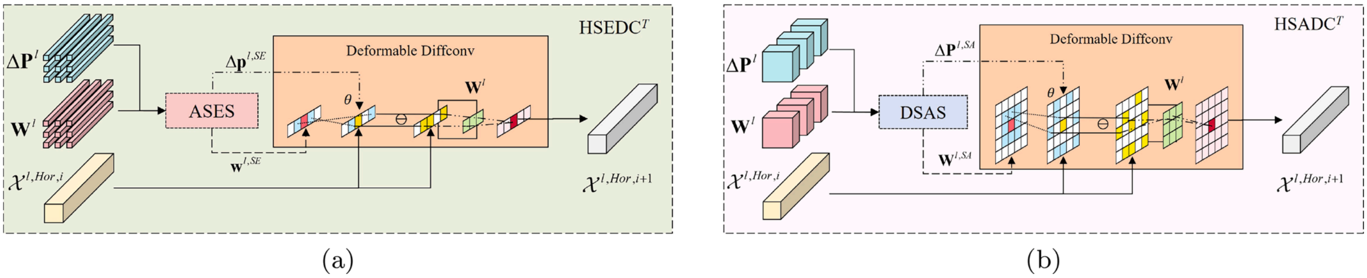


Fig. 3. High-order differential convolution module. (a) HSEDC. (b) HSADC.

3.2. HorMSDC

We define the input HSI data cube as $\mathcal{Z} \in \mathbb{R}^{H\times W\times B}$, where H and W represent the height and width of the HSI data, respectively, and B denotes the number of spectral bands. Let $\mathbf{z}_{h,w}$ represent the pixel located at spatial position (h, w), where $h = 1, 2, \dots, H$, $w = 1, 2, \dots, W$. First, a convolution layer is used to transform the number of spectral bands B of the input data $\mathcal{Z}$ into the internal channel dimension C, allowing for the preliminary extraction of features:

$$\mathcal{X}^0 = \mathrm{Conv}_{3\times 3}^{in}(\mathcal{Z}), \tag{1}$$

where $\mathcal{X}^0 \in \mathbb{R}^{H\times W\times C}$ denotes the features input into the HorMSDC block.

Subsequently, $\mathcal{X}^0$ is processed through multiple HorMSDC blocks. The designed HorMSDC block is capable of simultaneously processing features at multiple scales within a single layer, effectively addressing the limitations in feature extraction that may arise from a single convolution kernel:

$$\mathcal{X}^l = \mathrm{HorMSDC}^l\left(\mathcal{X}^{l-1}\right), l = 1, 2, \dots, L, \tag{2}$$

where $\mathcal{X}^l$ and $\mathcal{X}^{l-1}$ are the output feature maps at the l-th and (l-1)-th block, L represents the number of HorMSDC blocks in the HorD^{2}CN. The model maintains $\mathcal{X}^l \in \mathbb{R}^{H\times W\times C}$ and $\mathcal{X}^{l-1} \in \mathbb{R}^{H\times W\times C}$.

Each HorMSDC block consists of parallel branches via adaptive differential convolution and traditional convolution to integrate multi-scale features:

$$\begin{aligned}\mathrm{HorMSDC}^l\left(\mathcal{X}^{l-1}\right) = \mathrm{Conv}_{1\times1}^{l,out}\Big(\mathrm{Concat}\Big(&\mathrm{Conv}_{1\times1}^{l,in}\left(\mathcal{X}^{l-1}\right),\\ &\mathrm{HorDC}_{3\times3}^l\left(\mathcal{X}^{l-1}\right), \mathrm{HorDC}_{5\times5}^l\left(\mathcal{X}^{l-1}\right)\Big)\Big),\end{aligned} \tag{3}$$

where $\mathrm{HorDC}^l(\cdot)$ denotes the high-order differential convolution operation at the l-th block. In the HorMSDC block, this operation is performed using convolution kernels of size 3×3 and 5×5, respectively, to extract multi-scale spectral-spatial features.

Each HorDC branch processes features via deformable differential convolution, spectral differential convolution, and spatial differential convolution, followed by batch normalization and ReLU:

$$\begin{aligned}\mathrm{HorDC}_{k\times k}^l\left(\mathcal{X}^{l-1}\right) = \mathrm{ReLU}\Big(\mathrm{BN}\Big(&\mathrm{DConv}_{k\times k}^l\left(\mathcal{X}^{l-1}\right) + \mathrm{HSEDC}_{k\times k}^l\left(\mathcal{X}^{l-1}\right)\\ &+\mathrm{HSADC}_{k\times k}^l\left(\mathcal{X}^{l-1}\right)\Big)\Big),\end{aligned} \tag{4}$$

where k represents the convolution kernel size corresponding to the HorDC branch. HSEDC stands for high-order adaptive spectral differential convolution, which is used to extract spectral features from HSI, as shown in Fig. 3a. HSADC refers to high-order adaptive spatial differential convolution, which is employed to extract spatial features from HSI, as illustrated in Fig. 3b. DConv denotes deformable convolution, which provides input feature maps $\mathcal{X}^{l,Hor}$, convolution kernels $\mathcal{W}^l$, bias $\mathcal{B}^l$, and feature position offsets $\Delta\mathcal{P}^l$ to the HSEDC and HSADC modules, as described in Eq. (5).

$$\mathcal{X}^{l,Hor} = \mathrm{DConv}_{k\times k}^l\left(\mathcal{X}^{l-1}\right) = \mathrm{Offset}\left(\mathcal{X}^{l-1}, \Delta\mathcal{P}^l\right)\cdot \mathcal{W}^l + \mathcal{B}^l, \tag{5}$$

where $\mathrm{Offset}(\cdot)$ is the position offset operation, which calculates the new sampling positions for the convolution kernel $\mathcal{W}^l$ on the feature map $\mathcal{X}^{l-1}$ using the offset $\Delta\mathcal{P}^l$. The learnable offset $\Delta\mathcal{P}^l$ is derived from the input features via a dedicated convolution layer:

$$\Delta\mathcal{P}^l = \mathrm{Conv}_{\mathrm{offset}}^l(\mathcal{X}^{l-1}). \tag{6}$$

The $\mathrm{Offset}(\cdot)$ operation computes feature values at non-grid positions by bilinear interpolation of neighboring pixels:

$$\begin{aligned}\mathrm{Offset}_{h,w}\left(\mathcal{X}^{l-1}, \Delta\mathcal{P}^l\right) = &(1-\Delta\mathbf{p}_{h,w}^{l,H})(1-\Delta\mathbf{p}_{h,w}^{l,W})\mathbf{x}_{h,w}^{l-1}\\ &+\Delta\mathbf{p}_{h,w}^{l,W}(1-\Delta\mathbf{p}_{h,w}^{l,H})\mathbf{x}_{h+1,w}^{l-1}\\ &+(1-\Delta\mathbf{p}_{h,w}^{l,W})\Delta\mathbf{p}_{h,w}^{l,H}\mathbf{x}_{h,w+1}^{l-1}\\ &+\Delta\mathbf{p}_{h,w}^{l,W}\Delta\mathbf{p}_{h,w}^{l,H}\mathbf{x}_{h+1,w+1}^{l-1},\end{aligned} \tag{7}$$

where $\mathbf{x}_{h,w}^{l-1}$ denotes the feature vector at the spatial position (h, w) in the feature map. $\Delta\mathbf{p}_{h,w}^{l,H}$ and $\Delta\mathbf{p}_{h,w}^{l,W}$ represent the offset components along the H and W directions, respectively.

The new feature map $\mathcal{X}^{l,Hor}$, convolution kernel $\mathcal{W}^l$, and feature position offset $\Delta\mathcal{P}^l$ obtained from the deformable convolution, are then input into HSEDC and HSADC to extract spectral and spatial features, respectively.

3.3. HSEDC

The designed HSEDC module generates preliminary feature maps by integrating ASES with deformable convolution, simultaneously com-

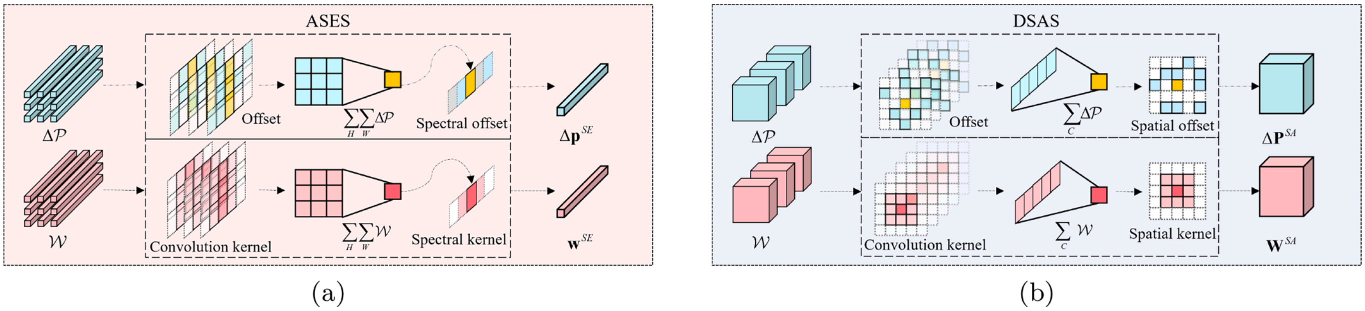


Fig. 4. Differential convolution kernels with offsets. (a) ASES. (b) DSAS.

puting the spectral differential convolution kernel $\mathbf{w}^{l,SE}$ and the corresponding spectral offset $\Delta\mathbf{p}^{l,SE}$ required for HSEDC, as illustrated in Fig. 4a. To intuitively explain the ASES operation, consider a 1D spectral vector: traditional convolution samples fixed positions, whereas ASES dynamically learns offset values to sample shifted positions. This adaptive sampling allows the model to align with subtle material-specific spectral variations. ASES effectively slides the convolution kernel across non-uniform, feature-aware positions along the spectral axis, capturing nuanced changes overlooked by fixed kernels.

First, the convolution kernel and offsets of the deformable convolution are utilized to compute the differential convolution kernel $\mathbf{w}^{l,SE}$ and the differential offset $\Delta\mathbf{p}^{l,SE}$, which are subsequently applied to deformable differential convolution operations at different orders:

$$w_c^{l,SE} = \sum_{i=1}^{k}\sum_{j=1}^{k} w_{i,j,c}^{l}, \tag{8}$$

where $w_c^{l,SE}$ represents the feature value of the differential convolution kernel at the c-th channel, and $w_{i,j,c}^{l}$ denotes the feature value of the deformable convolution weight $\mathcal{W}^l$ at the c-th channel and spatial position (i,j).

$$\Delta p_c^{l,SE} = \sum_{h=1}^{H}\sum_{w=1}^{W} \Delta p_{h,w,c}^{l}, \tag{9}$$

where $\Delta p_c^{l,SE}$ represents the differential offset at the c-th channel, $\Delta p_{h,w,c}^{l}$ denotes the deformable convolution offset $\Delta\mathcal{P}^l$ at the c-th channel and spatial position (h,w). Subsequently, the spectral differential convolution results are accumulated through high-order iterative operations. The results from each order of differential convolution are then combined through accumulation to obtain the high-order features corresponding to the module.

The high-order spectral features are obtained by accumulating differential convolution results through iterative operations:

$$\mathbf{x}_{h,w}^{l,HSEDC} = \mathbf{x}_{h,w}^{l,Hor} - \theta \sum_{t=0}^{T} \mathrm{SEDC}\left(\mathbf{x}_{h,w}^{l,Hor,t}\right), \tag{10}$$

where $\mathrm{SEDC}(\cdot)$ is the channel adaptive differential convolution, T denotes the order. $\mathbf{x}_{h,w}^{l,HSEDC}$ represents the feature vector at spatial position (h,w) of the output feature $\mathcal{X}^{l,HSEDC}$, and $\mathbf{x}_{h,w}^{l,Hor,t}$ is the feature vector computed at spatial position (h,w) during the t-th iteration. When $t=0$, $\mathbf{x}_{h,w}^{l,Hor,t}$ corresponds to the initial input $\mathbf{x}_{h,w}^{l,Hor}$. The calculation of $\mathrm{SEDC}(\cdot)$ is shown in Eq. (11).

$$\mathrm{SEDC}\left(\mathbf{x}_{h,w}^{l,Hor,t}\right) = \mathit{Offset}_E\left(\mathbf{x}_{h,w}^{l,Hor,t-1}, \Delta\mathbf{p}^{l,SE}\right) \cdot \mathbf{w}^{l,SE} + \mathbf{b}^{l,SE}, \tag{11}$$

where Offset_E represents the channel position offset operation, as shown in Eq. (12).

$$\mathit{Offset}_E\left(\mathbf{x}_{h,w}^{l,Hor,t-1}, \Delta\mathbf{p}^{l,SE}\right) = \left(1 - \Delta p_c^{l,SE}\right) x_{h,w,c}^{l,Hor} + \Delta p_c^{l,SE} x_{h,w,c+1}^{l,Hor}. \tag{12}$$

The incorporation of high-order differential iterations in HSEDC allows the model to emphasize curvature-level differences in spectral signatures, which are particularly important in HSI data where subtle shifts in spectral slope or shape often indicate different material classes. By using ASES, the HSEDC module avoids the limitations of fixed sampling, offering a flexible and learnable mechanism that adjusts sampling points to focus on more informative spectral variations, thereby improving the spectral sensitivity of the model.

3.4. HSADC

Similar to the HSEDC module, the HSADC module lies in the use of higher-order adaptive spatial differential convolution. First, DSAS integrates with deformable convolution to generate initial feature maps along with the corresponding convolution kernels and offsets, which are then used to compute the spatial differential convolution kernels and the corresponding spatial offsets needed in the HSADC, as shown in Fig. 4b.

Traditional 2D convolution utilizes uniformly spaced grid positions (e.g., a fixed 3 × 3 kernel with neighbors at fixed offsets). However, this rigid sampling is inadequate for modeling irregular object boundaries or shapes. The DSAS operation overcomes this limitation by learning spatial offsets that dynamically adjust the sampling positions based on content. As depicted in Fig. 4b, DSAS can shift the kernel sampling to follow the edge of an object or align with important structures in the spatial domain.

The spatial differential convolution kernel is computed by summing deformable convolution kernels across all channels

$$w_{i,j}^{l,SA} = \sum_{c=1}^{C} w_{i,j,c}^{l}, \tag{13}$$

where $w_{i,j}^{l,SA}$ represents the kernel value at spatial position (i,j), and $w_{i,j,c}^{l}$ is the deformable kernel value at the c-th channel.

The corresponding spatial offset is derived by summing deformable offsets across channels:

$$\Delta p_{h,w}^{l,SA} = \sum_{c=1}^{C} \Delta p_{h,w,c}^{l}, \tag{14}$$

where $\Delta p_{h,w}^{l,SA}$ represents the offsets at (h,w), and $\Delta p_{h,w,c}^{l}$ is the deformable offset at the c-th channel.

Subsequently, through a looping operation, deformable convolution is applied multiple times to progressively refine the feature representation and obtain higher-order spatial differential convolution results. The weight θ is used to adjust the results of the native deformable convolution and differential convolution, retaining detail information in each order. Finally, the differential convolution results from each order are merged through accumulation to obtain the higher-order features corresponding to the module.

$$\mathbf{X}_c^{l,HSADC} = \mathbf{X}_c^{l,Hor} - \theta \sum_{t=0}^{T} \mathrm{SADC}\left(\mathbf{X}_c^{l,Hor,t}\right), \tag{15}$$

where $\mathrm{SADC}(\cdot)$ is the spatially adaptive differential convolution, $\mathbf{X}_c^{l,HSADC}$ is the feature map of the output of HSADC $\mathcal{X}^{l,HSADC}$ at the c-th channel, and $\mathbf{X}_c^{l,Hor,t}$ is the feature map of the spatially adaptive differential convolution at the c-th channel in the t-th iteration. When $t=0$, $\mathbf{X}_c^{l,Hor,t}$ represents the initial input $\mathbf{x}_c^{l,Hor}$ to the HSADC module. The computation of $\mathrm{SADC}(\cdot)$ is given by Eq. (16):

$$\mathrm{SADC}\left(\mathbf{X}_c^{l,Hor,t}\right) = \mathit{Offset}_A\left(\mathbf{X}_c^{l,Hor,t-1}, \Delta\mathbf{P}^{l,SA}\right) \cdot \mathbf{W}^{l,SA} + \mathbf{B}^{l,SA}, \tag{16}$$

where Offset_A represents the spatial position offset operation. The spatial position offset operation is given by Eq. (17):

$$\begin{aligned}\mathit{Offset}_A\left(\mathbf{X}_c^{l,Hor,t-1}, \Delta\mathbf{P}^{l,SA}\right) = {} & (1 - \Delta p_{h,w}^{l,H,SA})(1 - \Delta p_{h,w}^{l,W,SA})\mathbf{x}_{h,w}^{l,Hor} \\ & + \Delta p_{h,w}^{l,W,SA}(1 - \Delta p_{h,w}^{l,H,SA})\mathbf{x}_{h+1,w}^{l,Hor} \\ & + (1 - \Delta p_{h,w}^{l,W,SA})\Delta p_{h,w}^{l,H,SA}\mathbf{x}_{h,w+1}^{l,Hor} \\ & + \Delta p_{h,w}^{l,W,SA}\Delta p_{h,w}^{l,H,SA}\mathbf{x}_{h+1,w+1}^{l,Hor},\end{aligned} \tag{17}$$

where, $\Delta p_{h,w}^{l,H,SA}$ and $\Delta p_{h,w}^{l,W,SA}$ represent the offset components along the H and W directions, respectively.

The recursive higher-order operation in HSADC enables the extraction of nuanced geometric features such as edge transitions, texture variations, and structural context, which are often overlooked by shallow convolution. This is conceptually related to high-order local descriptors, but is further enhanced here by deformable spatial shifts that dynamically adapt to object boundaries and complex layouts in hyperspectral scenes. With DSAS operation, the model performs flexible spatial sampling, adapting to complex geometries and avoiding the oversmooth effect of fixed kernels. This deformable mechanism results in improved boundary delineation and better spatial context modeling, especially for non-uniform terrain and man-made structures in hyperspectral scenes.

3.5. Classifier

To predict the category, we construct a classifier to classify the integrated features $\mathcal{X}^L$ output by the final HorMSDC block. The classifier consists of global average pooling (GAP(·)) and a multilayer perceptron (MLP(·)).

$$\hat{\boldsymbol{y}} = \mathrm{Softmax}\big(\mathrm{MLP}\big(\mathrm{GAP}\big(\mathcal{X}^L\big)\big)\big), \tag{18}$$

where $\hat{\boldsymbol{y}} \in \mathbb{R}^{1\times P}$ represents the predicted result, and P is the total number of classes. The Softmax function maps the output to class probabilities, from which the final prediction of the model is obtained.

4. Experiments

4.1. Experimental datasets

The experiments are conducted on five public HSI datasets. The false-color and ground truth maps information are shown in Fig. 5, with descriptive information provided in Table 1. The more detailed description of experimental datasets is as follows:

- The Indian Pines (IP) dataset (Larry L. Biehl & Landgrebe, 2015) was acquired in 1992 by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) over a region in northwestern Indiana, USA. This dataset comprises 145 × 145 pixels with 220 spectral bands (400–2500 nm). After removing water absorption bands, 200 spectral bands were retained. The spatial resolution is 20 m, and it includes 16 distinct land cover types.
- The Salinas Valley (SV) dataset (Gualtieri et al., 1999) was collected in 1998 by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over Salinas Valley, California, USA. It has a spatial resolution of 3.7 m and consists of 512 × 217 pixels with 224 spectral bands (400–2500 nm). After removing water absorption bands, 204 spectral bands were retained. It includes 16 distinct land cover classes.
- The WHU-Hi-LongKou (LK) dataset (Zhong et al., 2020), released by Wuhan University, was collected in 2017 using UAV over agricultural areas in Longkou Town, Hubei Province, China. It includes 9 land cover categories, with a spatial resolution of 0.463 m. This dataset consists of 550 × 400 pixels and 270 spectral bands (400–1000 nm).

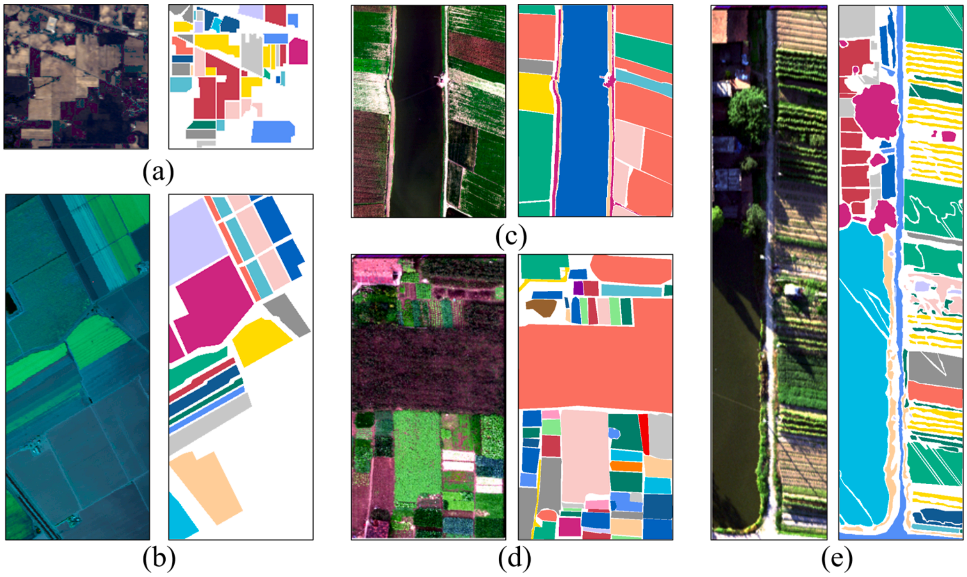


Fig. 5. The false-color images and ground-truth maps of the experimental datasets. (a) IP. (b) SA. (c) LK. (d) HH. (e) HC.

Table 1
The color, class, and sample number for each category of five HSI datasets.

	IP		SA		LK		HH		HC	
NO.	Class	Number	Class	Number	Class	Number	Class	Number	Class	Number
C01	Alfalfa	46	Brocoli green weeds 1	2009	Corn	34500	Red roof	14041	Strawberry	44735
C02	Corn notill	1428	Brocoli green weeds 2	3726	Cotton	8374	Road	3512	Cowpea	22753
C03	Corn mintil	830	Fallow	1976	Sesame	3031	Bare soil	21821	Soybean	10287
C04	Corn	237	Fallow rough plow	1394	Broad leaf soybean	63201	Cotton	163285	Sorghum	5353
C05	Grass pasture	483	Fallow smooth	2678	Narrow leaf soybean	4151	Cotton firewood	6218	Water spinach	1200
C06	Grass trees	730	Stubble	3959	Rice	11854	Rape	44557	Watermelon	4533
C07	Grass pasture mowed	28	Celery	3579	Water	67056	Chinese cabbage	24103	Greens	5903
C08	Hay windrowed	478	Grapes untrained	11271	Roads and houses	7124	Pakchoi	4054	Tress	17978
C09	Oats	20	Grapes untrained	6203	Mixed weed	5229	Cabbage	10819	Grass	9469
C10	Soybean notill	972	Corn senesced green weeds	3278			Tuber mustard	12394	Red roof	10516
C11	Soybean mintill	2455	Lettuce romaine 4wk	1068			Brassica parachinensis	11015	Gray roof	16911
C12	Soybean clean	593	Lettuce romaine 5wk	1927			Brassica chinensis	8954	Plastic	3679
C13	Wheat	205	Lettuce romaine 6wk	916			Small Brassica chinensis	22507	Bare soil	9116
C14	Woods	1265	Lettuce romaine 7wk	1070			Lactuca sativa	7356	Road	18560
C15	Buildings grass trees drives	386	Vinyard untrained	7268			Celtuce	1002	Bright object	1136
C16	Stone steel towers	93	Vinyard vertical trellis	1807			Film covered lettuce	7262	Water	75601
C17							Romaine lettuce	3010		
C18							Carrot	3217		
C19							White radish	8712		
C20							Garlic sprout	3486		
C21							Broad bean	1328		
C22							Tree	4040		
Total		10249		57129		204542		386693		257497

Table 2
Classification accuracy with different number of orders on the five HSI datasets.

Dataset	Metrics	order-1	order-2	order-3	order-4	order-5
IP	OA(%)	92.04	92.76	**92.81**	92.19	91.94
	AA(%)	93.76	94.76	**94.45**	94.29	93.71
	κ(×100)	91.06	91.74	**91.80**	91.10	90.82
SA	OA(%)	99.48	99.49	**99.50**	99.41	99.40
	AA(%)	99.46	99.47	**99.56**	99.40	99.22
	κ(×100)	99.43	99.43	**99.44**	99.26	99.15
LK	OA(%)	99.72	99.73	**99.76**	99.76	99.63
	AA(%)	99.28	99.30	**99.35**	99.33	98.96
	κ(×100)	99.63	99.64	**99.68**	99.68	99.51
HH	OA(%)	99.03	99.23	**99.31**	99.28	99.29
	AA(%)	98.37	98.50	98.41	98.23	**98.72**
	κ(×100)	99.08	99.08	**99.12**	99.04	99.03
HC	OA(%)	98.30	98.45	**98.69**	98.21	98.49
	AA(%)	98.22	98.33	**98.50**	98.01	98.30
	κ(×100)	98.90	99.03	**99.20**	99.11	99.09

- The WHU-Hi-HongHu (HH) dataset (Zhong et al., 2020) was acquired in 2017 using UAV over agricultural areas in Honghu City, Hubei Province, China. This dataset comprises 22 land cover categories, with a spatial resolution of 0.043 m. It consists of 940 × 475 pixels and 270 spectral bands (400–1000 nm).
- The WHU-Hi-Han-Chuan (HC) dataset (Zhong et al., 2020) was collected in Hanchuan City, Hubei Province. This dataset covers a wide range of feature types, including farmland, water bodies, buildings, forests, and roads. It provides accurate pixel-level labeling information, with 274 spectral bands (400–2500 nm) and 16 categories. The spatial resolution is 0.109 m. The dataset consists of 1217 × 303 pixels.

4.2. Experimental configurations

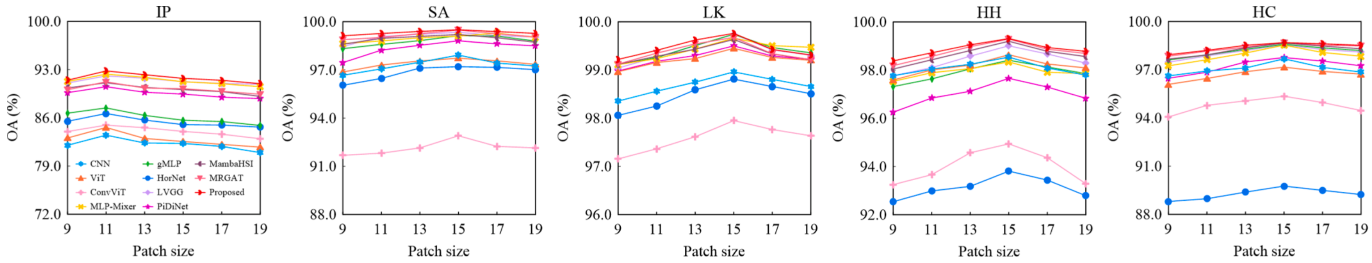
The experiments are conducted using Python 3.11 and PyTorch 2.2.2 on a workstation equipped with an Intel Core i9-13900KF CPU, 128G RAM, and an NVIDIA GeForce RTX 4090 24GB GPU. To evaluate the performance of our proposed HorD^{2}CN, we select ten advanced deep learning methods for comparison, i.e., CNN (Lee & Kwon, 2017), ViT (Dosovitskiy et al., 2021), ConvViT (Wu et al., 2021), MLP-Mixer (Tolstikhin et al., 2021), gMLP (Liu et al., 2021), HorNet (Rao et al., 2022), LVGG (Fei et al., 2024), PiDiNet (Su et al., 2021), MambaHSI (Li et al., 2024), and MRGAT (Ding et al., 2023).

We employ five evaluation metrics on the five HSI datasets in the experiments, including overall accuracy (OA), average accuracy (AA), Kappa coefficient (κ), model parameters (Params.), floating point operations (FLOPs), and training time per epoch (Time) (Li et al., 2024; Zhong et al., 2025).

4.3. Parameter validation

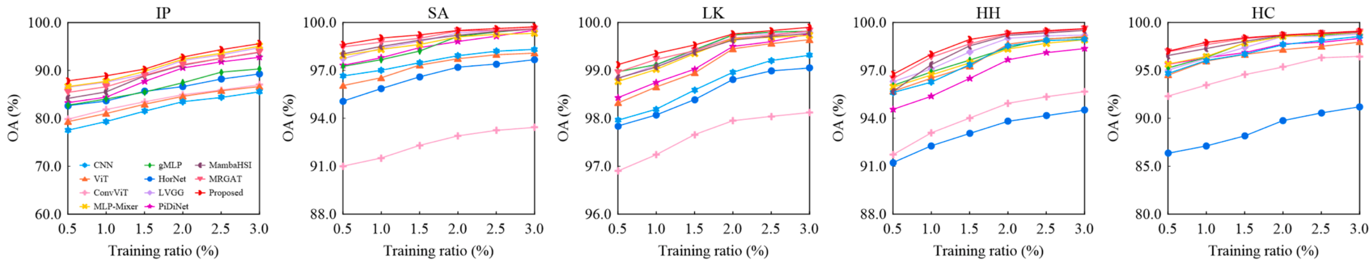
Order T in HorD^{2}CN: Table 2 presents the performance metrics on the five datasets for orders ranging from 1 to 5. The results indicate that all metrics improve as the order increases from 1 to 3, reaching their peak at order-3. Beyond order-3, performance begins to decline, suggesting that higher orders may lead to overfitting. For the IP dataset, the metrics show slight improvements from order-1 to order-3, but performance fluctuates overall. The SA and LK datasets exhibit stable performance across different orders, with a slight decline at higher orders (order-4 and order-5). In the HH dataset, the best OA and κ values occur at order-3, while AA is slightly higher at order-5. For the HC dataset, AA shows a marginal decrease from 98.41 % at order-3 to 98.30 % at order-5. Overall, order-3 is identified as the optimal choice based on the evaluation metrics.

Patch size: Fig. 6 illustrates the impact of patch size on OA across various datasets. For the IP dataset, OA increases as the patch size grows from 9 to 11, with HorD^{2}CN achieving the highest performance at a patch size of 11. In the SA dataset, OA increases rapidly with smaller patch sizes (9 to 15) before stabilizing at larger patch sizes (17 to 19), with optimal performance at a patch size of 15. The LK, HH, and HC datasets show similar trends, with OA increasing significantly as the patch size grows, but stabilizing at a patch size of 15. HorD^{2}CN achieves the best performance at this patch size, outperforming other models. Based on these observations, a patch size of 11 is selected for the IP dataset, while a patch size of 15 is chosen for the other four datasets.



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

Fig. 6. OAs of different patch sizes across five HSI datasets. (a) IP. (b) SA. (c) LK. (d) HH. (e) HC.



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

Fig. 7. OAs of different training ratios across five HSI datasets. (a) IP. (b) SA. (c) LK. (d) HH. (e) HC.

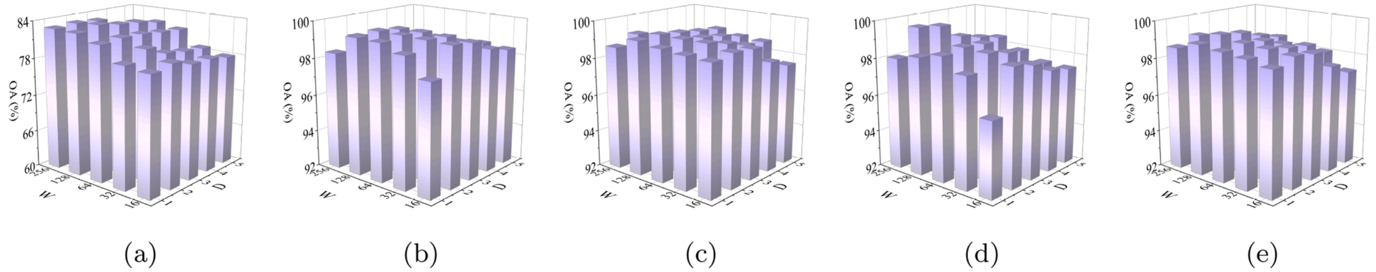


Fig. 8. OAs of HorD^{2}CN corresponding to different depth and width for HSI datasets. (a) IP. (b) SA. (c) LK. (d) HH. (e) HC.

Training ratio: This experiment evaluates six different training ratios (0.5 %, 1.0 %, 1.5 %, 2.0 %, 2.5 %, and 3.0 %). Fig. 7 illustrates the relationship between the training ratio and OA across the five datasets. Generally, as the training ratio increases, the OA of all models improves. HorD^{2}CN demonstrates significant advantages, particularly at low training ratios (0.5 % and 1.0 %), where its performance is notably superior to other models. CNN, ConvViT, and HorNet perform relatively poorly across all datasets, while MLP-Mixer and LVGG show similar performance, though still lagging behind HorD^{2}CN. As the training ratio increases, the performance gap between models narrows. When the training ratio reaches 2.0 %, the improvement slows, resulting in the selection of a 2.0 % training ratio for the experiment.

Depth and width of HorD^{2}CN: The depth and width of a deep learning model are critical hyperparameters that significantly influence its performance. Depth (D), defined by the number of layers, determines the capacity of the network to learn complex and hierarchical feature representations. However, increasing depth also raises computational costs and the risk of challenges such as vanishing gradients. Width (W), characterized by the number of neurons or channels per layer, governs the ability of the network to capture fine-grained details and intricate data interactions. Achieving an optimal balance between depth and width is essential for maximizing accuracy and efficiency, particularly under constrained computational resources.

To validate the relationship between model depth and width, we evaluate depths ranging from 1 to 5 and widths of 16, 32, 64, 128, and 256. As shown in Fig. 8, performance improves with increasing width and depth. The best results are achieved with a depth of 2 and a width of 128. Deeper and wider networks tend to overfit and increase computational demands, which is why this combination is selected for the experiment. Specifically, in the IP and HC datasets, OA varies significantly with changes in depth and width, indicating that these datasets are more sensitive to the model structure. In the SA dataset, the OA is relatively high and evenly distributed, suggesting that the model performs more stably on this dataset.

4.4. Comparison analysis

Tables 3–7 present the comparison results between HorD^{2}CN and other methods across the five datasets. Overall, HorD^{2}CN outperforms the other methods in terms of OA, AA, and $\kappa(\times 100)$ across all datasets. Compared to the second-best method, HorD^{2}CN achieves improvements in OA by 0.46 % (IP dataset), 0.03 % (SA dataset), 0.02 % (LK dataset), 0.02 % (HH dataset), and 0.01 % (HC dataset). CNN exhibits the poorest performance across all datasets, while ViT-based methods show relatively lower performance compared to MLP-based methods. PiDiNet achieves relatively satisfactory results due to its focus on extracting local edges and texture information, with high sensitivity to detailed features. However, PiDiNet has limitations in extracting spectral dimension features. LVGG, by retaining the multi-layer convolutional characteristics of the VGG network, enables the extraction of richer, multi-level features in both spatial and spectral dimensions, resulting in superior performance compared to PiDiNet.

In our experiments, MambaHSI and MRGAT demonstrate competitive performance in categories with complex spatial structures or subtle spectral differences (e.g., C02 and C03 in the IP dataset), yet

Table 3
Quantitative results of different deep learning methods on the IP dataset (2.0 % training ratio) with best results highlighted in bold.

Class	CNN	ViT	ConvViT	MLP-Mixer	gMLP	HorNet	LVGG	PiDiNet	MambaHSI	MRGAT	HorD^{2}CN
1	84.94±0.94	**100.0±0.00**	88.98±3.52	92.65±1.08	94.94±0.94	94.94±0.94	97.15±0.21	**100.0±0.00**	97.15±0.21	97.15±0.21	97.15±0.21
2	77.79±0.72	79.75±0.17	79.04±0.51	89.29±0.33	80.58±0.48	80.78±0.46	89.89±0.47	87.21±0.51	90.10±0.56	**92.00±0.42**	88.95±0.38
3	79.35±0.63	69.80±0.52	81.09±0.23	80.44±0.50	81.44±1.10	66.44±0.75	78.26±0.57	80.35±0.62	**87.04±0.58**	86.56±0.57	86.33±0.69
4	85.69 + 0.95	85.37±0.93	89.44±1.57	99.66±0.28	97.43±0.44	95.74±0.33	97.87±0.22	99.11±0.26	**100.0±0.00**	98.87±0.53	**100.0±0.00**
5	78.41±1.01	71.37±1.25	79.21±1.51	89.62±0.75	79.46±1.07	83.61±1.16	90.84±0.79	88.74±0.79	80.84±1.22	**91.16±0.93**	88.41±0.70
6	76.39±0.38	84.29±0.84	82.93±0.78	95.10±0.54	88.27±0.62	95.72±0.34	95.69±0.42	**95.87±0.52**	86.43±1.08	89.60±0.80	93.99±0.70
7	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**
8	89.83±0.14	**100.0±0.00**	90.71±1.02	99.56±0.14	98.90±0.25	99.83±0.13	99.67±0.27	97.91±0.28	**100.0±0.00**	**100.0±0.00**	99.28±0.29
9	**100.0±0.00**	91.68±4.80	86.23±4.95	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**
10	74.77±0.45	74.91±0.44	81.35±0.51	82.30±0.86	71.35±0.57	77.62±1.01	80.91±0.69	81.90±0.32	84.93±0.68	**89.12±0.28**	83.48±0.44
11	86.30±0.20	88.29±0.48	87.81±0.40	95.76±0.21	92.94±0.33	88.38±0.29	95.16±0.10	90.75±0.22	93.18±0.15	95.27±0.37	**96.70±0.21**
12	83.24±0.66	68.60±1.55	81.65±1.04	88.97±0.88	76.79±1.26	78.53±0.50	88.96±0.41	87.88±0.32	86.19±0.47	88.55±0.77	**90.61±0.52**
13	**100.0±0.00**	99.48±0.26	98.82±0.28	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	97.15±0.46	99.08±0.34
14	88.53±0.15	92.60±0.30	75.96±0.40	98.74±0.19	96.88±0.25	97.59±0.25	98.01±0.30	97.76±0.18	96.60±0.20	**99.24±0.11**	98.53±0.33
15	84.48±0.53	92.08±0.67	82.78±0.75	95.45±0.40	92.62±0.70	92.49±0.76	88.28±0.35	**93.16±1.02**	93.04±0.65	91.72±0.23	88.69±0.97
16	**100.0±0.00**	89.71±1.38	98.57±0.04	98.57±0.04	**100.0±0.00**	95.43±0.55	**100.0±0.00**	**100.0±0.00**	94.58±0.99	90.01±1.44	**100.0±0.00**
OA (%)	83.49±0.18	84.60±0.11	84.94±0.20	92.35±0.09	87.44±0.19	86.63±0.18	92.09±0.08	90.52±0.21	91.10±0.13	91.23±0.21	**92.81±0.12**
AA (%)	85.25±0.07	86.75±0.35	84.66±0.30	94.13±0.05	90.72±0.16	90.44±0.15	94.42±0.06	93.79±0.13	93.13±0.15	94.15±0.16	**94.45±0.09**
$\kappa\times100$	82.60±0.21	81.31±0.14	82.45±0.22	91.27±0.10	85.65±0.22	84.80±0.21	90.98±0.10	89.21±0.24	89.87±0.15	91.28±0.24	**91.80±0.14**
Param. (M)	0.34	0.61	0.75	7.44	7.52	0.57	0.84	4.06	**0.05**	1.51	0.58
Flops (G)	3.40	0.88	1.05	23.09	21.66	2.81	7.42	3.05	**0.33**	18.87	2.95
Time (s)	3.68	3.18	3.62	2.95	3.49	3.03	3.29	3.01	**2.53**	4.13	3.04

Table 4
Quantitative results of different deep learning methods on the SA dataset (2.0 % training ratio) with best results highlighted in bold.

Class	CNN	ViT	ConvViT	MLP-Mixer	gMLP	HorNet	LVGG	PiDiNet	MambaHSI	MRGAT	HorD^{2}CN
1	**100.0±0.00**	99.97±0.03	80.12±0.10	99.95±0.03	99.91±0.07	99.65±0.08	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	99.85±0.08	**100.0±0.00**
2	**100.0±0.00**	99.94±0.03	84.78±0.21	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	99.97±0.01	99.89±0.03
3	99.97±0.03	99.41±0.06	99.13±0.19	**100.0±0.00**	99.95±0.03	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**
4	99.10±0.14	97.19±0.22	96.95±0.28	99.92±0.04	**100.0±0.00**	99.83±0.04	99.38±0.05	97.46±0.18	96.34±0.26	99.81±0.06	99.92±0.04
5	98.57±0.05	92.96±0.20	92.15±0.45	96.23±0.14	96.47±0.13	98.19±0.12	**99.37±0.07**	98.22±0.05	97.87±0.10	97.92±0.13	98.19±0.08
6	**100.0±0.00**	99.92±0.03	99.78±0.03	**100.0±0.00**	**100.0±0.00**	99.95±0.02	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**
7	**100.0±0.00**	99.86±0.03	98.69±0.06	99.89±0.02	99.81±0.04	99.91±0.02	99.87±0.02	99.73±0.05	99.94±0.02	99.96±0.01	99.77±0.04
8	95.45±0.17	94.05±0.18	90.49±0.16	98.80±0.06	98.47±0.09	91.29±0.08	98.62±0.07	97.15±0.06	97.96±0.03	99.03±0.04	**99.30±0.06**
9	99.97±0.02	**100.0±0.00**	99.85±0.03	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	99.98±0.01	99.87±0.02	**100.0±0.00**
10	96.61±0.08	97.34±0.04	96.19±0.21	96.29±0.11	97.50±0.04	96.53±0.11	97.02±0.10	97.02±0.11	**99.25±0.08**	97.01±0.10	97.48±0.04
11	**100.0±0.00**	99.51±0.11	89.54±0.63	99.90±0.05	**100.0±0.00**	99.53±0.20	**100.0±0.00**	99.95±0.06	**100.0±0.00**	98.74±0.21	**100.0±0.00**
12	99.81±0.05	99.50±0.08	99.52±0.06	99.93±0.01	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	99.18±0.11	99.86±0.08	**100.0±0.00**	**100.0±0.00**
13	**100.0±0.00**	99.59±0.11	98.07±0.19	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	99.77±0.07	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**
14	99.55±0.13	99.60±0.09	96.61±0.44	99.82±0.06	99.72±0.05	99.72±0.09	**100.0±0.00**	96.76±0.36	**100.0±0.00**	98.91±0.17	98.84±0.21
15	93.98±0.13	97.64±0.12	86.93±0.18	98.39±0.10	98.52±0.04	95.31±0.12	98.75±0.06	99.08±0.06	99.17±0.04	99.45±0.09	**99.55±0.02**
16	99.96±0.04	98.69±0.17	88.04±0.26	**100.0±0.00**	99.88±0.06	99.73±0.08	**100.0±0.00**	99.17±0.10	99.84±0.03	99.59±0.03	**100.0±0.00**
OA	97.92±0.03	97.73±0.04	92.90±0.02	99.10±0.02	99.13±0.02	97.20±0.03	99.34±0.02	98.81±0.01	99.20±0.01	99.47±0.01	**99.50±0.02**
AA	98.94±0.01	98.45±0.04	93.55±0.05	99.32±0.01	99.39±0.01	98.73±0.03	**99.56±0.01**	98.97±0.02	99.39±0.02	99.38±0.02	**99.56±0.02**
$\kappa(\times 100)$	97.69±0.03	97.48±0.05	92.09±0.02	99.00±0.03	99.03±0.02	96.89±0.03	99.27±0.02	98.67±0.01	99.11±0.01	99.41±0.01	**99.45±0.02**
Param. (M)	0.35	0.62	0.76	7.45	7.52	0.57	0.85	4.09	**0.05**	1.51	0.58
Flops (G)	3.07	0.77	0.92	20.05	18.78	2.45	6.53	2.75	**0.30**	16.38	2.57
Time (s)	10.23	10.92	10.87	7.44	11.51	9.68	10.07	**7.20**	7.65	13.85	9.95

Table 5
Quantitative results of different deep learning methods on the LK dataset (2.0 % training ratio) with best results highlighted in bold.

Class	CNN	ViT	ConvViT	MLP-Mixer	gMLP	HorNet	LVGG	PiDiNet	MambaHSI	MRGAT	HorD^{2}CN
1	99.98±0.01	99.90±0.01	99.70±0.02	99.98±0.01	99.97±0.01	99.96±0.01	**100.0±0.00**	99.88±0.01	99.82±0.01	99.95±0.01	99.93±0.01
2	97.74±0.09	99.47±0.05	90.03±0.12	99.99±0.01	99.97±0.00	98.59±0.04	99.94±0.01	99.92±0.01	99.93±0.01	**100.0±0.00**	99.98±0.01
3	95.47±0.17	98.93±0.10	96.83±0.19	99.87±0.06	99.36±0.05	**99.91±0.03**	99.71±0.04	99.89±0.02	99.81±0.04	99.86±0.03	99.81±0.03
4	98.41±0.03	99.67±0.01	98.81±0.03	99.80±0.01	99.79±0.01	99.29±0.03	99.68±0.01	99.58±0.01	**99.84±0.01**	**99.84±0.01**	99.82±0.01
5	90.79±0.25	97.95±0.23	59.62±0.29	99.66±0.03	99.67±0.04	96.95±0.14	99.27±0.09	97.20±0.18	98.55±0.15	99.54±0.05	**99.77±0.02**
6	99.92±0.02	99.32±0.03	98.45±0.04	99.79±0.02	99.72±0.02	99.09±0.05	99.73±0.03	**99.86±0.01**	99.53±0.05	99.83±0.02	99.64±0.04
7	99.98±0.00	99.96±0.00	99.94±0.00	99.91±0.00	99.98±0.00	99.82±0.00	**99.99±0.00**	99.97±0.00	99.98±0.01	99.97±0.01	99.98±0.01
8	96.36±0.10	93.64±0.13	95.66±0.08	97.18±0.09	97.36±0.09	83.98±0.05	98.10±0.05	94.79±0.07	95.90±0.08	95.05±0.02	**98.34±0.02**
9	**97.70±0.11**	96.58±0.08	96.16±0.12	96.34±0.07	97.63±0.02	93.04±0.11	97.30±0.06	96.73±0.11	97.04±0.14	97.60±0.12	96.86±0.06
OA	98.96±0.01	99.45±0.01	97.95±0.02	99.66±0.02	99.74±0.01	98.81±0.01	99.74±0.00	99.50±0.01	99.63±0.01	99.68±0.01	**99.76±0.01**
AA	97.37±0.05	98.38±0.03	92.80±0.04	99.17±0.02	99.27±0.02	96.74±0.02	99.30±0.01	98.65±0.02	98.93±0.03	99.07±0.02	**99.35±0.01**
$\kappa(\times 100)$	98.64±0.01	99.28±0.01	97.30±0.02	99.64±0.01	99.66±0.01	98.43±0.02	99.66±0.01	99.35±0.01	99.52±0.01	99.57±0.01	**99.68±0.01**
Param. (M)	0.45	0.65	0.83	18.17	18.12	0.42	0.92	4.27	**0.05**	1.52	0.44
Flops (G)	4.32	0.87	1.08	21.90	20.31	2.72	7.87	3.78	**0.41**	17.83	2.86
Time (s)	26.84	14.73	15.45	12.02	16.85	13.21	14.74	12.25	**11.00**	20.16	14.51

their OA remains lower than that of HorD^{2}CN. MambaHSI leverages a structured state-space model to capture long-range dependencies and integrates spectral-spatial features via adaptive fusion modules. However, its complex architecture with multiple Mamba blocks may lead to overfitting on small datasets. MRGAT efficiently extracts local-global features and edge semantics by combining superpixel segmentation with multi-scale graph attention, but its classification accuracy is sensitive to superpixel segmentation parameters. In contrast, HorD^{2}CN achieves an optimal balance between accuracy and efficiency, delivering superior overall classification performance. Notably, while maintaining state-of-the-art accuracy, HorD^{2}CN requires significantly fewer parameters than models such as MLP-Mixer and gMLP. Despite incurring slightly higher computational overhead than MambaHSI, HorD^{2}CN provides markedly improved accuracy, highlighting its strong cost-effectiveness. These results underscore the advantages of the proposed HorD^{2}CN for efficient and accurate HSI classification.

Figs. 9–13 illustrate the classification maps corresponding to the results in Tables 3–7, respectively. Overall, HorD^{2}CN exhibits superior classification accuracy compared to baseline methods, primarily attributed to its integration of high-order differential convolution for both spectral and spatial feature extraction. Additionally, an adaptive shift operation is employed in the spectral domain to dynamically emphasize subtle variations while suppressing noise, thereby enhancing the robustness and discriminative power of the feature representation. In the spatial domain, deformable convolution improves spatial context awareness by adaptively focusing on detailed information through its flexible sampling properties, resulting in better alignment with the ground truth distribution and improved spatial consistency.

As illustrated in Fig. 9, HorD^{2}CN demonstrates significant advantages on the IP dataset, particularly for classes C11 (Soybean mintill), C12 (Soybean clean), and C16 (Stone steel towers). In contrast, comparative methods exhibit large-scale misclassifications in these categories (highlighted in red), all of which represent crop types with highly similar spectral signatures. The subtle differences in reflectance across certain spectral bands and complex spatial distribution make these categories challenging to distinguish. The ability of HorD^{2}CN to accurately differentiate between these classes underscores its effectiveness in capturing discriminative spectral features and spatial structures. In the SA dataset (Fig. 10), the visualized results reveal that HorD^{2}CN generates the least salt-and-pepper noise and achieves the highest classification accuracy. Notably, class C08 (Grapes untrained) is misclassified across all methods, but HorD^{2}CN shows the fewest misclassified pixels in this region, further highlighting its robustness.

Fig. 11 demonstrates the superior overall performance of the HorD^{2}CN model, which can be attributed to its dynamic high-order dif-

Table 6
Quantitative results of different deep learning methods on the HH dataset (2.0 % training ratio) with best results highlighted in bold.

Class	CNN	ViT	ConvViT	MLP-Mixer	gMLP	HorNet	LVGG	PiDiNet	MambaHSI	MRGAT	HorD^{2}CN
1	99.04±0.05	99.06±0.05	97.71±0.04	99.22±0.01	99.40±0.03	96.10±0.08	98.98±0.02	99.34±0.02	98.87±0.03	99.32±0.01	**99.41±0.03**
2	97.79±0.16	96.39±0.13	81.98±0.18	97.53±0.12	97.58±0.12	85.92±0.32	**98.83±0.09**	96.65±0.18	96.41±0.13	95.35±0.04	97.58±0.14
3	97.77±0.05	97.48±0.06	94.50±0.06	98.79±0.02	98.71±0.03	93.50±0.07	96.35±0.06	96.86±0.02	98.41±0.05	97.99±0.03	**99.11±0.02**
4	99.85±0.01	99.77±0.01	97.75±0.01	99.83±0.00	**99.90±0.00**	99.16±0.01	99.86±0.01	99.75±0.00	99.87±0.00	**99.90±0.00**	99.85±0.00
5	**99.76±0.03**	99.50±0.05	90.84±0.15	99.61±0.03	99.28±0.03	96.79±0.10	99.29±0.02	99.73±0.03	99.40±0.06	99.39±0.02	99.47±0.04
6	99.90±0.00	99.66±0.01	97.70±0.03	99.90±0.00	99.91±0.00	98.34±0.04	**99.94±0.00**	99.91±0.01	99.80±0.02	99.90±0.00	99.92±0.00
7	98.04±0.04	97.59±0.02	93.33±0.10	98.70±0.03	98.74±0.03	93.91±0.12	97.70±0.05	98.14±0.05	**98.81±0.04**	98.40±0.04	98.79±0.03
8	97.72±0.09	90.57±0.22	51.29±0.40	97.34±0.10	97.69±0.05	79.43±0.22	**97.99±0.08**	94.71±0.14	97.38±0.09	97.76±0.03	97.95±0.15
9	99.53±0.02	99.03±0.06	98.48±0.08	99.25±0.02	99.35±0.02	98.63±0.06	**99.60±0.01**	99.23±0.04	99.55±0.03	99.48±0.03	99.56±0.02
10	98.81±0.04	97.32±0.04	81.70±0.24	98.43±0.03	98.24±0.05	94.70±0.09	98.51±0.05	98.63±0.05	**99.09±0.05**	98.76±0.05	98.74±0.05
11	98.96±0.06	96.74±0.05	73.94±0.12	99.28±0.04	**99.29±0.04**	95.01±0.11	99.22±0.03	98.71±0.05	99.03±0.04	98.45±0.03	97.70±0.09
12	97.30±0.08	95.23±0.07	75.12±0.44	97.12±0.06	97.57±0.02	82.13±0.10	95.21±0.12	96.97±0.10	96.30±0.02	97.11±0.08	**97.66±0.07**
13	98.06±0.06	97.17±0.05	85.79±0.13	98.20±0.04	98.09±0.08	91.03±0.09	97.55±0.06	96.53±0.12	98.05±0.07	**98.34±0.05**	98.11±0.07
14	99.62±0.03	99.37±0.05	92.82±0.07	99.67±0.05	99.62±0.04	94.08±0.12	**99.88±0.02**	99.71±0.03	99.30±0.04	99.61±0.05	99.82±0.03
15	94.55±0.42	88.04±0.29	64.37±0.69	93.23±0.29	93.85±0.39	88.68±0.29	93.64±0.24	89.91±0.44	95.14±0.29	**96.21±0.18**	94.12±0.25
16	99.52±0.03	99.04±0.05	85.95±0.20	98.15±0.04	98.69±0.05	92.34±0.09	**99.66±0.03**	98.30±0.06	99.08±0.04	99.22±0.03	99.08±0.06
17	97.20±0.18	96.54±0.15	75.18±0.29	97.68±0.05	97.66±0.07	87.54±0.22	99.01±0.03	96.95±0.05	97.92±0.06	**98.24±0.09**	97.56±0.08
18	94.42±0.16	94.03±0.13	77.08±0.32	94.37±0.10	96.15±0.15	96.20±0.18	95.29±0.10	93.27±0.18	96.18±0.17	96.22±0.19	**96.29±0.16**
19	98.59±0.04	97.32±0.09	92.11±0.15	99.08±0.07	99.12±0.06	94.24±0.11	98.76±0.06	96.38±0.06	97.59±0.06	98.75±0.05	**99.23±0.05**
20	96.94±0.16	93.73±0.32	84.23±0.08	96.96±0.22	97.26±0.15	94.79±0.08	95.74±0.22	94.92±0.24	96.97±0.15	97.52±0.15	**97.85±0.11**
21	98.16±0.06	94.27±0.38	46.29±0.51	97.84±0.34	97.44±0.12	97.82±0.17	95.49±0.20	95.29±0.27	95.99±0.25	**98.92±0.22**	98.40±0.10
22	99.62±0.02	99.47±0.06	92.45±0.20	99.77±0.03	**100.0±0.00**	94.08±0.21	99.54±0.04	99.30±0.06	99.51±0.03	99.81±0.03	98.82±0.11
OA	98.54±0.01	98.64±0.01	94.94±0.03	98.34±0.01	98.41±0.01	93.81±0.01	99.01±0.01	97.66±0.01	99.20±0.01	99.29±0.01	**99.31±0.01**
AA	98.23±0.04	96.70±0.05	83.21±0.07	98.18±0.03	98.34±0.03	92.93±0.03	98.00±0.02	97.24±0.04	98.12±0.02	98.39±0.03	**98.41±0.02**
κ(×100)	99.04±0.01	98.28±0.02	91.08±0.03	98.37±0.01	99.28±0.01	95.24±0.02	98.75±0.01	98.52±0.02	98.99±0.01	99.11±0.01	**99.12±0.01**
Param. (M)	0.45	0.65	0.83	18.18	18.12	0.71	0.92	4.29	**0.06**	1.53	0.72
Flops (G)	3.54	0.72	0.89	21.59	20.02	2.25	6.46	3.10	**0.34**	14.64	2.36
Time (s)	51.59	31.14	30.40	23.47	31.94	26.13	28.68	24.28	**21.39**	40.01	28.58

Table 7
Quantitative results of different deep learning methods on the HC dataset (2.0 % training ratio) with best results highlighted in bold.

Class	CNN	ViT	ConvViT	MLP-Mixer	gMLP	HorNet	LVGG	PiDiNet	MambaHSI	MRGAT	HorD^{2}CN
1	99.91±0.00	99.43±0.02	95.80±0.02	99.76±0.01	99.83±0.00	98.49±0.03	99.24±0.02	99.40±0.02	99.72±0.01	99.83±0.01	**99.86±0.01**
2	99.05±0.04	99.09±0.03	93.26±0.05	**99.72±0.01**	99.66±0.02	93.00±0.11	99.15±0.04	99.38±0.02	99.40±0.01	99.27±0.03	99.41±0.02
3	99.71±0.04	99.31±0.06	90.67±0.13	99.97±0.01	**99.98±0.01**	94.70±0.10	99.06±0.06	99.88±0.01	99.55±0.04	99.85±0.01	99.91±0.02
4	96.15±0.11	99.58±0.06	92.96±0.19	99.64±0.06	99.34±0.07	94.93±0.11	**99.75±0.05**	98.82±0.11	99.65±0.05	99.49±0.06	99.51±0.08
5	**100.0±0.00**	98.94±0.17	43.45±0.95	99.37±0.11	98.80±0.11	82.09±0.60	99.35±0.06	99.34±0.12	99.48±0.11	**99.80±0.08**	99.46±0.14
6	83.07±0.35	87.65±0.27	14.96±0.16	94.62±0.17	95.11±0.04	81.06±0.26	88.43±0.20	89.38±0.13	95.71±0.20	**95.75±0.08**	94.70±0.15
7	96.29±0.11	98.74±0.04	88.10±0.13	98.31±0.08	**99.54±0.05**	94.05±0.15	98.31±0.08	98.89±0.08	98.18±0.05	98.49±0.05	98.61±0.10
8	94.54±0.11	94.35±0.08	81.16±0.21	99.03±0.02	**99.18±0.03**	83.33±0.19	97.73±0.07	97.67±0.05	98.54±0.03	98.63±0.03	98.77±0.03
9	95.92±0.08	95.35±0.11	68.01±0.23	98.03±0.08	99.00±0.02	89.15±0.10	98.24±0.04	96.31±0.05	98.12±0.03	**99.17±0.03**	98.62±0.08
10	98.57±0.05	98.37±0.04	95.62±0.15	98.14±0.05	98.69±0.03	98.13±0.08	98.56±0.09	97.78±0.05	98.75±0.06	98.69±0.05	**98.89±0.10**
11	95.95±0.08	99.22±0.02	94.76±0.07	99.60±0.01	**99.88±0.02**	98.31±0.04	99.53±0.05	99.42±0.06	99.69±0.02	99.85±0.02	99.80±0.01
12	99.24±0.06	99.78±0.00	73.47±0.25	99.85±0.03	99.94±0.01	83.44±0.22	99.39±0.04	99.22±0.07	99.23±0.07	**100.0±0.00**	**100.0±0.00**
13	87.92±0.11	91.94±0.17	70.46±0.27	95.86±0.09	95.78±0.07	75.24±0.27	94.80±0.10	91.74±0.13	94.16±0.11	95.83±0.08	**95.95±0.10**
14	95.84±0.08	97.64±0.04	89.83±0.05	99.43±0.02	97.18±0.02	93.80±0.10	99.36±0.01	98.46±0.04	99.26±0.03	**99.48±0.02**	98.94±0.02
15	89.59±0.41	92.71±0.29	44.50±0.50	92.08±0.28	88.10±0.32	79.13±0.47	92.69±0.22	**94.25±0.42**	93.34±0.30	90.29±0.30	93.69±0.26
16	99.65±0.01	99.89±0.01	96.92±0.02	98.95±0.00	**99.93±0.00**	99.43±0.01	99.84±0.01	99.88±0.00	99.80±0.00	99.86±0.00	99.91±0.01
OA	97.68±0.01	97.16±0.01	95.35±0.04	98.53±0.01	98.58±0.00	89.76±0.02	98.54±0.01	97.76±0.01	98.66±0.01	98.68±0.01	**98.69±0.01**
AA	95.71±0.03	97.00±0.03	77.12±0.10	98.27±0.01	98.12±0.02	89.89±0.07	97.71±0.02	97.49±0.03	98.29±0.03	98.39±0.02	**98.50±0.04**
κ(×100)	97.29±0.01	98.04±0.01	88.08±0.05	99.28±0.01	98.33±0.00	93.67±0.02	98.69±0.01	98.47±0.01	99.09±0.01	99.12±0.01	**99.20±0.01**
Param. (M)	0.46	0.66	0.83	7.58	7.55	0.58	0.93	4.29	**0.06**	1.53	0.59
Flops (G)	1.96	0.39	0.49	9.78	9.07	1.22	3.54	1.71	**0.19**	7.96	1.28
Time (s)	36.09	19.85	19.38	16.65	22.37	18.29	20.05	16.76	**15.04**	25.10	18.85

ferential convolutional architecture. Specifically, in classes C05 (Narrow leaf soybean) and C08 (Roads and houses), the model significantly reduces salt-and-pepper noise, enhancing the stability and reliability of the classification results. Moreover, HorD^{2}CN outperforms other methods in preserving feature edges, which contributes to improved class separability and overall classification accuracy.

The HH dataset, which contains the largest number of categories among the five datasets, presents challenges due to the spatial proximity of certain crop types. Classes C06 (Rape), C12 (Brassica chinensis), and C19 (White radish) are frequently misclassified, as shown in Fig. 12. This misclassification arises from the overlapping geographical distributions of these crops, leading to spatial confusion between adjacent pixels. HorD^{2}CN addresses this issue by leveraging high-order differential convolution with deformable operations in the spatial domain to extract features from land cover boundaries, effectively mitigating interference caused by neighboring pixels. As a result, HorD^{2}CN achieves the highest classification accuracy, as evidenced by the clear delineation of boundaries in the results.

The HC dataset encompasses various land cover types, including farmland, buildings, water bodies, and roads. As depicted in Fig. 7, HorD^{2}CN achieves the highest classification accuracy while demonstrating superior boundary preservation. In contrast, CNN exhibits comparatively weaker performance on this dataset. Although ViT and MLP-based methods offer significant improvements over CNN, they still suffer from misclassifications in certain regions. These findings highlight the effectiveness of HorD^{2}CN in addressing the complex spatial and spectral char-

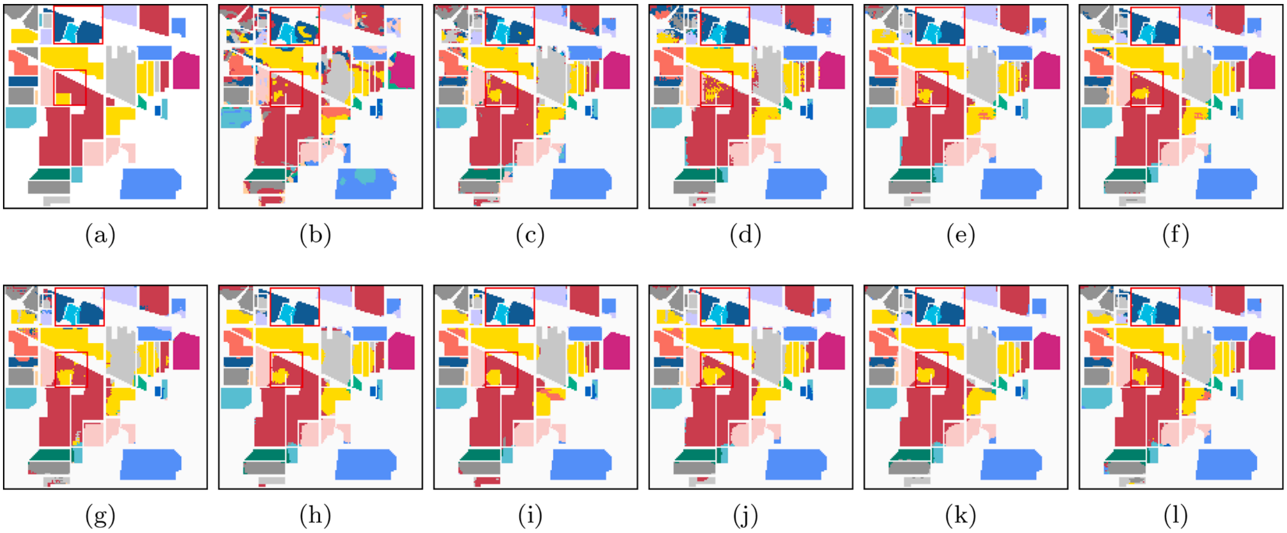


Fig. 9. Classification maps on the IP dataset at 2.0 % training ratio. (a) Ground truth. (b) CNN (83.65 %). (c) ViT (84.71 %). (d) ConvViT (85.14 %) (e) MLP-Mixer (93.23 %). (f) gMLP (87.56 %). (g) HorNet (86.75 %). (h) LVGG (92.17 %). (i) PiDiNet (90.73 %). (j) MambaHSI (91.20 %). (k) MRGAT (91.44 %). (l) **HorD^{2}CN (92.93 %)**.

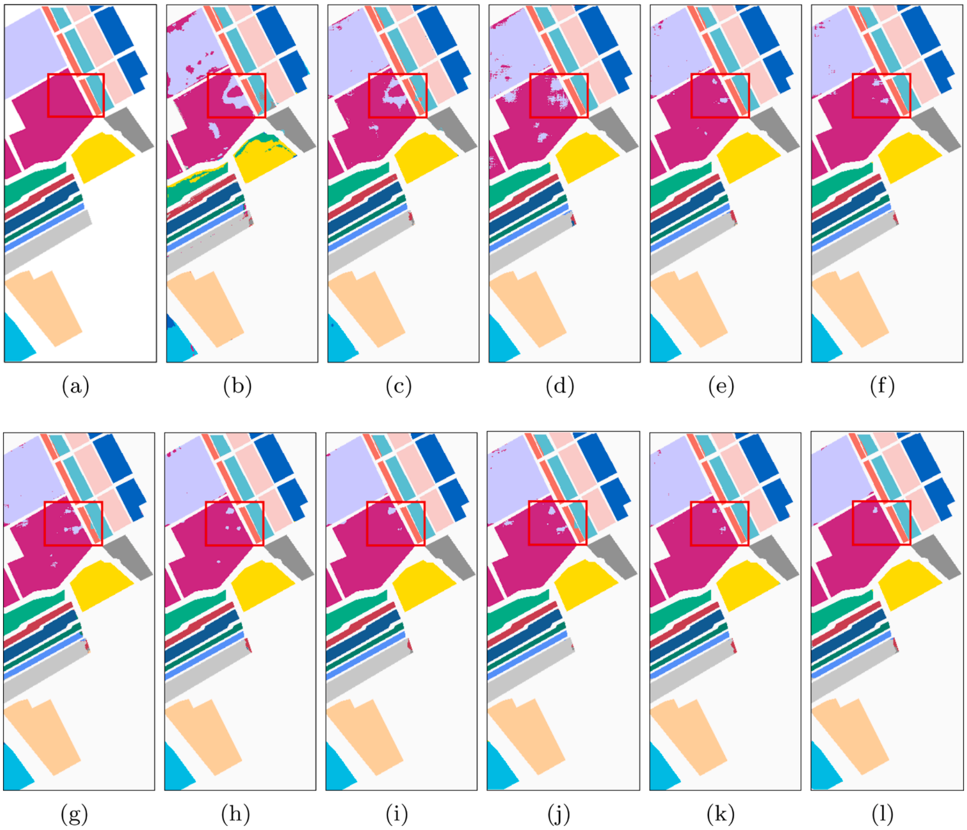


Fig. 10. Classification maps on the SA dataset at 2.0 % training ratio. (a) Ground truth. (b) CNN (97.95 %). (c) ViT (97.75 %). (d) ConvViT (92.92 %) (e) MLP-Mixer (99.12 %). (f) gMLP (99.15 %). (g) HorNet (97.23 %). (h) LVGG (99.36 %). (i) PiDiNet (98.82 %). (j) MambaHSI (99.21 %). (k) MRGAT (99.48 %). (l) **HorD^{2}CN (99.52 %)**.

acteristics of the dataset, ultimately enhancing classification accuracy and boundary delineation.

Fig. 14 provides a critical assessment of the feature separability learned by HorD^{2}CN and competing deep learning methods on the IP dataset, using t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten & Hinton, 2008) visualizations. Conventional CNN and ViT exhibit significant inter-class overlap, indicating their inability to disentangle complex spectral-spatial patterns. While frequency-aware models (LVGG, PiDiNet) and state-space model-based MambaHSI improve cluster separation, residual overlaps persist in spectrally similar agricultural categories (e.g., C02-C03). HorD^{2}CN achieves maximally compact intra-class distributions and distinct inter-class boundaries, as shown in Fig. 14(l), directly validating the efficacy of its architecture. The HSEDC module dynamically amplifies discriminative band-to-band differences, while the HSADC module adapts sampling positions to land cover boundaries. This synergistic operation, enabled by high-order dif-

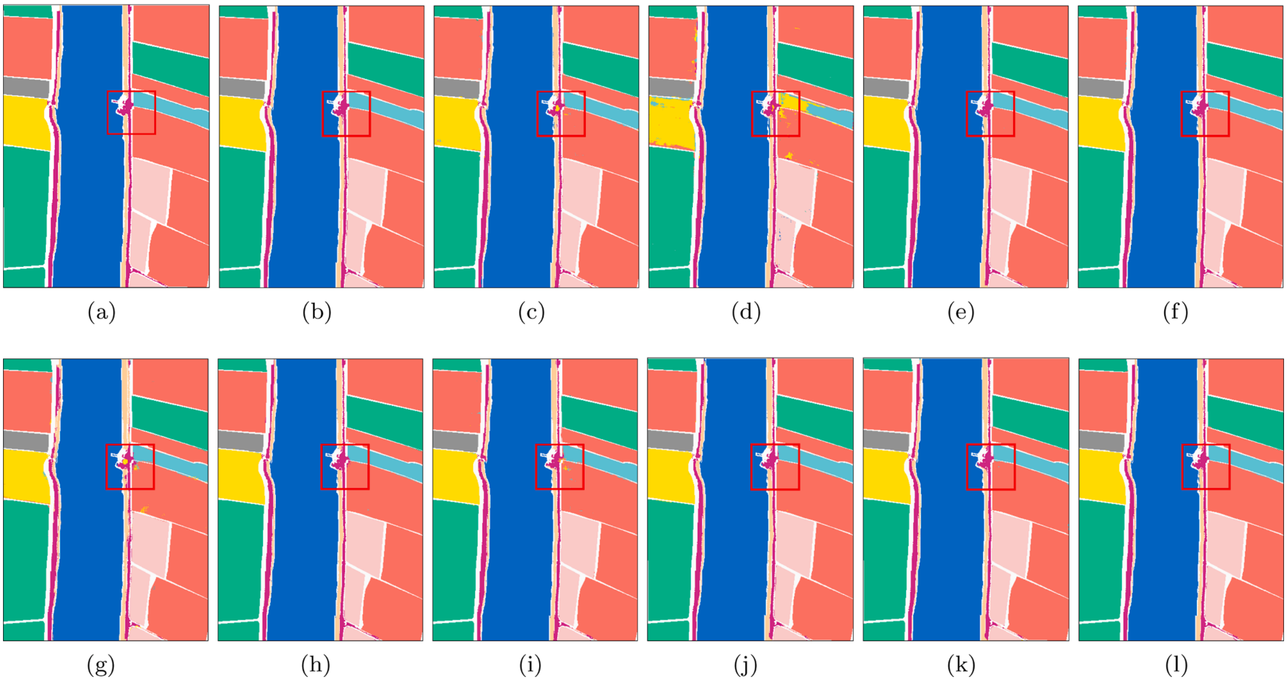


Fig. 11. Classification maps on the LK dataset at 2.0 % training ratio. (a) Ground truth. (b) CNN (98.67 %). (c) ViT (99.46 %). (d) ConvViT (97.97 %) (e) MLP-Mixer (99.75 %). (f) gMLP (99.75 %). (g) HorNet (98.82 %). (h) LVGG (99.74 %). (i) PiDiNet (99.51 %). (j) MambaHSI (99.64 %). (k) MRGAT (99.69 %). (l) **HorD^{2}CN (99.77 %)**.

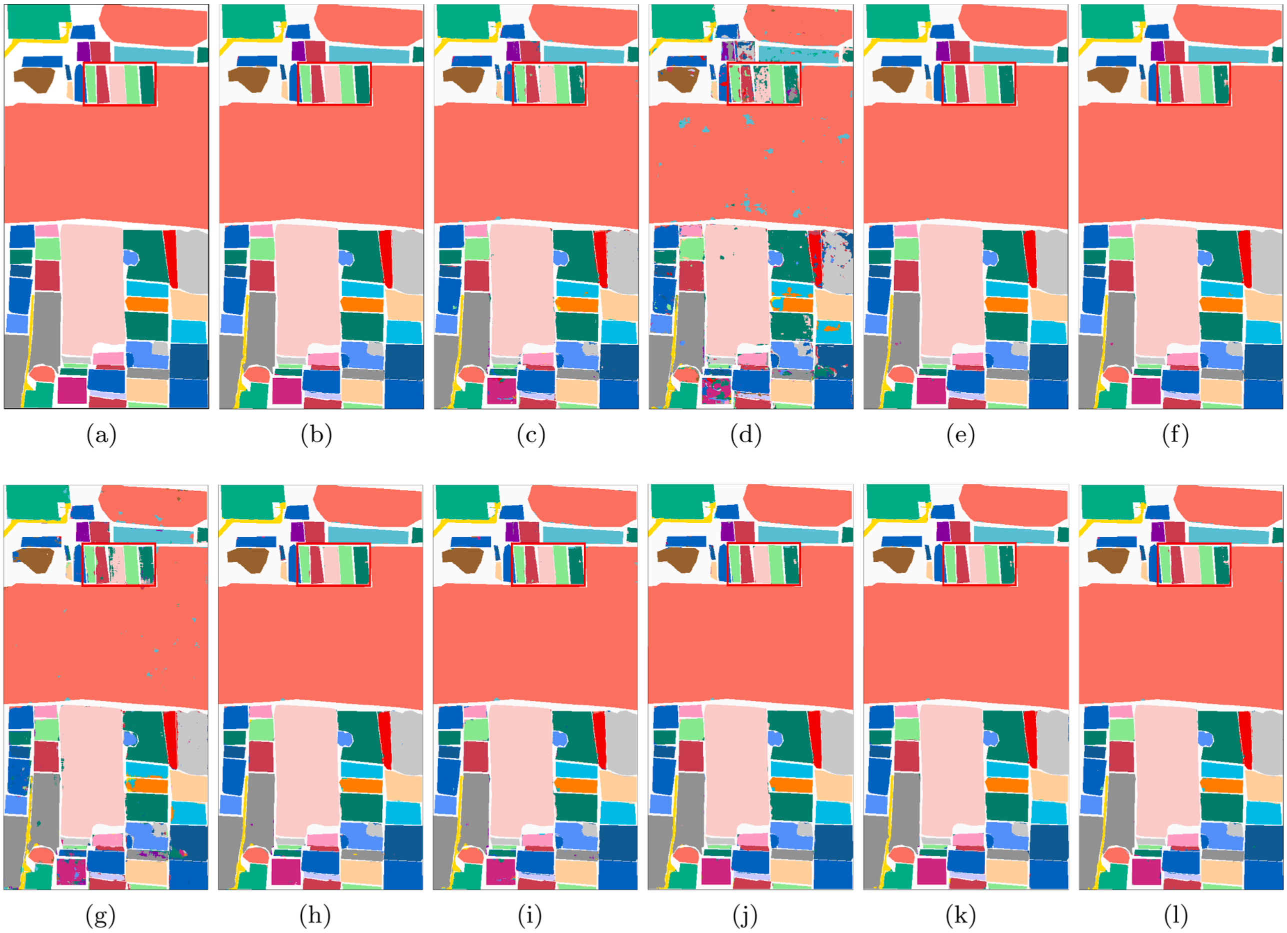


Fig. 12. Classification maps on the HH dataset at 2.0 % training ratio. (a) Ground truth. (b) CNN (98.55 %). (c) ViT (98.65 %). (d) ConvViT (94.97 %) (e) MLP-Mixer (98.35 %). (f) gMLP (98.42 %). (g) HorNet (93.82 %). (h) LVGG (98.55 %). (i) PiDiNet (97.67 %). (j) MambaHSI (99.21 %). (k) MRGAT (99.29 %). (l) **HorD^{2}CN (99.32 %)**.

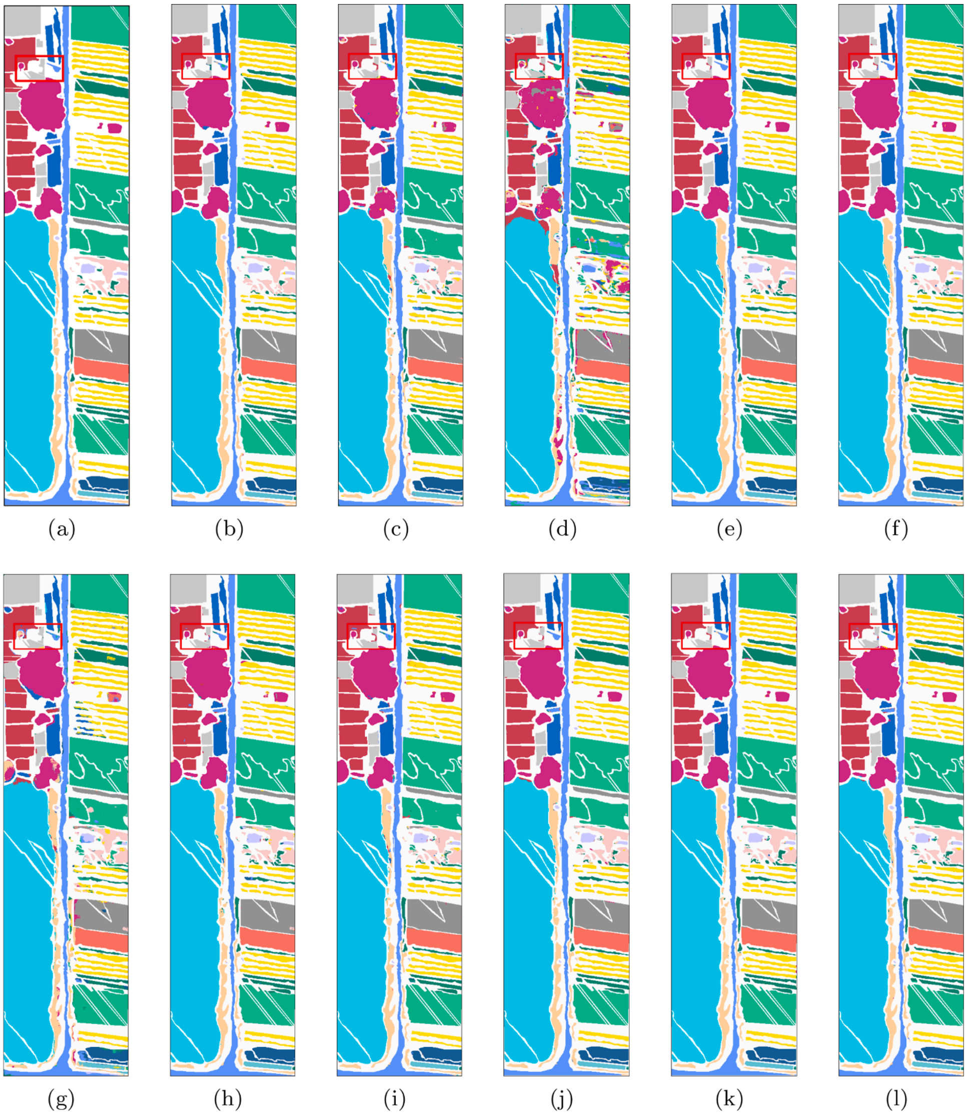


Fig. 13. Classification maps on the HC dataset at 2.0 % training ratio. (a) Groud truth (b) CNN (97.69 %). (c) ViT (97.17 %). (d) ConvViT (95.39 %) (e) MLP-Mixer (98.54 %). (f) gMLP (98.50 %). (g) HorNet (89.78 %). (h) LVGG (98.55 %). (i) PiDiNet (97.77 %). (j) MambaHSI (98.67 %). (k) MRGAT (98.69 %). (l) **HorD^{2}CN (98.70 %)**.

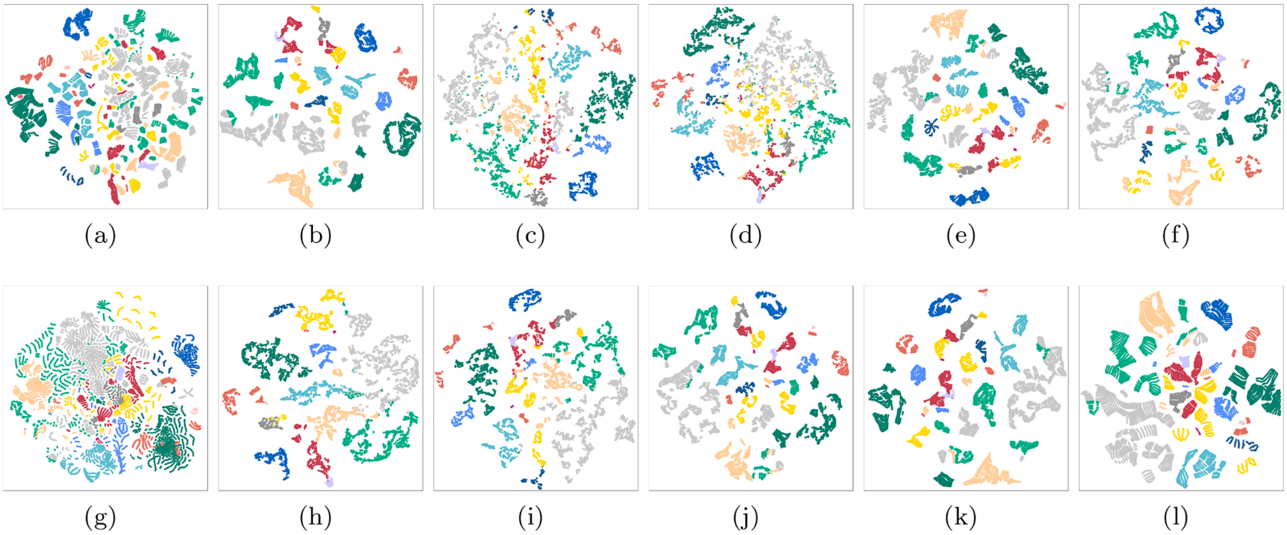


Fig. 14. The t-SNE plots of the IP dataset at 2.0 % training ratio. (a) Raw data. (b) CNN. (c) ViT. (d) ConvViT. (e) MLP-Mixer. (f) gMLP. (g) HorNet. (h) LVGG. (i) PiDiNet. (j) MambaHSI. (k) MRGAT. (l) HorD^{2}CN.

Table 8
Reference tables for different modules.

Component	Ours	HorD^{2}CN-A	HorD^{2}CN-B	HorD^{2}CN-C	HorD^{2}CN-D	HorD^{2}CN-E
HSEDC	✔	✘	✔	✔	✔	✘
ASES	✔	✘	✔	✘	✔	✘
HSADC	✔	✔	✘	✔	✔	✘
DSAS	✔	✔	✘	✔	✘	✘
DC	✘	✘	✘	✘	✘	✔

Table 9
Ablation studies on different experimental configuration of HSI dataset.

	Metric	Ours	HorD^{2}CN-A	HorD^{2}CN-B	HorD^{2}CN-C	HorD^{2}CN-D	HorD^{2}CN-E
IP	OA(%)	**92.81**	91.85	91.50	92.68	92.67	92.66
	AA(%)	**94.45**	92.32	93.40	94.42	94.06	94.08
	$\kappa(\times100)$	**91.80**	91.35	90.32	91.33	91.64	91.77
SA	OA(%)	**99.50**	99.44	99.38	99.18	99.46	99.48
	AA(%)	**99.56**	99.50	99.33	99.29	99.44	99.46
	$\kappa(\times100)$	**99.45**	99.36	99.33	99.32	99.36	99.39
LK	OA(%)	**98.69**	98.54	98.59	98.29	98.50	98.52
	AA(%)	**98.50**	98.41	98.23	98.35	98.02	98.40
	$\kappa(\times100)$	98.20	97.90	98.06	98.09	99.07	**99.17**
HH	OA(%)	**99.76**	99.25	99.28	99.31	99.42	99.60
	AA(%)	**99.35**	99.10	98.36	98.52	98.69	98.68
	$\kappa(\times100)$	**99.68**	99.23	99.12	99.20	99.32	99.61
HC	OA(%)	**98.66**	97.95	98.22	98.30	98.45	98.48
	AA(%)	**98.41**	97.71	97.42	98.15	98.15	98.19
	$\kappa(\times100)$	**99.12**	98.01	98.11	99.02	99.05	99.08

ferential convolution, resolves spectral confusion in challenging crop categories, thus substantiating the pivotal role of explicit high-order interaction learning for HSI classification.

4.5. Ablation study

To validate the necessity of each component in the proposed framework, an ablation study is conducted by removing key modules and operations and evaluating performance on five HSI datasets. Table 8 shows five ablated configurations, including HorD^{2}CN-A and HorD^{2}CN-B, which remove the HSEDC and HSADC modules from the proposed model, respectively. HorD^{2}CN-C, which removes the ASES operation from the HSEDC module; HorD^{2}CN-D, which removes the DSAS operation from the HSADC module; and HorD^{2}CN-E, which retains only the differential convolution (DC) module without high-order interaction.

The ablation study, as shown in Table 9, conclusively validates the critical contribution of each component in HorD^{2}CN to its superior performance. Taking the IP dataset as an example, removing the HSEDC module from HorD^{2}CN degrades results, with OA dropping by 0.96 %, demonstrating its spectral adaptive capability. Similarly, eliminating the HSADC module reduces OA by 1.31 %, underscoring its capacity to model spatial structural details. The performance gap between HorD^{2}CN and variants lacking deformable operations (e.g., HorD^{2}CN-C/D) further confirms that the integration of spectral adaptability and spatial deformability is paramount for robust feature extraction in HSI classification.

Furthermore, the high-order design of HorD^{2}CN proves indispensable. The high-order differential convolution in HorD^{2}CN outperforms the standard DC in HorD^{2}CN-E, demonstrating its superior capacity to model intricate nonlinear spectral-spatial interactions beyond second-order statistics. The consistent superiority of HorD^{2}CN across all five datasets and evaluation metrics affirms that its unified framework, integrating deformable operations for geometric adaptability and high-order differential convolution for fine-grained feature interaction, achieves an optimal balance for HSI classification.

5. Conclusions

In this paper, a novel high-order deformable differential convolution network (termed HorD^{2}CN) is proposed for HSI classification, designed to effectively capture the complex high-order features inherent in HSI data and generate a robust feature representation. The proposed HorMSDC block aggregates multi-scale spectral-spatial high-order features, comprising two key components: the HSEDC module, which extracts subtle spectral variations among different land cover types and facilitates the learning of discriminative spectral features, and the HSADC module, which models local spatial structural details and enhances spatial feature representation. Additionally, deformable operations dynamically adjust the positions of convolution kernels by introducing learnable offsets, enabling the model to better adapt to the complex geometric shapes and structural changes of ground objects. Experimental results on five publicly available HSI datasets demonstrate that HorD^{2}CN outperforms ten state-of-the-art deep learning methods, underscoring its superior performance in HSI classification. These results validate the crucial role of high-order feature interactions and the representation of detailed information in improving HSI classification accuracy.

Future work will focus on optimizing the differential operation methodology to improve computational efficiency, further enhancing the spectral-spatial high-order feature interactions, and expanding the framework to additional remote sensing tasks such as target detection and image segmentation to explore its broader applicability and potential.

Declaration of generative AI and AI-assisted technologies in the writing process

No applicable.

CRediT authorship contribution statement

Zitong Zhang: Methodology, Writing – original draft, Writing – review & editing, Investigation, Conceptualization; **Fujie Jiang:** Writing – review & editing, Visualization, Supervision, Validation; **Chengcheng Zhong:** Writing – review & editing, Validation; **Qiaoyu Ma:** Software, Methodology.

Data availability

Data will be made available on request.

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 supported by the Postdoctoral Fellowship Program and China Postdoctoral Science Foundation under Grant BX20250435.

References

Aburaed, N., Alkhatib, M. Q., Marshall, S., Zabalza, J., & Al Ahmad, H. (2023). A review of spatial enhancement of hyperspectral remote sensing imaging techniques. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, 2275–2300.

Ahmad, M., Distifano, S., Khan, A. M., Mazzara, M., Li, C., Yao, J., Li, H., Aryal, J., Vivone, G., & Hong, D. (2024a). A comprehensive survey for hyperspectral image classification: The evolution from conventional to transformers. arXiv preprint arXiv:2404.14955, .

Ahmad, M., Ghous, U., Usama, M., & Mazzara, M. (2024b). WaveFormer: Spectral–spatial wavelet transformer for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, *21*, 1–5.

Bian, L., Wang, Z., Zhang, Y., Li, L., Zhang, Y., Yang, C., Fang, W., Zhao, J., Zhu, C., Meng, Q. et al. (2024). A broadband hyperspectral image sensor with high spatio-temporal resolution. *Nature*, *635*(8037), 73–81.

Camps-Valls, G., Tuia, D., Bruzzone, L., & Benediktsson, J. A. (2013). Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Processing Magazine*, *31*(1), 45–54.

Chang, C.-I., Xiong, W., & Wen, C.-H. (2014). A theory of high-order statistics-based virtual dimensionality for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(1), 188–208. https://doi.org/10.1109/TGRS.2012.2237554

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., & Wei, Y. (2017). Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision (ICCV)*.

Ding, Y., Zhang, Z., Zhao, X., Hong, D., Cai, W., Yang, N., & Wang, B. (2023). Multi-scale receptive fields: Graph attention neural network for hyperspectral image classification. *Expert Systems with Applications*, *223*, 119858. https://doi.org/10.1016/j.eswa.2023.119858

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*. https://openreview.net/forum?id=YicbFdNTTy.

Fan, K.-C., & Hung, T.-Y. (2014). A novel local pattern descriptor-local vector pattern in high-order derivative space for face recognition. *IEEE Transactions on Image Processing*, *23*(7), 2877–2891. https://doi.org/10.1109/TIP.2014.2321495

Fang, Y., Sun, L., Zheng, Y., & Wu, Z. (2025). Deformable convolution-enhanced hierarchical transformer with spectral-spatial cluster attention for hyperspectral image classification. *IEEE Transactions on Image Processing*, *34*, 701–716. https://doi.org/10.1109/TIP.2024.3522809

Fei, X., Wu, S., Miao, J., Wang, G., & Sun, L. (2024). Lightweight-VGG: A fast deep learning architecture based on dimensionality reduction and nonlinear enhancement for hyperspectral image classification. *Remote Sensing*, *16*(2), 259.

Feng, S., Zhang, H., Xi, B., Zhao, C., Li, Y., & Chanussot, J. (2024). Cross-domain few-shot learning based on decoupled knowledge distillation for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–14.

Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752, .

Gu, A., Goel, K., & Ré, C. (2021). Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396, .

Gualtieri, A. G., Chettri, S., Cromp, R. F., & Johnson, L. F. (1999). Support vector machine classifiers as applied to AVIRIS data. https://api.semanticscholar.org/CorpusID:15743704.

Hasan, H., Shafri, H. Z. M., & Habshi, M. (2019). A comparison between support vector machine (SVM) and convolutional neural network (CNN) models for hyperspectral image classification. In *IOP Conference series: Earth and environmental science* (p. 012035). IOP Publishing (*vol. 357*).

Jiang, Y., Zhou, H., Zhang, Z., Zhang, C., & Zhang, K. (2023). S^{2}MoINet: Spectral-spatial multi-order interactions network for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, 7135–7150.

Jijón-Palma, M. E., Kern, J., Amisse, C., & Centeno, J. A. S. (2021). Improving stacked-autoencoders with 1D convolutional-nets for hyperspectral image land-cover classification. *Journal of Applied Remote Sensing*, *15*(2), 026506.

Kaur, P., SINGH, M. P., MISHRA, A. M., SHANKAR, A., SINGH, P., DIWAKAR, M., & NAYAK, S. R. (2023). DELM: Deep ensemble learning model for multiclass classification of super-resolution leaf disease images. *Turkish Journal of Agriculture and Forestry*, *47*(5), 727–745.

Larry L. Biehl, M. F. B., & Landgrebe, D. A. (2015). 220 Band AVIRIS hyperspectral image data set: June 12, 1992 Indian pine test site 3. https://doi.org/10.4231/R7RX991C

Lee, H., & Kwon, H. (2017). Going deeper with contextual CNN for hyperspectral image classification. *IEEE Transactions on Image Processing*, *26*(10), 4843–4855.

Li, G., Huang, Q., Wang, W., & Liu, L. (2025). Selective and multi-scale fusion Mamba for medical image segmentation. *Expert Systems with Applications*, *261*, 125518. https://doi.org/10.1016/j.eswa.2024.125518

Li, W., Chen, H., Liu, Q., Liu, H., Wang, Y., & Gui, G. (2022). Attention mechanism and depthwise separable convolution aided 3DCNN for hyperspectral remote sensing image classification. *Remote Sensing*, *14*(9), 2215.

Li, Y., Luo, Y., Zhang, L., Wang, Z., & Du, B. (2024). MambaHSI: Spatial-spectral Mamba for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–16. https://doi.org/10.1109/TGRS.2024.3430985

Liu, H., Dai, Z., So, D., & Le, Q. V. (2021). Pay attention to MLPs. *Advances in neural information processing systems*, *34*, 9204–9215.

Liu, Q., Yue, J., Fang, Y., Xia, S., & Fang, L. (2024). HyperMamba: A spectral-spatial adaptive Mamba for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–14.

Liu, Z., Qiu, C.-J., Song, Y.-Q., Liu, X.-H., Wang, J., & Sheng, V. S. (2019). Texture feature extraction from thyroid MR imaging using high-order derived mean CLBP. *Journal of Computer Science and Technology*, *34*, 35–46.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, *24*(7), 971–987.

Qin, B., Feng, S., Zhao, C., Xi, B., Li, W., & Tao, R. (2024a). FDGNet: Frequency disentanglement and data geometry for domain generalization in cross-scene hyperspectral image classification. *IEEE Transactions on Neural Networks and Learning Systems*, *36*(6),10297–10310.

Qin, B., Feng, S., Zhao, C., Xi, B., Li, W., Tao, R., & Li, Y. (2024b). Hyperspherical structural-aware distillation enhanced spatial-spectral bidirectional interaction network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *62*,1–14.

Rao, Y., Zhao, W., Tang, Y., Zhou, J., Lim, S. N., & Lu, J. (2022). HorNet: Efficient high-order spatial interactions with recursive gated convolutions. In *Advances in neural information processing systems*. (*vol. 35*).

Shao, Y., Liu, J., Yang, J., & Wu, Z. (2022). Spatial–spectral involution MLP network for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 9293–9310.

Su, Z., Liu, W., Yu, Z., Hu, D., Liao, Q., Tian, Q., Pietikäinen, M., & Liu, L. (2021). Pixel difference networks for efficient edge detection. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 5117–5127).

Tang, T., Liu, J., Luo, X., Gao, X., & Pan, X. (2022). Triple-branch ternary-attention mechanism network with deformable 3D convolution for hyperspectral image classification. *International Journal of Remote Sensing*, *43*(12), 4352–4377.

Tolstikhin, I., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., Lucic, M., & Dosovitskiy, A. (2021). MLP-Mixer: An all-MLP architecture for vision. (*vol. 34*). In *Advances in Neural Information Processing Systems* (pp. 24261–24272).

Torun, O., Yuksel, S. E., Erdem, E., Imamoglu, N., & Erdem, A. (2024). Hyperspectral image denoising via self-modulating convolutional neural networks. *Signal Processing*, *214*, 109248.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(11), 2579–2605.

Verma, A., & Yadav, A. K. (2025). FusionNet: Dual input feature fusion network with ensemble based filter feature selection for enhanced brain tumor classification. *Brain Research*, *1852*, 149507.

Wang, C., Liu, B., Liu, L., Zhu, Y., Hou, J., Liu, P., & Li, X. (2021). A review of deep learning used in the hyperspectral image analysis for agriculture. *Artificial Intelligence Review*, *54*(7), 5205–5253.

Wang, G., Zhang, X., Peng, Z., Zhang, T., & Jiao, L. (2025). S^{2}Mamba: A spatial-spectral state space model for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *63*, 1–13.

Waswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *NIPS* (pp. 5999–6009).

Wu, H., Xiao, B., Codella, N., Liu, M., Dai, X., Yuan, L., & Zhang, L. (2021). CVT: Introducing convolutions to vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 22–31).

Xie, Z., Zhang, W., Sheng, B., Li, P., & Chen, C. L. P. (2021). BaGFN: Broad attentive graph fusion network for high-order feature interactions. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(8), 4499–4513.

Yang, C., Hu, S., Tang, L., Deng, R., Zhou, G., Yi, J., & Chen, A. (2024). A barking emotion recognition method based on Mamba and synchrosqueezing short-time fourier transform. *Expert Systems with Applications*, *258*, 125213. https://doi.org/10.1016/j.eswa.2024.125213

Yang, J., Li, A., Qian, J., Qin, J., & Wang, L. (2023). A hyperspectral image classification method based on pyramid feature extraction with deformable-dilated convolution. *IEEE Geoscience and Remote Sensing Letters*, *21*, 1–5.

Yu, Z., Wan, J., Qin, Y., Li, X., Li, S. Z., & Zhao, G. (2020a). NAS-FAS: Static-dynamic central difference network search for face anti-spoofing. *IEEE transactions on pattern analysis and machine intelligence*, *43*(9), 3005–3023.

Yu, Z., Zhao, C., Wang, Z., Qin, Y., Su, Z., Li, X., Zhou, F., & Zhao, G. (2020b). Searching central difference convolutional networks for face anti-spoofing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5295–5305).

Yu, Z., Zhou, B., Wan, J., Wang, P., Chen, H., Liu, X., Li, S. Z., & Zhao, G. (2021). Searching multi-rate and multi-modal temporal enhanced networks for gesture recognition. *IEEE Transactions on Image Processing*, *30*, 5626–5640.

Zhang et al. (2024a). LDS^{2}MLP: A novel learnable dilated spectral-spatial MLP for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 17207–17220.

Zhang, B., Gao, Y., Zhao, S., & Liu, J. (2010). Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor. *IEEE Transactions on Image Processing*, *19*(2), 533–544. https://doi.org/10.1109/TIP.2009.2035882

Zhang, M., Liu, L., Jin, Y., Lei, Z., Wang, Z., & Jiao, L. (2024b). Tree-shaped multiobjective evolutionary CNN for hyperspectral image classification. *Applied Soft Computing*, *152*, 111176.

Zhang, R., Li, G., Qu, S., Wang, J., & Peng, J. (2025a). Mamba-GIE: A visual state space models-based generalized image extrapolation method via dual-level adaptive feature fusion. *Expert Systems with Applications*, *264*, 125961. https://doi.org/10.1016/j.eswa.2024.125961

Zhang, Y., Cao, G., Li, X., & Wang, B. (2018). Cascaded random forest for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(4), 1082–1094.

Zhang, Z., Feng, H., Zhang, C., Ma, Q., & Li, Y. (2024c). S^{2}DCN: Spectral-spatial difference convolution network for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 3053–3068.

Zhang, Z., Peng, B., & Zhao, T. (2025b). An ultra-lightweight network combining Mamba and frequency-domain feature extraction for pavement tiny-crack segmentation. *Expert Systems with Applications*, *264*, 125941. https://doi.org/10.1016/j.eswa.2024.125941

Zhao, C., Zhu, W., & Feng, S. (2022). Superpixel guided deformable convolution network for hyperspectral image classification. *IEEE Transactions on Image Processing*, *31*, 3838–3851.

Zhao, Z., Xu, X., Li, S., & Plaza, A. (2024). Hyperspectral image classification using groupwise separable convolutional vision transformer network. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–17.

Zhong, C., Gong, N., Zhang, Z., Jiang, Y., & Zhang, K. (2023). LiteCCLKNet: A lightweight criss-cross large kernel convolutional neural network for hyperspectral image classification. *IET Computer Vision*, *17*(7), 763–776.

Zhong, C., Zhang, K., Zhang, Z., Jiang, Y., & Zhang, C. (2025). DF^{2}Net: Deformable fourier filter network for hyperspectral image classification. *Applied Intelligence*, *55*(7), 1–20.

Zhong, Y., Hu, X., Luo, C., Wang, X., Zhao, J., & Zhang, L. (2020). WHU-Hi: UAV-Borne hyperspectral with high spatial resolution (H-2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. *Remote Sensing of Environment*, *250*. https://doi.org/10.1016/j.rse.2020.112012

Zhou, H., Zhang, X., Zhang, C., & Ma, Q. (2023). Vision transformer with contrastive learning for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, *20*, 1–5.

Zhu, J., Fang, L., & Ghamisi, P. (2018). Deformable convolutional neural networks for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, *15*(8), 1254–1258.

Zhu, X., Hu, H., Lin, S., & Dai, J. (2019). Deformable ConvNets V2: More deformable, better results. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*.