HyperMODE: A Continuous-Depth Spectral–Spatial Modeling Framework with Mamba and Neural Ordinary Differential Equations for Hyperspectral Image Classification

Jialin Tang[a], Yunduan Lou[b], Yanhui Guo[c] and Yu Bai[a,*]

[a]College of Engineering and Computer Science, California State University, Fullerton, Fullerton, CA, 92831, USA
[b]Computational Science Research Center, University of California, Irvine, Irvine, CA, 92697, USA
[c]School of Artificial Intelligence, Shandong Women's University, Jinan, 250300, China

ARTICLE INFO

Keywords:
Hyperspectral image classification
Mamba
Neural Ordinary Differential Equations
Continuous–discrete hybrid modeling
Deep learning

ABSTRACT

Hyperspectral image (HSI) classification requires jointly capturing fine-grained spectral–spatial patterns and long-range contextual dependencies. However, existing convolutional, Transformer-based, and recent state-space architectures mainly rely on fixed discrete layer-wise transformations, which may lead to fragmented intermediate feature updates and insufficient coordination between multi-scale local representation learning and global spatial propagation. To address this issue, we propose HyperMODE, a unified continuous–discrete spectral–spatial modeling framework for HSI classification. Specifically, HyperMODE first constructs multi-scale spectral–spatial representations from compact input patches, and then couples Neural ODE-based continuous-depth feature evolution with scale-diverse Mamba-based spatial propagation through an interval-coupled pipeline. In this framework, Neural ODE enables smoother and more consistent feature transformation in the reduced representation space, while the multi-branch Mamba architecture performs efficient long-range spatial propagation and cross-scale context interaction with linear complexity. By unifying continuous feature evolution and efficient spatial propagation, HyperMODE establishes a coherent hybrid paradigm for spectral–spatial representation learning. Extensive experiments on four public HSI benchmarks, including Pavia University, WHU-Hi-HanChuan, WHU-Hi-LongKou, and Houston 2013, demonstrate stable and competitive classification performance, validating the effectiveness of the proposed framework. The source code will be made available at https://github.com/JaylenTang/HyperMODE.

1. Introduction

Hyperspectral image (HSI) classification is a fundamental task in remote sensing, where hundreds of narrow and contiguous spectral bands provide rich material-specific signatures beyond conventional RGB imagery, supporting a wide range of applications such as precision agriculture, environmental monitoring, and mineral exploration (Wang et al., 2021; Lu et al., 2020; Peyghambari & Zhang, 2021). Despite its great potential, HSI classification remains challenging due to high spectral dimensionality, strong inter-band redundancy, complex spectral–spatial variability, and the scarcity of labeled samples, all of which place stringent demands on representation learning frameworks.

Convolutional neural networks (CNNs) have long served as important backbones for HSI analysis because of their effectiveness in local spectral–spatial feature extraction, as demonstrated by early deep CNN frameworks and subsequent spectral–spatial architectures (Chen et al., 2016; Paoletti et al., 2019; Ben Hamida et al., 2018; Roy et al., 2020). However, the locality of convolution operations inherently limits their receptive fields, making it difficult to capture long-range contextual interactions and global spectral–spatial dependencies in large-scale hyperspectral scenes. To overcome this limitation, Transformer-based models have been introduced into HSI classification (He et al., 2020; Hong et al., 2022; Peng et al., 2022; He et al., 2024a). By means of self-attention, Transformers explicitly model global interactions among spectral–spatial tokens and enable adaptive context aggregation beyond fixed convolutional receptive fields. Nevertheless, the quadratic computational and memory complexity of self-attention with respect to token length may hinder its efficiency in high-resolution scenes or dense spectral–spatial tokenization settings.

To improve efficiency while maintaining strong representation capability, recent studies have explored hybrid and enhanced architectures that combine local feature extraction with global dependency modeling (Wan et al., 2026; Ding et al., 2025a; Xu et al., 2024; Xi et al., 2025). These methods generally achieve a favorable trade-off between accuracy and computational cost. However, most existing CNN–Transformer or related hybrid frameworks still rely on stacked discrete layer-wise transformations, where intermediate spectral–spatial representations are updated through fixed-depth mappings. Such purely discrete updates may lead to fragmented feature transitions and insufficient coordination between local representation learning and global context propagation. At the same time, efficient long-range spatial modeling remains crucial for HSI classification, especially under limited computational budgets. Recent structured architectures, including efficient Transformers and

*Corresponding author
tjl_0516@csu.fullerton.edu (J. Tang); yundual@uci.edu (Y. Lou); guoyanhui@sdwu.edu.cn (Y. Guo); ybai@fullerton.edu (Y. Bai)
ORCID(s): 0009-0006-8813-0513 (J. Tang); 0000-0002-2303-1120 (Y. Bai)

Mamba-based state-space models, have shown promising ability to capture long-range dependencies with improved efficiency (Zhao et al., 2024; Wang et al., 2025a). Despite these advances, existing HSI frameworks have not sufficiently explored how to unify robust multi-scale spectral–spatial initialization, continuous-depth feature evolution, and efficient long-range spatial propagation within a coherent spectral–spatial modeling framework. To address this issue, we propose HyperMODE, a continuous–discrete spectral–spatial modeling framework for hyperspectral image classification. Specifically, a Scale-Diverse Feature Extraction (SDFE) front-end is introduced to provide robust multi-scale spectral–spatial initialization. On top of these representations, a Neural Ordinary Differential Equation (Neural ODE) module is employed to model continuous-depth feature evolution in the reduced representation space, thereby alleviating the fragmentation introduced by purely discrete layer-wise updates. To complement this continuous feature dynamics, a scale-diverse multi-branch Mamba architecture is adopted to perform efficient long-range spatial propagation and cross-scale context interaction with linear complexity. By integrating multi-scale initialization, continuous-depth feature evolution, and efficient state-space spatial propagation within a unified continuous–discrete pipeline, HyperMODE establishes a coherent framework for spectral–spatial representation learning. The main contributions of this work are summarized as follows:

- We propose HyperMODE, a unified continuous–discrete spectral–spatial modeling framework for hyperspectral image classification, which integrates Neural ODE-based continuous-depth feature evolution with Mamba-based state-space spatial propagation.
- We develop an interval-coupled hybrid pipeline that alternates continuous feature evolution and efficient state-space spatial propagation across scales, establishing a coherent connection between the two modeling paradigms.
- We integrate time-modulated Neural ODE evolution with a scale-diverse Mamba design to improve feature consistency, long-range spatial modeling, and cross-scale interaction with linear computational complexity.
- Extensive experiments on four public hyperspectral benchmarks demonstrate the effectiveness and competitiveness of the proposed framework.

2. Related Work

2.1. CNN-Based Methods for HSI Classification

Deep convolutional neural networks (CNNs) have long served as a fundamental backbone for hyperspectral image (HSI) classification due to their effectiveness in hierarchical spectral–spatial feature extraction. Early deep-learning-based hyperspectral studies explored spectral–spatial modeling using convolutional architectures and recurrent variants to capture contextual dependencies (Hu et al., 2020). Subsequent works introduced autoencoder-based and adversarial learning strategies to enhance representation learning and improve classification robustness under complex scene variations (Ma et al., 2016; Zhong et al., 2020).

To further strengthen feature learning, hierarchical and multiscale CNN frameworks were proposed to capture information across different receptive fields and spatial contexts (Yang et al., 2023; Li et al., 2019). Meanwhile, lightweight and efficient 3D CNN architectures were developed to reduce computational overhead while preserving spectral–spatial modeling capability (Ahmad et al., 2022). Fully contextual convolutional networks have also been explored to improve global scene understanding and semantic consistency in hyperspectral classification (Wang et al., 2022).

Despite these advances, CNN-based approaches remain inherently constrained by localized convolutional operations, limiting their ability to capture long-range spectral–spatial dependencies and motivating the exploration of more scalable non-local modeling architectures.

2.2. Transformer-Based Methods for HSI Classification

Transformer architectures have recently gained increasing attention in hyperspectral image (HSI) classification due to their ability to model global dependencies through self-attention mechanisms. Early attempts adapted Transformer encoders to spectral–spatial token sequences, demonstrating improved contextual modeling capability compared with convolutional networks (Mei et al., 2022; Yang et al., 2022). Subsequent studies introduced hybrid convolution–Transformer designs and morphological attention mechanisms to enhance spectral–spatial interaction and improve representation learning under complex scene variations (Zhang et al., 2022; Roy et al., 2023).

More recent works further explored hierarchical tokenization and multiscale attention strategies to improve long-range modeling efficiency, including multigranularity Transformers and cross-attention-based frameworks (Tang et al., 2025a; Sun et al., 2024). To enhance training efficiency and representation robustness, recent studies have explored masked spectral pretraining, graph-aware modeling, and factorized attention designs, which improve generalization capability in hyperspectral classification (Tian et al., 2024; Mohamed et al., 2024; Jiang et al., 2024).

However, the quadratic computational complexity of self-attention and its reliance on stacked discrete layer-wise transformations remain key challenges for modeling high-dimensional hyperspectral data efficiently, especially when both long-range dependency modeling and smooth intermediate feature evolution are desired.

2.3. State-Space Models and Mamba

State-space models (SSMs) have recently emerged as an efficient paradigm for long-sequence modeling by describing feature propagation through latent dynamical systems. The Mamba architecture introduces selective state-space modeling, enabling linear-time sequence processing while maintaining strong long-range dependency modeling capability (Gu & Dao, 2023). By dynamically modulating state transitions conditioned on input features, Mamba achieves scalable sequence modeling efficiency and alleviates the quadratic computational burden of self-attention.

Recent studies have extended Mamba to hyperspectral image classification by integrating spectral–spatial feature learning within state-space formulations. Early attempts such as MambaHSI and HyperMamba explored spectral–spatial sequence modeling using selective state transitions (Li et al., 2024; Liu et al., 2024). Subsequent works further introduced hierarchical or multi-branch structures, including 3DSS-Mamba and IGroupSS-Mamba, to enhance multi-scale representation learning and long-range dependency propagation (He et al., 2024b,c). More recent architectures, such as S2Mamba, SpiralMamba, and MambaMoE, incorporate advanced spatial scanning strategies or expert mixtures to improve modeling flexibility and robustness under complex hyperspectral scenes (Wang et al., 2025b; Tang et al., 2025b; Xu et al., 2026). Beyond pure Mamba-based designs, hybrid architectures have also emerged to leverage the complementary advantages of state-space models and attention mechanisms. For example, HSI-MFormer integrates Transformer experts for short-range modeling with Mamba experts for long-range sequence learning, achieving multiscale spectral–spatial representation through hybrid expert fusion (He et al., 2025).

Despite these advances, most existing Mamba-based HSI frameworks still rely on discrete token propagation and fixed-depth state transitions. As a result, they mainly emphasize efficient long-range dependency modeling, while the role of continuous-depth feature evolution in improving the consistency of intermediate spectral–spatial representations remains less explored.

2.4. Neural Ordinary Differential Equations and Continuous-Depth Feature Modeling

Neural Ordinary Differential Equations (Neural ODEs) introduce a continuous-depth modeling paradigm by parameterizing feature transformations as differential equations, allowing hidden representations to evolve along continuous trajectories rather than through fixed discrete layers (Chen et al., 2018). By reformulating deep networks as dynamical systems, Neural ODEs provide a flexible view of representation learning in which intermediate features are updated through continuous-depth dynamics instead of strictly layer-wise mappings. This perspective has shown promise for providing a more flexible and continuous parameterization of intermediate feature transformations in deep models.

Recent studies have extended ODE-based formulations to hyperspectral image analysis and broader remote sensing tasks. For example, Neural ODE frameworks have been explored for hyperspectral representation learning and remote sensing scene understanding under challenging observation conditions, while Dual ODE introduces coupled differential modeling to capture multidimensional feature interactions. In addition, recent HSI-specific studies such as HyperKAN-ODE further demonstrate the potential of Neural ODE-based continuous dynamic modeling for hyperspectral classification (Paoletti et al., 2020; Zhang et al., 2021; Metzger et al., 2022; Zhang et al., 2024; Chen & Ye, 2024; Zhang et al., 2026). These studies suggest that continuous-depth modeling provides a useful alternative to purely discrete architectures, particularly when smoother intermediate feature evolution and more flexible dynamic transformations are desired.

Despite these advantages, purely continuous modeling frameworks often lack efficient mechanisms for capturing complex long-range spatial dependencies and hierarchical contextual interactions. Consequently, recent studies have increasingly explored hybrid paradigms that combine continuous feature transformation with scalable discrete spatial modeling. Motivated by these observations, HyperMODE is designed as a unified continuous–discrete framework that couples Neural ODE-based continuous-depth feature evolution with Mamba-based state-space spatial propagation. In this design, the Neural ODE module is used to model continuous feature transformations in the reduced representation space, while the scale-diverse Mamba structure enables efficient long-range spatial propagation and cross-scale interaction, together establishing a unified spectral–spatial representation learning framework with improved efficiency and modeling flexibility.

3. Methodology

This section begins with the preliminaries of continuous-depth feature modeling and selective state-space propagation, which establish the theoretical foundation of the proposed continuous–discrete learning paradigm. Following this, we present an overview of the HyperMODE framework for hyperspectral image (HSI) classification, where shallow spectral–spatial representations are first extracted by a Scale-Diverse Feature Extraction Module (SDFE), followed by hierarchical representation learning across multiple continuous–discrete stages. Subsequently, a Neural Ordinary Differential Equation (Neural ODE)-based feature evolution module is introduced to model continuous-depth transformations in the reduced representation space, while a scale-diverse Mamba-based hybrid block performs efficient long-range spatial propagation via selective state-space modeling with linear computational complexity. Finally, an interval-coupled architecture that alternates continuous feature evolution and discrete state-space spatial propagation is presented, forming a unified spectral–spatial representation learning framework that balances modeling flexibility and computational efficiency.

3.1. Preliminaries

This subsection introduces the theoretical preliminaries of continuous-depth feature modeling and selective state-space propagation, which form the mathematical foundation of the proposed HyperMODE framework.

3.1.1. Continuous-Depth Feature Modeling

Instead of updating intermediate representations solely through fixed discrete layers, Neural Ordinary Differential Equations (Neural ODEs) describe feature transformation as a continuous-time dynamical system:

$$\frac{d\mathbf{h}(t)}{dt} = f_\theta(\mathbf{h}(t), t), \tag{1}$$

where $\mathbf{h}(t) \in \mathbb{R}^N$ denotes the hidden feature state at continuous depth t, and t denotes the continuous-depth variable (i.e., pseudo-time) used for feature evolution during ODE integration, rather than a physical wavelength index, spectral band index, or PCA component index. Here, $f_\theta(\cdot)$ represents a learnable nonlinear transformation parameterized by convolutional operators. In HyperMODE, this formulation is used to model continuous-depth feature evolution in the reduced representation space.

Unlike fixed layer-wise mappings, the ODE formulation enables the feature transformation to evolve continuously with depth through a time-dependent vector field $f_\theta(\mathbf{h}(t), t)$. In this work, the role of t is to parameterize continuous-depth feature evolution, so that the transformation behavior can vary smoothly along the integration trajectory.

To ensure computational efficiency and memory stability, the continuous dynamics are solved using a first-order explicit Euler integration scheme:

$$\mathbf{h}_{n+1} = \mathbf{h}_n + \Delta t\, f_\theta(\mathbf{h}_n, t_n), \tag{2}$$

where Δt denotes a fixed integration interval. Compared with deep discrete stacking, this formulation provides a lightweight approximation of continuous-depth feature transformation while preserving computational efficiency for high-dimensional hyperspectral data. In our implementation, HyperMODE does not rely on adaptive-step numerical solvers. Instead, the flexibility of the proposed continuous-depth modeling comes from the time-dependent vector field $f_\theta(\mathbf{h}(t), t)$, which allows the feature transformation pattern to change smoothly with continuous depth.

3.1.2. Selective State-Space Modeling

State-space models (SSMs) describe sequence propagation through latent dynamical states. Given an input token sequence $\mathbf{u}_t$, a discrete state-space system can be formulated as

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t, \tag{3}$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t, \tag{4}$$

where $\mathbf{x}_t$ denotes the hidden state and $\mathbf{A}, \mathbf{B}, \mathbf{C}$ represent state transition and projection matrices.

Unlike classical linear time-invariant SSMs, the selective state-space mechanism adopted in HyperMODE dynamically modulates transition parameters conditioned on the input sequence:

$$\mathbf{B}_t = \mathrm{Projection}_B(\mathbf{u}_t), \tag{5}$$

$$\mathbf{C}_t = \mathrm{Projection}_C(\mathbf{u}_t), \tag{6}$$

$$\delta_t = \tau\big(\mathrm{Projection}_\delta(\mathbf{u}_t)\big), \tag{7}$$

where δ_t denotes the adaptive discretization factor controlling the selective propagation behavior. It should be noted that this adaptive discretization belongs to the selective state-space propagation mechanism in Mamba, and is different from the fixed-step ODE integration used in the Neural ODE module. This formulation enables content-aware state transitions and efficient long-range spatial dependency modeling with linear computational complexity, consistent with the Mamba selective state-space architecture.

3.1.3. Continuous–Discrete Hybrid Formulation

While Neural ODEs model continuous-depth feature evolution, selective state-space models efficiently propagate long-range spatial dependencies. In HyperMODE, these two paradigms are coupled through an interval-based hybrid architecture. For notational simplicity, the branch index is omitted here, and the hybrid propagation is described at the stage level. At the l-th stage, features are first updated via continuous-depth transformation, followed by feed-forward refinement and scale-diverse Mamba propagation:

$$\widehat{\mathbf{Z}}^{(l)} = \mathrm{ODE}\big(\mathbf{Z}^{(l-1)}\big), \tag{8}$$

$$\widetilde{\mathbf{Z}}^{(l)} = \mathrm{FFN}\big(\widehat{\mathbf{Z}}^{(l)}\big), \tag{9}$$

$$\mathbf{Z}^{(l)} = \mathrm{Mamba}\big(\widetilde{\mathbf{Z}}^{(l)}\big). \tag{10}$$

This interval-coupled formulation reflects the practical implementation of HyperMODE, where continuous feature evolution and discrete state-space spatial propagation are alternated to achieve stable spectral–spatial representation learning.

3.2. HyperMODE: Overview

Given an input hyperspectral cube $\mathbf{X} \in \mathbb{R}^{H\times W\times B}$, principal component analysis (PCA) is first applied as a lightweight preprocessing step to reduce spectral redundancy and computational burden, yielding a compact representation $\mathbf{X}_{\mathrm{PCA}} \in \mathbb{R}^{H\times W\times L}$. The transformed image is further partitioned into a set of 3-D patch cubes $\{\mathbf{img} \in \mathbb{R}^{P\times P\times L}\}$, where each patch is labeled according to its central pixel.

The overall architecture of HyperMODE is illustrated in Fig. 1. The network consists of three main components: a scale-diverse feature extraction module, an interval-coupled continuous–discrete hybrid propagation module, and a prediction head.

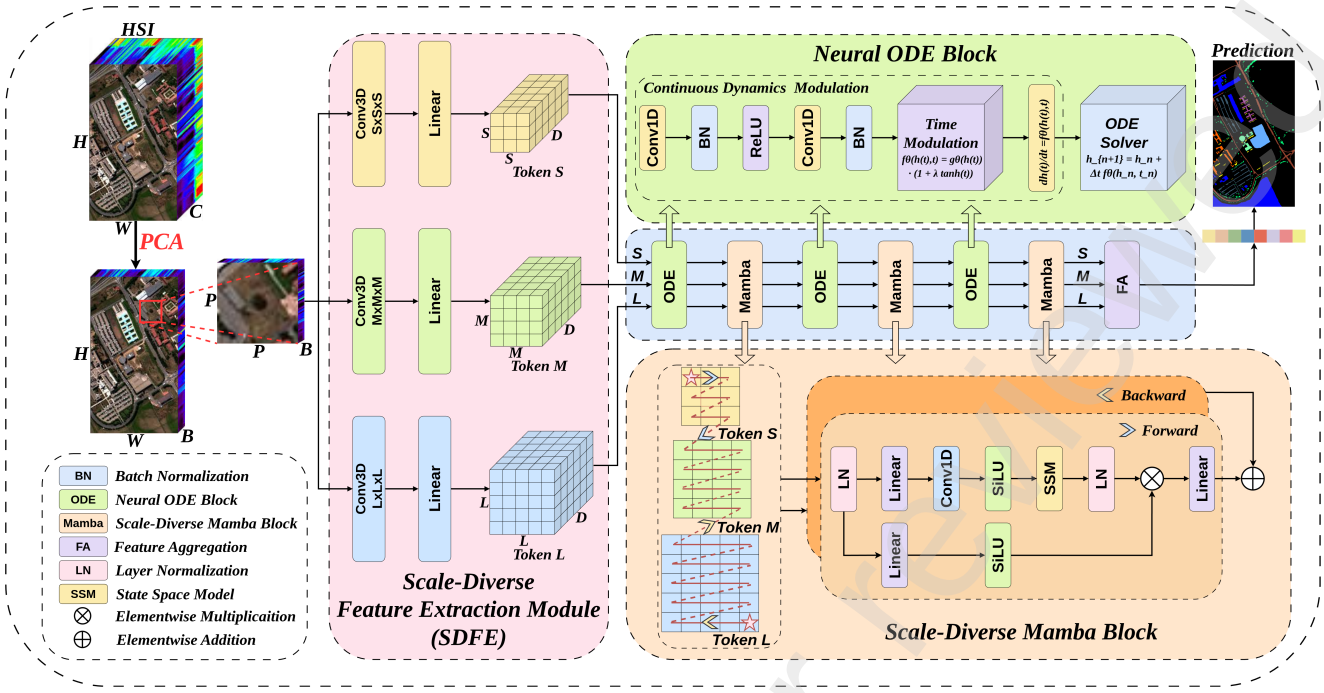


Figure 1: Overall framework of the proposed HyperMODE architecture. Three parallel multi-scale token branches (Token S/M/L) are first constructed by the scale-diverse feature extraction module. Within each branch, Neural ODE blocks and scale-diverse Mamba blocks are sequentially interleaved to form an interval-coupled continuous–discrete propagation pipeline, where the Neural ODE blocks model continuous-depth feature evolution in the reduced representation space and the Mamba blocks enable efficient long-range spatial propagation and cross-scale interaction.

3.2.1. Scale-Diverse Feature Extraction

Given the input patch cube **img**, shallow spectral–spatial representations are first generated by a Scale-Diverse Feature Extraction (SDFE) module. As shown in Fig. 1, the SDFE adopts three parallel 3D convolution branches with different receptive fields to produce multi-scale token groups $\{\mathbf{Z}_S, \mathbf{Z}_M, \mathbf{Z}_L\}$, providing multi-scale initialization for subsequent continuous–discrete representation learning.

3.2.2. Interval-Coupled Hybrid Propagation

After shallow embedding, the multi-scale tokens are fed into an interval-coupled hybrid propagation chain composed of Neural ODE blocks and scale-diverse Mamba blocks. The Neural ODE blocks perform continuous-depth feature evolution, while the Mamba blocks enable efficient long-range spatial propagation. These two modules are alternately arranged, forming the continuous–discrete hybrid structure illustrated in Fig. 1.

3.2.3. Feature Aggregation and Prediction

After hybrid propagation, the output tokens are fused by a feature aggregation (FA) module to obtain the final spectral–spatial representation. Global feature vectors are then fed into a linear classifier to produce the prediction results.

3.3. Scale-Diverse Feature Extraction Module (SDFE)

Hyperspectral scenes contain heterogeneous objects with varying spatial resolutions and spectral characteristics. To provide robust multi-scale initialization for the subsequent continuous–discrete modeling stages, HyperMODE employs a lightweight Scale-Diverse Feature Extraction (SDFE) module.

3.3.1. Multi-Branch 3D Encoding

Given an input hyperspectral patch $\mathbf{img} \in \mathbb{R}^{P\times P\times L}$, SDFE applies three parallel 3D convolution branches with different receptive fields:

$$\mathbf{F}_r = \sigma\big(\mathrm{BN}\big(\mathrm{Conv3D}_r(\mathbf{img})\big)\big), \tag{11}$$

where $r \in \{S, M, L\}$ denotes the branch index corresponding to the small-, medium-, and large-scale receptive fields, respectively, $\mathrm{BN}(\cdot)$ is batch normalization, and $\sigma(\cdot)$ is a nonlinear activation. These branches capture complementary spectral–spatial patterns at multiple scales.

3.3.2. Token Generation

The extracted volumetric features are projected into a shared embedding space:

$$\mathbf{E}_r = \mathrm{Embed}(\mathbf{F}_r), \tag{12}$$

and reshaped into token sequences

$$\mathbf{Z}_r^{(0)} \in \mathbb{R}^{N_r \times D}, \quad r \in \{S, M, L\}. \tag{13}$$

Here, D denotes the token embedding dimension in the shared embedding space after the projection Embed$(\cdot)$, and N_r is the number of tokens in the r-th branch after reshaping.

The resulting multi-scale tokens $\{\mathbf{Z}_S^{(0)}, \mathbf{Z}_M^{(0)}, \mathbf{Z}_L^{(0)}\}$ serve as the input to the interval-coupled Neural ODE and Mamba hybrid stages described next.

3.4. Neural ODE-Based Continuous-Depth Feature Evolution

3.4.1. Motivation

Most existing hyperspectral image classification models rely on stacked discrete layer-wise transformations, where intermediate representations are updated through fixed-depth mappings. Although effective, such purely discrete updates may lead to fragmented feature transitions and limit the flexibility of intermediate representation evolution. To address this issue, HyperMODE introduces a Neural Ordinary Differential Equation (Neural ODE)-based feature evolution mechanism, which reformulates intermediate feature transformation as a continuous-depth dynamical process. This formulation provides a continuous parameterization of feature evolution beyond rigid discrete stacking.

In HyperMODE, the Neural ODE module is not intended to model a literal physical differential process along the wavelength axis. Instead, it is used to describe continuous-depth feature evolution in the reduced representation space, where the dynamics are parameterized by lightweight convolutional operators. Compared with fixed discrete layer stacking, this design offers a more flexible feature transformation process and serves as a continuous counterpart to subsequent discrete spatial propagation. In this sense, the proposed formulation differs from simple residual stacking or feature smoothing, since it models feature evolution through a learnable vector field rather than a sequence of fixed discrete updates. Moreover, the proposed module does not depend on adaptive-step ODE solvers; instead, its flexibility arises from the explicit time dependence of the vector field, which enables depth-varying feature transformation under a lightweight fixed-step integration scheme.

3.4.2. Time-Modulated Feature Dynamics

To introduce depth-dependent modulation into the continuous feature evolution, the feature dynamics are parameterized as follows.

Given an input token sequence $\mathbf{Z} \in \mathbb{R}^{B \times N \times D}$, the ODE function is implemented as a compact feature mixing operator:

$$\mathbf{H}_1 = \mathrm{ReLU}\big(\mathrm{BN}(\mathrm{Conv1D}(\mathbf{Z}))\big), \tag{14}$$

$$\mathbf{H}_2 = \mathrm{BN}(\mathrm{Conv1D}(\mathbf{H}_1)). \tag{15}$$

To realize time-dependent feature transformation, a learnable scaling factor is applied:

$$f_\theta(\mathbf{Z}, t) = \mathbf{H}_2 \cdot (1 + \lambda \tanh(t)), \tag{16}$$

where λ controls the strength of the time modulation. This formulation introduces an explicit dependence on t, allowing the feature transformation behavior to vary continuously along depth during ODE integration. It should be noted that this design does not imply adaptive-step numerical integration; rather, it realizes time-varying continuous-depth feature dynamics through a depth-modulated vector field.

3.4.3. Explicit Euler Feature Evolution

The continuous dynamics are discretized using a first-order explicit Euler solver:

$$\mathbf{h}_{n+1} = \mathbf{h}_n + \Delta t \, f_\theta(\mathbf{h}_n, t_n), \tag{17}$$

which provides a lightweight approximation of continuous-depth feature transformation. Because a fixed-step explicit Euler scheme is adopted, the proposed ODE block should be interpreted as a lightweight continuous-depth feature evolution module with time-varying dynamics, rather than an adaptive-step Neural ODE solver. Compared with more complex adaptive or high-order solvers, the explicit Euler scheme significantly reduces memory overhead, making it suitable for high-dimensional hyperspectral representation learning. This formulation can also be viewed as a residual continuous-depth update, which stabilizes optimization while preserving feature identity mappings.

3.4.4. Neural ODE Integration within HyperMODE

Unlike standalone Neural ODE formulations that operate as independent continuous-depth networks, the proposed Neural ODE module is tightly coupled with the interval-based hybrid architecture of HyperMODE.

Given multi-scale token groups $\{\mathbf{Z}_S, \mathbf{Z}_M, \mathbf{Z}_L\}$, continuous-depth feature evolution is applied independently to each branch:

$$\widehat{\mathbf{Z}}_r = \mathrm{ODE}\big(\mathbf{Z}_r\big), \quad r \in \{S, M, L\}. \tag{18}$$

This branch-wise evolution preserves scale-specific representation patterns while maintaining consistent continuous transformations across hierarchical stages. The evolved tokens are subsequently propagated through selective state-space modeling, forming an interval-coupled continuous–discrete pipeline. Compared with conventional layer-wise token mixing, this design provides a more flexible feature evolution process while maintaining a unified continuous–discrete modeling paradigm within HyperMODE.

3.5. Scale-Diverse Mamba Hybrid Block

3.5.1. Motivation

Although Neural ODE modules provide continuous-depth feature evolution, spatial interactions in hyperspectral scenes remain highly non-local and scale-sensitive. Objects with different spatial resolutions require adaptive context aggregation beyond single-scale modeling. To address this limitation, we design a scale-diverse Mamba hybrid block that performs selective state-space propagation over architecturally independent branches while enabling cross-scale interaction. This design complements continuous feature dynamics by enabling efficient long-range spatial reasoning across hierarchical token groups.

3.5.2. Multi-Branch Selective State-Space Modeling

Let $\mathbf{Z}_S, \mathbf{Z}_M, \mathbf{Z}_L$ denote token groups generated by the SDFE module at different receptive-field scales. Each branch is independently processed using a Mamba-based state-space operator:

$$\widetilde{\mathbf{Z}}_r = \mathrm{Mamba}\big(\mathrm{LN}(\mathbf{Z}_r)\big), \quad r \in \{S, M, L\}, \tag{19}$$

where LN denotes layer normalization. The selective state-space mechanism performs bidirectional sequence propagation through forward and reverse scanning, allowing efficient modeling of global spatial context with linear computational complexity.

3.5.3. Spatial Reconstruction

The enhanced token sequences are reshaped into spatial feature maps:

$$\mathbf{X}_r = \mathrm{Reshape}(\widetilde{\mathbf{Z}}_r), \quad \mathbf{X}_r \in \mathbb{R}^{B\times C\times H_r\times W_r}, \quad r \in \{S, M, L\}. \tag{20}$$

This spatial reconstruction preserves spatial topology and prepares multi-scale features for subsequent cross-scale interaction.

3.5.4. Cross-Scale Feature Fusion

To enhance hierarchical representation learning, HyperMODE performs cross-scale fusion in which fine-resolution tokens integrate global contextual information from larger receptive fields. Given the fused token representation $\mathbf{Z}_r^{fuse}$, the output tokens are updated through a lightweight residual projection:

$$\mathbf{Z}_r^{out} = \mathbf{Z}_r + \mathrm{DropPath}\big(\mathrm{Proj}(\mathbf{Z}_r^{fuse})\big), \quad r \in \{S, M, L\}. \tag{21}$$

This formulation enables hierarchical spatial interaction while remaining consistent with the continuous-depth feature evolution introduced by the Neural ODE module.

4. Experiments

4.1. Datasets

To validate the classification performance of the proposed HyperMODE framework, extensive experimental comparisons are conducted on four publicly available hyperspectral image (HSI) benchmarks: Pavia University, WHU-Hi-HanChuan, WHU-Hi-LongKou, and Houston 2013. Table 1 summarizes the category definitions together with the numbers of training and testing samples for each class.

4.1.1. Pavia University

The Pavia University dataset was acquired by the ROSIS sensor over Pavia, Italy. After removing noisy and water-absorption bands, 103 spectral channels remain. The image size is 610×610 pixels with nine land-cover classes, making it a widely adopted benchmark for evaluating spectral–spatial representation learning. Due to its complex urban structure and spectral variability, it provides a challenging testbed for hyperspectral classification.

4.1.2. WHU-Hi-HanChuan

WHU-Hi-HanChuan is a UAV-based hyperspectral dataset collected over HanChuan City, China, covering 1217×303 pixels with more than 283k labeled samples. The dataset contains diverse land-cover categories, including vegetation and built-up regions, posing challenges for large-scale contextual modeling and long-range spatial dependency learning.

4.1.3. WHU-Hi-LongKou

WHU-Hi-LongKou was captured using a Headwall Nano-Hyperspec sensor mounted on a UAV platform. The spatial size is 550×400 pixels with over 204k labeled samples after preprocessing. Its agricultural–urban mixture and complex spectral–spatial variability make it a challenging benchmark for evaluating robust spectral–spatial representation learning.

4.1.4. Houston 2013

Houston 2013 was collected by the CASI sensor during the IEEE GRSS Data Fusion Contest. The dataset includes 144 spectral bands and a spatial size of 349×1905 pixels with 15 land-cover classes. Strong intra-class variability and frequent mixed pixels make it one of the most challenging urban hyperspectral benchmarks for spectral–spatial classification evaluation.

4.2. Experimental Settings

1) Evaluation Metrics: Following common hyperspectral image classification protocols, overall accuracy (OA), average accuracy (AA), and the kappa coefficient (Kappa) are adopted for quantitative evaluation. To reduce randomness caused by sampling and model initialization, each experiment is repeated ten times, and the mean performance together with standard deviation is reported.

2) Implementation Details: All experiments are implemented in Python 3.10 using the PyTorch 2.8.0 + CUDA

Table 1
Category names, color labels, and the numbers of training and testing samples per class on the Pavia University, WHU-Hi-HanChuan, WHU-Hi-LongKou, and Houston 2013 datasets.

	Pavia University			WHU-Hi-HanChuan			WHU-Hi-LongKou			Houston 2013		
No.	Category	Train	Test	Category	Train	Test	Category	Train	Test	Category	Train	Test
C1	Asphalt	66	6565	Strawberry	447	44288	Corn	345	34166	Healthy grass	125	1126
C2	Meadows	186	18463	Cowpea	227	22526	Cotton	84	8290	Stressed grass	125	1129
C3	Gravel	21	2078	Soybean	103	10184	Sesame	30	3001	Synthetic grass	70	627
C4	Trees	31	3033	Sorghum	54	5299	B-leaf soybean	632	62580	Trees	124	1120
C5	P-metal sheets	13	1332	Water spinach	12	1188	N-leaf soybean	42	4109	Soil	124	1118
C6	Bare Soil	50	4979	Watermelon	45	4488	Rice	119	11735	Water	33	292
C7	Bitumen	13	1317	Greens	59	5844	Water	670	66386	Residential	127	1141
C8	S-B-Bricks	37	3645	Trees	180	17798	R and houses	71	7053	Commercial	124	1120
C9	Shadows	10	937	Grass	95	9374	Mixed weed	52	5177	Road	125	1127
C10				Red roof	105	10411				Highway	123	1104
C11				Gray roof	169	16742				Railway	123	1112
C12				Plastic	37	3642				Parking Lot 1	123	1110
C13				Bare soil	91	9025				Parking Lot 2	47	422
C14				Road	186	18374				Tennis Court	43	385
C15				Bright object	11	1125				Running Track	66	594
C16				Water	754	74647						
Total		**427**	**42349**		**2575**	**254955**		**2045**	**202497**		**1502**	**13527**

12.8 framework and executed on a single NVIDIA RTX 3090 GPU (24GB). Models are trained for 100 epochs with a batch size of 64. Following realistic low-label settings, the training ratio is set to 1% for Pavia University, WHU-Hi-HanChuan, and WHU-Hi-LongKou, while a 10% split is adopted for Houston 2013 due to its limited annotated samples.

For the proposed HyperMODE framework, the scale-diverse feature extraction module employs three 3D convolution branches with kernel sizes of $5\times5\times5$, $7\times7\times7$, and $9\times9\times9$ to provide multi-scale spectral–spatial initialization. The Adam optimizer is adopted with an initial learning rate of 0.001, and the input patch size is fixed to 15×15. Principal Component Analysis (PCA) is applied as a preprocessing step to reduce spectral redundancy, with the reduced spectral dimension set to 30. Unless otherwise noted, this setting follows common practice in hyperspectral classification for fair comparison.

3) Competitive Approaches: To comprehensively evaluate the effectiveness of the proposed framework, several representative hyperspectral image classification methods from different architectural paradigms are selected for comparison, including recent Mamba-based state-space models and spectral–spatial Transformer-related or hybrid architectures. Specifically, the compared methods include MambaHSI (Li et al., 2024), 3DSS-Mamba (He et al., 2024b), HyperMamba (Liu et al., 2024), GraphGST (Jiang et al., 2024), Swin-MSP (Tian et al., 2024), MambaLG (Pan et al., 2025), LRDTN (Ding et al., 2025a), and HypyraMamba (Li et al., 2026), enabling a comprehensive comparison against recent representative state-of-the-art approaches.

4.3. Comparison With State-of-the-Art Methods

To comprehensively evaluate the proposed framework, quantitative comparisons between HyperMODE and recent competitive hyperspectral image classification methods are presented in Tables 2–5. Overall Accuracy (OA), Average Accuracy (AA), and the Kappa coefficient are adopted for evaluation under low-label training settings.

Pavia University (1% training samples): On the Pavia University dataset, HyperMODE achieves an OA of 97.78%, AA of 95.83%, and Kappa of 97.05, delivering the best overall performance under the same sampling ratio. Compared with recent Transformer- and Mamba-based architectures, the proposed framework also exhibits improved stability, reflected by smaller performance variance across repeated runs. As shown in Fig. 2, HyperMODE produces smoother class boundaries and reduces fragmented predictions in complex urban regions with strong spectral variability.

WHU-Hi-HanChuan (1% training samples): On the large-scale WHU-Hi-HanChuan dataset, HyperMODE achieves an OA of 97.71%, AA of 94.49%, and Kappa of 97.32 under low-label training settings. The results show clear improvements over recent state-space and Transformer-based models, especially on minority categories affected by class imbalance. Fig. 3 shows that the proposed method maintains more consistent spatial structures and fewer noisy regions across heterogeneous UAV scenes.

WHU-Hi-LongKou (1% training samples): On WHU-Hi-LongKou, HyperMODE reaches an OA of 99.55%, AA of 98.43%, and Kappa of 99.41. Although many methods already achieve high accuracy on this dataset, the proposed framework maintains superior stability with minimal standard deviation, indicating robust and consistent feature modeling under the low-label setting. The visual comparisons in Fig. 4 further confirm more coherent classification maps and reduced boundary artifacts.

Houston 2013 (10% training samples): On the Houston 2013 dataset, HyperMODE obtains an OA of 99.51%, AA of 99.56%, and Kappa of 99.47 under the 10% training split. Compared with the UAV datasets, Houston exhibits higher spectral dimensionality and stronger intra-class variability, posing greater challenges for stable feature modeling. The

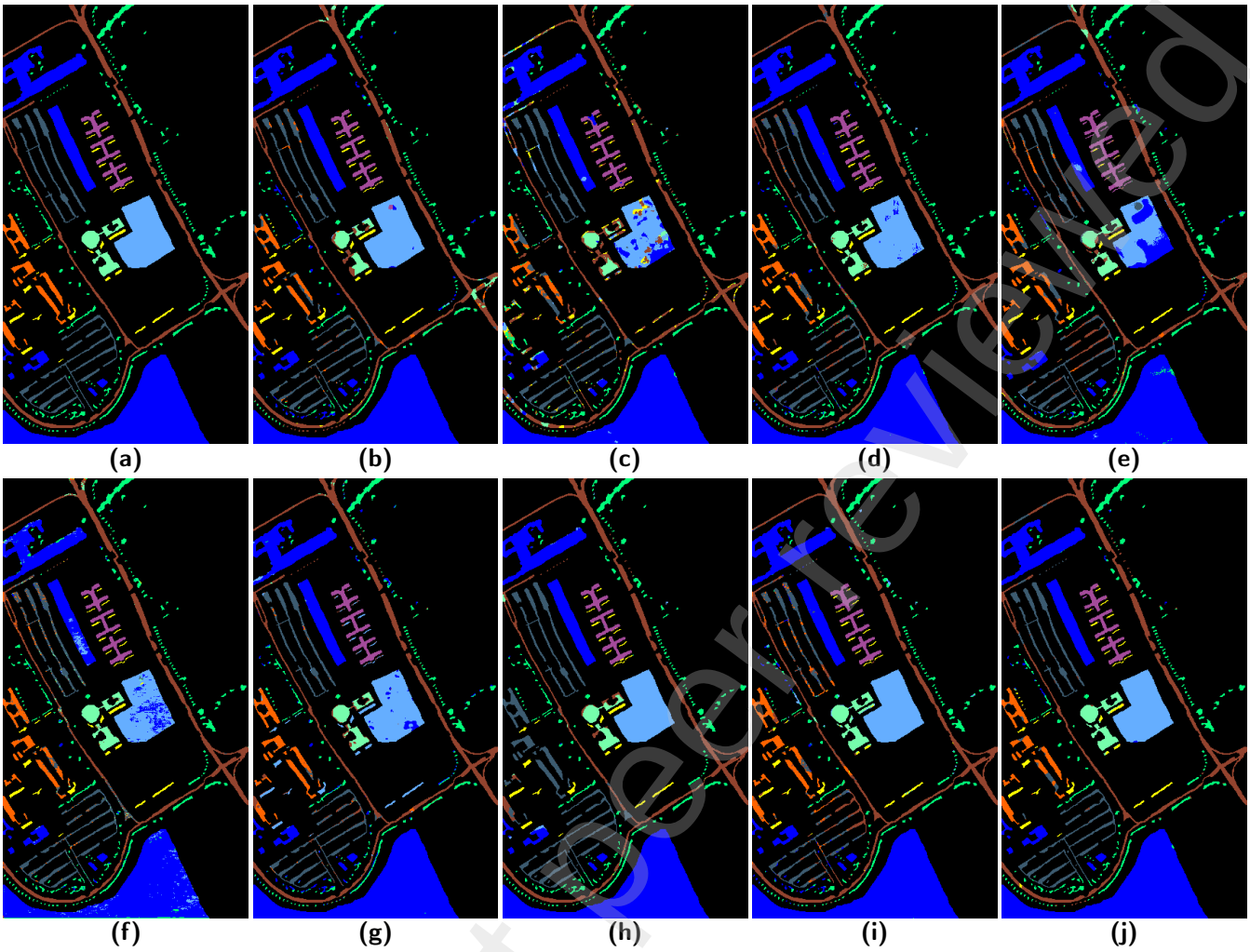


Figure 2: Classification maps of the Pavia University dataset produced by different methods under the 1% training-sample setting. (a) Ground Truth, (b) MambaHSI, (c) 3DSS-Mamba, (d) HyperMamba, (e) GraphGST, (f) Swin-MSP, (g) MambaLG, (h) LRDTN, (i) HypyraMamba, and (j) HyperMODE.

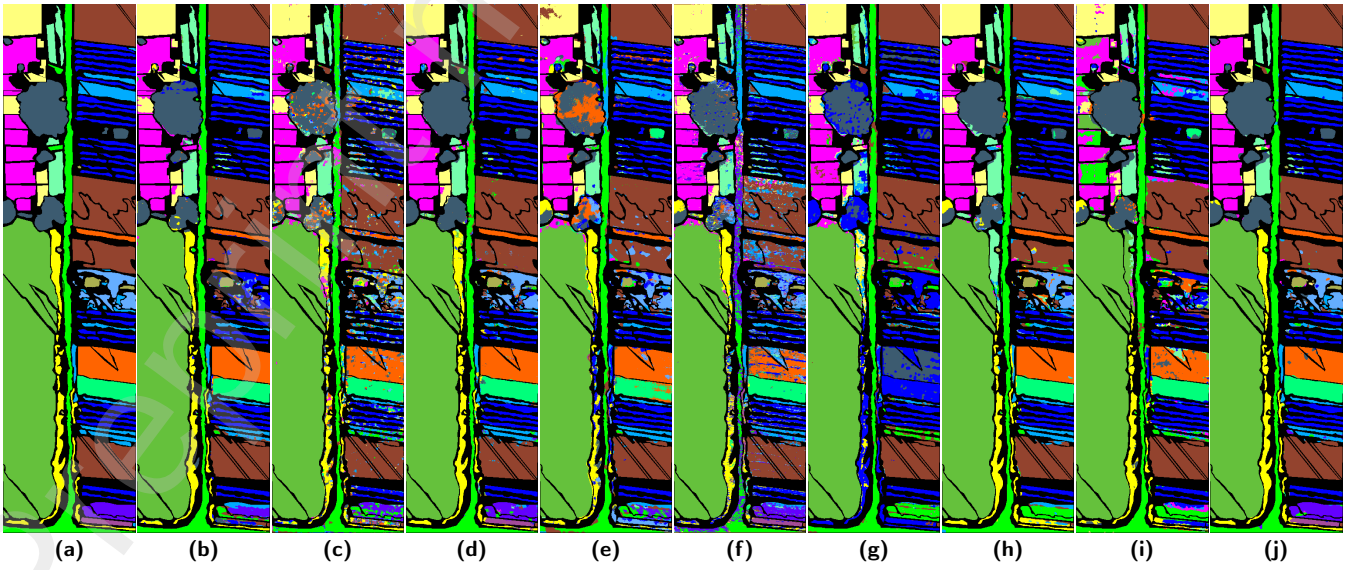


Figure 3: Classification maps of the WHU-Hi-HanChuan dataset produced by different methods under the 1% training-sample setting. (a) Ground Truth, (b) MambaHSI, (c) 3DSS-Mamba, (d) HyperMamba, (e) GraphGST, (f) Swin-MSP, (g) MambaLG, (h) LRDTN, (i) HypyraMamba, and (j) HyperMODE.

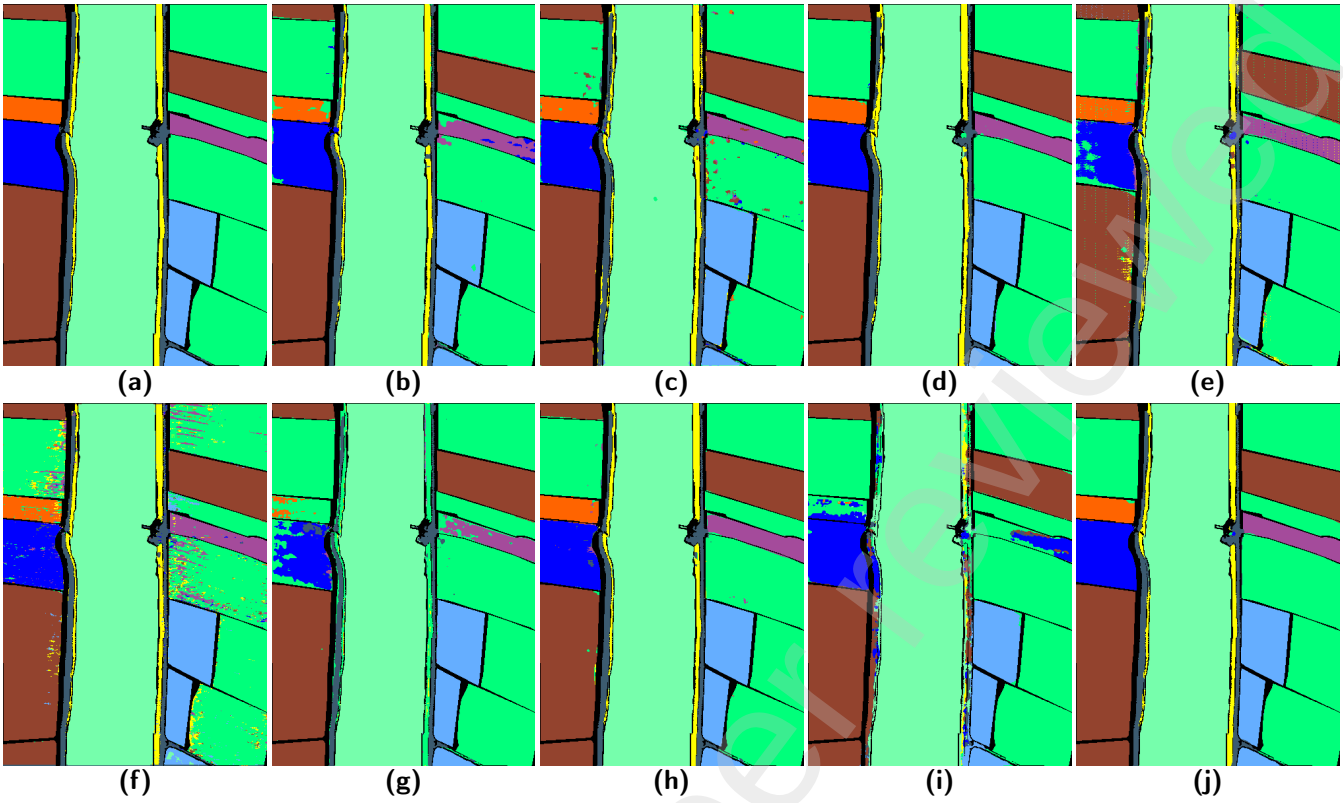


Figure 4: Classification maps of the WHU-Hi-LongKou dataset produced by different methods under the 1% training-sample setting. (a) Ground Truth, (b) MambaHSI, (c) 3DSS-Mamba, (d) HyperMamba, (e) GraphGST, (f) Swin-MSP, (g) MambaLG, (h) LRDTN, (i) HypyraMamba, and (j) HyperMODE.

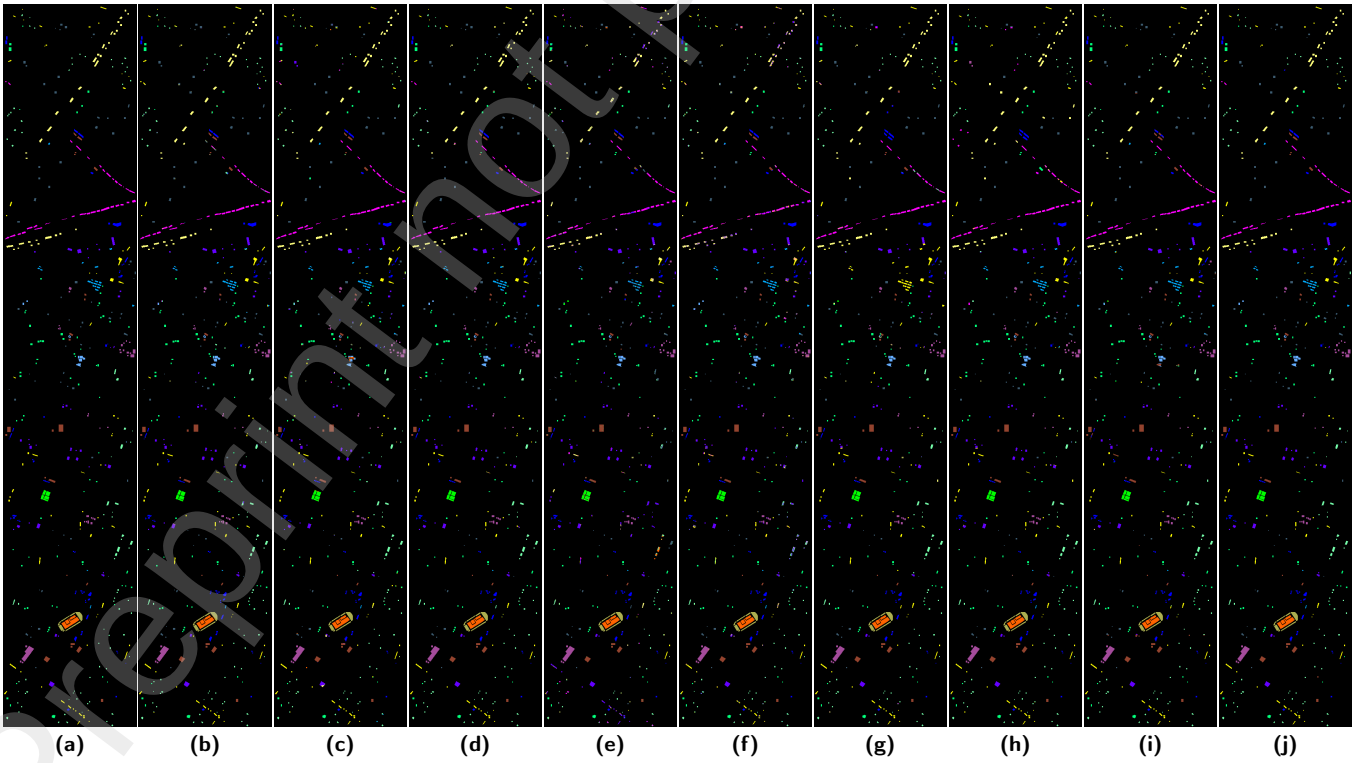


Figure 5: Classification maps of the Houston 2013 dataset produced by different methods under the 10% training-sample setting. (a) Ground Truth, (b) MambaHSI, (c) 3DSS-Mamba, (d) HyperMamba, (e) GraphGST, (f) Swin-MSP, (g) MambaLG, (h) LRDTN, (i) HypyraMamba, and (j) HyperMODE.

Table 2
Quantitative comparison on the Pavia University dataset under the 1% training-sample setting. Results are reported as mean ± standard deviation over ten runs. The best results are highlighted in bold.

No.	MambaHSI 2024	3DSS-Mamba 2024	HyperMamba 2024	GraphGST 2024	Swin-MSP 2024	MambaLG 2025	LRDTN 2025	HypyraMamba 2026	**HyperMODE Ours**
1	96.11±2.14	89.08±4.25	93.87±1.80	94.43±1.21	90.64±2.44	**97.61±1.53**	96.59±1.86	95.65±2.74	97.52±0.35
2	99.58±0.36	97.14±1.48	99.89±0.10	97.48±1.57	89.73±6.84	99.49±0.49	99.54±0.21	99.85±0.10	**99.95±0.03**
3	83.88±7.16	80.57±10.97	84.26±3.56	84.77±2.16	**86.64±11.67**	32.30±32.87	76.98±25.94	77.83±8.29	85.43±1.95
4	82.71±5.45	83.96±3.59	**93.65±2.81**	89.47±2.34	93.06±4.93	78.22±5.34	93.53±1.79	88.70±3.66	90.98±0.89
5	99.99±0.02	96.84±3.90	99.95±0.07	**100.00±0.00**	**100.00±0.00**	58.29±47.73	99.59±0.55	99.35±0.67	99.58±0.11
6	99.28±0.68	76.01±4.93	97.09±1.37	83.90±9.22	78.70±14.09	82.77±16.89	96.12±4.63	99.68±0.59	**99.74±0.12**
7	84.39±6.77	65.72±6.58	97.06±2.95	92.44±3.00	95.01±3.03	41.40±37.40	62.18±32.64	94.27±10.83	**99.77±0.22**
8	93.29±2.99	89.68±2.91	83.97±3.72	86.40±3.35	81.52±17.45	85.25±16.56	93.84±3.85	91.94±9.08	**97.29±1.09**
9	97.36±2.04	61.71±25.14	92.69±3.68	97.67±1.15	**99.94±0.10**	30.61±40.80	97.28±1.73	56.93±17.24	92.22±1.55
OA(%)	95.97±0.65	89.24±1.90	95.80±0.50	93.19±1.47	88.65±2.28	86.56±4.57	95.44±1.29	95.48±0.88	**97.78±0.20**
AA(%)	92.95±1.43	82.30±4.67	93.60±0.70	91.84±1.62	90.58±2.93	67.32±11.11	90.63±3.79	89.36±2.91	**95.83±0.38**
Kappa	93.52±3.34	85.67±2.55	94.41±0.67	90.91±2.00	85.06±2.97	81.66±6.45	93.94±1.73	94.38±1.84	**97.05±0.26**

Table 3
Quantitative comparison on the WHU-Hi-HanChuan dataset under the 1% training-sample setting. Results are reported as mean ± standard deviation over ten runs. The best results are highlighted in bold.

No.	MambaHSI 2024	3DSS-Mamba 2024	HyperMamba 2024	GraphGST 2024	Swin-MSP 2024	MambaLG 2025	LRDTN 2025	HypyraMamba 2026	**HyperMODE Ours**
1	98.67±0.78	94.88±1.60	98.54±0.42	97.35±0.62	78.13±8.62	97.67±0.82	98.90±0.52	98.26±1.38	**99.43±0.12**
2	94.17±2.70	83.33±6.16	95.78±1.09	90.26±2.99	73.43±10.73	65.23±35.95	**96.91±1.11**	94.78±2.81	96.56±0.51
3	**97.74±2.14**	78.67±6.25	96.74±1.21	88.39±5.57	64.00±15.06	26.94±29.56	95.70±3.02	87.84±12.08	97.29±0.28
4	96.12±1.70	92.56±2.03	97.70±1.44	92.54±4.92	96.92±2.34	24.06±34.28	87.23±29.17	95.82±1.98	**99.31±0.10**
5	32.94±27.36	55.69±13.69	75.29±8.80	61.11±30.20	**93.12±4.75**	0.05±0.15	0.00±0.00	0.00±0.00	92.96±1.45
6	31.04±19.96	44.33±7.42	82.95±2.76	64.91±9.01	54.66±13.05	0.82±1.82	71.82±9.47	3.11±3.46	**85.84±0.80**
7	92.59±4.04	86.98±3.59	**94.31±2.43**	90.03±3.10	86.56±6.83	19.63±24.36	92.16±3.10	62.92±21.25	93.21±0.64
8	93.25±2.11	77.48±5.38	95.07±1.25	76.60±12.21	66.27±9.55	63.80±27.06	87.95±3.88	88.52±4.37	**97.08±0.40**
9	90.67±5.71	65.56±7.94	93.34±1.36	82.89±11.05	67.30±12.48	7.28±9.18	91.63±7.71	63.25±14.35	**98.22±0.38**
10	94.36±1.09	94.52±2.53	99.01±0.51	97.78±3.22	94.08±2.23	66.47±38.47	96.83±1.54	87.26±7.25	**99.43±0.20**
11	97.09±2.20	92.70±2.92	**98.81±0.44**	95.54±2.11	88.86±2.87	71.93±21.42	94.36±3.75	87.87±13.63	98.74±0.20
12	51.86±18.06	71.96±9.88	89.00±5.61	72.50±13.77	67.03±9.33	0.00±0.00	45.63±37.72	10.07±29.98	**92.79±1.64**
13	74.45±5.97	65.32±6.21	82.99±2.88	69.73±5.50	62.82±18.82	20.70±22.01	82.73±2.86	49.14±24.63	**86.06±0.97**
14	**97.04±0.67**	88.23±2.30	95.12±1.01	91.96±3.23	70.23±14.64	77.84±26.38	94.35±1.94	92.80±3.44	96.94±0.28
15	58.14±4.06	52.48±8.98	77.47±7.42	55.97±8.56	**94.03±1.92**	0.78±2.26	83.58±4.43	0.43±1.29	78.09±4.66
16	99.56±0.29	99.50±0.19	99.85±0.07	99.35±0.67	97.51±0.95	98.80±1.43	99.63±0.25	89.90±29.97	**99.87±0.02**
OA(%)	94.03±1.35	88.77±1.09	96.69±0.19	91.96±1.86	82.01±3.66	72.20±3.81	94.33±0.77	87.31±1.65	**97.71±0.11**
AA(%)	81.23±4.16	77.76±2.68	92.00±0.65	82.93±4.51	78.43±3.94	40.12±4.57	82.46±3.20	62.47±2.70	**94.49±0.45**
Kappa	93.01±1.45	86.85±1.27	96.12±0.22	90.58±2.18	79.16±4.11	66.84±4.74	93.36±0.90	83.91±2.06	**97.32±0.13**

results suggest that coupling Neural ODE-based continuous-depth feature evolution with state-space spatial propagation improves robustness in complex urban scenes. As illustrated in Fig. 5, HyperMODE produces more coherent classification regions with fewer misclassified pixels.

Overall, HyperMODE consistently achieves superior or competitive performance across datasets with different spatial scales, spectral distributions, and sampling ratios, validating the effectiveness of the proposed continuous–discrete spectral–spatial modeling framework.

4.4. Computational Complexity

We further analyze the computational efficiency of HyperMODE in terms of parameters, FLOPs, training time, and testing time, as summarized in Table 6. HyperMODE achieves competitive classification performance while maintaining a favorable balance between accuracy and computational cost. In particular, HyperMODE requires 0.1877 M parameters and 0.0729 GFLOPs, indicating good parameter efficiency and a relatively lightweight overall design.

It is worth noting that the proposed framework adopts a multi-scale spectral–spatial design, where three scale branches are processed in parallel and each branch alternates Mamba-based state-space propagation with Neural ODE-based continuous-depth feature evolution across three stages. This design is introduced to preserve strong classification performance and improve robustness across heterogeneous hyperspectral scenes. Although the multi-scale parallel structure and repeated ODE–Mamba interactions inevitably increase computational overhead, the overall complexity remains manageable due to the lightweight Euler-based ODE solver and the linear-complexity Mamba backbone.

In terms of runtime performance, HyperMODE achieves a training time of 23.58 s and a testing time of 11.01 s. While some lightweight baselines exhibit faster inference

Table 4
Quantitative comparison on the WHU-Hi-LongKou dataset under the 1% training-sample setting. Results are reported as mean ± standard deviation over ten runs. The best results are highlighted in bold.

No.	MambaHSI 2024	3DSS-Mamba 2024	HyperMamba 2024	GraphGST 2024	Swin-MSP 2024	MambaLG 2025	LRDTN 2025	HypyraMamba 2026	**HyperMODE Ours**
1	99.60±0.12	99.75±0.13	99.97±0.02	99.44±0.54	97.42±1.79	99.72±0.13	99.81±0.08	99.90±0.07	**99.99±0.01**
2	97.24±0.76	97.97±1.38	99.41±0.56	85.94±8.77	94.38±4.04	32.24±33.23	99.37±0.38	94.99±7.63	**99.71±0.26**
3	86.74±3.17	92.00±4.38	**98.93±0.61**	87.89±17.32	89.73±11.49	19.80±35.35	97.60±0.51	47.00±21.83	97.48±1.00
4	99.36±0.14	99.38±0.54	99.66±0.09	98.14±1.08	89.05±4.66	99.11±0.47	99.56±0.14	99.76±0.14	**99.83±0.06**
5	89.92±5.47	94.83±1.71	96.66±1.48	82.08±14.76	92.49±7.06	11.54±24.28	95.10±1.28	20.67±20.01	**99.11±0.19**
6	99.47±0.19	97.98±0.93	**99.94±0.04**	97.71±2.32	97.82±0.25	71.03±42.01	99.29±0.21	97.47±2.21	99.66±0.08
7	**99.99±0.01**	99.82±0.09	99.95±0.03	99.98±0.01	99.58±0.19	**99.99±0.01**	99.97±0.00	**99.99±0.01**	99.94±0.02
8	94.05±1.95	93.38±2.95	95.26±1.04	94.02±1.64	81.66±26.49	68.88±32.25	**97.54±0.22**	67.62±21.09	95.96±0.56
9	93.28±2.89	89.09±5.04	90.91±1.61	79.07±12.88	89.39±6.77	27.05±30.36	**97.46±0.88**	44.38±18.84	94.19±1.54
OA(%)	98.81±0.12	98.78±0.28	99.36±0.07	97.33±1.05	94.48±2.34	89.29±3.12	99.49±0.06	94.61±1.47	**99.55±0.05**
AA(%)	95.52±0.67	96.02±0.84	97.85±0.31	91.59±5.00	92.39±4.42	58.82±9.33	98.42±0.26	74.64±6.59	**98.43±0.26**
Kappa	98.46±0.39	98.39±0.36	99.16±0.10	96.48±1.39	92.84±2.99	85.58±4.30	99.33±0.08	93.07±2.14	**99.41±0.07**

Table 5
Quantitative comparison on the Houston 2013 dataset under the 10% training-sample setting. Results are reported as mean ± standard deviation over ten runs. The best results are highlighted in bold.

No.	MambaHSI 2024	3DSS-Mamba 2024	HyperMamba 2024	GraphGST 2024	Swin-MSP 2024	MambaLG 2025	LRDTN 2025	HypyraMamba 2026	**HyperMODE Ours**
1	99.12±0.63	98.38±1.85	99.80±0.45	96.80±1.58	93.61±3.75	91.32±4.40	89.45±6.43	98.59±1.09	**99.92±0.05**
2	98.59±1.07	98.42±0.74	99.42±0.50	97.48±1.27	98.38±1.31	95.04±5.13	95.40±5.52	97.20±2.11	**99.98±0.04**
3	99.89±0.14	99.71±0.52	99.89±0.18	99.38±0.29	99.78±0.37	89.04±29.71	**99.97±0.06**	99.92±0.11	99.90±0.08
4	98.87±0.66	96.54±2.64	**99.60±0.48**	97.35±1.79	97.32±5.40	96.87±2.63	96.56±2.75	97.69±0.32	99.37±0.47
5	99.99±0.03	99.62±0.84	**100.00±0.00**	99.99±0.03	98.46±2.52	90.10±29.68	99.98±0.05	99.76±0.47	**100.00±0.00**
6	94.93±4.49	94.32±10.18	98.33±1.75	99.35±1.40	99.71±0.27	36.35±37.71	87.71±0.52	82.72±1.92	**100.00±0.00**
7	**98.13±1.53**	93.23±3.03	97.83±0.67	89.35±3.33	82.68±6.86	95.18±2.09	94.08±1.87	96.88±0.88	97.90±0.43
8	93.90±1.72	93.92±5.66	97.21±2.25	94.75±2.14	72.69±15.85	82.17±8.53	84.61±8.73	93.88±1.80	**97.75±0.09**
9	96.78±2.09	96.46±2.54	96.34±1.98	87.02±5.04	71.52±12.14	93.59±2.55	94.21±2.61	94.39±2.26	**99.96±0.11**
10	99.15±0.89	99.01±0.74	98.60±1.14	95.17±2.04	79.92±9.66	92.25±10.40	97.73±3.25	99.88±0.22	**100.00±0.00**
11	95.92±2.52	97.90±1.28	98.92±1.46	97.10±2.58	64.23±23.64	88.62±6.55	90.71±5.98	97.00±1.35	**100.00±0.00**
12	97.72±1.41	97.19±2.06	98.74±1.06	94.83±3.40	59.48±22.64	94.99±3.85	95.57±2.39	95.87±1.11	**99.64±0.00**
13	93.75±1.24	97.99±1.13	94.11±3.40	84.63±4.62	96.97±4.69	51.44±35.18	93.41±1.73	94.20±3.18	**99.00±0.51**
14	**100.00±0.00**	99.40±1.08	99.87±0.25	99.97±0.08	98.89±1.30	80.78±34.61	**100.00±0.00**	99.97±0.08	**100.00±0.00**
15	96.41±1.22	99.76±0.31	**100.00±0.00**	99.88±0.30	99.41±0.54	89.66±29.89	99.75±0.30	**100.00±0.00**	**100.00±0.00**
OA(%)	97.73±0.58	97.34±1.21	98.65±0.23	95.30±0.71	84.67±4.70	88.99±3.37	94.40±0.66	97.05±0.43	**99.51±0.04**
AA(%)	97.54±0.63	97.46±1.51	98.58±0.30	95.54±0.68	87.54±3.94	84.49±4.28	94.61±0.54	96.53±0.45	**99.56±0.05**
Kappa	97.88±2.26	97.12±1.31	98.54±0.24	94.92±0.77	83.43±5.08	88.07±3.66	93.94±0.72	97.43±1.86	**99.47±0.05**

due to simplified modeling capacity, HyperMODE provides a more balanced trade-off between efficiency and accuracy, avoiding the excessive latency observed in high-complexity architectures.

Overall, these results demonstrate that the proposed continuous–discrete spectral–spatial framework can achieve strong classification performance without incurring prohibitive computational cost.

4.5. Ablation Study

4.5.1. Effect of Patch Size

We first investigate the influence of spatial patch size on the performance of HyperMODE, as summarized in Table 7 and visualized in Fig. 6. Overall, moderate patch sizes lead to more stable and accurate predictions across datasets. Specifically, a patch size of 15×15 achieves the best overall performance on Pavia University, WHU-Hi-LongKou, and Houston 2013, while 17×17 shows a slight advantage on

Table 6
Computational complexity comparison of different methods in terms of parameters, FLOPs, training time, and testing time. All results are measured under the same experimental setting.

Metrics	MambaHSI 2024	3DSS-Mamba 2024	HyperMamba 2024	GraphGST 2024	Swin-MSP 2024	MambaLG 2025	LRDTN 2025	HyPyraMamba 2026	**HyperMODE Ours**
Params (M)	0.1681	0.0103	0.0810	0.6790	5.0455	0.1337	0.0880	1.4694	0.1877
FLOPs (G)	9.1976	0.0003	0.0030	0.0007	0.2272	25.6318	0.0059	79.0054	0.0729
Training Time (s)	120.15	2.02	134.69	4.69	11.83	63.54	411.27	492.73	23.58
Testing Time (s)	0.30	0.71	12.12	2.04	3.16	0.27	1.37	1156.12	11.01

WHU-Hi-HanChuan, suggesting that datasets with richer spatial variability may benefit from slightly larger receptive fields.

When the patch size is too small (e.g., 11×11), the model lacks sufficient spatial context, resulting in reduced AA and Kappa scores. Conversely, excessively large patches (e.g., 19×19) introduce redundant background information and may weaken discriminative spectral–spatial cues, leading to marginal performance degradation. The trends observed in Fig. 6 indicate that HyperMODE benefits from a balanced receptive field, where moderate spatial context effectively complements the continuous-depth feature evolution modeled by the Neural ODE module.

4.5.2. Effect of Kernel-Size Combinations

We further analyze the influence of different kernel-size combinations, as summarized in Table 8 and illustrated in Fig. 7. In general, medium-scale configurations provide the most balanced performance across datasets, whereas the best-performing kernel combination remains dataset-dependent.

The smaller-kernel configuration (3-4-5) achieves the best results on Pavia University and WHU-Hi-HanChuan, indicating strong modeling ability for fine-grained spatial patterns. The medium-scale configuration (5-7-9) delivers the most consistent overall behavior and performs particularly well on WHU-Hi-LongKou. In contrast, the larger-kernel configuration (7-9-11) is more suitable for large-scale scenes and achieves the best results on Houston 2013.

To further evaluate stability, we compute the average ranking across OA, AA, and Kappa over all datasets. The medium-scale kernel configuration achieves the most balanced ranking overall, supporting the design motivation of HyperMODE that moderate receptive fields provide effective spatial priors for continuous-depth feature evolution and long-range spatial dependency modeling.

4.5.3. Effectiveness of SDFE, Mamba, and ODE Modules

To validate the contribution of each component, we conduct an ablation study by selectively enabling the SDFE, Mamba, and ODE modules, as shown in Table 9. Using only the SDFE (Conv3D) module results in the lowest performance, indicating that purely local spectral–spatial modeling is insufficient for capturing complex contextual dependencies in hyperspectral data.

Introducing either the Mamba backbone or the Neural ODE module consistently improves OA, AA, and Kappa across all datasets, demonstrating that both efficient state-space spatial propagation and continuous-depth feature evolution contribute to more robust representation learning. In particular, the ODE module provides clear gains on datasets with more complex feature variability, while the Mamba branch strengthens long-range spatial modeling. Notably, the full HyperMODE configuration (Conv3D + Mamba + ODE) achieves the best performance on all four datasets, confirming the complementary nature of the two modules.

Together with the patch-size analysis, these results reveal a consistent design principle: HyperMODE favors moderate spatial receptive fields rather than extreme local or global configurations, highlighting the effectiveness of balanced spatial priors in continuous–discrete spectral–spatial modeling.

5. Discussion

5.1. Continuous-Depth Feature Evolution and Stability Analysis

Across datasets with varying spectral–spatial complexity and sampling conditions, HyperMODE demonstrates consistently stable behavior. In particular, the reduced performance variance observed in Tables 2–5 suggests that continuous-depth feature evolution contributes to stabilizing intermediate representation updates during training.

Unlike purely discrete layer-wise transformations, the Neural ODE module models feature transformation through a continuous-depth trajectory, which encourages smoother representation updates and mitigates abrupt feature fluctuations. This behavior appears particularly beneficial in challenging scenarios such as WHU-Hi-HanChuan, where class imbalance and limited training samples increase learning difficulty. The improvements in AA and Kappa under low-label settings indicate that continuous-depth feature evolution may serve as an implicit regularization mechanism, leading to more consistent predictions across minority classes.

Furthermore, the patch-size ablation results demonstrate that a balanced spatial scale is sufficient once continuous-depth feature refinement is introduced. This suggests that the observed stability gains arise from the synergy between moderate spatial priors and smoother feature evolution, rather than relying solely on excessively enlarged spatial context. These observations also provide the basis for its complementary integration with Mamba-based spatial propagation, where smoother intermediate feature evolution can be further enhanced by efficient long-range contextual reasoning.

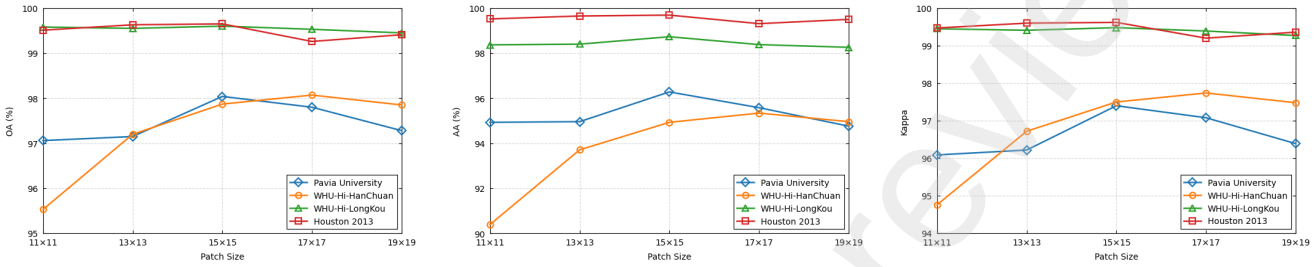
5.2. Complementarity Between Neural ODE and Mamba

The dataset-wise results highlight a clear division of roles between the Neural ODE module and Mamba-based state-space modeling. While the Neural ODE module focuses on continuous-depth feature evolution, the Mamba backbone efficiently propagates long-range spatial dependencies with linear complexity. This complementary interaction is reflected not only in their distinct functional roles, but also in their iterative cooperation within the interval-coupled pipeline, leading to consistent performance gains across both urban scenes (Pavia University, Houston 2013) and UAV datasets (WHU-Hi series).

Kernel-size ablation further supports this interpretation. Medium-scale kernel combinations achieve the most balanced performance across datasets, indicating that once

Table 7
Patch-size ablation study of HyperMODE on four hyperspectral datasets. The best results are highlighted in bold.

Patch Size	Pavia University			WHU-Hi-HanChuan			WHU-Hi-LongKou			Houston 2013		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
11×11	97.06	94.93	96.09	95.53	90.39	94.76	99.58	98.37	99.45	99.51	99.52	99.47
13×13	97.15	94.96	96.22	97.20	93.71	96.72	99.55	98.40	99.41	99.63	99.65	99.60
15×15	**98.04**	**96.28**	**97.40**	97.87	94.93	97.50	**99.60**	**98.73**	**99.48**	**99.65**	**99.69**	**99.62**
17×17	97.80	95.58	97.08	**98.07**	**95.34**	**97.74**	99.53	98.38	99.39	99.26	99.31	99.20
19×19	97.28	94.76	96.39	97.85	94.96	97.48	99.45	98.26	99.27	99.41	99.50	99.36

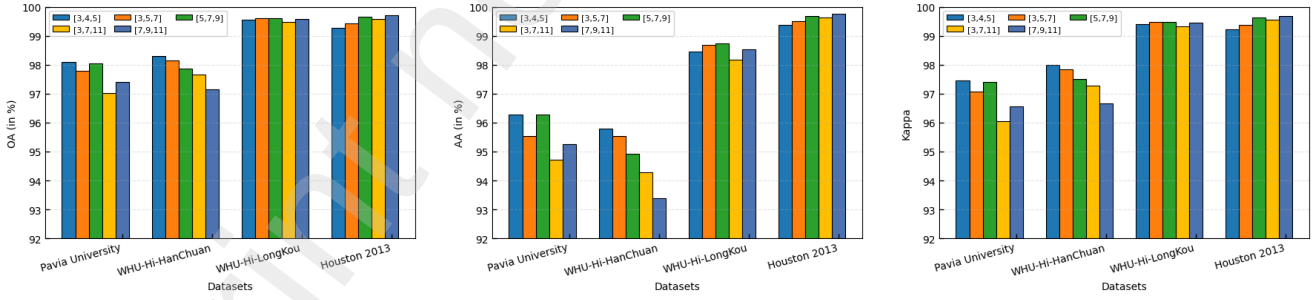


(a) OA (b) AA (c) Kappa

Figure 6: Effects of patch size on OA, AA, and Kappa across four hyperspectral datasets. Lines correspond to different datasets.

Table 8
Kernel-size combination ablation study on four hyperspectral datasets. OA and AA are reported in percentages (%), and Kappa is unitless. For each dataset, the best result in each metric column is highlighted in bold.

Kernel Size ($k \times k \times k$)						Pavia University			WHU-Hi-HanChuan			WHU-Hi-LongKou			Houston 2013		
3	4	5	7	9	11	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
✓	✓	✓	–	–	–	**98.09**	96.27	**97.46**	**98.29**	**95.78**	**98.00**	99.54	98.45	99.40	99.28	99.37	99.22
✓	–	✓	✓	–	–	97.79	95.54	97.07	98.14	95.54	97.83	**99.60**	98.69	99.47	99.43	99.49	99.38
–	–	✓	✓	✓	–	98.04	**96.28**	97.40	97.87	94.93	97.50	**99.60**	**98.73**	**99.48**	99.65	99.69	99.62
✓	–	–	✓	–	✓	97.03	94.72	96.06	97.67	94.29	97.28	99.48	98.17	99.31	99.59	99.62	99.55
–	–	–	✓	✓	✓	97.40	95.26	96.55	97.14	93.39	96.65	99.58	98.53	99.45	**99.70**	**99.76**	**99.67**



(a) OA (b) AA (c) Kappa

Figure 7: Effects of different kernel-size combinations on OA, AA, and Kappa across four hyperspectral datasets.

Table 9
Component ablation study of SDFE, Mamba, and ODE on four hyperspectral datasets. The best results are highlighted in bold.

SDFE	Mamba	ODE	Pavia University			WHU-Hi-HanChuan			WHU-Hi-LongKou			Houston 2013		
			OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
✓	–	–	96.40	93.27	95.21	96.39	91.73	95.77	99.50	98.23	99.34	99.08	99.15	99.01
✓	✓	–	97.40	95.42	96.55	97.55	94.62	97.13	99.51	98.17	99.36	99.44	99.49	99.39
✓	–	✓	97.69	95.70	96.93	97.74	94.22	97.35	99.52	98.38	99.37	99.57	99.66	99.54
✓	✓	✓	**98.04**	**96.28**	**97.40**	**97.87**	**94.93**	**97.50**	**99.60**	**98.73**	**99.48**	**99.65**	**99.69**	**99.62**

global context propagation is handled by Mamba, excessively large receptive fields become less critical. Instead, balanced multi-scale spatial priors combined with continuous-depth feature evolution provide sufficient representational capacity without introducing redundant complexity.

Rather than simply stacking two modeling paradigms, HyperMODE assigns distinct responsibilities to each module. The Neural ODE module is responsible for refining intermediate representations through continuous-depth transformation, whereas the Mamba module emphasizes efficient long-range spatial reasoning and cross-scale context propagation. This role separation improves representational efficiency and contributes to the favorable accuracy–complexity trade-off observed in Table 6.

5.3. Effectiveness of the Interval-Coupled Integration Strategy

A key architectural characteristic of HyperMODE is the interval-coupled integration strategy, where Neural ODE blocks and Mamba blocks are alternately applied across network depth. Experimental evidence from the ablation studies shows that removing either component leads to noticeable performance degradation, indicating that the effectiveness of HyperMODE does not arise from isolated module insertion, but from the coordinated interaction between continuous-depth feature refinement and efficient spatial context propagation.

From a modeling perspective, the alternating design enables progressive refinement: the Neural ODE blocks provide continuous-depth feature transformation, while the Mamba blocks expand spatial context through efficient state-space propagation. Repeated transitions between continuous and discrete modeling help balance local feature consistency and global contextual reasoning, which is particularly advantageous on datasets with uneven class distributions.

Although this staged integration increases computational depth, the efficiency analysis demonstrates that the resulting overhead remains manageable due to the lightweight Euler-based ODE solver and the linear-complexity Mamba backbone. This suggests that iterative continuous–discrete interaction provides a practical alternative to deeper attention-based architectures.

5.4. Efficiency–Accuracy Trade-Off and Design Implications

Beyond accuracy improvements, the results highlight an important efficiency–accuracy trade-off enabled by the hybrid design. Despite incorporating multi-scale branches and repeated ODE–Mamba alternation, HyperMODE maintains moderate parameter counts and computational cost compared with heavier Transformer-based methods.

Dataset-wise analysis indicates that class imbalance and scene complexity exert a strong influence on model behavior. Continuous-depth feature evolution improves robustness under low-sample conditions, while linear-time state-space propagation ensures scalability on large scenes. The patch-size and kernel ablation studies jointly suggest that balanced spatial context is more effective than aggressively increasing receptive fields.

Overall, these observations suggest that the effectiveness of HyperMODE stems from the complementary coupling of continuous-depth feature modeling and efficient discrete spatial propagation, and that similar continuous–discrete designs may offer a promising direction for future hyperspectral image classification research.

6. Conclusion

In this work, we propose HyperMODE, a unified continuous–discrete spectral–spatial modeling framework for hyperspectral image classification that couples Neural ODE-based continuous-depth feature evolution with Mamba-based state-space spatial propagation. By coupling continuous-depth feature evolution with efficient state-space spatial propagation, the proposed framework provides a flexible alternative to purely discrete layer-wise updates while maintaining effective long-range contextual modeling with manageable computational cost.

Extensive experiments on four benchmark datasets demonstrate that the proposed hybrid design achieves strong classification performance and a favorable efficiency–accuracy trade-off under diverse spectral–spatial conditions. From our study, two main observations can be drawn:

1. Continuous-depth feature evolution contributes to more stable and consistent intermediate representations than purely discrete transformations.
2. Linear-time state-space propagation offers an efficient alternative to attention-based architectures for capturing long-range spatial dependencies without incurring quadratic complexity.

These findings suggest that the effectiveness of HyperMODE lies in the coordinated integration of continuous-depth feature modeling and scalable state-space spatial propagation, and that such continuous–discrete hybrid designs provide a promising direction for future hyperspectral image classification research.

References

Ahmad, M., Khan, A. M., Mazzara, M., Distefano, S., Ali, M., & Sarfraz, M. S. (2022). A fast and compact 3-D CNN for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters, 19*, 1–5.

Ben Hamida, A., Benoit, A., Lambert, P., & Ben Amar, C. (2018). 3-D deep learning approach for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing, 56*(8), 4420–4434.

Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing, 54*(10), 6232–6251.

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. (2018). Neural ordinary differential equations. In *Advances in Neural Information Processing Systems* (Vol. 31).

Chen, T., & Ye, Z. (2024). FreqODEs: Frequency neural ODE networks for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1–12.

Ding, S., Liu, Z., Wang, Y., & Plaza, A. (2025). LRDTN: Spectral–spatial convolutional fusion long-range dependence transformer network for

hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 63*, 1–21.

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

He, J., Zhao, L., Yang, H., Zhang, M., & Li, W. (2020). HSI-BERT: Hyperspectral image classification using the bidirectional encoder representation from transformers. *IEEE Transactions on Geoscience and Remote Sensing, 58*(1), 165–178.

He, Y., Tu, B., Liu, B., Chen, Y., Li, J., & Plaza, A. (2024). Hybrid multiscale spatial–spectral transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1–15.

He, Y., Tu, B., Liu, B., Li, J., & Plaza, A. (2024). 3DSS-Mamba: 3D-spectral–spatial Mamba for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1–16.

He, Y., Tu, B., Jiang, P., Liu, B., Li, J., & Plaza, A. (2024). IGroupSS-Mamba: Interval group spatial–spectral Mamba for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1–17.

He, Y., Tu, B., Liu, B., Li, J., & Plaza, A. (2025). HSI-MFormer: Integrating Mamba and Transformer experts for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 63*, 1–16.

Hong, D., Han, Z., Yokoya, N., & Zhu, X. X. (2022). SpectralFormer: Rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–15.

Hu, W.-S., Li, H.-C., Pan, L., Li, W., Tao, R., & Du, Q. (2020). Spatial–spectral feature extraction via deep ConvLSTM neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 58*(6), 4237–4250.

Jiang, M., Su, Y., Gao, L., Plaza, A., Zhao, X.-L., Sun, X., & Liu, G. (2024). GraphGST: Graph generative structure-aware transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1–16.

Li, S., Zhu, X., & Bao, J. (2019). Hierarchical multi-scale convolutional neural networks for hyperspectral image classification. *Sensors, 19*(7), 1714.

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.

Li, D., Bhatti, U. A., Huang, M., Bruzzone, L., & Li, J. (2026). HyPyra-Mamba: A pyramid spectral attention and Mamba-based architecture for robust hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 64*, 1–16.

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.

Lu, B., Dao, P., Liu, J., He, Y., & Shang, J. (2020). Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sensing, 12*(16), 2659.

Ma, X., Wang, H., & Geng, J. (2016). Spectral–spatial classification of hyperspectral image based on deep auto-encoder. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9*(9), 4073–4085.

Mei, S., Song, C., Ma, M., & Xu, F. (2022). Hyperspectral image classification using group-aware hierarchical transformer. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–14.

Metzger, N., Turkoglu, M. O., D'Aronco, S., Wegner, J. D., & Schindler, K. (2022). Crop classification under varying cloud cover with neural ordinary differential equations. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–12.

Mohamed, S., Haghighat, M., Fernando, T., Sridharan, S., Fookes, C., & Moghadam, P. (2024). FactoFormer: Factorized hyperspectral transformers with self-supervised pretraining. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1–14.

Pan, Z., Li, C., Plaza, A., Chanussot, J., & Hong, D. (2025). Hyperspectral image classification with Mamba. *IEEE Transactions on Geoscience and Remote Sensing, 63*, 1–14.

Paoletti, M. E., Haut, J. M., Fernandez-Beltran, R., Plaza, J., Plaza, A. J., & Pla, F. (2019). Deep pyramidal residual networks for spectral–spatial hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 57*(2), 740–754.

Paoletti, M. E., Haut, J. M., Plaza, J., & Plaza, A. (2020). Neural ordinary differential equations for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 58*(3), 1718–1734.

Peng, Y., Zhang, Y., Tu, B., Li, Q., & Li, W. (2022). Spatial–spectral transformer with cross-attention for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–14.

Peyghambari, S., & Zhang, Y. (2021). Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: An updated review. *Journal of Applied Remote Sensing, 15*(3), 031501.

Roy, S. K., Krishna, G., Dubey, S. R., & Chaudhuri, B. B. (2020). HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters, 17*(2), 277–281.

Roy, S. K., Deria, A., Shah, C., Haut, J. M., Du, Q., & Plaza, A. (2023). Spectral–spatial morphological attention transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 61*.

Sun, L., Zhang, H., Zheng, Y., Wu, Z., Ye, Z., & Zhao, H. (2024). MASSFormer: Memory-augmented spectral–spatial transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1–15.

Tang, J., Ma, N., Jia, C., Tian, R., & Guo, Y. (2025). HyperEAST: An enhanced attention-based spectral–spatial transformer with self-supervised pretraining for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18*, 22241–22255.

Tang, X., et al. (2025). SpiralMamba: Spatial–spectral complementary Mamba with spatial spiral scan for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 63*, 1–19.

Tian, R., Liu, D., Bai, Y., Jin, Y., Wan, G., & Guo, Y. (2024). Swin-MSP: A shifted windows masked spectral pretraining model for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1–14.

Wan, G., Liu, D., Liu, Y., Yang, T., & Guo, Y. (2026). S2CaT: Spatial–spectral-based CNN and transformer network for hyperspectral image classification. *IEEE Access, 14*, 17434–17453.

Wang, C., Ma, J., Guo, X., & Zhang, Y. (2021). A review of deep learning used in the hyperspectral image analysis for agriculture. *Artificial Intelligence Review, 54*(7), 5205–5253.

Wang, D., Du, B., & Zhang, L. (2022). Fully contextual network for hyperspectral scene parsing. *IEEE Transactions on Geoscience and Remote Sensing, 60*.

Wang, H., Zhuang, P., Zhang, X., & Li, J. (2025). DBMGNet: A dual-branch Mamba-GCN network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 63*, 1–17.

Wang, G., Zhang, X., Peng, Z., Zhang, T., & Jiao, L. (2025). S2Mamba: A spatial–spectral state space model for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 63*, 1–13.

Xi, B., Zhang, Y., Li, H., & Plaza, A. (2025). MCTGCL: Mixed CNN–transformer for Mars hyperspectral image classification with graph contrastive learning. *IEEE Transactions on Geoscience and Remote Sensing, 63*, 1–14.

Xu, R., Dong, X.-M., Li, W., Peng, J., Sun, W., & Xu, Y. (2024). DBCTNet: Double branch convolution-transformer network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1–15.

Xu, Y., Wang, D., Jiao, H., Zhang, L., & Zhang, L. (2026). MambaMoE: Mixture-of-spectral–spatial-experts state space model for hyperspectral image classification. *Information Fusion, 127*, 102904.

Yang, X., Cao, W., Lu, Y., & Zhou, Y. (2022). Hyperspectral image transformer classification networks. *IEEE Transactions on Geoscience and Remote Sensing, 60*.

Yang, J., Du, B., & Zhang, L. (2023). From center to surrounding: An interactive learning framework for hyperspectral image classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 197*, 145–166.

Zhang, X., Wei, W., Zhang, L., & Ding, C. (2021). Neural stochastic differential equation for hyperspectral image classification. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 2357–2360).

Zhang, J., Meng, Z., Zhao, F., Liu, H., & Chang, Z. (2022). Convolution transformer mixer for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters, 19*, 1–5.

Zhang, X., Song, C., You, T., Bai, Q., Wei, W., & Zhang, L. (2024). Dual ODE: Spatial–spectral neural ordinary differential equations for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1–15.

Zhang, Z., Feng, H., Zhou, G., & Ma, Q. (2026). HyperKANODE: Kolmogorov-Arnold network with neural ordinary differential equations for hyperspectral image classification. *Expert Systems with Applications, 320*.

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, Z., Li, J., Clausi, D. A., & Wong, A. (2020). Generative adversarial networks and conditional random fields for hyperspectral image classification. *IEEE Transactions on Cybernetics, 50*(7), 3318–3329.

Jialin Tang received the B.M. degree in Information Management and Information Systems from Shandong University of Finance and Economics, China, in 2022, and the M.S. degree in Computer Science from California State University, Fullerton, USA, in 2026. He is currently pursuing the Ph.D. degree in Computational Science at the University of California, Irvine, USA.

His research interests include deep learning and image processing.

Yunduan Lou received the B.A. degree in Mathematics from Nazareth University, USA, in 2020, and the M.S. degree in Mathematics from the University of Connecticut, USA, in 2022. He is currently pursuing the Ph.D. degree in Computational Science at the University of California, Irvine, USA.

His research interests include deep learning and model compression.

Yanhui Guo (Member, IEEE) received the B.S. degree in Information Management and Information System from Xi'an University of Finance and Economics, Xi'an, China, in 2006, and the M.S. and Ph.D. degrees in Computer Software and Theory from Shaanxi Normal University, Xi'an, China, in 2009 and 2020, respectively.

Since 2009, he has been with the School of Artificial Intelligence, Shandong Women's University, Jinan, China, where he is currently a professor. His research interests include machine learning and remote sensing image processing.

Yu Bai (Senior Member, IEEE) received the Ph.D. degree in Electrical Engineering from the University of Central Florida, Orlando, FL, USA, in 2016, and the M.S. degree in Electrical and Computer Engineering from the University of Texas–Pan American, Edinburg, TX, USA, in 2011. Prior to his academic career, he was with Siemens Energy Inc.

He is currently a professor and director of the Intelligent Computing Research Lab, College of Engineering and Computer Science, California State University, Fullerton, Fullerton, CA, USA. His research interests include generative AI, AI for science, stochastic computing, neuromorphic computing, FPGA design, nanoscale computing systems with novel silicon and post-silicon devices, and low-power digital and mixed-signal CMOS circuit design.