MambaHSI+: Multidirectional State Propagation for Efficient Hyperspectral Image Classification

Yunbiao Wang, *Member, IEEE*, Lupeng Liu, Jun Xiao, Dongbo Yu, Ye Tao, and Wenniu Zhang

Abstract— **Hyperspectral image (HSI) classification faces significant challenges due to high-dimensional spectral redundancy and complex spatial–spectral dependencies. While existing MambaHSI models leverage state-space modeling to enhance representation learning, their unidirectional formulation fails to fully capture bidirectional spatial interactions and cross-band contextual dependencies. Moreover, the uniform projection mechanism struggles to effectively distinguish spectral variations across different wavelengths. To address these limitations, we propose MambaHSI+, a novel framework that integrates bidirectional state-space modeling with spectral trajectory learning (STL). The proposed architecture introduces three key innovations: 1) a bidirectional context modeling module, enhanced by reverse-order modeling, enables multidirectional spatial information propagation through recursive state transitions, facilitating comprehensive local–global feature aggregation while maintaining linear computational efficiency; 2) an STL paradigm that formulates spectral evolution as continuous state-space process with bidirectional propagation, effectively encoding cross-band relationships; and 3) a Mamba-enhanced channel attention mechanism that adaptively emphasizes discriminative spectral features via selective state-space transformations. Extensive experiments on four benchmark datasets demonstrate state-of-the-art (SOTA) performance, achieving an average accuracy (AA) improvement of 2.51% over MambaHSI. By integrating state-space systems with advanced Mamba mechanisms, MambaHSI+ establishes a new paradigm for spectral–spatial representation learning, significantly advancing classification accuracy while ensuring computational efficiency. The latest logs are now available at https://github.com/RockAilab/MambaHSI_Plus**

Index Terms— **Bidirectional propagation, hyperspectral image (HSI) classification, mamba architecture, spectral trajectory learning (STL), state-space models (SSMs).**

I. INTRODUCTION

HYPERSPECTRAL image (HSI) classification has attracted significant attention due to its broad applications in remote sensing, environmental monitoring, agriculture, and medical imaging. HSI provides rich spectral information, with each pixel being represented by a high-dimensional vector across multiple spectral bands [1], [2].
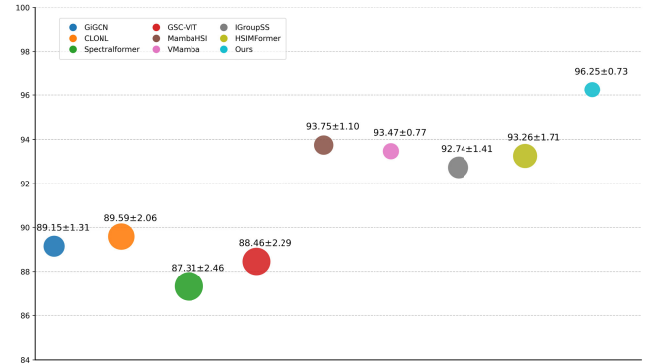
Fig. 1. Comparison of the mean OA metric across different methods on four datasets. Each circle represents a method, with the vertical position indicating the mean OA value and the circle size proportional to the STD (smaller circles indicate lower STD and higher stability). Our method MambaHSI+ achieves the highest mean OA value with the smallest STD, demonstrating superior accuracy and stability across the datasets.

The ability to capture detailed spectral features allows HSI to effectively differentiate various materials or objects in the scene, making it an indispensable tool for many real-world applications [3], [4]. However, the high-dimensional nature of HSI data presents several challenges, including high spectral redundancy, and complex spatial–spectral dependencies, which make effective classification a nontrivial task.

HSI classification has undergone transformative advancements through the integration of deep learning, particularly in addressing intricate spectral–spatial feature representation challenges. Current methodologies predominantly fall into two categories: traditional machine learning [5], [6] and deep learning approaches. The latter encompasses two dominant architectures: convolutional and graph neural network-based models [7], [8], [9], [10], [11], [12], [13], [14] and Transformer-based models [15], [16], [17], [18], [19]. While CNN-based methods excel in local spatial feature extraction via convolutional operations, their limited receptive fields inherently restrict their capacity to model long-range dependencies. In contrast, Transformer-based frameworks leverage self-attention mechanisms to achieve global contextual modeling, albeit at the expense of quadratic computational complexity relative to input sequence length. Beyond architectural designs, insights from related remote sensing tasks have also enriched spectral–spatial modeling strategies. In particular, spectral–spatial fusion strategies developed in multispectral image pansharpening [20], [21] emphasize the importance of prior modeling and degradation simulation,

offering valuable guidance for enhancing hyperspectral feature representation.

Recent hybrid architectures combining CNNs and Transformers [22], [23], [24], [25], [26] aim to synergize local feature extraction with global dependency modeling. While these frameworks improve spectral–spatial representation relative to standalone models, they inherit inherent limitations.

The emergence of state-space models (SSMs), notably structured state-space sequence (S4) models [27], [28], [29], [30], [31], [32], has introduced novel paradigms for efficient long-sequence analysis. Mamba [27], an enhanced SSM variant, further incorporates a selective mechanism that dynamically filters input-dependent information while preserving linear computational complexity. These advancements have demonstrated exceptional efficacy across diverse domains, including natural language processing [27], [32] and computer vision [29], [30], [33], [34], [35]. Notably, Mamba's hardware-optimized implementation achieves superior computational efficiency compared to Transformer architectures, rendering it particularly suitable for high-dimensional data processing. However, existing adaptations of Mamba for vision tasks [27], [30], [34], [35] while enhancing spatial context modeling through multidirectional scanning strategies exhibit suboptimal performance in the HSI classification. Their window-based partitioning mechanisms hinder precise pixel-wise attention required for dense prediction tasks and simultaneously neglect spectral continuity—a critical property of hyperspectral data characterized by gradual spectral transitions across adjacent bands. Recently, targeted adaptations such as IGroupSS-Mamba and HSI-MFormer [36], [37] have been proposed to enhance Mamba's performance for HSI by designing interval grouping strategies for spectral–spatial modeling. However, challenges remain in fully capturing spectral continuity and achieving fine-grained dense prediction. While HSI-MFormer [37] effectively combines multiscale spatial–spectral modeling through its hybrid Mamba-Transformer architecture, it still faces challenges in preserving fine-grained spectral continuity, achieving seamless dense prediction, and balancing computational efficiency with model complexity. Furthermore, HSIs typically contain hundreds of spectral bands, yet current methods uniformly treat all bands with equal priority [38], failing to adequately prioritize discriminative spectral bands.

In this article, we propose a novel framework to address the challenges of HSI classification. The proposed framework consists of three key components. First, we introduce a pixel-level bidirectional spatial modeling mechanism that leverages state-space propagation to capture bidirectional spatial dependencies. This allows the model to encode rich local context information for each pixel. To further enhance the spatial feature representation, we incorporate a reverse-order sequence augmentation strategy, which improves the consistency of spatial features and better preserves global context. Second, we propose spectral trajectory learning (STL), where the spectral dimension is modeled as a dynamic state trajectory. By employing bidirectional propagation along the spectral axis, our approach effectively captures interband correlations and reduces the impact of spectral noise. This modeling strategy ensures that each spectral band's final representation is a unique function of all input

bands, preserving the intrinsic spectral relationships. Finally, to enhance the discriminative power of the spatial and spectral features, we incorporate a mamba-enhanced channel attention mechanism. This attention module dynamically adjusts the importance of different spectral bands, allowing the model to focus on the most relevant features while suppressing redundant information.

The contribution of this article can mainly be summarized in the following points.

1) A spatial contextualization framework that synergizes local feature extraction with global context aggregation through recursive bidirectional propagation, addressing the limited receptive field problem in pixel-wise classification.
2) We present STL, which models the spectral dimension as a dynamic state trajectory to capture interband correlations.
3) An efficient channel attention mechanism using SSMs to achieve adaptive feature recalibration with linear complexity.
4) Extensive experimental validation across multiple HSI benchmarks, as shown in Fig. 1, demonstrating consistent performance improvements over state-of-the-art (SOTA) approaches.

The remainder of this article is structured as follows. In Section II, we review related work on HSI classification, focusing on methods for spatial and spectral modeling. In Section III, we detail the proposed framework and its key components. Section IV presents the experimental results and analyses, including ablation studies and comparisons with SOTA methods. Finally, Section V concludes this article and discusses future research directions.

II. RELATED WORK

A. HSI Classification

Deep learning has profoundly advanced HSI classification through three dominant architectural paradigms, each exhibiting distinct strengths and limitations in spectral–spatial relationship modeling. This section systematically analyzes their evolutionary trajectories and technical characteristics.

1) Local and Sequential Feature Extractors: This category encompasses CNN- and RNN-based approaches emphasizing localized pattern recognition. CNN architectures extract hierarchical features through 1-D spectral convolutions [39], 3-D spectral–spatial kernels [40], and multiscale image patches [41]. Enhanced variants integrate attention mechanisms [42] and spectral–spatial fully convolutional network (SSFCN) [43]. RNN-based methods model spectral sequences via cascaded architectures [44] and bidirectional LSTM networks [45], augmented by spatial sequencing mechanisms [46] and multiscanning strategies [47]. While these models efficiently capture local features, they struggle with limited receptive fields and inadequate long-range dependency modeling, restricting their ability to encode global spatial–spectral interactions [29].

2) Global Dependency Modeling Architectures: GCNs and Transformers address long-range dependencies through graph structures and self-attention mechanisms. GCN variants

include second-order neighborhood modeling [48], superpixel-enhanced graphs (Wan et al. [49], GiGCN [50]), and dynamic multiscale graph fusion [13]. Transformers evolve from spectral sequence modeling (SpectralFormer [16]) to spatial–spectral joint architectures [19], [26], [51], with advancements in multiscale interaction (CSIL [52]), lightweight design (SPFormer [53]), and local–global representation enhancement (GSC-ViT [19] and SPRLT-Net [54]). Both architectures face scalability challenges: GCNs incur high graph construction costs [55], while Transformers exhibit quadratic complexity in self-attention [17].

3) Hybrid and Cross-Modal Fusion Frameworks: Synergistic architectures combine CNNs/Transformers for complementary local–global modeling. Representative hybrids include Gaussian-weighted token Transformers [26], grouped pixel embedding modules (GAHT [25]), and dual-context encoding networks (DCN-T [56]). Cross-modal frameworks enhance few-shot learning via multihead encoders (DKDFN [57]) and locality-preserved embeddings [58]. Emerging innovations feature large-scale pretraining (HSIMAE [59]) and online spectral compensation (OSICN [60]).

Despite these advancements, existing models still face challenges in simultaneously capturing local–global spectral–spatial relationships while maintaining computational efficiency. CNNs fail to provide both a large receptive field and a lightweight structure, RNNs struggle with long-term dependencies, and Transformers suffer from high computational complexity. Hybrid CNN–Transformer models often overlook the intrinsic relationships between local modeling and global modeling. To address these challenges, we explore a parallel Mamba-convolution architecture, which offers low time-memory complexity and superior spectral–spatial feature extraction capabilities.

B. State-Space Models

SSMs have recently gained considerable attention due to their ability to model long-range dependencies while offering computational advantages over traditional models like CNNs and Transformers.

The S4 [32] model has demonstrated the ability to model long-range dependencies with linear complexity, making it an attractive alternative for sequence-based tasks. Extensions to S4, such as the S5 layer [61], have introduced more efficient methods for parallel processing. Additionally, Mamba [27] has shown superior performance to Transformers on large-scale real-world data, maintaining linear scaling in sequence length. Recent works [62] have further explored the potential of SSMs in improving sequence modeling, filling the performance gap between SSMs and Transformers in tasks like language modeling [63].

SSMs have also found applications in visual tasks, including image and video processing. Nguyen et al. [64] extended 1-D S4 to handle 2-D images and 3-D videos, while TranS4mer [65] combined the strengths of S4 with self-attention to achieve the SOTA performance in movie scene detection. Wang et al. [66] introduced a selectivity mechanism in S4 to enhance its performance on long-form video understanding with reduced memory footprint. The integration

of Mamba with U-shaped architectures for biomedical image segmentation [67] has further showcased the flexibility of SSMs in visual applications. To further enhance spatial feature extraction, Zhu et al. [29] introduced the Vision Mamba architecture, which employs bidirectional scanning mechanisms to convert images into ordered visual sequences. Building upon this, Liu et al. [28] proposed VMamba, incorporating a cross-selective scan technique that traverses images both horizontally and vertically. Zhao et al. [30] further advanced this line of research by developing an omnidirectional selective scan mechanism to capture extensive spatial features from multiple directions.

Recent studies [38], [68], [69], [70], [71] have adopted Mamba for HSI classification. SpectralMamba [68] introduced a PSS and GSSM module to capture sequential learning in the state domain and correct spectral distortions. He et al. [69] applied Mamba in the context of 3-D hyperspectral blocks, aiming to address the challenges associated with 3-D spectral data. Despite these advancements, Mamba-based models in HSI classification still exhibit several limitations. Existing methods primarily rely on window-based partitioning mechanisms [28], [29], which hinder precise pixel-wise attention required for dense prediction tasks and simultaneously neglect spectral continuity—a critical property of hyperspectral data characterized by gradual spectral transitions across adjacent bands. Furthermore, HSIs typically contain hundreds of spectral bands, yet current methods [70] uniformly treat all bands with equal priority, failing to adequately prioritize discriminative spectral bands. Finally, the sequential unfolding of HSIs [38] may disrupt their inherent spatial structures, leading to suboptimal feature extraction. These limitations highlight the need for improved methods that effectively balance spatial and spectral feature extraction while maintaining computational efficiency.

To address these challenges, we propose a novel approach that integrates bidirectional spatial modeling and STL to enhance HSI classification. Specifically, our method recursively propagates spatial information in four directions to preserve both local and global spatial context at the pixel level. Additionally, we model the spectral dimension as a dynamic state trajectory to effectively capture interband correlations. To further refine feature representation, we incorporate a lightweight channel attention mechanism that enhances the discriminative capability of fused spatial and spectral features. This design ensures balanced spatial–spectral extraction while maintaining computational efficiency and scalability.

III. METHODOLOGY

A. Motivation and Theoretical Framework

HSI classification presents a unique challenge as a pixel-wise discriminative task, requiring precise modeling of subtle spectral–spatial variations across high-dimensional data. While existing approaches acknowledge the significance of long-range dependency modeling, conventional CNN architectures are intrinsically limited by their local receptive fields, inherently restricting their capacity to capture global contextual patterns, and the long-range modeling of single sequences compromises the spatial integrity of the original image.
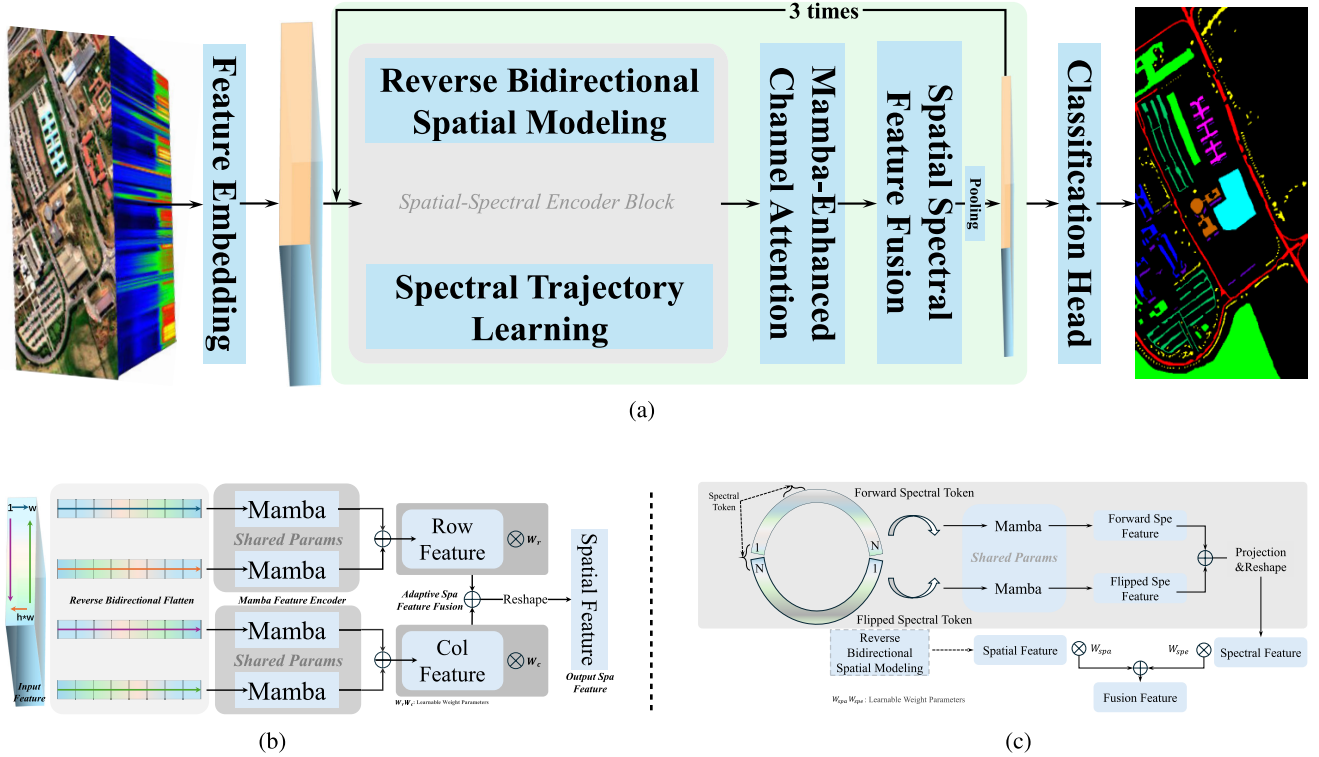
(a)

(b)

(c)

Fig. 2. MambaHSI+ framework for HSI classification, integrating bidirectional state-space modeling Section III-C and STL Section III-D. (a) Illustration of the proposed MambaHSI+ framework. The architecture consists of iterative feedback loops, where fused features undergo three successive rounds of spatial–spectral encoding before final classification. (b) Reverse bidirectional spatial modeling. (c) STL.

Transformer-based alternatives, though theoretically capable of modeling global interactions, suffer from prohibitive quadratic computational complexity ($O(T^2)$) in self-attention mechanisms. This complexity bottleneck forces practical implementations to adopt coarse patch-level tokenization, thereby sacrificing the granularity required for pixel-level discrimination.

Our theoretical breakthrough stems from reinterpreting HSI classification through the lens of state-space systems, which inherently balance global context modeling with linear computational scaling ($O(T)$). The proposed framework is grounded in two fundamental properties of hyperspectral data as follows.

1) *Noncausal Spatial–Spectral Correlations*: Pixel-wise dependencies extend bidirectionally across spatial and spectral dimensions, necessitating a hybrid state propagation mechanism

$$\mathbf{h}_t = \underbrace{\mathbf{A}^{\rightarrow}\mathbf{h}_{t-1}}_{\text{Forward Context}} + \underbrace{\mathbf{A}^{\leftarrow}\mathbf{h}_{t+1}}_{\text{Reverse Context}} + \mathbf{B}\mathbf{x}_t. \quad (1)$$

After the first stage, each input $\mathbf{x}$ results from a prior spatial–spectral feature fusion, embedding joint spatial and spectral cues [Fig. 2(a)]. Consequently, $\mathbf{x}$ and $\mathbf{h}$ lie in a unified spatial–spectral feature space.

2) *Spectral Continuity Prior*: The smooth spectral profiles enable state trajectory learning across bands

$$\mathbf{s}_\lambda = \mathbf{A}_\lambda \mathbf{s}_{\lambda-1} + \mathbf{B}_\lambda \mathbf{x}_\lambda, \forall \lambda \in \{1, \dots, C\} \quad (2)$$

with spectral transition matrices $\mathbf{A}_\lambda$ adaptively conditioned on input characteristics.

B. Architecture Synthesis

As illustrated in Fig. 2, our framework synthesizes three interconnected theoretical components.

1) *Bidirectional State-Space Modeling*: We develop a dual-path spatial modeling architecture where hidden states propagate in both raster-scan and reverse directions. This bidirectional propagation captures global spatial contexts while maintaining linear complexity through recursive state updates.

2) *Spectral Trajectory Learning*: The spectral dimension is modeled as a continuous state trajectory, where each band's hidden state $\mathbf{s}_\lambda$ evolves through learned dynamics matrices $\mathbf{A}_\lambda$, preserving interband correlations while filtering spectral noise.

3) *Adaptive Feature Fusion*: Spatial and spectral representations are fused through attention-guided recalibration

$$\mathbf{F}_{\text{fusion}} = \text{LayerNorm}\big(\alpha \odot \mathbf{F}_{\text{spa}} + (1 - \alpha) \odot \mathbf{F}_{\text{spe}}\big) \quad (3)$$

with attention weights α dynamically computed from cross-modality interactions. This theoretical formulation provides provable advantages in both discriminative power and computational efficiency, as demonstrated in Section IV.

C. Reverse Bidirectional Contextual Modeling

We propose a reverse bidirectional contextual modeling framework [Fig. 2(b)] for HSI classification, combining dual-directional propagation, reverse sequencing, and adaptive fusion. By restructuring input tensors bidirectionally, the model encodes horizontal and vertical spatial dependencies into pixel-level features while preserving spatial topology. The

reverse-order augmentation enhances neighborhood interactions without sacrificing spatial granularity, overcoming local context limitations of conventional methods.

Let $\mathbf{X} \in \mathbb{R}^{B \times C \times H \times W}$ denote the input tensor and $\mathbf{x}_{i,j} \in \mathbb{R}^C$ denote the feature at spatial location (i, j). To capture bidirectional spatial dependencies, we perform state-space propagation along rows and columns. In particular, the row-wise propagation is defined by

$$\mathbf{h}_{i,j}^{(r,\rightarrow)} = \mathbf{A}_r \, \mathbf{h}_{i,j-1}^{(r,\rightarrow)} + \mathbf{B}_r \, \mathbf{x}_{i,j}, \, j = 1, \ldots, W$$
$$[1mm]\mathbf{h}_{i,j}^{(r,\leftarrow)} = \mathbf{A}_r \, \mathbf{h}_{i,j+1}^{(r,\leftarrow)} + \mathbf{B}_r \, \mathbf{x}_{i,j}, \, j = W, \ldots, 1 \quad (4)$$

and these are fused via attention-based weighted summation

$$\mathbf{h}_{i,j}^{(r)} = \sum_{j'=1}^{W} \alpha_{i,j,j'}^{(r)} \left[\mathbf{h}_{i,j'}^{(r,\rightarrow)} + \mathbf{h}_{i,j'}^{(r,\leftarrow)} \right] \quad (5)$$

where

$$\alpha_{i,j,j'}^{(r)} = \text{softmax}\left(\mathbf{W}_r \left[\mathbf{h}_{i,j}^{(r,\rightarrow)}; \, \mathbf{h}_{i,j}^{(r,\leftarrow)} \right] \right). \quad (6)$$

Similarly, the column-wise propagation is given by

$$\mathbf{h}_{i,j}^{(c,\downarrow)} = \mathbf{A}_c \, \mathbf{h}_{i-1,j}^{(c,\downarrow)} + \mathbf{B}_c \, \mathbf{x}_{i,j}, \, i = 1, \ldots, H$$
$$[1mm]\mathbf{h}_{i,j}^{(c,\uparrow)} = \mathbf{A}_c \, \mathbf{h}_{i+1,j}^{(c,\uparrow)} + \mathbf{B}_c \, \mathbf{x}_{i,j}, \, i = H, \ldots, 1 \quad (7)$$

with its fusion expressed as

$$\mathbf{h}_{i,j}^{(c)} = \sum_{i'=1}^{H} \alpha_{i,j,i'}^{(c)} \left[\mathbf{h}_{i',j}^{(c,\downarrow)} + \mathbf{h}_{i',j}^{(c,\uparrow)} \right] \quad (8)$$

where

$$\alpha_{i,j,i'}^{(c)} = \text{softmax}\left(\mathbf{W}_c \left[\mathbf{h}_{i,j}^{(c,\downarrow)}; \, \mathbf{h}_{i,j}^{(c,\uparrow)} \right] \right). \quad (9)$$

The above propagation mechanisms are designed such that the forward and reverse directions recursively aggregate all preceding and succeeding features along a row (or column), ensuring that $\mathbf{h}_{i,j}^{(r)}$ and $\mathbf{h}_{i,j}^{(c)}$ are injective functions of all input features in the corresponding row and column, respectively. In other words, the row-wise features encode $\{\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,W}\}$ and the column-wise features encode $\{\mathbf{x}_{1,j}, \ldots, \mathbf{x}_{H,j}\}$, thereby preserving complete spatial context.

After performing reverse-order sequence augmentation (e.g., by flipping the sequence order prior to propagation and then re-flipping the outputs), the processed sequences are reshaped back to the original 2-D layout, i.e., $\mathbb{R}^{B \times C \times H \times W}$. Finally, the adaptive fusion of row and column features is performed as follows:

$$\mathbf{h}_{i,j} = \gamma_{i,j} \odot \mathbf{h}_{i,j}^{(r)} + \left(1 - \gamma_{i,j}\right) \odot \mathbf{h}_{i,j}^{(c)} \quad (10)$$

where

$$\gamma_{i,j} = \text{softmax}\left(\mathbf{W}_\gamma \left[\mathbf{h}_{i,j}^{(r)}; \, \mathbf{h}_{i,j}^{(c)} \right] \right). \quad (11)$$

The adaptive weight $\gamma_{i,j}$, computed from the concatenated row and column features via a softmax function, uniquely encodes the spatial configuration, ensuring that the final feature $\mathbf{h}_{i,j}$ of each pixel is an injective function of its entire row and column context. Consequently, the proposed removal of bidirectional spatial modeling (RBCM) framework not only aggregates sequential contextual information but also guarantees the unique preservation of local spatial relationships,

thus conferring comprehensive perceptive capability in all directions.

In summary, by integrating dual-directional state propagation, reverse-order sequence augmentation, and attention-based fusion into a unified framework, our method provides a rigorous mechanism to capture global and local spatial dependencies while uniquely encoding the 2-D structure of the input.

D. Spectral Trajectory Learning

To complement spatial modeling, we model the spectral dimension as a continuous state trajectory that aggregates information from all spectral bands [Fig. 2(c)]. Let $\mathbf{x}_\lambda \in \mathbb{R}^d$ denote the input feature for the λth spectral band, where $\lambda = 1, \ldots, C$. We define two state sequences along the spectral dimension: a forward state and a flipped state.

Compared to the unidirectional spectral propagation design in MambaHSI, we introduce a bidirectional spectral state propagation mechanism, ensuring that the final representation of each band comprehensively incorporates information from all spectral bands.

The forward spectral state is recursively defined as

$$\mathbf{s}_\lambda^{(\rightarrow)} = \mathbf{A}_\lambda^{(\rightarrow)} \, \mathbf{s}_{\lambda-1}^{(\rightarrow)} + \mathbf{B}_\lambda^{(\rightarrow)} \, \mathbf{x}_\lambda, \, \lambda = 1, \ldots, C \quad (12)$$

and the flipped spectral state is given by

$$\mathbf{s}_\lambda^{(\leftarrow)} = \mathbf{A}_\lambda^{(\leftarrow)} \, \mathbf{s}_{\lambda+1}^{(\leftarrow)} + \mathbf{B}_\lambda^{(\leftarrow)} \, \mathbf{x}_\lambda, \, \lambda = C, \ldots, 1 \quad (13)$$

where $\mathbf{A}_\lambda^{(\rightarrow)}$ and $\mathbf{A}_\lambda^{(\leftarrow)}$ are learnable dynamics matrices that control the evolution of the spectral states, while $\mathbf{B}_\lambda^{(\rightarrow)}$ and $\mathbf{B}_\lambda^{(\leftarrow)}$ weight the current input $\mathbf{x}_\lambda$.

To enhance the flexibility and representation capability of the dynamics matrices $\mathbf{A}_\lambda$, we further introduce a lightweight dynamic projection mechanism. Concretely, we parameterize the matrices $\mathbf{A}_\lambda$ as follows. First, we initialize $\mathbf{A}$ using a log-spaced setting and then apply a learnable linear projection to generate $\mathbf{A}_\lambda$ dynamically. Mathematically, this process can be described as

$$\mathbf{A}_\lambda = -\exp\left(\mathbf{W}_{\text{proj}} \cdot \log(\mathbf{A}_0) \right) + \mathbf{A}_0 \quad (14)$$

where $\mathbf{A}_0$ is the initial matrix with logarithmically spaced elements and $\mathbf{W}_{\text{proj}}$ is a learnable linear projection matrix. This design allows the model to adaptively adjust the spectral transition dynamics for different datasets and tasks, thereby improving the spectral modeling performance. Moreover, we connect the last and the first spectral bands, enabling the model to learn richer spectral features across the entire wavelength range [Fig. 2(c)].

To integrate these two trajectories and fully capture interband correlations, the final spectral feature at band λ is defined as a weighted sum over contributions from all bands

$$\mathbf{s}_\lambda = \sum_{k=1}^{C} \beta_{\lambda,k} \left[\mathbf{s}_k^{(\rightarrow)} + \mathbf{s}_k^{(\leftarrow)} \right] \quad (15)$$

where the attention weights are computed by

$$\beta_{\lambda,k} = \text{softmax}\left(\mathbf{W}_s \left[\mathbf{s}_\lambda^{(\rightarrow)}; \, \mathbf{s}_\lambda^{(\leftarrow)} \right] \right)_k \quad (16)$$

where $\mathbf{W}_s$ is a learnable parameter matrix and the notation $[\cdot ; \cdot]$ denotes the vector concatenation. The softmax operation normalizes the weights over k, ensuring that the fusion is an injective mapping of the entire spectral context.

This formulation guarantees that each spectral band's final representation $\mathbf{s}_\lambda$ is a unique function of all input features $\{\mathbf{x}_k\}_{k=1}^{C}$, thereby preserving intrinsic interband correlations while effectively attenuating spectral noise. In summary, by modeling the spectral dimension as a bidirectional state-space process and fusing the forward and reverse trajectories via attention-based weighting, our framework robustly captures the fine-grained spectral dynamics essential for HSI analysis.

E. Mamba-Enhanced Channel Attention Module

To enhance the discriminative capability of hyperspectral feature representations, we propose a novel channel attention mechanism that integrates SSMs for dynamic spectral band interaction. Given an input feature tensor $X \in \mathbb{R}^{B \times C \times H \times W}$, where B, C, and $H \times W$ denote the batch size, spectral channels, and spatial dimensions, respectively, the proposed module first aggregates global spatial information through adaptive average pooling

$$F_g = \text{AdaptiveAvgPool2d}(X) \in \mathbb{R}^{B \times C \times 1 \times 1}. \quad (17)$$

To facilitate spectral sequence modeling, the pooled features are reshaped via permutation and squeezing operations

$$F_t = \text{Permute}\big(\text{Squeeze}(F_g)\big) \in \mathbb{R}^{B \times 1 \times C}. \quad (18)$$

The core innovation lies in replacing conventional 1-D convolutions with a Mamba-based state-space model, which captures long-range spectral dependencies through hidden state evolution

$$h'(t) = Ah(t) + Bx(t) \quad (19)$$

where $A \in \mathbb{R}^{N \times N}$ governs state transitions, while B and C are projection matrices. To enhance parameter efficiency, we adopt an expansion strategy with expand $= 2$, enabling dynamic information flow modulation based on spectral channel significance

$$F_m = \text{SSM}_\theta(F_t) + \text{Conv1D}(F_t) \quad (20)$$

where SSM_θ denotes the learnable state-space projection. The resulting features undergo activation to produce channel-wise attention weights

$$W_c = \sigma(\text{Unsqueeze}(\text{Permute}(F_m))) \in \mathbb{R}^{B \times C \times 1 \times 1}. \quad (21)$$

Finally, the attention weights are applied to the input through element-wise multiplication

$$X' = X \odot W_c. \quad (22)$$

By leveraging tensor permutation operations, the proposed module aligns spectral channel dimensions with sequence modeling requirements while preserving spatial–spectral topology, achieving an efficient $\mathcal{O}(C)$ complexity for C spectral bands.

IV. EXPERIMENTS

In this section, we present a detailed experimental evaluation of the proposed method. We first introduce the experimental setup, including dataset descriptions, performance metrics, baseline comparisons, and implementation details. Subsequently, we conduct a comprehensive qualitative and quantitative comparison with SOTA methods. Finally, an extensive ablation study is performed to analyze the contribution of each module.

A. Experimental Setup

1) Datasets: To thoroughly assess the effectiveness of the proposed approach, we conduct experiments on four widely adopted and diverse hyperspectral datasets, namely Pavia University (PaviaU), WHU-Hi-HongHu (HongHu) [72], WHU-Hi-HanChuan (HanChuan) [72], and Houston. The detailed description of each dataset is shown in Fig. 3 and Table I.

a) Pavia University: The Pavia University dataset was collected using the Reflective Optics System Imaging Spectrometer (ROSIS) over the Pavia University campus. It consists of 103 spectral bands with a spatial resolution of 610×340. The dataset contains 42 776 labeled pixels spanning nine distinct land-cover classes. Fig. 3(a) provides an overview of the dataset, including the original HSI, ground-truth annotations, and the data split configuration.

b) HongHu: The WHU-Hi-HongHu dataset was collected on November 20, 2017, between 16:23 and 17:37 in Honghu, Hubei, China. The imagery was captured using a 17-mm-focal-length Headwall Nano-Hyperspec sensor mounted on a DJI Matrice 600 Pro UAV platform. The study area primarily consists of complex agricultural landscapes, where various crop types and cultivars of the same crop are distributed across the region. The dataset consists of 940×475 pixels and contains 270 spectral bands covering the 400–1000-nm wavelength range. Fig. 3(b) provides a visualization of the dataset.

c) HanChuan: The WHU-Hi-HanChuan dataset was acquired on June 17, 2016, between 17:57 and 18:46 in Hanchuan, Hubei, China. Data collection was performed using a 17-mm-focal length Headwall Nano-Hyperspec imaging sensor mounted on a Leica Aibot X6 UAV V1 platform. The study area represents a peri-urban region containing buildings, water bodies, and cultivated lands. It includes seven crop species: strawberry, cowpea, soybean, sorghum, water spinach, watermelon, and greens. The image dimensions are 1217×303 pixels, spanning 274 spectral bands within the 400–1000-nm range. An overview of this dataset is shown in Fig. 3(c).

d) Houston: The Houston dataset was captured over the University of Houston and its neighboring areas using the ITRES CASI 1500 hyperspectral imager. The image comprises 144 spectral bands with a spatial dimension of 349×1905. This dataset was officially released as part of the 2013 IEEE Geoscience and Remote Sensing Society (GRSS) Data Fusion Contest and contains 16 land-cover categories, including residential, commercial, and vegetation classes. Additional dataset details are illustrated in Fig. 3(d).

2) Evaluation Metrics: To ensure a fair and rigorous assessment, we adopt three widely used performance metrics
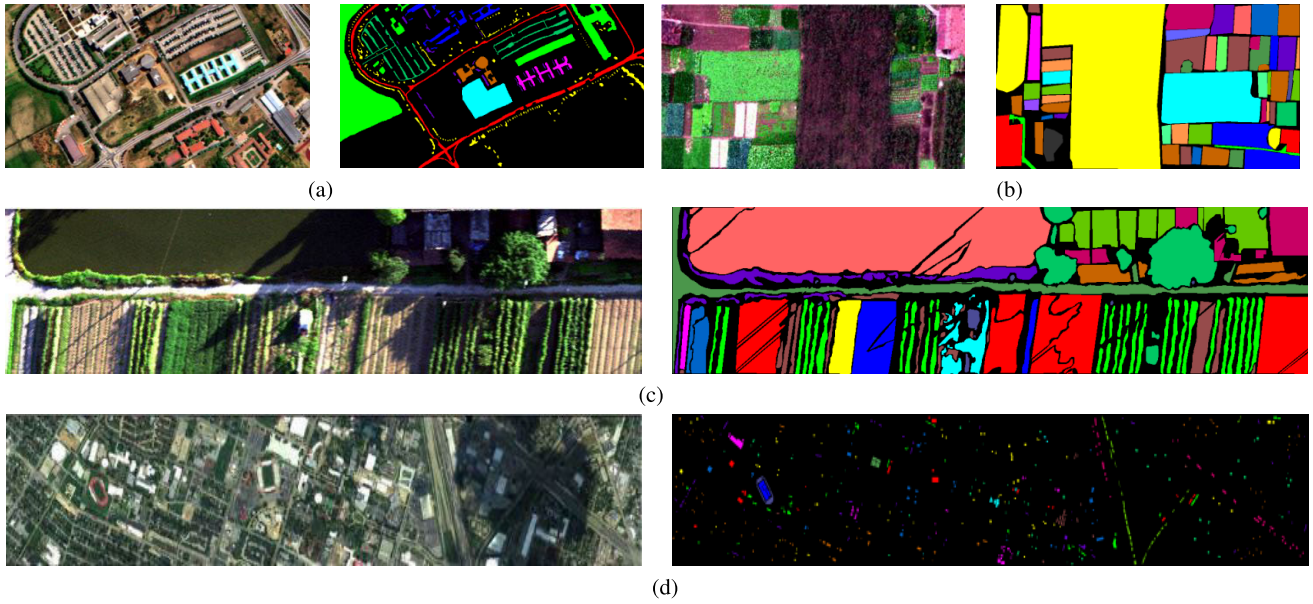
Fig. 3. Experimental HSIs: false-color images and ground truth across four regions. (a) Pavia University. (b) Hong Hu. (c) Han Chuan. (d) Houston.

TABLE I
DATASET DISTRIBUTION BY CATEGORY FOR PAVIA UNIVERSITY, HONG HU, HAN CHUAN, AND HOUSTON

Pavia University			Hong Hu			Han Chuan			Houston		
Color	Category	Total	Color	Category	Total	Color	Category	Total	Color	Category	Total
	Asphalt	6631		Red roof	14041		Strawberry	44735		Healthy Grass	1251
	Meadows	18649		Road	3512		Cowpea	22753		Stressed Grass	1254
	Gravel	2099		Bare soil	21821		Soybean	10287		Synthetic Grass	697
	Trees	3064		Cotton	163285		Sorghum	5353		Tree	1244
	Metal sheets	1345		Cotton firewood	6218		Water spinach	1200		Soil	1242
	Bare soil	5029		Rape	44557		Watermelon	4533		Water	325
	Bitumen	1330		Chinese cabbage	24103		Greens	5903		Residential	1268
	Bricks	3682		Pakchoi	4054		Trees	17978		Commercial	1244
	Shadows	947		Cabbage	10819		Grass	9469		Road	1252
				Tuber mustard	12394		Red roof	10516		Highway	1227
				Brassica parachinensis	11015		Gray roof	16911		Railway	1235
				Brassica chinensis	8954		Plastic	3679		Parking Lot 1	1233
				Small Brassica chinensis	22507		Bare soil	9116		Parking Lot 2	469
				Lactuce sativa	7356		Road	18560		Tennis Court	428
				Celtuce	1002		Bright object	1136		Running Track	660
				Film covered lettuce	7262		Water	75401			
				Romaine lettuce	3010						
				Carrot	3217						
				White radish	8712						
				Garlic sprout	3486						
				Broad bean	1328						
				Tree	4040						
Total		**42776**	**Total**		**386693**	**Total**		**257530**	**Total**		**15029**

in HSI classification: overall accuracy (OA), average accuracy (AA), and Kappa coefficient (κ). These metrics provide comprehensive insights into classification performance. Given that training sample selection may introduce variability, all primary comparison and ablation studies are repeated ten times, whereas additional analytical experiments are conducted five times. Higher mean values and lower standard deviations (STDs) in these metrics indicate superior model performance.

3) Implementation Details: The proposed method is implemented using PyTorch. To mitigate the impact of random sample selection and parameter initialization, all experiments are conducted ten times with different random seeds. In each trial, the training and validation sets are constructed by randomly selecting 30 and 10 samples per class from the dataset, respectively, while the remaining samples are used

as the test set. Since the proposed method processes the entire HSI in a single forward pass and generates predictions for all pixels simultaneously, the batch size is set to 1. The Adam optimizer is employed to optimize the model parameters, with an initial learning rate of 3×10^{-4}. The number of groups G and the hidden dimension D are set to 4 and 128, respectively. All experiments are conducted on the same computational platform, which consists of an NVIDIA A100 GPU, a ×64 Intel Xeon Gold 6348 CPU, and 256 GB of RAM.

B. Comparison With SOTA Methods

To comprehensively evaluate the effectiveness of the proposed method, we compare it against multiple SOTA HSI classification approaches from both quantitative and qualitative perspectives. The selected baselines cover various

TABLE II

QUANTITATIVE RESULT (ACC% ± STD%) OF PAVIA UNIVERSITY. BEST IN BOLD AND SECOND WITH UNDERLINE. IF TWO VALUES ARE IDENTICAL, THE ONE WITH THE SMALLER STD IS CONSIDERED THE BETTER. THESE NOTES ARE THE SAME TO OTHERS

|	GiGCN	CLONL	Spectralformer	GSC-VIT	MambaHSI	VMamba	IGroupSS	HSIMFormer	Ours
Asphalt	90.36±1.73	**98.27±1.00**	92.60±2.29	97.10±1.64	94.35±1.73	92.76±2.42	93.41±3.24	93.85±2.80	97.40±2.91
Meadows	98.86±0.78	**98.98±0.76**	93.58±2.14	97.68±1.32	96.43±1.92	97.24±1.14	94.86±2.53	95.11±2.31	98.25±1.27
Gravel	91.66±6.33	87.83±7.44	60.90±8.15	86.46±7.42	93.39±5.35	93.65±4.67	92.58±3.91	92.94±3.60	**97.62±3.77**
Trees	81.34±10.11	97.80±1.46	92.28±6.23	91.83±9.94	**97.83±5.60**	94.35±3.65	93.11±3.42	93.77±2.70	95.38±1.82
Metal sheets	91.15±1.31	98.00±1.72	93.50±2.43	97.65±1.24	**99.95±0.33**	99.75±0.22	99.23±0.45	99.26±0.38	**99.95±0.14**
Bare soil	98.87±1.90	85.27±7.50	63.92±1.24	79.65±9.24	98.60±3.93	97.58±0.87	95.76±1.88	96.10±1.72	**100.00±0.00**
Bitumen	99.28±0.61	90.93±4.12	48.11±8.67	89.50±1.70	94.40±2.93	98.42±1.12	96.43±2.16	96.97±1.93	**99.43±0.82**
Bricks	87.68±7.46	90.93±4.12	81.22±4.66	89.50±1.70	94.40±2.93	95.27±2.50	94.25±2.27	94.65±2.44	**98.39±1.42**
Shadows	97.21±3.36	97.12±0.33	88.78±5.92	**99.36±0.62**	**99.36±0.62**	98.63±1.36	97.89±0.76	98.03±1.19	98.85±0.96
OA(%)	94.47±0.97	94.97±1.49	91.10±2.60	92.65±2.30	95.74±0.90	95.43±0.85	94.18±1.12	94.67±1.06	**98.20±0.69**
AA(%)	92.93±1.07	93.53±1.53	79.93±6.7	91.89±1.73	95.86±1.11	96.43±0.93	95.98±0.89	95.74±0.83	**98.36±0.57**
Kappa(%)	92.67±1.26	93.38±1.84	77.40±4.41	90.38±2.95	95.00±2.24	96.12±1.04	92.40±1.35	93.01±1.22	**98.50±1.56**
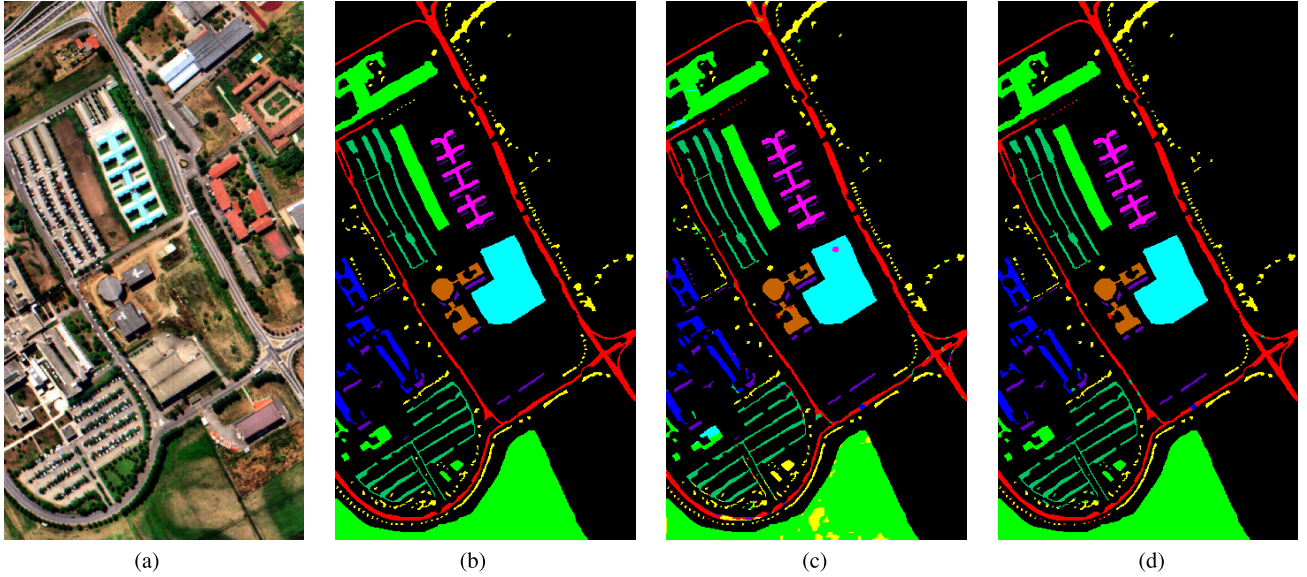
Fig. 4. Qualitative visualization of the classification map for Pavia University dataset. To clearly demonstrate the visual superiority of our method, comparative results are shown exclusively with the SOTA MambaHSI method. Extended visual comparisons with other methods can be found in MambaHSI [38]. (a) Color image. (b) Ground truth. (c) MambaHSI. (d) Ours.

model architectures, including GCN- and CNN-based [11], [50], Transformer-based [16], [19], and Mamba variants [28], [36], [37], [38] methods. The corresponding quantitative and qualitative results are reported in Tables II–V and Figs. 4–7, respectively.

1) Pavia University: Table II provides a detailed quantitative comparison of various methods on the Pavia University dataset, demonstrating the superior performance of our proposed approach. Our method consistently outperforms all competing methods in terms of OA, AA, and Kappa coefficient. Notably, it achieves the highest OA of 98.20% ± 0.69, surpassing the second-best method (MambaHSI) by a significant margin of 2.46 percentage points, underscoring its clear advantage in classification performance. Furthermore, in terms of AA, our method achieves 98.36% ± 0.57, outperforming the second-best method (VMamba), by 1.93 percentage points, further emphasizing its robustness and reliability across a wide range of categories.

As visually substantiated in Fig. 4, our method demonstrates clearer categorical boundaries compared to MambaHSI, particularly in distinguishing spectrally similar categories such as Meadows and Trees. While MambaHSI exhibits confusion

between these classes (e.g., erroneous tree pixels within meadow regions), our framework effectively resolves such ambiguities through enhanced feature discrimination. The proposed architecture reduces scattered misclassifications while maintaining spatial coherence, even in heterogeneous areas. This visual precision aligns with quantitative improvements, confirming our method's superior handling of spectral overlap and spatial complexity.

2) HongHu Dataset: The proposed method exhibits superior performance on the HongHu dataset across three critical evaluation metrics, achieving SOTA results with statistically significant margins. As quantitatively demonstrated in Table III, our framework attains an OA of 96.84% ± 0.45% and an AA of 97.02% ± 0.24%, surpassing the second-best competitor by substantial margins of 2.26 and 2.25 percentage points, respectively. Particularly noteworthy is the method's exceptional Kappa coefficient of 96.57% ± 0.90%, outperforming existing approaches by 2.29 percentage points. These comprehensive improvements not only validate the method's enhanced discriminative capability across diverse categories but also demonstrate its remarkable stability in cross-validation scenarios, establishing new benchmarks

TABLE III
QUANTITATIVE RESULT (ACC% ± STD%) OF HONGHU DATASET

| Class	GiGCN	CLONL	Spectralformer	GSC-ViT	MambaHSI	VMamba	IGroupSS	HSIMFormer	Ours
Red roof	90.93±2.64	95.36±5.56	96.66±1.63	**98.16±1.08**	96.47±1.50	96.04±1.31	94.41±2.17	92.88±7.48	96.55±1.02
Road	51.88±7.77	78.60±6.19	62.84±4.24	73.78±9.09	95.43±2.33	97.02±2.02	92.60±2.51	93.78±2.90	**98.11±0.81**
Bare soil	78.80±2.87	**98.22±1.32**	95.60±2.81	97.21±1.81	94.66±1.05	92.36±1.68	88.05±2.89	85.32±7.64	94.39±1.19
Cotton	99.37±0.32	99.48±0.37	99.42±0.30	**99.57±0.14**	96.96±1.79	98.67±1.04	94.82±2.35	96.07±1.28	98.66±0.80
Cotton firewood	62.41±8.43	41.33±18.92	42.14±16.25	37.03±9.33	96.20±1.77	98.17±1.06	94.43±2.14	90.48±12.37	**98.95±0.81**
Rape	89.61±3.17	96.09±1.07	95.73±1.02	95.72±1.83	93.90±2.47	94.41±1.68	93.35±2.45	93.26±1.96	**97.40±1.05**
Chinese cabbage	82.91±3.07	**92.72±1.96**	90.05±2.26	90.38±3.66	86.60±3.30	80.58±4.00	82.43±4.19	74.48±21.54	90.77±1.56
Pakchoi	75.44±8.66	46.44±6.02	35.59±5.04	32.12±4.83	96.08±2.90	94.44±3.55	88.99±3.01	87.19±16.56	**98.42±2.28**
Cabbage	85.39±3.04	97.04±2.05	97.89±1.07	97.06±2.73	94.78±1.69	95.94±2.02	96.14±1.90	95.15±2.14	**97.91±1.64**
Tuber mustard	92.42±2.80	84.88±3.47	78.81±4.58	79.23±7.53	90.45±2.72	87.99±1.66	87.07±5.44	84.46±12.08	**95.33±2.03**
Brassica parachinensis	75.17±4.50	72.03±8.35	69.29±7.09	72.41±8.95	91.96±2.87	85.97±7.43	84.17±3.15	87.66±3.83	**96.29±1.65**
Brassica chinensis	82.52±5.36	66.63±8.82	58.68±5.96	65.61±8.66	87.40±5.22	89.14±5.22	84.46±5.67	79.39±8.41	**93.41±3.49**
Small Brassica chinensis	89.28±2.70	82.67±5.83	76.20±6.07	77.49±8.47	89.12±3.13	83.20±2.02	77.25±4.13	77.41±3.39	**91.04±1.73**
Lactuca sativa	68.03±5.44	87.59±7.03	83.89±6.15	86.26±11.93	94.01±1.65	92.81±1.47	90.63±4.40	91.68±3.52	**97.34±1.18**
Celtuce	21.65±2.65	75.88±12.93	61.24±13.37	63.62±20.25	98.84±1.17	98.78±2.38	98.51±1.28	98.30±1.58	**99.68±0.45**
Film covered lettuce	71.70±6.79	97.39±1.73	95.23±2.99	97.42±1.66	96.83±1.60	**97.66±1.35**	94.12±2.15	93.73±3.66	97.54±1.77
Romaine lettuce	54.94±4.04	80.09±9.70	68.68±8.79	84.92±6.09	98.60±1.49	98.00±2.08	95.32±1.75	93.95±7.51	**99.61±0.53**
Carrot	67.65±8.17	65.55±15.89	60.99±7.01	67.44±10.50	96.72±1.82	97.52±1.72	98.61±0.97	93.25±16.31	**99.29±0.79**
White radish	72.71±8.86	82.53±7.23	82.77±4.78	85.10±8.11	91.30±3.07	93.38±2.71	94.71±1.74	93.95±3.31	**94.58±2.90**
Garlic sprout	66.48±6.47	81.38±6.60	71.18±8.29	71.64±10.46	99.18±0.42	98.54±1.16	96.22±2.23	91.97±10.71	**99.39±0.45**
Broad bean	65.20±14.58	30.19±12.06	28.91±8.18	25.92±7.91	99.91±0.18	98.16±3.42	99.70±0.41	99.78±0.31	**99.99±0.02**
Tree	85.35±8.84	55.02±10.89	58.01±15.28	46.26±11.15	99.54±0.51	99.47±0.87	96.64±3.79	96.47±4.12	**99.88±0.37**
OA(%)	87.22±0.60	87.93±2.11	86.04±2.83	86.18±1.86	94.58±1.01	94.37±0.44	91.62±1.03	91.14±3.32	**96.84±0.45**
AA(%)	74.08±1.50	77.60±2.25	73.17±2.45	74.74±1.64	94.77±0.61	94.01±0.44	91.94±0.60	90.48±5.84	**97.02±0.24**
Kappa(%)	83.95±0.73	85.06±2.49	82.72±3.22	82.93±2.16	94.00±1.88	89.62±2.10	89.53±1.23	88.89±4.18	**96.57±0.90**
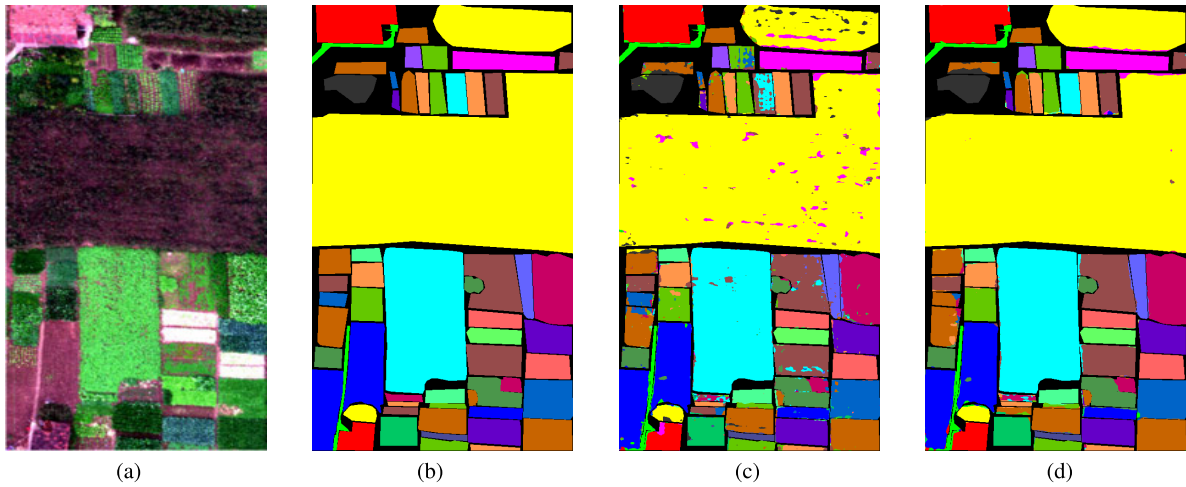
Fig. 5. Qualitative visualization of the classification map for the HongHu dataset. (a) Color image. (b) Ground truth. (c) MambaHSI. (d) Ours.

for classification reliability in complex remote sensing applications.

Visual comparisons reveal distinct advantages of our method in critical classification scenarios. As shown in Fig. 5, the MambaHSI approach produces noticeable misclassifications in rape fields (yellow regions) and cotton areas (pink regions), particularly along field boundaries where spectral mixing occurs. Significant edge confusion persists across different crop types, with errors propagating into homogeneous regions. In contrast, our method achieves near-perfect categorical consistency, precisely delineating crop boundaries even in geometrically complex transition zones. The limited residual errors are primarily confined to subpixel mixed areas at crop junctions, demonstrating remarkable robustness against spectral ambiguities. This visual superiority aligns with quantitative metrics, confirming our framework's enhanced capacity to resolve challenging interclass variations in agricultural landscapes.

3) HanChuan Dataset: As quantitatively demonstrated in Table IV, our method sets new SOTA benchmarks across all evaluation metrics, achieving an unprecedented OA of 95.02% ± 1.01%, AA of 93.91% ± 0.62%, and a Kappa coefficient of 92.60% ± 1.90%. These results represent statistically significant improvements of 4.81, 4.71, and 3.53 percentage points over the baseline method (MambaHSI) in OA, AA, and Kappa, respectively, demonstrating comprehensive superiority in both per-class discriminative power (as evidenced by AA gains) and classification consistency (validated through Kappa improvements).

Visual comparisons further confirm our method's superiority in boundary delineation between crop types. As shown in Fig. 6, MambaHSI exhibits blurred edges and fragmented misclassifications at junctions of spectrally similar crops (e.g., tree and roof fields), whereas our approach generates sharply defined boundaries with minimal edge artifacts. The proposed method significantly reduces confusion in transitional zones,

TABLE IV
QUANTITATIVE RESULT (ACC% ± STD%) OF HANCHUAN DATASET

| Class	GiGCN	CLONL	Spectralformer	GSC-ViT	MambaHSI	VMamba	IGroupSS	HSIMFormer	Ours
Strawberry	89.34±2.23	95.97±1.87	95.13±1.25	95.26±1.50	89.72±5.75	94.25±1.93	96.74±1.01	94.50±2.23	**97.03±1.27**
Cowpea	81.64±3.38	92.48±2.56	87.27±3.37	91.22±3.05	81.37±5.69	83.48±5.09	90.77±2.57	88.73±2.90	**96.54±1.92**
Soybean	81.06±6.56	80.03±9.18	70.36±6.15	78.37±7.47	94.88±3.73	83.51±4.97	88.96±3.87	92.09±3.49	**97.85±2.63**
Sorghum	67.09±4.82	91.12±7.72	88.62±6.67	85.17±12.38	98.46±1.35	98.72±0.92	97.55±1.01	98.86±0.55	**99.10±0.37**
Water spinach	34.98±2.76	56.82±12.89	42.01±11.48	46.57±13.03	99.54±0.85	91.31±6.88	98.83±0.79	97.84±0.88	**99.79±0.21**
Watermelon	53.81±9.81	38.07±8.70	34.68±6.40	35.97±7.01	76.99±4.95	73.77±7.99	88.63±4.52	90.46±4.16	**93.59±3.29**
Greens	80.84±8.87	68.87±17.46	67.69±10.29	66.90±15.67	92.61±4.63	94.47±4.07	97.12±3.76	98.74±1.75	**99.31±0.86**
Trees	**89.27±4.80**	81.52±7.87	83.17±4.16	82.39±5.23	70.73±7.10	84.15±4.41	84.99±3.78	88.85±3.60	88.67±3.26
Grass	65.14±6.79	68.59±6.61	63.00±4.53	60.64±17.62	89.01±3.62	86.65±4.01	91.25±2.52	86.19±5.91	**93.94±2.16**
Red roof	93.03±4.21	90.79±14.71	88.76±12.99	93.22±4.31	94.44±2.18	94.83±2.81	92.44±3.85	94.99±3.23	**96.55±1.77**
Gray roof	93.86±5.06	87.86±4.63	77.41±13.19	82.13±7.96	89.56±4.91	87.37±7.36	94.78±3.20	88.93±6.55	**95.11±2.13**
Plastic	89.22±7.45	63.48±14.80	51.40±6.89	51.92±6.01	88.83±6.91	85.11±3.85	88.56±4.33	89.39±5.18	**92.78±3.50**
Bare soil	66.66±6.60	50.21±6.71	47.62±9.17	53.28±10.88	78.69±3.73	78.88±7.10	84.92±4.91	85.18±4.26	**87.23±3.82**
Road	91.25±2.43	90.78±2.74	88.92±3.08	90.64±3.07	89.68±3.03	90.62±5.17	91.07±3.03	90.92±3.84	**93.36±2.76**
Bright object	46.57±17.42	62.42±10.67	52.51±8.59	70.05±12.35	94.42±2.41	95.22±2.61	96.65±2.30	95.13±2.47	**97.27±1.75**
Water	99.75±0.14	98.97±1.02	99.47±0.33	**99.77±0.23**	98.28±1.62	99.11±0.45	98.50±1.10	99.56±0.37	99.68±0.19
OA(%)	86.78±1.43	85.14±2.93	82.99±2.31	83.14±3.96	90.21±1.67	91.68±1.16	93.48±1.17	92.83±1.43	**95.02±1.01**
AA(%)	76.47±1.62	76.13±3.26	71.13±2.97	75.41±3.97	89.26±1.73	88.84±1.19	91.38±1.42	90.75±1.52	**94.48±0.97**
Kappa(%)	82.22±1.64	81.62±2.92	78.73±2.55	78.81±4.33	87.51±2.42	80.87±5.08	**91.07±1.47**	89.89±1.92	90.93±1.89
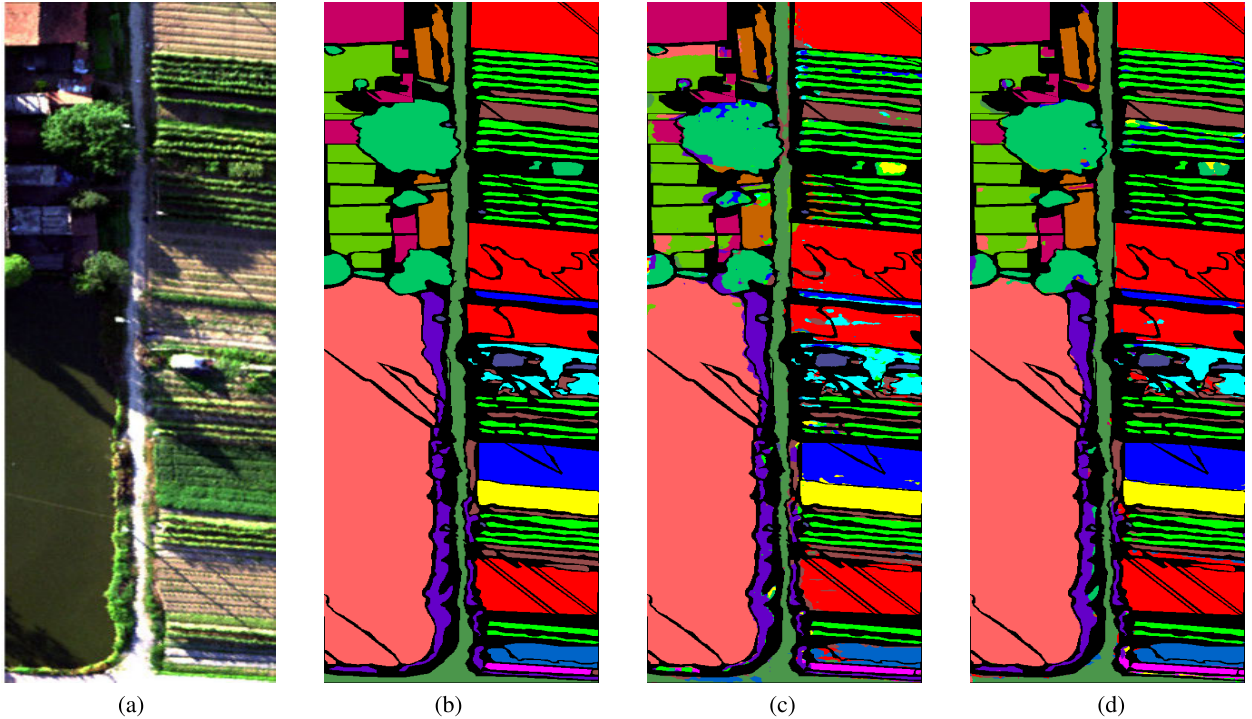
Fig. 6. Qualitative visualization of the classification map for the HanChuan dataset. (a) Color image. (b) Ground truth. (c) MambaHSI. (d) Ours.

achieving spatially coherent classification maps that align closely with ground-truth distributions. This visual precision complements the quantitative improvements in OA and Kappa metrics, reinforcing the method's robustness in complex agricultural scenarios.

4) Houston Dataset: As systematically benchmarked in Table V and Fig. 7, our method establishes the SOTA performance across all evaluation metrics on the Houston dataset. Achieving an OA of 94.96% ± 0.78%, our framework outperforms the second-best method (MambaHSI, 94.46% ± 0.82%) with statistically significant margins, demonstrating enhanced discriminative capacity across spectrally heterogeneous land-cover categories. The AA superiority (95.65% ± 0.75% versus MambaHSI's 95.34% ± 0.78%)

further validates our model's robustness in handling class-imbalanced distributions, particularly for underrepresented categories with limited training samples. Notably, the Kappa coefficient of 94.50% ± 1.60%—exceeding MambaHSI by 0.29 percentage points—provides rigorous statistical evidence of superior classification consistency, confirming minimized systematic bias in complex urban–natural interface modeling.

5) Robustness and Classification Consistency: The proposed method demonstrates enhanced robustness through superior metric stability, as manifested by its minimized STDs across all evaluation criteria compared to competing approaches. This reduced variability in OA, AA, and Kappa metrics confirms reliable performance under heterogeneous data conditions. Notably, the Kappa coefficient

TABLE V
QUANTITATIVE RESULT (ACC% ± STD%) OF HOUSTON DATASET

| Class	GiGCN	CLONL	Spectralformer	GSC-ViT	MambaHSI	VMamba	IGroupSS	HSIMFormer	Ours
Healthy Grass	83.21±5.03	88.24±9.00	91.29±5.38	90.16±6.48	**95.20±4.66**	90.43±3.68	91.40±4.62	92.33±4.92	92.73±4.54
Stressed Grass	82.83±7.41	93.41±5.99	92.12±6.99	95.77±2.95	**98.29±1.26**	91.05±2.81	91.48±5.35	97.82±1.58	94.85±1.55
Synthetic Grass	98.83±1.66	95.21±9.84	94.68±3.25	**99.83±0.26**	99.74±0.46	99.19±0.82	98.08±1.72	99.15±0.52	99.59±0.58
Tree	72.42±10.94	97.44±3.80	93.32±9.22	94.78±7.69	97.26±2.47	92.90±3.83	94.58±2.90	**97.57±1.92**	97.03±2.08
Soil	93.14±4.62	97.07±2.86	98.18±1.72	97.63±2.69	99.54±0.76	99.51±0.77	99.42±0.64	**99.96±0.12**	99.75±0.51
Water	88.99±6.78	94.85±4.72	87.21±8.11	96.03±2.59	97.47±2.33	97.40±3.40	97.09±2.56	**99.19±1.00**	98.84±2.06
Residential	82.34±5.24	94.02±2.94	91.47±3.54	93.11±2.39	93.03±2.25	90.68±1.93	89.28±3.44	87.69±3.93	**94.34±2.41**
Commercial	91.02±6.25	**93.80±3.94**	84.41±7.91	90.53±8.57	81.35±3.98	76.59±3.76	76.25±7.48	80.44±6.77	81.85±5.92
Road	85.51±6.43	88.32±2.83	87.75±3.83	85.20±4.96	90.03±3.40	**93.57±2.89**	92.48±3.02	89.16±4.92	93.33±2.16
Highway	94.92±2.69	76.55±6.96	82.09±5.12	85.28±4.22	96.55±1.49	98.64±2.03	95.95±3.28	98.13±2.28	**99.34±0.70**
Railway	96.52±3.78	91.57±4.40	85.90±4.63	93.08±3.33	92.74±2.18	92.38±3.82	89.51±7.38	**98.53±2.41**	95.62±2.87
Parking Lot 1	92.45±3.63	83.37±6.17	85.33±5.87	88.46±3.14	91.19±3.68	90.39±2.47	90.96±4.21	93.31±3.72	**94.45±2.52**
Parking Lot 2	88.10±5.13	89.11±4.64	86.81±6.28	90.01±4.69	**97.67±2.04**	93.19±2.73	89.65±5.58	96.40±2.91	96.83±2.46
Tennis Court	98.30±2.67	97.98±2.63	91.65±6.36	97.01±5.11	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**
Running Track	99.14±1.89	96.08±3.96	96.18±2.66	97.73±1.23	**100.00±0.00**	93.13±3.03	89.29±4.02	99.98±0.05	96.15±2.87
OA(%)	88.13±2.26	90.31±1.73	89.13±2.12	91.85±1.05	94.46±0.83	92.40±0.64	91.67±2.30	94.39±1.05	**94.96±0.78**
AA(%)	89.85±1.91	91.80±1.46	89.89±1.74	92.97±0.92	95.34±0.78	93.27±0.57	92.36±1.94	95.31±0.89	**95.65±0.75**
Kappa(%)	87.16±2.44	89.52±1.87	88.24±2.29	91.18±1.14	94.21±0.83	92.07±0.64	90.43±3.00	93.94±1.05	**94.50±0.78**

(a)

(b)

(c)

(d)
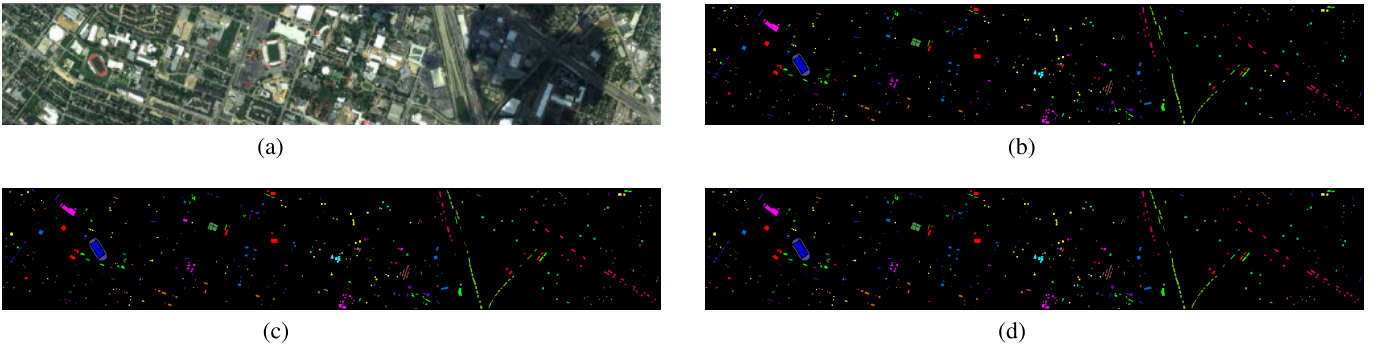
Fig. 7. Qualitative visualization of the classification map for the Houston dataset. (a) Color image. (b) Ground truth. (c) MambaHSI. (d) Ours.

improvement over baseline methods highlights strengthened classification consensus, reflecting precise alignment between spectral–spatial feature extraction and ground-truth distributions. Such dual optimization of precision stability (STD) and categorical agreement (Kappa) establishes our framework's capability to mitigate error propagation at category junctions while maintaining operational reliability.

In summary, systematic evaluations across four benchmark datasets (Pavia University, HongHu, HanChuan, and Houston) confirm our method's superior classification accuracy and robustness. The framework achieves SOTA performance in OA (94.96%–98.50%), AA (93.91%–97.02%), and Kappa coefficients (92.60%–96.57%), outperforming MambaHSI by 0.50–4.81 percentage points in OA across datasets. Remarkably, low STDs (OA STD<1.01%) demonstrate stability against spectral–spatial variations, while Kappa values exceeding 92.60% validate strong label consensus. These results collectively highlight the method's capability to balance precision and reliability in complex hyperspectral scenarios.

C. Ablation Study

To rigorously decompose the contributions of architectural components, we conduct systematic ablation experiments across four benchmark datasets: Pavia University, HongHu, HanChuan, and Houston. Table VI presents quantitative comparisons using OA, where each row removes one critical component while maintaining others.

1) *Bidirectional Spatial Modeling:* The RBCM ablation induces the most severe performance deterioration (0.51%–2.06% OA reduction), substantiating that bidirectional state transitions enable comprehensive spatial context modeling through synergistic forward–backward feature propagation.

2) *Reverse-Order Augmentation:* Performance degradation of 0.40%–2.50% across datasets confirms that sequence permutation strengthens spatial pattern learning by enforcing permutation-invariant feature representations.

3) *Adaptive Spatial Fusion:* The 0.24%–0.75% OA decrease with direct summation demonstrates ASF's superiority in dynamically weighting spatial features through context-dependent attention mechanisms.

4) *Spectral Trajectory Learning:* STL removal causes 0.95%–2.57% accuracy decline, emphasizing the necessity of modeling spectral–temporal evolution via differential state-space equations.

5) *Mamba Channel Attention:* The 0.39%–0.86% OA reduction verifies MCA's effectiveness in spectral feature recalibration through gated state-space transformations.

6) *Parameter Sharing Mechanism:* Compared to the nonshared variant, our shared-parameter design achieves slightly improved OA across three datasets (0.03%–0.13%), with a significant reduction in parameter count from 2.49 to 1.73 M. This demonstrates the design's

TABLE VI

SYSTEMATIC COMPONENT ABLATION ANALYSIS ACROSS HYPERSPECTRAL DATASETS (OA, %). SUBSCRIPTS INDICATE
ACCURACY DEGRADATION RELATIVE TO COMPLETE FRAMEWORK. VALUES COMPUTED AS $OA_{FULL} - OA_{ABLATED}$

Configuration	Pavia University	HongHu	HanChuan	Houston
Full Framework	98.20	96.84	95.02	94.96
w/o Bidirectional Spatial Modeling	$96.45_{(-1.75)}$	$94.78_{(-2.06)}$	$93.12_{(-1.90)}$	$94.45_{(-0.51)}$
w/o Reverse-Order Augmentation	$96.12_{(-2.08)}$	$94.34_{(-2.50)}$	$93.23_{(-1.79)}$	$94.56_{(-0.40)}$
w/o Adaptive Spatial Fusion	$97.45_{(-0.75)}$	$96.38_{(-0.46)}$	$94.67_{(-0.35)}$	$94.72_{(-0.24)}$
w/o Spectral Trajectory Learning	$96.78_{(-1.42)}$	$95.89_{(-0.95)}$	$92.45_{(-2.57)}$	$93.78_{(-1.18)}$
w/o Mamba Channel Attention	$97.34_{(-0.86)}$	$96.45_{(-0.39)}$	$94.32_{(-0.70)}$	$94.34_{(-0.62)}$
w/o Parameter Sharing	$98.17_{(-0.03)}$	$96.72_{(-0.12)}$	$95.11_{(+0.09)}$	$94.83_{(-0.13)}$

TABLE VII

COMPARISON OF MODEL EFFICIENCY IN TERMS OF PARAMETER COUNT, GFLOPS, AND INFERENCE TIME PER
IMAGE (MEASURED ON THE PAVIA UNIVERSITY DATASET)

Model	Params (M)	GFLOPs	Inference Time (s)	OA	AA	Kappa
GSC-ViT	**0.153**	2026.30	49.44	92.65	91.89	90.38
SpectralFormer	0.177	9086.19	62.43	91.10	79.93	77.40
MambaHSI	0.412	**39.18**	**0.03**	95.74	95.86	95.00
MambaHSI+ (Ours)	1.730	148.15	0.37	**98.20**	**98.36**	**98.50**

strong parameter efficiency and generalization capacity. Although a marginal performance drop (0.09%) occurs in some datasets, the tradeoff is worthwhile given the substantial computational savings.

The systematic ablation studies reveal critical insights into component contributions across four hyperspectral benchmarks. RBCM demonstrates the most pronounced impact (0.51%–2.06% OA reduction), particularly affecting Pavia University (−1.75%) and HongHu (−2.06%), confirming its essential role in spatial context modeling. Reverse-order augmentation shows dataset-dependent significance, achieving maximum effectiveness on HongHu (−2.50%) while minimally impacting Houston (−0.40%), suggesting inherent rotational variance differences in urban versus agricultural landscapes. Adaptive spatial fusion (ASF) exhibits consistent but moderate improvements (0.24%–0.75%), with peak performance on Pavia University, validating its spatial weighting superiority. STL emerges as the most influential spectral component, particularly crucial for HanChuan's complex vegetation patterns (−2.57% reduction). Mamba channel attention demonstrates universal effectiveness (0.39%–0.86% degradation), with maximal impact on Pavia University's spectral features.

D. Model Efficiency Analysis

To further verify the computational efficiency of our proposed method, we conduct a comparative study of model complexity and inference latency among several representative methods.

As shown in Table VII, MambaHSI+ consistently achieves superior classification performance, obtaining the highest OA (98.20%), AA (98.36%), and Kappa coefficient (98.50) among all compared methods. Although MambaHSI+ introduces a moderate increase in parameter count (1.73 M) compared to MambaHSI (0.41 M), it maintains an efficient inference time of only 0.37 s on the Pavia University dataset. More importantly, it significantly reduces the floating-point operations (GFLOPs) relative to mainstream Transformer-based methods such as GSC-ViT (2026.30 GFLOPs) and SpectralFormer

(9086.19 GFLOPs), demonstrating a substantial computational advantage. While slightly higher than the baseline MambaHSI (39.18 GFLOPs), our method (148.15 GFLOPs) offers markedly improved accuracy while preserving competitive efficiency.

Regarding the computational overhead introduced by the bidirectional propagation mechanism in MambaHSI+, we provide a supplementary discussion here. Compared to the unidirectional design in MambaHSI, the proposed bidirectional structure enhances contextual modeling capabilities but inevitably brings additional memory consumption due to the need to maintain forward and backward hidden states. The peak GPU memory usage on the Pavia University dataset increases from 1763 MB (MambaHSI) to 8172 MB (MambaHSI+). While this indicates a rise in resource demand, the overhead remains acceptable under modern hardware configurations and is justified by the significant accuracy gains achieved by MambaHSI+.

Furthermore, we analyze the theoretical time complexity of MambaHSI+. The bidirectional propagation mechanism operates sequentially over token positions, with a time complexity of $\mathcal{O}(N)$ for each direction, where N denotes the number of tokens. Since both directions are computed independently and their outputs fused afterward, the overall time complexity remains linear with respect to the sequence length.

To further validate this theoretical insight, we conducted experiments by feeding inputs of varying spatial sizes into the model and measuring the corresponding runtime. The results confirm that MambaHSI+ exhibits empirical linear scaling behavior, which is consistent with the theoretical complexity analysis.

These results confirm that MambaHSI+ achieves a favorable tradeoff between model complexity and runtime efficiency. The proposed model not only inherits the linear computational characteristics of the original MambaHSI design but also incorporates an enhanced feature encoding strategy to deliver the SOTA performance under practical computational constraints. This demonstrates the method's suitability for

deployment in real-world HSI classification scenarios where both accuracy and efficiency are critical.

V. Conclusion

This study presents MambaHSI+, an advanced framework for HSI classification that integrates bidirectional state-space modeling with STL. The architecture systematically addresses three key limitations in current approaches.

1) The reverse bidirectional contextual modeling module enhances spatial context modeling through four-directional propagation, enabling robust pixel-level feature aggregation without compromising computational efficiency.

2) The STL mechanism adaptively encodes cross-band correlations via bidirectional state transitions, reducing spectral redundancy while maintaining interband continuity.

3) A Mamba-driven attention scheme selectively prioritizes discriminative spectral features through state-space transformations, optimizing hardware utilization. Evaluations across multiple hyperspectral benchmarks confirm consistent improvements over existing methods, particularly in handling spectral ambiguities and spatially heterogeneous regions. By unifying spatial–spectral feature extraction with linear computational complexity, this framework provides a scalable solution for large-scale hyperspectral analysis. Future directions include extending the model to multimodal data integration and limited-sample learning, with applications in ecological monitoring and agricultural resource management.

References

[1] B. Du, Y. Zhang, L. Zhang, and D. Tao, "Beyond the sparsity-based target detector: A hybrid sparsity and statistics-based detector for hyperspectral images," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5345–5357, Nov. 2016.

[2] Q. Zhu et al., "A spectral–spatial-dependent global learning framework for insufficient and imbalanced hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 52, no. 11, pp. 11709–11723, Nov. 2022.

[3] M. Shimoni, R. Haelterman, and C. Perneel, "Hyperspectral imaging for military and security applications: Combining myriad processing and sensing techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 101–117, Jun. 2019.

[4] Z. Wu et al., "Scheduling-guided automatic processing of massive hyperspectral image classification on cloud computing architectures," *IEEE Trans. Cybern.*, vol. 51, no. 7, pp. 3588–3601, Jul. 2021.

[5] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004, doi: 10.1109/TGRS.2004.831865.

[6] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005, doi: 10.1109/TGRS.2004.842481.

[7] X. Yang, Y. Ye, X. Li, R. Y. K. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5408–5423, Sep. 2018.

[8] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.

[9] C. Yu, R. Han, M. Song, C. Liu, and C.-I. Chang, "Feedback attention-based dense CNN for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5501916.

[10] M. Zhu, J. Fan, Q. Yang, and T. Chen, "SC-EADNet: A self-supervised contrastive efficient asymmetric dilated network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5519517.

[11] C. Li, B. Rasti, X. Tang, P. Duan, J. Li, and Y. Peng, "Channel-layer-oriented lightweight spectral–spatial network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5504214, doi: 10.1109/TGRS.2024.3350055.

[12] M. Li, Y. Liu, G. Xue, Y. Huang, and G. Yang, "Exploring the relationship between center and neighborhoods: Central vector oriented self-similarity network for hyperspectral image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 4, pp. 1979–1993, Apr. 2023.

[13] Y. Yang et al., "Semi-supervised multiscale dynamic graph convolution network for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 5, pp. 6806–6820, May 2024.

[14] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[15] W. Zhou, S.-I. Kamata, H. Wang, P. Lu, and M. Zhang, "Segmented recurrent transformer with cubed 3-D-multiscanning strategy for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5513322.

[16] D. Hong et al., "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518615.

[17] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 5998–6008. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fb d053c1c4a845aa-Abstract.html

[18] X. Yang, W. Cao, Y. Lu, and Y. Zhou, "Hyperspectral image transformer classification networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5528715.

[19] Z. Zhao, X. Xu, S. Li, and A. Plaza, "Hyperspectral image classification using groupwise separable convolutional vision transformer network," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5511817.

[20] P. Wang, Z. He, B. Huang, M. Dalla Mura, H. Leung, and J. Chanussot, "VOGTNet: Variational optimization-guided two-stage network for multispectral and panchromatic image fusion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 5, pp. 9268–9282, May 2025.

[21] P. Wang et al., "Low-rank tensor completion pansharpening based on haze correction," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5405720.

[22] F. Xu, G. Zhang, C. Song, H. Wang, and S. Mei, "Multiscale and cross-level attention learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5501615.

[23] J. Zou, W. He, and H. Zhang, "LESSFormer: Local-enhanced spectral–spatial transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5535416.

[24] K. Wu, J. Fan, P. Ye, and M. Zhu, "Hyperspectral image classification using spectral–spatial token enhanced transformer with hash-based positional embedding," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5507016.

[25] S. Mei, C. Song, M. Ma, and F. Xu, "Hyperspectral image classification using group-aware hierarchical transformer," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5539014.

[26] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, "Spectral–spatial feature tokenization transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5522214.

[27] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," 2023, *arXiv:2312.00752*.

[28] Y. Liu et al., "VMamba: Visual state space model," 2024, *arXiv:2401.10166*.

[29] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, "Vision mamba: Efficient visual representation learning with bidirectional state space model," in *Proc. ICML*, Jan. 2024, pp. 1–10. [Online]. Available: https://openreview.net/forum?id=YbHCqn4qF4

[30] S. Zhao, H. Chen, X. Zhang, P. Xiao, L. Bai, and W. Ouyang, "RS-Mamba for large remote sensing image dense prediction," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024.

[31] A. Gu et al., "Combining recurrent, convolutional, and continuous-time models with linear state space layers," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, M. Ranzato, A. Beygelzimer, Y. N. Dauphin, P. Liang, and J. W. Vaughan, Eds., 2021, pp. 572–585. [Online]. Available: https://proceedings.neurips.cc/paper/2021/hash/05546b0e38ab9175cd 905eebcc6ebb76-Abstract.html

[32] A. Gu, K. Goel, and C. Ré, "Efficiently modeling long sequences with structured state spaces," in *Proc. 10th Int. Conf. Learn. Represent. (ICLR)*, Jan. 2021. [Online]. Available: https://openreview.net/ forum?id=uYLFoz1vlAC

[33] R. Xu, S. Yang, Y. Wang, Y. Cai, B. Du, and H. Chen, "Visual Mamba: A survey and new outlooks," 2024, *arXiv:2404.18861*.

[34] Y. Oshima, S. Taniguchi, M. Suzuki, and Y. Matsuo, "SSM meets video diffusion models: Efficient long-term video generation with structured state spaces," 2024, *arXiv:2403.07711*.

[35] D. Liang et al., "PointMamba: A simple state space model for point cloud analysis," in *Proc. Adv. Neural Inf. Process. Syst.*, Feb. 2024, pp. 1–11. [Online]. Available: http://papers.nips.cc/paperfiles/paper/2024/hash/395371f778ebd4854b88521100af30ad-Abstract-Conference.html

[36] Y. He, B. Tu, P. Jiang, B. Liu, J. Li, and A. Plaza, "IGroupSS-Mamba: Interval group spatial–spectral mamba for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5538817.

[37] Y. He, B. Tu, B. Liu, J. Li, and A. Plaza, "HSI-MFormer: Integrating mamba and transformer experts for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5621916.

[38] Y. Li, Y. Luo, L. Zhang, Z. Wang, and B. Du, "MambaHSI: Spatial–spectral Mamba for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5524216.

[39] S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing*, vol. 219, pp. 88–98, Jan. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231216310104

[40] M. Ahmad, A. M. Khan, M. Mazzara, S. Distefano, M. Ali, and M. S. Sarfraz, "A fast and compact 3-D CNN for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[41] S. Li, X. Zhu, and J. Bao, "Hierarchical multi-scale convolutional neural networks for hyperspectral image classification," *Sensors*, vol. 19, no. 7, p. 1714, Apr. 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/7/1714

[42] D. Wang, B. Du, and L. Zhang, "Fully contextual network for hyperspectral scene parsing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5501316.

[43] Y. Xu, B. Du, and L. Zhang, "Beyond the patchwise classification: Spectral–spatial fully convolutional networks for hyperspectral image classification," *IEEE Trans. Big Data*, vol. 6, no. 3, pp. 492–506, Sep. 2020.

[44] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, Aug. 2019.

[45] S. Mei, X. Li, X. Liu, H. Cai, and Q. Du, "Hyperspectral image classification using attention-based bidirectional long short-term memory network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5509612.

[46] X. Zhang, Y. Sun, K. Jiang, C. Li, L. Jiao, and H. Zhou, "Spatial sequential recurrent neural network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4141–4155, Nov. 2018.

[47] W. Zhou, S.-I. Kamata, Z. Luo, and H. Wang, "Multiscanning strategy-based recurrent neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5521018.

[48] A. Qin, Z. Shang, J. Tian, Y. Wang, T. Zhang, and Y. Y. Tang, "Spectral–spatial graph convolutional networks for semisupervised hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 241–245, Feb. 2019.

[49] S. Wan, C. Gong, P. Zhong, S. Pan, G. Li, and J. Yang, "Hyperspectral image classification with context-aware dynamic graph convolutional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 597–612, Jan. 2021.

[50] S. Jia, S. Jiang, S. Zhang, M. Xu, and X. Jia, "Graph-in-graph convolutional network for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 1, pp. 1157–1171, Jan. 2024.

[51] Y. Peng, Y. Zhang, B. Tu, Q. Li, and W. Li, "Spatial–spectral transformer with cross-attention for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5537415.

[52] J. Yang, B. Du, and L. Zhang, "From center to surrounding: An interactive learning framework for hyperspectral image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 197, pp. 145–166, Mar. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271623000308

[53] Z. Li, Z. Xue, Q. Xu, L. Zhang, T. Zhu, and M. Zhang, "SPFormer: Self-pooling transformer for few-shot hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5502019.

[54] Z. Xue, Q. Xu, and M. Zhang, "Local transformer with spatial partition restore for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4307–4325, 2022.

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

[56] D. Wang, J. Zhang, B. Du, L. Zhang, and D. Tao, "DCN-T: Dual context network with transformer for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 32, pp. 2536–2551, 2023.

[57] Y. Li, Y. Zhou, Y. Zhang, L. Zhong, J. Wang, and J. Chen, "DKDFN: Domain knowledge-guided deep collaborative fusion network for multimodal unitemporal remote sensing land cover classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 186, pp. 170–189, Apr. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271622000557

[58] Y. Li, Z. Zhu, J.-G. Yu, and Y. Zhang, "Learning deep cross-modal embedding networks for zero-shot remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10590–10603, Dec. 2021.

[59] Y. Wang et al., "HSIMAE: A unified masked autoencoder with large-scale pretraining for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 14064–14079, 2024.

[60] J. Yang, B. Du, Y. Xu, and L. Zhang, "Can spectral information work while extracting spatial distribution?—An online spectral information compensation network for HSI classification," *IEEE Trans. Image Process.*, vol. 32, pp. 2360–2373, 2023.

[61] J. T. H. Smith, A. Warrington, and S. W. Linderman, "Simplified state space layers for sequence modeling," in *Proc. 11th Int. Conf. Learn. Represent. (ICLR)*, Kigali, Rwanda, Jan. 2022, pp. 1–35. [Online]. Available: https://openreview.net/forum?id=Ai8Hw3AXqks

[62] H. Mehta, A. Gupta, A. Cutkosky, and B. Neyshabur, "Long range language modeling via gated state spaces," in *Proc. 11th Int. Conf. Learn. Represent. (ICLR)*, Kigali, Rwanda, Jan. 2022, pp. 1–15. [Online]. Available: https://openreview.net/forum?id=5MkYIYCbva

[63] T. Dao, D. Y. Fu, K. K. Saab, A. W. Thomas, A. Rudra, and C. Ré, "Hungry hungry hippos: Towards language modeling with state space models," in *Proc. 11th Int. Conf. Learn. Represent. (ICLR)*, Kigali, Rwanda, Jan. 2022. [Online]. Available: https://openreview.net/forum?id=COZDy0WYGg

[64] E. Nguyen et al., "S4ND: Modeling images and videos as multidimensional signals using state spaces," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., Red Hook, NY, USA: Curran Associates, Dec. 2022, pp. 13388–13400. [Online]. Available: https://proceedings.neurips.cc/paperfiles/paper/2022/file/13388efc819c09564c66ab2dc8463809-Paper-Conference.pdf

[65] M. M. Islam, M. Hasan, K. S. Athrey, T. Braskich, and G. Bertasius, "Efficient movie scene detection using state-space transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 18749–18758.

[66] J. Wang et al., "Selective structured state-spaces for long-form video understanding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 6387–6397.

[67] J. Ma, F. Li, and B. Wang, "U-mamba: Enhancing long-range dependency for biomedical image segmentation," 2024, *arXiv:2401.04722*.

[68] J. Yao, D. Hong, C. Li, and J. Chanussot, "SpectralMamba: Efficient mamba for hyperspectral image classification," 2024, *arXiv:2404.08489*.

[69] Y. He, B. Tu, B. Liu, J. Li, and A. Plaza, "3DSS-Mamba: 3D-spectral–spatial Mamba for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5534216.

[70] J. Sheng, J. Zhou, J. Wang, P. Ye, and J. Fan, "DualMamba: A lightweight spectral–spatial Mamba-convolution network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5501415.

[71] L. Huang, Y. Chen, and X. He, "Spectral–spatial Mamba for hyperspectral image classification," *Remote Sens.*, vol. 16, no. 13, p. 2449, Jul. 2024. [Online]. Available: https://www.mdpi.com/2072-4292/16/13/2449

[72] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, and L. Zhang, "WHU-hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF," *Remote Sens. Environ.*, vol. 250, Dec. 2020, Art. no. 112012.