MambaHSINet: A Dual-Branch Bidirectional State Space Network for Hyperspectral Tree Species Classification

Xinying Liu , Yanfeng Zhang , Junyang Wu , Tianyu Cai , Yumeng Li , Xinran Wang , Xinwei Li *

Posted Date: 25 May 2026

doi: 10.20944/preprints202605.1614.v1

Keywords: hyperspectral image classification; tree species classification; st ate space model; Mamba

 doi:10.20944/preprints202605.1614.v1

Article

MambaHSINet: A Dual-Branch Bidirectional State Space Network for Hyperspectral Tree Species Classification

Xinying Liu [1,†], Yanfeng Zhang[1,†], Junyang Wu [1], Tianyu Cai [1], Yumeng Li [1], Xinran Wang [2] and Xinwei Li [1,*]

[1] School of Sciences, Beijing Forestry University, Beijing 100083, China
[2] School of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China
* Correspondence: lixinwei@bjfu.edu.cn; Tel.: +86-18810457603
† These authors contributed equally to this work.

Abstract

Hyperspectral remote sensing provides rich spectral information and has been widely used in fine-grained land-cover classification and forest monitoring. However, accurate tree species classification remains challenging due to subtle interspecific spectral differences, similar spatial structures among related species, redundant spectral bands, and the limited ability of existing methods to model long-range spatial–spectral dependencies efficiently. In addition, many existing hyperspectral image classification methods rely on patch-based inputs and sliding-window inference, which often lead to redundant computation and insufficient utilization of global image context. To address these issues, this paper proposes MambaHSINet, a dual-branch bidirectional state space network for full-image pixel-wise hyperspectral tree species classification. Specifically, the proposed network employs a spectral branch and a spatial branch to explicitly extract complementary spectral responses and spatial structural features. A lightweight channel attention mechanism is introduced to emphasize informative spectral–spatial representations while suppressing redundant information. Subsequently, a bidirectional Mamba global modeling module based on selective state space modeling is adopted to capture long-range contextual dependencies in both forward and backward directions with linear computational complexity. Unlike conventional patch-based methods, MambaHSINet takes the entire hyperspectral image as input and produces full-resolution pixel-wise classification maps, thereby avoiding repeated cropping and redundant sliding-window inference. Experimental results on public and self-collected hyperspectral datasets demonstrate that the proposed method achieves a favorable balance among classification accuracy, inference efficiency, and model complexity, showing strong potential for practical tree species classification applications.The source code of this paper will be made available at https://github.com/YFENG-123/mambaHSI after the publication of the paper.

Keywords: hyperspectral image classification; tree species classification; state space model; Mamba

1. Introduction

Hyperspectral remote sensing has emerged as a cornerstone of modern Earth observation, offering unparalleled capabilities for forestry applications, such as forest structural inversion, tree species mapping, biomass estimation, and forest health monitoring [1–3]. By acquiring hundreds of contiguous narrow spectral bands, hyperspectral imagery (HSI) provides a wealth of information that enables the identification of subtle land-cover variations. Owing to its fine spectral resolution, it can characterize minor reflectance differences that are imperceptible to conventional multispectral sensors, thereby providing a robust foundation for characterizing complex vegetation canopies and precise forest resource monitoring [4,5].

 doi:10.20944/preprints202605.1614.v1

In the context of tree species classification, HSI presents distinct advantages as it captures unique spectral signatures dictated by leaf structure, pigment content, water status, and biochemical [6,7]. Compared with traditional field surveys, RGB-based visual interpretation and multispectral imagery, HSI offers enhanced separability for tree species identification within complex forest ecosystems [8–11]. Consequently, hyperspectral tree species classification has evolved into a pivotal research domain. Accurate tree species identification not only contributes to forest inventory and ecological assessment, but also provides critical data support for biodiversity conservation, carbon cycle modeling, and sustainable forest management.

The methodological evolution of hyperspectral image classification (HSIC) has progresssed from traditional machine learning to deep learning models. Classical classifiers, such as support vector machine (SVM), random forest (RF), k-nearest neighbor (KNN), and sparse representation classification (SRC), generally rely on hand-crafted features and have limited capability in extracting deep discriminative representations [12–15]. Although these methods perform adequately on simple datasets, their generalization capability is usually constrained when addressing complex scenes and fine-grained class categories.

To overcome these limitations, Convolutional Neural Networks (CNNs), including 1D-, 2D-, and 3D-CNNs, have been extensively deployed to learn spectral, spatial, and spectral–spatial features [16–18]. In general, 1D-CNNs focus on spectral signatures extraction, 2D-CNNs mainly capture spatial texture information, and 3D-CNNs jointly exploit spectral and spatial correlations. These methods have significantly improved HSIC performance compared with traditional machine learning approaches. However, CNNs are inherently limited by local receptive fields, rendering them less effective at modeling long-range dependencies, particularly in the scene with large homogeneous regions or intricate contextual structures.

To deal with the issue of local context, Transformer architectures have been introduced, leveraging self-attention mechanisms to capture global correlations [19–21]. By establishing pairwise relationships among tokens, Transformer-based models can effectively model long-range contextual dependencies and have achieved promising performance in HSIC. Nevertheless, their quadratic complexity, $O(N^2)$, results in high computational and memory costs for hyperspectral long-sequence data, which limits their efficiency and scalability in high-resolution hyperspectral classification tasks.

Recently, state space models (SSMs)-based methods, particularly Mamba, have revolutionized long-sequence modeling by achieving linear complexity, $O(N)$, while maintaining global modeling capabilities [22].Given the sequential nature of HSI, Mamba-based models are well-suited for HSIC. Several studies have integrating Mamba with convolutions, spatial–spectral attention, or dual-path structures to leverage local and global features synergistically [23–27]. These methods demonstrate that Mamba can serve as an efficient alternative to self-attention for large-scale sequence modeling [28, 29]. Nevertheless, the majority of existing Mamba-based methods are mainly designed for general land-cover classification, and their effectiveness for fine-grained tree species classification remains insufficiently explored.

Practical deployment of hyperspectral tree species classification faces several persistent challenges. First, closely related species often exhibit subtle spectral differences and similar spatial textures, making fine-grained discrimination extremely difficult. Second, existing methods usually lack explicit and efficient interaction between spectral and spatial features, leading to insufficient utilization of their complementarity [30]. Third, patch-based inference may introduce redundant computation and fail to fully exploit global contextual information [31]. In addition, ultra-high-resolution UAV-borne hyperspectral datasets for realistic fine-grained tree species classification remain limited, which also restricts the development and evaluation of advanced methods in this field [32,33].

To address these gaps, this study constructs a self-collected, manually annotated hyperspectral tree species dataset, named ZJK, and proposes a full-image pixel-wise classification network, MambaHSINet. The proposed network adopts a dual-branch architecture to separately extract spectral and spatial features, integrates channel attention module to enhance informative bands, and employs

 doi:10.20944/preprints202605.1614.v1

a bidirectional Mamba module for efficient global contextual modeling. Unlike patch-based methods, MambaHSINet avoids repeated cropping and sliding-window inference while maximizing the utilization of global context. In this way, the proposed method aims to achieve a superior balance between classification accuracy and computational efficiency for practical hyperspectral tree species classification tasks.

The main contributions of this study are summarized as follows:

1. A novel bidirectional Mamba dual-branch network (MambaHSINet) is proposed, explicitly decoupling spectral and spatial feature extraction to enhance discriminative representation for tree species with subtle spectral differences and similar spatial structures, while a bidirectional Mamba module is introduced to capture long-range spectral–spatial dependencies with linear complexity O(N).
2. A full-image pixel-wise inference paradigm is developed, fundamentally eliminating patch cropping and sliding-window redundancy to achieve a favorable balance among classification accuracy, inference efficiency, and model complexity.
3. A new UAV-borne hyperspectral tree species dataset (ZJK) with fine-grained manual annotations, publicly released as the first benchmark specifically designed for realistic fine-grained tree species classification under complex forest canopies.

The remainder of this paper is organized as follows. Section 2 introduces the related foundations. Section 3 describes the proposed MambaHSINet. Section 4 presents the experiments and analysis. Section 5 provides a detailed discussion of the experimental results and analyzed phenomena. Finally, Section 6 concludes this paper.

2. Related Work

This section provides a brief overview of algorithms related to Mamba. Regarding Mamba, we first introduce the State-Space Models (SSMs). Subsequently, Structured SSMs (S4) are introduced through discretization, followed by the addition of a selective mechanism to form S6. Finally, the Mamba model is constructed based on S6.

2.1. State-Space Models (SSMs)

State-space models (SSMs) are a class of systems that relate an input function $x(t) \in \mathbb{R}^L$ to an output response $y(t) \in \mathbb{R}^L$ through a set of latent state variables $\xi(t) \in \mathbb{R}^N$. Inspired by modern control theory, these models have recently been reintroduced to deep learning as an efficient alternative for sequence modeling. Generally, a continuous-time SSM is formulated as a linear first-order differential equation:

$$\dot{\xi}(t) = \mathbf{A}\xi(t) + \mathbf{B}x(t) \tag{1}$$

$$y(t) = \mathbf{C}\xi(t) + \mathbf{D}x(t) \tag{2}$$

where $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the evolution matrix that governs the system dynamics, $\mathbf{B} \in \mathbb{R}^{N\times L}$ and $\mathbf{C} \in \mathbb{R}^{L\times N}$ are the projection matrices, and $\mathbf{D} \in \mathbb{R}^{L\times L}$ represents the direct feedthrough component (usually skipped or handled as a residual connection in practical implementations). In hyperspectral image processing, this framework provides a rigorous mathematical basis for characterizing the spectral evolution of different tree species. By treating contiguous spectral bands as a continuous sequence, SSMs can capture the complex physical and biochemical signatures of forest canopies with linear computational complexity.

 doi:10.20944/preprints202605.1614.v1

2.2. Structured SSMs (S4)

To facilitate the processing of discrete digital HSI data, the continuous system must be transformed into a discrete-time representation. By introducing a timescale parameter $\Delta \in \mathbb{R}^+$, the continuous matrices $(\mathbf{A}, \mathbf{B})$ are discretized into $(\overline{\mathbf{A}}, \overline{\mathbf{B}})$ via the zero-order hold (ZOH) mechanism:

$$\overline{\mathbf{A}} = \exp(\Delta \mathbf{A}) \tag{3}$$

$$\overline{\mathbf{B}} = (\Delta \mathbf{A})^{-1}(\exp(\Delta \mathbf{A}) - \mathbf{I}) \cdot \Delta \mathbf{B} \tag{4}$$

The Structured SSM (S4) incorporates a specifically structured transition matrix (e.g., HiPPO initialization) to mitigate the vanishing gradient problem in long-range modeling. For HSI cubes with hundreds of spectral bands, S4 enables the extraction of global dependencies with $O(L \log L)$ computational complexity. This provides a significant efficiency advantage over Transformer-based architectures, which suffer from a quadratic $O(L^2)$ memory footprint when dealing with high-dimensional remote sensing data.

2.3. Selective Structured SSMs (S6)

Despite the efficiency of S4, its time-invariant nature means the transition parameters remain constant regardless of the input content, which limits its adaptability to complex land-cover types. The selective structured SSM (S6), as the core of the Mamba architecture, overcomes this by parameterizing $\mathbf{B}$, $\mathbf{C}$, and Δ as functions of the input x:

$$\mathbf{B} = \text{Linear}_B(x), \quad \mathbf{C} = \text{Linear}_C(x), \quad \Delta = \text{Softplus}(\text{Parameter} + \text{Linear}_\Delta(x)) \tag{5}$$

This content-aware selection mechanism allows MambaHSINet to adaptively focus on discriminative spectral features (e.g., the red-edge position or chlorophyll absorption pits) while suppressing redundant noise. Such a property is particularly vital for fine-grained tree species classification, where inter-class spectral differences are often subtle and localized.

2.4. Mamba Block and Bidirectional Modeling

The Mamba block integrates the S6 engine with gated linear units and depth-wise convolutions to exploit both local and global features. However, traditional Mamba blocks are designed for causal sequences, which is suboptimal for HSI data where spatial-spectral correlations are inherently non-causal and multi-directional. In MambaHSINet, we implement a bidirectional modeling strategy to scan the spatial-spectral tokens from both forward and backward directions. This approach ensures that the contextual information of the forest canopy is captured holistically. By merging the efficiency of SSMs with bidirectional feature interaction, the Mamba block facilitates a comprehensive representation of HSI data, leading to enhanced classification performance in complex forest ecosystems.

3. Proposed Method

This section introduces the proposed MambaHSINet. We first present the overall architecture and then describe its main components in detail, including the spectral branch, spatial branch, dual-branch feature fusion, bidirectional Mamba global modeling module, classification head, and training strategy. The proposed method takes the full hyperspectral image as input and jointly performs local spectral–spatial feature extraction and global contextual dependency modeling with relatively low computational cost.

3.1. Overall Framework

As shown in Figure 1, the proposed MambaHSINet mainly consists of five parts: 1) input preprocessing, 2) spectral branch, 3) spatial branch, 4) feature fusion and bidirectional Mamba global modeling, and 5) classification head. Unlike traditional patch-based classification methods, the proposed framework directly takes the whole hyperspectral image as input and outputs pixel-wise

 doi:10.20944/preprints202605.1614.v1

classification results, thereby avoiding repeated cropping and redundant inference while making better use of full-image contextual information [34].

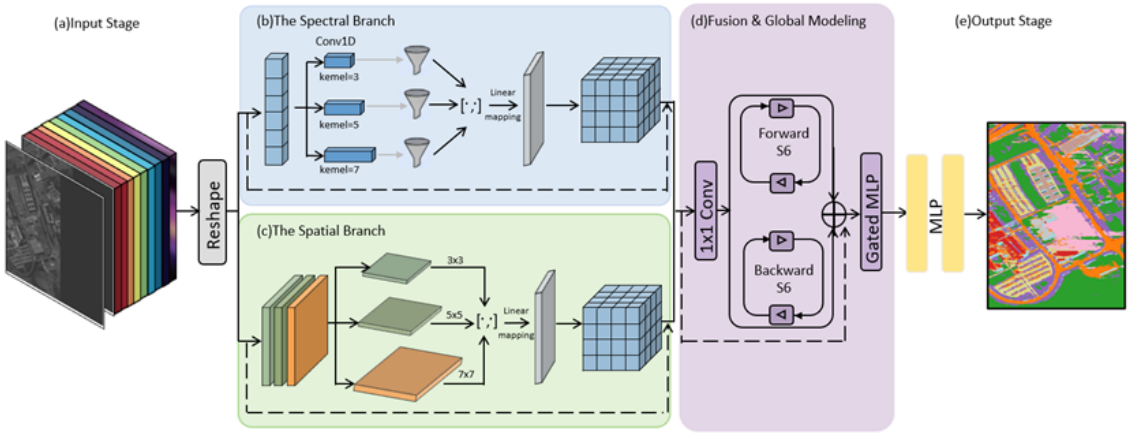

Figure 1. Overall Framework of the Proposed Method

Let the input hyperspectral image be denoted as

$$\mathbf{X} \in \mathbb{R}^{H \times W \times B}, \tag{6}$$

where H and W denote the spatial height and width, respectively, and B is the number of spectral bands. First, layer normalization is applied along the spectral dimension to obtain the normalized feature

$$\mathbf{X}_n = \mathrm{LN}(\mathbf{X}). \tag{7}$$

Then, $\mathbf{X}_n$ is fed into the spectral branch and the spatial branch to extract complementary spectral-domain and spatial-structure representations. The two features are concatenated and fused, rearranged into a sequence, and then sent into a bidirectional Mamba module for long-range contextual modeling. Finally, a classification head is used to output the category prediction for each pixel.

The whole framework can be formulated as

$$\mathbf{Y} = F_{\mathrm{cls}}\Big(F_m(F_{\mathrm{fus}}([F_{\mathrm{spe}}(\mathbf{X}_n), F_{\mathrm{spa}}(\mathbf{X}_n)]))\Big), \tag{8}$$

where $F_{\mathrm{spe}}(\cdot)$, $F_{\mathrm{spa}}(\cdot)$, $F_{\mathrm{fus}}(\cdot)$, $F_m(\cdot)$, and $F_{\mathrm{cls}}(\cdot)$ denote the spectral branch, spatial branch, fusion module, bidirectional Mamba module, and classification head, respectively; $[\cdot,\cdot]$ denotes channel-wise concatenation; and $\mathbf{Y}$ is the final pixel-wise classification output.

3.2. Spectral Branch

Each pixel in a hyperspectral image (HSI) possesses a high-dimensional spectral response, where subtle but decisive inter-class variances are often embedded within the continuous spectral curves. To effectively mine these intra-pixel correlations while mitigating the risk of feature degradation, we design a lightweight multi-scale spectral branch with an explicit emphasis on original information preservation.

 doi:10.20944/preprints202605.1614.v1

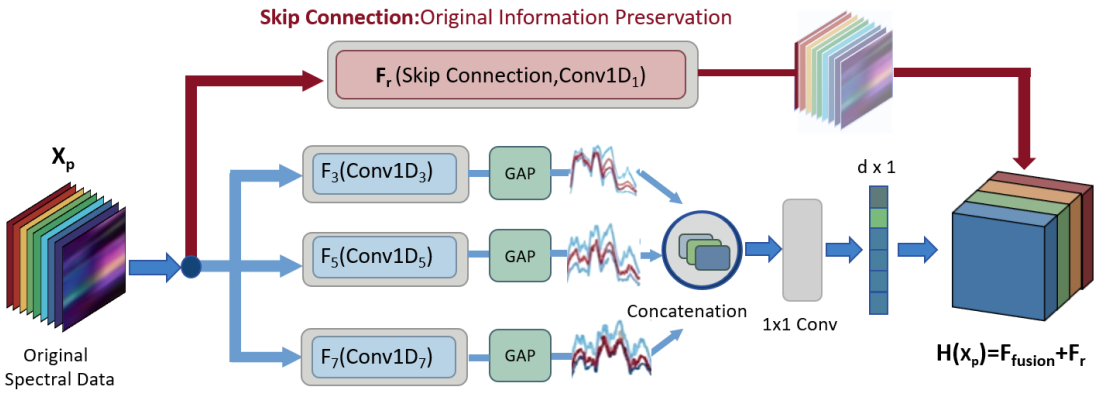

Figure 2. Spectral Branch Structure Diagram

Given the input feature $\mathbf{X}_n \in \mathbb{R}^{H\times W\times B}$, it is first reshaped into a pixel-wise spectral sequence $\mathbf{X}_p \in \mathbb{R}^{(HW)\times 1\times B}$, where B denotes the number of spectral bands. To capture local spectral dependencies at various scales, three parallel 1D convolution branches with kernel sizes of 3, 5, and 7 are employed [35]. Crucially, a 1×1 residual projection branch is integrated as a core component to serve as an "identity-like" anchor, ensuring that the intrinsic spectral signatures are not smoothed out or lost during multi-scale feature extraction. These branches are formally expressed as:

$$\mathbf{F}_3 = \text{Conv1D}_3(\mathbf{X}_p), \tag{9}$$

$$\mathbf{F}_5 = \text{Conv1D}_5(\mathbf{X}_p), \tag{10}$$

$$\mathbf{F}_7 = \text{Conv1D}_7(\mathbf{X}_p), \tag{11}$$

$$\mathbf{F}_r = \text{Conv1D}_1(\mathbf{X}_p), \tag{12}$$

where $\mathbf{F}_r$ represents the residual path that directly propagates the primary spectral response.

To maintain computational efficiency, global average pooling is applied along the spectral dimension of each branch to generate compact descriptors. The resulting features are concatenated and projected onto a unified d-dimensional space via linear mapping, yielding $\mathbf{X}_{\text{spe}}$, which is then rearranged back to $\mathbb{R}^{H\times W\times d}$. This architecture ensures that the model can exploit high-level multi-scale correlations without sacrificing the fidelity of the raw spectral data, which is vital for distinguishing tree species with highly overlapping spectral profiles.

3.3. Spatial Branch

Besides spectral information, hyperspectral images also contain strong spatial correlations among neighboring pixels. In particular, local textures, boundary structures, and neighborhood patterns are important for improving discriminability in complex scenes. Therefore, a multi-scale spatial branch is further introduced to extract local spatial contextual features.

First, the normalized input $\mathbf{X}_n$ is converted into a convolution-friendly tensor:

$$\mathbf{X}_s \in \mathbb{R}^{1\times B\times H\times W}. \tag{13}$$

Then, three parallel 2D convolution branches with kernel sizes of 3×3, 5×5, and 7×7 are adopted to extract local spatial information under different receptive fields. Meanwhile, a 1×1 convolution residual projection branch is introduced to complement the original spatial representation. The corresponding computations are

$$\mathbf{S}_3 = \text{Conv2D}_{3\times 3}(\mathbf{X}_s), \tag{14}$$

 doi:10.20944/preprints202605.1614.v1

$$\mathbf{S}_5 = \mathrm{Conv2D}_{5\times5}(\mathbf{X}_s), \tag{15}$$

$$\mathbf{S}_7 = \mathrm{Conv2D}_{7\times7}(\mathbf{X}_s), \tag{16}$$

$$\mathbf{S}_r = \mathrm{Conv2D}_{1\times1}(\mathbf{X}_s). \tag{17}$$

After that, the four feature maps are concatenated along the channel dimension and projected to the unified feature dimension d through a 1×1 convolution, resulting in the spatial feature $\mathbf{X}_{\mathrm{spa}}$. The output is finally rearranged into a tensor of size $H \times W \times d$.

Through this design, the spatial branch can sufficiently capture local texture patterns, edge structures, and neighborhood information at multiple scales, while the residual path helps preserve shallow spatial priors and improves training stability.

3.4. Dual-Branch Feature Fusion

The spectral branch focuses on intra-pixel spectral correlations, whereas the spatial branch emphasizes local structural patterns in neighboring regions. Since the two branches are highly complementary, an effective feature fusion strategy is required.

First, the two branch features are concatenated along the channel dimension:

$$\mathbf{X}_{\mathrm{cat}} = [\mathbf{X}_{\mathrm{spe}}, \mathbf{X}_{\mathrm{spa}}] \in \mathbb{R}^{H\times W\times 2d}. \tag{18}$$

Then, $\mathbf{X}_{\mathrm{cat}}$ is transformed into a convolutional tensor and passed through a 1×1 convolution for cross-channel interaction and dimensionality reduction, yielding the fused feature

$$\mathbf{X}_{\mathrm{fus}} \in \mathbb{R}^{H\times W\times d}. \tag{19}$$

Compared with simple weighted summation, this fusion strategy enables more sufficient semantic interaction between spectral and spatial features at relatively low cost. On the one hand, the 1×1 convolution explicitly models channel-wise correlations across branches; on the other hand, the channel number is compressed back to a unified dimension, which is beneficial for the subsequent global sequence modeling.

3.5. Bidirectional Mamba Global Modeling Module

Although multi-scale convolution is effective for local spectral–spatial feature extraction, its receptive field is still limited by the kernel size and thus insufficient for modeling long-range contextual dependencies. To further enhance the representation of full-image semantic relationships, a bidirectional Mamba module is introduced after feature fusion.

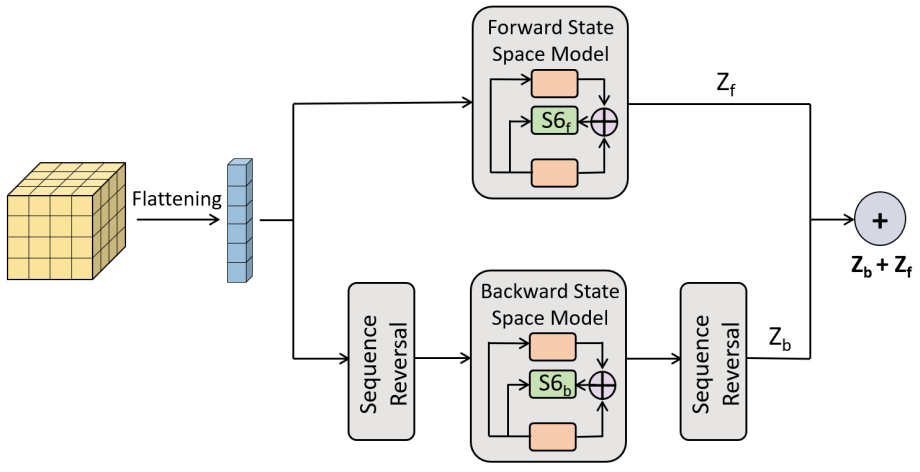

Figure 3. Bidirectional Mamba Modeling Module

 doi:10.20944/preprints202605.1614.v1

Given the fused feature

$$\mathbf{X}_{\text{fus}} \in \mathbb{R}^{H\times W\times d}, \tag{20}$$

it is first flattened into a sequence of length

$$L = H \times W, \tag{21}$$

resulting in

$$\mathbf{Z} \in \mathbb{R}^{1\times L\times d}. \tag{22}$$

To capture both forward and backward contextual information, two independent S6 state space modules are used for forward and reverse scanning [36]. The forward feature is defined as

$$\mathbf{Z}_f = \text{S6}_f(\mathbf{Z}), \tag{23}$$

and the backward feature is defined as

$$\mathbf{Z}_b = \text{Flip}(\text{S6}_b(\text{Flip}(\mathbf{Z}))), \tag{24}$$

where $\text{Flip}(\cdot)$ denotes sequence reversal. The two directional features are then fused by summation and combined with normalization, nonlinear activation, dropout, and residual connection to obtain the final globally enhanced feature:

$$\mathbf{Z}_m = \mathbf{Z} + G(\mathbf{Z}_f + \mathbf{Z}_b), \tag{25}$$

where $G(\cdot)$ denotes a nonlinear transformation composed of Layer Normalization, GELU activation, and Dropout.

The main advantages of this module are threefold. First, the state space model captures long-range dependencies with linear complexity. Second, the bidirectional scanning alleviates the information bias caused by unidirectional sequence modeling. Third, the residual connection and pre-normalization strategy improve both training stability and feature representation capability. Therefore, this module enables more complete pixel-wise contextual modeling over the whole image.

3.6. Classification Head

After obtaining the globally enhanced feature $\mathbf{Z}_m$, a lightweight multilayer perceptron is used as the classification head to map each feature vector to its category. The classification head consists of two fully connected layers with Layer Normalization, GELU activation, and Dropout inserted between them to improve nonlinear representation ability and suppress overfitting. It can be written as

$$\mathbf{P} = \text{FC}_2(\text{Drop}(\text{GELU}(\text{LN}(\text{FC}_1(\mathbf{Z}_m))))), \tag{26}$$

where $\mathbf{P}$ denotes the sequence-form prediction result. Finally, $\mathbf{P}$ is rearranged back to the image space to obtain the pixel-wise classification output:

$$\mathbf{Y} \in \mathbb{R}^{H\times W\times C}, \tag{27}$$

where C is the number of categories.

3.7. Loss Function and Training Strategy

Since the proposed framework adopts a full-image pixel-wise classification paradigm, the supervision loss is computed only on labeled pixels during training. Let

$$\mathbf{M} \in \{0,1\}^{H\times W} \tag{28}$$

 doi:10.20944/preprints202605.1614.v1

be the training mask, where only pixels satisfying $M(i,j) = 1$ participate in loss computation. For the pixel at location (i,j), let $\hat{\mathbf{y}}_{ij}$ denote the predicted class probability and y_{ij} denote the ground-truth label. The training objective is defined as a weighted cross-entropy loss:

$$\mathcal{L} = -\sum_{(i,j)\in\Omega} w_{y_{ij}} \log \hat{y}_{ij}^{(y_{ij})}, \tag{29}$$

where Ω is the set of labeled training pixels and $w_{y_{ij}}$ is the class weight used to alleviate class imbalance [37].

The model is trained in an end-to-end manner using gradient-based optimization. Gradient clipping is further employed to suppress gradient explosion and improve training stability. During inference, the model outputs the full-image prediction in a single forward pass, thereby avoiding the repeated computation introduced by traditional patch-based sliding-window inference.

4. Experiments

4.1. Datasets

This study utilizes the public Pavia Centre dataset alongside two subsets derived from a self-constructed Zhangjiakou (ZJK) tree-species collection, denoted as HuaiL_1 (HL_1) and HuaiL_2 (HL_2).

Table 1. Summary of the datasets used in this study.

No.	Dataset	Location	Size	Classes	Bands
1	Pavia Centre	Italy	1096×1096	9	102
2	HL_1	China	3465×130	7	164
3	HL_2	China	1240×265	6	164

4.1.1. Pavia Centre

The Pavia Centre dataset was acquired over the central urban area of Pavia, Italy [38], using the Reflective Optics System Imaging Spectrometer (ROSIS) sensor. It consists of a hyperspectral image with a spatial size of 1096×1096 pixels and 102 valid spectral bands. According to land-cover categories, the image was annotated into 9 classes, with a total of 7456 labeled samples. The main classes include typical urban objects such as roads, buildings, shadows, bare soil, and low vegetation.

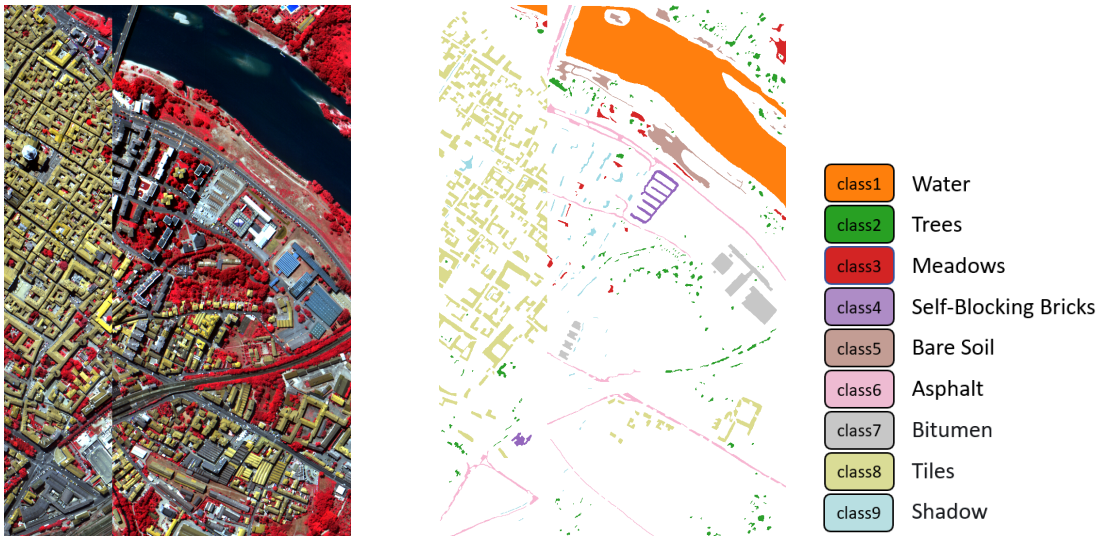

Figure 4. Pavia dataset: (a)RGB image, (b)Ground truth map and color legends.

 doi:10.20944/preprints202605.1614.v1

4.2. Zhangjiakou (ZJK)

The ZJK dataset is a high-spatial and high-spectral (H^2) benchmark dataset established by our team, specifically designed for fine-grained tree species identification in complex urban-fringe and "forest-agriculture" mosaic ecosystems.

4.2.1. Study Area Overview and Environmental Background

The ZJK dataset was collected in Donghuayuan Town, Huailai County, Zhangjiakou City, Hebei Province, with geographical coordinates of 40°17′ N, 115°53′ E and an elevation of 470–480 meters. The region is situated within the Huailai Basin and is characterized by a typical temperate continental monsoon climate, featuring four distinct seasons and concurrent rain and heat. The total area of the study site is approximately 400 mu, with flat micro-topography. The land cover encompasses artificial forests, nurseries, and seasonal agricultural crops, presenting typical transitional landscape characteristics. This provides a complex and diverse ecological background for tree species identification.

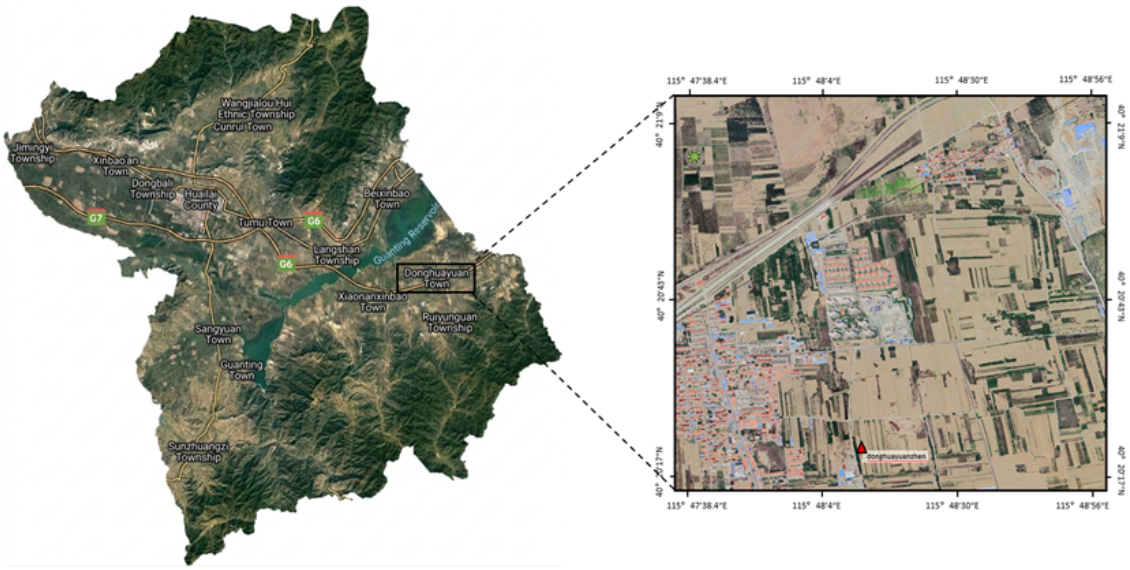

Figure 5. Geographical location and environmental setting of the study area: (left) topographic map of Huailai County, Hebei Province; (right) schematic diagram of the ZJK dataset site in Donghuayuan Town.

4.2.2. Spatial-Spectral Data Acquisition Platform

The imagery was acquired in October 2025 using an ultra-high-resolution multi-source remote sensing platform. A DJI Matrice 350 RTK UAV, equipped with an X20P-LIR integrated multi-source imaging system, served as the flight platform. To minimize shadow interference and ensure a high signal-to-noise ratio (SNR), flight missions were conducted between 11:00 and 14:00 under clear, cloudless weather conditions.

The UAV flight altitude was set at 80 m. The acquired hyperspectral imagery covers a spectral range of 350–1000 nm (consisting of 164 bands) with an image size of 940 × 475 pixels and a spatial resolution of approximately 0.027 m. Simultaneously, RGB imagery with 26 million pixels and thermal infrared (TIR) imagery with a resolution of 640 × 512 pixels were also obtained.

 doi:10.20944/preprints202605.1614.v1

Figure 6. The data acquisition platform and flight planning: (a) X20P-LIR system mounted on the DJI M300/350 RTK multi-rotor UAV; (b) Predefined flight route map.

4.2.3. Systematic Preprocessing and Ground Truth Construction

Following professional remote sensing standards, the raw hyperspectral data underwent a streamlined preprocessing and validation pipeline:

- **(1) Radiometric and Geometric Correction**: Raw Digital Number (DN) values were converted to surface reflectance using laboratory calibration parameters and the empirical line method with reference panels. Geometric distortions were corrected by integrating high-precision GNSS/IMU data with a digital elevation model (DEM) to achieve sub-decimeter spatial registration, ensuring the geometric fidelity of the extracted spectral curves.
- **(2) Collaborative Annotation and Sample Generation**: To manage large-scale imagery, a spatial partitioning strategy was employed where multiple annotators performed pixel-level delineation using Labelme. The scattered JSON annotation files were then integrated and processed via a self-developed data formatting tool. This tool performed coordinate transformation, topological stitching, and automated cropping, effectively transforming raw polygons into a seamless, standardized, and AI-ready dataset.
- **(3) Field-to-Image Verification**: To ensure taxonomic integrity, systematic plots were established for concurrent botanical surveys (identifying species, tree height, and DBH). These ground records were cross-validated with the annotated polygons to exclude ambiguous samples caused by shadows or spectral mixing, guaranteeing the botanical reliability of the final ground truth labels for fine-grained classification.

4.2.4. Representative Sub-scenes: HL_1 and HL_2

The complete ZJK dataset currently covers 20 land-cover categories, with various tree species serving as the core components, encompassing a total of 4,857 field-surveyed annotation samples. To rigorously verify the performance of MambaHSINet in fine-grained tree species identification, two representative sub-scenes, HuaiL_1 (HL_1) and HuaiL_2 (HL_2), were strategically selected from the full dataset as benchmark testing scenarios:

1. **HL_1 dataset**: This scene represents a complex urban-fringe "forest-building" composite landscape, containing 7 land-cover categories with a total of 450,450 labeled pixels. It is characterized by the interlocking distribution of typical tree species, such as *Malus spectabilis* and *Pinus tabuliformis*, with artificial structures. Detailed information regarding the number of categories and pixel statistics is provided in Table 2.

 doi:10.20944/preprints202605.1614.v1

Table 2. Land cover classes and sample statistics for each type in the HL_1 dataset.

No.	Class Name	Pixel count	Percentage(%)	Count
C1	Crabapple	56,589	12.56	2
C2	Chinese pine	20,730	4.60	4
C3	Grassland	55,988	12.43	5
C4	Mono maple	6,679	1.48	3
C5	Corn stubble	42,963	9.54	3
C6	Corn field	67,978	15.09	3
C7	Chinese scholar tree	48,344	10.73	2
BG	Background	151,179	33.56	/
Total	/	450,450	100.00	22

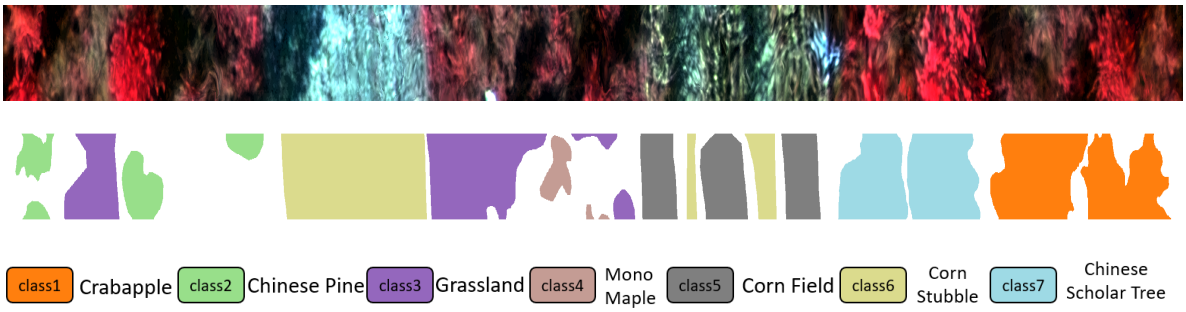

Figure 7. HL_1 dataset. (left) RGB image. (right) Ground truth map and color legends.

2. **HL_2 dataset**: This scene captures a typical "artificial forest-farmland" transition zone, consisting of 6 major categories with 328,600 labeled pixels. The core difficulty of HL_2 lies in the extreme spectral similarity among taxonomically related coniferous species, particularly *Pinus tabuliformis*, *Picea asperata*, and *Platycladus orientalis*. Detailed information regarding the number of categories and pixel statistics is provided in Table 3.

Table 3. Land cover classes and sample statistics for each type in the HL_2 dataset.

No.	Class Name	Pixel count	Percentage(%)	Count
C1	Crabapple	37,446	11.40	5
C2	Populus tomentosa	30,910	9.41	16
C3	Grassland	42,869	13.05	4
C4	Leaf litter	83,374	25.37	7
C5	Corn field	44,454	13.53	1
C6	Road	20,651	6.28	1
BG	Background	68,896	20.97	/
Total	/	328,600	100.00	34

 doi:10.20944/preprints202605.1614.v1

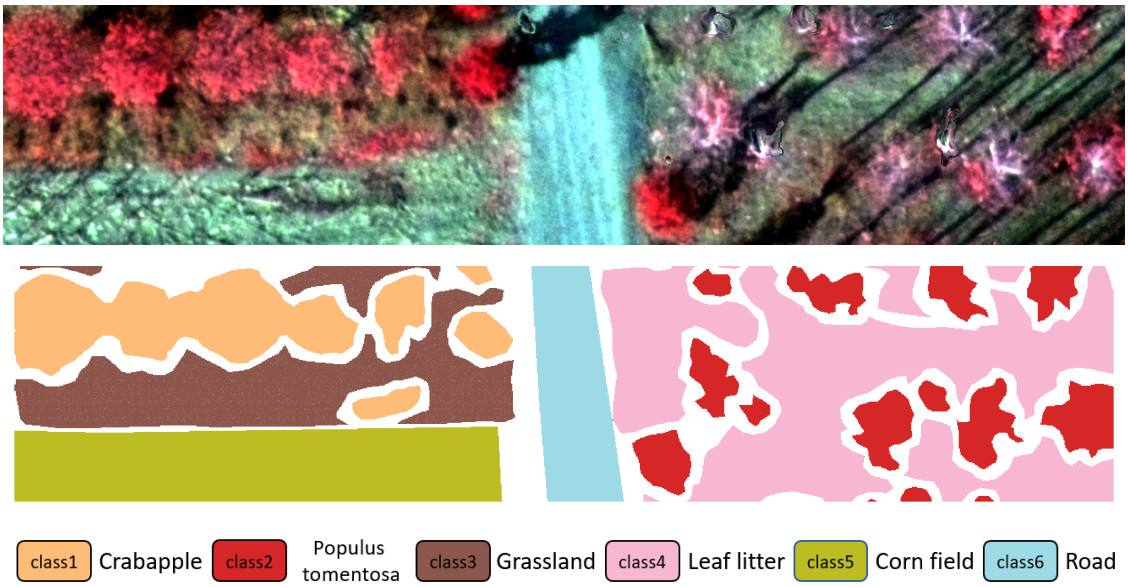

Figure 8. HL_2 dataset. (left) RGB image. (right) Ground truth map and color legends.

4.2.5. Spectral Characteristic Analysis

The mean spectral reflectance curves for the *HL_1* and *HL_2* datasets are illustrated in Figure 9 and Figure 10, respectively.

In the *HL_1* scene, the land-cover categories exhibit distinctive spectral signatures. The prominent "red edge" effect and near-infrared plateau features provide a solid physical foundation for the fine-grained identification of different tree species, such as *Malus spectabilis* and *Pinus tabuliformis*.

In contrast, the *HL_2* scene presents a more significant classification challenge. As shown in Figure 10, the spectral profiles of *Crabapple* (C1), *Populus tomentosa* (C2), and *Leaf litter* (C4) are highly bundled and overlapping across the entire range of 164 bands, particularly within the 100–140 band interval. This extreme intra-class spectral similarity—often referred to as the "same spectrum, different objects" phenomenon—constitutes the core difficulty of this dataset.

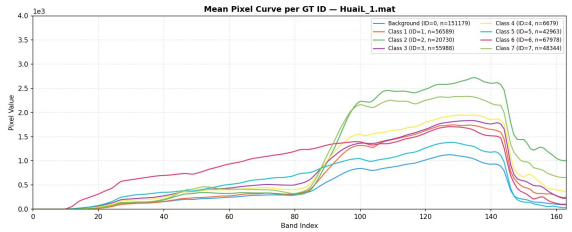

Figure 9. Spectral curves of HL_1

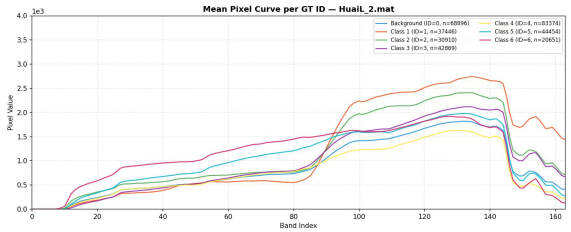

Figure 10. Spectral curves of HL_2

4.3. Experimental Settings

1. **Evaluation metrics:** Following existing HSIC studies, overall accuracy (OA), average accuracy (AA), and the Kappa coefficient were adopted as evaluation metrics.
2. **Compared methods:** To verify the effectiveness of the proposed novel network architecture combining a dual-branch structure with bidirectional Mamba, four categories of HSIC methods were selected for comparison, including the traditional machine learning method SVM [12], CNN-based methods (Tri-CNN [35] and HybridSN [39]), the Transformer-based method SS-FTT (Spectral-Spatial Feature Tokenization Transformer) [21], and the Mamba-based methods MambaHSI [25], 3DSS-Mamba [23], and SS-Mamba [40].

    a) **SVM:** SVM is based on structural risk minimization and uses kernel functions to handle high-dimensional nonlinear data for HSIC.

 doi:10.20944/preprints202605.1614.v1

b) **Tri-CNN:** Tri-CNN uses a multi-branch structure for feature fusion and can effectively extract multi-scale and multimodal information from images.
c) **HybridSN:** HybridSN combines 2D convolution and 3D convolution to jointly model spatial and spectral information, enabling deep spectral–spatial feature extraction.
d) **SSFTT:** SSFTT tokenizes the three-dimensional spectral–spatial cube and uses a dual-branch encoder to achieve end-to-end global joint feature learning.
e) **MambaHSI:** MambaHSI alternately stacks spatial and spectral Mamba modules to perform decoupled modeling and interactive fusion of long-range dependencies in hyperspectral images.
f) **3DSS-Mamba:** 3DSS-Mamba constructs three-dimensional hyperspectral data as long sequences and performs unified spectral–spatial global modeling through bidirectional Mamba blocks.
g) **SS-Mamba:** SS-Mamba introduces a spectral–spatial learning framework based on the Mamba model and performs feature extraction through a token generation mechanism and a feature enhancement module.

3. **Implementation details:** All experiments were conducted on the PyTorch platform. The hardware configuration included an NVIDIA GeForce RTX 4060 Ti GPU, an x64 Intel Core i7-14700KF CPU, and 128 GB RAM. In the experiments, the training set and validation set each accounted for 10% of the total samples, while the remaining samples were used for testing. Since the proposed dual-branch architecture with bidirectional Mamba takes the entire image as input, the batch size was set to 1. The number of training epochs was set to 100. To reduce randomness and improve the reliability and stability of the results, 10 different random seeds were used for repeated experiments, and the final results were reported as the average of the 10 runs.

4.3.1. Parameter Analysis

To evaluate the impact of key hyperparameters on classification performance, we conducted extensive sensitivity experiments using the Kappa coefficient across the *HL_1*, *HL_2*, and *Pavia* datasets. The analysis focuses on three critical parameters that govern the model's multi-scale feature extraction and global dependency modeling capabilities.

- **Intermediate Channels of Spectral Branch:** This parameter determines the channel capacity during the multi-scale spectral feature extraction stage. In the spectral branch, three multi-scale 1D convolutions (with kernel sizes of 3, 5, and 7) first map the spectral vector of each pixel to an intermediate-dimensional representation. Experimental results indicate that a limited number of intermediate channels restricts the model's ability to capture subtle spectral variations, as the advantages of multi-scale receptive fields cannot be fully exploited. While increasing this value enhances the expressive power of spectral features and improves classification accuracy, excessively high channel counts introduce redundant parameters, leading to higher overfitting risks and increased computational overhead. Thus, an appropriate balance between feature richness and model complexity is essential for optimal performance.
- **Intermediate Channels of Spatial Branch:** Symmetrical to the spectral branch, this parameter controls the channel dimensionality of the three multi-scale 2D convolutions (3×3, 5×5, and 7×7). When the number of intermediate channels is small, the spatial context information extracted by each branch remains constrained, rendering the fused spatial features insufficient for precise land-cover boundary delineation. Increasing this value improves the diversity and discriminative power of spatial features. However, similar to the spectral branch, an excessively large number of intermediate channels significantly elevates the computational load of 2D convolutions with diminishing marginal returns. Therefore, the selected configuration achieves a reasonable compromise between spatial modeling capability and computational efficiency.

 doi:10.20944/preprints202605.1614.v1

- **State Dimension of Mamba Layer:** The bidirectional state-space dimension directly determines the capacity for global dependency modeling within the flattened spatial-spectral joint sequence. This dimension governs the scale of the hidden state and the precision of modeling long-range dependencies. An insufficient state dimension prevents the effective transmission of distant correlations within the sequence, limiting the model's global perspective. Conversely, increasing this dimension significantly enhances the efficiency of utilizing global contextual information. Under the optimal state dimension setting, MambaHSINet achieved peak Kappa values of 99.39%, 97.98%, and 98.65% on the three datasets, respectively. However, an excessively large state dimension may lead to state-space explosion, resulting in increased memory consumption and slower training convergence. Thus, the final selection balances global modeling capability with overall training stability.

4.3.2. Ablation Study

To verify the effectiveness of the key components in MambaHSINet, we conducted comprehensive ablation experiments on three datasets. The quantitative results, including OA,AA and Kappa, are summarized in Table 4.

Table 4. Ablation study results of the key components in MambaHSINet on three datasets (%). The best results for each dataset are highlighted in **bold**.

Components			HL_1 (%)			HL_2 (%)			Pavia (%)		
Spatial	Spectral	Mamba	OA	AA	Kx100	OA	AA	Kx100	OA	AA	Kx100
×	✓	✓	96.91	97.43	96.27	94.61	95.36	93.28	98.40	96.13	97.74
✓	×	✓	99.44	99.52	99.32	98.39	98.58	97.98	98.82	96.82	98.33
✓	✓	×	95.68	95.81	94.78	96.03	96.48	95.05	97.61	92.93	96.61
✓	✓	✓	**99.50**	**99.62**	**99.39**	**98.39**	**98.58**	**97.98**	**99.05**	**97.16**	**98.65**

A. **Effectiveness of Dual-Branch Structure.** The complementarity between the spatial and spectral branches is fundamental for high-precision tree species classification. As shown in Table 4, removing either branch leads to a noticeable performance decline. Specifically, relying solely on spectral features results in OA values of 96.91%, 94.61%, and 98.40% on the three datasets. After incorporating the spatial branch, the OA of the complete model increases to 99.50%, 98.39%, and 99.05%, respectively. This improvement indicates that spatial structural information provides vital complementary cues that compensate for the limitations of spectral-only modeling in complex forest scenes.

B. **Significance of Bidirectional Mamba Module.** The Mamba module serves as the core for global contextual modeling in our architecture. When the Mamba module is removed and the model relies exclusively on local convolutional features, the OA on the *HL_1* dataset drops significantly from 99.50% to 95.68%, and the Kappa on the *Pavia* dataset decreases from 98.65% to 96.61%. These results demonstrate that while local convolutions effectively capture fine-grained textures, the bidirectional Mamba module is essential for aggregating long-range dependencies and establishing global semantic representations.

C. **Synergy of the Full Configuration.** The complete MambaHSINet consistently achieves the highest metrics across all evaluated datasets. Notably, on the *HL_1* dataset, which contains 7 distinct tree species categories, our full model reaches a near-perfect OA of 99.50%. This validates that the synergy between dual-branch feature extraction and bidirectional Mamba modeling allows the network to effectively mitigate the "same object, different spectra" and "different objects, same spectrum" phenomena, ensuring superior robustness and discriminative capability in fine-grained classification tasks.

4.4. Comparison with State-of-the-Art Methods

To evaluate MambaHSINet, we compared it with traditional (SVM), CNN-based (Tri-CNN, HybridSN), Transformer-based (SSFTT), and SSM-based (MambaHSI, 3DSS-Mamba, SS-Mamba)

 doi:10.20944/preprints202605.1614.v1

methods. Quantitative results for the *HL_1*, *HL_2*, and *Pavia* datasets are summarized in Tables 5, 6, and 7.

Table 5. Classification performance (%) of SVM, Tri-CNN, HybridSN, SSFTT, MambaHSI, 3DSS-Mamba, SS-Mamba, and the proposed MambaHSINet on the HL_1 dataset.

Class	SVM	Tri-CNN	HybridSN	SSFTT	MambaHSI	3DSS-Mamba	SS-Mamba	Proposed
C1	96.82	98.69	99.22	97.13	93.79	76.96	98.77	**99.93**
C2	94.92	96.70	99.04	96.17	93.27	64.04	96.34	**99.70**
C3	95.55	96.35	98.90	97.97	88.87	83.01	97.18	**99.36**
C4	79.85	85.46	94.70	86.92	97.44	0.00	84.19	**99.32**
C5	97.89	98.24	**99.76**	94.78	93.12	86.17	98.66	99.51
C6	98.12	99.11	**99.31**	99.00	95.93	92.04	98.98	98.94
C7	94.79	98.06	99.48	97.88	91.11	52.25	97.73	**99.95**
OA	96.24	97.75	99.10	97.20	92.87	76.23	97.84	**99.50**
AA	94.01	96.09	98.63	95.69	93.36	64.92	95.98	**99.62**
Kappa	95.44	97.27	99.03	96.60	93.33	71.00	97.39	**99.39**

Table 6. Classification performance (%) of SVM, Tri-CNN, HybridSN, SSFTT, MambaHSI, 3DSS-Mamba, SS-Mamba, and the proposed MambaHSINet on the HL_2 dataset.

Class	SVM	Tri-CNN	HybridSN	SSFTT	MambaHSI	3DSS-Mamba	SS-Mamba	Proposed
C1	97.89	98.71	98.08	96.88	92.67	92.86	97.25	**98.94**
C2	91.53	94.16	96.59	92.21	86.25	69.12	91.58	**96.65**
C3	95.02	95.49	98.21	94.25	92.73	83.99	93.15	**98.64**
C4	96.76	97.58	**98.36**	93.95	89.26	79.58	96.53	97.44
C5	97.79	99.01	99.27	97.05	92.73	88.97	97.91	**99.89**
C6	97.91	99.39	99.05	97.65	96.90	82.08	97.77	**99.71**
OA	96.33	97.38	98.29	95.04	91.17	82.78	95.82	**98.39**
AA	96.18	97.39	98.26	95.33	91.76	82.76	95.70	**98.58**
Kappa	95.40	96.72	97.86	93.80	90.00	78.41	94.76	**97.98**

Table 7. Classification performance (%) of SVM, Tri-CNN, HybridSN, SSFTT, MambaHSI, 3DSS-Mamba, SS-Mamba, and the proposed MambaHSINet on the Pavia dataset.

Class	SVM	Tri-CNN	HybridSN	SSFTT	MambaHSI	3DSS-Mamba	SS-Mamba	Proposed
C1	99.99	**100.00**	**100.00**	99.90	99.55	99.96	**100.00**	99.99
C2	95.11	95.56	93.14	95.27	93.80	**99.22**	94.26	95.52
C3	89.07	93.91	71.88	**96.2**9	87.89	0.00	92.25	88.66
C4	72.90	88.37	78.54	91.50	**98.54**	0.53	88.51	98.12
C5	97.79	97.43	80.33	94.71	94.11	39.04	97.14	**99.19**
C6	93.24	**99.30**	96.87	98.49	97.10	68.41	97.49	96.10
C7	91.56	96.53	96.80	94.39	91.70	94.36	94.39	**97.38**
C8	99.37	**99.92**	99.65	99.27	99.19	98.45	99.88	99.91
C9	**99.97**	97.22	82.64	95.33	99.60	0.00	99.96	99.57
OA	97.98	99.03	97.01	98.58	98.12	88.72	98.74	**99.05**
AA	93.15	96.47	88.87	96.13	95.72	55.55	95.99	**97.16**
Kappa	97.13	98.62	95.75	97.98	94.38	83.55	98.21	**98.65**

Table 8 distinguishes MambaHSINet from existing Mamba-based SOTA. While methods like SS-Mamba and MambaHSI focus on local patch enhancement via SSM, MambaHSINet introduces a structural shift to full-image bidirectional modeling. By explicitly decoupling spectral-spatial branches, our architecture captures wide-range zonal dependencies and eliminates the sliding-window redundancy inherent in patch-based SSMs, ensuring a non-incremental advancement in both efficiency and global context awareness.

 doi:10.20944/preprints202605.1614.v1

Table 8. Architectural comparison between MambaHSINet and SOTA Mamba methods.

Model	Input Paradigm	Feature Extraction	Context Scope
SS-Mamba	Patch-based	Mixed Spectral-Spatial	Patch-limited
MambaHSI	Patch-based	Multi-scale Spatial	Local-scale
3DSS-Mamba	Patch-based	3D Conv + SSM	Local-scale
MambaHSINet	**Full-image**	**Decoupled Dual-branch**	**Full-image Global**

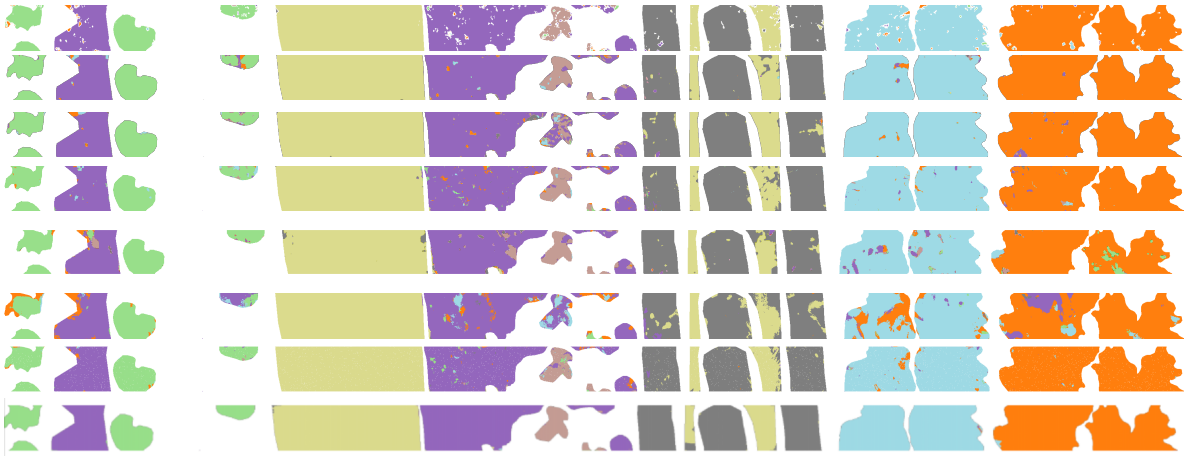

Figure 11. Overview of annotated images and classification performance of SVM, Tri-CNN, HybridSN, SSFTT, MambaHSI, 3DSS-Mamba,SS-Mamba, and the proposed MambaHSINet on the HuaiL_1 dataset.

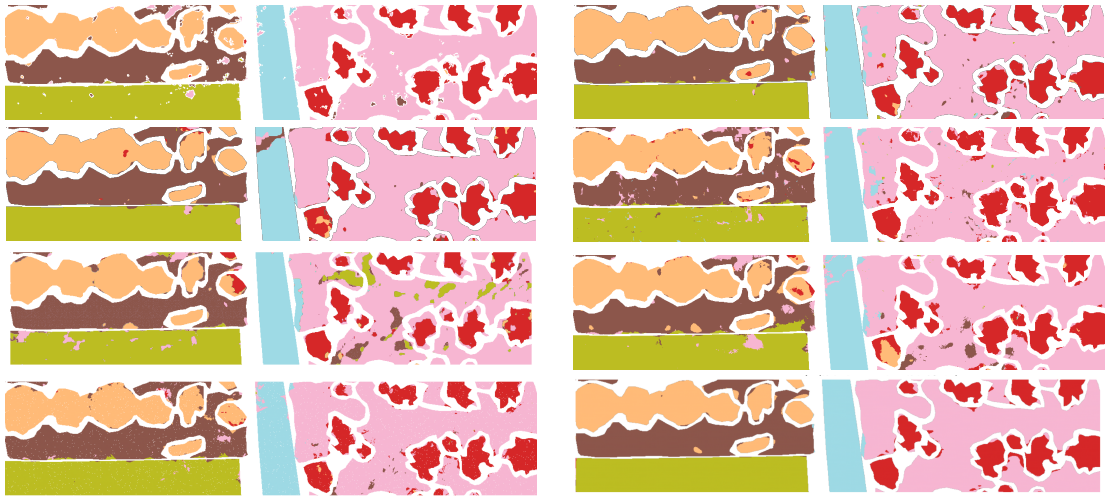

Figure 12. Overview of annotated images and classification performance of SVM, Tri-CNN, HybridSN, SSFTT, MambaHSI, 3DSS-Mamba, SS-Mamba, and the proposed MambaHSINet on the HuaiL_2 dataset.

 doi:10.20944/preprints202605.1614.v1

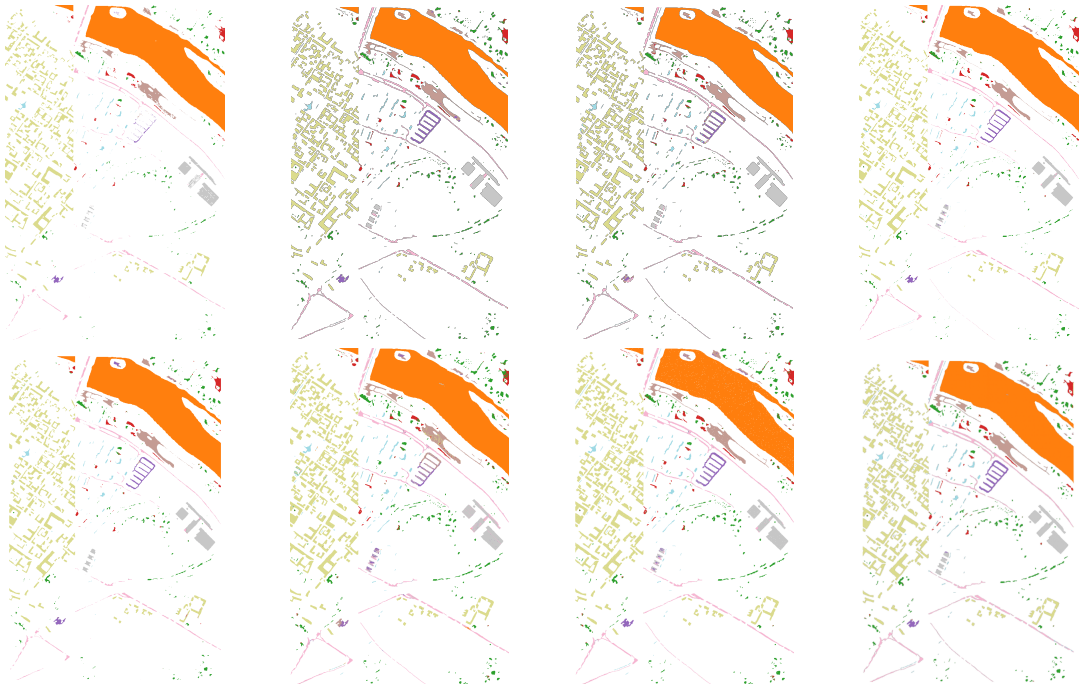

Figure 13. Overview of annotated images and classification performance of SVM, Tri-CNN, HybridSN, SSFTT, MambaHSI, 3DSS-Mamba, SS-Mamba, and the proposed MambaHSINet on the Pavia dataset.

1) HL_1 Dataset

Table 5 presents the classification results on the *HL_1* dataset. The proposed MambaHSINet achieves the state-of-the-art performance with OA, AA, and Kappa reaching 99.50%, 99.62%, and 99.39%, respectively. Compared with the strong 3D-CNN baseline HybridSN, our method improves OA and AA by 0.40% and 0.99%, respectively. This demonstrates that for complex forest scenes with extremely high spectral redundancy, our dual-branch structure can better decouple spectral and spatial features than pure convolutional networks. Furthermore, MambaHSINet outperforms SS-Mamba, demonstrating its superior capability in modeling long-range dependencies for fine-grained tree species classification.

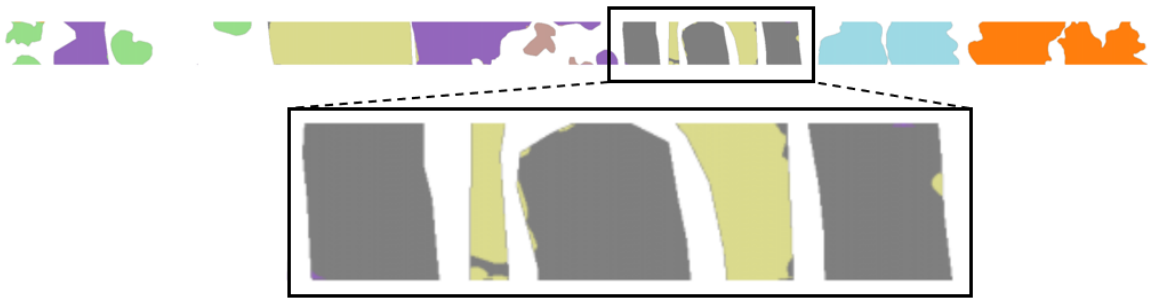

Figure 14. Detailed visualization of the long-straight boundary issue in the HL_1 dataset.

In the *HL_1* dataset, Category 5 (C5) and Category 6 (C6) both represent classes with long-straight boundaries that are cross-arranged and interfere with each other. Comparative experimental results indicate that the original MambaHSI method achieves accuracies of 93.12% for C5 and 95.93% for C6, both of which are lower than other methods such as SVM (97.89% and 98.12%) and HybridSN (99.76% and 99.31%).

This phenomenon reveals a significant long-straight boundary issue: when C5 and C6 are cross-arranged with straight edges, heterogeneous pixel sequence patterns frequently appear within the same scanning line. This poses an interference to the state update mechanism of Mamba, leading to the mutual penetration of feature representations between classes on both sides of the boundary area.

 doi:10.20944/preprints202605.1614.v1

MambaHSINet improved the accuracy to 99.51% for C5 and 98.94% for C6, demonstrating that optimizations to the dual-branch architecture—specifically deeper residual connections and enhanced Squeeze-and-Excitation (SE) attention mechanisms—effectively alleviate the boundary confusion problem. Consequently, the overall accuracy (OA) reached 99.48% and the average accuracy (AA) reached 99.53%, outperforming all comparative methods.

2) HL_2 Dataset

Table 6 reports the classification results on the HL_2 dataset. The proposed MambaHSINet achieves the highest performance with a robust OA of 98.39%, maintaining a significant lead over other advanced models such as the Transformer-based SSFTT (95.04%) and the Mamba-based SS-Mamba (95.82%). Specifically, MambaHSINet improves the OA by 3.35% and 2.57% compared to these two methods, respectively. While HybridSN shows a slight advantage in class C4 (98.36%), our method achieves the highest accuracy in all other five categories, particularly exceeding 99.7% in classes C5 and C6. This consistent superiority across diverse categories confirms that our dual-branch architecture is highly adaptable to datasets with varying spatial scales and complex category distributions, effectively capturing both local textures and long-range dependencies.

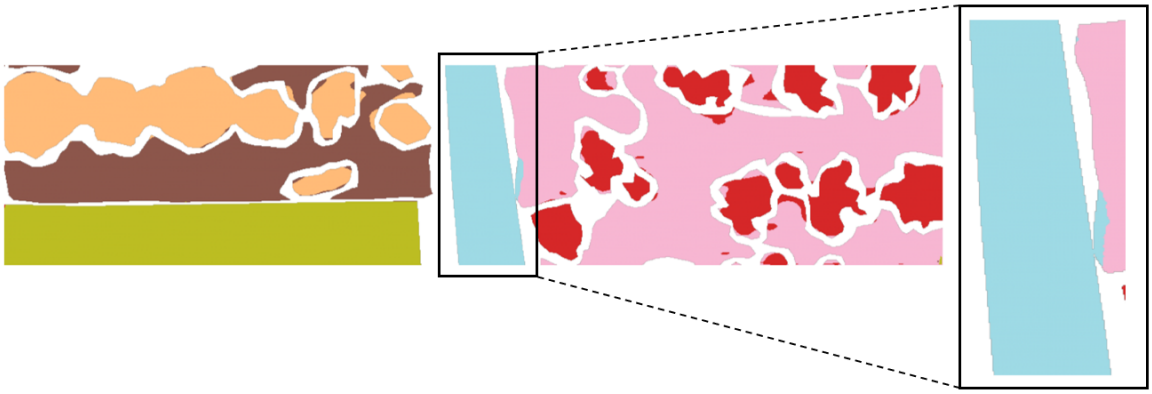

Figure 15. Detailed visualization of the long-straight boundary issue in the HL_2 dataset.

In the *HL_2* dataset, Category 4 (C4) is significantly interfered with by the adjacent long-straight boundary class C6, leading to feature confusion within the spatial branch. The classification accuracy of the original MambaHSI on C4 is only 89.26%, whereas all other comparative methods achieve over 96%, representing a highly significant discrepancy.

Notably, the low accuracy of C4 is not an isolated case; the overall accuracy (OA) and average accuracy (AA) of the original MambaHSI are only 91.17% and 91.76%, respectively, both of which are markedly lower than those of other methods. This indicates that the interference from long-straight boundaries in the *HL_2* dataset has a global impact. Due to the presence of long-straight boundary features across multiple categories, the feature confusion at the boundaries within the spatial branch generates a cross-class propagation effect, resulting in a systematic degradation of the model's overall performance.

MambaHSINet improved the accuracy of C4 to 97.44%, the OA to 98.39%, and the AA to 98.58%, further validating the effectiveness of the architectural optimization in mitigating long-straight boundary issues.

3) Pavia Dataset

Table 7 presents the results on the Pavia dataset. MambaHSINet achieves the highest overall performance with an OA of 99.05% and AA of 97.16%, outperforming all comparative methods. While our model maintains a robust balance across most categories, specific class-wise variations reveal the architectural sensitivities of different models.

 doi:10.20944/preprints202605.1614.v1

Figure 16. Detailed visualization of the long-straight boundary issue in the Pavia dataset.

In the *Pavia* dataset, Category 3 (C3) occupies a relatively small area, resulting in significant performance variations across different algorithms, primarily constrained by statistical fluctuations due to limited training samples. In contrast, Category 6 (C6) consistently exhibits long-straight boundary characteristics and is subjected to strong interference from surrounding objects.

The classification results for C6 reveal a striking disparity among the algorithms: 3DSS-Mamba achieved only 0.53% (rendering it nearly ineffective), while the original MambaHSI reached 97.10% and Tri-CNN attained 99.30%. This substantial gap demonstrates that classes with long-straight boundaries are critical scenarios that lead to performance divergence among different model architectures. Specifically, the combination of 3D convolution and Mamba in 3DSS-Mamba caused severe feature confusion in this context.

MambaHSINet achieved 96.10% on C6, with an overall accuracy (OA) of 98.65% and an average accuracy (AA) of 97.16%. Its superior comprehensive performance across all comparative methods indicates that the proposed architectural improvements possess a certain degree of resistance to long-straight boundary interference.

5. Discussion

5.1. Analysis of the Impact of Long-Straight Boundary Classes on Classification Performance

Based on the comparative experimental results, the classification accuracy of classes with long-straight boundaries exhibits significant variations across different algorithms. This phenomenon is representative in hyperspectral image (HSI) classification and merits in-depth analysis.

MambaHSINet utilizes a dual-branch architecture to extract spectral and spatial features, followed by bidirectional Mamba layers to establish global contextual dependencies. While this design performs excellently across most categories, noteworthy fluctuations occur in specific classes characterized by long-straight boundaries. Considering the architectural characteristics, the primary reason is conjectured to be that the Mamba layer flattens 2D spatial data into 1D sequences for selective scanning. When a long-straight boundary is parallel to the scanning direction, heterogeneous pixels from both sides of the boundary are forced into adjacent positions in the sequence, which interacts with Mamba's state update mechanism. Specifically:

- **First**, the current spatial branch employs multi-scale square convolutions (3×3, 5×5, and 7×7). Their maximum receptive fields only cover local neighborhoods, making it difficult to establish global contextual connections across long-straight boundaries during the spatial feature extraction stage.
- **Second**, although Mamba's selective scanning possesses global modeling capabilities, its sequential nature inherently traverses the image in a fixed direction (e.g., row-major). When processing large-scale regular boundaries, the state updates for pixels on either side are constrained by the sequential order.
- **Third**, while bidirectional Mamba (forward and backward) partially mitigates the bias of unidirectional scanning, it remains a form of 1D directional modeling. It cannot fully cover the arbitrary spatial connection patterns in HSIs. When a boundary aligns perfectly with a scanning direction, the "false adjacency effect" of pixels across the boundary becomes more pronounced in the sequence.

 doi:10.20944/preprints202605.1614.v1

5.2. Applicability Discussion in Tree Species Classification Scenarios

It is essential to highlight that the impact of the aforementioned long-straight boundary issue on MambaHSINet is highly dependent on the specific application scenario. There are fundamental differences in spatial distribution characteristics between tree species classification and general land-cover classification. The boundaries of general land-cover types (e.g., farmlands, water bodies, and construction land) are typically determined by human activities or topography, presenting large-scale, regular, and long-straight geometries. In contrast, tree species usually appear as scattered canopy patches in remote sensing images. Their boundaries are governed by biological growth and ecological competition, resulting in highly irregular curvilinear shapes. Furthermore, ecological transition zones often exist between adjacent species, making category changes gradual rather than abrupt.

In tree species classification, the dual-branch architecture of MambaHSINet is fully utilized: the spectral branch accurately distinguishes the spectral signatures of different species, the spatial branch captures multi-scale textural features of the canopy, and the bidirectional Mamba establishes consistent spatial-spectral contexts within scattered patches. The long-straight boundary issue is rarely triggered in this context because tree species lack large-scale regular boundaries and high-frequency category jumps within the same scanning line. Therefore, applying MambaHSINet to tree species classification is highly rational, allowing its advantages in scattered and irregular patch scenarios to be fully realized.

5.3. Future Research Directions

Although MambaHSINet demonstrates superior performance in tree species classification, the long-straight boundary issue must be addressed if the model is to be extended to general HSI classification tasks, such as agricultural plot zoning or urban land-use mapping. One feasible approach is to introduce multi-directional *stripe convolutions* (horizontal and vertical) into the spatial branch, working in parallel with the existing multi-scale square convolutions. By establishing direct contextual links across long-straight boundaries prior to Mamba's serialized scanning, the false adjacency effect can be mitigated at the spatial branch stage. This improvement is expected to make the model compatible with both scattered patches and large-scale regular plots, facilitating its evolution toward a universal high-precision HSI classification model.

6. Conclusion

This paper proposed the MambaHSINet, a full-image-input classification framework, and constructed the ZJK tree-species dataset oriented toward complex real-world scenarios. By integrating dual-branch spatial–spectral feature extraction with bidirectional Mamba global modeling, the model effectively captures multi-scale structural information and long-range contextual dependencies while maintaining low computational complexity.

Experimental results demonstrate that MambaHSINet exhibits superior and robust performance across multiple datasets, significantly outperforming comparative methods, particularly when handling scattered and irregularly shaped tree-species patches. Addressing the "long-straight boundary issue" inherent in sequential scanning mechanisms, MambaHSINet effectively alleviates such interference through optimized residual connections and attention mechanisms. Given that natural forest canopies predominantly present irregular curvilinear boundaries, this model possesses extremely high application value in practical forestry investigations. Future research will explore the introduction of multi-directional stripe convolutions to further enhance the model's robustness in universal land-cover classification tasks.

References

1. Zhong, L.; Dai, Z.; Fang, P.; Cao, Y.; Wang, L. A Review: Tree Species Classification Based on Remote Sensing Data and Classic Deep Learning-Based Methods. *Forests* **2024**, *15*, 852.
2. Ni-Meister, W.; Albanese, A.; Lingo, F. Assessing Data Preparation and Machine Learning for Tree Species Classification Using Hyperspectral Imagery. *Remote Sensing* **2024**, *16*, 3313. https://doi.org/10.3390/rs16173313.

 doi:10.20944/preprints202605.1614.v1

3. Zhao, X.; Qi, J.; Huang, H. Sensitivity Testing Analysis of Airborne Hyperspectral Lidar Signals for Monitoring Insects and Diseases Based on 3D Radiative Transfer Model. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Pasadena, CA, USA, 2023; pp. 4613–4616.
4. Goetz, A. Three Decades of Hyperspectral Remote Sensing of the Earth: A Personal View. *Remote Sens. Environ.* **2009**, *113*, S5–S16. https://doi.org/10.1016/j.rse.2007.12.014.
5. Dalponte, M.; Ørka, H.; Gobakken, T.; Gianelle, D.; Næsset, E. Tree Species Classification in Boreal Forests with Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2012**, *51*, 2632–2645. https://doi.org/10.1109/TGRS.2012.2216272.
6. Yang, R.; Kan, J. Classification of Tree Species in Different Seasons and Regions Based on Leaf Hyperspectral Images. *Remote Sensing* **2022**, *14*, 1524.
7. Hou, C.; Liu, Z.; Chen, Y.; Wang, S.; Liu, A. Tree Species Classification from Airborne Hyperspectral Images Using Spatial–Spectral Network. *Remote Sensing* **2023**, *15*, 5679.
8. Clark, M.L.; Roberts, D.A.; Clark, D.B. Hyperspectral Discrimination of Tropical Rain Forest Tree Species at Leaf to Crown Scales. *Remote Sensing of Environment* **2005**, *96*, 375–398.
9. Dalponte, M.; Bruzzone, L.; Gianelle, D. Tree Species Classification in the Southern Alps Based on the Fusion of Very High Geometrical Resolution Multispectral/Hyperspectral Images and LiDAR Data. *Remote Sensing of Environment* **2012**, *123*, 258–270.
10. Immitzer, M.; Atzberger, C.; Koukal, T. Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing* **2012**, *4*, 2661–2693.
11. Yel, S.G.; Tunc Gormus, E. Exploiting Hyperspectral and Multispectral Images in the Detection of Tree Species: A Review. *Frontiers in Remote Sensing* **2023**, *4*, 1136289.
12. Melgani, F.; Bruzzone, L. Classification of Hyperspectral Remote Sensing Images with Support Vector Machines. *IEEE Transactions on Geoscience and Remote Sensing* **2004**, *42*, 1778–1790.
13. Mountrakis, G.; Im, J.; Ogole, C. Support Vector Machines in Remote Sensing: A Review. *ISPRS Journal of Photogrammetry and Remote Sensing* **2011**, *66*, 247–259.
14. Tong, F.; Zhang, Y. Spectral–Spatial and Cascaded Multilayer Random Forests for Tree Species Classification in Airborne Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing* **2022**, *60*, 1–11. Art. no. 4411711.
15. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Sparse Representation for Target Detection in Hyperspectral Imagery. *IEEE Journal of Selected Topics in Signal Processing* **2011**, *5*, 629–640.
16. Hu, W.; Huang, Y.; Wei, L.; et al. Deep Convolutional Neural Networks for Hyperspectral Image Classification. *Journal of Sensors* **2015**, *2015*, 258619.
17. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing* **2016**, *54*, 6232–6251.
18. Li, W.; Wu, G.; Zhang, F.; Du, Q. Hyperspectral Image Classification Using Deep Pixel-Pair Features. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *55*, 844–853.
19. Vaswani, A.; Shazeer, N.; Parmar, N.; et al. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems, 2017, Vol. 30.
20. Hong, D.; Han, Z.; Yao, J.; et al. SpectralFormer: Rethinking Hyperspectral Image Classification with Transformers. *IEEE Transactions on Geoscience and Remote Sensing* **2021**, *60*, 1–15.
21. Sun, L.; Zhao, G.; Zheng, Y.; et al. Spectral–Spatial Feature Tokenization Transformer for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing* **2022**, *60*, 1–14.
22. Gu, A.; Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. In Proceedings of the First Conference on Language Modeling, 2024.
23. He, Y.; Tu, B.; Liu, B.; et al. 3DSS-Mamba: 3D-Spectral-Spatial Mamba for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing* **2024**.
24. Liang, L.; Zhang, J.; Duan, P.; et al. LKMA: Learnable Kernel and Mamba with Spatial-Spectral Attention Fusion for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing* **2025**.
25. Li, Y.; Luo, Y.; Zhang, L.; Wang, Z.; Du, B. MambaHSI: Spatial–Spectral Mamba for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing* **2024**, *62*, 1–16. Art. no. 5524216.
26. Zhou, Z.; Guo, X.; Xiong, Y.; Xia, C. Kalman-SSM: Modeling Long-Term Time Series with Kalman Filter Structured State Spaces. *IEEE Signal Process. Lett.* **2024**, *31*, 2470–2474. https://doi.org/10.1109/LSP.2024.3461421.

 doi:10.20944/preprints202605.1614.v1

27. Huang, L.; Chen, Y.; He, X. Spectral–Spatial Mamba for Hyperspectral Image Classification. *Remote Sens.* **2024**, *16*, 2449. https://doi.org/10.3390/rs16132449.
28. Liao, J.; Wang, L. HyperspectralMamba: A Novel State Space Model Architecture for Hyperspectral Image Classification. *Remote Sens.* **2025**, *17*, 2577. https://doi.org/10.3390/rs17152577.
29. Liu, Q.; Yue, J.; Fang, Y.; Xia, S.; Fang, L. HyperMamba: A Spectral–Spatial Adaptive Mamba for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–14. https://doi.org/10.1109/TGRS.2024.3482473.
30. Peng, H.; Lin, K.; Liu, H. HS-Mamba: Full-Field Interaction Multi-Groups Mamba for Hyperspectral Image Classification. *arXiv* **2025**, [2504.15612].
31. Ahmad, M.; Usama, M.; Mazzara, M.; Distefano, S. WaveMamba: Spatial–Spectral Wavelet Mamba for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2024**, *22*, 1–5.
32. Wang, G.; Zhang, X.; Peng, Z.; Zhang, T.; Jiao, L. S^2 Mamba: A Spatial–Spectral State Space Model for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 1–13. https://doi.org/10.1109/TGRS.2024.3431668.
33. Pan, Z.; Li, C.; Plaza, A.; Chanussot, J.; Hong, D. Hyperspectral Image Classification With Mamba. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 1–14. https://doi.org/10.1109/TGRS.2024.3521411.
34. Liang, L.; Xie, P.; Zhang, Y.; Li, J.; Zhang, Z.; Li, J.; Plaza, A. DBMLLA: Double-Branch Mamba-Like Linear Attention Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**.
35. Alkhatib, M.; Al-Saad, M.; Aburaed, N.; Almansoori, S.; Zabalza, J.; Marshall, S.; Al-Ahmad, H. Tri-CNN: A Three Branch Model for Hyperspectral Image Classification. *Remote Sens.* **2023**, *15*, 316.
36. Jiang, X.; Han, C.; Mesgarani, N. Dual-Path Mamba: Short and Long-Term Bidirectional Selective Structured State Space Models for Speech Separation **2025**. pp. 1–5.
37. Li, S.; Song, W.; Fang, L.; Chen, Y.; Ghamisi, P.; Benediktsson, J. Deep Learning for Hyperspectral Image Classification: An Overview. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6690–6709.
38. Fauvel, M.; Benediktsson, J.; Chanussot, J.; Sveinsson, J. Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 3804–3814.
39. Roy, S.; Krishna, G.; Dubey, S.; Chaudhuri, B. HybridSN: Exploring 3-D–2-D CNN Feature Hierarchy for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 277–281.
40. Li, J.; Li, C.; Wang, S.; Yan, J.; Fu, H.; Gao, L.; Liao, W. SS-Mamba: A Spectral–Spatial Mamba-Based Learning Framework for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–14.