3DSS-Mamba: 3D-Spectral–Spatial Mamba for Hyperspectral Image Classification

Yan He, *Student Member, IEEE*, Bing Tu, *Senior Member, IEEE*, Bo Liu, *Member, IEEE*, Jun Li, *Fellow, IEEE*, and Antonio Plaza, *Fellow, IEEE*

***Abstract*— Hyperspectral image (HSI) classification constitutes the fundamental research in remote sensing fields. Convolutional neural networks (CNNs) and Transformers have demonstrated impressive capability in capturing spectral–spatial contextual dependencies. However, these architectures suffer from limited receptive fields and quadratic computational complexity, respectively. Fortunately, recent Mamba architectures built upon the state space models (SSMs) integrate the advantages of long-range sequence modeling and linear computational efficiency, exhibiting substantial potential in low-dimensional scenarios. Motivated by this, we propose a novel 3D-spectral–spatial mamba (3DSS-Mamba) framework for HSI classification, allowing for global spectral–spatial relationship modeling with greater computational efficiency. Technically, a spectral–spatial token generation (SSTG) module is designed to convert the HSI cube into a set of 3-D spectral–spatial tokens. To overcome the limitations of traditional Mamba, which is confined to modeling causal sequences and inadaptable to high-dimensional scenarios, a 3D-spectral–spatial selective scanning (3DSS) mechanism is introduced, which performs pixel-wise selective scanning on 3-D hyperspectral tokens along the spectral and spatial dimensions. Five scanning routes are constructed to investigate the impact of dimension prioritization. The 3DSS scanning mechanism combined with conventional mapping operations forms the 3D-spectral–spatial mamba block (3DMB), enabling the extraction of global spectral–spatial semantic representations. Experimental results and analysis demonstrate that the proposed method outperforms the state-of-the-art methods on HSI classification benchmarks. The code is available at https://github.com/IIP-Team/3DSS-Mamba.**

***Index Terms*— Hyperspectral image (HSI) classification, Mamba, spectral–spatial modeling.**

I. Introduction

HYPERSPECTRAL images (HSI) are represented by hundreds of continuous spectral bands in the electromagnetic spectrum, encompassing abundant spectral and spatial information. Compared with natural images, HSI performs widespread applications in various remote sensing scenarios, such as mineral exploration [1], [2], military reconnaissance [3], [4], and environmental monitoring [5], [6]. As a fundamental task for HSI processing, HSI classification focuses on pixel-level category distinguishing for ground objects, which has received considerable attention in remote sensing [7], [8].

Traditional research on HSI classification typically draws upon spectral feature extraction with hand-crafted descriptors and subspace learning, such as support vector machine (SVM) [9], linear discriminant analysis (LDA) [5], and manifold learning [10], [11]. To cope with the challenges of spectral variability and spectral confusion, several efforts integrate complementary spatial contextual information with spectral features for precise HSI classification mapping, including extended morphological profiles (EMPs) [12], extended multiattribute profiles (EMAPs) [13], and sparse manifold representations [14]. However, these methods heavily rely on prior parameter settings, which exhibit insufficient data fitting and description capabilities when confronting complex environmental conditions.

The rapid development of deep learning technology has brought significant paradigms for HSI classification task. Representative models encompass autoencoders (AEs) [15], [16], convolutional neural networks (CNNs) [17], [18], [19], [20], [21], recurrent neural networks (RNNs) [22], [23], and graph convolutional networks (GCNs) [24], [25]. Building upon the properties of local receptive fields and parameter sharing, CNN architectures progressively demonstrate predominant status in HSI classification. For instance, Hu et al. [26] first presented a hierarchical 1-D CNN network to extract the high-level spectral features along the spectral dimension of hyperspectral data. Given the characteristics of abundant spectral channels and strong spatial correlation in HSIs, Yang et al. [27] constructed a dual-branch architecture that combines 1-D CNN and 2-D CNN to simultaneously capture finer spectral and spatial features for HSI classification. Compared with the 2-D convolutional paradigm restricted to spatial dimension, the 3-D convolutional kernels enjoy the advantage of spectral–spatial joint feature extraction. Classically, Zhong et al. [28] developed a 3-D CNN-based spectral–spatial residual network (SSRN), which is capable of capturing deep spectral–spatial blocks directly from raw 3-D HSI cubes without additional feature engineering. Despite achieving encouraging performance compared to traditional approaches,

Received 2 June 2024; revised 30 July 2024 and 23 September 2024; accepted 29 September 2024. Date of publication 2 October 2024; date of current version 17 October 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62271200 and in part by the Start-Up Foundation for Introducing Talent of Nanjing University of Information Science and Technology (NUIST) under Grant 2023r091. *(Corresponding author: Bing Tu.)*

Yan He, Bing Tu, and Bo Liu are with the Institute of Optics and Electronics, the State Key Laboratory Cultivation Base of Atmospheric Optoelectronic Detection and Information Fusion, and the Jiangsu International Joint Laboratory on Meteorological Photonics and Optoelectronic Detection, and the Jiangsu Engineering Research Center for Intelligent Optoelectronic Sensing Technology of Atmosphere, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: tubing@nuist.edu.cn).

Jun Li is with the Faculty of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: lijuncug@cug.edu.cn).

Antonio Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politecnica, University of Extremadura, 10003 Cáceres, Spain (e-mail: aplaza@unex.es).

Digital Object Identifier 10.1109/TGRS.2024.3472091

1558-0644 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

CNN-based models struggle to establish long-range dependencies between pixels, failing to capture global spectral and spatial characteristics.

Benefiting from the powerful long-distance sequence modeling capability based on attention mechanism, Transformer architecture has been adeptly investigated for HSI classification task [29], [30], [31], [32], [33]. The vision transformer (ViT) [34] treats each pixel within the HSI cube as a sequence input to the standard Transformer model, capturing the correlations between pixels through the self-attention mechanism. Derived from this, He et al. [35] proposed a bidirectional encoder representation transformer network (HSI-BERT) for HSI classification, which overcomes the restrictions of spatial distance through global receptive fields. Given the significance of long-range dependencies in spectral dimensions, Hong et al. [36] devised a novel Transformer-based SpectralFormer (SF) network, which constructs group-wise spectral embeddings to capture the spectral sequence information between neighboring HSI bands. To comprehensively integrate both spectral and spatial information for HSI classification, Zhong et al. [37] developed a spectral–spatial transformer network (SSTN), which breaks the long-range limitations by integrating spatial attention and spectral association modules, and incorporates a factorized architecture search model to determine the layer-level operations and block-level orders. Additionally, Peng et al. [38] constructed a dual-branch spatial–spectral transformer with cross-attention (CASST), where the spectral branch establishes dependencies among spectral sequences and the spatial branch captures fine-grained spatial contexts. The interaction between spatial and spectral information is performed within each transformer block through a cross-attention mechanism. Although Transformer architecture has exhibited impressive capability in HSI classification, its inherent self-attention mechanism is characterized by quadratic computation complexity $\mathcal{O}(N^2)$, which poses significant challenges in modeling efficiency and memory overhead.

Comparatively, recent Mamba [39] built on state-space models (SSMs) establishes long-distance dependency through state transitions, enjoying the promising attributes of linear computational complexity and scalability. As an effective alternative to the Transformer, Mamba introduces the selective SSMs for 1-D sequence modeling along specific orientation, which demonstrates substantial potential in natural language processing (NLP) tasks. To accommodate vision scenarios involving 2-D-spatial awareness, Vim [40] and VMamba [41] extend the Mamba architecture by introducing a multidirectional scanning mechanism to achieve global contextual modeling, showcasing great efficiency and effectiveness in 2-D visual tasks, such as object detection and semantic segmentation. Although Mamba architectures have demonstrated substantial potential in low-dimensional scenarios, the adaptability to high-dimensional HSI classification tasks involving 3-D hyperspectral data requires further exploration, as depicted in Fig. 1.

To this end, this work investigates 3D-spectral–spatial Mamba (3DSS-Mamba), an efficient global spectral–spatial contextual modeling framework based on the SSMs for HSI classification. The 3DSS-Mamba consists of a spectral–spatial token generation (SSTG) module, multiple stacked 3D-spectral–spatial mamba blocks (3DMBs), and a prediction module. Specifically, SSTG converts the HSI cube into a set of spectral–spatial tokens by introducing a 3-D convolution block, with each token maintaining the 3-D structure. To address the inadaptability of traditional Mamba for high-dimensional hyperspectral scenarios, a 3D-spectral–spatial selective scanning (3DSS) mechanism is customized as the core component of 3DMB to achieve spectral–spatial sequence modeling. The 3DSS first performs pixel-wise sequence flattening on each 3-D hyperspectral token along the spectral and spatial dimensions, then introduces the S6 model to conduct selective scanning to facilitate interactions among adjacent pixels. Five scanning routes are constructed to investigate the impact of dimension prioritization, including spectral-priority, spatial-priority, cross spectral–spatial, cross spatial–spectral, and parallel spectral–spatial. Multiple 3DMBs are stacked to extract comprehensive spectral–spatial semantic features, followed by a classifier for final classification. Compared to existing methods, the proposed 3DSS-Mamba exhibits superior capabilities in capturing global spectral–spatial contextual dependencies with greater computational efficiency.

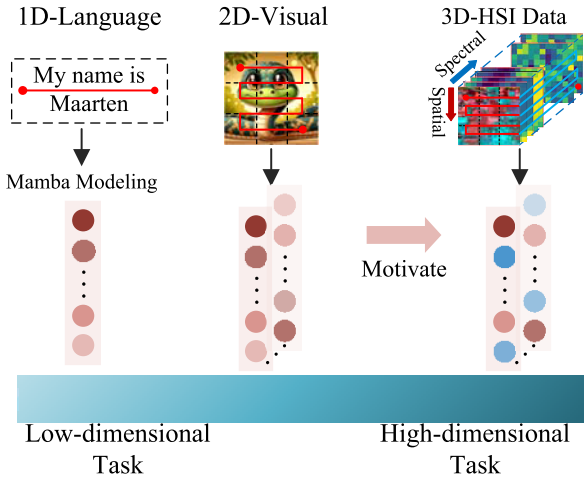

Fig. 1. Motivation of the proposed 3DSS-Mamba. Mamba modeling has demonstrated substantial potential in low-dimensional scenarios such as 1-D language and 2-D visual tasks, motivating its adaptability to high-dimensional HSI classification task.

The main contributions are summarized as follows.

1) A novel 3DSS-Mamba framework based on the SSMs is proposed for HSI classification, which can explicitly establish long-range spectral–spatial contextual dependencies with linear computational complexity.
2) A 3DSS mechanism tailored for high-dimensional hyperspectral scenarios is introduced. By performing pixel-wise selective scanning on 3-D hyperspectral tokens along the spectral and spatial dimensions, the spectral reflectance and spatial regularity can be adequately explored from the sequence modeling perspective.
3) Extensive experiments are verified on three public hyperspectral datasets. The results indicate the effectiveness and superiority of the proposed method.

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

The remaining sections of this article are organized as follows. Section II provides a comprehensive description of the proposed method. Section III outlines the experimental results and analyses. Conclusions and future work are discussed in Section IV.

II. Proposed Network

This section commences with the preliminaries associated with SSMs. Following this, we investigate a novel 3DSS mechanism specifically tailored for the 3-D HSI data, followed by the establishment of 3DMB. Building upon these submodules, the overall architecture of the proposed 3DSS-Mamba framework for HSI classification is meticulously introduced.

A. Preliminaries

1) State Space Models: The concept of SSMs [42] originates from continuous linear time-invariant systems. Taking a 1-D signal $x(t) \in \mathbb{R}$ as input, SSMs are dedicated to mapping it into a sequence $y(t) \in \mathbb{R}$ via an intermediate hidden state $h(t) \in \mathbb{R}^N$. This procedure can be formulated through the following linear ordinary differential equation (ODE)

$$\begin{aligned} h'(t) &= \mathbf{A}h(t) + \mathbf{B}x(t) \\ y(t) &= \mathbf{C}h(t) \end{aligned} \tag{1}$$

where $h(t)' \in \mathbb{R}^N$ refers to the time derivative of $h(t)$, $\mathbf{A} \in \mathbb{R}^{N\times N}$ denotes the state transition matrix, and $\mathbf{B} \in \mathbb{R}^{N\times 1}$, $\mathbf{C} \in \mathbb{R}^{N\times 1}$ represent the projection matrices. The first equation defines the evolution of the hidden state $h(t)$, while the second specifies that the output is a linear transformation of the hidden state $h(t)$.

The continuous-time system delineated by (1) generally encounters challenges when integrating into discrete sequence-based deep models. To this end, the zero-order hold (ZOH) technique [43] with a time-scale parameter Δ is subsequently employed to facilitate a straightforward discretization step, which converts the continuous parameters $\mathbf{A}$ and $\mathbf{B}$ into their discrete counterparts $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$

$$\begin{aligned} \overline{\mathbf{A}} &= \exp(\Delta\mathbf{A}) \\ \overline{\mathbf{B}} &= (\Delta\mathbf{A})^{-1}(\exp(\Delta\mathbf{A}) - \mathbf{I}) \cdot \Delta\mathbf{B}. \end{aligned} \tag{2}$$

The detailed discretization process can be referred to [39] and [44]. B is modulated by the dynamics of A through the exponential term, reflecting the influence of state transitions on the input. After discretization, the discretized SSM system can be formulated as follows:

$$\begin{aligned} h_t &= \overline{\mathbf{A}}h_{t-1} + \overline{\mathbf{B}}x_t \\ y_t &= \mathbf{C}h_t. \end{aligned} \tag{3}$$

To enhance the computational efficiency and scalability, the convolution operation $*$ is harnessed to expedite the linear recurrence process outlined above. Consequently, the ultimate output can be synthesized as

$$\begin{aligned} \overline{\mathbf{K}} &= \left(\mathbf{C}\overline{\mathbf{B}}, \mathbf{C}\overline{\mathbf{A}\mathbf{B}}, \ldots, \mathbf{C}\overline{\mathbf{A}}^{\mathbf{L}-1}\overline{\mathbf{B}}\right) \\ \mathbf{y} &= \mathbf{x} * \overline{\mathbf{K}} \end{aligned} \tag{4}$$

where L denotes the length of input sequence, and $\overline{\mathbf{K}} \in \mathbb{R}^L$ serves as the structured convolutional kernel.

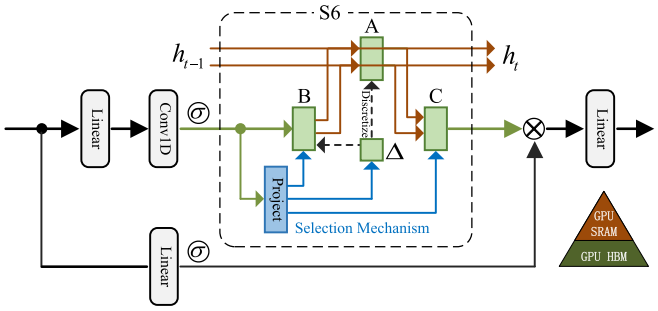

Fig. 2. Architecture of the Mamba block.

2) Mamba: Traditional SSMs predominantly rely on the simplifying assumption of linear time-invariant, which enjoy the advantage of linear time complexity but struggle to capture the contextual information within input sequences. To break this limitation, Mamba [39] introduces a selection mechanism and proposes the selective SSMs (S6) to achieve dynamic interactions between sequential states. The detailed architecture of the Mamba block is illustrated in Fig. 2. Different from traditional SSMs integrated with static parameterization, S6 allows the projection matrices to be modified as input-dependent, which achieves selective attention on each sequence unit. Concretely, the parameters $\mathbf{B}$, $\mathbf{C}$, and interval Δ are dynamically projected from the input sequence $x \in \mathbb{R}^{B\times L\times D}$, which can be formulated as

$$\begin{aligned} \mathbf{B} &= \text{Projection}_B(x) \\ \mathbf{C} &= \text{Projection}_C(x) \\ \Delta &= \tau_\Delta(\text{Parameter} + \text{Projection}_\Delta(x)) \end{aligned} \tag{5}$$

where $\mathbf{B} \in \mathbb{R}^{B\times L\times N}$, $\mathbf{C} \in \mathbb{R}^{B\times L\times N}$, and $\Delta \in \mathbb{R}^{B\times L\times D}$. The functions $\text{Projection}_B(\cdot)$ and $\text{Projection}_C(\cdot)$ perform linear projections to the N-dimensional space. The symbol τ_Δ represents the Softplus activation function [45], and $\text{Projection}_\Delta = \text{Broadcast}_D(\text{Linear}_1(x))$, which transforms x to a size of $\mathbb{R}^{B\times L\times 1}$ and then broadcasts to the D-dimension. This selective mechanism allows Mamba to effectively filter out irrelevant noise in time series tasks, while selectively retaining or discarding information pertinent to the current input. Additionally, Mamba introduces a hardware-aware algorithm that computes the model recurrently with a scan, ensuring both effectiveness and efficiency in capturing global contextual information.

B. 3D-Spectral–Spatial Selective Scanning

The original Mamba processes data along a specific orientation, which is effectively employed for the causal modeling of 1-D input sequences. To accommodate vision tasks involving 2-D-spatial awareness, recent VMamba [41] introduces a 2-D-selective-scanning technique. The cross-scanning mechanism rearranges tokens along spatial dimensions and then transmits them into the S6 model for sequence modeling. Although the above scanning techniques have demonstrated commendable application in language data and natural image, they may encounter substantial challenges when adapting to 3-D hyperspectral data that exhibit inherent visual spatial and

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

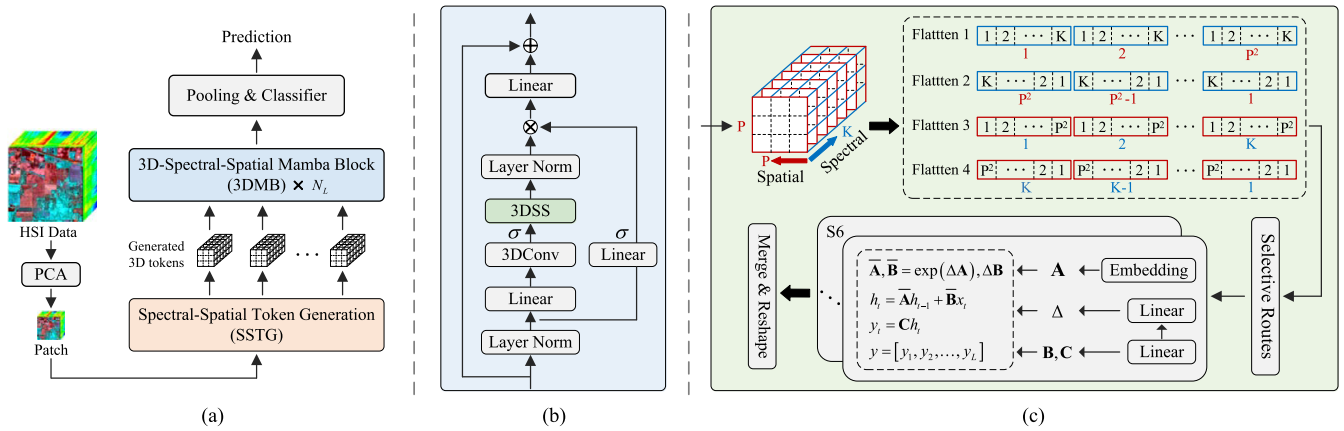

Fig. 3. (a) Overall architecture of the proposed 3DSS-Mamba for HSI classification, which consists of a SSTG, N_L stacked 3DMBs, and a classifier module. (b) Structural flow of proposed 3DMB. (c) Computational procedure of proposed 3DSS.

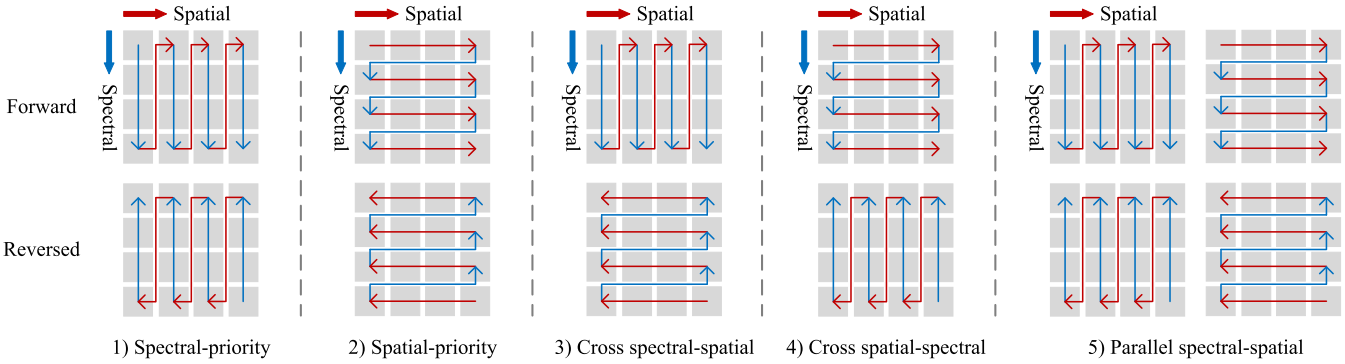

Fig. 4. Five flattening routes are constructed to explore the impact of dimension prioritization.

continuous spectral characteristics. To address this issue, this article proposes a 3DSS module, which performs pixel-wise sequential scanning for 3-D hyperspectral input to achieve global spectral–spatial relationship modeling.

Inspired by VMamba [41], the developed 3DSS is depicted in Fig. 3(c), primarily comprising two stages: 3D-spectral–spatial sequence flattening and selective scanning with S6 mechanism.

1) 3D-Spectral–Spatial Sequence Flattening: Unlike conventional 2-D scanning that solely emphasizes spatial information, 3DSS performs pixel-wise sequential scanning for 3-D hyperspectral tokens along both spectral and spatial dimensions, generating the flattened forward 1-D sequence. To adequately capture spectral–spatial contextual details, 3DSS additionally performs flipping operations on the forward sequence to enable bidirectional sequence scanning.

To explore the impact of dimension prioritization, five flattening routes are delineated as illustrated in Fig. 4. First, spectral-priority: initially unfolds the 3-D hyperspectral token along the spectral dimension and then arranges them in spatial order. Given the input of hyperspectral tokens $F = \{S_1, S_2, \ldots, S_M\}$, $S_i \in \mathbb{R}^{P\times P\times K}$, where P and K denote the patch size and spectral band number of token cube, respectively, the flattening process for each 3-D token cube S_i can be formulated as

$$\begin{aligned}\mathcal{S}_i^{\text{spe-fwd}} &= \left[\left[\mathcal{S}_{i,1}^1, \ldots, \mathcal{S}_{i,1}^K\right], \ldots, \left[\mathcal{S}_{i,P^2}^1, \ldots, \mathcal{S}_{i,P^2}^K\right]\right] \\ \mathcal{S}_i^{\text{spe-rvs}} &= \Phi_{\text{revert}}\left(\mathcal{S}_i^{\text{spe-fwd}}\right)\end{aligned} \tag{6}$$

where $\mathcal{S}_i^{\text{spe-fwd}} \in \mathbb{R}^{1\times(P^2\cdot K)}$ indicates the generated forward sequence, and $\mathcal{S}_i^{\text{spe-rvs}} \in \mathbb{R}^{1\times(P^2\cdot K)}$ represents the reversed sequence with flipping function Φ_{revert}. As a result, the bidirectional sequences driven by spectral-priority can be expressed as $\mathcal{S}_i^{\text{seq}} = \{\mathcal{S}_i^{\text{spe-fwd}}, \mathcal{S}_i^{\text{spe-rvs}}\}$. Second, spatial-priority: first organizes the token cube by spatial location and then stacks them band by band. The reversed sequence is constructed with flipping operations. Technically, the flattened sequences can be generated by

$$\begin{aligned}\mathcal{S}_i^{\text{spa-fwd}} &= \left[\left[S_{i,1}^1, \ldots, \mathcal{S}_{i,P^2}^1\right], \ldots, \left[\mathcal{S}_{i,1}^K, \ldots, \mathcal{S}_{i,P^2}^K\right]\right] \\ \mathcal{S}_i^{\text{spa-rvs}} &= \Phi_{\text{revert}}\left(\mathcal{S}_i^{\text{spa-fwd}}\right).\end{aligned} \tag{7}$$

Similarly, the ultimate bidirectional sequences guided by spectral-priority is $\mathcal{S}_i^{\text{seq}} = \{\mathcal{S}_i^{\text{spa-fwd}}, \mathcal{S}_i^{\text{spa-rvs}}\}$. Third, cross spectral–spatial: a hybrid pattern integrating forward Spectral-priority and reversed Spatial-priority, which can be expressed as $\mathcal{S}_i^{\text{seq}} = \{\mathcal{S}_i^{\text{spe-fwd}}, \mathcal{S}_i^{\text{spa-rvs}}\}$. Fourth, cross spatial-spectral: a hybrid pattern combining forward spatial-priority and reversed spectral-priority, i.e., $\mathcal{S}_i^{\text{seq}} = \{\mathcal{S}_i^{\text{spa-fwd}}, \mathcal{S}_i^{\text{spe-rvs}}\}$. Fifth, parallel spectral–spatial: integrates both forward and reversed spatial-priority and spectral-priority routes. The generated sequences can be represented as $\mathcal{S}_i^{\text{seq}} = \{\mathcal{S}_i^{\text{spa-fwd}}, \mathcal{S}_i^{\text{spa-rvs}}, \mathcal{S}_i^{\text{spe-fwd}}, \mathcal{S}_i^{\text{spe-rvs}}\}$. These five routes facilitate pixel interactions among adjacent spatial and spectral positions in different dimension priorities, and their effectiveness will be analyzed and compared in the experimental section.

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

After complimenting the flattening operation following the preset route, the generated sequences S_i^{seq} are transmitted into the subsequent S6 model for sequence modeling.

2) Selective Scanning With S6 Model: The selective scanning model S6 [39] maintains the advantages of dynamic weights (i.e., selectivity) and linear computational complexity. Building on this, the S6 model is extended to multisequence hyperspectral scenario for learning spectral–spatial sequence modeling expression. Specifically, we devise multiple parallel S6 models to independently process input sequences, and eventually merge the resultant to form the output response.

Taking spatial-priority route $\mathcal{S}_i^{\text{seq}} = \left\{\mathcal{S}_i^{\text{spa-fwd}}, \mathcal{S}_i^{\text{spa-rvs}}\right\}$ as example, the scanning procedure for individual sequence can be formulated as

$$\begin{aligned} Y_i^{\text{spa-fwd}} &= \Phi_{S6\text{-fwd}}\left(\mathcal{S}_i^{\text{spa-fwd}}\right) \\ Y_i^{\text{spa-rvs}} &= \Phi_{S6\text{-rvs}}\left(\mathcal{S}_i^{\text{spa-rvs}}\right) \end{aligned} \tag{8}$$

where $\Phi_{S6\text{-}*}$ represents the S6 model, with the detailed computation referenced in (3). After scanning, these generated 1-D mapping sequences are reshaped into 3-D structure and subsequently merged

$$Y_i = \Phi_{\text{merge}}\left(Y_i^{\text{spa-fwd}}, \Phi_{\text{revert}}\left(Y_i^{\text{spa-rvs}}\right)\right). \tag{9}$$

As a result, the ultimate transformed output tokens can be expressed as $F^{\text{out}} = \{Y_1, Y_2, \ldots, Y_M\}, Y_i \in \mathbb{R}^{P\times P\times K}$.

C. 3D-Spectral–Spatial Mamba Block

The 3DMB takes the 3DSS mechanism as its core computing unit, with the objective of capturing global spectral–spatial semantic information. The detailed structure is illustrated in Fig. 3(b).

Specifically, the 3DMB commences with a normalization layer to enhance the model training stability. Following this, two parallel linear embedding layers are stacked, with one branch followed by an activation function for gating signal generation, and the other branch undergoes a 3-D convolution operation with kernel $1 \times 1 \times 1$ for feature extraction. The procedure can be formulated as

$$z = \sigma(\Phi_{\text{linear}}(\Phi_{\text{norm}}(T))) \tag{10}$$

$$F = \sigma(\Phi_{\text{3DConv}}(\Phi_{\text{linear}}(\Phi_{\text{norm}}(T)))) \tag{11}$$

where $T \in \mathbb{R}^{M\times P\times P\times K}$ denotes the input tokens, and σ denotes the Silu [46] activation operation. After this stage, the generated F passes through the pivotal 3DSS mechanism, executing selective scanning as previously described. Subsequently, the output of 3DSS undergoes layer normalization and a gating operation. Finally, the features are transmitted to the ultimate linear layer, followed by a residual connection

$$R = \Phi_{\text{linear}}(\Phi_{\text{norm}}(\text{3DSS}(F)) \otimes z) + T. \tag{12}$$

Notably, the 3DMB enjoys linear computation complexity benefiting from the 3DSS mechanism, allowing for more stackings with similar budgets compared to the Transformer.

Algorithm 1 Pseudo Procedure of the Proposed 3DSS-Mamba

Input: Input HSI data $X \in \mathbb{R}^{H\times W\times V}$, patch size B, depth of 3DMB N_L; Preset $route$ from {Spectral-priority, Spatial-priority, Cross spectral–spatial, Cross spatial–spectral, Parallel spectral–spatial}.

Output: Predicted labels of the testing samples.

Apply PCA for dimension reduction $I_{PCA} \in \mathbb{R}^{H\times W\times d}$, with $d = 30$.
Crop patch $x \in \mathbb{R}^{B\times B\times d}$, with labels determined by the central pixel.
SSTG module:
Perform pixel embedding on x based on Eq. (13), generating a set of 3-D spectral–spatial tokens T.
3DMB blocks:
for depth $j = 1$ to N_L **do**
  Apply linear embedding on normalized T based on Eq. (10), generating gating signal z.
  Apply linear embedding and 3D convolution on normalized T based on Eq. (11), generating feature F.
  3DSS mechanism:
  Perform 3D-spectral–spatial sequence flattening on F following the preset $route$, generating sequences $\mathcal{S}_i^{seq}$ based on Eqs. (6-7).
  for each subsequence in $\mathcal{S}_i^{seq}$ **do**
    Generate parameter matrixes Δ_i, $\mathbf{A}_i$, $\mathbf{B}_i$, $\mathbf{C}_i$ based on Eq. (5).
    Perform selective scanning with the S6 model, with the detailed computation process based on Eq. (3).
  end for
  Merge output sequences based on Eq. (9).
  Apply layer normalization and a gating operation on the merged output F^{out}, generating R based on Eq. (12).
  Apply $T^{j+1} = R^j$ based on Eq. (14).
end for
Classifier:
Perform ultimate classification on spectral–spatial feature R^{N_L}, generating results $pred$ based on Eq. (15).

D. 3D-Spectral–Spatial Mamba: Overview

The architecture of the proposed 3DSS-Mamba for HSI classification is illustrated in Fig. 3(a). It consists of a SSTG module, multiple stacked 3DMBs, and a prediction module. Initially, the cropped patch cube is fed into the SSTG to acquire a series of 3-D spectral–spatial tokens. Subsequently, the generated tokens are input into the stacked 3DMB to capture discriminative spectral–spatial semantic representations. Ultimately, the extracted spectral–spatial features are transmitted to the prediction module to accomplish classification.

Assume that the original hyperspectral data are expressed as $I \in \mathbb{R}^{H\times W\times V}$, where H and W represent the spatial dimensions, and V denotes the spectral dimension. To mitigate the potential Hughes phenomenon [47] caused by high dimensionality, dimensionality reduction is initially conducted on the original hyperspectral data through principal component analysis (PCA) [48]. The modified image is represented as $I_{\text{PCA}} \in \mathbb{R}^{H\times W\times d}$, where d refers to the reduced spectral

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

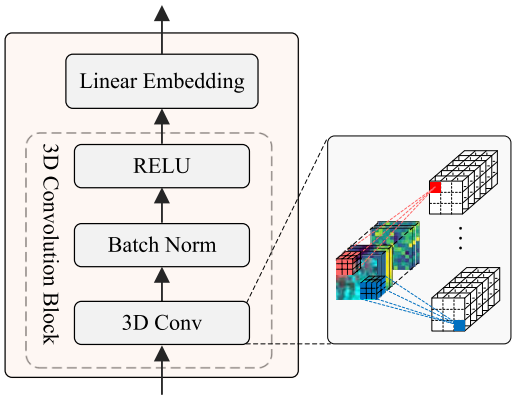

Fig. 5. Detailed structure of SSTG module.

dimension. Given that the adjacent pixels can supplement spatial information for the central pixel, the modified image is further divided into a series of 3-D patch cubes $\{x_i \in \mathbb{R}^{B\times B\times d}\}_{i=1}^{H\times W}$ as input for 3DSS-Mamba, with the labels determined by the central pixel of each patch.

The developed SSTG module is constructed by a 3-D convolution block and an embedding operation, which projects the HSI patch cube into spectral–spatial tokens. The detailed structure is shown in Fig. 5. Taking the cropped pixel-wise patch cube $x \in \mathbb{R}^{B\times B\times d}$ as input, the tokenization process can be formulated as

$$T = \Phi_{\text{linear}}(\Phi_{\text{3DConv}}(x)) \tag{13}$$

where $T \in \mathbb{R}^{M\times P\times P\times K}$ denotes the generated spectral–spatial token. The 3-D convolution block consists of a 3-D convolution layer, a batch normalization layer, and a ReLU activation function. The embedding operation involves a linear layer for dimension transformation.

Subsequently, the generated tokens are fed into by N_L stacked 3DMBs for spectral–spatial semantic extraction. This procedure can be iteratively delineated as follows:

$$\begin{aligned} R^j &= \Phi^j_{\text{3DMB}}\left(T^j\right) \\ T^{j+1} &= R^j \end{aligned} \tag{14}$$

where Φ^j_{3DMB} signifies the jth 3DMB block, and $R^j \in \mathbb{R}^{M\times P\times P\times K}$ represents the corresponding output.

After completing the 3DMB modeling, the spectral–spatial feature R^{N_L} undergoes an average pooling operation, and then passes through the classifier $\Phi_{\text{classifier}}$ comprised of a multilayer perceptron layer to yield the ultimate classification results

$$\text{pred} = \Phi_{\text{classifier}}\left(\Phi_{\text{avg}}\left(R^{N_L}\right)\right). \tag{15}$$

The procedural steps of the proposed 3DSS-Mamba is summarized in Algorithm 1.

III. Performance Evaluation

A. Datasets Description

To illustrate the classification capabilities of the proposed 3DSS-Mamba, three publicly available HSI databases are utilized for comprehensive evaluation, including Pavia University, Indian Pines, and Houston 2013. Detailed descriptions are provided below.

TABLE I
NAME AND NUMBER OF SAMPLES OF EACH CLASS ON THE PAVIA UNIVERSITY DATASETS WITH 5% LABELED DATA

Pavia University dataset			
No.	Class Name	Train	Test
1	Asphalt	332	6299
2	Meadows	932	17717
3	Gravel	105	1994
4	Trees	153	2911
5	Painted metal sheets	67	1278
6	Bare soil	251	4778
7	Bitumen	67	1263
8	Self-blocking bricks	184	3498
9	Shadows	47	900
	Total	2138	40638

TABLE II
NAME AND NUMBER OF SAMPLES OF EACH CLASS ON THE INDIAN PINES DATASETS WITH 10% LABELED DATA

Indian Pines dataset			
No.	Class Name	Train	Test
1	Alfalfa	5	41
2	Corn-notill	143	1285
3	Corn-mintill	83	747
4	Corn	24	213
5	Grass-pasture-mowed	48	435
6	Grass-trees	73	657
7	Grass-pasture	3	25
8	Hay-windrowed	48	430
9	Oats	2	18
10	Soybean-notill	97	875
11	Soybean-mintill	245	2210
12	Soybean-clean	59	534
13	Wheat	20	185
14	Woods	126	1139
15	Buildings	39	347
16	Stone	6	84
	Total	1024	9225

1) Pavia University: The dataset was collected by the Reflective Optics System Imaging Spectrometer (ROSIS) over Pavia, Northern Italy. The imaging wavelength of the spectrometer ranges from 0.43 to 0.86 μm. After removing the noisy bands, the dataset consists of 103 spectral bands and 610 × 340 pixels, with a spatial resolution of 1.3 m per pixel. There are totally 42 776 ground sample points, categorized into nine types including asphalt and gravel. Table I provides the division details for training and testing sets.

2) Indian Pines: The dataset was acquired by the Airborne/Visible Infrared Imaging Spectrometer (AVIRIS) imaging an Indian pine tree over Northwestern Indiana in 1992. The imaging wavelength of the spectrometer ranges from 0.4 to 2.5 μm. The dataset encompasses 200 spectral bands and 145 × 145 pixels after removing the water absorption channels, with a spatial resolution of 20 m per pixel. There are a total of 10249 ground object pixels, representing 16 distinct categories including Alfalfa and Corn-notill. The detailed data division in the experiment is described in Table II.

3) Houston 2013: The dataset was captured by the ITRES CASI-1500 sensor over the University of Houston campus and its surrounding areas, provided by the 2013 GRSS Data

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

TABLE III
NAME AND NUMBER OF SAMPLES OF EACH CLASS ON THE HOUSTON 2013 DATASETS WITH 10% LABELED DATA

Houston 2013 dataset			
No.	Class Name	Train	Test
1	Healthy grass	125	1238
2	Stressed grass	125	1241
3	Synthetic grass	70	690
4	Trees	124	1231
5	Soil	124	1229
6	Water	33	321
7	Residential	127	1255
8	Commercial	124	1231
9	Road	125	1239
10	Highway	123	1214
11	Railway	123	1222
12	Parking Lot 1	123	1220
13	Parking Lot 2	47	464
14	Tennis Court	43	423
15	Running Track	66	653
	Total	1502	14871

Fusion Contest. The image comprises 144 spectral bands ranging in wavelength from 0.38 to 1.05 μm, and consists of 340 $\times$ 1905 pixels with a spatial resolution of 2.5 m per pixel. There are 16 373 sample pixels, covering 15 challenging land cover categories. The precise splitting of training and testing data is exhibited in Table III.

B. Experimental Settings

1) Evaluation Metrics: Following the state-of-the-art HSI classification approaches, overall classification accuracy (OA), average classification accuracy (AA), and kappa coefficient (Kappa) are employed as the evaluation metrics. To guarantee fairness in comparison, all experiments are performed under identical experimental conditions, and the reported results are averaged over five consecutive experiments.

2) Implementation Details: All the experiments are implemented on the PyTorch platform with one RTX 3090Ti GPU. The training epochs and batch size are set as 100 and 64, respectively. The Adam gradient descent optimizer with learning rate 0.001 is exploited for parameter optimization. The PCA dimension for reduction is set to 30. Following the default hyperparameters in VMamba [41], the state dimension and expansion ratio in the 3DSS mechanism are fixed at 16 and 2, respectively.

3) Competitive Approaches: To demonstrate the effectiveness of the proposed 3DSS-Mamba, three kinds of representative HSI classification architectures are selected for comprehensive comparison, including conventional methods (SVM [9]), CNN-based methods (1D-CNN [26], 2D-CNN [49], 3D-CNN [28]), and Transformer-based methods (ViT [34], SF [36], HSI-BERT [35], and DCTN [50]).

a) SVM: The algorithm is implemented with the radial basis function (RBF) kernel on the libsvm toolbox platform. In the RBF, two hyperparameters σ and λ are optimally determined through fivefold cross-validation on the training set in the range of $\sigma = [2^{-3}, 2^{-2}, \ldots, 2^{4}]$ and $\lambda = [10^{-2}, 2^{-1}, \ldots, 10^{4}]$, respectively.

b) 1-D CNN: This model aims to capture spectral features along the spectral dimension, which consists of two 1-D convolutional blocks followed by a fully connected (FC) layer. Each convolutional block comprises a 1-D convolutional layer, a ReLU activation function, and a max-pooling layer. The feature dimensions for the two convolutional layers are set to 32 and 64, respectively, with a kernel size of 1 $\times$ 1.

c) 2-D CNN: This architecture contains two 2-D convolutional blocks and one FC linear layer. Similar to 1-D CNN, the convolution block includes a 2-D conventional layer, a ReLU activation function, and a max-pooling layer. The convolutional layers contain eight and 32 2-D filters with a kernel size of 3 $\times$ 3, respectively. The input patch size is set to 5 $\times$ 5 for all datasets.

d) 3-D CNN: Two 3-D convolution blocks are stacked to extract the spatial–spectral features. Each convolution layer is followed by a ReLU activation function and a max-pooling layer. The final classification output is obtained through one FC linear layer. The feature dimension of the two 3-D convolutional layers are configured to 8 and 16, respectively, with the kernel size of 3 $\times$ 3 $\times$ 3. The input patch size is set to 5 $\times$ 5.

e) ViT (Pixel): The pixel-level ViT utilizes 1-D spectral sequence as input to achieve spectral feature extraction and classification. Three multihead self-attention (MHSA) blocks are stacked, with the feature dimensions set to 32, 64, and 128, respectively.

f) ViT (Patch): The patch-level ViT is extended for spatial–spectral feature extraction from HSI patch cubes. Similarly, it consists of three stacked MHSA blocks, with feature dimensions set to 32, 64, and 128, respectively. The input patch size is 13 $\times$ 13.

g) SF: This architecture focuses on spectral sequence attribute, which consists of five transformer encoders with each incorporating a cross-layer adaptive fusion module. The feature embedding for each encoder is fixed at 64.

h) HSI-BERT: This model comprises a PE block for vector transformation and two stacked bidirectional transformer-based encoders. The learned features are subsequently fed into a single FC linear layer for label prediction.

i) DCTN: This model consists of two stacked transformer-based encoders and four stacked convolution operations. The feature embedding dimensions for transformer encoders are set to 320 and 512.

C. Ablation Study

1) Effectiveness of Different Scanning Routes in 3DSS: Acknowledging the impact of scanning dimension priority on modeling capability, this section explores the effectiveness of constructed five scanning routes, including spectral-priority, spatial-priority, cross spectral–spatial, cross spatial–spectral, and parallel spectral–spatial. The experiments are performed under the setting that the stacked depth of 3DMB is set as 6, with the embedding dimension of 32. This process focuses on comparing the classification performance of different scanning paths under the same settings, without considering the model parameters and computational burden. Table IV illustrates the classification results in terms of accuracy metrics.

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
ABLATION STUDY ON THE ACCURACY METRICS FOR DIFFERENT SCANNING ROUTES IN 3DSS

Routes		Pavia University			Indian Pines			Houston 2013		
		OA(%)	AA(%)	Kappa	OA(%)	AA(%)	Kappa	OA(%)	AA(%)	Kappa
1	Spectral-priority	98.34	97.12	97.81	93.21	84.73	92.25	97.95	98.07	97.78
2	Spatial-priority	99.18	98.14	98.91	95.49	88.06	94.85	98.31	98.39	98.17
3	Cross spectral-spatial	99.07	98.5	98.77	96.16	89.41	95.61	98.5	98.57	98.38
4	Cross spatial-spectral	99.32	98.74	**99.14**	96.16	90.22	95.62	98.84	**99.04**	**98.86**
5	Parallel spectral-spatial	**99.34**	**98.95**	99.12	**96.47**	**93.09**	**95.97**	**98.93**	98.92	98.85

TABLE V
ABLATION STUDY ON THE ACCURACY METRICS FOR DIFFERENT DIMENSIONALITY REDUCTION TECHNIQUES

	Pavia University			Indian Pines			Houston 2013		
	OA (%)	AA (%)	Kappa (%)	OA (%)	AA (%)	Kappa (%)	OA (%)	AA (%)	Kappa (%)
Baseline	95.79	94.72	94.41	92.15	83.44	91.02	95.94	95.84	95.61
MNF	98.61	97.64	98.16	96.16	92.47	95.62	98.56	98.71	98.45
PCA	98.48	97.56	97.98	95.82	90.83	95.23	98.37	98.44	98.24

As can be observed, all scanning routes achieve significant classification performance, demonstrating the superiority of 3DSS in modeling global spectral–spatial contextual relationships. Comparatively, the spatial-priority scanning route demonstrates more competitive advantages than the spectrum-prioritized mechanism. The integration of spatial and spectral information further contributes to the enhancement of classification capability. Notably, the parallel spectral–spatial route showcases the optimal performance across all three datasets, benefiting from both the spatial and spectral priorities with bidirectional modeling. Taking the Pavia University dataset as an example, the parallel spectral–spatial route surpasses the basic spectral-priority by margins of 1.0%, 1.83%, and 1.31% for OA, AA, and Kappa, respectively. As a result, the parallel spectral–spatial route is selected for subsequent experiments.

2) Effectiveness to Different Dimensionality Reduction Techniques: This section explores the effectiveness of different dimensionality reduction techniques for 3DSS-Mamba, including PCA and minimum noise fraction (MNF). Table V illustrates the classification performance on three datasets in terms of OA (%), AA (%), and Kappa (%). As can be observed, both the MNF and PCA techniques are beneficial for enhancing classification performance compared to the Baseline scenario without dimensionality reduction. Notably, MNF demonstrates a marginal improvement over PCA, with an average increase of 0.22% in OA, 0.66% in AA, and 0.26% in the Kappa coefficient. This indicates MNF's slightly superior capability in preserving critical spectral–spatial information while reducing noise, thus further optimizing classification accuracy. PCA transforms the original high-dimensional data into a set of orthogonal components, reducing the overall dimensionality by retaining the most informative variance. However, it does not inherently focus on noise reduction, which could limit its effectiveness in noisy HSI datasets. In contrast, the MNF provides a more noise-aware approach. MNF incorporates the noise covariance matrix, and the feature components are sorted based on the signal-to-noise ratio (SNR) [51], [52]. By maximizing the SNR in the transformed components, MNF reduces dimensionality while prioritizing noise suppression, making it particularly advantageous in challenging noisy scenarios. In our experiments, PCA has been utilized for dimensionality reduction, and we will consider including MNF in future work.

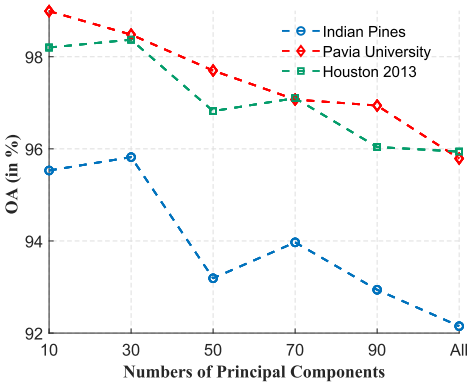

Fig. 6. Sensitivity analysis for the proposed method with different numbers of principal components in PCA.

D. Parameter Analysis

In this section, a series of experiments are carried out to analyze and determine the optimal parameters for 3DSS-Mamba, including the input patch sizes, the 3-D convolution scales in SSTG, and the embedding dims and depths for 3DMB.

1) Different Numbers of Principal Components in PCA: To determine the optimal number of principal components, we conduct ablation experiments by truncating different numbers of principal components. Fig. 6 depicts the classification performance on three datasets, where the number of principal components is set to {10, 30, 50, 70, 90, all(w/oPCA)}. As can be observed, performance initially exhibits progressive improvement as the number of principal components increases on Indian Pines and Houston 2013 datasets. However, when the dimensionality exceeds 30, the Hughes phenomenon may occur, suffering from classification accuracy degradation. Table VI presents the variations in variance contribution and computational efficiency (model Flops and inference time). It can be observed that the cumulative variance contribution

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

TABLE VI
COMPUTATIONAL ANALYSIS WITH DIFFERENT NUMBERS OF PRINCIPAL COMPONENTS IN TERMS OF VARIANCE CONTRIBUTION, FLOPS, AND INFERENCE TIME

	Metrics	PCA=10	PCA=30	PCA=50	PCA=70	PCA=90	All (w/o PCA)
Pavia University	Variance (%)	99.814	99.964	99.990	99.998	99.999	100.00
	Flops (G)	0.0039	0.0139	0.0239	0.0339	0.0438	0.0503
	Inference time (s)	1.53	4.50	8.11	11.29	14.52	17.11
Indian Pines	Variance (%)	96.917	99.248	99.730	99.910	99.977	100.00
	Flops (G)	0.0065	0.0230	0.0395	0.056	0.0725	0.1632
	Inference time (s)	0.54	1.83	2.87	4.37	5.72	13.06
Houston 2013	Variance (%)	99.950	99.989	99.995	99.997	99.998	100.00
	Flops (G)	0.0065	0.0230	0.0395	0.0560	0.0725	0.1013
	Inference time (s)	0.77	2.61	4.31	6.21	8.10	11.43

exhibits slight fluctuations. Even with the principal component dimensionality reduced to 10, almost all information is preserved, with the variance contribution reaching 99.950% on Houston 2013 dataset. Meanwhile, fewer principal component dimensions generally contribute to higher computational efficiency. Therefore, we select the top 30 principal components for 3DSS-Mamba, which strikes a balance between classification performance and computational efficiency.

The effectiveness of PCA in HSI classification is primarily attributed to the following properties.

1) Mitigating the Hughes Phenomenon [47]: classification accuracy gradually increases in the beginning as the number of spectral bands or dimensions increases, but decreases dramatically when the band number reaches some value. PCA employs orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables, effectively reducing the number of spectral bands and thus mitigating the curse of dimensionality.
2) By focusing on the lower order principal components that encapsulate the most significant variance, PCA effectively filters out the noise and redundant information, while preserving more informative and compact spectral components. Consequently, the patch-level 3DSS-Mamba trained on PCA-transformed data tends to capture the most informative and robust spectral–spatial representations, contributing to both classification performance and computational efficiency.

2) Different Input Patch Sizes: Fig. 7 depicts the classification performance across different input patch sizes, ranging from 9×9 to 17×17 with a growth interval of 2. As observed from the figures, the Indian Pines and Houston 2013 exhibit a similar variation tendency of consistently increasing and then decreasing, with the maximum peak at the identified point 13×13. For the Pavia University dataset, accuracy diminishes as the input patch size increases to 11. Accordingly, a patch size of 13×13 is leveraged for the Indian Pines and Houston 2013 datasets, and 11×11 is employed for the Pavia University dataset.

3) Different Scales of 3-D Convolution Kernel in SSTG: The generated 3-D spectral–spatial tokens are determined by the scale of 3-D convolution kernels within the SSTG module. Fig. 8 illustrates the classification sensitivity achieved with distinct 3-D kernels on the three datasets. It can be

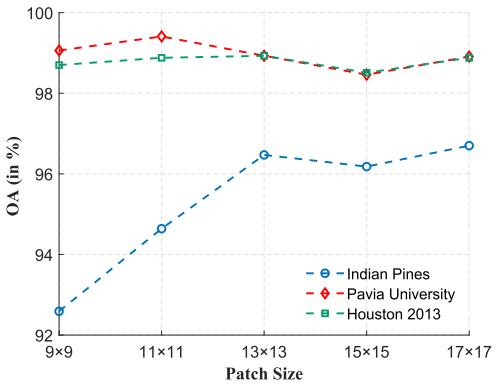

Fig. 7. Sensitivity analysis for the proposed method with different sizes of input patches.

observed that appropriately increasing the scale of 3-D_Conv kernels contributes to capturing richer spectral–spatial contextual information. Based on the results, the optimal kernel scale for all three datasets is established as 32 & [3, 5, 5].

4) Different Embedding Dims and Depths for 3DMB: As the core of 3DSS-Mamba, the 3DMB module based on 3DSS scanning mechanism is iteratively stacked to achieve the extraction of global spectral–spatial semantic representations. To explore the optimal structure of 3DSS-Mamba for classification, mixed experiments are carried out by simultaneously adjusting the embedding dimensions in 3DSS and the stacked depth of 3DMB. Fig. 9 demonstrates the classification sensitivity on three datasets. The range of embedding dim is set to {8, 16, 24, 32, 48}, and the depth covers an interval of {1, 3, 6, 12, 18}. The optimal combination is highlighted with a green point. As observed across all three datasets, lower embedding dimensions can lead to performance degradation due to underfitting. Conversely, excessively high dimensions and deeper depths provide limited accuracy improvements but computational burdens. As can be observed in Table VII, the model parameters and Flops exhibit a gradually increasing trend with the rise of stacking depth on all the datasets. When the depth is set to 1, the model achieves the highest computational efficiency, with the fewest parameters and Flops. By trading off these metrics, the proposed 3DSS-Mamba is constructed as a lightweight structure, where the embedding dimension is determined as 32, and the stacked depth is selected as 1. The feature dimensions of the linear

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

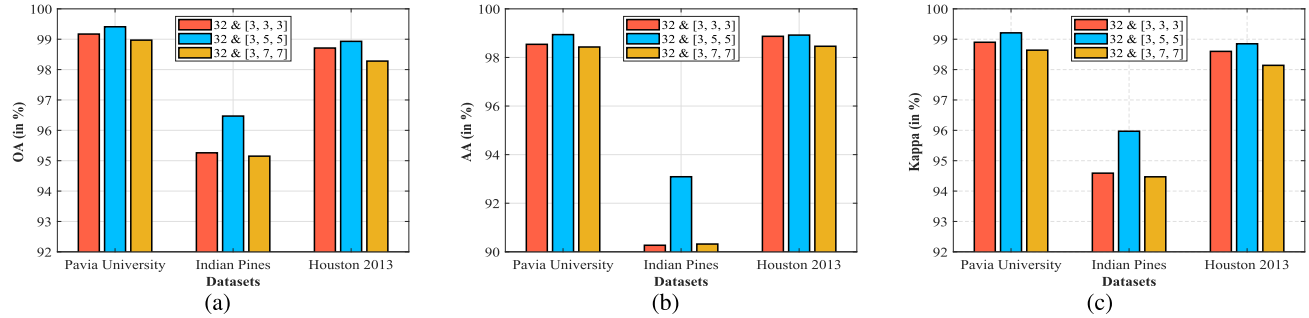

Fig. 8. Sensitivity analysis for the proposed method with different scales of 3-D_Conv kernel in SSTG in terms of (a) OA, (b) AA, and (c) Kappa.

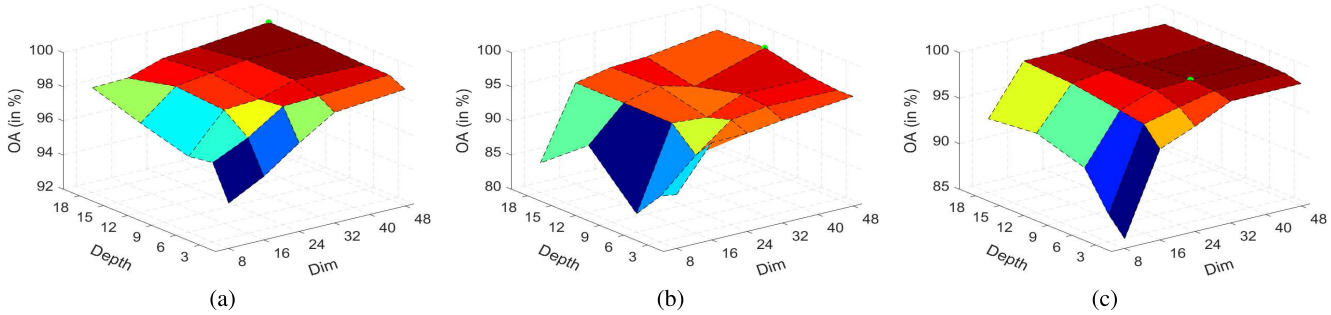

Fig. 9. Sensitivity analysis for the proposed method with different embedding dims and depths for 3DMB. (a) Pavia University. (b) Indian Pines. (c) Houston 2013.

TABLE VII
COMPUTATIONAL ANALYSIS WITH DIFFERENT DEPTHS FOR 3DMB IN TERMS OF MODEL PARAMETERS AND FLOPS

Depth	Pavia University		Indian Pines		Houston 2013	
	Params (M)	Flops (G)	Params (M)	Flops (G)	Params (M)	Flops (G)
1	0.0103	0.0139	0.0105	0.0231	0.0105	0.023
3	0.0231	0.032	0.0233	0.0529	0.0233	0.0529
6	0.0423	0.0591	0.0425	0.0978	0.0425	0.0978
12	0.0807	0.1134	0.0809	0.1875	0.0809	0.1875
18	0.1191	0.1677	0.1193	0.2772	0.1193	0.2772

layers in (10), (11), and (13) are set to be 2, 2, and 1 times the embedding dimension, respectively, i.e., 64, 64, and 32. The dimension for $\Phi_{\text{classifier}}$ in (15) is consistent with the embedding dimension.

E. Experimental Comparison With Competitive Approaches

The quantitative accuracies [OA (%), AA (%), and Kappa (%)] on the Pavia University, Indian Pines, and Houston 2013 datasets are summarized in Tables VIII–X, with the best results highlighted in bold. Corresponding visualization maps are provided in Figs. 10–12.

1) Pavia University Dataset: The classification experiments on the Pavia University dataset are conducted with 5% of the reference samples. Table VIII provides the quantitative comparison results with each competitive approach. As can be observed, the proposed 3DSS-Mamba yields the most competitive performance in comparison with other studied methods. Restricted by hand-crafted feature descriptors, conventional approaches exhibit limitations in handling HSI data with complex contextual semantics, resulting in unsatisfactory classification performance. Despite achieving encouraging results, CNN-based models struggle to establish long-range dependencies due to their limited receptive fields. The Transformer-based architecture, such as DCTN, employs a dual-branch convolutional Transformer architecture with interactive self-attention, enabling the integration of local and global spectral–spatial information. Comparatively, 3DSS-Mamba performs global spectral–spatial semantic extraction from the 3-D sequence modeling perspective. Compared to the suboptimal DCTN, 3DSS-Mamba achieves quantitative improvements in terms of OA, AA, and Kappa up to 2.25%, 3.08%, and 3.0%, respectively. Notably, the proposed 3DSS-Mamba exhibits significantly enhanced classification performance in challenging cases such as classes 3 and 7, potentially attributed to its selective 3-D spectral–spatial contextual modeling capabilities. Mamba is established upon the selective SSM (S6). Compared to the Transformer processing all inputs at each time step without distinction, Mamba allows the projection matrices to be modified as input-dependent, which achieves selective attention on each sequence unit. This selective mechanism enables Mamba to highlight the most critical and informative spectral–spatial

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

TABLE VIII

CLASSIFICATION ACCURACIES OF THE COMPARED METHODS IN TERMS OF OA, AA, AND κ, AND THE ACCURACIES OF EACH CLASSES FOR THE PAVIA UNIVERSITY DATASET. THE BEST ACCURACIES ARE PRESENTED IN BOLD

Class	Conventional	CNN-based Methods			Transformer-based Methods					3DSS-Mamba
	SVM	1D-CNN	2D-CNN	3D-CNN	VIT(Pixel)	VIT(Patch)	SF	HSI-BERT	DCTN	
1	91.63	83.11	96.31	95.94	82.87	92.46	94.75	97.63	95.36	**99.33**
2	97.56	92.78	99.37	99.71	94.00	96.53	98.14	**99.93**	99.51	99.34
3	73.90	60.03	82.22	89.76	75.91	93.73	84.70	86.51	**96.94**	95.09
4	92.10	85.57	94.35	97.36	83.39	95.64	96.81	97.53	**99.45**	95.91
5	98.51	97.77	99.85	**100.0**	99.14	**100.0**	99.22	99.61	98.44	99.69
6	86.04	70.77	93.65	97.41	51.49	98.47	94.37	93.05	87.80	**99.43**
7	84.09	44.21	89.17	**96.69**	41.34	94.62	84.56	95.80	79.18	95.88
8	90.25	66.38	88.73	93.92	76.48	**98.60**	92.65	97.86	94.05	96.20
9	99.22	80.89	97.57	**100.0**	99.34	98.11	**100.0**	99.67	99.56	97.22
OA (%)	92.75	82.68	95.77	97.62	82.77	96.18	96.60	97.61	96.23	**98.48**
AA (%)	90.37	75.72	93.47	96.75	78.22	96.46	93.91	96.40	94.48	**97.56**
Kappa	90.35	76.92	94.37	96.84	76.77	94.97	94.17	96.82	94.98	**97.98**

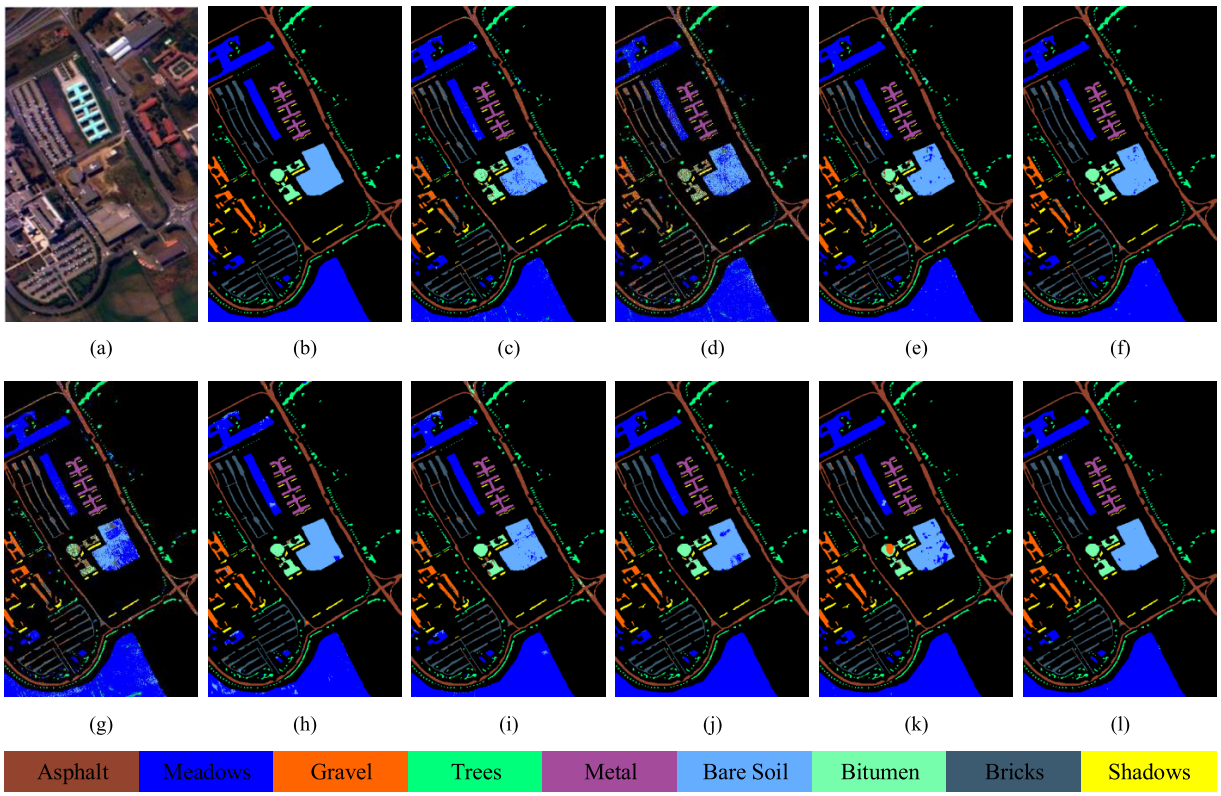

Fig. 10. Classification maps obtained by the compared methods on the Pavia University dataset. (a) Original data. (b) Reference map. (c) SVM. (d) 1D-CNN. (e) 2D-CNN. (f) 3D-CNN. (g) ViT (Pixel). (h) ViT (Patch). (i) SF. (j) HSI-BERT. (k) DCTN. (l) Proposed 3DSS-Mamba.

patterns while suppressing sensitivity to irrelevant or misleading information. Additionally, the proposed 3-D spectral–spatial scanning mechanism enables 3DSS-Mamba to integrate the 3-D characteristics of HSIs adequately. By considering different dimensionality priorities, pixel interactions among adjacent spatial and spectral positions can be effectively facilitated. This is crucial for accurate classification within the challenging scenarios.

The corresponding classification maps obtained by different approaches are visualized in Fig. 10. As can be observed, conventional architecture such as SVM generally brings noticeable noises, which can be attributed to the limited feature extraction capability of hand-crafted descriptors. Typical CNN-based and Transformer-based approaches such as 1D-CNN and ViT (Pixel) suffer from significant misclassifications, particularly evident in the bare soil category. In contrast, the proposed 3DSS-Mamba achieves the most consistent results with the ground truth, which presents the clearest category boundaries with minimal noise.

2) Indian Pines Dataset: The experiments on the Indian Pines dataset are performed with 10% of the reference samples. The quantitative classification accuracies are reported in Table IX. Based on the results, the proposed 3DSS-Mamba achieves the highest recognition performance in comparison with other competitive approaches, and exhibits excellent

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

TABLE IX

CLASSIFICATION ACCURACIES OF THE COMPARED METHODS IN TERMS OF OA, AA, AND κ, AND THE ACCURACIES OF EACH CLASSES FOR THE INDIAN PINES DATASET. THE BEST ACCURACIES ARE PRESENTED IN BOLD

Class	Conventional	CNN-based Methods			Transformer-based Methods					3DSS-Mamba
	SVM	1D-CNN	2D-CNN	3D-CNN	VIT(Pixel)	VIT(patch)	SF	HSI-BERT	DCTN	
1	20.73	15.22	76.09	95.65	0.00	70.73	**100.0**	48.78	78.05	78.05
2	73.07	64.43	**92.16**	88.31	36.65	59.22	78.37	78.44	92.14	90.27
3	62.99	50.48	87.23	84.94	0.94	60.51	89.29	80.58	**97.32**	93.04
4	50.70	28.69	64.98	87.76	4.23	90.14	81.22	51.17	91.08	**93.43**
5	92.64	76.60	96.48	96.48	20.46	64.14	89.89	91.24	**97.93**	97.70
6	94.90	85.89	98.77	96.16	94.22	97.72	97.41	96.19	**99.39**	97.26
7	76.00	32.14	32.14	78.57	0.00	**100.0**	84.00	20.00	96.00	96.00
8	96.63	86.40	**100.0**	**100.0**	99.53	97.91	**100.0**	98.83	**100.0**	99.30
9	33.33	15.00	30.00	**70.00**	0.00	0.00	27.78	0.00	55.56	38.89
10	68.69	60.08	91.87	92.80	33.26	85.71	94.97	78.37	57.83	**96.80**
11	85.16	69.61	95.93	91.20	88.96	92.81	93.44	92.85	92.17	**98.91**
12	64.89	56.32	**93.42**	88.03	4.12	90.07	81.65	62.47	92.88	91.39
13	97.03	89.27	97.56	94.15	82.70	98.38	87.03	96.19	**100.0**	95.68
14	96.49	79.45	98.81	99.05	98.24	97.01	94.82	94.46	**99.91**	99.65
15	55.33	49.48	81.35	89.38	11.53	**99.71**	96.83	87.89	89.91	94.52
16	93.93	40.86	97.85	**100.0**	91.67	**100.0**	95.24	65.06	82.14	84.52
OA (%)	79.82	67.13	93.34	92.17	57.35	84.55	90.67	85.45	91.31	**95.82**
AA (%)	72.03	56.25	83.41	90.78	41.65	81.50	87.00	71.41	88.90	**90.33**
Kappa	76.84	62.44	92.39	91.08	48.99	82.32	89.36	83.36	90.12	**95.23**

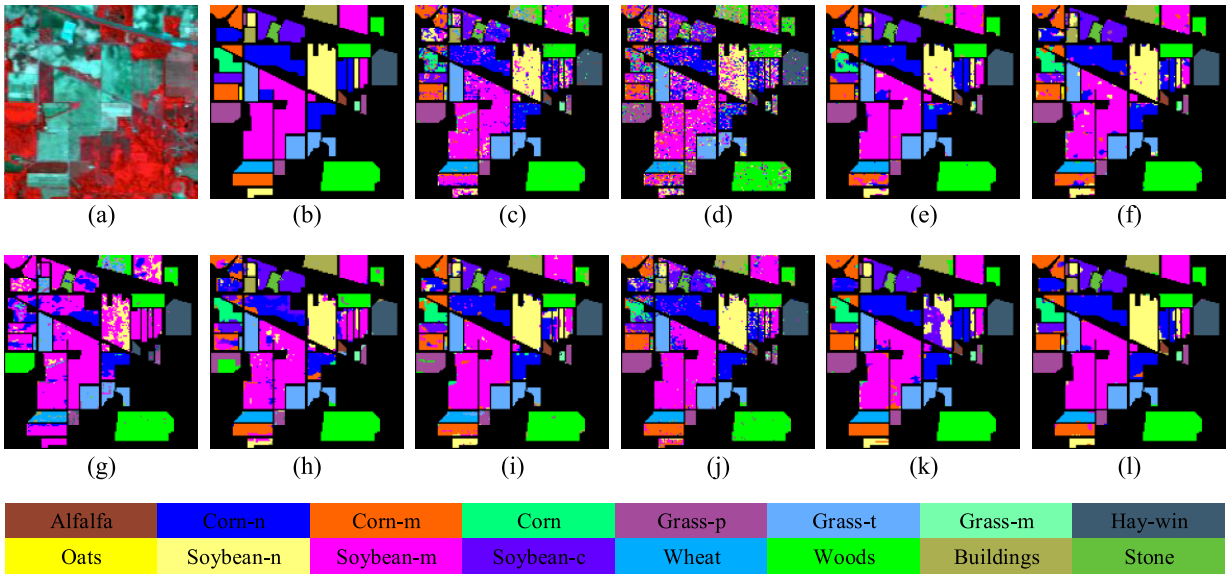

Fig. 11. Classification maps obtained by the compared methods on the Indian Pines dataset. (a) Original data. (b) Reference map. (c) SVM. (d) 1D-CNN. (e) 2D-CNN. (f) 3D-CNN. (g) ViT (Pixel). (h) ViT (Patch). (i) SF. (j) HSI-BERT. (k) DCTN. (l) Proposed 3DSS-Mamba.

improvements in several categories, such as Corn-notill, Corn-mintil, and Soybean-notill. Due to the lack of consideration of spatial information, the Transformer-based ViT (Pixel) method inevitably suffers from undesirable classification consequences, with the reduction compared to 3DSS-Mamba reaching 38.47% for OA, 49.18% for AA, and 46.24% for Kappa, respectively. To highlight the differences in classification results, Fig. 11 further provides the visualization maps of various methods. Benefiting from the extraction of global spectral–spatial semantic information with the 3-D sequential modeling mechanism, the proposed 3DSS-Mamba exhibits the smoothest and clearest classifications across most regions despite slight edge information confusion. These phenomena further reveal the effectiveness and preeminence of the proposed method.

3) Houston 2013 Dataset: The experiments on the Houston 2013 dataset are executed with 10% of the labeled samples. As evident from the classification results in Table X, the proposed 3DSS-Mamba consistently outperforms other techniques by substantial margins, demonstrating the highest quantities across all three metrics. In contrast to the suboptimal DCTN, 3DSS-Mamba achieves significant improvements in OA, AA, and Kappa by 1.05%, 1.27%, and 1.13%, respectively. Regarding the visualization maps in Fig. 12, 3DSS-Mamba provides the most precise prediction details, even though this scenario is predominantly distributed with discrete and local sample targets. These excellent improvements further verify the potential application of the sequence scanning model in HSI classification.

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

TABLE X

CLASSIFICATION ACCURACIES OF THE COMPARED METHODS IN TERMS OF OA, AA, AND κ, AND THE ACCURACIES OF EACH CLASSES FOR THE HOUSTON 2013 DATASET. THE BEST ACCURACIES ARE PRESENTED IN BOLD

Class	Conventional	CNN-based Methods			Transformer-based Methods					3DSS-Mamba
	SVM	1D-CNN	2D-CNN	3D-CNN	VIT	VIT(patch)	SF	HIS-BERT	DCTN	
1	98.53	90.25	98.40	97.04	93.34	91.56	**99.29**	92.26	98.49	99.02
2	98.05	95.14	99.20	98.48	99.29	94.69	98.05	83.15	**99.91**	99.56
3	98.88	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	99.20	99.84	**100.0**
4	97.81	93.89	98.31	**98.95**	87.41	98.48	96.43	93.47	96.43	97.68
5	98.57	93.40	99.28	99.44	99.37	92.75	**100.0**	99.73	**100.0**	99.91
6	92.64	99.38	92.62	**100.0**	97.26	**100.0**	95.89	78.76	99.32	98.29
7	91.98	80.84	94.01	93.38	85.28	97.11	96.06	91.06	96.32	**97.55**
8	91.96	70.26	89.63	94.29	67.50	91.52	91.70	86.68	95.09	**96.16**
9	85.09	67.65	83.87	93.61	70.63	87.67	94.59	89.60	**98.23**	96.27
10	93.16	67.64	94.62	96.33	79.89	89.04	97.46	89.03	95.65	**99.55**
11	87.01	69.64	90.53	96.84	58.63	93.88	90.56	95.31	97.21	**98.20**
12	87.25	62.94	89.38	95.46	51.44	87.84	**99.28**	88.63	96.31	98.38
13	32.35	49.04	77.83	94.24	6.64	**96.68**	56.40	93.12	84.83	**96.68**
14	98.91	97.20	99.30	98.60	79.48	99.74	98.44	97.92	**100.0**	99.48
15	85.08	99.09	**100.0**	99.85	98.65	99.66	99.66	99.66	**100.0**	**100.0**
OA (%)	91.74	81.06	93.93	96.77	79.27	93.63	95.46	91.67	97.32	**98.37**
AA (%)	90.03	82.42	93.80	97.10	78.32	94.70	94.25	91.84	97.17	**98.44**
Kappa	91.06	79.51	93.44	96.51	77.54	93.12	95.09	90.99	97.11	**98.24**

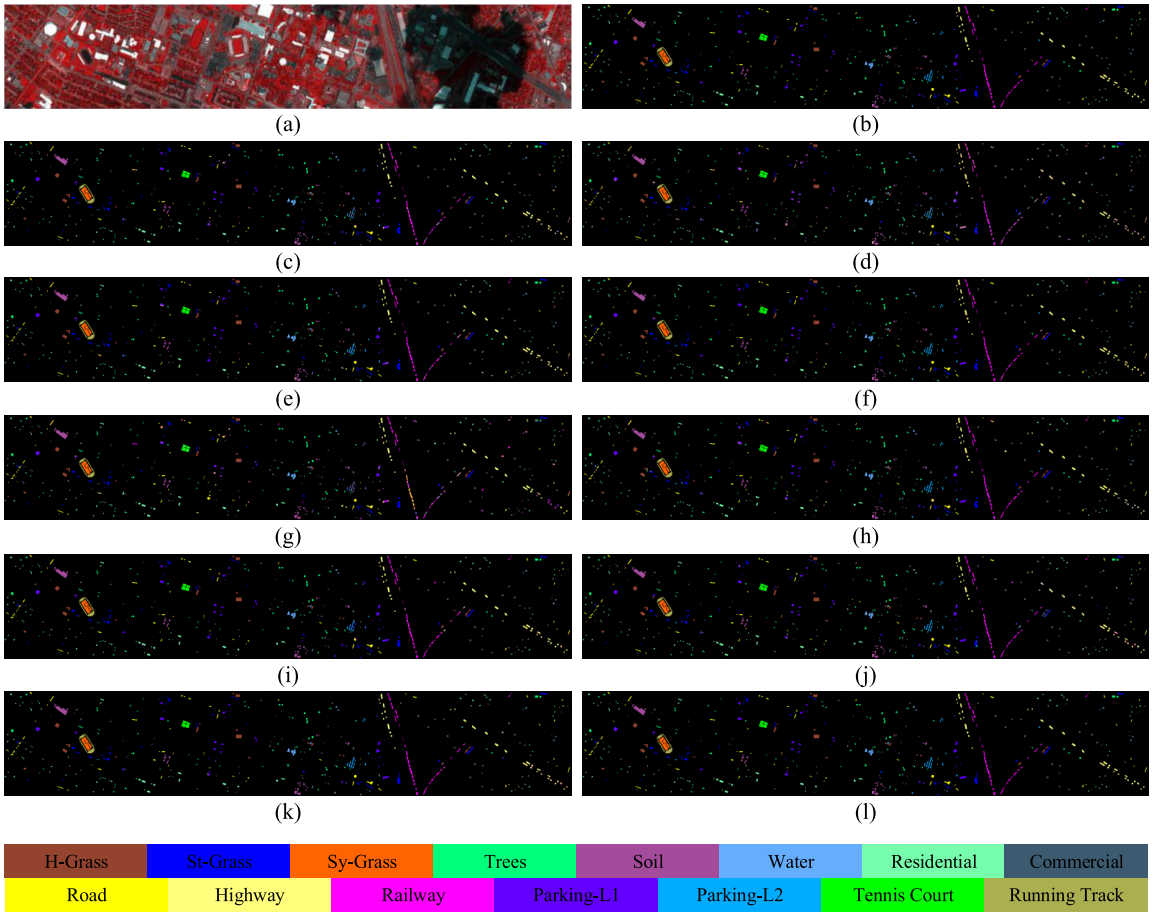

Fig. 12. Classification maps obtained by the compared methods on the Houston 2013 dataset. (a) Original data. (b) Reference map. (c) SVM. (d) 1D-CNN. (e) 2D-CNN. (f) 3D-CNN. (g) ViT (Pixel). (h) ViT (Patch). (i) SF. (j) HSI-BERT. (k) DCTN. (l) Proposed 3DSS-Mamba.

F. Robustness Assessment

To demonstrate the robustness of the proposed 3DSS-Mamba, extensive experiments are conducted considering various proportions of training samples. Specifically, the selected percentage for the Indian Pines and Houston 2013 datasets covers an interval of $\{1.0\%, 2.0\%, \ldots, 10.0\%\}$, and for the Pavia University is set to $\{0.5\%, 1.0\%, \ldots, 5.0\%\}$. Fig. 13 displays the performance variations with different percentages

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

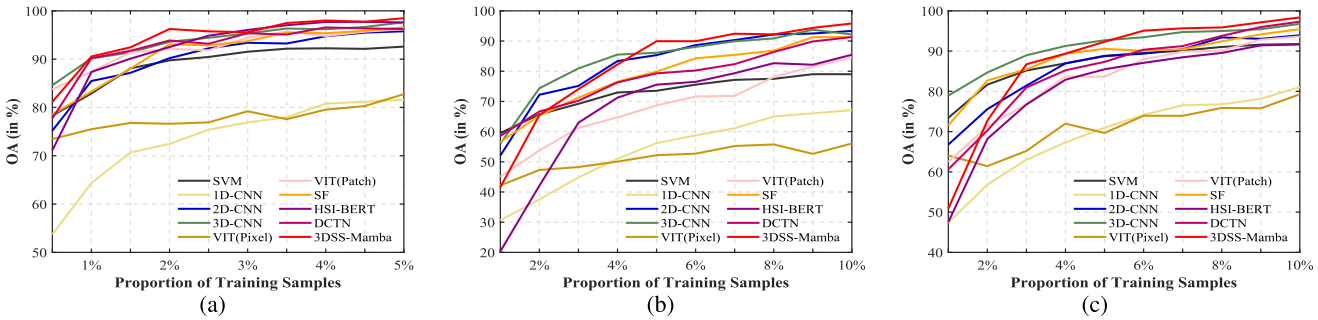

Fig. 13. Classification performance of the compared methods with different numbers of labeled data on three datasets. (a) Pavia University. (b) Indian Pines. (c) Houston 2013.

TABLE XI
COMPUTATIONAL ANALYSIS OF VARIOUS COMPARISON APPROACHES IN TERMS OF PARAMETERS, FLOPS, AND INFERENCE TIME

	Metrics	1D-CNN	2D-CNN	3D-CNN	VIT(Pixel)	VIT(Patch)	SF	HIS-BERT	DCTN	3DSS-Mamba
Pavia University	Params (M)	0.0083	0.0101	0.0073	0.1106	0.1106	0.1222	0.3006	11.057	0.0103
	Flops (G)	2.3008	0.0001	0.0012	0.0002	0.0185	0.0126	5.441	0.2077	0.0139
	Inference time (s)	0.50	0.72	0.76	1.64	3.40	6.55	36.13	5.19	4.50
Indian Pines	Params (M)	0.0137	0.0173	0.0165	0.1147	0.1146	0.1226	0.3951	11.067	0.0105
	Flops (G)	3.4560	0.0003	0.0024	0.0002	0.0191	0.0245	7.1523	0.2818	0.0230
	Inference time (s)	0.12	0.20	0.24	0.57	0.97	1.82	9.02	1.95	1.83
Houston 2013	Params (M)	0.0115	0.0132	0.0123	0.1127	0.1127	0.1133	0.3773	11.0634	0.0105
	Flops (G)	2.8864	0.0002	0.0017	0.0002	0.0135	0.0163	6.8135	0.239	0.0230
	Inference time (s)	0.32	0.25	0.42	0.84	1.03	1.77	12.21	5.37	2.61

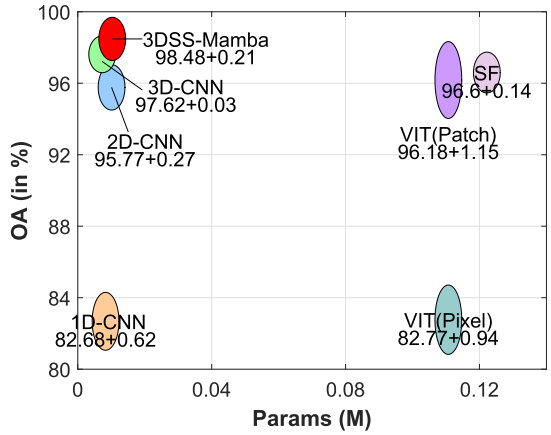

Fig. 14. Computational parameter analysis of various comparison approaches on the Pavia University dataset.

on the four HSI datasets, with the proposed 3DSS-Mamba highlighted by red curves. There is a basically reasonable trend that the classification accuracy of 3DSS-Mamba steadily increases with the percentages of training samples, which exhibit substantial robustness. Furthermore, 3DSS-Mamba consistently outperforms other competitive methods across most percentages on the three testing datasets. These notable advantages further demonstrate the feasibility and superiority of the proposed method.

G. Analysis of Computational Complexity

This section investigates the computational complexity of the proposed 3DSS-Mamba, focusing on model parameters, Flops, and inference time. Fig. 14 illustrates the model parameter sizes of various comparison approaches on the Pavia University dataset. Attributed to the inherent linear sequential modeling mechanism, the proposed 3DSS-Mamba achieves the best classification performance with significantly fewer computational parameters. The variance is calculated by repeating the experiments five times, which achieves a lower fluctuation of 0.21, indicating the superiority and stability of our method. Although the CNN-based methods enjoy slight computational burdens, their capability of capturing long-range dependencies is constrained by their local receptive field, restricting further performance improvement. Besides, Transformer-based methods generally suffer from higher resource consumption due to the series of MHSA modules, and are associated with larger variance fluctuations in performance. While ViT (Patch) and SF methods provide competitive performance, their model parameters are almost ten times that of 3DSS-Mamba. The DCTN suffers from higher resource consumption due to its dual-branch architecture and MHSA module. Table XI further presents the detailed model parameters, Flops, and inference time. The proposed 3DSS-Mamba consistently demonstrates lower computational parameters and Flops across various datasets. The inference time is also comparatively acceptable relative to other approaches. In summary, the proposed 3DSS-Mamba exhibits competitive advantages in balancing computational efficiency and classification effectiveness, highlighting significant potentiality and viability for HSI classification tasks.

IV. CONCLUSION

In this article, we introduce 3DSS-Mamba, a novel architecture based on the SSMs for HSI classification. Benefiting from the integrated SSTG module and 3DSS mechanism, 3DSS-Mamba achieves the substantial advantages of both

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

global spectral–spatial contextual modeling and linear computational complexity from the sequence modeling perspective. Extensive experiments demonstrate that the proposed 3DSS-Mamba efficiently breaks the performance and efficiency bottlenecks of state-of-the-art CNN-based and Transformer-based HSI architectures. This research offers a feasible solution for the HSI classification task. Future work will endeavor to explore the scalability of the Mamba model across a wider range of hyperspectral scenarios.

REFERENCES

[1] L. Ni, H. Xu, and X. Zhou, "Mineral identification and mapping by synthesis of hyperspectral VNIR/SWIR and multispectral TIR remotely sensed data with different classifiers," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3155–3163, 2020.

[2] K. Siebels, K. Goïta, and M. Germain, "Estimation of mineral abundance from hyperspectral data using a new supervised neighbor-band ratio unmixing approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 6754–6766, Oct. 2020.

[3] J.-P. Ardouin, J. Lévesque, and T. A. Rea, "A demonstration of hyperspectral image exploitation for military applications," in *Proc. 10th Int. Conf. Inf. Fusion*, Jul. 2007, pp. 1–8.

[4] S. Peyghambari and Y. Zhang, "Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: An updated review," *J. Appl. Remote Sens.*, vol. 15, no. 3, Jul. 2021, Art. no. 031501.

[5] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.

[6] X. Li, M. Ding, Y. Gu, and A. Pižurica, "An end-to-end framework for joint denoising and classification of hyperspectral images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 7, pp. 3269–3283, Jul. 2023.

[7] 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.

[8] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[9] 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.

[10] H. Huang, G. Shi, H. He, Y. Duan, and F. Luo, "Dimensionality reduction of hyperspectral imagery based on spatial–spectral manifold learning," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2604–2616, Jun. 2020.

[11] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 55–66, Jan. 2014.

[12] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.

[13] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5975–5991, Dec. 2010.

[14] Y. Duan, H. Huang, and T. Wang, "Semisupervised feature extraction of hyperspectral image using nonlinear geodesic sparse hypergraphs," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5515115.

[15] L. Gao, J. Li, K. Zheng, and X. Jia, "Enhanced autoencoders with attention-embedded degradation learning for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5509417.

[16] S. Mei, J. Ji, Y. Geng, Z. Zhang, X. Li, and Q. Du, "Unsupervised spatial–spectral feature learning by 3D convolutional autoencoder for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6808–6820, Sep. 2019.

[17] H. Sun, X. Zheng, X. Lu, and S. Wu, "Spectral–spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3232–3245, 2019.

[18] M. Zhu, L. Jiao, F. Liu, S. Yang, and J. Wang, "Residual spectral–spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 449–462, May 2020.

[19] X. Li, M. Ding, and A. Pižurica, "Deep feature fusion via two-stream convolutional neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2615–2629, Apr. 2020.

[20] G. Yue, L. Zhang, Y. Zhou, Y. Wang, and Z. Xue, "S2TNet: Spectral–spatial triplet network for few-shot hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.

[21] X. Li, Y. Gu, and A. Pižurica, "A unified multiview spectral feature learning framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5540614.

[22] W.-S. Hu, H.-C. Li, L. Pan, W. Li, R. Tao, and Q. Du, "Spatial–spectral feature extraction via deep ConvLSTM neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4237–4250, Jun. 2020.

[23] Q. Liu, F. Zhou, R. Hang, and X. Yuan, "Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification," *Remote Sens.*, vol. 9, no. 12, p. 1330, Dec. 2017.

[24] X. Liao, B. Tu, J. Li, and A. Plaza, "Class-wise graph embedding-based active learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5522813.

[25] Y. Su et al., "ACGT-Net: Adaptive cuckoo refinement-based graph transfer network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5521314.

[26] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, pp. 1–12, Nov. 2015.

[27] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Learning and transferring deep joint spectral–spatial features for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4729–4742, Aug. 2017.

[28] 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.

[29] 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.

[30] S. K. Roy, A. Deria, C. Shah, J. M. Haut, Q. Du, and A. Plaza, "Spectral–spatial morphological attention transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5503615.

[31] J. Zhang, Z. Meng, F. Zhao, H. Liu, and Z. Chang, "Convolution transformer mixer for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[32] E. Ouyang, B. Li, W. Hu, G. Zhang, L. Zhao, and J. Wu, "When multigranularity meets spatial–spectral attention: A hybrid transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4401118.

[33] Y. He, B. Tu, B. Liu, Y. Chen, J. Li, and A. Plaza, "Hybrid multiscale spatial–spectral transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5527918.

[34] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent.*, 2020.

[35] J. He, L. Zhao, H. Yang, M. Zhang, and W. Li, "HSI-BERT: Hyperspectral image classification using the bidirectional encoder representation from transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 165–178, Sep. 2019.

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

[37] Z. Zhong, Y. Li, L. Ma, J. Li, and W.-S. Zheng, "Spectral–spatial transformer network for hyperspectral image classification: A factorized architecture search framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5514715.

[38] 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.

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

[40] 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. 41st Int. Conf. Mach. Learn.*, 2024.

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.

[41] Y. Liu et al., "VMamba: Visual state space model," 2024, *arXiv:2401.10166*.
[42] A. Gu, K. Goel, and C. Re, "Efficiently modeling long sequences with structured state spaces," in *Proc. Int. Conf. Learn. Represent.*, 2021.
[43] G. Pechlivanidou and N. Karampetakis, "Zero-order hold discretization of general state space systems with input delay," *IMA J. Math. Control Inf.*, vol. 39, no. 2, pp. 708–730, Jun. 2022.
[44] Z. He and Y.-H. Wang, "Mamba meets crack segmentation," 2024, *arXiv:2407.15714*.
[45] H. Zheng, Z. Yang, W. Liu, J. Liang, and Y. Li, "Improving deep neural networks using softplus units," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2015, pp. 1–4.
[46] S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural Netw.*, vol. 107, pp. 3–11, Nov. 2018.
[47] W. Ma, C. Gong, Y. Hu, P. Meng, and F. Xu, "The Hughes phenomenon in hyperspectral classification based on the ground spectrum of grasslands in the region around Qinghai lake," *Proc. SPIE*, vol. 8910, pp. 363–373, Aug. 2013.
[48] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.
[49] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
[50] Y. Zhou, X. Huang, X. Yang, J. Peng, and Y. Ban, "DCTN: Dual-branch convolutional transformer network with efficient interactive self-attention for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5508616.
[51] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Feb. 1988.
[52] G. Lixin, X. Weixin, and P. Jihong, "Segmented minimum noise fraction transformation for efficient feature extraction of hyperspectral images," *Pattern Recognit.*, vol. 48, no. 10, pp. 3216–3226, Oct. 2015.

Yan He (Student Member, IEEE) received the M.S. degree in computer science from Hunan University, Changsha, Hunan, China, in 2020.

Her research focuses on hyperspectral image processing and remote sensing image registration.

Bing Tu (Senior Member, IEEE) received the M.S. degree in control science and engineering from Guilin University of Technology, Guilin, China, in 2009, and the Ph.D. degree in mechatronic engineering from Beijing University of Technology, Beijing, China, in 2013.

From 2015 to 2016, he was a Visiting Researcher with the Department of Computer Science and Engineering, University of Nevada, Reno, NV, USA, which is supported by China Scholarship Council. He is currently a Full Professor with Nanjing University of Information Science and Technology, Nanjing, China. His research interests include sparse representation, pattern recognition, and analysis in remote sensing.

Dr. Tu is an Associate Editor of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Bo Liu (Member, IEEE) received the B.S. and Ph.D. degrees in optical engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2008 and 2013, respectively.

He is currently a Professor with the School of Physics and Optoelectronics, Nanjing University of Information Science and Technology (NUIST), Nanjing, China. His research interests include all optical signal processing, radio over fiber, and broadband optical communication.

Jun Li (Fellow, IEEE) received the B.S. degree in geographic information systems from Hunan Normal University, Changsha, China, in 2004, the M.E. degree in remote sensing from Peking University, Beijing, China, in 2007, and the Ph.D. degree in electrical engineering from the Instituto de Telecomunicações, Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisbon, Portugal, in 2011.

From 2013 to 2021, she was a Full Professor with Sun Yat-sen University, Guangzhou, China. Since 2022, she has been with China University of Geosciences, Wuhan, China, as a Full Professor. She has received several prestigious funding grants at the national and international levels. She has authored more than 160 journal citation report (JCR) articles, 60 international conference papers, and a book chapter.

Dr. Li has been serving as the Editor-in-Chief for the *Journal of Selected Topics in Applied Earth Observations and Remote Sensing* since 2021.

Antonio Plaza (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura, Cáceres, Spain, in 1999 and 2002, respectively.

He is currently the Head of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura. He has authored more than 600 publications, including over 200 JCR journal articles (over 160 in IEEE journals), 23 book chapters, and around 300 peer-reviewed conference proceeding papers. His research interests include hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza was a member of the Editorial Board of the IEEE Geoscience and Remote Sensing Newsletter from 2011 to 2012 and *IEEE Geoscience and Remote Sensing Magazine* in 2013. He was also a member of the Steering Committee of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He is also a fellow of IEEE for contributions to hyperspectral data processing and parallel computing of Earth observation data. He was a recipient of the recognition of Best Reviewers of IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2009 and the Best Reviewer of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2010, for which he has served as an Associate Editor from 2007 to 2012. He was also a recipient of the Best Column Award of *IEEE Signal Processing Magazine* in 2015, the 2013 Best Paper Award of the JSTARS journal, and the Most Highly Cited Paper from 2005 to 2010 in *Journal of Parallel and Distributed Computing*. He received the Best Paper Awards at the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology. He has served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) from 2011 to 2012 and the President of the Spanish Chapter of IEEE GRSS from 2012 to 2016. He has reviewed more than 500 papers for over 50 different journals. He has served as the Editor-in-Chief for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2013 to 2017. He has guest edited ten special issues on hyperspectral remote sensing for different journals. He is also an Associate Editor of IEEE ACCESS (receiving the recognition as an Outstanding Associate Editor of the journal in 2017). Additional information is available at http://www.umbc.edu/rssipl/people/aplaza.

Authorized licensed use limited to: National Institute of Technology Andhra Pradesh. Downloaded on May 22,2025 at 09:56:56 UTC from IEEE Xplore. Restrictions apply.