remote sensing

Article

SSUM: Spatial–Spectral Unified Mamba for Hyperspectral Image Classification

Song Lu [1], Min Zhang [1,*], Yu Huo [1], Chenhao Wang [1], Jingwen Wang [2] and Chenyu Gao [2]

[1] School of Aerospace Science and Technology, Xidian University, Xi'an 710126, China
[2] Beijing Research Institute of Telemetry, Beijing 100076, China
[*] Correspondence: minzhang@xidian.edu.cn

Abstract: How to effectively extract spectral and spatial information and apply it to hyperspectral image classification (HSIC) has been a hot research topic. In recent years, the transformer-based HSIC models have attracted much interest due to their advantages in long-distance modeling of spatial and spectral features in hyperspectral images (HSIs). However, the transformer-based method suffers from high computational complexity, especially in HSIC tasks that require processing large amounts of data. In addition, the spatial variability inherent in HSIs limits the performance improvement of HSIC. To handle these challenges, a novel Spectral–Spatial Unified Mamba (SSUM) model is proposed, which introduces the State Space Model (SSM) into HSIC tasks to reduce computational complexity and improve model performance. The SSUM model is composed of two branches, i.e., the Spectral Mamba branch and the Spatial Mamba branch, designed to extract the features of HSIs from both spectral and spatial perspectives. Specifically, in the Spectral Mamba branch, a nearest-neighbor spectrum fusion (NSF) strategy is proposed to alleviate the interference caused by the spatial variability (i.e., same object having different spectra). In addition, a novel sub-spectrum scanning (SS) mechanism is proposed, which scans along the sub-spectrum dimension to enhance the model's perception of subtle spectral details. In the Spatial Mamba branch, a Spatial Mamba (SM) module is designed by combining a 2D Selective Scan Module (SS2D) and Spatial Attention (SA) into a unified network to sufficiently extract the spatial features of HSIs. Finally, the classification results are derived by uniting the output feature of the Spectral Mamba and Spatial Mamba branch, thus improving the comprehensive performance of HSIC. The ablation studies verify the effectiveness of the proposed NSF, SS, and SM. Comparison experiments on four public HSI datasets show the superior of the proposed SSUM.

Keywords: hyperspectral image classification; space state model; nearest-neighbor spectrum fusion strategy; sub-spectrum scanning mechanism; Spatial Mamba module

check for updates

Citation: Lu, S.; Zhang, M.; Huo, Y.; Wang, C.; Wang, J.; Gao, C. SSUM: Spatial–Spectral Unified Mamba for Hyperspectral Image Classification. *Remote Sens.* **2024**, *16*, 4653. https://doi.org/10.3390/rs16244653

Academic Editors: Chein-I Chang and Salah Bourennane

Received: 21 October 2024
Revised: 5 December 2024
Accepted: 10 December 2024
Published: 12 December 2024

1. Introduction

A hyperspectral image (HSI) is a special type of remote sensing image that can capture a very wide range of spectral bands [1–3], typically from the near-ultraviolet to the near-infrared wavelengths [1,3]. The resulting three-dimensional (3D) HSI data cube contains nearly continuous spectral profiles for each spatial resolution element [4], thereby allowing for more accurate quantification, identification, and discrimination of the imaged content [5]. This unique characteristic of HSIs sets them apart from traditional remote sensing images and makes them particularly valuable in various applications.

The hyperspectral imaging technology has found application in a variety of practical scenarios [6], encompassing, but not limited to, atmospheric, environmental, urban, agricultural, geological, and mineral exploration [7]. Among these applications, hyperspectral image classification (HSIC) stands out as a pivotal one. HSIC has become a focal point in the domain of remote sensing research, eliciting considerable academic and practical interest [8–10]. The ability to classify and understand the data captured by HSIs is crucial for extracting meaningful information and supporting decision-making processes.

In the early phases of HSIC exploration, there was a notable trend towards utilizing statistical methods. These encompassed methods such as Principal Component Analysis (PCA) [11], Independent Component Analysis (ICA) [12], K-Nearest Neighbors (KNN) [13], and Random Forests [14,15], which were instrumental in the proficient processing and analysis of HSI data. Among these, KNN, a straightforward and intuitive supervised learning algorithm, does not necessitate feature extraction [13]. Amini introduced the application of the Random Forest [14] algorithm in HSIC, which is an ensemble learning algorithm known for its efficiency, stability, robustness, and interpretability, enhancing classification accuracy and stability through the construction of multiple decision trees [14]. Additionally, Fang et al. utilized local covariance matrices to encapsulate the inter-connections amongst various spectral intervals [16], employing these matrices for HSI training and classification using Support Vector Machines (SVMs) [17]. Furthermore, traditional deep learning models encompass stacked autoencoders (SAEs) [18], deep belief networks (DBNs) [19], and others. Moreover, feature extraction techniques such as Extended Morphological Profile (EMP) [20], Extended Multi-Attribute Profile (EMAP) [21], Gabor filtering [22], and sparse representation [23] have also been integrated with various classifiers, forming a diverse array of methods. However, these traditional methods, which solely rely on spectral features without considering spatial information [24], are vulnerable to the effects of spectral variability, thereby restricting classification performance.

The advancement in deep learning (DL) has notably propelled the evolution of hyperspectral image classification (HSIC) tasks. This progress is largely due to the introduction of Convolutional Neural Networks (CNNs), which have been pivotal in this evolution, effectively capturing both spatial and spectral aspects of HSI data [1,25]. These networks are proficient in generating hierarchical representations from HSI data [26], enabling the detection of complex patterns which traditional methods often ignore [27]. Initially, the 1D-CNN [18] architecture was introduced for HSIC, treating the spectrum as the classification subject and employing convolution along the spectral direction to extract features. Roy et al. introduced HybridSN [28], an integrated strategy that merges 2D-CNN and 3D-CNN and is adopted to harness the advantages of each, utilizing 2D-CNN for capturing spatial features and 3D-CNN for extracting spectral features. In a novel approach, Zhong et al. introduced the Spectral–Spatial Residual Network (SSRN) [29]. Within the SSRN framework, residual blocks are designed with identity mappings to bridge all the 3D convolutional layers. Xu et al. introduced the SSUN network [30], integrating spectral, spatial feature extraction, and classification into a unified framework, employing a self-developed MSCNN for spatial feature extraction. Paoletti et al. have proposed the Deep Pyramidal Residual Networks (DPRNs) [31] especially for the HSI data. Zhong et al. introduced SSFCNS [32]; it incorporates reciprocal loops, transforming the CNN into a tightly integrated network configuration, in contrast to the conventional CNN, which operates as a straightforward open feed-forward architecture. Chang et al. introduced IRTS-CNN [33], which significantly enhances HSIC performance, especially with limited training samples, by integrating CNN with Iterative Training Sampling Spectral–Spatial Classification (IRTS-SSC) and utilizing a feedback system. Hong et al. proposed FuNet [34], which extracts the HSI features by locally preserving the graph structure in one batch. Li et al. introduced a pixel-block pair (PBP)-based data augmentation technique [35] to generalize the deep learning for HSI classification. A shared trait among these methods is their reliance on convolution for feature extraction. However, constrained by the local receptive field, CNNs are unable to comprehensively grasp continuous spectral attributes [36].

In recent years, sequence models exemplified by transformers [37] and Recurrent Neural Networks (RNNs) have been increasingly utilized in HSIC tasks. Unlike Convolutional Neural Network (CNN) models that primarily focus on local feature extraction, sequence models excel at capturing long-range dependencies. This capability enables them to assist classifiers in comprehensively learning the spatial and spectral relationships inherent within HSI [31]. For instance, Hang et al. introduced CasRNN [38], a technique that employs a cascaded RNN equipped with Gated Recurrent Units to explore the redundant and complementary information present in hyperspectral data. Zhang et al. developed SSRNN [39], which utilizes the Local Spatial Sequence (LSS) approach to extract structural details before feeding this information into an RNN for classification purposes. Hong et al. presented SpectralFormer [36], a method adept at extracting spectral local sequence knowledge from contiguous spectral bands within HSI, resulting in clustered spectral vector representations. Jiang et al. proposed GraphGST [40], a method designed to capture local-to-global correlations that enhance positional encoding for transformers. Sun et al. introduced SSFTT [41], which synergizes the strengths of CNNs and transformers for improved performance. He et al.'s HSI-BERT [42] treats each pixel within a specified HSI cube as a token, allowing the transformer to encapsulate global context; this approach is recognized as one of the first instances of employing a transformer-based model for classification tasks, achieving accuracies comparable to leading methods. Lastly, Yang et al. proposed the HSI transformer network (HiT) [43], integrating convolution operations into a transformer architecture aimed at enhancing HSIC through effective capture of both spectral and spatial details [44]. Zou et al. proposed a local-enhanced spectral–spatial transformer [45], equipped with the HSI2Token module and local-enhanced transformer encoder [46], that excels in HSI classification by effectively capturing both local features and long-range dependencies, surpassing other state-of-the-art networks. Gai et al. proposed a mask-guided spectral-wise transformer (MST) [47] for improved HSI reconstruction by leveraging spectral-wise similarity and the guidance effect of coded apertures. Qi et al. proposed a global–local 3D convolutional transformer network [48], which is proposed to address the limitations of CNNs and vision transformers in HSI classification by embedding 3D convolution within a dual-branch transformer to capture global–local spectral and spatial associations. Roy et al. proposed morphFormer [49], a novel transformer model that enhances HSI classification by integrating learnable spectral and spatial morphological convolutions with attention mechanisms, outperforming traditional CNNs and existing transformers. Ibañez et al. proposed MAEST [50], a ViT-based encoder–decoder model that uses a masking auto-encoding strategy to dynamically uncover the most robust features and employs transformer decoders to reconstruct these features. Sequence models have demonstrated superior proficiency in capturing the nonlinear dynamics of data within complex remote sensing scenarios compared to Convolutional Neural Networks (CNNs). However, Recurrent Neural Networks (RNNs) face considerable computational challenges during both the training and inference phases [51,52], limiting their efficacy in handling lengthy sequences and large datasets. By employing the self-attention mechanism, transformers are skilled at effectively modeling the complex interactions across various spectral bands and spatial domains, leading to more accurate and robust classification results [53,54]. Yet, the self-attention mechanism inherent to transformers necessitates a large number of pairwise multiplication operations during inference, which imposes a significant computational burden. This excessive computational load leads to performance issues that can critically affect HSIC tasks.

Recently, Mamba, based on the State Space Model (SSM) [55], has been proposed and has swiftly garnered widespread attention from the academic community. Mamba exhibits outstanding computational efficiency and robust feature extraction skills, notably in its proficiency to identify long-range dependencies akin to transformers, while also boasting superior computational performance. Gu et al. introduced the HIPPO [56] matrix to solve the problem of long-distance modeling within the limited storage space of the SSM, and the individual parameters of the SSM are associated with the neural network to make it learnable [57]. Liu et al. introduced VMamba [58], a method that incorporates an SSM-based sequential scanning mechanism: SS2D, which helps to bridge the gap between the orderly nature of one-dimensional selective scanning and the non-sequential structure of two-dimensional visual data. Chen et al. proposed RSMamba [59], a technique that employs a dynamic multipath activation strategy to improve Mamba's representation of non-causal data and also pioneers the application of the SSM within the realm of remote sensing. Ge et al. proposed MambaTSR [60], which combines the SSM mechanism with the TSR model for traffic sign recognition. However, these models have primarily been applied to natural image processing [61–64], and the question of how to employ the SSM for HSI processing remains an area worthy of exploration.

Recently, several methods for using SSM models for HSIC tasks have also been proposed. Huang et al. proposed SS-Mamba [65], an efficient deep learning architecture that integrates spectral and spatial features to enhance HSIC performance. Yang et al. proposed GraphMamba [66], an efficient HSIC framework that deeply mines spatial–spectral information and achieves optimal performance through graph structure learning and adaptive spatial context awareness. Li et al. proposed MambaHSI [67], which is a novel model that leverages the long-range interaction capabilities and linear computational complexity of the Mamba architecture to achieve outstanding performance through adaptive fusion of spatial and spectral features.

To diminish the computational intricacy of the model and mitigate the impact of spatial heterogeneity, a Spectral–Spatial Unified Mamba (SSUM) model is proposed to reduce computational complexity and improve model performance. The SSUM model is composed of a Spectral Mamba branch and a Spatial Mamba branch, which comprehensively leverage the features of both spectral and spatial domains to enhance the model performance, as shown in Figure 1. Specifically, in the Spectral Mamba branch, a nearest-neighbor spectrum fusion (NSF) strategy is proposed to mitigate the interference arising from the spatial variability. Furthermore, a novel sub-spectrum scanning (SS) mechanism is developed, which scans across the sub-spectrum dimension to better learn the details of spectral features. Within the Spatial Mamba branch, a Spatial Mamba (SM) module is developed by integrating a 2D Selective Scan Module (SS2D) with Spatial Attention (SA) into a unified network to effectively capture the spatial features of HSI. At the same time, the SM part obtains spatial information from a larger scale, which reduces the influence of spatial variability of classification edge position. Finally, the features obtained from the Spectral Mamba and Spatial Mamba branches are fused at the decision level to achieve the final classification of HSI. In addition, the computational complexity of the SSM is low, which improves the detection speed of SSUM.

The main contributions of this paper are as follows:

1. A novel backbone network based on the State Space Model, named Spectral–Spatial Unified Mamba (SSUM), is proposed for HSIC. This framework integrates the Spectral Mamba, Spatial Mamba, and classifier into a unified neural network. This is the first time that the SSM has been used in conjunction with the conventional spatial–spectral combining algorithm for HSIC.
2. In the Spectral Mamba branch, a novel NSF strategy is proposed to mitigate the interference arising from the spatial variability (i.e., same object having different spectra). In addition, a novel SS mechanism is developed, which scans across each sub-spectrum dimension to better learn the details of spectral features.
3. In the Spatial Mamba branch, an SM module is developed by integrating an SS2D with SA into a unified network, which can effectively capture the spatial features of HSI and alleviate the effect of spatial variability.

The remaining of this paper is organized as follows: Section 2 introduces the State Space Model and the methodological analysis of our SSUM. Section 3 details the experiments, including dataset selection, detection speed, comparison results and analysis, ablation study, and parameter analysis. Section 4 discusses the efficiency and classification details of the methodology. Finally, Section 5 concludes this work and points out reasonable future directions.

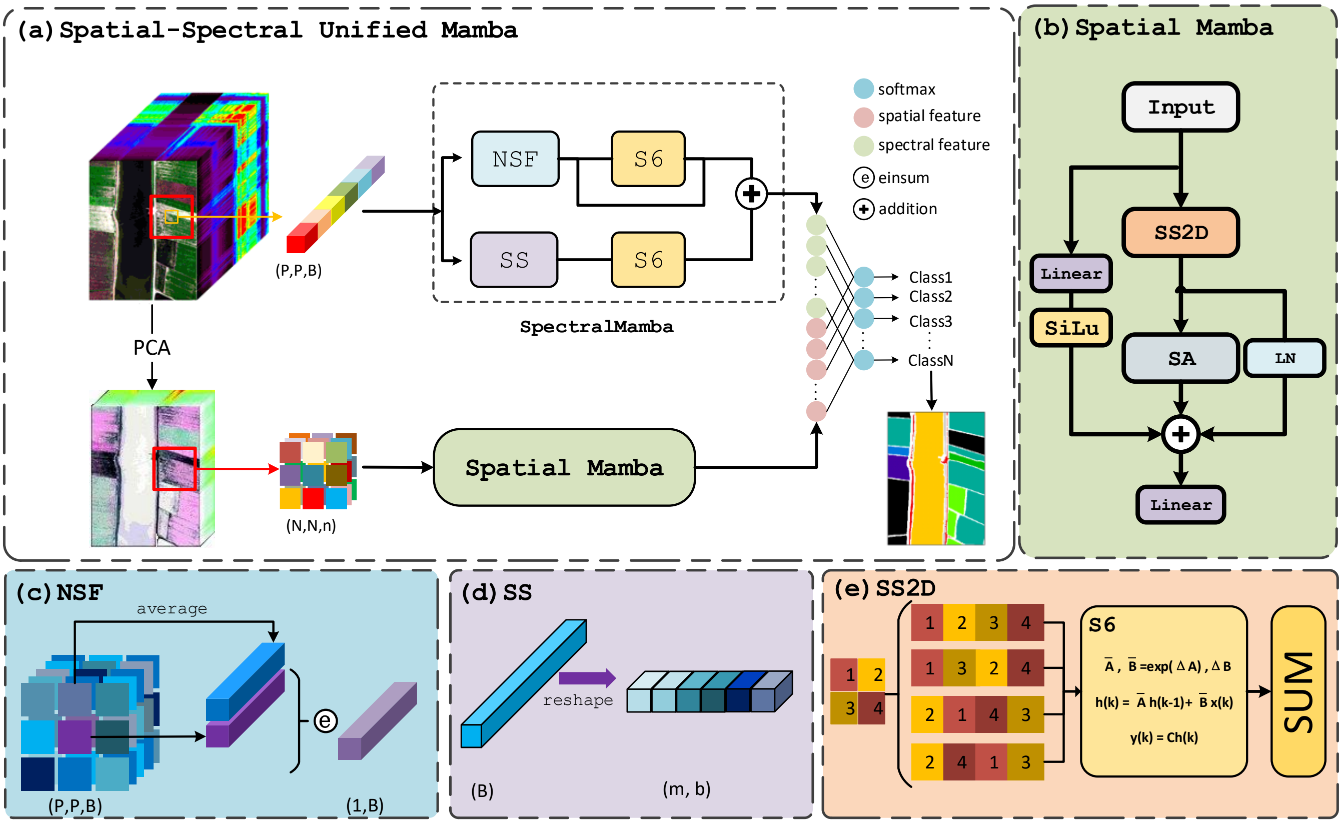


Figure 1. (**a**) The overall architecture of the proposed SSUM, including (**b**) Spatial Mamba; (**c**) nearest-neighbor spectrum fusion (NSF) strategy; (**d**) sub-spectrum scanning (SS) mechanism; and (**e**) 2D Selective Scan Module (SS2D). Specifically, (**a**) denotes the overall architecture of the SSUM; (**b**) denotes the Spatial Mamba, which corresponds to the Spatial Mamba in (**a**); (**c**) denotes the nearest-neighbor spectrum fusion (NSF) strategy, which corresponds to the NSF in (**a**); (**d**) denotes the sub-spectrum scanning (SS) mechanism, which corresponds to the SS in (**a**); (**e**) denotes the 2D selective scan module (SS2D), which corresponds to the SS2D in (**b**).

2. Methods

2.1. Overview

The overall architecture of the proposed SSUM model is demonstrated as shown in Figure 1, which integrates the Spectral Mamba branch and the Spatial Mamba branch into a unified network for classification in HSIs. Figure 1a is the framework of the entire algorithm, and Figure 1b–d are details of the algorithm in Figure 1a. Figure 1e is the implementation detail of the SS2D. Both the two branches and the classifier are trained to share a unified objective function and can simultaneously optimize all parameters within the network. Specifically, the Spectral Mamba branch receives a small-scale HSI patch with the full spectrum and employs a dual-branch structure. This structure includes an NSF strategy and an SS mechanism for extracting comprehensive spectral features. The Spatial Mamba branch consists of an SS2D model, an SA module, and several residual connections and receives an HSI patch processed by PCA. Similar to the Spectral Mamba branch, the spatial features extracted by the Spatial Mamba branch are finally passed through a fully connected layer to generate a classification result. The two classification results are fused at the decision level, generating the final classification result within the Multi-Layer Perceptron (MLP) classifier.

2.2. Selective Scan Space State Sequential Model (S6)

The structured State Space Model has recently emerged [57], attracting considerable attention in the field of sequential data modeling. This class of models commonly originates from continuous-time systems, where the system translates input functions or sequences $x(t) \in R^L$ into output response signals $y(t) \in R^L$ via an implicit latent state $h(t) \in R^N$, where N and L denote the dimensions of the latent space and the sequence, respectively [58]. The procedural mathematics can be expressed using the following differential equations of standard form:

$$h'(t) = Ah(t) + Bx(t) \tag{1}$$

$$y(t) = Ch(t) + Dx(t) \tag{2}$$

where $h'(t) = \frac{dh(t)}{dt}$ refers to the time derivative of $h(t)$; $A \in R^{N\times N}$, $B \in R^{\mathrm{N}\times\mathrm{L}}$, and $C \in R^{L\times N}$ represent the system parameters; and D is often omitted in practical applications.

S4 is the discrete variant of the SSM system, which enables the integration of SSM into deep learning algorithms by sampling input signals at fixed time intervals to obtain their discrete-time counterparts. It is commonly expressed as follows:

$$h(k) = \overline{A}h(k-1) + \overline{B}x(k) \tag{3}$$

$$y(k) = Ch(k) \tag{4}$$

The computation of matrices $\overline{A}$ and $\overline{B}$ is carried out as follows:

$$\overline{A} = \exp(\Delta A) \tag{5}$$

$$\overline{B} = (\Delta A)^{-1}(\exp(\Delta A) - I) \odot \Delta B \tag{6}$$

where the time scaling parameter $\Delta \in R^L$ is employed to transform the continuous parameters $\overline{A}$ and $\overline{B}$ into their discrete counterparts $\overline{A} \in R^{N\times N}$ and $\overline{B} \in R^{N\times L}$.

Traditional SSMs are characterized by their linear time-invariance, implying that the projection matrices remain constant irrespective of the input signal, which results in an undiscriminating focus across all sequence components. The selective scan mechanism, however, correlates parameters with the inputs, thereby augmenting the proficiency in handling intricate sequences and converting the SSM into a linear time-variant architecture. The linear time-varying structure that is linked to the input is termed S6.

2.3. 2D Selective Scan Module (SS2D)

To address the issue of substantial computational overhead in transformer-based models for visual tasks, Yue Liu introduced a novel State Space Model-based architecture termed VMamba [37]. The core of VMamba consists of a stack of SS2D (Structured State Space for 2D) modules with 2D selective scanning. The sequential processing inherent in the scanning operations within the S6 framework is advantageous for NLP tasks that deal with temporal sequences. However, this approach encounters considerable difficulties when extended to visual data, which are fundamentally non-sequential and incorporate spatial characteristics. As depicted in Figure 1e, the data flow in the SS2D model encompasses a tripartite process: initial cross-scanning, subsequent selective scanning of S6 blocks, and final cross-merging.

By traversing along four scanning paths, SS2D aids in bridging the gap between the ordered nature of 1D selective scanning and the non-sequential structure of 2D visual data, which enables the gathering of contextual data from diverse origins and viewpoints. Specifically, the characteristics of two-dimensional imagery are converted into a linear array and then traversed across four unique orientations: commencing at the uppermost left and concluding at the lowermost right, reversing this path from the lowermost right to the uppermost left, beginning at the lowermost left and terminating at the uppermost right, and finally from the uppermost right to the lowermost left. Then, the aforementioned S6 is used to capture the long-term dependencies of each sequence. Finally, all sequences are combined using a summation operation and reshaped back into a 2D structure. In the concluding stage, the sequences are aggregated through summation operations and transformed into a two-dimensional format. The SS2D model, by employing mutually replenishing one-dimensional traversal paths, facilitates the effective integration of information from every pixel across the image with pixels in various orientations. This approach aids in the construction of a comprehensive acceptance domain within the two-dimensional spatial context.

2.4. Spectral Mamba Branch

Considering that the SSM has the advantage of processing long sequence information, we designed the Spectral Mamba to process full band information in a small range. As shown in Figure 1a, the Spectral Mamba branch takes a small-scale full-spectrum patch $X_{patch} \in R^{P\times P\times B}$ as input and employs a dual-branch structure, where P represents the width and height of the taken patch and b represents the band of the HSI. We propose two strategies for processing X_{patch}.

2.4.1. Nearest-Neighbor Spectrum Fusion (NSF) Strategy

HSIs exhibit the characteristic of spatial variability. To mitigate the phenomenon of spectrally similar objects appearing different due to spatial variability, it is necessary to incorporate the neighbor spectrum information for classification. As shown in Figure 1c, firstly, for the input patch $X_{patch} \in R^{P\times P\times B}$, where the central pixel x_{center} is the pixel to be classified, the average features of X_{patch} need to be considered.

$$X_{patch} = [x_0, x_1, x_2 \ldots\ldots x_n] \tag{7}$$

$$x_{avg} = \frac{\sum\limits_{i=0}^{\mathrm{N_p}} x_i}{\mathrm{N_p}} \tag{8}$$

where $\mathrm{N_p} = P \times P$ and x_{avg} denotes the average value of all pixels within X_{patch}.

Then, a fused spectral vector $x_{fus} \in R^{1\times 1\times B}$ is calculated by leveraging the central pixel $x_{center} \in R^{1\times 1\times B}$ of X_{patch} and the average vector $x_{avg} \in R^{1\times 1\times B}$, employing the Einstein summation convention.

$$x_{fus} = einsum(x_{center}, x_{avg}) \tag{9}$$

Finally, the x_{fus} is fed into the S6 model for feature extraction, and following a connection through a residual layer, the resulting feature is denoted as F_{nsf}.

$$F_{nsf} = S6(x_{fus}) + x_{fus} \tag{10}$$

2.4.2. Sub-Spectrum Scanning (SS) Mechanism

HSIs possess the characteristic of data redundancy. Since the size of the hidden state in State Space Models is determined by the input size, State Space Modeling can focus on the features of specific wavelengths through a parameterization method that depends on the input. At the same time, the decomposition of hundreds of spectral bands into segments makes the framework more computationally friendly. To minimize the data redundancy in HSI and to facilitate the State Space Model's attention to fine distinctions within the spectral domain, we propose a novel SS mechanism along with the sub-spectrum scanning to fully utilize the reflectance characteristics of different types of ground objects.

For the input $x_{center} \in R^{1\times 1\times B}$, we split it into a few sub-spectrums of length B', then reassemble them into a new 2D tensor $x'_{center} \in R^{m\times B'}$, where m is the number of splits.

$$x'_{center} = reshape(x_{center}) \tag{11}$$

Then, the x'_{center} is fed into the S6 model for extracting the spatial feature:

$$F_{ss} = reshape[S6(x'_{center})] \tag{12}$$

Finally, the features F_{ss} are concatenated with F_{nsf} from Equation (1) and passed through a linear layer to obtain the final Spectral Mamba feature F_f.

2.5. Spatial Mamba Branch

In order to improve the processing efficiency and reduce the computational burden, many papers show that Mamba has good image processing ability [58–60]. The Spatial Mamba module is designed by combining an SS2D and SA into a unified network to sufficiently extract the spatial features of HISs to improve the processing efficiency and reduce the computational burden. Specifically, in the input section of the Spatial Mamba, we first apply Principal Component Analysis (PCA) to the complete HSI, reducing its dimensionality to n dimensions. This is performed because the Spatial Mamba does not focus on spectral information, and the purpose is to decrease the computational load of the model. Following this, we extract the pixel of interest for classification, along with its neighboring pixels, with the size defined as N. In our method, N is significantly greater than n. The extracted pixel block, which is broad but thin, is represented as $X_{patch} \in R^{N \times N \times n}$.

$$X_{PCA} = PCA(X_{HSI}) \tag{13}$$

where $X_{PCA} \in R^{W \times L \times \mathrm{n}}$ and $X_{HSI} \in R^{W \times L \times B}$. W and L represent the width and length of the original HSI, B denotes the number of bands (dimensions) of the original HSI, and n signifies the dimensionality of the HSI after the dimensionality reduction process.

As shown in the Figure 1b, the Spatial Mamba model is primarily composed of the SS2D layer, a Spatial Attention mechanism, and several residual blocks. By employing complementary 1D traversal paths, SS2D enables each pixel in the image to effectively integrate information from all other pixels in various orientations, thereby facilitating the establishment of a global receptive field in 2D space. However, the four-directional scanning approach of SS2D has certain limitations. In particular, some contiguous pixels within the two-dimensional feature map are notably disjoint in the one-dimensional token sequence. For instance, during left and right scanning, pixels that are close vertically are far apart in the 1D token sequence. Such long distances can lead to the neglect of local pixel relationships. Therefore, it is necessary to introduce SA mechanisms to restore the neighborhood similarity. Specifically, we use the pooling operation to compress the feature map, capture the global context information of the spatial feature map mapped by the SS2D module, and then use the convolutional layer to compensate the local features. By emphasizing the regions related to image details in the feature map, SA can help the network to better learn the spatial characteristics of the feature image, which is crucial for recovering the spatial location information of the feature map. For our task, the SA mechanism addresses the inherent shortcoming of SS2D models.

As shown in the Figure 2, we design an improved Spatial Attention mechanism to assist the SS2D model in restoring neighborhood similarity. First, the results of max pooling and average pooling are computed for the feature block, and after concatenating these results, a convolution operation is performed to obtain the Spatial Attention features. The convolution kernel size is set to 3, and the step size is set to 2. In this paper, the obtained Spatial Attention features are multiplied by the input to ensure an output F_{mamba} with the same dimensions as the input. Subsequent to this, the following computations were conducted.

$$F_{PCA} = SiLU[Linear(X_{patch})] \tag{14}$$

$$F_{mamba} = SA[SS2D(X_{patch})] \tag{15}$$

$$F_{\mathrm{ln}} = LayerNorm[SS2D(X_{patch})] \tag{16}$$

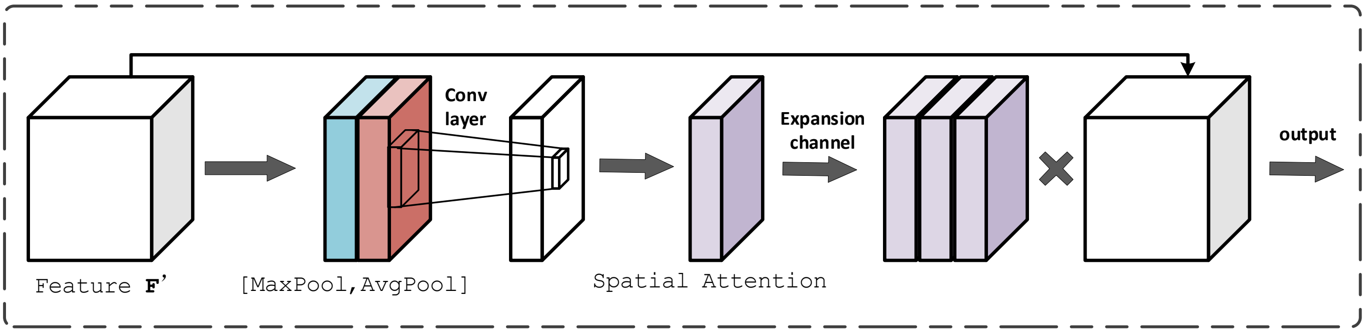


Figure 2. Improved Spatial Attention mechanism applied to this method.

As indicated by the formula, the output of the SS2D is subjected to layer normalization to obtain the feature F_{ln}. This action is performed to preserve the characteristics of SS2D while bolstering the model's generalization abilities. The original image is passed through a fully connected layer followed by the activation function SiLU to obtain F_{ori}, which aids subsequent linear layers in better learning the features. Finally, the three features are summed together and passed through a simple linear layer to yield the final classification results of the Spatial Mamba model.

2.6. Feature Fusion at the Decision Level

The Spectral Mamba and Spatial Mamba can perform pixel classification independently. The Spectral Mamba uses a small range of full-band spectral information, and the Spatial Mamba uses a large range of spatial information after dimensionality reduction. This way of information utilization is more comprehensive. For the purpose of realizing comprehensive spectral–spatial classification, we merge the final fully connected (FC) layer from the Spectral Mamba with the corresponding layer from the Spatial Mamba, thereby creating a novel FC layer. This concatenated layer is then fed into a simple MLP classifier to realize the final classification. The cross-entropy loss is employed to train the SSUM network, which is defined as follows:

$$Loss = -\frac{1}{K}\sum_{i=1}^{K}\mathbf{y}_i \log \mathbf{z}_i \tag{17}$$

where $\mathbf{y}_i$ is the one-hot encoding of true labels, which is considered as a one-hot vector, $\mathbf{z}_i$ is the class prediction of SSUM, and K represents the number of training samples.

In summary, the NSF strategy is proposed to alleviate the interference caused by the spatial variability (i.e., same object having different spectra). The SS mechanism is proposed, which scans along the sub-spectrum dimension to enhance the model's perception of subtle spectral details. The Spatial Mamba module is designed by combining an SS2D and SA into a unified network to sufficiently extract the spatial features of HSIs. Finally, the classification results are derived by uniting the output feature of the Spectral Mamba and Spatial Mamba branch, thus improving the comprehensive performance of HSIC.

3. Results

3.1. Experimental Setup

In the course of this study, all experimental protocols were conducted within the PyTorch framework on a solitary NVIDIA GeForce RTX 2080 graphics processing unit (GPU) that boasts 12 gigabytes of memory. The initialization of the proposed SSUM model entailed populating its parameters with stochastic values sampled from a normal distribution with a mean set at zero and a standard deviation of 0.01. For the optimization of the SSUM, the Adam optimization algorithm was adopted in conjunction with an exponential decay learning rate policy that initiated at 0.0001. The training phase of the model spanned 200 epochs, with each training batch consisting of 64 samples. The input size of the Spectral Mamba is set to 3 × 3 × N_b, where N_b denotes the number of bands, and the input size of the Spatial Mamba is set to 40 × 40 × 3.

To illustrate the efficacy of the introduced SSUM, we selected and assessed 11 prominent hyperspectral imaging (HSI) classification techniques as benchmarks for comparison. These eight approaches consist of KNN [13], RF [14], 1DCNN, 2DCNN [62], HybridSN [28], IRTS-3DCNN [33], CasRNN [38], VIT [46], SpectralFormer [36], GraphGST [40], and SS-Mamba [65]. These methods are all supervised. The array of comparative methods spans a diverse range of techniques, including conventional algorithms and RNN-based, CNN-based, and transformer-based methods, offering a thorough and comprehensive framework for evaluation. The specific implementation details for these methods can be found in their respective original publications. Each method was subjected to testing under the optimal experimental conditions as documented in their papers, or they were replicated using the authors' official code repositories.

3.2. Datasets Description

We evaluated the SSUM on four openly accessible datasets: Indian Pines, Pavia University, Salinas Valley, and WHU-Hi-LongKou. Randomly selected training sets are consistently used across experiments to ensure fairness.

(1) Indian Pines: This dataset predominantly documents an agricultural region in the northwestern part of Indiana, United States, and was acquired in June of 1992 through the Airborne Visible Infrared Imaging Spectrometer (AVIRIS). Comprising an image of 145 by 145 pixels, each with a spatial resolution of 20 m, the dataset includes 220 spectral bands that span a wavelength spectrum from 400 nanometers to 2500 nanometers. For utility in real-world scenarios, 20 bands that exhibited a low signal-to-noise ratio were excluded, leaving 200 bands for analysis. The dataset encompasses 16 distinct types of land cover. A pseudo-color representation of the image and the ground truth data are depicted in the accompanying figure, while the allocation of training and test sets is detailed in Table 1. Due to the small amount of samples in the Indian Pines dataset, we selected 5% of the samples in each category as the training set and the remaining 95% as the test set. Its false-color map and ground truth map are shown in Figure 3.

Table 1. The numbers of samples in the Indian Pines dataset (5% of the labeled samples in each category are randomly selected for training).

No.	Category	Training	Testing
1	Alfalfa	2	44
2	Cron Notill	71	1357
3	Cron Mintill	41	789
4	Cron	11	226
5	Grass—Pasture	24	459
6	Grass—Trees	36	694
7	Grass Pasture Mowed	1	27
8	Hay Windrowed	23	455
9	Oats	1	19
10	Soybean Notill	48	924
11	Soybean Mintill	122	2333
12	Soybean Clean	29	564
13	Wheat	10	195
14	Woods	63	1202
15	Buildings Grass Trees Drivers	19	367
16	Stone Steel Towers	4	89
	Total	505	9744

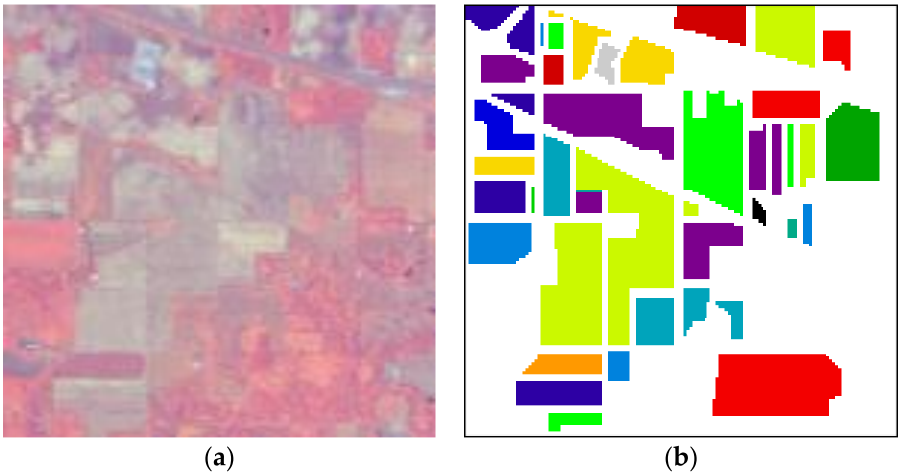


Figure 3. Indian Pines dataset. (**a**) False-color map. (**b**) Ground truth.

(2) Pavia University: This dataset collected via the Reflective Optics System Imaging Spectrometer (RO-SIS) at the University of Pavia's campus in Italy includes an image measuring 610 by 340 pixels, where each pixel has a spatial resolution of 1.3 m. It includes nine different land cover categories and encompasses 103 spectral bands. A visual representation of the dataset in pseudo-color, along with the ground truth information, is illustrated in the figure. Additionally, the distribution of training and test sets for the dataset is outlined in Table 2. We select 1% of the samples for each category as the training set and the remaining 99% as the test set. Its false-color map and ground truth map are shown in Figure 4.

Table 2. The numbers of samples in the Pavia University dataset (1% of the labeled samples in each category are randomly selected for training).

No.	Category	Training	Testing
1	Asphalt	66	6565
2	Meadows	186	18,463
3	Gravel	20	2079
4	Trees	30	3034
5	Mental sheets	13	1332
6	Bare soil	50	4979
7	Bitumen	13	1317
8	Bricks	36	3646
9	Shadow	9	938
	Total	423	42,353

(3) Salinas Valley: This dataset is collected by the 224-band AVIRIS sensor in the Salinas Valley, California, and is characterized by its high spatial resolution (3.7 m pixels). The image size is 512 by 217 pixels, containing 224 bands. The dataset has a size of 512 × 217, with 204 effective bands, and the spatial resolution of the image is 3.7 m. The wavelength coverage spans from 400 to 2500 nanometers. The dataset includes 16 land cover classes. The pseudo-color image and ground truth of the dataset are shown in the figure, and the training and test set divisions are presented in Table 3. We select 1% of the samples for each class as the training set and the remaining 99% as the test set. Its false-color map and ground truth map are shown in Figure 5.

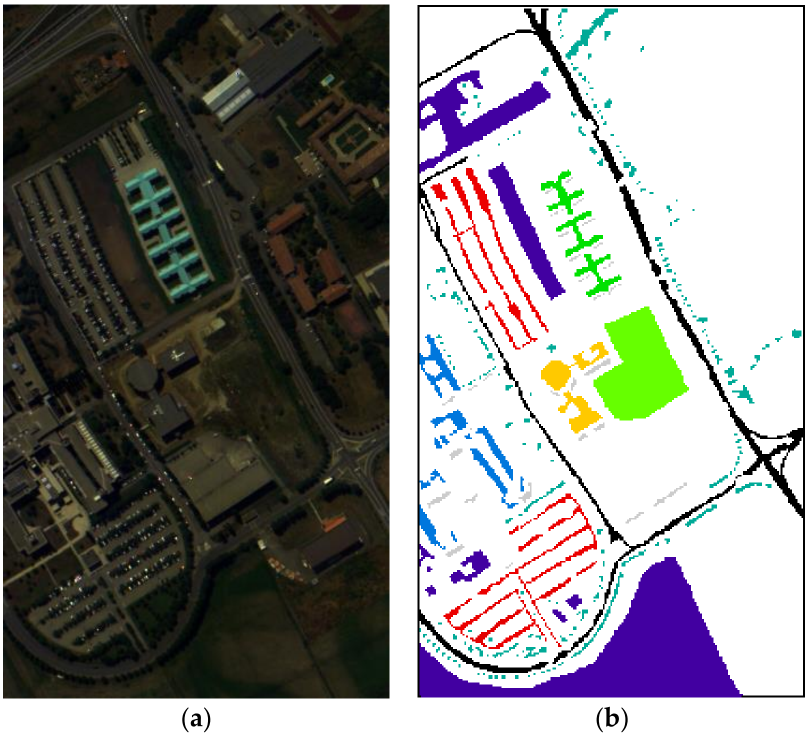


Figure 4. Pavia University dataset. (**a**) False-color map. (**b**) Ground truth.

Table 3. The numbers of samples in the Salinas Valley dataset (1% of the labeled samples in each category are randomly selected for training).

No.	Category	Training	Testing
1	Brocoli green weeds 1	20	1989
2	Cron Mintill	37	3689
3	Fallow	19	1957
4	Fallow rough plow	13	1381
5	Fallow smooth	26	2652
6	Stubble	39	3920
7	Celery	35	3544
8	Grapes untrained	112	11,159
9	Soil vinyard develop	62	6147
10	Corn senesced green weeds	32	3246
11	Lettuce romaine 4wk	10	1058
12	Lettuce romaine 5wk	19	1908
13	Lettuce romaine 6wk	9	907
14	Lettuce romaine 7wk	10	1060
15	Vinyard untrained	72	7196
16	Vinyard vertical trellis	18	1789
	Total	533	53,596

(4) WHU-Hi-LongKou: This dataset is collected in LongKou Town, Hubei Province, China, using the Headwall Nano-Hyperspectral imaging sensor with an 8 mm focal length mounted on a DJI Matrice 600 Pro (DJI M600 Pro, DJI, Hong Kong, China) unmanned aerial vehicle (UAV) platform [68]. The dataset comprises a size of 550 by 400 pixels, including 270 bands, with a wavelength range covering 400–1000 nm. There are a total of nine land cover classes, comprising six crop types and three terrain types. The pseudo-color image and ground truth of the dataset are shown in the figure, and the training and test set divisions are presented in Table 4. We select 0.5%

of the samples for each category as the training set and the remaining 99.5% as the test set. Its false-color map and ground truth map are shown in Figure 6.

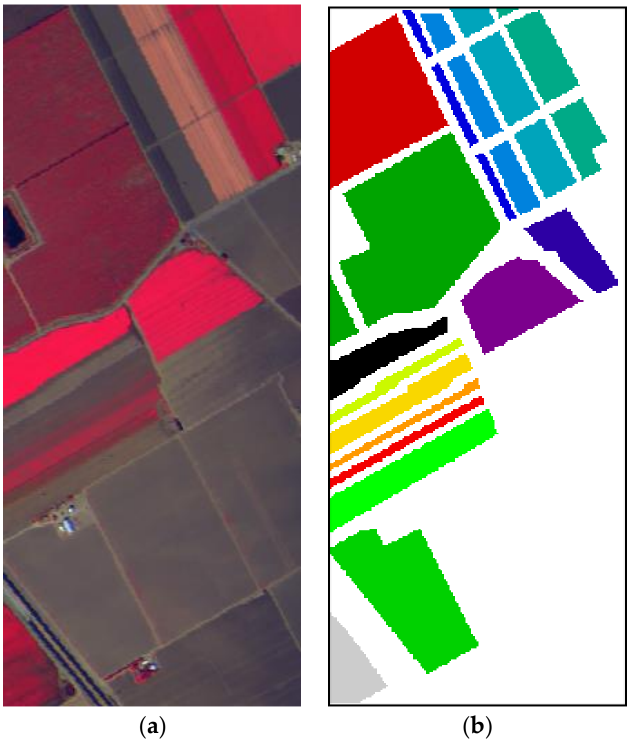


Figure 5. Salinas Valley dataset. (**a**) False-color map. (**b**) Ground truth.

Table 4. The numbers of samples in the WHU-Hi-Long Kou dataset (0.5% of the labeled samples in each category are randomly selected for training).

No.	Category	Training	Testing
1	Corn	172	34,339
2	Cotton	41	8333
3	Sesame	15	3016
4	Broad-leaf soybean	316	62,896
5	Narrow-leaf soybean	20	4131
6	Rice	59	11,795
7	Water	335	66,721
8	Roads and houses	35	7089
9	Mixed weed	26	5203
	Total	1019	203,523

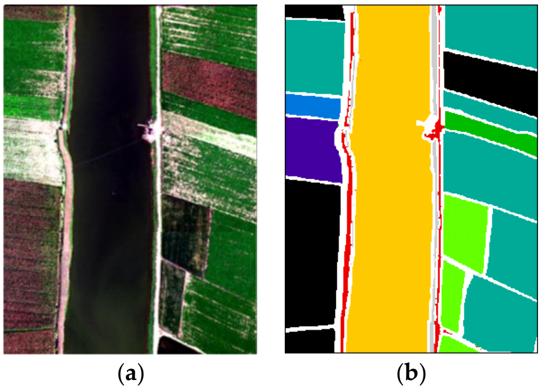


Figure 6. WHU-Hi-Long Kou dataset. (**a**) False-color map. (**b**) Ground truth.

3.3. Comparative Experimentation

It can be seen from Tables 5–8 that three standard evaluation metrics, including overall accuracy (OA %), average accuracy (AA %), and kappa coefficient (κ) [27,69,70], are reported in the last four lines. The first line lists our method and the comparison methods, while the subsequent lines indicate the accuracy (%) for each category across the different methods. It can be seen that hyperspectral classification methods based on deep learning have higher accuracy than traditional methods on the whole. As can be seen from the figure, the method of simply extracting spectral information for classification often cannot overcome the influence of spatial variability, resulting in a large number of noise points generated by misclassification. In contrast, the classification results of the fusion of spatial and spectral information are smoother, which accords with the general expectation of HSIC. Figures 7–10 show the classification results of various methods on the four datasets.

Table 5. Results of the comparison for the Indian Pines test set (5% of the labeled samples in each category are randomly selected for training).

Category	KNN	RF	1D CNN	2D CNN	Hybrid SN	IRST 3DCNN	CasRNN	ViT	Spectral Former	GraphGST	SS-Mamba	SSUM
1	0.00	2.27	13.64	18.18	36.36	50.00	36.36	2.27	20.45	**97.72**	95.34	93.18
2	43.99	50.26	59.17	68.83	83.20	81.35	42.00	51.14	76.41	**92.40**	85.39	91.89
3	37.26	45.63	43.35	58.56	95.06	90.49	42.21	30.92	68.06	89.98	87.43	**96.07**
4	2.21	21.68	33.63	38.50	67.70	69.02	17.26	28.76	42.92	75.66	75.11	**91.15**
5	43.79	78.21	80.17	59.48	83.22	93.02	63.40	66.44	73.42	92.15	86.68	**94.99**
6	97.55	97.69	86.02	88.18	98.85	99.71	69.31	89.48	95.96	99.85	**100.00**	94.09
7	0.00	0.00	25.93	11.11	48.15	85.18	14.81	29.62	25.92	**100.00**	88.46	40.74
8	99.56	95.82	96.26	94.07	**100.00**	**100.00**	91.65	93.84	99.34	**100.00**	**100.00**	99.78
9	0.00	0.00	21.05	73.68	15.79	26.31	10.53	10.52	5.26	**100.00**	0.00	76.47
10	43.72	52.38	69.70	56.39	94.81	90.90	47.51	52.27	79.43	94.58	66.73	**95.89**
11	74.15	80.88	70.21	82.73	84.26	89.66	64.47	62.40	80.19	90.87	94.08	**98.59**
12	20.04	37.59	54.61	54.96	79.43	74.64	23.94	28.54	66.84	82.80	**97.86**	96.45
13	85.13	96.41	87.69	90.77	96.41	**100.00**	83.23	85.12	96.92	**100.00**	98.96	95.90
14	95.01	89.60	90.35	91.10	98.75	99.08	86.61	72.04	93.59	96.25	98.58	**99.33**
15	4.90	21.53	39.78	82.02	99.46	80.65	30.52	28.61	57.76	97.00	**99.18**	98.64
16	50.56	59.55	76.40	84.27	88.76	88.76	53.93	65.16	52.80	91.01	**98.86**	95.51
OA (%)	59.99	67.17	68.78	74.20	91.64	89.40	57.52	58.11	79.00	92.83	90.56	**96.25**
AA (%)	43.62	51.84	59.25	65.80	80.01	82.42	48.90	49.83	64.71	**93.77**	85.79	91.17
k	0.5349	0.62	0.6433	0.7036	0.9045	0.8792	0.5159	0.5213	0.7611	0.9183	0.8924	**0.9573**

Table 6. Results of the comparison for the Pavia University test set (1% of the labeled samples in each category are randomly selected for training).

Category	KNN	RF	1D CNN	2D CNN	Hybrid SN	IRST 3DCNN	CasRNN	ViT	Spectral Former	GraphGST	SS-Mamba	SSUM
1	78.29	81.66	81.33	90.21	79.01	90.84	83.37	77.02	81.76	93.37	**99.95**	86.90
2	98.94	97.89	95.52	98.94	96.65	99.51	89.28	95.47	96.08	99.51	**100.00**	99.96
3	20.06	36.99	61.86	80.86	68.35	92.20	62.34	51.56	67.91	79.02	**100.00**	89.85
4	47.30	79.80	90.71	93.57	99.74	92.08	84.77	73.07	86.65	91.59	96.07	98.06
5	98.95	98.72	99.40	100.00	99.62	99.92	99.47	99.47	**100.00**	**100.00**	**100.00**	**100.00**
6	20.73	33.92	47.12	90.84	99.26	99.21	53.89	32.39	38.60	94.31	80.03	**100.00**
7	62.34	67.20	80.56	84.66	97.27	67.12	81.70	79.11	85.49	96.35	56.00	**97.95**
8	89.88	89.66	91.14	93.03	98.16	**99.14**	86.01	82.94	92.37	96.16	83.23	87.66
9	99.68	99.25	95.84	97.87	99.57	**100.00**	99.04	**100.00**	97.65	99.46	99.57	98.40
OA (%)	77.07	81.96	94.92	94.42	93.36	**96.22**	82.56	80.08	84.56	96.01	94.54	96.15
AA (%)	68.46	76.12	82.61	92.22	93.07	93.34	82.21	76.79	82.95	94.42	90.54	**95.42**
k	67.81	0.7523	0.7964	0.9261	0.9128	0.949	0.7691	0.7279	0.7902	0.947	0.9266	**0.9491**

Among these selected comparison methods, HybridSN and IRTS-3DCNN are 3DCNN-based methods, while SpectralFormer and GraphGST are transformer-based DL methods. It can be observed that these four SOTA methods achieve good results on all four datasets. Compared to the SpectralFormer method based on ViT, GraphGST performs better, especially GraphGST. This is mainly because the method based on ViT uses patches as tokens and adopts a patch-to-patch strategy to achieve training, preventing them from distinguishing heterogeneous pixels within the same patch. While the classic transformer lacks the ability to capture local features, it can be integrated with a local feature extractor to obtain good classification results. HybridSN and IRTS-3DCNN achieve close classification results, because they are both based on 3DCNN models, which extract spectral information while maintaining local features in space. Different from the above four SOTA methods, our proposed SSUM considers spatial information and spectral information separately and uses multiple state space models with different sizes to extract features. Although spatial information and spectral information are used, the five models use them in very different ways. In addition, KNN, RF, 1DCNN, and other methods only analyze the spectral characteristics of a single HSI pixel, which cannot reduce the influence of spatial variability, resulting in considerable noise points in the obtained classification map. The spatial–spectral integration method effectively solves most of the classification errors caused by spatial heterogeneity. It is worth noting that the classification graphs generated by the proposed SSUM method are in close agreement with those of the most advanced (SOTA) algorithms GraphGST and IRTS-3DCNN, all of which show superior classification performance. Using the same set of training samples, the SSUM model produced in this study achieved the highest accuracy on all four datasets compared to the other models in the attached table. For example, on the Pavia University dataset, the SSUM model is 12.87%, 12.92%, and 0.1746 higher in overall accuracy (OA), average accuracy (AA), and Kappa (κ), respectively, than the SpectralFormer. It also outperformed the current SOTA method GraphGST in OA (96.93% vs. 94.84%), AA (95.27% vs. 93.76%), and κ (0.9593 vs. 0.9318). The results show the effectiveness of the SSUM model proposed in this study.

Table 7. Results of the comparison for the Salinas Valley test set (1% of the labeled samples in each category are randomly selected for training).

Category	KNN	RF	1D CNN	2D CNN	Hybrid SN	IRST 3DCNN	CasRNN	ViT	Spectral Former	GraphGST	SS-Mamba	SSUM
1	45.85	98.89	98.89	82.91	**100.00**	87.93	97.44	99.79	96.53	99.89	98.89	89.84
2	88.26	99.81	85.69	98.45	**100.00**	97.56	82.52	57.57	99.10	93.95	**100.00**	81.84
3	57.08	94.69	77.41	86.71	**99.85**	95.19	85.64	69.85	77.36	94.17	99.59	98.93
4	96.23	98.48	99.78	97.97	99.28	99.34	91.75	97.75	97.24	98.76	**99.63**	98.7
5	94.76	95.48	82.24	96.53	98.00	97.88	97.47	96.19	92.57	97.92	**99.43**	98.27
6	97.91	98.11	99.11	98.72	99.21	98.85	99.06	99.41	99.56	99.79	**100.00**	99.74
7	82.48	98.90	99.35	99.21	98.39	99.60	98.98	96.72	87.69	94.69	**100.00**	99.18
8	80.67	84.59	85.03	79.20	93.84	88.85	74.45	74.92	81.29	85.58	**99.74**	88.88
9	97.93	95.88	97.41	99.54	100.00	99.02	96.12	95.81	97.45	99.31	**100.00**	**100.00**
10	55.05	86.66	67.87	85.86	96.00	96.11	75.97	60.32	77.32	96.85	93.12	**99.69**
11	58.98	87.33	76.37	86.67	96.03	95.46	81.76	83.83	86.86	90.54	91.86	**99.81**
12	72.12	99.37	99.27	96.86	**100.00**	**100.00**	99.42	66.19	82.38	98.63	97.85	93.61
13	79.38	97.91	98.57	100.00	85.45	96.69	91.62	94.48	98.89	94.7	**100.00**	93.50
14	35.47	90.66	83.58	99.53	**100.00**	97.73	89.25	90.37	93.01	96.50	92.91	98.96
15	31.31	43.77	44.05	66.44	63.02	90.75	42.52	38.47	55.71	76.26	57.51	**94.39**
16	1.57	94.63	87.37	83.57	**100.00**	97.65	87.37	68.69	83.51	97.78	99.72	98.04
OA (%)	71.03	86.57	83.02	87.73	92.90	94.8	82.17	76.25	84.60	91.94	93.34	**95.31**
AA (%)	67.19	91.57	86.37	91.14	95.57	96.16	81.61	80.65	87.91	94.71	95.64	**96.46**
κ	0.6723	0.85	0.8099	0.8633	0.9207	0.9421	0.7952	0.7348	0.8282	0.9103	0.9255	**0.9479**

Table 8. Results of the comparison for the WHU-Hi-LongKou test set (0.5% of the labeled samples in each category are randomly selected for training).

Category	KNN	RF	1D CNN	2D CNN	Hybrid SN	IRST 3DCNN	CasRNN	ViT	Spectral Former	GraphGST	SS-Mamba	SSUM
1	96.75	95.33	93.72	99.42	99.66	99.63	91.47	97.34	93.94	99.56	**99.79**	**99.79**
2	51.29	56.77	46.14	96.70	98.38	87.56	50.37	63.20	46.93	95.11	**99.05**	95.58
3	0.86	42.31	63.03	94.23	90.15	83.09	41.45	46.85	55.63	**96.18**	64.67	91.41
4	94.12	95.62	93.28	97.70	98.05	98.49	91.76	91.04	93.48	97.29	94.77	**98.66**
5	30.69	43.14	33.99	80.49	87.00	62.59	50.30	53.20	43.69	86.54	9.87	**90.83**
6	84.08	97.94	97.27	98.55	98.45	**99.65**	97.65	98.23	99.26	98.97	92.37	96.48
7	99.98	99.93	99.52	99.97	99.98	99.98	99.90	99.81	99.94	99.96	**100.00**	99.83
8	73.24	79.74	73.16	88.80	90.51	90.49	81.93	80.61	88.99	94.20	**97.67**	91.61
9	0.35	35.10	68.27	87.60	**93.20**	84.58	57.29	44.22	52.68	89.73	85.44	92.81
OA (%)	88.36	91.57	90.71	97.70	98.26	97.20	83.46	91.28	91.34	98.02	95.06	**98.32**
AA (%)	59.04	71.76	74.26	93.72	95.04	89.56	73.50	74.95	74.95	**95.28**	82.63	95.22
κ	0.8439	0.888	0.8777	0.9708	0.9772	0.9631	0.8714	0.8852	0.8858	0.9741	0.9352	**0.9779**

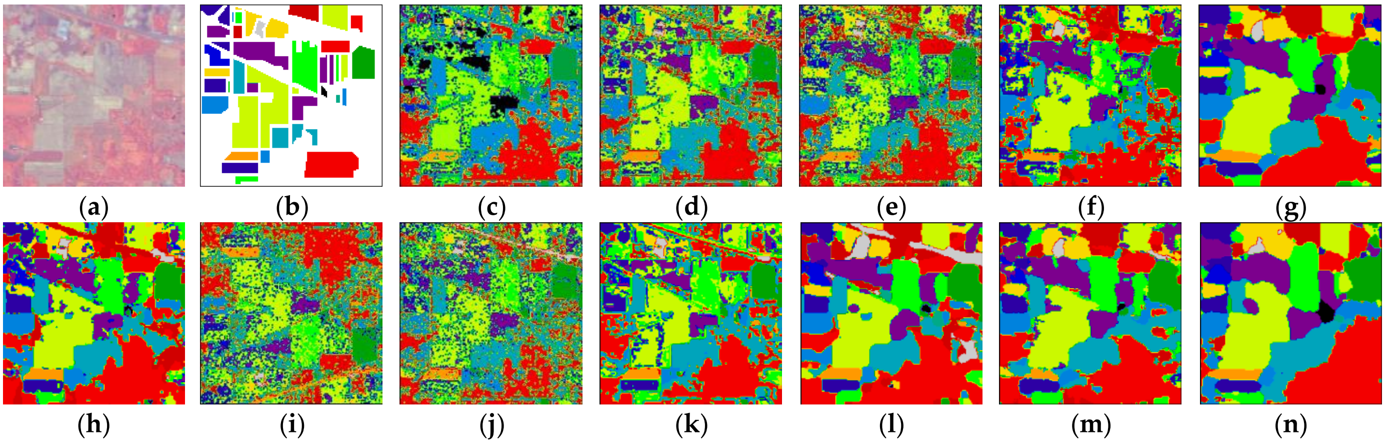


Figure 7. Classification maps produced by various methods applied to the Indian Pines dataset: (**a**) false-color map, (**b**) ground truth, (**c**) KNN, (**d**) RF, (**e**) 1DCNN, (**f**) 2DCNN, (**g**) HybridSN, (**h**) IRTS-3DCNN, (**i**) CasRNN, (**j**) ViT, (**k**) SpectralFormer, (**l**) GraphGST, (**m**) SS-Mamba, and (**n**) SSUM.

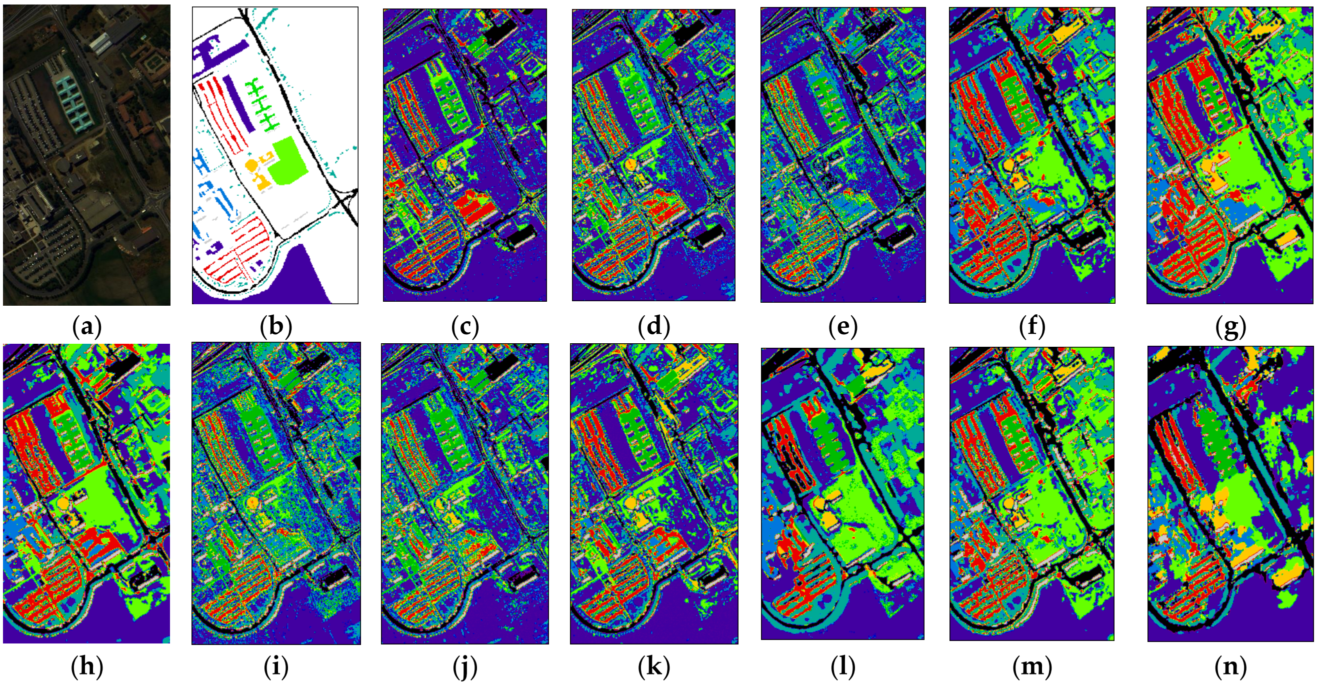


Figure 8. Classification maps produced by various methods applied to the Pavia University dataset: (**a**) false-color map, (**b**) ground truth, (**c**) KNN, (**d**) RF, (**e**) 1DCNN, (**f**) 2DCNN, (**g**) HybridSN, (**h**) IRTS-3DCNN, (**i**) CasRNN, (**j**) ViT, (**k**) SpectralFormer, (**l**) GraphGST, (**m**) SS-Mamba, and (**n**) SSUM.

For the Pavia University dataset, as delineated in Table 6, our approach yields the superlative outcomes concerning AA and κ, and OA also maintained a high level. It markedly surpasses competing methodologies in the discrimination of painted bare soil, bitumen, and mental sheets. The Pavia University dataset encompasses more nuanced textural data, and our proposed SM technique is adept at discerning a broader spectrum of textural nuances, thereby markedly improving classification precision over alternative methods. In the case of the Indian Pines dataset, as noted, our method also attains premier results in OA and κ, with particular prowess in the differentiation of Hay Windrowed, Soybean Notill, and Soybean Mintill. The Indian Pines dataset is characterized by a greater number of spectral bands, and our method adeptly distinguishes the nuanced contrasts among akin categories like Soybean Notill and Soybean Mintill, resulting in precise classification. For the Salinas Valley dataset, the outcomes presented in Table 7 reveal that the methodology introduced in this treatise achieves flawless identification for nine categories and also excels in OA, AA, and κ. The data in Table 8 for the LongKou dataset signify that our strategy markedly outstrips other image classification procedures across seven categories, while also securing the highest rankings in OA and κ. The LongKou dataset is notable for its extensive band count, and the results imply that our dual-branch strategy for spectral information extraction holds a distinct advantage when applied to datasets with an extended spectral range. As can be seen from Figures 7–10, on the whole, HybridSN and IRTS-3DCNN have clearer texture details in the classification results of the four datasets, which is because 3DCNN has a smaller receptive field compared with our method and can learn the joint distribution of the empty spectrum in a smaller range. However, this does not mean that its classification accuracy is higher, and because the receptive field is too small, it may not be able to overcome the influence caused by the space spectrum variability, which may cause its score to be inferior to the SSUM method proposed by us.

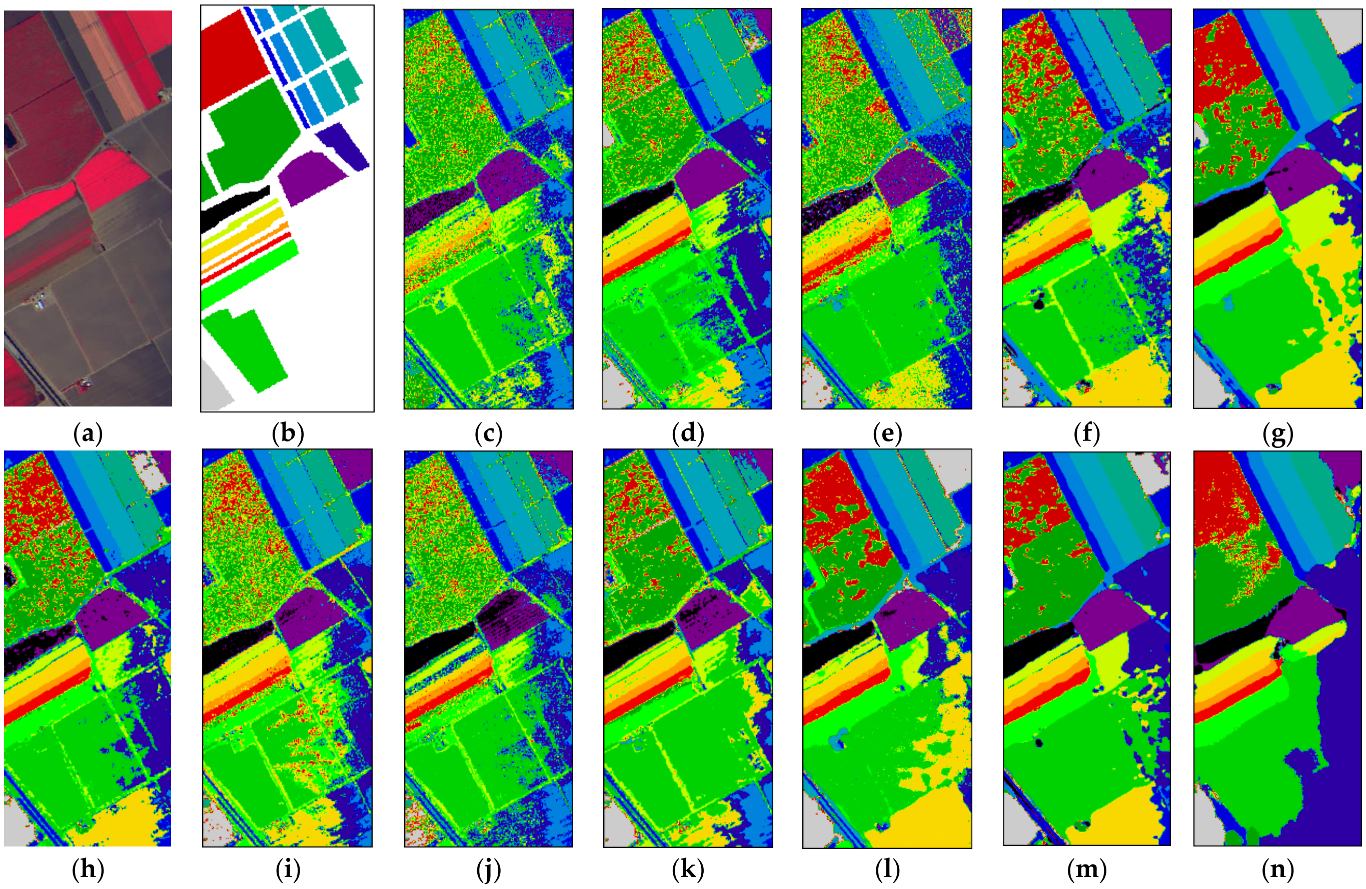


Figure 9. Classification maps produced by various methods applied to the Salinas Valley dataset: (**a**) false-color map, (**b**) ground truth, (**c**) KNN, (**d**) RF, (**e**) 1DCNN, (**f**) 2DCNN, (**g**) HybridSN, (**h**) IRTS-3DCNN, (**i**) CasRNN, (**j**) ViT, (**k**) SpectralFormer, (**l**) GraphGST, (**m**) SS-Mamba, and (**n**) SSUM.

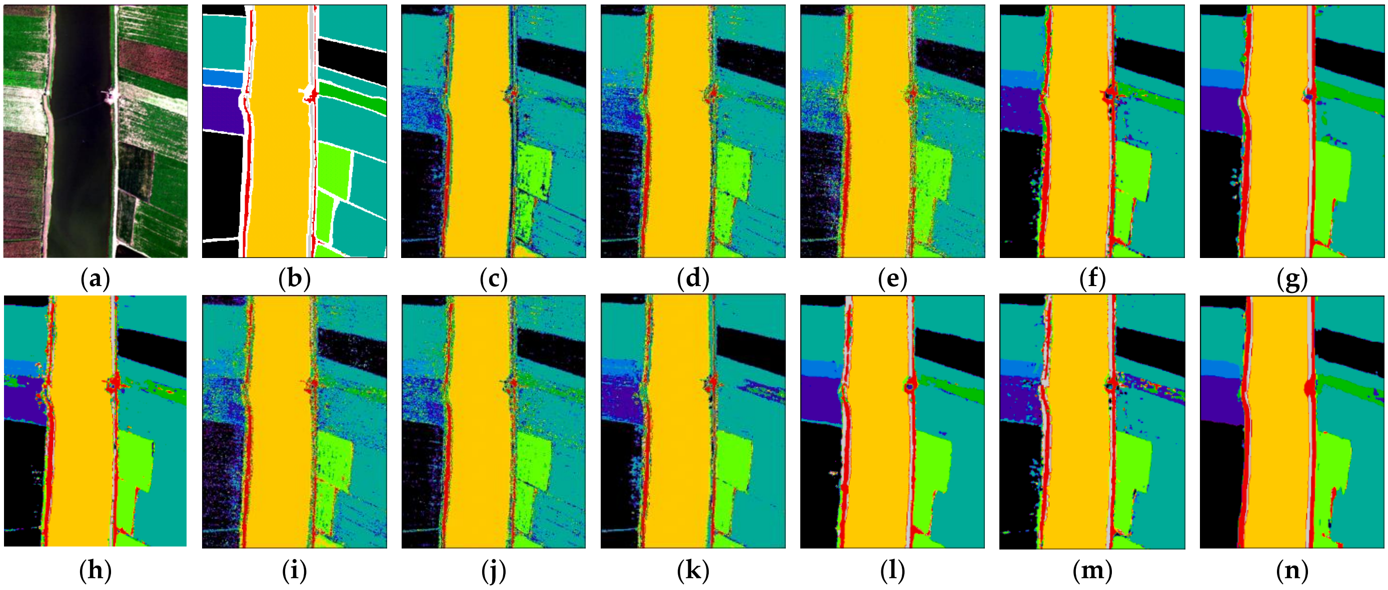

(a) **(b)** **(c)** **(d)** **(e)** **(f)** **(g)**
(h) **(i)** **(j)** **(k)** **(l)** **(m)** **(n)**

Figure 10. Classification maps produced by various methods applied to the WHU-Hi-LongKou dataset: (**a**) false-color map, (**b**) ground truth, (**c**) KNN, (**d**) RF, (**e**) 1DCNN, (**f**) 2DCNN, (**g**) HybridSN, (**h**) IRTS-3DCNN, (**i**) CasRNN, (**j**) ViT, (**k**) SpectralFormer, (**l**) GraphGST, (**m**) SS-Mamba, and (**n**) SSUM.

3.4. Ablation Study

Table 9 shows the impact of each component in the SSUM. Ablation study is used to verify the validity of NSF, SS, SM, and their combinations.

Table 9. Result of ablation study.

NSF	SS	SM	Indian Pines			Pavia University			Salinas Valley			WHU-Hi-LongKou		
			OA	AA	κ	OA	AA	κ	OA	AA	κ	OA	AA	κ
			70.11	71.63	0.6801	86.66	85.84	0.8125	79.25	80.34	0.8044	89.33	84.47	0.8680
√			71.25	79.22	0.7001	88.17	92.00	0.8515	83.55	85.01	0.8857	91.02	87.01	0.8791
√	√		81.44	84.33	0.7654	89.00	90.15	0.8823	85.22	85.43	0.9013	93.77	88.88	0.9202
		√	91.44	89.17	0.9055	91.94	91.95	0.9029	86.09	92.82	0.8463	95.41	93.19	0.9532
√		√	93.09	90.56	0.9211	94.21	93.84	0.9347	90.66	93.49	0.9017	97.13	94.09	0.9625
	√	√	91.80	90.05	0.9062	93.44	92.70	0.9331	87.73	95.35	0.8637	97.27	94.44	0.9599
√	√	√	96.25	91.17	0.9573	96.15	95.42	0.9491	95.31	96.46	0.9479	98.32	95.22	0.9779

The first row presents the classification results when only the central vector is fed directly into the S6 model for classification. This serves as the baseline for analyzing the role of each module.

The second row shows the classification results after the addition of the NSF mechanism, which mitigates the spatial variability of HSIs by calculating the average value of the neighbor pixels. After the NSF mechanism, the S6 model can learn small-scale spatial information of the spectral vectors. It can be observed that the evaluation metrics have improved across all four datasets.

The third row displays the classification results after incorporating the SS mechanism. The SS mechanism enables the S6 model to focus on subtle spectral differences, taking full advantage of S6's effective ability to model long sequences. After the combination of the NSF and SS mechanisms, there are different degrees of improvement in four datasets, which further verifies the effectiveness of the SS mechanism. The evaluation metrics show improvement on all four datasets, indicating that SS can provide more discriminative detail clues by segmenting continuous spectral bands.

The fourth row presents the classification outcomes utilizing the SM mechanism alone. It is observable that satisfactory classification results can also be attained by merely extracting spatial information. In this case, the classification effect is proximate to that of the spatial spectrum combination on datasets with vast space, such as WHU-Hi-LongKou. This is attributed to the fact that the WHU-Hi-LongKou dataset has fewer classification types within a larger spatial range, and its ultimate classification results are more reliant on spatial information. On the contrary, in the datasets with a greater number of classification types, such as Indian Pines and Salinas Valley, there exists a considerable disparity between the classification results using only the SM module and the final eighth row. The fifth and sixth rows, respectively, combine the SM mechanism with NSF and SS. It can be perceived that after integrating spatial information with spectral information, the evaluation index is enhanced on the four datasets. Among them, NSF+SM extracts the comprehensive spectral information, which is superior to SS+SM. This is because the NSF mechanism suppresses small-scale spatial variability, which cannot be overcome solely by the SS mechanism.

Finally, by fusing the three mechanisms, the classifier obtains the HSI features including the complete spectrum, spectral details, and large-scale spatial information, which helps us further improve the classification effect to the highest value on the four datasets. This confirms the effectiveness of the large-scale spatial texture features extracted by the SM module for classification, the effectiveness of the NSF for suppressing spatial variability, and the effectiveness of the SS for sensing spectral details.

3.5. Parameters Analyzed

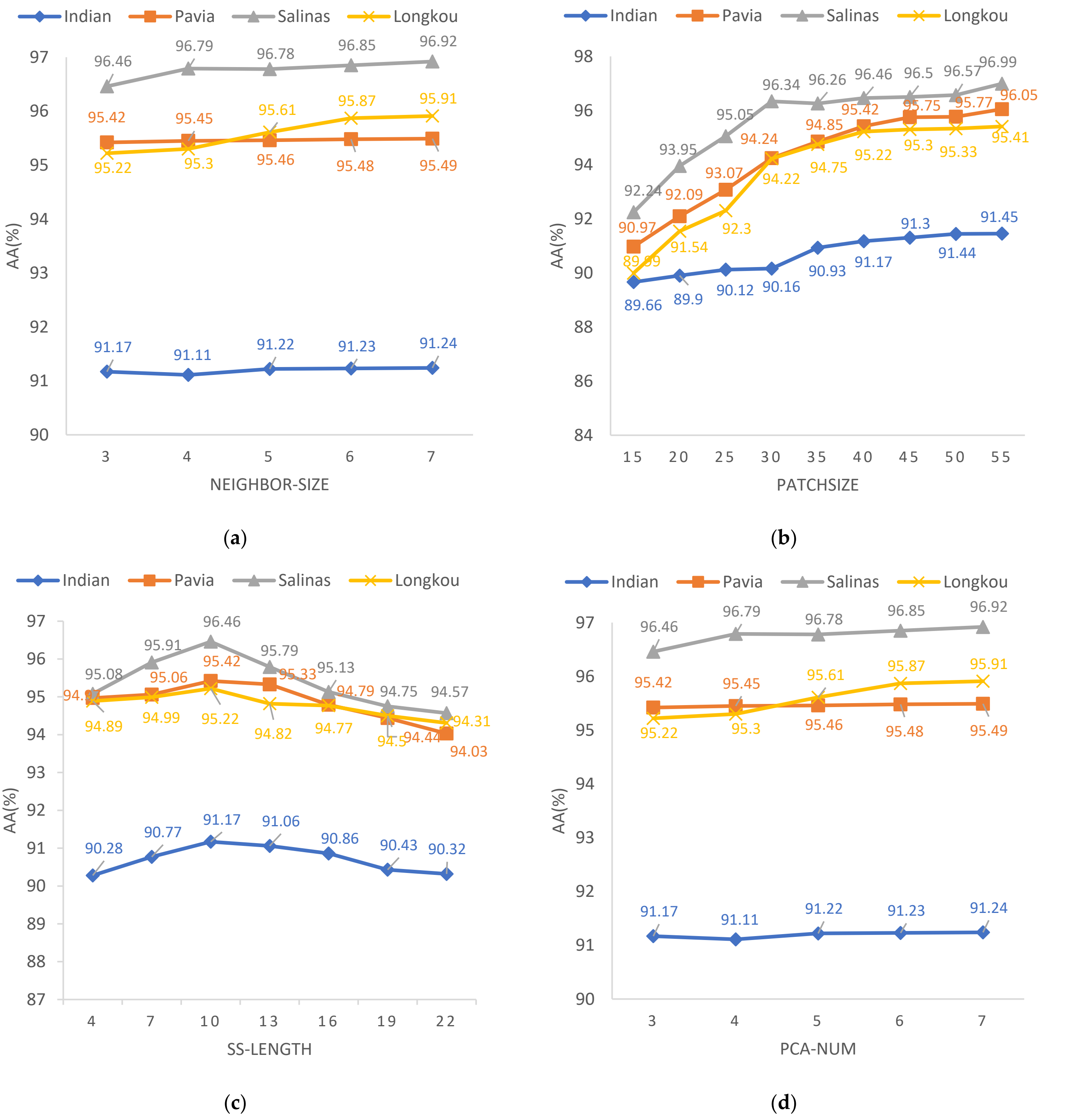
We conducted experiments to analyze the hyperparameters in the proposed SSUM. Specifically, we analyzed the following:

(a) Impact of the neighborhood size. The neighborhood size is quantitatively analyzed in Figure 11a. It can be seen that, as the neighborhood size increased, the AA values did not produce a large change. Meanwhile, because the NSF method needs to use all the bands in the neighborhood, too large neighborhood size will lead to too large memory occupation in the training process. Therefore, the neighborhood size is set to three in our experiments.

(b) Impact of the patch size. The patch size is quantitatively analyzed in Figure 11b. It can be seen that, as the patch size increased, the AA values progressively enhance and approach stabilization at a patch size of 40. Considering the memory consumption and computational load, the patch size is set to 40 in our experiments.

(c) Impact of the sub-spectrum length. The length of the sub-spectrum is quantitatively analyzed in Figure 11c. It can be seen that the highest classification accuracy is attained when the sub-spectrum is set to 10.

(d) Impact of the number of bands after PCA. The number of bands after PCA is quantitatively analyzed in Figure 11d. It can be seen that, as the patch size increased, the AA values are relatively stable, but with the increase in patch size, the model parameters will significantly increase, resulting in additional computational burden. Therefore, the number of bands after PCA is set to three in our experiments.



(**a**) (**b**)

(**c**) (**d**)

Figure 11. Impacts of the different parameters for the proposed SSUM. (**a**) Impact of the neighborhood size. (**b**) Impact of the patch size. (**c**) Impact of the sub-spectrum length. (**d**) Impact of the number of bands after PCA.

3.6. Precision Rate Analysis

Considering that there are a lot of background (BKG) samples on HSIs, a new criterion, called precision rate (PR) [33], was introduced to solve the BKG problem. Specifically, PR is the method that can include all data samples for evaluation. For a certain category C_m, PR is calculated as follows:

$$P_{PR}(C_m) = \frac{n_{mm}}{n_m} \tag{18}$$

where $n_m = \sum_{j=1}^{M} n_{mj}$ is the total number of samples that were classified as category C_m and n_{mj} is the total number of samples that should have been classified as category C_m but were incorrectly classified as category C_j. To exploit the PR of each class for the overall analysis, we use overall precision rate (OPR). It is defined as follows:

$$P_{OPR} = \sum_{m=1}^{M} \frac{n_m}{N} P_{PR}(C_m) \tag{19}$$

where M is the total number of classes in the dataset and $N = \sum_{m=1}^{M} n_m$ is the total number of samples that have been classified.

As shown in Table 10, it can be seen that SSUM obtains the highest OPR on the WHU-Hi-LongKou dataset, followed by the Salinas Valley and Indian Pines datasets. In addition, SSUM obtains a poor OPR on the Pavia University dataset, because the dataset has a large number of background samples.

Table 10. PR analysis considering the BKG on four datasets.

Category	Indian Pines	Pavia University	Salinas Valley	WHU-Hi-LongKou
1	35.00	19.51	58.71	92.26
2	57.06	22.28	54.31	84.18
3	66.18	15.03	11.89	76.62
4	74.55	10.10	44.18	93.50
5	46.13	20.61	75.19	94.06
6	36.42	27.22	76.01	91.25
7	32.55	18.56	52.43	96.67
8	60.38	25.22	83.29	63.69
9	36.36	12.34	21.57	71.55
10	58.62		72.15	
11	61.88		44.00	
12	48.86		56.93	
13	39.25		62.00	
14	22.92		58.03	
15	46.88		80.65	
16	61.71		70.75	
OPR (%)	46.66	19.92	47.44	91.36

4. Discussion

4.1. Discussion of the Run Time

For the purpose of evaluating the inference speed of various models, we chose test subsets from four distinct datasets for category prediction. As depicted in Table 11, the SSUM approach demonstrates a notable speed benefit over the other three transformer-based techniques. This advantage stems from the SSM mechanism's superior computational efficiency in contrast to the self-attention mechanism. The SSM model, characterized by its high computational capacity for long-sequence modeling, is particularly well suited for handling the extensive HSI data encountered in HSIC tasks.

Table 11. Comparison of the run time of the transformer-based models and the SSUM.

Datasets	ViT	SpectralFormer	GraphGST	SSUM
Indian Pines	1.95 s	2.71 s	1.59 s	1.20 s
Pavia University	7.35 s	9.08 s	18.02 s	3.17 s
Salinas Valley	12.14 s	13.61 s	9.05 s	7.18 s
LongKou	30.89 s	32.17 s	35.24 s	26.61 s

4.2. Discussion of the Classification Maps

Utilizing the developed spectral–spatial learning architecture in conjunction with Mamba's feature extraction proficiency, the spectral–spatial information can be fully utilized to improve the performance of HSI classification. For example, in the Salinas Valley dataset, the distinction between vineyard and grape fields is difficult, possibly because the close reflectivity of the leaves. This takes into account subtle differences between the spectral. As can be seen from Figure 12, compared with GraphGST, our SSUM performed better in distinguishing fallow land and grape land, which means that the SS mechanism we designed played a role. For instance, in the lower left corner of the LongKou dataset, as it shows in Figure 12c,d, adjacent to the right edge of the cornfield, other methods frequently misclassify the edge of the cornfield as a road. This error is evidently due to the influence of spectral mixing phenomena in the area where the cornfield meets the road. As illustrated in Figure 12, the NSF mechanism we designed effectively addresses this issue at this location. Compared to other models, including GraphGST, our SSUM model prevents this misclassification.

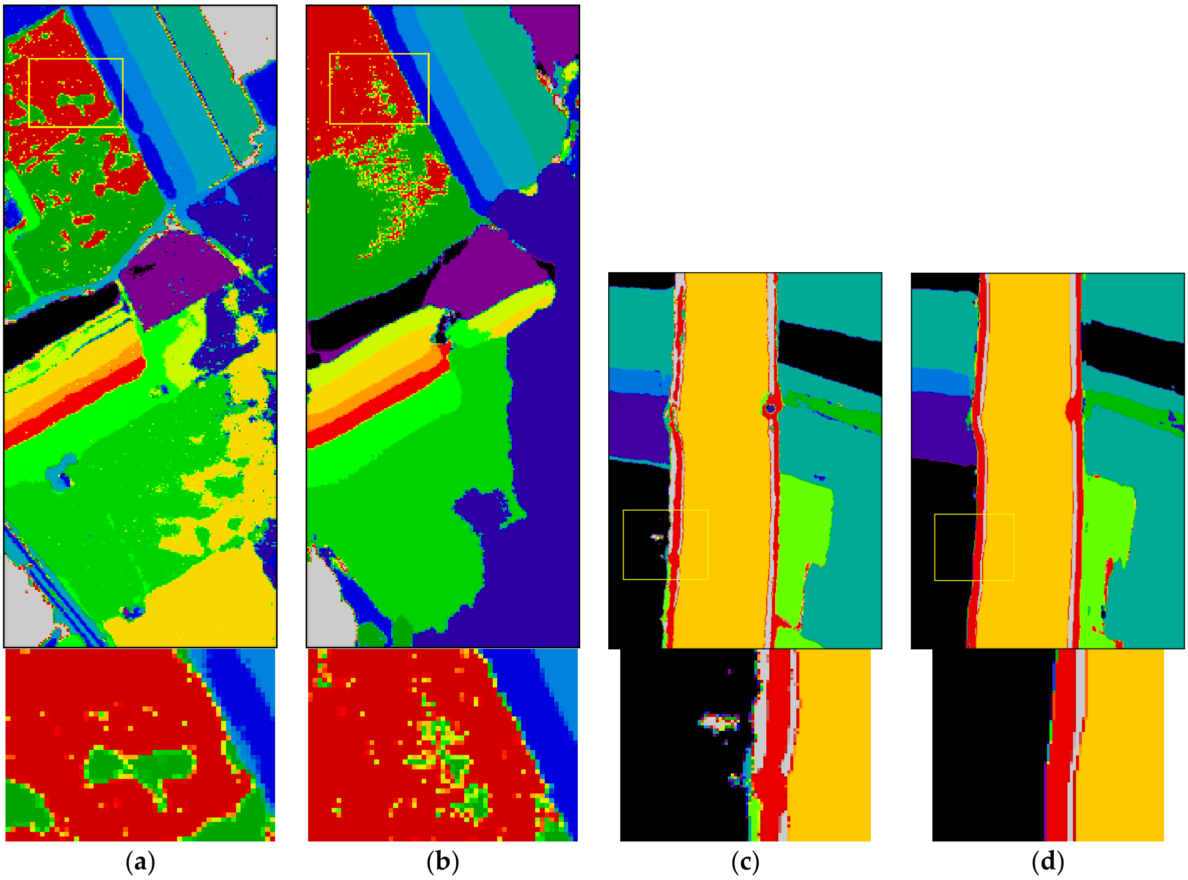


Figure 12. Classification map detail analysis. (**a**) Salinas Valley classified by GraphGST. (**b**) Salinas Valley classified by SSUM. (**c**) Long Kou classified by GraphGST. (**d**) Long Kou classified by SSUM.

4.3. Discussion of the Limitations

The SSUM model attained SOTA performance across four datasets; however, this approach still presents certain limitations. For instance, in the decision layer, the method employs an MLP to integrate the spectral features for attaining the ultimate classification outcome, potentially resulting in the loss of some underlying information beneficial for decision-making. The subsequent step is to deliberate on the feasibility of fusing spatial and spectral features at the feature extraction level through the utilization of a Spatial Mamba and a Spectral Mamba. Additionally, the application of PCA dimensionality reduction suffers from the issue of being sensitive to noise, which is detrimental to the feature extraction of the Spatial Mamba. The next measure is to contemplate using band selection techniques to select bands with complete spatial information instead of the PCA dimensionality reduction method. In addition, we analyze the parameter quantity of the models involved in the comparative experimentation using the Indian Pines dataset, as presented in Table 12. It is evident that HybridSN possesses the highest parameter quantity, a characteristic inherent to the nature of 3DCNN itself. Our proposed SSUM exhibits the second highest parameter quantity, which is comparable to that of GraphGST. This can be attributed to the necessity for parameters A, B, and C within the S6 model to be adjusted by the fully connected (FC) layer, thereby increasing its parameter burden. Fortunately, this increase does not impede the detection speed of the SSUM method due to SSM's low computational complexity. Future work may consider optimizing the S6 model through techniques such as pruning to reduce its required parameter quantity.

Table 12. Comparison of the storage space.

Model.	Hybird SN	IRTS-3DCNN	CasRNN	ViT	SS-Mamba	Spectral Former	Graph GST	SSUM
storage space	4.068 M	0.7130 M	0.3511 M	0.3462 M	0.4700 M	0.3463 M	3.5314 M	3.7340 M

4.4. Discussion of the Training Set Selection

At the same time, we conducted a comparison experiment with the same number of training sets for each category. Tables 13–16 show the comparison of experimental results under the two strategies of selecting training samples. The different training samples for each category were selected referring to Tables 1–4. The same training samples for each category were selected as follows: for the Indian Pines dataset, 50 samples from each category were selected as training samples (15 samples with less than 50 samples were selected as training sets in total). For the Pavia University dataset, 200 samples were drawn from each category as training samples. For the Salinas Valley dataset, 100 samples from each category were taken as training samples. For the LongKou dataset, 100 samples of each type were taken as training samples. It can be seen that whether we choose the same training samples or different training samples, our method can achieve a better performance than other advanced methods.

Table 13. Performance comparisons between same training samples for each category (refer to the second row) and different training samples for each category (refer to the third row) on the Indian Pines dataset.

Methods	KNN	RF	1D CNN	2D CNN	Hybrid SN	IRST 3DCNN	CasRNN	ViT	Spectral Former	GraphGST	SS-Mamba	SSUM
OA (%)	59.22	70.60	73.55	78.31	95.81	96.09	68.44	64.66	78.87	95.30	93.73	**96.51**
AA (%)	63.69	77.74	82.31	85.77	97.50	94.91	75.73	74.67	84.48	97.70	96.91	**98.32**
k	0.5402	0.6681	0.7029	0.7525	0.952	0.9553	0.6374	0.604	0.7605	0.9462	0.9285	**0.9599**
OA (%)	59.99	67.17	68.78	74.20	91.64	89.40	57.52	58.11	79.00	92.83	90.56	**96.25**
AA (%)	43.62	51.84	59.25	65.80	80.01	82.42	48.98	49.83	64.71	**93.77**	85.79	91.17
k	0.5349	0.62	0.6433	0.7036	0.9045	0.8792	0.5159	0.5213	0.7611	0.9183	0.8924	**0.9573**

Table 14. Performance comparisons between same training samples for each category (refer to the second row) and different training samples for each category (refer to the third row) on the Pavia University dataset.

Methods	KNN	RF	1D CNN	2D CNN	Hybrid SN	IRST 3DCNN	CasRNN	ViT	Spectral Former	GraphGST	SS-Mamba	SSUM
OA (%)	78.54	81.66	88.11	95.66	97.95	96.42	83.89	83.60	93.20	98.39	96.28	**98.60**
AA (%)	84.10	86.97	88.78	95.76	98.31	96.35	87.86	87.35	93.84	98.20	98.54	**99.12**
k	0.7194	0.7618	0.8398	0.9415	0.9725	0.9522	0.7831	0.7848	0.9083	0.9784	0.9585	**0.9812**
OA (%)	77.07	81.96	94.92	94.42	93.36	**96.22**	82.56	80.08	84.56	96.01	94.54	96.15
AA (%)	68.46	76.12	82.61	92.22	93.07	93.34	82.21	76.79	82.95	94.42	90.54	**95.42**
k	67.81	0.7523	0.7964	0.9261	0.9128	0.949	0.7691	0.7279	0.7902	0.947	0.9266	**0.9491**

Table 15. Performance comparisons between same training samples for each category (refer to the second row) and different training samples for each category (refer to the third row) on the Salinas Valley dataset.

Methods	KNN	RF	1D CNN	2D CNN	Hybrid SN	IRST 3DCNN	CasRNN	ViT	Spectral Former	GraphGST	SS-Mamba	SSUM
OA (%)	75.19	87.35	86.18	89.71	95.78	95.53	84.37	85.71	91.98	93.42	95.07	**96.51**
AA (%)	80.74	93.72	90.81	95.40	98.29	97.32	91.61	91.80	95.43	96.97	97.76	**98.32**
κ	0.7239	0.8593	0.8454	0.8855	0.9529	0.9502	0.8255	0.8407	0.9109	0.9265	0.9449	**0.9599**
OA (%)	71.03	86.57	83.02	87.73	92.90	94.80	82.17	76.25	84.60	91.94	93.34	**95.31**
AA (%)	67.19	91.57	86.37	91.14	95.57	96.16	81.61	80.65	87.91	94.71	95.64	**96.46**
κ	0.6723	0.85	0.8099	0.8633	0.9207	0.9421	0.7952	0.7348	0.8282	0.9103	0.9255	**0.9479**

Table 16. Performance comparisons between same training samples for each category (refer to the second row) and different training samples for each category (refer to the third row) on the WHU-Hi-Long Kou dataset.

Methods	KNN	RF	1D CNN	2D CNN	Hybrid SN	IRST 3DCNN	CasRNN	ViT	Spectral Former	GraphGST	SS-Mamba	SSUM
OA (%)	80.57	87.56	91.35	96.90	97.13	97.82	86.75	89.40	91.34	97.74	97.73	**98.00**
AA (%)	76.13	83.11	85.40	96.21	97.84	95.00	82.13	82.19	89.75	97.59	97.34	**98.48**
κ	0.7558	0.8405	0.8879	0.9594	0.9626	0.9646	0.8257	0.7254	0.8885	0.9704	0.9702	**0.9738**
OA (%)	88.36	91.57	90.71	97.70	98.26	97.20	83.46	91.28	91.34	98.02	95.06	**98.32**
AA (%)	59.04	71.76	74.26	93.72	95.04	89.56	73.57	74.95	74.95	**95.28**	82.63	95.22
κ	0.8439	0.888	0.8777	0.9708	0.9772	0.9631	0.8714	0.8852	0.8858	0.9741	0.9352	**0.9779**

5. Conclusions

In this article, a Mamba-based SSUM model is proposed to extract spectral and spatial features, which can reduce the computational complexity and improve the model's performance. Specifically, in the Spectral Mamba branch, an NSF strategy is proposed to reduce the interference arising from the spatial variability. Additionally, an innovative SS mechanism is introduced, which functions by scanning across the sub-spectrum dimension to better learn the details of spectral features. In the Spatial Mamba branch, an SM module is developed by integrating an SS2D with SA within a cohesive framework to effectively extract the spatial features of HSIs. Finally, the output feature of the Spectral Mamba and Spatial Mamba branch is united to comprehensively determine the category of the HSI. The ablation study validated the effectiveness of the proposed NSF, SS, and SM. Abundant comparison experiments with other developed HSIC methods demonstrated the superiority of the proposed SSUM on four HSI datasets. Specifically, our method achieved increases of 26.14%, 9.49%, 16.06%, and 8.99% in terms of OA compared to the baseline method on the Indian Pines, Pavia University, Salinas Valley, and WHU-Hi-LongKou datasets, respectively. HSIC has a multitude of potential applications across various fields, such as agriculture, environmental monitoring, mineral exploration, etc.

Author Contributions: S.L., M.Z., Y.H., C.W., J.W. and C.G. provided the methodology; S.L. wrote the original draft; S.L., M.Z., Y.H. and C.W. performed experiments; S.L., M.Z., Y.H., C.W., J.W. and C.G. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The Indian Pines, Pavia University, and Salinas Valley datasets are available at http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes (accessed on 1 September 2020). The WHU-Hi-LongKou dataset is available at http://rsidea.whu.edu.cn/resource_WHUHi_sharing.htm (accessed on 1 December 2020).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liao, D.; Shi, C.; Wang, L. A Spectral–Spatial Fusion transformer Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5515216. [CrossRef]
2. Lin, S.; Zhang, M.; Cheng, X.; Wang, L.; Xu, M.; Wang, H. Hyperspectral Anomaly Detection via Dual Dictionaries Construction Guided by Two-Stage Complementary Decision. *Remote Sens.* **2022**, *14*, 1784. [CrossRef]
3. Huo, Y.; Cheng, X.; Lin, S.; Zhang, M.; Wang, H. Memory-Augmented Autoencoder with Adaptive Reconstruction and Sample Attribution Mining for Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5518118. [CrossRef]
4. Xi, B.; Li, J.; Li, Y.; Song, R.; Hong, D.; Chanussot, J. Few-Shot Learning with Class-Covariance Metric for Hyperspectral Image Classification. *IEEE Trans. Image Process.* **2022**, *31*, 5079–5092. [CrossRef] [PubMed]
5. Lin, S.; Zhang, M.; Cheng, X.; Zhou, K.; Zhao, S.; Wang, H. Hyperspectral Anomaly Detection via Sparse Representation and Collaborative Representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 946–961. [CrossRef]
6. Cheng, X.; Zhang, M.; Lin, S.; Li, Y.; Wang, H. Deep Self-Representation Learning Framework for Hyperspectral Anomaly Detection. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5002016. [CrossRef]
7. Ahmad, M.; Shabbir, S.; Roy, S.K.; Hong, D.; Wu, X.; Yao, J. Hyperspectral Image Classification—Traditional to Deep Models: A Survey for Future Prospects. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 968–999. [CrossRef]
8. Chang, C.-I. *Hyperspectral Data Exploitation: Theory and Applications*; Wiley-Interscience eBooks: Hoboken, NJ, USA, 2007.
9. Chen, M.; Feng, S.; Zhao, C.; Qu, B.; Su, N.; Li, W.; Tao, R. Fractional Fourier-Based Frequency-Spatial–Spectral Prototype Network for Agricultural Hyperspectral Image Open-Set Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5514014. [CrossRef]
10. Yang, S.; Zhang, Y.; Jia, Y.; Zhang, W. Local Low-Rank Approximation with Superpixel-Guided Locality Preserving Graph for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 7741–7754. [CrossRef]
11. Kang, X.; Xiang, X.; Li, S.; Benediktsson, J.A. PCA-Based Edge-Preserving Features for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 7140–7151. [CrossRef]
12. Stone, J.V. Overview of Independent Component Analysis. In *Independent Component Analysis: A Tutorial Introduction*; MIT Press: Cambridge, MA, USA, 2004; pp. 5–11.
13. Li, W.; Du, Q.; Zhang, F.; Hu, W. Collaborative Representation Based K-nearest Neighbor Classifier for Hyperspectral Imagery. In Proceedings of the 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lausanne, Switzerland, 24–27 June 2014. [CrossRef]
14. Amini, S.; Homayouni, S.; Safari, A.; Darvishsefat, A.A. Object-based classification of hyperspectral data using Random Forest algorithm. *Geo-Spat. Inf. Sci.* **2018**, *21*, 127–138. [CrossRef]
15. Cheng, X.; Zhang, M.; Lin, S.; Zhou, K.; Zhao, S.; Wang, H. Two-Stream Isolation Forest Based on Deep Features for Hyperspectral Anomaly Detection. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 5504205. [CrossRef]
16. Fang, L.; He, N.; Li, S.; Plaza, A.J.; Plaza, J. A New Spatial–Spectral Feature Extraction Method for Hyperspectral Images Using Local Covariance Matrix Representation. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3534–3546. [CrossRef]
17. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [CrossRef]
18. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep Learning-Based Classification of Hyperspectral Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2094–2107. [CrossRef]
19. Chen, C.; Ma, Y.; Ren, G. Hyperspectral Classification Using Deep Belief Networks Based on Conjugate Gradient Update and Pixel-Centric Spectral Block Features. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4060–4069. [CrossRef]
20. Benediktsson, J.A.; Palmason, J.A.; Sveinsson, J.R. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 480–491. [CrossRef]
21. Tang, Y.; Feng, S.; Zhao, C.; Fan, Y.; Shi, Q.; Li, W.; Tao, R. An Object Fine-Grained Change Detection Method Based on Frequency Decoupling Interaction for High-Resolution Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5600213. [CrossRef]
22. Jia, S.; Shen, L.; Li, Q. Gabor Feature-Based Collaborative Representation for Hyperspectral Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1118–1129. [CrossRef]
23. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral Image Classification Using Dictionary-Based Sparse Representation. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3973–3985. [CrossRef]
24. Lin, S.; Zhang, M.; Cheng, X.; Shi, L.; Gamba, P.; Wang, H. Dynamic Low-Rank and Sparse Priors Constrained Deep Autoencoders for Hyperspectral Anomaly Detection. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 2500518. [CrossRef]
25. Huo, Y.; Qian, X.; Li, C.; Wang, W. Multiple Instance Complementary Detection and Difficulty Evaluation for Weakly Supervised Object Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 6006505. [CrossRef]
26. Lin, S.; Zhang, M.; Cheng, X.; Zhao, S.; Shi, L.; Wang, H. Hyperspectral Anomaly Detection Using Spatial–Spectral-Based Union Dictionary and Improved Saliency Weight. *Remote Sens.* **2023**, *15*, 3609. [CrossRef]
27. Hong, D.; Gao, L.; Yao, J.; Zhang, B.; Plaza, A.; Chanussot, J. Graph Convolutional Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5966–5978. [CrossRef]
28. Roy, S.K.; Krishna, G.; Dubey, S.R.; Chaudhuri, B.B. HybridSN: Exploring 3-D–2-D CNN Feature Hierarchy for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 277–281. [CrossRef]
29. Zhong, Z.; Li, J.; Luo, Z.; Chapman, M. Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 847–858. [CrossRef]

30. Xu, Y.; Zhang, L.; Du, B.; Zhang, F. Spectral–Spatial Unified Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5893–5909. [CrossRef]
31. Han, D.; Kim, J.; Kim, J. Deep Pyramidal Residual Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6307–6315. [CrossRef]
32. Zhong, S.; Zhang, Y.; Chang, C.-I. A Spectral–Spatial Feedback Close Network System for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 10056–10069. [CrossRef]
33. Chang, C.-I.; Ma, K.Y.; Liang, C.-C.; Kuo, Y.-M.; Chen, S.; Zhong, S. Iterative Random Training Sampling Spectral Spatial Classification for Hyperspectral Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3986–4007. [CrossRef]
34. Cai, Y.; Liu, X.; Cai, Z. BS-Nets: An End-to-End Framework for Band Selection of Hyperspectral Image. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 1969–1984. [CrossRef]
35. Li, W.; Chen, C.; Zhang, M.; Li, H.; Du, Q. Data augmentation for hyperspectral image classification with deep CNN. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 593–597. [CrossRef]
36. Hong, D.; Han, Z.; Yao, J.; Gao, L.; Zhang, B.; Plaza, A.; Chanussot, J. SpectralFormer: Rethinking Hyperspectral Image Classification with Transformers. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5518615. [CrossRef]
37. Vaswani, A.; Shazee, N.; Parmar, N.; Uszkorei, J.; Jones, L.; Gomes, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In Proceedings of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
38. Hang, R.; Liu, Q.; Hong, D.; Ghamisi, P. Cascaded Recurrent Neural Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5384–5394. [CrossRef]
39. Zhang, X.; Sun, Y.; Jiang, K.; Li, C.; Jiao, L.; Zhou, H. Spatial Sequential Recurrent Neural Network for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4141–4155. [CrossRef]
40. Jiang, M.; Su, Y.; Gao, L.; Plaza, A.; Zhao, X.L.; Sun, X.; Liu, G. GraphGST: Graph Generative Structure-Aware transformer for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5504016. [CrossRef]
41. Sun, L.; Zhao, G.; Zheng, Y.; Wu, Z. Spectral–Spatial Feature Tokenization transformer for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5522214. [CrossRef]
42. He, J.; Zhao, L.; Yang, H.; Zhang, M.; Li, W. HSI-BERT: Hyperspectral Image Classification Using the Bidirectional Encoder Representation from Transformers. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 165–178. [CrossRef]
43. Yang, X.; Cao, W.; Lu, Y.; Zhou, Y. Hyperspectral Image Transformer Classification Networks. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5528715. [CrossRef]
44. Wang, D.; Zhuang, L.; Gao, L.; Sun, X.; Huang, M.; Plaza, A. BockNet: Blind-Block Reconstruction Network with a Guard Window for Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5531916. [CrossRef]
45. Zou, J.; He, W.; Zhang, H. LESSFormer: Local-Enhanced Spectral-Spatial Transformer for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5535416. [CrossRef]
46. Li, J.; Zheng, K.; Liu, W.; Li, Z.; Yu, H.; Ni, L. Model-Guided Coarse-to-Fine Fusion Network for Unsupervised Hyperspectral Image Super-Resolution. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 5508605. [CrossRef]
47. Cai, Y.; Lin, J.; Hu, X.; Wang, H.; Yuan, X.; Zhang, Y.; Timofte, R.; Van Gool, L. Mask-guided Spectral-wise Transformer for Efficient Hyperspectral Image Reconstruction. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 17481–17490. [CrossRef]
48. Qi, W.; Huang, C.; Wang, Y.; Zhang, X.; Sun, W.; Zhang, L. Global–Local 3-D Convolutional Transformer Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5510820. [CrossRef]
49. Roy, S.K.; Deria, A.; Shah, C.; Haut, J.M.; Du, Q.; Plaza, A. Spectral–Spatial Morphological Attention Transformer for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5503615. [CrossRef]
50. Iban, D.; Fernandez-Beltran, R.; Pla, F.; Yokoya, N. Masked auto-encoding spectral–spatial transformer for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5542614.
51. Wang, D.; Zhuang, L.; Gao, L.; Sun, X.; Zhao, X.; Plaza, A. Sliding Dual-Window-Inspired Reconstruction Network for Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5504115. [CrossRef]
52. Xiang, P.; Ali, S.; Zhang, J.; Jung, S.K.; Zhou, H. Pixel-associated autoencoder for hyperspectral anomaly detection. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *129*, 103816. [CrossRef]
53. Li, J.; Zheng, K.; Gao, L.; Ni, L.; Huang, M.; Chanussot, J. Model-Informed Multistage Unsupervised Network for Hyperspectral Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5516117. [CrossRef]
54. Peng, Y.; Zhang, Y.; Tu, B.; Li, Q.; Li, W. Spatial–Spectral transformer with Cross-Attention for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5537415. [CrossRef]
55. Gu, A.; Johnson, I.; Goel, K.; Saab, K.; Dao, T.; Rudra, A.; Ré, C. Combining Recurrent, Convolutional, and Continuous-time Models with Linear State-Space Layers. *arXiv* **2021**, arXiv:2110.13985.
56. Gu, A.; Dao, T.; Ermon, S.; Rudra, A.; Ré, C. HiPPO: Recurrent Memory with Optimal Polynomial Projections. *arXiv* **2020**, arXiv:2008.07669.
57. Gu, A.; Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. *arXiv* **2023**, arXiv:2312.00752.
58. Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; Liu, Y. VMamba: Visual State Space Model. *arXiv* **2024**, arXiv:2401.10166.
59. Chen, K.; Chen, B.; Liu, C.; Li, W.; Zou, Z.; Shi, Z. RSMamba: Remote Sensing Image Classification with State Space Model. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 8002605. [CrossRef]

60. Ge, Y.; Chen, Z.; Yu, M.; Yue, Q.; You, R.; Zhu, L. MambaTSR: You only need 90k parameters for traffic sign recognition. *Neurocomputing* **2024**, *599*, 128104. [CrossRef]
61. Lin, S.; Zhang, M.; Cheng, X.; Zhou, K.; Zhao, S.; Wang, H. Dual Collaborative Constraints Regularized Low-Rank and Sparse Representation via Robust Dictionaries Construction for Hyperspectral Anomaly Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 2009–2024. [CrossRef]
62. Zhu, X.; Cheng, D.; Zhang, Z.; Lin, S.; Dai, J. An Empirical Study of Spatial Attention Mechanisms in Deep Networks. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6687–6696. [CrossRef]
63. Yuan, Z.; Gong, J.; Guo, B.; Wang, C.; Liao, N.; Song, J.; Wu, Q. Small Object Detection in UAV Remote Sensing Images Based on Intra-Group Multi-Scale Fusion Attention and Adaptive Weighted Feature Fusion Mechanism. *Remote Sens.* **2024**, *16*, 4265. [CrossRef]
64. Gu, A.; Goel, K.; Ré, C. Efficiently Modeling Long Sequences with Structured State Spaces. *arXiv* **2021**, arXiv:2111.00396.
65. Huang, L.; Chen, Y.; He, X. Spectral-Spatial Mamba for Hyperspectral Image Classification. *Remote Sens.* **2024**, *16*, 2449. [CrossRef]
66. Yang, A.; Li, M.; Ding, Y.; Fang, L.; Cai, Y.; He, Y. GraphMamba: An Efficient Graph Structure Learning Vision Mamba for Hyperspectral Image Classification. *arXiv* **2024**, arXiv:2407.08255v1. [CrossRef]
67. Li, Y.; Luo, Y.; Zhang, L.; Wang, Z.; Du, B. MambaHSI: Spatial–Spectral Mamba for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5524216. [CrossRef]
68. Zhong, Y.; Hu, X.; Luo, C.; Wang, X.; Zhao, J.; Zhang, L. WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. *Remote Sens. Environ.* **2020**, *250*, 112012. [CrossRef]
69. Cheng, X.; Huo, Y.; Lin, S.; Dong, Y.; Zhao, S.; Zhang, M.; Wang, H. Deep Feature Aggregation Network for Hyperspectral Anomaly Detection. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5033016. [CrossRef]
70. Lin, S.; Cheng, X.; Zeng, Y.; Huo, Y.; Zhang, M.; Wang, H. Low-Rank and Sparse Representation Inspired Interpretable Network for Hyperspectral Anomaly Detection. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5033116. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.