MCAM: A Multi-scale Cyclic Adaptive Mamba Network for Hyperspectral Image Classification

Yihang ZOU[1] Lina XU[1] Yanni DONG[2]

[1]Hubei Subsurface Multi-scale Imaging Key Laboratory, School of Geophysics and Geomatics,
China University of Geosciences, Wuhan, 430074, China, 757835177@qq.com; xulina@cug.edu.cn

[2]School of Resource and Environmental Sciences, Wuhan University, Wuhan, 430079, China, dongyanni@whu.edu.cn

Keywords: Hyperspectral Image Classification, Mamba, Deep learning, Spectral-spatial learning

Abstract:

Hyperspectral image (HSI) classification is one of the core tasks in the field of remote sensing, whose key lies in the effective fusion of spectral and spatial information. Among existing methods, convolutional neural networks (CNNs) are limited by their local receptive field, making it difficult to model long-range spectral dependencies, while Transformers, although capable of capturing global relationships, suffer from high quadratic computational complexity. To address these issues, this paper proposes a Multi-scale Cyclic Adaptive Mamba Network (MCAM) based on state-space models (SSM) for hyperspectral image classification. First, a multi-scale feature convolution block is introduced to extract spatial features from local to global levels in parallel, thereby enhancing feature representation. Subsequently, a cyclic adaptive scan module is incorporated to strengthen the modeling of long-range spectral–spatial dependencies. Furthermore, a combination of triplet loss and classification loss is adopted to improve the model’s discriminative ability in few-shot learning scenarios. Experiments conducted on the Indian Pines and Liao Ning-01 datasets demonstrate that MCAM outperforms existing mainstream methods in terms of overall accuracy (OA), average accuracy (AA), and Kappa coefficient, particularly excelling in class boundary clarity and spatial consistency. This study validates the efficiency and potential of the Mamba architecture in HSI classification and provides new insights for subsequent related research.

1. Introduction

Hyperspectral images (HSI) capture hundreds of narrow spectral bands across the electromagnetic spectrum, providing rich spatial and spectral information (Paoletti, 2019). Compared with conventional remote sensing imagery, HSI provides more detailed surface coverage data and enables the detection of subtle spectral differences among ground objects. As a result, hyperspectral remote sensing has been widely applied in precision agriculture, environmental monitoring(Calin, 2014), and mineral resource exploration (Bioucas-Dias, 2013).

The development of deep learning has significantly advanced HSI classification. Convolutional neural networks (CNNs) have demonstrated strong capabilities in extracting spectral and spatial features to improve classification accuracy. For example, Hu et al. (Hu, 2015) proposed a 1D-CNN to capture the spectral characteristics of HSI. These methods enable deeper integration of spectral and spatial information, thereby facilitating more accurate classification. However, due to the limited receptive field of convolutional operations, CNNs struggle to model long-range spectral dependencies (Ma, 2023;Duan, 2023).

Inspired by advances in natural language processing, Vision Transformer and its variants have been introduced into computer vision and rapidly applied to hyperspectral analysis(Sun,2022). Through the self-attention mechanism, Transformer can explicitly model global relationships among all pixels or spectral bands, capturing both long-range spatial correlations and remote dependencies along the spectral sequence (Gao, 2022). Nevertheless, the computational complexity and memory consumption of self-attention grow quadratically with the input sequence length. For hyperspectral images containing hundreds of bands, this results in prohibitive computational costs and training burdens, limiting their deployment in large-scale practical or resource-constrained scenarios (Gu, 2021).

Recently, the emergence of the Mamba architecture, based on structured state-space models (SSMs), has provided a revolutionary solution for long-sequence modeling. Mamba dynamically determines whether to propagate or discard

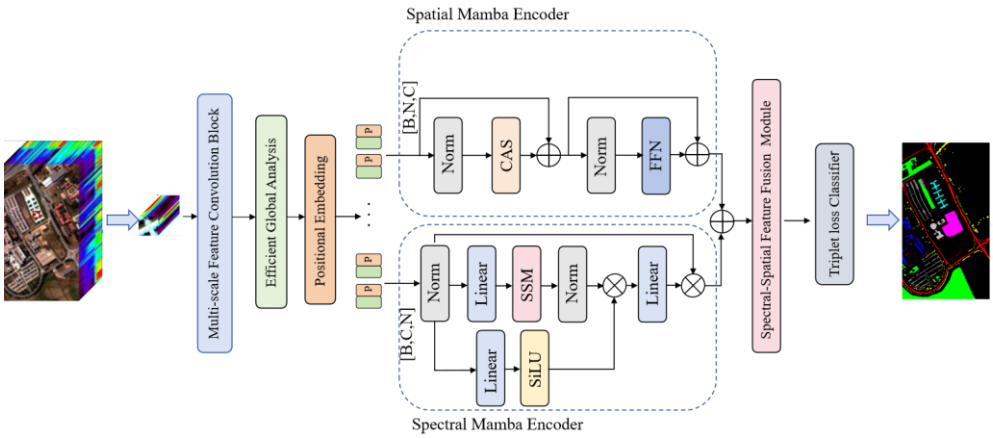


Figure 1. The illustration for the design details of the proposed MCAM-working flow.

information based on the input, enabling more flexible modeling of long-range dependencies. Crucially, Mamba achieves linear time complexity for sequence modeling, offering significantly superior computational efficiency compared to Transformer (Gu, 2023), while demonstrating stronger feature extraction capabilities on multiple long-sequence benchmarks.

Despite these advantages, directly applying Mamba to hyperspectral image classification remains challenging. This requires the model to possess not only strong sequence modeling capabilities but also adaptive perception of spatial context and the ability to effectively integrate multi-scale features.

To address these issues, this study proposes a multi-scale cyclic adaptive mamba method (MCAM). The proposed method is specifically designed to tackle key challenges in HSI classification, including modeling long-range spectral dependencies, extracting spatial multi-scale features, and handling high intra-class variability alongside high inter-class similarity.

2. Method

This experiment proposes a multi-scale cyclic adaptive mamba method (MCAM) for hyperspectral image classification, introducing a Cyclic Adaptive Scan Model that adapts to varying data distributions and feature samples. It generates adaptive feature maps based on spectral feature weights, enabling intelligent multi-view fusion and efficient deep feature extraction. Additionally, to optimize spatial-spectral features, a Multi-scale Feature Convolution Block is introduced to collaboratively model different-sized ground object regions and their rich spectral characteristics. Additionally, addressing the challenges of significant intra-class variability and high inter-class similarity in HSIs, the study combines triplet loss with classification loss to enhance the model's capability to distinguish confusing land cover categories in few-shot scenarios. The overall flowchart of the proposed MCAM method is shown in Figure 1.

2.1 Multi-scale Feature Convolution Block

Multi-scale Feature Convolution Block: Features output from the cyclic adaptive scanning module are fed into a multi-scale convolutional feature block. This block utilizes three parallel convolutional kernels of different sizes to simultaneously extract fine-grained local structures, medium-scale contextual information, and global spatial profiles. The extracted multi-scale features are then concatenated and nonlinearly fused, before being aggregated with the original input features via a residual connection. This process generates enhanced features that concurrently encapsulate rich spectral dependencies and multi-scale spatial contexts, and its structure is shown in Figure 2.

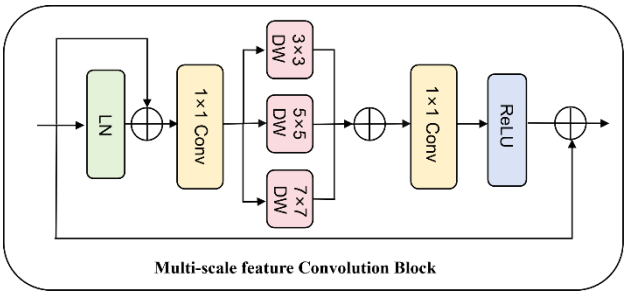


Figure 2. Multi-scale feature Convolution Block

To enable adaptive calibration of the input features' distribution and scale, LayerNorm and learnable scaling factors are incorporated at the input stage. Given the input feature x_0 the normalized feature x_{norm} is calculated as:

$$x_{\text{norm}} = \text{LayerNorm}(x_0) + \gamma \cdot x_0 \quad (1)$$

γ is a learnable parameter that balances the contribution between the normalized and the original input features.

We designed a multi-branch depthwise separable convolution-based visual filter bank to capture spatial contextual information

at different scales. The normalized features first undergo channel reduction via a 1×1 convolution:

$$f_{\text{down}} = \text{Conv}_{1\times1}(x_{\text{norm}}). \quad (2)$$

Subsequently, three depthwise separable convolutions (DWConv) with different kernel sizes are applied in parallel, where the kernel sizes are 3×3, 5×5, and 7×7, respectively. Their corresponding weights are denoted as: $\omega_{\text{dw}}^{i} \in \mathbb{R}^{C_{\text{in}}\times K_i\times K_i\times C_{\text{out}}}$ where i∈{1,2,3}. The outputs of the three branches are averaged and then added to the dimension-reduced features via a skip connection:

$$f_{\text{dw}} = f_{\text{down}} + \alpha \cdot \arg\left(\sum_{i=1}^{3} \omega_{\text{dw}}^{i} \,\widehat{\otimes}\, f_{\text{down}}\right) \quad (3)$$

where $\widehat{\otimes}$ denotes the depthwise separable convolution operation, and α is a learnable scaling factor.

In the feature aggregation and nonlinear activation stage, the fused features are further processed by a 1×1 pointwise convolution (PWConv) for cross-channel feature integration, with weights denoted as $\omega_{\text{pw}} \in \mathbb{R}^{C_{\text{in}}\times1\times1\times C_{\text{out}}}$ and introduce the skip connection:

$$f_{\text{pw}} = f_{\text{dw}} + \omega_{\text{pw}} \,\bar{\otimes}\, f_{\text{dw}}, \quad (4)$$

where $\bar{\otimes}$ denotes the pointwise convolution operation.The features are then passed through a ReLU activation function for nonlinear transformation.

$$f_{\text{out}} = \text{ReLU}(f_{\text{pw}}) \quad (5)$$

To preserve the original feature information and facilitate gradient flow, the module output is added to the original input via a residual connection:

$$x_{\text{out}} = x_0 + f_{\text{out}}. \quad (6)$$

2.2 Cyclic Adaptive Scan Model

To efficiently capture multi-directional contextual information without introducing the significant computational overhead of traditional multi-directional scanning (e.g., bidirectional or four-directional scanning), we propose a cyclic adaptive scanning model. Its core design is a lightweight module that performs adaptive feature fusion through multiple geometric transformations guided by global weights, enabling efficient modeling within a cyclic scanning framework, and its structure is shown in Figure 3.

We define a set of four fundamental reversible geometric transformation strategies, denoted as $T = \{T_0, T_1, T_2, T_3\}$, where T_0 represents the identity transformation, T_1denotes the transpose operation, T_2indicates horizontal flipping, and T_3 indicates vertical flipping. These transformations form the basis for observing

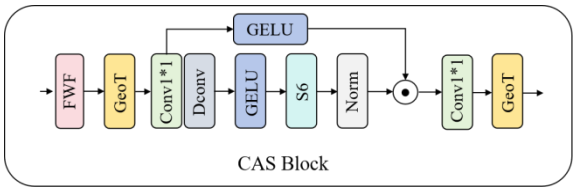


Figure 3. Cyclic Adaptive Scan Model

the feature map from different spatial perspectives. By alternately applying these geometric transformations to modify the spatial arrangement, the subsequent state-space scanning can capture spatial dependencies along different directions. The schematic graph of the geometric transformation is shown in Figure 4.

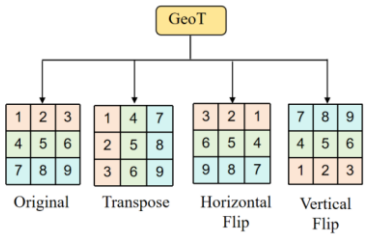


Figure 4. schematic graph of the geometric transformation

We further design a global weight generation module. For the input feature map $F_{\text{in}} \in \mathbb{R}^{H\times W\times C}$, global spatial information is first aggregated through global average pooling, followed by a two-layer multilayer perceptron (MLP). This lightweight module generates four corresponding global weight vectors $\text{w}_0, \text{w}_1, \text{w}_2, \text{w}_3 \in \mathbb{R}^{C}$, each associated with one transformation in T:

$$\text{w}_0, \text{w}_1, \text{w}_2, \text{w}_3 = \text{MLP}\big(\text{GAP}(F_{\text{in}})\big) \quad (7)$$

Adaptive Feature Fusion and Cyclic Scanning Structure.The model performs cyclic execution over four steps indexed by k∈{0,1,2,3}. At step k, the transformation T_k is designated as the dominant transformation.$w_j$are the global weight vectors corresponding to other non-dominant transformations.The corresponding weight w_k is fused with the other weights to form the adaptive fusion weight $\text{w}_{\text{fused}}^{k}$.

$$\text{w}_{\text{fused}}^{k} = \text{w}_k + \lambda \sum_{j\neq k} w_j \quad (8)$$

where λ is a learnable scaling factor. The input features are then modulated channel-wise using the fused weight and subsequently processed through the dominant geometric transformation.

$$F_{\text{mod}}^{k} = T_k\big(F_{\text{in}} \otimes \text{w}_{\text{fused}}^{k}\big) \quad (9)$$

Here, $\otimes$ denotes channel-wise multiplication. The transformed features F_{mod}^{k} are then processed by the selective scanning module. Finally, the inverse transformation T_k^{-1} is applied to the output of the S6 module to restore the features to their original spatial arrangement, yielding the output of step k. This cyclic process ensures comprehensive multi-directional perception.

2.3 Triplet loss Classifier Model

To enhance intra-class compactness and inter-class separability, we introduce a triplet-loss classification head after the feature extraction network. This module combines metric learning with classification supervision by constructing triplets of anchor, positive, and negative samples, enabling the network to learn more discriminative class-specific features.A schematic illustration of the triplet loss is shown in Figure 5.

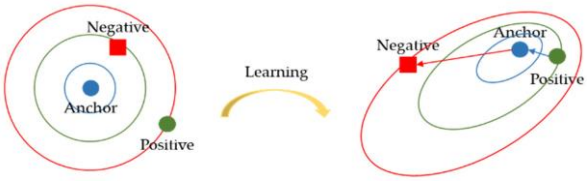


Figure 5. Triplet loss Classifier

Triplet construction strategy.Given a training batch, we randomly sample N triplets$\{(x_a^{(i)}, x_p^{(i)}, x_n^{(i)})\}_{i=1}^{N}$,where $x_a^{(i)}$ denotes the anchor sample, $, x_p^{(i)}$ is a positive sample belonging to the same class as the anchor, and $x_n^{(i)}$ is a negative sample from a different class. All samples are first mapped into feature vectors through the feature extraction network $\phi(\cdot)$.

The triplet loss function is formulated based on the Mahalanobis distance and is defined as follows:

$$\mathcal{L}_{\text{triplet}} = \frac{1}{N}\sum_{i=1}^{N} \max\left(d_M\left(\phi\left(\mathbf{x}_a^{(i)}\right), \phi\left(\mathbf{x}_p^{(i)}\right)\right) - d_M\left(\phi\left(\mathbf{x}_a^{(i)}\right), \phi\left(\mathbf{x}_n^{(i)}\right)\right) + \alpha, 0\right) \quad (10)$$

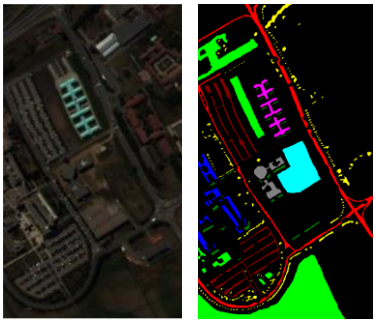
where $d_M(\mathrm{u}, \mathrm{v}) = (\mathrm{u} - \mathrm{v})^{\mathrm{T}}\mathrm{M}(\mathrm{u} - \mathrm{v})$ denotes the Mahalanobis distance, M is a positive semi-definite matrix that can be learned by the network, and α is a margin hyperparameter used to control the distance gap between positive and negative sample pairs.

3. Experiment

3.1 Data Description

To comprehensively evaluate the effectiveness of the proposed method, experiments are conducted on two widely used hyperspectral datasets: the University of Pavia (PaviaU) and Indian Pines.

1)Pavia University: The Pavia University dataset was acquired in 2003 over Pavia, Italy, using the Reflective Optics System Imaging Spectrometer (ROSIS-03). The sensor captured 115 continuous spectral bands within the wavelength range of 0.43–0.86 μm, with a spatial resolution of 1.3 m. After removing 12 noisy bands, the dataset has a spatial size of 610 × 340 pixels, totaling 2,207,400 pixels. These pixels are categorized into nine classes, including trees, asphalt, bricks, grass, and others. Figure 6. shows the false color and ground-truth maps, the detailed description of Pavia University dataset is shown in Table 1.



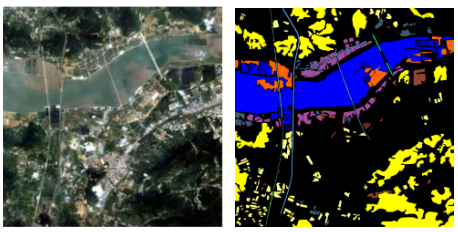
(a) (b)

Figure 6. Pavia University dataset: (a) false color map, (b) ground truth

Table 1. Dataset distribution by category Pavia University

No.	Color	Class	Total
1		Asphalt	6631
2		Meadows	18649
3		Gravel	2099
4		Trees	3064
5		Metal sheets	1345
6		Bare soil	5029
7		Bitumen	1330
8		Bricks	3682
9		Shadows	947
Total			42776

2) Liao Ning-01: The Liao Ning-01 (LN) dataset is generated by fusing multispectral and hyperspectral images acquired by the Zhongxing-1 02D satellite over the northwestern region of Jinpu New District, Dalian, Liaoning Province. The fusion is performed using an unsupervised deep spectral–spatial collaborative constraint method. The fused data have dimensions of 900 × 900 × 166, as shown in Figure 6.



(a) (b)

Figure 7. LN dataset: (a) false color map, (b) ground truth

A total of 265,004 pixels are annotated with ground-truth labels and categorized into 10 distinct classes. Detailed information about each class and the dataset partition is provided in Table 2.

Table 2. Dataset distribution by category LN

No.	Color	Class	Total
1		Reservoir	5476
2		Seawater	100276
3		Sandy soil	14587
4		Broken bridge	534
5		Barren grass	16499
6		Bare soil	5583
7		Highway	1430
8		Railway	6287
9		Mountain vegetation	95071
10		Arable land	19261
Total			265004

3.2 Experimental Setup

The method is implemented using the PyTorch framework and conducted on a system equipped with an Intel i7-10700K CPU, 32 GB RAM, and an NVIDIA GeForce RTX 3090 GPU. The model is trained for 200 epochs using the Adam optimizer. The initial learning rate for all datasets is set to 0.0005.

To quantitatively evaluate the performance of HSI classification, three metrics are adopted: Overall Accuracy (OA), Average Accuracy (AA), and the kappa coefficient (Kappa). OA is defined as the percentage of correctly classified samples over the total number of samples across all classes. AA is computed as the mean accuracy over all classes. The kappa coefficient measures the agreement between the ground truth (GT) and the classification results.

To evaluate the classification performance of MCAM, it is compared with several representative models, including CNN-based approaches, as well as two hybrid models that combine Transformer and Mamba, namely HSIMFormer (He, 2025) and 3DSS-Mamba (He, 2024), and a pure Transformer-based method, SQSFormer (Chen, 2024). All experiments are conducted using the hyperparameters recommended in the original papers of these methods.

3.3 Classification results

1) Pavia University dataset: Table 3 presents the detailed quantitative comparison results of different methods on the University of Pavia dataset, clearly demonstrating the superior performance of the proposed approach.

Specifically, our method achieves the highest OA of 98.95% ± 0.54, which is 3.21 percentage points higher than the second-best method (HSIMFormer), highlighting its significant advantage in overall classification performance. In terms of AA, our method reaches 97.32% ± 1.29, outperforming the second-best method (HSIMFormer) by 2.65 percentage points, further demonstrating its robustness and reliability in multi-class classification tasks. The Kappa coefficient also ranks the highest at 97.68 ± 1.63, which is substantially better than the other methods, indicating stronger consistency and stability in the classification results.

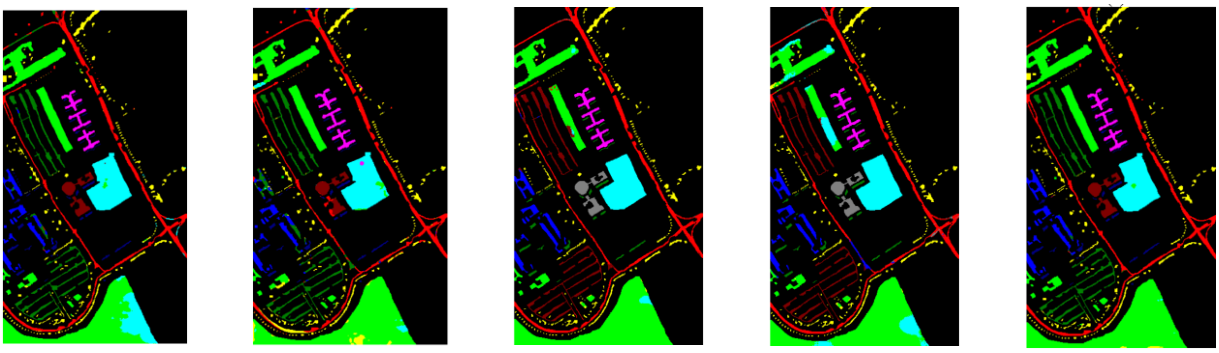
Table 3. Testing data classification results (mean ± standard deviation) on the Pavia University dataset

NO	CNN	3DSS-Mamba	HSIMFormer	SQSFormer	MCAM
1	81.27 ± 6.60	90.59 ± 5.26	93.85±2.80	90.83 ± 2.23	**94.15± 3.78**
2	83.13 ± 3.26	94.62 ± 4.88	95.11±2.31	95.17 ± 0.79	**95.42± 3.17**
3	81.60 ± 4.51	94.39 ± 4.19	92.94±3.60	66.64 ± 3.69	**99.71± 1.43**
4	95.29 ± 2.24	84.61 ± 6.87	93.77±2.70	91.51 ± 1.65	**99.276 ± 4.46**
5	99.26 ± 0.46	99.18 ± 0.60	99.26±0.38	99.93 ± 0.09	**100.00 ± 0.00**
6	80.27 ± 6.31	99.53 ± 1.08	96.10±1.72	97.27 ± 0.85	**99.50 ± 0.82**
7	93.11 ± 1.56	97.34 ± 0.67	96.97±1.93	99.14 ± 1.89	**99.82 ± 0.96**
8	85.86 ± 3.96	99.10 ± 2.75	94.65±2.44	88.14 ± 2.85	**98.81 ± 3.78**
9	99.85 ± 0.12	93.95 ± 3.14	98.03±1.19	99.94 ± 0.06	**99.92± 2.30**
AA(%)	84.53 ± 2.22	94.33 ± 2.85	94.67±1.06	92.85 ± 0.44	**97.32± 1.29**
OA(%)	88.63 ± 1.57	94.81 ± 1.77	95.74±0.83	92.06 ± 0.58	**98.95± 0.54**
K×100	80.00 ± 2.76	92.60 ± 3.63	93.01±1.22	90.61 ± 0.57	**97.68 ± 1.63**

The visual classification maps (as shown in Figure 7.) further corroborate the quantitative advantages described above. Our method exhibits clearer inter-class boundaries and higher regional consistency, particularly among classes with similar spectral characteristics. By enhancing feature discriminability and contextual perception, the proposed architecture effectively reduces scattered misclassifications and improves the overall spatial coherence and intra-class purity of the classification maps. Even in regions with complex structures and severe spectral overlap, the model maintains stable classification performance.

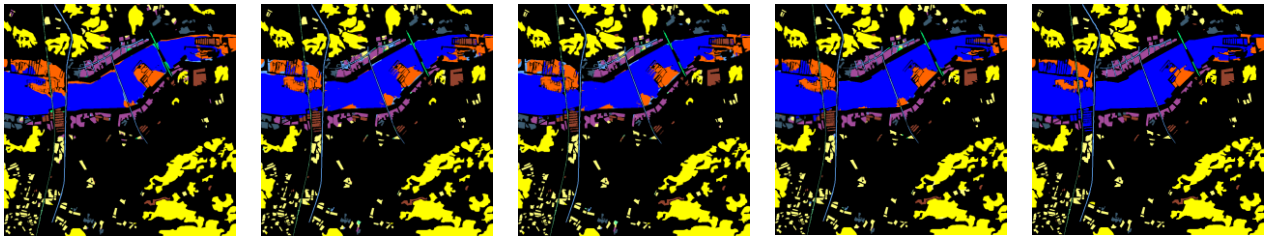
2) LN Dataset: The LN dataset contains a large number of samples and exhibits highly complex distributions of ground object features, posing greater challenges to the model’s discriminative capability and generalization performance. As shown in Table 4, the dataset consists of ten typical landcover classes with diverse sample compositions and intricate feature distributions, which significantly increases the difficulty of feature extraction and classification.

Experimental results demonstrate that our method consistently outperforms all competing approaches in terms of the three core



(a) (b) (c) (d) (e)

Figure 7. Classification maps using different methods on the Pavia University dataset. (a) CNN,(b) 3DSS-Mamba,(c) SQSFormer, (d) HSIMFormer,(e) MCAM.



(a) (b) (c) (d) (e)

Figure 8. Classification maps using different methods on the LN dataset. (a) CNN,(b) 3DSS-Mamba,(c) SQSFormer, (d) HSIM-Former,(e) MCAM.

Table 4.Testing data classification results on the LN dataset

NO	CNN	3DSS-Mamba	SQSFormer	HSIMFormer	MCAM
1	85.89±2.04	88.76±3.41	94.17±2.70	97.51±3.73	**99.23±1.27**
2	91.19±0.84	90.84±4.27	92.82±1.99	86.1±1.35	**93.89±1.92**
3	92.69±2.52	95.77±0.54	98.74±3.86	93.22±0.85	**98.89±2.63**
4	92.23±1.20	89.96±0.55	96.33±0.45	98.57±4.95	**99.35±0.37**
5	88.43±2.31	95.79±0.00	**98.08±1.89**	90.22±4.63	94.02±0.21
6	98.85±7.24	89.33±1.63	**97.68±5.78**	84.3±3.03	94.86±3.29
7	88.86±2.20	91.84±0.20	94.59±8.24	93.27±3.62	**97.87±0.86**
8	80.56±7.59	91.49±0.95	89.4±2.08	92.11±2.18	**95.23±2.76**
9	87.96±1.64	83.1±0.05	90.98±1.21	92.21±2.41	**94.97±1.75**
10	95.32±6.61	96.06±0.12	**95.61±2.40**	92.84±1.62	91.78±0.19
AA(%)	85.63±1.18	92.63±2.27	93.42±1.40	91.98±1.67	**95.18±1.01**
OA(%)	91.61±1.93	90.23±0.77	94.21±1.53	92.64±1.73	**94.84±0.97**
K×100	82.36±1.32	88.42±2.92	91.68±1.88	89.07±2.42	**92.65±1.59**

Table 5. Progressive ablation study of the MCAM framework on the Pavia University dataset

Backbone	CAS	MFC	Triplet	OA	AA	Kappa
✓				93.24	94.93	94.02
✓	✓			95.69	96.71	95.54
✓	✓	✓		96.68	97.48	96.19
✓	✓	✓	✓	97.32	98.95	97.68

metrics: Overall Accuracy (OA), Average Accuracy (AA), and the Kappa coefficient. Specifically, our method achieves the highest OA of 94.84% ± 0.97, surpassing the second-best method (SQSFormer) by 0.63 percentage points. In terms of AA, it reaches 95.18% ± 1.01, which is 1.76 percentage points higher than that of SQSFormer (93.42% ± 1.40). Moreover, our method attains the highest Kappa coefficient (92.65 ± 1.59), further confirming its superior classification consistency and reliability.

The visualization results in Figure 8. further illustrate that, through multi-level feature fusion and high-order contextual modeling, the proposed method significantly improves the spatial coherence and intra-class consistency of the classification maps. In particular, it demonstrates clearer boundary preservation between easily confused classes, such as roads and water bodies, as well as among different vegetation types.

In summary, the proposed method not only achieves overall superiority in quantitative metrics but also demonstrates better detail preservation and structural integrity in the visual classification maps, validating its strong discriminative capability and stability when handling complex scenes and diverse feature variations.

3.4 Ablation Experiments

To further verify the effectiveness of each component in the proposed MCAM framework, ablation experiments were conducted on the Pavia University dataset. Three key modules were individually removed from the full model, including the Cyclic Adaptive Scan (CAS) module, the Multi-scale Feature Convolution (MFC) block, and the triplet loss, while keeping the remaining structure unchanged. The experimental results are summarized in Table 5. Specifically, the Cyclic Adaptive Scan Model enables the model to perceive spatial–spectral dependencies from multiple geometric perspectives without introducing excessive computational overhead. By adaptively fusing features under different transformations, the model establishes long-range contextual relationships that cannot be captured by conventional sequential scanning or local convolution operations.

The multi-branch convolutional structure of the MFC block aggregates spatial information across different receptive fields, allowing the model to simultaneously capture fine-grained local structures of ground objects as well as broader contextual profiles. Finally, by incorporating the triplet loss to form the complete MCAM framework, the model achieves optimal performance across all evaluation metrics. The metric learning constraint enhances intra-class compactness and inter-class separability in the feature space, and the combination of classification supervision and metric learning further strengthens the discriminative capability of the learned representations.

4. Conclusion

This study addresses several key challenges in hyperspectral image (HSI) classification, including the difficulty of modeling long-range spectral dependencies, insufficient extraction of spatial multi-scale features, and the issues of high intra-class variability and high inter-class similarity. To this end, we propose a MCAM. The core innovation of the proposed method lies in the integration and optimization of cyclic adaptive scanning and multi-scale feature fusion mechanisms, which effectively enhance the spectral-spatial feature representation and classification accuracy of HSI.The experimental results demonstrate that, by integrating the advanced Mamba architecture with the proposed multi-scale cyclic adaptive design, MCAM provides an efficient and accurate solution for hyperspectral image classification. In future work, we will further explore the potential of hybrid Mamba architectures and extend their application to other vision tasks.

References

Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral remote sensing data analysis and future challenges. IEEE Geosci. Remote Sens. Mag. 2013, 1, 6–36.

Gao, Z.; Shi, X.; Wang, H.; Zhu, Y.; Wang, Y.B.; Li, M.; Yeung, D.-Y. Earthformer: Exploring space-time transformers for earth system forecasting. Adv. Neural Inf. Process. Syst. 2022, 35, 25390–25403.

Gu, A.; Goel, K.; Ré, C. Efficiently modeling long sequences with structured state spaces. arXiv 2021, arXiv:2111.00396.

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

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

M. A. Calin, S. V. Parasca, D. Savastru, and D. Manea, “Hyperspectral imaging in the medical field: Present and future,” Appl. Spectrosc. Rev.,vol. 49, no. 6, pp. 435–447, Aug. 2014.

Ma, X.; Kang, X.; Qin, H.; Wang, W.; Ren, G.; Wang, J.; Liu, B. A Lightweight Hybrid Convolutional Neural Network for Hyperspectral Image Classification. IEEE Trans. Geosci. Remote Sens. 2023, 61, 1–14.

N. Chen, L. Fang, Y. Xia, S. Xia, H. Liu and J. Yue, "Spectral Query Spatial: Revisiting the Role of Center Pixel in Transformer for Hyperspectral Image Classification," in IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1-14, 2024, Art no. 5402714, doi: 10.1109/TGRS.2024.3361652.

Paoletti, M.E.; Haut, J.M.; Plaza, J.; Plaza, A. Deep learning classifiers for hyperspectral imaging: A review. ISPRS J. Photogramm Remote Sens. 2019, 158, 279–317.

Q. Zhu et al., “Samba: Semantic segmentation of remotely sensed images with state space model,” *arXiv:2404.01705*, 2024.

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, Jan. 2015.

Y. Duan, F. Luo, M. Fu, Y. Niu, and X. Gong, “Classification via structure-preserved hypergraph convolution network for hyperspectral image,” IEEE Trans. Geosci. Remote Sens., vol. 61, 2023, Art. no.5507113.

Y. He, B. Tu, B. Liu, J. Li and A. Plaza, "3DSS-Mamba: 3D-Spectral-Spatial Mamba for Hyperspectral Image Classification,"

in I*EEE Transactions on Geoscience and Remote Sensing,* vol. 62, pp. 1-16, 2024

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