Supplementary Material

Multi-Scale Lightweight Spectral Attention Network for Hyperspectral Image Classification

This file includes:

1.Attention Module Interpretability Analysis

2.Confusion matrix on the IP dataset

3.Comparative Analysis of Inference Efficiency

4.Effectiveness Analysis of Structural Improvements

5.Table S3. Comparison of LPPA Components and SimAM on the Indian Pines Dataset

6.Table S4. Channel Masking Analysis Using LPPA Scores

7.Table S5. Sensitivity Analysis of the Offset Term in LPPA

8.Table S6. Sensitivity Analysis of Training Sample Proportion on the Indian Pines Dataset

9.Table S7. Sensitivity Analysis of Training Sample Proportions on the Pavia Centre Dataset

10.Table S8. Sensitivity Analysis of Training Sample Proportions on the Salinas Valley Dataset

1.Attention Module Interpretability Analysis

To further interpret the superior classification performance of MLSANet, this study visualizes attention heatmaps of the LPPA and PADA modules on samples from the Indian Pines dataset. The LPPA heatmaps are used to characterize the model's ability to suppress spectral redundancy and select informative channel responses, while the PADA heatmaps are used to capture spatial context enhancement and pixel-wise response modulation. In the heatmaps, warmer colors (yellow/orange-red) indicate higher attention responses, whereas cooler colors (dark blue) correspond to lower responses. It should be noted that these attention heatmaps are primarily used to qualitatively illustrate the response distribution characteristics of the proposed modules, while quantitative performance is evaluated using metrics such as OA, AA, and Kappa.

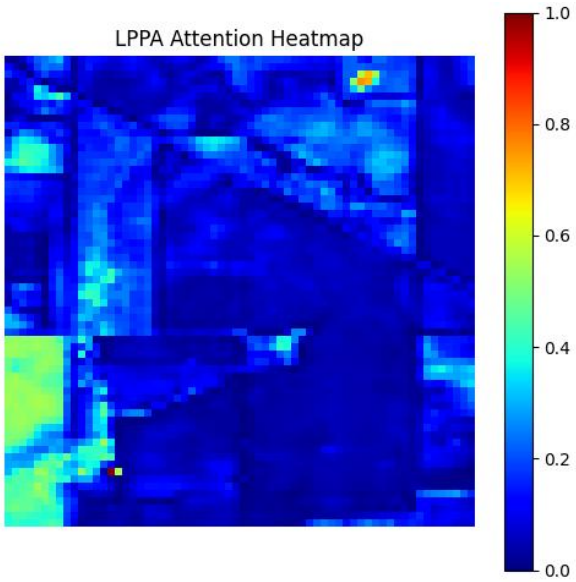


Figure S1 Attention heatmap of the LPPA spectral response pruning module

Figure S1 shows the attention heatmap of the LPPA pruning mechanism. The LPPA module is designed to address the strong spectral redundancy and high inter-channel response correlation in hyperspectral data; thus, its heatmaps primarily reflect spectral response selection and redundancy suppression behavior. It can be observed that the attention distribution exhibits a clearly sparse pattern, where most low-response regions are effectively suppressed, while high attention weights are retained only at a few key feature locations. This suggests that the threshold-based pruning mechanism can effectively filter low-contribution responses, thereby reducing interference from redundant information in subsequent classification. These observations are consistent with the design objective of LPPA, which aims to enhance discriminative feature responses and suppress redundant spectral information through channel-wise energy modeling and pruning operations.

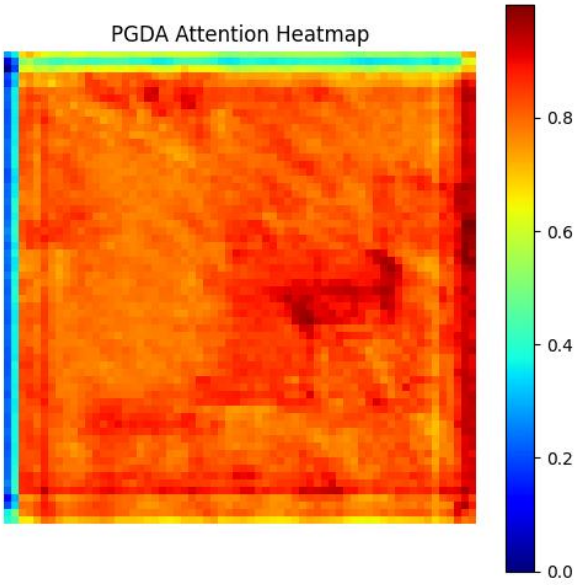


Figure S2 Pixel-wise attention heatmap of the PADA module

Figure S2 shows the attention heatmap of the PADA module. The PADA module is designed to address spatial context modeling in hyperspectral images; therefore, its heatmaps primarily reflect the adaptive allocation of importance across different spatial locations. It can be observed that high-attention regions are mainly concentrated in areas with more pronounced structural information within each patch, while boundary regions or weakly correlated areas exhibit relatively lower responses. This indicates that PADA can adaptively modulate responses at different spatial locations through dilated convolutions, multi-scale spatial feature extraction, and pixel-wise attention gating. Unlike the stronger pruning-based suppression in LPPA, PADA tends to enhance key spatial regions through continuous soft-weight allocation while retaining certain weak responses, making it more suitable for contextual modeling in complex spatial structures.

Overall, LPPA and PADA exhibit distinct functional emphases in their attention distributions. LPPA focuses on suppressing redundant responses and selecting key channel features, whereas PADA emphasizes spatial context enhancement and pixel-wise response modulation. Together, they are complementary from the perspectives of spectral redundancy suppression and spatial structure modeling, providing a visual explanation for the improvement in classification performance.

2.Confusion matrix on the IP dataset

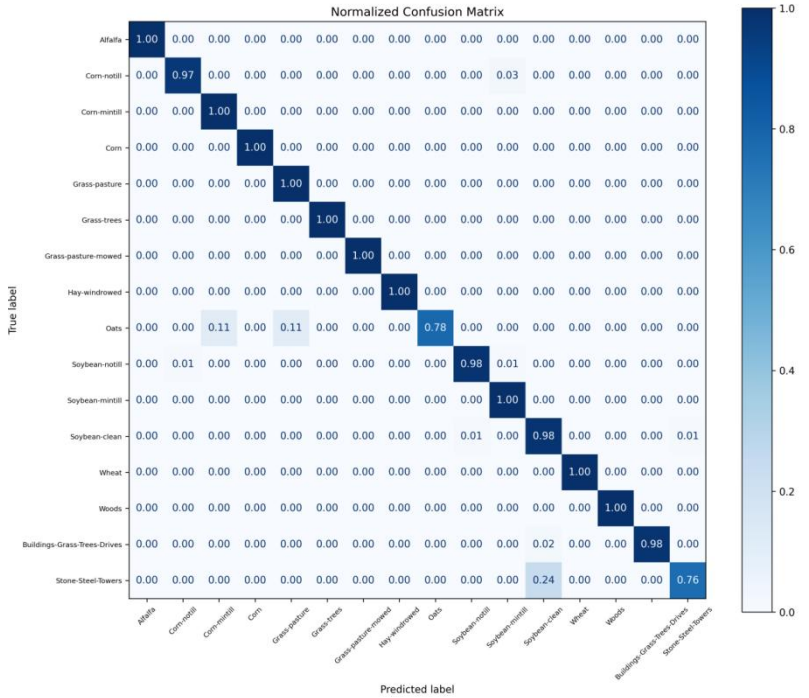


Figure S3 Confusion matrix of the baseline method on the Indian Pines dataset (IP)

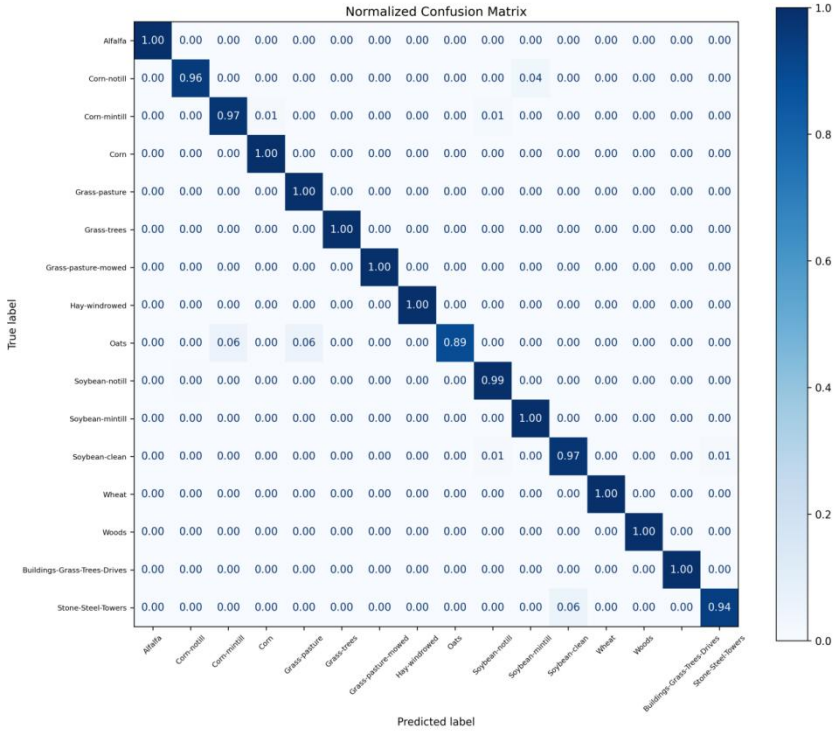


Figure S4 Confusion matrix of the full model on the Indian Pines dataset (IP)

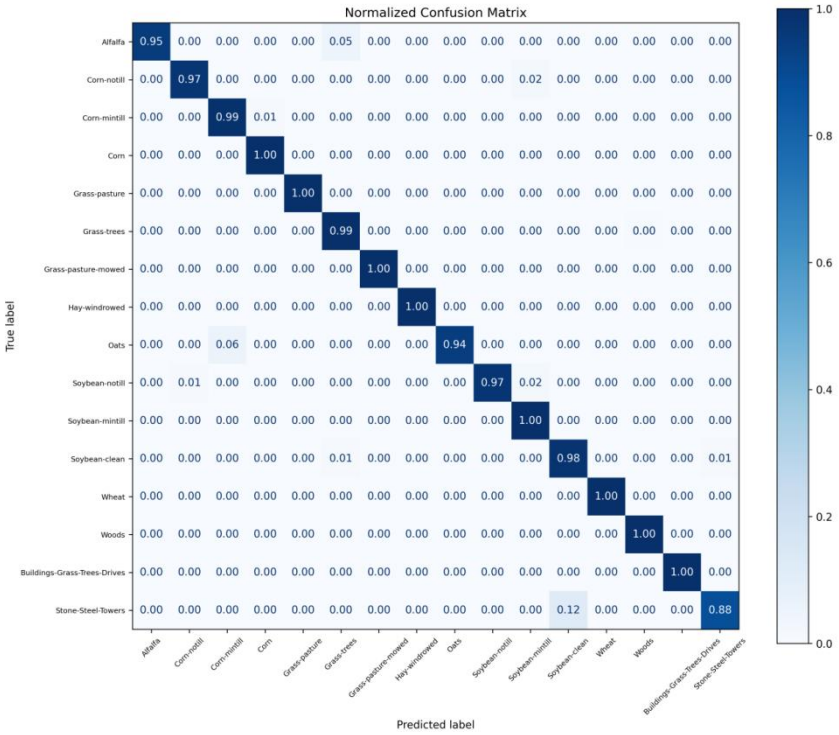


Figure S5 Confusion matrix of the model without the LMLP module on the Indian Pines dataset (IP).

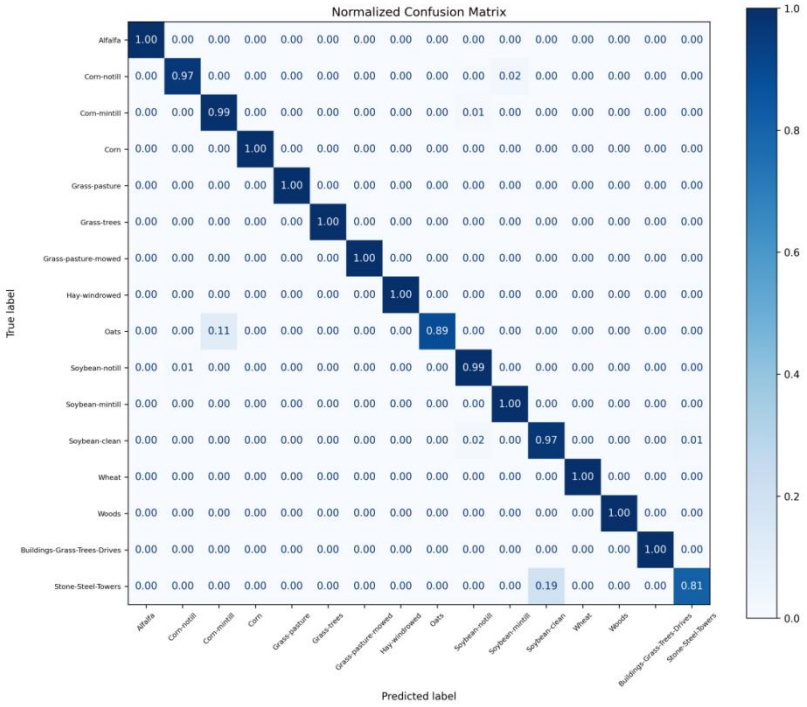


Figure S6 Confusion matrix of the model without the LPPA module on the Indian Pines dataset (IP).

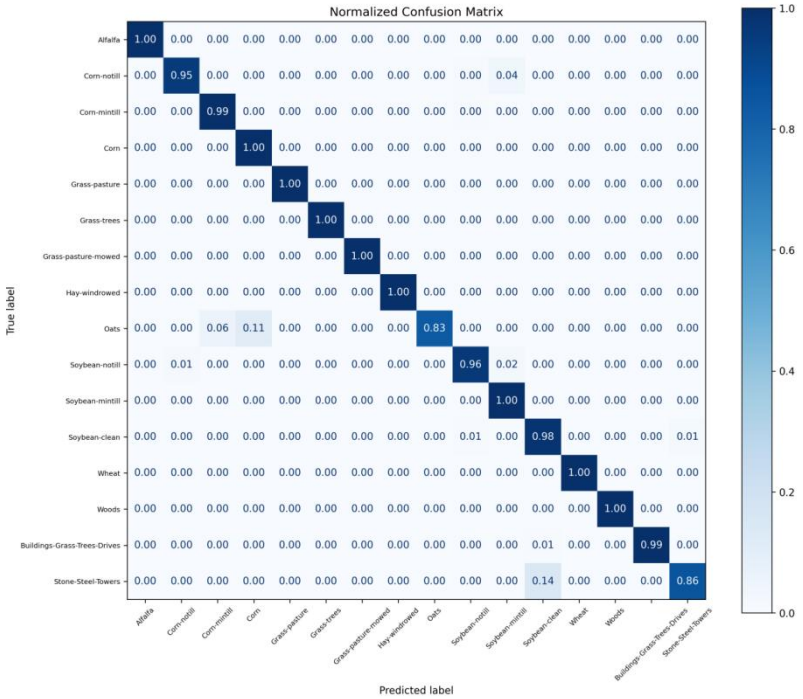


Figure S7 Confusion matrix of the model without the PADA module on the Indian Pines dataset (IP)

3.Comparative Analysis of Inference Efficiency

The inference efficiency comparison in Table S1 shows that MLSANet achieves favorable inference performance on different hardware platforms. It should be noted that the baseline denotes the basic network structure without the proposed core modules, including LPPA, PADA, and the structure-aware representation layer, while Ours (MLSANet) denotes the complete model with these modules.

On the GPU platform with an NVIDIA 1080Ti, the inference time of the baseline is 83 ms per image, whereas MLSANet reduces the inference time to 42 ms per image. On the CPU platform with an Intel i7 processor, the inference time decreases from 348 ms for the baseline to 180 ms for MLSANet. These results indicate that MLSANet can maintain high inference efficiency after introducing the spectral redundancy suppression, spatial context enhancement, and structure-aware modeling modules.

Overall, compared with the baseline without the proposed core modules, MLSANet achieves lower inference time on both GPU and CPU platforms. This demonstrates that the lightweight feature extraction and compact structural design of MLSANet can effectively reduce redundant computation and improve inference efficiency, showing good applicability across different hardware platforms.

Table S1 Comparative Analysis of Inference Efficiency

Model Variant	Inference Performance	InferenceTime (ms/image)	Speedup Ratio
baseline	GPU(1080Ti)	83ms	1.00×
MLSANet	GPU(1080Ti)	42ms	1.98×
baseline	CPU(i7)	348ms	1.00×
MLSANet	CPU(i7)	180ms	1.93×

4.Effectiveness Analysis of Structural Improvements

As shown in Table S2, MLSANet significantly reduces the computational complexity compared with the baseline on all three datasets. It should be noted that the baseline denotes the basic network structure without the proposed core modules, including LPPA, PADA, and the structure-aware representation layer, while MLSANet denotes the complete model with these modules.

Specifically, on the IP dataset, the FLOPs of MLSANet decrease from 25.92M to 8.90M, while OA, AA, and Kappa increase from 95.48%, 95.59%, and 94.85% to 99.01%, 98.32%, and 98.87%, respectively. This indicates that the proposed modules can effectively improve classification performance while reducing computational cost. On the PC dataset, the FLOPs decrease from 23.61M to 7.32M, and OA, AA, and Kappa increase to 98.66%, 96.29%, and 98.10%, respectively, demonstrating that the proposed structure has strong feature representation capability under small-sample and complex class-distribution conditions.

On the SA dataset, the FLOPs decrease from 25.52M to 6.58M, and AA increases from 97.21% to 98.26%, indicating that the model improves the average recognition ability across different categories. However, OA and Kappa show a slight decrease compared with the baseline, suggesting that there remains a trade-off between overall accuracy and model lightweight design on this dataset. Overall, MLSANet can substantially reduce FLOPs while maintaining or improving classification performance on most metrics, demonstrating that the proposed lightweight structure achieves a favorable balance between computational efficiency and classification accuracy.

Table S2 Impact of Lightweight Architecture on Classification Performance

Dataset	Model	FLOPs	OA(%)	AA(%)	Kappa(%)
IP	baseline	25.92	95.48	95.59	94.85
	MLSANet	8.902	99.01	98.32	98.87
PC	baseline	23.61	94.97	89.00	94.69
	MLSANet	7.3220	98.66	96.29	98.10
SA	baseline	25.52	96.02	97.21	95.57
	MLSANet	6.5811	95.88	98.26	95.42

5.Table S3. Comparison of LPPA Components and SimAM on the Indian Pines Dataset

Table S3 Effectiveness evaluation of LPPA components compared with SimAM (IP)

Method	SimAM	**Variance Weighting**	**Pruning**	**Gamma**	**OA (%)**	**AA (%)**	**Kappa(%)**
reference（baseline）	×	×	×	×	97.97	98.56	98.83
SimAM	√	×	×	×	**98.88**	98.22	98.73
LPPA-variance	×	√	×	×	98.96	97.97	98.81
LPPA-pruning	×	√	√	×	98.96	98.43	98.81
LPPA-gamma	×	√	×	√	98.55	98.71	98.34
Full LPPA	×	√	√	√	**99.01**	98.32	98.87

6.Table S4. Channel Masking Analysis Using LPPA Scores

Table S4 Channel masking analysis between LPPA scores and classification performance

Setting	Masked Channels	OA (%)	OA Drop (%)	AA (%)	Kappa (%)
Original full features	0.00	98.54	0.00	96.15	98.33
Mask top 10% LPPA-score channels	6.00	98.44	0.10	97.45	98.22
Mask top 25% LPPA-score channels	16.00	**80.16**	18.37	91.57	77.79
Mask bottom 10% LPPA-score channels	6.00	98.08	0.46	91.75	97.81
Mask bottom 25% LPPA-score channels	16.00	92.59	5.95	68.14	91.50
Mask random 10% channels	6.00	98.25	0.28	98.53	98.01
Mask random 25% channels	16.00	97.56	0.98	86.30	97.22

7.Table S5. Sensitivity Analysis of the Offset Term in LPPA

Table S5 Sensitivity analysis of the offset term in LPPA

Offset	**OA (%)**	**AA (%)**	**Kappa (%)**
0.40	98.81	97.54	98.64
0.50	98.86	97.79	98.70
0.60	96.81	94.59	96.36

8.Table S6. Sensitivity Analysis of Training Sample Proportion on the Indian Pines Dataset

Table S6 Sensitivity analysis of training sample proportions on the IP dataset

IP	**5%**	**10%**	**15%**	**20%**
OA	95.27	99.00	96.71	97.65
AA	80.36	98.50	95.05	95.68
Kappa	94.59	98.56	96.25	97.33

9.Table S7. Sensitivity Analysis of Training Sample Proportions on the Pavia Centre Dataset

Table S7 Sensitivity analysis of training sample proportions on the PC dataset

PC	**0.1%**	**0.2%**	**0.3%**	**0.4%**	**0.5%**	**0.7%**	**1%**
OA	94.30	95.60	94.05	98.58	98.17	98.23	98.24
AA	85.64	83.76	80.92	96.16	93.55	93.55	94.38
Kappa	91.92	93.76	91.52	98.42	97.40	97.50	97.50

10.Table S8. Sensitivity Analysis of Training Sample Proportions on the Salinas Valley Dataset

Table S8 Sensitivity analysis of training sample proportions on the SA dataset

SA	**0.1%**	**0.3%**	**0.5%**	**0.7%**	**1%**
OA	83.83	86.61	97.02	97.98	98.20
AA	88.54	90.36	97.88	98.72	99.09
Kappa	81.90	84.98	96.69	97.76	98.00