KGLP-NET: KALMAN-RELIABLE GLOBAL-LOCAL LEARNING WITH GROUP FISHER PRUNING FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Existing global–local HSI classifiers rarely account for the relative uncertainty of complementary information sources, while model compression typically ignores the uncertainty estimated during fusion. To address this limitation, we propose **Kalman-Reliable Global–Local Learning with Group Fisher Pruning (KGLP-Net)**, which couples uncertainty-aware fusion with reliability-guided structured pruning. KGLP-Net employs an asymmetric dual-view architecture, where a Global Mamba Encoder captures long-range contextual dependencies and a Local CNN Encoder preserves fine-grained target-neighborhood information, providing complementary representations with distinct uncertainty characteristics. **Kalman-Reliable Fusion (KRF)** treats the global and local representations as a prior and an observation, respectively, and performs an uncertainty-guided, Kalman-inspired posterior update to jointly estimate a fused posterior representation and its uncertainty. **Kalman-Reliable Group Fisher Pruning (KRGF)** further transfers the posterior uncertainty into group-level reliability via normalized posterior precision and integrates it with Group Fisher importance, allowing the uncertainty to directly inform structured pruning. On all four public HSI datasets, Compact48, the KRGF-compressed model with hidden width 48, achieves the highest OA among ten representative competing classifiers. Compared with Dense64, it reduces parameters, FLOPs, and inference time by 38.5%, 34.8%, and 37.6% on average, respectively, while incurring only a 0.03 percentage point decrease in mean OA.

1 INTRODUCTION

Hyperspectral images (HSIs) provide rich spectral signatures for fine-grained land-cover recognition, yet classification remains challenging because of within-class spectral variability, inter-class similarity, and mixed pixels near class boundaries (Ghamisi et al., 2018; He et al., 2018; Li et al., 2019). Deep HSI classifiers initially focused on learned spectral representations (Chen et al., 2014) and subsequently advanced toward joint spectral–spatial and 3-D convolutional modeling (Ben Hamida et al., 2018; Zhong et al., 2018; Roy et al., 2020). Transformers, rooted in self-attention (Vaswani et al., 2017), enable broader contextual modeling of spectral–spatial dependencies (Hong et al., 2022; Sun et al., 2022; Mei et al., 2022). More recently, advances in state-space sequence modeling, from structured SSMs such as S4 to selective state-space models such as Mamba and their visual adaptations (Gu et al., 2022; Gu & Dao, 2024; Zhu et al., 2024; Liu et al., 2024b), have motivated efficient long-range HSI models such as MambaHSI (Li et al., 2024b). Despite these advances, achieving an effective balance between broad contextual information and fine-grained spectral–spatial cues around the target remains challenging in patch-wise HSI classification.

Recent HSI architectures exploit complementary information through global–local or dual-branch designs (Qi et al., 2023; Zhou et al., 2024; Meng et al., 2025; Pan et al., 2025; Sheng et al., 2025), while center–neighborhood relationships have also been explicitly modeled (Li et al., 2023). **Complementarity, however, does not imply equal reliability.** Global information can provide useful structural cues but may introduce irrelevant background, whereas the local information is more target-focused but may lack sufficient context. Their relative uncertainty may therefore vary across inputs and feature representations. Common fusion mechanisms such as summation, concatenation,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

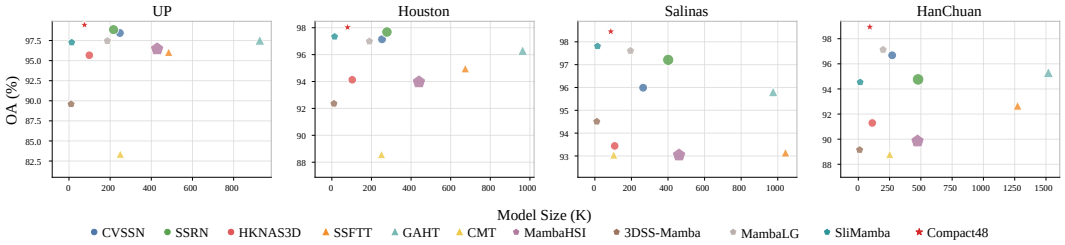


Figure 1: Accuracy–compactness comparison on four HSI datasets. The horizontal axis denotes parameter count, the vertical axis denotes OA, and bubble size represents FLOPs.

attention, and gating primarily operate on feature responses, leaving source-level uncertainty implicit. This distinction motivates uncertainty-aware global–local fusion.

Efficiency is also important for resource-constrained HSI deployment. Lightweight HSI models reduce complexity through compact CNNs, morphology-inspired modules, channel/layer redesign, and groupwise separable Transformers (Jia et al., 2021; Zhou et al., 2023; Li et al., 2024a; Zhao et al., 2024). Model compression has progressed from unstructured pruning and quantization (Han et al., 2016) to structured channel/filter pruning (He et al., 2017; Luo et al., 2017; Li et al., 2017; Liu et al., 2017) and task-aware or dependency-aware criteria (Molchanov et al., 2019; Liu et al., 2021; Fang et al., 2023). These criteria generally do not explicitly incorporate uncertainty estimated during feature fusion. Since task importance measures classification sensitivity whereas posterior reliability measures relative precision, uncertainty can provide a complementary signal for structure selection.

Based on these observations, we propose **Kalman-Reliable Global–Local Learning with Group Fisher Pruning (KGLP-Net)**, which links uncertainty-aware global–local fusion to reliability-guided structured pruning. KRF estimates relative uncertainty during fusion, while KRGF propagates the resulting posterior reliability into Group Fisher-based structure selection. As shown in Figure 1, Compact48 achieves a favorable accuracy–compactness trade-off across four public HSI datasets. Our main contributions are summarized as follows:

- **Uncertainty-aware global–local representation learning.** We construct a dual-view framework with a Global Mamba Encoder and a Local CNN Encoder to learn complementary representations from different spatial scales and modeling mechanisms. KRF further estimates their relative uncertainties and uses them to guide a Kalman-inspired fusion update, retaining posterior uncertainty for subsequent reliability estimation.
- **Reliability-guided structured pruning.** We introduce Kalman-coupled groups that jointly organize the corresponding global-prior, local-observation, and posterior channels. We further propose KRGF, which combines Group Fisher importance with posterior reliability for the selection of coupled groups and compact-network reconstruction.
- **Favorable accuracy–efficiency trade-off.** Extensive experiments and mechanism analyses demonstrate that uncertainty-derived reliability consistently refines Fisher-based pruning, yielding compact models with a favorable accuracy–efficiency trade-off. Moreover, Compact48 outperforms ten representative competing classifiers on four public HSI datasets.

2 Related Work

Global–Local Hyperspectral Representation Learning. CNN-based architectures effectively extract local spectral–spatial features (Zhong et al., 2018; Roy et al., 2020), while center–neighborhood relationships have been explicitly modeled to account for the distinct contributions of neighboring pixels (Li et al., 2023). Transformers exploit self-attention to model broader contextual dependencies (Vaswani et al., 2017; Hong et al., 2022; Sun et al., 2022; Mei et al., 2022; Qi et al., 2023). Hybrid architectures such as DCTN and GLMGT further combine local and global representations to exploit their complementary properties (Zhou et al., 2024; Meng et al., 2025).

More recently, state-space models (SSMs) have provided an efficient alternative for long-range sequence modeling, evolving from structured SSMs such as S4 (Gu et al., 2022) to selective Mamba (Gu & Dao, 2024) and its visual adaptations (Zhu et al., 2024; Liu et al., 2024b). Recent HSI classification models, including MambaHSI, 3DSS-Mamba, HyperMamba, and S2Mamba, adapt

such long-range modeling to spectral–spatial representations (Li et al., 2024b; He et al., 2024; Liu et al., 2024a; Wang et al., 2025a), while MambaLG and DualMamba further integrate global and local information (Pan et al., 2025; Sheng et al., 2025). These studies demonstrate the utility of complementary contextual representations, but primarily focus on how such features are extracted and combined rather than explicitly characterizing their relative uncertainty during fusion.

Uncertainty and Reliability Modeling. Uncertainty modeling characterizes ambiguity arising from data and limited model knowledge, with representative approaches including Bayesian approximation (Gal & Ghahramani, 2016), aleatoric–epistemic modeling (Kendall & Gal, 2017), deep ensembles (Lakshminarayanan et al., 2017), and evidential learning (Sensoy et al., 2018). A related line of work studies reliability, including whether predictions or individual information sources can be trusted. Predictive reliability has been examined through confidence calibration (Guo et al., 2017) and uncertainty under dataset shift (Ovadia et al., 2019).

More recently, trusted multi-view methods explicitly consider source reliability. TMC uses view-specific uncertainty to assess trustworthiness and dynamically fuse evidence (Han et al., 2023), while trust-based discounting models instance-wise view reliability to resolve conflicting information (Lu et al., 2025). In remote sensing, uncertainty has also been used to dynamically weight different modalities during fusion (Wang et al., 2025b). For HSI classification, uncertainty has been exploited for out-of-distribution and misclassification detection (Yu et al., 2024), as well as uncertainty-guided sampling and pseudo-label refinement (Yang et al., 2026). In contrast, KGLP-Net propagates uncertainty beyond fusion by deriving group-level reliability from posterior uncertainty through normalized posterior precision and using it for coupled-group selection. Its posterior update is inspired by the Kalman measurement update (Kalman, 1960).

Structured Pruning. Network pruning reduces model redundancy by removing parameters or structural units. While unstructured (connection-level) pruning produces sparse parameterizations (Han et al., 2016), structured pruning removes whole channels or filters to obtain regular compact networks (Li et al., 2017; He et al., 2017; Luo et al., 2017). Representative approaches include channel-sparsity-based Network Slimming, Taylor-based importance estimation, dependency-aware structural pruning, and Group Fisher importance for coupled channels (Liu et al., 2017; Molchanov et al., 2019; Fang et al., 2023; Liu et al., 2021). Prior work has also examined the roles of network architecture and retraining in pruning performance (Liu et al., 2019). In HSI classification, efficiency is additionally pursued through lightweight CNNs, morphology-inspired operators, channel/layer redesign, and groupwise separable Transformer designs (Jia et al., 2021; Zhou et al., 2023; Li et al., 2024a; Zhao et al., 2024). These pruning criteria mainly characterize parameter importance, task sensitivity, or structural dependency and do not explicitly incorporate reliability derived from feature-fusion uncertainty into structure selection.

Taken together, complementary representation learning, uncertainty modeling, and structured pruning have largely been studied separately. KGLP-Net connects them by deriving group-level reliability from the posterior uncertainty of global–local fusion and integrating this reliability with Group Fisher importance for structure selection.

3 METHOD

3.1 FRAMEWORK OVERVIEW

For a target pixel at spatial location p, we construct two center-aligned hyperspectral views with different spatial extents, $\mathbf{X}_p^g \in \mathbb{R}^{B\times s_g\times s_g}$ and $\mathbf{X}_p^l \in \mathbb{R}^{B\times s_l\times s_l}$, where B denotes the number of spectral bands and $s_g > s_l$. The larger view is processed by a Global Mamba Encoder $\mathcal{E}_g$, while the smaller center-aligned view is processed by a Local CNN Encoder $\mathcal{E}_l$:

$$\mathbf{G} = \mathcal{E}_g(\mathbf{X}_p^g), \qquad \mathbf{L} = \mathcal{E}_l(\mathbf{X}_p^l), \tag{1}$$

where both representations have hidden width D.

As illustrated in Figure 2, KGLP-Net comprises asymmetric global–local representation learning, KRF, and KRGF. With the global representation as the prior and the local representation as the observation, KRF performs an uncertainty-guided, Kalman-inspired feature update and estimates

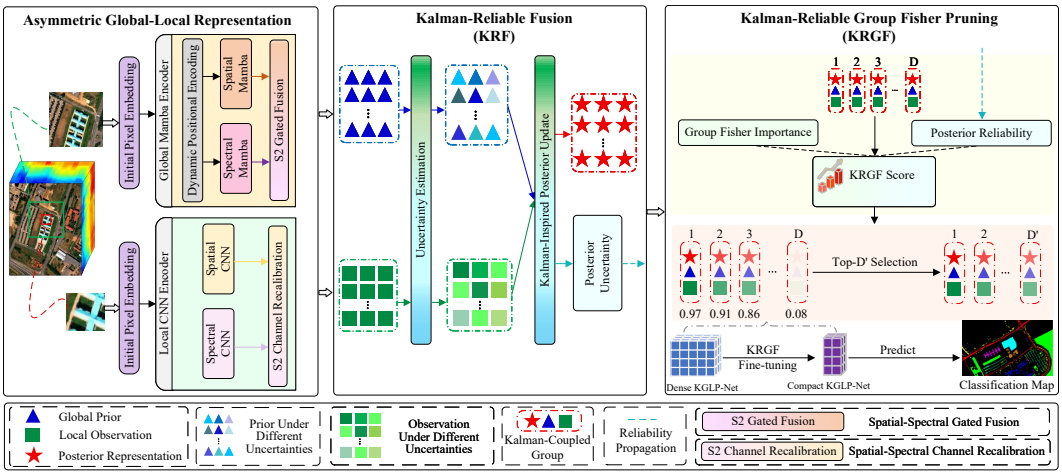


Figure 2: Overview of KGLP-Net, comprising asymmetric global–local representation learning, Kalman-Reliable Fusion (KRF), and Kalman-Reliable Group Fisher Pruning (KRGF).

posterior uncertainty. KRGF derives group-level reliability from this uncertainty and uses it to calibrate Group Fisher selection of Kalman-coupled groups for compact reconstruction. The fused representation is finally mapped to class logits by global average pooling and a linear classifier.

3.2 Asymmetric Global–Local Representation

The two branches learn complementary representations from different spatial scopes and modeling mechanisms. The Global Mamba Encoder operates on the larger view. After feature embedding and dynamic positional encoding, parallel spatial and spectral-channel Mamba pathways model contextual dependencies, and their responses are integrated by gated spatial–spectral fusion. In contrast, the Local CNN Encoder processes the center-aligned local view using parallel spectral and spatial convolutions, followed by pointwise feature integration and channel recalibration. Architectural details of both encoders are provided in Appendix A.

In patch-wise HSI classification, the center pixel is the prediction target, while surrounding pixels provide auxiliary context (Li et al., 2023), motivating distinct estimation roles for the two views. The larger global view aggregates broader context and provides a context-conditioned estimate of the target representation, whereas the center-aligned local view preserves finer spectral–spatial cues around the target. KRF therefore treats the global representation as the contextual prior estimate and the local representation as the observation, and obtains the posterior estimate by correcting the prior toward the observation according to their learned relative uncertainties.

3.3 Kalman-Reliable Fusion

KRF performs a feature-space analogue of the Kalman measurement update (Kalman, 1960). Let $\mathbf{G} \in \mathbb{R}^{D\times H_g\times W_g}$ and $\mathbf{L} \in \mathbb{R}^{D\times H_l\times W_l}$ denote the global and local representations, respectively. When their spatial resolutions differ, $\mathbf{G}$ is bilinearly aligned to the resolution of $\mathbf{L}$. For notational simplicity, the aligned representation is still denoted by $\mathbf{G}$ hereafter, such that $\mathbf{G}, \mathbf{L} \in \mathbb{R}^{D\times H\times W}$.

Interaction-Aware Uncertainty Estimation. The reliability of the two representations cannot be determined from either source independently. KRF therefore constructs a joint interaction representation

$$\mathbf{C} = [\mathbf{L}, \mathbf{G}, \mathbf{L} \odot \mathbf{G}, |\mathbf{L} - \mathbf{G}|] \in \mathbb{R}^{4D\times H\times W}, \tag{2}$$

where $[\cdot]$ denotes channel-wise concatenation and $\odot$ denotes element-wise multiplication. Besides the two original representations, the product term captures their multiplicative interaction and the absolute difference characterizes their discrepancy.

A lightweight uncertainty estimator $\mathcal{U}$ maps $\mathbf{C}$ to two positive element-wise uncertainty estimates:

$$[\mathbf{P}_g, \mathbf{R}_l] = \mathrm{Softplus}\left(\mathrm{Split}(\mathcal{U}(\mathbf{C}))\right) + \delta, \tag{3}$$

where $\delta = 10^{-6}$ ensures positivity and numerical stability. $\mathbf{P}_g, \mathbf{R}_l \in \mathbb{R}^{D\times H\times W}$ denote the uncertainty terms associated with the global prior estimate and local observation, respectively. In implementation, $\mathcal{U}$ consists of a 1×1 projection, a depthwise 3×3 convolution, and a final 1×1 projection, with batch normalization and SiLU activation after the first two operations. The final $2D$ channels are split into $\mathbf{P}_g$ and $\mathbf{R}_l$.

Uncertainty-Guided Posterior Update. Given $\mathbf{P}_g$ and $\mathbf{R}_l$, KRF computes an element-wise Kalman gain

$$\mathbf{K} = \frac{\mathbf{P}_g}{\mathbf{P}_g + \mathbf{R}_l + \delta}, \tag{4}$$

and obtains the posterior feature estimate as

$$\bar{\mathbf{F}} = \mathbf{G} + \mathbf{K} \odot (\mathbf{L} - \mathbf{G}). \tag{5}$$

Here, a larger $\mathbf{K}$ favors the local observation, whereas a smaller $\mathbf{K}$ retains more of the global prior, with the balance determined by their relative uncertainties.

To complement the uncertainty-guided posterior update with nonlinear global–local interactions, KRF introduces a lightweight residual path:

$$\mathbf{F} = \Phi_o\left(\bar{\mathbf{F}} + \mathcal{D}\left(\boldsymbol{\alpha} \odot \Phi_i(\mathbf{C})\right)\right), \tag{6}$$

where Φ_i reduces the $4D$-channel interaction representation to D channels using a 1×1 projection followed by a depthwise 3×3 convolution, while Φ_o is a 1×1 output projection. Batch normalization and SiLU activation are used within these mappings. $\boldsymbol{\alpha} \in \mathbb{R}^{D\times1\times1}$ is a learnable channel-wise scale, and $\mathcal{D}$ denotes dropout during training. The resulting $\mathbf{F}$ serves as the fused representation for classification.

Posterior Uncertainty. Using the same uncertainty terms, KRF computes the element-wise posterior uncertainty estimate associated with $\bar{\mathbf{F}}$:

$$\mathbf{P}^+ = \frac{\mathbf{P}_g \odot \mathbf{R}_l}{\mathbf{P}_g + \mathbf{R}_l + \delta}. \tag{7}$$

While $\mathbf{F}$ is used for classification, $\mathbf{P}^+$ provides the posterior uncertainty associated with $\bar{\mathbf{F}}$, which KRGF subsequently aggregates into group-level reliability.

3.4 KALMAN-RELIABLE GROUP FISHER PRUNING

KRF produces both a fused representation and posterior uncertainty, while the dense network can still contain redundant hidden structures. KRGF propagates the uncertainty estimated during fusion into structured compression by combining posterior reliability with Group Fisher task sensitivity.

Kalman-Coupled Group. The same channel index is maintained across the global, local, and fused representations in KRF. We therefore define the c-th Kalman-coupled group as

$$\mathcal{S}_c = \{\mathbf{G}_c, \mathbf{L}_c, \mathbf{F}_c\}, \qquad c = 1, \ldots, D. \tag{8}$$

During group evaluation, a shared virtual gate m_c is applied to the c-th global and local channels before KRF and to the corresponding fused channel before classification. Consequently, the sensitivity of m_c reflects the joint task contribution of the complete coupled hidden path rather than that of an isolated feature channel.

Group Fisher Importance. Following the shared-mask formulation of Group Fisher (Liu et al., 2021), we estimate the task sensitivity of each coupled group using the squared gradient of the classification objective with respect to its shared virtual gate. Let M denote the number of calibration mini-batches. The importance of $\mathcal{S}_c$ is

$$\mathcal{I}_c = \frac{1}{M}\sum_{b=1}^{M}\left(\frac{\partial \mathcal{L}_{\text{cls}}^{(b)}}{\partial m_c}\right)^2, \tag{9}$$

where $\mathcal{L}_{\text{cls}}^{(b)}$ is the supervised classification loss on the b-th calibration mini-batch. Network parameters are fixed during this evaluation and only the virtual gate is used to measure sensitivity. A larger

$\mathcal{I}_c$ indicates a greater influence of the corresponding coupled path on the classification objective. Since all candidate groups have the same structural width, they are ranked directly by their Fisher sensitivity.

Posterior Reliability. Task sensitivity alone does not describe how reliably the corresponding posterior information is estimated. KRGF therefore complements $\mathcal{I}_c$ with a reliability score derived from the posterior uncertainty of KRF. For channel c, posterior uncertainty is first averaged over samples and spatial positions within each calibration mini-batch and then over all M mini-batches:

$$\bar{P}_c^+ = \frac{1}{M}\sum_{b=1}^{M} \mathrm{Mean}_{n,h,w}\left[P_{b,n,c,h,w}^+\right], \tag{10}$$

where n indexes samples within mini-batch b. Its reciprocal defines the posterior precision

$$\pi_c = \frac{1}{\bar{P}_c^+ + \delta}. \tag{11}$$

To obtain a relative reliability measure across coupled groups, the precision is normalized by its mean over the D hidden groups:

$$\rho_c = \frac{\pi_c}{\frac{1}{D}\sum_{j=1}^{D}\pi_j + \varepsilon}, \tag{12}$$

where $\varepsilon = 10^{-12}$ ensures numerical stability. We use ρ_c as the relative posterior-reliability score of $\mathcal{S}_c$. A larger ρ_c indicates higher posterior precision relative to the other coupled groups.

Reliability-Calibrated Group Selection. Group Fisher importance and posterior reliability characterize different properties of the same structural group. Group Fisher importance $\mathcal{I}_c$ measures its sensitivity to the classification objective, whereas posterior reliability ρ_c reflects the normalized precision of the posterior information produced by the prior–observation update. KRGF combines them multiplicatively:

$$Q_c = \mathcal{I}_c \rho_c, \tag{13}$$

so that posterior reliability reweights the task-related importance of each coupled group. Given a target hidden width D', KRGF retains the Kalman-coupled groups with the largest scores:

$$\Omega = \mathrm{TopK}\left(\{Q_c\}_{c=1}^{D}, D'\right). \tag{14}$$

The selected groups define the compact model, with posterior reliability ρ_c guiding coupled-group selection and Group Fisher importance preserving task relevance.

3.5 Training and Compact Reconstruction

KGLP-Net follows a three-stage training and compression procedure comprising dense-model training, progressive KRGF compression, and compact-model fine-tuning. After Dense64 is trained end-to-end, the hidden width is progressively reduced through 48, 40, and 32, with successive reductions of 16, 8, and 8 channels. Compact48, Compact40, and Compact32 therefore follow $64 \to 48$, $64 \to 48 \to 40$, and $64 \to 48 \to 40 \to 32$, respectively. At each stage, KRGF scores are recomputed on the current model and the D' highest-scoring groups are retained to reconstruct the narrower model. Its BatchNorm statistics are recalibrated with one complete pass over the training loader before the next compression stage. After reaching the target width, the compact model is fine-tuned using the supervised classification objective.

4 Experiments

4.1 Experimental Setup

Datasets and Evaluation Metrics. We select four public datasets for the experiments, namely Pavia University (UP), Houston 2013 (Houston), Salinas, and WHU-Hi-HanChuan (HanChuan) (Zhong et al., 2020). These datasets cover urban, agricultural, and UAV-based remote-sensing scenes. The class-wise sample counts and complete train/validation/test splits are provided in Appendix B. Classification performance is evaluated using standard HSI metrics, including overall accuracy

Table 1: OA (%), AA (%), and $\kappa \times 100$ statistics on four HSI datasets.The best and second-best results are shown in **bold** and underline, respectively.

		CNN-Based			Transformer-Based			SSM-Based				
Dataset	Metric	CVSSN	SSRN	HKNAS3D	SSFTT	GAHT	CMT	MambaHSI	3DSS-Mamba	MambaLG	SliMamba	Compact48 (Proposed)
UP	OA	98.44±0.23	98.86±0.16	95.67±0.88	96.06±1.07	97.54±0.44	83.37±4.94	96.46±0.46	89.59±1.13	97.46±0.31	97.27±0.39	**99.45±0.12**
	AA	98.02±0.45	98.36±0.28	94.89±0.83	95.90±0.71	96.49±0.70	84.53±3.60	96.20±0.87	83.66±1.96	95.31±0.81	95.66±0.60	**99.28±0.12**
	$\kappa \times 100$	97.93±0.31	98.49±0.21	94.29±1.14	94.81±1.38	96.74±0.58	78.53±5.87	96.38±2.05	86.16±1.50	96.62±0.41	96.39±0.51	**99.28±0.16**
Houston	OA	97.14±0.40	97.68±0.41	94.13±0.77	94.97±0.81	96.31±0.60	88.58±2.10	93.97±1.08	92.36±1.52	96.99±0.53	97.34±0.42	**98.03±0.21**
	AA	97.53±0.41	97.69±0.53	94.97±0.66	95.67±0.75	96.40±0.61	90.44±1.65	94.94±0.93	92.79±1.48	97.05±0.42	97.43±0.52	**98.14±0.42**
	$\kappa \times 100$	96.91±0.43	97.49±0.44	93.66±0.83	94.56±0.87	96.01±0.65	87.66±2.27	94.00±2.31	91.74±1.65	96.75±0.57	97.13±0.46	**97.87±0.23**
Salinas	OA	95.99±0.37	97.21±0.45	93.44±0.57	93.15±1.12	95.81±0.31	93.05±1.48	93.04±2.80	94.51±1.03	97.61±0.60	97.81±0.55	**98.45±0.83**
	AA	98.10±0.29	98.29±0.32	96.66±0.34	96.43±0.63	97.38±0.44	95.65±1.06	94.14±2.87	94.05±1.22	97.00±0.74	98.73±0.31	**99.18±0.47**
	$\kappa \times 100$	95.55±0.42	96.90±0.49	92.71±0.63	92.39±1.25	95.34±0.34	92.27±1.64	92.67±3.53	93.88±1.15	97.34±0.66	97.56±0.62	**98.27±0.92**
HanChuan	OA	96.68±0.53	94.76±0.45	91.29±0.79	92.67±0.41	95.31±0.40	88.79±1.83	89.85±1.08	89.15±0.68	97.13±0.77	94.54±0.38	**98.94±0.14**
	AA	94.10±1.14	88.30±1.03	89.92±0.66	87.43±0.96	90.75±0.83	82.87±2.61	88.13±1.16	77.54±1.30	88.45±3.98	90.75±0.65	**98.05±0.42**
	$\kappa \times 100$	96.11±0.62	93.87±0.53	89.87±0.90	91.43±0.47	94.51±0.46	86.92±2.12	86.67±3.17	87.30±0.79	96.64±0.90	93.61±0.44	**98.76±0.16**

(OA), average accuracy (AA), and Cohen's Kappa coefficient (κ) (Cohen, 1960). All experiments are independently repeated 10 times to reduce sampling bias. The main paper reports the overall classification results, while complete class-wise accuracy results are provided in Appendix E.

Implementation Details. We compare KGLP-Net with ten representative HSI classification methods, including 1) CNN-based methods: SSRN (Zhong et al., 2018), CVSSN (Li et al., 2023), and HKNAS (Wang et al., 2024); 2) Transformer-based methods: SSFTT (Sun et al., 2022), GAHT (Mei et al., 2022), and CMT (Jia et al., 2024); 3) SSM-based methods: MambaHSI (Li et al., 2024b), 3DSS-Mamba (He et al., 2024), MambaLG (Pan et al., 2025), and SliMamba (Zhao et al., 2026). For Dense64, the hidden width and Mamba state dimension are set to 64 and 16. The global/local view sizes are set to 21/7 for UP and Houston and 25/9 for Salinas and HanChuan based on validation, with sensitivity analysis provided in Appendix C. For Dense64 training, the initial learning rate, weight decay, batch size, and epochs are set to 0.0005, 0.05, 64, and 180. AdamW (Loshchilov & Hutter, 2019) and cosine annealing (Loshchilov & Hutter, 2017) are used for parameter optimization and learning-rate scheduling, respectively, with Poly-1 loss (Leng et al., 2022). In addition, the reconstructed compact model is fine-tuned for up to 80 epochs at 0.0002 using the same optimizer and scheduler, and all experiments are implemented in PyTorch.

4.2 Comparison with Representative Methods

Quantitative Results. Table 1 summarizes OA, AA, and κ for all methods on four datasets. Compact48 (hidden width $= 48$) outperforms all competing methods and achieves the highest OA, AA, and κ on every dataset, demonstrating stable classification performance across scenes with different spatial structures and land-cover compositions. The class-wise results in Appendix E further show that this advantage is distributed across multiple categories rather than being dominated by a small subset of classes.

Accuracy–Compactness Trade-off. Figure 1 compares the competing methods in terms of classification performance, parameter count, and computational complexity. Compact48 consistently occupies a favorable high-accuracy, low-parameter region on all four datasets and also exhibits relatively low computational complexity. Although 3DSS-Mamba and SliMamba have smaller model sizes, their classification accuracy degrades noticeably on some datasets, while several substantially larger models do not achieve higher accuracy. These results indicate that Compact48 achieves a favorable balance between discriminative performance and model size. The specific effect of KRGF on this accuracy–compactness relationship is further analyzed in Sec. 4.4.

Qualitative Results. For each dataset, we visualize the strongest CNN, Transformer, and SSM baselines using the run closest to their mean OA. Complete comparisons are provided in Appendix F. Figure 3 shows that Compact48 better matches the ground truth in localized regions. On UP, it yields fewer isolated misclassifications around homogeneous and narrow structures. On Salinas, it reduces scattered errors within large agricultural fields while preserving field boundaries. On Houston, the marked regions show more accurate class assignment and spatial extent for small isolated regions. On HanChuan, Compact48 better preserves large homogeneous regions, the elongated central structure, and dense narrow regions with fewer local errors.

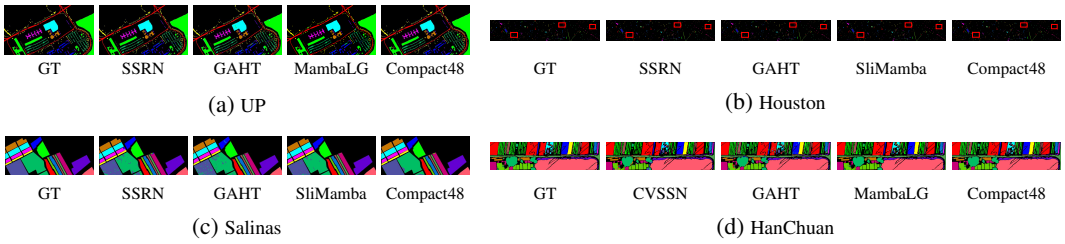


Figure 3: Qualitative classification maps on four HSI datasets. Within each subfigure, the panels from left to right show the ground truth, the selected CNN-based baseline, Transformer-based baseline, SSM-based baseline, and Compact48.

Table 2: Ablation of asymmetric dual-view representations on four HSI datasets in terms of OA (%). The best results are shown in **bold**.

Local View	Global View	UP	Houston	Salinas	HanChuan
✓	✗	96.71±0.85	96.25±0.37	94.84±1.11	96.82±0.31
✗	✓	95.14±1.01	94.21±0.77	96.23±0.82	97.92±0.19
✓	✓	**99.51±0.12**	**98.04±0.31**	**98.44±0.90**	**99.00±0.14**

Table 3: Ablation of Kalman-Reliable Fusion on four HSI datasets in terms of OA (%). The best results are shown in **bold**.

	Fusion Strategy			
Dataset	Sum	Concat	Gate	**KRF**
UP	99.43±0.21	99.33±0.45	99.35±0.30	**99.51±0.12**
Houston	97.59±0.59	97.64±0.54	97.82±0.44	**98.04±0.31**
Salinas	97.85±1.12	96.95±2.58	97.94±1.12	**98.44±0.90**
HanChuan	98.72±0.17	98.82±0.17	98.94±0.11	**99.00±0.14**

Table 4: Ablation of the KRGF assessment criteria under the common $64 \rightarrow 32$ compression setting in terms of OA (%). The best results are shown in **bold**.

	Assessment Criteria		
Dataset	Group Fisher Importance ($\mathcal{I}_c$)	Posterior Reliability (ρ_c)	**KRGF** (Q_c)
UP	98.62±0.71	97.36±0.99	**98.98±0.32**
Houston	97.04±0.56	96.66±0.61	**97.33±0.32**
Salinas	95.41±2.47	92.18±4.98	**96.89±0.86**
HanChuan	98.02±0.31	98.22±0.28	**98.62±0.20**

4.3 Ablation and Reliability Analysis

Global–Local Complementarity. To analyze the role of asymmetric dual-view representation, we construct three variants: Local-only, Global-only, and Dual-view. As shown in Table 2, Local-only performs better on UP and Houston, whereas Global-only is stronger on Salinas and HanChuan. Dual-view achieves the best performance on all four datasets. These results show that broad regional context and fine-grained information preserved around the center neighborhood provide stable complementary cues, while their relative importance varies with scene structure.

Kalman-Reliable Fusion. Building on the dual-view representation, we compare KRF with Sum, Concat, and Gate fusion. As shown in Table 3, KRF achieves the highest OA on all four datasets, with gains of 0.08–0.50 percentage points over the strongest alternative. By using learned uncertainty estimates to determine the relative contributions of the global prior and local observation to the posterior update, KRF consistently improves over the comparison fusion strategies while providing posterior uncertainty for subsequent reliability estimation in KRGF.

Group Fisher Importance vs. Posterior Reliability. Under the common $64 \rightarrow 32$ strong-compression setting, we compare Group Fisher importance ($\mathcal{I}_c$), posterior reliability (ρ_c), and KRGF (Q_c). As shown in Table 4, KRGF outperforms either individual criterion on all four datasets. Posterior reliability alone is generally weaker than Group Fisher importance, while their combination consistently improves classification performance, confirming their complementarity.

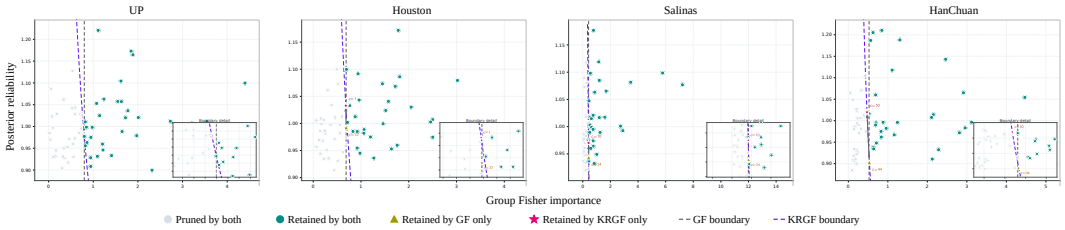


Figure 4: Coupled-group selections produced by Group Fisher and KRGF under $64 \rightarrow 32$ compression for representative runs whose KRGF OA is closest to the corresponding ten-run mean. Insets show pruning-boundary regions, and numerical labels denote the group index c.

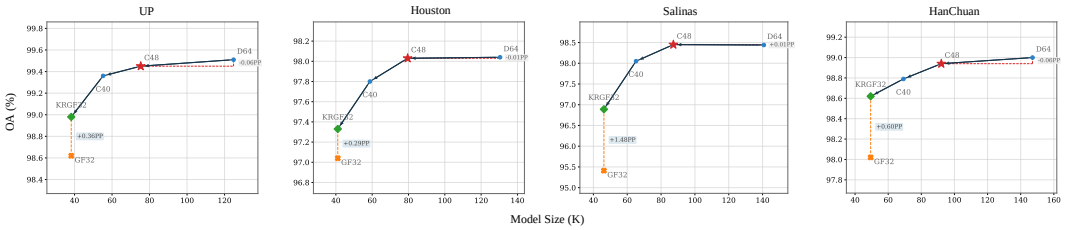


Figure 5: Accuracy–complexity trajectories under progressive KRGF compression. GF32 uses $\mathcal{I}_c$ alone at width 32.

4.4 KRGF Mechanism and Compression Efficiency Analysis

Posterior Reliability Calibrates Group Fisher Selection. Figure 4 visualizes the coupled-group selections produced by Group Fisher and KRGF under $64 \rightarrow 32$ compression for representative runs. Their differences occur mainly near the pruning boundary, where groups have similar I_c. Since KRGF ranks groups by $Q_c = I_c \rho_c$, posterior reliability ρ_c reweights Fisher importance and can alter the priority of such near-boundary groups. These observations suggest that posterior reliability has a greater influence on the ranking of coupled groups when Fisher importance is less discriminative near the pruning boundary, whereas groups with clearly high or low Fisher importance tend to retain their original selection status. For example, on Houston, $c = 22$ is retained only by Group Fisher, whereas $c = 1$ is retained only by KRGF. Similar boundary-level changes are observed on Salinas and HanChuan, while both criteria yield the same Top-32 set in the representative UP run. Overall, posterior reliability primarily refines Fisher-based ranking near the pruning boundary rather than replacing Fisher importance as the main task-related criterion.

Compression and Efficiency. Figure 5 shows the accuracy–complexity trajectory from Dense64 to Compact48, Compact40, and Compact32, with complete results in Appendix D. Compared with Dense64, Compact48 reduces parameters, FLOPs, and inference time by 38.5%, 34.8%, and 37.6% on average, respectively, while reducing mean OA by only 0.03 percentage points. We therefore use Compact48 as the accuracy–efficiency operating point, while stronger compression to widths 40 and 32 incurs larger accuracy drops. Under the common $64 \rightarrow 32$ setting, KRGF32 consistently outperforms the GF32 baseline using Fisher importance alone, supporting reliability-calibrated Group Fisher selection under strong compression.

5 Conclusion

This work presents **KGLP-Net**, which couples uncertainty-aware global–local fusion with reliability-guided structured pruning for efficient HSI classification. Experiments on four public datasets demonstrate a favorable accuracy–efficiency trade-off, while pruning ablations support posterior reliability as a complementary signal for structure selection. More broadly, linking representation uncertainty directly to structured pruning suggests a general design principle for lightweight neural networks, where uncertainty-derived reliability can guide structural retention. Future work will explore cross-channel uncertainty modeling, adaptive spatial scales, and the extension of uncertainty-guided pruning to other architectures and tasks.

AI USE STATEMENT

Generative AI tools, including large language models (LLMs), were used solely to assist with the writing and presentation of this manuscript. Specifically, they were used for language polishing, sentence rephrasing, grammar and readability improvements, LaTeX formatting, and checks for consistency with the ICLR submission template.

Generative AI tools were not used for research ideation, the development of the KGLP-Net methodology, mathematical formulation, experimental design, method implementation, dataset processing, or the interpretation of experimental results. All research concepts, methodological designs, experiments, analyses, and scientific conclusions were developed and conducted by the authors.

All AI-assisted edits were carefully reviewed by the authors to ensure technical accuracy and consistency with the proposed methodology, experimental results, source materials, and cited literature. The authors take full responsibility for the final content of this work, including all text, claims, and artifacts produced with the assistance of generative AI.

REPRODUCIBILITY STATEMENT

We have made every effort to facilitate the reproducibility of KGLP-Net. The proposed global–local representation, Kalman-Reliable Fusion (KRF), and Kalman-Reliable Group Fisher Pruning (KRGF) are described in detail in the main paper, with additional architectural specifications of the global and local encoders provided in Appendix A. The complete experimental protocol, evaluation metrics, training configurations, and model settings are reported in Sec. 4.1.

All experiments are conducted on four publicly available hyperspectral image classification datasets: Pavia University, Houston 2013, Salinas, and WHU-Hi-HanChuan. Detailed dataset descriptions, class-wise train/validation/test splits, and the resampling protocol used for the ten independent runs are provided in Appendix B. Additional experiments on patch-size sensitivity, compression and efficiency, class-wise classification performance, and full qualitative classification maps are reported in Appendices C–F, providing further information for verifying the reported results.

An anonymized implementation of KGLP-Net, together with the complete experimental configurations and instructions for reproducing the main experiments, is provided in the supplementary material to facilitate independent reproduction and verification of the reported results.

REFERENCES

Amina Ben Hamida, Alexandre Benoit, Patrick Lambert, and Chokri Ben Amar. 3-D deep learning approach for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8):4420–4434, 2018. doi: 10.1109/TGRS.2018.2818945.

Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2094–2107, 2014. doi: 10.1109/JSTARS.2014.2329330.

Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960. doi: 10.1177/001316446002000104.

Gongfan Fang, Xinyin Ma, Mingli Song, Michael Bi Mi, and Xinchao Wang. DepGraph: Towards any structural pruning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16091–16101, 2023. doi: 10.1109/CVPR52729.2023.01544.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 1050–1059. PMLR, 2016.

Pedram Ghamisi, Emmanuel Maggiori, Shutao Li, Roberto Souza, Yuliya Tarabalka, Gabriele Moser, Andrea De Giorgi, Leyuan Fang, Yushi Chen, Mingmin Chi, Sebastiano B. Serpico, and Jon Atli Benediktsson. New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, markov random fields, segmentation, sparse representation, and deep learning. *IEEE Geoscience and Remote Sensing Magazine*, 6(3):10–43, 2018. doi: 10.1109/MGRS.2018.2854840.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling*, 2024.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations*, 2022.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1321–1330. PMLR, 2017.

Song Han, Huizi Mao, and William J. Dally. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. In *International Conference on Learning Representations*, 2016.

Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification with dynamic evidential fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):2551–2566, 2023. doi: 10.1109/TPAMI.2022.3171983.

Lin He, Jun Li, Chenying Liu, and Shutao Li. Recent advances on spectral–spatial hyperspectral image classification: An overview and new guidelines. *IEEE Transactions on Geoscience and Remote Sensing*, 56(3):1579–1597, 2018. doi: 10.1109/TGRS.2017.2765364.

Yan He, Bing Tu, Bo Liu, Jun Li, and Antonio Plaza. 3DSS-Mamba: 3D-spectral–spatial Mamba for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–16, 2024. doi: 10.1109/TGRS.2024.3472091. Article 5534216.

Yihui He, Xiangyu Zhang, and Jian Sun. Channel pruning for accelerating very deep neural networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1398–1406, 2017. doi: 10.1109/ICCV.2017.155.

Danfeng Hong, Zhu Han, Jing Yao, Lianru Gao, Bing Zhang, Antonio Plaza, and Jocelyn Chanussot. SpectralFormer: Rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022. doi: 10.1109/TGRS.2021.3130716. Article 5518615.

Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7132–7141, 2018. doi: 10.1109/CVPR.2018.00745.

Sen Jia, Zhijie Lin, Meng Xu, Qiang Huang, Jun Zhou, Xiuping Jia, and Qingquan Li. A lightweight convolutional neural network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(5):4150–4163, 2021. doi: 10.1109/TGRS.2020.3014313.

Sen Jia, Yifan Wang, Shuguo Jiang, and Ruyan He. A center-masked transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–16, 2024. doi: 10.1109/TGRS.2024.3369075. Article 5510416.

Rudolf E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960. doi: 10.1115/1.3662552.

Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems*, volume 30, pp. 5574–5584, 2017.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*, volume 30, pp. 6402–6413, 2017.

Zhaoqi Leng, Mingxing Tan, Chenxi Liu, Ekin Dogus Cubuk, Jay Shi, Shuyang Cheng, and Dragomir Anguelov. PolyLoss: A polynomial expansion perspective of classification loss functions. In *International Conference on Learning Representations*, 2022.

Chunchao Li, Behnood Rasti, Xuebin Tang, Puhong Duan, Jun Li, and Yuanxi Peng. Channel-layer-oriented lightweight spectral–spatial network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024a. doi: 10.1109/TGRS.2024.3350055. Article 5504214.

Hao Li, Asim Kadav, Igor Durdanovic, Hanan Samet, and Hans Peter Graf. Pruning filters for efficient ConvNets. In *International Conference on Learning Representations*, 2017.

Mingsong Li, Yikun Liu, Guangkuo Xue, Yuwen Huang, and Gongping Yang. Exploring the relationship between center and neighborhoods: Central vector oriented self-similarity network for hyperspectral image classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(4):1979–1993, 2023. doi: 10.1109/TCSVT.2022.3218284.

Shutao Li, Weiwei Song, Leyuan Fang, Yushi Chen, Pedram Ghamisi, and Jon Atli Benediktsson. Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6690–6709, 2019. doi: 10.1109/TGRS.2019.2907932.

Yapeng Li, Yong Luo, Lefei Zhang, Zengmao Wang, and Bo Du. MambaHSI: Spatial–spectral Mamba for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–16, 2024b. doi: 10.1109/TGRS.2024.3430985. Article 5524216.

Liyang Liu, Shilong Zhang, Zhanghui Kuang, Aojun Zhou, Jing-Hao Xue, Xinjiang Wang, Yimin Chen, Wenming Yang, Qingmin Liao, and Wayne Zhang. Group fisher pruning for practical network compression. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 7021–7032. PMLR, 2021.

Qiang Liu, Jun Yue, Yi Fang, Shaobo Xia, and Leyuan Fang. HyperMamba: A spectral–spatial adaptive Mamba for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024a. doi: 10.1109/TGRS.2024.3482473. Article 5536514.

Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. VMamba: Visual state space model. In *Advances in Neural Information Processing Systems*, volume 37, pp. 103031–103063, 2024b. doi: 10.52202/079017-3273.

Zhuang Liu, Jianguo Li, Zhiqiang Shen, Gao Huang, Shoumeng Yan, and Changshui Zhang. Learning efficient convolutional networks through network slimming. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2755–2763, 2017. doi: 10.1109/ICCV.2017.298.

Zhuang Liu, Mingjie Sun, Tinghui Zhou, Gao Huang, and Trevor Darrell. Rethinking the value of network pruning. In *International Conference on Learning Representations*, 2019.

Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. In *International Conference on Learning Representations*, 2017.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Jueqing Lu, Wray Buntine, Yuanyuan Qi, Joanna Dipnall, Belinda Gabbe, and Lan Du. Navigating conflicting views: Harnessing trust for learning. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 40411–40435. PMLR, 2025.

Jian-Hao Luo, Jianxin Wu, and Weiyao Lin. ThiNet: A filter level pruning method for deep neural network compression. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 5068–5076, 2017. doi: 10.1109/ICCV.2017.541.

Shaohui Mei, Chao Song, Mingyang Ma, and Fulin Xu. Hyperspectral image classification using group-aware hierarchical transformer. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–14, 2022. doi: 10.1109/TGRS.2022.3207933. Article 5539014.

Zhe Meng, Qian Yan, Feng Zhao, Gaige Chen, Wenqiang Hua, and Miaomiao Liang. Global–local multigranularity transformer for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18:112–131, 2025. doi: 10.1109/JSTARS.2024.3491294.

Pavlo Molchanov, Arun Mallya, Stephen Tyree, Iuri Frosio, and Jan Kautz. Importance estimation for neural network pruning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11264–11272, 2019. doi: 10.1109/CVPR.2019.01152.

Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, D. Sculley, Sebastian Nowozin, Joshua V. Dillon, Balaji Lakshminarayanan, and Jasper Snoek. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. In *Advances in Neural Information Processing Systems*, volume 32, pp. 13991–14002, 2019.

Zhaojie Pan, Chenyu Li, Antonio Plaza, Jocelyn Chanussot, and Danfeng Hong. Hyperspectral image classification with Mamba. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–14, 2025. doi: 10.1109/TGRS.2024.3521411. Article 5602814.

Wenchao Qi, Changping Huang, Yibo Wang, Xia Zhang, Weiwei Sun, and Lifu Zhang. Global–local 3-D convolutional transformer network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–20, 2023. doi: 10.1109/TGRS.2023.3272885. Article 5510820.

Swalpa Kumar Roy, Gopal Krishna, Shiv Ram Dubey, and Bidyut B. Chaudhuri. HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 17(2):277–281, 2020. doi: 10.1109/LGRS.2019.2918719.

Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. In *Advances in Neural Information Processing Systems*, volume 31, pp. 3179–3189, 2018.

Jiamu Sheng, Jingyi Zhou, Jiong Wang, Peng Ye, and Jiayuan Fan. DualMamba: A lightweight spectral–spatial mamba-convolution network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–15, 2025. doi: 10.1109/TGRS.2024.3516817. Article 5501415.

Le Sun, Guangrui Zhao, Yuhui Zheng, and Zebin Wu. Spectral–spatial feature tokenization transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022. doi: 10.1109/TGRS.2022.3144158. Article 5522214.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5998–6008, 2017.

Di Wang, Bo Du, Liangpei Zhang, and Dacheng Tao. HKNAS: Classification of hyperspectral imagery based on hyper kernel neural architecture search. *IEEE Transactions on Neural Networks and Learning Systems*, 35(10):13631–13645, 2024. doi: 10.1109/TNNLS.2023.3270369.

Guanchun Wang, Xiangrong Zhang, Zelin Peng, Tianyang Zhang, and Licheng Jiao. S2Mamba: A spatial–spectral state space model for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–13, 2025a. doi: 10.1109/TGRS.2025.3530993. Article 5511413.

Hui Wang, Youxiang Huang, Hao Huang, Yu Wang, Jun Li, and Guan Gui. Uncertainty-aware dynamic fusion network with criss-cross attention for multimodal remote sensing land cover classification. *Information Fusion*, 123:103249, 2025b. doi: 10.1016/j.inffus.2025.103249.

Muzhou Yang, Wuzhou Quan, and Mingqiang Wei. Perceive, act and correct: Confidence is not enough for hyperspectral classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 11631–11639, 2026. doi: 10.1609/aaai.v40i14.38147.

Linlin Yu, Yifei Lou, and Feng Chen. Uncertainty-aware graph-based hyperspectral image classification. In *International Conference on Learning Representations*, 2024.

Zhuoyi Zhao, Xiang Xu, Shutao Li, and Antonio Plaza. Hyperspectral image classification using groupwise separable convolutional vision transformer network. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–17, 2024. doi: 10.1109/TGRS.2024.3377610. Article 5511817.

Zhuoyi Zhao, Xiang Xu, Chuiyi Deng, Junwei Li, and Antonio Plaza. Spectral selection convolution and overlapped-centering mamba network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–17, 2026. doi: 10.1109/TGRS.2026.3656272. Article 5502417.

Yanfei Zhong, Xin Hu, Chang Luo, Xinyu Wang, Ji Zhao, and Liangpei Zhang. WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H^2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. *Remote Sensing of Environment*, 250:112012, 2020. doi: 10.1016/j.rse.2020.112012.

Zilong Zhong, Jonathan Li, Zhiming Luo, and Michael Chapman. Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2):847–858, 2018. doi: 10.1109/TGRS.2017.2755542.

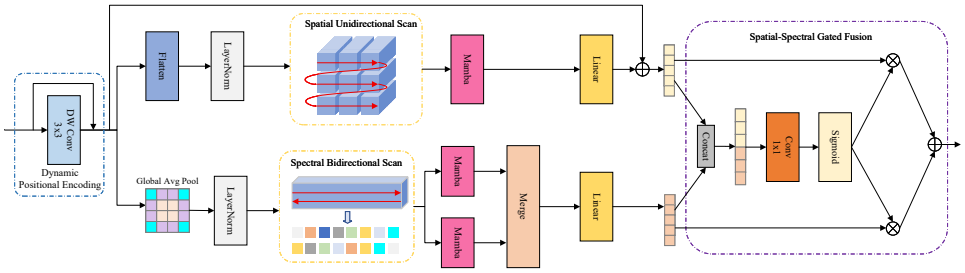
Fei Zhou, Xin Sun, Chengze Sun, Junyu Dong, and Xiao Xiang Zhu. Adaptive morphology filter: A lightweight module for deep hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–16, 2023. doi: 10.1109/TGRS.2023.3327418. Article 5529316.

Yunfei Zhou, Xiaohui Huang, Xiaofei Yang, Jiangtao Peng, and Yifang Ban. DCTN: Dual-branch convolutional transformer network with efficient interactive self-attention for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–16, 2024. doi: 10.1109/TGRS.2024.3364143. Article 5508616.

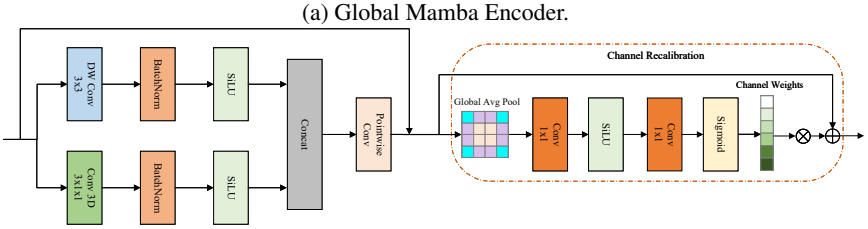
Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision Mamba: Efficient visual representation learning with bidirectional state space model. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 62429–62442. PMLR, 2024.

Appendix

A Encoder Architecture Details



(a) Global Mamba Encoder.



(b) Local CNN Encoder.

Figure 6: Detailed architectures of the global and local encoders.

Global Mamba Encoder. After embedding and dynamic positional encoding, the global feature is processed by parallel spatial and spectral Mamba pathways. The spatial pathway flattens the $H \times W$ feature grid in row-major order and performs unidirectional Mamba modeling to capture spatial context. The spectral pathway applies global average pooling followed by bidirectional Mamba modeling to capture complementary spectral information. Their responses are combined by spatial–spectral gated fusion, where a learned sigmoid gate adaptively balances the two pathways before residual addition.

Local CNN Encoder. The local encoder extracts complementary spectral and spatial responses using a spectral $3 \times 1 \times 1$ 3-D convolution and a depthwise spatial 3×3 convolution, respectively. The resulting responses are concatenated along the channel dimension and integrated through a pointwise convolution. Channel recalibration then generates channel-wise weights from the integrated representation through global average pooling, two pointwise projections with an intermediate SiLU activation, and a sigmoid gate. The resulting weights adaptively recalibrate the integrated feature before residual addition, following the general principle of channel-wise feature recalibration (Hu et al., 2018).

B Dataset Details

We summarize the acquisition information, spatial–spectral characteristics, and class composition of the four HSI datasets used in our experiments.

1. **Pavia University (UP).** The UP scene was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) during a flight campaign over Pavia, northern Italy. The original image contains 610×610 pixels and 115 spectral bands covering 430–860 nm. After removing the uninformative spatial region and 12 noisy bands, the commonly used version contains 610×340 pixels and 103 spectral bands with a spatial resolution of 1.3 m. The ground truth comprises nine land-cover classes and 42,776 labeled pixels.
2. **Houston 2013 (Houston).** The Houston 2013 hyperspectral scene was acquired in June 2012 using the ITRES CASI-1500 sensor over the University of Houston campus and its surrounding urban area in Houston, Texas, USA, and was released through the 2013 IEEE GRSS Data Fusion Contest. The hyperspectral image has a spatial size of 349×1905

pixels and consists of 144 spectral bands spanning 380–1050 nm, with a ground sampling distance of 2.5 m. The scene represents a heterogeneous urban environment containing vegetation, buildings, transportation infrastructure, and other man-made surfaces. Its ground truth comprises 15 land-cover classes with a total of 15,029 labeled pixels.

3. **Salinas.** The Salinas scene was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over Salinas Valley, California. The original image contains 512×217 pixels, 224 spectral bands covering approximately 400–2500 nm, and has a spatial resolution of approximately 3.7 m. The ground truth comprises 16 land-cover classes with 54,129 labeled pixels.
4. **WHU-Hi-HanChuan (HanChuan).** The HanChuan scene was acquired on June 17, 2016, in Hanchuan, Hubei Province, China, using a Headwall Nano-Hyperspec imaging sensor with a 17-mm focal length mounted on a Leica Aibot X6 UAV V1 platform (Zhong et al., 2020). The study area is a rural–urban fringe containing buildings, water bodies, and cultivated land. The image contains 1217×303 pixels and 274 spectral bands spanning 400–1000 nm, with a spatial resolution of approximately 0.109 m. Its ground truth comprises 16 land-cover classes with 257,530 labeled pixels, including seven crop categories: strawberry, cowpea, soybean, sorghum, water spinach, watermelon, and greens.

Split protocol. We follow the sampling protocol described in Sec. 4.1. For UP, Salinas, and HanChuan, 1% of the labeled samples are used for training, whereas Houston uses 5% for training. For all four datasets, 5% of the labeled samples are used for validation and the remaining labeled samples are reserved for testing. Background pixels are excluded throughout. Tables 5–8 report the exact class-wise split used in our experiments. Across the 10 independent runs, sample identities are resampled while the class-wise sample counts remain fixed.

Table 5: Class-wise train/validation/test split for UP.

Class	Color	Train	Validation	Test	Total
1 Asphalt		66	331	6234	6631
2 Meadows		186	932	17531	18649
3 Gravel		20	104	1975	2099
4 Trees		30	153	2881	3064
5 Painted metal sheets		13	67	1265	1345
6 Bare Soil		50	251	4728	5029
7 Bitumen		13	66	1251	1330
8 Self-Blocking Bricks		36	184	3462	3682
9 Shadows		9	47	891	947
Total		**423**	**2135**	**40218**	**42776**

Table 6: Class-wise train/validation/test split for Houston.

Class	Color	Train	Validation	Test	Total
1 Healthy grass		62	62	1127	1251
2 Stressed grass		62	62	1130	1254
3 Synthetic grass		34	34	629	697
4 Trees		62	62	1120	1244
5 Soil		62	62	1118	1242
6 Water		16	16	293	325
7 Residential		63	63	1142	1268
8 Commercial		62	62	1120	1244
9 Road		62	62	1128	1252
10 Highway		61	61	1105	1227
11 Railway		61	61	1113	1235
12 Parking Lot 1		61	61	1111	1233
13 Parking Lot 2		23	23	423	469
14 Tennis Court		21	21	386	428
15 Running Track		33	33	594	660
Total		**745**	**745**	**13539**	**15029**

Table 7: Class-wise train/validation/test split for Salinas.

Class	Color	Train	Validation	Test	Total
1 Brocoli green weeds 1		20	100	1889	2009
2 Brocoli green weeds 2		37	186	3503	3726
3 Fallow		19	98	1859	1976
4 Fallow rough plow		13	69	1312	1394
5 Fallow smooth		26	133	2519	2678
6 Stubble		39	197	3723	3959
7 Celery		35	178	3366	3579
8 Grapes untrained		112	563	10596	11271
9 Soil vinyard develop		62	310	5831	6203
10 Corn senesced green weeds		32	163	3083	3278
11 Lettuce romaine 4wk		10	53	1005	1068
12 Lettuce romaine 5wk		19	96	1812	1927
13 Lettuce romaine 6wk		9	45	862	916
14 Lettuce romaine 7wk		10	53	1007	1070
15 Vinyard untrained		72	363	6833	7268
16 Vinyard vertical trellis		18	90	1699	1807
Total		**533**	**2697**	**50899**	**54129**

Table 8: Class-wise train/validation/test split for HanChuan.

Class	Color	Train	Validation	Test	Total
1 Strawberry		447	2236	42052	44735
2 Cowpea		227	1137	21389	22753
3 Soybean		102	514	9671	10287
4 Sorghum		53	267	5033	5353
5 Water spinach		12	60	1128	1200
6 Watermelon		45	226	4262	4533
7 Greens		59	295	5549	5903
8 Trees		179	898	16901	17978
9 Grass		94	473	8902	9469
10 Red roof		105	525	9886	10516
11 Gray roof		169	845	15897	16911
12 Plastic		36	183	3460	3679
13 Bare soil		91	455	8570	9116
14 Road		185	928	17447	18560
15 Bright object		11	56	1069	1136
16 Water		754	3770	70877	75401
Total		**2569**	**12868**	**242093**	**257530**

C Patch-Size Sensitivity Analysis

Table 9 evaluates the sensitivity of the asymmetric dual-view architecture to global/local view sizes. UP and Houston perform best with moderate views, whereas Salinas and HanChuan benefit from larger views, showing that the optimal spatial support is dataset dependent. Based on validation performance, we use $21/7$ for UP and Houston and $25/9$ for Salinas and HanChuan. Changing the view sizes adjusts the contextual range without changing the network width or parameter count.

Table 9: Sensitivity to global/local view sizes. The best OA, AA, and $\kappa \times 100$ for each dataset are shown in bold.

Dataset	Global	Local	OA (%)	AA (%)	$\kappa \times 100$	Params (K)	FLOPs (M)
UP	15	5	99.02±0.29	98.83±0.38	98.71±0.38	124.80	11.24
	21	7	**99.51±0.12**	**99.32±0.17**	**99.35±0.16**	124.80	21.60
	25	9	99.50±0.10	99.26±0.30	99.34±0.13	124.80	31.96
Houston	15	5	97.89±0.38	98.00±0.56	97.72±0.41	130.44	12.55
	21	7	**98.04±0.31**	**98.12±0.47**	**97.88±0.33**	130.44	24.18
	25	9	97.98±0.44	98.05±0.63	97.81±0.48	130.44	35.66
Salinas	15	5	96.90±1.17	98.55±0.55	96.55±1.30	140.74	15.11
	21	7	97.60±0.99	98.86±0.49	97.33±1.10	140.74	29.19
	25	9	**98.44±0.90**	**99.20±0.47**	**98.27±1.00**	140.74	42.89
HanChuan	15	5	98.05±0.22	96.70±0.55	97.72±0.26	147.14	16.71
	21	7	98.71±0.19	97.80±0.49	98.50±0.22	147.14	32.33
	25	9	**99.00±0.14**	**98.12±0.43**	**98.83±0.16**	147.14	47.41

D Detailed Compression and Efficiency Results

Table 10 reports the complete results underlying Fig. 5. KRGF substantially reduces model complexity from Dense64 to Compact48 while largely preserving classification performance. Further width reduction produces larger accuracy drops, supporting Compact48 as the accuracy–efficiency operating point. The measured inference time generally follows the reduction in computational cost, although small non-monotonic variations between adjacent compact widths remain, reflecting the additional effects of operator implementation and hardware utilization on wall-clock efficiency.

Table 10: Compression performance and efficiency of KGLP-Net on four HSI datasets. The last row reports the average change of Compact48 relative to Dense64.

Dataset	Model	OA (%)	Params (K)	FLOPs (M)	Inference Time (s)
UP	Dense64	99.51±0.12	124.80	21.60	2.02
	Compact48	99.45±0.12	75.31	13.66	0.90
	Compact40	99.36±0.20	55.22	10.33	0.76
	Compact32	98.98±0.32	38.24	7.41	0.63
Houston	Dense64	98.04±0.31	130.44	24.18	0.70
	Compact48	98.03±0.21	79.54	15.59	0.46
	Compact40	97.80±0.36	58.75	11.93	0.39
	Compact32	97.33±0.32	41.06	8.70	0.33
Salinas	Dense64	98.44±0.90	140.74	42.89	3.78
	Compact48	98.45±0.83	87.27	28.35	2.79
	Compact40	98.05±0.85	65.19	22.03	2.66
	Compact32	96.89±0.86	46.21	16.35	2.41
HanChuan	Dense64	99.00±0.14	147.14	47.41	18.32
	Compact48	98.94±0.14	92.07	31.74	11.99
	Compact40	98.79±0.16	69.19	24.86	12.22
	Compact32	98.62±0.20	49.41	18.61	12.04
Avg. Compact48 vs. Dense64		-0.03 **pp**	-38.5%	-34.8%	-37.6%

E DETAILED CLASS-WISE RESULTS

Tables 11–14 provide class-wise results corresponding to the aggregate comparison in Table 1. Compact48 maintains consistently strong performance across diverse land-cover categories rather than deriving its overall advantage from only a few classes. Strong baselines remain competitive on several individual categories, indicating that the overall gain reflects more balanced cross-class discrimination. All entries are reported as mean ± standard deviation over 10 independent runs. The best and second-best mean accuracies are shown in **bold** and underline, respectively.

Table 11: Class-wise accuracy and aggregate classification metrics on UP.

	CNN-Based			Transformer-Based			SSM-Based				
Class	CVSSN	SSRN	HKNAS3D	SSFTT	GAHT	CMT	MambaHSI	3DSS-Mamba	MambaLG	SliMamba	Compact48 (Proposed)
1 Asphalt	97.06±1.43	**99.45±0.48**	95.53±1.11	93.51±2.39	96.12±1.26	81.78±18.24	94.12±2.27	88.17±2.64	97.67±0.94	97.23±1.04	99.20±0.55
2 Meadows	99.51±0.45	99.67±0.26	97.10±1.30	97.08±1.70	99.50±0.26	85.01±10.93	97.59±1.04	96.90±0.67	99.66±0.22	99.69±0.13	**99.80±0.16**
3 Gravel	95.80±2.75	94.22±2.20	88.33±5.22	90.51±1.76	88.76±3.83	72.52±19.27	90.46±4.30	75.34±6.33	90.65±3.84	83.97±7.25	**98.34±1.13**
4 Trees	98.46±0.54	96.48±1.73	95.81±2.31	**98.51±0.83**	96.89±1.32	94.24±2.60	92.86±3.47	86.27±4.32	92.00±3.05	95.85±1.27	98.44±0.47
5 Painted metal sheets	99.98±0.05	**100.00±0.00**	99.75±0.43	**100.00±0.00**	**100.00±0.00**	98.32±1.88	**100.00±0.00**	97.17±2.83	99.85±0.27	99.85±0.22	**100.00±0.00**
6 Bare Soil	98.72±1.73	99.34±0.79	96.55±2.18	97.87±1.31	98.37±1.05	79.15±25.84	99.00±0.80	80.07±5.35	99.35±0.81	98.21±1.49	**99.95±0.10**
7 Bitumen	96.61±2.51	99.32±0.87	90.58±5.64	93.33±3.99	95.22±1.99	82.02±24.63	96.62±2.62	62.89±9.02	89.28±6.98	98.06±1.92	**99.92±0.17**
8 Self-Blocking Bricks	96.32±2.02	96.85±1.70	91.00±5.10	92.85±3.86	93.96±1.82	72.83±22.34	95.79±1.62	89.05±3.99	94.47±2.80	91.61±3.97	**98.30±1.51**
9 Shadows	99.71±0.39	**99.89±0.16**	99.38±0.50	99.44±0.64	99.61±0.42	94.89±3.84	99.37±0.67	77.13±19.71	94.89±3.78	96.52±1.28	99.63±0.67
OA	98.44±0.23	98.86±0.16	95.67±0.88	96.06±1.07	97.54±0.44	83.37±4.94	96.46±0.46	89.59±1.13	97.46±0.31	97.27±0.39	**99.45±0.12**
AA	98.02±0.45	98.36±0.28	94.89±0.83	95.90±0.71	96.49±0.70	84.53±3.60	96.20±0.87	83.66±1.96	95.31±0.81	95.66±0.60	**99.28±0.12**
$\kappa \times 100$	97.93±0.31	98.49±0.21	94.29±1.14	94.81±1.38	96.74±0.58	78.53±5.87	96.38±2.05	86.16±1.50	96.62±0.41	96.39±0.51	**99.28±0.16**

Table 12: Class-wise accuracy and aggregate classification metrics on Houston.

	CNN-Based			Transformer-Based			SSM-Based				
Class	CVSSN	SSRN	HKNAS3D	SSFTT	GAHT	CMT	MambaHSI	3DSS-Mamba	MambaLG	SliMamba	Compact48 (Proposed)
1 Healthy grass	98.28±2.73	**99.06±0.84**	96.40±3.41	97.79±2.81	98.80±0.94	96.67±2.37	95.37±4.53	94.23±3.03	98.15±1.91	97.35±1.76	98.91±1.53
2 Stressed grass	98.63±1.09	99.09±0.92	97.44±2.10	98.22±1.08	98.18±1.41	92.70±7.47	97.27±1.42	94.06±3.08	**99.30±0.53**	98.48±0.55	98.02±1.68
3 Synthetic grass	99.92±0.19	**100.00±0.00**	99.87±0.25	**100.00±0.00**	99.94±0.11	99.98±0.05	99.70±0.44	98.36±1.92	99.05±0.79	99.90±0.16	**100.00±0.00**
4 Trees	94.41±0.58	**99.58±0.49**	94.88±1.81	94.96±1.78	96.80±2.61	97.39±2.64	97.12±2.32	92.18±3.38	98.21±1.23	98.74±1.04	97.12±1.68
5 Soil	99.90±0.24	99.85±0.46	99.11±2.04	99.76±0.43	99.97±0.06	99.67±0.82	99.92±0.13	99.19±0.67	99.93±0.10	99.94±0.19	**99.98±0.04**
6 Water	98.36±4.69	95.22±5.44	98.02±4.50	97.54±4.36	90.58±3.79	**98.63±3.55**	98.46±1.48	89.90±5.79	95.70±4.05	97.95±4.85	95.73±5.61
7 Residential	95.46±1.88	**98.05±0.53**	91.89±3.03	90.72±3.31	93.61±2.47	92.06±3.46	92.91±4.49	84.77±5.38	96.42±2.23	96.70±2.27	97.78±1.21
8 Commercial	91.62±2.35	92.23±2.77	87.56±3.05	90.91±2.97	91.17±3.37	88.70±5.60	79.86±5.57	85.98±3.18	91.20±2.66	92.29±2.58	**92.78±2.11**
9 Road	95.00±2.78	93.64±2.83	88.21±3.60	90.73±2.88	91.94±2.47	78.68±16.31	90.64±2.21	88.22±4.17	92.89±3.77	94.71±2.44	**97.37±1.78**
10 Highway	99.08±1.52	98.85±1.46	95.20±3.13	96.77±2.44	97.56±1.69	81.10±22.36	95.89±2.57	93.38±4.48	98.24±1.87	98.88±1.41	**99.81±0.38**
11 Railway	97.90±1.62	97.61±1.30	88.48±4.37	87.39±4.54	96.01±1.32	68.24±16.54	90.92±4.95	93.02±2.50	96.86±2.50	**99.26±0.89**	98.10±0.98
12 Parking Lot 1	96.55±1.07	97.20±2.76	92.42±3.41	94.23±1.91	95.53±1.59	69.69±15.19	90.35±3.47	91.51±4.66	96.80±2.67	95.18±3.28	**98.12±1.62**
13 Parking Lot 2	97.87±2.14	94.89±1.79	95.32±1.22	96.12±1.28	96.05±1.26	95.13±2.50	98.11±2.15	89.50±3.75	94.00±2.12	92.10±2.92	**98.44±2.00**
14 Tennis Court	**100.00±0.00**	**100.00±0.00**	99.82±0.33	**100.00±0.00**	99.82±0.28	99.84±0.39	**100.00±0.00**	98.99±0.76	99.22±1.11	**100.00±0.00**	**100.00±0.00**
15 Running Track	**100.00±0.00**	**100.00±0.00**	99.90±0.22	99.85±0.45	**100.00±0.00**	98.13±3.73	97.61±1.42	98.55±1.31	99.81±0.21	99.98±0.05	**100.00±0.00**
OA	97.14±0.40	97.68±0.41	94.13±0.77	94.97±0.81	96.31±0.60	88.58±2.10	93.97±1.08	92.36±1.52	96.99±0.53	97.34±0.42	**98.03±0.21**
AA	97.53±0.41	97.69±0.53	94.97±0.66	95.67±0.75	96.40±0.61	90.44±1.65	94.94±0.93	92.79±1.48	97.05±0.42	97.43±0.52	**98.14±0.42**
$\kappa \times 100$	96.91±0.43	97.49±0.44	93.66±0.83	94.56±0.87	96.01±0.65	87.66±2.27	94.00±2.31	91.74±1.65	96.75±0.57	97.13±0.46	**97.87±0.23**

Table 13: Class-wise accuracy and aggregate classification metrics on Salinas.

	CNN-Based			Transformer-Based			SSM-Based				
Class	CVSSN	SSRN	HKNAS3D	SSFTT	GAHT	CMT	MambaHSI	3DSS-Mamba	MambaLG	SliMamba	Compact48 (Proposed)
1 Brocoli green weeds 1	99.95±0.11	**100.00±0.00**	98.96±1.61	99.47±0.74	99.28±1.26	99.80±0.29	99.97±0.08	98.58±1.56	99.63±0.48	99.97±0.05	99.97±0.10
2 Brocoli green weeds 2	99.96±0.04	**100.00±0.00**	99.39±1.35	99.98±0.05	99.85±0.26	99.65±0.52	98.43±3.45	97.04±1.96	99.99±0.03	99.96±0.08	**100.00±0.00**
3 Fallow	99.70±0.28	98.49±2.77	97.70±1.69	97.73±2.06	97.64±2.43	96.63±9.10	75.34±20.96	92.79±5.21	93.51±9.46	99.87±0.20	**99.96±0.13**
4 Fallow rough plow	99.35±0.48	**99.92±0.20**	98.11±2.04	99.73±0.65	99.60±0.50	94.67±8.32	99.90±0.10	94.56±3.45	99.27±0.40	99.27±0.74	**99.92±0.12**
5 Fallow smooth	95.54±3.07	97.58±1.31	97.46±1.75	97.25±1.95	97.36±1.43	96.67±2.50	95.44±3.31	96.17±2.08	97.36±1.63	97.84±1.73	**98.22±0.89**
6 Stubble	99.95±0.09	99.99±0.02	99.79±0.33	99.54±0.66	99.95±0.12	99.88±0.21	99.78±0.35	99.40±0.53	**100.00±0.01**	99.96±0.07	99.97±0.09
7 Celery	99.83±0.20	**100.00±0.01**	98.53±1.56	99.84±0.16	99.83±0.20	99.77±0.30	99.57±0.59	97.77±0.93	99.82±0.20	99.66±0.32	99.98±0.03
8 Grapes untrained	87.97±1.62	93.88±3.10	81.61±3.28	79.79±4.43	90.65±1.60	83.88±9.47	85.59±6.85	92.42±2.47	**97.49±1.44**	94.40±1.58	95.15±2.63
9 Soil vinyard develop	99.47±0.43	99.98±0.06	98.21±2.29	99.71±0.31	99.92±0.14	99.49±1.38	98.37±4.89	98.79±0.74	99.95±0.08	99.98±0.02	**100.00±0.00**
10 Corn senesced green weeds	98.58±0.80	96.57±1.19	97.30±1.66	97.33±1.49	95.76±2.80	95.35±4.12	83.17±6.75	92.86±2.50	96.09±1.71	**98.69±0.69**	97.99±1.16
11 Lettuce romaine 4wk	99.31±1.15	98.44±1.37	96.65±2.86	94.89±4.17	96.93±2.72	98.61±2.38	95.32±10.08	87.14±3.13	91.99±4.37	99.44±0.72	**99.84±0.35**
12 Lettuce romaine 5wk	99.65±0.32	**100.00±0.00**	99.06±1.51	99.64±0.36	99.88±0.22	94.92±5.42	91.35±11.73	92.45±3.83	99.07±0.94	99.61±0.44	99.89±0.10
13 Lettuce romaine 6wk	99.30±1.02	99.06±1.42	99.18±2.07	96.90±4.16	98.13±2.39	92.38±12.78	96.64±5.98	96.16±4.81	93.70±11.52	99.57±0.66	**99.74±0.32**
14 Lettuce romaine 7wk	99.40±0.61	98.28±1.30	98.88±1.59	97.02±3.28	98.03±1.61	98.14±2.86	95.28±5.85	86.78±8.40	93.76±4.49	98.18±1.80	**99.63±0.24**
15 Vinyard untrained	92.49±2.02	92.82±2.53	87.13±6.47	86.12±3.80	89.40±2.25	82.63±13.91	93.61±7.83	91.58±3.20	95.34±3.34	95.18±2.04	**98.09±2.16**
16 Vinyard vertical trellis	**99.12±0.64**	97.65±3.58	98.57±1.35	97.91±1.23	95.92±5.04	97.89±1.46	98.47±2.71	90.34±5.28	95.07±5.91	98.05±1.87	98.56±4.00
OA	95.99±0.37	97.21±0.45	93.44±0.57	93.15±1.12	95.81±0.31	93.05±1.48	93.04±2.80	94.51±1.03	97.61±0.60	97.81±0.55	**98.45±0.83**
AA	98.10±0.29	98.29±0.32	96.66±0.34	96.43±0.63	97.38±0.44	95.65±1.06	94.14±2.87	94.05±1.22	97.00±0.74	98.73±0.31	**99.18±0.47**
$\kappa \times 100$	95.55±0.42	96.90±0.49	92.71±0.63	92.39±1.25	95.34±0.34	92.27±1.64	92.67±3.53	93.88±1.15	97.34±0.66	97.56±0.62	**98.27±0.92**

Table 14: Class-wise accuracy and aggregate classification metrics on HanChuan.

	CNN-Based			Transformer-Based			SSM-Based				
Class	CVSSN	SSRN	HKNAS3D	SSFTT	GAHT	CMT	MambaHSI	3DSS-Mamba	MambaLG	SliMamba	Compact48 (Proposed)
1 Strawberry	98.05±0.53	97.92±0.52	89.94±2.89	95.20±1.14	97.81±1.07	89.58±3.75	89.57±5.74	93.88±1.05	99.00±0.34	96.75±0.60	**99.27±0.27**
2 Cowpea	97.41±0.59	94.33±0.94	88.13±5.11	90.88±1.37	96.62±0.72	84.49±7.95	81.69±4.83	86.63±3.02	98.07±0.64	92.14±1.39	**99.03±0.65**
3 Soybean	98.64±0.68	90.40±3.64	93.24±3.22	87.62±2.17	95.73±1.77	86.70±5.37	94.47±2.88	80.01±5.72	99.30±0.38	95.41±1.04	**99.91±0.07**
4 Sorghum	98.23±1.16	97.65±0.68	96.72±2.74	96.60±1.21	98.07±1.24	96.68±1.80	98.40±1.64	93.44±1.42	96.26±1.53	97.34±1.13	**99.55±0.43**
5 Water spinach	94.82±4.10	81.17±7.85	93.30±5.95	76.70±8.00	91.14±5.53	78.68±15.51	**99.67±0.44**	40.93±11.70	59.39±33.46	93.69±3.75	99.33±1.30
6 Watermelon	81.38±4.98	64.45±5.70	81.69±5.12	67.61±2.37	69.90±4.41	62.45±10.38	60.40±11.74	41.32±5.11	78.66±15.45	74.28±4.65	**94.40±3.01**
7 Greens	95.13±2.02	91.86±1.59	93.28±3.26	88.54±2.79	89.65±2.87	77.13±16.13	94.53±3.72	82.22±5.40	93.67±2.98	90.83±2.93	**99.16±0.92**
8 Trees	90.13±3.13	92.59±2.28	78.67±10.66	86.97±2.39	90.44±1.58	76.74±9.17	76.97±6.65	79.52±4.21	96.91±1.24	89.83±0.98	**97.97±0.50**
9 Grass	96.10±1.29	94.49±1.53	89.22±2.61	89.33±2.79	91.24±2.65	87.37±4.15	86.67±4.87	71.19±3.76	95.23±2.04	89.13±1.95	**98.97±0.58**
10 Red roof	98.46±0.71	97.06±1.53	97.19±1.19	96.03±1.46	98.42±0.89	93.26±7.36	91.76±3.17	92.14±1.69	98.25±1.16	97.38±1.25	**99.38±0.27**
11 Gray roof	96.28±2.97	96.80±0.95	90.12±5.01	95.72±0.98	95.17±1.80	84.96±12.75	89.89±6.81	94.45±1.38	97.68±1.34	97.15±0.73	**99.18±0.38**
12 Plastic	91.27±4.78	76.15±5.45	90.59±5.84	78.64±4.13	83.99±4.13	70.53±19.03	84.96±8.19	74.03±5.93	89.89±11.51	89.13±3.76	**98.64±1.17**
13 Bare soil	90.97±1.84	76.09±4.48	86.61±3.21	77.70±2.81	81.67±3.66	70.23±9.38	78.10±4.70	68.31±2.35	88.86±2.26	78.33±3.69	**94.96±1.29**
14 Road	94.88±1.50	92.52±1.76	87.04±4.69	89.62±2.02	93.21±3.06	86.94±8.94	91.39±2.51	87.85±2.84	97.55±0.57	91.74±1.73	**98.83±0.66**
15 Bright object	84.42±6.46	69.80±9.25	85.73±6.25	83.56±6.64	79.46±9.03	81.19±7.08	**94.80±2.09**	55.49±3.64	26.70±27.41	79.74±6.24	90.68±5.19
16 Water	99.39±0.20	99.54±0.30	97.23±3.25	98.21±0.31	99.45±0.37	99.00±0.53	96.87±3.23	99.27±0.28	**99.82±0.06**	99.10±0.67	99.55±0.23
OA	96.68±0.53	94.76±0.45	91.29±0.79	92.67±0.41	95.31±0.40	88.79±1.83	89.85±1.08	89.15±0.68	97.13±0.77	94.54±0.38	**98.94±0.14**
AA	94.10±1.14	88.30±1.03	89.92±0.66	87.43±0.96	90.75±0.83	82.87±2.61	88.13±1.16	77.54±1.30	88.45±3.98	90.75±0.65	**98.05±0.42**
$\kappa \times 100$	96.11±0.62	93.87±0.53	89.87±0.90	91.43±0.47	94.51±0.46	86.92±2.12	86.67±3.17	87.30±0.79	96.64±0.90	93.61±0.44	**98.76±0.16**

F Full Qualitative Classification Maps

Figures 7–10 extend Fig. 3 to all ten competing methods and Compact48. For each method, we show the run whose OA is closest to its mean over 10 independent runs, providing full-scene comparisons of classification errors.

On UP, several methods produce isolated errors within homogeneous regions and around narrow structures, whereas Compact48 more closely follows the ground truth in these areas and along local class boundaries.

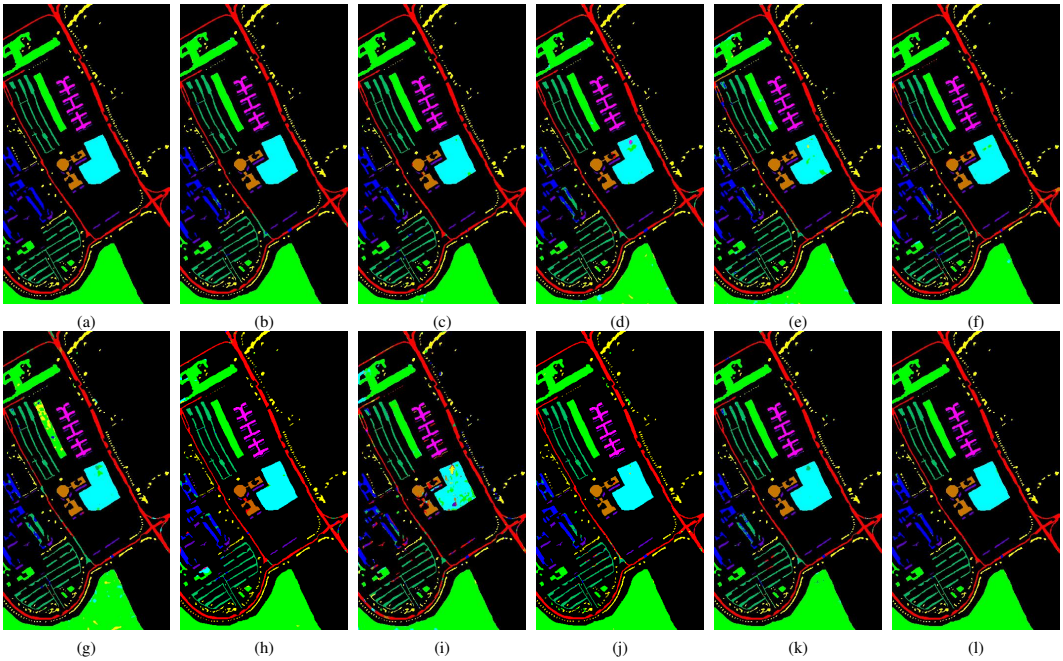


Figure 7: Classification maps of the UP dataset (corresponding to Table 11). (a) Ground truth. (b) CVSSN. (c) SSRN. (d) HKNAS3D. (e) SSFTT. (f) GAHT. (g) CMT. (h) MambaHSI. (i) 3DSS-Mamba. (j) MambaLG. (k) SliMamba. (l) Compact48.

On Houston, the labeled regions are spatially dispersed and often small, so differences are mainly manifested as localized class substitutions in small regions and short linear structures rather than large contiguous errors. Compact48 shows closer agreement with the ground truth in these locations.

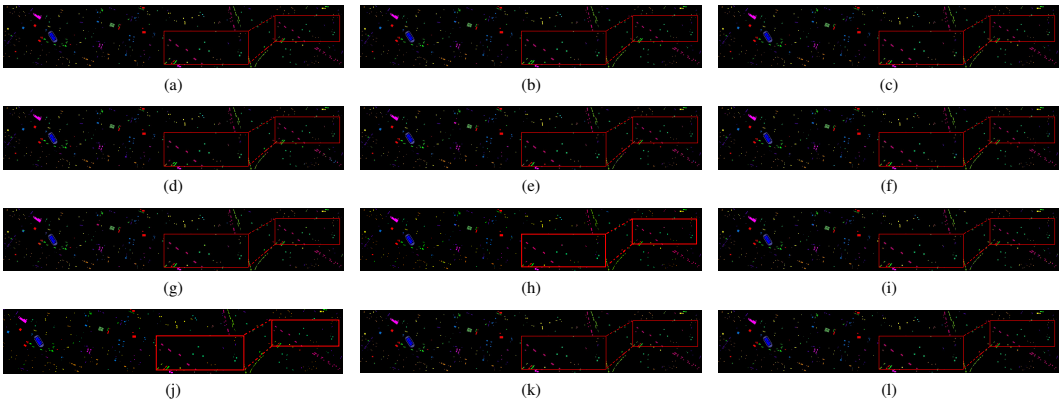


Figure 8: Classification maps of the Houston dataset (corresponding to Table 12). (a) Ground truth. (b) CVSSN. (c) SSRN. (d) HKNAS3D. (e) SSFTT. (f) GAHT. (g) CMT. (h) MambaHSI. (i) 3DSS-Mamba. (j) MambaLG. (k) SliMamba. (l) Compact48.

On Salinas, scattered misclassifications are clearly visible within large homogeneous agricultural fields for several competing methods. Compact48 substantially reduces these errors while preserving the boundaries and continuity of narrow field regions.

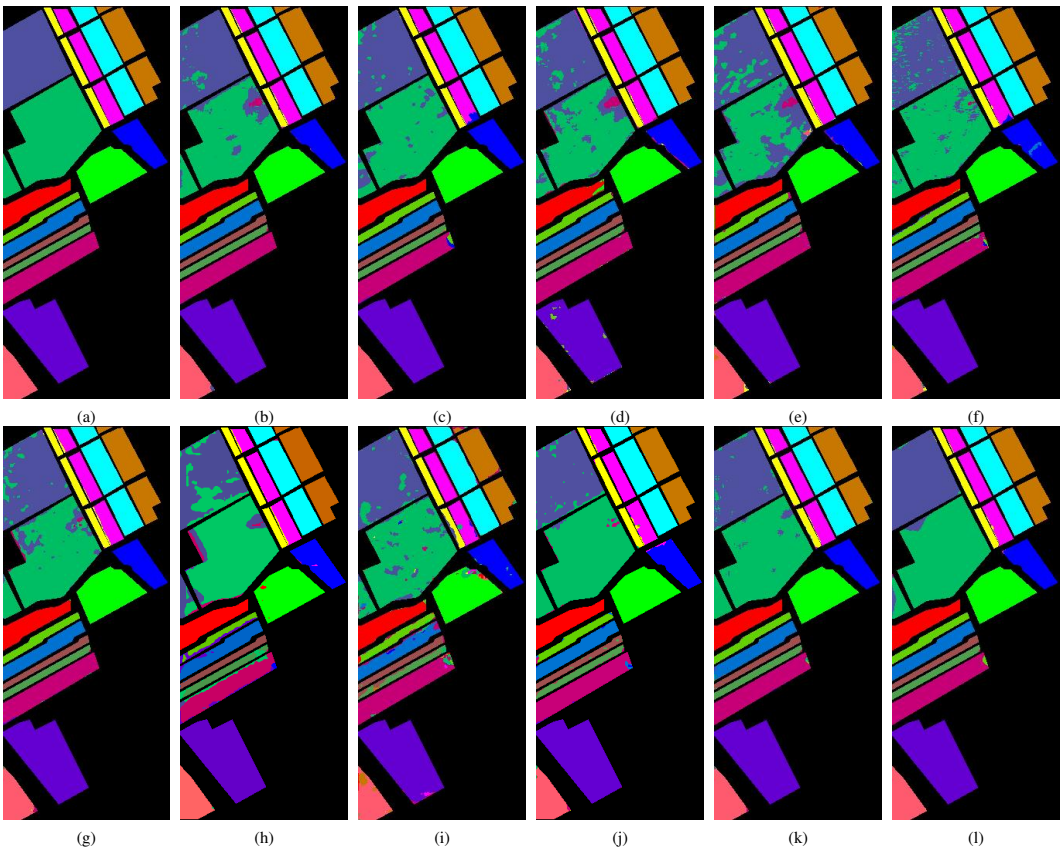


Figure 9: Classification maps of the Salinas dataset (corresponding to Table 13). (a) Ground truth. (b) CVSSN. (c) SSRN. (d) HKNAS3D. (e) SSFTT. (f) GAHT. (g) CMT. (h) MambaHSI. (i) 3DSS-Mamba. (j) MambaLG. (k) SliMamba. (l) Compact48.

HanChuan presents a more complex combination of large homogeneous regions, an elongated central structure, and densely arranged narrow regions. Several methods exhibit scattered misclassifications

or local fragmentation in these areas, while Compact48 more consistently preserves both large-region homogeneity and fine spatial structures.

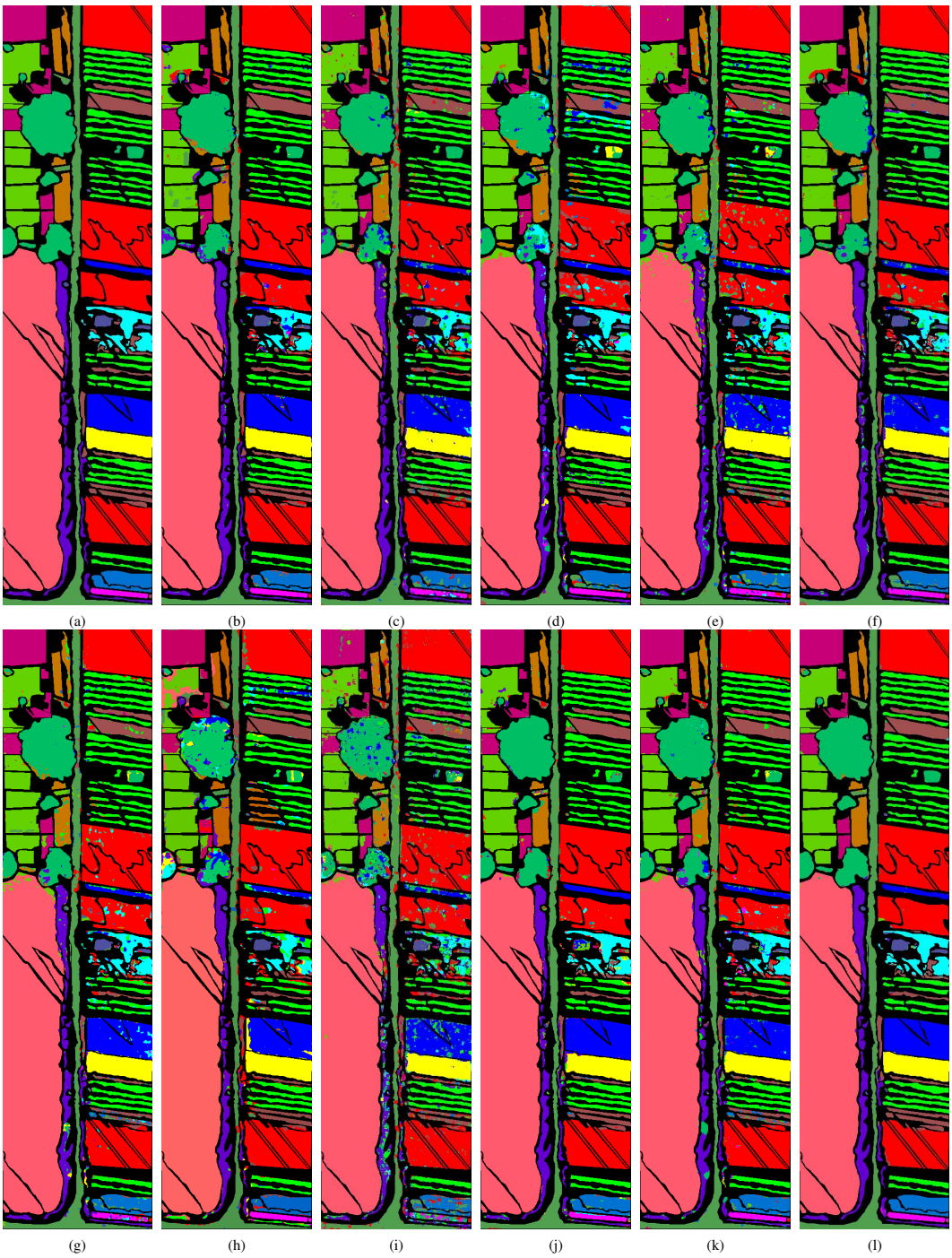


Figure 10: Classification maps of the HanChuan dataset (corresponding to Table 14). (a) Ground truth. (b) CVSSN. (c) SSRN. (d) HKNAS3D. (e) SSFTT. (f) GAHT. (g) CMT. (h) MambaHSI. (i) 3DSS-Mamba. (j) MambaLG. (k) SliMamba. (l) Compact48.

Overall, the full-scene comparison shows that the qualitative advantage of Compact48 is primarily reflected in fewer localized misclassifications, improved intra-region consistency, and better preservation of class boundaries and narrow structures. These observations complement the representative comparisons in Fig. 3 and are consistent with the quantitative results in Table 1.