Cross-Domain Hyperspectral Image Classification via Mamba-CNN and Knowledge Distillation

Aoyan Du[1], Guixin Zhao[1], **Yuyao Wang[2],**
Aimei Dong[1], Xiangjun Dong[1]

[1]Shandong Computer Science Center
[2]Department of Mathematics and Statistics, Boston University

IEEE TRGS 2025

Outline

1. Introduction
   - Background
   - Motivation
2. Method
3. Experiments
4. Conclusion

Background

- Remote sensing data: **HSI, LiDAR, SAR**
- HSI → **fine-grained spectral info**, accurate land cover classification
- Problem: **scene-specific methods**, poor generalization
- Need: apply model from **Source Domain (SD)** → **Target Domain (TD)**

Motivation → Our Solution

1. Feature Extraction Challenge

Problem:

CNN → good at **local details**, weak at long-range.

Transformer → good at **global context**, but costly & miss fine details.

Mamba → efficient for long-range, weak for local.

Our Solution:

Hybrid Network (MKDnet) = Mamba + CNN

→ **joint global + local feature extraction**

→ capture **long-range dependencies + fine details** in HSI.

Motivation → Our Solution

2. Inter-Class Feature Consistency

Problem:

Most DA methods align **marginal distributions** only.

→ Ignore **class-level relationships**, leading to confusion between classes.

Our Solution:

Knowledge Distillation (STKD) with **soft labels**

→ preserve **inter-class similarities & differences**

→ stronger alignment, less class confusion.

Motivation → Our Solution

3. Robust Domain Alignment

Problem:

- Traditional alignment ignores **graph topology**.
- Large SD–TD discrepancy → risk of **negative transfer**.

Our Solution:

- **Graph Alignment with GCN + Consistency Constraints**
- → dynamically model **spectral–spatial topology**
- → enforce consistency in **distribution + graph alignment**
- → robust against domain shift.

Outline

1. Introduction
2. Method
3. Experiments
4. Conclusion

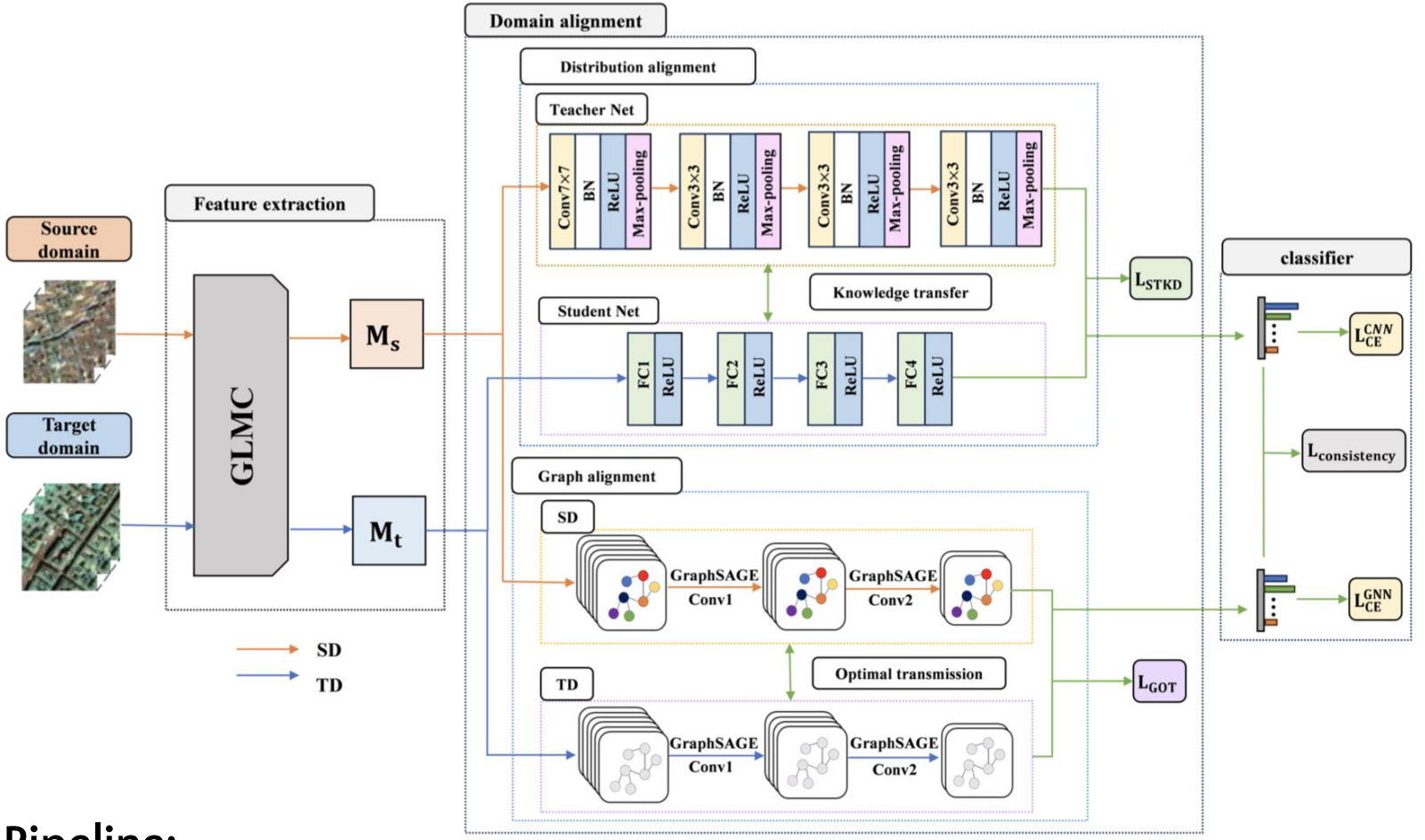
Overview

- **Mamba**
  - New **State Space Model (SSM)** → combines control theory + DL
  - **Strength**: captures long-range dependencies efficiently
  - **Gap**: underexplored for **high-dim HSI** → potential for global feature modeling
- **Knowledge Distillation (KD)**
  - **Student learns from teacher** → hard + soft labels
  - **Soft labels**: encode inter-class similarity, improve generalization
  - **Gap**: limited use in cross-domain HSI → can enhance **inter-class consistency**
- **Graph Neural Networks (GNNs)**
  - **Key idea**: aggregate neighbor info recursively, designed for **graph-structured data**
  - **GraphSAGE**: iterative neighbor sampling + aggregation
  - **Relevance**: model **spectral–spatial topology** in HSI, support alignment

Method:

The proposed **MKDnet** consists of three key modules:

- **Feature extractor** – jointly captures global and local HSI features.
- **Domain alignment** – includes distribution alignment (via knowledge distillation) and graph alignment (via GOT).
- **Classifier** – ensures domain-invariant prediction.



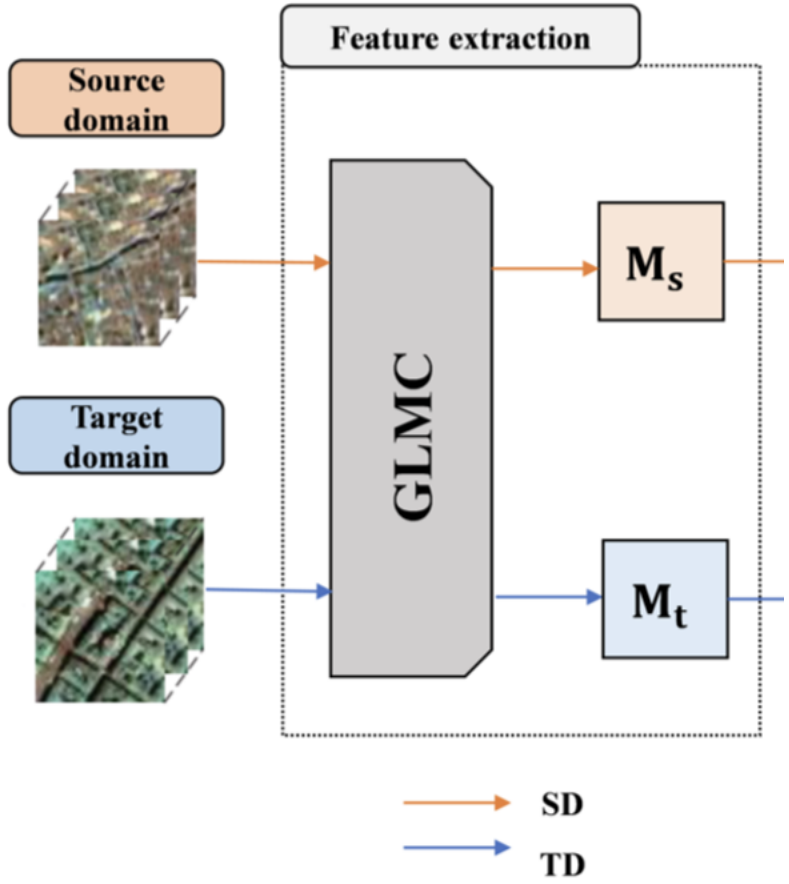
Pipeline:

- Input labeled patches from SD + unlabeled patches from TD
- Extract features → align (distribution + graph) → classify TD

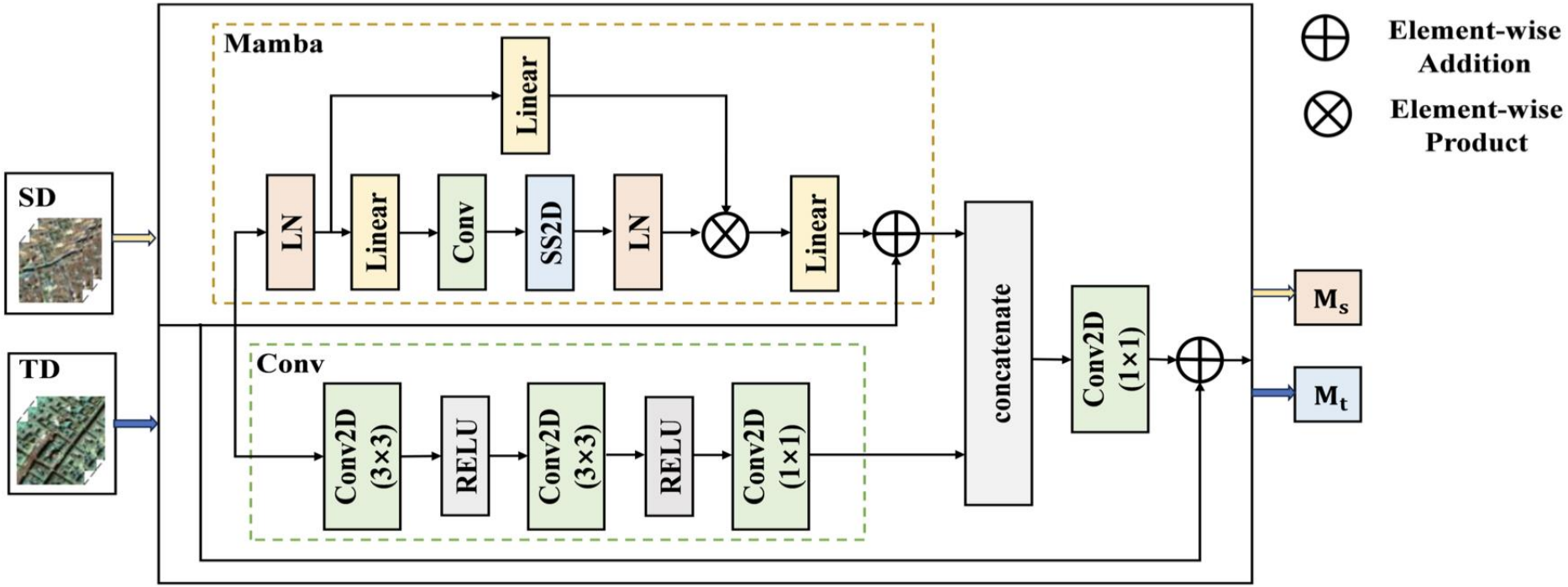
Method: A. Feature Extractor (GLMC)

- **Motivation**: HSI requires **global + local features**
  - Global → capture long-range dependencies
  - Local → capture fine-grained details

- **Output**: Rich feature representations
  - M_s for source, M_t for target
  - Passed to **domain alignment** stage



Method: A. Feature Extractor (GLMC)

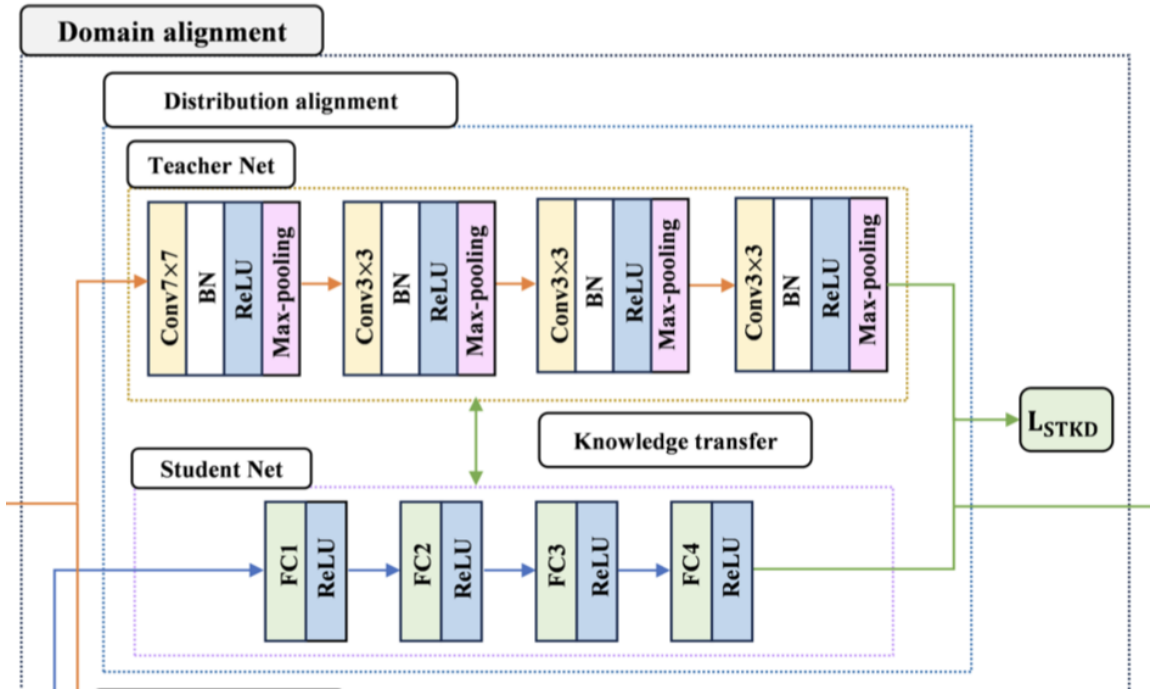


- **Design**: Dual-branch module
  - **Mamba branch**: models **global dependencies** via selective scanning (SS2D), linear + conv layers, residual mixing
  - **CNN branch**: three Conv2D layers for **local spatial patterns**
- **Fusion**:
  - Concatenate outputs from both branches
  - Apply 1×1Conv to fuse channels → joint feature maps

Method: B. Domain Alignment

1) Distribution Alignment (STKD)

- **Problem**: Domain shift → SD and TD feature distributions differ.
- **Idea**: Use **knowledge distillation** to transfer richer feature info from SD → TD.
- **Design**:
  - Teacher network (trained on SD) produces **soft labels** → capture inter-class relationships.
  - Student network (on TD) learns to match teacher outputs via **distillation loss**.
  - Combine **distillation loss + alignment loss** for balanced training.

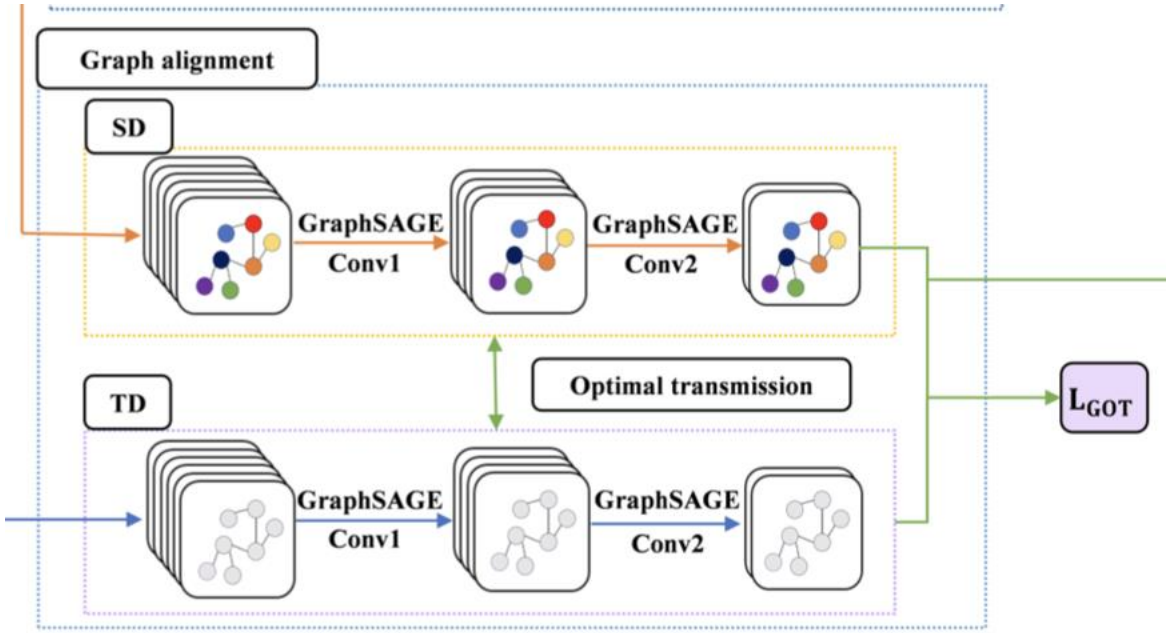


- **Effect**: Enhances **inter-class consistency**, reduces domain gap.

Method: B. Domain Alignment

2) Graph Alignment

- **Problem**: Distribution alignment alone ignores **topological structure** in HSI.
- **Idea**: Use **GraphSAGE** to build subgraphs from features M_s, M_t.
- **Design**:
  - Recursive neighbor aggregation → capture local + higher-order structure.
  - Apply **Graph Optimal Transport (GOT)** to align subgraph features layer by layer.
- **Effect**: Preserves **spectral–spatial topology** across domains, improves robustness against domain shift.



Method: C. Classifiers of MKDnet

- **Problem**: Need to leverage both **image features (CNN)** and **graph features (GNN)** for robust classification.
- **Design**:
  - **CNN classifier**: handles spectral–spatial image features.
  - **GNN classifier**: handles topology-aware graph features.
  - Both trained with **cross-entropy loss** on source data.
- **Consistency Constraint**:
  - Enforce probability outputs of CNN & GNN to be consistent.
  - Prevents feature imbalance, ensures stable predictions.
- **Overall Loss**:

$$L = L_{STKD} + L_{GOT} + L_{CE}^{CNN} + L_{CE}^{GNN} + L_{consistency}$$

- **Effect**: Combines **distribution alignment + graph alignment + dual classifiers**,
  → reduces domain shift, improves cross-domain HSI classification accuracy.

Outline

1. Introduction
2. Method
3. Experiments
4. Conclusion

Experiments Datasets

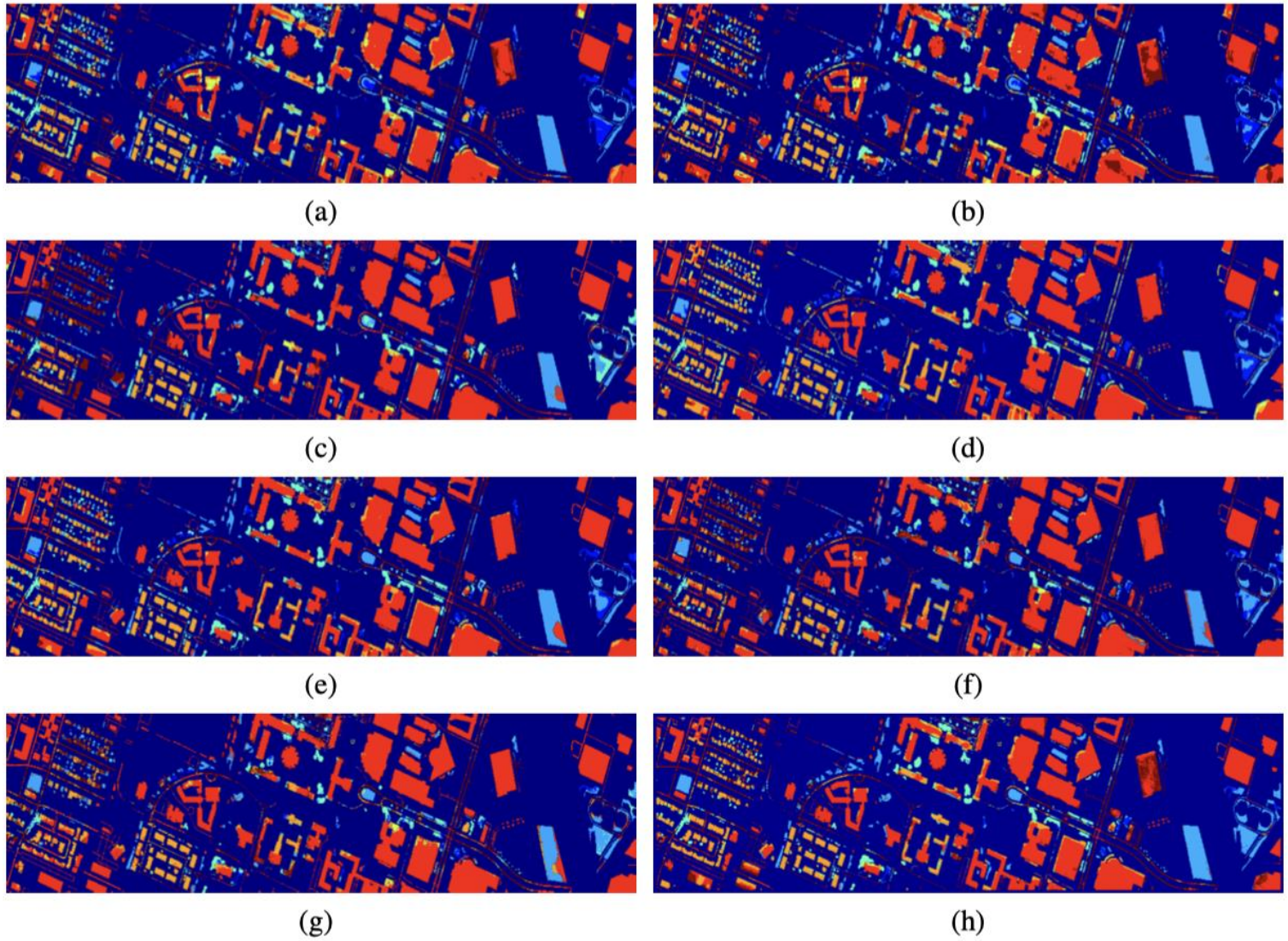
- **Houston (2013 & 2018)**
  - Different sensors, temporal gap → clear **domain shift**
  - 7 common classes used for DA task
- **HyRANK (Dioni & Loukia)**
  - Hyperion sensor, 176 bands
  - 12 common classes → vegetation prone to misclassification
- **Pavia (University & Center)**
  - Acquired by ROSIS sensor, Italy
  - 7 consistent classes → urban variation, same city but different scenes

Experiments Setup

- **Compared methods**:
  - DA-based: **DAN, MRAN, DSAN, TSTnet, MKDnet**
  - DG-based: **SagNet, SDEnet, S²AMSnet**
- **Evaluation metrics**: **CA (Class Accuracy)**, **OA (Overall Accuracy)**, **KC (Kappa Coefficient)**
- **Training settings**: DA: 80% SD labeled + all TD; DG: 80% SD labeled (20% validation)
- **Implementation**:
  - PyTorch + PyTorch Geometric (GraphSAGE)
  - Optimizer: SGD + momentum 0.9
  - Iterations: 500, LR schedule $\eta_w = \eta_0 / \left(1 + \alpha w\right)^{\beta}$

Experiments Results: Houston.

- Visual results of **Houston**
  - MKDnet produces cleaner, more accurate classification maps across **Houston**.
  - Outperforms 7 advanced DA/DG methods.



(a) DAN. (b) MRAN. (c) SagNet. (d) TSTnet.
(e) DSAN. (f) SDEnet. (g) S^2 AMSnet. (h) MKDnet.

Experiments Results: Houston.

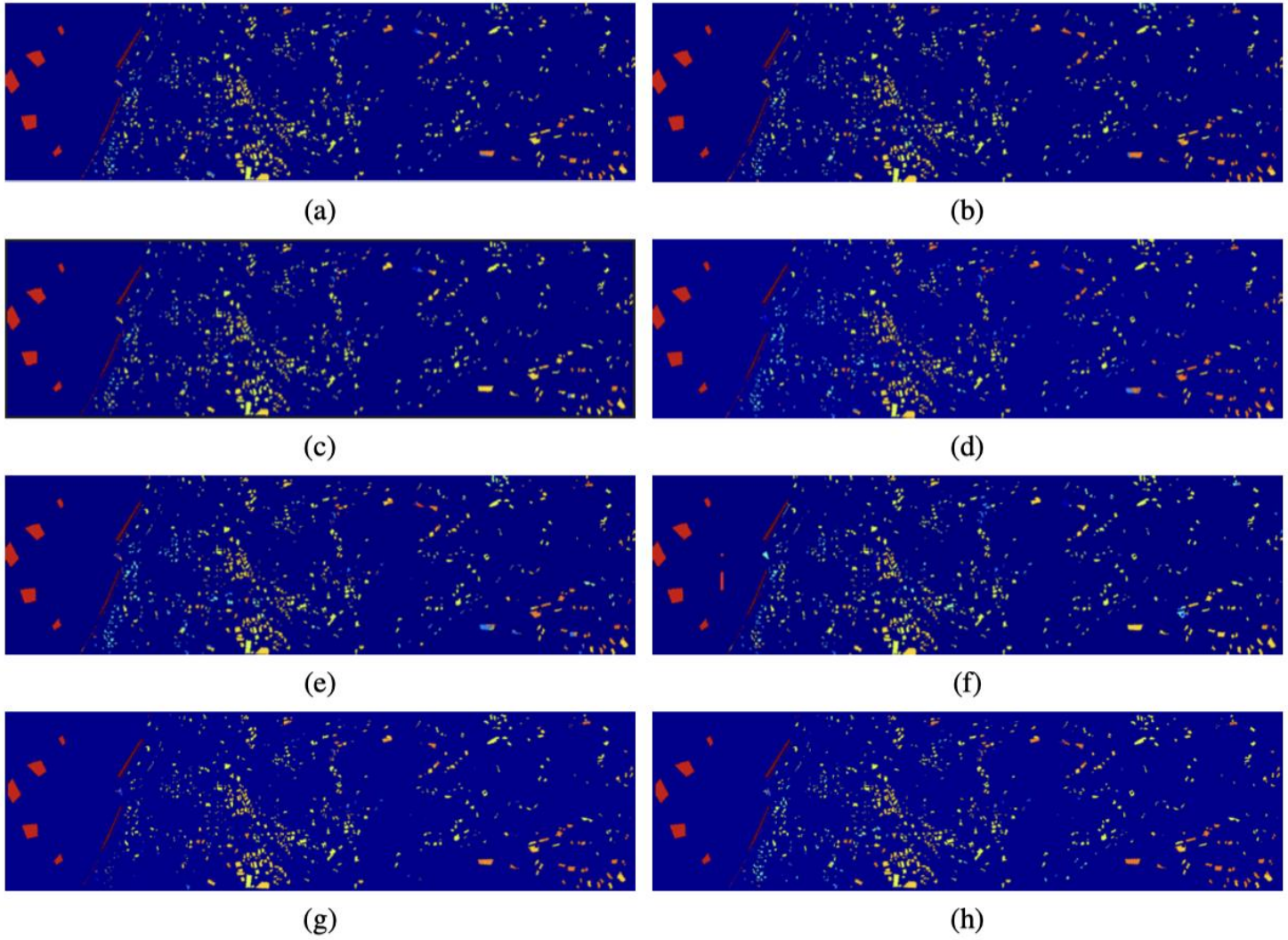
Table; Classification Results on Pavia Center Dataset

Class	DAN	SagNet	MRAN	TSTNet	DSAN	SDEnet	S^2AMSNet	MKDnet
1	50.85	24.79	54.39	82.03	57.95	25.49	35.79	38.59
2	72.38	62.19	75.77	68.73	67.90	83.26	59.99	85.29
3	61.77	48.66	68.79	52.82	71.69	58.31	77.64	69.08
4	100	81.02	64.33	100	81.82	100	100	76.36
5	67.43	59.57	68.50	61.71	61.79	62.34	45.95	75.58
6	73.11	88.29	70.54	84.44	70.26	85.36	90.12	77.92
7	63.99	33.99	54.21	53.07	54.53	55.16	64.48	54.10
OA(%)	67.77	73.11	68.99	75.34	67.08	75.99	75.73	76.11
KC(x100)	51.17	55.39	52.67	59.69	52.05	60.77	60.54	60.87

- MKDnet surpasses baselines by **1%–9% OA** on key target scenes on Houston.

Experiments Results: HyRANK.

- Visual results of **Houston**
  - MKDnet produces cleaner, more accurate classification maps across **HyRANK**.
  - (Outperforms 7 advanced DA/DG methods.



(a) DAN. (b) MRAN. (c) SagNet. (d) TSTnet.
(e) DSAN. (f) SDEnet. (g) S^2 AMSnet. (h) MKDnet.

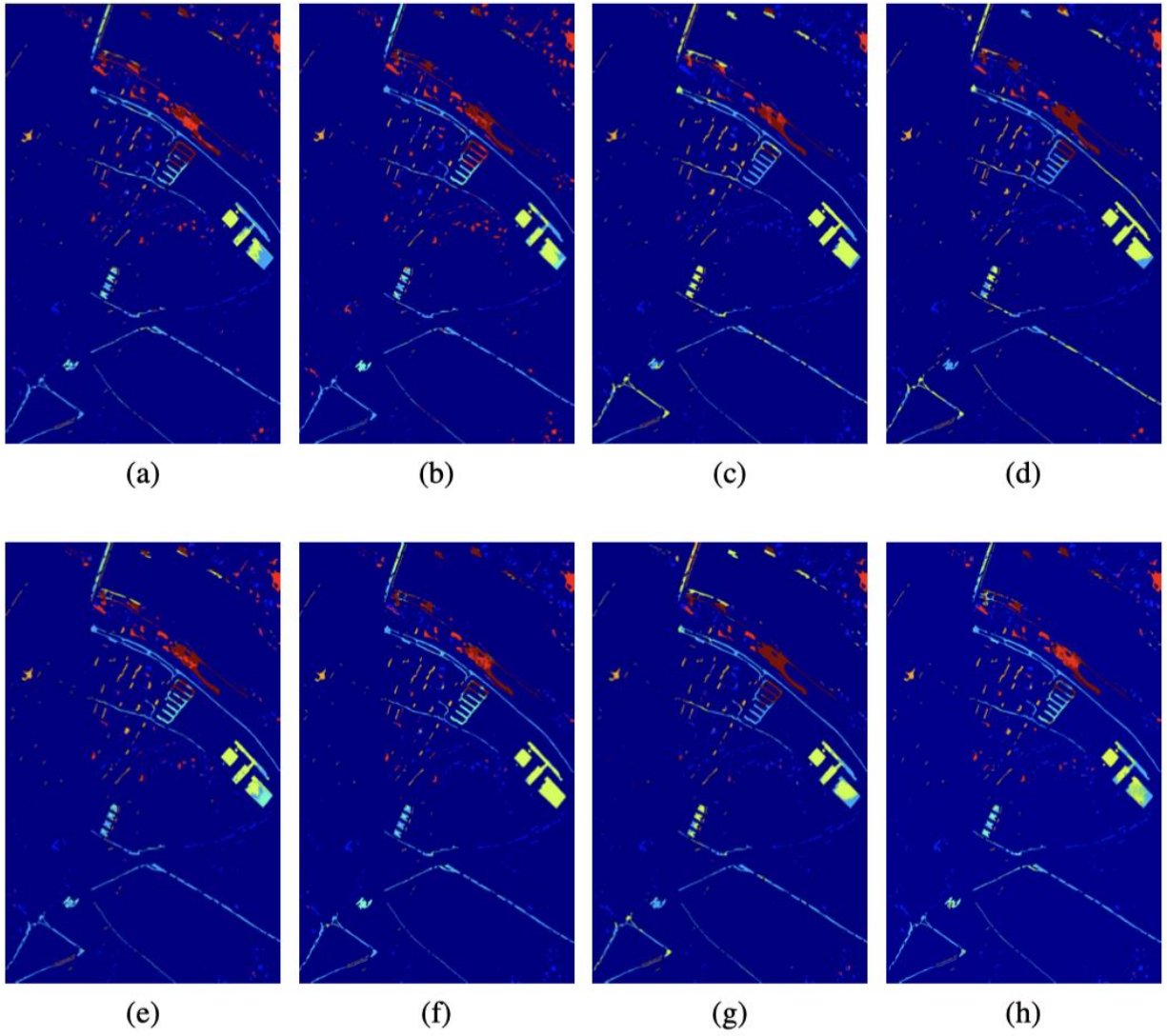
Experiments Results: HyRANK.

Table; Classification Results on Loukia Center Dataset

Class	DAN	SagNet	MRAN	TSTNet	DSAN	SDEnet	S²AMSNet	MKDnet
1	11.66	21.99	12.78	39.81	26.21	35.32	39.39	37.72
2	43.44	7.14	0	24.07	18.52	13.63	0	46.11
3	22.07	49.70	5.11	14.79	23.71	60.44	51.34	23.47
4	11.77	16.46	0	1.27	0	8.78	0	25.70
5	3.22	0.22	28.43	28.01	48.69	2.63	0	1.30
6	20.71	34.65	40.51	5.45	45.97	23.36	26.69	25.40
7	71.66	80.77	64.23	64.57	60.58	72.98	55.26	73.47
8	63.22	62.71	70.99	62.60	67.26	70.63	53.82	64.34
9	53.11	59.46	7.02	73.68	37.09	19.80	12.23	37.09
10	0	1.10	0.22	2.21	4.86	18.77	0	4.21
11	99.79	100	100	100	100	100	100	100
12	100	100	100	100	100	100	90.93	100
OA(%)	56.43	61.77	58.65	62.66	60.92	61.37	55.75	63.01
KC(x100)	44.69	52.12	49.32	54.20	52.44	52.99	45.78	54.98

Experiments Results: Pavia

- Visual results of **Houston**
  - MKDnet produces cleaner, more accurate classification maps across **Pavia**.
  - (Outperforms 7 advanced DA/DG methods.



(a) DAN. (b) MRAN. (c) SagNet. (d) TSTnet.
(e) DSAN. (f) SDEnet. (g) S^2 AMSnet. (h) MKDnet.

Experiments Results: Pavia

Table; Classification Results on Pavia Center Dataset

Class	DAN	SagNet	MRAN	TSTNet	DSAN	SDEnet	S^2AMSNet	MKDnet
1	73.79	96.34	60.16	51.82	84.39	86.32	89.42	78.48
2	78.65	56.87	78.77	79.76	76.34	80.32	77.12	81.96
3	80.77	6.31	54.75	60.41	54.77	69.53	25.36	51.10
4	66.43	79.66	67.32	31.44	74.65	70.31	70.98	35.06
5	96.23	90.40	68.63	74.64	98.66	76.32	92.36	86.18
6	73.46	50.33	86.45	86.99	74.98	75.91	42.18	77.62
7	70.13	57.99	62.31	51.09	67.36	60.34	67.62	60.14
OA(%)	73.99	69.71	68.87	69.63	72.96	74.36	72.87	74.94
KC(x100)	68.89	63.14	62.36	64.08	68.32	69.12	67.65	69.73

- MKDnet surpasses baselines by **1%–9% OA** on key target scenes on Pavia Center.

Outline

1. Introduction
2. Method
3. Experiments
4. Conclusion

Conclusion

1. Understanding the Gap

- Cross-domain HSI classification suffers from **domain shift** (sensor, time, environment).
- Existing DA/DG methods are limited:
  - Focus only on **marginal distribution** → ignore inter-class consistency.
  - Biased toward **local** (CNN) or **global** (Transformer, SSM) features → lack comprehensive extraction.
- **Gap**: Need **joint global+local features** + **robust domain alignment**.

2. Our Method (MKDnet)

- **Hybrid extractor (Mamba + CNN)** → captures **long-range dependencies + fine local details**.
- **STKD distillation** → preserves **inter-class feature consistency** with soft labels.
- **Graph alignment (GCN + GOT)** → models spectral–spatial topology, enhances alignment robustness.

Unified loss combines **feature extraction + distribution alignment + graph alignment**.

3. Experimental Results: MKDnet effectively addresses **feature extraction + domain alignment**, providing **robust cross-domain HSI classification**.