国防科技大学
National University of Defense Technology

CVPR
JUNE 3-7, 2026
DENVER
COLORADO

Local Precise Refinement: A Dual-Gated Mixture-of-Experts for Enhancing Foundation Model Generalization against Spectral Shifts

Xi Chen

https://nudt-sawlab.github.io/SpectralMoE/

1.Background & Motivation

1.1 Background

Spectral Remote Sensing Imagery： Encompassing Hyperspectral, Multispectral, and RGB Data

Hyperspectral

Bands: 32
Resolution: 10 m

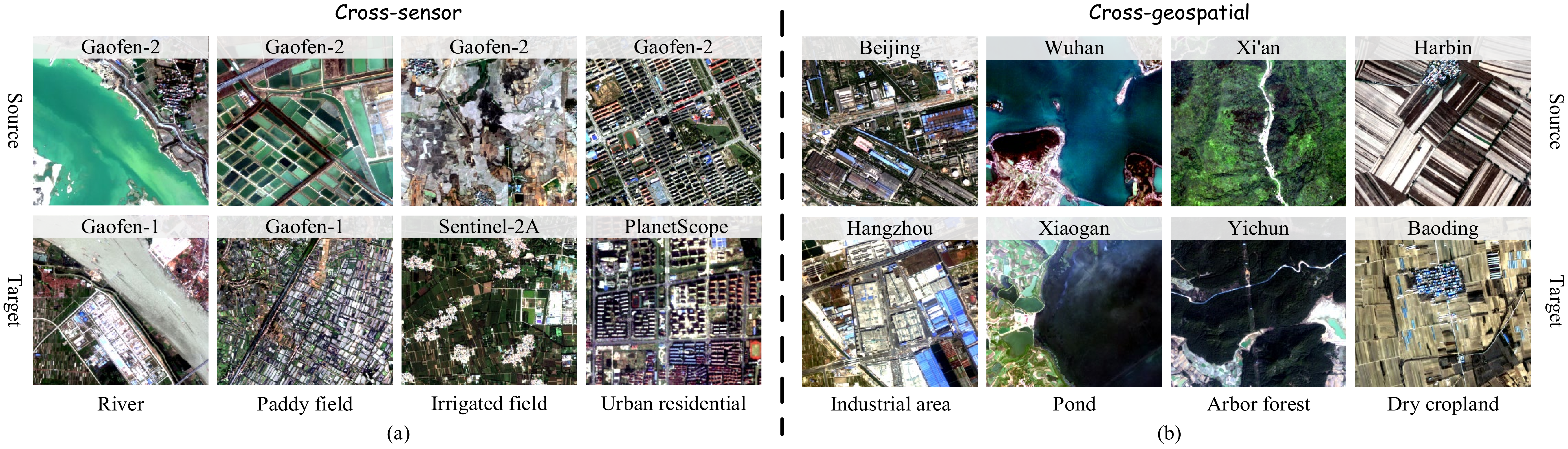
Multispectral

Bands: 4
Resolution: 4 m

RGB

Bands: 3
Resolution: 0.1 m

Spectral Shifts in Practical Mapping: Cross-Sensor and Cross-Geospatial Variations



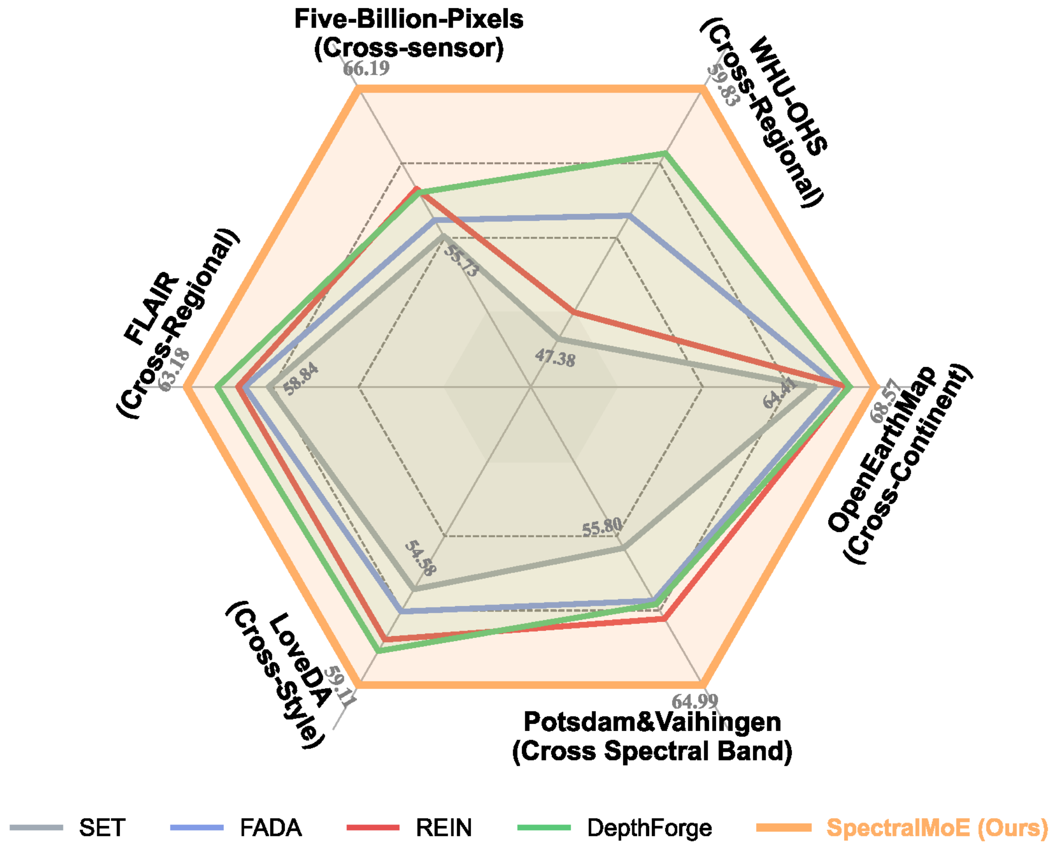
CVPR JUNE 3-7, 2026 DENVER COLORADO

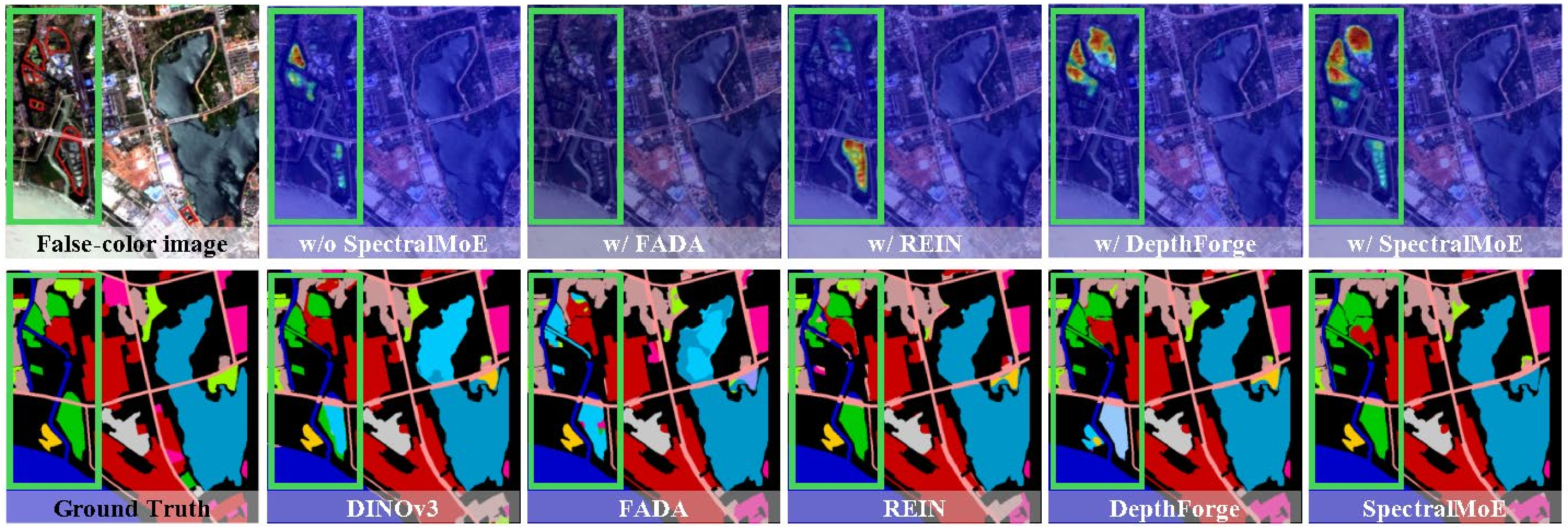
- **Most existing fine-tuning approaches rely on global, homogeneous adjustments**

Limitation:

- **High Spatial Heterogeneity:** Remote sensing images are highly heterogeneous (e.g., "paddy fields" and "ponds" are both spectrally similar and spatially adjacent).
- **The "Ripple Effect":** Global adjustments lack localized control and easily trigger unintended widespread changes, leading to Semantic Confusion.

- ***Our Paradigm Shift: Local Precise Refinement***



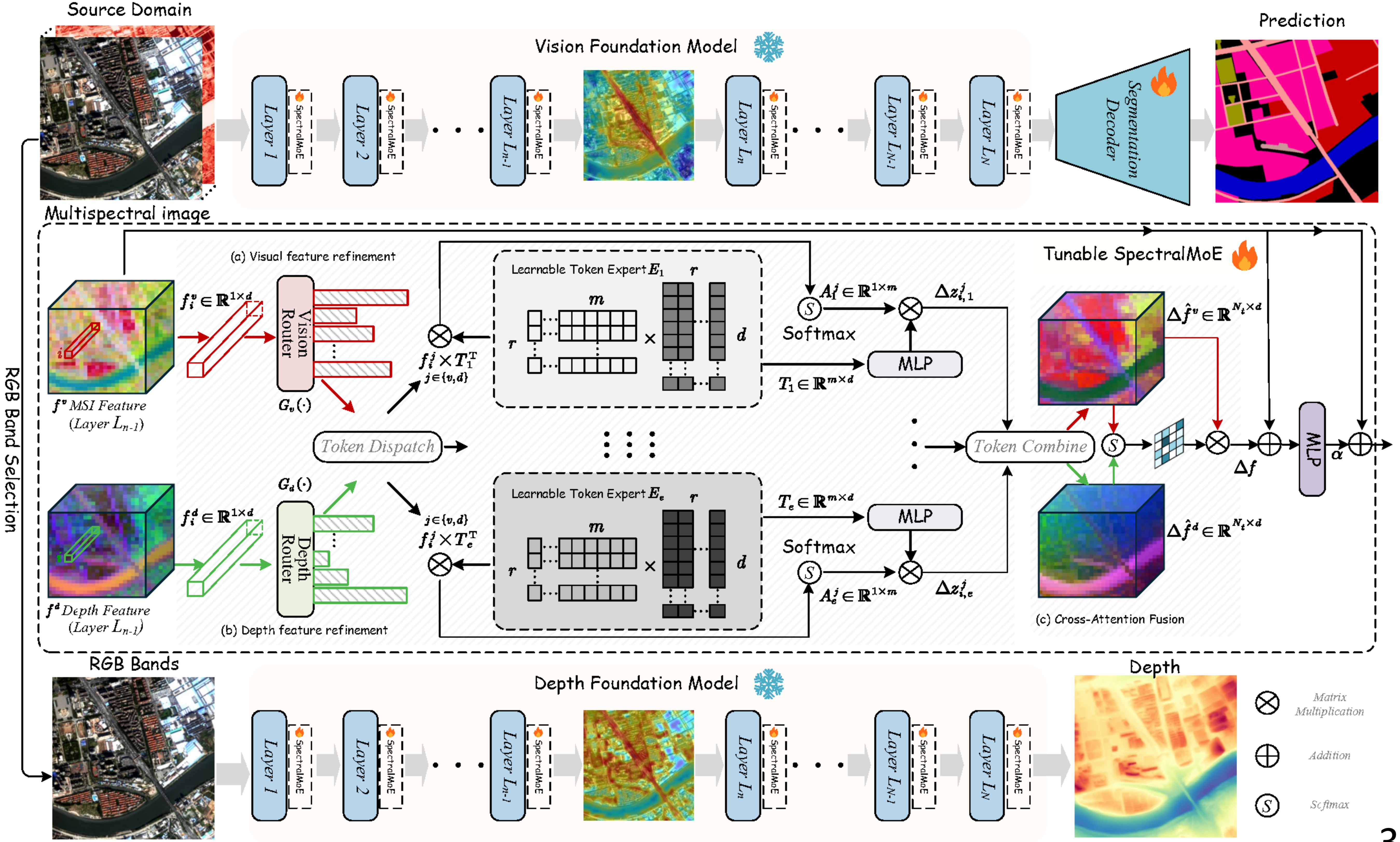


2. Methodology: SpectralMoE

2.1 SpectralMoE Architecture

SpectralMoE is inserted as a lightweight plugin into each layer of frozen VFMs and DFMs.

- SpectralMoE, a fine-tuning framework for spectral RS DGSS across hyperspectral, multispectral, and RGB remote sensing imagery.
- A dual-gated MoE that independently routes visual and depth tokens to expert adapters for local precise refinement.
- A cross-attention fusion module that lets visual adjustments query depthderived structural priors, mitigating semantic ambiguity from spectral similarity.



3. Experiments & Results

3.1 Experimental Setup & Cross-Domain Benchmarks

Evaluating 7 DGSS Tasks across 3 modalities (Hyperspectral, Multispectral, RGB).

Generalization Scenarios: Cross-Sensor, Cross-Region, Cross-Style, etc.

1. Hyperspectral Imagery (HSI)

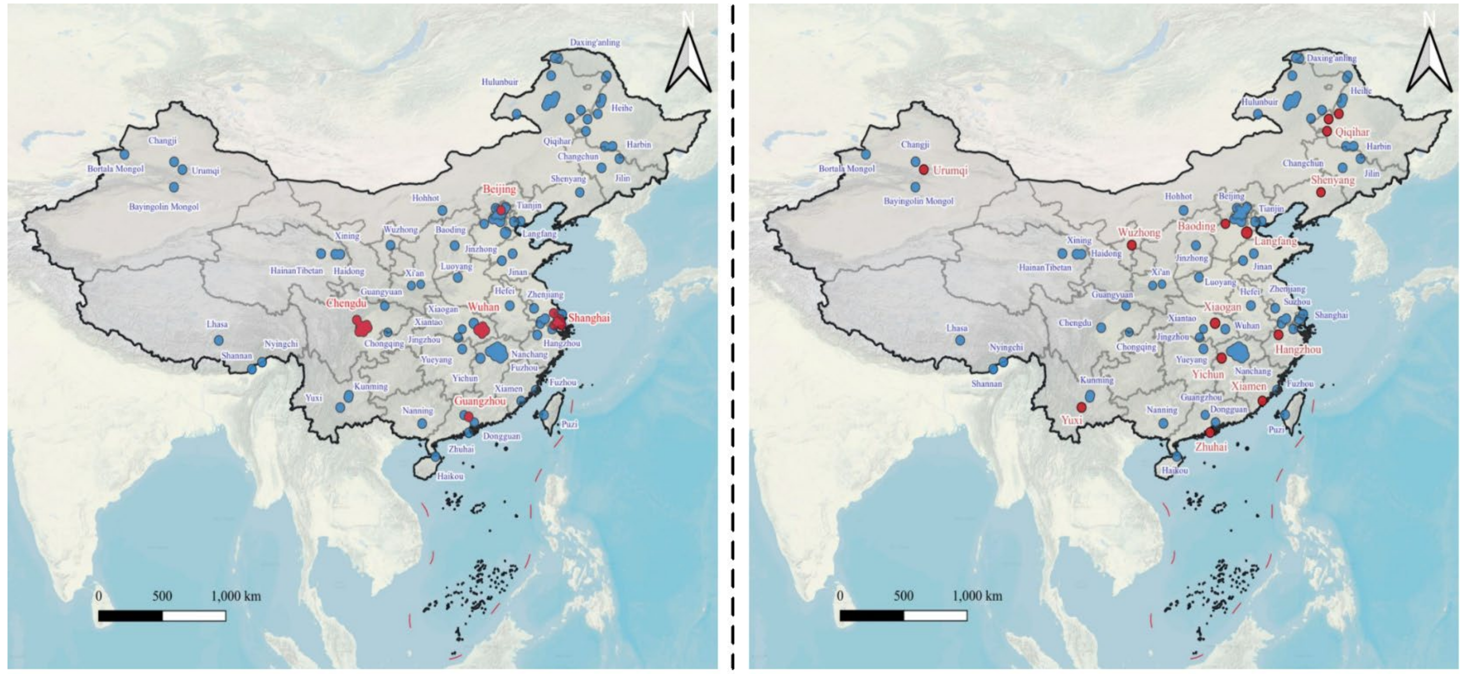
- Cross-Region: WHU-OHS dataset (covering 40 diverse locations across China).

2. Multispectral Imagery (MSI)

- Cross-Sensor: Five-Billion-Pixels (Source: GF-2 ➔ Target: PlanetScope / GF-1 / Sentinel-2).
- Cross-Region (China): Five-Billion-Pixels (Source: 50 regions ➔ Target: 12 unseen cities).
- Cross-Region (Europe): FLAIR (covering 50 distinct spatial domains in France).

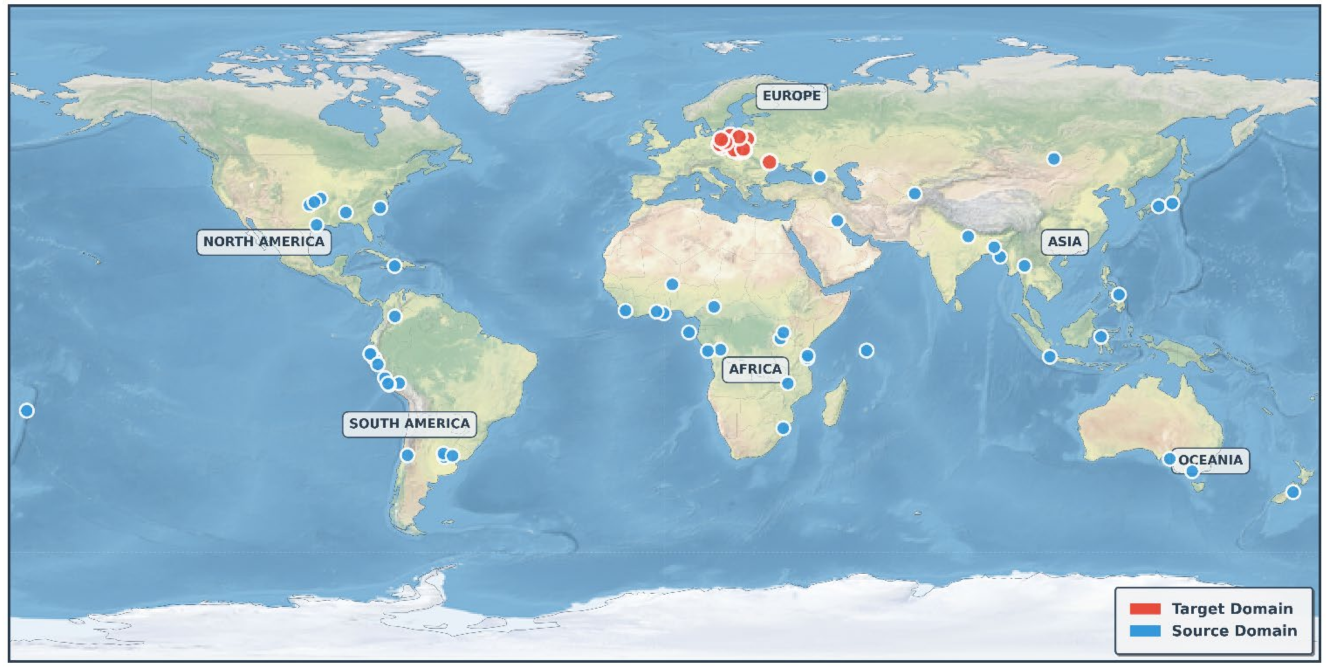
3. RGB Imagery

- Cross-Style (Rural ➔ Urban): LoveDA dataset.
- Cross-Spectral-Band: Potsdam (RGB) ➔ Vaihingen (NIR-R-G).
- Cross-Continent: OpenEarthMap (Source: 5 Continents ➔ Target: Europe).



(a) Five-Billion-Pixels (Cross-sensor)

(b) Five-Billion-Pixels (Cross-Regional)



3.2 Quantitative Results: Outperforming SOTA

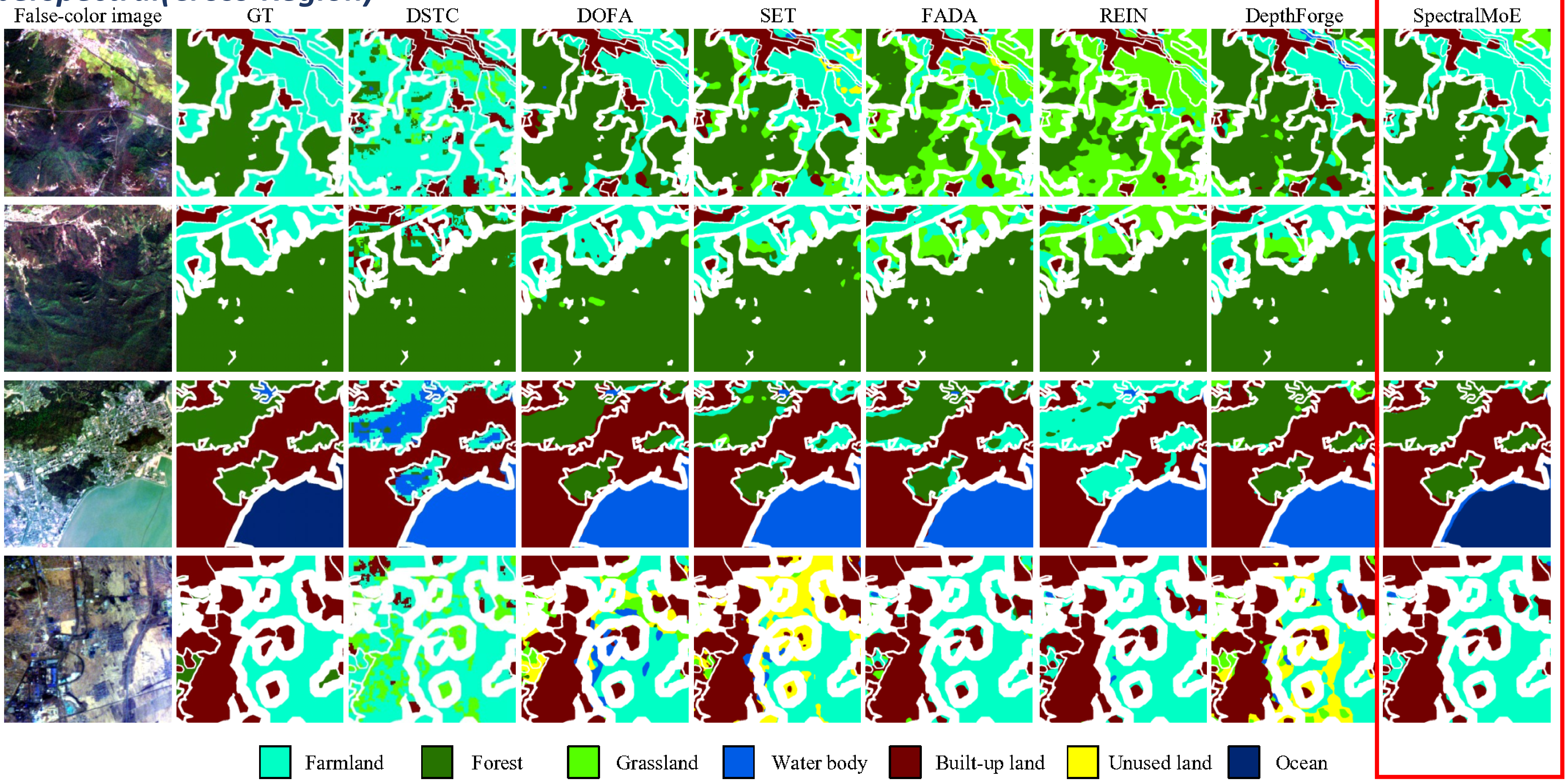
CVPR JUNE 3-7, 2026 DENVER COLORADO

- ***Boosts mIoU significantly on leading foundation models (DINOv3, DOFA).***
- ***Breakthrough performance on the Five-Billion-Pixels (Cross-Sensor) benchmark.***

Method	*Hyperspectral DGSS Tasks*		*Multispectral DGSS Tasks*						*RGB DGSS Tasks*					
	WHU-OHS (Cross-Regional)		Five-Billion-Pixels (Cross-sensor)		Five-Billion-Pixels (Cross-Regional)		FLAIR (Cross-Regional)		LoveDA (Cross-Style)		Potsdam&Vaihingen (Cross Spectral Band)		OpenEarthMap (Cross-Continent)	
	mIoU	mAcc	mIoU	mAcc	mIoU	mAcc	mIoU	mAcc	mIoU	mAcc	mIoU	mAcc	mIoU	mAcc
Spectral Remote Sensing Domain Generalized Semantic Segmentation Models														
DSTC [31]	43.09	57.81	29.79	40.61	46.27	59.71	53.42	68.05	47.78	62.90	16.62	32.54	46.49	63.17
Remote Sensing Foundation Models (Freeze Backbone)														
SoftCon [44]	–	–	29.59	40.81	43.24	58.65	53.21	68.27	41.01	54.91	11.27	24.05	39.63	60.15
Galileo [42]	–	–	31.68	43.36	38.54	52.63	52.07	67.58	32.45	45.49	11.69	27.12	35.95	57.12
SenPaMAE [37]	45.76	59.01	32.28	45.50	43.31	58.94	55.48	70.34	38.71	53.10	19.60	34.40	44.13	60.56
Copernicus [45]	42.87	56.83	33.84	42.44	43.56	60.35	56.15	71.84	43.96	58.16	15.91	30.76	48.37	69.16
DOFA [49]	46.03	59.73	44.60	58.66	49.48	67.38	58.25	72.67	52.41	67.54	38.10	58.74	58.49	73.70
Visual Foundation Models (Freeze Backbone)														
CLIP [38]	–	–	46.30	61.59	48.10	64.51	56.81	70.24	51.62	66.95	40.84	64.17	61.93	78.12
SAM [24]	–	–	44.98	59.35	47.14	64.68	56.65	71.92	52.17	66.82	40.48	63.57	56.45	73.32
EVA02 [15, 16]	–	–	45.48	59.11	49.33	62.66	57.52	71.59	53.04	65.70	42.81	65.27	61.83	75.45
DINOv2 [34]	–	–	53.37	69.10	52.48	69.92	58.38	73.03	54.10	68.39	55.39	77.68	64.37	76.79
DINOv3 [39]	–	–	55.54	69.18	54.44	70.71	59.60	73.40	55.75	74.83	58.79	80.27	65.48	77.02
Foundation model-based DGSS Models														
SET [52]	47.38	60.61	55.73	70.24	55.61	70.98	58.84	73.23	54.58	72.38	55.80	78.85	64.41	78.60
FADA [1]	53.51	67.54	56.84	73.44	55.35	71.99	60.12	74.11	55.63	70.85	59.33	80.11	66.08	78.87
REIN [47]	48.71	63.58	59.06	73.44	55.27	71.87	60.46	73.76	56.96	71.37	60.54	81.36	66.76	78.81
DepthForge [5]	56.61	69.76	58.79	75.08	54.92	71.43	61.56	74.14	57.50	71.92	59.57	81.51	66.85	79.70
Ours														
SpectralMoE	**59.83**	**73.13**	**66.19**	**77.26**	**60.32**	**75.78**	**63.18**	**76.52**	**59.11**	**75.38**	**64.99**	**85.43**	**68.57**	**80.17**

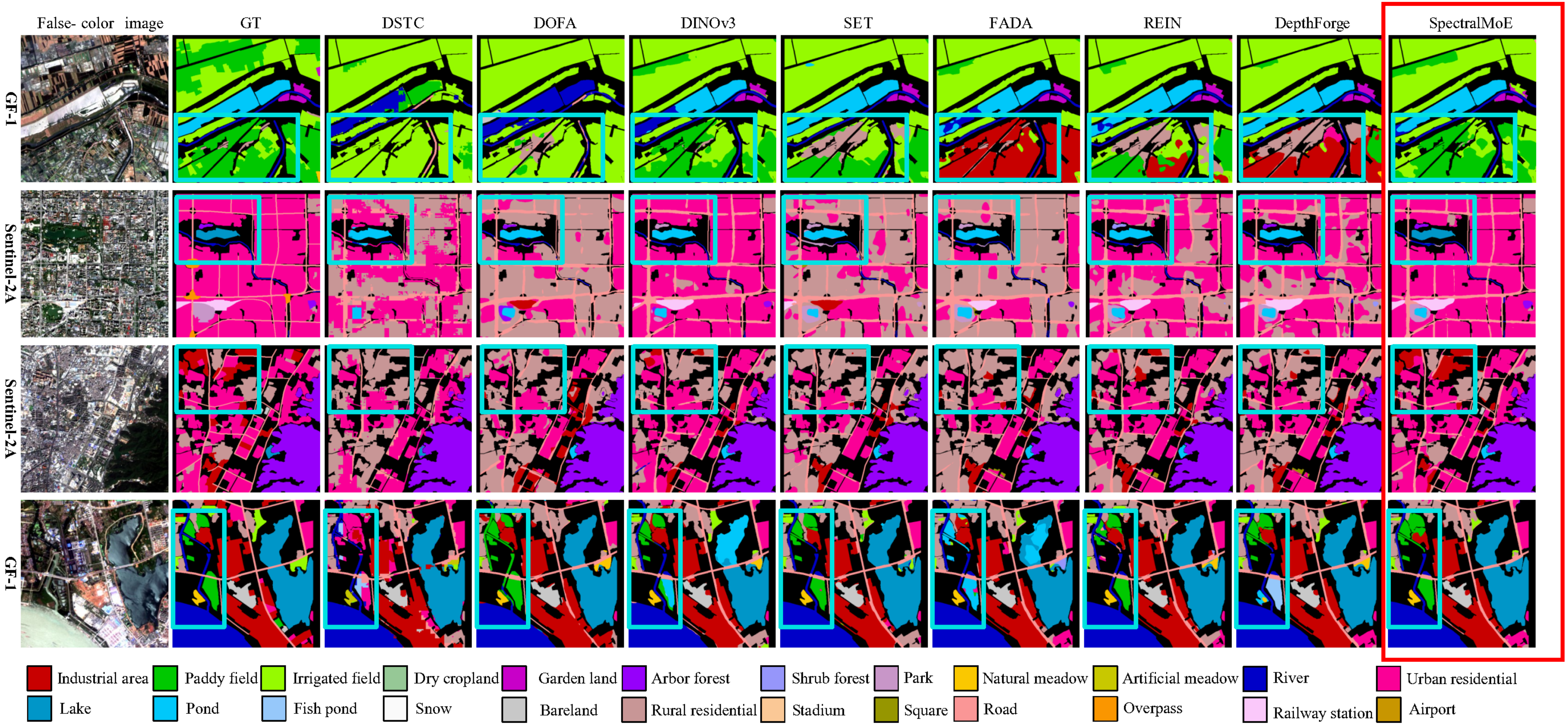
3.3 Qualitative Results: Finer Local Details

Hyperspectral(Cross-Region)



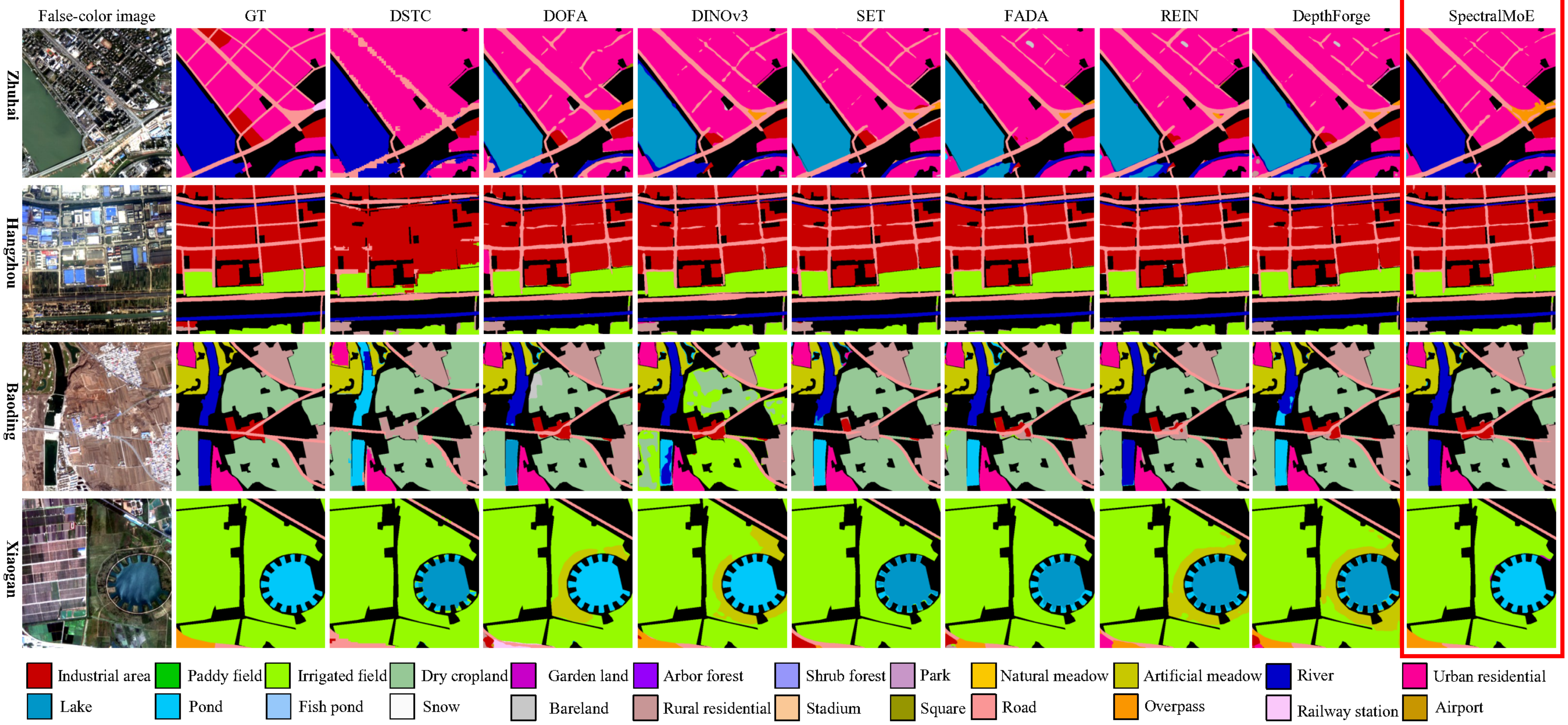
3.3 Qualitative Results: Finer Local Details

Multispectral(Cross-Sensor)



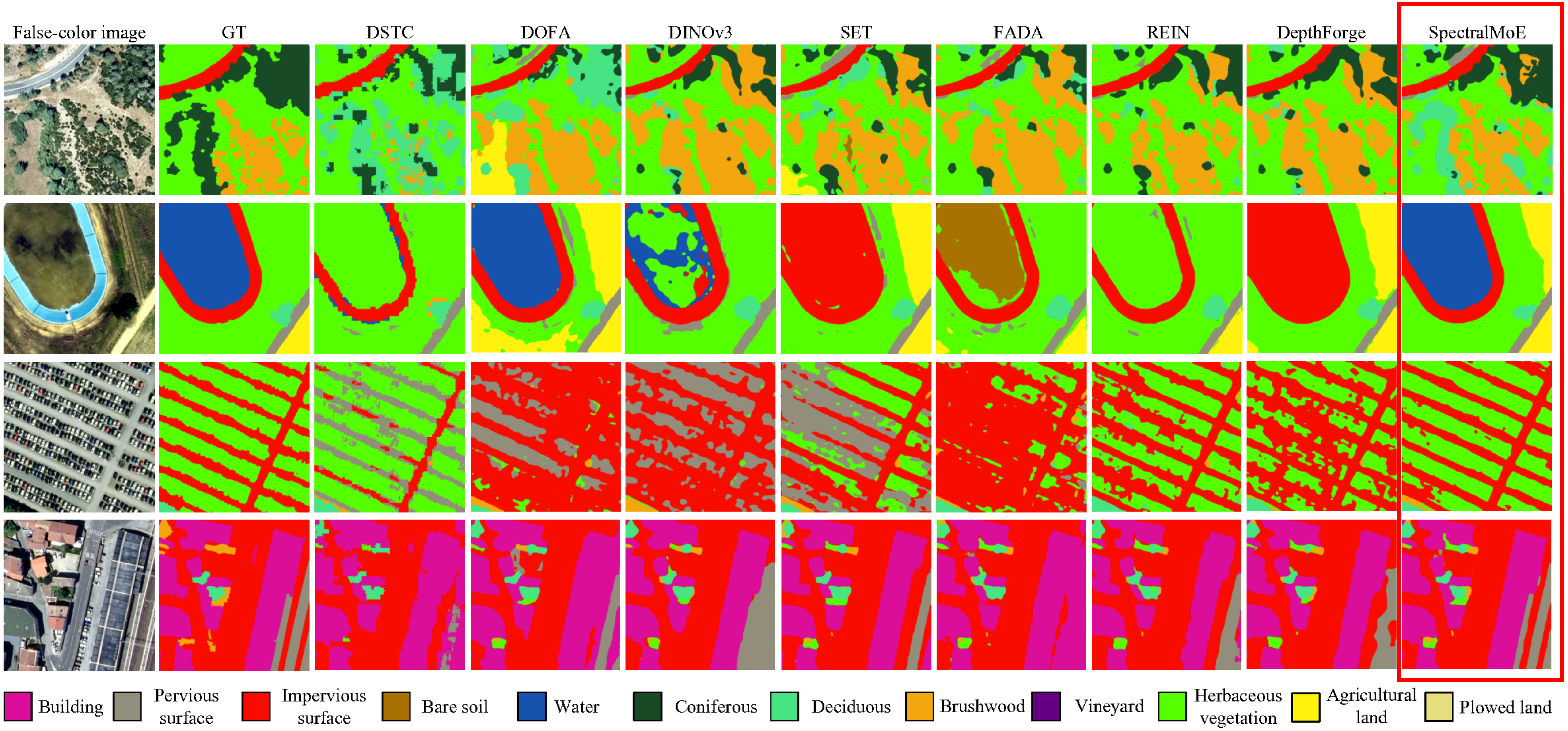
3.3 Qualitative Results: Finer Local Details

Multispectral(Cross-Region)



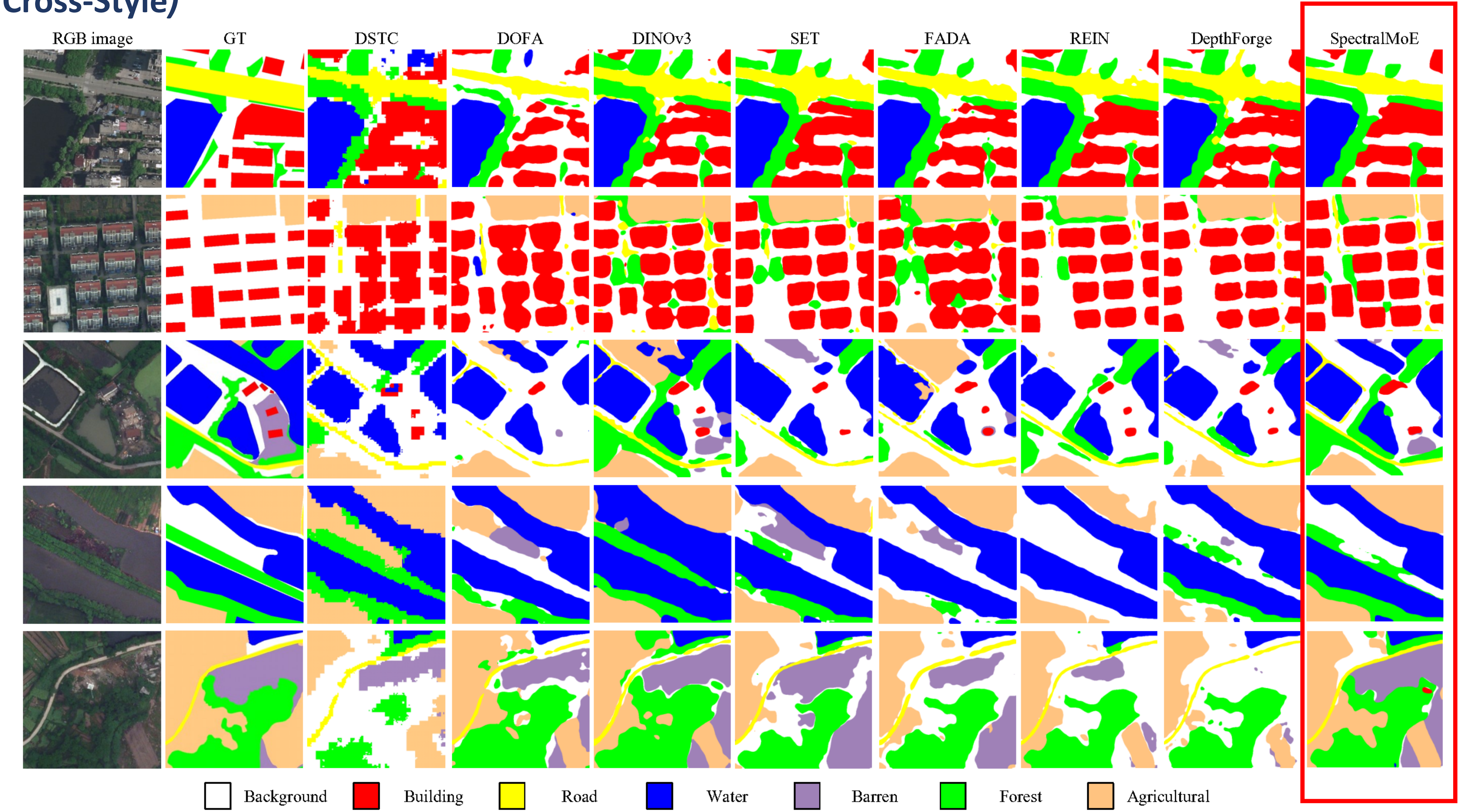
3.3 Qualitative Results: Finer Local Details

Multispectral(Cross-Region)



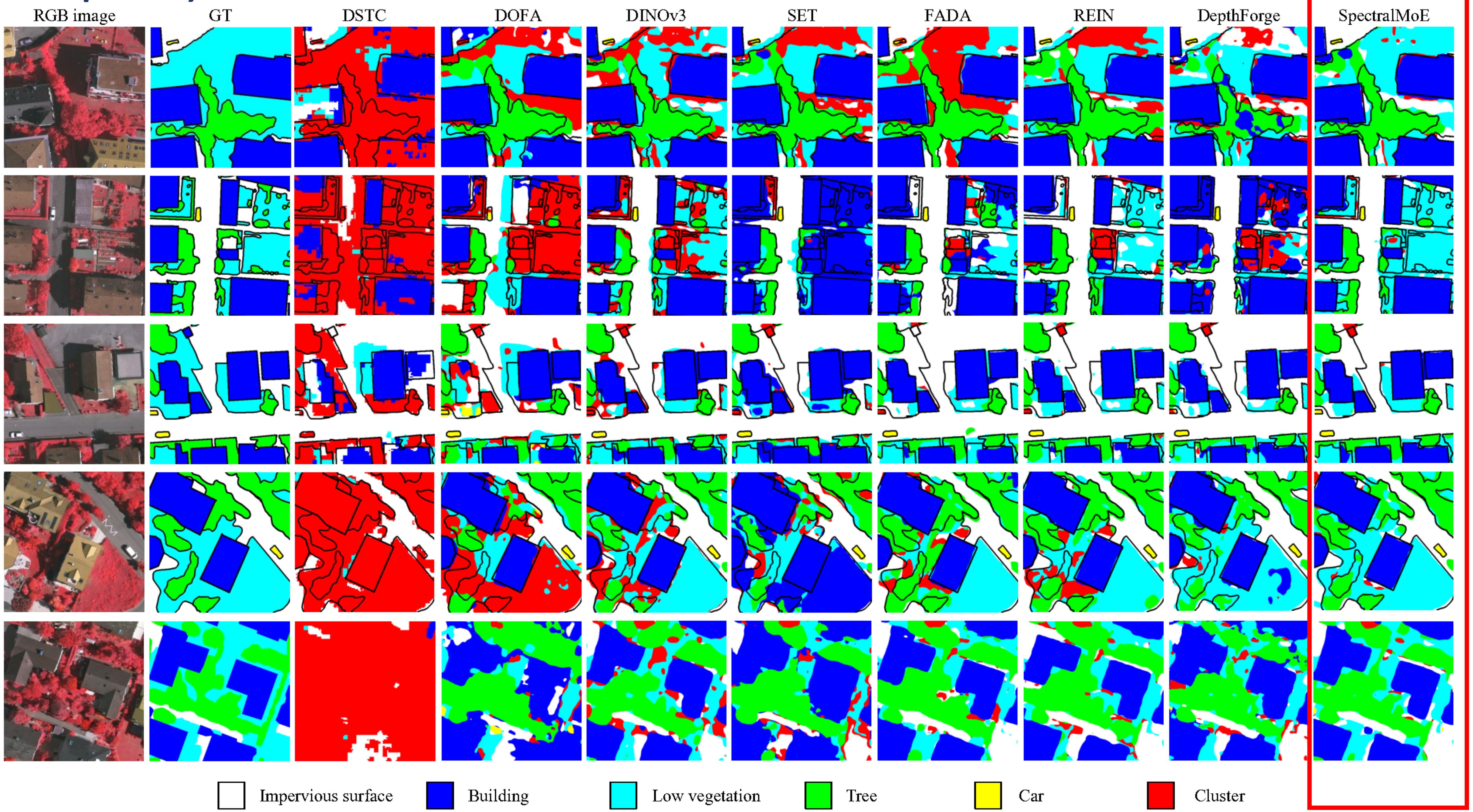
3.3 Qualitative Results: Finer Local Details

RGB(Cross-Style)



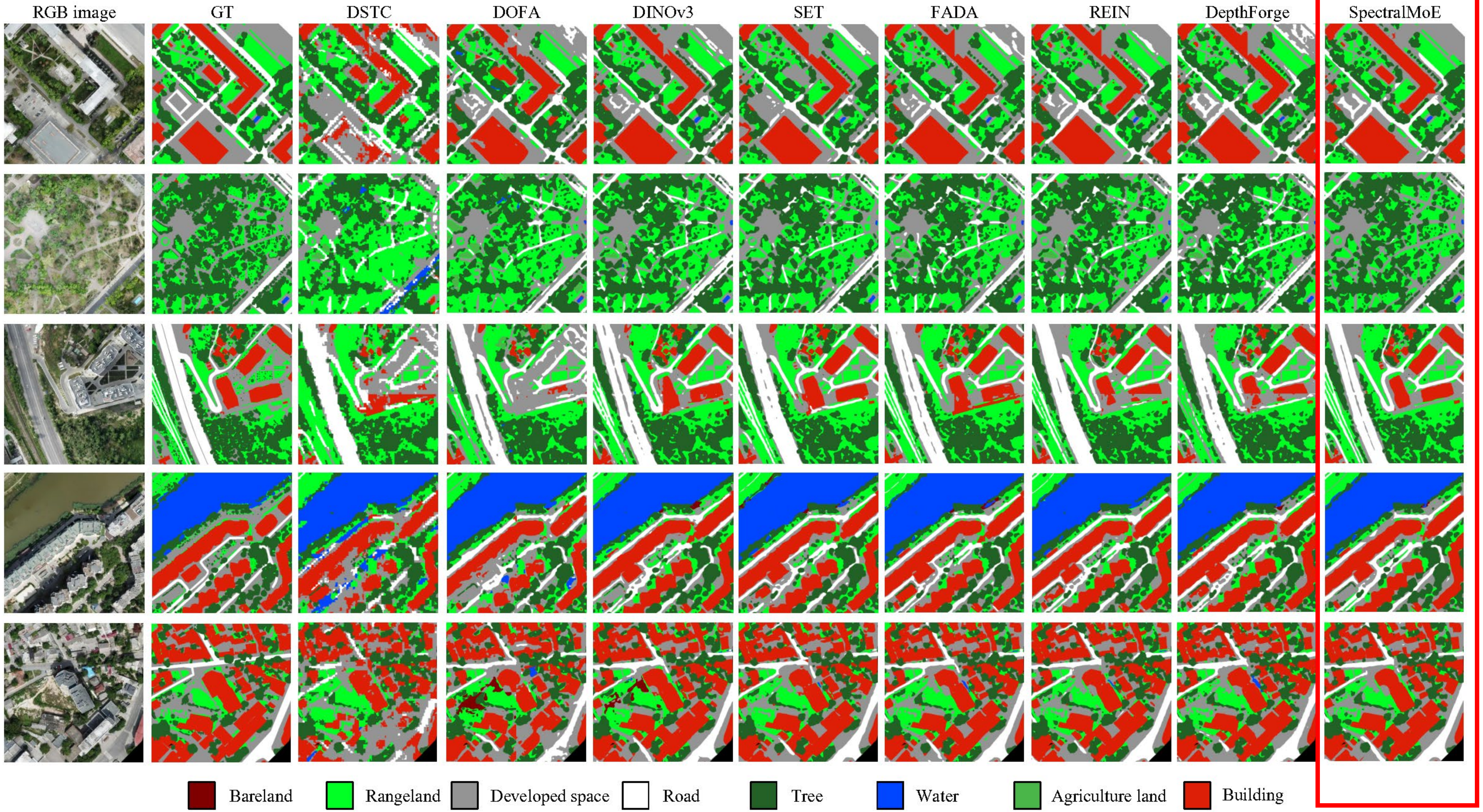
CVPR JUNE 3-7, 2026 DENVER COLORADO

*RGB(*Cross-Spectral*)*



3.3 Qualitative Results: Finer Local Details

RGB(Cross-Continent)



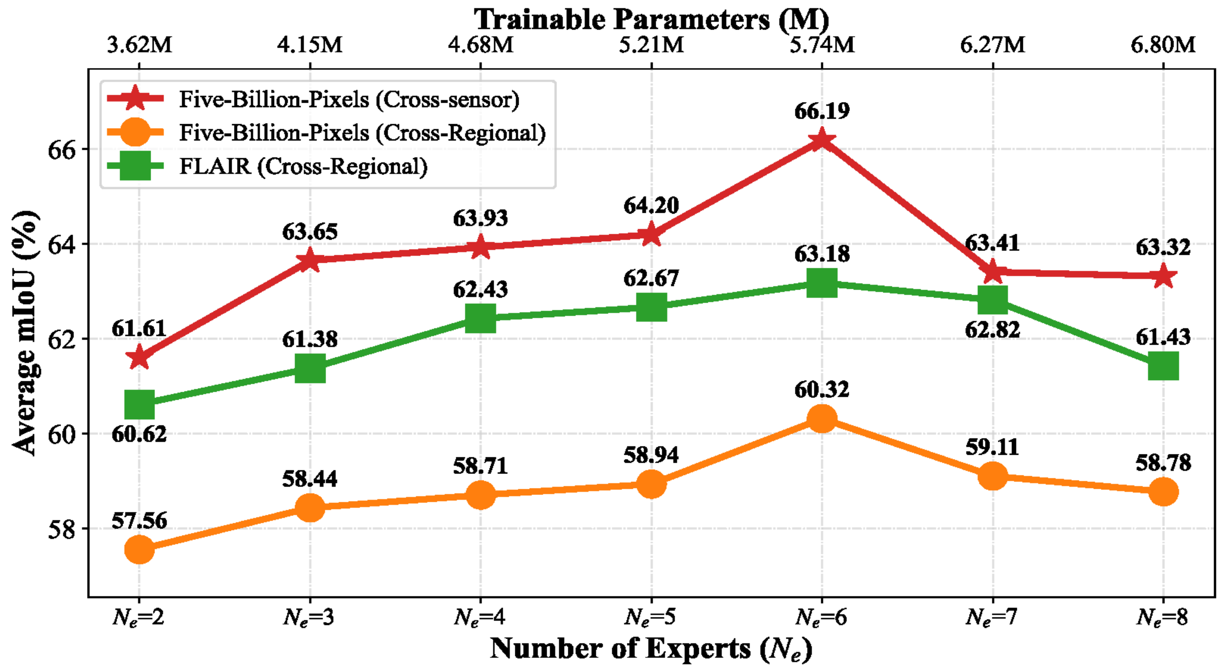
4. Ablation Studies

CVPR JUNE 3-7, 2026 DENVER COLORADO

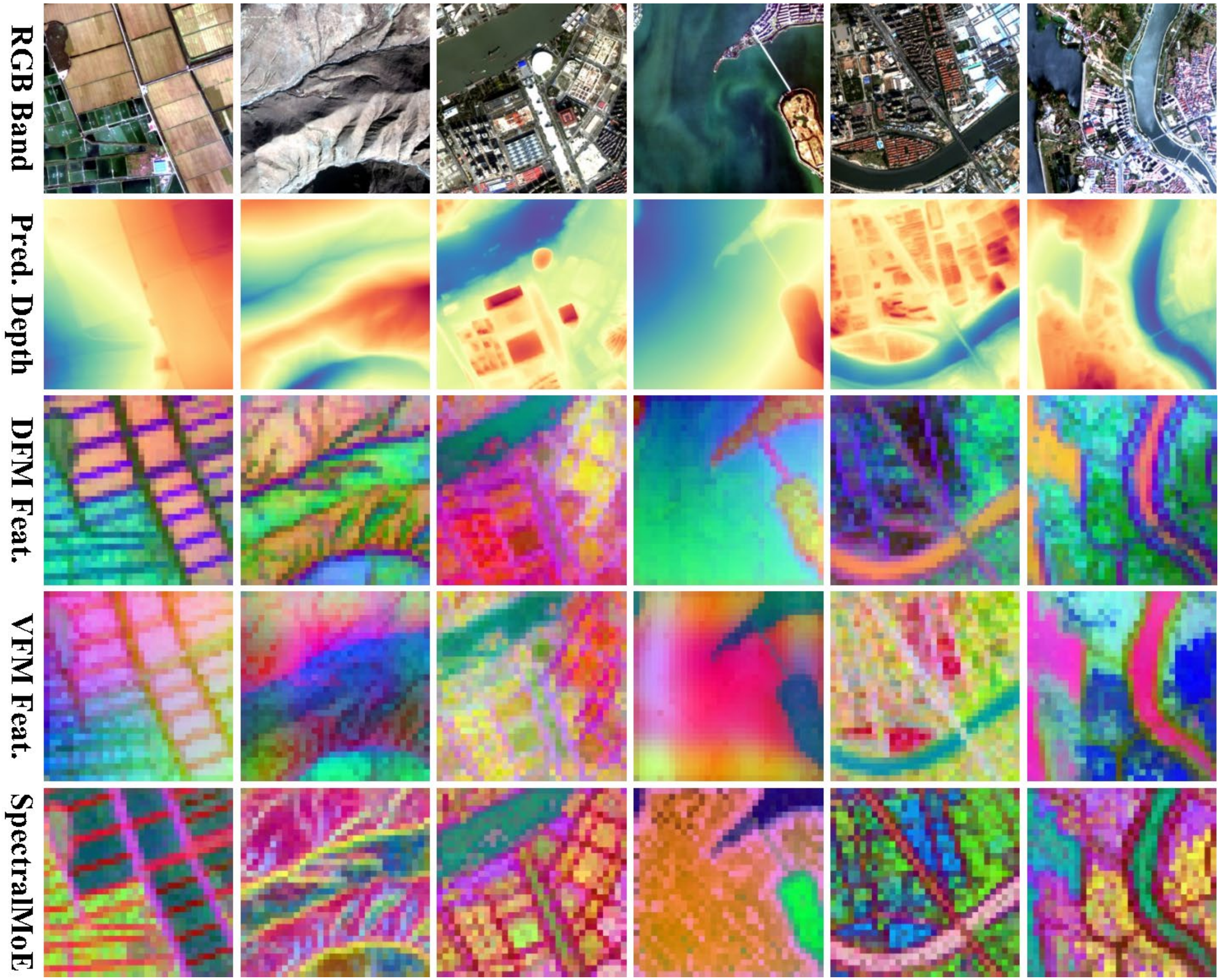
Effectiveness of Key Modules

Backbone	Configuration	mIoU (%) Five-Billion-Pixels (Cross-Sensor)	mIoU (%) Five-Billion-Pixels (Cross-Regional)	mIoU (%) FLAIR (Cross-Regional)
	w/o MoE	63.41	59.37	60.82
VFM	w/o Dual Gating	63.44	58.97	61.95
DINOv3	w/o Depth Feature	63.52	59.01	63.07
(Large) [39]	w/o Cross-Attention	63.48	58.35	61.30
	SpectralMoE (ours)	**66.19**	**60.32**	**63.18**
	w/o MoE	55.69	54.58	59.94
RSFM	w/o Dual Gating	53.66	54.10	60.18
DOFA	w/o Depth Feature	54.71	54.44	61.31
(Large) [49]	w/o Cross-Attention	54.20	53.17	60.10
	SpectralMoE (ours)	**57.73**	**55.41**	**61.50**

Impact of the Number of Experts



Superiority of Fine-Tuning Foundation Model Features



4.2 Versatility & Superior Performance

CVPR JUNE 3-7, 2026 DENVER COLORADO

Adaptability Across Foundation Models

VFM	Method	Params (M)	mIoU (%) FBP (CS)	FBP (CR)	FLAIR (CR)	Avg.
CLIP (Large) [38]	Freeze	0.00	46.30	48.10	56.81	50.40
	+ SET [52]	6.13	55.00	52.62	59.88	55.83
	+ FADA [1]	11.65	55.03	51.86	60.52	55.80
	+ REIN [47]	2.99	56.70	53.37	60.97	57.01
	+ DepthForge [5]	2.99	54.77	53.89	61.17	56.61
	+ SpectralMoE	5.74	**61.33**	**54.13**	**61.47**	**58.98**
SAM (Huge) [24]	Freeze	0.00	44.98	47.14	56.65	49.59
	+ SET [52]	9.21	50.31	52.10	58.17	53.53
	+ FADA [1]	16.59	49.74	50.04	59.38	53.05
	+ REIN [47]	4.51	50.93	50.27	59.36	53.52
	+ DepthForge [5]	4.51	48.15	52.08	59.26	53.16
	+ SpectralMoE	10.29	**59.16**	**56.79**	**59.41**	**58.45**
EVA02 (Large) [15, 16]	Freeze	0.00	45.48	49.33	57.52	50.78
	+ SET [52]	6.13	46.29	51.57	50.38	49.41
	+ FADA [1]	11.65	46.49	51.27	59.37	52.38
	+ REIN [47]	2.99	50.81	51.33	59.46	53.87
	+ DepthForge [5]	2.99	40.90	47.60	58.81	49.10
	+ SpectralMoE	5.74	**60.20**	**56.29**	**62.09**	**59.53**
DINOv2 (Large) [34]	Freeze	0.00	53.37	52.48	58.38	54.74
	+ SET [52]	6.13	55.73	55.61	58.84	56.73
	+ FADA [1]	11.65	56.84	55.35	60.12	57.44
	+ REIN [47]	2.99	59.06	55.27	60.46	58.26
	+ DepthForge [5]	2.99	58.79	54.92	61.56	58.42
	+ SpectralMoE	5.74	**63.77**	**57.77**	**62.62**	**61.39**

Outperforms State-of-the-Art PEFT Methods

Dataset	Method	Params (M)	DINOv3 Large [39] (VFM) mIoU	mAcc	DOFA Large [49] (RSFM) mIoU	mAcc
FBP (CS)	Freeze	0.00	53.37	69.10	46.98	62.83
	+ LoRA [21]	0.79	59.64	74.17	49.49	62.12
	+ AdaptFormer [4]	3.22	57.73	73.22	49.88	65.89
	+ VPT [22]	3.69	58.14	72.50	49.70	60.96
	+ SET	6.13	58.02	72.73	49.28	61.51
	+ FADA	11.65	58.27	73.72	53.93	67.32
	+ REIN	2.99	61.79	76.10	52.72	64.50
	+ DepthForge	2.99	62.16	74.91	52.76	65.74
	+ SpectralMoE	5.74	**66.19**	**77.26**	**57.73**	**70.37**
FBP (CR)	Freeze	0.00	52.48	69.92	49.16	65.41
	+ LoRA [21]	0.79	56.15	73.33	50.60	65.09
	+ AdaptFormer [4]	3.22	55.82	74.89	50.32	64.99
	+ VPT [22]	3.69	56.04	73.75	49.43	68.60
	+ SET	6.13	56.06	72.53	50.11	64.64
	+ FADA	11.65	56.25	72.99	54.20	70.83
	+ REIN	2.99	57.09	74.00	53.74	70.88
	+ DepthForge	2.99	57.54	72.14	53.95	69.33
	+ SpectralMoE	5.74	**60.32**	**75.78**	**55.41**	**72.71**
FLAIR (CR)	Freeze	0.00	58.38	73.03	58.25	72.67
	+ LoRA [21]	0.79	61.12	75.14	60.31	74.03
	+ AdaptFormer [4]	3.22	60.19	72.76	59.68	74.19
	+ VPT [22]	3.69	60.16	74.11	59.47	73.36
	+ SET	6.13	60.32	72.35	59.93	74.79
	+ FADA	11.65	61.61	76.31	60.15	74.54
	+ REIN	2.99	62.79	76.49	60.94	74.90
	+ DepthForge	2.99	62.47	76.68	60.52	74.57
	+ SpectralMoE	5.74	**63.18**	**76.52**	**61.50**	**75.43**

Thank You for Your Attention!

Xi Chen

xi_chen@nudt.edu.cn