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

Supplementary Material

7. Construction Details of DGSS Tasks for Spectral Remote Sensing

7.1. Data Sources

Our experiments address DGSS across three spectral RS modalities: hyperspectral, multispectral, and RGB.

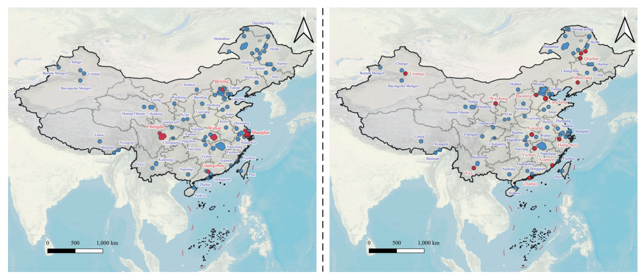
For the hyperspectral DGSS task, we utilize Orbita hyperspectral satellite imagery from the WHU-OHS dataset [28], covering 40 locations across China. These hyperspectral images (HSIs) span a spectral range of 400–1000 nm with an average spectral resolution of 2.5 nm and a spatial resolution of 10 m, comprising 32 spectral bands.

For the multispectral DGSS task, we use multispectral images (MSIs) from four satellite platforms (GF-2, PlanetScope, GF-1, and Sentinel-2) and aerial imagery from the FLAIR dataset.

- **GF-2**, the second satellite of the High-Definition Earth Observation System (HDEOS) led by the China National Space Administration (CNSA), provides MSIs with a 4 m spatial resolution. Its sensor captures four bands: blue (0.45–0.52 μm), green (0.52–0.59 μm), red (0.63–0.69 μm), and near-infrared (0.77–0.89 μm) [42].
- **PlanetScope**, operated by Planet Labs (USA), is a constellation of approximately 130 CubeSats. It delivers MSIs at a 3.7–4.1 m spatial resolution, with bands covering blue (0.46–0.52 μm), green (0.50–0.59 μm), red (0.59–0.67 μm), and near-infrared (0.78–0.86 μm) [42].
- **GF-1**, the first satellite of China's HDEOS program, carries a multispectral sensor with an 8 m spatial resolution. Its spectral band ranges are identical to those of GF-2: blue (0.45–0.52 μm), green (0.52–0.59 μm), red (0.63–0.69 μm), and near-infrared (0.77–0.89 μm) [41].
- **Sentinel-2**, from the European Union's Copernicus programme, provides MSIs. We selected its 10 m resolution bands for our experiments: blue (central wavelength 0.49 μm, Band 2), green (central wavelength 0.56 μm, Band 3), red (central wavelength 0.66 μm, Band 4), and near-infrared (central wavelength 0.83 μm, Band 8) [41].
- **FLAIR** dataset provides high-resolution (0.2 m) aerial MSIs captured by Vexcel's Ultracam Eagle Mark3 and IGN's CAMv2. These MSIs include blue, green, red, and near-infrared bands, cover 50 spatial domains in France, and were acquired between April and November from 2018 to 2021 [18].

For the RGB RS DGSS task, we utilize the following datasets:

- **LoveDA** [44] is constructed from 0.3 m resolution imagery sourced from the Google Earth platform. Following its official rural-to-urban generalization setup, we establish a rural-to-urban (cross-style) task. For this, we use images from the Pukou, Lishui, Gaochun, and Jiangxia regions as the source domain (rural), and images from the Yuhuatai and Jintan regions as the target domain (urban).
- **Potsdam and Vaihingen** are used to establish the cross-spectral-band task. This task involves training on the RGB bands of Potsdam and generalizing to the Near-Infrared, Red, and Green (NIR-R-G) bands of Vaihingen.
- **OpenEarthMap** [49] is used to construct the cross-continent generalization task, utilizing its built-in geographical splits.



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

Figure 6. **Geographical distribution of source and target domains for the constructed Five-Billion-Pixels (cross-sensor) and Five-Billion-Pixels (cross-regional) generalization tasks.** Subfigure (a) illustrates the domain distribution for the **Five-Billion-Pixels (cross-sensor)** task, where locations corresponding to the source domain (GF-2 MSIs) are marked by solid blue circles, and those for the target domains (PlanetScope, GF-1, and Sentinel-2 MSIs) are indicated by solid red circles. Subfigure (b) shows the domain distribution for the **Five-Billion-Pixels (cross-regional)** task, with the source domain (GF-2 MSIs from various regions) represented by solid blue circles and the target domain (GF-2 MSIs from designated cities) denoted by solid red circles.

7.2. Details of Five-Billion-Pixels (Cross-sensor) Multispectral DGSS Task

In the **Five-Billion-Pixels (Cross-sensor)** multispectral DGSS task, we select GF-2 MSIs from the Five-Billion-Pixels [42] dataset as the source domain training data. MSIs acquired by PlanetScope, GF-1, and Sentinel-2 serve as the target domains to evaluate the model's generalization performance under sensor shifts. As illustrated in Fig. 6(a), the city distribution of the source domain (GF-2 imagery) is marked by solid blue circles, whereas the city distributions for the target domain data (PlanetScope, GF-1, and Sentinel-2) are indicated by red circles. Specifically, the PlanetScope target domain data covers the cities of Chengdu and Shanghai; the GF-1 target domain data is sourced from Wuhan; and the Sentinel-2 target domain data is collected from Beijing and Guangzhou.

7.3. Details of Five-Billion-Pixels (Cross-Regional) Multispectral DGSS Task

The **Five-Billion-Pixels (Cross-Regional)** multispectral DGSS task is constructed based on 150 GF-2 MSIs sourced from 62 administrative regions across China, as provided in the Five-Billion-Pixels [42] dataset. We designate the images from the administrative regions of Yichun, Baoding, Xiaogan, Langfang, Yuxi, Qiqihar, Hangzhou, Zhuhai, Urumqi, Wuzhong, Xiamen, and Shenyang as the target domain. Their geographical distribution is indicated by solid red circles in Fig. 6(b). The remaining MSIs, from administrative regions geographically distinct from the target domain locations, are utilized as the source domain training data. Their distributions are represented by solid blue circles in Fig. 6(b).

7.4. Details of OpenEarthMap(Cross-Continent) RGB RS DGSS Task

The **OpenEarthMap (Cross-Continent)** RGB RS DGSS task is constructed using the OpenEarthMap dataset [49], which covers 44 countries across 6 continents. For this task, we designate the images from North America, South America, Africa, Asia, and Oceania as the source domain. The images from Europe are set as the target domain to evaluate the model's generalization capability on a global scale.

8. Additional Parameter Analysis

8.1. Impact of the Gating Function in SpectralMoE

We analyze the choice of the gating function, a critical component of our Dual-Gated Network (Section 3, Equation (1)). We compare three strategies: standard **Softmax Gating**, **Gaussian Gating** (our distance-based function with L_2-norm, $p = 2$), and our adopted **Laplacian Gating** (L_1-norm, $p = 1$).

As shown in Tab. 5, we evaluate these strategies on both a state-of-the-art VFM (DINOv3) and RSFM (DOFA). The results show that the performance of standard Softmax and Gaussian Gating (L_2) is mixed and inconsistent across backbones. For instance, while Gaussian Gating (L_2) is superior on the VFM, it underperforms Softmax on the RSFM's Cross-Sensor task (54.80 vs. 56.11). In stark contrast, Laplacian Gating (L_1) consistently yields the best mIoU across all benchmarks and backbones, outperforming both alternatives. This validates its selection as the optimal choice for SpectralMoE, as its superior robustness to outliers enables more precise and reliable expert assignment.

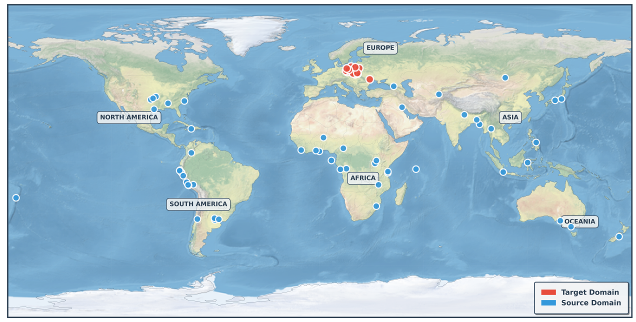


Figure 7. The spatial distribution for the OpenEarthMap (cross-continent) task. The source domain comprises North America, South America, Africa, Asia, and Oceania, while the target domain is Europe.

Backbone	Gating Function	mIoU (%)		
		Five-Billion-Pixels (Cross-Sensor)	Five-Billion-Pixels (Cross-Regional)	FLAIR (Cross-Regional)
VFM DINOv3	Softmax	64.54	59.59	62.69
	Gaussian	65.35	59.88	62.82
	Laplace	**66.19**	**60.32**	**63.18**
RSFM DOFA	Softmax	56.11	54.68	60.75
	Gaussian	54.80	55.18	61.31
	Laplace	**57.73**	**55.41**	**61.50**

Table 5. **Ablation study on gating functions.** The metric is mIoU (%). Best results for each backbone are in **bold**.

8.2. Impact of the Number of Activated Experts in SpectralMoE

We investigate the influence of k, the number of selected "Top-k" experts for each token, as defined in our gating mechanism (Section 3, Equation (2)). This hyperparameter controls the sparsity of the expert activation. A setting of $k = 1$ implies that each token is routed to only the single best expert, whereas $k > 1$ allows tokens to be processed by multiple experts.

We conduct experiments varying k from 1 to 5, with the results presented in Tab. 6. The experimental data clearly shows that setting $k = 1$ achieves the best mIoU performance across all three multispectral benchmarks. Increasing k to 2 or higher consistently leads to a decline in performance. This suggests that forcing a sparse, single-expert selection ($k = 1$) encourages a higher degree of expert specialization, which is more beneficial for generalization than activating and averaging multiple experts. Therefore, we adopt $k = 1$ as the optimal setting in our final model.

8.3. Impact of the Rank of Learnable Token Experts in SpectralMoE

We analyze the impact of the rank r used in the Low-Rank Decomposition for our expert parameterization, as introduced in Section 3. This rank r ($T_e = A_e \cdot B_e$, where $A_e \in \mathbb{R}^{m \times r}, B_e \in \mathbb{R}^{r \times d}$) governs the trade-off between

Top-k	mIoU (%)		
	Five-Billion-Pixels (Cross-sensor)	Five-Billion-Pixels (Cross-Regional)	FLAIR (Cross-Regional)
1	**66.19**	**60.32**	**63.18**
2	64.86	59.83	62.38
3	64.41	59.43	62.35
4	64.31	59.39	62.32
5	64.57	59.51	62.40

Table 6. Ablation study on the number of activated experts (*Top-k*). Best results are in bold.

Rank r	Params (M)	Five-Billion-Pixels (Cross-Sensor)		Five-Billion-Pixels (Cross-Regional)		FLAIR (Cross-Regional)	
		mIoU	mAcc	mIoU	mAcc	mIoU	mAcc
4	3.80	62.14	74.90	57.23	73.10	61.90	75.86
8	4.44	62.22	75.24	58.85	75.18	62.93	76.31
16	5.74	**66.19**	**77.26**	**60.32**	**75.78**	**63.18**	**76.52**
32	8.33	64.07	76.93	59.20	74.34	62.63	76.23
64	13.51	63.13	75.55	58.93	74.30	62.08	76.06

Table 7. **Ablation study on the rank parameter r.** The metric is mIoU (%) and mAcc (%). The best configuration ($r = 16$) is highlighted in **bold**. Trainable parameters exclude the backbone.

the expressive capacity of the adaptive tokens and the number of learnable parameters.

To determine the optimal value, we conduct experiments varying r from 4 to 64, with results presented in Tab. 7. The model's performance, measured by mIoU and mAcc, improves as the rank increases from $r = 4$ to $r = 16$. At $r = 16$, the model achieves the best results across all three multispectral benchmarks, with a parameter count of 5.74M.

However, increasing the rank further to $r = 32$ and $r = 64$ leads to a noticeable performance degradation, while significantly increasing the parameter count (to 8.33M and 13.51M, respectively). This suggests that a higher rank may lead to overfitting or parameter redundancy. Therefore, we select $r = 16$ as the optimal configuration, as it provides the best balance between model performance and parameter efficiency.

8.4. Impact of the Length of Learnable Token Experts in SpectralMoE

We evaluate the impact of the adaptive token length m, a key hyperparameter in our expert parameterization ($T_e \in \mathbb{R}^{m \times d}$), as introduced in Section 3. This parameter m determines the number of adaptive tokens within each expert, directly influencing the expert's expressive capacity and its associated parameter count.

To identify the optimal value, we conduct experiments varying m from 25 to 250. The results, shown in Fig. 8, demonstrate a clear and consistent trend across both the DINOv3 (VFM) and DOFA (RSFM) backbones. The model's

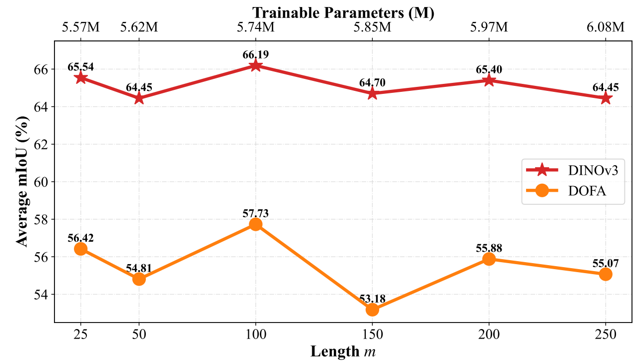


Figure 8. **Ablation study on the length m of learnable token experts on the Five-Billion-Pixels (Cross-sensor) DGSS task.**

performance (mIoU) improves as the token length increases from $m = 25$, peaking at $m = 100$. Beyond this optimal point, increasing the token length further (e.g., to 150, 200, or 250) results in a steady performance degradation.

This non-monotonic behavior suggests that while a sufficient number of tokens is necessary for expressive feature adjustment, an excessive length leads to parameter redundancy or overfitting, diminishing generalization. As the parameter count scales linearly with m, we select $m = 100$ as the optimal configuration, providing the best trade-off between model capacity and performance.

9. Additional Results on DGSS for Spectral Remote Sensing

This section presents supplementary qualitative results to further validate the generalization capability of our proposed SpectralMoE. Complementing the visualization provided in the main text, we present visual comparisons of the predicted land cover classification maps for the remaining benchmarks, spanning hyperspectral, multispectral, and RGB modalities.

Specifically, we show results for: the hyperspectral WHU-OHS (Cross-Regional) DGSS task (Fig. 9); the multispectral Five-Billion-Pixels (Cross-Regional) DGSS task (Fig. 10) and FLAIR (Cross-Regional) DGSS task (Fig. 11); and the RGB RS DGSS tasks including LoveDA (Cross-Style) (Fig. 12), Potsdam&Vaihingen (Cross Spectral Band) (Fig. 13), and OpenEarthMap (Cross-Continent) (Fig. 14).

As illustrated in Figs. 9–14, SpectralMoE consistently produces segmentation maps that are visually superior to competing methods across all these diverse and challenging domain shifts. Our model demonstrates a robust ability to mitigate domain gaps, yielding more precise semantic boundaries, reduced inter-class confusion (e.g., between spectrally similar classes), and fewer false positives. These results visually substantiate the effectiveness of our fine-grained, adaptive feature modulation and fusion strategy.

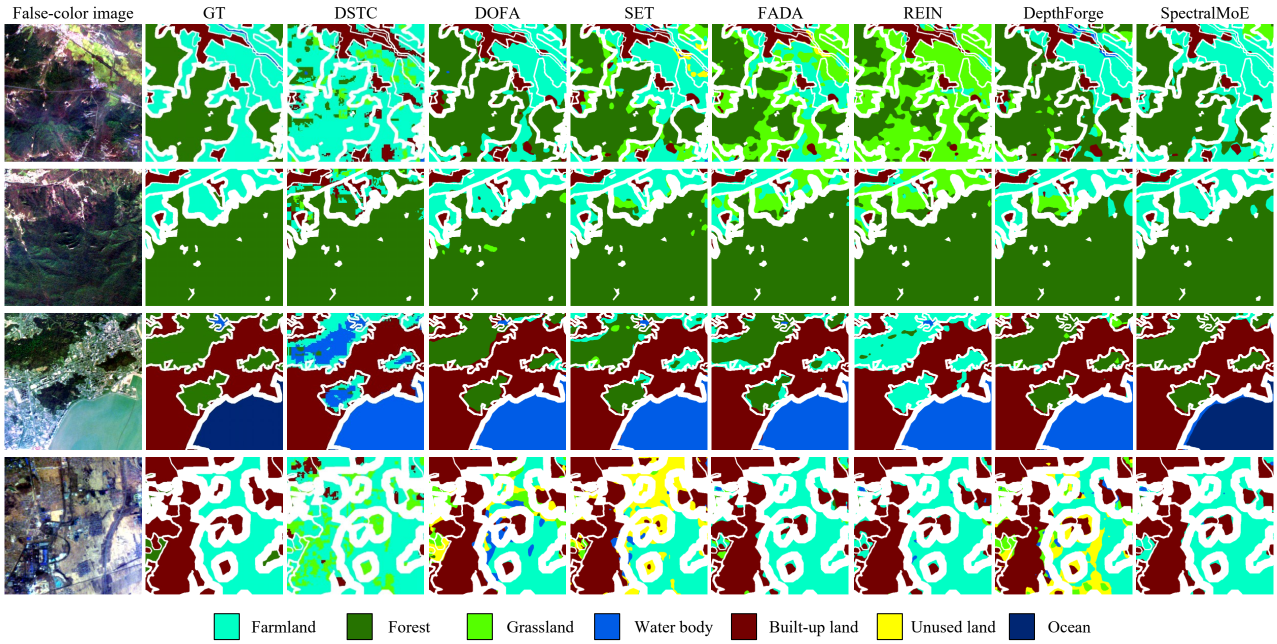


Figure 9. **Qualitative results for the WHU-OHS (Cross-Regional) hyperspectral DGSS task.** This figure provides a visual comparison of the land cover segmentation maps predicted by SpectralMoE against leading baseline methods.

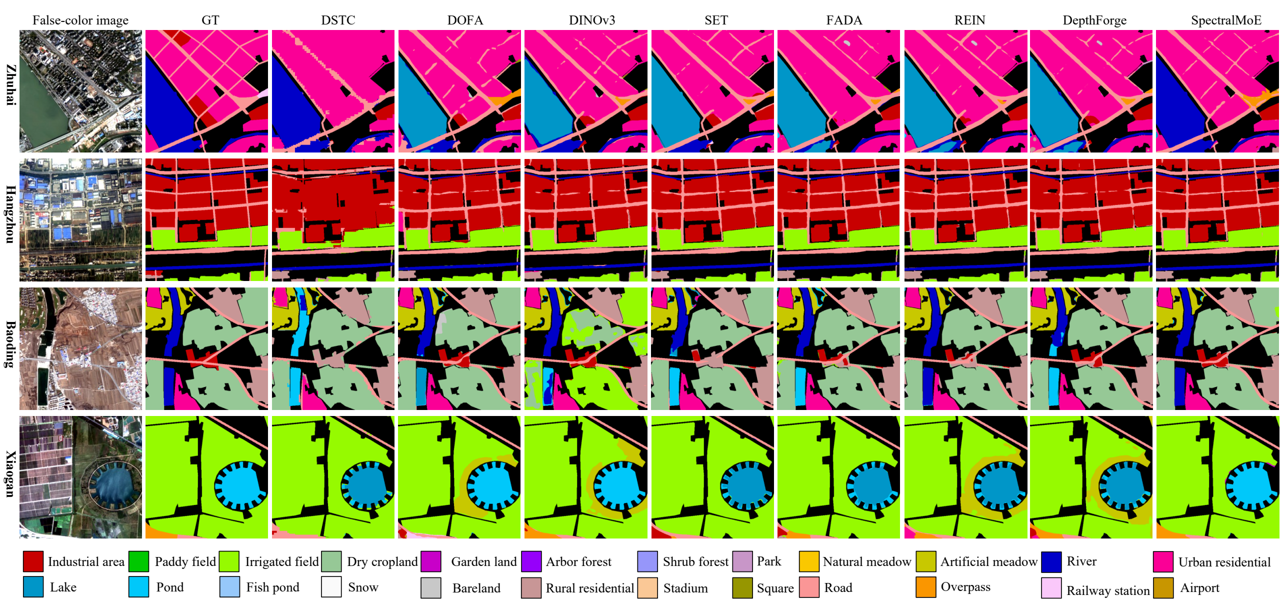


Figure 10. **Qualitative results for the Five-Billion-Pixels (Cross-Regional) multispectral DGSS task.** The figure compares the segmentation performance of SpectralMoE with competing methods, demonstrating its robustness across different regions.

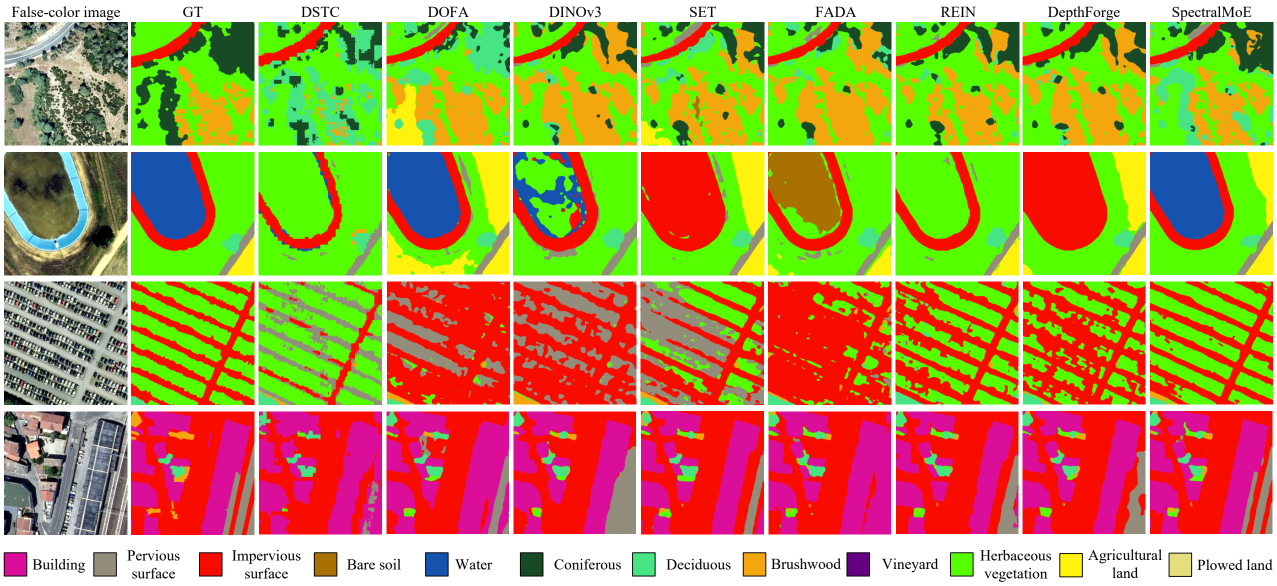


Figure 11. **Qualitative results for the FLAIR (Cross-Regional) multispectral DGSS task.** Visual comparison of SpectralMoE and baseline methods on high-resolution aerial imagery.

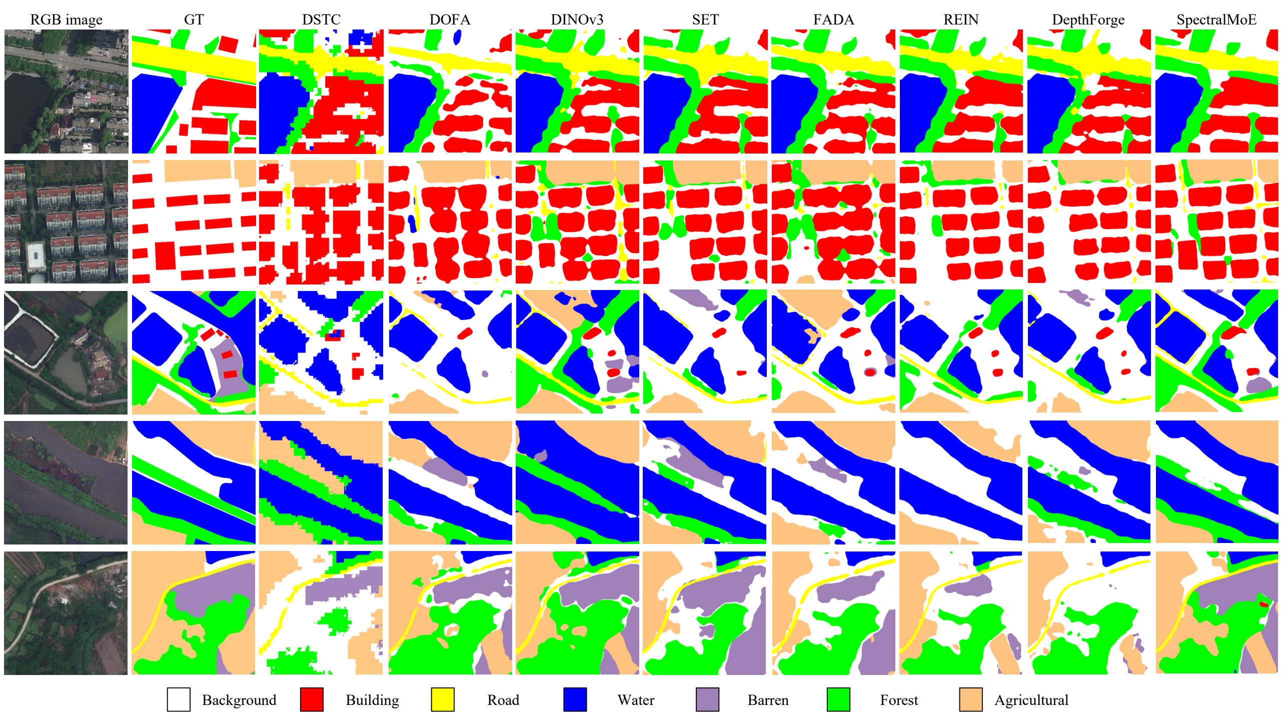


Figure 12. **Qualitative results for the LoveDA (Cross-Style) RGB DGSS task.** Visual comparison of segmentation maps in the challenging rural-to-urban generalization scenario.

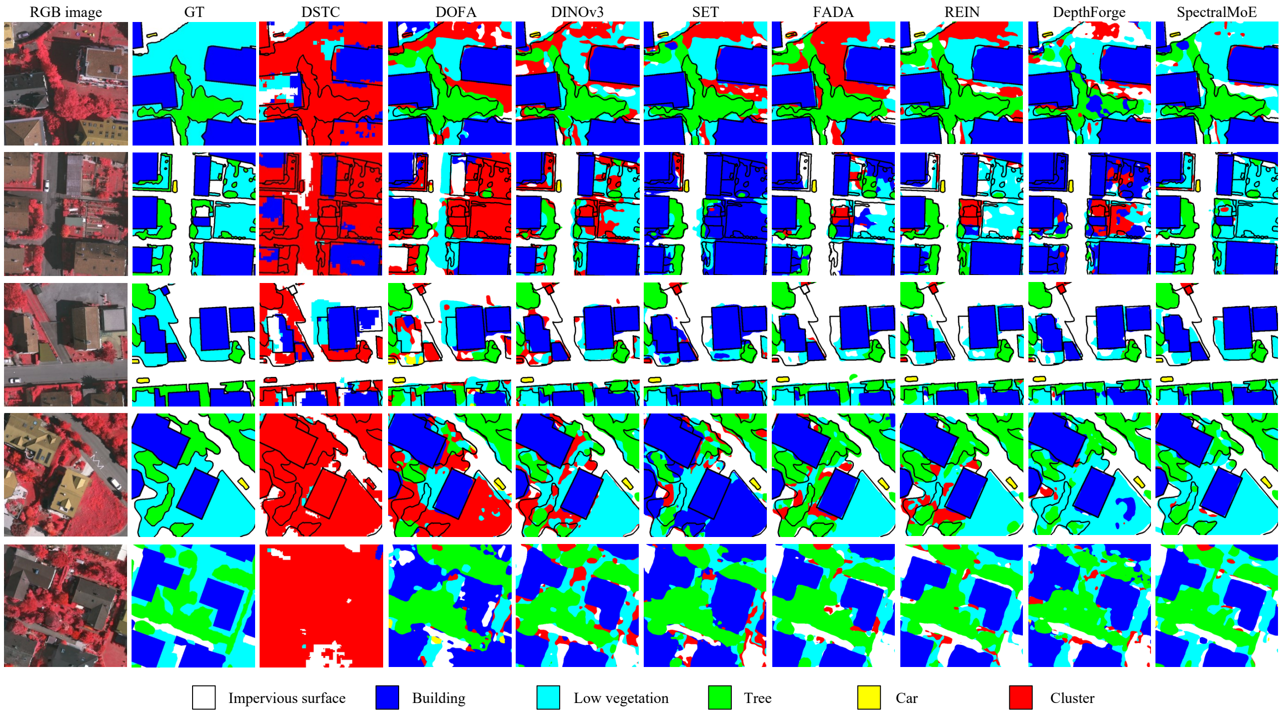


Figure 13. **Qualitative results for the Potsdam & Vaihingen (Cross-Spectral-Band) RGB DGSS task.** The figure illustrates the generalization capability of SpectralMoE when transferring from RGB to NIR-R-G spectral bands.

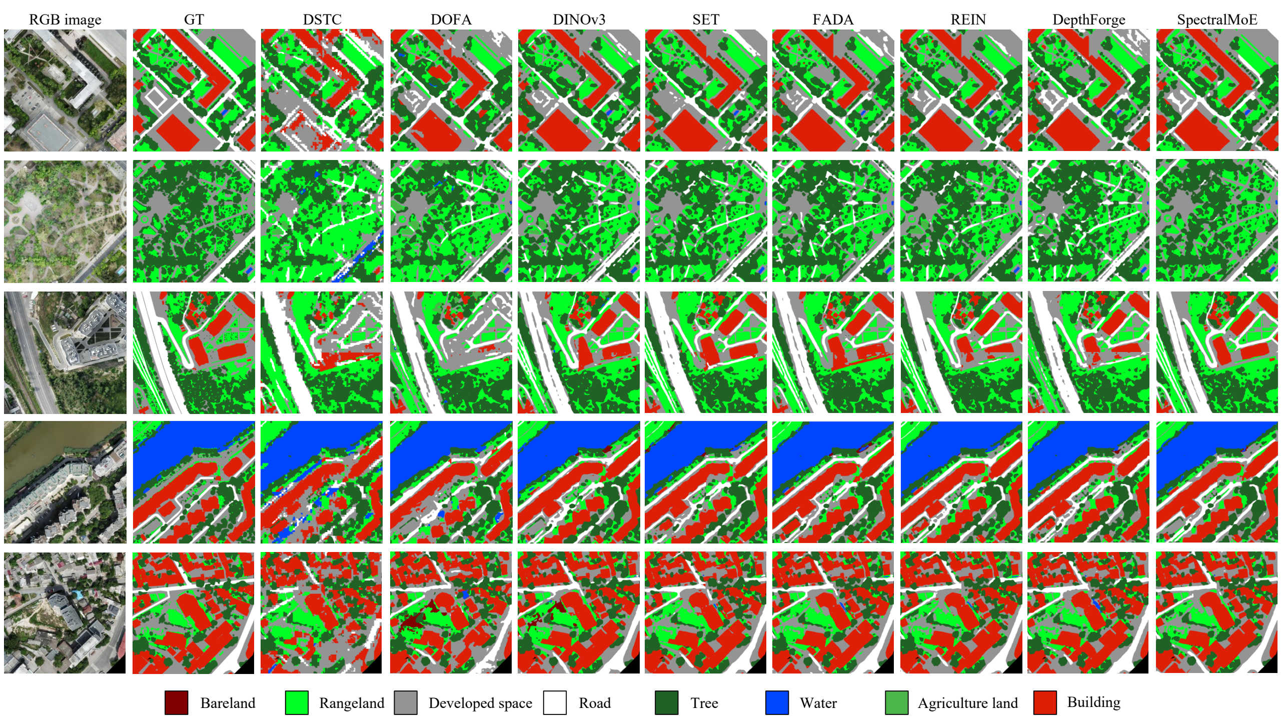


Figure 14. **Qualitative results for the OpenEarthMap (Cross-Continent) RGB DGSS task.** Visual comparison showing the model's performance under large-scale continental distribution shifts.