جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of Science and Technology

Flexible Multimodal Foundation Learning for Heterogeneous Earth Observation and Environmental Monitoring

Understanding Earth across sensors and scales

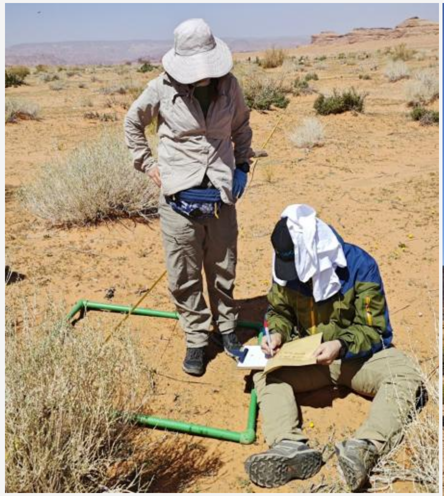
Jorge L. Rodríguez, Victor Angulo Morales, Areej Alwahas, Mariana Elías Lara, Fida Mohammad Thoker, Kasper Johansen, Bernard Ghanem, Fernando T. Maestre, Matthew F. McCabe

Jorge L. Rodríguez
King Abdullah University of Science and Technology | Presenting author

From local measurements to regional ecological inference

Field ecology

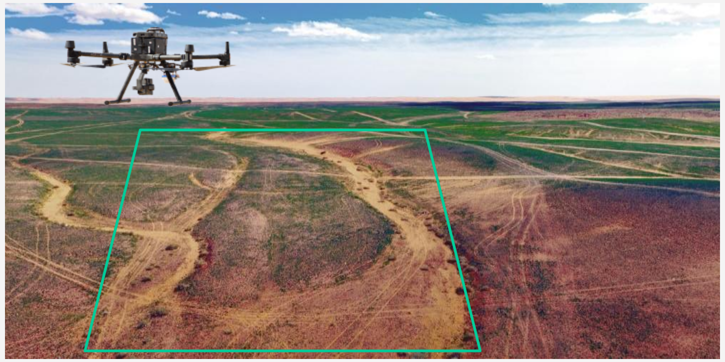
Ground measurements of ecosystems and biodiversity



- Vegetation structure
- Biogeochemistry
- Biomass and forage

UAV data acquisition

High resolution observations at local scale



- RGB
- Multispectral
- LiDAR

Scaling to satellites

From local to regional and national coverage

High-resolution satellites (SkySat)

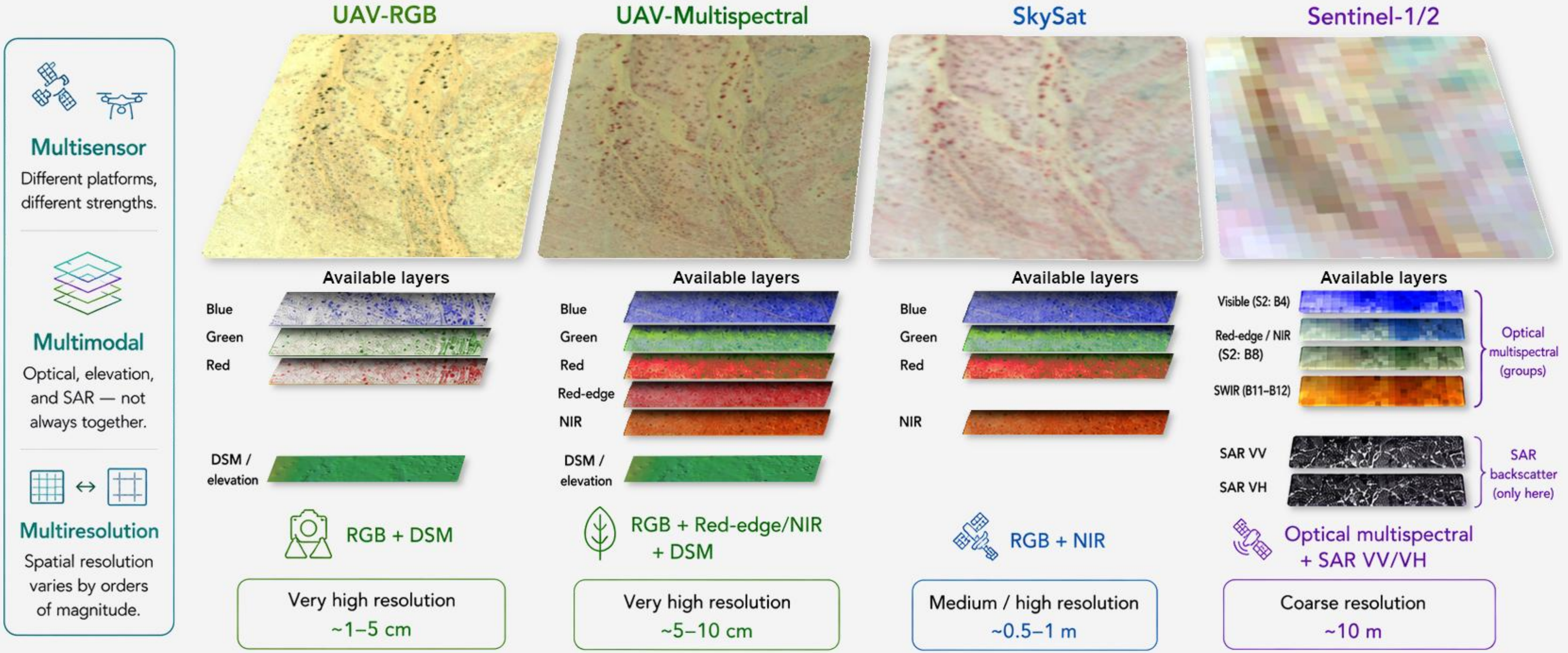
Medium-resolution satellites (Sentinel-1/2)

- Large-area coverage
- Frequent revisits
- Multi-source and complimentary

Our research integrates **field measurements, UAV data acquisition,** and **satellite data** across scales to better understand and monitor ecosystems at regional to global scales.

Multisensor, multimodal, multiresolution ecological data

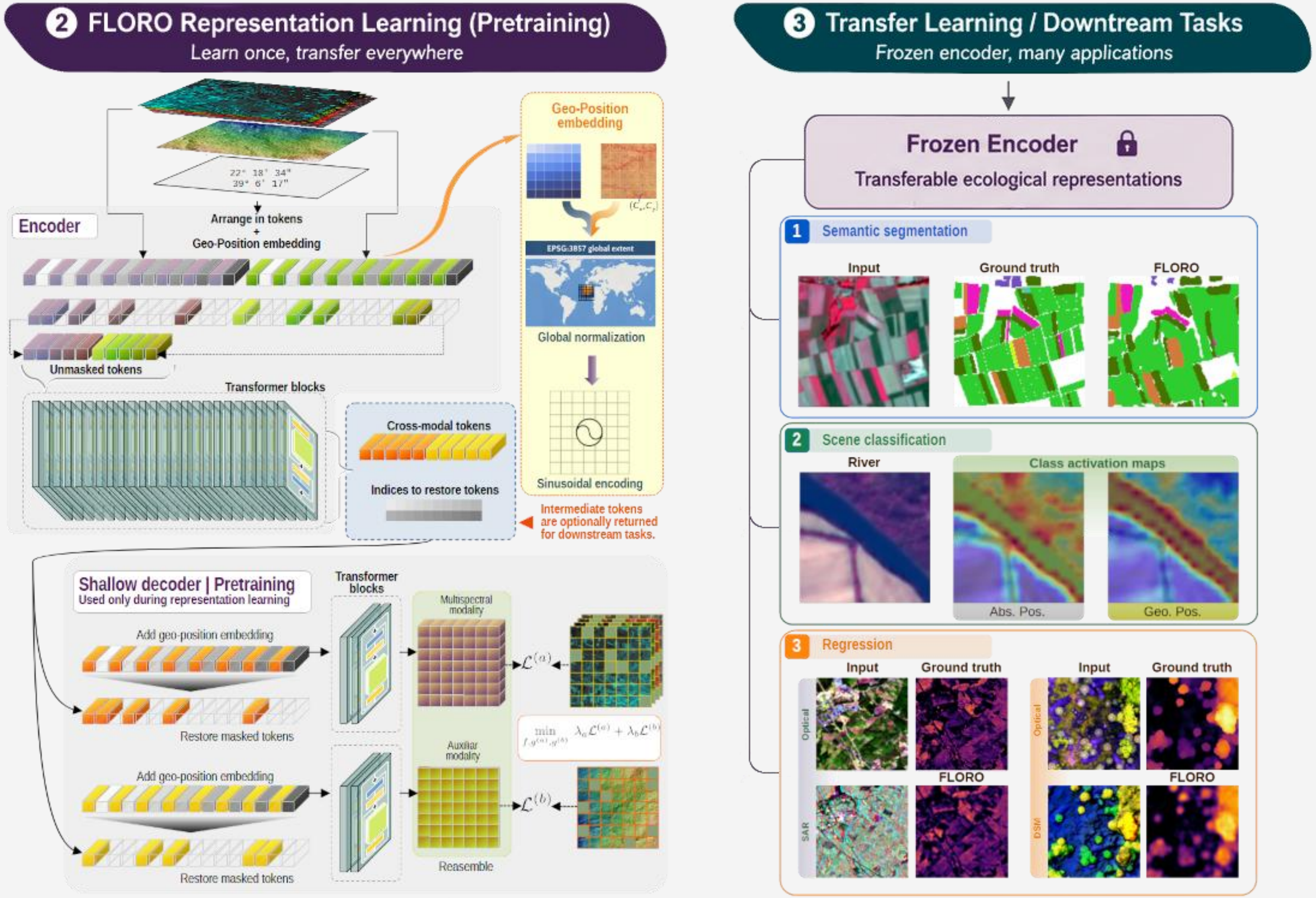
Common workflows do not share a fixed set of spectral bands or modalities.



Ecological remote sensing requires models that can work across incomplete, sensor-specific inputs.

FLORO: A Multimodal Geospatial Foundation Model for Ecological Remote Sensing Across Sensors and Scales

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology



Ecological remote sensing requires models that can work across incomplete, sensor-specific inputs.

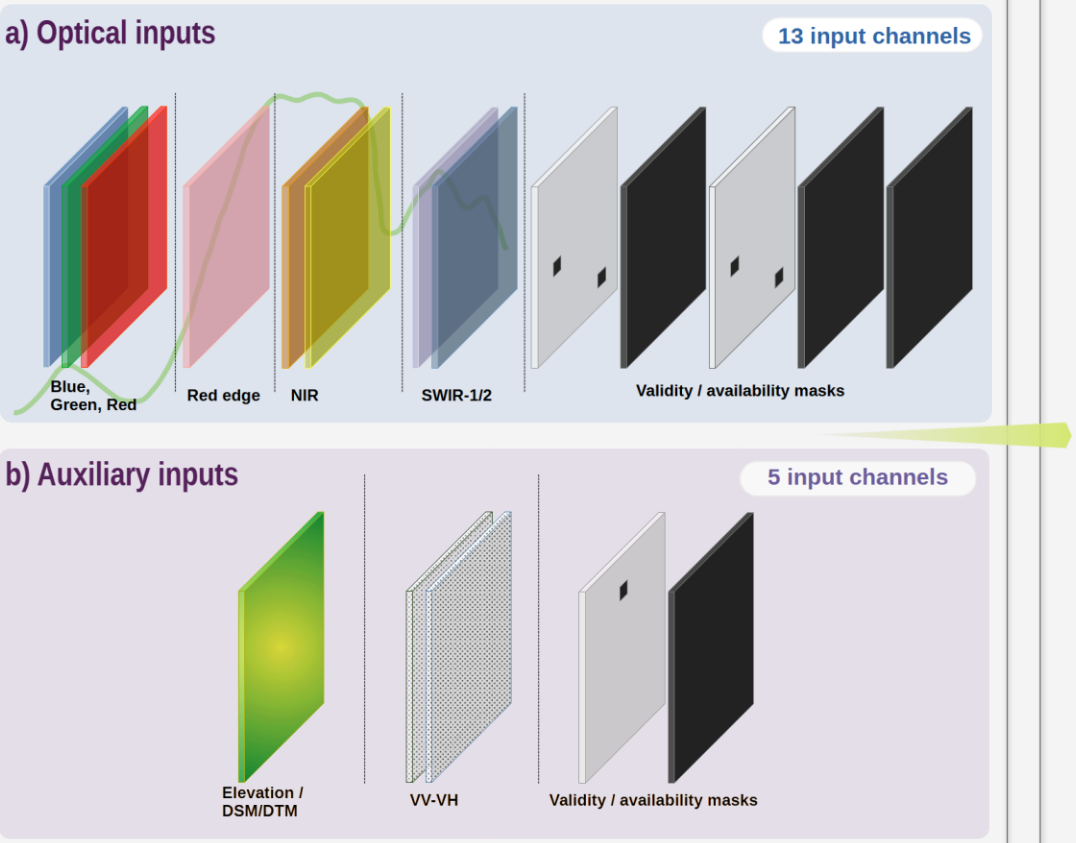
Pretraining

Sensor availability (data are heterogeneous)

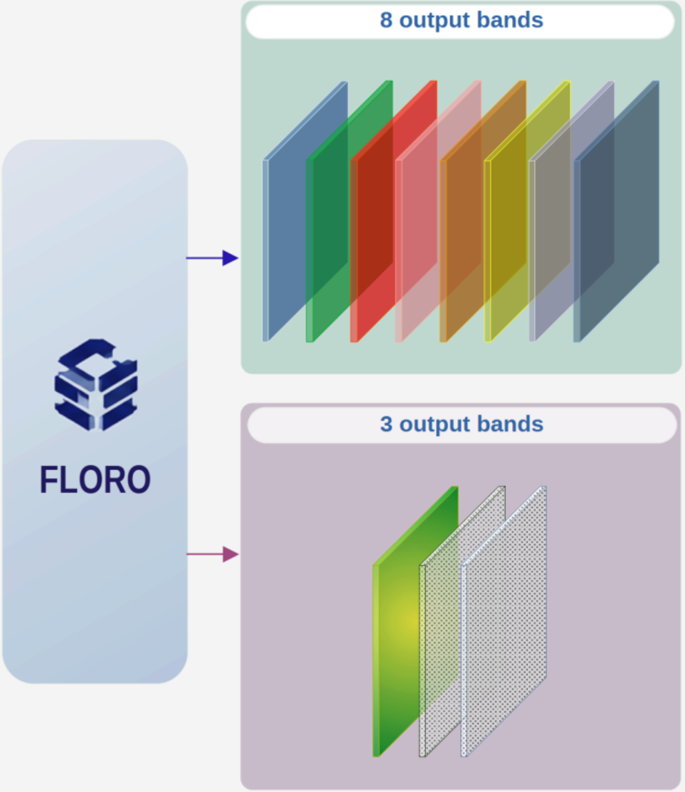
	UAV-RGB	UAV-MS	HR Sat.	S1 / S2
Optical				
Blue	✓	✓	✓	✓
Green	✓	✓	✓	✓
Red	✓	✓	✓	✓
Red edge		✓		✓
Near infrared		✓	✓	✓
Near infrared (B8A)				✓
SWIR 1				✓
SWIR 2				✓
Modalities				
Elevation	✓	✓	✓	✓
VV				✓
VH				✓

✓ Available ☐ Not available

Input representation (FLORO inputs)



Representation learning (reconstruction targets)

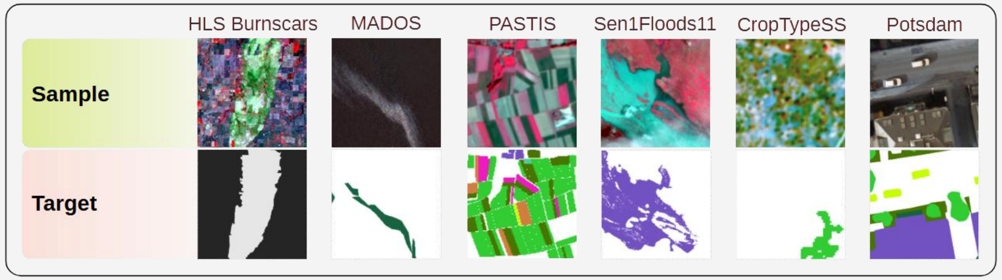


Validity / availability masks indicate which bands are present for a sample. They modulate contributions by modality and band group, but do not remove valid pixels within an available group.

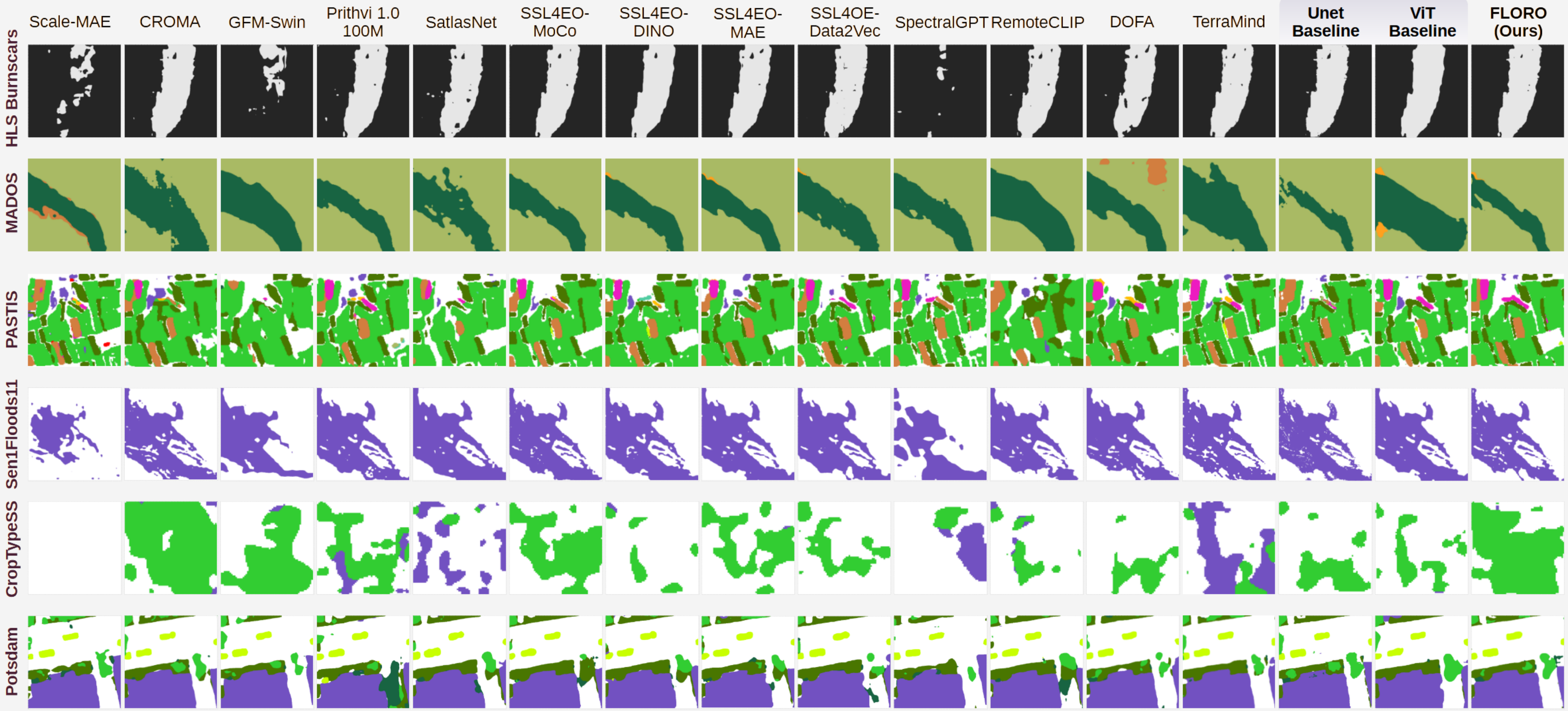
During pretraining FLORO learns representations from incomplete, sensor-specific optical, elevation and SAR data.

Segmentation transfer across diverse benchmarks

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of Science and Technology



FLORO provides competitive and consistent segmentation transfer across heterogeneous sensors, spatial resolutions, and land-cover types.



FLORO

Avg. rank: **4.0**

Competitive transfer across **six segmentation benchmarks**

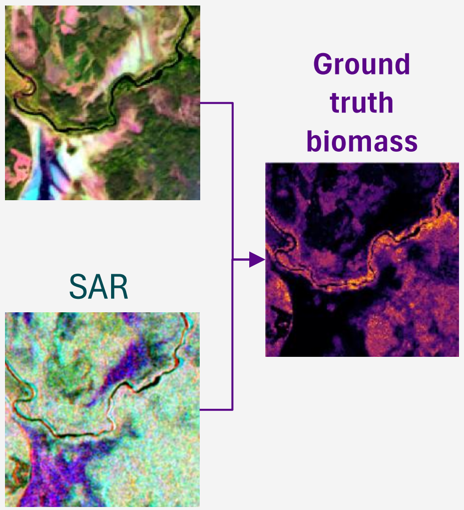
- Multiple sensors
- Different spatial resolutions
- Diverse land-cover types

Regression transfer on BioMassters

Biomass estimation from optical and SAR data

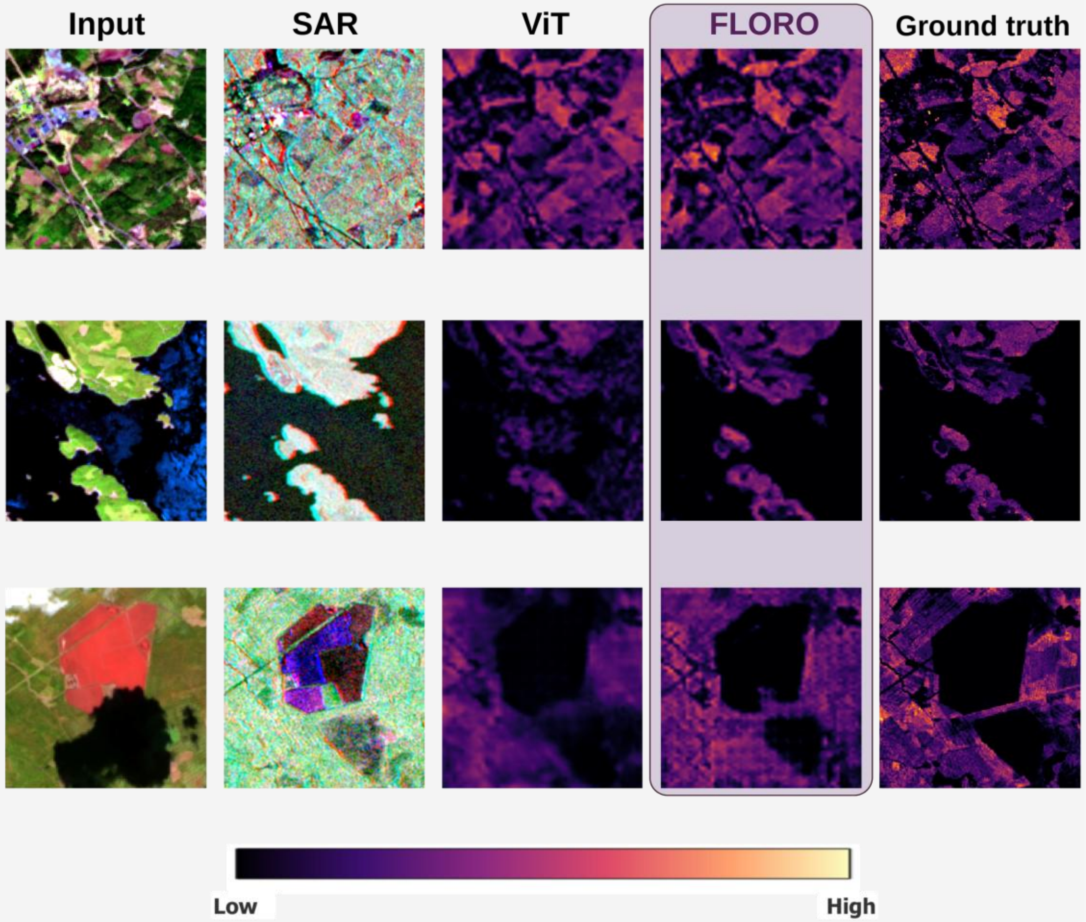
جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

Benchmark inputs

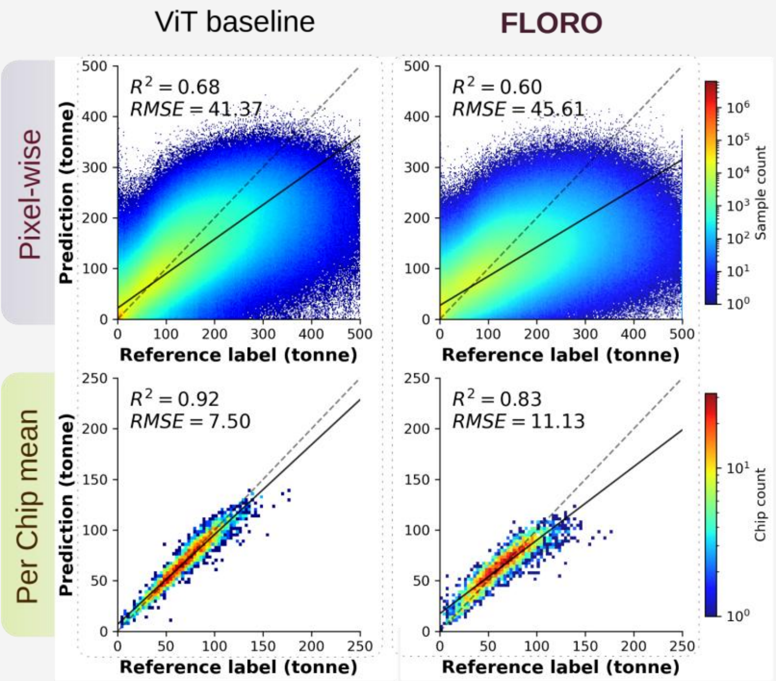


- Paired optical-SAR inputs
- Pixel-wise height estimation
- Evaluation at pixel-wise and chip-level scales

Some qualitative examples



Quantitative summary



- Strong chip-level agreement
- Competitive pixel-wise estimation
- Benefits from multimodal pretraining

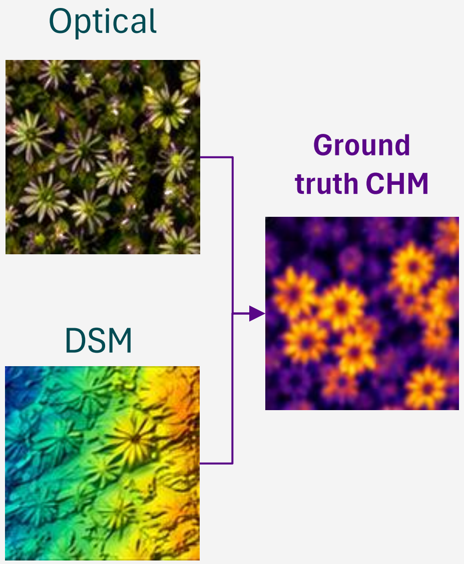
FLORO provides competitive biomass regression transfer, capturing meaningful spatial patterns from optical and SAR inputs.

Regression transfer on Espeletia

Canopy height estimation from optical and elevation data

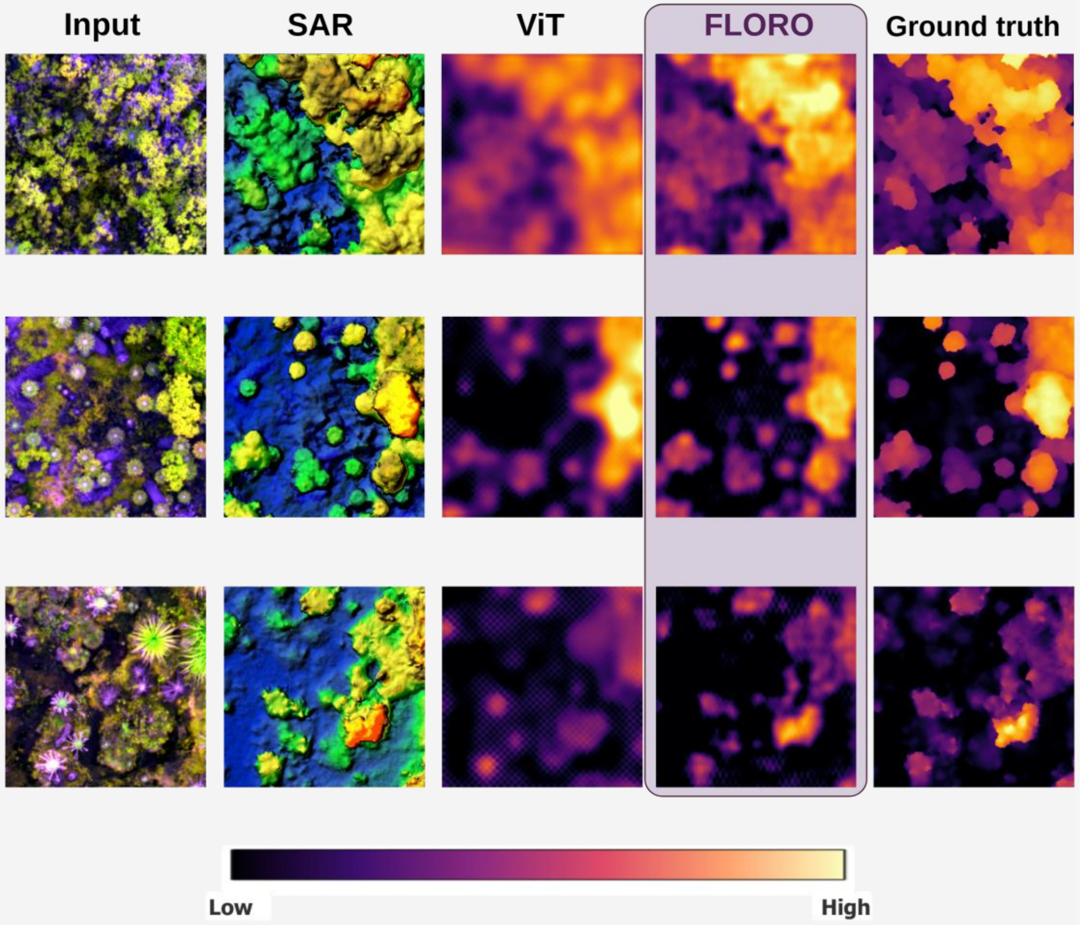
جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

Benchmark inputs

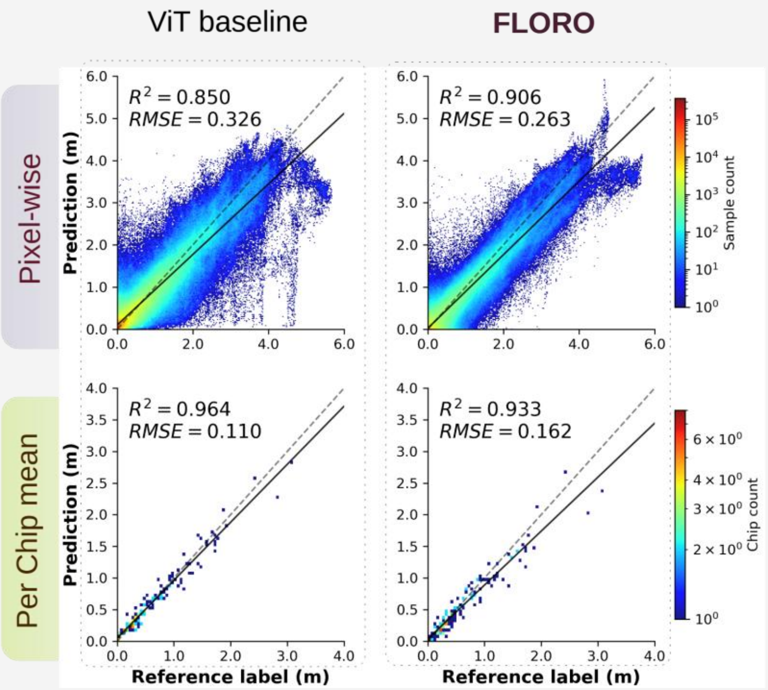


- Paired optical-DSM inputs
- Pixel-wise height estimation
- Evaluation at pixel-wise and chip-level scales

Some qualitative examples



Quantitative summary



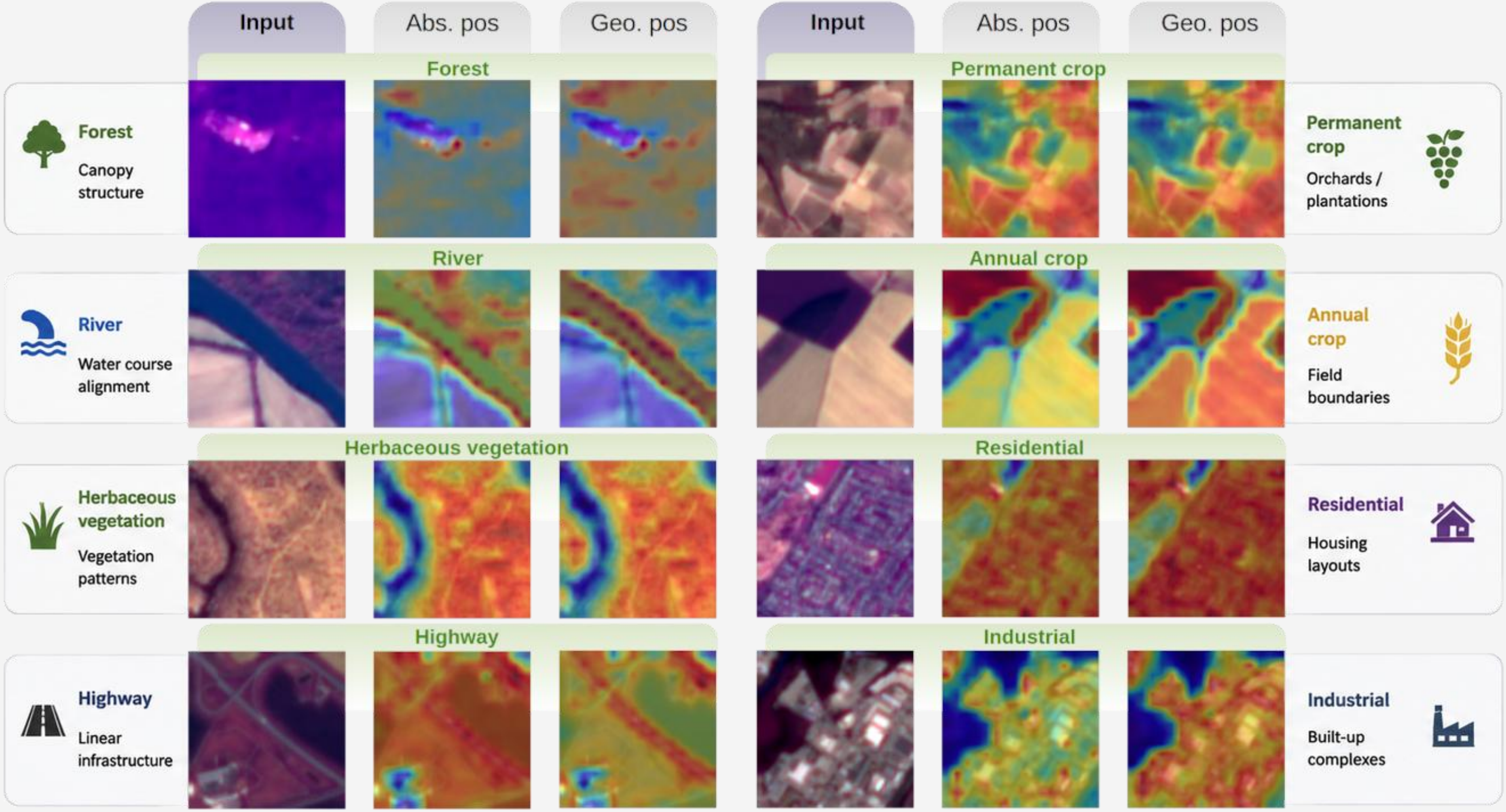
- Stronger pixel-wise estimation
- Competitive chip-level agreement
- Benefits from multimodal pretraining

FLORO provides strong transfer for canopy-height estimation, capturing meaningful spatial structure from optical and DSM inputs.

Geo-position embeddings matter for CAMs

Location helps the model focus on the right places

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology



Key takeaways

Better localization
Geo-pos embeddings produce CAMs that focus on the correct objects and boundaries.

Stronger generalization
Helps the model transfer across locations, sensors, and spatial layouts.

Spatial awareness is key
Absolute position alone is insufficient; geographic context provides the missing signal.

Low activation — High activation

Geo-position embeddings improve the spatial focus and interpretability of CAMs across land cover types.

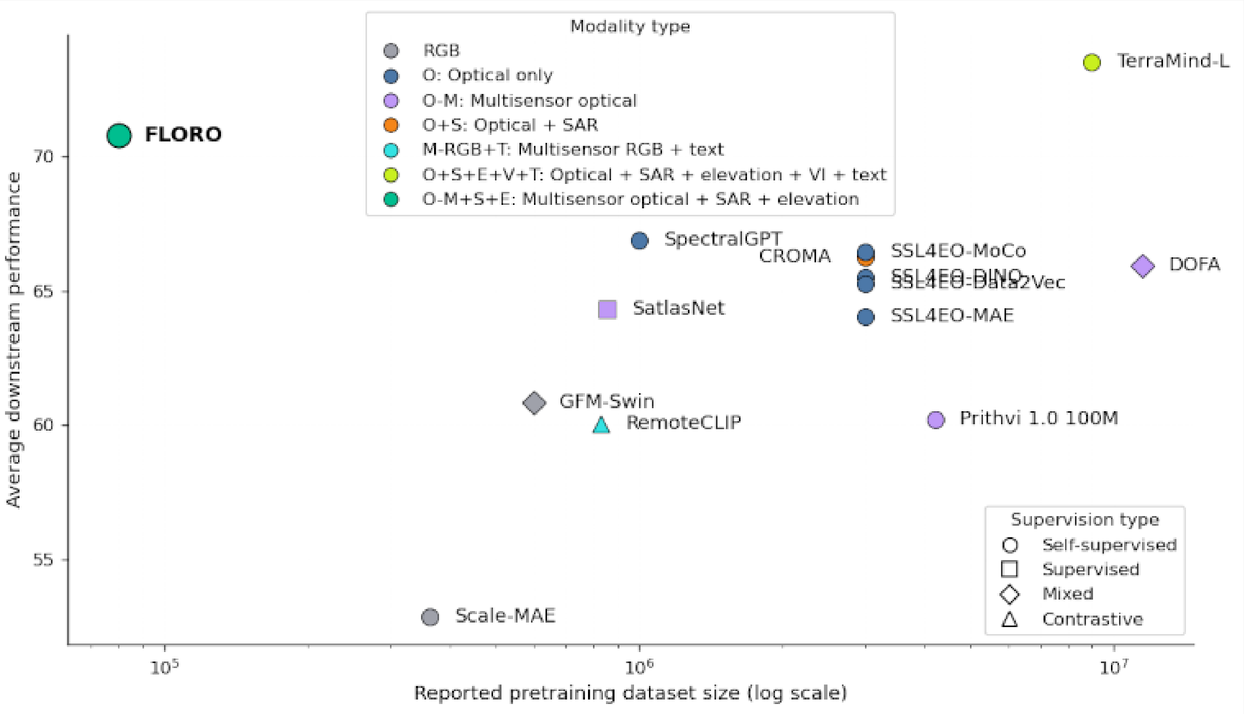
Diverse pretraining, not just massive pretraining

FLORO achieves competitive transfer with ~ 112.5x fewer pretraining images than TerraMind-L

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

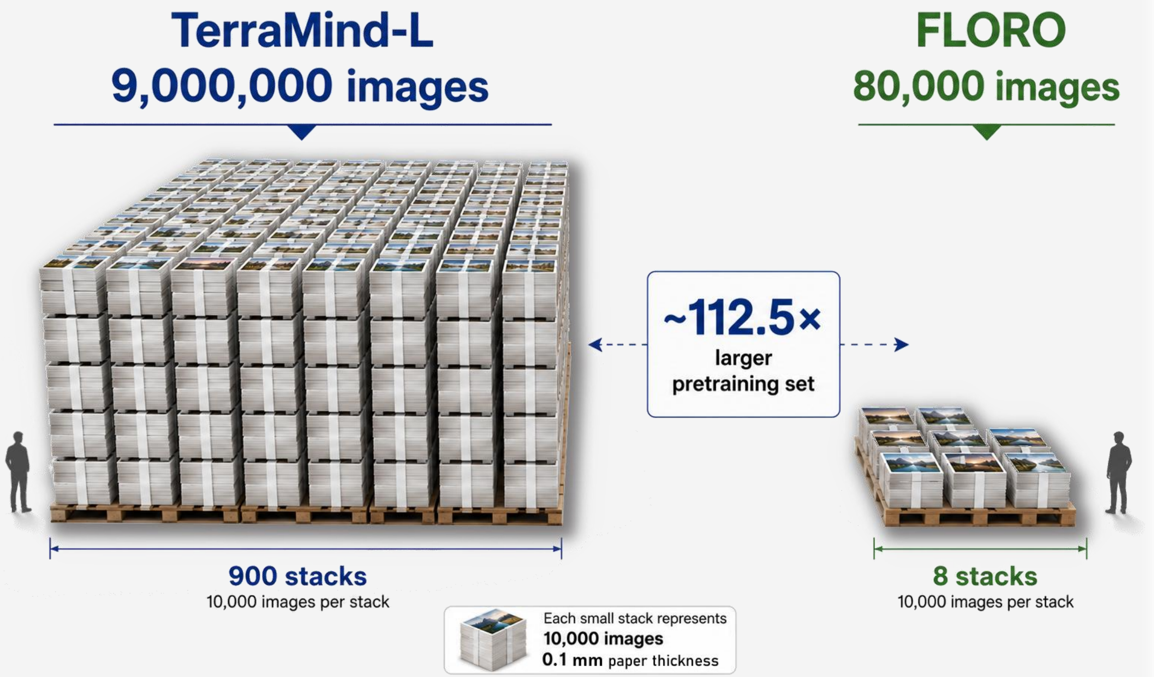
Cross-benchmark performance versus pretraining dataset size

FLORO delivers strong average downstream performance while using the smallest pretraining dataset



Visualizing scale comparison

FLORO was pretrained on substantially fewer images than similarly sized models



Competitive transfer
Top-tier average performance across 11 diverse benchmarks

Scale efficiency
~112.5× fewer images than TerraMind-L

Sensor & task diversity
Optical, SAR, elevation, and multiple ecological tasks

Practical for ecology
Enables foundation models when massive homogeneous corpora are not available

Takeaway: Carefully curated, diverse multimodal pretraining can deliver strong generalization without massive scale

Thanks!

Pre-print available:

FLORO: A Multimodal Geospatial Foundation Model for Ecological Remote Sensing Across Sensors and Scales

https://doi.org/10.48550/arXiv.2605.28174

jorge.rodriguezgalvis@kaust.edu.sa