SatOSM: Training geospatial foundation models with the Earth's largest open ground truth

Chenhui Zhang* Adam Yang Ruizhe Huang Jordi Laguarta Soler Xinyi Tong Jonathan Giezendanner Sherrie Wang
MIT, Earth Intelligence Lab

Contributions

SatOSM: A large-scale, high-resolution dataset for Earth observation.
- Object-level supervision using OSM masks and tags.
- Semantically diverse, open-vocabulary classes.

SatOSM-Net: Novel pretraining framework.
- Geographically and semantically grounded training architecture based on OSM.
- Outperforms baselines and existing GFMs in downstream tasks.

What is OpenStreetMap (OSM)?

- Voluntary geospatial database with billions of annotated objects.
- Open vocabulary annotation system with millions of distinct tags.

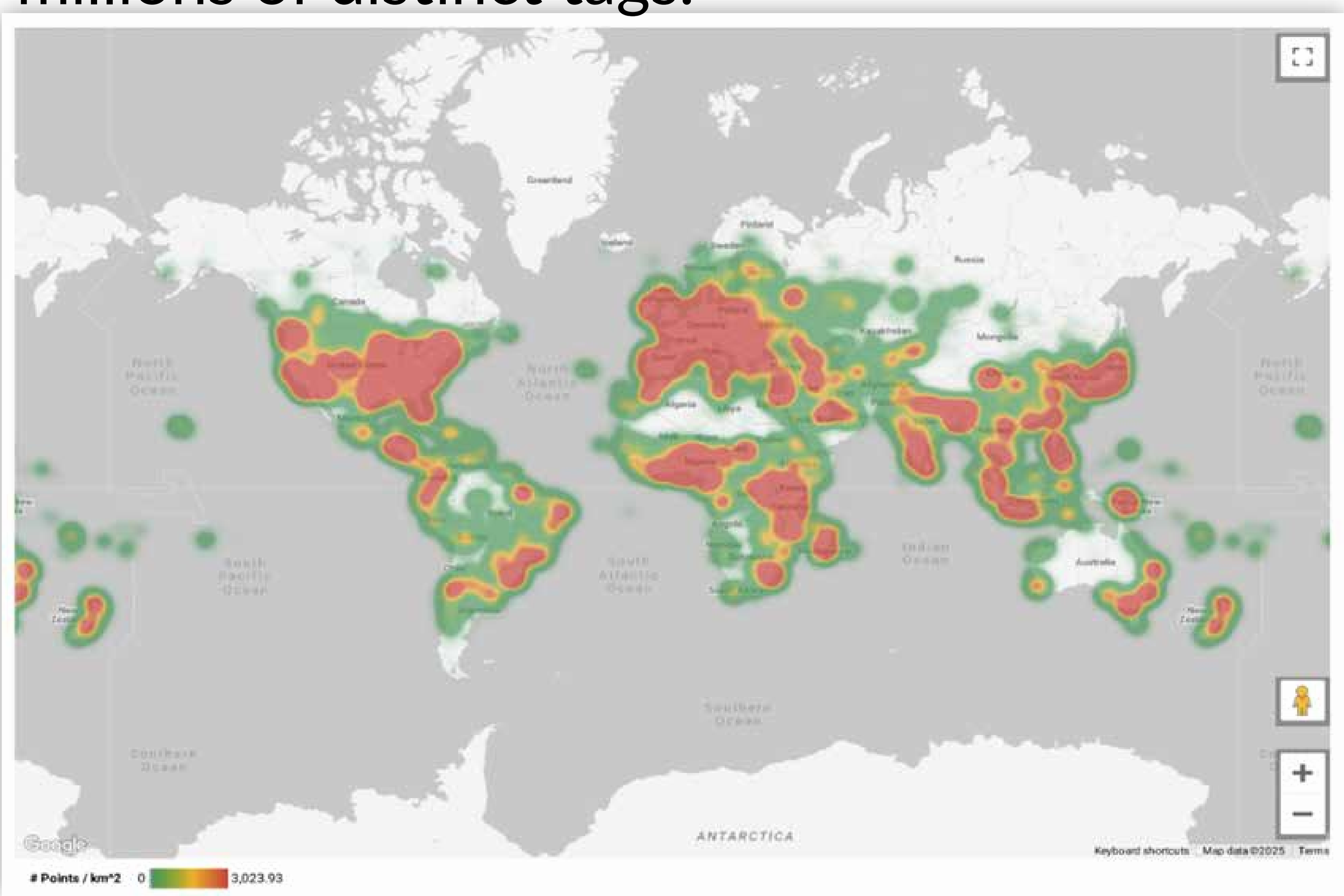

Fig 1. Annotation density of OSM across globe.

How do we use OSM?

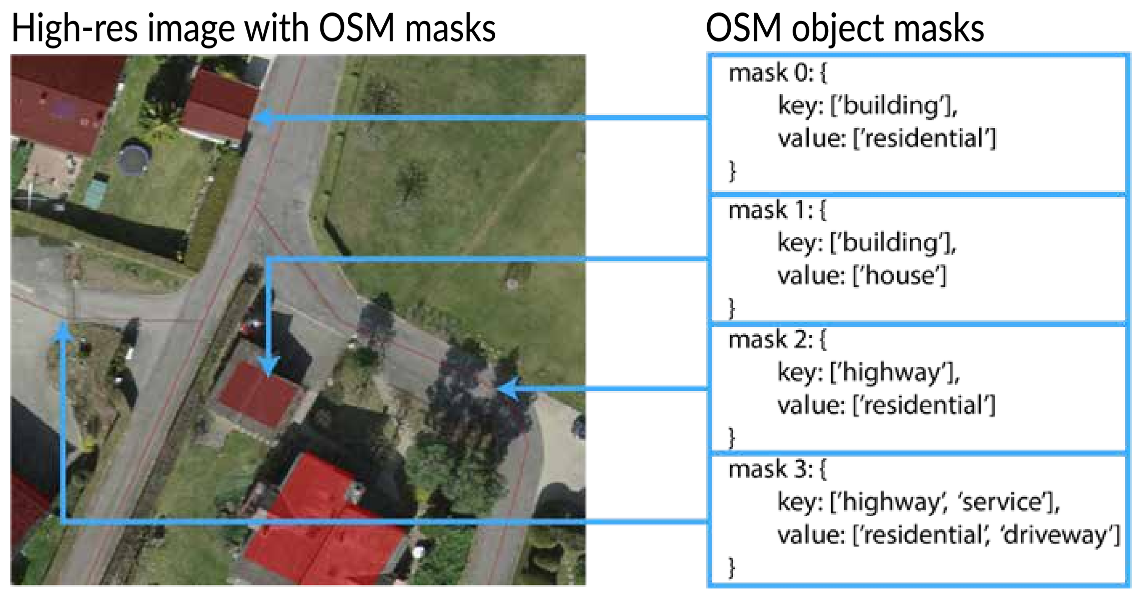

Fig 2. Sample from SatOSM which includes an image, OSM object masks, and OSM tags.

SatOSM

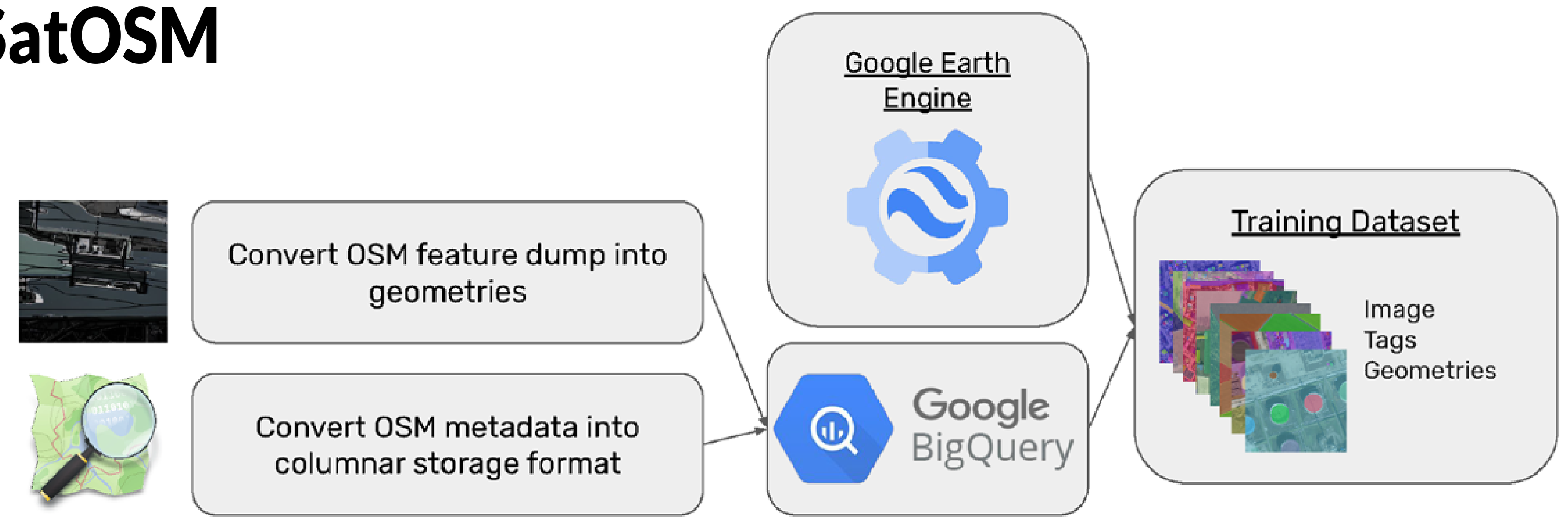

Fig 3. SatOSM data collection workflow.

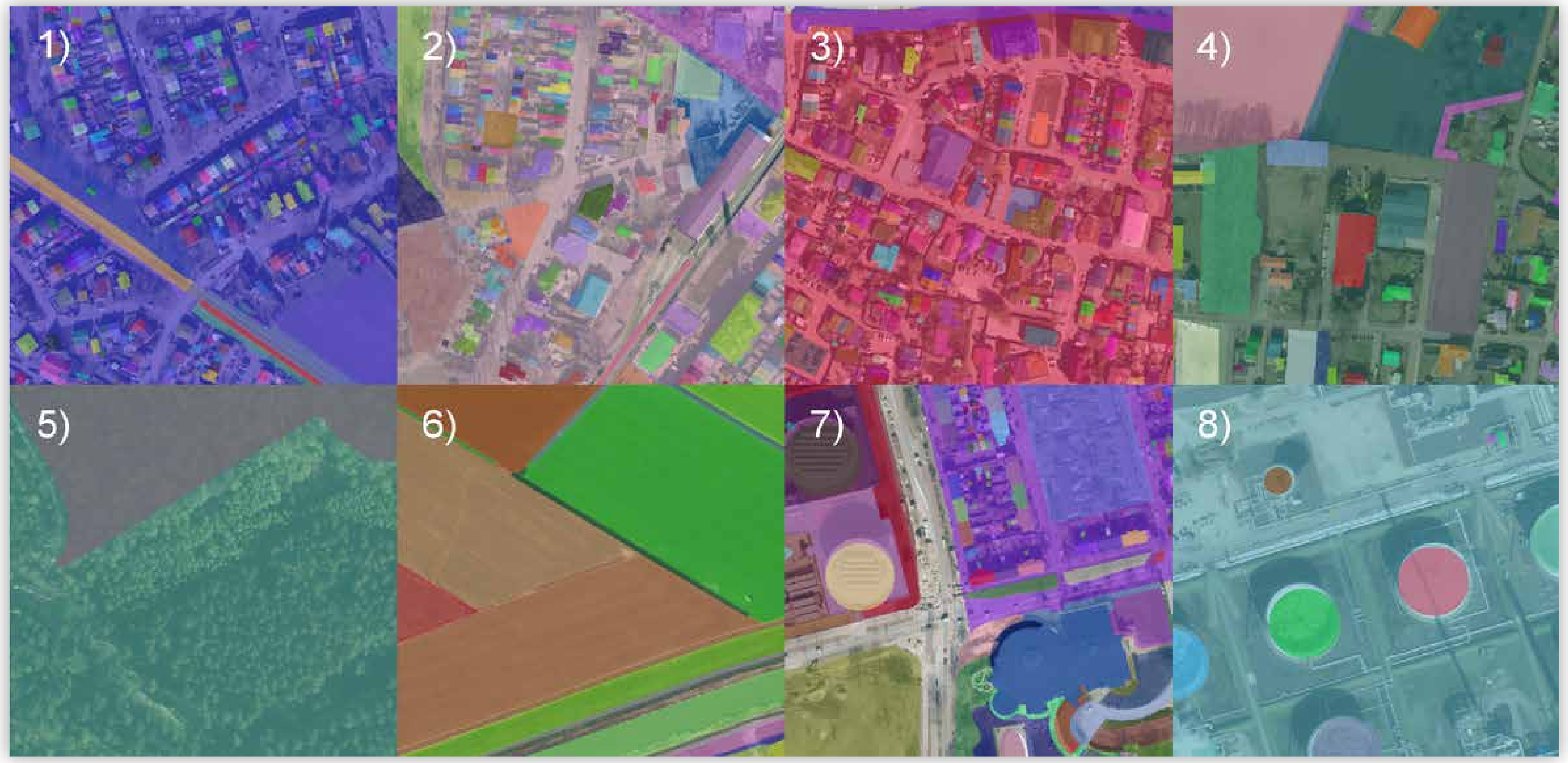

Fig 4. SatOSM samples across diverse regions. Tags in images above include 'building=industrial', 'landuse=forest', and 'amenity=parking'.

- Semantically diverse: 2,219 unique object classes.
- Large scale: 34 million high-res images with 122 million objects across 8 countries.

SatOSM-Net Pretraining

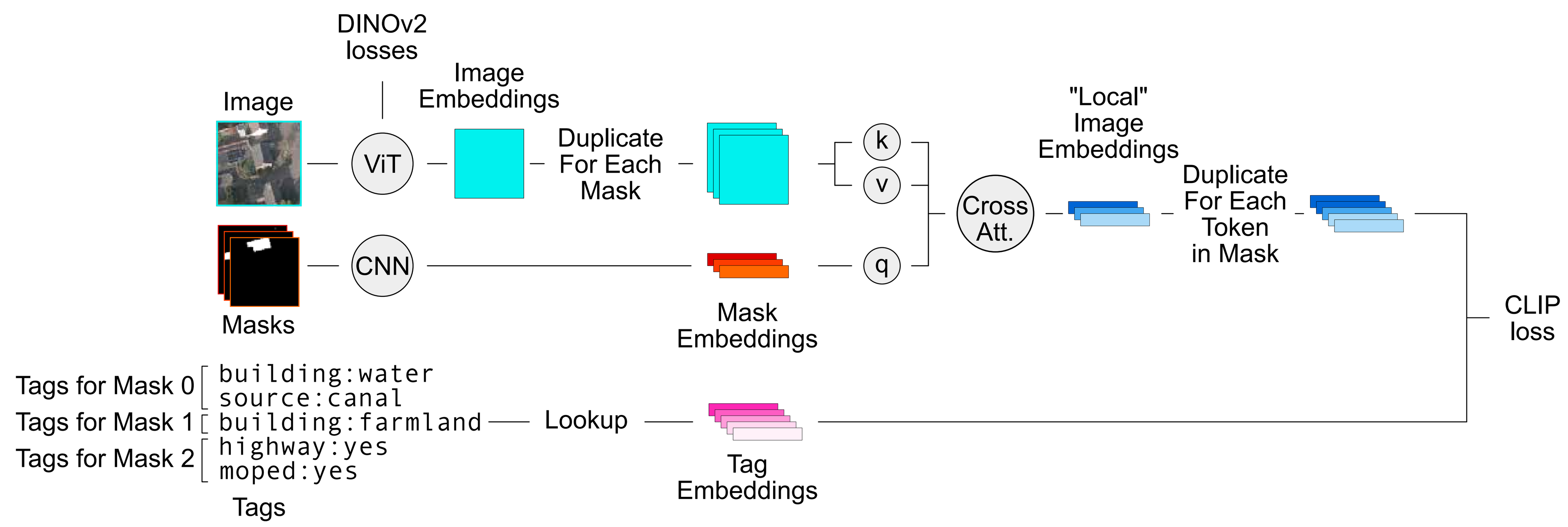

Fig 5. SatOSM-Net training architecture.

- Grounded image embeddings via cross attention with OSM masks.
- Learnable tag embeddings for semantic alignment.
- Two fold loss using CLIP and DINOv2.

Results

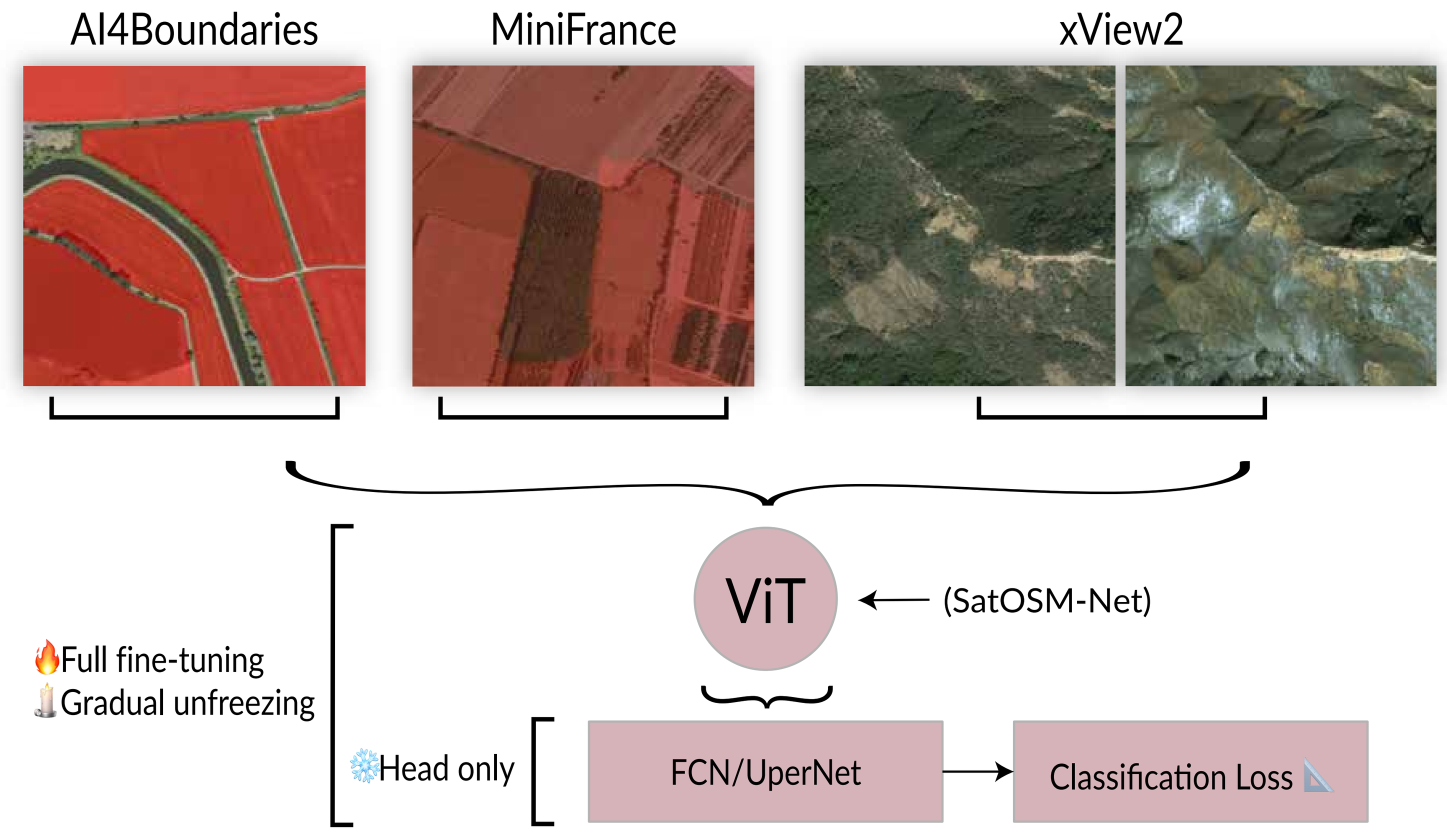

Fig 6. Fine-tuning setup for AI4Boundaries, MiniFrance, and xView2. SatOSM-Net evaluations performed using full fine-tuning, gradual unfreezing, and head only.

Downstream evaluation results

	AI4Boundaries*		MiniFrance		xView2	
Model	IoU	mAP@0.5	mIoU	FWIoU	mIoU	FWIoU
Non-pretrained						
U-Net	69.14	65.49	45.58	55.15	60.06	79.85
ViT	54.82	48.11	32.33	45.94	56.02	77.18
Pretrained GFMs						
Scale-MAE	64.51	59.17	43.96	53.61	53.71	75.59
SkyCLIP-50	64.59	52.76	53.00	58.05	**65.19**	**82.43**
DOFA-CLIP	75.34	71.93	53.22	58.39	63.57	81.39
SatOSM-Net	**77.48**	**74.47**	**56.76**	**60.35**	64.77	81.48

Fig 7. Comparison between baselines and existing GFMs. All GFMs fine-tuned using gradual unfreezing.

Comparison of fine-tuning strategies

	AI4Boundaries*		MiniFrance		xView2	
Method	IoU	mAP@0.5	mIoU	FWIoU	mIoU	FWIoU
Head only	74.49	70.13	52.42	57.90	62.93	81.26
Full fine-tune	69.07	64.33	44.02	53.13	53.71	76.13
Gradual unfreezing	**77.48**	**74.47**	**56.76**	**60.35**	**64.77**	**81.48**

Fig 8. Comparison between SatOSM-Net fine-tuning strategies.

- SatOSM-Net outperforms GFMs on AI4Boundaries and MiniFrance and is second best on xView2.
- Gradual unfreezing is the best performing fine-tuning strategy.

*AI4Boundaries evaluated with SatOSM-Net trained on a Netherlands-only SatOSM subset.

Future work

- Expand spatial coverage: SatOSM currently spans a handful of countries in EU.
- Possible SatOSM-Net designs not fully explored.

REO @ Eurips 2025