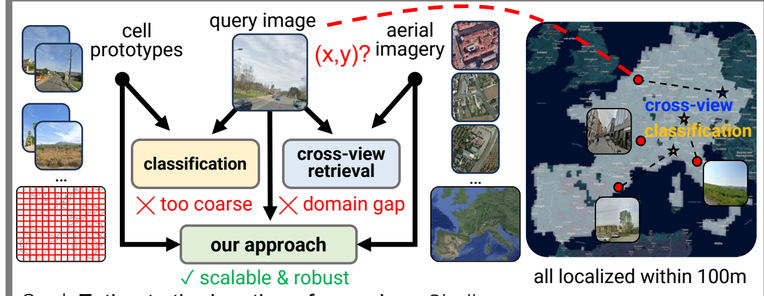
Scaling Image Geo-Localization to Continent Level

Philipp Lindenberger[1] Paul-Edouard Sarlin[2] Jan Hosang[2]
Matteo Balice[3] Marc Pollefeys[1] Simon Lynen[2] Eduard Trulls[2]

[1] ETH zürich [2] Google [3] POLITECNICO MILANO 1863

NEURAL INFORMATION PROCESSING SYSTEMS

1. Overview



Goal: Estimate the location of a random image **up to ~100m** over **10 countries**

Insight: Hybrid **classification** and **cross-view retrieval** **synergize**

Result: **69% Top-1 Recall @ 200m**

Challenges:
- lack of visual features (rural)
- temporal / seasonal changes
- limited coverage
- scale

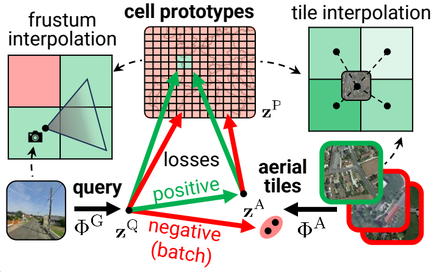
Evaluation: temporal split (2023)

2. Supervision

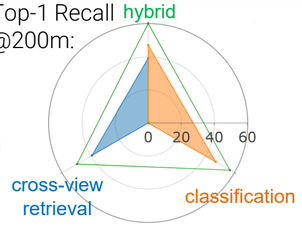
Loss: Combined Multi-Similarity Loss

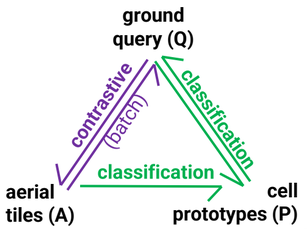
$$\mathcal{L}_{pos,i} = \frac{1}{\alpha} log\left(1 + \sum_{\mathrm{IJ}\in\{\mathrm{QA,QP,AP}\}} e^{-\alpha(s_{ii}^{\mathrm{IJ}}-\lambda)}\right)$$

$$\mathcal{L}_{neg,i} = \frac{1}{\beta} log\left(1 + \sum_{\mathrm{IJ}\in\{\mathrm{QA,AQ,QP,AP}\}} \sum_{j\in\mathcal{N}_i^{\mathrm{IJ}}} e^{\beta(s_{ij}^{\mathrm{IJ}}-\lambda)}\right)$$

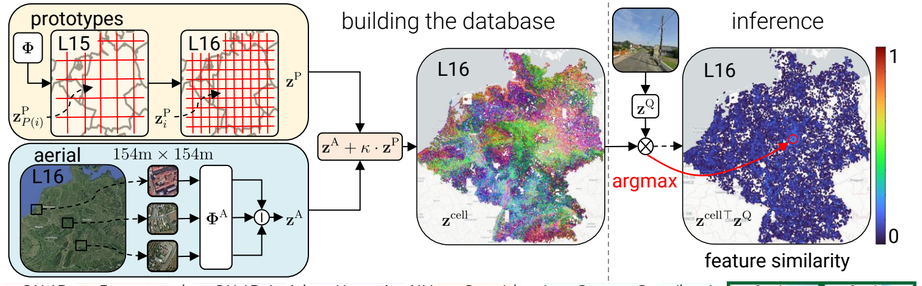


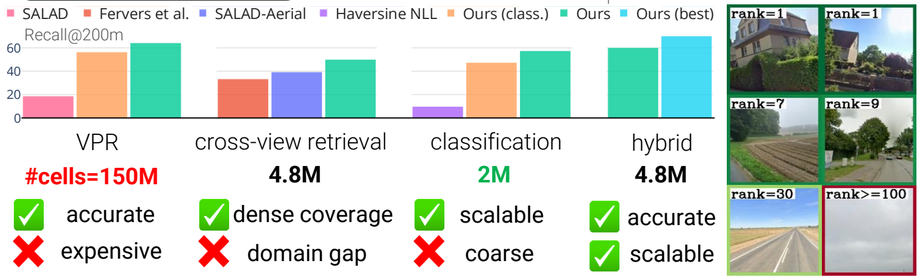
Top-1 Recall @200m:





3. Hybrid Localization



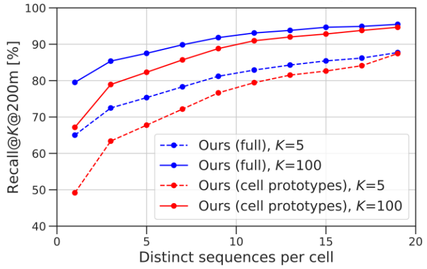


VPR	cross-view retrieval	classification	hybrid
#cells=150M	**4.8M**	**2M**	**4.8M**
✓ accurate	✓ dense coverage	✓ scalable	✓ accurate
✗ expensive	✗ domain gap	✗ coarse	✓ scalable

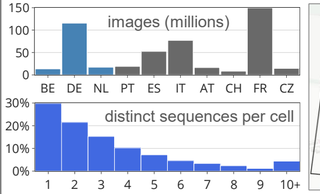
4. Data

classification and VPR can not handle missing data

aerial imagery anchors prototypes & enables localization in missing cells

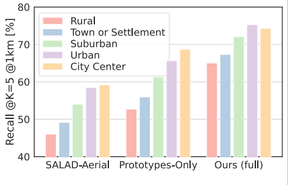


coverage (training split)



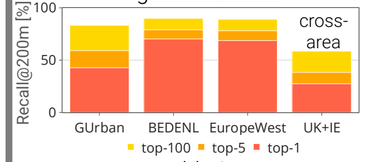
space-time sampling

data variety

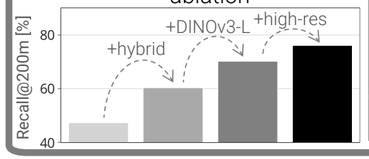


5. Insights

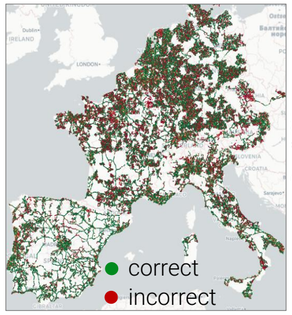
generalization



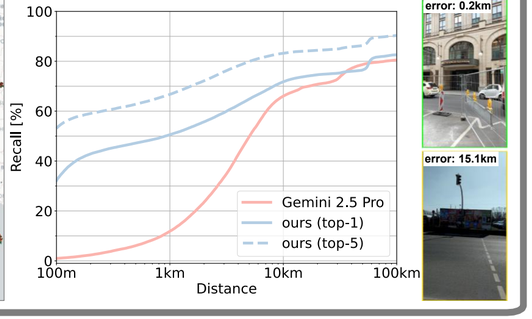
ablation



error distribution

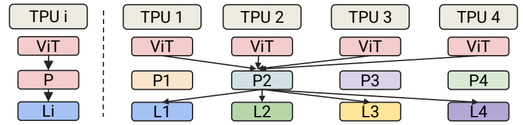


comparison to VLMs on GoogleUrban



6. Scaling

encoder replicated, prototypes sharded



finer cells = higher accuracy (same budget)