



PlanarCityLoc: Breaking Silhouette Ambiguity in LoD Aerial Localization via 3D Planes

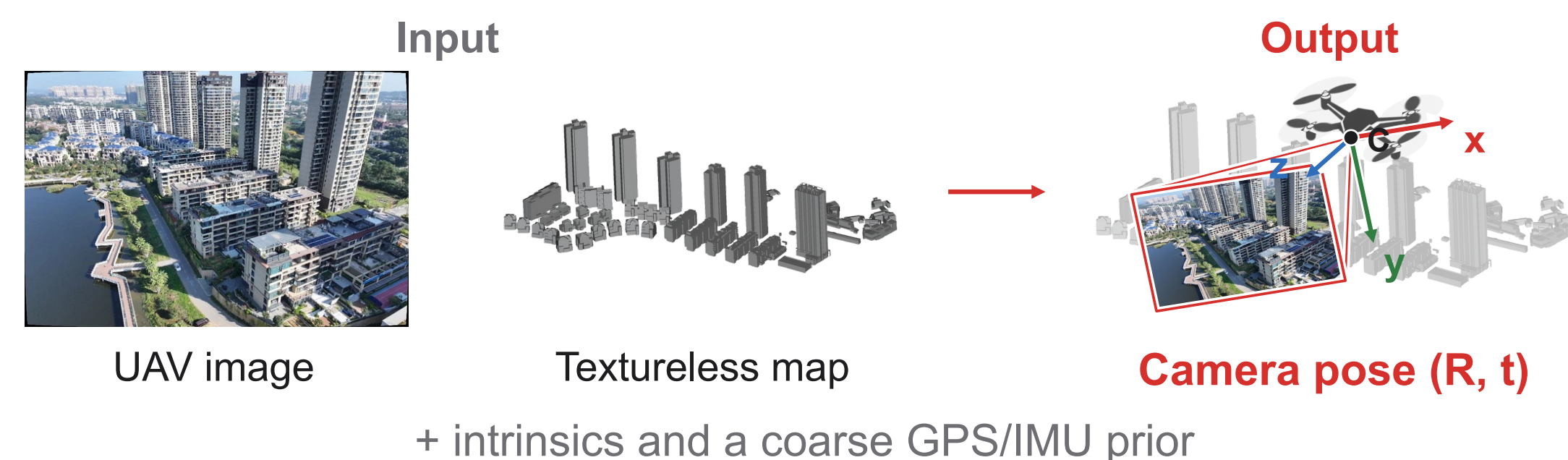
Oussema Dhaouadi^{1,2,4,5,6} Shao Jie Hu Chen^{1,2} Jakob Gollreiter^{1,2} Samir Jusufi^{1,3} Nikolas Rieger^{1,2,7} Caroline Wendlandt^{1,2}
 Christian Wender^{1,8} Jingfeng Mao^{1,2,4} Guangming Wang^{1,4} Zuria Bauer⁵ Marc Pollefeys^{5,9} Daniel Cremers^{2,6} Olaf Wysocki^{1,4}

¹ TUM.ai ² Technical University of Munich ³ LMU Munich ⁴ University of Cambridge ⁵ ETH Zürich ⁶ MCML ⁷ Konrad Zuse School of Excellence in Reliable AI (relAI) ⁸ BMW Group ⁹ Microsoft



Goal & motivation

Goal: 6-DoF UAV localization in LoD maps



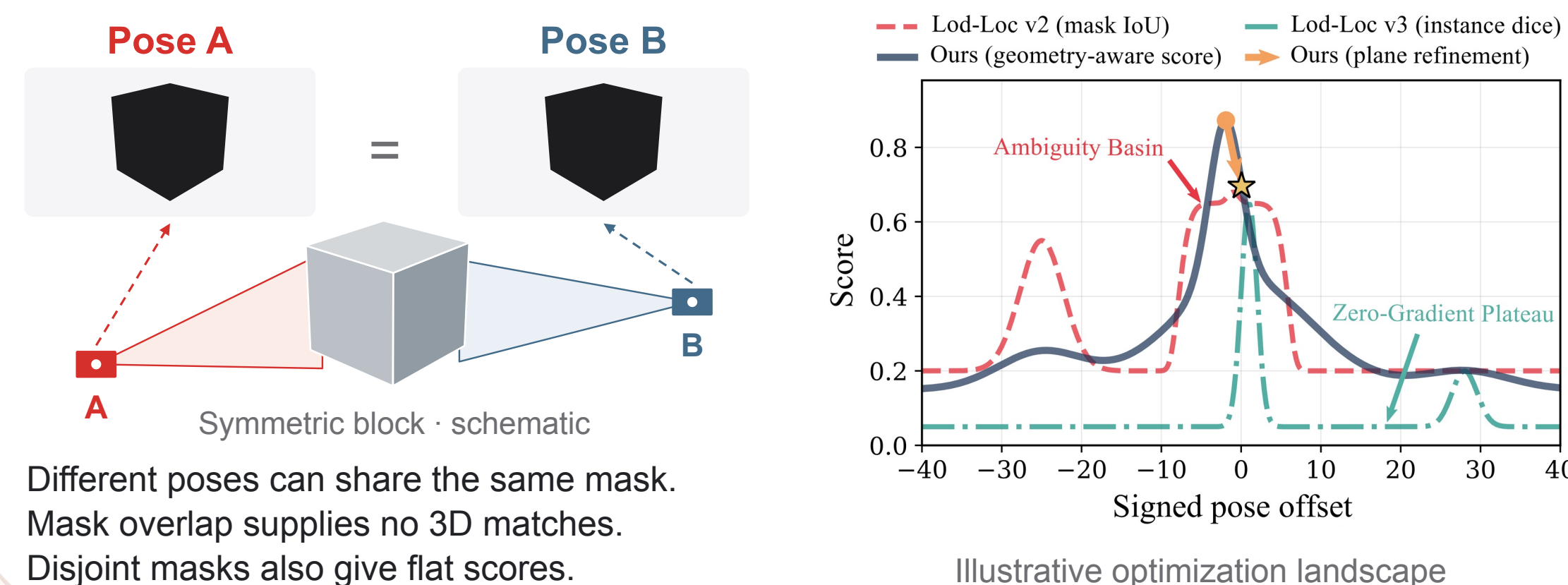
Why? Reliable positioning for UAV applications



Urban mapping, Building inspection, Urban navigation

Application illustrations

Limitation: silhouette ambiguity in LoD-Loc v2 [3] and LoD-Loc v3 [4]



Contributions & takeaway

- Planar grounding**
3D map patches $\leftrightarrow$ 2D image regions.
- Geometry-aware pose search**
Masks, edges and depth guide initialization.
- Off-the-shelf visual models**
No dataset-specific model fine-tuning.

Address silhouette ambiguity through 3D plane grounding.

TL;DR

We present PlanarCityLoc to locate a drone in a 3D city map: building outlines alone can fit different viewpoints. 3D planes help resolve this silhouette ambiguity; our results improve recall over off-the-shelf outline matching.

Method

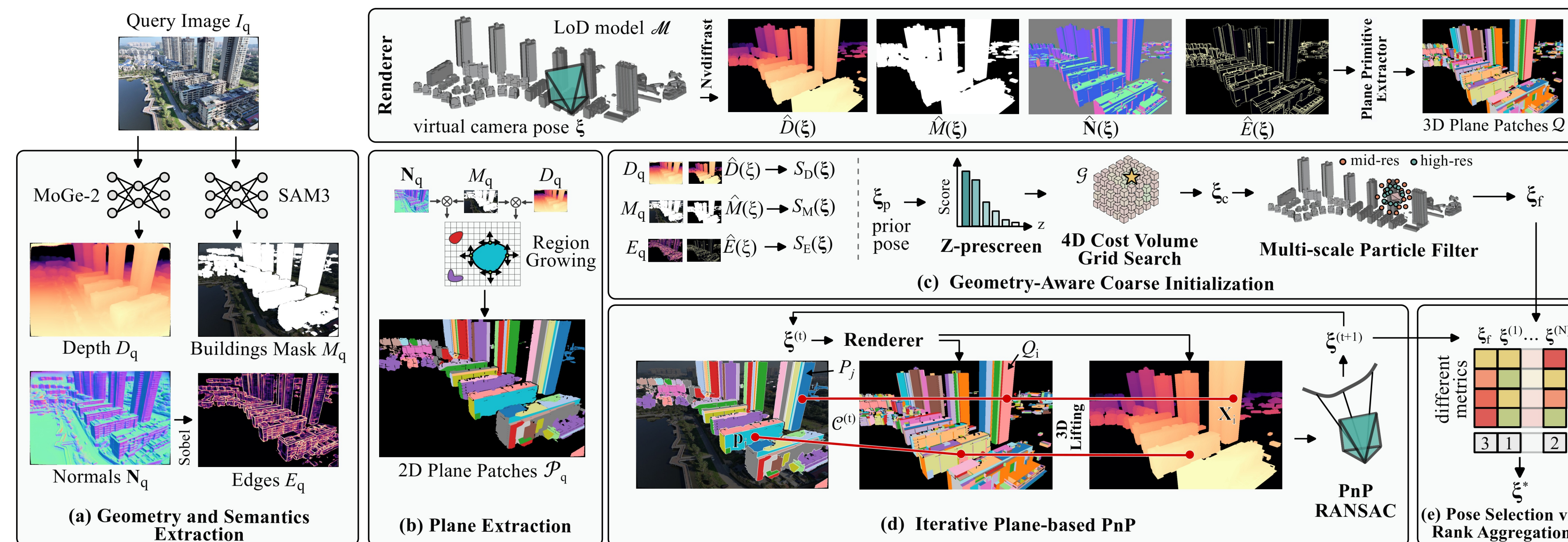
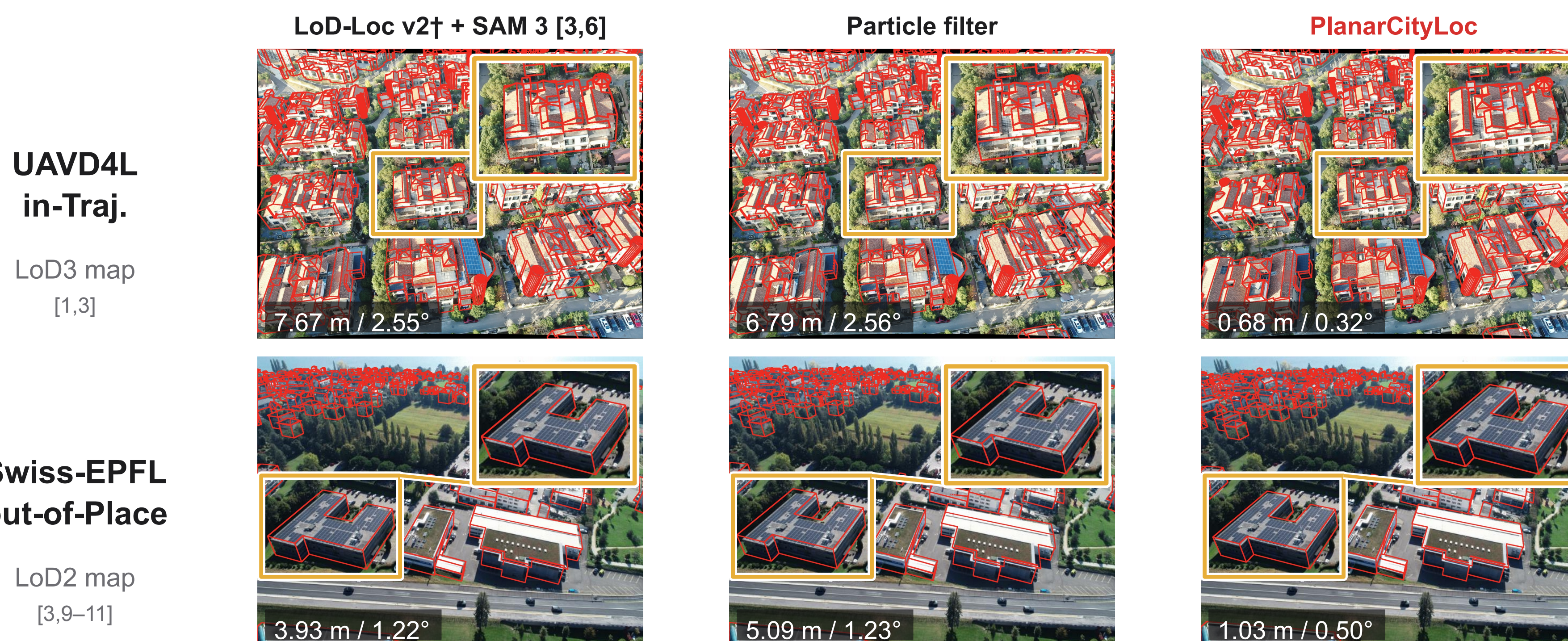


Image cues: SAM 3 [6] and MoGe-2 ViT-S [7]. Robust refinement: plane-based PnP–RANSAC [13].

Qualitative localization



Red: projected LoD edges. Labels: translation / rotation error.

Localization recall

Joint recall at 2 m / 2° (%) ↑

Method	UAVD4L-LoDv2 [1,3]		Swiss-EPFL-v2 [3,9–11]	
	in-Traj.	out-Traj.	in-Place	out-Place
Trainable methods				
LoD-Loc · trained [2]	49.6	54.2	36.8	14.2
LoD-Loc v2† · oracle GT [3]	57.2	64.9	74.9	43.3
LoD-Loc v3† · oracle GT [4]	54.5	41.3	56.3	40.1
Off-the-shelf methods				
MC-Loc · DINOv2 [5,12]	1.2	2.4	0.9	2.9
LoD-Loc v2† · SAM 3 [3,6]	35.7	64.3	9.4	17.4
Cost volume (ours)	27.6	34.9	15.4	11.3
+ particle filter (ours)	36.0	60.8	24.9	14.5
PlanarCityLoc	75.4	80.6	20.6	48.5

† Adapted re-runs: GT = oracle masks (not trained predictions); SAM 3 = off-the-shelf predictions. LoD-Loc [2]: published trained result. Map/search setups differ. Published trained v2/v3 results: see paper. Bold/underline: best/second within oracle or off-the-shelf runs, not across setups.

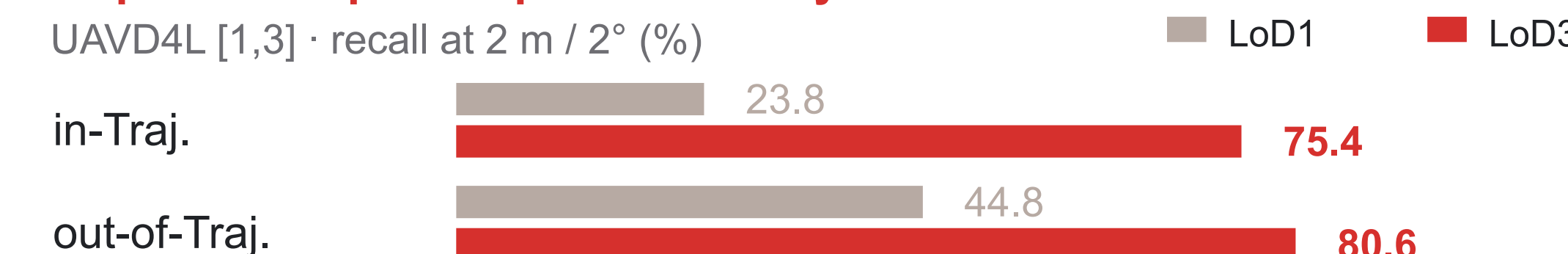
Pose accuracy

Split	3 m / 3° ↑	5 m / 5° ↑	T.e. (m) ↓	R.e. (°) ↓
UAV · in-Traj.	90.8	97.2	1.28	0.68
UAV · out-Traj.	95.1	99.0	1.26	0.69
Swiss · in-Place	39.0	59.7	3.70	1.90
Swiss · out-Place	72.3	89.4	1.98	1.21

PlanarCityLoc: recall (%) and median translation / rotation errors.

Ablations

Map detail improves pose accuracy



Depth and normals matter

Depth + normal predictor	in-Traj.	out-of-Traj.
Metric3D v2 · ViT-S [8]	33.0	55.0
MoGe-2 · ViT-S [7]	75.0	84.0

Geometry: 100 queries/split, SAM 3 [6] fixed. Map detail: full splits.

[1] Wu et al. UAVD4L: A Large-Scale Dataset for UAV 6-DoF Localization. 3DV, 2024.

[2] Zhu et al. LoD-Loc: Aerial Visual Localization using LoD 3D Map with Neural Wireframe Alignment. NeurIPS, 2024.

[3] Zhu et al. LoD-Loc v2: Aerial Visual Localization over Low Level-of-Detail City Models using Explicit Silhouette Alignment. ICCV, 2025.

[4] Peng et al. LoD-Loc v3: Generalized Aerial Localization in Dense Cities Using Instance Silhouette Alignment. CVPR, 2026.

[5] Trivigno et al. The Unreasonable Effectiveness of Pre-Trained Features for Camera Pose Refinement. CVPR, 2024.

[6] Canon et al. SAM 3: Segment Anything with Concepts. ICLR, 2026.

[7] Wang et al. MoGe-2: Accurate Monocular Geometry with Metric Scale and Sharp Details. NeurIPS, 2025.

[8] Hu et al. Metric3D v2: A Versatile Monocular Geometric Foundation Model for Zero-Shot Metric Depth and Surface Normal Estimation. TPAMI, 2024.

[9] Yan et al. CrossLoc: Scalable Aerial Localization Assisted by Multimodal Synthetic Data. CVPR, 2022.

[10] Doytchinov et al. CrossLoc Benchmark Datasets. Dryad, 2022. doi:10.5061/dryad.mgqnk991c.

[11] © swisstopo. swissBUILDINGS3D 3.0 Beta. Swiss LoD city-map geodata.

[12] Oquab et al. DINOv2: Learning Robust Visual Features without Supervision. TMLR, 2024.

[13] Fischler & Bolles. Random Sample Consensus. Communications of the ACM, 1981.

Application illustrations: AI-generated. All scientific figures and results use the paper's recorded sources.