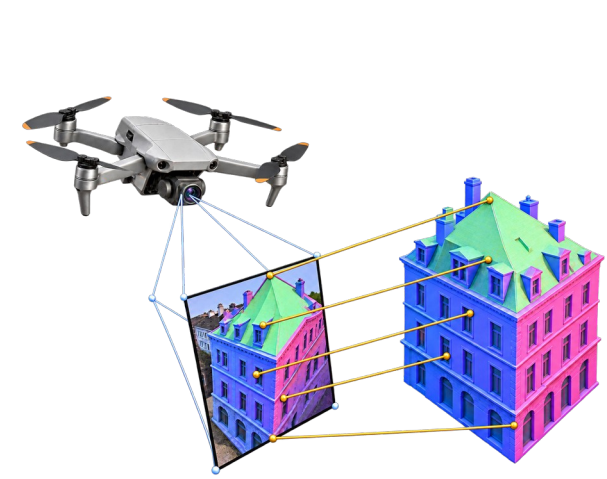




MatchCityLoc: Aerial Localization in LoD City Models via Surface Normal Matching

Oussema Dhaouadi^{1,2,3,4} Zuria Bauer³ Johannes Meier^{2,4} Marc Pollefeys^{3,5} Daniel Cremers^{2,4} Olaf Wysocki¹

¹ University of Cambridge ² Technical University of Munich ³ ETH Zürich

⁴ Munich Center for Machine Learning ⁵ Microsoft

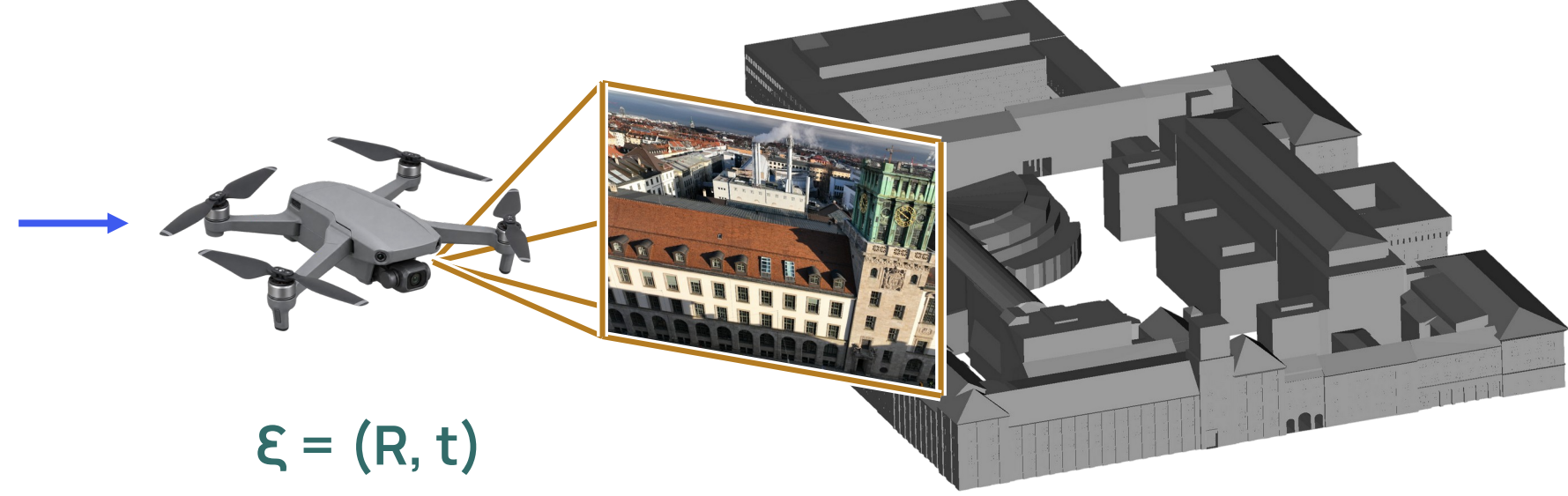


Goal & Motivation

Goal: refine a UAV camera's 6-DoF pose in a textureless city model.

UAV image

Camera pose in the LoD model



$$\xi = (R, t)$$

Known intrinsics + position-and-orientation prior

Potential Applications



Urban navigation



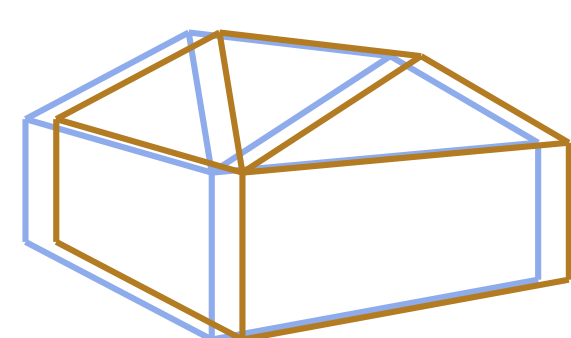
Building inspection



Urban mapping

Previous Work

LoD-Loc [1]



Predicted / projected edges

Wireframe alignment

Align predicted edges with projected 3D lines.

Limit: edge quality; dataset-specific training.

LoD-Loc v2/v3 [2, 9]
SemCityLoc [3]



Silhouette alignment

Score pose hypotheses by mask/surface overlap.

Limit: ambiguous or inaccurate masks.

PlanarCityLoc [4]



3D plane alignment

Align image-derived planes with model planes.

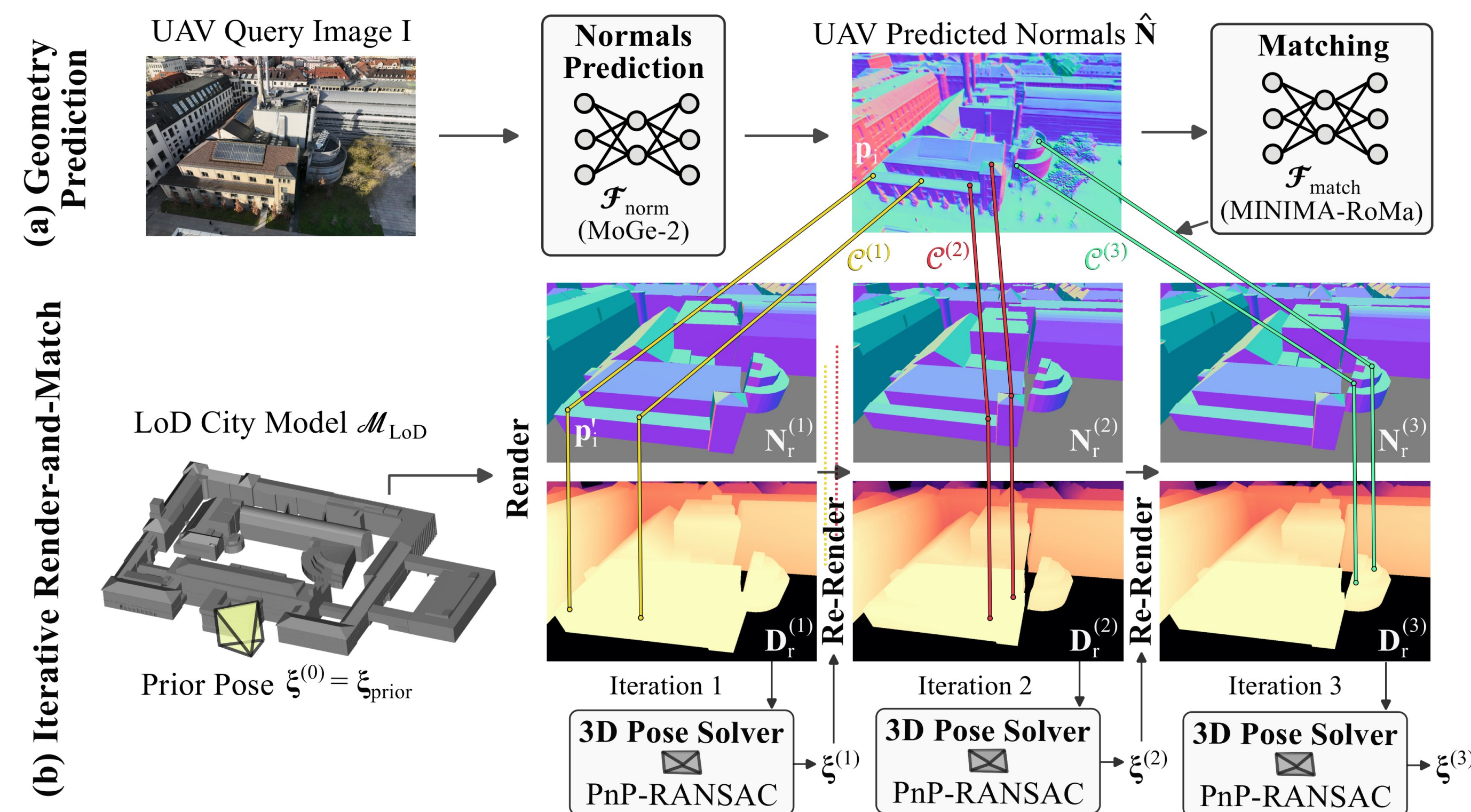
Limit: unreliable plane extraction or association.

Silhouette ambiguity: different poses can produce similar building outlines. Overlap scores provide no explicit 2D-3D correspondences.

TL;DR

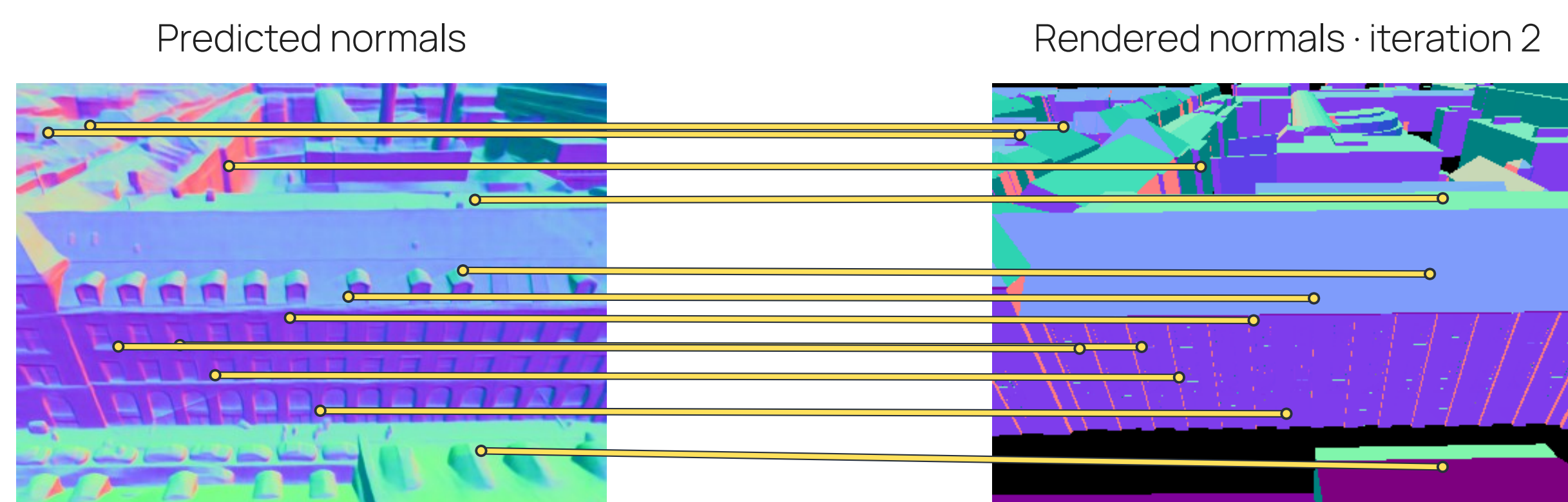
We present MatchCityLoc: surface-normal matching to refine drone poses in textureless city models. Prior silhouette methods score 2D overlap; we solve pose from 2D-3D matches to reduce ambiguity.

Render-Match-Refine



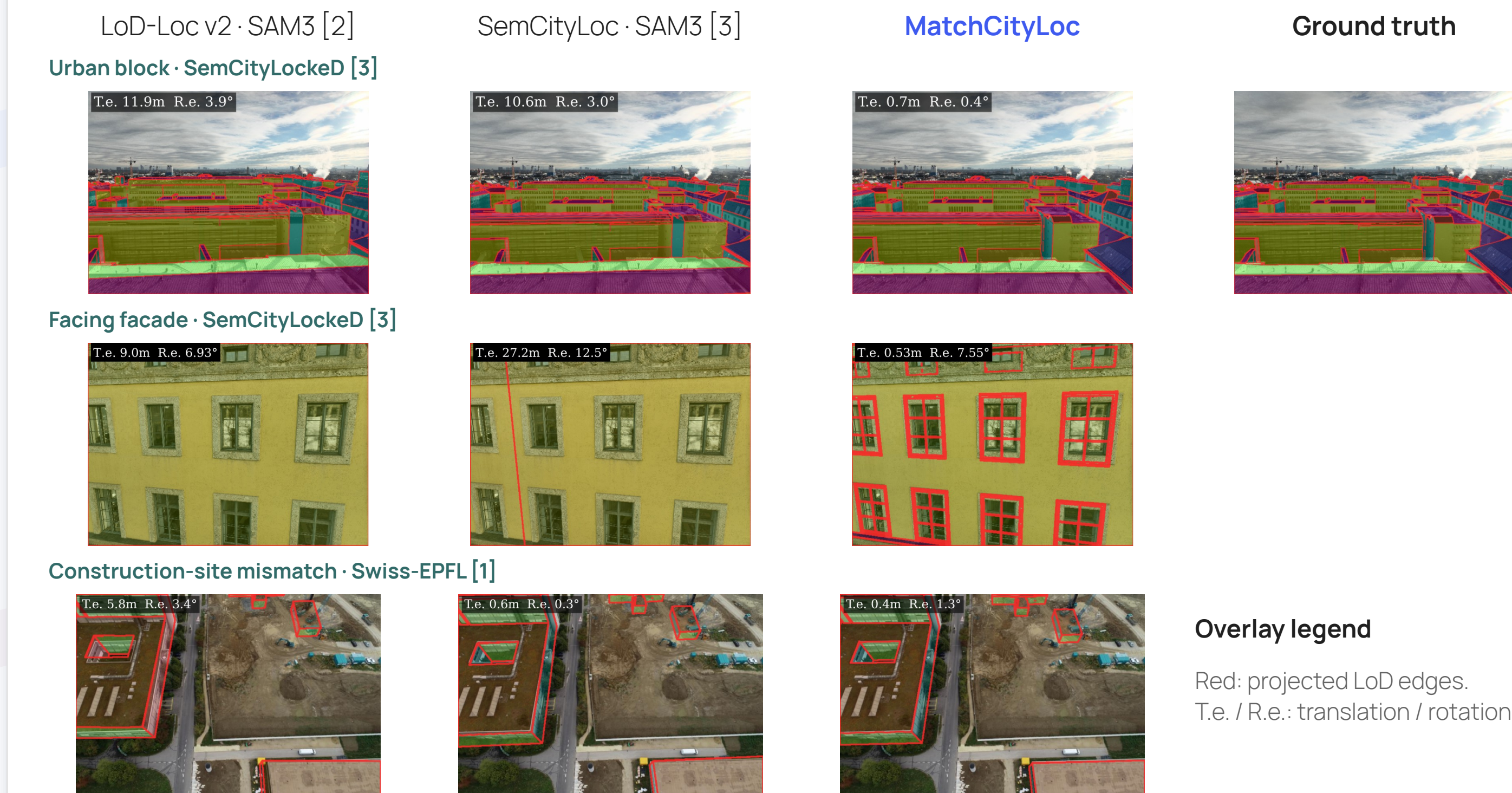
- 1 Predict normals with MoGe-2 [5].
 - 2 Render LoD normals/depth [8].
 - 3 Match with MINIMA (RoMa) [6, 7], lift to 3D, and solve with PnP.
- Repeat with a fixed update budget and validity guards.

Normal-to-Normal Matches



Matched normal-map pixels are lifted to 3D using rendered depth.

Qualitative Comparisons



Benchmark Results

Localization accuracy

Method	UAVD4L-LoD (LoD3) [1]		Swiss-EPFL (LoD2) [1]	
	in-Traj	out-Traj	in-Place	out-Place
LoD-Loc v2 · SAM3+ [2]	35.72 (57.17)	64.28 (64.92)	9.41 (74.86)	17.41 (43.27)
SemCityLoc · SAM3+ [3]	63.84 (94.76)	69.07 (92.20)	0.00 (70.51)	0.00 (72.82)
PlanarCityLoc [4]	75.40	80.60	20.60	48.50
MatchCityLoc (ours)	92.64	94.66	45.51	72.82

SemCityLocked [3]

Method	LoD2	LoD3
LoD-Loc v2 · SAM3+ [2]	37.89 (79.08)	30.42 (78.01)
SemCityLoc · SAM3+ [3]	49.73 (90.07)	49.63 (93.28)
MatchCityLoc (ours)	59.12	70.33

Off-the-shelf (GT-mask oracle; not a guaranteed upper bound). † Adapted re-runs.
Updates: UAVID4L-LoD 3 · Swiss-EPFL / SemCityLocked 5.

Why Normals?

Recall at 2 m / 2° (%) ↑

Query → model	in-Traj	out-Traj
RGB → Depth	57.61	61.22
RGB → Normals	77.81	83.90
Depth → Normals	73.38	79.24
Depth → Depth	83.42	88.32
Normals → Normals	93.64	95.48

UAVID4L (LoD3) [1], 3-update runs.

How Much LoD Detail?

SemCityLocked [3] · recall at 2 m / 2° (%) ↑

LoD1	14.09
LoD2	59.12
LoD3	70.33

Coarse geometry limits normal matching.