LoD-Loc: Aerial Visual Localization using LoD 3D Map with Neural Wireframe Alignment

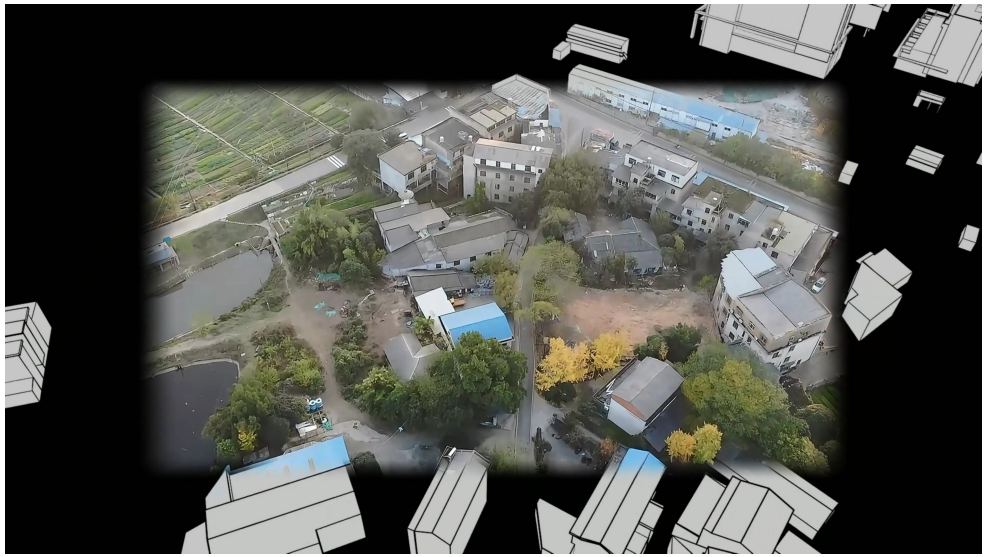
Juelin Zhu[1*] Shen Yan[1*] Long Wang[2] Shengyue Zhang[1]
Yu Liu[1] Maojun Zhang[1†]

[1] National University of Defense Technology
[2] SenseTime Research

Background

The Aerial Visual Localization Problem

6-DoF Pose Estimation
(x, y, z, yaw, pitch, roll)



Compute the camera **translation** and **orientation** from a given image

Background

Challenge: state-of-the-art visual localization methods rely on complex 3D representations

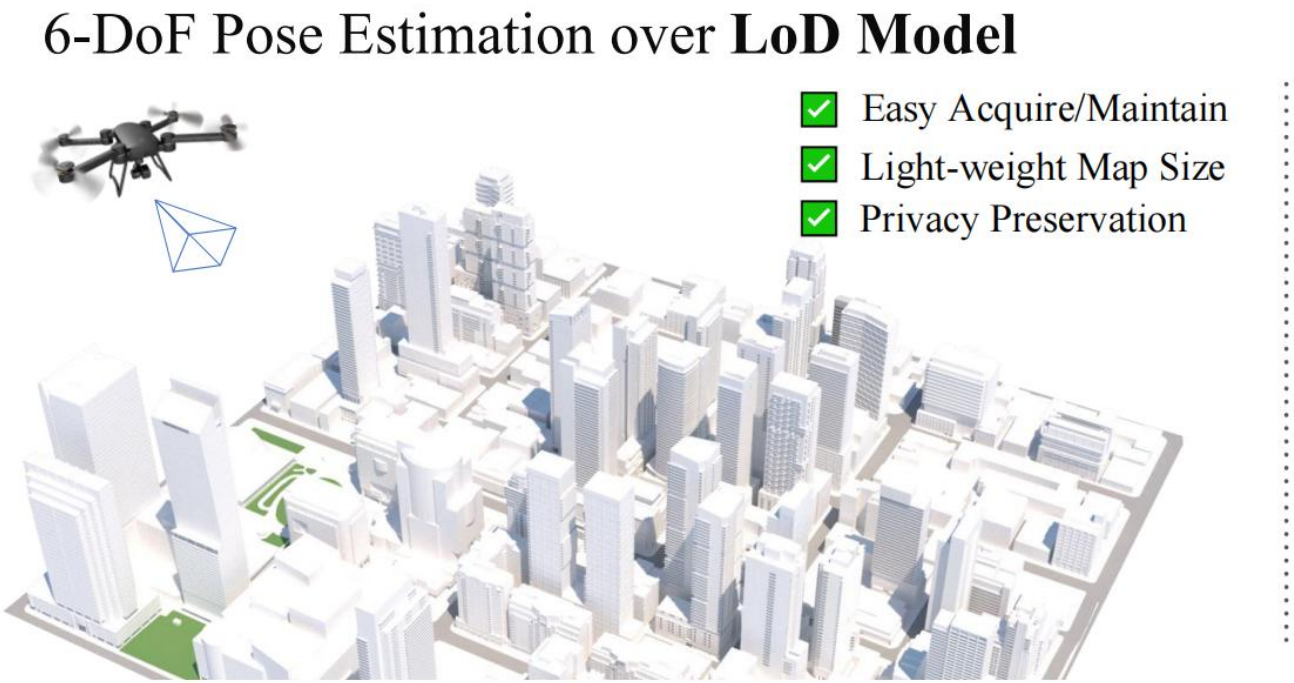
Structure-from-Motion

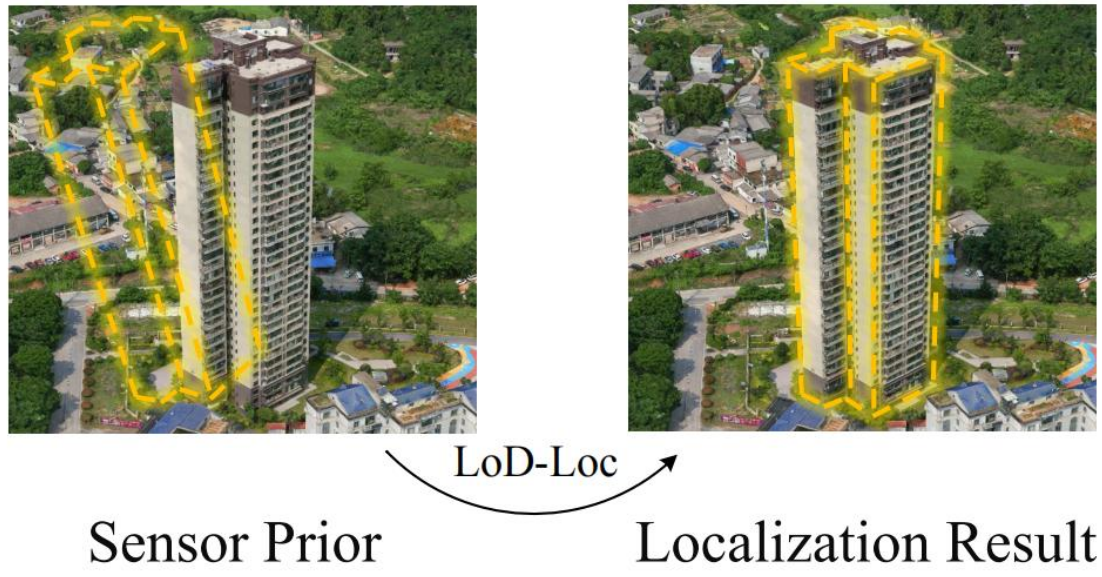
Mesh Model

⚠️ - **High Maintenance Costs:** Expensive to maintain on a global scale.
⚠️ - **Frequent Updates Required:** Needs constant updates to stay relevant.
⚠️ - **Costly to store:** Requires significant storage capacity due to the high data volume.
⚠️ - **Privacy Concerns:** High-resolution 3D maps reveal detailed information.

Motivation

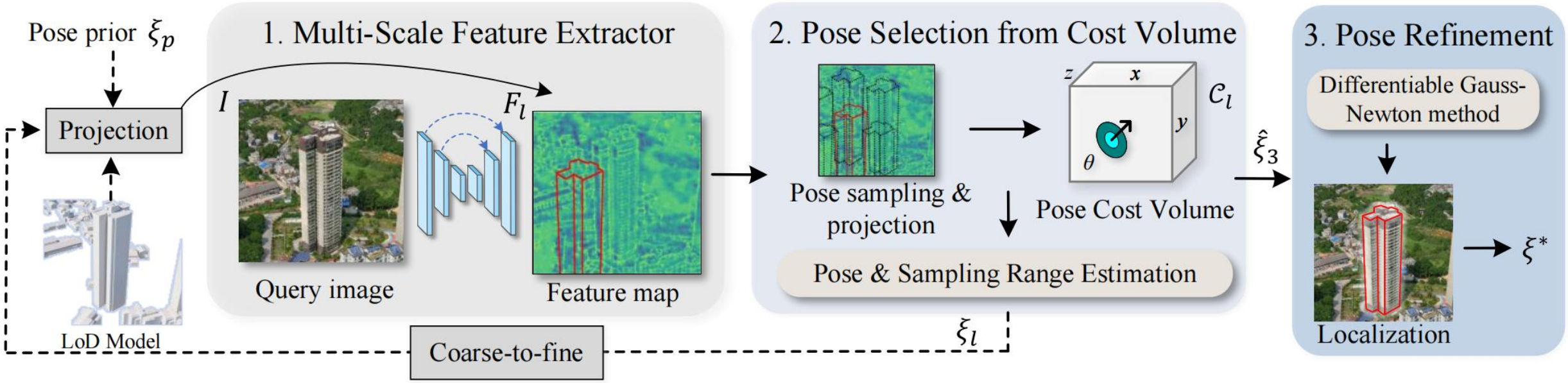
- Level of Detail (LoD) 3D models are **Easy Acquire/Maintain**, **Light-weight Size**, **Privacy Preservation**
- **Alignment**: Predicted building wireframes align with projections from the LoD 3D model when the pose is correct. 💡





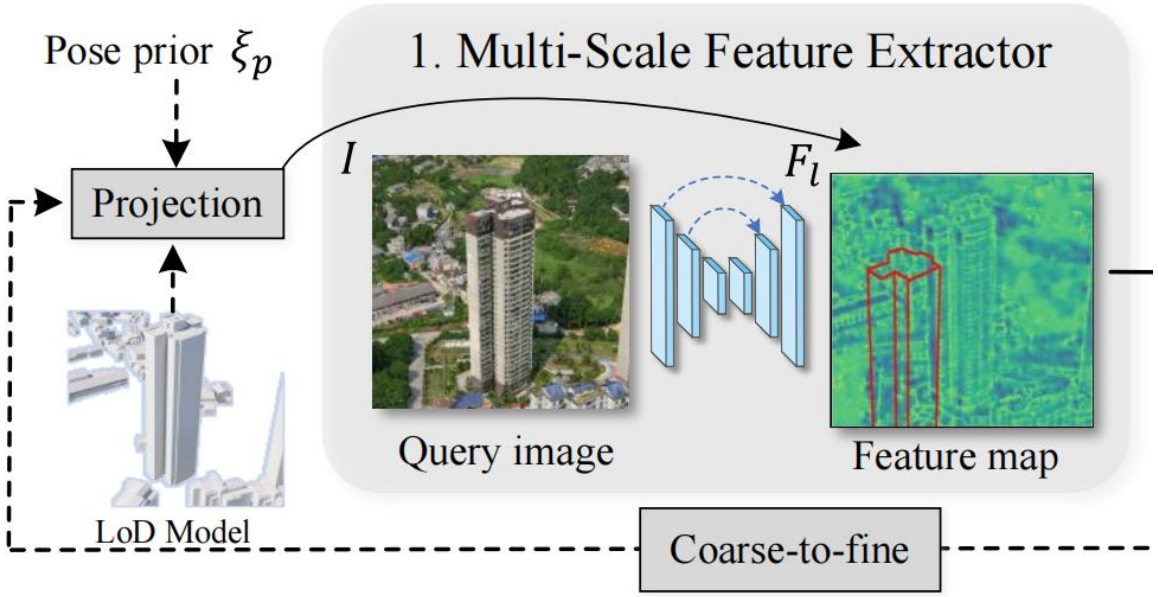
LoD-Loc

Pipeline overview



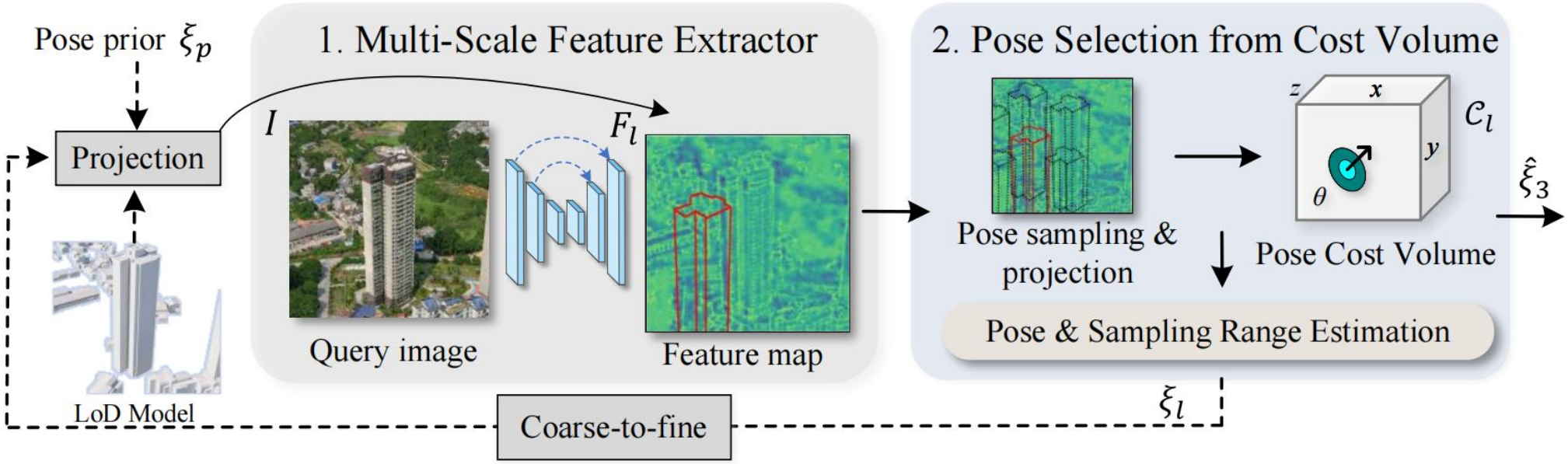
LoD-Loc

Pipeline overview



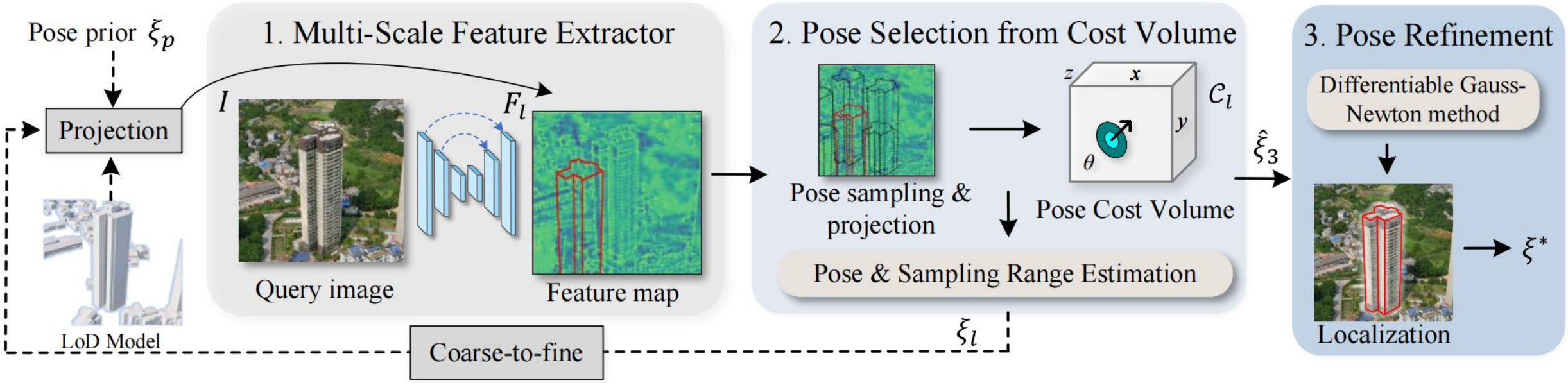
LoD-Loc

Pipeline overview



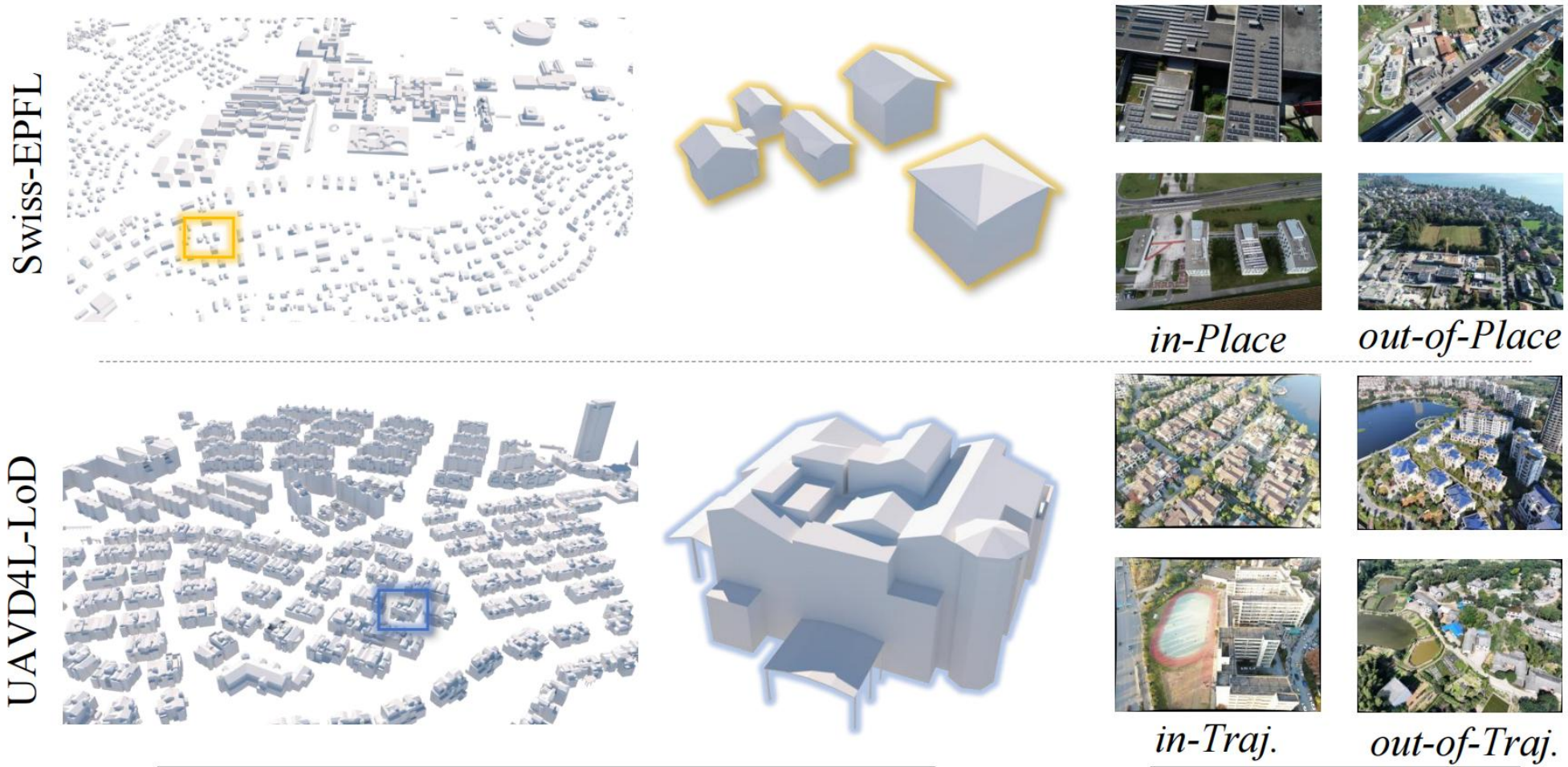
LoD-Loc

Pipeline overview



LoD-Loc

Dataset overview



Dataset

Query image collection

in-Traj.

out-of-Traj.

Table 5: **Key distinctions between the *in-Traj.* and *out-of-Traj.* sequences.**

Name	Capture device	Capture pitch angle	Capture height	Capture route
in-Traj.	DJI M300+H20t	0° or 45°	120m	Zig-zag flight on a selected region
out-of-Traj.	DJI Mavic3 Pro	30° $\sim$ 60°	90m $\sim$ 150m	Manually controlled flight on the map

Experiment

- Results over the UAVD4L-LoD dataset.

Table 2: **Quantitative comparison results over the UAVD4L-LoD dataset.**

Method		*in-Traj.*			*out-of-Traj.*		
		2m-2°	3m-3°	5m-5°	2m-2°	3m-3°	5m-5°
Sensor Priors		0	0	4.3	0	0	0.36
UAVD4L *Mesh model*	SIFT+NN	73.13	78.62	80.42	82.39	85.13	86.36
	SPP+SPG	91.71	92.02	92.14	93.43	93.70	93.80
	LoFTR	84.98	88.09	88.90	91.56	92.02	92.11
	e-LoFTR	84.47	88.21	88.96	91.06	91.93	92.02
	RoMA	**93.27**	**93.70**	93.77	95.03	95.53	95.53
CadLoc *LoD model*	SIFT+NN	0	0	0	0	0	0
	SPP+SPG	0	0	0	0	0	0
	LoFTR	0	0	0	0	0	0
	e-LoFTR	0.37	0.87	1.31	0.41	0.78	1.37
	RoMA	2.18	2.87	3.68	6.93	8.76	10.40
	SOLD2	0	0	0	0	0	0
	DeepLSD+SOLD2	0	0	0	0	0	0
	DeepLSD+GlueStick	0	0	0	0	0	0
Ours *LoD model*	no NWE	10.41	16.21	24.19	6.93	12.64	21.62
	no USR	70.39	85.47	95.32	82.62	94.71	97.63
	no $Refine$	51.31	76.06	86.78	74.27	97.95	99.36
	Full model	84.41	91.77	**96.95**	**95.94**	**99.00**	**99.36**

Experiment

- Results over the Swiss-EPFL dataset.

Table 3: **Quantitative comparison results over the Swiss-EPFL dataset**.

Method		*in-Place*			*out-of-Place*		
		2m-2°	3m-3°	5m-5°	2m-2°	3m-3°	5m-5°
Generated Priors		0	0	0.56	0	0	1.06
UAVD4L *Mesh model*	SIFT+NN	14.47	23.31	36.52	32.98	54.35	71.50
	SPP+SPG	34.83	60.39	77.25	77.04	89.71	92.35
	LoFTR	27.67	49.58	66.43	68.87	81.00	84.96
	e-LoFTR	37.64	60.96	76.40	81.53	91.03	93.93
	RoMA	45.98	**66.77**	**80.73**	**89.18**	**98.68**	**98.94**
CadLoc *LoD model*	SIFT+NN	0	0	0	0	0	0
	SPP+SPG	0	0	0	0	0	0
	LoFTR	0	0	0	0	0	0
	e-LoFTR	0	0.14	0.14	0	0	0.53
	RoMA	0.98	1.97	2.67	2.37	5.01	6.33
	SOLD2	0	0	0	0	0	0
	DeepLSD+SOLD2	0	0	0	0	0	0
	DeepLSD+GlueStick	0	0	0	0	0	0
Ours *LoD model*	no NWE	11.37	21.35	33.57	18.99	31.39	45.91
	no USR	42.42	58.29	71.21	31.40	48.81	70.45
	no $Refine$	36.10	58.01	76.97	18.21	39.31	66.23
	Full model	**48.60**	65.31	79.78	37.73	57.26	77.57

Thanks for listening

Paper link: https://arxiv.org/abs/2410.12269
Project link: https://victorzoo.github.io/LoD-Loc.github.io/