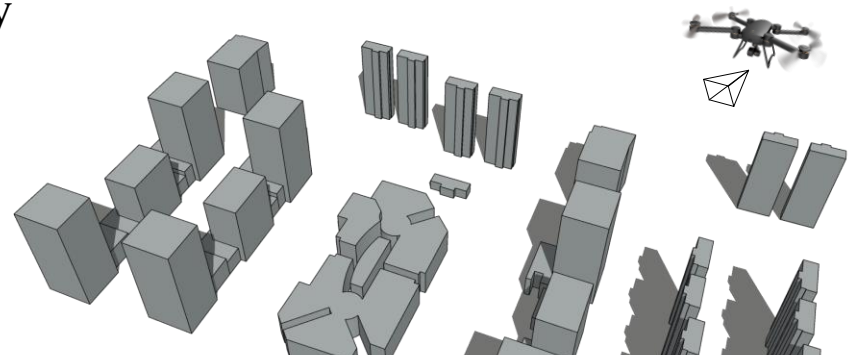
LoD-Loc v2: Aerial Visual Localization over Low Level-of-Detail City Models using Explicit Silhouette Alignment

Juelin Zhu[1] Shuaibang Peng[1] Long Wang[2] Hanlin Tan[1]
Yu Liu[1] Maojun Zhang[1] Shen Yan[1*]

[1] National University of Defense Technology
[2] Westlake University



Background

The Aerial Visual Localization Problem

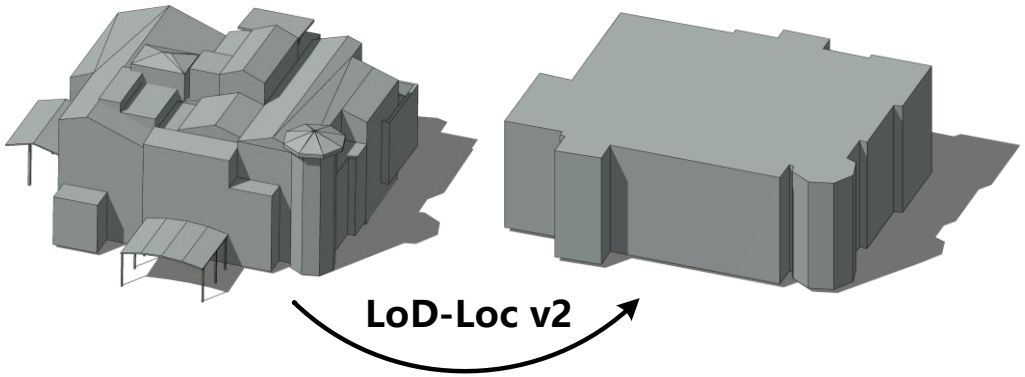
6-DoF Pose Estimation
(x, y, z, yaw, pitch, roll)

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

Background

Challenge: state-of-the-art visual localization methods rely on Level-of-Detail(LoD) City Models

LoD3 Model



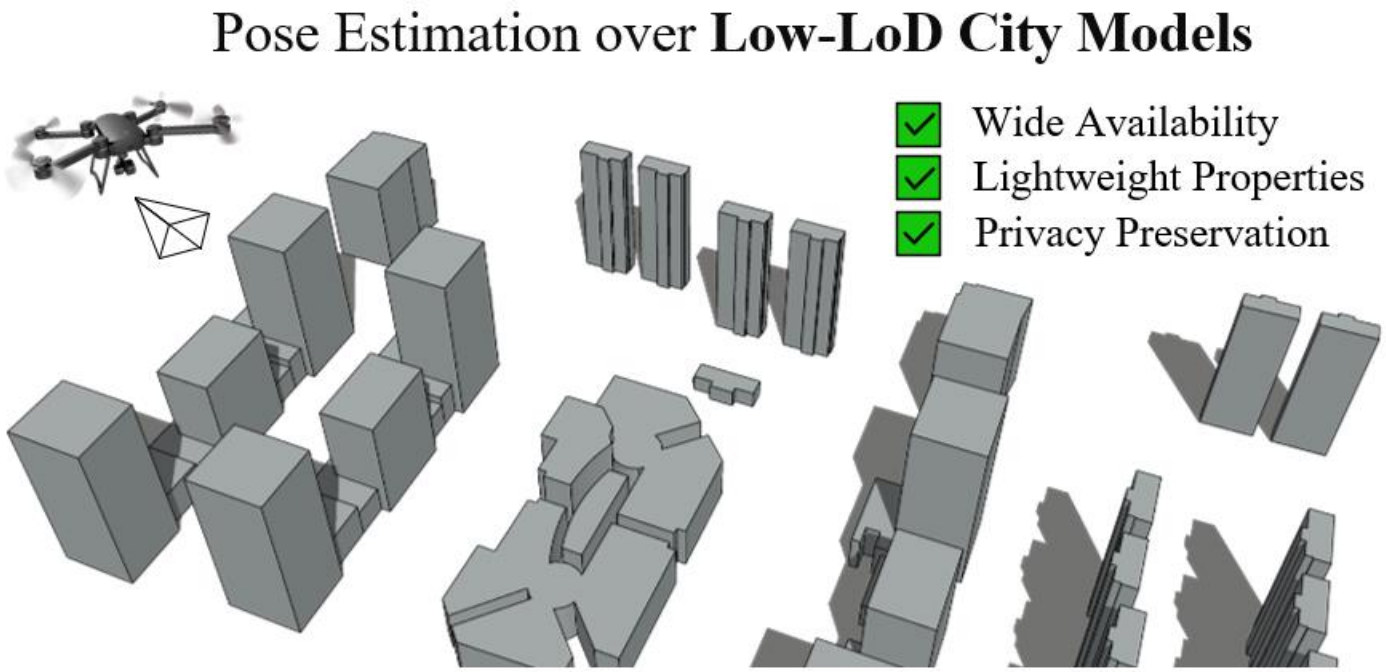
LoD3 LoD1

LoD1 vs. LoD2/3

- **Wide availability:** Most openly LoD1, not LoD3/LoD2.
- **Lightweight Properties:** Lower cost to more accurace localization.
- **Privacy Perservation:** LoD1 lose significant structural details to LoD3/LoD2.

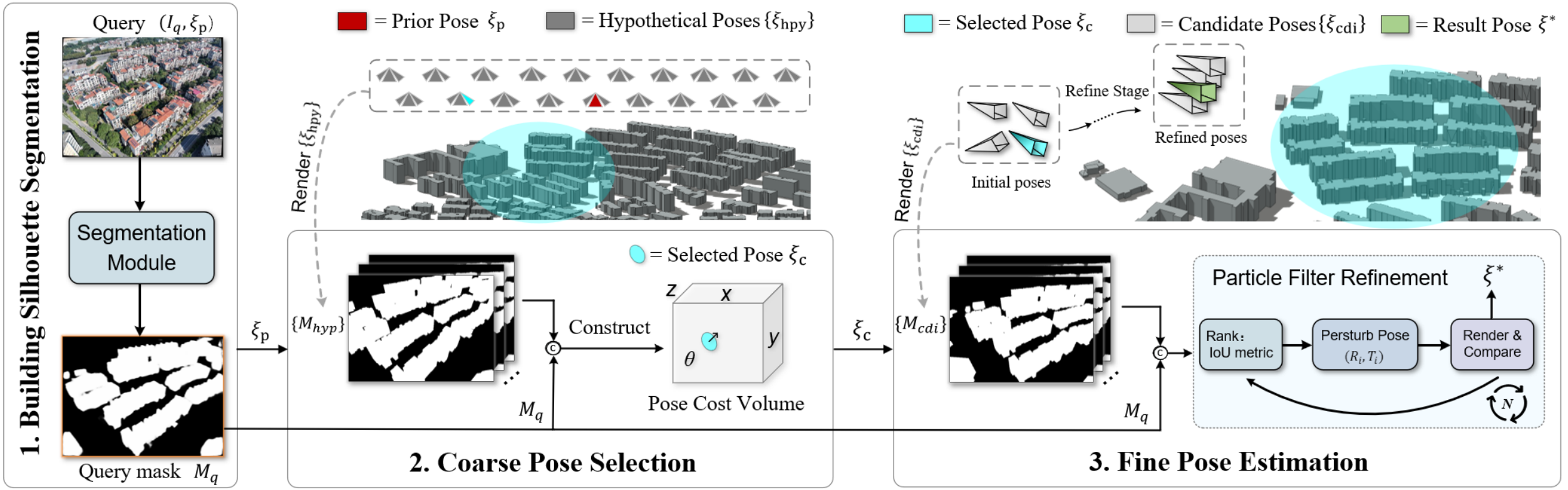
Motivation

- LoD1 3D models are more **Wide Availability**, **Lightweight Properties**, **Privacy Preservation**
- **Coarse-to-fine**: Predicted silhouette align with projected silhouette from the LoD model when the pose is correct.



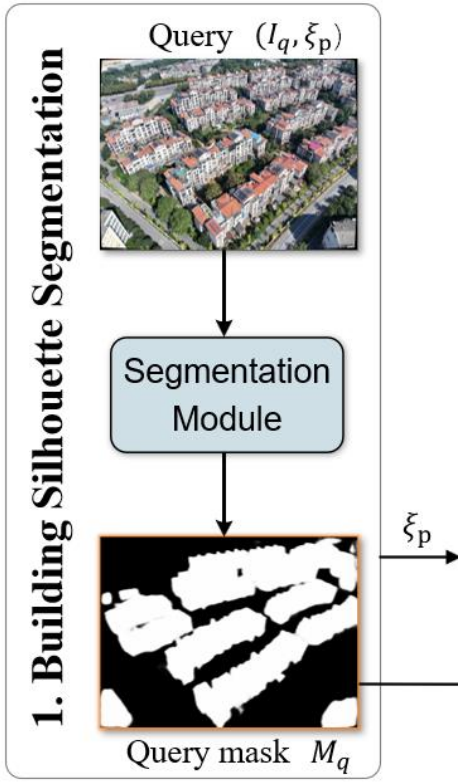
LoD-Loc v2

Pipeline overview



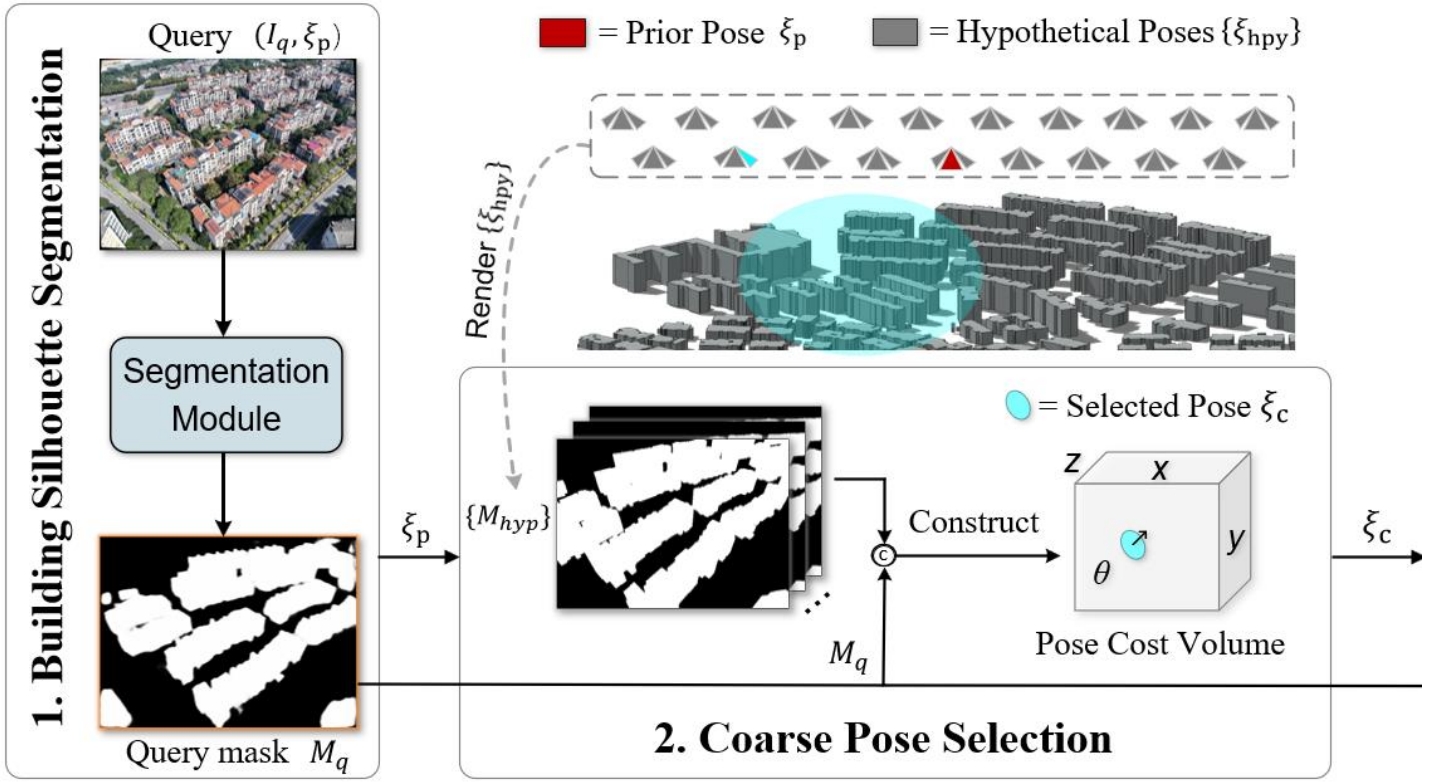
LoD-Loc v2

Pipeline overview



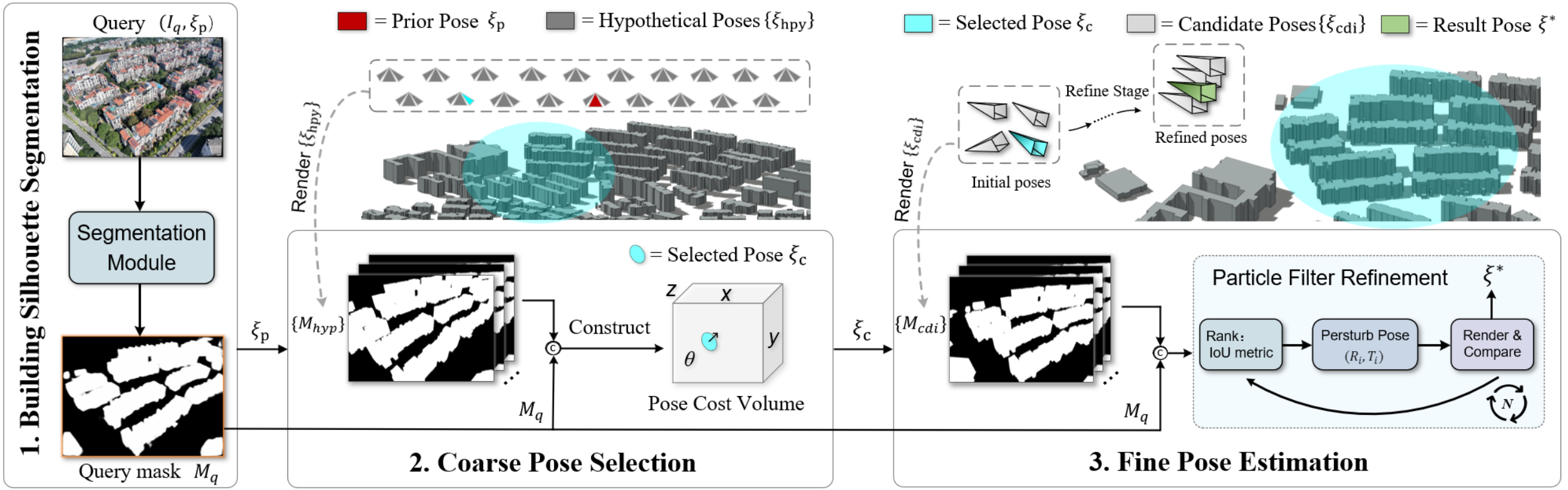
LoD-Loc v2

Pipeline overview



LoD-Loc v2

Pipeline overview

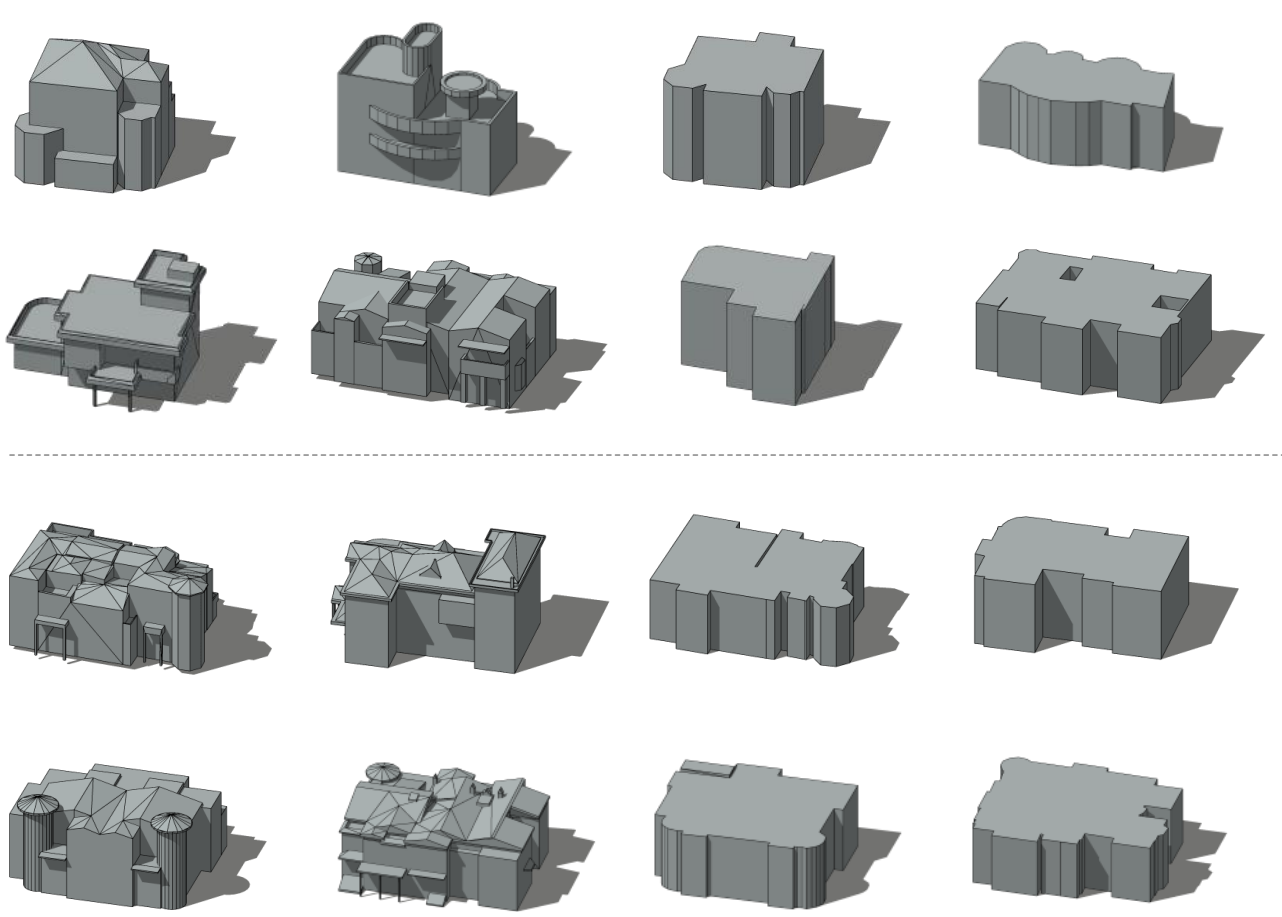


LoD-Loc v2

Dataset overview

Swiss-EPFLv2

UAVD4L-LoDv2



in-Place *out-of-Place*

in-Traj. *out-of-Traj.*

LoD3 models with details

LoD1 models with details

Query samples

Dataset

Query image collection

UAVD4L-LoD v2

Swiss-EPFLv2

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

Table 2. Differences between the *in-Traj.* and *out-of-Traj.* sequences.

Experiment

- Results over the UAVD4L-LoDv2 dataset.

Method		*in-Traj.*				*out-of-Traj.*			
		2m-2°	3m-3°	5m-5°	T.e./R.e.	2m-2°	3m-3°	5m-5°	T.e./R.e.
Prior		0	0	4.3	6.48/1.63	0	0	0.36	11.1/0.92
UAVD4L [88] *Texture model*	SIFT+NN	73.13	78.62	80.42	1.13/0.44	82.39	85.13	86.36	0.87/0.29
	SPP+SPG	91.71	92.02	92.14	0.79/0.29	93.43	93.70	93.80	0.74/0.19
	LoFTR	84.98	88.09	88.90	0.81/0.29	91.56	92.02	92.11	0.79/0.20
	e-LoFTR	84.47	88.21	88.96	0.96/0.35	91.06	91.93	92.02	0.86/0.22
	RoMA	93.27	93.70	93.77	0.78/0.25	95.03	95.53	95.53	0.73/0.22
CAD-Loc [60] *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	0	-	0	0	0	-
	RoMA	0	0	0	-	0	0	0	-
MC-Loc [82] *LoD model*	DINOv2	1.20	4.10	17.40	8.29/2.58	2.40	7.40	26.10	7.02/2.29
	RoMa	0.10	0.60	3.30	10.6/8.60	0.20	0.90	3.30	16.9/3.88
LoD-Loc [99] *LoD model*	-	49.56	71.82	89.09	3.32/1.48	54.20	75.05	89.51	3.33/1.18
LoD-Loc v2 *LoD model*	no refine	0	0	23.38	6.19/0.67	11.68	29.88	51.14	4.78/0.92
	no select	93.50	98.40	99.50	0.74/0.17	90.50	94.80	96.90	0.77/0.16
	Full	**93.70**	**98.40**	**99.50**	**0.72/0.15**	**97.90**	**99.80**	**100.00**	**0.71/0.14**

Table 2. **Quantitative comparison results of different methods over UAVD4L-LoDv2 dataset.** T.e. and R.e. denote median translation error (m) and median rotation error (°), respectively,

Experiment

- Results over the Swiss-EPFLv2 dataset.

Method		*in-Place.*				*out-of-Place.*			
		2m-2°	3m-3°	5m-5°	T.e./R.e.	2m-2°	3m-3°	5m-5°	T.e./R.e.
Prior		0	0	0.56	17.6/3.87	0	0	1.06	17.9/3.94
UAVD4L [88] *Texture model*	SIFT+NN	15.17	23.74	35.11	2.57/1.54	32.98	54.35	71.50	2.76/1.59
	SPP+SPG	33.85	56.32	72.75	2.57/1.54	77.04	89.71	92.35	1.12/1.17
	LoFTR	26.40	46.21	62.22	3.06/1.86	68.87	81.00	84.96	1.12/1.08
	e-LoFTR	37.64	60.96	76.40	2.38/1.45	81.53	91.03	93.93	0.91/1.08
	RoMA	45.95	66.77	80.73	2.08/1.29	**89.18**	**89.68**	**98.84**	**0.77**/1.04
CAD-Loc [60] *LoD model*	*same**	**0**	**0**	**0**	-	**0**	**0**	**0**	-
MC-Loc [82] *LoD model*	DINOv2	0.90	4.40	17.50	8.18/2.54	2.90	9.00	30.20	6.23/1.97
	RoMa	0.20	1.20	4.80	9.80/2.65	0.70	2.10	11.5	10.3/3.97
LoD-Loc [99] *LoD model*	-	36.79	50.56	69.77	2.87/1.78	14.24	31.39	59.89	8.73/2.78
LoD-Loc v2 *LoD model*	no refine	0.56	3.73	20.79	7.37/3.76	0.53	2.11	11.35	8.92/3.90
	no select	52.10	72.10	88.30	1.90/0.89	31.10	55.90	81.30	2.73/0.73
	Full	**54.20**	**74.60**	**92.00**	**1.83/0.85**	31.40	58.53	86.30	2.64/**0.73**

Table 3. **Quantitative comparison results of different methods over Swiss-EPFLv2 dataset.** The *same** indicates that the variants are identical to those in Tab. 2. T.e. and R.e. have the same meanings as those in Tab. 2.

Thanks for listening

Paper link: https://arxiv.org/abs/2507.00659
Project link: https://github.com/VictorZoo/LoD-Loc-v2