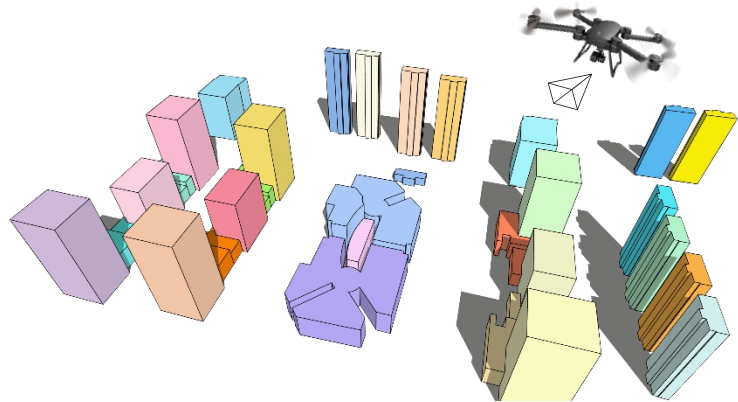
LoD-Loc v3: Generalized Aerial Localization in Dense Cities using Instance Silhouette Alignment

Shuaibang Peng[1] Juelin Zhu[1†] Xia Li[1] Kun Yang[2]
Yu Liu[1] Maojun Zhang[1] Shen Yan[1†]

[1] National University of Defense Technology
[2] Northwestern Polytechnical University

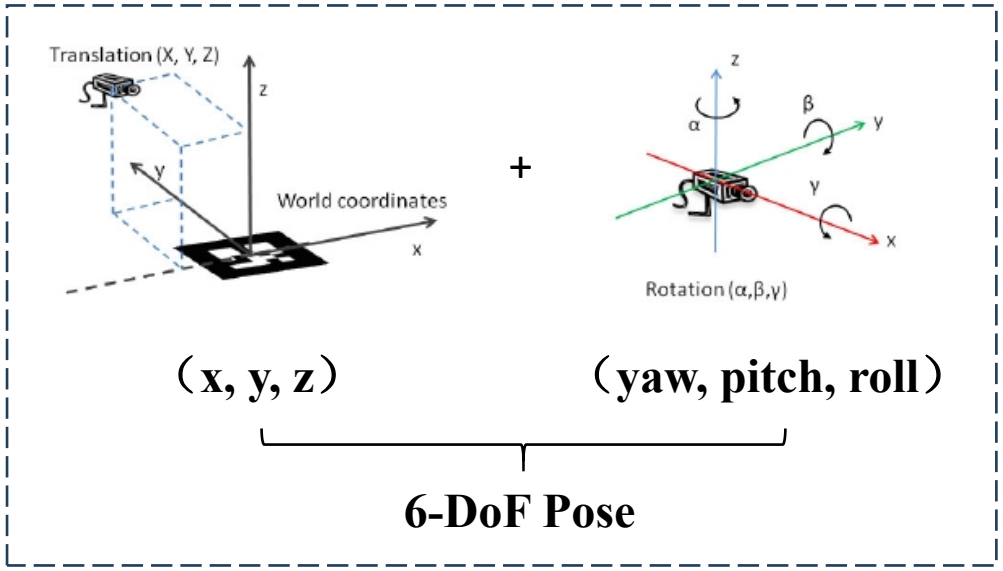


Background

The Aerial Visual Localization Problem

Query image

Compute



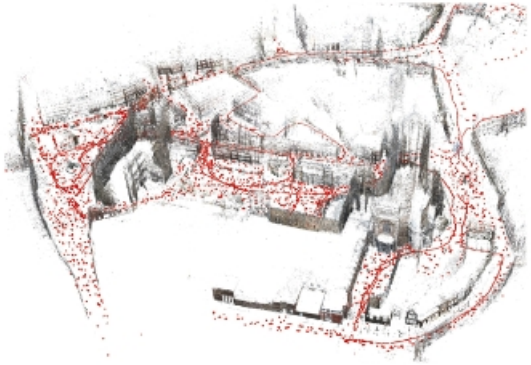
Input

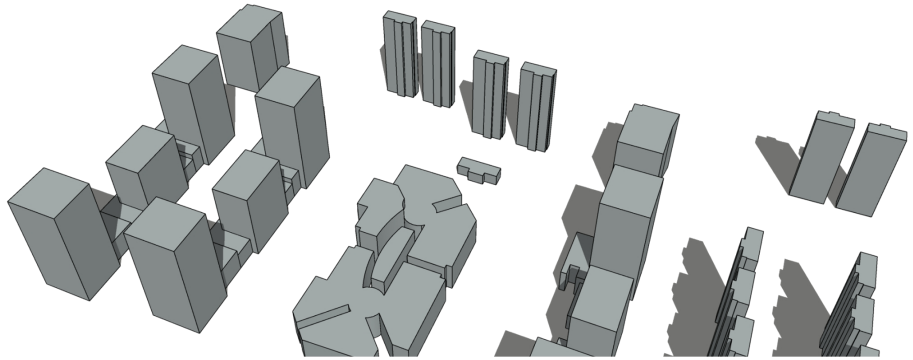
Output

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

Background

Why Localize over LoD City Models?





Traditional 3D Maps

LoD City Models

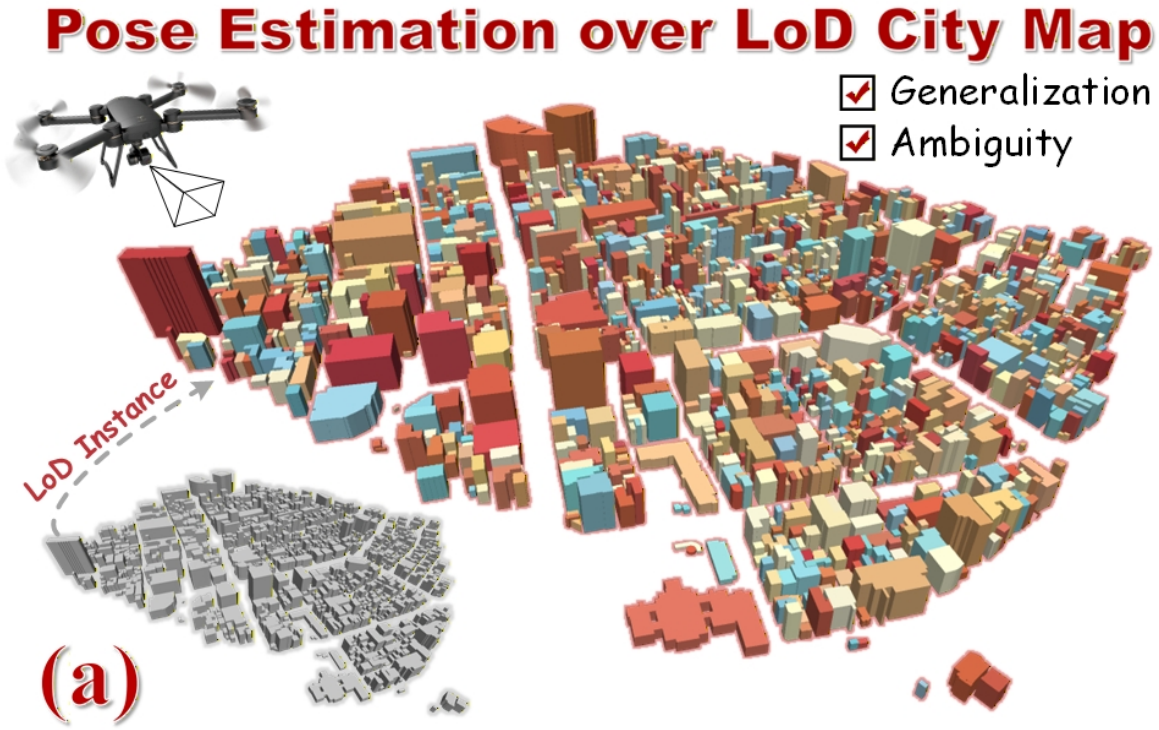
SFM/Mesh vs. LoD

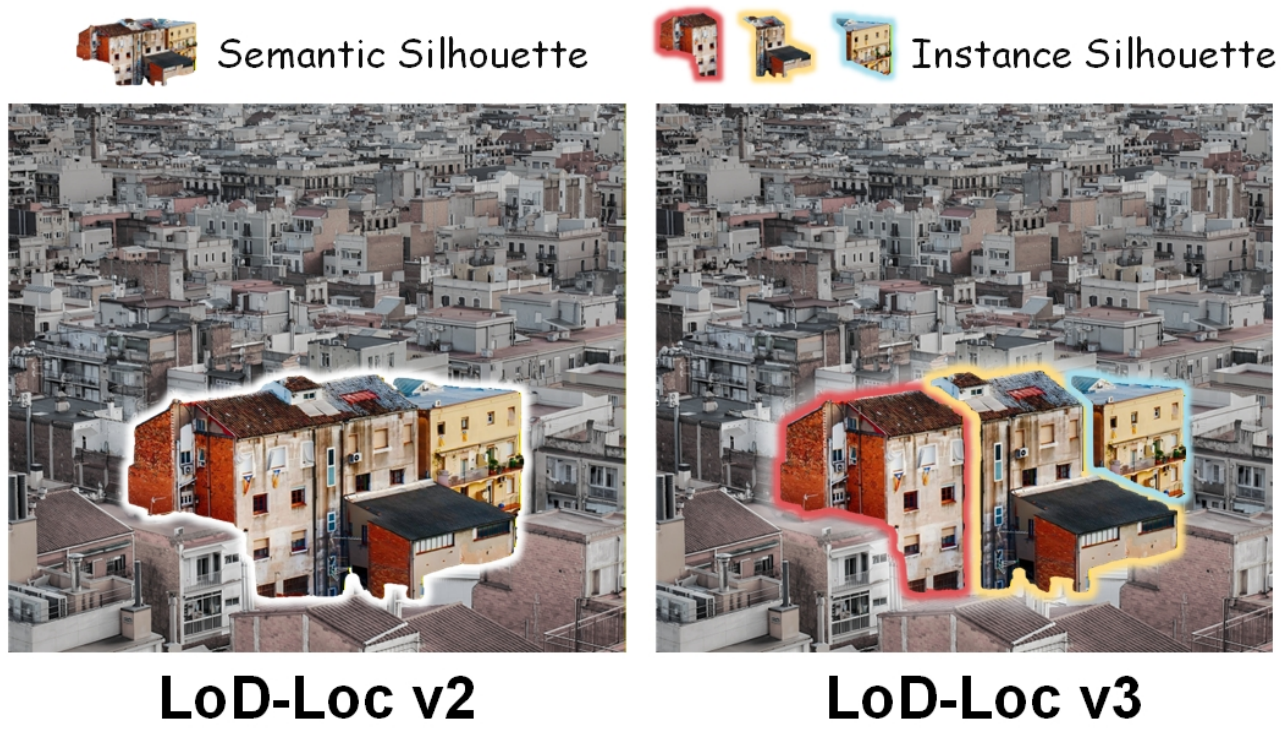
- **Lightweight**
- **Easier to acquire and maintain**
- **Increasingly available worldwide**
- **Privacy Perservation**

Limitations of LoD-Loc v2

- Poor cross-scene generalization
- Ambiguity in dense urban scenes

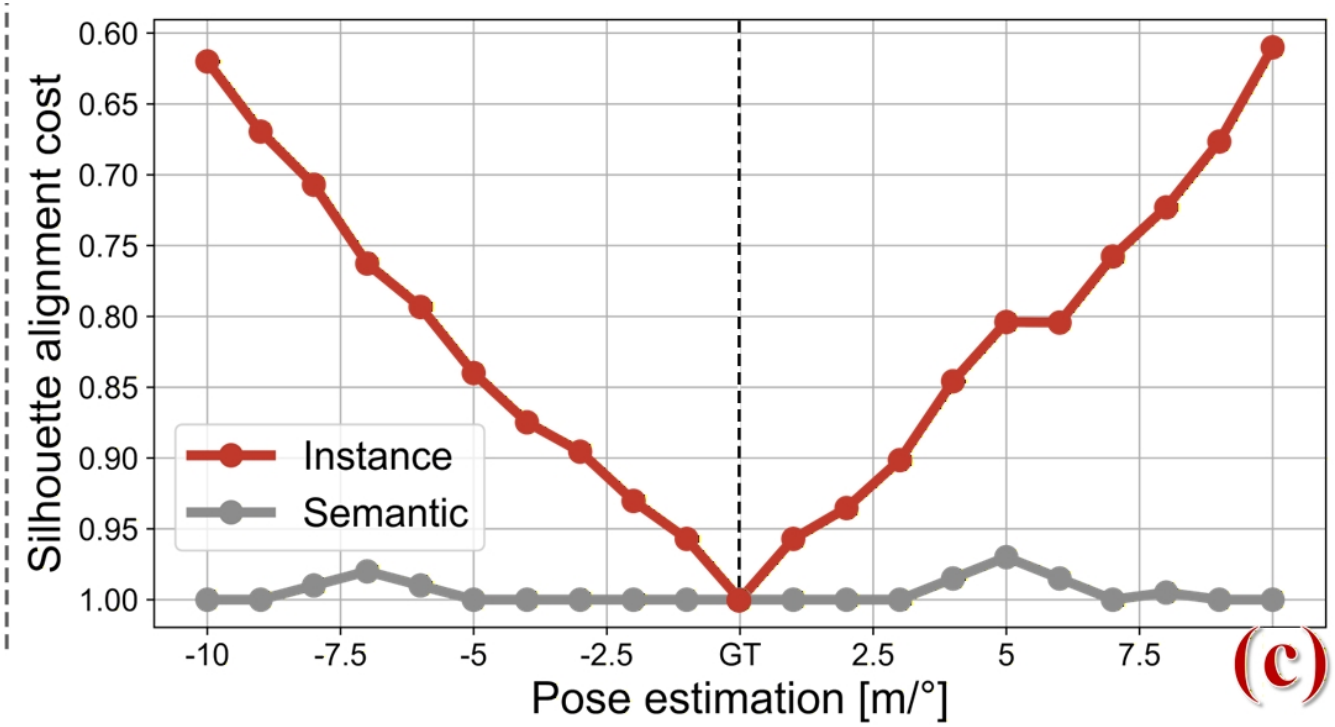
How can we localization generalize better and work reliably in dense areas?





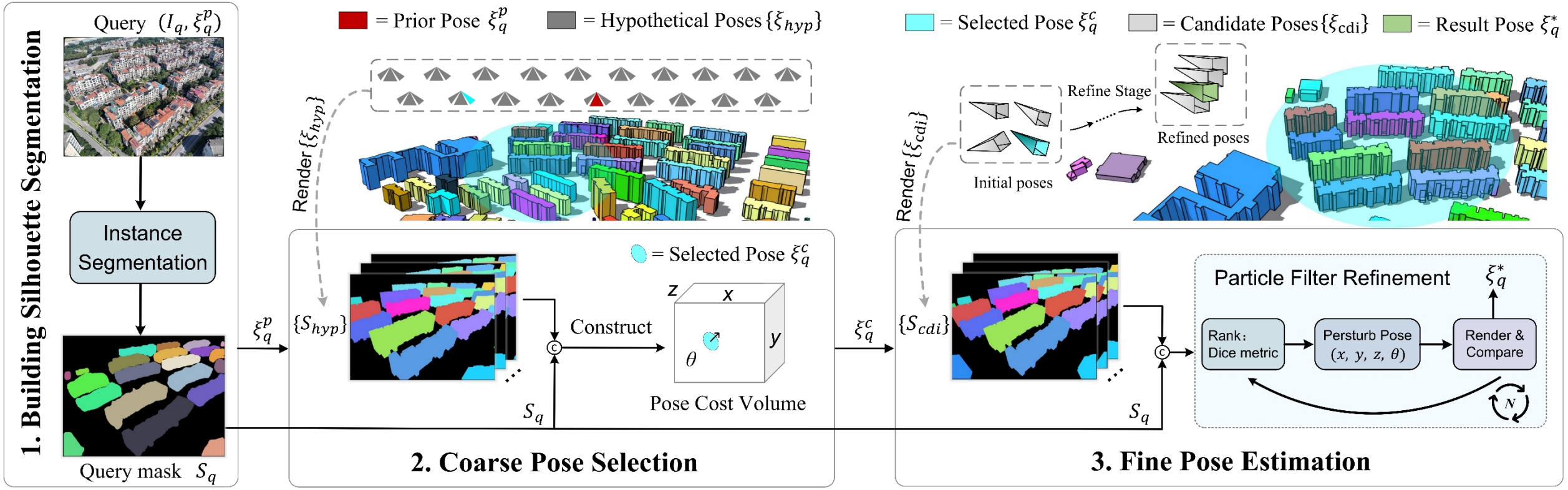
Key Ideas of LoD-Loc v3

- **Synthetic Dataset**
  Construct the largest aerial instance segmentation dataset
- **Instance Alignment** 💡
  Shift from semantic silhouettes to instance silhouettes



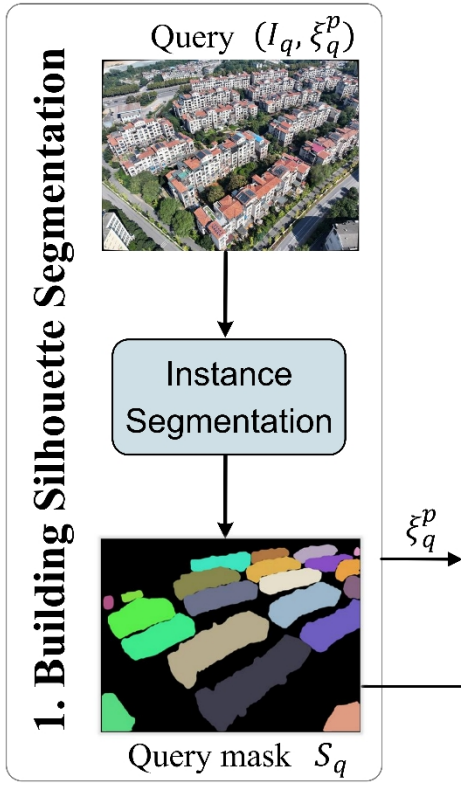
LoD-Loc v3

Pipeline overview



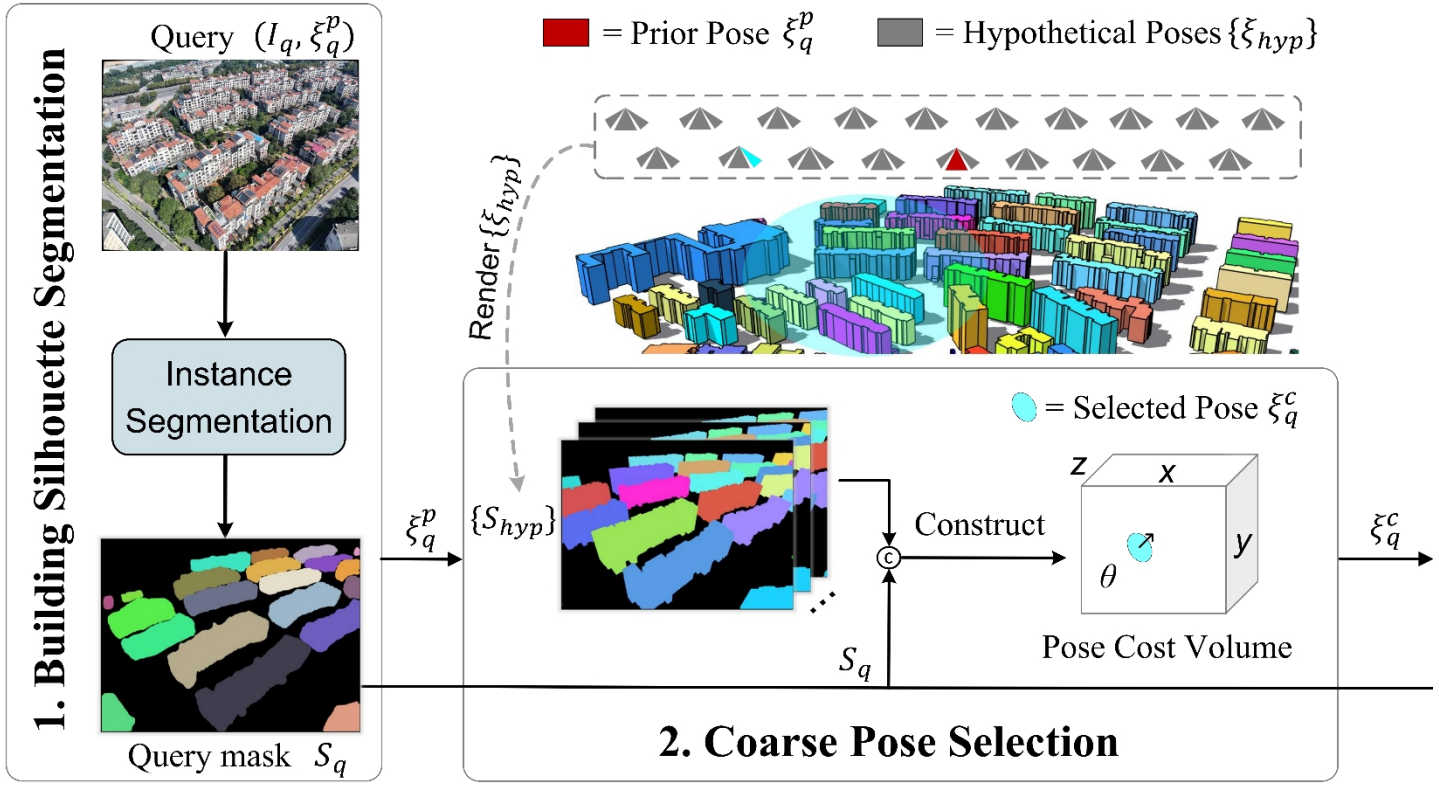
LoD-Loc v3

Pipeline overview



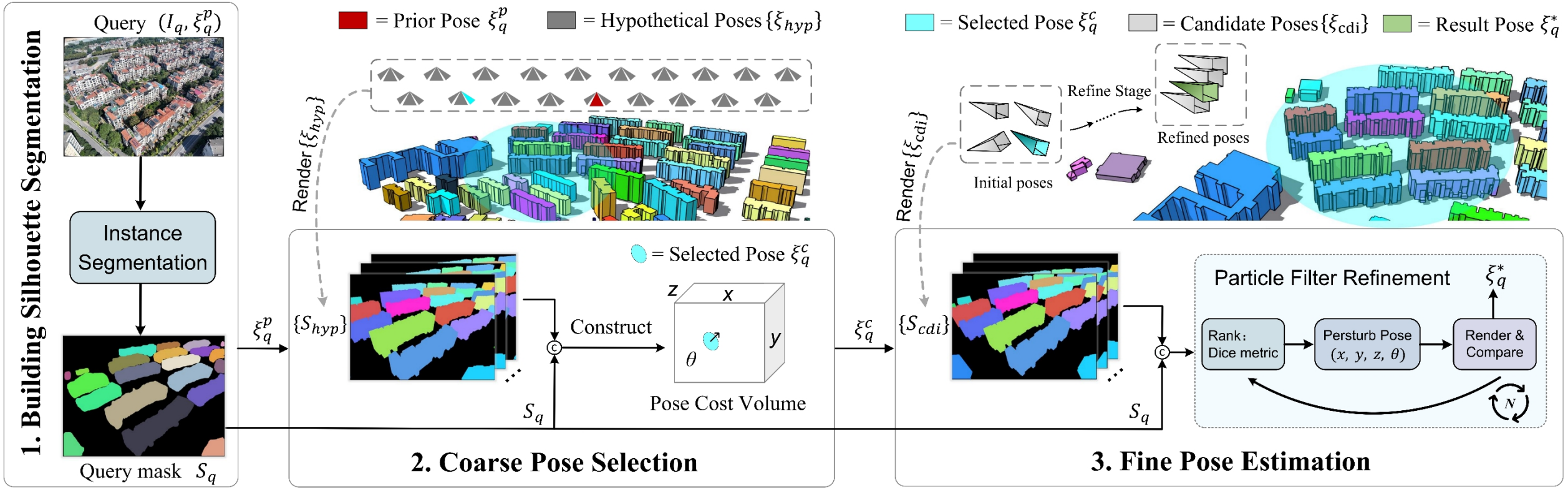
LoD-Loc v3

Pipeline overview



LoD-Loc v3

Pipeline overview



Dataset Generation Pipeline

- **UE5 + Cesium + Google 3D Tiles**
  Generate photorealistic RGB UAV images
- **AirSim** Control camera pose, altitude, trajectory, view angle
- **Instanced LoD models + OpenSceneGraph**
  Render aligned instance masks with the same camera parameters

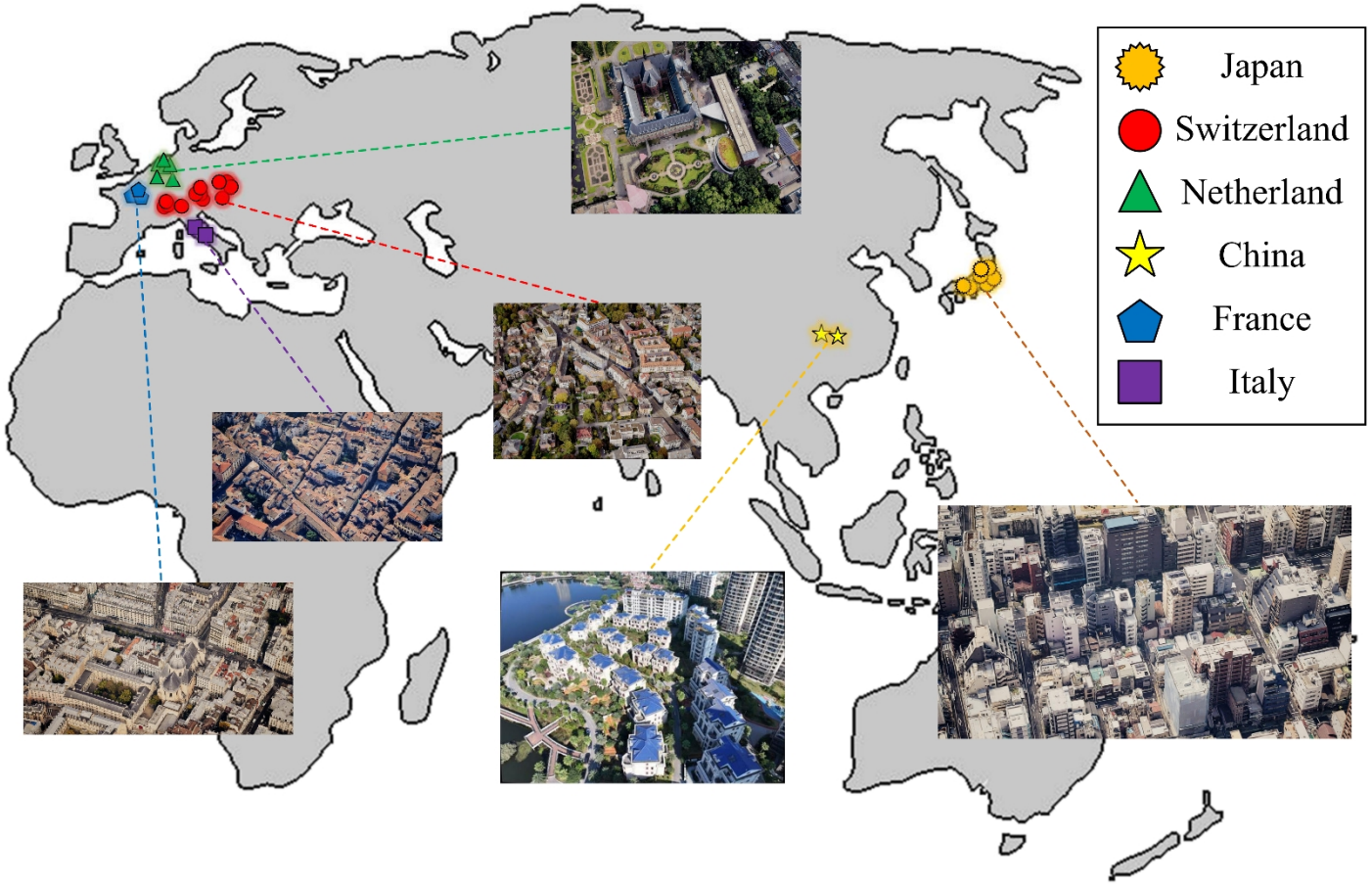
Camera	Camera Configuration			Sampling Methodology			
	Resolution[px]	Sensor Size	FOV	Strategy	Path	View Angle	Height[m]
Camera 1	1600×1200	16mm×12mm	45°	Sequence	Irregular	[0°, 70°]	[200, 500]
Camera 2	1920×1080	23.76mm×13.365mm	25°	Grid	Uniform	0°, 45°	[200, 500]
Camera 3	1920×1080	23.76mm×13.365mm	25°	Sequence	Regular	0°, 45°	[400, 500]

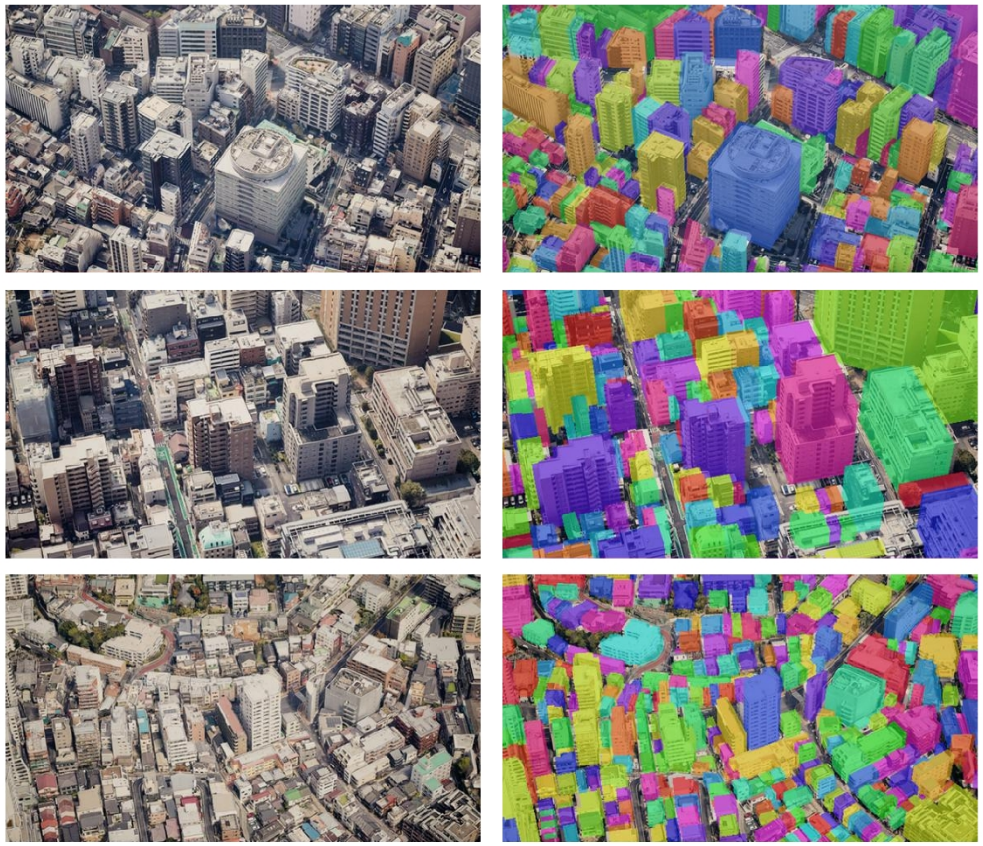
Table 1. Camera configurations and sampling methodologies used for data acquisition.

InsLoD-Loc Dataset Overview

Geographic

40 distinct UAV flight areas across 6 countries

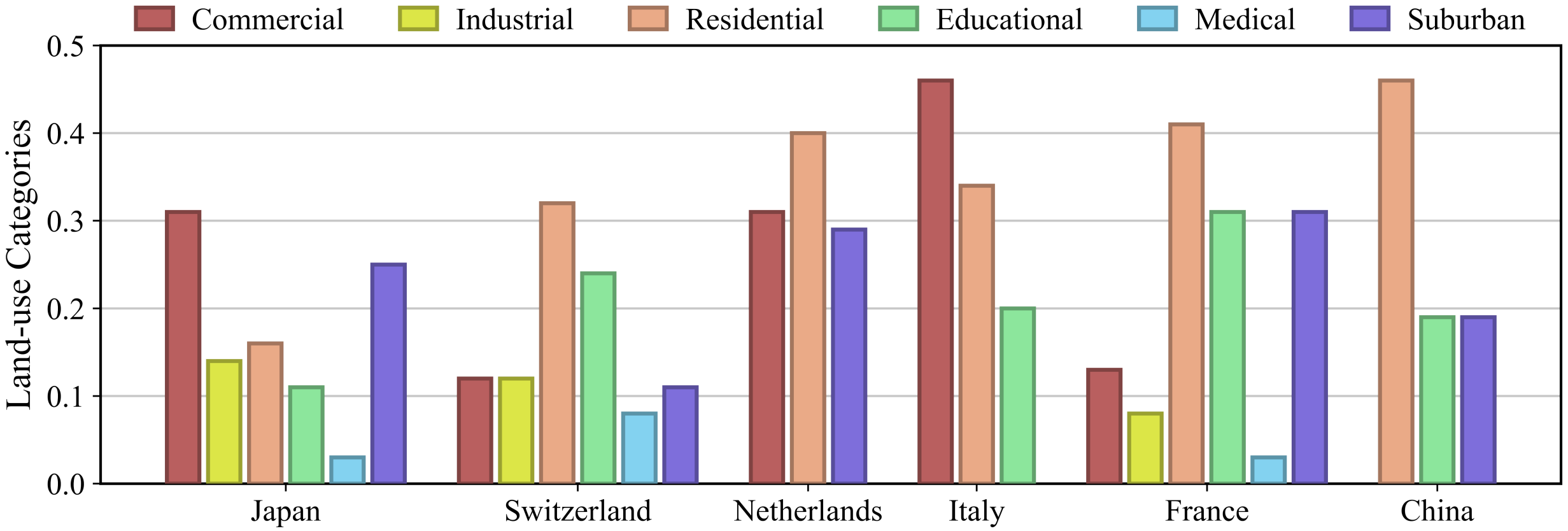




InsLoD-Loc Dataset Overview

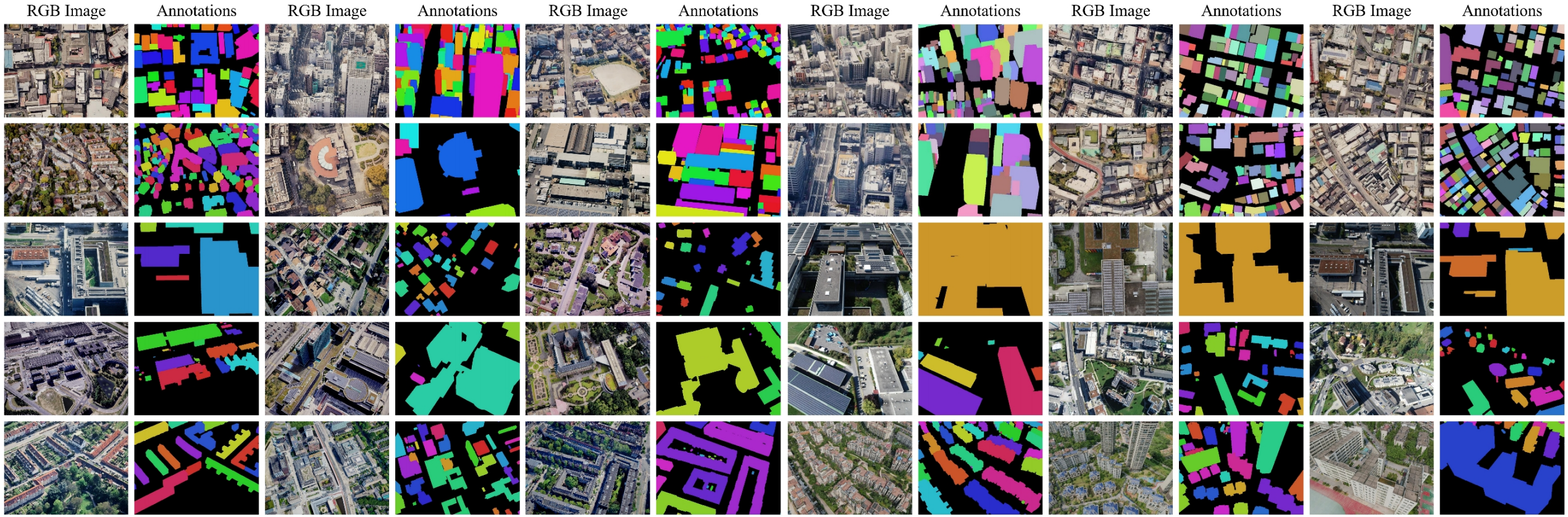
Environmental Diversity

Diverse land-use categories: Commercial, industrial, residential, educational, medical, and suburban zones



InsLoD-Loc Dataset Overview

- **A Large-Scale Dataset for Generalization**
  108,109 images & 2.67 Million annotations
  Benchmarks: UAVD4L-LoDv2, Swiss-EPFLv2, and Tokyo-LoDv3



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	4.30	6.48/1.63	0	0	0.36	11.10/0.92
CAD-Loc [65]	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 [85]	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† [98]	-	49.56	71.82	89.09	3.32/1.48	54.20	75.05	89.51	3.33/1.18
LoD-Loc v2† [99]	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
LoD-Loc v3	no refine	0	0	24.10	6.09/0.67	11.70	30.80	52.00	4.58/0.91
	no select	97.40	99.10	99.80	0.50/0.12	96.10	97.80	98.30	0.61/0.12
	Full	**97.60**	**98.90**	**99.70**	**0.49/0.13**	97.40	99.00	99.40	**0.60/0.12**

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 (°). † indicates models trained in-distribution on this dataset. Our method utilizes area-based weighting.

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
CAD-Loc	*same**	**0**	**0**	**0**	**-**	**0**	**0**	**0**	**-**
MC-Loc	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†	-	36.79	50.56	69.77	2.87/1.78	14.24	31.39	59.89	8.73/2.78
LoD-Loc v2†	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
LoD-Loc v3	no refine	0.60	2.90	16.20	8.20/3.79	0.50	1.80	11.10	9.02/3.90
	no select	58.60	77.50	90.60	1.68/0.80	34.60	57.50	80.20	2.65/0.88
	Full	**58.60**	**79.90**	**95.40**	**1.61/0.77**	**36.90**	**64.10**	**88.90**	**2.43/0.83**

Table 3. **Quantitative comparison results of different methods over Swiss-EPFLv2 dataset.** The *same** indicates that the statistics are identical to those in Tab. 2. †, T.e., and R.e. have the same meanings as those in Tab. 2. Our method utilizes area-based weighting.

Experiment

Results over the Tokyo-LoDv3 dataset

Method		*Grid-Traj.*				*Sequence-Traj.*			
		2m-2°	3m-3°	5m-5°	T.e./R.e.	2m-2°	3m-3°	5m-5°	T.e./R.e.
Prior		0.10	1.70	8.90	8.03/1.78	0.30	0.80	8.80	8.19//1.76
CAD-Loc	*same**	**0**	**0**	**0**	**-**	**0**	**0**	**0**	**-**
MC-Loc	DINOv2	0	0.26	1.61	31.95/1.88	3.42	10.50	29.81	5.16/0.49
	RoMa	0	0.14	0.39	32.36/1.89	3.58	10.72	31.12	5.16/0.48
LoD-Loc	-	10.52	20.95	33.80	8.03/1.78	0	0	5.06	8.19/1.76
LoD-Loc v2	no refine	4.00	12.50	39.70	5.91/1.15	9.00	23.10	57.40	4.56/0.91
	no select	2.50	7.50	23.50	8.04/1.56	2.90	7.30	24.60	8.53/1.43
	Full	2.70	8.10	22.70	7.86/1.48	2.30	6.80	24.00	8.75/1.52
LoD-Loc v3$_c$	no refine	7.90	21.60	51.70	4.89/0.98	18.60	38.20	75.60	3.58/0.72
	no select	35.00	62.00	83.40	2.53/0.30	41.60	69.10	89.80	2.24/0.28
	Full	**39.30**	**68.00**	**89.90**	**2.29/0.27**	**50.30**	**79.90**	**97.30**	**1.98/0.23**
LoD-Loc v3$_a$	no refine	7.90	21.60	51.70	4.89/0.98	18.60	38.20	75.60	3.58/0.72
	no select	35.50	61.70	85.10	2.49/0.29	42.30	71.70	92.70	2.21/0.27
	Full	**38.10**	**65.40**	**86.40**	**2.42/0.27**	**49.80**	**79.90**	**95.80**	**2.00/0.23**

Table 4. **Quantitative comparison results of different methods over Tokyo-LoDv3 dataset.** The *same** indicates that the statistics are identical to those in Tab. 2. T.e. and R.e. have the same meanings as in Tab. 2. LoD-Loc v3$_c$ method utilizes confidence-based weighting, and LoD-Loc v3$_a$ method utilizes area-based weighting.

Thanks for listening

Paper link: https://arxiv.org/abs/2603.19609
Project link: https://github.com/nudt-sawlab/LoD-Loc-v3