GSplatLoc: Grounding Keypoint Descriptors into 3D Gaussian Splatting for Improved Visual Localization

Gennady Sidorov [1,2] Malik Mohrat [1,2] Denis Gridusov [1]

Ruslan Rakhimov [3] Sergey Kolyubin [1]

[1] ITMO [2] SberRoboticsCenter [3] T-Tech

Project page

Modeling Stage — **Test Stage**
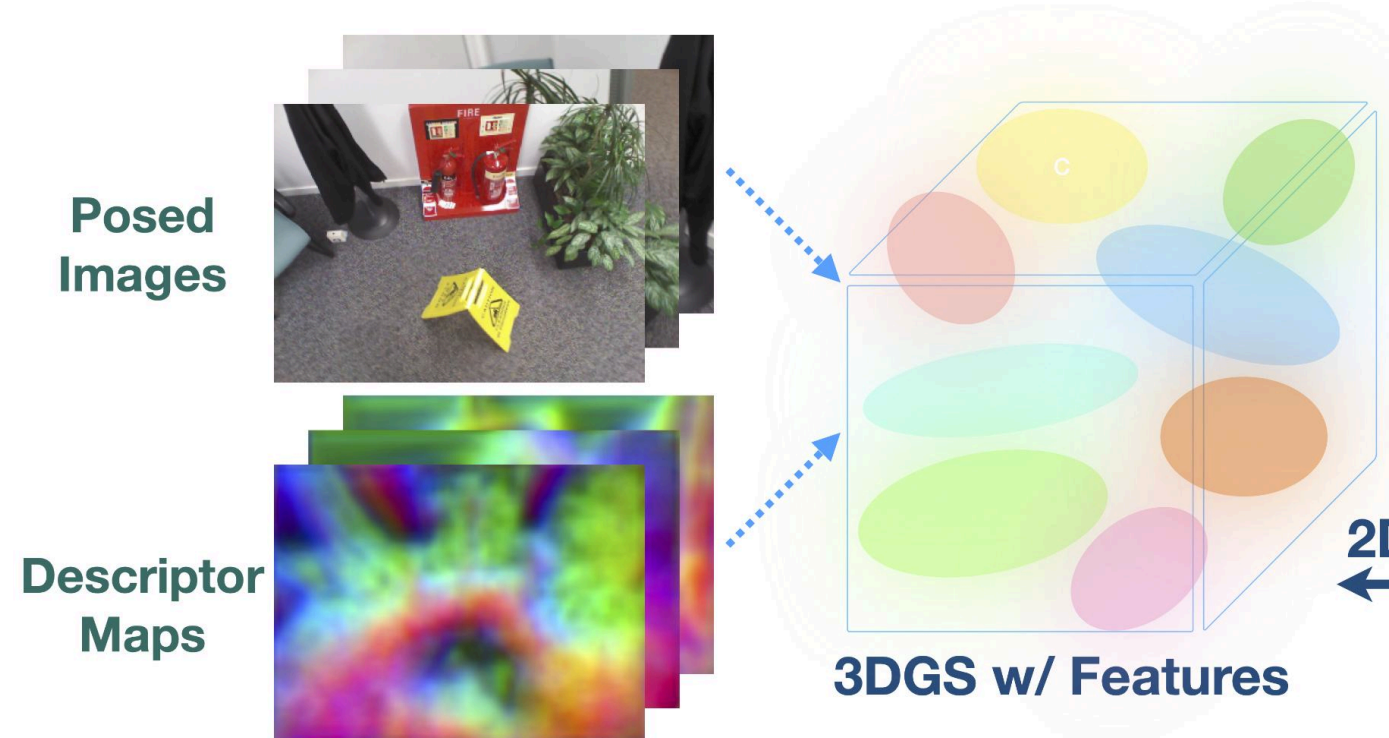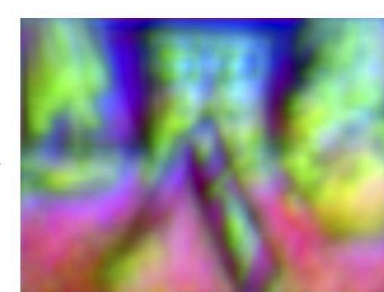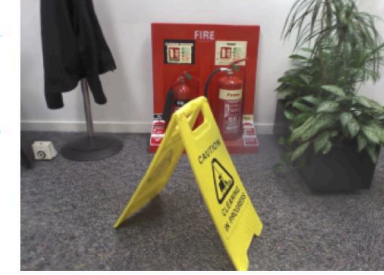
Posed Images / Descriptor Maps / 3DGS w/ Features / Coarse Pose / PnP + RANSAC / 2D-3D Correspondences / Visual Reference Image / Query Image / $\mathcal{L}_{rgb-warp}$ / Refined Pose / GT Pose / Refined Pose / Coarse Pose
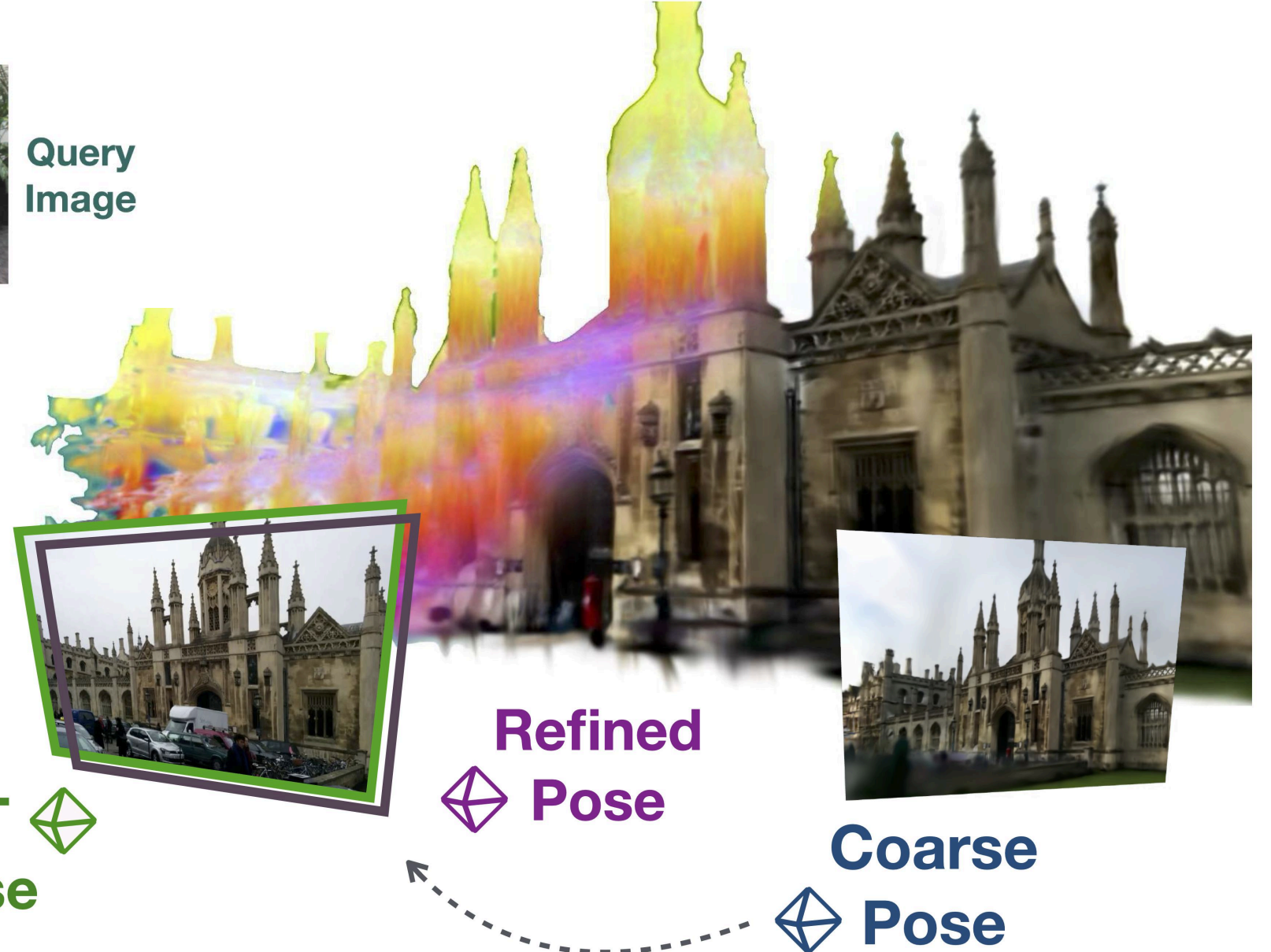
Motivation

Visual localization methods force a trade-off between efficient but specialized models (SCR) and versatile but based on slow representations (NeRF).

GSplatLoc uses 3D Gaussian Splatting to create a single, unified representation for both fast localization and other robotics tasks.

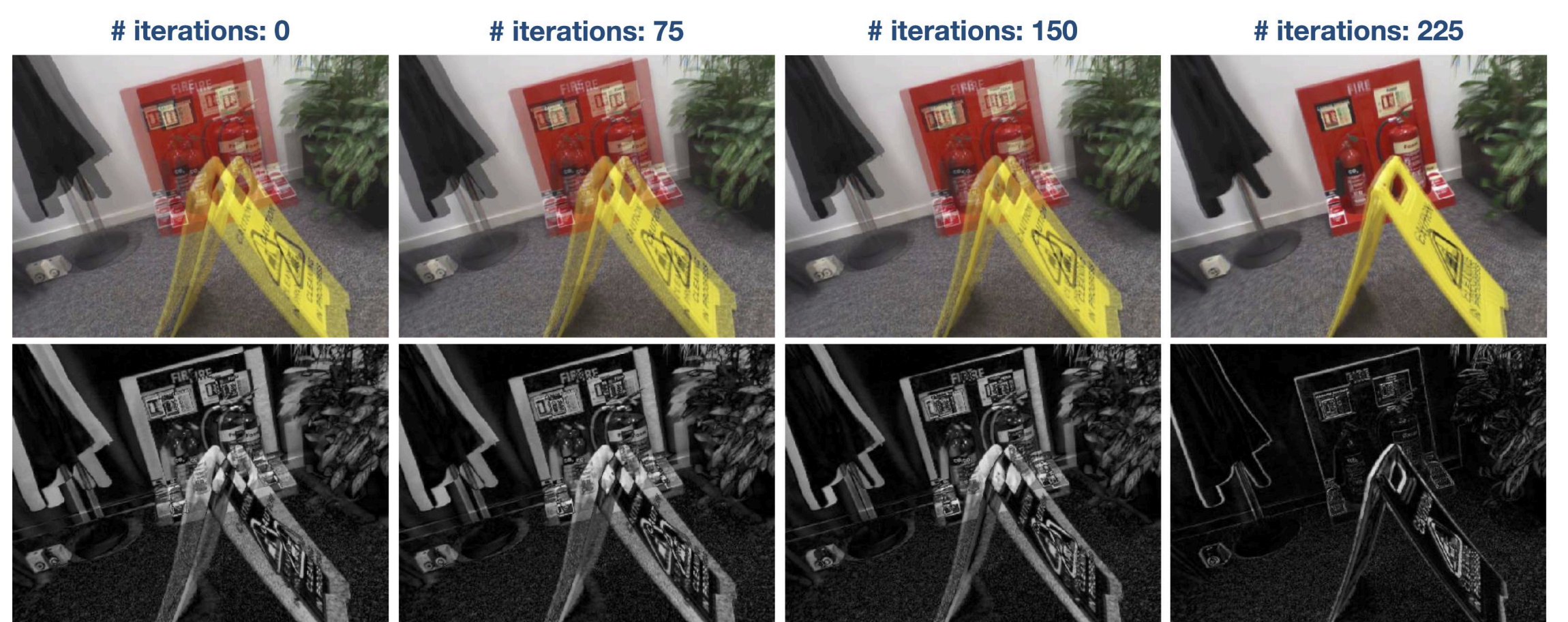## Methodology

1. Scene Modelling:
We build a 3DGS map with distilled 2D feature descriptors (XFeat) embedded into each 3D Gaussian.

2. Pose Estimation (Progressive):
→ Coarse (2D-3D matching + PnP/RANSAC)
→ Base (+ photometric refinement)
→ Fine (+ iterative feature-based refinement)

# iterations: 0 / # iterations: 75 / # iterations: 150 / # iterations: 225

Warped / GT Image

Absolute Error

Key Findings

- State-of-the-art accuracy among Neural Render Pose (NRP) methods on the indoor 7Scenes dataset.

	Methods	Chess	Fire	Heads	Office	Pumpkin	Redkitchen	Stairs	Avg. ↓ [cm/°]
APR	PoseNet [13]	10/4.02	27/10.0	18/13.0	17/5.97	19/4.67	22/5.91	35/10.5	21/7.74
	MS-Transformer [43]	11/6.38	23/11.5	13/13.0	18/8.14	17/8.42	16/8.92	29/10.3	18/9.51
	DFNet [19]	3/1.12	6/2.30	4/2.29	6/1.54	7/1.92	7/1.74	12/2.63	6/1.93
	Marepo [21]	1.9/0.83	2.3/0.92	2.1/1.24	2.9/0.93	2.5/0.88	2.9/0.98	5.9/1.48	2.9/1.04
SCR	ACE [44]	0.5/0.18	0.8/0.33	0.5/0.33	1.0/0.29	1/0.22	0.8/0.2	2.9/0.81	1.1/0.34
NRP	FQN-MN [45]	4.1/1.31	10.5/2.97	9.2/2.45	3.6/2.36	4.6/1.76	16.1/4.42	139.5/34.67	28/7.3
	CrossFire [46]	1/0.4	5/1.9	3/2.3	5/1.6	3/0.8	2/0.8	12/1.9	4.4/1.38
	PNeRFLoc [33]	2/0.8	2/0.88	1/0.83	3/1.05	6/1.51	5/1.54	32/5.73	7.28/1.76
	NeRFMatch [34]	0.9/0.3	1.1/0.4	1.5/1.0	3.0/0.8	2.2/0.6	1.0/0.3	10.1/1.7	2.8/0.7
	GSplatLoc (Coarse)	3.17/0.49	3.34/0.7	1.96/0.76	3.8/0.62	5.12/0.7	4.54/0.64	10.97/2.63	4.7/0.94
	GSplatLoc (Base)	0.43/0.16	1.03/0.32	1.06/0.62	1.85/0.4	1.80/0.35	2.71/0.55	8.83/2.34	2.53/0.68
	GSplatLoc (Fine)	0.39/0.13	0.91/0.29	0.94/0.50	1.41/0.32	1.41/0.26	1.32/0.29	3.44/0.82	1.40/0.37

- Outperforms specialized SCR methods (ACE) in dynamic outdoor scenes (Cambridge Landmarks).

	Methods	Kings	Hospital	Shop	Church	Avg. ↓ [cm/°]
APR	PoseNet [13]	93/2.73	224/7.88	147/6.62	237/5.94	175/5.79
	MS-Transformer [43]	85/1.45	175/2.43	88/3.20	166/4.12	129/2.80
	LENS [31]	33/0.5	44/0.9	27/1.6	53/1.6	39/1.15
	DFNet [19]	73/2.37	200/2.98	67/2.21	137/4.02	119/2.90
SCR	ACE [44]	29/0.38	31/0.61	5/0.3	19/0.6	21/0.47
NRP	FQN-MN [45]	28/0.4	54/0.8	13/0.6	58/2	38/1
	CrossFire [46]	47/0.7	43/0.7	20/1.2	39/1.4	37/1
	PNeRFLoc [33]	24/0.29	28/0.37	6/0.27	40/0.55	24.5/0.37
	GSplatLoc (Coarse)	41/0.50	32/0.87	11/0.40	31/0.72	29/0.62
	GSplatLoc (Base)	27/0.46	20/0.71	5/0.36	16/0.61	17/0.53
	GSplatLoc (Fine)	31/0.49	16/0.68	4/0.34	14/0.42	16/0.48
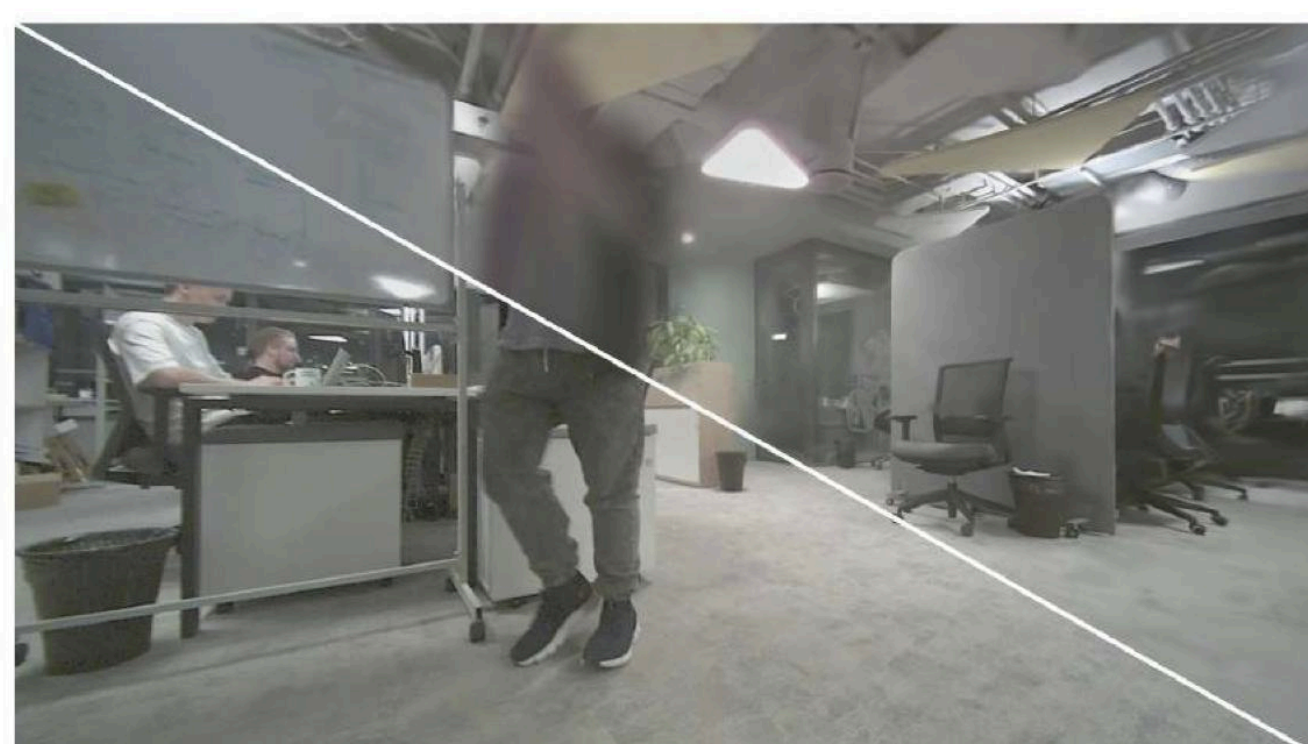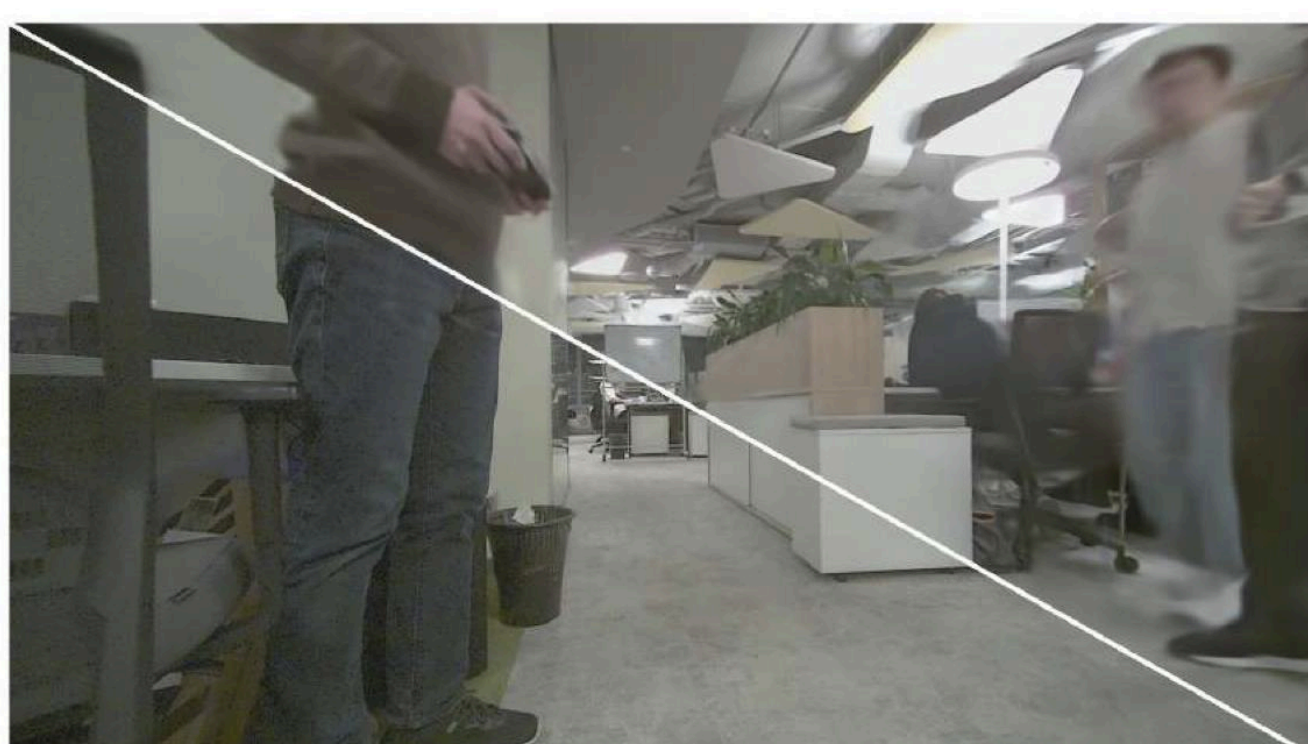
- Highly efficient: provides a refined pose in under 1 second, significantly faster than NeRF-based methods.

Method	Feature Est. (s)	Rendering (s)	Refinement (s)	Overall Query (s)
PNeRFLoc	N/A	N/A	N/A	5.560
NeRFMatch	0.157	0.141	0.846	1.144
GSplatLoc (Base)	0.018	0.140	0.651	0.809
GSplatLoc (Fine)	0.018	0.140	1.911	2.069