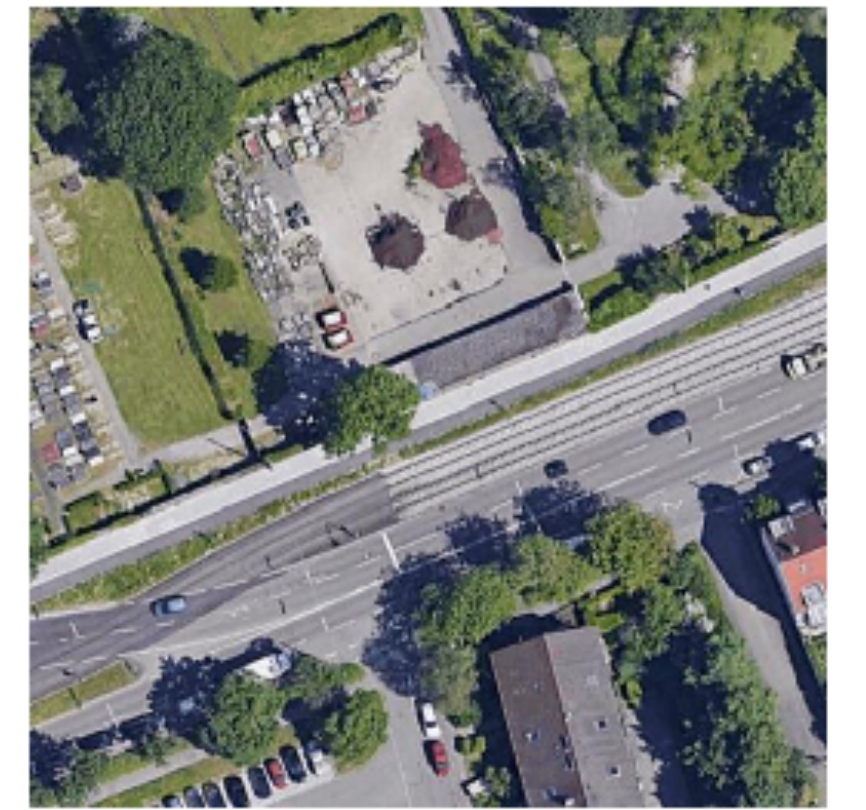
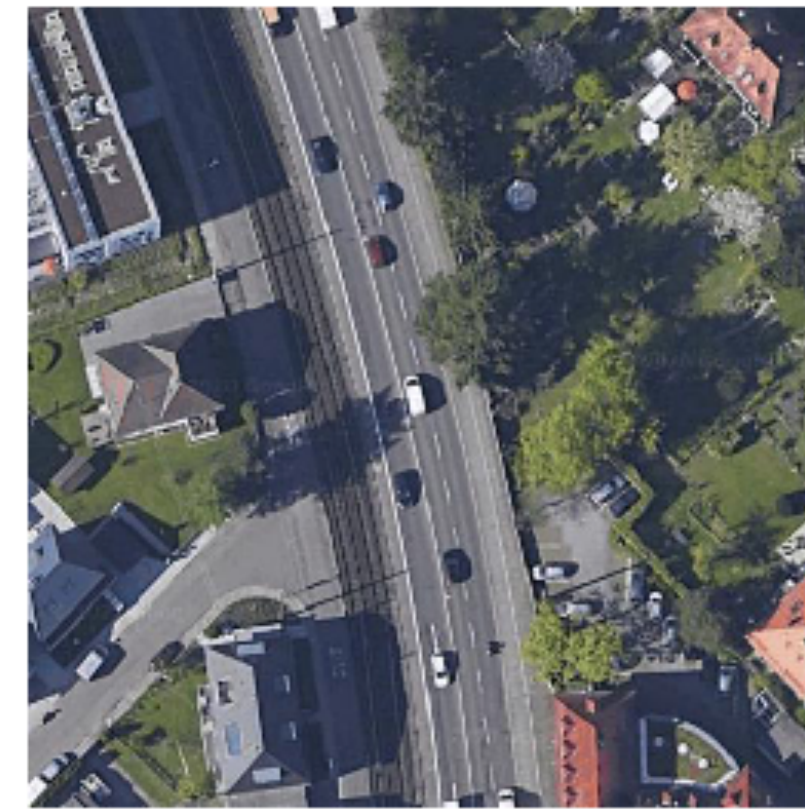
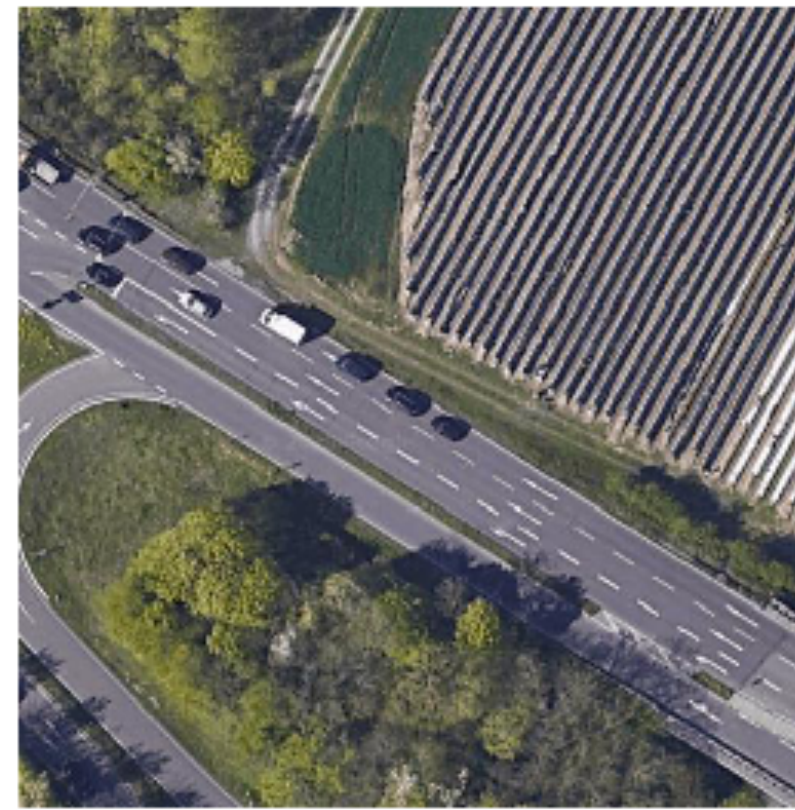
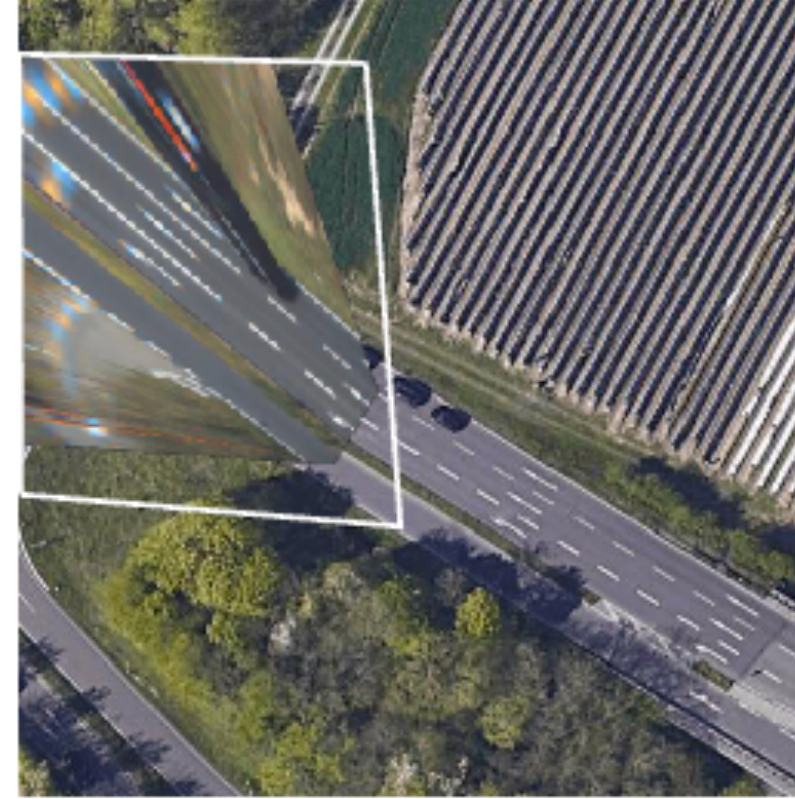
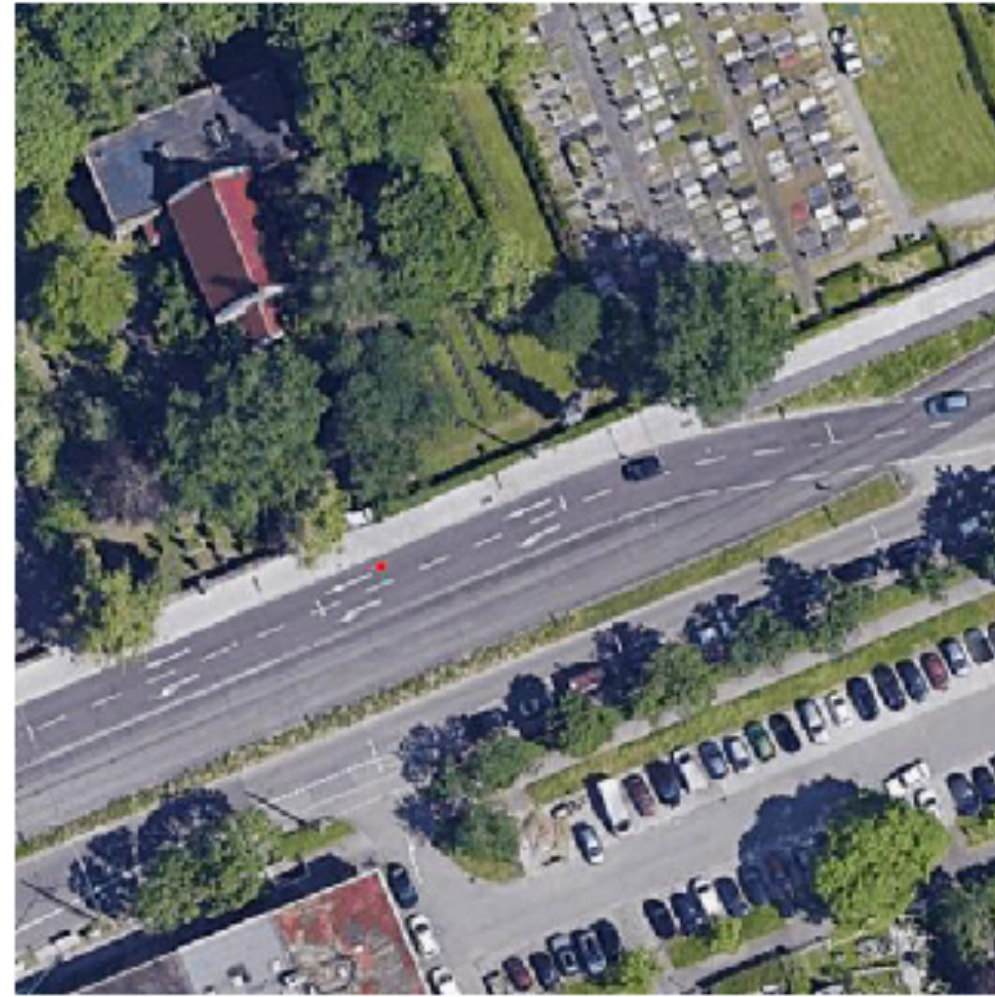
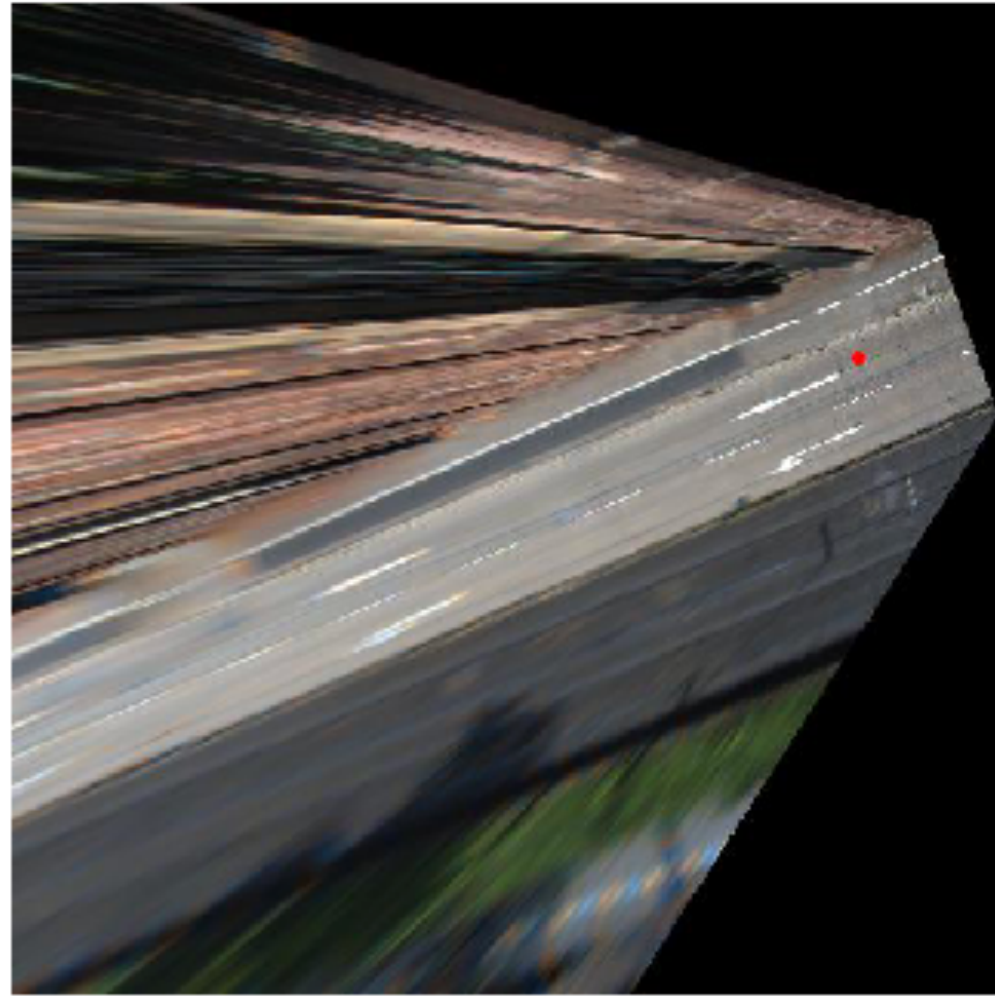

BevSplat: Resolving Height Ambiguity via Feature-Based Gaussian Primitives for Weakly-Supervised Cross-View Localization

Qiwei Wang¹, Shaoxun Wu¹, Yujiao Shi¹

¹ShanghaiTech University, Shanghai, China. Correspondence to: Yujiao Shi
<shiyj2@shanghaiTech.edu.cn>

Traditional IPM Results

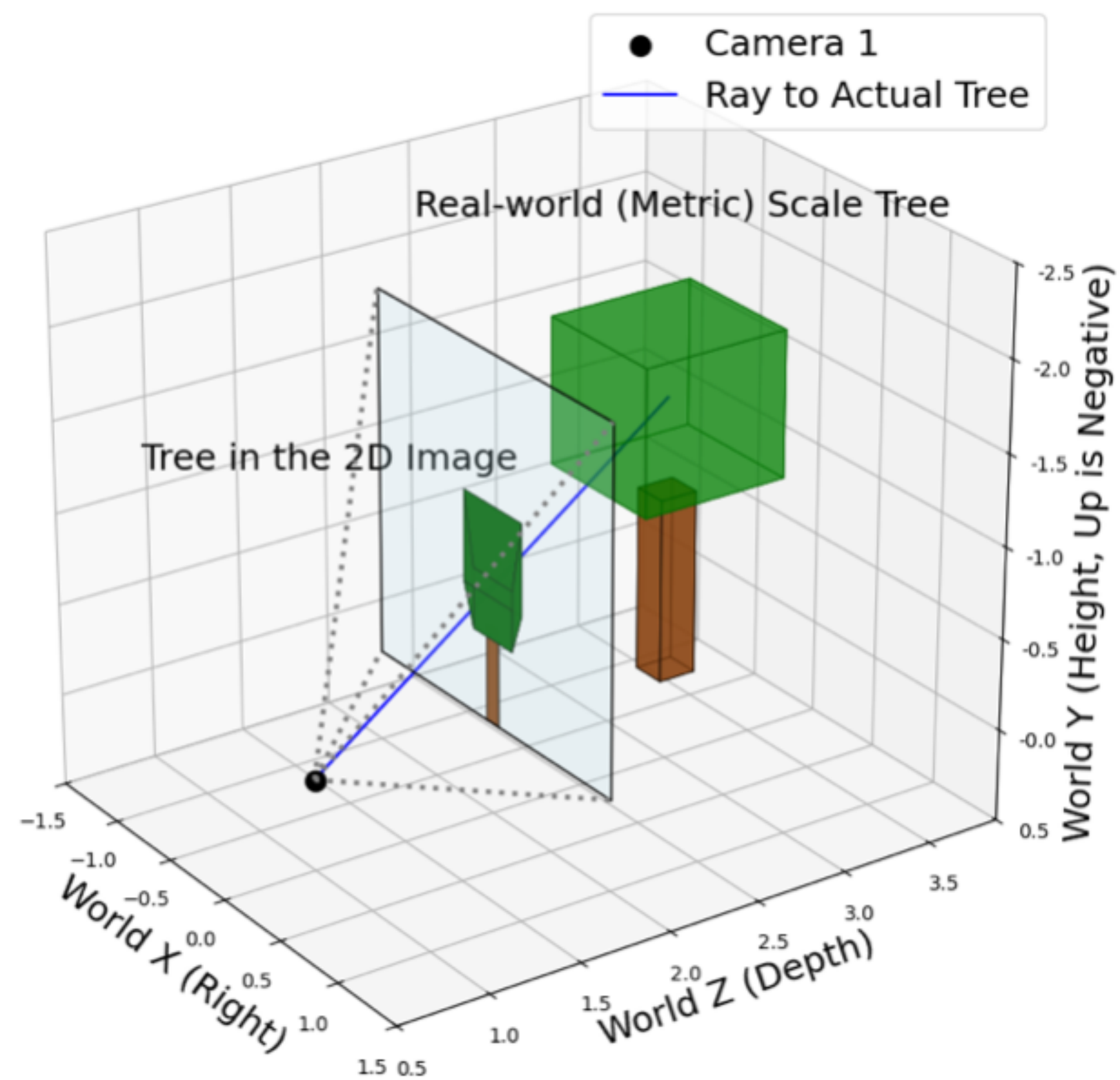


(a) Projection of KITTI

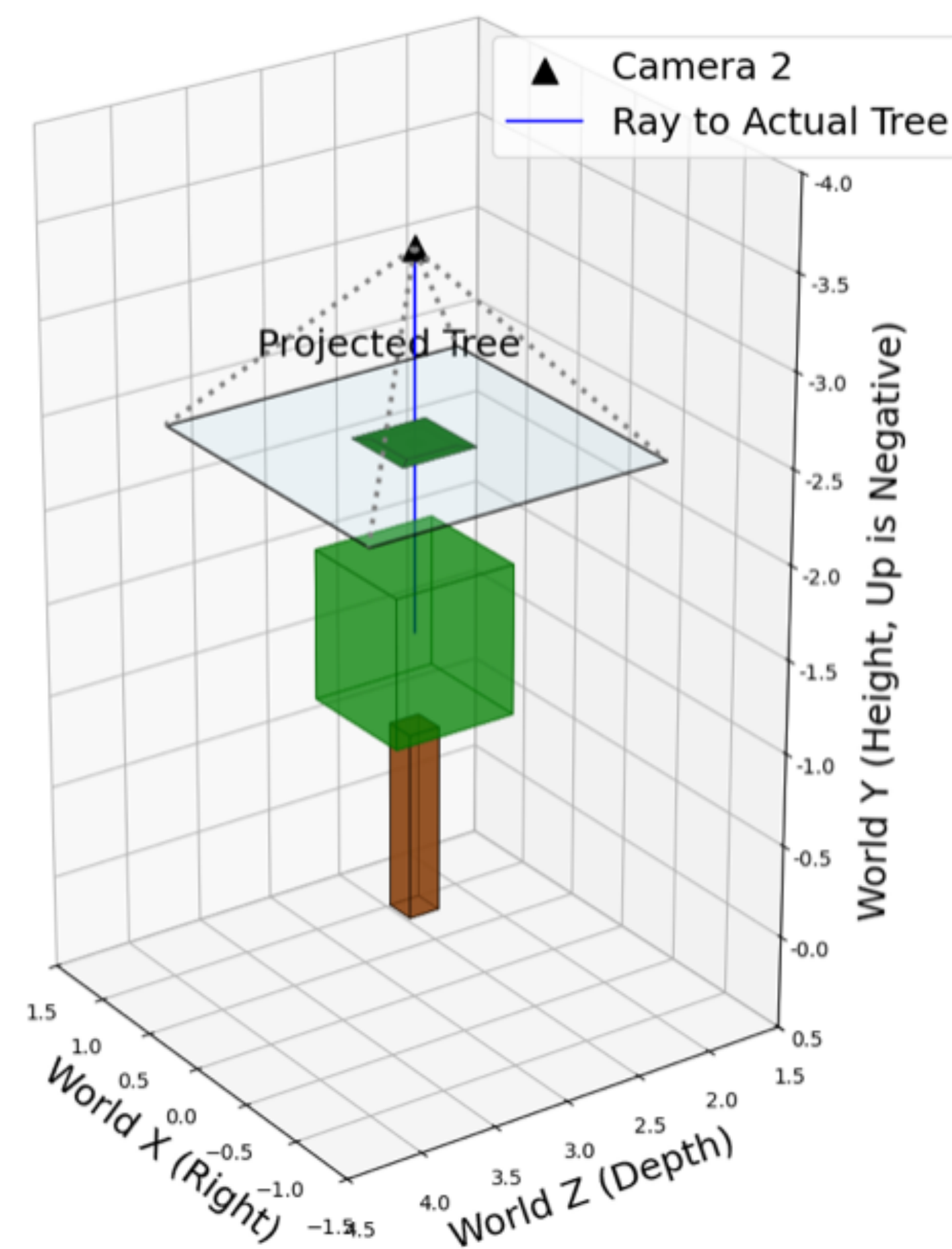
(b) Localization Results

Two Steps of Our BevSplat

Projecting the Tree from the Ground Image into 3D



Projecting the 3D Tree into the Bird's-Eye View (BEV)



BevSplat vs. IPM

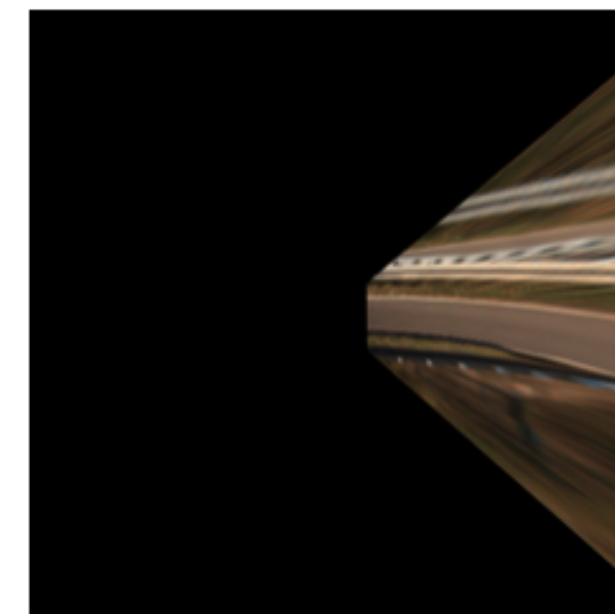


Input Ground Image

BevSplat



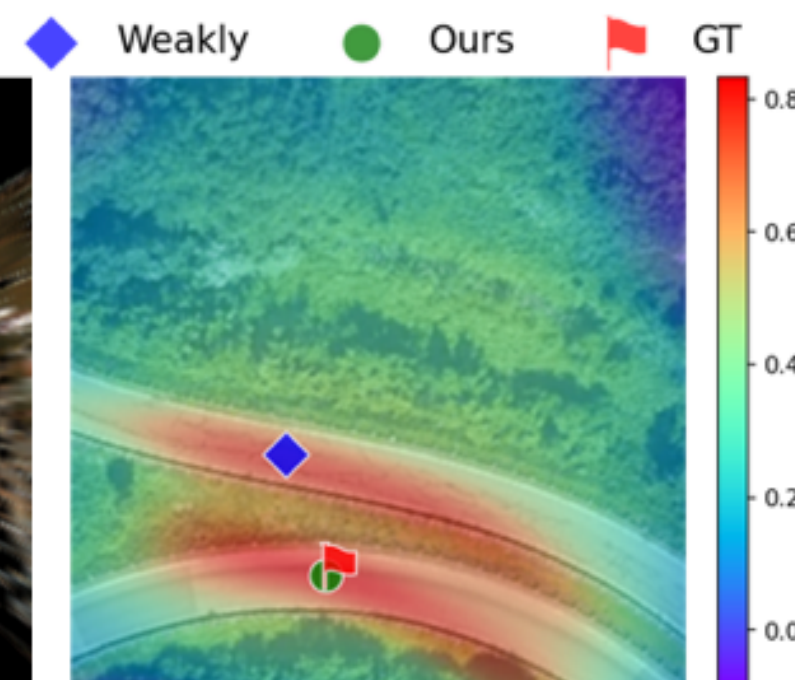
3D Reconstruction



IMP Projection



BevSplat Projection

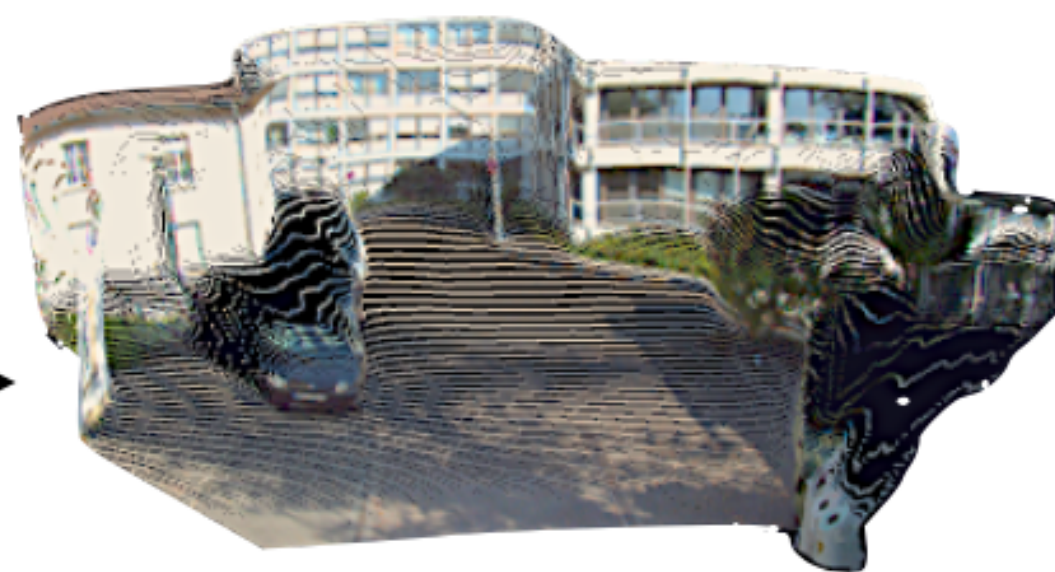


Location Map



Input Ground Image

BevSplat



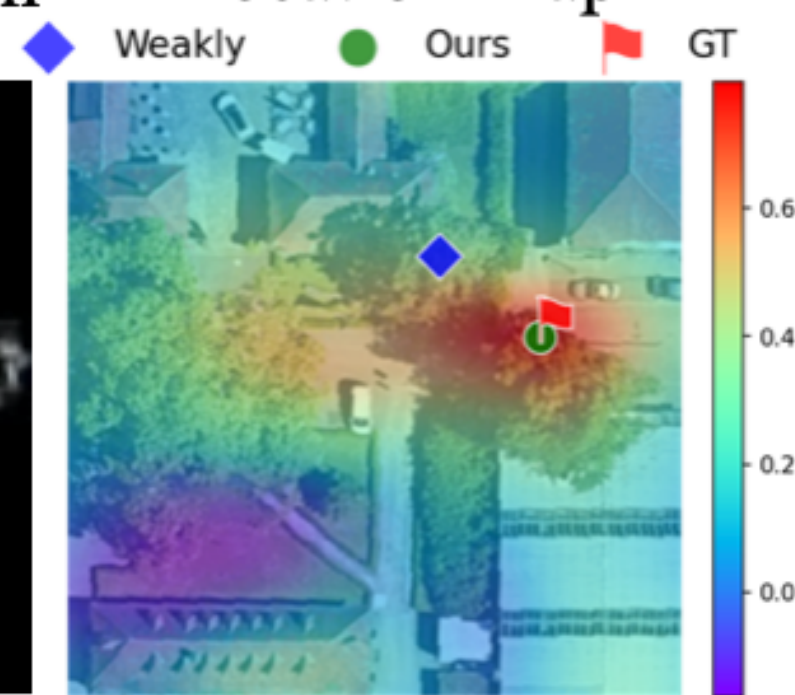
3D Reconstruction



IMP Projection

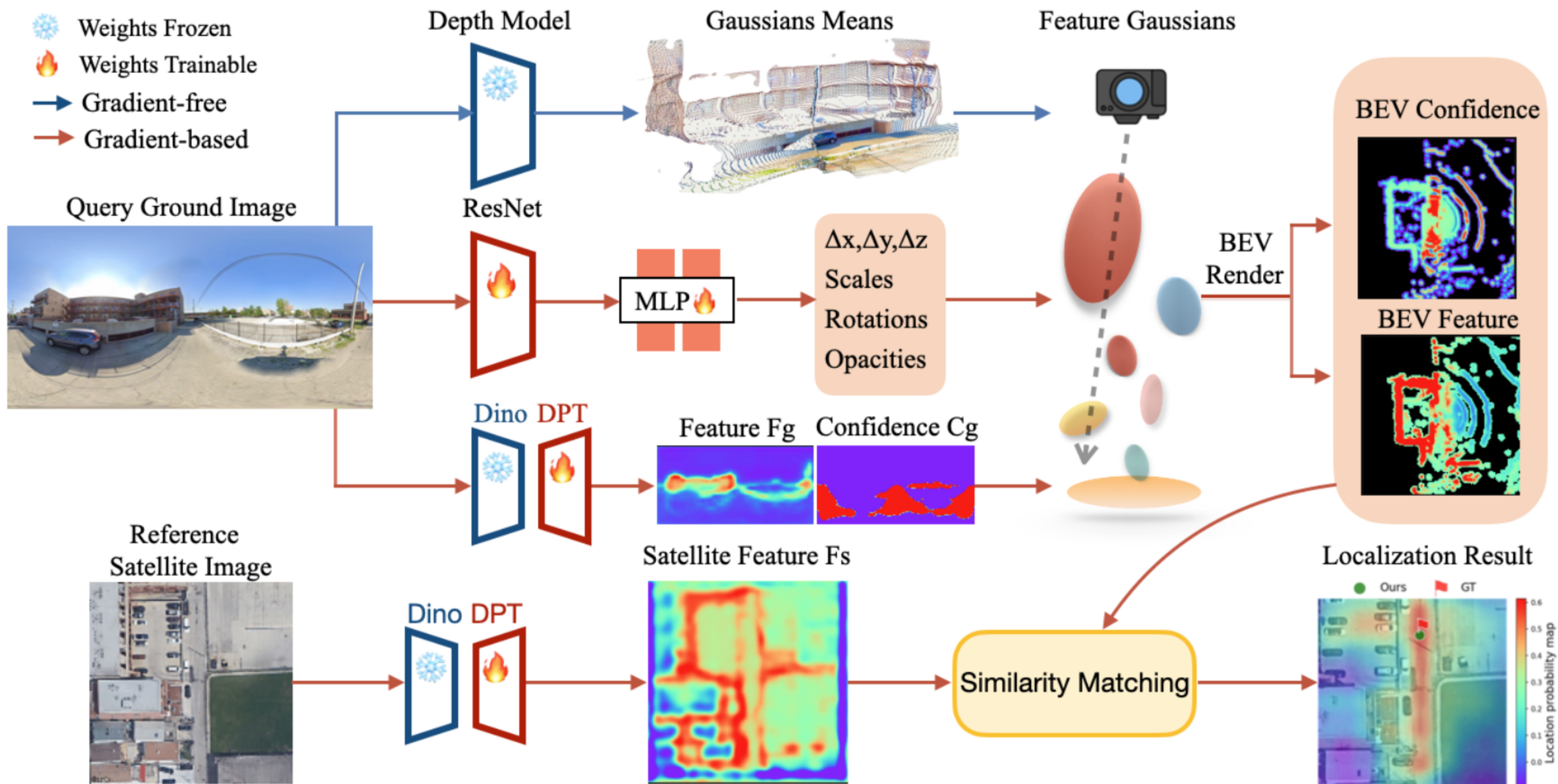


BevSplat Projection



Location Map

Framework

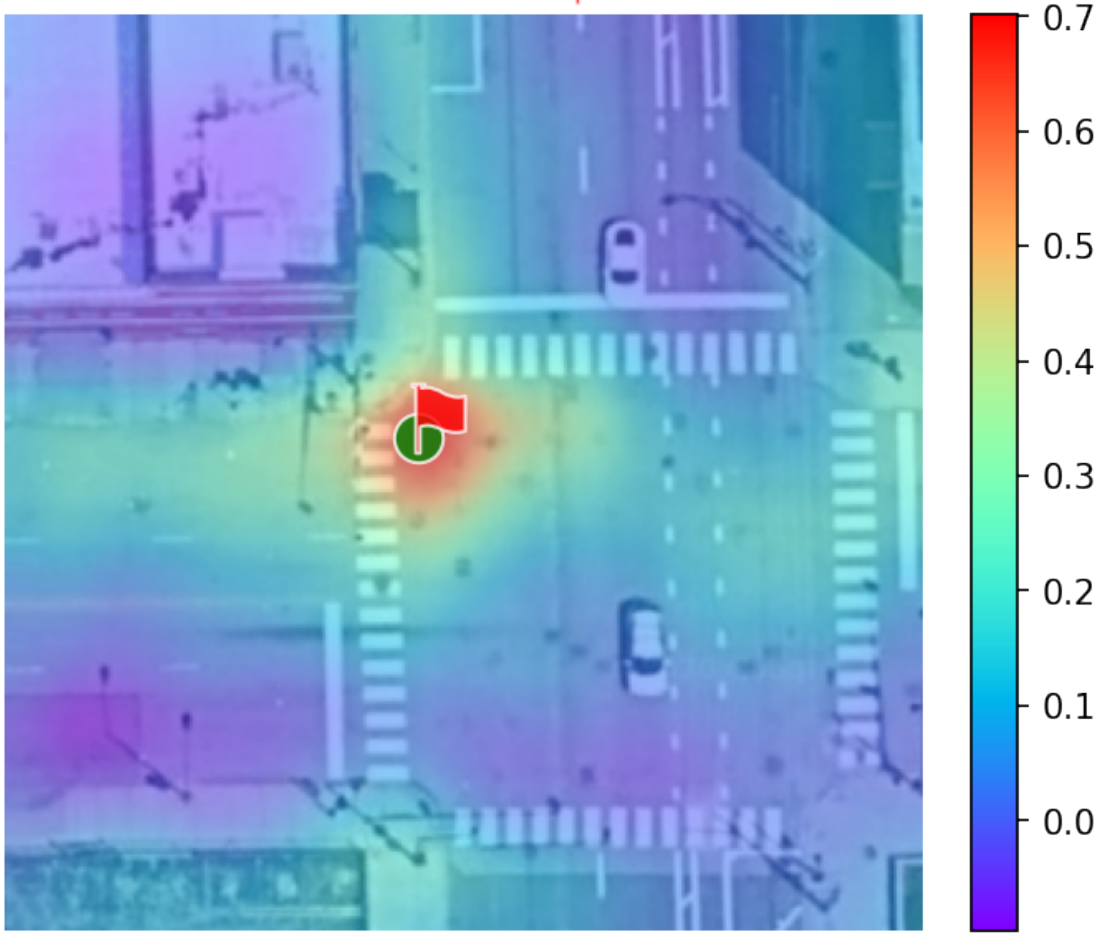
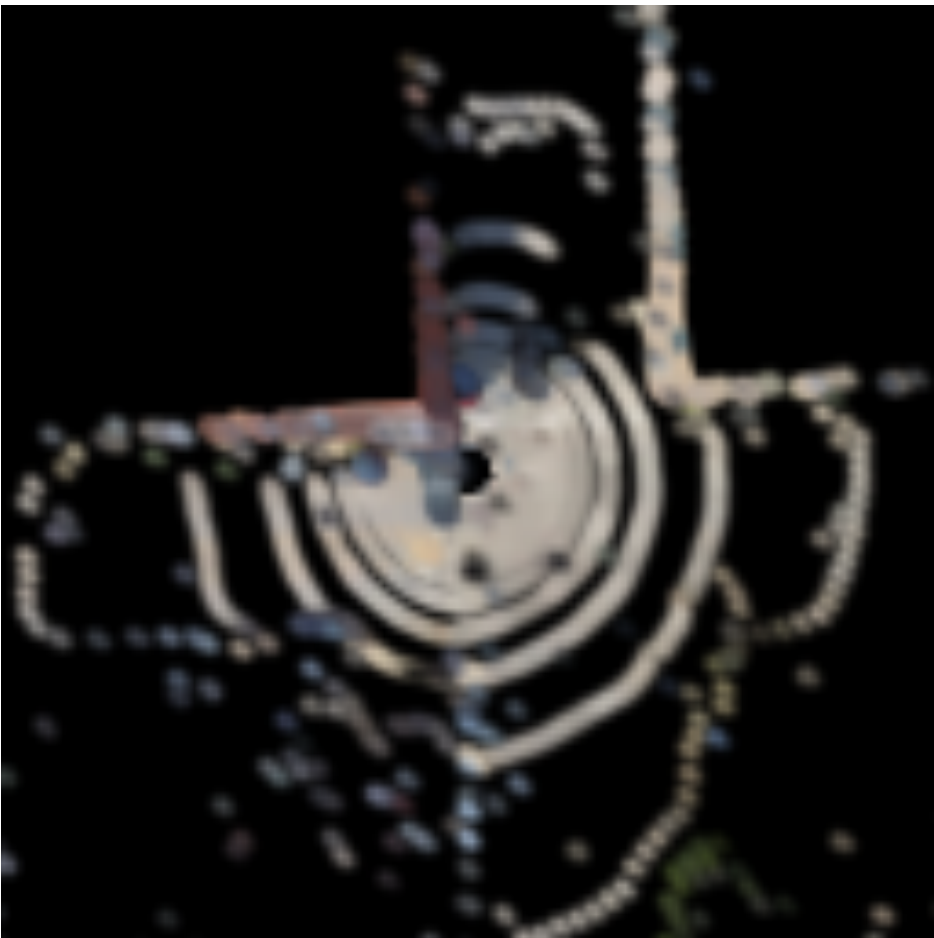
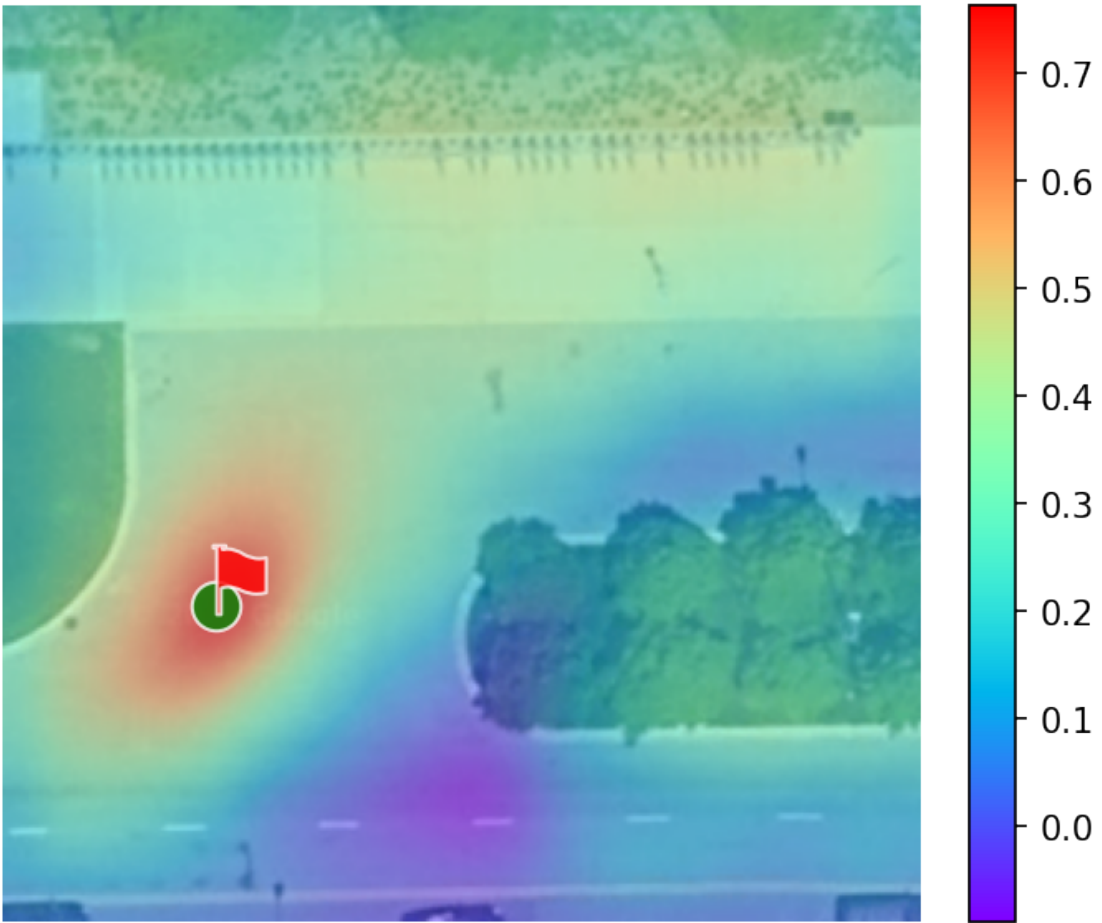
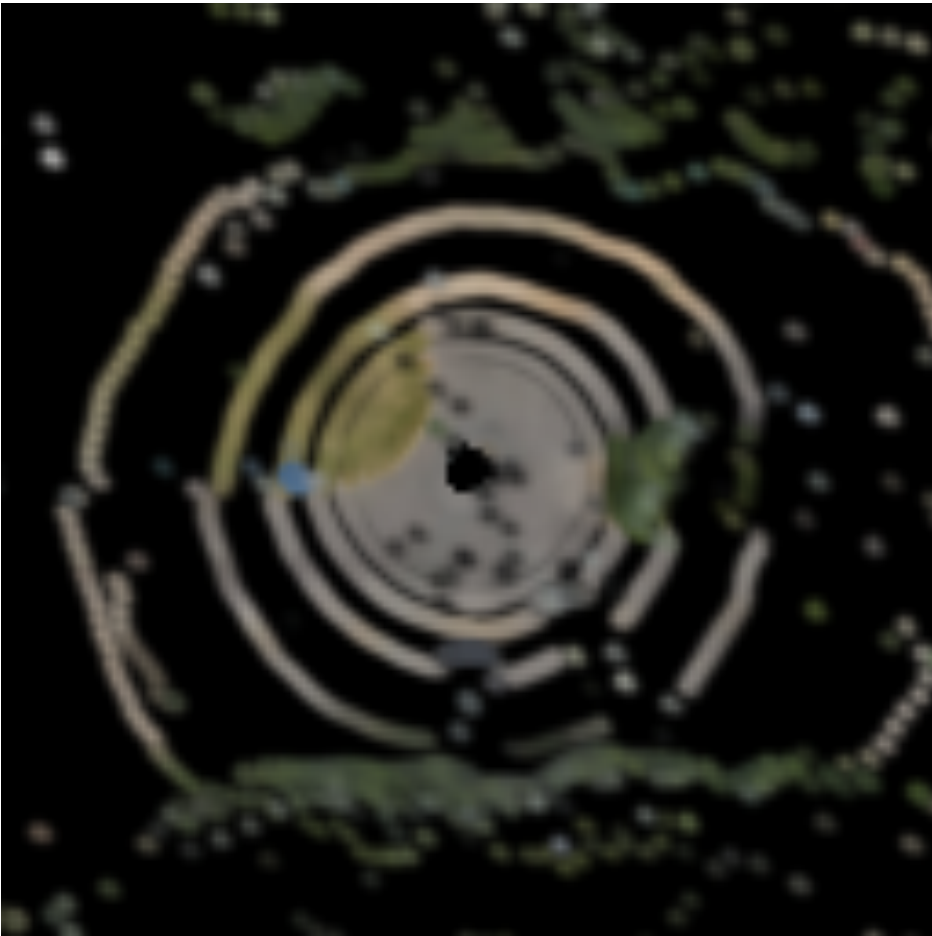


Visualization of Panorama



● Ours 🚩 GT

● Ours 🚩 GT



BevSplat Image

Location Map

BevSplat Image

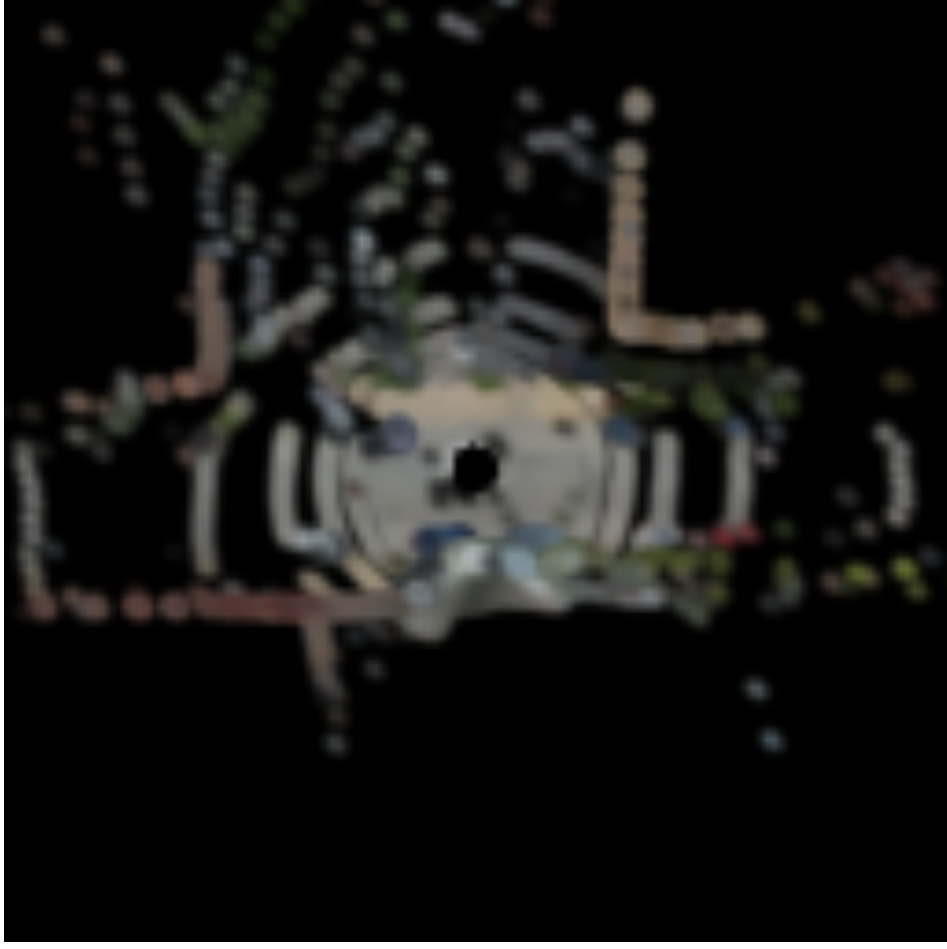
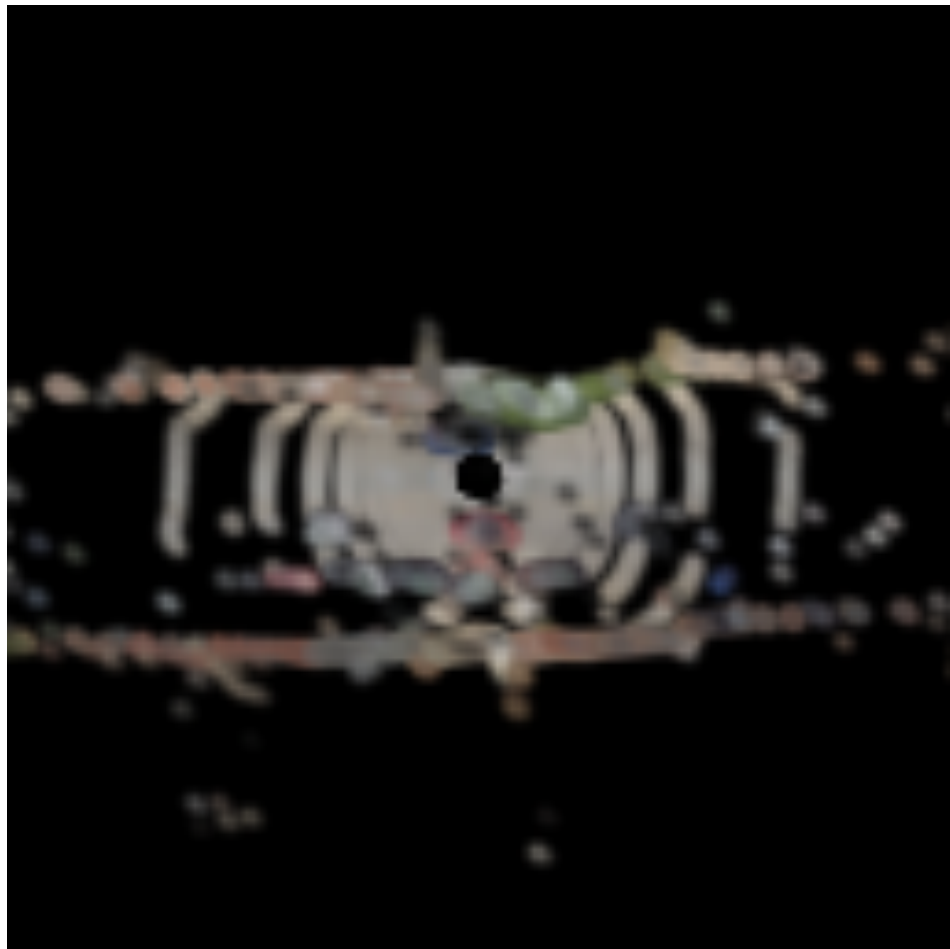
Location Map

Visualization of Panorama



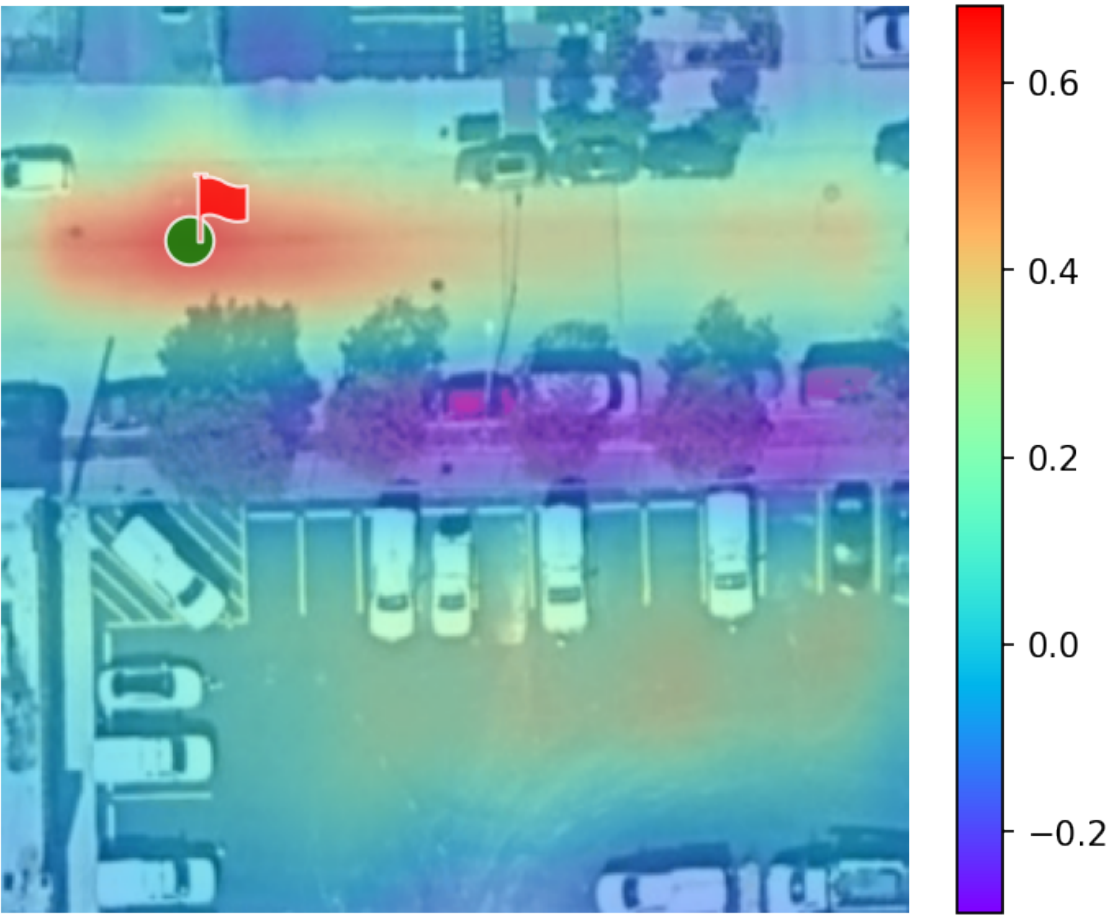
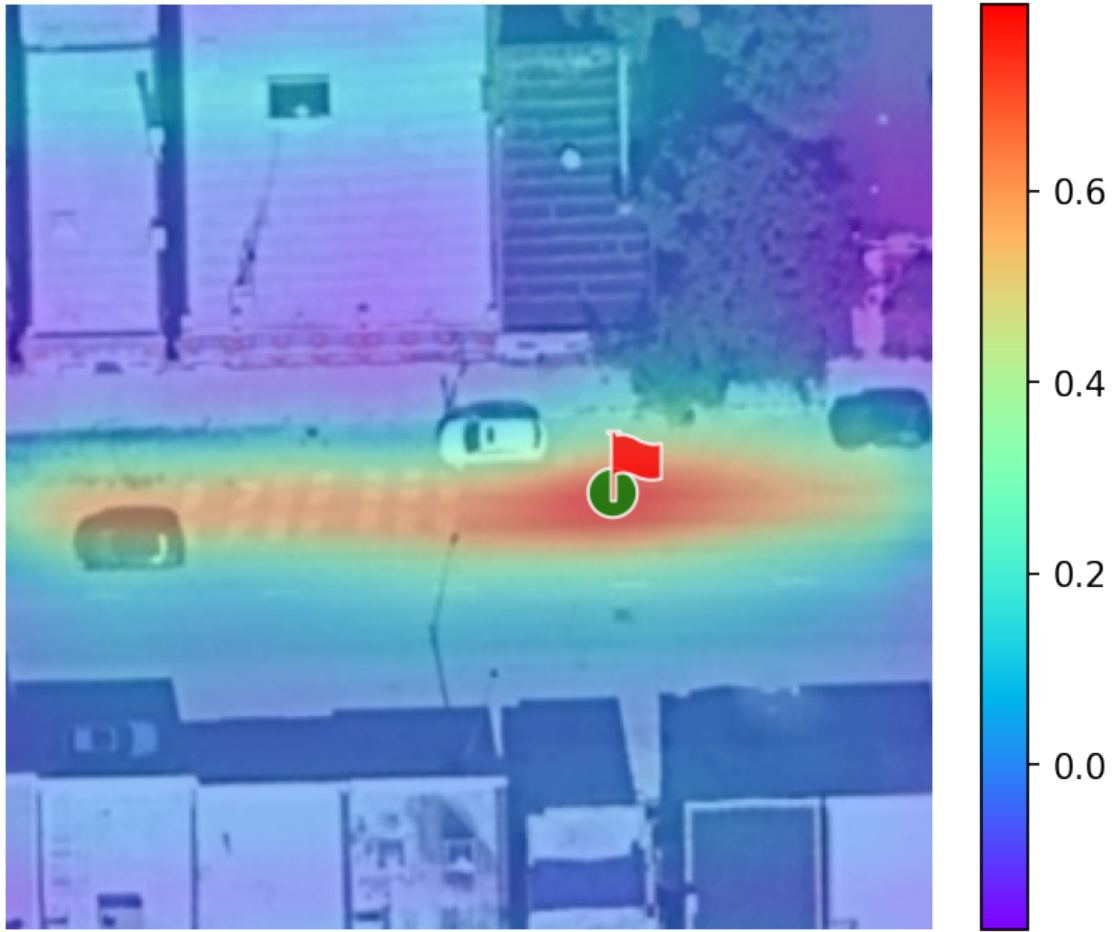
● Ours 🚩 GT

● Ours 🚩 GT



BevSplat Image

BevSplat Image



Location Map

Location Map

Visualization of KITTI

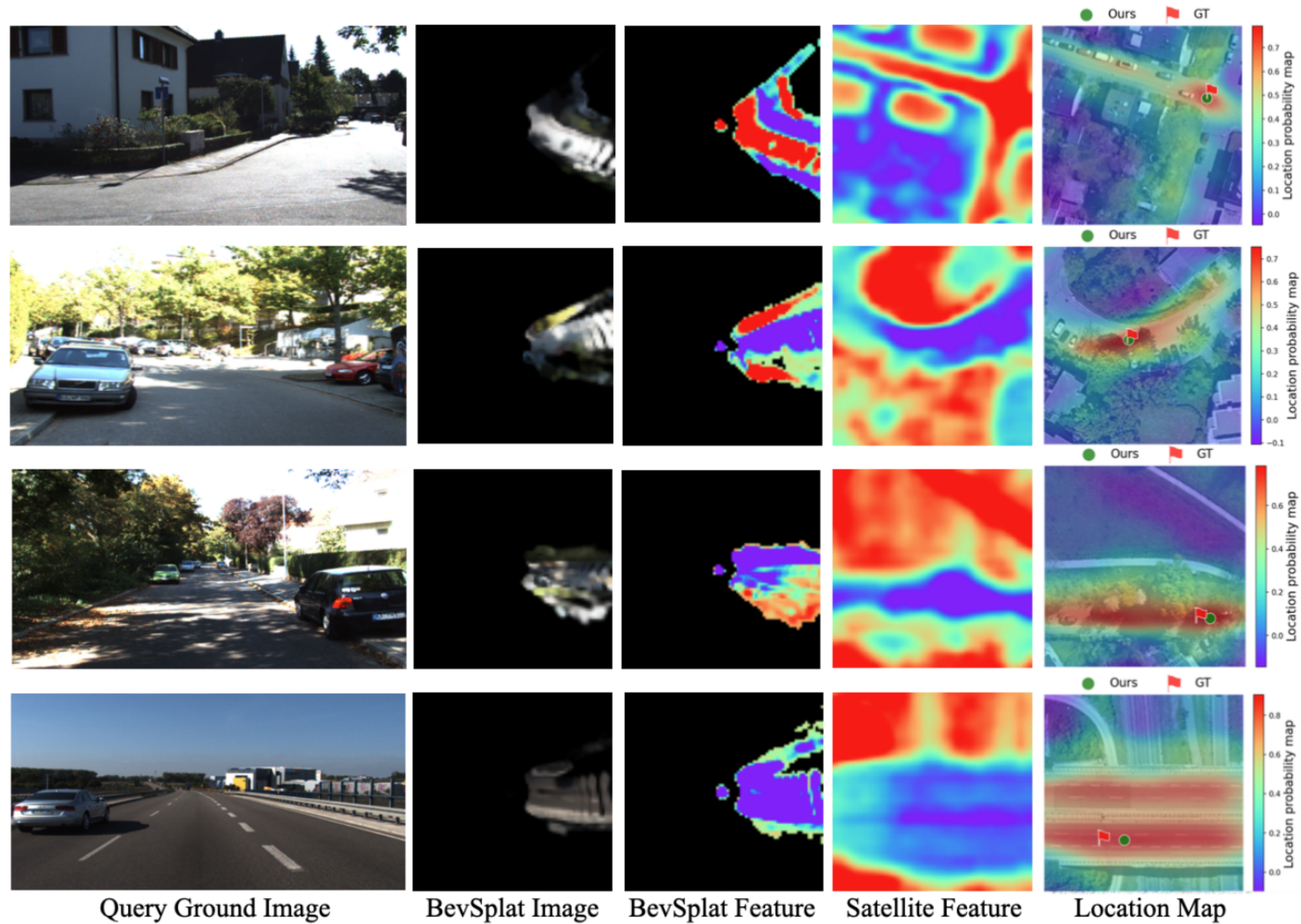


Figure 7: Qualitative results for BevSplat-based single-image localization on KITTI. Top two rows: successful examples; bottom two rows: failure examples.

IPv vs. Point Cloud vs. BevSplat

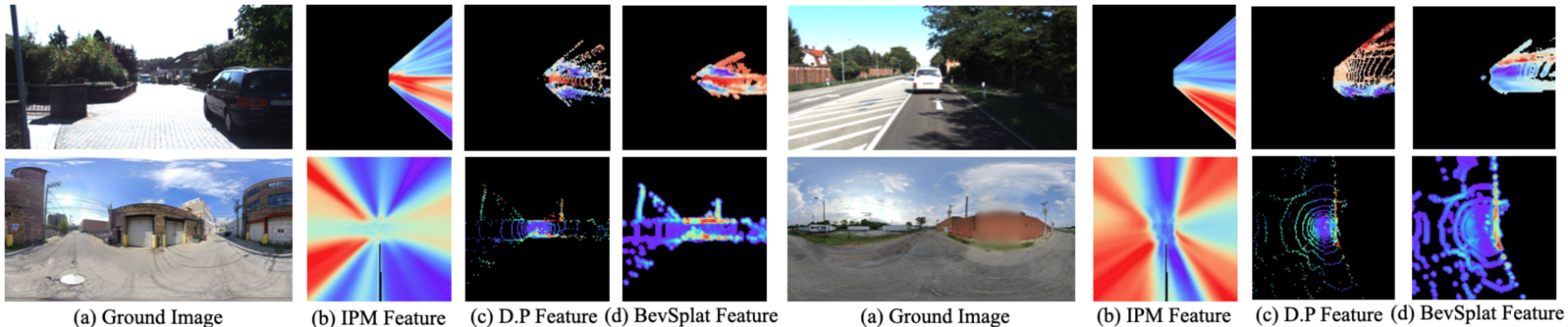


Figure 4: Visualization of query ground images (a), the corresponding BEV feature maps synthesized by IPM (b), by direct projection (c), and by the proposed BevSplat (d). The top two examples are from the KITTI dataset, while the bottom two are from the VIGOR dataset.

Table 1: Comparison with the most recent state-of-the-art (SOTA) on KITTI (* denotes supervised learning algorithms). Our weakly supervised approach slightly outperforms SOTA supervised methods in Cross-Area evaluations.

Algorithms	λ_1	Test Area	localization		Lateral		Longitudinal		Azimuth			
			mean(m)↓	median(m)↓	d=1m↑	d=3m↑	d=1m↑	d=3m↑	$\theta = 1^\circ \uparrow$	$\theta = 3^\circ \uparrow$	mean($^\circ$)↓	median($^\circ$)↓
Boosting [4]*	-	Same-Area	-	-	76.44	96.34	23.54	50.57	99.10	100.00	-	-
CCVPE [55]*	-		1.22	0.62	97.35	98.65	77.13	96.08	77.39	99.47	0.67	0.54
HC-Net [6]*	-		0.80	0.50	99.01	-	92.20	-	91.35	99.84	0.45	0.33
DenseFlow [18]*	-		1.48	0.47	95.47	-	87.89	-	89.40	-	0.49	0.30
G2SWeakly [1]	0		12.03	8.10	59.58	85.74	11.37	31.94	99.99	100.00	0.33	0.28
Ours	0		5.82	2.85	60.04	91.54	24.06	56.82	99.99	100.00	0.33	0.28
G2SWeakly [1]	1	Cross-Area	6.81	3.39	66.07	94.22	16.51	49.96	99.99	100.00	0.33	0.28
Ours	1		2.87	2.06	60.28	94.24	35.62	76.57	99.99	100.00	0.33	0.28
Boosting [4]*	-	Cross-Area	-	-	57.72	86.77	14.15	34.59	98.98	100.00	-	-
CCVPE [55]*	-		9.16	3.33	44.06	81.72	23.08	52.85	57.72	92.34	1.55	0.84
HC-Net [6]*	-		8.47	4.57	75.00	-	58.93	-	33.58	83.78	3.22	1.63
DenseFlow [18]*	-		7.97	3.52	54.19	-	23.10	-	43.44	-	2.17	1.21
G2SWeakly [1]	0		13.87	10.24	62.73	86.53	9.98	29.67	99.99	100.00	0.33	0.28
Ours	0		7.05	3.22	63.06	92.62	23.08	51.61	99.99	100.00	0.33	0.28
G2SWeakly [1]	1	Cross-Area	12.15	7.16	64.74	86.18	11.81	34.77	99.99	100.00	0.33	0.28
Ours	1		6.20	2.51	60.01	95.17	27.41	60.45	99.99	100.00	0.33	0.28

Table 2: Comparison with the most recent state-of-the-art (SOTA) on VIGOR (* denotes supervised learning algorithms). Our weakly supervised approach slightly outperforms SOTA supervised methods in Same-Area evaluations and comprehensively surpasses them in Cross-Area evaluations.

Method	λ_1	Same-Area				Cross-Area			
		Aligned-orientation		Unknown-orientation		Aligned-orientation		Unknown-orientation	
		mean(m)↓	median(m)↓	mean(m)↓	median(m)↓	mean(m)↓	median(m)↓	mean(m)↓	median(m)↓
Boosting [4]*	-	4.12	1.34	-	-	5.16	1.40	-	-
CCVPE [55]*	-	3.60	1.36	3.74	1.42	4.97	1.68	5.41	1.89
HC-Net [6]*	-	2.65	1.17	-	-	3.35	1.59	-	-
DenseFlow [18]*	-	3.03	0.97	4.97	1.90	5.01	2.42	7.67	3.67
G2SWeakly [1]	0	5.22	1.97	5.33	2.09	5.37	1.93	5.37	1.93
Ours	0	2.96	1.62	2.98	1.74	2.90	1.65	2.93	1.73
G2SWeakly [1]	1	4.19	1.68	4.18	1.66	4.70	1.68	4.52	1.65
Ours	1	2.57	1.40	2.61	1.41	2.63	1.38	2.67	1.46

Results of Mutil-Frame Localization

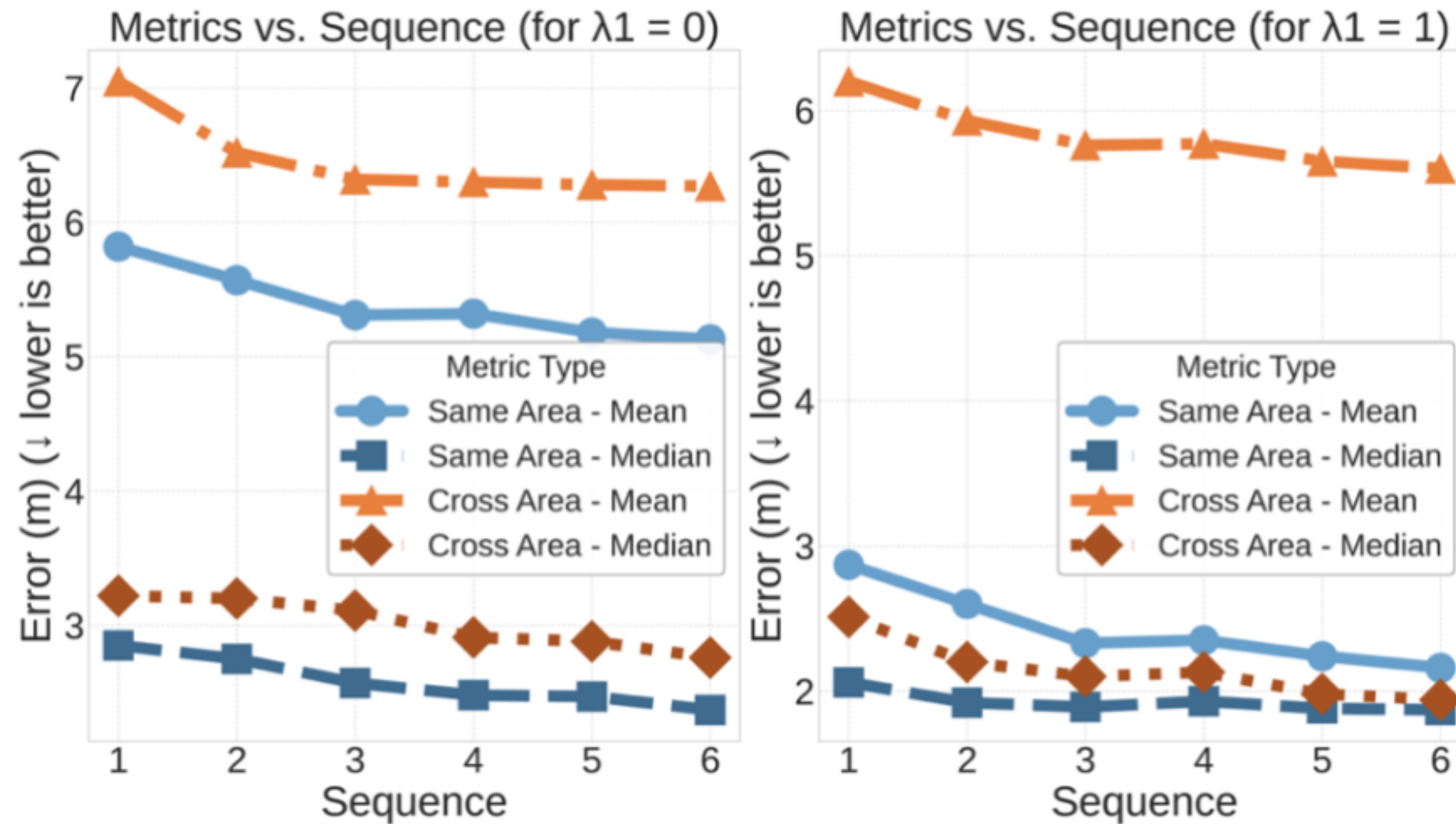
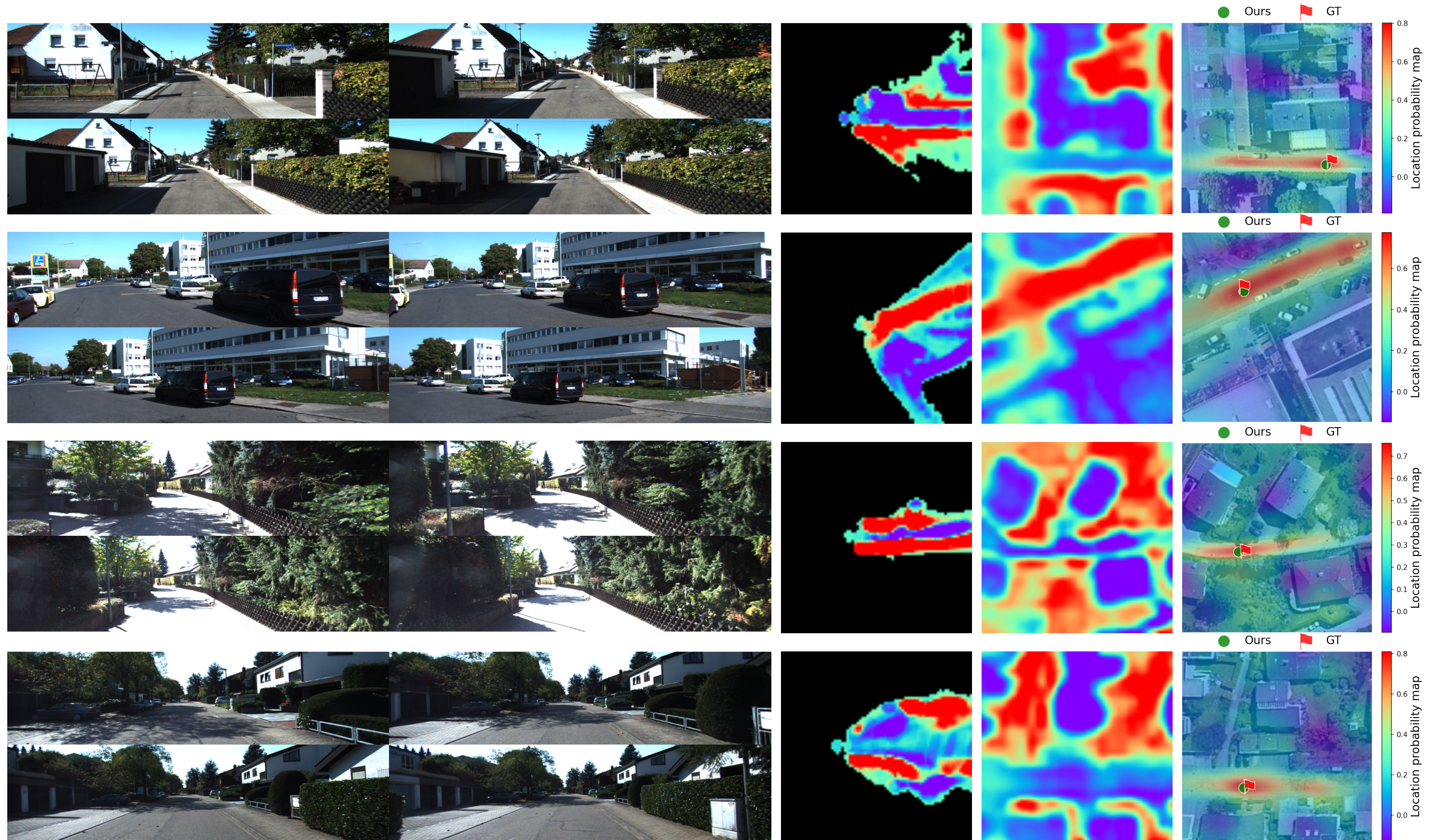


Figure 6: Location error with increasing sequence length on KITTI dataset.

Visualization of Mutil-Frame Localization



Input Sequence of Ground Images (Frames 1 to 4)

Fuse Feature

Satellite Feature

Location Map