SGLoc: Scene Geometry Encoding for Outdoor LiDAR Localization

Wen Li[1,*] Shangshu Yu[1,*] Cheng Wang[1,†] Guosheng Hu[2] Siqi Shen[1] Chenglu Wen[1]
[1] Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University, China
[2] Oosto, Belfast, UK

In this supplementary material, we first describe more details on datasets (Sec. 1). Then, we describe the network architecture (Sec. 2, and training procedure (Sec.3). We further provide additional results (Sec.4). Finally, we show more visualizations on the Oxford, quality-enhanced Oxford, and NCLT datasets (Sec. 5).

1. Dataset Details

The Oxford Radar RobotCar [1] dataset was gathered in January 2019 on a central Oxford route. This dataset includes substantial observation in various weather conditions, *e.g.*, sunny, overcast, and different lighting conditions, *e.g.*, dim and glare roadwork, making localization difficult. In our experiments, we use the data of 11-14-02-26, 14-12-05-52, 14-14-48-55, and 18-15-20-12 as the training set. The data of 15-13-06-37, 17-13-26-39, 17-14-03-00, and 18-14-14-42 are applied as the test data. More details about trajectories can be found in Tab. 1.

The NCLT [4] dataset is collected approximately bi-weekly, between January 8, 2012 and April 5, 2013, on the University of Michigan's North Campus. This dataset contains various environmental changes, *e.g.*, season, lighting, and building structure changes. Moreover, the NCLT dataset covers indoor and outdoor scenes, making it quite challenging. In our experiments, the data of 2012-01-22, 2012-02-02, 2012-02-18, and 2012-05-11 are treated as the training set, and the data of 2012-02-12, 2012-02-19, 2012-03-31, and 2012-05-26 are used as the test set. More details can be found in Tab. 2.

2. Network Architecture

The details parameter setting of SGLoc is shown in Tab. 3, except for the tri-scale spatial feature aggregation module. Specifically, we use block1 to block8 to implement the feature extractor. We use three convolution blocks for the regressor with kernel sizes of $1\times1\times1$. Note that SGLoc with three convolution layers with stride 2, thus the output resolution is reduced by a factor of 8.

Sequence	Length	Tag	Training	Test
11-14-02-26	9.37km	sunny	✓	
14-12-05-52	9.22km	overcast	✓	
14-14-48-55	9.04km	overcast	✓	
18-15-20-12	9.04km	overcast	✓	
15-13-06-37	8.85km	overcast		✓
17-13-26-39	9.02km	sunny		✓
17-14-03-00	9.02km	sunny		✓
18-14-14-42	9.04km	overcast		✓

Table 1. Dataset Descriptions on the Oxford dataset.

Sequence	Length	Tag	Training	Test
2012-01-22	6.1km	overcast	✓	
2012-02-02	6.2km	sunny	✓	
2012-02-18	6.2km	sunny	✓	
2012-05-11	6.0km	sunny	✓	
2012-02-12	5.8km	sunny		✓
2012-02-19	6.2km	overcast		✓
2012-03-31	6.0km	overcast		✓
2012-05-26	6.3km	sunny		✓

Table 2. Dataset descriptions on the NCLT dataset.

Layer name	Kernel size	Stride	Channel dimension
Block1	5×5×5	1	64
Block2	3×3×3	2	128
Block3	3×3×3	2	128
Block4	3×3×3	1	256
Block5	3×3×3	2	256
Block6	3×3×3	1	512
Block7	3×3×3	1	512
Block8	3×3×3	1	512
Block9	1×1×1	1	4096
Block10	1×1×1	1	4096
Block11	1×1×1	1	3

Table 3. Parameter setting of the network. Each block contains two convolution layers and a residual connection.

*Equal contribution.
†Corresponding author.

Methods	PNVLAD	DCP	PosePN	PosePN++	PoseSOE	PoseMinkLoc	PointLoc	**SGLoc**
15-13-06-37	9.34m, 2.65°	8.75m, 2.41°	8.53m, 3.02°	4.01m, 2.03°	3.29m, 1.75°	6.10m, 1.87°	9.70m, 2.37°	**1.58m**, **1.10°**
17-13-26-39	13.22m, 2.45°	9.53m, 1.96°	12.04m, 2.26°	5.90m, 1.74°	4.99m, 1.53°	7.98m, 1.72°	10.08m, 2.06°	**1.56m**, **1.16°**
17-14-03-00	9.81m, 2.15°	8.93m, 1.88°	7.69m, 2.19°	4.32m, 1.52°	4.27m, 1.69°	7.37m, 1.66°	9.51m, 1.86°	**1.10m**, **1.18°**
18-14-14-42	7.68m, 1.80°	7.64m, 1.93°	5.56m, 1.72°	4.18m, 1.71°	3.26m, 1.55°	5.85m, 2.01°	8.84m, 2.14°	**0.99m**, **1.04°**
Average	10.01m, 2.26°	8.71m, 2.05°	8.46m, 2.30°	4.60m, 1.75°	3.95m, 1.63°	6.83m, 1.82°	9.53m, 2.11°	**1.31m**, **1.12°**

Table 4. Position error (m) and orientation error (°) for various methods with PGO on the quality-enhanced Oxford dataset.

Methods	PNVLAD	DCP	PosePN	PosePN++	PoseSOE	PoseMinkLoc	PointLoc	**SGLoc**
2012-02-12	6.11m, 5.50°	7.22m, 6.84°	8.03m, 6.53°	4.34m, 3.18°	10.04m, 6.74°	5.25m, 4.22°	6.19m, 4.03°	**0.88m**, **2.35°**
2012-02-19	5.99m, 5.11°	5.73m, 4.52°	5.19m, 4.95°	3.16m, 2.09°	4.30m, 3.15°	3.96m, 3.47°	5.36m, 3.10°	**0.85m**, **2.06°**
2012-03-31	5.59m, 5.37°	6.42m, 5.22°	4.89m, 5.09°	3.79m, 2.85°	3.74m, 3.62°	3.44m, 3.70°	5.25m, 2.94°	**0.79m**, **2.34°**
2012-05-26	12.71m, 7.10°	13.06m, 6.62°	12.18m, 7.42°	8.78m, 3.86°	10.01m, 6.34°	9.26m, 6.24°	8.82m, 4.44°	**3.25m**, **3.52°**
Average	7.60m, 5.77°	8.11m, 5.80°	7.57m, 6.00°	5.02m, 3.00°	7.02m, 4.96°	5.48m, 4.41°	6.41m, 3.63°	**1.44m**, **2.57°**

Table 5. Position error (m) and orientation error (°) for various methods with PGO on the NCLT dataset.

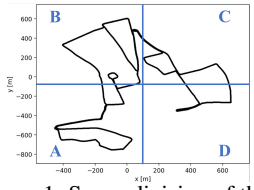

Figure 1. Scene division of the Oxford dataset.

Area	Error
A	3.25m/2.24°
AB	3.09m/2.05°
ABC	3.03m/1.97°
ABCD	3.14m/1.88°

Table 6. Position error (m) and orientation error (∘) on the divided Oxford dataset.

3. Training Procedure

Similarly to [3], we use the pre-trained SPVCNN [5] and set the voxel size to 0.15m to segment the movable objects generating the binary masks for training. Note that this mask is only used in the loss calculation stage. For the Oxford and quality-enhanced Oxford dataset, SGLoc is trained from scratch for 50 epochs with an initial learning rate of 0.001 using Adam. The batch size, decay step, and decay rate are set to 35, 1200, and 0.95, respectively. For the NCLT dataset, we train the network for 40 epochs, and the batch size, decay step, and decay rate are set to 30, 1000, and 0.95, respectively.

4. Additional Results

4.1. Localization on various-scale scenes

We have conducted the effectiveness of the proposed method on two large outdoor datasets, Oxford and NCLT. To further demonstrate the robustness of SGLoc on various-scale scenes, as shown in Fig. 1, we divide the Oxford dataset into four parts (Area A, B, C, and D) according to the center point of the trajectory. As shown in Tab. 6, the localization performance of SGLoc is robust as the scene coverage grows from 40hm^2 (A) to 200 hm^2 (ABCD), demonstrating its applicability to various-scale scenes. We believe the key reason is SGLoc can capture scene geometry by recovering the scenes in the world coordinate frame.

4.2. Localization Accuracy at Sub-meter Level

As mentioned in the main paper, we utilize pose graph optimization (PGO) [2] as post-processing to further improve localization results. Tab. 4 and Tab. 5 show the localization results of various methods with PGO on the quality-enhanced Oxford and NCLT datasets. Clearly, the existing methods cannot achieve accuracy at the sub-meter level, even on some trajectories. To our knowledge, SGLoc is the first regression-based method to reduce the error to the level of the sub-meter on some trajectories, which demonstrates its effectiveness.

5. Visualization

We show more visualization results in Fig. 2, Fig. 3, and Fig. 4 for Oxford, quality-enhanced Oxford, and NCLT datasets, respectively.

References

[1] Dan Barnes, Matthew Gadd, Paul Murcutt, Paul Newman, and Ingmar Posner. The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset. In *ICRA*, pages 6433–6438, 2020. 1

[2] Luca Carlone, Giuseppe C. Calafiore, Carlo Tommolillo, and Frank Dellaert. Planar pose graph optimization: Duality, optimal solutions, and verification. *IEEE TR*, 32:545–565, 2017. 2

[3] Zhaoyang Huang, Yan Xu, Jianping Shi, Xiaowei Zhou, Hujun Bao, and Guofeng Zhang. Prior guided dropout for robust visual localization in dynamic environments. In *ICCV*, pages 2791–2800, 2019. 2

[4] Carlevaris-Bianco Nicholas, K. Ushani Arash, and M. Eustice Ryan. University of michigan north campus long-term vision and lidar dataset. *Int. J. of Rob. Res.*, 35:545–565, 2015. 1

[5] Haotian Tang, Zhijian Liu, Shengyu Zhao, Yujun Lin, Ji Lin, Hanrui Wang, and Song Han. Searching efficient 3d architectures with sparse point-voxel convolution. In *ECCV*, pages 685–702, 2020. 2

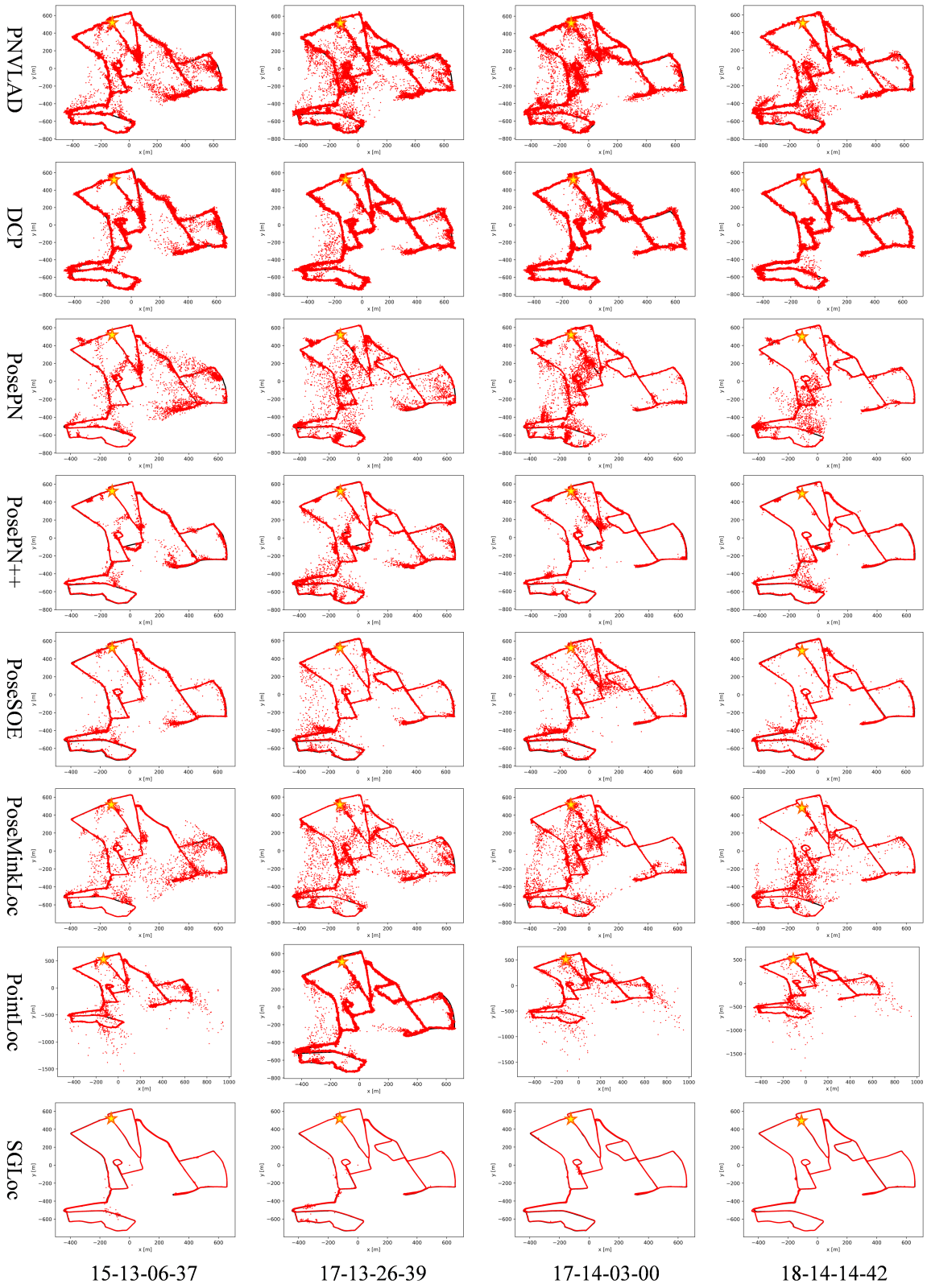

Figure 2. Trajectories of the baselines and the proposed method on the Oxford dataset. The ground truth and predictions are shown in black and red, respectively. The star indicates the starting position.

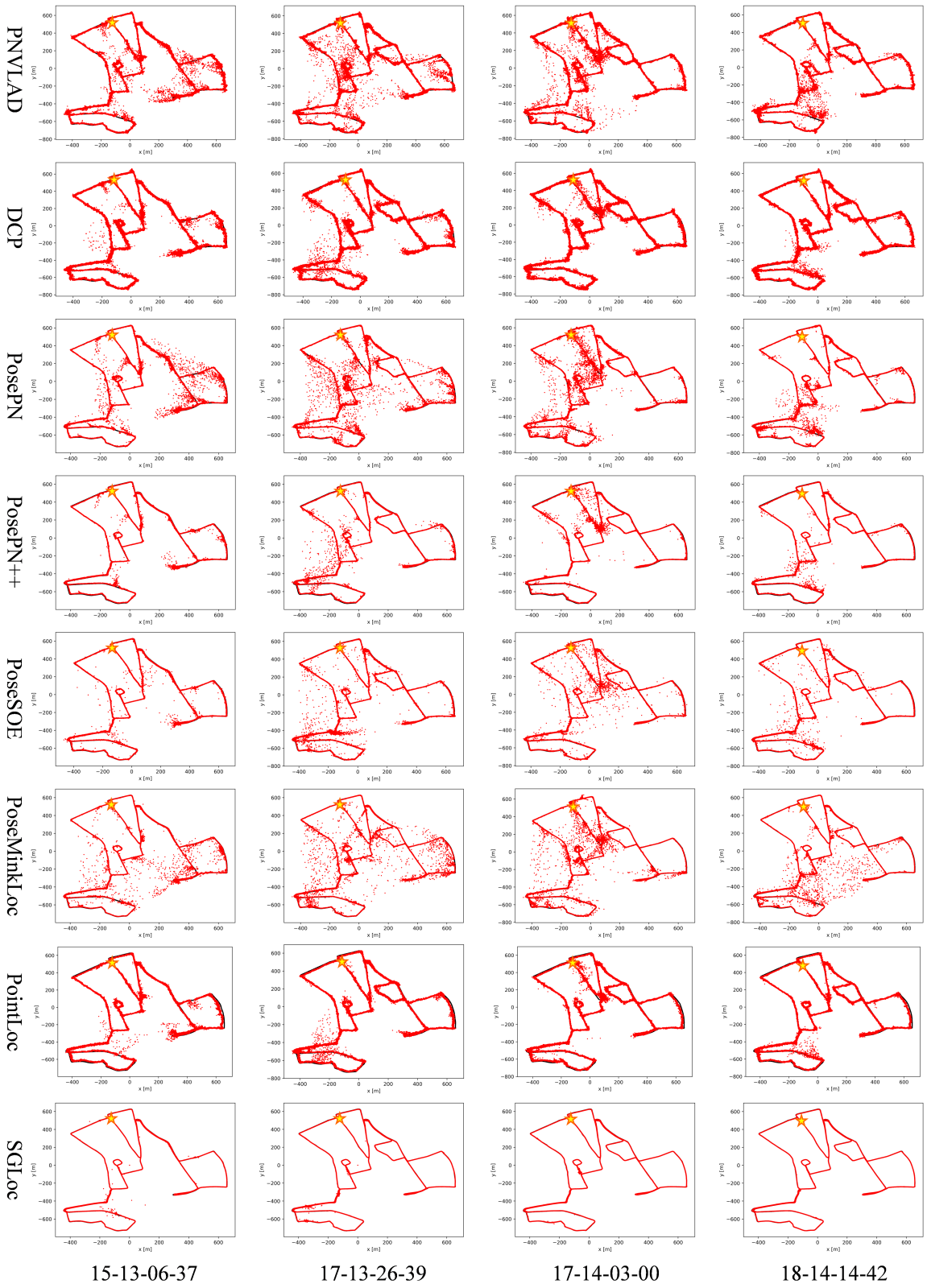

Figure 3. Trajectories of the baselines and the proposed method on the quality-enhanced Oxford dataset. The ground truth and predictions are shown in black and red, respectively. The star indicates the starting position.

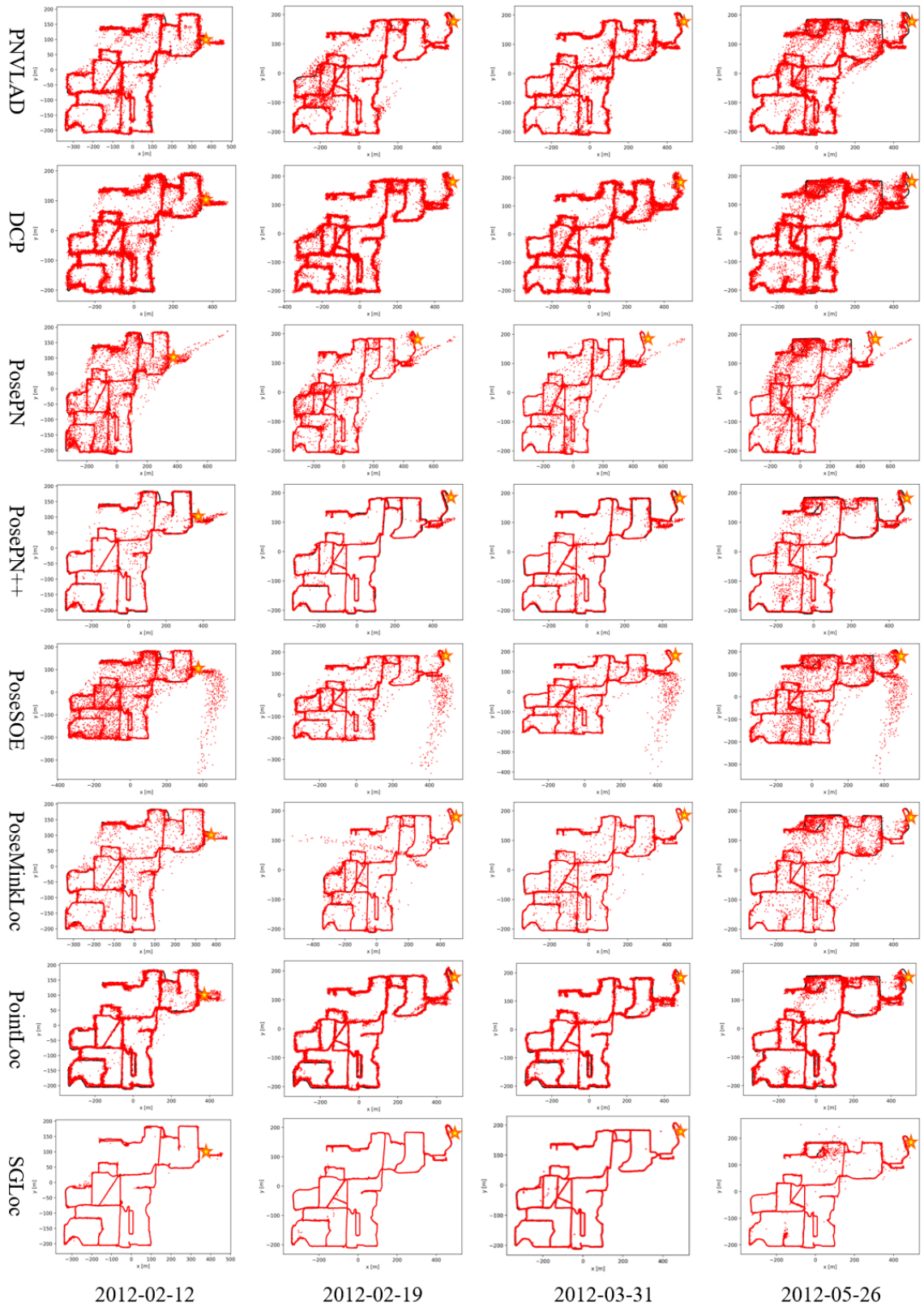

Figure 4. Trajectories of the baselines and the proposed method on the NCLT dataset. The ground truth and predictions are shown in black and red, respectively. The star indicates the starting position.