Supplementary Material for TransLoc4D: Transformer-based 4D Radar Place Recognition

Due to the space limitations of the manuscript, in this supplementary material, we provide additional details and experiments to support our proposed approach.

1. Descriptions on the Employed 4D Radars

The NTU4DRadLM [11] dataset is collected by an Oculii EAGLE 4D Imaging Radar, which is a 2-chip hardware platform. Each chip is automotive-grade radar that has 6 transmitting antennas and 8 receiving antennas (6T8R). Based on Oculii's proprietary AI-powered Virtual Aperture Imaging technology, a 50X increase in angular resolution is enabled. The Oculii EAGLE 4D radar works in the frequency band between 77-79 GHz, and can output a frame of point cloud with about 4000 points every 50ms.

In SJTU4D dataset [7], ZF FRGen21 4D Radar is used. It has 12 transmitting antennas and 16 receiving antennas (12T16R) to generate a total of 192 channels. It works in the frequency band from 76 GHz to 77 GHz. A frame of point cloud including about 400 to 1400 points can be obtained every 60ms.

The detailed specifications of the two radars employed in NTU4DPR and SJTU4DPR are presented in Tab. 1. Compared with 3D LiDAR, 4D radar has a limited Field-Of-View (FOV) and a limited number of points per frame. These bring greater challenges to the place recognition task based on 4D radar.

Table 1. Specifications of the two types of 4D radar employed.

		Field	Resolution	Accuracy
Oculii Eagle	Range	0∼400m	≤0.86m	≤0.16m
	Azimuth	-56.5°∼56.5°	0.5°	0.44°
	Elevation	-22.5°∼22.5°	1.0°	0.175°
	Velocity	-86.8∼86.8 m/s	0.27 m/s	0.09 m/s
	Frenquency	77-79GHz	-	-
	Framerate	15Hz	-	-
	Points/frame	around 4000	-	-
ZF FRGen21	Range	0∼350m	≤0.2m	≤0.02m
	Azimuth	-75°∼75°	1.5°	0.15°
	Elevation	-15°∼15°	1.5°	0.3°
	Velocity	-40∼40 m/s	0.1 m/s	0.01 m/s
	Frenquency	76-77GHz	-	-
	Framerate	16Hz	-	-
	Points/frame	400∼1400	-	-

2. More Details on the Proposed Datasets

2.1. Expanded Test Sets of NTU4DPR

Given that the NTU4DRadLM dataset (from which the NTU4DPR is generated) only comprises data gathered during sunny daytime, we expand the NTU4DPR with additional trajectories collected under different weather conditions, to validate 4DRPR and our proposed method across diverse environments. Specifically, two additional subsets are collected at Nanyang Technological University (NTU), in the Sports and Recreation Center (SRC) and at Nanyang Link (NYL). Depicted in Fig. 1, the path within SRC spans approximately 1.1 km, while the route in NYL is over 1 km.

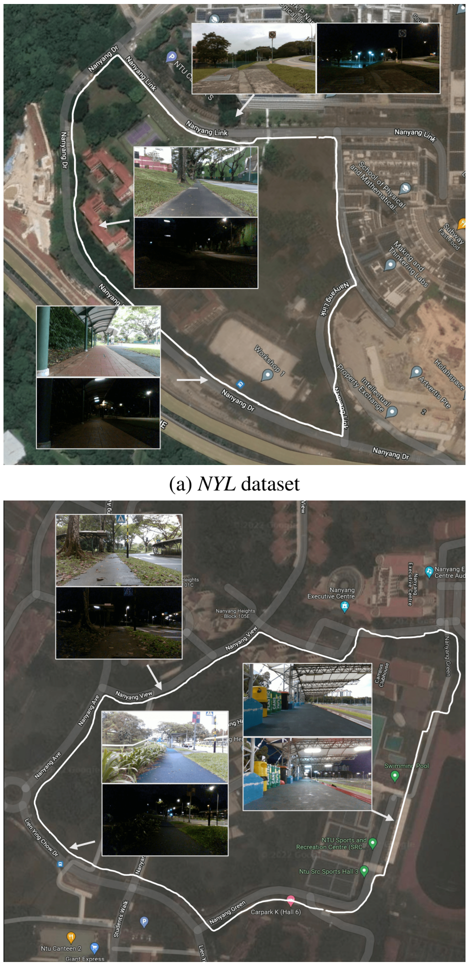
The two new subsets, namely NTU4DPR-NYL and NTU4DPR-SRC, are collected on sidewalks instead of main roads. Characterized by similar and repetitive structures, sidewalk scenes pose a challenge for place recognition. Noticeably, data collection spans various periods, including day and night, and extends to a range of weather conditions, from sunny days to light and moderate rainfall. NTU4DPR-NYL contains three trajectories along a repeated route in NYL on cloudy, night, and rainy days respectively. The nighttime and rainy trajectories are sampled as two query splits, NYL-Night and NYL-Rain. The remaining split collected during cloudy daytime is sampled as the database. Similarly, the NTU4DPR-SRC contains two repeated trajectories of SRC collected during daytime and nighttime. They are sampled as the query split SRC-Night and the corresponding database respectively. As expanded test sets of NTU4DPR, NYL-Night, NYL-Rain, and SRC-Night are used as challenging queries to evaluate the cross-domain robustness of the comparative models. The statistics of each new data split can refer to Tab. 1 in the main text.

2.2. Datasets Statistics

The statistics of the proposed 4DRPR datasets, summarized in Tab. 2, provide the distribution of point-based measurements within each dataset. The 'Min Points' represents the minimal number of points recorded in a frame across all samples in the subset. The 'Max Points' reflects the maximal number of points, and the 'Mean

Table 2. Summary of Dataset Statistics with Database and Query Subsets

Dataset	Subset	Min Points	Max Points	Mean Points	Std Dev Points	Min Velocity	Max Velocity	Min Intensity	Max Intensity
NTU-Train	Database	1067	6824	3388.91	976.72	−27.06	11.99	0.00	33.87
	Query	692	6099	3203.90	954.29	−34.20	14.65	0.00	33.20
NTU-Test	Database	491	9120	3603.99	1610.18	−28.95	29.04	0.00	35.34
	Query	848	8830	3610.02	1437.11	−28.42	17.40	0.00	34.48
NYL-Night	Database	710	7876	3994.43	1259.63	−18.50	23.84	0.00	42.05
	Query	1154	7904	4120.79	1306.73	−12.49	17.45	0.00	41.79
NYL-Rain	Database	710	7876	3994.43	1259.63	−18.50	23.84	0.00	42.05
	Query	392	6631	3179.70	1245.04	−18.20	23.98	0.00	40.60
SRC-Night	Database	821	8081	4886.37	1522.73	−10.70	17.58	0.00	42.21
	Query	867	8122	5124.07	1430.19	−9.86	8.79	0.00	40.97
SJTU-TestA	Database	638	1606	1132.22	172.33	−26.42	26.41	50.00	121.80
	Query	392	1515	1066.90	157.73	−26.44	26.42	50.00	122.37
SJTU-TestB	Database	378	1613	1084.05	218.58	−26.43	26.43	50.00	121.79
	Query	145	1543	1087.45	227.13	−26.44	26.43	50.00	134.76



(a) *NYL* dataset

(b) *SRC* dataset

Figure 1. Satellite views of *NTU4DPR-SRC* and *NTU4DPR-NYL*

.

Table 3. The comparisons of taking the original relative radial velocity (-V^d) and the proposed relative azimuth angle (-V) as the model input.

Method	NTU4DPR-Test			SJTU4DPR-TestA		
	r@1	r@5	r@10	r@1	r@5	r@10
TransLoc4D-R	92.7	94.5	95.1	88.6	93.2	94.0
TransLoc4D-R-V^d	94.7	95.7	96.2	55.4	64.2	67.3
TransLoc4D-R-V	94.3	95.9	96.5	89.5	93.2	94.1
TransLoc4D-R-VI	95.5	96.3	96.6	89.0	92.4	93.3
Method	NYL-Rain			SRC-Night		
	r@1	r@5	r@10	r@1	r@5	r@10
TransLoc4D-R	81.0	88.4	91.5	89.0	94.5	96.4
TransLoc4D-R-V^d	68.8	81.0	85.9	68.4	82.1	87.3
TransLoc4D-R-V	83.3	89.5	93.4	93.6	97.3	98.3
TransLoc4D-R-VI	82.5	89.7	92.1	94.4	96.9	97.9

Points' indicates the average number of points across all samples in the subset. The NTU4DPR and SJTU4DPR were captured using different radars, which can be evidenced by the stark differences in statistics. For instance, the SJTU-TestB exhibits significantly fewer mean points (1084.05) than the SRC-Night (5124.07), which poses challenges in cross-dataset generalization due to the disparity in point density and distribution. Additionally, the SJTU subsets have a higher minimum intensity of 50.00 when compared to the other NTU subsets with a minimum intensity of 0.0. These discrepancies imply the necessity of handling source-specific characteristics when developing models that generalize well across different datasets. Therefore, in our experiments, intensity readings are normalized to zero mean and 0.1 standard deviation to mitigate divergence caused by different radar used.

3. More Experimental Results

3.1. Utilization of Velocity Attribute

As stationary points in a scene have the same relative velocity to the radar, the difference in their relative radial

Table 4. Comparisons with Scan Context, Intensity Scan Context, and other variants on 4D radar datasets.

Method	NTU4DPR-Test			NYL-Night			NYL-Rain			SRC-Night			SJTU4DPR-TestA			SJTU4DPR-TestB		
	r@1	r@5	r@10	r@1	r@5	r@10	r@1	r@5	r@10	r@1	r@5	r@10	r@1	r@5	r@10	r@1	r@5	r@10
Scan Context [5]	74.8	88.4	91.9	23.4	48.2	63.4	12.4	28.4	40.0	23.8	48.2	60.3	35.8	72.4	80.8	61.9	80.4	86.3
Azimuth Scan Context	78.5	88.8	91.6	85.4	92.3	94.0	58.7	73.6	79.1	71.0	84.0	89.2	79.7	84.8	86.6	79.0	85.5	88.4
Intensity Scan Context [9]	90.0	93.4	94.3	87.2	92.8	94.7	69.1	80.3	**94.3**	68.1	82.2	86.9	67.9	79.8	83.7	78.0	86.5	90.2
TransLoc4D-RVIT (ours)	**95.1**	**96.1**	**96.4**	**97.1**	**98.4**	**98.7**	**86.8**	**91.8**	94.0	**94.5**	**97.0**	**98.0**	**90.8**	**92.9**	**93.4**	**85.9**	**88.7**	**90.5**

velocities implicitly reflects their positions, where geometric patterns could be mined. Therefore, the velocity attribute of 4D radar scans may contain potentially beneficial information for 4DRPR task. With this intuition, we attempt to explore the velocity attribute for scene description.

Specifically, we propose the relative azimuth angle derived from the velocity attribute of 4D radar as a numerical feature. Table.1 in the main text shows TransLoc4D-R-V steadily outperforms TransLoc4D-R, verifying the effectiveness of the numerical representation of velocity attribute (-V) coupled with point cloud refinement (-R). The numerical velocity representation (-V) employed here refers to our proposed relative azimuth angle, which eliminates the bias caused by ego-velocity of the 4D radar.

To further validate the new attribute of relative azimuth, we compare the two TransLoc4D-R-V variants respectively taking the original relative radial velocity v^d and the relative azimuth attribute s as input. The benchmark with original velocity is denoted as TransLoc4D-R-V^d.

As can be seen in Tab. 3, TransLoc4D-R-V^d constantly underperforms TransLoc4D-R. A drastic performance drop of over 10% can be observed on SJTU4DPR-TestA, NYL-Rain, and SRC-Night. This proves our conjecture that directly incorporating radial relative velocity v^d into feature embedding may introduce bias, causing the model to learn harmful tricks. Besides, TransLoc4D-R-V leads TransLoc4D-R-V^d by a large margin on all datasets. It demonstrates the advantages of the new attribute s over the original velocity attribute v^d. It also verifies the rationality of decoupling the speed and direction of velocity to formulate a new attribute of relative azimuth independent of 4D radar ego-velocity.

Combined with the experiments in the main text, a conclusion can be drawn that the velocity attribute of 4D radar can be effective in eliminating dynamic interference and bringing robustness to the 4DRPR task.

3.2. Comparisons with 4D Radar Scan Context

In Sec.4.4 of the main text, we adapt the State-Of-The-Arts (SOTA) learning architectures for 3D LiDAR place recognition, MinkLoc3Dv2 [6], TransLoc3D [10], and PTC-Net [4] to the 4D radar place recognition task. On this basis, we compare the proposed TransLoc4D with the adapted benchmark architectures and demonstrate its superiority.

Besides learning-based architectures that learn scene description directly from 3D points, another widely recognized handcrafted 3D point cloud descriptor is Scan Context [5]. It converts point clouds into polar coordinate images and proposes bird's-eye view partitioning to construct regional maximum height features. Its representative variant, Intensity Scan Context [9], uses the maximum intensity instead of height to construct the scan context matrix.

In order to compare them with our proposed TransLoc4D in 4D radar place recognition, we adapt Scan Context and Intensity Scan Context to 4D radar first. While 3D LiDAR has a 360°azimuth Field Of View (FOV), 4D radar is with a narrower FOV (110°for NTU4DPR and 150°for SJTU4DPR). Therefore, we divide the 110°sector area into 40 rings and 20 sectors accordingly. The feature value of each bin is chosen as the maximum height and the maximum intensity respectively for Scan Context and Intensity Scan Context. Other encoding steps remain the same as in Vanilla 3D LiDAR Scan Context. Considering the velocity attribute of 4D radar scanning, we set up an Azimuth Scan Context similar to the Intensity Scan Context.

As in Tab. 4, Azimuth Scan Context significantly surpasses the baseline Scan Context, showing that the azimuth attribute is a better feature than the height for 4D radar Scan Context. Intensity Scan Context outperforms Azimuth Scan Context on most subsets, except for SRC-Night and SJTU4DPR-TestA. It demonstrates the intensity attribute to be discriminative for the 4DRPR task. Although Scan Context variants demonstrate good adaptability to the 4D radar place recognition task, they are handcrafted descriptors that are not learnable. Our TransLoc4D is an end-to-end encoding architecture that is differentiable for fine-tuning. TransLoc4D outperforms the second-best Scan Context variant by a large margin of over 10% on NYL-Rain, SRC-Night, and SJTU4DPR-TestA, which can be attributed to the deep model and data-driven fine-tuning. This set of experiments shows the potential of our TransLoc4D as a replacement for the Scan Context family in robotic tasks, such as loop closure detection in 4D radar Simultaneous Localization and Mapping (SLAM) [12, 13].

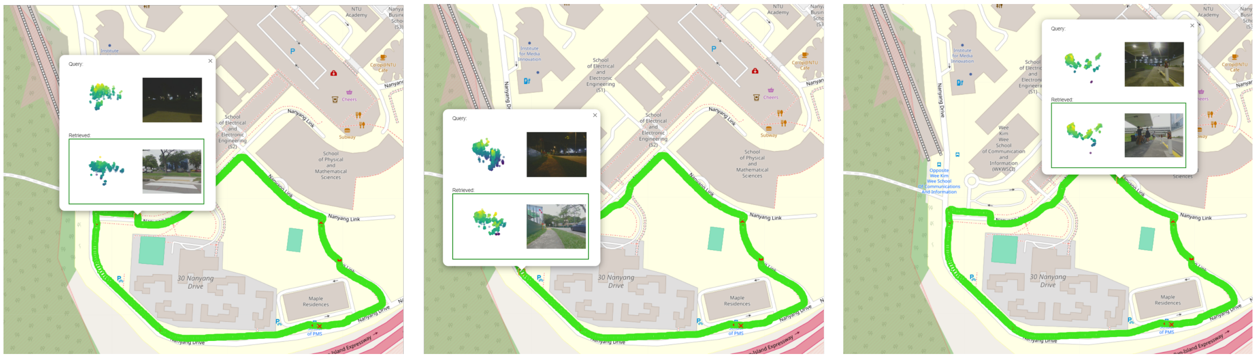


Figure 2. Instances of TransLoc4D retrieval on NTU4DPR-NYL dataset with nighttime queries. By 4D point could descriptor matching, TransLoc4D is able to correctly retrieve the reference point cloud when images exhibit drastic appearance differences.

Table 5. Comparison with SOTA 2D image place recognition methods on 4D radar datasets.

NTU4DPR-Test	recall@1	recall@5	recall@10
CosPlace [2]	97.9	98.0	98.1
EigenPlaces [3]	**98.1**	**98.2**	**98.3**
TransLoc4D (ours)	95.1	96.1	96.4
NYL-Night	recall@1	recall@5	recall@10
CosPlace [2]	62.1	67.1	70.4
EigenPlaces [3]	73.6	79.6	82.3
TransLoc4D (ours)	**97.1**	**98.4**	**98.7**
SRC-Night	recall@1	recall@5	recall@10
CosPlace [2]	84.5	89.2	90.8
EigenPlaces [3]	87.3	91.8	93.5
TransLoc4D (ours)	**94.5**	**97.0**	**98.0**

3.3. Comparisons with 2D Image Place Recognition

2D image place recognition (2DVPR) is more popular than 3D point cloud-based methods due to the easy availability and established solutions of RGB sensors. Despite the lack of 3D information, the rich features of the RGB sensor guarantee reliable recognition capabilities. However, the effectiveness of 2DVPR is compromised by sensitivity to lighting changes, where drastic illumination shifts significantly affect data distribution and challenge its robustness.

Contemporary image-based algorithms typically employ a CNN [1] or Transformer [8] for feature extraction from images, followed by aggregation using NetVLAD [1] or GeM [2] pooling to create a descriptor vector. Recent advancements, notably CosPlace [2] and EigenPlace [3], have significantly improved VPR benchmarks. Their advantages can be attributed to training models on categorization tasks with Large Margin Cosine Loss and using UTM coordinates and image orientations to refine the training process.

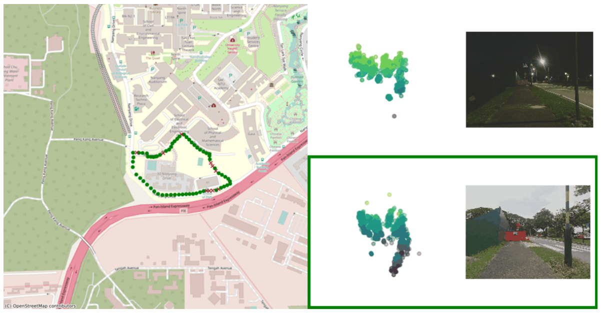


Figure 3. GIF visualization of retrieval results on NTU4DPR-NYL dataset. The challenging nighttime queries retrieve the daytime database. Red and green circles (left) and bounding boxes (right-bottom) indicate incorrect and correct retrievals, respectively. Long press the mouse on the GIF to pause and view the point cloud and image retrieved by a single frame.

To benchmark our proposed TransLoc4D against leading 2D image-based place recognition algorithms, we reimplemented CosPlace [2], EigenPlaces [3] and utilized their publicly available checkpoint. Given that NTU4DPR offers both images and 4D radar data, it allows for a direct comparison of these single-source algorithms. All data, including images, 4D radar, and GPS, are synchronized using their respective timestamps. The evaluation of these algorithms was conducted across three distinct subsets of the NTU4DPR dataset: NTU4DPR-Test, NYL-Night, and SRC-Night. Both the NYL-Night and SRC-Night subsets utilize nighttime queries, whereas the test set comprises queries captured during daytime.

As illustrated in Tab. 5, EigenPlaces [3] demonstrates strong performance in same-domain retrieval. However, its performance is significantly reduced when query and reference images are from different domains. In contrast, our TransLoc4D maintains nearly consistent performance across different scenarios (day to night, main roads to sidewalks, vehicle mounted to handheld), which demonstrates its outstanding robustness against adverse environments.

4. Additional Visualization Results

Fig. 3 gives a GIF visualization of retrieval results of our TransLoc4D on the challenging NTU4DPR-NYL dataset. It can be seen that TransLoc4D can correctly retrieve most frames when the query and reference are from different data domains. It verifies the feasibility of our TransLoc4D for 4D radar place recognition tasks. In Fig. 2, when images exhibit drastic appearance differences, 4D point clouds still show stable similarities, which reflects the advantages of 4D radar over 2D cameras in handling harsh environments and dynamic objects. Fig. 4 presents some challenging queries in NTU4DPR-NYL dataset and the top retrieved images using different models. As can be seen, when other benchmark models fail, TransLoc4D can still retrieve the queries correctly. This demonstrates the better robustness of our proposed model. Overall, it can be inferred from the additional visualization results that, taking advantage of the 4D radar attributes, TransLoc4D can easily cope with practical challenges in the place recognition task, such as illumination changes, and dynamic occlusions.

5. Application Scenarios and Future Work

The value of 4DRPR mainly lies in its ability to enable robust perception in harsh conditions, such as heavy rain, snow, smoke, fog, and dust. Specifically, 4DRPR enables robust localization and re-localization in such adverse conditions, where traditional camera and LiDAR will fail to perform such tasks. One most straightforward application of 4DRPR is it can be used in 4D radar SLAM [12] as the loop closure module, since correct loop closure is of key importance to the back-end optimization of SLAM system.

The potential application scenarios include but are not limited to the following:

- Firefighting robot which operates in heavy smoke and fog environment.
- Unmanned vehicles which run in heavy fog or heavy snow environments.
- Autonomous bulldozers that operate in heavy dust environments.
- Mining robots that run in an underground environment with heavy dust and low illumination.

For future work, two directions are worth exploring: 1) Pre-processing of 4D point cloud. As mentioned before, the point cloud of 4D radar is sparse, cluttered and noisy, which poses challenges to the 4DRPR task. Thus, a better strategy to pre-process the 4D point cloud could be researched to improve the model performance. For example, densification or completion of sparse 4D point clouds. 2) Multi-modal based place recognition. Considering that different sensors have their own advantages and disadvantages, in some ways they can complement each other. Thus, it is worth exploring to fuse other sensors together with 4D radar for the place recognition task, to achieve all-weather place recognition with superior performance.

References

[1] Relja Arandjelovic, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic. Netvlad: Cnn architecture for weakly supervised place recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5297–5307, 2016. 4

[2] Gabriele Berton, Carlo Masone, and Barbara Caputo. Rethinking visual geo-localization for large-scale applications. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4868–4878, 2022. 4

[3] Gabriele Berton, Gabriele Trivigno, Barbara Caputo, and Carlo Masone. Eigenplaces: Training viewpoint robust models for visual place recognition. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11046–11056, 2023. 4

[4] Lineng Chen, Huan Wang, Hui Kong, Wankou Yang, and Mingwu Ren. Ptc-net: Point-wise transformer with sparse convolution network for place recognition. *IEEE Robotics and Automation Letters*, 8(6):3414–3421, 2023. 3

[5] Giseop Kim and Ayoung Kim. Scan Context: Egocentric Spatial Descriptor for Place Recognition Within 3D Point Cloud Map. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4802–4809, 2018. 3

[6] Jacek Komorowski. Improving point cloud based place recognition with ranking-based loss and large batch training. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 3699–3705, 2022. 3

[7] Xingyi Li, Han Zhang, and Weidong Chen. 4d radar-based pose graph slam with ego-velocity preintegration factor. *IEEE Robotics and Automation Letters*, 8(8):5124–5131, 2023. 1

[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Neural Information Processing Systems (NeurIPS)*, 2017. 4

[9] Han Wang, Chen Wang, and Lihua Xie. Intensity Scan Context: Coding Intensity and Geometry Relations for Loop Closure Detection. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2095–2101, 2020. 3

[10] Tianhan Xu, Yuanchen Guo, Yu-Kun Lai, and Song-Hai Zhang. Transloc3d : Point cloud based large-scale place

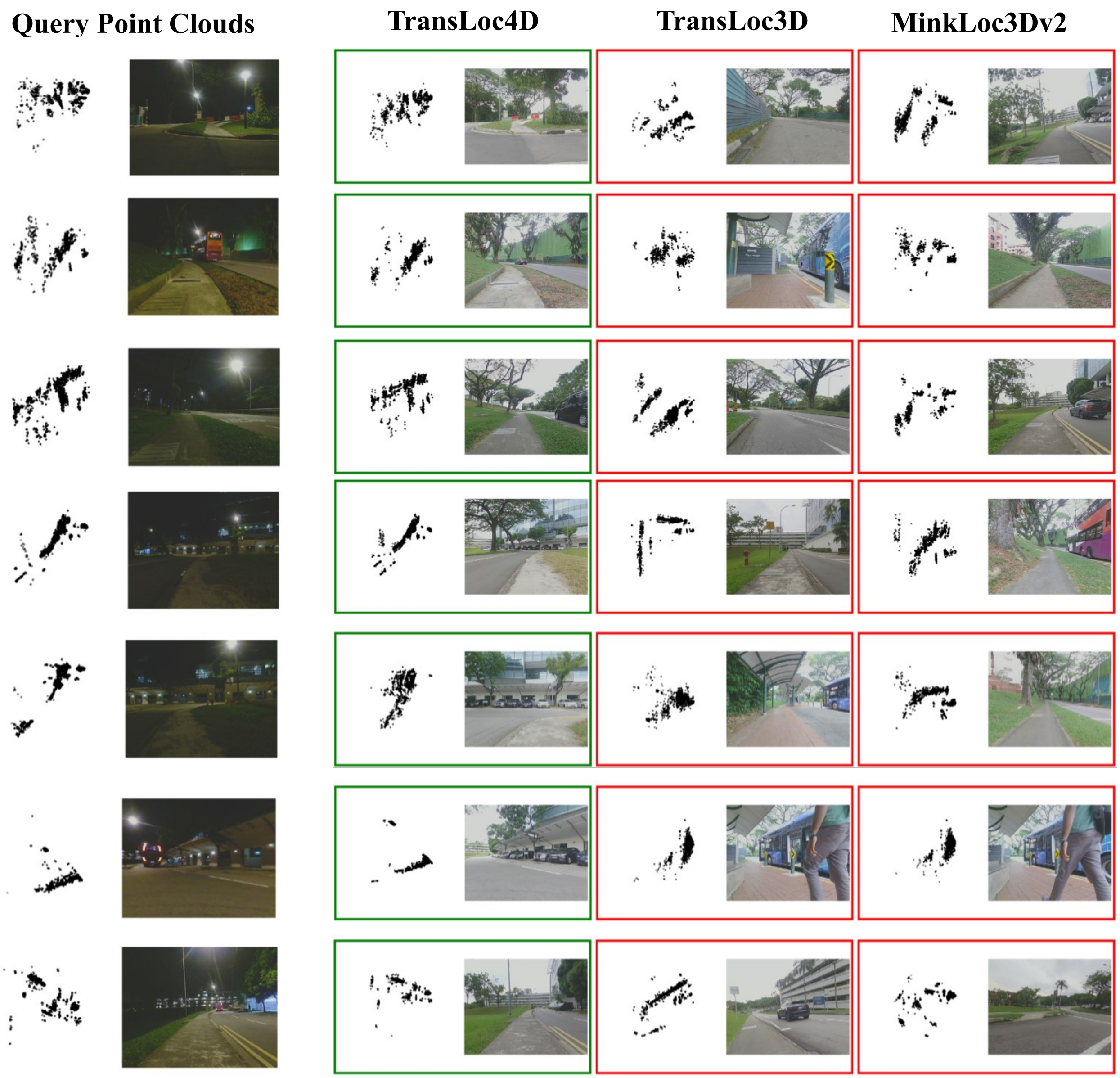


Figure 4. Example retrieval results. From left to right: query frames each with a pair of point cloud and image, the top retrieved frame using our method (TransLoc4D), the top retrieved frame using the adapted SOTAs (TransLoc3D and MinkLoc3Dv2). Green and red borders indicate correct and incorrect retrieved results respectively.

recognition using adaptive receptive fields. *Communications in Information and Systems*, 23:57–83, 2021. 3

[11] Jun Zhang, Huayang Zhuge, Yiyao Liu, Guohao Peng, Zhenyu Wu, Haoyuan Zhang, Qiyang Lyu, Heshan Li, Chunyang Zhao, Dogan Kircali, Sanat Mharolkar, Xun Yang, Su Yi, Yuanzhe Wang, and Danwei Wang. Ntu4dradlm: 4d radar-centric multi-modal dataset for localization and mapping. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, pages 4291–4296, 2023. 1

[12] Jun Zhang, Huayang Zhuge, Zhenyu Wu, Guohao Peng, Mingxing Wen, Yiyao Liu, and Danwei Wang. 4dradarslam: A 4d imaging radar slam system for large-scale environments based on pose graph optimization. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8333–8340, 2023. 3, 5

[13] Yuan Zhuang, Binliang Wang, Jianzhu Huai, and Miao Li. 4d iriom: 4d imaging radar inertial odometry and mapping. *IEEE Robotics and Automation Letters*, 8(6):3246–3253, 2023. 3