

# RALoc: Enhancing Outdoor LiDAR Localization via Rotation Awareness

Yuyang Yang<sup>1,2\*</sup> Wen Li<sup>1,2\*</sup> Sheng Ao<sup>1,2</sup> Qingshan Xu<sup>3</sup> Shangshu Yu<sup>4</sup> Yu Guo<sup>1,2</sup>  
Yin Zhou<sup>5</sup> Siqi Shen<sup>1,2</sup> Cheng Wang<sup>1,2†</sup>

<sup>1</sup>Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University

<sup>2</sup>Key Laboratory of Multimedia Trusted Perception and Efficient Computing,  
Ministry of Education of China, Xiamen University

<sup>3</sup>Nanyang Technological University <sup>4</sup>Northeastern University <sup>5</sup>GAC R&D Center

## Abstract

*LiDAR localization is a fundamental task in autonomous driving and robotics. Scene Coordinate Regression (SCR) exhibits leading pose accuracy, achieving impressive results in learning-based localization. We observe that the real-world LiDAR scans captured from different viewpoints usually result in the catastrophic collapse of SCR. However, existing LiDAR localization methods have largely overlooked the issue of rotation sensitivity in SCR. In this paper, we present RALoc, an outdoor LiDAR localization method with rotation awareness to achieve accurate localization. The key to our approach is to design a Point Cloud Canonicalization module, which leverages a powerful equivariant key feature aggregation to transform the input LiDAR scan towards a consistent orientation, effectively eliminating the adverse effects of rotation. This proposed module has promising scalability and can be seamlessly integrated with the existing LiDAR localization network. Moreover, we propose the **Bidirectional LiDAR Localization (BiLiLo)** dataset as a benchmark to evaluate the performance of various methods in large outdoor scenes with significant rotation changes. Extensive experiments show that RALoc significantly improves localization performance in scenarios with large rotation changes, and also achieves competitive performance in the Oxford Radar RobotCar dataset. Our project is available at <https://etheryangyy.github.io/raloc.github.io>.*

## 1. Introduction

LiDAR localization is a crucial task that plays an important role in various applications, such as autonomous driving [34], robotics and augmented reality. The objective of LiDAR localization is to use the point cloud of the scene

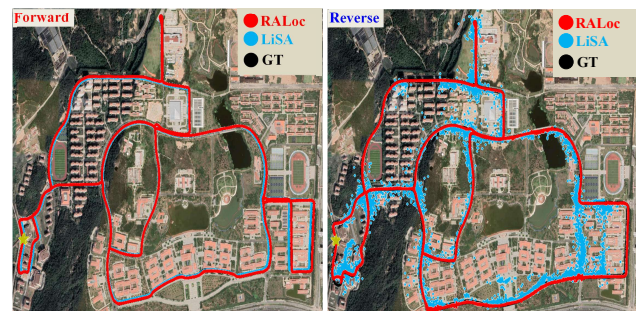


Figure 1. **Teaser.** This paper primarily investigates the problem of rotation-robust LiDAR localization. We present ground truth and predicted trajectories on the collected rotation changes dataset. Left: the forward trajectory (aligned with the training set driving direction). Right: the reverse trajectory (completely opposite to the training set driving direction). Compared to the current state-of-the-art method, LiSA [55], RALoc retains high accuracy in the forward trajectory and avoids catastrophic failures due to rotation in the reverse trajectory.

captured by the LiDAR as input and output its 6 degrees-of-freedom (DoF) pose. However, due to uneven data quality (e.g., noise distribution, non-uniform density, and varying viewing angles), it remains challenging to estimate accurate 6-DoF pose in large-scale real-world applications.

Most existing methods [12, 27, 45, 54] typically follow the retrieval-registration paradigm for self-localization, which involves retrieving candidate scenes from a database, verifying their correctness, and subsequently estimating the 6-DoF relative pose via point cloud registration [35, 40]. However, when dealing with a large amount of LiDAR point cloud data, this approach not only imposes significant demands on storage and communication resources, but also struggles to achieve efficient localization from scratch [57].

Learning-based regression has been shown to be a promising research direction to address the aforementioned issues by memorizing specific scenes within a neural network [57]. Existing methods can be categorized into Abso-

\*Equal contribution

†Corresponding author

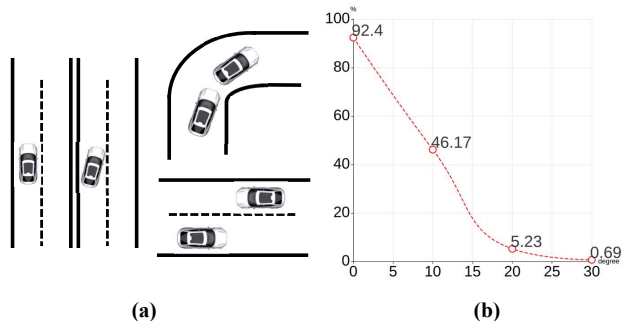


Figure 2. **Motivation.** (a) The rotation problems caused by changing lanes, turning, and reversing. (b) The impact of Yaw angle variation on LiSA’s [55] recall rate within 5 meters and 3 degrees.

lute Pose Regression (APR) and Scene Coordinate Regression (SCR). APR [25, 31, 50] directly estimates the global pose of the sensor in an end-to-end manner. Despite its effectiveness, sub-optimal results are achieved due to the highly abstract global feature used to represent the scene. Unlike the APR method, the SCR [7–9, 30] technique regresses the correspondences between the point cloud in the LiDAR coordinate and the world coordinate frame, followed by RANSAC [19] algorithm to estimate a precise 6-DoF pose. By explicitly capturing scene geometry, SCR achieves significant performance improvements.

Nonetheless, as shown in the right side of Fig. 1, we observe that the learning-based regression method suffers from catastrophic collapse due to large viewpoint discrepancies when the scanned point cloud is rotated relative to the training data. In practical applications, factors such as oncoming traffic, lane changes, and turns often result in unpredictable changes in vehicle direction. To illustrate the effect of rotation changes, we take a recently proposed superior method, LiSA [55], to uncover this problem. As shown in Fig. 2, we conducted a pilot study with various rotation changes and recorded the localization recall. Specifically, we randomly rotate the query point cloud by the Yaw axis to simulate the challenge of a real driving scenario. Clearly, recall decreases significantly as the rotation increases. When rotations exceed 30 degrees, the recall drops below 1%, indicating that localization algorithms fail in such cases. These experiments demonstrate that rotation changes negatively impact the performance of existing localization models.

We believe that the main reason is that the neural network used lacks robustness to rotations [47]. Although data augmentation can improve the rotational robustness of networks to some extent, it compromises the overall performance of the network and necessitates larger capacities along with more complex architectures [1, 6, 52]. In contrast, equivariant models [38, 39] provide a strong inductive bias for the network, which is expected to fundamen-

tally address the issue of regression methods being sensitive to rotations. Furthermore, we found that test scenarios in commonly used benchmarks [4, 11] are relatively simple and rarely account for situations where the vehicle direction changes. To close this gap, we contribute a large-scale outdoor localization dataset with substantial rotation variations, which can offer a more accurate basis for assessing model robustness under real-world conditions.

In this paper, we propose a novel localization framework, RALoc, to address the issue of current LiDAR localization models being vulnerable to rotations. Explicitly, we introduce an equivariant network to canonicalize the input point clouds. Additionally, to address the presence of noise interference in outdoor LiDAR localization tasks, we design an equivariant key feature aggregation module that effectively extracts scene features to better capture rotation changes. Moreover, to evaluate the rotational robustness of SCR in real-world driving scenarios, we establish equipment setups and collect a large-scale outdoor LiDAR localization dataset characterized by significant rotational variations.

Our contributions can be summarized as follows:

- We reveal that learning-based regression methods are sensitive to rotation changes, which greatly weakens their localization performance in real-world applications. To this end, we propose RALoc, the first attempt to introduce an equivariant network to SCR.
- We introduce a novel equivariant key feature aggregation module designed to mitigate interference from noisy point clouds, enhancing our approach with robust rotation-equivariant awareness.
- We collect a large-scale outdoor LiDAR localization dataset with significant rotation variations, providing an effective benchmark to facilitate related research.
- Extensive experiments on the proposed rotation challenge and Oxford Radar RobotCar [4] datasets demonstrate the effectiveness of our method. In particular, we show that RALoc achieves robust performance even with significant rotational variations (see Fig. 1).

## 2. Related work

### 2.1. Conventional Localization

Conventional localization methods aim to match the query point cloud with a pre-built 3D map, which can be classified into retrieval-based methods [29, 48, 53] and matching-based methods [23, 42, 46]. Retrieval-based methods cast localization as a problem of place recognition, which requires a pre-built database with feature descriptors and relies on querying the most similar point cloud in the database [37, 59]. Matching-based methods are typically based on either searching for similar frames in a pre-built global descriptor database or 3D point clouds and match features between the query frame and the map [2, 33]. Both

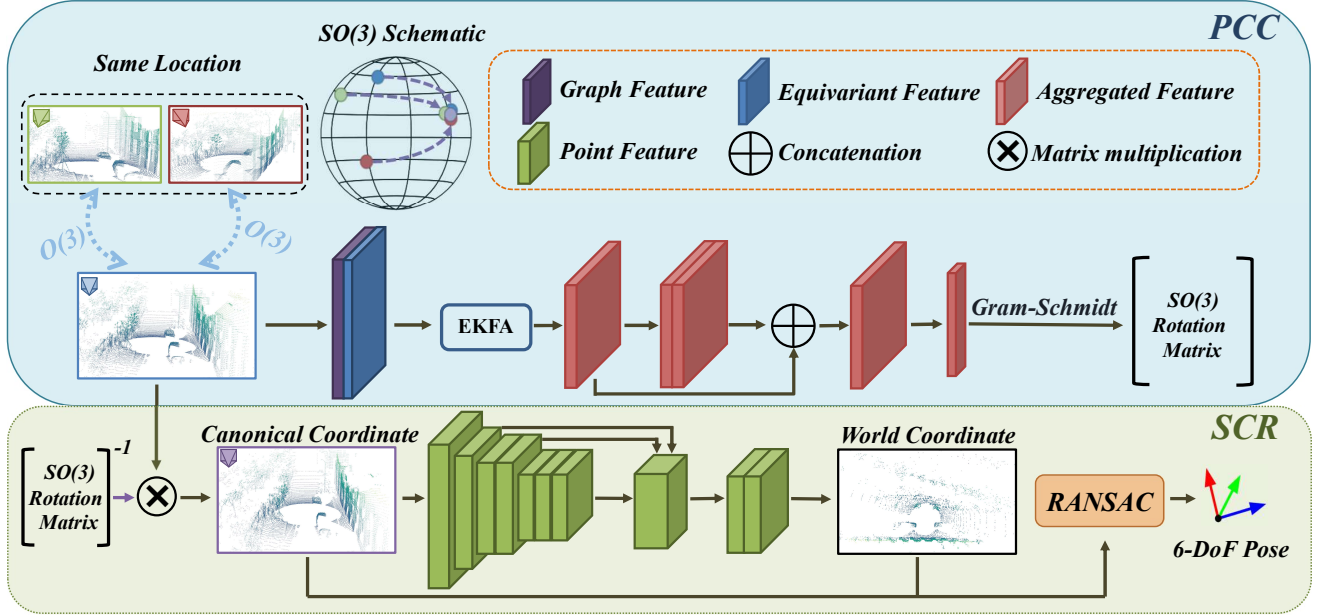


Figure 3. **The pipeline of RALoc.** RALoc consists of two main components: (1) the Point Cloud Canonicalization (PCC) network and (2) the Scene Coordinate Regression (SCR) network. Specifically, the PCC network captured from various rotations in the scene to a canonical view. The canonicalized point clouds are then fed into the SCR network, which outputs the 6-DoF pose.

methods depend on pre-stored maps, resulting in challenges related to storage and communication.

## 2.2. Regression-based Localization

Recent advances in deep learning have spurred interest in regression-based localization methods that directly predict 6-DoF poses end-to-end. PoseNet [25] is a widely used APR method that regresses camera poses via a feature encoder and regressor. Given LiDAR’s robustness to lighting and weather, APR has also been extended to LiDAR [58]. PointLoc [50] introduced LiDAR-based APR, and HypLiLoc [49] leveraged multi-level features to capture richer scene characteristics. DiffLoc [31] improved accuracy by transforming the pose regression process into multiple iterative steps through the diffusion model [22].

However, as mentioned in [30, 44], APR methods focus on high-dimensional pose-feature mappings, like retrieval-based techniques, but overlook scene geometry, limiting their accuracy. To address this issue, SCR predicts the coordinates of each point in the world coordinate, effectively encoding scene geometry. Ultimately, RANSAC is used to estimate the pose. SGLoc [30] is the first to incorporate SCR into LiDAR localization, followed by LiSA [55], which mitigates the impact of disruptive points in the scene by introducing semantic information, further enhancing accuracy. Although APR and SCR have achieved commendable accuracy in LiDAR localization, these methods struggle with rotation-related issues in realistic scenarios.

## 2.3. SO(3)-Equivariant Network

With growing performance issues due to rotation, equivariant networks have seen notable advances in recent years [14, 20, 32, 43, 51, 56]. Although data augmentation can mitigate rotation issues during training, this often affects the overall performance of the model when dealing with large datasets [6, 38, 52]. Therefore, maintaining rotational equivariance is essential when processing point cloud data. Various studies [13, 36, 61] have effectively addressed this problem through different approaches. Among them, the Vector Neurons [16] (VNN) is a lightweight equivariant framework that integrates well with existing point cloud processing networks, and many subsequent works [3, 24, 60] have expanded upon it. Consequently, we have constructed an equivariant network for processing outdoor LiDAR point clouds using VNN as the foundation. Moreover, we design an equivariant key feature aggregation module to further boost rotation robustness.

## 3. Method

Recent SCR methods [30, 55] have achieved impressive results in outdoor LiDAR localization. However, in scenarios with vehicular trajectory variations, as shown in Fig. 2 (a), they fail due to rotational deviations between scanned point clouds and the training data. In this paper, we propose RALoc, which utilizes the equivariant network to solve the problem of rotation sensitivity by point cloud canonicaliza-

tion (Sec. 3.1). In addition, to evaluate the rotation robustness of localization methods in the real driving scenario, we propose a large-scale outdoor dataset with significant rotation variations (Sec. 3.2).

### 3.1. RALoc

We now introduce RALoc, as shown in Fig. 3, which can be divided into (1) **Point Cloud Canonicalization (PCC)**: aligning LiDAR point cloud data with different orientations, efficiently sensing rotations through an equivariant network, and reorienting the LiDAR point clouds accordingly. (2) **Scene Coordinate Regression (SCR)**: estimate the global pose by matching the scanned point cloud to the regressed point cloud in the world coordinate [30].

In this work, we design a VNN-based [16] equivariant network to implement PCC. This network extracts LiDAR point cloud features using VN-MLP, aggregates scene features through Equivariant Key Feature Aggregation (EKFA), and then regresses the rotation matrix with respect to scene rotation changes.

**Point cloud canonicalization.** Here we formulate the framework of PCC. For point clouds captured from diverse orientations, we aim to align them into a consistent reference frame, termed as the canonicalization coordinate.

To establish this coordinate, we leverage the trajectory data from the training set. For a fixed position in the scene, the canonicalization coordinate is defined as the orientation of the LiDAR at that position in the training set.

Specifically, given a point cloud scanned from arbitrary vehicle direction  $Rx$ , where  $R \in SO(3)$ ,  $x \in \mathbb{R}^{N \times 3}$  represent the point cloud in the LiDAR coordinate, PCC can be represented as follows:

$$PCC(Rx) = \phi(Rx)^{-1} Rx, \quad (1)$$

where  $\phi$  denotes a VNN-based equivariant network. This network ensures that the PCC produces consistent outputs by leveraging the properties of equivariant networks:

$$\phi(Rx) = R\phi(x). \quad (2)$$

As shown in Fig. 3, we utilize the equivariant network to perceive the rotational state of the vehicle at the feature level, and regress the matrix. Then, we use the Schmidt orthogonalization, which is also equivariant (as proved in the supplementary), to ensure its a rotation matrix. Finally, we apply the inverse transformation of this rotation to the input point cloud, resulting in a canonical point cloud. This canonical point cloud can then be fed into the subsequent LiDAR localization network, enhancing the robustness of the localization network against rotation. More details on the architecture can be found in the supplementary.

**Equivariant key feature aggregation.** As described above, the task of equivariant network  $\phi$  is to ensure that

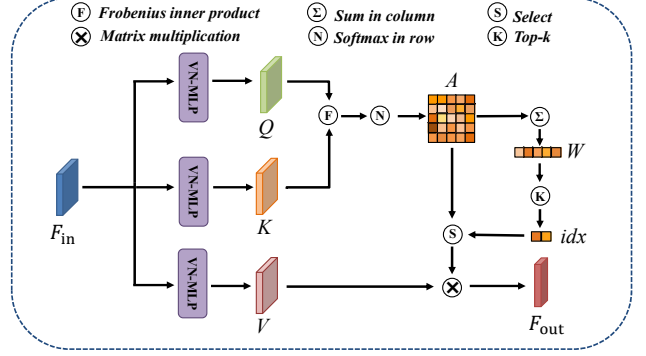


Figure 4. Equivariant Key Feature Aggregation (EKFA) module.

point clouds scanned from different orientations of the same scene exhibit the same rotational relationship in their extracted equivariant features. However, due to the challenges of outdoor scenes, such as dynamic objects, not all points benefit the extraction of robust features [55]. Therefore, to effectively select the useful features, we introduce an equivariant key feature aggregation module.

As shown in Fig. 4, the EKFA module takes the equivariant features  $F_{in} \in \mathbb{R}^{C \times 3 \times N}$  as input. We first use VN-MLP to generate features, denoted as  $Q$ ,  $K$ , and  $V$ , each with the same dimension of  $F_{in}$ . Then, to ensure strict equivariance, we compute the Frobenius inner product [3] of  $Q$  and  $K$ , divide it by  $\sqrt{3C}$ , and apply softmax to obtain the attention matrix  $A$ . This process can be expressed as follows:

$$A = \text{Softmax} \left( \frac{\langle Q, K \rangle_F}{\sqrt{3C}} \right), \quad (3)$$

where  $\langle \cdot \rangle_F$  is the Frobenius inner product, which is rotation invariant. The square root of the vector neurons features dimension count  $\sqrt{3C}$  serves as a scaling factor, and the shape of  $A$  is  $N \times N$ .

To obtain the aggregation weights  $W$ , we sum the attention matrix  $A$  along its last dimension (i.e., across columns), sort  $W \in \mathbb{R}^N$  in descending order, and select the top  $k$  indices to construct a key feature mask  $A'$  from  $A$ :

$$idx = \text{topk}(W), \quad (4)$$

$$A' = A[idx, :]. \quad (5)$$

Finally, we perform a matrix multiplication with  $V$  to obtain aggregated features enriched with key information after filtering out scene noise. The output can be expressed by matrix multiplication:

$$F_{out} = V \times A'^T. \quad (6)$$

**Scene coordinate regression.** After obtaining the canonicalized point cloud, the next step is to input it into SCR,

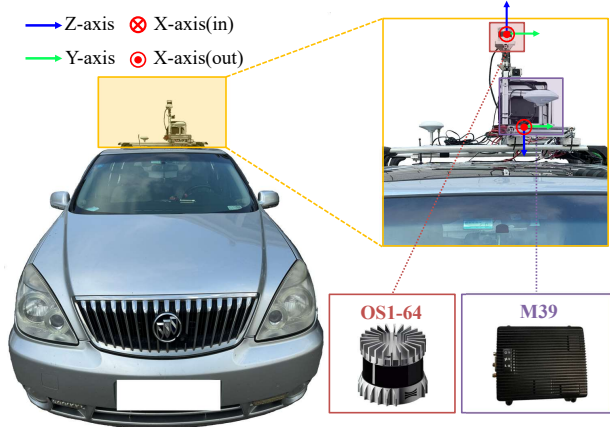


Figure 5. We scan point clouds and record ground truth using a vehicle equipped with an Ouster OS1-64 spanning LiDAR and a M39 GPS/IMU device. Coordinate frames illustrate the direction of each sensor on the vehicle with the convention: X-axis (red), Y-axis (green), and Z-axis (blue).

SGLoc [30] in this work, to get the global pose. Specifically, SCR regresses the corresponding point cloud in the world coordinate, and matches them inside a RANSAC loop for robust pose estimation.

**Loss function.** The output of RALoc consists of canonicalized point clouds and poses. Therefore, the total loss function  $\mathcal{L}$  should contain both canonicalization  $\mathcal{L}_{pcc}$  and regression  $\mathcal{L}_{reg}$  components. Following the definition of previous work [24, 38]. We choose the original LiDAR coordinate system in the training data as the canonical orientation, and utilize  $\mathcal{L}_{pcc}$  to keep point cloud canonicalization output point cloud orientation as constant as possible. Specifically, we use Mean Squared Error loss to optimize the point cloud canonicalization:

$$\mathcal{L}_{pcc} = \frac{1}{n} \sum_{i=1}^n (R_c - I)^2, R_c \in SO_3, \quad (7)$$

where the rotation matrix  $R_c$  is obtained by applying Schmidt orthogonalization to the equivariant features and  $I$  denotes the identity matrix.

For the pose estimation, we employ  $L_1$  loss to guide SCR prediction correspondences to achieve localization.

The overall loss function of the entire network is:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{reg} + \lambda_2 \mathcal{L}_{pcc}, \quad (8)$$

where  $\lambda_1$  and  $\lambda_2$  are hyper-parameters to balance the losses.

### 3.2. Dataset

At first glance, SCR yields satisfactory results. However, accuracy degrades markedly with changes in vehicle orientation (see Fig. 2 (a)). In real-world settings, avoiding such

Dataset	LiDAR	GT	Reverse	Unobstructed view	Long Interval
KITTI [21]	1×64	GPS/IMU	✗	✓	✗
Oxford [4]	2×32	GPS/IMU	✗	✗	✗
NCLT [11]	1×32	SLAM	✗	✓	✓
Boreas [10]	1×128	GPS/IMU	✗	✓	✓
MulRan [26]	1×64	SLAM	✓	✗	✗
<b>BiLiLo (Ours)</b>	1×64	GPS/IMU	✓	✓	✓

Table 1. **Related Datasets.** GT: Ground truth poses. Reverse: Presence of trajectories in the opposite direction. Unobstructed View: Minimal obstructions in the LiDAR’s field of view. Long Interval: Extended time intervals between data collections.

directional shifts during lane changes and turns proves difficult (Fig. 2 (b)). Due to the lack of localization datasets with significant directional variation, previous work [18] simulates these situations by rotating point clouds. While these simulations can somewhat validate the model’s robustness to rotation, there are still discrepancies with the real scanned data, such as missing data caused by changes in scanning direction, etc. This motivates us to create a dataset with significant rotation changes in real-world scenarios to better evaluate existing methods.

**Sensors setup.** Our dataset mainly consists of two sensors, as shown in Fig. 5, including (1) Ouster OS1-64, a 64-beam spinning LiDAR to collect point cloud data, and (2) M39, a GPS/IMU device, to obtain the ego-motion pose. We set the FOV, resolution, and FPS as  $[-22.5^\circ, 22.5^\circ]$ ,  $64 \times 1024$  and 10, respectively. The FPS of M39 is set to 200, to provide dense and smoothing trajectories. We design a rigid support system for securing sensors to ensure reliably.

**Sensors calibration.** Following KITTI [21]. We use ICP [5] to accurately register point clouds of a parking sequence, which provides a wide variety of orientations and translations essential for effectively conditioning the minimization problem. Next, we randomly sample 1,000 pairs of poses from this sequence and obtain the desired result using [17] to achieve LiDAR-to-GPS/IMU calibration.

**Target environments.** Our dataset is collected in a campus environment featuring multiple distinguishable structures and a broad field of view. To assess the robustness of localization methods against translation and rotation, we deliberately incorporated multiple reverse revisits and revisits with lane-level translations.

Furthermore, to enhance data diversity, the collection period spanned both high-traffic intervals between classes and low-traffic periods during class sessions on campus. Note that we include two reverse trajectories, which pose challenges for localization due to missing data resulting from changes in scanning direction.

The dataset comprises 8 sequences for training and testing, each approximately 10 kilometers long, including 6 forward trajectories and 2 reverse trajectories. More details can be found in the supplementary.

**Comparison with other datasets.** Tab. 1 summarizes the

Methods	SO(3) Rotation Augmentation			Without Augmentation			
	DiffLoc	SGLoc	LiSA	DiffLoc	SGLoc	LiSA	RALoc
15-13-06-37	4.87m, <u>1.62</u> <sup>°</sup>	3.79m, 5.58 <sup>°</sup>	<u>3.78m</u> , 5.83 <sup>°</sup>	4.09m, <b>1.58</b> <sup>°</sup>	45.82m, 27.77 <sup>°</sup>	52.81m, 21.47 <sup>°</sup>	<b>3.17m</b> , 4.08 <sup>°</sup>
17-13-26-39	7.35m, <u>2.01</u> <sup>°</sup>	4.44m, 5.00 <sup>°</sup>	4.48m, 4.92 <sup>°</sup>	<u>4.17m</u> , <b>1.30</b> <sup>°</sup>	42.59m, 25.14 <sup>°</sup>	56.02m, 22.89 <sup>°</sup>	<b>3.87m</b> , 3.96 <sup>°</sup>
17-14-03-00	6.52m, <u>2.14</u> <sup>°</sup>	<u>3.66m</u> , 4.98 <sup>°</sup>	3.93m, 5.22 <sup>°</sup>	4.61m, <b>1.26</b> <sup>°</sup>	41.05m, 25.32 <sup>°</sup>	50.65m, 22.54 <sup>°</sup>	<b>3.32m</b> , 3.83 <sup>°</sup>
18-14-14-42	3.83m, <u>1.43</u> <sup>°</sup>	<u>2.76m</u> , 4.38 <sup>°</sup>	2.90m, 4.25 <sup>°</sup>	3.22m, <b>1.23</b> <sup>°</sup>	38.51m, 24.75 <sup>°</sup>	51.25m, 21.94 <sup>°</sup>	<b>2.58m</b> , 3.70 <sup>°</sup>
Average	5.64m, <u>1.80</u> <sup>°</sup>	<u>3.66m</u> , 4.99 <sup>°</sup>	3.77m, 5.06 <sup>°</sup>	4.02m, <b>1.34</b> <sup>°</sup>	41.99m, 25.75 <sup>°</sup>	52.68m, 22.21 <sup>°</sup>	<b>3.24m</b> , 3.89 <sup>°</sup>

Table 2. **Quantitative results on the Oxford dataset with small rotation.** Mean position error (m) and mean orientation error (<sup>°</sup>) are reported, with best results in **bold** and second best results underlined. Here, **small rotation** refers to a random rotation of  $[-10^{\circ}, 10^{\circ}]$  degrees applied to Roll, Pitch, and Yaw axes. SO(3) rotation augmentation involves random rotations of training data within  $[-180^{\circ}, 180^{\circ}]$ .

Methods	SO(3) Rotation Augmentation			Without Augmentation			
	DiffLoc	SGLoc	LiSA	DiffLoc	SGLoc	LiSA	RALoc
15-13-06-37	4.90m, <b>1.61</b> <sup>°</sup>	4.18m, 5.66 <sup>°</sup>	<u>3.66m</u> , 5.43 <sup>°</sup>	122.70m, 43.57 <sup>°</sup>	341.79m, 94.04 <sup>°</sup>	332.43m, 79.40 <sup>°</sup>	<b>3.19m</b> , <u>4.07</u> <sup>°</sup>
17-13-26-39	8.03m, <b>2.09</b> <sup>°</sup>	<u>4.55m</u> , 5.05 <sup>°</sup>	4.64m, 5.04 <sup>°</sup>	153.37m, 43.03 <sup>°</sup>	341.70m, 89.57 <sup>°</sup>	324.07m, 76.47 <sup>°</sup>	<b>3.87m</b> , <u>3.96</u> <sup>°</sup>
17-14-03-00	7.05m, <b>2.26</b> <sup>°</sup>	<u>3.70m</u> , 5.11 <sup>°</sup>	4.00m, 5.37 <sup>°</sup>	138.43m, 43.16 <sup>°</sup>	334.34m, 88.43 <sup>°</sup>	323.45m, 76.94 <sup>°</sup>	<b>3.33m</b> , <u>3.88</u> <sup>°</sup>
18-14-14-42	4.11m, <b>1.52</b> <sup>°</sup>	<u>2.82m</u> , 4.45 <sup>°</sup>	2.94m, 4.38 <sup>°</sup>	129.65m, 41.08 <sup>°</sup>	343.08m, 91.18 <sup>°</sup>	336.76m, 78.59 <sup>°</sup>	<b>2.58m</b> , <u>3.73</u> <sup>°</sup>
Average	6.02m, <b>1.87</b> <sup>°</sup>	<u>3.81m</u> , 5.07 <sup>°</sup>	<u>3.81m</u> , 5.06 <sup>°</sup>	136.04m, 42.71 <sup>°</sup>	340.23m, 90.81 <sup>°</sup>	329.18m, 77.85 <sup>°</sup>	<b>3.24m</b> , <u>3.91</u> <sup>°</sup>

Table 3. **Quantitative results on the Oxford dataset with large rotation.** Mean position error (m) and mean orientation error (<sup>°</sup>) are reported, with best results in **bold** and second best results underlined. Here, **large rotation** refers to a random rotation of  $[-180^{\circ}, 180^{\circ}]$  degrees applied to the Yaw axis. SO(3) rotation augmentation involves random rotations of training data within  $[-180^{\circ}, 180^{\circ}]$ .

comparison of our dataset with the existing widely used datasets. Our objective is to verify the method’s robustness to rotation by collecting datasets with rotation changes. Compared to Oxford Radar Robotcar [4] and NCLT [11], which are widely used for localization evaluation, we provide multiple reverse trajectories. MulRan [26] offers reverse direction data but suffers from Radar occlusion that restricts visibility and challenges localization consistency. In contrast, our dataset emphasizes the following features: (1) sensors are vertically mounted to reduce inter-sensor occlusions, ensuring near-complete 360-degree scene coverage. (2) data collection spans three months, offering diverse conditions for a robust evaluation of localization methods. See the supplementary for detailed comparisons.

## 4. Experiment

### 4.1. Setup

**Datasets and metrics.** We evaluate RALoc for LiDAR localization on two large-scale outdoor datasets: Oxford Radar RobotCar [4] and BiLiLo datasets. During the evaluation, we select the mean position and orientation error as the evaluation metrics.

**Oxford Radar RobotCar.** The Oxford Radar RobotCar (Oxford) dataset is a critical benchmark for urban scene localization. It covers a 2 km<sup>2</sup> area, with each trajectory spanning nearly 10 km. The dataset collects data from multiple sensors, such as LiDAR, cameras, radar, and GPS/INS. Our method only utilizes the information from LiDAR. The Oxford dataset captures a wide range of traffic conditions,

making it an ideal resource for evaluating LiDAR-based localization models. Following previous works [30, 31, 55], we use data from 11-14-02-26, 14-12-05-52, 14-14-48-55, and 18-15-20-12 for training, and data from 15-13-06-37, 17-13-26-39, 17-14-03-00, and 18-14-14-42 for testing.

**BiLiLo.** Our collected BiLiLo dataset is an outdoor campus scene localization benchmark, where each trajectory is nearly 10 km in length. It contains 8 trajectories obtained through LiDAR, with 2 trajectories specifically designed to have driving directions entirely opposite to the others. This makes the BiLiLo dataset a complementary resource to other localization datasets, enabling a more comprehensive evaluation of model performance across various aspects. We use the data from 11-28-11-20, 11-28-14-54, 11-28-15-26, and 12-06-13-57 for training, and data from 12-06-10-30, 12-06-10-59, 03-04-16-04, and 03-04-16-29 for testing, where the last two trajectories in each set are reverse data. Note the naming convention for the trajectories is based on the data collection start time, formatted as **month-date-hour-minute**.

**Implementation.** RALoc is implemented by Pytorch [41] and Spconv [15]. We conduct our experiments on a server equipped with an Intel Xeon(R) Gold 6342 CPU, and 4 NVIDIA RTX 4090 GPUs. For the point cloud canonicalization process, we randomly down sample 4,096 point clouds from the scene as input, and the subsequent LiDAR localization network is based on SGLoc. During training, we use the Adam [28] optimizer with a learning rate of 0.002 and a batch size of 120. On Oxford Radar RobotCar dataset, we set EKFA with  $k = 2048$ , and set the loss

Methods	SO(3) Rotation Augmentation			Without Augmentation			
	DiffLoc	SGLoc	LiSA	DiffLoc	SGLoc	LiSA	<b>RALoc</b>
15-13-06-37	4.88m, 1.62°	4.42m, 6.17°	3.77m, 5.81°	3.57m, <b>0.88°</b>	<u>3.01m</u> , 1.91°	<b>2.36m</b> , <u>1.29°</u>	3.19m, 4.10°
17-13-26-39	7.89m, 2.01°	4.62m, 5.60°	4.58m, 5.47°	<u>3.65m</u> , <b>0.68°</b>	4.07m, 2.07°	<b>3.47m</b> , <u>1.43°</u>	3.87m, 3.96°
17-14-03-00	7.16m, 2.26°	3.84m, 5.63°	4.08m, 5.67°	4.03m, <b>0.70°</b>	3.37m, 1.89°	<b>3.19m</b> , <u>1.34°</u>	<u>3.32m</u> , 3.87°
18-14-14-42	4.22m, 1.40°	2.89m, 4.83°	3.05m, 4.67°	2.86m, <b>0.60°</b>	<u>2.12m</u> , 1.66°	<b>1.95m</b> , <u>1.23°</u>	2.59m, 3.71°
Average	6.04m, 1.82°	3.94m, 5.56°	3.87m, 5.41°	3.53m, <b>0.72°</b>	<u>3.14m</u> , 1.88°	<b>2.74m</b> , <u>1.32°</u>	3.24m, 3.91°

Table 4. **Quantitative results on the Oxford dataset without rotation.** Mean position error (m) and mean orientation error (°) for various methods are reported, with best results in **bold** and second best results underlined. SO(3) rotation augmentation involves random rotations of training data along Roll, Pitch, and Yaw axes within  $[-180^\circ, 180^\circ]$ .

Methods	SO(3) Rotation Augmentation			Without Augmentation			
	DiffLoc	SGLoc	LiSA	DiffLoc	SGLoc	LiSA	<b>RALoc</b>
12-06-10-30	6.93m, 10.26°	3.35m, 2.93°	2.87m, 1.96°	2.78m, <b>0.77°</b>	2.14m, 1.81°	<u>1.92m</u> , <u>1.02°</u>	<b>1.91m</b> , 2.90°
12-06-10-59	6.69m, 9.66°	3.29m, 2.88°	2.73m, 1.80°	2.62m, <b>0.73°</b>	2.13m, 1.69°	<b>1.85m</b> , 0.96°	1.89m, 2.75°
03-04-16-04	27.99m, 23.75°	<u>10.12m</u> , <u>10.47°</u>	24.93m, 62.85°	383.85m, 72.70°	379.96m, 90.73°	322.93m, 94.56°	<b>5.24m</b> , <b>9.59°</b>
03-04-16-29	45.61m, 29.46°	<u>11.08m</u> , <u>13.60°</u>	27.66m, 66.42°	395.29m, 78.82°	378.96m, 92.43°	322.32m, 95.81°	<b>6.97m</b> , <b>11.02°</b>
Average	21.81m, 18.28°	<u>6.96m</u> , <u>7.47°</u>	14.55m, 33.26°	196.14m, 38.26°	190.80m, 46.67°	162.26m, 48.09°	<b>4.00m</b> , <b>6.57°</b>

Table 5. **Quantitative results on the BiLiLo dataset.** Mean position error (m) and mean orientation error (°) for various methods are reported, with best results in **bold** and second best results underlined. SO(3) rotation augmentation involves random rotations of training data along Yaw, Pitch, and Roll axes within  $[-180^\circ, 180^\circ]$ . The **red**-marked trajectories are aligned with the overall driving direction during training, while the **blue** ones are in the reverse direction.

weights to  $\lambda_1 = 1$ ,  $\lambda_2 = 1200$ . On BiLiLo dataset, we set  $k = 512$ ,  $\lambda_1 = 1$  and  $\lambda_2 = 1600$ .

**Baselines.** To more effectively evaluate our model’s performance, we select several representative and competitive baselines. For the APR method, we include the current state-of-the-art DiffLoc [31] for comprehensive comparison. For the SCR method, we choose SGLoc [30] and LiSA [55], where LiSA incorporates semantics into SGLoc through distillation while preserving a consistent framework. Note LiSA is the current state-of-the-art regression-based LiDAR localization method.

## 4.2. Results

**Results on Oxford with Rotation.** In this experiment, we introduce random rotations within 10 degrees in Roll, Pitch, and Yaw axes to the point cloud. Tab. 2 presents quantitative results on the Oxford dataset with small rotations (within  $[-10^\circ, 10^\circ]$ ), comparing baseline performance without augmentation to results with random rotation augmentation applied during training. RALoc achieves an average error of 3.24m/3.89°. It is clear that the results of the previous method experienced a significant decline without the introduction of rotation augmentation. Even after introducing data augmentation, our method still outperforms our baseline, SGLoc, with 3.24m/3.89° vs 3.66m/4.99°.

Additionally, to account for the potential need for frequent direction changes during vehicle operation, we apply random rotations around the Yaw axis (within  $[-180^\circ, 180^\circ]$ ) to the point clouds and adjust the ground-truth values accordingly. As shown in Tab. 3, RALoc achieves

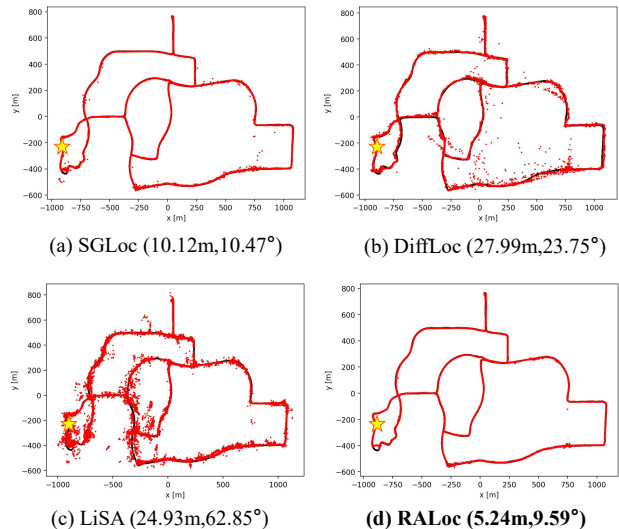


Figure 6. **Visualization of different methods on BiLiLo.** We select a reverse trajectory for testing. The black and red points represent the ground-truth and estimation poses, respectively, and the star indicates the first frame. The caption of each subfigure shows the mean position error (m) and orientation error (°).

3.24m/3.91°, reaching state-of-the-art performance in position, with no observable accuracy loss compared to previous evaluations. This result effectively demonstrates our model’s resilience to significant rotational changes.

**Results on Oxford.** We also evaluate RALoc on the Oxford dataset without rotation interference. Tab. 4 reports

the results of RALoc and all other baselines with and without random rotation augmentation during training. RALoc achieves 3.24m/3.91° average error. Although the proposed method shows a decrease in orientation accuracy compared to our baseline, SGLoc, it becomes more robust across various rotation scenarios. SGLoc has significantly lower accuracy than us after adding rotation augmentation. Meanwhile, our method is not significantly affected in terms of position accuracy and even outperforms our baseline on certain trajectories.

**Results on BiLiLo.** As mentioned earlier, simulated experiments struggle to comprehensively evaluate model performance, so we collected a real dataset with significant rotational variations for validation. Tab. 5 presents the evaluation results on the BiLiLo dataset. In particular, trajectories 03-04-16-04 and 03-04-16-29 have driving directions entirely opposite to the training data, enabling a rigorous evaluation of each model’s rotational robustness. Our method achieves 4.00m/6.57° average error. Compared to previous methods, RALoc achieves a substantial improvement in position accuracy on the forward trajectories, reaching state-of-the-art performance. On the reverse trajectories, our model produces superior results, significantly outperforming other baselines.

Fig. 6 illustrates the qualitative results of compartments with rotation augmentation on trajectory 03-04-16-04 with mean position error (m) and orientation error (°). It is clear that RALoc offers a clean trajectory, indicating its capacity to produce more robust results.

**Speed and parameter.** Real-time performance is a critical metric for localization tasks. Given that commonly used LiDAR sensors operate at a frequency of 20Hz, our inference time must remain below 50ms. We quantitatively evaluate the inference speed and parameter count of our proposed model, RALoc. The average inference time on the two datasets is 26ms. Additionally, in terms of total parameter count, we employ a relatively lightweight approach to achieve equivariance in the localization model. This requires only an additional 0.4M parameters compared to SGLoc, representing an increase of only **0.0043%**.

### 4.3. Ablation study

**Study on Rotation Augmentation.** As shown in Tab. 6 between Row 1 and Row 2, using data augmentation can, to some extent, mitigate catastrophic performance losses caused by rotation. However, as shown in Tab. 5, for the normal trajectories, it also brings a negative impact. By contrast, our method (Row 4, Tab. 6), outperforms it, achieving 4.00m/6.57° vs. 6.96m/7.47°, while avoiding significant localization accuracy losses on normal trajectories. These results highlight the superior robustness of our approach.

**Study on PCC.** We further conduct ablation experiments to demonstrate the importance of PCC. The comparison be-

	Augmentation	PCC	EKFA	Average
1				190.80m, 46.67°
2	✓			6.96m, 7.47°
3		✓		5.79m, 8.02°
4		✓	✓	4.00m, 6.57°

Table 6. **Ablation study on the BiLiLo dataset.** **Augmentation:** using data augmentation to increase the rotational diversity of training data. **PCC:** using Point Cloud Canonicalization process to perceptive rotation and align point cloud into canonicalization coordinate. **EKFA:** using Equivariant Key Feature Aggregation module to guide feature aggregation.

tween Row 1 and Row 3 show PCC can solve the challenges brought by rotation changes. Moreover, from Row 2 and Row 3, when compared to data augmentation, PCC also takes an average improvement of 16.8% on position accuracy. This shows that PCC can get a better result than rotation augmentation.

**Study on EKFA.** We evaluate the EKFA module’s impact on handling noise in real-world LiDAR point clouds. The EKFA module is designed to mitigate scene noise, aiming to enhance the equivariant network’s robustness to rotation and reduce noise-induced localization errors. Compared to Row 3 and Row 4, it yields a process of 30.9%/18.1% on position and orientation accuracy, respectively. Compared to the data augmentation, it also get an improvement of 42.5%/12.0%. Moreover, as shown in Tab. 5, integrating EKFA into PCC improves position accuracy by an average of 11.0% on normal trajectories compared to the vanilla method (SGLoc in this paper). These results show that the EKFA module effectively extracts scene-relevant information, reducing noise impact on the equivariant network and enhancing localization robustness.

## 5. Conclusion

In this paper, we reveal that current learning-based regression methods are sensitive to rotations, resulting in significant accuracy loss under rotational conditions. We propose RALoc, a novel rotation-robust scene coordinate regression framework for LiDAR-based localization. To more realistically evaluate the rotational robustness of the localization method, we establish equipment setups and collect a large-scale outdoor localization dataset, BiLiLo, with significant rotational variations. Extensive experiments demonstrate that the proposed RALoc significantly outperforms the state-of-the-art methods in scenarios with large rotations. In the future, we aim to enhance the training efficiency of RALoc and integrate multimodal information to further improve localization accuracy.

**Acknowledgements.** This work is supported in part by the Fundamental Research Funds for the Central Universities (No. 20720230033) and Xiaomi Young Talents Program.

## References

- [1] Sheng Ao, Qingyong Hu, Bo Yang, Andrew Markham, and Yulan Guo. Spinnet: Learning a general surface descriptor for 3d point cloud registration. In *CVPR*, pages 11753–11762, 2021. 2
- [2] Sheng Ao, Qingyong Hu, Hanyun Wang, Kai Xu, and Yulan Guo. Buffer: Balancing accuracy, efficiency, and generalizability in point cloud registration. In *CVPR*, pages 1255–1264, 2023. 2
- [3] Serge Assaad, Carlton Downey, Rami Al-Rfou, Nigamaa Nayakanti, and Ben Sapp. Vn-transformer: Rotation-equivariant attention for vector neurons. *arXiv preprint arXiv:2206.04176*, 2022. 3, 4
- [4] 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. IEEE, 2020. 2, 5, 6
- [5] Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, pages 586–606. Spie, 1992. 5
- [6] Georg Bökman, Johan Edstedt, Michael Felsberg, and Fredrik Kahl. Steerers: A framework for rotation equivariant keypoint descriptors. In *CVPR*, pages 4885–4895, 2024. 2, 3
- [7] Eric Brachmann and Carsten Rother. Visual camera re-localization from rgb and rgb-d images using dsac. *IEEE TPAMI*, 44(9):5847–5865, 2021. 2
- [8] Eric Brachmann, Alexander Krull, Sebastian Nowozin, Jamie Shotton, Frank Michel, Stefan Gumhold, and Carsten Rother. Dsac-differentiable ransac for camera localization. In *CVPR*, pages 6684–6692, 2017.
- [9] Eric Brachmann, Tommaso Cavallari, and Victor Adrian Prisacariu. Accelerated coordinate encoding: Learning to relocalize in minutes using rgb and poses. In *CVPR*, pages 5044–5053, 2023. 2
- [10] Keenan Burnett, David J Yoon, Yuchen Wu, Andrew Z Li, Haowei Zhang, Shichen Lu, Jingxing Qian, Wei-Kang Tseng, Andrew Lambert, Keith YK Leung, et al. Boreas: A multi-season autonomous driving dataset. *The International Journal of Robotics Research*, 42(1-2):33–42, 2023. 5
- [11] Nicholas Carlevaris-Bianco, Arash K Ushani, and Ryan M Eustice. University of michigan north campus long-term vision and lidar dataset. *The International Journal of Robotics Research*, 35(9):1023–1035, 2016. 2, 5, 6
- [12] Daniele Cattaneo, Matteo Vaghi, and Abhinav Valada. Lcd-net: Deep loop closure detection and point cloud registration for lidar slam. *IEEE T-RO*, 38(4):2074–2093, 2022. 1
- [13] Haiwei Chen, Shichen Liu, Weikai Chen, Hao Li, and Randall Hill. Equivariant point network for 3d point cloud analysis. In *CVPR*, pages 14514–14523, 2021. 3
- [14] Yunlu Chen, Basura Fernando, Hakan Bilen, Matthias Nießner, and Efstratios Gavves. 3d equivariant graph implicit functions. In *ECCV*, pages 485–502. Springer, 2022. 3
- [15] Spconv Contributors. Spconv: Spatially sparse convolution library. <https://github.com/traveller59/spconv>, 2022. 6
- [16] Congyue Deng, Or Litany, Yueqi Duan, Adrien Poulencard, Andrea Tagliasacchi, and Leonidas J Guibas. Vector neurons: A general framework for so (3)-equivariant networks. In *ICCV*, pages 12200–12209, 2021. 3, 4
- [17] Fadi Dornaika and Radu Horaud. Simultaneous robot-world and hand-eye calibration. *IEEE transactions on Robotics and Automation*, 14(4):617–622, 1998. 5
- [18] Zhaoxin Fan, Zhenbo Song, Wenping Zhang, Hongyan Liu, Jun He, and Xiaoyong Du. Rpr-net: A point cloud-based rotation-aware large scale place recognition network. In *ECCV*, pages 709–725. Springer, 2022. 5
- [19] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 15:381–395, 1981. 2
- [20] Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. Se (3)-transformers: 3d roto-translation equivariant attention networks. *NeurIPS*, 33:1970–1981, 2020. 3
- [21] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *CVPR*, pages 3354–3361. IEEE, 2012. 5
- [22] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 33:6840–6851, 2020. 3
- [23] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3d point clouds with low overlap. In *CVPR*, pages 4267–4276, 2021. 2
- [24] Oren Katzir, Dani Lischinski, and Daniel Cohen-Or. Shapepose disentanglement using se (3)-equivariant vector neurons. In *ECCV*, pages 468–484. Springer, 2022. 3, 5
- [25] Alex Kendall, Matthew Grimes, and Roberto Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *ICCV*, pages 2938–2946, 2015. 2, 3
- [26] Giseop Kim, Yeong Sang Park, Younghun Cho, Jinyong Jeong, and Ayoung Kim. Mulran: Multimodal range dataset for urban place recognition. In *ICRA*, pages 6246–6253. IEEE, 2020. 5, 6
- [27] Giseop Kim, Sunwook Choi, and Ayoung Kim. Scan context++: Structural place recognition robust to rotation and lateral variations in urban environments. *IEEE T-RO*, 38(3): 1856–1874, 2021. 1
- [28] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6
- [29] Jacek Komorowski. Minkloc3d: Point cloud based large-scale place recognition. In *WACV*, pages 1790–1799, 2021. 2
- [30] Wen Li, Shangshu Yu, Cheng Wang, Guosheng Hu, Siqi Shen, and Chenglu Wen. Sgloc: Scene geometry encoding for outdoor lidar localization. In *CVPR*, pages 9286–9295, 2023. 2, 3, 4, 5, 6, 7
- [31] Wen Li, Yuyang Yang, Shangshu Yu, Guosheng Hu, Chenglu Wen, Ming Cheng, and Cheng Wang. Diffloc: Diffusion model for outdoor lidar localization. In *CVPR*, pages 15045–15054, 2024. 2, 3, 6, 7
- [32] Chien Erh Lin, Jingwei Song, Ray Zhang, Minghan Zhu, and Maani Ghaffari. Se (3)-equivariant point cloud-based place recognition. pages 1520–1530. PMLR, 2023. 3

- [33] Quan Liu, Hongzi Zhu, Zhenxi Wang, Yunsong Zhou, Shan Chang, and Minyi Guo. Extend your own correspondences: Unsupervised distant point cloud registration by progressive distance extension. In *CVPR*, pages 20816–20826, 2024. 2
- [34] Weixin Lu, Yao Zhou, Guowei Wan, Shenhua Hou, and Shiyu Song. L3-net: Towards learning based lidar localization for autonomous driving. In *CVPR*, pages 6389–6398, 2019. 1
- [35] Lun Luo, Shuhang Zheng, Yixuan Li, Yongzhi Fan, Beinan Yu, Si-Yuan Cao, Junwei Li, and Hui-Liang Shen. Bevplace: Learning lidar-based place recognition using bird’s eye view images. In *ICCV*, pages 8700–8709, 2023. 1
- [36] Shitong Luo, Jiahao Li, Jiaqi Guan, Yufeng Su, Chaoran Cheng, Jian Peng, and Jianzhu Ma. Equivariant point cloud analysis via learning orientations for message passing. In *CVPR*, pages 18932–18941, 2022. 3
- [37] Junyi Ma, Jun Zhang, Jintao Xu, Rui Ai, Weihao Gu, and Xieyuanli Chen. Overlaptransformer: An efficient and yaw-angle-invariant transformer network for lidar-based place recognition. 7(3):6958–6965, 2022. 2
- [38] Arnab Kumar Mondal, Siba Smarak Panigrahi, Oumar Kaba, Sai Rajeswar Mudumba, and Siamak Ravanbakhsh. Equivariant adaptation of large pretrained models. *NeurIPS*, 36: 50293–50309, 2023. 2, 3, 5
- [39] Mohamed Adel Musallam, Vincent Gaudilliere, Miguel Ortiz Del Castillo, Kassem Al Ismaeil, and Djamila Aouada. Leveraging equivariant features for absolute pose regression. In *CVPR*, pages 6876–6886, 2022. 2
- [40] Yue Pan, Xingguang Zhong, Louis Wiesmann, Thorbjörn Posewsky, Jens Behley, and Cyrill Stachniss. Pin-slam: Lidar slam using a point-based implicit neural representation for achieving global map consistency. *IEEE T-RO*, 2024. 1
- [41] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *NeurIPS*, 32, 2019. 6
- [42] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In *CVPR*, pages 11143–11152, 2022. 2
- [43] Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *ICML*, pages 9323–9332. PMLR, 2021. 3
- [44] Torsten Sattler, Qunjie Zhou, Marc Pollefeys, and Laura Leal-Taixe. Understanding the limitations of cnn-based absolute camera pose regression. In *CVPR*, pages 3302–3312, 2019. 3
- [45] Chenghao Shi, Xieyuanli Chen, Junhao Xiao, Bin Dai, and Huimin Lu. Fast and accurate deep loop closing and relocalization for reliable lidar slam. *IEEE T-RO*, 2024. 1
- [46] Linus Svärm, Olof Enqvist, Fredrik Kahl, and Magnus Oskarsson. City-scale localization for cameras with known vertical direction. *IEEE TPAMI*, 39(7):1455–1461, 2016. 2
- [47] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018. 2
- [48] Mikaela Angelina Uy and Gim Hee Lee. Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition. In *CVPR*, pages 4470–4479, 2018. 2
- [49] Sijie Wang, Qiyu Kang, Rui She, Wei Wang, Kai Zhao, Yang Song, and Wee Peng Tay. Hypliloc: Towards effective lidar pose regression with hyperbolic fusion. In *CVPR*, pages 5176–5185, 2023. 3
- [50] Wei Wang, Bing Wang, Peijun Zhao, Changhao Chen, Ronald Clark, Bo Yang, Andrew Markham, and Niki Trigoni. Pointloc: Deep pose regressor for lidar point cloud localization. *IEEE Sensors Journal*, 22(1):959–968, 2021. 2, 3
- [51] Maurice Weiler and Gabriele Cesa. General e (2)-equivariant steerable cnns. *NeurIPS*, 32, 2019. 3
- [52] Hai Wu, Chenglu Wen, Wei Li, Xin Li, Ruigang Yang, and Cheng Wang. Transformation-equivariant 3d object detection for autonomous driving. In *AAAI*, pages 2795–2802, 2023. 2, 3
- [53] Yan Xia, Mariia Gladkova, Rui Wang, Qianyun Li, Uwe Stilla, Joao F Henriques, and Daniel Cremers. Casspr: Cross attention single scan place recognition. In *ICCV*, pages 8461–8472, 2023. 2
- [54] Xuecheng Xu, Sha Lu, Jun Wu, Haojian Lu, Qiuguo Zhu, Yiyi Liao, Rong Xiong, and Yue Wang. Ring++: Rotation-invariant gram for global localization on a sparse scan map. *IEEE T-RO*, 2023. 1
- [55] Bochun Yang, Zijun Li, Wen Li, Zhipeng Cai, Chenglu Wen, Yu Zang, Matthias Muller, and Cheng Wang. Lisa: Lidar localization with semantic awareness. In *CVPR*, pages 15271–15280, 2024. 1, 2, 3, 4, 6, 7
- [56] Jingyun Yang, Congyue Deng, Jimmy Wu, Rika Antonova, Leonidas Guibas, and Jeannette Bohg. Equivact: Sim (3)-equivariant visuomotor policies beyond rigid object manipulation. In *ICRA*, pages 9249–9255. IEEE, 2024. 3
- [57] Huan Yin, Xuecheng Xu, Sha Lu, Xieyuanli Chen, Rong Xiong, Shaojie Shen, Cyrill Stachniss, and Yue Wang. A survey on global lidar localization: Challenges, advances and open problems. *IJCV*, 132(8):3139–3171, 2024. 1
- [58] Shangshu Yu, Xiaotian Sun, Wen Li, Chenglu Wen, Yunuo Yang, Bailu Si, Guosheng Hu, and Cheng Wang. Nidaloc: Neurobiologically inspired deep lidar localization. *IEEE TITS*, 2023. 3
- [59] Chongjian Yuan, Jiarong Lin, Zuhao Zou, Xiaoping Hong, and Fu Zhang. Std: Stable triangle descriptor for 3d place recognition. In *ICRA*, pages 1897–1903, 2023. 2
- [60] Liyuan Zhu, Shengyu Huang, Konrad Schindler, and Iro Armeni. Living scenes: Multi-object relocalization and reconstruction in changing 3d environments. In *CVPR*, pages 28014–28024, 2024. 3
- [61] Minghan Zhu, Maani Ghaffari, William A Clark, and Hui Peng. E2pn: Efficient se (3)-equivariant point network. In *CVPR*, pages 1223–1232, 2023. 3