IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

LighterBEV: LiDAR Global Localization Meets Online Learning

Binhong Liu[1], Tao Yang[1*], Haoji Cao[1], Shuqi Fu[3], Yangwang Fang[1], and Zhi Yan[2]

***Abstract*—LiDAR-based global localization provides accurate robot pose estimates against a prior map. Existing deep-learning methods, however, demand heavy computation and long training or inference times and degrade sharply when faced with domain shifts. This letter presents LighterBEV, a lightweight, fast, and generalizable localization method. An Informative Compression Module achieves a fourfold reduction in local-feature dimensionality while improving accuracy. We further integrate online learning to enable rapid post-deployment adaptation, mitigating degradation under distribution shift. Extensive experiments on four large-scale datasets show that LighterBEV achieves state-of-the-art performance with limited training data, maintains high accuracy under domain shift, and runs in real time on resource-constrained hardware—supporting both inference and online updates. To our knowledge, LighterBEV is the first LiDAR global localization approach to incorporate online learning for automatic adaptation to new environments, thereby narrowing the domain gap. Code will be released at: https://github.com/npu-ius-lab/LighterBEV.**

I. INTRODUCTION

Global localization estimates a robot's global pose from observations and a prior map, underpinning loop closure and relocalization in SLAM. Most systems follow two stages: coarse place recognition with global descriptors, then precise pose estimation through pairwise registration. Deep learning based LiDAR methods perform well on large-scale benchmarks, but models trained in one domain and evaluated in another degrade when the training and test distributions are not independent and identically distributed. A direct remedy is domain adaptation with additional data [1], which increases collection and training cost. Continual learning [2], [3] adapts sequentially across domains, yet it often responds too slowly to rapid environmental change, which limits timely deployment.

Online learning [4] enables a deployed model to continuously update its parameters from streaming data, allowing real-time adaptation to changing environments and data distributions. This paradigm has begun to appear in deep visual odometry. Self-supervised VO [5] introduces an online meta-learning scheme that continually adapts visual odometry networks without manual labels. Adaptive VIO [6] combines online learning with nonlinear optimization. With differentiable IMU preintegration, it allows a visual–inertial odometry system to adapt to new environments. Applying online learning to LiDAR global localization can quickly reduce the performance drop caused by domain shift. However, this approach also increases on-board computational load. A robot must generate and select training samples and update the model while operating under limited resources and sustaining real-time inference. Given these considerations, desirable attributes for LiDAR global localization with online learning include the following:

- The model should be generalizable across domains and robust to changes in viewpoint and LiDAR sensors.
- The model should be efficient for training and inference.
- The model should be able to learn from newly streaming data and quickly adapt to new domains.

Therefore, we propose LighterBEV, a lightweight, fast, and generalizable LiDAR global localization method that meets the computational requirements of online learning and real-time inference. It maintains an online buffer for rapid on-device adaptation under limited resources. Fig. 1 illustrates the pipeline. When integrated with a SLAM system, LighterBEV assists loop closure and uses certified SLAM-derived relative positions or together with GPS signals to update after deployment, further improving global localization, particularly when the training and test domains differ substantially. The main contributions of this paper are:

- We introduce **LighterBEV**, a lightweight BEV-based LiDAR global localization method that outperforms existing methods in accuracy while reducing computational cost.
- We present the first *online learning* paradigm for LiDAR global localization, enabling rapid post-deployment adaptation to new environments and mitigating domain-shift degradation. The source code will be released.
- We conduct extensive experiments demonstrating LighterBEV's high accuracy, generalization, inference speed, and low GPU usage; with online learning, robots can

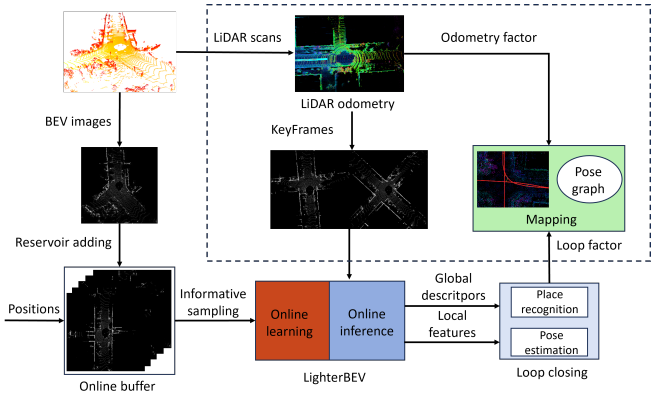

Fig. 1. Overview of the LighterBEV online learning pipeline integrated with a LiDAR-based SLAM system. The dashed box denotes the SLAM module, while the remaining components correspond to the online learning–based LighterBEV. LighterBEV online updates its parameters during the SLAM process, adapting to new environments over time.

Manuscript received: August 4, 2025; Revised: September 29, 2025; Accepted: November 25, 2025. This paper was recommended for publication by Editor Javier Civera upon evaluation of the Associate Editor and Reviewers' comments.

[1]Binhong Liu, Tao Yang, Haoji Cao, and Yangwang Fang are with Unmanned System Research Institute, National Key Laboratory of Unmanned Aerial Vehicle Technology, Integrated Research and Development Platform of Unmanned Aerial Vehicle Technology, Northwestern Polytechnical University, 710072 Xi'an, China. *Corresponding author: yangtao@nwpu.edu.cn

[2]Zhi Yan is with U2IS, ENSTA, Institut Polytechnique de Paris, France.

[3]Shuqi Fu is with the College of Engineering, Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

automatically adapt to dynamic environments.

II. Related Work

A. Learning-based LiDAR Global Localization

Existing learning-based LiDAR global localization methods can be broadly divided into three categories: point-based [7], voxel-based [8], and projection-based [9]. LCDNet [7] uses a point-based network to extract local features, aggregates them into a compact global descriptor with NetVLAD [10], and recovers pose via RANSAC on matched feature correspondences. EgoNN [8] maps points to absolute cylindrical coordinates and uses sparse 3D voxel convolutions for keypoint regression and description. BEVPlace++ [9] converts point clouds into BEV images, applies a rotation-equivariant module for feature extraction, aggregates features with NetVLAD into an 8192-dimensional global descriptor, and upsamples feature maps to obtain local descriptors. Among these categories, projection-based BEV methods are generally regarded as lightweight and robust, with strong cross-domain generalization.

B. Feature Dimensionality Reduction

Reducing feature dimensionality improves computational efficiency and mitigates redundancy in deep models. Nonlinear methods, such as variational autoencoders [11], jointly learn an encoder–decoder under a variational objective to capture complex data distributions and produce smooth latent embeddings. However, the encoder–decoder pipeline and stochastic sampling incur substantial computation and memory overhead, and their nonlinear mappings can disrupt the rotation-equivariant structure of feature maps, which is critical for LiDAR-based localization. Linear methods preserve equivariance but may be hard to preserve important information. LCDNet [7] uses a trainable linear projection to compress high-dimensional global descriptors into a 256-dimensional latent space; this is efficient and easy to integrate but can degrade accuracy by losing salient features. Principal Component Analysis (PCA) identifies orthogonal directions of maximum variance and projects onto the leading components, yielding a compact, decorrelated representation that retains the most informative structure. PCA thus provides a principled trade-off between efficiency and information retention. However, effectively integrating unsupervised PCA with deep networks remains an open challenge.

C. Online Learning for Robotics

Existing offline-trained LiDAR localization models [7], [8], [12], [9] deliver strong performance and generalization. However, their accuracy still degrades when deployed in new environments, necessitating offline, domain-specific retraining. In the LiDAR 3D detection task, OCL3D [13] transfer and adapt 2D visual detectors to 3D data online without offline retraining. In visual odometry task, self-supervised VO [5] and Adaptive VIO [6] employ online meta-learning and differentiable IMU preintegration to adapt continuously under self-supervision. In this work, we propose an online learning framework for LiDAR global localization. It maintains an online buffer for rapid domain adaptation and enables efficient model updates under limited computational resources, improving robustness to distribution drift without offline retraining.

III. Problem Formalization and Preliminaries

A. BEV Image-based Global Localization Pipeline

1) Place Recognition

Given a LiDAR point cloud $P \in \mathbb{R}^{N\times 3}$ at position p, we form a BEV image $\mathbf{I}$ by orthographic projection. A backbone f_θ encodes $\mathbf{I}$ into a global descriptor $\mathbf{v} = f_\theta(\mathbf{I})$. For place recognition, descriptors should preserve the ordering induced by geographic proximity, that is:

$$D_g(P_i, P_j) < D_g(P_i, P_k) \;\Rightarrow\; \|\mathbf{v}_i - \mathbf{v}_j\| < \|\mathbf{v}_i - \mathbf{v}_k\|, \quad (1)$$

where $D_g(\cdot,\cdot)$ is the geographic distance between point clouds and $\|\cdot\|_2$ is the Euclidean norm in the embedding space.

2) Pose Estimation

After place recognition, we estimate the pose with a two-stage pipeline. First, FAST [14] keypoints are detected in the query and reference BEV images, and compact local descriptors are computed at each keypoint. Brute-force nearest-neighbor search provides initial matches. Second, RANSAC refines correspondences and robustly estimates the rigid transform $\mathbf{T}$ by maximizing inlier consensus, yielding reliable alignment between the images.

B. Rotation Equivariance and Invariance

Given a BEV image $\mathbf{I}$ and rotation operator $g(\cdot)$, the backbone f is rotation equivariant if

$$f(g(\mathbf{I})) = g(f(\mathbf{I})). \quad (2)$$

and is rotation invariant if

$$\mathbf{F} = f(\mathbf{I}) = f(g(\mathbf{I})). \quad (3)$$

Equivariance guarantees that rotating the input and then extracting features is equivalent to rotating the feature map, while invariance ensures that feature extraction is unchanged by input rotations.

IV. Proposed Method

In this section, we first present LighterBEV: a fast, generalizable, and computationally efficient LiDAR-based global localization method. Then we further integrate an online learning paradigm into LighterBEV, enabling rapid domain adaptation. The architecture of LighterBEV is detailed in Fig. 2.

A. Random Sector Masking

In real-world deployments, LiDAR point clouds are often occluded by foreground objects, which disrupts appearance continuity between adjacent scenes. To improve robustness to such occlusions, we apply random sector masking during preprocessing. For each BEV image, we take the image center as the origin, sample a random angle, and mask the corresponding circular sector, thereby simulating partial blockage by obstacles.

B. Lighter Rotation Equivariant Module

We propose the Lighter Rotation Equivariant Module (LighterREM), which consists of a Rotation Equivariant Module (REM [9]) and an Informative Compression Module (ICM).

1) Rotation Equivariant Module

We adopt the Rotation Equivariant Module (REM), which is first introduced in BEVPlace++ [9], to extract distinctive and rotation-equivariant features. Given an input BEV image $\mathbf{I}$, it is

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

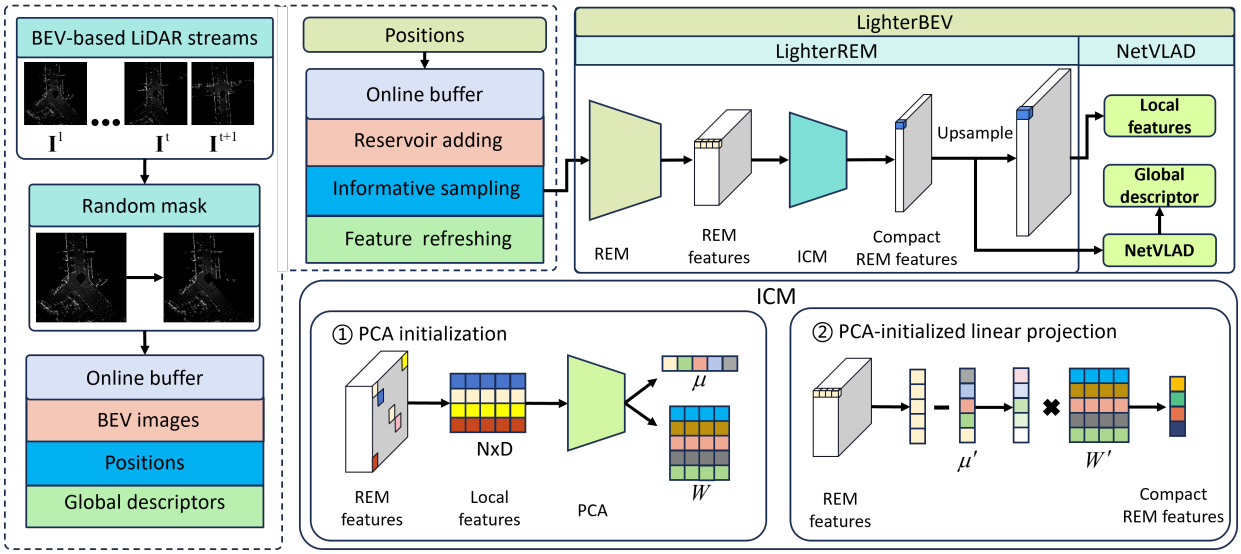

Fig. 2. Overall architecture of LighterBEV (modules within solid boxes) and its online learning mechanism (modules within dashed boxes). BEV-based LiDAR streams with position information are stored in the online buffer via reservoir adding. The buffer performs informative sampling and feature refreshing to select representative samples for online training. LighterBEV consists of the LighterREM encoder and the NetVLAD aggregator, which extract compact local features and generate global descriptors, respectively. The LighterREM encoder further comprises the REM [9] module and the proposed Information Compression Module (ICM) for efficient dimensionality reduction and feature refinement.

initially transformed using N_R distinct rotation angles selected from the set $\mathcal{R} = \{0, \frac{2\pi}{N_R}, \ldots, (N-1)\frac{2\pi}{N_R}\}$. For each rotated image, local features are extracted using residual convolutional modules with shared weights, after which the feature map is rotated back to its original orientation. The rotation-equivariant local features are then obtained by applying max pooling across the feature maps of the N rotated versions. For a more detailed description of the REM module, refer to [9].

2) Informative Compression Module

To improve computational efficiency, we compress REM features while preserving rotation equivariance and discriminative content. Before training, we sample a subset of BEV images and extract their REM feature maps. From these maps we randomly draw N local features $\{\mathbf{v}_i\}_{i=1}^{N}$, with $\mathbf{v}_i \in \mathbb{R}^D$, to form a feature set for learning the lower-dimensional projection. We apply PCA to the sampled feature set to obtain a D'-dimensional subspace, yielding a projection matrix $\mathbf{W} \in \mathbb{R}^{D\times D'}$ and a mean vector $\boldsymbol{\mu} \in \mathbb{R}^D$. During training, each feature map $\mathbf{F} \in \mathbb{R}^{D\times H\times W}$ is treated as HW vectors $\mathbf{v} \in \mathbb{R}^D$, which are compressed by

$$\mathbf{v}' = (\mathbf{v} - \boldsymbol{\mu}') \times \mathbf{W}', \tag{4}$$

where $\mathbf{W}'$ and $\boldsymbol{\mu}'$ are trainable parameters initialized with $\mathbf{W}$ and $\boldsymbol{\mu}$ and updated via backpropagation.

3) Rotation Equivariance Analysis

Let f represent the function that extracts the REM features $\mathbf{F}$, and f' represent the function that extracts the compact REM features $\mathbf{F}'$. The rotation transformation is denoted as $g(\cdot)$. The compact REM features can be expressed as:

$$\mathbf{F}' = f'(\mathbf{I}) = \big(f(\mathbf{I}) - \mathbf{V}\big)\,\mathbf{W}', \tag{5}$$

where $\mathbf{V} = \boldsymbol{\mu}' \otimes \mathbf{1}_{H\times W}$. Next, consider the effect of the rotation transformation on the compact REM features:

$$f'(g(\mathbf{I})) = f(g(\mathbf{I})) \cdot \mathbf{W}' - \mathbf{V} \cdot \mathbf{W}', \tag{6}$$
$$g(f'(\mathbf{I})) = g(f(\mathbf{I}) \cdot \mathbf{W}') - g(\mathbf{V} \cdot \mathbf{W}'). \tag{7}$$

Since the rotation equivariance of the REM model has been proven in [9], the following equation holds:

$$g(f(\mathbf{I})) = f(g(\mathbf{I})). \tag{8}$$

Moreover, $\mathbf{V}$ is constructed by broadcasting the mean vector $\boldsymbol{\mu}'$ to all spatial positions, it remains unchanged under rotations: $g(\mathbf{V}) = \mathbf{V}$, which leads to the conclusion:

$$f'(g(\mathbf{I})) = g(f'(\mathbf{I})). \tag{9}$$

Thus, we have shown that the compact REM features are rotation equivariant.

C. Online Learning for LighterBEV

To train LighterBEV online, we design an online buffer, which is a fixed-size memory space. In this memory space, we mainly store three types of data: BEV image representations of LiDAR scans, positions (x, y, z), and the global descriptors corresponding to each image. The online buffer contains three main operations: reservoir adding, informative sampling and feature refreshing. The full pipeline of the online learning process is detailed in Algorithm 1.

1) Reservoir Adding

The reservoir adding operation is designed to efficiently manage and store data in the online buffer. It ensures that the buffer maintains a representative sample of the incoming data stream. Specifically, as new data arrives, it is added to the buffer with a probability that depends on both the current size of the buffer and the total number of LiDAR scans observed so far. This approach helps prevent bias toward storing only the most recent data, instead providing a well-balanced and diverse set of samples for training.

2) Informative Sampling

To construct informative triplets for online learning, we prioritize negative samples that are similar in feature space. Using global descriptors stored in the online buffer, we select the m hardest negatives per query. For each query image,

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Algorithm 1 Online Learning with LighterBEV

Input: Initial backbone f_θ from the model trained on KITTI dataset.
Parameter: online buffer B with fixed-size k.
for each new BEV image x and position p **do**
  Calculate the global descriptor $\mathbf{v} = f(x)$;
  Reservoir Adding $(x, p, \mathbf{v})$ to B;
  Informative Sampling from B to form triplets (q, p, n);
  Calculate loss L using sampled triplets (q, p, n) based on Eq. 10;
  Calculate gradient $\frac{dL}{d\theta}$ then update model f with parameters θ by gradient descent;
  Feature Refreshing for online buffer;
end for

positives and negatives are first determined by relative positions. Then we compute feature distances to all negatives, choose the m with the smallest distances, and randomly select one positive to complete the triplet.

3) Feature Refreshing

During the online learning process, the parameters θ of the backbone f are updated after each backpropagation, which can result in outdated global descriptors stored in the online buffer. This leads to the informative sampling operation being unable to identify truly hard negative samples, thereby reducing learning efficiency. To address this, we refresh the features in the online buffer at a fixed frequency. The Feature Refreshing operation complements the informative sampling process by updating the outdated global descriptors of BEV images in the buffer, ensuring it consistently holds up-to-date feature representations.

D. Loss Function

We use the lazy triplet loss [15], [9] to train the LighterBEV. This loss function focuses on maximizing the feature distance between a query and its hardest negative samples in a triplet, that is:

$$\mathcal{L} = \max_j \big(\max(\delta + ||\mathbf{v}_q - \mathbf{v}_p||_2 - ||\mathbf{v}_q - \mathbf{v}_{nj}||_2, 0)\big), \quad (10)$$

where $\mathbf{v}_q$, $\mathbf{v}_p$ and $\mathbf{v}_{nj}$ are the global descriptors of the query sample, its positive sample, and the j-th negative sample, respectively. The tunable margin hyperparameter δ helps to enforce a minimum separation between the positive and negative pairs in the embedding space.

V. Experimental Settings

A. Datasets

To evaluate the performance of LighterBEV, we use four large-scale datasets, including KITTI, NCLT, Oxford RobotCar [16], and In-house [15] dataset, which were collected in different environments.

KITTI and Rotated KITTI. The KITTI odometry dataset is a widely used large-scale benchmark consisting of 11 sequences of point clouds collected with a Velodyne HDL-64E LiDAR, along with ground-truth poses. In our experiments we use sequences *00*, *02*, *05*, *06*, and *08*. We follow the frame-partitioning protocol of [9], listed in Table I. We create Rotated KITTI by randomly rotating each frame to test the robustness to viewpoint changes.

NCLT [17]. The NCLT dataset provides multi-season LiDAR data collected in campus environment using a Velodyne HDL-32E, along with ground-truth poses for long-term global localization benchmarking. We evaluate our method on seven seasonally distinct sessions: *2012-01-15*, *2012-02-04*, *2012-03-17*, *2012-06-15*, *2012-09-28*, *2012-11-16*, and *2013-02-23*.

TABLE I
Database and querybase partition of the KITTI dataset.

Sequence	00	02	05	06	08
Database	0-3000	0-3400	0-1000	0-600	0-3000
Querybase	3200-4650	3600-4661	1200-2751	800-1100	3200-4645

Oxford RobotCar [16] and In-house [15]. The Oxford RobotCar dataset offers year-long LiDAR traverses of Oxford using a SICK LD-MRS scanner, released as position-annotated submaps. The In-house dataset, collected with a Velodyne HDL-64E, includes five runs across three areas in Singapore—university sector (U.S.), residential area (R.A.), and business district (B.D.)—in the same submap format. Following [9], we convert both datasets into BEV images for evaluation.

B. Online Learning Protocol

To simulate real-world deployment, we design an online learning protocol in which a pretrained LighterBEV model updates from a non-i.i.d. LiDAR stream. Each scan can be fed into the system only once and stored in a small memory buffer. Because data arrive sequentially, future positions are unavailable at update time, precluding offline selection of positives and negatives by position. This constraint creates a realistic and demanding scenario that emphasizes efficient adaptation to distribution shifts under strict streaming conditions.

C. Evaluation Metrics

To assess LighterBEV across its subtasks, we adopt task-specific metrics. We establish a 5m threshold to determine whether a match is true positive (TP), false negative (FN), or false positives (FP). For place recognition, we report Recall which can be expressed as follows:

$$\text{Recall} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \quad (11)$$

where N_{TP} and N_{FN} denote the numbers of true positives and false negatives, respectively. For loop closing, we use the average precision (AP), F1 max score, max recall at 100% precision to evaluate the loop closure detection performance. The Average Precision (AP) metric quantifies the area under the Precision-Recall curve generated by varying the descriptor distance threshold. The Precision can be expressed as follows:

$$\text{Precision} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}. \quad (12)$$

The F1 score provides a balanced measure between precision and recall. We report the maximum F1 score on all descriptor distance thresholds to characterize the optimal operating point of our system. The F1 score can be expressed as follows:

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (13)$$

The max recall represents the highest attainable recall rate while maintaining zero false positives, which is formally defined as:

$$\text{max recall} = \max \left\{ \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \;\middle|\; N_{\text{FP}} = 0 \right\}. \quad (14)$$

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE II
RECALL@1 ON KITTI AND NCLT USING MODELS TRAINED ON KITTI; BASELINE RESULTS FOLLOW [9].

Method	KITTI						NCLT						
	00 ↑	02 ↑	05 ↑	06 ↑	08 ↑	Mean ↑	2012-02-04 ↑	2012-03-17 ↑	2012-06-15 ↑	2012-09-28 ↑	2012-11-16 ↑	2013-02-23 ↑	Mean ↑
M2DP [18]	92.9	69.3	80.7	94.8	34.4	74.4	63.2	58.0	42.4	40.6	49.3	27.9	46.9
BoW3D [19]	71.4	15.5	58.7	91.8	57.0	58.9	14.9	10.7	6.5	5.0	5.2	7.5	8.3
CVTNet [20]	98.7	87.1	93.5	97.8	83.7	92.1	89.2	88.0	81.2	74.9	77.1	80.3	81.8
Logg3D [21]	99.6	96.1	97.5	**100.0**	93.5	97.3	69.9	19.6	11.0	8.7	10.9	25.6	24.3
LCDNet [7]	99.9	97.7	95.3	**100.0**	94.4	97.4	60.5	54.2	44.2	34.9	31.7	10.9	39.4
BEVPlace [12]	99.7	98.1	99.3	**100.0**	92.0	97.8	93.5	92.7	87.4	87.8	88.9	86.2	89.4
BEVPlace++ [9]	**100.0**	99.3	**100.0**	**100.0**	99.1	99.7	95.3	94.2	90.2	88.9	91.3	87.8	91.3
LighterBEV (ours)	99.9	**100.0**	99.3	**100.0**	**100.0**	**99.8**	**96.6**	**95.3**	**93.0**	**90.7**	**92.6**	**91.2**	**93.2**

The mean translation errors ($\hat{e}_t$) and the mean rotation errors ($\hat{e}_r$) are calculated by comparing the estimated pose with their corresponding ground truth from the correctly found loop closure pairs. For multi-session global localization, we employ Recall@1 for retrieval evaluation and report both mean translation and rotation errors for registration accuracy. We additionally calculate the Success Rate (SR) defined as the percentage of queries localized under $(2m, 5^\circ)$ thresholds.

D. Implementation details

For each point cloud, we generate a BEV image with a size of 200×200. We train LighterBEV for 40 epochs, and the learning rate is init to $1e^{-4}$ and decays by a factor of 2 every 10 epochs. For the offline training, we select 10 negative samples for each query sample, with a positive sample threshold of 5m and a negative sample threshold of 7m. For the online protocol, we select 3 negative samples for each query sample, with a positive sample threshold of 5m and a negative sample threshold of 10m. The online buffer size k is set to 500, and Feature Refreshing is performed every 200 samples.

VI. EXPERIMENTS

In this section, we first evaluate LighterBEV under the offline fashion on all aforementioned datasets and subtasks. We then apply the online learning protocol that allows the model to adapt to unseen environments.

TABLE III
RECALL AT TOP-1 ON THE ROTATED KITTI DATASET.

Sequence	00 ↑	02 ↑	05 ↑	06 ↑	08 ↑	Mean ↑
M2DP [18]	92.9	69.3	80.7	94.8	34.4	74.4
Logg3D[21]	99.4	96.4	97.3	99.6	92.0	96.9
CVTNet [20]	98.7	87.4	93.3	98.5	85.8	92.7
BoW3D [19]	19.2	9.1	13.5	13.4	1.5	11.3
LCDNet [7]	99.7	98.1	95.5	**100.0**	94.7	97.6
BEVPlace [12]	99.6	93.5	98.9	**100.0**	92.0	96.8
BEVPlace++ [9]	99.7	97.1	98.9	**100.0**	97.3	98.6
LighterBEV (ours)	**99.9**	**100.0**	**99.1**	**100.0**	**100.0**	**99.8**

TABLE IV
AVERAGE RECALL@1 ON THE OXFORD AND IN-HOUSE DATASET.

	Oxford ↑	U.S. ↑	R.A. ↑	B.D. ↑	Mean ↑
PNV [15]	62.8	63.2	56.1	57.2	59.8
LPD-Net [22]	86.3	87.0	83.1	82.5	84.7
PPT-Net [23]	93.5	90.1	84.1	84.6	88.1
MinkLoc3Dv2 [24]	96.3	90.9	86.5	86.3	90.0
BEVPlace [12]	96.5	96.9	92.3	95.3	95.3
BEVPlace++ [9]	96.2	**97.1**	92.7	95.6	95.4
LighterBEV (ours)	**96.7**	96.7	**93.3**	**95.8**	**95.6**

A. Place Recognition Evaluation

KITTI and NCLT dataset. Table II summarizes Recall@1 on KITTI and NCLT. On the KITTI side (left), LighterBEV outperforms all baselines on every sequence and achieves 100% Recall@1 on sequences *02* and *08*, which contain numerous complex reverse loops. These results on the most challenging sequences underscore LighterBEV's robustness to rotational variations, while using global descriptors that are only one quarter the length of those in BEVPlace++. On the right-hand side of Table II, we report a cross-dataset evaluation in which a model trained on KITTI is tested directly on NCLT without any fine-tuning. The NCLT sequence *2012-01-15* is used as the reference database, and all remaining sequences serve as the query set. As shown, LighterBEV consistently achieves the highest Recall@1 on NCLT. Maintaining high accuracy across these distinct environments demonstrates LighterBEV's strong cross-dataset generalization.

Rotated KITTI dataset. To further evaluate the robustness to view changes, we use the model trained on the KITTI dataset and test on the Rotated KITTI dataset. Table III shows the comparisons of different methods. The results demonstrate that LighterBEV still outperforms baseline methods and shows the robustness to large viewpoint changes.

Oxford RobotCar and In-house dataset. These two datasets serve as standard benchmarks for recent sparse point-based place recognition methods. We train our LighterBEV on the Oxford RobotCar training set, and test it on its test set and In-house dataset. Table IV presents the results with recall@1 metric, with all results directly sourced from their respective original publications. The results demonstrate that LighterBEV not only achieves state-of-the-art performance on the Oxford dataset but also shows superior generalization capabilities on the In-house dataset in terms of overall performance.

B. Loop Closing Evaluation

We evaluate loop closing using models pre-trained on KITTI sequence *00*. The pipeline has two stages: (i) loop-closure detection via place recognition, and (ii) pose estimation. To avoid trivial matches, the nearest 100 database frames to each query are excluded. Table V reports Average Precision (AP), F1 max score, max recall at 100% precision, mean translation error ($\hat{e}_t$), and mean rotation error ($\hat{e}_r$). On KITTI, LCDNet, BEVPlace++, and LighterBEV deliver similarly strong detection performance and clearly surpass M2DP, Logg3D, CVTNet, and BoW3D. On the challenging KITTI *08* sequence, LCDNet lags behind BEVPlace++ and LighterBEV, reflecting the advantage of rotation-invariant global descriptors. Moreover,

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE V
AVERAGE PRECISION AND F1 MAX SCORE OF LOOP CLOSURE ON KITTI AND NCLT.

Sequence	KITTI 00					KITTI 02					KITTI 05					KITTI 06				
	AP	max F1	max Recall	$\hat{e}_t$ (m)	$\hat{e}_r$ (°)	AP	max F1	max Recall	$\hat{e}_t$ (m)	$\hat{e}_r$ (°)	AP	max F1	max Recall	$\hat{e}_t$ (m)	$\hat{e}_r$ (°)	AP	max F1	max Recall	$\hat{e}_t$ (m)	$\hat{e}_r$ (°)
M2DP [18]	0.982	0.936	86.7	-	-	0.884	0.844	0.0	-	-	0.946	0.897	68.1	-	-	0.974	0.938	76.2	-	-
Logg3D[21]	0.995	0.976	55.2	-	-	0.983	0.927	82.7	-	-	0.995	0.975	86.2	-	-	0.996	0.970	91.9	-	-
CVTNet [20]	0.994	0.965	84.8	-	-	0.931	0.898	64.6	-	-	0.975	0.933	96.2	-	-	0.996	0.981	96.2	-	-
BoW3D [19]	0.979	0.897	46.5	0.54	1.20	0.559	0.546	10.6	0.74	0.55	0.957	0.857	47.8	0.69	0.72	0.992	0.968	48.1	0.62	0.73
LCDNet [7]	0.997	0.974	94.1	0.10	0.14	0.976	0.928	**83.7**	0.65	**0.44**	0.994	0.964	93.0	0.12	0.17	0.999	0.997	99.6	0.11	0.17
BEVPlace++ [9]	**0.999**	**0.995**	**98.4**	0.08	**0.11**	**0.977**	**0.934**	70.0	**0.38**	0.70	0.994	**0.982**	**96.2**	0.12	**0.09**	0.999	0.999	**100.0**	0.18	**0.08**
LighterBEV(ours)	**0.999**	**0.995**	**98.4**	**0.07**	**0.11**	0.971	0.899	73.8	0.62	0.45	**0.996**	0.979	94.8	**0.08**	0.13	**1.000**	**1.000**	**100.0**	**0.09**	0.17
	KITTI 08					NCLT 2012-01-15					NCLT 2012-02-04					NCLT 2012-03-17				
	AP	max F1	max Recall	$\hat{e}_t$ (m)	$\hat{e}_r$ (°)	AP	max F1	max Recall	$\hat{e}_t$ (m)	$\hat{e}_r$ (°)	AP	max F1	max Recall	$\hat{e}_t$ (m)	$\hat{e}_r$ (°)	AP	max F1	max Recall	$\hat{e}_t$ (m)	$\hat{e}_r$ (°)
M2DP [18]	0.081	0.162	0.0	-	-	0.783	0.695	4.8	-	-	0.700	0.620	3.7	-	-	0.654	0.621	4.0	-	-
Logg3D[21]	0.958	0.929	2.7	-	-	0.679	0.592	1.0	-	-	0.575	0.517	0.6	-	-	0.570	0.530	1.4	-	
CVTNet [20]	0.848	0.721	26.0	-	-	0.947	0.876	20.5	-	-	0.923	0.863	30.3	-	-	0.907	0.836	11.2	-	-
BoW3D [19]	0.905	0.829	14.4	1.44	2.81	0.000	0.000	0.0	-	-	0.000	0.000	0.0	-	-	0.000	0.000	0.0	-	-
LCDNet [7]	0.952	0.918	12.2	0.21	0.47	0.633	0.342	0.0	0.39	1.20	0.621	0.362	0.0	0.37	1.16	0.684	0.321	0.0	0.37	1.26
BEVPlace++ [9]	**0.999**	0.984	76.4	0.35	0.57	0.963	0.912	24.9	0.34	**1.09**	0.969	0.916	**34.5**	0.36	1.19	0.935	0.859	**31.2**	0.40	**1.17**
LighterBEV(ours)	**0.999**	**0.987**	**89.3**	**0.12**	**0.36**	**0.983**	**0.939**	**54.5**	**0.33**	1.11	**0.970**	**0.927**	17.4	**0.30**	**1.01**	**0.960**	**0.911**	10.0	**0.32**	1.23
	NCLT 2012-06-15					NCLT 2012-09-28					NCLT 2012-11-16					NCLT 2013-02-23				
	AP	max F1	max Recall	$\hat{e}_t$ (m)	$\hat{e}_r$ (°)	AP	max F1	max Recall	$\hat{e}_t$ (m)	$\hat{e}_r$ (°)	AP	max F1	max Recall	$\hat{e}_t$ (m)	$\hat{e}_r$ (°)	AP	max F1	max Recall	$\hat{e}_t$ (m)	$\hat{e}_r$ (°)
M2DP [18]	0.666	0.617	1.9	-	-	0.676	0.602	4.2	-	-	0.281	0.380	0.0	-	-	0.700	0.656	1.3	-	-
Logg3D[21]	0.427	0.413	0.3	-	-	0.509	0.476	1.0	-	-	0.282	0.279	0.0	-	-	0.511	0.472	0.2	-	-
CVTNet [20]	0.937	0.869	36.8	-	-	0.920	0.840	19.7	-	-	0.784	0.719	8.1	-	-	0.897	0.823	15.4	-	-
BoW3D [19]	0.024	0.102	0.0	-	-	0.000	0.000	0.0	-	-	0.000	0.000	0.0	-	-	0.000	0.000	0.0	-	-
LCDNet [7]	0.628	0.288	0.0	0.50	1.30	0.552	0.244	0.0	0.44	1.27	0.243	0.039	0.0	0.47	1.55	0.231	0.191	0.0	0.52	1.68
BEVPlace++ [9]	0.955	0.901	**63.4**	0.40	1.19	0.957	0.894	**45.3**	0.40	1.61	**0.839**	0.733	15.8	**0.40**	**1.10**	0.959	0.887	45.5	0.44	1.05
LighterBEV(ours)	**0.963**	**0.946**	59.3	**0.31**	**0.96**	**0.966**	**0.902**	17.9	**0.35**	**1.11**	0.819	**0.745**	**21.9**	0.49	2.07	**0.973**	**0.919**	**54.3**	**0.37**	**0.91**

TABLE VI
COMPLETE GLOBAL LOCALIZATION PERFORMANCE ON NCLT DATASET.

Dataset	Sequence	LCDNet				BEVPlace++				LighterBEV			
		Recall (%) ↑	SR(%) ↑	$\bar{e}_t$ (m) ↓	$\bar{e}_r$ (°) ↓	Recall (%) ↑	SR(%) ↑	$\bar{e}_t$ (m) ↓	$\bar{e}_r$ (°) ↓	Recall (%) ↑	SR(%) ↑	$\bar{e}_t$ (m) ↓	$\bar{e}_r$ (°) ↓
NCLT	2012-02-04	60.5	58.5	0.37	1.15	95.3	95.6	0.32	**1.06**	**96.6**	**97.2**	**0.29**	1.07
	2012-03-17	54.2	52.0	0.37	1.26	94.2	95.1	0.33	1.18	**95.3**	**95.8**	**0.30**	**1.17**
	2012-06-15	44.2	40.0	0.49	1.28	90.2	90.9	0.42	1.11	**93.0**	**93.3**	**0.38**	**1.10**
	2012-09-28	34.9	32.2	0.44	1.27	88.9	**89.8**	0.46	**1.23**	**90.7**	89.2	**0.42**	1.25
	2012-11-16	31.7	28.8	0.47	**1.54**	91.3	**90.2**	0.44	1.65	**92.6**	89.1	**0.40**	1.58
	2013-02-23	10.9	6.8	0.50	1.62	87.8	88.5	0.37	**1.05**	**91.2**	**91.4**	**0.35**	**1.05**

on *08*, LighterBEV attains higher max recall and lower $\hat{e}_t$/$\hat{e}_r$ than BEVPlace++, indicating perfect-precision detection with improved pose accuracy. These gains stem from compact yet informative local features that reduce redundancy and enhance distinctiveness. In cross-dataset testing on NCLT, all methods drop due to domain shift. Even so, CVTNet, BEVPlace++, and LighterBEV retain most of their detection performance, with LighterBEV achieving the best AP, F1 max score, and max recall. For pose estimation, LCDNet, BEVPlace++, and LighterBEV are comparable; notably, LighterBEV uses only 32-D local features (vs. 128 for BEVPlace++ and 640 for LCDNet), yielding faster registration and improved deployability.

C. Multi-Session Global Localization Evaluation

Multi-session global localization re-localizes LiDAR scans on an outdated database without an initial pose. We use NCLT *2012-01-15* as the database and treat all other sequences as queries. Each query first undergoes place recognition, followed by pose estimation to recover a 3-DoF pose. We evaluate LCDNet, BEVPlace++, and LighterBEV using models trained on KITTI *00*. As summarized in Table VI, domain shift causes degradation for all methods, with LCDNet deteriorating most—its recall and registration success decline over later sessions. In contrast, LighterBEV achieves a recall rate of over 90% across all sessions, demonstrating superior generalization and robustness to long-term appearance changes.

D. Online Learning Evaluation

In online learning scenarios, precise 6-DoF poses are often unavailable while coarse 3-DoF positions from the certified SLAM module or together with GPS signals can be provided. We therefore evaluate BEVPlace++ and LighterBEV under this constraint. Although LighterBEV generalizes strongly on NCLT, it still incurs a 6.8% retrieval-failure rate, highlighting the gap between closed-world training and open-world deployment. To address this, we initialize both models on KITTI sequence *00* and perform online learning on NCLT sequence *2013-02-23*, using the open-source BEVPlace++ implementation with default settings. Table VII reports mean results over five runs and seven metrics before and after online learning; both methods show consistent improvements, confirming the effectiveness of online adaptation in unseen environments.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE VII
COMPLETE GLOBAL LOCALIZATION PERFORMANCE ON NCLT DATASET WITH ONLINE LEARNING.

Dataset	Sequence	BEVPlace++							BEVPlace++*						
		Recall (%) ↑	AP ↑	max F1 ↑	max Recall ↑	SR(%) ↑	$\bar{e}_t$ (m) ↓	$\bar{e}_r$ (°) ↓	Recall (%) ↑	AP ↑	max F1 ↑	max Recall ↑	SR(%) ↑	$\bar{e}_t$ (m) ↓	$\bar{e}_r$ (°) ↓
NCLT	2012-01-15	-	0.944	0.872	17.6	-	-	-	-	0.984	0.932	48.5	-	-	-
	2012-02-04	94.7	0.929	0.859	30.4	96.1	0.29	1.08	96.9	0.975	0.920	38.6	97.1	0.28	1.07
	2012-03-17	92.8	0.915	0.841	38.4	93.5	0.30	1.18	96.0	0.963	0.887	43.5	95.8	**0.29**	1.17
	2012-06-15	90.0	0.914	0.827	33.7	90.8	0.41	1.10	95.2	0.985	0.943	67.9	95.2	0.39	1.12
	2012-09-28	87.3	0.914	0.827	33.7	87.1	0.44	1.27	93.8	0.983	0.933	48.9	92.1	0.42	**1.24**
	2012-11-16	90.4	0.609	0.592	9.2	87.7	0.41	1.59	94.2	0.878	0.809	11.3	90.8	**0.39**	**1.57**
	2013-02-23	87.3	0.877	0.783	31.3	88.7	0.34	1.06	92.5	0.985	0.940	55.9	92.1	**0.34**	1.05
Dataset	Sequence	LighterBEV							LighterBEV*						
		Recall (%) ↑	AP ↑	max F1 ↑	max Recall ↑	SR(%) ↑	$\bar{e}_t$ (m) ↓	$\bar{e}_r$ (°) ↓	Recall (%) ↑	AP ↑	max F1 ↑	max Recall ↑	SR(%) ↑	$\bar{e}_t$ (m) ↓	$\bar{e}_r$ (°) ↓
NCLT	2012-01-15	-	0.983	0.939	54.5	-	-	-	-	**0.992**	**0.957**	**56.8**	-	-	-
	2012-02-04	96.6	0.970	0.927	17.4	97.2	0.29	1.07	**97.5**	**0.985**	**0.946**	**41.2**	**97.6**	**0.27**	**1.06**
	2012-03-17	95.3	0.960	0.911	10.0	95.8	0.30	1.17	**96.8**	**0.981**	**0.926**	**61.6**	**96.4**	**0.29**	**1.16**
	2012-06-15	93.0	0.963	0.946	59.3	93.3	0.38	1.10	**95.9**	**0.994**	**0.967**	**74.7**	**95.3**	**0.37**	**1.09**
	2012-09-28	90.7	0.966	0.902	17.9	89.2	0.42	1.25	**94.1**	**0.993**	**0.959**	**70.4**	**92.3**	**0.41**	1.25
	2012-11-16	92.6	0.819	0.745	21.9	89.1	0.40	1.58	**95.2**	**0.929**	**0.857**	**37.2**	**91.4**	**0.39**	**1.57**
	2013-02-23	91.2	0.973	0.919	54.3	91.4	0.35	1.05	**93.3**	**0.991**	**0.955**	**76.8**	**92.5**	**0.34**	**1.04**

TABLE VIII
RUNTIME ANALYSIS.

	Feature Extract. (ms)	Place Retrieval. (ms)	Pose Estim. (ms)
BEVPlace++ [9]	12.7	0.26	80.9
LighterBEV(ours)	**12.5**	**0.066**	68.5
LighterBEV†(ours)	**12.5**	**0.066**	**4.0**

E. Efficiency Analysis

Table VIII compares BEVPlace++ and LighterBEV on a desktop with an i9-14900K CPU and an RTX 4090D GPU. For pose estimation, both methods use RANSAC with 1000 maximum iterations. LighterBEV achieves faster feature extraction thanks to reduced feature dimensionality, which accelerates NetVLAD aggregation. In retrieval, both methods use FAISS [25], but LighterBEV is four times faster due to the Informative Compression Module reducing descriptor size. With the same RANSAC, it is 12.4 ms faster in pose estimation (15% speedup), and with an optimized backend (LighterBEV†), pose estimation takes only 4 ms. Overall, LighterBEV† achieves a runtime speed exceeding 30 Hz.

TABLE IX
EFFICIENCY ANALYSIS FOR ONLINE LEARNING.

	Training time (s)	GPU Usage (GB)	Platform
BEVPlace++ [9]	188	11.6	desktop
BEVPlace++ [9]	623	6.9	mobile
LighterBEV(ours)	**155**	7.3	desktop
LighterBEV(ours)	465	**5.8**	mobile

We then evaluate the efficiency of online learning. Table IX reports training time and GPU usage during online learning on two platforms: the RTX 4090D desktop and a mobile device with an RTX 4060 GPU. The NCLT sequence *2013-02-23* contains 4649 BEV images. Assuming the LiDAR operates at 10 Hz, the entire data stream spans approximately 464.9 s. Both methods achieve real-time online learning on the desktop. On the mobile platform, only LighterBEV maintains near real-time performance with a 0.1 s delay. Due to differences in PyTorch memory management across GPUs, memory usage varies by platform. LighterBEV attains real-time online learning with substantially lower GPU memory consumption, which improves deployability.

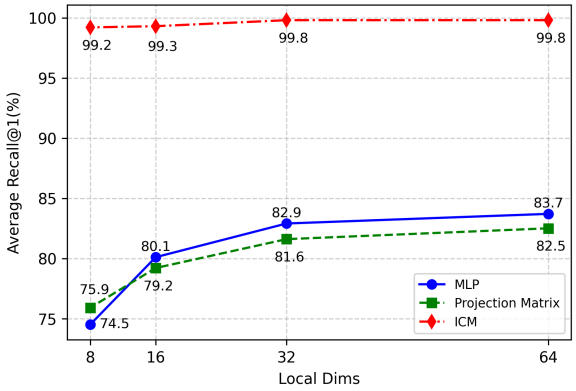

Fig. 3. Ablation study for different compression modules with varying local feature dimensions.

F. Ablation Study

In this section, we conduct ablation studies to evaluate the impact of the ICM module in LighterBEV and the online learning algorithm used for LighterBEV. For the ablation studies on the ICM module, we replace the ICM module with either a randomly initialized MLP module or a randomly initialized projection matrix module and vary the compression dimensions across 64, 32, 16, and 8. Fig. 3 shows the results of LighterBEV with different feature compression modules on the KITTI dataset. We use average recall@1 across all sequences for evaluation. It can be observed that with randomly initialized compression modules, the model suffers significant performance drops. In practice, both the MLP module and the projection matrix module encounter convergence challenges. In contrast, the ICM consistently preserves accuracy under compression. Even with a sixteenfold reduction (8-D local features), it attains 99.2% average Recall@1 on KITTI, with only a 0.6% performance drop relative to the best one.

We evaluate the online buffer by isolating its three operations.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE X
ABLATION STUDY FOR ONLINE LEARNING.

	Reservoir Adding	Informative Sampling	Feature Refreshing	AR@1↑	$\bar{e}_t$ (m) ↓	$\bar{e}_r$ (°) ↓
[a]	×	×	×	94.0	0.35	1.20
[b]	√	×	×	94.2	0.35	1.20
[c]	√	√	×	94.9	0.35	1.20
[d]	√	√	√	**95.5**	0.35	1.20
[e]	√	×	√	94.3	0.35	**1.19**
[f]	×	√	√	94.7	**0.34**	1.21

Reservoir adding is replaced with a First-In-First-Out (FIFO) based adding strategy, informative sampling with random sampling, and feature refreshing is toggled on and off. Table X reports average Recall@1 on NCLT for place recognition and mean translation and rotation errors for pose estimation. Configuration [a], which uses none of the operations, performs worst. Adding only reservoir adding ([b]) improves results, and adding informative sampling on top of reservoir adding ([c]) improves further. Using all three operations ([d]) yields the best place recognition performance, whereas variants [e] and [f], which omit informative sampling and reservoir adding, respectively, achieve lower recall rates than [d]. And for pose estimation, because precise pose constraints are unavailable, translation and rotation errors remain broadly similar during online learning.

VII. CONCLUSIONS

We presented LighterBEV, a lightweight LiDAR global localization framework that achieves state-of-the-art accuracy on four large-scale datasets and integrates online learning to enable automatic adaptation to new environments after deployment. Nevertheless, dynamic objects such as vehicles and pedestrians can introduce geometric inconsistencies in highly dynamic scenes. To further enhance robustness, future work will incorporate LiDAR-based moving object segmentation [26], [27] to mitigate their influence. In addition, catastrophic forgetting remains a major challenge in online learning, and we plan to investigate continual adaptation strategies that preserve prior knowledge while supporting real-time updates.

REFERENCES

[1] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine learning*, vol. 79, pp. 151–175, 2010.

[2] B. Liu, T. Yang, Y. Fang, and Z. Yan, "Micl: Mutual information guided continual learning for lidar place recognition," *IEEE Robotics and Automation Letters*, vol. 9, no. 11, pp. 10 463–10 470, 2024.

[3] J. Cui and X. Chen, "CCL: Continual Contrastive Learning for LiDAR Place Recognition," *IEEE Robotics and Automation Letters (RA-L)*, 2023.

[4] Z. Yan, L. Sun, T. Krajnik, T. Duckett, and N. Bellotto, "Towards long-term autonomy: A perspective from robot learning," *arXiv preprint arXiv:2212.12798*, 2022.

[5] S. Li, X. Wang, Y. Cao, F. Xue, Z. Yan, and H. Zha, "Self-supervised deep visual odometry with online adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6339–6348.

[6] Y. Pan, W. Zhou, Y. Cao, and H. Zha, "Adaptive vio: Deep visual-inertial odometry with online continual learning," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2024, pp. 18 019–18 028.

[7] D. Cattaneo, M. Vaghi, and A. Valada, "Lcdnet: Deep loop closure detection and point cloud registration for lidar slam," *IEEE Transactions on Robotics*, vol. 38, no. 4, pp. 2074–2093, Aug 2022.

[8] J. Komorowski, M. Wysoczanska, and T. Trzcinski, "Egonn: Egocentric neural network for point cloud based 6dof relocalization at the city scale," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 722–729, 2022.

[9] L. Luo, S.-Y. Cao, X. Li, J. Xu, R. Ai, Z. Yu, and X. Chen, "Bevplace++: Fast, robust, and lightweight lidar global localization for unmanned ground vehicles," *IEEE Transactions on Robotics (T-RO)*, vol. 41, pp. 4479–4498, 2025.

[10] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "Netvlad: Cnn architecture for weakly supervised place recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5297–5307.

[11] D. P. Kingma, M. Welling, *et al.*, "Auto-encoding variational bayes," 2013.

[12] L. Luo, S. Zheng, Y. Li, Y. Fan, B. Yu, S.-Y. Cao, J. Li, and H.-L. Shen, "Bevplace: Learning lidar-based place recognition using bird's eye view images," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 8666–8675.

[13] B. Liu, D. Yao, R. Yang, Z. Yan, and T. Yang, "Semi-supervised online continual learning for 3d object detection in mobile robotics," *Journal of Intelligent & Robotic Systems*, vol. 110, no. 4, pp. 1–16, 2024.

[14] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9*. Springer, 2006, pp. 430–443.

[15] M. A. Uy and G. H. Lee, "Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4470–4479.

[16] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The oxford robotcar dataset," *The International Journal of Robotics Research*, vol. 36, no. 1, pp. 3–15, 2017.

[17] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, "University of michigan north campus long-term vision and LiDAR dataset," *The International Journal of Robotics Research*, vol. 35, no. 9, pp. 1023–1035, 2016.

[18] L. He, X. Wang, and H. Zhang, "M2dp: A novel 3d point cloud descriptor and its application in loop closure detection," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 231–237.

[19] Y. Cui, X. Chen, Y. Zhang, J. Dong, Q. Wu, and F. Zhu, "Bow3d: Bag of words for real-time loop closing in 3d lidar slam," *IEEE Robotics and Automation Letters*, vol. 8, no. 5, pp. 2828–2835, 2022.

[20] J. Ma, G. Xiong, J. Xu, and X. Chen, "Cvtnet: A cross-view transformer network for lidar-based place recognition in autonomous driving environments," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 3, pp. 4039–4048, 2024.

[21] K. Vidanapathirana, M. Ramezani, P. Moghadam, S. Sridharan, and C. Fookes, "Logg3d-net: Locally guided global descriptor learning for 3d place recognition," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2215–2221.

[22] Z. Liu, S. Zhou, C. Suo, P. Yin, W. Chen, H. Wang, H. Li, and Y.-H. Liu, "Lpd-net: 3d point cloud learning for large-scale place recognition and environment analysis," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 2831–2840.

[23] L. Hui, H. Yang, M. Cheng, J. Xie, and J. Yang, "Pyramid point cloud transformer for large-scale place recognition," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 6078–6087.

[24] J. Komorowski, "Improving point cloud based place recognition with ranking-based loss and large batch training," in *2022 26th International Conference on Pattern Recognition (ICPR)*, 2022, pp. 3699–3705.

[25] J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with GPUs," *IEEE Transactions on Big Data*, vol. 7, no. 3, pp. 535–547, 2019.

[26] X. Chen, J. Cui, Y. Liu, X. Zhang, J. Sun, R. Ai, W. Gu, J. Xu, and H. Lu, "Joint scene flow estimation and moving object segmentation on rotational lidar data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 11, pp. 17 733–17 743, 2024.

[27] N. Wang, C. Shi, R. Guo, H. Lu, Z. Zheng, and X. Chen, "Insmos: Instance-aware moving object segmentation in lidar data," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 7598–7605.