

LPR-Mate: A Lightweight Universal Reranking-based Optimizer for LiDAR Place Recognition in Challenging Environments

Zhengkua Zhang ^{a, *}, Mingcong Shu ^b, Meng Sun ^a

^a School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China;

^b Huaiyin Normal University, Huaian 223300, China;

Keywords: LiDAR place recognition; Reranking optimization; Localization enhancement; Spatial consistency

Abstract

LiDAR place recognition (LPR) plays a critical role in simultaneous localization and mapping (SLAM) and autonomous driving systems. However, current LPR methods exhibit significant performance degradation under rotational shifts, noise interference, point cloud sparsity, and long-term environmental changes. This limitation stems from their reliance on fixed-length global descriptors, which lack the capacity to preserve comprehensive scene information in complex scenarios. To address these challenges, we propose LPR-Mate, a lightweight universal reranking-based optimizer that enhances the robustness of existing LPR frameworks in challenging environments. LPR-Mate processes top-k retrieval candidates from baseline LPR methods through a dual-stage pipeline: (1) A fast trigger mechanism evaluates spatial consistency between query and candidate scenes, selectively activating reranking only for low-confidence matches; (2) An independent reranking network refines candidate rankings by fusing local features, global descriptors, and spatial consistency scores through group and channel attention mechanisms. Extensive experiments on the Oxford RobotCar, NUS-Inhouse, and MulRan datasets demonstrate that LPR-Mate achieves >96% recall in localization accuracy validation and delivers a 32.34% average improvement in Recall@1 under rotational shifts, sparsity, and noise perturbations, while maintaining robustness for raw point clouds and long-term scenarios. As a plug-and-play module, LPR-Mate integrates seamlessly with diverse LPR architectures—including region-sampling and sparse-voxelization-based methods—without requiring retraining or structural modifications, ensuring computational efficiency and cross-architectural universality.

1. Introduction

Place recognition involves retrieving geotagged databases to identify the submap most similar to captured scene data, thereby determining the current location. This capability is critical for robotic applications such as loop closure in simultaneous localization and mapping (SLAM) systems (Xu et al. 2022; Lehtola et al. 2016) and global localization in GPS-denied environments (Yu et al. 2021). While vision-based methods have seen significant advancements, their performance remains susceptible to environmental changes, including illumination variations, viewport shifts, and adverse weather conditions. In contrast, LiDAR place recognition (LPR) demonstrates greater robustness, benefiting from LiDAR’s inherent insensitivity to lighting variations and weather conditions.

Most LPR methods encode scene point clouds into fixed-length global descriptors to enable rapid retrieval and localization through feature distance comparisons. Early approaches relied on handcrafted features, which exhibited limited robustness to scene variations and required manual parameter tuning. Recent advances in deep neural networks have significantly improved LPR accuracy by leveraging diverse LiDAR data representations, including raw point clouds (Uy and Lee 2018; Zhang et al. 2023), voxels (Komorowski 2021; Xu et al. 2021), bird’s-eye views (Luo et al. 2023), and range images (Stathouloupoulos et al. 2024). However, despite these advancements, global descriptor-based methods—though achieving strong performance on benchmark datasets such as Oxford RobotCar (Maddern et al. 2017) and NUS Inhouse (Maddern et al. 2017)—demonstrate notable accuracy degradation under real-world conditions due to temporal and spatial discrepancies between database construction and user-side localization. Two critical challenges arise: First, existing methods exhibit performance degradation when query point clouds are rotated relative to database submaps (evidenced in Table. 2). Such rotational discrepancies frequently occur during lane changes, turns, or multi-directional traversals, necessitating robust rotation-invariant retrieval capabilities. Second, environmental dynamics over time—including weather fluctuations, seasonal transitions, and structural alterations—introduce point density variations and noise distribution disparities between query and database submaps, compromising long-term system reliability.

The core limitation of current methods lies in their reliance on fixed-length global descriptors, which lack the granularity to

resolve the aforementioned challenges. This dependency necessitates preprocessing steps (e.g., ground removal, downsampling) to eliminate redundant data, thereby preventing direct processing of raw point clouds. Moreover, the emphasis on global features motivates the development of complex local geometric extractors, yielding oversized models incompatible with resource-constrained devices. Recent works have introduced re-ranking methodologies like TReR (global feature enhancement) and SpectralGV (local spatial consistency evaluation) to refine localization. However, these methods neglect the integration of global and local features, creating suboptimal representation synergy; Additionally, the absence of localization quality assessment mechanisms prompts indiscriminate optimization attempts, yielding unwarranted computational costs without performance guarantees.

To address these limitations, we propose LPR-Mate, a lightweight universal reranking mechanism designed to reduce the reliance of existing LPR methods on global descriptors and enhance their robustness in challenging environments. Operating as a plug-and-play post-optimizer, LPR-Mate processes the local and global features from the top k initial retrieval candidates (k=25 in this work) generated by existing LPR methods through two sequential phases: First, a training-free trigger mechanism computes spatial consistency scores between query and candidate scenes to identify potential mismatches. If activated, the system initiates reranking. Second, an independent reranking network refines candidate rankings by fusing local features, global descriptors, and spatial consistency scores via spatial-channel attention mechanisms, thereby improving localization accuracy. Comprehensive evaluations across three large-scale datasets—Oxford RobotCar (Maddern et al. 2017), NUS-Inhouse (Uy and Lee 2018), and MulRan (Jung et al. 2023; Kim et al. 2020)—confirm the framework’s superior performance. LPR-Mate achieves a >96% recall rate in validating localization accuracy and delivers a 32.34% average improvement in Recall@1 metrics under rotational shifts, noise, sparsity, and raw point cloud conditions, significantly enhancing the robustness of existing LPR methods in challenging environments.

2. Related Works

Learning global descriptor for LPR task: Global descriptors compress scene information to enable efficient database retrieval in LPR task, addressing the computational infeasibility of direct pairwise submap comparisons. PointNetVLAD (Uy and Lee

2018) pioneered this approach by combining PointNet-based local feature extraction with NetVLAD-based global descriptor fusion. Subsequent advancements refined this framework: PCAN (Zhang and Xiao 2019) improved accuracy via a point contextual attention module, LPD-Net (Liu et al. 2019) integrated hand-crafted and graph-based features, and EPC-Net (Hui et al. 2022) enhanced efficiency with lightweight ProxyConv. However, reliance on computationally intensive operations like farthest point sampling and k-nearest neighbor (KNN) searches limited scalability for large-scale point cloud. Recent lightweight architectures, such as LWR-Net (Zhang et al. 2023) and MinkLoc3D (Komorowski 2021), reduced overhead using random sampling or sparse voxelized representations. Extensions like MinkLoc++ (Komorowski, Wysoczańska, and Trzcinski 2021) fused multimodal data, while MinkLoc3Dv2 (Komorowski 2022) introduced ranking-based losses. Despite progress, these methods exhibit sensitivity to rotational variances, sparse distributions, and noise, as validated by our experiments.

Reranking-based optimization for LPR task: Reranking methods refine initial retrieval results by reordering candidates to prioritize correct matches. While widely explored in image/text retrieval through techniques like RANSAC-based geometric verification (Avrithis and Tolas 2014) or transformer-driven local feature alignment (Tan, Yuan, and Ordonez 2021), these approaches face significant challenges in LiDAR applications due to point clouds' massive data volume and complex 3D geometry. Recent attempts to balance efficiency and accuracy include Query Expansion (QE) (Radenović, Tolas, and Chum 2019) that aggregates database features to augment queries, and SuperGlobal (Shao et al. 2023) which enhances global descriptors via GeM pooling variants (GeM+, Scale-GeM, Regional-GeM) with top-k neighbor aggregation. However, such methods remain computationally inefficient or insufficiently robust for large-scale LiDAR scenarios. Most LPR research focuses on improving global descriptor learning, leaving reranking based on local features underexplored – existing solutions like TRER's transformer-based global descriptor reranking (Barros et al. 2023) or GIDP's training-set relationship (Fan, Song, Liu, et al. 2023) mimicry achieve efficiency but suffer from accuracy limitations in complex environments due to fixed-length descriptor constraints. Conversely, SpectralGV (Vidanapathirana et al. 2023) employs spectral geometric verification with local features but neglects global-local feature synergy. Crucially, current frameworks universally lack adaptive triggering mechanisms, applying resource-intensive reranking indiscriminately regardless of initial retrieval quality, thereby incurring unnecessary computational costs.

3. Methodology

We propose LPR-Mate, a lightweight universal reranking framework designed to enhance the localization accuracy of existing LPR methods in challenging environments. LPR-Mate processes the top k retrieval candidates' information ($k=25$ in this work), which include local regional information (denoted as $\mathbf{L} = \{\mathbf{P}_{lr}, \mathbf{F}_{lr}\}$) and global descriptors ($\mathbf{G}$), where $\mathbf{P}_{lr} \in \mathbb{R}^{N_s \times 3}$ represents local region coordinates and $\mathbf{F}_{lr} \in \mathbb{R}^{N_s \times 256}$ corresponds to local features. LPR-Mate operates through two sequential phases: First, a training-free trigger mechanism evaluates retrieval reliability by computing spatial consistency scores between query and candidate scenes. If triggered (indicating low-confidence retrieval), the system initiates reranking (Section 3.1); Second, an independent reranking network optimizes candidate rankings by fusing local features, global descriptors, and spatial consistency scores via spatial and channel attention mechanisms, prioritizing geometrically coherent matches (Section 3.2). Implementation details are elaborated in Section 3.3.

Notably, existing LPR methods adopt an encoding paradigm of "local feature extraction $\rightarrow$ global feature aggregation", inherently encapsulating both local and global feature learning. LPR-Mate directly leverages the pre-trained network's local and global features generated during this process, eliminating the need for architectural modifications or retraining of the original LPR networks.

3.1 Trigger phase

The Trigger phase evaluates the reliability of retrieved candidates to avoid unnecessary computational overhead from indiscriminate reranking. Inspired by SpectralGV (Vidanapathirana et al. 2023), we construct an undirected graph $D = (V, \mathcal{E})$ between the query and candidate scenes using local regional positions $\mathbf{P}_{lr}$ and features $\mathbf{F}_{lr}$. Vertices $V = \{v_1, v_2, \dots, v_{N_s}\}$ represent correspondences identified via nearest-neighbour searches on local feature distances, where $v_t = (P_t^q, P_t^{ci})$ links query and candidate local regions. Edges $\mathcal{E}$ encode pairwise geometric compatibility via matching scores:

$$ms_{v_1 v_2} = \left[1 - \frac{d_{v_1 v_2}^2}{d_{thr}^2} \right]_+, \quad (1)$$

where $[\cdot]_+ = \max(\cdot, 0)$ ensures non-negative scores, $\|\cdot\|$ denotes L2 distance, and d_{thr} is a threshold. Correspondence pairs with $d_{v_1 v_2} > d_{thr}$ receive zero scores (geometric incompatibility, Fig. 1b), while preserved spatial relationships yield high scores (Fig. 1a).

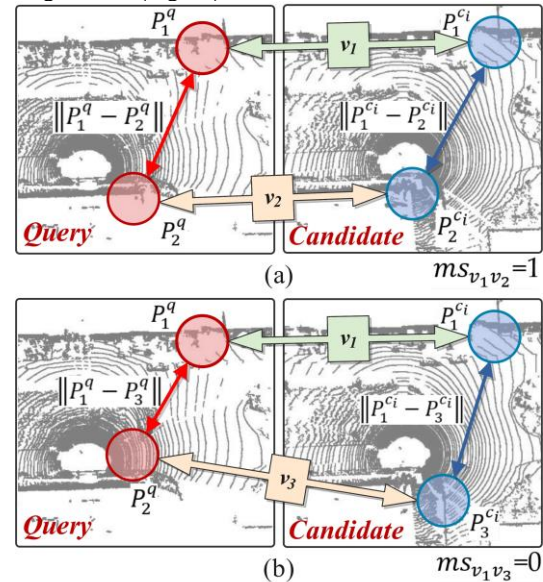


Figure 1. Pairwise geometric compatibility analysis of retrieval candidates: (a) Vertices with a high matching score; (b) Vertices with a zero matching score.

The matching scores populate matrix $\mathbf{M} \in \mathbb{R}^{N_s \times N_s}$, and we seek the inter-cluster $\mathbf{C}$ maximizing the Rayleigh quotient:

$$\hat{c} = \operatorname{argmax} \left(\frac{c^T \mathbf{M} c}{c^T c} \right) \quad (2)$$

Where $c \in \mathbb{R}^{N_s \times 1}$ is a binary cluster indicator vector. The principal eigenvector of $\mathbf{M}$ approximates $\hat{c}$, and the spatial consistency (SC) score is computed as:

$$SC \text{ score} = \hat{c}^T \mathbf{M} \hat{c} \quad (3)$$

Unlike SpectralGV, which directly reranks candidates using spatial compatibility, our method activates reranking only if the top candidate's SC score is not the highest. This lightweight, training-free mechanism ensures computational efficiency while preserving accuracy.

3.2 Reranking phase

The reranking phase performs fine-grained scene comparisons by leveraging multi-level features (local region features $\mathbf{L} = \{\mathbf{P}_{lr}, \mathbf{F}_{lr}\}$, global descriptors $\mathbf{G}$, and SC scores) rather than relying solely on global descriptors. An independent network processes these inputs to generate a probability vector $\mathbf{Prob} \in \mathbb{R}^{k \times 1}$, representing the likelihood of each candidate being a correct localization result (Fig. 2).

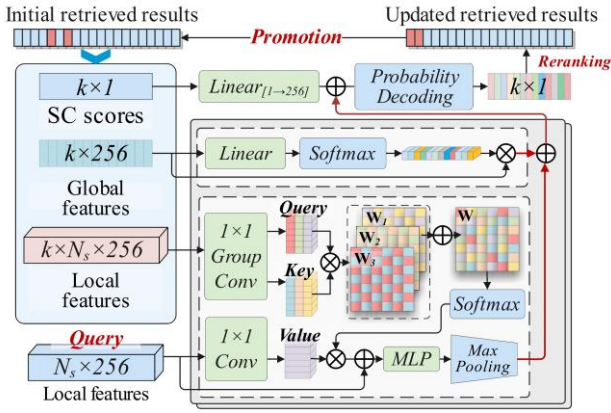


Figure 2. The architecture of reranking network

The network employs three parallel branches for multi-level feature encoding and fusion:

(1) Local Feature Enhancement: The local feature enhancement branch refines candidate scene features through a grouped attention mechanism interacting with query features:

$$\mathbf{Q}_g = W_q^g(\mathbf{F}_{lr}^{C_i}), \mathbf{K}_g = W_k^g(\mathbf{F}_{lr}^q), \mathbf{V} = W_v(\mathbf{F}_{lr}^{C_i}), \quad (4)$$

$$\mathbf{W} = \sum_{g=1}^G \mathbf{Q}_g \mathbf{K}_g^T, \quad (5)$$

$$\mathbf{F}_{lr,enhanced}^{C_i} = \text{MLP}(\text{Softmax}(\frac{\mathbf{W}}{\sqrt{C}})) \mathbf{V} + \mathbf{F}_{lr}^{C_i}. \quad (6)$$

where $\mathbf{F}_{lr}^q$ and $\mathbf{F}_{lr}^{C_i}$ denote local features of the query and i -th candidate, $W_q^g(\cdot)$ and $W_k^g(\cdot)$ are group-wise 1×1 convolutional layers that partition the feature map into G groups, denoted as $\{\mathbf{Q}_g, \mathbf{K}_g \in \mathbb{R}^{N_s \times (\frac{C}{G})} | g = 1, \dots, G\}$; and $W_v(\cdot)$ is a 1×1 convolutional layer to generate the value feature map $\mathbf{V} \in \mathbb{R}^{N_s \times C}$. This grouped attention reduces computational complexity from $O(m_h^2 \cdot C)$ to $O(m_h^2 \cdot (\frac{C}{G} + G))$ while retaining discriminative power. Where m_h represents the number of heads in the multi-head attention mechanism. To maintain the network's lightweight design, we utilize a single-head self-attention mechanism with $m_h = 1$.

(2) Global Feature Enhancement: For the global descriptor used in the retrieval phase, each channel is treated uniformly. However, given that global descriptors encode abstract, compact scene representations, feature channels inherently carry unequal discriminative weights. To address this limitation, the global feature enhancement branch introduces channel-wise attention to address uniform channel treatment in global descriptors. As formalized in Equations (7) and (8), we first compute the global feature difference for the i -th candidate scene:

$$\mathbf{F}_{gl,diff}^{C_i} = \mathbf{F}_{gl}^{C_i} - \mathbf{F}_{gl}^q \quad (7)$$

where $\mathbf{F}_{gl}^q$ and $\mathbf{F}_{gl}^{C_i}$ denote the query and candidate scene global descriptors, respectively. A linear layer $\text{Linear}(\cdot)$ followed by a sigmoid activation generates channel-wise attention weights of dimension 1×256 . The refined global feature $\mathbf{F}_{gl,enhanced}^{C_i} \in \mathbb{R}^{256}$ is then obtained via:

$$\mathbf{F}_{gl,enhanced}^{C_i} = \text{Sigmoid}(\text{Linear}(\mathbf{F}_{gl,diff}^{C_i})) \cdot \mathbf{F}_{gl,diff}^{C_i}. \quad (8)$$

(3) SC Score Fusion: The SC score fusion branch encodes spatial consistency scores into the feature space:

$$\mathbf{F}_{Final}^{C_i} = \text{Linear}_{[1 \rightarrow 256]}(\mathbf{F}_{SC}^{C_i}) + \mathbf{F}_{gl,enhanced}^{C_i} + \text{Max}(\mathbf{F}_{lr,enhanced}^{C_i}). \quad (9)$$

where $\mathbf{F}_{SC}^{C_i} \in \mathbb{R}^1$ denote the SC score of i -th candidate scene, which is subsequently expanded from a dimension of 1 to 256 using $\text{Linear}_{[1 \rightarrow 256]}(\cdot)$. The $\text{Max}(\cdot)$ represents max pooling. The enhanced local feature and global feature are fused with the encoded SC score through a summarization process to produce final feature of the i -th candidate scene, denoted as $\mathbf{F}_{Final}^{C_i} \in \mathbb{R}^{256}$. The updated features of all candidate scenes are denoted as $\mathbf{F}_{enhanced} \in \mathbb{R}^{k \times 256}$, which is decoded into probabilities Equation (10). The linear mapping layer followed by the Sigmoid function to transform the feature into the probability of each scene being correctly localized. Based on this probability, the candidate scenes can be reorganized.

$$\mathbf{Prob} = \text{Sigmoid}(\text{Linear}(\mathbf{F}_{enhanced})). \quad (10)$$

Due to the imbalance of correctly and incorrectly located scene in the initial candidates, the re-ranking network utilize the Focal loss function (Lin et al. 2017) for training, as shown in Equation (11). Where Prob_t represents the estimated probability of the t -th scene in $\mathbf{Prob}$, GtLab_t denotes the ground truth label of the t -th scene, and p_t represents the probability assigned by the model to the correct localization. The parameter $\alpha_t \in [0, 1]$ acts as a weighting parameter for adjusting the proportion of loss between positive and negative samples, while γ serves as a focusing parameter that can be tuned to control the focusing effect. When γ is set to a value greater than 0, more emphasis is placed on challenging, misclassified examples; however, a value of $\gamma = 0$ reduces it to the standard cross-entropy loss.

$$\text{FocalLoss}(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t), p_t = \begin{cases} \text{Prob}_t & \text{if } \text{GtLab}_t = 1 \\ 1 - \text{Prob}_t & \text{otherwise} \end{cases}. \quad (11)$$

3.3 Implementation details

Two hierarchical encoding layers are implemented in the reranking network, each maintaining a feature dimension of 256. The local feature enhancement branch adopts grouped attention with $G = 4$ groups. The probability decoding stage incorporates three linear layers with output dimensions of 128, 64, and 1, respectively. The Focal loss function is configured with $\alpha_t = 0.25$ and $\gamma = 2$, following the recommendations of (Lin et al. 2017). All experiments are conducted on an Intel i7-9700k CPU and an NVIDIA RTX 2070S GPU.

4. Experiments

4.1 Experimental setting

Benchmark datasets: Three large-scale datasets—Oxford RobotCar (Maddern et al. 2017), NUS In-house (Uy and Lee 2018), and MulRan (Jung et al. 2023)—were utilized to evaluate the proposed method. For comprehensive comparisons under standardized conditions, the Oxford RobotCar and NUS In-house datasets (collectively referred to as standardized point cloud datasets) were employed. The Oxford RobotCar dataset comprises 10 km urban trajectories captured in central Oxford using a vehicle-mounted Sick LMS-151 LiDAR. The NUS In-house dataset includes three segments: University Sector (U.S., 10 km), Residential Area (R.A., 8 km), and Business District (B.D., 5 km), collected with a Velodyne HDL-64 LiDAR. Both datasets underwent preprocessing: non-informative ground points were removed, point clouds were uniformly downsampled to 4,096 points, and normalized to zero mean.

within $[-1, 1]$. Each submap was georeferenced using UTM coordinates from GPS/INS data.

The MulRan dataset (raw point cloud dataset) was adopted to assess performance under real-world challenges. It comprises Ouster-64 LiDAR data collected across four distinct environments: Daejeon Convention Center (DCC), KAIST, Riverside, and Sejong City. To rigorously evaluate long-term place recognition capabilities, the recently proposed MulRan-Longterm subset (Jung et al. 2023) was constructed by integrating Ouster-128 LiDAR data captured four years after the original dataset. This subset also introduces sensor variance, as the original database (Ouster-64) and query scans (Ouster-128) differ in resolution. Notably, the MulRan dataset retains raw, unprocessed point clouds—no ground removal or downsampling was applied—to preserve real-world fidelity for robust evaluation.

To generate training tuples, point cloud pairs with a distance of less than 10m were defined as positive pairs, while more than 50m were defined as negative pairs. We further applied default random rotations to each query submap to evaluate the performance of methods under point cloud rotation, including rotations around the Z-axis within ± 30 degrees and rotations around the X-axis and Y-axis within ± 5 degrees.

Evaluation metrics: Performance was evaluated using the Recall@K metric, —specifically Recall@1 (top match) and Recall@1% (top 1% of candidates)—was applied. A query was considered successfully localized if at least one of the top-K retrieved database scans fell within 25m of the ground-truth position. Computational efficiency was assessed via average running time and model size (in millions of parameters, M).

4.2 Enhancement offered by LPR-Mate

We selected four LPR methods as backbone—LWR-Net (Zhang et al. 2023), MinkLoc3D (Komorowski 2021), SVT-Net (Fan et al. 2022), and MinkLoc-v2 (Komorowski 2022)—to evaluate LPR-Mate’s optimization efficacy across diverse LPR architectures. Our selection prioritized divergent local feature perception mechanisms: LWR-Net employs a region-sampling and neighborhood-grouping strategy, while MinkLoc3D, SVT-Net, and MinkLoc-v2 rely on sparse-voxelization-based regional extraction. This methodological dichotomy between regional sampling and voxelization paradigms ensures validation of LPR-Mate’s adaptability.

We first evaluated the effectiveness of LPR-Mate’s trigger mechanism on existing methods under three environmental challenges: rotational variance, noise interference, and point cloud sparsity. Controlled perturbations were applied to query scenes using three protocols: (1) pose variation via random rotations ($\pm 30^\circ$ about the Z-axis, $\pm 5^\circ$ about X/Y-axes); (2) sparsity simulation by downsampling original point clouds at predefined ratios; and (3) noise simulation through injection of outlier points, constrained by a 5 cm Euclidean distance threshold from structural elements to ensure realistic noise patterns.

As demonstrated in Table 1, the trigger mechanism achieves high activation recall rates (94.64% under rotation, 94.71% under noise, 95.52% under sparsity) across diverse LPR backbone, regardless of input feature types (local regions or sparse voxels). A negative correlation exists between the trigger’s activation ratio and the backbone’s Recall@1. This indicates that when a method excels in the retrieval stage (with a high Recall@1 rate), the trigger phase exhibits a low activation ratio, thereby avoiding the subsequent re-ranking computation and conserving computational resources. Conversely, when a method performs poorly in the retrieval stage, the trigger stage is promptly activated to optimize the current localization performance through the re-ranking stage.

Methods	With random rotation		
	Recall@1 (%)	Trigger Ratio (%)	Trigger Recall (%)
LWR-Net	39.85	68.74	92.46
MinkLoc3D	74.26	45.46	95.41
SVT-Net	72.12	49.92	96.08
MinkLoc-v2	84.57	36.73	94.60
Methods	With 80% downsampling ratio		
	Recall@1 (%)	Trigger Ratio (%)	Trigger Recall (%)
LWR-Net	33.89	73.81	91.89
MinkLoc3D	55.24	61.83	95.50
SVT-Net	44.23	68.10	97.35
MinkLoc-v2	70.95	48.68	94.10
Methods	With 256 outlier points		
	Recall@1 (%)	Trigger Ratio (%)	Trigger Recall (%)
LWR-Net	33.89	73.81	91.89
MinkLoc3D	60.91	58.30	96.29
SVT-Net	66.77	52.72	97.70
MinkLoc-v2	79.58	45.63	96.20

Table 1. Evaluation of trigger mechanism on Oxford RobotCar Dataset

Figure 3 demonstrates LPR-Mate’s systematic performance gains when integrated with existing methods. Across rotational perturbations, extreme sparsity, and outlier corruption scenarios, the framework achieves 32.34% average (48.40% peak) Recall@1 improvements - independent of base encoders’ structural paradigms (regional geometric neighborhoods in LWR-Net or sparse voxel hierarchies in MinkLoc3D). While effectively enhancing weak retrievers (+17.46% Recall@1 for LWR-Net under 90° rotations), performance remains candidate-dependent: Ground-truth exclusion in initial retrievals necessitates either candidate pool expansion (top-50 vs. top-25) or robust backbone selection for extreme conditions.

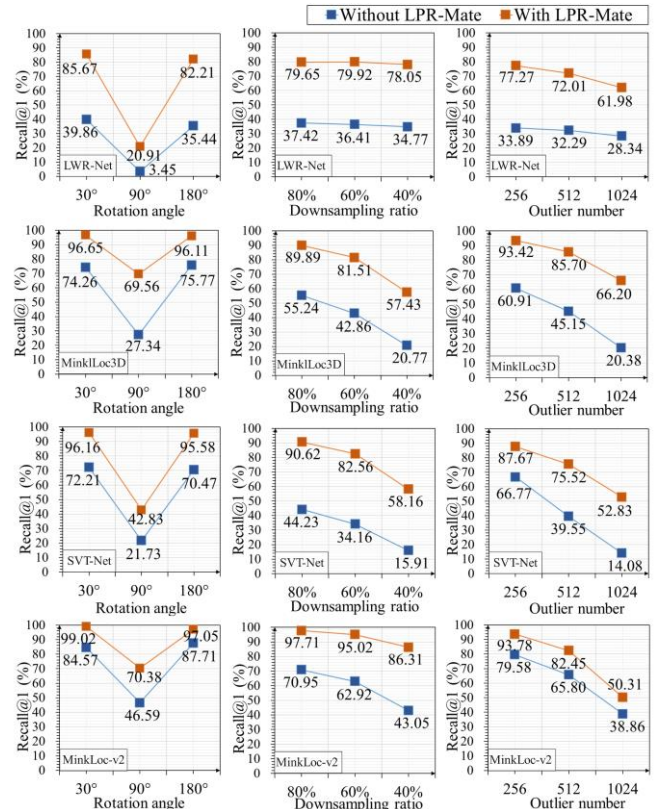


Figure 3. Retrieval enhancement of existing methods by the LPR-Mate under varying conditions of point cloud rotation, outlier, and sparsity levels.

4.3 Experiments on standardized datasets

In this experiment, we employed our previously proposed method, LR-Net (Zhang et al. 2024), as the backbone to quantify performance enhancements enabled by LPR-Mate optimization (denoted as LR-Mate in comparative analyses). LR-Mate was benchmarked against 16 state-of-the-art methods: PointNetVLAD (PNVLAD) (Uy and Lee 2018), PCAN (Zhang and Xiao 2019), LPD-Net (Liu et al. 2019), SOE-Net (Xia et al. 2021), EPC-Net (Hui et al. 2022), MinkLoc3D (MinkL3D) (Komorowski 2021), MinkLoc++ (MinkL++) (Komorowski, Wysoczańska, and Trzcinski 2021), MinkLoc-v2 (MinkLv2) (Komorowski 2022), SVT-Net (Fan et al. 2022), PPT-Net (Hui et al. 2021), LWR-Net (Zhang et al. 2023), ERI-Net (Weng, Zhang, and Li 2022), RPR-Net (Fan, Song, Zhang, et al. 2023), VNI-Net (Tian et al. 2023), and RIA-Net (Hao, Zhang, and Su 2024). All methods were uniformly trained to generate 256-dimensional global descriptors. Evaluations utilized authors’ publicly released code and pre-trained models where available. For PPT-Net, model re-training was conducted owing to unavailability of pre-trained weights. For ERI-Net, RPR-Net, VNI-Net, and RIA-Net, results were reproduced from prior studies under identical evaluation protocols, as original implementations were partially incomplete or inaccessible. The evaluation results on the Oxford RobotCar dataset (Table 2) demonstrate that most existing methods—excluding those explicitly designed for rotation invariance (e.g., ERI-Net, RPR-Net)—suffer accuracy degradation under query point cloud rotations (within $\pm 30^\circ$). For instance, PNVLAD and PCAN exhibit $>60\%$ drops in Recall@1% under rotation. In contrast, LR-Mate achieves state-of-the-art performance in both rotated (Recall@1: 98.45%) and non-rotated (Recall@1: 98.52%) scenarios, confirming its rotation invariance and superiority.

Methods	Recall @1 (%)		Recall@1% (%)	
	With rotation	Without rotation	With rotation	Without rotation
PNVLAD	15.01	62.76	31.31	81.01
PCAN	14.81	69.05	32.81	83.81
EPC-Net	59.78	86.84	80.93	95.19
MinkL3D	74.26	93.48	90.14	97.85
MinkL++	80.79	93.90	92.75	98.15
MinkL-v2	82.68	96.25	95.15	98.87
PPT-Net	61.55	92.21	82.82	97.50
SVT-Net	63.52	93.70	84.11	97.80
SOE-Net	86.72	89.37	95.60	96.40
LWR-Net	40.41	93.76	62.67	98.32
ERI-Net	85.30	92.84	95.31	97.89
RPR-Net	81.10	81.00	92.20	92.20
VNI-Net	85.50	85.50	94.00	94.00
RIA-Net	89.74	91.23	96.22	96.84
LR-Net	88.75	89.00	96.17	96.19
LR-Mate	98.45	98.52	99.46	99.52

Table 2. Test on the Oxford RobotCar dataset with random rotation within 30°

As shown in Table 3, we compare model size and average runtime between LR-Mate and all baseline methods on the Oxford RobotCar dataset. While LR-Mate incurs elevated computational overhead (19ms average runtime compared to LR-Net’s 10ms baseline) attributable to its two-stage architecture (retrieval: 10ms; reranking: 6ms), its trigger mechanism (3ms latency, 96.77% error recall rate in this experiment) ensures operational efficiency. Critically, only 33.02% of queries activate the reranking phase, meaning 66.98% of queries resolve within 13ms (retrieval + trigger latency). This adaptive framework achieves an effective equilibrium between localization robustness and operational efficiency.

Methods	Average running time	Model size (million)
PNVLAD	15ms	19.78
PCAN	55ms	20.42
EPC-Net	26ms	4.70
MinkLoc3D	12ms	1.06
MinkLoc++	12ms	1.06
MinkLoc-v2	19ms	2.66
PPT-Net	22ms	13.39
SVT-Net	11ms	0.94
SOE-Net	22ms	19.40
LWR-Net	10ms	0.44
RPR-Net	238ms	1.10
VNI-Net	574ms	2.20
RIA-Net	/	18.39
LR-Net	10ms	0.46
LR-Mate	19ms	0.97

Table 3. Model size and average running time on Oxford RobotCar dataset

We further evaluated LR-Mate’s robustness against point cloud sparsity and outlier perturbations. As demonstrated in Figure 4, LR-Mate outperforms existing methods in maintaining recognition accuracy under varying density and noise conditions. Most current approaches, constrained by reliance on 256-channel global descriptors, exhibit sensitivity to noise interference and sparsity variations due to inherent dimensionality constraints. In contrast, LR-Mate optimizes localization results through the retrieval candidates, enhanced by trigger and re-ranking mechanisms, which decouples robustness from descriptor compactness, enabling reliable performance in degraded environments.

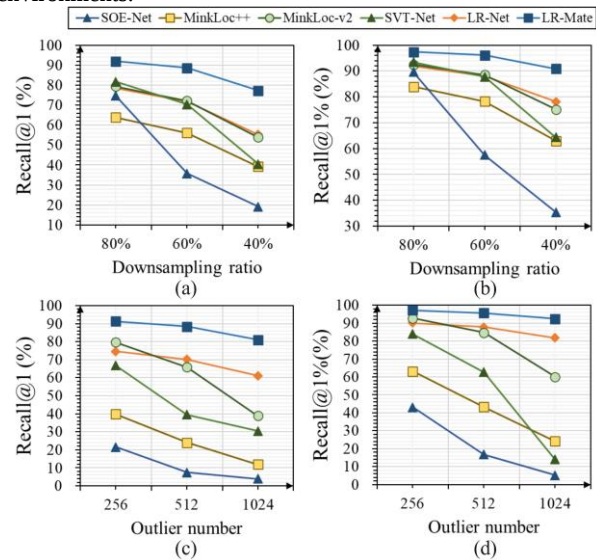


Figure 4. Robustness to the point cloud noise and sparsity on the Oxford RobotCar dataset: (a) Average recall @1 changes with point cloud sparsity; (b) Average recall @1% changes with point cloud sparsity; (c) Average recall @1 changes with point cloud noise; (d) Average recall @1% changes with point cloud noise.

To verify the generalization capability of each method, we evaluated models trained on the Oxford RobotCar dataset directly on the NUS In-house datasets. As shown in Table 4, LPR-Mate achieves state-of-the-art performance across all three scenes, outperforming recent methods like RIA-Net and VNI-Net. While the LR-Net (LR-Mate’s backbone) exhibits weaker generalization (e.g., 83.74% Recall@1 in U.S. vs. RIA-Net’s 88.21%), the trigger and reranking components bridge this gap, enhancing Recall@1 by 5.10% (U.S.), 4.61% (R.A.), and 5.62%

(B.D.). This underscores the paradigm’s ability to refine initial candidates and mitigate overfitting to training data.

Methods	U.S.		R.A.		B.D.	
	R@1 (%)	R@1% (%)	R@1 (%)	R@1% (%)	R@1 (%)	R@1% (%)
PNVLAD	18.16	33.76	15.17	28.42	17.39	26.12
PCAN	10.56	25.00	11.95	23.31	12.99	19.42
EPC-Net	60.10	80.54	51.89	70.77	57.48	69.44
MinkL3D	57.17	68.27	44.88	60.54	46.13	63.79
MinkL++	65.91	81.83	57.11	76.20	65.73	75.94
MinkLv2	61.47	80.80	47.37	64.38	63.83	72.99
PPT-Net	42.53	69.08	40.59	61.56	54.83	68.14
SVT-Net	53.47	69.00	52.28	68.34	67.44	78.76
SOE-Net	77.32	90.71	76.08	89.78	76.35	81.32
LWR-Net	55.02	66.04	45.22	60.88	56.08	68.15
RPR-Net	83.80	93.80	83.30	91.30	80.40	86.80
VNI-Net	85.30	95.00	83.30	91.50	81.40	86.80
RIA-Net	88.21	96.58	86.36	92.54	83.46	89.03
LR-Net	83.74	95.79	82.16	91.53	79.80	86.22
LR-Mate	88.84	96.92	86.77	94.85	85.42	92.27

Table 4. Test on NUS In-house datasets using the model trained on the Oxford RobotCar dataset

4.4 Experiments on raw point cloud dataset

Due to hardware constraints imposed by the large-scale raw point clouds, many existing methods face challenges in directly utilizing such data for training and testing. Consequently, we adopt lightweight models—MinkLoc3D, MinkLoc-v2, SVT-Net, and LR-Net—for comparative analysis. As illustrated in Table 5 (where R@1 denotes Recall@1 and R@1% denotes Recall@1%), existing methods exhibit varying degrees of accuracy degradation when applied to raw point clouds. This decline arises primarily from the abundance of low-texture redundant features, particularly from ground points, which weaken the discriminative capacity of global descriptors aggregated by conventional frameworks. In contrast, LR-Mate achieves superior performance in unprocessed raw point cloud scenarios by integrating global feature retrieval with local region comparisons to enhance scene matching precision. This improvement highlights LR-Mate’s efficacy in optimizing scene recognition methodologies under complex environmental conditions.

Methods	DCC		KAIST		RiverSide	
	R@1 (%)	R@1% (%)	R@1 (%)	R@1% (%)	R@1 (%)	R@1% (%)
MinkL3D	16.06	39.91	17.95	43.38	15.04	40.65
SVT-Net	27.71	71.05	29.35	80.23	19.46	58.88
MinkLv2	31.61	80.35	37.96	90.81	24.18	55.30
LR-Net	86.39	97.79	81.85	95.20	61.95	84.05
LR-Mate	95.61	99.95	93.13	96.80	75.90	85.83

Table 4. Test on original MulRan dataset

To further evaluate the performance of existing methods in long-term LPR scenarios, we conducted experiments using the recently updated MulRan dataset, which incorporates 4-year longitudinal data featuring structural environmental changes and sensor configuration shifts (from OS1-64 to OS2-128 LiDAR sensors). As shown in Table 5, all benchmarked methods exhibit substantial performance degradation under these challenging conditions. For instance, MinkLoc-v2’s Recall@1% at KAIST drops from 90.81% to 34.99%, whereas the proposed method retains relatively robust accuracy. This resilience is attributed to the LPR-Mate’s ability to correct temporal misalignments via local feature comparison, even when initial candidates from LR-Net are suboptimal. The results underscore the LPR-Mate’s adaptability to environmental and sensor variations, though

performance gaps in long-term scenarios indicate room for improvement.

Methods	DCC		KAIST		RiverSide	
	R@1 (%)	R@1% (%)	R@1 (%)	R@1% (%)	R@1 (%)	R@1% (%)
MinkL3D	0.67	18.01	4.72	31.43	2.28	25.94
SVT-Net	4.55	40.53	4.23	51.42	2.99	31.39
MinkLv2	12.12	44.59	7.71	34.99	2.63	27.12
LR-Net	21.10	70.96	6.56	42.46	9.48	46.14
LR-Mate	60.78	72.03	41.06	44.89	36.31	45.75

Table 5. Test on long-term MulRan dataset

4.5 Ablation study

Given that LPR-Mate utilizes a non-learning trigger stage, the evaluation focuses on the contributions of distinct learning branches within its reranking network. To quantify their individual impacts, three ablated variants of the re-ranking subnetwork were generated: (1) LPRMate-WoLF (excluding the local feature enhancement branch), (2) LPRMate-WoGF (excluding the global feature enhancement branch), and (3) LPRMate-WoSC (excluding the SC-score fusion branch). As summarized in Table 6, removing any branch reduced overall performance, with the most pronounced degradation observed for LPRMate-WoLF, underscoring the critical role of local feature enhancement in reranking optimization.

Model	Recall@1 (%)	Recall@1% (%)
LPR-Mate	98.45	99.46
LPRMate-WoLF	90.90	96.31
LPRMate-WoGF	96.66	99.31
LPRMate-WoSC	93.73	98.11

Table 6. Ablation study on the Oxford RobotCar dataset

4.6 Discussion

The experimental results demonstrate that LPR-Mate offers several advantages. First, LPR-Mate can adapt to diverse existing LPR methods—whether local-region-based or sparse-voxelization-based—and effectively supports optimization. Second, the non-learning-based trigger phase accurately assesses initial retrieval correctness, enabling non-blind optimization by activating reranking only when necessary, thereby improving efficiency. Third, the reranking network effectively addresses challenging scenarios (e.g., point cloud rotation, noise, sparsity, or temporal variations), achieving state-of-the-art retrieval accuracy in complex environments. Furthermore, LPR-Mate operates as a plug-and-play module: it directly utilizes local features and global descriptors from existing LPR networks without requiring architectural or training adjustments, ensuring seamless integration with pre-trained models. Figure 5 illustrates LPR-mate’s performance under rotational perturbations, noise, sparsity, and raw point clouds. While global descriptor-based initial retrieval may produce structurally similar false candidates under adverse conditions, LPR-Mate enables accurate final matching through local feature comparison.

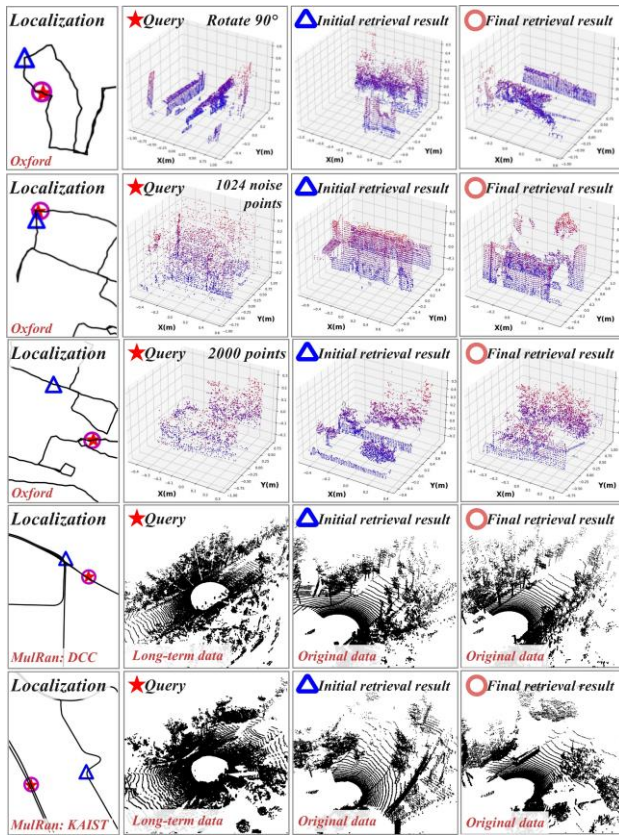


Figure 9. Retrieval and localization results of LPR-Mate under the influence of point cloud rotation, noise, sparsity, and raw point cloud.

Despite these strengths, LPR-Mate has limitations. First, its reranking efficacy depends on the initial retrieval candidates; if correct scenes are absent from these candidates (e.g., in extreme environments), reranking cannot recover them. Expanding the candidate pool mitigates this issue. Second, storing local region features increases storage demands compared to conventional LPR methods. Future work could explore feature compression techniques to reduce memory overhead while preserving optimization performance.

5. Conclusions

In this study, we propose LPR-Mate, a universal post-optimization method designed to enhance localization performance in challenging scenarios. LPR-Mate processes the top-k retrieval candidates' information ($k=25$ in this work) generated by existing LPR methods through a two-stage pipeline: trigger phase and reranking phase. The trigger phase evaluates candidate spatial consistency across local regions without requiring training, thereby enabling selective activation of reranking. When initial retrievals are accurate, reranking remains inactive to preserve computational efficiency. In cases requiring refinement, the reranking phase integrates local features, global descriptors, and spatial consistency scores. A group-transformer architecture facilitates local-region feature interaction, while channel attention dynamically reweights global descriptor channels to optimize candidate rankings. Extensive experiments validate LPR-Mate's state-of-the-art robustness against point cloud rotation, noise, sparsity, and long-term environmental variations. As a plug-and-play post-optimization module, LPR-Mate seamlessly integrates with existing LPR pipelines—including both local-region and sparse-voxelization-based approaches—without requiring retraining or structural modifications.

Despite these advancements, certain limitations persist. First, LPR-Mate's efficacy depends on the quality of initial retrieval candidates; scenarios lacking correct candidates within the input pool may compromise reranking efficacy. Second, the storage overhead for local features poses practical constraints. Future work could investigate lightweight alternatives, such as compressed local feature representations or enhanced spatial consistency metrics, to reduce memory demands. Additionally, integrating LPR-Mate into SLAM systems could further demonstrate its applicability in real-world robotic navigation tasks.

Acknowledgements

This research was funded by the Natural Science Foundation of Jiangsu Province (Grant No.BK20241646), the Fundamental Research Funds for the Central Universities (Grant No. 2024QN11082), the China Postdoctoral Science Foundation (Grant No. 2022M723377), and the National Natural Science Foundation of China (Grant No. 42394065, 42394060).

References

- Avrithis, Yannis, and Giorgos Tolias. 2014. "Hough Pyramid Matching: Speeded-Up Geometry Re-ranking for Large Scale Image Retrieval." *International Journal of Computer Vision* 107 (1):1-19. <https://doi.org/10.1007/s11263-013-0659-3>.
- Barros, T., L. Garrote, M. Aleksandrov, C. Prenebida, and U. J. Nunes. 2023. TRer: A Lightweight Transformer Re-Ranking Approach for 3D LiDAR Place Recognition. Paper presented at the 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), 24-28 Sept. 2023.
- Fan, Z., Z. Song, H. Liu, and J. He. 2023. GIDP: Learning a Good Initialization and Inducing Descriptor Post-enhancing for Large-scale Place Recognition. Paper presented at the 2023 IEEE International Conference on Robotics and Automation (ICRA), 29 May-2 June 2023.
- Fan, Zhaoxin, Zhenbo Song, Hongyan Liu, Zhiwu Lu, Jun He, and Xiaoyong Du. 2022. "SVT-Net: Super Light-Weight Sparse Voxel Transformer for Large Scale Place Recognition." *Proceedings of the AAAI Conference on Artificial Intelligence* 36 (1):551-560. <https://doi.org/10.1609/aaai.v36i1.19934>.
- Fan, Zhaoxin, Zhenbo Song, Wenping Zhang, Hongyan Liu, Jun He, and Xiaoyong Du. 2023. RPR-Net: A Point Cloud-Based Rotation-Aware Large Scale Place Recognition Network. Paper presented at the Computer Vision – ECCV 2022 Workshops, Cham, 2023//.
- Hao, W., W. Zhang, and H. Su. 2024. "RIA-Net: Rotation Invariant Aware 3D Point Cloud for Large-Scale Place Recognition." *IEEE Robotics and Automation Letters* 9 (6):5014-5021. <https://doi.org/10.1109/LRA.2024.3384887>.
- Hui, L., M. Cheng, J. Xie, J. Yang, and M. M. Cheng. 2022. "Efficient 3D Point Cloud Feature Learning for Large-Scale Place Recognition." *IEEE Transactions on Image Processing* 31:1258-1270. <https://doi.org/https://doi.org/10.1109/TIP.2021.3136714>.
- Hui, L., H. Yang, M. Cheng, J. Xie, and J. Yang. 2021. Pyramid Point Cloud Transformer for Large-Scale Place Recognition. Paper presented at the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 10-17 Oct. 2021.
- Jung, Minwoo, Woosong Yang, Dongjae Lee, Hyeonjae Gil, Giseop Kim, and Ayoung Kim. 2023. "HeLiPR: Heterogeneous LiDAR Dataset for inter-LiDAR Place

- Recognition under Spatial and Temporal Variations." *ArXiv* abs/2309.14590. <https://doi.org/10.48550/arXiv.2309.14590>.
- Kim, G., Y. S. Park, Y. Cho, J. Jeong, and A. Kim. 2020. MulRan: Multimodal Range Dataset for Urban Place Recognition. Paper presented at the 2020 IEEE International Conference on Robotics and Automation (ICRA), 31 May-31 Aug. 2020.
- Komorowski, J. 2021. MinkLoc3D: Point Cloud Based Large-Scale Place Recognition. Paper presented at the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 3-8 Jan. 2021.
- . 2022. Improving Point Cloud Based Place Recognition with Ranking-based Loss and Large Batch Training. Paper presented at the 2022 26th International Conference on Pattern Recognition (ICPR), 21-25 Aug. 2022.
- Komorowski, J., M. Wysoczańska, and T. Trzcinski. 2021. MinkLoc++: Lidar and Monocular Image Fusion for Place Recognition. Paper presented at the 2021 International Joint Conference on Neural Networks (IJCNN), 18-22 July 2021.
- Lehtola, Ville V., Juho-Pekka Virtanen, Matti T. Vaaja, Hannu Hyypä, and Andreas Nüchter. 2016. "Localization of a mobile laser scanner via dimensional reduction." *ISPRS Journal of Photogrammetry and Remote Sensing* 121:48-59. <https://doi.org/https://doi.org/10.1016/j.isprsjprs.2016.09.004>.
- Lin, T. Y., P. Goyal, R. Girshick, K. He, and P. Dollár. 2017. Focal Loss for Dense Object Detection. Paper presented at the 2017 IEEE International Conference on Computer Vision (ICCV), 22-29 Oct. 2017.
- Liu, Z., S. Zhou, C. Suo, P. Yin, W. Chen, H. Wang, H. Li, and Y. Liu. 2019. LPD-Net: 3D Point Cloud Learning for Large-Scale Place Recognition and Environment Analysis. Paper presented at the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 27 Oct.-2 Nov. 2019.
- Luo, L., S. Zheng, Y. Li, Y. Fan, B. Yu, S. Y. Cao, J. Li, and H. L. Shen. 2023. BEVPlace: Learning LiDAR-based Place Recognition using Bird's Eye View Images. Paper presented at the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 1-6 Oct. 2023.
- Maddern, Will, Geoffrey Pascoe, Chris Linegar, and Paul Newman. 2017. "1 year, 1000 km: The Oxford RobotCar dataset." *The International Journal of Robotics Research* 36 (1):3-15. <https://doi.org/10.1177/0278364916679498>.
- Radenović, F., G. Tolas, and O. Chum. 2019. "Fine-Tuning CNN Image Retrieval with No Human Annotation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41 (7):1655-1668. <https://doi.org/10.1109/TPAMI.2018.2846566>.
- Shao, S., K. Chen, A. Karpur, Q. Cui, A. Araujo, and B. Cao. 2023. Global Features are All You Need for Image Retrieval and Reranking. Paper presented at the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 1-6 Oct. 2023.
- Stathouloupoulos, N., M. A. V. Saucedo, A. Koval, and G. Nikolakopoulos. 2024. RecNet: An Invertible Point Cloud Encoding through Range Image Embeddings for Multi-Robot Map Sharing and Reconstruction. Paper presented at the 2024 IEEE International Conference on Robotics and Automation (ICRA), 13-17 May 2024.
- Tan, F., J. Yuan, and V. Ordonez. 2021. Instance-level Image Retrieval using Reranking Transformers. Paper presented at the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 10-17 Oct. 2021.
- Tian, Geng, Junqiao Zhao, Yingfeng Cai, Fenglin Zhang, Wenjie Mu, and Chen Ye. 2023. "VNI-Net: Vector Neurons-based Rotation-Invariant Descriptor for LiDAR Place Recognition." *ArXiv* abs/2308.12870.
- Uy, M. A., and G. H. Lee. 2018. PointNetVLAD: Deep Point Cloud Based Retrieval for Large-Scale Place Recognition. Paper presented at the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18-23 June 2018.
- Vidanapathirana, K., P. Moghadam, S. Sridharan, and C. Fookes. 2023. "Spectral Geometric Verification: Re-Ranking Point Cloud Retrieval for Metric Localization." *IEEE Robotics and Automation Letters* 8 (5):2494-2501. <https://doi.org/10.1109/LRA.2023.3255560>.
- Weng, S., R. Zhang, and G. Li. 2022. ERINet: Effective Rotation Invariant Network for Point Cloud based Place Recognition. Paper presented at the 2022 IEEE International Conference on Visual Communications and Image Processing (VCIP), 13-16 Dec. 2022.
- Xia, Y., Y. Xu, S. Li, R. Wang, J. Du, D. Cremers, and U. Stilla. 2021. SOE-Net: A Self-Attention and Orientation Encoding Network for Point Cloud based Place Recognition. Paper presented at the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 20-25 June 2021.
- Xu, Dong, Jingbin Liu, Yifan Liang, Xuanfan Lv, and Juha Hyypä. 2022. "A LiDAR-based single-shot global localization solution using a cross-section shape context descriptor." *ISPRS Journal of Photogrammetry and Remote Sensing* 189:272-288. <https://doi.org/https://doi.org/10.1016/j.isprsjprs.2022.05.005>.
- Xu, Tian-Xing, Yuan-Chen Guo, Zhiqiang Li, Ge Yu, Yu-Kun Lai, and Song-Hai Zhang. 2021. "TransLoc3D: Point cloud based large-scale place recognition using adaptive receptive fields." *arXiv preprint arXiv:2105.11605*.
- Yu, Shangshu, Cheng Wang, Zenglei Yu, Xin Li, Ming Cheng, and Yu Zang. 2021. "Deep regression for LiDAR-based localization in dense urban areas." *ISPRS Journal of Photogrammetry and Remote Sensing* 172:240-252. <https://doi.org/https://doi.org/10.1016/j.isprsjprs.2020.12.013>.
- Zhang, W., and C. Xiao. 2019. PCAN: 3D Attention Map Learning Using Contextual Information for Point Cloud Based Retrieval. Paper presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 15-20 June 2019.
- Zhang, Zhenghua, Guoliang Chen, Mingcong Shu, and Xuan Wang. 2023. "LWR-Net: Robust and Lightweight Place Recognition Network for Noisy and Low-Density Point Clouds." *Sensors* 23 (21):8664.
- Zhang, Zhenghua, Hu Liu, Xuan Wang, Mingcong Shu, Guoliang Chen, and Qiuzhao Zhang. 2024. "Lightweight and rotation-invariant place recognition network for large-scale raw point clouds." *ISPRS Journal of Photogrammetry and Remote Sensing* 212:58-72. <https://doi.org/https://doi.org/10.1016/j.isprsjprs.2024.04.030>.