LiDAR-Based Place Recognition For Autonomous Driving: A Survey

YONGJUN ZHANG, Wuhan University, Wuhan, China
PENGCHENG SHI, Wuhan University, Wuhan, China
JIAYUAN LI, Wuhan University, Wuhan, China

LiDAR has gained popularity in autonomous driving due to advantages like long measurement distance, rich three-dimensional information, and stability in harsh environments. Place Recognition (PR) enables vehicles to identify previously visited locations despite variations in appearance, weather, and viewpoints, even determining their global location within prior maps. This capability is crucial for accurate localization in autonomous driving. Consequently, LiDAR-based Place Recognition (LPR) has emerged as a research hotspot in robotics. However, existing reviews predominantly concentrate on Visual Place Recognition, leaving a gap in systematic reviews on LPR. This article bridges this gap by providing a comprehensive review of LPR methods, thus facilitating and encouraging further research. We commence by exploring the relationship between PR and autonomous driving components. Then, we delve into the problem formulation of LPR, challenges, and relations to previous surveys. Subsequently, we conduct an in-depth review of related research, which offers detailed classifications, strengths and weaknesses, and architectures. Finally, we summarize existing datasets and evaluation metrics and envision promising future directions. This article can serve as a valuable tutorial for newcomers entering the field of place recognition. We plan to maintain an up-to-date project on https://github.com/ShiPC-AI/LPR-Survey.

CCS Concepts: • **General and reference** → **Surveys and overviews**; • **Computer systems organization** → **Robotic autonomy**; • **Networks** → *Network performance analysis*; • **Theory of computation** → *Computational geometry;*

Additional Key Words and Phrases: LiDAR-based place recognition, robotics, autonomous driving, simultaneous localization and mapping, loop closure detection, map localization

ACM Reference Format:
Yongjun Zhang, Pengcheng Shi, and Jiayuan Li. 2024. LiDAR-Based Place Recognition For Autonomous Driving: A Survey. *ACM Comput. Surv.* 57, 4, Article 106 (December 2024), 36 pages. https://doi.org/10.1145/3707446

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 42030102 and 42271444, and the Science and Technology Major Project of Hubei Province under Grant 2021AAA010.
Authors' Contact Information: Yongjun Zhang, Wuhan University, Wuhan, Hubei, China; e-mail: zhangyj@whu.edu.cn; Pengcheng Shi (Corresponding author), Wuhan University, Wuhan, Hubei, China; e-mail: shipc_2021@whu.edu.cn; Jiayuan Li (Corresponding author), Wuhan University, Wuhan, Hubei, China; e-mail: ljy_whu_2012@whu.edu.cn.

ACM 0360-0300/2024/12-ART106
https://doi.org/10.1145/3707446

1 Introduction

1.1 Background

In recent years, autonomous driving has rapidly advanced with applications in logistics, public transportation, food delivery, warehousing, and medical rescue. The autonomous driving system integrates complex modules such as sensors, perception, localization, planning, control, and communication. These modules employ techniques like computer vision, sensor technology, and data processing to enable vehicle autonomy. Location information is a prerequisite for the entire system, helping vehicles understand their environment for better navigation. Notably, **Place Recognition (PR)** [1–5] is a vital technique for obtaining and expressing this information.

As shown in Figure 1, we summarize PR's functions into two aspects. (1) It addresses the problem of "*where have I ever been*," also known as **Loop Closure Detection (LCD)** [6–8]. When the robot explores unknown environments, PR compares newly captured sensor data with historical data to identify revisited locations. The loop closure data are then passed to the backend system to establish constraints and mitigate pose drifts. In this context, PR and localization are interdependent, with PR enhancing localization accuracy through loop closure detection and error correction. (2) It tackles the issue of "*where am I*," also known as global localization [9, 10]. When the vehicle travels within a predefined map, PR correlates newly captured sensor data with maps to pinpoint the vehicle's locations. Subsequently, the planning module utilizes this location to optimize path prediction, while the control module generates corresponding vehicle commands. In this context, PR is a specialized localization method that directly provides the vehicle's global pose.

To date, place recognition remains a challenging and ongoing research problem. Traditional **Global Positioning System (GPS)** [11] provides absolute positions but is easily blocked in underground parking lots and canyons, limiting their effectiveness. **Inertia Measurement Unit (IMU)** [12] offers high-frequency attitude data but suffers from accuracy degradation over time due to cumulative errors. Wheel odometers [13] provide stable displacement information, but slippery roads or sudden acceleration can distort data. In contrast, measurement sensors like cameras and LiDAR provide high-resolution environmental information with frequent updates. Cameras [14] excel in capturing visual details like color, shape, and texture but are susceptible to illumination changes and a narrow **Field Of View (FOV)**. In contrast, LiDAR [15] generates rich three-dimensional (3D) data and operates reliably in harsh conditions (e.g., night, rain, and fog), signal-limited environments, and complex terrain. These attributes greatly enhance LiDAR's suitability for autonomous driving applications, leading to growing interest in **LiDAR-based Place Recognition (LPR)**.

1.2 Relation to Autonomous Driving

1.2.1 Overview of Autonomous Driving Components. Autonomous driving [16] enables vehicles to execute tasks independently using sensors, computer systems, and artificial intelligence, managing acceleration, braking, and steering while ensuring safety and efficiency. The American Society of Autonomous Driving Engineers classifies autonomous driving into six levels [17], ranging from full human control (L0) to complete automation (L5). The key to vehicle automation and intelligence is ensuring the core system components effectively interact and collaborate. The industry and academia generally adopt a modular approach, viewing autonomous driving systems as extensions of mobile robot architectures [18], including perception, localization, planning, control, and **Human–Machine Interface (HMI)**.

Perception. The perception module [19] acquires real-time environmental data through multiple sensors. It performs target detection [20], classification [21], and tracking [22] to provide high-precision environmental information for localization, planning, and control. It identifies traffic signs, lanes, and lights to evaluate road conditions. Additionally, it detects and tracks obstacles to determine their proximity to maintain a safe driving distance.

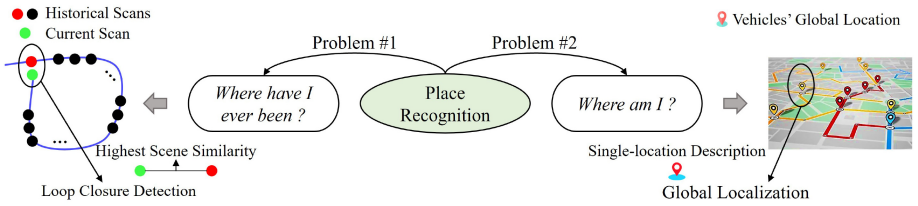


Fig. 1. Two key problems addressed by PR. On the left, the blue line denotes the vehicle trajectory, with solid circles indicating collected scans by sensors over time. The green circle marks the current scan, while the others represent past scans. The green and red circles are geographically close, and their scans exhibit the highest scene similarity, forming a closed loop. On the right, the black-marked area shows a vehicle's global location, only providing a single-location description within maps.

Localization. The localization module [23] determines the vehicle's position to support navigation, decision, and control. It typically relies on multiple sensors, including LiDAR, cameras, GPS, and IMU, to manage complex scenarios. GPS provides location data in open areas, while **High-Definition (HD)** maps [24] offer detailed road information in GPS-denied environments, helping anticipate hazards like sharp turns and steep slopes. IMU [12] measures acceleration and angular velocity to help infer the vehicle's motion in high-speed scenarios.

Planning. The planning module [25] connects perception, localization, and control, which employs path optimization and decision-making technologies to develop smooth and safe driving routes in complex environments. It establishes a global path from the start to the destination based on criteria such as shortest distance or minimum energy consumption. Simultaneously, it dynamically adjusts the local path using real-time perception data and vehicle location to avoid obstacles and address emergencies.

Control. The control module [26], interfacing directly with the vehicle, employs perception data, vehicle location, and planning instructions to generate operational commands. It controls acceleration, braking, and steering to guide vehicles to follow the predefined and dynamically adjusted paths, ensuring smooth and safe driving. Additionally, it receives perception data and location information to manage honking and lighting to alert pedestrians and other vehicles to avoid accidents.

Human–Machine Interface. The HMI module [27] facilitates communication between users and vehicles, enhancing the transparency of autonomous driving systems and user experience. It enables interactions through voice, touch, or gesture, allowing users to convey instructions and understand system responses. By providing intuitive interfaces and feedback, HMI helps users grasp system status, warnings, and decisions, offering clear guidance at critical moments to ensure safety and reliability in diverse scenarios.

1.2.2 Relationship between PR and Each Component. As illustrated in Figure 2, we summarize the relationship between PR and autonomous driving components:

High-Precision Localization. In autonomous driving systems with HD maps, PR is typically viewed as a component of the localization module. It integrates sensor data [28], perceived landmarks [29], and HD map elements [30] to accurately pinpoint the vehicle's location, even in GPS-denied areas. HD maps, in turn, use the vehicle's location and sensor data to monitor environmental changes and update the map if significant alterations are detected.

Environment Perception Support. PR and perception modules exchange information and feedback with each other, facilitating data fusion and error correction to enhance perception and localization accuracy [31]. The perception identifies landmarks like lanes and traffic signs,

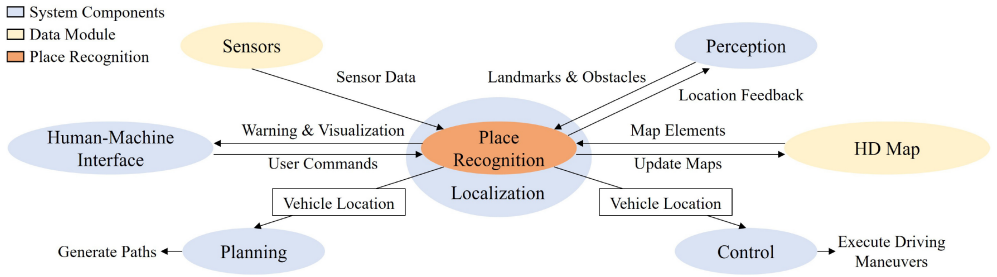


Fig. 2. The relationship between PR and key components of autonomous driving. HD maps, being external information, are grouped with sensors under data modules.

supplying key features for PR to use in localization. PR offers location feedback to the perception, comparing real-time data with maps to correct perception errors.

Path Planning Foundation. The precise location data from PR aids the planning module in selecting the optimal path based on current road conditions and adjusting it for real-time traffic changes [32]. It also supports precise obstacle avoidance and lane-keeping in complex scenarios, ensuring safe vehicle operation.

Accurate Control Assistance. PR provides the control module with crucial location data, enabling precise control commands and real-time adjustments [33]. These data help the module determine the vehicle's position and direction, allowing accurate acceleration, braking, and steering to follow the planned path. If deviating from the trajectory, then the control module swiftly corrects the course to keep vehicles on the correct route.

Enhanced Human–Machine Interactivity. PR aids the HMI in showing the vehicle's precise location, route, and destination, which enhances user understanding of the vehicle's status [34]. In abnormal situations, HMI employs PR's feedback to issue warnings or provide suggestions, which helps users deal with current driving conditions. Users can enter destinations, preferences, and modes through HMI, and PR updates navigation and optimizes strategies based on real-time data, improving the user experience.

Therefore, PR is crucial in autonomous driving, supporting localization, perception, planning, control, and HMI modules by providing accurate vehicle location data. It boosts localization accuracy, enables the system to adapt to complex environments, ensures efficient and safe vehicle operation, and improves user experience. Effective collaboration among these modules creates a comprehensive solution that allows vehicles to navigate safely in diverse and dynamic scenarios.

1.3 Contributions

In this article, we present a comprehensive review of LPR research, accompanied by a detailed methodological taxonomy depicted in Figure 3. We categorize methods into handcrafted and learning-based types, further subdividing them, and present detailed introductions to pioneering works. This well-organized layout enhances reading efficiency and facilitates a better understanding of the relevant technologies in place recognition. Our main contributions are as follows:

— **Filling Survey Gap.** Existing reviews focus on Visual Place Recognition or discuss general place recognition issues with little emphasis on LiDAR. Our dedicated survey on LPR bridges this gap, helping researchers understand and disseminate **State-of-the-Art (SOTA)** LPR methods.

— **Detailed Definition and Method Classification.** We offer a comprehensive definition of place recognition, including implicit loop closure detection and explicit global localization.

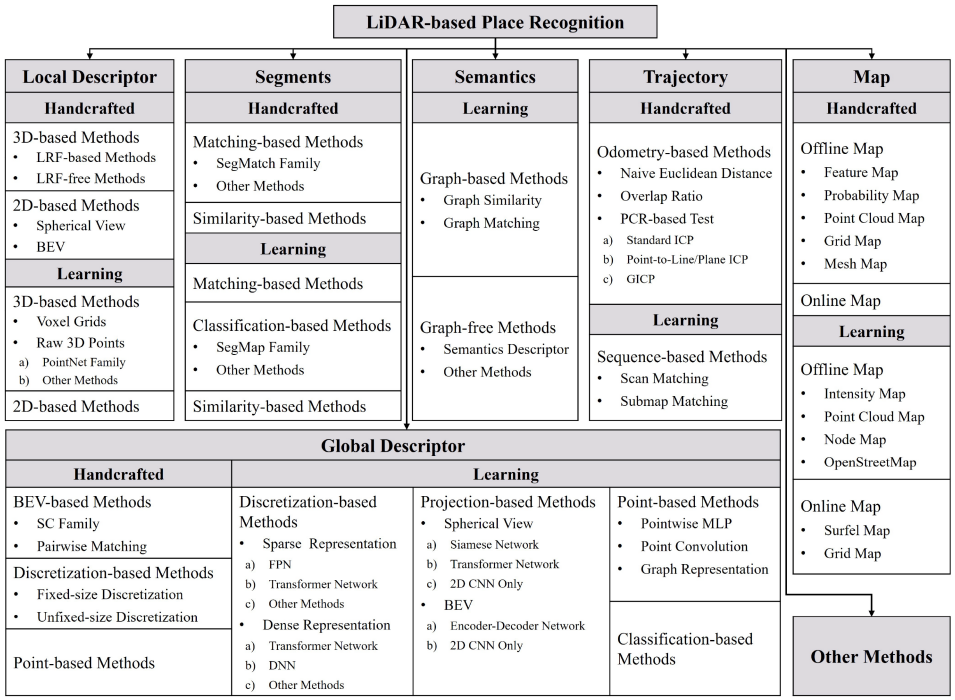


Fig. 3. Proposed taxonomy of LPR methods.

The proposed taxonomy of LPR methods includes seven parts, each split into handcrafted and learning-based branches, with further subcategories based on specific rules. We also outline their advantages and disadvantages. The efforts may help researchers understand the applicability of methods and inspire the development of place recognition techniques for challenging scenarios like docks, parks, and woodlands.

— **Future Direction.** We propose promising directions for advancing LPR research: innovative solutions like cloud computing, quantum technology, bionic localization, and applications in space exploration, polar research, and underwater robotics. These areas remain unexplored in previous surveys. We also highlight multimodal information sources such as WiFi, voice, and **Radio Frequency Identification (RFID)**, along with advanced sensors like solid-state lidar, event cameras, and millimeter-wave radar. Additionally, we recommend evaluating methods based on scale, efficiency, long-term performance, and developing standard datasets.

— **Open source Project.** We maintain an up-to-date project on our website, allowing the robotics community to stay current with SOTA LPR technologies. This resource helps newcomers quickly grasp essential information such as datasets, evaluation metrics, and mainstream methods.

2 Definition and Challenges

2.1 Problem Formulation

As summarized in Reference [35], PR has two prevalent definitions: overlap-based and distance-based. The overlap-based definition [36] emphasizes visual similarity, identifying two images depicting the same place if they show significant visual overlap. However, the distance-based

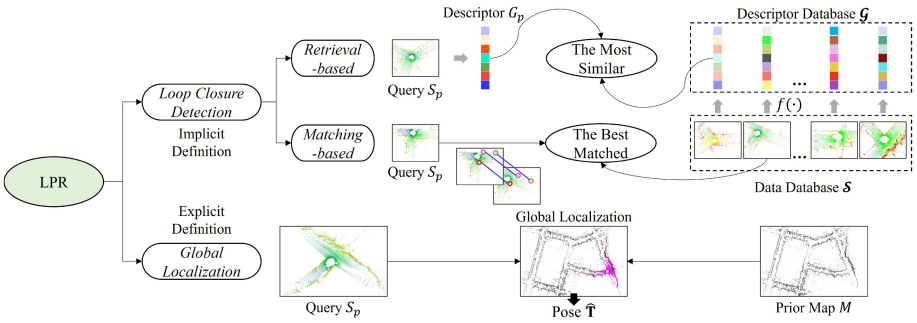


Fig. 4. Definition of LiDAR-based place recognition. We classify the definition into implicit loop closure detection and explicit global localization. Since global localization provides global vehicle poses, we term it explicit place recognition. In contrast, loop closure detection identifies revisited poses through frame-to-frame comparison without a global pose, making it implicit.

definition [30] relies on geographical proximity, defining places as identical if the distance between two vehicles is below a user-defined threshold. As our survey focuses on LiDAR sensors, we follow the distance-based definition. Differently, we present two definitions for LPR: implicit loop closure detection and explicit global localization. As shown in Figure 4, loop closure detection solves the revisit problem by determining if two locations are close based on similarity or matching ratio. Due to the absence of direct global poses, it is called implicit place recognition. In contrast, global localization provides the vehicle's poses within a pre-existing map, termed explicit place recognition.

2.1.1 Loop Closure Detection. We follow the definition in Reference [35]. Let S_p denote the current sensor data, L_p the current robot location, $\mathcal{L} = \{L_1, L_2, \ldots, L_k\}$ previously visited locations, and $\mathcal{S} = \{S_1, S_2, \ldots, S_k\}$ previously collected data, with subscripts indicating the data index. Given another sensor data $S_q \in \mathcal{S}$ collected at location $L_q \in \mathcal{L}$, a loop closure occurs when L_p and L_q are geographically close. Two geographically close locations typically have similar environmental layouts, which result in their sensor data with high matching ratios:

$$dist(L_p, L_q) < \lambda_l, \quad score(S_p, S_q) > \lambda_s, \tag{1}$$

where $dist(L_p, L_q)$ denotes the geographical distance between L_p and L_q, $score(S_p, S_q)$ means the matching ratio for S_p and S_q, λ_l, and λ_s are two user-defined thresholds based on a specific application. Identifying the closest location is then transformed into searching the data with the highest similarity score to the current data,

$$\widehat{L}_q = \underset{L_q \in \mathcal{L}}{\arg\min}\, dist(L_p, L_q) \Rightarrow \widehat{S}_q = \underset{S_q \in \mathcal{S}}{\arg\max}\, score(S_p, S_q), \tag{2}$$

where $\widehat{L}_q$ is the closest location and $\widehat{S}_q$ represents the most similar or best-matched sensor data.

Some researchers [4, 37–40] treat the above searching for best-matched data as a retrieval task. These methods typically encode sensor data into a global descriptor G and aggregate historical descriptors into a database $\mathcal{G}$. During retrieval, the current descriptor G_p is matched against the database $\mathcal{G}$ to find the most similar descriptor $\widehat{G}_q$:

$$\widehat{L}_q = \underset{L_q \in \mathcal{L}}{\arg\min}\, dist(L_p, L_q) \Rightarrow \widehat{G}_q = \underset{G_q \in \mathcal{G}}{\arg\min}\, \delta(G_p, G_q), G_p = f(S_p), G_q = f(S_q), \tag{3}$$

where $\delta(\cdot)$ denotes the descriptor distance, typically the L_2 norm, and $f(\cdot)$ represents a descriptor encoding process. Since encoding methods vary widely without a unified rule, we do not provide

a specific definition. Unlike image retrieval [41] in computer vision, this operates directly on point clouds and uses the geographical distance between vehicles as the metric.

2.1.2 Global Localization. Since our survey focuses on LiDAR sensors, our definition of global localization excludes GPS-based methods [42]. Let S_p represent the current sensor data and M represent a prior map. Global localization [43] establishes associations between S_p and M to pinpoint the vehicle's global pose:

$$\widehat{\mathbf{T}} = \arg\min_{\mathbf{T}} loss(\mathbf{T}, S_p, M), \tag{4}$$

where $\mathbf{T}$ is a transformation with six or three **Degree-of-Freedoms (DoF)**, *loss* denotes the loss function between the current sensor data S_p and the map data M calculated based on the pose $\mathbf{T}$, generally the point-to-point or point-to-surface distance.

2.2 Challenges of LPR

Numerous recent LPR methods have explored techniques like **Bird-Eye-View (BEV)** [40, 44], histograms [45, 46], image representations [47, 48], and graph theory [49, 50] to enhance performance. They achieve rotational invariance through height similarity [7, 40] and frequency domain analysis [44, 51]. Additional approaches employ pose proximity [52], sequence matching [53], and **Point Cloud Registration (PCR)** [54] techniques to improve recognition accuracy. However, these methods struggle in dynamic and highly occluded environments. Traditional methods rely on low-level features (coordinates [55], normals [56], intensities [57], and range [58]), while learning-based approaches gradually show promising results using neural networks [59], attention mechanisms [60], and semantics [61]. Furthermore, diverse map representations, such as point clouds [62], semantics [63], and mesh [64], have been successfully applied in map localization. Despite the impressive results claimed by these methods, several challenges persist that require further attention:

Motion Distortions. Vehicle motion inevitably distorts the point cloud, severely affecting feature matching and registration between scans [65]. Several methods [66, 67] employ a constant velocity motion model based on the previous pose to correct this. Although effective in most scenarios, this approach is inadequate for sudden direction changes [68] or rapid acceleration [69].

Viewpoint Differences. Lane-level horizontal deviations may exist when a robot revisits a historical place from different directions. While a few methods [7, 40, 48] address rotation invariance, they overlook the impact of translation on place recognition.

Weather Conditions. Laser signals exhibit varying behavior under different weather conditions [70]. They attenuate less and travel farther on sunny days but decay significantly in rainy and foggy weather.

Perceptual Aliasing. Distinct places in confined corridors [71, 72] may exhibit similar point cloud data, which introduces ambiguous interpretations.

Appearance Changes. Long-term navigation applications [29, 48] often involve significant environmental changes, which leads to potential failures.

Sensor Characteristics. Mechanical LiDAR [73] produces point clouds in a format of multiple scan lines, resulting in vertical sparsity. Solid-state LiDAR [74] provides limited horizontal FOV and requires specific considerations.

2.3 Relation to Previous Surveys

In recent years, there has been a proliferation of reviews on visual technologies, addressing various topics like place recognition [6, 36, 75, 76], localization [30, 77], tracking [78, 79], and SLAM [80–82]. Despite their valuable contributions to the progress of research in robotics and autonomous

driving, these reviews have unfortunately overlooked the technology of LiDAR. Cadena et al. [83] extensively reviewed the current state of SLAM and delved into potential future directions. Yin et al. [35] provided a comprehensive place recognition survey encompassing cameras, LiDAR, radar, and joint sensors. However, the section dedicated to discussing LiDAR was relatively limited. Yin et al. [84] offered an informative overview of the recent progress in LiDAR-based global localization, while it merely represented a specialized branch of place recognition.

Comparatively, our survey distinguishes itself through the following features: (1) It is the first survey dedicated solely to LPR research, filling a gap in the field and advancing the dissemination of SOTA LPR methods. (2) It offers a more comprehensive problem formulation, method classification, and summary, enhancing understanding of methodological advantages, disadvantages, and applicability. The efforts facilitates the development of LPR techniques, especially in challenging scenarios. (3) We envision promising future directions to advance LPR research, including innovations such as cloud computing, quantum technology, and bionic localization, as well as applications in space exploration, polar research, and underwater robotics. (4) We maintain an up-to-date project to keep the robotics community updated on SOTA LPR technologies and aid newcomers in quickly grasping essential information like datasets, evaluation metrics, and methodologies.

3 LPR Techniques: Local Descriptor

The local descriptor is a compact representation of regions or points, capturing distinctive characteristics such as texture, color, density, or shape. Local descriptor-based methods typically extract keypoints and employ local descriptors to characterize their surrounding context. They generally fall into either 3D-based or 2D-based categories based on the nature of descriptors. Table 1 contains a systematic summary.

3.1 Handcrafted Methods

3.1.1 3D-based Methods. We roughly categorized handcrafted 3D local descriptors into two groups based on a **Local Reference Frame (LRF)**.

LRF-based Methods. LRF rigidly transforms the patch into canonical representation by selecting neighborhood points to build a covariance matrix and computing the eigenvector as reference axes. While initially designed for PCR, it is also applicable to place recognition. Several methods focus on encoding geometric information, such as normal [57], height [92], and mesh [90], within the LRF to achieve precise geometric descriptions. Others enhance the stability of LRF using weighted projection vectors [91] or sign disambiguation [89].

LRF-free Methods. Other methods ensure rotation invariance by avoiding LRF construction and focusing solely on the underlying geometry of the local surface. Early methods directly count surrounding geometric information, such as height [55], distance [87], and density [85]. Subsequently, several methods encode point distribution [88] and sparse triangulated landmark [86].

3.1.2 2D-based Methods. These methods build handcrafted local descriptors from the projected 2D image, followed by an image matching problem. Spherical view and BEV are two representative projection approaches.

Spherical View. Projecting point clouds into a spherical or range image effectively mitigate orientation ambiguity. Steder et al. [93] pioneer the projection of point clouds to range images for place recognition. They extract the local descriptor vector [117] and evaluated candidate transformations through keypoint reprojection. Afterward, several works extend the method [93] using Normal Aligned Radial Features [8], Speeded Up Robust Features [94], and ORB [95, 96].

BEV. Several works incorporate proposal-wise features from BEV images into image matching. BVMatch [51] extract the maximum index map of the Log-Gabor filter responses, employing BEV

Table 1. A Summary of Local Descriptor-based Methods

Handcrafted Methods							
Methods			**Year**	**Feature**	**Size**	**Similarity Metric**	**Code**
3D-based Methods	LRF-free	Spin Image [85]	1999	Density	153	L2 Distance	✓
		3DGestalt [55]	2013	Height	32×10	Voting	
		NBLD [86]	2016	Density	16×4×8	Voting	
		GLAROT-3D [87]	2017	Orientation+Range	1880	Rotated L1 Norm	
		HoPPF [88]	2020	Angle+Distance	600		
	LRF-based	USC [89]	2010	Density	1960	Euclidean Distance	✓
		RoPS [90]	2013	Density	135	L2 Distance	✓
		TOLDI [91]	2017	Depth	3×20×20		
		ISHOT [57]	2019	Angle+Intensity	1344	Voting	✓
		Sun et al. [92]	2020	Height	20×20	Euclidean Distance	
2D-based Methods	Spherical View	Steder et al. [93]	2010	Range+Curvature		Euclidean Distance	
		Steder et al. [8]	2011	Range+Curvature	36	Manhattan Distance	
		Zhuang et al. [94]	2013	Space		Matching Score	
		Cao et al. [95]	2018	Position	600×391	L1-Norm	
		Shan et al. [96]	2021	Intensity	1024×128	L1 Distance+Hamming Distance	✓
	BEV	BVMatch [51]	2021	Density	6×6×6	2D Rigid Pose	✓
		HOPN [46]	2022	Normal+Density	6×6×6	2D Rigid Pose	✓
Learning-based Methods								
Methods			**Year**	**Backbone**	**Size**	**Loss**	**EtE**	**Code**
3D-based Methods	Voxel Grids	3DShapeNet [97]	2015	Convolutional BDN	24×24×24	Contrastive Divergence		✓
		VolumetricCNN [21]	2016	CNN	512	Classification	✓	
		3DMatch [98]	2017	3D ConvNet	512	Contrastive		✓
		3DSmoothNet [99]	2019	CNN	16	Batch Hard		✓
		SpinNet [100, 101]	2021	Transformer+3DCCN	32	Contrastive	✓	✓
	Raw 3D Points	PointNet [102]	2017	CNN	1024	Regularization Softmax		✓
		PointNet++ [103]	2017	PointNet		Cross Entropy	✓	✓
		CGF [104]	2017	DNN	32	Triplet		✓
		PPFNet [105]	2018	PointNet	64	N-tuple		✓
		PPF-FoldNet [106]	2018	MLP	512	N-tuple		✓
		DeepVcp [107]	2019	PointNet++	32	L1+L2	✓	✓
		RelativeNet [108]	2019	PPF-FoldNet		Chamfer	✓	
		L3Ds [109]	2022	TNet+PointNet	32	Contrastive		✓
		Poiesi et al. [110]	2022	QNet+PointNet++	32	Hardest-contrastive		✓
		LEAD [111]	2022	Spherical CNN	512	Chamfer Distance	✓	
2D-based Methods		LORAX [112]	2017	DNN	1032	Pixelwise Error+ICP		✓
		MVDesc [113]	2018	MatchNet	32	Double-margin Contrastive		
		Li et al. [114]	2020	CNN	32	Batch-hard triplet	✓	✓
		Gojcic et al. [115]	2020	FCGF	32	Hardest Contrastive	✓	✓
		DeLightLCD [116]	2022	DNN	1×300×32	Binary Cross Entropy	✓	

Size and **EtE** denotes the descriptor size and end-to-end learning, respectively.

feature transform and BoW for place recognition. It provides relative poses and effectively overcomes sparsity and intensity distortion. Luo et al. [46] apply FAST [118] detectors on the BEV image and construct a global descriptor using 3D normals. It showcases superior localization capability in large-scale scenarios.

3.2 Learning-based Methods

3.2.1 3D-based Methods. Learning-based 3D local descriptors typically employ 3D CNNs to encode point cloud patches, divided into voxel grids and raw 3D points according to network inputs.

Voxel Grids. The pioneering work 3DMatch [98] transform patches into voxel grids of Truncated Signed Distance Function and employed eight convolutional layers to learn the descriptor.

Several works extend this idea to more informative encoding manner such as binary occupancy [21], multi-label occupancy [97], smoothed density value [99], and spherical voxelization [100, 101].

Raw 3D Points. An alternative method involves direct processing of the raw point cloud data.

(a) PointNet Family. PointNet [102] is a pioneering work of learning from unordered point clouds, which learns a symmetry function approximated by a **Multi-Layer Perceptron (MLP)** to handle detailed shapes. Subsequently, several works extend PointNet [102] by ball query search [103], normals [105], FoldingNet decoder [106], orientation [108], LRF [109, 110], and semantics [107].

(b) Other Methods. **Compact Geometric Features (CGF)** [104] trains a deep network to map from the high-dimensional space of spherical histograms to a low-dimensional Euclidean space. LEAD [111] combines spherical CNNs to learn the equivariant representation.

3.2.2 2D-based Methods. Several works infer local descriptors using well-established 2D CNNs from projected 2D images. They demonstrate superior performance in the task of 3D shape recognition and retrieval. LORAX [112] and DeLightLCD [116] employ a **Deep Neural Network (DNN)** auto-encoder and attention to enhance descriptor descriptiveness on 2D depth images, respectively. Another spectrum of research fuses multi-view features into descriptors by soft-view pooling [114], graphical model [113], and spectral relaxation [115].

3.3 Observations and Implications

This section review some representative methods and more comprehensive surveys of local descriptors are available in Reference [119]. Although local descriptors find wide applications in tasks such as registration and object recognition, they are not the preferred methods for PR. There are mainly the following reasons:

(i) Viewpoint changes can affect the accuracy of 3D keypoints, rendering them unsuitable for matching. Moreover, they may not effectively handle data noise and object occlusions. (ii) The usage of 3D local descriptors [55, 86] can be challenging as it requires dense point clouds, which is computationally expensive and may not work well with sensors like Velodyne VLP-16 [58] that produce sparse point clouds. (iii) While converting point clouds into images [51, 93, 116] can use mature image processing techniques, this results in the loss of geometric information, making it unsuitable for large-scale scenarios.

4 LPR Techniques: Global Descriptor

The global descriptor captures the overall features of a scene, providing a holistic view of the data rather than focusing on specific regions or points. Table 2 contains a systematic summary of global descriptor-based methods.

4.1 Handcrafted Methods

4.1.1 BEV-based Methods. BEV projection gains significant attention in the robotics community due to its ability to enhance algorithm efficiency through dimension reduction, making it highly suitable for real-time applications. **Scan Context (SC)** [40] family and pairwise matching are two mainstream methods.

SC Family. The pioneering work SC [40] partitions the horizontal space into discrete bins while maintaining the points' maximum height to generate a 2D matrix descriptor. It utilizes the ring key to search for potential matches and conducts a columnwise comparison to identify the closest one. This method demonstrates promising performance but may fail when dealing with significant lateral offsets. Subsequently, researchers propose a series of SC-based variant methods,

Table 2. A Summary of Global Descriptor-based Methods

Handcrafted Methods							
Methods			**Year**	**Metric**	**Size**	**Feature**	**Code**
BEV	SC Family	SC [40]	2018	L0 Norm+Cosine Distance	60×20	Height	✓
		ISC [7]	2020	Cosine Distance	60×20	Intensity	✓
		SC++ [4]	2021	L1 Norm+Cosine Distance	60×20	Height	✓
		FreSCo [120]	2022	L1 Norm+Cosine Distance	20×120	Height	✓
		FSC [121]	2022	F-Norm		Height+Intensity	
		Ou et al. [122]	2023	Cosine Distance	5×20×60	Density+Height	
	Pairwise Matching	LiDAR Iris [44]	2020	Hamming Distance	80×360	Height	✓
		RING [123]	2022	Circular Cross-correlation	120×120	Occupancy	✓
		RING++ [124]	2022	Circular Cross-correlation	120×120	Height+Occupancy	✓
		PGHCI [24]	2022	JS Divergence+Pixel Values	40×20	Height	
Discretization	Fixed-size	Magnusson et al. [125]	2009	Euclidean Distance		Shape	
		Lin et al. [74]	2019	Normalized Cross-correlation	60×60	Shape	✓
		Cao et al. [126]	2021	Euclidean Distance	360×180	Context+Layout	
	Unfixed-size	DELIGHT [127]	2018	Chi-squared Test	256	Intensity	
		Mo et al. [128]	2020	L2-Norm+Chi-square Test		Density+Intensity+Height	✓
Point		Z-Projection [56]	2011	χ^2 Distance+Sørensen Distance	101	Normal	
		Fast Histogram [45]	2015	Wasserstein Metric	80	Height	
		M2DP [129]	2016	L2-Norm	192	Density	✓
		C-M2DP [130]	2019	L2 Distance	576	Color+Shape	
Learning-based Methods							
Methods			**Year**	**Backbone**	**Aggregator**	**Size**	**Loss**	**EtE**
Point	Pointwise MLP	Pointnetvlad [37]	2018	PointNet	NetVLAD	256	Lazy triplet and quadruplet	✓
		PCAN [38]	2019	PointNet	NetVLAD	256	Lazy quadruplet	✓
		SOE-Net [131]	2021	PointOE	NetVLAD	256	HPHN quadruplet	✓
		LCD-Net [132]	2022	PV-RCNN	NetVLAD	256	Triplet	✓
	Point Conv	DH3D [60]	2020	FlexConv+SE block	NetVLAD	256	N-tuple	✓
		EPC-Net [59]	2022	PPCNN	VLAD	256	Lazy quadruplet	✓
	Graph	LPD-Net [2]	2019	PointNet	NetVLAD	256	Lazy quadruplet	✓
		DAGC [39]	2020	ResGCN	NetVLAD	256	Lazy quadruplet	
		SR-Net [133]	2020	SGC+SAM	NetVLAD	1024	Lazy quadruplet	
		vLPD-Net [134]	2021	LPD-Net+S-ARN	MinkPool		Joint loss	
		PPT-Net [135]	2021	Transformer	VLAD	256	Lazy quadruplet	✓
Discretization	Sparse Repre	MinkLoc3D [136]	2021	FPN	GeM	256	Triplet margin	✓
		MinkLoc++ [137]	2021	ResNet18+FPN	GeM	256	Triplet margin	✓
		EgoNN [138]	2021	CNN	GeM	256	Triplet margin	✓
		TransLoc3D [139]	2021	Transformer	NetVLAD	256	Triplet margin	✓
		MinkLoc3Dv2 [140]	2022	FPN	GeM	256	Modified Smooth-AP	✓
		MinkLoc3D-SI [141]	2022	FPN	GeM	256	Triplet margin	✓
		SVTNet [3]	2022	Transformer	GeM	256	Triple	✓
		LoGG3D-Net [142]	2022	U-Net	O2P+ePN	256	Contrastive+Quadruplet	✓
	Dense Repre	SpoxelNet [143]	2020	CNN	NetVLAD		Lazy quadruplet	✓
		VBRL [144]	2020				Modality norm	
		HiTPR [145]	2022	Transformer	Max pooling	1024	Lazy quadruplet	
		NDT-Transformer [146]	2022	Transformer	NetVLAD	256	Lazy quadruplet	✓
Projection	Spherical View	Yin et al. [147]	2017	DNN			Contrastive	
		MMCS-Net [148]	2022	Siamese CNNs	NetVLAD		Contrastive	
		SeqOT [149]	2022	Transformer	GeM	256	Triplet	✓
		OverlapTransformer [150]	2022	Transformer	NetVLAD	256	Lazy triplet	✓
		AttDLNet [151]	2021	DarkNet53	Max pooling	1024	Cosine similarity	✓
		OREOS [47]	2019	CNN			Triplet	
	BEV	SCI [48]	2019	LeNet			Categorical cross-entropy	✓
		DiSCO [152]	2021	U-Net		1024	Quadruplet+KL divergence	✓

Size and **EtE** denotes the descriptor size and end-to-end learning, respectively. **Sparse/dense repre** denotes sparse/dense representation. **Point Conv** means point convolution.

which employ the polar and cart context [4], intensities [7], frequency domain [120], F-norm [121], and Spatial Binary Pattern [122] to enhance performance.

Pairwise Matching. LiDAR Iris [44] draws inspiration from human iris signatures, utilizing LoG-Gabor filtering and thresholding to create binary signature images, then measuring descriptor similarities using Hamming distance. Some methods also employ weighted distances [24] and orientation-invariant metrics [123, 124] for pairwise similarity computation.

4.1.2 Discretization-based Methods. The discretization processing transforms the point cloud into 3D discrete representations, categorized into fixed and unfixed size-based approaches.

Fixed-size Discretization. Magnusson et al. [125] maps the point cloud to **Normal Distribution Transform (NDT)** voxels and creates a histogram based on the probability density function of the local surface. They calculate the descriptor similarities using weighted Euclidean distances. This method demonstrates the potential of NDT descriptors for place recognition. Subsequently, researchers introduce techniques like k-means++ clustering [153] and normalized cross-correlation metrics [74] to enhance generalization. Cao et al. [126] also generate a numerical descriptor by detecting contours and computing spectrum energies from wedge-shaped voxels.

Unfixed-size Discretization. DELIGHT [127, 128] divides the support region into two concentric spheres and gets non-overlapping bins by horizontal and azimuthal divisions. It computes intensity histograms for each bin and assesses descriptor similarities by chi-squared tests. Notably, the descriptor can be local or global based on the descriptor's radius and center point.

4.1.3 Point-based Methods. Several methods treat place recognition as a histogram matching problem, encoding angle and height information and calculating histogram similarities using Wasserstein metric [45], Sørensen [56], and χ^2 distances [56]. They achieve rotation invariance and overcome noises. A parallel track of works follow a projection-based scheme. M2DP [129] and C-M2DP [130] project the point cloud onto 2D planes using azimuth and elevation angles, counting point densities to create the descriptor. Multi-view density signatures enable accurate descriptions with fewer computational resources, making it particularly effective for sparse point clouds.

4.2 Learning-based Methods

4.2.1 Point-based Methods. One prevalent approach directly utilizes the inherent 3D spatial information for LiDAR point cloud processing, involving pointwise MLP, point convolution, and graph representation.

Pointwise MLP. The pioneering work PointNetVLAD [37] combines PointNet [102] for local feature extraction and NetVLAD [154] for global descriptor generation. It employs metric learning and introduces the lazy triplet and quadruplet loss functions to enhance generality. Afterward, PCAN [38] and SOE-Net [131] improve high-dimensional feature representation by incorporating attention mechanisms. LCD-Net [132] utilizes the **PointVoxel-RCNN (PV-RCNN)** [155] architecture and combines the feature extraction capabilities of DNN with transport theory algorithms.

Point Convolution. DH3D [60] introduces a Siamese network for local feature detection, description, and global descriptor extraction in a single forward pass. It incorporates multi-level spatial contextual information and channelwise feature correlations. EPC-Net [59] is a compact model based on edge convolution, simplifying the process with spatial-adjacent matrices and proxy points. It achieves excellent performance while significantly reducing computational memory.

Graph Representation. Graph networks efficiently captures underlying geometric and shape properties, allowing for feature comparison across multiple locations within graphs. LPD-Net [2] and vLPD-Net [134] extract multiple local features, including curvature, height, and density, and employ a **Graph Neural Network (GNN)** for feature aggregation. Several works integrate an

attention module to discern task-relevant features [39], learn spatial relationships between regions [135], and mitigated the influence of movable noises [133].

4.2.2 Discretization-based Methods. We divide discretization-based methods into sparse and dense representations based on voxel density.

Sparse Representation. Three approaches for generating global descriptors using 3D CNNs on sparse volume representations are **Feature Pyramid Network (FPN)**, transformer network, and other methods.

(a) FPN. MinkLoc3D [136], a pioneering work based on sparse voxelization, employs a 3D FPN [156] and **Generalized-Mean (GeM)** [157] pooling for global descriptor generation. It showcases a simple and elegant architecture, highlighting the potential of sparse voxelized representation and sparse convolutions for efficient 3D feature extraction. Later works extend enhance recognition performance by incorporating intensity [141], image [137], and attention [138, 140].
(b) Transformer Network. TransLoc3D [139] re-weights features from multiple receptive scales using an attention map and incorporates external attention layers for capturing long-range contextual information. SVT-Net [3], introduces two types of transformers to capture short-range local features and long-range contextual features, respectively. Despite a shallow network architecture, it generates descriptive descriptors.
(c) Other Methods. LoGG3D-Net [142] utilizes the sparse point-voxel convolution for high-dimensional feature embedding and introduces a local consistency loss for feature similarity maximization. It exhibits superior end-to-end performance, operating in near real-time.

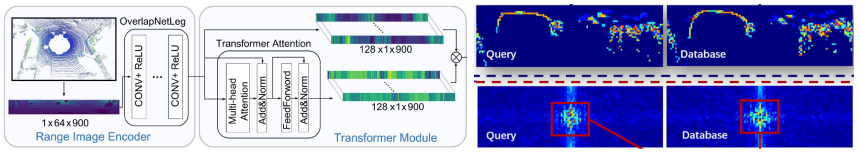
Dense Representation. Likewise, we broadly categorize dense representation-based approaches into transformer network, DNN, and other methods.

(a) Transformer Network. HiTPR [145] utilizes a short-range transformer to extract local features within cells and a long-range transformer to encode global relations among the cells. It enhances the relevance of local neighbors and global contextual dependencies. NDT-Transformer [146] introduces a novel network with three stacked transformer encoders, learning a global descriptor from discrete NDT cells. It is a valuable addition to NDT-based SLAM and **Monte Carlo Localization (MCL)** methods.
(b) DNN. SpoxelNet [143] voxelizes the point cloud in spherical coordinates, representing voxel occupancy using ternary values. It extracts multi-scale structural features and generates a global descriptor by concatenating features from various directions. This method effectively handles occlusion and moving objects in crowded indoor spaces.
(c) Other Methods. **Voxel-based Representation Learning (VBRL)** [144] tackles long-term place recognition by jointly learning voxel importance and feature modalities using structured sparsity-inducing norms. It integrates all features into a unified regularized optimization formulation.

4.2.3 Classification-based Methods. Several approaches address the place recognition problem using classifiers. FastLCD [158] encodes multi-modality features into a global descriptor, detecting candidate loop closures using supervised learning and rejecting false positives through cross-validation and post-verification. Habich et al. [159] perform loop searches within a variable radius based on the eigenvalue of the position covariance matrix and predict loops using a classifier.

4.2.4 Projection-based Methods. Following the taxonomy of learning-based 2D local descriptors (Section 3.1.2), this section also categorizes projection methods into spherical view and BEV.

Spherical View. Using spherical projection images as input, three representative methods include siamese network, transformer network, and 2D CNN only.



(a) Spherical View (b) BEV

Fig. 5. Two projection-based methods. Panels (a) and (b) are originally shown in References [150] and [152], respectively.

(a) Siamese Network. Yin et al. [147] transform the point cloud into a one-channel image and use a Siamese CNN to convert LCD into a similarity modeling problem, improving search efficiency by combining Euclidean metric and kd-tree. MMCS-Net [148] incorporates a Siamese CNN with shallow-deep feature fusion and a cascaded attention mechanism to handle pseudo images. It effectively strikes a favorable balance between effectiveness and efficiency.

(b) Transformer Network. SeqOT [149] employs multi-scale transformers to generate sub-descriptors that fuse spatial and temporal information from sequential LiDAR range images. It ensures robustness to viewpoint changes and scan order, enabling reliable place recognition even in opposite directions. As depicted in Figure 5(a), OverlapTransformer [150] extracts features from range images and integrates a transformer to capture relative feature locations, demonstrating fast running speed and robust generalization. AttDLNet [151] incorporates a four-layer attention network to capture long-range context and inter-feature relationships.

(c) 2D CNN Only. OREOS [47] uses 2D convolutional and max pooling layers to extract features from 2D range images, enhancing the descriptor performance via a triple loss function and strong negative mining strategy. It efficiently computes the descriptor while enabling long-term three-DoF metric localization in outdoor environments.

BEV. Two representative BEV-based methods are encoder-decoder network and 2D CNN only.

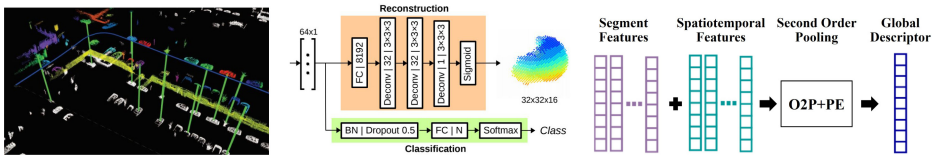
(a) Encoder-decoder network. As shown in Figure 5(b), DiSCO [152] employs an encoder-decoder network to extract descriptors and estimates relative orientation through Fourier-Mellin Transform and differentiable phase correlation. It enhances the interpretability and efficiency of the feature extractor.

(b) 2D CNN only. **Scan Context Image (SCI)** [48] extends SC [40] into three channels, enabling robot localization on a grid map through a convolutional neural network-based place classification. It demonstrates robust year-round localization with only a single day of learning.

4.3 Observations and Implications

Global descriptors are currently the most popular place recognition method, which can provide information about the entire scene, unaffected by local changes. The progress of deep learning in 3D computer vision paves the way for data-driven methods in LPR. Several observations are summarized as follows:

For the handcrafted part: (i) BEV [4, 40] demonstrates superior performance in flat structural environments but may yield poor results when the LiDAR's z-axis changes, as these methods assume local planar vehicle motion. (ii) Discretization-based methods [74, 125, 126] can describe the local surface using robust mathematical theories. However, increasing the resolution will significantly incur a heavy computational burden. (iii) Point-based methods [45, 56, 129] are the most



(a) Matching (b) Classification (c) Similarity

Fig. 6. An illustration of three representative segment-based methods. Panels (a)–(c) are originally shown in References [160], [161], and [162], respectively.

basic global descriptor methods. However, they require expensive neighbor searching to establish topological relationships. Furthermore, projection operations may result in information loss and cause potential false positives.

For the learning part: (i) Learning-based methods are efficient and accurate but require large, clean datasets and often necessitate transfer learning to address real-world issues like noise and occlusions. (ii) Transformers [145, 146] excel at capturing contextual relationships, enabling reliable recognition in cluttered environments. However, their substantial computational demands constrain the batch size for metric learning. Sparse convolutional architectures [136, 141] excel at generating informative local features while struggling to discriminate feature size in dynamic scenarios. (iii) Point-based methods [37, 38] handle unordered data well but may miss local spatial details. Classification-based methods [158, 159] assign higher weights to informative features during training. However, the specific contribution of each weak classifier to the overall prediction may be less interpretable. Projection-based methods [47, 148] are efficient and interpretable but may lose information due to dimensionality reduction.

5 LPR Techniques: Segments

Segments are meaningful region divisions characterized by similar geometric properties. These methods divide the point cloud into segments and three typical methods are shown in Figure 6.

5.1 Handcrafted Methods

5.1.1 Matching-based Methods. Segment-based matching for finding correspondences primarily comprises the SegMatch [160] family and other methods.

SegMatch Family. The pioneering work SegMatch [160], depicted in Figure 6(a), employs Euclidean clustering to partition the point cloud into segments and extracts eigenvalue-based features. It effectively identifies potential correspondences using random forest and **Random Sample Consensus (RANSAC)** [163]-based geometric verification. Dubé et al. [164] enhance SegMatch [160] by tracking a single segment using region-growing-based incremental segmentation. Moreover, certain studies effectively integrate SegMatch [160] into traditional LiDAR SLAM [165] and multi-robot systems [166].

Other Methods. RDC-SLAM [167] combines an eigenvalue-based segment descriptor, **K Nearest Neighbors (KNN)** search, and RANSAC-based verification [163] to refine relative poses. Gong et al. [168] construct a spatial relation graph to represent segments, effectively capturing general spatial relations between irregular clusters.

5.1.2 Similarity-based Methods. Seed [169] develops a segmentation-based egocentric descriptor, incorporating topological information into SC-based place recognition [40]. It achieves translation and rotation invariance by utilizing the inner topological structure of segmented objects.

5.2 Learning-based Methods

5.2.1 Matching-based Methods. Tinchev et al. [170] encode geometric properties and point distribution of segments to extract repeatable oriented key poses, which are matched using reliable shape descriptors and a Random Forest. However, significant changes in the sensor's vantage point could negatively impact segment-matching performance. Tinchev et al. [171] utilize convolution to obtain an embedding space suitable for urban and natural scenarios. They subsequently estimate match quality through probabilistic geometric validation.

5.2.2 Classification-based Methods. Several works compute the category of segments by performing classification in the descriptor space, categorized into SegMap family and other methods.

SegMap Family. As shown in Figure 6(b), the pioneering work SegMap [161] incrementally clusters point clouds to create a global segment map. It employs segmentwise KNN retrieval with a data-driven descriptor extractor comprising three convolutional and two fully connected layers, then assigns a classification score using a fully connected network. It enables high compression rates in environment reconstruction and facilitates large-scale 3D LiDAR SLAM. Subsequently, researchers successfully integrate SegMap [161] into LiDAR SLAM [172] and segment-based mapping framework [173].

Other Methods. Wietrzykowski et al. [174] propose a DNN that learns visual context from synthetic LiDAR intensity images. They claim that using the latest LiDAR and ambient images can yield additional performance improvements. OneShot [58] employs a range image-based method for segment extraction and a custom-tailored neural network to extract LiDAR-Vision descriptors.

5.2.3 Similarity-based Methods. Another approach constructs segment-based descriptors and assessed their similarities, thus combining the advantages of segments and global descriptors. As shown in Figure 6(c), Locus [162] encodes topological and temporal information of segments to create a global descriptor using second-order pooling and nonlinear transformation. It avoids global map construction, achieving robustness to viewpoint changes and occlusions.

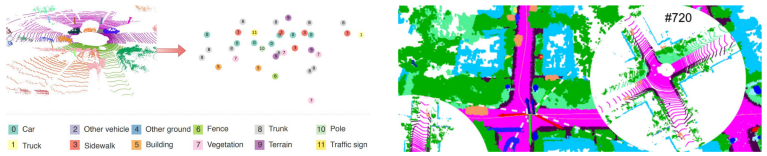
5.3 Observations and Implications

Traditional point cloud descriptors rely on low-level properties [7, 45, 56, 93, 127] to encode the point cloud, but local descriptors lack description ability, and global descriptors struggle with rotation and translation invariance. Fortunately, segments offer a good compromise between the two. Several observations are summarized as follows:

(i) Segments [160, 161, 173] offer a potential solution to reduce feature computation by avoiding processing the entire point cloud. Nevertheless, several approaches rely on point cloud aggregation or map construction, leading to inefficiencies when dealing with large-scale environments. (ii) Segment-based methods show promise in enhancing accuracy by incorporating geometric, color, and semantic information of segments. However, they require rich 3D geometry structures for segmentation, which may not always be available, thus limiting their applicability. (iii) Segment-based methods are well-known for their resilience to environmental changes, encompassing illumination, weather, and seasonal variations. However, they offer limited insights into the underlying 3D structures, resulting in subpar segmentation performance during long-term localization scenarios with numerous moving objects.

6 LPR Techniques: Semantics

Semantics refers to labels or categories that divide point clouds into various instances using learning-based segmentation technology, facilitating semantic-level place recognition. Thus, semantics-based place recognition falls under the category of learning-based methods. Based on



(a) Graph-based Methods (b) Graph-free Methods

Fig. 7. Two semantics-based methods. Panels (a) and (b) were originally shown in References [49] and [175], respectively.

the approach used for semantics association, they can be classified into two types: graph-based and graph-free.

6.1 Learning-based Methods

6.1.1 Graph-based Methods. Semantic graphs intuitively depict the location and topological information of objects. Graph similarity and graph matching are two typical graph operations.

Graph Similarity. As depicted in Figure 7(a), SGPR [49, 176] represents semantic categories and centroids of points as nodes, capturing node feature relations through edges. It develops a GNN-based graph network with node embedding, graph embedding, and graph-graph interaction to compute graph similarity. SGPR demonstrates robustness against occlusion and viewpoint changes, especially for reverse loops.

Graph Matching. GOSMatch [50] introduces an object-based place recognition approach for urban environments, which employs graph descriptors for candidate search and vertex descriptors for one-to-one correspondence calculation. BoxGraph [177] stores object shapes in vertices and simplifies place recognition to an optimal vertex assignment problem. It employs bounding boxes as appearance embeddings for vertex entities and extends them for pose estimation.

6.1.2 Graph-free Methods. Other works avoid semantic graph construction and mainly fall into two categories: semantic descriptors and other methods.

Semantics Descriptors. As shown in Figure 7(b), Semantic Scan Context [175] enhances SC [40] by utilizing semantics instead of height. Object Scan Context [178] improves SC [40] by constructing the descriptor around uniformly distributed objects (e.g., street lights and trash cans). Seq-Ndt [179] extends the NDT-based histogram descriptor[180] by incorporating semantic information and utilizes the **Kullback–Leibler (KL)** divergence to measure similarity. RINet [181] develops a lightweight siamese network with convolution, down-sampling, and attention mechanisms to compute descriptor similarities. It prioritizes scene learning over point cloud orientation and is highly efficient, allowing for deployment on resource-constrained platforms.

Other Methods. Recent studies enhance semantic-based LPR through innovative technologies and theories, including multiple hypothesis trees [182], siamese neural network [61], spherical convolution [183], and neural tensor network [184].

6.2 Observations and Implications

Inspired by human perception, semantic-based methods utilize pre-defined knowledge databases to categorize objects and identify their topological relationships. However, these methods are still relatively new and immature, because they required advanced semantic segmentation technology. Several key observations are summarized below:

(i) Graph-based methods [49, 50, 176] have streamlined point cloud comprehension but exhibited three limitations. First, potential loss of specific features, like object size. Second, inability to differentiate between parts of the same category leading to information loss. Third, computing metrics

between two graphs remains NP-complete, hindering the precise distance calculation within a reasonable timeframe. (ii) Semantic labels outperform using only geometric features, offering more interpretable and intuitive results. They demonstrate greater resilience to occlusion and viewpoint changes, especially in reverse LCD. However, predefined semantic labels in test datasets are limited, failing to encompass various categories in real-life scenarios. (iii) In dynamic or cluttered environments, leveraging objects and their topological information can enhance recognition accuracy. These methods heavily rely on the outcomes of semantic segmentation, which may lead to poor performance in diverse scenarios. Despite these challenges, they hold promise in applications where traditional methods fall short.

7 LPR Techniques: Trajectory

Trajectory information enables correlating current and recent historical scans for place recognition. Odometry (handcrafted) and sequence (learning) are two prominent methods for historical data.

7.1 Odometry-based Methods

SLAM systems with LCD modules often adopt handcrafted approaches, utilizing front-end poses or traditional registration techniques [185, 186] for place recognition to reduce system complexity. They can be further categorized into naive Euclidean distance, overlap ratio, and PCR-based test.

7.1.1 Naive Euclidean Distance. Comparing the Euclidean distance between real-time and historical poses enables rough loop closure detection. Some works use piecewise orientation functions [187] and global factor graphs [52] for pose similarity comparison, while others employ multi-sensor calibration and mapping [188].

7.1.2 Overlap Ratio. The overlap ratio can assess place similarity, with a higher value indicating closer proximity. S4-SLAM [189] stores historical poses using a kd tree and evaluates candidate loops based on overlap rate. It balances real-time performance and accuracy, demonstrating robustness even with limited feature points and high moving speeds. Mendes et al. [190] utilize an overlap criterion to generate new keyframes and implement a graphical model layer over LiDAR odometry to reduce drifts through graph-level loop closing.

7.1.3 PCR-based Test. PCR techniques verify candidate loops using relative poses, such as standard ICP, point-to-line/plane ICP, and **Generalized ICP (GICP)**.

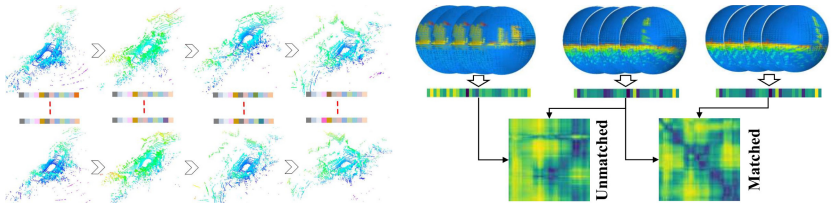
Standard ICP. IN2LAAMA [69] devises an offline probabilistic framework that identifies loop closures using poses and validates candidates with an ICP test [191], proficiently handling motion distortion without an explicit motion model.

Point-to-Line/Plane ICP. Lego-LOAM [73] compares historical scans with pose constraints and refines transformations with ICP [191]. It is a pioneering work to incorporate LCD into LiDAR SLAM, making it well suited for long-duration navigation tasks. LILO [192] extends this idea to a LiDAR-IMU system. Other works enhance registration robustness using plane graphs [193], intuitive weighting [22], and KL divergence [194], respectively.

GICP. LAMP [195] develops a multi-robot LiDAR SLAM system for challenging subterranean environments that utilizes GICP [196] to register nearby scans and proposes an Incremental Consistent Measurement set maximization to reject outlying loop closures.

7.2 Sequence-based Methods

SeqSLAM [197] pioneers visual feature similarity comparisons over time to integrate sequence information and identify the best match within local sequences, showcasing exceptional



(a) Scan Matching (b) Submap Matching

Fig. 8. Two sequence-based methods. Panels (a) and (b) are originally shown in References [198] and [199], respectively.

performance in extreme environmental changes and providing insights for LiDAR-based solutions. The point cloud sequence matching incurs higher computational costs than image-based alternatives. Consequently, approaches such as scan matching and submap matching integrate neural networks with GPUs to enhance efficiency.

7.2.1 Scan Matching. As shown in Figure 8(a), SeqLPD [198] employs LPD-Net [2] for global descriptor extraction and selects super keyframes based on feature space distribution. It combines super keyframe-based coarse matching with the local sequence fine matching to improve detection accuracy and efficiency. The trained model can be directly applied in real-world scenarios without additional training, facilitating practical applications.

7.2.2 Submap Matching. As illustrated in Figure 8(b), SeqSphereVLAD [199] and Yin et al. [53] utilize a spherical convolution module to extract orientation-equivariant local features across multiple layers of spherical perspectives. It effectively handles changing viewpoints and addresses large-scale SLAM challenges. FusionVLAD [200] proposes a multi-view fusion network that encodes top-down and spherical-view features from the local map, enhancing feature combination through a parallel fusion module for end-to-end training. It is well suited for large-scale mapping tasks with limited computation resources.

7.3 Observations and Implications

Traditional frame-to-frame comparison methods yield an intuitive similarity score but intend to degrade in closed, symmetric, and dynamic environments. The trajectory-based approaches incorporate both spatial and temporal information to address this limitation. Two observations are summarized as follows:

(i) LiDAR SLAM [73, 195] employs a straightforward LCD method based on pose proximity, followed by PCR for calculating relative transformations. Despite satisfactory results, two limitations remain. Cumulative errors affect the reliability of odometry poses in large-scale scenarios. Furthermore, the local optimality of PCR impeds the integration of loop constraints into global optimization. (ii) Sequence-based methods exhibit versatility, since they effectively incorporate diverse place recognition techniques, such as local and global descriptors [53, 198], semantics, and segments. While visual sequence-based methods have been well-studied, LiDAR-based approaches are still in the early stages. Furthermore, the expensive calculations required for matching and feature fusion restrict their practical applicability.

8 LPR Techniques: Map

Map-assisted methods provide global metric localization to achieve place recognition. They generally fall into two groups based on the map construction timing: offline and online maps.

8.1 Handcrafted Methods

8.1.1 Offline Map-based Methods. As the vehicle moves, the static offline map confines its motion within predefined boundaries. Handcrafted methods mainly involve five map types: feature, probability, point cloud, grid, and mesh.

Feature Map. Dong et al. [29] employ range image-based pole extraction to build a global map and utilize MCL to update particle weights based on pole matching. Shi et al. [71] use RANSAC [163] to extract walls from the offline map and online scans, applying point-to-point and point-to-line distance constraints to compute vehicle poses.

Probability Map. Schmiedel et al. [201] characterize surface patches in NDT maps using curvature and object shape. They match descriptors between online scans and the global map, apply RANSAC [163] for outlier detection, and evaluate matches using a normalized inlier ratio.

Point Cloud Map. Xu et al. [62] introduce a cross-section shape context descriptor that describes spatial distribution using elevation and point density, improving recognition performance with two-stage similarity estimation and the nearest cluster distance ratio. Shi et al. [72, 202] create an offline map database with a kd tree to simulate vehicle orientations and develop a binary loss function to improve localization accuracy.

Grid Map. Aldibaja et al. [203] convert LiDAR scans into image-like representations of road surfaces, incorporating elevation and irradiation data. They employ a shared ID-based XY correlation matrix to represent loop-closure events among map nodes, facilitating large-scale map processing and map-combiner event detection independent of the driving trajectory.

Mesh Map. Chen et al. [64] employ Poisson surface reconstruction to generate a mesh map, developing an observation model of an MCL framework. It showcases robust generalization across different LiDAR sensors, eliminating the need for additional training data in varying environments.

8.1.2 Online Map-based Methods. SLAM dynamically constructs and updates an online point cloud map of the surrounding environment as the vehicle navigates within unknown terrain. MULLS [67] incorporates TEASER [204] for loop verification and employs map-to-map ICP [191] to enhance inter-submap edges with accurate transformations. CT-ICP [68] projects a local map onto an elevation image, estimates a 2D transformation using RANSAC [163], and computes a six-DoF pose through ICP [191] to identify potential loop closures. Liu et al. [205] introduce a real-time 6D SLAM for large-scale natural terrains, which combines rotation histogram matching with a branch and bound search-based ICP [191] to achieve real-time LCD.

8.2 Learning-based Methods

8.2.1 Offline Map-based Methods. Researchers explore four types of offline maps in learning-based map localization: intensity, point cloud, node, and **OpenStreetMap (OSM)**.

Intensity Map. Barsan et al. [206] embed LiDAR intensity maps and online scans into a joint space to determine the vehicle's position. This method achieves centimeter-level accuracy and showcases robustness in handling uncalibrated data.

Point Cloud Map. Some works [207, 208] employ DNN for feature learning and achieve global localization through the MCL framework. They address the non-conjugate issue between the Gaussian model and MCL, enhancing long-term localization performance. L3-Net [209] captures temporal motion dynamics using deep Recurrent Neural Networks, achieving comparable localization accuracy to SOTA methods. Retriever [210] aggregates compact features with a perceiver for place recognition, enhancing computation efficiency by avoiding computation-heavy decompression.

Node Map. S4-SLAM2 [211] constructs a node map comprising point cloud, feature vectors, and location information. It extracts geometric and statistical features to create multi-modal descriptors and classified loop closures with a random forest classifier.

OpenStreetMap. OSM offers comprehensive geographic details such as streets, railways, water systems, and buildings. Several approaches [212, 213] integrates semantic information extracted from OSM into a particle filter framework. Cho et al. [214] generate a descriptor by calculating the distances to buildings at regular angles.

8.2.2 Online Map-based Methods. These methods are roughly divided into surfel-based and grid map-based methods.

Surfel Map. SuMa [215] and SuMa++ [63] employ range images and surfel-based maps for data association, detecting candidate loops by combining radius search and frame-to-model ICP [191]. They verify loops by tracking poses, which ensures robust detection even with low overlap.

Grid Map. Yin et al. [216] generate a BEV map from the local occupancy map, considering vehicle motion errors. Furthermore, they introduce an additional GAN [217] with conditional entropy reduction to enhance unsupervised feature learning for long-term recognition applications.

8.3 Observations and Implications

Maps [72, 208, 209] have been widely used in robot localization and path planning as they offer precise and detailed representations of the environment. Remarkably, map-based methods excel in recognizing topologically similar localization, providing pose information, and recovering kidnapped robots effectively. Several observations are summarized as follows:

(i) Map representations enhance global consistency and reduced localization errors. However, their large memory requirements result in time-consuming loading, communication, and processing. (ii) Map can overcome noise and partial occlusions, ensuring robust recognition even in challenging scenarios. However, the significant density difference poses difficulties in registering online scans to maps. (iii) A robust prior map facilitates long-term localization in a consistent environment. However, significant environmental changes can cause the existing map to be outdated, resulting in localization errors.

9 LPR Techniques: Other Methods

InCloud [218] distills the angular relationship between global representations, preserving the complex structure of the embedding space between training steps. Granström et al. [219] encode the point cloud using geometry features and range histograms, detecting loops with a trained classifier. While achieving high precision and recall rates, it requires the ordered point cloud.

10 Datasets and Metrics

10.1 Datasets

A large number of datasets have been collected to evaluate the performance of LPR methods. Table 3 provided a summary of these datasets. Their characteristics are summarized as follows:

Long-term Collection. [222, 223, 228] repeatedly gather the same scenario along similar routes in different seasons or times.

Multi-modal Data. In addition to LiDAR sensors, radar is used in References [226, 228] and cameras are mounted in References [150, 183, 220–223, 227–229]. Semantic information is also available in Reference [225].

LiDAR Sparsity. These datasets cover various density LiDAR sensors, such as mechanical 16-line [224, 227], 32-line [222, 228], 64-line [220, 221], and 128-line LiDAR [121], as well as solid-state LiDAR [74].

Table 3. A Summary of Existing Datasets for LPR

Year	Name	Seq	Trajectory (KM)	Type	Sensor Modilty	Model of LiDAR	Loop	GT	LT	Public
2009	Hannover2 [125]	1	1.24	Out	L		S+O	✓		✓
2010	Freiburg [93]	1	0.723	Out	L	SICK LMS	S	✓		✓
2011	Ford Campus [220]			Out	C+L	Velodyne HDL-64E		✓		✓
2012	KITTI Odometry [221]	22	39.2	Out	C+L	Velodyne HDL-64E	S+O	✓		✓
2016	NCLT [222]	27	147.4	In+Out	C+L	Velodyne HDL-32+Hokuyo UTM-30LX+Hokuyo URG-04LX	S+O	✓	✓	✓
2017	Oxford RobotCar [223]	> 130	>1000	Out	C+L	SICK LD-MRS+SICK LMS-151	S+O	✓	✓	✓
2018	Complex Urban [224]	19	158.82	Out	L	Velodyne VLP-16+SICK LMS-511	S+O	✓		✓
2018	In-House [37]	3		Out	L	Velodyne HDL-64E		✓	✓	✓
2019	Semantic KITTI [225]	22	39.2	Out	L	Velodyne HDL-64E	S+O	✓		✓
2019	Apollo-SouthBay [209]		>380	Out	L	Velodyne HDL-64E		✓	✓	✓
2019	HKUST [74]			In+Out		Livox-MID40				✓
2020	MulRan [226]	12	41.2	Out	L+R	Ouster OS1-64	S+O	✓	✓	✓
2020	USyd [227]	>50		Out	C+L	Velodyne VLP-16	S+O	✓	✓	✓
2020	Oxford Radar Robotcar [228]	>32	>280	Out	C+L+R	Velodyne HDL-32E+SICK LMS-151+Navtech CTS350-X	S+O	✓	✓	✓
2021	DUT-AS [126]	30		Out	L	SICK LMS 511	S+O	✓	✓	
2021	CMU Dataset [183]	11	2.0	Out	C+L	Velodyne VLP-16		✓		
2021	Pittsburgh Dataset [183]	12	12.0	Out	C+L	Velodyne VLP-16		✓		
2022	HAOMO [150]	5		Out	C+L	HESAI PandarXT-32	S+O	✓	✓	✓
2022	Campus [53]	11	2	Out	L	Velodyne-VLP 16		✓		
2022	City [53]	13	11	Out	L	Velodyne-VLP 16		✓		
2022	KITTI-360 [229]	9	73.7	Out	C+L	Velodyne HDL-64E	S+O	✓		✓
2022	CHDloop [121]	5	1.519	Out	L	RoboSense RS-Ruby 128	S+O	✓		
2022	LGSVL [148]			Out		Velodyne HDL-64E		✓	✓	
2022	Real Vehicle [148]			Out	C+L	Velodyne VLP-32C	S+O	✓		

Seq, **GT** and **LT** represent sequence, ground-truth, and long-term, respectively. **In** and **Out** mean indoor and outdoor, respectively. **C**, **L**, and **R** denote camera, LiDAR, and radar, respectively. **S** and **O** represent the same and oppo-direction loop, respectively.

Viewpoint Change. In addition to same-direction revisits, References [121, 126, 148, 150, 221–229] contain reverse loops.

Scenario Diversity. These datasets are generally divided into two categories: indoor and outdoor datasets. Outdoor datasets are the most widely used, mainly including campuses [222], highways [224], rural areas [221], cities [227], and riversides [226].

10.2 Evaluation Metrics

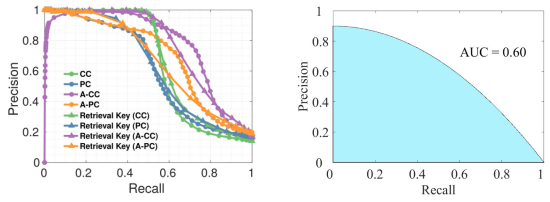
Different evaluation metrics have been proposed to test LPR methods, summarized as follows:

Revisit Criteria. A distance threshold is defined before evaluation to determine whether the query and candidate belong to the same place.

Precision-recall (PR) Curves [230]. As depicted in Figure 9(a), this curve measures the relationship between Pecision (P) and Recall (R) under different threshold parameters. P measures the ratio of correct matches to the total of predicted positive instances, while R quantifies the proportion of real positive cases correctly identified as positive matches,

$$P = \frac{TP}{TP + FP}, \; R = \frac{TP}{TP + FN}, \tag{5}$$

where TP, FP, and FN represent true positive, false positive, and false negative, respectively.



(a) PR curve (b) AUC

Fig. 9. An illustration of PR curve and AUC. (a) is originally shown in Reference [4]. The AUC in (b) corresponds to the area of the blue region.

Area Under the PR Curve (AUC) [48]. As illustrated in Figure 9(b), it reflects the discrimination power of a LPR method and a larger AUC means more places are recognized with fewer errors. However, it does not retain any information regarding the features of the original PR Curve.

Recall @Top-N. It evaluates the accuracy of place recognition methods in identifying the correct places among the top-k retrieved matches. A higher value indicates better performance. TOP 1% [51] and TOP 1 [123] are the two most frequently used metrics.

F_β **Score**. It is the harmonic mean of precision and recall. A high value indicates the system struck a good balance between them as follows:

$$F_\beta = \left(1+\beta^2\right) \times \frac{P \times R}{\beta^2 P + R}, \tag{6}$$

where P and R represent precision and recall, respectively. β is a parameter that determines the weights of P and R. F_1 score [49, 61, 158, 162, 168, 175, 177] is the most frequently used metric:

$$F_1 = 2 \times \frac{P \times R}{P + R}, \tag{7}$$

where F_1 treats P and R as equally important. The maximum F_1 score (F_1^m) is then calculated as follows:

$$F_1^m = \underset{\tau}{max}\ 2 \times \frac{P_\tau \times R_\tau}{P_\tau + R_\tau}, \tag{8}$$

where τ refers to a user-defined threshold for matching score or distance. For place recognition, we identify the best candidate for the query scan and compute their matching score or distance. We compare this score with τ to classify the result as TP, FP, or FN. We then calculate the precision, recall, and F_1 score for the entire sequence. Finally, we evaluate different τ values to determine the highest F_1 score as the final result F_1^m.

Extended Precision (EP). It provides more comprehensive insights by simultaneously considering the lower and upper-performance bounds of an LPR method [162, 175, 231]:

$$EP = \frac{1}{2}\left(P_{R0} + R_{P100}\right), \tag{9}$$

where P_{R0} is the precision at minimum recall, and R_{P100} is the max recall at 100% precision.

Translation and Rotation Error. In global localization-related tasks, the translation error e_t quantifies the difference between the estimated and ground-truth translations, reflecting the accuracy of the robot's position. The rotation error e_r measures the discrepancy between the estimated and actual rotations, indicating the accuracy of the robot's attitude,

$$e_t = \boldsymbol{t}_{es} - \boldsymbol{t}_{gt}, \quad e_r = \arccos(trace\left(\mathbf{R}_{gt}^{\mathrm{T}}\mathbf{R}_{es}\right)/2), \tag{10}$$

where $\boldsymbol{t}_{es}$ and $\mathbf{R}_{es}$ denote the estimated translation and rotation and $\boldsymbol{t}_{gt}$ and $\mathbf{R}_{gt}$ denote the ground-truth values. $trace$ refers to the matrix's trace.

Localization Success Rate. The success rate $rate_{success}$ is the proportion of successfully localized cases $N_{success}$ to the total cases N_{total}. A higher success rate indicates a more robust system capable of accurately localizing the robot from its initial state:

$$rate_{success} = \frac{N_{success}}{N_{total}}, \tag{11}$$

where a localization is successful only if the translation error e_t is less than ϕ_t and the rotation error e_r is less than ϕ_t; ϕ_t and ϕ_t are thresholds for translation and rotation errors, respectively.

Running Efficiency. Runtime is crucial for online SLAM. Descriptor-based methods typically involve feature extraction and search, while map-based methods require map processing and matching. Data-driven methods, however, necessitate training and inference.

11 Future Directions

11.1 Multi-modality Information

Multi-modality information offers the opportunity to leverage complementary features and enhance the robustness of localization. Several novel solutions are as follows:

WiFi. WiFi-based localization retrieves the Media Access Control address of routers via the client device to calculate positions. This solution offers extensive coverage and achieves accurate indoor localization, overcoming GPS signal limitations. Moreover, it is easy to deploy and provides fast localization, holding significant potential for advancing indoor localization industries.

Voice. Voice-based localization uses microphone arrays to capture audio signals, employing delay estimation or spectral analysis to determine sound source locations. It offers advantages like low computational requirements, high concealment, and strong compatibility. Furthermore, it can seamlessly integrate with human-computer interaction systems, smart homes, and other voice-controlled applications, enhancing vehicle situational awareness.

Radio Frequency Identification. RFID employs radio signals to identify and track objects without physical contact. Its anti-interference capability will ensure reliable autonomous driving in harsh environments. Additionally, the long lifespan can enhance the stability of long-term localization systems. Applying cryptographic encryption to tag data can strengthen system security.

11.2 Innovative Solutions

Breaking free from conventional solutions and incorporating interdisciplinary new technologies into autonomous driving holds the potential for unforeseen improvements:

Cloud Computing. Cloud computing offers high-performance shared computing resources to users. Offloading computing tasks to cloud servers enhances robot localization efficiency. With access to powerful computing resources, robots can process diverse high-precision sensor data, thus improving their localization capabilities.

Quantum Technology. Quantum technology revolutionizes information calculation, encoding, and transmission. Quantum sensors can capture subtle changes, delivering ultra-high-precision measurements. Integrating such sensors into robotic navigation will greatly enhance localization and mapping in complex environments.

Bio-inspired Localization. Bio-inspired navigation and group behavior provide novel insights for robot localization. Inspired by turtles' navigation behavior, robots can enhance localization robustness by employing magnetic field sensors. Multi-robot systems can improve formation stability by emulating the cooperative behavior seen in bird flocks.

11.3 Advanced Sensors

Equipping robots with advanced sensors can enhance their navigation capabilities. We present several promising sensors as follows:

Solid-State LiDAR. Solid-state LiDAR employs a Micro-Electro-Mechanical System, Optical Phased Array, or flash technology for signal transmission and reception. It boasts a compact size, high resolution, fast scanning speed, and extended measurement range. It enables precise identification of buildings, vehicles, and traffic signs, effectively ensuring autonomous driving safety, supporting smart transportation data, and monitoring traffic accidents.

Event Camera. Event cameras exclusively generate an asynchronous event stream when notable visual changes occur. They offer numerous benefits, including high time resolution, low latency, wide dynamic range, and low power consumption. Equipping vehicles with event cameras enhances obstacle avoidance in high-speed scenarios, enables navigation through scenes with abrupt light changes, and facilitates handling emergencies.

Millimeter Wave Radar. Millimeter-wave radar employs Frequency-Modulated Continuous Wave signals and mixers to measure speed, distance, and direction, providing cost-effectiveness, precise longitudinal ranging, accurate object detection, weather resistance, and high bandwidth. It will find wide applications in blind spot detection, object detection and tracking, parking assistance, and adaptive cruise control.

11.4 Significant Applications

As a new scientific and technological revolution unfolds, robot technology will spearhead advancements in several critical fields shaping the fate of humanity:

Space Exploration. In inhospitable environments like the moon, robots will assume the role of humans, undertaking tasks such as terrain mapping, mineral identification, house construction, and 3D printing. They will aid human understanding of deep space and other planets.

Polar Research. Robots advance polar research by enhancing data collection capabilities. They collect data on glaciers, weather, and temperature, facilitating continuous environmental monitoring. High-resolution mapping of polar topography helps identify landform changes, glacier movement, and geological processes.

Underwater Robots. Underwater robots are advanced submersibles tailored for extreme underwater operations. The fiber optic gyroscope and Doppler log will greatly enhance localization performance, benefiting port construction and naval defense. Sonar detection technology will further improve task efficiency like underwater rescue and pipeline maintenance.

11.5 Approach Evaluations

A fair and thorough evaluation is crucial for adapting robot products, algorithms, and scenarios. Here are three pivotal considerations for future algorithm evaluation.

Scalability and Efficiency. The growing affordability and accessibility of LiDAR sensors have spurred the demand for large-scale place recognition. This necessitates the development of scalable algorithms for handling large-scale point clouds and the design of efficient algorithms for the real-time processing of point cloud data on resource-constrained platforms.

Long-term Place Recognition. Long-term place recognition refers to the ability of a system to identify places over extended periods, despite appearance and weather variations. It is a crucial capability for autonomous navigation. Designing algorithms that can handle seasons, weather, appearances, and dynamic objects, will drive significant advancements in this field.

Standardized Datasets. A good dataset should possess sufficient size, high-quality data, reliable ground truth, annotations or labels, ethical considerations, normalized form, and clear

instructions. Creating diverse data encompassing various sensor modalities, environmental conditions, and weather changes is also highly valuable.

12 Conclusions

As high-level autonomous driving advances, robust navigation systems are essential for navigating complex environments. Place recognition enables vehicles to recognize previously visited locations despite changes in appearance, weather, and viewpoints, even determining their global location within prior maps. It is becoming increasingly important in autonomous driving. LiDAR, with its rich 3D data, long-range measurement, and stability in harsh conditions, has made LPR a research hotspot. To address the gap in this field, we propose the first LPR survey. We first discuss the main problems solved by place recognition and analyze LPR's role in autonomous driving. Then, we offer a comprehensive problem statement, which divides LPR into implicit loop closure detection and explicit global localization to help readers understand the function of LPR in different requirements.

In recent years, many regions have implemented strict safety requirements for autonomous driving. LiDAR, as a reliable sensor, enhances system safety and reliability, making it an ideal choice. With technological advancements and increased production, LiDAR costs have decreased, broadening its application scope. Consequently, many LPR methods have emerged. We comprehensively classify these methods, describe their principles, and summarize their architectures, advantages, and disadvantages. These detailed analyses aim to help researchers understand each method's applicability and inspire advancements in place recognition technology for challenging environments such as docks, parks, and woodlands.

In the future, ongoing technological innovations will enhance LiDAR performance with higher resolution, faster scanning speeds, and longer detection distances. Large-scale production and usage of new materials will reduce costs, broadening LiDAR's applications. The fusion of LiDAR with other sensors (e.g., cameras, radars, GPS) will advance, providing more comprehensive and accurate localization and place recognition. To promote the further development of LPR, we summarize commonly used datasets, evaluation metrics, and promising future directions. Additionally, we maintain a project to collect SOTA LPR technologies, keeping the robotics community up to date.

References

[1] Akihiko Torii, Josef Sivic, Masatoshi Okutomi, and Tomas Pajdla. 2015. Visual place recognition with repetitive structures. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 11 (2015), 2346–2359.

[2] Zhe Liu, Shunbo Zhou, Chuanzhe Suo, Peng Yin, Wen Chen, Hesheng Wang, Haoang Li, and Yun-Hui Liu. 2019. 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*. 2831–2840.

[3] Zhaoxin Fan, 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. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 551–560.

[4] Giseop Kim, Sunwook Choi, and Ayoung Kim. 2022. Scan context++: Structural place recognition robust to rotation and lateral variations in urban environments. *IEEE Trans. Robot.* 38, 3 (2022), 1856–1874.

[5] Pengcheng Shi, Yilin Xiao, Wenqing Chen, Jiayuan Li, and Yongjun Zhang. 2024. A new horizon: Employing map clustering similarity for lidar-based place recognition. *IEEE Trans. Intell. Vehic.* Early access, Jan. 30 (2024). DOI: 10.1109/TIV.2024.3360321

[6] Konstantinos A. Tsintotas, Loukas Bampis, and Antonios Gasteratos. 2022. The revisiting problem in simultaneous localization and mapping: A survey on visual loop closure detection. *IEEE Trans. Intell. Transport. Syst.* 23, 11 (2022), 19929–19953.

[7] Han Wang, Chen Wang, and Lihua Xie. 2020. Intensity scan context: Coding intensity and geometry relations for loop closure detection. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'20)*. IEEE, 2095–2101.

[8] Bastian Steder, Michael Ruhnke, Slawomir Grzonka, and Wolfram Burgard. 2011. Place recognition in 3D scans using a combination of bag of words and point feature based relative pose estimation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. 1249–1255.

[9] Yuri D. V. Yasuda, Luiz Eduardo G. Martins, and Fabio A. M. Cappabianco. 2020. Autonomous visual navigation for mobile robots: A systematic literature review. *ACM Comput. Surv.* 53, 1 (2020), 1–34.

[10] Pengcheng Shi, Jiayuan Li, Xinyi Liu, and Yongjun Zhang. 2024. Indoor cylinders guided LiDAR global localization and loop closure detection. *Geomat. Inf. Sci. Wuhan Univ.* 49, 7 (2024), 1088–1099.

[11] David Schleicher, Luis M. Bergasa, Manuel Ocaña, Rafael Barea, and María Elena López. 2009. Real-time hierarchical outdoor SLAM based on stereovision and GPS fusion. *IEEE Trans. Intell. Transport. Syst.* 10, 3 (2009), 440–452.

[12] Xin Xia, Ehsan Hashemi, Lu Xiong, Amir Khajepour, and Nan Xu. 2021. Autonomous vehicles sideslip angle estimation: Single antenna GNSS/IMU fusion with observability analysis. *IEEE IoT J.* 8, 19 (2021), 14845–14859.

[13] Mingming Zhang, Xingxing Zuo, Yiming Chen, Yong Liu, and Mingyang Li. 2021. Pose estimation for ground robots: On manifold representation, integration, reparameterization, and optimization. *IEEE Trans. Robot.* 37, 4 (2021), 1081–1099.

[14] Ralf Möller, Michael Horst, and David Fleer. 2014. Illumination tolerance for visual navigation with the holistic min-warping method. *Robotics* 3, 1 (2014), 22–67.

[15] Yongjun Zhang, Pengcheng Shi, and Jiayuan Li. 2024. 3D LiDAR SLAM: A survey. *Photogram. Rec.* 39, 186 (2024), 457–517.

[16] Claudine Badue, Rânik Guidolini, Raphael Vivacqua Carneiro, Pedro Azevedo, Vinicius B Cardoso, Avelino Forechi, Luan Jesus, Rodrigo Berriel, Thiago M Paixao, Filipe Mutz, et al. 2021. Self-driving cars: A survey. *Expert Syst. Appl.* 165 (2021), 113816.

[17] SAE International. 2018. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. *SAE Int.* 4970, 724 (2018), 1–5.

[18] Ardi Tampuu, Tambet Matiisen, Maksym Semikin, Dmytro Fishman, and Naveed Muhammad. 2020. A survey of end-to-end driving: Architectures and training methods. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 4 (2020), 1364–1384.

[19] Éloi Zablocki, Hédi Ben-Younes, Patrick Pérez, and Matthieu Cord. 2022. Explainability of deep vision-based autonomous driving systems: Review and challenges. *Int. J. Comput. Vis.* 130, 10 (2022), 2425–2452.

[20] Alberto Y. Hata and Denis F. Wolf. 2016. Feature detection for vehicle localization in urban environments using a multilayer LIDAR. *IEEE Trans. Intell. Transport. Syst.* 17, 2 (2016), 420–429.

[21] Charles R. Qi, Hao Su, Matthias Nießner, Angela Dai, Mengyuan Yan, and Leonidas J. Guibas. 2016. Volumetric and multi-view cnns for object classification on 3d data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5648–5656.

[22] Masashi Yokozuka, Kenji Koide, Shuji Oishi, and Atsuhiko Banno. 2020. LiTAMIN: LiDAR-based tracking and mapping by stabilized ICP for geometry approximation with normal distributions. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'20)*. 5143–5150.

[23] Suining He and Kang G. Shin. 2017. Geomagnetism for smartphone-based indoor localization: Challenges, advances, and comparisons. *ACM Comput. Surv.* 50, 6 (2017), 1–37.

[24] Dong Xu, Jingbin Liu, Juha Hyyppä, Yifan Liang, and Wuyong Tao. 2022. A heterogeneous 3D map-based place recognition solution using virtual LiDAR and a polar grid height coding image descriptor. *ISPRS J. Photogram. Remote Sens.* 183 (2022), 1–18.

[25] Laurene Claussmann, Marc Revilloud, Dominique Gruyer, and Sébastien Glaser. 2019. A review of motion planning for highway autonomous driving. *IEEE Trans. Intell. Transport. Syst.* 21, 5 (2019), 1826–1848.

[26] Jingyuan Zhao, Wenyi Zhao, Bo Deng, Zhenghong Wang, Feng Zhang, Wenxiang Zheng, Wanke Cao, Jinrui Nan, Yubo Lian, and Andrew F Burke. 2023. Autonomous driving system: A comprehensive survey. *Expert Syst. Appl.* 242 (2023), 122836.

[27] Dominique Gruyer, Valentin Magnier, Karima Hamdi, Laurène Claussmann, Olivier Orfila, and Andry Rakotonirainy. 2017. Perception, information processing and modeling: Critical stages for autonomous driving applications. *Ann. Rev. Contr.* 44 (2017), 323–341.

[28] Haryong Song, Wonsub Choi, and Haedong Kim. 2016. Robust vision-based relative-localization approach using an RGB-depth camera and LiDAR sensor fusion. *IEEE Trans. Industr. Electr.* 63, 6 (2016), 3725–3736.

[29] Hao Dong, Xieyuanli Chen, and Cyrill Stachniss. 2021. Online range image-based pole extractor for long-term LiDAR localization in urban environments. In *Proceedings of the European Conference on Mobile Robots (ECMR'21)*. IEEE, 1–6.

[30] Elhousni Mahdi and Huang Xinming. 2022. A survey on visual map localization using LiDARs and cameras. arXiv:2208.03376. Retrieved from https://arxiv.org/abs/2208.03376

[31] Chenyi Chen, Ari Seff, Alain Kornhauser, and Jianxiong Xiao. 2015. Deepdriving: Learning affordance for direct perception in autonomous driving. In *Proceedings of the IEEE International Conference on Computer Vision*. 2722–2730.

[32] Mrinal R. Bachute and Javed M. Subhedar. 2021. Autonomous driving architectures: Insights of machine learning and deep learning algorithms. *Mach. Learn. Appl.* 6 (2021), 100164.

[33] Dvij Kalaria, Qin Lin, and John M Dolan. 2024. Delay-aware robust control for safe autonomous driving and racing. *IEEE Trans. Intell. Transport. Syst.* 25, 7 (2024), 7140–7150.

[34] Daniel Omeiza, Helena Webb, Marina Jirotka, and Lars Kunze. 2021. Explanations in autonomous driving: A survey. *IEEE Trans. Intell. Transport. Syst.* 23, 8 (2021), 10142–10162.

[35] Peng Yin, Shiqi Zhao, Ivan Cisneros, Abulikemu Abuduweili, Guoquan Huang, Micheal Milford, Changliu Liu, Howie Choset, and Sebastian Scherer. 2022. General place recognition survey: Towards the real-world autonomy age. arXiv:2209.04497. Retrieved from https://arxiv.org/abs/2209.04497

[36] Xiwu Zhang, Lei Wang, and Yan Su. 2021. Visual place recognition: A survey from deep learning perspective. *Pattern Recogn.* 113 (2021), 107760.

[37] Mikaela Angelina Uy and Gim Hee Lee. 2018. Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4470–4479.

[38] Wenxiao Zhang and Chunxia Xiao. 2019. PCAN: 3D attention map learning using contextual information for point cloud based retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12436–12445.

[39] Qi Sun, Hongyan Liu, Jun He, Zhaoxin Fan, and Xiaoyong Du. 2020. Dagc: Employing dual attention and graph convolution for point cloud based place recognition. In *Proceedings of the International Conference on Multimedia Retrieval*. 224–232.

[40] Giseop Kim and Ayoung Kim. 2018. Scan context: Egocentric spatial descriptor for place recognition within 3d point cloud map. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'18)*. IEEE, 4802–4809.

[41] Filip Radenović, Giorgos Tolias, and Ondřej Chum. 2019. Fine-tuning CNN image retrieval with no human annotation. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 7 (2019), 1655–1668.

[42] Nabil M. Drawil, Haitham M. Amar, and Otman A. Basir. 2013. GPS localization accuracy classification: A context-based approach. *IEEE Trans. Intell. Transport. Syst.* 14, 1 (2013), 262–273.

[43] Su-Yong An and Jaeyoung Kim. 2022. Extracting statistical signatures of geometry and structure in 2D occupancy grid maps for global localization. *IEEE Robot. Autom. Lett.* 7, 2 (2022), 4291–4298.

[44] Ying Wang, Zezhou Sun, Cheng-Zhong Xu, Sanjay E Sarma, Jian Yang, and Hui Kong. 2020. Lidar iris for loop-closure detection. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'20)*. IEEE, 5769–5775.

[45] Timo Röhling, Jennifer Mack, and Dirk Schulz. 2015. A fast histogram-based similarity measure for detecting loop closures in 3-d lidar data. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'15)*. IEEE, 736–741.

[46] Lun Luo, Si-Yuan Cao, Zehua Sheng, and Hui-Liang Shen. 2022. LiDAR-based global localization using histogram of orientations of principal normals. *IEEE Trans. Intell. Vehic.* 7, 3 (2022), 771–782.

[47] Lukas Schaupp, Mathias Bürki, Renaud Dubé, Roland Siegwart, and Cesar Cadena. 2019. OREOS: Oriented recognition of 3D point clouds in outdoor scenarios. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'19)*. IEEE, 3255–3261.

[48] Giseop Kim, Byungjae Park, and Ayoung Kim. 2019. 1-Day learning, 1-year localization: Long-term LiDAR localization using scan context image. *IEEE Robot. Autom. Lett.* 4, 2 (2019), 1948–1955.

[49] Xin Kong, Xuemeng Yang, Guangyao Zhai, Xiangrui Zhao, Xianfang Zeng, Mengmeng Wang, Yong Liu, Wanlong Li, and Feng Wen. 2020. Semantic graph based place recognition for 3d point clouds. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'20)*. IEEE, 8216–8223.

[50] Yachen Zhu, Yanyang Ma, Long Chen, Cong Liu, Maosheng Ye, and Lingxi Li. 2020. Gosmatch: Graph-of-semantics matching for detecting loop closures in 3d lidar data. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'20)*. IEEE, 5151–5157.

[51] Lun Luo, Si-Yuan Cao, Bin Han, Hui-Liang Shen, and Junwei Li. 2021. BVMatch: Lidar-based place recognition using bird's-eye view images. *IEEE Robot. Autom. Lett.* 6, 3 (2021), 6076–6083.

[52] Tixiao Shan, Brendan Englot, Drew Meyers, Wei Wang, Carlo Ratti, and Daniela Rus. 2020. LIO-SAM: Tightly-coupled lidar inertial odometry via smoothing and mapping. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'20)*. 5135–5142.

[53] Peng Yin, Fuying Wang, Anton Egorov, Jiafan Hou, Zhenzhong Jia, and Jianda Han. 2022. Fast sequence-matching enhanced viewpoint-invariant 3-D place recognition. *IEEE Trans. Industr. Electr.* 69, 2 (2022), 2127–2135.

[54] Jiayuan Li, Pengcheng Shi, Qingwu Hu, and Yongjun Zhang. 2023. QGORE: Quadratic-time guaranteed outlier removal for point cloud registration. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 9 (2023), 11136–11151.
[55] Michael Bosse and Robert Zlot. 2013. Place recognition using keypoint voting in large 3D lidar datasets. In *Proceedings of the IEEE International Conference on Robotics and Automation*. 2677–2684.
[56] Naveed Muhammad and Simon Lacroix. 2011. Loop closure detection using small-sized signatures from 3D LIDAR data. In *Proceedings of the IEEE International Symposium on Safety, Security, and Rescue Robotics*. IEEE, 333–338.
[57] Jiadong Guo, Paulo VK Borges, Chanoh Park, and Abel Gawel. 2019. Local descriptor for robust place recognition using lidar intensity. *IEEE Robot. Autom. Lett.* 4, 2 (2019), 1470–1477.
[58] Sebastian Ratz, Marcin Dymczyk, Roland Siegwart, and Renaud Dubé. 2020. Oneshot global localization: Instant lidar-visual pose estimation. In *Proceedings of the IEEE International conference on Robotics and Automation (ICRA'20)*. IEEE, 5415–5421.
[59] Le Hui, Mingmei Cheng, Jin Xie, Jian Yang, and Ming-Ming Cheng. 2022. Efficient 3D point cloud feature learning for large-scale place recognition. *IEEE Trans. Image Process.* 31 (2022), 1258–1270.
[60] Juan Du, Rui Wang, and Daniel Cremers. 2020. Dh3d: Deep hierarchical 3d descriptors for robust large-scale 6dof relocalization. In *European Conference on Computer Vision*. Springer, 744–762.
[61] Xieyuanli Chen, Thomas Läbe, Andres Milioto, Timo Röhling, Jens Behley, and Cyrill Stachniss. 2022. OverlapNet: A siamese network for computing LiDAR scan similarity with applications to loop closing and localization. *Autonom. Robot.* 46, 1 (2022), 61–81.
[62] Dong Xu, Jingbin Liu, Yifan Liang, Xuanfan Lv, and Juha Hyyppä. 2022. A LiDAR-based single-shot global localization solution using a cross-section shape context descriptor. *ISPRS J. Photogram. Remote Sens.* 189 (2022), 272–288.
[63] Xieyuanli Chen, Andres Milioto, Emanuele Palazzolo, Philippe Giguere, Jens Behley, and Cyrill Stachniss. 2019. Suma++: Efficient lidar-based semantic slam. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'19)*. IEEE, 4530–4537.
[64] Xieyuanli Chen, Ignacio Vizzo, Thomas Läbe, Jens Behley, and Cyrill Stachniss. 2021. Range image-based LiDAR localization for autonomous vehicles. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'21)*. 5802–5808.
[65] Matthew McDermott and Jason Rife. 2024. Correcting motion distortion for LIDAR scan-to-map registration. *IEEE Robot. Autom. Lett.* 9, 2 (2024), 1516–1523.
[66] Han Wang, Chen Wang, Chun-Lin Chen, and Lihua Xie. 2021. F-LOAM : Fast LiDAR odometry and mapping. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'21)*. 4390–4396.
[67] Yue Pan, Pengchuan Xiao, Yujie He, Zhenlei Shao, and Zesong Li. 2021. MULLS: Versatile LiDAR SLAM via multi-metric linear least square. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'21)*. IEEE, 11633–11640.
[68] Pierre Dellenbach, Jean-Emmanuel Deschaud, Bastien Jacquet, and François Goulette. 2022. CT-ICP: Real-time elastic LiDAR odometry with loop closure. In *Proceedings of the International Conference on Robotics and Automation (ICRA'22)*. 5580–5586.
[69] Cedric Le Gentil, Teresa Vidal-Calleja, and Shoudong Huang. 2020. IN2LAAMA: Inertial lidar localization autocalibration and mapping. *IEEE Trans. Robot.* 37, 1 (2020), 275–290.
[70] Kamil Żywanowski, Adam Banaszczyk, and Michał R. Nowicki. 2020. Comparison of camera-based and 3D LiDAR-based place recognition across weather conditions. In *Proceedings of the 16th International Conference on Control, Automation, Robotics and Vision (ICARCV'20)*. 886–891.
[71] Pengcheng Shi, Qin Ye, Zhang Shaoming, and Deng Haifeng. 2021. Localization initialization for multi-beam LiDAR considering indoor scene feature. *Acta Geodaet. Cartogr. Sin.* 50 (2021), 1594–1604.
[72] Pengcheng Shi, Jiayuan Li, and Yongjun Zhang. 2023. LiDAR localization at 100 FPS: A map-aided and template descriptor-based global method. *Int. J. Appl. Earth Observ. Geoinf.* 120 (2023), 103336.
[73] Tixiao Shan and Brendan Englot. 2018. Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'18)*. IEEE, 4758–4765.
[74] Jiarong Lin and Fu Zhang. 2019. A fast, complete, point cloud based loop closure for LiDAR odometry and mapping. arXiv:1909.11811. Retrieved from https://arxiv.org/abs/1909.11811
[75] Michael Horst and Ralf Möller. 2017. Visual place recognition for autonomous mobile robots. *Robotics* 6, 2 (2017), 9.
[76] Stephanie Lowry, Niko Sünderhauf, Paul Newman, John J. Leonard, David Cox, Peter Corke, and Michael J. Milford. 2016. Visual place recognition: A survey. *IEEE Trans. Robot.* 32, 1 (2016), 1–19.
[77] Daniel Wilson, Xiaohan Zhang, Waqas Sultani, and Safwan Wshah. 2021. Visual and object geo-localization: A comprehensive survey. arXiv:2112.15202. Retrieved from https://arxiv.org/abs/2112.15202
[78] Fei Chen, Xiaodong Wang, Yunxiang Zhao, Shaohe Lv, and Xin Niu. 2022. Visual object tracking: A survey. *Comput. Vis. Image Understand.* 222 (2022), 103508.

[79] Seyed Mojtaba Marvasti-Zadeh, Li Cheng, Hossein Ghanei-Yakhdan, and Shohreh Kasaei. 2022. Deep learning for visual tracking: A comprehensive survey. *IEEE Trans. Intell. Transport. Syst.* 23, 5 (2022), 3943–3968.
[80] Iman Abaspur Kazerouni, Luke Fitzgerald, Gerard Dooly, and Daniel Toal. 2022. A survey of state-of-the-art on visual SLAM. *Expert Syst. Appl.* 205 (2022), 117734.
[81] Song Zhang, Shili Zhao, Dong An, Jincun Liu, He Wang, Yu Feng, Daoliang Li, and Ran Zhao. 2022. Visual SLAM for underwater vehicles: A survey. *Comput. Sci. Rev.* 46 (2022), 100510.
[82] Kunping Huang, Sen Zhang, Jing Zhang, and Dacheng Tao. 2023. Event-based simultaneous localization and mapping: A comprehensive survey. arXiv:2304.09793. Retrieved from https://arxiv.org/abs/2304.09793
[83] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, José Neira, Ian Reid, and John J Leonard. 2016. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Trans. Robot.* 32, 6 (2016), 1309–1332.
[84] Huan Yin, Xuecheng Xu, Sha Lu, Xieyuanli Chen, Rong Xiong, Shaojie Shen, Cyrill Stachniss, and Yue Wang. 2024. A survey on global LiDAR localization: Challenges, advances and open problems. *Int. J. Comput. Vis.* 132 (2024), 3139–3171.
[85] Andrew E Johnson and Martial Hebert. 1999. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* 21, 5 (1999), 433–449.
[86] Titus Cieslewski, Elena Stumm, Abel Gawel, Mike Bosse, Simon Lynen, and Roland Siegwart. 2016. Point cloud descriptors for place recognition using sparse visual information. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'16)*. IEEE, 4830–4836.
[87] Dario Lodi Rizzini. 2017. Place recognition of 3D landmarks based on geometric relations. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'17)*. IEEE, 648–654.
[88] Huan Zhao, Minjie Tang, and Han Ding. 2020. HoPPF: A novel local surface descriptor for 3D object recognition. *Pattern Recogn.* 103 (2020), 107272.
[89] Federico Tombari, Samuele Salti, and Luigi Di Stefano. 2010. Unique shape context for 3D data description. In *Proceedings of the ACM Workshop on 3D Object Retrieval*. 57–62.
[90] Yulan Guo, Ferdous Sohel, Mohammed Bennamoun, Min Lu, and Jianwei Wan. 2013. Rotational projection statistics for 3D local surface description and object recognition. *Int. J. Comput. Vis.* 105, 1 (2013), 63–86.
[91] Jiaqi Yang, Qian Zhang, Yang Xiao, and Zhiguo Cao. 2017. TOLDI: An effective and robust approach for 3D local shape description. *Pattern Recogn.* 65 (2017), 175–187.
[92] Tiecheng Sun, Guanghui Liu, Shuaicheng Liu, Fanman Meng, Liaoyuan Zeng, and Ru Li. 2020. An efficient and compact 3D local descriptor based on the weighted height image. *Inf. Sci.* 520 (2020), 209–231.
[93] Bastian Steder, Giorgio Grisetti, and Wolfram Burgard. 2010. Robust place recognition for 3D range data based on point features. In *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 1400–1405.
[94] Yan Zhuang, Nan Jiang, Huosheng Hu, and Fei Yan. 2013. 3-D-laser-based scene measurement and place recognition for mobile robots in dynamic indoor environments. *IEEE Trans. Instrum. Meas.* 62, 2 (2013), 438–450.
[95] Fengkui Cao, Yan Zhuang, Hong Zhang, and Wei Wang. 2018. Robust place recognition and loop closing in laser-based SLAM for UGVs in urban environments. *IEEE Sens. J.* 18, 10 (2018), 4242–4252.
[96] Tixiao Shan, Brendan Englot, Fábio Duarte, Carlo Ratti, and Daniela Rus. 2021. Robust place recognition using an imaging lidar. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'21)*. 5469–5475.
[97] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 2015. 3D ShapeNets: A deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)*. 1912–1920.
[98] Andy Zeng, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 2017. 3DMatch: Learning local geometric descriptors from RGB-D reconstructions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17)*. 199–208.
[99] Zan Gojcic, Caifa Zhou, Jan D. Wegner, and Andreas Wieser. 2019. The perfect match: 3D point cloud matching with smoothed densities. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'19)*. 5540–5549.
[100] Sheng Ao, Qingyong Hu, Bo Yang, Andrew Markham, and Yulan Guo. 2021. Spinnet: Learning a general surface descriptor for 3d point cloud registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11753–11762.
[101] Sheng Ao, Yulan Guo, Qingyong Hu, Bo Yang, Andrew Markham, and Zengping Chen. 2023. You only train once: Learning general and distinctive 3D local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 3 (2023), 3949–3967.
[102] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 652–660.

[103] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems*, Vol. 30.

[104] Marc Khoury, Qian-Yi Zhou, and Vladlen Koltun. 2017. Learning compact geometric features. In *Proceedings of the IEEE International Conference on Computer Vision*. 153–161.

[105] Haowen Deng, Tolga Birdal, and Slobodan Ilic. 2018. PPFNet: Global context aware local features for robust 3D point matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 195–205.

[106] Haowen Deng, Tolga Birdal, and Slobodan Ilic. 2018. Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors. In *Proceedings of the European Conference on Computer Vision (ECCV'18)*. 602–618.

[107] Weixin Lu, Guowei Wan, Yao Zhou, Xiangyu Fu, Pengfei Yuan, and Shiyu Song. 2019. Deepvcp: An end-to-end deep neural network for point cloud registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12–21.

[108] Haowen Deng, Tolga Birdal, and Slobodan Ilic. 2019. 3d local features for direct pairwise registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3244–3253.

[109] Youjie Zhou, Yiming Wang, Fabio Poiesi, Qi Qin, and Yi Wan. 2022. Loop closure detection using local 3D deep descriptors. *IEEE Robot. Autom. Lett.* 7, 3 (2022), 6335–6342.

[110] Fabio Poiesi and Davide Boscaini. 2022. Learning general and distinctive 3D local deep descriptors for point cloud registration. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 3 (2022), 3979–3985.

[111] Marlon Marcon, Riccardo Spezialetti, Samuele Salti, Luciano Silva, and Luigi Di Stefano. 2022. Unsupervised learning of local equivariant descriptors for point clouds. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 12 (2022), 9687–9702.

[112] Gil Elbaz, Tamar Avraham, and Anath Fischer. 2017. 3D point cloud registration for localization using a deep neural network auto-encoder. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4631–4640.

[113] Lei Zhou, Siyu Zhu, Zixin Luo, Tianwei Shen, Runze Zhang, Mingmin Zhen, Tian Fang, and Long Quan. 2018. Learning and matching multi-view descriptors for registration of point clouds. In *Proceedings of the European Conference on Computer Vision (ECCV'18)*. 505–522.

[114] Lei Li, Siyu Zhu, Hongbo Fu, Ping Tan, and Chiew-Lan Tai. 2020. End-to-end learning local multi-view descriptors for 3D point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'20)*. 1916–1925.

[115] Zan Gojcic, Caifa Zhou, Jan D. Wegner, Leonidas J. Guibas, and Tolga Birdal. 2020. Learning multiview 3D point cloud registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'20)*. 1756–1766.

[116] Haodong Xiang, Xiaosheng Zhu, Wenzhong Shi, Wenzheng Fan, Pengxin Chen, and Sheng Bao. 2022. DeLightLCD: A deep and lightweight network for loop closure detection in LiDAR SLAM. *IEEE Sens. J.* 22, 21 (2022), 20761–20772.

[117] Bastian Steder, Giorgio Grisetti, Mark Van Loock, and Wolfram Burgard. 2009. Robust on-line model-based object detection from range images. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 4739–4744.

[118] Edward Rosten and Tom Drummond. 2006. Machine learning for high-speed corner detection. In *Proceedings of the 9th European Conference on Computer Vision Part I 9*. Springer, 430–443.

[119] Yulan Guo, Mohammed Bennamoun, Ferdous Sohel, Min Lu, and Jianwei Wan. 2014. 3D object recognition in cluttered scenes with local surface features: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 11 (2014), 2270–2287.

[120] Yongzhi Fan, Xin Du, Lun Luo, and Jizhong Shen. 2022. FreSCo: Frequency-domain scan context for LiDAR-based place recognition with translation and rotation invariance. In *Proceedings of the 17th International Conference on Control, Automation, Robotics and Vision (ICARCV'22)*. IEEE, 576–583.

[121] Wuqi Wang, Haigen Min, Xia Wu, Xinmeng Hou, Yao Li, and Xiangmo Zhao. 2023. High accuracy and low complexity LiDAR place recognition using unitary invariant frobenius norm. *IEEE Sens. J.* 23, 11 (2023), 11205–11217.

[122] Fang Ou, Yunhui Li, and Zhonghua Miao. 2023. Place recognition of large-scale unstructured orchards with attention score maps. *IEEE Robot. Autom. Lett.* 8, 2 (2023), 958–965.

[123] Sha Lu, Xuecheng Xu, Huan Yin, Zexi Chen, Rong Xiong, and Yue Wang. 2022. One RING to rule them all: Radon sinogram for place recognition, orientation and translation estimation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'22)*. 2778–2785.

[124] Xuecheng Xu, Sha Lu, Jun Wu, Haojian Lu, Qiuguo Zhu, Yiyi Liao, Rong Xiong, and Yue Wang. 2023. RING++: Roto-translation invariant gram for global localization on a sparse scan map. *IEEE Trans. Robot.* 39, 6 (2023), 4616–4635.

[125] Martin Magnusson, Henrik Andreasson, Andreas Nuchter, and Achim J. Lilienthal. 2009. Appearance-based loop detection from 3D laser data using the normal distributions transform. In *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 23–28.

[126] Fengkui Cao, Fei Yan, Sen Wang, Yan Zhuang, and Wei Wang. 2021. Season-invariant and viewpoint-tolerant LiDAR place recognition in GPS-denied environments. *IEEE Trans. Industr. Electr.* 68, 1 (2021), 563–574.

[127] Konrad P. Cop, Paulo V. K. Borges, and Renaud Dubé. 2018. Delight: An efficient descriptor for global localisation using lidar intensities. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'18)*. IEEE, 3653–3660.

[128] Jiawei Mo and Junaed Sattar. 2020. A fast and robust place recognition approach for stereo visual odometry using LiDAR descriptors. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'20)*. IEEE, 5893–5900.

[129] Li He, Xiaolong Wang, and Hong Zhang. 2016. M2DP: A novel 3D point cloud descriptor and its application in loop closure detection. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'16)*. IEEE, 231–237.

[130] Leonardo Perdomo, Diego Pittol, Mathias Mantelli, Renan Maffei, Mariana Kolberg, and Edson Prestes. 2019. c-M2DP: A fast point cloud descriptor with color information to perform loop closure detection. In *Proceedings of the IEEE 15th International Conference on Automation Science and Engineering (CASE'19)*. IEEE, 1145–1150.

[131] Yan Xia, Yusheng Xu, Shuang Li, Rui Wang, Juan Du, Daniel Cremers, and Uwe Stilla. 2021. SOE-Net: A self-attention and orientation encoding network for point cloud based place recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'21)*. IEEE, 11343–11352.

[132] Daniele Cattaneo, Matteo Vaghi, and Abhinav Valada. 2022. LCDNet: Deep loop closure detection and point cloud registration for LiDAR SLAM. *IEEE Trans. Robot.* 38, 4 (2022), 2074–2093.

[133] Zhaoxin Fan, Hongyan Liu, Jun He, Qi Sun, and Xiaoyong Du. 2020. Srnet: A 3d scene recognition network using static graph and dense semantic fusion. In *Computer Graphics Forum*, Vol. 39. Wiley Online Library, 301–311.

[134] Zhijian Qiao, Hanjiang Hu, Weiang Shi, Siyuan Chen, Zhe Liu, and Hesheng Wang. 2021. A registration-aided domain adaptation network for 3D point cloud based place recognition. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'21)*. IEEE, 1317–1322.

[135] Le Hui, Hang Yang, Mingmei Cheng, Jin Xie, and Jian Yang. 2021. Pyramid point cloud transformer for large-scale place recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 6098–6107.

[136] Jacek Komorowski. 2021. Minkloc3d: Point cloud based large-scale place recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1790–1799.

[137] Jacek Komorowski, Monika Wysoczańska, and Tomasz Trzcinski. 2021. MinkLoc++: Lidar and monocular image fusion for place recognition. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN'21)*. 1–8.

[138] Jacek Komorowski, Monika Wysoczanska, and Tomasz Trzcinski. 2021. Egonn: Egocentric neural network for point cloud based 6dof relocalization at the city scale. *IEEE Robot. Autom. Lett.* 7, 2 (2021), 722–729.

[139] Tian-Xing Xu, Yuan-Chen Guo, Yu-Kun Lai, and Song-Hai Zhang. 2021. TransLoc3D: Point cloud based large-scale place recognition using adaptive receptive fields. arXiv:2105.11605. Retrieved from https://arxiv.org/abs/2105.11605

[140] Jacek Komorowski. 2022. Improving point cloud based place recognition with ranking-based loss and large batch training. In *Proceedings of the 26th International Conference on Pattern Recognition (ICPR'22)*. IEEE, 3699–3705.

[141] Kamil Żywanowski, Adam Banaszczyk, Michał R. Nowicki, and Jacek Komorowski. 2022. MinkLoc3D-SI: 3D LiDAR place recognition with sparse convolutions, spherical coordinates, and intensity. *IEEE Robot. Autom. Lett.* 7, 2 (2022), 1079–1086.

[142] Kavisha Vidanapathirana, Milad Ramezani, Peyman Moghadam, Sridha Sridharan, and Clinton Fookes. 2022. LoGG3D-Net: Locally guided global descriptor learning for 3D place recognition. In *Proceedings of the International Conference on Robotics and Automation (ICRA'22)*. IEEE, 2215–2221.

[143] Min Young Chang, Suyong Yeon, Soohyun Ryu, and Donghwan Lee. 2020. Spoxelnet: Spherical voxel-based deep place recognition for 3d point clouds of crowded indoor spaces. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'20)*. IEEE, 8564–8570.

[144] Sriram Siva, Zachary Nahman, and Hao Zhang. 2020. Voxel-based representation learning for place recognition based on 3D point clouds. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'20)*. 8351–8357.

[145] Zhixing Hou, Yan Yan, Chengzhong Xu, and Hui Kong. 2022. HiTPR: Hierarchical transformer for place recognition in point cloud. In *Proceedings of the International Conference on Robotics and Automation (ICRA'22)*. IEEE, 2612–2618.

[146] Zhicheng Zhou, Cheng Zhao, Daniel Adolfsson, Songzhi Su, Yang Gao, Tom Duckett, and Li Sun. 2021. Ndt-transformer: Large-scale 3d point cloud localisation using the normal distribution transform representation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'21)*. IEEE, 5654–5660.

[147] Huan Yin, Xiaqing Ding, Li Tang, Yue Wang, and Rong Xiong. 2017. Efficient 3D LIDAR based loop closing using deep neural network. In *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO'17)*. IEEE, 481–486.

[148] Dong Kong, Xu Li, Yanqing Cen, Qimin Xu, and Aimin Wang. 2023. Simultaneous viewpoint- and condition-invariant loop closure detection based on LiDAR descriptor for outdoor large-scale environments. *IEEE Trans. Industr. Electr.* 70, 2 (2023), 2117–2127.

[149] Junyi Ma, Xieyuanli Chen, Jingyi Xu, and Guangming Xiong. 2023. SeqOT: A spatial-temporal transformer network for place recognition using sequential LiDAR data. *IEEE Trans. Industr. Electr.* 70, 8 (2023), 8225–8234.

[150] Junyi Ma, Jun Zhang, Jintao Xu, Rui Ai, Weihao Gu, and Xieyuanli Chen. 2022. OverlapTransformer: An efficient and yaw-angle-invariant transformer network for LiDAR-based place recognition. *IEEE Robot. Autom. Lett.* 7, 3 (2022), 6958–6965.

[151] Tiago Barros, Luis Garrote, Ricardo Pereira, Cristiano Premebida, and Urbano J. Nunes. 2023. AttDLNet: Attention-based deep network for 3d lidar place recognition. In *Proceedings of the Fifth Iberian Robotics Conference*. Springer, 309–320.

[152] Xuecheng Xu, Huan Yin, Zexi Chen, Yuehua Li, Yue Wang, and Rong Xiong. 2021. Disco: Differentiable scan context with orientation. *IEEE Robot. Autom. Lett.* 6, 2 (2021), 2791–2798.

[153] Martin Magnusson, Tomasz Piotr Kucner, Saeed Gholami Shahbandi, Henrik Andreasson, and Achim J Lilienthal. 2017. Semi-supervised 3d place categorisation by descriptor clustering. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'17)*. IEEE, 620–625.

[154] Relja Arandjelovic, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic. 2016. NetVLAD: CNN architecture for weakly supervised place recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5297–5307.

[155] Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. 2020. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10529–10538.

[156] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2117–2125.

[157] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2117–2125.

[158] Haodong Xiang, Wenzhong Shi, Wenzheng Fan, Pengxin Chen, Sheng Bao, and Mingyan Nie. 2021. FastLCD: A fast and compact loop closure detection approach using 3D point cloud for indoor mobile mapping. *Int. J. Appl. Earth Observ. Geoinf.* 102 (2021), 102430.

[159] Tim-Lukas Habich, Marvin Stuede, Mathieu Labbé, and Svenja Spindeldreier. 2021. Have i been here before? Learning to close the loop with LiDAR data in graph-based SLAM. In *Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM'21)*. IEEE, 504–510.

[160] Renaud Dubé, Daniel Dugas, Elena Stumm, Juan Nieto, Roland Siegwart, and Cesar Cadena. 2017. SegMatch: Segment based place recognition in 3D point clouds. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'17)*. 5266–5272.

[161] Renaud Dube, Andre Cramariuc, Daniel Dugas, Juan Nieto, Roland Siegwart, and Cesar Cadena. 2018. SegMap: 3D segment mapping using data-driven descriptors. In *Robotics: Science and Systems*.

[162] Kavisha Vidanapathirana, Peyman Moghadam, Ben Harwood, Muming Zhao, Sridha Sridharan, and Clinton Fookes. 2021. Locus: Lidar-based place recognition using spatiotemporal higher-order pooling. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'21)*. IEEE, 5075–5081.

[163] Martin A. Fischler and Robert C. Bolles. 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 6 (1981), 381–395.

[164] Renaud Dubé, Mattia G. Gollub, Hannes Sommer, Igor Gilitschenski, Roland Siegwart, Cesar Cadena, and Juan Nieto. 2018. Incremental-segment-based localization in 3-D point clouds. *IEEE Robot. Autom. Lett.* 3, 3 (2018), 1832–1839.

[165] Xingliang Ji, Lin Zuo, Changhua Zhang, and Yu Liu. 2019. Lloam: Lidar odometry and mapping with loop-closure detection based correction. In *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA'19)*. IEEE, 2475–2480.

[166] Renaud Dubé, Abel Gawel, Hannes Sommer, Juan Nieto, Roland Siegwart, and Cesar Cadena. 2017. An online multi-robot SLAM system for 3D LiDARs. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'17)*. 1004–1011.

[167] Yuting Xie, Yachen Zhang, Long Chen, Hui Cheng, Wei Tu, Dongpu Cao, and Qingquan Li. 2021. RDC-SLAM: A real-time distributed cooperative SLAM system based on 3D LiDAR. *IEEE Trans. Intell. Transport. Syst.* (2021).

[168] Yansong Gong, Fengchi Sun, Jing Yuan, Wenbin Zhu, and Qinxuan Sun. 2021. A two-level framework for place recognition with 3D LiDAR based on spatial relation graph. *Pattern Recogn.* 120 (2021), 108171.

[169] Yunfeng Fan, Yichang He, and U-Xuan Tan. 2020. Seed: A segmentation-based egocentric 3D point cloud descriptor for loop closure detection. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'20)*. IEEE, 5158–5163.

[170] Georgi Tinchev, Simona Nobili, and Maurice Fallon. 2018. Seeing the wood for the trees: Reliable localization in urban and natural environments. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'18)*. 8239–8246.

[171] Georgi Tinchev, Adrian Penate-Sanchez, and Maurice Fallon. 2019. Learning to see the wood for the trees: Deep laser localization in urban and natural environments on a CPU. *IEEE Robot. Autom. Lett.* 4, 2 (2019), 1327–1334.

[172] Dávid Rozenberszki and András L. Majdik. 2020. LOL: Lidar-only odometry and localization in 3D point cloud maps. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'20)*. 4379–4385.

[173] Andrei Cramariuc, Florian Tschopp, Nikhilesh Alatur, Stefan Benz, Tillmann Falck, Marius Brühlmeier, Benjamin Hahn, Juan Nieto, and Roland Siegwart. 2021. Semsegmap–3d segment-based semantic localization. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'21)*. IEEE, 1183–1190.

[174] Jan Wietrzykowski and Piotr Skrzypczyński. 2021. On the descriptive power of LiDAR intensity images for segment-based loop closing in 3-D SLAM. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'21)*. 79–85.

[175] Lin Li, Xin Kong, Xiangrui Zhao, Tianxin Huang, Wanlong Li, Feng Wen, Hongbo Zhang, and Yong Liu. 2021. SSC: Semantic scan context for large-scale place recognition. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'21)*. IEEE, 2092–2099.

[176] Lin Li, Xin Kong, Xiangrui Zhao, Wanlong Li, Feng Wen, Hongbo Zhang, and Yong Liu. 2021. SA-LOAM: Semantic-aided LiDAR SLAM with loop closure. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'21)*. IEEE, 7627–7634.

[177] Georgi Pramatarov, Daniele De Martini, Matthew Gadd, and Paul Newman. 2022. BoxGraph: Semantic place recognition and pose estimation from 3D LiDAR. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'22)*. IEEE, 7004–2011.

[178] Haodong Yuan, Yudong Zhang, Shengyin Fan, Xue Li, and Jian Wang. 2022. Object scan context: Object-centric spatial descriptor for place recognition within 3D point cloud map. arXiv:2206.03062. Retrieved from https://arxiv.org/abs/2206.03062

[179] Anestis Zaganidis, Alexandros Zerntev, Tom Duckett, and Grzegorz Cielniak. 2019. Semantically assisted loop closure in slam using ndt histograms. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'19)*. IEEE, 4562–4568.

[180] Martin Magnusson, Henrik Andreasson, Andreas Nuchter, and Achim J Lilienthal. 2009. Automatic appearance-based loop detection from three-dimensional laser data using the normal distributions transform. *J. Field Robot.* 26, 11-12 (2009), 892–914.

[181] Lin Li, Xin Kong, Xiangrui Zhao, Tianxin Huang, Wanlong Li, Feng Wen, Hongbo Zhang, and Yong Liu. 2022. RINet: Efficient 3D lidar-based place recognition using rotation invariant neural network. *IEEE Robot. Autom. Lett.* 7, 2 (2022), 4321–4328.

[182] Lukas Bernreiter, Abel Gawel, Hannes Sommer, Juan Nieto, Roland Siegwart, and Cesar Cadena Lerma. 2019. Multiple hypothesis semantic mapping for robust data association. *IEEE Robot. Autom. Lett.* 4, 4 (2019), 3255–3262.

[183] Peng Yin, Lingyun Xu, Ziyue Feng, Anton Egorov, and Bing Li. 2022. PSE-Match: A viewpoint-free place recognition method with parallel semantic embedding. *IEEE Trans. Intell. Transport. Syst.* 23, 8 (2022), 11249–11260.

[184] Deyun Dai, Jikai Wang, Zonghai Chen, and Peng Bao. 2022. SC-LPR: Spatiotemporal context based LiDAR place recognition. *Pattern Recogn. Lett.* 156 (2022), 160–166.

[185] Jing Tao, Qin Ye, and Pengcheng Shi. 2021. A novel robust point cloud registration method based on directional feature weighted constraint. In *Proceedings of the 12th International Conference on Information Optics and Photonics*, Vol. 12057. SPIE, 1123–1130.

[186] Pengcheng Shi, Shaocheng Yan, Yilin Xiao, Xinyi Liu, Yongjun Zhang, and Jiayuan Li. 2024. RANSAC back to SOTA: A two-stage consensus filtering for real-time 3D registration. *IEEE Robot. Autom. Lett.* 9, 12 (2024), 11881–11888. DOI: http://doi.org/10.1109/LRA.2024.3502056

[187] Nils Rottmann, Ralf Bruder, Achim Schweikard, and Elmar Rueckert. 2019. Loop closure detection in closed environments. In *Proceedings of the European Conference on Mobile Robots (ECMR'19)*. IEEE, 1–8.

[188] Chenglu Wen, Yudi Dai, Yan Xia, Yuhan Lian, Jinbin Tan, Cheng Wang, and Jonathan Li. 2020. Toward efficient 3-D colored mapping in GPS-/GNSS-Denied environments. *IEEE Geosci. Remote Sens. Lett.* 17, 1 (2020), 147–151.

[189] Bo Zhou, Yi He, Kun Qian, Xudong Ma, and Xiaomao Li. 2021. S4-SLAM: A real-time 3D LIDAR SLAM system for ground/watersurface multi-scene outdoor applications. *Auton. Robot.* 45, 1 (2021), 77–98.

[190] Ellon Mendes, Pierrick Koch, and Simon Lacroix. 2016. ICP-based pose-graph SLAM. In *Proceedings of the IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR'16)*. IEEE, 195–200.

[191] P. J. Besl and Neil D. McKay. 1992. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* 14, 2 (1992), 239–256.
[192] Yi Zhang. 2022. LILO: A novel Lidar-IMU SLAM System with Loop Optimization. *IEEE Trans. Aerosp. Electron. Syst.* 58, 4 (2022), 2649–2659.
[193] Jianwen Jiang, Jikai Wang, Peng Wang, Peng Bao, and Zonghai Chen. 2020. LiPMatch: LiDAR point cloud plane based loop-closure. *IEEE Robot. Autom. Lett.* 5, 4 (2020), 6861–6868.
[194] Masashi Yokozuka, Kenji Koide, Shuji Oishi, and Atsuhiko Banno. 2021. LiTAMIN2: Ultra Light LiDAR-based SLAM using Geometric Approximation applied with KL-Divergence. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'21)*. 11619–11625.
[195] Kamak Ebadi, Yun Chang, Matteo Palieri, Alex Stephens, Alex Hatteland, Eric Heiden, Abhishek Thakur, Nobuhiro Funabiki, Benjamin Morrell, Sally Wood, Luca Carlone, and Ali-akbar Agha-mohammadi. 2020. LAMP: Large-scale autonomous mapping and positioning for exploration of perceptually-degraded subterranean environments. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'20)*. 80–86.
[196] Aleksandr Segal, Dirk Haehnel, and Sebastian Thrun. 2009. Generalized-icp. In *Robotics: Science and Systems*, Vol. 2. Seattle, WA, 435.
[197] Michael J. Milford and Gordon F. Wyeth. 2012. SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights. In *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 1643–1649.
[198] Zhe Liu, Chuanzhe Suo, Shunbo Zhou, Fan Xu, Huanshu Wei, Wen Chen, Hesheng Wang, Xinwu Liang, and Yun-Hui Liu. 2019. Seqlpd: Sequence matching enhanced loop-closure detection based on large-scale point cloud description for self-driving vehicles. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'19)*. IEEE, 1218–1223.
[199] Peng Yin, Fuying Wang, Anton Egorov, Jiafan Hou, Ji Zhang, and Howie Choset. 2020. SeqSphereVLAD: Sequence matching enhanced orientation-invariant place recognition. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'20)*. 5024–5029.
[200] Peng Yin, Lingyun Xu, Ji Zhang, and Howie Choset. 2021. Fusionvlad: A multi-view deep fusion networks for viewpoint-free 3d place recognition. *IEEE Robot. Autom. Lett.* 6, 2 (2021), 2304–2310.
[201] Thomas Schmiedel, Erik Einhorn, and Horst-Michael Gross. 2015. IRON: A fast interest point descriptor for robust NDT-map matching and its application to robot localization. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'15)*. IEEE, 3144–3151.
[202] Pengcheng Shi, Jiayuan Li, and Yongjun Zhang. 2023. A fast LiDAR place recognition and localization method by fusing local and global search. *ISPRS J. Photogram. Remote Sens.* 202 (2023), 637–651.
[203] Mohammad Aldibaja, Naoki Suganuma, Ryo Yanase, Lu Cao, Keisuke Yoneda, and Akisue Kuramoto. 2020. Loop-closure and map-combiner detection strategy based on LIDAR reflectance and elevation maps. In *Proceedings of the IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC'20)*. IEEE, 1–7.
[204] Heng Yang, Jingnan Shi, and Luca Carlone. 2021. TEASER: Fast and certifiable point cloud registration. *IEEE Trans. Robot.* 37, 2 (2021), 314–333.
[205] Zhongze Liu, Huiyan Chen, Huijun Di, Yi Tao, Jianwei Gong, Guangming Xiong, and Jianyong Qi. 2018. Real-time 6d lidar slam in large scale natural terrains for ugv. In *Proceedings of the IEEE Intelligent Vehicles Symposium (IV'18)*. IEEE, 662–667.
[206] Ioan Andrei Barsan, Shenlong Wang, Andrei Pokrovsky, and Raquel Urtasun. 2018. Learning to localize using a LiDAR intensity map. In *Conference on Robot Learning*. PMLR, 605–616.
[207] Huan Yin, Yue Wang, Xiaqing Ding, Li Tang, Shoudong Huang, and Rong Xiong. 2020. 3d lidar-based global localization using siamese neural network. *IEEE Trans. Intell. Transport. Syst.* 21, 4 (2020), 1380–1392.
[208] Li Sun, Daniel Adolfsson, Martin Magnusson, Henrik Andreasson, Ingmar Posner, and Tom Duckett. 2020. Localising faster: Efficient and precise lidar-based robot localisation in large-scale environments. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'20)*. IEEE, 4386–4392.
[209] Weixin Lu, Yao Zhou, Guowei Wan, Shenhua Hou, and Shiyu Song. 2019. L3-net: Towards learning based lidar localization for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6389–6398.
[210] Louis Wiesmann, Rodrigo Marcuzzi, Cyrill Stachniss, and Jens Behley. 2022. Retriever: Point cloud retrieval in compressed 3D maps. In *Proceedings of the International Conference on Robotics and Automation (ICRA'22)*. 10925–10932.
[211] Bo Zhou, Yi He, Wenchao Huang, Xiang Yu, Fang Fang, and Xiaomao Li. 2022. Place recognition and navigation of outdoor mobile robots based on random Forest learning with a 3D LiDAR. *J. Intell. Robot. Syst.* 104, 4 (2022), 1–26.
[212] Benjamin Suger and Wolfram Burgard. 2017. Global outer-urban navigation with openstreetmap. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'17)*. IEEE, 1417–1422.

[213] Fan Yan, Olga Vysotska, and Cyrill Stachniss. 2019. Global localization on openstreetmap using 4-bit semantic descriptors. In *Proceedings of the European Conference on Mobile Robots (ECMR'19)*. IEEE, 1–7.
[214] Younghun Cho, Giseop Kim, Sangmin Lee, and Jee-Hwan Ryu. 2022. Openstreetmap-based lidar global localization in urban environment without a prior lidar map. *IEEE Robot. Autom. Lett.* 7, 2 (2022), 4999–5006.
[215] Jens Behley and Cyrill Stachniss. 2018. Efficient surfel-based SLAM using 3D laser range data in urban environments.. In *Robotics: Science and Systems*, Vol. 2018. 59.
[216] Peng Yin, Lingyun Xu, Zhe Liu, Lu Li, Hadi Salman, Yuqing He, Weiliang Xu, Hesheng Wang, and Howie Choset. 2018. Stabilize an unsupervised feature learning for LiDAR-based place recognition. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'18)*. IEEE, 1162–1167.
[217] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.
[218] Joshua Knights, Peyman Moghadam, Milad Ramezani, Sridha Sridharan, and Clinton Fookes. 2022. Incloud: Incremental learning for point cloud place recognition. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'22)*. IEEE, 8559–8566.
[219] Karl Granström and Thomas B. Schön. 2010. Learning to close the loop from 3D point clouds. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2089–2095.
[220] Gaurav Pandey, James R McBride, and Ryan M Eustice. 2011. Ford campus vision and lidar data set. *Int. J. Robot. Res.* 30, 13 (2011), 1543–1552.
[221] Andreas Geiger, Philip Lenz, and Raquel Urtasun. 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3354–3361.
[222] Nicholas Carlevaris-Bianco, Arash K. Ushani, and Ryan M. Eustice. 2016. University of Michigan North Campus long-term vision and lidar dataset. *Int. J. Robot. Res.* 35, 9 (2016), 1023–1035.
[223] Will Maddern, Geoffrey Pascoe, Chris Linegar, and Paul Newman. 2017. 1 year, 1000 km: The Oxford RobotCar dataset. *Int. J. Robot. Res.* 36, 1 (2017), 3–15.
[224] Jinyong Jeong, Younggun Cho, Young-Sik Shin, Hyunchul Roh, and Ayoung Kim. 2018. Complex urban LiDAR data set. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'18)*. 6344–6351.
[225] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. 2019. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9297–9307.
[226] Giseop Kim, Yeong Sang Park, Younghun Cho, Jinyong Jeong, and Ayoung Kim. 2020. MulRan: Multimodal range dataset for urban place recognition. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'20)*. 6246–6253.
[227] Wei Zhou, Julie Stephany Berrio, Charika De Alvis, Mao Shan, Stewart Worrall, James Ward, and Eduardo Nebot. 2020. Developing and testing robust autonomy: The University of Sydney campus data set. *IEEE Intell. Transport. Syst. Mag.* 12, 4 (2020), 23–40.
[228] Dan Barnes, Matthew Gadd, Paul Murcutt, Paul Newman, and Ingmar Posner. 2020. The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'20)*. IEEE, 6433–6438.
[229] Yiyi Liao, Jun Xie, and Andreas Geiger. 2023. KITTI-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 3 (2023), 3292–3310.
[230] Dorian Gálvez-López and Juan D Tardos. 2012. Bags of binary words for fast place recognition in image sequences. *IEEE Trans. Robot.* 28, 5 (2012), 1188–1197.
[231] Bruno Ferrarini, Maria Waheed, Sania Waheed, Shoaib Ehsan, Michael J Milford, and Klaus D McDonald-Maier. 2020. Exploring performance bounds of visual place recognition using extended precision. *IEEE Robot. Autom. Lett.* 5, 2 (2020), 1688–1695.

Received 7 September 2023; revised 12 November 2024; accepted 28 November 2024