2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Reliable LiDAR Loop Detection through Structural Descriptors and Semantic Graph Matching

Yujie Tang[1†], Sibo Zuo[1†], Meiling Wang[1], Jianyu Dou[1], Jiahui Wang[1], Yufeng Yue[1*]

***Abstract*— Outdoor loop closure detection is essential for mitigating accumulated drift in SLAM and generating a global consistent map. Semantic graph matching methods utilize object-level topology for distinctive scene representation but rely on environments with rich and distinguishable objects. Moreover, accurately matching nodes remains difficult due to ambiguities among same-class semantic nodes. These challenges limit their effectiveness in varied road environments, highlighting the need for representations that are both robust and adaptable. To address this, we introduce SD-SGM, a novel loop closure detection framework combining the powerful context-adaptation capabilities of structural descriptors with the high-level semantic reasoning abilities of semantic graphs. Initially, we extract semantic graphs alongside global structural descriptors from point clouds. Distinctive local graph features are then used to generate candidate node pairs, and the maximal clique algorithm identifies correspondences that are globally consistent. The similarity scores of both methods are then evaluated and a cross-validation mechanism assesses their reliability and adaptively weights them. Extensive loop closure detection experiments on various datasets demonstrate that SD-SGM achieves state-of-the-art (SOTA) performance compared to strong baselines. Additionally, we verify its effectiveness in improving SLAM trajectory accuracy. We provide the code at: `https://github.com/BIT-TYJ/SD-SGM`.**

I. Introduction

In outdoor environments, robots often perform tasks such as navigation [1], [2] and rescue [3], where accurate mapping [4]–[8] and reliable self-localization are critical. This requires a robust LiDAR SLAM framework [9]–[11]. However, in long-term deployments, state estimation errors inevitably accumulate. Loop closure detection [12], [13] is essential for correcting accumulated drift [14], [15], by recognizing previously visited places [16], [17].

Existing LiDAR loop closure detection methods fall mainly into two categories: (geometric) structural descriptors [18]–[27], and semantic graph matching [28]–[35], which extract point cloud features to assess scene similarity. Structural descriptors function across diverse environments but are constrained by LiDAR sparsity, noise, and lack of semantic awareness. Semantic graph matching provides compact, object-level scene representations more robust to sparsity and noise, yet performs well only in object-rich environments. Due to their complementary advantages, several works have explored combining the two representations to enhance the robustness of loop closure detection. SectionKey [31] and SGLC [33] mainly perform loop closure detection through semantic graph matching, with structural descriptors used for auxiliary verification. Although beneficial, such hybrid methods often exhibit limited interaction between the two modalities or rely on heuristic fusion strategies.

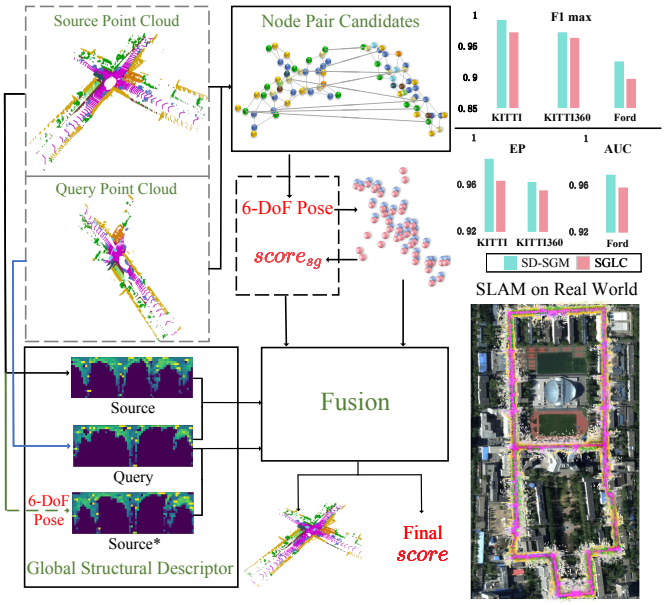

Fig. 1: Given two LiDAR point clouds from outdoor road environments, we construct and compare their semantic node graphs and global structural descriptors, then fuse the comparison results to produce the final similarity score and 6-DoF pose estimation. Our method, **SD-SGM**, outperforms the novel method SGLC in loop detection metrics across different datasets. We also demonstrate the consistent semantic map optimized by SD-SGM after SLAM trajectory correction on the real world campus.

This raises the question: Can semantic graph–based and structural descriptor–based methods be combined in a principled way to fully exploit their complementary strengths? Our objective is to develop a loop closure detection framework adaptable to diverse scenes while ensuring strong scene discrimination. However, the following challenges are faced: **1. Due to the use of different similarity metrics, the scores produced by the two methods are not directly**

This work is supported by the National Natural Science Foundation of China under Grant No. NSFC 62473050, 92370203.(Corresponding Author: Yufeng Yue, yueyufeng@bit.edu.cn)

[1]Yujie Tang, Sibo Zuo, Meiling Wang, Jianyu Dou, Jiahui Wang, Yufeng Yue are with School of Automation, Beijing Institute of Technology, Beijing, 100081, China.

†: Equal contribution.

∗: Corresponding Author.

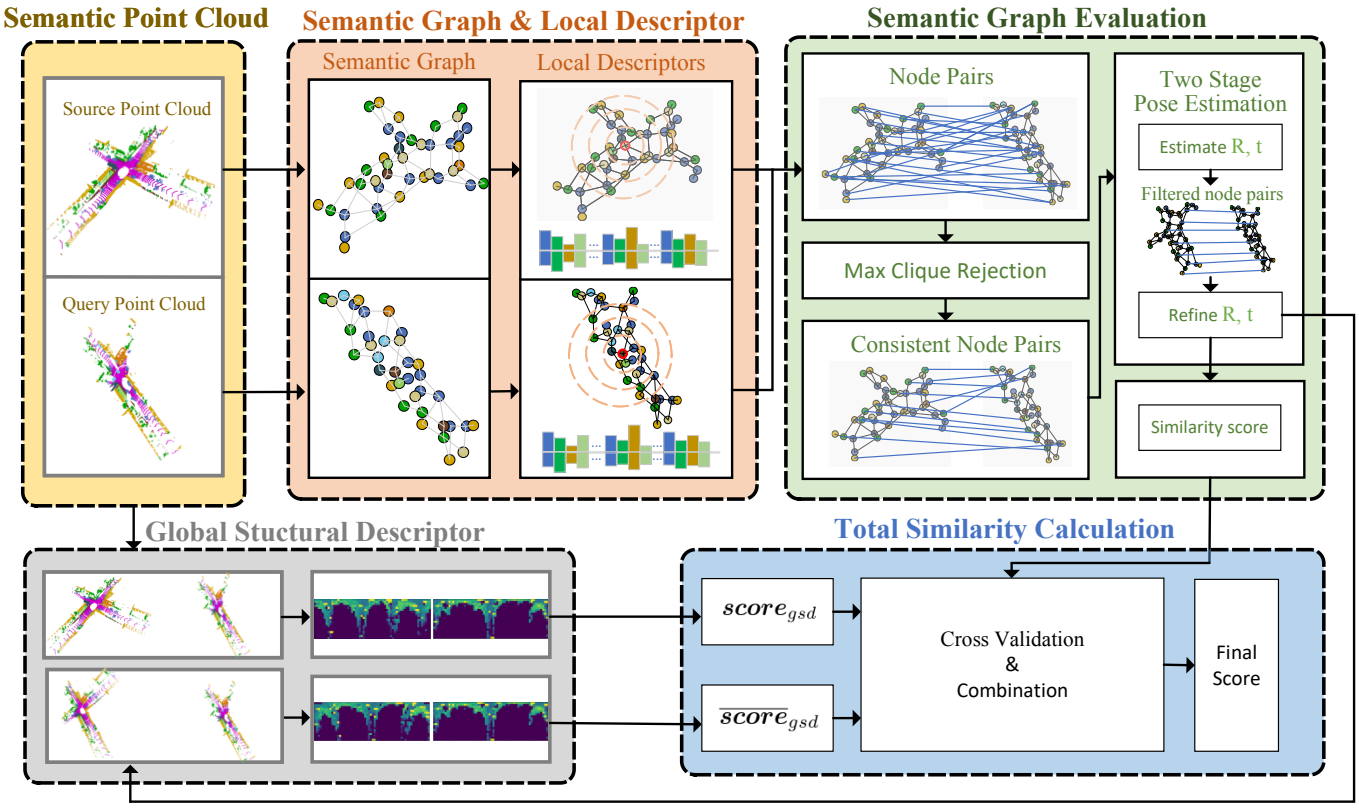

Fig. 2: The framework of SD-SGM consists of three models: Semantic graph matching (pictures above), global structural descriptor comparison (the bottom left), and the fusion model. Semantic graphs are extracted for node correspondence via local features and a maximal clique algorithm, followed by two-stage pose estimation to yield a 6-DoF pose and similarity score. In parallel, global structural descriptors provide another score, and both are fused with intermediate variables to obtain the final similarity score.

comparable, which complicates the assignment of reliable confidence weights. To guide the reliability assessment, we institute a cross-validation mechanism to evaluate their loop closure detection results. Intermediate variables such as node richness and yaw discrepancy are further leveraged for adaptive weighting of the two modalities. **2. Reliably resolving node correspondences in semantic graphs remains challenging in the presence of repetitive structures, hindering accurate graph alignment and loop pose estimation.** To address this, multi-scale local graph features are used to identify high-quality candidate node pairs, while outlier rejection is formulated as a maximum clique problem to ensure globally consistent one-to-one correspondences.

Based on the above analysis, we propose SD-SGM, a loop closure detection framework that fuses semantic graphs with global structural descriptors for robust and accurate detection in outdoor road environments. As shown in Fig. 1, semantic graphs and structural descriptors are extracted from point clouds, high-quality node correspondences enable accurate matching, and an adaptive evaluation balances their contributions to produce a unified similarity score and 6-DoF relative pose. Comparisons with the state-of-the-art method demonstrate the superiority and effectiveness of our approach. Our main contributions are summarized as follows.

- A SOTA LiDAR loop closure detection method is proposed, achieving superior performance across diverse datasets.
- We propose an adaptive fusion framework that integrates semantic graph matching and global structural descriptors, where a cross-validation mechanism adaptively weights similarity scores using reliability cues such as node richness and yaw discrepancy.
- We propose a semantic graph matching method that enhances node distinctiveness with volume and multi-scale features, rejects outliers via a maximum clique formulation, and employs a two-stage pose estimation strategy for accurate and robust loop closure alignment.

II. Related Work

A. Geometry-based or Semantic Graph-based Methods for Loop Closure Detection

Geometry-based methods [18]–[26] typically extract structural features directly from raw point clouds and construct descriptors for comparison. M2DP [18] projects a 3D point cloud on multiple 2D planes and uses the singular vectors of the resulting density signatures to construct a compact descriptor. Scan Context [21] and its variants [22] project 3D point clouds into a 2D polar coordinate system, encoding each bin with height or other local features. Rotation invariance is achieved by circularly shifting the descriptor columns. Overall, geometry-based methods are robust across environments, but their dependence on low-level features

makes them vulnerable to sparsity and noise, reducing the accuracy of loop closure detection.

Semantic graph-based methods [28], [29], [31], [33], [34], [36], [37] transform point clouds into semantic graphs, capturing object-level semantics and spatial relationships to provide a more abstract representation of the environment. Specially, SGLC [33] extracts semantic graphs from foreground point clouds for initial loop pose estimation, and refines the pose using background point clouds. SGT-LLC [34] encodes semantic and geometric structures via triangular topology graphs, allowing scalable place recognition and accurate 6-DoF pose estimation with a robust descriptor and efficient search. However, semantic graph-based methods often struggle in environments with sparse semantics or dominated by indistinct classes (e.g., vegetation) due to their reliance on rich semantic cues and stable nodes.

B. Semantic-Geometric Integration for Loop Detection

Recent efforts have also explored combining semantic and geometric cues for more robust loop closure detection. SSC [25] improves structural descriptors with semantic labels but lacks object-level scene understanding. GOSMatch [29] performs semantic graph matching followed by geometric verification based on pairwise point distances. SGLC [33] is based on semantic graph matching and employs background structural descriptors for loop verification; however, it simply combines the two representations without a deeper integration. SectionKey [31] and SETD [36] represent and match semantic graphs in a structural-descriptor-like manner. However, these methods fall short of deeply integrating semantics and geometry, limiting their representational power and adaptability in diverse environments.

III. Method

This chapter details the specific approach SD-SGM used for loop closure detection; the system framework is shown in Fig. 2.

A. Semantic Descriptor Construction

Semantic node graphs, which serve as the foundation for semantic graph matching, are constructed from raw point clouds as described in this subsection. Given a raw point cloud $\mathcal{P}$, where each point is defined exclusively by its three-dimensional coordinates and corresponding intensity value, we employ a semantic segmentation model [38] to assign a semantic label to each point. This process results in a semantically annotated point cloud $\mathcal{S}\_\mathcal{P}$. Next, we apply a clustering algorithm [39] to the semantic point cloud $\mathcal{S}\_\mathcal{P}$ to generate the corresponding semantic node graph $\mathcal{G}$, a set of object nodes $\{n_1, n_2, ..., n_j, ...\}$. In addition to coordinates and semantic labels, the nodes in $\mathcal{G}$ also contain bounding box information. In this manner, both semantic and partial geometric information can be extracted from the raw point cloud $\mathcal{P}$, which is crucial for scene understanding.

In order to match two semantic graphs, node correspondences need to be established across the graphs. To this end, a semantic descriptor is generated for each node, providing a compact representation of its properties. Subsequently, node similarities are computed by comparing their descriptors.

As an example, for a node n_i in $\mathcal{G}$, firstly, we select the nodes in $\mathcal{G}$ whose spatial distances to node n_i are less than $30m$, and group them into three subsets N_1, N_2, and N_3 according to their distance to n_i. This hierarchical grouping method captures the multi-scale spatial context surrounding n_i. Second, we divide the nodes in each subset N_i into c disjoint sub-subsets based on their semantic categories, where c is the number of semantic categories predefined in our method. Specifically, the subsets N_1, N_2, and N_3 are partitioned into $\{N_{1i}\}_{i=1}^{c}$, $\{N_{2j}\}_{j=1}^{c}$, and $\{N_{3k}\}_{k=1}^{c}$, respectively. In total, this results in $3c$ semantic sub-subsets. Finally, for every sub-subset, we determine the average size of the node bounding boxes and, in parallel, compute the mean spatial distance from node n_i to all nodes within the sub-subset. This yields a descriptor vector $\boldsymbol{f}_i$ for the node n_i with a dimensionality of $6c$, where each pair of dimensions encodes two key properties: (1) the spatial distribution of neighboring nodes; and (2) the geometric scale of the associated instances.

B. Semantic Graph Matching

Given two semantic graphs $\mathcal{G}_1 = \{n_1^1, n_2^1, \ldots, n_i^1, \ldots\}$, $\mathcal{G}_2 = \{n_1^2, n_2^2, \ldots, n_i^2, \ldots\}$, we compute the similarity between nodes from $\mathcal{G}_1$ and $\mathcal{G}_2$ that share the same semantic label. We begin by defining the semantic descriptor difference in Eq. (1).

$$\mathbf{Err}_{i,j} = \left(\boldsymbol{f}_i^1 - \boldsymbol{f}_j^2\right) \oslash \left(\boldsymbol{f}_i^1 + \boldsymbol{f}_j^2\right) \tag{1}$$

where $\boldsymbol{f}_i^1$ and $\boldsymbol{f}_j^2$ denote the semantic descriptors of n_i^1 and n_j^2, respectively, while $\oslash$ indicates the element-wise vector division. And the similarity can be calculated by:

$$\mathbf{Sim}\,(i,j) = 1 - \frac{1}{6c} \sum_{a_n \in \mathbf{Err}_{i,j}} |a_n| \tag{2}$$

If the similarity $\mathbf{Sim}\,(i,j)$ surpass a threshold, then the corresponding nodes n_i^1 and n_j^2 are considered a node pair. For simplicity, we unify the notation of node pair indices and denote the node pair set as $P = \left\{\boldsymbol{pair}_i = (n_i^1, n_i^2)\right\}$.

However, due to the limited discriminative ability of the semantic features, the node pairs in P may include mismatched pairs. Therefore, we apply a global consistency criterion to filter out these mismatched node pairs. Specially, we construct a *vertex-edge graph*, where each vertex represents a node pair, and an edge is added between two vertices if their respective node pairs are globally consistent [12]. Specifically, if nodes n_i^1, n_i^2 from $\boldsymbol{pair}_i$ and n_j^1, n_j^2 from $\boldsymbol{pair}_j$ satisfy $|\text{dist}(n_i^1, n_j^1) - \text{dist}(n_i^2, n_j^2)| < \phi$, then the two node pairs are globally consistent (dist$(,)$ means the Euclidean distance). After get the *vertex-edge graph*, we use a maximum clique algorithm [40] to find the largest complete subgraph. In our constructed *vertex-edge graph*, a complete subgraph—where any two vertices in the subgraph are connected by an edge—implies all node pairs corresponding to the vertices in this subgraph are globally consistent.

Consequently, the solution produced by the maximum clique algorithm yields the most plausible set of node pairs, denoted as P^*, under the global consistency constraint.

Based on the node pairs set P^*, if the number of node pairs in P^* is more than 3, we can estimate the transformation (rotation $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ and translation $\boldsymbol{t} \in \mathbb{R}^3$) between $\mathcal{G}_1$ and $\mathcal{G}_2$ by minimizing the sum of Euclidean distances between matched nodes in P^*:

$$\boldsymbol{R}, \boldsymbol{t} = \arg\min_{\boldsymbol{R},\boldsymbol{t}} \sum_{\boldsymbol{pair}_i \in P^*} \left(\boldsymbol{R} \cdot n_i^1 + \boldsymbol{t} - n_i^2\right) \tag{3}$$

To further enhance the robustness of pose estimation, we refine the set of node pairs using the estimated transformation $\boldsymbol{R}$, $\boldsymbol{t}$. Specifically, for node n_i^1 in $\mathcal{G}_1$, if there is one node n_j^2 in $\mathcal{G}_2$ satisfying $\text{dist}\left(\boldsymbol{R} \cdot n_i^1 + \boldsymbol{t}, n_j^2\right) < 0.75$, they are considered to be a matched node pair. Thus, we obtain the resulting refined set P_{refine}.

The refined set P_{refine} is then used to solve the transformation again as formulated in Eq. (4), yielding $\boldsymbol{R}_{refine} \in \mathbb{R}^{3\times3}, \boldsymbol{t}_{refine} \in \mathbb{R}^3$ as the estimated 6-DoF relative pose between the two semantic graphs. To further refine the pose at the point cloud level, an additional fine alignment can be performed using the ICP (Iterative Closest Point) algorithm. Besides, the rotation matrix $\boldsymbol{R}_{refine}$ can be decomposed into roll, pitch, and yaw angles, denoted $\boldsymbol{roll}_{sg}, \boldsymbol{pitch}_{sg}, \boldsymbol{yaw}_{sg}$, which will be used in the following fusion module.

$$\boldsymbol{R}_{refine}, \boldsymbol{t}_{refine} = \arg\min_{\boldsymbol{R},\boldsymbol{t}} \sum_{\boldsymbol{pair}_i \in P_{refine}} \left(\boldsymbol{R} \cdot n_i^1 + \boldsymbol{t} - n_i^2\right) \tag{4}$$

Next, we calculate the similarity score between the two graphs. We first transform the coordinate of nodes in $\mathcal{G}_1$ accordingly and evaluate the alignment by measuring the Euclidean distance between each transformed node n_i^1 and its nearest node in $\mathcal{G}_2$. Considering that some nodes may have no matched nodes even under a true loop closure, we truncated the distance at $1m$ to avoid a substantial $\boldsymbol{loss}$ as defined in Eq. (5).

$$\boldsymbol{loss}(n_i^1) = \min\left(\text{dist}\left(\boldsymbol{R}_{refine} \cdot n_i^1 + \boldsymbol{t}_{refine}, n_{matched}^2\right), 1\right) \tag{5}$$

where $n_{matched}^2$ denotes the closest node in $\mathcal{G}_2$ to the transformed node n_i^1, under the constraints of semantic label consistency and similar bounding box size. In addition, if $\boldsymbol{loss}_i < 1$, n_i^1 is considered a *well-aligned* node.

The similarity score between $\mathcal{G}_1$ and $\mathcal{G}_2$ is defined below:

$$\boldsymbol{score}_{\text{sg}} = \exp\left(-\frac{\sum_{n_i^1 \in \mathcal{G}_1} \boldsymbol{loss}(n_i^1)}{\boldsymbol{n}_{well}}\right) \tag{6}$$

where $\boldsymbol{n}_{well}$ denotes the number of *well-aligned* nodes in $\mathcal{G}_1$, that is, those with $\boldsymbol{loss}(n_i^1) < 1$. In addition, $\boldsymbol{score}_{\text{sg}}$ is set to 0 if P^* or P_{refine} contains fewer than 3 pairs, as the relative pose cannot be reliably estimated (as in Eq. (3) or (4)), and node-wise distances become unreliable.

C. Global Structural Descriptor

When the number of nodes is small, the semantic graph's capacity to represent the environment is substantially weakened, resulting in degraded performance. To address this, we draw inspiration from scan context-based methods [21], [25] and introduce a global structural descriptor module to strengthen point cloud representation. Specifically, we adopt an ROI partitioning strategy based on the polar-range coordinate that divides the space into concentric rings and angular sectors. Let L_r and L_θ denote the resolutions of the radial and angular divisions, respectively, and R the maximum radius centered at the point cloud's origin. The total numbers of radial and angular sectors are defined below:

$$N_r = \left\lceil \frac{R}{L_r} \right\rceil, \quad N_\theta = \left\lceil \frac{2\pi}{L_\theta} \right\rceil \tag{7}$$

Given any point (x, y, z) in the point cloud, its corresponding radial and angular sector indices (r, θ) are computed as follows. Each (r, θ) defines a fan-shaped bin that may contain multiple points falling within its spatial extent.

$$r = \left\lceil \frac{\sqrt{x^2+y^2}}{L_r} \right\rceil, \; \theta = \left\lceil \frac{\text{atan2}(y,x) + 2\pi \cdot \mathbf{1}_{\text{atan2}(y,x)<0}}{L_\theta} \right\rceil \tag{8}$$

where $\mathbf{1}_{\text{atan2}(y,x)<0} = 1$ if $\text{atan2}(y, x) < 0$, and 0 otherwise.

Next, we assign each fan-shaped bin (r, θ) a representative value that captures the dominant feature of the points within it, e.g., maximum intensity [21] or the most frequent semantic category [25]. We adopt the latter.

For a fixed angular sector θ_i, the representative values of all its radial bins form a 1D vector c_i. By collecting the 1D vectors c_i of all angular sectors, we obtain two global descriptors $S_1 = \left\{c_1^1, c_2^1, \ldots, c_{N_\theta}^1\right\}$ and $S_2 = \left\{c_1^2, c_2^2, \ldots, c_{N_\theta}^2\right\}$ for the two point clouds corresponding to $\mathcal{G}_1$ and $\mathcal{G}_2$. We then perform column-wise shifting on S_1 to simulate the rotation of the point cloud until we find the optimal rotation angle that maximizes the similarity between the two descriptors. The maximum similarity and the optimal rotation angle are denoted by $\boldsymbol{score}_{gsd}$ and $\boldsymbol{yaw}_{gsd}$, respectively. The detailed computation process is omitted here and can be found in the analogous calculation in Scan Context [21].

D. Fusion of Semantic Graph and Global Structural Descriptor

In order to combine the advantage of semantic graph matching and global structural descriptor for a more robust loop closure detection, we have adopted a new method for similarity calculation, which fully considers the effective information obtained by the two.

If $(\boldsymbol{R}_{refine}, \boldsymbol{t}_{refine})$ is reliably estimated, we first apply $(\boldsymbol{R}_{refine}, \boldsymbol{t}_{refine})$ to the point cloud $\mathcal{P}_1$, resulting in the transformed point cloud $\bar{\mathcal{P}}_1$. Next, we construct a global structural descriptor $\bar{S}_1 = \left\{\bar{c}_1^1, \bar{c}_2^1, \ldots, \bar{c}_{N_\theta}^1\right\}$ for $\bar{\mathcal{P}}_1$. Since $\bar{\mathcal{P}}_1$ has been spatially aligned to $\mathcal{P}_2$ by using the transformation matrix from semantic graph matching, the similarity

TABLE I: The F_1 max and EP results on KITTI and KITTI-360 datasets

Method	KITTI							KITTI-360	
	00	02	05	06	07	08	Mean	0002	0009
PV	0.779 0.641	0.727 0.691	0.541 0.536	0.852 0.767	0.631 0.591	0.037 0.500	0.595 0.621	0.349 0.515	0.330 0.510
SGPR*	0.720 0.507	0.823 0.531	0.720 0.552	0.680 0.524	0.700 0.500	0.683 0.506	0.721 0.520	0.862 0.506	0.845 0.503
OT	0.873 0.800	0.810 0.725	0.837 0.772	0.876 0.809	0.625 0.505	0.667 0.510	0.781 0.687	0.796 0.507	0.879 0.528
BEVPlace	0.960 0.849	0.845 0.819	0.885 0.815	0.895 0.815	0.917 0.687	0.967 0.868	0.912 0.809	0.920 0.504	0.938 0.684
GOSMatch*	0.916 0.535	0.694 0.575	0.785 0.611	0.491 0.518	0.947 0.913	0.908 0.571	0.790 0.621	0.694 0.500	0.766 0.575
SSC*	0.955 0.865	0.933 0.768	0.966 0.956	0.983 0.978	0.894 0.785	0.950 0.940	0.947 0.882	**0.965** 0.793	0.966 0.873
BOW3D	0.977 0.981	0.578 0.704	0.965 0.969	0.985 0.985	0.906 0.929	0.900 0.866	0.885 0.906	0.210 0.560	0.682 0.761
CC	0.977 0.964	0.903 0.568	0.958 0.901	0.993 0.959	0.906 0.811	0.823 0.575	0.927 0.796	0.717 0.560	0.886 0.620
SGLC*	0.998 0.986	0.888 0.899	0.969 0.967	0.995 0.963	0.993 0.991	0.988 0.980	0.972 0.964	0.932 0.934	**0.994** 0.978
SGT-LLC*	0.986 0.615	0.751 0.801	0.970 0.550	0.875 0.504	0.964 0.816	0.952 0.918	0.916 0.701	0.769 0.813	0.938 0.603
SD-SGM*	**1.000 1.000**	**0.983 0.943**	**0.979 0.972**	**0.998 0.993**	**1.000 1.000**	**0.992 0.992**	**0.992 0.983**	0.937 **0.940**	0.992 **0.985**

*The best results are in bold. * denotes semantic-assisted method.*

between $\bar{S}_1$ and S_2 can be calculated directly as in Eq. (9), without circular shifting.

$$\overline{\boldsymbol{score}}_{gsd} = \sum_{i=1}^{N_\theta} \frac{(\bar{c}_i^1 - \bar{c}^1) \cdot (c_i^2 - c^2)}{\|\bar{c}_i^1 - \bar{c}^1\| \cdot \|c_i^2 - c^2\|} \tag{9}$$

where $\bar{c}^1$ and c^2 represent the means of all columns in $\bar{S}_1$ and S_2, respectively.

Next, we jointly consider indicators derived from both the semantic graph and the global structural descriptor, and we evaluate the final similarity $\boldsymbol{score}_{final}$ between point clouds $\mathcal{P}_1$ and $\mathcal{P}_2$ as follows.

$$\boldsymbol{score}_{final} = \begin{cases} \boldsymbol{score}_{\text{sg}}, & \overline{\boldsymbol{score}}_{gsd} > \boldsymbol{score}_{gsd} \\ \boldsymbol{score}_{\text{sg}} \cdot \exp\left(-\frac{|\Delta \boldsymbol{yaw}|}{30 * \boldsymbol{score}_{gsd}}\right), & \text{otherwise} \end{cases} \tag{10}$$

where $\Delta \boldsymbol{yaw} = \boldsymbol{yaw}_{sg} - \boldsymbol{yaw}_{gsd}$. Specifically, we compare $\overline{\boldsymbol{score}}_{gsd}$ with $\boldsymbol{score}_{gsd}$. If $\overline{\boldsymbol{score}}_{gsd} > \boldsymbol{score}_{gsd}$, it implies that the 6-DoF pose estimated from the semantic graph is accurate. In this case, the semantic graph-based method is deemed more reliable, and we set $\boldsymbol{score}_{final}$ directly as $\boldsymbol{score}_{\text{sg}}$. Otherwise, we refine $\boldsymbol{score}_{\text{sg}}$ by jointly consider the yaw discrepancy $\Delta \boldsymbol{yaw}$ and the global descriptor similarity $\boldsymbol{score}_{gsd}$. A large $\Delta \boldsymbol{yaw}$ implies a significant inconsistency between the yaw estimates of the two methods, thus leading to a reduction of $\boldsymbol{score}_{\text{sg}}$, while a higher $\boldsymbol{score}_{gsd}$ reflects stronger geometric alignment and improves $\boldsymbol{score}_{final}$.

Otherwise, if the matched node pairs from the semantic graph matching is less than 3. In this case, we can only use the similarity of global structural descriptor $\boldsymbol{score}_{gsd}$ to determine $\boldsymbol{score}_{final}$. A node richness factor is introduced to further penalize $\boldsymbol{score}_{final}$. The final score is decreased when both semantic graphs are rich in nodes but fail to yield meaningful correspondences, which indicates a likely dissimilarity between the two graphs.

$$\boldsymbol{score}_{final} = r \cdot \frac{\boldsymbol{score}_{gsd}}{log\left(\mathbf{min}(n_1, n_2)\right)} \tag{11}$$

where n_1 and n_2 are the number of nodes in $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively.

IV. Experimental Results

This section evaluates SD-SGM for loop closure detection across multiple public outdoor datasets, benchmarked against recent state-of-the-art methods. Since our framework builds upon semantic graphs, semantic segmentation serves as a prerequisite. To better reflect practical deployment scenarios, we employ SegNet4D [38], a real-time semantic segmentation network, to produce semantic point clouds instead of ground-truth annotations. In our experiments, semantic graphs are mainly constructed using static object categories, including parked cars, trunks, and poles.

A. Loop Detection on KITTI and KITTI-360 Datasets

Experiments were first carried out on the KITTI [41] and KITTI-360 [42] datasets, both acquired with a 64-beam LiDAR and featuring diverse outdoor driving environments. From each dataset, we selected multiple sequences containing loop closures for quantitative evaluation. To ensure comparability with state-of-the-art approaches, we strictly followed the evaluation protocol proposed in SSC [25]: a detected loop closure is considered a true positive when the position distance is within 3 meters, and as a false loop closure when the distance exceeds 20 meters. We utilized the frame pairs from SSC [25], which include the complete set of positive loop closure pairs and a negative sample set 100 times larger in size. The benchmarked methods are PV [19], SGPR [28], OT [23], BEVPlace [43], GOSMatch [29], SSC [25], BOW3D [24], CC [44], SGLC [33] and SGT-LLC [34]. Loop closure performance was evaluated using standard metrics, including the maximum F_1 score (F_1 max) and Extended Precision (EP). The F_1 score, which balances precision and recall, is computed as:

$$F_1 = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

F_1 max is the highest F_1 score around all precision and recall pairs. EP is defined by:

$$\text{EP} = \frac{P_0 + R_1}{2} \tag{13}$$

where P_0 is the precision when the recall is 0% and R_1 is the recall when the precision is 100%.

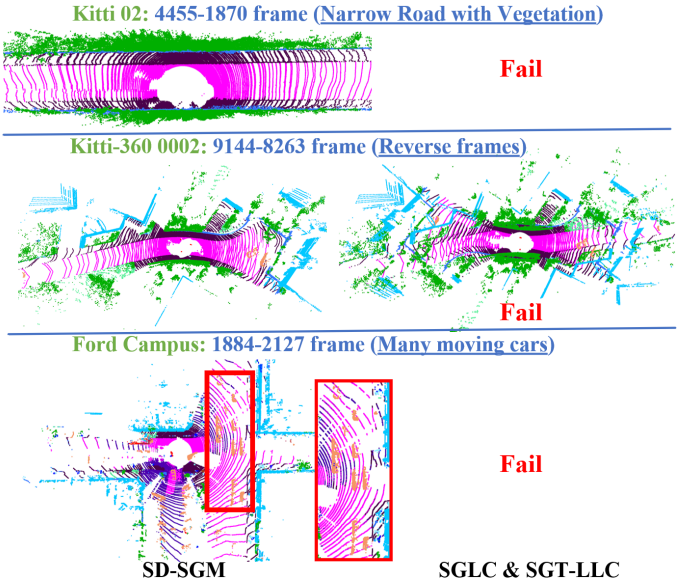

Fig. 3: This figure shows failure cases of SGLC and SGT-LLC—background dominance (e.g., vegetation) in KITTI-02 and KITTI-360 0002, and **many dynamic vehicles in Ford Campus** cause missed loops. In contrast, SD-SGM reliably detects loops and registers poses, even under reverse-frame conditions in KITTI-360.

TABLE II: Loop Closure Performance on Ford Campus

Method	Ford Campus		
	AUC↑	F_1 max↑	Recall@1↑
SGPR	0.412	0.439	0.467
GOSMATCH	0.752	0.632	0.820
SSC	0.924	0.865	0.915
PADLoC	0.938	0.873	0.910
SGLC†	0.955	0.908	0.920
SGLC	0.959	0.897	0.896
SGT-LLC	0.941	0.881	0.879
SD-SGM	**0.970**	**0.925**	**0.927**

↑ *means larger is better.*
† *denotes the vehicle nodes is not included in the semantic graph.*

Tab. I summarizes the results, where SD-SGM consistently surpasses all baselines in terms of evaluation metrics. Particularly in the difficult KITTI 02 and KITTI-360 0002 sequences—characterized by narrow roads and dense vegetation that hinder semantic instance visibility—SOTA semantic graph-based approaches such as SGLC and SGT-LLC exhibit pronounced performance degradation. In contrast, SD-SGM maintains robustness under these adverse conditions (see Fig. 3) by jointly leveraging semantic graph reasoning and global structural descriptors. Across the remaining sequences, SD-SGM further outperforms existing methods, largely due to two key designs: (1) the maximum clique algorithm for selecting spatially consistent node correspondences, (2) a two-stage pose estimation that enhances graph similarity evaluation through node-pair distances. Together, these mechanisms improve the discrimination between true and false loop closures.

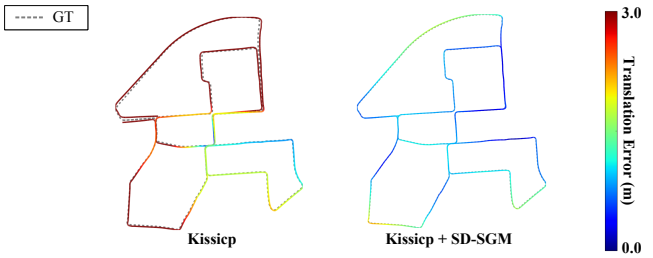

Fig. 4: The figure shows the SLAM trajectories on the KITTI 00 dataset. The **gray dashed line** represents the ground truth (GT) trajectories. Kissicp represents the odometry trajectory by Kissicp, and Kissicp+SD-SGM includes loop closure refinement from SD-SGM.

B. Generalization Testing

To further assess the generalization ability of SD-SGM in loop closure detection, we performed experiments on sequences from the Ford Campus dataset [45]. This dataset is particularly challenging due to moving vehicles, imperfect semantic segmentation, and frequent occlusions (see Fig. 3). The evaluation was carried out using three metrics: AUC, maximum F_1 score, and Recall@1, with the outcomes summarized in Tab. II. AUC is the area under the PR curve. Recall@1 measures the proportion of point cloud frames with true loop closures for which the top-ranked candidate (with the highest similarity score) is indeed correct. In comparison with state-of-the-art baselines, SD-SGM consistently delivers superior performance, highlighting its robustness in complex scenarios. This advantage primarily stems from its capability to accurately establish true node pairs within the semantic graph matching module. Additionally, qualitative results on the Ford Campus dataset are illustrated in Fig. 3. They demonstrate that SD-SGM can reliably identify true loop closures and align point clouds effectively, even under challenging conditions with poor semantic segmentation and numerous dynamic vehicles.

C. SLAM Validation

To evaluate SD-SGM's effectiveness in loop closure detection within SLAM, we conducted experiments on publicly available datasets. Fig. 4 shows the SLAM trajectory on KITTI 00, where odometry was obtained using Kissicp [46] and loop closures were detected by SD-SGM. Incorporating SD-SGM (Kissicp+SD-SGM) notably enhances trajectory accuracy compared to using Kissicp alone.

In addition, we outline the loop retrieval procedure used in SLAM evaluation. For each keyframe, a k-d tree search is conducted with a Scan Context-like descriptor [21] to identify the top-n most similar frames (commonly $n = 50$). SD-SGM subsequently performs loop verification and estimates the relative pose. The complete process requires an average of only 1086 **ms**.

V. CONCLUSIONS

In this paper, we present SD-SGM, a novel loop closure detection framework that effectively combines semantic

graph matching with global structural descriptors to achieve robust and accurate place recognition in outdoor environments. We introduce a cross-validation mechanism that dynamically balances the contributions of semantic graphs and global descriptors based on their reliability. This is achieved by evaluating intermediate variables, such as node richness and yaw discrepancy, enabling SD-SGM to maintain robustness across diverse environments. Extensive experiments on multiple datasets demonstrate that SD-SGM achieves state-of-the-art performance in loop closure detection. In future work, we will endeavor to further enhance the robustness of the integrated framework while simultaneously improving algorithmic efficiency.

REFERENCES

[1] O. Mayuku, B. W. Surgenor, and J. A. Marshall, "A self-supervised near-to-far approach for terrain-adaptive off-road autonomous driving," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 14 054–14 060.

[2] Y. Tang, M. Wang, Y. Deng, Z. Zheng, J. Zhong, T. Liu, C. Zhao, and Y. Yue, "Openobject-nav: Open-vocabulary object-oriented navigation based on dynamic carrier-relationship scene graph," in *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2025, pp. 20 182–20 188.

[3] A. Romero, C. Delgado, L. Zanzi, R. Suárez, and X. Costa-Pérez, "Cellular-enabled collaborative robots planning and operations for search-and-rescue scenarios," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 5942–5948.

[4] Y. Yue, C. Zhao, Z. Wu, C. Yang, Y. Wang, and D. Wang, "Collaborative semantic understanding and mapping framework for autonomous systems," *IEEE/ASME Transactions on Mechatronics*, vol. 26, no. 2, pp. 978–989, 2020.

[5] Q. Zhang, M. Wang, Y. Yue, and T. Liu, "Lcr-smm: Large convergence region semantic map matching through expectation maximization," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 5, pp. 3029–3040, 2021.

[6] Y. Deng, Y. Tang, Y. Yang, D. Wang, and Y. Yue, "Macim: Multi-agent collaborative implicit mapping," *IEEE Robotics and Automation Letters*, vol. 9, no. 5, pp. 4369–4376, 2024.

[7] Z. Wu, W. Wang, Y. Yue, J. Zhang, H. Shen, and D. Wang, "Mag-MM: Magnetic-enhanced multisession mapping in repetitive environments," *IEEE/ASME Trans. Mechatron. (TMECH)*, 2025, Early Access.

[8] J. Dou, Y. Deng, J. Wang, X. Tang, Y. Yang, and Y. Yue, "Openmulti: Open-vocabulary instance-level multi-agent distributed implicit mapping," *IEEE Robotics and Automation Letters*, 2025.

[9] P. Pfreundschuh, H. F. Hendrikx, V. Reijgwart, R. Dubé, R. Siegwart, and A. Cramariuc, "Dynamic object aware lidar slam based on automatic generation of training data," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 11 641–11 647.

[10] X. Lin, S. Ji, Z. Liao, L. He, X. Zhou, and H. Zhang, "Pmo-slam: Enhancing incremental optimization with probabilistic metrics for object slam," *IEEE/ASME Transactions on Mechatronics*, 2025.

[11] N. Wang, H. Lu, Z. Zheng, H. Wang, Y.-H. Liu, and X. Chen, "Leveraging semantic graphs for efficient and robust lidar slam," *arXiv preprint arXiv:2503.11145*, 2025.

[12] Y. Tang, M. Wang, Y. Yang, Z. Lan, and Y. Yue, "Robust large-scale collaborative localization based on semantic submaps with extreme outliers," *IEEE/ASME Transactions on Mechatronics*, 2023.

[13] Y. Tang, M. Wang, Y. Deng, Y. Yang, and Y. Yue, "Ssgm: Spatial semantic graph matching for loop closure detection in indoor environments," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 9163–9168.

[14] L. Li, X. Kong, X. Zhao, W. Li, F. Wen, H. Zhang, and Y. Liu, "Saloam: Semantic-aided lidar slam with loop closure," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 7627–7634.

[15] S. Gupta, T. Guadagnino, B. Mersch, I. Vizzo, and C. Stachniss, "Effectively detecting loop closures using point cloud density maps," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 10 260–10 266.

[16] Z. Wu, W. Wang, J. Zhang, Y. Wang, G. Peng, and D. Wang, "Mglt: Magnetic-lead global localization and tracking in degenerated repetitive environments," *IEEE/ASME Transactions on Mechatronics*, 2024.

[17] Y. Tang, M. Wang, Y. Deng, Y. Yang, Z. Lan, and Y. Yue, "Multi-view robust collaborative localization in high outlier ratio scenes based on semantic features," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 11 042–11 047.

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

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

[20] Y. Wang, Z. Sun, C.-Z. Xu, S. E. Sarma, J. Yang, and H. Kong, "Lidar iris for loop-closure detection," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5769–5775.

[21] G. Kim and A. Kim, "Scan context: Egocentric spatial descriptor for place recognition within 3d point cloud map," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 4802–4809.

[22] H. Wang, C. Wang, and L. Xie, "Intensity scan context: Coding intensity and geometry relations for loop closure detection," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2095–2101.

[23] J. Ma, J. Zhang, J. Xu, R. Ai, W. Gu, and X. Chen, "Overlaptransformer: An efficient and yaw-angle-invariant transformer network for lidar-based place recognition," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6958–6965, 2022.

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

[25] L. Li, X. Kong, X. Zhao, T. Huang, and Y. Liu, "Semantic scan context: a novel semantic-based loop-closure method for lidar slam," *Autonomous Robots*, vol. 46, no. 4, pp. 535–551, 2022.

[26] X. Xu and et al., "Ring++: Roto-translation invariant gram for global localization on a sparse scan map," *IEEE Transactions on Robotics*, vol. 39, no. 6, pp. 4616–4635, 2023.

[27] H. Li, Y. Mai, M. Gao, J. He, Z. Liu, and H. Wang, "Large-scale lidar-based loop closing via combination of equivariance and invariance on se (3)," *IEEE/ASME Transactions on Mechatronics*, 2025.

[28] X. Kong, X. Yang, G. Zhai, X. Zhao, X. Zeng, M. Wang, Y. Liu, W. Li, and F. Wen, "Semantic graph based place recognition for 3d point clouds," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 8216–8223.

[29] Y. Zhu, Y. Ma, L. Chen, C. Liu, M. Ye, and L. Li, "Gosmatch: Graph-of-semantics matching for detecting loop closures in 3d lidar data," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5151–5157.

[30] F. Nie, W. Zhang, Y. Wang, Y. Shi, and Q. Huang, "A forest 3-d lidar slam system for rubber-tapping robot based on trunk center atlas," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 5, pp. 2623–2633, 2021.

[31] S. Jin, Z. Wu, C. Zhao, J. Zhang, G. Peng, and D. Wang, "Sectionkey: 3-d semantic point cloud descriptor for place recognition," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 9905–9910.

[32] W. Dong, C. Gong, G. Chen, X. Sheng, and X. Zhu, "Computationally efficient topological mapping with layered spanning trees," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 5, pp. 4067–4077, 2022.

[33] N. Wang, X. Chen, C. Shi, Z. Zheng, H. Yu, and H. Lu, "Sglc: Semantic graph-guided coarse-fine-refine full loop closing for lidar slam," *IEEE Robotics and Automation Letters*, 2024.

[34] S. Wang, F. Cao, T. Wang, X. Chen, and S. Shao, "Sgt-llc: Lidar loop closing based on semantic graph with triangular spatial topology," *IEEE Robotics and Automation Letters*, 2025.

[35] M. Wang, Y. Tang, Y. Deng, H. Lu, S. Zuo, and Y. Yue, "Segram: Aligned coordinate system aided semantic graph matching method for loop closure detection," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 24 024–24 035, 2025.

[36] M. Liao and et al., "Semantic topological descriptor for loop closure detection within 3d point clouds in outdoor environment," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 2856–2863.

[37] Y. Tang, M. Wang, H. Lu, J. Zhong, S. Zuo, Y. Deng, and Y. Yue, "Sloop: Aligned coordinate system-aided lidar loop closure detection based on semantic node graph matching," in *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2025, pp. 15 719–15 725.

[38] N. Wang, R. Guo, C. Shi, Z. Wang, H. Zhang, H. Lu, Z. Zheng, and X. Chen, "Segnet4d: Efficient instance-aware 4d semantic segmentation for lidar point cloud," *IEEE Transactions on Automation Science and Engineering*, 2025.

[39] S. Park, S. Wang, H. Lim, and U. Kang, "Curved-voxel clustering for accurate segmentation of 3d lidar point clouds with real-time performance," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 6459–6464.

[40] Q. Wu and J.-K. Hao, "A review on algorithms for maximum clique problems," *European Journal of Operational Research*, vol. 242, no. 3, pp. 693–709, 2015.

[41] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "Semantickitti: A dataset for semantic scene understanding of lidar sequences," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9297–9307.

[42] Y. Liao, J. Xie, and A. Geiger, "Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3292–3310, 2022.

[43] L. Luo, S. Zheng, Y. Li, Y. Fan, B. Yu, S.-Y. Cao, J. Li, and H.-L. Shen, "Bevplace: Learning lidar-based place recognition using bird's eye view images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 8700–8709.

[44] B. Jiang and S. Shen, "Contour context: Abstract structural distribution for 3d lidar loop detection and metric pose estimation," in *2023 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2023, pp. 8386–8392.

[45] G. Pandey, J. R. McBride, and R. M. Eustice, "Ford campus vision and lidar data set," *The International Journal of Robotics Research*, vol. 30, no. 13, pp. 1543–1552, 2011.

[46] I. Vizzo, T. Guadagnino, B. Mersch, L. Wiesmann, J. Behley, and C. Stachniss, "Kiss-icp: In defense of point-to-point icp–simple, accurate, and robust registration if done the right way," *IEEE Robotics and Automation Letters*, vol. 8, no. 2, pp. 1029–1036, 2023.