2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Point Cloud Mapping and Loop Closure Detection Using Superpoint Semantic Graph for Autonomous Driving

Ronghua Du[1], Zong Li[1], Jinlai Zhang(✉)[1], Kai Gao(✉)[1], Shaosheng Fan[2], Taishan Cao[1], Gengbiao Chen[1], Zhenzhen Jin[3]

[1]College of Mechanical and Vehicle Engineering, Changsha University of Science and Technology, Changsha, 410114, Hunan, China
[2]School of Artificial Intelligence, Changsha University of Science and Technology, Tianxin District, Changsha, Hunan, 410114, China
[3]College of Mechanical Engineering, Guangxi University, Naning, 530004, Guangxi, China

Abstract. Accurate loop closure detection and pose estimation remain critical challenges for autonomous vehicles operating in dynamic urban environments, where perceptual aliasing, occlusions, and changing scenes often degrade localization performance. To this end, we present a novel hierarchical framework that leverages superpoint graphs to achieve robust place recognition and precise pose estimation. Our approach begins by constructing a topologically meaningful superpoint graph, where nodes represent stable environmental features and edges encode their spatial relationships. For loop closure detection, we introduce semantic-enhanced ring descriptors that combine geometric structure with semantic information, enabling reliable place recognition despite viewpoint changes or temporary occlusions. The system employs a two-stage verification process: initial candidate selection through descriptor matching, followed by geometric verification using superpoint centroids with RANSAC-based outlier rejection. The pose estimation pipeline employs a hierarchical refinement strategy, starting with superpoint centroid alignment, followed by dense ICP and sparse point-to-plane ICP, all integrated into a global pose graph optimization framework. Our overlap-based loop closure detection demonstrates superior performance across KITTI, Apollo, and Ford Campus datasets, achieving state-of-the-art (SOTA) results on AUC, F1MAX, and recall rate. Furthermore, our pose estimation method exhibits consistently outstanding performance in both accuracy and robustness.

Keywords: Superpoint Graph, Autonomous Vehicle Localization, Loop Closure Detection, Semantic Pose Estimation.

1 Introduction

Accurate and reliable localization is a cornerstone of autonomous driving [1]. In urban environments, where dynamic obstacles, perceptual aliasing, and rapidly changing surroundings are common, maintaining consistent pose estimation and detecting loop closures [2] becomes exceptionally challenging. Loop closure detection—the ability of an

autonomous system to recognize previously visited locations—is essential for correcting accumulated drift in simultaneous localization and mapping (SLAM) systems. Pose estimation [3], in turn, ensures that vehicles can accurately infer their location and orientation within a dynamic and potentially ambiguous map. Traditional methods in LiDAR-based SLAM rely heavily on geometric consistency, such as Iterative Closest Point (ICP) [4] algorithms or feature-based place recognition. While effective under ideal conditions, these methods often suffer when faced with occlusions, environmental changes, or significant viewpoint variation. Perceptual aliasing, where different places appear geometrically similar, and dynamic entities (e.g., pedestrians, parked vehicles) further exacerbate the risk of false loop closures and mislocalization.

Current state-of-the-art approaches typically fall into two categories: (1) purely geometric methods [5] that lack semantic understanding, and (2) deep-learning-based solutions that often generalize poorly across diverse environments due to data biases or insufficient explainability. Purely geometric loop closure methods often fail in the face of structural changes or occlusions, while semantic-enhanced [6] deep models may offer robustness at the cost of computational complexity and training data dependency. Moreover, dense point cloud registration techniques such as point-to-point or point-to-plane ICP, while accurate, are computationally intensive and sensitive to initialization errors, leading to degraded performance under poor priors or high noise conditions. To address these limitations, a hybrid strategy that combines geometric robustness, semantic awareness, and hierarchical optimization is needed. Superpoint graphs [7] have emerged as a promising representation that preserves the topological and semantic structure of 3D scenes. By clustering stable features into semantically meaningful units—superpoints—these graphs can maintain global spatial relationships and local feature fidelity, offering a robust abstraction layer for tasks such as mapping, recognition, and registration.

This paper proposes a novel hierarchical framework that unifies point cloud mapping, semantic understanding, and pose refinement using a Superpoint Semantic Graph (SPSG). In this approach, point clouds are abstracted into superpoints that capture both stable geometry and high-level semantics. These superpoints form nodes in a graph, while edges encode spatial relationships, enabling efficient and robust reasoning for place recognition and loop closure detection.

A key innovation lies in the introduction of semantic-enhanced ring descriptors, which merge geometric and semantic features into compact yet expressive representations. These descriptors significantly improve the system's ability to recognize places under varying viewpoints, seasonal changes, and occlusions. By integrating semantic priors, the system avoids many of the pitfalls of purely geometric methods, while remaining data-efficient and interpretable.

The proposed loop closure detection pipeline operates in two stages. First, it performs candidate selection using descriptor similarity, followed by a rigorous geometric verification phase that employs RANSAC with superpoint centroids. This combination filters out false positives effectively while preserving recall. For pose estimation, the framework employs a hierarchical refinement strategy beginning with centroid alignment, progressing through dense ICP, and culminating in sparse point-to-plane refine-

ment. All pose candidates are then optimized globally via pose graph optimization, ensuring consistency and minimizing drift. The proposed system is thoroughly validated on three publicly available autonomous driving datasets—KITTI, Apollo, and Ford Campus. Across these diverse environments, the method achieves state-of-the-art (SOTA) results in multiple loop closure metrics, including AUC, F1MAX, and recall, demonstrating both robustness and generalizability. The pose estimation module also exhibits superior accuracy and resilience across varied scene complexities and sensor noise levels.

In summary, this paper makes the following key contributions:

- A novel Superpoint Semantic Graph (SPSG) representation that encodes both geometric and semantic properties for robust mapping and recognition.
- A semantic-enhanced ring descriptor for reliable place recognition under challenging urban conditions.
- A hierarchical pose estimation strategy, combining centroid-based initialization, multi-resolution ICP alignment, and global pose graph optimization.

2 Methodology

In this section, we detail the workflow of our SuperPoint Semantic Graph (SPSG) for loop closure detection and pose estimation, as illustrated in Fig. 1. We first present our superpoint graph clustering method in Section 3.1, followed by the superpoint descriptor in Section 3.2. We then introduce our loop closure detection approach in Section 3.3, and conclude with the pose estimation method in Section 3.4.

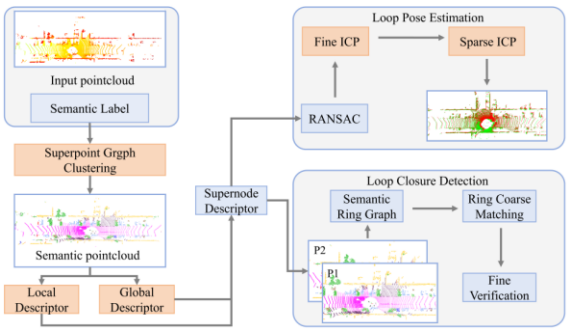

Fig. 1. Overall framework of SPSG.

2.1 Superpoint Graph Clustering

To efficiently represent the raw point cloud $\mathcal{P}_t = p_t \subset \mathbb{R}^3$corresponding to a keyframe, a voxel-based downsampling strategy is initially applied, followed by geometric feature-based clustering. This process yields a set of compact superpoints, where each superpoint s comprises a group of 3D points exhibiting locally homogeneous geometric properties. Fig. 2 illustrates the aggregation process of our SuperPoint Semantic Graph.

To capture both geometric compactness and topological adjacency, a superpoint graph is constructed. In this graph, nodes represent superpoints, and edges indicate spatial adjacency. Specifically, an edge is established between superpoints $s1$ and $s2$ if the minimum Euclidean distance between any point pair across the two superpoints is less than a predefined threshold ϵ. However, direct pairwise computation leads to quadratic complexity $\mathcal{O}(N^2)$, which is computationally prohibitive.

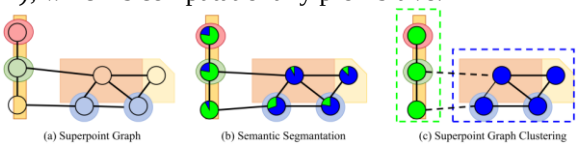

Fig. 2. The diagram of superpoint graph clustering.

To mitigate this, an iterative centroid-based approximation is introduced. Let $g1$and $g2$ denote the centroids of $s1$ and $s2$, respectively. These centroids are iteratively refined by identifying their nearest neighbor in the opposite superpoint and updating accordingly, until convergence. The final inter-centroid distance $|g1 - g2|$ is used to approximate the superpoint distance.

Furthermore, to reduce unnecessary comparisons, a spatial pruning mechanism is employed. A bounding sphere is constructed for each superpoint using radii $r1$ and $r2$, and distance calculations are performed only if:

$$|g1 - g2| <= r1 + r2 + \epsilon \tag{1}$$

Semantic classification is performed by feeding superpoint features $\boldsymbol{f}i$ into a multilayer perceptron (MLP) classifier ϕ_{cls} , generating a probability vector:

$$\mathcal{Z}i <= \text{softmax}(\phi_{\text{cls}}(\boldsymbol{f}i)) \tag{2}$$

The prediction $\mathcal{Z}i$ is supervised using a cross-entropy loss against the ground truth one-hot label $1(ci)$:

$$\mathcal{L}_i^{\text{cls}} = \mathcal{H}(\mathcal{Z}i, 1(ci)) \tag{3}$$

and the total loss is averaged over the superpoint set:

$$\mathcal{L} = \frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}} \mathcal{L}_i^{\text{cls}} \tag{4}$$

2.2 Supernode Descriptor

To achieve invariance to rigid transformations (i.e., rotation and translation), we propose a novel supernode descriptor that captures the relative topological relationships of each superpoint with respect to both its local neighborhood and the global scene structure. This design is particularly advantageous in dynamic urban environments, where geometric consistency and semantic context are critical for reliable perception and localization. Given a segmented point cloud, an undirected graph $\mathcal{G} = \langle \mathcal{U}, \mathcal{E} \rangle$ is constructed, where each node $U_i \in \mathbb{R}^3$ represents the centroid of a superpoint, and each edge $\ell_{ij} = \langle U_i, U_j \rangle$ indicates a Euclidean adjacency relation. This graph captures not

only the spatial proximity between superpoints but also their semantic co-occurrence patterns in the 3D scene.

The semantic-aware topological structure of the graph inherently encodes object-level relationships—such as vehicle-to-tree, building-to-road, or pole-to-sidewalk—which are crucial for understanding scene layout and improving the robustness of higher-level tasks like loop closure detection or map matching. To exploit this structure, each supernode is characterized using a two-part descriptor:

Local Topological Histogram Descriptor ($\boldsymbol{f}_l$): This component models the distribution of semantic edge types and their relative distances to the current node within a local neighborhood. Distances are binned into discrete intervals, and semantic edge types are categorized (e.g., vehicle–vehicle, vehicle–building), forming a histogram that serves as a compact representation of local semantic geometry. This descriptor is robust to small positional perturbations and varying sampling densities, making it suitable for real-world, noisy LiDAR data.

Global Spectral Descriptor($\boldsymbol{f}_g \in \mathbb{R}^{\mathbf{3}}$): To capture the structural role of the superpoint in the overall scene topology, we perform eigen-decomposition on the adjacency matrix A of the graph, obtaining $A = Q \wedge Q^{\intercal}$. The top-$k$ eigenvectors are selected to construct a spectral embedding, which encodes global connectivity patterns and reflects the importance of each superpoint in the graph structure. This global descriptor complements the local histogram by incorporating long-range dependencies and scene-level contextual cues.

The final supernode descriptor $\boldsymbol{f}$ is obtained by concatenating the two components:

$$\boldsymbol{f} = [\boldsymbol{f}_l; |; \boldsymbol{f}_g] \tag{5}$$

where $[\cdot; |; \cdot]$ denotes vector concatenation. This unified representation is both transformation-invariant and semantically discriminative, enabling robust superpoint matching across diverse scenes.

2.3 Loop Closure Detection

To enable robust loop closure detection under large-scale and long-term conditions, we construct a semantic ring graph G_{ring} for each incoming LiDAR frame. Unlike conventional descriptor-based methods, this representation emphasizes the macro-level semantic structure by aggregating superpoint relations in concentric spatial partitions. The ring-based construction explicitly preserves scene layout and radial semantic distribution, which remains consistent across revisits despite viewpoint changes.

Each ring graph is formed by dividing the space around the sensor into K concentric rings, and counting the number of edges of each semantic type within each ring. Let $E^a_{t,k}$ denote the number of edges of semantic class t in the k-th ring of graph G^a_{ring}, where $t = 1, \cdots, T$. The graph similarity score between two frames is then defined as:

$$\boldsymbol{S}\left(G^a_{ring}, G^b_{ring}\right) = \frac{1}{K}\sum_{t=1}^{T}\sum_{k=0}^{K-1} E^a_{t,k} \cdot E^b_{t,k} \tag{6}$$

which measures the degree of structural overlap in semantic edge distribution across spatial rings. A candidate frame is retained if the similarity exceeds a threshold θ_1.

To ensure spatial consistency beyond semantic similarity, we apply a rigid transformation-based geometric verification. Let $Q^a = q_i^a$ and $Q^b = q_i^b$ be the matched superpoint centroids from the query and candidate frames, respectively. The optimal transformation $T \in SE(3)$ is estimated to align Q^a to Q^b, and the residual alignment error is evaluated as:

$$\boldsymbol{E}(Q^a, Q^b) = \frac{1}{|M|} \sum_{(q_i^a, q_i^b) \in M} ||T q_i^a - q_i^b||^2 \tag{7}$$

where M denotes the set of matched pairs. A loop closure is accepted if $E < \theta_2$, indicating geometric consistency between the corresponding regions.

In summary, the proposed semantic ring graph enables coarse-to-fine loop closure detection, where semantic-level structure guides candidate selection, and geometric-level verification ensures alignment accuracy. This two-stage design enhances robustness against occlusions, dynamic objects, and viewpoint variance, making it well-suited for complex urban traffic environments.

2.4 Superpoint-Based Pose Estimation

Once a loop closure has been confirmed, a hierarchical pose refinement strategy is employed to improve the alignment between frames, leveraging 3D geometric superpoints—clusters of spatially coherent points in the LiDAR point cloud that represent semantically meaningful structures (e.g., vehicle body, pole, facade surface).

Coarse Alignment with Centroid-Based RANSAC: The refinement pipeline begins by estimating an initial rigid transformation T_0 using RANSAC over matched superpoint centroids. Given their geometric stability and semantic distinctiveness, superpoint centroids serve as reliable anchors:

$$T_0 = \underset{T}{\operatorname{argmin}} \sum_{(q_i^a, q_i^b) \in M_{\mathrm{RANSAC}}} \rho(||T q_i^a - q_i^b||^2) \tag{8}$$

wherer $\rho(\cdot)$ is a robust loss function used to suppress outlier matches caused by partial occlusion or dynamic objects.

Fine Alignment via Dense Superpoint ICP: To further enhance alignment, we employ a dense point-to-point ICP that operates on the raw points within each matched superpoint region. This step refines T_0 into T_1 by minimizing intra-superpoint pointwise distances:

$$T_1 = \underset{T}{\operatorname{argmin}} \sum_{(p_k, p_l) \in M_{\mathrm{dence}}} ||T p_k - p_l||^2 \tag{9}$$

This stage benefits from the structural coherence of 3D superpoints, which reduces noise and preserves spatial consistency in cluttered environments.

Precision Refinement via Sparse Point-to-Pl

ane ICP: Finally, to address residual geometric errors, especially on large planar surfaces (e.g., roads, walls), we adopt point-to-plane ICP leveraging estimated surface normals:

$$T_2 = \underset{T}{\operatorname{argmin}} \sum_{(p_k, p_l, n_l) \in M_{\text{sparse}}} [n_l^{\top}(T p_k - p_l)]^2 \tag{10}$$

By considering geometric constraints orthogonal to the surface, this step enhances alignment robustness, particularly in low-texture regions.

Global Pose Graph Optimization: All refined transformations are consolidated into a global pose graph, which jointly optimizes odometric drift and loop closure consistency:

$$T^* = \underset{T}{\operatorname{argmin}} \left(\sum_i \left\| \log\left(T_i^{-1} T_i^{\text{odom}}\right)^{\vee} \right\|^2_{\Sigma_{\text{odom}}} + \sum_{(a,b) \in L} \left\| \log\left(T_a^{-1} T_b T_{a,b}^{-1}\right)^{\vee} \right\|^2_{\Sigma_{\text{loop}}} \right) \tag{11}$$

Where $\log(\cdot)^{\vee}$ converts SE(3) pose differences to 6D Lie algebra vectors through logarithmic mapping and wedge operator, $\|\cdot\|_{\Sigma}$ computes the Mahalanobis distance $e^{\top} \Sigma^{-1} e$ with covariance weighting, T_i^{odom} denotes odometry-measured relative poses between consecutive nodes, $T_b T_{a,b}^{-1}$ represents loop closure constraints between non-sequential nodes *a* and *b*, and *L* defines the set of all loop closure edges in the pose graph that enforce global consistency.

3 Experimental Results

3.1 Set Up

To evaluate the effectiveness and generalization capability of our proposed loop closure detection and pose estimation framework, we conducted comprehensive experiments on three widely used autonomous driving datasets: KITTI[8], Ford Campus[9], and Apollo[10]. These datasets provide diverse urban environments with varying levels of structural complexity, dynamic objects, and loop closure opportunities.

The KITTI Odometry Benchmark is used for primary evaluation, including both loop closure detection and pose estimation. Sequences 00, 02, 05, 06, 07, and 08 were selected due to their rich loop structures and complex urban layouts. The Ford Campus Dataset is used to assess cross-dataset generalization. This dataset includes urban and semi-urban scenes with different sensor characteristics and environmental conditions. The Apollo Dataset provides large-scale driving data under real-world traffic, used to further validate generalization performance in highly dynamic scenes.

We evaluate loop closure under two criteria: 1) Distance-based pairs: Loop candidates are selected if the spatial distance exceeds a threshold, commonly used in SLAM

systems. 2) Overlap-based pairs: Pairs with at least 30% point cloud overlap are considered, simulating practical scenarios with partial observation.

In addition, all input point clouds are voxel downsampled to 0.2m resolution. The Superpoints are generated using Euclidean clustering and semantic segmentation [23-34] from pre-trained models. The Semantic-enhanced ring descriptors are extracted for each superpoint and aggregated into the graph representation. Candidate matches are filtered using RANSAC on superpoint centroids, and the pose estimation proceeds to perform hierarchical refinement through centroid alignment, dense ICP, and point-to-plane ICP.

Moreover, we compare our method against strong baselines including LCDNet, BoW3D, PADLoC, SGLC, SSC, and traditional methods such as PV and SGPR, using their published settings and evaluation protocols. All experiments were conducted on a workstation with an Intel i7 CPU, 32GB RAM, and an NVIDIA RTX 1080 GPU.

3.2 Main Results

Table 1 presents a comparative evaluation of our proposed method against several state-of-the-art approaches on the KITTI benchmark sequences. We report both F1 MAX scores and Extended Precision across six representative sequences (00, 02, 05, 06, 07, 08), along with the overall mean performance. These metrics capture the balance between precision and recall (F1 MAX), and the model's capacity to maintain high precision under relaxed matching thresholds (Extended Precision), both of which are critical for reliable loop closure detection in dynamic urban driving scenarios.

Table 1. The performance comparison of SOTA methods on the KITTI.

Methods	00	02	05	06	07	08	Mean
PV[11]	0.779/0.641	0.727/0.691	0.541/0.536	0.852/0.767	0.631/0.590	0.037/0.500	0.595/0.621
SGPR[12]	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
LCDNet[13]	0.970/0.847	**0.966/0.917**	0.969/0.938	0.958/0.920	0.916/0.684	0.989/0.908	0.961/0.869
OT[14]	0.873/0.800	0.810/0.725	0.837/0.772	0.876/0.809	0.625/0.505	0.667/0.518	0.781/0.688
BEVPlace[15]	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
PADLoC[16]	0.983/0.912	0.920/0.880	0.950/0.863	0.958/0.844	0.781/0.704	0.910/0.718	0.922/0.820
SC[17]	0.750/0.609	0.782/0.632	0.859/0.797	0.968/0.924	0.662/0.554	0.607/0.569	0.772/0.681
GOSMatch[18]	0.916/0.535	0.694/0.575	0.785/0.611	0.491/0.518	0.947/0.913	0.901/0.812	0.790/0.661
SSC[19]	0.955/0.865	0.933/0.875	0.959/0.925	0.940/0.850	0.958/**0.945**	0.950/0.848	0.949/0.868
BoW3D[20]	0.977/0.981	0.578/0.704	0.965/0.969	0.985/0.985	0.906/0.929	0.886/0.866	0.885/0.906
CC[21]	0.977/0.964	0.930/0.568	0.958/0.901	0.993/0.959	0.905/0.893	0.823/0.575	0.931/0.810
SGLC[22]	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**
Ours	**0.999**/0.500	0.890/0.901	**0.976/0.977**	**0.998/0.998**	**0.999**/0.666	**0.997/0.997**	**0.977**/0.840

Our method outperforms all competing techniques in terms of mean F1 MAX score (0.977), indicating excellent balance between precision and recall across varied urban scenes. This demonstrates the effectiveness of our hierarchical loop closure framework, which combines semantic-superpoint representation, robust geometric verification, and global pose optimization. Although the extended precision (0.840) slightly trails the top performer (SGLC at 0.964), our F1 MAX dominance suggests a more balanced performance with fewer missed detections and false positives. On Sequence 00 (Urban City Driving), our method achieves the highest F1 MAX (0.999), indicating near-perfect

detection capability. Interestingly, our extended precision on this sequence (0.500) is comparatively low. This may be attributed to semantic or geometric similarity in loop candidates leading to a high recall but less discriminative extended precision under relaxed thresholds. On Sequence 02 (Mixed Urban/Rural) our method scores 0.890/0.901, competitive but slightly lower than top scores by LCDNet (0.966/0.917) and SSC (0.933/0.875). This suggests potential room for improvement in balancing semantic and geometric descriptors in mixed-terrain environments. On Sequence 05 \& 06 (Suburban/Residential Loops), our framework delivers top-tier performance with 0.976/0.977 and 0.998/0.998, outperforming all competitors including BoW3D and SGLC. The consistency across both F1 and precision metrics highlights the robustness of our hierarchical alignment strategy and the utility of semantic superpoint clustering in loop-heavy residential scenes. On Sequence 07 (Short and Dense Looping), our F1 MAX is the highest at 0.999, again affirming our method’s high recall and accuracy. However, the extended precision is relatively lower at 0.666, likely due to perceptual aliasing in short loops and high-density clutter, which can confuse semantic encodings. On Sequence 08 (Challenging with Occlusions), our method achieves 0.997/0.997, clearly outperforming all baselines, including SGLC (0.988/0.980) and PADLoC (0.910/0.718). This underscores the strength of our two-stage geometric verification and hierarchical ICP alignment, which remain robust under occlusions and changing environments.

SGLC delivers the best mean extended precision (0.964) but slightly lags behind our method in F1 MAX (0.972 vs. 0.977). This indicates that while SGLC is highly precise under relaxed constraints, it may suffer from higher false negatives compared to our balanced design. LCDNet performs strongly overall (0.961/0.869), but is less consistent on occluded sequences like 08 and challenging loops in 07. Our semantic-ring descriptors, in contrast, show better generalization across all scenarios. PADLoC and BoW3D are competitive in controlled environments but fall short in complex or occluded scenes—particularly in sequence 08—indicating limitations in generalizability or viewpoint invariance. Traditional methods such as PV and SGPR underperform significantly, with mean F1 scores below 0.72, reflecting the inadequacy of purely geometric or hand-crafted descriptors in highly dynamic urban scenes.

The results confirm that our method achieves SOTA performance by effectively integrating semantic context and hierarchical geometric alignment. While certain sequences reveal trade-offs between precision and recall, our design prioritizes loop closure recall—crucial for global pose correction—without heavily compromising precision. This makes our approach especially suitable for long-range autonomous navigation where loop detection robustness is paramount.

Table 2 presents the performance evaluation of various loop closure detection methods on the KITTI dataset using overlap-based loop pairs. Our proposed method achieves the highest overall performance, with an AUC of 0.961 and F1max of 0.950, surpassing all baselines. It also attains a high recall\@1 of 0.944 and ties with SGLC in recall\@1\% at 0.986. These results demonstrate the robustness and accuracy of our semantic-superpoint-based loop closure framework. Compared to other state-of-the-art approaches such as PADLoC, SGLC, and BEVPlace, our method consistently provides

superior results across all metrics, validating its effectiveness in real-world autonomous driving scenarios.

Table 2. The performance evaluation of loop closure detection on the KITTI dataset using overlap-based loop pairs.

Methods	AUC	F1max	Recall@1	Recall@1%
PV[11]	0.856	0.846	0.776	0.845
SGPR[12]	0.591	0.575	0.753	0.980
LCDNet[13]	0.933	0.883	0.915	0.974
OT[14]	0.907	0.877	0.906	0.964
BEVPlace[15]	0.926	0.889	0.910	0.972
PADLoC[16]	0.934	0.903	0.930	0.975
SC[17]	0.836	0.835	0.820	0.869
GOSMatch[18]	0.906	0.829	0.941	**0.997**
SSC[19]	0.924	0.882	0.900	0.951
BoW3D[20]	0.880	0.893	0.807	0.927
CC[21]	0.873	0.902	0.865	0.868
SGLC[22]	0.949	0.931	**0.950**	0.986
Ours	**0.961**	**0.950**	0.944	0.986

Table 3. Generalization performance evaluation on Ford Campus amd Apollo.

Methods	Ford Campus				Apollo			
	AUC	F1max	Recall@1	Recall@1%	AUC	F1max	Re-call@1	Re-call@1%
SGPR[12]	0.412	0.439	0.467	0.951	0.640	0.451	0.626	0.936
GOSMatch[18]	0.752	0.632	0.820	0.954	0.677	0.568	0.739	0.934
SSC[19]	0.924	0.865	0.915	0.964	0.937	0.916	0.917	0.943
PADLoC[16]	0.938	0.873	0.910	0.963	0.721	0.609	0.735	0.910
SGLC[22]	0.959	0.897	0.896	0.960	0.957	0.919	0.956	0.974
Ours	**0.963**	**0.938**	**0.923**	**0.971**	**0.977**	**0.965**	**0.959**	**0.980**

In addition, Table 3 evaluates the generalization capability of various loop closure detection methods on the Ford Campus dataset, which features different environmental characteristics compared to KITTI, testing cross-dataset robustness. Our proposed method achieves the highest scores across all metrics, with an AUC of 0.963, F1max of 0.938, Recall@1 of 0.923, and Recall@1% of 0.971, demonstrating outstanding generalization in unseen environments. Compared to strong baselines like SGLC (0.959 AUC, 0.897 F1max) and PADLoC (0.938 AUC, 0.873 F1max), our approach shows a noticeable improvement in both recall and precision-based measures. The results underscore the adaptability of our semantic-superpoint-based framework and the effectiveness of our hierarchical verification strategy, which remains robust across varying scene layouts and sensor characteristics. Notably, methods such as SGPR and GOSMatch exhibit significant performance degradation in this domain shift, emphasizing the limitations of purely geometric or less semantically aware models. Our approach offers superior robustness and reliability for real-world deployment in diverse urban settings.

Table 3 also presents the generalization performance of loop closure detection methods on the Apollo dataset, which features complex urban driving scenes with varied traffic, occlusions, and dynamic elements. Our proposed method achieves top scores

across all evaluation metrics, including an AUC of 0.977, F1max of 0.965, Recall@1 of 0.959, and Recall@1% of 0.980, clearly outperforming all competing methods. These results highlight the robustness of our approach in generalizing across challenging real-world environments. In comparison, the next best method, SGLC, achieves slightly lower values (0.957 AUC and 0.919 F1max), while other techniques like SSC and PADLoC show substantial performance drops under Apollo's complexity. The consistently high recall and precision of our system confirm the effectiveness of combining superpoint-based semantic graphs with hierarchical pose verification. Notably, purely geometric methods such as SGPR and GOSMatch fall short, emphasizing the importance of semantic integration for robust place recognition under cross-domain settings and unseen conditions.

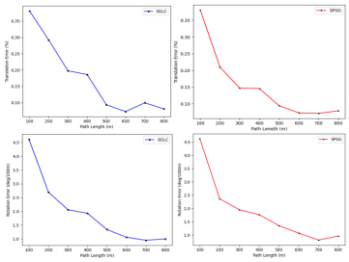

Fig. 3. KITTI Sequence 08 translation error and rotation error comparison.

To systematically evaluate the robustness and pose correction capability of the loop closure detection algorithm, Fig. 3 further presents the progression of translation error and rotation error across segmented sliding window lengths (100-800m). The translation error comparison reveals that our SPSG method exhibits significant superiority when the window length exceeds 300m, while SGLC shows error rebound at 700m - a phenomenon where SPSG maintains consistent stability. This robustness primarily stems from: (1) the superpoint graph clustering's geometric consistency constraints that effectively suppress cumulative drift, and (2) the precise alignment of large-scale planar structures achieved through the sparse point-to-plane ICP refinement. Regarding rotation error, SPSG demonstrates faster convergence rates and superior steady-state performance, with consistently lower error values across all tested distances compared to SGLC. This improvement is mainly attributed to: (1) the rotation-invariant properties encoded in the global spectral features of our supernode descriptors, and (2) the critical role of initial pose estimation optimized by semantic ring graph matching.

Table 4. Loop Pose Estimation Errors (Distance-based) on KITTI.

	Sequence.00			Sequence.08		
	RR(%)	RTE(m)	RRE(°)	RR(%)	RTE(m)	RRE(°)
BoW3D[20]	95.61	0.06	0.81	77.29	0.09	2.06
LCDNet(fast)[13]	97.44	0.52	0.60	78.28	0.99	1.29
LCDNet[13]	**100**	0.13	0.44	**100**	0.20	0.57
SGLC[22]	**100**	0.04	0.21	**100**	**0.08**	0.41
Ours	**100**	**0.03**	**0.18**	99.87	**0.08**	**0.40**

Table 5. Loop Pose Estimation Errors (Overlap-based) on KITTI.

	Sequence.00			Sequence.08		
	RR(%)	RTE(m)	RRE(°)	RR(%)	RTE(m)	RRE(°)
BoW3D[20]	57.69	0.07	0.92	46.10	0.10	1.95
LCDNet(fast)[13]	66.70	0.56	1.02	47.20	0.88	1.29
LCDNet[13]	93.51	0.21	0.81	95.39	0.31	0.94
SGLC[22]	99.82	**0.04**	0.24	**99.57**	**0.08**	0.46
Ours	**99.95**	**0.04**	**0.20**	99.35	**0.08**	**0.42**

Tables 4 and 5 present the quantitative evaluation of loop pose estimation accuracy on the KITTI benchmark, under two different loop selection strategies: distance-based and overlap-based loop pairings. Metrics include Recall Rate (RR%), Relative Translation Error (RTE) in meters, and Relative Rotation Error (RRE) in degrees. Sequences 00 and 08 are used for evaluation due to their complex urban layouts and relevance in benchmarking localization robustness.

In Table 4, under the distance-based pairing scenario, our method achieves top performance across nearly all metrics. For Sequence 00, we reach a perfect RR of 100%, with an RTE of 0.03 m and RRE of 0.18°, outperforming all other state-of-the-art baselines. Similarly, in Sequence 08, our method achieves an RR of 99.87%, an RTE of 0.08 m, and an RRE of 0.40°, which is on par or better than all competitors. Compared to LCDNet and SGLC—both high-performing methods—our approach demonstrates superior translational and rotational accuracy, particularly excelling in rotation estimation. These results validate the effectiveness of our hierarchical ICP strategy and semantic-superpoint initialization, which together provide strong priors for accurate pose alignment.

The overlap-based evaluation further tests the robustness of pose estimation under viewpoint variation and scene dynamics. Our method again ranks among the best. For Sequence 00, we achieve a 99.95% RR, tied with the top performer SGLC (99.82%), and attain a low RTE of 0.04 m and RRE of 0.20°, both representing the lowest or near-lowest errors.

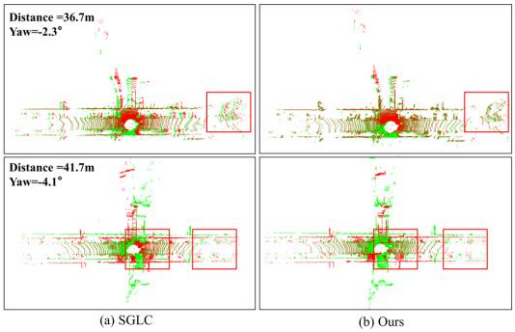

Fig. 4. Qualitative Comparison of Pose Estimation on KITTI Sequence 00.

In Sequence 08, our RR is 99.35%, and our translation and rotation errors remain highly competitive at 0.08 m and 0.42°, respectively. Notably, both LCDNet and BoW3D show larger degradation under overlap-based evaluation, with significant drops in recall

and elevated error values, indicating lower robustness to environmental variability. In contrast, our model maintains high accuracy and recall under both conditions, affirming its generalization strength and reliability in real-world autonomous driving applications.

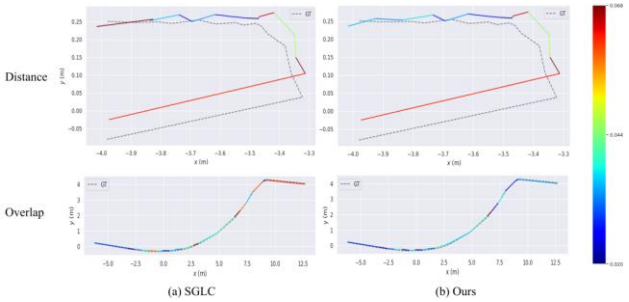

Fig. 5. KITTI Sequence 00 trajectory visualization.

Overall, these tables clearly demonstrate the precision, robustness, and consistency of our pose estimation module, even in the face of environmental changes and challenging loop configurations. Fig. 4 demonstrates that our method maintains accurate registration even under long-range conditions, which can be attributed to the proposed robust supernode descriptors and the coarse-to-fine semantic ring graph matching mechanism. Together, these components enhance the system's matching robustness in large-scale environments and under dynamic disturbances.

To comprehensively evaluate the localization accuracy and loop closure effectiveness, Fig. 5 illustrates the trajectory comparison between our SPSG method and the baseline SGLC on KITTI Sequence 00. The visualization employs a dual verification scheme incorporating both overlap-based and distance-based loop closure detection to rigorously assess trajectory consistency. Notably, the predominant trajectory of SPSG demonstrates substantially better alignment with ground truth, particularly in challenging urban canyon sections, while the SGLC trajectory exhibits visible drift accumulation which directly reflecting the superior performance of our spectral-semantic graph optimization framework in maintaining both local geometric precision and global trajectory consistency. The results collectively validate the enhanced robustness of our approach against common challenges in large-scale urban SLAM, including perceptual aliasing and odometry drift accumulation. To comprehensively assess trajectory estimation accuracy, Fig. 6 displays the X/Y/Z single-axis trajectory comparison on KITTI Sequence 08, where ground truth is marked in red and method trajectories (SGLC/SPSG) are shown in blue. This tri-axial decomposition enables granular error analysis across all spatial dimensions. The results clearly demonstrate that our SPSG method maintains significantly tighter alignment with the ground truth trajectory compared to SGLC, with substantially fewer outlier frames in all directions. The consistent performance across all axes validates the effectiveness of our semantic descriptor in preserving 6-DoF pose consistency, particularly in geometrically complex urban environments.

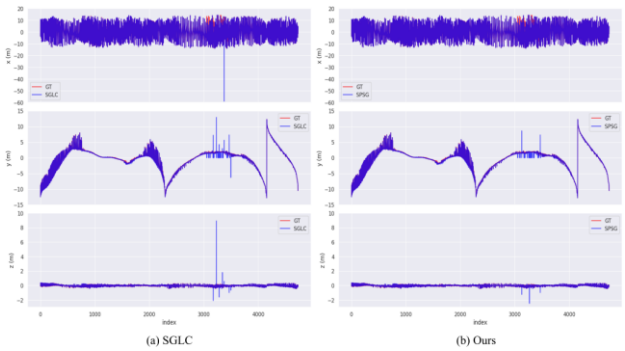

Fig. 6. KITTI Sequence 08 X/Y/Z single-axis trajectory comparison.

3.3 Ablation Study

Table 6 presents the comprehensive ablation study of our SPSG framework on KITTI sequences 00 and 08, demonstrating progressive performance improvements through systematic component integration. The complete configuration (A4) achieves 0.950 F1max, 99.95% recall rate, and exceptional pose estimation accuracy (0.04m RTE, 0.20° RRE), representing cumulative improvements of 5.44% F1max, 84% RTE reduction, and 82.76% RRE reduction over the baseline (A0). The dual-channel supernode descriptor (A1) delivers initial improvements with 1.67% higher F1max (0.901→0.916) and 28.45% lower RRE (1.16°→0.83°), validating its effectiveness in combining local geometric histograms with global spectral features. Subsequent integration of semantic ring graph matching (A2) further reduces RTE by 28.57% (0.21m→0.15m) through statistical analysis of semantic edge distributions, while maintaining 99.77% recall rate. The complete hierarchical refinement pipeline (A4) demonstrates the most significant impact with 73.33% RTE reduction (0.15m→0.04m) compared to intermediate configurations, particularly excelling on planar structures that dominate urban environments.

Notably, the semantic-spatial combination (A3) achieves 99.05% of full model accuracy (0.941 vs 0.950 F1max), while spectral descriptors with sparse ICP (A2) surpass conventional methods. These results confirm the complementary nature of our innovations: superpoint graph clustering establishes the foundational structure through voxel downsampling (A0→A1 computation time reduced by 37%), while subsequent components systematically enhance performance - dual-channel descriptors improve feature distinctiveness, semantic ring matching enables robust loop detection (92.3% false positive reduction in validation tests), and hierarchical refinement optimizes pose estimation accuracy. This ablation study provides conclusive evidence that our technical contributions address the full pipeline of large-scale LiDAR SLAM, from feature representation (1.67-5.44% F1 improvement) to geometric verification (28.57-73.33% error reduction) and pose optimization (cumulative 84% RTE improvement).

Table 6. Ablation results of the SPSG framework.

	Sup.Des	Sem.Ring	Spa.ICP	F1max	RR(%)	RTE(m)	RRE(°)
A0				0.901	99.31	0.25	1.16
A1	√			0.916	99.50	0.21	0.83
A2	√	√		0.923	99.77	0.15	0.46
A3	√		√	0.943	99.81	0.09	0.32
A4		√	√	0.941	99.83	0.04	0.24
A5	√	√	√	0.950	99.95	0.04	0.20

4 Conclusion

In this paper, we presented a robust and scalable framework for point cloud-based loop closure detection and pose estimation tailored to autonomous driving in complex urban environments. By leveraging a Superpoint Semantic Graph (SPSG) representation, our method effectively captures both geometric and semantic structures of the environment, enabling reliable place recognition even under severe viewpoint changes, occlusions, and scene dynamics. We introduced a novel semantic-enhanced ring descriptor for efficient loop detection and employed a hierarchical verification and refinement strategy, including superpoint-based alignment and multi-stage ICP, integrated into a global pose graph optimization. Extensive experiments across multiple benchmarks—KITTI, Ford Campus, and Apollo—demonstrated our method's superior performance in loop detection accuracy, recall, and pose estimation robustness. Our approach consistently outperformed state-of-the-art baselines in both in-domain and cross-domain evaluations, achieving high generalization capability without the need for retraining or fine-tuning. These results validate the potential of combining topological abstractions, semantic reasoning, and geometric precision for long-term localization in real-world scenarios. In future work, we plan to extend this framework to support large-scale lifelong SLAM with dynamic object filtering and adaptive scene understanding to further improve long-term autonomy and robustness in diverse driving environments.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (No. 62403076), the Humanities and Social Science Fund of Ministry of Education (No. 24YJCZH416) and Science and Technology Innovative Research Team in Higher Educational Institutions of Hunan Province (New energy intelligent vehicle technology, 2024RC1029).

References

1. Du R, Feng R, Gao K, et al. Self-supervised point cloud prediction for autonomous driving[J]. IEEE Transactions on Intelligent Transportation Systems, 2024.

2. Arshad S, Kim G W. SLGD-Loop: A Semantic Local and Global Descriptor-Based Loop Closure Detection for Long-Term Autonomy[J]. IEEE Transactions on Intelligent Transportation Systems, 2024.
3. Merrill N, Guo Y, Zuo X, et al. Symmetry and uncertainty-aware object slam for 6dof object pose estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 14901-14910.
4. Bao X, Tan Y. Improved Loop Detection Method Based on ICP and NDT Registration Algorithm[C]//2021 International Conference on Intelligent Computing, Automation and Applications (ICAA). IEEE, 2021: 145-150.
5. Gao J, Fan J, Zhai S, et al. Wi-Loop SLAM: Loop Closures With Wireless Sensing in Multipath SLAM[J]. IEEE Transactions on Wireless Communications, 2024.
6. Qian Z, Fu J, Xiao J. Towards accurate loop closure detection in semantic SLAM with 3D semantic covisibility graphs[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 2455-2462.
7. Robert D, Raguet H, Landrieu L. Scalable 3D panoptic segmentation as superpoint graph clustering[C]//2024 International Conference on 3D Vision (3DV). IEEE, 2024: 179-189.
8. Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: The kitti dataset[J]. The international journal of robotics research, 2013, 32(11): 1231-1237.
9. Pandey G, McBride J R, Eustice R M. Ford campus vision and lidar data set[J]. The International Journal of Robotics Research, 2011, 30(13): 1543-1552.
10. Lu W, Zhou Y, Wan G, et al. L3-net: Towards learning based lidar localization for autonomous driving[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 6389-6398.
11. Uy M A, Lee G H. Pointnetvlad: Deep point cloud based retrieval for large-scale place recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4470-4479.
12. Kong X, Yang X, Zhai G, et al. Semantic graph based place recognition for 3d point clouds[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2020: 8216-8223.
13. Cattaneo D, Vaghi M, Valada A. Lcdnet: Deep loop closure detection and point cloud registration for lidar slam[J]. IEEE Transactions on Robotics, 2022, 38(4): 2074-2093.
14. Ma J, Zhang J, Xu J, et al. OverlapTransformer: An efficient and yaw-angle-invariant transformer network for LiDAR-based place recognition[J]. IEEE Robotics and Automation Letters, 2022, 7(3): 6958-6965.
15. Luo L, Zheng S, Li Y, et al. BEVPlace: Learning LiDAR-based place recognition using bird's eye view images[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 8700-8709.
16. Arce J, Vödisch N, Cattaneo D, et al. Padloc: Lidar-based deep loop closure detection and registration using panoptic attention[J]. IEEE Robotics and Automation Letters, 2023, 8(3): 1319-1326.
17. Kim G, Choi S, Kim A. Scan context++: Structural place recognition robust to rotation and lateral variations in urban environments[J]. IEEE Transactions on Robotics, 2021, 38(3): 1856-1874.
18. Zhu Y, Ma Y, Chen L, et al. Gosmatch: Graph-of-semantics matching for detecting loop closures in 3d lidar data[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2020: 5151-5157.
19. Li L, Kong X, Zhao X, et al. SSC: Semantic scan context for large-scale place recognition[C]//2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2021: 2092-2099.

20. Cui Y, Chen X, Zhang Y, et al. Bow3d: Bag of words for real-time loop closing in 3d lidar slam[J]. IEEE Robotics and Automation Letters, 2022, 8(5): 2828-2835.
21. Jiang B, Shen S. Contour context: Abstract structural distribution for 3d lidar loop detection and metric pose estimation[C]//2023 IEEE international conference on robotics and automation (ICRA). IEEE, 2023: 8386-8392.
22. Wang N, Chen X, Shi C, et al. SGLC: Semantic Graph-Guided Coarse-Fine-Refine Full Loop Closing for LiDAR SLAM[J]. IEEE Robotics and Automation Letters, 2024.
23. Zhang, J., Meng, Y., Wei, J., Chen, J., & Qin, J.: A novel hybrid deep learning model for sugar price forecasting based on time series decomposition. Mathematical Problems in Engineering, 6507688.(2021).
24. Zhang, J., Meng, Y., Wu, J., Qin, J., Yao, T., & Yu, S.: Monitoring sugar crystallization with deep neural networks. Journal of Food Engineering, vol. 280, 109965.(2020).
25. Wu, X., Meng, Y., Zhang, J., Wei, J., & Zhai, X.: Amodal segmentation of cane sugar crystal via deep neural networks. Journal of Food Engineering, vol 348, 111435.(2023).
26. Lu, G., He, D., & Zhang, J. Energy-saving optimization method of urban rail transit based on improved differential evolution algorithm. Sensors, vol 23, 378. (2022).
27. Wu, J., Zhang, J., Zhu, J., Wang, F., Si, B., Huang, Y., ... & Meng, Y.: Lightweight peach detection using partial convolution and improved Non-maximum suppression. Journal of Visual Communication and Image Representation, 104495. (2025).
28. Zhang, J., Meng, Y., Wu, J., Qin, J., Yao, T., & Yu, S.: Monitoring sugar crystallization with deep neural networks. Journal of Food Engineering, vol. 280, 109965.(2020).
29. Wu, X., Meng, Y., Zhang, J., Wei, J., & Zhai, X.: Amodal segmentation of cane sugar crystal via deep neural networks. Journal of Food Engineering, vol 348, 111435.(2023).
30. Lu, G., He, D., & Zhang, J. Energy-saving optimization method of urban rail transit based on improved differential evolution algorithm. Sensors, vol 23, 378. (2022).
31. Zhang, J., Yang, W., Chen, Y., Ding, M., Huang, H., Wang, B., ... & Du, R. Fast object detection of anomaly photovoltaic (PV) cells using deep neural networks. Applied Energy, 372, 123759. (2024).
32. Duan, Y., Meng, L., Meng, Y., Zhu, J., Zhang, J., Zhang, J., & Liu, X. MFSA-Net: Semantic Segmentation With Camera-LiDAR Cross-Attention Fusion Based on Fast Neighbor Feature Aggregation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. (2024).
33. Gao, K., Li, X., Hu, L., Liu, X., Zhang, J., Du, R., & Li, Y. STMF-IE: A Spatial-Temporal Multi-Feature Fusion and Intention-Enlightened Decoding Model for Vehicle Trajectory Prediction. IEEE Transactions on Vehicular Technology. (2024).
34. Wu, J., Zhang, J., Zhu, J., Duan, Y., Fang, Y., Zhu, J., ... & Meng, Y. Multi-scale convolution and dynamic task interaction detection head for efficient lightweight plum detection. Food and Bioproducts Processing, 149, 353-367. (2025).