Rethink Semantic Maps as Living Priors: A Lifecycle Framework for Semantic-Aware Navigation in Dynamic Urban Environments

Anonymous author(s)

***Abstract*— Deploying robots in dynamic urban environments demands navigation systems covering the full operational lifecycle: from traversing unknown spaces to building semantic representations, to navigating with them as conditions change. Existing approaches cover only fragments, either assuming a pre-built static map or relying on reactive planners without long-range semantic reasoning. We present a lifecycle framework unifying first-traversal and map-guided navigation. For first-traversal, the robot incrementally constructs an instance-level metric-semantic map via semantic-aware target selection, guided by sparse waypoints. In map-based navigation, a global semantic path on the prior map encodes long-range intent, while target search reconciles it with real-time perception. Both phases share a terrain-aware path search and constrained trajectory optimizer, supported by an online maintenance module that incrementally updates the semantic map to reflect observed changes. We benchmark our mapping module on SemanticKITTI dataset and validate the full navigation stack in Isaac Sim on an urban scenario.**

I. Introduction

Quadrupedal robots are increasingly deployed in urban environments for delivery, patrol, and inspection [1], [2], [3], where traversability depends on semantic context, not geometry alone [4], [5]. Such reasoning is straightforward given a detailed semantic map [6], [7], [8], yet none exists upon first deployment [9], forcing the robot to simultaneously explore, build a representation, and progress toward its goal. Even once built, the map degrades over time due to moving objects, construction, or shifting terrain [10]. Navigation is thus a continuous operational lifecycle [11] spanning semantic exploration, long-range intent-driven navigation, and dynamic updates to handle environmental change.

Current approaches address only fragments of this lifecycle. Metric-semantic mapping systems [6], [12], [7], [8] build increasingly rich semantic and spatio-temporal representations, yet focus on the map as a deliverable and with little attention to how it is consumed by or co-evolves with the planner. On the planning side, three families each cover a different fragment. Exploration planners such as TARE [13] and GBPlanner [14] efficiently discover unknown space but optimize volumetric completeness rather than semantic terrain or a specific goal. Map-based planners built on D* Lite [15], DWA [16], and layered costmaps [17] cover long distances with semantic-aware costs but assume a static, pre-built map. Reactive local planners such as ViPlanner [18], CRESTE [19], BehAV [20], and Sem-NaVAE [21] reason about semantic traversability from onboard perception but are short-horizon and cannot leverage a prior map. The common *lifecycle gap*: no system maintains semantic terrain reasoning and route validity from first traversal of an unknown environment to repeated navigation of an evolving one.

We present a navigation framework addressing the *full lifecycle* from first-time traverse to repeated map-guided navigation under changing conditions. It comprises two phases:

- **Phase I—Exploratory Navigation.** With no prior map, the robot incrementally builds an instance-level metric-semantic submap collection while navigating towards the goal, with the guidance direction set to the next waypoint to steer the robot onto semantically preferred surfaces in unknown territory.
- **Phase II—Map-Guided Navigation with Dynamic Adaptation.** Given a prior map, a global path along preferred terrain encodes long-range intent, against which the planner balances live observations to select local actions; a map maintenance module reconciles the prior with current perception and removes stale regions.

Our contributions are:

1) An instance-level metric-semantic mapping system with *online maintenance*.
2) A two-phase hierarchical planning architecture that unifies the full lifecycle, sharing a common backbone of semantics-aware path search, and differing only in how the guidance direction is set.
3) Validating mapping on SemanticKITTI and closed-loop planning in Isaac Sim on urban scenarios.

II. Related Work

A. Metric-Semantic Mapping for Urban Navigation

A robot navigating urban terrain must distinguish geometrically similar yet semantically distinct surfaces, beyond classical occupancy grids [22]. Dense semantic mapping bridges this gap by fusing geometry with per-element labels. Early indoor systems such as SemanticFusion [23] combined segmentation with TSDF volumes on Voxblox [24], yielding instance-aware maps for small static scenes. Kimera [6] builds real-time metric-semantic mesh and 3D dynamic scene graphs with category-level labels. [7], [12] use TSDF submaps with panoptic labels. Khronos [8] unifies motion tracking with long-term change detection. Despite these advances, existing systems focus on representation rather than planner coupling, treating the map as a deliverable.

B. Autonomous Exploration and Long-Range Navigation

Long-distance navigation in unknown environments requires more than local obstacle avoidance. Frontier-based exploration [25] and its extensions, including information-gain sampling [26] and topological roadmaps [14], [13],

target volumetric coverage without semantics or goal directedness. The classical global-local planning stack (e.g., D* Lite [15] with DWA [16] or layered costmaps [17]) assumes a static prior map with no mechanism for environmental change. Foundation-model approaches such as GNM [27], ViNT [28], and NoMaD [29] learn visual navigation policies from large datasets but build no explicit map and remain bounded by instantaneous perception. Our work unifies these strands into a lifecycle framework spanning first-time traversal of unknown space to repeated navigation under change.

C. Semantics-Driven Path Planning

A growing body of work incorporates semantics into planning via learned costs or end-to-end policies. Traversability analysis has a long history in field robotics [4]. [30] adds semantic segmentation to elevation maps distinguishes geometrically similar terrain. ViPlanner [18] trains an imperative-learning local planner jointly on depth and semantic image. BehAV [20] builds a behavioral costmap from vision-language outputs. CRESTE [19] distills visual foundation features into a BEV reward map via counterfactual inverse RL. Sem-NaVAE [21] generates trajectory hypotheses with a conditional VAE and selects via open-vocabulary segmentation. These methods rely on current observations alone, whereas our framework fuses semantics accumulated across past traversals with real-time perception.

III. Problem Formulation

Consider a robot navigating a dynamic urban environment characterized by a semantic map $S : \mathbf{p} \to \mathcal{S}$, where each class in $\mathcal{S}$ incurs a traversal cost $c : \mathcal{S} \to \mathbb{R}^+$. We seek an optimal trajectory $\mathcal{P}^*$ in the collision-free space $\mathcal{Q}_{\text{free}}$ from $\mathbf{x}_0$ to goal g minimizing the cumulative semantic cost:

$$\min_{\mathcal{P} \subset \mathcal{Q}_{\text{free}}} J(\mathcal{P}) = \int_{\mathcal{P}} c(S(\mathbf{p}))\, ds. \tag{1}$$

This process decomposes into two phases: **First-Traversal Semantic Navigation.** With S unknown and built online, the robot selects $\mathbf{q}_t^*$ along exploration boundary Ω_t, balancing progress with semantic preference:

$$\mathbf{q}_t^* = \arg\min_{\mathbf{q} \in \Omega_t} \Big[\underbrace{d(\mathbf{q}, g)}_{\text{Goal Progress}} + \lambda \cdot \underbrace{c(S(\mathbf{q}))}_{\text{Semantic Preference}} \Big], \tag{2}$$

Map-Guided Navigation with Dynamic updates. Given a prior map S, a global path $\mathcal{P}^g$ is pre-planned, and operator Φ fuses current observations $\mathcal{O}_t$ into the working map:

$$S_{t+1}(\mathbf{p}) = \Phi(S_t(\mathbf{p}), \mathcal{O}_t)\,. \tag{3}$$

The robot selects $\mathbf{q}_t^*$ in perception field $\mathcal{B}_t$, balancing deviation from $\mathcal{P}^g$ against the semantic cost J_{loc}:

$$\mathbf{q}_t^* = \arg\min_{\mathbf{q} \in \mathcal{B}_t} \Big[\underbrace{\lceil(\mathbf{q}, \mathcal{P}^g)}_{\text{global intent}} + \lambda \cdot \underbrace{c(S(\mathbf{q}))}_{\text{Semantic Preference}} \Big]. \tag{4}$$

IV. Methodology

Our framework (Fig. 1) unifies semantic mapping with a two-phase planner to bridge the operational lifecycle from initial exploration to persistent navigation.

A. Incremental Metric-Semantic Mapping

The environment is a collection of TSDF submaps $\mathcal{M} = \{M_k\}_{k=1}^K$, where each submap corresponds to a spatially coherent object with panoptic label $\ell_k = (s_k, i_k)$. This representation enables efficient spatial indexing and change-aware updates without reprocessing the global map.

State Estimator. The robot's 6-DoF pose $\mathbf{x}_t = (\mathbf{R}_{I_t}^w, \mathbf{t}_{I_t}^w)$ is estimated by tightly-coupled LiDAR-Inertial Odometry [31], which registers raw LiDAR scan against a global map through an on-manifold iterated error-state Kalman filter minimizing point ${}^L\mathbf{p}_j$ to plane $(\mathbf{u}_j^w, \mathbf{q}_j^w)$ residuals:

$$0 = \mathbf{u}_j^{w\top} \big[\mathbf{R}_{I_t}^w \, {}^I\mathbf{T}_L \, {}^L\mathbf{p}_j + \mathbf{t}_{I_t}^w - \mathbf{q}_j^w \big], \tag{5}$$

Perception. A panoptic segmentation network [32] produces per-pixel semantic class $s \in \mathcal{S}$ and instance ID $i \in \mathbb{N}$ for each camera frame I_t. Labeled pixels are lifted to 3D via LiDAR returns using the camera-LiDAR extrinsic ${}^C\mathbf{T}_L$, and points sharing the same (s, i) label form a segment.

Mapping. The incoming segment associated with existing submaps based on semantic labels and 3D IoU:

$$k^* = \begin{cases} \{k \mid s_k = s\}, & s \in \mathcal{S}_{\text{stuff}}, \\ \arg\max_k \ \text{IoU}_{\text{3D}}(\text{seg}, M_k), & s \in \mathcal{S}_{\text{thing}}. \end{cases} \tag{6}$$

Unmatched segments with enough points create submaps. Labeled points are integrated into their associated submaps through bundled ray-tracing. Every voxel v traversed by a ray with signed distance d_v and weight w_v updates its TSDF by weighted averaging:

$$D_v \leftarrow \frac{W_v D_v + w_v \, \Gamma(d_v, \tau)}{W_v + w_v}, \quad W_v \leftarrow W_v + w_v, \tag{7}$$

where $\Gamma(\cdot, \tau)$ clamps the signed distance to $[-\tau, \tau]$. Per-voxel label histograms are resolved by maximum-count voting, and the submap label ℓ_k takes the dominant voxel label. This temporal aggregation filters out transient errors and yields instance-consistent geometry and semantics. Independent submap creation can produce multiple submaps occupying the same physical region, inflating retrieval cost. To enforce spatial exclusivity, co-located voxels $v_a \in M_a$ and $v_b \in M_b$ are periodically fused

$$D_{v^*} = \frac{W_{v_a} D_{v_a} + W_{v_b} D_{v_b}}{W_{v_a} + W_{v_b}}, \quad W_{v^*} = W_{v_a} + W_{v_b}, \tag{8}$$

and reassigned to the submap matching its dominant instance ID, ensuring each voxel is owned by exactly one instance.

Submap-level Change Updates. For long-term deployment, the map must reflect the current states and remove stale information. Every iso-surface point $p \in \mathcal{P}_k$ in untracked submaps M_k queries its signed distance $\text{sdf}_{\text{act}}(p)$ in overlapping traclked submaps and is classified as

$$\phi(p) = \begin{cases} \text{conflict} & |\text{sdf}_{\text{act}}(p)| > \xi_{sdf}, \\ \text{agree} & |\text{sdf}_{\text{act}}(p)| \le \xi_{sdf}, \end{cases} \tag{9}$$

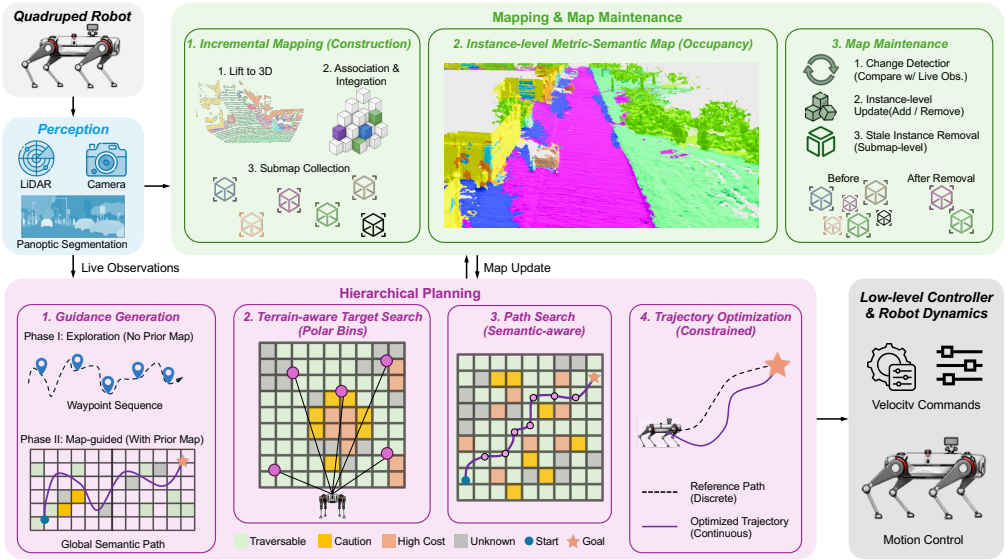

Fig. 1: The framework unifies instance-level metric-semantic mapping, online map maintenance, and two-phase planning. Multimodal perception incrementally builds an instance-level metric-semantic map during first traversal and maintains it via submap-level change detection upon redeployment. Both phases share a common backbone: following either sparse waypoints for exploration or a global semantic path on the prior map, and employing semantic-aware target search, followed by path search, and constrained trajectory optimization.

Submap is removed as a whole when conflict ratio $R_k^{\text{con}} = \frac{N_k^{\text{con}}}{|\mathcal{P}_k^{\text{obs}}|} > \rho_{\text{thr}}$, preserving instance-level consistency.

B. Hierarchical Goal-Conditioned Planning

The planning framework distills global mission objectives into local reactive execution, supporting both first-time traversal and map-guided deployment. Each cycle proceeds as: (i) select a guidance direction representing global intent; (ii) polar swaths produce reachable, terrain-admissible candidates, from which a local search retains the minimum-cost path; (iii) the path serves as reference for downstream trajectory optimization and tracking.

Guidance Direction. The guidance direction $\theta^\star$ steers the subsequent search by pointing toward a guide point $\mathbf{g}$:

$$\theta^\star = \operatorname{atan2}(\mathbf{g} - \mathbf{x}), \tag{10}$$

Without a prior map, g is the direct bearing from the current pose $\mathbf{x}$ to the next waypoint $\mathbf{w}_j$; when a prior map and pre-computed global path $\mathcal{P}^{\text{g}}$ are available, $\theta^\star$ instead points toward the guide point g, defined as the intersection of $\mathcal{P}^{\text{g}}$ with the local map.

Candidate Scoring and Path Search. The $\pm 90^\circ$ field around $\theta^\star$ is split into bins of width $2R_{\text{robot}}$. Each bin carries a terminal label ℓ, collision-free reach ρ, and mean traversal cost $\bar{c}$. Bins are ranked by

$$C = \underbrace{w_1\,\tilde{\delta}}_{\text{direction}} + \underbrace{w_2\,w(\ell)}_{\text{terminal}} + \underbrace{w_3\,\bar{c}}_{\text{corridor}} + \underbrace{w_4\,(\rho_{\max} - \rho)}_{\text{reach}}, \tag{11}$$

where $\tilde{\delta}$ is the normalized angular deviation from $\theta^\star$ and $w(\ell)$ is the per-class traversal weight. The top-K bins are expanded by A* with edge costs combining geometric distance and $w(\ell)$, and the lowest-cost collision-free path is retained as the local reference $\mathcal{P}_t^L$.

Trajectory Optimization and Tracking. $\mathcal{P}_t^L$ is handed to a downstream optimizer that produces a time-parameterized, smooth trajectory respecting kinodynamic limits and obstacle clearance. The trajectory is then executed by a high-frequency tracking controller.

V. Experiments

We benchmark our panoptic mapping module on SemanticKITTI (Sec. V-A) and deploy the full navigation stack in Isaac Sim on an urban sidewalk scenario (Sec. V-B).

A. Mapping Evaluation

Dataset. We evaluate on SemanticKITTI [33], a large-scale urban driving benchmark camera images, LiDAR point clouds, and panoptic labels.

Baselines. We compare against Kimera [6], which builds a single TSDF map with only semantic labels, and Panmap [7], which also maintains panoptic submaps.

Metrics. For label quality we report 3D Panoptic, Recognition, and Segmentation Quality (PQ, RQ, SQ), mIoU and pixel Accuracy, and instance mAP at IoU $\in \{0.30, 0.50\}$. For geometry we report Chamfer-L1 (C-L1), mesh Accuracy and Completion, and F-score.

Results. Table I shows our method attaining the best PQ (0.568) and SQ (0.340), outperforming Panmap by 6.4 % and 38.2 % respectively, which we attribute to inter-submap label fusion that resolves boundary inconsistencies in overlapping submaps. Instance mAP also leads at every IoU threshold ($\text{mAP}_{50} = 0.758$), confirming reliable instance-level association. For geometry, the lowest C-L1 (0.0549 m) and highest F-score (56.34) indicate that panoptic-aware fusion does not compromise the underlying surface quality. Kimera retains a small edge on mIoU and Acc. due to its single-volume formulation, but cannot disentangle instances, which is a capability essential for downstream navigation.

TABLE I: Mapping quality comparison on SemanticKITTI. We report panoptic metrics (PQ, RQ, SQ), semantic metrics (mIoU, Accuracy), instance metrics (mAP at IoU = 0.30 and 0.50), and geometric reconstruction quality (Acc., Comp., C-L1, F-score). **Bold**: best; underline: second best. Kimera produces only semantic labels and dense geometry; panoptic/instance metrics are not applicable ("–").

Method	Panoptic			Semantic		Instance		Geometry			
	PQ↑	RQ↑	SQ↑	mIoU↑	Acc.↑	mAP_{30}↑	mAP_{50}↑	Acc.↓	Comp.↓	C-L1↓	F↑
Kimera	–	–	–	**0.338**	**0.770**	–	–	**0.0430**	0.1871	0.1151	55.54
Panmap	0.534	**0.804**	0.246	0.271	0.679	0.546	0.532	0.0688	0.3119	0.1904	9.16
Ours	**0.568**	0.701	**0.340**	0.263	0.713	**0.769**	**0.758**	0.0506	**0.0584**	**0.0549**	**56.34**

TABLE II: Terrain Classification and Walkability Mapping

Walkability	Terrain Types
Preferred	Sidewalk, Crosswalk
Allowed	Grass, Curb ramp, Other walkable
Forbidden	Road, Non-walkable surfaces
Obstacle	Static/dynamic obstacles

TABLE III: Navigation performance and trajectory semantic distribution. SR: Success Rate; Coll.: Collision Rate; DTG: Distance to Goal; Phase I: first-traversal navigation without prior map; Phase II: map-guided navigation using the prior map constructed by our Phase I module. **Bold**: best per phase.

Method	SR (%)↑	Coll. (%)↓	DTG (m)↓	Pref. (%)↑	Allow. (%)	Forb. (%)↓	Unkn. (%)↓
Phase I — First-Traversal Navigation (no prior map)							
GH	33.3	22.2	8.71	60.4	0.0	3.1	36.5
Ours	**66.7**	**11.1**	**0.37**	**88.2**	0.3	**1.6**	**9.9**
Phase II — Map-Guided Navigation (with prior map)							
PF	55.6	22.2	8.58	59.0	0.0	4.0	37.0
Ours	**77.8**	**11.1**	**0.43**	**86.4**	0.0	**4.5**	**9.0**

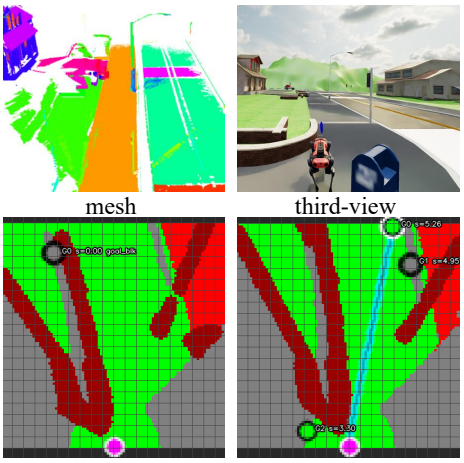

Fig. 2: Top: panoptic mesh (colors denote distinct submaps) and corresponding simulator view. Bottom: local costmaps for GH (left) and Ours (right); green = preferred, red = forbidden, gray = unknown. GH selects guide point at the goal bearing but finds it blocked. Ours evaluates multiple candidatess and routes through semantically preferred terrain (cyan path).

B. Planning Evaluation

Environment and cost. We use an urban sidewalk scene with coexisting sidewalk, road, crosswalk, and grass terrains and moderate obstacle density. Each terrain class is assigned a traversal level reflecting walkability (Table II).

Scenario. Sparse waypoints are placed such that each consecutive pair admits geometrically feasible paths with distinct semantic costs. We run 9 trials per method per phase.

Baselines. We evaluate the planner in two regimes, each against a baseline. *Phase I.* With only local perception, the guide point $\mathbf{g}$ must lie inside the local map. We compare against **Greedy Heading** (GH), which sets $\mathbf{g}$ to the direct bearing from the robot to the next waypoint. *Phase II.* A prior map and pre-computed global path $\mathcal{P}^{\mathrm{g}}$ are available, but the environment may change at execution time. We compare against **Pure-Follower** (PF), which directly tracks the slice of $\mathcal{P}^{\mathrm{g}}$ ahead of the robot with no local replanning.

Metrics. *Navigation performance:* Success Rate (SR↑), Collision Rate (Coll.↓), and Distance-To-Goal (DTG↓). *Semantic quality:* traversal distribution over Preferred (↑), Allowed, Forbidden (↓), and Unknown (↓) classes.

Results. Table III reports navigation performance alongside the semantic distribution of traversed trajectories. In Phase I, our method doubles the success rate of GH (66.7% vs. 33.3%) while halving the collision rate, and maintains 88.2% preferred-terrain coverage versus 60.4% for GH. The DTG gap is striking: GH averages 8.71 m from the goal ring while our method reaches 0.37 m, showing that semantic-aware frontier ranking not only improves success but also enables substantially better terrain compliance. Fig. 2 illustrates a representative case. In Phase II, the prior map improves goal-reaching for both methods (SR rises to 77.8% for ours and 55.6% for PF), yet PF's collision rate remains at 22.2% and its semantic profile barely changes (Pref. 59.0%, Unkn. 37.0%). A prior map and global path improve goal-reaching for both methods, but the environment may change between traversals; only a planner that reasons over terrain costs and updates the map online can translate map knowledge into consistently safer and socially compliant trajectories.

VI. Conclusion

We addressed the *lifecycle gap* in semantic urban navigation by treating the semantic map as a continually evolving asset. Our framework couples instance-level metric-semantic mapping, built incrementally and maintained against live observations, with a two-phase hierarchical planner sharing a common backbone. Experiments on SemanticKITTI and in Isaac Sim validate the approach. Future work includes pedestrian intent prediction, interactive avoidance of dynamic agents, and uncertainty-aware planning that incorporates semantic confidence into cost estimation planning.

REFERENCES

[1] C. Gehring, P. Fankhauser, L. Isler, R. Diethelm, S. Bachmann, M. Potz, L. Gerstenberg, and M. Hutter, "Anymal in the field: Solving industrial inspection of an offshore hvdc platform with a quadrupedal robot," in *Field and service robotics: results of the 12th international conference*. Springer, 2021, pp. 247–260.

[2] M. Tranzatto, T. Miki, M. Dharmadhikari, L. Bernreiter, M. Kulkarni, F. Mascarich, O. Andersson, S. Khattak, M. Hutter, R. Siegwart *et al.*, "Cerberus in the darpa subterranean challenge," *Science Robotics*, vol. 7, no. 66, p. eabp9742, 2022.

[3] Y. Fan, Z. Pei, C. Wang, M. Li, Z. Tang, and Q. Liu, "A review of quadruped robots: Structure, control, and autonomous motion," *Advanced Intelligent Systems*, vol. 6, no. 6, p. 2300783, 2024.

[4] P. Papadakis, "Terrain traversability analysis methods for unmanned ground vehicles: A survey," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 4, pp. 1373–1385, 2013.

[5] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2017.

[6] A. Rosinol, M. Abate, Y. Chang, and L. Carlone, "Kimera: an open-source library for real-time metric-semantic localization and mapping," in *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2020, pp. 1689–1696.

[7] L. Schmid, J. Delmerico, J. L. Schönberger, J. Nieto, M. Pollefeys, R. Siegwart, and C. Cadena, "Panoptic multi-tsdfs: a flexible representation for online multi-resolution volumetric mapping and long-term dynamic scene consistency," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 8018–8024.

[8] L. Schmid, M. Abate, Y. Chang, and L. Carlone, "Khronos: A unified approach for spatio-temporal metric-semantic slam in dynamic environments," *arXiv preprint arXiv:2402.13817*, 2024.

[9] J. A. Placed, J. Strader, H. Carrillo, N. Atanasov, V. Indelman, L. Carlone, and J. A. Castellanos, "A survey on active simultaneous localization and mapping: State of the art and new frontiers," *IEEE Transactions on Robotics*, vol. 39, no. 3, pp. 1686–1705, 2023.

[10] R. B. Sousa, H. M. Sobreira, and A. P. Moreira, "A systematic literature review on long-term localization and mapping for mobile robots," *Journal of Field Robotics*, vol. 40, no. 5, pp. 1245–1322, 2023.

[11] L. Kunze, N. Hawes, T. Duckett, M. Hanheide, and T. Krajník, "Artificial intelligence for long-term robot autonomy: A survey," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4023–4030, 2018.

[12] T. Hu, J. Jiao, Y. Xu, H. Liu, S. Wang, and M. Liu, "Dhp-mapping: A dense panoptic mapping system with hierarchical world representation and label optimization techniques," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 1101–1107.

[13] C. Cao, H. Zhu, H. Choset, and J. Zhang, "Tare: A hierarchical framework for efficiently exploring complex 3d environments." in *Robotics: Science and Systems*, vol. 5, 2021, p. 2.

[14] T. Dang, M. Tranzatto, S. Khattak, F. Mascarich, K. Alexis, and M. Hutter, "Graph-based subterranean exploration path planning using aerial and legged robots," *Journal of Field Robotics*, vol. 37, no. 8, pp. 1363–1388, 2020.

[15] S. Koenig and M. Likhachev, "D* lite," in *Eighteenth national conference on Artificial intelligence*, 2002, pp. 476–483.

[16] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE robotics & automation magazine*, vol. 4, no. 1, pp. 23–33, 2002.

[17] D. V. Lu, D. Hershberger, and W. D. Smart, "Layered costmaps for context-sensitive navigation," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 709–715.

[18] P. Roth, J. Nubert, F. Yang, M. Mittal, and M. Hutter, "Viplanner: Visual semantic imperative learning for local navigation," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 5243–5249.

[19] A. Zhang, H. Sikchi, A. Zhang, and J. Biswas, "Creste: scalable mapless navigation with internet scale priors and counterfactual guidance," *arXiv preprint arXiv:2503.03921*, 2025.

[20] K. Weerakoon, M. Elnoor, G. Seneviratne, V. Rajagopal, S. H. Arul, J. Liang, M. K. M. Jaffar, and D. Manocha, "Behav: Behavioral rule guided autonomy using vlms for robot navigation in outdoor scenes," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 7044–7051.

[21] G. Olguin and J. Ruiz-del Solar, "Sem-navae: Semantically-guided outdoor mapless navigation via generative trajectory priors," *arXiv preprint arXiv:2602.01429*, 2026.

[22] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, no. 6, pp. 46–57, 2002.

[23] J. McCormac, A. Handa, A. Davison, and S. Leutenegger, "Semanticfusion: Dense 3d semantic mapping with convolutional neural networks," in *2017 IEEE International Conference on Robotics and automation (ICRA)*. IEEE, 2017, pp. 4628–4635.

[24] H. Oleynikova, Z. Taylor, M. Fehr, R. Siegwart, and J. Nieto, "Voxblox: Incremental 3d euclidean signed distance fields for on-board mav planning," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 1366–1373.

[25] A. Topiwala, P. Inani, and A. Kathpal, "Frontier based exploration for autonomous robot," *arXiv preprint arXiv:1806.03581*, 2018.

[26] A. Bircher, M. Kamel, K. Alexis, H. Oleynikova, and R. Siegwart, "Receding horizon" next-best-view" planner for 3d exploration," in *2016 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2016, pp. 1462–1468.

[27] D. Shah, A. Sridhar, A. Bhorkar, N. Hirose, and S. Levine, "Gnm: A general navigation model to drive any robot," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 7226–7233.

[28] D. Shah, A. Sridhar, N. Dashora, K. Stachowicz, K. Black, N. Hirose, and S. Levine, "Vint: A foundation model for visual navigation," *arXiv preprint arXiv:2306.14846*, 2023.

[29] A. Sridhar, D. Shah, C. Glossop, and S. Levine, "Nomad: Goal masked diffusion policies for navigation and exploration," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 63–70.

[30] D. Maturana, P.-W. Chou, M. Uenoyama, and S. Scherer, "Real-time semantic mapping for autonomous off-road navigation," in *Field and Service Robotics: Results of the 11th International Conference*. Springer, 2017, pp. 335–350.

[31] W. Xu, Y. Cai, D. He, J. Lin, and F. Zhang, "Fast-lio2: Fast direct lidar-inertial odometry," *IEEE Transactions on Robotics*, vol. 38, no. 4, pp. 2053–2073, 2022.

[32] K. M. Gunnarsson, "Panoptic segmentation in video," 2022.

[33] 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.