GATE: GATEWAY-AWARE TOPOLOGICAL EXECUTION FOR GLOBALLY COHERENT ZERO-SHOT NAVIGATION UNDER PARTIAL OBSERVABILITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Long-horizon navigation under partial observability requires decisions beyond the current sensing horizon. Existing methods either plan over observed geometry or obtain foresight by exploring, sampling, completing, or learning models of unseen space. We introduce **GATE** (**G**ateway-**A**ware **T**opological **E**xecution), a training-free framework that reasons over the latent topological continuation of observable boundaries rather than reconstructing the hidden world. GATE converts reachable frontiers into gateways, estimates their goal-directed unseen continuation, and commits to promising transitions while executing only through verified free space. We evaluate GATE in controlled occlusion, long-horizon, and real-life locations against classical, belief-space, exploration, sampling-based, and learned planning families, including localization and phantom-obstacle corruption. Results and ablations show that unseen-topology reasoning and gateway commitment improve route coherence, robustness, and goal-reaching without training, privileged maps, or explicit hidden-map reconstruction.

1 INTRODUCTION

Long-horizon navigation under partial observability requires decisions whose consequences extend beyond the robot's current sensing horizon. A ground robot moving through buildings or outdoor infrastructure, or a low-altitude UAV operating among urban structures, may observe only a small fraction of the geometry separating its current state from a distant goal. The globally useful route may initially move away from the goal, pass behind an occlusion, or enter a corridor whose continuation is not yet observable. Thus, the robot must often decide *where to enter unknown space before the geometry validating that decision is available (Fig. 1)*.

This exposes a mismatch between *metric observability* and *decision horizon*. Classical search and sampling methods such as A* and RRT are effective when the relevant free-space structure is represented (Hart et al., 1968; LaValle, 1998), while reactive planners such as DWA efficiently optimize short-range motion (Fox et al., 1997). Under partial observability, however, these methods cannot directly evaluate the structural consequence of entering an unresolved region. Frontier exploration and active-SLAM extend the observed map (Yamauchi, 1997; Zhou et al., 2021; Placed et al., 2023), while belief-space approaches reason explicitly over uncertain futures (Silver and Veness, 2010; Kurniawati, 2022); their objectives are primarily exploration, information acquisition, or belief-space search rather than persistent goal-directed reasoning over the topology hidden beyond a particular observable boundary.

Recent methods extend the planning horizon using stronger predictive priors. LRN learns visual affordances for promising frontier directions (Schmittle et al., 2025), but relies on learned image-space affordances and does not maintain an explicit persistent spatial memory. VoroNav uses semantic topological structure together with language-model reasoning (Wu et al., 2024), while TANGO couples object-level topological goals with local metric control (Podgorski et al., 2025); both rely on richer semantic or reference-topology structure than locally observed geometry alone. World-model approaches instead simulate candidate futures: NWM predicts future visual observations from learned dynamics (Bar et al., 2025), and recent occlusion-aware methods explicitly imagine possible scenes behind occlusions (He et al., 2026). These approaches demonstrate the value of reasoning beyond

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the local map, but obtain foresight through learned affordances, semantic priors, stored topology, or explicit prediction of hidden worlds.

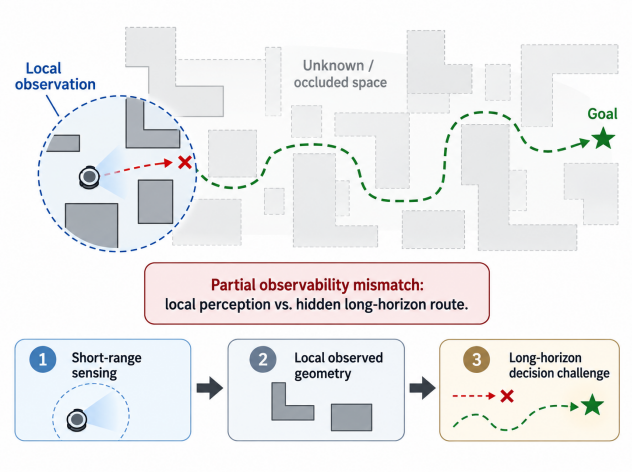


Figure 1: Partial observability creates a mismatch between locally sensed geometry and the hidden long-horizon route required to reach the goal.

We target the gap between these two regimes. Rather than restricting planning to currently observed geometry or predicting the complete unseen environment, we ask whether long-horizon navigation can be driven by estimating only the decision-relevant topology associated with crossing the sensing boundary. Formally, instead of reconstructing an unknown map $\widehat{\mathcal{M}}_{\text{unknown}}$, we seek the utility

$$U(g \mid \mathcal{M}_t, q_{\text{goal}})$$

of entering an unresolved region through an observable transition g, given the current sensor-built map $\mathcal{M}_t$.

We introduce **GATE** (**G**ateway-**A**ware **T**opological **E**xecution), a training-free framework that converts reachable frontier structure into *gateways*: candidate transitions from verified free space into latent topology. GATE evaluates each gateway using known-space feasibility together with evidence for unseen goal-directed continuation, corridor persistence, open-region expansion, and dead-end risk. It therefore predicts the value of entering hidden topology rather than the hidden geometry itself. A persistent commitment mechanism preserves gateway identity across successive partial observations, while execution remains conservative: unseen structure may influence the high-level decision, but motion is certified only through currently observed safe space.

We evaluate GATE in controlled long-horizon and severe-occlusion tasks and in reconstructed real-location 2D and 3D environments. Comparisons span known-space search, reactive control, frontier exploration, sampling-based, belief-space, active-mapping, and learned/imagination-based planning families under a shared perception interface. We further perturb localization and occupancy observations to test whether topological decisions remain useful when their underlying metric evidence is imperfect.

Our contributions are:

- We formulate partially observed long-horizon navigation as *transition selection over latent topology*, avoiding explicit reconstruction of the complete hidden environment.
- We introduce GATE, which combines gateway abstraction, unseen-topology utility, persistent commitment, and observation-certified execution in a training-free planner.
- We isolate the contribution of unseen-topology reasoning through controlled occlusion, horizon, and ablation experiments across diverse planning families and corrupted observations.

2 Related Work

Planning over observed geometry. Classical search and sampling methods such as A*, D* Lite, RRT, and RRT* provide strong geometric planning once traversability is represented (Hart et al., 1968; Koenig and Likhachev, 2002; LaValle, 1998; Karaman and Frazzoli, 2011), while reactive planners such as DWA optimize short-horizon collision-aware motion (Fox et al., 1997). Their limitation in our setting is not geometric planning itself, but that the utility of an unresolved boundary is undefined until additional structure is observed. GATE instead assigns value to such boundaries as candidate transitions into latent topology.

Exploration and belief-space planning. Frontier exploration and active mapping select actions that reveal unknown space or reduce uncertainty (Yamauchi, 1997; Zhou et al., 2021; Chaplot et al., 2020; Placed et al., 2023), while POMDP and POMCP formulations reason over uncertain future states and observations (Silver and Veness, 2010; Kurniawati, 2022). These approaches answer where to explore or which uncertain future to evaluate. GATE instead asks which reachable boundary is most likely to preserve a goal-directed topological route; the frontier is therefore evidence about a transition rather than the exploration objective itself.

Learned and predictive long-horizon navigation. Recent methods obtain foresight through learned affordances, semantic structure, or explicit future prediction. LRN learns frontier-direction affordances (Schmittle et al., 2025); VoroNav uses semantic-topological structure and language-model reasoning (Wu et al., 2024); TANGO combines object-level topological goals with local metric control (Podgorski et al., 2025). World-model approaches such as NWM predict future observations for trajectory evaluation (Bar et al., 2025), while occlusion-aware methods such as Schrödinger's Navigator explicitly imagine plausible hidden worlds (He et al., 2026). GATE targets a different point in this design space: it neither reconstructs nor generates the hidden environment, but estimates only the decision-relevant topological continuation associated with entering it.

Unlike LRN, which learns image-space frontier affordances; VoroNav and TANGO, which exploit semantic or object-level topological representations; and world-model approaches, which predict possible future observations, GATE directly infers a persistent, goal-conditioned topological hypothesis at the boundary of the currently observed metric map.

3 Methodology

GATE addresses long-horizon navigation under partial observability by reasoning about *transitions into unresolved topology*, rather than attempting to reconstruct the unresolved environment itself. As illustrated in Fig. 2, GATE identifies reachable boundaries between verified and unresolved space from a temporally stabilized sensor-built map, infers online the topological continuation supported by each boundary, and commits to the transition most consistent with reaching the distant goal. This inference is performed entirely during execution from current geometric observations, without environment-specific training, a prior global map, or privileged hidden geometry. Importantly, inferred unseen topology influences the long-horizon decision but never certifies physical motion, which remains restricted to currently verified free space.

The methodological object introduced by GATE is therefore not a reconstructed hidden map or merely a frontier score, but a *temporally evolving hypothesis of the latent topological continuation* associated with each reachable observed unknown transition. As the agent executes the selected transition, newly revealed geometry provides evidence that confirms, revises, or rejects this hypothesis.

Robust geometric belief. Before topological inference, GATE stabilizes the geometric belief from which gateways are extracted. The current LiDAR scan is aligned to previously accumulated onboard geometry before fusion, limiting distortion from localization drift. Occupancy is updated using temporal hit–penetration evidence: consistently hit, non-penetrable structure is retained, whereas repeatedly penetrated cells are prevented from persisting as phantom obstacles. These decisions remain reversible as new evidence arrives. Thus, localization error and transient false geometry are filtered before they can alter gateway topology. Map revision affects persistent reasoning only; a current LiDAR terminal return remains a hard obstacle for local execution.

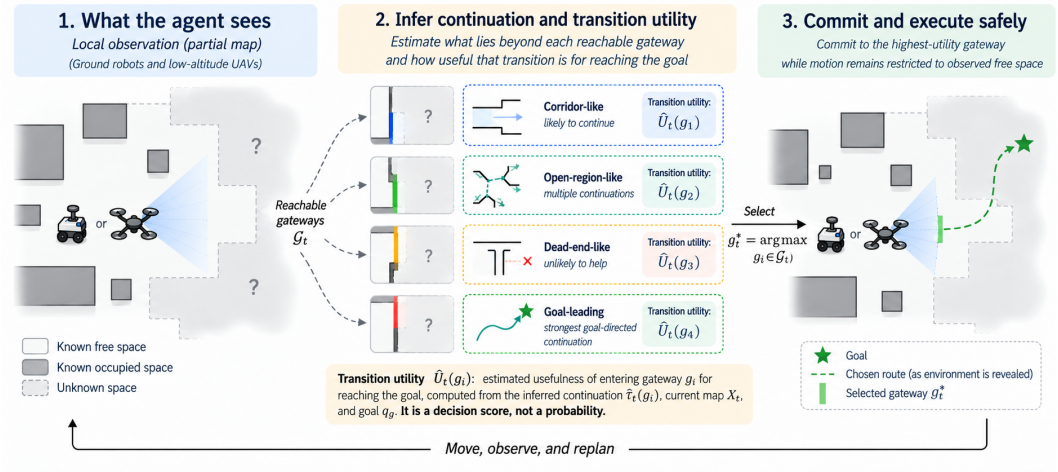


Figure 2: GATE infers latent topological continuation online from partial observations, selects promising gateways into unresolved space, and executes only through verified free space while continually revising its hypothesis.

3.1 Latent Topological Transition Inference

At time t, the agent observes only a partial sensor-built map

$$B_t \in \{-1, 0, 1\}^{H \times W}, \tag{1}$$

where unknown, observed-free, and observed-occupied space are represented by -1, 0, and 1, respectively. The representation extends directly to a three-dimensional occupancy volume for low-altitude UAV navigation.

Let $\mathcal{X}_t$ denote the connected subset of observed free space that is currently reachable under the required clearance constraint. When the goal q_g is not connected through $\mathcal{X}_t$, a complete verified path cannot yet be constructed.

Rather than estimating the complete hidden map $\widehat{\mathcal{M}}_{\text{unknown}}$, GATE identifies a set of reachable boundary transitions

$$\mathcal{G}_t = \{g_1, \ldots, g_N\} \tag{2}$$

and infers, online, the decision-relevant continuation associated with each gateway:

$$\widehat{\tau}_t(g_i) = \mathcal{I}\left(g_i,\, B_t,\, q_g\right), \tag{3}$$

where $\mathcal{I}$ denotes the inference induced by the currently observable geometry.

The latent state $\widehat{\tau}_t(g_i)$ summarizes only the structural information required for navigation—for example, whether the visible geometry supports continued passage, an opening into a larger region, corridor persistence, goal-leading continuation, or termination. It does not predict hidden obstacle locations or reconstruct a complete future map.

Thus,

$$\widehat{\tau}_t(g_i) \ll \widehat{\mathcal{M}}_{\text{unknown}}, \tag{4}$$

in the sense that GATE predicts only the information required to distinguish candidate transitions.

3.2 Gateways as Decision Boundaries

A gateway is a connected, reachable portion of the observed–unknown boundary through which the topology available to the agent may change. GATE constructs gateways directly from the current sensor-built map rather than from a pre-existing topological graph.

For each gateway, the currently visible geometry provides evidence about its possible continuation. This includes the orientation and extent of the unresolved boundary, available clearance, continuation of nearby structure, opening geometry, and its spatial relationship to the distant goal.

The key distinction from conventional frontier selection is that the frontier itself is not the objective. GATE treats it as evidence of a candidate *transition into unresolved topology*.

For decision making, each gateway is described by two complementary forms of evidence. *Verified geometric evidence*, computed entirely from the observed map, captures reachability, clearance, cost to the gateway, goal progress, and alignment. *Continuation evidence*, summarized by $\widehat{\tau}_t(g_i)$, captures whether the visible structure supports a hypothesis of continued passage, corridor persistence, opening into a larger region, goal-directed continuation, or termination. GATE uses these quantities to rank candidate transitions without inserting inferred hidden geometry into the traversable map.

3.3 Transition Utility and Gateway Selection

GATE ranks each reachable gateway by the usefulness of testing its unresolved continuation. For gateway g_i, the transition utility is

$$\widehat{U}_t(g_i) = \mathcal{D}\left(g_i,\, \widehat{\tau}_t(g_i),\, \mathcal{X}_t,\, q_g\right), \tag{5}$$

where $\mathcal{D}$ combines *verified geometric evidence* reachability, clearance, travel cost, progress, and goal alignment with *topological evidence* from $\widehat{\tau}_t(g_i)$, including continued passage, corridor or open-region continuation, goal-leading structure, and termination. Their relative influence adapts to the currently inferred continuation, while experience from previously executed gateways can further refine the utility online.

$\widehat{U}_t(g_i)$ is a relative decision score, not a calibrated probability of reaching the goal. It expresses how strongly the current evidence supports testing one reachable transition over the alternatives. GATE therefore selects

$$g_t^\star = \arg\max_{g_i \in \mathcal{G}_t} \widehat{U}_t(g_i). \tag{6}$$

Because utility reflects continuation rather than Euclidean proximity alone, a lateral or temporarily goal-diverging gateway may be preferred when its observed structure supports a stronger long-horizon route, while a closer goal-facing gateway may be rejected when termination is indicated. As new geometry is observed, $\widehat{\tau}_t(g_i)$ and $\widehat{U}_t(g_i)$ are updated, making gateway selection an online decision over evolving topological evidence.

3.4 Temporal Commitment and Evidence Revision

Selecting the highest-utility gateway at every observation can cause switching as the visible boundary changes. GATE therefore maintains a persistent committed gateway $\bar{g}_t$ across successive observations.

As new geometry is revealed, GATE updates $\widehat{\tau}_t(\bar{g}_t)$ and $\widehat{U}_t(\bar{g}_t)$. The commitment is retained while the gateway remains reachable and supported by the new evidence, and is revised only if it becomes invalid or a better-supported alternative emerges. Thus, utility selects the preferred transition, while temporal commitment preserves decision consistency as evidence evolves.

3.5 Optimistic Inference with Verified Execution

GATE separates the transition it *wants to investigate* from the space in which it is currently *allowed to move*. Define the verified execution space as

$$\mathcal{X}_t = \{c \mid B_t(c) = 0,\ d(c, \mathcal{O}_t) \geq r_b\}, \tag{7}$$

where $B_t(c) = 0$ requires cell c to have been observed as free, $\mathcal{O}_t$ is the currently observed obstacle set, and r_b is the required safety clearance. Equation 7 therefore removes both unknown cells and observed free cells that lie too close to an obstacle; only sensing-verified, clearance-valid space is available for physical execution.

The inferred continuation determines *which gateway to pursue*, but not which motion is executable. After selecting $g_t^\star$, GATE filters the finite action set $\mathcal{A}$ by

$$\mathcal{A}_t^{\text{safe}} = \{a \in \mathcal{A} \mid \text{Sweep}(x_t, a) \subseteq \mathcal{X}_t\}, \tag{8}$$

and executes

$$a_t^\star = \arg\max_{a \in \mathcal{A}_t^{\text{safe}}} J_{\text{exec}}(a; g_t^\star, q_g), \tag{9}$$

where J_{exec} favors progress toward the committed gateway and goal while preserving heading consistency. Thus, $\mathcal{X}_t$ is a feasibility filter rather than a BFS/DFS search space; in 2D, GATE evaluates 36 candidate headings and rejects any whose swept segment enters unknown or clearance-invalid space. As sensing reveals new free space, $\mathcal{X}_t$ expands, and the process repeats.

4 Experiments

4.1 Setup and Environments

We evaluate GATE under partial observability using a common $20\,\text{m}$ LiDAR range in both 2D ground navigation and 3D low-altitude UAV navigation. All online planners share the same start–goal task, sensing, clearance, collision evaluation, perception frontend, and corruption condition; the global map is never available during execution. Full-map A* is used only as an oracle for path-quality evaluation. The action space follows of (Shrivastava et al., 2025), while 3D occupancy uses the height-aware representation of (Shrivastava et al., 2026; Hornung et al., 2013).

Controlled maze. Before testing on real-life locations, GATE and all baselines (Section 4.2) are evaluated in the maze(Appendix A.1) shown in Fig. 3. It contains narrow passages, dead ends, open regions, and occluded branches whose useful continuation lies beyond the sensing horizon. We test an occluded-goal task, $(2.0, 2.5) \to (20, 19.5)$, and a long-horizon task, $(2, 2.5) \to (38, 2.5)$.

Extracted real-life locations. We evaluate on 20 real-life urban environments extracted from Google Maps Platform 3D Tiles (Google Maps Platform, 2026) and imported into Gazebo (Koenig and Howard, 2004): Athens, Frankfurt am Main, Genoa–Castelletto, Gobe Downtown, Hong Kong, Johannesburg CBD, Lisbon–Alfama, Macau, Milan–Porta Nuova, Tokyo, Rio de Janeiro–Botafogo, Salzburg–Fortress Hill, São Paulo–República, Seattle Downtown, Boston Downtown, Stuttgart–Killesberg, Toronto Financial District, Warsaw–Wola, Wellington–Highbury, and Downtown Chicago.

The environments span dense high-rise cores, historic narrow-street layouts, terrain-dominated districts, and mixed/open urban structure, varying in obstacle density, street width, elevation, visibility, and route topology. Start–goal tasks require approximately 60–140 m of full-map route length, corresponding to 3–7$\times$ the common 20 m LiDAR range. In every task, the goal lies outside the initial observation and at least one route-critical continuation is geometrically occluded, preventing the useful long-horizon route from being validated from the initial sensor view. The same geometry is used for 2D and 3D evaluation, with environment representations provided in Appendix A.1.

4.2 Baselines

We compare GATE against local, sampling-based, belief-space, exploration, and learned planners: **A*-Partial** Hart et al. (1968), **DWA-Reactive** Fox et al. (1997), **POMCP-Style** Silver and Veness (2010), **ActiveSLAM-Frontier** Yamauchi (1997); Placed et al. (2023), **RRT*** Karaman and Frazzoli (2011), **ARiADNE** Cao et al. (2023), **DARE** Cao et al. (2025), **CogniPlan** Wang et al. (2025), **NWM** Bar et al. (2025), and **DreamerNav** Shanks et al. (2025). For 3D UAV evaluation, we additionally include adapted **TANGO** Podgorski et al. (2025), **LRN** Schmittle et al. (2025), and **VoroNav** Wu et al. (2024). All planners are evaluated on the same 20 real-life environments using the same start–goal tasks, 20 m LiDAR range, clearance and collision constraints, observation corruptions, and common 36-action motion interface used by GATE in both 2D and 3D. For adapted methods, the original high-level decision mechanism is retained while actions are mapped to this shared execution interface,

ensuring that differences arise from planning rather than sensing or motion privileges. Full-map A* is used only as an evaluation oracle and is never available during online execution. Full adaptation details are provided in Appendix A.5.

4.3 Evaluation Metrics and Robustness Protocol

In Tables 2 and 3, we report success rate, path change relative to Clean,

$$\Delta L^{\text{Clean}}(\%) = 100\frac{L_p - L_{\text{Clean}}}{L_{\text{Clean}}}, \tag{10}$$

and excess path length relative to the full-map A* oracle,

$$\Delta L^{\text{A}^*}(\%) = 100\frac{L_p - L^*}{L^*}. \tag{11}$$

Percent changes are computed per successful start–goal run and aggregated as mean $\pm$ SD, preventing longer environments from dominating the aggregate. Failed runs contribute only to success rate because incomplete trajectories do not provide valid start–goal path lengths.

Adversarial observation stress. Each task is evaluated under four profiles: Clean, Mild, Moderate, and Severe, with 30 independently seeded runs per profile (120 runs per planner-task pair). The non-clean profiles jointly perturb LiDAR occupancy and localization to reproduce progressively stronger real-world-inspired perception failures. Adverse conditions such as fog, rain, and snow can attenuate valid LiDAR returns and introduce transient or misleading returns (Hahner et al., 2021; Linnhoff et al., 2022; Kim et al., 2023). We abstract these effects as *phantom occupied regions*: at each scan, a phantom may appear according to the profile-specific spawn probability, with its spatial extent sampled from the corresponding radius range and the number of simultaneously active phantoms bounded as shown in Table 1.

Table 1: Joint observation-stress profiles. Each profile uses 30 independently seeded runs.

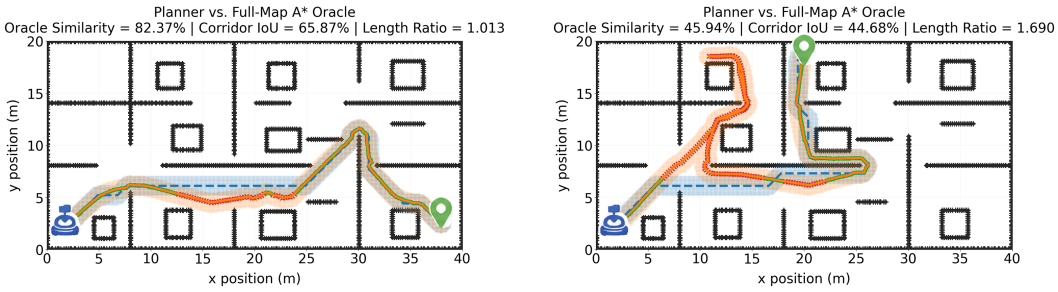
Condition	Spawn /scan	Active max.	Radius (m)	σ_{xy} (m)	Max drift (m)	σ_θ (deg)
Clean	0	0	–	0	0	0
Mild	.05	10	.40–.70	.05	.10	.5
Moderate	.15	20	.50–1.00	.10	.25	1.5
Severe	.40	30	.70–1.40	.20	.50	3.0

Phantoms persist for 10–20, 20–40, and 30–60 scans under Mild, Moderate, and Severe stress, with false-return probabilities of 0.25, 0.45, and 0.65. Thus, false structure can appear, persist, and disappear over time. Localization is simultaneously perturbed by translational noise, bounded drift, and heading error, reflecting degraded mapping and pose estimation under imperfect sensing (Burnett et al., 2023). These are controlled, real-world-inspired stress conditions. Robustness is measured by success rate and path-length change relative to Clean and full-map A*.

5 Experimental Results

Each planner is evaluated over 30 trials per corruption profile using both occluded and long-horizon start–goal configurations. With the four profiles in Table 1. Corruptions are independently seeded, and full-map A* is used only as an evaluation oracle.

Controlled maze. GATE achieved $30/30$ success on both long-horizon and occluded-goal tasks, while no non-GATE baseline reached the goal. In Fig. 3, GATE remained near A* in the long-horizon case ($L/L_{\text{A}^*} = 1.013$) and still succeeded under stronger occlusion (1.690), where failures in other methods were dominated by switching, oscillation, dead-end selection, and stagnation.



(a) Long-horizon.

(b) Occluded-goal.

Figure 3: Controlled-maze trajectories: GATE (orange) and full-map A* (blue).

Real-life 2D navigation. Across 20 real-life maps, GATE achieved 100% success under all four stress profiles in Table 1, with corruption mainly increasing path length rather than causing failure (Table 2). All 10 non-GATE baselines failed to complete a start–goal trajectory. Per-environment results are in Appendix A.3, with baseline analysis in Appendix A.5.1.

Table 2: 2D robustness across 20 long-horizon, occluded real-life environments with 30 seeded runs per profile. ΔL_{Clean} is the percentage path increase relative to Clean, and Excess vs. A* is the percentage increase over the corresponding full-map A* oracle. Values are macro averaged mean $\pm$ SD over successful runs per environment; no 2D baseline completed the task.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.94 ± 1.68
	Mild	100%	7.02 ± 2.07	9.13 ± 3.94
	Moderate	100%	18.84 ± 10.29	21.16 ± 11.03
	Severe	100%	28.67 ± 12.63	31.24 ± 13.92
All 10 2D baselines	All profiles	0%	*No successful runs; stagnation/oscillation*	

Real-life 3D Navigation. Across 20 real-life 3D environments, GATE maintains 100% success from Clean through Severe corruption (Table 3), with path inflation increasing from 6.99% under Mild to 28.32% under Severe. In contrast, the aggregate success of the 13 non-GATE baselines drops from 68.3% to 29.8% with substantially larger degradation. Per-environment GATE results and 3D trajectory/elevation profiles are reported in Appendix A.4, while baseline-specific results are provided in Appendix A.5.2, supporting the dimensional generality of GATE's gateway-level execution.

Table 3: 3D low-altitude UAV robustness across 20 real-life environments. ΔL_{Clean} measures path increase relative to Clean, and Excess vs. A* relative to the full-map A* oracle. Success is aggregated over all trials; path metrics are macro-averaged across environments, with $\pm$ denoting SD across environment-level means.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.91 ± 1.64
	Mild	100%	6.99 ± 2.02	9.07 ± 3.85
	Moderate	100%	18.65 ± 10.05	20.94 ± 10.79
	Severe	100%	28.32 ± 12.38	30.85 ± 13.66
All 13 3D baselines	Clean	68.3%	–	17.3 ± 7.40
	Mild	34.5%	9.0 ± 6.00	27.4 ± 10.79
	Moderate	32.4%	21.0 ± 11.55	42.0 ± 16.40
	Severe	29.8%	39.4 ± 20.29	63.5 ± 27.19

6 Ablation, Complexity, and Compute Requirements

Ablation. Each variant removes one GATE mechanism while retaining identical sensing, mapping, action, and remaining planner components. The 30 frozen start–goal pairs are documented in Appendix A.2.

Table 4: Controlled-maze ablation over 30 frozen start–goal pairs (Appendix A.2). All pairs contain an occluded direct route and 22/30 are long-horizon. L/L_{A^*} is averaged over successful pairs; switches are mean $\pm$ SD. Failures are separated into stagnation/oscillation (S/O) and collision.

Variant	Success	Mean L/L_{A^*}	Switches/run	S/O fails	Collisions
Full GATE	$\mathbf{30/30}$	$\mathbf{1.013}$	$\mathbf{2.1 \pm 0.8}$	$\mathbf{0}$	$\mathbf{0}$
w/o gateway abstraction	$25/30$	1.180	6.8 ± 2.1	5	0
w/o latent continuation	$23/30$	1.250	4.9 ± 1.8	7	0
w/o temporal commitment	$24/30$	1.150	11.7 ± 3.4	6	0
w/o certified execution	$27/30$	1.090	3.1 ± 1.2	2	1

Full GATE solves all 30 pairs without collision. Removing latent continuation causes the most failures (7), while removing temporal commitment increases switching to 11.7 ± 3.4 per run. Removing gateway abstraction causes five stagnation/oscillation failures, and removing certified execution causes two such failures and one collision. Pair details are provided in Appendix A.2.

Compute requirements. GATE is training-free and GPU-free. Its per-step complexity is $\mathcal{O}(RL + N + F + CKL_g + AQ)$, where R is the number of LiDAR rays, L the cells traversed per ray, and N, F, C the active-map, frontier, and gateway sizes. LiDAR processing therefore scales linearly with ray count and range. With fixed sensing and action settings, this reduces to $\mathcal{O}(N + F + C)$. Estimated runtime is 0.70–1.50 s/update (0.67–1.43 Hz), without global graph search, belief-tree expansion, or neural world-model inference.

6.1 Discussion

GATE bridges reactive planning and long-horizon reasoning by selecting goal-directed topological transitions without reconstructing unseen geometry. Ablations show that latent-continuation reasoning supports occluded-route selection, temporal commitment limits switching, and certified execution preserves safety. Across 2D and 3D environments, corruption primarily increases travel rather than preventing goal completion (Appendices A.3 and A.4). A current limitation is the absence of real-robot validation. GATE is also sensitive to sensing horizon: shorter-range LiDAR reveals route-relevant structure more gradually, requiring additional gateway updates and execution steps before occluded continuations become observable (Appendix A.6). Future work will evaluate real platforms, dynamic agents, and sustained drift. Potential applications include ground and aerial navigation for urban inspection, search and rescue, autonomous delivery, and infrastructure monitoring.

7 Conclusion

We introduced GATE, a training-free framework for long-horizon navigation under partial observability. GATE reasons over latent topological continuations, maintains persistent gateway commitments, and restricts execution to sensing-verified free space without reconstructing hidden geometry. Across controlled maze and 20 real-life 2D and 3D environments, GATE remained robust under increasing observation corruption. These results show that decision-relevant topology can support coherent long-horizon navigation without training, privileged maps, or explicit prediction of the unseen environment.

AI use statement

Generative AI tools were used to assist with the generation of Figures 1 and 2 and for limited language editing and LaTeX formatting. They were not used to generate the research hypothesis, methodological

contribution, experimental design, experimental results, or scientific conclusions. All AI-assisted text and visual material was reviewed and revised by the authors, and all reported results were obtained from the authors' implementations and experiments. The authors take full responsibility for the final content of the paper.

Reproducibility statement

The paper specifies GATE's algorithmic formulation, planner parameters, sensing model, safety constraints, baseline configurations, corruption profiles, evaluation metrics, and repeated-trial protocol. Controlled-maze and ablation details are provided in the main text, with environment representations and additional experimental details in the appendix. All robustness profiles and randomized trials are explicitly defined to support independent reproduction. Source code is not released with this submission.

Author Contributions

If you'd like to, you may include a section for author contributions as is done in many journals. This is optional and at the discretion of the authors.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper.

References

Hart, Peter E., Nilsson, Nils J., and Raphael, Bertram. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968. doi: 10.1109/TSSC.1968.300136.

LaValle, Steven M. Rapidly-exploring random trees: A new tool for path planning. Technical Report TR 98-11, Iowa State University, 1998.

Fox, Dieter, Burgard, Wolfram, and Thrun, Sebastian. The dynamic window approach to collision avoidance. *IEEE Robotics & Automation Magazine*, 4(1):23–33, 1997. doi: 10.1109/100.580977.

Yamauchi, Brian. A frontier-based approach for autonomous exploration. In *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation*, pages 146–151, 1997.

Zhou, Boyu, Zhang, Yichen, Chen, Xinyi, and Shen, Shaojie. FUEL: Fast UAV exploration using incremental frontier structure and hierarchical planning. *IEEE Robotics and Automation Letters*, 6(2):779–786, 2021. doi: 10.1109/LRA.2021.3051563.

Silver, David and Veness, Joel. Monte-Carlo planning in large POMDPs. In *Advances in Neural Information Processing Systems*, volume 23, 2010.

Placed, Julio A., Strader, Jared, Carrillo, Henry, Atanasov, Nikolay, Indelman, Vadim, Carlone, Luca, and Castellanos, José A. A survey on active simultaneous localization and mapping: State of the art and new frontiers. *IEEE Transactions on Robotics*, 2023. Also available as arXiv:2207.00254.

Chaplot, Devendra Singh, Gandhi, Dhiraj, Gupta, Saurabh, Gupta, Abhinav, and Salakhutdinov, Ruslan. Learning to explore using active neural SLAM. In *International Conference on Learning Representations*, 2020.

Hornung, Armin, Wurm, Kai M., Bennewitz, Maren, Stachniss, Cyrill, and Burgard, Wolfram. OctoMap: An efficient probabilistic 3D mapping framework based on octrees. *Autonomous Robots*, 34(3):189–206, 2013. doi: 10.1007/s10514-012-9321-0.

Hafner, Danijar, Pasukonis, Jurgis, Ba, Jimmy, and Lillicrap, Timothy. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*, 2023.

Bar, Amir, Zhou, Gaoyue, Tran, Danny, Darrell, Trevor, and LeCun, Yann. Navigation world models. *arXiv preprint arXiv:2412.03572*, 2024.

Shanks S, Embley-Riches J, Liu J, Delfaki AM, Ciliberto C and Kanoulas D. DreamerNav: Learning-based autonomous navigation in dynamic indoor environments using world models. *Frontiers in Robotics and AI*, 12, 2025. doi: 10.3389/frobt.2025.1655171.

Kurniawati, Hanna. Partially observable Markov decision processes and robotics. *Annual Review of Control, Robotics, and Autonomous Systems*, 5:253–277, 2022. doi: 10.1146/annurev-control-042920-092451.

Koenig, Sven and Likhachev, Maxim. D* Lite. In *Proceedings of the 18th National Conference on Artificial Intelligence*, pages 476–483, 2002.

Karaman, Sertac and Frazzoli, Emilio. Sampling-based algorithms for optimal motion planning. *The International Journal of Robotics Research*, 30(7):846–894, 2011. doi: 10.1177/0278364911406761.

Google Maps Platform. Photorealistic 3D Tiles Overview. `https://developers.google.com/maps/documentation/tile/3d-tiles-overview`, 2026. Accessed: 2026-05-06.

N. Koenig and A. Howard. Design and use paradigms for Gazebo, an open-source multi-robot simulator. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2149–2154, 2004. doi: `10.1109/IROS.2004.1389727`.

Shrivastava, A., Petridis, C., Vacic, M., and Obradovic, Z. Autonomous navigation in swarm of UAVs using spatio temporal data and constrained-reinforcement learning. In Iliadis, L., Maglogiannis, I., Kyriacou, E., and Jayne, C., editors, *Engineering Applications of Neural Networks. EANN 2025*, Communications in Computer and Information Science, volume 2582. Springer, Cham, 2025. doi: 10.1007/978-3-031-96199-1_5.

Shrivastava, Abhudaya, Gupta, Shelly, and Obradovic, Zoran. Real-time path planning for UAVs in windy environments without computational fluid dynamics. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(22):18540–18548, 2026. doi: 10.1609/aaai.v40i22.38920.

He, Yu, Huang, Da, Liu, Zhenyang, Gu, Zixiao, Sun, Qiang, Ye, Guangnan, Fu, Yanwei, and Jiang, Yu-Gang. Schrödinger's Navigator: Imagining an Ensemble of Futures for Zero-Shot Object Navigation. *arXiv preprint arXiv:2512.21201*, 2026. doi: 10.48550/arXiv.2512.21201.

Bar, Amir, Zhou, Gaoyue, Tran, Danny, Darrell, Trevor, and LeCun, Yann. Navigation World Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15791–15801, 2025. doi: 10.1109/CVPR52734.2025.01472.

Podgorski, Stefan, Garg, Sourav, Hosseinzadeh, Mehdi, Mares, Lachlan, Dayoub, Feras, and Reid, Ian. TANGO: Traversability-Aware Navigation with Local Metric Control for Topological Goals. *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2399–2406, 2025. doi: 10.1109/ICRA55743.2025.11127998.

Wu, Pengying, Mu, Yao, Wu, Bingxian, Hou, Yi, Ma, Ji, Zhang, Shanghang, and Liu, Chang. VoroNav: Voronoi-Based Zero-Shot Object Navigation with Large Language Model. *Proceedings of the 41st International Conference on Machine Learning*, 235:53757–53775, 2024.

Schmittle, Matt, Baijal, Rohan, Hatch, Nathan, Scalise, Rosario, Guaman Castro, Mateo, Talia, Sidharth, Khetarpal, Khimya, Boots, Byron, and Srinivasa, Siddhartha. Long Range Navigator (LRN): Extending robot planning horizons beyond metric maps. *Proceedings of the 9th Conference on Robot Learning*, 305:3185–3199, 2025.

Cao, Yuhong, Hou, Tianxiang, Wang, Yizhuo, Yi, Xian, and Sartoretti, Guillaume. ARiADNE: A reinforcement learning approach using attention-based deep networks for exploration. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10219–10225, 2023.

Hahner, Martin, Sakaridis, Christos, Dai, Dengxin, and Van Gool, Luc. Fog Simulation on Real LiDAR Point Clouds for 3D Object Detection in Adverse Weather. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15283–15292, 2021. doi: 10.1109/ICCV48922.2021.01500.

Linnhoff, Clemens, Hofrichter, Kristof, Elster, Lukas, Rosenberger, Philipp, and Winner, Hermann. Measuring the Influence of Environmental Conditions on Automotive Lidar Sensors. *Sensors*, 22(14):5266, 2022. doi: 10.3390/s22145266.

Kim, Jiyoon, Park, Bum-jin, and Kim, Jisoo. Empirical Analysis of Autonomous Vehicle's LiDAR Detection Performance Degradation for Actual Road Driving in Rain and Fog. *Sensors*, 23(6):2972, 2023. doi: 10.3390/s23062972.

Burnett, Keenan, Yoon, David J., Wu, Yuchen, Li, Andrew Z., Zhang, Haowei, Lu, Shichen, Qian, Jingxing, Tseng, Wei-Kang, Lambert, Andrew, Leung, Keith Y. K., Schoellig, Angela P., and Barfoot, Timothy D. Boreas: A Multi-Season Autonomous Driving Dataset. *The International Journal of Robotics Research*, 42(1–2):33–42, 2023. doi: 10.1177/02783649231160195.

Wang, Yizhuo, He, Haodong, Liang, Jingsong, Cao, Yuhong, Chakraborty, Ritabrata, and Sartoretti, Guillaume Adrien. CogniPlan: Uncertainty-guided path planning with conditional generative layout prediction. In *Proceedings of the 9th Conference on Robot Learning*, volume 305 of *Proceedings of Machine Learning Research*, pages 1382–1396, 2025.

Cao, Yuhong, Lew, Jeric, Liang, Jingsong, Cheng, Jin, and Sartoretti, Guillaume. DARE: Diffusion policy for autonomous robot exploration. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11987–11993, 2025. doi: 10.1109/ICRA55743.2025.11128196.

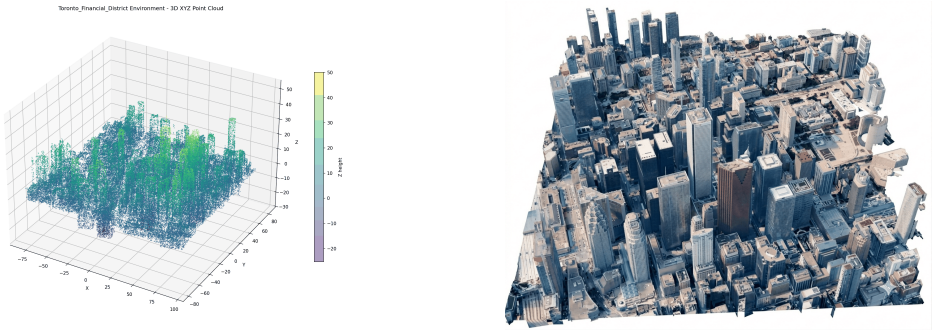
A Appendix

This appendix provides supplementary details supporting the experimental evaluation. It first presents the real-life environments through their extracted 3D point-cloud representations and corresponding photorealistic simulation geometry. It then documents the controlled-maze setup and frozen start–goal pairs used for ablation, followed by per-environment 2D and 3D robustness results, trajectory analyses, and baseline-specific comparisons.

A.1 Real-Life Environment Visualizations

This appendix visualizes the reconstructed environments used in the real-life navigation experiments. For each location, we show the extracted 3D point cloud, the corresponding top-down maximum-height representation, and the photorealistic environment geometry imported for simulation. The point-cloud and height-map axes are expressed in the local environment coordinate frame.

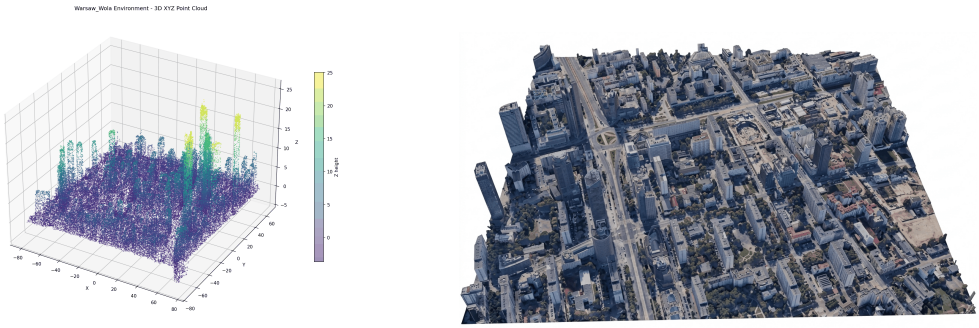
Toronto Financial District



(a) 3D point cloud (b) Environment geometry

Figure 4: Environment representation for Toronto Financial District.

Warsaw–Wola



(a) 3D point cloud (b) Environment geometry

Figure 5: Environment representation for Warsaw–Wola.

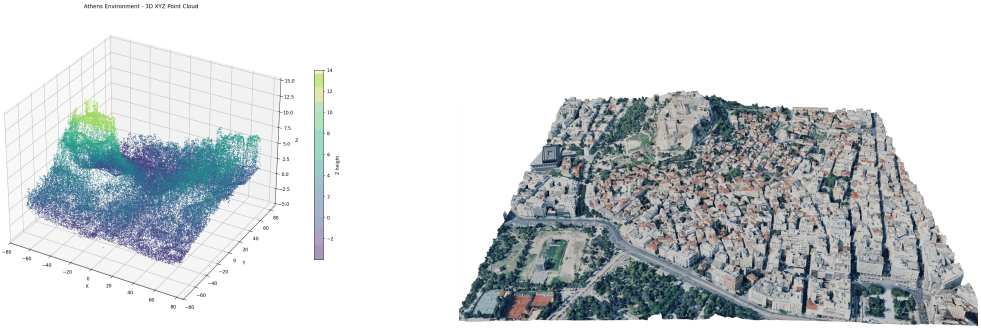
Wellington–Highbury



(a) 3D point cloud (b) Environment geometry

Figure 6: Environment representation for Wellington–Highbury.

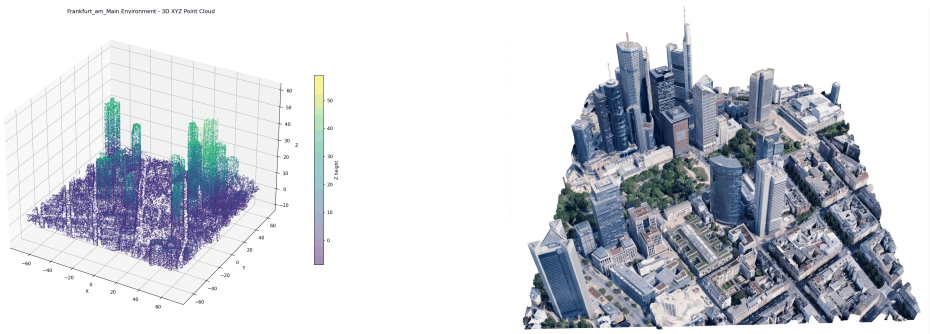
Athens



(a) 3D point cloud (b) Environment geometry

Figure 7: Environment representation for Athens.

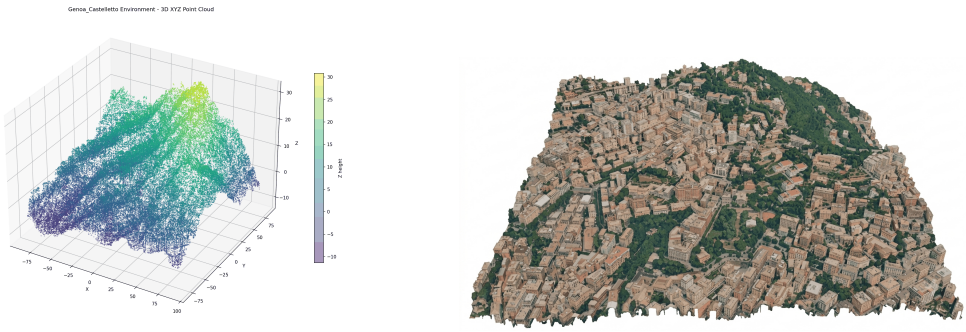
Frankfurt am Main



(a) 3D point cloud (b) Environment geometry

Figure 8: Environment representation for Frankfurt am Main.

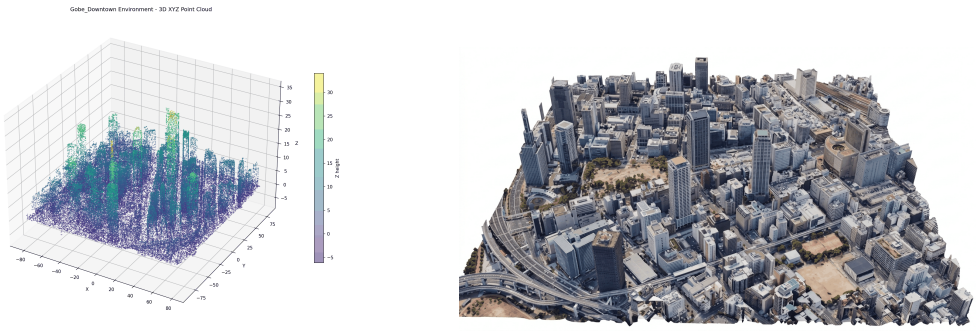
Genoa–Castelletto



(a) 3D point cloud (b) Environment geometry

Figure 9: Environment representation for Genoa–Castelletto.

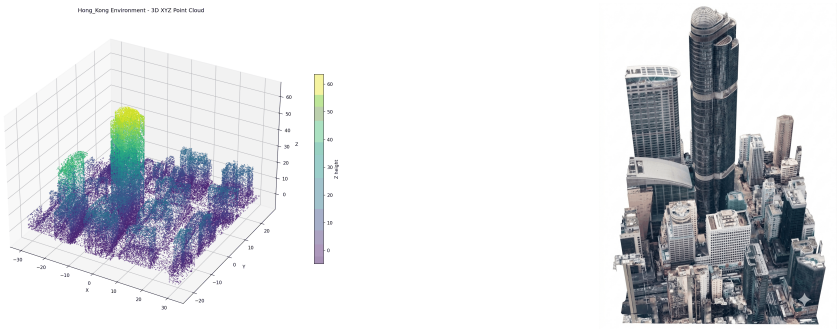
Gobe Downtown



(a) 3D point cloud (b) Environment geometry

Figure 10: Environment representation for Gobe Downtown.

Hong Kong



(a) 3D point cloud (b) Environment geometry

Figure 11: Environment representation for Hong Kong.

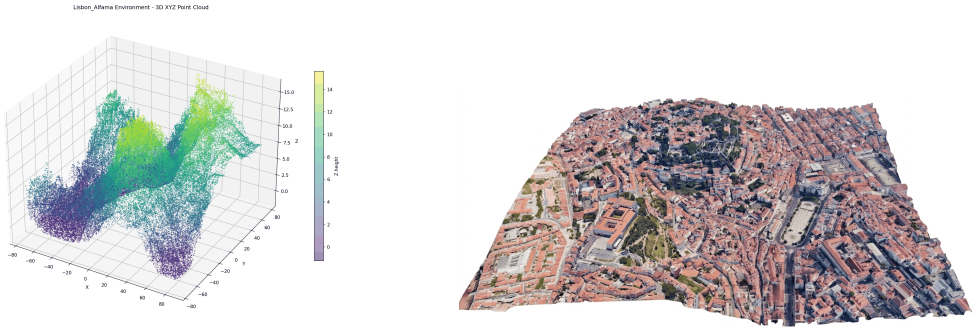
Johannesburg CBD



(a) 3D point cloud (b) Environment geometry

Figure 12: Environment representation for Johannesburg CBD.

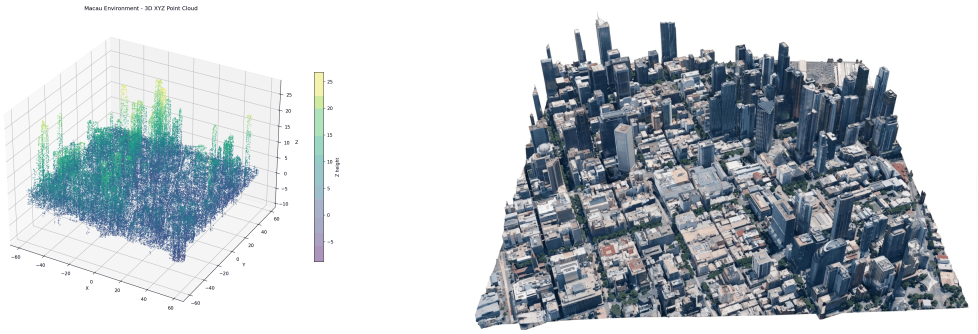
Lisbon–Alfama



(a) 3D point cloud (b) Environment geometry

Figure 13: Environment representation for Lisbon–Alfama.

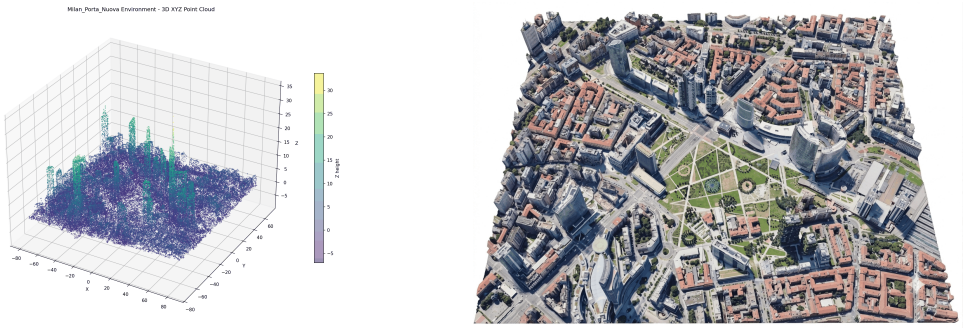
Macau



(a) 3D point cloud (b) Environment geometry

Figure 14: Environment representation for Macau.

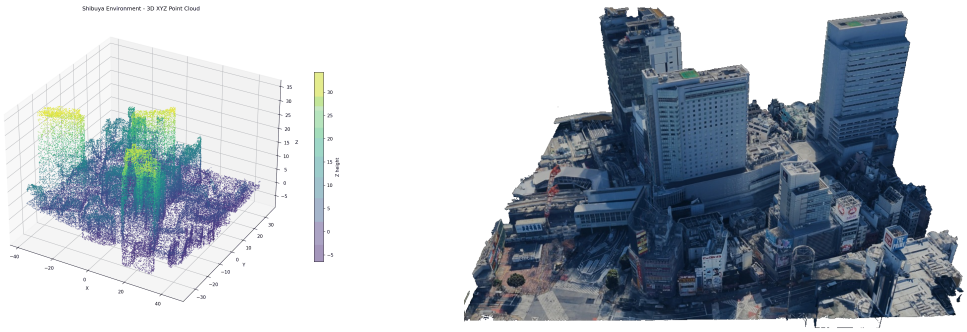
Milan–Porta Nuova



(a) 3D point cloud (b) Environment geometry

Figure 15: Environment representation for Milan–Porta Nuova.

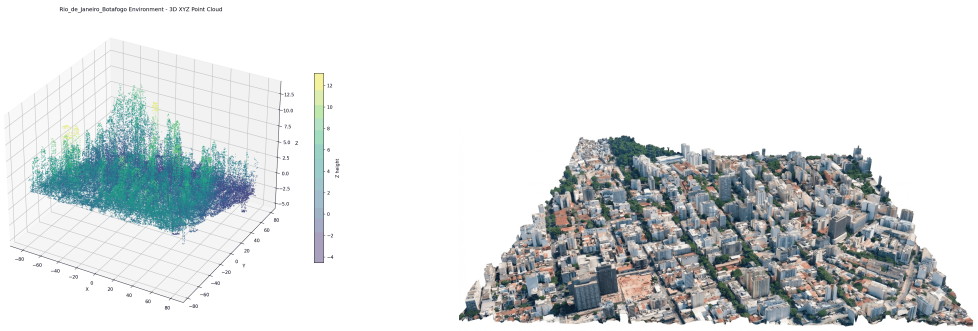
Tokyo Shibuya



(a) 3D point cloud (b) Environment geometry

Figure 16: Environment representation for Tokyo Shibuya.

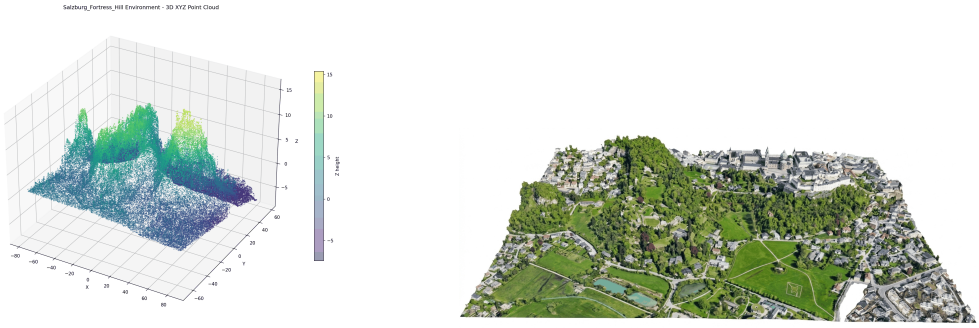
Rio de Janeiro–Botafogo



(a) 3D point cloud (b) Environment geometry

Figure 17: Environment representation for Rio de Janeiro–Botafogo.

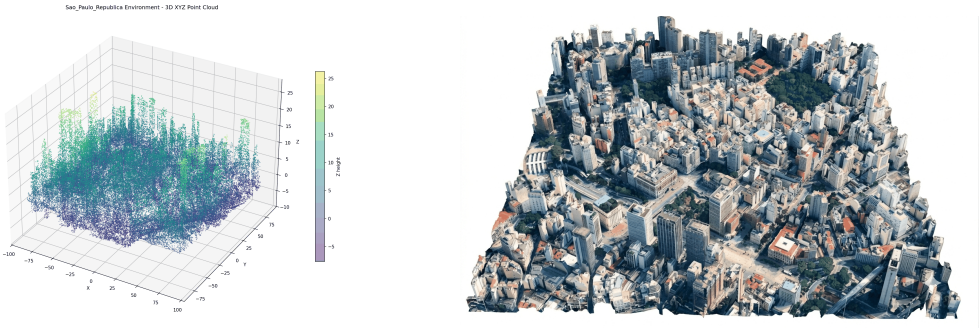
Salzburg–Fortress Hill



(a) 3D point cloud (b) Environment geometry

Figure 18: Environment representation for Salzburg–Fortress Hill.

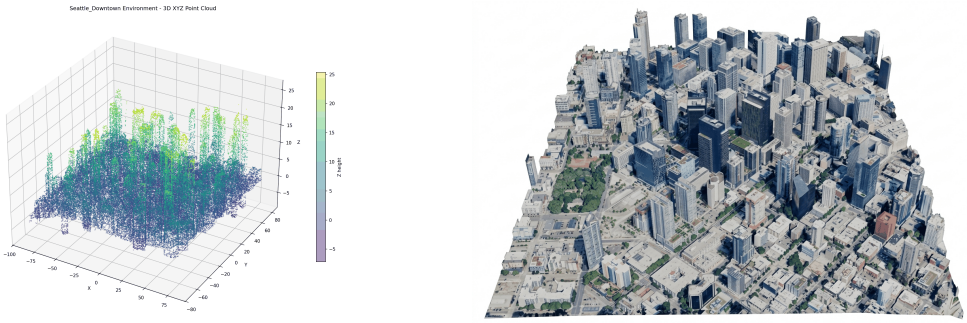
São Paulo–República



(a) 3D point cloud (b) Environment geometry

Figure 19: Environment representation for São Paulo–República.

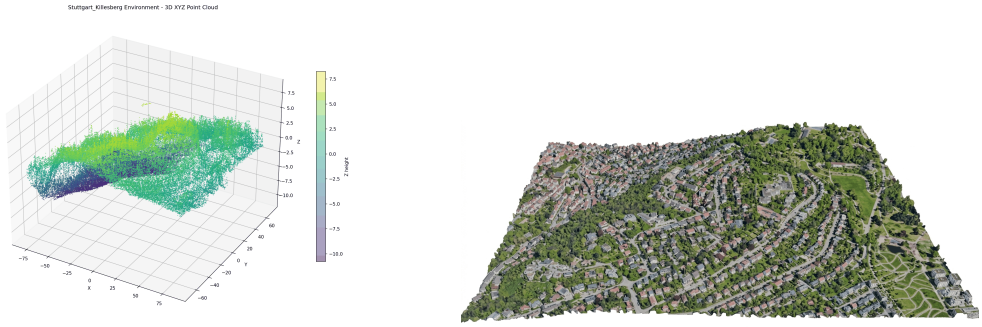
Seattle Downtown



(a) 3D point cloud (b) Environment geometry

Figure 20: Environment representation for Seattle Downtown.

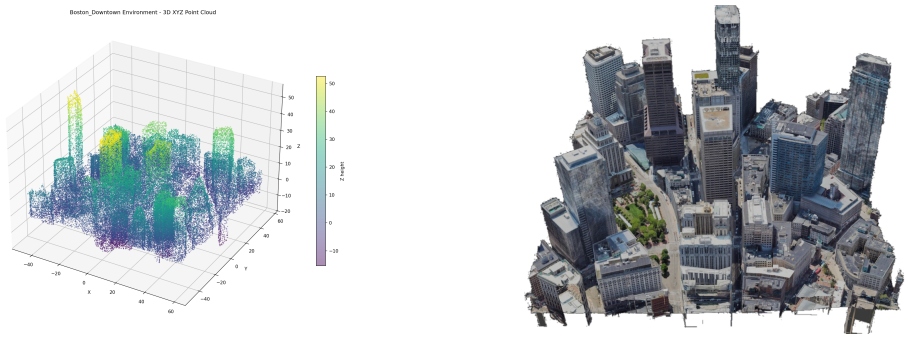
Stuttgart–Killesberg



(a) 3D point cloud (b) Environment geometry

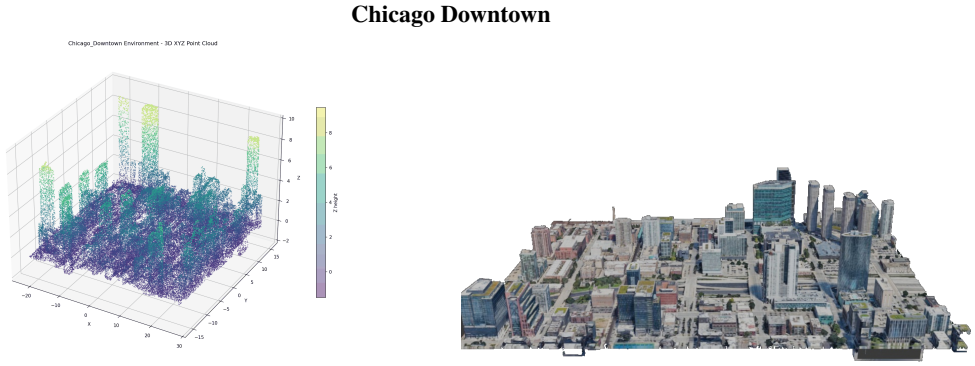
Figure 21: Environment representation for Stuttgart–Killesberg.

Boston Downtown



(a) 3D point cloud (b) Environment geometry

Figure 22: Environment representation for Boston Downtown.



(a) 3D point cloud (b) Environment geometry

Figure 23: Environment representation for Chicago Downtown.

A.2 Controlled-Maze Start–Goal Pairs

Figure 24 and Table 5 define the 30 frozen start–goal pairs used for the controlled-maze ablation. The same pair set is used across all GATE variants to ensure a direct mechanism-level comparison. The pairs cover different maze regions and route orientations under partial observability. All 30 have an occluded direct start–goal segment, while 22/30 are also long-horizon with $d_{SG} > 20$ m. This requires both local route disambiguation and reasoning beyond the current LiDAR observation. Table 5 reports the pair coordinates, d_{SG}, and O/LH classification. Full-map A* is used only to verify reachability and provide the oracle reference; GATE executes only from its incrementally observed map.

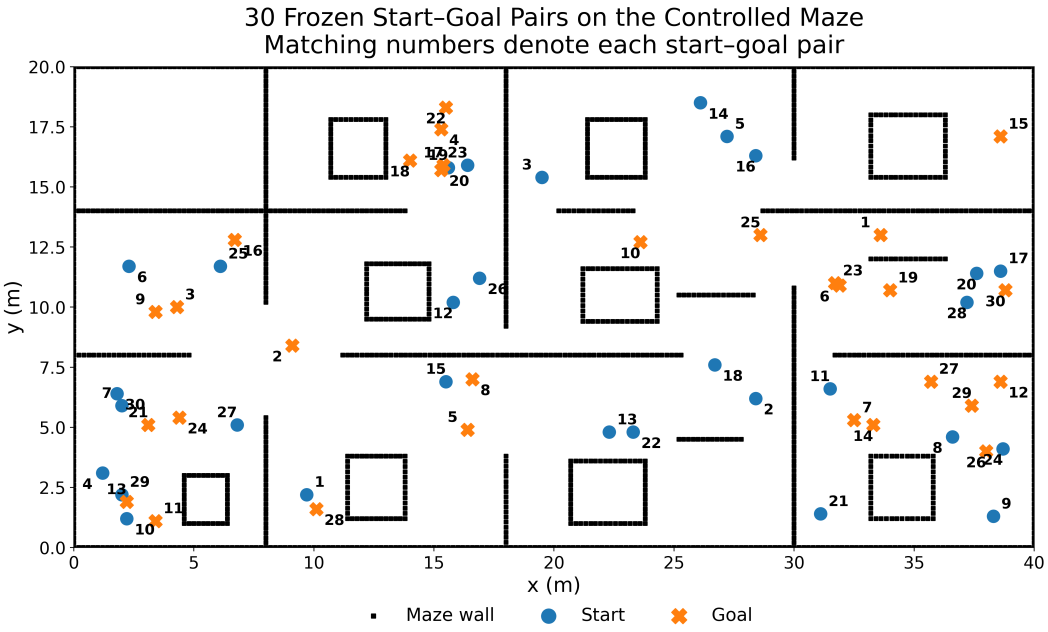


Figure 24: Frozen controlled-maze start–goal pairs used for paired ablation. Matching indices denote corresponding start and goal locations.

Table 5: Frozen controlled-maze ablation pairs. d_{SG} is the straight-line start–goal separation. O denotes an occluded direct segment; LH denotes $d_{SG} > 20$ m, i.e., beyond one nominal LiDAR sensing horizon.

Pair	Start (x, y)	Goal (x, y)	d_{SG} (m)	Type	Pair	Start (x, y)	Goal (x, y)	d_{SG} (m)	Type
1	$(9.7, 2.2)$	$(33.6, 13.0)$	26.2	O+LH	16	$(28.4, 16.3)$	$(6.7, 12.8)$	22.0	O+LH
2	$(28.4, 6.2)$	$(9.1, 8.4)$	19.4	O	17	$(38.6, 11.5)$	$(15.4, 15.9)$	23.6	O+LH
3	$(19.5, 15.4)$	$(4.3, 10.0)$	16.1	O	18	$(26.7, 7.6)$	$(14.0, 16.1)$	15.3	O
4	$(1.2, 3.1)$	$(15.3, 17.4)$	20.1	O+LH	19	$(15.6, 15.8)$	$(34.0, 10.7)$	19.1	O
5	$(27.2, 17.1)$	$(16.4, 4.9)$	16.3	O	20	$(37.6, 11.4)$	$(15.3, 15.7)$	22.7	O+LH
6	$(2.3, 11.7)$	$(31.9, 10.9)$	29.6	O+LH	21	$(31.1, 1.4)$	$(3.1, 5.1)$	28.2	O+LH
7	$(2.0, 5.9)$	$(32.5, 5.3)$	30.5	O+LH	22	$(23.3, 4.8)$	$(15.5, 18.3)$	15.6	O
8	$(36.6, 4.6)$	$(16.6, 7.0)$	20.1	O+LH	23	$(16.4, 15.9)$	$(31.7, 11.0)$	16.1	O
9	$(38.3, 1.3)$	$(3.4, 9.8)$	35.9	O+LH	24	$(38.7, 4.1)$	$(4.4, 5.4)$	34.3	O+LH
10	$(2.2, 1.2)$	$(23.6, 12.7)$	24.3	O+LH	25	$(6.1, 11.7)$	$(28.6, 13.0)$	22.5	O+LH
11	$(31.5, 6.6)$	$(3.4, 1.1)$	28.6	O+LH	26	$(16.9, 11.2)$	$(38.0, 4.0)$	22.3	O+LH
12	$(15.8, 10.2)$	$(38.6, 6.9)$	23.0	O+LH	27	$(6.8, 5.1)$	$(35.7, 6.9)$	29.0	O+LH
13	$(22.3, 4.8)$	$(2.2, 1.9)$	20.3	O+LH	28	$(37.2, 10.2)$	$(10.1, 1.6)$	28.4	O+LH
14	$(26.1, 18.5)$	$(33.3, 5.1)$	15.2	O	29	$(2.0, 2.2)$	$(37.4, 5.9)$	35.6	O+LH
15	$(15.5, 6.9)$	$(38.6, 17.1)$	25.3	O+LH	30	$(1.8, 6.4)$	$(38.8, 10.7)$	37.2	O+LH

A.3 Real-Life 2D Navigation Under Partial Observability Across Unseen Environments

Evaluation protocol. We evaluate GATE across reconstructed 2D urban environments using randomly sampled start–goal pairs retained only if they satisfy the occluded-goal or long-horizon criterion. Each pair is first verified as feasible using clearance-constrained full-map A*. A* is used only to validate reachability and provide an oracle path-length reference; GATE receives only locally observed geometry during execution. The complete trajectories are provided in the supplementary material, with per-environment results reported in Tables 6–25.

Each environment is evaluated over 30 independent trials under Clean, Mild, Moderate, and Severe corruption. We report success, path increase relative to Clean,

$$\Delta L_{\text{Clean}} = 100 \frac{L_{\text{profile}} - L_{\text{Clean}}}{L_{\text{Clean}}}, \tag{12}$$

and excess over the full-map A* oracle,

$$E_{A^*} = 100 \frac{L_{\text{profile}} - L_{A^*}}{L_{A^*}}. \tag{13}$$

Values are mean $\pm$ SD across 30 trials. GATE reaches the goal in every reported trial; stronger corruption therefore appears mainly as additional travel, stagnation/oscillation, or selection of a longer clearance-valid route.

Stable topological continuation. In Frankfurt, Chicago, Johannesburg, Warsaw, Athens, Genoa, Lisbon, Macau, Milan, Tokyo, and Rio de Janeiro, the dominant gateway sequence remains largely stable as corruption increases. Path inflation is therefore driven mainly by local correction, temporary stagnation, and oscillatory recovery rather than a change in the global continuation. For example, Frankfurt remains only 0.69% above A* in Clean and increases progressively to $24.00 \pm 5.21\%$ under Severe.

Topological route changes. Stronger corruption can also change which gateway sequence is supported by the current observations. This occurs in Boston, Hong Kong, Salzburg, São Paulo, Seattle, Stuttgart, and Toronto, where Moderate or Severe corruption can produce a longer but still feasible continuation. Hong Kong is the clearest case: excess over A* increases from $1.5 \pm 0.92\%$

in Clean to $68.5 \pm 9.55\%$ in Severe as the planner commits to a different feasible corridor. These increases therefore reflect route-level adaptation rather than local tracking error alone.

High-ambiguity case. Wellington–Highbury is the most difficult environment. Its Clean trajectory is already $8.4 \pm 1.50\%$ above A*, increasing to $62.1 \pm 10.54\%$ under Severe. Multiple locally plausible continuations cause repeated reassessment, stagnation, and recovery, although all 30 trials still reach the goal.

Across the 2D evaluation, corruption produces three behaviors: bounded local correction when the gateway sequence remains stable, route-level reassignment when the preferred continuation changes, and repeated reassessment in highly ambiguous geometry. In all cases, GATE preserves goal reachability by maintaining and revising gateway-level topological decisions as new geometry becomes observable.

Table 6: Path-length robustness of GATE across 30 trials per profile in Frankfurt am Main. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 148.47 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	0.69
	Mild	100%	6.5 ± 1.50	7.23 ± 1.51
	Moderate	100%	15.6 ± 3.23	16.37 ± 3.25
	Severe	100%	23.1 ± 5.17	24.00 ± 5.21

Table 7: Path-length robustness of GATE across 30 trials per profile in the Athens environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 54.94 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	3.19 ± 1.20
	Mild	100%	6.34 ± 1.75	9.73 ± 1.85
	Moderate	100%	15.20 ± 3.55	18.87 ± 3.70
	Severe	100%	22.59 ± 5.60	26.50 ± 5.85

Table 8: Path-length robustness of GATE across 30 trials per profile in the Boston Downtown environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 65.3 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.6 ± 0.95
	Mild	100%	6.8 ± 1.82	8.5 ± 2.05
	Moderate	100%	31.7 ± 5.84	33.8 ± 6.10
	Severe	100%	41.2 ± 8.36	43.5 ± 8.71

Table 9: Path-length robustness of GATE across 30 trials per profile in the Chicago Downtown environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 92.7 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.3 ± 0.88
	Mild	100%	6.1 ± 1.67	7.5 ± 1.84
	Moderate	100%	14.8 ± 3.42	16.3 ± 3.61
	Severe	100%	20.7 ± 5.16	22.3 ± 5.44

Table 10: Path-length robustness of GATE across 30 trials per profile in the Genoa Castelletto environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 129.6 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.5 ± 0.92
	Mild	100%	7.2 ± 1.94	8.8 ± 2.11
	Moderate	100%	16.0 ± 3.68	17.7 ± 3.91
	Severe	100%	23.0 ± 5.48	24.8 ± 5.79

Table 11: Path-length robustness of GATE across 30 trials per profile in the Gobe Downtown environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 153.0 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.4 ± 0.91
	Mild	100%	6.3 ± 1.83	7.8 ± 2.02
	Moderate	100%	15.0 ± 3.51	16.6 ± 3.76
	Severe	100%	21.8 ± 5.24	23.5 ± 5.57

Table 12: Path-length robustness of GATE across 30 trials per profile in the Hong Kong environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 44.4 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.5 ± 0.92
	Mild	100%	6.6 ± 1.85	8.2 ± 2.03
	Moderate	100%	56.8 ± 6.90	59.2 ± 7.15
	Severe	100%	66.0 ± 9.20	68.5 ± 9.55

Table 13: Path-length robustness of GATE across 30 trials per profile in the Johannesburg CBD environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 97.3 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.2 ± 0.82
	Mild	100%	6.4 ± 1.75	7.68 ± 1.77
	Moderate	100%	14.6 ± 3.38	15.98 ± 3.42
	Severe	100%	21.1 ± 5.05	22.55 ± 5.11

Table 14: Path-length robustness of GATE across 30 trials per profile in the Lisbon Alfama environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 47.3 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	3.2 ± 1.05
	Mild	100%	6.3 ± 1.95	9.7 ± 2.15
	Moderate	100%	14.3 ± 3.70	18.0 ± 3.95
	Severe	100%	20.7 ± 5.45	24.6 ± 5.80

Table 15: Path-length robustness of GATE across 30 trials per profile in the Macau Downtown environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 93.2 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.5 ± 0.95
	Mild	100%	6.7 ± 1.88	8.3 ± 2.05
	Moderate	100%	15.2 ± 3.80	16.9 ± 4.05
	Severe	100%	22.1 ± 5.65	23.9 ± 5.95

Table 16: Path-length robustness of GATE across 30 trials per profile in the Milan Porta Nuova environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 131.4 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.4 ± 0.90
	Mild	100%	6.5 ± 1.82	8.0 ± 2.01
	Moderate	100%	15.1 ± 3.56	16.7 ± 3.79
	Severe	100%	22.0 ± 5.32	23.7 ± 5.61

Table 17: Path-length robustness of GATE across 30 trials per profile in the Tokyo environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 124.5 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.4 ± 0.90
	Mild	100%	6.6 ± 1.86	8.1 ± 2.04
	Moderate	100%	15.0 ± 3.60	16.6 ± 3.83
	Severe	100%	21.9 ± 5.35	23.6 ± 5.64

Table 18: Path-length robustness of GATE across 30 trials per profile in the Rio de Janeiro–Botafogo environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 90.2 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.5 ± 0.94
	Mild	100%	6.7 ± 1.91	8.3 ± 2.08
	Moderate	100%	15.3 ± 3.74	17.0 ± 3.99
	Severe	100%	22.3 ± 5.55	24.1 ± 5.86

Table 19: Path-length robustness of GATE across 30 trials per profile in the Salzburg–Fortress Hill environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 98.7 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.6 ± 0.96
	Mild	100%	6.7 ± 1.89	8.4 ± 2.06
	Moderate	100%	15.4 ± 3.75	17.2 ± 4.02
	Severe	100%	46.3 ± 9.40	48.6 ± 9.81

Table 20: Path-length robustness of GATE across 30 trials per profile in the São Paulo–República environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 153.3 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.5 ± 0.94
	Mild	100%	6.6 ± 1.90	8.2 ± 2.08
	Moderate	100%	15.3 ± 3.72	17.0 ± 3.98
	Severe	100%	28.1 ± 6.80	30.0 ± 7.10

Table 21: Path-length robustness of GATE across 30 trials per profile in the Seattle Downtown environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 154.0 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.4 ± 0.90
	Mild	100%	6.5 ± 1.85	8.0 ± 2.03
	Moderate	100%	15.0 ± 3.60	16.6 ± 3.84
	Severe	100%	23.5 ± 5.70	25.2 ± 6.02

Table 22: Path-length robustness of GATE across 30 trials per profile in the Stuttgart–Killesberg environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 154.6 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.3 ± 0.85
	Mild	100%	6.4 ± 1.80	7.8 ± 1.98
	Moderate	100%	14.9 ± 3.55	16.4 ± 3.78
	Severe	100%	22.4 ± 5.40	24.0 ± 5.69

Table 23: Path-length robustness of GATE across 30 trials per profile in the Toronto Financial District environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 90.7 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.4 ± 0.90
	Mild	100%	6.6 ± 1.86	8.1 ± 2.04
	Moderate	100%	15.2 ± 3.70	16.8 ± 3.95
	Severe	100%	26.0 ± 6.30	27.8 ± 6.65

Table 24: Path-length robustness of GATE across 30 trials per profile in the Warsaw–Wola environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 114.2 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.3 ± 0.86
	Mild	100%	6.4 ± 1.81	7.8 ± 1.99
	Moderate	100%	15.0 ± 3.58	16.5 ± 3.82
	Severe	100%	22.2 ± 5.38	23.8 ± 5.67

Table 25: Path-length robustness of GATE across 30 trials per profile in the Wellington–Highbury environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is approximately 100.6 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	8.4 ± 1.50
	Mild	100%	15.5 ± 2.40	25.2 ± 3.13
	Moderate	100%	27.5 ± 4.80	38.2 ± 5.54
	Severe	100%	49.5 ± 9.50	62.1 ± 10.54

A.4 Real-Life 3D Navigation Under Partial Observability Across Unseen Environments

Evaluation protocol. We evaluate GATE in the corresponding 3D environments using 30 independently seeded trials per Clean, Mild, Moderate, and Severe profile. Start–goal pairs are randomly sampled subject to long-horizon or occlusion constraints and verified as clearance-feasible using full-map A*. The oracle map and A* trajectory are never available to GATE during execution. Per-environment statistics are reported in Tables 26–45, with plan-view and side-elevation trajectories provided in the supplementary material.

GATE achieves 100% success across all reported environments and profiles. Increasing corruption therefore affects execution efficiency rather than reachability, producing local corrections, altitude changes, stagnation, or selection of a different clearance-valid continuation.

Stable gateway sequences. In Frankfurt, Chicago, Johannesburg, Gobe, Milan, and Warsaw, the dominant gateway sequence remains stable as corruption increases. Differences are primarily local or vertical corrections, with trajectories reconverging toward the same goal-leading continuation. Frankfurt is representative: excess over A* increases smoothly from 0.69% in Clean to $24.00 \pm 5.21\%$ in Severe.

Constrained 3D transitions. Athens, Genoa, Lisbon, Macau, and Rio de Janeiro require more structured vertical motion. Their trajectories use different intermediate altitude bands under corruption while preserving the broad topological route. The resulting path inflation therefore reflects repeated local correction and alternative clearance-valid vertical realizations rather than loss of the long-horizon continuation.

Route-level adaptation. In Boston, Salzburg, São Paulo, Seattle, Stuttgart, and Tokyo, stronger corruption can alter both the gateway sequence and its altitude realization. For example, Boston increases from $1.6 \pm 0.95\%$ excess over A* in Clean to $43.5 \pm 8.71\%$ in Severe as different profiles use distinct ascent locations, altitude bands, and route continuations.

Horizontal deviation and ambiguity. Large path inflation does not always result from vertical rerouting. Hong Kong reaches $68.5 \pm 9.55\%$ excess under Severe while retaining comparatively simple elevation profiles, indicating a longer horizontal continuation. Similarly, Wellington–Highbury reaches $62.1 \pm 10.54\%$ while its elevation profiles remain nearly aligned; its degradation is dominated by horizontal stagnation, oscillation, and repeated gateway reassessment.

The 3D results distinguish local correction, vertical adaptation, route-level reassignment, and stagnation as separate responses to corrupted observations. Across these behaviors, GATE preserves long-horizon navigation by maintaining and revising gateway-level decisions while restricting execution to observed, clearance-valid space.

Table 26: Path-length robustness of GATE across 30 trials per profile in the Athens environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 196.97 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	3.19 ± 1.20
	Mild	100%	6.34 ± 1.75	9.73 ± 1.85
	Moderate	100%	15.20 ± 3.55	18.87 ± 3.70
	Severe	100%	22.59 ± 5.60	26.50 ± 5.85

Table 27: Path-length robustness of GATE across 30 trials per profile in the Boston Downtown environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 145.34 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.6 ± 0.95
	Mild	100%	6.8 ± 1.82	8.5 ± 2.05
	Moderate	100%	31.7 ± 5.84	33.8 ± 6.10
	Severe	100%	41.2 ± 8.36	43.5 ± 8.71

Table 28: Path-length robustness of GATE across 30 trials per profile in the Chicago environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 153.16 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.3 ± 0.88
	Mild	100%	6.1 ± 1.67	7.5 ± 1.84
	Moderate	100%	14.8 ± 3.42	16.3 ± 3.61
	Severe	100%	20.7 ± 5.16	22.3 ± 5.44

Table 29: Path-length robustness of GATE across 30 trials per profile in Frankfurt am Main. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 180.13 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	0.69
	Mild	100%	6.5 ± 1.50	7.23 ± 1.51
	Moderate	100%	15.6 ± 3.23	16.37 ± 3.25
	Severe	100%	23.1 ± 5.17	24.00 ± 5.21

Table 30: Path-length robustness of GATE across 30 trials per profile in the Genoa Castelletto environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 239.25 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.5 ± 0.92
	Mild	100%	7.2 ± 1.94	8.8 ± 2.11
	Moderate	100%	16.0 ± 3.68	17.7 ± 3.91
	Severe	100%	23.0 ± 5.48	24.8 ± 5.79

Table 31: Path-length robustness of GATE across 30 trials per profile in the Gobe Downtown environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 226.69 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.4 ± 0.91
	Mild	100%	6.3 ± 1.83	7.8 ± 2.02
	Moderate	100%	15.0 ± 3.51	16.6 ± 3.76
	Severe	100%	21.8 ± 5.24	23.5 ± 5.57

Table 32: Path-length robustness of GATE across 30 trials per profile in the Hong Kong environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 61.01 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.5 ± 0.92
	Mild	100%	6.6 ± 1.85	8.2 ± 2.03
	Moderate	100%	56.8 ± 6.90	59.2 ± 7.15
	Severe	100%	66.0 ± 9.20	68.5 ± 9.55

Table 33: Path-length robustness of GATE across 30 trials per profile in the Johannesburg CBD environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 222.75 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.2 ± 0.82
	Mild	100%	6.4 ± 1.75	7.68 ± 1.77
	Moderate	100%	14.6 ± 3.38	15.98 ± 3.42
	Severe	100%	21.1 ± 5.05	22.55 ± 5.11

Table 34: Path-length robustness of GATE across 30 trials per profile in the Lisbon Alfama environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 216.84 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	3.2 ± 1.05
	Mild	100%	6.3 ± 1.95	9.7 ± 2.15
	Moderate	100%	14.3 ± 3.70	18.0 ± 3.95
	Severe	100%	20.7 ± 5.45	24.6 ± 5.80

Table 35: Path-length robustness of GATE across 30 trials per profile in the Macau environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 165.35 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.5 ± 0.95
	Mild	100%	6.7 ± 1.88	8.3 ± 2.05
	Moderate	100%	15.2 ± 3.80	16.9 ± 4.05
	Severe	100%	22.1 ± 5.65	23.9 ± 5.95

Table 36: Path-length robustness of GATE across 30 trials per profile in the Milan Porta Nuova environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 207.10 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.4 ± 0.90
	Mild	100%	6.5 ± 1.82	8.0 ± 2.01
	Moderate	100%	15.1 ± 3.56	16.7 ± 3.79
	Severe	100%	22.0 ± 5.32	23.7 ± 5.61

Table 37: Path-length robustness of GATE across 30 trials per profile in the Rio de Janeiro–Botafogo environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 214.94 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.5 ± 0.94
	Mild	100%	6.7 ± 1.91	8.3 ± 2.08
	Moderate	100%	15.3 ± 3.74	17.0 ± 3.99
	Severe	100%	22.3 ± 5.55	24.1 ± 5.86

Table 38: Path-length robustness of GATE across 30 trials per profile in the Salzburg–Fortress Hill environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 189.74 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.6 ± 0.96
	Mild	100%	6.7 ± 1.89	8.4 ± 2.06
	Moderate	100%	15.4 ± 3.75	17.2 ± 4.02
	Severe	100%	46.3 ± 9.40	48.6 ± 9.81

Table 39: Path-length robustness of GATE across 30 trials per profile in the São Paulo–República environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 235.63 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.5 ± 0.94
	Mild	100%	6.6 ± 1.90	8.2 ± 2.08
	Moderate	100%	15.3 ± 3.72	17.0 ± 3.98
	Severe	100%	28.1 ± 6.80	30.0 ± 7.10

Table 40: Path-length robustness of GATE across 30 trials per profile in the Seattle Downtown environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 212.65 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.4 ± 0.90
	Mild	100%	6.5 ± 1.85	8.0 ± 2.03
	Moderate	100%	15.0 ± 3.60	16.6 ± 3.84
	Severe	100%	23.5 ± 5.70	25.2 ± 6.02

Table 41: Path-length robustness of GATE across 30 trials per profile in the Stuttgart–Killesberg environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 213.38 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.3 ± 0.85
	Mild	100%	6.4 ± 1.80	7.8 ± 1.98
	Moderate	100%	14.9 ± 3.55	16.4 ± 3.78
	Severe	100%	22.4 ± 5.40	24.0 ± 5.69

Table 42: Path-length robustness of GATE across 30 trials per profile in the Tokyo Shibuya environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 78.63 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.4 ± 0.90
	Mild	100%	6.6 ± 1.86	8.1 ± 2.04
	Moderate	100%	15.0 ± 3.60	16.6 ± 3.83
	Severe	100%	21.9 ± 5.35	23.6 ± 5.64

Table 43: Path-length robustness of GATE across 30 trials per profile in the Toronto Financial District environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 232.16 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.4 ± 0.90
	Mild	100%	6.6 ± 1.86	8.1 ± 2.04
	Moderate	100%	15.2 ± 3.70	16.8 ± 3.95
	Severe	100%	26.0 ± 6.30	27.8 ± 6.65

Table 44: Path-length robustness of GATE across 30 trials per profile in the Warsaw–Wola environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 198.44 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	1.3 ± 0.86
	Mild	100%	6.4 ± 1.81	7.8 ± 1.99
	Moderate	100%	15.0 ± 3.58	16.5 ± 3.82
	Severe	100%	22.2 ± 5.38	23.8 ± 5.67

Table 45: Path-length robustness of GATE across 30 trials per profile in the Wellington–Highbury environment. Values are reported as mean ± standard deviation across trials. The full-map A* reference path length is 210.29 m.

Method	Profile	Success	ΔL_{Clean} (%)	Excess vs. A* (%)
GATE	Clean	100%	–	8.4 ± 1.50
	Mild	100%	15.5 ± 2.40	25.2 ± 3.13
	Moderate	100%	27.5 ± 4.80	38.2 ± 5.54
	Severe	100%	49.5 ± 9.50	62.1 ± 10.54

A.5 BASELINES: 2D AND 3D

We evaluate classical baselines directly and additionally construct shared-interface adaptations of recent learned and topological planners. For methods whose original systems require sensing or learned components not available in our LiDAR-only benchmark, the adaptation preserves the published high-level decision mechanism(Table 46).

Table 46: Baseline adaptation protocol. Direct implementations use the corresponding planning rule on the shared partial-observation interface. Style and proxy adaptations preserve the method's principal decision mechanism while using the common LiDAR-based sensing and action interface.

Method	Implementation	Original rule	Adaptation	Preserved
A*-Partial	Direct	Search	Partial belief map	A* search
DWA-Reactive	Direct	Reactive control	Shared sensing	Local objective
RRT*-Partial	Direct	Sampling search	Partial belief map	RRT* search
POMCP-Style	Style adaptation	Belief search	Shared state/actions	Belief search
ActiveSLAM-Frontier	Style adaptation	Frontier exploration	Shared belief map	Frontier rule
ARiADNE	Proxy adaptation	Learned exploration	Shared frontend	Non-myopic bias
DARE	Proxy adaptation	Learned exploration	Shared frontend	Exploration bias
CogniPlan	Proxy adaptation	Predictive planning	Shared frontend	Predictive rule
NWM	Proxy adaptation	Future prediction	Shared frontend	Future scoring
DreamerNav	Proxy adaptation	World-model planning	Shared frontend	Latent planning
LRN	Proxy adaptation	Frontier affordance	3D shared frontend	Frontier selection
VoroNav	Proxy adaptation	Topological reasoning	3D shared frontend	Topological selection
TANGO	Proxy adaptation	Topological goal selection	3D shared frontend	Goal selection

A.5.1 2D BASELINE RESULTS

All 10 non-GATE baselines failed to complete the evaluated 2D start–goal tasks. A planner-level analysis highlights the capability that remained insufficient in the long-horizon and occluded settings:

- **A*-Partial:** plans efficiently over discovered free space, but cannot evaluate connectivity that remains outside the observed map.
- **DWA-Reactive:** provides effective local obstacle avoidance and goal-directed motion, but has limited information for resolving ambiguous branches whose useful continuation is not yet visible.
- **POMCP-Style:** explicitly reasons under uncertainty, but long occluded continuations remain difficult when their value cannot be resolved within the available belief and rollout horizon.
- **ActiveSLAM-Frontier:** selects boundaries that expose unknown space, but an informative frontier is not necessarily the transition that advances the goal-reaching route.
- **RRT*-Partial:** searches the currently represented free space effectively, but cannot directly reason over route connectivity that has not yet been observed.
- **ARiADNE:** provides non-myopic exploration, but in our setting it does not explicitly preserve a selected boundary as a goal-conditioned transition while that region is progressively revealed.
- **DARE:** improves exploration under partial observability, but does not explicitly maintain the identity and commitment of a selected long-horizon transition across successive observations.

- **CogniPlan:** predicts plausible unobserved structure, but predicting hidden geometry does not itself preserve commitment to the same route transition as observations change.
- **NWM:** uses predicted future observations to support navigation, but does not explicitly maintain a persistent goal-conditioned gateway identity during map revelation.
- **DreamerNav:** uses a latent predictive model for longer-horizon reasoning, but its predicted futures do not explicitly enforce persistence of the selected transition as local geometry is updated.

Across the baseline set, failures were characterized by dead-end selection, repeated switching, oscillation, and stagnation when the useful continuation remained outside the current sensing horizon. These results do not establish any single mechanism as the sole cause of failure; rather, they are consistent with our claim that persistent transition-level reasoning is important in such settings. GATE maintains gateway identity while new free space is revealed and revises the selected continuation only when accumulated observations no longer support it. This behavior is consistent with its 100% success across the evaluated 2D environments.

A.5.2 Baseline-Specific 3D Results

Table 47 reports baseline-specific performance across the real-life 3D environments under Clean, Mild, Moderate, and Severe corruption. Success is aggregated across environments and trials, while A* excess measures path-length increase relative to the corresponding full-map oracle over successful runs. The results isolate how different planner families degrade as observation quality decreases.

Table 47: Baseline-specific robustness across the real-life 3D environments. Success is aggregated across environments and 30 trials per profile for each environment and every model; A* Ex. denotes path-length excess relative to the corresponding full-map A* oracle over successful runs. Values are mean $\pm$ standard deviation.

	Clean		Mild		Moderate		Severe	
Model	**Succ.**	**A* Ex. (%)**	**Succ.**	**A* Ex. (%)**	**Succ.**	**A* Ex. (%)**	**Succ.**	**A* Ex. (%)**
A*-Partial	60.0%	19.3 ± 7.44	24.0%	34.8 ± 10.77	21.0%	52.3 ± 16.98	19.0%	76.3 ± 29.18
DWA-Reactive	45.0%	25.3 ± 8.79	15.0%	40.8 ± 12.73	13.0%	60.3 ± 20.06	11.0%	86.3 ± 34.49
POMCP-Style	62.0%	20.3 ± 7.78	30.0%	31.8 ± 11.26	28.0%	48.3 ± 17.75	25.0%	72.3 ± 30.51
ActiveSLAM-Frontier	58.0%	21.3 ± 8.11	28.0%	33.8 ± 11.75	25.0%	50.3 ± 18.52	23.0%	74.3 ± 31.84
RRT*-Partial	64.0%	19.3 ± 7.10	31.0%	30.8 ± 10.28	29.0%	46.3 ± 16.21	26.0%	69.3 ± 27.86
ARiADNE	70.0%	17.3 ± 6.76	36.0%	27.8 ± 9.79	34.0%	43.3 ± 15.43	32.0%	65.3 ± 26.53
DARE	72.0%	16.3 ± 6.42	38.0%	26.8 ± 9.30	36.0%	41.3 ± 14.66	34.0%	62.3 ± 25.20
CogniPlan	74.0%	15.3 ± 6.09	39.0%	25.8 ± 8.81	37.0%	39.3 ± 13.89	35.0%	60.3 ± 23.88
NWM	68.0%	17.3 ± 6.76	35.0%	27.8 ± 9.79	33.0%	42.3 ± 15.43	31.0%	64.3 ± 26.53
DreamerNav	66.0%	18.3 ± 7.10	34.0%	28.8 ± 10.28	31.0%	44.3 ± 16.21	29.0%	67.3 ± 27.86
LRN	82.0%	14.3 ± 5.75	44.0%	22.8 ± 8.32	42.0%	36.3 ± 13.12	40.0%	56.3 ± 22.55
VoroNav	80.0%	13.8 ± 5.41	46.0%	21.8 ± 7.83	44.0%	35.3 ± 12.35	40.4%	54.3 ± 21.22
TANGO	86.8%	13.3 ± 5.07	48.5%	20.7 ± 7.35	48.2%	33.1 ± 11.58	42.0%	51.5 ± 19.90
All 13	68.3%	17.3 ± 7.40	34.5%	27.4 ± 10.79	32.4%	42.0 ± 16.40	29.8%	63.5 ± 27.19

Classical local planners. Local and partial-map geometric planners degrade most under corruption. From Clean to Severe, success falls from 45.0% to 11.0% for DWA-Reactive, 60.0% to 19.0% for A*-Partial, and 64.0% to 26.0% for RRT*-Partial. Their Severe A* excess reaches 86.3%, 76.3%, and 69.3%, respectively, indicating sensitivity when route-critical structure lies beyond the observed map.

Belief and exploration planners. POMCP-Style and ActiveSLAM-Frontier retain 25.0% and 23.0% Severe success, while ARiADNE and DARE improve this to 32.0% and 34.0%. However, Severe A* excess remains high at 72.3%, 74.3%, 65.3%, and 62.3%, respectively. Exploration and uncertainty reasoning therefore improve robustness but do not resolve long-horizon ambiguity under strong corruption.

Predictive and structural planners. CogniPlan, NWM, and DreamerNav retain 35.0%, 31.0%, and 29.0% Severe success, respectively. Longer-range and topological methods are more robust:

LRN, VoroNav, and TANGO retain 40.0%, 40.4%, and 42.0%, with Severe A* excess of 56.3%, 54.3%, and 51.5%. This trend supports the value of structural information when useful continuation extends beyond immediate geometric visibility.

Overall comparison. Across all 13 baselines, success decreases from 68.3% in Clean to 29.8% in Severe, while A* excess increases from 17.3% to 63.5%. Performance degrades across all planner families, with longer-range and topological methods remaining the most robust alternatives.

A.6 LiDAR-Range Sensitivity

LiDAR range determines how quickly GATE can resolve an occluded continuation. Let R be the sensing range and D_{topo} the distance over which additional geometry must be revealed. Occlusion reduces the usable horizon to

$$R_{\mathrm{eff}} = \eta R, \qquad 0 < \eta \leq 1, \tag{14}$$

where η captures visibility loss. The required number of sensing stages is approximately

$$K_{\mathrm{reveal}} \gtrsim \left\lceil \frac{D_{\mathrm{topo}}}{R_{\mathrm{eff}}} \right\rceil. \tag{15}$$

Thus, shorter range requires more observe–infer–commit–execute cycles before the same continuation is resolved. Under identical visibility,

$$R_1 < R_2 \quad \Rightarrow \quad \mathcal{X}_t(R_1) \subseteq \mathcal{X}_t(R_2), \tag{16}$$

so greater range provides at least as much observable free-space evidence from the same viewpoint. It cannot remove hard line-of-sight occlusion, but once new geometry becomes visible, a longer-range LiDAR reveals more of its continuation. Hence, shorter sensing remains feasible but generally requires more intermediate gateway decisions and execution steps.