SPM-NAV: PERSISTENT SPACE–PORTAL MEMORY FOR ZERO-SHOT OBJECT NAVIGATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Zero-shot object navigation requires long-horizon search in unseen environments, where effective decisions depend on both target likelihood and accumulated search history. However, existing spatial representations often leave search evidence and progress fragmented, limiting the persistence and reuse of search experience across revisits. To address this issue, SPM-Nav presents a zero-shot object navigation framework centered on Persistent Space–Portal Memory. Each Space maintains a persistent regional search state that consolidates regional semantics, target evidence, historical observations, negative evidence, and search progress across locations and viewpoints. Portals represent traversal-validated inter-Space transitions and preserve reusable connectivity between regions. Building upon this memory, Adaptive Space Search dynamically adjusts local search effort according to accumulated Space states, reducing redundant inspection while preserving essential inter-region connectivity. SPM-Nav is evaluated on HM3D-v1, HM3D-v2, and MP3D, achieving success rates of 70.6%, 83.8%, and 55.7%, respectively. Component ablations and mechanism analyses further validate the effectiveness of persistent search states, reusable inter-Space transitions, and memory-conditioned adaptive allocation.

1 INTRODUCTION

Zero-Shot Object Navigation (ZSON) requires embodied agents to autonomously explore unseen environments and locate target instances specified by object categories Batra et al. (2020); Kuang et al. (2024). Recent advances in open-vocabulary vision models and vision-language models have substantially improved agents' ability to estimate target relevance from visual observations and environmental context, enabling navigation without task-specific training for target categories Zhou et al. (2023); Yokoyama et al. (2024); Zhang et al. (2025b). However, object search is inherently a long-horizon process. Since target-related evidence is often distributed across different locations and viewpoints, effective decisions depend not only on where the target may be, but also on what has already been searched. Agents must therefore turn observations accumulated throughout exploration into reusable search experience.

Existing spatial representations organize geometry, semantics, and navigation experience through metric and semantic maps He et al. (2026) or topological graphs, effectively supporting local navigation and long-range planning Chaplot et al. (2020); An et al. (2025). However, search evidence and progress can remain distributed across locations and viewpoints within the same physical region, while exploration decisions are often tied to transient candidates such as frontiers whose identities change as exploration proceeds Yamauchi (1997). As frontiers move, split, merge, or disappear with the explored–unexplored boundary, accumulated search experience may lose a stable spatial association and become fragmented over time.

This limitation is particularly important for object search because semantic relevance alone does not indicate whether further inspection remains worthwhile or how thoroughly a region has already been searched. The same physical region may be observed from multiple viewpoints or navigation locations, each providing partial evidence about the target. Without a persistent regional search state, these observations remain associated with transient exploration decisions and cannot be effectively reused during revisits. Figure 1 illustrates this issue. Existing exploration cues, such as frontiers, represent transient opportunities whose identities change as exploration proceeds, causing accu-

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.

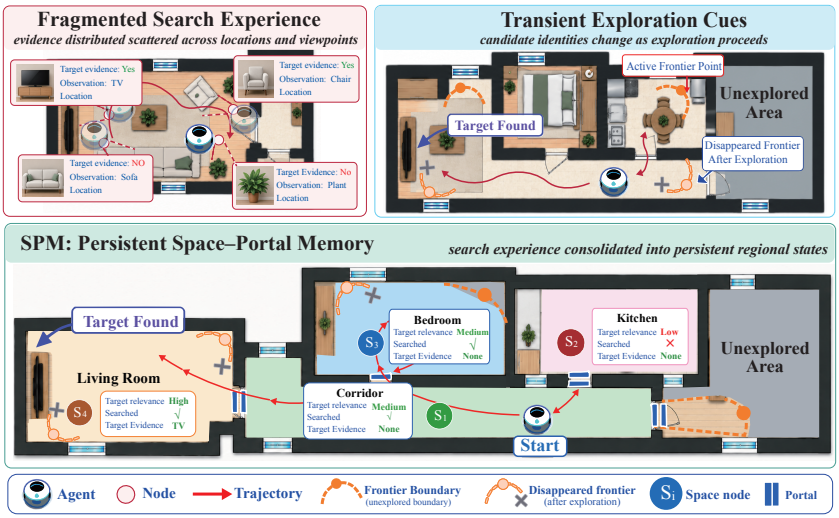

Figure 1: **From Fragmented Search Experience to Persistent Space–Portal Memory.** Observations collected from different locations and viewpoints within the same physical region can remain fragmented when exploration decisions rely on transient cues such as frontiers Zhao et al. (2025). SPM-Nav consolidates such evidence and search progress into persistent Space states, while Portals preserve validated inter-Space transitions and reusable connectivity across regions.

mulated evidence to become fragmented. SPM-Nav instead consolidates observations and search progress into persistent Space states while maintaining reusable inter-Space connectivity through Portals.

To address this problem, SPM-Nav is introduced as a zero-shot object navigation framework centered on **Persistent Space–Portal Memory**. Each **Space** is maintained as a persistent region-level unit rather than an individual navigation location, and consolidates regional semantics, target evidence, historical observations, negative evidence, and search progress across locations, viewpoints, and exploration stages. This allows accumulated search experience to remain associated with the same region and inform subsequent search decisions. Region-level search states alone are insufficient, since regions that no longer warrant further inspection may still be important for reaching other parts of the environment. SPM-Nav therefore introduces **Portals** to represent traversable region transitions, such as doorways and openings, which are retained after traversal validation. Portals separate the search states of adjacent Spaces while preserving reusable inter-region connectivity. Together, Spaces and Portals maintain both persistent regional search states and the transition structure required for subsequent exploration. Building on Persistent Space–Portal Memory, SPM-Nav further introduces **Adaptive Space Search**, which adjusts local search effort according to accumulated Space states. Rather than repeatedly allocating similar effort based only on current semantic relevance, the agent accounts for previously accumulated evidence and search progress, reducing redundant inspection while preserving access to downstream regions.

The main contributions are summarized as follows:

1. **Persistent Space–Portal Memory.** A persistent spatial memory is introduced for zero-shot object navigation that consolidates accumulated search experience into region-level search states while preserving validated and reusable inter-Space transitions.

2. **Memory-Conditioned Adaptive Space Search.** Building on the proposed memory, Adaptive Space Search regulates local exploration effort according to regional relevance and accumulated search states, while preserving necessary inter-region connectivity.

3. **Comprehensive Evaluation and Analysis.** Experiments on three ObjectNav benchmarks, together with component ablations and mechanism analyses, validate persistent search-state reuse, reusable inter-Space connectivity, and memory-conditioned adaptive search.

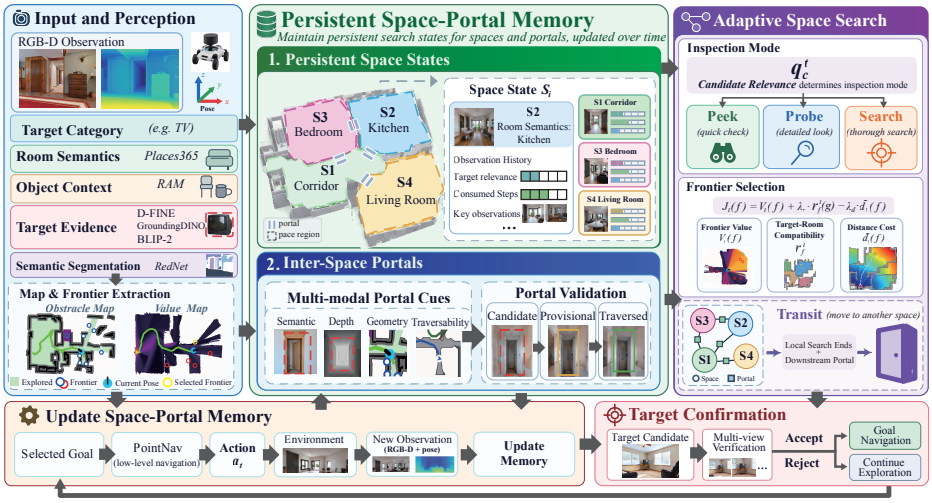

Figure 2: **Overview of SPM-Nav.** SPM-Nav constructs a Persistent Space–Portal Memory from online RGB–D perception and mapping. Spaces aggregate persistent region-level search states, while Portals encode validated and reusable inter-Space transitions. Adaptive Space Search queries this memory to regulate inspection depth, local frontier selection, and connectivity-preserving Transit. Selected navigation goals are executed by a pretrained PointNav policy, while observations and target verification continuously update the memory, forming a closed-loop navigation process.

2 Related Work

Spatial Memory and Structured Representations. Metric and semantic maps aggregate geometry and scene semantics into a unified spatial map Chaplot et al. (2020), while topological approaches organize navigation experience around landmarks, viewpoints, or waypoints for long-range planning Kim et al. (2023); Chen et al. (2022); An et al. (2025). TopoNav organizes navigation memory around discrete topological nodes with semantic and frontier attributes Liu et al. (2025). BSC-Nav combines landmark memory with trajectory-based allocentric cognitive maps to support subsequent navigation and reasoning Ruan et al. (2026). EvoMemNav uses a room–view–object hierarchy and associates navigation priors with room, anchor-view, and frontier-view nodes Ge et al. (2026). Nevertheless, search experience can remain distributed across locations, views, and transient exploration candidates. SPM-Nav instead organizes accumulated search information in a Persistent Space–Portal Memory that couples regional search states with validated inter-region transitions.

Zero-Shot Object Navigation and Semantic Exploration. Training-free zero-shot object navigation methods leverage pretrained visual-semantic models to guide exploration without task-specific training Gadre et al. (2023); Zhou et al. (2023); Yokoyama et al. (2024). Recent approaches further improve navigation through structured semantic reasoning and adaptive use of semantic and geometric cues Gong et al. (2026); Zhang et al. (2025b). These methods primarily improve the estimation and selection of promising exploration targets from current observations. Building on Persistent Space–Portal Memory, SPM-Nav instead regulates subsequent search effort according to accumulated regional search information.

3 Method

3.1 Overview

Given a target object category g, SPM-Nav explores an unseen environment from continuous RGB–D observations and agent poses, with the goal of navigating to a target instance. Figure 2 illustrates the overall framework. Online perception and mapping provide regional semantics, target evidence, and local exploration candidates. SPM-Nav accumulates regional information in a Persistent Space–Portal Memory, while Adaptive Space Search uses the resulting memory states together with current exploration candidates to regulate local inspection and inter-Space navigation. Selected navigation

goals are executed by a pretrained PointNav policy, and the resulting observations are fed back to update the memory, forming a closed-loop exploration process. The memory is incrementally maintained and represented at time step t as:

$$\mathcal{M}_t^{SP} = (\mathcal{S}_t, \mathcal{P}_t)\,, \tag{1}$$

where $\mathcal{S}_t$ and $\mathcal{P}_t$ denote the sets of Spaces and Portals established by time step t, respectively. A **Space** is a persistent region-level unit rather than an individual observation or navigation location, and aggregates regional semantics, target evidence, historical observations, and search progress collected across locations and viewpoints. A **Portal** is an explicitly localized and traversal-validated inter-Space transition that preserves reusable connectivity together with navigable anchors on both sides. In this way, region-level search states and their inter-region connectivity are jointly organized within a persistent memory. Building on this memory, **Adaptive Space Search** regulates local search effort and inter-Space transitions according to accumulated Space states. Selected navigation goals are executed by a pretrained PointNav policy, and newly acquired observations are used to update the Space–Portal Memory, forming a closed-loop online exploration process.

3.2 Persistent Space–Portal Memory

Persistent Space–Portal Memory couples persistent regional search states with validated inter-Space transitions. Spaces maintain accumulated search information within regions, while Portals define their transition boundaries and reusable connectivity.

3.2.1 Persistent Space States

SPM-Nav uses Spaces as persistent region-level units of search state. Unlike viewpoint- or location-based topological representations, changes in robot pose, observation viewpoint, or local navigation position do not by themselves instantiate new Spaces.

Space Formation and Identity Maintenance. A new Space hypothesis is introduced only when the agent encounters evidence of an inter-region transition. The hypothesis is initially maintained provisionally, while subsequent observations accumulate its spatial support and regional semantics before the Space is stabilized. Observations are associated with an existing Space when they remain consistent with its accumulated spatial support and no confirmed inter-region transition has been crossed. The active Space is switched only after such a transition is confirmed. Once stabilized, an existing Space identity is reused when the region is re-associated during subsequent revisits, allowing previously accumulated observations and search states to remain available for future decisions. Detailed association and stabilization criteria are provided in Appendix A. The state of the i-th Space at time step t is defined as

$$S_i^t = \left(p_i^t, \mathcal{V}_i^t, s_i^t(g), z_i^t, B_i^t, L_i^t, N_i^t\right), \tag{2}$$

where p_i^t denotes the regional semantics accumulated over time, $\mathcal{V}_i^t$ is the set of valid historical observations, and $s_i^t(g)$ denotes the persistent relevance of the Space to target category g. The variables z_i^t, B_i^t, and L_i^t denote the search priority, local search budget, and consumed local-search steps, respectively, while N_i^t denotes the accumulated number of independent negative observations.

Regional Semantics and Target Relevance. To reduce the sensitivity of single-frame scene recognition to viewpoint variation and limited field of view, SPM-Nav combines scene semantics from Places365 with object context extracted by RAM to obtain the room-semantic distribution $\tilde{p}_t$ for the current observation. Valid observations associated with the same Space are continuously aggregated into persistent regional semantics p_i^t. For target category g, let $\pi_g(r)$ denote its target–room prior Gong et al. (2026). We define the compatibility between a room-semantic distribution p and the target prior as

$$\mathrm{Compat}(p, \pi_g) = \frac{\sum_{r\in\mathcal{R}} p(r)\pi_g(r)}{\max_{r\in\mathcal{R}} \pi_g(r)}, \tag{3}$$

where $\mathcal{R}$ denotes the set of room-semantic categories.

The observation-level compatibility score is recursively updated with the historical relevance using an exponentially weighted moving average (EWMA) Qi et al. (2026), yielding the Space-level target

relevance:

$$s_i^t(g) = (1-\lambda)s_i^{t-1}(g) + \lambda \, \mathrm{Compat}\left(\tilde{p}_t, \pi_g\right), \tag{4}$$

where $\lambda \in (0, 1]$ controls the update weight of the current observation. Here, p_i^t represents persistent regional semantics that are independent of the current target category, whereas $s_i^t(g)$ captures the persistent target relevance of Space S_i with respect to target category g.

Search Evidence and Progress. Semantic relevance reflects the potential search value of a region, but does not indicate how thoroughly the region has already been inspected. SPM-Nav therefore maintains valid historical observations acquired from distinct positions and viewing directions:

$$\mathcal{V}_i^t = \{(\mathbf{x}_k, \theta_k, e_k)\}_{k \in \mathcal{K}_i^t}, \tag{5}$$

where $\mathbf{x}_k$, θ_k, and e_k denote the observation position, viewing direction, and visual target evidence, respectively. A new observation is counted as independent only if it is sufficiently distinct from existing records in position or viewing direction. Among these independent observations, those whose target evidence falls below a threshold are counted toward N_i^t as negative evidence. Such evidence indicates that the region has been inspected from distinct informative positions or viewing directions without obtaining sufficient support for the target. Meanwhile, L_i^t continuously records the number of local-search steps already executed within the current Space, which is later used to determine its remaining search budget (see Sec. 3.3.2). These states are persistently associated with the Space identity, allowing previously accumulated search progress to be directly reused upon revisiting the region.

3.2.2 Inter-Space Portals

Space-level states additionally require explicit transition boundaries and reusable inter-region connectivity. SPM-Nav therefore represents each such transition as a **Portal**. Unlike an abstract graph edge that only records connectivity, a Portal explicitly localizes and validates the physical transition through which Space membership changes. Each confirmed Portal retains the transition location and navigable anchors on both sides, allowing it to serve both as the boundary between adjacent Space states and as a reusable connection between them. Potential Portals are proposed using complementary semantic, depth, local geometric, and traversability cues, including doorway/opening semantics, opening structures observed in RGB–D input, local transition geometry, and navigability on both sides of the transition.

To reduce incorrect spatial partitioning caused by transient perception errors, SPM-Nav adopts **delayed commitment** for Portal construction. A candidate transition is first validated by subsequent observations and traversability evidence, while navigable anchors on both sides are identified as the agent approaches the transition. The relation is stored as a persistent inter-Space transition only after an actual traversal has been confirmed. The j-th Portal is represented as

$$P_j^t = \left(\mathbf{q}_j, S_j^-, S_j^+, \mathbf{a}_j^-, \mathbf{a}_j^+, \sigma_j^t\right), \tag{6}$$

where $\mathbf{q}_j$ denotes the transition location, S_j^- and S_j^+ denote the Spaces on the two sides of the Portal, $\mathbf{a}_j^-$ and $\mathbf{a}_j^+$ are the corresponding navigable anchors, and σ_j^t denotes the confirmation state.

Before traversal is confirmed, newly acquired observations continue to update the source Space S_j^-. After the transition is confirmed, subsequent semantic observations, target evidence, and search progress are associated with the destination Space S_j^+. Once confirmed, a Portal persistently retains its two-sided anchors and transition relation, allowing the connection to be directly reused during subsequent revisits, inter-Space navigation, and recovery.

3.3 Adaptive Space Search

Persistent Space–Portal Memory maintains continuously updated search states for different regions. Based on these states, SPM-Nav employs **Adaptive Space Search** to regulate local search effort across regions while distinguishing regional search value from inter-region transit value.

3.3.1 Adaptive Inspection of Candidate Regions

For a candidate region c that has not yet formed a stable Space, SPM-Nav estimates its search relevance by combining room semantics, object context, direct target evidence, and observation history:

$$q_c^t = w_r r_c^t + w_o o_c^t + w_v v_c^t + w_h h_c^t, \qquad \sum_{\ell \in \{r,o,v,h\}} w_\ell = 1, \tag{7}$$

where r_c^t and o_c^t measure the compatibility of scene-level room semantics and object-derived contextual semantics with the target–room prior, respectively; v_c^t represents direct target evidence from the current observation, and h_c^t decreases as the number of independent historical observations of the region increases, thereby discouraging redundant inspection. According to q_c^t, the system adjusts inspection intensity among **Peek**, **Probe**, and **Search**. Peek performs low-cost observation near the region entrance; Probe obtains additional evidence through limited entry; and Search conducts more thorough local exploration in highly relevant candidate regions. Newly acquired observations continuously update the candidate state, allowing the inspection intensity to adapt as evidence accumulates.

3.3.2 Space-level Search and Connectivity-Preserving Transit

Once a candidate region forms a stable Space, its subsequent search effort is jointly determined by persistent target relevance and accumulated search state. To suppress priority chattering near decision boundaries, SPM-Nav updates the Space priority through a state-dependent hysteretic mapping Hespanha et al. (2003):

$$z_i^t = Q\left(s_i^t(g), z_i^{t-1}\right) \in \{\mathrm{LOW}, \mathrm{MEDIUM}, \mathrm{HIGH}\}, \tag{8}$$

which prevents frequent switching caused by small relevance fluctuations near priority boundaries. Higher search priorities are assigned larger local-search budgets $B(z_i^t)$. If the current Space has already consumed L_i^t local-search steps, its remaining budget is

$$B_{i,\mathrm{rem}}^t = \max\left(0, B(z_i^t) - L_i^t\right). \tag{9}$$

Accordingly, newly acquired target evidence can adjust the amount of future search effort through the updated priority, while previously consumed search cost is preserved across revisits. In addition to the remaining budget, accumulated negative evidence N_i^t provides an early-stopping signal when the region has already been sufficiently inspected without obtaining target support. During local Search, SPM-Nav separates Space-level search allocation from local frontier selection. The Space state determines whether Search should continue and how much search effort remains, while frontiers serve as local navigation candidates within the current Space. For a candidate frontier f, let $V_t(f)$ denote its exploration value estimated by the vision-language model, $r_f^t = \mathrm{Compat}(p_f^t, \pi_g)$ denote its target–room compatibility, and $\bar{d}_t(f)$ denote its normalized navigation cost. The local frontier score is defined as

$$J_t(f) = V_t(f) + \lambda_r r_f^t - \lambda_d \bar{d}_t(f), \tag{10}$$

The highest-scoring frontier is then selected as the next local navigation goal.

When local Search terminates due to limited remaining budget or sufficient negative evidence, SPM-Nav may still preserve the current Space for inter-region navigation. If an existing Portal leads to an unvisited or still relevant downstream Space, the agent switches to Transit and navigates toward the corresponding Portal anchor, avoiding further local inspection while preserving connectivity to downstream regions. If no suitable transition is available, the agent performs Backtrack through previously established Portals and historical trajectories to resume exploration from an earlier Space. Backtrack is used only for navigation recovery rather than as an independent search behavior.

4 Experiments

Datasets and Evaluation Protocol. SPM-Nav is evaluated on HM3D-v1 Ramakrishnan et al. (2021), HM3D-v2 Yadav et al. (2023), and MP3D Chang et al. (2017) in Habitat Savva et al. (2019). HM3D-v1 contains 2,000 validation episodes from 20 scenes with six target categories, while HM3D-v2 contains 1,000 episodes from 36 scenes using the same target categories. MP3D

Table 1: Comparison with prior methods on HM3D-v1, HM3D-v2, and MP3D. Best and second-best results among training-free methods are highlighted in bold and underline, respectively.

Method	Training-free	Venue	Foundation Models		HM3D-v1		HM3D-v2		MP3D	
			Vision	Language	SR↑	SPL↑	SR↑	SPL↑	SR↑	SPL↑
Uni-NaVid Zhang et al. (2025a)	×	RSS 2025	–	–	–	–	73.7	37.1	–	–
ImagineNav Zhao et al. (2025)	×	ICLR 2025	PolyOculus	GPT-4o-mini	53.0	23.8	–	–	–	–
MTU3D Zhu et al. (2025)	×	ICCV 2025	–	–	–	–	–	–	–	–
CompassNav Li et al. (2026)	×	ICLR 2026	–	–	–	–	61.6	27.8	42.0	17.5
OVSegDT Zemskova et al. (2026)	×	CVPR 2026	–	–	–	–	–	–	–	–
TrajRAG Wang et al. (2026)	×	CVPR 2026	–	–	–	–	78.1	40.2	42.6	18.0
CoW Gadre et al. (2023)	✓	CVPR 2023	–	–	–	–	–	–	9.2	4.9
L3MVN Yu et al. (2023)	✓	IROS 2023	–	RoBERTa-large	50.4	23.1	36.3	15.7	34.9	14.5
VLFM Yokoyama et al. (2024)	✓	ICRA 2024	BLIP-2	–	52.5	30.4	63.6	32.5	36.4	17.5
SG-Nav Yin et al. (2024)	✓	NeurIPS 2024	LLaVA-1.6-7B	GPT-4	54.0	24.9	49.6	25.5	40.2	16.0
InstructNav Long et al. (2024)	✓	CoRL 2024	GPT-4V	GPT-4	58.0	20.9	58.0	20.9	–	–
BeliefMapNav Zhou et al. (2025)	✓	NeurIPS 2025	–	–	–	–	–	–	37.3	17.6
ApexNav Zhang et al. (2025b)	✓	IEEE RA-L 2025	BLIP-2	DeepSeek-V3	59.6	33.0	<u>76.2</u>	<u>38.0</u>	39.2	<u>17.8</u>
WMNav Nie et al. (2025)	✓	IROS 2025	–	–	58.1	31.2	72.1	33.3	45.4	17.2
TopoNav Liu et al. (2025)	✓	arXiv 2025	GLEE	GPT-4o	60.1	<u>34.6</u>	–	–	<u>45.5</u>	16.8
MSGNav Huang et al. (2026)	✓	CVPR 2026	GPT-4o	GPT-4o	63.0	31.4	74.1	33.4	–	–
ASCENT Gong et al. (2026)	✓	IEEE RA-L 2026	BLIP-2	Qwen2.5-7B	<u>65.4</u>	33.5	–	–	44.5	15.5
EvoMemNav Ge et al. (2026)	✓	arXiv 2026	YOLOv8-World + SAM	Qwen3-VL-8B	59.2	33.6	63.8	**39.4**	–	–
SPM-Nav (Ours)	✓	–	BLIP-2	–	**70.6**	**34.8**	**83.8**	37.6	**55.7**	**24.7**

Table 2: Core-component ablation on HM3D-v2. *w/o Persistent Memory* retains the Space–Portal structure but removes the historical state stored in each Space.

Variant	PSM	Portal	ASS	SR↑	SPL↑	Avg. Steps↓	ΔSPL [95% CI]	ΔSteps [95% CI]
w/o Persistent Memory	×	✓	✓	82.1	37.2	231.57	$-0.40\,[-1.11, +0.29]$	$+13.40\,[+9.44, +17.48]$
w/o Portals	✓	×	✓	79.1	35.6	234.19	$\mathbf{-2.00}\,[\mathbf{-2.79}, \mathbf{-1.23}]$	$\mathbf{+16.02}\,[\mathbf{+11.29}, \mathbf{+20.84}]$
w/o Adaptive Search	✓	✓	×	81.8	36.4	226.17	$-1.20\,[-2.49, +0.07]$	$\mathbf{+8.00}\,[\mathbf{+0.82}, \mathbf{+14.99}]$
Full SPM-Nav	✓	✓	✓	**83.8**	**37.6**	**218.17**	–	–

contains 2,195 validation episodes from 11 scenes with 21 target categories. SPM-Nav performs online navigation using only pretrained models, without task-specific training or fine-tuning on the evaluation environments. The maximum episode length is set to 500 steps, and an episode is considered successful when the agent issues the STOP action within 1.0 m of the target.

Evaluation Metrics. Success Rate (SR) and Success weighted by Path Length (SPL) are used as the primary evaluation metrics Anderson et al. (2018). For the core ablations, the average number of executed actions (Avg. Steps) is additionally reported, while SoftSPL is included in the connectivity analysis. SR, SPL, and SoftSPL are reported as percentages (%). For the core ablations, all variants are evaluated on the same episodes using paired comparisons. The reported ΔSPL and ΔSteps are computed as $M_{\text{ablated}} - M_{\text{full}}$, and 95% confidence intervals are estimated using 10,000 episode-level paired bootstrap resamples. Paired differences in SR are evaluated using the exact McNemar test. All remaining mechanism analyses are conducted on a fixed 50-episode diagnostic subset.

Implementation Details. Places365 is used to estimate room semantics, RAM provides contextual semantic cues, D-FINE and Grounding DINO are used for object detection, RedNet provides semantic segmentation, and BLIP-2 is used for visual–language target verification. Visual perception is combined with online geometric mapping, frontier extraction, and a pretrained PointNav policy to update the Space–Portal memory and execute navigation. SPM-Nav does not use a language foundation model; semantic guidance beyond visual perception is limited to an object–room prior. All experiments are conducted on a single NVIDIA RTX 4090 GPU. Additional implementation details and search settings are provided in Appendix A.

4.1 Comparison with Prior Methods

Table 1 compares SPM-Nav with prior training-based and training-free methods on three ObjectNav benchmarks. TopoNav Liu et al. (2025) evaluates HM3D using the 2,000-episode, 20-scene split; we therefore report its results under HM3D-v1, while leaving HM3D-v2 unreported. EvoMemNav Ge et al. (2026) reports results on both standard HM3D splits but does not report MP3D results. ImagineNav Zhao et al. (2025) evaluates on the 20-scene HM3D validation split, so we report its results under HM3D-v1; it does not report HM3D-v2 or MP3D navigation results. Although ImagineNav performs zero-shot open-vocabulary navigation on HM3D, we mark it as not training-free because its Where2Imagine ResNet-18 module is trained from scratch on Habitat-Web demonstrations. SPM-Nav achieves SRs of 70.6%, 83.8%, and 55.7% on HM3D-v1, HM3D-v2, and MP3D, respec-

tively. It further obtains an SPL of 34.8% on HM3D-v1 and 24.7% on MP3D, showing consistent navigation performance across different environments and target-category settings. On HM3D-v2, SPM-Nav achieves an SR of 83.8%, outperforming all training-free baselines listed in the table. Its SPL of 37.6% is slightly below the 39.4% of EvoMemNav and the 38.0% of ApexNav. Therefore, the main advantage of SPM-Nav on HM3D-v2 lies in its substantially higher success rate while retaining competitive path efficiency. Paired qualitative comparisons with ASCENT on identical navigation episodes are provided in Appendix E.2.

4.2 Ablation Studies

Core-component ablation on HM3D-v2. We independently remove Persistent Space Memory, Inter-Space Portals, and Adaptive Space Search to examine the contribution of each component. For *w/o Persistent Memory*, the Space states and the structural information required for Portal construction are retained, while the historical state accumulated across visits is removed. This setting isolates the effect of persistent Space-level search memory without simultaneously disrupting the inter-Space topology. As shown in Table 2, the full SPM-Nav achieves 83.8% SR and 37.6% SPL while requiring the fewest actions on average. Removing Inter-Space Portals causes the largest overall degradation: SPL decreases by 2.00 percentage points and the average episode length increases by 16.02 steps, with both paired confidence intervals excluding zero. These results demonstrate the importance of persistent inter-Space connectivity for efficient navigation. Removing Persistent Space Memory results in a smaller SPL change of -0.40 percentage points, but increases the average episode length by 13.40 steps, indicating that persistent search states primarily help reduce redundant exploration. Removing Adaptive Space Search increases the average episode length by 8.00 steps, while the point estimate of SPL decreases by 1.20 percentage points. Overall, Space-level memory, Portal connectivity, and adaptive search provide complementary benefits through historical-state reuse, persistent inter-region connectivity, and adaptive allocation of search effort, respectively.

4.3 Mechanism Analysis

4.3.1 Space–Portal Memory

Persistence under Revisits. The analysis first examines whether Persistent Space Memory enables previously accumulated search experience to be preserved and reused when a Space is revisited. The evaluation is conducted on the fixed 50-episode HM3D-v1 diagnostic subset. To isolate the effect of persistent search states, *w/o Persistent Memory* retains Space identity and the Portal connectivity structure, while clearing the previously accumulated search state upon each revisit. *State Retained* counts revisits for which the historical search state remains available. *Revisit Steps* measures the average number of additional local-search steps executed after a revisit, and *Reactivation Rate* measures the fraction of revisits that trigger renewed local search. As shown in Table 3, SPM-Nav retains the historical search state in all 65 Space revisits and performs no additional local search after revisiting. In contrast, removing Persistent Memory results in an average of 2.7 additional search steps per revisit, while 36.76% of revisits reactivate local inspection. These results show that persistent Space states allow previously accumulated search progress to be directly reused, thereby reducing redundant inspection of already explored regions.

Table 3: Persistent-memory behavior under Space revisits on the fixed 50-episode HM3D-v1 diagnostic subset.

Metric	SPM-Nav	w/o Persistent Memory
State Retained ↑	**65/65**	0/68
Revisit Steps ↓	**0.000**	2.7
Reactivation Rate ↓	**0.00%**	36.76%

Inter-Space Portal Analysis. The analysis next examines whether Inter-Space Portals provide reusable connectivity across Spaces. The *w/o Portals* variant still allows the agent to move between regions using the current map and frontiers, but does not retain persistent inter-Space transition memory. *Reuse Events* counts repeated traversals of previously used Portals, while *Reuse Rate* denotes the fraction of confirmed Portals that are reused at least once. *Backtracks* and *Backtrack Steps* measure the number of navigation-recovery events and their execution cost, respectively.

Table 4: Portal reuse and recovery on the fixed 50-episode HM3D-v1 diagnostic subset.

Metric	SPM-Nav	w/o Portals
Confirmed Portals	53	N/A
Portal Traversals	112	N/A
Reuse Events	59	N/A
Reused Portals	19	N/A
Reuse Rate	35.85%	N/A
Backtracks ↓	**10**	18
Backtrack Steps ↓	**175**	332

As shown in Table 4, SPM-Nav establishes 53 confirmed inter-Space Portals over the fixed 50 episodes. Among them, 19 are reused, corresponding to a reuse rate of 35.85%, with 59 reuse events in total. When Portal memory is removed, the number of Backtracks increases from 10 to 18, while Backtrack Steps increase from 175 to 332. These results indicate that Portals not only preserve inter-Space connectivity but also enable previously established transitions to be reused during subsequent navigation and recovery, reducing the cost of repeated cross-Space transitions. Episode-level Portal statistics are provided in Appendix D.

4.3.2 Memory-Conditioned Adaptive Search

Adaptive Space Search adjusts local search intensity according to Space relevance. Visited Spaces are grouped into LOW-, MEDIUM-, and HIGH-relevance categories, and the number of local-search steps executed within each group is compared together with the corresponding Peek and Probe behaviors. As shown in Fig. 3, SPM-Nav spends an average of 0.52, 10.93, and 31.00 local-search steps in LOW-, MEDIUM-, and HIGH-relevance Spaces, respectively, exhibiting a clear relevance-dependent allocation of search effort. Peek and Probe are also concentrated primarily in MEDIUM- and HIGH-relevance Spaces. After removing Adaptive Space Search, the corresponding average search efforts become 7.90, 7.93, and 12.60 steps, showing substantially weaker variation across relevance levels. These results indicate that Adaptive Space Search uses the accumulated Space state to regulate local search effort, reducing unnecessary inspection in low-relevance regions while allocating more search resources to more relevant Spaces. Additional analysis of connectivity-preserving inter-Space Transit is provided in Appendix C.

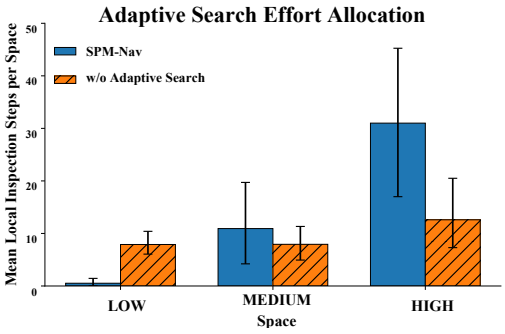

Figure 3: Local search effort across Space-relevance levels.

Qualitative trajectory visualizations illustrating the evolution and reuse of the Space–Portal memory, relevance-conditioned search allocation, and representative failure modes are provided in Appendix E.

5 Conclusion

SPM-Nav is a training-free object navigation framework that decouples transient frontier candidates from persistent region-level search states through a persistent Space–Portal Memory. Spaces aggregate regional semantics, target evidence, and search progress, while Portals maintain reusable transitions between Spaces. Built upon this memory, Adaptive Space Search allocates local search effort according to the accumulated regional state while preserving the connectivity required to reach downstream regions. Experiments on three ObjectNav benchmarks, together with component ablations and mechanism analyses, show that the proposed design reduces redundant exploration, enables the reuse of inter-Space connectivity, and improves the allocation of search effort. Overall, SPM-Nav provides a framework that shifts zero-shot object navigation from transient frontier-level decisions toward persistent region-level search memory.

AI USE STATEMENT

We used OpenAI's ChatGPT/Codex to assist with Chinese-to-English translation and software implementation. We additionally used these tools to edit portions of the manuscript for clarity and concision, locate and summarize relevant literature, format LaTeX and references, and check the numerical and cross-reference consistency of the manuscript. The research method and experimental protocol were determined by the authors. Generative AI was not used to generate, select, or process experimental data, and all reported measurements were obtained from experiments conducted by the authors. All AI-assisted text and code were manually reviewed, revised, and verified by the authors. Citations were checked against the original sources, and AI-assisted code was validated through author-run experiments and tests. The authors take full responsibility for the final manuscript, code, claims, and results.

REPRODUCIBILITY STATEMENT

Section 4 specifies the datasets, evaluation protocol, and metrics. Appendix A documents the implementation and search settings, including the open-source codebase on which SPM-Nav builds. Appendix B describes the ablation configurations and statistical analysis. Additional diagnostic results are provided in Appendices C and D.

REFERENCES

Dong An, Hanqing Wang, Wenguan Wang, Zun Wang, Yan Huang, Keji He, and Liang Wang. ETPNav: Evolving topological planning for vision-language navigation in continuous environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7):5130–5145, 2025. doi: 10.1109/TPAMI.2024.3386695.

Peter Anderson, Angel Chang, Devendra Singh Chaplot, Alexey Dosovitskiy, Saurabh Gupta, Vladlen Koltun, Jana Kosecka, Jitendra Malik, Roozbeh Mottaghi, Manolis Savva, and Amir R. Zamir. On evaluation of embodied navigation agents. *arXiv preprint arXiv:1807.06757*, 2018.

Dhruv Batra, Aaron Gokaslan, Aniruddha Kembhavi, Oleksandr Maksymets, Roozbeh Mottaghi, Manolis Savva, Alexander Toshev, and Erik Wijmans. Objectnav revisited: On evaluation of embodied agents navigating to objects. *arXiv preprint arXiv:2006.13171*, 2020.

Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Nießner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3d: Learning from RGB-D data in indoor environments. In *International Conference on 3D Vision*, pp. 667–676, 2017. doi: 10.1109/3DV.2017.00081.

Devendra Singh Chaplot, Dhiraj Gandhi, Abhinav Gupta, and Ruslan Salakhutdinov. Object goal navigation using goal-oriented semantic exploration. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, 2020.

Shizhe Chen, Pierre-Louis Guhur, Makarand Tapaswi, Cordelia Schmid, and Ivan Laptev. Think global, act local: Dual-scale graph transformer for vision-and-language navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16537–16547, 2022.

Samir Yitzhak Gadre, Mitchell Wortsman, Gabriel Ilharco, Ludwig Schmidt, and Shuran Song. Cows on pasture: Baselines and benchmarks for language-driven zero-shot object navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23171–23181, 2023.

Zuhao Ge, Xiaosong Jia, Chao Wu, Yuchen Zhou, Zuxuan Wu, and Yu-Gang Jiang. EvoMemNav: Efficient self-evolving fine-grained memory for zero-shot embodied navigation. *arXiv preprint arXiv:2606.03509*, 2026. URL https://arxiv.org/abs/2606.03509.

Zeying Gong, Rong Li, Tianshuai Hu, Ronghe Qiu, Lingdong Kong, Lingfeng Zhang, Guoyang Zhao, Yiyi Ding, and Junwei Liang. Stairway to success: An online floor-aware zero-shot object-goal navigation framework via LLM-driven coarse-to-fine exploration. *IEEE Robotics and Automation Letters*, 11(3):2943–2950, 2026. doi: 10.1109/LRA.2026.3655265.

Zizhan He, Maxime Daigle, and Pouya Bashivan. Building spatial world models from sparse transitional episodic memories. In *International Conference on Learning Representations*, 2026.

João P. Hespanha, Daniel Liberzon, and A. Stephen Morse. Hysteresis-based switching algorithms for supervisory control of uncertain systems. *Automatica*, 39(2):263–272, 2003. doi: 10.1016/S0005-1098(02)00241-8.

Xun Huang, Shijia Zhao, Yunxiang Wang, Xin Lu, Wanfa Zhang, Rongsheng Qu, Weixin Li, Yunhong Wang, and Chenglu Wen. MSGNav: Unleashing the power of multi-modal 3d scene graph for zero-shot embodied navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 37154–37163, 2026.

Nuri Kim, Obin Kwon, Hwiyeon Yoo, Yunho Choi, Jeongho Park, and Songhwai Oh. Topological semantic graph memory for image-goal navigation. In *Proceedings of the Conference on Robot Learning (CoRL)*, volume 205 of *Proceedings of Machine Learning Research*, pp. 393–402, 2023.

Yuxuan Kuang, Hai Lin, and Meng Jiang. OpenFMNav: Towards open-set zero-shot object navigation via vision–language foundation models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=l1eI6NrXYX.

LinFeng Li, Jian Zhao, Yuan Xie, Xin Tan, and Xuelong Li. CompassNav: Steering from path imitation to decision understanding in navigation. In *International Conference on Learning Representations*, 2026.

Peiran Liu, Qiang Zhang, Daojie Peng, Lingfeng Zhang, Yihao Qin, Hang Zhou, Jun Ma, Renjing Xu, and Yiding Ji. TopoNav: Topological graphs as a key enabler for advanced object navigation. *arXiv preprint arXiv:2509.01364*, 2025. URL https://arxiv.org/abs/2509.01364.

Yuxing Long, Wenzhe Cai, Hongcheng Wang, Guanqi Zhan, and Hao Dong. InstructNav: Zero-shot system for generic instruction navigation in unexplored environment. In *Proceedings of the Conference on Robot Learning*, volume 270, pp. 2049–2060, 2024.

Dujun Nie, Xianda Guo, Yiqun Duan, Ruijun Zhang, and Long Chen. WMNav: Integrating vision–language models into world models for object goal navigation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2392–2399, 2025. doi: 10.1109/IROS60139.2025.11246684.

Dekang Qi, Shuang Zeng, Xinyuan Chang, Feng Xiong, Shichao Xie, Xiaolong Wu, and Mu Xu. MerNav: A highly generalizable memory–execute–review framework for zero-shot object goal navigation. *arXiv preprint arXiv:2602.05467*, 2026.

Santhosh Kumar Ramakrishnan, Aaron Gokaslan, Erik Wijmans, Oleksandr Maksymets, Alexander Clegg, John M. Turner, Eric Undersander, Wojciech Galuba, Andrew Westbury, Angel X. Chang, Manolis Savva, Yili Zhao, and Dhruv Batra. Habitat–matterport 3d dataset (HM3D): 1000 large-scale 3d environments for embodied AI. *arXiv preprint arXiv:2109.08238*, 2021.

Shouwei Ruan, Liyuan Wang, Caixin Kang, Qihui Zhu, Songming Liu, Xingxing Wei, and Hang Su. Brain-inspired spatial intelligence for embodied agents. *Nature Communications*, 17(1):8061, 2026. doi: 10.1038/s41467-026-74358-5.

Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, Devi Parikh, and Dhruv Batra. Habitat: A platform for embodied AI research. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9339–9347, 2019. doi: 10.1109/ICCV.2019.00943.

Yiyao Wang, Sixian Zhang, Keming Zhang, Xinhang Song, Songjie Du, and Shuqiang Jiang. TrajRAG: Retrieving geometric-semantic experience for zero-shot object navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15166–15176, 2026.

Karmesh Yadav, Ram Ramrakhya, Santhosh Kumar Ramakrishnan, Théo Gervet, John Turner, Aaron Gokaslan, Noah Maestre, Angel X. Chang, Dhruv Batra, Manolis Savva, Alexander Clegg, and Devendra Singh Chaplot. Habitat–matterport 3d semantics dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4927–4936, 2023.

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

Hang Yin, Xiuwei Xu, Zhenyu Wu, Jie Zhou, and Jiwen Lu. SG-Nav: Online 3d scene graph prompting for LLM-based zero-shot object navigation. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-0171.

Naoki Yokoyama, Sehoon Ha, Dhruv Batra, Jiuguang Wang, and Bernadette Bucher. VLFM: Vision-language frontier maps for zero-shot semantic navigation. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2024. doi: 10.1109/ICRA57147.2024.10610712.

Bangguo Yu, Hamidreza Kasaei, and Ming Cao. L3MVN: Leveraging large language models for visual target navigation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3554–3560, 2023. doi: 10.1109/IROS55552.2023.10342512.

Tatiana Zemskova, Aleksei Staroverov, Dmitry Yudin, and Aleksandr Panov. OVSegDT: Segmenting transformer for open-vocabulary object goal navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8120–8129, 2026.

Jiazhao Zhang, Kunyu Wang, Shaoan Wang, Minghan Li, Haoran Liu, Songlin Wei, Zhongyuan Wang, Zhizheng Zhang, and He Wang. Uni-NaVid: A video-based vision–language–action model for unifying embodied navigation tasks. In *Proceedings of Robotics: Science and Systems*, 2025a. doi: 10.15607/RSS.2025.XXI.013.

Mingjie Zhang, Yuheng Du, Chengkai Wu, Jinni Zhou, Zhenchao Qi, Jun Ma, and Boyu Zhou. ApexNav: An adaptive exploration strategy for zero-shot object navigation with target-centric semantic fusion. *IEEE Robotics and Automation Letters*, 10(11):11530–11537, 2025b. doi: 10.1109/LRA.2025.3606388.

Xinxin Zhao, Wenzhe Cai, Likun Tang, and Teng Wang. ImagineNav: Prompting vision–language models as embodied navigator through scene imagination. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=vQFw9ryKyK.

Kaiwen Zhou, Kaizhi Zheng, Connor Pryor, Yilin Shen, Hongxia Jin, Lise Getoor, and Xin Eric Wang. ESC: Exploration with soft commonsense constraints for zero-shot object navigation. In *International Conference on Machine Learning*, pp. 42829–42842, 2023.

Zibo Zhou, Yue Hu, Lingkai Zhang, Zonglin Li, and Siheng Chen. BeliefMapNav: 3d voxel-based belief map for zero-shot object navigation. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-2782.

Ziyu Zhu, Xilin Wang, Yixuan Li, Zhuofan Zhang, Xiaojian Ma, Yixin Chen, Baoxiong Jia, Wei Liang, Qian Yu, Zhidong Deng, Siyuan Huang, and Qing Li. Move to understand a 3d scene: Bridging visual grounding and exploration for efficient and versatile embodied navigation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8120–8132, 2025.

A Implementation Details and Search Settings

Codebase. Our implementation builds on the open-source codebase of ASCENT Gong et al. (2026). We extend its navigation and perception pipeline with Persistent Space–Portal Memory and Adaptive Space Search. The LLM-based exploration component of ASCENT is not used in our experiments.

Perception and navigation. SPM-Nav processes RGB–D observations and agent poses online. Places365 estimates room semantics; RAM extracts contextual semantic cues; D-FINE and Grounding DINO provide object detection; RedNet provides semantic segmentation; and BLIP-2 supports visual–language target verification. Navigation goals are executed by a pretrained DD-PPO PointNav policy. The exploration policy does not invoke a language model.

Space formation and stabilization. When evidence of an inter-region transition is encountered, SPM-Nav initializes a provisional destination Space using the current robot position or the stable destination-side trajectory obtained during transition probing. Each Space maintains an entry reference, a semantic anchor, and a running centroid of its supporting observations. The entry reference remains fixed for Portal geometry and transition reasoning, whereas the semantic anchor is selected from actually visited positions inside the region and may be refined as additional evidence is accumulated. Selecting an observed position rather than a geometric mean prevents the representative anchor from falling inside obstacles or non-traversable areas.

A provisional Space is stabilized only after sufficient spatial and semantic support has been accumulated. In the default configuration, stabilization typically requires at least three spatially distinct observations, a trajectory span of at least 0.60 m, observations extending approximately 0.80 m beyond the entry region, a maximum Jensen–Shannon divergence of 0.15 among room-semantic distributions, and a dominant-room probability of at least 0.35. After six supporting positions, sufficient geometric coverage can stabilize a Space when room semantics remain ambiguous; a relaxed condition is used for small regions whose transitions have already been confirmed.

Before introducing a new Space, the system searches for an existing Space on the same floor using its entry reference, semantic anchor, observation positions, and keyframe locations, with a spatial association radius of approximately 0.40 m. Room semantics are used only to rank spatially plausible candidates and cannot merge spatially separated Spaces by themselves. If subsequent probing indicates that a candidate transition does not separate two regions, the provisional destination Space is collapsed back into the source Space rather than being retained as an independent topological node.

Target verification. A target candidate is checked across multiple frames and viewpoints before visual-language verification.

Search parameters. Table 5 lists the principal settings. The Space-relevance update uses the current observation with weight $\lambda = 0.30$ and the previous estimate with weight 0.70. This coefficient is distinct from the temporal weight used to fuse candidate room distributions.

Negative evidence. Table 5 summarizes the thresholds for counting independent negative views. A view is counted only when its target-presence score is below 0.20, its valid-depth fraction is at least 0.30, and the agent is not Backtracking. Local search may stop after 3, 5, or 6 counted views for LOW-, MEDIUM-, or HIGH-relevance Spaces, respectively; the HIGH condition additionally requires observations from at least two positions separated by 0.50 m.

B Ablation and Statistical Details

Ablation configurations. The *w/o Persistent Memory* variant retains Space identities, geometry, and Portal topology but does not restore accumulated search state upon a Space revisit. The *w/o Portals* variant removes persistent inter-Space transition memory while retaining navigation through the current map and frontiers. The *w/o Adaptive Search* variant disables adaptive budgets and progressive Peek/Probe inspection, using a uniform 30-step local-search budget.

Table 5: Implementation and search settings.

Setting	Value
Candidate-relevance weights (w_r, w_o, w_v, w_h)	$(0.55, 0.25, 0.10, 0.10)$
Space-relevance EWMA weight λ	0.30
Target–room prior top-K / threshold	3 / 0.15
Space-relevance boundaries	0.30 / 0.60
Local-search budgets (LOW/MEDIUM/HIGH)	0 / 12 / 30 steps
Peek viewing offsets	$-30°, 0°, +30°$
Probe depth	0.50 m
Independent-view separation	0.50 m or 25°
Negative-evidence target-presence threshold	0.20
Minimum valid-depth fraction	0.30
Negative-view stopping counts (LOW/MEDIUM/HIGH)	3 / 5 / 6
Maximum exit-discovery turns	4
Episode length / success distance	500 steps / 1.0 m

Paired analysis. HM3D-v2 ablations are compared with SPM-Nav episode by episode. Confidence intervals for differences in SPL and steps are computed from 10,000 paired bootstrap resamples, and success differences are assessed using the exact McNemar test. In the Portal ablation, 62 episodes succeed only with SPM-Nav, compared with 26 that succeed only without Portals ($p = 0.0004$).

C Connectivity-Preserving Transit

A Space with limited remaining value for local inspection may still connect the agent to unexplored regions. To isolate this role, the *w/o Connectivity Preservation* variant disables exit-oriented traversal after local inspection is deemed unwarranted, while retaining Space–Portal memory and the other search mechanisms. Transit here denotes an exit-oriented inter-Space movement, executed by the same low-level navigation policy as other goals.

On the fixed 50-episode HM3D-v1 diagnostic subset, 73.42% of exit-oriented traversals in SPM-Nav reach another Space; the rate is 73.33% when restricted to LOW-relevance source Spaces. The corresponding rates without connectivity preservation are 2.13% and 2.38%. SPM-Nav reaches 36 distinct downstream Spaces, compared with 1 for the ablation. Figure 4 visualizes this difference.

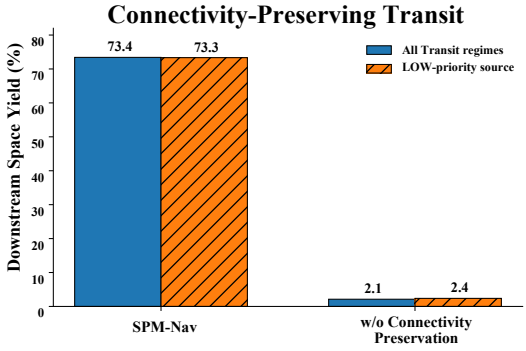

Figure 4: Downstream Space reachability with and without connectivity-preserving Transit on the fixed 50-episode HM3D-v1 subset.

D Portal Distribution Across Episodes

Table 6 reports confirmed inter-Space Portals per episode on the fixed 50-episode HM3D-v1 subset. The count excludes stair Portals. Of the 22 episodes without a confirmed inter-Space Portal, 4 terminate before a valid Space is established and 18 visit only one Space. Every episode that visits multiple Spaces records at least one confirmed inter-Space Portal.

Table 6: Distribution of confirmed inter-Space Portals across the fixed 50-episode HM3D-v1 diagnostic subset.

Portals per episode	Episodes	Fraction
0	22	44%
1	14	28%
2	7	14%
3	4	8%
4	2	4%
5	1	2%
Total	50	100%

E Qualitative Analysis of Navigation Trajectories

E.1 Mechanism-Focused Analysis

Figures 5(a)–(c) illustrate the three core mechanisms of SPM-Nav: Portal-based inter-Space connectivity, persistent Space states across backtracking, and relevance-conditioned Adaptive Search. In Fig. 5(a), the discovered Space–Portal structure expands from a single Space to two Spaces connected by a validated Portal, forming $S_0 \rightarrow P_0 \rightarrow S_1$ and illustrating persistent inter-Space connectivity. In Fig. 5(b), the agent searches in S_1 and subsequently backtracks toward S_0, while the previously established S_0–P_0–S_1 structure remains available, illustrating persistence across revisits and recovery. Figure 5(c) shows the effect of Adaptive Search: a LOW-relevance region receives no local search budget, whereas a HIGH-relevance bedroom is assigned substantially more search effort before the target is found.

Figures 5(d)–(f) further illustrate long-horizon topology accumulation and representative failure modes. In Fig. 5(d), the discovered Space–Portal structure grows from a single Space to a multi-region structure with five Spaces and seven Portals, showing progressive accumulation of regional connectivity over a long episode. The relatively low SPL also indicates that a more extensive Space–Portal structure does not necessarily imply efficient navigation. In Fig. 5(e), four high-scoring target candidates are rejected, preventing premature stopping but revealing remaining challenges in semantic ambiguity and confidence calibration. In Fig. 5(f), the agent explores multiple floors and establishes two Spaces and three Portals but never observes the target, reflecting the combined difficulty of floor-level coverage, target visibility, and incomplete regional exploration. These examples provide qualitative illustrations of the proposed mechanisms, while their overall effects are evaluated quantitatively in the ablation and mechanism analyses.

E.2 Paired Comparison with ASCENT

Figures 6 and 7 present a paired qualitative comparison between ASCENT and SPM-Nav on three selected episodes with identical scene, episode, and target specifications. The examples are intended to illustrate representative navigation behaviors rather than to provide statistical evidence. Figure 6 shows the failure trajectories of ASCENT, while Fig. 7 shows the corresponding successful trajectories of SPM-Nav on the same tasks.

Episodes 32 and 33 mainly highlight recovery from ambiguous semantic evidence. In Episode 32, ASCENT exhausts the episode budget and terminates with a false-positive failure, whereas SPM-Nav continues region-level exploration and eventually verifies the bed. In Episode 33, ASCENT exhibits repeated local exploration and again ends with a false-positive failure, while SPM-Nav reaches a higher-relevance region through continued cross-floor search and successfully verifies the TV. These two examples qualitatively suggest that persistent Space–Portal state together with continued target verification can prevent an isolated ambiguous observation from irreversibly dominating the subsequent search process.

Episode 77 highlights a different advantage. ASCENT extends exploration across floors without ever observing the toilet and ultimately exhausts the 500-step budget. In contrast, SPM-Nav succeeds in 49 steps with an SPL of 89%. As the agent moves into increasingly relevant regions, Adaptive

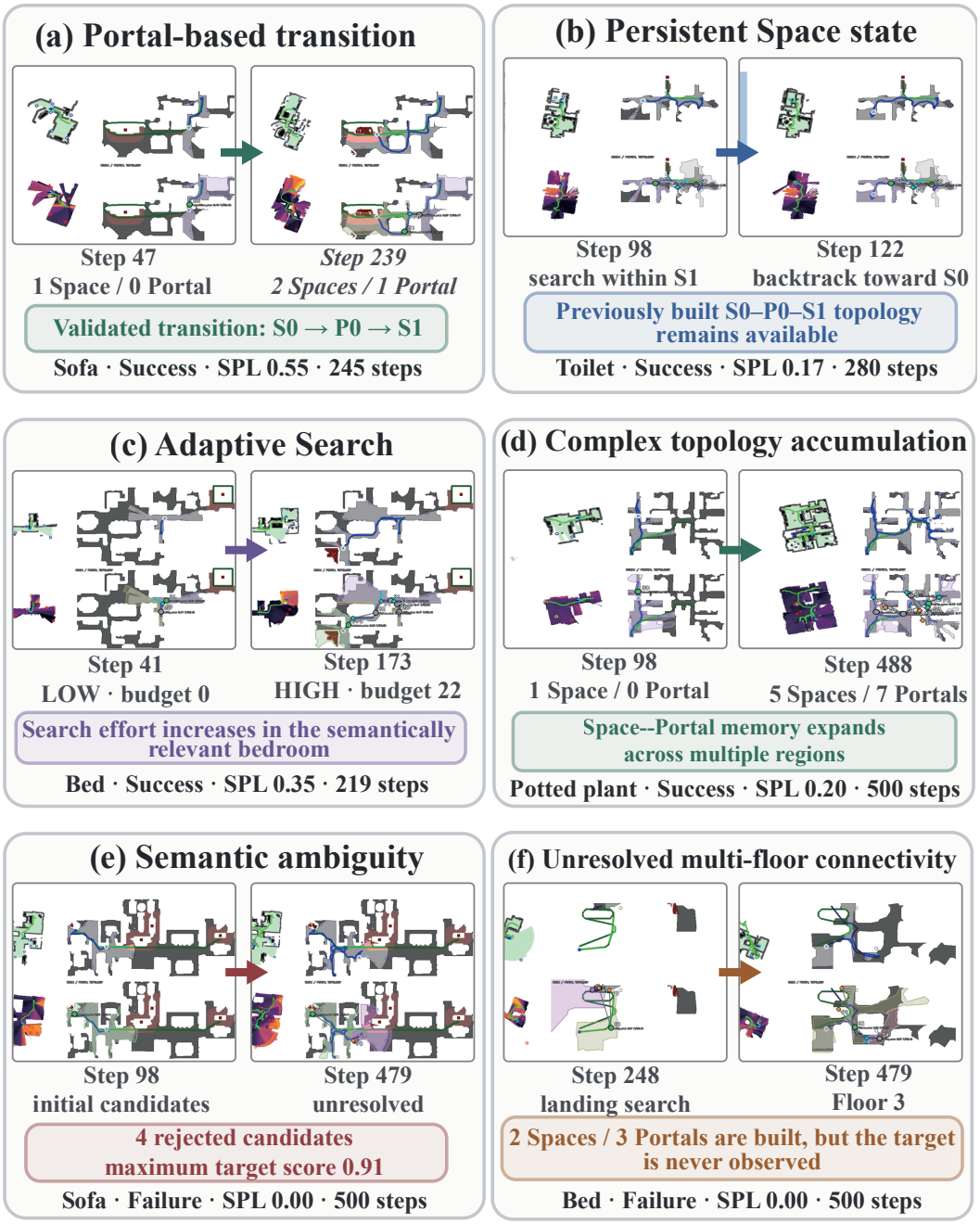

Figure 5: **Qualitative analysis of SPM-Nav navigation trajectories.** (a)–(c) illustrate Portal-based inter-Space connectivity, persistent Space states during backtracking, and relevance-conditioned Adaptive Search. (d)–(f) show long-horizon Space–Portal accumulation and representative failure cases involving semantic ambiguity and incomplete multi-floor exploration. Space and Portal counts denote discovered memory entities rather than traversal counts.

Search raises the local search budget and concentrates search effort where the target is more likely to appear. This example therefore provides a direct qualitative illustration of relevance-conditioned search allocation. Overall, these cases complement the aggregate quantitative results by visualizing how persistent regional state, continued verification, and adaptive budget allocation affect navigation behavior in individual episodes.



Figure 6: **ASCENT trajectories on three selected paired episodes.** Episodes 32 and 33 terminate with recorded false-positive failures, while Episode 77 exhausts the episode budget without observing the target. The corresponding SPM-Nav runs are shown in Fig. 7.



Figure 7: **SPM-Nav trajectories on the same paired episodes as Fig. 6.** SPM-Nav successfully verifies the target in all three selected cases. Episodes 32–33 illustrate continued exploration after ambiguous observations, while Episode 77 highlights relevance-conditioned Adaptive Search. Space and Portal counts denote discovered memory entities rather than traversal counts.