All Roads Lead to Rome: Flow-driven Multi-Anchor Exploration for Open-Environment Active 3D Mapping

Yang Li[1] **Aming Wu**[2]* **Zihao Zhang**[1] **Ziju Han**[3] **Sijia Zhang**[1] **Yahong Han**[1]*

[1]School of Artificial Intelligence, Tianjin University, China
[2]School of Computer Science and Information Engineering, Hefei University of Technology, China
[3]School of Cybersecurity, Tianjin University, China

{liyang1389, zhangzihao2490, hanziju, 3024244296, yahong}@tju.edu.cn
amwu@hfut.edu.cn

Abstract

To advance the development of embodied intelligence, Open-Environment Active 3D Mapping has attracted increasing attention, aiming to perform a long-horizon and shortest trajectory exploration for reconstructing unseen scenarios. Since only limited information about unseen environments is available, methods built on the closed-set assumption, i.e., assuming that the test environments are similar to those seen during training, cannot generalize satisfactorily. In existing active mapping methods, long-horizon exploration is often guided by predicting a coarse long-range goal and then converting it into an executable path. However, this stage is usually formulated as single-point prediction. Under partial observability, the same local observation may correspond to multiple plausible exploration directions, making such deterministic prediction prone to brittle decisions and degraded performance in unseen scenarios. Our experiments further verify that this is a key factor underlying their weak generalization. To address this issue, we reformulate long-horizon target prediction as conditional multimodal anchor generation using Conditional Flow Matching. Instead of predicting a single goal, our method learns a conditional distribution over coarse exploration anchors from the current mapping state. These anchors are first converted into executable candidate paths through obstacle-aware planning. We then apply exploration-mode clustering to compress geometrically similar trajectories and reduce candidate redundancy. Finally, a hierarchical selection module selects the most promising mode and reranks paths within it to produce the final executable trajectory. Experiments show that our method improves generalization and reconstruction efficiency in open environments.

1 Introduction

Active 3D Mapping is a core problem in robotics and digital-twin applications, where an embodied agent must actively move, acquire observations, and reconstruct the geometry of an unknown scene as efficiently as possible. Unlike passive reconstruction [1, 2], active mapping [3, 4] requires not only accurate surface reconstruction but also effective exploration decisions under a limited sensing budget. Thus, the key challenge lies not only in *how to reconstruct*, but also in *how to plan*. Early works [5–7] are typically built on the closed-set assumption, i.e., assuming that the test environments are similar to those seen during training. However, in open environments, only limited information about unseen scenes is available at test time, which makes such methods difficult to generalize satisfactorily.

*Corresponding authors: Aming Wu and Yahong Han.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

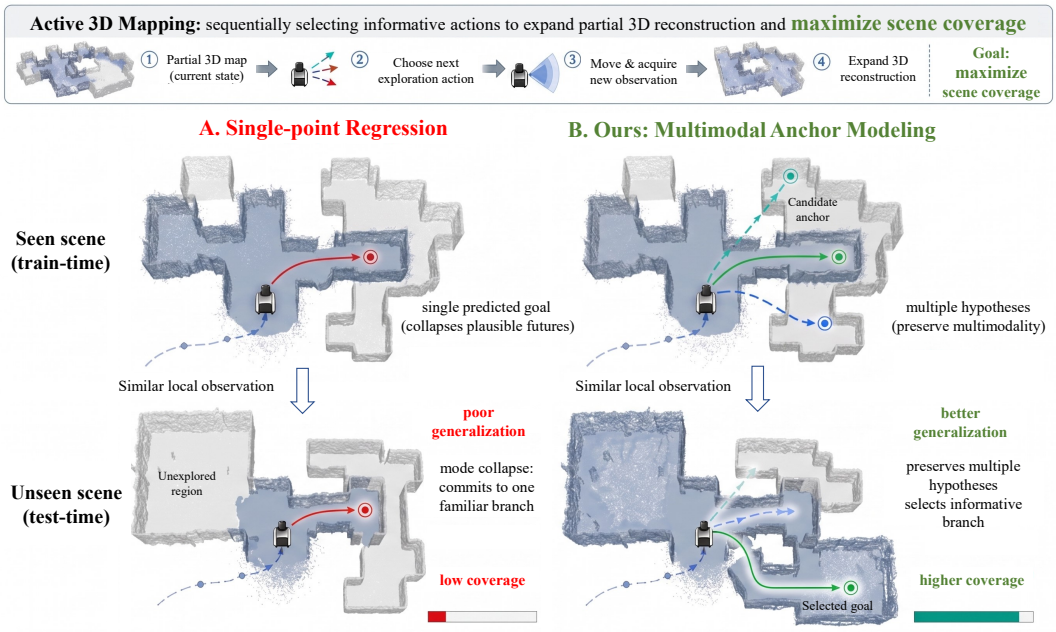

Figure 1: **Motivation.** Under partial observability, a similar local observation may support multiple plausible long-horizon exploration directions. Single-goal prediction collapses these futures into one dominant branch and generalizes poorly when the unseen scene continuation changes. In contrast, multimodal anchor modeling preserves alternative hypotheses before downstream planning, leading to better generalization and higher coverage.

Recently, open-environment benchmarks [8, 9] have extended active 3D mapping to cross-scene and cross-difficulty settings, placing much stronger demands on generalization. In these settings, long-horizon exploration is often guided by first predicting a coarse long-range goal. However, existing methods on these benchmarks [8, 9] still largely formulate this stage as single-point prediction. This formulation is fundamentally limited under partial observability, where the same local observation may admit multiple plausible long-horizon exploration directions. As shown in Fig. 1, single-point regression collapses multiple plausible long-horizon futures into one dominant goal. When the unseen scene continuation changes at test time, it tends to commit to a familiar branch, leading to brittle decisions and weak generalization.

This observation motivates us to move beyond single-goal prediction and instead model long-horizon exploration with multiple plausible goal hypotheses. Recent progress in generative modeling [10, 11] suggests that conditional generation is particularly effective for representing multimodal futures. In particular, Conditional Flow Matching (CFM) [10] learns a continuous transport from a simple prior to a conditional target distribution, making it both stable and efficient for multimodal modeling. However, existing flow-based models in embodied decision-making tasks [12–15] are primarily designed for high-dimensional trajectory or action generation, whereas active 3D mapping fundamentally requires modeling long-horizon exploration intent under partial observability rather than directly synthesizing full behaviors.

To this end, we reformulate long-horizon target prediction in active 3D mapping as a conditional multimodal proposal generation problem, and adapt Conditional Flow Matching (CFM) to this setting in a task-specific manner so as to model multiple plausible long-horizon exploration hypotheses in a low-dimensional anchor space. The sampled anchors are then converted into executable candidate paths through obstacle-aware planning, compressed into exploration modes by anchor-guided clustering, and finally resolved by a hierarchical selection module to produce the final trajectory. Beyond the architectural design, we provide theoretical analysis showing why deterministic single-goal prediction is intrinsically mismatched to multimodal long-horizon exploration under partial observability, and why CFM offers a principled way to learn the resulting conditional anchor distribution. Extensive experiments, including within-dataset difficulty shift and cross-dataset transfer, demonstrate that the proposed framework achieves more robust generalization, higher reconstruction quality, and more efficient exploration in open environments.

Our contributions are threefold. ***First***, we identify deterministic single-goal prediction as a key limitation of open-environment active 3D mapping under partial observability, and reformulate long-horizon exploration as conditional multimodal proposal generation. ***Second***, we propose a flow-driven multi-anchor exploration framework that learns diverse long-horizon exploration anchors with Conditional Flow Matching, followed by obstacle-aware planning, exploration-mode clustering, and hierarchical path selection. ***Third***, experiments on AiMDoom and Matterport3D under within-dataset difficulty shift and cross-dataset transfer demonstrate consistent improvements in generalization, reconstruction quality, and exploration efficiency.

2 Active 3D Mapping with Open-Environment Generalization

Let $\mathcal{S}$ denote an unknown scene, and let $\mathcal{X}^{\mathrm{GT}} \subset \mathbb{R}^3$ denote its ground-truth surface. Active 3D mapping seeks to control an agent to acquire observations and reconstruct the scene as completely as possible under a finite budget T. At each time step t, the agent receives an RGB-D observation I_t and predicts the next pose $c_t = (c_t^{\mathrm{pos}}, c_t^{\mathrm{rot}})$ based on the history $\mathcal{H}_t = \{(I_0, c_0), \ldots, (I_t, c_t)\}$. This defines a partially observable sequential decision problem, where the high-level action satisfies $a_t \sim \pi(\cdot \mid \mathcal{H}_t)$. In this work, a_t is instantiated as a long-horizon exploration anchor, and executable candidate paths are recovered from sampled anchors using an explicit obstacle-aware planner.

Beyond within-distribution mapping performance, we focus on generalization to unseen scenes in open environments. Let $p_{\mathrm{tr}}(\mathcal{S})$ and $p_{\mathrm{te}}(\mathcal{S})$ denote the training and test scene distributions, where $p_{\mathrm{tr}}(\mathcal{S}) \neq p_{\mathrm{te}}(\mathcal{S})$ is allowed. Given the reconstructed point cloud $\mathcal{P}_t$, we measure mapping quality by the surface coverage $\mathrm{Cov}(\mathcal{P}_t, \mathcal{X}^{\mathrm{GT}})$, and use $\mathrm{AUC}(\pi_\theta) = \frac{1}{T}\sum_{t=1}^{T} \mathrm{Cov}(\mathcal{P}_t, \mathcal{X}^{\mathrm{GT}})$ to quantify intermediate mapping efficiency. The overall objective is

$$\max_{\theta} \; \mathbb{E}_{\mathcal{S}\sim p_{\mathrm{te}}(\mathcal{S})}\big[\lambda_1 \,\mathrm{Cov}(\mathcal{P}_T, \mathcal{X}^{\mathrm{GT}}) + \lambda_2 \,\mathrm{AUC}(\pi_\theta)\big]. \tag{1}$$

3 Method

3.1 Overview

Active mapping in open environments exhibits two key challenges. First, the same local observation may admit multiple plausible long-horizon exploration directions under partial observability. Second, after geometric planning, these hypotheses often produce many similar candidates, which increases redundancy in downstream decision-making. To address this, we decompose long-horizon exploration into a hierarchical process. We first represent the current mapping progress by a 2D projected state. Specifically, let $\mathcal{P}_t$ denote the accumulated reconstructed point cloud at time t. We partition $\mathcal{P}_t$ into K horizontal slices and project them into K local density maps, while the trajectory history is projected into a visitation map, yielding $E_t = \{I_{t,1}^{pc}, \ldots, I_{t,K}^{pc}, I_t^{\mathrm{hist}}\}$. The shared context feature is then obtained by $z_t = f_\phi(E_t)$, where f_ϕ is a mapping-progress encoder. Based on z_t, we predict a local obstacle map $\hat{O}_t = g_\psi(z_t)$, which is used for subsequent feasible path planning. Built on the context feature z_t, our method proceeds in four stages, as illustrated in Fig. 3:

$$E_t \xrightarrow{\text{Encoder}} z_t \xrightarrow{\text{CFM}} \mathcal{A}_t \xrightarrow{\text{Planning+Clustering}} \{\mathcal{M}_t^{(m)}\}_{m=1}^{M_t} \xrightarrow{\text{Selection}} \tau_t^\star. \tag{2}$$

Conditional Flow Matching (CFM) first generates a multimodal anchor set $\mathcal{A}_t$ for long-horizon exploration. The anchors are then converted into candidate paths, clustered into exploration modes $\{\mathcal{M}_t^{(m)}\}_{m=1}^{M_t}$, and hierarchically selected to obtain the final path $\tau_t^\star$.

3.2 Conditional Flow Matching for Multimodal Anchors

The purpose of this subsection is two-fold. First, we show why deterministic single-goal prediction is intrinsically mismatched to multimodal long-horizon exploration under partial observability. Second, based on this observation, we show why learning a conditional distribution over coarse anchors naturally leads to a CFM-based formulation in a low-dimensional decision space.

Revisiting the learning target. In open-environment active 3D mapping, the same local observation may correspond to multiple plausible long-horizon exploration directions. Let $p(a \mid z_t)$ denote the conditional distribution of feasible long-horizon exploration anchors under the current context z_t. Suppose one instead learns a deterministic predictor $\hat{a}(z_t)$ under the standard squared loss:

$$\mathcal{R}(\hat{a}; z_t) = \mathbb{E}_{a\sim p(a|z_t)}\left[\|\hat{a}(z_t) - a\|_2^2\right]. \tag{3}$$

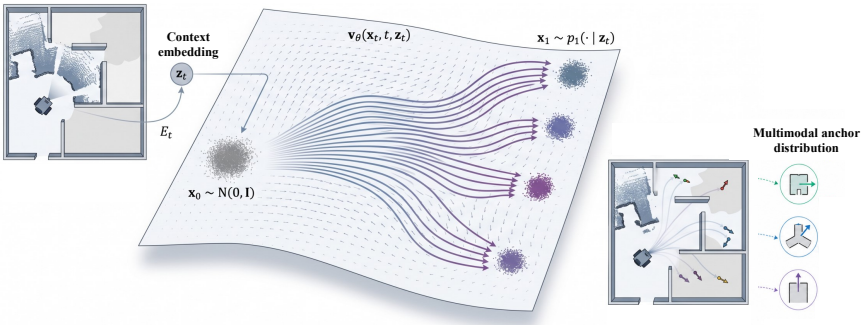

Figure 2: **Conditional Flow Matching for multimodal long-horizon anchors.** CFM transports samples from a simple prior $x_0 \sim \mathcal{N}(0, I)$ toward the target anchor distribution $x_1 \sim p_1(\cdot \mid z_t)$, yielding multiple plausible long-horizon exploration hypotheses under the same partial observation. Unlike directly generating full trajectories, this formulation models coarse exploration intent in a compact space and remains naturally compatible with downstream planning.

Proposition 3.1 (Deterministic single-goal collapse)**.** *The Bayes-optimal deterministic predictor of Eq. 3 is the conditional mean*

$$\hat{a}^\star(z_t) = \mathbb{E}[a \mid z_t]. \tag{4}$$

Moreover, when $p(a \mid z_t)$ is multimodal and its feasible support is nonconvex due to scene geometry, $\hat{a}^\star(z_t)$ may lie outside any high-density mode of $p(a \mid z_t)$ and may even become infeasible for downstream planning.

Proof sketch. Eq. 3 is a standard conditional least-squares objective, whose minimizer is the conditional mean in Eq. 4. When the anchor distribution contains several separated feasible modes, however, this mean generally lies between them. Under obstacle-induced nonconvex geometry, such an averaged anchor can fall into a blocked or low-probability region, making it neither representative nor executable.

Proposition 3.1 formalizes the mismatch of single-goal prediction: under multimodal long-horizon futures, deterministic regression is mean-seeking rather than mode-preserving. Therefore, the correct learning target is not a single long-horizon goal, but the full conditional anchor distribution $p(a \mid z_t)$.

A low-dimensional conditional anchor space. A way to preserve multimodality would be to generate full trajectories. However, the role of the long-horizon proposal is to express *where to explore next* at a coarse semantic level, while detailed geometric feasibility is naturally handled by the explicit planner using the obstacle map $\hat{O}_t$. This suggests separating high-level long-horizon intent from low-level path realization, and introducing a compact variable for decision making.

We therefore define the long-horizon exploration anchor as

$$a = (a_{xy}, a_\theta) \in \mathbb{R}^4, \qquad a_{xy} = (x_a, y_a), \quad a_\theta = (\sin\theta_a, \cos\theta_a), \tag{5}$$

where a_{xy} and a_θ represent the position and orientation of the long-horizon goal, respectively. This low-dimensional parameterization isolates long-horizon exploration intent from detailed path geometry, leaving fine-grained execution to the downstream planner.

Conditional transport via flow matching. Having established that the desired learning target is the conditional anchor distribution $p(a \mid z_t)$, we next need a principled way to model it. Conditional Flow Matching (CFM) is particularly suitable here, because it directly learns a conditional transport from a simple prior to the target distribution in continuous space, as illustrated in Fig. 2.

During training, we set the source distribution as a standard Gaussian and the target samples as oracle anchors:

$$x_0 \sim p_0 = \mathcal{N}(0, I), \qquad x_1 \sim p_1(\cdot \mid z_t), \tag{6}$$

where x_1 is extracted from anchors in a high-reward path bank. Using the linear interpolation path

$$x_t = (1-t)x_0 + t x_1, \qquad t \sim \mathcal{U}(0, 1), \tag{7}$$

the target velocity is

$$u_t = x_1 - x_0. \tag{8}$$

We then train a conditional velocity field $v_\theta(x_t, t, z_t)$ by minimizing

$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E}_{x_0, x_1, t}\left[\|v_\theta(x_t, t, z_t) - u_t\|_2^2\right]. \tag{9}$$

The key point is that Eq. 9 is not merely a heuristic diversity objective; it is a conditional regression problem whose population optimum recovers the transport field of the anchor distribution.

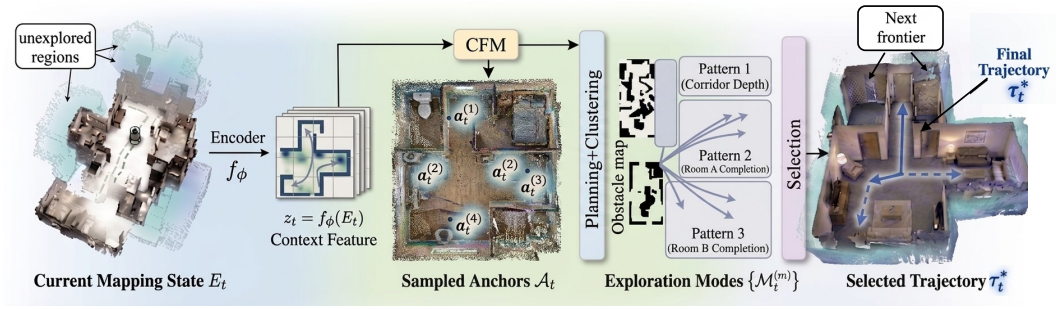

Figure 3: **Overview of our method.** Given the current mapping state E_t, a mapping-progress encoder produces the context feature z_t. Conditional Flow Matching (CFM) then generates a multimodal set of long-horizon exploration anchors $\mathcal{A}_t$, which are converted into executable candidate paths and clustered into exploration modes $\{\mathcal{M}_t^{(m)}\}_{m=1}^{M_t}$. Finally, a hierarchical selection module first selects the most promising mode and then reranks paths within that mode to output the final trajectory $\tau_t^\star$.

Proposition 3.2 (CFM learns the conditional anchor transport field)**.** *Let $p_t(\cdot \mid z_t)$ be the path induced by Eq. 7 between p_0 and $p_1(\cdot \mid z_t)$. Then the population minimizer of Eq. 9 is*

$$v^\star(x_t, t, z_t) = \mathbb{E}[u_t \mid x_t, t, z_t]. \tag{10}$$

Thus, CFM learns the conditional transport field induced by the entire anchor distribution $p_1(\cdot \mid z_t)$, rather than regressing a single deterministic long-horizon target.

Proof sketch. Eq. 9 is a standard conditional least-squares objective. By the orthogonal decomposition of regression risk,

$$\mathbb{E}\big[\|v - u_t\|_2^2\big] = \mathbb{E}\big[\|v - \mathbb{E}[u_t \mid x_t, t, z_t]\|_2^2\big] + \mathbb{E}\big[\|u_t - \mathbb{E}[u_t \mid x_t, t, z_t]\|_2^2\big] .$$

The second term is independent of v, so the minimum is achieved by Eq. 10.

Proposition 3.2 shows why CFM is a principled choice for our setting: once the learning target is reformulated from a deterministic goal to a conditional anchor distribution, CFM directly learns its conditional transport in the low-dimensional anchor space. Compared with directly generating full trajectories, this formulation is easier to optimize, preserves multimodal long-horizon hypotheses, and remains naturally compatible with explicit planning.

During inference, we sample K anchors from the learned conditional prior:

$$\mathcal{A}_t = \{a_t^{(1)}, \ldots, a_t^{(K)}\}, \qquad a_t^{(k)} \sim p_\theta(a \mid z_t). \tag{11}$$

These samples constitute the multimodal long-horizon exploration hypotheses at the current time step. To encourage coverage over distinct long-horizon exploration modes, we further introduce a diversity regularization term:

$$\mathcal{L}_{\text{div}} = \frac{1}{K(K-1)} \sum_{i \neq j} \exp\left(-\frac{\|a_t^{(i)} - a_t^{(j)}\|_2^2}{\sigma^2} \right). \tag{12}$$

Minimizing this term encourages different anchors to remain sufficiently dispersed, thereby covering diverse long-horizon exploration modes. In this way, the proposed CFM module provides a set of diverse, context-consistent exploration hypotheses for downstream planning and selection.

3.3 Planning and Anchor-Guided Mode Construction.

The previous subsection motivates modeling long-horizon exploration as a conditional anchor distribution rather than a single deterministic goal. However, anchors are only coarse intents and cannot be executed directly, and different anchors may still produce highly similar paths after planning. We therefore recover anchors into executable candidate paths via obstacle-aware planning, and further cluster them into representative exploration modes to reduce redundancy before hierarchical selection.

From Anchors to Executable Paths. Given the current pose c_t, anchor $a_t^{(k)}$, and obstacle map $\hat{O}_t$, we obtain the candidate path via an explicit planner:

$$\tau_t^{(k)} = \mathit{Plan}(c_t, a_t^{(k)}, \hat{O}_t), \tag{13}$$

where *Plan* is implemented as an A*-based [16] shortest path planner. To characterize each candidate path for downstream reasoning, we further extract a compact path descriptor

$$\varphi(\tau_t^{(k)}) = [g_k,\ r_k,\ c_k,\ n_k,\ h_k,\ u_k], \tag{14}$$

which respectively characterize the gain proxy, risk, cost, novelty, continuity deviation, and uncertainty statistics. They serve as structured intermediate features for the learned hierarchical selector in Sec. 3.4.

From Candidate Paths to Exploration Modes. If selection is performed directly path-by-path on $\{\tau_t^{(k)}\}_{k=1}^{K}$, redundant candidates will interfere with decision-making. To circumvent this, we first discover more abstract exploration modes over the path set. Denoting the candidate path set as $\mathcal{T}_t = \{\tau_t^{(1)}, \ldots, \tau_t^{(K)}\}$, we define the distance between any two paths τ_i, τ_j in the path space as:

$$d(\tau_i, \tau_j) = \lambda_1 d_{\text{end}}(\tau_i, \tau_j) + \lambda_2 d_{\text{shape}}(\tau_i, \tau_j) + \lambda_3 d_{\text{prefix}}(\tau_i, \tau_j) + \lambda_4 d_{\text{topo}}(\tau_i, \tau_j), \tag{15}$$

where d_{end} measures the endpoint discrepancy, d_{shape} measures the geometric shape discrepancy of the paths after resampling, d_{prefix} measures the prefix trajectory discrepancy to characterize short-term execution behaviors, and d_{topo} measures the topological difference between the two paths in terms of obstacle circumvention.

Under the distance metric, we cluster the candidate paths into M_t exploration modes:

$$\mathcal{T}_t \longrightarrow \{\mathcal{M}_t^{(1)}, \ldots, \mathcal{M}_t^{(M_t)}\}. \tag{16}$$

Each mode $\mathcal{M}_t^{(m)}$ corresponds to a class of long-horizon exploration hypotheses, such as "*prioritize entering the left room*", "*continue deep along the main corridor*", or "*complete the current local region first before expanding farther*". Furthermore, we define a representative path $\tilde{\tau}_t^{(m)} \in \mathcal{M}_t^{(m)}$ for each mode, which is implemented as the medoid within that mode.

Simultaneously, we define mode statistics:

$$\chi_t^{(m)} = \left[|\mathcal{M}_t^{(m)}|,\ \bar{g}_m,\ \bar{r}_m,\ \mathrm{Var}(g)_m,\ \mathrm{Var}(r)_m,\ \rho_m\right], \tag{17}$$

where $|\mathcal{M}_t^{(m)}|$ denotes the mode size, $\bar{g}_m$ and $\bar{r}_m$ denote the average gain and average risk within the mode, $\mathrm{Var}(g)_m$ and $\mathrm{Var}(r)_m$ measure the dispersion of gain and risk, and ρ_m denotes the mode radius. Together with the representative path, these statistics provide a compact summary of the mode's quality and internal consistency. As with the path descriptors above, they are used as auxiliary structured inputs to the downstream learned scorer, rather than as standalone selection criteria.

Therefore, this module elevates multimodal anchor hypotheses into executable, comparable, and hierarchically structured exploration mode representations.

3.4 Hierarchical Mode Selection

The previous subsection organizes multimodal anchors into a set of exploration modes, but mode construction alone does not resolve the final decision. In open-environment active 3D mapping, the agent's goal is to make robust long-horizon exploration decisions under partial observability and scene shift, rather than to score many redundant paths independently. If selection is still performed directly over all candidate paths, geometrically similar trajectories can interfere with scoring and bias the policy toward unstable local choices, which weakens generalization and reduces exploration efficiency. We therefore adopt a hierarchical selection strategy: first selecting a promising mode at the strategy level, and then reranking paths within that mode to obtain the final executable trajectory.

Mode-Level Scoring. For each mode, we construct a mode representation:

$$z_t^{(m)} = \Gamma(z_t, \tilde{\tau}_t^{(m)}, \chi_t^{(m)}), \tag{18}$$

where z_t is the shared context output by the context encoder, and $\Gamma(\cdot)$ represents the mode encoder used to fuse the shared context, representative path, and mode statistics.

In open environments, a mode worth executing should not only possess a high expected gain but also simultaneously satisfy the criteria of controllable risk, low uncertainty, and continuity with existing execution plans, while preserving a degree of exploration diversity when necessary. To this end, we predict multi-head scores via a mode scorer:

$$(G_m, R_m, U_m, C_m, D_m) = s_\eta(z_t^{(m)}), \tag{19}$$

where G_m, R_m, U_m, C_m, and D_m represent the expected gain, execution risk, uncertainty, mode switching or cost, and exploration diversity reward of the mode, respectively. Here, $s_\eta(\cdot)$ represents a learnable mode-level scoring network. Accordingly, the overall mode-level score is defined as:

$$S_m^{\text{mode}} = w_G G_m - w_R R_m - w_U U_m - w_C C_m + w_D D_m, \tag{20}$$

where $w_G, w_R, w_U, w_C, w_D \geq 0$ are learnable weights. Thus, the optimal mode selection at the current time step is:

$$m_t^\star = \arg\max_m S_m^{\text{mode}}. \tag{21}$$

Intra-Mode Local Reranking. Within the selected mode $\mathcal{M}_t^{(m_t^\star)}$, for any candidate path τ, we construct a path-level representation:

$$\zeta_t(\tau) = \Lambda(z_t, \tau, \varphi(\tau), \chi_t^{(m_t^\star)}), \tag{22}$$

where $\Lambda(\cdot)$ represents the feature fusion function of the local reranker. Subsequently, the local reranker outputs four path-level scoring heads $\hat{G}(\tau), \hat{R}(\tau), \hat{C}(\tau), \hat{H}(\tau)$, which respectively characterize path gain, path risk, path cost, and continuity penalty. Consequently, the overall local path score is defined as:

$$S^{\text{local}}(\tau) = \hat{G}(\tau) - \lambda_R \hat{R}(\tau) - \lambda_C \hat{C}(\tau) - \lambda_H \hat{H}(\tau). \tag{23}$$

The final execution path is thus given by:

$$\tau_t^\star = \arg \max_{\tau \in \mathcal{M}_t^{(m_t^\star)}} S^{\text{local}}(\tau). \tag{24}$$

Thus, this section decouples long-range strategy reasoning from local path execution via hierarchical selection, improving stability and robustness in open-environment exploration.

4 Experiments

4.1 Experimental Setup

We evaluate our method on Matterport3D (MP3D) [17] and AiMDoom [8]. For AiMDoom, we adopt a 70/30 train/test split within each difficulty level. On AiMDoom, we report two standard active-mapping metrics: *Final Coverage*, which measures the final scene coverage at the end of an episode, and *AUC*, which evaluates reconstruction efficiency by the area under the coverage curve over time. Surface coverage is computed against the ground-truth mesh following [8]. We evaluate five trajectories per scene using identical random initial poses for different methods, and report the mean and standard deviation over all testing trajectories. For MP3D, we additionally report the completion metrics used in prior work, namely *Comp. (%)* and *Comp. (cm)*.

We follow the same mapping-state construction and basic training protocol as [8] for fair comparison. Our model takes projected local map observations and trajectory history as input, and predicts an obstacle map for downstream proposal generation and planning.

4.2 Comparison with the State of the Art

AiMDoom dataset. Table 1 shows that our method achieves the best performance on all four AiMDoom difficulty levels. Compared with NBP [8], it consistently improves both Final Coverage and AUC, with larger gains on the harder splits. In particular, Final Coverage improves from 0.734 to 0.781 on Normal, from 0.618 to 0.662 on Hard, and from 0.472 to 0.534 on Insane, while AUC also improves consistently. This trend is particularly meaningful because AiMDoom is designed to increase scene ambiguity and long-horizon planning difficulty as the level becomes harder. The results therefore suggest that preserving multiple plausible long-horizon hypotheses makes exploration more robust under increasing uncertainty, rather than committing prematurely to a single predicted branch.

MP3D dataset. As shown in Table 2, our method achieves the best results on MP3D, reaching 82.67 in Comp. (%) and 5.67 in Comp. (cm). Compared with NBP [8], this corresponds to a gain of 3.29 points in completion ratio and a reduction of 1.11 cm in completion error. The improvement in both metrics is important: the higher completion ratio indicates better scene coverage, while the lower completion error shows that the reconstructed geometry is also more accurate. These results suggest that the multimodal long-horizon modeling is not only effective on synthetic open-environment benchmarks, but also transfers well to realistic indoor scenes with more complex layouts.

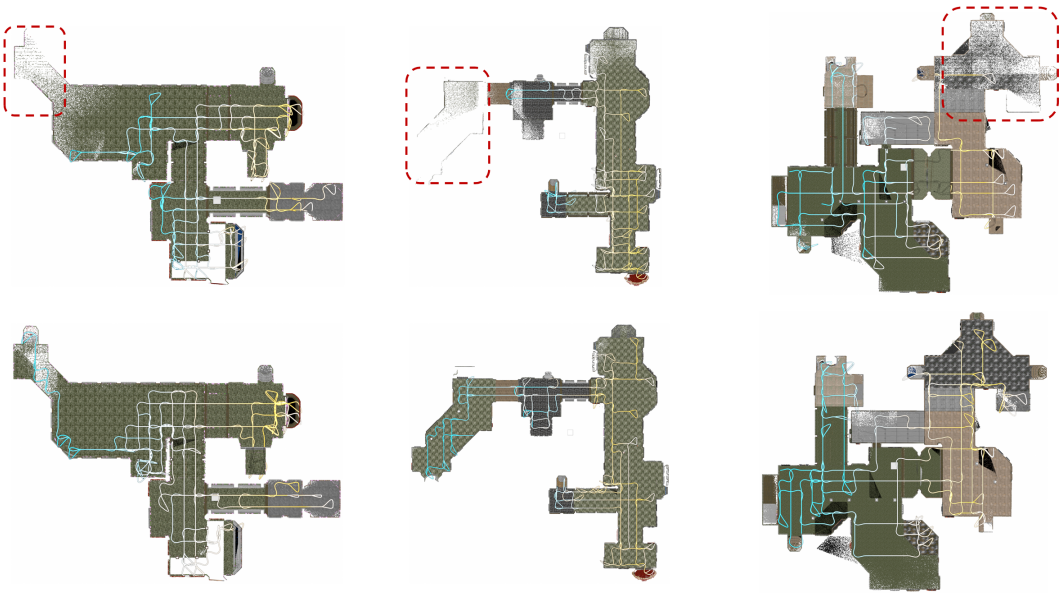

Figure 4: **Qualitative comparison between the SOTA method NBP (Top) and Ours (Bottom).** For each scene, both methods start from the same initial pose. The dashed boxes highlight representative regions that remain unexplored by NBP but are successfully recovered by our method. This difference is not merely local: NBP tends to commit early to a single dominant long-horizon branch, which induces a narrower global coverage pattern and leaves weakly indicated but still informative continuations unexplored. In contrast, our method preserves multiple plausible long-horizon exploration hypotheses before downstream planning and selection, enabling more balanced expansion across alternative scene continuations and ultimately yielding more complete scene coverage.

4.3 Generalization under Out-of-Distribution Settings

We evaluate OOD generalization from two perspectives: *within-dataset difficulty shift* on AiMDoom and *cross-dataset transfer* from AiMDoom to MP3D. The results consistently support our central claim: preserving multiple long-horizon exploration hypotheses improves generalization, whereas single-goal methods generalize poorly under scene shift.

Table 1: **Evaluation results on AiMDoom Dataset.** For each difficulty level, all baseline models, including ours, are trained from scratch on the corresponding training set to ensure a fair comparison.

	Simple		**Normal**		**Hard**		**Insane**	
	Final Cov.	AUCs	Final Cov.	AUCs	Final Cov.	AUCs	Final Cov.	AUCs
Random	0.323±0.156	0.270±0.135	0.190±0.124	0.152±0.103	0.124±0.082	0.088±0.060	0.074±0.048	0.050±0.035
FBE [18]	0.760±0.174	0.605±0.171	0.565±0.139	0.415±0.109	0.425±0.114	0.311±0.080	0.330±0.097	0.239±0.079
SCONE [19]	0.577±0.173	0.483±0.138	0.412±0.114	0.313±0.087	0.290±0.093	0.210±0.072	0.196±0.079	0.140±0.060
MACARONS [7]	0.599±0.200	0.479±0.172	0.418±0.120	0.314±0.088	0.302±0.097	0.218±0.070	0.192±0.078	0.139±0.058
NBP [8]	0.879±0.142	0.692±0.156	0.734±0.142	0.526±0.112	0.618±0.153	0.432±0.115	0.472±0.095	0.312±0.073
Ours	**0.913**±0.138	**0.707**±0.141	**0.781**±0.125	**0.554**±0.096	**0.662**±0.139	**0.446**±0.112	**0.534**±0.073	**0.325**±0.089

Table 2: **Comparison on the MP3D dataset.**

Method	**Comp. (%) ↑**	**Comp. (cm) ↓**
Random	45.67	26.53
FBE [18]	71.18	9.78
UPEN [3]	69.06	10.60
OccAnt [6]	71.72	9.40
ANM [4]	73.15	9.11
NBP [8]	79.38	6.78
Ours	**82.67**	**5.67**

Table 3: **Train on AiMDoom test on MP3D.**

Method	**Comp. (%) ↑**	**Comp. (cm) ↓**
Random	32.12	29.36
FBE [18]	56.73	14.45
UPEN [3]	54.70	13.09
OccAnt [6]	58.49	14.18
ANM [4]	60.22	11.78
NBP [8]	61.81	10.22
Ours	**68.45**	**9.11**

Table 4: **Generalization under difficulty shift on AiMDoom.** We report two complementary protocols. Protocol A evaluates cross-difficulty transfer by training on the *Simple* split only and directly testing on all difficulty levels . Protocol B evaluates leave-one-difficulty-out generalization by training on *Simple* + *Normal* + *Hard* and directly testing on the held-out *Insane* split.

		Simple		Normal		Hard		Insane	
Protocol	**Method**	Final Cov.	AUC	Final Cov.	AUC	Final Cov.	AUC	Final Cov.	AUC
Protocol A: Train on Simple, test on all difficulty levels									
	Random	0.323±0.156	0.270±0.135	0.174±0.112	0.132±0.091	0.101±0.075	0.072±0.054	0.062±0.041	0.041±0.030
	FBE [18]	0.760±0.174	0.605±0.171	0.512±0.155	0.364±0.124	0.368±0.132	0.255±0.098	0.284±0.110	0.188±0.088
	SCONE [19]	0.577±0.173	0.483±0.138	0.354±0.128	0.245±0.104	0.213±0.111	0.142±0.082	0.145±0.092	0.094±0.071
	MACARONS [7]	0.599±0.200	0.479±0.172	0.368±0.141	0.252±0.115	0.225±0.124	0.148±0.087	0.151±0.096	0.098±0.075
	NBP [8]	0.879±0.142	0.692±0.156	0.695±0.158	0.502±0.128	0.562±0.161	0.401±0.122	0.434±0.114	0.275±0.098
	Ours	**0.913**±0.138	**0.707**±0.141	**0.727**±0.152	**0.526**±0.133	**0.597**±0.174	**0.403**±0.121	**0.467**±0.106	**0.313**±0.101
Protocol B: Train on Simple + Normal + Hard, test on held-out Insane									
	Random	–	–	–	–	–	–	0.068±0.045	0.048±0.038
	FBE [18]	–	–	–	–	–	–	0.312±0.124	0.201±0.105
	SCONE [19]	–	–	–	–	–	–	0.168±0.101	0.112±0.082
	MACARONS [7]	–	–	–	–	–	–	0.182±0.115	0.115±0.088
	NBP [8]	–	–	–	–	–	–	0.421±0.128	0.302±0.112
	Ours	–	–	–	–	–	–	**0.438**±0.119	**0.317**±0.089

Cross-dataset transfer from AiMDoom to MP3D. Table 3 reports zero-shot transfer from AiMDoom to MP3D. Despite the large domain gap, our method achieves the best performance, reaching 68.45 in Comp. (%) and 9.11 in Comp. (cm), improving over NBP by 6.64 points in completion ratio and 1.11 cm in completion error. This result suggests that multimodal long-horizon modeling transfers better across unseen scene distributions than single-goal prediction.

Within-dataset difficulty shift on AiMDoom. Table 4 reports two OOD protocols on AiMDoom. Under both settings, our method consistently outperforms NBP, with especially clear gains on the harder splits. This suggests that preserving multiple plausible long-horizon exploration paths is more robust to increasing scene ambiguity than single-goal prediction, which tends to commit early to one branch and thus generalizes poorly under difficulty shift.

4.4 Ablation Study

To verify the generalization of our method to unseen open-world scenes, all experiments in this subsection are conducted under a cross-dataset setting: *trained on AiMDoom and tested on MP3D*.

Component Analysis. Table 5 reports a progressive validation of the proposed pipeline, where MAG, EMC, and HMS denote multimodal anchor generation, exploration-mode clustering, and hierarchical mode/path selection, respectively. As modules are introduced stage by stage, the performance improves consistently under the cross-dataset OOD setting. The largest gain comes from adding MAG, which improves Comp. (%) from 61.81 to 66.21 and reduces Comp. (cm) from 10.22 to 9.64. This is consistent with our central claim that the main limitation of existing methods lies in deterministic single-goal prediction: replacing it with multimodal anchor generation already yields a much stronger long-horizon decision mechanism. Adding EMC further improves the results to 67.60 and 9.36, showing that organizing raw candidate paths into exploration modes is beneficial for reducing redundancy among geometrically similar proposals. Finally, adding HMS yields the best performance of 68.45 and 9.11, suggesting that hierarchical mode/path selection further stabilizes downstream decision making by decoupling high-level strategy choice from local path reranking.

5 Conclusion

In this paper, we study active 3D mapping in open environments, where partial observability and scene shift make deterministic single-goal prediction difficult to generalize. To address this issue, we reformulate long-horizon decision making as a conditional multimodal proposal generation problem and propose a flow-matching-based framework for multimodal anchor modeling. The generated anchors are fur-

Table 5: **Progressive validation of the proposed pipeline under the cross-dataset OOD setting.**

Method				OOD Setting	
Baseline	MAG	EMC	HMS	Comp. (%) ↑	Comp. (cm) ↓
✓				61.81	10.22
✓	✓			66.21	9.64
✓	✓	✓		67.60	9.36
✓	✓	✓	✓	**68.45**	**9.11**

ther converted into executable candidate paths, compressed into exploration modes, and selected through a hierarchical decision process. Extensive experiments show that our method consistently improves reconstruction quality, exploration efficiency, and robustness under both difficulty shift and cross-dataset transfer. These results suggest that preserving multimodal long-horizon exploration hypotheses is a promising direction for open-environment active mapping.

Limitations. The current framework has several limitations. Since our method models multimodal exploration in a coarse anchor space rather than directly reasoning over full future map completion, the utility of distant but weakly indicated regions may not always be captured well. In addition, the downstream planning and clustering stages depend on the quality of the predicted obstacle map and candidate-path structure, so errors in these intermediate representations can affect mode construction and final selection. Future work may benefit from stronger global uncertainty estimation, more topology-aware proposal generation, and tighter coupling between multimodal hypothesis generation and downstream planning.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (Grant No. 62636008, 62376186, 62472333, and 61932009).

References

[1] Joseph Ortiz, Alexander Clegg, Jing Dong, Edgar Sucar, David Novotny, Michael Zollhoefer, and Mustafa Mukadam. isdf: Real-time neural signed distance fields for robot perception. *arXiv preprint arXiv:2204.02296*, 2022.

[2] Edgar Sucar, Shikun Liu, Joseph Ortiz, and Andrew J Davison. imap: Implicit mapping and positioning in real-time. In *ICCV*, pages 6229–6238, 2021.

[3] Georgios Georgakis, Bernadette Bucher, Anton Arapin, Karl Schmeckpeper, Nikolai Matni, and Kostas Daniilidis. Uncertainty-driven planner for exploration and navigation. In *ICRA*, pages 11295–11302, 2022.

[4] Zike Yan, Haoxiang Yang, and Hongbin Zha. Active neural mapping. In *ICCV*, pages 10981–10992, 2023.

[5] Daryl Peralta, Joel Casimiro, Aldrin Michael Nilles, Justine Aletta Aguilar, Rowel Atienza, and Rhandley Cajote. Next-best view policy for 3d reconstruction. In *ECCV Workshops*, pages 558–573, 2020.

[6] Santhosh K Ramakrishnan, Ziad Al-Halah, and Kristen Grauman. Occupancy anticipation for efficient exploration and navigation. In *ECCV*, pages 400–418, 2020.

[7] Antoine Guédon, Tom Monnier, Pascal Monasse, and Vincent Lepetit. Macarons: Mapping and coverage anticipation with rgb online self-supervision. In *CVPR*, pages 940–951, 2023.

[8] Shiyao Li, Antoine Guedon, Clémentin Boittiaux, Shizhe Chen, and Vincent Lepetit. Nextbestpath: Efficient 3d mapping of unseen environments. In *ICLR*, 2025.

[9] Xiao Chen, Tai Wang, Quanyi Li, Tao Huang, Jiangmiao Pang, and Tianfan Xue. Gleam: Learning generalizable exploration policy for active mapping in complex 3d indoor scene. In *ICCV*, pages 5558–5568, 2025.

[10] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

[11] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[12] Jeongyong Yang, Seunghwan Jang, and SooJean Han. Safeflowmatcher: Safe and fast planning using flow matching with control barrier functions, 2026. URL `https://arxiv.org/abs/2509.24243`.

[13] Samiran Gode, Abhijeet Nayak, Débora N.P. Oliveira, Michael Krawez, Cordelia Schmid, and Wolfram Burgard. Flownav: Combining flow matching and depth priors for efficient navigation. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 17762–17768, 2025. doi: 10.1109/IROS60139.2025.11246634.

[14] Sen Wang, Le Wang, Sanping Zhou, Jingyi Tian, Jiayi Li, Haowen Sun, and Wei Tang. Flowram: Grounding flow matching policy with region-aware mamba framework for robotic manipulation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 12176–12186, 2025.

[15] Zebin Xing, Xingyu Zhang, Yang Hu, Bo Jiang, Tong He, Qian Zhang, Xiaoxiao Long, and Wei Yin. Goalflow: Goal-driven flow matching for multimodal trajectories generation in end-to-end autonomous driving. In *CVPR*, pages 1602–1611, 2025.

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

[17] 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. *arXiv preprint arXiv:1709.06158*, 2017.

[18] Brian Yamauchi. A frontier-based approach for autonomous exploration. In *CIRA*, pages 146–151, 1997.

[19] Antoine Guédon, Pascal Monasse, and Vincent Lepetit. Scone: Surface coverage optimization in unknown environments by volumetric integration. In *NeurIPS*, 2022.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We clarify the research contributions and scope of the study in the abstract and introduction of the paper.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations and future directions are discussed in the Conclusion.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [No]

Justification: The main text includes theoretical analysis in Sec. 3.2, where Proposition 3.1 and Proposition 3.2 justify the formulation of multimodal anchor modeling and the use of CFM. The main text provides proof sketches; complete proofs are omitted from this version.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [No]

   Justification: The extended network and hyperparameter details are not included in this version.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The code will be released upon acceptance. The datasets used derive from AimDooM and MP3d; they are already public and available online.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The details necessary to faithfully reproduce our experiments (training procedure, data splits, optimizer, etc.) are included in the paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We report average results of multiple runs in our experimental section. Our paper does not report error bars.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: This version does not provide detailed hardware and runtime information for each experiment.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We reviewed and ensured that the present work respects the NeurIPS Code of Ethics at each individual part.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: Our paper is not highly related to societal impacts.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This paper does not pose any such risk.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All the datasets used to train our models, and the code for the papers utilized as benchmarks to evaluate our algorithms are cited.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: We do not release any new assets with this submission. However, we will make the code as well as trained models publicly available if/when the paper is accepted.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This research did not involve any crowdsourcing experiments or studies with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This research did not involve studies with human subjects

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLM is not an important, original, or non-standard component of the core methods in this research.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.