ACTIVE GAUSSIAN SCENE RECONSTRUCTION VIA POLICY LEARNING ON HYBRID VOXEL MAPS

Anonymous authors
Paper under double-blind review

ABSTRACT

Active reconstruction requires an agent to decide both where to move and where to look to explore unknown regions and refine poorly reconstructed ones. However, existing methods often struggle to balance exploration and refinement, while greedy next view selection over jointly sampled positions and viewing directions overlooks future exploration gains. To address these issues, we present a reinforcement learning framework based on hybrid voxel maps that align exploration state and reconstruction uncertainty on a common voxel grid. The reconstruction uncertainty is estimated from the Fisher information of individual Gaussian primitives and aggregated onto the common voxel grid, allowing exploration and refinement gains to be evaluated in a common voxel space for position selection, view scoring, and policy rewards. We further decouple position selection from viewing direction selection, with position selection learned on the hybrid voxel maps to account for long term gains, and viewing direction selection performed by evaluating candidate directions based on their gains on the same maps at the selected position. This decomposition reduces the policy action space and simplifies policy learning. Experiments on indoor scenes show that our method achieves better reconstruction performance than existing active reconstruction methods under the same observation budget and reconstruction configuration.

1 INTRODUCTION

Active reconstruction involves deciding where an agent should move and where it should look to explore unknown regions and refine poorly reconstructed ones. Recent advances in Gaussian Splatting and dense RGB-D SLAM have made incremental high fidelity scene reconstruction increasingly practical. However, reconstruction quality still strongly depends on where new observations are acquired, making view planning an important component of active reconstruction.

Early active reconstruction methods built scene models from point clouds, surfels, or volumetric maps (Monica & Aleotti, 2018; Pito, 1999; Isler et al., 2016), with view selection often guided by information gain over unobserved space (Stachniss et al., 2005; Bircher et al., 2016). In Gaussian Splatting (Kerbl et al., 2023), primitives are progressively optimized from accumulated observations, yet observed regions may remain poorly reconstructed when available views provide insufficient constraints. View selection must therefore account for both unexplored regions and observed regions requiring further refinement.

Existing Gaussian based active reconstruction methods extend conventional information gain with reconstruction related measures such as Fisher information (Jiang et al., 2024; Xu et al., 2025), mutual information (Xie et al., 2025), entropy (Lee et al., 2022; Zhan et al., 2022), or coverage heuristics (Jin et al., 2025; Chen et al., 2025). However balancing exploration and refinement remains difficult because they involve distinct states and gains. Moreover, greedy next-view selection can undervalue actions whose benefits emerge through subsequent observations, while jointly sampling positions and viewing directions can require extensive search over the camera-pose space. These challenges motivate a shared representation for exploration and refinement and an efficient decision process that accounts for long-term reconstruction progress.

To address these challenges, we present a reinforcement learning framework for active Gaussian scene reconstruction based on hybrid voxel maps, as illustrated in Fig. 1. These maps align exploration state and reconstruction uncertainty on a common voxel grid. The exploration voxel map

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.

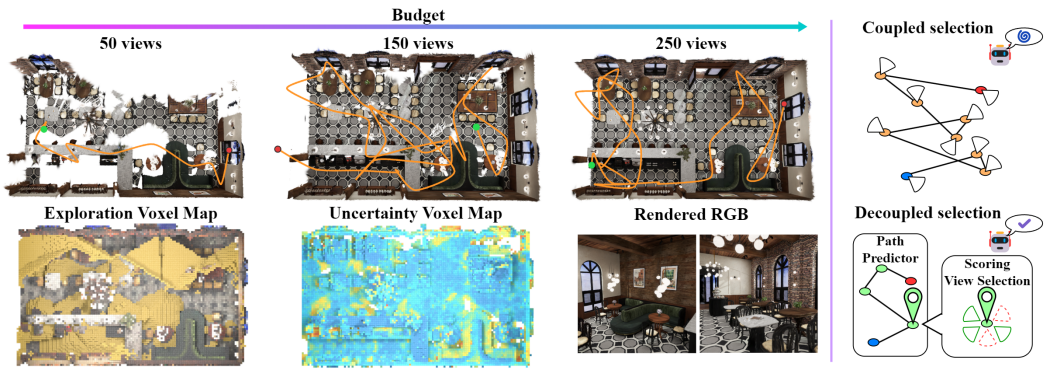

Figure 1: **Active reconstruction with decoupled position and view selection.** Our hybrid voxel maps align exploration state and reconstruction uncertainty on a common voxel maps, providing a shared basis for decision making. We decouple position and viewing direction selection, avoiding the large policy action space induced by joint camera-pose selection.

records free, occupied, and unknown space, while the uncertainty voxel map aggregates reconstruction uncertainty estimated from the Fisher information of individual Gaussian primitives. This representation allows exploration and refinement gains to be evaluated in a common spatial unit for position selection, view scoring, and policy rewards. We further decouple position selection from viewing direction selection to reduce the policy action space. Position selection is learned from the hybrid voxel maps to account for long term gains, while viewing direction selection evaluates candidate directions on the same maps at the selected target position. To train the policy toward both exploration and refinement, we define voxel based progress rewards from changes in the voxel representation, rewarding the discovery of unknown regions as well as the reduction of reconstruction uncertainty. Our contributions are threefold.

- We propose hybrid voxel maps that align exploration state and Gaussian reconstruction uncertainty on a common voxel grid, providing a unified representation for evaluating exploration and refinement gains in active Gaussian scene reconstruction.
- We introduce a decoupled decision framework that learns position selection from the hybrid voxel maps for long horizon decision making and evaluates candidate viewing directions at the selected position, avoiding joint decisions over the full camera pose space.
- We design voxel based progress rewards that jointly reflect exploration and refinement progress, enabling the policy to learn position selection toward long term active reconstruction objectives.

2 Related Works

2.1 Active Scene Reconstruction.

Active reconstruction methods select views to explore unknown regions and improve incomplete reconstruction. Recent Gaussian Splatting based systems show that Gaussian maps can serve as reconstruction outputs and spatial signals for downstream decisions (Keetha et al., 2024; Yan et al., 2024; Yugay et al., 2023). Existing active reconstruction methods mainly optimize view selection through uncertainty or coverage criteria. FisherRF (Jiang et al., 2024), GauSS-MI (Xie et al., 2025), and HGS-Planner (Xu et al., 2025) use information based objectives to estimate candidate view utility, while ActiveGS (Jin et al., 2025), ActiveGAMER (Chen et al., 2025), and GS-Planner (Jin et al., 2024) focus on geometric completeness or rendering quality. However, these methods typically evaluate views independently and rely on greedy selection. Recent methods extend the planning horizon through hierarchical planning, trajectory search, or optimization, including ActiveSplat (Li et al., 2025b), MAGICIAN (Li et al., 2026), and TRACE (Zheng et al., 2026), but the decision cost

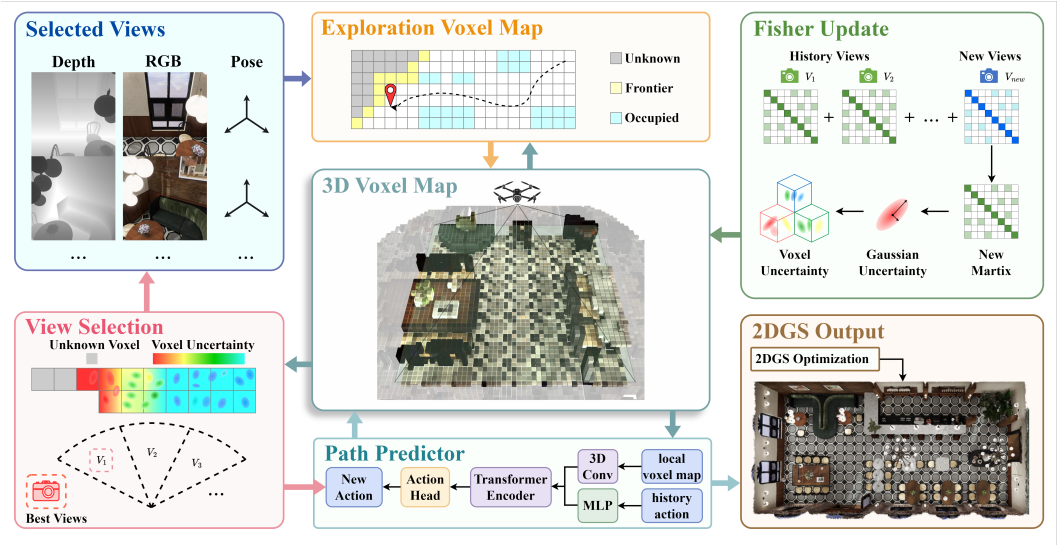

Figure 2: **Overview of our framework.** Our framework integrates exploration state and reconstruction uncertainty into hybrid voxel maps and decouples target-position selection from viewing-direction selection. Selected RGB-D observations update the exploration voxel map and the 2D Gaussian representation, whose per Gaussian uncertainty is aggregated into the uncertainty voxel map. The Path Predictor selects the next target position, while view selection determines the viewing directions based on unknown-space discovery and expected uncertainty reduction.

increases with the planning horizon. Our method learns a policy over a reconstruction aware voxel state, encoding long horizon decisions into the policy while maintaining bounded inference cost.

2.2 Learning-Based Exploration Policies.

Learning based methods have been studied for active exploration and reconstruction by learning decision policies from spatial representations. Active Neural Mapping (Chaplot et al., 2020) and OccAnt (Ramakrishnan et al., 2020) learn exploration strategies from learned maps or occupancy representations, but mainly focus on discovering unknown regions. For active reconstruction, GenNBV (Chen et al., 2024) learns a next best view policy guided by surface coverage, which is effective for object scale scanning but does not explicitly model the refinement of observed regions in indoor reconstruction. Uncertainty has also been incorporated into active view selection. ActiveNeRF Pan et al. (2022) and NVF Xue et al. (2024) use uncertainty estimation to guide view selection for neural representations. For longer horizons, NextBestPath Li et al. (2025a) learns accumulated coverage gain for distant goal selection, while ActiveRMAP Zhan et al. (2022) considers longer horizon decisions through trajectory optimization. However, existing methods mainly use uncertainty for individual view selection or local optimization. Our approach instead transforms Gaussian reconstruction uncertainty into a voxel representation that serves as the shared state for policy observation, view scoring, and reward computation.

3 Method

We propose hybrid voxel maps that integrate exploration state and reconstruction uncertainty into a unified representation. Based on this representation, we decouple target position and viewing direction selection and train a policy with voxel based progress rewards.

3.1 2DGS Map Representation

We represent the scene using a set of 2D Gaussian primitives following 2D Gaussian Splatting(Huang et al., 2024). Each Gaussian $\mathcal{G}_i = \{\boldsymbol{\mu}_i, \mathbf{q}_i, \mathbf{s}_i, \alpha_i, \mathbf{h}_i\}$ is parameterized by its center

$\boldsymbol{\mu}_i$, orientation $\mathbf{q}_i$, spatial extent $\mathbf{s}_i$, opacity α_i, and spherical harmonic coefficients $\mathbf{h}_i$ for appearance modeling. The orientation defines a local tangent frame, and each primitive represents a locally planar surface element:

$$\mathbf{P}_i(u, v) = \boldsymbol{\mu}_i + s_{u,i}\mathbf{t}_{u,i}u + s_{v,i}\mathbf{t}_{v,i}v. \tag{1}$$

This explicit geometric representation provides the surface attributes for uncertainty estimation. Following Keetha et al. (2024), we reconstruct the map incrementally by initializing Gaussians from depth observations and optimizing them with keyframes. The mapping objective is defined as

$$\mathcal{L}_{\text{map}} = \mathcal{L}_{\text{rgb}} + 0.8\,\mathcal{L}_{\text{depth}} + 0.1\,\mathcal{L}_{\text{normal}}, \tag{2}$$

where $\mathcal{L}_{\text{rgb}}$ combines an L_1 photometric loss with SSIM, $\mathcal{L}_{\text{depth}}$ enforces depth consistency, and $\mathcal{L}_{\text{normal}}$ encourages surface-normal consistency. The optimized Gaussian map is then used to update the uncertainty voxel map.

3.2 Hybrid Voxel Maps

We maintain hybrid voxel maps that combine reconstruction uncertainty and exploration states in a shared spatial representation.

Uncertainty Voxel Map The uncertainty of a Gaussian measures how well its geometry is constrained by the collected observations. We quantify this uncertainty using the Fisher information of its geometric parameters, following the per-Gaussian Fisher formulation of Pop-GS (Wilson et al., 2025). For Gaussian g, the information from one observation is

$$\mathbf{F}_g = \mathbf{J}_g^\top \mathbf{J}_g, \tag{3}$$

where $\mathbf{J}_g$ is the Jacobian of rendered color and depth with respect to the Gaussian geometry. We maintain an incremental approximation $\bar{\mathbf{F}}_g$ during mapping and convert it into a scalar uncertainty using the regularized A-optimal criterion

$$\nu_g = \frac{\lambda}{d}\,\mathrm{tr}((\bar{\mathbf{F}}_g + \lambda I)^{-1}), \tag{4}$$

where $d = 8$ is the dimension of the geometric parameter vector. The Gaussian uncertainties are then aggregated into the voxel space. Let $\mathcal{G}(v)$ denote the set of Gaussians whose centers lie in voxel v. For a nonempty voxel, its uncertainty is

$$U(v) = \frac{1}{|\mathcal{G}(v)|} \sum_{g \in \mathcal{G}(v)} \nu_g. \tag{5}$$

For voxels without Gaussians, we set $U(v) = 1$. The resulting uncertainty field provides the reconstruction state used by subsequent decision making.

Exploration Voxel Map The exploration voxel map records the observed state of the scene and complements the uncertainty field with spatial information about unexplored regions. Each voxel is labeled as free, occupied, or unknown and is initialized as unknown. Given an RGB-D observation, depth pixels are back-projected and the voxels traversed by the sensor rays are marked as free, while the voxels containing depth endpoints are marked as occupied.

The boundary between free and unknown regions forms the frontier, which indicates potential exploration directions. The exploration voxel map therefore provides the spatial context required for navigation and target position selection. Together with the uncertainty voxel map, it forms the hybrid voxel representation used by the policy, view scoring, and reward computation.

3.3 Decoupled Action Space

Selecting camera poses in the full SE(3) space requires jointly optimizing position and viewing direction, resulting in a large and non-convex action space. We therefore decompose the decision into target position selection and viewing direction selection. The policy predicts target positions from the hybrid voxel maps, while viewing directions are selected by scoring candidate views on the same maps. This decomposition reduces the policy action space and facilitates long horizon decision making. The following sections describe the path decision and viewpoint decision in detail.

Path Decision The Path Predictor is a policy network that predicts a target position at each decision step to maximize exploration and refinement gains. It observes an ego centric local crop of the hybrid voxel maps, together with recent position history. The voxel maps are encoded by 3D convolutions into spatial features, while the pose history is embedded by an MLP. These features are fused through a cross-modal transformer, and an action head predicts a normalized position displacement $\mathbf{a}_t = (\Delta x_t, \Delta y_t, \Delta z_t)$ within the local region. The displacement is scaled by the extent of the local region and transformed into a target position in the world coordinate system. The policy therefore determines where to move based on exploration state, reconstruction uncertainty, and recent motion.

Before execution, the predicted target is checked for reachability using the exploration voxel map. We first test the direct path from the current position to the target against occupied voxels. If the direct path is blocked, a graph search over free voxels is used to find an alternative path. Targets without a valid path are rejected without acquiring an RGB-D observation or updating the maps. Once a valid target is reached, view selection determines the observation directions.

View Selection After reaching the target position, we generate a fixed set of candidate viewing directions and evaluate each candidate using the hybrid voxel maps. The score considers both the refinement of observed surfaces and the discovery of unknown regions.

For the refinement term, we score each candidate by the expected uncertainty reduction it produces. Specifically, we render the expected observation from the current Gaussian map and compute the Fisher contribution $\mathbf{C}_g$ of the candidate view for each visible Gaussian. Only Gaussians already present in the current map participate in this refinement prediction. Let $\mathbf{H}_g = \bar{\mathbf{F}}_g + \lambda \mathbf{I}$. The expected uncertainty reduction of Gaussian g is

$$\Delta\nu_g = \frac{\lambda}{d}\left[\mathrm{tr}\left(\mathbf{H}_g^{-1}\right) - \mathrm{tr}\left((\mathbf{H}_g + \mathbf{C}_g)^{-1}\right)\right]. \tag{6}$$

The expected uncertainty reduction of each voxel is computed by aggregating the predicted reductions of its Gaussians. The refinement gain $R_{\mathrm{known}}(i)$ is then obtained by summing the uncertainty reductions over the voxels visible from candidate view i.

For the exploration term, we cast rays from the candidate view through the exploration voxel map within the sensing range. Each ray terminates at the first voxel that is not labeled free, and an unknown endpoint contributes to the exploration gain $R_{\mathrm{unknown}}(i)$. Rays that remain in free space up to the sensing limit contribute no exploration gain.

The final score of candidate i is

$$S(i) = R_{\mathrm{unknown}}(i) + \lambda_o R_{\mathrm{known}}(i), \tag{7}$$

where both terms are measured in voxel space. The coefficient λ_o is the fraction of free and occupied voxels in the local region, allowing the balance between exploration and refinement to adapt to the current scene state. Candidate directions are ranked by $S(i)$, and the selected directions are used to acquire RGB-D observations and update the Gaussian map and both voxel fields.

3.4 Training Strategy

We train the Path Predictor with proximal policy optimization (PPO) (Schulman et al., 2017) to optimize exploration and refinement progress measured from the hybrid voxel maps. Since active reconstruction requires both discovering unknown regions and improving observed surfaces, we define voxel based progress rewards using three complementary progress measures in the same voxel space.

Let $\mathcal{V}_{\mathrm{gt}}$ denote the set of ground-truth surface voxels available during training, and let N_{surf} and N_{exp} be fixed normalization constants for coverage and exploration progress, respectively. Let $\mathcal{V}_{\mathrm{cov}}(t)$ denote the set of surface voxels covered by depth observations up to time t. A surface voxel is considered covered when at least one depth measurement is projected into the corresponding voxel. The coverage progress is defined as

$$P_{\mathrm{cov}}(t) = \frac{|\mathcal{V}_{\mathrm{cov}}(t)|}{N_{\mathrm{surf}}}. \tag{8}$$

This term encourages the policy to acquire observations of previously unseen surfaces. Similarly, let $\mathcal{V}_{\text{exp}}(t)$ denote the set of voxels labeled as free or occupied by time t. The exploration progress is defined as

$$P_{\text{exp}}(t) = \frac{|\mathcal{V}_{\text{exp}}(t)|}{N_{\text{exp}}}. \tag{9}$$

Coverage and exploration measure what has been observed, but do not indicate whether the observed surfaces are sufficiently constrained. We therefore define the refinement progress from the uncertainty voxel map as

$$P_{\text{unc}}(t) = \frac{1}{N_{\text{surf}}} \sum_{v \in \mathcal{V}_{\text{gt}}} \left(1 - U(v,t)\right). \tag{10}$$

With the default $U(v,t) = 1$ defined in Sec. 3.2, voxels without Gaussians contribute zero to P_{unc}. Like existing Gaussians, newly initialized Gaussians are included in the voxel uncertainty update after all selected observations are incorporated. All progress measures are evaluated before each policy decision and after incorporating all observations acquired at the selected target. The predicted per Gaussian reduction $\Delta\nu_g$ is used only for candidate view scoring in Sec. 3.3, whereas P_{unc} is computed from the updated voxel uncertainty field after the selected observations are incorporated. The reward is defined from the improvement of the three progress measures:

$$\Delta P_k(t) = P_k(t) - P_k(t-1), \qquad R_t = \sum_{k \in \{\text{cov},\text{exp},\text{unc}\}} \beta_k \, \Delta P_k(t). \tag{11}$$

where β_k balances the three objectives. The three terms encourage surface discovery, spatial exploration, and reconstruction refinement within the same voxel representation. Under the undiscounted training setting, each progress difference telescopes over an episode, so its accumulated reward corresponds to the final improvement of the associated progress measure while providing dense feedback during exploration.

4 Experiments

4.1 Experimental Setup

Datasets. We conduct experiments on InteriorGS (SpatialVerse Research Team, 2025),a dataset introduced by Miao et al. (2026), and Replica (Straub et al., 2019). InteriorGS contains 1,000 synthetic indoor scenes and is simulated in Isaac Sim 5.0 (NVIDIA). We use 900 scenes for training and 8 held-out scenes for evaluation. Some surfaces in InteriorGS cannot be observed from navigable regions and are still included in the ground truth, so the reported coverage and completeness may not reach their theoretical limits. We additionally evaluate the trained policy on Replica using Habitat-Sim (Puig et al., 2024) without fine-tuning to examine transfer from the larger InteriorGS scenes to the smaller indoor scenes commonly used by existing methods.

Baselines. On InteriorGS, we compare with ActiveGS (Jin et al., 2025), ActiveGAMER (Chen et al., 2025), and FisherRF (Jiang et al., 2024). ActiveGS and ActiveGAMER represent coverage oriented Gaussian reconstruction methods, while FisherRF selects views using Fisher information. We use the official active reconstruction implementation of FisherRF. On Replica, we additionally compare with NARUTO (Feng et al., 2024), a neural implicit active reconstruction method. FisherRF is evaluated without a depth limit, following its original implementation, since it behaves abnormally under the 0.1 –3.2 m depth range used by the other methods.

Evaluation Metrics. For geometry evaluation, we uniformly sample the ground truth mesh and compare it with points extracted from the reconstructed Gaussian map. Coverage measures the percentage of ground truth points with a reconstructed point within distance τ, while Completeness measures the mean distance from ground truth points to the reconstruction. For rendering evaluation, we render 500 viewpoints sampled from the free space and report PSNR, SSIM, and LPIPS against the ground truth images.

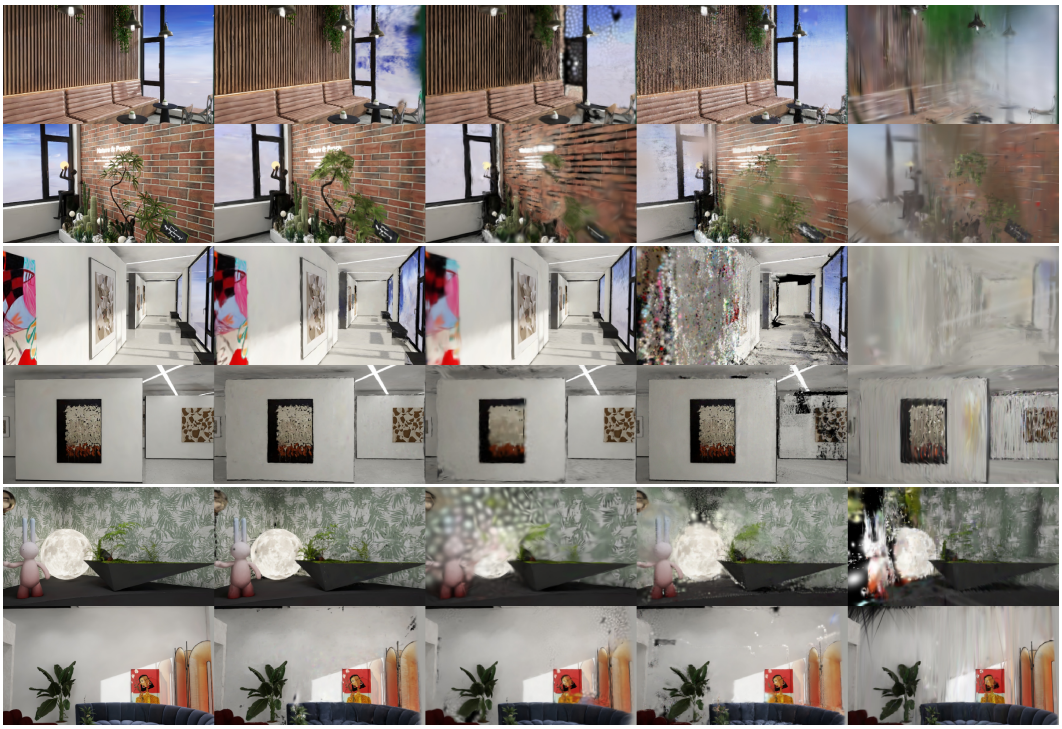

Figure 3: Qualitative rendering comparisons on InteriorGS. Each pair of rows shows two test views from scenes 0051, 0070, and 0098, respectively. Our method yields sharper details, fewer artifacts, and better preservation of scene structures than the baselines.

4.2 Implementation Details

Reconstruction and sensing configuration. Unless otherwise stated, all methods use a camera resolution of 512×512, a field of view of $60^\circ \times 60^\circ$, and a depth range of 0.1–3.2 m. We use a common budget of 500 RGB-D observations per scene and 30 mapping iterations per fused frame. For Gaussian uncertainty estimation, we use regularization $\lambda = 10^{-5}$. Rejected targets count as policy decision steps but do not consume the observation budget.

Path Predictor architecture. The Path Predictor uses a three-layer 3D convolutional encoder with a feature dimension of 64 to encode a local crop of the hybrid voxel maps. A history of the 30 most recent poses is encoded by an MLP with hidden dimension 256. The resulting spatial and pose features form 257 tokens, which are fused by a two-layer cross-modal transformer (Yang et al., 2022) with hidden dimension 64. The action head predicts the normalized displacement defined in Path Decision.

View selection. At each valid target position, we evaluate 18 candidate viewing directions using Eq. (7) and execute up to six highest-scoring directions above a threshold of 150.0. If fewer than two exceed the threshold, the top two are selected.

Training details and PPO hyperparameters are provided in Appendix C.

4.3 Reconstruction Results

Results on InteriorGS. Table 1 reports the final reconstruction results on InteriorGS and Replica under equal observation budgets. On InteriorGS, our method achieves the best performance across all evaluated metrics. The two coverage-driven baselines, ActiveGS and ActiveGAMER, score views according to unobserved regions, so their selection signal becomes less informative once

Table 1: Final reconstruction results on InteriorGS and Replica under equal observation budgets. Our method achieves strong performance in both geometric reconstruction and rendering quality across both benchmarks. Best results for each dataset are shown in bold.

Method	Coverage ↑	Comp. (m) ↓	PSNR ↑	SSIM ↑	LPIPS ↓
InteriorGS					
ActiveGS	0.572	0.204	20.96	0.754	0.342
ActiveGAMER	0.621	0.127	19.96	0.686	0.433
FisherRF	0.595	0.143	17.92	0.676	0.504
Ours	**0.692**	**0.085**	**24.37**	**0.830**	**0.199**
Replica					
ActiveGS	**0.967**	0.020	**31.43**	**0.923**	0.149
ActiveGAMER	0.864	0.060	23.78	0.787	0.324
NARUTO	0.890	0.021	24.85	0.801	0.404
Ours	0.965	**0.019**	31.13	0.902	**0.114**

Table 2: Reconstruction progress on InteriorGS across observation budgets. Our method consistently achieves strong reconstruction performance. Best results are shown in bold.

	ActiveGS		ActiveGAMER		FisherRF		Ours	
Budget	Comp.↓	PSNR↑	Comp.↓	PSNR↑	Comp.↓	PSNR↑	Comp.↓	PSNR↑
20%	1.266	11.55	1.619	10.71	0.439	15.08	0.866	14.79
40%	0.607	14.92	0.703	13.72	0.222	17.06	0.252	19.61
60%	0.400	17.55	0.278	16.33	0.179	17.75	0.110	22.89
80%	0.287	19.30	0.191	18.48	0.150	18.16	0.094	23.74
100%	0.204	20.96	0.127	19.96	0.143	17.92	**0.085**	**24.37**

most of the accessible scene has been explored. FisherRF instead scores views using Fisher information and targets reconstruction quality, but selects views greedily in isolation. In contrast, our policy jointly exploits exploration state and reconstruction uncertainty from the hybrid voxel maps and optimizes decisions over the entire exploration process.Figure 3 presents qualitative rendering comparisons on representative InteriorGS scenes, and Figure. 4 compares the reconstruction completeness across different methods.

To examine how reconstruction quality evolves with the observation budget, Table 2 reports completeness and PSNR throughout the reconstruction process. FisherRF performs better at the early stage, while our method becomes substantially stronger as more observations are acquired and eventually achieves the best completeness and PSNR among all compared methods. These results indicate that the learned policy increasingly focuses the observation budget on views that improve both geometry and rendering quality as exploration progresses.

Results on Replica. The Replica results in Table 1 evaluate cross-dataset transfer without fine-tuning. Despite being trained only on the larger InteriorGS environments, our method maintains strong reconstruction quality on the substantially smaller Replica scenes and remains competitive with methods designed for this setting. These results suggest that the learned policy transfers effectively across differences in scene scale and environment distribution without additional adaptation.

Efficiency. We profile the mapping pipeline on an InteriorGS scene (0403) using a single RTX PRO 6000 GPU. The Gaussian optimization and the Gaussian pruning account for 77.0% and 16.9% of the mapping time, respectively, while the Fisher based uncertainty update introduced by our method accounts for only 5.2%, and frame insertion for 0.7%. Since the optimization and pruning costs are shared by all Gaussian-based active reconstruction methods, the additional overhead of maintaining the uncertainty field is marginal.

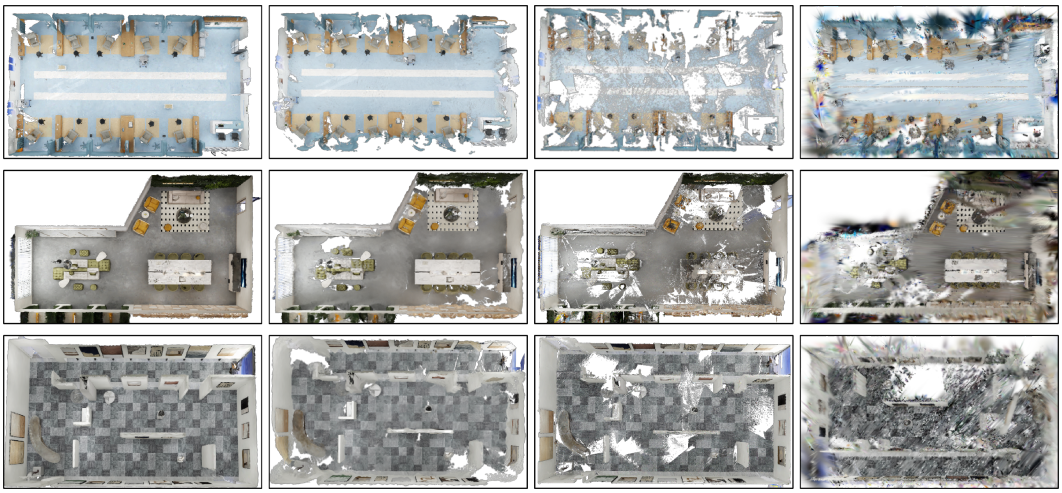

Figure 4: Bird's eye view comparison of reconstruction completeness on InteriorGS scenes 0042, 0073 and 0049. Each row shows above views of one scene across methods, highlighting completeness and missing regions across methods. Our method reconstructs more complete structures.

Table 3: Ablation on one InteriorGS test scene under an equal observation budget. The full model achieves the best performance across all metrics. Best results are shown in bold.

Variant	Coverage ↑	Comp. (m) ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Random	0.422	0.627	14.47	0.576	0.430
Exploration only	0.540	0.241	19.41	0.733	0.292
Coverage only	0.568	0.166	20.68	0.779	0.256
Cov.+Exp.	0.608	0.149	22.99	0.811	0.228
Full	**0.615**	**0.146**	**24.13**	**0.832**	**0.195**

4.4 Ablation Study

We ablate the reward design and reconstruction uncertainty on one representative InteriorGS test scene under the same mapping pipeline and observation budget. *Random* uses random target positions and viewing directions. *Coverage only*, *Exploration only*, and *Cov.+Exp.* use only the exploration voxel map as policy observation and exclude uncertainty from view scoring, while being trained with the corresponding coverage, exploration, or combined rewards. The *Full* model further incorporates the uncertainty voxel map, uncertainty reward, and uncertainty-based view scoring. Table 3 shows that combining coverage and exploration outperforms either reward alone, while incorporating reconstruction uncertainty further improves reconstruction quality. The substantial degradation of *Random* confirms the importance of learned position and view selection.

5 Conclusion

In this paper, we proposed a reinforcement learning framework for active Gaussian scene reconstruction based on hybrid voxel maps. By integrating exploration states with reconstruction uncertainty, our method provides a unified spatial representation for exploring unknown regions and refining poorly reconstructed surfaces. We further decouple target-position selection from viewing-direction selection, using a learned policy for long horizon decisions and uncertainty-based scoring for candidate views. Voxel-based progress rewards guide policy learning toward both exploration and reconstruction refinement. Experiments on InteriorGS demonstrate improvements in geometric and rendering quality over existing active reconstruction methods under the same observation budget. Evaluation on Replica further demonstrates the transferability of our learned policy to smaller indoor scenes without fine-tuning.

AI USE STATEMENT

In this work, we used generative AI tools for assisting with translation. We have not used generative AI tools for helping develop theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, implementing methods, cleaning and reformatting dataset, supporting qualitative and thematic data analysis, interpreting results, and generating synthetic data sets are not applicable to this work.

Additionally, we used generative AI tools for creating or editing software code and editing a research paper to improve readability.

We have reviewed all AI-assisted work. AI-generated translations and polished text were reviewed by the authors to ensure that the original meaning and terminology were accurately conveyed. We also used AI-generated code in our experiments, and its correctness was verified by the authors.

We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We provide detailed descriptions of the proposed method, implementation details, and experimental setup in the main paper. Additional hyperparameter settings for reinforcement learning training are provided in the appendix to facilitate reproduction of our results.

REFERENCES

Andreas Bircher, Mina Kamel, Kostas Alexis, Helen Oleynikova, and Roland Siegwart. Receding horizon" next-best-view" planner for 3d exploration. In *2016 IEEE international conference on robotics and automation (ICRA)*, pp. 1462–1468. IEEE, 2016.

Devendra Singh Chaplot, Dhiraj Gandhi, Saurabh Gupta, Abhinav Gupta, and Ruslan Salakhutdinov. Learning to explore using active neural slam. *arXiv preprint arXiv:2004.05155*, 2020.

Liyan Chen, Huangying Zhan, Kevin Chen, Xiangyu Xu, Qingan Yan, Changjiang Cai, and Yi Xu. Activegamer: Active gaussian mapping through efficient rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16486–16497, June 2025.

Xiao Chen, Quanyi Li, Tai Wang, Tianfan Xue, and Jiangmiao Pang. Gennbv: Generalizable next-best-view policy for active 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16436–16445, June 2024.

Ziyue Feng, Huangying Zhan, Zheng Chen, Qingan Yan, Xiangyu Xu, Changjiang Cai, Bing Li, Qilun Zhu, and Yi Xu. Naruto: Neural active reconstruction from uncertain target observations. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21572–21583. IEEE, 2024.

Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *SIGGRAPH 2024 Conference Papers*. Association for Computing Machinery, 2024. doi: 10.1145/3641519.3657428.

Stefan Isler, Reza Sabzevari, Jeffrey Delmerico, and Davide Scaramuzza. An information gain formulation for active volumetric 3d reconstruction. In *2016 IEEE international conference on robotics and automation (ICRA)*, pp. 3477–3484. IEEE, 2016.

Wen Jiang, Boshu Lei, and Kostas Daniilidis. FisherRF: Active view selection and mapping with radiance fields using fisher information. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024. doi: 10.1007/978-3-031-72624-8_24.

Liren Jin, Xingguang Zhong, Yue Pan, Jens Behley, Cyrill Stachniss, and Marija Popović. ActiveGS: Active Scene Reconstruction Using Gaussian Splatting. *IEEE Robotics and Automation Letters*, 10(5):4866–4873, 2025. doi: 10.1109/LRA.2025.3555149.

Rui Jin, Yuman Gao, Yingjian Wang, Yuze Wu, Haojian Lu, Chao Xu, and Fei Gao. Gs-planner: A gaussian-splatting-based planning framework for active high-fidelity reconstruction. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 11202–11209. IEEE, 2024.

Nikhil Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian Scherer, Deva Ramanan, and Jonathon Luiten. Splatam: Splat track & map 3d gaussians for dense rgb-d slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21357–21366, June 2024.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (SIGGRAPH Conference Proceedings)*, 42(4), 2023. URL https://www-sop.inria.fr/reves/Basilic/2023/KKLD23.

Soomin Lee, Le Chen, Jiahao Wang, Alexander Liniger, Suryansh Kumar, and Fisher Yu. Uncertainty guided policy for active robotic 3d reconstruction using neural radiance fields. *IEEE Robotics and Automation Letters*, 7(4):12070–12077, 2022. doi: 10.1109/LRA.2022.3212668.

Shiyao Li, Antoine Guédon, Clémentin Boittiaux, Shizhe Chen, and Vincent Lepetit. Nextbestpath: Efficient 3d mapping of unseen environments. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025a. URL https://openreview.net/forum?id=7WaRh4gCXp.

Shiyao Li, Antoine Guédon, Shizhe Chen, and Vincent Lepetit. Magician: Efficient long-term planning with imagined gaussians for active mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21606–21615, June 2026.

Yuetao Li, Zijia Kuang, Ting Li, Qun Hao, Zike Yan, Guyue Zhou, and Shaohui Zhang. Activesplat: High-fidelity scene reconstruction through active gaussian splatting. *IEEE Robotics and Automation Letters*, 10(8):8099–8106, 2025b. doi: 10.1109/LRA.2025.3580331.

Bingchen Miao, Rong Wei, Zhiqi Ge, Xiaoquan Sun, Shiqi Gao, Jingzhe Zhu, Renhan Wang, Siliang Tang, Jun Xiao, Rui Tang, and Juncheng Li. Towards physically executable 3D gaussian for embodied navigation. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=HB6KvsqcAn.

Riccardo Monica and Jacopo Aleotti. Surfel-based next best view planning. *IEEE Robotics and Automation Letters*, 3(4):3324–3331, 2018.

NVIDIA. Isaac sim. URL https://github.com/isaac-sim/IsaacSim.

Xuran Pan, Zihang Lai, Shiji Song, and Gao Huang. ActiveNeRF: Learning where to see with uncertainty estimation. In *Computer Vision – ECCV 2022*, volume 13693 of *Lecture Notes in Computer Science*, pp. 230–246. Springer, 2022. doi: 10.1007/978-3-031-19827-4_14.

Richard Pito. A solution to the next best view problem for automated surface acquisition. *IEEE Transactions on pattern analysis and machine intelligence*, 21(10):1016–1030, 1999.

Xavier Puig, Eric Undersander, Andrew Szot, Mikael Dallaire Cote, Tsung-Yen Yang, Ruslan Partsey, Ruta Desai, Alexander Clegg, Michal Hlavac, So Yeon Min, Vladimír Vondruš, Theophile Gervet, Vincent-Pierre Berges, John M Turner, Oleksandr Maksymets, Zsolt Kira, Mrinal Kalakrishnan, Jitendra Malik, Devendra Singh Chaplot, Unnat Jain, Dhruv Batra, Akshara Rai, and Roozbeh Mottaghi. Habitat 3.0: A Co-Habitat for Humans, Avatars, and Robots. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://mlanthology.org/iclr/2024/puig2024iclr-habitat/.

Santhosh K. Ramakrishnan, Ziad Al-Halah, and Kristen Grauman. Occupancy anticipation for efficient exploration and navigation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 400–418, 2020. doi: 10.1007/978-3-030-58558-7_24.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. doi: 10.48550/arXiv.1707.06347.

Manycore Tech Inc. SpatialVerse Research Team. Interiorgs: A 3d gaussian splatting dataset of semantically labeled indoor scenes. https://huggingface.co/datasets/spatialverse/InteriorGS, 2025.

Cyrill Stachniss, Giorgio Grisetti, and Wolfram Burgard. Information gain-based exploration using Rao-Blackwellized particle filters. In *Proceedings of Robotics: Science and Systems*, Cambridge, USA, June 2005. doi: 10.15607/RSS.2005.I.009.

Julian Straub, Thomas Whelan, Lingni Ma, Yufan Chen, Erik Wijmans, Simon Green, Jakob J. Engel, Raul Mur-Artal, Carl Ren, Shobhit Verma, Anton Clarkson, Mingfei Yan, Brian Budge, Yajie Yan, Xiaqing Pan, June Yon, Yuyang Zou, Kimberly Leon, Nigel Carter, Jesus Briales, Tyler Gillingham, Elias Mueggler, Luis Pesqueira, Manolis Savva, Dhruv Batra, Hauke M. Strasdat, Renzo De Nardi, Michael Goesele, Steven Lovegrove, and Richard Newcombe. The replica dataset: A digital replica of indoor spaces. *arXiv preprint arXiv:1906.05797*, 2019. doi: 10.48550/arXiv.1906.05797.

Joey Wilson, Marcelino Almeida, Sachit Mahajan, Martin Labrie, Maani Ghaffari, Omid Ghasemalizadeh, Min Sun, Cheng-Hao Kuo, and Arnab Sen. Pop-gs: Next best view in 3d-gaussian splatting with p-optimality. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3646–3655. IEEE, 2025.

Yuhan Xie, Yixi Cai, Yinqiang Zhang, Lei Yang, and Jia Pan. Gauss-mi: Gaussian splatting shannon mutual information for active 3d reconstruction. In *Robotics: Science and Systems (RSS)*, 2025. doi: 10.15607/RSS.2025.XXI.030.

Zijun Xu, Rui Jin, Ke Wu, Yi Zhao, Zhiwei Zhang, Jieru Zhao, Fei Gao, Zhongxue Gan, and Wenchao Ding. HGS-Planner: Hierarchical planning framework for active scene reconstruction using 3D gaussian splatting. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 14161–14167, 2025.

Shangjie Xue, Jesse Dill, Pranay Mathur, Frank Dellaert, Panagiotis Tsiotra, and Danfei Xu. Neural visibility field for uncertainty-driven active mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18122–18132, June 2024.

Chi Yan, Delin Qu, Dan Xu, Bin Zhao, Zhigang Wang, Dong Wang, and Xuelong Li. Gs-slam: Dense visual slam with 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19595–19604, June 2024.

Ruihan Yang, Minghao Zhang, Nicklas Hansen, Huazhe Xu, and Xiaolong Wang. Learning Vision-Guided Quadrupedal Locomotion End-to-End with Cross-Modal Transformers. In *International Conference on Learning Representations*, 2022. URL https://mlanthology.org/iclr/2022/yang2022iclr-learning-a/.

Vladimir Yugay, Yue Li, Theo Gevers, and Martin R. Oswald. Gaussian-slam: Photo-realistic dense slam with gaussian splatting, 2023.

Huangying Zhan, Jiyang Zheng, Yi Xu, Ian Reid, and Hamid Rezatofighi. Activermap: Radiance field for active mapping and planning. *arXiv preprint arXiv:2211.12656*, 2022. doi: 10.48550/arXiv.2211.12656.

Ziyue Zheng, Linli Shi, Bingkun He, Wen Jiang, and Ziyun Wang. Trace: Ergodic trajectory optimization for active scene reconstruction. *arXiv preprint arXiv:2608.02304*, 2026.

A Implementation Details of Gaussian Uncertainty Estimation

A.1 Geometric Parameterization and Fisher Computation

We compute Fisher information only for the geometric parameters of each Gaussian. For Gaussian g, the parameter vector is

$$\boldsymbol{\theta}_g = (\boldsymbol{\mu}_g, \boldsymbol{\ell}_g, \boldsymbol{\phi}_g) \in \mathbb{R}^8, \tag{A1}$$

where $\boldsymbol{\mu}_g \in \mathbb{R}^3$ denotes the center, $\boldsymbol{\ell}_g = \log \mathbf{s}_g \in \mathbb{R}^2$ denotes the logarithmic spatial extent, and $\boldsymbol{\phi}_g \in \mathbb{R}^3$ denotes the rotation perturbation. We exclude appearance parameters because additional viewpoints primarily improve the geometric constraints considered by the active reconstruction policy.

For an observation, the Jacobian $\mathbf{J}_g$ contains derivatives of the rendered color and depth with respect to $\boldsymbol{\theta}_g$. With the residual precision matrix $\mathbf{W}$, the corresponding Fisher contribution is

$$\mathbf{F}_g = \mathbf{J}_g^\top \mathbf{W} \mathbf{J}_g. \tag{A2}$$

After normalizing the residual channels, we use the simplified form $\mathbf{F}_g = \mathbf{J}_g^\top \mathbf{J}_g$ reported in the main paper.

A.2 Incremental Fisher Update

Recomputing the Fisher information from all historical observations after every map update would require repeatedly evaluating historical Jacobians. We instead maintain an incremental approximation $\bar{\mathbf{F}}_g$ for each Gaussian. Let $\mathcal{B}$ denote the stored keyframes revisited in the current mapping update. Their previous contributions are replaced by the values recomputed at the current Gaussian parameters, while the newly captured frame t contributes $\mathbf{F}_{t,g}$:

$$\bar{\mathbf{F}}_g \leftarrow \bar{\mathbf{F}}_g - \sum_{k \in \mathcal{B}} \mathbf{F}_{k,g} + \sum_{k \in \mathcal{B}} \mathbf{F}_{k,g}^{\text{new}} + \mathbf{F}_{t,g}. \tag{A3}$$

A keyframe contribution is stored separately only while the frame remains active for mapping. Once it is no longer revisited, its most recent contribution remains in the aggregate without further relinearization. Consequently, $\bar{\mathbf{F}}_g$ is a persistent approximation rather than the exact historical Fisher information evaluated at the current Gaussian parameters. This avoids replaying all previous observations while retaining their accumulated geometric constraints.

A.3 Coordinate Normalization

The geometric parameter vector contains translation, logarithmic scale, and rotation variables with different units. Before evaluating the A-optimal uncertainty, we normalize the translation coordinates using a fixed characteristic length ℓ. Equivalently, each Fisher contribution is transformed by

$$\mathbf{F}_g \leftarrow \mathbf{D}\mathbf{F}_g\mathbf{D}, \qquad \mathbf{D} = \operatorname{diag}(\ell, \ell, \ell, 1, 1, 1, 1, 1). \tag{A4}$$

This normalization makes the parameter dimensions comparable when evaluating $\operatorname{tr}((\bar{\mathbf{F}}_g + \lambda \mathbf{I})^{-1})$. We use a characteristic length $\ell = 0.1, \mathrm{m}$ in all experiments.

B Reachability Check Details

The target position predicted by the Path Predictor is verified using the current exploration voxel map before execution. We first examine the straight path from the current position to the target and reject it as a direct path if it intersects occupied or unavailable voxels. When the direct path is blocked, we perform a graph search over the connected free voxels to determine whether a collision-free alternative path exists.

A target is accepted only if such a path is found. Otherwise, the action is rejected without acquiring an RGB-D observation or updating the Gaussian and voxel maps. During training, the rejected action still constitutes a policy decision, while the observation budget is unchanged because no new observation is acquired.

Table A1: PPO hyperparameters used for training the Path Predictor.

Hyperparameter	Value
Optimizer	Adam
Optimization batch size	120
Learning rate	5×10^{-5}
Training iterations	1000
Training environments	7
Rollout steps (N)	240
PPO epochs	5
Buffer size	30
Value coefficient	0.8
Entropy coefficient	0.01
Discount factor γ	1.0
GAE parameter λ	0.97
PPO clipping parameter	0.2

C Training Details

Each training episode starts from a random valid pose, and an initial set of RGB-D observations is used to initialize the Gaussian and voxel maps. We train the Path Predictor with PPO using reward weights $\beta_{\text{cov}} = 2.0$, $\beta_{\text{exp}} = 2.0$, and $\beta_{\text{unc}} = 3.2$. The progress measures use fixed normalization constants $N_{\text{surf}} = 50{,}000$ and $N_{\text{exp}} = 200{,}000$, and the return is undiscounted with $\gamma = 1$.

Each episode contains at most 500 policy decision steps and terminates early when progress stagnates over the last ten executed decisions or when ten consecutive targets are rejected. The Path Predictor is trained on 8 NVIDIA RTX PRO 6000 GPUs, while the 2D Gaussian map is reconstructed online alongside policy training. We optimize the policy using PPO, with the remaining training hyperparameters summarized in Table A1.