DynActiveGS: Active Gaussian Splatting for Dynamic Scene Reconstruction

Anonymous Author(s)
Submission Id: 664

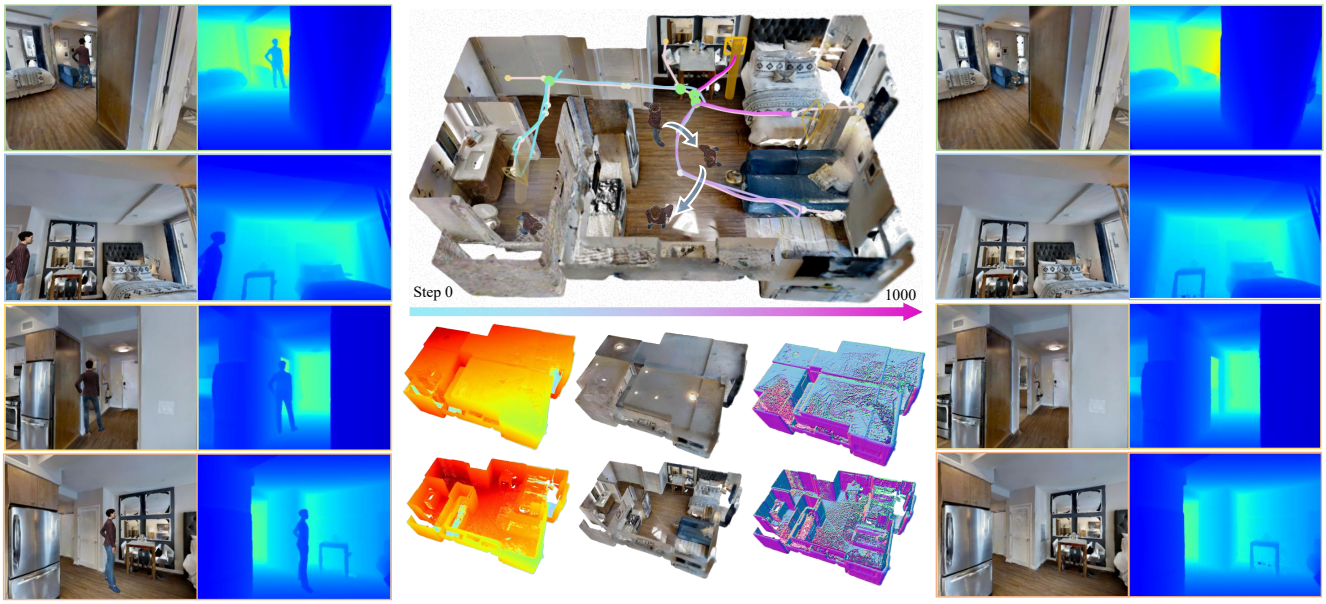

Figure 1: DynActiveGS in action. The robot actively explores a dynamic indoor scene and progressively improves reconstruction completeness. From left to right, we show RGB-D inputs with dynamic disturbances, intermediate reconstruction states, and final renderings after 1000 steps. DynActiveGS preserves stable geometry and rendering quality despite dynamic interference.

Abstract

We present DynActiveGS, a dynamic-aware active reconstruction framework based on 3D Gaussian Splatting (3DGS) for autonomous exploration in dynamic environments. The framework incrementally reconstructs a 3D Gaussian scene representation while suppressing motion-corrupted observations through online uncertainty prediction and uncertainty-weighted Gaussian optimization. A key component of DynActiveGS is the explicit decomposition of uncertainty into structural uncertainty and motion-induced uncertainty, which enables the system to distinguish under-reconstructed static regions from dynamically unreliable areas. Based on these uncertainty fields, DynActiveGS performs dynamic-aware viewpoint selection and dynamic-constrained path planning to favor informative yet stable observations during exploration. The resulting system forms a unified closed-loop pipeline for robust active reconstruction in dynamic scenes. Extensive experiments on challenging dynamic benchmarks demonstrate consistent improvements over existing active reconstruction baselines in reconstruction accuracy, completeness, rendering quality, and exploration efficiency.

CCS Concepts

• **Computing methodologies** → **Reconstruction**; **Vision for robotics**; **Robotic planning**.

Keywords

Active Reconstruction, 3D Gaussian Splatting, Dynamic Scenes

1 Introduction

The ability to reconstruct 3D scenes from visual observations is a cornerstone of computer vision and robotics [10, 35, 50]. It enables intelligent agents to perceive, reason about, and interact with the physical world. Beyond passive perception, an ideal system should actively decide where to move and what to observe [14] in order to efficiently acquire informative data and build high-fidelity scene representations [22]. This paradigm, commonly referred to as *active 3D reconstruction* [1], lies at the intersection of perception, uncertainty modeling, and motion planning [7].

Recent advances in 3DGS [21] have significantly reshaped the landscape of real-time 3D scene representation. Compared to traditional multi-view geometry and implicit neural representations [34], 3DGS provides an explicit, differentiable, and efficient scene model, enabling high-quality rendering and incremental optimization. These properties have inspired a series of active reconstruction frameworks that leverage 3DGS to guide exploration via next-best-view (NBV) [39, 40] planning and uncertainty [25, 38, 42] driven decision making. However, existing active reconstruction approaches [6,

11, 17, 18, 29, 30, 41, 53, 55] are fundamentally designed under the assumption of static environments.

In real-world scenarios, dynamic objects such as pedestrians, vehicles, and articulated agents are ubiquitous. Their presence fundamentally challenges active reconstruction [13, 57]. Motion-induced artifacts corrupt Gaussian optimization [37, 54], uncertainty estimation becomes unreliable [24, 43], and viewpoint planning [16, 23] is often misled by transient observations. Meanwhile, recent efforts on dynamic 3DGS and Gaussian-based SLAM primarily focus on passive mapping and dynamic object filtering [44, 54, 59], without considering active perception and long-horizon exploration. As a result, current methods are unable to achieve robust and complete reconstruction in dynamic environments. This reveals a critical gap in active reconstruction under dynamic scenes.

In this paper, we present *DynActiveGS*, a dynamic-aware active reconstruction framework for 3D Gaussian Splatting in dynamic scenes. As illustrated in Fig. 1, our method first performs uncertainty-aware Gaussian reconstruction by predicting a per-pixel uncertainty map and using it to suppress motion-corrupted observations during online map optimization. Based on the reconstructed Gaussian map, we explicitly disentangle structural uncertainty from motion-induced uncertainty to build map-level uncertainty fields for active exploration. These uncertainty fields are then used to guide dynamic-aware viewpoint selection through local-global scoring on a Voronoi graph, while a motion-constrained path planner favors geometrically efficient yet dynamically stable trajectories. Together, these components form a unified perception–planning–reconstruction pipeline for robust closed-loop active reconstruction in dynamic environments.

We evaluate DynActiveGS on diverse dynamic benchmarks. Experimental results show that our method consistently outperforms existing active reconstruction baselines in reconstruction accuracy, completeness, rendering quality, and exploration efficiency. Our work advances active 3D reconstruction beyond the static-scene assumption and takes a step toward more capable embodied perception in dynamic real-world environments. Our **contributions** are summarized as follows:

- We propose DynActiveGS, a dynamic-aware active reconstruction framework for 3D Gaussian Splatting in dynamic scenes.
- We introduce an uncertainty-aware Gaussian reconstruction strategy that predicts frame-level uncertainty online, suppresses motion-corrupted observations during map optimization, and explicitly disentangles structural uncertainty from motion-induced uncertainty.
- We develop a unified closed-loop system that integrates dynamic-aware viewpoint selection and dynamic-constrained path planning for robust active exploration, and demonstrate strong performance on challenging dynamic reconstruction benchmarks.

2 Related Work

2.1 3D Scene Representation

Scene representation determines the balance among reconstruction fidelity, efficiency, and suitability for online perception. Classical 3D reconstruction is largely built upon structure-from-motion (SfM) pipelines [33], which recover camera poses and scene geometry from image collections or video sequences [20, 31, 55] through feature extraction and matching [44, 46], geometric verification, and bundle adjustment [9]. Although accurate, these methods are computationally expensive and typically assume static scenes. Neural scene representations further advance this line of research: NeRFs [34] and their variants [2, 5, 26] enable high-quality view synthesis but remain costly for real-time and closed-loop deployment. In contrast, 3DGS [21] provides an explicit, differentiable, and efficient representation with real-time rendering and incremental optimization, making it attractive for online reconstruction and robotic perception. Recent extensions to large-scale mapping and SLAM further demonstrate its practical potential, although most existing Gaussian-based systems are still developed under static-scene assumptions.

2.2 Active 3D Reconstruction

Active reconstruction aims to improve scene quality by autonomously selecting informative viewpoints [8, 28]. Early methods formulate next-best-view planning using geometric uncertainty [58], occupancy reasoning, or volumetric information gain [16, 38], while more recent approaches combine learned scene representations with uncertainty estimation to guide exploration [11, 23, 25, 56]. With the emergence of neural and Gaussian-based representations, several methods explore active reconstruction using NeRFs and 3DGS [17, 18, 29, 55]. Despite this progress, existing active reconstruction frameworks are almost exclusively designed for static environments: dynamic objects are usually ignored or treated as noise, which can corrupt optimization, distort uncertainty estimation, and mislead viewpoint selection in dynamic scenes.

2.3 Dynamic Scene Reconstruction

Dynamic environments introduce substantial challenges to 3D reconstruction. Traditional SLAM systems address moving objects through motion segmentation or robust optimization [3, 15], while more recent neural and Gaussian-based approaches explicitly model or filter dynamic content to improve robustness [44, 54, 59]. In particular, uncertainty-aware Gaussian mapping has shown promising performance for dynamic reconstruction. However, these methods mainly focus on passive mapping and tracking, rather than active viewpoint selection and long-horizon exploration [20, 60]. Thus, although they improve robustness under a given trajectory, they do not address how an embodied agent should actively move to acquire reliable and informative observations in dynamic environments. In contrast, our work unifies dynamic scene modeling and active viewpoint planning within a Gaussian splatting framework, enabling robust and efficient active reconstruction in dynamic scenes.

3 Method

As shown in Fig. 2, DynActiveGS operates as a closed-loop active reconstruction system in dynamic environments. At each timestep t, the embodied agent acquires an RGB-D observation (I_t, D_t) and an estimated camera pose $\mathbf{T}_t \in SE(3)$. The system updates the 3D Gaussian scene representation through uncertainty-aware reconstruction, constructs structural uncertainty U_s and motion uncertainty U_m, selects the next informative viewpoint, and executes a

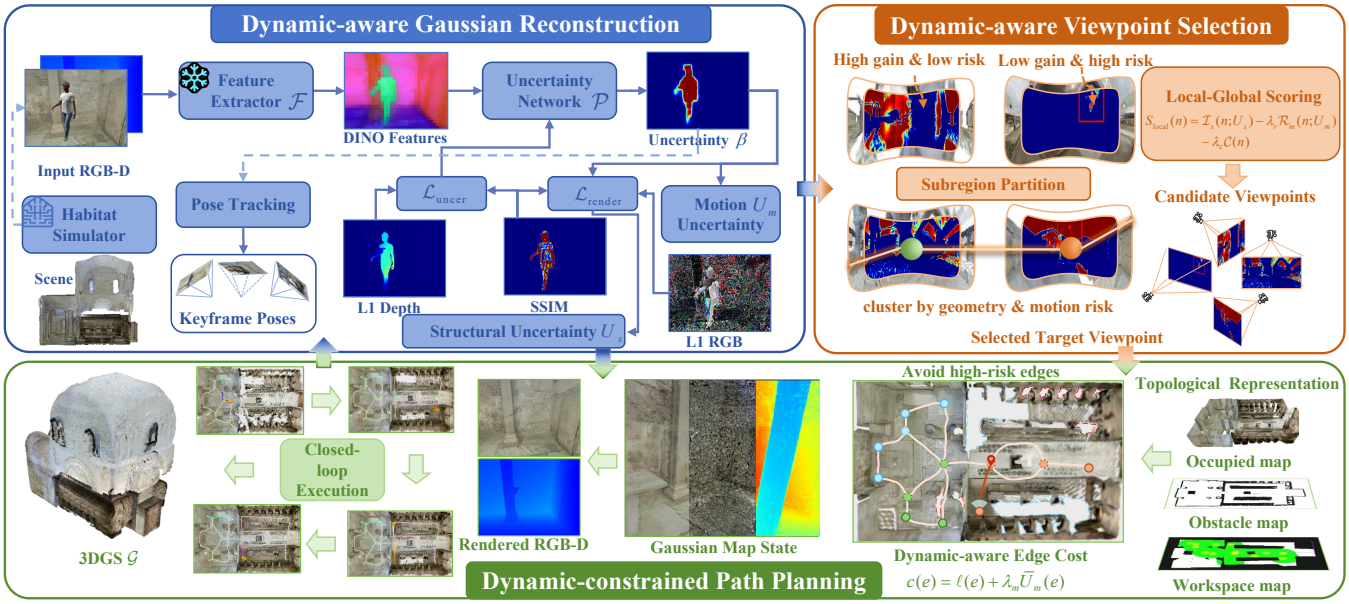

Figure 2: Overview of DynActiveGS. Given RGB-D observations in dynamic environments, DynActiveGS first performs dynamic-aware Gaussian reconstruction by predicting a per-pixel uncertainty map and updating the 3D Gaussian representation through uncertainty-weighted optimization. It then constructs structural uncertainty U_s and motion uncertainty U_m to guide dynamic-aware viewpoint selection via local-global scoring on a Voronoi graph. Finally, a dynamic-constrained path planner computes stable trajectories using motion-aware edge costs, enabling closed-loop active reconstruction under dynamic disturbances.

dynamically stable path toward it. The loop continues until the exploration budget of S steps is exhausted.

3.1 Preliminary: 3D Gaussian Splatting

We represent the scene as a set of anisotropic 3D Gaussian primitives $\mathcal{G} = \{g_i\}_{i=1}^{N}$, where each Gaussian g_i is parameterized by its mean $\boldsymbol{\mu}_i \in \mathbb{R}^3$, color $\mathbf{c}_i \in \mathbb{R}^3$, opacity $\eta_i \in [0,1]$, scale $\mathbf{S}_i \in \mathbb{R}^{3\times3}$, and rotation $\mathbf{R}_i \in \mathrm{SO}(3)$. Its covariance is given by

$$\Sigma_i = \mathbf{R}_i \mathbf{S}_i \mathbf{S}_i^\top \mathbf{R}_i^\top. \tag{1}$$

The rendered color is obtained via alpha compositing:

$$\hat{C}(\mathbf{p}) = \sum_{i=1}^{N} \alpha_i \mathbf{c}_i \prod_{j=1}^{i-1} (1-\alpha_j), \tag{2}$$

where

$$\alpha_i = \eta_i \exp\left(-\tfrac{1}{2}(\mathbf{p}-\boldsymbol{\mu}_i)^\top \Sigma_i^{-1} (\mathbf{p}-\boldsymbol{\mu}_i)\right). \tag{3}$$

This formulation is fully differentiable and supports efficient online optimization.

3.2 Dynamic-aware Gaussian Reconstruction

Dynamic environments violate the static-scene assumption: moving objects and transient occlusions introduce unreliable observations, which can corrupt Gaussian optimization and further mislead active planning. To address this issue, we develop an uncertainty-aware reconstruction strategy for dynamic active reconstruction. Specifically, we predict a per-pixel uncertainty map online, use it to suppress dynamically corrupted observations during Gaussian map optimization, and further lift it into map-level structural and motion uncertainty fields for downstream planning.

3.2.1 Frame-level uncertainty prediction. For each input frame I_t, we extract dense visual features using a pre-trained DINOv3 feature extractor [36, 47]:

$$F_t = \mathcal{F}(I_t), \tag{4}$$

where $\mathcal{F}$ denotes the DINOv3 encoder and F_t is the resulting feature map. These features are fed into a lightweight uncertainty prediction network $\mathcal{P}$ to estimate a per-pixel uncertainty map

$$\beta_t = \mathcal{P}(F_t), \tag{5}$$

where $\beta_t \in \mathbb{R}^{H\times W}$ is bilinearly upsampled to the input resolution. Here, $\beta_t(\mathbf{u})$ measures the reliability of pixel $\mathbf{u}$, and pixels affected by dynamic motion or transient occlusion tend to have higher uncertainty values. The uncertainty predictor is supervised by appearance and depth consistency; the full objective is provided in the supplementary material.

3.2.2 Uncertainty-weighted Gaussian mapping. Let $\mathcal{G}_t$ denote the current Gaussian map. Given the camera pose $\mathbf{T}_t$ and intrinsics $\mathbf{K}$, we render the predicted RGB image $\hat{I}_t$ and depth map $\hat{D}_t$ from $\mathcal{G}_t$. The Gaussian map is optimized using an uncertainty-weighted rendering loss

$$\mathcal{L}_{\text{render}} = \frac{\lambda_1 \mathcal{L}_{\text{color}} + \lambda_2 \mathcal{L}_{\text{depth}}}{\beta_t^2} + \lambda_3 \mathcal{L}_{\text{iso}}, \tag{6}$$

where $\mathcal{L}_{\text{color}}$ combines pixel-wise ℓ_1 and SSIM losses, and $\mathcal{L}_{\text{iso}}$ denotes isotropic Gaussian regularization. Here, the uncertainty map β_t acts as a pixel-wise weighting factor, so that unreliable

observations contribute less to map optimization [27, 43, 59]. This encourages the optimizer to focus on stable static structures while suppressing dynamic distractors during Gaussian updates.

3.2.3 Map-level structural and motion uncertainty. To support dynamic-aware planning, we maintain two map-level uncertainty fields over a discrete 3D query set Ω:

$$U_s : \Omega \to \mathbb{R}_{\geq 0}, \qquad U_m : \Omega \to \mathbb{R}_{\geq 0}, \tag{7}$$

where $U_s(\mathbf{x})$ measures structural uncertainty and $U_m(\mathbf{x})$ measures motion-induced unreliability at location $\mathbf{x}$.

We estimate structural uncertainty from both insufficient observation and reconstruction mismatch. For each $\mathbf{x} \in \Omega$, let $n_t(\mathbf{x})$ denote the effective observation count and let $\bar{e}_t(\mathbf{x})$ denote an aggregated rendering residual. We define

$$U_s(\mathbf{x}) = \lambda_c \frac{1}{\sqrt{n_t(\mathbf{x}) + \epsilon}} + \lambda_r \, \mathrm{clip}(\bar{e}_t(\mathbf{x}), 0, e_{\max}), \tag{8}$$

where the first term highlights under-observed regions and the second term emphasizes locations that are poorly explained by the current Gaussian map.

Motion uncertainty is obtained by lifting the frame-level uncertainty map β_t into 3D and accumulating it over time. For each query location $\mathbf{x} \in \Omega$, let $\mathbf{u}_t = \Pi(\mathbf{T}_t, \mathbf{x})$ denote its projection into frame t, and let $\mathbf{1}_t(\mathbf{x})$ indicate whether it is observable in that frame. We update motion uncertainty using

$$U_m(\mathbf{x}) \leftarrow (1 - \alpha)\, U_m(\mathbf{x}) + \alpha\, \mathbf{1}_t(\mathbf{x})\, \beta_t(\mathbf{u}_t), \tag{9}$$

where $\alpha \in (0, 1)$ controls the update rate. As a result, U_s identifies where more static evidence is needed, while U_m identifies where future observations are likely to be unreliable due to dynamic disturbances.

3.3 Dynamic-aware Viewpoint Selection

Given the map-level structural uncertainty U_s and motion uncertainty U_m, we select viewpoints that are both informative and dynamically reliable. To improve long-horizon exploration efficiency in multi-room environments, we adopt a hierarchical planning strategy on a Voronoi graph [51] $G = (\mathcal{V}, \mathcal{E})$, where each node $n \in \mathcal{V}$ denotes a reachable region and each edge encodes traversal connectivity.

3.3.1 Dynamic-aware subregion partition. Rather than planning over all nodes uniformly, we partition the graph into subregions using agglomerative clustering. To account for dynamic disturbances, the pairwise distance between nodes n_i and n_j is defined as

$$d_{\mathrm{dyn}}(n_i, n_j) = \lambda_e d_E(n_i, n_j) + \lambda_p d_P(n_i, n_j) + \lambda_m d_M(n_i, n_j), \tag{10}$$

where d_E is the Euclidean distance, d_P is the graph travel distance, and d_M is the accumulated motion risk along the connecting path. This produces a set of subregions $\{R_k\}$, including the current local subregion R_l containing the agent.

3.3.2 Local-global viewpoint selection. We prioritize local exploration before switching to global exploration. For each candidate node $n \in R_l$, we define a local score

$$S_{\mathrm{local}}(n) = \mathcal{I}_s(n; U_s) - \lambda_r \mathcal{R}_m(n; U_m) - \lambda_c C(n), \tag{11}$$

where $C(n)$ is the travel cost from the current pose to node n. The structural utility term is

$$\mathcal{I}_s(n; U_s) = \sum_{\mathbf{x} \in \Omega} \mathrm{Vis}(n, \mathbf{x})\, U_s(\mathbf{x}), \tag{12}$$

and the motion risk term is

$$\mathcal{R}_m(n; U_m) = \sum_{\mathbf{x} \in \Omega} \mathrm{Vis}(n, \mathbf{x})\, U_m(\mathbf{x}), \tag{13}$$

where $\mathrm{Vis}(n, \mathbf{x}) \in [0, 1]$ is a soft visibility score under the current Gaussian map $\mathcal{G}_t$.

The best local node is selected iteratively as long as its score remains above a threshold. When the local region is either sufficiently explored or becomes unstable due to high motion uncertainty, the planner switches to global selection. For each candidate subregion $R_k \neq R_l$, we define

$$\begin{aligned} S_{\mathrm{global}}(R_k) = \sum_{n \in R_k} \big(\mathcal{I}_s(n; U_s) - \lambda_r \mathcal{R}_m(n; U_m) \big) \\ - \lambda_d\, d(R_l, R_k) - \lambda_v\, P_{\mathrm{visit}}(R_k). \end{aligned} \tag{14}$$

where $d(R_l, R_k)$ is the inter-region travel cost and $P_{\mathrm{visit}}(R_k)$ penalizes repeatedly visited regions. The region with the highest score is selected as the next exploration target, and a local candidate view action $a_t^\star = (\mathbf{p}_t^\star, \boldsymbol{\theta}_t^\star)$ is chosen around the selected node.

3.4 Dynamic-constrained Path Planning

After selecting the next target viewpoint, we compute a feasible path that jointly accounts for geometry and dynamic risk. Unlike static planning, the objective here is not simply to minimize travel distance, but also to avoid routes likely to produce unstable or motion-corrupted observations.

For each graph edge $e \in \mathcal{E}$, we define a dynamic-constrained traversal cost

$$c(e) = \ell(e) + \lambda_m \bar{U}_m(e), \tag{15}$$

where $\ell(e)$ is the geometric edge length and $\bar{U}_m(e)$ is the average motion uncertainty along the edge. The cost of a path π is then

$$C_{\mathrm{path}}(\pi) = \sum_{e \in \pi} c(e). \tag{16}$$

This encourages the robot to favor dynamically stable routes instead of purely shortest paths.

To improve temporal robustness, motion uncertainty is accumulated over time using an exponential moving average:

$$U_m^{(t)}(\mathbf{x}) = (1 - \alpha)\, U_m^{(t-1)}(\mathbf{x}) + \alpha\, \beta_t(\mathbf{x}), \tag{17}$$

where β_t is the frame-level uncertainty map predicted in Sec. 3.2. As a result, transient disturbances do not dominate planning, while persistently dynamic regions are consistently down-weighted.

By combining dynamic-aware viewpoint selection with motion-constrained path planning, DynActiveGS forms a closed-loop active reconstruction system. At each exploration step, the robot updates the Gaussian map and uncertainty fields from the latest RGB-D observation, selects the next informative target, and executes a dynamically stable path toward it. The loop continues until the exploration step budget S is exhausted, as summarized in the supplementary material Algorithm 1.

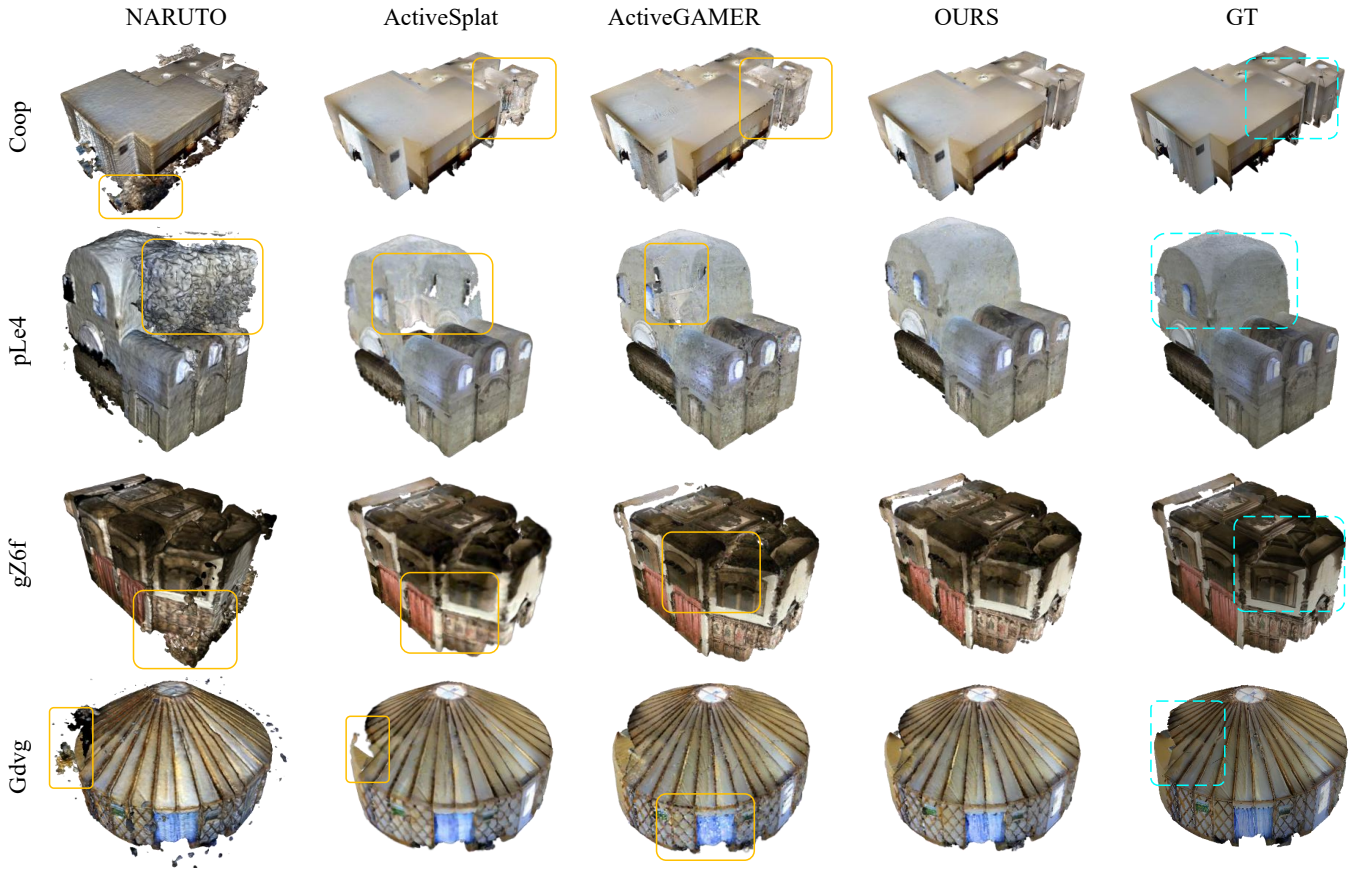

Figure 3: Qualitative comparison of 3D reconstruction results on representative scenes from Social-MP3D and Dynamic Gibson. Blue bounding boxes indicate reference areas for easier comparison, while orange ones highlight low-quality reconstruction.

4 Experiments

4.1 Experimental Setup

Simulator and Dynamic Scenes. All experiments are conducted in the Habitat simulator [45]. To evaluate dynamic active reconstruction, we use indoor scenes with runtime human dynamics in two settings. For Matterport3D (MP3D) [4], we adopt the public Social-MP3D benchmark [12], which augments MP3D scenes with dynamic humans. For Gibson [52], since no public benchmark provides runtime humans for active reconstruction, we construct a reproducible *Dynamic Gibson Protocol* in Habitat-Sim, following the human simulation paradigm of HabiCrowd [49]. In both settings, virtual pedestrians are instantiated at runtime with randomized start–goal waypoints and simulated using ORCA [48] and UPL++ [19, 49] crowd dynamics. The dynamic humans are not part of the static scene geometry used for reconstruction evaluation. We use fixed scene splits, shared human configurations, and standardized random seeds across all compared methods.

Baselines. We compare DynActiveGS against state-of-the-art active reconstruction methods, including 3DGS-based approaches (ActiveGAMER [6], ActiveGS [17], ActiveSplat [29]) and neural active mapping methods (NARUTO [11]). Although some baselines employ free-flying 6DoF cameras, we evaluate them using their official implementations under identical observation budgets. We also include dynamic passive mapping baselines without next-best-view planning.

Geometric Metrics. We evaluate geometric reconstruction using Accuracy (Acc), Completion (Com), and Completion Ratio (C.R.) with a 5 cm threshold. Metrics are computed by uniformly sampling points from static ground-truth meshes and comparing them with point clouds extracted from reconstructed Gaussian maps. Since dynamic humans are instantiated at runtime and are excluded from the ground-truth meshes, these metrics measure robustness of static scene reconstruction under dynamic disturbances.

Rendering Metrics. For novel view rendering evaluation, we follow the protocol of the baseline ActiveGAMER [6] and use same predefined novel-view trajectories for testing. Specifically, we compare the rendered RGB images against static ground-truth renderings at identical poses, and report PSNR, SSIM, and LPIPS. For depth rendering performance, we additionally use Depth L1 distance as the evaluation metric.

Implementation Details. We consider a realistic ground mobile robot with a fixed camera height. At each step, the agent executes one discrete navigation action and acquires one posed RGB-D observation along the planned viewpoint. The camera field of view is

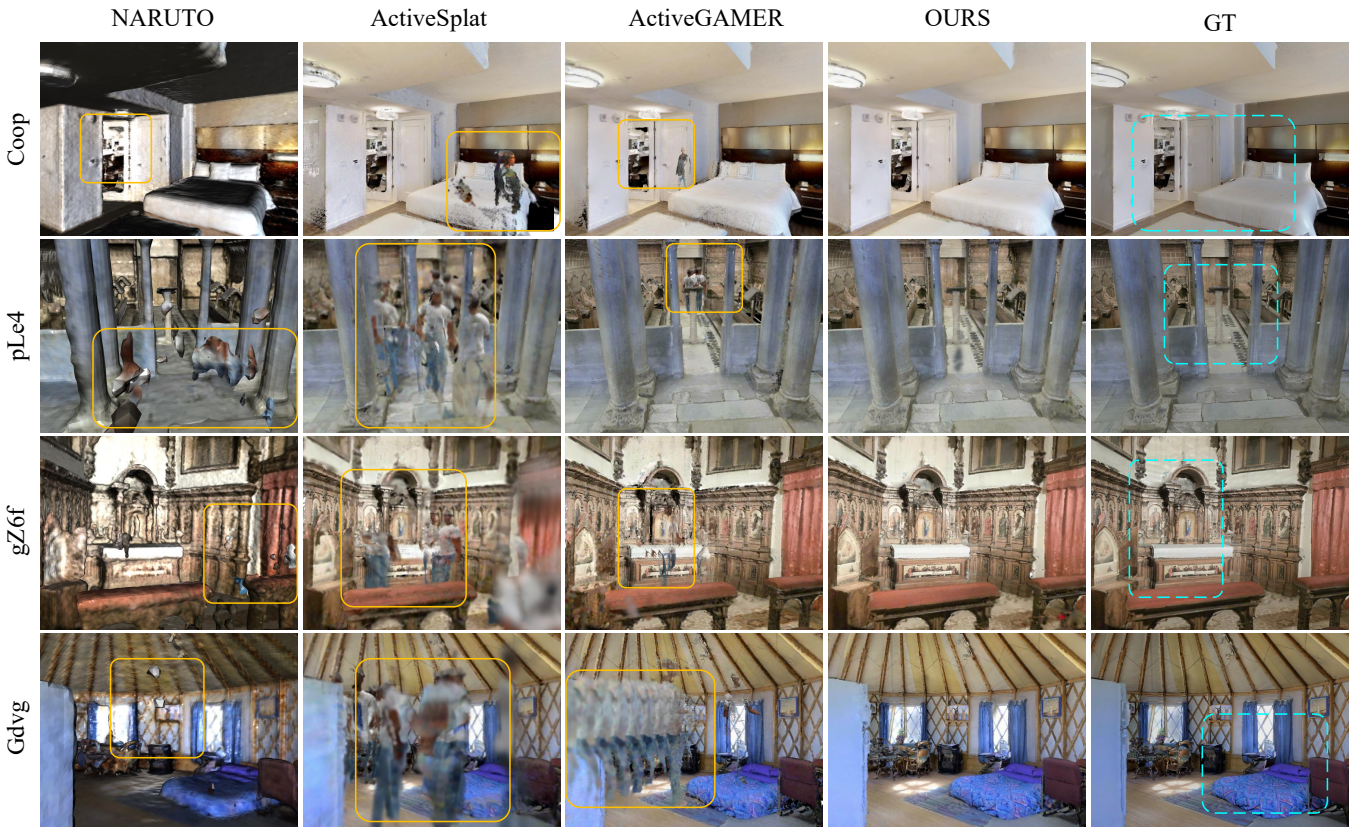

Figure 4: Novel view synthesis results on representative scenes from Social-MP3D and Dynamic Gibson. The evaluated viewpoints are held-out views that do not appear in the training trajectories of any compared method. Blue bounding boxes indicate reference regions for easier comparison, while orange bounding boxes highlight low-quality renderings.

set to 60° vertically and 90° horizontally, and the system processes image sequences online with on-policy planning and incremental reconstruction. Unlike prior works that assume free-flying 6DoF cameras, our embodiment better reflects practical ground robot constraints. All methods are evaluated under an equal observation budget: each method is allowed 1000 exploration steps, and each step produces exactly one RGB-D frame, ensuring identical input image counts across all approaches.

4.2 Comparison Results

Results on Social-MP3D. As shown in Tab. 1, the advantage of DynActiveGS becomes even more evident on the larger and more heavily occluded Social-MP3D scenes. Under the same observation budget, static-scene baselines suffer substantial performance degradation in complex layouts with frequent human occlusions, whereas our method maintains stable reconstruction quality.

Results on Dynamic Gibson. Tab. 2 reports the quantitative results on Gibson scenes augmented with our pedestrian injection protocol. Although Gibson has traditionally been used as a static active reconstruction benchmark, introducing dynamic humans leads to frequent occlusions and motion-corrupted observations, which pose significant challenges to next-best-view planning.

DynActiveGS consistently outperforms all baselines in both geometric and photometric metrics. Compared with the strongest static-scene active reconstruction baseline, our method reduces reconstruction error and improves completion ratio. More importantly, the rendering quality is substantially improved, as reflected by higher PSNR and significantly lower LPIPS.

These results demonstrate that the proposed motion-aware uncertainty modeling generalizes effectively to active reconstruction in dynamic scenes. While static-scene baselines continue to expand coverage, they cannot filter motion-contaminated observations, often resulting in duplicated structures and texture inconsistencies. By contrast, DynActiveGS explicitly disentangles structural uncertainty from motion-induced uncertainty, enabling the agent to avoid high-risk viewpoints and maintain stable map updates.

More importantly, this experiment shows that the gains of our method are not tied to a specific dataset. Even when dynamic agents are introduced through an external augmentation protocol rather than being natively embedded in the dataset, our framework still exhibits strong robustness and consistent reconstruction quality.

Qualitative Comparison. Fig. 3 and Fig. 4 present qualitative comparisons. When humans move across the field of view, static baselines tend to produce duplicated structures and motion streaks.

Table 1: Quantitative comparison on the Social-MP3D dataset for 3D reconstruction and novel view synthesis.

Method	Metric	Gdvg	gZ6f	HxpK	pLe4	YmJk	Avg.
NARUTO [11]	Acc (cm)↓	6.62	5.65	10.87	6.35	13.02	8.50
	Com. (cm)↓	7.91	4.26	9.41	5.98	9.73	7.46
	C.R. (%)↑	82.46	85.11	83.27	79.42	76.76	81.40
	PSNR↑	18.63	18.51	15.46	20.24	16.45	17.86
	SSIM↑	0.624	0.548	0.432	0.697	0.386	0.537
	LPIPS↓	0.431	0.453	0.561	0.482	0.496	0.485
ActiveSplat [29]	Acc (cm)↓	4.64	3.05	5.98	6.74	9.27	5.94
	Com. (cm)↓	4.72	3.48	6.75	4.22	5.91	5.02
	C.R. (%)↑	88.27	87.16	85.54	86.64	85.22	86.57
	PSNR↑	19.65	19.12	21.85	21.57	20.47	20.53
	SSIM↑	0.662	0.646	0.788	0.763	0.655	0.703
	LPIPS↓	0.516	0.552	0.387	0.423	0.458	0.467
ActiveGAMER [6]	Acc (cm)↓	3.71	3.37	3.78	5.43	5.42	4.34
	Com. (cm)↓	5.76	3.91	4.11	6.08	6.64	5.30
	C.R. (%)↑	92.18	92.74	91.46	88.12	86.35	90.17
	PSNR↑	19.76	20.23	22.83	23.07	21.52	21.48
	SSIM↑	0.670	0.693	0.816	0.826	0.721	0.745
	LPIPS↓	0.445	0.420	0.372	0.425	0.386	0.410
Ours	Acc (cm)↓	**2.54**	**2.11**	**3.06**	**2.62**	**3.97**	**2.86**
	Com. (cm)↓	**2.87**	**2.46**	**3.82**	**3.02**	**4.28**	**3.29**
	C.R. (%)↑	**94.41**	**93.02**	**95.58**	**94.77**	**92.63**	**94.08**
	PSNR↑	**21.36**	**20.74**	**26.23**	**24.98**	**21.87**	**23.04**
	SSIM↑	**0.722**	**0.703**	**0.894**	**0.862**	**0.755**	**0.787**
	LPIPS↓	**0.316**	**0.338**	**0.185**	**0.362**	**0.196**	**0.279**

Table 2: Quantitative comparison on the Dynamic Gibson for 3D reconstruction and novel view synthesis.

Method	Metric	Beac	Coop	Denm	Elmi	Eudo	Grei	Home	Ribe	Avg.
NARUTO [11]	Acc (cm)↓	5.42	4.75	4.62	6.31	5.14	5.26	4.58	4.38	5.06
	Com. (cm)↓	5.72	5.36	5.18	7.02	5.72	5.88	5.12	4.95	5.66
	C.R. (%)↑	84.38	86.95	87.42	81.64	85.91	84.72	88.62	88.76	86.05
	PSNR↑	18.94	18.68	17.91	16.85	18.36	16.97	19.31	20.82	18.48
	SSIM↑	0.691	0.737	0.682	0.659	0.676	0.586	0.687	0.763	0.685
	LPIPS↓	0.472	0.449	0.468	0.496	0.531	0.488	0.469	0.456	0.479
ActiveSplat [29]	Acc (cm)↓	5.61	4.52	4.05	4.78	4.18	4.65	3.96	3.87	4.45
	Com. (cm)↓	6.25	5.03	4.58	5.41	4.72	5.16	4.34	4.26	4.97
	C.R. (%)↑	84.26	88.37	89.94	86.91	89.18	87.05	90.36	90.82	88.36
	PSNR↑	22.22	21.05	19.34	20.17	20.88	22.43	21.81	20.72	21.08
	SSIM↑	0.797	0.833	0.789	0.714	0.761	0.736	0.731	0.782	0.768
	LPIPS↓	0.323	0.369	0.519	0.445	0.531	0.339	0.456	0.451	0.429
ActiveGAMER [6]	Acc (cm)↓	5.08	3.72	4.14	4.31	4.08	3.61	3.51	3.42	3.98
	Com. (cm)↓	5.74	4.28	4.71	4.92	4.56	4.15	3.96	3.88	4.53
	C.R. (%)↑	87.84	92.03	90.47	90.12	91.35	92.76	93.11	93.48	91.40
	PSNR↑	23.47	22.11	20.42	21.34	21.96	22.58	23.95	22.08	22.24
	SSIM↑	0.794	0.832	0.747	0.723	0.756	0.729	0.748	0.785	0.764
	LPIPS↓	0.338	0.391	0.486	0.414	0.398	0.366	0.409	0.372	0.397
Ours	Acc (cm)↓	**2.14**	**2.31**	**2.24**	**2.68**	**3.52**	**2.76**	**3.45**	**2.03**	**2.52**
	Com. (cm)↓	**2.56**	**2.78**	**2.71**	**3.26**	**3.04**	**3.35**	**3.98**	**2.48**	**3.02**
	C.R. (%)↑	**96.18**	**95.37**	**95.84**	**93.86**	**94.92**	**93.94**	**91.75**	**96.73**	**94.82**
	PSNR↑	**25.28**	**24.34**	**21.86**	**23.91**	**24.18**	**24.86**	**25.65**	**24.62**	**24.34**
	SSIM↑	**0.866**	**0.858**	**0.811**	**0.771**	**0.842**	**0.835**	**0.841**	**0.852**	**0.835**
	LPIPS↓	**0.254**	**0.243**	**0.339**	**0.262**	**0.298**	**0.286**	**0.311**	**0.307**	**0.288**

DynActiveGS effectively suppresses these artifacts while preserving consistent geometry and texture fidelity. These observations further validate the benefit of risk-aware planning under dynamic disturbances. These results suggest that the primary limitation of conventional active reconstruction in dynamic environments is not insufficient view sampling capacity, but the lack of motion-aware decision-making.

Reconstruction Progress over Exploration. As shown in Fig. 5, DynActiveGS exhibits the fastest improvement in both rendering quality and reconstruction completeness on the Social-MP3D *HxpK* scene. Compared with NARUTO [11], ActiveSplat [29], and ActiveGAMER [6], our method achieves higher PSNR and completion ratio across nearly all exploration stages, further validating the effectiveness of the proposed dynamic-aware active reconstruction pipeline.

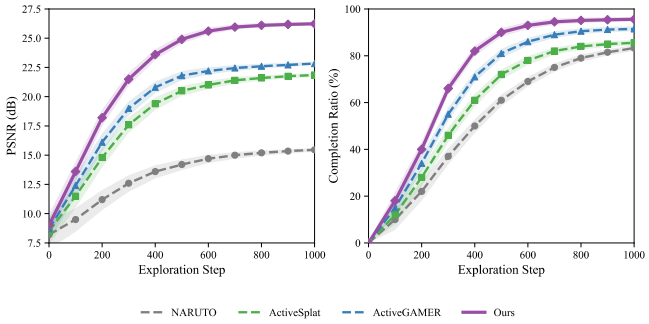

Figure 5: Reconstruction progress on Social-MP3D *HxpK*. DynActiveGS achieves the fastest improvement in both PSNR and completion ratio throughout the exploration process.

Figure 6: Rendering quality across reconstruction stages on Dynamic Gibson *Denm* scene. The proposed uncertainty-aware optimization progressively suppresses dynamic artifacts and improves reconstruction quality from the initial online result to the final refined result.

4.3 Ablation Study

We conduct ablation studies on representative dynamic scenes from Social-MP3D and Dynamic Gibson to validate the key components of DynActiveGS. Our analysis covers three aspects: (1) core module ablation, (2) motion-aware viewpoint selection, and (3) dynamic-constrained planning.

Core module ablation. We first evaluate the contribution of the main modules, including uncertainty-weighted Gaussian mapping, motion-aware viewpoint selection, and dynamic-constrained planning, using the Dynamic Gibson *Denm* scene. As shown in Tab. 3, each component consistently improves both reconstruction and rendering quality. In particular, uncertainty-weighted mapping suppresses motion-corrupted observations, motion-aware viewpoint selection improves observation quality, and dynamic-constrained

Table 3: Ablation of core components. We progressively enable uncertainty-weighted Gaussian mapping, motion-aware viewpoint selection, and dynamic-constrained planning. Each component brings consistent gains, and the full model achieves the best balance between geometric reconstruction and rendering quality.

Variant	Weighted Map	Motion-aware View	Dyn. Planner	Acc↓	Com↓	C.R.↑	PSNR↑	SSIM↑	LPIPS↓	L1-D↓
Baseline	✗	✗	✗	3.21	3.84	90.62	18.71	0.503	0.505	6.94
+ Weighted Mapping	✓	✗	✗	2.98	3.56	91.85	19.83	0.621	0.471	5.12
+ Motion-aware View	✓	✓	✗	2.76	3.28	92.93	20.66	0.733	0.409	4.08
Full (Ours)	✓	✓	✓	**2.24**	**2.71**	**95.84**	**21.86**	**0.811**	**0.339**	**2.92**

Table 4: Ablation of motion-aware viewpoint selection.

Variant	U_s	U_m	Cost	Acc↓	C.R.↑	PSNR↑
Random	✗	✗	✗	3.08	91.24	19.37
Static-Gain	✓	✗	✗	2.81	92.15	20.14
Gain+Cost	✓	✗	✓	2.63	93.08	20.73
Ours	✓	✓	✓	**2.03**	**96.73**	**24.62**

Table 5: Ablation of dynamic-constrained planning.

Path Ratio	Static Hier.			+ Dyn. Edge Cost			**Ours**		
	Com.↓	C.R.↑	PSNR↑	Com.↓	C.R.↑	PSNR↑	Com.↓	C.R.↑	PSNR↑
25%	6.12	88.47	18.95	5.86	89.73	19.42	**5.54**	**90.68**	**20.91**
50%	5.36	90.22	19.71	5.08	91.84	20.74	**4.28**	**92.63**	**21.87**
75%	4.25	91.67	20.82	3.96	92.31	21.26	**3.61**	**93.46**	**22.72**
100%	3.89	92.28	21.97	3.66	93.02	22.63	**2.96**	**94.41**	**24.17**

planning further enhances exploration efficiency. The full model achieves the best overall performance across all metrics.

Ablation of motion-aware viewpoint selection. We compare random viewpoint sampling (*Random*), structural-gain-only selection (*Static-Gain*), structural gain with travel cost (*Gain+Cost*), and our full motion-aware formulation (*Ours*) on the Dynamic Gibson *Ribe* scene. Tab. 4 shows that using structural uncertainty alone is insufficient in dynamic scenes, since the agent is often attracted to regions with high gain but low observation reliability. Introducing travel cost improves efficiency, while explicitly modeling motion uncertainty further improves both reconstruction accuracy and rendering fidelity. These results confirm the importance of motion-aware viewpoint utility under human disturbances.

Ablation of dynamic-constrained planning. We further analyze the planner by comparing a static hierarchical planner (*Static Hier.*), a planner with dynamic edge cost but without temporal accumulation (*+ Dyn. Edge Cost*), and our full dynamic-constrained planner (*Ours*). To evaluate long-horizon efficiency, we extend the exploration horizon to 2000 steps on the Social-MP3D *YmJk* scene and report results at different path ratios in Tab. 5. Hierarchical planning already improves long-horizon exploration over naive shortest-path execution, while dynamic edge cost reduces unstable routes through highly dynamic regions. Temporal accumulation yields the best final performance, suggesting that persistent motion patterns are more informative for planning than instantaneous disturbances.

Stage-wise rendering analysis. Fig. 6 visualizes reconstruction quality at different stages of the pipeline. The initial online result (*Online Before Opt*) still contains strong dynamic artifacts. After uncertainty-guided online optimization (*Online After Opt*), these artifacts are substantially reduced. The final offline refinement further improves consistency and produces the best visual quality. Together with the uncertainty visualization, this result shows that the learned uncertainty map effectively highlights motion-corrupted regions and supports robust Gaussian optimization.

Overall, these ablations consistently validate the design of DynActiveGS. Uncertainty-weighted mapping improves reconstruction robustness, motion-aware viewpoint selection improves observation quality, and dynamic-constrained planning enhances long-horizon exploration efficiency. Together, these components enable robust active reconstruction in dynamic environments.

5 Conclusion and Future Work

We introduced **DynActiveGS**, a dynamic-aware active reconstruction framework based on 3D Gaussian Splatting. By explicitly disentangling structural uncertainty from motion-induced uncertainty, and by integrating dynamic-aware viewpoint selection with dynamic-constrained path planning, our method enables robust closed-loop exploration and high-fidelity reconstruction in dynamic environments. Extensive experiments on Social-MP3D and dynamically augmented Gibson scenes demonstrate clear and consistent improvements over existing baselines in reconstruction accuracy, completeness, and rendering quality.

Limitations. Our current evaluation is conducted in simulation with controllable human dynamics, whereas real-world environments may exhibit more diverse, irregular, and less predictable motion patterns. In addition, the current framework has not been explicitly designed for active reconstruction in large-scale multi-floor environments with more complex spatial connectivity.

Future Work. Promising directions for future research include deployment on real robotic platforms, the integration of online motion prediction, and extending dynamic-aware active reconstruction toward 4D scene representations [32] to model persistent scene dynamics over time. We believe that dynamic-aware active reconstruction is an important step toward more capable embodied perception systems in real-world environments.

References

[1] John Aloimonos, Isaac Weiss, and Amit Bandyopadhyay. 1988. Active vision. *International journal of computer vision* 1, 4 (1988), 333–356.

[2] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5470–5479.

[3] Berta Bescos, José M Fácil, Javier Civera, and José Neira. 2018. DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes. *IEEE robotics and automation letters* 3, 4 (2018), 4076–4083.

[4] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niebner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. 2018. Matterport3D: Learning from RGB-D data in indoor environments. In *7th IEEE International Conference on 3D Vision, 3DV 2017*. 667–676.

[5] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. 2022. Tensorf: Tensorial radiance fields. In *European conference on computer vision*. Springer, 333–350.

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

[7] Shengyong Chen, Youfu Li, and Ngai Ming Kwok. 2011. Active vision in robotic systems: A survey of recent developments. *The International Journal of Robotics Research* 30, 11 (2011), 1343–1377.

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

[9] Shin-Fang Chng, Sameera Ramasinghe, Jamie Sherrah, and Simon Lucey. 2022. Gaussian activated neural radiance fields for high fidelity reconstruction and pose estimation. In *European Conference on Computer Vision*. Springer, 264–280.

[10] Angela Dai, Matthias Nießner, Michael Zollhöfer, Shahram Izadi, and Christian Theobalt. 2017. Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface reintegration. *ACM Transactions on Graphics (ToG)* 36, 4 (2017), 1.

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

[12] Zeying Gong, Tianshuai Hu, Ronghe Qiu, and Junwei Liang. 2025. From cognition to precognition: A future-aware framework for social navigation. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 9122–9129.

[13] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2024. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 conference papers*. 1–11.

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

[15] Haochen Jiang, Yueming Xu, Kejie Li, Jianfeng Feng, and Li Zhang. 2024. Rodyn-slam: Robust dynamic dense rgb-d slam with neural radiance fields. *IEEE Robotics and Automation Letters* 9, 9 (2024), 7509–7516.

[16] Wen Jiang, Boshu Lei, and Kostas Daniilidis. 2024. Fisherrf: Active view selection and mapping with radiance fields using fisher information. In *European Conference on Computer Vision*. Springer, 422–440.

[17] Liren Jin, Xingguang Zhong, Yue Pan, Jens Behley, Cyrill Stachniss, and Marija Popović. 2025. Activegs: Active scene reconstruction using gaussian splatting. *IEEE Robotics and Automation Letters* (2025).

[18] Rui Jin, Yuman Gao, Yingjian Wang, Yuze Wu, Haojian Lu, Chao Xu, and Fei Gao. 2024. 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)*. IEEE, 11202–11209.

[19] Ioannis Karamouzas, Brian Skinner, and Stephen J Guy. 2014. Universal power law governing pedestrian interactions. *Physical review letters* 113, 23 (2014), 238701.

[20] Nikhil Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian Scherer, Deva Ramanan, and Jonathon Luiten. 2024. Splatam: Splat track & map 3d gaussians for dense rgb-d slam. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 21357–21366.

[21] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* 42, 4 (2023), 139–1.

[22] Andreas Kirsch, Joost Van Amersfoort, and Yarin Gal. 2019. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. *Advances in neural information processing systems* 32 (2019).

[23] Zijia Kuang, Zike Yan, Hao Zhao, Guyue Zhou, and Hongbin Zha. 2024. Active neural mapping at scale. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 7152–7159.

[24] Jonas Kulhanek, Songyou Peng, Zuzana Kukelova, Marc Pollefeys, and Torsten Sattler. 2024. WildGaussians: 3D Gaussian Splatting In the Wild. *Advances in Neural Information Processing Systems* 37 (2024), 21271–21288.

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

[26] Kejie Li, Yansong Tang, Victor Adrian Prisacariu, and Philip HS Torr. 2022. Bnv-fusion: Dense 3d reconstruction using bi-level neural volume fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6166–6175.

[27] Moyang Li, Zihan Zhu, Marc Pollefeys, and Daniel Barath. 2026. DROID-SLAM in the Wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[28] Shiyao Li, Antoine Guédon, Clémentin Boittiaux, Shizhe Chen, and Vincent Lepetit. 2025. NextBestPath: Efficient 3D Mapping of Unseen Environments. In *ICLR 2025-Thirteenth International Conference on Learning Representations*.

[29] Yuetao Li, Zijia Kuang, Ting Li, Qun Hao, Zike Yan, Guyue Zhou, and Shaohui Zhang. 2025. ActiveSplat: High-Fidelity Scene Reconstruction through Active Gaussian Splatting. *IEEE Robotics and Automation Letters* 10, 8 (Aug. 2025), 8099–8106. doi:10.1109/LRA.2025.3580331 arXiv:2410.21955 [cs].

[30] Yan Li, Yingzhao Li, and Gim Hee Lee. 2025. Active3D: Active High-Fidelity 3D Reconstruction via Hierarchical Uncertainty Quantification. doi:10.48550/arXiv.2511.20050 arXiv:2511.20050 [cs].

[31] Yanyan Li, Chenyu Lyu, Yan Di, Guangyao Zhai, Gim Hee Lee, and Federico Tombari. 2024. Geogaussian: Geometry-aware gaussian splatting for scene rendering. In *European conference on computer vision*. Springer, 441–457.

[32] Chenguo Lin, Yuchen Lin, Panwang Pan, Yifan Yu, Tao Hu, Honglei Yan, Katerina Fragkiadaki, and Yadong Mu. 2026. MoVieS: Motion-Aware 4D Dynamic View Synthesis in One Second. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*.

[33] Jiachen Liu, Pan Ji, Nitin Bansal, Changjiang Cai, Qingan Yan, Xiaolei Huang, and Yi Xu. 2022. Planemvs: 3d plane reconstruction from multi-view stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8665–8675.

[34] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (2021), 99–106.

[35] Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. 2011. Kinectfusion: Real-time dense surface mapping and tracking. In *2011 10th IEEE international symposium on mixed and augmented reality*. Ieee, 127–136.

[36] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. 2023. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193* (2023).

[37] Emanuele Palazzolo, Jens Behley, Philipp Lottes, Philippe Giguere, and Cyrill Stachniss. 2019. ReFusion: 3D reconstruction in dynamic environments for RGB-D cameras exploiting residuals. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 7855–7862.

[38] Xuran Pan, Zihang Lai, Shiji Song, and Gao Huang. 2022. Activenerf: Learning where to see with uncertainty estimation. In *European Conference on Computer Vision*. Springer, 230–246.

[39] Daryl Peralta, Joel Casimiro, Aldrin Michael Nilles, Justine Aletta Aguilar, Rowel Atienza, and Rhandley Cajote. 2020. Next-best view policy for 3d reconstruction. In *European Conference on Computer Vision*. Springer, 558–573.

[40] Richard Pito. 2002. A solution to the next best view problem for automated surface acquisition. *IEEE Transactions on pattern analysis and machine intelligence* 21, 10 (2002), 1016–1030.

[41] Konstantinos D. Polyzos, Athanasios Bacharis, Saketh Madhuvarasu, Nikos Papanikolopoulos, and Tara Javidi. 2025. ActiveInitSplat: How Active Image Selection Helps Gaussian Splatting. doi:10.48550/arXiv.2503.06859 arXiv:2503.06859 [cs].

[42] Yunlong Ran, Jing Zeng, Shibo He, Jiming Chen, Lincheng Li, Yingfeng Chen, Gimhee Lee, and Qi Ye. 2023. Neurar: Neural uncertainty for autonomous 3d reconstruction with implicit neural representations. *IEEE Robotics and Automation Letters* 8, 2 (2023), 1125–1132.

[43] Weining Ren, Zihan Zhu, Boyang Sun, Jiaqi Chen, Marc Pollefeys, and Songyou Peng. 2024. Nerf on-the-go: Exploiting uncertainty for distractor-free nerfs in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8931–8940.

[44] Erik Sandström, Ganlin Zhang, Keisuke Tateno, Michael Oechsle, Michael Niemeyer, Youmin Zhang, Manthan Patel, Luc Van Gool, Martin Oswald, and Federico Tombari. 2025. Splat-slam: Globally optimized rgb-only slam with 3d gaussians. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 1680–1691.

[45] Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, et al. 2019. Habitat: A platform for embodied ai research. In *Proceedings of the IEEE/CVF international conference on computer vision*. 9339–9347.

[46] Johannes L Schonberger and Jan-Michael Frahm. 2016. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4104–4113.

[47] Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. 2025. Dinov3. *arXiv preprint arXiv:2508.10104* (2025).

[48] Jur Van Den Berg, Stephen J Guy, Ming Lin, and Dinesh Manocha. 2011. Reciprocal n-body collision avoidance. In *Robotics research: the 14th international symposium ISRR*. Springer, 3–19.

[49] An Vuong, Toan Nguyen, Minh Nhat Vu, Baoru Huang, Huynh Thi Thanh Binh, Thieu Vo, and Anh Nguyen. 2024. Habicrowd: A high performance simulator for crowd-aware visual navigation. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 5821–5827.

[50] Thomas Whelan, Stefan Leutenegger, Renato F Salas-Moreno, Ben Glocker, and Andrew J Davison. 2015. ElasticFusion: Dense SLAM without a pose graph.. In *Robotics: science and systems*, Vol. 11. Rome, Italy.

[51] Pengying Wu, Yao Mu, Bingxian Wu, Yi Hou, Ji Ma, Shanghang Zhang, and Chang Liu. 2024. VoroNav: Voronoi-based Zero-shot Object Navigation with Large Language Model. In *International Conference on Machine Learning*. PMLR, 53757–53775.

[52] Fei Xia, Amir R Zamir, Zhiyang He, Alexander Sax, Jitendra Malik, and Silvio Savarese. 2018. Gibson env: Real-world perception for embodied agents. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 9068–9079.

[53] Yuhan Xie, Yixi Cai, Yinqiang Zhang, Lei Yang, and Jia Pan. 2025. GauSS-MI: Gaussian Splatting Shannon Mutual Information for Active 3D Reconstruction. In *Proceedings of Robotics: Science and Systems*. LosAngeles, CA, USA. doi:10.15607/RSS.2025.XXI.030

[54] Yueming Xu, Haochen Jiang, Zhongyang Xiao, Jianfeng Feng, and Li Zhang. 2024. Dg-slam: Robust dynamic gaussian splatting slam with hybrid pose optimization. *Advances in Neural Information Processing Systems* 37 (2024), 51577–51596.

[55] Zijun Xu, Rui Jin, Ke Wu, Yi Zhao, Zhiwei Zhang, Jieru Zhao, Fei Gao, Zhongxue Gan, and Wenchao Ding. 2025. HGS-Planner: Hierarchical Planning Framework for Active Scene Reconstruction Using 3D Gaussian Splatting. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*. 14161–14167. doi:10.1109/ICRA55743.2025.11127649

[56] Zike Yan, Haoxiang Yang, and Hongbin Zha. 2023. Active Neural Mapping. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Paris, France, 10947–10958. doi:10.1109/ICCV51070.2023.01008

[57] Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. 2024. Mip-splatting: Alias-free 3d gaussian splatting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 19447–19456.

[58] Zhengquan Zhang, Feng Xu, and Mengmi Zhang. 2025. Peering into the Unknown: Active View Selection with Neural Uncertainty Maps for 3D Reconstruction. *arXiv preprint arXiv:2506.14856* (2025).

[59] Jianhao Zheng, Zihan Zhu, Valentin Bieri, Marc Pollefeys, Songyou Peng, and Iro Armeni. 2025. Wildgs-slam: Monocular gaussian splatting slam in dynamic environments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11461–11471.

[60] Zihan Zhu, Wei Zhang, Norbert Haala, Marc Pollefeys, and Daniel Barath. 2025. VIGS-SLAM: Visual Inertial Gaussian Splatting SLAM. doi:10.48550/arXiv.2512.02293 arXiv:2512.02293 [cs].