

ActiveSplat: High-Fidelity Scene Reconstruction through Active Gaussian Splatting

Yuetao Li^{1,2*}, Zijia Kuang^{2*}, Ting Li², Qun Hao¹, Zike Yan^{2†}, Guyue Zhou^{2,3}, Shaohui Zhang^{1†}

Abstract—We propose ActiveSplat, an autonomous high-fidelity reconstruction system leveraging Gaussian splatting. Taking advantage of efficient and realistic rendering, the system establishes a unified framework for online mapping, viewpoint selection, and path planning. The key to ActiveSplat is a hybrid map representation that integrates both dense information about the environment and a sparse abstraction of the workspace. Therefore, the system leverages sparse topology for efficient viewpoint sampling and path planning, while exploiting view-dependent dense prediction for viewpoint selection, facilitating efficient decision-making with promising accuracy and completeness. A hierarchical planning strategy based on the topological map is adopted to mitigate repetitive trajectories and improve local granularity given limited time budgets, ensuring high-fidelity reconstruction with photorealistic view synthesis. Extensive experiments and ablation studies validate the efficacy of the proposed method in terms of reconstruction accuracy, data coverage, and exploration efficiency. The released code will be available on our project page¹.

Index Terms—Autonomous Agents, Mapping, RGB-D perception

I. INTRODUCTION

FINE-GRAINED reconstruction of three-dimensional environments has long been a central research focus in robotics, computer vision, and computer graphics. Within the robotics community, there is a growing demand for high-fidelity digitization of the physical world, not only to facilitate immersive applications like teleoperation [1], but also to narrow the sim-to-real gap, advancing generalizable robot autonomy through photorealistic simulation [2].

Recent progress in differentiable rendering has significantly improved the quality of the reconstructed environments. Neural radiance fields (NeRF) [3] and its variant [4] leverage neural networks as compact scene representations, using volume rendering to synthesize high-quality novel views. However, the computational inefficiencies caused by volume integration

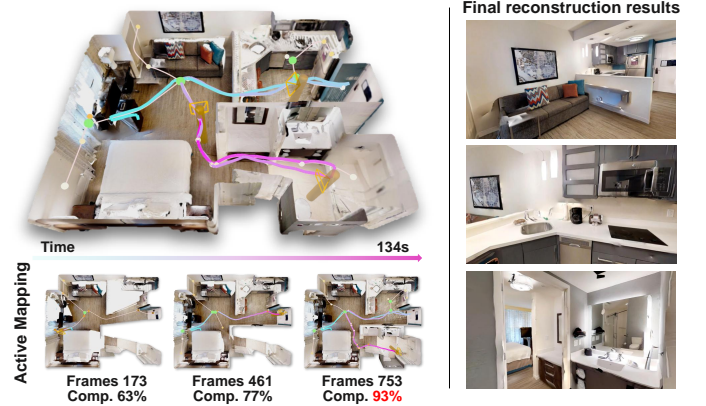


Fig. 1. The agent explores the environment autonomously to build a 3D map on the fly. The integration of a Gaussian map and a Voronoi graph ensures efficient and complete exploration, resulting in high-fidelity reconstruction and a high completion ratio (Comp. %).

along rays pose challenges in terms of memory and processing. To address these limitations, Gaussian splatting [5], [6] has been introduced, enabling efficient rasterization and achieving promising rendering quality through α -blending. Despite these advances, scene-specific and data-dependent optimization makes these methods highly sensitive to noise and artifacts, which often emerge due to insufficient view coverage, especially without ground-truth supervision during data collection.

In this work, we aim to address these issues through active mapping, where a mobile agent reconstructs the environment on the fly, assesses the instant quality of the map, and plans its path to cover the entire environment. We find Gaussian splatting to be particularly suitable for high-fidelity active mapping, owing to its capability for *view-dependent dense predictions*. This characteristic enables the system to efficiently and accurately extract the workspace² within the horizontal plane, while also quantifying data coverage in a unified manner by splatting Gaussians of interest onto the actively sampled views. The proposed system, dubbed **ActiveSplat**, incrementally updates a renderable Gaussian map through gradient-based optimization, progressively refining and completing the scene representation with high fidelity.

To balance reconstruction accuracy and exploration efficiency, our system adopts a hybrid map representation inspired by [7], but replaces the volumetric neural fields with explicit 3D Gaussians, enabling significantly faster convergence and

Manuscript received February 22, 2025; accepted May 27, 2025. This paper was recommended for publication by Editors Tamim Asfour and Sven Behnke upon evaluation by the Associate Editor and reviewers.

This work was supported by Development and Ecosystem Construction of Robot Simulation and End-to-End Framework, National Key Research and Development Program of China (2021YFC2202404), and National Natural Science Foundation of China (62275020).

*Equal contributions, †Equal advising.

¹School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China.

²Institute for AI Industry Research (AIR), Tsinghua University, Beijing 100084, China.

³School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China. Digital Object Identifier (DOI): 10.1109/LRA.2025.3580331.

¹<https://li-yuetao.github.io/ActiveSplat/>

²The workspace refers to the navigable free space from a top-down perspective (see Sec. III-A).

real-time rendering essential for online mapping. A set of 3D Gaussians is maintained as a dense map to provide view-dependent dense predictions, while a Voronoi graph is extracted as a topological map to represent the abstraction of the workspace. Sparse yet representative view positions are derived from this graph, guiding the agent to extend the boundaries of the workspace. Meanwhile, the viewing orientation at each position is determined by view-dependent completeness measures of the Gaussian map. Based on this, a viewpoint decoupling method is proposed to reduce the infinite number of possible viewpoints in free space to a manageable set of positions and rotation angles, ensuring efficient and safe traversal. Additionally, a hierarchical planning strategy based on the topological map is employed to reduce redundant trajectories during global exploration and improve the overall efficiency of the autonomy process. The key contributions of the letter can then be categorized as follows:

- A novel system that actively splats Gaussians of interest to build a unified, autonomous, and high-fidelity reconstruction system.
- A hybrid map representation combining dense predictions with Gaussians and sparse abstraction of Voronoi graph for comprehensive viewpoint selection and safe path planning.
- A hierarchical planning strategy based on the Voronoi graph, which prioritizes local areas to minimize redundant exploration, decoupling viewpoint selection to balance exploration efficiency and reconstruction accuracy.

II. RELATED WORK

A. Autonomous Exploration

In the robotic community, autonomous exploration aims to best acquire observations to cover the entire space traversed by the robot. Existing strategies seek to balance exploration completeness and efficiency, and can be broadly categorized as frontier-based methods and sampling-based methods. Frontier-based methods [8], [9] focus on expanding the exploration area by navigating to the boundary between explored and unexplored regions until full coverage is achieved. However, these methods rely on the discrete grid representation to discern the decision boundary, thus lacking adaptive granularity given diverse geometry complexity. In contrast, sampling-based methods [7], [10], [11] sample candidate viewpoints and prioritize those that maximize uncertainty reduction or expected information gain, thus improving scene coverage by reducing environmental uncertainty. Efforts are made to design proper sampling strategies for efficiency and precise scoring techniques given the samples. TARE [12] introduces a hierarchical strategy for LiDAR-based exploration, where a local subspace is traversed at a fine-grained level while global target goals at a coarse level are maintained, thus achieving a balance between exploration efficiency and mapping completeness. Similarly, [13] adopts fine-grained next-best-view planning for local exploration, while leveraging frontier-based strategies for global coverage. Topology maps have become widely used as sparse scene abstractions in autonomous exploration [7], [14]. Recently, most relevant to ours is [7], which also leverages

a hybrid representation containing a dense neural map and a topology map for exploration. However, the neural map suffers from slow convergence and inefficient volume rendering. We, on the other hand, leverage the efficient optimization and rendering of Gaussian primitives to achieve reconstruction with high fidelity. Furthermore, we adopt a decoupled viewpoint selection to separately address translation and rotation during exploration, along with a topology-based hierarchical planning strategy, ultimately achieving a balance between exploration efficiency and completeness.

B. High-Fidelity Scene Reconstruction

Recent progress in differentiable rendering attracts significant attention in the research community. Parameterized by implicit NeRF representations [3], [4] or explicit 2D/3D Gaussian representations [5], [6], photorealistic images of novel views can be rendered with promising efficiency. Gradient-based optimization has also been applied in an online setting to incrementally update the neural map [15], [16] or Gaussian parameters [17]–[19] through differentiable rendering. Recently, continual learning of the neural map has turned into an active fashion through uncertainty-guided autonomous exploration [7], [11], [20]–[22]. However, the neural representation has to sacrifice model capacity to achieve real-time performance, requiring inevitable tradeoffs between accuracy and convergence efficiency with different network architectures. In contrast, we adopt a set of Gaussian primitives as the scene representation, which allows consistent optimization of Gaussian parameters in an online update or offline post-processing setting. Relevant works have also developed Gaussian splatting-based mapping systems using unmanned ground vehicle (UGV) [23]–[26] and unmanned aerial vehicle (UAV) [27], [28] platforms. Notably, both the safe navigation system [23] and HGS-Planner [25] build upon FisherRF [29], which utilizes Fisher information for quantifying the uncertainty of Gaussians. ActiveGS [28] tackles similar challenges by modeling Gaussian confidence and integrating splatting with voxel maps to improve scene reconstruction. The major difference lies in our proposed hybrid map representation, which enforces safe and hierarchical path planning based on a topological map.

III. METHODOLOGY

Our ActiveSplat system is a Gaussian splatting-based active mapping framework that aims to maximize scene completeness and reconstruction accuracy through autonomous exploration using posed RGB-D input. The hybrid map and hierarchical planning are introduced to improve exploration efficiency within a limited number of steps, balancing the tradeoff between exploration path length and reconstruction completeness. The overview of the system is illustrated in Fig. 2, which shows that Gaussians of interest are splatted onto the image plane, serving as a consistent technique for online map updating (Sec. III-A), viewpoint selection (Sec. III-B), and path planning (Sec. III-C).

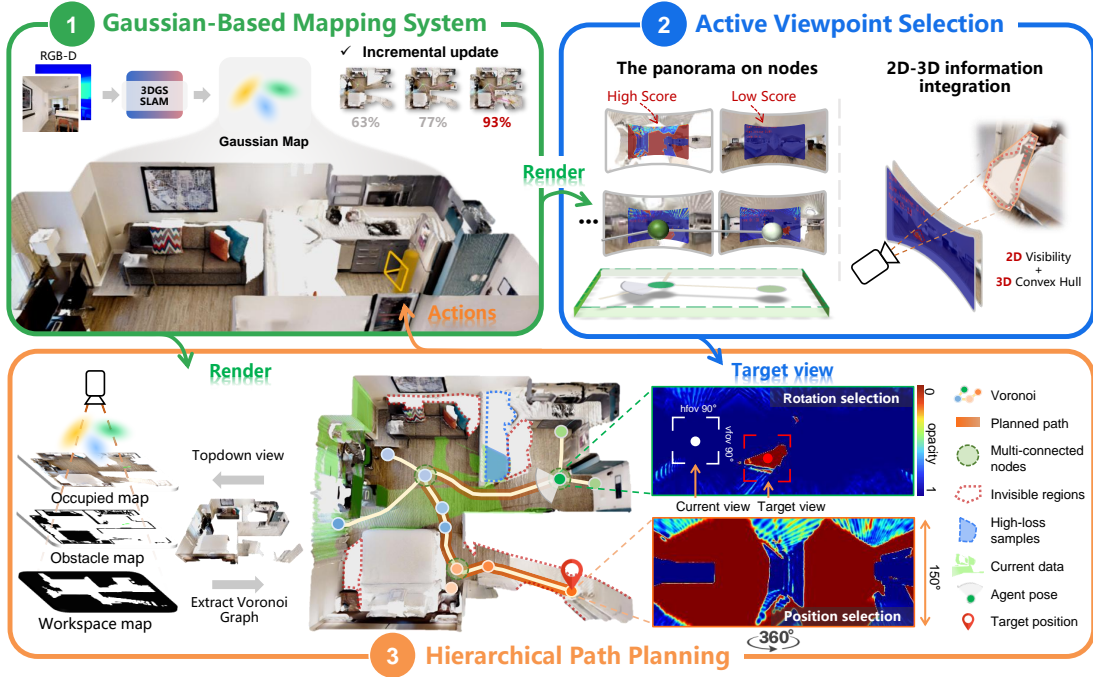


Fig. 2. **Overview of ActiveSplat:** The proposed active mapping system achieves high-fidelity reconstruction through a perception-action closed loop, utilizing a hybrid map that combines dense Gaussians with topological abstractions. Splatting Gaussians of interest on the fly provides a consistent approach for **online map updating**, **viewpoint selection**, and **path planning**. **Note:** Subregions are distinguished by node color, with node scores indicated by color intensity.

A. Hybrid Map Updating

Central to the proposed ActiveSplat system is a hybrid map representation containing both Gaussian primitives, which allow dense prediction, and a topological structure, which provides sparse abstraction of the workspace. A Gaussian primitive is an explicit representation parameterized by color \mathbf{c} , center position $\boldsymbol{\mu}$, anisotropic covariance Σ , and opacity o , where the influence of each Gaussian can be expressed as:

$$f_i(\mathbf{u}_k) = o \cdot \exp\left(-\frac{1}{2}(\mathbf{x}(\mathbf{u}_k) - \boldsymbol{\mu}_i)^\top \Sigma^{-1}(\mathbf{x}(\mathbf{u}_k) - \boldsymbol{\mu}_i)\right). \quad (1)$$

View synthesis can then be implemented through splatting given the Gaussian map and a camera pose, where the color of each pixel \mathbf{u} is linearly affected by the projected 3D Gaussians as:

$$\hat{C}_k = \sum_{i=1}^{n_k} \mathbf{c}_i f_i(\mathbf{u}_k) \prod_{j=1}^{i-1} (1 - f_j(\mathbf{u}_k)). \quad (2)$$

Similarly, the differentiable rendering can also be applied for depth and visibility (accumulated opacity) estimation:

$$\hat{D}_k = \sum_{i=1}^{n_k} d_i f_i(\mathbf{u}_k) \prod_{j=1}^{i-1} (1 - f_j(\mathbf{u}_k)), \quad (3)$$

$$\hat{O}_k = \sum_{i=1}^{n_k} f_i(\mathbf{u}_k) \prod_{j=1}^{i-1} (1 - f_j(\mathbf{u}_k)), \quad (4)$$

where d_i is the depth of the Gaussian center in the camera coordinate.

The optimization of the Gaussian map is performed given photometric and geometric losses defined in [17]:

$$L_{pho} = \lambda_1 |C_k - \hat{C}_k| + \lambda_2 (1 - \text{SSIM}(C_k, \hat{C}_k)), \quad (5)$$

$$L_{geo} = |D_k - \hat{D}_k|, \quad (6)$$

$$L = w_c L_{pho} + w_d L_{geo}, \quad (7)$$

where $\lambda_1 = 0.8$, $\lambda_2 = 0.2$, $w_c = 0.5$, $w_d = 1.0$ and C_k, D_k are the captured RGB-D images of the k -th frame.

During the online mapping process, new Gaussians are dynamically initialized to cover newly observed areas, and redundant Gaussians with near-zero opacity or large covariances are removed as in [5]. Following [17], areas with low accumulated opacity or geometric deviations are identified as newly-observed areas:

$$M_k = \left(\hat{O}_k < \tau_{o1}\right) \vee \left((D_k < \hat{D}_k) \wedge (|D_k - \hat{D}_k| > \epsilon_{\text{MDE}})\right), \quad (8)$$

where $\tau_{o1} = 0.98$, and ϵ_{MDE} represents 50 times median depth error.

The dense prediction of the Gaussian map allows convenient extraction of the workspace and the obstacles. As illustrated in Box 3 of Fig. 2, the top-down view can be efficiently rendered using a large focal length as the orthographic projection of the dense map. The region with sufficient accumulated opacity is taken as occupied, where areas above the ground and within the agent's height represent the obstacles. Navigable workspace can then be extracted as occupied areas at the ground level, excluding the obstacles. A Voronoi graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is generated through Voronoi tessellation [30], with edges \mathcal{E} being equidistant from the obstacles and nodes \mathcal{N} as intersections where the edges terminate.

The Gaussian map and the Voronoi graph are complementary: The Gaussian map provides dense and complete information about previously visited regions of the scene, while the Voronoi graph offers a sparse structure of the workspace,

which can also be seen as a strong deformation retract of the global free space [31] in topology. We show in the following that the integration of the two representations leads to an adaptive granularity of the environment and guarantees an effective trade-off between efficiency and accuracy during the autonomous reconstruction process.

B. Active Viewpoint Selection

The objective of active mapping is to traverse the workspace and best capture the information of previously unseen areas. This is usually achieved by iteratively selecting target views, where the sampling strategy of accessible viewpoint candidates and the selection criteria are crucial to efficiency and overall coverage. We sample viewpoints on the Voronoi graph to maintain a compact and accessible set that covers the entire scene. Besides, the Voronoi graph generates a path that stays as far away from the obstacles as possible, thus guaranteeing a safe traversal [7].

1) *Decoupled Position and Rotation Candidates*: To best exploit the compact structure of the Voronoi graph along with the dense geometry and appearance of the Gaussian map, we propose decoupling the position and rotation candidates, rather than uniformly [25] or randomly [28] sampling viewpoints across the workspace. This approach reduces the search dimensionality while ensuring thorough observations. The dynamically updated map leads to a progressively refined graph that completely describes the partially-observed workspace. We iteratively select Voronoi nodes as viewpoint position candidates, where the node with the highest score (calculated as in Eq. 10) favors unvisited areas that push the boundary of the workspace for traversal. Regarding the view rotation, we adopt the yaw and pitch rotations at the selected view positions to get the observation toward a specific region. The target node position and the target rotation angle are determined in a view-dependent manner as follows.

2) *Coverage Evaluation*: The actions regarding translation and rotation during the autonomous exploration process undergo different granularity. We aim to efficiently traverse the entire set of Voronoi nodes to maintain complete coverage, while conducting careful inspection in an area with intricate intersections of paths [7]. In practice, we splat the constructed Gaussians to quickly obtain surrounding information of a specified pose, render panoramic images of the visibility at all unvisited nodes, as illustrated in Fig. 2. To generate a sparse and representative set of rotation candidates, we apply the DBSCAN algorithm to cluster low-visibility regions (highlighted in red) within the panoramic view, where the pixel coordinate of the cluster center indicates yaw and pitch angles, as elaborated in Sec. III-D2.

Note that the view-dependent accumulated opacity does not precisely reflect the reconstruction accuracy of the entire space. Firstly, the proportion of low-visibility areas in the image domain does not reflect the actual unseen space in three dimensions, as a node close to the unseen area results in a large amount of invisible pixels. Secondly, the accumulated opacity leads to an over-confident evaluation of completeness, as the splatting from backside geometry can also result in

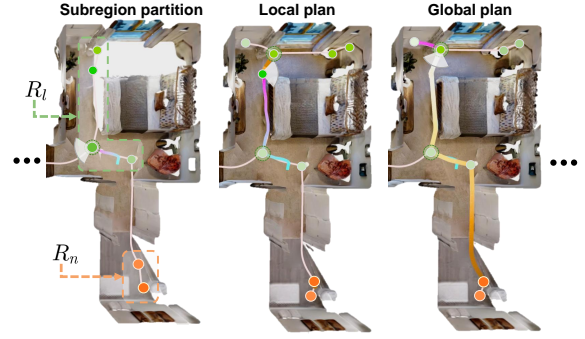


Fig. 3. Once the agent gets sufficient observations within a local region R_l (the green nodes), it selects the next sub-area R_n within the highest score (the orange node) globally for further exploration.

high accumulated opacity. To address these challenges, our method back-projects contour pixels, utilizing convex hull [32] to get the approximated volume of the incomplete geometry. We further maintain a set of high-loss samples through joint analysis of observation depth D_k and rendered depth \hat{D}_k :

$$M_h = (\hat{O}_k > \tau_{o2}) \wedge (D_k < \hat{D}_k) \wedge (|D_k - \hat{D}_k| > \epsilon_1), \quad (9)$$

where $\tau_{o2} = 0.8$, and $\epsilon_1 = 0.3$. The high-loss areas at each frame, before densification, are clustered with temporal propagation to keep track the newly observed regions.

3) *Determination of Target Views*: The viewpoint selection is then conducted in two stages. The agent first selects the node with the most invisible areas, taking both panorama visibility measures and convex hull volumes into consideration. This strategy forces the agent to reach a closed space by fast marching the nodes with high scores, as detailed in Sec. III-D3, therefore expanding the workspace efficiently. Once the agent arrives at its goal position, the panorama image and the maintained high-loss samples guide the agent to rotate, as illustrated in Fig. 2 and Eq. 9, where invisible and high-loss areas get observations. The proposed method keeps the agent along the Voronoi graph that compresses potential viewpoints into a finite sparse set, reducing the computational complexity while guaranteeing completeness and safety.

C. Hierarchical Planning with Voronoi Graph

In active mapping of multi-room environments, we observed that the prioritization of areas with high scores overlooks local coverage, resulting in paths that are unnecessarily visited multiple times, as shown in Fig. 5. The nodes of a Voronoi graph represent distinct reachable regions, and the edges efficiently assess traversal costs between them. Therefore, we propose a hierarchical planning strategy based on Voronoi graphs, which includes subregion partitioning and local-global path selection, as illustrated in Fig. 3.

1) *Subregion Partition*: Building upon the topological structure of the Voronoi graph, we aim to dynamically partition it during the exploration into n subregions R_n , including a local subregion R_l where the agent is currently located, ensuring fine-grained local granularity with global guidance. In practice, we adopt the agglomerative hierarchical clustering method (UPGMA) [33] for partitioning, with pairwise

distances based on both Euclidean distance and travel path length. The hierarchy allows the flexibility to choose partitions at different levels and adapts to spatial data, without specifying the number of clusters beforehand.

2) *Local-Global Goal Selection*: The proposed method favors local exploration before global exploration. Once local areas are thoroughly explored, the next-best-subregion with the highest score is selected. Local planning is conducted by quantifying the incomplete score, as specified in Eq. 10, within the local subregion. The node with the highest score above a threshold is selected iteratively. Once the maximum score of the nodes within the local subregion falls below the threshold or all nodes are visited, the agent performs global planning by selecting the node outside the local subregion with the highest score. The global score not only takes the coverage into account, but also considers the distance cost along with the visited probabilities during past exploration. The active mapping process is then conducted to iterate between rigorous local mapping and coarse global exploration, to balance the reconstruction accuracy and efficiency.

D. Implementation Details

Given the selected target positions and rotations, the agent actively explores the unknown environment and captures new information. In the following, we discuss the details regarding bootstrapping, panorama rendering, path planning, and post-processing.

1) *Bootstrapping*: Due to the limited field of view of the camera, we force the agent to look around at the very beginning. The agent takes discrete actions to execute 360 degrees of yaw rotation to obtain a complete ambient view.

2) *Panorama Rendering*: As the Gaussian splatting technique allows efficient rendering of pinhole images, we use three virtual cameras with 150 degrees of FOV vertically and 120 degrees of FOV horizontally to get the panoramic images, holistically quantifying the node-wise scores. The size of each panoramic image is set to be 360×150 to allow convenient selection of the rotation angle.

3) *Path Planning*: Once the target goal position is determined, the shortest path can then be found through Dijkstra's algorithm. The score of the i -th node, denoted as S_i , is implemented as a weighted sum of the following factors:

$$S_i = w_o \cdot s_o(i) + w_c \cdot s_c(i) + w_u \cdot s_u(i) + w_h \cdot s_h(i), \quad (10)$$

where $w_o = 20$, $w_c = 10$, $w_u = w_h = 10$. s_o and s_c are the portion of areas regarding the 2D invisible subregion and the 3D convex hull. s_u and s_h are the boolean values of unvisited and in-horizon states. Nodes with the same score are ranked according to their distance from the agent, where the nearer node is favored. We set a fixed distance threshold of 2 meters to control the granularity of subregion partitioning. We also enforce rotation once the agent arrives at multi-connected nodes in the graph as they are often intersection points between regions that require careful decision-making. Experiments indicate the efficacy of this strategy, achieving better efficiency for thorough exploration compared to previous state-of-the-art approaches [7].

TABLE I
COMPARISON AGAINST RELEVANT METHODS REGARDING THE COMPLETENESS OF THE OBSERVED DATA

	Gibson		Matterport3D	
	%↑	cm↓	%↑	cm↓
FBE [8]	68.30	14.42	74.30	9.29
UPEN [20]	63.30	21.09	75.56	9.72
ANM [11]	80.45	7.44	79.36	7.40
NARUTO [21]	79.16	3.52	84.90	5.94
ANM-S [7]	92.10	2.83	89.74	4.14
Ours	92.24	2.43	92.48	2.84

4) *Post-Processing*: Unlike NeRF-based SLAM algorithms that sacrifice model capacity for fast convergence to meet real-time demand, Gaussian splatting-based approaches maintain a consistent parameter space that allows post-processing. Therefore, we further apply adaptive density controls, as well as depth and normal regularization [5], [6] to refine the online-constructed map given stored keyframe data. The reconstruction results are shown in Fig. 6.

IV. EXPERIMENTS

The experiments are conducted on a desktop PC with an Intel Core i9-12900K CPU and an NVIDIA RTX 3090 GPU. The reported results are averaged across 5 trials.

A. Experimental Setup

To ensure fairness across exploration strategies, we perform qualitative and quantitative evaluations on the visually realistic Gibson [34] and Matterport3D [35] datasets using the Habitat simulator [36], following the protocol of [11]. The single-floor scenes in the test/validation set are divided into small (less than 5 rooms) and medium (5–10 rooms) scenes. Unless otherwise specified, the agent collects posed RGB-D data at a resolution of 256×256 and performs discrete actions: MOVE_FORWARD by 6.5 cm, TURN_LEFT and TURN_RIGHT by 10° , TURN_UP and TURN_DOWN by 15° , and STOP. The agent height is set to 1.25 m, with vertical and horizontal FOV of 90° . Additionally, the agent takes a 45° downward pitch rotation to ensure a closed ground surface before departure.

B. Evaluation Metrics

Following [21] and [11], we evaluate exploration coverage using completion ratio (%) and completion (cm) across different map representations. To evaluate rendering quality, we use PSNR (dB), SSIM, and LPIPS for RGB rendering [5], and Depth L1 (cm) distance for depth rendering performance. Additionally, we further evaluate the path length traveled by the agent during exploration.

C. Comparison to Other Methods

We first evaluate the exploration coverage across 13 different scenes following the setup of [11]. As shown in Tab. I, the proposed system outperforms all relevant methods within the limited steps (1000 for small scenes and 2000 for medium-scale scenes). NARUTO [21] achieves 92% coverage in simple

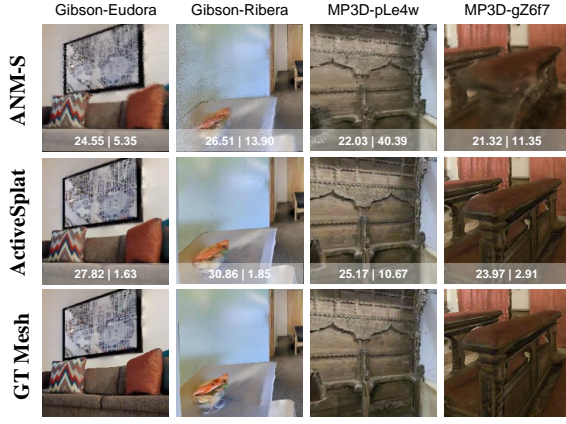


Fig. 4. The novel view synthesis results of ours compared to the NeRF-based active mapping [7] on Gibson and MP3D datasets. The bottom row of each picture shows the average PSNR (dB) and Depth L1 error (cm) of the scene from 50 randomly sampled test views.

TABLE II
ABLATION OF EXPLORATION STRATEGY

	Gibson		Matterport3D	
	% \uparrow	cm \downarrow	% \uparrow	cm \downarrow
Random	84.20	6.13	83.91	5.50
Position	90.41	2.74	89.54	3.67
Viewpoint	91.76	2.30	92.38	2.85
Ours	92.24	2.43	92.48	2.84

scenes with a single room (e.g., MP3D-gZ6f7), thanks to its unrestricted movement capabilities, but is limited in complex scenes with multiple rooms (e.g., Cantwell and Eastville) due to its greedy strategy. Even though [7] adopts a similar strategy of topology-guided exploration, the proposed method outperforms the baseline due to hierarchical planning that balances local reconstruction granularity and global scene coverage. We also performed a qualitative evaluation of novel view synthesis. As shown in Fig. 4, the proposed method achieves better rendering quality of novel views when compared to the NeRF-based system [7], exhibiting sharp edges and ample textures, as detailed in the supplementary Sec. ??.

D. Ablation Study and System Performance

To validate the rationale behind our solution, we conduct ablation studies and performance analyses to justify the effectiveness of different modules.

1) *Exploration Strategy*: We first analyze the impact of different exploration strategies for thorough exploration, including a randomly sampled baseline (Random), exploration with node selection while ignoring target rotations (Position), decoupled selection of view positions and rotations (Viewpoint), and our proposed approach further incorporating multi-connected regions and hierarchical planning (Ours). As shown in Tab. II, different exploration strategies lead to diverse behaviors for efficiency-accuracy tradeoffs. The Random baseline indicates that the Voronoi graph guarantees complete exploration. Nevertheless, the traversal of all nodes without proper order is inefficient and overlooks certain areas. This issue remains for the greedy Position strategy as it only strives

TABLE III
ABLATION OF HIERARCHICAL PLANNING

Path Ratio	Gibson				Matterport3D			
	w/o HP		Ours		w/o HP		Ours	
	% \uparrow	cm \downarrow	% \uparrow	cm \downarrow	% \uparrow	cm \downarrow	% \uparrow	cm \downarrow
25%	87.48	4.96	88.73	4.49	92.61	2.73	91.50	3.04
50%	93.00	2.15	94.26	1.75	95.14	2.18	95.07	2.21
75%	95.75	1.20	96.20	1.04	95.28	2.14	95.31	2.14
100%	96.38	1.01	96.77	0.90	95.33	2.11	95.42	2.11

Completion ratio (%) and completion error (cm) under different allowed path lengths (expressed as percentages of the full trajectory).

TABLE IV
ABLATION OF COVERAGE EVALUATION

	Gibson		Matterport3D	
	% \uparrow	cm \downarrow	% \uparrow	cm \downarrow
Visibility only	90.19	3.09	91.40	3.15
Convex hull only	91.09	2.86	91.50	2.98
Ours	92.24	2.43	92.48	2.84

TABLE V
ABLATION OF POST-PROCESSING ON GIBSON-DENMARK

	Depth loss	Split	Depth L1 \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Online	True	Train	1.91	25.28	0.83	0.22
		Test	9.01	21.72	0.76	0.29
Refined with 3DGS	False	Train	4.49	39.10	0.98	0.03
		Test	11.2	26.27	0.86	0.19
	True	Train	0.77	38.90	0.99	0.03
		Test	7.81	26.87	0.88	0.17
Refined with 2DGS	False	Train	3.94	38.84	0.99	0.04
		Test	10.10	27.13	0.88	0.18
	True	Train	0.80	38.90	0.99	0.04
		Test	7.56	27.58	0.88	0.17

to push the boundaries of the workspace toward thorough traversal. The completeness can be improved with the rotation involved. We provide more detailed scene-wise results in the supplementary Sec. ?? . Finally, the careful treatment of multi-connected nodes and the hierarchical planning (HP) strategy bring further advantages due to different inspection granularity locally and globally. Tab. III shows the results of an ablation in which we increase the number of steps to 4000 (in all scenes) and compare the results with and without hierarchical planning at different stages. Our method benefits from fine-grained observations within local subregions, enabling higher completeness in the later stages. Additionally, we visualize the exploration trajectory when exploring a medium-sized environment. As illustrated in Fig. 5, even though the greedy strategy leads to a rapid increase of completion in the beginning, coarsely exploring the neighboring areas results in repetitive trajectories.

2) *Coverage Evaluation*: As mentioned in Sec. III-B2, the quantification of visibility in the panoramic view guides the agent to push the boundary of the workspace. To verify the effectiveness of the integration of both the invisible mask area and convex hull volume, we compare different visibility quantification methods: guiding the agent to the node with the largest rendered area of invisible regions (Visibility only) and favoring the node with the largest 3D convex hull of the invisible boundary (Convex hull only). As shown in Tab. IV, using the view-dependent 2D results or the 3D volume

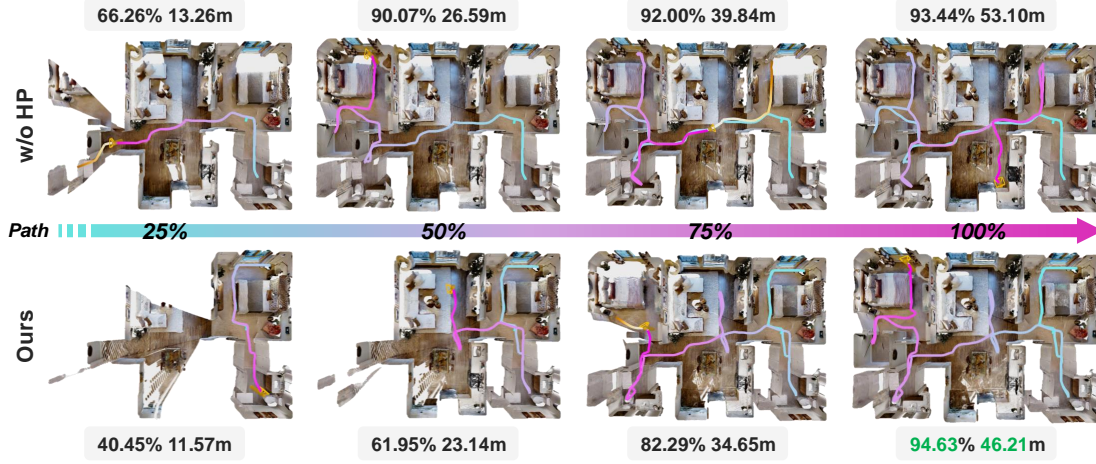


Fig. 5. **Ablation of hierarchical planning (Gibson Quantico):** The online reconstruction progress with increased completion ratio (%) and path length (m) at different stages. The hierarchical planning strategy results in better completeness and reduced path length during the exploration.

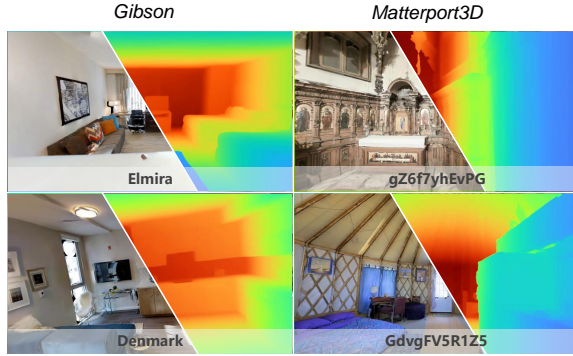


Fig. 6. **Reconstruction results:** The autonomous reconstruction lead to photorealistic rendering and accurate geometry at a resolution of 512×512 . The left and right sides of each image show rendered RGB and depth.

quantification alone may not best evaluate the candidate nodes. The integration of both strategies (Ours) as weighted averaging of the normalized scores takes the relative extent of invisible areas near the Voronoi nodes into consideration, achieving promising completeness during the active mapping.

3) *Post-Processing:* The recent Gaussian splatting technique allows convenient post-processing with the stored keyframe buffer. We here compare our results before and after the post-processing using 3DGS [5] and 2DGS [6] on Gibson Denmark. 50 frames of observations are selected uniformly as the train split, and 50 images with randomly sampled camera poses within the free space are taken as the test split. As shown in Tab. V, refinement can drastically enhance the reconstruction quality in terms of both geometry and appearance with the incorporation of RGB and depth during the optimization. The online autonomous exploration process allows active data capture for complete and high-fidelity reconstruction. It can be noted that the two-dimensional flattened Gaussian parameter representation of 2DGS [6], along with the geometric regularization terms, shows better results in the test split, while 3DGS [5] indicates better overfitting in the train split. Besides, the use of depth loss, as defined in Eq. 6, not only leads to better geometry (lower Depth L1) in both train and test splits, but also enhances the generalization of the map

TABLE VI
AVERAGE PROCESSING TIME PER STEP

Mapper	Get workspace	Get Voronoi	Get subregion
45.14ms	43.87ms	0.69ms	1.56ms
Rotation selection	Position selection	Path planner	Visualizer (optional)
2.76ms	6.46ms	0.20ms	67.43ms

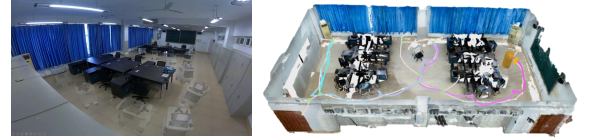


Fig. 7. Real-world experimental result in an office scene.

(better quality in the test split). Refinement without depth loss may result in more realistic view synthesis results on training views (severe overfitting), but the geometry may deteriorate due to ambiguities in textureless areas.

4) *System Performance:* The average processing time per step for each module is presented in Tab. VI. The proposed system runs at 8 fps in a headless mode, demonstrating real-time capable performance. Furthermore, the sparse decision-making ensures efficient viewpoint selection, while path planning constitutes only a small portion of the overall processing load. The major costs lie in the mapping and workspace extraction modules, which can be further accelerated with compressed Gaussian parameters.

E. Deployment in Real World

To evaluate the practical usage of the proposed system in the real world, we deploy it on an omnidirectional mobile robot (Agile-X Ranger Mini) equipped with an Azure Kinect RGB-D sensor, where the camera pose is estimated by a line-based SLAM system [37] in a parallel thread. Compared to simulation settings, we adjusted the radius hyperparameter of visited nodes for more detailed active mapping. The online reconstruction results are shown in Fig. 7, depicting the complete exploration that progressively and autonomously reconstructs

the indoor environment. The detailed reconstruction process with additional results is provided in the supplementary video.

V. CONCLUSION

In this paper, we introduce an active mapping system for high-fidelity reconstruction of indoor scenes. Benefiting from the accurate dense prediction of a Gaussian splatting-based differentiable renderer and the workspace abstraction through Voronoi graph extraction, we employ the hybrid map along with a novel topology-based hierarchical planning strategy to achieve promising tradeoffs between exploration efficiency and completeness. Detailed experimental results indicate that the proposed system achieves effective tradeoffs between exploration coverage and reconstruction quality.

While ActiveSplat performs well, future work could further enhance its robustness, scalability, and real-world applicability. Integrating LiDAR into the Gaussian splatting-based SLAM system would enable robust tracking and more accurate geometric reconstruction in large-scale scenes. Moreover, high-fidelity exploration in unknown indoor environments opens new avenues for research in robotic autonomy, extending to tasks like lifelong navigation and mobile manipulation.

REFERENCES

- [1] V. Patil and M. Hutter, "Radiance fields for robotic teleoperation," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2024, pp. 13 861–13 868.
- [2] M. Torne, A. Simeonov, Z. Li, A. Chan, T. Chen, A. Gupta, and P. Agrawal, "Reconciling reality through simulation: A real-to-sim-to-real approach for robust manipulation," in *Robotics: Science and Systems (RSS)*, 2024.
- [3] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Commun. ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [4] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and p. Hedman, "Zip-nerf: Anti-aliased grid-based neural radiance fields," in *Intl. Conf. on Computer Vision (ICCV)*, 2023, pp. 19 697–19 705.
- [5] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM Trans. Graphics*, vol. 42, no. 4, pp. 1–14, 2023.
- [6] B. Huang, Z. Yu, A. Chen, A. Geiger, and S. Gao, "2d gaussian splatting for geometrically accurate radiance fields," in *SIGGRAPH*. Association for Computing Machinery, 2024.
- [7] Z. Kuang, Z. Yan, H. Zhao, G. Zhou, and H. Zha, "Active neural mapping at scale," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2024.
- [8] B. Yamauchi, "A frontier-based approach for autonomous exploration," in *IEEE Intl. Sym. on Computational Intelligence in Robotics and Automation (CIRA)*, 1997, pp. 146–151.
- [9] H. Umari and S. Mukhopadhyay, "Autonomous robotic exploration based on multiple rapidly-exploring randomized trees," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2017, pp. 1396–1402.
- [10] L. Schmid, M. Pantic, R. Khanna, L. Ott, R. Siegwart, and J. Nieto, "An efficient sampling-based method for online informative path planning in unknown environments," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1500–1507, 2020.
- [11] Z. Yan, H. Yang, and H. Zha, "Active neural mapping," in *Intl. Conf. on Computer Vision (ICCV)*, 2023.
- [12] C. Cao, H. Zhu, H. Choset, and J. Zhang, "Tare: A hierarchical framework for efficiently exploring complex 3d environments," in *Robotics: Science and Systems (RSS)*, vol. 5, 2021, p. 2.
- [13] M. Selin, M. Tiger, D. Duberg, F. Heintz, and P. Jensfelt, "Efficient autonomous exploration planning of large-scale 3-d environments," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1699–1706, 2019.
- [14] Q. Dong, H. Xi, S. Zhang, Q. Bi, T. Li, Z. Wang, and X. Zhang, "Fast and communication-efficient multi-uav exploration via voronoi partition on dynamic topological graph," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2024, pp. 14 063–14 070.
- [15] Z. Zhu, S. Peng, V. Larsson, Z. Cui, M. R. Oswald, A. Geiger, and M. Pollefeys, "Nicer-slam: Neural implicit scene encoding for rgb slam," in *Intl. Conf. on 3D Vision (3DV)*, March 2024.
- [16] C. Jiang, H. Zhang, P. Liu, Z. Yu, H. Cheng, B. Zhou, and S. Shen, "H2-mapping: Real-time dense mapping using hierarchical hybrid representation," *IEEE Robotics and Automation Letters*, 2023.
- [17] N. Keetha, J. Karhade, K. M. Jatavallabhula, G. Yang, S. Scherer, D. Ramanan, and J. Luiten, "Splatam: Splat track & map 3d gaussians for dense rgb-d slam," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 21 357–21 366.
- [18] H. Matsuki, R. Murai, P. H. Kelly, and A. J. Davison, "Gaussian splatting slam," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 18 039–18 048.
- [19] J. Wei and S. Leutenegger, "GSfusion: Online rgb-d mapping where gaussian splatting meets tsdf fusion," *IEEE Robotics and Automation Letters*, 2024.
- [20] G. Georgakis, B. Bucher, A. Arapin, K. Schmeckpeper, N. Matni, and K. Daniilidis, "Uncertainty-driven planner for exploration and navigation," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2022.
- [21] Z. Feng, H. Zhan, Z. Chen, Q. Yan, X. Xu, C. Cai, B. Li, Q. Zhu, and Y. Xu, "Naruto: Neural active reconstruction from uncertain target observations," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 21 572–21 583.
- [22] X. Pan, Z. Lai, S. Song, and G. Huang, "Activenerf: Learning where to see with uncertainty estimation," in *European Conf. on Computer Vision (ECCV)*. Springer, 2022, pp. 230–246.
- [23] G. Liu, W. Jiang, B. Lei, V. Pandey, K. Daniilidis, and N. Motee, "Beyond uncertainty: Risk-aware active view acquisition for safe robot navigation and 3d scene understanding with fisherrf," *arXiv preprint arXiv:2403.11396*, 2024.
- [24] R. Jin, Y. Gao, Y. Wang, Y. Wu, H. Lu, C. Xu, and F. Gao, "Gs-planner: A gaussian-splatting-based planning framework for active high-fidelity reconstruction," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2024, pp. 11 202–11 209.
- [25] Z. Xu, R. Jin, K. Wu, Y. Zhao, Z. Zhang, J. Zhao, F. Gao, Z. Gan, and W. Ding, "Hgs-planner: Hierarchical planning framework for active scene reconstruction using 3d gaussian splatting," *arXiv preprint arXiv:2409.17624*, 2024.
- [26] Y. Tao, D. Ong, V. Murali, I. Spasojevic, P. Chaudhari, and V. Kumar, "Rt-guide: Real-time gaussian splatting for information-driven exploration," *arXiv preprint arXiv:2409.18122*, 2024.
- [27] T. Chen, O. Shorinwa, J. Bruno, A. Swann, J. Yu, W. Zeng, K. Nagami, P. Dames, and M. Schwager, "Splat-nav: Safe real-time robot navigation in gaussian splatting maps," *IEEE Trans. Robotics*, 2025.
- [28] L. Jin, X. Zhong, Y. Pan, J. Behley, C. Stachniss, and M. Popović, "Activevegs: Active scene reconstruction using gaussian splatting," *IEEE Robotics and Automation Letters*, 2025.
- [29] W. Jiang, B. Lei, and K. Daniilidis, "Fisherrf: Active view selection and mapping with radiance fields using fisher information," in *European Conf. on Computer Vision (ECCV)*. Springer, 2024, pp. 422–440.
- [30] Q. Du, V. Faber, and M. Gunzburger, "Centroidal voronoi tessellations: Applications and algorithms," *SIAM review*, vol. 41, no. 4, pp. 637–676, 1999.
- [31] J. Canny and B. Donald, "Simplified voronoi diagrams," in *Proceedings of the third annual symposium on Computational geometry*, 1987, pp. 153–161.
- [32] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Trans. Math. Softw.*, vol. 22, no. 4, pp. 469–483, 1996.
- [33] O. Arslan, D. P. Guralnik, and D. E. Koditschek, "Coordinated robot navigation via hierarchical clustering," *IEEE Trans. Robotics*, vol. 32, no. 2, pp. 352–371, 2016.
- [34] F. Xia, A. R. Zamir, Z. He, A. Sax, J. Malik, and S. Savarese, "Gibson Env: real-world perception for embodied agents," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [35] A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, "Matterport3d: Learning from rgb-d data in indoor environments," in *Intl. Conf. on 3D Vision (3DV)*, 2017.
- [36] M. Savva, A. Kadian, O. Maksymets, Y. Zhao, E. Wijmans, B. Jain, J. Straub, J. Liu, V. Koltun, J. Malik et al., "Habitat: A platform for embodied ai research," in *Intl. Conf. on Computer Vision (ICCV)*, 2019, pp. 9339–9347.
- [37] Q. Wang, Z. Yan, J. Wang, F. Xue, W. Ma, and H. Zha, "Line flow based simultaneous localization and mapping," *IEEE Trans. Robotics*, vol. 37, no. 5, pp. 1416–1432, 2021.