ActiveSplat: High-Fidelity Scene Reconstruction through Active Gaussian Splatting

Yuetao Li[1,2⋆], Zijia Kuang[2⋆], Ting Li[2], Guyue Zhou[2], Shaohui Zhang[1,†], Zike Yan[2,†]

***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 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. Project page: https://li-yuetao.github.io/ActiveSplat/.**

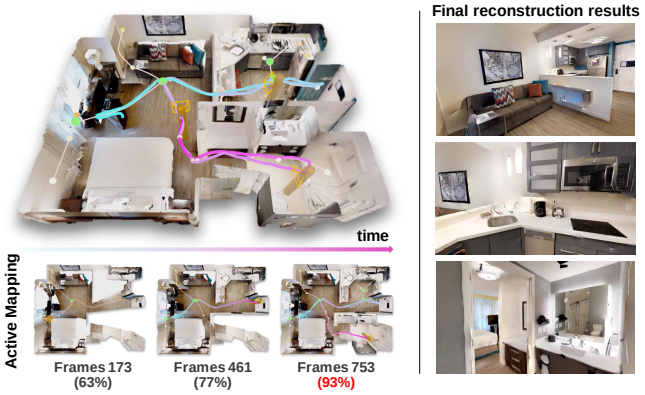

Fig. 1: The agent explores the environment autonomously to build a 3D map on the fly. The integration of a Gaussian map and a Voronoi graph assures efficient and complete exploration with high-fidelity reconstruction results.

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 photo-realistic simulation [2].

Recent progress in differentiable rendering has significantly improved the quality of reconstructed environments. Neural Radiance Fields (NeRF) [3] and its variants [4], [5], [6] leverage neural networks as compact scene representations, using volume rendering to synthesize high-quality novel views. However, the computational inefficiencies caused by volume integration along rays pose challenges in terms of memory and processing. To address these limitations, Gaussian splatting [7], [8], [9], [10], [11] has been introduced, enabling efficient rasterization and achieving promising rendering quality through α-blending. Despite these advances, scene-specific representations and data-dependent optimization make these methods highly sensitive to the captured observations, where noise and artifacts can easily emerge due to insufficient view coverage, especially without direct feedback during data collection [12], [13].

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 both working space and obstacles 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 [14]. 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 working space. Sparse yet representative view positions are derived from this graph, guiding the agent to extend the boundaries of the working space. Meanwhile, the viewing orientation at each position is determined by view-dependent completeness measures. This approach reduces 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

⋆ The co-first authors contributed equally.
† The corresponding author: zhangshaohui@bit.edu.cn, yanzike@air.tsinghua.edu.cn
[1] School of Optics and Photonics, Beijing Institute of Technology, China
[2] Institute for AI Industry Research (AIR), Tsinghua University, China.

autonomy process. The key contributions of the paper 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 of Gaussians and sparse abstraction of Voronoi graph for comprehensive viewpoint selection and safe path planning.
- A hierarchical planning strategy based on the Voronoi graph prioritizes local areas to minimize redundant exploration, decoupling viewpoint selection to balance exploration efficiency and reconstruction accuracy.

II. Related Work

A. Autonomous exploration

Autonomous exploration aims to best acquire observations to cover the entire space with limited steps. Existing strategies can be broadly categorized as frontier-based methods and sampling-based methods. Frontier-based methods [15], [16] 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 [17], [18], [14] 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 [19] introduces a hierarchical strategy for LiDAR-based exploration, where local subspace is traversed at a fine-grained level while global target goals at a coase level are maintained, thus achieving a balance between exploration efficiency and mapping completeness. Similarly, [20] adopts a fine-grained Next-best-view planning for local exploration, while leveraging frontier-based strategies for global coverage. Most relevant to ours is [14] 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 provides a coarse and cumbersome assessment of reconstruction quality. We, on the other hand, leverage the efficient optimization and rendering of 3DGS to achieve autonomous reconstruction in high fidelity.

B. High-fidelity scene reconstruction

Recent progress in differentiable rendering attracts significant attention in the research community. Parameterized by implicit NeRF [3], [4], [5], [6] representations or explicit 2D/3D Gaussian [7], [8], [9], [10], [11] representations, photo-realistic images of novel views can be rendered with promising efficiency. The gradient-based optimization has also been applied in an online setting to incrementally update the neural map [21], [22], [23], [24], [25] or the Gaussian parameters [26], [27], [28], [29] through differentiable rendering. Recently, continual learning of the neural map has turned into an active fashion through uncertainty-guided autonomous exploration [18], [14], [30], [31], [32]. However, the computationally expensive training process forces the map to balance between accuracy and convergence efficiency with different network architectures. We adopt a set of Gaussian primitives as the scene representation that allows consistent optimization in an online updating or offline post-processing setting. Relevant work includes a safe navigation system [33] builds upon FisherRF [34] that utilizes Fisher information for quantifying the uncertainty of Gaussians, and GS-Planner [35], an active mapping system using the Gaussian splatting technique. The major differences lie in our hybrid map representation that allows safe and hierarchical path planning with thorough data acquisition.

III. Methodology

The overview of our ActiveSplat system is illustrated in Fig. 2, where Gaussians of interest are splatted onto the image plane as a consistent technique utilized for online map updating, viewpoint selection, and path planning.

A. Hybrid map updating

Central to the proposed ActiveSplat system is a hybrid map representation containing both Gaussian primitives that allow dense prediction and a topological structure that provides sparse abstraction of the workspace. 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:

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

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

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

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

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

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

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 [26]:

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

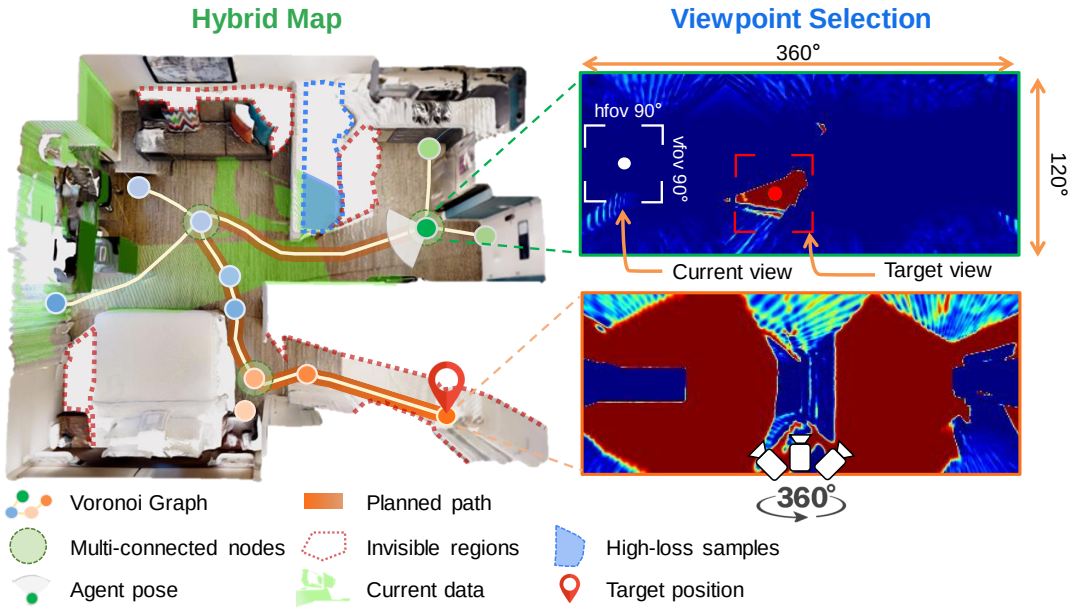

Fig. 2: **Overview of ActiveSplat**: The autonomous reconstruction system employs a hybrid map that incorporates both dense predictions and sparse topological abstractions, balancing exploration efficiency with reconstruction accuracy. Splatting Gaussians of interest on the fly forms a consistent manner for online map updating, viewpoint selection, and path planning. **Note:** Subregions are distinguished by node color, with node scores indicated by color intensity.

$$L_{geo} = |D_k - \hat{D}_k|, \tag{6}$$

$$L = w_c L_{pho} + w_d L_{geo}, \tag{7}$$

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

During the online mapping process, new Gaussians will be dynamically initialized to cover newly observed areas, and redundant Gaussians with near-zero opacity or large covariances will be removed as [7]. We follow [26] to define newly observed areas as a view-dependent binary mask with small accumulated opacity or incoming geometry in front of the maintained map:

$$M_k = (O_k < \tau_o) \vee \Big((D_k < \hat{D}_k) \wedge (|D_k - \hat{D}_k| > \epsilon_{\text{MDE}}) \Big), \tag{8}$$

where $\tau_o = 0.98, \epsilon_{\text{MDE}}$ equals 50 times median depth error.

The dense prediction of the Gaussian map allows convenient extraction of the workspace and the obstacles. As illustrated in Fig. 3, the top-down view can be efficiently rendered given a large focal length as the orthographic projection of the dense map. The region with sufficient accumulated opacity is taken as occupied, where occupied areas above the ground indicate the obstacles. Navigable workspace can then be extracted as occupied areas near the ground excluding the obstacles. An undirected Voronoi graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ can then be extracted through Voronoi tessellation [36] to obtain edges $\mathcal{E}$ equidistant to obstacles and nodes $\mathcal{N}$ as intersections where the edges $\mathcal{E}$ terminate.

The Gaussian map and the Voronoi graph are complementary: Gaussian map provides dense and complete information about past experience, while the Voronoi graph offers a sparse structure of the workspace. We will show as follows that the integration of the two formats leads to an adaptive granularity of the environment and guarantees a nice trade-off between efficiency and accuracy during the autonomous reconstruction process.

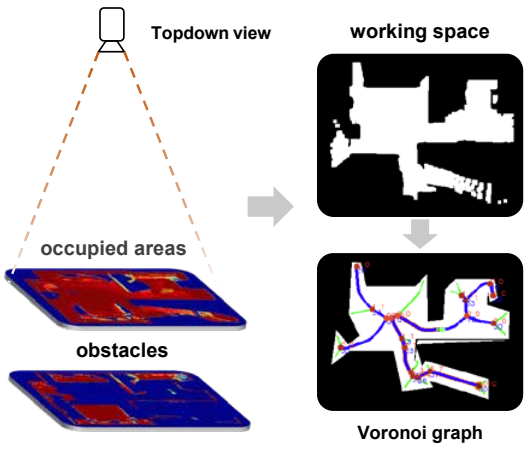

Fig. 3: The extraction of the Voronoi graph.

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 vital to efficiency and overall coverage. One nice property of Voronoi graph is that it can be seen as a strong deformation retract of the global free space [37]. Besides, Voronoi graph generates a path that stays as far away from the obstacles as possible,

thus guaranteeing a safe traversal. Consequently, we wish to sample viewpoints on the Voronoi graph to maintain a compact and accessible set that covers the entire scene.

a) Decoupled position and rotation candidates: To best exploit the compact structure of the Voronoi graph along with the rich information inherited in the Gaussian map, we propose to decouple the position and rotation candidates to enforce sparsity while maintaining thorough observations. The dynamically updated workspace leads to an incremental-augmented graph that completely describes the partially-observed workspace. We iteratively select the Voronoi node as viewpoint position candidates, where the node with the most information gain pushes 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.

b) 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. In practice, we render panoptic images regarding the visibility at all nodes as illustrated in Fig. 2, and cluster the low-visibility areas through DBSCAN algorithm [38] for generating rotation angle candidates. 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 areas will result in a large amount of invisible pixels. Besides, the accumulated opacity leads to an over-confident evaluation of completeness as the splatting from backside geometry can also result in high accumulated opacity. As illustrated in Fig. 2, we project the contour pixels of low-visibility areas into the 3D space, where convex hull [39] is applied to get the approximated volume regarding the unseen geometry. We further maintain a set of high-loss samples according to Eq. 8. The high-loss areas at each frame before densification will be clustered with temporal propagation to keep track of the newly observed region.

c) Determination of target views: The viewpoint selection is then conducted in two stages. The agent will first select 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 information gains, therefore expanding the workspace efficiently. Once the agent arrives at the target goal position, the panorama image and the maintained high-loss samples guide the agent to rotate, as illustrated in Fig. 2, where invisible and high-loss areas will get observations accordingly. The proposed method keeps the agent along the Voronoi graph that compresses potential viewpoints into a finite sparse set while guaranteeing completeness and safety.

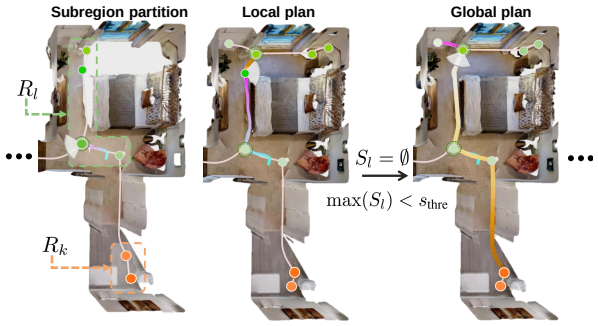

Fig. 4: Once the agent gets sufficient observations within a local region R_l (the green nodes), it selects the next sub-area R_k within most information gains (the orange nodes) globally for further exploration.

C. Hierarchical planning with Voronoi Graph

To enhance the overall efficiency of agent exploration and avoid revisiting past areas, we propose a hierarchical planning strategy based on Voronoi graphs, which include subregion partition and local-global goal selection (as illustrated in Fig. 4).

a) Subregion partition: Building upon the topological structure of the Voronoi graph, we aim to dynamically partition the graph into subregions during the exploration process to ensure fine-grained local granularity with global guidance. In practice, we adopt the agglomerative hierarchical clustering method (UPGMA) [40], [41] for subregion partition, where the pairwise distance takes both Euclidean and geodesic distance metrics. The hierarchy allows the flexibility to choose partitions at different levels and adapts to spatial data. As the agglomeration clustering considers the average distance between nodes, it produces balanced clusters with promising efficiency given the sparse graph structure.

b) Local-global goal selection: The selection of next-best-subregion follows similar criteria as Sec. III-B(c) that directs the agent towards nodes to rapidly explore the surroundings. During the exploration, local areas around the agent are prioritized for detailed inspection, while the next-best-subregion with the most information gain will be selected once local areas are thoroughly explored. The local planning is conducted by quantifying the aforementioned incomplete score within the local subregion. The node with the highest score above a threshold will be selected iteratively. Once the score of any node within the local horizon is below the threshold, the agent will execute global planning by finding 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 mapping locally and coarse exploration globally to balance the reconstruction accuracy and efficiency.

D. Implementation details

Given the selected target positions and the target rotations, the agent will actively explore the unknown environment and capture new information. We further discuss the details regarding bootstrapping, panorama rendering, path planning, and post-processing.

a) Boostrapping: Due to the limited field of view, 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. In the simulation, the agent will additionally take 45 degrees of downward pitch rotation to ensure a closed ground surface before departure.

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

c) Path planning: Once the target goal position is determined, the shortest path can then be found through Dijkstra's algorithm. The node scoring is implemented as a weighted sum of the factors, where the maximum score of each factor is 1. The scores regarding 2D invisible subareas and convex hulls are normalized by the maximum area/volume, and the rest are binary values. Nodes with the same score will be ranked according to their distance from the agent, where the nearer node will be favored. We set a fixed distance threshold of 90 for dendrogram pruning 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 with better efficiency for thorough exploration.

d) Post-processing: Unlike NeRF-based SLAM algorithms that sacrifice model capacity for fast convergence to meet real-time demand, Gaussian-based approaches maintain a consistent parameter space that allows post-processing. We further apply adaptive density controls and additional optimization [7], [8] to refine the online-constructed map given stored keyframe data (as illustrated in Fig. 7).

IV. Experiment

A. Experiment setup

The experiments are conducted on a desktop with a Intel Core i9-12900K CPU and an NVIDIA RTX 3090 GPU. Following the protocol of [18], we utilize the Habitat simulator [42] with Gibson [43] and Matterport3D [44] datasets for qualitative and quantitative evaluation. The agent collects posed RGB-D data at a resolution of 256×256 and performs discrete actions of `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 the vertical and horizontal fields of view of 90°.

	Gibson		Matterport3D	
	%↑	cm↓	%↑	cm↓
FBE [15]	68.30	14.42	74.30	9.29
UPEN [45]	63.30	21.09	75.56	9.72
ANM [18]	80.45	7.44	79.36	7.40
ANM-S [14]	92.10	2.83	89.74	4.14
Ours	**92.24**	**2.43**	**92.48**	**2.84**

TABLE I: Comparison against relevant methods regarding the completeness of the observed data.

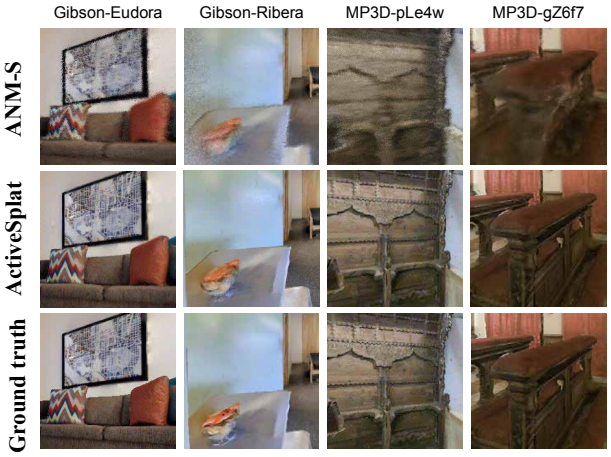

Fig. 5: The novel view synthesis results of ours compared to the NeRF-based active mapping [14] on Gibson and MP3D datasets.

B. Comparison to other methods

We first evaluate the exploration coverage across 13 different scenes following the setup of [18] and calculate the completion ratio (%) and completion (cm) for quantitative assessment. As shown in Tab. I, the proposed system outperforms all relevant methods within the limited steps (1000 for small scenes and 2000 for large-scale scenes). Even though [14] adopts a similar strategy of topology-guided exploration, the hierarchical planning strategy balances between local reconstruction granularity and global scene coverage.

We also performed a qualitative evaluation regarding the novel view synthesis. As shown in Fig. 5, the proposed ActiveSplat takes advantage of Gaussian splatting technique and achieves significant improvement regarding the novel view synthesis task compared to the NeRF-based system [14].

C. Ablation study

To validate the rationale behind our solution, we conduct ablation studies in different modules to justify the effectiveness of each strategy for high-fidelity reconstruction.

a) Exploration strategy: We first analyze the effects of different strategies for thorough exploration. As demonstrated in Tab. II, 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

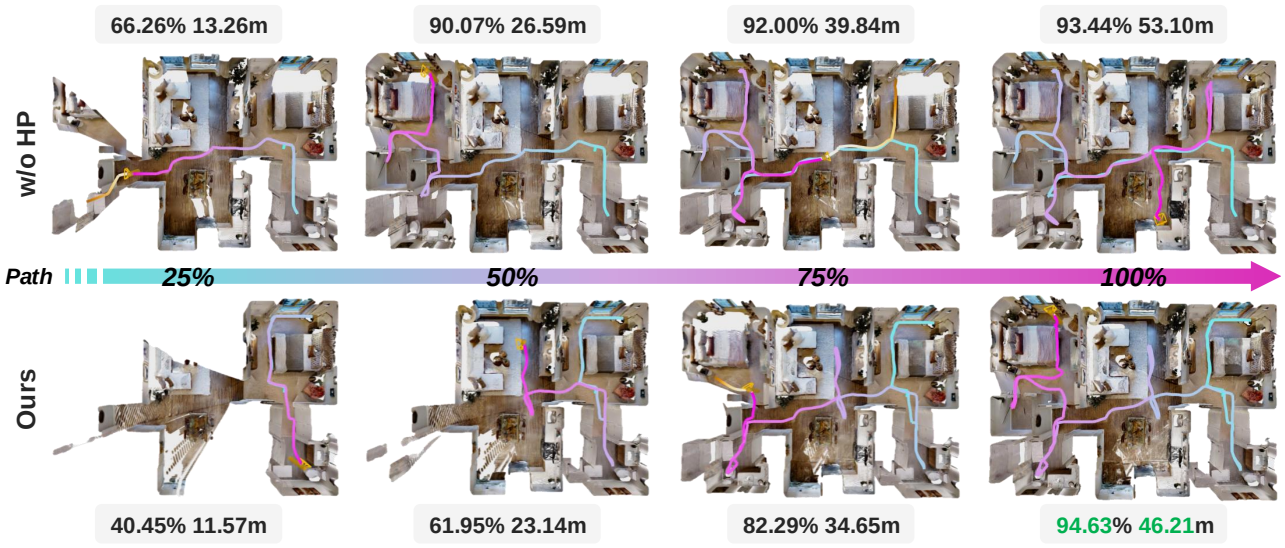

Fig. 6: **Ablation of hierarchical planning (scene ID: `Quantico`):** The online reconstruction progress with increased completeness (%) and path length (m) at different stages ($25\%, 50\%, 75\%, 100\%$). The hierarchical planning strategy results in better completeness and reduced path length during the exploration.

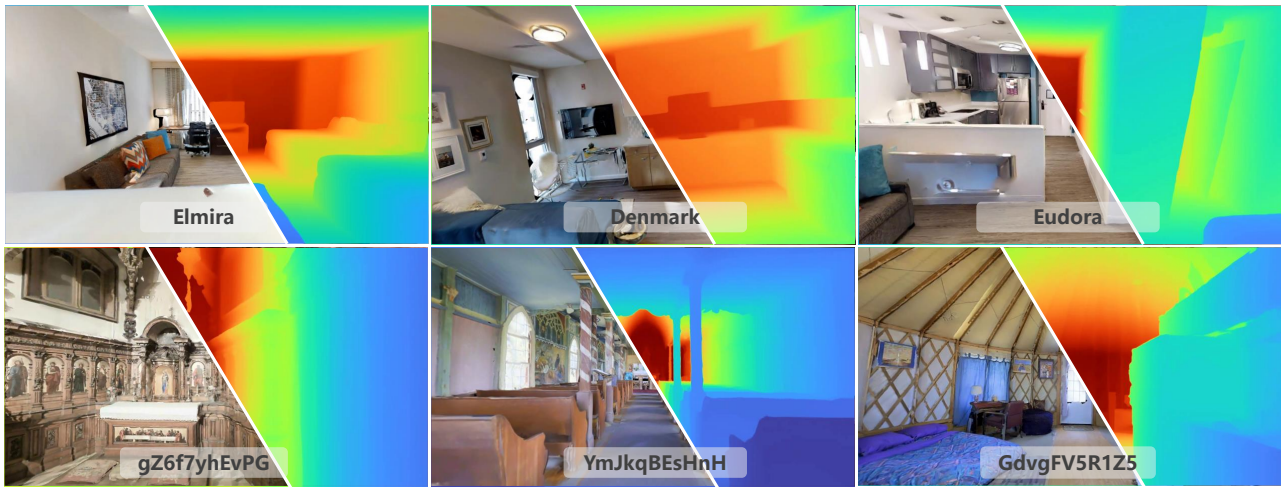

Fig. 7: **Reconstruction results**: The autonomous reconstruction lead to photorealistic rendering and accurate geometry.

for the greedy Position strategy as it only strives to push the boundaries of working space toward thorough traversal. The completeness can be improved with the rotation involved. Finally, the careful treatment of multi-connected nodes and the hierarchical planning strategy bring further advantages due to different inspection granularity locally and globally. As shown in Tab. III and Fig. 6, even though the greedy strategy leads to rapid ascendance of completeness in the beginning, coarsely exploring the neighboring areas results in repetitive trajectories. The proposed hierarchical planning strategy assures a smaller path length with higher completeness compared to the greedy baseline after traversal.

b) Coverage evaluation: As mentioned in Sec. III-B(b), the quantification of visibility in the panoramic view guides the agent to push the boundary of the working space. To verify the effectiveness of the integration of both the invisible mask area and convex hull volume, we evaluate the results with the following settings. "Visibility only" guides the agent to the node with the largest area of invisible regions, and "Convex hull only" favors the unvisited areas with the largest 3D convex hull. As shown in Tab. IV, merely using the view-dependent 2D results or the 3D volume quantification does not best capture the information gain at the corresponding node. For instance, the "Visibility only" strategy cannot reflect the actual size of the unvisited regions, while "Convex hull only" strategy does not quantify the exact information gain viewed from the node position. The integration of both strategies (Ours) as a normalized scoring manner serves as a complementary solution to take the relative extent of invisible areas near the Voronoi nodes into consideration, achieving promising completeness during the active mapping.

c) Post-processing: The unified representation of Gaussians allows convenient post-processing. We here compare our results before and after the post-processing

	Gibson		Matterport3D	
	%↑	cm↓	%↑	cm↓
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**

TABLE II: **Ablation of exploration strategy:** (1) "Random" denotes the random selection of unvisited Voronoi nodes; (2) "Position" denotes exploration with only node selection but ignoring the target rotations; (3) "Viewpoint" denotes the decoupled selection of both view positions and rotations; (4) "Ours" incorporates multi-connected regions and hierarchical planning during the navigation. Different exploration strategies lead to diverse behaviors for efficiency-accuracy tradeoffs.

	Gibson				Matterport3D			
	w/o HP		**Ours**		**w/o HP**		**Ours**	
Path Ratio	% ↑	cm ↓	% ↑	cm ↓	% ↑	cm ↓	% ↑	cm ↓
25%	**66.58**	**15.95**	66.40	16.24	**76.66**	**7.75**	71.80	9.57
50%	75.94	11.00	**78.82**	**8.60**	**85.68**	**4.63**	80.85	6.29
75%	**85.38**	**5.31**	82.31	7.35	**88.80**	**3.81**	87.60	4.28
100%	**92.96**	**1.98**	92.24	2.43	91.39	3.01	**92.48**	**2.84**
150%	94.24	1.55	**95.24**	**1.17**	93.33	2.53	**94.11**	**2.40**
P.L.(m ↓)	41.56		**37.12**		31.12		**27.41**	

TABLE III: **Ablation of hierarchical planning:** Hierarchical Planning (HP) enables the agent to achieve higher reconstruction completeness with shorter path lengths (P.L.).

using 3DGS [7] and 2DGS [8]. 50 captured frames are selected uniformly as the train split for each scene, and randomly sample 50 images given randomly sampled camera poses within the free space are taken as the test split. As shown in Tab. V, further refinement can drastically enhance the reconstruction quality in terms of both geometry and appearance if both RGB and depth observations are utilized during the optimization. The online feedback allows active data captures for complete and high-fidelity reconstruction. It can be noted that the two-dimensional flattened Gaussian parameter representation of 2DGS [8] along with the geometric regularization terms show better results in the test split, while 3DGS [7] indicates better overfitting in the train split. Besides, depth image not only leads to better geometry (lower Depth L1) in both train and test splits, but also enhances the generalization of the map (better quality in the test split). Refinement without depth loss may result in more realistic view synthesis results (severe overfitting), but the geometry may deteriorate due to ambiguities in textureless areas.

V. Conclusion

In this paper, we introduce ActiveSplat, an active mapping system for high-fidelity reconstruction of indoor scenes. Benefiting from the accurate dense prediction of the differentiable rendering and the working space abstraction through Voronoi graph extraction, the system achieves promising tradeoffs between exploration efficiency and completeness. Detailed experimental results validate the efficacy of the proposed system. The ability to explore in unknown indoor environments with high-fidelity reconstruction opens the door for further research on robotic autonomy and active perception, which can be further extended to complex tasks such as lifelong navigation and mobile manipulation.

	Gibson		Matterport3D	
	%↑	cm↓	%↑	cm↓
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 IV: **Ablation of coverage evaluation:** Merely 2D or 3D way of quantifying invisible areas does not best guide the agent.

	Depth loss	**Split**	**Depth L1↓**	**PSNR↑**	**SSIM↑**	**LPIPS↓**
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

TABLE V: **Ablation of post-processing:** The reconstruction results before and after post-processing.

References

[1] V. Patil and M. Hutter, "Radiance fields for robotic teleoperation," *arXiv preprint arXiv:2407.20194*, 2024.

[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," *arXiv preprint arXiv:2403.03949*, 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," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[4] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, "Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields," in *Intl. Conf. on Computer Vision (ICCV)*, 2021, pp. 5855–5864.

[5] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Mip-nerf 360: Unbounded anti-aliased neural radiance fields," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5470–5479.

[6] 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.

[7] 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.

[8] 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.

[9] Z. Yu, A. Chen, B. Huang, T. Sattler, and A. Geiger, "Mip-splatting: Alias-free 3d gaussian splatting," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 19 447–19 456.

[10] A. Guédon and V. Lepetit, "Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 5354–5363.

[11] K. Cheng, X. Long, K. Yang, Y. Yao, W. Yin, Y. Ma, W. Wang, and X. Chen, "Gaussianpro: 3d gaussian splatting with progressive propagation," in *Intl. Conf. on Machine Learning (ICML)*, 2024.

[12] G. Chen and W. Wang, "A survey on 3d gaussian splatting," *arXiv preprint arXiv:2401.03890*, 2024.

[13] J. Zhang, J. Li, X. Yu, L. Huang, L. Gu, J. Zheng, and X. Bai, "Cor-gs: sparse-view 3d gaussian splatting via co-regularization," in *European Conference on Computer Vision*. Springer, 2025, pp. 335–352.

[14] 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.

[15] B. Yamauchi, "A frontier-based approach for autonomous exploration," in *Proceedings 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'97. 'Towards New Computational Principles for Robotics and Automation'*. IEEE, 1997, pp. 146–151.

[16] 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)*. IEEE, 2017, pp. 1396–1402.

[17] 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.

[18] Z. Yan, H. Yang, and H. Zha, "Active neural mapping," in *Intl. Conf. on Computer Vision (ICCV)*, 2023.

[19] 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.

[20] 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.

[21] E. Sucar, S. Liu, J. Ortiz, and A. J. Davison, "imap: Implicit mapping and positioning in real-time," in *Intl. Conf. on Computer Vision (ICCV)*, 2021, pp. 6229–6238.

[22] Z. Zhu, S. Peng, V. Larsson, W. Xu, H. Bao, Z. Cui, M. R. Oswald, and M. Pollefeys, "Nice-slam: Neural implicit scalable encoding for slam," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 12 786–12 796.

[23] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Trans. Graphics*, vol. 41, no. 4, pp. 1–15, 2022.

[24] 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.

[25] C. Jiang, Y. Luo, B. Zhou, and S. Shen, "H3-mapping: Quasi-heterogeneous feature grids for real-time dense mapping using hierarchical hybrid representation," *arXiv preprint arXiv:2403.10821*, 2024.

[26] 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.

[27] C. Yan, D. Qu, D. Xu, B. Zhao, Z. Wang, D. Wang, and X. Li, "Gs-slam: Dense visual slam with 3d gaussian splatting," in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 19 595–19 604.

[28] 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.

[29] J. Wei and S. Leutenegger, "Gsfusion: Online rgb-d mapping where gaussian splatting meets tsdf fusion," 2024. [Online]. Available: https://arxiv.org/abs/2408.12677

[30] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 21 572–21 583.

[31] 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.

[32] S. Lee, L. Chen, J. Wang, A. Liniger, S. Kumar, and F. Yu, "Uncertainty guided policy for active robotic 3d reconstruction using neural radiance fields," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 12 070–12 077, 2022.

[33] 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.

[34] W. Jiang, B. Lei, and K. Daniilidis, "Fisherrf: Active view selection and uncertainty quantification for radiance fields using fisher information," *arXiv*, 2023.

[35] R. Jin, Y. Gao, H. Lu, and F. Gao, "Gs-planner: A gaussian-splatting-based planning framework for active high-fidelity reconstruction," *arXiv preprint arXiv:2405.10142*, 2024.

[36] Q. Du, V. Faber, and M. Gunzburger, "Centroidal voronoi tessellations: Applications and algorithms," *SIAM review*, vol. 41, no. 4, pp. 637–676, 1999.

[37] J. Canny and B. Donald, "Simplified voronoi diagrams," in *Proceedings of the third annual symposium on Computational geometry*, 1987, pp. 153–161.

[38] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "Dbscan revisited, revisited: why and how you should (still) use dbscan," *ACM Transactions on Database Systems (TODS)*, vol. 42, no. 3, pp. 1–21, 2017.

[39] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Transactions on Mathematical Software (TOMS)*, vol. 22, no. 4, pp. 469–483, 1996.

[40] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, 2012.

[41] 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.

[42] 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 *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 9339–9347.

[43] F. Xia, A. R. Zamir, Z. He, A. Sax, J. Malik, and S. Savarese, "Gibson Env: real-world perception for embodied agents," in *Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on*. IEEE, 2018.

[44] 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," *International Conference on 3D Vision (3DV)*, 2017.

[45] 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.