ActiveGAMER: Active GAussian Mapping through Efficient Rendering

Supplementary Material

6. Overview

In this supplementary material, we provide a detailed outline structured as follows: Sec. 7 delves into additional implementation specifics of ActiveGAMER. Sec. 8 examines the computation costs associated with each module. Complementing the results in Sec. 4, Sec. 9 extends our analysis for MP3D and Replica, including per-scene reconstruction and novel view rendering results, both quantitative and qualitative. We also include trajectory visualizations to illustrate the planned trajectories, and evaluate the trajectory lengths across scenes. Lastly, we present analysis and visualizations of common failure cases observed in our method.

7. Implementation Details

7.1. Hardware Requirements

We conduct our experiments on a desktop PC equipped with a 2.2GHz Intel Xeon E5-2698 CPU and an NVIDIA V100 GPU. Memory consumption varies depending on the scene size. For reference, in an $80\,\mathrm{m}^3$ scene, the GPU memory and RAM usage are as follows across different stages:

- **Coarse-Exploration**: GPU memory consumption is approximately 5.3GB, and RAM usage is 1GB.
- **Fine-Exploration**: GPU memory increases to 6.2GB due to the addition of more Gaussians.
- **Post-Refinement**: GPU memory reaches 6.5GB, and RAM usage increases to 1.8GB as more optimization iterations are performed.

7.2. Coarse-to-fine Exploration Details

Avoiding candidate sampling near surfaces. While building the Exploration Candidate Pool, we intentionally avoid sampling candidates from free spaces close to surfaces for two key reasons: (1) Rendering Limitations: The Gaussian Map renderer ignores 3D Gaussians near the camera. When candidates are close to surfaces (e.g., walls), the most informative viewpoints tend to face the surface directly. This results in low-visibility masks due to the renderer ignoring nearby Gaussians, leading to incorrect candidate selection. (2) Collision Prevention: To simulate collision avoidance between the agent/camera and the surface, we maintain a buffer of free space between them.

7.3. Post-Refinement Details

During the exploration stages, each Gaussian Map update step uses only 15 optimization iterations, and a quarter of the image resolution is employed for densification. In the post-refinement stage, the number of optimization iterations increases from 15 to 60, and the full image resolution is utilized for densification.

8. Runtime Analysis

In this section, we present a runtime analysis of the three major modules in ActiveGAMER:

- **Data Generation**: A simulator generates RGB-D data.
- **Gaussian Mapping Module**: Updates the Gaussian Map.
- **Rendering-Based Planning Module**: Includes a global planner for goal searching and a local planner for path planning.

For data generation, HabitatSim is required to generate 680×1200 RGB-D data per iteration. The detailed runtime for each stage is presented in Tab. 4. We compute the statistics using average.

Two types of statistics are reported in the table:

- **Per-iteration timing**: The runtime for each individual module when it is activated.
- **Overall timing per stage**: The average runtime considering the infrequent activation of certain modules.

The **Gaussian Mapping Module** is active only during keyframe steps (*i.e.* every five frames) in the exploration stage, reducing its average runtime. The **Planning Module** is triggered on demand, contributing minimal runtime overhead. In contrast, **Post-Refinement** involves more intensive computation with additional iterations.

9. Additional Experimental Results

9.1. Per-scene Results on Replica

3D Reconstruction In Tab. 5, we present the 3D reconstruction results on the Replica dataset [60], comparing our method with the state-of-the-art active mapping approach, NARUTO [18]. Our method demonstrates comparable reconstruction performance to NARUTO, achieving higher accuracy and completeness on average while maintaining a similar completeness ratio. The slightly lower completeness ratio in our results may be attributed to the extrapolation capability inherent in NARUTO's neural radiance field representation.

Qualitative Results In Fig. 6, we present a qualitative evaluation of our method's rendering performance against the ground truth across various scenes in the Replica dataset. The evaluation involves a novel view trajectory with a 360° circular movement (an outside-in trajectory looking at a specific point in the scene). To provide a

Stage	Module	Per Iteration (s)	Per Stage (s)
All	HabitatSim	0.244	0.244
Coarse-Exploration	Gaussian Mapping	0.687	0.181
	Planning - Global	1.126	0.071
	Planning - Local	0.018	
Fine-Exploration	Gaussian Mapping	0.759	0.184
	Planning - Global	5.411	0.283
	Planning - Local	0.023	
Post-Refinement	Gaussian Mapping	2.580	2.580

Table 4. **Runtime Analysis of ActiveGAMER Modules on Replica-room0.**

Methods	Metrics	office0	office1	office2	office3	office4	room0	room1	room2	Avg.
NARUTO[18]	Acc. (cm) ↓	1.30	1.03	2.25	2.29	1.75	1.56	1.25	1.47	1.61
	Comp. (cm) ↓	1.39	1.53	1.69	2.27	1.79	1.68	1.43	1.48	1.66
	Comp. Ratio ↑	98.17	95.26	97.54	93.91	97.93	98.28	98.04	98.47	**97.20**
Ours	Acc. (cm) ↓	1.35	1.07	1.06	1.03	0.84	1.22	1.40	1.31	**1.16**
	Comp. (cm) ↓	1.93	1.26	1.31	1.13	1.02	1.89	2.06	1.89	**1.56**
	Comp. Ratio (%) ↑	96.29	98.14	97.91	98.12	98.18	94.33	93.93	95.15	96.50

Table 5. **3D Reconstruction Results on Replica[60].**

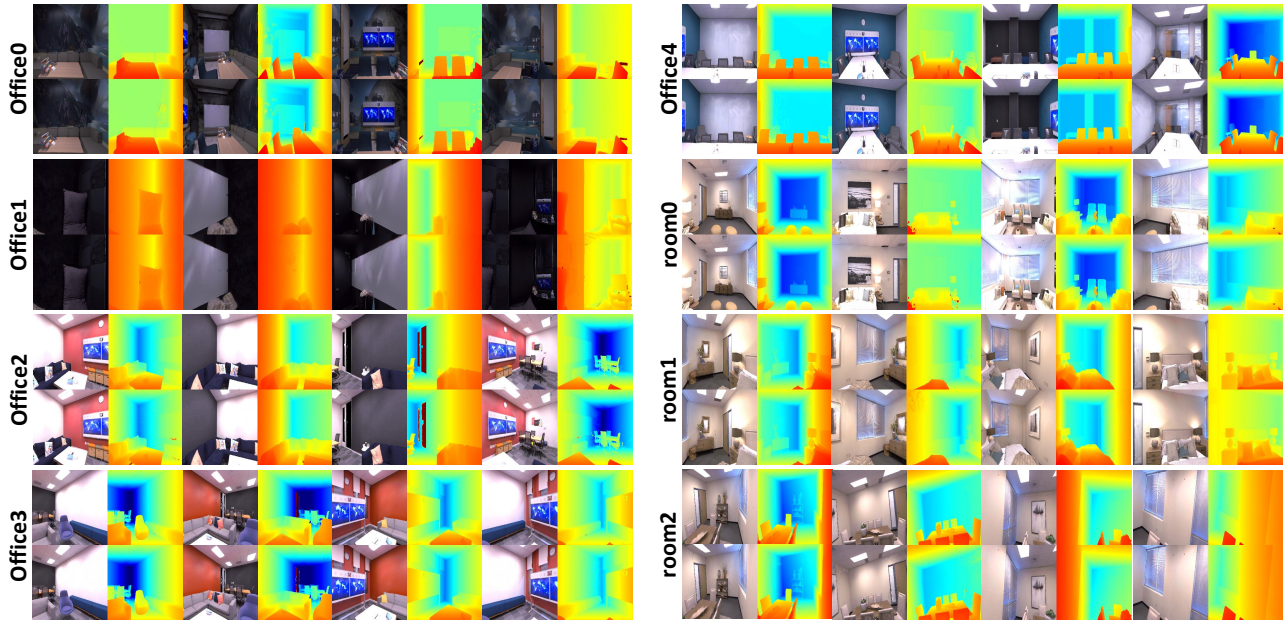

Figure 6. **Novel View Rendering Results on Replica.** This figure provides a side-by-side comparison of rendering results on the Replica dataset. Upper rows show the ground truth RGB-D images, while bottom rows present the rendered RGB-D images.

comprehensive visualization, we present the rendering results from four directions: front, left, right, and back. Our method demonstrates high fidelity in both RGB and depth rendering.

9.2. Per-scene Results on MP3D

3D Reconstruction In Tab. 6, we present a comparative analysis of our method against the state-of-the-art approaches, Active Neural Mapping (ANM) [69] and

Methods	Metrics	Gdvg	gZ6f	HxpK	pLe4	YmJk	Avg.
ANM [69]	Acc. (cm) ↓	5.09	4.15	15.60	5.56	8.61	7.80
	Comp. (cm) ↓	5.69	7.43	15.96	8.03	8.46	9.11
	Comp. Ratio ↑	80.99	80.68	48.34	76.41	79.35	73.15
NARUTO[18]	Acc. (cm) ↓	3.78	3.36	9.24	5.15	10.04	6.31
	Comp. (cm) ↓	2.91	2.31	2.67	3.24	3.86	3.00
	Comp. Ratio ↑	91.15	95.63	91.62	87.76	84.74	90.18
Ours	Acc. (cm) ↓	1.28	1.55	1.46	2.13	1.87	**1.66**
	Comp. (cm) ↓	2.62	1.74	1.68	2.63	2.83	**2.30**
	Comp. Ratio ↑	97.62	97.74	97.29	93.96	89.87	**95.32**

Table 6. **3D Reconstruction Results on Matterport3D [7] dataset**. Our method achieves consistently better reconstruction than prior methods.

Methods	Metrics	Gdvg	gZ6f	HxpK	pLe4	YmJk	Avg.
NARUTO [18]	PSNR ↑	22.40	22.65	14.74	21.63	21.20	20.52
	SSIM ↑	0.81	0.85	0.45	0.74	0.75	0.72
	LPIPS ↓	0.50	0.48	0.76	0.62	0.56	0.58
	L1-D ↓	5.11	6.00	9.82	9.91	8.94	7.95
Ours	PSNR ↑	25.51	26.87	22.48	26.75	22.20	**24.76**
	SSIM ↑	0.92	0.96	0.82	0.94	0.84	**0.90**
	LPIPS ↓	0.20	0.15	0.34	0.19	0.37	**0.25**
	L1-D ↓	1.92	1.54	6.42	3.59	10.67	**4.83**

Table 7. **Novel View Rendering results on Matterport3D[7] dataset.** Our method shows consistently better rendering result than NARUTO.

NARUTO [18]. The results clearly demonstrate that our method outperforms both ANM and NARUTO across all evaluated metrics. Notably, our approach achieves significant improvements in reconstruction quality and completeness, surpassing the benchmarks established by prior methods, thanks to the explicit point cloud representation (extracted from Gaussian Map). This consistent performance advantage highlights the effectiveness of our method in tackling challenging reconstruction scenarios.

Novel View Rendering We further evaluate the novel view rendering performance on MP3D, as shown in Tab. 7. For this evaluation, we generate a circular novel view trajectory in each selected scene and compare the performance of NARUTO [18] and our proposed method. Leveraging the Gaussian Mapping representation, our method consistently demonstrates superior rendering performance compared to NARUTO, which relies on a neural radiance field representation.

Method	Avg	O0	O1	O2	O3	O4	R0	R1	R2
NARUTO	74.83	81.27	30.02	90.20	88.59	96.36	73.91	96.99	41.31
Ours	**66.59**	46.13	35.18	64.16	95.13	78.87	93.92	53.42	65.87

Table 8. Trajectory length evaluation on Replica dataset

Qualitative Results In Fig. 7, we present a qualitative evaluation of our method compared to the ground truth for various scenes in the Matterport3D dataset. The odd-numbered rows display the ground truth data, while the even-numbered rows showcase our method's geometric and photometric reconstructions. Each scene is labeled with a unique code (*e.g.*, "Gdvg", "gZ6f") on the left. The first column highlights the exterior reconstruction; the second and fourth columns present RGB rendering results of the interior space; and the third and fifth columns show the corresponding depth rendering results. This result provides a comprehensive visual comparison, effectively illustrating the geometric and photometric performance of our method across the scenes.

9.3. Trajectory Results

We visualize the planned trajectories in Replica scenes, as shown in Fig. 8. During Coarse-Exploration, the planned camera movement is restricted to a 2D plane, as candidate sampling is limited to a single height level. In Fine-Exploration, two height levels are defined, allowing the camera to search for candidates across these levels.

We further compare the trajectory lengths of ours versus NARUTO [18] on the Replica dataset, with the results detailed in Tab. 8. Our approach exhibits a shorter average trajectory length than NARUTO, underscoring our efficiency in terms of travel distance.

9.4. Failure Cases

Although our method demonstrates high-quality reconstruction and high-fidelity rendering, some failure cases are observed, as illustrated in Fig. 9.

Ignored Candidates Near Surface Regions. As discussed in Sec. 7.2, we avoid sampling candidates from regions near surfaces. Consequently, regions that can only be observed from such candidates remain unobserved, leading to incomplete reconstruction in these areas.

Double-Sided Objects. Our method relies on rendering-based exploration information derived from the rendered visibility mask. For double-sided objects, rendering from the back side does not provide additional information, and such viewpoints are not considered informative. This results in incomplete reconstruction of back-side regions. To address this limitation, we plan to incorporate surface information into the rendering-based information gain in future work, enhancing the robustness of our exploration strategy.

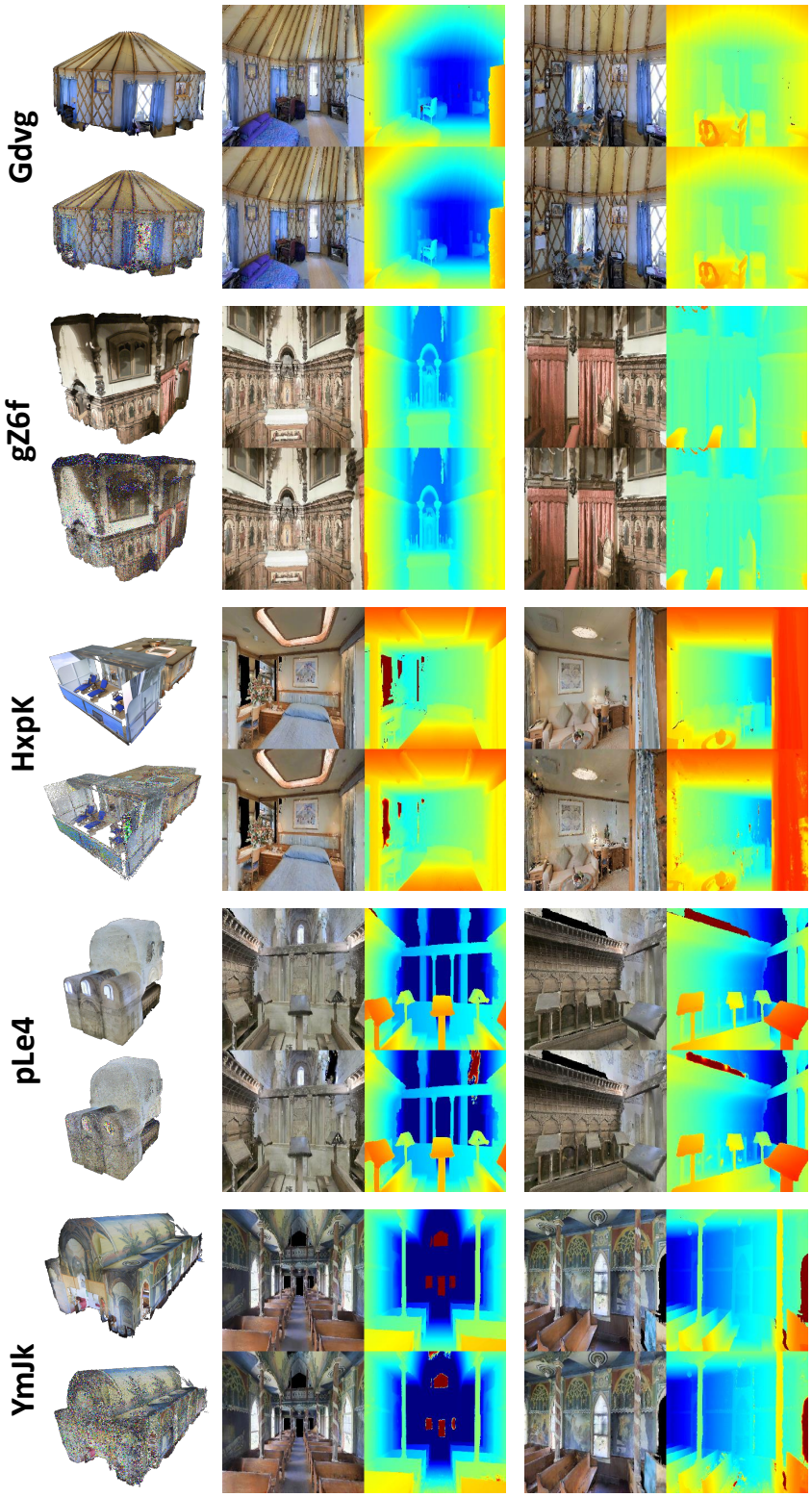

Figure 7. **MP3D Reconstruction and Rendering Results.** This figure provides a side-by-side comparison of reconstruction results on the MP3D dataset. Odd-numbered rows show the ground truth meshes and RGB-D images, while even-numbered rows present the 3D point cloud extracted from the Gaussian Map and the rendered RGB-D images. The noisy points represent low-weight Gaussians and do not affect actual rendering quality. Our results demonstrate high-quality and complete reconstructions, closely aligning with the ground truths in both geometric and photometric aspects, highlighting the effectiveness of our method in accurately reconstructing complex spatial geometries.

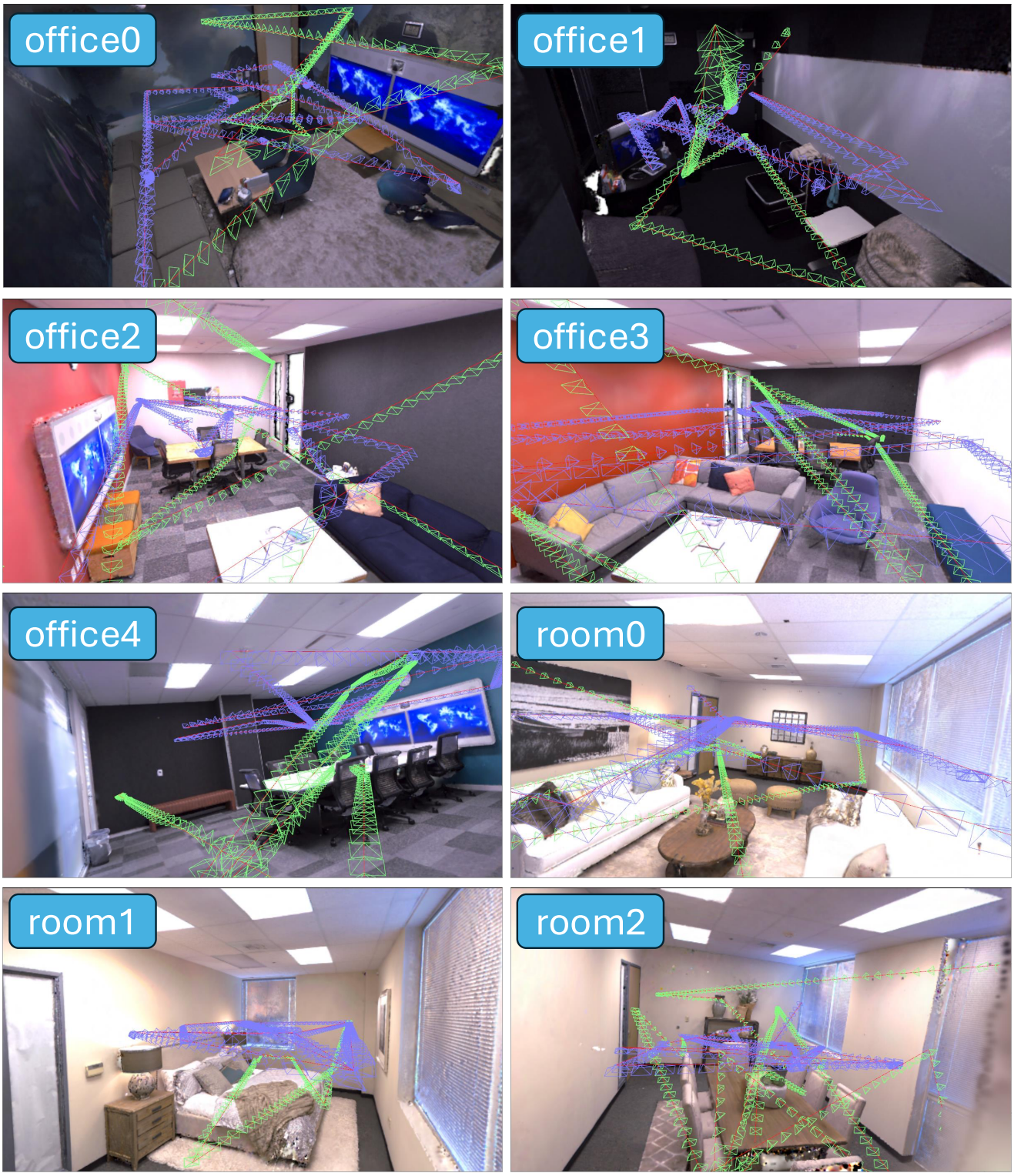

Figure 8. The exploration trajectories for selected scenes in Replica are visualized, showing only keyframe cameras. Camera colors indicate the transition from coarse to fine exploration stages.

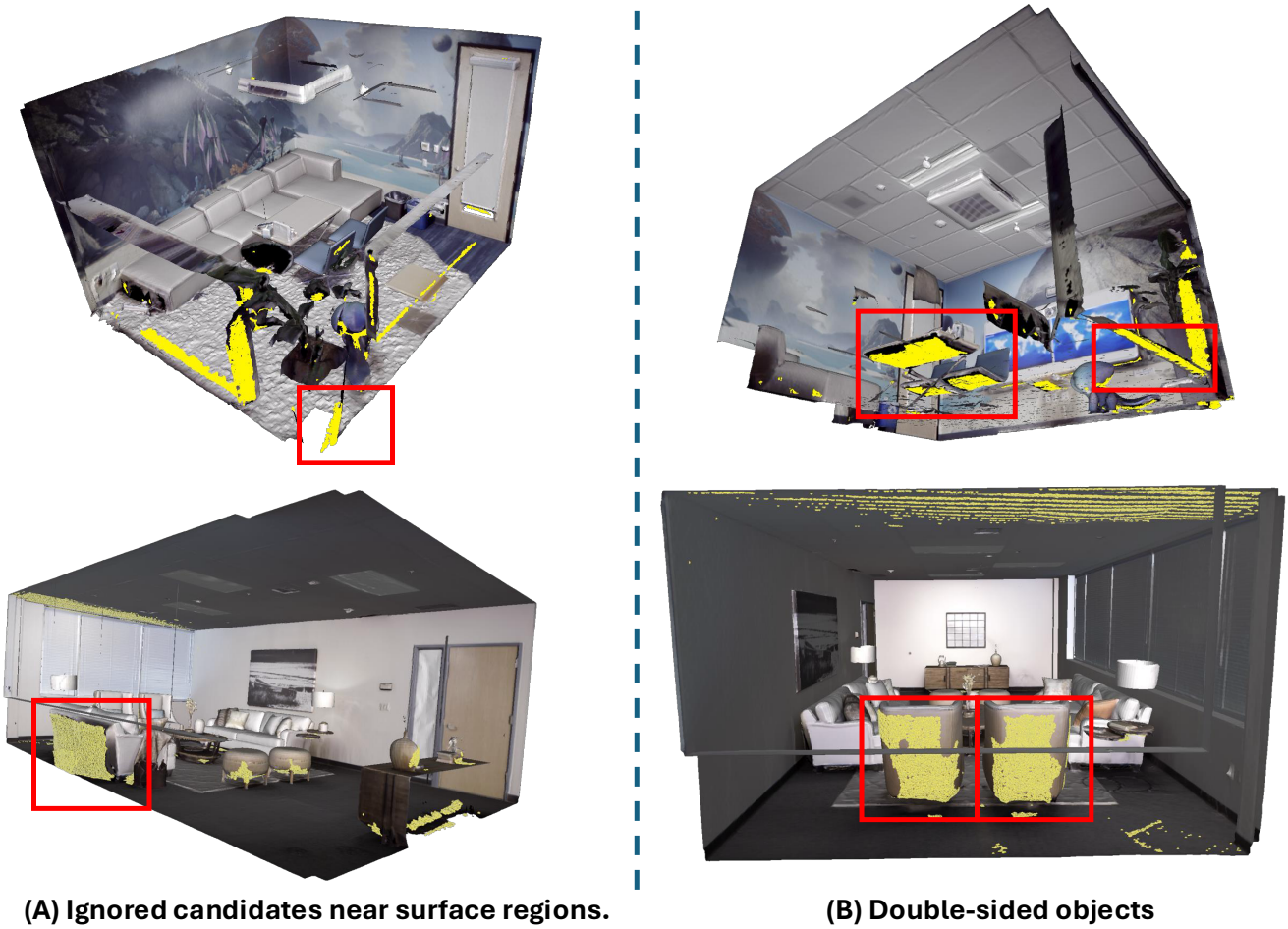

Figure 9. **Reconstruction Failure Cases.** Our method encounters two common failure scenarios: (A) Insufficient observations due to ignored candidates near surface regions. (B) Unreconstructed back sides of double-sided objects, as these regions are not captured by the rendering-based information gain. The yellow-highlighted areas indicate regions with either no reconstruction or low completeness.