MAGICIAN: Efficient Long-Term Planning with Imagined Gaussians for Active Mapping

Supplementary Material

In Sec. A, we present the details of the occupancy module, and the complete formulation of the coverage gain computation along with the analytical derivation of the depth-dependent weighting. In Sec. B, we provide additional implementation details, detailed tables, additional quantitative comparisons, and additional ablation studies. In Sec. C, we discuss observed failure cases and provide an analysis.

A. Method

A.1. Neural Occupancy Prediction

Here, we provide additional architectural details of the volume occupancy module $\hat{\sigma}(\mathbf{x} \mid \mathbf{C}_t)$.

At each time step t, the occupancy module receives a 3D query point $\mathbf{x}$, the reconstructed surface point cloud S_t, and the previously visited camera poses $\mathbf{C}_t$, and predicts an occupancy value in $[0, 1]$ for $\mathbf{x}$.

To capture the local geometry around $\mathbf{x}$, we compute its k-nearest neighbors in S_t and encode this neighborhood using a self-attention unit followed by pooling. To capture larger-scale structure, we repeat this procedure on progressively downsampled versions of S_t: at each scale, we recompute the neighbors of $\mathbf{x}$ and process them with an additional self-attention–pooling block. Coarser scales naturally expand the receptive field, allowing the model to integrate fine-grained and global geometric information.

The multi-scale features are concatenated and fed into an MLP to predict the occupancy value $\hat{\sigma}(\mathbf{x} \mid \mathbf{C}_t)$. Because the architecture operates solely on local neighborhoods at each scale, it can be applied efficiently to large point clouds while still preserving fine geometric details. In practice, we set $k = 16$ and use three neighborhood scales.

We adopt this model architecture from [19] without modification. The diagram of this model architecture is presented in Figure 7 of that work.

A.2. Coverage Gain Computation

A.2.1. Coverage Gain Formulation

For each candidate camera pose $\mathbf{c}$, we compute the coverage gain $G_{\text{rendered}}(\mathbf{c})$ by rendering depth and novelty maps from the current Imagined Gaussian state using volumetric rendering (Eq. (4) in the main paper):

$$G_{\text{rendered}}(\mathbf{c}) = \sum_{\mathbf{p} \in \mathcal{P}_{\text{valid}}} w_{\text{depth}}(\mathbf{p})\, I_{\text{novelty}}(\mathbf{p}), \tag{5}$$

where $\mathcal{P}_{\text{valid}} = \{\mathbf{p} \mid D(\mathbf{p}) > 0\}$ denotes pixels with valid depth $D(\mathbf{p})$, $I_{\text{novelty}}(\mathbf{p})$ is the rendered novelty value, and $w_{\text{depth}}(\mathbf{p})$ is a depth-dependent weighting factor:

$$w_{\text{depth}}(\mathbf{p}) = \min\left(1, \left(\frac{D(\mathbf{p})}{D_{\text{th}}}\right)^2\right), \tag{6}$$

where D_{th} denotes a threshold and is set to half of the estimated scene scale. This weighting term mitigates oversampling at close range, where the pixel sampling density of the depth sensor exceeds the resolution required for faithful surface reconstruction.

A.2.2. Analytical Derivation of Depth Weighting

Target surface density. Surface coverage becomes well defined only after specifying a target spatial resolution. For large urban scenes, one point per square decimeter may suffice, whereas tabletop objects typically require several points per square centimeter. We denote this desired sampling resolution as the *target surface density* r_{target}, representing the minimum number of points per unit surface area required for adequate reconstruction. This concept is commonly used in existing methods [18, 19].

Since the depth sensor always captures a fixed number of samples $N = H \times W$ per frame, the local surface sampling density depends solely on the distance between the camera and the observed surface. By Thales's theorem, this density decays quadratically with depth. Consequently, when the sensor is too close to a surface, the resulting sample density exceeds r_{target} and provides no additional benefit for coverage. Thus, r_{target} naturally induces a threshold depth D_{th}, below which moving the camera closer becomes inefficient.

Mathematical derivation. Consider a square patch of the depth map with side length s centered at pixel $\mathbf{p}$. This patch contains

$$n_{\text{captured}} = s^2 \tag{7}$$

captured depth samples. By Thales's theorem, the corresponding 3D surface region has area:

$$A = \left(\frac{s\, D(\mathbf{p})}{f}\right)^2, \tag{8}$$

where f is the focal length in pixel units. The resulting surface sampling density is therefore:

$$r(\mathbf{p}) = \frac{n_{\text{captured}}}{A} = \left(\frac{f}{D(\mathbf{p})}\right)^2, \tag{9}$$

confirming the inverse-square relationship with depth. The depth at which $r(\mathbf{p})$ equals the target density r_{target} is ob-

Table 6. **AUCs of full scenes on the Macarons++ dataset.**

Scene	Rand. Walk	SCONE [18]	MACARONS[19]	FisherRF [26]	**MAGICIAN (Ours)**
Dunnottar Castle	0.149	0.366	0.618	0.500	0.745
Colosseum	0.219	0.589	0.656	0.551	0.704
Bannerman Castle	0.192	0.559	0.575	0.595	0.761
Pantheon	0.198	0.465	0.601	0.270	0.644
Christ the Redeemer	0.439	0.772	0.859	0.727	0.793
Statue of Liberty	0.323	0.632	0.711	0.553	0.797
Pisa Cathedral	0.290	0.486	0.678	0.486	0.723
Fushimi Castle	0.279	0.689	0.718	0.565	0.766
Alhambra Palace	0.126	0.369	0.567	0.462	0.631
Neuschwanstein Castle	0.184	0.325	0.452	0.375	0.608
Eiffel Tower	0.333	0.683	0.709	0.616	0.754
Manhattan Bridge	0.258	0.632	0.750	0.637	0.705
St. Sofia Church	0.280	0.532	0.621	0.608	0.710
Barts	0.214	0.551	0.660	0.673	0.831
Sestino Museum	0.132	0.367	0.537	0.571	0.637
Average	0.241	0.534	0.647	0.546	0.721

Table 7. **Final Coverages of full scenes on the Macarons++ dataset.**

Scene	Rand. Walk	SCONE [18]	MACARONS [19]	FisherRF [26]	**MAGICIAN (Ours)**
Dunnottar Castle	0.225	0.527	0.820	0.809	0.975
Colosseum	0.272	0.755	0.794	0.757	0.872
Bannerman Castle	0.240	0.722	0.834	0.801	0.917
Pantheon	0.309	0.610	0.796	0.444	0.842
Christ the Redeemer	0.581	0.924	0.967	0.876	0.973
Statue of Liberty	0.443	0.819	0.909	0.819	0.947
Pisa Cathedral	0.353	0.566	0.865	0.776	0.941
Fushimi Castle	0.449	0.844	0.853	0.814	0.931
Alhambra Palace	0.162	0.473	0.775	0.615	0.852
Neuschwanstein Castle	0.223	0.444	0.551	0.582	0.848
Eiffel Tower	0.541	0.856	0.915	0.827	0.923
Manhattan Bridge	0.356	0.781	0.924	0.877	0.955
St. Sofia Church	0.331	0.619	0.795	0.865	0.891
Barts	0.240	0.677	0.768	0.878	0.996
Sestino Museum	0.141	0.430	0.713	0.842	0.924
Average	0.324	0.670	0.819	0.786	0.919

tained by solving $r(\mathbf{p}) = r_{\text{target}}$:

$$D_{\text{th}} = \frac{f}{\sqrt{r_{\text{target}}}}. \tag{10}$$

For depths $D(\mathbf{p}) < D_{\text{th}}$, the captured sample density is unnecessarily high. In this regime, although the patch contains $n_{\text{captured}} = s^2$ samples, only $A\, r_{\text{target}}$ samples are needed to meet the target surface density. The fraction of samples that meaningfully contribute to coverage is thus:

$$p(\mathbf{p}) = \frac{A\, r_{\text{target}}}{n_{\text{captured}}} = \frac{r_{\text{target}}}{r(\mathbf{p})} = \left(\frac{D(\mathbf{p})}{D_{\text{th}}}\right)^2. \tag{11}$$

Pixels observed at depths smaller than D_{th} should therefore contribute only proportionally to $p(\mathbf{p})$, reflecting the redundancy introduced by oversampling in this regime.

Conversely, when $D(\mathbf{p}) \geq D_{\text{th}}$, the sampling density satisfies $r(\mathbf{p}) \leq r_{\text{target}}$, meaning that all captured samples are necessary and should contribute fully. Combining both regimes yields the depth-dependent weighting function:

$$w_{\text{depth}}(\mathbf{p}) = \min\left(1, \left(\frac{D(\mathbf{p})}{D_{\text{th}}}\right)^2\right). \tag{12}$$

This weighting strategy, used in Eq. (5), prevents the planner from favoring near-surface viewpoints that artificially inflate point counts without improving effective surface coverage. As a result, the exploration process is guided toward trajectories that yield more efficient and informative observations.

B. Experiments

B.1. Implementation Details

Our simulation is built on PyTorch3D [39], which supports differentiable rendering and ray casting to generate RGB-D data from arbitrary camera viewpoints. The pretrained occupancy model was trained using four NVIDIA Tesla V100 SXM2 32 GB GPUs, while inference was performed on a single V100 GPU.

In our experiments on the Macarons++ dataset, we evaluated Final Coverage and AUC scores using ground-truth point clouds. However, unlike prior work [19, 32] that directly samples point clouds from the ground-truth mesh, which may include invisible points (e.g., points inside Pisa Cathedral), we generated the ground-truth point cloud by rendering depth maps from all accessible viewpoints and projecting them into a 3D point cloud.

For each scene, we evaluate 15 novel views. To obtain a set of novel views that cover the entire ground-truth mesh, we use a submodular optimization–based selection procedure. At each iteration, we randomly sample 100 candidate 6D poses within the scene's bounding box and, for each pose, count how many ground-truth points are visible from that viewpoint. We then select the pose that observes the largest number of previously unseen ground-truth points and mask out those newly observed points from the ground-truth point cloud. We repeat this process by sampling a new batch of 100 candidate poses and again selecting the pose that reveals the most remaining unseen points, until 15 novel views are selected.

B.2. Comparison with State-of-the-Art Methods

In this section, we provide detailed evaluation results on the Macarons++ dataset, along with additional qualitative comparisons and analyses.

From Tab. 6 and Tab. 7, we observe that the state-of-the-art NBV-based method MACARONS [19] remains a very strong baseline in relatively simple scenes such as Manhattan Bridge and Christ the Redeemer. However, due to its lack of long-term planning, it struggles to escape already fully explored local regions, which leads to poor performance in indoor environments. FisherRF [26], which relies on frontier detection and Fisher information, performs reasonably better in indoor environments due to its frontier-based exploration. However, the frontier mechanism also introduces unnecessary movements, leading to inefficient trajectories, particularly in outdoor scenes. In contrast, our method is neither restricted by frontier heuristics nor hampered by short-sighted planning. By performing the tree search to identify full trajectories that maximize coverage gain, our method achieves state-of-the-art performance in both indoor and outdoor scenes.

As we mentioned in the main paper, during the evaluation stage, the five starting poses in each scene are randomly sampled. To more rigorously evaluate the stability of each method under this randomness, we compute the standard deviation of the final coverage for each method in each scene, and further compute their average across all scenes to summarize the overall variability. The results shown in Fig. 9 demonstrate that our method exhibits consistently low values in this metric, indicating that its performance is highly robust: despite different random initial poses, it reliably achieves high final coverage. In contrast, the other methods exhibit substantially larger variance, suggesting that their performance is highly sensitive to the initial pose and the corresponding early observations.

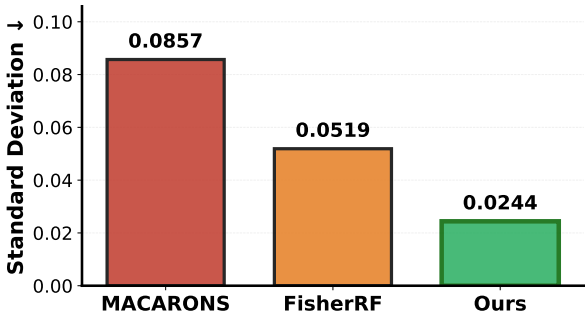


Figure 9. **Standard deviation of the final coverage across different methods and scenes.** Our method achieves consistently low values for this metric, indicating strong robustness to random starting poses, whereas other methods exhibit much larger variability.

Table 8. Ablation study on look-ahead steps N_d.

N_d	10	15	20	25	30	50
AUC	0.652	**0.673**	0.664	0.662	0.658	0.652
Cov.	0.888	0.887	**0.892**	0.879	0.881	0.878

In Fig. 10 and Fig. 11, we present visualizations of the exploration trajectories generated by different methods, where for each scene all methods start from the same initial pose, along with qualitative comparisons of novel view synthesis and mesh-based normal maps. Under an identical movement budget, our method achieves thorough exploration in both indoor and outdoor environments, resulting in high-quality reconstructions, whereas incomplete exploration by the other methods leads to noticeably inferior reconstruction quality.

B.3. Additional Ablation Study

Impact of longer-range look-ahead steps N_d. Table 8 presents the results for increased look-ahead steps $N_d > 10$. Performance peaks when $N_d = 15$–20; while it slightly declines for larger values, it remains superior to shorter look-ahead steps, as shown in Figure 7.

Robustness under pose uncertainty. We corrupt cam-

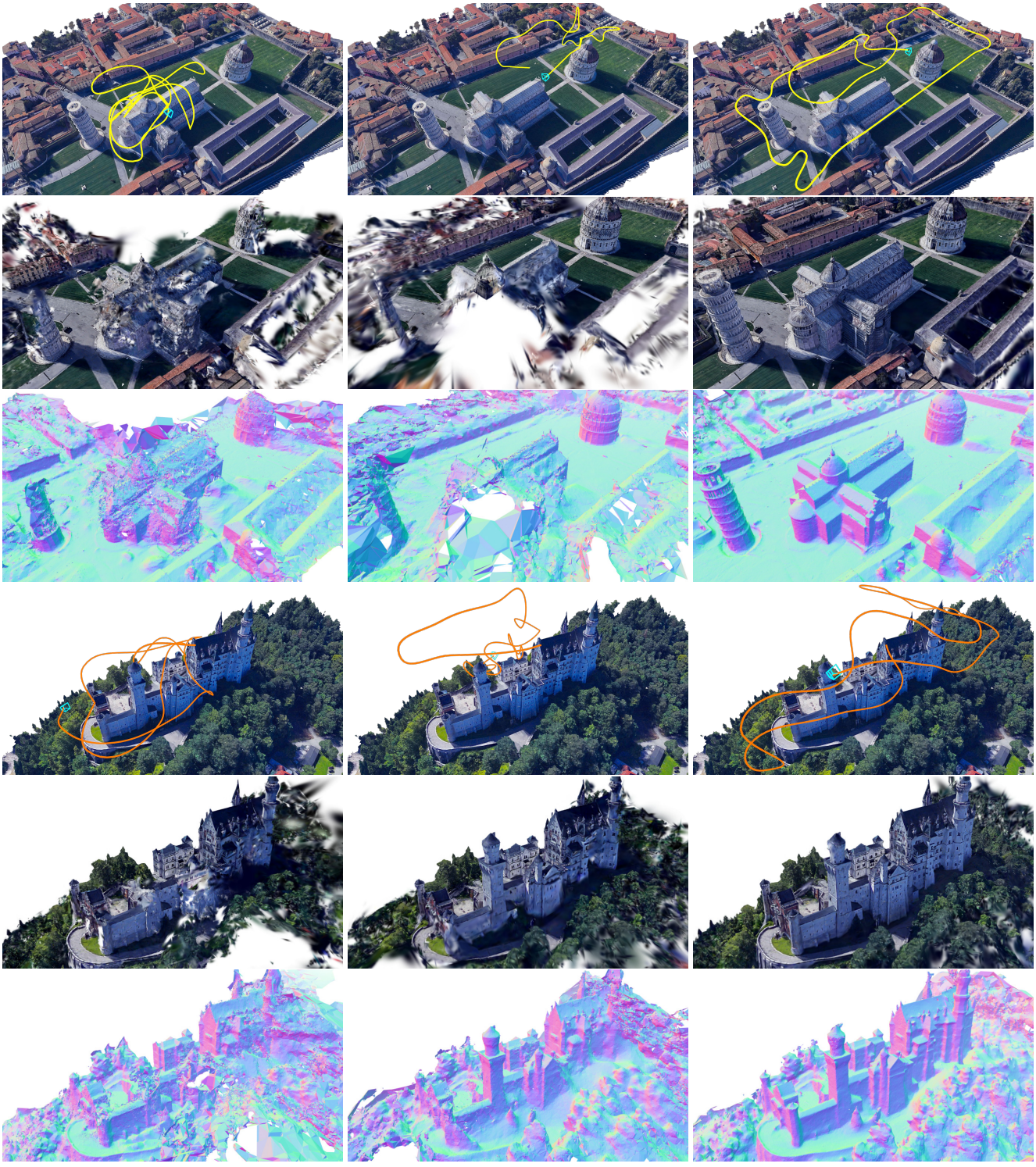
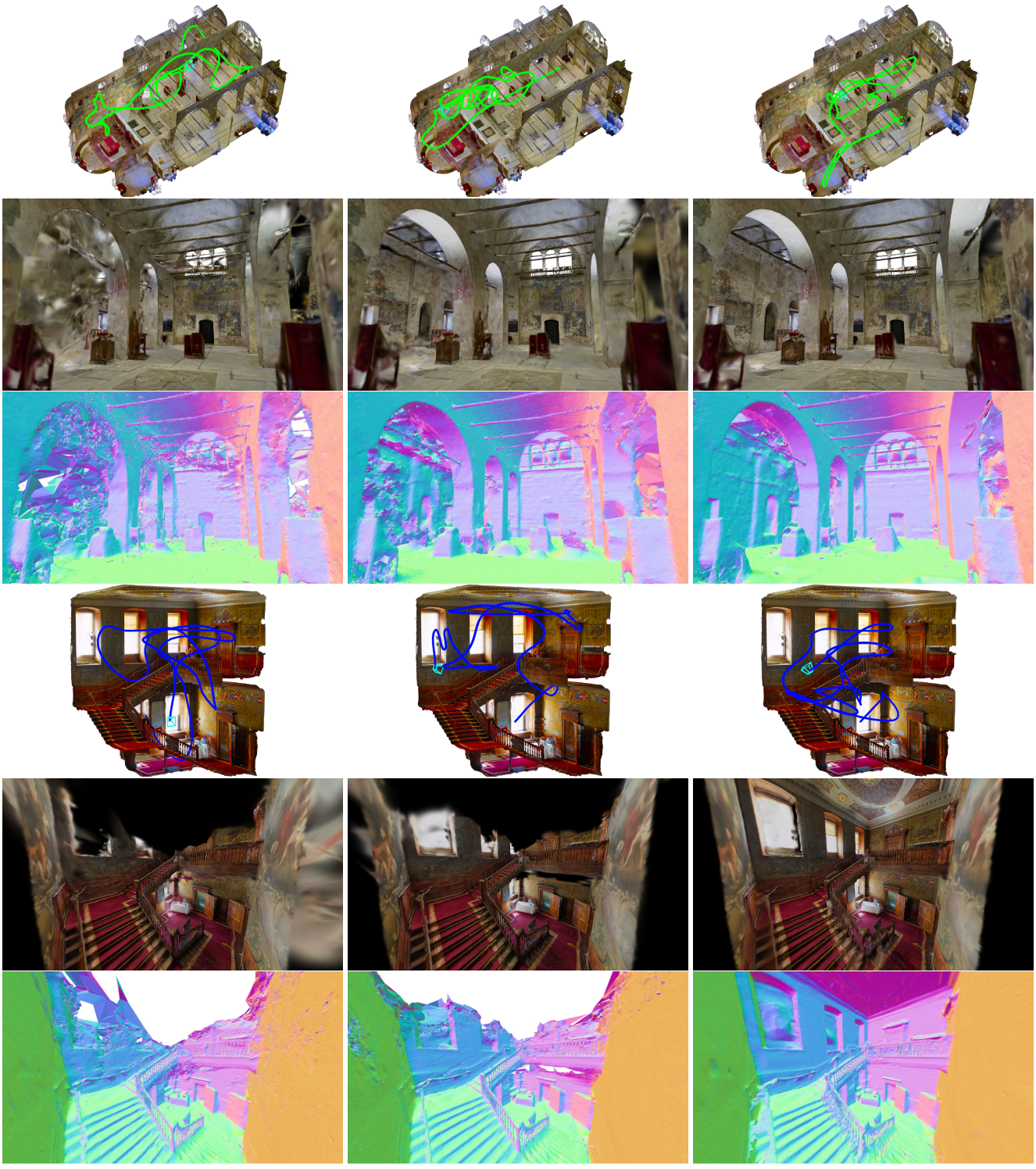


Figure 10. **Visualization of exploration trajectories and qualitative comparisons of novel view synthesis and surface reconstruction in outdoor scenes. From top to bottom, the scenes are Pisa Cathedral and Neuschwanstein Castle.** In the same scene, all methods start from the same initial camera pose, and for each trajectory visualization, we additionally show the final camera pose at the end of the trajectory. Our trajectory planning method yields more accurate and complete reconstructions, resulting in higher-quality renderings and effectively preventing holes or noise in the reconstructed surfaces.



FisherRF [26] MACARONS [19] Ours

Figure 11. **Visualization of exploration trajectories and qualitative comparisons of novel view synthesis and surface reconstruction in indoor scenes. From top to bottom, the scenes are St. Sofia Church and Barts.** In the same scene, all methods start from the same initial camera pose, and for each trajectory visualization, we additionally show the final camera pose at the end of the trajectory. Our trajectory planning method yields more accurate and complete reconstructions, resulting in higher-quality renderings and effectively preventing holes or noise in the reconstructed surfaces.

Table 9. Ablation study on proxy point sampling density ($1\times$ indicates the original density).

Density	$0.5\times$	$1\times$	$2\times$	$4\times$
AUC	0.640	0.652	0.672	**0.685**
Cov.	0.848	0.888	0.895	**0.905**

era poses with Gaussian noise (σ = 0.5m translation, 3° rotation) during planning. These are deliberately larger than typical localization errors to rigorously stress-test the method. Under this setting, performance decreases only marginally, with AUC dropping from 0.652 to 0.649 (-0.28 pp) and Cov. decreasing from 0.888 to 0.877 (-1.12 pp), demonstrating strong robustness to substantial pose uncertainty.

Effect of proxy point sampling density. Table 9 shows that while increasing proxy point density leads to steady improvements in AUC and Cov. by refining coverage gain estimates, the performance remains relatively stable across a broad range of densities. This suggests that our method is robust to sampling density, with a $1\times$ density already providing a strong balance between estimation accuracy and computational overhead.

C. Failure Case and Analysis

In a few scenes, we observe that the occupancy model exhibits reduced accuracy during the early stages of exploration, which leads to lower initial exploration efficiency. This limitation arises because the occupancy model is fundamentally geometric, relying on features extracted from local 3D neighborhoods. While such local geometric priors are effective at capturing generalizable primitives across scales and domains, they may be insufficient to provide a reliable global understanding when observations are sparse. As a result, the planner may not accurately identify the most informative regions at the beginning, leading to suboptimal estimation of coverage gain. However, as more observations are accumulated, the environment representation is progressively refined, and the system mitigates this issue through frequent closed-loop replanning, ultimately improving exploration performance over time.

References

[1] Shi Bai, Jinkun Wang, Fanfei Chen, and Brendan Englot. Information-Theoretic Exploration with Bayesian Optimization. In *International Conference on Intelligent Robots and Systems*, pages 1816–1822, 2016. 2

[2] Joseph E. Banta, L. R. Wong, Christophe Dumont, and Mongi A. Abidi. A Next-Best-View System for Autonomous 3D Object Reconstruction. *IEEE Transactions on Systems, Man, and Cybernetics*, 30(5):589–598, 2000. 2

[3] Ana Batinovic, Tamara Petrovic, Antun Ivanovic, Frano Petric, and Stjepan Bogdan. A Multi-Resolution Frontier-Based Planner for Autonomous 3D Exploration. *IEEE Robotics and Automation Letters*, 6(3):4528–4535, 2021. 2

[4] Andreas Bircher, Mina Kamel, Kostas Alexis, Helen Oleynikova, and Roland Siegwart. Receding Horizon" Next-Best-View" Planner for 3D Exploration. In *International Conference on Robotics and Automation*, pages 1462–1468, 2016. 2, 3

[5] Frederic Bourgault, Alexei A. Makarenko, Stefan B. Williams, Ben Grocholsky, and Hugh F. Durrant-Whyte. Information Based Adaptive Robotic Exploration. In *International Conference on Intelligent Robots and Systems*, pages 540–545, 2002. 2

[6] Chao Cao, Ji Zhang, Matt Travers, and Howie Choset. Hierarchical coverage path planning in complex 3d environments. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3206–3212. IEEE, 2020. 2

[7] Chao Cao, Hongbiao Zhu, Howie Choset, and Ji Zhang. TARE: A Hierarchical Framework for Efficiently Exploring Complex 3D Environments. *Robotics: Science and Systems*, 5:2, 2021. 2

[8] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3D: Learning from RGB-D Data in Indoor Environments. In *arXiv Preprint*, 2017. 2, 5

[9] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. ShapeNet: An Information-Rich 3D Model Repository. In *arXiv Preprint*, 2015. 4

[10] Liyan Chen, Huangying Zhan, Kevin Chen, Xiangyu Xu, Qingan Yan, Changjiang Cai, and Yi Xu. ActiveGamer: Active Gaussian Mapping through Efficient Rendering. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 16486–16497, 2025. 2, 3, 6, 8

[11] Liyan Chen, Huangying Zhan, Hairong Yin, Yi Xu, and Philippos Mordohai. Understanding while exploring: Semantics-driven active mapping. *arXiv preprint arXiv:2506.00225*, 2025. 8

[12] Xiao Chen, Tai Wang, Quanyi Li, Tao Huang, Jiangmiao Pang, and Tianfan Xue. GLEAM: Learning Generalizable Exploration Policy for Active Mapping in Complex 3D Indoor Scenes. In *arXiv Preprint*, 2025. 2, 3

[13] Anna Dai, Sotiris Papatheodorou, Nils Funk, Dimos Tzoumanikas, and Stefan Leutenegger. Fast Frontier-Based Information-Driven Autonomous Exploration with an Mav. In *International Conference on Robotics and Automation*, pages 9570–9576, 2020. 2

[14] Andrew J. Davison, Ian D. Reid, Nicholas D. Molton, and Olivier Stasse. MonoSLAM: Real-Time Single Camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052–1067, 2007. 2

[15] Ziyue Feng, Huangying Zhan, Zheng Chen, Qingan Yan, Xiangyu Xu, Changjiang Cai, Bing Li, Qilun Zhu, and Yi Xu. NARUTO: Neural Active Reconstruction from Uncertain Target Observations. In *Conference on Computer Vision and Pattern Recognition*, pages 21572–21583, 2024. 2, 3, 6, 8

[16] Georgios Georgakis, Bernadette Bucher, Anton Arapin, Karl Schmeckpeper, Nikolai Matni, and Kostas Daniilidis. Uncertainty-Driven Planner for Exploration and Navigation. In *International Conference on Robotics and Automation*, pages 11295–11302, 2022. 6, 8

[17] Antoine Guédon and Vincent Lepetit. Sugar: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering. In *Conference on Computer Vision and Pattern Recognition*, pages 5354–5363, 2024. 3

[18] Antoine Guédon, Pascal Monasse, and Vincent Lepetit. SCONE: Surface Coverage Optimization In Unknown Environments by Volumetric Integration. In *Advances in Neural Information Processing Systems*, page NIPS, 2022. 2, 3, 4, 5, 6, 7, 1

[19] Antoine Guédon, Tom Monnier, Pascal Monasse, and Vincent Lepetit. MACARONS: Mapping And Coverage Anticipation with RGB Online Self-Supervision. In *Conference on Computer Vision and Pattern Recognition*, pages 940–951, 2023. 2, 3, 4, 5, 6, 7, 1

[20] Antoine Guédon, Diego Gomez, Nissim Maruani, Bingchen Gong, George Drettakis, and Maks Ovsjanikov. MILo: Mesh-In-the-Loop Gaussian Splatting for Detailed and Efficient Surface Reconstruction. In *arXiv Preprint*, pages arXiv–2506, 2025. 6, 7

[21] Guillaume Hardouin, Julien Moras, Fabio Morbidi, Julien Marzat, and El Mustapha Mouaddib. Next-Best-View Planning for Surface Reconstruction of Large-Scale 3D Environments with Multiple UAVs. In *International Conference on Intelligent Robots and Systems*, pages 1567–1574, 2020. 2

[22] Peter E Hart, Nils J Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics*, 4 (2):100–107, 1968. 2

[23] Lionel Heng, Alkis Gotovos, Andreas Krause, and Marc Pollefeys. Efficient Visual Exploration and Coverage with a Micro Aerial Vehicle in Unknown Environments. In *International Conference on Robotics and Automation*, pages 1071–1078, 2015. 2

[24] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2D Gaussian Splatting for Geometrically Accurate Radiance Fields. In *ACM SIGGRAPH*, pages 1–11, 2024. 3

[25] Stefan Isler, Reza Sabzevari, Jeffrey Delmerico, and Davide Scaramuzza. An Information Gain Formulation for Active

Volumetric 3D Reconstruction. In *International Conference on Robotics and Automation*, pages 3477–3484, 2016. 2

[26] Wen Jiang, Boshu Lei, and Kostas Daniilidis. FisherRF: Active View Selection and Mapping with Radiance Fields Using Fisher Information. In *European Conference on Computer Vision*, pages 422–440, 2024. 2, 3, 6, 7, 4, 5

[27] Liren Jin, Xingguang Zhong, Yue Pan, Jens Behley, Cyrill Stachniss, and Marija Popović. Activegs: Active Scene Reconstruction Using Gaussian Splatting. *IEEE Robotics and Automation Letters*, 2025. 2, 3

[28] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, 42(4):1–14, 2023. 3

[29] Simon Kriegel, Christian Rink, Tim Bodenmüller, Alexander Narr, Michael Suppa, and Gerd Hirzinger. Next-Best-Scan Planning for Autonomous 3D Modeling. In *International Conference on Intelligent Robots and Systems*, pages 2850–2856, 2012. 2

[30] Steven LaValle. Rapidly-exploring random trees: A new tool for path planning. *Research Report 9811*, 1998. 2

[31] Soomin Lee, Le Chen, Jiahao Wang, Alexander Liniger, Suryansh Kumar, and Fisher Yu. Uncertainty Guided Policy for Active Robotic 3D Reconstruction Using Neural Radiance Fields. *IEEE Robotics and Automation Letters*, 7(4): 12070–12077, 2022. 2

[32] Shiyao Li, Antoine Guedon, Clémentin Boittiaux, Shizhe Chen, and Vincent Lepetit. NextBestPath: Efficient 3D Mapping of Unseen Environments. In *International Conference on Learning Representations*, 2025. 2, 3, 6, 8

[33] Yuetao Li, Zijia Kuang, Ting Li, Qun Hao, Zike Yan, Guyue Zhou, and Shaohui Zhang. Activesplat: High-fidelity scene reconstruction through active gaussian splatting. *IEEE Robotics and Automation Letters*, 2025. 2, 3

[34] Miguel Mendoza, Juan Irving Vasquez-Gomez, Hind Taud, Luis Enrique Sucar, and Carolina Reta. Supervised Learning of the Next-Best-View for 3D Object Reconstruction. *Pattern Recognition Letters*, 2020. 2, 3

[35] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing Scenes As Neural Radiance Fields for View Synthesis. In *European Conference on Computer Vision*, pages 405–421. Springer, 2020. 3, 4

[36] Raul Mur-Artal and Juan D Tardos. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 31(5):1147–1163, 2015. 2

[37] Xuran Pan, Zihang Lai, Shiji Song, and Gao Huang. Activenerf: Learning Where to See with Uncertainty Estimation. In *European Conference on Computer Vision*, pages 230–246, 2022. 3

[38] Santhosh K. Ramakrishnan, Ziad Al-Halah, and Kristen Grauman. Occupancy Anticipation for Efficient Exploration and Navigation. In *European Conference on Computer Vision*, pages 400–418, 2020. 6, 8

[39] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3d deep learning with pytorch3d. *arXiv preprint arXiv:2007.08501*, 2020. 3

[40] Mike Roberts, Debadeepta Dey, Anh Truong, Sudipta Sinha, Shital Shah, Ashish Kapoor, Pat Hanrahan, and Neel Joshi. Submodular Trajectory Optimization for Aerial 3D Scanning. In *International Conference on Computer Vision*, pages 5324–5333, 2017. 2

[41] Cyrill Stachniss, Giorgio Grisetti, and Wolfram Burgard. Information Gain-Based Exploration Using Rao-Blackwellized Particle Filters. *Robotics: Science and systems*, 2(1):65–72, 2005. 2

[42] Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. Vggt: Visual geometry grounded transformer. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 5294–5306, 2025. 8

[43] Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. Dust3r: Geometric 3d vision made easy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20697–20709, 2024. 8

[44] Brian 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'*, pages 146–151, 1997. 2, 8

[45] Zike Yan, Haoxiang Yang, and Hongbin Zha. Active Neural Mapping. In *International Conference on Computer Vision*, pages 10981–10992, 2023. 2, 3, 6, 8

[46] Rui Zeng, Wang Zhao, and Yong-Jin Liu. PC-NBV: A Point Cloud Based Deep Network for Efficient Next Best View Planning. In *International Conference on Intelligent Robots and Systems*, 2020. 2, 3

[47] Baowen Zhang, Chuan Fang, Rakesh Shrestha, Yixun Liang, Xiaoxiao Long, and Ping Tan. RaDe-GS: Rasterizing Depth in Gaussian Splatting. *arXiv Preprint*, 2024. 6

[48] Boyu Zhou, Yichen Zhang, Xinyi Chen, and Shaojie Shen. Fuel: Fast uav exploration using incremental frontier structure and hierarchical planning. *IEEE Robotics and Automation Letters*, 6(2):779–786, 2021. 3