2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Active Scene Reconstruction with Topological Reasoning and Semantic-Augmented Reinforcement Learning

Yiqing Yuan[1*], Zhi Li[2*], Hao Ren[1], Kairao Zheng[1], Hui Cheng[1,†]

***Abstract*— Active scene reconstruction aims to autonomously recover the fine-grained appearance and structural details of a complex unknown scenes. Existing approaches based on 2D topological or voxel-based abstractions often scale poorly to large environments and rely heavily on handcrafted features and heuristic rules, limiting scalability and robustness. To address these challenges, using a RGB-D camera on a mobile robot, we present a graph-based planning framework by integrating skeleton-derived topology, Bird's-Eye-View (BEV)-augmented graph inference, and offline Reinforcement Learning (RL) for policy optimization. The 3D skeleton graph captures full spatial connectivity, overcoming the limitations of 2D representations. BEV-augmented graph inference enriches node embeddings with semantic context, avoiding handcrafted feature design. The offline RL approach replaces heuristic planning with data-driven decision-making, while an additional Maximum Mean Discrepancy (MMD) term mitigates distributional shift before and after feature injection, improving stability. Extensive simulation results validate the efficacy of the proposed method. Real-world experiments demonstrate the zero-shot transferability of the learned policy, highlighting its potential for scalable, fine-grained scene reconstruction.**

I. Introduction

Active mapping is a key capability for mobile robots deployed in unknown environments, enabling them to incrementally acquire spatial knowledge and construct maps that support navigation, planning, and interaction. Over the past decade, extensive research has explored active mapping based on point cloud and voxel representations [1], [2], [3], [4], achieving promising performance in exploration and occupancy mapping. However, such representations remain inherently limited in their ability to capture fine-grained appearance and structural details of complex 3D scenes. As a result, the resulting maps are often coarse, lack photorealistic fidelity, and fail to meet the demands of applications requiring high-quality reconstructions, such as augmented/virtual reality (AR/VR) [5] and digital twin system [6].

Recent advances in neural implicit scene representations, notably Neural Radiance Fields (NeRF) [7] and Gaussian Splatting (GS) [8], can encode continuous 3D scenes with photorealistic fidelity even from sparse observations, inspiring new paradigms for *active scene reconstruction*. Several recent works have explored active mapping based on NeRF or GS [9], [10], [11], demonstrating the potential of implicit representations for active reconstruction while exposing key challenges. In particular, NeRF-based approaches rely on computationally expensive volumetric rendering, making online incremental mapping prohibitively slow and impractical for real-time deployment. To address this limitation, GS has emerged as a promising alternative: its rasterization-based rendering pipeline delivers significant speedups while preserving high-fidelity scene representation.

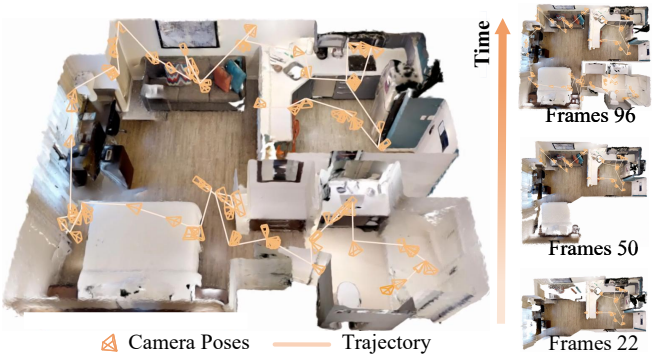


Fig. 1. Autonomous exploration for detailed scene reconstruction: Left shows the final result with trajectory and camera poses, while Right illustrates the intermediate reconstruction process.

Despite its efficiency advantages, GS-based active scene reconstruction remains challenging. Existing GS-based approaches often combine GS with voxel-based confidence modeling or hierarchical planning [10], [11], [12], which depend heavily on heuristic information-gain measures and struggle to scale in cluttered environments. Other methods that rely on information-theoretic objectives or 2D topological abstractions [13], [14] either incur high computation cost or fail to capture full 3D connectivity, leading to suboptimal planning. These limitations reveal that, while GS provides a practical rendering backbone, its effective coupling with planning and decision-making is still an open problem.

RL offers a principled solution to these challenging issues. By formulating viewpoint selection as a sequential decision-making problem, RL can adaptively balance exploration and exploitation, eliminating the need for handcrafted heuristics or expensive information-theoretic objectives. RL has already validated effective in active mapping with point cloud and voxel-based maps [15], [16], [17], demonstrating data-efficient performance. However, these methods typically rely on online RL, which requires extensive interaction with the environment—an approach that is computationally prohibitive when paired with GS-based reconstruction. To overcome these challenges, we propose a scalable graph-based framework for active scene reconstruction and employ offline RL for policy

* Equal contribution. † Corresponding Author.
[1]Yiqing Yuan, Hao Ren, Kairao Zheng, Hui Cheng are with the School of Computer Science and Engineering, Sun Yat-sen University. [2]Zhi Li is with the School of Systems Science and Engineering, Sun Yat-sen University.
This work was supported in part by the National Key R&D Program of China under Grant 2022ZD0119602. Corresponding to chengh9@mail.sysu.edu.cn

optimization. Our approach enables stable learning from limited data while avoiding costly online interaction. Fig. 1 illustrates representative results of our approach, showing the robot trajectory, camera poses, and intermediate reconstruction stages. Our key contributions are summarized as follows:

- 3D Skeleton-based graph construction. We develop a compact 3D skeleton graph that encodes global connectivity, overcoming the limitations of 2D abstractions and enabling reasoning and semantic augmentation via Graph Neural Networks (GNNs).
- BEV-augmented graph inference with offline RL. We introduce a BEV-augmented graph inference pipeline with offline RL for policy optimization, allowing semantic feature fusion into node embeddings, removing the need for handcrafted features, and replacing heuristic planning with data-driven decision-making.
- Comprehensive validation. We conduct large-scale simulations showing consistent improvements over state-of-the-art baselines, perform ablation studies to confirm the contribution of each component, and validate zero-shot transfer of the learned policy in real-world experiments, demonstrating the practicality and scalability of our approach.

II. Related Work

A. Active Scene Reconstruction

The emergence of neural implicit representations, such as NeRF [7] and GS [8], has facilitate active scene reconstruction with photorealistic fidelity from multiple viewpoints. A variety of NeRF or GS-based methods have been presented to integrate implicit representations with exploration strategies to enhance reconstruction quality and efficiency. Hybrid approaches such as NARUTO [9] jointly learn an uncertainty-aware occupancy map with a neural scene representation, enabling targeted exploration of uncertain regions. However, NeRF-based methods often suffer from inefficient map updates and catastrophic forgetting when applied to incremental mapping. To address these limitations, recently several methods leverage GS for active mapping has been proposed due to its differentiable and efficient rendering properties.

GS-Planner [10] integrates unknown voxels into the GS pipeline to identify unexplored regions and improve exploration efficiency. Li et al. [14] construct a Voronoi-based topological graph from the GS map to support path planning, but the approach is limited to 2D topology, restricting coverage and navigation efficiency. In contrast, our 3D topological graph extraction captures traversable regions with minimal redundancy, enabling more effective GNN-based scene reasoning. FisherRF [13] evaluates candidate viewpoints by computing the Fisher information of GS parameters, but this requires expensive gradient calculations for each candidate, limiting scalability. ActiveGS [12] models the confidence of each Gaussian primitive to balance scene completeness and reconstruction quality during planning. Inspired by these methods, we incorporate per-primitive opacity and BEV-derived semantic features into node attributes, yielding richer and more robust environment modeling for active scene reconstruction.

B. Reinforcement Learning for Active Mapping

RL has shown strong potential for active mapping tasks, with most prior work [15], [16], [17] focusing on online RL methods such as Proximal Policy Optimization (PPO)) [18] and Soft Actor-Critic (SAC) [19]. These approaches demonstrate that policy-driven viewpoint planning can significantly improve data efficiency. However, online training requires frequent interaction with the environment and continuous data collection—an approach that becomes computationally prohibitive when combined with the expensive rendering pipelines of NeRF and GS-based reconstruction.

Offline RL offers a promising alternative, as methods such as Batch-Constrained Deep Q-learning (BCQ) [20], Conservative Q-learning (CQL) [21], and Implicit Q-learning (IQL) [22] learn policies entirely from pre-collected datasets, thereby reducing interaction costs and improving sample efficiency. Despite these advantages, offline RL remains underexplored for active scene reconstruction, in part due to concerns over generalization and representation power when learning from limited data. In this work, we build upon IQL and introduce a feature-level regularization term into the advantage-weighted policy objective. This regularization mitigates the distributional shift induced by BEV-based cross-attention, stabilizes training, and enhances generalization, enabling robust and data-efficient offline policy optimization for active scene reconstruction.

III. Preliminaries

This section covers the Partially Observable Markov Decision Process (POMDP) formulation for autonomous scene reconstruction, as well as the GS representation used for environmental modeling.

A. POMDP Formulation for Active Scene Reconstruction

Active scene reconstruction aims to generate a fine-grained and geometrically accurate 3D map of an initially unknown environment with minimal trajectory cost. We consider the setting where the robot incrementally acquires observations using an onboard RGB-D camera. To bridge the sim-to-real gap associated with raw sensor data, we represent the environment as a topological graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ extracted from the active scene reconstruction process. In this representation, each node $v \in \mathcal{V}$ encodes essential features including geometry, semantic confidence, and occupancy attributes derived from both the voxel grid and GS maps. Edges $e \in \mathcal{E}$ are added only after identifying traversable regions via the skeletonized map, ensuring that connectivity reflects feasible paths within the accessible environment. The evolving topological graph, along with its associated BEV, jointly reflect the robot's current, partial knowledge of the scene.

Given the evolving topological graph and its associated BEV which together serve as the robot's observation of the current state, we formulate the autonomous exploration

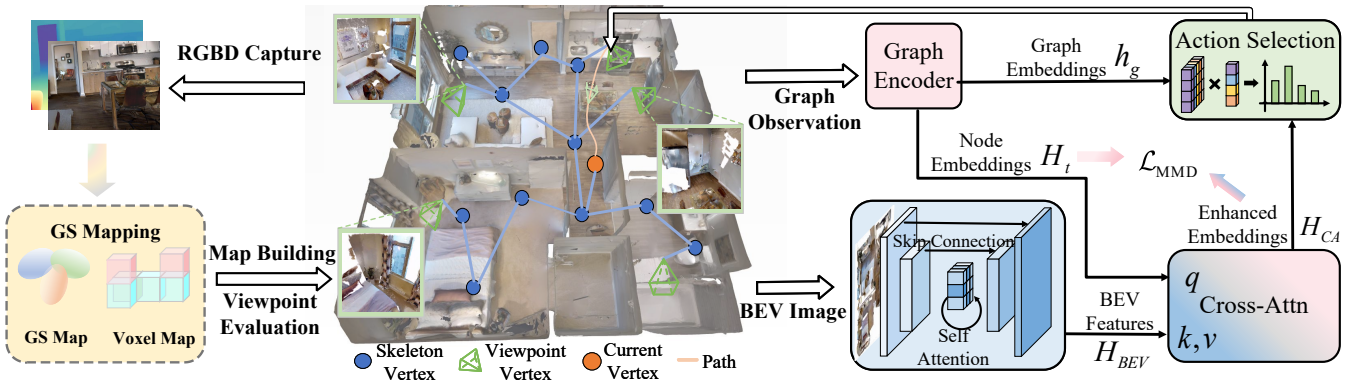


Fig. 2. Overview of the proposed BEV-augmented topological graph learning pipeline for autonomous scene reconstruction. The pipeline begins with RGBD capture, GS-based mapping, and voxel map construction, from which a skeleton-based topological graph with candidate viewpoints is extracted. The graph and BEV image are encoded by a GNN and a U-Net with self-attention, respectively, and then fused via cross-attention, which injects semantic BEV context into the node representations. A pointer network finally selects actions, leveraging both structural and semantic features for viewpoint planning.

and reconstruction process as a POMDP, defined by the tuple $\langle S, A, P, O, R, \gamma \rangle$. Here, S represents the state space capturing the environment's complete spatial information, A is the discrete action space of candidate viewpoints, O is the observation space, $P(s'|s, a)$ models the transition dynamics, and $R(s, a, s')$ is the reward function, with $\gamma \in [0, 1)$ as the discount factor.

The reward function is given by:

$$R(s, a, s') = -\lambda \cdot R_c(s, s') + R_f(s'), \tag{1}$$

where $R_c(s, s')$ measures the travel cost between viewpoints, and $R_f(s')$ is a terminal bonus when the skeletonized graph yields no additional reachable viewpoints, indicating completion of reconstruction, and the coefficient λ prioritizes completeness over minimizing path length by ensuring $\lambda \cdot R_c < \gamma^T R_f$, where T is the truncation horizon.

B. Gaussian Splatting Representation

We adopt a GS map based on the Gaussian surfel representation [23]. Each surfel is parameterized by its position, orientation, anisotropic scale, RGB color, opacity, and a learnable confidence score.

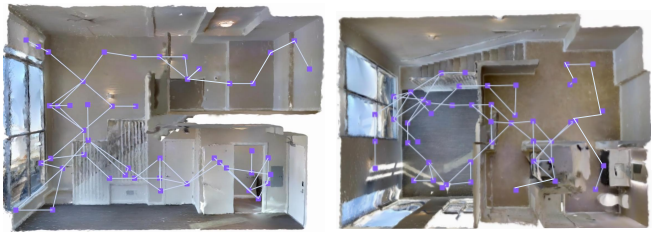
Formally, each Gaussian $g_i = (x_i, q_i, s_i, c_i, o_i, k_i)$ includes the 3D center x_i, rotation q_i, scaling s_i, color c_i, opacity o_i, and confidence k_i. This structure enables differentiable rasterization to produce not only RGB and depth maps but also confidence and opacity maps, attributes naturally supported by the Gaussian surfel formulation and explicitly leveraged in our system for guiding exploration and evaluating reconstruction quality. This expressive representation enables efficient, high-fidelity scene reconstruction for autonomous robotics.

IV. Method

We present an integrated framework for active scene reconstruction that combines skeleton-based topological graph construction, BEV-augmented feature fusion, and MMD regularization for robust viewpoint policy optimization, as illustrated in Fig. 2. The method extracts a skeleton-based topological graph from the reconstructed environment and inserts candidate viewpoints as nodes with informative features. The topological graph is first processed by an encoder to produce node embeddings that capture geometric structure and connectivity. In parallel, the semantic BEV map is encoded into a compact feature representation. These two modalities are fused via a cross-attention decoder that injects BEV features into node embeddings, integrating geometric and semantic information. Action selection and value estimation are performed via a pointer-based decoder and twin-critic architecture. The end-to-end network is trained using IQL with an additional cross-attention regularization loss to stabilize BEV feature injection.

A. Topological Graph Construction

We construct the topological graph through two primary stages: skeleton extraction and viewpoint insertion.



(a) Front View (b) Top View

Fig. 3. Skeleton-based topological graph of the environment. (a) Front view and (b) top view, where points denote skeleton nodes and lines represent their connectivity. Despite the sparsity of nodes, the graph compactly captures the entire environment structure.

1) Skeleton Extraction: We preprocess the occupancy grid $\mathbb{Z}^3 \to \{0, 1\}$, with 1 denoting obstacles and 0 free space. Let $\mathcal{O}_{\text{free}}$ and $\mathcal{O}_{\text{obs}}$ denote the sets of free and obstacle voxels, respectively. For each free voxel $p \in \mathcal{O}_{\text{free}}$, the Euclidean distance transform assigns its distance to the nearest obstacle voxel $q \in \mathcal{O}_{\text{obs}}$:

$$D(p) = \min_{q \in \mathcal{O}_{\text{obs}}} \|p - q\|_2, \quad p \in \mathcal{O}_{\text{free}}. \tag{2}$$

After preprocessing, we extract a surface skeleton via slice-wise 2D medial axis transforms. For each axis $a \in \{x, y, z\}$ and coordinate $k \in a$, we define the free-space slice $\mathcal{O}_{\text{free}}^{(a,k)}$ as the set of voxels lying on the plane orthogonal to a at coordinate k. Within each slice, we select voxels that locally maximize the distance field:

$$S_{a,k} = \Big\{ p \in \mathcal{O}_{\text{free}}^{(a,k)} \;\Big|\; D(p) \geq D(q),\ \forall q \in \mathcal{N}(p) \Big\}. \quad (3)$$

Then we aggregate selected voxels $S_{a,k}$ across slices and axes to obtain:

$$S_{\text{surf}} = \cup_{a\in\{x,y,z\}} \cup_k S_{a,k}, \quad (4)$$

where $\cup$ denotes set union. The resulting skeleton S_{surf} integrates three orthogonal views and mitigates directional artifacts.

We further refine S_{surf} into a one-voxel-wide line skeleton through sequential axis-wise thinning. For each axis $a \in \{x, y, z\}$ and coordinate $k \in a$, let $S_{\text{surf}}^{(a,k)}$ denote the slice of S_{surf} orthogonal to a. We then obtain its thinned skeleton by applying a topology-preserving skeletonization operator $\text{Skel}(\cdot)$ [24]:

$$T_{a,k} = \text{Skel}\Big(S_{\text{surf}}^{(a,k)}\Big). \quad (5)$$

Aggregating across slices and composing along axes yields:

$$\text{Thin}_a(S) = \bigcup_k T_{a,k}, \quad (6)$$

$$S_{\text{line}} = \text{Thin}_z\Big(\text{Thin}_y\big(\text{Thin}_x(S_{\text{surf}})\big)\Big). \quad (7)$$

By sequentially applying Thin_a across axes and using each stage's output as the next input, S_{surf} is progressively refined into a one-voxel-wide line skeleton. This ensures S_{line} remains topology-preserving while capturing free-space connectivity.

Finally, we construct the skeleton graph $G_{\text{skel}} = (\mathcal{V}_{\text{skel}}, \mathcal{E}_{\text{skel}})$ under the 26-neighborhood, providing a compact representation of the environment topology. As shown in Fig. 3, the resulting graph in a two-floor residential environment effectively captures the main traversable structure.

2) Viewpoint Generation and Integration: We adopt the Regions of Interest (ROI)-driven candidate viewpoint generation strategy from ActiveGS [12]. To identify the ROIs, ActiveGS detects frontier voxels and low-confidence Gaussian primitives, defining the regions as those containing any of them. This ensures that the robot prioritizes areas requiring further inspection or exploration, targeting both reconstruction completeness and quality. However, the large number of sampled viewpoints can increase the decision burden during planning. To mitigate this, we introduce a Jaccard similarity-based clustering method. For a pair of candidate viewpoints v_i and v_j, let $\mathcal{R}_i$ and $\mathcal{R}_j$ denote their visible ROI sets, respectively. The Jaccard similarity between them is:

$$J_{ij} = \frac{|\mathcal{R}_i \cap \mathcal{R}_j|}{|\mathcal{R}_i \cup \mathcal{R}_j|}. \quad (8)$$

A threshold is applied to J_{ij}, grouping viewpoints with sufficiently high similarity. Only the representative viewpoint with maximal ROI coverage is preserved. This clustering reduces the number of viewpoints considered at each decision step without compromising the final completeness of the reconstruction. In the later stages of reconstruction, as the viewpoints become sparser, no further merging is necessary. Finally, all candidates, along with the current position, are inserted as nodes into the skeleton graph, connected to their nearest skeleton points.

3) Node Feature Construction: Each node i is assigned a 10-dimensional feature vector u_i, which includes its 3D position, direction, and a binary flag indicating whether it has been visited. For candidate viewpoints, u_i additionally includes the number of visible frontier voxels and the sums of rendered confidence and opacity values (across all pixels), while for non-viewpoint nodes these entries are padded as zero. All features are normalized to $[0, 1]$ for stable training.

B. BEV Augmented Graph Inference Framework

For active scene reconstruction, we adopt an encoder-decoder GNN tailored for large-scale topological graphs. The graph branch provides structural cues, while a BEV branch supplies complementary semantic context. The two streams are fused via cross-attention.

1) Graph Encoding: Following SGFormer [25], we adopt a hybrid encoder that combines a local Graph Convolutional Network (GCN) [26] for neighborhood-level aggregation and a Graph Attention Network (GAT) [27] for capturing global dependencies. Given a topological graph $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t)$ with node features $\{u_i\}_{i\in\mathcal{V}_t}$, the local and global embeddings for each node are computed as:

$$h_i^{\text{local}} = \text{GCN}(u_i, \mathcal{N}(i)), \quad (9)$$

$$h_i^{\text{global}} = \text{GAT}(\{u_j\}_{j\in\mathcal{V}_t}), \quad (10)$$

$$\gamma_i = \sigma\big(W_\gamma [\, h_i^{\text{global}} \,\|\, h_i^{\text{local}} \,]\big), \quad (11)$$

$$h_i^{\text{fused}} = (1-\gamma_i)\, h_i^{\text{global}} + \gamma_i\, h_i^{\text{local}}, \quad (12)$$

where W_γ is a learnable weight matrix and $\sigma(\cdot)$ is the sigmoid function. The gating coefficient γ_i adaptively balances local structural information and global contextual information for each node. Stacking all fused node features yields the node embeddings:

$$H_t = [h_1^{fused}, \ldots, h_{|\mathcal{V}_t|}^{\text{fused}}]. \quad (13)$$

The graph embeddings is then obtained by attention pooling over feasible viewpoint nodes:

$$h_g = \text{AttnPool}(H_t). \quad (14)$$

2) Implicit Feature Injection Mechanism: A U-Net [28] processes the BEV input to produce BEV features H_{BEV}; self-attention at the bottleneck captures long-range context. We then apply cross-attention with queries from the node embeddings and keys/values from the BEV features, yielding cross-attention enhanced embeddings:

$$H_{\text{CA}} = \text{CrossAttn}(q{=}H_t,\ k{=}H_{\text{BEV}},\ v{=}H_{\text{BEV}}), \quad (15)$$

this injection enriches each node with BEV semantics and mitigates sparsity/ambiguity in purely geometric descriptors.

3) Action Selection: For viewpoint selection, we combine the graph embeddings and the cross-attention enhanced embeddings to score feasible actions:

$$\textit{logits} = \text{Pointer}\,(H_{\text{CA}},\, h_g,\, \textit{mask}), \quad (16)$$

where *mask* encodes action feasibility. Applying a softmax over *logits* yields the final action distribution. This design jointly exploits semantic context and global connectivity for robust viewpoint planning.

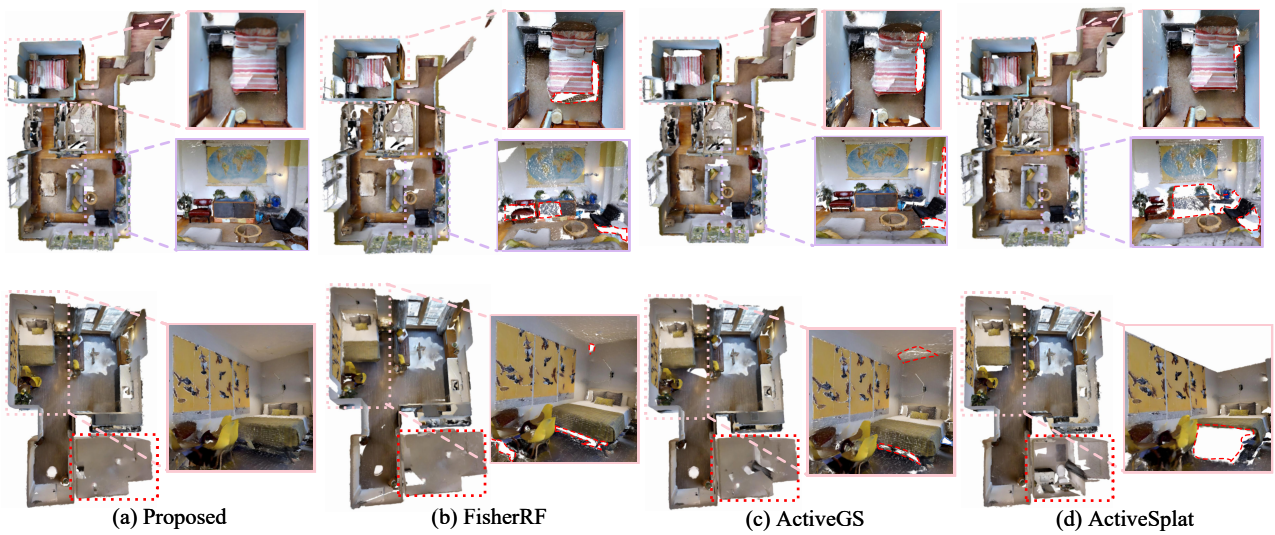


Fig. 4. Comparison of reconstructed scenes using different Gaussian-based methods. For each method, the left panel shows the reconstructed top-down view, while the right panels present zoom-in views of cluttered regions. Red regions are marked to facilitate visual comparison across different methods.

C. Offline RL for Policy Optimization

Our method extends the original IQL [22] framework with a modular actor-critic architecture tailored for offline 3D scene reconstruction. The overall system consists of a twin-critic module for value estimation and a triple-regularized actor for robust policy learning.

1) Value and Q-Value Estimation: Following the original IQL algorithm, the critic consists of two heads: a value estimator $V(s)$ and a Q-value estimator $Q(s,a)$. For each candidate viewpoint, we concatenate the corresponding features from H_t and H_{CA} as input to lightweight MLP heads: the Q-value head predicts $Q(s,a)$ for each candidate, while the value head computes $V(s)$ by pooling Q-values across all candidates. The value head is trained to approximate the maximal in-distribution value:

$$\mathcal{L}_V(\psi) = \mathbb{E}_{(s,a)\sim D}\left[\mathcal{L}_\tau\left(Q_\theta(s,a) - V_\psi(s)\right)\right], \tag{17}$$

where D is the offline dataset and $\mathcal{L}_\tau$ is the asymmetric expectile regression loss with parameter $\tau \in (0,1)$. The Q head is updated by Bellman regression with the learned value as the target:

$$\mathcal{L}_Q(\theta) = \mathbb{E}_{(s,a,s')\sim D}\left[\left(r(s,a) + \gamma V_\psi(s') - Q_\theta(s,a)\right)^2\right]. \tag{18}$$

This design ensures stable and efficient value function learning under offline RL.

2) Policy Learning with Cross-Attention Regularization: In the original IQL framework, the actor objective combines behavior cloning (BC) with advantage-weighted learning (AW) to stabilize training. Formally, the loss is defined as:

$$\begin{aligned}\mathcal{L}_{\text{actor}}(\theta) = {} & \lambda_{\text{BC}}\,\mathbb{E}_{(s,a)\sim D}\left[-\log \pi_\theta(a|s)\right] \\ & + \lambda_{\text{AW}}\,\mathbb{E}_{(s,a)\sim D}\left[-\exp\left(\tfrac{A(s,a)}{\tau}\right)\log \pi_\theta(a|s)\right],\end{aligned} \tag{19}$$

where $A(s,a) = Q(s,a) - V(s)$ denotes the action advantage term. BC leverages the dataset's behavior policy to stabilize training and prevent drift from the data distribution; advantage-weighted encourages the policy to prefer high-advantage actions via temperature-scaled exponential weighting, improving exploitation of promising behaviors while maintaining dataset coverage. However, when BEV-derived features are injected via cross-attention, the training process often becomes unstable due to distributional shifts between semantic and structural representations. To address this issue, we introduce an additional regularization term:

$$\mathcal{L}_{\text{MMD}}(\phi) = \mathbb{E}\left[\|\mathbb{E}[\phi(H_{\text{CA}})] - \mathbb{E}[\phi(H_{\text{t}})]\|^2\right], \tag{20}$$

which constrains the discrepancy between the kernel-mapped representations $\phi(H_{\text{CA}})$ of the cross-attention outputs and $\phi(H_{\text{t}})$ of the original node embeddings. This regularization enforces their alignment and ensures consistent feature fusion.

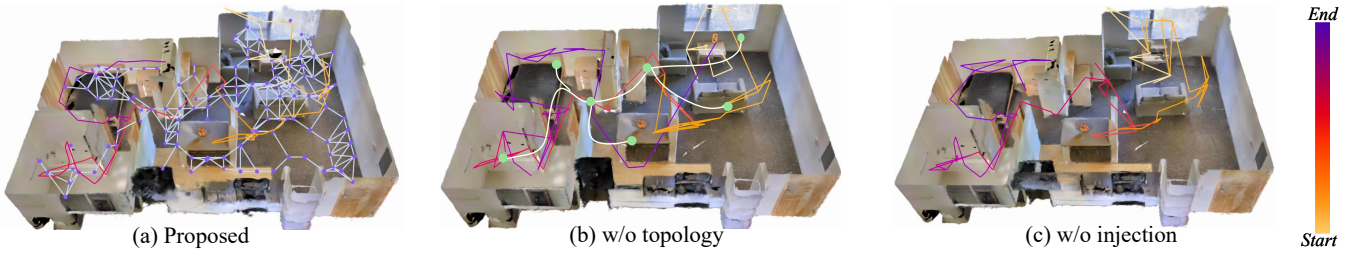
The three components are jointly optimized, with their relative contributions balanced by the coefficients λ_{BC}, λ_{AW}, and λ_{MMD}.

V. Experiments

We extensively evaluate our framework for active scene reconstruction on both simulated and real-world environments. Performance is compared with representative baselines and recent state-of-the-art methods. We further conduct a series of ablation studies, systematically analyzing the impact of BEV-augmented feature injection and the role of topological graph connectivity. These experiments clarify the individual contribution of each module and validate the overall design effectiveness.

A. Environmental Setup

Datasets: For a fair comparison, our model is trained on scenes from *Matterport3D (MP3D)* [29] and comparative experiments are conducted on scenes from *Gibson* [30] and *Replica* [31]. This setup allows us to assess the generalization ability of our method across different scene datasets and mitigates overfitting to a particular domain.



(a) Proposed (b) w/o topology (c) w/o injection

Fig. 5. Ablation study on semantic injection and topology. (a) Proposed method with a 3D skeleton-based graph, (b) variant with a 2D Voronoi graph replacing the skeleton, and (c) variant without BEV feature injection. In all cases, trajectories are shown together with the underlying graph, and the color gradient from yellow to purple indicates the progression of each trajectory.

TABLE I

QUANTITATIVE COMPARISON WITH BASELINES AND ABLATIONS ON THE REPLICA AND GIBSON DATASETS.

Method	Replica						Gibson						Tve [s] ↓
	Len. ↓	PSNR ↑	SSIM ↑	LPIPS ↓	DMAE ↓	Comp. ↑	Len. ↓	PSNR ↑	SSIM ↑	LPIPS ↓	DMAE ↓	Comp. ↑	
ActiveGS	30.3±2.8	27.87	0.965	0.074	0.129	**0.957**	75.3±6.5	26.32	0.954	0.085	0.135	0.914	0.232
ActiveSplat	**23.2±1.8**	21.25	0.897	0.183	0.153	0.872	65.6±3.2	19.84	0.872	0.193	0.148	0.851	0.257
FisherRF	33.2±2.1	26.61	0.934	0.097	0.132	0.924	72.6±6.2	25.19	0.918	0.112	0.130	0.901	0.322
NARUTO	40.7±2.7	25.32	0.923	0.094	0.167	0.902	80.1±4.8	23.63	0.912	0.104	0.154	0.880	0.336
w/o injection	25.8±1.8	28.02	0.964	0.077	0.128	0.944	66.5±3.1	27.12	0.950	0.086	0.119	0.923	0.233
w/o topology	32.2±2.9	24.31	0.922	0.089	0.134	0.937	76.8±5.9	22.98	0.908	0.099	0.129	0.912	**0.225**
Ours (full)	25.2±1.3	**28.64**	**0.973**	**0.065**	**0.125**	0.944	**62.3±2.2**	**27.89**	**0.961**	**0.074**	**0.112**	**0.936**	0.245

Model Training: We collected training data using the ActiveGS [12] approach, yielding 500 trajectories. The reward function was configured with a completion bonus of $R_f(s') = 100$ and a trade-off parameter of $\lambda = 1$. Both the policy and value networks in our RL module share a common encoder architecture based on heterogeneous graphs. Training and evaluation were performed on a server equipped with two AMD EPYC 7773X 64-Core CPUs and three NVIDIA RTX 3090 GPUs. The model was trained for 8,000 steps to ensure convergence, with the entire process taking nearly 60 hours.

Baselines: We compare our method against several representative approaches. NARUTO [9] is a state-of-the-art NeRF-based active scene reconstruction pipeline that leverages uncertainty estimation for view selection. FisherRF [13], ActiveGS [12], and ActiveSplat [14] are all GS-based baselines. FisherRF utilizes Fisher information on frontier voxels for candidate view selection; ActiveGS combines a GS map with a voxel map and confidence modeling for adaptive exploration; ActiveSplat overlays dense GS mapping with a sparse 2D Voronoi graph, enabling viewpoint selection and planning.

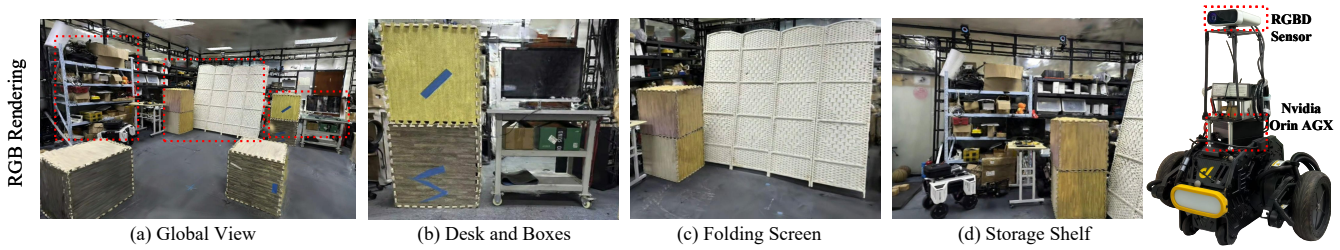
Evaluation Metrics: We assess both path efficiency and map quality using the following metrics: Path Length (Len., m), the total trajectory length required to achieve the reconstruction target; Decision time (Tve, s), the average time per decision step; RGB rendering quality, measured by Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS), all evaluated on novel RGB viewpoints; Depth accuracy, expressed as the Mean Absolute Error (DMAE) between the reconstructed and ground-truth depths; and Completeness (Comp.), the fraction of the scene surface that is successfully reconstructed. For NARUTO [9], we adopt the original definition of Completeness. For GS-based methods, it is measured following the HGS-Planner protocol [11], which extracts geometric centroids from Gaussian spheres to approximate mesh surfaces.

B. Simulation Experiments

In simulation, we comprehensively compare our framework with all baseline methods as well as two ablation variants: one without the top-down implicit feature injection, and one without topological graph connectivity. The simulated robot uses a velocity of $1, \mathrm{m/s}$, with a field of view set to $[60°, 60°]$. RGB-D observations are captured at a resolution of 512×512 pixels, with depth values ranging from 0.3 to 4 meters. Both the voxel map resolution and the movement step size are set to 0.3 meters, and each episode is limited to a maximum of 200 frames. The remaining hyperparameters follow the default settings.

Table I summarizes the quantitative results on both the Replica and Gibson datasets. For each dataset, we randomly select ten scenes. For path efficiency and coverage, Path length (Len.) and Completeness (Comp.) are averaged over all episodes within each dataset, while the decision time (Tve) is reported as the mean across all scenes over both datasets. For rendering and depth quality (PSNR, SSIM, LPIPS, and DMAE), we randomly sample five RGB-D reconstructions per scene, yielding 50 samples per dataset (10 scenes $\times$ 5 samples), and report the averages over these samples.

As shown in Table I, our method achieves best or near-best performance across most metrics. On the Gibson dataset, with larger and more complex scenes, our approach yields clear improvements in trajectory length, RGB/depth reconstruction quality, and surface completeness. On the smaller Replica dataset, some baselines remain competitive, but ours consistently attains the highest RGB quality and DMAE scores



(a) Global View (b) Desk and Boxes (c) Folding Screen (d) Storage Shelf

Fig. 6. Real-world reconstruction in a cluttered indoor scene: (a) shows the global view; (b), (c), and (d) present zoom-in details of regions with dense object arrangements. Experiments were conducted on a wheeled-leg robot equipped with an RGB-D camera.

while producing shorter trajectories than most alternatives. Although not the lowest in Tve, our average latency of 0.245 s per step remains below most baselines and sufficient for real-time use, given the accuracy and robustness gains. Compared with prior works, ActiveGS struggles in cluttered layouts due to viewpoint-confidence modelling, ActiveSplat's 2D Voronoi abstraction limits 3D connectivity, FisherRF incurs heavy computation in large scenes, and NARUTO suffers from slow training and scalability. Qualitative results in Fig. 4 confirm these trends: our reconstructions preserve structural boundaries, better cover cluttered regions, and show fewer artifacts, while baselines often fail in such cases.

TABLE II

COMPARISON OF DIFFERENT METHODS WITH POST-PROCESSING.

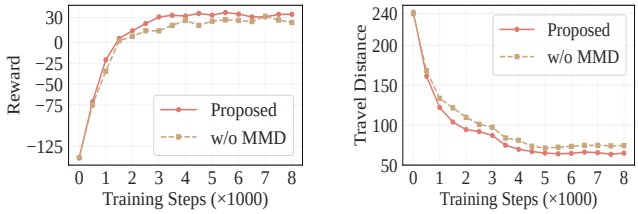
Method	PSNR ↑	SSIM ↑	LPIPS ↓	DMAE ↓	Comp. ↑
ActiveGS	30.66	0.936	0.076	0.133	0.911
ActiveSplat	23.38	0.870	0.188	0.158	0.847
FisherRF	29.27	0.905	0.100	0.136	0.898
w/o injection	30.40	0.912	0.070	0.131	0.912
w/o topology	28.12	0.883	0.082	0.142	0.901
Ours (full)	**31.50**	**0.964**	**0.067**	**0.129**	**0.932**

Table II summarizes the results of methods after applying Gaussian-based post-processing, where metrics are averaged across all scenes. It can be observed that our method achieves the best performance on all RGB-D evaluation metrics. It is also worth noting that the overall completeness (Comp.) metric shows a slight decrease for all methods. This is due to the post-processing step, which prunes a portion of the Gaussian points to improve rendering-oriented losses, but reduces the completeness of the reconstructed map.

C. Ablation Studies

We first evaluate two ablated variants to clarify the role of semantic injection and topological connectivity, as illustrated in Table I and Fig. 5. The variant without BEV injection leaves skeleton nodes with only geometric and frontier-based attributes, resulting in degraded rendering fidelity (lower PSNR/SSIM, higher LPIPS/DMAE) and inefficient exploration, as the policy is biased toward locally high-value viewpoints. The variant replacing our 3D skeleton graph with a 2D Voronoi abstraction achieves the lowest decision time due to reduced per-step computation, but at the expense of longer trajectories, lower completeness, and poorer reconstruction, since 2D connectivity fails to capture cross-room and vertical traversability. These results confirm that semantic injection and 3D skeleton-based topology are essential for consistent planning and high-fidelity reconstruction.

To further validate the effectiveness of our proposed policy optimization, we perform an additional ablation targeting the cross-attention regularization. We evaluate performance by testing saved checkpoints every 500 training steps. As shown in Fig. 7, removing MMD leads to unstable convergence, with lower rewards and longer travel distances, as BEV cross-attention introduces distributional shifts that destabilize policy learning. In contrast, the proposed regularization consistently improves stability and efficiency, demonstrating its key role in enabling robust policy optimization when integrating BEV-derived semantic features.



(a) Reward curve (b) Distance curve

Fig. 7. Ablation on cross-attention regularization.

D. Real-World Experiments

To validate our approach in real-world settings, we deployed it on a wheeled-leg robot platform (operating in wheeled mode in this study) equipped with an NVIDIA Jetson Orin AGX for onboard computation and an Azure Kinect DK RGB-D sensor for onboard perception. Accurate localization is provided by a motion capture system. The 3D voxel map is generated by converting OctoMap [32] representations. Notably, the policy is trained entirely in simulation and transferred to the real robot in a zero-shot manner, operating without any additional fine-tuning or adaptation.

Compared to simulation, the real-world setup introduced additional constraints. Viewpoint sampling is restricted in both height and pitch angle, and the robot's maximum velocity and acceleration are limited to 0.5 m/s and 0.2 m/s^2, respectively, to ensure safe navigation. As shown in Fig. 6, our framework successfully reconstructs cluttered indoor scenes with high fidelity, despite onboard computational limitations and real-world noise. The results validate the practicality and zero-shot transferability of our approach in real environments.

VI. CONCLUSION

In this paper, we propose a graph-based framework for active scene reconstruction that integrates 3D skeleton-based

topology, BEV-augmented graph inference, and offline RL for policy optimization. The skeleton graph captures full 3D connectivity, enabling effective reasoning and semantic augmentation, while BEV-augmented inference combined with offline RL results in a more robust viewpoint selection policy. Extensive simulation results and ablation studies confirm the effectiveness of our approach, and real-world experiments demonstrate its zero-shot transferability without performance degradation. Despite these promising results, our current framework assumes access to accurate localization, which limits tis applicability in real-world deployments. Future work will focus on relaxing this assumption by exploring GS-SLAM formulations that jointly optimize poses and reconstruction, and advancing GNN-based reasoning to further enhance active scene reconstruction in complex, large-scale environments.

REFERENCES

[1] C. Cao, H. Zhu, H. Choset, and J. Zhang, "Tare: A hierarchical framework for efficiently exploring complex 3d environments." in *Robotics: Science and Systems*, vol. 5, 2021, p. 2.

[2] J. Huang, B. Zhou, Z. Fan, Y. Zhu, Y. Jie, L. Li, and H. Cheng, "Fael: Fast autonomous exploration for large-scale environments with a mobile robot," *IEEE Robotics and Automation Letters*, vol. 8, no. 3, pp. 1667–1674, 2023.

[3] S. Geng, Z. Ning, F. Zhang, and B. Zhou, "Epic: A lightweight lidar-based aav exploration framework for large-scale scenarios," *IEEE Robotics and Automation Letters*, vol. 10, no. 5, pp. 5090–5097, 2025.

[4] X. Chen, T. Wang, Q. Li, T. Huang, J. Pang, and T. Xue, "Gleam: Learning generalizable exploration policy for active mapping in complex 3d indoor scenes," *arXiv preprint arXiv:2505.20294*, 2025.

[5] Y. Jiang, C. Yu, T. Xie, X. Li, Y. Feng, H. Wang, and et al., "Vr-gs: A physical dynamics-aware interactive gaussian splatting system in virtual reality," *arXiv preprint arXiv:2401.16663*, 2024.

[6] J. Guo, Y. Xin, G. Liu, K. Xu, L. Liu, and R. Hu, "Articulatedgs: Self-supervised digital twin modeling of articulated objects using 3d gaussian splatting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2025, pp. 27 144–27 153.

[7] 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. [Online]. Available: https://doi.org/10.1145/3503250

[8] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, pp. 1–14, 2023. [Online]. Available: https://doi.org/10.1145/3592433

[9] 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 (CVPR)*, June 2024, pp. 21 572–21 583.

[10] 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 *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024, pp. 11 202–11 209.

[11] 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," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2025, pp. 14 161–14 167.

[12] L. Jin, X. Zhong, Y. Pan, J. Behley, C. Stachniss, and M. Popović, "Activegs: Active scene reconstruction using gaussian splatting," *IEEE Robotics and Automation Letters*, vol. 10, no. 5, pp. 4866–4873, 2025.

[13] W. Jiang, B. Lei, and K. Daniilidis, "FisherRF: Active View Selection and Mapping with Radiance Fields Using Fisher Information," in *Computer Vision – ECCV 2024*, ser. Lecture Notes in Computer Science, vol. 15071. Cham: Springer, 2025, pp. 415–432. [Online]. Available: https://doi.org/10.1007/978-3-031-72624-8_24

[14] Y. Li, Z. Kuang, T. Li, Q. Hao, Z. Yan, G. Zhou, and S. Zhang, "Activesplat: High-fidelity scene reconstruction through active gaussian splatting," *IEEE Robotics and Automation Letters*, vol. 10, no. 8, pp. 8099–8106, 2025.

[15] Y. Cao, R. Zhao, Y. Wang, B. Xiang, and G. Sartoretti, "Deep reinforcement learning-based large-scale robot exploration," *IEEE Robotics and Automation Letters*, vol. 9, no. 5, pp. 4631–4638, 2024.

[16] Y. Xu, J. Yu, J. Tang, J. Qiu, J. Wang, Y. Shen, Y. Wang, and H. Yang, "Explore-bench: Data sets, metrics and evaluations for frontier-based and deep-reinforcement-learning-based autonomous exploration," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 6225–6231.

[17] J. Liang, Y. Cao, Y. Ma, H. Zhao, and G. Sartoretti, "Hdplanner: Advancing autonomous deployments in unknown environments through hierarchical decision networks," *IEEE Robotics and Automation Letters*, vol. 10, no. 1, pp. 256–263, 2025.

[18] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," in *Proc. 34th Int. Conf. Machine Learning (ICML)*, 2017, pp. 1–12, arXiv:1707.06347 [cs.LG].

[19] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, and S. Levine, "Soft actor-critic algorithms and applications," in *Proc. 36th Int. Conf. Machine Learning (ICML)*, 2019, pp. 1–15, arXiv:1812.05905 [cs.LG].

[20] S. Fujimoto, D. Meger, and D. Precup, "Off-Policy Deep Reinforcement Learning without Exploration," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, ser. Proceedings of Machine Learning Research, vol. 97. PMLR, 2019, pp. 2052–2062. [Online]. Available: https://proceedings.mlr.press/v97/fujimoto19a.html

[21] A. Kumar, A. Zhou, G. Tucker, and S. Levine, "Conservative q-learning for offline reinforcement learning," *Advances in neural information processing systems*, vol. 33, pp. 1179–1191, 2020.

[22] I. Kostrikov, A. Nair, and S. Levine, "Offline reinforcement learning with implicit q-learning," in *International Conference on Learning Representations*, 2022. [Online]. Available: https://openreview.net/forum?id=68n2s9ZJWF8

[23] P. Dai, J. Xu, W. Xie, X. Liu, H. Wang, and W. Xu, "High-Quality Surface Reconstruction Using Gaussian Surfels," in *Proceedings of the ACM SIGGRAPH Conference Papers*. Denver, CO, USA: Association for Computing Machinery, 2024. [Online]. Available: https://doi.org/10.1145/3641519.3657441

[24] T. Noël, A. Lehuger, E. Marchand, and F. Chaumette, "Skeleton disk-graph roadmap: A sparse deterministic roadmap for safe 2d navigation and exploration," *IEEE Robotics and Automation Letters*, vol. 9, no. 1, pp. 555–562, 2024.

[25] Q. Wu, W. Zhao, C. Yang, H. Zhang, F. Nie, H. Jiang, Y. Bian, and J. Yan, "Sgformer: Simplifying and empowering transformers for large-graph representations," *Advances in Neural Information Processing Systems*, vol. 36, pp. 64 753–64 773, 2023.

[26] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," *arXiv preprint arXiv:1609.02907*, 2017. [Online]. Available: https://arxiv.org/abs/1609.02907

[27] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," *arXiv preprint arXiv:1710.10903*, 2018. [Online]. Available: https://arxiv.org/abs/1710.10903

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, ser. Lecture Notes in Computer Science, vol. 9351. Cham: Springer, 2015, pp. 234–241.

[29] 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.

[30] F. Xia, A. R. Zamir, Z. He, A. Sax, J. Malik, and S. Savarese, "Gibson env: Real-world perception for embodied agents," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[31] J. Straub, T. Whelan, L. Ma, Y. Chen, E. Wijmans, S. Green *et al.*, "The Replica Dataset: A Digital Replica of Indoor Spaces," *arXiv preprint arXiv:1906.05797*, 2019. [Online]. Available: https://arxiv.org/abs/1906.05797

[32] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "Octomap: An efficient probabilistic 3d mapping framework based on octrees," *Autonomous Robots*, vol. 34, no. 3, pp. 189–206, 2013.