ActiveStructure: Plane Scene Graph–Guided Active 3D Gaussian Splatting

Yingzhao Li[1,2], Yan Li[3], Yanjie Liu[1], Hanyu Zhou[3], Lijun Zhao[1,2]⋆, and Gim Hee Lee[3]

[1] Harbin Institute of Technology, Harbin, China
[2] Yangtze River Delta HIT Robot Technology Research Institute, Wuhu, China
[3] National University of Singapore, Singapore, Singapore

Abstract. Active 3D Gaussian Splatting methods select views using factorised per-pixel metrics, which cannot represent relationships between surfaces. Incompleteness that arises from unverified geometric relationships therefore goes undetected. We present **ActiveStructure**, which lifts information gain from pixels to scene-level relational topology through three innovations: (1) a *Plane Scene Graph* whose topological health, captured through node connectivity, edge completeness, and angular consistency, quantifies a form of structural completeness that no pixel-level criterion can reach; (2) a *gain decomposition* that splits the value of a viewpoint into a rendering-fidelity term and a structural-integrity term that can be optimised independently; (3) *informational polarity reversal*, which lets scaffold planes switch from exploration attractors to suppressors and provides state-driven phase control tied to structural objectives. Experiments on Replica and Matterport3D demonstrate state-of-the-art reconstruction accuracy and rendering quality.

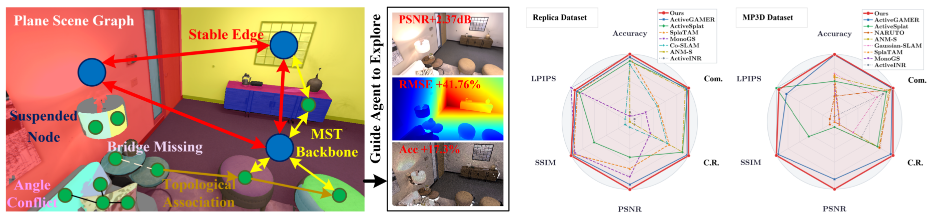
Keywords: 3D Gaussian Splatting · Active Reconstruction · Next-Best-View

1 Introduction

3D Gaussian Splatting (3DGS) [14] enables real-time photorealistic rendering through explicit anisotropic Gaussian primitives. Extensions to geometrically faithful surfaces [3, 7, 8, 24] and online dense SLAM [13, 20] have broadened its applicability to embodied perception. However, these methods share a passive assumption, namely that the input images are collected in advance and the algorithm has no say in which viewpoints are observed. In embodied scenarios, reconstruction quality is inseparable from the observation strategy.

Recent active 3DGS methods [4, 10, 11, 19, 30, 32, 33] couple next-best-view (NBV) selection with incremental Gaussian mapping through pixel- or Gaussian-level information gain, rendering-based coverage [4], hybrid voxel-Gaussian utility [19], Fisher information [10], P-Optimality [30], and Shannon mutual information [32]. Despite this diversity, all rely on a *factorised observation model* in

⋆ Corresponding author: Lijun Zhao (zhaolj@hit.edu.cn)



(a) PSG utilizes graph structures to identify unstable mapping regions. (b) Performance comparison on mainstream benchmarks.

Fig. 1: We propose **ActiveStructure**, which leverages PSG to guide adaptive progressive exploration in unknown environments, incrementally reconstructing 3D Gaussian maps while jointly optimizing reconstruction accuracy and rendering quality.

which each pixel or primitive is evaluated independently, which leaves them unable to capture relationships between surfaces (e.g., Fig. 2). The orthogonality of adjacent walls, the parallelism between floor and ceiling, and the coplanarity of fragmented observations are defined over tuples of primitives, lying outside the representational capacity of any per-element metric [4, 32].

Planar constraints have been exploited as geometric regularisers in reconstruction [3, 8, 17, 31] and neural SLAM [12, 26, 29], and spatial scene graphs [6, 9] organise primitives into hierarchical structures for downstream reasoning. However, planes universally serve as *passive consumers* of observations that improve reconstruction post-hoc, and scene graphs are built from completed reconstructions for semantic querying; neither assigns geometric structures a generative role in producing planning signals.

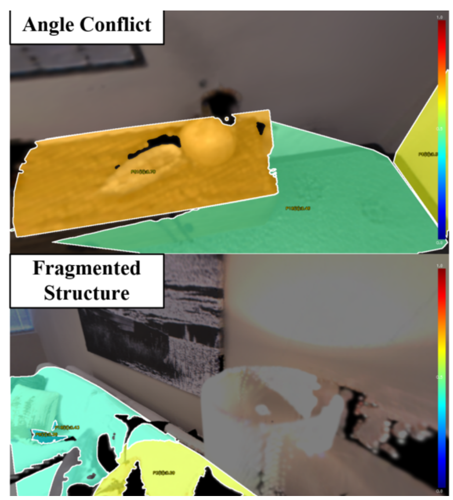


Fig. 2: Although these viewpoints offer sufficient pixel coverage, traditional NBV strategies consider them reconstructed, yet structural conflicts remain severe and thus need higher weight.

We present **ActiveStructure**, which closes this gap by lifting information gain from pixels to scene-level relational topology (Fig. 1). We construct a *Plane Scene Graph* (PSG), a typed, weighted graph over detected planar surfaces, whose topological health directly quantifies structural completeness: dangling nodes signal under-anchored surfaces, missing edges indicate unverified inter-surface constraints, and angular conflicts reveal geometric inconsistencies. We split the value of a viewpoint into two terms that can be acted on independently, a pixel-level appearance gain and a structural gain derived from the PSG. To balance scaffold construction against detail refinement, we introduce *informational polarity* reversal. During initial mapping, large structural planes act as exploration attractors and draw the agent toward under-constrained walls and ceilings. Once the scaffold is in place, the same planes act as suppressors and redirect the budget toward smaller object surfaces.

Our contributions are:

- The **Plane Scene Graph**, whose topological invariants (node connectivity, edge completeness, angular consistency) provide a structural information-gain signal complementary to and independently actionable from pixel-level assessment.
- A **per-plane quality metric** grounded in failure-mode analysis with 3D deficiency localisation, coupled with three graph-structural anomaly indicators whose multiplicative fusion translates topological violations into viewpoint-level planning signals.
- A **hierarchical exploration strategy with informational polarity reversal** that drives state-based transitions between scaffold construction and detail refinement, with each phase ending once its structural objective is met.

Experiments on Replica [28] and Matterport3D [2] demonstrate that ActiveStructure consistently outperforms state-of-the-art methods in both geometric accuracy and rendering quality.

2 Related Work

Gaussian and neural radiance fields. NeRF [21] and variants [1, 22] achieve photorealistic synthesis via implicit volumetric optimisation. 3DGS [14] replaces the implicit volume with explicit anisotropic Gaussians for real-time rasterisation, with extensions improving fidelity through planar regularisation [3, 8], uncertainty pruning [7], and sparse-view robustness [24]. Online Gaussian SLAM [13, 20] demonstrates incremental mapping from sequential RGB-D input. These advances refine representation quality given fixed views but do not address view selection.

Active reconstruction with radiance fields. Uncertainty-guided NBV methods in the NeRF era [16, 23, 25] select views via pixel- or volume-level uncertainty. With the shift to Gaussians, ActiveGAMER [4] uses rendering-based pixel information gain; ActiveSplat [19] combines hybrid voxel–Gaussian maps with hierarchical planning; GS-Planner [11] and HGS-Planner [33] formulate Gaussian-level utility; FisherRF [10] quantifies per-view Fisher information; POp-GS [30] generalises to P-Optimality; GauSS-MI [32] derives Shannon mutual information. All of these methods evaluate quality through a factorised observation model, so none of them can detect inter-surface relational deficits. ActiveStructure introduces a complementary relational assessment layer providing structural information gain independently from the pixel-level channel. A closely related graph-based selector, ActivePolicy [18], builds a spectral graph over candidate viewpoints and reads a next best view from its Fiedler centrality, and it further suppresses floaters with a depth-manifold regulariser. ActiveStructure instead builds a persistent graph over scene surfaces whose inter-plane topology forms a structural channel, which view-level selection provably cannot reach.

Planar priors in reconstruction and SLAM. PGSR [3] regularises Gaussians toward planar configurations; 2DGS [8] uses planar disks for faithful rendering; PlanarRecon [31] incrementally regresses 3D planes from posed RGB. Neural SLAM methods [12,26,29] integrate implicit planar priors for low-texture geometry. In all cases, planes are *passive constraints* improving reconstruction post-hoc but never generating planning signals. Spatial scene graphs [6,9] organise primitives for task planning from completed reconstructions. Our PSG differs in two important ways. It is *self-supervising*, since node classification follows from the graph topology alone, and it is *action-oriented*, since the NBV planner continuously uses its topological anomalies within a closed exploration-and-assessment loop.

3 Method

As shown in Fig. 3, an embodied RGB-D agent explores an unknown indoor environment to produce a high-fidelity 3DGS reconstruction. At timestep t it acquires an RGB-D pair $(\mathbf{I}_t, \mathbf{D}_t)$ at pose $\mathbf{T}_t \in \mathrm{SE}(3)$ and maintains three complementary representations: a *3DGS radiance map* $\mathcal{M}$ for photometric appearance (§3.1), a *Plane Scene Graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ for inter-surface topology (§3.2), and a *signed distance field* (SDF) for traversability (§3.4). $\mathcal{M}$ supplies per-pixel rendering quality; $\mathcal{G}$ lifts assessment to the relational level, exposing structurally fragile surfaces invisible to any 2D test; the SDF translates selected viewpoints into obstacle-aware trajectories.

Decoupling rendering from completeness. Unlike spectral selection over a viewpoint graph [18], our structural term is computed over a graph of scene surfaces, so it scores relational completeness rather than the centrality of candidate poses. Rather than conflating both into a single pixel-level score, we decompose candidate viewpoint information gain into two largely complementary terms:

$$\mathrm{IG}(\mathbf{T}_c; \mathcal{M}, \mathcal{G}) \;=\; \underbrace{\mathrm{IG}_{\mathrm{px}}(\mathbf{T}_c; \mathcal{M})}_{\text{2D appearance}} + \omega^{(\psi)} \underbrace{\mathrm{IG}_{\mathrm{str}}(\mathbf{T}_c; \mathcal{G})}_{\text{3D structure}}, \tag{1}$$

where $\omega^{(\psi)}$ is a phase-dependent weight. The pixel term mainly drives discovery of unmapped regions, while the structural term consolidates relationally weak surfaces, including ones that already render well. The two terms dominate at different stages, with the pixel gain leading during early exploration and the structural gain becoming the main way to separate candidates once coverage saturates.

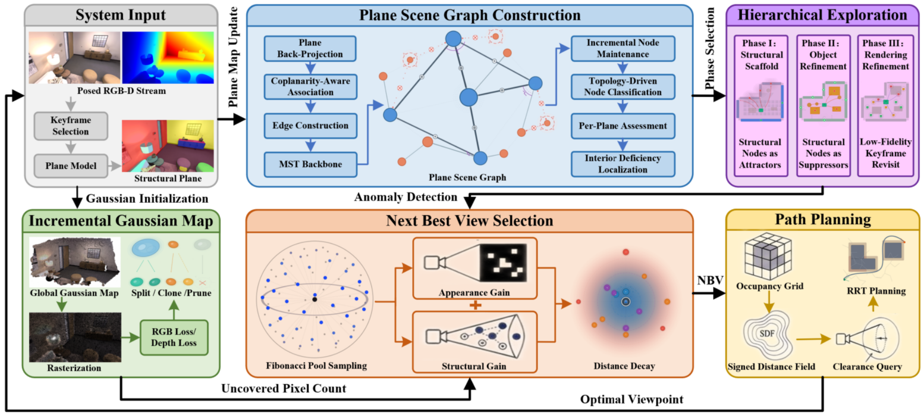


Fig. 3: ActiveStructure processes posed RGB-D frames, selecting keyframes to detect planar structures. RGB-D images incrementally project Gaussian primitives into the global map, with loss-triggered splitting/cloning/pruning; detected 2D planes are projected into 3D to update the Plane Scene Graph (PSG) nodes/edges, using attractors/repulsors per exploration phase. Exploration relies on NBV selection: candidate views are sampled globally; photometric gain (via rasterization) and PSG-derived geometric gain are weighted-summed with distance decay to determine the NBV. Path planning over the SDF map then yields the next view via interpolation. The system iterates until exploration ends.

3.1 Incremental 3D Gaussian Map

The scene is encoded as anisotropic 3D Gaussians $\{\mathcal{G}_k\}_{k=1}^{K}$, each with centre $\boldsymbol{\mu}_k$, covariance $\boldsymbol{\Sigma}_k$, opacity α_k, and spherical-harmonic colour coefficients. Differentiable splatting [14] renders colour and depth via alpha compositing:

$$\hat{X}(\mathbf{p}) = \sum_{k} x_k\, \alpha_k\, \sigma_k(\mathbf{p}) \prod_{j<k} \big(1 - \alpha_j \sigma_j(\mathbf{p})\big), \quad X \in \{\mathbf{I}, \mathbf{D}\}, \tag{2}$$

where x_k denotes colour $\mathbf{c}_k$ or depth d_k, and $\sigma_k(\mathbf{p})$ is the projected 2D Gaussian response. The map is maintained via an online *tracking–densification–mapping* loop: incoming frames refine camera pose; keyframes spawn new Gaussians at uncovered back-projections; parameters are jointly optimised using

$$\mathcal{L} = \lambda_{\mathrm{c}} \|\hat{\mathbf{I}} - \mathbf{I}\|_1 + \lambda_{\mathrm{d}} \|\hat{\mathbf{D}} - \mathbf{D}\|_1\,, \tag{3}$$

ensuring that $\mathcal{M}$, and therefore $\mathrm{IG}_{\mathrm{px}}$, always reflects the latest state.

3.2 Plane Scene Graph as Structural Oracle

The radiance map says nothing about relationships between surfaces. Two adjacent planes can each render faithfully while the constraint between them stays unverified. Closing this gap requires a representation in which 3D structural integrity is itself a measurable quantity that produces information-gain signals directly. The PSG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ addresses this through what we call the structural oracle principle, where its topological health, expressed by node connectivity, edge completeness, and angular consistency, serves as a strong practical proxy for structural completeness in plane-dominated indoor scenes. When any of these invariants is violated, it reveals a form of incompleteness that the radiance map alone is unlikely to detect, which makes the PSG a complementary information channel.

Incremental node maintenance. At periodic keyframes, RGB-D plane segments are lifted to 3D and fitted with plane equations $\boldsymbol{\pi}_i = (\mathbf{n}_i, d_i)$ by SVD. New segments are matched to existing nodes through a *coplanarity-aware dual-mode* criterion. Coplanar candidates, which must satisfy strict normal alignment and point-to-plane proximity, are allowed a relaxed centroid threshold so that distant fragments of a single surface are correctly unified, whereas non-coplanar candidates are gated by a stricter centroid distance to avoid merging parallel but disjoint surfaces. A depth discontinuity test further rejects candidates with significant depth offsets, preventing fusion of geometrically similar but physically separated surfaces. Matched segments update nodes via exponential moving-average fusion; unmatched segments initialise new nodes. Each node $v_i \in \mathcal{V}$ maintains the plane equation $\boldsymbol{\pi}_i$, point set $\mathcal{P}_i$, area a_i, observation count o_i, centroid $\bar{\mathbf{p}}_i$, and normal history.

Edge construction and topology-driven classification. Edges encode geometric affinity via a convex combination of normal alignment and spatial proximity:

$$w_{ij} = 1 - \left[\beta\,(\mathbf{n}_i^\top \mathbf{n}_j)^2 + (1-\beta)\exp(-\lambda\|\bar{\mathbf{p}}_i - \bar{\mathbf{p}}_j\|)\right], \tag{4}$$

where small weights indicate reliable relationships. The *minimum spanning tree* $\mathcal{T} = (\mathcal{V}, \mathcal{E}_\mathcal{T})$ concentrates the strongest pairwise relationships, naturally surfacing the scene's structural backbone. We recover the natural hierarchy that separates large structural surfaces anchoring the layout from smaller object surfaces using graph topology alone, computing the betweenness centrality c_i of each node in $\mathcal{T}$ and classifying v_i as *structural* ($v_i \in \mathcal{V}_\mathrm{S}$) when:

$$c_i \geq \tau_c \;\vee\; \deg_\mathcal{T}(i) \geq \tau_d \;\vee\; \big(a_i \geq \tau_a \wedge |\mathcal{P}_i| \geq \tau_p\big), \tag{5}$$

with $\mathcal{V}_\mathrm{O} = \mathcal{V} \setminus \mathcal{V}_\mathrm{S}$. The three criteria capture complementary aspects: centrality identifies relationship-mediating surfaces; MST degree identifies relational hubs; area–density conjunction captures physically dominant surfaces not yet topologically connected. This self-supervising classification adapts as the graph evolves.

3.3 Structural Quality and Anomaly Assessment

Each PSG node is evaluated in two ways. An intrinsic quality estimate detects per-plane reconstruction deficiency, while a relational assessment detects incompleteness from violations of the topological invariants that a complete reconstruction should satisfy. Combining the two yields a structural information deficit that is large only when a plane is both under-reconstructed and poorly anchored, signalling incompleteness that pixel-level assessment cannot reach.

Failure-mode quality vector. An incrementally reconstructed plane can fail via eight independently detectable modes: (i) spatial under-coverage, (ii) density insufficiency, (iii) fitting instability, (iv) observation scarcity, (v) boundary incompleteness, (vi) normal inconsistency, (vii) temporal instability, and (viii) interior deficiency, yielding $\mathbf{s}_i \in [0,1]^8$ spanning the observable failure space without redundancy. Mode (vi) accumulates a sliding window of per-observation normals, measuring alignment with the running average to distinguish transient fitting accuracy from genuine convergence. Mode (viii) produces explicit 3D planning targets by projecting points onto the tangent frame via SVD,

$$\mathcal{P}_i - \bar{\mathbf{p}}_i = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top, \quad \mathbf{p}_j^{(2\mathrm{D})} = (\mathbf{p}_j - \bar{\mathbf{p}}_i)^\top \mathbf{V}_{:,1:2}\,, \tag{6}$$

discretised into a grid with morphological closing to obtain interior mask $\mathbf{I}_i$. The deficiency ratio $g_i = |\mathbf{I}_i \setminus \mathbf{O}_i|/(|\mathbf{I}_i| + \epsilon)$, where $\mathbf{O}_i$ denotes occupied cells, measures interior completeness; unoccupied cells are inverted to 3D as deficiency centres $\mathcal{C}_i^{\mathrm{def}} \subset \mathbb{R}^3$, transforming detected gaps into active viewpoint attractors for the NBV module (§3.4). The scalar quality aggregates with type-dependent weights:

$$q_i = \mathbf{w}_{\tau(i)}^\top \mathbf{s}_i\,, \quad \tau(i) = \begin{cases} \mathrm{S} & v_i \in \mathcal{V}_\mathrm{S} \\ \mathrm{O} & v_i \in \mathcal{V}_\mathrm{O} \end{cases}, \tag{7}$$

where $\mathbf{w}_\mathrm{S}$ emphasises coverage, interior completeness, and normal consistency (modes i, viii, vi), while $\mathbf{w}_\mathrm{O}$ emphasises density, boundary sharpness, and observation count (modes ii, v, iv), reflecting the distinct roles of structural versus object surfaces.

Graph-structural anomaly indicators. A plane may achieve high intrinsic quality while remaining poorly anchored relationally. We formalise three topological deviations as *structural surprise* indicators, each detecting a distinct, independently addressable mode. *Connectivity surprise* $\phi_i^\mathrm{c} = h(\deg_\mathcal{T}(i))$, monotonically decreasing in MST degree with saturation floor $\phi_{\min}$, penalises dangling nodes that have been observed but not contextualised. *Neighbourhood surprise* detects missing bridges among spatially proximate but unlinked nodes. Let $\mathcal{N}_r(i) = \{v_j : \|\bar{\mathbf{p}}_i - \bar{\mathbf{p}}_j\| < r\}$:

$$\phi_i^\mathrm{r} = \phi_{\min} + (1-\phi_{\min})\Big(1 - \frac{|\{v_j \in \mathcal{N}_r(i) : e_{ij} \in \mathcal{E}\}|}{|\mathcal{N}_r(i)| + \epsilon}\Big). \tag{8}$$

Geometric surprise exploits the observation that indoor inter-plane angles cluster near canonical set $\mathcal{C} = \{0°, 90°, 180°\}$; persistent deviations signal uncertain normals or incomplete observation:

$$\phi_i^\mathrm{g} = \phi_{\min} + \frac{1-\phi_{\min}}{|\mathcal{N}_\mathcal{T}(i)|} \sum_{j \in \mathcal{N}_\mathcal{T}(i)} \frac{\min_{c \in \mathcal{C}} |\theta_{ij} - c|}{\theta_{\max}}\,, \tag{9}$$

where $\theta_{ij} = \arccos |\mathbf{n}_i^\top \mathbf{n}_j|$.

Structural information deficit. Multiplicative fusion implements a soft conjunction: $\Phi_i = \phi_i^{\mathrm{c}} \cdot \phi_i^{\mathrm{r}} \cdot \phi_i^{\mathrm{g}}$ is large only when all three invariants are simultaneously violated, screening out benign deviations. The *structural information deficit* combines intrinsic quality with relational surprise:

$$\gamma_i = (1 - q_i) \cdot \Phi_i \,, \tag{10}$$

which is large only when plane v_i is both under-reconstructed *and* poorly integrated. For viewpoint-level aggregation, anomaly scores over visible planes are blended:

$$\bar{\Phi}_c = \eta \, \mathrm{mean}_{v_i \in \mathcal{V}_c^{\mathrm{vis}}} \Phi_i + (1-\eta) \max_{v_i \in \mathcal{V}_c^{\mathrm{vis}}} \Phi_i \,, \tag{11}$$

balancing region-wide weakness against isolated critical fragility.

3.4 Structure-Aware Exploration and Planning

We now operationalise Eq. (1) with computable proxies for both terms and couple the resulting NBV selector with a hierarchical phase strategy that governs the *informational role* of each surface type.

Candidate generation and appearance term. Free voxels combined with K uniformly sampled directions (Fibonacci lattice on $\mathbb{S}^2$) populate a candidate pool $\mathcal{E}_{\mathrm{pool}}$. The pixel-level gain is $\mathrm{IG}_{\mathrm{px}}(\mathbf{T}_c) = |\{\mathbf{p} : \mathbf{M}_c(\mathbf{p}) = 0\}|$, where $\mathbf{M}_c(\mathbf{p}) \in \{0,1\}$ indicates confident Gaussian coverage. A structure-dependent discount $\rho_c \in [0, \rho_{\max}^{(\psi)}]$ replaces structurally blind pixel thresholds:

$$\widetilde{\mathrm{IG}}_{\mathrm{px}}(\mathbf{T}_c) = \mathrm{IG}_{\mathrm{px}}(\mathbf{T}_c) \cdot (1 - \rho_c), \tag{12}$$

where ρ_c grows with visible tracked planes and inversely with structural deficit, coupling pool dynamics to reconstruction state.

Structural term with deficiency targeting. Visible planes $\mathcal{V}_c^{\mathrm{vis}} \subseteq \mathcal{V}$ are identified via frustum culling, distance gating, and view-angle filtering:

$$\mathrm{IG}_{\mathrm{str}}(\mathbf{T}_c) = \sum_{v_i \in \mathcal{V}_c^{\mathrm{vis}}} \alpha_i^{(\psi)} \gamma_i \; + \; \omega_d \, \mathrm{IG}_{\mathrm{def}}(\mathbf{T}_c) \,, \tag{13}$$

where $\alpha_i^{(\psi)}$ are phase-dependent type weights and $\mathrm{IG}_{\mathrm{def}}$ evaluates 3D deficiency centres from §3.3 against the viewing frustum, quantifying each candidate's potential to fill localised interior voids. Candidates viewing tracked planes also receive a non-zero *visibility floor*. The composite information gain and next-best-view are:

$$\mathrm{IG}(\mathbf{T}_c) = \widetilde{\mathrm{IG}}_{\mathrm{px}}(\mathbf{T}_c) + \omega^{(\psi)} \, \mathrm{IG}_{\mathrm{str}}(\mathbf{T}_c), \tag{14}$$

$$\mathbf{T}^* = \arg \max_{\mathbf{T}_c \in \mathcal{E}_{\mathrm{pool}}} \; \mathrm{softmax}\big(\log \mathrm{IG}(\mathbf{T}_c)\big) \cdot g(\mathbf{T}_c, \mathbf{T}_{\mathrm{cur}}), \tag{15}$$

where $g(\cdot, \cdot)$ favours reachable candidates. As pixel coverage saturates, the structural term increasingly differentiates candidates. The standard 3DGS optimisation loop then densifies and refines the structurally deficient region, so $\mathrm{IG}_{\mathrm{str}}$ steers attention toward supervision-starved regions without changing the reconstruction pipeline itself. A collision-free path from $\mathbf{T}_{\mathrm{cur}}$ to $\mathbf{T}^*$ is computed via RRT on the discretised SDF with minimum clearance constraints, followed by a smoothing pass interpolating translation and rotation.

Hierarchical exploration with informational polarity reversal. The same structural plane that serves as an exploration *attractor* during scaffold construction becomes a *suppressor* once established. We term this *informational polarity reversal* and instantiate it as a three-phase hierarchy, where each phase controls type weights $\alpha_i^{(\psi)}$, coupling $\omega^{(\psi)}$, and maximum discount $\rho_{\max}^{(\psi)}$:

- **Phase I: Structural scaffold** ($\psi = 1$). Weights emphasise $\mathcal{V}_{\mathrm{S}}$: structural deficit acts as *attractor* toward under-constrained walls, floors, and ceilings. Pruning is conservative. Terminates when the pool is exhausted.
- **Phase II: Object refinement with suppression** ($\psi = 2$). Structural planes undergo *polarity reversal*: they become *termination signals.* Weighting shifts to $\mathcal{V}_{\mathrm{O}}$; discount cap increases. When a candidate's FOV is dominated by well-reconstructed structure, additional discounting redirects budget toward object surfaces benefiting from scaffold context.
- **Phase III: Post-refinement** ($\psi = 3$). Keyframes with lowest rendering fidelity are revisited with increased mapping iterations. Terminates when target keyframes exceed a quality threshold.

Each phase transitions when its *structural objective* is satisfied, coupling phase boundaries directly to reconstruction state rather than predetermined frame counts. The polarity reversal ensures the hierarchy mirrors indoor scene composition.

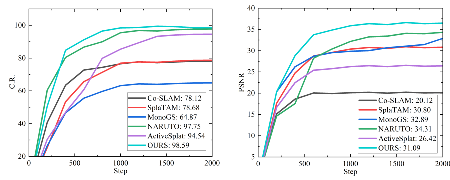
4 Experiments

4.1 Implementation Details

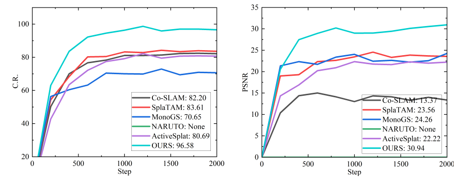
We implement ActiveStructure in Habitat [27] with a $60^\circ{\times}90^\circ$ FOV RGB-D sensor returning posed observations. Each method runs for 2000 frames per sequence (all three stages of ActiveStructure are included.). For the passive reconstruction systems [13, 20, 29], we directly equip them with ground-truth poses and remove the Tracking stage to ensure a fair comparison. All experiments use a single RTX 4090.

Datasets. Following [4, 5, 19]: (i) **Replica** [28] (8 photorealistic indoor scenes); (ii) **Matterport3D** [2] (5 large real-world environments with complex geometry and textureless regions).

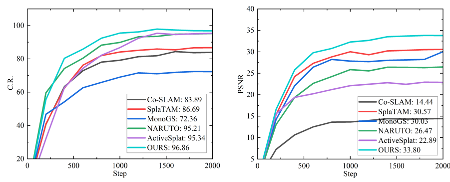
Metrics. Geometric: accuracy (Acc, cm), completion (Com., cm), coverage ratio (C.R., %) at 5 cm threshold. Rendering: PSNR, SSIM, LPIPS on novel views from [4].



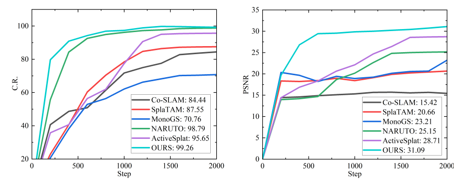
(a) Office1



(b) Office2



(c) Office4



(d) Room2

Fig. 4: Convergence curves of reconstruction completeness and rendering quality for different methods.

Baselines. Active: ActiveINR [34], ANM-S [15], NARUTO [5], ActiveSplat [19], ActiveGAMER [4]. Passive SLAM: Co-SLAM [29], MonoGS [20], SplaTAM [13]. All active GS reconstruction frameworks adopt SplaTAM as the mapping backbone without modifying Gaussian attributes, so performance differences stem solely from their exploration strategies. The top-3 results are highlighted (1st, 2nd, 3rd).

4.2 Reconstruction Quality

Tables 1 and 2 present geometric results.

Replica. ActiveStructure achieves the best average accuracy (1.16 cm), completion (1.49 cm), and coverage (97.47%), surpassing ActiveGAMER (1.28 cm / 1.74 cm / 96.59%) by consistent margins. On the geometrically complex Off2, accuracy improves from 1.37 to 1.22 cm and completion from 2.35 to 1.65 cm, a joint gain that pixel-level methods cannot achieve because they fail to distinguish structurally fragile regions. Coverage on R2 reaches 99.26%, the highest across all methods, confirming that PSG-guided exploration eliminates interior voids eluding pixel-counting criteria.

MP3D. The advantage amplifies on MP3D (Table 2): best average accuracy (1.57 cm), completion (1.91 cm), and coverage (96.23%), improving upon ActiveGAMER by 3.7%, 24.8%, and 1.7%. Gains are most pronounced on structured buildings, where PSG topological indicators detect inter-surface incompleteness persisting despite adequate pixel coverage, and deficiency targeting translates these into gap-resolving viewpoints.

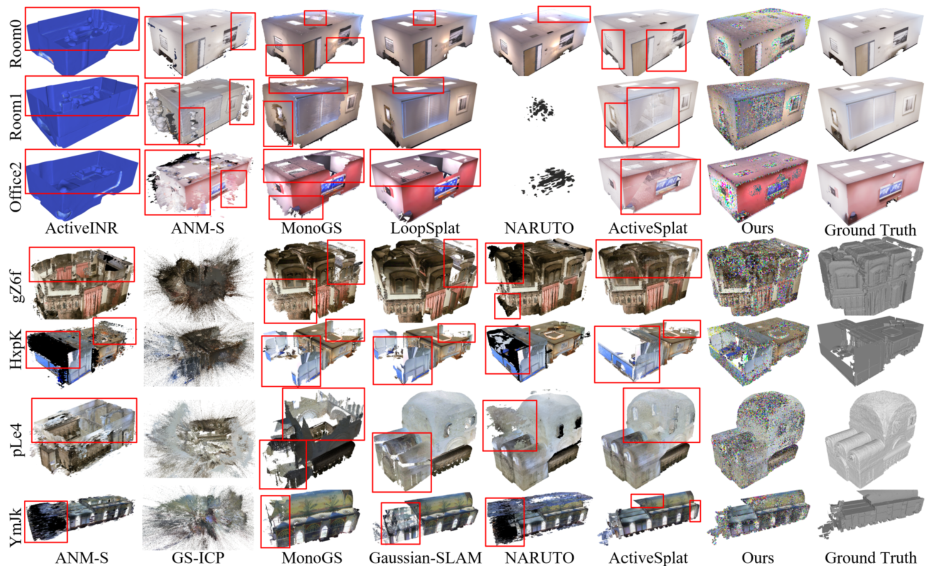


Fig. 5: We present qualitative comparisons of 3D reconstructions, showing that our ActiveStructure achieves high visual quality. Unlike other methods that rely on offline surface reconstruction from captured RGB-D data after exploration, ours is evaluated directly using online Gaussian models.

Qualitative analysis. Fig. 5 shows ActiveStructure produces coherent surfaces with fewer holes and sharper boundaries, particularly at wall–floor junctions. Results are evaluated directly from the online Gaussian model, confirming improvements stem from better observation selection.

4.3 Rendering Quality

Replica. ActiveStructure achieves the best average PSNR (32.13 dB) and SSIM (0.974), and second-best LPIPS (0.116) in Table 1. The +1.14 dB gain over ActiveGAMER exceeds 2 dB on Off2 (+2.21) and Off4 (+2.72), stemming from structural-term-driven revisitation providing multi-view supervision. MonoGS achieves the best LPIPS (0.085) but at only 73.20% coverage versus our 97.47%.

MP3D. ActiveStructure achieves the best PSNR (26.08 dB) and SSIM (0.920) in Table 2, surpassing ActiveGAMER and ActiveSplat by clear margins. Fig.6 compares the rendering quality across all scenes.

Convergence. Fig. 4 shows faster convergence in both coverage and PSNR. ActiveStructure achieves the fastest convergence across scenes by prioritizing dominant structures in Stage I, with object-level exploration in Stage II further improving completeness. Stage III optimizes for PSNR at the cost of accuracy and completeness, yet our method strikes an effective balance between 2D rendering and 3D reconstruction despite their inherent conflict.

Method	Metric	Off0	Off1	Off2	Off3	Off4	R0	R1	R2	Avg.
ActiveINR [34]	Acc (cm) ↓	1.36	1.27	1.66	11.23	3.59	1.77	5.31	1.21	3.43
	Com. (cm) ↓	7.69	7.76	7.04	14.17	12.06	6.87	11.73	4.85	9.02
	C.R. (%) ↑	72.18	70.84	67.34	31.53	49.00	70.61	39.00	80.68	60.15
ANM-S [15]	Acc (cm) ↓	1.44	1.03	1.60	1.80	1.50	1.47	1.29	1.28	1.43
	Com. (cm) ↓	1.98	1.55	6.65	1.13	1.08	0.91	1.02	0.85	1.90
	C.R. (%) ↑	95.43	92.66	79.20	94.98	95.35	96.71	95.66	96.79	93.35
Co-SLAM [29]	Acc (cm) ↓	1.28	1.02	2.09	2.24	1.65	1.55	1.24	1.37	1.56
	Com. (cm) ↓	9.97	9.21	5.29	3.85	4.42	3.56	3.26	4.71	5.53
	C.R. (%) ↑	80.42	78.12	82.20	84.50	83.89	88.49	88.65	84.44	83.84
	PSNR ↑	19.80	20.12	13.37	13.45	14.44	17.23	18.62	15.42	16.56
	SSIM ↑	0.609	0.619	0.468	0.418	0.451	0.491	0.551	0.481	0.511
	LPIPS ↓	0.668	0.626	0.705	0.774	0.766	0.733	0.690	0.716	0.710
MonoGS [20]	Acc (cm) ↓	2.72	4.87	3.88	3.89	3.73	3.49	3.92	3.98	3.81
	Com. (cm) ↓	10.36	10.27	7.12	5.23	5.71	4.86	4.75	7.52	6.98
	C.R. (%) ↑	77.89	64.87	70.65	73.56	72.36	78.47	77.04	70.76	73.20
	PSNR ↑	31.28	32.89	24.26	26.27	30.03	34.21	32.06	23.21	29.28
	SSIM ↑	0.931	0.942	0.907	0.933	0.965	0.964	0.944	0.924	0.939
	LPIPS ↓	0.082	0.095	0.129	0.087	0.056	0.047	0.067	0.113	0.085
SplaTAM [13]	Acc (cm) ↓	1.01	0.88	1.43	1.61	1.49	1.46	1.08	1.23	1.27
	Com. (cm) ↓	10.07	9.24	5.26	3.63	3.83	3.57	3.26	4.17	5.38
	C.R. (%) ↑	81.27	78.68	83.61	86.25	86.69	88.66	88.61	87.55	85.17
	PSNR ↑	29.59	30.80	23.56	26.29	30.57	28.93	28.24	20.66	27.33
	SSIM ↑	0.947	0.959	0.929	0.946	0.977	0.946	0.956	0.911	0.946
	LPIPS ↓	0.134	0.156	0.194	0.189	0.108	0.156	0.113	0.182	0.154
NARUTO [5]	Acc (cm) ↓	1.28	1.03	-	2.79	1.65	1.56	-	1.50	-
	Com. (cm) ↓	1.41	1.20	-	2.41	1.97	1.82	-	1.47	-
	C.R. (%) ↑	97.57	97.75	-	91.75	95.21	96.97	-	98.79	-
	PSNR ↑	31.61	34.31	-	26.45	26.47	26.45	-	25.15	-
	SSIM ↑	0.903	0.923	-	0.835	0.878	0.801	-	0.750	-
	LPIPS ↓	0.169	0.117	-	0.204	0.180	0.219	-	0.283	-
ActiveGAMER [4]	Acc (cm) ↓	0.97	0.87	1.37	1.49	1.51	1.54	1.19	1.26	1.28
	Com. (cm) ↓	1.16	1.07	2.35	2.09	1.97	2.06	1.57	1.62	1.74
	C.R. (%) ↑	98.23	98.42	96.87	96.24	94.72	93.96	97.12	97.17	96.59
	PSNR ↑	34.41	34.79	28.73	31.94	31.08	27.97	28.66	30.37	30.99
	SSIM ↑	0.975	0.967	0.964	0.968	0.974	0.931	0.939	0.963	0.960
	LPIPS ↓	0.101	0.140	0.154	0.084	0.118	0.217	0.171	0.113	0.137
ActiveSplat [19]	Acc (cm) ↓	1.16	1.11	1.47	1.70	1.50	1.67	1.43	1.36	1.43
	Com. (cm) ↓	0.63	0.94	5.59	1.83	1.06	0.84	0.74	1.04	1.58
	C.R. (%) ↑	97.54	94.54	80.69	91.49	95.34	97.04	96.84	95.65	93.64
	PSNR ↑	23.79	26.42	22.22	20.96	22.89	26.36	26.39	28.71	24.72
	SSIM ↑	0.841	0.864	0.881	0.804	0.866	0.830	0.878	0.870	0.854
	LPIPS ↓	0.162	0.129	0.116	0.232	0.275	0.195	0.137	0.116	0.170
OURS	Acc (cm) ↓	0.91	0.81	1.22	1.32	1.31	1.33	1.16	1.19	1.16
	Com. (cm) ↓	1.02	1.04	1.65	2.21	1.67	1.74	1.42	1.21	1.49
	C.R. (%) ↑	98.48	98.59	96.58	95.05	96.86	96.75	98.17	99.26	97.47
	PSNR ↑	35.77	36.47	30.94	29.93	33.80	28.82	29.84	31.49	32.13
	SSIM ↑	0.980	0.975	0.976	0.969	0.985	0.942	0.985	0.982	0.974
	LPIPS ↓	0.084	0.115	0.115	0.127	0.074	0.174	0.148	0.089	0.116

Table 1: Quantitative comparison of 3D reconstruction and view synthesis on the Replica dataset. "-" denotes failure to complete exploration within five trials.

4.4 Ablation Study

Table 3 incrementally activates each component on Replica Off2.

(A) Pixel-only baseline. Worst accuracy (1.51 cm), completion (4.65 cm), and coverage (91.24%) confirm that factorised per-pixel criteria miss inter-surface incompleteness.

(A→B) Naïve PSG. Raw PSG gain yields only marginal gains (+0.09 cm Acc, +0.83% C.R.), lacking discriminative power without anomaly indicators.

(B→C) Anomaly detection. The largest single-component improvement: completion 3.45→1.68 cm, coverage 92.07→96.98%. The multiplicative deficit γ_i isolates surfaces simultaneously under-reconstructed and topologically isolated.

Method	Metric	Gdvg	gZ6f	HxpK	pLe4	YmJk	Avg.
ActiveINR [34]	Acc (cm) ↓	5.09	4.15	15.60	5.56	8.61	7.80
	Com. (cm) ↓	5.69	7.43	15.96	8.03	8.46	9.11
	C.R. (%) ↑	80.99	80.68	48.34	76.41	79.35	73.15
ANM-S [15]	Acc (cm) ↓	5.52	1.62	2.13	4.54	4.50	3.66
	Com. (cm) ↓	3.95	2.01	12.49	2.51	3.53	4.90
	C.R. (%) ↑	91.00	94.58	60.39	95.02	88.65	85.93
MonoGS [20]	Acc (cm) ↓	3.71	7.69	5.44	2.14	3.17	4.43
	Com. (cm) ↓	11.93	8.45	19.44	57.92	121.41	43.83
	C.R. (%) ↑	60.23	48.93	57.14	53.01	47.58	53.38
SplaTAM [13]	Acc (cm) ↓	2.17	2.98	3.07	4.88	4.11	3.44
	Com. (cm) ↓	10.87	5.17	22.92	61.77	111.52	42.45
	C.R. (%) ↑	68.11	76.31	50.87	46.00	43.69	57.00
Gaussian-SLAM [35]	Acc (cm) ↓	2.63	3.42	2.54	3.77	4.32	3.34
	Com. (cm) ↓	12.41	5.87	22.71	6.42	19.41	13.36
	C.R. (%) ↑	66.24	76.11	62.51	73.88	56.43	67.03
NARUTO [5]	Acc (cm) ↓	2.34	3.57	7.29	4.46	9.52	5.44
	Com. (cm) ↓	4.93	2.47	2.84	3.14	5.68	3.81
	C.R. (%) ↑	84.88	93.26	92.15	82.67	78.99	86.39
	PSNR ↑	22.27	21.98	21.96	22.20	19.72	21.63
	SSIM ↑	0.642	0.596	0.668	0.548	0.558	0.602
	LPIPS ↓	0.461	0.463	0.440	0.536	0.493	0.479
ActiveGAMER [4]	Acc (cm) ↓	1.26	1.42	1.31	2.15	2.03	1.63
	Com. (cm) ↓	1.63	1.63	1.62	2.38	5.43	2.54
	C.R. (%) ↑	96.85	97.85	96.72	93.37	88.35	94.63
	PSNR ↑	26.56	26.54	25.71	26.42	21.93	25.43
	SSIM ↑	0.939	0.949	0.911	0.916	0.847	0.912
	LPIPS ↓	0.223	0.224	0.290	0.315	0.374	0.285
ActiveSplat [19]	Acc (cm) ↓	2.39	1.74	2.53	4.09	9.52	4.05
	Com. (cm) ↓	3.76	1.34	24.28	1.07	2.84	6.66
	C.R. (%) ↑	92.11	97.61	44.45	99.10	90.78	84.81
	PSNR ↑	22.77	19.03	19.07	23.49	24.57	21.79
	SSIM ↑	0.700	0.688	0.780	0.667	0.852	0.737
	LPIPS ↓	0.264	0.273	0.227	0.345	0.156	0.253
OURS	Acc (cm) ↓	1.17	1.37	1.25	2.04	2.00	1.57
	Com. (cm) ↓	1.35	1.57	1.56	2.33	2.74	1.91
	C.R. (%) ↑	98.53	97.95	96.56	96.21	91.91	96.23
	PSNR ↑	27.42	27.19	26.72	26.85	22.23	26.08
	SSIM ↑	0.951	0.958	0.928	0.928	0.837	0.920
	LPIPS ↓	0.182	0.186	0.215	0.300	0.396	0.256

Table 2: Quantitative comparison on MP3D for 3D reconstruction and novel view synthesis.

Variant	Acc↓	Com.↓	C.R.↑	PSNR↑	SSIM↑	LPIPS↓	Depth↓
(A) Pixel-only*	1.51	4.65	91.24	28.41	0.962	0.163	1.47
(B) + PSG gain	1.42	3.45	92.07	28.65	0.963	0.157	1.39
(C) + Anomaly detection	1.33	1.68	96.98	29.12	0.964	0.135	1.22
(D) + Deficiency targeting	1.17	1.56	97.21	29.03	0.963	0.131	1.07
(E) Full model	1.22	1.65	96.58	30.94	0.976	0.115	0.91

Table 3: Ablation study on Replica Off2. * Pixel-only baseline selects views by valid-pixel counts alone. Each row cumulatively adds one component.

(C→D) Deficiency targeting. 3D void localisation pushes accuracy to the best (1.17 cm) and coverage to 97.21%, as the agent persistently explores and reconstructs structural voids. However, the sustained exploration focus leaves insufficient optimisation iterations for appearance refinement, resulting in a slight PSNR drop (−0.09 dB).

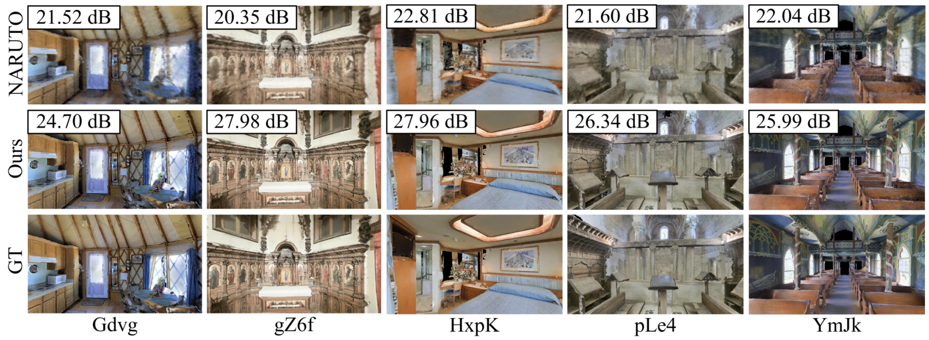


Fig. 6: Qualitative comparison of rendering quality. Our ActiveStructure achieves higher fidelity than previous state-of-the-art methods. The numbers in the top-left corner indicate PSNR.

	Room0				Room1			
Method	MonoGS	ActiveSplat	ActiveGAMER	Ours	MonoGS	ActiveSplat	ActiveGAMER	Ours
Time (min)	32.8	25.7	114.5	20.5	35.6	28.9	99.1	17.6

Table 4: Computational efficiency on Replica dataset.

(D→E) Polarity reversal. The full model achieves the best rendering quality (PSNR 30.94, SSIM 0.976, LPIPS 0.115) and depth accuracy (0.91 cm) with a minor geometric trade-off. Phase II redirects budget to object surfaces via polarity reversal, and Phase III dedicates iterations to appearance optimisation. The modest regression in accuracy (1.17→1.22 cm) and coverage (97.21→96.58%) is substantially outweighed by rendering gains (+1.91 dB PSNR, +0.013 SSIM), confirming the hierarchical strategy effectively resolves the exploration–optimisation tension.

4.5 Computational Efficiency

Table 4 compares wall-clock time on Replica. ActiveStructure achieves a 5.6× speedup over ActiveGAMER (20.5 min vs. 114.5 min on Room0), owing to PSG-guided candidate pruning (Eq. 12) that discards structurally redundant viewpoints and a hierarchical phase strategy that concentrates iterations on deficit-bearing surfaces. It is also 1.3–1.6× faster than ActiveSplat and 1.6–2.0× faster than MonoGS, since the PSG operates over dozens of planar nodes, extracted only from keyframes, rather than thousands of voxels or Gaussians, which keeps the per-step overhead negligible.

5 Conclusion

We presented ActiveStructure, which elevates active 3DGS from pixel-level assessment to scene-level relational reasoning through the Plane Scene Graph. The PSG uses three topological invariants, node connectivity, edge completeness, and angular consistency, as a practical indicator of 3D structural integrity, and it detects forms of incompleteness that no factorised per-pixel metric can reach. A complementary information-gain decomposition mitigates the tension between rendering fidelity and geometric completeness, while informational polarity reversal yields budget-free phase transitions that mirror the compositional hierarchy of indoor scenes. Experiments on Replica and Matterport3D demonstrate consistent state-of-the-art accuracy and rendering quality; ablations confirm each component contributes independently and cumulatively.

Acknowledgment

This work is supported by the Self-Planned Task of SKLRS (HIT) of China (SKLRS202501E), the Heilongjiang Province's Key R&D Program: 'Leading the Charge with Open Competition' (2023ZXJ01A02), the Key Special Projects of Heilongjiang Province's Key R&D Program (2023ZX01A01), the National Research Foundation, Singapore under its Competitive Research Program Award NRF-CRP23-2019- 0003, and the Tier 2 grant MOET2EP20124-0015 from the Singapore Ministry of Education.

References

1. Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P.: Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields. In: ICCV. pp. 5855–5864 (2021)
2. Chang, A., Dai, A., Funkhouser, T., Halber, M., Niessner, M., Savva, M., Song, S., Zeng, A., Zhang, Y.: Matterport3D: Learning from RGB-D data in indoor environments. arXiv preprint arXiv:1709.06158 (2017)
3. Chen, D., Li, H., Ye, W., Wang, Y., Xie, W., Zhai, S., Wang, N., Liu, H., Bao, H., Zhang, G.: Pgsr: Planar-based gaussian splatting for efficient and high-fidelity surface reconstruction. IEEE Transactions on Visualization and Computer Graphics **31**(9), 6100–6111 (2024)
4. Chen, L., Zhan, H., Chen, K., Xu, X., Yan, Q., Cai, C., Xu, Y.: ActiveGAMER: Active gaussian mapping through efficient rendering. In: CVPR. pp. 16486–16497 (2025)
5. Feng, Z., Zhan, H., Chen, Z., Yan, Q., Xu, X., Cai, C., Li, B., Zhu, Q., Xu, Y.: NARUTO: Neural active reconstruction from uncertain target observations. In: CVPR. pp. 21572–21583 (2024)
6. Gu, Q., Kuwajerwala, A., Morin, S., Jatavallabhula, K.M., Sen, B., Agarwal, A., Rivera, C., Paul, W., Ellis, K., Cherian, R., et al.: ConceptGraphs: Open-vocabulary 3D scene graphs for perception and planning. In: ICRA. pp. 6694–6701 (2024)

7. Hanson, A., Tu, A., Singla, V., Jayawardhana, M., Zwicker, M., Goldstein, T.: PUP 3D-GS: Principled uncertainty pruning for 3d gaussian splatting. In: CVPR. pp. 5949–5958 (2025)
8. Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S.: 2d gaussian splatting for geometrically accurate radiance fields. In: ACM SIGGRAPH. pp. 1–11 (2024)
9. Hughes, N., Chang, Y., Carlone, L.: Hydra: A real-time spatial perception system for 3D scene graph construction and optimization. In: Robot. Sci. Syst. (2022)
10. Jiang, W., Lei, B., Daniilidis, K.: FisherRF: Active view selection and mapping with radiance fields using fisher information. In: ECCV. pp. 422–440 (2024)
11. Jin, R., Gao, Y., Wang, Y., Wu, Y., Lu, H., Xu, C., Gao, F.: GS-Planner: A gaussian-splatting-based planning framework for active high-fidelity reconstruction. In: IROS. pp. 11202–11209 (2024)
12. Johari, M.M., Carta, C., Fleuret, F.: ESLAM: Efficient dense SLAM system based on hybrid representation of signed distance fields. In: CVPR. pp. 17408–17419 (2023)
13. Keetha, N., Karhade, J., Jatavallabhula, K.M., Yang, G., Scherer, S., Ramanan, D., Luiten, J.: SplaTAM: Splat track & map 3d gaussians for dense RGB-D SLAM. In: CVPR. pp. 21357–21366 (2024)
14. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4), 139:1–139:14 (2023)
15. Kuang, Z., Yan, Z., Zhao, H., Zhou, G., Zha, H.: Active neural mapping at scale. In: IROS. pp. 7152–7159 (2024)
16. Lee, S., Chen, L., Wang, J., Liniger, A., Kumar, S., Yu, F.: Uncertainty guided policy for active robotic 3D reconstruction using neural radiance fields. IEEE Robot. Autom. Lett. **7**(4), 12070–12077 (2022)
17. Li, Y., Lyu, C., Di, Y., Zhai, G., Lee, G.H., Tombari, F.: GeoGaussian: Geometry-aware gaussian splatting for scene rendering. In: ECCV. pp. 441–457 (2024)
18. Li, Y., Liu, Y., Zhao, L.: Activepolicy: Active gaussian reconstruction and optimization strategy based on global-local information gain. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2026)
19. Li, Y., Kuang, Z., Li, T., Hao, Q., Yan, Z., Zhou, G., Zhang, S.: ActiveSplat: High-fidelity scene reconstruction through active gaussian splatting. IEEE Robot. Autom. Lett. **10**(8), 8099–8106 (2025)
20. Matsuki, H., Murai, R., Kelly, P.H., Davison, A.J.: Gaussian splatting SLAM. In: CVPR. pp. 18039–18048 (2024)
21. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing scenes as neural radiance fields for view synthesis. In: ECCV. pp. 405–421 (2020)
22. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM Trans. Graph. **41**(4), 1–15 (2022)
23. Pan, X., Lai, Z., Song, S., Huang, G.: ActiveNeRF: Learning where to see with uncertainty estimation. In: ECCV. pp. 230–246 (2022)
24. Park, H., Ryu, G., Kim, W.: DropGaussian: Structural regularization for sparse-view gaussian splatting. In: CVPR. pp. 21600–21609 (2025)
25. Ran, Y., Zeng, J., He, S., Chen, J., Li, L., Chen, Y., Lee, G., Ye, Q.: NeurAR: Neural uncertainty for autonomous 3D reconstruction with implicit neural representations. IEEE Robot. Autom. Lett. **8**(2), 1125–1132 (2023)
26. Sandström, E., Li, Y., Van Gool, L., Oswald, M.R.: Point-SLAM: Dense neural point cloud-based SLAM. In: ICCV. pp. 18433–18444 (2023)

27. Savva, M., Kadian, A., Maksymets, O., Zhao, Y., Wijmans, E., Jain, B., Straub, J., Liu, J., Koltun, V., Malik, J., et al.: Habitat: A platform for embodied AI research. In: ICCV. pp. 9339–9347 (2019)
28. Straub, J., Whelan, T., Ma, L., Chen, Y., Wijmans, E., Green, S., Engel, J.J., Mur-Artal, R., Ren, C., Verma, S., et al.: The Replica dataset: A digital replica of indoor spaces. arXiv preprint arXiv:1906.05797 (2019)
29. Wang, H., Wang, J., Agapito, L.: Co-SLAM: Joint coordinate and sparse parametric encodings for neural real-time SLAM. In: CVPR. pp. 13293–13302 (2023)
30. Wilson, J., Almeida, M., Mahajan, S., Labrie, M., Ghaffari, M., Ghasemalizadeh, O., Sun, M., Kuo, C.H., Sen, A.: POp-GS: Next best view in 3D-gaussian splatting with P-optimality. In: CVPR. pp. 3646–3655 (2025)
31. Xie, Y., Gadelha, M., Yang, F., Zhou, X., Jiang, H.: PlanarRecon: Real-time 3D plane detection and reconstruction from posed monocular videos. In: CVPR. pp. 6219–6228 (2022)
32. Xie, Y., Cai, Y., Zhang, Y., Yang, L., Pan, J.: GauSS-MI: Gaussian splatting shannon mutual information for active 3D reconstruction (2025)
33. Xu, Z., Jin, R., Wu, K., Zhao, Y., Zhang, Z., Zhao, J., Gao, F., Gan, Z., Ding, W.: HGS-Planner: Hierarchical planning framework for active scene reconstruction using 3D gaussian splatting. In: ICRA. pp. 14161–14167 (2025)
34. Yan, Z., Yang, H., Zha, H.: Active neural mapping. In: ICCV. pp. 10981–10992 (2023)
35. Yugay, V., Li, Y., Gevers, T., Oswald, M.R.: Gaussian-slam: Photo-realistic dense slam with gaussian splatting (2023)