DC-DLO: DECOUPLING LOCAL REPAIR AND POSE DISCOVERY FOR 3D GAUSSIAN SPLATTING

Anonymous authors
Paper under double-blind review

ABSTRACT

3D Gaussian Splatting (3DGS) has shown strong performance in novel-view synthesis, but few-view reconstruction remains highly sensitive to which observations are acquired under a limited budget. Active view selection addresses this problem by sequentially choosing additional views. Applying sequential selection to 3DGS is difficult, because optimization, densification, and pruning continually change the underlying Gaussian primitive set. A view-selection state indexed by the current primitives therefore lacks stable spatial indexing across acquisition rounds. We propose DC-DLO (**D**iscovery-**C**onstrained **D**ynamic **L**ocal **O**bservability), an active view selection framework that separates reconstruction-dependent local repair from a pose-only diversity mechanism. In this framework, we first introduce World-Space Local Repair (WSLR), which organizes the repair state over fixed world-space cells and rebuilds it from the current 3DGS after each reconstruction update. The spatial indexing remains fixed while the state is recomputed. Each occupied cell stores a compact 3×3 directional support matrix summarizing support from the acquired cameras. Candidate views are evaluated by marginal log-determinant gain, and an Adaptive Top-K module is proposed to aggregate the largest local gains into the repair utility. Next, we introduce the Pose-Only Discovery mechanism, which independently restricts the feasible candidate set using only camera poses and does not modify the repair utility. Candidate RGB is revealed only after a view has been selected. We evaluate DC-DLO on the full nine-scene Mip-NeRF 360 benchmark, the seven-scene Extended NBV benchmark, and the four-scene Light-Field benchmark. DC-DLO achieves the highest aggregate PSNR and SSIM and the lowest LPIPS among the evaluated selectors on all three benchmarks.

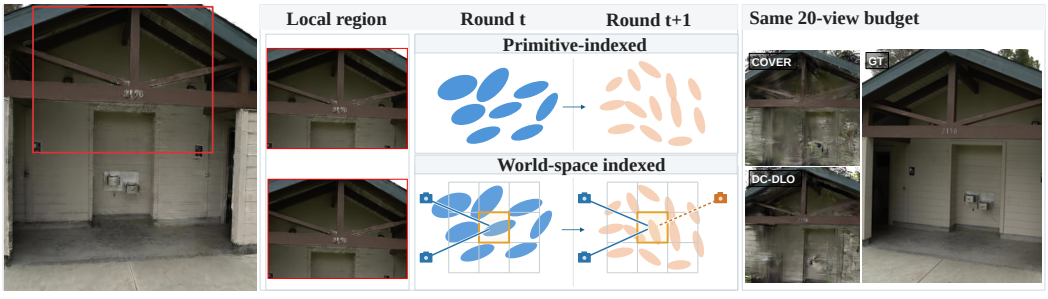

Figure 1: Motivation for fixed spatial indexing. Primitive-indexed scoring units change with the evolving Gaussian representation, whereas fixed world-space cells preserve spatial correspondence across acquisition rounds. Right: representative held-out renderings under the same 20-view budget.

1 INTRODUCTION

3D reconstruction underpins novel-view synthesis, robotic perception, and interactive scene capture. Neural Radiance Fields (NeRFs) (Mildenhall et al., 2020) showed that high-quality scene representations can be learned from posed images, while 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) provides an explicit representation with high rendering quality and efficient optimization. Despite these advances, reconstruction quality remains sensitive to which observations are acquired, espe-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

cially under a limited image budget. View selection therefore becomes an integral part of few-view reconstruction.

Active view-selection methods commonly use the current reconstruction to guide the acquisition of new observations. In 3DGS, however, the representation itself evolves after each acquisition and reconstruction update. Optimization, densification, and pruning continually change the number, locations, and spatial extent of Gaussian primitives, so the same physical region may be represented by different Gaussians across rounds. A view-selection state indexed by the current primitives therefore inherits this changing spatial indexing. Recomputing the state after each update does not remove this dependence because the indexing still follows the current Gaussian set rather than fixed locations in physical space. Figure 1 illustrates this mismatch. This motivates treating the spatial indexing of acquisition evidence as an explicit design choice in sequential 3DGS. Here, acquisition evidence denotes directional-support statistics derived from acquired views over the current reconstruction.

We introduce DC-DLO based on this principle and organize acquisition evidence by fixed world-space locations rather than by evolving Gaussian primitives. World-Space Local Repair (WSLR) rebuilds a per-cell repair state on a fixed voxel grid after each reconstruction update. Here, repair refers to improving directional support over regions represented by the current reconstruction. Each occupied cell stores representative geometry and a compact 3×3 matrix summarizing directional support from the acquired cameras, and candidate views are evaluated by marginal log-determinant gain. Although the Gaussians assigned to a cell may change between rounds, the cell retains the same spatial meaning.

Two practical issues remain. First, summing local gains over all cells can favor candidates that accumulate many weak improvements. Therefore we propose an Adaptive Top-K module to reduce this bias by averaging a shared number of the largest local gains in each round. Second, repair utility is defined only over regions represented by the current reconstruction and cannot score regions absent from the model. The Pose-Only Discovery mechanism complements WSLR by restricting the feasible candidate set using pose novelty, independently of the repair utility. Repair utility then ranks candidates within this set.

We evaluate DC-DLO on Mip-NeRF 360 (Barron et al., 2022), Extended NBV (Kim et al., 2026), and Light-Field (Yücer et al., 2016), spanning different scene types and capture regimes. DC-DLO achieves the highest aggregate PSNR and SSIM and the lowest aggregate LPIPS among the evaluated selectors on all three benchmarks. A controlled spatial-indexing comparison on Mip-NeRF 360 further shows a mean PSNR gain of 0.4643 dB and an LPIPS reduction of 0.01113 when Gaussian-indexed scoring units are replaced by fixed world-space cells. Together, these results support treating the indexing domain of acquisition evidence as an explicit design choice in sequential 3DGS and provide empirical evidence for anchoring repair evidence to fixed physical locations as the Gaussian representation evolves. We further evaluate the selected reconstructions on real-query camera registration using NAVI (Jampani et al., 2023) and an illustrative qualitative grasping case.

Our contributions are as follows.

- We identify the spatial indexing of acquisition evidence as an explicit design choice in sequential 3DGS and introduce **World-Space Local Repair (WSLR)**. WSLR rebuilds directional repair states over fixed world-space cells after each reconstruction update, preserving a stable spatial reference while the underlying Gaussian representation evolves.

- We complement WSLR with the **Pose-Only Discovery mechanism**, which uses only candidate poses and the acquired pose history to define candidate feasibility. Pose novelty restricts the feasible set, while repair utility ranks candidates within that set, keeping pose-diversity control separate from reconstruction-dependent repair utility.

- We provide a controlled comparison of spatial-indexing choices, in which fixed world-space scoring units yield higher mean PSNR and SSIM and lower mean LPIPS than Gaussian-indexed scoring units. Across Mip-NeRF 360, Extended NBV, and Light-Field, the complete DC-DLO system achieves the highest aggregate PSNR and SSIM and the lowest aggregate LPIPS among the evaluated selectors.

2 Related Work

Uncertainty and information-driven view selection. Active view selection for neural scene representations commonly estimates uncertainty or expected information from the current reconstruction. For NeRF-based representations, ActiveNeRF (Pan et al., 2022) uses predictive uncertainty, NeU-NBV (Jin et al., 2023) uses uncertainty in image-based neural rendering, and FisherRF (Jiang et al., 2024) estimates expected information gain in radiance-field parameter space. Related uncertainty-estimation methods include Neural Visibility Field (NVF) (Xue et al., 2024), which models visibility-aware uncertainty for active mapping, and Bayes' Rays (Goli et al., 2024), which provides post-hoc spatial uncertainty for pretrained NeRFs.

For 3DGS, GauSS-MI (Xie et al., 2025) models visual uncertainty per Gaussian and ranks views by mutual information. OUGS (Li et al., 2026a) propagates uncertainty over Gaussian geometry and appearance parameters through the rendering Jacobian and combines image-space uncertainty with semantic masks. POp-GS (Wilson et al., 2025) studies optimal-design criteria over 3DGS parameters, PRIMU (Gottwald et al., 2026) projects training error and directional coverage onto primitives, and GAVIS (Xue et al., 2026) represents anisotropic particle visibility with spherical harmonics. These methods primarily differ in how uncertainty or acquisition utility is estimated from the current scene representation. Our question is different: how should the view-selection state be spatially indexed when the Gaussian primitive set changes across acquisition rounds? DC-DLO uses fixed world-space cells as scoring units, and its 3×3 matrices represent directional geometric support rather than uncertainty covariance or Fisher information over 3DGS parameters.

Coverage and spatial acquisition states. Coverage provides an alternative to explicit uncertainty estimation. NeRF Director (Xiao et al., 2024) uses camera-space coverage for active neural rendering, while COVER (Chen et al., 2026) derives a lightweight coverage surrogate from an approximation to Fisher information and extends it to viewing directions. Spatial representations can also maintain observation statistics: $\mathbb{R}^3$-RECON (Jin et al., 2026) maintains online statistics in a voxel map and aggregates them into scalar renderability. Although closely related to our use of spatial scoring units, its voxel state plays a different role. $\mathbb{R}^3$-RECON persistently updates statistics from RGB-D observations, whereas DC-DLO rebuilds its repair state from the current 3DGS after each reconstruction update. World-space cells therefore provide fixed spatial indexing rather than a persistent map: the Gaussians assigned to a cell may change while its spatial meaning remains fixed.

Active reconstruction and broader next-best-view planning. Broader active-reconstruction methods combine view selection with learned policies, Gaussian mapping, continuous planning, robotic exploration, or training-side augmentation (Chen et al., 2024; 2025; Jin et al., 2025; Li et al., 2025; 2026b;c; Kim et al., 2026). DC-DLO instead considers pool-based sequential selection from a fixed calibrated candidate set, where candidate poses are known but candidate RGB remains unavailable before selection. Within this setting, WSLR ranks modeled regions by repair utility, while the Pose-Only Discovery mechanism independently constrains candidate feasibility using pose novelty. Appendix Table A8 compares the state-indexing units and information-access boundaries of related methods.

3 Method

Section 3.1 defines the pool-based sequential view-selection problem and the closed-loop acquisition setting. Section 3.2 presents World-Space Local Repair, which rebuilds the repair state over fixed world-space scoring units and evaluates candidates through directional local gain. Section 3.3 introduces the Pose-Only Discovery mechanism, which defines the feasible candidate set from pose novelty independently of the current reconstruction. Section 3.4 combines the two branches into the deterministic selection and reconstruction update loop.

3.1 Preliminaries and Overview

Problem setting. At acquisition round t, the observed set $\mathcal{D}_t$ contains the acquired calibrated RGB images used to optimize the current reconstruction $\mathcal{G}_t$, while the remaining selectable poses form $\mathcal{V}_t$. The selector accesses $\mathcal{G}_t$, the acquired pose history, and candidate poses and calibration, and chooses $v_t^\star \in \mathcal{V}_t$. Candidate and held-out RGB images remain unavailable during ranking.

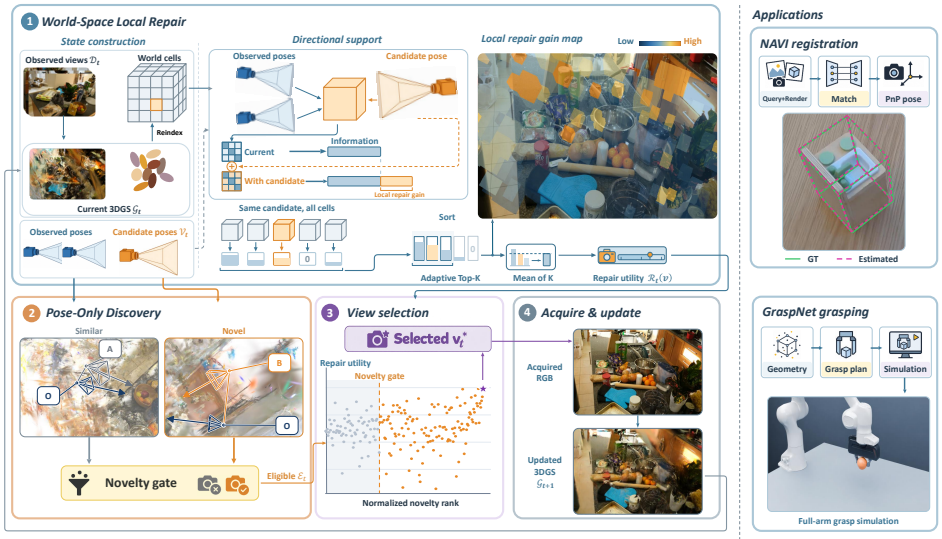

Figure 2: Closed-loop DC-DLO selection. WSLR rebuilds world-space directional repair evidence from the current 3DGS, while Pose-Only Discovery independently defines candidate feasibility. Repair utility ranks feasible poses, and the state is rebuilt after acquiring the selected RGB.

After selection, the chosen RGB image is added to $\mathcal{D}_t$, its pose is removed from $\mathcal{V}_t$, and the updated observed set is used to optimize $\mathcal{G}_{t+1}$ before the next round.

Overview of DC-DLO. As illustrated in Figure 2, DC-DLO separates view selection into two branches. The World-Space Local Repair branch rebuilds the repair state on a fixed world-space voxel grid from the current 3DGS and evaluates how much each candidate complements the directional support of modeled regions. The Pose-Only Discovery mechanism, in contrast, uses only camera poses to define the feasible candidate set. The feasible candidates are then ranked by repair utility.

3.2 World-Space Local Repair

World-space repair state. The 3DGS primitive set changes during optimization through adaptive density control (Kerbl et al., 2023). As densification, pruning, and optimization change the number and placement of Gaussians, the same spatial region may be represented by different scoring units from one round to the next. A Gaussian-indexed state therefore inherits this change in its spatial indexing.

We instead use occupied cells of a fixed world-space voxel grid as scoring units and rebuild the view-selection state at every round. The current Gaussians are filtered and assigned to these cells. Within each occupied cell, the representative point is computed as the activated-opacity-weighted mean of its Gaussian centers, and an oriented local normal is estimated from neighboring cell representatives. These quantities form the rebuilt world-space repair state $\mathcal{C}_t$. Each acquired camera contributes at most once per cell.

The key distinction is between fixed spatial indexing and a rebuilt state. The world-space cell locations remain fixed, while occupied cells, representative points, normals, visibility relations, and directional support matrices are recomputed from $\mathcal{G}_t$ at every round. The voxel grid therefore keeps the spatial indexing stable without requiring Gaussian correspondence or transfer of Gaussian-indexed state across rounds. Exact filtering, normal estimation, and visibility rules are given in Appendix A.5.

Directional support. Let $\mathcal{O}_\ell$ denote the acquired cameras from which cell ℓ passes a coarse geometric visibility check and the front-facing criterion. For each $j \in \mathcal{O}_\ell$, let $\boldsymbol{r}_{\ell j}$ be the unit direction from the representative point to camera j, and let $\alpha_\ell(j)$ be its visibility-gated facing weight.

We define the directional support contributed by camera j to cell ℓ as

$$\boldsymbol{J}_\ell(j) = \alpha_\ell(j)\left(\boldsymbol{I} - \boldsymbol{r}_{\ell j}\boldsymbol{r}_{\ell j}^{\mathsf{T}}\right). \tag{1}$$

For a unit viewing direction, the projector $\boldsymbol{I} - \boldsymbol{r}\boldsymbol{r}^\top$ removes the component along the viewing ray and retains the two orthogonal directions (Zhao & Zelazo, 2019). Thus, $\boldsymbol{J}_\ell(j)$ describes the local directional support provided by camera j, with $\alpha_\ell(j)$ weighting that contribution according to visibility and surface facing.

We accumulate the contributions from all acquired cameras to form the directional support state of cell ℓ:

$$\boldsymbol{H}_\ell = \lambda \boldsymbol{I} + \sum_{j \in \mathcal{O}_\ell} \boldsymbol{J}_\ell(j), \tag{2}$$

where $\lambda > 0$ is an isotropic regularizer that keeps $\boldsymbol{H}_\ell$ positive definite. Intuitively, $\boldsymbol{H}_\ell$ summarizes how well the cell has already been observed from different directions. We use it as a compact geometric directional-support matrix, distinct from renderer Fisher information and uncertainty covariance.

For a candidate pose v, we construct $\boldsymbol{J}_\ell(v)$ using the same form and the current cell geometry; it is zero when the candidate fails either the coarse geometric visibility check or the front-facing criterion. A candidate is therefore more useful when it contributes support along directions that are weakly represented in the current $\boldsymbol{H}_\ell$, rather than repeatedly reinforcing directions that are already well supported. The round index is omitted on cell-level quantities for readability. Appendix A.6 derives the bearing form and states precisely which metric factors are retained.

Local log-determinant gain. The candidate contribution $\boldsymbol{J}_\ell(v)$ depends only on its pose, calibration, and current modeled geometry, and therefore requires no candidate RGB. Motivated by determinant-based criteria in D-optimal experimental design (Boyd & Vandenberghe, 2004) and log-determinant information gain used in active radiance-field selection (Jiang et al., 2024), we measure additional directional support by

$$\Delta_\ell(v) = \log\det(\boldsymbol{H}_\ell + \boldsymbol{J}_\ell(v)) - \log\det \boldsymbol{H}_\ell. \tag{3}$$

Here $\boldsymbol{H}_\ell$ is a direction-only geometric proxy rather than a renderer Fisher information matrix. Since $\boldsymbol{H}_\ell \succ \boldsymbol{0}$ and $\boldsymbol{J}_\ell(v) \succeq \boldsymbol{0}$, the exact gain is nonnegative. With comparable facing weights, contributions that complement weakly supported directions tend to obtain larger gains than those that reinforce already well-supported directions. Unlike idealized bearing information, our local gain intentionally omits the inverse-squared range factor d^{-2} so that utility emphasizes viewing-direction complementarity rather than metric triangulation precision; Appendix Table A9 evaluates this modeling choice.

Adaptive Top-K aggregation. A raw sum over cells can favor a candidate by accumulating many weak gains. For each candidate, let

$$N_t(v) = \left|\{\ell : \Delta_\ell(v) > \varepsilon_\Delta\}\right|. \tag{4}$$

This quantity denotes the number of cells whose gain exceeds ε_Δ. We set a common truncation budget K_t from the rounded median of $N_t(v)$ over the full remaining pool, clamped to at least one. If $\Delta_{(k)}(v)$ is the k-th largest local gain over occupied cells, the repair utility is

$$R_t(v) = \frac{1}{K_t} \sum_{k=1}^{K_t} \Delta_{(k)}(v). \tag{5}$$

The shared cutoff limits accumulation of many small local gains while comparing all candidates at a common aggregation width. Because K_t is computed from the full remaining pool before the Pose-Only Discovery mechanism is applied, the resulting feasibility filtering does not alter the repair aggregation rule. Exact rounding and numerical handling are given in Appendix A.7.

3.3 Pose-Only Discovery

The repair utility $R_t(v)$ evaluates directional complementarity only in regions represented by the current reconstruction. Regions absent from the model have no occupied cells and therefore receive no local repair utility; errors in representative geometry, normals, or coarse visibility can also affect the utility. Fixed world-space indexing avoids tying the spatial organization of the scoring units to individual Gaussian primitives, while the repair utility still depends on the quality of the current

reconstruction. We therefore introduce the Pose-Only Discovery mechanism, which defines candidate feasibility using only camera poses and independently of the current reconstruction.

For each candidate v, let $\nu_t(v)$ denote its novelty with respect to the acquired pose history, defined by its distance to the nearest acquired pose under a pose-distance function combining normalized camera-center separation and optical-axis separation. Let $p_t(v) \in [0, 1]$ be the normalized average rank of $\nu_t(v)$ within the full remaining pool $\mathcal{V}_t$, with larger values indicating greater pose novelty. Pose normalization is fixed once from the initial acquired and selectable pose pool, and both novelty and ranking use only supplied acquisition poses and the acquired pose history.

Let $\bar{\nu}_t$ denote the mean novelty of the current remaining pool. We compute a fixed reference $\bar{\nu}_0$ once before sequential acquisition from the initial acquired set and the initial remaining candidate pool. The relative novelty depletion is then defined as

$$\rho_t = \mathrm{clip}\left(1 - \frac{\bar{\nu}_t}{\bar{\nu}_0 + \varepsilon_\nu}, 0, 1\right), \tag{6}$$

where $\varepsilon_\nu = 10^{-12}$ is a fixed numerical stabilizer that prevents division by zero. The feasible candidate set is then defined as

$$\mathcal{E}_t = \left\{ v \in \mathcal{V}_t : p_t(v) \geq \frac{1 - \rho_t}{2} \right\}. \tag{7}$$

Since $\rho_t \in [0, 1]$, the rank threshold $g_t = (1 - \rho_t)/2$ lies in $[0, \frac{1}{2}]$. When the pool mean novelty $\bar{\nu}_t$ remains close to its initial reference $\bar{\nu}_0$, the threshold stays near $\frac{1}{2}$. Lower mean novelty increases ρ_t and reduces g_t, thereby admitting a larger fraction of the remaining pool. Because ρ_t is recomputed from the current remaining pool, g_t need not evolve monotonically across acquisition rounds.

The rank $p_t(v)$ orders candidates within the current pool, whereas ρ_t controls the feasibility threshold used by the Pose-Only Discovery mechanism. Pose novelty does not enter the repair utility $R_t(v)$; it affects feasibility and is used secondarily for tie breaking and the near-zero-repair fallback. Appendix A.8 specifies the exact normalization, ranking, and edge-case rules.

3.4 Discovery-Constrained View Selection

At the final decision, the Pose-Only Discovery mechanism defines the feasible candidate set, while repair utility ranks candidates within that set,

$$v_t^\star = \arg\max_{v \in \mathcal{E}_t} R_t(v). \tag{8}$$

Normalized novelty rank $p_t(v)$ and then the lower frame identifier resolve repair-utility ties. If every candidate in the full remaining pool has $R_t(v) \leq \epsilon_0$, where $\epsilon_0 = 10^{-12}$, numerically negligible repair-utility differences are ignored and selection falls back to novelty rank and frame identifier. The deterministic ordering is given in Appendix A.8.

Only after selection do we reveal the chosen RGB image, add it to the observed set to form $\mathcal{D}_{t+1}$, and remove its pose from the candidate pool to form $\mathcal{V}_{t+1}$. The updated observed set is used to optimize $\mathcal{G}_{t+1}$, after which the world-space repair state is rebuilt. Beyond the current reconstruction and fixed initialization statistics $(\boldsymbol{m}_c, s_c, \bar{\nu}_0)$, only the acquired indices are carried across rounds.

The directional state stores one 3×3 matrix per occupied cell, and each log-determinant evaluation involves only a 3×3 matrix. The selector constructs neither a renderer-parameter Hessian nor a global covariance over Gaussian parameters.

4 Experiments

We evaluate DC-DLO on three reconstruction benchmarks, controlled ablations, and downstream tasks. The ablations test fixed world-space indexing, the directional local state, and the Pose-Only Discovery mechanism under matched protocols. We also measure selector runtime and evaluate real-query registration and qualitative grasp execution.

4.1 Experimental Setup

Datasets. We evaluate on three public benchmarks covering different capture regimes. **Mip-NeRF 360** (Barron et al., 2022) contains nine real-world scenes, including bounded indoor environments and unbounded outdoor captures with clutter, foliage, occlusion, and large depth variation. **Extended NBV**, introduced by SA-ResGS (Kim et al., 2026), contains seven challenging scenes curated from Deep Blending (Hedman et al., 2018) and Tanks and Temples (Knapitsch et al., 2017), with irregular camera trajectories, different scene scales, and non-uniform observations. We also use the four-scene **Light-Field** benchmark (Yücer et al., 2016)—Torch, Statue, Basket, and Africa—following recent 3DGS active-view evaluation (Li et al., 2026a). Its object-centric capture geometry differs from the scene-level settings of the other two benchmarks.

Protocol and baselines. Mip-NeRF 360 uses two seeds per scene. Each run starts from four initial views, selects 16 additional views sequentially, and optimizes the final 20-view 3DGS for 30K iterations. Extended NBV and Light-Field use the same 4+16-view acquisition protocol and 30K final optimization with seed 0. We compare Random, COVER (Chen et al., 2026), FisherRF (Jiang et al., 2024), GauSS-MI (Xie et al., 2025), and DC-DLO.

Within each benchmark, all methods share the same initialization, candidate pool, held-out split, reconstruction backend, metric implementation, and information-access boundary: candidate RGB, held-out RGB, depth, and ground-truth geometry are unavailable during ranking. COVER uses the official implementation under this common protocol, while GauSS-MI is adapted to the same RGB-only pool-based setting using only acquired RGB observations and the current 3DGS. Further protocol details are given in Appendix A.1. Bicycle, Garden, and Room form the targeted subset used for the component and sensitivity diagnostics. Full split, subset protocol, and reconstruction settings are reported in Appendix A.2.

Metrics. We report held-out PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018). Metrics are first averaged over held-out views for each combination of scene and seed and then averaged across all scene–seed combinations.

4.2 Reconstruction Quality

Table 1: Final-budget reconstruction quality across the three evaluation sets. Best and second-best values within each benchmark are **bold** and underlined.

	Mip-NeRF 360			Extended NBV			Light-Field		
Method	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Random	<u>15.780</u>	<u>0.3661</u>	<u>0.4804</u>	14.324	<u>0.3826</u>	<u>0.4827</u>	<u>13.408</u>	<u>0.3319</u>	<u>0.4800</u>
FisherRF (ECCV'24)	15.523	0.3401	0.4966	<u>14.385</u>	0.3764	0.4912	12.771	0.2899	0.5085
GauSS-MI (RSS'25)	15.265	0.3358	0.4988	14.145	0.3647	0.5014	13.174	0.3248	0.4851
COVER (CVPR'26)	15.436	0.3400	0.4966	14.173	0.3767	0.4858	13.272	0.3224	0.4870
DC-DLO (Ours)	**16.224**	**0.3739**	**0.4738**	**14.584**	**0.3895**	**0.4764**	**13.425**	**0.3392**	**0.4761**

Cross-benchmark reconstruction performance. Table 1 shows that DC-DLO achieves the highest aggregate PSNR and SSIM and the lowest LPIPS on all three benchmarks. On Mip-NeRF 360, DC-DLO improves mean PSNR over Random by 0.443 dB. Notably, Random outperforms several model-based selectors in aggregate. We conjecture that the predefined candidate trajectories already provide broad spatial and angular coverage, making unguided sampling competitive. DC-DLO further combines reconstruction-dependent local repair with pose-diversity preservation, which is consistent with its stronger aggregate performance over Random. The evaluation spans indoor and outdoor scenes, irregular camera trajectories, and object-centric captures. On the five-scene Extended NBV runtime subset, DC-DLO is faster than COVER and FisherRF and remains close to GauSS-MI in selector time (Appendix Table A12).

Qualitative comparison. Figure 3 shows that, at the same 20-view budget, DC-DLO retains fine structure and local appearance in the highlighted Mip-NeRF 360 renderings.

4.3 Ablation Studies

We first examine the components of WSLR and their interaction with the Pose-Only Discovery mechanism on three scenes. Within WSLR, we separately vary the world-space repair state and

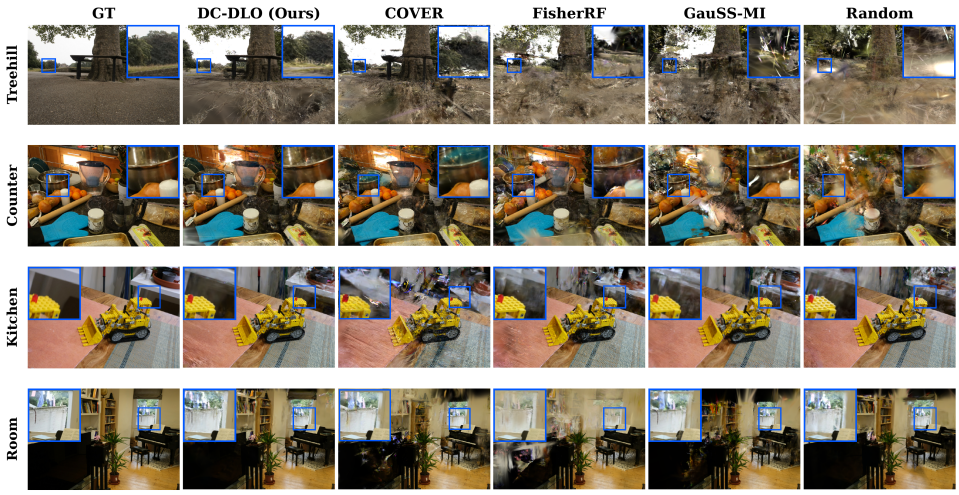

Figure 3: Qualitative comparison on held-out Mip-NeRF 360 views. Rows show Treehill, Counter, Kitchen, and Room; columns show GT, DC-DLO, COVER, FisherRF, GauSS-MI, and Random. Blue boxes mark identical detail regions across methods, with enlarged crops shown for visual comparison.

Adaptive Top-K while keeping the directional matrix objective fixed. We then isolate the spatial-indexing choice on the full nine-scene Mip-NeRF 360 benchmark. A scalar-versus-matrix comparison with mixed metric effects is reported in Appendix A.4, with paired statistics in Appendix C.

Table 2: **Component ablation** on Bicycle, Garden, and Room. WSLR components and the Pose-Only Discovery mechanism are varied with the directional matrix objective fixed.

Exp.	WSLR		Pose-Only Discovery	PSNR ↑	SSIM ↑	LPIPS ↓
	World-space repair state	Adaptive Top-K				
1	×	×	×	15.1943	0.3628	0.4873
2	✓	×	×	14.5340	0.3320	0.5080
3	×	✓	×	16.1813	0.3920	0.4671
4	✓	×	✓	15.4110	0.3600	0.4820
5	✓	✓	×	15.9835	0.3896	0.4655
6	×	✓	✓	16.2621	**0.3982**	0.4661
7	✓	✓	✓	**16.4028**	0.3956	**0.4586**

Component ablation. Comparisons between Exp. 1, Exp. 2, and Exp. 3 show that the world-space repair state alone is insufficient to improve the all-off baseline on this subset, whereas Adaptive Top-K provides a clear gain by limiting the accumulation of weak local gains.

Comparisons between Exp. 2, Exp. 4, Exp. 5, and Exp. 7 show that Adaptive Top-K and the Pose-Only Discovery mechanism play complementary roles within the world-space repair route. Either component improves over Exp. 2, while their combination in Exp. 7 achieves the highest PSNR and lowest LPIPS.

Comparisons between Exp. 3 and Exp. 6 further show that the Pose-Only Discovery mechanism is also beneficial with Gaussian-indexed repair, supporting its role as a pose-diversity mechanism independent of the repair-state indexing.

Finally, comparisons between Exp. 6 and Exp. 7 show that the complete world-space configuration further improves PSNR and LPIPS, although Exp. 6 retains the highest SSIM. Since these configurations differ in more than spatial indexing, we next isolate the indexing choice under matched non-indexing components.

Controlled spatial indexing. Table 3 matches the source Gaussians and all non-indexing components while changing only the indexing domain. Fixed world-space cells yield a mean PSNR gain of 0.4643 dB and reduce LPIPS by 0.01113, with the corresponding 95% confidence intervals

Table 3: **Controlled spatial-indexing comparison** on the full nine-scene Mip-NeRF 360 benchmark. Differences are fixed world-space cells minus Gaussian-indexed scoring units. The LPIPS column reports $-\Delta$LPIPS.

	ΔPSNR ↑	ΔSSIM ↑	$-\Delta$LPIPS ↑
Mean effect	$+0.4643$	$+0.01277$	$+0.01113$
95% CI	$[0.0064, 0.9711]$	$[-0.00301, 0.03016]$	$[0.00059, 0.02172]$

excluding zero; the SSIM interval crosses zero. These results support fixed world-space indexing as an independent design choice under the evaluated protocol.

4.4 Downstream Task Utility

Table 4: NAVI real-query pose recovery. Success requires rotation error $\leq 5^\circ$ and translation error $\leq 0.05D$.

Method	Success ↑	Median rot. ($^\circ$) ↓	Median trans./D ↓
Random	30.47%	1.4338	0.0746
COVER	21.29%	1.5770	0.0847
FisherRF	30.08%	1.3721	0.0688
GauSS-MI	26.76%	1.5194	0.0763
DC-DLO	**32.13%**	**1.2857**	**0.0672**

Real-query pose recovery. We evaluate NAVI (Jampani et al., 2023) over 5,120 trials across four objects and two seeds, without ground-truth depth. DC-DLO achieves the highest pooled success rate at 32.13% (Table 4). Because registration uses an independent query image followed by PnP, the result shows that the selected reconstructions retain useful structure outside the acquisition pool. Detailed statistics are reported in Appendix B.

Grasping. To probe whether the selected 20-view reconstructions can support downstream manipulation, we present one matched simulated grasping trial after view acquisition. Under the same perturbation, the reconstruction selected by DC-DLO supports a successful grasp, whereas the FisherRF reconstruction fails (Figure 4). Grasp outcomes do not inform view selection. This single case illustrates downstream feasibility; aggregate grasp-success rates remain unmeasured. Full sequences are provided in Appendix B.7.

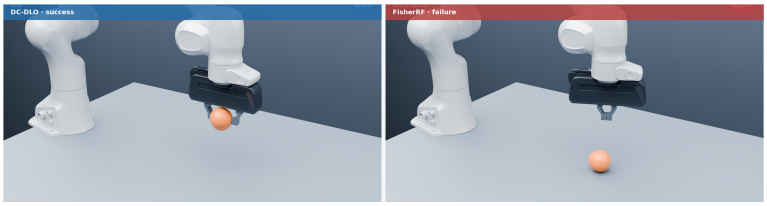

Figure 4: **Illustrative matched grasping case.** Under the same perturbation, the reconstruction selected by DC-DLO supports a successful grasp, whereas the FisherRF reconstruction fails.

5 Conclusion

We presented DC-DLO, a pool-based 3DGS view selector that organizes repair evidence over fixed world-space cells. WSLR ranks candidates by directional local repair with Adaptive Top-K, while the Pose-Only Discovery mechanism separately preserves pose diversity. Across three reconstruction benchmarks, DC-DLO achieves the strongest aggregate reconstruction quality among the evaluated selectors, and a controlled comparison supports fixed world-space indexing under matched non-indexing components. The selected reconstructions also support out-of-pool registration and an illustrative grasping case.

The current method remains limited by reconstruction-dependent repair, fixed-scale voxelization, and calibrated pool-based candidate poses. Future work could investigate adaptive spatial partitions and continuous viewpoint planning.

AI USE STATEMENT

Generative AI tools were used to assist with language editing and manuscript organization. They were not used as experimental evidence. The authors remain responsible for verifying every technical claim, derivation, result, and citation in the submitted version.

REFERENCES

Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5470–5479, June 2022.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Liyan Chen, Huangying Zhan, Kevin Chen, Xiangyu Xu, Qingan Yan, Changjiang Cai, and Yi Xu. ActiveGAMER: Active GAussian mapping through efficient rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16486–16497, 2025.

Timothy Chen, Adam Dai, Maximilian Adang, Grace Gao, and Mac Schwager. Coverage optimization for camera view selection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19443–19451, 2026.

Xiao Chen, Quanyi Li, Tai Wang, Tianfan Xue, and Jiangmiao Pang. GenNBV: Generalizable next-best-view policy for active 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16436–16445, 2024.

Lily Goli, Cody Reading, Silvia Sellán, Alec Jacobson, and Andrea Tagliasacchi. Bayes' rays: Uncertainty quantification for neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20061–20070, June 2024.

Thomas Gottwald, Edgar Heinert, Peter Stehr, Chamuditha Jayanga Galappaththige, and Matthias Rottmann. PRIMU: Uncertainty estimation for novel views in gaussian splatting from primitive-based representations of error and coverage. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11871–11880, June 2026.

Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel J. Brostow. Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics*, 37(6): 257:1–257:15, 2018. doi: 10.1145/3272127.3275084.

Varun Jampani, Kevis-Kokitsi Maninis, Andreas Engelhardt, Arjun Karpur, Karen Truong, Kyle Sargent, Stefan Popov, Andre Araujo, Ricardo Martin-Brualla, Kaushal Patel, Daniel Vlasic, Vittorio Ferrari, Ameesh Makadia, Ce Liu, Yuanzhen Li, and Howard Zhou. NAVI: Category-agnostic image collections with high-quality 3d shape and pose annotations. In *NeurIPS*, 2023. URL `https://navidataset.github.io/`.

Wen Jiang, Boshu Lei, and Kostas Daniilidis. FisherRF: Active view selection and mapping with radiance fields using fisher information. In *European Conference on Computer Vision*, 2024.

Liren Jin, Xieyuanli Chen, Julius Rückin, and Marija Popović. NeU-NBV: Next best view planning using uncertainty estimation in image-based neural rendering. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 11305–11312, 2023.

Liren Jin, Xingguang Zhong, Yue Pan, Jens Behley, Cyrill Stachniss, and Marija Popović. ActiveGS: Active scene reconstruction using gaussian splatting. *IEEE Robotics and Automation Letters*, 10 (5):4866–4873, 2025. doi: 10.1109/LRA.2025.3555149.

Xiaofeng Jin, Matteo Frosi, Yiran Guo, and Matteo Matteucci. $\mathbb{R}^3$-RECON: Radiance-field-free active reconstruction via renderability. *arXiv preprint arXiv:2601.07484*, 2026. URL `https://arxiv.org/abs/2601.07484`.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, 42(4):139:1–139:14, 2023.

Jun-Seong Kim, Tae-Hyun Oh, Eduardo Pérez-Pellitero, and Youngkyoon Jang. SA-ResGS: Self-augmented residual 3D Gaussian Splatting for next best view selection. In *European Conference on Computer Vision (ECCV)*, pp. 300–319, 2026.

Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4):78:1–78:13, 2017. doi: 10.1145/3072959.3073599.

Haiyi Li, Qi Chen, Denis Kalkofen, and Hsiang-Ting Chen. OUGS: Active view selection via object-aware uncertainty estimation in 3DGS. *Computer Graphics Forum*, 2026a.

Shiyao Li, Antoine Guédon, Shizhe Chen, and Vincent Lepetit. MAGICIAN: Efficient long-term planning with imagined gaussians for active mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21606–21615, June 2026b.

Yingzhao Li, Yanjie Liu, and Lijun Zhao. 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*, 2026c.

Yuetao Li, Zijia Kuang, Ting Li, Qun Hao, Zike Yan, Guyue Zhou, and Shaohui Zhang. ActiveSplat: High-fidelity scene reconstruction through active gaussian splatting. *IEEE Robotics and Automation Letters*, 10(8):8099–8106, 2025.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, pp. 405–421, 2020.

Xuran Pan, Zihang Lai, Shiji Song, and Gao Huang. ActiveNeRF: Learning where to see with uncertainty estimation. In *European Conference on Computer Vision (ECCV)*, pp. 230–246, 2022.

Johannes L. Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

Joey Wilson, Marcelino Almeida, Sachit Mahajan, Martin Labrie, Maani Ghaffari, Omid Ghasemalizadeh, Min Sun, Cheng-Hao Kuo, and Arnab Sen. POp-GS: Next best view in 3D-gaussian splatting with P-optimality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3646–3655, June 2025.

Wenhui Xiao, Rodrigo Santa Cruz, David Ahmedt-Aristizabal, Olivier Salvado, Clinton Fookes, and Leo Lebrat. NeRF Director: Revisiting view selection in neural volume rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20742–20751, 2024.

Yuhan Xie, Yixi Cai, Yinqiang Zhang, Lei Yang, and Jia Pan. GauSS-MI: Gaussian splatting shannon mutual information for active 3d reconstruction. In *Proceedings of Robotics: Science and Systems*, 2025.

Shangjie Xue, Jesse Dill, Pranay Mathur, Frank Dellaert, Panagiotis Tsiotras, and Danfei Xu. Neural visibility field for uncertainty-driven active mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18122–18132, 2024.

Shangjie Xue, Jesse Dill, Dhruv Ahuja, Frank Dellaert, Panagiotis Tsiotras, and Danfei Xu. Uncertainty-driven 3D Gaussian Splatting active mapping via anisotropic visibility field. *arXiv preprint arXiv:2605.30342*, 2026. URL `https://arxiv.org/abs/2605.30342`.

Kaan Yücer, Alexander Sorkine-Hornung, Oliver Wang, and Olga Sorkine-Hornung. Efficient 3d object segmentation from densely sampled light fields with applications to 3d reconstruction. *ACM Transactions on Graphics*, 35(3):22:1–22:15, 2016.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

Shiyu Zhao and Daniel Zelazo. Bearing rigidity theory and its applications for control and estimation of network systems: Life beyond distance rigidity. *IEEE Control Systems Magazine*, 39(2):66–83, April 2019.

A Reproducibility Details

A.1 Acquisition Protocol and Information Access

Let $\mathcal{I}$ denote the finite acquisition-view pool containing the initial and selectable views, $\mathcal{S}_t \subseteq \mathcal{I}$ the acquired frame indices at round t, and $\mathcal{R}_t = \mathcal{I} \setminus \mathcal{S}_t$ the remaining selectable frame indices. These index sets induce the observed set $\mathcal{D}_t$ and candidate set $\mathcal{V}_t$ used in the main text. Candidate camera pose and calibration are available before selection, but candidate RGB is not opened until the candidate is selected. Held-out evaluation views are disjoint from the acquisition-view pool and are excluded from pose normalization, selector inputs, and reconstruction optimization. Their RGB images may be read only by the separate evaluation routine for reporting metrics and never affect selector scores, training loss, backpropagation, or model updates.

At every dynamic round, the implementation first rebuilds the current world-space repair state from the 3DGS optimized on the acquired views, evaluates cell-level candidate gains over the full remaining pool, computes the shared support width K_t and repair utilities $R_t(v)$, then computes pose novelty and the feasibility mask of the Pose-Only Discovery mechanism before applying the final deterministic selection rule.

GauSS-MI protocol adaptation. For the GauSS-MI baseline, we use an RGB-only pool-based adaptation under the same information-access boundary as the other selectors. Its Gaussian-level uncertainty and mutual-information ranking are computed from the current 3DGS state obtained from already acquired RGB observations. Candidate RGB and held-out RGB are unavailable during ranking, and no depth or ground-truth geometry is provided to the selector. After a view is selected, only its newly acquired RGB image becomes available for the subsequent reconstruction and baseline update. This adaptation preserves the Gaussian-conditioned acquisition principle of GauSS-MI while matching the RGB-only sensing protocol used in our evaluation. Throughout the main text and tables, this protocol-matched variant is referred to simply as GauSS-MI.

A.2 Evaluation Splits and Shared Reconstruction Settings

Mip-NeRF 360 evaluation split. The primary benchmark uses a fixed modulo-8 held-out split over the `images_4` directory of each scene. Ordinary image files are sorted by their full filenames using Python's stable, case-sensitive string ordering; with zero-based indices and offset zero, an image is held out exactly when its index satisfies $i \bmod 8 = 0$. The same held-out IDs are used for both seeds and every method, and they are excluded from pose normalization, selection, and reconstruction training. Table A1 reports the resulting split for all nine Mip-NeRF 360 scenes used in the Full-9 evaluation. For Stump, the rule yields 16 held-out views at indices $0, 8, \ldots, 120$.

Table A1: Fixed Full-9 Mip-NeRF 360 held-out split at the `images_4` resolution.

Scene	Total images	Held-out views	Image size (W×H)
Bicycle	194	25	1237×822
Flowers	173	22	1256×828
Garden	185	24	1297×840
Treehill	141	18	1267×832
Bonsai	292	37	780×520
Counter	240	30	779×519
Kitchen	279	35	779×520
Room	311	39	779×519
Stump	125	16	1245×825

Transfer benchmarks. Extended NBV is evaluated on all seven benchmark scenes introduced by SA-ResGS (Kim et al., 2026) and curated from Deep Blending (Hedman et al., 2018) and Tanks and Temples (Knapitsch et al., 2017): Horse, Truck, Francis, Ballroom, Barn, Ponche, and Playroom. Light-Field is evaluated on Torch, Statue, Basket, and Africa from the densely sampled Light-Field data (Yücer et al., 2016), following its use in recent 3DGS active-view evaluation (Li et al., 2026a). Both benchmarks use seed 0, four shared initial views, 16 sequential acquisitions, and a final 20-view, 30K reconstruction. Candidate pools and held-out splits are fixed before selection and shared across methods. Candidate RGB, held-out RGB, depth, and ground-truth geometry are excluded from selector inputs.

Table A2: Protocol summary for the two transfer benchmarks.

Benchmark	Scenes	Seeds	Initial	Acquired	Final views
Extended NBV	7	1	4	16	20
Light-Field	4	1	4	16	20

Shared 3DGS backend. All methods use the same fixed Inria-3DGS-derived reconstruction backend and the same shared initialization for a given scene and protocol seed. The reconstruction and evaluation settings used by the reported experiments are summarized in Table A3.

Table A3: Shared 3DGS reconstruction and evaluation settings.

Component	Setting
Images	$4\times$ downsampled inputs; Mip-NeRF 360 uses `images_4`; no additional runtime resize
Optimization	shared initialization checkpoint at 1K; round models advance by 300 iterations per acquisition (1.3K–5.8K); final 20-view reconstruction at 30K
SH / background	maximum SH degree 3; fixed black background
Loss	$0.8\,L_1 + 0.2(1 - \mathrm{SSIM})$; depth branch inactive
Position LR	$1.6\times10^{-4} \rightarrow 1.6\times10^{-6}$, delay multiplier 0.01, max steps 30K
Other LRs	feature 2.5×10^{-3}; opacity 2.5×10^{-2}; scale 5×10^{-3}; rotation 1×10^{-3}
Densification	iterations 501–14999, every 100 iterations, gradient threshold 2×10^{-4}
Opacity reset	every 3000 iterations
Evaluation	iteration 30K, held-out test cameras only, full RGB, black background, no crop/mask/extra resize
Metrics	PSNR from all-pixel RGB MSE; `skimage` SSIM; LPIPS-VGG (`lpipsPyTorch`, version 0.1)

A.3 Controlled Spatial-Indexing Diagnostic

The controlled spatial-indexing experiment isolates the mapping from the current reconstructed Gaussians to scoring units. Both variants use the same retained-source Gaussian prefilter, temporary camera-facing normal construction, front-facing threshold, fallback behavior, KNN-30 normal estimation, historical and candidate visibility rules, directional matrix objective, Adaptive Top-K, the Pose-Only Discovery mechanism, deterministic tie-breaking, reconstruction backend, and evaluation protocol. The only intentional change is the indexing rule: the Gaussian-indexed condition keeps each retained Gaussian center as a scoring unit, whereas the world-space condition groups the same retained Gaussians using the fixed $h = 0.25$ world-space grid and uses opacity-weighted cell representatives.

The experiment covers all nine Mip-NeRF 360 scenes and both protocol seeds. It follows the same information-access boundary as the main evaluation; the main benchmark uses the complete WSLR implementation.

Table A4: Controlled spatial-indexing effects on the full nine-scene Mip-NeRF 360 benchmark. Differences are fixed world-space cells minus Gaussian-indexed scoring units.

Metric	Mean effect	95% CI
ΔPSNR $\uparrow$	$+0.46434$	$[0.00638, 0.97109]$
ΔSSIM $\uparrow$	$+0.012770$	$[-0.003009, 0.030155]$
$-\Delta$LPIPS $\uparrow$	$+0.011128$	$[0.000586, 0.021723]$

The PSNR win/tie/loss count over the 18 paired scene–seed cells is $12/0/6$. Averaging the two seeds within each scene, the PSNR differences (world minus Gaussian) are Bicycle -0.367, Flowers

$+0.258$, Garden $+0.129$, Treehill $+0.208$, Bonsai $+1.499$, Counter $+0.791$, Kitchen $+0.317$, Room $+1.846$, and Stump -0.502 dB. Seven of the nine scene means are positive. Because all non-indexing components are matched, this experiment isolates spatial indexing as an empirical design variable in sequential 3DGS view selection.

A.4 Core State-Ablation Variants

Matched protocol and shared decision pipeline. The state-construction ablation compares three matched variants: Gaussian-indexed matrix, world-space scalar, and world-space matrix. In every variant, the view-selection state is rebuilt from the current acquired-RGB-trained 3DGS after each acquisition; no local matrix or scalar state is transferred across rounds. The variants use the same four initial views, ordered candidate pools, 16 sequential acquisitions, reconstruction backend, round-wise training schedule, final 30K reconstruction at 20 views, held-out evaluation split, and PSNR/SSIM/LPIPS implementations. They also use the same Adaptive Top-K aggregation, the Pose-Only Discovery mechanism, and deterministic repair-utility–novelty–frame-ID ordering. Selection uses acquired RGB and the current reconstruction together with acquired and candidate poses/calibration; it receives no candidate RGB, held-out RGB, depth, ground-truth geometry, or all-view geometry.

Table A5: Definitions of the three core state-construction variants. The state-builder column is included because the Gaussian-indexed and world-space matrix variants differ in more than spatial indexing.

Variant	Scoring units	Local state	State builder	Local gain
Gaussian-indexed	Retained Gaussian centers	Matrix	Gaussian builder: geometry-only acquired-view prefilter; KNN-30 primitive-center normals	Marginal log-determinant in Equation (3)
World-space scalar	World cells, $\lfloor \boldsymbol{x}/0.25 \rfloor$	Scalar	World-cell builder: temporary camera-facing normals, $q \geq 0.1$ filtering, opacity-weighted representatives, and fallback	$\log(c_i + w_i(v) + 0.01) - \log(c_i + 0.01)$
World-space matrix	World cells, $\lfloor \boldsymbol{x}/0.25 \rfloor$	Matrix	The same world-cell builder as the scalar variant	Marginal log-determinant in Equation (3)

Gaussian-indexed matrix. The Gaussian builder first retains primitives with finite centers, scales, and opacities, strictly positive scale components, and opacity at least 0.05. It then requires geometry-only frustum/z-buffer visibility from at least one acquired camera; this builder-stage test does not apply the $q \geq 0.1$ front-facing threshold. The retained Gaussian centers are the scoring units. Their normals are estimated from KNN-30 neighborhoods of those centers and oriented using their geometry-visible acquired-camera memberships. The builder terminates if fewer than three Gaussians survive. Downstream, the selector uses the matrix contribution in Equation (1), the regularized directional state in Equation (2) with $\lambda = 0.01$, and the marginal log-determinant gain in Equation (3), clamped to be nonnegative.

Shared world-cell builder. Both world-space variants use exactly the same builder, whose geometric details are given in Appendix A.5. It first applies the same finite-value, positive-scale, and opacity-0.05 checks. Because the input Gaussians do not carry geometric normals, the builder assigns each Gaussian a temporary normal pointing toward the mean acquired-camera center and applies the shared frustum/z-buffer test together with the $q \geq 0.1$ front-facing threshold. If no Gaussian survives but at least one passes the finite/scale/opacity checks, the fallback retains all such valid Gaussians. Surviving centers are indexed componentwise by $\lfloor \boldsymbol{x}/0.25 \rfloor$ in the native world coordinate system, and each occupied cell is represented by the opacity-weighted mean of its member centers. Final normals are estimated from KNN-30 neighborhoods of the cell representatives and, when recorded support cameras are nonempty, oriented toward the aggregate direction of those acquired cameras. Consequently, the builder-stage visibility and fallback also affect the support memberships used for normal orientation.

World-space scalar and matrix states. For the scalar variant, let $w_i(j)$ be the visibility-gated facing weight of acquired camera j at world cell i. Its historical state is exactly $c_i = \sum_{j \in \mathcal{S}_t} w_i(j)$, and its candidate-local gain is $\log(c_i + w_i(v) + 0.01) - \log(c_i + 0.01)$, clamped to be nonnegative. The world-space matrix variant uses the same world-cell builder, representatives, normals, and support

memberships, but replaces this scalar state and gain with the directional matrix in Equations (1) and (2) and the marginal log-determinant gain in Equation (3). The implementation likewise clamps the matrix gain to be nonnegative. The world-space matrix versus world-space scalar pair is therefore the cleaner comparison of local-state form.

Comparison scope. The Gaussian-indexed and world-space matrix variants share the downstream matrix objective, candidate scoring, aggregation, the Pose-Only Discovery mechanism, deterministic decision pipeline, and acquisition/evaluation protocol, while each uses its corresponding state constructor. Table A6 therefore compares the complete state-construction alternatives. The controlled experiment in Appendix A.3 separately isolates the spatial-indexing choice, while the world-space scalar–matrix pair provides the cleaner comparison of local-state form within the same world-cell builder.

Local-state representation. The preceding experiment isolates where repair evidence is indexed. We next examine what is stored at each world-space scoring unit. The scalar and matrix variants use the same world-space builder and differ only in the local support state, while the Gaussian-indexed matrix row provides a reference for the alternative state construction.

Table A6: **State-construction and local-state comparison** on the full nine-scene Mip-NeRF 360 benchmark (9 scenes $\times$ 2 seeds). Adaptive Top-K and the Pose-Only Discovery mechanism are fixed in every row. The two world-space rows share the same state builder.

Scoring units	Local state	PSNR ↑	SSIM ↑	LPIPS ↓
Gaussian primitives	Matrix	15.9035	0.373370	0.474405
World-space cells	Scalar	<u>16.0585</u>	**0.374018**	**0.471734**
World-space cells	**Matrix**	**16.2235**	<u>0.373918</u>	<u>0.473768</u>

Both world-space variants achieve higher PSNR than the Gaussian-indexed reference. More importantly, within the shared world-space builder, replacing the scalar support state with the directional matrix increases PSNR from 16.0585 to 16.2235 dB, a gain of 0.1650 dB. The scalar variant is slightly better in SSIM and LPIPS, so the matrix formulation is not uniformly superior across image metrics.

A scalar records the accumulated amount of local support, whereas the 3×3 matrix retains how that support is distributed across viewing directions. Consequently, two cells with similar total support can respond differently to a candidate view depending on whether it complements a weakly observed direction or repeats an already supported one. This matrix structure allows the marginal log-determinant objective to distinguish complementary directional support from repeated support. We therefore retain the matrix formulation because it matches the directional structure of the repair problem, not because it dominates every image metric.

A.5 Cell Construction and Coarse Visibility

Repair frame and Gaussian filter. At dynamic round t, the Repair state is rebuilt from the current 3DGS PLY trained using the acquired images. Gaussian centers and camera centers share the native per-scene COLMAP (Schonberger & Frahm, 2016)/3DGS (Kerbl et al., 2023) world frame; no scene normalization is applied to Repair coordinates. Let $\boldsymbol{\mu}_g \in \mathbb{R}^3$ denote the current Gaussian center, and let $\boldsymbol{s}_g \in \mathbb{R}^3$ and $o_g \in [0, 1]$ denote the activated scale and opacity used by the current 3DGS. Before voxelization, we require all three quantities to be finite, every component of $\boldsymbol{s}_g$ to be strictly positive, and $o_g \geq o_{\min} = 0.05$. We additionally retain only Gaussians visible from at least one acquired camera under the coarse visibility test below. For this pre-filter, the z-buffer is built from the otherwise valid Gaussian centers, and each temporary normal points from the Gaussian center toward the mean acquired-camera center. If this visibility test rejects every otherwise valid Gaussian, the implementation falls back to retaining the otherwise valid set.

World-space voxel grid and representative point. The voxel grid has fixed origin $\boldsymbol{o} = \mathbf{0}$ and width $h = 0.25$ in native scene-coordinate units. Each retained Gaussian is assigned componentwise by

$$\kappa_h(\boldsymbol{\mu}_g) = \left\lfloor \frac{\boldsymbol{\mu}_g}{h} \right\rfloor \in \mathbb{Z}^3. \tag{A1}$$

There is no clamping or coordinate offset, and empty cells are omitted. Let $\mathcal{G}_{t,\ell}$ denote the retained Gaussians assigned to occupied cell ℓ. Its representative point is the activated-opacity-weighted mean

$$\boldsymbol{x}_{t,\ell} = \frac{\sum_{g\in\mathcal{G}_{t,\ell}} o_g \boldsymbol{\mu}_g}{\sum_{g\in\mathcal{G}_{t,\ell}} o_g}. \tag{A2}$$

Thus, a singleton cell is represented by its single Gaussian center. Although the implementation records member counts and source-camera metadata, neither count weights the subsequent directional contribution.

Collecting the occupied coordinates in $\mathcal{L}_t$, the rebuilt view-selection state is

$$\mathcal{C}_t = \{c_{t,\ell} : \ell \in \mathcal{L}_t\}, \tag{A3}$$

where each $c_{t,\ell}$ contains the representative geometry and source-camera metadata required for normal orientation. The state is rebuilt after every reconstruction update; no Gaussian identity is transferred across rounds.

Representative normal and orientation. Throughout this subsection, $\mathrm{unit}(\cdot)$ denotes the implementation's guarded vector normalization; the corresponding 10^{-12} numerical guard is reported in Table A7. Normals are estimated from the full set of current cell representatives. Let $\mathcal{N}_{t,\ell}$ be the Euclidean KNN set returned for $\boldsymbol{x}_{t,\ell}$, including the query point, with

$$K = \min(\max(K_n, 3), |\mathcal{L}_t|), \qquad K_n = 30. \tag{A4}$$

Writing $\bar{\boldsymbol{x}}_{t,\ell}$ for the neighborhood mean, the implemented sample covariance is

$$\boldsymbol{\Sigma}_{t,\ell} = \frac{1}{K-1} \sum_{\boldsymbol{y}\in\mathcal{N}_{t,\ell}} (\boldsymbol{y} - \bar{\boldsymbol{x}}_{t,\ell})(\boldsymbol{y} - \bar{\boldsymbol{x}}_{t,\ell})^{\mathsf{T}}. \tag{A5}$$

The unoriented normal is the unit eigenvector associated with the smallest eigenvalue of $\boldsymbol{\Sigma}_{t,\ell}$. Let $\mathcal{S}_{t,\ell}$ be the deduplicated union of acquired cameras from which member Gaussians were visible during the state pre-filter, and define

$$\bar{\boldsymbol{r}}_{t,\ell} = \mathrm{unit}\left(\sum_{j\in\mathcal{S}_{t,\ell}} \mathrm{unit}(\boldsymbol{C}_j - \boldsymbol{x}_{t,\ell})\right). \tag{A6}$$

The PCA normal is multiplied by -1 exactly when its dot product with $\bar{\boldsymbol{r}}_{t,\ell}$ is negative. The same resulting $\boldsymbol{n}_{t,\ell}$ is used for historical and candidate cameras in round t. The implementation requires at least three global cell representatives and applies no additional eigen-gap or rank-degeneracy rule.

Coarse projection and visibility. Visibility uses a deterministic point z-buffer at an image downscale of four, with image dimensions and intrinsic parameters scaled consistently; it does not use rendered depth. Let $\boldsymbol{f}_j$ denote the camera forward axis. The implementation reconstructs a roll-free basis $\boldsymbol{a}_j, \boldsymbol{b}_j$ from $\boldsymbol{f}_j$ and world-up $(0,0,1)^{\mathsf{T}}$, switching the reference to $(0,1,0)^{\mathsf{T}}$ when the absolute forward/up dot product exceeds 0.98. For $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{C}_j$,

$$z_j(\boldsymbol{x}) = \boldsymbol{f}_j^{\mathsf{T}} \boldsymbol{y}, \tag{A7}$$

$$u_j(\boldsymbol{x}) = f_{x,j} \frac{\boldsymbol{a}_j^{\mathsf{T}} \boldsymbol{y}}{\mathrm{safe}(z_j)} + c_{x,j}, \tag{A8}$$

$$v_j(\boldsymbol{x}) = f_{y,j} \frac{\boldsymbol{b}_j^{\mathsf{T}} \boldsymbol{y}}{\mathrm{safe}(z_j)} + c_{y,j}, \tag{A9}$$

where $\mathrm{safe}(z) = 10^{-9}$ when $|z| < 10^{-9}$ and equals z otherwise. Each point with positive depth and continuous image coordinates inside the image splats its axial depth by a minimum reduction into a radius-3 pixel disk centered at its componentwise rounded pixel. Pixel rounding follows round-to-nearest with exact half-integers rounded to the even integer.

Let D_j be the resulting float64 depth buffer initialized to $+\infty$. Writing $\mathrm{rint}_{\mathrm{even}}$ for this rounding rule, the queried pixel is

$$\hat{u} = \mathrm{clip}(\mathrm{rint}_{\mathrm{even}}(u), 0, W_j - 1), \tag{A10}$$

$$\hat{v} = \mathrm{clip}(\mathrm{rint}_{\mathrm{even}}(v), 0, H_j - 1). \tag{A11}$$

The geometric visibility predicate is

$$\widetilde{Z}_{t,\ell}(j) = \mathbf{1}[z > 0]\,\mathbf{1}[0 \le u < W_j]\,\mathbf{1}[0 \le v < H_j] \times \mathbf{1}[\neg(\mathrm{finite}(D_j[\hat{v},\hat{u}]) \wedge z > D_j[\hat{v},\hat{u}] + \tau_z)], \qquad \tau_z = 0.01. \tag{A12}$$

Here τ_z is an absolute camera-axis depth tolerance in native scene coordinates. The state pre-filter builds D_j from Gaussian centers, whereas the selector rebuilds it from current cell representatives.

A.6 Origin and Scope of the Directional Matrix

Bearing-derived directional form. Let $\boldsymbol{x}$ be a local scene point, $\boldsymbol{c}$ a fixed camera center, and $d = \|\boldsymbol{c} - \boldsymbol{x}\|_2 > 0$. Define the unit direction $\boldsymbol{r} = (\boldsymbol{c} - \boldsymbol{x})/d$. Differentiating the normalization with respect to the scene point gives

$$\boldsymbol{B}(\boldsymbol{x}) = \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{x}} = -\frac{1}{d}\left(\boldsymbol{I} - \boldsymbol{r}\boldsymbol{r}^{\mathsf{T}}\right). \tag{A13}$$

The projector annihilates motion along the ray and retains perturbations in its orthogonal plane. Under an idealized isotropic bearing-noise model with variance σ_b^2 in that tangent plane, the corresponding local Gauss–Newton matrix is

$$\frac{1}{\sigma_b^2}\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B} = \frac{1}{\sigma_b^2 d^2}\left(\boldsymbol{I} - \boldsymbol{r}\boldsymbol{r}^{\mathsf{T}}\right). \tag{A14}$$

This follows because the directional projector is symmetric and idempotent. The sign convention of the ray does not affect this directional matrix because $\boldsymbol{I} - (-\boldsymbol{r})(-\boldsymbol{r})^{\mathsf{T}} = \boldsymbol{I} - \boldsymbol{r}\boldsymbol{r}^{\mathsf{T}}$. This derivation motivates the directional form used by the acquisition surrogate. DC-DLO omits the range factor d^{-2} and noise variance σ_b^2 so that the directional surrogate emphasizes viewing-direction complementarity; retaining d^{-2} would systematically increase the weight of nearby cameras even when their viewing directions largely repeat existing support.

Directional weighting and support. For a historical or candidate camera j, define

$$\boldsymbol{r}_{t,\ell j} = \frac{\boldsymbol{C}_j - \boldsymbol{x}_{t,\ell}}{\max(\|\boldsymbol{C}_j - \boldsymbol{x}_{t,\ell}\|_2, 10^{-12})}, \tag{A15}$$

$$Q_{t,\ell}(j) = \max\left(0, \boldsymbol{n}_{t,\ell}^{\mathsf{T}}\boldsymbol{r}_{t,\ell j}\right), \tag{A16}$$

$$Z_{t,\ell}(j) = \widetilde{Z}_{t,\ell}(j)\mathbf{1}[Q_{t,\ell}(j) \ge q_{\min}], \qquad q_{\min} = 0.1, \tag{A17}$$

$$\alpha_{t,\ell}(j) = Z_{t,\ell}(j)Q_{t,\ell}(j), \tag{A18}$$

$$\boldsymbol{J}_{t,\ell}(j) = \alpha_{t,\ell}(j)\left(\boldsymbol{I} - \boldsymbol{r}_{t,\ell j}\boldsymbol{r}_{t,\ell j}^{\mathsf{T}}\right), \tag{A19}$$

$$\boldsymbol{H}_{t,\ell} = \lambda\boldsymbol{I} + \sum_{j \in \mathcal{S}_t} \boldsymbol{J}_{t,\ell}(j), \qquad \lambda = 0.01. \tag{A20}$$

The visibility/facing factor α is an acquisition weighting rule and is not derived from the idealized bearing-noise model above. Camera count, Gaussian count, range, photometric residuals, uncertainty, and measurement covariance do not appear as multiplicative weights in this directional block. Image resolution affects the coarse visibility test but does not multiply the directional contribution itself.

The candidate gain is

$$\Delta_{t,\ell}(v) = \log\det(\boldsymbol{H}_{t,\ell} + \boldsymbol{J}_{t,\ell}(v)) - \log\det\boldsymbol{H}_{t,\ell}. \tag{A21}$$

Both log determinants are evaluated by Cholesky factorization as twice the sum of the logarithms of the Cholesky diagonal. Values below -10^{-10} trigger a numerical error; otherwise the computed gain is clamped to zero from below. Since $\boldsymbol{H}_{t,\ell} \succ \mathbf{0}$ and $\boldsymbol{J}_{t,\ell}(v) \succeq \mathbf{0}$, the exact gain also satisfies

$$\Delta_{t,\ell}(v) = \log\det\left(\boldsymbol{I} + \boldsymbol{H}_{t,\ell}^{-1/2}\boldsymbol{J}_{t,\ell}(v)\boldsymbol{H}_{t,\ell}^{-1/2}\right) \ge 0. \tag{A22}$$

Omitting inverse-squared range is a modeling choice rather than removal of a common constant: observations at different distances can therefore change their relative weighting compared with a full bearing-information model. The resulting local gain is intended to measure directional complementarity for the current modeled geometry.

A.7 Adaptive Support Budget

All candidates in the full remaining pool $\mathcal{R}_t$ are scored before the feasibility mask of the Pose-Only Discovery mechanism is applied. With $\varepsilon_\Delta = 10^{-12}$,

$$N_t(v) = \sum_{\ell \in \mathcal{L}_t} \mathbf{1}[\Delta_{t,\ell}(v) > \varepsilon_\Delta]. \tag{A23}$$

Let $\mathrm{rint}_{\mathrm{even}}$ denote round-to-nearest with exact half-integers rounded to the even integer. The shared support width is

$$K_t = \max(1, \mathrm{int}[\mathrm{rint}_{\mathrm{even}}(\mathrm{median}_{v \in \mathcal{R}_t} N_t(v))]) \,. \tag{A24}$$

For each candidate, TopK_{K_t} selects the K_t largest entries of the complete nonnegative clamped gain vector, including positive entries not exceeding ε_Δ, and

$$R_t(v) = \frac{1}{K_t} \sum_{\delta \in \mathrm{TopK}_{K_t}(\{\Delta_{t,\ell}(v)\}_{\ell \in \mathcal{L}_t})} \delta. \tag{A25}$$

Thus, ε_Δ controls the support counts and hence K_t, but does not hard-zero sub-threshold positive gains. Because K_t is shared by all candidates in a round, the factor $1/K_t$ does not alter within-round ranking. If $N_t(v) = 0$ for every candidate, then $K_t = 1$ and every repair utility is at most 10^{-12}, so the novelty-only fallback in Appendix A.8 is used. Computing K_t before applying the Pose-Only Discovery mechanism prevents it from altering the Repair aggregation rule.

A.8 Pose Novelty, Feasibility Adaptation, and Deterministic Selection

Pose normalization and distance. The fixed normalization pool is

$$\mathcal{P} = \mathcal{S}_0 \cup \mathcal{R}_0, \tag{A26}$$

i.e., the initial acquired poses plus every selectable candidate pose. This is exactly the protocol's acquisition/training-view pose pool and excludes all held-out evaluation poses. Let $M = |\mathcal{P}|$. Camera centers are normalized once by

$$\boldsymbol{m}_c = \mathrm{median}^{\mathrm{coordinate}}_{j \in \mathcal{P}}(\boldsymbol{C}_j), \tag{A27}$$

$$s_c = \sqrt{\frac{1}{M} \sum_{j \in \mathcal{P}} \|\boldsymbol{C}_j - \boldsymbol{m}_c\|_2^2}, \tag{A28}$$

$$\widetilde{\boldsymbol{C}}_j = (\boldsymbol{C}_j - \boldsymbol{m}_c)/s_c. \tag{A29}$$

There is no additive normalization epsilon; the implementation aborts if $s_c < 10^{-12}$.

Let $\boldsymbol{R}_j$ denote COLMAP's world-to-camera rotation. The unit optical axis is the viewing $+Z$ row expressed in world coordinates,

$$\boldsymbol{d}_j = \mathrm{unit}(\boldsymbol{R}_j[2, :]). \tag{A30}$$

This deliberately ignores in-plane roll: the pose metric measures coarse viewpoint diversity through camera-center and optical-axis separation rather than full SO(3) pose distance. Using the untuned symmetric default $w_c = w_d = 1$, we write the pose-distance rule as

$$d_{\mathrm{pose}}(a, b) = \sqrt{w_c \|\widetilde{\boldsymbol{C}}_a - \widetilde{\boldsymbol{C}}_b\|_2^2 + w_d \|\boldsymbol{d}_a - \boldsymbol{d}_b\|_2^2}, \qquad w_c = w_d = 1. \tag{A31}$$

The orientation term is chordal optical-axis separation, $\|\boldsymbol{d}_a - \boldsymbol{d}_b\|_2 = 2\sin(\theta/2)$; it is not an angular distance. We use the untuned symmetric default $w_c = w_d = 1$ throughout the reported experiments; Table A9 characterizes the sensitivity to w_d.

Novelty and normalized average rank. For the acquired set $\mathcal{S}_t$ and remaining pool $\mathcal{R}_t$,

$$\nu_t(v) = \min_{j \in \mathcal{S}_t} d_{\mathrm{pose}}(v, j), \qquad v \in \mathcal{R}_t. \tag{A32}$$

Let $M_t = |\mathcal{R}_t|$ and let $\mathrm{rank}_{\mathrm{avg}}$ be the ascending one-based rank that assigns tied values the average of their occupied ordinal positions. The normalized novelty rank is

$$p_t(v) = \begin{cases} 1, & M_t = 1, \\ \dfrac{\mathrm{rank}_{\mathrm{avg}}(\nu_t(v)) - 1}{M_t - 1}, & M_t > 1. \end{cases} \tag{A33}$$

Thus, a top tie need not attain one; if all novelties tie and $M_t > 1$, every candidate has $p_t(v) = 1/2$.

The current mean novelty and fixed initial reference are

$$\bar{\nu}_t = \frac{1}{|\mathcal{R}_t|} \sum_{v \in \mathcal{R}_t} \nu_t(v), \tag{A34}$$

$$\bar{\nu}_0 = \frac{1}{|\mathcal{R}_0|} \sum_{v \in \mathcal{R}_0} \min_{j \in \mathcal{S}_0} d_{\mathrm{pose}}(v, j). \tag{A35}$$

The implementation uses the fixed numerical stabilizer $\varepsilon_\nu = 10^{-12}$ and directly recomputes

$$\rho_t = \mathrm{clip}\left(1 - \frac{\bar{\nu}_t}{\bar{\nu}_0 + \varepsilon_\nu}, 0, 1\right), \tag{A36}$$

$$g_t = \frac{1 - \rho_t}{2}, \tag{A37}$$

$$\mathcal{E}_t = \{v \in \mathcal{R}_t : p_t(v) \geq g_t\}. \tag{A38}$$

The stabilizer prevents division by zero when $\bar{\nu}_0 = 0$ and is fixed across scenes and rounds. There is no running maximum or stored previous relaxation state. At the first dynamic round, $\rho_0 = \varepsilon_\nu/(\bar{\nu}_0 + \varepsilon_\nu) > 0$ whenever $\bar{\nu}_0 > 0$. Neither ρ_t nor $\mathcal{E}_t$ is asserted to be monotone across rounds.

For every nonempty remaining pool, $g_t \leq 1/2$ and the highest average-rank group has normalized rank at least $1/2$; hence $\mathcal{E}_t \neq \varnothing$. For a singleton pool, $p_t(v) = 1$ explicitly. For numerical robustness, the implementation also guards against an empty feasibility mask by marking the first candidate attaining maximum raw novelty as feasible. Under the finite rank conditions above, the analytical feasible set is already nonempty.

Deterministic selection and reveal. Let $\epsilon_0 = 10^{-12}$. If at least one candidate in the full remaining pool $\mathcal{R}_t$ has $R_t(v) > \epsilon_0$, feasible candidates are ordered lexicographically by

$$\big(-R_t(v), -p_t(v), \mathrm{int64}(v)\big), \qquad v \in \mathcal{E}_t. \tag{A39}$$

Thus, Repair utility is primary, normalized novelty rank is secondary, and the smaller integer frame identifier is tertiary. If every candidate in $\mathcal{R}_t$ satisfies $R_t(v) \leq \epsilon_0$, score differences are ignored and feasible candidates are ordered by

$$\big(-p_t(v), \mathrm{int64}(v)\big), \qquad v \in \mathcal{E}_t. \tag{A40}$$

Only after this choice is committed is the selected RGB image linked into the acquired training set. The next reconstruction is then optimized using the updated acquired set, after which the world-space repair state is rebuilt from the updated 3DGS PLY.

A.9 Fixed Selector Settings

Table A7 summarizes the fixed method settings and numerical conventions of the reported selector. The same $h = 0.25$ and $w_c = w_d = 1$ defaults are used across scenes and datasets. Rule-derived quantities K_t, $p_t(v)$, and ρ_t are recomputed during acquisition. Appendix Table A9 reports sensitivity to h, the omitted range factor, and w_d.

Table A7: Fixed selector settings and numerical conventions used by DC-DLO. Training and evaluation settings are specified separately in the experimental protocol.

Setting	Symbol	Value	Role
Voxel width	h	0.25	Native scene-coordinate spatial scale
Grid origin	$\boldsymbol{o}$	$\mathbf{0}$	Fixed world-space indexing origin
Opacity threshold	$o_{\min}$	0.05	Filters near-transparent Gaussians
Normal neighbors	K_n	30	KNN/PCA representative normal
Front-facing threshold	$q_{\min}$	0.1	Rejects grazing support
Isotropic regularizer	λ	0.01	Makes each local support matrix positive definite
Coarse-depth tolerance	τ_z	0.01	Absolute axial-depth tolerance
Image downscale	–	4	Coarse projection resolution
Z-buffer splat radius	–	3 px	Point-depth buffer construction
Gain-count threshold	ε_Δ	10^{-12}	Defines cells counted in $N_t(v)$ and hence K_t
Adaptive-K rounding	–	ties-to-even	Rounds the median support count
Pose-center weight	w_c	1.0	Translation term in pose distance
Pose-axis weight	w_d	1.0	Optical-axis term in pose distance
Novelty-depletion stabilizer	ε_ν	10^{-12}	Prevents zero denominator in ρ_t
Pose-scale guard	–	10^{-12}	Rejects degenerate normalization scale
Unit-vector guard	–	10^{-12}	Stabilizes vector normalization
Projection safe depth	–	10^{-9}	Stabilizes image projection division
Negative-gain tolerance	–	-10^{-10}	Errors below this value; otherwise clamp to zero
Near-zero repair threshold	ϵ_0	10^{-12}	Activates novelty-only fallback
Novelty-rank ties	–	average rank	Deterministic normalized rank convention
Tie-breaking order	–	$R\downarrow, p\downarrow, \mathrm{id}\uparrow$	Primary selection ordering

B Additional Diagnostics

B.1 View-Selection State Indexing

Table A8: View-selection state indexing and information-access boundaries. "Update" describes how the selection state changes between rounds; candidate input is information required before acquisition. The table does not assume that other methods preserve primitive identities.

Method	State indexing units	State evolution	Candidate input	Separate pose mechanism
FisherRF	Renderer parameters	Re-estimated	Pose	No
GauSS-MI	Gaussian primitives	Acquired-RGB update	Pose	No
OUGS	Gaussian primitives	Online EMA	Pose + mask	No
PRIMU	Gaussian primitives	Projected statistics	Pose	No
GAVIS	Gaussian primitives	Visibility field	Pose	No
COVER	Primitive coverage	Accumulated	Pose	No
$\mathbb{R}^3$-RECON	World voxels	Online statistics	Pose	No
DC-DLO	Fixed world-space cells	Rebuilt	Pose	**Yes**

B.2 Sensitivity of Fixed Design Choices

We perform a sensitivity diagnostic on Bicycle, Garden, and Room at seed 0. All variants use the same initial views, candidate pools, information-access boundary, 20-view budget, and final 30K reconstruction. The reported default is $h = 0.25$, direction-only support without d^{-2}, and $w_c = w_d = 1$. These experiments characterize the fixed design choices while leaving the main configuration unchanged.

Table A9: Three-scene seed-0 sensitivity of fixed design choices. "Default" denotes the setting used by all reported main experiments.

Factor	Setting	PSNR ↑	SSIM ↑	LPIPS ↓
Voxel width	$h = 0.125$	16.3625	0.4063	0.4492
	$h = 0.25$ (default)	16.4028	0.3956	0.4586
	$h = 0.5$	15.6480	0.3673	0.4865
Range factor	direction-only (default)	16.4028	0.3956	0.4586
	d^{-2}	16.1983	0.3890	0.4666
Pose-axis weight	$w_d = 0.5$	16.5233	0.4090	0.4540
	$w_d = 1$ (default)	16.4028	0.3956	0.4586
	$w_d = 2$	16.7901	0.4110	0.4450

The voxel-width response is asymmetric: reducing the default from 0.25 to 0.125 changes PSNR by only -0.040 dB, whereas increasing it to 0.5 reduces PSNR by 0.755 dB and degrades SSIM/LPIPS. Thus, the method is not scale-free and overly coarse aggregation is harmful, although the default and a finer grid are close in PSNR on this diagnostic. Restoring the inverse-squared range factor decreases PSNR by 0.204 dB, SSIM by 0.0066, and worsens LPIPS by 0.0080, supporting the implemented direction-only choice on this test. Finally, changing w_d from the default changes macro PSNR by $+0.121$ dB at 0.5 and $+0.387$ dB at 2. The pose weighting is therefore moderately sensitive; the main experiments retain the original untuned symmetric default $w_d = 1$.

B.3 Static Geometry Evaluation Protocol

We evaluate static geometry using a method-independent full-view 3DGS proxy rather than ground-truth geometry. For each scene, an all-available-view 30K model and every final-20-view method model are rendered at the same held-out cameras and fused in the common COLMAP frame, with no post-hoc alignment. Per-pixel depth is the expected camera-z value, $\sum_i \alpha_i z_i / \max(\sum_i \alpha_i, 10^{-8})$. We retain finite depths with $0.01 \leq z$ and alpha at least 0.5. The shared maximum depth is set to 1.2 times the 99th percentile of valid held-out reference depths. Back-projection uses the FoV intrinsics with principal point $(W/2, H/2)$ and the stored COLMAP camera transform. We then apply a reference-derived axis-wise 1st–99th percentile spatial ROI expanded by $0.02d$, where d is the reference bounding-box diagonal before margin expansion, and voxelize all surfaces with width $0.0025d$. Global precision and recall use the threshold $\tau = 0.005d$; their harmonic mean is the Global F-score. Chamfer L1 is $\frac{1}{2}$ times the sum of the two directional mean nearest-neighbor distances and remains in the shared COLMAP world-coordinate units.

The local-repair mask contains the 20% of reference points farthest from the common four-view initial reconstruction, measured by nearest-surface distance. The discovery mask contains reference points invisible from all four initial cameras; when that set contains fewer than $\max(5000, \lceil 0.02|\mathcal{R}| \rceil)$ points, the preregistered fallback uses points visible from at most one initial camera. Both masks are determined without using any final method output. Local F-score uses the same $\tau = 0.005d$ precision/recall rule inside the repair mask, while Discovery Recall measures the fraction of discovery-mask reference points lying within τ of the predicted surface. Metrics are computed per scene–seed cell and macro-averaged equally over the 18 cells.

Table A10: Supporting geometry diagnostic on Mip-NeRF 360. The reference is an all-available-view 30K 3DGS depth-fusion proxy rather than ground-truth geometry. COVER-GT† is the privileged oracle-style geometry selector used only in this diagnostic.

Method	Global F ↑	Chamfer L1 ↓	Local Repair F ↑	Discovery Recall ↑
Random	0.3563	**1.2357**	0.0740	0.3123
FisherRF	0.3601	1.3413	**0.1095**	<u>0.3366</u>
GauSS-MI	0.3532	1.3939	0.0647	0.3072
COVER-GT†	0.3541	<u>1.2773</u>	<u>0.1033</u>	0.3282
DC-DLO	**0.3765**	1.2920	0.0909	**0.3428**

DC-DLO has the highest Global F-score and Discovery Recall. Random has the lowest Chamfer distance, and FisherRF has the highest Local Repair F-score.

B.4 Spatial Attribution of the Full Policy

For each held-out view, we define a fixed mask containing the 10% of pixels with the largest RGB residual under the reconstruction shared by all configurations before acquisition. The mask is held fixed across methods and used only for post-hoc evaluation. We compare DC-DLO with Raw DLO, which retains world-space rebuilding and local marginal log-determinant gain but disables both Adaptive Top-K and the Pose-Only Discovery mechanism. This comparison therefore attributes the combined policy difference, not an individual component.

Table A11: Spatial attribution of the full policy. Entries are DC-DLO minus Raw DLO PSNR in dB. Hard-region masks are fixed from the shared initial reconstruction and never exposed to either selector.

Scene	Full image	Fixed hard region
Bicycle	$+1.040$	$+1.155$
Garden	$+0.456$	$+0.731$
Room	$+1.522$	$+2.071$
Macro	$+1.006$	$+1.319$

The larger gains inside the fixed masks are consistent with the full policy placing more acquisition value on difficult support regions.

B.5 Representative Selection Cost

Selector-only latency is measured on a five-scene Extended NBV subset. Each value is the total selector time across 16 acquisition rounds divided by 16 and macro-averaged over the five scenes. Per-round dispersion is unavailable, so we report the comparison using the mean time per round.

Table A12: Mean selector-only time per acquisition round on the five-scene Extended NBV runtime subset.

Method	Time/round (s) ↓
FisherRF	17.2884
GauSS-MI	**4.5246**
COVER	7.1573
DC-DLO	4.9905

In this implementation, DC-DLO requires approximately five seconds of selector computation per round, substantially less than FisherRF and COVER, while GauSS-MI is slightly faster. The measurements are implementation-specific and characterize the computational cost of the evaluated pipeline.

B.6 NAVI Real-Query Registration

The NAVI evaluation contains four objects, two seeds, five selectors, 16 independent queries, and eight perturbations per query, yielding 5,120 trials in total and 1,024 trials per method. Query RGB is used only by the downstream registration procedure and is never available during view selection. No ground-truth depth is used in the registration pipeline.

Table A13: NAVI real-query pose recovery. Success requires rotation error $\leq 5^\circ$ and translation error $\leq 0.05D$.

Method	Success@5°/$0.05D$ ↑
Random	<u>30.47%</u>
COVER	21.29%
FisherRF	30.08%
GauSS-MI	26.76%
DC-DLO	**32.13%**

A trial succeeds when the estimated pose has rotation error at most 5° and translation error at most $0.05D$, where D is the diagonal of the object's scaled axis-aligned bounding box in mapping coordinates. All trials are retained, including PnP failures.

Table A14: Detailed NAVI evaluation. "Valid PnP" counts trials for which PnP returns a pose. Success additionally requires both pose-error thresholds to be satisfied.

Method	Trials	Valid PnP	PnP fail	Success	Rate	Median rot. (°)	Median trans./D
Random	1024	998	26	312	30.47%	1.4338	0.0746
COVER	1024	1002	22	218	21.29%	1.5770	0.0847
FisherRF	1024	1000	24	308	30.08%	1.3721	0.0688
GauSS-MI	1024	1001	23	274	26.76%	1.5194	0.0763
DC-DLO	1024	999	25	**329**	**32.13%**	**1.2857**	**0.0672**

The pooled success-rate differences of DC-DLO are $+2.05$ percentage points relative to FisherRF, $+5.37$ relative to GauSS-MI, $+1.66$ relative to Random, and $+10.84$ relative to COVER. At the object level, DC-DLO records two wins and two losses against FisherRF, GauSS-MI, and Random, and four wins against COVER. The object-cluster bootstrap intervals for the differences versus FisherRF and GauSS-MI are $[-4.69, 10.25]$ and $[-4.49, 15.23]$ percentage points, respectively. The object-cluster bootstrap intervals versus FisherRF and GauSS-MI both include zero, and the object-level comparisons are mixed. The pooled result therefore provides evidence of downstream utility without implying a uniform object-level advantage.

B.7 Qualitative Grasping Sequences

The grasping examples are downstream-only demonstrations: the grasp pipeline runs after view acquisition, and neither grasp outcome nor execution state is available to the selector. They are included to illustrate a possible task-level consequence of reconstruction quality and are not used for aggregate method ranking.

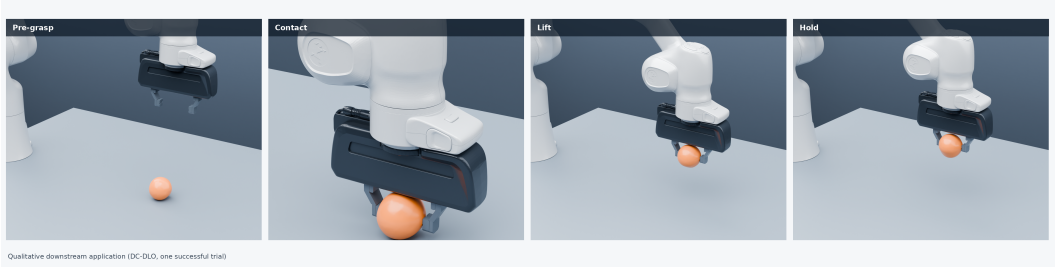

Figure A1: DC-DLO qualitative grasping sequence for one successful trial: pre-grasp, contact, lift, and stable hold.

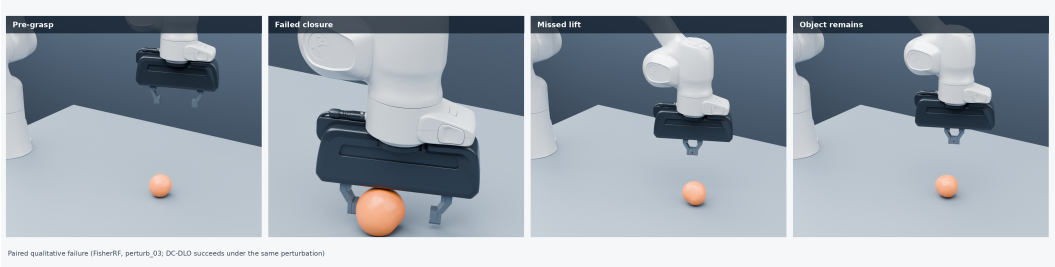

Figure A2: FisherRF qualitative failure sequence for the matched perturbation: pre-grasp, failed closure, missed lift, and the object remaining on the table. The matched DC-DLO success sequence is shown in Figure A1.

C Paired Effects and Representation Diagnostics

Paired statistics preserve scene–seed correspondence, and confidence intervals are obtained by bootstrapping scenes while retaining both protocol seeds from each sampled scene. Positive differences favor DC-DLO. The baseline, complete state-construction, and comparisons involving the Pose-Only Discovery mechanism in this section use the full nine-scene Mip-NeRF 360 benchmark. The controlled spatial-indexing experiment is reported separately in Appendix A.3.

Table A15: Paired effects of DC-DLO relative to each baseline on the full nine-scene Mip-NeRF 360 benchmark. Each entry reports the mean difference and a scene-clustered 95% bootstrap interval. PSNR wins count positive differences over the 18 scene–seed pairs.

Baseline	ΔPSNR $\uparrow$ (95% CI)	ΔSSIM $\uparrow$ (95% CI)	$-\Delta$LPIPS $\uparrow$ (95% CI)	PSNR W/T/L
Random	0.443 [0.086, 0.754]	0.0078 [−0.0071, 0.0218]	0.0066 [−0.0043, 0.0173]	13/0/5
COVER	0.806 [0.230, 1.384]	0.0345 [0.0093, 0.0601]	0.0236 [0.0060, 0.0418]	13/0/5
FisherRF	0.701 [0.238, 1.131]	0.0338 [0.0108, 0.0562]	0.0229 [0.0064, 0.0386]	14/0/4
GauSS-MI	0.959 [0.602, 1.270]	0.0381 [0.0222, 0.0540]	0.0250 [0.0148, 0.0353]	16/0/2

Paired effects of state construction and the Pose-Only Discovery mechanism. The main-paper component study evaluates WSLR and the Pose-Only Discovery mechanism jointly on the targeted three-scene subset. We additionally report full nine-scene paired comparisons for the complete state-construction alternatives and for the Pose-Only Discovery mechanism once the world-space matrix and Adaptive Top-K repair route are active.

Table A16: Full-nine-scene paired state-construction and Pose-Only Discovery effects on Mip-NeRF 360. Differences are left minus right; the LPIPS column reports $-\Delta$LPIPS so that positive values favor the left variant. Intervals use 100,000 scene-cluster bootstrap resamples retaining both seeds.

Comparison	ΔPSNR $\uparrow$ (95% CI)	ΔSSIM $\uparrow$ (95% CI)	$-\Delta$LPIPS $\uparrow$ (95% CI)	PSNR W/T/L
World matrix − Gaussian matrix	0.3200 [−0.0974, 0.7702]	0.00055 [−0.02007, 0.02238]	0.00064 [−0.01177, 0.01259]	10/0/8
Matrix world-space − Scalar world-space	0.1650 [−0.1632, 0.5186]	−0.00010 [−0.01513, 0.01665]	−0.00203 [−0.01104, 0.00800]	9/0/9
With Discovery − Without Discovery	−0.0634 [−0.3807, 0.2075]	−0.00448 [−0.01461, 0.00406]	−0.00411 [−0.01202, 0.00221]	8/0/10

Across all nine scenes and two seeds, the marginal effect of the Pose-Only Discovery mechanism after the complete WSLR+Top-K repair route is active is small in the aggregate and varies substantially with scene and candidate-pool geometry. This broader result complements the targeted component study by showing that the Pose-Only Discovery mechanism should be interpreted as a pose-diversity feasibility mechanism rather than as a universally positive repair-utility term.

Table A17: Scene-level marginal PSNR effect of the Pose-Only Discovery mechanism on the full nine-scene Mip-NeRF 360 benchmark. Values are Full minus Without-Discovery, averaged over the two protocol seeds.

Scene	ΔPSNR (dB) $\uparrow$
Bicycle	+0.467
Flowers	−0.051
Garden	+0.049
Treehill	+0.391
Bonsai	−1.066
Counter	−0.233
Kitchen	+0.040
Room	+0.314
Stump	−0.484
Macro	−0.063

C.1 Representation-Change Diagnostic

We evaluate 39 consecutive-round transitions—13 each from Bicycle, Garden, and Room. Gaussian centers are associated for this diagnostic only by mutual-nearest spatial matching at threshold 0.125, half the edge of the diagnostic voxel grid. We report the fractions of current and previous primitives that remain unmatched under this rule, together with the fraction of previous primitives that receive a spatial match and the occupied-support IoU at voxel width 0.25. These are spatial rematching statistics rather than tracked Gaussian birth/death identities.

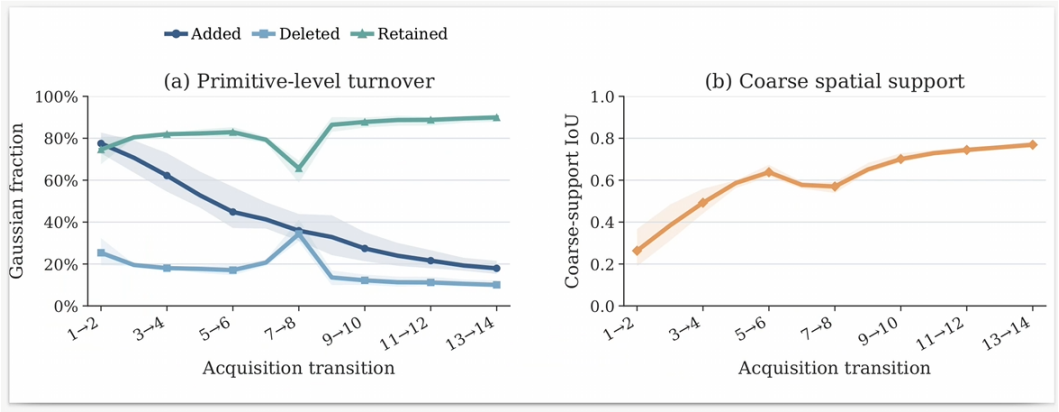

Figure A3: Spatial rematching and coarse support change across acquisition transitions. Solid lines are scene-macro means and shaded bands span the minimum and maximum across scenes.

Across the three scenes, the mean current-unmatched and previous-unmatched fractions are 40.63% and 17.04%, while the mean occupied-voxel IoU is 0.605. These measurements characterize substantial change in the spatial primitive decomposition and occupied support across acquisition rounds.

Table A18: Per-scene representation-change diagnostic. Current-unmatched, previous-unmatched, and spatially matched fractions use the fixed mutual-nearest rule; coarse IoU compares occupied voxel support.

Scene	Current unmatched (%)	Previous unmatched (%)	Spatially matched (%)	Coarse IoU
Bicycle	48.34	19.63	80.37	0.577
Garden	37.22	15.86	84.14	0.626
Room	36.34	15.64	84.36	0.611
Mean	40.63	17.04	82.96	0.605

D Primary Benchmark Per-Scene Results

Tables A19–A21 report the nine-scene Mip-NeRF 360 results after averaging the two protocol seeds within each scene. The macro-averages of the underlying 18 scene–seed cells reproduce Table 1.

Table A19: Per-scene PSNR on the primary Mip-NeRF 360 benchmark, averaged over two seeds.

Scene	Random	COVER	FisherRF	GauSS-MI	DC-DLO
Bicycle	13.291	**13.513**	13.177	11.885	12.728
Flowers	11.747	11.906	11.582	11.635	**12.388**
Garden	17.104	17.312	17.160	16.722	**17.670**
Treehill	14.328	13.967	13.766	13.375	**14.935**
Bonsai	16.615	15.101	15.420	15.883	**17.240**
Counter	16.255	16.138	15.897	15.843	**16.751**
Kitchen	18.716	17.823	18.998	19.005	**19.834**
Room	18.183	17.639	17.891	17.425	**19.055**
Stump	15.782	15.524	**15.815**	15.611	15.410

Table A20: Per-scene SSIM on the primary Mip-NeRF 360 benchmark, averaged over two seeds.

Scene	Random	COVER	FisherRF	GauSS-MI	DC-DLO
Bicycle	0.1622	**0.1660**	0.1578	0.1243	0.1331
Flowers	0.0938	0.0957	0.0920	0.0899	**0.1165**
Garden	0.3910	0.4216	0.3793	0.3830	**0.4261**
Treehill	**0.2687**	0.2290	0.2130	0.2207	0.2594
Bonsai	**0.4945**	0.3958	0.3964	0.4488	0.4894
Counter	0.4714	0.4412	0.4526	0.4325	**0.4896**
Kitchen	0.6358	0.5729	0.6359	0.6163	**0.6662**
Room	0.5905	0.5715	0.5496	0.5338	**0.6170**
Stump	**0.1874**	0.1659	0.1847	0.1730	0.1682

Table A21: Per-scene LPIPS on the primary Mip-NeRF 360 benchmark, averaged over two seeds.

Scene	Random	COVER	FisherRF	GauSS-MI	DC-DLO
Bicycle	0.5951	0.5883	**0.5868**	0.6166	0.6089
Flowers	0.6131	0.6064	0.6163	0.6238	**0.5927**
Garden	0.4245	0.4060	0.4342	0.4340	**0.4012**
Treehill	**0.5416**	0.5566	0.5630	0.5670	0.5500
Bonsai	**0.4397**	0.5072	0.5098	0.4614	0.4470
Counter	0.4379	0.4602	0.4513	0.4641	**0.4232**
Kitchen	0.3108	0.3538	0.3105	0.3174	**0.2877**
Room	0.3987	0.4177	0.4265	0.4279	**0.3754**
Stump	**0.5620**	0.5732	0.5712	0.5770	0.5779

E Additional Component Diagnostics

E.1 Matched World-Space-State–Pose-Only Discovery Study

As a matched subset of the main component ablation, we isolate a 2×2 comparison on Bicycle, Garden, and Room. The directional matrix objective and Adaptive Top-K are fixed in every configuration, while the repair-state construction and the Pose-Only Discovery mechanism are varied independently. The Gaussian-indexed baseline was newly executed under the same 4+16-view, final-30K protocol; the other three configurations reused existing final models and were re-evaluated with the same held-out evaluator. The reused evaluations reproduced their stored metrics exactly.

Table A22: Matched three-scene repair-state and Pose-Only Discovery study. The directional matrix objective and Adaptive Top-K are fixed in all configurations.

Variant	World-space state	Pose-Only Discovery	PSNR ↑	SSIM ↑	LPIPS ↓
Gaussian-indexed	No	No	16.1813	0.3920	0.4671
World-space	Yes	No	15.9835	0.3896	0.4655
Gaussian + Discovery	No	Yes	16.2621	0.3982	0.4661
DC-DLO	Yes	Yes	16.4028	0.3956	0.4586

Table A23: Factorial effects in the matched three-scene repair-state and Pose-Only Discovery study. The LPIPS column reports $-\Delta$LPIPS so that positive values are favorable.

Contrast	ΔPSNR ↑	ΔSSIM ↑	−ΔLPIPS ↑
World-space state without Discovery	-0.1978	-0.0024	$+0.0016$
Discovery with Gaussian-indexed state	$+0.0808$	$+0.0062$	$+0.0010$
Discovery with world-space state	$+0.4193$	$+0.0060$	$+0.0069$
World-space state with Discovery	$+0.1407$	-0.0026	$+0.0075$
DC-DLO − Gaussian-indexed/no-Discovery	$+0.2215$	$+0.0036$	$+0.0085$

On this targeted subset, the Pose-Only Discovery mechanism is favorable under both repair states, with a larger effect under the world-space state. The world-space state alone does not improve PSNR or SSIM, whereas with the Pose-Only Discovery mechanism enabled it improves PSNR and LPIPS. These effects are not well described as independent additive gains.