RECOVERABILITY-AWARE GAUSSIAN PRUNING FOR 3D GAUSSIAN SPLATTING COMPRESSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Gaussian pruning for 3D Gaussian Splatting (3DGS) typically retains primitives according to importance or estimated removal effect at the model state. Yet pruning quality is evaluated only after the remaining Gaussians are further optimized, leaving a key factor unmodeled: how well the retained representation adapts to removed information. We call this *post-pruning recoverability*. We study recoverability through controlled counterfactual interventions. Across matched recovery trajectories, recoverability exhibits reproducible structure across recovery schedules, with more consistent cross-schedule agreement at longer recovery horizons, while freezing retained Gaussians with strong geometric responses reduces subsequent recovery. Directly using this response as a pruning criterion, however, degrades reconstruction quality, showing that counterfactual compensation evidence complements rather than replaces primitive importance. Based on this observation, we introduce Recoverability-Aware Gaussian Pruning (RAGP), a fixed-budget refinement framework for Gaussian importance criteria. RAGP uses counterfactual pruning previews to measure the response of the surviving representation to localized information removal. The resulting geometric response is whitened and fused with the host importance prior through bounded host-relative calibration, while preserving the native host-score construction, Gaussian budget, and post-pruning optimization protocol. Thus, RAGP changes only the retained Gaussian identities, adding no parameters or inference-time rendering operations. Across five OMG pruning budgets and 13 scenes, RAGP improves mean PSNR by 0.061–0.148 dB under matched Gaussian counts across datasets, consistently outperforming the pruning pipeline at the same compression budget.

1 INTRODUCTION

3D Gaussian Splatting (3DGS) represents scenes with explicit anisotropic Gaussian primitives and enables high-quality novel-view synthesis through differentiable rasterization (Kerbl et al., 2023). Following the progress of neural rendering and radiance field representations (Mildenhall et al., 2021; Barron et al., 2021; 2022), 3DGS provides an explicit alternative with improved rendering efficiency. Recent efforts have further explored Gaussian representations through compression, acceleration, and adaptive optimization strategies (Liu et al., 2026; Lu et al., 2024; Ye et al., 2025).

Most Gaussian pruning methods select primitives according to their importance measured at a single model snapshot. Existing criteria typically rely on primitive attributes, rendering contributions, or instantaneous reconstruction sensitivity (Lee et al., 2024a; Fan et al., 2024; Girish et al., 2024; Hanson et al., 2025b; Niedermayr et al., 2024; Ali et al., 2024; Lee et al., 2025). This paradigm overlooks an important property revealed during post-pruning optimization: the quality of a pruned representation is not solely determined by the immediate damage caused by removal, but also by its ability to recover through subsequent adaptation. A similar issue has been observed in neural network compression, where static parameter importance fails to fully characterize post-compression adaptability (LeCun et al., 1989; Han et al., 2015). Therefore, current Gaussian pruning methods suffer from a **selection-evaluation gap**: primitives are selected based on their instantaneous impact, while performance is ultimately determined after recovery optimization. Two pruning decisions with similar initial degradation may lead to different final rendering quality because the retained Gaussians possess different compensation capabilities.

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.



(a) Masked (b) Compensation response (c) Recoverability-aware pruning outcome

Figure 1: **Illustration of localized counterfactual probing.** Localized removal produces the Masked rendering (a), exposes the compensation response (b), and yields recoverability-aware pruning under a matched Gaussian budget (c). Warmer regions indicate stronger responses.

We study this missing factor from a counterfactual perspective through **post-pruning recoverability**: the extent to which the surviving representation can adapt after information becomes unavailable. Rather than treating pruning as a static removal decision, we view it as an intervention on the rendering evidence to the model. This perspective extends conventional importance estimation. Beyond asking *what is costly to remove now*, it asks *how the surviving representation responds when the removed information is no longer available*.

Our experiments reveal a consistent distinction between these two signals. First, normalized recoverability exhibits structured agreement across independent recovery-camera schedules. Second, when we freeze retained Gaussians with strong counterfactual geometric responses, subsequent local recovery is suppressed relative to matched random and low-response controls. Third, ranking Gaussians by this response alone degrades reconstruction quality. Taken together, these observations indicate that counterfactual response captures *compensatory participation* but does not replace global primitive importance. This leads to our central principle: *compensation evidence should complement rather than replace the host importance prior*. Fig. 1 illustrates the motivation of our counterfactual pruning preview: localized information removal exposes the compensation response of the surviving representation, which is then incorporated into fixed-budget Gaussian pruning.

Based on this principle, we develop Recoverability-Aware Gaussian Pruning (RAGP), which uses controlled counterfactual pruning previews to measure how the surviving representation responds when localized information becomes unavailable before the final pruning decision. These responses are aggregated into geometric compensation evidence that reflects local compensatory participation of the retained representation. Rather than using this response as a replacement importance criterion, RAGP refines the final ranking through bounded calibration while preserving the native host-score construction. Because different host criteria respond differently to geometric and rendering normalization, we use a lightweight host-relative calibration before incorporating the counterfactual response. This calibration preserves the native host ranking as the prior while correcting only a bounded projected-area normalization mismatch.

Across five OMG operating points, this fixed-budget refinement yields positive 13-scene average PSNR improvements, ranging from **+0.061** to **+0.148 dB** under the same fixed budget. Because every Official–RAGP pair shares the same source state, host prior, Gaussian count, and recovery protocol, these gains can be attributed to differences in the retained identities under the matched operating point rather than changes in compression ratio or optimization budget. The smaller gain at XL is consistent with diminishing selection headroom in the high-retention regime. We further evaluate the same recoverability-aware refinement under seven heterogeneous Gaussian importance criteria within a pruning-and-recovery protocol. Across all seven host criteria, the 13-scene average PSNR change remains positive, indicating that counterfactual compensation evidence provides complementary information beyond the host importance formulation. An overview of the framework is shown in Fig. 3. The contributions are the following:

- We identify *post-pruning recoverability* as a pruning-relevant factor distinct from instantaneous importance and removal sensitivity. Controlled recovery and freezing interventions show that recoverability exhibits reproducible structure and that high-response retained Gaussians participate disproportionately in subsequent compensation.
- We develop a counterfactual pruning-preview mechanism that measures the response of surviving Gaussians to information removal before the pruning decision, providing com-

pensation evidence about the retained representation. We integrate this evidence through a bounded, host-preserving refinement rather than replacing the importance criterion.

- We validate RAGP across five OMG pruning budgets and seven host importance criteria. The 13-scene ΔPSNR remains positive across all five OMG budgets and under every host criterion, supporting recoverability as complementary evidence beyond an importance formulation, without introducing inference-time parameters or renderer operations.

2 Related Work

2.1 3D Gaussian Splatting Compression

Existing 3DGS compression methods mainly reduce representation cost through compact attribute encoding and primitive reduction. Attribute-oriented approaches compress Gaussian parameters through quantization and compact feature representations (Lee et al., 2024a; Girish et al., 2024; Niedermayr et al., 2024; Navaneet et al., 2024; Xue et al., 2026; Chen et al., 2026b), as well as hash-based and entropy-coded representations (Chen et al., 2024; Xie et al., 2025; Shin et al., 2025; Chen et al., 2026a). Recent works further explore structured Gaussian representations and compact scene parameterization through anchors and adaptive primitives. Primitive-reduction methods instead remove redundant Gaussians, thereby reducing both representation size and rasterization cost. Representative approaches combine importance-based pruning with attribute compression or subsequent optimization recovery (Fan et al., 2024; Lee et al., 2025; Zhang and Sui, 2026; Wang et al., 2026). Our work focuses on the latter and studies how the optimization behavior of the retained representation can inform Gaussian selection. Beyond Gaussian-specific compression, related efforts in efficient neural rendering investigate compact scene representations and accelerated inference (Reiser et al., 2021; Yu et al., 2021; Chen et al., 2022; Le et al., 2026).

2.2 Gaussian Pruning and Importance Estimation

Gaussian pruning methods primarily differ in how they estimate primitive importance. Existing approaches use opacity, rendering contribution, cross-view significance, or learnable masks to identify redundant Gaussians (Lee et al., 2024a; Fan et al., 2024; Girish et al., 2024; Liu et al., 2025a; Tang et al., 2026). Related pruning formulations have also been explored (Ali et al., 2024; Zhang et al., 2024). More recent methods develop increasingly principled importance and sensitivity measures. PUP 3D-GS (Hanson et al., 2025b) estimates second-order reconstruction sensitivity with respect to Gaussian spatial parameters, while REFINE (Chen et al., 2026c) uses a rendering-aware Hessian approximation to estimate removal-induced error. OMG (Lee et al., 2025) employs scene-adaptive importance estimation for aggressive primitive reduction. Other recent approaches investigate improved pruning functions or efficient importance prediction (Lee et al., 2024b; Yang et al., 2026). These methods substantially improve the estimation of primitive contribution or removal sensitivity. Our work considers a complementary source of information: how retained Gaussians adapt when local information is removed. Similar importance estimation principles have been extensively studied in sparse representation and parameter selection (Molchanov et al., 2016; Liu et al., 2024).

2.3 Optimization-Coupled Gaussian Pruning and Recovery

Recent work has increasingly recognized that Gaussian selection and optimization are coupled (Bai et al., 2026; Mallick et al., 2024; Hanson et al., 2025a). LightGaussian (Fan et al., 2024) performs recovery optimization after pruning. Mini-Splatting (Fang and Wang, 2024) observes that pre-simplification importance does not necessarily preserve the spatial structure needed for subsequent optimization and therefore adopts importance-weighted sampling. MaskGaussian (Liu et al., 2025b) avoids making Gaussian existence a permanent decision from a single snapshot by allowing probabilistically masked primitives to continue receiving gradients. GaussianSpa (Zhang et al., 2025) alternates representation optimization and sparsification, while Prune Wisely, Reconstruct Sharply (Wang et al., 2026) adapts pruning and refinement to the evolving reconstruction state.

These methods establish that pruning and optimization should not be treated as independent stages, but they address that coupling in a different way from our work (Zhu and Gupta, 2017; Evci et al., 2020; Xiang et al., 2024). Sensitivity-based methods such as PUP and REFINE estimate the cost of removing a primitive, while optimization-coupled methods modify the recovery, existence, sparsification, or pruning procedure itself. We instead keep the post-pruning optimization objective fixed and use controlled local interventions to *measure the response of the surviving representation* to missing information. The resulting signal is interpreted as evidence of compensatory adaptability and is used only to calibrate an existing host ranking. Thus, our contribution is not another recovery

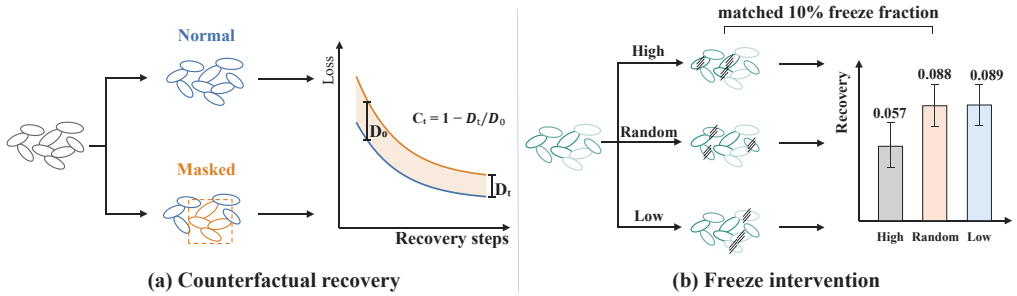

Figure 2: **Evidence for post-pruning recoverability and compensation.** (a) Paired Normal and Masked recovery trajectories. (b) Freezing high-response Gaussians suppresses local recovery. Bars show scene means with 95% bootstrap CIs. Additional details are provided in Appendix F.2.

schedule or standalone importance function, but a measurement of a previously unmodeled factor that connects pruning decisions to subsequent adaptation.

3 Post-Pruning Recoverability

Diagnostic definition. Conventional pruning criteria characterize a Gaussian's contribution or immediate removal effect. However, pruning quality is evaluated after the retained representation undergoes optimization. We therefore distinguish *removal damage*, i.e., the instantaneous perturbation introduced by pruning, from *post-pruning recoverability*, i.e., how much damage can be recovered through adaptation of retained Gaussians. Consequently, similar removal damage need not imply similar post-pruning quality. From this viewpoint, Normal and Masked branches represent counterfactual states of the same representation. The Normal branch follows the optimization trajectory with complete rendering evidence, whereas the Masked branch simulates a localized intervention where rendering evidence becomes unavailable. Their difference probes how optimization dynamics respond to missing information. To measure this distinction, we construct paired Normal and Masked recovery trajectories, as illustrated in Fig. 2(a). For a Gaussian unit U, let $L_N^t(U)$ and $L_M^t(U)$ denote the reconstruction losses after t recovery steps. We define:

$$D_0(U) = L_M^0(U) - L_N^0(U), \qquad D_T(U) = L_M^T(U) - L_N^T(U), \tag{1}$$

and measure the fraction of the initial gap closed through optimization as:

$$\mathcal{C}_T(U) = 1 - \frac{D_T(U)}{D_0(U)}, \qquad D_0(U) > \delta, \tag{2}$$

where δ excludes interventions below the measurement noise floor. Importantly, $\mathcal{C}_T$ is a retrospective diagnostic of recoverability, defined by the adaptation trajectory after information removal. Using $\mathcal{C}_T$ for pruning selection would require candidate-specific recovery rollouts. Therefore, R_i^{geo} is not an estimator of $\mathcal{C}_T$. It is a pre-pruning intervention signal of local compensatory participation.

Counterfactual evidence. We next examine whether recoverability exhibits structured adaptive behavior and whether high-response retained Gaussians participate disproportionately in subsequent compensation. Varying only the recovery-camera schedule while fixing all other factors, pairwise Spearman correlations of normalized recoverability range from 0.432 to 0.831 at $T = 100$ and from 0.637 to 0.818 at $T = 500$, indicating more consistently positive agreement at the longer horizon. We then freeze the top 10% high-response retained Gaussians during matched recovery (Fig. 2(b)). At $T = 100$, this reduces local recovery by 0.031 relative to random freezing and by 0.032 relative to low-response freezing, with the same direction in $12/13$ scenes ($p < 0.001$). At $T = 500$, both mean differences are -0.011, again in $12/13$ scenes. Full results are provided in Appendix F.

These findings support interpreting R_i^{geo} as evidence of *local compensatory participation*, rather than standalone importance. This distinction is critical: ranking Gaussians by response alone degrades reconstruction by **0.098 dB** on average (ΔSSIM $= -0.0189$, ΔLPIPS $= +0.0276$), whereas using the response to refine the host criterion yields positive gains. Thus, *counterfactual response should complement, rather than replace, the host importance prior.*

4 Recoverability-Aware Gaussian Pruning

4.1 Overview

Fig. 3 summarizes the final recoverability-aware pruning framework. Motivated by the intervention evidence in Sec. 3, RAGP seeks to obtain pre-pruning evidence of local compensatory participation

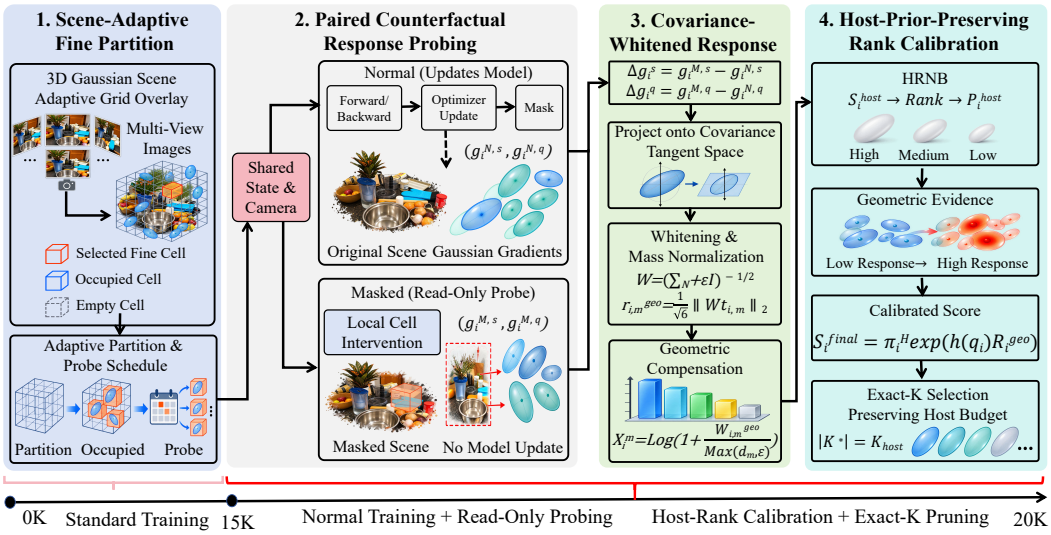

Figure 3: Overview of the Recoverability-Aware Gaussian Pruning (RAGP) framework.

without executing candidate-specific recovery trajectories. Instead of predicting the final recovery outcome, RAGP performs lightweight counterfactual pruning previews and measures how the surviving representation responds to localized information removal.

Let S_i^{host} denote the native importance assigned to Gaussian i. Our framework performs a counterfactual pruning preview by measuring how surviving Gaussians respond when rendering evidence is removed. The response is not treated as a direct importance measurement. Instead, it is transformed into a geometry-aware and optimization-normalized compensation signal R_i^{geo} that reflects local compensatory participation of retained Gaussians. This signal complements the host prior during selection. Because host constructions may encode parts of the same rendering trend, we use a lightweight Host-Relative Normalization Bridge (HRNB) to calibrate the host ranking before the compensation evidence is incorporated. The resulting selection can be summarized as:

$$S^{\text{host}} \xrightarrow{\text{HRNB}} \widetilde{\ell} \xrightarrow{R^{\text{geo}}} S^{\text{final}} \xrightarrow{\text{TopK}} \mathcal{K}^{\star}. \tag{3}$$

The native host formula and pruning budget remain unchanged throughout. We measure compensatory response using a single scene-adaptive Fine partition. The finest spatial resolution $R^{\star}$ satisfying the frozen training-view coverage requirement is selected. Grouping and scheduling details are given in Appendix A.1 and Appendix A.2.

4.2 Counterfactual Whitened Fine Response

For each local cell and camera, Normal and Masked branches start from the same Gaussian state, forming a paired intervention. The Normal branch follows the optimization dynamics, while the Masked branch removes rendering evidence and is evaluated through read-only gradients without parameter updates. Therefore, the trajectory is determined by the Normal branch. For Gaussian i under intervention m, we form the paired scale and rotation gradient contrasts:

$$\Delta \mathbf{g}_{i,m}^{s} = \mathbf{g}_{i,m}^{M,s} - \mathbf{g}_{i,m}^{N,s}, \qquad \Delta \mathbf{g}_{i,m}^{q} = \mathbf{g}_{i,m}^{M,q} - \mathbf{g}_{i,m}^{N,q}. \tag{4}$$

The counterfactual gradient contrast measures how the rendering intervention alters the optimization signal, but remains in heterogeneous scale and rotation spaces and is affected by native gradient statistics. We therefore map it into Gaussian geometric space and normalize it relative to the Normal dynamics rather than use it directly as the recoverability signal. Using the normalized projected directions in Appendix A.3, we map the contrast into each Gaussian's covariance tangent space:

$$\mathbf{t}_{i,m} = \mathcal{T}\left(\mathbf{d}_{i,m}^{s}, \mathbf{d}_{i,m}^{q}; \mathbf{s}_i, \mathbf{q}_i\right). \tag{5}$$

This projection maps heterogeneous parameter-space changes into a geometry-aware representation. The covariance-tangent formulation expresses changes in each Gaussian's covariance geometry, providing a common space for geometric response across heterogeneous parameter update directions. The tangent response is whitened using Normal-branch statistics, normalizing the counterfactual response relative to the covariance structure of normal optimization. This suppresses high-variance

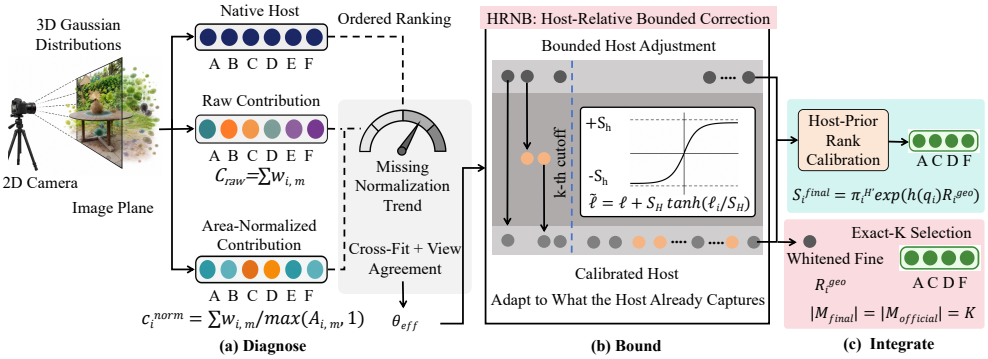

Figure 4: Overview of the Host-Relative Normalization Bridge (HRNB).

directions in ordinary optimization and yields a covariance-aware representation of geometric response. It is then normalized by the estimated removed contribution mass and accumulated across interventions. This produces a conservative per-Gaussian geometric response statistic L_i, which is converted to centered geometric compensation evidence:

$$R_i^{\mathrm{geo}} = 2\widehat{F}_{\mathrm{mid}}(L_i) - 1. \tag{6}$$

Gaussians without sufficient reliable observations receive zero response to avoid unstable compensation estimates. The whitening transform and mass normalization are detailed in Appendix A.4, while online statistics and reliability criteria are given in Appendix A.5.

4.3 Host-Relative Normalization Bridge

As shown in Fig. 4, HRNB diagnoses normalization trends and applies a bounded correction before exact-K selection. HRNB retains the native host score as the ranking prior and serves a role distinct from the counterfactual response. Different host criteria exhibit different sensitivity to projected-area normalization, so directly adding an auxiliary response can produce host-dependent ranking shifts. HRNB therefore estimates how strongly the host ranking already reflects a projected-area-normalized contribution trend and applies only a bounded correction along this normalization direction. The counterfactual response is incorporated afterward. Using all training views at the pruning state, we compute the raw and projected-area-normalized rendering contributions:

$$C_i^{\mathrm{raw}} = \sum_m w_{i,m}, \qquad C_i^{\mathrm{norm}} = \sum_m \frac{w_{i,m}}{\max(A_{i,m}, 1)}. \tag{7}$$

Their ranking change provides a normalization companion rather than a new importance score. A two-fold rank-space estimate with split-view consistency determines a cross-fitted reliability factor $\eta_{\mathrm{eff}}^{(f(i))} \in [0,1]$ for the fold containing Gaussian i. Details are provided in Appendix A.6. Let:

$$\ell_i = \log S_i^{\mathrm{host}}, \qquad r_i = \log C_i^{\mathrm{norm}} - \log C_i^{\mathrm{raw}}, \qquad \widetilde{r}_i = r_i - \mathrm{median}(r), \tag{8}$$

and let $s_H = 1.4826\ \mathrm{MAD}(\ell)$ denote the robust scale of the native log-host scores. HRNB applies:

$$\widetilde{\ell}_i = \ell_i + s_H \tanh\left(\frac{\eta_{\mathrm{eff}}^{(f(i))}\widetilde{r}_i}{s_H}\right). \tag{9}$$

Median centering removes the common offset from the normalization residual before correction, while the bounded update limits its influence on the native host prior. HRNB uses only training-view statistics and leaves the host formula and budget unchanged.

4.4 Final Host-Preserving Rank Calibration

Let r_i^H denote the descending rank induced by the HRNB-calibrated log-host score $\widetilde{\ell}_i$. We convert this ranking into a positive host priority and a smooth rank coordinate:

$$\pi_i^H = \frac{N - r_i^H + 1}{N}, \qquad q_i = \frac{r_i^H - \frac{1}{2}}{N}, \qquad h(q_i) = q_i^2(3 - 2q_i). \tag{10}$$

The geometric compensation evidence then refines this calibrated host priority:

$$S_i^{\mathrm{final}} = \pi_i^H \exp(h(q_i) R_i^{\mathrm{geo}}), \qquad \mathcal{K}^\star = \mathrm{TopK}\left(S^{\mathrm{final}}, K_{\mathrm{official}}\right). \tag{11}$$

Here K_{official} is inherited directly from the native host pruning rule. Thus, the final method changes only which Gaussians are retained, not how many are retained.

Table 1: Quantitative comparison of RAGP with Gaussian pruning and compression methods on three benchmark datasets. We report PSNR, SSIM, LPIPS, and model size in MB.

	Mip-NeRF 360				Tanks & Temples				Deep Blending			
Method	PSNR↑	SSIM↑	LPIPS↓	Size↓	PSNR↑	SSIM↑	LPIPS↓	Size↓	PSNR↑	SSIM↑	LPIPS↓	Size↓
3DGS	27.44	0.813	0.218	822.6	23.67	0.844	0.179	452.4	29.48	0.900	0.246	692.5
Scaffold-GS	27.66	0.812	0.223	187.3	24.11	0.855	0.165	154.3	30.28	0.907	0.243	121.2
Mini-Splatting	27.39	0.822	0.216	119.5	23.41	0.846	0.180	67.6	30.04	0.910	0.241	124.9
CompGS	27.04	0.804	0.243	22.93	23.29	0.835	0.201	14.23	29.89	0.907	0.253	15.15
Compact-3DGS	26.95	0.797	0.244	26.31	23.33	0.831	0.202	18.97	29.71	0.901	0.257	21.75
C3DGS	27.09	0.802	0.237	29.98	23.52	0.837	0.188	18.58	29.53	0.899	0.254	24.96
LightGaussian	26.90	0.800	0.240	53.96	23.32	0.829	0.204	29.94	29.12	0.895	0.262	45.25
PUP 3D-GS	26.67	0.786	0.272	74.65	22.72	0.801	0.244	43.33	28.85	0.881	0.302	69.92
EAGLES	27.10	0.807	0.234	59.49	23.14	0.833	0.203	30.18	29.72	0.906	0.249	54.45
SOG	27.01	0.800	0.226	43.77	23.54	0.833	0.188	24.42	29.21	0.891	0.271	19.32
HAC	27.49	0.807	0.236	16.95	24.08	0.846	0.186	8.42	29.99	0.902	0.268	4.51
LocoGS-S	27.04	0.806	0.232	7.90	23.63	0.847	0.169	6.59	30.06	0.904	0.249	7.64
LocoGS-L	27.33	0.814	0.219	13.89	23.84	0.852	0.161	12.34	30.11	0.906	0.243	13.38
OMG-XS	27.10	0.807	0.245	4.03	23.07	0.833	0.206	2.42	29.85	0.905	0.266	3.30
OMG-X	27.19	0.812	0.236	4.76	23.23	0.838	0.197	2.84	29.90	0.907	0.261	3.86
OMG-M	27.24	0.815	0.230	5.28	23.37	0.842	0.191	3.16	29.92	0.908	0.257	4.32
OMG-L	27.31	0.818	0.222	6.39	23.48	0.846	0.182	3.93	30.04	0.910	0.252	5.18
OMG-XL	27.35	0.819	0.219	6.83	23.54	0.847	0.180	4.28	30.07	0.911	0.251	5.51
RAGP-XS (Ours)	27.21	0.808	0.244	**4.04**	23.42	0.836	0.204	**2.43**	29.98	0.908	0.261	**3.29**
RAGP-X (Ours)	27.29	0.813	0.235	**4.76**	23.43	0.840	0.195	**2.86**	30.03	0.909	0.257	**3.86**
RAGP-M (Ours)	27.36	0.816	0.229	5.29	23.54	0.843	0.190	3.18	30.05	0.910	0.254	4.32
RAGP-L (Ours)	**27.43**	**0.819**	0.221	6.39	23.58	**0.847**	**0.182**	3.94	30.14	0.911	0.250	5.18
RAGP-XL (Ours)	27.40	0.819	**0.219**	6.84	**23.60**	0.848	**0.180**	4.30	**30.16**	**0.912**	**0.249**	5.51

5 EXPERIMENTS

Our experiments address three questions. First, does recoverability-aware selection improve reconstruction across pruning budgets? Second, how do compensation evidence and host-relative calibration contribute to the result? Third, does recoverability remain complementary under heterogeneous importance formulations? We use OMG for the multi-budget and component analyses, and evaluate seven host importance criteria under a unified pruning-and-recovery protocol.

5.1 EXPERIMENTAL SETUP

We evaluate on Tanks & Temples, Deep Blending, and Mip-NeRF 360 using PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018). The comparison methods and configurations reported in Table 1 are summarized in Appendix B.2. OMG is the primary host for multi-budget and component evaluation. Cross-criterion generalization is evaluated by replacing the host importance ranking with heterogeneous criteria while fixing the remaining pruning, recovery, and evaluation protocol.

Paired evaluation protocol. RAGP is evaluated as a fixed-budget refinement of the native host rather than as a replacement pruning criterion. Within every Official–RAGP pair, both branches start from the same source state, use the same native host-prior construction, retain exactly the same number of Gaussians K, and follow the same post-pruning optimization protocol. Consequently, the comparison changes only the identities of the retained Gaussians. Improvements under this protocol therefore measure selection headroom within a fixed host operating point, rather than gains obtained by changing model size, pruning budget, or recovery optimization. Selection does not use test-set metrics or future recovery trajectories. We denote the five matched OMG importance-threshold settings by **XS**, **X**, **M**, **L**, and **XL**, corresponding to 0.96, 0.98, 0.99, 0.999, and 0.9999, respectively.

Host-prior attribution. All primary OMG comparisons use the native OMG host construction for each dataset. Tanks & Temples uses the Outdoor-style score, Deep Blending uses the Indoor-style score, and Mip-NeRF 360 follows the native OMG Indoor/Outdoor mapping. Within every paired comparison, Official OMG and the recoverability-aware branch use the same host prior before calibration is applied. Appendix H separately analyzes the interaction between compensation evidence and alternative OMG host constructions. Within every scene–budget pair, the paired branches share the same 20K source state, exact-K budget, permanent IDs, RNG state, camera schedule, and post-

Table 2: **Generalization across heterogeneous host importance criteria.** Each entry reports *Host* → *Host+RAGP* under an OMG pruning-and-recovery protocol at budget X. Only the host criterion changes, while the checkpoint, Gaussian budget, recovery schedule, and evaluation remain fixed.

Metric	Host criterion	Tanks & Temples	Deep Blending	Mip-NeRF 360
PSNR ↑	Opacity	22.943 → **23.166**	29.928 → **30.196**	26.898 → **27.010**
	FoV-style	22.346 → **22.876**	29.508 → **29.833**	26.862 → **26.988**
	MesonGS	22.783 → **23.081**	29.582 → **29.816**	26.754 → **26.850**
	GoDe grad-all	22.657 → **23.182**	30.238 → **30.295**	27.048 → **27.295**
	PUP Fisher	21.187 → **22.163**	29.619 → **29.861**	26.147 → **26.617**
	PWR	23.226 → **23.260**	30.006 → **30.191**	27.080 → **27.147**
	FastGS-style	21.576 → **22.665**	27.338 → **30.000**	24.792 → **26.939**
SSIM ↑	Opacity	**0.8265** → 0.8235	0.9049 → **0.9060**	0.8046 → **0.8079**
	FoV-style	0.8090 → **0.8219**	0.9033 → **0.9071**	0.8049 → **0.8069**
	MesonGS	0.8095 → **0.8203**	0.8969 → **0.8997**	0.7934 → **0.7987**
	GoDe grad-all	0.8199 → **0.8336**	0.9117 → **0.9119**	0.8102 → **0.8144**
	PUP Fisher	0.7731 → **0.7977**	0.9004 → **0.9035**	0.7868 → **0.7964**
	PWR	0.8196 → **0.8249**	0.9026 → **0.9046**	0.8017 → **0.8055**
	FastGS-style	0.7849 → **0.8088**	0.8730 → **0.9038**	0.7639 → **0.8067**
LPIPS ↓	Opacity	**0.2122** → 0.2222	**0.2663** → 0.2669	0.2416 → **0.2412**
	FoV-style	0.2338 → **0.2232**	0.2692 → **0.2622**	0.2429 → **0.2411**
	MesonGS	0.2374 → **0.2277**	0.2818 → **0.2786**	0.2567 → **0.2512**
	GoDe grad-all	0.2178 → **0.2056**	0.2533 → **0.2529**	0.2420 → **0.2356**
	PUP Fisher	0.2640 → **0.2430**	0.2754 → **0.2698**	0.2678 → **0.2574**
	PWR	0.2361 → **0.2250**	0.2799 → **0.2728**	0.2546 → **0.2484**
	FastGS-style	0.2571 → **0.2328**	0.3138 → **0.2713**	0.2908 → **0.2431**

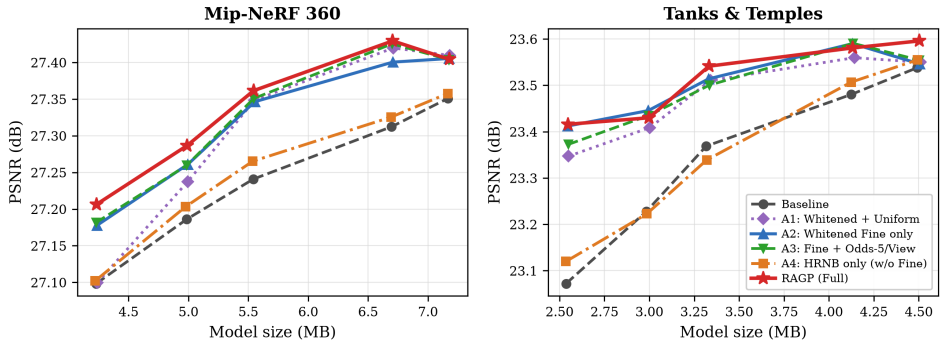

Figure 5: Component behavior across pruning budgets.

pruning recovery/evaluation protocol. Complete training, probing, implementation, and hardware details are provided in Appendices A and B. **HRNB** uses only training-view statistics at the frozen 20K state and does not modify the native host formula, Gaussian budget, or recovery protocol.

5.2 Performance Comparison

Main results. Table 1 compares RAGP with representative Gaussian pruning and compression methods. For OMG, RAGP is evaluated as a host-preserving refinement, with each paired comparison using the same native host prior, exact Gaussian count, and recovery protocol, differing only in the retained identities. RAGP improves the 13-scene average ΔPSNR across all OMG operating points, with gains ranging from **+0.061 dB to +0.148 dB**, while average SSIM and LPIPS are also favorable throughout. The smaller gain at XL may reflect reduced selection headroom at higher retention rates, where fewer pruning decisions are available for correction. Nevertheless, 12/13 scenes still improve in PSNR. These gains therefore isolate improvements associated with the retained Gaussian identities within the host-defined operating point, rather than changes in model size or pruning budget. Complete scene-level results are provided in Appendix C.

Generalization across importance criteria. As shown in Table 2, we evaluate seven Gaussian importance criteria while fixing the remaining pruning-and-recovery pipeline, including the 20K source state, budget X, exact Gaussian count, OMG recovery schedule, and evaluation protocol. Only the host ranking and retained identities differ. Criterion details are provided in Appendix B.5.

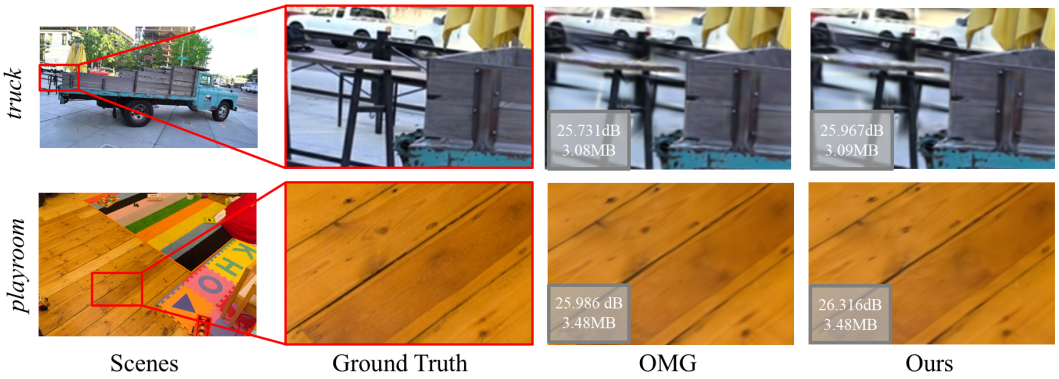

Figure 6: Qualitative comparison under a matched Gaussian budget.

Recoverability-aware selection yields positive dataset-level PSNR changes on Tanks & Temples, Deep Blending, and Mip-NeRF 360 for every criterion, with **positive 13-scene average PSNR gains across all seven host criteria**. Average SSIM also improves for all seven criteria. LPIPS improves for six criteria, with a small perceptual trade-off under Opacity. Average FPS changes vary in both directions across criteria, indicating no systematic increase in rendering cost. Aggregate quality and FPS changes are summarized in Appendix E. These results support the complementarity of counterfactual compensation evidence across heterogeneous Gaussian importance formulations.

5.3 Component Analysis

The final method combines counterfactual compensation evidence with a lightweight host-relative calibration. We examine their roles across five OMG pruning budgets using the same host prior and matched Gaussian budgets. Whitened Fine provides the dominant standalone improvement. At budget X, it improves the 13-scene average PSNR by **+0.1059 dB**, compared with **+0.0123 dB** for HRNB alone, while their combination reaches **+0.1208 dB**. Fig. 5 shows that the contribution of HRNB varies across datasets and pruning regimes, supporting its role as a complementary host-relative calibration rather than a replacement for the counterfactual compensation signal. A final-mask analysis shows that the calibration remains local in its effect on the retained set. Relative to Whitened Fine alone, the full method changes only **4.11%** of the selected Gaussians while preserving **95.89%** overlap. Together with the exact-K paired protocol, this supports viewing the method as a bounded refinement of host-preserving selection rather than a replacement pruning criterion.

5.4 Design Controls and Efficiency

Design controls. We examine several alternatives within the final pipeline at budget X. Replacing the final response with a gradient-norm response gives **+0.0953 dB**, while using the covariance-tangent representation without whitening gives **+0.1075 dB**. Using the whitened response with uniform rather than host-rank integration gives **+0.0972 dB**, and removing the reliability constraint gives **+0.0922 dB**. All remain below the **+0.1208 dB** improvement of the final method. Additional response-construction, spatial-granularity, and multi-scale ablations are provided in Appendix G.

Efficiency. Across train, playroom, and room, the method introduces an average training-time overhead of **36.02%** relative to Official. Most of the cost comes from counterfactual Whitened Fine measurement, while the subsequent host-relative calibration contributes only a small fraction. The method introduces no additional Gaussian attributes, learned parameters, or renderer operations at inference time. Detailed profiling is provided in Appendix J. Fig. 6 provides a qualitative comparison under the same retained Gaussian budget. The recoverability-aware result preserves fine structures and local appearance while keeping the representation size matched to the host baseline.

6 Conclusion

This work exposes a selection–evaluation gap in Gaussian pruning: primitives are selected before, but evaluated after, post-pruning adaptation. Controlled interventions show that counterfactual geometric responses identify retained Gaussians that participate in subsequent compensation, while response-only pruning performs poorly. Based on this observation, RAGP combines intervention-derived compensation evidence with the native host criterion under a fixed Gaussian budget. Across five OMG operating points and seven host criteria, this refinement consistently improves PSNR without changing inference-time representation or rendering operations. These results suggest that post-pruning adaptation provides useful information beyond instantaneous removal importance.

AI USE STATEMENT

The authors used generative AI tools to assist with literature organization, drafting, LaTeX formatting, and code review. All AI-assisted text, equations, and code were reviewed by the authors, and all numbers were checked against the project's experiment records. The authors take responsibility for the final content of this work.

ETHICS STATEMENT

This work uses publicly available scene datasets and does not involve human subjects. No additional ethical concerns were identified.

REPRODUCIBILITY STATEMENT

The main paper specifies the paired evaluation protocol and the principal components of RAGP. The appendix provides implementation details, host-criterion construction, recovery schedules, hyperparameters, scene-level results, and efficiency measurements required to reproduce the experiments.

REFERENCES

Muhammad Salman Ali, Maryam Qamar, Sung-Ho Bae, and Enzo Tartaglione. Trimming the Fat: Efficient Compression of 3D Gaussian Splats through Pruning. In *British Machine Vision Conference*, 2024.

Tian Bai, Zheng Qiu, Haojie Chen, and Ziyang Dai. Plug-and-play Optimization for 3D Gaussian Splatting Compression: Distribution Regularization, Probabilistic Pruning and Detail Compensation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 2372–2380, 2026.

Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-NeRF: A Multiscale Representation for Anti-aliasing Neural Radiance Fields. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5835–5844. IEEE, 2021.

Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-NeRF 360: Unbounded Anti-aliased Neural Radiance Fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5460–5469. IEEE, 2022.

Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. TensoRF: Tensorial Radiance Fields. In *European Conference on Computer Vision*, pages 333–350. Springer, 2022.

Yihang Chen, Qianyi Wu, Weiyao Lin, Mehrtash Harandi, and Jianfei Cai. HAC: Hash-Grid Assisted Context for 3D Gaussian Splatting Compression. In *European Conference on Computer Vision*, pages 422–438, 2024.

Yihang Chen, Mengyao Li, Qianyi Wu, Weiyao Lin, Mehrtash Harandi, and Jianfei Cai. PCGS: Progressive Compression of 3D Gaussian Splatting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 3111–3119, 2026a.

Yihang Chen, Qianyi Wu, Mengyao Li, Weiyao Lin, Junhui Hou, Mehrtash Harandi, and Jianfei Cai. Feedforward Compression of Static and Streamable 3D Gaussian Splatting. *IEEE Transactions on Circuits and Systems for Video Technology*, 2026b.

Zhang Chen, Shuai Wan, Mengting Yu, Fuzheng Yang, and Junhui Hou. REFINE: Super-Efficient 3D Gaussian Splatting Pruning via Rendering-Free Primitive Importance, 2026c. URL https://arxiv.org/abs/2606.09074.

Francesco Di Sario, Riccardo Renzulli, Marco Grangetto, Akihiro Sugimoto, and Enzo Tartaglione. GoDe: Gaussians on Demand for Progressive Level of Detail and Scalable Compression. *arXiv preprint arXiv:2501.13558*, 2025. URL https://arxiv.org/abs/2501.13558.

Utku Evci, Trevor Gale, Jacob Menick, Pablo Samuel Castro, and Erich Elsen. Rigging the Lottery: Making all Tickets Winners. In *International Conference on Machine Learning*, pages 2943–2952. PMLR, 2020.

Zhiwen Fan, Kevin Wang, Kairun Wen, Zehao Zhu, Dejia Xu, and Zhangyang Wang. LightGaussian: Unbounded 3D Gaussian Compression with 15x Reduction and 200+ FPS. In *Advances in Neural Information Processing Systems*, volume 37, pages 140138–140158, 2024.

Guangchi Fang and Bing Wang. Mini-Splatting: Representing Scenes with a Constrained Number of Gaussians. In *European Conference on Computer Vision*, pages 165–181, 2024.

Sharath Girish, Kamal Gupta, and Abhinav Shrivastava. EAGLES: Efficient Accelerated 3D Gaussians with Lightweight EncodingS. In *European Conference on Computer Vision*, pages 54–71, 2024.

Song Han, Jeff Pool, John Tran, and William Dally. Learning both Weights and Connections for Efficient Neural Network. *Advances in Neural Information Processing Systems*, 28, 2015.

Alex Hanson, Allen Tu, Geng Lin, Vasu Singla, Matthias Zwicker, and Tom Goldstein. Speedysplat: Fast 3D Gaussian Splatting with Sparse Pixels and Sparse Primitives. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21537–21546. IEEE, 2025a.

Alex Hanson, Allen Tu, Vasu Singla, Mayuka Jayawardhana, Matthias Zwicker, and Tom Goldstein. PUP 3D-GS: Principled Uncertainty Pruning for 3D Gaussian Splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5949–5958, 2025b.

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.

Thanh-Hai Le, Hoang-Hau Tran, and Trong-Nghia Vu. Few TensoRF: Enhance the Few-shot on Tensorial Radiance Fields. *arXiv preprint arXiv:2603.25008*, 2026.

Yann LeCun, John Denker, and Sara Solla. Optimal Brain Damage. *Advances in Neural Information Processing Systems*, 2, 1989.

Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, and Eunbyung Park. Compact 3D Gaussian Representation for Radiance Field. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21719–21728, 2024a.

Joo Chan Lee, Jong Hwan Ko, and Eunbyung Park. Optimized Minimal 3D Gaussian Splatting. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pages 135864–135888. Curran Associates, Inc., 2025. doi: 10.52202/085713-4534. URL `https://arxiv.org/abs/2503.16924`.

Yongjae Lee, Zhaoliang Zhang, and Deliang Fan. SafeguardGS: 3D Gaussian Primitive Pruning While Avoiding Catastrophic Scene Destruction, 2024b. URL `https://arxiv.org/abs/2405.17793`.

Weikai Lin, Yu Feng, and Yuhao Zhu. MetaSapiens: Real-Time Neural Rendering with Efficiency-Aware Pruning and Accelerated Foveated Rendering. In *Proceedings of the 30th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, pages 669–682, 2025.

Bohan Liu, Zijie Zhang, Peixiong He, Zhensen Wang, Yang Xiao, Ruimeng Ye, Yang Zhou, Wei-Shinn Ku, and Bo Hui. A Survey of Lottery Ticket Hypothesis. *arXiv preprint arXiv:2403.04861*, 2024.

Hengyu Liu, Yuehao Wang, Chenxin Li, Ruisi Cai, Kevin Wang, Wuyang Li, Pavlo Molchanov, Peihao Wang, and Zhangyang Wang. FlexGS: Train Once, Deploy Everywhere with Many-in-one Flexible 3D Gaussian Splatting. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16336–16345. IEEE, 2025a.

Xiang Liu, Yimin Zhou, Jinxiang Wang, Yujun Huang, Shuzhao Xie, Shiyu Qin, Mingyao Hong, Jiawei Li, Yaowei Wang, Zhi Wang, et al. Splatwizard: A Benchmark Toolkit for 3D Gaussian Splatting Compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2261–2271, 2026.

Yifei Liu, Zhihang Zhong, Yifan Zhan, Sheng Xu, and Xiao Sun. MaskGaussian: Adaptive 3D Gaussian Representation from Probabilistic Masks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 681–690, 2025b.

Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. Scaffold-GS: Structured 3D Gaussians for View-adaptive Rendering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20654–20664. IEEE, 2024.

Saswat Subhajyoti Mallick, Rahul Goel, Bernhard Kerbl, Markus Steinberger, Francisco Vicente Carrasco, and Fernando De La Torre. Taming 3DGS: High-quality Radiance Fields with Limited Resources. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–11, 2024.

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. *Communications of the ACM*, 65(1):99–106, 2021.

Pavlo Molchanov, Stephen Tyree, Tero Karras, Timo Aila, and Jan Kautz. Pruning Convolutional Neural Networks for Resource Efficient Inference. *arXiv preprint arXiv:1611.06440*, 2016.

Wieland Morgenstern, Florian Barthel, Anna Hilsmann, and Peter Eisert. Compact 3D Scene Representation via Self-Organizing Gaussian Grids. In *Computer Vision – ECCV 2024*, pages 18–34, Cham, 2025. Springer Nature Switzerland. doi: 10.1007/978-3-031-73013-9_2. URL https://fraunhoferhhi.github.io/Self-Organizing-Gaussians/.

K. L. Navaneet, Kossar Pourahmadi Meibodi, Soroush Abbasi Koohpayegani, and Hamed Pirsiavash. CompGS: Smaller and Faster Gaussian Splatting with Vector Quantization. In *European Conference on Computer Vision*, pages 330–349, 2024.

Simon Niedermayr, Josef Stumpfegger, and Rüdiger Westermann. Compressed 3D Gaussian Splatting for Accelerated Novel View Synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10349–10358, 2024.

Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. KiloNeRF: Speeding up Neural Radiance Fields with Thousands of Tiny MLPs. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14315–14325. IEEE, 2021.

Shiwei Ren, Tianci Wen, Yongchun Fang, and Biao Lu. FastGS: Training 3D Gaussian Splatting in 100 Seconds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26094–26103, 2026.

Seungjoo Shin, Jaesik Park, and Sunghyun Cho. Locality-aware Gaussian Compression for Fast and High-quality Rendering. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025.

Zhenyu Tang, Chaoran Feng, Xinhua Cheng, Wangbo Yu, Junwu Zhang, Yuan Liu, Xiao-Xiao Long, Wenping Wang, and Li Yuan. NeuralGS: Bridging Neural Fields and 3D Gaussian Splatting for Compact 3D Representations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 9493–9501, 2026.

Haoran Wang, Guoxi Huang, Fan Zhang, David Bull, and Nantheera Anantrasirichai. Prune wisely, reconstruct sharply: Compact 3D gaussian splatting via adaptive pruning and difference-of-gaussian primitives. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11716–11725, June 2026.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image Quality Assessment: from Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004.

Mingcan Xiang, Jiaxun Tang, Qizheng Yang, Hui Guan, and Tongping Liu. Adapmtl: Adaptive Pruning Framework for Multitask Learning Model. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 5121–5130, 2024.

Shuzhao Xie, Weixiang Zhang, Chen Tang, Yunpeng Bai, Rongwei Lu, Shijia Ge, and Zhi Wang. MesonGS: Post-training Compression of 3D Gaussians via Efficient Attribute Transformation. In *European Conference on Computer Vision*, pages 434–452. Springer, 2024.

Yijia Xie, Yuhang Lin, Laijian Li, Lina Liu, Xiaobin Wei, Yong Liu, and Jiajun Lv. Hash-GS: Anchor-Based 3D Gaussian Splatting with Multi-Resolution Hash Encoding for Efficient Scene Reconstruction. In *IEEE International Conference on Robotics and Automation*, 2025.

Ruixiang Xue, Tong Chen, and Zhan Ma. 3D Gaussian Splatting Compression with Object Scalability. In *European Conference on Computer Vision*, pages 1–20. Springer, 2026.

Kaifa Yang, Qi Yang, Yiling Xu, and Zhu Li. RAP: Fast Feedforward Rendering-Free Attribute-Guided Primitive Importance Score Prediction for Efficient 3D Gaussian Splatting Processing, 2026. URL `https://arxiv.org/abs/2602.19753`.

Vickie Ye, Ruilong Li, Justin Kerr, Matias Turkulainen, Brent Yi, Zhuoyang Pan, Otto Seiskari, Jianbo Ye, Jeffrey Hu, Matthew Tancik, et al. gsplat: An Open-source Library for Gaussian Splatting. *Journal of Machine Learning Research*, 26(34):1–17, 2025.

Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. PlenOctrees for Real-time Rendering of Neural Radiance Fields. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5732–5741. IEEE, 2021.

Baobing Zhang and Wanxin Sui. GETA-3DGS: Automatic Joint Structured Pruning and Quantization for 3D Gaussian Splatting. *arXiv preprint arXiv:2605.02086*, 2026.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 586–595. IEEE, 2018.

Yangming Zhang, Wenqi Jia, Wei Niu, and Miao Yin. GaussianSpa: An "Optimizing-Sparsifying" Simplification Framework for Compact and High-Quality 3D Gaussian Splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26673–26682, 2025.

Zhaoliang Zhang, Tianchen Song, Yongjae Lee, Li Yang, Cheng Peng, Rama Chellappa, and Deliang Fan. LP-3DGS: Learning to Prune 3D Gaussian Splatting. In *Advances in Neural Information Processing Systems*, volume 37, pages 122434–122457, 2024.

Michael Zhu and Suyog Gupta. To Prune, or not to Prune: Exploring the Efficacy of Pruning for Model Compression. *arXiv preprint arXiv:1710.01878*, 2017.

A Complete RAGP Method Details

This section provides the implementation details omitted from the main paper. Unless stated, all quantities below correspond to the frozen single-level Fine configuration used by the final method.

A.1 Fine-Grid Construction and Scene-Adaptive Resolution

The final scene-adaptive resolutions and occupied-cell counts are summarized in Table 3. At the frozen probing state, Gaussian centers are normalized independently along the three axes of the scene bounding box:

$$\bar{\mathbf{x}}_i = \mathrm{clip}\left(\frac{\mathbf{x}_i - \mathbf{x}_{\min}}{\mathbf{x}_{\max} - \mathbf{x}_{\min}}, 0, 1-\epsilon\right). \tag{12}$$

The final method constructs only one spatial level:

$$\mathbf{c}_i = \lfloor R^\star \bar{\mathbf{x}}_i \rfloor, \tag{13}$$

where only occupied cells are instantiated. The grid is regular in normalized scene coordinates and is therefore not required to be isotropic in world coordinates.

To determine $R^\star$, we first identify an initial resolution whose number of occupied cells is closest to a target of 128 groups. Starting from this resolution, increasingly finer candidate resolutions are examined consecutively. For each candidate, the frozen training-camera schedule over the probing window is simulated. A candidate is feasible only if every eligible Fine cell can receive at least two actual Masked observations. When ordinary scheduling is insufficient, a frozen deadline-aware reservation rule prioritizes under-observed cells that are approaching their final visible opportunities. The selected resolution is:

$$R^\star = \max\{R : R_0, \ldots, R \text{ all satisfy the required coverage}\}. \tag{14}$$

If the coverage requirement cannot be satisfied, the preflight check fails rather than silently accepting an incompletely measured partition.

Table 3: Scene-adaptive Fine resolutions and occupied-cell counts of the frozen Fine partition.

Dataset	Scene	$R^\star$	Occupied Fine cells
Tanks & Temples	train	11	187
	truck	22	1,227
Deep Blending	drjohnson	23	1,364
	playroom	19	1,309
Mip-NeRF 360	bicycle	27	1,204
	bonsai	23	1,318
	counter	21	903
	flowers	26	1,470
	garden	20	483
	kitchen	14	227
	room	35	1,223
	stump	20	911
	treehill	39	1,524

A.2 Counterfactual Probe Scheduling and Attenuation

The complete probing, pruning, and recovery timeline is summarized in Table 4. The primary OMG experiments enter probing from the same frozen 15K source state. The first part of the measurement window is used to collect Normal-gradient statistics for the whitening baseline, after which counterfactual probes are evaluated until pruning at 20K. Only the Normal branch performs the update.

For a selected Fine cell m under training camera v, let $C_{i,v}$ denote the rendering contribution of Gaussian g_i. Its contribution fraction is:

$$a_{m,v} = \frac{\sum_{i \in m} C_{i,v}}{\sum_j C_{j,v}}. \tag{15}$$

Table 4: Counterfactual probing schedule used by the final method.

Stage	Iterations	Configuration
Initialization	15K	Frozen Fine partition and permanent IDs
Normal-baseline warm-up	15K–15.5K	Normal gradients only
Counterfactual probing	15.5K–20K	Paired Normal/Masked measurements
Exact-K pruning	20K	One global deterministic selection
Matched recovery	after 20K	Same host recovery protocol

Given the scheduler-provided minimum attenuation multiplier b and dataset-specific target intervention mass ρ, we use:

$$\eta_{m,v} = \min\left(1 - b, \frac{\rho}{a_{m,v}}\right), \qquad \gamma_{m,v} = 1 - \eta_{m,v}, \tag{16}$$

We define the contribution-based intervention mass proxy as:

$$d_{m,v} = a_{m,v}\eta_{m,v} \le \rho. \tag{17}$$

This proxy scales the pre-intervention contribution fraction by the attenuation strength; it does not measure the exact rendering change after alpha compositing. Thus, ρ bounds the intervention mass proxy rather than the resulting image difference. The frozen dataset-specific values are:

$$\rho = \begin{cases} 0.020, & \text{Tanks \& Temples,} \\ 0.015, & \text{Deep Blending,} \\ 0.020, & \text{Mip-NeRF 360.} \end{cases} \tag{18}$$

Each probe uses the same Gaussian state and the same training camera for the two branches. The Normal branch performs the ordinary forward/backward pass. Its gradients and random-number-generator states are saved before the Masked rendering is evaluated. Masked gradients are obtained with `autograd.grad`. The Normal gradients and random states are then restored, and only the Normal optimizer step is executed. Therefore, the Masked branch does not modify model parameters, optimizer state, training gradients, random-number state, or the camera schedule.

A.3 Covariance-Tangent Geometry Representation

Let $\mathbf{s}_i$ denote raw log-scales and $\widetilde{\mathbf{q}}_i$ the raw quaternion in (w, x, y, z) order. For intervention m, the paired gradient contrasts are:

$$\Delta\mathbf{g}^s_{i,m} = \mathbf{g}^{M,s}_{i,m} - \mathbf{g}^{N,s}_{i,m}, \qquad \Delta\mathbf{g}^q_{i,m} = \mathbf{g}^{M,q}_{i,m} - \mathbf{g}^{N,q}_{i,m}. \tag{19}$$

Let $\mathcal{R}_m$ contain Gaussians visible in either branch. We normalize the two parameter blocks using Normal-gradient statistics:

$$\nu^a_m = \mathrm{median}_{j\in\mathcal{R}_m} \|\mathbf{g}^{N,a}_{j,m}\|_2 + \epsilon, \qquad a \in \{s, q\}, \quad \epsilon = 10^{-8}. \tag{20}$$

The measurement directions are:

$$\mathbf{d}^s_{i,m} = -\Delta\mathbf{g}^s_{i,m}/\nu^s_m, \qquad \mathbf{d}^{\widetilde{q}}_{i,m} = -\Delta\mathbf{g}^q_{i,m}/\nu^q_m. \tag{21}$$

These are normalized negative-gradient contrasts, not actual Adam updates. We normalize the quaternion and remove its radial component:

$$z_i = \max(\|\widetilde{\mathbf{q}}_i\|_2, \epsilon), \qquad \mathbf{q}_i = \widetilde{\mathbf{q}}_i/z_i, \qquad \mathbf{d}^q_{i,m} = \frac{(\mathbf{I} - \mathbf{q}_i\mathbf{q}_i^\top)\mathbf{d}^{\widetilde{q}}_{i,m}}{z_i}. \tag{22}$$

The corresponding local angular direction is:

$$\boldsymbol{\omega}_{i,m} = 2\ \mathrm{vec}\left(\overline{\mathbf{q}}_i \otimes \mathbf{d}^q_{i,m}\right), \tag{23}$$

where $\overline{\mathbf{q}}_i$ is the quaternion conjugate and vec extracts the vector part. For $a < b$, define $\delta_{ab} = \mathrm{clip}(s_{ib} - s_{ia}, -10, 10)$. The off-diagonal responses are:

$$a_{01} = -2\omega_z \sinh(\delta_{01}), \quad a_{02} = 2\omega_y \sinh(\delta_{02}), \quad a_{12} = -2\omega_x \sinh(\delta_{12}). \tag{24}$$

The six-dimensional geometric mapping is:

$$\mathbf{t}_{i,m} = \mathcal{T}(\mathbf{d}^s_{i,m}, \mathbf{d}^q_{i,m}; \mathbf{s}_i, \mathbf{q}_i) = \begin{bmatrix} 2d^s_{i0} & 2d^s_{i1} & 2d^s_{i2} & \sqrt{2}a_{01} & \sqrt{2}a_{02} & \sqrt{2}a_{12} \end{bmatrix}^\top. \tag{25}$$

To interpret this construction, write $\mathbf{\Sigma}_i = \mathbf{R}_i \mathbf{D}_i^2 \mathbf{R}_i^\top$, where $\mathbf{D}_i = \mathrm{diag}(\exp(\mathbf{s}_i))$. Without log-scale-difference clipping, the vector above encodes:

$$\mathbf{T}_{i,m} = \mathbf{D}_i^{-1} \mathbf{R}_i^\top \dot{\mathbf{\Sigma}}_{i,m} \mathbf{R}_i \mathbf{D}_i^{-1}, \tag{26}$$

where $\dot{\mathbf{\Sigma}}_{i,m}$ is the covariance derivative along the measurement direction. The $\sqrt{2}$ factors preserve the Frobenius norm of this symmetric matrix under vectorization. Clipping provides a numerically stabilized implementation. Non-finite mapped components are set to zero.

A.4 Normal-Baseline Whitening

During the 500-step Normal-only warm-up, tangent samples are collected before each optimizer update from Gaussians visible in the current training view for numerically stable online covariance estimation. For each parameter block $a \in \{s, q\}$, we define:

$$\nu^{N,a}_k = \mathrm{median}_{j \in \mathcal{V}^N_k} \|\mathbf{g}^{N,a}_{j,k}\|_2 + \epsilon, \qquad \mathbf{d}^{N,a}_{i,k} = -\mathbf{g}^{N,a}_{i,k} / \nu^{N,a}_k, \tag{27}$$

where $\mathcal{V}^N_k$ is the Normal visibility set and $\epsilon = 10^{-8}$. For the quaternion block, the normalized raw direction is projected as in Appendix A.3. The same mapping $\mathcal{T}$ then produces $\mathbf{t}^N_{i,k}$.

Each visible Gaussian–iteration pair contributes one sample. Repeated observations of the same Gaussian are retained. Zero gradients contribute zero tangents. The baseline uses neither Masked gradients nor removed-mass normalization.

For the n accumulated samples, we compute in double precision:

$$\mathbf{u} = \sum_{j=1}^{n} \mathbf{t}^N_j, \qquad \mathbf{G} = \sum_{j=1}^{n} \mathbf{t}^N_j (\mathbf{t}^N_j)^\top, \qquad \mathbf{\Sigma}_N = \frac{\mathbf{G} - \mathbf{u}\mathbf{u}^\top / n}{n-1}. \tag{28}$$

At least two samples are required. After explicit symmetrization, let:

$$\mathbf{\Sigma}_N = \mathbf{U}\, \mathrm{diag}(\lambda_1, \ldots, \lambda_6) \mathbf{U}^\top. \tag{29}$$

We apply an eigenvalue floor:

$$\bar{\lambda} = \max\left(\frac{1}{6} \sum_{j=1}^{6} \lambda_j, 10^{-8} \right), \qquad \tau = \max\left(\bar{\lambda} \sqrt{\epsilon^{(64)}_{\mathrm{mach}}}, 10^{-8} \right). \tag{30}$$

The frozen whitening matrix is:

$$\mathbf{W}_N = \mathbf{U}\, \mathrm{diag}\left(\max(\lambda_1, \tau)^{-1/2}, \ldots, \max(\lambda_6, \tau)^{-1/2} \right) \mathbf{U}^\top. \tag{31}$$

Here $\epsilon^{(64)}_{\mathrm{mach}}$ is the machine precision of float64. For a surviving Gaussian g_i under counterfactual intervention m, its single-event geometric response is:

$$W^{\mathrm{geo}}_{i,m} = \frac{1}{\sqrt{6}} \|\mathbf{W}_N \mathbf{t}_{i,m}\|_2. \tag{32}$$

Gaussians belonging to the perturbed Fine cell are excluded from the responder set. Because different interventions remove different amounts of rendering contribution, we normalize each response by the realized removed contribution mass d_m:

$$x_{i,m} = \log\left(1 + \frac{W^{\mathrm{geo}}_{i,m}}{\max(d_m, \epsilon)} \right). \tag{33}$$

Whitening and removed-mass normalization are fixed parts of the response estimator used in all final experiments.

A.5 Reliable Single-Level Response Statistics

For each Gaussian, valid mass-normalized observations $x_{i,m}$ are accumulated online using Welford statistics for numerically stable streaming estimation:

$$n_i, \qquad \bar{x}_i, \qquad M_{2,i}. \tag{34}$$

The sample variance estimate is:

$$\widehat{\sigma}_i^2 = \frac{M_{2,i}}{\max(n_i - 1, 1)}. \tag{35}$$

We use a conservative response statistic:

$$L_i = \left[\bar{x}_i - \gamma \sqrt{\frac{\widehat{\sigma}_i^2}{n_i + \kappa}} \right]_+ \frac{n_i}{n_i + \kappa}, \tag{36}$$

where $[z]_+ = \max(z, 0)$, $\gamma = 1$, and $\kappa = 3$. This statistic is computed once at least two valid observations are available, while a Gaussian contributes nonzero final compensation evidence only when $n_i \geq 4$.

Among Gaussians satisfying this final observation requirement, we use a tie-preserving empirical midpoint CDF. Let $\mathcal{E}$ denote the eligible set. Then:

$$\widehat{F}_{\text{mid}}(L_i) = \frac{|\{j \in \mathcal{E} : L_j < L_i\}| + \frac{1}{2}|\{j \in \mathcal{E} : L_j = L_i\}|}{|\mathcal{E}|}. \tag{37}$$

The frozen geometric compensation evidence is:

$$R_i^{\text{geo}} = \begin{cases} 2\widehat{F}_{\text{mid}}(L_i) - 1, & n_i \geq 4, \\ 0, & \text{otherwise.} \end{cases} \tag{38}$$

Because the final method uses only the scene-adaptive Fine level, no cross-scale aggregation is performed.

A.6 HRNB Cross-Fit and View Consistency

At the frozen 20K state, we render all training cameras and accumulate two contribution statistics:

$$C_i^{\text{raw}} = \sum_m w_{i,m}, \qquad C_i^{\text{norm}} = \sum_m \frac{w_{i,m}}{\max(A_{i,m}, 1)}, \tag{39}$$

where $w_{i,m}$ is the rendering contribution of Gaussian i in training view m and $A_{i,m}$ is its projected area. HRNB uses the relative change from raw to projected-area-normalized contribution. It does not introduce an independent area-based importance score. For valid positive host scores, we compute tie-preserving rank CDF coordinates q_i^H, q_i^R, and q_i^N for the native host, raw contribution, and projected-area-normalized contribution, respectively, and define:

$$d_i = q_i^N - q_i^R. \tag{40}$$

Visibility reliability is:

$$\omega_i = \frac{v_i}{v_i + 4}, \tag{41}$$

where v_i is the number of training views in which Gaussian i is visible. Permanent Gaussian IDs are split by parity. For each held-out fold f, the complementary fold $\overline{f}$ estimates:

$$\eta^{(f)} = \frac{\sum_{i \in \overline{f}} \omega_i d_i (q_i^N - q_i^H)}{\max\left(\sum_{i \in \overline{f}} \omega_i d_i^2, \epsilon_{\text{mach}}^{(64)} \right)}, \qquad \bar{\eta}^{(f)} = \text{clip}(\eta^{(f)}, 0, 1). \tag{42}$$

The symmetric Odds-5 gate is:

$$g_5(\bar{\eta}) = \frac{\bar{\eta}^5}{\bar{\eta}^5 + (1 - \bar{\eta})^5}. \tag{43}$$

Training cameras are sorted by image name and split into two interleaved subsets $\mathcal{V}_0$ and $\mathcal{V}_1$. For each subset a, we compute raw and area-normalized contributions $C_i^{R,a}$ and $C_i^{N,a}$ using Eq. (39), and define:

$$d_i^a = F_{\text{rank}}(C_i^{N,a}) - F_{\text{rank}}(C_i^{R,a}), \tag{44}$$

where F_{rank} preserves ties. Consistency is evaluated near the native exact-K cutoff. Let p_i^H be the zero-based descending host position, with ties resolved by permanent IDs, and let $b = \max(100, \text{round}(0.05N))$. The comparison set is:

$$\mathcal{B} = \left\{ i : K - b \le p_i^H < K + b,\ C_i^{R,0}, C_i^{N,0}, C_i^{R,1}, C_i^{N,1} > 0 \right\}. \tag{45}$$

At least four valid Gaussians are required. Writing $x_i = d_i^0$, $y_i = d_i^1$, and $\mathcal{S} = \{i \in \mathcal{B} : x_i y_i > 0\}$, we compute:

$$c_{\text{corr}} = \text{clip}\left(\text{Corr}_{i\in\mathcal{B}}(x_i, y_i), 0, 1\right), \tag{46}$$

$$c_{\text{sign}} = \frac{|\mathcal{S}|}{|\mathcal{B}|}, \tag{47}$$

$$c_{\text{mag}} = \begin{cases} \dfrac{1}{|\mathcal{S}|} \displaystyle\sum_{i\in\mathcal{S}} \frac{2\min(|x_i|, |y_i|)}{\max(|x_i| + |y_i|, \epsilon_{\text{mach}}^{(64)})}, & |\mathcal{S}| > 0, \\ 0, & |\mathcal{S}| = 0. \end{cases} \tag{48}$$

Here Corr denotes Pearson correlation. The view-consistency factor and cross-fitted correction strengths are:

$$c_{\text{view}} = c_{\text{corr}} c_{\text{sign}} c_{\text{mag}}, \qquad \eta_{\text{eff}}^{(f)} = g_5(\bar{\eta}^{(f)}) c_{\text{view}}. \tag{49}$$

For log-host coordinate $\ell_i = \log S_i^{\text{host}}$, define:

$$r_i = \log C_i^{\text{norm}} - \log C_i^{\text{raw}}, \qquad \widetilde{r}_i = r_i - \text{median}(r), \tag{50}$$

and

$$s_H = 1.4826\ \max\left(\text{MAD}(\ell), \epsilon_{\text{mach}}^{(64)}\right). \tag{51}$$

The factor 1.4826 is the fixed Gaussian-consistency correction for MAD and is not a tuned hyperparameter. The calibrated log-host score is:

$$\widetilde{\ell}_i = \ell_i + s_H \tanh\left(\frac{\eta_{\text{eff}}^{(f(i))} \widetilde{r}_i}{s_H}\right). \tag{52}$$

Here $f(i)$ denotes the permanent-ID fold containing Gaussian i. Its correction coefficient is fitted on the complementary fold.

A.7 Final Host-Preserving Rank Calibration and Exact-K Selection

Let r_i^H denote the descending rank induced by $\widetilde{\ell}_i$. We define:

$$\pi_i^H = \frac{N - r_i^H + 1}{N}, \qquad q_i = \frac{r_i^H - \frac{1}{2}}{N}, \tag{53}$$

with

$$h(q_i) = q_i^2(3 - 2q_i). \tag{54}$$

The final priority is:

$$S_i^{\text{final}} = \pi_i^H \exp\left(h(q_i) R_i^{\text{geo}}\right). \tag{55}$$

For each pruning budget:

$$\mathcal{K}^\star = \text{TopK}\left(S_i^{\text{final}}, K_{\text{official}}\right), \tag{56}$$

where K_{official} is inherited from the native host mask and immutable Gaussian IDs provide deterministic tie breaking.

Table 5: Final training, probing, pruning, and recovery protocol.

Item	Value
Fine partition / permanent IDs	15K
Normal-baseline warm-up	15K–15.5K
Counterfactual measurement	15.5K–20K
Native host scoring and pruning	20K
OMG operating points	XS / X / M / L / XL
Matched recovery endpoint	30K

B Experimental and Implementation Details

Unless otherwise stated, every paired comparison shares the same source state, native host prior, Gaussian budget, post-pruning recovery schedule, and evaluation protocol. Our method changes only the retained Gaussian identities within the host-defined budget. Counterfactual response estimation uses only training-time information and does not access test-set metrics or future recovery trajectories during selection.

B.1 Datasets and Evaluation

We evaluate on Tanks & Temples, Deep Blending, and Mip-NeRF 360. Tanks & Temples contains train and truck; Deep Blending contains drjohnson and playroom; and Mip-NeRF 360 contains bicycle, bonsai, counter, flowers, garden, kitchen, room, stump, and treehill. Novel-view reconstruction quality is evaluated using PSNR, SSIM, and LPIPS.

B.2 Comparison Methods

Table 1 includes representative 3D Gaussian Splatting pruning and compression baselines to provide broader context for the quality–size trade-off. We compare against 3DGS (Kerbl et al., 2023), Scaffold-GS (Lu et al., 2024), Mini-Splatting (Fang and Wang, 2024), CompGS (Navaneet et al., 2024), Compact-3DGS (Lee et al., 2024a), C3DGS (Niedermayr et al., 2024), LightGaussian (Fan et al., 2024), PUP 3D-GS (Hanson et al., 2025b), EAGLES (Girish et al., 2024), SOG (Morgenstern et al., 2025), HAC (Chen et al., 2024), and LocoGS (Shin et al., 2025). For LocoGS, both the S and L operating configurations are reported as LocoGS-S and LocoGS-L, respectively.

OMG (Lee et al., 2025) serves as the primary host for our controlled fixed-budget comparison. We evaluate its five operating points, denoted XS, X, M, L, and XL, corresponding to importance-threshold settings 0.96, 0.98, 0.99, 0.999, and 0.9999, respectively. For every OMG operating point, the corresponding RAGP variant starts from the same source state, uses the same native OMG host prior, retains exactly the same number of Gaussians, and follows the same post-pruning recovery protocol. We therefore denote the paired variants as RAGP-XS, RAGP-X, RAGP-M, RAGP-L, and RAGP-XL. The non-OMG methods in Table 1 provide broader cross-method context because their representations, compression mechanisms, and training protocols differ. The primary controlled comparison for assessing RAGP is therefore the matched OMG–RAGP pair at each operating point.

B.3 Training and Recovery Protocol

The final training and recovery protocol is summarized in Table 5. All OMG variants follow a common training trajectory to 15K, where the scene-adaptive Fine partition and permanent Gaussian identities are fixed. From 15K to 15.5K, ordinary Normal updates are used to estimate the covariance-tangent whitening statistics. From 15.5K to 20K, Normal and read-only Masked branches are evaluated in pairs, while only the Normal branch updates the model.

At 20K, the native OMG importance score and its official budget rule are evaluated on the common source state. The Official and recoverability-aware branches then retain exactly the same number of Gaussians and follow the same post-pruning recovery schedule to 30K. We evaluate all five OMG operating points, denoted XS, X, M, L, and XL.

Table 6: Host importance criteria used in Table 2. Scores are computed from the shared OMG 20K state. Only the scoring criteria are transferred into the common pruning-and-recovery pipeline.

Host criterion	Score construction	Implementation details
Opacity	Activated per-Gaussian opacity.	Evaluated directly at the shared checkpoint.
FoV-style	$s_i = \max_v \frac{C_{iv}}{N_{iv} + 10^{-7}}$.	All training views. C_{iv} and N_{iv} are the rasterizer contribution and participation count. The view-wise score is zero if $N_{iv} < 1$.
MesonGS	$s_i = \left(\frac{V_i}{V_{\text{ref}}}\right)^{\beta} \sum_v C_{iv}^{\text{Meson}}$, where $V_i = \prod_{d=1}^{3} \sigma_{i,d}$.	All training views. V_{ref} is selected at index $\lfloor 0.9N \rfloor$ after descending volume sorting. $\beta = 0.03$ for bicycle, 0.01 for stump, and 0.1 otherwise.
GoDe grad-all	$s_i = \sum_{\theta \in \mathcal{A}} \left\| \sum_v \nabla_{\theta_i} \mathcal{L}_v \right\|_2$.	All training views. Gradients are accumulated across views before taking each attribute-group norm. $\mathcal{L}_v = 0.8\mathcal{L}_1 + 0.2(1 - \text{SSIM})$.
PUP Fisher	$s_i = \log\det(\mathbf{F}_i)$, where $\mathbf{F}_i$ is the 6×6 Fisher block for position and scaling.	All training views; `fisher_resolution`=4. Singular scores are placed below all finite scores before ranking.
PWR	Pre-DoG squared kernel sensitivity accumulated over training views, using a rendering objective weighted by detached image detail.	Detail coefficient 0.10; detail weights are mean-normalized and clipped to $[0, 3]$. Antialiasing is disabled; DoG primitives are not used.
FastGS-style	$q_i = \sum_{v \in \mathcal{V}_{10}} \ell_v N_{iv}^{\text{err}}$, $s_i = 1 - \text{minmax}(q_i)$.	Ten training views, camera seed 0. $\ell_v = 0.8\,\text{MAE} + 0.2(1 - \text{SSIM})$. Normalized pixel-error threshold 0.1; `mult`=0.5.

$\mathcal{A}$ comprises position, opacity logit, raw scaling, rotation, SH DC coefficients, and remaining SH coefficients. N_{iv}^{err} counts overlaps with high-error pixels. Raw scores are converted to tie-preserving empirical ranks before host-relative calibration. Host and Host+RAGP retain exactly the same number of Gaussians.

B.4 OMG Host-Importance Construction

To isolate recoverability-aware selection, each Official–Ours pair uses the same native OMG host prior: Outdoor-style for Tanks & Temples, Indoor-style for Deep Blending, and the native OMG Indoor/Outdoor mapping for Mip-NeRF 360. For reproducibility, we give the Outdoor-style construction used by Tanks & Temples and by the diagnostic host-prior interaction experiments below. Let $C_{i,v}$ and $A_{i,v}$ denote the rendering contribution and projected area terms used by the frozen implementation. The Outdoor-style contribution term is:

$$I_i^{\text{out}} = \sum_v \mathbf{1}[A_{i,v}^{\max} \neq 0] \frac{C_{i,v}}{A_{i,v}}, \tag{57}$$

where the indicator matches the visibility/update mask of the frozen implementation.

Let $\boldsymbol{\pi} = (\pi_0, \ldots, \pi_{N-1})$ denote the Morton-order index array returned by the official OMG implementation and let $\widetilde{\mathbf{f}}_k = \mathbf{f}_{\pi_k}$. The legacy indexing used by the reported experiments defines:

$$\ell_i = \max(\pi_i - 1, 0), \qquad r_i = \min(\pi_i + 1, N - 1), \tag{58}$$

and

$$d_i^{\text{OMG}} = \frac{1}{D} \left(\|\widetilde{\mathbf{f}}_{\ell_i} - \mathbf{f}_i\|_1 + \|\widetilde{\mathbf{f}}_{r_i} - \mathbf{f}_i\|_1 \right). \tag{59}$$

For scenes using the Outdoor-style construction, the host score is:

$$S_i^{\text{host}} = I_i^{\text{out}} \left(d_i^{\text{OMG}}\right)^{\lambda_{\text{LD}}}, \qquad \lambda_{\text{LD}} = 2. \tag{60}$$

Table 7: Hardware and software configuration used in the current experiments.

Component	Configuration
GPU	NVIDIA GeForce RTX 5090
GPU memory	32,607 MiB per GPU
CUDA runtime	12.8
PyTorch	2.11.0+cu128
NVIDIA driver	580.159.03
Python	3.11.15
CPU	Intel Core i9-14900K
CPU cores / threads	24 / 32
System memory	94 GiB

This expression intentionally mirrors the official OMG legacy index calculation used by the experiments. Replacing it with an inverse-permutation Morton-neighbor formula would define a different host prior. Appendix H separately studies how the same compensation evidence interacts with alternative OMG host-prior constructions. This analysis is diagnostic: it examines host dependence without treating either construction as intrinsically preferable.

B.5 Integration with Heterogeneous Host Importance Criteria

To isolate criterion-level generalization, all cross-importance experiments are evaluated within a unified OMG pruning-and-recovery pipeline at budget X. For each criterion, the host importance ranking is computed from the same 20K source state. The corresponding Host and Host+RAGP branches retain exactly the same number of Gaussians and follow the same OMG recovery and evaluation protocol. Only the host importance ranking is changed.

We evaluate opacity and six criteria derived from MetaSapiens (Lin et al., 2025), MesonGS (Xie et al., 2024), GoDe (Di Sario et al., 2025), PUP (Hanson et al., 2025b), Prune Wisely, Reconstruct Sharply (PWR) (Wang et al., 2026), and FastGS (Ren et al., 2026). Their transferred signals are summarized in Table 6. Source-specific pruning schedules, recovery procedures, quantization, encoding, and other compression components are not transferred.

This controlled protocol isolates whether recoverability-aware compensation evidence remains complementary when the underlying definition of Gaussian importance changes.

B.6 Hardware and Software

The hardware and software environment is summarized in Table 7.

C Complete RAGP Results

Tables 8–12 report the complete 13-scene results across the five OMG pruning budgets XS, X, M, L, and XL. All comparisons use the native OMG host mapping and exactly matched Gaussian counts. For each scene, we report Official OMG, Ours, and their differences in PSNR, SSIM, and LPIPS.

Across the five budgets, the 13-scene average PSNR gains are **+0.14762**, **+0.12083**, **+0.12909**, **+0.11162**, and **+0.06078 dB** at XS, X, M, L, and XL, respectively. PSNR improves on all 13 scenes at XS, X, M, and L, and on 12/13 scenes at XL. The corresponding average SSIM and LPIPS changes are also favorable at every budget, while scene-level perceptual metrics remain slightly heterogeneous.

D Cross-Budget Result Consistency

Table 13 summarizes the aggregate behavior across the five OMG pruning budgets. The 13-scene average PSNR improvement is positive at every budget. PSNR improves on all 13 scenes at XS, X, M, and L, and on 12/13 scenes at XL. Average SSIM and LPIPS also improve at every budget, although a small number of individual scenes exhibit perceptual-metric trade-offs.

Table 8: Complete scene-level OMG results at pruning budget XS. All Δ values are computed as Ours minus Official.

Dataset	Scene	PSNR ↑			SSIM ↑			LPIPS ↓		
		Official	Ours	Δ	Official	Ours	Δ	Official	Ours	Δ
Tanks & Temples	train	21.2055	21.6090	+0.4035	0.7949	0.7981	+0.0031	0.2520	0.2499	-0.0021
	truck	24.9385	25.2230	+0.2844	0.8711	0.8744	+0.0034	0.1595	0.1580	-0.0015
Deep Blending	drjohnson	29.3866	29.4636	+0.0770	0.9020	0.9047	+0.0027	0.2678	0.2631	-0.0047
	playroom	30.3061	30.4872	+0.1811	0.9083	0.9108	+0.0025	0.2642	0.2598	-0.0043
Mip-NeRF 360	bicycle	24.9547	24.9832	+0.0285	0.7419	0.7422	+0.0003	0.2780	0.2799	+0.0019
	bonsai	30.9010	30.9783	+0.0773	0.9321	0.9344	+0.0023	0.2079	0.2051	-0.0029
	counter	28.4659	28.5318	+0.0659	0.8999	0.9031	+0.0032	0.2081	0.2030	-0.0051
	flowers	21.3494	21.5213	+0.1719	0.5976	0.5956	-0.0020	0.3682	0.3707	+0.0025
	garden	26.2808	26.5268	+0.2460	0.8173	0.8221	+0.0048	0.1908	0.1867	-0.0042
	kitchen	30.9648	31.0115	+0.0466	0.9195	0.9207	+0.0012	0.1377	0.1347	-0.0030
	room	31.1558	31.2859	+0.1300	0.9199	0.9228	+0.0029	0.2107	0.2056	-0.0051
	stump	27.1430	27.1552	+0.0122	0.7884	0.7872	-0.0012	0.2482	0.2509	+0.0027
	treehill	22.6694	22.8642	+0.1948	0.6474	0.6467	-0.0008	0.3578	0.3598	+0.0020

Table 9: Complete scene-level OMG results at pruning budget X. All Δ values are computed as Ours minus Official.

Dataset	Scene	PSNR ↑			SSIM ↑			LPIPS ↓		
		Official	Ours	Δ	Official	Ours	Δ	Official	Ours	Δ
Tanks & Temples	train	21.2888	21.5252	+0.2365	0.8010	0.8025	+0.0015	0.2424	0.2406	-0.0018
	truck	25.1666	25.3348	+0.1682	0.8758	0.8780	+0.0022	0.1510	0.1496	-0.0014
Deep Blending	drjohnson	29.4305	29.5647	+0.1342	0.9042	0.9065	+0.0023	0.2619	0.2579	-0.0039
	playroom	30.3597	30.4877	+0.1280	0.9097	0.9113	+0.0017	0.2598	0.2557	-0.0041
Mip-NeRF 360	bicycle	25.0922	25.1155	+0.0232	0.7504	0.7510	+0.0006	0.2637	0.2638	+0.0002
	bonsai	31.1393	31.2457	+0.1064	0.9352	0.9377	+0.0025	0.1988	0.1962	-0.0026
	counter	28.5540	28.6211	+0.0672	0.9035	0.9058	+0.0023	0.2012	0.1974	-0.0038
	flowers	21.3780	21.5181	+0.1402	0.6049	0.6034	-0.0015	0.3575	0.3600	+0.0025
	garden	26.5389	26.7040	+0.1651	0.8253	0.8288	+0.0034	0.1777	0.1747	-0.0031
	kitchen	31.0135	31.1015	+0.0880	0.9221	0.9231	+0.0010	0.1331	0.1310	-0.0021
	room	31.1375	31.2217	+0.0842	0.9233	0.9247	+0.0014	0.2040	0.2010	-0.0030
	stump	27.1556	27.1944	+0.0389	0.7931	0.7924	-0.0006	0.2382	0.2397	+0.0014
	treehill	22.6671	22.8578	+0.1907	0.6507	0.6506	-0.0001	0.3461	0.3469	+0.0008

Table 10: Complete scene-level OMG results at pruning budget M. All Δ values are computed as Ours minus Official.

Dataset	Scene	PSNR ↑			SSIM ↑			LPIPS ↓		
		Official	Ours	Δ	Official	Ours	Δ	Official	Ours	Δ
Tanks & Temples	train	21.4665	21.6918	+0.2253	0.8047	0.8059	+0.0011	0.2361	0.2356	-0.0005
	truck	25.2711	25.3910	+0.1199	0.8785	0.8798	+0.0013	0.1456	0.1449	-0.0007
Deep Blending	drjohnson	29.4803	29.5797	+0.0995	0.9057	0.9075	+0.0018	0.2581	0.2550	-0.0031
	playroom	30.3692	30.5167	+0.1475	0.9102	0.9115	+0.0014	0.2566	0.2533	-0.0033
Mip-NeRF 360	bicycle	25.1021	25.1195	+0.0174	0.7550	0.7559	+0.0009	0.2559	0.2560	+0.0001
	bonsai	31.1914	31.2651	+0.0737	0.9383	0.9393	+0.0010	0.1954	0.1932	-0.0022
	counter	28.6692	28.7390	+0.0699	0.9058	0.9077	+0.0018	0.1960	0.1933	-0.0027
	flowers	21.3884	21.6727	+0.2844	0.6082	0.6083	+0.0001	0.3498	0.3514	+0.0016
	garden	26.6620	26.8184	+0.1564	0.8316	0.8341	+0.0025	0.1699	0.1671	-0.0027
	kitchen	31.1289	31.2084	+0.0795	0.9238	0.9244	+0.0006	0.1304	0.1290	-0.0014
	room	31.1761	31.4003	+0.2242	0.9223	0.9243	+0.0020	0.2020	0.1997	-0.0023
	stump	27.1954	27.2151	+0.0197	0.7951	0.7952	+0.0001	0.2327	0.2335	+0.0008
	treehill	22.6543	22.8154	+0.1611	0.6523	0.6529	+0.0006	0.3395	0.3386	-0.0009

Table 11: Complete scene-level OMG results at pruning budget L. All Δ values are computed as Ours minus Official.

Dataset	Scene	PSNR ↑			SSIM ↑			LPIPS ↓		
		Official	Ours	Δ	Official	Ours	Δ	Official	Ours	Δ
Tanks & Temples	train	21.5867	21.7150	+0.1283	0.8109	0.8119	+0.0010	0.2262	0.2263	+0.0002
	truck	25.3736	25.4467	+0.0731	0.8819	0.8824	+0.0005	0.1379	0.1377	-0.0002
Deep Blending	drjohnson	29.5474	29.6486	+0.1012	0.9079	0.9088	+0.0009	0.2520	0.2508	-0.0012
	playroom	30.5349	30.6291	+0.0942	0.9130	0.9141	+0.0010	0.2522	0.2496	-0.0026
Mip-NeRF 360	bicycle	25.1697	25.2002	+0.0305	0.7620	0.7623	+0.0003	0.2419	0.2427	+0.0008
	bonsai	31.3264	31.4030	+0.0766	0.9412	0.9416	+0.0004	0.1890	0.1873	-0.0017
	counter	28.7437	28.7982	+0.0546	0.9086	0.9095	+0.0009	0.1903	0.1888	-0.0015
	flowers	21.3972	21.6791	+0.2818	0.6133	0.6144	+0.0011	0.3383	0.3393	+0.0010
	garden	26.8430	26.9107	+0.0677	0.8380	0.8389	+0.0009	0.1596	0.1582	-0.0014
	kitchen	31.2237	31.3045	+0.0808	0.9258	0.9259	+0.0001	0.1270	0.1263	-0.0007
	room	31.2188	31.4872	+0.2683	0.9246	0.9259	+0.0013	0.1966	0.1949	-0.0018
	stump	27.1079	27.1583	+0.0504	0.7961	0.7965	+0.0004	0.2264	0.2267	+0.0003
	treehill	22.7797	22.9233	+0.1436	0.6537	0.6548	+0.0011	0.3295	0.3286	-0.0009

Table 12: Complete scene-level OMG results at pruning budget XL. All Δ values are computed as Ours minus Official.

Dataset	Scene	PSNR ↑			SSIM ↑			LPIPS ↓		
		Official	Ours	Δ	Official	Ours	Δ	Official	Ours	Δ
Tanks & Temples	train	21.6558	21.7172	+0.0614	0.8118	0.8122	+0.0004	0.2241	0.2241	+0.0000
	truck	25.4201	25.4747	+0.0546	0.8829	0.8831	+0.0002	0.1355	0.1356	+0.0001
Deep Blending	drjohnson	29.5787	29.6639	+0.0852	0.9086	0.9091	+0.0006	0.2504	0.2499	-0.0005
	playroom	30.5562	30.6640	+0.1078	0.9134	0.9141	+0.0007	0.2510	0.2486	-0.0024
Mip-NeRF 360	bicycle	25.2611	25.2587	-0.0025	0.7635	0.7635	0.0000	0.2385	0.2389	+0.0004
	bonsai	31.4143	31.4362	+0.0219	0.9420	0.9422	+0.0001	0.1874	0.1857	-0.0017
	counter	28.7512	28.8153	+0.0641	0.9093	0.9099	+0.0006	0.1889	0.1876	-0.0013
	flowers	21.4855	21.5974	+0.1119	0.6140	0.6150	+0.0010	0.3337	0.3351	+0.0014
	garden	26.8343	26.8723	+0.0380	0.8392	0.8395	+0.0003	0.1567	0.1564	-0.0004
	kitchen	31.2845	31.3251	+0.0406	0.9263	0.9263	-0.0000	0.1262	0.1257	-0.0005
	room	31.3209	31.4392	+0.1183	0.9269	0.9274	+0.0005	0.1928	0.1920	-0.0009
	stump	27.1053	27.1406	+0.0353	0.7961	0.7966	+0.0005	0.2246	0.2249	+0.0003
	treehill	22.7008	22.7543	+0.0536	0.6538	0.6547	+0.0010	0.3253	0.3245	-0.0008

Table 13: Aggregate consistency across five OMG pruning budgets.

Budget	ΔPSNR ↑	ΔSSIM ↑	ΔLPIPS ↓	PSNR-positive scenes
XS	+0.14762	+0.001722	-0.001839	13/13
X	+0.12083	+0.001289	-0.001609	13/13
M	+0.12909	+0.001164	-0.001326	13/13
L	+0.11162	+0.000761	-0.000742	13/13
XL	+0.06078	+0.000451	-0.000479	12/13

Averaged across the five budget-level results, the PSNR improvement is $+0.11399$ dB. More importantly, the improvement remains positive at every evaluated operating point, while the average SSIM and LPIPS directions are also favorable throughout the budget range. Scene-level perceptual metrics, however, remain slightly heterogeneous.

E Cross-Importance Generalization Results

Table 14 summarizes aggregate cross-importance results under the unified protocol described in Appendix B.5. All comparisons use budget X and share the same 20K source state, exact Gaussian

Table 14: Cross-importance generalization over all 13 scenes at budget X. Quality changes are computed as RAGP minus the corresponding host criterion.

Host criterion	ΔPSNR ↑	ΔSSIM ↑	ΔLPIPS ↓	ΔFPS ↑
Opacity	+0.153	+0.0020	+0.0013	-2.07%
FoV-style	+0.218	+0.0039	-0.0039	-0.54%
MesonGS	+0.148	+0.0058	-0.0058	+6.67%
GoDe grad-all	+0.261	+0.0050	-0.0064	-0.77%
PUP Fisher	+0.513	+0.0109	-0.0113	-1.90%
PWR	+0.080	+0.0037	-0.0071	+1.69%
FastGS-style	+2.064	+0.0380	-0.0433	-7.13%

budget, OMG recovery schedule, and evaluation protocol. The source methods contribute only their importance criteria. Their native pruning or compression pipelines are not transferred.

The 13-scene average PSNR change is positive under all seven importance criteria. Average SSIM also improves for every criterion. Average LPIPS improves for six criteria, while the Opacity criterion exhibits a small perceptual trade-off. The aggregate FPS change varies in both directions across criteria, indicating no systematic increase in rendering cost.

F Additional Recoverability and Compensation Analysis

Algorithm 1 summarizes the recoverability-aware pruning procedure described in Sec. 4, including counterfactual response measurement, host-relative calibration, and exact-K Gaussian selection.

Algorithm 1 Recoverability-Aware Gaussian Pruning (RAGP)

Require: Gaussian set $\mathcal{G}$, training views $\mathcal{V}$, host score S^{host}, native pruning rule
Ensure: Retained Gaussian set $\mathcal{G}^\star$

Train to 15K and construct adaptive Fine groups
Estimate Normal tangent statistics during 15K–15.5K
Fine Response Estimation:
for $t = 15.5\text{K}, \ldots, 20\text{K}$ **do**
  Evaluate paired Normal and Masked branches
  Compute and whiten geometric gradient contrast
  if the paired response is valid **then**
    Update recoverability response R_i^{geo}
  end if
  Apply only the Normal optimizer update
end for
Freeze R_i^{geo} and the 20K source state
Obtain the native host mask and exact budget K
Host Calibration:
Compute raw and area-normalized contributions
Form host/raw/normalized ranks (q^H, q^R, q^N)
Estimate host-relative correction by cross-fitting
if cross-view consistency is satisfied **then**
  Apply the bounded HRNB correction
else
  Retain the native host calibration
end if
Compute final score $S_i^{\text{final}} = \pi_i^H \exp(h(q_i) R_i^{\text{geo}})$
$\mathcal{G}^\star \leftarrow \text{TopK}(S^{\text{final}}, K)$
return $\mathcal{G}^\star$

F.1 Recoverability across Recovery Schedules

Pairwise recoverability agreement across recovery-camera schedules is reported in Table 15. We keep the checkpoint, optimizer, learning-rate schedule, evaluation views, and interventions fixed and vary only the recovery-camera schedule. Each comparison contains $n = 23$ valid interventions. Agreement is stronger at the longer horizon, indicating that normalized recoverability contains reproducible structure that is not specific to a single recovery-camera schedule.

Table 15: Pairwise agreement of normalized recoverability across recovery-camera schedules.

Horizon	Schedule pair	Spearman ρ	95% CI
100	seed 0 vs. 1	0.831	[0.524, 0.991]
	seed 0 vs. 2	0.480	[-0.069, 0.948]
	seed 1 vs. 2	0.432	[-0.115, 0.933]
500	seed 0 vs. 1	0.818	[0.491, 0.985]
	seed 0 vs. 2	0.755	[0.415, 0.930]
	seed 1 vs. 2	0.637	[0.178, 0.925]

F.2 High-Response Freezing Intervention

Aggregate freezing effects are reported in Table 16, with per-scene results in Table 17. We freeze the top 10% of retained Gaussians according to the frozen R_i^{geo} ranking and compare them with low-response and equal-count random controls. Each scene contains 12 intervention masks, with three random controls per mask. Freezing applies only to the selected Gaussian parameter rows and their Adam moments in the Masked branch. All conditions share the same Normal reference trajectory, which evolves without the freezing intervention. Densification and topology changes are disabled. Selection does not use test-set metrics or future optimization trajectories.

Local recovery measure. For intervention U, let $I_N^t(v,p)$ and $I_{M,b}^t(v,p)$ denote the RGB renderings at pixel p in view v, where $b \in \{\text{High}, \text{Random}, \text{Low}\}$ identifies the freezing condition. The fixed evaluation set $\mathcal{V}_{\text{eval}}$ contains training views excluded from the recovery-camera schedule. The initial renderings are shared across freezing conditions. We construct the damage footprint as:

$$a_U(v,p) = \frac{1}{3}\left\|I_M^0(v,p) - I_N^0(v,p)\right\|_1, \qquad Z_U = \sum_{v\in\mathcal{V}_{\text{eval}}}\sum_p a_U(v,p),$$

and define:

$$w_U(v,p) = \begin{cases} a_U(v,p)/Z_U, & Z_U > 10^{-12}, \\ 0, & \text{otherwise.} \end{cases}$$

These weights define a soft local region without a manually selected crop and remain fixed throughout recovery. For branch B, the local reconstruction loss is:

$$L_B^{t,\text{local}}(U) = \sum_{v\in\mathcal{V}_{\text{eval}}}\sum_p w_U(v,p)\,\frac{\|I_B^t(v,p) - I_{\text{GT}}(v,p)\|_1}{3}.$$

Suppressing U for brevity, we define:

$$D_{t,b}^{\text{local}} = L_{M,b}^{t,\text{local}} - L_N^{t,\text{local}}, \qquad G_{T,b}^{\text{local}} = D_0^{\text{local}} - D_{T,b}^{\text{local}}.$$

Thus, $G_{T,b}^{\text{local}}$ measures the reduction in the initial local loss gap, rather than the normalized recoverability $\mathcal{C}_T$ defined in Sec. 3. Zero-footprint interventions, if present, contribute zero and remain in the average.

Aggregation and uncertainty. Fig. 2(b) reports $G_{100,b}^{\text{local}}$. Tables 16 and 17 report the paired differences $G_{T,\text{High}}^{\text{local}} - G_{T,\text{Random}}^{\text{local}}$ and $G_{T,\text{High}}^{\text{local}} - G_{T,\text{Low}}^{\text{local}}$. Negative differences indicate weaker local recovery under High freezing. Random first averages its three controls within each mask. We then average 12 masks within each scene and the 13 scene means equally. Confidence intervals use the

Table 16: High-response freezing intervention. Negative differences indicate weaker local recovery after freezing high-response Gaussians.

T	Comparison	Mean	Median	95% CI	Neg. scenes
100	High – Random	-0.031	-0.030	[-0.042, -0.020]	12/13
	High – Low	-0.032	-0.030	[-0.043, -0.020]	12/13
500	High – Random	-0.011	-0.007	[-0.016, -0.007]	12/13
	High – Low	-0.011	-0.008	[-0.017, -0.007]	12/13

Table 17: Per-scene paired differences in local recovery gain. Negative values indicate weaker recovery under High freezing.

Dataset	Scene	T	High–Random	High–Low
Tanks & Temples	train	100	-0.059	-0.059
		500	-0.007	-0.008
	truck	100	-0.053	-0.054
		500	-0.030	-0.031
Deep Blending	drjohnson	100	-0.002	-0.002
		500	-0.001	-0.001
	playroom	100	-0.033	-0.034
		500	-0.009	-0.009
Mip-NeRF 360	bicycle	100	-0.047	-0.050
		500	-0.026	-0.026
	bonsai	100	-0.030	-0.030
		500	-0.006	-0.006
	counter	100	-0.047	-0.048
		500	-0.018	-0.018
	flowers	100	-0.026	-0.026
		500	-0.015	-0.016
	garden	100	-0.025	-0.025
		500	-0.007	-0.006
	kitchen	100	-0.013	-0.014
		500	-0.005	-0.005
	room	100	-0.060	-0.061
		500	-0.019	-0.019
	stump	100	0.000	0.000
		500	0.000	0.000
	treehill	100	-0.008	-0.009
		500	-0.004	-0.004

2.5th and 97.5th percentiles of 20,000 bootstrap means obtained by resampling scenes with replacement. For Fig. 2(b), these are computed from scene-level recovery gains; for Table 16, they are computed from scene-level paired differences.

All four comparisons yield a one-sided Wilcoxon signed-rank test with $p < 0.001$ at the scene level. Freezing high-response Gaussians weakens subsequent local recovery relative to equal-count random and low-response controls, supporting the counterfactual response as evidence of compensatory participation. Stump is the only scene whose average direction differs from the remaining 12 scenes. The intervention uses a single freeze fraction of 10%. We therefore do not claim a dose–response relationship across freeze ratios.

G Detailed Ablation Studies

Unless otherwise specified, all ablations use matched source states, host priors, Gaussian budgets, post-pruning recovery schedules, and evaluation protocols. We first report component and design controls evaluated within the final pipeline at budget X. Additional diagnostic suites examine response integration, geometric representation, spatial granularity, multi-scale aggregation, and coverage feasibility. When diagnostic suites use independent matched source chains, comparisons should be interpreted within each suite rather than across tables.

G.1 Final-Pipeline Component and Design Controls

Table 18 evaluates the principal components and design choices within the final pipeline at budget X. All variants share the same 20K source state, native host prior, exact-K budget, recovery schedule, and evaluation protocol.

Table 18: Final-pipeline component and design controls at OMG budget X over all 13 scenes. All Δ metrics are relative to the paired Official branch.

Variant	ΔPSNR ↑	ΔSSIM ↑	ΔLPIPS ↓
Baseline (Official OMG)	0	0	0
A2: Whitened Fine only	+0.10592	+0.000656	-0.000444
A3: Whitened Fine + Odds-5/View	+0.10478	+0.001226	-0.001612
A4: HRNB only (w/o Fine)	+0.01233	+0.000499	-0.000802
RAGP (Full)	**+0.12083**	**+0.001289**	**-0.001609**
Gradient norm + Host-rank	+0.09530	+0.000133	+0.000313
Covariance tangent + Host-rank	+0.10748	+0.000564	-0.000249
A1: Whitened + Uniform	+0.09715	-0.000931	+0.002017
Whitened w/o reliability	+0.09223	+0.000495	-0.000316

The Baseline, A1–A4, and RAGP (Full) configurations correspond to those in Fig. 5. This table reports 13-scene mean differences at budget X, whereas the figure shows dataset-level absolute results across five budgets.

Whitened Fine provides the dominant standalone improvement, whereas HRNB alone produces a smaller positive gain. Their combination increases the 13-scene average ΔPSNR from $+0.10592$ to $+0.12083$ dB. The remaining controls show that gradient magnitude, the unwhitened covariance-tangent response, uniform integration, and removal of the reliability constraint are all weaker than the final configuration.

G.2 Final-Mask Scope of Host-Relative Calibration

We further examine how much the host-relative calibration changes the final retained Gaussian set. Table 19 compares Full with both the native Official selection and Whitened Fine alone. All comparisons use permanent Gaussian identities and exactly matched retained counts.

Table 19: Scope of the final-mask modification. Rank shifts are measured in ordinal rank coordinates relative to the corresponding pre-calibration ranking.

Comparison	Promoted	Swap/K	Overlap/K	Promoted shift	Demoted shift
Official → Full	740385	0.1302	0.8698	-0.0285	+0.0472
Fine → Full	195818	**0.0411**	**0.9589**	-0.0608	+0.0605

Relative to Whitened Fine alone, the full method exchanges only **4.11%** of the retained set on average and preserves **95.89%** overlap. This confirms that the host-relative calibration acts as a bounded local adjustment rather than rewriting the compensation-aware selection.

G.3 Response-Only and Calibration Diagnostic

In a separate matched diagnostic suite, we test whether the frozen counterfactual response should replace or complement a fixed host ranking. Table 20 reports this comparison. This experiment isolates the interpretation of the response signal and is not intended to be numerically compared with the final-pipeline ablation in Table 18.

Table 20: Response-integration diagnostic at OMG budget X over all 13 scenes. Δ metrics are relative to the paired Official branch within this matched diagnostic suite.

Variant	ΔPSNR ↑	ΔSSIM ↑	ΔLPIPS ↓
Response-only ranking	-0.098	-0.0189	+0.0276
Uniform calibration	**+0.222**	-0.0005	+0.0013
Host-rank calibration	+0.220	**+0.0015**	**-0.0021**

Response-only ranking degrades all three quality metrics on average, demonstrating that counterfactual response should not replace primitive importance. Incorporating the same response as complementary evidence to a host ranking instead yields positive reconstruction gains. This diagnostic therefore supports interpreting the response as evidence of compensatory participation rather than as a standalone importance score.

G.4 Geometry Response Construction

This representation study is conducted as an independent matched diagnostic suite. Table 21 compares the alternative geometric-response constructions and should be interpreted within this suite rather than numerically compared with Table 18.

Table 21: Effect of geometric-response representation. All variants use the same Fine probing protocol, host-rank calibration, exact Gaussian budget, and recovery configuration.

Response	ΔPSNR ↑	ΔSSIM ↑	ΔLPIPS ↓
Gradient-norm response	+0.2153	+0.00113	-0.00147
Covariance tangent, no whitening	+0.2140	+0.00148	-0.00197
Covariance tangent + whitening	**+0.2200**	**+0.00154**	**-0.00207**

The gradient-norm response provides a strong baseline. Mapping scale and rotation changes to the covariance tangent space produces comparable PSNR while improving SSIM and LPIPS. Whitening the covariance-tangent response using Normal-branch statistics yields the strongest result in this suite. These comparisons support covariance whitening as a normalization of optimization variation, rather than claiming that covariance parameterization alone produces a large PSNR gain.

G.5 Spatial Granularity

In a separate matched diagnostic suite, Fine, Medium, and Coarse are evaluated as independent strict one-level measurements. Quality changes are reported in Table 22, while the corresponding partition and scheduled-probe statistics are reported in Table 23.

All variants share the same source-state protocol, measurement window, minimum-observation requirement, response estimator, host calibration, exact-K pruning budget, and recovery procedure. Because the number and visibility of eligible cells vary with spatial granularity, each variant constructs its own coverage-feasible adaptive schedule. The resulting probe counts are therefore not forced to be identical and are reported explicitly in Table 23.

This comparison evaluates the complete granularity-specific probing configurations under a common coverage rule. It should not be interpreted as an iso-compute ablation. All listed configurations satisfy the frozen coverage criterion, with zero undercovered eligible cells. Realized probe counts differ because each granularity constructs its own coverage-feasible adaptive schedule. This table therefore characterizes matched coverage rather than matched compute.

Table 22: Spatial granularity under a shared scene-adaptive construction and the same minimum-coverage criterion. Each granularity uses its own coverage-feasible schedule. Realized probe counts are reported in Table 23.

Granularity	ΔPSNR ↑	ΔSSIM ↑	ΔLPIPS ↓
Fine ($R^\star$)	**+0.24205**	**+0.001637**	**-0.002119**
Medium	+0.22675	+0.001374	-0.001812
Coarse	+0.20462	+0.001101	-0.001554

Table 23: Partition and scheduled probing statistics for the spatial-granularity study, obtained from the scheduler preflight. For each scene and granularity, we report the selected adaptive resolution $R^\star$, occupied and eligible cell counts, scheduled probes, and the minimum scheduled observations per eligible cell. The last column reports a scheduling outcome, not the minimum coverage requirement. Probe counts may differ across granularities.

Dataset	Scene	Granularity	$R^\star$	Occupied cells	Eligible cells	Probes	Minimum observations
Tanks & Temples	train	Fine	11	187	79	702	2
		Medium	11	54	28	651	2
		Coarse	11	14	11	639	4
	truck	Fine	22	1,227	1,038	2,196	2
		Medium	22	302	257	904	2
		Coarse	22	80	67	637	3
Deep Blending	drjohnson	Fine	23	1,364	1,310	2,636	2
		Medium	23	326	322	1,013	2
		Coarse	23	82	82	637	5
	playroom	Fine	19	1,309	1,284	2,594	2
		Medium	19	298	297	971	2
		Coarse	19	71	71	637	8
Mip-NeRF 360	bicycle	Fine	27	1,204	1,047	2,209	2
		Medium	27	319	279	941	2
		Coarse	27	88	81	641	4
	bonsai	Fine	23	1,318	1,293	2,609	2
		Medium	23	289	287	954	2
		Coarse	23	72	71	640	4
	counter	Fine	21	903	827	1,852	2
		Medium	21	221	207	819	2
		Coarse	21	59	54	639	3
	flowers	Fine	26	1,470	1,318	2,649	2
		Medium	26	363	332	1,030	2
		Coarse	26	91	83	637	5
	garden	Fine	20	483	448	1,221	2
		Medium	20	134	131	696	2
		Coarse	20	40	39	637	11
	kitchen	Fine	14	227	201	808	2
		Medium	14	69	61	646	3
		Coarse	14	19	17	637	13
	room	Fine	35	1,223	1,121	2,331	2
		Medium	35	343	318	1,006	2
		Coarse	35	92	87	646	3
	stump	Fine	20	911	793	1,804	2
		Medium	20	217	195	838	2
		Coarse	20	56	55	637	7
	treehill	Fine	39	1,524	1,277	2,583	2
		Medium	39	441	373	1,099	2
		Coarse	39	126	109	653	3

Table 24: Representative multi-scale aggregation strategies under a fixed hierarchy. All Δ metrics are 13-scene averages relative to the paired Official branch.

Aggregation strategy	ΔPSNR ↑	ΔSSIM ↑	ΔLPIPS ↓
Fine only	**+0.2421**	**+0.00164**	**-0.00212**
Arithmetic mean	+0.2193	+0.00144	-0.00198
Reliability-weighted mean	+0.2249	+0.00151	-0.00209
Weighted median	+0.2238	+0.00140	-0.00189

G.6 Representative Multi-Scale Aggregation Strategies

Multi-scale probing is not a unique baseline because both the hierarchy and the fusion rule can vary. We therefore fix the same Fine/Medium/Coarse hierarchy used above and compare several representative deterministic aggregation rules. These experiments test whether combining broader spatial context improves the response under our framework. They do not claim to exhaust the multi-scale design space. The resulting comparison is reported in Table 24.

Fine-only probing gives the strongest aggregate change across all three quality metrics among these representative aggregation rules. The multi-scale fusion variants remain competitive but do not improve over the Fine-only response in this study. We therefore retain the single Fine response in the final method, without claiming that these rules exhaust the multi-scale design space.

G.7 Coverage Feasibility Around the Adaptive Fine Resolution

The adaptive rule selects the finest contiguous resolution satisfying the frozen minimum-observation requirement. We examine the immediately adjacent resolutions without changing the camera schedule or coverage criterion. The coarser neighbor $R^\star - 1$ remains feasible in all scenes, whereas $R^\star + 1$ produces undercovered eligible cells in every evaluated scene. We therefore interpret $R^\star$ as the empirical coverage-feasibility boundary and do not report reconstruction results for the infeasible finer candidate. The adjacent-resolution analysis is summarized in Table 25.

Table 25: Coverage feasibility around the selected Fine-grid resolution. $R^\star$ is the finest feasible resolution under the frozen camera schedule and minimum-observation constraint.

Dataset	Scene	$R^\star - 1$	Selected $R^\star$	$R^\star + 1$	Undercovered cells at $R^\star + 1$
Tanks & Temples	train	10	11	12	3
	truck	21	22	23	26
Deep Blending	drjohnson	22	23	24	87
	playroom	18	19	20	138
Mip-NeRF 360	bicycle	26	27	28	3
	bonsai	22	23	24	65
	counter	20	21	22	2
	flowers	25	26	27	81
	garden	19	20	21	33
	kitchen	13	14	15	1
	room	34	35	36	33
	stump	19	20	21	54
	treehill	38	39	40	26

H Interaction with Host-Prior Construction

We examine how Whitened Fine compensation evidence interacts with Indoor-style and Outdoor-style OMG host priors, and whether HRNB provides an additional benefit. The diagnostic experiments cover the four indoor scenes of Mip-NeRF 360 and the two Deep Blending scenes at budget X. Within each scene, we keep the source state, Fine response, Gaussian budget, and recovery protocol

fixed while varying the host construction and enabling or disabling HRNB. The primary experiments retain the native OMG host mapping.

H.1 Host-Prior and Fine Interaction

Table 26 compares each host alone, with Fine, and with Fine plus HRNB. Fine denotes the complete whitened response. HRNB is enabled only in the Host+Fine+HRNB configurations.

Table 26: Host-prior interaction and incremental HRNB effects at budget X. We report mean PSNR (dB) on the four indoor scenes of Mip-NeRF 360 and the two Deep Blending scenes. Δ_{HRNB} denotes Host+Fine+HRNB minus the corresponding Host+Fine result.

Configuration	Mip-NeRF 360 indoor	Deep Blending
Indoor Host	30.46793	29.92078
Indoor + Fine	30.48862	30.00451
Indoor + Fine + HRNB	**30.50101**	**30.01498**
Outdoor Host	30.22100	29.79660
Outdoor + Fine	30.47985	29.98318
Outdoor + Fine + HRNB	30.48088	30.00279
Δ_{HRNB} (Indoor)	+0.01239	+0.01047
Δ_{HRNB} (Outdoor)	+0.00103	+0.01961

Both datasets exhibit the same qualitative pattern. Fine produces larger PSNR gains under the Outdoor-style host: 0.25885 dB on Mip-NeRF 360 indoor and 0.18658 dB on Deep Blending, compared with 0.02069 and 0.08373 dB under the Indoor-style host. Nevertheless, Indoor+Fine retains higher final PSNR on both datasets. The larger Outdoor gains therefore close much of the initial gap between the host constructions without reversing their final PSNR ordering. These results show that the benefit of compensation evidence depends on the host prior. A larger improvement alone does not justify replacing the native host.

H.2 Incremental Effect of Host-Relative Calibration

Adding HRNB improves mean PSNR for both host constructions on both datasets. Under the native Indoor-style host, HRNB adds 0.01239 dB on Mip-NeRF 360 indoor and 0.01047 dB on Deep Blending. Under the Outdoor-style host, the corresponding gains are 0.00103 and 0.01961 dB, indicating that the magnitude of the benefit varies with the host and dataset. Indoor+Fine+HRNB achieves the highest PSNR among the evaluated configurations on both datasets.

Together, these results motivate retaining the native host prior while calibrating its integration with compensation evidence. The host comparison establishes that Fine gains are host-dependent, while the HRNB ablation demonstrates an additional benefit beyond Fine-based selection. This supports HRNB as a complementary, bounded calibration rather than a replacement for the native importance criterion.

I Qualitative Results and Metric Trade-offs

I.1 Representative Qualitative Results

We provide additional qualitative comparisons on Tanks & Temples, Deep Blending, and Mip-NeRF 360. Within each comparison, Official OMG and Ours use the same retained Gaussian budget and are rendered from the same test view with identical crops. These examples complement the quantitative results by illustrating local reconstruction differences that may be difficult to capture with scene-level averages alone. The corresponding comparisons are shown in Figs. 7, 8, and 9.

Qualitative evaluation protocol. PSNR annotations in Figs. 1, 6, 7, 8, 9, and 10 refer to the full image at the displayed viewpoint, rather than the enlarged crop or a scene average. The comparisons in Figs. 6, 7, 8, 9, and 10 use budget X (0.98).

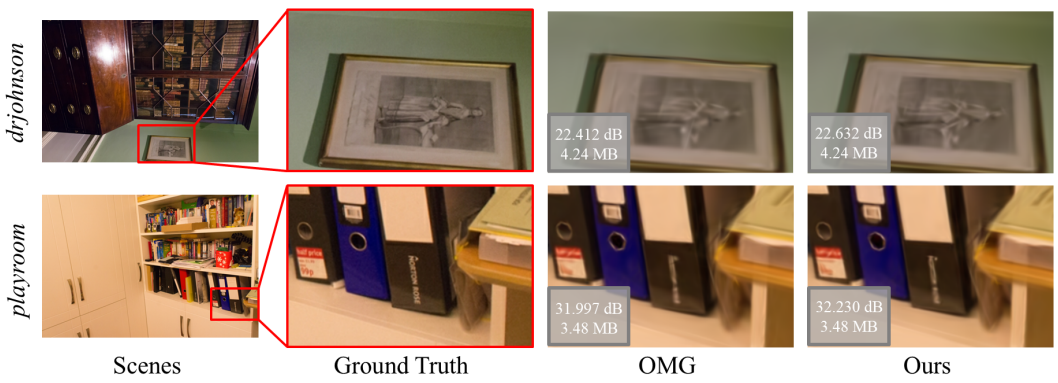

Figure 7: **Qualitative comparison on Deep Blending.** Representative ground-truth, Official OMG, and Ours renderings under a matched Gaussian budget.

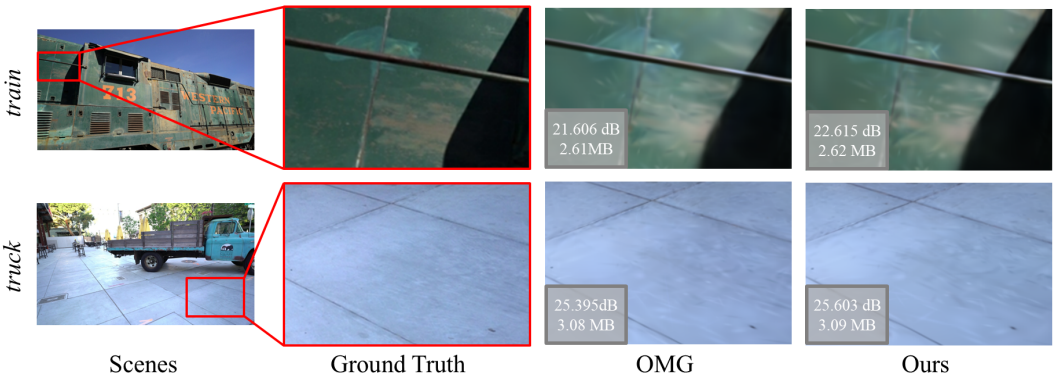

Figure 8: **Qualitative comparison on Tanks & Temples.** Representative ground-truth, Official OMG, and Ours renderings under a matched Gaussian budget.

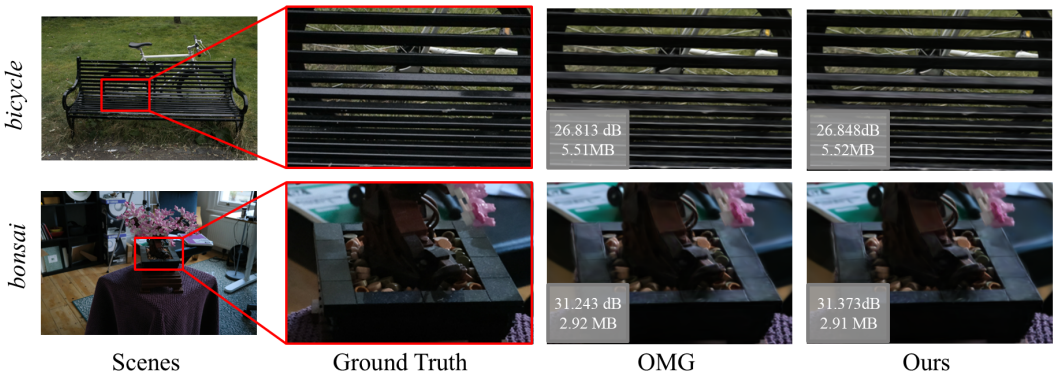

Figure 9: **Qualitative comparison on Mip-NeRF 360.** Representative ground-truth, Official OMG, and Ours renderings under a matched Gaussian budget.

I.2 Qualitative Comparisons across Importance Criteria

Fig. 10 presents qualitative comparisons for selected host importance criteria at budget X. For each criterion, Host and Host+RAGP share the same Gaussian budget, recovery schedule, test view, and crop regions. These comparisons complement the quantitative results in Tables 2 and 14. Only the importance criteria are transferred into the common OMG pruning-and-recovery pipeline, not the source methods' complete pipelines.

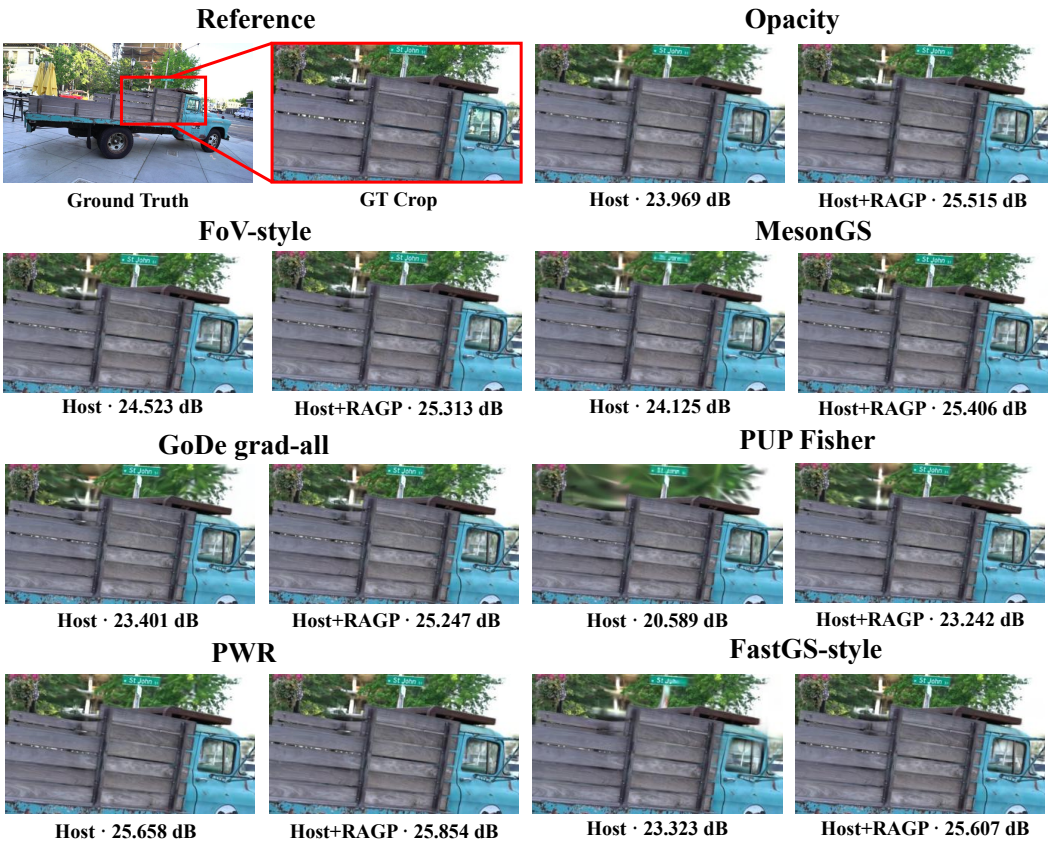

Figure 10: **Qualitative comparisons across host importance criteria on the truck scene from Tanks & Temples.** Ground-truth, Host, and Host+RAGP renderings are shown at budget X. Each comparison is labeled by its host criterion. Corresponding crops show the same image region, with Host and Host+RAGP retaining exactly the same number of Gaussians.

Table 27: Training-time efficiency of the complete method on three representative scenes. Runtime covers optimization and measurement through the matched 30K recovery endpoint, including their required rendering operations, but excludes final test-view rendering and metric evaluation. Peak VRAM is the maximum memory usage over the sequential Whitened Fine and HRNB stages.

Scene	Official (s)	Ours (s)	Extra time (s)	Overhead	Peak VRAM (MiB)
train	652.10	843.67	+191.57	+29.38%	7454
playroom	705.37	947.20	+241.83	+34.28%	8788
room	1073.68	1550.34	+476.66	+44.40%	13228
Mean	–	–	–	**+36.02%**	–

J Efficiency Analysis

We evaluate the training-time cost of the complete method on three representative scenes. Whitened Fine performs the counterfactual measurement during training, after which HRNB computes training-view contribution statistics and the final host-relative calibration. Because these stages are executed sequentially under the same profiling environment, their runtimes are combined to obtain the total training time of the final pipeline. The complete profiling results are reported in Table 27.

Across the three representative scenes, the complete method introduces an average training-time overhead of **36.02%** relative to Official. Most of this additional cost comes from the counterfactual Whitened Fine measurement stage, while the subsequent host-relative calibration contributes only a small fraction of the total runtime.

HRNB does not increase the overall peak-memory requirement in our profiling. Its contribution-statistics and score-materialization stages are executed after the Whitened Fine measurement and exhibit lower peak memory than the preceding Fine stage on all three representative scenes. Consequently, the peak memory of the complete pipeline is determined by Whitened Fine rather than by the host-relative calibration stage.

The additional computation is confined to training-time measurement and selection. After pruning, the method preserves the host Gaussian count and introduces no additional Gaussian attributes, learned parameters, or renderer operations. Inference therefore uses the same Gaussian representation and rendering pipeline as the corresponding native host.