GENERAL POST-TRAINING 3DGS COMPRESSION: COORDINATING QUANTIZATION AND SEARCH FOR QUALITY-CONSTRAINED SIZE MINIMIZATION

Anonymous authors
Paper under double-blind review

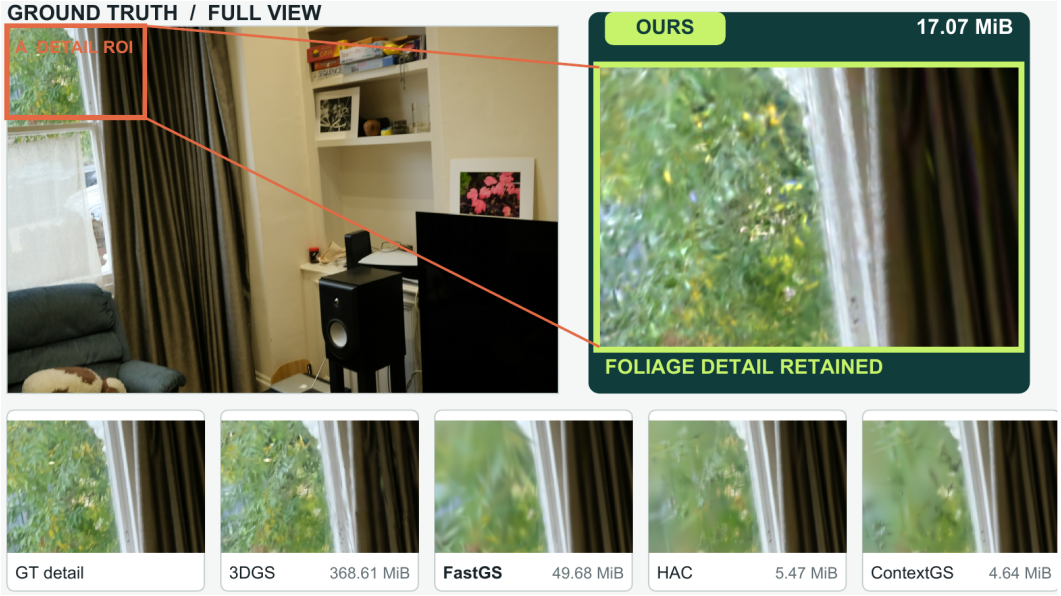


Figure 1: Visual comparison on the Mip-NeRF 360 room scene. The orange box marks the region enlarged for every method. Our delivered representation retains visible foliage detail at 17.07 MiB, compared with 49.68 MiB for FastGS. Sizes refer to complete scene representations.

ABSTRACT

The large file sizes of trained 3D Gaussian Splatting (3DGS) scenes increase storage and distribution costs. We propose a general post-training compression framework to reduce model size under rendering-quality constraints and a limited evaluation budget. Our key idea is to coordinate nonuniform quantization with configuration search. We apply signed-power companding to higher-order spherical harmonic coefficients within spatial blocks. This reduces quantization error and allows lower precision or larger blocks within the quality limits. The search jointly selects the pretrained base model, block size, and coefficient precision without retraining or fine-tuning. An independent view set checks the selected model against mean and tail PSNR-loss limits before delivery. Failed confirmation triggers a lossless reference fallback. Across 192 matched configurations, companding increases the number of candidates satisfying the search quality criteria from 75 to 119. At predetermined default settings on eight benchmark scenes, our complete system reduces mean model size by 97.37% and 79.60% compared with 3DGS and FastGS, respectively. Mean PSNR improves by 0.257 dB and 0.173 dB, respectively. The average rendering speed reaches 1,285.5 FPS.

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.

1 INTRODUCTION

3D Gaussian Splatting (3DGS) enables high-quality, real-time novel-view rendering (Kerbl et al., 2023). Once a scene has been trained, its model must be stored and distributed for deployment. Explicit Gaussian attributes can require hundreds of megabytes per scene, making storage and transmission costly (Niedermayr et al., 2024). This motivates post-training compression: turn an existing scene model into a compact deliverable while keeping visual quality within an acceptable loss.

Existing methods reduce storage through pruning, quantization, and compact representations (Lee et al., 2024; Fan et al., 2024; Niedermayr et al., 2024). Context models and rate–distortion optimization improve coding efficiency (Chen et al., 2025b; Wang et al., 2025). Feedforward methods also compress pretrained scenes without per-scene fine-tuning (Chen et al., 2025a). Deployment also requires choosing a compression configuration for each scene. We ask: *how can we obtain a smaller model within an acceptable quality loss and a limited number of candidate evaluations?*

The choice involves coupled decisions. Base-model size, spatial block size, and spherical harmonic (SH) bit widths jointly affect storage and decoded quality. A smaller base leaves fewer bytes to encode, but can also leave less room for quantization error. Lower SH precision saves bits, but may violate the quality limits. Each queried candidate must be encoded, decoded, and rendered to measure these effects. Quantization and search should therefore work together: better quantization makes more configurations feasible, and search uses a limited budget to find a small feasible combination.

We present a *general post-training compression framework* built around this coordination. Its compression mechanism operates on fixed Gaussian attributes and does not depend on a specific training objective. Our implementation supports degree-3 SH representations; our experiments use FastGS-based pretrained models (Ren et al., 2026). We combine spatial block quantization with signed-power companding of non-DC SH coefficients, a form of nonuniform scalar quantization (Gray & Neuhoff, 1998). Companding reduces mean search-view loss in all 192 matched comparisons and increases the guarded-feasible count from 75 to 119. Predictions fitted separately for each base guide the search over base size, block size, and SH precision. No candidate is retrained or fine-tuned during compression.

The search selects the smallest measured stream that passes empirical mean and tail PSNR-loss limits. The tail statistic averages the worst 10% of evaluated views. We then freeze this configuration and check its decoded quality once on disjoint views. If confirmation fails, or no feasible candidate is found, we deliver a losslessly encoded reference. We evaluate deployment storage using the actual delivered bytes, including metadata and any fallback. This procedure connects configuration selection to the model ultimately stored and distributed. The quality checks apply to evaluated views; they do not guarantee fidelity at every unseen viewpoint.

Our contributions are threefold:

- We introduce a general post-training compression framework for compatible 3DGS representations. Its compression mechanism couples spatial block quantization and SH companding with budgeted configuration search. SH companding expands the feasible candidate set. Search coordinates base size, block size, and SH precision to minimize measured model size under quality constraints.
- We formulate post-training 3DGS compression as a budgeted *delivery* problem. Evaluation uses the complete bitstream ultimately delivered. Empirical mean and tail PSNR losses constrain decoded quality. Evaluation cost, independent confirmation, and lossless-reference fallback are explicit parts of the formulation.
- We demonstrate strong storage–PSNR trade-offs and fast rendering. On eight existing-domain scenes at default settings, mean model size falls by factors of 38.04 and 4.90 relative to 3DGS and FastGS. Mean RGB8 PSNR rises by 0.257 and 0.173 dB, respectively. Mean rendering throughput is 1,285.5 FPS, or $4.49\times$ and $4.70\times$ the throughput of HAC and ContextGS. Across 12 external-comparison scenes, the post-training compression procedure averages 58.25 s, including reference rendering, confirmation, and any fallback. Base construction is timed separately. These comparisons use predetermined operating points; they do not imply gains in every perceptual metric or at matched quality.

2 Related Work

Compact Gaussian representations. The explicit geometry and appearance attributes of 3D Gaussian Splatting (Kerbl et al., 2023) motivate compression at both the primitive and attribute levels. Compact 3DGS uses masking and vector quantization (Lee et al., 2024); LightGaussian combines pruning, SH distillation, and quantization (Fan et al., 2024). Compressed 3DGS uses sensitivity-aware clustering (Niedermayr et al., 2024), while Papantonakis et al. (2024) adapt pruning and SH bands to resolution. Optimized Minimal 3DGS also reduces primitives and their representation cost (Lee et al., 2025). We instead hold two FastGS-trained bases fixed (Ren et al., 2026) and study which codec configuration to measure and deliver. Our SH transform is a form of classical nonuniform scalar quantization (Gray & Neuhoff, 1998).

Context modeling and efficient coding. Learned probability models exploit attribute dependencies: HAC and HAC++ use spatial context (Chen et al., 2025b;c), ContextGS models anchor-level context (Wang et al., 2024), and CAT-3DGS uses triplanes and autoregressive coding (Zhan et al., 2025). RDO-Gaussian optimizes a rate–distortion objective (Wang et al., 2025); FCGS compresses pretrained scenes without per-scene fine-tuning (Chen et al., 2025a); PCGS supports progressive quality layers (Chen et al., 2026). These codecs offer different operating points. Our question is which measured point to deliver when evaluations are limited.

Budgeted operating-point selection. Constrained black-box optimization uses measured objective and feasibility values to guide costly queries (Gardner et al., 2014). Here, candidates are discrete base-model/codec pairs, and each query requires encoding, decoding, and rendering. Local predictions prioritize queries; actual bytes and guarded distortion determine selection. The query count is distinct from base-training and per-query execution time.

Tail distortion and delivery. Average distortion can hide losses concentrated in a few views. We use empirical CVaR (Rockafellar & Uryasev, 2002) to measure the worst 10% of per-view PSNR losses, then check the selected stream on disjoint views. Failure delivers a losslessly encoded reference, whose bytes enter the result. These statistics describe evaluated views and do not guarantee quality for unseen viewpoints.

3 Method

Our procedure couples measured bitstream size with the quality of the decoded representation. Figure 2 summarizes four stages: fixed-base construction, SH-aware candidate encoding, budgeted selection, and independent confirmation before delivery. The exact decision sequence and stream accounting appear in Appendix A.

3.1 Problem formulation

Let $\{\mathcal{G}_s\}$ be fixed pretrained Gaussian representations and $\mathcal{G}_{\mathrm{ref}}$ the reference. A candidate $c = (s, L, \mathbf{b})$ specifies a base model, spatial block size L, and SH precisions $\mathbf{b} = (b_1, b_2, b_3)$. Its bitstream, reconstruction, and storage cost are

$$z_c = \mathcal{E}_c(\mathcal{G}_s), \qquad \widehat{\mathcal{G}}_c = \mathcal{D}(z_c), \qquad R(c) = |z_c|_{\mathrm{bytes}}. \tag{1}$$

Here $\mathcal{E}_c$ and $\mathcal{D}$ denote encoding and decoding. Let $P_v(\mathcal{G})$ be the PSNR of the image rendered from $\mathcal{G}$ at view v, measured against the ground-truth image. We define

$$d_v(c) = P_v(\mathcal{G}_{\mathrm{ref}}) - P_v(\mathcal{D}(\mathcal{E}_c(\mathcal{G}_s))) \,. \tag{2}$$

Negative values indicate improved PSNR and are retained. Held-out views are partitioned into disjoint search and confirmation sets, V_{S} and V_{C}; neither participates in base training. For search and confirmation, P_v is computed from floating-point renderings and the per-channel PSNR values are averaged. External comparisons instead use independently decoded RGB8 images and PSNR from the full RGB mean squared error; these two metric conventions are kept separate. For a view

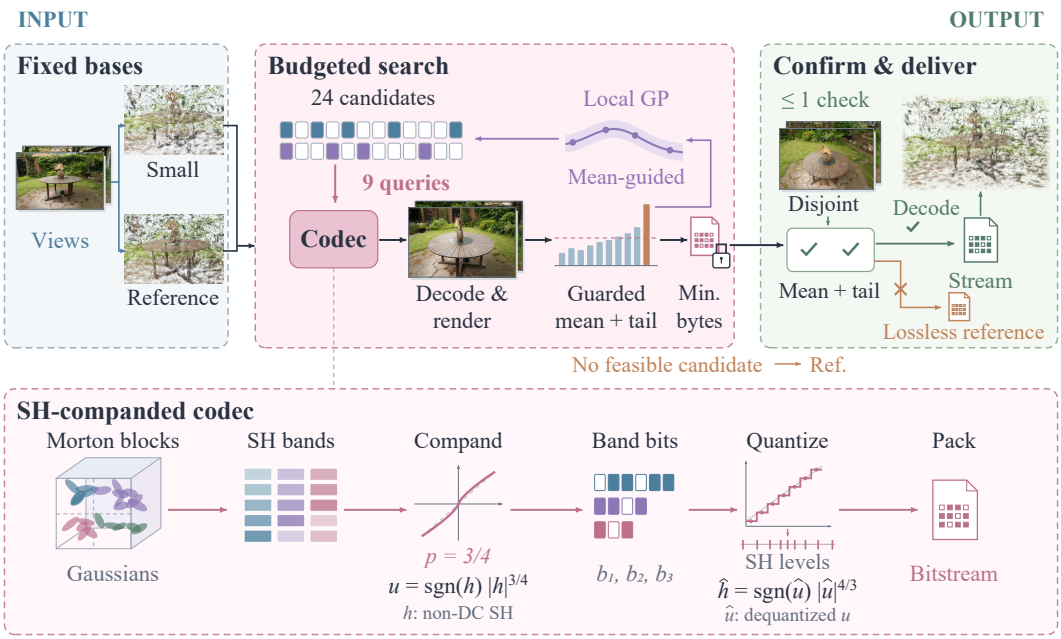


Figure 2: Overview of the budgeted selection and delivery framework. Two fixed bases form a finite SH-aware candidate pool; measured encode–decode–render queries guide the search toward a small feasible stream. The selected candidate is frozen and checked once on disjoint views, after which the candidate or a lossless reference is delivered and the actual bytes are recorded.

set V, we measure mean loss and empirical upper-tail CVaR (Rockafellar & Uryasev, 2002):

$$\mu_V(c) = \frac{1}{|V|}\sum_{v\in V} d_v(c),$$
$$\tau_{\alpha,V}(c) = \min_{\eta\in\mathbb{R}}\left\{\eta + \frac{1}{\alpha|V|}\sum_{v\in V}[d_v(c)-\eta]_+\right\}, \tag{3}$$

where $[x]_+ = \max(x, 0)$ and $\alpha = 0.1$. This definition retains exactly the worst 10% empirical probability mass, including a fractional boundary weight.

Given a budget Q, we reserve one evaluation for confirmation and construct a measured candidate set $\mathcal{O}$ with $|\mathcal{O}| \le Q - 1$. Our search selects

$$c^\star = \arg\min_{c\in\mathcal{O}} R(c),$$
$$\text{subject to} \qquad \widetilde{\mu}_S(c) \le \epsilon_\mu - \gamma, \qquad \widetilde{\tau}_{\alpha,S}(c) \le \epsilon_\tau - \gamma. \tag{4}$$

The tildes denote the empirical search guards defined in Section 3.3. We use $(\epsilon_\mu, \epsilon_\tau) = (1.0, 1.5)$ dB and $\gamma = 0.1$ dB. Equation 4 minimizes actual bytes among measured feasible candidates; the query policy determines which candidates enter $\mathcal{O}$. An empty feasible set triggers reference delivery.

3.2 Spatial and SH-aware candidate construction

We obtain two base representations with different densification schedules using the FastGS-based training pipeline (Ren et al., 2026), then keep them fixed during search. Each is paired with 12 predefined $(L, \mathbf{b})$ configurations, giving 24 candidates with $L \in \{1024, 4096\}$. These paired configurations vary spatial granularity and appearance precision without retraining a candidate.

For each base, we sort Gaussians by Morton order and partition the ordered sequence into blocks of L primitives. Quantization ranges are computed per block and attribute channel. For non-DC SH

coefficients, we use signed-power companding within the standard framework of nonuniform scalar quantization (Gray & Neuhoff, 1998):

$$u = T_p(h) = \mathrm{sgn}(h)|h|^p, \qquad T_p^{-1}(u) = \mathrm{sgn}(u)|u|^{1/p}, \qquad p = \tfrac{3}{4}. \tag{5}$$

For channel j in block k, let a_{kj} and m_{kj} be its minimum and maximum transformed values. Quantization and reconstruction are

$$\begin{aligned} \Delta_{kj} &= \frac{m_{kj} - a_{kj}}{2^{b(j)} - 1}, \qquad q_{ij} = \mathrm{round}\left[\mathrm{clip}\left(\frac{u_{ij} - a_{kj}}{\Delta_{kj}}, 0, 2^{b(j)} - 1\right)\right], \\ \widehat{h}_{ij} &= T_p^{-1}(a_{kj} + \Delta_{kj} q_{ij}). \end{aligned} \tag{6}$$

For constant channels, we set $q_{ij} = 0$ and reconstruct $T_p^{-1}(a_{kj})$. SH degrees 1, 2, and 3 use b_1, b_2, and b_3, respectively, across all color channels. Other attributes use the identity transform and fixed precisions. For $p < 1$, the transform allocates finer reconstruction resolution near zero. The exponent is fixed before search; uniform quantization uses $p = 1$ in the matched control. The resulting distortion reduction can make lower SH precision or larger blocks feasible; Section 4.2 tests this mechanism at measured byte costs.

Each stream contains quantized arrays, block minima and steps, and the metadata required for decoding. We measure $R(c)$ from the complete compressed archive, including this side information. Every queried stream is decoded and rendered before its loss is recorded.

3.3 Budgeted selection

The search first evaluates one initial configuration, $(L, \mathbf{b}) = (4096, (6, 6, 6))$, for each base model. Both evaluations count toward Q. Subsequent queries combine observed search-view losses and file sizes with candidate descriptors.

Guarded feasibility. For each measured loss vector, we draw 512 bootstrap resamples of its views. For $\rho \in \{\mu, \tau_\alpha\}$, the search guard is

$$\widetilde{\rho}_{\mathrm{S}}(c) = \max\left\{\rho_{\mathrm{S}}(c), \mathrm{Quantile}_{0.95}\left(\{\rho(\mathbf{d}^{*(j)}(c))\}_{j=1}^{512}\right)\right\}. \tag{7}$$

This empirical guard makes selection more conservative when the observed losses are unstable under resampling. It is not a coverage bound.

Predictions within each base model. Because base models can have different quality floors, we fit a separate fixed-kernel Gaussian-process surrogate to each model's observed per-view loss vectors. The inputs describe SH precision and block size. Let $\widehat{\mathbf{d}}_t(c)$ and $\boldsymbol{\sigma}_t(c)$ denote the predicted vector and its exploration scales. The mean and tail scales are $s_{\mu,t}(c) = \mathrm{mean}_v\, \sigma_{t,v}(c)$ and $s_{\tau,t}(c) = \max_v \sigma_{t,v}(c)$. For either risk statistic, we form

$$\widehat{\rho}_t^{\,g}(c) = \rho(\widehat{\mathbf{d}}_t(c)) + \frac{1}{|\mathcal{O}_{t,s}|} \sum_{c_i \in \mathcal{O}_{t,s}} \left[\widetilde{\rho}_{\mathrm{S}}(c_i) - \rho_{\mathrm{S}}(c_i)\right], \tag{8}$$

where $\mathcal{O}_{t,s}$ contains the measured candidates from base s. Thus, acquisition retains the observed resampling penalty. We also calibrate a nominal pre-encoding size $R_{\mathrm{nom}}(c)$ separately within each base:

$$\widehat{R}_t(c) = R_{\mathrm{nom}}(c)\ \mathrm{median}_{c_i \in \mathcal{O}_{t,s}} \frac{R(c_i)}{R_{\mathrm{nom}}(c_i)}. \tag{9}$$

The nominal size reflects point count, attribute precision, and block overhead; only measured bytes enter the final selection.

Query acquisition. Following the use of feasibility scores in constrained optimization (Gardner et al., 2014), our primary policy uses predicted mean loss to prioritize candidates. With Φ the standard normal CDF, it computes

$$\begin{aligned} p_t(c) &= \Phi\left(\frac{\epsilon_\mu - \gamma - \widehat{\mu}_t^{\,g}(c)}{\max\{s_{\mu,t}(c), 0.05\}}\right), \\ a_t(c) &= p_t(c)\left[\left(\log \frac{R_t^\star}{\max\{\widehat{R}_t(c), 1\}}\right)_+ + 0.05\, \frac{s_{\mu,t}(c) + s_{\tau,t}(c)}{2}\right], \end{aligned} \tag{10}$$

where $R_t^\star$ is the smallest measured size satisfying Equation 4. Before a feasible candidate is found, we use $a_t(c) = p_t(c)$. When possible, queries are restricted to base models with an observed feasible candidate; otherwise all unqueried candidates are eligible. We query the eligible candidate with maximum a_t, update the observations, and stop after $Q - 1$ search evaluations. These scores guide exploration rather than provide calibrated confirmation probabilities. Although acquisition uses mean feasibility, Equation 4 always checks both mean and tail loss.

3.4 Independent confirmation and delivery

We freeze $c^\star$ before accessing V_{C}. If it exists, we decode its stream and evaluate it once on these views. Confirmation passes when $\mu_{\mathrm{C}}(c^\star) \le \epsilon_\mu$ and $\tau_{\alpha,\mathrm{C}}(c^\star) \le \epsilon_\tau$, using the empirical statistics in Equation 3 without the search guard. Let $A(c^\star)$ indicate this decision and let $R_0 = |\mathcal{E}_{\mathrm{lossless}}(\mathcal{G}_{\mathrm{ref}})|_{\mathrm{bytes}}$. The actual delivered size is

$$R_{\mathrm{del}} = \begin{cases} R(c^\star), & c^\star \text{ exists and } A(c^\star) = 1, \\ R_0, & \text{otherwise.} \end{cases} \tag{11}$$

We deliver the corresponding candidate stream or a losslessly encoded reference, with no reselection after confirmation. A query evaluates one candidate over a view set, so $Q_{\mathrm{used}} = |\mathcal{O}| + \mathbf{1}[c^\star \text{ exists}] \le Q$. Our setting uses $Q = 10$; base training, reference rendering, and fallback encoding are accounted for separately. Final test views are used only after delivery is fixed, and reported storage includes the actual fallback bytes.

4 Experiments

We evaluate three links in the compression pipeline: the configurations made feasible by companding, the candidates found under a query budget, and the artifacts actually delivered after confirmation.

4.1 Experimental Setup

Data and evaluation. The evaluation comprises 31 original scenes, an eight-scene study with three training seeds, and a three-seed extension to all eight Synthetic-NSVF and four Blended-MVS scenes in our collection. External default-point comparisons use eight existing-domain and four additional-domain scenes with one matched training seed per scene. All methods share training and test images; search and confirmation views are disjoint and excluded from training and fine-tuning. External comparisons and the additional-domain extension use independently decoded RGB8 images, full-RGB PSNR, SSIM, and VGG LPIPS. Original 31-scene and repeatability results retain their historical floating-point metrics. Internal risk uses channel-averaged floating-point PSNR (Section 3.1). We average seeds within scenes; pooled byte savings are identified separately. Appendix B.1 specifies coverage and aggregation. All delivered sizes include metadata and actual reference fallback.

Baselines and implementation. Default-point baselines are 3DGS (Kerbl et al., 2023), FastGS (Ren et al., 2026), LightGaussian (Fan et al., 2024), Compact-3DGS (Lee et al., 2024), C3DGS (Niedermayr et al., 2024), HAC (Chen et al., 2025b), and ContextGS (Wang et al., 2024). An additional six-scene comparison evaluates 36 FCGS (Chen et al., 2025a) points on the same cached small/reference inputs. Its numerical compatibility amendment and shared-decoder accounting are disclosed in Appendix C.12. Ours uses companding and local-mean acquisition; Uniform changes only p to 1. Both bases are independently trained for 30,000 iterations with densification intervals 100 and 500. Defaults are $p = 0.75$, 24 candidates, $Q = 10$ (nine search queries and at most one confirmation), $(\epsilon_\mu, \epsilon_\tau) = (1.0, 1.5)$ dB, and $\gamma = 0.1$ dB. Search performs no gradient-based fine-tuning. Experiments use NVIDIA RTX 5880 Ada GPUs.

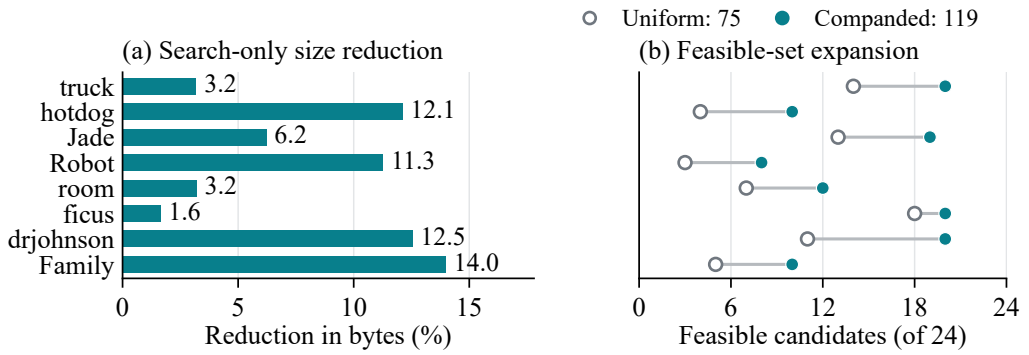


Figure 3: Companding improves the available compression configurations. Across all eight full-pool scenes, the smallest search-feasible stream shrinks by 1.64–13.98% and feasible configurations increase from 75 to 119. These search-only diagnostics precede budgeted selection, confirmation, and fallback.

4.2 Companding Expands the Feasible Configuration Space

Controlled distortion improvement. We isolate the transform in 192 matched configurations across eight scenes, holding the base, block size, and SH precisions fixed and changing only $p = 1$ to $p = 0.75$. Companding reduces SH coefficient NMSE and mean search-view PSNR loss in all 192 pairs; the median relative reduction in SH coefficient NMSE is 44.88%. Tail loss decreases in 189 pairs. Guarded-feasible configurations increase from 75 to 119 (58.67%), with gains in every scene and no loss of a previously feasible configuration.

From lower distortion to smaller feasible streams. Reanalysis of the complete measured pools shows that their minimum search-feasible size decreases in *all eight scenes*, by 1.64–13.98% (Figure 3). The reduction is 12.11% on hotdog, 12.54% on drjohnson, and 13.98% on Family. This connects the distortion benefit to a concrete compression opportunity: lower SH precision or larger blocks become feasible. At identical configurations, the median paired size increase is 4.96%. Thus the gain comes from better feasible settings, not lower bytes at unchanged settings. Appendix C.5 reports every scene. These full-pool search diagnostics do not imply that a $Q = 10$ search finds the minimum or that confirmation accepts it.

4.3 Delivered Compression and Operating-Point Competitiveness

Matched delivery gains across additional domains. On all eight Synthetic-NSVF scenes and three training seeds, Ours reduces mean delivered size from 2.553 to 2.489 MiB (2.50%) relative to Uniform, while improving all three mean image metrics: PSNR 36.607 to 36.669 dB, SSIM 0.9849 to 0.9853, and LPIPS 0.0252 to 0.0244. Across all twelve additional-domain scenes, Ours delivers fewer bytes in 24/36 pairs and all 36 deliveries meet the empirical final-test risk limits. Pooled storage decreases by 27.24%; one avoided Uniform fallback on Statues drives most of that gain. Excluding the entire Statues scene still leaves 3.71% fewer delivered bytes across the remaining eleven scenes. Appendix C.11 separates ordinary candidate savings from fallback effects.

Default-point system comparison. Figure 4 shows the highest mean RGB8 PSNR and second-highest mean FPS among the nine default-point methods in both subsets. On the existing-domain subset, Ours reduces FastGS's mean size from 50.562 to 10.315 MiB (79.60%), gains 0.173 dB, and retains 85.7% of its throughput (1285.5 versus 1500.0 FPS). The additional-domain subset saves 71.18% and gains 0.718 dB. Relative to C3DGS on the existing subset, Ours saves 33.78% and gains 0.455 dB. HAC and ContextGS have smaller default files. These are system-level operating points, not universal matched-quality gains; SSIM/LPIPS do not improve uniformly (Table 3).

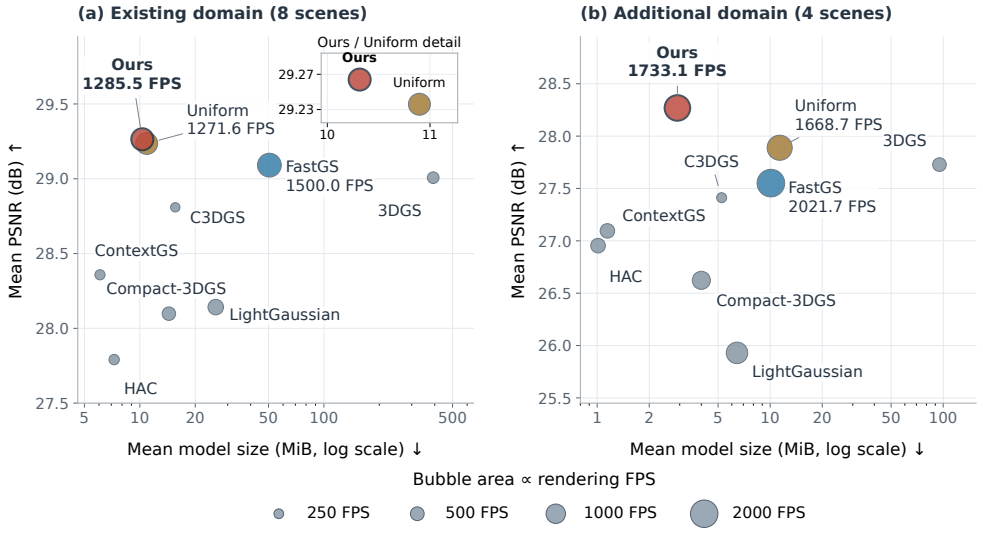


Figure 4: Delivered storage, RGB8 PSNR, and rendering throughput at predetermined default points. Panels show means for the eight existing-domain and four additional-domain scenes, including fallback. Bubble area uses a shared FPS scale; size axes are logarithmic except in the inset. FPS excludes loading and decoding. Axis ranges differ between subsets. These are default-point, not matched-quality comparisons.

Table 1: Matched codec–policy ablation on eight scenes and three training seeds (24 runs per row). $\text{PSNR}_{\text{float}}$ uses the historical evaluator; F and V count fallbacks and final-test mean/tail violations.

Codec	Policy	MiB ↓	$\text{PSNR}_{\text{float}}$ ↑	F/24	V/24
Uniform	Greedy	27.716	29.329	7	1
Companded	Greedy	17.904	29.050	8	3
Uniform	Local mean	15.680	29.090	3	2
Companded	Local mean	16.856	29.070	6	3

Measured comparison with feedforward compression. Across six scenes, we compare three quality-limit settings with FCGS on both cached inputs at three checkpoints. On truck, our default uses 10.822 MiB at 25.441 dB, versus 11.651 MiB at 25.050 dB for FCGS's small-input $\lambda = 4 \times 10^{-4}$ point. On Robot, our relaxed point uses 1.817 MiB at 38.192 dB versus 1.836 MiB at 38.102 dB. These are competitive measured byte–PSNR regions; FCGS wins elsewhere, including garden and Jade. Appendix C.12 retains all 36 points, perceptual metrics, and costs; Appendices C.8–C.9 retain other measured curves and quality-matched diagnostics.

4.4 Ablation and Budgeted Selection

Candidate gains and delivery costs. With companding fixed, local-mean search reduces delivered bytes by 5.85% relative to greedy (Table 1). Against Uniform local-mean, Ours wins 17/24 pairs and reduces total selected-candidate bytes by 5.47%, but additional fallback raises delivered bytes by 7.50%. The decomposition is explicit: 13.478 MiB of candidate savings is outweighed by 41.711 MiB of extra fallback increments (Appendix C.11). Across the original 31-scene single-run study, 24 scenes have smaller deliveries and pooled bytes decrease 3.08%; repeated-run results and their variability remain in Appendices C.2–C.4.

Limited query budgets. We compare local-mean, greedy, and five-seed random search at $Q \in \{5, 10, 20\}$ with the same two initial queries. At $Q = 10$, local-mean reduces delivered size versus greedy by 24.49% on bonsai and 18.25% on hotdog, with PSNR decreases of 0.422 and 0.278 dB. On bonsai it reaches the artifact greedy reaches at $Q = 20$. Greedy is smaller on bicycle; ficus exposes

Table 2: Tail and confirmation ablations: six scenes, three training seeds. Losses are floating-point PSNR differences from the reference; CVaR_{10} is computed per run before averaging.

Variant	MiB	Mean loss	CVaR_{10}	F/18	V/18
Full	9.232	0.244	0.944	3	3
Mean-only	9.173	0.257	0.977	3	3
No confirmation	6.927	0.332	1.134	0	3

confirmation-induced fallback. A separate four-scene search-only oracle gives a mean $Q = 10$ gap of 6.98%; it excludes confirmation and is not a deliverable optimum. Appendix C.6 reports all budgets and gaps.

4.5 Quality Control and Delivery Reliability

Mean-only removes tail feasibility and acceptance limits. No confirmation delivers the same candidate selected by the full method's nine search queries (Table 2). Confirmation lowers mean loss from 0.332 to 0.244 dB and mean scene-level CVaR_{10} from 1.134 to 0.944 dB, at a 33.27% increase in delivered size. All variants still exceed test limits in 3/18 runs: the full method falls back on ficus, whereas drjohnson exceeds the limits after passing confirmation. The tail constraint adds only a small aggregate gain over mean-only. Thus confirmation is an empirical check with a storage cost, not an unseen-view guarantee. Appendix C.7 retains all rejected candidates and their test losses.

4.6 Computational and Deployment Costs

Across the twelve external-comparison scenes, constructing both bases takes a recorded mean of 345.63 s; search, confirmation, and any fallback take 58.25 s. Decoding adds 2.42 s. Bases are reusable across quality limits. On the existing-domain subset, rendering averages 1285.5 FPS versus 286.3 for HAC and 273.6 for ContextGS; those methods have smaller default files. Throughput reflects complete representations and native renderers, not companding alone. We use ten warm-up frames and three CUDA-synchronized test passes, excluding decoding. Appendix C.10 includes teacher and base costs. Historical/new timings mix shared-machine loads and do not establish controlled training speedups.

5 Conclusion

This paper presents a general post-training compression framework for 3D Gaussian Splatting. We coordinate SH companding with budgeted configuration search to reduce model size under quality constraints. Companding lowers quantization error and expands the feasible configuration space. The search jointly selects the pretrained base model, block size, and SH precision without retraining or fine-tuning. Experiments demonstrate the effectiveness of this coordinated design. At predetermined default settings on eight benchmark scenes, our complete system achieves substantially smaller models and higher mean PSNR than 3DGS and FastGS. It also maintains fast rendering. These results support the practical use of our framework for compact storage and efficient distribution of trained 3DGS scenes.

AI use statement

Generative AI tools assisted with language editing, appendix drafting, and arranging existing experimental images for comparison. All reported measurements and renderings come from the authors' recorded experiments. The authors checked the final text, figures, and numerical claims and take responsibility for the manuscript.

References

Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Confer-*

ence on Computer Vision and Pattern Recognition, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Barron_Mip-NeRF_360_Unbounded_Anti-Aliased_Neural_Radiance_Fields_CVPR_2022_paper.html.

Yihang Chen, Qianyi Wu, Mengyao Li, Weiyao Lin, Mehrtash Harandi, and Jianfei Cai. Fast feed-forward 3D gaussian splatting compression. In *International Conference on Learning Representations*, 2025a. URL https://openreview.net/forum?id=DCandSZ2F1.

Yihang Chen, Qianyi Wu, Weiyao Lin, Mehrtash Harandi, and Jianfei Cai. HAC: Hash-grid assisted context for 3D gaussian splatting compression. In *Computer Vision – ECCV 2024*, pp. 422–438. Springer Nature Switzerland, 2025b. doi: 10.1007/978-3-031-72667-5_24. URL https://www.ecva.net/papers/eccv_2024/papers_ECCV/html/1178_ECCV_2024_paper.php.

Yihang Chen, Qianyi Wu, Weiyao Lin, Mehrtash Harandi, and Jianfei Cai. HAC++: Towards 100x compression of 3D gaussian splatting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 10210–10226, 2025c. doi: 10.1109/TPAMI.2025.3594066.

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*, pp. 3111–3119, 2026. doi: 10.1609/aaai.v40i4.37304.

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, pp. 140138–140158, 2024. doi: 10.52202/079017-4447. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/fd881d3b625437354d4421818f81058f-Abstract-Conference.html.

Jacob R. Gardner, Matt J. Kusner, Zhixiang Xu, Kilian Q. Weinberger, and John P. Cunningham. Bayesian optimization with inequality constraints. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pp. 937–945. PMLR, 2014. URL https://proceedings.mlr.press/v32/gardner14.html.

R. M. Gray and D. L. Neuhoff. Quantization. *IEEE Transactions on Information Theory*, 44 (6):2325–2383, 1998. doi: 10.1109/18.720541. URL https://doi.org/10.1109/18.720541.

Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow. Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics*, 37(6), 2018.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023. URL https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/.

Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4), 2017. URL https://www.tanksandtemples.org/.

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*, pp. 21719–21728, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Lee_Compact_3D_Gaussian_Representation_for_Radiance_Field_CVPR_2024_paper.html.

Joo Chan Lee, Jong Hwan Ko, and Eunbyung Park. Optimized minimal 3D gaussian splatting. In *Advances in Neural Information Processing Systems*, volume 38, pp. 150615–150639, 2025. doi: 10.52202/085713-4534.

Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. Neural sparse voxel fields. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://arxiv.org/abs/2007.11571.

Ben Mildenhall, Pratul P. Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics*, 38(4), 2019. URL https://arxiv.org/abs/1905.00889.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *Computer Vision – ECCV 2020*, 2020. doi: 10.1007/978-3-030-58452-8_24.

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*, pp. 10349–10358, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Niedermayr_Compressed_3D_Gaussian_Splatting_for_Accelerated_Novel_View_Synthesis_CVPR_2024_paper.html.

Panagiotis Papantonakis, Georgios Kopanas, Bernhard Kerbl, Alexandre Lanvin, and George Drettakis. Reducing the memory footprint of 3D gaussian splatting. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 7(1), 2024. doi: 10.1145/3651282. URL https://repo-sam.inria.fr/fungraph/reduced_3dgs/.

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*, pp. 26094–26103, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Ren_FastGS_Training_3D_Gaussian_Splatting_in_100_Seconds_CVPR_2026_paper.html.

R. Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002. doi: 10.1016/S0378-4266(02)00271-6. URL https://doi.org/10.1016/S0378-4266(02)00271-6.

Henan Wang, Hanxin Zhu, Tianyu He, Runsen Feng, Jiajun Deng, Jiang Bian, and Zhibo Chen. End-to-end rate-distortion optimized 3D gaussian representation. In *Computer Vision – ECCV 2024*, pp. 76–92. Springer Nature Switzerland, 2025. doi: 10.1007/978-3-031-73636-0_5. URL https://www.ecva.net/papers/eccv_2024/papers_ECCV/html/7526_ECCV_2024_paper.php.

Yufei Wang, Zhihao Li, Lanqing Guo, Wenhan Yang, Alex C. Kot, and Bihan Wen. ContextGS: Compact 3D gaussian splatting with anchor level context model. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 51532–51551. Curran Associates, Inc., 2024. doi: 10.52202/079017-1631. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/5c20ca4b0b20b0bd2f1d839dc605e70f-Abstract-Conference.html.

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. doi: 10.1109/TIP.2003.819861.

Yao Yao, Zixin Luo, Shiwei Li, Jingyang Zhang, Yufan Ren, Lei Zhou, Tian Fang, and Long Quan. BlendedMVS: A large-scale dataset for generalized multi-view stereo networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Yao_BlendedMVS_A_Large-Scale_Dataset_for_Generalized_Multi-View_Stereo_Networks_CVPR_2020_paper.html.

Yu-Ting Zhan, Cheng-Yuan Ho, Hebi Yang, Yi-Hsin Chen, Jui Chiu Chiang, Yu-Lun Liu, and Wen-Hsiao Peng. CAT-3DGS: A context-adaptive triplane approach to rate-distortion-optimized 3DGS compression. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2503.00357.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_The_Unreasonable_Effectiveness_CVPR_2018_paper.html.

A Method and Algorithm Details

A.1 Decision sequence and delivered artifact

The two base models are trained before compression search and remain fixed. For a scene, the candidate pool is the 12 predefined block/precision pairs on each base, not a Cartesian product of every listed block size and bit width. The following sequence makes the query and delivery accounting explicit:

1. Encode, decode, and render the initial configuration on each base. These two measurements count toward the nine search queries.
2. Fit one surrogate per base from its measured per-view losses and update the base-specific ratio of measured to nominal stream size. Select and measure the next eligible configuration using Equation 10; repeat until nine candidates have been measured. Acquisition ranks unmeasured configurations, while both guarded mean and tail limits govern feasibility of measured ones.
3. Among measured feasible streams, freeze the one with the fewest actual bytes. If none exists, deliver the losslessly encoded reference.
4. Otherwise check the frozen stream once on disjoint confirmation views. Deliver it if both unguarded limits pass; on failure deliver the losslessly encoded reference. Do not search again after confirmation.

The reported rate is the complete archive actually delivered. In particular, a rejected candidate's encoded bytes are not substituted for the reference bytes in a result table. The fallback is a lossless encoding of the reference Gaussian representation; it is distinct from both the uncompressed reference PLY and an accepted lossy candidate.

A.2 Blockwise codec and candidate pool

The candidate pool combines two block sizes with predetermined precision allocations across SH degrees, yielding 12 configurations per base. Spatial ordering precedes block partitioning. Each delivered stream includes the quantized attributes and all metadata required for decoding, so rate measurements reflect the complete representation. The inverse signed-power transform is applied after dequantization of higher-order SH coefficients. All other attribute transforms and their precisions are identical in the matched uniform and companded configurations.

A.3 Search prediction and empirical risk

The per-base surrogate uses descriptors for the block size and SH precisions. Each predicted view loss contributes to the mean and upper-tail summaries in Equation 8. The observed bootstrap penalty is added within the same base, and a median measured-to-nominal byte ratio calibrates predicted size. Neither prediction replaces a decoded measurement when choosing the final artifact. Acquisition uses predicted mean feasibility to spend queries, then final search selection uses the guarded mean and CVaR_{10} limits together.

For n observed losses sorted from largest to smallest, empirical CVaR_{10} averages the largest $0.1n$ units of probability mass. When $0.1n$ is nonintegral, the boundary loss receives fractional weight. With fewer than ten views the statistic is the worst observed loss. The bootstrap search guard reflects sensitivity to resampling the available views; it is not a formal guarantee on unseen views. Confirmation uses a disjoint set and the original unguarded limits.

B Experimental Protocol and Reproducibility

B.1 Protocols, scene coverage, and aggregation

Scene coverage. The 31-scene study covers five benchmark families under six data/preprocessing groups: Mip-NeRF 360 (Barron et al., 2022) (bicycle, bonsai, counter, garden, kitchen, room, stump), Deep Blending (Hedman et al., 2018) (drjohnson, playroom), Tanks and Temples (Knapitsch et al., 2017) (train, truck, Barn, Caterpillar, Family, Ignatius), LLFF (Mildenhall et al., 2019) (fern, flower, fortress, horns, leaves, orchids, room, trex), and NeRF Synthetic (Mildenhall et al., 2020) (chair, drums, ficus, hotdog, lego, materials, mic, ship). The two Tanks and Temples preprocessing groups do not duplicate truck. The repeatability study uses bonsai, bicycle, drjohnson, truck, Family, LLFF room, hotdog, and ficus, each with three independent training runs. These scenes were selected from previously observed results, so the repeated runs assess repeatability rather than unseen-scene generalization. The additional-domain extension uses three training runs on Synthetic-NSVF (Liu et al., 2020) (Bike, Lifestyle, Palace, Robot, Spaceship, Steamtrain, Toad, Wineholder) and BlendedMVS (Yao et al., 2020) (Character, Fountain, Jade, Statues).

The external existing-domain subset comprises bonsai, garden, room, and counter from Mip-NeRF 360, together with truck, hotdog, Family, and Barn. Its additional-domain subset comprises Robot, Wineholder, Jade, and Statues. These fixed subsets should not be interpreted as complete benchmark averages. The six rate–distortion scenes are bonsai, garden, truck, Mip-NeRF 360 room, Robot, and Jade. The confirmation ablation uses hotdog, truck, Mip-NeRF 360 room, Family, ficus, and drjohnson with three training runs per scene.

Image splits and preprocessing. Search and confirmation views are disjoint and excluded from training and any baseline postprocessing in the unified external comparison. Test views are held out from both selection stages. We retain the benchmark-specific image preprocessing and native test splits; validation views are partitioned between search and confirmation where applicable. All methods receive the same images, cameras, and initial points for a given comparison scene.

Metrics and storage. The external-comparison and additional-domain results use RGB8 images, full-RGB mean-squared-error PSNR, the fixed 3DGS SSIM implementation (Wang et al., 2004), and VGG LPIPS v0.1 (Zhang et al., 2018) with inputs mapped to $[-1, 1]$. The original 31-scene and repeatability studies use their frozen floating-point evaluator. Their PSNR averages channel-wise logarithmic scores and must not be pooled with RGB8 PSNR. Internal risk always uses the floating-point reference-minus-candidate PSNR and fractional empirical CVaR_{10}. Reported test violations occur when mean loss exceeds 1.0 dB or CVaR_{10} exceeds 1.5 dB, unless the operating-point limits are explicitly varied. These audits do not trigger reselection. One MiB is 2^{20} bytes. Delivered size includes required metadata and the actual lossless-reference stream on fallback; it is not the decoded GPU memory footprint.

Aggregation. All scheduled valid outcomes, including reference fallbacks, enter the reported aggregates. With $R^C_{s,k}$ and $R^U_{s,k}$ denoting companded and uniform delivered bytes for scene s and training seed k, the paired percentage saving is $100(1 - R^C_{s,k}/R^U_{s,k})$. We average this percentage over seeds within each scene and then over scenes. This differs from the pooled saving $100(1 - \sum_{s,k} R^C_{s,k} / \sum_{s,k} R^U_{s,k})$. The latter weights comparisons by the control's file size and can be strongly affected by reference fallback. Standard deviations are sample standard deviations over the stated seeds; candidates and views are not treated as independent scenes.

Baselines and candidate configuration. External baselines use fixed default operating points and their standard training or postprocessing procedures. For additional domains, these are predetermined transfer settings, not claims of author-tuned optima. The supplementary FCGS operating-point protocol is reported separately in Appendix C.12. Both internal bases use the FastGS-based training pipeline with periodic pruning and no depth supervision. The smaller base is independently trained, not obtained by pruning the reference during search. Each base is paired with the same 12 codec configurations. Both mean and tail checks remain active under local-mean acquisition.

Table 3: Comparison at predetermined default operating points. Entries are scene means within each fixed subset; all scheduled scenes and fallback artifacts are included. PSNR, SSIM, and LPIPS use the unified RGB8 evaluation. FPS excludes loading and decoding. The best, second-best, and third-best entries in each metric column within each subset are highlighted in red, orange, and yellow, respectively.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	MiB ↓	FPS ↑
Existing-domain subset (8 scenes)					
3DGS (Kerbl et al., 2023)	29.007	0.9200	0.1716	392.407	359.4
FastGS (Ren et al., 2026)	29.091	0.9129	0.1987	50.562	1500.0
LightGaussian (Fan et al., 2024)	28.142	0.9071	0.1947	25.842	639.7
Compact-3DGS (Lee et al., 2024)	28.098	0.9091	0.1906	14.379	482.5
C3DGS (Niedermayr et al., 2024)	28.809	0.9139	0.1827	15.576	232.4
HAC (Chen et al., 2025b)	27.791	0.8998	0.2097	7.252	286.3
ContextGS (Wang et al., 2024)	28.358	0.9057	0.2031	6.072	273.6
Uniform	29.236	0.9111	0.2028	10.897	1271.6
Ours	29.264	0.9109	0.2015	10.315	1285.5
Additional-domain subset (4 scenes)					
3DGS (Kerbl et al., 2023)	27.728	0.9141	0.1054	95.135	494.7
FastGS (Ren et al., 2026)	27.551	0.9376	0.0867	10.093	2021.7
LightGaussian (Fan et al., 2024)	25.930	0.9064	0.1193	6.431	1223.6
Compact-3DGS (Lee et al., 2024)	26.623	0.9013	0.1220	4.000	870.9
C3DGS (Niedermayr et al., 2024)	27.412	0.9124	0.1069	5.242	264.0
HAC (Chen et al., 2025b)	26.954	0.9163	0.1129	1.012	564.6
ContextGS (Wang et al., 2024)	27.095	0.9114	0.1171	1.146	557.3
Uniform	27.889	0.9142	0.1070	11.348	1668.7
Ours	28.269	0.9126	0.1370	2.909	1733.1

C Additional Quantitative Results

C.1 External comparison at fixed operating points

Table 3 reports all default-point metrics behind Figure 4. The eight existing-domain and four additional-domain scenes are aggregated separately. Ours has the highest mean RGB8 PSNR and second-highest rendering FPS in both subsets; HAC and ContextGS use fewer bytes, while FastGS has higher SSIM and lower LPIPS. These are complete delivered artifacts, including any fallback. The 79.60% existing-domain storage reduction relative to FastGS is a system-level comparison, not a matched-codec companding effect.

C.2 Complete 31-scene matched comparison

Table 4 reports the original single-run matched comparison. The total delivered size is 319.390 MiB for Uniform and 309.561 MiB for Ours. The median scene-wise saving is 3.79%, whereas the mean scene-wise percentage is -19.34% because reference fallback can produce large relative increases. Both methods fall back twice, but on different scenes. Ours exceeds the test CVaR limit on five of 31 scenes (bonsai, playroom, materials, Caterpillar, and Ignatius), despite all test mean losses remaining below 1.0 dB.

Table 4: All 31 scenes from one training run. U: Uniform; C: Ours. PSNR uses the historical floating-point evaluator. Superscript F marks a lossless-reference fallback.

Scene	PSNR_U	PSNR_C	MiB_U	MiB_C
Mip-NeRF 360				
bicycle	23.851	24.080	32.347	33.687
bonsai	32.227	32.314	11.353	10.923
counter	29.305	29.155	10.911	10.457
garden	26.921	26.801	25.631	23.896
kitchen	31.723	31.926	18.280	18.962
room	31.230	31.277	18.055	17.069
stump	24.545	24.888	21.346	22.364
Deep Blending				
drjohnson	28.384	28.354	15.468	14.533
playroom	29.123	29.102	12.302	11.739
Tanks and Temples: train/truck				
train	22.181	22.182	15.297	15.703
truck	25.477	25.448	11.392	10.822
Tanks and Temples: additional scenes				
Barn	24.354	24.487	5.586	5.408
Caterpillar	21.677	21.349	4.411	4.354
Family	27.446	27.592	2.442	2.270
Ignatius	27.569	27.259	2.272	2.227
LLFF				
fern	16.942	16.981	5.167	5.076
flower	20.706	20.721	5.565	5.369
fortress	22.854	22.870	3.675	3.578
horns	20.471	20.508	5.820	5.599
leaves	15.036	15.040	7.586	7.177
orchids	14.626	14.513	8.194	7.154
room	24.069	23.897	1.968	1.662
trex	19.601	19.734	4.385	4.600
NeRF Synthetic				
chair	35.538	35.328	20.618^{F}	2.472
drums	25.913	25.947	3.262	3.220
ficus	34.087	34.310	1.751	15.512^{F}
hotdog	37.150	37.260	1.803	1.678
lego	34.840	34.736	3.341	3.101
materials	29.415	29.238	2.599	2.425
mic	35.543	35.593	2.113	2.071
ship	30.196	30.196	34.451^{F}	34.451^{F}

C.3 Three-run repeatability and search-guard ablation

The paired statistics in Table 5 compare Ours with Uniform over three training runs. Ficus and LLFF room account for the strongest adverse delivery effects; their outcomes are retained. Family's large average saving includes a Uniform fallback in one run. Hotdog and truck show much smaller variation across runs.

Removing only the bootstrap search guard, while retaining the margin and mean/tail confirmation limits, changes one of eight deliveries. Full and unguarded mean sizes are 12.967 and 13.091 MiB, respectively; both have two fallbacks and one test violation. The changed scene is drjohnson: removing the guard increases size by 7.62% and PSNR by 0.2735 dB, while lowering test CVaR_{10} from 2.3299 to 1.9899 dB. Both outcomes exceed the tail limit. This experiment does not establish a general reliability improvement from bootstrap guarding.

Table 5: Paired size savings and PSNR differences (Ours minus Uniform). Values are means ± sample standard deviations over three training runs. F and V show Ours/Uniform fallback and test-violation counts, each out of three.

Scene	Saving (%)	$\Delta\text{PSNR}_{\text{float}}$	F (C/U)	V (C/U)
bonsai	2.76 ± 2.47	-0.085 ± 0.191	1/1	0/0
drjohnson	10.33 ± 3.84	-0.253 ± 0.147	0/0	3/2
truck	4.84 ± 0.13	-0.031 ± 0.016	0/0	0/0
Family	34.98 ± 48.42	0.046 ± 0.103	0/1	0/0
LLFF room	-297.48 ± 518.64	-0.174 ± 0.426	2/1	0/0
hotdog	6.89 ± 0.02	0.100 ± 0.004	0/0	0/0
bicycle	3.49 ± 6.67	0.022 ± 0.191	0/0	0/0
ficus	-776.09 ± 1.27	0.215 ± 0.008	3/0	0/0

C.4 Additional-domain results

Table 6 gives complete dataset means for the four matched codec–policy combinations. Table 7 provides all per-scene paired savings. The pooled byte savings for Ours versus Uniform are 2.50% on Synthetic-NSVF and 47.89% on BlendedMVS, compared with scene-averaged paired percentage savings of 2.89% and 12.19%. One Statues run contributes strongly to the latter dataset's pooled reduction: Uniform delivers a 35.844 MiB reference fallback, while Ours delivers 2.492 MiB. Its RGB8 LPIPS changes from 0.2773 to 0.4004, illustrating that the PSNR-based acceptance rule does not ensure an improvement in every perceptual metric.

Table 6: Dataset means under unified RGB8 evaluation. F and V count fallbacks and floating-point test-risk violations out of 24 runs for Synthetic-NSVF and 12 for BlendedMVS. All means include every scene and training run.

Method	PSNR	SSIM	LPIPS	MiB	F	V
Synthetic-NSVF: 8 scenes, 3 runs						
Uniform + greedy	36.828	0.9858	0.0230	2.904	0	0
Companded + greedy	36.624	0.9847	0.0254	2.592	0	0
Uniform	36.607	0.9849	0.0252	2.553	0	0
Ours	36.669	0.9853	0.0244	2.489	0	0
BlendedMVS: 4 scenes, 3 runs						
Uniform + greedy	23.220	0.8823	0.1837	6.373	1	2
Companded + greedy	23.309	0.8802	0.1986	3.054	0	0
Uniform	23.256	0.8825	0.1857	6.114	1	2
Ours	23.407	0.8827	0.1924	3.186	0	0

C.5 Matched quantization analysis

The mechanism analysis uses 24 configurations per scene: two fixed bases and 12 block/SH-precision configurations. Each pair changes only the companding exponent and is evaluated on search views. Across all 192 pairs, mean search loss decreases from 2.016 to 0.981 dB and mean CVaR_{10} from 3.205 to 1.763 dB. These are configuration-weighted diagnostics, not final-test dataset scores. Configuration-level measurements do not constitute 192 independent scenes. We additionally reanalyse these archived full-pool measurements at their actual archive byte costs. This is a codec-level search-feasible envelope, not a new independent scene study or a delivery oracle.

At fixed configurations all 192 companded archives are larger, with a 4.96% median paired increase. Nevertheless, the distortion improvement permits different feasible configurations. On hotdog the minimizing setting changes from SH bits $(6, 5, 4)$ with block size 1024 to $(4, 4, 4)$ with block size 4096. On truck, room, and ficus the minimizing SH bits stay unchanged while block size increases from 1024 to 4096. The improved envelope therefore reflects both appearance precision and spatial overhead, rather than a same-configuration entropy-coding gain.

Table 7: Ours versus Uniform across all additional-domain scenes. Saving and RGB8 PSNR differences are paired means ± sample standard deviations across three training runs.

Scene	Saving (%)	ΔPSNR (dB)
Synthetic-NSVF		
Bike	-4.65 ± 0.01	0.368 ± 0.003
Lifestyle	0.07 ± 4.00	0.212 ± 0.100
Palace	2.46 ± 5.10	0.030 ± 0.164
Robot	11.24 ± 0.02	-0.233 ± 0.004
Spaceship	-3.23 ± 0.12	0.215 ± 0.013
Steamtrain	18.42 ± 3.36	-0.319 ± 0.106
Toad	-3.63 ± 0.03	0.169 ± 0.007
Wineholder	2.43 ± 0.18	0.050 ± 0.004
BlendedMVS		
Character	9.98 ± 5.78	0.031 ± 0.343
Fountain	6.59 ± 0.18	-0.165 ± 0.016
Jade	1.84 ± 0.13	0.021 ± 0.003
Statues	30.34 ± 54.48	0.718 ± 0.926

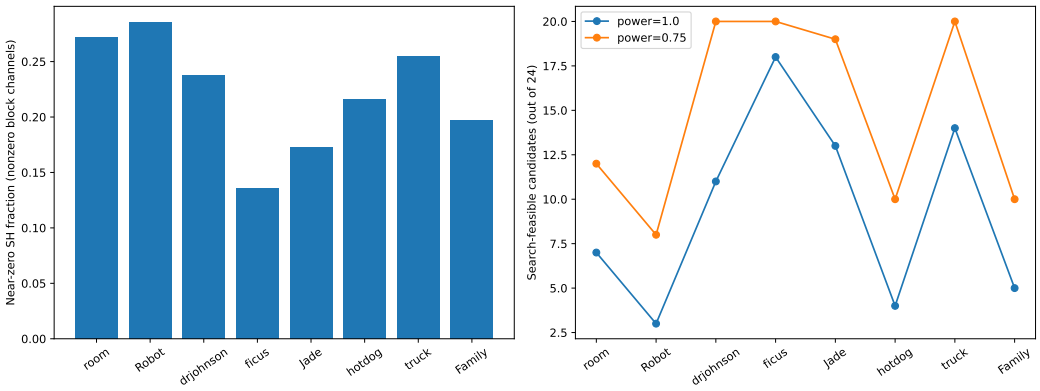


Figure 5: Matched-codec diagnostic across eight analyzed scenes. The left panel shows the measured fraction of near-zero SH coefficients; the right panel counts configurations satisfying the guarded search mean and tail limits (out of 24 per scene). Power 1.0 is Uniform and 0.75 is Ours. Feasible counts describe the full measured candidate pool, not the candidates visited by a budgeted search.

Table 8: Complete-pool search-feasible envelope: U is Uniform and C is Companded. Counts are out of 24 per scene; minima use actual archives. Companding improves both feasible count and minimum size in every scene. These search-only quantities exclude confirmation and fallback.

	Feasible count		Minimum MiB		
Scene	U	C	U	C	Saving (%)
360 room	7	12	12.121	11.732	3.21
Jade	13	19	3.444	3.230	6.23
Robot	3	8	2.317	2.056	11.27
Family	5	10	2.475	2.129	13.98
drjohnson	11	20	14.884	13.017	12.54
ficus	18	20	1.542	1.516	1.64
hotdog	4	10	1.790	1.573	12.11
truck	14	20	10.433	10.100	3.19

C.6 QUERY-BUDGET AND ORACLE RESULTS

All budget experiments use the same trained models. Random search shares the two initial queries with the other policies and uses five independent search repetitions, which are not additional training runs. Table 9 reports delivered size, including reference fallback.

Table 9: Delivered MiB at each query cap. Random shows mean and sample SD over five independent searches. The last column gives fallback percentages for Ours/greedy/random, respectively. Increasing the budget can trigger fallback after a smaller candidate is selected.

Scene	Q	Ours	Greedy	Random	F (%)
bonsai	5	12.256	14.338	13.505 ± 1.140	0/0/0
bonsai	10	10.826	14.338	12.417 ± 1.338	0/0/0
bonsai	20	10.826	10.826	11.160 ± 0.457	0/0/0
hotdog	5	2.039	2.039	1.926 ± 0.154	0/0/0
hotdog	10	1.667	2.039	1.776 ± 0.178	0/0/0
hotdog	20	1.573	1.573	1.611 ± 0.051	0/0/0
bicycle	5	42.657	50.718	38.337 ± 8.356	0/0/0
bicycle	10	32.555	30.206	32.823 ± 2.923	0/0/0
bicycle	20	30.206	30.206	31.145 ± 1.287	0/0/0
ficus	5	2.382	2.382	7.398 ± 7.297	0/0/40
ficus	10	15.388	15.388	9.942 ± 7.458	100/100/60
ficus	20	15.388	15.388	15.388 ± 0.000	100/100/100

Table 10: Gap to the smallest search-feasible companded configuration in the full pool. Ours and greedy use $Q = 10$. The oracle is evaluated only on search views and excludes confirmation and fallback.

Scene	Oracle MiB	Ours gap (%)	Greedy gap (%)
Family	2.129	8.31	0.00
hotdog	1.573	5.92	29.57
bicycle	30.206	7.78	0.00
ficus	1.516	5.92	0.00

Ours does not exactly hit the companded oracle in these four cases; greedy hits it in three but has a 29.57% gap on hotdog. Their mean gaps are 6.98% and 7.39%, respectively. This small diagnostic supports nearby search operating points, not global optimality or a general ranking of the policies. With 24 search queries and one confirmation, a full-pool delivery evaluation would require $Q = 25$; $Q = 20$ is not exhaustive.

C.7 CONFIRMATION AND TEST-RISK DETAILS

The no-confirmation control reuses the full method's selected candidate after nine queries, so it is a matched-candidate ablation, not a ten-search-query competitor. Mean-only removes tail limits from feasibility, active-base filtering, final selection, and confirmation; it retains the original local-mean exploration term and records tail statistics for analysis. The full method's three fallbacks occur on ficus. Its rejected candidates would not exceed the test tail limit, whereas all three drjohnson runs pass confirmation but exceed that limit on test views. Table 11 reports every full-method run together with the corresponding unconfirmed candidate, making this distinction explicit.

Table 11: Candidate and delivered outcomes for all 18 full-method runs. Runs 1–3 denote independent training realizations. Candidate columns describe the frozen candidate before confirmation; delivery columns include lossless fallback (F). Both CVaR columns are final-test diagnostics and are not used to revise delivery.

Scene	Run	Cand. MiB	Del. MiB	Cand. CVaR	Del. CVaR	F
360 room	1	11.732	11.732	1.228	1.228	N
360 room	2	11.006	11.006	1.003	1.003	N
360 room	3	11.506	11.506	0.923	0.923	N
Family	1	2.306	2.306	0.595	0.595	N
Family	2	1.569	1.569	0.572	0.572	N
Family	3	2.290	2.290	0.872	0.872	N
drjohnson	1	13.017	13.017	2.330	2.330	N
drjohnson	2	14.690	14.690	2.134	2.134	N
drjohnson	3	13.664	13.664	1.607	1.607	N
ficus	1	1.606	15.388	1.130	0.000	Y
ficus	2	1.610	15.488	1.145	0.000	Y
ficus	3	1.611	15.435	1.155	0.000	Y
hotdog	1	1.667	1.667	1.010	1.010	N
hotdog	2	1.641	1.641	1.166	1.166	N
hotdog	3	1.658	1.658	1.241	1.241	N
truck	1	10.968	10.968	0.724	0.724	N
truck	2	11.137	11.137	0.764	0.764	N
truck	3	11.008	11.008	0.820	0.820	N

C.8 MEASURED RATE–DISTORTION OPERATING POINTS

Each method is measured at three predetermined operating points. For Ours and Uniform, the points vary quality limits under the same query budget; the external baselines vary their native rate controls. Point indices are method-specific settings and do not imply matched quality or matched rate. Tables 12 and 13 retain all 90 evaluations, including repeated points and reference fallbacks. No extrapolated BD-rate is used.

Table 12: All measured RD points, shown as RGB8 PSNR (dB) / delivered MiB. F denotes reference fallback.

Method	Point 1	Point 2	Point 3
bonsai			
Uniform	32.636 / 14.117	32.169 / 11.353	32.169 / 11.353
Ours	32.672 / 14.499	32.256 / 10.923	31.884 / 10.829
HAC	32.834 / 15.925	32.425 / 10.841	32.127 / 8.617
ContextGS	32.685 / 12.939	32.589 / 9.163	32.301 / 7.332
C3DGS	31.570 / 12.665	31.693 / 13.098	31.744 / 13.497
garden			
Uniform	27.158 / 31.194	26.916 / 25.631	26.649 / 25.550
Ours	27.183 / 28.914	26.795 / 23.896	26.795 / 23.896
HAC	27.592 / 38.654	27.502 / 29.030	27.365 / 23.971
ContextGS	27.567 / 33.199	27.496 / 24.106	27.405 / 19.457
C3DGS	26.814 / 45.171	26.847 / 47.100	26.874 / 48.516
truck			
Uniform	25.892 / 128.476^{F}	25.470 / 11.392	25.050 / 10.292
Ours	25.892 / 128.476^{F}	25.441 / 10.822	25.441 / 10.822
HAC	26.165 / 15.136	26.065 / 11.362	25.933 / 9.111
ContextGS	25.990 / 12.542	26.027 / 9.112	25.943 / 7.774
C3DGS	25.199 / 20.924	25.217 / 21.647	25.223 / 22.186
Mip-NeRF 360 room			
Uniform	31.580 / 20.006	31.195 / 18.055	31.882 / 14.576
Ours	31.628 / 19.328	31.239 / 17.069	31.775 / 12.225
HAC	31.923 / 9.621	31.729 / 6.668	31.345 / 5.471
ContextGS	31.878 / 8.341	31.769 / 5.813	31.619 / 4.644
C3DGS	31.111 / 14.500	31.153 / 14.908	31.141 / 15.284
Robot			
Uniform	39.265 / 3.455	38.751 / 2.317	38.030 / 2.074
Ours	39.001 / 2.885	38.514 / 2.056	38.192 / 1.817
HAC	38.350 / 0.956	36.245 / 0.708	34.634 / 0.603
ContextGS	38.336 / 1.001	36.795 / 0.768	35.511 / 0.684
C3DGS	38.095 / 2.783	38.338 / 2.943	38.498 / 3.144
Jade			
Uniform	27.813 / 4.215	27.768 / 3.752	27.768 / 3.752
Ours	27.826 / 3.902	27.792 / 3.688	27.547 / 3.230
HAC	27.450 / 0.565	27.545 / 0.436	27.477 / 0.396
ContextGS	27.354 / 0.615	27.447 / 0.532	27.336 / 0.487
C3DGS	28.190 / 5.999	28.187 / 6.198	28.182 / 6.450

Table 13: SSIM / LPIPS for the same measured operating points as Table 12, with identical ordering and RGB8 evaluation.

Method	Point 1	Point 2	Point 3
bonsai			
Uniform	0.9401 / 0.2587	0.9337 / 0.2739	0.9337 / 0.2739
Ours	0.9405 / 0.2575	0.9358 / 0.2708	0.9321 / 0.2797
HAC	0.9434 / 0.2576	0.9398 / 0.2632	0.9362 / 0.2685
ContextGS	0.9430 / 0.2588	0.9415 / 0.2620	0.9392 / 0.2668
C3DGS	0.9335 / 0.2645	0.9347 / 0.2627	0.9354 / 0.2618
garden			
Uniform	0.8433 / 0.1838	0.8350 / 0.2032	0.8266 / 0.2216
Ours	0.8446 / 0.1813	0.8318 / 0.2108	0.8318 / 0.2108
HAC	0.8609 / 0.1398	0.8581 / 0.1457	0.8535 / 0.1536
ContextGS	0.8610 / 0.1391	0.8588 / 0.1440	0.8563 / 0.1504
C3DGS	0.8458 / 0.1621	0.8475 / 0.1592	0.8486 / 0.1576
truck			
Uniform	0.8877 / 0.1696	0.8715 / 0.2129	0.8575 / 0.2386
Ours	0.8877 / 0.1696	0.8700 / 0.2143	0.8700 / 0.2143
HAC	0.8899 / 0.1715	0.8872 / 0.1777	0.8841 / 0.1851
ContextGS	0.8870 / 0.1770	0.8869 / 0.1797	0.8842 / 0.1855
C3DGS	0.8754 / 0.1889	0.8761 / 0.1878	0.8765 / 0.1871
Mip-NeRF 360 room			
Uniform	0.9127 / 0.2957	0.9058 / 0.3111	0.9165 / 0.2953
Ours	0.9159 / 0.2933	0.9067 / 0.3099	0.9148 / 0.2996
HAC	0.9208 / 0.2880	0.9172 / 0.2979	0.9131 / 0.3055
ContextGS	0.9203 / 0.2896	0.9184 / 0.2955	0.9152 / 0.3021
C3DGS	0.9118 / 0.2992	0.9125 / 0.2977	0.9127 / 0.2971
Robot			
Uniform	0.9954 / 0.0103	0.9949 / 0.0116	0.9937 / 0.0143
Ours	0.9949 / 0.0112	0.9945 / 0.0125	0.9938 / 0.0139
HAC	0.9948 / 0.0111	0.9927 / 0.0136	0.9900 / 0.0165
ContextGS	0.9948 / 0.0111	0.9933 / 0.0129	0.9915 / 0.0149
C3DGS	0.9946 / 0.0108	0.9948 / 0.0106	0.9950 / 0.0104
Jade			
Uniform	0.9204 / 0.1005	0.9185 / 0.1050	0.9185 / 0.1050
Ours	0.9209 / 0.0993	0.9196 / 0.1023	0.9115 / 0.1177
HAC	0.9167 / 0.1186	0.9167 / 0.1214	0.9157 / 0.1250
ContextGS	0.9160 / 0.1201	0.9167 / 0.1209	0.9143 / 0.1262
C3DGS	0.9347 / 0.0807	0.9348 / 0.0805	0.9347 / 0.0801

C.9 Exploratory quality-matched follow-up

The 90 predetermined rate–distortion points above are measured outputs, not quality-matched method pairs. Across six scenes, the PSNR ranges of Ours and another method overlap in 16 of 30 scene–method comparisons; 14 have no overlap. At the midpoint of each overlapping range, descriptive interpolation of log bytes favors Ours in eight comparisons and the other method in eight. Interpolation does not add measured operating points or support a universal matched-quality claim.

After inspecting these gaps, we evaluated five additional Ours settings on garden, Mip-NeRF 360 room, and Jade, choosing the settings on validation views before testing. Three illustrative pairs have RGB8 PSNR gaps below 0.1 dB: on garden, a strict limit causes reference fallback (260.903 MiB versus 19.457 MiB for ContextGS, a 0.013 dB gap); on room, Ours uses 10.093 MiB versus 14.908 MiB for C3DGS (0.031 dB gap); on Jade, Ours uses 3.024 MiB versus 0.487 MiB for ContextGS (0.021 dB gap). These exploratory comparisons were motivated by previously observed test gaps and are not an independent prospective evaluation. They expose an advantage at one matched point and substantial losses where fallback or a stronger competing codec dominates; the remaining unmatched ranges are not extrapolated.

C.10 PREPARATION, DECODING, AND RENDERING COSTS

Table 14 separates preparation from decoding. Teacher time is added for postprocessing methods; Ours and Uniform include both independently trained bases. The recorded search total includes reference rendering, candidate encoding/decoding and evaluation, confirmation, and fallback as applicable. Existing models can be reused when changing quality limits; their historical training cost remains part of from-scratch accounting. Timings include reused historical runs and runs under concurrent machine load, and are descriptive rather than controlled cross-method training speedups. Rendering throughput in Figure 4 excludes all preparation and decoding.

Table 14: Recorded mean times (s) over all 12 external-comparison scenes. Own denotes method training/postprocessing or our two-base library; Teacher is additional upstream training. Prep. is the recorded sum of Own, Teacher, and Search, excluding Decode.

Method	Own	Teacher	Search	Prep.	Decode
3DGS	707.83	0.00	0.00	707.83	2.11
FastGS	142.12	0.00	0.00	142.12	0.51
LightGaussian	270.68	707.83	0.00	978.51	0.85
Compact-3DGS	883.61	0.00	0.00	883.61	100.74
C3DGS	131.30	707.83	0.00	839.12	0.51
HAC	1043.61	0.00	0.00	1043.61	6.15
ContextGS	1817.86	0.00	0.00	1817.86	31.01
Uniform	345.63	0.00	56.42	402.05	1.89
Ours	345.63	0.00	58.25	403.88	2.42

C.11 PAIRED DELIVERY GAINS AND THEIR SOURCE

The matched full-pool analysis in Table 8 establishes the codec mechanism: companding expands the guarded-feasible set and lowers its minimum-byte envelope in all eight evaluated scenes. Here we examine how this advantage transfers to the representation actually delivered. We reanalyse the archived default-protocol runs for training seeds 19, 29, and 39: eight existing-domain scenes (24 pairs) and twelve additional-domain scenes (36 pairs). Every pair uses the same scene, training seed, base library, candidate pool, quality limits, and query budget; only the SH transform changes. Both codecs use the local-mean policy. These are paired reanalyses of existing runs, not newly trained replicates, and the two cohorts are kept separate.

From candidate savings to delivered savings. For each run, let the fallback increment be $P = R_{\mathrm{del}} - R(c)$, which is zero when the selected candidate is delivered. Consequently, $\Delta R_{\mathrm{del}} = \Delta R(c) + \Delta P$ for the companded-minus-uniform difference. Table 15 reports sums over all runs, retaining every fallback. In the additional domains, companding reduces selected-candidate bytes by 3.201 MiB and avoids 33.471 MiB of fallback increments, yielding **27.24% fewer total delivered bytes**. It delivers fewer bytes in 24/36 paired runs; all 36 companded deliveries satisfy both the empirical final-test mean-loss and CVaR_{10} thresholds. The full per-scene and perceptual-quality results remain in Appendix C.4.

Table 15: Paired delivery accounting for three training seeds. All byte columns are *totals* in MiB, not per-run means. Candidate plus fallback increment equals delivered size. F counts reference fallbacks; no fallback is removed from the totals. Both arms use local-mean search.

Cohort	Codec	Candidate	Fallback incr.	Delivered	F
Existing domain	Uniform	246.535	129.781	376.316	3/24
	Companded	233.057	171.492	404.548	6/24
Additional domains	Uniform	101.164	33.471	134.634	1/36
	Companded	97.963	0.000	97.963	0/36

Sensitivity and scope of the gain. One avoided Uniform fallback on Statues (seed 19) contributes 91.27% of the additional-domain total byte saving. Removing the *entire* Statues scene, including all three seeds, still leaves **3.71% fewer delivered bytes** over the remaining eleven scenes (33 pairs). Thus the direction of the aggregate gain survives this sensitivity check, while its magnitude is much smaller. The large 27.24% reduction should be interpreted as the observed combination of ordinary candidate savings and one avoided high-cost fallback, rather than a typical per-scene saving. It is not an improvement in every image metric: the Statues fallback comparison also increases LPIPS, as reported in Appendix C.4.

The existing-domain cohort illustrates the converse. Its 13.478 MiB candidate saving is outweighed by 41.711 MiB of additional fallback increments, increasing total delivered bytes by 7.50%, despite smaller deliveries in 17/24 pairs. We report these cohorts separately because their delivery behavior differs. Scene-cluster 95% bootstrap intervals for the delivered-byte ratio (Companded/Uniform), with all three seeds retained within each sampled scene, are $[0.866, 1.607]$ and $[0.467, 0.986]$ for the existing and additional domains, respectively. With only eight or twelve scenes, these intervals are descriptive sensitivity summaries. The evidence supports the codec's expanded feasible envelope and a delivery benefit in the additional domains; it does not establish universal delivery savings or a statistical guarantee on unseen views.

C.12 COMPLETE MEASURED COMPARISON WITH FEEDFORWARD COMPRESSION

Frozen operating points and common inputs. We add FCGS (Chen et al., 2025a), a feedforward codec for pretrained Gaussian scenes, to the measured operating-point comparison. The matrix contains six scenes, both cached inputs (small and reference), and three predeclared checkpoints, $\lambda \in \{1, 4, 16\} \times 10^{-4}$: 36 points in total. The scenes are bonsai, garden, truck, room, Robot, and Jade. The first four use training seed 9 and the latter two seed 19, matching the corresponding archived comparison cohort. Both input bases are reported independently; no input or checkpoint is selected by final-test quality. Decoded models enter the same pinned RGB8 evaluator used for the other measured curves. Table 16 lists all 36 points, and Figures 6 and 7 show the measured PSNR and LPIPS trade-offs, respectively. Only actual measured points are plotted; no interpolation or common quality target is imposed.

Numerical compatibility and actual decoding. The unmodified implementation completed all 36 processes, but 19 streams failed an attribute-level comparison between encoder reconstruction and the actual decoder output. Those damaged outputs are excluded from rate–distortion claims and remain archived with their failure labels. The issue was traced to batch-shape-dependent rounding in affine layers, propagated through entropy contexts. We apply one compatibility amendment uniformly to all 36 points: contiguous padded batches of 4,096 rows for rowwise linear layers, with TF32 disabled. Checkpoint weights, quantization, chunk limits, and scene/camera settings are unchanged. The amendment was accepted using attribute roundtrip consistency, without image-quality selection. All 36 corrected streams pass the attribute audit after encoding and decoding in separate processes. Expected encoder tensors are consumed only by a post-decoding observer; they are not decoder inputs. We therefore identify the plotted baseline as *FCGS (adapted)*. This is a reproducible compatibility repair on the evaluated setup, not a claim that upstream FCGS guarantees numerical determinism on every device.

Scene bytes, shared assets, and point-specific costs. FCGS rates count the complete scene stream. Each selected checkpoint requires a 10,223,425-byte shared decoder asset, reported separately under a preinstalled-decoder assumption. A fresh installation serving one scene must additionally distribute the required checkpoint; retaining all three checkpoints requires three such assets. The PSNR and LPIPS plots use scene-stream bytes for FCGS and complete delivered scene artifacts for our method. Shared decoder assets are excluded from the per-scene comparison under this assumption. Table 16 reports the measured encoding, decoding, and rendering costs of each operating point. Codec entrypoint times exclude the separately measured diagnostic observer; they are not full preparation times and exclude base training. Historical and new runs were collected under shared-machine load, so these data do not support a controlled cross-method timing speedup.

Where the measured trade-offs are competitive. The added comparison provides concrete regions of competitiveness for our method. On truck, our default delivered artifact uses 10.822 MiB at 25.441 dB, versus 11.651 MiB at 25.050 dB for the FCGS small-input 4×10^{-4} point: fewer scene bytes and higher RGB8 PSNR. On Robot, our relaxed operating point uses 1.817 MiB at 38.192 dB, compared with 1.836 MiB at 38.102 dB for the same FCGS checkpoint on the small input. These comparisons concern the measured byte–PSNR coordinates only. They are balanced by opposing examples: the FCGS small-input 16×10^{-4} point on garden uses 19.556 MiB at 26.874 dB, versus our default 23.896 MiB at 26.795 dB; on Jade, the FCGS small-input 4×10^{-4} point uses 3.522 MiB at 27.837 dB, versus our default 3.688 MiB at 27.792 dB. The complete curves and table retain both directions. Accordingly, the contribution is a competitive set of post-training operating points on some scenes, rather than universal dominance over feedforward compression. Perceptual quality and throughput remain separate coordinates; a byte–PSNR advantage is not interpreted as a win on every metric.

Table 16: All 36 corrected FCGS operating points. λ is in units of 10^{-4}. Size is scene-stream MiB; PSNR is in dB; Enc. and Dec. are codec-entrypoint seconds. Quality uses the common RGB8 evaluator. FPS is measured separately at each point. Shared decoder weights (10,223,425 bytes per checkpoint) are additional. No point is omitted on the basis of rate, quality, or runtime.

Input	λ	MiB	PSNR	SSIM	LPIPS	FPS	Enc.	Dec.
bonsai (Mip-NeRF 360; seed 9)								
Small	1	11.778	32.313	0.9364	0.2605	1115.8	4.91	5.57
Small	4	9.798	32.232	0.9354	0.2631	1071.7	3.70	5.16
Small	16	7.331	31.819	0.9305	0.2718	1083.2	3.65	5.02
Reference	1	17.671	32.517	0.9407	0.2468	953.8	5.44	8.48
Reference	4	14.584	32.434	0.9395	0.2498	931.4	4.41	7.10
Reference	16	10.842	32.016	0.9341	0.2597	901.4	4.88	6.78
garden (Mip-NeRF 360; seed 9)								
Small	1	31.788	27.154	0.8422	0.1769	843.0	9.49	12.51
Small	4	26.670	27.104	0.8407	0.1797	850.4	7.27	10.73
Small	16	19.556	26.874	0.8338	0.1909	843.6	6.07	10.45
Reference	1	46.845	27.297	0.8550	0.1478	677.2	14.09	18.55
Reference	4	38.780	27.250	0.8535	0.1507	683.8	10.53	16.42
Reference	16	28.192	27.049	0.8469	0.1622	686.7	9.59	14.13
truck (Tanks and Temples; seed 9)								
Small	1	14.387	25.090	0.8533	0.2102	1160.9	6.33	7.92
Small	4	11.651	25.050	0.8520	0.2123	1109.3	4.22	5.69
Small	16	8.333	24.909	0.8466	0.2198	1160.0	3.54	6.01
Reference	1	21.570	24.117	0.8057	0.2324	990.7	6.59	10.64
Reference	4	17.301	24.094	0.8046	0.2342	947.2	6.34	8.18
Reference	16	12.179	24.003	0.8003	0.2408	984.6	5.24	8.28
room (Mip-NeRF 360; seed 9)								
Small	1	11.275	31.539	0.9098	0.3011	1089.7	4.76	5.83
Small	4	9.146	31.479	0.9086	0.3037	1086.3	3.42	5.46
Small	16	6.603	31.178	0.9034	0.3126	1107.1	4.23	4.73
Reference	1	19.251	31.812	0.9205	0.2833	961.3	6.82	9.02
Reference	4	15.291	31.729	0.9188	0.2869	972.0	5.28	8.43
Reference	16	10.863	31.417	0.9125	0.2990	949.2	4.76	8.17
Robot (Synthetic-NSVF; seed 19)								
Small	1	2.211	38.738	0.9949	0.0115	1869.2	1.46	1.97
Small	4	1.836	38.102	0.9938	0.0131	1970.0	1.44	2.00
Small	16	1.319	36.206	0.9909	0.0167	1977.7	0.85	1.88
Reference	1	3.151	39.239	0.9953	0.0104	1760.1	1.78	2.11
Reference	4	2.590	38.656	0.9945	0.0117	1712.4	1.57	2.20
Reference	16	1.787	36.604	0.9918	0.0149	1822.1	0.99	2.10
Jade (BlendedMVS; seed 19)								
Small	1	4.410	27.894	0.9235	0.0930	1766.2	1.84	2.62
Small	4	3.522	27.837	0.9215	0.0961	1691.6	1.82	2.45
Small	16	2.117	27.498	0.9124	0.1093	1701.0	1.76	2.34
Reference	1	6.054	28.088	0.9234	0.0907	1437.7	2.16	3.09
Reference	4	4.802	28.038	0.9214	0.0938	1434.7	1.80	2.77
Reference	16	2.793	27.714	0.9119	0.1069	1490.6	1.68	2.82

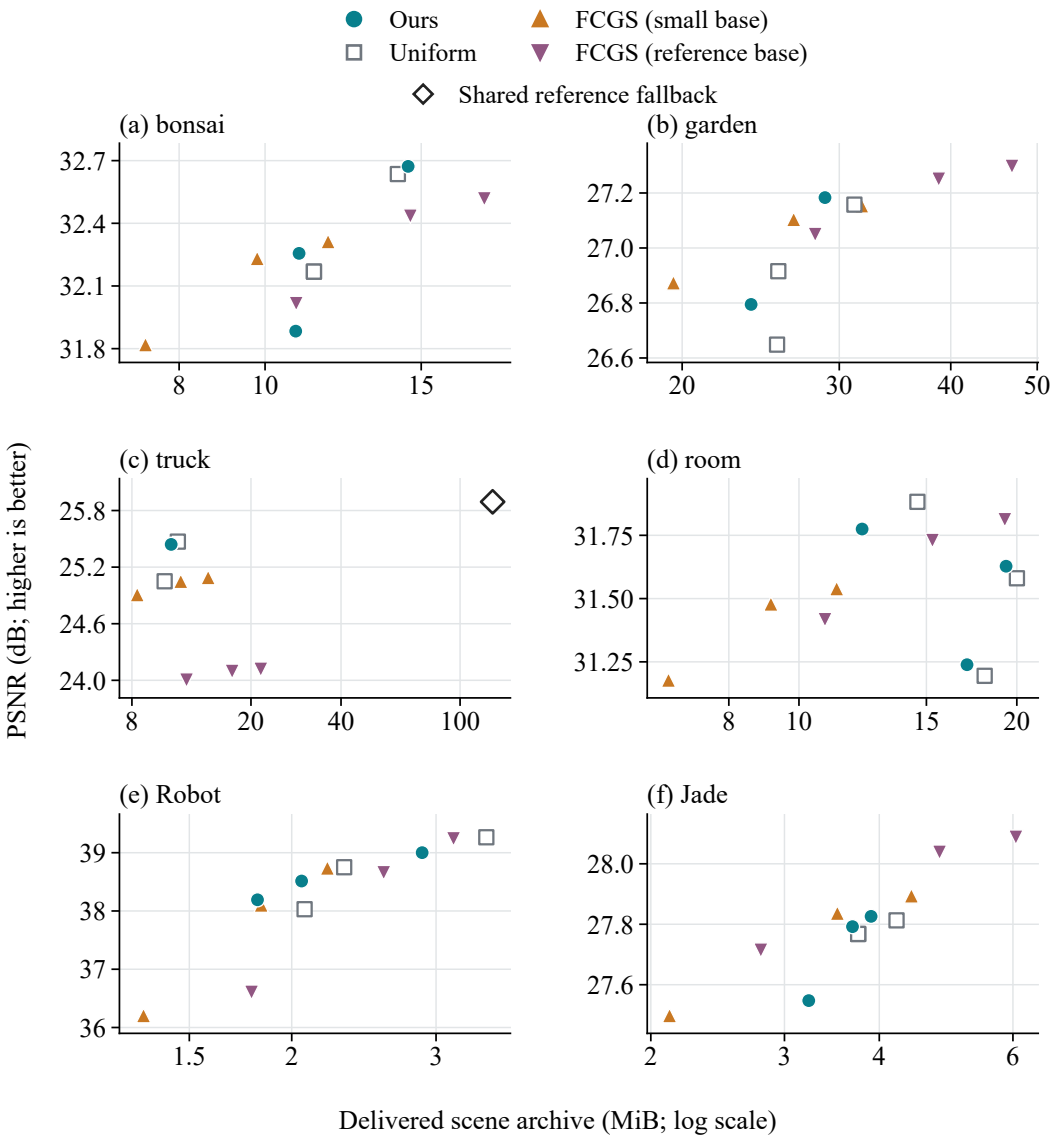


Figure 6: Measured scene-stream size versus RGB8 PSNR on the six frozen FCGS scenes. Both FCGS inputs and all three checkpoints are retained; FCGS uses the uniformly corrected implementation and a preinstalled shared decoder. Our points include any reference fallback and use the same archived quality-limit settings as the existing measured curves. Only measured operating points are plotted. Each panel has its own axis range; byte–PSNR competitiveness varies with scene and setting.

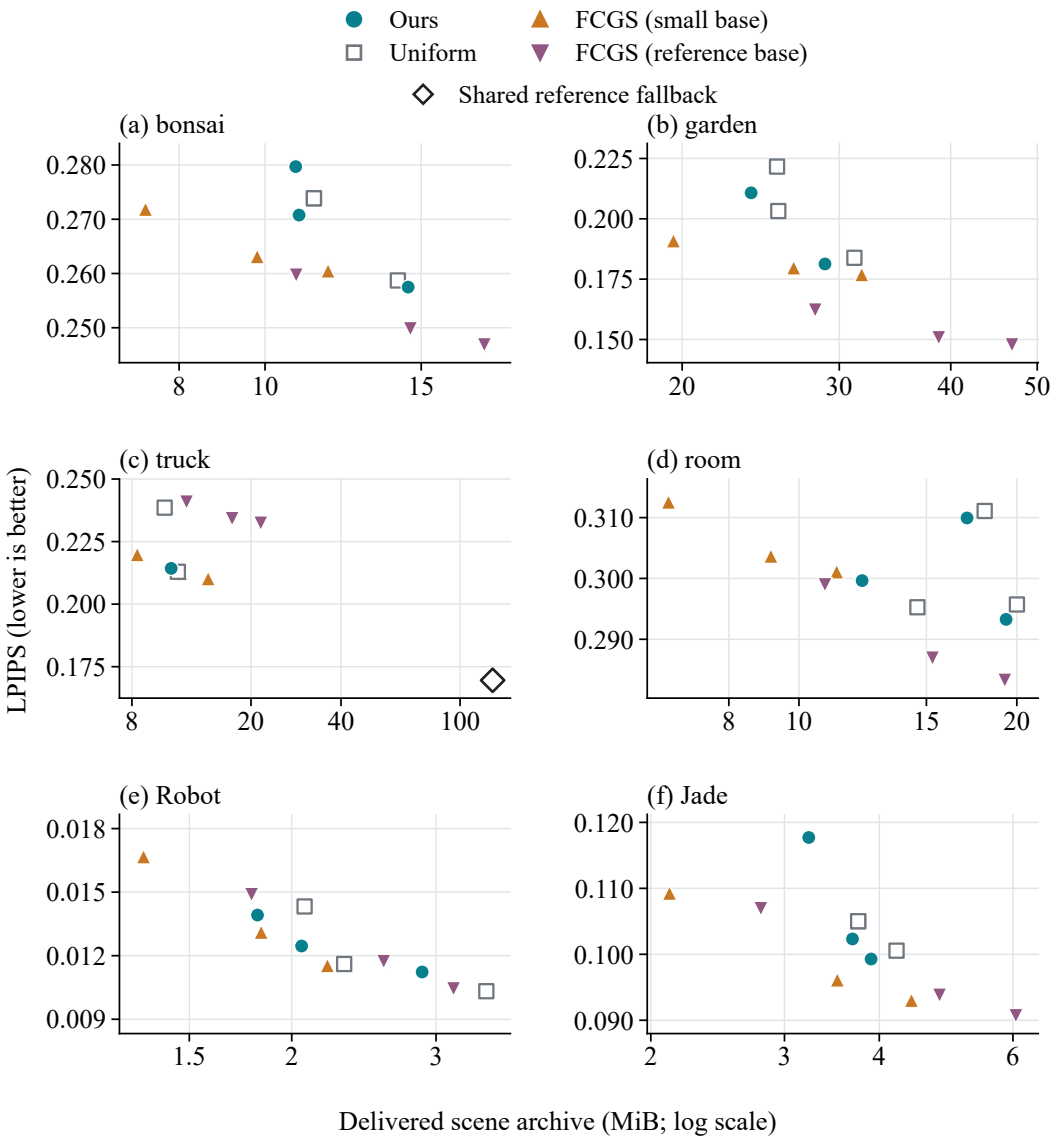


Figure 7: Measured scene-stream size versus VGG LPIPS for the same operating points as Figure 6; lower LPIPS is better. Reporting the perceptual coordinate separately avoids treating PSNR-based feasibility or byte–PSNR competitiveness as an improvement in every image metric. FCGS checkpoint assets use the same separately disclosed accounting as in Figure 6.

D ADDITIONAL QUALITATIVE COMPARISONS

Figures 8–11 compare delivered reconstructions with four external baselines per scene, covering all seven external methods in Table 3 across the examples. The figures use held-out test views. Within each scene, every panel uses the same camera and image crop; the boxed region is enlarged in the inset. The displayed views are representative examples rather than view-wise performance rankings. Table 17 reports complete artifact sizes and scene-mean RGB8 PSNR for all compared methods. Each reconstruction panel also labels the complete model size in MiB; ground truth has no model size. The scene means should not be read as the PSNR of a single displayed view. All four Ours artifacts passed confirmation and were delivered as compressed candidates.

Table 17: Complete delivered model size and scene-mean RGB8 PSNR for the displayed scenes. The baselines and Ours are evaluated on matched test views; all Ours artifacts are accepted compressed candidates. MiB is 2^{20} bytes.

Method	Hotdog	Family	Truck	Robot
Complete model size (MiB) ↓				
3DGS	34.446	53.612	600.807	67.041
FastGS	4.729	5.507	58.149	5.392
LightGaussian	2.413	3.696	39.679	4.570
Compact-3DGS	1.576	5.648	20.032	1.617
C3DGS	2.832	3.533	21.647	2.943
HAC	0.639	1.071	9.111	0.603
ContextGS	0.743	1.101	7.774	0.684
Uniform	1.803	2.442	11.392	2.317
Ours	**1.678**	**2.270**	**10.822**	**2.056**
Scene-mean RGB8 PSNR (dB) ↑				
3DGS	37.552	24.911	25.387	39.457
FastGS	36.476	25.462	25.776	36.314
LightGaussian	34.980	24.411	25.097	35.363
Compact-3DGS	36.498	22.054	25.015	36.992
C3DGS	37.120	25.380	25.217	38.338
HAC	36.344	15.743	25.933	34.634
ContextGS	36.335	20.671	25.943	35.511
Uniform	37.069	27.430	25.470	38.751
Ours	**37.174**	**27.587**	**25.441**	**38.514**



Figure 8: NeRF Synthetic Hotdog. Ground truth and four external baselines are shown alongside the delivered Ours reconstruction. Each inset enlarges the same boxed region.

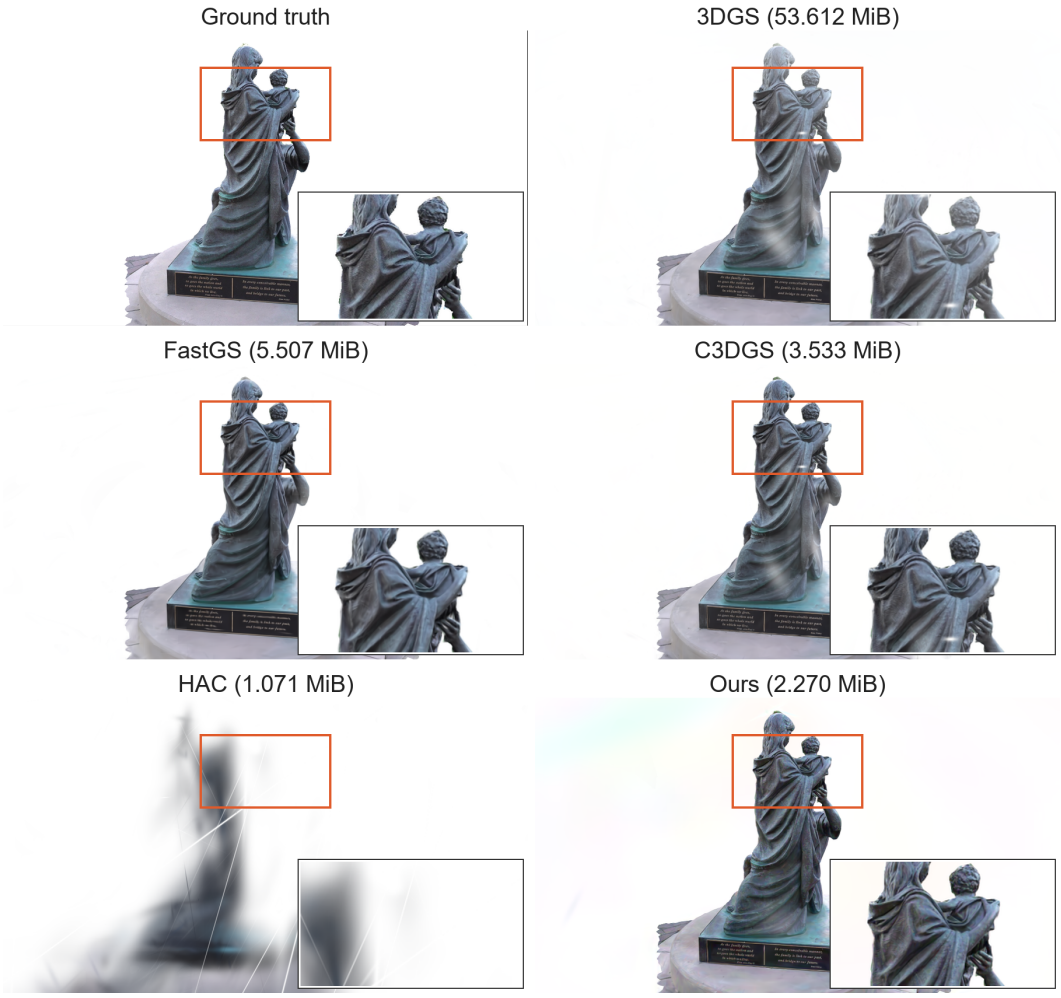


Figure 9: Tanks and Temples Family. The shared test view and boxed crop allow direct comparison of scene detail across delivered methods.



Figure 10: Tanks and Temples Truck. Ours has a lower scene-mean PSNR than FastGS at a substantially smaller delivered size; see Table 17.

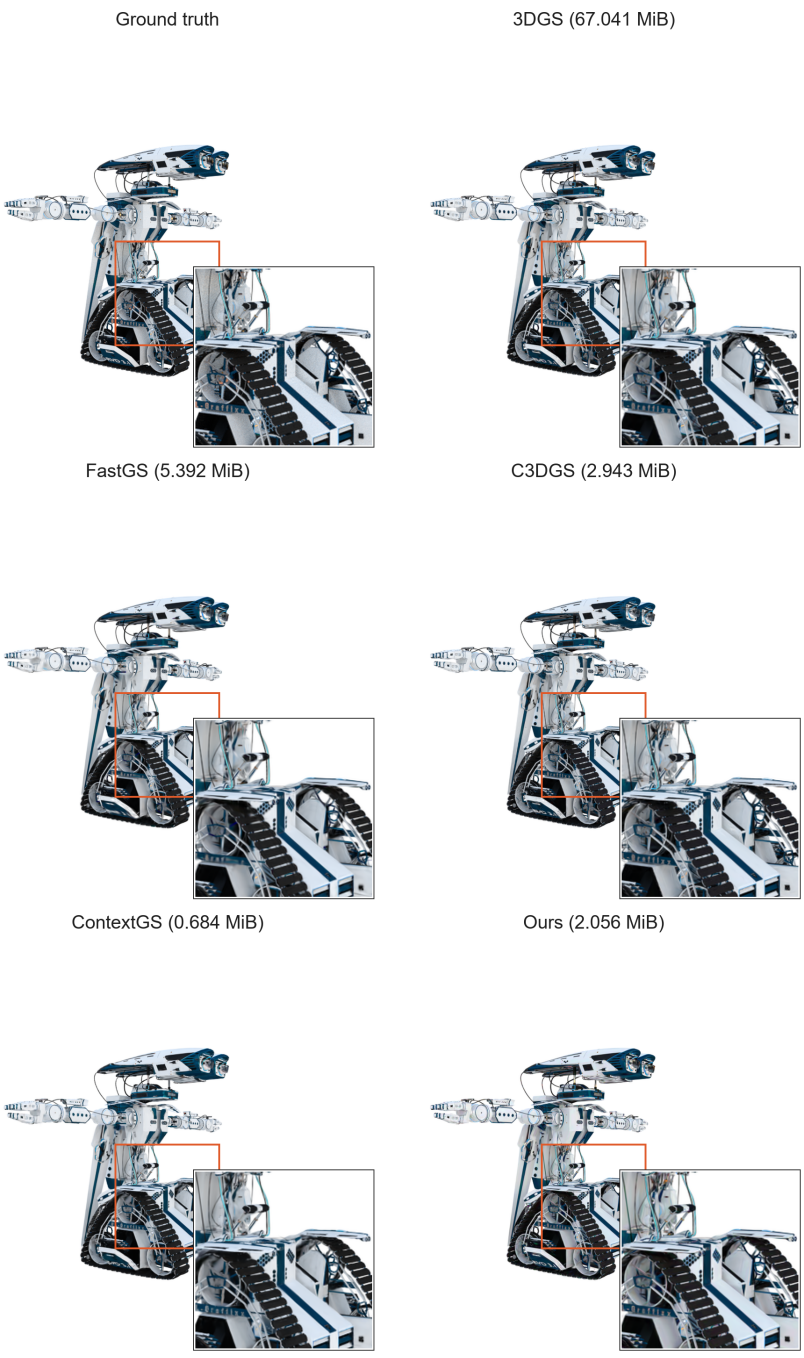


Figure 11: Synthetic-NSVF Robot. All methods are shown on the same held-out test view, with a matched detail crop.