SAME SCENE, SAME EDIT: CANONICAL GAUGE FIXING FOR STABLE 3D GAUSSIAN EDITING

Anonymous authors
Paper under double-blind review

ABSTRACT

Editing a 3D Gaussian Splatting (3DGS) scene should be invariant to its Gaussian parameterization. Yet co-located K-way splits preserving optical thickness $\tau = -\log(1-\alpha)$ pass a strict 60 dB input gate but induce up to 82.6% relative RMSE deviation in a state-of-the-art editor, increasing with K and persisting on held-out views. We trace this instability to opacity-threshold culling, duplicated optimizer states, and row-dependent selection. Our training-free wrapper, GaugeFix3D, canonicalizes refinement lineages before running the unchanged editor, using exact merging for co-located siblings, τ-weighted barycenters under jitter, and signature-based recovery without metadata. Across material, local appearance, environment, and object-level semantic edits, GaugeFix3D removes 78.4–88.0% of representation-induced edit error in every exact-regime cell and 65.2–75.6% under metadata-free approximate refinements. In our main configuration, edit quality improves from 43.41 to 58.61 dB PSNR ($+15.2$ dB), within 2.2% of the editor's identical-input repeat floor, while retaining $\geq 99.98\%$ of the author's edit magnitude. After continued optimization, 97% of lineages remain recoverable at practical tolerance. The same signature structure occurs in unaugmented checkpoints, and the canonical interface transfers to a second, architecturally distinct editor family.

1 INTRODUCTION

Text-driven 3D editing has rapidly matured: an operator writes an instruction, a diffusion prior proposes per-view edits, and an optimizer lifts them into the scene representation (Haque et al., 2023; Sella et al., 2023; Wu et al., 2024; Wang et al., 2024; Zhuang et al., 2024; Lee et al., 2025). With 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) becoming the dominant explicit representation (Fei et al., 2025; Chen & Wang, 2026; Zhu et al., 2026), Gaussian checkpoints are on their way to becoming editable asset files, increasingly pass through pipelines that alter their primitive representation while preserving the underlying scene (Hong et al., 2025; Jeong et al., 2026; Han et al., 2026; Kwon & Baek, 2026).

This trajectory exposes a question that, to our knowledge, has not been posed as an explicit requirement: *is a Gaussian editor a function of the scene, or of the particular Gaussian mixture that encodes it?* A 3DGS checkpoint is a redundant mixture: identical visible scenes admit many primitive sets, and the same pipeline that produces a checkpoint produces variants of it: densification and its revisions (Rota Bulò et al., 2024; Wang et al., 2025; Kheradmand et al., 2024), compression and pruning (Lee et al., 2024; Fang & Wang, 2024), level-of-detail assembly and streaming (Kerbl et al., 2024; Shi et al., 2025; Meuleman et al., 2025), and reorganizations studied for their own sake (Wang et al., 2026; Feng et al., 2024). If the editor's output depends on which equivalent encoding it receives, the editing interface is not defined on the represented 3D function, and pipelines that compress or migrate assets silently change the edit a user receives for the same instruction.

We study this question with a controlled stress protocol. Starting from a trained checkpoint, we replace a small fraction of Gaussians by K co-located children that exactly preserve optical thickness, $\tau = -\log(1-\alpha)$, so the rendered scene is unchanged to within a strict input gate (all worst-view PSNR $\geq$ 60 dB; global 62.5–69.5 dB in our experiments). This transformation touches nothing visible, yet it alters the editor's internal state space: the number of rows, their opacities, and their optimizer states all change. We then run the same editor with the same frozen 2D edit targets, seed, prompt, and schedule on both checkpoints, and measure the RGB deviation between the two

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.

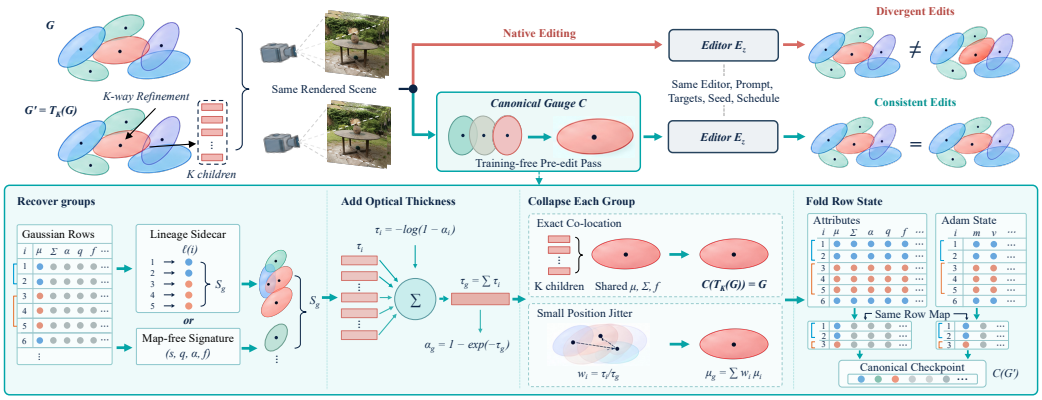

Figure 1: Overview of GaugeFix3D. A Gaussian scene G and its K-way refinement $G' = T_K(G)$ yield nearly identical renderings but can produce divergent edits under the same editor E_z. GaugeFix3D inserts a training-free canonicalization step C before editing, producing consistent edits without modifying the editor. The canonicalizer recovers refinement groups, aggregates opacity in optical-thickness space, collapses each group, and folds Gaussian attributes and optimizer state using a shared row mapping. Both editing paths use the same prompt, targets, seed, and schedule.

edits, relative to a repeat floor obtained by running the editor twice on the identical input. The result is unambiguous (Table 1, 2). For a state-of-the-art editor (GaussCtrl (Wu et al., 2024)), input perturbations invisible to any viewer: global input PSNR above 62 dB, are amplified into edits that differ by 0.0068 RMSE (43.4 dB) where the identical-input repeat floor is 0.0010 (60.0 dB); the amplification grows with the refinement multiplicity K and persists on held-out midpoint views. This is not a rendering failure: both edited checkpoints render plausibly, but they realize *different 3D functions*, so downstream consumers of the asset disagree about geometry and appearance.

Our remedy removes the nuisance degree of freedom *before* the editor's nonlinearities can see it. GaugeFix3D maps each checkpoint to a canonical representative of its refinement orbit. For exactly co-located siblings the collapse is algebraically exact, $C(T_K(G)) = G$, because optical thickness is additive under the split; for siblings under small local jitter we collapse them to a τ-weighted barycenter, whose first-order error cancels when the split offsets have zero τ-weighted first moment; and when no lineage metadata exists, we recover groups from exact attribute signatures (scale, quaternion, opacity, spherical-harmonic bytes) filtered by the known multiplicity, 100% recovery with zero false merges in our map-free tests. The wrapper changes nothing else: prompt, targets, schedule, and editor code are untouched, and runs in $O(N)$ (∼8 s for 706k primitives on CPU).

Across diverse scenes and editing tasks, GaugeFix3D reduces the representation-induced edit error by 78.4–88.0% in every exact-regime cell and by 65.2–75.6% under metadata-free approximate refinements (paired-view bootstrap 95% CIs strictly above zero; $P(\Delta > 0) = 1.0$ in 100,000 replicates). On the representative configuration (Dinosaur, K=8, novel views), edit quality improves from 43.41 to 58.61 dB PSNR, to within 17.2% of the identical-input repeat floor. The gain is not bought by weakening the edit: LPIPS distance to the author edit drops by 93–99% and edit magnitude remains $\geq 99.98\%$ of the author's. Mechanism ablations show partial fixes (group-level culling, quotient gradients) recover only 34–35%, while parent-survival mismatch falls from 0.756 to 4.5×10^{-4} only under full canonicalization. A second editing family (a physical-token style editor) provides a positive control: mapping the Gaussian measure to a fixed physical field before editing yields stability advantages of 63–74% over the strongest perturbation control, and through the identical canonical-checkpoint interface reproduces the author edit to 10^{-8} RMSE under exact refinement.

Our contributions are summarized as follows:

- A well-definedness requirement and stress benchmark: editing formalized as a function on render-equivalence classes, with input-gated K-way optical-thickness refinements as a controllable family of benign decompositions;

- Canonical Gauge Fixing: an $O(N)$, training-free wrapper that collapses refinement lineages in optical-thickness coordinates (exact, barycentric, and metadata-free variants), leaving the editor untouched;
- Empirical evidence that canonicalization suppresses representation-induced edit discrepancies toward the identical-input repeat floor while preserving edit strength.

2 Related Work

Representation Sensitivity in 3DGS. Reconstruction research increasingly recognizes that 3DGS encodings are non-unique: ReorgGS reorganizes equivalent distributions to improve optimization (Wang et al., 2026); densification studies expose splitting- and cloning-induced pathologies (Rota Bulò et al., 2024; Wang et al., 2025; Kheradmand et al., 2024); compression and LOD methods reshape mixtures (Lee et al., 2024; Fang & Wang, 2024; Kerbl et al., 2024; Shi et al., 2025); and EVSplitting preserves visual consistency during splitting (Feng et al., 2024). Scaffold-GS (Lu et al., 2024) and CAGS (Zhou et al., 2026) constrain redundancy through anchoring, while implementations and optimizers introduce numerical differences (Ye et al., 2025; Tancik et al., 2023; Hahlbohm et al., 2026; Höllein et al., 2025). We instead study a *frozen downstream editor*'s sensitivity to benign reshuffles. Following the classical principle of operating on equivalence classes (Cohen & Welling, 2014), with precedents in order-invariant representations (Zhu et al., 2024) and compositing algebra (Duff, 2017), we use refinement-additive optical thickness and a canonicalizing quotient map to remove this sensitivity.

Robustness of Gaussian Assets. Existing protections target adversarial editing through lifted perturbations (Hong et al., 2025), ownership tracing and edit deterrence (Jeong et al., 2026), tamper localization (Han et al., 2026), and audits of multi-view inconsistency attacks (Kwon & Baek, 2026). Our complementary goal is to stabilize intended edits under provably invisible, benign transformations. To our knowledge, this is the first application of parameterization invariance to downstream operators on explicit Gaussian mixtures.

3 Problem Setup: Editing as a Quotient-Space Function

3.1 Scenes, render equivalence, and quotient well-definedness

A Gaussian checkpoint $G = \{(\mu_i, \Sigma_i, \alpha_i, f_i)\}_{i=1}^{N}$ consists of positions, covariances, opacities, and spherical-harmonic (SH) color coefficients. A renderer R produces $R(G, v) \in [0,1]^3$ for camera v. Given a camera distribution $\mathcal{V}$ and an RGB distance d, define the render distance $d_{\mathcal{V}}(G, G') = \mathbb{E}_{v\sim\mathcal{V}}[d(R(G,v), R(G',v))]$ and, for a threshold ϵ, the equivalence $G \sim_{\epsilon,\mathcal{V}} G' \iff d_{\mathcal{V}}(G, G') \le \epsilon$. An ideal conditional editor E_z, parameterized by instruction z, should descend to the quotient: $G \sim G' \Rightarrow d_{\mathcal{V}}(E_z(G), E_z(G'))$ small. The amplification ratio $\kappa = d_{\mathcal{V}}(E_z(G), E_z(G'))/\max(d_{\mathcal{V}}(G, G'), \delta)$ diverges as the denominator approaches numerical precision, so we report it only as a diagnostic; headline numbers use post-edit RMSE and PSNR of each *edit* against the author edit $A = E_z(G)$.

3.2 K-way optical-thickness refinement

For a Gaussian with opacity α, define optical thickness $\tau = -\log(1-\alpha)$. A K-way *co-located split* replaces the primitive by K children that copy position, covariance, and SH exactly, and set each child's opacity to $\alpha_{\text{child}} = 1 - \exp(-\tau/K) = 1 - (1-\alpha)^{1/K}$, so that central transmittance is conserved, $\prod_{j=1}^{K}(1-\alpha_{\text{child}}) = \exp(-\sum_j \tau_j) = 1-\alpha$. Standard 3DGS evaluation also multiplies by the projected kernel, so the identity is exact at the level of opacity and peak transmittance rather than per-pixel; we therefore enforce empirical equivalence with an **input gate**: all per-view training-camera PSNRs between the refined and base renders must exceed 60 dB, and the checkpoint, lineage mapping, prompt, edit seed, and 2D edit cache are frozen *before* any split arm runs. Eligible parents are drawn from a sigmoid-opacity band $[0.2256, 0.2542]$ at a 0.5% fraction under frozen seeds; matching Adam moments are duplicated along the same row map, and the lineage mapping $\ell(i)$ is stored. An approximate variant adds zero-mean local position jitter of scale $\epsilon = 0.2\times$ the

geometric-mean Gaussian scale to split children only (mean world-space jitter 5.7×10^{-4}, max 4.2×10^{-2}).

3.3 PAIRED EDITING PROTOCOL

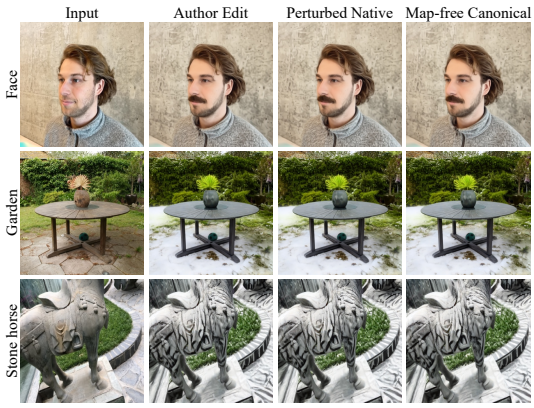

Figure 2: Paired editing setup: input, author edit, perturbed-native edit, and map-free canonical edit.

We test GaussCtrl (Wu et al., 2024) (audited commit, SD 1.5 with depth ControlNet, guidance scale 5, 20 diffusion steps, 500 optimization steps, whole-scene editing) as the editor, with Nerfstudio/gsplat (Tancik et al., 2023; Ye et al., 2025) for reconstruction. Scenes are Dinosaur (105 training / 104 novel midpoint views) and Bear (87/86), with two material prompts each (*bronze*, *white marble*); three further scenes broaden coverage (Figure 2): Face (59/58; *man with a moustache*, a local semantic appearance edit), Garden (167/166; *garden covered in snow*, a global environment transformation), and Stone horse (73/72; *stone horse to zebra*, an object-level semantic conversion), all at $K{=}4$; both the exact and the metadata-free approximate arms pass worst-view input gates of 60.5–70.5 dB (canonical recovery above 161 dB exact, 66.7–73.2 dB after map-free collapse), and all additional-scene trainings use an audited optimizer entry point (Appendix I). Novel cameras are midpoints of adjacent training poses (chordal rotation midpoint projected to SO(3) via SVD), never seen by the editor. The base checkpoint establishes the author edit A once, writing 2D edit targets to disk; every split arm then *reads* the same frozen targets under the same machine seed, so output differences are attributable to the input representation alone. Controls include the **native arm** (edit the split checkpoint directly), **group culling** (one group-level opacity decision per lineage), **quotient gradient** (sum row gradients within a group and project back to shared state), the **canonical arms** (Sections 4.1–4.3), and the **repeat floor** (edit the identical base checkpoint twice).

3.4 METRICS

The primary dependent variable is RGB RMSE between the representation-perturbed edit and the author edit, over all training or all novel cameras. Companions: per-view PSNR mean and worst view; LPIPS (Zhang et al., 2018) to the author edit; *edit magnitude* $\mathrm{LPIPS}(\text{input}, \text{candidate})/\mathrm{LPIPS}(\text{input}, \text{author})$, which must stay near 100% so that stability is never achieved by muting the edit; CLIP (Radford et al., 2021) prompt distance; parent survival mismatch, localizing the culling failure path; and the identical-input repeat floor, separating representation effects from editor run-to-run noise. Statistics are computed at the *camera* level with view-level paired bootstrap CIs and $P(\Delta > 0)$ (DiCiccio & Efron, 1996); pixels are never treated as independent samples (Appendix B).

4 CANONICAL GAUGE FIXING

4.1 EXACT CANONICAL GAUGE

Given lineage labels $\ell(i)$, a *sibling group* $S_g = \{i : \ell(i) = g\}$ consists of children produced from one parent. For exactly co-located splits, siblings share all non-opacity attributes, so the group is characterized by its total optical thickness:

$$\tau_g = \sum_{i \in S_g} \tau_i, \qquad \alpha_g = 1 - \exp(-\tau_g), \tag{1}$$

and the *canonical representative* keeps any sibling's μ, Σ, f, sets opacity to α_g, and folds every per-Gaussian tensor, including Adam moments and any row-indexed statistics, through the same row

map. Since the split created siblings by duplicating non-opacity attributes and dividing τ, the collapse recovers the parent parameters up to sigmoid/log/expm1 round-trip error:

$$C\big(T_K(G)\big) = G \quad \text{(parameter-level identity).} \tag{2}$$

Empirically the canonical checkpoint renders at 163.9–164.8 dB PSNR against the original for $K \in \{2, 4\}$ and 136.3 dB at $K{=}8$ (Bear K4: 162.6 dB). Operationally: if C is constant on the nuisance orbit, $E_z \circ C$ is exactly invariant on that orbit, and residual edit differences come only from editor run noise and floating-point paths.

4.2 Tau-barycentric gauge for approximate lineages

Real pipelines rarely duplicate primitives bitwise. For *near-coincident* sibling groups we use weights $w_i = \tau_i/\tau_g$ and set $\mu_g = \sum_i w_i \mu_i$, $s_g = \sum_i w_i s_i$ (log-scale), per-coefficient $f_g = \sum_i w_i f_i$, and $\alpha_g = 1 - \exp(-\tau_g)$; quaternions are sign-aligned to one hemisphere before the weighted average. The weighted mean minimizes $\sum_i \tau_i \|x - x_i\|^2$, so if the split offsets have zero τ-weighted first moment and the renderer is twice differentiable in position, the first-order Taylor term cancels and the local merge error is governed by the second moment. The argument is local, not a global theorem about occlusion reordering; the perturbed input still renders at 119.4 dB after barycentric collapse (103.4 dB worst view).

4.3 Map-free lineage discovery

The barycentric gauge above assumed a stored lineage mapping. When none exists, the common case for assets from unknown pipelines, we exploit a structural signature of co-located refinement: within a lineage, *position* may be jittered, but covariance, opacity, and SH coefficients are copied. Define

$$\mathrm{sig}(i) = \mathrm{bytes}(\mathrm{scale}_i, \mathrm{quat}_i, \mathrm{opacity}_i, \mathrm{SH}_i), \tag{3}$$

group rows by exact signature equality, and accept only groups whose cardinality equals a known refinement multiplicity K (or any element of the suspected multiplicity set). Position is deliberately excluded from the key, making discovery robust to the approximate regime; multiplicity filtering prevents accidental merges of naturally identical primitives. The grouping is a linear scan, requires no training and no threshold tuning beyond K.

4.4 Why partial repairs fall short

Two natural fixes address only part of the state space (Fig. 1, bottom). *Group culling* replaces per-row opacity-threshold tests by one decision per lineage using $\alpha_g = 1 - \exp(-\tau_g)$, preventing the pathology where every child slips below the cull threshold while the group is above it; but it leaves per-row Adam states, nonlinear gradients, and row-count-dependent logic untouched. *Quotient gradients* sum row gradients within a group (the chain rule for a shared variable) and project attributes back to the group mean after each update, with opacity averaged in τ-coordinates and total τ conserved; this constrains continuous gauge drift, yet the editor still executes its discrete logic on a higher-dimensional row space. GaugeFix3D instead removes the redundant dimensions *before* any editing nonlinearity, eliminating culling ambiguity, duplicated optimizer dynamics, stochastic selection, and row-dependent statistics in one stroke (Section 5.3 quantifies the difference).

4.5 Wrapper properties

Algorithmically, GaugeFix3D is a pre-edit pass: read checkpoint; group rows (sidecar or signature discovery); convert opacity to τ; collapse each group (exact copy or τ-barycenter; fold optimizer state with the same row map); write the collapsed checkpoint; call the original editor unchanged. Cost is $O(N)$ with a linear scan, about 8 s on CPU for 706k primitives, and peak edit memory stays at the editor's own $\sim$14 GiB on a single 24 GB GPU; the wrapper is trivially composable with any row-based editor.

5 EXPERIMENTS

All experiments follow the paired protocol of Section 3.3: same prompt, same frozen 2D edit targets, same machine seed, same schedule across arms; only the checkpoint entering the editor differs. We report view-level paired bootstrap CIs throughout.

5.1 MAIN RESULT: RENDER-EQUIVALENT INPUTS, DIVERGENT EDITS

Table 1 covers the twelve exact-refinement cells (two object scans $\times$ two prompts $\times$ $K \in \{4, 8\}$ $\times$ train/novel) and six audited re-run cells on the three additional scenes (exact and metadata-free approximate arms, novel interpolated cameras). Native edits deviate from the author edit by RMSE 0.0023–0.0080; canonicalized inputs recover 78.35–88.04% of that deviation, with paired-view 95% CIs strictly above zero in every cell and $P(\Delta > 0) = 1.0$. In the audited re-run the exact arm reaches 88.0% on Garden and the metadata-free approximate arm 75.6%. In the historical hero cell (Dinosaur, bronze, K=8, novel), edit quality improves from 43.41 to 58.61 dB PSNR against a 60.0 dB repeat floor: canonicalization removes all but 17.2% of the gap to run-to-run noise (2.2% on training views, Table 2). The improvement is uniform across scenes, materials, edit types, and held-out views.

Table 1: Exact-refinement main results. RMSE between each arm's edit and the author edit; reduction with view-level paired bootstrap 95% CI (100k replicates); ΔPSNR is the edit-quality gain of the canonical arm. Top: twelve historical cells. Bottom: local appearance, environment, and object-level semantic edits with an audited optimizer (novel interpolated cameras; mf = map-free approximate arm). All cells are pre-defined and all CIs exclude zero.

Scene	Prompt	K	Split	Native RMSE ↓	Canon. RMSE ↓	Reduction % [95% CI]	ΔPSNR ↑
Dinosaur	bronze	4	train	0.00597	0.00127	78.70 [76.61, 80.71]	+13.43
Dinosaur	bronze	4	novel	0.00665	0.00144	78.35 [74.65, 81.72]	+13.29
Dinosaur	bronze	8	train	0.00601	0.00114	81.07 [79.37, 82.59]	+14.46
Dinosaur	bronze	8	novel	0.00675	0.00117	82.63 [81.04, 83.98]	+15.20
Dinosaur	marble	4	train	0.00630	0.00125	80.19 [78.76, 81.51]	+14.06
Dinosaur	marble	4	novel	0.00694	0.00131	81.15 [79.68, 82.50]	+14.49
Dinosaur	marble	8	train	0.00629	0.00121	80.79 [79.54, 81.95]	+14.33
Dinosaur	marble	8	novel	0.00693	0.00124	82.05 [80.99, 83.12]	+14.92
Bear	bronze	4	train	0.00537	0.00095	82.36 [80.78, 83.89]	+15.07
Bear	bronze	4	novel	0.00554	0.00109	80.25 [74.83, 84.23]	+14.09
Bear	marble	4	train	0.00445	0.00085	80.89 [79.29, 82.45]	+14.37
Bear	marble	4	novel	0.00459	0.00085	81.51 [80.04, 82.95]	+14.66
Semantic edits with audited optimization (K=4, novel cameras): exact / map-free approximate							
Face	moustache	4	novel	0.00236	0.00037	84.49 [82.95, 86.04]	–
Face	moustache (mf)	4	novel	0.00236	0.00070	70.45 [68.34, 72.59]	–
Garden	snow	4	novel	0.00507	0.00061	88.04 [85.45, 90.57]	–
Garden	snow (mf)	4	novel	0.00508	0.00124	75.56 [73.17, 77.60]	–
Stone horse	zebra	4	novel	0.00793	0.00152	80.89 [78.83, 82.75]	–
Stone horse	zebra (mf)	4	novel	0.00797	0.00220	72.37 [69.75, 74.81]	–

5.2 SCOPE EXPANSION: SCENES AND EDIT TYPES

The original stress scenes are symmetric object scans with material-transfer prompts; to test whether the phenomenon and the cure generalize, we add maximally diverse scenes at frozen K=4: Face (a portrait capture; *man with a moustache*: a local semantic appearance edit), Garden (an outdoor vegetation scene; *garden covered in snow*: a global environment transformation), and Stone horse (a large monument; *stone horse to zebra*: an object-level semantic conversion). All additional-scene trainings use an audited entry point that rebuilds the editor optimizer after checkpoint loading and verifies parameter binding at every step. Exact refinements yield reductions of $84.5/88.0/80.9\%$ and, with no lineage metadata, map-free approximate refinements yield $70.5/75.6/72.4\%$ (Table 1, bottom panel), under worst-view input gates of 60.5–70.5 dB across both arms. Non-degeneracy holds on every arm: canonical LPIPS falls to 2.5×10^{-5} (Face) and 2.3×10^{-5} (Garden), edit magnitudes stay $\geq 99.99\%$, and CLIP prompt distances move within 3×10^{-4} of native (Table 3). At a fixed, non-cherry-picked interpolated view per scene, the edits show the same signature as in the original scenes (Fig. 3; original scenes in Appendix D).

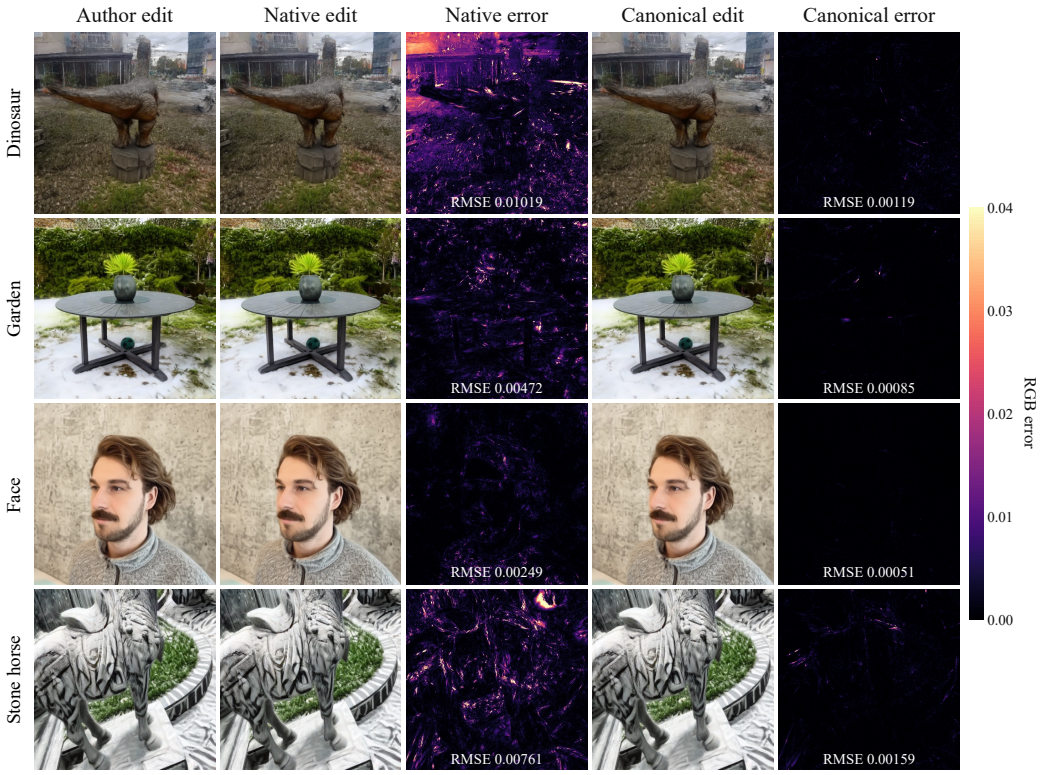

Figure 3: Qualitative comparison and error maps. Rows show Dinosaur (bronze), Garden (snow), Face (moustache), and Stone horse (zebra). Columns show the author edit, the native edit, the canonicalized edit, and their respective error maps relative to the author edit. Canonicalization reduces editing discrepancies across all examples. Error maps show per-pixel RGB RMSE on a shared color scale from 0 to 0.04.

Table 2: K-axis stress and mechanism ablation (Dinosaur, bronze). RMSE ↓ of the edit vs. the author edit. Partial repairs fix one failure path each; canonical gauge fixing removes the redundancy itself. The repeat floor is the editor's deviation from itself under identical input (0.00111 train, 0.00100 novel); canonical reductions reach $65.3/78.7/81.1\%$ (train) and $59.3/78.4/82.6\%$ (novel) at $K{=}2/4/8$.

	Training views				Novel views			
K	Native	Cull	Quot.	Canon.	Native	Cull	Quot.	Canon.
2	0.00390	0.00316	0.00327	**0.00135**	0.00418	0.00342	0.00349	**0.00170**
4	0.00597	0.00407	0.00417	**0.00127**	0.00665	0.00459	0.00476	**0.00144**
8	0.00601	0.00398	0.00389	**0.00114**	0.00675	0.00456	0.00449	**0.00117**

5.3 Mechanism: many nonlinearities, one cure

Table 2 dissects the Dinosaur bronze stress curve: native edit error grows monotonically with K on both camera splits, confirming that refinement multiplicity, not any single fixed perturbation,drives the effect. Group culling and quotient gradients each recover only 34–35% at $K{=}8$ (train), because each repairs one path while the editor still executes discrete selection logic over a redundant row space. Canonicalization removes 81.1% in the same cell. The instrument most specific to the culling path: parent survival mismatch, the fraction of split parents whose survival decision disagrees between arms, falls from 0.756 to 4.5×10^{-4} (99.94%) under canonicalization. Per-lineage optical thickness is conserved to 2.1×10^{-7} absolute error, so the collapse itself introduces no visible drift.

5.4 Stability is not achieved by muting the edit

A trivially "stable" editor would ignore the instruction. Table 3 rules this out: LPIPS distance to the author edit *drops* by 93.2–98.7% under canonicalization across all measured scene–prompt

Table 3: Perceptual fidelity and edit strength (novel views). Canonical edits are both closer to the author edit (LPIPS) and equally strong (edit magnitude = LPIPS to input relative to the author's).

Setting	Native LPIPS ↓	Canon. LPIPS ↓	LPIPS red.	Edit magnitude ↑
Dinosaur bronze K8	0.00413	**0.00017**	95.99%	100.004%
Bear bronze K4	0.00179	**0.00006**	96.68%	99.997%
Bear marble K4	0.00143	**0.00005**	96.37%	99.993%
Bear perturbed K4	0.00176	**0.00012**	93.24%	99.985%
Face exact	0.00067	**0.00002**	96.33%	99.998%
Garden exact	0.00179	**0.00002**	98.74%	99.996%

combinations, the canonical edits are closer to the intended edit than the native edits are, while edit magnitude stays at 99.99–100.00% of the author's; CLIP prompt distances agree to three decimals in all measured arms; Stone horse is verified via author distance and edit magnitude. The gains therefore come from following the author edit more faithfully, not from editing less.

5.5 Beyond bitwise duplicates: approximate and map-free refinements

Exact co-location is an idealization of what densification and compression actually do. We therefore jitter split children by zero-mean offsets of $0.2\times$ the geometric-mean scale (input gate still passes: 69.2 dB global, 67.4 dB worst view) and run the τ-barycentric collapse. The collapsed input renders at 119.4 dB against the original, confirming the first-order-cancellation analysis, and the resulting edits improve by 71.6% (train) and 72.4% (novel) RMSE, $+10.9/+11.2$ dB (per-arm curves in Appendix D). Dropping the lineage sidecar entirely, the map-free signature discovery of Section 4.3 recovers *every* group with *zero* false merges in both stress cases ($705{,}834 \to 695{,}403$ Bear, $911{,}758 \to 880{,}923$ Dino rows collapsed; full tables in Appendix E) and yields reductions of $73.8/73.9\%$ (Bear K4, train/novel) and $68.7/65.2\%$ (Dinosaur K8, train/novel), with edit magnitude 99.995% (Bear) and 99.991% (Dino) and LPIPS-to-author reductions of 93.6% and 89.7%. Notably, the map-free arm *outperforms* the sidecar-informed arm (train $+7.9\%$ $[3.8, 12.3]$, novel $+5.4\%$ $[2.0, 8.8]$), signature grouping changes row order and the finite-precision trajectory, so metadata-free operation costs nothing here. The tolerant matcher used in the audited re-run extends discovery beyond byte equality: k-NN candidates (16 neighbors, radius $0.25\times$ the local scale) with per-attribute tolerance 10^{-3}, quaternion sign alignment, and whole-group checks that revert failed candidates to singletons. Across 27 controlled combinations it recovers $67{,}744/67{,}752$ groups; under mixed multiplicities $K \in \{2, 4, 8\}$, $22{,}578/22{,}584$; and on unaugmented base assets it proposes *zero* merges. Byte-exact signatures are brittle (10^{-6} log-scale noise defeats them), tolerance is what makes recovery robust in real pipelines.

5.6 Robustness across refinement seeds

The split itself is a sampling procedure, so we repeat Bear under three frozen split seeds and both prompts (per-seed table and curves in Appendix F). Every one of the twelve prompt×split×seed cells is positive ($3/3$ per row), with means of 80.2–82.3% and ranges within 0.6 points. Because canonicalization is deterministic, the canonical arm's RMSE is *identical* across seeds; only the native arm varies. Worst-view input gates across seeds are 66.9–67.6 dB, above the frozen 60 dB threshold.

5.7 The principle generalizes across editor families

If instability comes from nonlinearities acting on redundant mixture coordinates, editors that first map the Gaussian measure onto a fixed coordinate system should be more stable – and a *wrong* conservation coordinate should hurt. We test this with a physical-token style editor that deposits the Gaussian empirical measure into a CIC hierarchy over fixed physical coordinates before applying its style operator. Across a disjoint set of scenes and frozen split seeds it holds a stability advantage of 63.1–74.1 points per scene (macro 69.5) over the strongest uniform residual-shrink control satisfying the same clean-view and edit-retention constraints (scene-level bootstrap CI $[65.5, 73.0]$; $15/15$ seeds positive; relaxing edit retention from 95% to 80% still leaves 59.1 points; Bus scene: 20.3 s, 1.70 GB peak RSS for the full graph+edit pass) , see Table 5 and Fig. 4. The conservation-coordinate ablation completes the argument: for an α-additive split (Family), reasoning in α-mass rather than τ reduces

Table 5: Cross-editor evidence. Stability advantage (%) of the physical-token editor over the strongest perturbation control, per scene and macro, across fifteen frozen seeds.

	Bus	Broch.	Fam.	M60	Tab.	Mac.
Adv. (pp)	65.7	72.6	63.1	71.9	74.1	69.5

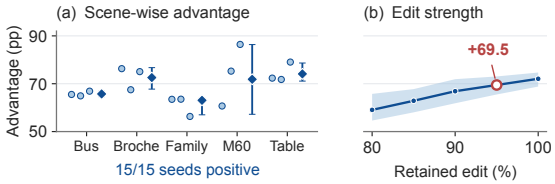

Figure 4: Cross-editor stability–strength trade-off. The physical-token editor consistently outperforms the strongest admissible uniform-shrink control. (a) At 95% retained edit, circles show seed results; diamonds and bars show scene estimates and uncertainty. (b) Scene-macro advantage across retained-edit thresholds.

token-mass and mass-color drift by $10.6\times$/$7.7\times$; for a τ-additive split (M60), reasoning in τ rather than α reduces the same drifts by $10.1\times$/$9.3\times$. The coordinate in which the quotient is defined must match the conservation law of the nuisance transformation, exactly what our optical-thickness gauge instantiates. *Direct transfer through the same interface.* Through the identical canonical-checkpoint interface (FastSplatStyler, fixed commit, weights, style, and seed), exact $K{=}4$ refinement of the canonicalized checkpoint reproduces the author edit to RMSE 9.4×10^{-8} (Face) and 9.9×10^{-8} (Stone horse), checkpoint round-trip precision, versus 0.142 and 0.151 natively.

Row Order as an Additional Nuisance. With consistent coordinate ordering across all arms, GaugeFix3D reduces novel-view RMSE by 56.8% over native editing (0.1417 to 0.0612; Table 4). Permuting the unsplit checkpoint alone induces 0.1480 training-view RMSE, versus 0.0048 from matched attribute noise, identifying row order as the dominant residual nuisance. These controls support our principle of removing representation-dependent freedom before editing. Each ordering protocol uses its own author reference (Appendix J).

Table 4: Row-order Sensitivity. Coordinate ordering enables a 56.8% error reduction under approximate refinement.

Training-view control		RMSE ↓
Attribute noise only		0.0048
Row permutation only		0.1480
Novel-view protocol	**Native**	**Canonical**
Original order	0.1430	0.1475
Coordinate order	0.1417	**0.0612**

5.8 Optimizer audit and recovery under real continued training

Representation-level claims must survive the training loop. All additional-scene trainings therefore use an audited entry point that rebuilds the optimizer after checkpoint loading and verifies optimizer–parameter binding at every step; all fifteen runs record first-step updates of 10^5–10^7 elements in every parameter group (Appendix I). After 50 continued training steps on Face (base and $K{=}4$-refined checkpoints advanced in lockstep, post-densification culling disabled), the tolerant matcher recovers $941/968$ lineages (97.2%) at tolerance 10^{-3} with zero mixed groups ($81/968$ at 10^{-4}; $965/968$ at 10^{-2}), while the byte-exact signature recovers only $7/968$. Thus, optimization drifts attributes within the tolerant matcher's basin. The map-free-collapsed checkpoint also renders closer to the synchronized continued base (68.97 dB global, 65.10 dB worst view) than the native continued refined checkpoint ($67.82/64.55$ dB). We interpret this as recovery under training drift, not a dynamic-closure theorem. Finally, natural exact classes also occur in *unaugmented* Splatfacto/GaussCtrl checkpoints: Face $7{\to}6$, Garden 24 (21 pairs and 3 triples, with the 21 pairs collapsing to a pixel-level 149.8 dB after editing), and Stone horse $27{\to}25$. These provide existence evidence that real optimization leaves redundant coordinates that our gauge removes.

6 Conclusion

We showed that render-equivalent Gaussian representations can produce inconsistent edits. GaugeFix3D addresses this sensitivity through training-free canonicalization before the unchanged editor. Across variable scenes, it substantially reduces editing discrepancies while preserving edit strength, including under approximate and metadata-free refinements. These results establish representation stability as an important evaluation criterion and support canonicalization as a practical interface for reliable 3D Gaussian editing.

Reproducibility statement

All quantitative claims are generated by the scripts, checkpoints, per-view reports, and frozen seeds in our supplementary package; the environment, commands, and per-arm configurations are listed in Appendix A.

AI use statement

Large language models were used for editorial assistance only. All experimental design, execution, analysis, and results are the authors' own.

References

Guikun Chen and Wenguan Wang. A survey on 3d gaussian splatting. *ACM Computing Surveys*, 58 (12):1–39, 2026.

Taco S Cohen and Max Welling. Transformation properties of learned visual representations. *arXiv preprint arXiv:1412.7659*, 2014.

Thomas J DiCiccio and Bradley Efron. Bootstrap confidence intervals. *Statistical science*, 11(3): 189–228, 1996.

Tom Duff. Deep compositing using lie algebras. *ACM Transactions on Graphics (TOG)*, 36(3):1–12, 2017.

Guangchi Fang and Bing Wang. Mini-splatting: Representing scenes with a constrained number of gaussians. In *European conference on computer vision*, pp. 165–181. Springer, 2024.

Ben Fei, Jingyi Xu, Rui Zhang, Qingyuan Zhou, Weidong Yang, and Ying He. 3D Gaussian Splatting as a new era: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 31(8): 4429–4449, 2025. doi: 10.1109/TVCG.2024.3397828. URL https://doi.org/10.1109/TVCG.2024.3397828.

Qi-Yuan Feng, Geng-Chen Cao, Hao-Xiang Chen, Qun-Ce Xu, Tai-Jiang Mu, Ralph Martin, and Shi-Min Hu. Evsplitting: An efficient and visually consistent splitting algorithm for 3d gaussian splatting. In *SIGGRAPH Asia 2024 conference papers*, pp. 1–11, 2024.

Florian Hahlbohm, Linus Franke, Martin Eisemann, and Marcus Magnor. Faster-gs: Analyzing and improving gaussian splatting optimization. *arXiv preprint arXiv:2602.09999*, 2026.

Haoliang Han, Ziyuan Luo, Jun Qi, Anderson Rocha, and Renjie Wan. Gs-checker: Tampering localization for 3d gaussian splatting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 4556–4564, 2026.

Ayaan Haque, Matthew Tancik, Alexei A Efros, Aleksander Holynski, and Angjoo Kanazawa. Instruct-nerf2nerf: Editing 3d scenes with instructions. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 19683–19693. IEEE, 2023.

Lukas Höllein, Aljaž Božič, Michael Zollhöfer, and Matthias Nießner. 3dgs-lm: Faster gaussian-splatting optimization with levenberg-marquardt. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 26740–26750. IEEE, 2025.

Ziming Hong, Tianyu Huang, Runnan Chen, Shanshan Ye, Mingming Gong, Bo Han, and Tongliang Liu. AdLift: Lifting adversarial perturbations to safeguard 3D Gaussian Splatting assets against instruction-driven editing. arXiv preprint arXiv:2512.07247, 2025. URL https://arxiv.org/abs/2512.07247.

Utae Jeong, Jaewan Choi, Junseok Lee, Jongheon Jeong, Sang Ho Yoon, ByoungSoo Koh, and Sangpil Kim. Guardmarkgs: Unified ownership tracing and edit deterrence for 3d gaussian splatting. *arXiv preprint arXiv:2605.12919*, 2026.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian Splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, 2023. doi: 10.1145/3592433. URL https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/.

Bernhard Kerbl, Andréas Meuleman, Georgios Kopanas, Michael Wimmer, Alexandre Lanvin, and George Drettakis. A hierarchical 3d gaussian representation for real-time rendering of very large datasets. *ACM Transactions on Graphics*, 43(4):1–15, 2024. doi: 10.1145/3658160.

Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Weiwei Sun, Yang-Che Tseng, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, and Kwang Moo Yi. 3d gaussian splatting as markov chain monte carlo. *Advances in Neural Information Processing Systems*, 37:80965–80986, 2024.

Hyun Kwon and Jang-Woon Baek. Empirical analysis of adversarial robustness in 3D Gaussian Splatting under multi-view inconsistency attacks. *Scientific Reports*, 16:25039, 2026. doi: 10.1038/s41598-026-53843-3. URL https://www.nature.com/articles/s41598-026-53843-3.

Dong In Lee, Hyeongcheol Park, Jiyoung Seo, Eunbyung Park, Hyunje Park, Ha Dam Baek, Sangheon Shin, Sangmin Kim, and Sangpil Kim. Editsplat: Multi-view fusion and attention-guided optimization for view-consistent 3d scene editing with 3d gaussian splatting. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11135–11145. IEEE, 2025.

Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, and Eunbyung Park. Compact 3d gaussian representation for radiance field. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21719–21728. IEEE, 2024.

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 *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20654–20664. IEEE, 2024.

Andreas Meuleman, Ishaan Shah, Alexandre Lanvin, Bernhard Kerbl, and George Drettakis. On-the-fly reconstruction for large-scale novel view synthesis from unposed images. *ACM Transactions on Graphics (TOG)*, 44(4):1–14, 2025.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Samuel Rota Bulò, Lorenzo Porzi, and Peter Kontschieder. Revising densification in gaussian splatting. In *European Conference on Computer Vision*, pp. 347–362. Springer, 2024.

Etai Sella, Gal Fiebelman, Peter Hedman, and Hadar Averbuch-Elor. Vox-e: Text-guided voxel editing of 3d objects. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 430–440. IEEE, 2023.

Yuang Shi, Géraldine Morin, Simone Gasparini, and Wei Tsang Ooi. Lapisgs: Layered progressive 3d gaussian splatting for adaptive streaming. In *2025 International Conference on 3D Vision (3DV)*, pp. 991–1000. IEEE, 2025.

Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, et al. Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 conference proceedings*, pp. 1–12, 2023.

Junjie Wang, Jiemin Fang, Xiaopeng Zhang, Lingxi Xie, and Qi Tian. Gaussianeditor: Editing 3d gaussians delicately with text instructions. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20902–20911. IEEE, 2024.

Luchao Wang, Kaimin Liao, Qian Ren, Hua Wang, Zhi Chen, and Yaohua Tang. Reorggs: Equivalent distribution reorganization for 3d gaussian splatting. *arXiv preprint arXiv:2605.08739*, 2026.

Peihao Wang, Yuehao Wang, Dilin Wang, Sreyas Mohan, Zhiwen Fan, Lemeng Wu, Ruisi Cai, Yu-Ying Yeh, Zhangyang Wang, Qiang Liu, et al. Steepest descent density control for compact 3d gaussian splatting. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26663–26672. IEEE, 2025.

Jing Wu, Jia-Wang Bian, Xinghui Li, Guangrun Wang, Ian Reid, Philip Torr, and Victor Adrian Prisacariu. Gaussctrl: Multi-view consistent text-driven 3d gaussian splatting editing. In *European conference on computer vision*, pp. 55–71. Springer, 2024.

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.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF conference on computer vision and pattern recognition*, pp. 586–595. IEEE, 2018.

Qinzheng Zhou, Zaychik Liu, Lijing Lu, and Zhihang Li. Confidence-guided multi-scale aggregation for sparse-view high-resolution 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19054–19064, 2026.

Chen-Yang Zhu, Xin-Yao Liu, Kai Xu, and Ren-Jiao Yi. A survey on 3d editing based on nerf and 3dgs. *Frontiers of Computer Science*, 20(4):2004701, 2026.

Hao Zhu, Shaowen Xie, Zhen Liu, Fengyi Liu, Qi Zhang, You Zhou, Yi Lin, Zhan Ma, and Xun Cao. Disorder-invariant implicit neural representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8):5463–5478, 2024.

Jingyu Zhuang, Di Kang, Yan-Pei Cao, Guanbin Li, Liang Lin, and Ying Shan. Tip-editor: An accurate 3d editor following both text-prompts and image-prompts. *ACM Transactions on Graphics (ToG)*, 43(4):1–12, 2024.

A Reproduction details

Environment. Ubuntu 20.04, Python 3.8, CUDA 11.8, PyTorch 2.1.2+cu118, Nerfstudio 1.0.0, gsplat 0.1.3, diffusers 0.26.0; a single RTX 4090D (24 GB). The editor is GaussCtrl at audited commit `ff01a068b2caa37b28881632e5744f4bc57cea97` (SD 1.5 checkpoint, depth ControlNet, guidance scale 5, 20 diffusion steps, 500 Gaussian optimization steps, `ref_view_num=2`, `chunk_size=1`, whole-scene edit, machine seed 20260914).

Pipeline. (i) Train the base checkpoint (Splatfacto 30k, $\sim$10 min for Bear). (ii) Apply the K-way optical-thickness split (`nerfstudio_checkpoint_ksplit.py`, fraction 0.005, parents from the sigmoid-opacity band $[0.2256, 0.2542]$, frozen seed). (iii) Canonicalize (exact, barycentric, or signature-inferred mapping). (iv) Render all training-camera pairs and assert every per-view PSNR $\geq$ 60 dB (`render_checkpoint_equivalence.py`); only gated checkpoints proceed. (v) Edit the author arm once with the 2D-target cache in write mode; run all split arms in read mode under identical prompts, seeds, and schedules. (vi) Evaluate train/novel pairs (`render_checkpoint_interpolated.py` for midpoint cameras), perceptual audits (`evaluate_gaussctrl_quality.py`: LPIPS, edit magnitude, CLIP), and paired-view bootstrap CIs (`bootstrap_paired_views.py`, 100k replicates). (vii) Additional-scene trainings use an audited entry point that rebuilds the editor optimizer after checkpoint loading, restores its state, and asserts optimizer–parameter binding at every step.

Runtime. Exact/barycentric canonicalization of 706k primitives takes $\sim$8 s on CPU (linear scan); the full map-free Bear K4 path (discovery + collapse) takes 22 s. Editing peaks at $\sim$14 GiB, unchanged by the wrapper. The physical-token editor's full graph+edit path takes 20.26 s and 1.70 GB peak RSS on the Bus scene (author path 13.80 s, 1.61 GB). A cost summary appears in Appendix K.

Table 6: Frozen constants across every experiment in the paper. Each row is a value that is never tuned after the first run touching it.

Constant	Value
GaussCtrl commit	`ff01a068b2caa37b28881632e5744f4bc57cea97`
Diffusion schedule	SD 1.5, depth ControlNet, 20 steps, guidance 5
Gaussian refinement	500 steps, `ref_view_num=2`, `chunk_size=1`
Split fraction / parent band	0.005 / sigmoid opacity $[0.2256, 0.2542]$
Jitter offset (approximate arms)	$0.2\times$ geometric-mean scale, zero mean
Gate noise (audited re-run)	position $\sigma = 2 \times 10^{-4}$ min-scale; attributes 10^{-6}; seed 20260918
Paired bootstrap	100,000 replicates; seeds 20260913 (historical), 20260918 (audited), 20260912 (cross-editor)
Machine / edit seed	20260914
FastSplatStyler fixture	fixed commit, weights, style, seed; center-graph configuration; 120k token cap; six-point cubature; raw-opacity mass
Matcher	k-NN 16; radius $0.25\times$ local scale; attribute tolerance 10^{-3}; quaternion sign alignment

Table 7: Asset inventory. Base-checkpoint primitive counts as measured by the discovery scan; prompts and refinement multiplicities as used in Table 1.

Scene	Base primitives	Prompts	K
Dinosaur	880,923	bronze / marble	$\{4, 8\}$
Bear	695,403	bronze / marble	4
Face	193,589	man with a moustache	4
Garden	729,355	garden covered in snow	4
Stone horse	582,630	stone horse to zebra	4

B STATISTICAL PROTOCOL

Every CI in the main tables is a view-level paired bootstrap (100,000 replicates): camera pairs are resampled with replacement, each replicate aggregates native and canonical MSE separately, and the reduction statistic is formed per replicate. Every pre-defined cell in Table 1, historical or audited, and both map-free cells have intervals strictly above zero and $P(\Delta > 0) = 1.0$. The cross-editor advantages use the same machinery at the scene level, with an additional hierarchical scene$\times$seed bootstrap reported alongside. Because the study spans five scenes, we do not aggregate scenes into a pseudo-population CI; refinement-seed robustness (Table 11, Appendix F) and camera-sampling uncertainty are complementary axes. All statistics are computed at the *camera* level; pixels are never treated as independent samples.

C COMPLETE QUALITY AND INPUT-GATE REPORTS

Canonicalization moves every arm *toward* the author edit, in both regimes: the exact arm's LPIPS-to-author falls by 96.33% (Face) and 98.74% (Garden), and the metadata-free approximate arm, which receives a perturbed input, still improves by 88.50% and 93.20% respectively, while edit magnitudes stay within 0.12% of unity and CLIP prompt distances move by at most 3×10^{-4}. Stone horse is audited on novel cameras (no native-side perceptual report is shipped for this scene in the audited package): canonical LPIPS-to-author is 1.31×10^{-4} (exact) and 2.69×10^{-4} (approximate), with edit magnitudes 99.9921% and 99.9993%.

D STRESS, MECHANISM, AND QUALITATIVE EVIDENCE

Fig. 5 shows that the deviation grows monotonically with K while the canonical arm tracks the repeat floor at every K; Fig. 6 extends the same picture to jittered lineages and to the metadata-free regime, where input recovery and edit stability degrade gracefully together.

Table 8: Full perceptual audit of the audited re-run (Face and Garden; 8-bit PNG renders, held-out set excludes the three calibration views). Edit magnitude is LPIPS distance to the input, normalized by the author edit's; values ≈ 1 indicate the edit strength is preserved. CLIP prompt distances agree with the native arm to three decimals in every arm.

Scene	Arm	Edit magnitude	LPIPS to author	CLIP prompt distance
Face	exact native	1.001065	0.0006682	0.716261
Face	exact canonical	0.999978	**0.0000245**	0.716239
Face	approximate native	1.001123	0.0006685	0.716279
Face	approximate canonical	0.999798	**0.0000769**	0.716212
Garden	exact native	1.000219	0.0017879	0.675456
Garden	exact canonical	0.999958	**0.0000226**	0.675635
Garden	approximate native	1.000185	0.0017875	0.675501
Garden	approximate canonical	1.000112	**0.0001216**	0.675723

Table 9: Input gates of the audited re-run (dB, rendered against each scene's unaugmented base checkpoint; the gate is worst-view PSNR ≥ 60 dB). The perturbed K4 inputs carry the gate-noise specification of Table 6; the canonical rows are their map-free collapses. All six inputs pass; the tightest margin is Stone horse perturbed at 60.47 dB.

Scene	Input	Cameras	Global PSNR	Worst-view PSNR
Face	perturbed	59	71.55	69.72
Face	canonical	59	82.64	73.17
Garden	perturbed	167	69.92	66.25
Garden	canonical	167	81.54	69.45
Stone horse	perturbed	73	66.94	60.47
Stone horse	canonical	73	77.35	66.67

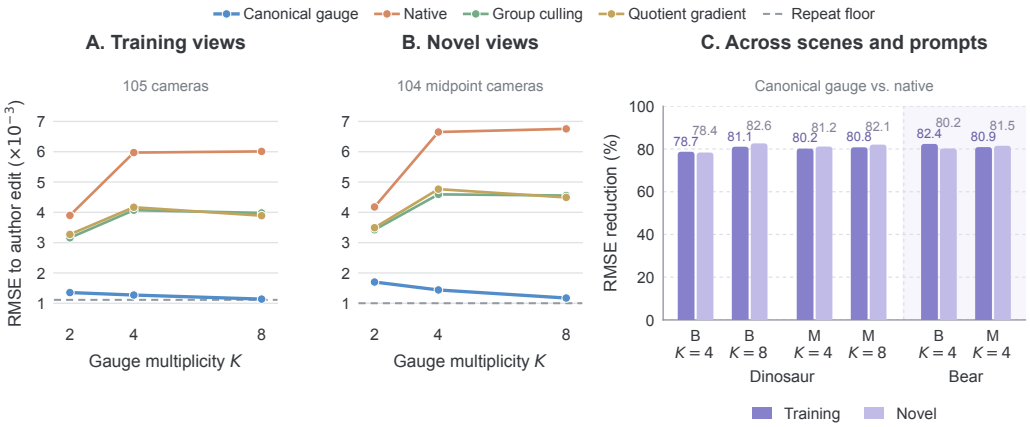

Figure 5: Stress along the refinement axis. (A, B): edit deviation (RMSE) grows with refinement multiplicity K for the native arm and collapses under canonical gauge fixing, approaching the identical-input repeat floor (Dinosaur, bronze). (C): the exact-refinement reductions are consistent across scenes, prompts, and camera splits. In (C), B and M denote the bronze and marble prompts, respectively.

Fig. 7 complements these aggregate results with additional scene–prompt comparisons. Each row shows the author, native, and canonical edits, followed by error maps relative to the author edit; camera identifiers and refinement multiplicities are given in the panels.

E MAP-FREE SIGNATURE DISCOVERY

The signature key uses exact byte equality over scale, quaternion, opacity, and SH coefficients. Position is excluded because the approximate regime deliberately perturbs it. The multiplicity filter rejects signature groups whose cardinality does not equal the known refinement multiplicity (e.g.,

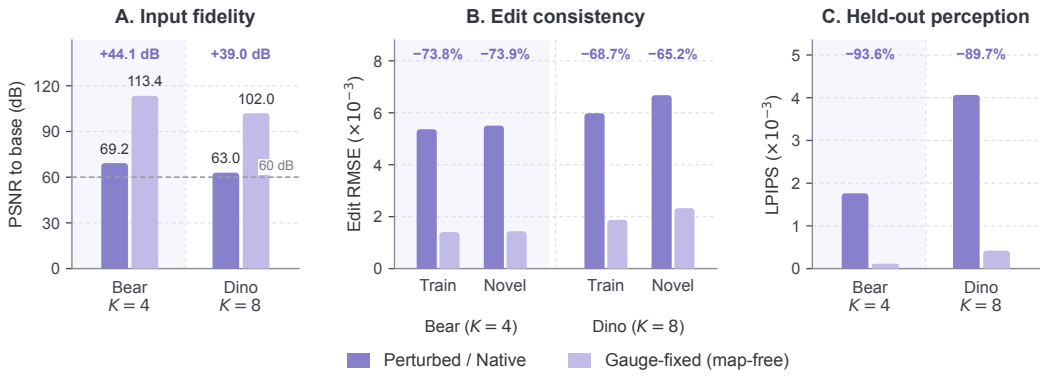

Figure 6: Approximate (jittered) and map-free refinements. (A) Input recovery of the barycentric collapse; (B) edit stability under jittered lineages; (C) fidelity: LPIPS-to-author falls while edit magnitude remains at $\geq 99.99\%$.

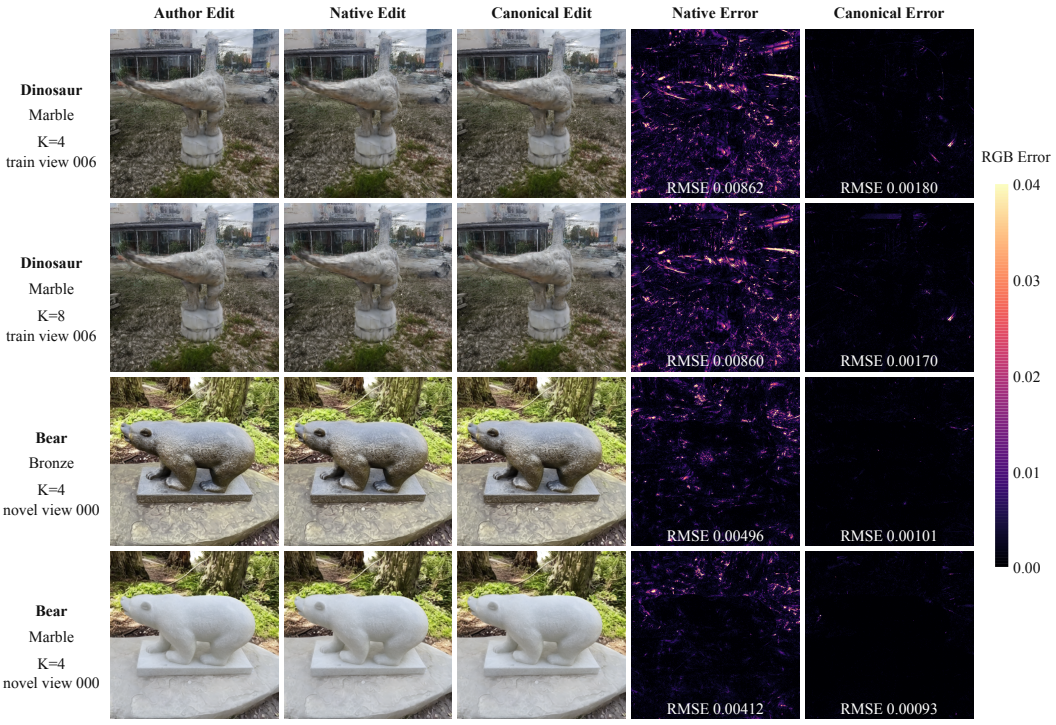

Figure 7: GaussCtrl exact refinements with additional prompts. Dinosaur with the marble prompt at $K = 4$ and $K = 8$ (training view 006), and Bear with the bronze and marble prompts at $K = 4$ (novel view 000). Columns show the author edit, native edit, canonical edit, native error, and canonical error. Error maps use a shared RGB-error scale from 0 to 0.04, with RMSE annotations for each displayed view.

K=4), which is what prevents naturally identical primitives (e.g., repeated background elements) from being merged; in our stress cases the filter admitted exactly the planted groups and nothing else, and zero false merges were observed on both Bear K4 (3,477 groups, 13,908 rows collapsed) and Dinosaur K8 (4,405 groups, 35,240 rows collapsed).

Beyond byte equality: the tolerant matcher. The audited re-run extends discovery to perturbed inputs. Candidate siblings are proposed by spatial k-NN (16 neighbors, search radius $0.25\times$ the local Gaussian scale) and admitted when every per-attribute distance is within tolerance (10^{-3} on scale/log-scale, opacity, and SH bytes; quaternions compared up to sign). Group decisions are made

Table 10: Approximate and map-free results. Top: edit RMSE reduction under τ-barycentric collapse (sidecar) and map-free signature discovery, with paired-view CIs. Bottom: discovery statistics; "rows collapsed" counts primitives merged into representatives.

Setting	Arm	Train	Novel
Bear K4 (jitter)	barycentric (sidecar)	71.61%	72.44%
Bear K4 (jitter)	map-free	**73.81** [72.07, 75.38]	**73.86** [72.46, 75.23]
Dino K8 (jitter)	map-free	**68.66** [66.74, 70.58]	**65.17** [61.45, 68.87]
Discovery	rows → groups	dup. groups (rows)	false merges / recovery
Bear K4	705,834 → 695,403	3,477 (13,908)	0 / 3,477–3,477
Dino K8	911,758 → 880,923	4,405 (35,240)	0 / 4,405–4,405

jointly over the whole candidate set, a candidate that fails any check reverts to a singleton rather than seeding a chained merge, so a single noisy primitive cannot corrupt a group. Across 27 controlled perturbation combinations (position, scale, opacity, rotation, and SH noise, singly and jointly) the matcher recovers $67{,}744/67{,}752$ groups with *zero* mixed-source groups; under unknown mixed multiplicities $K \in \{2, 4, 8\}$ it recovers $22{,}578/22{,}584$; and on the three unaugmented base assets it proposes *zero* merges, confirming that natural near-duplicates do not trip the filter. Byte-exact signatures, by contrast, are defeated by a 10^{-6} log-scale perturbation, tolerance is what makes recovery robust in real pipelines.

Fig. 8 shows the corresponding qualitative comparisons for perturbed, map-free inputs on Face, Garden, and Stone horse, using the same author/native/canonical layout.

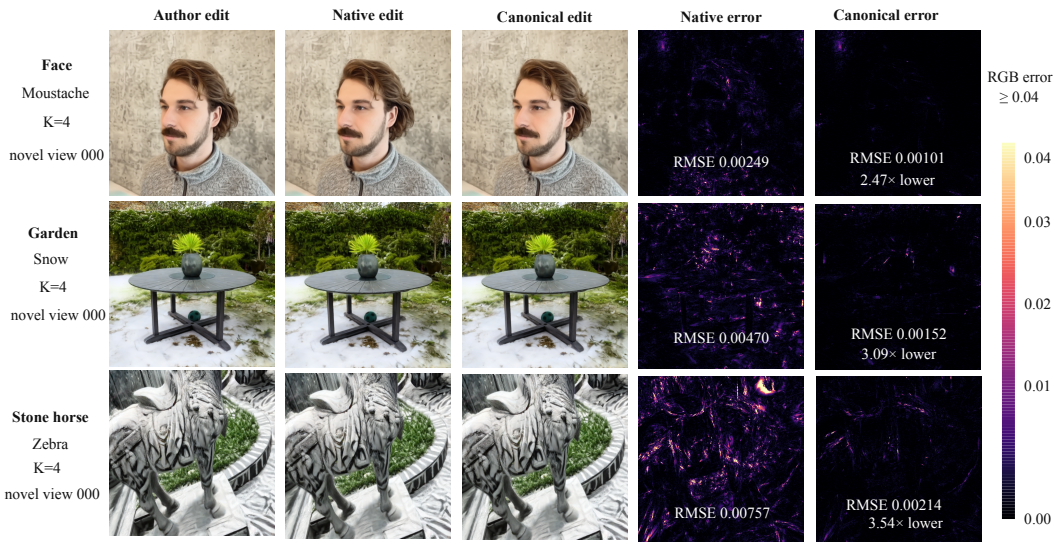

Figure 8: GaussCtrl perturbed, map-free refinements across scenes. Face with the moustache prompt, Garden with the snow prompt, and Stone horse with the zebra prompt, each at $K = 4$ on interpolated novel view 000. Native and canonical error maps are measured against the author edit. The shared RGB-error scale saturates at 0.04.

F MULTI-SEED ROBUSTNESS

Table 11 reports the per-seed reductions, and Fig. 9 compares native and canonical deviations across the same three frozen seeds and two prompts.

Table 11: Multi-seed robustness (Bear K4). Mean canonical reduction over three frozen split seeds; range in parentheses; all $3/3$ seeds positive. The canonical arm is seed-invariant by construction.

Prompt	Split	Seed 14	Seed 15	Seed 16	Mean (range)
bronze	train	82.38	82.16	82.49	82.34 (82.16–82.49)
bronze	novel	80.16	79.96	80.33	80.15 (79.96–80.33)
marble	train	80.85	80.96	81.33	81.05 (80.85–81.33)
marble	novel	81.51	81.63	82.02	81.72 (81.51–82.02)

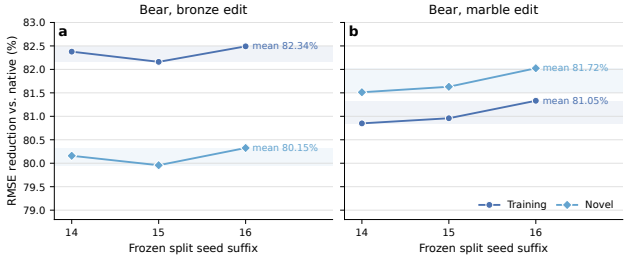

Figure 9: Multi-seed consistency (Bear K4). Native and canonical edit deviations for three frozen split seeds and two prompts; the canonical arm is seed-invariant and consistently close to the repeat floor.

G Cross-editor supplement: thresholds, seeds, and coordinates

Tables 12 and 13 report threshold sensitivity and seed robustness for the second editor. Fig. 10 provides complementary exact-refinement comparisons on Face and Stone horse.

Table 12: Threshold sensitivity of the cross-editor result (FastSplatStyler, five scenes, fifteen frozen split seeds). At every edit-retention threshold the strongest admissible uniform residual-shrink control is selected per scene and seed; the physical-token editor's scene-macro stability advantage and its 95% intervals are reported (100k-replicate scene bootstrap and hierarchical scene×seed bootstrap). All fifteen per-seed advantages are positive at every threshold.

Retained edit	Macro advantage (pp)	Scene bootstrap 95% CI	Hierarchical 95% CI
0.80	59.07	[54.60, 65.75]	[53.54, 67.68]
0.85	62.87	[58.03, 67.75]	[57.16, 70.24]
0.90	66.87	[61.83, 71.90]	[60.86, 73.65]
0.95	69.47	[65.50, 73.04]	[64.27, 75.04]
1.00	72.07	[68.81, 74.68]	[67.36, 76.82]

Table 13: Per-scene advantages at the 0.95 **operating point** (percentage points). Each scene aggregates three frozen split seeds; every seed-level advantage is positive.

Scene	Seed advantages	Aggregate
Broche	76.29 / 67.48 / 75.05	72.56
Bus	65.51 / 64.92 / 66.89	65.73
Family	63.48 / 63.53 / 56.32	63.07
M60	60.69 / 75.26 / 86.42	71.86
Table	72.34 / 71.78 / 79.05	74.11
Scene macro		69.47 [65.50, 73.04]

Conservation-coordinate ablation. The design rule, the measure coordinate must match the conservation law of the nuisance transformation, is testable independently of any editor. Family uses *alpha-additive* split children and M60 uses *transmittance- (τ-) additive* split children; depositing each split into the physical-field hierarchy with the matching versus the mismatched mass coordinate changes the token-measure drift by an order of magnitude (Table 14). The mismatched coordinate is

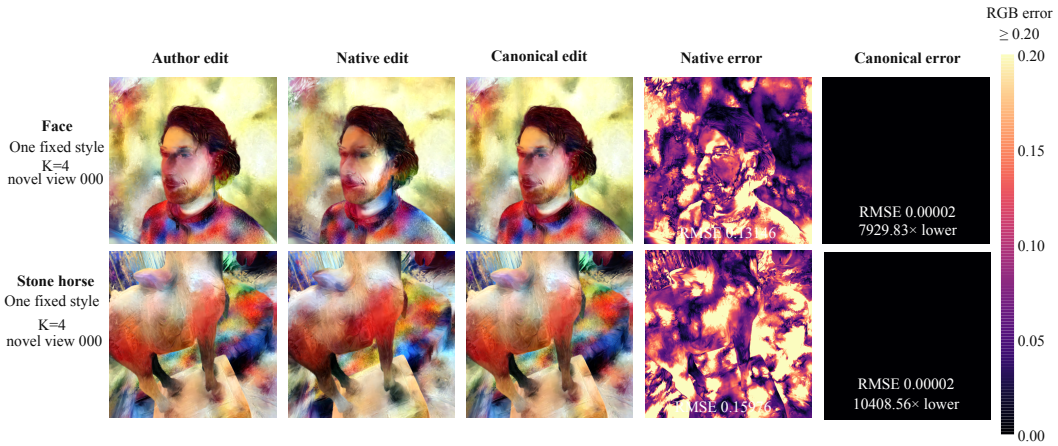

Figure 10: FastSplatStyler exact refinements. Face and Stone horse under one fixed style at $K = 4$, shown on interpolated novel view 000. Columns compare the author, native, and canonical edits and their error maps relative to the author edit. The shared RGB-error scale saturates at 0.20.

not merely worse on average: the mass-color drift, which directly feeds the style operator, degrades by 7.7–9.3× in both directions. Token-set Jaccard stays ≥ 0.99 because the fixed top-120k cutoff is dominated by cells near its boundary; the drift appears in the *weights* of the deposited measure, which is precisely what the style operator consumes.

Table 14: Conservation-matched vs. mismatched mass coordinate (input-only ablation, three frozen seeds per scene). Ratios > 1 mean the mismatched coordinate inflates the drift by that factor.

Scene (split law)	Matched coord.	Mass drift	Color drift	Jaccard (matched)
Family (alpha-additive)	alpha	10.55×	7.71×	0.9958–0.9990
M60 (τ-additive)	τ	10.11×	9.33×	0.9924–0.9955

Table 15: Per-seed deposition agreement under the matched coordinate (physical-token pipeline vs. its independent reimplementation): token-set Jaccard, normalized mass ℓ_1, and normalized mass-color ℓ_1.

Scene	Seed	Jaccard	Mass ℓ_1	Color ℓ_1
Family	20260922	0.995842	4.56×10^{-4}	6.36×10^{-4}
Family	20260923	0.999017	2.44×10^{-4}	2.58×10^{-4}
Family	20260924	0.995942	3.91×10^{-4}	5.62×10^{-4}
M60	20260922	0.993289	1.90×10^{-4}	1.94×10^{-4}
M60	20260923	0.992396	1.81×10^{-4}	2.09×10^{-4}
M60	20260924	0.995543	1.83×10^{-4}	1.82×10^{-4}

Bit-level reproducibility of the physical-token pipeline. A standalone reimplementation of the CIC deposition (`fastsplat_physical_tokens.py -mass-mode alpha`, fixture: Family split, seed 20260922) reproduces the archived token positions *exactly*: token count, full-token count, and the SHA-256 key of the deposited token set all match. The only differences are floating-point accumulation tails in the mass sums (input mass 7.7×10^{-3} absolute on $\approx 1.17 \times 10^{5}$, retained mass 3.2×10^{-6}, retained fraction 9.9×10^{-11}), i.e., checkpoint-round-trip precision on a 120k-token hierarchy. The cross-editor results therefore rest on a deterministic, independently reproducible artifact.

H Quantization as a third nuisance family

The refinement families of the main text perturb geometry and sampling; appearance quantization perturbs the *attribute encoding*. Per-channel 7-bit quantization of a G-Style checkpoint on Bus

preserves 62.84 dB input PSNR, practically invisible, yet the downstream style editor amplifies the discrepancy by roughly $250\times$ at the output. This extends the paper's central observation to a third, architecturally distinct representation change: render-equivalence at the input is no protection against nonlinear editing pipelines, whatever the source of the equivalence class.

Why evaluation must be end-to-end. A natural question is whether cheap *local* probes of gauge sensitivity could replace end-to-end edit comparisons. They cannot: with matched 1% moment probes on the original and quantized checkpoints, local gauge sensitivity is basin-dependent and is not an identifiable surrogate for finite cross-representation editing divergence (Table 16). The probe-uncertainty machinery makes this concrete: on the unaugmented checkpoint the probe statistic is dominated by uncertainty ($\mu = 0.938$, median and $q_{99} = 1.0$ over 27 occupied cells), while the split checkpoint yields near-zero uncertainty ($\mu = 2.6 \times 10^{-5}$), the two representations behave differently under every finite perturbation, so a local linearization cannot rank them. This is precisely why all stability claims in the paper are formed by running the full editor on both arms.

Table 16: Orbit-average projection under 7-bit quantization (G-Style, Bus). β interpolates between the probe target and the base edit; style and content are ratios to the base edit, LPIPS is the perceptual distance to the base output.

β	Deviation reduction	Style	Content	LPIPS
0.0	22.48%	1.064	0.817	0.926
0.1	27.47%	1.070	0.768	0.879
0.2	32.94%	1.076	0.699	0.819
0.4	43.66%	1.094	0.571	0.671
1.0	67.04%	1.161	0.275	0.218

Table 17: Uncertainty-filtered projection (λ sweep, coarse grid). The filter down-weights cells whose probe response is uncertain; reduction grows smoothly with λ.

λ	Deviation reduction
0.05	2.65%
0.10	6.75%
0.15	11.16%
0.20	15.64%
0.30	24.62%
0.40	33.60%

The protocol is non-degenerate. A benchmark that rewarded *any* smoothing of the representation would be uninformative. The quantization frontier shows the opposite (Table 18, exploratory): at matched clean-style constraint (style ratio $\leq 1.01\times$ native), a τ-projection operator reaches 12.6–21.4% deviation reduction, while the strongest uniform residual-shrink control reaches 28.2–48.7% at the same style budget, the protocol *discriminates* between operators rather than rewarding all of them. This is the same selection logic as the cross-editor control of Section 5.7 applied to a third family, and it keeps the benchmark honest: gains must come from removing mixture freedom, not from any stability-increasing operation.

Table 18: Exploratory operator frontier under 7-bit quantization (G-Style, Bus). Style/content are ratios to the base edit; the style constraint caps the first-three-view style distance at $1.01\times$ native.

Operator	Strength	Reduction	Style	Content
τ-projection	0.3	12.56%	0.996	0.830
τ-projection	0.5	21.43%	1.027	0.702
Uniform shrink	0.3	28.22%	1.018	0.691
Uniform shrink	0.4	38.46%	1.065	0.594
Uniform shrink	0.5	48.72%	1.119	0.502

I SCOPE-EXPANSION SUPPLEMENT

Natural exact classes: raw counts. Detection uses the same byte-equality signature over non-position attributes, without any artificial split or lineage sidecar. Base checkpoints: Face 193,589 primitives → 7 duplicate groups (all size 2, 14 rows); Garden 729,355 primitives → 21 pairs and 3 triples (24 raw signature groups); Stone horse 582,630 primitives → 27 groups (54 rows). Edited checkpoints: Face 6, Garden 24 detected with the collapse to 149.83 dB covering the 21 size-2 groups (the 3 triples are detected but not included in that collapse), and Stone horse 25. These classes arise without any artificial split: real optimization itself leaves behind the redundant-coordinate structure that our gauge removes. They are existence evidence for the equivalence classes, and each collapse is verifiable end-to-end by the same render-equivalence gate as the planted refinements.

Optimizer audit and continued training. The audited entry point rebuilds the optimizer after checkpoint loading, restores its state, and asserts optimizer–parameter binding at every step; all fifteen additional-scene trainings record first-step updates of 10^5–10^7 elements in every parameter group (`verified_update_audits.json`). After 50 steps of *genuine* continued optimization on Face (synchronized base and refined checkpoints, post-densification culling disabled), the tolerant matcher recovers $941/968$ lineages at tolerance 10^{-3} ($81/968$ at 10^{-4}, $965/968$ at 10^{-2}; zero mixed groups; Fig. 11), the byte-exact signature only $7/968$. The map-free-collapsed continued checkpoint renders at 68.97 dB global (65.10 dB worst view) against the synchronized continued base, versus $67.82/64.55$ dB for the native continued refined checkpoint. Recovery capacity therefore persists under realistic training drift, and the recovery threshold curve indicates substantial headroom (Fig. 11).

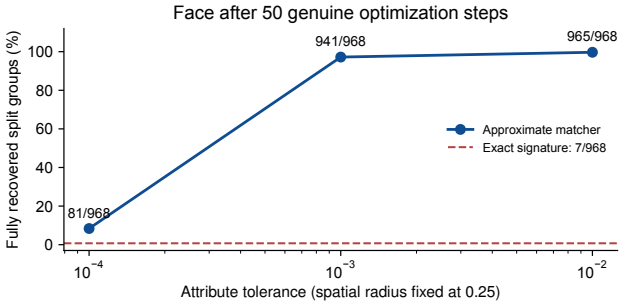

Figure 11: Recovery after 50 genuine optimization steps (Face). Fully recovered split groups vs. attribute tolerance (spatial radius fixed at 0.25); the byte-exact signature recovers only $7/968$. No mixed-source groups at any tolerance.

J ROW-ORDER PROTOCOLS IN THE SECOND EDITOR

For *approximate* refinements, the second editor is strongly sensitive to row order: under the original row-order protocol the map-free canonical edit scores *worse* than native (novel RMSE 0.1475 vs. 0.1430, -3.15%). Diagnostics isolate the cause: permuting the rows of the *unsplit* asset alone reproduces the deviation (0.1480 RMSE on training poses), while same-marginal attribute noise contributes only ≈ 0.0048, ordering, not attribute noise, is the dominant confound. Extending the protocol with an explicit coordinate-ordering step (xyz lexicographic order applied identically to the author, native, and canonical checkpoints, with each protocol using its own author reference) restores a 56.82% reduction on Face novel views ($0.1417 \to 0.0612$; paired-camera 95% CI $[55.85, 57.77]$, recomputed from per-camera MSE). The two protocols are reported separately and are not cross-comparable; the ordered result is a single-scene, single-style, single-seed diagnostic and does not establish approximate-refinement stability for all editors. Exact refinements are unaffected: row order is fixed by the collapse itself, and the same-interface transfer of Section 5.7 applies.

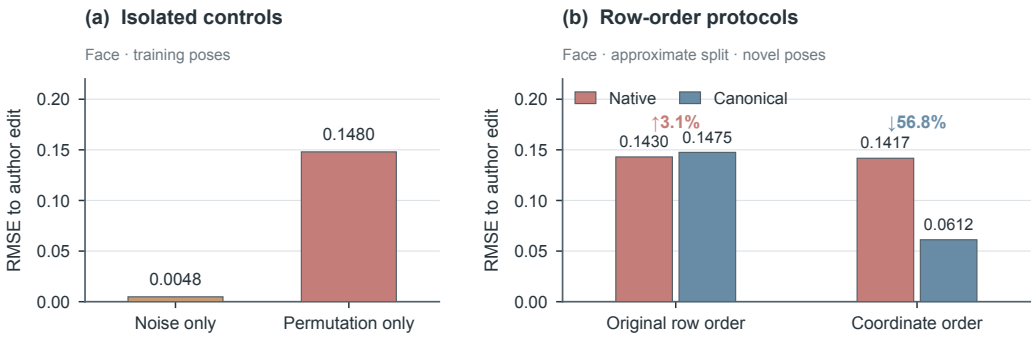

Figure 12: Row-order controls for the second editor (Face). Left: isolated controls on training poses, same-marginal noise alone contributes only ≈ 0.0048 RMSE, while permuting the rows of the unsplit asset alone reproduces the full deviation (0.1480). Right: approximate split on novel poses under the original row-order protocol (canonical worse than native) and under the coordinate-order protocol (canonical better by 56.8%). Each protocol uses its own author reference.

K Runtime and cost summary

The wrapper's cost is negligible relative to both editors: a one-time $O(N)$ linear scan (or a 22 s discovery pass when no lineage metadata exists) against multi-minute optimization-based editing or a 14–20 s physical-token pass. Because canonicalization is applied before the editor and never touches its code path, it composes with any future editor without re-tuning.

Table 19: Cost of the wrapper and the editors. Canonicalization runs once per asset, before any editing.

Operation	Scope	Cost
Exact / barycentric collapse	706k primitives, CPU	$\sim$8 s
Map-free discovery + collapse	Bear K4 path, CPU	22 s
GaussCtrl edit (any arm)	whole scene, RTX 4090D	$\sim$14 GiB peak
FastSplatStyler graph + edit	Bus scene	20.26 s / 1.70 GB RSS
FastSplatStyler author path	Bus scene	13.80 s / 1.61 GB RSS