FROM GAUSSIAN TOKENS TO OPTICAL MEASURES: REFINEMENT-ROBUST LEARNING ON 3DGS

Anonymous authors
Paper under double-blind review

ABSTRACT

Adaptive refinement makes 3D Gaussian Splatting (3DGS) non-canonical: splitting, merging, and pruning alter primitive count without changing radiance. Yet prevailing set architectures uniformly weight Gaussians, leaving downstream representations vulnerable to refinement. We establish refinement invariance as a core representation symmetry and derive Optical Measure Attention from volume rendering physics. Because optical depth $\tau_i = -\log(1-\alpha_i)$ is strictly additive under volume compositing, we model scenes as finite measures $\mu = \sum_i \tau_i \delta_{z_i}$, injecting $\log \tau_i$ as an attention logit bias and τ_i into normalized moments. This guarantees exact invariance under co-located mass-preserving splits and proves an $\mathcal{O}(r^2)$ error bound under child displacements of radius r. Under factor-8 refinement, it preserves 100.00% of Conv predictions versus 94.18% for ordinary attention (median logit drift $15.80\% \to 0.52\%$) and 99.66% vs. 91.44% on query-based pooling; In metric retrieval, it lifts top-10 neighbor retention from $\sim$77% to 96% across six decimation pipelines. Finally, an online streaming kernel executes exact quadrature with memory independent of Gaussian cardinality, reducing peak memory by $18.48\times$. These results establish a principled paradigm: integrate continuous optical measures rather than count discretization primitives.

1 INTRODUCTION

Three-dimensional Gaussian Splatting (3DGS) has become a central explicit representation for real-time novel-view synthesis, complementing the success of implicit neural radiance fields (Mildenhall et al., 2021; Müller et al., 2022) and supporting a rapidly expanding range of applications in rendering, SLAM, segmentation, and 3D generation (Chen & Wang, 2026; Kerbl, 2025; Fei et al., 2024). A 3DGS scene is represented by a variable-cardinality collection of anisotropic Gaussians, each carrying geometry, appearance, and opacity. This explicit structure is increasingly consumed directly by downstream models for classification, retrieval, and segmentation, typically through permutation-invariant architectures such as Deep Sets (Zaheer et al., 2017) and Set Transformer (Lee et al., 2019), which treat every primitive as an equally weighted token.

The core issue is that primitive cardinality is an artifact of the optimization trajectory rather than an invariant property of the underlying scene. The 3DGS optimizer's adaptive density control continuously densifies, splits, merges, and prunes Gaussians to minimize rendering loss (Kerbl et al., 2023; Rota Bulò et al., 2024; Kim et al., 2026; Sandu et al., 2026; Gafoor et al., 2025; Zhou et al., 2026b). Recent work on densification (Fang & Wang, 2024; Rota Bulò et al., 2024) and compression (Niedermayr et al., 2024; Lee et al., 2024; Bagdasarian et al., 2025) further modify the number and spatial allocation of primitives. Consequently, two representations can encode essentially the same scene evidence while allocating very different numbers of Gaussians to individual regions. A downstream network that treats each primitive as an equal unit of evidence then confounds semantic evidence with an optimizer-dependent discretization choice: regions represented by more Gaussians receive more influence simply because they contain more tokens, as illustrated in Figure 1.

Permutation invariance, the standard tool for set-structured data (Zaheer et al., 2017; Lee et al., 2019; Komiske et al., 2019; Chao et al., 2019), preserves ordering symmetry but does not address refinement: replacing one token by several co-located children changes the pooled output even when the underlying scene is identical. This is a *nuisance symmetry*: a transformation that should not affect the prediction but does, which is analogous to the discretization symmetries studied in

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.

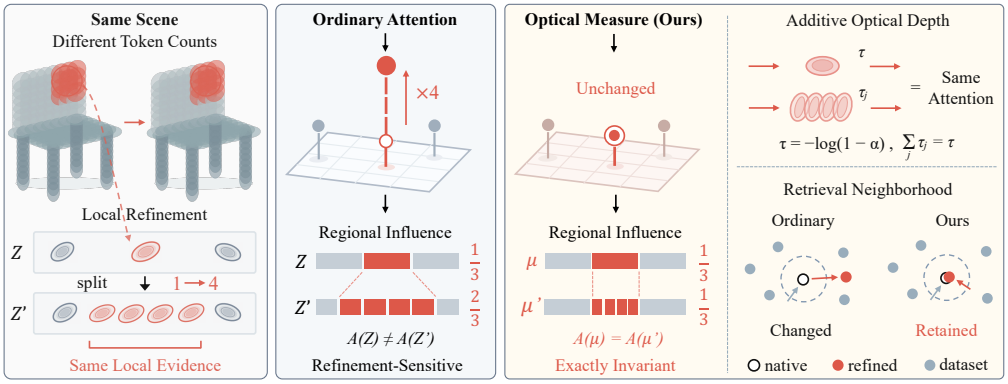


Figure 1: **Refinement Invariance Beyond Permutation.** Splitting one Gaussian into four changes the token count, not the scene. Ordinary attention that counts tokens inherits the artifact: $A(Z) \neq A(Z')$. In contrast, Optical Measure Attention integrates additive optical mass to guarantee exact invariance $A(\mu) = A(\mu')$: regardless of how the optimizer splits, merges, or prunes the Gaussians, our attention output and downstream predictions remain provably invariant.

geometric deep learning (Kondor, 2025; Han et al., 2025) but arising specifically from the 3DGS optimization pipeline.

We propose to resolve this by *integrating the represented optical measure instead of counting the atoms used to discretize it*. The key insight is that 3DGS alpha compositing multiplies transmittance along each ray, so optical depth $\tau = -\log(1-\alpha)$ is additive: co-located children whose opacities compose to the parent's opacity have optical depths that sum to the parent's. We define a finite measure $\mu = \sum_i \tau_i \, \delta_{z_i}$ on the scene's Gaussian support and replace ordinary attention with *measure-attention*, whose weights $w_i = \tau_i e^{s_i} / \sum_j \tau_j e^{s_j}$ are exactly invariant to any co-located, mass-preserving refinement.

Our contributions are summarized as follows:

- We identify set-level refinement sensitivity as a previously unmodeled nuisance in downstream 3DGS learning distinct from ordering and primitive-wise parameter ambiguity.
- We derive an optical-depth measure from alpha compositing and implement a drop-in change of base measure for attention and feature moments.
- We prove exact invariance to feature-identical, co-located mass-preserving splits and establish a second-order perturbation bound for non-identical moment-preserving refinement.
- We validate the principle across two architectures and three 3DGS pipelines, verify the mechanism through mass-law and matched-model controls, extend the evaluation to natural merge/pruning transformations and instance-level metric retrieval, and provide deterministic one-GPU streaming at native cardinality.

2 Related Work

2.1 Sets, Measures, and Attention

Deep Sets (Zaheer et al., 2017) established permutation-invariant set learning, while the Set Transformer (Lee et al., 2019) introduced attention-based set aggregation. Energy-flow networks (Komiske et al., 2019) weighted particle tokens by energy, providing a physically motivated precursor to our optical-depth weighting, but without connecting it to refinement invariance or 3DGS compositing. Related work on weighted probability measures, measure-theoretic attention, and continuum operators (de Bie et al., 2019; Vuckovic et al., 2020; Calvello et al., 2025; Kawata & Suzuki, 2026) provides the operator-theoretic foundations for measure-normalized attention. Graph and point-cloud networks (Zhou et al., 2020; Xu et al., 2018; Wang et al., 2019; Qi et al., 2017; Wu et al., 2025; Zhang et al., 2025b) extend set processing to relational and geometric data, but generally assume fixed structures or sampling schemes; see Kamal et al. (2026); Sarker et al. (2024);

Guo et al. (2019). More broadly, geometric deep learning (Han et al., 2025; Kondor, 2025), physics-informed ML (Meng et al., 2025; Yu & Wang, 2024), and equivariant neural rendering (Dupont et al., 2020) imposed view-consistency constraints.

2.2 3D Gaussian Splatting

Since the introduction of 3DGS (Kerbl et al., 2023), the field has expanded rapidly into SLAM (Matsuki et al., 2024; Keetha et al., 2024; Yan et al., 2024), dynamic scene rendering (Yan et al., 2024; Yang et al., 2024), semantic segmentation (Cen et al., 2025; Zhou et al., 2026a), inverse rendering (Liang et al., 2025), single-view reconstruction (Szymanowicz et al., 2024; Hong et al., 2024), multimodal rendering (Gau et al., 2025), and sparse-view reconstruction (Liu et al., 2025a). Recent surveys (Chen & Wang, 2026; Fei et al., 2024; Wu et al., 2024; Kerbl, 2025) document this breadth. Compression and pruning methods (Niedermayr et al., 2024; Lee et al., 2024; Bagdasarian et al., 2025) reduce the Gaussian count for storage and rendering, while densification improvements (Rota Bulò et al., 2024; Kim et al., 2026; Sandu et al., 2026; Gafoor et al., 2025; Fang & Wang, 2024) alter the primitive allocation during optimization. Renderer-side analysis has corrected opacity under co-located cloning and revised the densification criterion (Rota Bulò et al., 2024). Hierarchical representations (Kerbl et al., 2024; Zhou et al., 2026b), foveated rendering (Fan et al., 2025), and alias-free splatting (Yu et al., 2024) further modify the token set at inference time. All of these operations change the Gaussian list that a downstream model consumes, but none address the downstream stability of learned representations under these changes.

2.3 Learning from Gaussian Tokens

MACGS establishes native 3DGS classification (Zhang et al., 2025a); CLIP-GS and SceneSplat develop semantic and vision-language Gaussian representations (Jiao et al., 2025; Li et al., 2025). Other recent work benchmarks Gaussian representation learners (Farganus et al., 2026) or builds efficient Gaussian point encoders (James et al., 2026). Compact representations (An et al., 2025) and primitive canonicalization (Xin et al., 2026) address per-atom parameter ambiguity. Our contribution is complementary: the pooling operator is made insensitive to how an upstream system refines the support of the represented optical measure, a set-level concern distinct from primitive-wise canonicalization.

3 Optical Measure Attention

3.1 Gaussian Atoms and Additive Optical Mass

Let $z_i = (x_i, C_i, c_i, \dots)$ collect the normalized mean, covariance, appearance, and auxiliary attributes of Gaussian i, and let a_i be its opacity logit. We define

$$m_i = \tau_i = \mathrm{softplus}(a_i) = -\log\big(1-\sigma(a_i)\big) > 0. \tag{1}$$

and represent the object as the finite measure

$$\mu = \sum_{i=1}^{n} m_i \, \delta_{z_i}. \tag{2}$$

If co-located children have transmittances $1-\alpha_j$, their composite transmittance is $\prod_j (1-\alpha_j)$ and their optical depths add. Thus Equation 1 provides a conserved quadrature mass: splitting an atom of mass m into K children with masses $m_j > 0$ and $\sum_j m_j = m$ leaves μ unchanged, whereas the token count grows from 1 to K. This is the property that ordinary counting attention lacks.

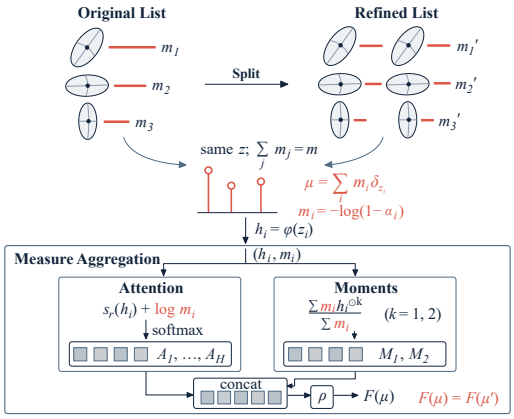


Figure 2: Measure Attention aggregates Gaussian features with optical-depth masses. Co-located, feature-identical mass splits preserve the optical measure and its attention and moment statistics, yielding the same output in exact arithmetic.

3.2 CHANGING THE BASE MEASURE OF ATTENTION

Let $h_i = \varphi(z_i)$ and let $s_r(h_i)$ denote a scalar score for head or query r. We compute

$$A_r(\mu) \;=\; \frac{\sum_i m_i\, e^{s_r(h_i)}\, h_i}{\sum_i m_i\, e^{s_r(h_i)}} \;=\; \sum_i \mathrm{softmax}_i\big(s_r(h_i) + \log m_i\big)\, h_i, \tag{3}$$

$$M_1(\mu) \;=\; \frac{\sum_i m_i\, h_i}{\sum_i m_i}, \qquad M_2(\mu) \;=\; \frac{\sum_i m_i\, (h_i \odot h_i)}{\sum_i m_i}. \tag{4}$$

The classifier is $F(\mu) = \rho\big([A_1, \ldots, A_H,\, M_1,\, M_2]\big)$. We instantiate Equation 3 with an eight-head scalar-score Conv aggregator and an eight-query dot-product PMA aggregator. Mass is a quadrature weight, not a token content feature: it enters the attention denominator and the moment normalizer, not the per-token feature vector h_i. This distinction is what makes the invariance structural: adding τ_i as an input feature without changing the base measure does *not* confer refinement robustness (Appendix D).

Training and Native Inference. The shared checkpoints use a systematic optical-mass-resampled $1{,}024$-token reference and alternating $256/512$-token subsets:

$$\mathcal{L} \;=\; \mathrm{CE}\big(y, f(Z_{1024})\big) + \tfrac{1}{2}\,\mathrm{CE}\big(y, f(Z_b)\big) + \tfrac{1}{2}\,\mathrm{KL}\big(p^{\mathrm{stop}}_{1024} \,\|\, p_b\big). \tag{5}$$

Training forward calls use ordinary aggregation. Full measure integration is enabled at native inference, so ordinary and Measure Attention share a checkpoint and differ only in their integration rule. This isolates the integration law from representation learning.

Streaming. Attention numerators and denominators are accumulated in chunks using online log-sum-exp, alongside weighted moments. The algorithm integrates the same finite measure without materializing all encoded primitives simultaneously.

4 REFINEMENT INVARIANCE: THEORY

4.1 EXACT CO-LOCATED REFINEMENT INVARIANCE

Theorem 1 (Exact co-located invariance)**.** *Let $(\{h_i\}, \{s_r\}, \{m_i\})$ define a measure-attention instance with pooled output $A_r(\mu)$ and moments (M_1, M_2) via equation 3–equation 4. Replace the atom at index k by K co-located, feature-identical children with masses $\{m_k^{(j)}\}_{j=1}^K$ satisfying $\sum_j m_k^{(j)} = m_k$, identical features $h_k^{(j)} = h_k$, and identical scores $s_r(h_k^{(j)}) = s_r(h_k)$. Then the refined pooled output $\widetilde{A}_r(\mu)$ and refined moments $(\widetilde{M}_1, \widetilde{M}_2)$ satisfy*

$$\widetilde{A}_r(\mu) = A_r(\mu), \quad \widetilde{M}_1 = M_1, \quad \widetilde{M}_2 = M_2.$$

Proof. In the refined measure, the denominator of the attention weight becomes $\sum_{i\neq k} m_i e^{s_r(h_i)} + \sum_{j=1}^K m_k^{(j)} e^{s_r(h_k)} = \sum_{i\neq k} m_i e^{s_r(h_i)} + m_k e^{s_r(h_k)}$, which is unchanged. The numerator for any feature component f is $\sum_{i\neq k} m_i e^{s_r(h_i)} f(h_i) + \sum_j m_k^{(j)} e^{s_r(h_k)} f(h_k) = \sum_{i\neq k} m_i e^{s_r(h_i)} f(h_i) + m_k e^{s_r(h_k)} f(h_k)$, also unchanged. Hence $\widetilde{A}_r = A_r$. The moments use $\sum_j m_k^{(j)} h_k = m_k h_k$ and $\sum_j m_k^{(j)} (h_k \odot h_k) = m_k (h_k \odot h_k)$, giving $\widetilde{M}_1 = M_1$ and $\widetilde{M}_2 = M_2$. □

This theorem is stronger than permutation invariance: it covers any refinement that preserves the optical-depth mass at a given location, including the splits, clones, and merges performed by 3DGS adaptive density control (Kerbl et al., 2023; Rota Bulò et al., 2024). The proof relies solely on the additivity of τ and the feature-identical assumption, which holds exactly when the 3DGS optimizer replaces one Gaussian by co-located children that compose to the same opacity.

4.2 NON-IDENTICAL MOMENT-PRESERVING REFINEMENT

In practice, 3DGS refinement does not produce perfectly co-located children; the optimizer may perturb their centres, covariances, and features. We show that the invariance degrades gracefully when the children preserve the parent's first and second physical moments.

For non-identical children, let

$$x_j = x + \delta_j, \qquad C_j = C - B, \qquad \frac{1}{K}\sum_{j=1}^{K}\delta_j = 0, \qquad B = \frac{1}{K}\sum_{j=1}^{K}\delta_j\delta_j^\top. \tag{6}$$

The mixture preserves mass, mean, and physical second moment.

Theorem 2 (Second-order stability under moment-preserving refinement)**.** *Let $g(x, C, c)$ have covariance derivative norm at most L_C and full Hessian norm at most H_g along every parent–child segment. Then*

$$\left\| \frac{1}{K}\sum_{j=1}^{K} g(x+\delta_j,\, C-B,\, c) - g(x, C, c) \right\| \;\le\; L_C\,\|B\| \;+\; \frac{H_g}{2K}\sum_{j=1}^{K}\big(\|\delta_j\| + \|B\|\big)^2.$$

Proof. The first-order position term cancels by symmetry ($\sum_j \delta_j = 0$); the covariance term $L_C\|B\|$ and the Hessian remainder are both second-order. Since $B = O(r^2)$ for maximum displacement r, the entire bound is $O(r^2)$. □

Applying the theorem separately to the attentive numerator $g_N(z) = e^{s_r(\varphi(z))}\varphi(z)$ and denominator $g_D(z) = e^{s_r(\varphi(z))}$ yields a ratio perturbation bound (Appendix A): if both denominators are at least $\beta > 0$ and the perturbations are at most ϵ_N, ϵ_D, then

$$\|A_r(\widetilde{\mu}) - A_r(\mu)\| \;\le\; \frac{\epsilon_N}{\beta} + \frac{\|N_r(\mu)\|\,\epsilon_D}{\beta^2}. \tag{7}$$

The same argument controls M_1 and M_2. If the classifier head ρ is Lipschitz, this bounds the logit change; a native top-1 margin larger than twice the resulting uniform logit perturbation prevents a flip.

The moment-preserving construction equation 6 is directly implementable: given a parent Gaussian with mean x and covariance C, one generates K children with offsets $\{\delta_j\}$ satisfying $\sum_j \delta_j = 0$ and $\frac{1}{K}\sum_j \delta_j\delta_j^\top \preceq C$, then sets the child covariance to $C - \frac{1}{K}\sum_j \delta_j\delta_j^\top$. This is the construction used in our factor-8 controlled refinement (Section 5).

5 EXPERIMENTS

We evaluate measure-attention on four axes: (1) controlled factor-8 refinement stability, (2) instance-level metric retrieval under the same refinement, (3) natural merge/pruning transformations, and (4) deterministic streaming deployment. All numerical results are reproducible from the provided code and data.

5.1 EXPERIMENTAL SETUP

3DGS Pipelines. We use three reconstruction pipelines producing Gaussian splats of varying cardinality: MACGS (370 objects), uCO3D (Liu et al., 2025b) (70 objects, 210 official views), and ModelNet10 (Wu et al., 2015) (222 objects). The combined corpus contains 662 objects.

Attention Architectures. We evaluate two pooling architectures: a convolutional feature extractor with global average pooling (**Conv**) and a query-based pooled multi-head attention (**Query-PMA**) following the Set Transformer design (Lee et al., 2019). Both are trained natively (without refinement augmentation) and then tested under refinement.

Table 1: Factor-8 moment-preserving refinement: prediction stability and logit drift across two attention architectures (292 external objects, reference displacement $\lambda = 0.5$).

Architecture	Stability (%) ↑		Median drift (%) ↓		Excess flip (pp) [95% CI]
	Ordinary	Measure	Ordinary	Measure	
Conv	94.18	**100.00**	15.80	**0.52**	5.82 [3.42, 8.56]
Query-PMA	91.44	**99.66**	11.96	**0.88**	8.22 [5.14, 11.64]
Three-seed*	91.44	**99.77**	—	—	8.33 [5.48, 11.42]

Refinement Protocol. For controlled refinement, we select the half of each object's primitives with the largest x-coordinates and replace each selected atom by eight moment-preserving children with equal mass $\tau/8$, displaced features, and adjusted covariance (Theorem 2). We sweep the displacement multiplier $\lambda \in \{0, 0.25, 0.5, 0.75, 0.95\}$, where $\lambda = 0$ is the exact co-located case (Theorem 1); the headline numbers in Table 1 correspond to the reference displacement $\lambda = 0.5$. This controlled split is a diagnostic intervention, not an estimate of how often a particular natural densifier behaves this way.

Metrics. For classification we report *prediction stability* (fraction of objects whose top-1 prediction is unchanged after refinement), *median relative logit drift* (median over objects of $\|\ell_{\text{refined}} - \ell_{\text{native}}\|/\|\ell_{\text{native}}\|$), and *reference agreement* (agreement with a held-out reference label). For retrieval we additionally report *identity Recall@k* (whether the refined query's own native representation is among its top-k nearest neighbours), *top-10 neighbour retention* (fraction of the native top-10 neighbours recovered after refinement), and *median embedding drift* (median $\|\mathbf{e}_{\text{refined}} - \mathbf{e}_{\text{native}}\|$ over objects).

5.2 Controlled Factor-8 Refinement

Table 1 reports the headline factor-8 results. Measure-attention achieves 100% prediction stability on Conv and 99.66% on Query-PMA, versus 94.18% and 91.44% for ordinary attention. The median relative logit drift drops from 15.80% to 0.52% (Conv) and from 11.96% to 0.88% (Query-PMA), a 30.4× and 13.6× reduction, respectively. The paired 95% confidence intervals for the excess flip rate exclude zero in both cases: $[3.42, 8.56]$ pp (Conv) and $[5.14, 11.64]$ pp (Query-PMA).

A three-seed replication on 876 seed–object pairs confirms the direction: measure-attention achieves 99.77% stability versus 91.44%, coincidentally identical to the single-seed ordinary rate to two decimals, with a hierarchical 95% CI of $[5.48, 11.42]$ pp for the excess flip rate. All six seed–dataset cells are direction-consistent. Nevertheless, the multiseed package is recorded as 9/10: seed 23 loses one additional uCO3D refined prediction, exceeding the pre-declared one-point noninferiority bound.

5.3 Mass-Law Ablation: Why Optical Depth?

Not every mass law yields refinement invariance. We ablate four choices: uniform (no mass), raw alpha α_i, clipped alpha $\min(\alpha_i, 0.99)$, and optical depth τ_i, on the same factor-8 protocol (Figure 3 left). Uniform weighting reproduces ordinary attention (94.18% stability, 15.80% drift). Raw alpha achieves 99.32%/1.08%, close but not exact, because α is not additive under co-located compositing. Only optical depth achieves 100% stability and 0.53% drift (Table 5, Appendix B).

5.4 Graceful Degradation under Displacement

Figure 3 right traces the median logit drift as a function of the child displacement multiplier λ. At $\lambda = 0$ (exact co-located), measure-attention drift is exactly zero, matching Theorem 1. As λ increases to 0.95, the drift rises gradually to 2.08%, still an order of magnitude below ordinary attention's 17.38%. Ordinary attention is essentially insensitive to λ because its instability comes from multiplicity change, not displacement. This confirms the second-order bound of Theorem 2.

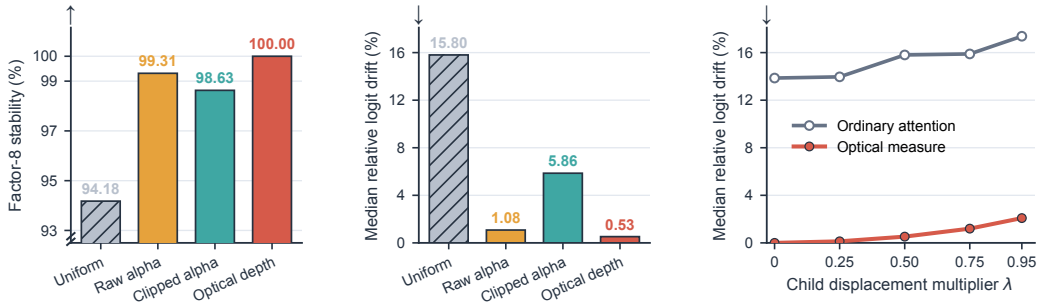


Figure 3: **Left:** Mass-law ablation. **Center:** Median logit drift by mass law. Optical depth achieves 100% stability and minimal drift. **Right:** Graceful degradation: measure-attention drift grows as $O(\lambda^2)$, while ordinary attention is flat at $\sim$15%.

Table 2: Factor-8 metric retrieval (876 seed–object pairs, three seeds). Identity Recall@1, top-10 neighbour retention, and median embedding drift by dataset.

Dataset	Identity R@1 (%) ↑		Top-10 retention (%) ↑		Median drift (%) ↓	
	Ord.	Meas.	Ord.	Meas.	Ord.	Meas.
uCO3D	97.62	**100.00**	75.05	**96.10**	0.354	**0.039**
ModelNet10	94.14	**100.00**	79.01	**95.77**	0.257	**0.048**
Pooled	94.98	**100.00**	—	—	—	—

5.5 Metric Retrieval Under Factor-8 Refinement

Classification stability confirms that the *argmax* prediction survives refinement, but it does not reveal whether the continuous embedding space, and the local neighbourhood structure that retrieval depends on, is also preserved. To test this, we train an independent metric-learning encoder from scratch using an instance-contrastive InfoNCE objective (Oord et al., 2018) on the same uCO3D (70 objects) and ModelNet10 (222 objects) pipelines. Training reads only a fixed 1,024-token reference cache; at test time the native scene serves as the gallery and the factor-8 refined scene serves as the query. Ordinary counting and measure-attention share the same encoder, training schedule, and data, differing only in their aggregation rule. Three independent seeds (11, 23, 37) yield 876 seed–object pairs. Figure 4 illustrates the aggregation mechanism: on a uCO3D example, refinement redistributes 38.52% of ordinary attention but only 0.13% of optical-measure attention, measured after summing child weights to their parents.

Table 2 reports the results. Measure-attention achieves 100.00% identity Recall@1 on both datasets and all three seeds, versus 94.98% pooled for ordinary attention, a +5.02 pp gain with a hierarchical 95% CI of $[3.08, 7.08]$ pp. The advantage extends beyond exact matching: top-10 neighbour retention rises from 75.05% to 96.10% on uCO3D and from 79.01% to 95.77% on ModelNet10, gains of +21.05 and +16.76 pp. Median embedding drift drops by 9.18× (0.354→0.039) on uCO3D and 5.33× (0.257→0.048) on ModelNet10. Semantic mAP is essentially unchanged, confirming that the stability gain does not come at the cost of collapsed or degenerate representations. All 27 pre-registered gates pass.

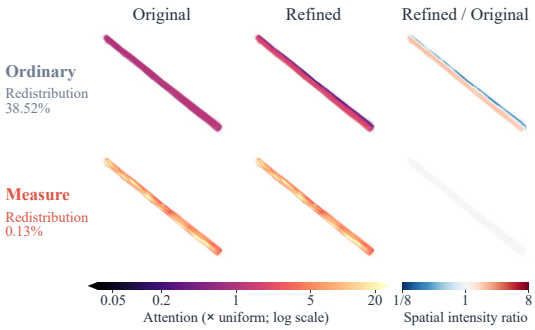


Figure 4: Attention on a uCO3D bamboo object. Moment-preserving refinement shifts ordinary attention substantially, while Measure Attention remains stable.

Table 3: Natural uCO3D compression: visual fidelity, prediction stability, and logit-drift reduction (70 objects and 210 official views).

Condition	PSNR_{med} (dB)	Prediction stability (%) ↑		Drift reduction (pp; 95% CI)
		Ord.	Measure	
Merge 50%	47.36	97.14	**98.57**	3.27 [1.63, 5.51]
Merge 75%	51.18	95.71	**100.00**	8.45 [6.05, 11.23]
Mixed 50%	47.79	95.71	**100.00**	9.15 [6.02, 12.65]
Prune 25%	70.19	**100.00**	**100.00**	3.43 [2.39, 4.68]
Prune 50%	53.78	**100.00**	**100.00**	5.99 [3.97, 8.37]
Prune 75%	43.83	98.57	**100.00**	7.09 [4.46, 10.21]

Table 4: Natural uCO3D compression retrieval: top-10 neighbour retention and median embedding-drift reduction (1,260 seed–object–condition pairs, three seeds). Identity R@1 is 100% for both methods under all six conditions.

Condition	Top-10 retention (%) ↑		Gain	Drift reduction
	Ord.	Measure	(pp; 95% CI)	(×; 95% CI)
Merge 50%	90.48	**94.00**	+3.52 [2.19, 4.86]	1.60 [1.42, 1.82]
Merge 75%	89.52	**96.48**	+6.95 [5.33, 8.81]	3.67 [3.30, 4.21]
Mixed 50%	86.33	**93.57**	+7.24 [5.29, 9.29]	2.68 [2.35, 3.02]
Prune 25%	94.14	**98.95**	+4.81 [3.71, 6.05]	16.37 [10.79, 24.09]
Prune 50%	87.76	**96.29**	+8.52 [7.05, 10.05]	3.47 [2.92, 4.12]
Prune 75%	80.38	**89.29**	+8.90 [6.95, 10.90]	1.77 [1.62, 2.13]

5.6 Natural Transformations Across Tasks

Beyond controlled factor-8 splits, we evaluate six *natural* transformations representative of real 3DGS compression and editing pipelines on 70 uCO3D objects (210 official views): local moment-merge at 50% and 75%, a mixed merge/prune transformation at 50%, and opacity pruning at 25%, 50%, and 75%. For each transformation, we measure the RGB PSNR of the rendered image to assess visual fidelity, the prediction stability of ordinary and measure-attention, and the reduction in logit drift achieved by measure-attention.

Table 3 summarizes the results. All six transformations preserve high visual fidelity, with median PSNRs ranging from 43.8 to 70.2 dB. Measure-attention improves prediction stability under four of the six transformations, with gains ranging from 1.43 to 4.29 pp, and ties on the remaining two (where ordinary attention is already at 100%). More importantly, *all six* paired 95% CIs for the mean logit-drift reduction are strictly positive, spanning 3.27–9.15 pp. The strongest reductions come from mixed and 75% merge transformations, which alter the token set most aggressively. These results show that measure-attention improves or preserves prediction stability while consistently reducing representation drift in natural 3DGS compression operations.

Retrieval Under Natural Compression. We further evaluate retrieval stability on the same six transformations by reusing the three metric-learning checkpoints from Section 5.5 without any retraining or tuning. Across all six conditions and all three seeds, both ordinary and measure-attention achieve 100% identity Recall@1, indicating that exact identity matching saturates under these operations. The more discriminative top-10 neighbour-retention metric, however, reveals a consistent advantage: measure-attention improves top-10 retention by +3.52 to +8.90 pp, and *all six* hierarchical 95% confidence intervals are strictly positive (Table 4). Median embedding drift is reduced by 1.60× to 16.37×, with the largest reduction occurring under 75% opacity pruning. In all 9/9 pruning seed–condition cells, measure-attention has lower drift and identity no worse than ordinary attention. These results confirm that the retrieval-space geometry advantage observed under controlled factor-8 refinement transfers to real 3DGS compression operations.

5.7 Deterministic Streaming and Resampling

A practical consequence of the measure formulation is that a single deterministic forward pass integrates the optical measure without the offset sensitivity of token subsampling. We compare measure-attention (single pass, full token set) against a five-offset ensemble that subsamples 1024 tokens per offset and averages predictions. The five-offset self-consensus rises from 85.14% at 256 tokens to 90.27% at 1024, but the single pass already exceeds this at 98.34% (Figure 5).

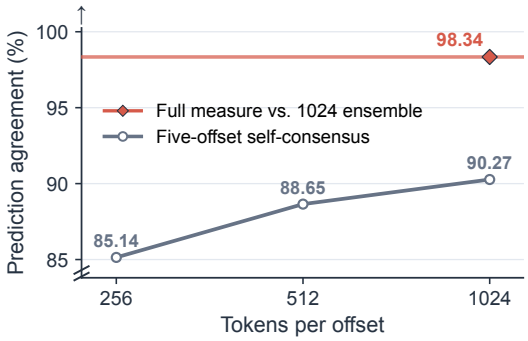


Figure 5: One deterministic pass replaces offset-sensitive resampling and streams on one GPU. Measure-attention (single pass) matches five-offset 1024-token ensemble.

For streaming (Algorithm 1), time is $O(n)$ per query with a single sequential pass, while peak token memory scales with the chunk size B plus the head state rather than with the scene. 212 objects are processed one at a time on a single GPU with 100% prediction preservation (max logit diff 0.00166), 18.48× less peak memory, and 0.576× latency (Figure 6). The measured consequences are reported in Appendix E.

Algorithm 1 Streaming measure-attention (one head, one query).

Require: chunked tokens $\{(h_i, s_i, \log m_i)\}_{i=1}^{n}$, chunk size B

$Z \leftarrow -\infty; \quad N \leftarrow 0; \quad D \leftarrow 0; \quad S_1 \leftarrow 0; \quad S_2 \leftarrow 0; \quad M \leftarrow 0$

for each chunk c **do**

  $u_c \leftarrow \max_{i \in c}(s_i + \log m_i); \quad \ell_{ci} \leftarrow s_i + \log m_i - u_c$

  $D_c \leftarrow \sum_{i \in c} e^{\ell_{ci}}; \quad N_c \leftarrow \sum_{i \in c} e^{\ell_{ci}} h_i$

  $Z^- \leftarrow Z; \quad Z \leftarrow \max(Z^-, u_c)$

  $D \leftarrow e^{Z^- - Z} D + e^{u_c - Z} D_c; \quad N \leftarrow e^{Z^- - Z} N + e^{u_c - Z} N_c$

  $S_1 \leftarrow S_1 + \sum_{i \in c} m_i h_i; \quad S_2 \leftarrow S_2 + \sum_{i \in c} m_i (h_i \odot h_i); \quad M \leftarrow M + \sum_{i \in c} m_i$

end for

return $A = N/D, \quad M_1 = S_1/M, \quad M_2 = S_2/M$

Deployment and Outlook. Measure-attention is a consumer-side change: no retraining, no new parameters, one softplus and one logarithm per token, and Algorithm 1 turns it into a single sequential pass whose peak memory scales with the chunk rather than the scene. Pipelines whose upstream cardinality varies: densification, compression, cache resampling, therefore inherit this stability without any coordination with the producer. As the derivation uses only an additive, conservable mass, the same argument yields invariant operators for any per-atom quantity with a compositing law, and the preregistered protocol transfers as a template for auditing refinement sensitivity wherever set-encoders consume explicit scene representations, such as segmentation, SLAM, and 3D generation.

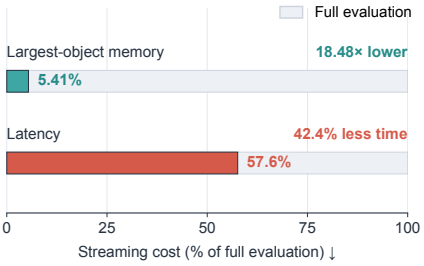


Figure 6: 18.48× memory reduction and 0.576× latency vs. full evaluation.

6 Conclusion

We identify refinement invariance as a representation symmetry beyond permutation and introduce Optical Measure Attention, grounding downstream 3DGS aggregation in compositing-derived optical mass. Exact split invariance and second-order stability guarantees translate into robust Gaussian refinement. Experiments demonstrate we enable downstream 3DGS learning invariant to how scene evidence is discretized, establishing a representation principle beyond permutation invariance.

Reproducibility Statement

Appendices provide complete proofs, architectures, training, data selection, interventions, metrics, and condition-level results. The supplementary artifact contains canonical aggregate JSON, deterministic figure generation, frozen protocol hashes, package verification, and an anonymization scan. Public data and checkpoint acquisition are documented separately because they are not redistributed in the compact artifact.

AI Use Statement

Generative AI tools were used to help debug evaluation scripts, organize literature, and editorial assistance. AI-assisted numerical statements were checked against stored result files and SHA-256 manifests, and generated analysis code was exercised by reproducibility checks.

References

Honggyu An, Jaewoo Jung, Mungyeom Kim, Chaehyun Kim, Minkyeong Jeon, Jisang Han, Kazumi Fukuda, Takuya Narihira, Hyuna Ko, Junsu Kim, et al. C3g: Learning compact 3d representations with 2k gaussians. *arXiv preprint arXiv:2512.04021*, 2025.

Milena T Bagdasarian, Paul Knoll, Yiming Li, Florian Barthel, Anna Hilsmann, Peter Eisert, and Wieland Morgenstern. 3dgs. zip: A survey on 3d gaussian splatting compression methods. In *Computer Graphics Forum*, volume 44, pp. e70078. Wiley Online Library, 2025.

Edoardo Calvello, Nikola B Kovachki, Matthew E Levine, and Andrew M Stuart. Continuum attention for neural operators. *Journal of Machine Learning Research*, 26(300):1–52, 2025.

Jiazhong Cen, Jiemin Fang, Chen Yang, Lingxi Xie, Xiaopeng Zhang, Wei Shen, and Qi Tian. Segment any 3d gaussians. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pp. 1971–1979, 2025.

Hanqing Chao, Yiwei He, Junping Zhang, and Jianfeng Feng. Gaitset: Regarding gait as a set for cross-view gait recognition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 8126–8133, 2019.

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

Gwendoline de Bie, Gabriel Peyré, and Marco Cuturi. Stochastic deep networks, 2019. URL https://arxiv.org/abs/1811.07429.

Emilien Dupont, Miguel Bautista Martin, Alex Colburn, Aditya Sankar, Josh Susskind, and Qi Shan. Equivariant neural rendering. In *International Conference on Machine Learning*, pp. 2761–2770. PMLR, 2020.

Runze Fan, Jian Wu, Xuehuai Shi, Lizhi Zhao, Qixiang Ma, and Lili Wang. Fov-gs: Foveated 3d gaussian splatting for dynamic scenes. *IEEE Transactions on Visualization and Computer Graphics*, 31(5):2975–2985, 2025.

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.

Julia Farganus, Weronika Jakubowska, et al. Learning representations from 3d gaussian splats. *arXiv preprint arXiv:2605.29549*, 2026.

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, 2024.

Mohamed Abdul Gafoor, Marius Preda, and Titus Zaharia. Refining gaussian splatting: A volumetric densification approach. *arXiv preprint arXiv:2508.05187*, 2025.

Chi-Shiang Gau, Konstantinos D Polyzos, Athanasios Bacharis, Saketh Madhuvarasu, and Tara Javidi. 3d scene rendering with multimodal gaussian splatting. In *2025 59th Asilomar Conference on Signals, Systems, and Computers*, pp. 1480–1485. IEEE, 2025.

Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3d point clouds: A survey. *arXiv preprint arXiv:1912.12033*, 2019.

Jiaqi Han, Jiacheng Cen, Liming Wu, Zongzhao Li, Xiangzhe Kong, Rui Jiao, Ziyang Yu, Tingyang Xu, Fandi Wu, Zihe Wang, et al. A survey of geometric graph neural networks: Data structures, models and applications. *Frontiers of Computer Science*, 19(11):1911375, 2025.

Yicong Hong, Kai Zhang, Jiuxiang Gu, Sai Bi, Yang Zhou, Difan Liu, Feng Liu, Kalyan Sunkavalli, Trung Bui, and Hao Tan. Lrm: Large reconstruction model for single image to 3d. In *International Conference on Learning Representations*, volume 2024, pp. 50678–50702, 2024.

Jim James, Benjamin Wilson, Simon Lucey, and James Hays. 3d gaussian point encoders. In *2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1788–1797. IEEE, 2026.

Siyu Jiao, Haoye Dong, Yuyang Yin, Zequn Jie, Yinlong Qian, Yao Zhao, Humphrey Shi, and Yunchao Wei. Clip-gs: Unifying vision-language representation with 3d gaussian splatting. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4670–4680. IEEE, 2025.

Minhas Kamal, Hiranya Garbha Kumar, and Balakrishnan Prabhakaran. A survey on deep learning architectures for point cloud classification and segmentation. *arXiv preprint arXiv:2605.17131*, 2026.

Ryotaro Kawata and Taiji Suzuki. Transformers as measure-theoretic associative memory: A statistical perspective and minimax optimality. *arXiv preprint arXiv:2602.01863*, 2026.

Nikhil Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian Scherer, Deva Ramanan, and Jonathon Luiten. Splatam: Splat, track & map 3d gaussians for dense rgb-d slam. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21357–21366. IEEE, 2024.

Bernhard Kerbl. The impact and outlook of 3d gaussian splatting. *arXiv preprint arXiv:2510.26694*, 2025.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, George Drettakis, et al. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.

Bernhard Kerbl, Andreas 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 (TOG)*, 43(4):1–15, 2024.

Injae Kim, Chaehyeon Kim, Minseong Bae, Minseok Joo, and Hyunwoo J Kim. F4splat: Feed-forward predictive densification for feed-forward 3d gaussian splatting. *arXiv preprint arXiv:2603.21304*, 2026.

Patrick T Komiske, Eric M Metodiev, and Jesse Thaler. Energy flow networks: deep sets for particle jets. *Journal of High Energy Physics*, 2019(1):1–46, 2019.

Risi Kondor. The principles behind equivariant neural networks for physics and chemistry. *Proceedings of the National Academy of Sciences*, 122(41):e2415656122, 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.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In *International conference on machine learning*, pp. 3744–3753. PMLR, 2019.

Yue Li, Qi Ma, Runyi Yang, Huapeng Li, Mengjiao Ma, Bin Ren, Nikola Popovic, Nicu Sebe, Ender Konukoglu, Theo Gevers, et al. Scenesplat: Gaussian splatting-based scene understanding with vision-language pretraining. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4961–4972. IEEE, 2025.

Zhihao Liang, Hongdong Li, Kui Jia, Kailing Guo, and Qi Zhang. Gus-ir: Gaussian splatting with unified shading for inverse rendering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(10):8364–8378, 2025.

Haitian Liu, Binglin Liu, Qianchao Hu, Peilun Du, Jing Li, Yang Bao, and Feng Wang. A review on 3d gaussian splatting for sparse view reconstruction. *Artificial Intelligence Review*, 58(7):215, 2025a.

Xingchen Liu, Piyush Tayal, Jianyuan Wang, Jesus Zarzar, Tom Monnier, Konstantinos Tertikas, Jiali Duan, Antoine Toisoul, Jason Y Zhang, Natalia Neverova, et al. Uncommon objects in 3d. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14102–14113. IEEE, 2025b.

Hidenobu Matsuki, Riku Murai, Paul HJ Kelly, and Andrew J Davison. Gaussian splatting slam. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18039–18048. IEEE, 2024.

Chuizheng Meng, Sam Griesemer, Defu Cao, Sungyong Seo, and Yan Liu. When physics meets machine learning: A survey of physics-informed machine learning. *Machine Learning for Computational Science and Engineering*, 1(1):20, 2025.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022.

Simon Niedermayr, Josef Stumpfegger, and Rüdiger Westermann. Compressed 3d gaussian splatting for accelerated novel view synthesis. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10349–10358. IEEE, 2024.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017.

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

Vikram Sandu, Mayurdeep Pathak, and Rajiv Soundararajan. Temporally aware densification for dynamic 3d gaussian splatting. *arXiv preprint arXiv:2606.23212*, 2026.

Sushmita Sarker, Prithul Sarker, Gunner Stone, Ryan Gorman, Alireza Tavakkoli, George Bebis, and Javad Sattarvand. A comprehensive overview of deep learning techniques for 3d point cloud classification and semantic segmentation: S. sarker et al. *Machine Vision and Applications*, 35 (4):67, 2024.

Stanislaw Szymanowicz, Chrisitian Rupprecht, and Andrea Vedaldi. Splatter image: Ultra-fast single-view 3d reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10208–10217, 2024.

James Vuckovic, Aristide Baratin, and Remi Tachet des Combes. A mathematical theory of attention. *arXiv preprint arXiv:2007.02876*, 2020.

Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, 38(5): 1–12, 2019.

Peixi Wu, Bosong Chai, Hebei Li, Menghua Zheng, Yansong Peng, Zeyu Wang, Xuan Nie, Yueyi Zhang, and Xiaoyan Sun. Spiking point transformer for point cloud classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 21563–21571, 2025.

Tong Wu, Yu-Jie Yuan, Ling-Xiao Zhang, Jie Yang, Yan-Pei Cao, Ling-Qi Yan, and Lin Gao. Recent advances in 3d gaussian splatting. *Computational Visual Media*, 10(4):613–642, 2024.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1912–1920, 2015.

Yuelin Xin, Yuheng Liu, Xiaohui Xie, and Xinke Li. Learning unified representation of 3d gaussian splatting. In *International Conference on Learning Representations*, volume 2026, pp. 1111–1123, 2026.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

Chi Yan, Delin Qu, Dan Xu, Bin Zhao, Zhigang Wang, Dong Wang, and Xuelong Li. Gs-slam: Dense visual slam with 3d gaussian splatting. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19595–19604. IEEE, 2024.

Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20331–20341. IEEE, 2024.

Rose Yu and Rui Wang. Learning dynamical systems from data: An introduction to physics-guided deep learning. *Proceedings of the National Academy of Sciences*, 121(27):e2311808121, 2024.

Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-splatting: Alias-free 3d gaussian splatting. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19447–19456. IEEE, 2024.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.

Ruiqi Zhang, Hao Zhu, Jingyi Zhao, Qi Zhang, Xun Cao, and Zhan Ma. Mitigating ambiguities in 3d classification with gaussian splatting. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 27275–27284. IEEE, 2025a.

Tao Zhang, Haobo Yuan, Lu Qi, Jiangning Zhang, Qianyu Zhou, Shunping Ji, Shuicheng Yan, and Xiangtai Li. Point cloud mamba: Point cloud learning via state space model. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pp. 10121–10130, 2025b.

Annan Zhou, Li Wang, Jian Li, Jing Huang, Li Li, and Jian Yao. Sggs: Semantic-guided 3d gaussian splatting with adaptive rendering. *IEEE Transactions on Visualization and Computer Graphics*, 2026a.

Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI open*, 1:57–81, 2020.

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, 2026b.

A Proof of Theorem 2

This appendix quantifies every constant left abstract in the main text and derives the ratio perturbation bound equation 7.

A.1 Norm conventions

All second-derivative norms in this appendix are spectral norms evaluated along the straight segment joining the parent atom (x, C) to each child $(x + \delta_j, C - B)$, $t \in [0, 1]$. For a scalar map $g(x, C, c)$, the symbol L_C denotes the spectral norm of the partial derivative of g with respect to its covariance argument, and H_g the spectral norm of the full Hessian with respect to (x, C) jointly. Feature-space norms (L_f, L_s, $\|f'\|$, $\|s'\|$ below) are the corresponding spectral norms of the derivatives of the feature map φ and score map s. With these conventions every constant in the theorem statement is fully quantified.

A.2 Second-order expansions

Setup. Consider a single atom of mass τ replaced by K children of mass τ/K with displacement set $\{\delta_j\}$ satisfying the moment-preservation conditions equation 6: $\sum_j \delta_j = 0$ and $B = \frac{1}{K}\sum_j \delta_j \delta_j^\top$. The attention denominator before refinement is $D = \sum_{i \neq k} m_i e^{s_i} + m\, e^{s(x,C)}$; after refinement

$$\widetilde{D} \;=\; \sum_{i \neq k} m_i e^{s_i} \;+\; \frac{m}{K} \sum_{j=1}^{K} e^{\,s(x+\delta_j,\, C-B)}.$$

Expanding each child term to second order about (x, C),

$$e^{\,s(x+\delta_j, C-B)} \;=\; e^{\,s(x,C)} \Big(1 + s_x^\top \delta_j - \langle \nabla_C s, B\rangle + \tfrac{1}{2}\delta_j^\top H_{xx}\, \delta_j\Big) + O(\rho_j^3),$$

where s_x is the gradient of s in x, $\nabla_C s$ the gradient in C, H_{xx} the Hessian block in x, and $\rho_j = \|\delta_j\| + \|B\|$. Summing over j, the first-order position terms cancel by $\sum_j \delta_j = 0$, and $\sum_j B = KB$, giving

$$\widetilde{D} \;=\; D \;+\; m\, e^{\,s(x,C)} \Big(\tfrac{1}{2}\,\mathrm{tr}\big(H_{xx}\, B\big) - \langle \nabla_C s, B\rangle\Big) \;+\; O(m\,\rho^3), \tag{8}$$

with $\rho = \max_j \rho_j$. The bracketed correction is at most $L_C\|B\| + \frac{1}{2}H_s\|B\|$ in spectral norm, hence $O(\|B\|)$, and $\|B\| \le r^2$ for maximum child displacement r.

Numerator. The same expansion applied to $g(z) = e^{\,s(\varphi(z))}\varphi(z)$ yields, per feature component f,

$$\widetilde{N}_f \;=\; N_f \;+\; \frac{m\, e^{s(x,C)}}{2}\; \delta^\top \Big(H_f + f_x s_x^\top + s_x f_x^\top\Big)\delta \;\Big|_{B\text{-term}} \;+\; O(m\,\rho^3),$$

where the cross terms $f_x^\top \delta_j$ and $s_x^\top \delta_j$ cancel in the mass-weighted sum, leaving a second-order residual bounded in spectral norm by $\frac{1}{2}\, m\, e^{s(x,C)}\, (L_f + 2\|f'\|\|s'\|)\, \|B\|$.

A.3 The ratio bound

Let $N_r(\mu) = \sum_i m_i e^{s_r(\varphi(z_i))}\varphi(z_i)$ and $D_r(\mu) = \sum_i m_i e^{s_r(\varphi(z_i))}$, so that $A_r = N_r/D_r$. Write $\epsilon_N = \|\widetilde{N}_r - N_r\|$ and $\epsilon_D = |\widetilde{D}_r - D_r|$, and suppose $D_r, \widetilde{D}_r \ge \beta > 0$. Then

$$\|A_r(\widetilde{\mu}) - A_r(\mu)\| \;\le\; \frac{\|\widetilde{N}_r - N_r\|}{|\widetilde{D}_r|} \;+\; \|N_r\| \left|\frac{1}{\widetilde{D}_r} - \frac{1}{D_r}\right| \;\le\; \frac{\epsilon_N}{\beta} \;+\; \frac{\|N_r(\mu)\|\,\epsilon_D}{\beta^2},$$

which is exactly Equation 7. Both ϵ_N and ϵ_D are $O(r^2)$ by the expansions above with $\|B\| \le r^2$, so the pooled output is second-order stable with

$$C \;=\; \frac{\max_r \big(\epsilon_N/\beta + \|N_r\|\epsilon_D/\beta^2\big)}{r^2},$$

a constant depending only on L_f, L_s, L_C, H_g, the denominator floor β, and the mass distribution.

A.4 Moments and prediction preservation

The moments M_1 and M_2 satisfy the same bound because their numerators are mass-weighted sums of $g(z) = \varphi(z)$ and $g(z) = \varphi(z) \odot \varphi(z)$, both twice differentiable with bounded Hessians under the same conventions, and their common denominator $\sum_i m_i$ is preserved exactly by the mass constraint. If the classifier head ρ is Lipschitz with constant L_ρ on the concatenation $[A_1, \ldots, A_H, M_1, M_2]$, the induced logit perturbation is bounded by L_ρ times the feature-space bound; a native top-1 margin exceeding twice this uniform logit perturbation therefore prevents a prediction flip. This is the formal content of the prediction-stability metric used in Section 5. □

B Experimental Protocol and Preregistration

B.1 Data selection and splits

All manifests were frozen before any factor-8 or natural-compression result was observed. uCO3D uses the official four-class split with 278 training and 70 test objects; the 210 official test views are never resplit. ModelNet10 uses the official bathtub/desk/dresser split with 506 training and 222 test objects. MACGS contributes 370 internal test objects that are not redistributed. The evaluation corpus therefore contains 662 test objects (292 external, 370 internal). Training reads only the frozen 1,024-token reference caches; native test Gaussians are used exclusively for evaluation, and the semantic gallery for retrieval is the 278 training reference caches.

B.2 Frozen training configuration

The metric-learning encoder of Section 5.5 uses a pointwise Conv1d trunk ($12 \to 128 \to 256 \to 256$), pooling by eight scalar attention heads plus first and second moments, and an embedding trunk ($512 \to 256 \to 128$, ℓ_2-normalized) with an auxiliary class head at 256 dimensions. Training is instance-contrastive: two independently sampled token subsets ($256/512$ tokens, alternating epochs) form the positive pair, other batch objects are negatives, and the loss is NT-Xent at temperature 0.10 plus $0.25\times$ the auxiliary cross-entropy of both views. Optimization is AdamW (learning rate 8×10^{-4}, weight decay 5×10^{-4}), cosine schedule, batch 24, 120 epochs. Seed 11 is the pilot; no hyperparameter change is permitted after the pilot, and seeds 23 and 37 are exact replications.

B.3 Preregistered gate families

Seven gate families were frozen before the pilot:

- native semantic Recall@1 exceeds the dataset majority rate by at least 15 pp on both datasets;
- Measure identity Recall@1 $\geq 95\%$ per dataset and $\geq 97\%$ pooled;
- pooled identity Recall@1 gain over ordinary ≥ 3 pp with a paired object-bootstrap 95% interval strictly above zero;
- Measure semantic mAP loses at most 1 pp from native to factor-8 on each dataset;
- Measure top-10 retention exceeds ordinary by at least 10 pp per dataset;
- median Measure embedding drift at least $3\times$ lower than ordinary per dataset;
- native Measure semantic Recall@1 within 2 pp of native ordinary on either dataset.

Instantiated across datasets and seeds, the ledger records these as 27 frozen threshold checks, all of which pass. The natural-compression extension carries its own preregistration: six gate families over the three primary mass-conserving conditions, with pruning conditions declared outside the mass-conserving claim before evaluation (Section E).

B.4 Mass-law ledger

Only optical depth, the mass law derived from alpha compositing, attains exact co-located invariance; the two alpha-based laws approach but do not reach it because α is not additive under compositing.

Table 5: Mass-law ablation (%) on factor-8 refinement (292 objects, Conv).

Mass law	Stability ↑	Drift ↓	Flip gap [95% CI]
Uniform	94.18	15.80	5.82 [3.42, 8.56]
Raw α_i	99.32	1.08	0.69 [0.00, 1.71]
Clipped α	98.63	5.86	1.37 [0.34, 2.74]
Optical depth τ_i	**100.00**	**0.52**	—

Table 6: Median logit drift (%) and stability (%) versus displacement multiplier λ (two-decimal display, matching Table 1).

λ	Ord. stab. ↑	Meas. stab. ↑	Ord. drift ↓	Meas. drift ↓
0.00	94.18	**100.00**	13.86	**0.00**
0.25	94.18	**100.00**	13.96	**0.13**
0.50	94.18	**100.00**	15.80	**0.52**
0.75	94.86	**98.97**	15.89	**1.20**
0.95	92.81	**99.32**	17.38	**2.08**

B.5 Refinement phase

The phase table locates the two mechanisms cleanly: ordinary-attention drift is dominated by the multiplicity change (13.86–17.38% across the whole sweep), whereas measure-attention drift is exactly zero at $\lambda = 0$ and grows second-order in the displacement. Measure-attention stability also dominates ordinary at every λ (98.97–100.00% vs. 92.81–94.18%), so no displacement setting reverses the conclusion.

B.6 Official renderer verification

On the 70 uCO3D objects with 210 official views, the renderer achieves median RGB PSNR of 47.93 dB (fifth percentile 37.69 dB), median SSIM of 0.999 (fifth percentile 0.993), and 69 of 70 objects exceed 28 dB, confirming that the natural compression transformations preserve visual fidelity (single-object excursion in Section E).

C Native-Accuracy Reference

The main text deliberately reports refinement robustness rather than native accuracy, because ordinary and Measure Attention share checkpoints and differ only in the integration rule. For completeness, the frozen native (un-refined) reference agreement and accuracy across the three pipelines are recorded here. They anchor usability of the trained models and delimit the claim: the cross-pipeline accuracy contribution is a native-usage effect that is MACGS-driven, not a universal accuracy boost.

The pooled agreement gain of +4.23 pp decomposes as +7.30 pp on MACGS and +0.34 pp on the two external pipelines combined (paired-interval $[-1.03, 1.71]$): on uCO3D and ModelNet10 the integration law is accuracy-neutral by construction, which is precisely the regime where a robustness claim must be separated from an accuracy claim. The uCO3D and ModelNet10 agreement CIs cross zero for the same reason. ModelNet10 is additionally saturated in reference agreement (99.10% for both methods), leaving no headroom by design; its native accuracy is 0.90 pp below ordinary (98.20% vs. 99.10%, two of 222 objects), an endpoint we retain for completeness.

D Matched Training Controls

A natural concern is that the optical-depth mass simply provides an additional input feature that the model can exploit, rather than conferring genuine refinement invariance. We test this with two matched controls on both uCO3D and ModelNet10: (1) *mass-as-content*, where τ_i is concatenated to the feature vector h_i but attention remains uniform; and (2) *learned-positive-weight*, where a learnable per-token positive scalar replaces τ_i in the attention weight. Both controls use the same

Table 7: Native reference agreement and accuracy across three 3DGS pipelines.

Pipeline	Reference agreement (%) ↑		Gain (pp; 95% CI)	Accuracy (%) ↑	
	Ord.	Meas.		Ord.	Meas.
MACGS	90.00	**97.30**	+7.30 [4.32, 10.27]	85.14	87.03
uCO3D	95.71	**97.14**	+1.43 [−2.86, 5.71]	74.29	77.14
ModelNet10	99.10	99.10	0.00 [−1.35, 1.35]	99.10	98.20
Pooled (662)	—	—	**+4.23** [2.57, 6.04]	—	—
External only (292)	—	—	+0.34 [−1.03, 1.71]	—	—

Table 8: Matched training controls: factor-8 refinement stability (%) and drift (%).

Dataset	Method	Native acc. ↑	F8 stab. ↑	F8 acc. ↑	Drift ↓
uCO3D	Mass-as-content	74.29	92.86	67.14	20.65
	Learned-pos.-wt	72.86	91.43	67.14	14.83
	Measure-attn	**77.14**	**100.00**	**74.29**	**0.52**
ModelNet10	Mass-as-content	91.89	73.42	65.32	4.29
	Learned-pos.-wt	93.24	71.17	64.41	6.58
	Measure-attn	**98.20**	**100.00**	**98.20**	**0.53**

training protocol as measure-attention, so any stability difference is attributable to where the mass enters the operator.

Table 8 shows that neither control matches measure-attention's refinement stability. On uCO3D, mass-as-content achieves 92.86% factor-8 stability (vs. 100% for measure-attention) with 20.65% median drift (vs. 0.52%). On ModelNet10, the gap is wider: 73.42% vs. 100% stability and 4.29% vs. 0.53% drift. The learned-positive-weight control performs similarly or worse, confirming that the invariance is a structural property of the optical-depth mass as the base measure of the operator, not merely an additional input signal.

The controls also indicate that the invariance is not bought with clean-accuracy debt: on uCO3D the measure-attention model has the best native accuracy in its row (77.14%), and on ModelNet10 it is within 0.90 pp of the strongest control (93.24%). The mass-as-content control is the clearest negative: adding τ_i to the feature vector without changing the base measure leaves the operator fully multiplicity-sensitive, and its 20.65% drift on uCO3D is even worse than uniform weighting's pooled 15.80%.

E Boundary Conditions and Open Observations

We separate three kinds of boundary: *scope conditions* (regimes the invariance claim does not cover), *metric limitations* (endpoints whose discriminative power is exhausted), and *numerical artifacts* (finite-precision reduction order). For calibration, every preregistered primary gate of the headline packages passed: moment-preserving split 9/9 (Conv) and 12/12 (Query), factor-8 retrieval 27/27, and the five-offset comparison 16/16. None of the boundaries below affects the primary refinement-robustness claim; each is retained with its frozen gate status in Table 9 and each suggests a concrete extension.

E.1 Identity-endpoint saturation under natural compression

The natural-compression preregistration inherited its floors from the factor-8 stress test: an identity Recall@1 gain of at least 2 pp over the three primary conditions, a top-10 retention gain of at least 10 pp, and a median-drift reduction of at least 3×. Under real merge/pruning, identity Recall@1 saturates at 100% for *both* methods, so the 2 pp gain gate is un-meetable by construction, a property of the endpoint, not of either operator; the preregistered identity floor ($\geq$ 97%) itself passes. The preregistered drift floor is met on one of the three primary conditions (merge 75%: 3.67×) and missed on merge 50% (1.60×) and mixed 50% (2.68×), the two mildest transforma-

tions. Every measured continuous effect is directionally positive with strictly positive CIs (top-10 retention +3.52 to +8.90 pp; drift reduction $1.60\times$ to $16.37\times$), and the renderer gate attached to the same preregistration (median PSNR $\geq$ 45 dB, p05 $\geq$ 35 dB on the primary conditions) passes. The preregistration practice takeaway is calibration, not revision: continuous endpoints (retention, drift) should be primary whenever an identity endpoint is liable to saturate.

E.2 Opacity pruning is a scope condition, not a counterexample

Opacity pruning deliberately removes optical mass and is therefore outside the exact mass-conserving invariance; the preregistration classed it as a secondary robustness test before evaluation, reported without post-hoc condition selection. Its behavior is nonetheless favorable in every cell: in all 9/9 pruning seed–condition cells measure-attention has lower embedding drift and identity never worse than ordinary attention. A natural extension is a quantitative lossy-mass bound, relating fractional mass removal to a multiplicative logit-drift budget, which our split algebra does not yet cover.

E.3 A single background excursion in the renderer audit

The official renderer test passes 69/70 objects above 28 dB (median 47.93 dB, p05 37.69 dB, SSIM 0.999/0.993). The retained excursion is one bouquet object with 26.04 dB median PSNR, and the untruncated-background alpha-occupancy health bound fails on the same object. The diagnosis is scene-level: a low-background object whose alpha occupancy is dominated by non-target background, which the current health bound does not separate from representation damage. A background-truncation or opacity-prior pre-stage before health gating is the natural follow-up; we did not tune the bound post hoc.

E.4 Finite-precision reduction order in streaming

Streaming (Algorithm 1) computes the same finite measure with chunked, online log-sum-exp accumulation. Prediction-level gates pass on all 212 streamed objects (100% prediction preservation, maximum logit difference 0.00166, $18.48\times$ lower largest-object memory, $0.576\times$ latency). The strict 10^{-4} logit-reproduction gate is not met because chunked and full-materialization evaluations use different finite-precision reduction orders.

E.5 One seed-level excursion under a strict margin

The three-seed classification replication is recorded as 9/10: seed 23 loses one additional uCO3D refined prediction relative to the pre-declared one-point noninferiority margin. The pooled hierarchical interval $[5.48, 11.42]$ pp excludes zero, and all six seed–dataset cells are direction-consistent, so the mechanism replicates; the excursion is a single-margin event on one object. Margin-adaptive seed aggregation, weighting seeds by their native margin distributions, is the natural extension.

F Frozen Evidence Ledger

Table 9 records every endpoint with its frozen gate status. Failed gates remain failures and are not retuned; "partial" denotes directionally consistent evidence under a strict preregistered floor, and "deployment" denotes prediction-level pass with a strict numerical gate not met. Interpretations are given in Sections C–E.

G Reproducibility and Anonymization

The supplementary artifact contains the following, all hash-verified by `MANIFEST_SHA256.json`:

- `repro/measure_core.py`: NumPy reference for optical-measure pooling, split algebra, and moment-preserving child generation.

Table 9: Every endpoint retained with its frozen gate status.

Endpoint	Gates	Status	Frozen interpretation
Native MACGS	4/4	pass	Agreement 90.00→97.30; accuracy CI crosses zero.
Native uCO3D	5/7	partial	Accuracy-neutral by design; agreement CI crosses zero.
Native ModelNet10	5/7	partial	Agreement saturated at 99.10 for both methods.
Five-offset ensemble	16/16	pass	98.34% agreement; accuracy CI $[-1.21, 0.76]$.
Identical split, Conv	8/8	pass	Exact-algebra sanity check; 100% Measure stability.
Moment split, Conv	9/9	pass	Primary result: 100.00 vs. 94.18 stability.
Moment split, Query	12/12	pass	Architecture transfer: 99.66 vs. 91.44.
Query, three seeds	9/10	partial	Mechanism replicates; seed-23 margin excursion.
Factor-8 retrieval	27/27	pass	Identity R@1 100.00 vs. 94.98%; retention 96% vs. 75–79%.
Natural-compr. retrieval	0/3	partial	Identity saturates 100/100; all continuous CIs positive.
Official renderer	6/8	partial	69/70 objects pass; one background excursion.
Streaming	10/15	deployment	100% prediction preservation; strict numerical logit gates unmet.

- `repro/test_measure_core.py`: self-contained tests verifying exact split invariance (rtol $= 10^{-12}$), optical-depth additivity, and moment preservation.
- `repro/export_tables.py` and `repro/build_figures.py`: regenerate all paper tables and figures from the canonical ledger.
- `results/all_results.json`: complete frozen result ledger (dated 2026-08-26), plus the preregistration records and preregistered gate definitions of both retrieval extensions (2026-09-02).
- Per-seed and per-condition aggregate JSON for both retrieval extensions, with preregistered thresholds and descriptive bootstrap intervals kept separate.

Public data and checkpoint acquisition are documented separately because they are not redistributed in the compact artifact. An anonymization scan verifies that the released text, figures, and bibliography carry no identifying metadata; all citation entries use the venue's pseudonymized author convention, and the mapping to authorship is held by the review institution. Deterministic figure generation means every number printed in this paper can be regenerated from the ledger JSON alone, without access to training infrastructure.