Decoupled Reprojection Consistency for Diagnosing 3D Gaussian Splatting Failures

Jin-Hyeong Park

Independent Researcher

Abstract

*This paper introduces **Decoupled Reprojection Consistency (DRC)**, a training-free diagnostic for 3D Gaussian Splatting (3DGS) that decomposes an ambiguous reprojection error into three interpretable maps: a geometry/visibility mismatch E_{depth}, a base photometric inconsistency E_{base} computed under SH truncation (L=0), and a view-dependent residual E_{vd}. By comparing cross-view reprojection errors from zeroth-order (*base*) and trained-degree (*full*) SH renders of the same model, DRC separates geometry/visibility-driven inconsistency from appearance-driven instability. Paired with a bivariate fingerprint (quadrant occupancy) and coverage reporting, DRC turns reprojection consistency into actionable per-pixel triage. The method is demonstrated on synthetic and real scenes, separating glossy regions from geometry failures in our examples.*

CCS Concepts

• ***Computing methodologies*** → ***Rendering;*** *Image-based rendering;*

1. Introduction

Reprojection consistency—warping one view into another and measuring photometric agreement—is a standard sanity check for radiance-field reconstructions [WBF*17]. However, a single error map is ambiguous: elevated error can stem from geometry/visibility failures (depth errors, occlusions, floaters) *or* from view-dependent appearance that does not warp well (glossy highlights, reflections). In 3DGS [KKLD23], view dependence is encoded via per-Gaussian spherical harmonics (SH), so reflective regions may look plausible per view yet become inconsistent under cross-view warping. Existing training-free consistency tools [STGP26, GRS*24] typically produce a single scalar map per view pair—useful for ranking but ambiguous for debugging.

A simple internal ablation addresses this: rendering the trained model with SH truncated to L=0 (*base*) versus the trained SH degree (*full*), and comparing reprojection inconsistencies under both renders to obtain a cause-separating diagnostic axis. Note that L=0 is used solely as an internal basis truncation for diagnostics, not as a physical diffuse or albedo decomposition. Concretely, the contributions are: (1) a training-free tri-map decomposition into E_{depth} (geometry/visibility), E_{base} (SH0/base), and E_{vd} (view-dependent residual); (2) coverage-aware reporting with a bivariate fingerprint and quadrant labeling for actionable triage; and (3) demonstrations on synthetic and real scenes from public datasets.

2. Method

Inputs. A trained 3DGS model, calibrated cameras $\{K_i, R_i, t_i\}$, and per-view buffers $\{I_i^{\text{base}}, I_i^{\text{full}}, D_i, A_i\}$. The base render uses SH truncation to the DC term only (L=0); the full render uses the model's trained SH degree. In 3DGS [KKLD23], outgoing radiance for direction $\mathbf{d}$ is $\mathbf{c}(\mathbf{d}) = \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} \mathbf{c}_{\ell m} Y_\ell^m(\mathbf{d})$.

Neighbor selection and warping. For each target view i, neighbors $j \in \mathcal{N}(i)$ are selected via camera-center kNN. Pixels in i are back-projected using D_i to 3D, reprojected into each j, and bilinearly sampled from I_j^{base} and I_j^{full} to obtain warped images $\tilde{I}_j^{\text{base}\to i}$ and $\tilde{I}_j^{\text{full}\to i}$.

Masks and coverage. Two validity masks are used per pair (i, j): (a) M^{geo}: in-bounds $\wedge$ alpha gating—used for E_{depth}; and (b) M^{photo}: $M^{\text{geo}} \wedge$ depth-consistency—used for E_{base} and E_{full}, where photometric comparison is meaningful only when geometry agrees.

Photometric coverage C_i^{photo}, the fraction of pixels that have at least one valid photometric neighbor, is reported alongside every result.

Error maps. Let $\rho(\cdot,\cdot)$ be the L1 photometric error. For each target view i, inconsistencies across neighbors are robustly aggregated (median) on valid pixels:

- $E_{\text{depth}}^{(i)}(x)$: median relative depth mismatch $\left|D_j(\tilde{x}) - z^{\text{pred}}\right| / (z^{\text{pred}} + \epsilon)$ across j on M^{geo}.
- $E_{\text{base}}^{(i)}(x)$ and $E_{\text{full}}^{(i)}(x)$: median $\rho\left(I_i^{\text{base/full}}(x), \tilde{I}_j^{\text{base/full}\to i}(x)\right)$ across j on M^{photo}.
- $E_{\text{vd}}^{(i)}(x) = \max\left(0, E_{\text{full}}^{(i)}(x) - E_{\text{base}}^{(i)}(x)\right)$.

Normalization and quadrant labels. E_{base} and E_{full} are normalized by a shared per-view scale $s^{(i)} = q_{90}\left(\{E_{\text{base}}^{(i)}\} \cup \{E_{\text{full}}^{(i)}\}\right) + \epsilon$,

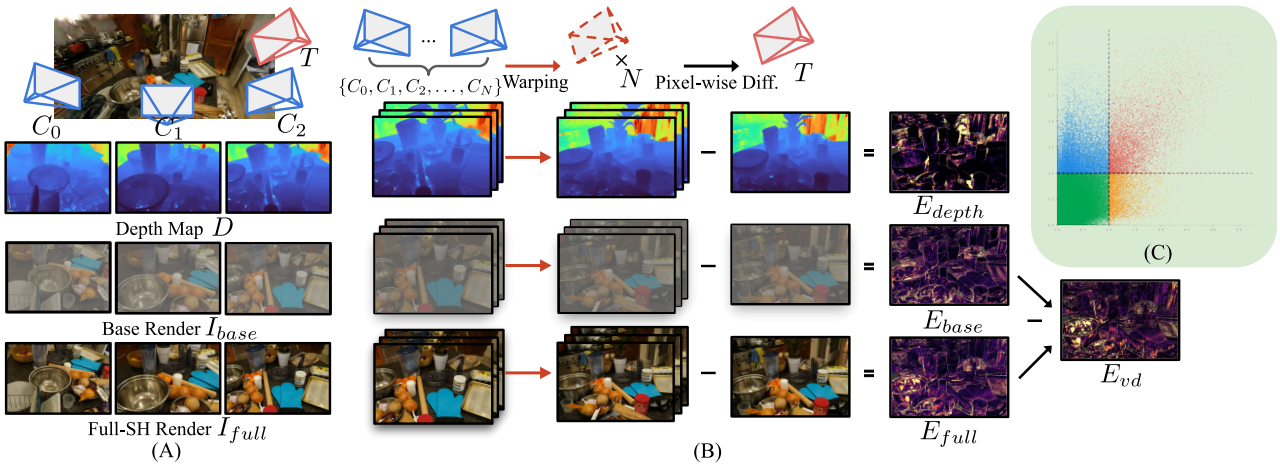


Figure 1: ***DRC pipeline.*** ***(A)*** *For target view T and context views $\{C_j\}$, render depth D, base color I_{base} (L=0), and full color I_{full} (trained SH degree).* ***(B)*** *Backward-warp context views into T using D_T; compute E_{depth}, E_{base}, E_{full}, and the residual $E_{\text{vd}}=\max(0, E_{\text{full}}-E_{\text{base}})$.* ***(C)*** *Per-view fingerprint: normalized $(E_{\text{base}}, E_{\text{full}})$ scatter with quadrant partitioning.*

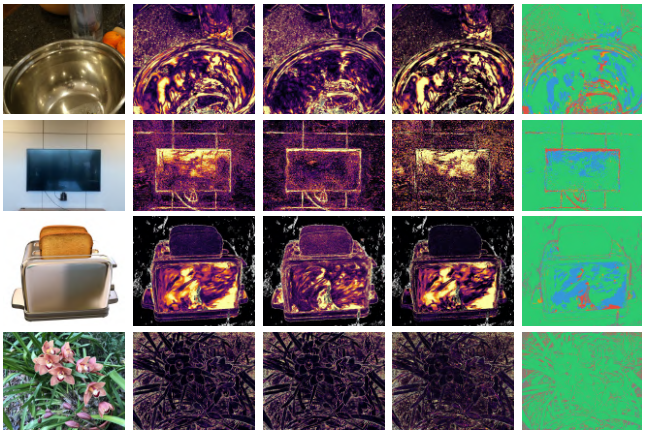


Figure 2: ***Qualitative triage.*** *Rows: Mip-NeRF 360* COUNTER *(stainless bowl ROI), LLFF* ROOM *(TV ROI), NeRF-Synthetic* TOASTER*, LLFF* ORCHIDS*. Columns: (a) RGB, (b) E_{full}, (c) E_{base}, (d) E_{vd}, (e) quadrant labels. Gray pixels are excluded by masking.*

computed on valid pixels, and define $E_{\bullet,n} = E_\bullet / s^{(i)}$. With threshold t=1, each valid pixel is assigned to one quadrant: *Consistent* (both low), *Geo/Vis-like* (both high), *Appearance-driven* ($E_{\text{base},n} \le t < E_{\text{full},n}$), or *Debug/Unexpected* ($E_{\text{base},n} > t \ge E_{\text{full},n}$).

3. Experiments

Experiments use Mip-NeRF 360 COUNTER, LLFF ROOM and ORCHIDS, and NeRF-Synthetic TOASTER. Fig. 2 shows that the single-map E_{full} conflates photometric mismatch and view-dependent effects: on COUNTER, marble elevates both E_{full} and E_{base}, while the stainless bowl is most prominent in E_{vd}; on ROOM and TOASTER, glossy surfaces (TV, toaster body) are labeled appearance-driven while edges fall into the Geo/Vis-like regime; on the matte ORCHIDS, appearance-driven pixels are sparse. Table 1 confirms high coverage ($\bar{C}_{\text{photo}} \approx 0.85$–$0.97$) and that the appearance-driven fraction is consistent with scene glossiness.

Scene (N)	$\bar{C}_{\text{photo}}$	Cons.	Geo	App	Dbg
M360/counter (216)	0.947	85.8	5.8	6.1	2.2
LLFF/room (37)	0.967	85.1	5.0	8.1	1.8
Syn/toaster (100)	0.854	85.0	5.0	8.2	1.8
LLFF/orchids (23)	0.852	87.7	7.7	3.1	1.4

Table 1: ***Coverage and quadrant occupancy on full target views.*** *$\bar{C}_{\text{photo}}$: mean photometric coverage. Cons./Geo/App/Dbg: quadrant occupancy (%) on valid pixels with t=1.*

4. Limitations and Takeaway

DRC is a diagnostic heuristic, not a ground-truth classifier: perfect mirrors and transparent glass can confound depth-based warping, and L=0 truncation is not a physical intrinsic decomposition. Low coverage from extreme pose gaps yields sparse maps; therefore, C^{photo} is always reported. Despite these caveats, DRC turns an ambiguous single error map into a training-free, coverage-aware tri-map diagnostic that separates geometry/visibility failures from view-dependent instability for practical 3DGS debugging.

References

[GRS*24] GOLI L., READING C., SELLÁN S., JACOBSON A., TAGLIASACCHI A.: Bayes' Rays: Uncertainty quantification for neural radiance fields. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2024), pp. 20061–20070. doi:10.1109/CVPR52733.2024.01896. 1

[KKLD23] KERBL B., KOPANAS G., LEIMKÜHLER T., DRETTAKIS G.: 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (Proc. SIGGRAPH) 42*, 4 (2023). doi:10.1145/3592433. 1

[STGP26] SAFADOUST S., TOSI F., GÜNEY F., POGGI M.: WarpRF: Multi-view consistency for training-free uncertainty quantification and applications in radiance fields. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (March 2026), pp. 5226–5235. 1

[WBF*17] WAECHTER M., BELJAN M., FUHRMANN S., MOEHRLE N., KOPF J., GOESELE M.: Virtual rephotography: Novel view prediction error for 3D reconstruction. *ACM Transactions on Graphics 36*, 4 (2017). doi:10.1145/3072959.2999533. 1

Decoupled Reprojection Consistency for Diagnosing 3D Gaussian Splatting Failures

Supplementary Material

Jin-Hyeong Park
Independent Researcher

This document provides supplementary material for the accompanying paper. Sec. A defines warping, masks, coverage, normalization, and quadrants for reproducibility. Sec. B reports extended quantitative summaries (coverage tails and error statistics). Sec. C provides additional qualitative galleries to reduce cherry-picking concerns, and Sec. D summarizes sanity checks and limitations.

A Definitions and Implementation Details

A.1 Warping notation

Let i denote a target view and $j \in \mathcal{N}(i)$ a neighboring view. For a pixel $x = (u, v)$ in view i with depth $D_i(x)$, the point is back-projected to 3D, transformed to world coordinates, and projected into view j. The projected coordinate is denoted by $\tilde{x}$, and neighbor images are bilinearly sampled at $\tilde{x}$ to obtain warped values.

Neighbor selection. For each target view i, the neighbor set $\mathcal{N}(i)$ consists of the k=6 nearest context views selected by camera-center kNN, excluding the target view itself.

Predicted neighbor depth. Let z^{pred} denote the depth of the back-projected 3D point expressed in the coordinate system of view j. Depth-consistency compares z^{pred} to the rendered depth $D_j(\tilde{x})$.

A.2 Masks and photometric coverage

Following the main paper, two validity masks are used per pair (i, j):

- **Geometry mask** $M^{i,j}_{\text{geo}}(x)$: in-bounds $\wedge$ alpha gate.
- **Photometric mask** $M^{i,j}_{\text{photo}}(x)$: $M^{i,j}_{\text{geo}}(x) \wedge$ depth-consistency.

Alpha gate. Let $A_i(x)$ denote the alpha/accumulation value at pixel x. Pixels with $A_i(x) > 0.05$ are retained. The same threshold is applied to sampled neighbor alpha.

Depth-consistency. A relative test is used:

$$\frac{|D_j(\tilde{x}) - z^{\text{pred}}|}{z^{\text{pred}} + \epsilon} < \tau_z. \tag{1}$$

Photometric coverage. For each target view i, coverage is the fraction of pixels that have at least one valid photometric neighbor:

$$C^{(i)}_{\text{photo}} = \mathbb{E}_x \left[\sum_{j \in \mathcal{N}(i)} \mathbf{1}\Big(M^{i,j}_{\text{photo}}(x) \Big) > 0 \right]. \tag{2}$$

$\bar{C}_{\text{photo}}$ is reported as the mean over target views.

Scene	N	mean	min	p10	median	p90	max
M360/counter	216	0.947	0.643	0.907	0.958	0.976	0.987
LLFF/room	37	0.967	0.887	0.944	0.972	0.989	0.989
Syn/toaster	100	0.854	0.432	0.702	0.889	0.955	0.978
LLFF/orchids	23	0.852	0.799	0.821	0.856	0.882	0.892

Table S1: **Distribution of photometric coverage across target views.** Low coverage indicates that few pixels have any valid photometric neighbor under M_{photo} (in-bounds $\wedge$ alpha gate $\wedge$ depth-consistency) and should be treated as a diagnostic warning rather than suppressed.

A.3 Error maps and residual

The following quantities are computed:

- E_{depth}: geometry/visibility mismatch using depth reprojection residuals on M_{geo}.
- E_{base}: photometric inconsistency under SH truncation (L=0) on M_{photo}.
- E_{full}: photometric inconsistency under trained SH on M_{photo}.

The view-dependent residual is:

$$E_{\text{vd}} = \max(0,\, E_{\text{full}} - E_{\text{base}}). \tag{3}$$

Photometric error and aggregation. Photometric inconsistency is computed using the L1 error. For each target view i, per-neighbor photometric inconsistencies are aggregated across $j \in \mathcal{N}(i)$ using the median on valid pixels under M_{photo}. Depth reprojection residuals are likewise aggregated using the median on valid pixels under M_{geo}.

A.4 Normalization and quadrant labeling

For quadrant labeling, E_{base} and E_{full} are normalized by a shared per-view scale:

$$s^{(i)} = q_{90}\Big(\{E^{(i)}_{\text{base}}\} \cup \{E^{(i)}_{\text{full}}\}\Big) + \epsilon, \quad E^{(i)}_{\bullet,n} = E^{(i)}_{\bullet}/s^{(i)}. \tag{4}$$

With threshold t=1, each valid pixel is assigned to one quadrant:

$$\text{Consistent: } E_{\text{base},n} \le t \ \wedge \ E_{\text{full},n} \le t \tag{5}$$

$$\text{Geo/Vis-like: } E_{\text{base},n} > t \ \wedge \ E_{\text{full},n} > t \tag{6}$$

$$\text{Appearance-driven: } E_{\text{base},n} \le t < E_{\text{full},n} \tag{7}$$

$$\text{Debug/Unexpected: } E_{\text{base},n} > t \ge E_{\text{full},n}. \tag{8}$$

Scope note. As stated in the main paper, L=0 is an internal SH basis truncation for diagnostics, not a physical intrinsic decomposition.

B Extended Quantitative Summaries

B.1 Coverage tails (distribution across views)

Table S1 reports per-scene distributions of $C^{(i)}_{\text{photo}}$ across target views. This complements the main paper (which reports only $\bar{C}_{\text{photo}}$) and helps detect low-coverage tails.

B.2 Per-scene error statistics and normalization scale

Table S2 reports averages of per-view robust quantiles (q50/q90) and the per-view normalization scale $s^{(i)}$. These values are not intended as a benchmark, but document the dynamic ranges used for normalization and visualization.

Scene	$\bar{s}$	E_{base} q50	q90	E_{full} q50	q90	E_{depth} q50	q90
M360/counter	0.0172	0.0031	0.0143	0.0045	0.0197	0.0032	0.0972
LLFF/room	0.0095	0.0018	0.0069	0.0033	0.0117	0.0044	0.0628
Syn/toaster	0.0455	0.0013	0.0341	0.0022	0.0560	0.0465	0.3524
LLFF/orchids	0.0450	0.0053	0.0412	0.0074	0.0490	0.0176	0.3461

Table S2: **Additional error statistics (averaged over target views).** $\bar{s}$ is the mean per-view normalization scale $s^{(i)}$. $q50/q90$ are per-view quantiles averaged over target views.

Scene	Cons. (%)	Geo (%)	App (%)	Dbg (%)
M360/counter	85.84±0.41	5.82±0.40	6.11±0.93	2.23±0.31
LLFF/room	85.12±0.40	5.01±0.41	8.10±1.22	1.77±0.48
Syn/toaster	85.16±1.56	5.15±1.56	7.94±3.09	1.76±0.47
LLFF/orchids	87.73±0.07	7.73±0.07	3.15±0.16	1.39±0.15

Table S3: **Quadrant occupancy variation across target views (mean±std).** All values are computed on valid pixels under M_{photo} and normalized with t=1.

Bin (TOASTER)	N pixels	Cons. (%)	Geo (%)	App (%)	Dbg (%)
Low σ^2_{proxy} ($\leq$ p10)	15,257,992	99.79	0.13	0.03	0.05
High σ^2_{proxy} ($\geq$ p90)	4,240,868	53.76	12.33	29.25	4.66

Table S4: **Proxy view-dependence bins vs DRC quadrant occupancy (toaster).** Quadrant occupancy is computed on valid pixels ($M_{\text{photo}} \wedge$ finite σ^2_{proxy}) and binned by per-view percentiles (low$\leq$p10, high$\geq$p90). High-variance pixels show a substantially higher appearance-driven fraction. *Note:* this proxy may include warping/occlusion effects and is used only as a sanity check.

B.3 Stability of quadrant occupancy across views

Table S3 reports mean and standard deviation of quadrant occupancy across target views. Synthetic TOASTER exhibits higher variance, consistent with its heavier coverage tail.

B.4 Proxy view-dependence bin test (NeRF-Synthetic toaster)

As a sanity check, a proxy view-dependence map σ^2_{proxy} is constructed from multi-view GT color variance after warping GT views into each target view using 3DGS depth. Quadrant occupancy changes between low- and high-variance pixels are then measured (Table S4).

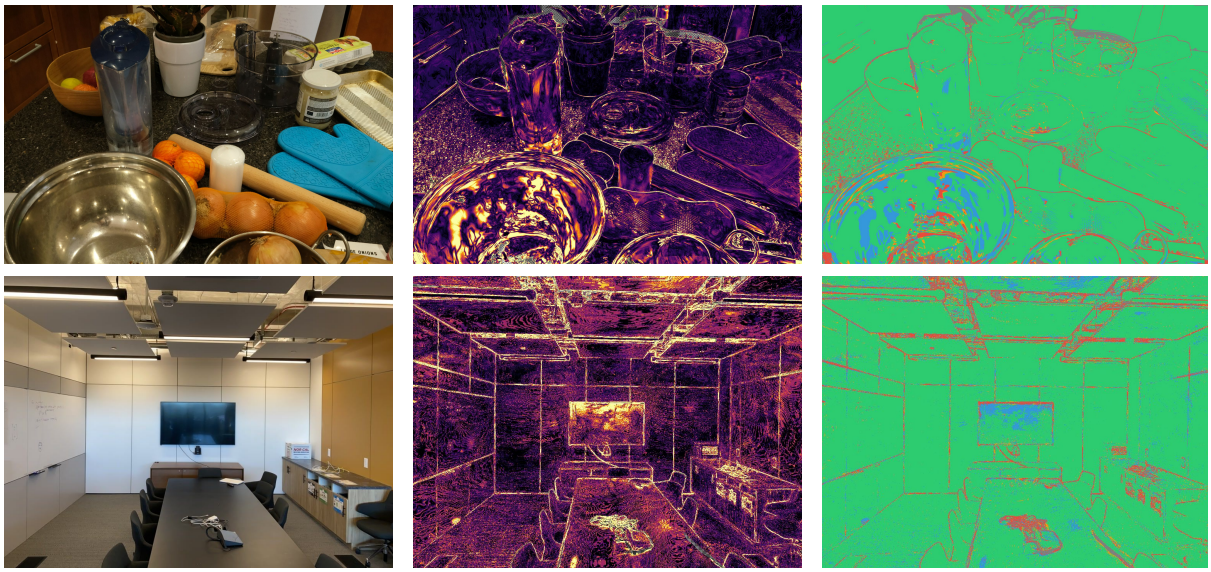
C Additional Qualitative Results

To reduce cherry-picking concerns for the ROI visualizations in the main paper (Fig. 2), uncropped full-view verification for the ROI scenes is provided (Fig. S1). For the main paper's Fig. 2, ROIs are used only for visualization in COUNTER and ROOM; all statistics are computed on full target views.

D Sanity Checks and Limitations

D.1 Sanity checks (brief)

Internal consistency is checked through: (i) identity warp ($i \leftarrow i$) producing near-zero error, (ii) backproject–project round-trip stability, (iii) depth-consistency statistics (in-bounds ratio and agreement between z^{pred} and $D_j(\tilde{x})$), and (iv) explicit coverage reporting ($C^{(i)}_{\text{photo}}$) for every view.



(a) (b) (c)

Figure S1: **Full-view (uncropped) verification for ROI scenes.** For the two scenes where ROIs are used for visualization in Fig. 2 (Mip-NeRF 360 COUNTER and LLFF ROOM), the corresponding *uncropped* target views are shown: (a) rendered RGB, (b) the baseline single-map inconsistency E_{full}, and (c) the quadrant overlay derived from $(E_{\text{base},n}, E_{\text{full},n})$ with t=1. This confirms that ROI cropping is used only for visualization; all statistics (coverage and quadrant occupancy) are computed on full target views. Gray pixels are excluded under M_{photo} (in-bounds/alpha/depth-consistency failures).

D.2 Limitations

DRC is a diagnostic heuristic, not a ground-truth classifier. Very sharp/high-frequency reflections may be partially absorbed into low-order components ("baked-in"), reducing separability under SH truncation. Object-only scenes or large pose gaps can yield low-coverage tails; DRC reports this via C_{photo}.