SIGNED PAIR COMPOSITING FOR GAUSSIAN SPLATTING

Anonymous authors
Paper under double-blind review

ABSTRACT

3D Gaussian Splatting (3DGS) renders a scene by projecting Gaussians to 2D elliptical footprints, sorting them by center depth, and alpha-compositing them front to back. Compositing these footprints in center order assumes that neighboring Gaussians do not overlap in depth, that their center order holds at every pixel, and that their coverages are uncorrelated. Most methods that address overlap relax the first two assumptions by re-sorting Gaussians per pixel or by evaluating them in 3D along each ray, but still treat coverages as independent. Instead, we keep the rasterization pipeline of 3DGS and restore the depth row of the projection Jacobian, which 3DGS sets to zero. This gives each Gaussian, at every pixel, a depth distribution derived from its existing 3D covariance, without changing its footprint. Building on these distributions, our signed pair compositing groups the sorted Gaussians at each pixel into disjoint adjacent pairs. Within a pair, the depth distributions give the probability that each Gaussian lies in front of the other, a soft per-pixel order. The shapes and positions of the two footprints relative to the pixel set the sign of a bounded correlation between their coverages. The compositing weights depend only on the existing geometry and opacities of the Gaussians, without learnable parameters. We prove that signed pair compositing leaves the coverage of every Gaussian unchanged and approaches standard compositing in the correct depth order for pairs well separated in depth. It adds a constant cost per pair and can render models trained with standard compositing without retraining. We also replace the per-channel L_1 loss with the nuclear norm of the per-view color residual matrix, whose gradient weights every principal direction of the color error equally.

1 INTRODUCTION

3D Gaussian Splatting (3DGS; Kerbl et al., 2023) represents a scene with anisotropic 3D Gaussians. It projects each Gaussian to a footprint on the image plane by elliptical weighted average (EWA) splatting (Zwicker et al., 2001; 2002), sorts the footprints by center depth, and alpha-composites them front to back. Unlike ray-marched neural fields (Mildenhall et al., 2020), this pipeline runs in real time, and the same compositing rule produces both the image and the gradients that shape the scene.

After projection, 3DGS treats each Gaussian as a flat ellipse at the depth of its center. Compositing these ellipses in center order assumes that Gaussians along a ray do not overlap in depth, that their center order holds at every pixel, and that their coverages of a pixel are uncorrelated (Max, 1995; Porter & Duff, 1984; Mai et al., 2025). Where Gaussians interpenetrate, the first two assumptions break. The center order can then be wrong over part of the footprint, which misplaces color and causes popping as the camera rotates (Radl et al., 2024). The third assumption can break even when the order is correct. Neighboring Gaussians that tile a surface cover complementary parts of the region around a pixel, but independent coverages let part of what lies behind leak through the seam between them.

Prior work relaxes the first two assumptions in two ways. StopThePop (Radl et al., 2024) approximately re-sorts Gaussians per pixel within a hierarchical tile rasterizer, which removes most popping. Other methods account for the extent of each Gaussian along the ray. Vol3DGS (Talegaonkar et al., 2025) folds this extent into the opacity inside the rasterizer, while ray-traced Gaussians

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.

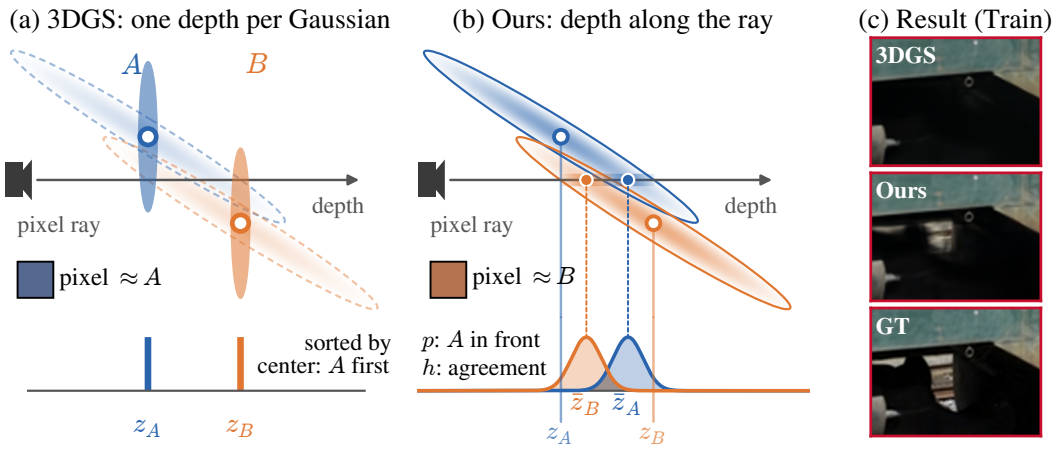


Figure 1: **One depth per Gaussian versus depth along the ray.** (a) 3DGS treats each Gaussian as a flat ellipse at the depth of its center and composites in that order, so the pixel takes mostly the color of A even though the ray meets B first. (b) Restoring the depth row of the projection gives each Gaussian a depth distribution along the ray. These yield a probability p that A is in front and a depth agreement h, and signed pair compositing gives the pixel mostly the color of B. (c) In Train, our full method (compositing and loss) preserves more detail behind the frame of the locomotive (Figure 3).

(Moenne-Loccoz et al., 2024) sample each Gaussian at its peak along the ray, and EVER (Mai et al., 2025) resolves overlap by ray tracing. Both groups still combine Gaussians through the product of their individual transmittances, that is, as independent coverages. For Gaussian splatting, non-exponential transmittance has been defined as a function of accumulated opacity and evaluated with a ray-tracing renderer (Speierer & Jarabo, 2026).

In this work, we handle overlap inside the sorted rasterizer and keep its 2D ellipses. For a single Gaussian, flatness loses nothing, because under the affine approximation of EWA its footprint already integrates out its extent along the ray. When Gaussians overlap, however, their position and thickness along each ray matter, and flatness discards them. This information is in the depth block of the projected covariance, which 3DGS drops by setting the depth row of the projection Jacobian to zero. Restoring that row gives each Gaussian a depth distribution, instead of a single depth, at every pixel it covers. The footprint is unchanged, and the mean and variance of the distribution follow from the existing covariance.

We introduce *signed pair compositing* (Figure 1), which uses these distributions to composite Gaussians in pairs. At each pixel, it groups the sorted Gaussians into disjoint adjacent pairs: the first with the second, the third with the fourth, and so on. Within a pair, the depth distributions determine the probability that each Gaussian lies in front of the other and how closely the two overlap in depth. Where they overlap, the shapes and positions of the two footprints indicate whether the Gaussians tend to cover the same or complementary parts around the pixel. This gives a soft per-pixel order of the pair and a signed footprint coupling that sets a bounded coverage correlation. Together they set how the two colors blend without changing either coverage. Signed pair compositing departs from 3DGS mainly where the members of a pair overlap in depth or are misordered at the pixel (Proposition 1).

We also change which color errors drive training, independently of the compositing. For a color residual E, the per-channel L_1 loss of 3DGS has gradient $\mathrm{sign}(E)$, which depends on the RGB basis. Where a brightness error dominates, the three channels typically share its sign, so the gradient there points along brightness and has no chromatic component. We replace L_1 with the nuclear norm of the color residual matrix of a view. Its gradient weights every principal direction of the color error equally, so a small but systematic chromatic error, such as a color cast, receives the same gradient weight as the dominant brightness error (Section 4.3).

Our contributions are:

- **Pixel-conditional depth:** restoring the depth row of the projection gives each Gaussian, at every pixel it covers, a depth mean that follows its tilt and a depth variance that measures its thickness. The mean has the affine form used for depth maps by RaDe-GS (Zhang et al., 2024a), and we add the variance. Both are stored as three values per Gaussian and evaluated with one dot product per pixel (Section 4.1).
- **Signed pair compositing:** a parameter-free rule for a sorted Gaussian rasterizer that composites two adjacent Gaussians in a soft per-pixel order set by their pixel-conditional depths. A signed footprint coupling adds a bounded coverage correlation. The rule preserves each coverage and, for pairs well separated in depth, approaches standard compositing in their correct depth order (Proposition 1, proved in Appendix C). Its cost per pair is constant, and it can render models trained with standard compositing without retraining (Table 3).
- **Nuclear-norm loss:** a color loss, invariant to orthogonal changes of the color basis, whose gradient weights all principal color-error directions equally (Section 4.3).

2 Related Work

Rasterized Gaussian splatting. Many follow-ups of 3DGS (Kerbl et al., 2023) improve anti-aliasing (Yu et al., 2024) or speed (Hanson et al., 2025) and keep the sorted compositing that our method modifies. Sort-free splatting (Hou et al., 2025) replaces the sort with learned order-independent weights, related to weighted blended order-independent transparency (McGuire & Bavoil, 2013). Real-time order-independent transparency also resolves fragments in small per-pixel buffers such as the k-buffer (Bavoil et al., 2007). Our method buffers a single fragment per pixel to form each pair and, unlike sort-free splatting, keeps the sort and adds no learnable parameters.

Depth inside the rasterizer. RaDe-GS (Zhang et al., 2024a) rasterizes the depth of each Gaussian as an affine function of the pixel to render depth and normal maps, and a Gaussian SLAM system uses the same depth (Zhang et al., 2025). StopThePop (Radl et al., 2024) sorts by the per-pixel peak depth of each Gaussian, and Vol3DGS (Talegaonkar et al., 2025) derives opacity from its density integrated along the ray. Our pixel-conditional depth (Section 4.1) summarizes the depth profile of each Gaussian along the ray. Its mean has the affine form of the RaDe-GS depth and, up to the low-pass dilation, is a first-order (EWA) estimate of the peak depth. Its variance estimates the spread along the ray. StopThePop and Vol3DGS evaluate each Gaussian on its own along the ray; we compare the depth distributions of two Gaussians, which gives a soft per-pixel order. Z-compositing (Duff, 1985) mixes A over B and B over A by the fraction of the pixel where A is nearer, estimated from subpixel depth slopes. Proposition 1(d) has this form, with our soft order from the depth spread in place of that fraction.

Volumetric and ray-traced Gaussians. Sorted compositing assumes that the primitives do not overlap along the ray (Max, 1995). Ray-traced Gaussian particles (Moenne-Loccoz et al., 2024) and EVER (Mai et al., 2025) evaluate primitives along each ray outside the rasterizer. EVER renders constant-density ellipsoids exactly. Unlike both, our method stays in the rasterizer and changes how adjacent contributors combine.

Correlated transmittance. Prior work models correlated, non-exponential transmittance for volumetric media (Bitterli et al., 2018; Vicini et al., 2021) and, for Gaussian splatting, as a function of the opacity accumulated along the ray (Speierer & Jarabo, 2026). We couple two adjacent Gaussians at a pixel through their depth distributions and footprints, inside the sorted rasterizer. The resulting pair transmittance τ in Eq. (9) generally differs from the product of the two member transmittances and equals it when the coupling vanishes.

Photometric losses. 3DGS trains with per-channel L_1 and D-SSIM (Kerbl et al., 2023), close to the L_1 + MS-SSIM loss of Zhao et al. (2017). FreGS (Zhang et al., 2024b) adds a frequency-domain term. These losses act per channel or along spatial and frequency axes. Nuclear-norm regression penalizes the singular values of a spatial error image (Yang et al., 2017). Our nuclear norm instead acts on the pixel-by-channel residual matrix, along the color axis.

3 BACKGROUND

3D Gaussian Splatting. Following Kerbl et al. (2023), a scene is a set of 3D Gaussians with mean $\boldsymbol{\mu}_i \in \mathbb{R}^3$ and covariance Σ_i, evaluated at a 3D point $\mathbf{y}$ as

$$\mathcal{G}_i(\mathbf{y}) = \exp\Big(-\tfrac{1}{2}(\mathbf{y}-\boldsymbol{\mu}_i)^\top \Sigma_i^{-1}(\mathbf{y}-\boldsymbol{\mu}_i)\Big), \qquad \Sigma_i = R_i S_i S_i^\top R_i^\top, \tag{1}$$

with diagonal scaling S_i and rotation R_i from a unit quaternion. Each Gaussian also has an opacity $o_i \in [0,1]$ and spherical-harmonic coefficients for its view-dependent color $\mathbf{c}_i$.

Projection. With view transform W and the Jacobian J_i of the affine projection approximation at $\boldsymbol{\mu}_i$, the projected covariance is (Zwicker et al., 2001)

$$\Sigma_i' = J_i W \Sigma_i W^\top J_i^\top = \begin{pmatrix} \Sigma_i^{\mathrm{2D}} & \boldsymbol{\sigma}_i \\ \boldsymbol{\sigma}_i^\top & \sigma_{z,i}^2 \end{pmatrix}. \tag{2}$$

3DGS zeroes the third (depth) row of J_i and keeps only the image-plane block Σ_i^{2D}. In color compositing, the depth z_i of the projected mean is only a sort key. With projected mean $\mathbf{m}_i$ and $Q_i = (\Sigma_i^{\mathrm{2D}} + 0.3I)^{-1}$, where $0.3I$ is a low-pass dilation, the opacity at pixel $\mathbf{x}$ is

$$\alpha_i(\mathbf{x}) = \min\Big(0.99,\ o_i \exp\big(-\tfrac{1}{2}(\mathbf{x}-\mathbf{m}_i)^\top Q_i (\mathbf{x}-\mathbf{m}_i)\big)\Big). \tag{3}$$

Alpha compositing. The contributors $\mathcal{S}$ of a pixel, the Gaussians with $\alpha_i(\mathbf{x}) \geq 1/255$, are indexed $1, \ldots, |\mathcal{S}|$ in order of z_i and blended front to back,

$$C(\mathbf{x}) = \sum_{i \in \mathcal{S}} \mathbf{c}_i\, \alpha_i \prod_{j=1}^{i-1} (1-\alpha_j), \tag{4}$$

where the product is the transmittance T_i. Blending stops before T_i drops below 10^{-4}. Read as coverage, Eq. (4) is the *over* operator (Porter & Duff, 1984). It assumes that the coverages of different layers are uncorrelated, an assumption that subpixel coverage masks such as the A-buffer avoid (Carpenter, 1984).

Training. 3DGS minimizes $\mathcal{L} = (1-\lambda)\mathcal{L}_1 + \lambda \mathcal{L}_{\text{D-SSIM}}$, where $\mathcal{L}_1$ is the per-channel L_1 loss and λ weights D-SSIM.

4 METHOD

We keep the sorting and densification of 3DGS unchanged. Figure 2 summarizes the pipeline, and Table 4 in Appendix A lists the notation.

4.1 PIXEL-CONDITIONAL DEPTH

To give each Gaussian a depth at every pixel it covers, we set the third row of J_i to $(0,0,1)$ instead of zero. The third row and column of Σ_i' in Eq. (2) then carry the camera depth, and we keep them. Conditioning the projected Gaussian on the image-plane position $\mathbf{x}$ gives a one-dimensional Gaussian in depth with mean and variance

$$\bar{z}_i(\mathbf{x}) = z_i + \mathbf{r}_i^\top(\mathbf{x}-\mathbf{m}_i), \qquad \mathbf{r}_i = Q_i \boldsymbol{\sigma}_i, \qquad V_i = \sigma_{z,i}^2 - \boldsymbol{\sigma}_i^\top Q_i \boldsymbol{\sigma}_i, \tag{5}$$

where $\mathbf{r}_i$ is the depth slope across the image plane. The Schur complement V_i is the depth variance of the dilated Gaussian along the pixel ray, i.e., its thickness. The affine approximation takes every pixel ray parallel to the viewing ray through $\boldsymbol{\mu}_i$, so V_i is the same for every $\mathbf{x}$. The mean $\bar{z}_i$ follows the tilt of the Gaussian across its footprint. Both come from the Σ_i' that the rasterizer already computes. We store $\mathbf{r}_i$ and V_i, three values per Gaussian, during preprocessing, so evaluating the distribution at a pixel costs one dot product. Appendix B derives them and shows that the image-plane footprint is unchanged.

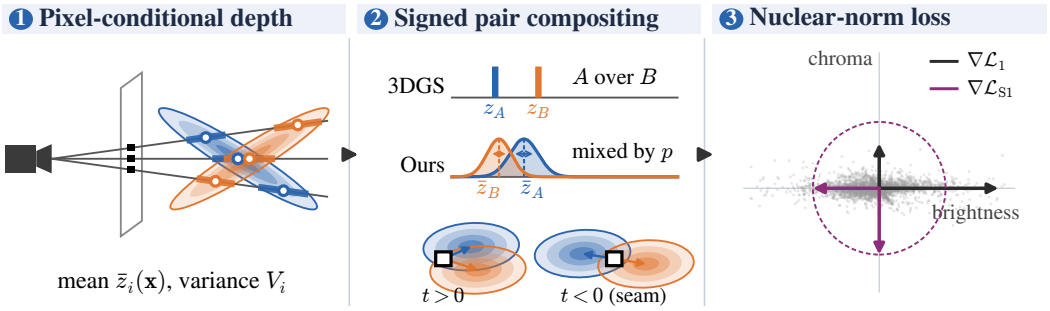


Figure 2: **Pipeline.** Left: along each pixel ray, each Gaussian has a depth distribution with a mean (dot) and a variance (bar); A (blue) is in front on the top ray and B (orange) on the bottom ray. Middle: 3DGS composites A over B by center depth, while we mix both orders by p from their per-pixel depth means and variances. Bottom middle: for isotropic footprints, the bounded coupling has $t > 0$ when both footprint centers lie on the same side of the pixel (white square), where they tend to cover the same area, and $t < 0$ when they lie on opposite sides, at a seam where less background leaks through. Right: each pixel of the residual E is a point in color space (Train). Arrows show how hard each loss pushes along each principal direction of E: L_1 mostly along brightness, the nuclear norm equally along all.

4.2 Signed Pair Compositing

Signed pair compositing uses these depth distributions to composite overlapping Gaussians in pairs. Figure 2 (middle) contrasts this with 3DGS.

Pairing. At each pixel, we group the sorted contributors into disjoint adjacent pairs (A, B), where A comes first in the sort: the first contributor is paired with the second, the third with the fourth, and so on. Pairs need only one buffered contributor per pixel and are the smallest groups in which coverages can be correlated. Because the pairs are disjoint, each contributor is weighted once, and at every pixel the weights of all contributors, including an unpaired final one, sum to one with the final transmittance, even under early termination. Neighbors in different pairs are composited as in 3DGS.

Depth order and agreement. A pair is summarized by its soft depth order p, its depth agreement h, and its signed footprint coupling t. The first two are

$$\delta = \bar{z}_A - \bar{z}_B, \qquad V_\delta = V_A + V_B, \qquad p = \Phi\big(-\delta/\sqrt{V_\delta}\big), \qquad h = \frac{2\sqrt{V_A V_B}}{V_\delta}\exp\Big(-\frac{\delta^2}{2V_\delta}\Big), \tag{6}$$

where Φ is the standard normal CDF, δ is the signed depth gap, and V_δ is its variance. Just as opacity is the probability that a sample is covered, we read Eq. (5) as the distribution of the depth at which the Gaussian covers the sample, given that it does. Then p is the probability that A covers the sample in front of B. The sort uses z_i, whereas p uses $\bar{z}_i$, so p can fall below $1/2$. The order within a pair is thus soft, and only the order between pairs follows the sort. Here $h \in (0, 1]$ is the squared Bhattacharyya coefficient of the two conditional depth Gaussians and equals one only when they coincide. Whether two footprints cover the same or complementary sample points matters mainly when both occupy the same depth range along the ray, as for neighbors that tile one surface. As a modeling choice, we therefore scale the footprint coupling t, defined next, by h. The factor h is small both for pairs far apart in depth and for a thin Gaussian inside a much thicker one ($2\sqrt{V_A V_B}/V_\delta \to 0$ as $V_A/V_B \to 0$ or ∞). Such pairs are ordered by p with nearly independent coverages (Proposition 1(d,e)).

Footprint coupling. The coupling t measures whether the two footprint centers lie on the same side of the pixel or on opposite sides. Let $\mathbf{d}_i = \mathbf{m}_i - \mathbf{x}$, $\mathbf{u} = Q_A\mathbf{d}_A$, $\mathbf{v} = Q_B\mathbf{d}_B$, and $K = (Q_A + Q_B)^{-1}$. Then

$$t = \frac{2\,\mathbf{u}^\top K\mathbf{v}}{2 + \mathbf{u}^\top K\mathbf{u} + \mathbf{v}^\top K\mathbf{v}}. \tag{7}$$

Here $\mathbf{u}$ and $\mathbf{v}$ are the gradients of the two log-footprints at the pixel, and K is the covariance of the normalized product of the two footprint Gaussians.

For isotropic footprints, t has the sign of $\mathbf{d}_A^\top \mathbf{d}_B$. We treat coverage as the event that a sample point near $\mathbf{x}$, in the region of covariance K, is covered. Same-side footprints then tend to cover the same sample points, and opposite-side footprints complementary ones. The coupling is symmetric in A and B, vanishes when the pixel is at either center, and is invariant to affine changes of image coordinates (Appendix D).

Joint coverage and weights. With opacities $a = \alpha_A$ and $b = \alpha_B$ from Eq. (3), the joint coverage of the pair is the independent value ab plus a covariance scaled by ht, the footprint coupling weighted by the depth agreement,

$$O = ab + h\,t\,a(1-a)\,b(1-b). \tag{8}$$

When the two coverage events are Bernoulli variables with means a and b and this covariance, O is the probability that both members cover a sample drawn near $\mathbf{x}$. It has the Farlie–Gumbel–Morgenstern form with dependence parameter ht (Nelsen, 2006). As in stochastic transparency (Enderton et al., 2010), standard compositing corresponds to independent coverage masks, and ht sets their dependence. We choose this covariance because it is polynomial in a and b, vanishes whenever either coverage is certain, and keeps O feasible for all $|ht| \le 1$ (Proposition 1(a,b); Appendix C). Each color weight is the probability that its member is visible, and the pair transmittance is the probability that neither member covers the sample:

$$w_A = a - (1-p)O, \qquad w_B = b - pO, \qquad \tau = 1 - a - b + O. \tag{9}$$

With the depth order independent of the coverage events, member A is visible when it covers and B does not (probability $a - O$) or when both cover and A is in front (probability pO). This gives w_A, and w_B follows by symmetry. A positive t raises O and τ. A negative t lowers both, so less of what lies behind shows through the seam where two footprints meet from opposite sides. Because $\tau = (1-a)(1-b)(1+ht\,ab)$, the coupling changes the leak $(1-a)(1-b)$ by the relative amount $ht\,ab$. This amount stays below 0.21 in magnitude for any two footprints, since $|t|$ approaches one only where both footprints have decayed at the pixel (Appendix C). Accumulation proceeds pair by pair:

$$C \leftarrow C + T\,(w_A \mathbf{c}_A + w_B \mathbf{c}_B), \qquad T \leftarrow T\,\tau. \tag{10}$$

Signed pair compositing has the following properties (proofs in Appendix C).

Proposition 1 (Signed pair compositing)**.** For $a, b \in [0,1)$, $p \in [0,1]$, h from Eq. (6), and t from Eq. (7):

(a) $|t| < 1$ and $h \le 1$, so $|ht| < 1$;

(b) $w_A,\ w_B,\ \tau \ge 0$ and $w_A + w_B + \tau = 1$;

(c) $w_A + (1-p)\,O = a$ and $w_B + p\,O = b$, so each member keeps its 3DGS coverage;

(d) if $ht = 0$, the weights are the p-weighted average of standard compositing in the orders (A, B) and (B, A), so p acts as a soft per-pixel order;

(e) as $\delta/\sqrt{V_\delta} \to -\infty$ $(+\infty)$, Eq. (10) tends to Eq. (4) applied to A then B (B then A).

We differentiate signed pair compositing exactly, except where the opacity clamp or cutoff switches or the sort order changes. Through $\bar{z}_i$ and V_i, the gradient reaches $\boldsymbol{\sigma}_i$ and $\sigma^2_{z,i}$, which 3DGS discards. In 3DGS, the first two rows of J_i annihilate the viewing direction through $\boldsymbol{\mu}_i$, so the image of a view is invariant to stretching or tilting a Gaussian along it. With the depth row restored, the loss of a view also shapes the extent of a Gaussian along its rays wherever it overlaps its partner.

4.3 Training Objective

Our second change replaces the per-channel L_1 term of the training loss. Let $E \in \mathbb{R}^{P\times 3}$ stack the per-pixel RGB residuals between a rendered view and its ground truth. We use the nuclear (Schatten-1) norm,

$$\mathcal{L}_{\mathrm{S1}} = \frac{\|E\|_*}{3\sqrt{P}}, \qquad \|E\|_* = \inf_{\Omega \succ 0} \tfrac{1}{2}\big[\,\mathrm{tr}(E^\top E\,\Omega^{-1}) + \mathrm{tr}\,\Omega\big], \tag{11}$$

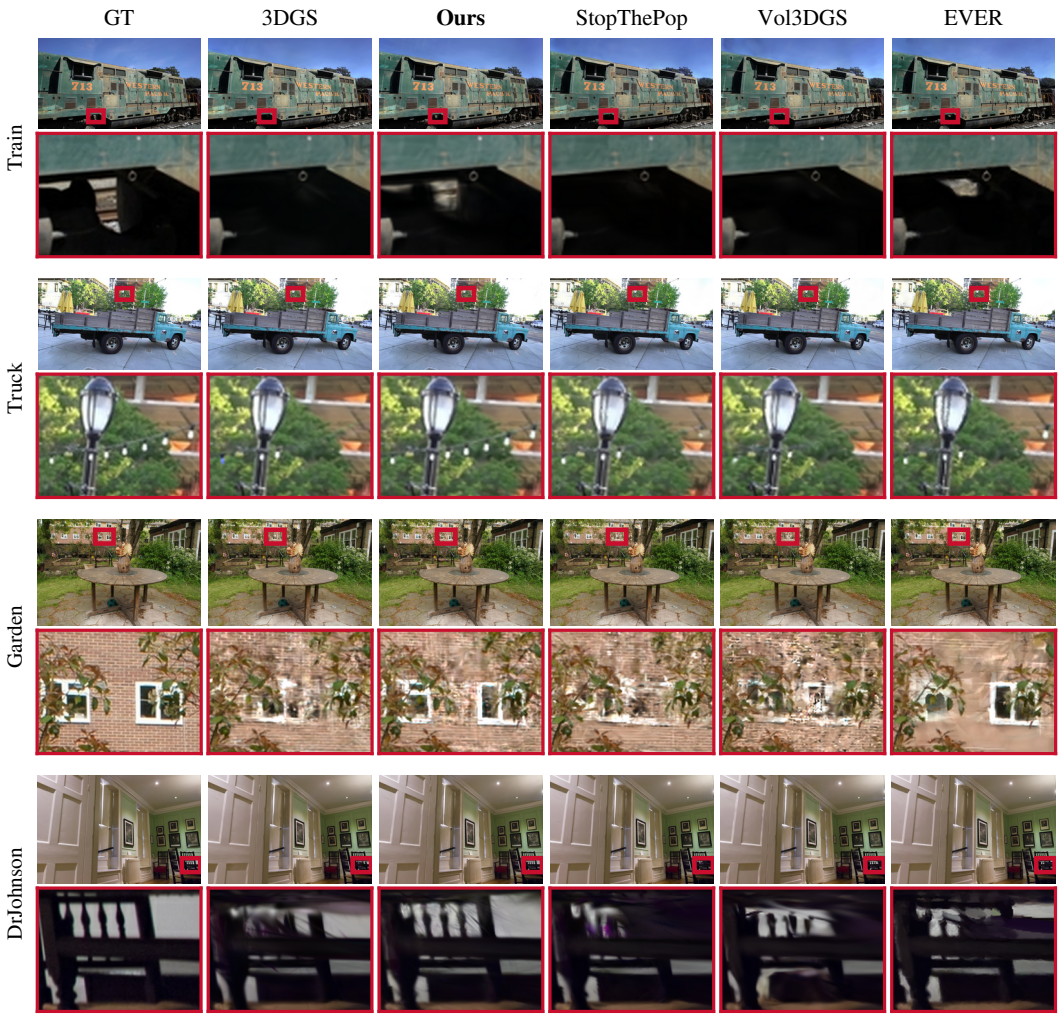


Figure 3: **Qualitative comparison.** Held-out views from Tanks&Temples (Train, Truck), Mip-NeRF360 (Garden), and Deep Blending (DrJohnson). For each scene, the top row shows the full view and the bottom row the region marked in red. Appendix H shows three more views.

and the total loss $(1-\lambda)\mathcal{L}_{\text{S1}} + \lambda\mathcal{L}_{\text{D-SSIM}}$ with $\lambda = 0.2$. For invertible $E^\top E$, the minimizer $\Omega = (E^\top E)^{1/2}$ has a closed form per view, so $\|E\|_*$ is a sum of squared Mahalanobis norms of the pixel residuals under a color metric fitted to the view. For isotropic errors, $\mathcal{L}_{\text{S1}}$ equals the RMS error. Unlike the per-channel L_1 term, $\mathcal{L}_{\text{S1}}$ is invariant to orthogonal changes of the color basis. Among scaled Schatten norms, it is the only one whose gradient $E(E^\top E)^{-1/2}$ weights every principal color-error direction equally (Appendix E).

4.4 Implementation

Signed pair compositing changes only the per-pixel blending and does not depend on tile assignment. It needs only 3D covariances, sorted contributors, and per-pixel opacities, and beyond standard compositing each pair costs one 2×2 inverse, one error function, one exponential, and two square roots. We assign tiles with the tile test of Speedy-Splat (Hanson et al., 2025), without its pruning, whereas the 3DGS baseline keeps its original tile assignment. The test assigns each Gaussian to the tiles that intersect the ellipse where its opacity reaches the cutoff $1/255$, instead of the tiles in its 3σ bounding box. Rendering our trained models with the 3σ bounding box leaves PSNR, SSIM, and LPIPS unchanged at the reported precision. All other settings follow the 3DGS reference implementation, and Appendix F gives the remaining details.

Table 1: Novel-view synthesis at 30K iterations, averaged over the scenes of each dataset (D-SSIM weight $\lambda = 0.2$, same splits and resolutions for all methods). FPS: rendering speed on one NVIDIA L40 GPU (Appendix F). Bold marks the best value in each column among the methods in this table. Table 2 separates the contributions of the compositing and the loss. Ours uses signed pair compositing and the nuclear-norm loss.

	Mip-NeRF360				Tanks&Temples				Deep Blending			
Method	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑
3DGS	27.54	0.816	0.215	114	23.80	0.853	0.169	154	29.84	0.907	0.238	**117**
StopThePop	27.34	0.816	0.212	97	23.13	0.846	0.171	144	**29.96**	**0.909**	**0.232**	104
EVER	27.51	**0.824**	**0.198**	15	23.53	**0.869**	**0.135**	14	29.48	0.907	0.240	20
Vol3DGS	27.33	0.818	0.200	59	23.61	0.851	0.174	102	29.64	0.905	0.242	60
Ours	**27.66**	0.816	0.213	**122**	**24.10**	0.855	0.165	**155**	29.74	0.906	0.236	106

Table 2: Ablation at 30K iterations, averaged over the scenes of each dataset. Pair: signed pair compositing (Section 4.2); $\mathcal{L}_{\mathrm{S1}}$: nuclear-norm loss (Section 4.3). Rows 1 and 4 are 3DGS and Ours in Table 1; row 1 is the official 3DGS, and rows 2–4 use the tile test (Section 4.4).

		Mip-NeRF360			Tanks&Temples			Deep Blending		
Pair	$\mathcal{L}_{\mathrm{S1}}$	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
		27.54	0.816	0.215	23.80	0.853	0.169	29.84	0.907	0.238
✓		27.53	0.816	0.212	23.79	0.853	0.166	29.80	0.907	0.236
	✓	27.73	0.816	0.215	24.07	0.855	0.167	29.78	0.906	0.239
✓	✓	27.66	0.816	0.213	24.10	0.855	0.165	29.74	0.906	0.236

Table 3: Same trained models, rendered with a different compositing. Each trained model of Table 2 is kept fixed, and only the compositing used to render it is changed, without retraining. Standard: 3DGS compositing, Eq. (4); †: the compositing the model was trained with. In the first column, + denotes 3DGS with signed pair compositing or with $\mathcal{L}_{\mathrm{S1}}$, as in Table 2. Averages over the scenes of each dataset at 30K iterations.

		Mip-NeRF360			Tanks&Temples			Deep Blending		
Trained	Rendered	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
3DGS	standard†	27.54	0.816	0.215	23.80	0.853	0.169	29.84	0.907	0.238
	signed pair	27.52	0.815	0.213	23.78	0.851	0.167	29.85	0.907	0.236
+ signed pair	signed pair†	27.53	0.816	0.212	23.79	0.853	0.166	29.80	0.907	0.236
	standard	27.53	0.817	0.214	23.80	0.853	0.168	29.79	0.907	0.237
+ $\mathcal{L}_{\mathrm{S1}}$	standard†	27.73	0.816	0.215	24.07	0.855	0.167	29.78	0.906	0.239
	signed pair	27.71	0.816	0.212	24.05	0.854	0.164	29.78	0.906	0.236
Ours	signed pair†	27.66	0.816	0.213	24.10	0.855	0.165	29.74	0.906	0.236
	standard	27.67	0.816	0.215	24.11	0.856	0.167	29.73	0.906	0.238

5 Experiments

5.1 Setup

We evaluate on all nine Mip-NeRF360 scenes (Barron et al., 2022), Train and Truck from Tanks&Temples (Knapitsch et al., 2017), and DrJohnson and Playroom from Deep Blending (Hedman et al., 2018). We use the splits and resolutions of Kerbl et al. (2023) and report PSNR, SSIM, and LPIPS (VGG; Zhang et al., 2018). We compare with 3DGS (Kerbl et al., 2023) and with three methods that relax its compositing assumptions. StopThePop (Radl et al., 2024) re-sorts Gaussians per pixel, and EVER (Mai et al., 2025) and Vol3DGS (Talegaonkar et al., 2025) account for the extent of each Gaussian along the ray. We train 3DGS, EVER, and Vol3DGS with their released settings, and StopThePop with the default settings of its code. These defaults keep the opacity reset of 3DGS rather than the opacity decay used for the results in its paper (Appendix F). All methods train for 30K iterations on a single NVIDIA L40 GPU. Appendix F lists the remaining settings.

5.2 Main Results

Table 1 compares all methods, including rendering speed. Appendix I gives per-scene values, and Table 5 (Appendix F) reports storage and the number of Gaussians that each method reaches under its own densification, whereas Table 2 compares the components under the same schedule.

5.3 Ablation

Table 2 starts from 3DGS and adds signed pair compositing and the nuclear-norm loss, first separately and then together. Table 3 keeps each trained model fixed and changes only the compositing used to render it. Switching the compositing leaves PSNR and SSIM nearly unchanged, while LPIPS changes with the compositing used for rendering rather than the one used for training. Signed pair compositing can therefore render models trained with standard compositing without retraining, and models trained with it can still be rendered with standard compositing. Appendix G renders our models with $t = 0$ and shows where each pair term acts: uncertain and reversed pair orders cover large parts of the surfaces in the views shown there, whereas the coupling concentrates along structure boundaries.

5.4 Qualitative Results

Figure 3 compares held-out views. Our method preserves more detail behind the frame of the locomotive in Train, in the string of bulbs in Truck, in the windows behind the branches in Garden, and in the balusters of a chair back in DrJohnson.

6 Limitations and Conclusion

Limitations. Disjoint pairing is a tractable local model of depth order and coverage dependence. Pairs are accumulated through transmittance, but dependencies across pair boundaries or among three or more Gaussians are not modeled explicitly, and the pairing changes discontinuously with the opacity cutoff and the sort. Overlapping pairs could extend this scope but would need a consistent treatment of shared contributors (Appendix C). The Farlie–Gumbel–Morgenstern form cannot represent exact seams or full nesting. Pixel-conditional depth relies on the affine approximation of EWA, and the compositing treats each Gaussian as a coverage event rather than an emission–absorption density (Max, 1995). The nuclear-norm loss has no spatial structure and relies on D-SSIM for it. Both changes use only inputs that sorted splatting pipelines already have, and the loss applies to other differentiable renderers; we evaluate them on static 3DGS optimized per scene.

Conclusion. Restoring the depth row of the projection lets the sorted rasterizer composite overlapping Gaussians in pairs from their per-pixel depth distributions. Signed pair compositing keeps every coverage and approaches standard compositing in the correct depth order for pairs well separated in depth. It can also render models trained with standard compositing without retraining (Table 3). The gradient of the nuclear-norm loss weights all principal color-error directions equally. The independence of coverages in the *over* operator is a modeling choice that the sorted rasterizer can relax at constant cost per pair, without ray tracing.

Reproducibility Statement

The compositing rule and its pair quantities are defined in Section 4 and summarized in Algorithm 1 (Appendix C), and implementation details are given in Section 4.4 and Appendix F. Proposition 1 is proved in Appendix C, and data, training, baseline, and metric settings are listed in Appendix F. The training objective is given in Section 4.3 and derived in Appendix E.

AI Use Statement

In this work, we used generative AI tools to assist with implementing and debugging parts of the code and with drafting and checking the proofs in the appendix. We have not used generative AI tools to propose the research idea, the method, or its hypotheses, and generating or cleaning data,

translation, and qualitative analysis are not applicable to this work. Additionally, we used generative AI tools for literature search, editing the text, preparing figures, and formatting references. We have reviewed all AI-assisted work, including re-checking the derivations and testing the code. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REFERENCES

Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In *CVPR*, 2022.

Louis Bavoil, Steven P. Callahan, Aaron Lefohn, João L. D. Comba, and Cláudio T. Silva. Multi-fragment effects on the GPU using the k-buffer. In *ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D)*, pp. 97–104, 2007.

Benedikt Bitterli, Srinath Ravichandran, Thomas Müller, Magnus Wrenninge, Jan Novák, Steve Marschner, and Wojciech Jarosz. A radiative transfer framework for non-exponential media. *ACM Transactions on Graphics*, 37(6), 2018.

Loren Carpenter. The A-buffer, an antialiased hidden surface method. In *Computer Graphics (Proceedings of SIGGRAPH)*, volume 18, pp. 103–108, 1984.

Tom Duff. Compositing 3-D rendered images. In *Computer Graphics (Proceedings of SIGGRAPH)*, volume 19, pp. 41–44, 1985.

Eric Enderton, Erik Sintorn, Peter Shirley, and David Luebke. Stochastic transparency. In *ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D)*, pp. 157–164, 2010.

Alex Hanson, Allen Tu, Geng Lin, Vasu Singla, Matthias Zwicker, and Tom Goldstein. Speedy-Splat: Fast 3D Gaussian splatting with sparse pixels and sparse primitives. In *CVPR*, 2025.

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

Qiqi Hou, Randall Rauwendaal, Zifeng Li, Hoang Le, Farzad Farhadzadeh, Fatih Porikli, Alexei Bourd, and Amir Said. Sort-free Gaussian splatting via weighted sum rendering. In *ICLR*, 2025.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023.

Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and Temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4), 2017.

Lawrence I-Kuei Lin. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45(1):255–268, 1989.

Alexander Mai, Peter Hedman, George Kopanas, Dor Verbin, David Futschik, Qiangeng Xu, Falko Kuester, Jonathan T. Barron, and Yinda Zhang. EVER: Exact volumetric ellipsoid rendering for real-time view synthesis. In *ICCV*, 2025.

Nelson Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2):99–108, 1995.

Morgan McGuire and Louis Bavoil. Weighted blended order-independent transparency. *Journal of Computer Graphics Techniques*, 2(2):122–141, 2013.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.

Nicolas Moenne-Loccoz, Ashkan Mirzaei, Or Perel, Riccardo de Lutio, Janick Martinez Esturo, Gavriel State, Sanja Fidler, Nicholas Sharp, and Zan Gojcic. 3D Gaussian ray tracing: Fast tracing of particle scenes. *ACM Transactions on Graphics*, 43(6), 2024.

Roger B. Nelsen. *An Introduction to Copulas*. Springer, 2nd edition, 2006.

Thomas Porter and Tom Duff. Compositing digital images. In *Computer Graphics (Proceedings of SIGGRAPH)*, volume 18, pp. 253–259, 1984.

Lukas Radl, Michael Steiner, Mathias Parger, Alexander Weinrauch, Bernhard Kerbl, and Markus Steinberger. StopThePop: Sorted Gaussian splatting for view-consistent real-time rendering. *ACM Transactions on Graphics*, 43(4), 2024.

Sébastien Speierer and Adrian Jarabo. Generalized non-exponential Gaussian splatting. *arXiv preprint arXiv:2603.02887*, 2026.

Chinmay Talegaonkar, Yash Belhe, Ravi Ramamoorthi, and Nicholas Antipa. Volumetrically consistent 3D Gaussian rasterization. In *CVPR*, 2025.

Delio Vicini, Wenzel Jakob, and Anton Kaplanyan. A non-exponential transmittance model for volumetric scene representations. *ACM Transactions on Graphics*, 40(4), 2021.

Jian Yang, Lei Luo, Jianjun Qian, Ying Tai, Fanlong Zhang, and Yong Xu. Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(1):156–171, 2017.

Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-Splatting: Alias-free 3D Gaussian splatting. In *CVPR*, 2024.

Baowen Zhang, Chuan Fang, Rakesh Shrestha, Yixun Liang, Xiaoxiao Long, and Ping Tan. RaDe-GS: Rasterizing depth in Gaussian splatting. *arXiv preprint arXiv:2406.01467*, 2024a.

Jiahui Zhang, Fangneng Zhan, Muyu Xu, Shijian Lu, and Eric Xing. FreGS: 3D Gaussian splatting with progressive frequency regularization. In *CVPR*, 2024b.

Richard Zhang, Phillip Isola, Alexei A. Efros, and Eli Shechtman. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.

Wei Zhang, Qing Cheng, David Skuddis, Niclas Zeller, Daniel Cremers, and Norbert Haala. HI-SLAM2: Geometry-aware Gaussian SLAM for fast monocular scene reconstruction. *IEEE Transactions on Robotics*, 41:6478–6493, 2025.

Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57, 2017.

Matthias Zwicker, Hanspeter Pfister, Jeroen van Baar, and Markus Gross. EWA volume splatting. In *IEEE Visualization*, 2001.

Matthias Zwicker, Hanspeter Pfister, Jeroen van Baar, and Markus Gross. EWA splatting. *IEEE Transactions on Visualization and Computer Graphics*, 8(3):223–238, 2002.

A NOTATION

Table 4 lists the main symbols.

Table 4: Notation for signed pair compositing and the training objective.

Symbol	Meaning
Σ'_i, Σ_i^{2D}, $\boldsymbol{\sigma}_i$, $\sigma^2_{z,i}$	projected covariance and its image, cross, and depth blocks (Eq. (2))
Q_i, $\mathbf{m}_i$, z_i	dilated image-plane precision, projected mean, depth of the mean
$\bar{z}_i(\mathbf{x})$, V_i, $\mathbf{r}_i$	mean of the depth distribution, its variance (dilated thickness along the ray), its image-plane slope (Eq. (5))
$\mathcal{S}$, Φ	contributors of a pixel (Eq. (4)), standard normal CDF
(A, B), a, b	disjoint adjacent pair (A earlier in the sort) and their opacities α_A, α_B
δ, V_δ	signed depth gap $\bar{z}_A - \bar{z}_B$ and its variance $V_A + V_B$
p	soft depth order: probability that a point of A's distribution lies in front of one of B's (Eq. (6))
h	depth agreement: squared Bhattacharyya coefficient of the two distributions (Eq. (6))
$\mathbf{d}_i$, $\mathbf{u}$, $\mathbf{v}$, K	offset $\mathbf{m}_i - \mathbf{x}$, log-footprint gradients, $(Q_A + Q_B)^{-1}$ (sample neighborhood)
t	signed footprint coupling (Eq. (7))
O	joint coverage of the pair (Eq. (8))
w_A, w_B, τ	color weights and transmittance of the pair (Eq. (9))
E, P	$P \times 3$ color residual matrix of a view, number of pixels
$\mathcal{L}_{\mathrm{S1}}$, λ, Ω	nuclear-norm loss, D-SSIM weight, matrix variable of its variational form (Eq. (11))
U, s_k, Γ, q	left singular vectors, singular values, right singular vectors, Schatten order (Appendix E)
κ_A, κ_B	isotropic precisions, $Q_i = \kappa_i I$ (Appendix D)
C, T	accumulated color and transmittance (Eq. (10))
k	pending first member in Algorithm 1

B PIXEL-CONDITIONAL DEPTH FROM THE FULL PROJECTION

This appendix derives Eq. (5) and the pair quantities of Eq. (6) from the projection that 3DGS already performs.

Keeping the third row. Let $\boldsymbol{\nu} = (\nu_x, \nu_y, \nu_z)$ be the position of $\boldsymbol{\mu}_i$ in camera coordinates. EWA splatting linearizes the map from camera coordinates to image coordinates around $\boldsymbol{\nu}$. We extend the image coordinates (x, y) by the camera depth z. A camera-space point $\mathbf{y}'$ near $\boldsymbol{\nu}$ then maps to $(x, y, z) = \phi(\boldsymbol{\nu}) + J_i(\mathbf{y}' - \boldsymbol{\nu})$, where ϕ is the perspective projection extended by depth, f_x, f_y are the focal lengths, and

$$J_i = \begin{pmatrix} f_x/\nu_z & 0 & -f_x\nu_x/\nu_z^2 \\ 0 & f_y/\nu_z & -f_y\nu_y/\nu_z^2 \\ 0 & 0 & 1 \end{pmatrix}. \tag{12}$$

The first two rows are those of 3DGS, including its clamping of ν_x/ν_z and ν_y/ν_z to slightly beyond the field of view. The reference implementation sets the third row to zero; we set it to $(0, 0, 1)$ to keep the depth coordinate. Under this affine map the Gaussian of Eq. (1) becomes a Gaussian in (x, y, z) with mean $(\mathbf{m}_i, z_i)$ and covariance Σ'_i of Eq. (2).

Marginal and conditional. Adding the low-pass dilation $\mathrm{diag}(0.3, 0.3, 0)$ to Σ'_i and marginalizing z leaves $\Sigma_i^{\mathrm{2D}} + 0.3I = Q_i^{-1}$, so the image-plane footprint, and hence α_i in Eq. (3), is exactly that of 3DGS. Conditioning on the image position $\mathbf{x}$ instead gives, by the standard Gaussian conditioning formula,

$$z \mid \mathbf{x} \;\sim\; \mathcal{N}\big(z_i + \boldsymbol{\sigma}_i^\top Q_i(\mathbf{x} - \mathbf{m}_i),\; \sigma^2_{z,i} - \boldsymbol{\sigma}_i^\top Q_i \boldsymbol{\sigma}_i\big), \tag{13}$$

which is Eq. (5) with $\mathbf{r}_i = Q_i\boldsymbol{\sigma}_i$. The mean varies linearly across the footprint. A Gaussian on an oblique surface is therefore nearer on one side of its footprint than on the other, and two such

Gaussians can swap depth order inside their overlap. As for any Gaussian conditional, the variance does not depend on $\mathbf{x}$; under the affine approximation, the lines of constant $\mathbf{x}$ are parallel to the viewing ray through $\boldsymbol{\mu}_i$. This variance is the thickness, in camera depth, of the low-pass-dilated Gaussian along that direction. It is a property of the existing covariance, not an uncertainty about its position. Both quantities come from entries of Σ_i' that the reference rasterizer already forms as a full 3×3 matrix. The extra cost per Gaussian is one matrix–vector product for $\mathbf{r}_i$ and one inner product for V_i.

Pair quantities. For the two members of a pair, draw independent points $z_A \sim \mathcal{N}(\bar{z}_A, V_A)$ and $z_B \sim \mathcal{N}(\bar{z}_B, V_B)$ from their depth distributions. Then $z_A - z_B \sim \mathcal{N}(\delta, V_\delta)$ with $\delta = \bar{z}_A - \bar{z}_B$ and $V_\delta = V_A + V_B$, and hence $p = \Pr[z_A < z_B] = \Phi(-\delta/\sqrt{V_\delta})$. The Bhattacharyya coefficient of the two Gaussians is $\mathrm{BC} = \sqrt{2\sqrt{V_A V_B}/V_\delta}\,\exp(-\delta^2/(4V_\delta))$, and $h = \mathrm{BC}^2$ gives Eq. (6). By the arithmetic–geometric mean inequality, $2\sqrt{V_A V_B} \le V_\delta$, so $h \in (0, 1]$, with $h = 1$ only for identical Gaussians. As $|\delta|/\sqrt{V_\delta} \to \infty$, the depth agreement h tends to 0 and p to 1 or 0 according to the sign of δ. Equation (10) then tends to standard compositing in sort order or in reversed order (Proposition 1(e)).

Camera depth versus ray distance. We measure depth along the camera axis, whereas EWA volume splatting uses the distance along the ray (Zwicker et al., 2001). On a single pixel ray, the two differ by a constant factor. Depth enters the pair quantities of Section 4.2 only through $\delta/\sqrt{V_\delta}$ and V_A/V_B, so the choice leaves them unchanged up to the difference between the affine approximations at the two means.

C Properties of Signed Pair Compositing

This appendix proves Proposition 1; Appendix D discusses the footprint coupling t.

(a) Bound on the coupling. We show $|t| < 1$ and $h \le 1$. Since $Q_A, Q_B \succ 0$, the matrix $K = (Q_A + Q_B)^{-1}$ is positive definite and defines an inner product. The Cauchy–Schwarz and arithmetic–geometric mean inequalities give $2|\mathbf{u}^\top K \mathbf{v}| \le 2\sqrt{\mathbf{u}^\top K \mathbf{u}\; \mathbf{v}^\top K \mathbf{v}} \le \mathbf{u}^\top K \mathbf{u} + \mathbf{v}^\top K \mathbf{v}$. The denominator of t exceeds this bound by 2, hence $|t| < 1$. The bound $h \le 1$ follows from $2\sqrt{V_A V_B} \le V_\delta$ (Appendix B).

(b) Nonnegativity and partition of unity. We show $w_A, w_B, \tau \ge 0$ and $w_A + w_B + \tau = 1$. Substituting Eq. (8) into Eq. (9) and factoring gives

$$\tau = (1-a)(1-b)\,(1 + ht\,ab), \tag{14}$$

$$w_A = a\,\big[1 - (1-p)\,b\,\big(1 + ht(1-a)(1-b)\big)\big], \tag{15}$$

$$w_B = b\,\big[1 - p\,a\,\big(1 + ht(1-a)(1-b)\big)\big]. \tag{16}$$

Since $|ht| < 1$ by (a), we have $\tau \ge (1-a)(1-b)(1-ab) \ge 0$. For w_A, the factor $1 + ht(1-a)(1-b)$ lies in $[0,\, 1 + (1-a)(1-b)]$, again by (a). Moreover, $b\,(1 + (1-a)(1-b)) = b + b(1-a)(1-b) \le b + (1-b) = 1$, since $b(1-a) \le 1$. With $1 - p \le 1$, the bracket in w_A is therefore nonnegative. The same argument with a and p in place of b and $1-p$ gives $w_B \ge 0$. Because $\alpha_i \le 0.99$ and $|ht| < 1$, the pair transmittance is strictly positive, so T never reaches zero exactly. Finally, summing Eq. (9) gives $w_A + w_B + \tau = a + b - O + 1 - a - b + O = 1$. By (a), $O = ab\,[1 + ht(1-a)(1-b)]$, $a - O = a(1-b)\,[1 - ht(1-a)b]$, and $b - O = b(1-a)\,[1 - ht\,a(1-b)]$ are also nonnegative, so together with τ they form a feasible joint distribution of the two coverage events.

(c) Marginal coverages. Rearranging Eq. (9) gives $w_A + (1-p)O = a$ and $w_B + pO = b$. The joint coverage O is thus split between the two members in proportion to the probability that each is in front. The total coverage of each member, visible or occluded by its partner, stays at its 3DGS value.

(d) Soft order without coupling. If $ht = 0$, then $O = ab$ and

$$w_A = p\,a + (1-p)\,a(1-b), \qquad w_B = p\,b(1-a) + (1-p)\,b, \qquad \tau = (1-a)(1-b), \tag{17}$$

Algorithm 1 Signed pair compositing at pixel $\mathbf{x}$.

Require: contributors in sort order with α_i, $\mathbf{c}_i$, $\bar{z}_i$, V_i, Q_i, $\mathbf{d}_i$; background color $\mathbf{c}_{\mathrm{bg}}$

$C \leftarrow \mathbf{0}, T \leftarrow 1, k \leftarrow \varnothing$ ▷ k: pending first member
for each contributor i with $\alpha_i \geq 1/255$ **do**
  if $k = \varnothing$ **then**
    $k \leftarrow i$
  else
    $(A, B) \leftarrow (k, i)$; compute p, h (Eq. (6)), t (Eq. (7)), O (Eq. (8)), w_A, w_B, τ (Eq. (9))
    $C \leftarrow C + T\,(w_A \mathbf{c}_A + w_B \mathbf{c}_B)$; $T \leftarrow T\,\tau$; $k \leftarrow \varnothing$
    if $T < 10^{-4}$ **then**
      break ▷ tested after each pair (Appendix F)
if $k \neq \varnothing$ **then** ▷ odd tail, Eq. (4)
  $C \leftarrow C + T\alpha_k \mathbf{c}_k$; $T \leftarrow T(1 - \alpha_k)$
return $C + T\,\mathbf{c}_{\mathrm{bg}}$

which are the weights of standard compositing in the orders (A, B) and (B, A), mixed with probabilities p and $1 - p$.

(e) Limits. As $|\delta|/\sqrt{V_\delta} \to \infty$, the exponential in Eq. (6) gives $h \to 0$, and $|t| < 1$ by (a), so $O \to ab$. Hence the weights tend to those of (d), with $p \to 1$ for $\delta/\sqrt{V_\delta} \to -\infty$ and $p \to 0$ for $\delta/\sqrt{V_\delta} \to +\infty$. The first limit gives $w_A = a$ and $w_B = b(1 - a)$, the second gives $w_A = a(1 - b)$ and $w_B = b$, and both give $\tau = (1 - a)(1 - b)$.

Bound on the coupled leak. Let $\beta_A = \mathbf{d}_A^\top Q_A \mathbf{d}_A$, $\beta_B = \mathbf{d}_B^\top Q_B \mathbf{d}_B$, and $\beta = \beta_A + \beta_B$. By Eq. (3), $a \leq e^{-\beta_A/2}$ and $b \leq e^{-\beta_B/2}$, so $ab \leq e^{-\beta/2}$. Since $K = (Q_A + Q_B)^{-1} \preceq Q_A^{-1}$, we have $\mathbf{u}^\top K \mathbf{u} = \mathbf{d}_A^\top Q_A K Q_A \mathbf{d}_A \leq \beta_A$, and likewise $\mathbf{v}^\top K \mathbf{v} \leq \beta_B$. With $\eta = \mathbf{u}^\top K \mathbf{u} + \mathbf{v}^\top K \mathbf{v} \leq \beta$, the proof of (a) gives $|t| \leq \eta/(2 + \eta) \leq \beta/(2 + \beta)$. With $h \leq 1$, $|ht|\,ab \leq g(\beta) = \beta\, e^{-\beta/2}/(2 + \beta)$. On $\beta \geq 0$, g is maximized at $\beta = \sqrt{5} - 1$, the positive root of $\beta^2 + 2\beta - 4 = 0$, where $g < 0.21$. Hence the coupling changes the pair transmittance $\tau = (1 - a)(1 - b)(1 + ht\,ab)$ by at most 21% of its independent value.

Why disjoint pairs. Overlapping windows would count each contributor twice. Coupling every contributor to both neighbors would need a joint law for three or more coverages. Disjoint pairs avoid both problems with a closed-form joint coverage and an exact partition.

Per-pixel loop. Algorithm 1 summarizes signed pair compositing at one pixel.

D The Footprint Coupling

The coupling t of Eq. (7) is a closed-form function of the two footprints and the pixel. It has the following properties.

- **Symmetry.** Exchanging A and B leaves t unchanged, so t does not depend on the sort order.
- **Vanishing at a center.** If the pixel lies at the center of either footprint, then $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$ and $t = 0$.
- **Affine invariance.** Under an invertible affine change of image coordinates $\mathbf{x} \mapsto L\mathbf{x} + \mathbf{x}_0$, the offsets become $L\mathbf{d}_i$, the precisions $L^{-\top} Q_i L^{-1}$, the gradients $L^{-\top}\mathbf{u}$ and $L^{-\top}\mathbf{v}$, and K becomes $LKL^\top$. The three quadratic forms in Eq. (7), and hence t, are unchanged.
- **Sign.** For isotropic footprints $Q_A = \kappa_A I$ and $Q_B = \kappa_B I$, we have $\mathbf{u}^\top K \mathbf{v} = \kappa_A \kappa_B\, \mathbf{d}_A^\top \mathbf{d}_B/(\kappa_A + \kappa_B)$. Hence t has the sign of $\mathbf{d}_A^\top \mathbf{d}_B$: positive when both centers lie on the same side of the pixel and negative when they lie on opposite sides.

A reading of t. As in Section 4.2, read coverage as the event that a sample point near $\mathbf{x}$ is covered. Let $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, K)$ be the offset of that sample from $\mathbf{x}$, so that the covariance K of the normalized

product of the two footprints sets the size of the neighborhood. Let $Y_A = \mathbf{u}^\top \boldsymbol{\xi} + \zeta_A$ and $Y_B = \mathbf{v}^\top \boldsymbol{\xi} + \zeta_B$ with independent $\zeta_A, \zeta_B \sim \mathcal{N}(0,1)$. To first order, $\mathbf{u}^\top \boldsymbol{\xi}$ and $\mathbf{v}^\top \boldsymbol{\xi}$ are the changes of the two log-footprints under the displacement. Then $\mathrm{Cov}(Y_A, Y_B) = \mathbf{u}^\top K \mathbf{v}$, $\mathrm{Var}\, Y_A = 1 + \mathbf{u}^\top K \mathbf{u}$, and $\mathrm{Var}\, Y_B = 1 + \mathbf{v}^\top K \mathbf{v}$, so $t = 2\,\mathrm{Cov}(Y_A, Y_B)/(\mathrm{Var}\, Y_A + \mathrm{Var}\, Y_B)$ is the concordance correlation coefficient of Y_A and Y_B (Lin, 1989). The coupling is positive when a displacement that brightens one footprint also brightens the other, so that the two tend to cover the same sample points, and negative when it brightens one and dims the other, so that they tend to cover complementary ones. Proposition 1 does not depend on this reading.

Choice of the joint coverage. The Fréchet–Hoeffding bounds $\min(a, b)$ and $\max(0, a + b - 1)$ represent full nesting and exact seams. Copulas that reach them are either non-differentiable where $a = b$ or $a + b = 1$ (e.g., mixtures toward those bounds) or not polynomial in a and b. The Farlie–Gumbel–Morgenstern form of Eq. (8) trades this range for a cheap smooth form (Section 6).

E Derivation of the Nuclear-Norm Loss

A family of spectral losses. The per-channel L_1 loss of 3DGS treats the three color channels of every pixel independently, and its gradient $\mathrm{sign}(E)$ depends on the choice of the RGB basis. Stacking the residuals of a view into $E \in \mathbb{R}^{P \times 3}$ instead exposes the principal directions of the color error. With the singular value decomposition $E = U\,\mathrm{diag}(\mathbf{s})\Gamma^\top$, $s_1 \geq s_2 \geq s_3 \geq 0$, these directions are the columns of Γ. They can carry very different energy, for instance when a brightness error dominates small chromatic errors. Losses that depend on E only through its singular values are invariant to permuting pixels and to any orthogonal change of the color basis, $E \mapsto E\Theta$ with Θ orthogonal. Their one-homogeneous members include a natural one-parameter family, the scaled Schatten norms,

$$\mathcal{L}_{\mathrm{S}q}(E) = \frac{\|E\|_{\mathrm{S}q}}{3^{1/q}\sqrt{P}}, \qquad \|E\|_{\mathrm{S}q} = \Big(\sum_{k=1}^{3} s_k^q\Big)^{1/q}, \qquad q \geq 1, \tag{18}$$

where the scale makes $\mathcal{L}_{\mathrm{S}q}$ equal the RMS error when $s_1 = s_2 = s_3$. The member $q = 2$ is the RMS error itself, and $q = 1$ is $\mathcal{L}_{\mathrm{S}1}$.

Why $q = 1$. For distinct nonzero singular values, the gradient of the Schatten norm is

$$\frac{\partial \|E\|_{\mathrm{S}q}}{\partial E} = U\,\mathrm{diag}\Big(\frac{s_k^{q-1}}{\|E\|_{\mathrm{S}q}^{q-1}}\Big)\Gamma^\top, \tag{19}$$

so the push along the k-th error direction is proportional to s_k^{q-1}. Relative to the dominant direction, a minor direction receives the fraction $(s_k/s_1)^{q-1}$. For $q > 1$, this ratio decays as the spectrum becomes more uneven, so the chromatic directions receive a vanishing share of the gradient once brightness errors dominate. For $q < 1$, it diverges as $s_k \to 0$, and the loss is no longer a norm. Within this family, only $q = 1$ gives every error direction the same push regardless of the spectrum. The gradient is then the whitening form $U\Gamma^\top = E(E^\top E)^{-1/2}$, derived below for full column rank. The choice thus follows from the gradient structure, not from tuning.

Variational form. Let $\Psi = E^\top E = \Gamma\,\mathrm{diag}(s_1^2, s_2^2, s_3^2)\Gamma^\top$, and assume first that $\Psi \succ 0$. We show that $\|E\|_*$ is the minimum of

$$f(\Omega) = \tfrac{1}{2}\big[\,\mathrm{tr}(\Psi\Omega^{-1}) + \mathrm{tr}\,\Omega\big], \qquad \Omega \succ 0. \tag{20}$$

Both terms are convex on the positive-definite cone, so any stationary point is a global minimizer. Since $\mathrm{d}\,\mathrm{tr}(\Psi\Omega^{-1}) = -\,\mathrm{tr}(\Omega^{-1}\Psi\Omega^{-1}\,\mathrm{d}\Omega)$ and $\mathrm{d}\,\mathrm{tr}\,\Omega = \mathrm{tr}\,\mathrm{d}\Omega$, stationarity reads

$$\Omega^{-1}\Psi\Omega^{-1} = I \quad \Longleftrightarrow \quad \Omega^2 = \Psi, \tag{21}$$

whose unique positive-definite solution is $\Omega^\star = \Psi^{1/2} = \Gamma\,\mathrm{diag}(s_1, s_2, s_3)\Gamma^\top$. Substituting gives $f(\Omega^\star) = \frac{1}{2}[\mathrm{tr}\,\Psi^{1/2} + \mathrm{tr}\,\Psi^{1/2}] = \sum_k s_k = \|E\|_*$. The two terms of $f(\Omega^\star)$ are equal, so $\|E\|_* = \mathrm{tr}(\Psi(\Omega^\star)^{-1}) = \sum_{\mathbf{x}} \boldsymbol{\varepsilon}_{\mathbf{x}}^\top (\Omega^\star)^{-1} \boldsymbol{\varepsilon}_{\mathbf{x}}$, where $\boldsymbol{\varepsilon}_{\mathbf{x}}$ is the residual of pixel $\mathbf{x}$. This is a sum of squared Mahalanobis norms under a per-view color metric. If Ψ is singular, the minimum is not attained, but the infimum is still $\|E\|_*$, approached by $\Omega = \Psi^{1/2} + \rho I$ as $\rho \to 0^+$.

Gradient. By the envelope theorem, the derivative of the minimum with respect to E is the partial derivative of f at $\Omega^\star$ with Ω held fixed:

$$\frac{\partial \|E\|_*}{\partial E} = E\,(\Omega^\star)^{-1} = E\,(E^\top E)^{-1/2}. \tag{22}$$

This agrees with the known gradient $U\Gamma^\top$ of the nuclear norm for full column rank. Each pixel row is multiplied by the same 3×3 matrix, so the gradient whitens across color channels.

Isotropic case. If $E^\top E = \gamma I$, then $\|E\|_* = 3\sqrt{\gamma}$ and $\|E\|_F = \sqrt{3\gamma}$, so $\mathcal{L}_{\mathrm{S1}} = \sqrt{\gamma/P} = \|E\|_F/\sqrt{3P}$, the RMS error.

Computation. In training, we compute the nuclear norm with a standard single-precision matrix-norm routine and differentiate it by automatic differentiation. Only the singular values are differentiated, so the backward pass is $U\,\mathrm{diag}(\partial\mathcal{L}/\partial\mathbf{s})\,\Gamma^\top$. It contains none of the $1/(s_k^2 - s_l^2)$ terms of the singular-vector derivative, so repeated singular values, including the isotropic case above, need no special handling. The rank-deficient case, in which only a subgradient exists, receives no special treatment.

F Experimental Details

Data. We use the test views of Kerbl et al. (2023), every eighth image of each scene, and train on the remaining images. The Mip-NeRF360 outdoor scenes (Bicycle, Flowers, Garden, Stump, Treehill) use the images downsampled by four, and the indoor scenes (Room, Counter, Kitchen, Bonsai) use those downsampled by two. Tanks&Temples and Deep Blending use the images as provided. All methods read the same image folders, and all use a black background.

Training. 3DGS, the ablation rows, and our method train for 30K iterations with the densification schedule, learning rates, and spherical-harmonic degree 3 of Kerbl et al. (2023), and with $\lambda = 0.2$ for the D-SSIM term. Apart from the tile test below, our method changes only the compositing (Section 4.2) and the loss (Section 4.3), and each ablation row changes only the components it marks. The ablation rows and our full method use the tile test of Speedy-Splat (Hanson et al., 2025), without its pruning; the 3DGS row uses the original tile assignment of 3DGS. The test assigns a Gaussian to the tiles that its cutoff ellipse $\alpha_i(\mathbf{x}) \geq 1/255$ intersects, which for opacities above about 0.35 can include tiles beyond the 3σ bounding box of 3DGS. Rendering the trained models of our method with the bounding box instead leaves PSNR, SSIM, and LPIPS unchanged at the reported precision in every scene.

Implementation. Restoring the depth row adds no matrix product to the projection, since the reference rasterizer already forms the full 3×3 product. Computing $\mathbf{r}_i$ and V_i adds one matrix–vector product and one inner product per Gaussian (Appendix B). The three values $\mathbf{r}_i$ and V_i are stored in one four-float record per Gaussian. The opacity clamp, the opacity cutoff $1/255$, the low-pass dilation, and the termination threshold 10^{-4} follow the reference implementation of 3DGS, and V_i is floored at 10^{-6}. Early termination is tested after each pair has been accumulated, when T changes.

Baselines. 3DGS uses the official implementation. StopThePop uses its hierarchical per-pixel resorting with the queue sizes of its release and the default training settings of its code, which keep the opacity reset of 3DGS. Its paper reports results with an opacity decay of 0.9995 instead, which reduces the number of Gaussians. EVER and Vol3DGS use their released code and settings. All baselines use the D-SSIM weight $\lambda = 0.2$ by default. We use the public implementations of 3DGS, StopThePop, EVER, and Vol3DGS.

Metrics. PSNR, SSIM, and LPIPS with the VGG backbone are computed on the test views at the training resolution with the evaluation script of Kerbl et al. (2023), which each baseline repository includes with the same LPIPS module. Storage is the size of the saved point cloud, and the number of Gaussians is counted at 30K iterations. Rendering speed is the inverse of the mean time of one render call, including spherical-harmonic evaluation. We time it with CUDA events over three passes through the test views after one warm-up pass.

Table 5: Storage (MiB, saved point cloud), number of Gaussians (#G, millions) at 30K iterations, and rendering speed (FPS) on one NVIDIA L40 GPU, averaged over the scenes of each dataset. All methods are timed with the same protocol, and FPS repeats Table 1.

	Mip-NeRF360			Tanks&Temples			Deep Blending		
Method	MiB↓	#G	FPS↑	MiB↓	#G	FPS↑	MiB↓	#G	FPS↑
3DGS	647	2.73	114	372	1.57	154	585	2.47	117
StopThePop	783	3.31	97	428	1.81	144	666	2.82	104
EVER	1107	4.68	15	1709	7.22	14	662	2.80	20
Vol3DGS	748	3.11	59	256	1.06	102	862	3.59	60
Ours	666	2.81	122	395	1.67	155	713	3.01	106

G WHERE THE PAIR TERMS ACT

Figure 4 shows, for two held-out views of our trained models, the share of the composited weight $T(w_A+w_B)$ at each pixel that falls in pairs whose order is uncertain ($0.1 \le p \le 0.9$), in pairs whose per-pixel order is reversed with respect to the sort ($p < 0.1$), and in pairs with a coupling $|ht| > 0.1$. The maps are rendered with the same pairing and pair quantities as signed pair compositing. Uncertain and reversed orders occur over large parts of the surfaces in both views, and the coupling concentrates along structure boundaries.

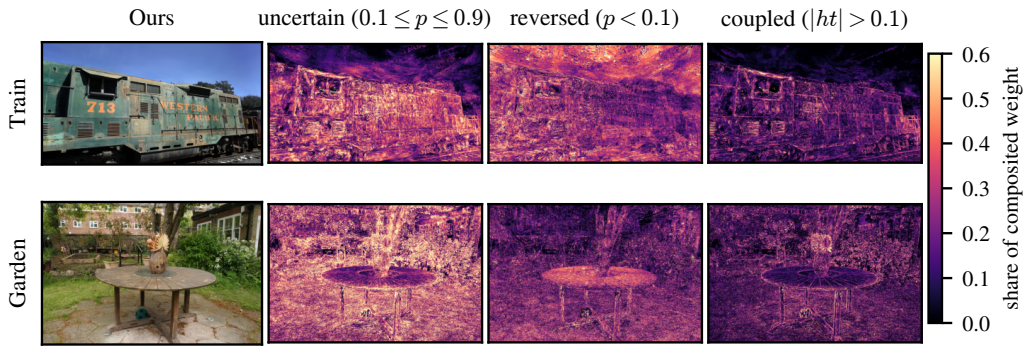


Figure 4: Share of the composited weight in pairs with an uncertain order ($0.1 \le p \le 0.9$), a reversed order ($p < 0.1$), and a coupling $|ht| > 0.1$, for held-out views of Train (top) and Garden (bottom). Reversed orders concentrate in the sky above the train and on the tabletop, and the coupling is largest along edges.

We render our trained models with $t = 0$ (Table 6), which keeps the pairing and the soft order p and sets the joint coverage to $O = ab$ (Proposition 1(d)). The dataset averages are unchanged at the reported precision, consistent with the bound on the coupled leak (Appendix C) and with the coupling being largest along structure boundaries (Figure 4). The models were trained with the coupling, so this test does not measure its effect on training.

Table 6: Our trained models rendered with signed pair compositing, with the coupling and with $t = 0$, which keeps p and sets $O = ab$ (Section 4.2). The first row repeats the Ours row of Table 3. Averages over the scenes of each dataset at 30K iterations.

	Mip-NeRF360			Tanks&Temples			Deep Blending		
Rendered	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
signed pair	27.66	0.816	0.213	24.10	0.855	0.165	29.74	0.906	0.236
signed pair, $t = 0$	27.66	0.816	0.213	24.10	0.855	0.165	29.74	0.906	0.236

H ADDITIONAL QUALITATIVE RESULTS

Figure 5 shows three more held-out views, one region each. In these regions, our method preserves more detail, such as the highlights in Bonsai and Counter.

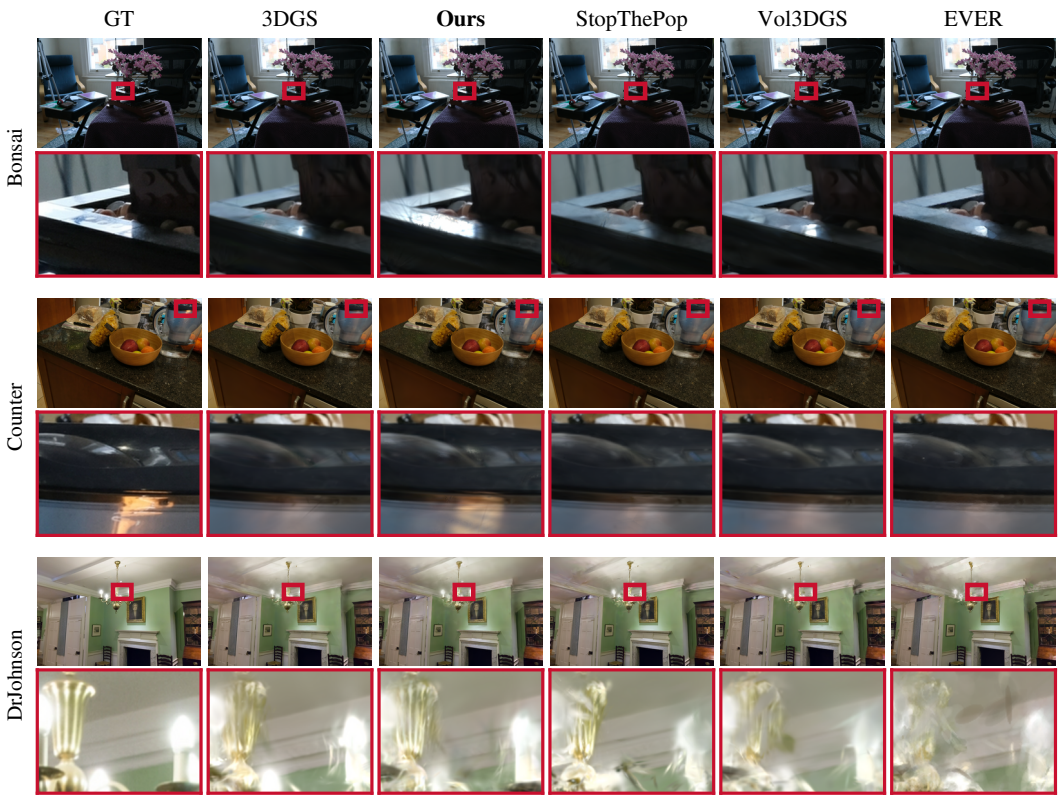


Figure 5: Additional held-out views. From top to bottom: Bonsai and Counter (Mip-NeRF360) and DrJohnson (Deep Blending).

Figure 6 shows the regions of Figure 3 for the rows of Table 2, which add signed pair compositing and the nuclear-norm loss to 3DGS separately and together.

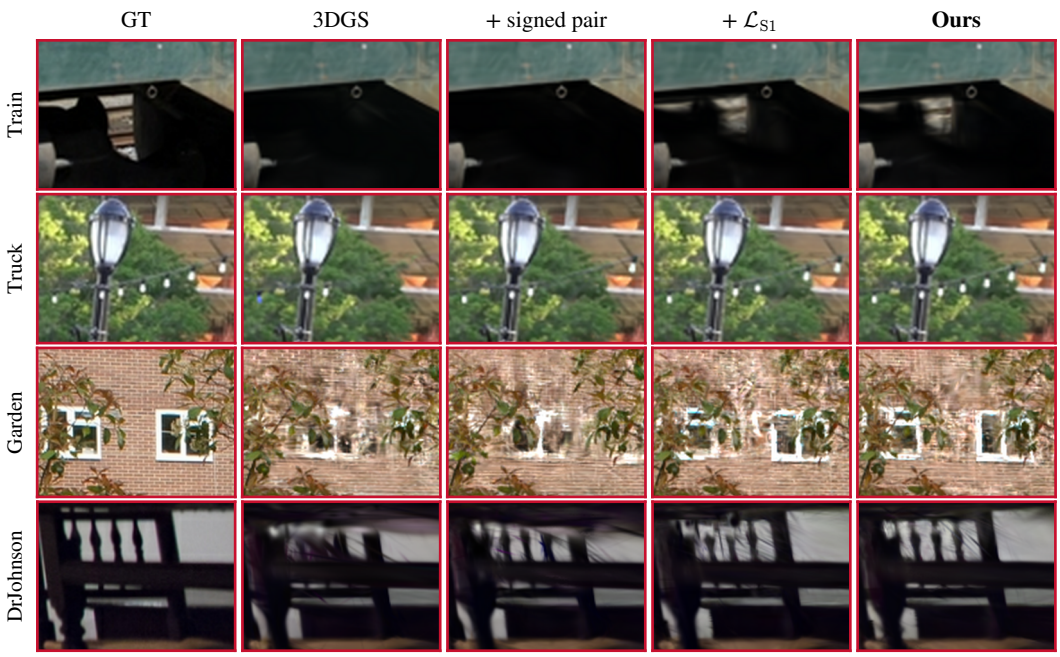


Figure 6: Regions of Figure 3 for the ablation rows of Table 2. From top to bottom: Train, Truck, Garden, and DrJohnson.

I Per-Scene Results

Tables 7–12 list the per-scene values averaged in Tables 1 and 2; the last three rows of each table are rows 2–4 of Table 2. Averages are computed from unrounded per-scene values.

Table 7: Per-scene PSNR↑ on Mip-NeRF360.

Method	Bicycle	Flowers	Garden	Stump	Treehill	Room	Counter	Kitchen	Bonsai	Avg.
3DGS	25.24	21.52	27.47	26.65	22.58	31.64	29.08	31.37	32.28	27.54
StopThePop	25.24	21.59	27.29	26.66	22.42	30.84	28.82	31.19	31.98	27.34
EVER	25.27	21.78	27.37	26.43	22.82	31.43	28.87	31.30	32.33	27.51
Vol3DGS	25.00	21.41	27.21	26.33	21.81	31.82	28.86	31.39	32.13	27.33
3DGS + signed pair	25.22	21.55	27.44	26.64	22.58	31.60	29.06	31.44	32.26	27.53
3DGS + $\mathcal{L}_{S1}$	25.35	21.80	27.63	26.69	22.51	31.97	29.22	31.83	32.58	27.73
Ours	25.34	21.85	27.63	26.73	22.42	31.72	29.24	31.72	32.31	27.66

Table 8: Per-scene PSNR↑ on Tanks&Temples and Deep Blending.

	Tanks&Temples			Deep Blending		
Method	Train	Truck	Avg.	DrJohnson	Playroom	Avg.
3DGS	22.13	25.48	23.80	29.48	30.20	29.84
StopThePop	21.43	24.83	23.13	29.51	30.41	29.96
EVER	21.97	25.09	23.53	29.47	29.49	29.48
Vol3DGS	22.00	25.21	23.61	29.29	29.98	29.64
3DGS + signed pair	22.13	25.46	23.79	29.45	30.15	29.80
3DGS + $\mathcal{L}_{S1}$	22.32	25.82	24.07	29.33	30.23	29.78
Ours	22.42	25.78	24.10	29.34	30.14	29.74

Table 9: Per-scene SSIM↑ on Mip-NeRF360.

Method	Bicycle	Flowers	Garden	Stump	Treehill	Room	Counter	Kitchen	Bonsai	Avg.
3DGS	0.766	0.605	0.868	0.772	0.634	0.921	0.909	0.928	0.942	0.816
StopThePop	0.768	0.606	0.866	0.775	0.637	0.919	0.908	0.927	0.942	0.816
EVER	0.773	0.638	0.866	0.781	0.656	0.922	0.910	0.926	0.944	0.824
Vol3DGS	0.769	0.626	0.864	0.773	0.619	0.928	0.914	0.930	0.945	0.818
3DGS + signed pair	0.768	0.606	0.868	0.772	0.635	0.920	0.909	0.928	0.942	0.816
3DGS + $\mathcal{L}_{\mathrm{S1}}$	0.765	0.611	0.868	0.773	0.632	0.921	0.907	0.929	0.942	0.816
Ours	0.765	0.612	0.868	0.773	0.630	0.920	0.907	0.928	0.939	0.816

Table 10: Per-scene SSIM↑ on Tanks&Temples and Deep Blending.

	Tanks&Temples			Deep Blending		
Method	Train	Truck	Avg.	DrJohnson	Playroom	Avg.
3DGS	0.821	0.885	0.853	0.905	0.909	0.907
StopThePop	0.811	0.882	0.846	0.906	0.911	0.909
EVER	0.846	0.892	0.869	0.906	0.907	0.907
Vol3DGS	0.819	0.883	0.851	0.903	0.908	0.905
3DGS + signed pair	0.820	0.885	0.853	0.905	0.909	0.907
3DGS + $\mathcal{L}_{\mathrm{S1}}$	0.824	0.887	0.855	0.903	0.908	0.906
Ours	0.824	0.886	0.855	0.903	0.908	0.906

Table 11: Per-scene LPIPS↓ on Mip-NeRF360.

Method	Bicycle	Flowers	Garden	Stump	Treehill	Room	Counter	Kitchen	Bonsai	Avg.
3DGS	0.208	0.336	0.106	0.215	0.324	0.217	0.199	0.126	0.203	0.215
StopThePop	0.205	0.334	0.105	0.211	0.317	0.214	0.196	0.125	0.200	0.212
EVER	0.199	0.287	0.109	0.197	0.282	0.207	0.185	0.125	0.188	0.198
Vol3DGS	0.193	0.284	0.109	0.205	0.306	0.203	0.184	0.123	0.190	0.200
3DGS + signed pair	0.205	0.334	0.106	0.213	0.320	0.213	0.196	0.125	0.201	0.212
3DGS + $\mathcal{L}_{\mathrm{S1}}$	0.210	0.332	0.108	0.214	0.324	0.216	0.200	0.125	0.205	0.215
Ours	0.210	0.329	0.107	0.213	0.324	0.212	0.198	0.124	0.203	0.213

Table 12: Per-scene LPIPS↓ on Tanks&Temples and Deep Blending.

	Tanks&Temples			Deep Blending		
Method	Train	Truck	Avg.	DrJohnson	Playroom	Avg.
3DGS	0.196	0.142	0.169	0.236	0.240	0.238
StopThePop	0.202	0.140	0.171	0.231	0.232	0.232
EVER	0.151	0.118	0.135	0.235	0.246	0.240
Vol3DGS	0.201	0.147	0.174	0.240	0.243	0.242
3DGS + signed pair	0.193	0.139	0.166	0.234	0.237	0.236
3DGS + $\mathcal{L}_{\mathrm{S1}}$	0.193	0.141	0.167	0.238	0.240	0.239
Ours	0.190	0.139	0.165	0.236	0.236	0.236