SalientGS: Supplementary Material

Tianyu Xiong
School of Computer Science, Northwestern Polytechnical University
Xi'an, China

Rui Li
CEMSE, King Abdullah University of Science and Technology
Thuwal, Saudi Arabia

Suning Ge
School of Computer Science, Northwestern Polytechnical University
Xi'an, China

Jiaqi Yang
School of Computer Science, Northwestern Polytechnical University
Xi'an, China

1 Overview

This supplementary material accompanies the main paper *"SalientGS: Unified SfM-to-3DGS with Importance-Guided MCMC Gaussian Allocation."* We provide:

- A released-code cross-benchmark verification with an explicit macro-average protocol (§2).
- A large-scene benchmark comparison (§3).
- Full SfM runtime tables for the Courthouse scene (§4).
- Full per-scene qualitative comparisons on all nine Mip-NeRF 360 scenes (§5).
- Per-image LPIPS analysis across ablation variants (§6).
- Threshold-sensitivity sweeps for τ_{imp} and the number of scoring views K (§7).
- Analysis of the SalientGS front-end: track-length distributions and a controlled track-cap ablation establishing that sufficient multi-view track coverage is a prerequisite for high reconstruction quality (§8).

2 Released-Code Cross-Benchmark Verification

Table 1 macro-averages the three dataset-level entries in the main paper, assigning equal weight to Mip-NeRF 360, Deep Blending, and Tanks & Temples. This is a cross-dataset summary, not a replacement for the individual benchmark results. Failed scenes remain in the corresponding dataset aggregate; notably, the *drjohnson* failures of GloSplat-A and VGGT-X are included rather than discarded. Under this explicit protocol, SalientGS has the best macro-average quality and runtime in the comparison. Its advantage reflects consistency across datasets: it is the only method with a top-two LPIPS result on every benchmark.

The released per-scene CSV uses seed 42, exactly 30K optimization steps, and a 1.5M Gaussian cap. All 13 runs execute feature extraction, matching, SfM, and training; reference sparse models are not used. For Mip-NeRF 360, the scene-equal aggregate used in the main comparison is 28.822 / 0.853 / 0.148 (PSNR / SSIM / LPIPS). Pooling all held-out images instead gives 29.607 / 0.874 / 0.140 because scenes contain different numbers of test images. We report the latter only as a diagnostic and never compare it against scene-averaged baselines.

Table 1: Cross-benchmark macro-average of the three dataset-level entries in the main comparison. Time is end-to-end minutes; each benchmark receives equal weight.

Method	Time↓	PSNR↑	SSIM↑	LPIPS↓
3DGS	28.35	26.98	0.855	0.211
3DGS-MCMC	28.85	27.40	0.872	0.191
Mini-Splatting	24.37	26.92	0.857	0.214
Speedy-splat	21.15	26.57	0.832	0.270
Taming-3DGS	14.71	26.96	0.840	0.251
DashGaussian	15.93	27.13	0.859	0.214
FastGS-big	13.54	27.48	0.861	0.211
GloSplat-A	22.01	23.15	0.750	0.265
VGGT-X	64.02	22.60	0.741	0.287
SalientGS	**10.62**	**27.65**	**0.876**	**0.147**

3 Large-Scene Benchmark Comparison

Table 3 reports SSIM, PSNR, and LPIPS on Building, Rubble, Residence, and Sci-Art; each competing method is cited in the first column. Our method achieves the best PSNR on Building, Rubble, and Residence, the best LPIPS on Building, Residence, and Sci-Art, and the best SSIM on Rubble, indicating that the unified SfM-to-3DGS pipeline remains highly competitive even on challenging large-scale scenes. Under the same large-scene training setup, SalientGS averages only about 1 h wall-clock training and ∼5 GB peak GPU memory, compared to about 3 h and ∼14 GB for CityGS-X [5].

4 SfM Runtime Details

Table 4 reports the full SfM component runtime on the Courthouse scene with 250–1000 input images, including GloSplat-F (XFeat+LightGlue with retrieval top-5; see §8 for details). These numbers correspond to the trend summarized by Figure 5 in the main paper.

5 Per-Scene Qualitative Comparisons

We present qualitative comparisons on all nine Mip-NeRF 360 [1] scenes. Each figure shows the rendered output from SalientGS (Ours) alongside ablation variants (Vanilla MCMC, Standard ADC, Freeze Poses) and a baseline method, with zoomed crops at the region of maximum LPIPS difference.

Table 2: Released-code verification by scene. Time includes feature extraction, matching, SfM, and 30K-step training on one RTX PRO 6000 Blackwell GPU. Every final model contains 1.5M Gaussians.

Dataset	Scene	PSNR↑	SSIM↑	LPIPS↓	Time (min)↓
Mip-NeRF 360	bicycle	27.125	0.828	0.135	10.37
	bonsai	33.207	0.955	0.106	13.53
	counter	29.604	0.919	0.136	14.03
	flowers	22.826	0.661	0.274	10.48
	garden	29.188	0.894	0.074	10.33
	kitchen	32.786	0.943	0.076	13.57
	room	31.945	0.928	0.150	12.72
	stump	28.310	0.836	0.123	10.86
	treehill	24.407	0.714	0.260	10.20
Deep Blending	drjohnson	28.824	0.892	0.210	10.69
	playroom	30.153	0.920	0.156	9.39
Tanks & Temples	train	23.003	0.846	0.128	10.32
	truck	26.306	0.893	0.090	9.73

Table 3: Large-scene benchmark comparison. Metrics are SSIM↑, PSNR↑, and LPIPS↓. Red / orange / yellow denote the best / second-best / third-best values in each column.

	Building			Rubble			Residence			Sci-Art		
Method	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS
w/ Geometric Optimization												
SuGaR [6]	0.507	17.76	0.455	0.577	20.69	0.453	0.603	18.74	0.406	0.698	18.60	0.349
NeuS [15]	0.463	18.01	0.611	0.480	20.46	0.618	0.503	17.85	0.533	0.633	18.62	0.472
Neuralangelo [9]	0.582	17.89	0.322	0.625	20.18	0.314	0.644	18.03	0.263	0.769	19.10	0.231
PGSR [2]	0.480	16.12	0.573	0.728	23.09	0.334	0.746	20.57	0.289	0.799	19.72	0.275
VCR-Gaus [3]	0.502	19.56	0.502	0.541	21.34	0.428	0.623	20.59	0.359	0.665	19.31	0.465
CityGaussianV2 [12]	0.650	19.07	0.397	0.720	23.75	0.322	0.769	21.15	0.234	0.810	20.66	0.266
UrbanGS [8]	0.802	22.82	0.208	0.791	26.25	**0.210**	**0.823**	22.48	0.205	0.824	22.62	0.279
CityGS-X [5]	**0.817**	22.76	0.191	0.823	26.15	**0.210**	0.819	22.44	0.194	**0.867**	22.77	0.179
w/o Geometric Optimization												
Mega-NeRF [14]	0.547	20.92	0.454	0.553	24.06	0.508	0.628	22.08	0.401	0.770	25.60	0.312
Switch-NeRF [13]	0.579	21.54	0.397	0.562	24.31	0.478	0.654	22.57	0.352	0.795	**26.51**	0.271
VastGaussian [10]	0.728	21.80	0.225	0.742	25.20	0.264	0.699	21.01	0.261	0.761	22.64	0.261
3DGS [7]	0.738	22.53	0.214	0.725	25.51	0.316	0.745	22.36	0.247	0.791	24.13	0.262
DoGaussian [4]	0.759	22.73	0.204	0.765	25.78	0.257	0.740	21.94	0.244	0.804	24.42	0.219
CityGaussian [11]	0.778	21.55	0.246	0.813	25.77	0.228	0.813	22.00	0.211	0.837	21.39	0.230
SalientGS (Ours)	0.804	**24.19**	**0.181**	**0.824**	**26.98**	0.213	0.816	**22.67**	**0.193**	0.857	22.91	**0.173**

Table 4: SfM runtime comparison (seconds). We compare the SfM component runtime across different methods on the Courthouse scene with varying numbers of input images. SalientGS combines Fisher Vector retrieval with first-order SfM optimization to achieve significant speedups over traditional and learning-based approaches.

# Images	COLMAP	MASt3R	GloSplat-A	VGGT-X	GloSplat-F	SalientGS
250	644.6	312.2	222.2	53.9	100.2	**47.8**
500	1165.0	638.0	408.2	113.3	191.9	**98.0**
750	2226.3	978.5	855.7	249.6	250.0	**138.5**
1000	4349.2	1308.2	1371.3	398.9	327.8	**186.1**

Outdoor Scenes. Figures 1–5 cover the five outdoor scenes. SalientGS consistently resolves fine-grained texture—foliage, bark, grass—that Vanilla MCMC blurs and Standard ADC can render with color artifacts on most scenes. The advantage is most pronounced on *stump* and *flowers*, where high-frequency detail benefits most from importance-guided capacity allocation; on *garden*, SalientGS and Standard ADC perform comparably.

Indoor Scenes. Figures 6–9 cover the four indoor scenes. On *counter*, *bonsai*, and *kitchen*, SalientGS outperforms all ablation variants, though the margins are smaller than on outdoor scenes due to simpler geometry and more uniform textures. On *room*, SalientGS underperforms VGGT-X on flat, textureless surfaces where first-order SfM initialization

Table 5: Mean per-image LPIPS (↓) on Mip-NeRF 360 scenes. Scene rows show per-scene means. We report both the scene-equal mean used in the main comparison and the pooled test-image diagnostic. Best displayed result per scene in bold.

Scene	SalientGS	Vanilla MCMC	Std. ADC	Freeze Poses
Outdoor Scenes				
bicycle	**0.135**	0.152	0.153	0.173
flowers	0.274	**0.264**	0.309	0.293
garden	0.074	0.074	**0.066**	0.081
stump	**0.123**	0.140	0.145	0.151
treehill	**0.260**	**0.260**	0.269	0.270
Indoor Scenes				
bonsai	**0.106**	0.169	0.171	0.170
counter	0.136	**0.132**	0.148	0.134
kitchen	**0.076**	0.236	0.167	0.173
room	**0.150**	0.271	0.252	0.271
Scene mean	**0.148**	0.189	0.187	0.192
Pooled images	**0.140**	0.189	0.187	0.191

Table 6: Threshold-sensitivity sweep on Mip-NeRF 360. We vary the importance threshold τ_{imp} and the number of scoring views K. The selected full-model configuration is marked as *default*.

Group	Value	PSNR↑	SSIM↑	LPIPS↓
τ_{imp}	1	28.688	0.8444	0.159
	3	28.685	0.8448	0.157
	5 (*default*)	**28.822**	**0.8531**	**0.148**
	10	28.732	0.8466	0.153
	20	28.759	0.8477	0.151
K	3	28.696	0.8457	0.156
	5	28.689	0.8455	0.156
	10 (*default*)	**28.822**	**0.8531**	**0.148**
	15	28.714	0.8450	0.157
	20	28.679	0.8446	0.157

provides weaker geometric constraints than learned pose estimators (see main paper, Qualitative Results and Conclusion), though it still leads all ablation variants in the per-image LPIPS table.

6 Per-Image LPIPS Analysis

Table 5 combines the released SalientGS verification with the controlled camera-ready ablation measurements. At full precision, SalientGS improves over vanilla MCMC on seven scenes and trails it on *flowers* and *counter*; *garden* and *treehill* appear tied after rounding. Its nine-scene mean is substantially lower (0.148 vs. 0.189). Because the columns come from two documented experiment snapshots, this table is descriptive; causal allocator conclusions are drawn from the matched controlled tables in §9.

7 Threshold Sensitivity Sweep

Table 6 reports the Mip-NeRF 360 sensitivity sweep referenced in the main paper. We vary the importance threshold τ_{imp} and the number of views K used for multi-view score computation while keeping the rest of the pipeline fixed.

The released setting uses τ_{imp}=5 and K=10. PSNR and SSIM vary mildly across nearby settings, while LPIPS is more sensitive to perceptual blur and high-frequency detail; we therefore report the three metrics separately rather than summarizing robustness with a single PSNR range.

8 Front-End Track Analysis and Track-Length Ablation

A natural question is whether the quality advantage of SalientGS over retrieval-only front-ends (such as GloSplat-F, which uses XFeat+LightGlue with top-5 retrieval) originates primarily from better Gaussian allocation or from a richer multi-view track structure. We investigate this through two complementary experiments: (1) a comparison of the 3D-point track-length distributions produced by three front-ends on the `bicycle` scene, and (2) a controlled track-length cap ablation that holds the front-end fixed and varies only the maximum number of observations retained per track before global positioning and bundle adjustment.

Track-Length Distributions. Table 7 compares the track-length distributions produced on the `bicycle` scene by three pipelines: GloSplat-A (SIFT exhaustive), GloSplat-F (XFeat+LightGlue, retrieval top-5), and SalientGS (SIFT + Fisher Vector retrieval, k=20). SalientGS reconstructs ∼4% fewer 3D points than GloSplat-A (51.8k vs. 54.3k), but its track-length distribution remains very close: median and max are nearly identical, the mean is only marginally lower (4.47 vs. 4.69), and all four length-bucket percentages differ by at most 0.6 pp. In contrast, GloSplat-F—which uses only 5 retrieval neighbors—yields ∼16% fewer points with a substantially shorter track distribution (median 3 vs. 4, max 20 vs. 74), and long tracks (≥9 observations) are nearly absent (2.1% vs. 9.5%). The MST connectivity guarantee built into SalientGS's Fisher Vector retrieval (Sec. 3.1 of the main paper) is responsible for preserving the exhaustive-like track structure: the k=20 candidate graph, augmented with MST edges, achieves far denser coverage than a bare top-5 list.

Table 7: Track-length distribution on the `bicycle` scene for GloSplat-A (SIFT exhaustive), GloSplat-F (XFeat+LightGlue, retrieval top-5), and SalientGS (Fisher Vector, k=20).

	GloSplat-A	GloSplat-F	SalientGS
Total points	54,275	45,507	51,847
Mean / Median	4.69 / 4	3.75 / 3	4.47 / 4
Max	74	20	71
2–3	26,861 (49.5%)	27,182 (59.7%)	25,814 (49.8%)
4–8	22,036 (40.6%)	17,366 (38.2%)	21,124 (40.7%)
9–16	4,694 (8.6%)	948 (2.1%)	4,133 (8.0%)
17+	684 (1.3%)	11 (0.0%)	776 (1.5%)

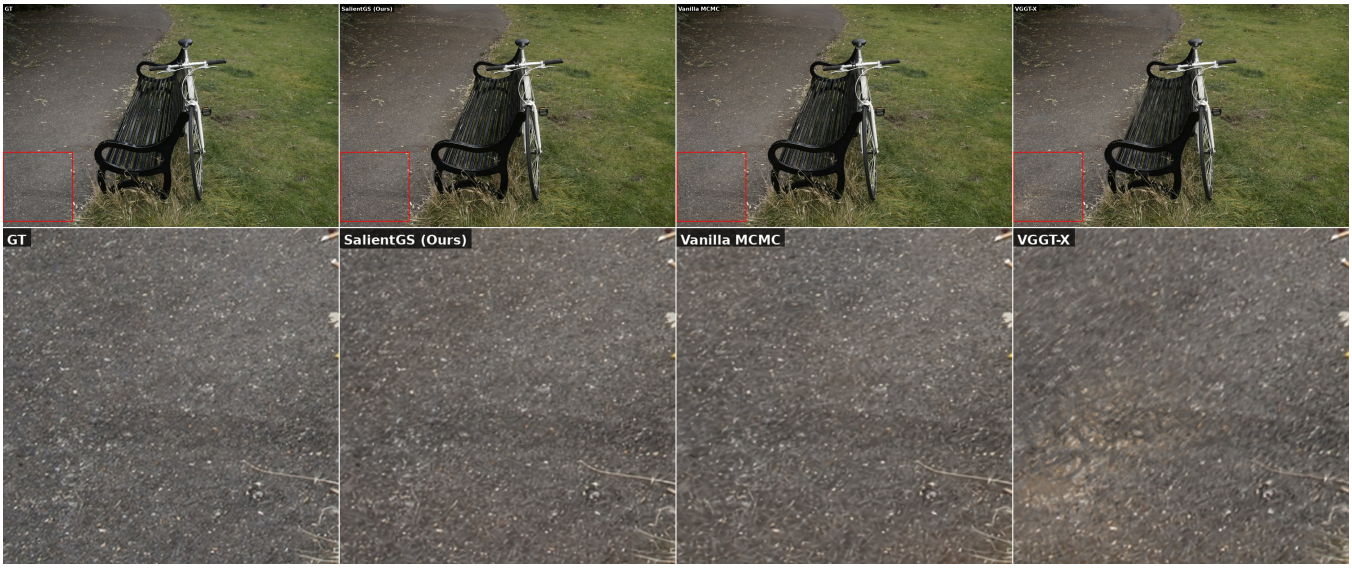

Figure 1: Per-scene qualitative comparison: *bicycle*.

Figure 2: Per-scene qualitative comparison: *flowers*.

Track-Length Cap Ablation. To establish causality between track length and downstream rendering quality, we conduct a controlled ablation: we fix each method's front-end (i.e., the same pair graph and matches) and vary only the maximum number of observations retained per track before global positioning and bundle adjustment. Table 8 reports the effect on Mip-NeRF 360 (averaged over 9 scenes). The L=2 condition reduces every track to a pairwise observation—simulating a worst-case version of the track shortening seen in GloSplat-F—while L=all retains the full distribution. For GloSplat (SIFT exhaustive front-end), capping at L=2 causes a 1.40 dB PSNR drop (28.86 → 27.46), confirming that long multi-view tracks provide substantial geometric constraints that improve triangulation robustness and stabilize bundle adjustment. In the matched track-cap experiment, SalientGS exhibits a similar degradation of 1.29 dB (28.82 → 27.53) under the same L=2 cap. Once every track is reduced to a pairwise observation, both pipelines operate on geometrically equivalent constraints regardless of their front-end or allocation strategy, leaving the underlying bundle adjustment quality as the dominant factor.

Taken together, these results support two conclusions. First, SalientGS's Fisher Vector retrieval with k=20 and MST connectivity is sufficient to recover an exhaustive-quality

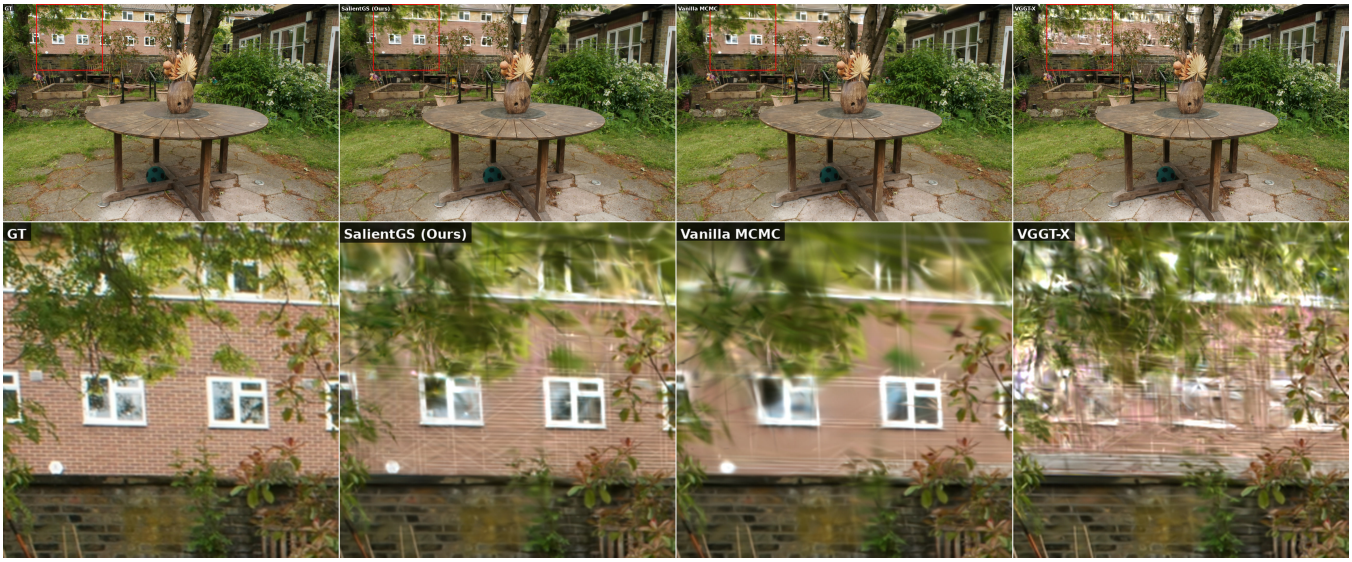

Figure 3: Per-scene qualitative comparison: *garden*.

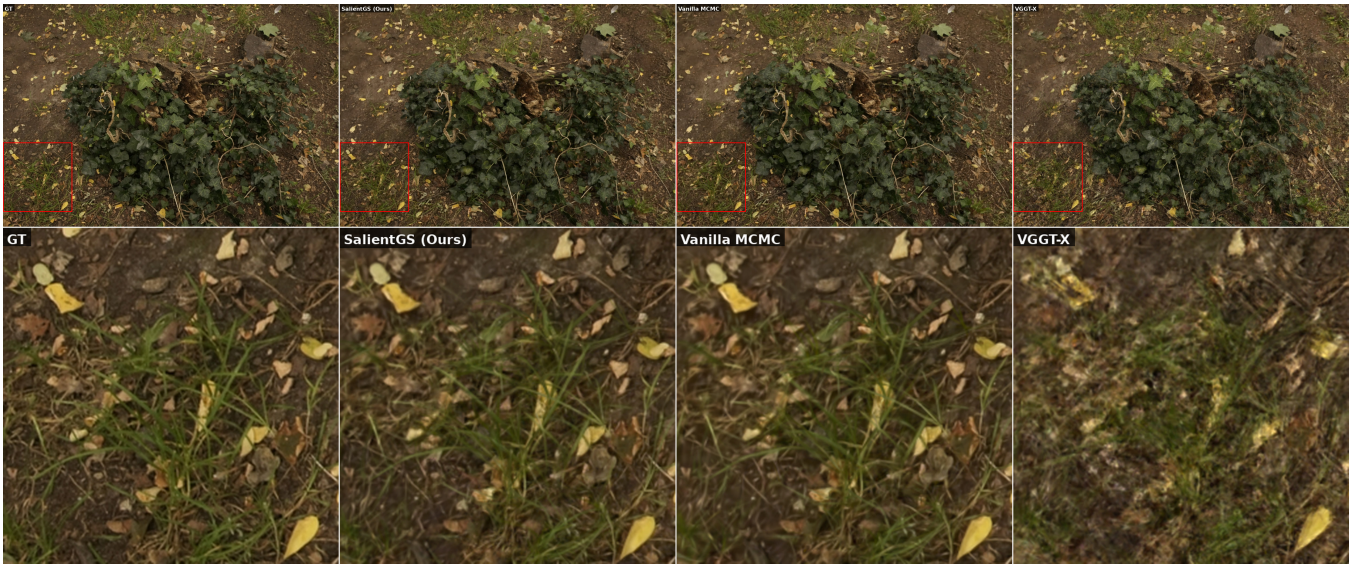

Figure 4: Per-scene qualitative comparison: *stump*.

track distribution at a fraction of the matching cost, closing the structural gap that limits GloSplat-F. Second, the near-identical L=2 scores across both methods confirm that the gain from full tracks is a property of multi-view triangulation geometry, not of any particular front-end or allocation design—making rich track coverage a necessary condition for high-quality reconstruction.

9 Additional Camera-Ready Ablations

This section records the ablations completed during the author response. All occurrences of the complete default model use the corrected released aggregate, 28.82 dB / 0.853 / 0.148; the remaining variant measurements are unchanged.

Table 9 shows that guided allocation improves both initializers: LPIPS changes from 0.151 to 0.133 for GLOMAP and from 0.149 to 0.148 for FastMap. This supports an allocator contribution that is not specific to one camera initializer.

10 Additional Robustness and Failure Diagnostics

The diagnostics in Table 17 identify the texture-poor *room* case through collapsed track length, greater MST reliance,

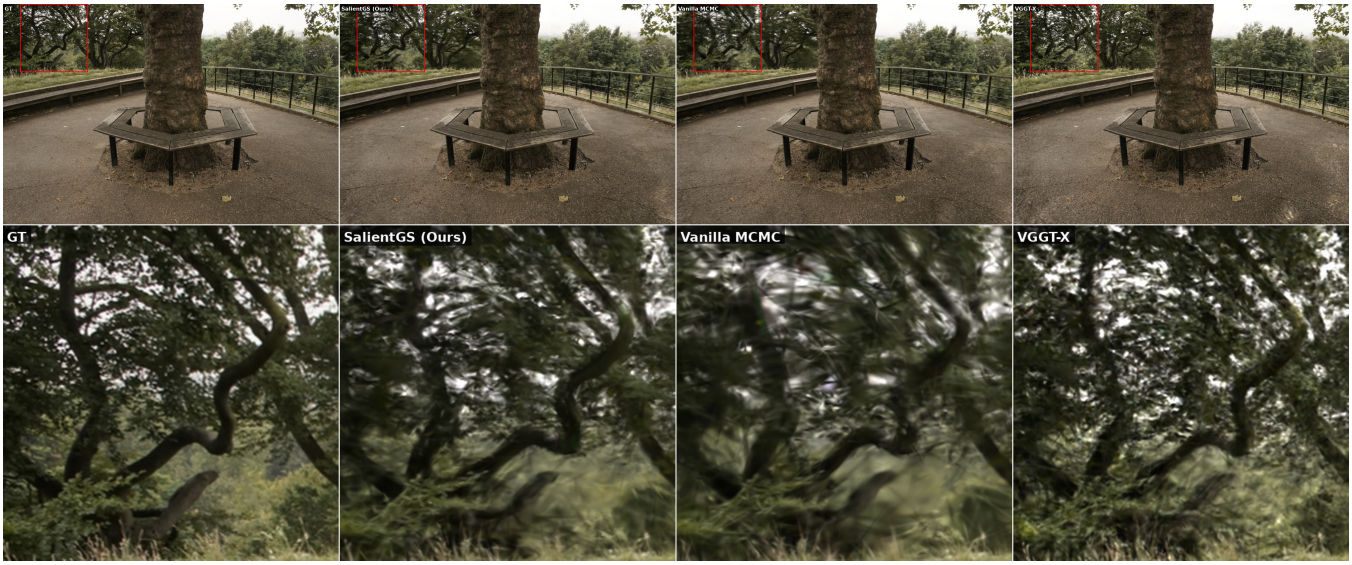

Figure 5: Per-scene qualitative comparison: *treehill*.

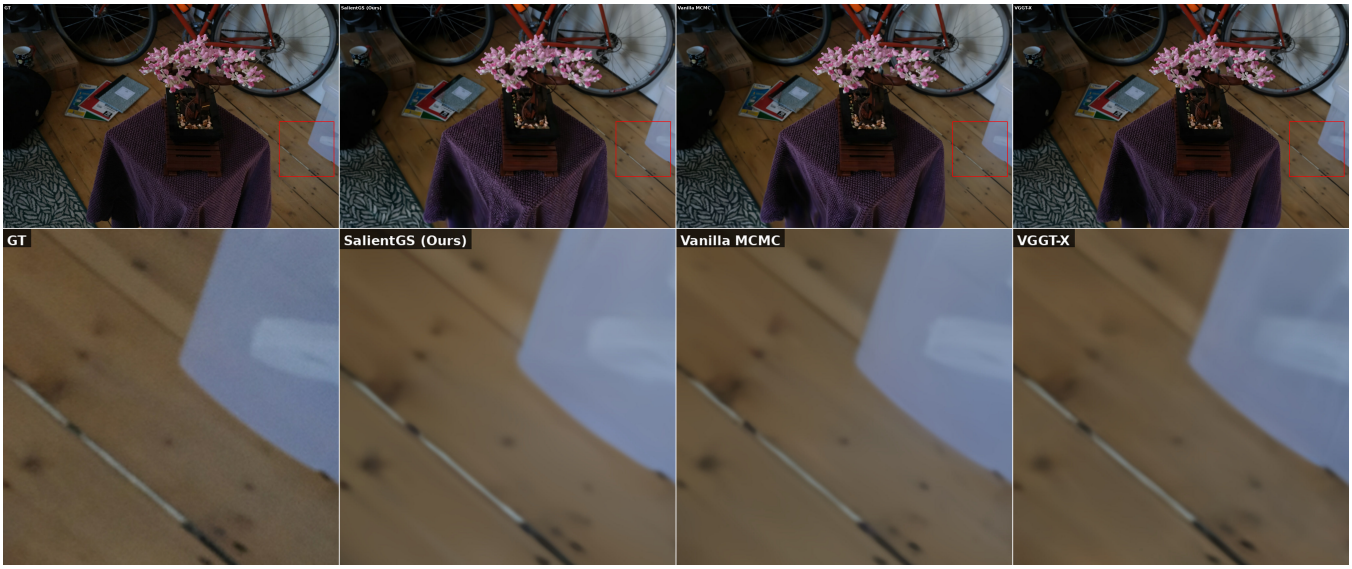

Figure 6: Per-scene qualitative comparison: *bonsai*.

and a high early residual. A feed-forward fallback improves the failure-view LPIPS from 0.312 to 0.235, slightly better than the VGGT-X reference of 0.244. Sparse/few-view reconstruction and severe repetitive or symmetric appearance remain limitations; learned retrieval and adaptive geometry-aware budgets are natural future extensions.

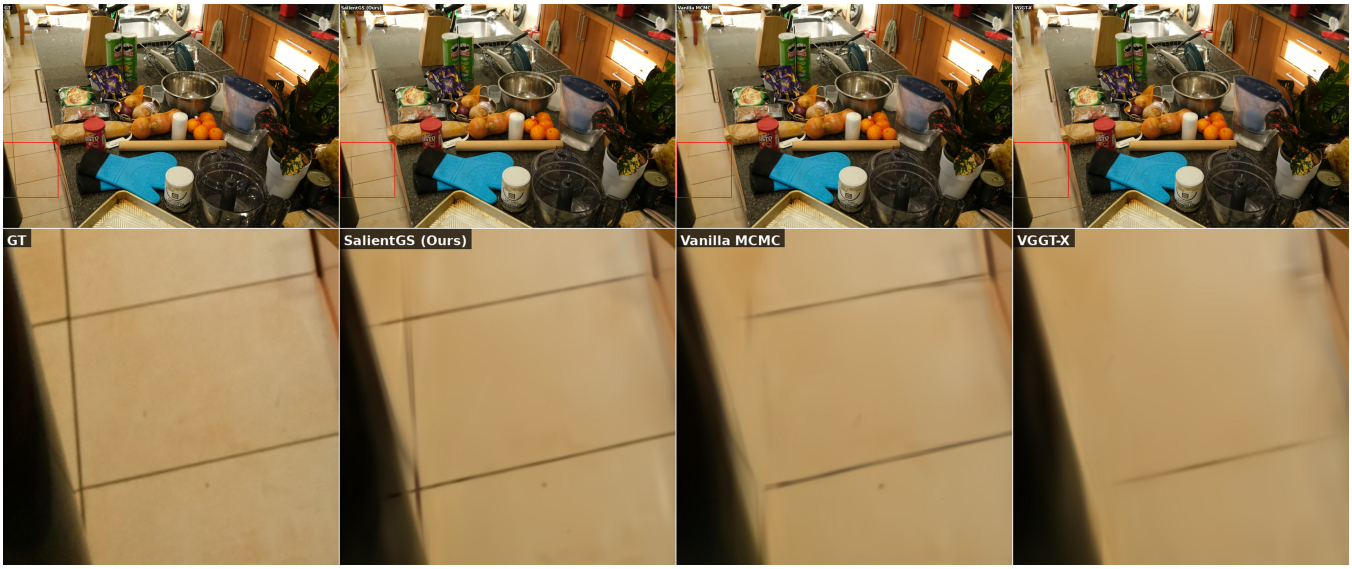

Figure 7: Per-scene qualitative comparison: *counter*.

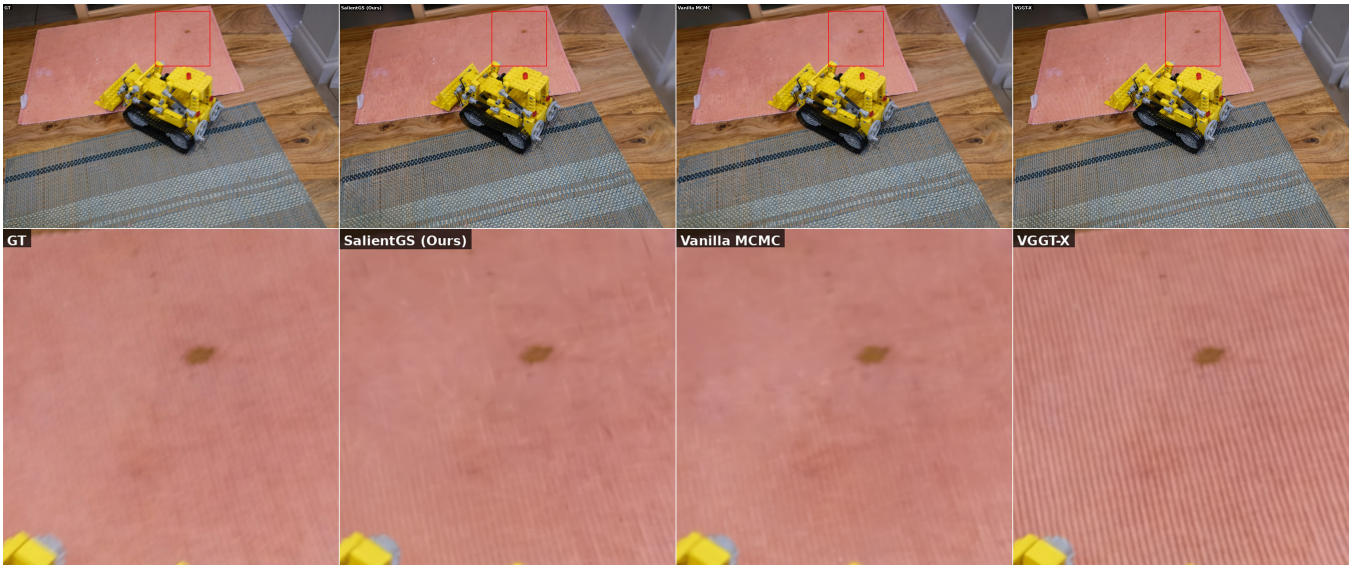

Figure 8: Per-scene qualitative comparison: *kitchen*.

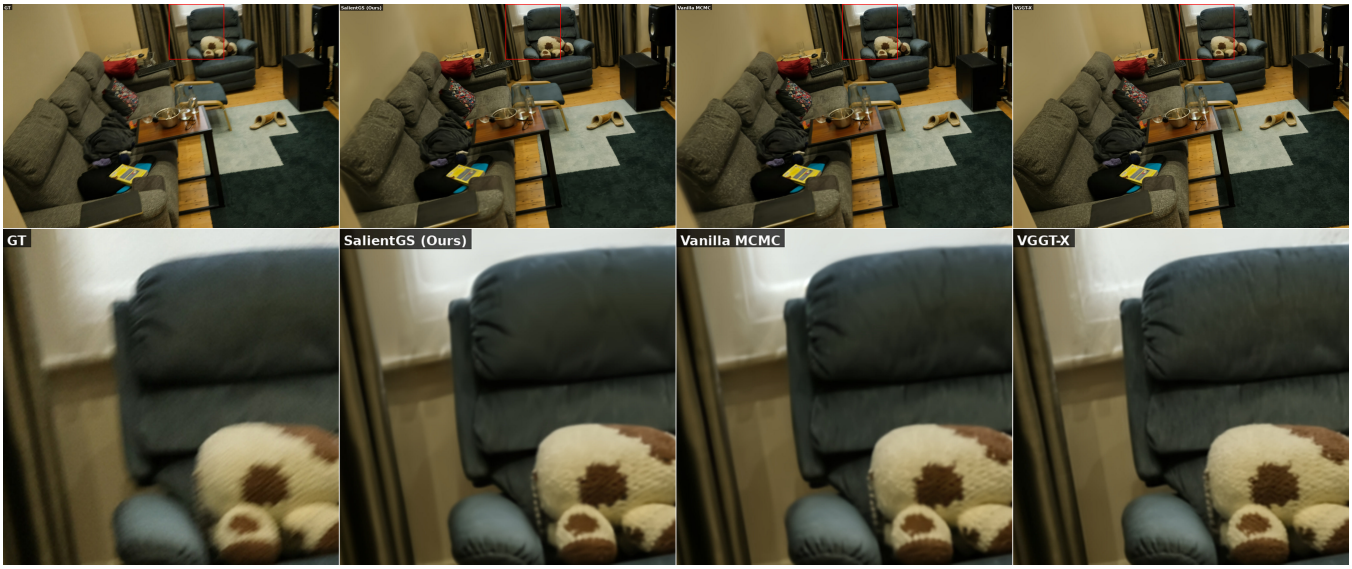

Figure 9: Per-scene qualitative comparison: *room*. SalientGS underperforms on this indoor scene with uniform textures, where first-order SfM provides weaker geometric constraints.

Table 8: Track-length cap ablation on MipNeRF360 (9-scene avg.). The front-end is fixed for each method; only the maximum retained observations per track varies.

Max track length	PSNR↑	SSIM↑	LPIPS↓
GloSplat (SIFT exhaustive front-end)			
GloSplat-F (retrieval top-5, ref.)	*27.77*	*0.818*	*0.164*
$L = 2$	27.46	0.804	0.174
$L =$ all	28.86	0.862	0.139
GloSplat-A (exhaustive, ref.)	*28.86*	*0.862*	*0.139*
SalientGS (Fisher Vector k=20 front-end)			
$L = 2$	27.53	0.808	0.172
$L =$ all (full model)	**28.82**	**0.853**	**0.148**

Table 9: SfM initialization versus Gaussian allocator on Mip-NeRF 360. All rows use the same joint training protocol.

SfM initialization	Allocator	PSNR↑	SSIM↑	LPIPS↓
GLOMAP	ADC	27.52	0.827	0.168
GLOMAP	Vanilla MCMC	28.70	0.846	0.151
GLOMAP	Guided MCMC	**28.87**	**0.855**	**0.133**
FastMap	ADC	27.67	0.828	0.162
FastMap	Vanilla MCMC	28.72	0.848	0.149
FastMap	Guided MCMC (full model)	28.82	0.853	0.148

Table 10: Importance-score schedule and opacity mixing on Mip-NeRF 360.

Setting	PSNR↑	SSIM↑	LPIPS↓
$T = 100$	28.83	0.855	0.135
$T = 200$	**28.87**	**0.856**	**0.133**
$T = 500$ (default/full)	28.82	0.853	0.148
$T = 1000$	28.80	0.852	0.137
score at iter. 0	28.78	0.851	**0.138**
score at iter. 3000 (default/full)	**28.82**	**0.853**	0.148
$\lambda_{\text{mix}} = 0$	28.78	0.851	**0.141**
$\lambda_{\text{mix}} = 0.05$ (default/full)	**28.82**	**0.853**	0.148
$\lambda_{\text{mix}} = 0.20$	28.75	0.850	0.145

Table 11: Robust quantile normalization and selection-quantile interaction on Mip-NeRF 360. The final row reports the fraction of Gaussians in the high-error mass at 30K iterations.

Setting	PSNR↑	SSIM↑	LPIPS↓
Plain min–max	28.71	0.850	0.150
Default quantiles (full)	**28.82**	**0.853**	0.148
$(q_{\text{hi}}, q_{\text{lo}}) = (.80, .20)$	28.79	0.852	0.139
$(q_{\text{hi}}, q_{\text{lo}}) = (.95, .05)$	**28.82**	0.851	**0.137**
High-error mass at 30K	robust: 9.1%; min–max: 42.5%		

Table 12: Pose error before and after joint optimization on the ETH3D SLAM evaluation. Errors are in centimeters; percentages are fractions of cameras below the indicated threshold.

Method	Median↓	Mean↓	< 5cm↑	< 10cm↑
FastMap init	9.8	18.5	31.6	50.5
FastMap + joint	**6.9**	13.5	**38.2**	**57.0**
GLOMAP init	8.5	14.6	35.5	55.5
GLOMAP + joint	7.1	**12.6**	37.8	56.5

Table 13: Per-scene Deep Blending results. The competing methods fail on *drjohnson*, which explains the averaged values in the main table.

Scene	Method	PSNR↑	SSIM↑	LPIPS↓
playroom	GloSplat-A	29.38	0.903	0.224
playroom	SalientGS	**30.15**	**0.920**	**0.156**
drjohnson	GloSplat-A	7.52	0.263	0.792
drjohnson	VGGT-X	7.65	0.337	0.858
drjohnson	SalientGS	**28.82**	**0.892**	**0.210**

Table 14: Budget efficiency on Mip-NeRF 360.

N_{GS}	Vanilla MCMC	Guided MCMC	ΔPSNR
500K	28.19	28.46	+0.27
1.0M	28.59	28.81	+0.22
1.5M	28.72	28.82	+0.10
2.0M	28.80	28.93	+0.13
3.0M	28.88	28.97	+0.09

Table 15: Dense indoor robustness. SG denotes SalientGS and V+M denotes the VGGT-X+MCMC comparison.

Scene	SG PSNR	V+M PSNR	SG LPIPS	V+M LPIPS
bonsai	33.21	30.71	0.106	0.139
counter	29.60	28.80	0.136	0.158
kitchen	32.79	29.68	0.076	0.113
room	31.95	30.76	0.150	0.165
Average	**31.89**	29.99	**0.117**	0.144

Table 16: Additional large-scene and drone-capture evidence.

Setting	Result
Mill-19 Building	Best PSNR, 24.19 dB
Mill-19 Rubble	Best PSNR, 26.98 dB
Four large scenes	Best PSNR/LPIPS on 3 of 4 scenes
Runtime/memory	∼1 h / ∼5 GB vs. ∼3 h / ∼14 GB for CityGS-X

Table 17: Failure diagnostics and fallback on representative Mip-NeRF 360 scenes. Residual is the normalized early epipolar residual.

Scene	Median track	Mean degree	MST reliance	Residual
garden	4	18.4	4%	0.20
kitchen	3	13.4	11%	0.43
room	2	9.7	21%	0.57
room LPIPS: original 0.312; fallback 0.235; VGGT-X 0.244				

Table 18: Sequential Tanks & Temples comparison with CF-3DGS.

Scene	Method	PSNR↑	SSIM↑	LPIPS↓
Family	CF-3DGS	33.12	0.960	0.051
Family	SalientGS	**36.80**	**0.970**	**0.024**
Francis	CF-3DGS	32.78	0.915	0.135
Francis	SalientGS	**36.29**	**0.947**	**0.056**

References

[1] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 5470–5479.

[2] Danpeng Chen, Hai Li, Weicai Ye, Yifan Wang, Weijian Xie, Shangjin Zhai, Nan Wang, Haomin Liu, Hujun Bao, and Guofeng Zhang. 2025. PGSR: Planar-Based Gaussian Splatting for Efficient and High-Fidelity Surface Reconstruction. *IEEE Transactions on Visualization and Computer Graphics* 31, 12 (2025), 6100–6111. doi:10.1109/TVCG.2024.3494046

[3] Hanlin Chen, Fangyin Wei, Chen Li, Tianxin Huang, Yunsong Wang, and Gim Hee Lee. 2024. VCR-GauS: View Consistent Depth-Normal Regularizer for Gaussian Surface Reconstruction. In *Advances in Neural Information Processing Systems*, Vol. 37. Curran Associates, Inc., Red Hook, NY, USA.

[4] Yu Chen and Gim Hee Lee. 2024. DOGS: Distributed-Oriented Gaussian Splatting for Large-Scale 3D Reconstruction Via Gaussian Consensus. In *Advances in Neural Information Processing Systems*, Vol. 37. Curran Associates, Inc., Red Hook, NY, USA.

[5] Yuanyuan Gao, Hao Li, Jiaqi Chen, Zhengyu Zou, Zhihang Zhong, Dingwen Zhang, Xiao Sun, and Junwei Han. 2025. CityGS-X: A Scalable Architecture for Efficient and Geometrically Accurate Large-Scale Scene Reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Piscataway, NJ, USA.

[6] Antoine Guédon and Vincent Lepetit. 2024. SuGaR: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 5354–5363.

[7] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics* 42, 4, Article 139 (July 2023), 14 pages. https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/

[8] Changbai Li, Haodong Zhu, Hanlin Chen, Xiuping Liang, Tongfei Chen, Shuwei Shao, Linlin Yang, Huobin Tan, and Baochang Zhang. 2026. UrbanGS: Efficient and Scalable Architecture for Geometrically Accurate Large-Scene Reconstruction. In *International Conference on Learning Representations*. https://openreview.net/forum?id=L3utaw6SD9

[9] Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H. Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. 2023. Neuralangelo: High-Fidelity Neural Surface Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 8456–8465.

[10] Jiaqi Lin, Zhihao Li, Xiao Tang, Jianzhuang Liu, Shiyong Liu, Jiayue Liu, Yangdi Lu, Xiaofei Wu, Songcen Xu, Youliang Yan, and Wenming Yang. 2024. VastGaussian: Vast 3D Gaussians for Large Scene Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA.

[11] Yang Liu, He Guan, Chuanchen Luo, Lue Fan, Naiyan Wang, Junran Peng, and Zhaoxiang Zhang. 2024. CityGaussian: Real-time High-quality Large-Scale Scene Rendering with Gaussians. In *European Conference on Computer Vision*. Springer, Cham.

[12] Yang Liu, Chuanchen Luo, Zhongkai Mao, Junran Peng, and Zhaoxiang Zhang. 2025. CityGaussianV2: Efficient and Geometrically Accurate Reconstruction for Large-Scale Scenes. In *International Conference on Learning Representations*. https://openreview.net/forum?id=a3ptUbuzbW

[13] Zhenxing Mi and Dan Xu. 2023. Switch-NeRF: Learning Scene Decomposition with Mixture of Experts for Large-scale Neural Radiance Fields. In *International Conference on Learning Representations*. https://openreview.net/forum?id=PQ2zoIZqvm

[14] Haithem Turki, Deva Ramanan, and Mahadev Satyanarayanan. 2022. Mega-NeRF: Scalable Construction of Large-Scale NeRFs for Virtual Fly-Throughs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 12922–12931.

[15] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. 2021. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. In *Advances in Neural Information Processing Systems*, Vol. 34. Curran Associates, Inc., Red Hook, NY, USA.