

GDoFS: Gaussian DoF Separation for Plausible 3D Geometry in Sparse-View 3DGS

Supplementary Material

A. Note on Supplementary Videos

The suppression of rendering artifacts resulting from our geometric improvements is most clearly demonstrated in video. **We highly recommend viewing the videos in the supplement material** to fully appreciate the reduction in rendering artifacts discussed in the main paper.

B. Implementation Details

B.1. Initialization Detail

We use MAST3R [3] and VGGT [4] to obtain the initial depth map and camera parameters. MAST3R prunes points with low confidence and returns a point cloud. However, this pruning can create empty spaces in the scene. To address this, we initialize our model using all pixels of the depth map, regardless of confidence.

C. Additional Experiments

C.1. Per-Scene Quantitative Results

We present the performance analysis for each scene, as summarized in Tabs. S1 to S3. Our method demonstrates results comparable to those of the baselines for novel-view synthesis performance, varying across scenes. However, the patch-wise depth Pearson correlation (PDC) performance consistently exhibits superior results for all scenes. This indicates that our method achieves robust geometry improvements without being affected by scene characteristics.

D. Additional Analysis

D.1. Opacity Distribution Changes

Sparsity of Gaussians. Fig. S1 shows the opacity changes over the training steps. Starting from an initial opacity of 0.1, Gaussians quickly become distinguishable as opaque or transparent within the first 1000 steps. To an-

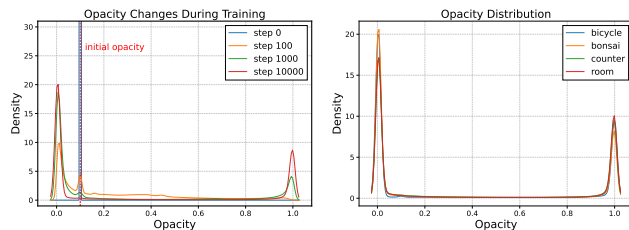


Figure S1. **Opacity distribution.** (Left) Opacity distribution changes during training and (Right) distribution after training across different scenes.

alyze opacities, we estimate the opacity distribution using kernel density estimation with a covariance factor of 0.025 on the Bonsai scene from Mip-NeRF 360 [1]. Thanks to this sparsity, the simple pruning method described in Sec. 7.3 results in a 46.4% increase in training speed.

Semi-transparent objects. Visibility loss treats occluded Gaussians as a failure case in reprojection-based DoF separation. However, semi-transparent objects may occlude other Gaussians while still producing plausible geometry. To analyze this, we estimate the opacity distribution on the Mip-NeRF 360 dataset [1]. As shown in Fig. S1, most Gaussians' opacities are binarized to either zero or one, validating our method in the majority of scenarios.

D.2. Effect of Visibility Loss Scheduling

We linearly decrease the coefficient of visibility loss from 0.1 to 0 during training. Tab. S4 presents the performance differences with and without linear scheduling on the Tanks and Temples dataset. The application of scheduling does not result in significant changes in patch-wise depth Pearson correlation (PDC) performance. However, a notable decrease in PSNR is observed. This suggests that controlling degrees of freedom (DoFs) through visibility loss is helpful in the early training stages, but relaxing constraints after establishing sufficient geometry improves scene representation.

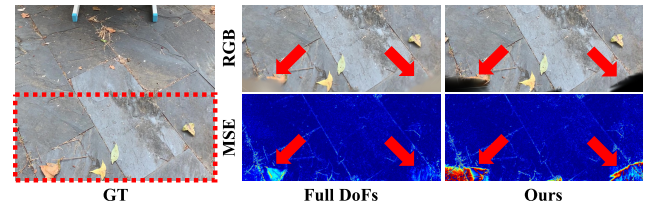


Figure S2. **Comparison of rendered RGB and MSE for unseen regions.** Most of the novel view synthesis errors occur in unseen regions that were not observed during training.

D.3. Challenges in Rendering Unseen Regions.

We enforce Gaussians to remain near the corresponding rays by imposing constraints on image-plane-parallel depths of field (DoFs), and extensive experiments demonstrate that this constraint contributes to improved geometric quality. However, this pipeline may degrade novel view synthesis evaluation performance. For instance, as shown in Fig. S2, since our method does not generate Gaussians in unseen regions of the training views, these regions remain empty when rendering test views. In contrast, base-

	MASt3R [3] + Full DoFs				VGGT [4] + Full DoFs				MASt3R [3] + GDoFS				VGGT [4] + GDoFS			
	PSNR	SSIM	LPIPS	PDC	PSNR	SSIM	LPIPS	PDC	PSNR	SSIM	LPIPS	PDC	PSNR	SSIM	LPIPS	PDC
Bicycle	13.70	0.1676	0.5615	0.0971	14.57	0.2381	0.5211	0.2044	14.14	0.1714	0.5481	0.2119	15.16	0.2320	0.4690	0.3269
Bonsai	26.59	0.8727	0.1341	0.3038	26.61	0.8730	0.1347	0.4233	26.41	0.8815	0.1178	0.5283	26.06	0.8766	0.1229	0.5299
Counter	24.06	0.7653	0.2085	0.3502	24.41	0.7794	0.2111	0.3900	24.01	0.7392	0.2067	0.5084	24.47	0.7658	0.2020	0.5029
Flowers	12.54	0.1265	0.6082	0.0265	12.78	0.1561	0.6820	0.0277	13.22	0.1389	0.6120	0.1464	12.72	0.1475	0.6310	0.1116
Garden	16.32	0.2781	0.4432	0.2310	18.15	0.4467	0.3758	0.3453	17.25	0.2709	0.4420	0.4506	18.54	0.4595	0.3325	0.4516
Kitchen	21.97	0.7153	0.2031	0.3451	17.40	0.5635	0.3218	0.2967	21.93	0.6918	0.1832	0.5721	19.25	0.6804	0.2098	0.4174
Room	25.18	0.8335	0.1864	0.5126	27.49	0.8675	0.1530	0.6010	26.35	0.8409	0.1506	0.6600	27.28	0.8508	0.1481	0.6770
Stump	16.90	0.2617	0.5430	0.0919	16.89	0.2936	0.5523	0.0894	17.44	0.2725	0.5299	0.1870	17.44	0.2824	0.5277	0.1792
Treehill	12.83	0.2485	0.5551	0.0953	13.05	0.2698	0.5611	0.1098	13.89	0.2822	0.5357	0.2886	13.85	0.2905	0.5260	0.2542
Avg.	18.90	0.4744	0.3826	0.2282	19.04	0.4986	0.3903	0.2764	19.41	0.4766	0.3696	0.3948	19.42	0.5095	0.3521	0.3834

Table S1. **Per-scene quantitative comparison on Mip-NeRF 360 [1].** All scenes are trained with 12 views. PDC denotes patch-wise depth Pearson correlation. The best results are highlighted in bold.

	MASt3R [3] + Full DoFs				VGGT [4] + Full DoFs				MASt3R [3] + GDoFS				VGGT [4] + GDoFS			
	PSNR	SSIM	LPIPS	PDC	PSNR	SSIM	LPIPS	PDC	PSNR	SSIM	LPIPS	PDC	PSNR	SSIM	LPIPS	PDC
Bench	23.47	0.7626	0.2140	0.3436	21.39	0.6376	0.2946	0.2439	23.99	0.7880	0.1886	0.5213	22.89	0.7395	0.2139	0.3681
Bicycle	11.20	0.1421	0.5655	0.1694	20.43	0.6277	0.3038	0.2561	11.27	0.1437	0.5728	0.2257	20.79	0.6759	0.2392	0.3215
Car	29.50	0.9252	0.1078	0.4074	25.40	0.8544	0.1632	0.2151	29.20	0.9216	0.1034	0.5066	26.38	0.8729	0.1374	0.3427
SUV	27.92	0.8685	0.1410	0.4381	27.64	0.8820	0.1330	0.4775	29.21	0.9016	0.1118	0.6592	27.62	0.8864	0.1174	0.5698
Chair	26.55	0.8333	0.2201	0.2641					27.26	0.8546	0.1836	0.3240				
Ladder	18.45	0.5933	0.3056	0.2062					18.63	0.5946	0.2979	0.3409				
Table	25.04	0.8471	0.1748	0.2379					24.91	0.8447	0.1653	0.3038				
Avg.	23.16	0.7103	0.2470	0.2952	23.72	0.7504	0.2236	0.2981	23.49	0.7213	0.2319	0.4116	24.42	0.7937	0.1770	0.4005

Table S2. **Per-scene quantitative comparison on MVImgNet [5].** All scenes are trained with 12 views. PDC denotes patch-wise depth Pearson correlation. The best results are highlighted in bold.

	MASt3R [3] + Full DoFs				VGGT [4] + Full DoFs				MASt3R [3] + GDoFS				VGGT [4] + GDoFS			
	PSNR	SSIM	LPIPS	PDC	PSNR	SSIM	LPIPS	PDC	PSNR	SSIM	LPIPS	PDC	PSNR	SSIM	LPIPS	PDC
Ballroom	24.12	0.8295	0.1078	0.6062	23.63	0.8330	0.1135	0.6049	24.20	0.8341	0.1117	0.7790	24.17	0.8399	0.1075	0.7123
Barn	20.11	0.6884	0.2160	0.3514	20.48	0.6932	0.2227	0.3739	22.03	0.6962	0.2245	0.5416	22.02	0.6885	0.2336	0.4507
Church	16.93	0.6510	0.2422	0.5646	19.18	0.6691	0.2487	0.5458	19.17	0.6598	0.2421	0.7029	19.30	0.6652	0.2450	0.6513
Family	23.29	0.8171	0.1277	0.3157	23.43	0.8087	0.1467	0.3485	24.16	0.8179	0.1349	0.6452	24.30	0.8199	0.1316	0.5292
Francis	19.39	0.7060	0.2725	0.2155	19.45	0.6241	0.3664	0.1527	24.03	0.7625	0.2417	0.3716	22.48	0.6840	0.3109	0.2614
Horse	22.94	0.8158	0.1336	0.2977	22.01	0.7904	0.1829	0.2796	23.31	0.8127	0.1499	0.4281	23.62	0.8249	0.1403	0.3877
Ignatius	22.81	0.6962	0.1791	0.3508	22.53	0.6749	0.2032	0.4040	23.39	0.7003	0.1768	0.7186	22.79	0.6936	0.1861	0.5282
Museum	22.88	0.7912	0.1260	0.5720	21.93	0.7828	0.1390	0.5028	22.72	0.7789	0.1514	0.6078	22.96	0.7869	0.1495	0.5305
Avg.	21.56	0.7494	0.1756	0.4092	21.58	0.7345	0.2029	0.4015	22.88	0.7578	0.1791	0.5993	22.71	0.7504	0.1880	0.5064

Table S3. **Per-scene quantitative comparison on Tanks and Temples [2].** All scenes are trained with 3 views. PDC denotes patch-wise depth Pearson correlation. The best results are highlighted in bold.

	w/o Depth Scheduling		Ours	
	PSNR	PDC	PSNR	PDC
Ballroom	30.18 (−0.38)	0.8122 (+0.0008)	30.56	0.8114
Barn	27.86 (−0.47)	0.5919 (−0.0012)	28.33	0.5931
Church	23.60 (+0.05)	0.7689 (+0.0092)	23.55	0.7597
Family	28.87 (−0.62)	0.6891 (+0.0021)	29.49	0.6870
Francis	30.35 (−1.08)	0.3992 (−0.0030)	31.43	0.4022
Horse	28.14 (−0.35)	0.4621 (+0.0019)	28.49	0.4602
Ignatius	26.58 (−1.23)	0.7484 (−0.0040)	27.81	0.7525
Museum	28.61 (−0.02)	0.6762 (+0.0033)	28.63	0.6729
Avg.	28.02 (−0.51)	0.6435 (+0.0011)	28.54	0.6424

Table S4. **Effect of visibility loss scheduling.** The values in parentheses indicate the variation compared to when visibility loss scheduling is used. Without visibility loss scheduling, PSNR decreases by up to 1.23 depending on the scene, while PDC shows negligible changes.

line methods allow them to populate such regions. As a result, the majority of PSNR degradation occurs in areas that were not observed during training. This suggests that our method’s novel view synthesis performance is underestimated.

References

- [1] 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. 1, 2
- [2] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM TOG*, 36(4):1–13, 2017. 2
- [3] Vincent Leroy, Yohann Cabon, and Jérôme Revaud. Grounding image matching in 3d with mast3r. In *ECCV*, pages 71–91, 2024. 1, 2
- [4] Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. Vggt: Visual geometry grounded transformer. In *CVPR*, pages 5294–5306, 2025. 1, 2
- [5] Xianggang Yu, Mutian Xu, Yidan Zhang, Haolin Liu, Chongjie Ye, Yushuang Wu, Zizheng Yan, Chenming Zhu, Zhangyang Xiong, Tianyou Liang, et al. Mvimngnet: A large-scale dataset of multi-view images. In *CVPR*, pages 9150–9161, 2023. 2