HoGS: Unified Near and Far Object Reconstruction via Homogeneous Gaussian Splatting

Supplementary Material

This supplementary material first provides a detailed comparison of our homogeneous representation with existing inverted spherical representation [9] (Sec. A). We also analyze the methods' convergence behavior (Sec. B) and potential design choices in methods and experiments (Sec. C). An experiment in Sec. D assesses whether our method can represent an *infinitely* far object (*i.e.*, the Moon) in our dataset. Finally, Sec. E shows additional and detailed results in each scene we tested.

The supplementary video visualizes the results in a more intuitive manner, and we strongly encourage the reader to refer to it.

A. Inverted Spherical vs. Homogeneous Representations

Similar to the homogeneous representation, the inverted spherical representation in NeRF++ [9] is also designed to represent distant objects effectively. We compare the performance of inverted spherical and homogeneous representation in 3D scene representation.

Inverted Spherical Representation [9]. We define a point $\boldsymbol{p} = [x, y, z]^\top \in \mathbb{R}^3$ to be represented with the inverted spherical representation $\boldsymbol{p'} = [\theta, \phi, w']^\top$ as:

$$\begin{cases} \theta = \arctan\left(\frac{y}{x}\right), \\ \phi = \arcsin\left(\frac{z}{\|r\|}\right), \\ w' = \frac{1}{r}, \end{cases} \tag{S1}$$

where w' is the inverted depth and $r = \sqrt{x^2 + y^2 + z^2} > 1$. The inverted spherical representation can be converted to Cartesian by:

$$\begin{cases} x = \frac{\sin\phi\cos\theta}{w'}, \\ y = \frac{\sin\phi\sin\theta}{w'}, \\ z = \frac{\cos\phi}{w'}. \end{cases} \tag{S2}$$

Comparison Results. The quantitative results presented in Table S1 and Table S2 show that homogeneous representation consistently outperforms inverted spherical representation in indoor and near-object scenarios. Although inverted depth w' effectively represents distant points in its range $(0, 1]$, its mapping of points with depth within 1 to its range $[1, +\infty)$ hinders the performance on nearby objects. In contrast, homogeneous coordinates offer a balanced representation of near and far objects, using the weight w to account for depths.

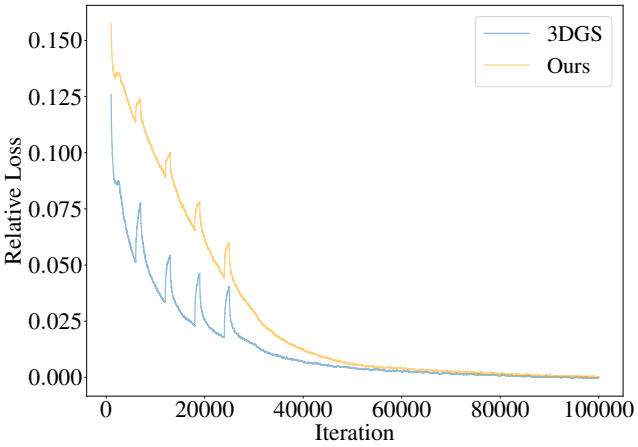
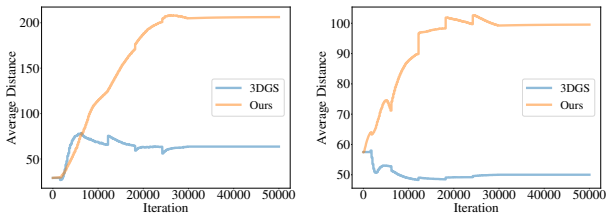


Figure S1. **Loss convergence of 3DGS and HoGS.** The loss curves indicate that with our setup on unbounded scenes, both 3DGS and HoGS have not fully converged by $30,000$ iterations, while with $50,000$ iterations they become nearly minimal.



(a) TRAIN (b) DL3 21

Figure S2. **Average distance of 10% farthest points.** In both the TRAIN and DL3DV 21 scenes, the 10% farthest points in our method reach significantly farther distances compared to 3DGS. By the end of the training, the farthest points in our method reach approximately $1,200$ meters in the TRAIN dataset (compared to 410 meters in 3DGS) and 520 meters in DL3 Scene 21 (compared to 97 meters in 3DGS) in the physical space.

B. More Convergence Analysis

While the main paper presents a convergence analysis in a simple setup, we further analyze the convergence behaviors.

Extended Training. The main paper uses $50,000$ iterations for both the original 3DGS and our HoGS. To validate the number of iterations for unbounded scenes, we present the training process of $100,000$ iterations in an unbounded scene (the TRAIN scene in the Tank&Temples dataset [3]) in Fig. S1. The loss curves in Fig. S1 show the relative loss, where the loss values are normalized by setting the loss at $50,000$ iterations as the reference point. Both 3DGS and

Dataset	Mip-NeRF 360 Dataset			Tanks&Temples			DL3DV-10K Benchmark		
Method | Metric	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
Inverted Spherical	**0.828**	**27.93**	0.196	0.855	24.15	0.169	0.918	29.85	0.117
HoGS (Ours)	**0.828**	27.92	**0.194**	**0.858**	**24.27**	**0.166**	**0.919**	**29.93**	**0.114**

Table S1. **Inverted spherical vs. homogeneous representations.** This table reports the performance of Inverted Spherical and Homogeneous (HoGS) methods on three datasets: Mip-NeRF 360 [1], Tanks&Temples [3], and DL3DV-10K Benchmark [4]. Results are evaluated using three metrics: SSIM, PSNR, and LPIPS. HoGS consistently shows competitive or better performance across all datasets.

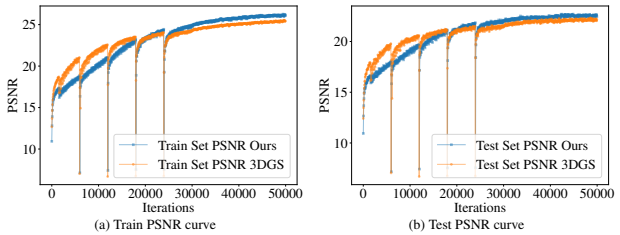
Method	Tanks&Temples (**Near**)			Tanks&Temples (**Far**)		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
Inverted Spherical	0.882	25.54	0.137	**0.976**	**30.78**	0.032
HoGS (Ours)	**0.883**	**25.69**	**0.135**	**0.976**	30.77	**0.031**
Method	DL3DV-10K (**Near**)			DL3DV-10K (**Far**)		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
Inverted Spherical	0.926	30.46	0.101	**0.993**	40.43	**0.013**
HoGS (Ours)	**0.927**	**30.53**	**0.098**	**0.993**	**40.55**	**0.013**
Method	Mip-NeRF360 (**Indoor**)			Mip-NeRF360 (**Outdoor**)		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
Inverted Spherical	0.917	30.40	**0.157**	0.746	25.13	0.223
HoGS (Ours)	**0.930**	**31.38**	0.163	**0.747**	**25.14**	**0.219**

Table S2. **Performance of inverted spherical and homogeneous methods across different scenarios.** This table highlights the performance of Homogeneous and Spherical representations on Tanks&Temples, DL3DV-10K, and Mip-NeRF360 datasets. Homogeneous representation consistently outperforms the inverted Spherical representation in indoor and near-object scenarios.

our method reduce the training losses until around $50,000$ iterations, which are almost converging there, demonstrating that $50,000$ iterations are reasonable for unbounded scene reconstruction in the main paper's experiments.

Convergence Behavior in Real Scenes. Here, we analyze the convergence behaviors in real scenes by tracking the average displacement of the Gaussian centers farthest 10% from the world origin over iterations. Figure S2 show that far points in our method move more quickly to distant positions than the original 3DGS in unbounded scenes (TRAIN [3] and DL3DV Scene 21 [4]), underscoring the efficiency of our approach in reconstructing distant objects.

PSNR Convergence Analysis. We also show a PSNR-iteration curve comparing HoGS and 3DGS as shown in Fig. S3. HoGS shows a slower initial convergence due to the homogeneous optimization of near and far objects, which requires Gaussians to propagate across the entire scene. However, as training progresses, HoGS surpasses 3DGS in PSNR, demonstrating superior reconstruction fidelity and global consistency.



(a) Train PSNR curve (b) Test PSNR curve

Figure S3. PSNR-iteration curve of our HoGS and 3DGS. HoGS exhibits a slower initial convergence due to the homogeneous optimization of near and far objects, requiring Gaussians to propagate across the scene. However, as training progresses, HoGS surpasses 3DGS in PSNR, demonstrating its superior reconstruction fidelity and global consistency.

C. More Design Choices

C.1. Skybox Initialization

Following the methods in [2, 7], we can essentially use skybox [2] initialization to enhance the points in distant scenes and the 3DGS and our proposed methods. We thus test the skybox initialization to both the 3DGS and our proposed methods. Specifically, we added $100,000$ blue points as initial points on the upper hemisphere, with a radius of $1,000$ unit distance, which corresponds to approximately 1.1 kilometers in DL3DV Scene 21 (estimated using the base width of the George Washington statue in front of the Indiana Statehouse) and 2.1 kilometers in the TRAIN dataset (estimated from the width of railway tracks). As shown in Table S3, the results indicate that incorporating a skybox improves the accuracy of distant scenes in the 3DGS method. This enhancement arises because the skybox provides additional guidance for representing distant elements like clouds, otherwise challenging for 3DGS to capture.

In contrast, adding a skybox to our method does not improve the accuracy of distant scenes. This is because our approach, even without skybox initialization, effectively leverages the attributes of homogeneous coordinates to represent both near and distant objects equally. Our method accurately represents distant areas without additional priors by inherently keeping consistent scaling across varying depths.

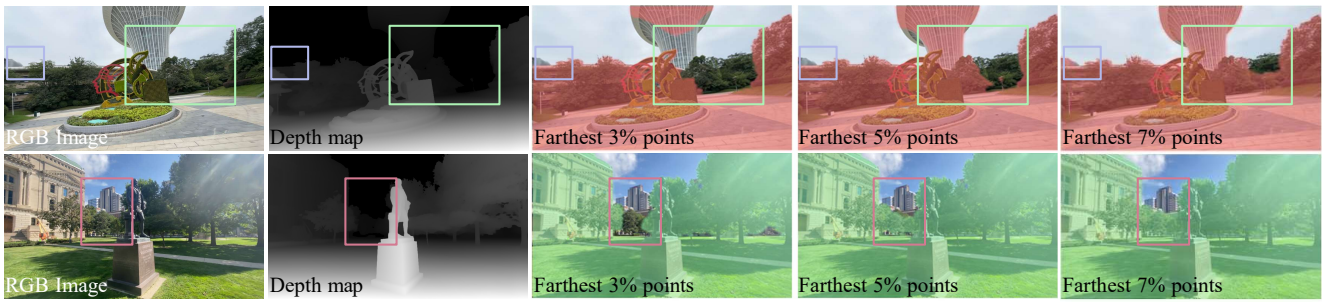


Figure S4. **Impact of threshold selection on near and far object classification.** We evaluate the effects of changing the threshold used to classify far and near objects in the depth mask used for the experiment. When the threshold is reduced to 3%, near objects such as parts of buildings and shrubs (green bounding box) and park trees (red bounding box) are mistakenly classified as far. Conversely, increasing the threshold to 7% results in some far objects, like distant buildings, being misclassified as near, as shown by the blue bounding box. To balance accuracy and meaningful far-object coverage, we select the 5 % threshold as the optimal trade-off.

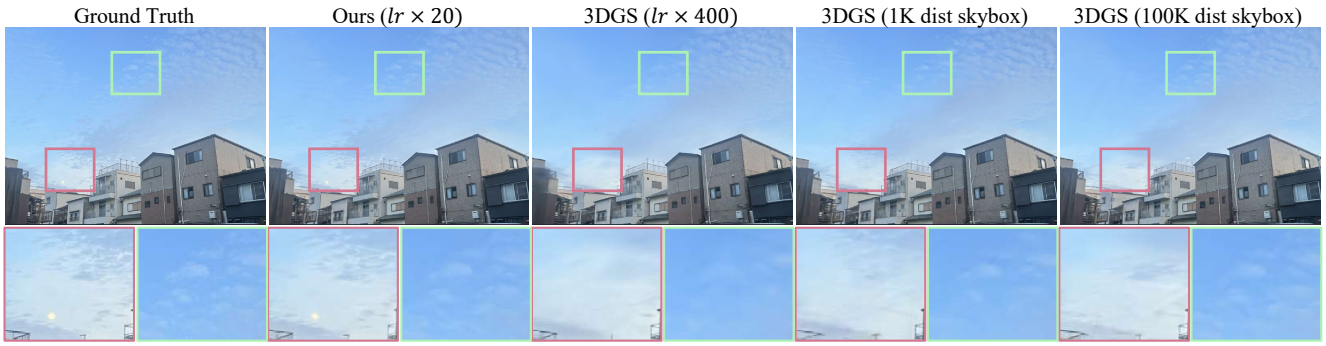


Figure S5. **Reconstruction of an infinitely far (*i.e.*, the Moon) scene.** With a 20x increase in the learning rate (lr), our method successfully represents the Moon, even at its (near-)infinity distances, as highlighted by the red bounding box. In contrast, 3DGS fails to achieve similar results, even when the lr is largely increased, or when incorporating a skybox with different $(1,000$ and $100,000)$ unit distances. Notably, while adding a skybox to 3DGS improves the fidelity of cloud details compared to simply increasing lr, it remains incapable of generating a realistic appearance of the Moon.

Method	TRAIN (**Near**)			TRAIN (**Far**)		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
3DGS w/o skybox	**0.854**	**23.86**	0.151	0.960	28.49	0.063
3DGS w/ skybox	0.851	23.67	0.153	0.962	28.58	0.063
Ours w/o skybox	**0.854**	23.70	**0.147**	**0.976**	**30.42**	**0.044**
Ours w/ skybox	0.849	23.33	0.149	0.975	30.23	0.046
Method	DL3DV Scene 21 (**Near**)			DL3DV Scene 21 (**Far**)		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
3DGS w/o skybox	0.870	27.72	0.147	0.972	31.23	0.042
3DGS w/ skybox	0.867	27.78	0.152	0.976	32.16	0.038
Ours w/o skybox	0.889	28.81	0.122	**0.986**	**35.48**	**0.022**
Ours w/ skybox	**0.890**	**28.83**	**0.120**	**0.986**	35.15	**0.022**

Table S3. **Impact of skybox on near and far scenarios.** This table compares the performance of 3DGS and our method (w/ and w/o skybox) on TRAIN and DL3DV Scene 21 datasets across Near and Far scenarios. Incorporating a skybox initialization can slightly improve the accuracy of distant scenes in the original 3DGS method but not in our method. Our method effectively represents both near and far objects even without skybox initialization.

Scene	Threshold	**Near**			**Far**		
		SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
TRAIN	7 %	0.858	23.82	0.143	0.972	29.94	0.048
	5 %	0.854	23.70	0.147	0.976	30.43	0.044
	3 %	0.848	23.53	0.151	0.982	31.25	0.039
DL3DV Scene 21	7 %	0.893	29.08	0.118	0.982	34.37	0.025
	5 %	0.889	28.81	0.122	0.986	35.48	0.022
	3 %	0.885	28.60	0.125	0.989	36.80	0.019
DL3DV Scene 97	7 %	0.965	34.63	0.047	0.989	38.69	0.043
	5 %	0.962	34.12	0.051	0.992	40.41	0.039
	3 %	0.959	33.72	0.054	0.995	42.48	0.036

Table S4. **Impact of depth threshold selection on near and far object accuracy.** This table evaluates the performance of near and far objects across different depth thresholds (7 %, 5 %, and 3 %) on the TRAIN, DL3DV Scene 21, and DL3DV Scene 97 datasets. As the threshold decreases, the accuracy for far objects consistently improves (*e.g.*, PSNR increases from 29.94 to 31.25 on TRAIN and from 34.37 to 36.80 on DL3DV Scene 21), driven by a greater share of sky pixels characterized by uniform colors and high accuracy. However, this improvement in the metrics does not take what we really want to evaluate, *e.g.*, buildings and mountains, into account.

C.2. Threshold for Near vs. Far in Experiment

In the main paper's experiment, to evaluate the accuracy for near and far objects, we use Depth Anything V2 [6] to gen-

Method	BICYCLE			FLOWERS			GARDEN			STUMP			TREEHILL		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
3DGS	0.770	25.59	0.219	0.609	21.36	0.347	0.869	27.68	0.110	0.775	26.66	0.223	0.649	22.63	0.332
Scaffold-GS	0.768	25.57	0.228	0.610	21.67	0.337	0.869	**27.92**	0.111	0.777	26.88	0.228	**0.664**	**23.53**	0.312
HoGS (Ours)	**0.790**	**25.84**	**0.183**	**0.638**	**22.18**	**0.311**	**0.874**	27.83	**0.098**	**0.785**	**26.92**	**0.195**	0.649	22.95	**0.310**

Method	BONSAI			COUNTER			KITCHEN			ROOM		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
3DGS	0.946	32.28	0.178	0.914	29.33	0.182	0.931	31.41	0.115	0.925	31.93	0.198
Scaffold-GS	**0.950**	**33.08**	0.169	**0.919**	**29.81**	**0.177**	**0.934**	31.94	**0.113**	**0.932**	**32.34**	**0.181**
HoGS (Ours)	0.948	32.56	**0.164**	0.913	29.25	0.179	0.932	**32.04**	0.114	0.925	31.67	0.196

Table S5. **Comparisons for Mip-NeRF360 outdoor and indoor scenes.** 3DGS-based methods' SSIM, PSNR and LPIPS scores for Mip-NeRF360 scenes. Outdoor scenes are listed above, while indoor scenes are listed below.

Dataset / Method / Metric	Mip-NeRF 360 Dataset			Tanks&Temples			DL3DV-10K Benchmark		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
Mip-Splatting [8]	0.835	27.97	0.182	0.856	24.01	0.160	0.917	29.44	0.116
Multi-Scale 3DGS [5]	0.821	27.73	0.211	0.846	23.92	0.180	0.893	28.23	0.148
HoGS (Ours)	0.828	27.92	0.194	0.858	24.27	0.166	0.919	29.93	0.114

Table S6. **Comparison with Mip-Splatting and Multi-Scale 3D Gaussian Splatting.**

erate depth maps from input images at their original resolution. We define distant areas as these maps' farthest 5 % of depth values. This threshold is effective because adopting the 0 % threshold would misclassify distant mountains and buildings as near objects, evaluating far objects' accuracy only on limited sky regions. Here, we additionally tested thresholds of 3 % and 7 % as shown in Fig. S4 and Table S4. As the threshold changes from 7 % to 5 % to 3 %, the accuracy for far objects improves. This is because the number of pixels classified as distant objects decreases while the number of pixels representing the sky remains constant. Typically characterized by uniform colors and high accuracy, the sky becomes a more significant proportion of the far-object pixels as the threshold is reduced. Consequently, the overall accuracy improves with an increasing share of sky pixels in the far-object category.

However, it is essential to note that while reducing the threshold improves the accuracy metrics, it does not measure the representation of other *distant* objects, such as mountains or buildings, which are critical for photorealistic scene reconstruction (see Fig. S4). Hence, we set a 5 % threshold in the main paper's experiment, which provides a balanced trade-off between including enough distant objects for meaningful evaluation and avoiding the misclassification of nearby elements.

D. Representing Infinitely Far Objects

We collected a custom dataset containing scenes with the Moon to assess the reconstruction capability of objects at *infinitely* far away[1]. Figure S5 shows a visual comparison. With our method, increasing the learning rate (lr) for the weight parameter w by 20x enabled the accurate reconstruction of the Moon. In contrast, 3DGS failed to reconstruct the Moon even with a large lr, with an increase of 400x, producing artifacts that disappear when the viewpoint changes. Adding a skybox with a $1,000$ unit or $100,000$ unit distance to 3DGS also failed to produce a realistic appearance of the Moon, where these distances are approximately 500 m and $50,000$ m, respectively, in the physical space.

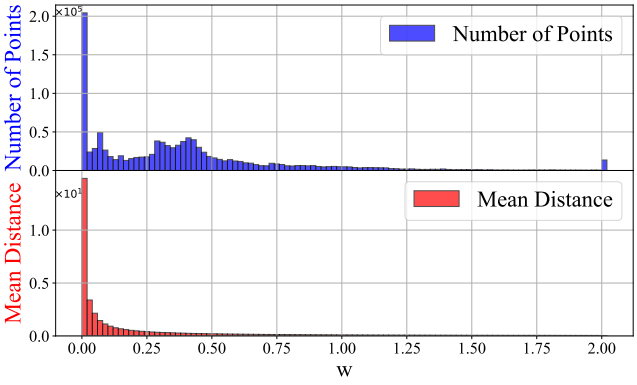


Figure S6. Histogram of w.

E. Additional Results

We provide quantitative comparisons per scene in Tables S5–S9, and qualitative results in Figs. S7 and S8 to supplement the main paper. These include detailed figures and tables showcasing the performance of our method across various scenes.

[1]Due to its proprietary nature, this dataset is used exclusively for experimental purposes.

Scene	Method	Overall			Near			Far		
		SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
TRAIN	3DGS	0.811	22.16	0.214	0.854	23.86	0.151	0.960	28.49	0.063
	Scaffold-GS	0.825	**22.69**	0.200	**0.859**	**23.87**	**0.146**	0.969	30.36	0.053
	HoGS (Ours)	**0.828**	22.60	**0.191**	0.854	23.70	0.147	**0.976**	**30.43**	**0.044**
TRUCK	3DGS	0.878	25.50	0.152	0.908	27.42	0.130	0.973	30.31	0.022
	Scaffold-GS	0.883	25.92	**0.139**	0.911	**27.87**	**0.118**	0.974	30.71	0.020
	HoGS (Ours)	**0.887**	**25.94**	0.141	**0.913**	27.68	0.123	**0.977**	**31.11**	**0.018**

Table S7. **Comparisons for Tanks&Temples scenes with near/far metrics.** 3DGS-based methods' SSIM, PSNR and LPIPS scores for Tanks&Temples scenes.

Scene	Method	Overall			Near			Far		
		SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
24	3DGS	0.944	31.77	0.094	0.951	32.35	0.085	0.995	41.85	0.006
	Scaffold-GS	0.946	31.98	0.090	0.950	32.32	0.084	0.997	44.33	0.003
	HoGS (Ours)	**0.953**	**32.70**	**0.078**	**0.955**	**32.95**	**0.075**	**0.998**	**46.82**	**0.002**
26	3DGS	0.904	29.48	0.185	0.911	30.13	0.176	0.994	39.67	0.007
	Scaffold-GS	0.895	29.66	0.194	0.902	30.17	0.186	0.995	40.68	0.006
	HoGS (Ours)	**0.913**	**30.20**	**0.165**	**0.918**	**30.64**	**0.159**	**0.996**	**41.71**	**0.005**
101	3DGS	0.924	27.54	0.091	0.931	27.97	0.083	0.994	39.59	0.007
	Scaffold-GS	0.907	26.94	0.106	0.911	27.20	0.102	0.997	41.01	0.003
	HoGS (Ours)	**0.936**	**28.26**	**0.071**	**0.939**	**28.50**	**0.068**	**0.998**	**42.40**	**0.002**

Table S8. **Comparisons for DL3DV Benchmark indoor scenes.** 3DGS-based methods' SSIM, PSNR and LPIPS scores for DL3DV Benchmark indoor scenes.

E.1. Distribution of w

We provide a histogram of the parameter w and the mean distance for each w on the TRAIN scene from Tanks & Temples [3]. indicates that w correctly converges to small values (*i.e.*, $w \sim 0$) for distant scenes, and *vice versa*.

E.2. Comparison with Mip-Splatting and Multi-Scale 3D Gaussian Splatting

We provide a comparison with 3DGS-based methods [5, 8] that focus on anti-aliasing and multi-scale representations. Even with these methods, our method still maintains state-of-the-art performance across most metrics.

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 *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2

[2] Bernhard Kerbl, Andreas Meuleman, Georgios Kopanas, Michael Wimmer, Alexandre Lanvin, and George Drettakis. A hierarchical 3D gaussian representation for real-time rendering of very large datasets. *ACM Transactions on Graphics (TOG)*, 43(4):62:1–62:15, 2024. 2

[3] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (TOG)*, 36(4): 78:1–78:13, 2017. 1, 2, 5

[4] Lu Ling, Yichen Sheng, Zhi Tu, Wentian Zhao, Cheng Xin, Kun Wan, Lantao Yu, Qianyu Guo, Zixun Yu, Yawen Lu, et al. DL3DV-10k: A large-scale scene dataset for deep learning-based 3d vision. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 2

[5] Zhiwen Yan, Weng Fei Low, Yu Chen, and Gim Hee Lee. Multi-scale 3D gaussian splatting for anti-aliased rendering. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 4, 5

[6] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth Anything v2. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. 3

[7] Chongjie Ye, Yinyu Nie, Jiahao Chang, Yuantao Chen, Yihao Zhi, and Xiaoguang Han. GauStudio: A modular framework for 3D gaussian splatting and beyond. *arXiv:2403.19632*, 2024. 2

[8] Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-splatting: Alias-free 3D gaussian splat-

Scene	Method	Overall			Near			Far		
		SSIM$^{\uparrow}$	PSNR$^{\uparrow}$	LPIPS$^{\downarrow}$	SSIM$^{\uparrow}$	PSNR$^{\uparrow}$	LPIPS$^{\downarrow}$	SSIM$^{\uparrow}$	PSNR$^{\uparrow}$	LPIPS$^{\downarrow}$
21	3DGS	0.838	25.99	0.196	0.870	27.72	0.147	0.972	31.23	0.042
	Scaffold-GS	0.846	27.00	0.181	0.866	28.11	0.148	0.983	34.11	0.028
	HoGS (Ours)	**0.872**	**27.80**	**0.148**	**0.889**	**28.81**	**0.122**	**0.986**	**35.48**	**0.022**
69	3DGS	0.858	26.41	0.165	0.891	27.77	0.131	0.972	32.83	0.027
	Scaffold-GS	0.870	27.23	0.152	0.890	28.11	0.132	0.985	35.14	0.015
	HoGS (Ours)	**0.888**	**27.50**	**0.129**	**0.902**	**28.19**	**0.115**	**0.990**	**36.50**	**0.010**
97	3DGS	0.930	29.62	0.127	0.952	32.47	0.066	0.980	33.57	0.058
	Scaffold-GS	0.945	31.82	0.099	0.956	32.87	0.056	0.991	38.81	0.042
	HoGS (Ours)	**0.953**	**33.12**	**0.092**	**0.962**	**34.12**	**0.051**	**0.992**	**40.41**	**0.039**

Table S9. **Comparisons for DL3DV Benchmark outdoor scenes.** 3DGS-based methods' SSIM, PSNR and LPIPS scores for DL3DV Benchmark outdoor scenes.

ting. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 4, 5

[9] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. NeRF++: Analyzing and improving neural radiance fields. *arXiv:2010.07492*, 2020. 1

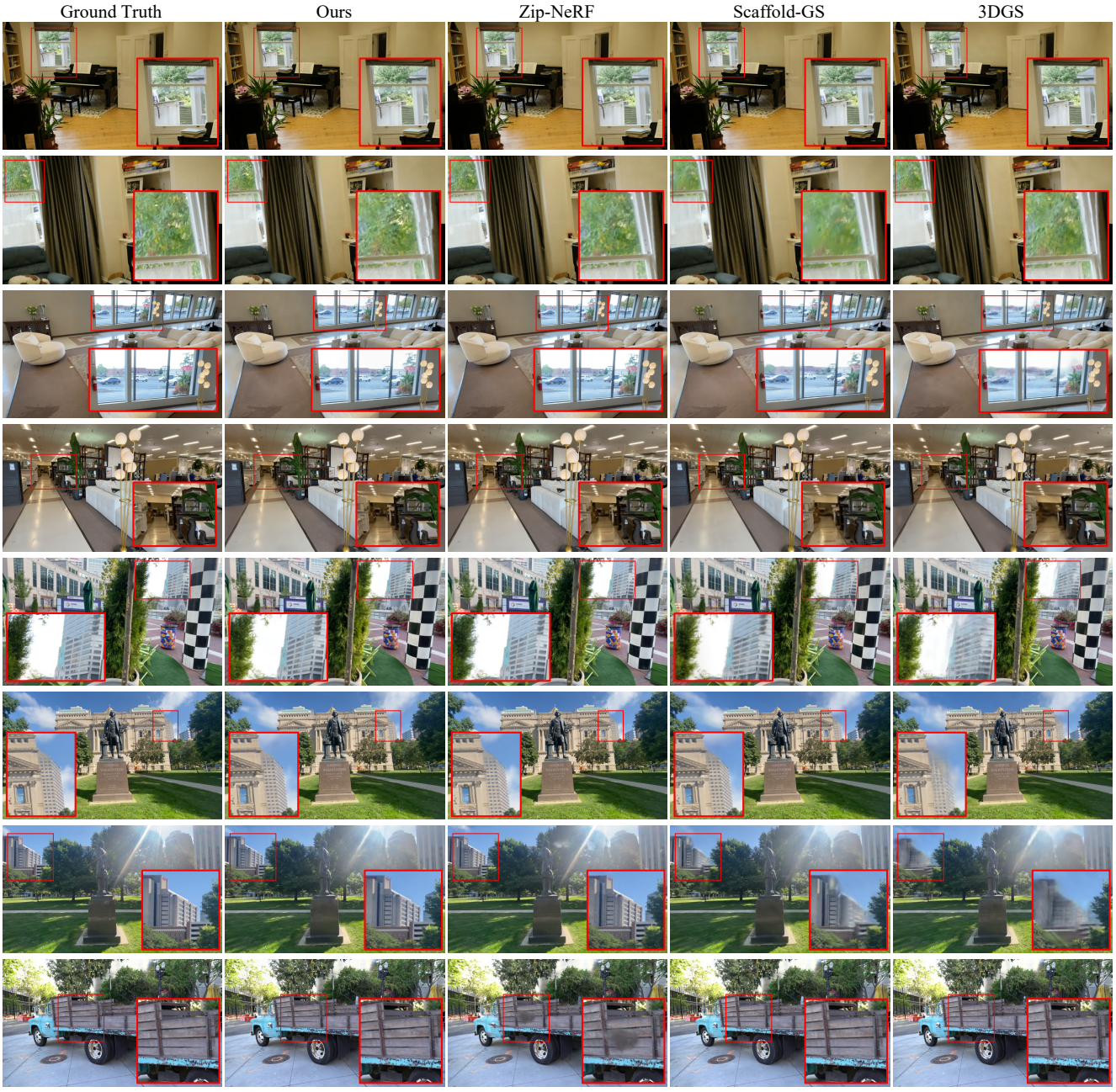


Figure S7. **Extra visual comparisons on novel view synthesis 1.** We present visual comparisons between methods on more test views. The figure includes ROOM from the Mip-NeRF360 dataset, TRUCK from Tanks&Temples, and scenes 21, 26, and 69 from the DL3DV benchmark. Key differences in quality are highlighted by insets.

Figure S8. **Extra visual comparisons on novel view synthesis 2.** We present visual comparisons between methods on more test views. The figure includes STUMP, BICYCLE, and TREEHILL from the Mip-NeRF360 dataset; and TRAIN from Tanks&Temples. Key differences in quality are highlighted by insets.