OccluGaussian: Occlusion-Aware Gaussian Splatting for Large Scene Reconstruction and Rendering

Supplementary Material

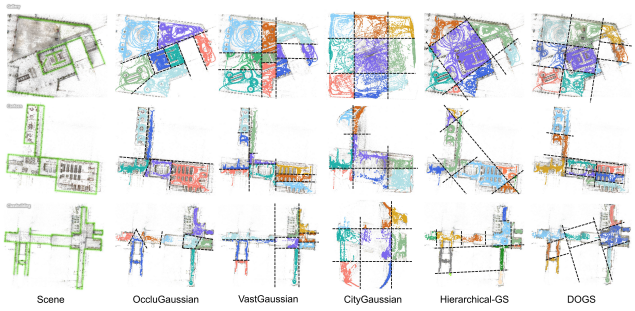

Figure 1. Comparative analysis of scene division on GALLERY, CANTEEN and CLASSBUILDING in the OcclusionScene3D dataset. Notably, CityGaussian's division is projected onto a contracted space.

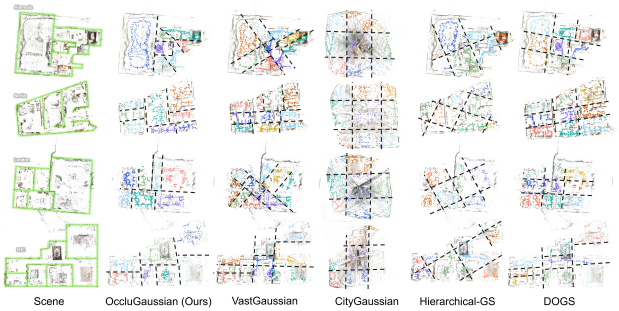

Figure 2. Comparative analysis of scene division on ALAMEDA, BERLIN, LONDON and NYC in the Zip-NeRF dataset. Notably, CityGaussian's division is projected onto a contracted space.

1. More Details of the OccluScene3D Dataset

There are three scenes in the OccluScene3D dataset: GALLERY, CANTEEN and CLASSBUILDING. All the videos are recorded by a mobile phone with the wide-angle mode and landscape orientation at a frame rate of 60 Hz. We use COLMAP [8] to estimate the camera intrinsic and extrinsic parameters. More details are shown in Tab. 1.

2. Additional Experimental Results

2.1. Additional Quantitative Analysis

Scene division strategy analysis. We present a comparison of scene division on the OccluScene3D dataset and the Zip-NeRF dataset among VastGaussian [5], CityGaussian [7], Hierarchical-GS [4], DOGS [2], and OccluGaussian in Fig. 1 and Fig. 2. The results demonstrate that the scene divisions produced by OccluGaussian better align with scene layouts. Consequently, OccluGaussian achieves superior reconstruction quality, which is proven by performing 3DGS optimization with the same hyperparameters across different scene division strategies, as illustrated in Fig. 3. We also present the division results of OccluGaussion on the Mill-19 dataset [9] and UrbanScene3D dataset [6] in Fig. 4, which demonstrate that our method also generalizes well in aerial capture scenes without occlusions.

Scene	Area (m^2)	#Video	Duration (min)	#Image
GALLERY	2500	71	174	9881
CANTEEN	1500	46	111	9034
CLASSBUILDING	1000	18	76	9118

Table 1. Statistics of the OccluScene3D dataset with three real scenes. Area: covered area; #Video: total number of videos; Duration: total duration of all the videos; #Image: total number of sampled images for reconstruction.

	PSNR	Train (m)	Mem (GB)	#GS (M)	FPS
VastGaussian	24.58	49.97	15.0	5.5	232
CityGaussian	20.96	106.35	**8.2**	6.4	80
Hierarchical-GS	22.94	190.36	30.5	40.0	205
3DGS	20.88	86.8	23.9	**0.9**	**422**
OccluGaussian	**25.46**	**48.26**	12.2	3.4	313

Table 2. More quantitative results on OccluScene3D dataset.

Detailed quantitative analysis of OccluGaussian. We present the average PSNR, training time, allocated memory, number of gaussians and FPS on OccluScene3D dataset in Tab. 2. We validate the methods on the Mill-19 [9], UrbanScene3D [6] and Zip-NeRF [1] datasets in Tab. 3 and Tab. 4. It can be observed that OccluGaussian outperforms others in terms of LPIPS among all the datasets. It also overall holds a clear advantage over existing methods in other metrics, highlighting its generality. These results show the effectiveness of our scene division strategy, which strengthens the correlations among training cameras within each region, and achieves a higher average contribution to the reconstruction results.

Extended camera ratios. We further compare the extended camera ratios of division strategy with those from VastGaussian [5], CityGaussian [7], Hierarchical-GS [4], DOGS [2] and OccluGaussian, as shown in Tab. 5. This ratio is defined as the total number of extended cameras across all regions divided by the total number of training cameras. A smaller ratio of extended cameras indicates that each

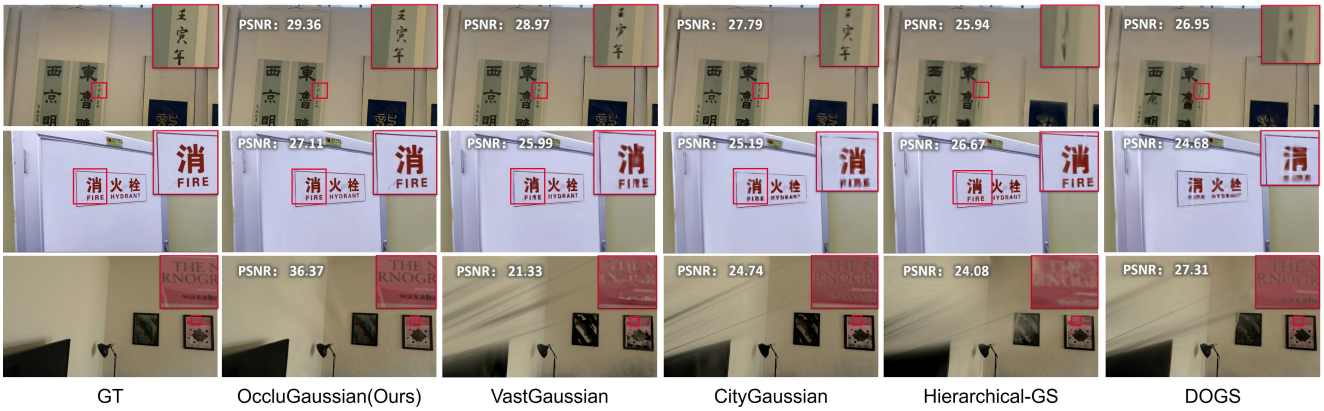

Figure 3. Quantitative evaluation of different division strategies applied under the same 3DGS optimization hyperparameters. Lines from top to bottom are the GALLERY, and CLASSBUILDING scenes from the OccluScene3D dataset, as well as the BERLIN scene from the Zip-NeRF dataset.

Scene	MILL-19						URBANSCENE3D								
	BUILDING			RUBBLE			CAMPUS			RESIDENCE			SCI-ART		
Metrics	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
VastGaussian [5]	23.50	0.804	0.130	26.92	0.823	0.132	26.00	0.816	0.151	**24.25**	**0.852**	0.124	**26.81**	0.885	0.121
CityGaussian [7]	22.59	0.757	0.174	25.98	0.784	0.158	-	-	-	23.25	0.806	0.156	24.49	0.836	0.167
Hierarchical-GS [4]	21.25	0.723	0.297	24.64	0.755	0.284	-	-	-	-	-	-	-	-	-
3DGS [3]	23.01	0.769	0.164	26.78	0.800	0.161	23.89	0.712	0.289	23.40	0.825	0.142	25.24	0.843	0.166
OccluGaussian [7]	**24.77**	**0.853**	**0.100**	**27.16**	**0.854**	**0.105**	**26.60**	**0.836**	**0.139**	24.24	0.846	**0.122**	26.42	**0.890**	**0.113**

Table 3. Quantitative evaluation of our method compared to previous work on the Mill-19 [9] and UrbanScene3D [6] datasets. The **best** and second best results are highlighted. Due to an out-of-memory issue, we were unable to test Hierarchical-GS [22] on the URBANSCENE3D dataset and CityGaussian on the CAMPUS scene.

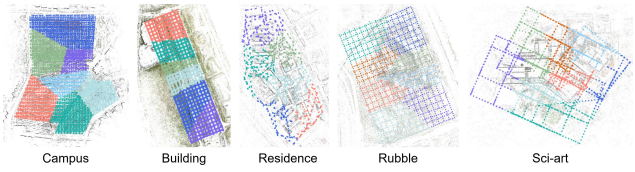

Figure 4. Scene division results by OccluGaussian on the Mill-19 and UrbanScene3D datasets.

partitioned region is more self-contained and can be well-reconstructed with fewer training iterations. Conversely, if region A and B are divided into occlusion-agnostic regions, many extended cameras for region A will be inside region B, and vice versa. In other words, these two sets of training cameras are almost the same (all cameras), making it difficult to achieve good reconstruction with limited training iterations. Compared to VastGaussian, CityGaussian, Hierarchical-GS and DOGS, our extended cameras are greatly reduced by our occlusion-aware division strategy, which leads to a higher average contribution of all training cameras for the reconstruction of each region. This advantage eventually leads to improved reconstruction quality, as proven by the results obtained when replacing our

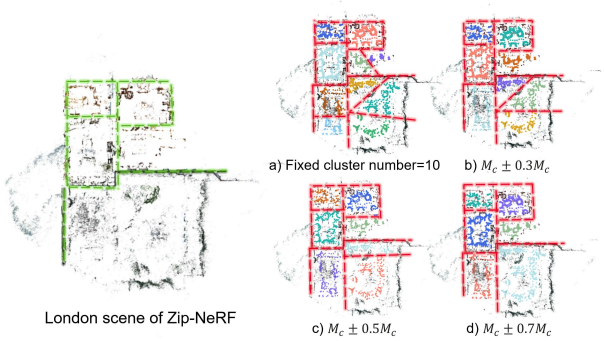

Figure 5. Division results of different camera count ranges for each cluster. The green lines denote the physical walls, and the red lines denote the boundaries of the divided regions.

scene division strategy with others in our 3DGS optimization pipeline, as shown in Tab. 5.

2.2. Additional Ablations

Cluster number refinement. To obtain our final scene division, we iteratively refine the result from the graph clustering algorithm by splitting or ignoring clusters until the

	BERLIN			LONDON			NYC			ALAMEDA		
Metrics	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
3DGS [3]	28.52	0.887	0.325	27.05	0.829	0.342	28.21	0.844	0.321	25.35	0.758	0.37
SMERF [**?**]	28.52	0.887	0.325	27.05	0.829	0.342	28.21	0.844	0.321	25.35	0.758	0.37
Zip-NeRF [1]	28.59	0.891	0.297	27.06	0.835	0.304	28.42	0.850	0.281	**25.41**	0.767	0.338
OccluGaussian	**30.37**	**0.937**	**0.076**	**28.06**	**0.868**	**0.141**	**31.33**	**0.902**	**0.121**	24.75	**0.814**	**0.201**

Table 4. Quantitative evaluation of our method compared to previous work on the Zip-NeRF [1] datasets.

	PSNR	Base cam.	Extended cam.	Ratio
VastGaussian [5]	24.58	9119	8921	97.8%
CityGaussian [7]	22.50	9119	5371	58.90%
Hierarchical-GS [4]	23.93	9119	6808	74.66%
DOGS [2]	25.11	9119	8331	91.36%
OccluGaussian	**25.81**	9119	**4232**	**46.41%**

Table 5. Extended camera comparison on OccluScene3D.

	#Clusters		Metrics		
	Initial K	Final K	PSNR	SSIM	LPIPS
Fixed cluster number	10	10	28.092	0.870	0.141
$M_c \pm 0.3M_c$	10	8	28.090	0.870	0.141
$M_c \pm 0.5M_c$	10	7	28.060	0.868	0.141
$M_c \pm 0.7M_c$	10	6	28.043	0.862	0.142

Table 6. Quantitative results of different camera count ranges.

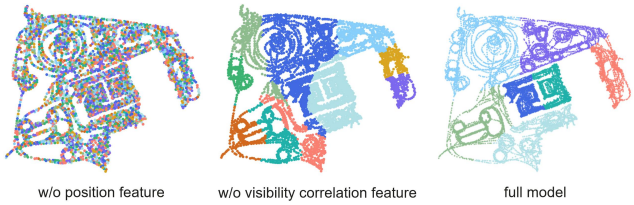

Figure 6. Ablation study of features in our attributed scene graph.

camera count in each cluster falls within $[M_c - \sigma_c M_c, M_c + \sigma_c M_c]$, where M_c is the average camera count across all clusters. Here we explore the influence with different σ_c choices, as shown in Fig. 5 and Tab. 6. The division results remain occlusion-aware, and similar performances are obtained. Note that without this refinement step, comparable reconstruction quality can still be achieved; however, more GPU resources are required as more regions need to be reconstructed.

Features in the scene graph. As shown in Fig. 6, both the position features in the node attribute and the visibility correlations encoded by the edge weight in the scene graph are important to achieve occlusion-aware division.

Region-based rendering. The ablation study of our region-based rendering is conducted on the OccluScene3D dataset, and the visual comparisons are shown in Fig. 7. Our region-based culling strategy substantially enhances rendering speeds without any perceptible loss in visual quality. Moreover, the proposed region subdivision technique further accelerates rendering.

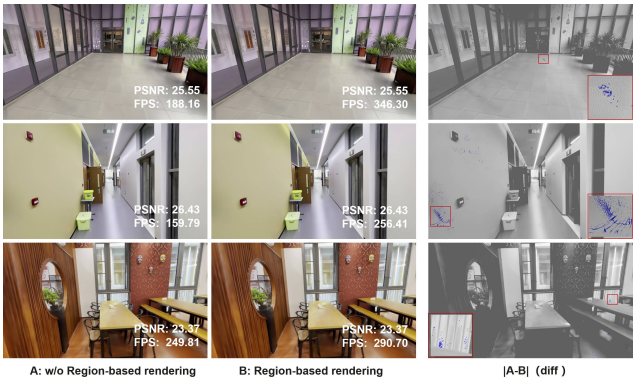

Figure 7. Quantitative evaluation of region-based rendering. The right-most image uses color coding to represent pixel differences: red for $|A - B| > 1$, blue for $|A - B| = 1$, and gray for A-B = 0. The difference between A and B is almost less than 1 pixel. Our approach achieves substantial enhancements in rendering speed without a perceptible loss in image quality.

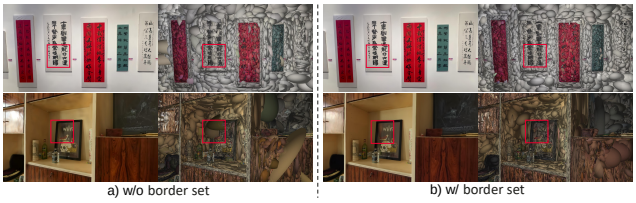

Figure 8. Ablation study of border set in training camera selection.

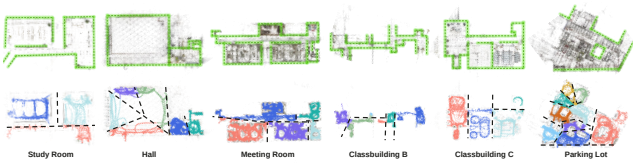

Figure 9. Scene division on the OcclusionScene3D-E dataset.

Training camera selection.The ablation study of border set in training camera selection is shown in Fig. 8. Without the border set, Gaussian primitives can become excessively

Figure 10. Quantitative Evaluation on the OcclusionScene3D-E Dataset. The first line shows the ground truth, while the second line presents the rendering results of our OccluGaussian method.

Scene	Area (m^2)	#Video	Duration (min)	#Image
STUDY ROOM	500	39	105	4752
HALL	800	33	97	6000
MEETING ROOM	1500	36	127	10000
CLASSBUILDING B	1000	27	63	7603
CLASSBUILDING C	1000	25	76	6435
PARKING LOT	500	38	78	6060

Table 7. Statistics of the OccluScene3D-E dataset with six real scenes.

Scene	PSNR	SSIM	LPIPS	#GS (M)	FPS
STUDY ROOM	22.52	0.84	0.190	2.6	314.73
HALL	22.70	0.79	0.212	1.7	341.76
MEETING ROOM	27.00	0.90	0.12	3.5	407.51
CLASSBUILDING B	22.24	0.85	0.158	1.8	595.38
CLASSBUILDING C	23.90	0.88	0.123	1.4	420.94
PARKING LOT	22.46	0.86	0.146	1.6	314.73

Table 8. Quantitative results on the OccluScene3D-E dataset.

large or elongated, leading to floaters or artifacts.

3. Extended Dataset

To further advance research in occluded scene reconstruction, we introduce an extended version of the OccluScene3D dataset named OccluScene3D-E. It encompasses six diverse scenes: STUDY ROOM, HALL, MEETING ROOM, CLASSBUILDING B, CLASSBUILDING C, and PARKING LOT, with detailed information in Tab. 7. As demonstrated in Fig. 9, applying OccluGaussian's scene division method to OccluScene3D-E yields results that closely align with the underlying scene structures. Tab. 8 details the quantitative metrics, including PSNR, SSIM, LPIPS, the number of gaussians, and FPS for OccluScene3D-E. The rendering outcomes for each scene, illustrated in Fig. 10, collectively confirm the robustness and effectiveness of our scene division strategy, as well as its capacity to produce high-fidelity visual results.

Initial K	Final K	PSNR	SSIM	LPIPS
1	1	25.18	0.882	0.145
5	5	25.35	0.895	0.114
7	7	25.59	0.902	0.103
10	7	25.81	0.903	0.099
15	10	25.33	0.899	0.101

Table 9. Different initial clustering numbers K on the Gallery scene of OccluScene3D dataset.

References

[1] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Zip-NeRF: Anti-aliased grid-based neural radiance fields. In *ICCV*, 2023. 1, 3

[2] Yu Chen and Gim Hee Lee. DOGS: Distributed-oriented gaussian splatting for large-scale 3d reconstruction via gaussian consensus. *NeurIPS*, 37:34487–34512, 2025. 1, 3

[3] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *TOG*, 2023. 2, 3

[4] 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. *TOG*, 2024. 1, 2, 3

[5] Jiaqi Lin, Zhihao Li, Xiao Tang, Jianzhuang Liu, Shiyong Liu, Jiayue Liu, Yangdi Lu, Xiaofei Wu, Songcen Xu, Youliang Yan, et al. VastGaussian: Vast 3D Gaussians for large scene reconstruction. In *CVPR*, 2024. 1, 2, 3

[6] Liqiang Lin, Yilin Liu, Yue Hu, Xingguang Yan, Ke Xie, and Hui Huang. Capturing, reconstructing, and simulating: the urbanscene3d dataset. In *ECCV*, 2022. 1, 2

[7] Yang Liu, Chuanchen Luo, Lue Fan, Naiyan Wang, Junran Peng, and Zhaoxiang Zhang. Citygaussian: Real-time high-quality large-scale scene rendering with gaussians. In *ECCV*. Springer, 2024. 1, 2, 3

[8] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *CVPR*, 2016. 1

[9] Haithem Turki, Deva Ramanan, and Mahadev Satyanarayanan. Mega-NeRF: Scalable construction of large-scale NeRFs for virtual fly-throughs. In *CVPR*, 2022. 1, 2