

S2Gaussian: Sparse-View Super-Resolution 3D Gaussian Splatting

- Supplementary Material

1. More Ablation Studies

Rendering Speed. As shown in Tab. 1, our method maintains the high rendering speed characteristic of Gaussian splatting frameworks. While the FPS is slightly lower than that of 3DGS and FSGS due to the inclusion of more Gaussian primitives for representing detailed high-resolution scenes, our approach has been demonstrated to significantly outperform existing methods, effectively addressing sparse views and clarity deficiencies simultaneously.

Table 1. Rendering speed comparison with NeRF-SR [6], 3DGS [3], and FSGS [9].

Method	NeRF-SR	3DGS	FSGS+SRGS	S2Gaussian
FPS	<1	438	428	387

Number of Pseudo Views. We also performed experiments to investigate the effect of the number of pseudo views and the results of which are presented in Tab. 2. It is observed that incorporating 72 pseudo views per scene, equivalent to three times the number of known views yields better performance. While more pseudo-views would introduce more potential improper supervision and thus observe performance degradation. Based on these results, we have chosen pseudo views amounting to three times the number of known views as our default setup.

Table 2. Ablation study of the number of pseudo views on Mip-NeRF 360 \times 4 (24 known views).

Pseudo Views	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FID \downarrow
24	21.39	0.632	0.351	61.23
48	21.87	0.674	0.311	52.63
72	22.05	0.687	0.296	43.51
96	21.96	0.682	0.298	45.65

Effect of α and λ in Gaussian Shuffle. In Fig. 1, we illustrate the impact of α on the densified 3D scenes. A larger α value enables the sub-Gaussians to disperse more widely, which facilitates high-resolution scene fitting but also exacerbates scene degradation. Based on this trade-off, we set $\alpha = 0.5$, as higher values lead to noticeable quality deterioration, undermining our objectives.

Additionally, we also explore the impact of λ on densified 3D scenes. Since the visual changes are minimal, we calculated the MSE values before and after the Gaussian Shuffle to quantify the effect. As shown in Table 3, too large or too small λ can lead to serious damage to the scene while the values between 1.8 and 2.0 are relatively optimal, hence we set $\lambda = 1.9$ which minimizes the changes to the scene, reducing the possible corruption of the 3D representation.

Table 3. Ablation study of λ , where MSE is calculated to quantify the effects before and after Gaussian Shuffle.

λ	1.0	1.8	1.9	2.0	3.0
MSE	95.7	43.9	42.8	43.6	148.6

Effect of ϵ in 3D Robust Optimization. As shown in Tab. 4, setting ϵ too small diminishes its ability to modulate update trends, while setting it too large weakens control over undesired gradients, reducing its effectiveness in suppressing their influence. Consequently, we set ϵ to 0.1 as the default value in all experiments.

Table 4. Ablation study of ϵ on Mip-NeRF 360 \times 4 (24 views).

ϵ	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FID \downarrow
0.01	21.29	0.631	0.347	59.69
0.05	21.79	0.666	0.325	44.12
0.1	22.05	0.687	0.296	43.51
0.2	21.85	0.675	0.316	47.28
0.3	21.63	0.657	0.311	46.37

2. More Experimental Results

Since most 2D super-resolution models such as ResShift [8] are typically trained on $\times 4$ data, we primarily experiment on $4 \times$ super-resolution tasks in the maintext. In the Suppl. Mat., we further performed experiments $2 \times$ and $8 \times$ super-resolution tasks. Specifically, for the $2 \times$ super-resolution task we concatenate a bicubic interpolation downsampling operation after ResShift to get the $\times 2$ results, and for the $8 \times$ super-resolution task we concatenate a bicubic interpolation upsampling operation after ResShift to get the $\times 8$ results.

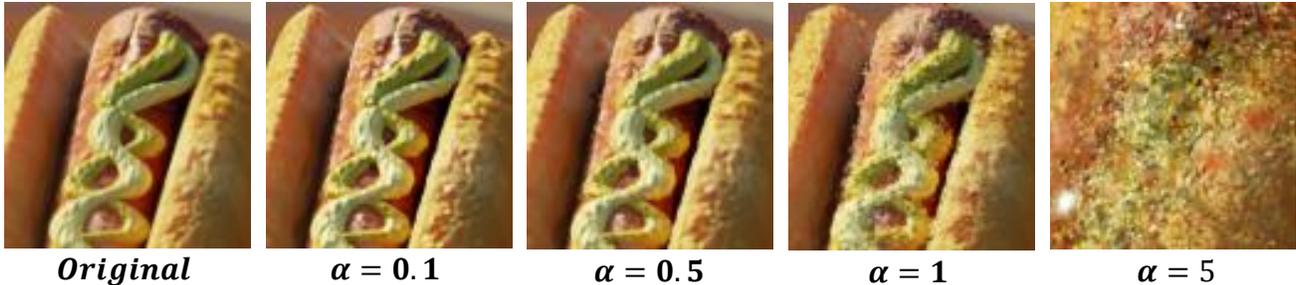


Figure 1. Visual comparison with different α in Gaussian Shuffle.

Table 5. Quantitative comparison on Blender $\times 2$ (8 views), LLFF $\times 2$ (3 views), and Mip-NeRF360 $\times 2$ (24 views). The best, second best, and third best entries are marked in red, orange, and yellow, respectively.

Method	Blender $\times 2$ (8 views)				LLFF $\times 2$ (3 views)				Mip-NeRF 360 $\times 2$ (24 views)			
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FID \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FID \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FID \downarrow
3DGS [3]	22.47	0.856	0.109	27.54	14.84	0.371	0.368	94.81	17.11	0.424	0.382	72.32
SRGS [2]	24.08	0.883	0.085	21.28	19.26	0.587	0.227	58.63	19.02	0.509	0.414	69.32
Mip-Splatting [7]	24.18	0.887	0.074	21.48	15.58	0.430	0.334	90.71	17.68	0.469	0.366	64.61
FSGS [9]	23.15	0.866	0.091	29.06	19.92	0.603	0.149	45.79	19.72	0.529	0.373	87.55
FSGS [9]+SRGS [2]	23.98	0.880	0.099	20.63	20.47	0.622	0.133	44.56	19.91	0.536	0.422	82.04
S2Gaussian (Ours)	24.58	0.891	0.071	19.65	20.98	0.649	0.119	40.15	22.53	0.712	0.280	42.56

Table 6. Quantitative comparison on Blender $\times 8$ (8 views), LLFF $\times 8$ (3 views), and Mip-NeRF360 $\times 8$ (24 views). The best, second best, and third best entries are marked in red, orange, and yellow, respectively.

Method	Blender $\times 8$ (8 views)				LLFF $\times 8$ (3 views)				Mip-NeRF 360 $\times 8$ (24 views)			
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FID \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FID \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FID \downarrow
3DGS [3]	18.86	0.818	0.210	68.67	10.68	0.211	0.635	160.11	15.06	0.310	0.567	125.93
SRGS [2]	22.25	0.864	0.126	34.28	18.65	0.512	0.387	125.32	18.09	0.435	0.502	106.75
Mip-Splatting [7]	22.08	0.863	0.159	51.86	15.72	0.449	0.536	142.94	18.29	0.459	0.517	94.95
FSGS [9]	19.53	0.818	0.190	72.35	16.63	0.446	0.412	143.85	17.29	0.407	0.500	117.17
FSGS [9]+SRGS [2]	22.43	0.864	0.119	31.68	19.30	0.530	0.362	100.37	19.43	0.477	0.520	103.42
S2Gaussian (Ours)	23.46	0.869	0.112	29.27	20.15	0.615	0.219	75.28	21.26	0.612	0.345	58.92

Quantitative Evaluation. Table 5 and Table 6 presents quantitative comparison results for $2 \times$ and $8 \times$ sparse-view super-resolution novel view synthesis tasks on the Blender [5], LLFF [4], and Mip-NeRF 360 datasets [1]. It can be observed that our method still achieves state-of-the-art performance, particularly on the more challenging $8 \times$ super-resolution task, where S2Gaussian outperforms FSGS+SRGS by 1.83 dB in PSNR, while reducing FID by nearly half.

Qualitative Evaluation. We also demonstrate visual comparisons in Fig. 2 and Fig. 3. It can be observed that existing techniques struggle to capture underlying scene details, even on $2 \times$ super-resolution tasks. In contrast, our method accurately reconstructs fine-grained, intricate details in both synthetic and real-world large-scale scenes, demonstrating the reliability and practicality of our method in realistic applications.

More Visual Comparison. Additionally, we provide visual comparisons of the rendering results across continuous viewpoints to further demonstrate the superiority of our

method in a more comprehensive and thorough manner. As shown in Fig. 4, our method reconstructs 3D scenes with more structurally accurate and detail-rich representations, consistently across different viewpoints. Especially its superiority becomes even more pronounced in large-scale, complex real-world scenes (the third column), demonstrating valuable practicality and applicability.

3. Impact, Limitation, and Future Work

S2Gaussian makes a substantial contribution to advancing 3D scene reconstruction by effectively addressing the challenges posed by sparse and low-resolution views. Its capability to generate geometrically precise and richly detailed reconstructions makes it a valuable tool for real-world applications, particularly where high-quality input views are limited. Furthermore, by reducing reliance on dense and high-resolution views, S2Gaussian lowers the entry barriers for organizations and researchers with constrained resources. The framework’s innovative use of Gaussian Shuf-

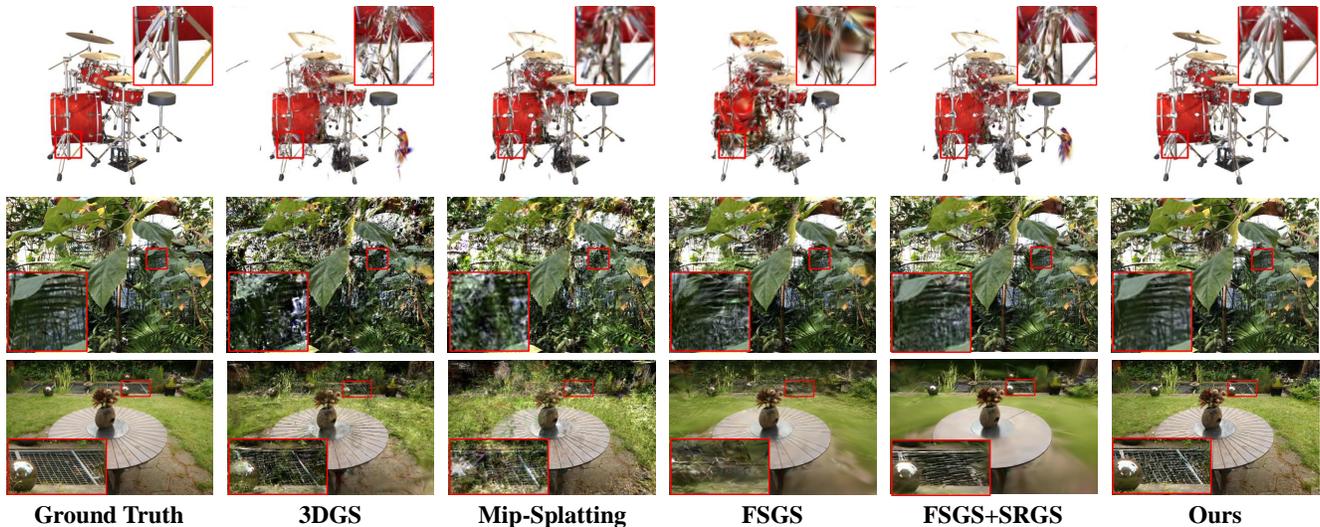


Figure 2. Qualitative comparisons on Blender $\times 2$ (8 views), LLFF $\times 2$ (3 views), and Mip-NeRF360 $\times 2$ (24 views).

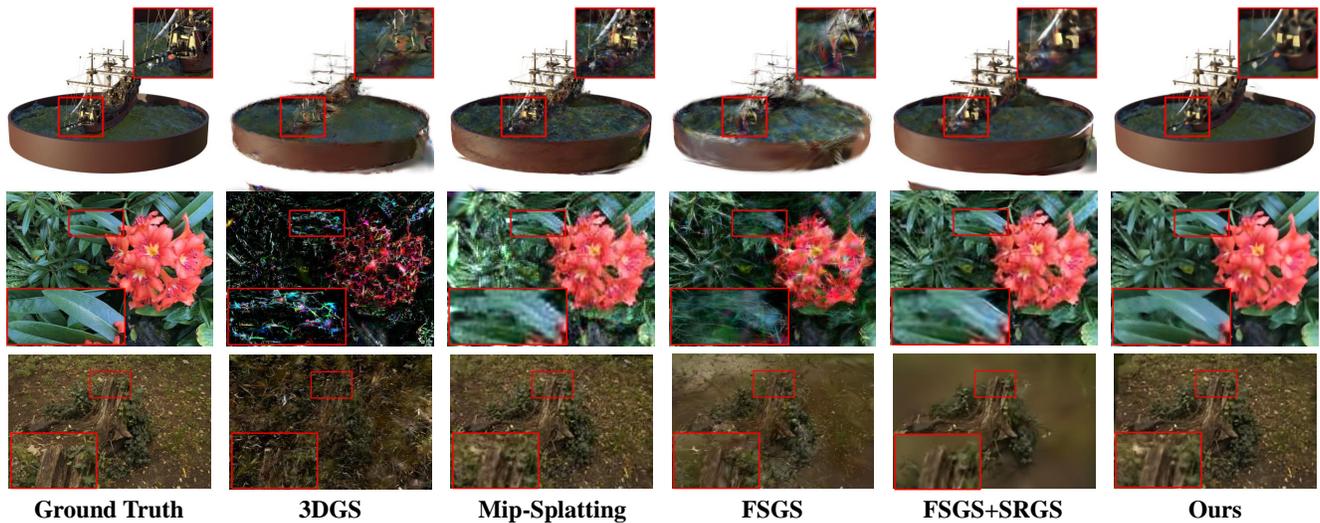


Figure 3. Qualitative comparisons on Blender $\times 8$ (8 views), LLFF $\times 8$ (3 views), and Mip-NeRF360 $\times 8$ (24 views).

fle and dual-stage optimization offers fresh insights into 3D Gaussian Splatting, paving the way for further advancements in reconstructing complex 3D scenes from more challenging input views.

One potential limitation of our current method is its focus on the reconstruction of static 3D scenes and is not yet well-suited for handling dynamic time-varying scenes. While the framework demonstrates robust performance for static environments, extending its capabilities to reconstruct dynamic 4D scenes where temporal variations and motion need to be accurately captured—remains an uncharted yet promising direction. Integrating our approach into the emerging paradigm of 4D Gaussian Splatting could provide a potential solution, enabling it to model both spatial and temporal dynamics with sparse and low-resolution time-

varying views.

Moving forward, we envision S2Gaussian as a powerful tool capable of unlocking diverse applications in real-world 3D scene reconstruction. As we chart the course ahead, one avenue of exploration involves extending our method to dynamic 4D scene reconstruction, enabling more immersive rendering with broad applications in interactive AR/VR environments and dynamic navigation in robotics. In essence, our journey continues not only in refining our method but also in pushing the boundaries of what 3DGS can offer to the broader field of 3D vision.

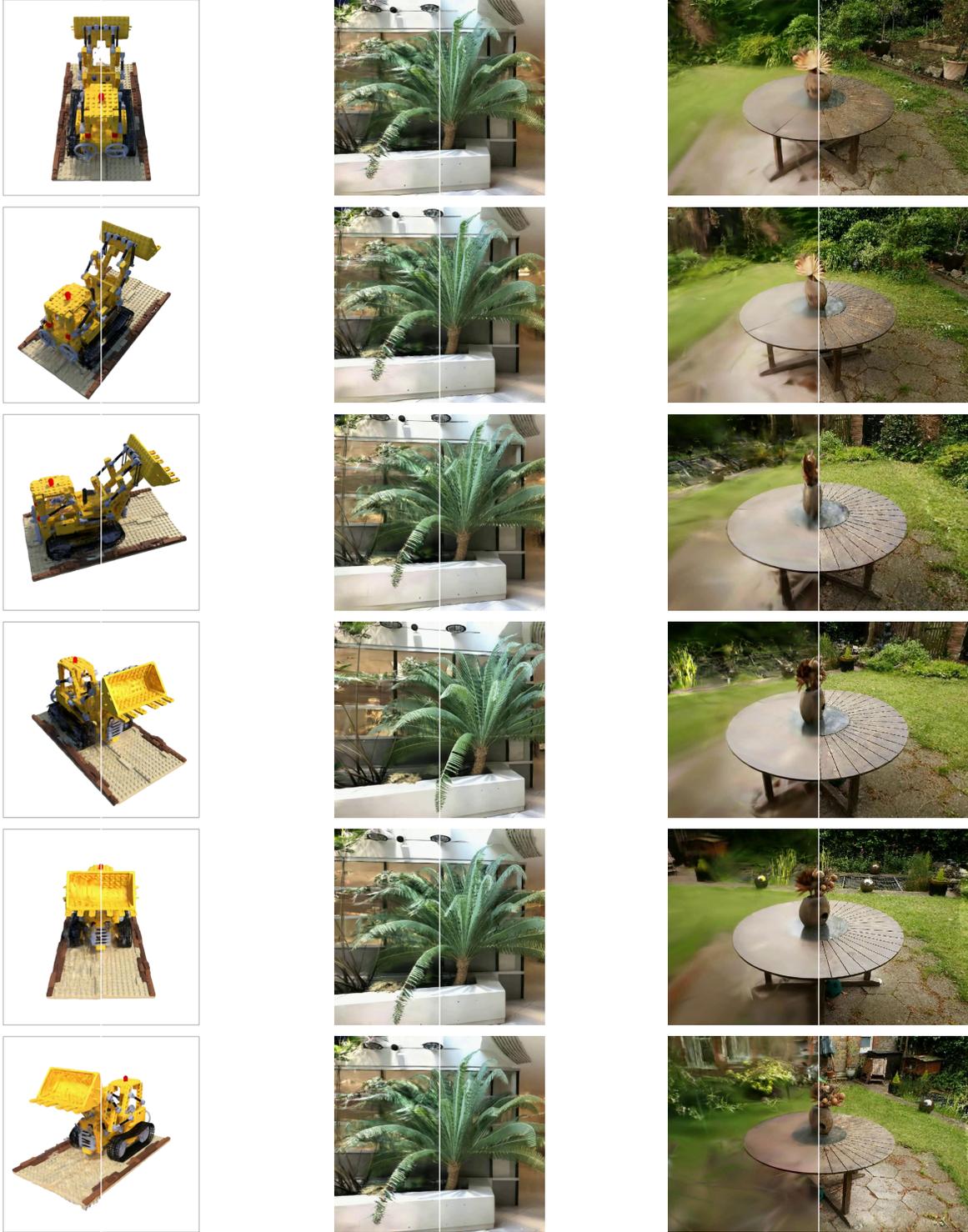


Figure 4. Visual comparison of consecutive views between combined method FSGS+SRGS (left) and our proposed S2Gaussian (right). Our method exhibits significantly superior performance in complex real-world scenes.

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 the IEEE/CVF conference on computer vision and pattern recognition*, pages 5470–5479, 2022. [2](#)
- [2] Xiang Feng, Yongbo He, Yubo Wang, Yan Yang, Zhenzhong Kuang, Yu Jun, Jianping Fan, et al. Srgs: Super-resolution 3d gaussian splatting. *arXiv preprint arXiv:2404.10318*, 2024. [2](#)
- [3] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. [1](#), [2](#)
- [4] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (ToG)*, 38(4):1–14, 2019. [2](#)
- [5] 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. *Communications of the ACM*, 65(1):99–106, 2021. [2](#)
- [6] Chen Wang, Xian Wu, Yuan-Chen Guo, Song-Hai Zhang, Yu-Wing Tai, and Shi-Min Hu. Nerf-sr: High quality neural radiance fields using supersampling. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6445–6454, 2022. [1](#)
- [7] Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-splatting: Alias-free 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19447–19456, 2024. [2](#)
- [8] Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. *Advances in Neural Information Processing Systems*, 36, 2024. [1](#)
- [9] Zehao Zhu, Zhiwen Fan, Yifan Jiang, and Zhangyang Wang. Fsgs: Real-time few-shot view synthesis using gaussian splatting. In *European Conference on Computer Vision*, pages 145–163. Springer, 2025. [1](#), [2](#)