
VA-GS: Enhancing the Geometric Representation of Gaussian Splatting via View Alignment

– Supplemental Material –

Qing Li¹ Huifang Feng^{2*} Xun Gong¹ Yu-Shen Liu³

¹ Southwest Jiaotong University, Chengdu, China

² Xihua University, Chengdu, China ³ Tsinghua University, Beijing, China

qingli@swjtu.edu.cn fhf@xhu.edu.cn xgong@swjtu.edu.cn liuyushen@tsinghua.edu.cn

This supplementary document is organized as follows: (1) We first present an overview of the baseline methods used in our evaluation experiments. (2) We then provide additional qualitative and quantitative experimental results to complement those in the main paper. (3) Further ablation studies are included to analyze the impact of key components in our method. (4) More discussion is provided to give a precise understanding of our approach. (5) We discuss the current limitations of our approach and outline potential directions for future research. (6) Finally, we consider the potential negative societal impacts associated with our work.

1 Baseline Methods

We compare our approach with state-of-the-art surface reconstruction and novel view synthesis methods based on 3D Gaussian Splatting, such as 3DGS [7], 2DGS [5], GOF [13], and PGSR [2]. In addition, we evaluate against implicit NeRF-based methods, such as NeuS [11] and Neuralangelo [9], which utilize Signed Distance Functions (SDFs) to represent scenes and convert them into opacity fields for volume rendering via ray tracing.

In our main paper, we report baseline results using values provided in the respective original publications whenever available. For visual comparisons shown in our figures, we generate the results using the official implementations released by the authors.

2 More Evaluation Results

2.1 Comparisons on the Deep Blending Dataset

In order to further test the performance of the algorithm on different real-world data, we evaluate both novel view synthesis and surface reconstruction on two commonly used indoor scenes, Dr Johnson and Playroom, from the Deep Blending dataset [4]. Table 1 presents quantitative comparisons against state-of-the-art Gaussian-based methods on novel view synthesis. All results, unless otherwise noted, are obtained from our own re-implementations using the official code released by the respective authors. For 3DGS [7], we use the pretrained models provided by the authors. For SuGaR [3], we report numbers directly from the original paper to avoid potential inconsistencies. Our method consistently outperforms all baselines across all metrics, demonstrating superior rendering quality and generalization to unseen viewpoints. These improvements can be attributed to our enhanced geometric representation, which yields better visual fidelity. Consistent with observations in the evaluation on the Mip-NeRF 360 dataset [1] presented in the main paper, methods such as 2DGS [5], SuGaR [3], and GS-Pull [16] underperform compared to vanilla 3DGS [7], suggesting that imposing planar Gaussian constraints may hinder performance in complex scenes.

*Corresponding author

Table 1: Quantitative comparison on the Deep Blending dataset. The best results are highlighted as 1st, 2nd and 3rd. * indicates results copied from the original paper.

	Dr Johnson			Playroom			Mean		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
3DGS [7]	28.94	0.896	0.248	29.93	0.901	0.244	29.43	0.898	0.246
SuGaR* [3]	28.71	0.889	0.273	30.12	0.898	0.261	29.41	0.893	0.267
2DGS [5]	28.89	0.898	0.259	29.88	0.901	0.259	29.38	0.899	0.259
GS-Pull [16]	25.69	0.830	0.387	25.89	0.838	0.375	25.79	0.834	0.381
GOF [13]	27.85	0.893	0.257	30.16	0.904	0.242	29.01	0.899	0.250
RaDe-GS [14]	27.83	0.896	0.257	30.04	0.905	0.243	28.94	0.901	0.250
PGSR [2]	28.61	0.891	0.251	29.92	0.903	0.243	29.27	0.897	0.247
Ours	29.10	0.900	0.241	30.22	0.905	0.241	29.66	0.903	0.241

Table 2: Quantitative results on the Mip-NeRF 360 dataset. We report PSNR, SSIM, and LPIPS for each scene. All results are from our own runs using official code, except for 3DGS, where we use the authors’ pre-trained models. The best results are highlighted as 1st, 2nd and 3rd.

	bicycle	flowers	garden	stump	treehill	room	counter	kitchen	bonsai	Mean
PSNR \uparrow										
3DGS [7]	25.17	21.45	27.18	26.56	22.30	31.34	28.89	30.71	31.98	27.29
2DGS [5]	24.71	21.03	26.61	26.11	22.27	30.72	28.08	30.30	31.24	26.79
GS-Pull [16]	24.19	20.56	26.08	25.24	22.60	30.84	26.41	26.06	29.10	25.68
RaDe-GS [14]	25.56	21.67	27.39	27.10	22.34	30.84	28.73	31.26	31.73	27.40
GOF [13]	25.44	21.59	27.27	26.93	22.40	30.42	28.62	30.64	31.50	27.20
PGSR [2]	25.66	21.52	27.49	26.98	22.29	30.06	28.31	30.80	31.55	27.18
Ours	25.89	21.71	27.69	27.33	22.38	30.73	28.77	31.09	31.93	27.50
SSIM \uparrow										
3DGS [7]	0.762	0.602	0.861	0.770	0.631	0.916	0.905	0.923	0.938	0.812
2DGS [5]	0.729	0.569	0.838	0.753	0.614	0.906	0.892	0.915	0.929	0.794
GS-Pull [16]	0.660	0.498	0.776	0.661	0.617	0.896	0.846	0.838	0.897	0.743
RaDe-GS [14]	0.793	0.640	0.870	0.801	0.648	0.916	0.905	0.924	0.939	0.826
GOF [13]	0.786	0.634	0.864	0.790	0.641	0.911	0.900	0.915	0.935	0.820
PGSR [2]	0.793	0.635	0.873	0.798	0.660	0.927	0.914	0.932	0.945	0.831
Ours	0.803	0.648	0.877	0.810	0.660	0.930	0.919	0.934	0.948	0.837
LPIPS \downarrow										
3DGS [7]	0.216	0.341	0.115	0.219	0.328	0.223	0.204	0.130	0.208	0.220
2DGS [5]	0.276	0.380	0.149	0.263	0.381	0.243	0.230	0.146	0.228	0.255
GS-Pull [16]	0.331	0.443	0.223	0.362	0.414	0.239	0.269	0.248	0.242	0.308
RaDe-GS [14]	0.176	0.286	0.103	0.190	0.279	0.218	0.205	0.130	0.204	0.199
GOF [13]	0.182	0.282	0.109	0.197	0.282	0.221	0.205	0.137	0.200	0.202
PGSR [2]	0.186	0.264	0.103	0.192	0.273	0.180	0.172	0.113	0.169	0.184
Ours	0.169	0.258	0.098	0.177	0.254	0.175	0.164	0.109	0.163	0.174

Fig. 1 shows qualitative comparisons of reconstructed meshes on the Deep Blending dataset. Our method recovers more accurate and complete surfaces, handling both dark and bright regions effectively. Competing methods often exhibit noise, oversmoothing, or missing geometry, especially near object boundaries. Results from GS-Pull are omitted due to its poor mesh quality and potential need for significant parameter tuning. For other baselines, we use the same default parameters as used in their evaluations on the Mip-NeRF 360 dataset.

2.2 More Experimental Results

Novel view synthesis. To provide a more comprehensive evaluation of novel view synthesis, we report per-scene quality metrics (PSNR, SSIM, and LPIPS) for Gaussian-based methods on the Mip-NeRF 360 dataset [1], as shown in Table 2. All results are obtained from our own runs using the official implementations of prior methods, except for 3DGS, where we use the pretrained models provided by the authors to avoid training inconsistencies. Our method consistently achieves the highest PSNR and SSIM scores on most scenes and has the best average scores. Notably, it yields

Table 3: Ablations on the TNT dataset.

	Precision \uparrow	Recall \uparrow	F1-score \uparrow
3×3	0.49	0.58	0.52
5×5	0.50	0.60	0.54
9×9	0.51	0.60	0.54
Random cube init	0.44	0.59	0.49
Random sphere init	0.47	0.59	0.51
$\tau = 0.001$	0.50	0.60	0.53
$\tau = 0.03$	0.50	0.59	0.53
Ours	0.51	0.60	0.54

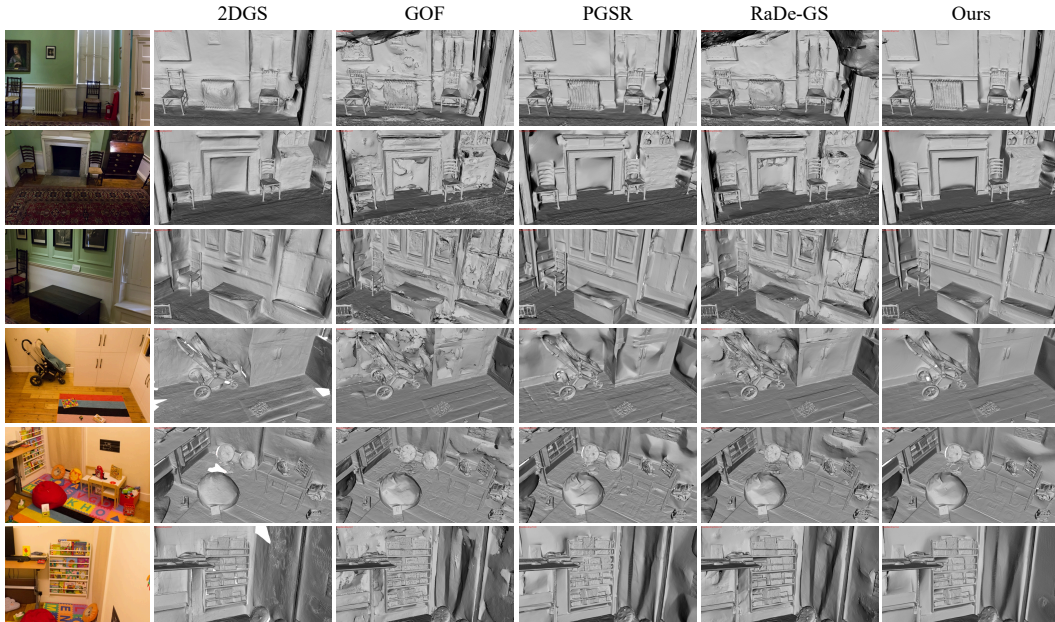


Figure 1: Visual comparison of surface reconstruction results on the Deep Blending dataset. Our method effectively handles the challenges posed by complex lighting conditions and ambiguous boundaries. GS-Pull is omitted as it fails to produce reasonable reconstructions.

significant improvements in LPIPS across all scenes, highlighting its ability to capture high-frequency details and produce perceptually superior renderings.

Qualitative comparisons are shown in Fig. 2, where we visualize novel view synthesis results on the Mip-NeRF 360 dataset. As highlighted in the boxed regions, our method produces sharper and more detailed renderings, particularly in complex areas such as grass and foliage, where competing methods often yield blurry outputs.

Surface reconstruction. Fig. 3 presents additional qualitative comparisons of reconstructed surfaces on real-world indoor and outdoor scenes from the TNT [8] and Mip-NeRF 360 [1] datasets. Compared to baseline methods, our approach produces more complete and continuous meshes with well-preserved high-frequency details. It also effectively avoids local minima, maintaining fine structures such as holes and sharp edges. By contrast, 2DGS [5] and GOF [13] often yield non-manifold meshes with broken topology, while GS-Pull [16], which extracts surfaces from a learned SDF, tends to produce overly smooth geometry.

Additional comparisons of surface reconstruction on the DTU dataset [6] are shown in Fig. 4. Given the relatively simple geometry of the target objects, most methods achieve visually complete reconstructions. However, the differences lie in the surface quality and detail. Our method demonstrates superior performance in handling reflective surfaces and preserves fine-scale features, producing reconstructions that are both smoother and more accurate, thereby improving visual fidelity and geometric realism.

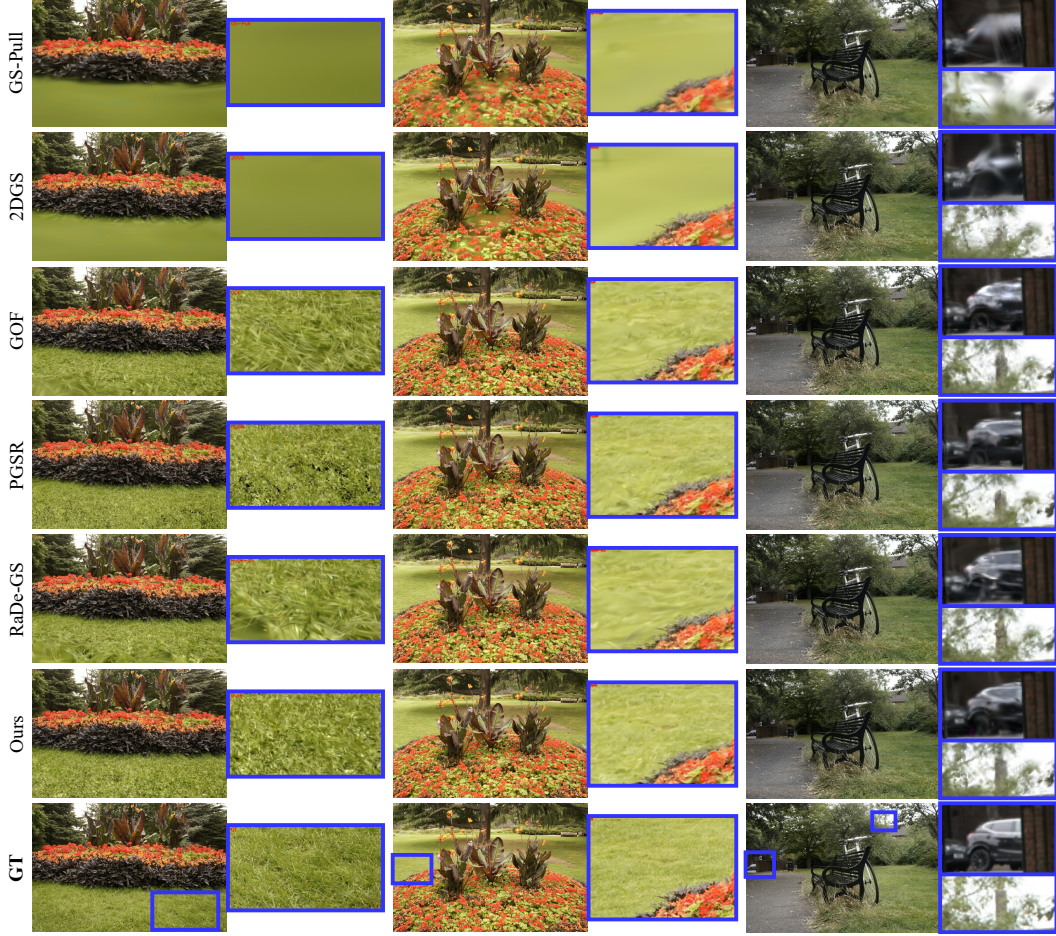


Figure 2: Comparison of our method with prior Gaussian-based approaches on novel view synthesis on the Mip-NeRF 360 dataset. Our method produces high-quality renderings with clear details, while previous methods generate blurry results in complex regions such as grass, vegetation, and cars, as highlighted in the framed areas.

3 More Ablation Results

To thoroughly evaluate the effectiveness and individual contributions of the components proposed in our method, we further conducted a series of ablation studies. These experiments systematically replace or remove specific modules and vary key hyperparameters to assess their impact on overall performance. The quantitative results are reported in Table 3.

(1) *Patch size in \mathcal{L}_p* . In our multi-view photometric alignment loss \mathcal{L}_p , we use a default patch size of 7×7 . We also evaluate alternative patch sizes, including 3×3 , 5×5 , and 9×9 . As shown in Table 3, smaller patches lead to suboptimal performance due to limited spatial context. Larger patches, such as 9×9 , offer no significant improvement while increasing computational cost. This confirms that our chosen patch size achieves a good balance between accuracy and efficiency.

(2) *Gaussian Initialization*. Similar to prior works, we use COLMAP [10] to generate a sparse point cloud from the input images, which serves as the initialization for our 3D Gaussians. To evaluate the impact of initialization quality, we experiment with randomly generated point clouds using two alternative strategies: (i) uniform sampling within a cube, and (ii) sampling points on a spherical surface. As shown in Table 3, both alternatives degrade the final reconstruction quality compared to COLMAP initialization. However, the spherical initialization performs better than the cube sampling, likely due to its more uniform coverage and approximate enclosure of the scene geometry.

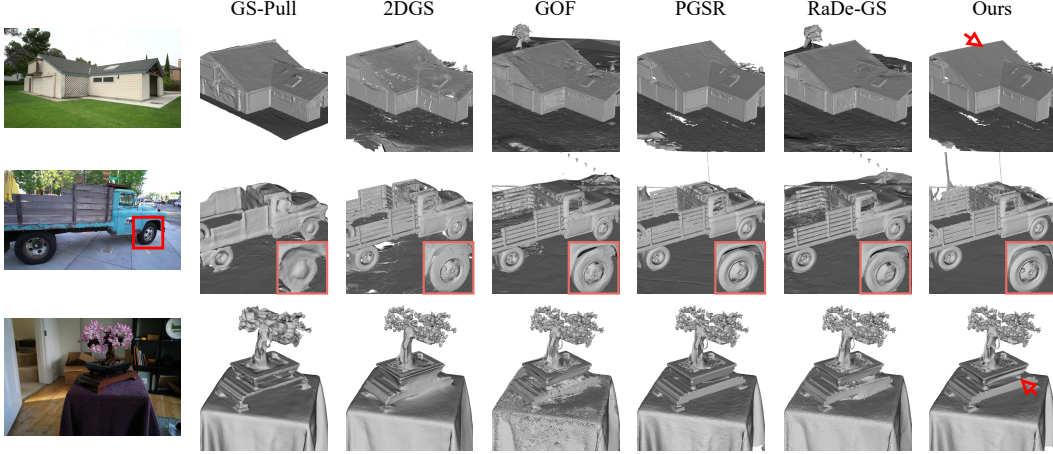


Figure 3: Visual comparison of surface reconstruction results on the TNT (first two rows) and Mip-NeRF 360 (third row) datasets. Our method produces more complete and accurate surfaces with clearer object boundaries and fewer artifacts, effectively handling challenges such as complex lighting conditions and ambiguous geometric structures.

Table 4: Ablation study on different loss terms.

	3DGS	Only \mathcal{L}_I	w/o \mathcal{L}_{nc}	w/o \mathcal{L}_{ns}	w/o \mathcal{L}_p	w/o \mathcal{L}_f	Full
F1-score \uparrow	0.09	0.13	0.52	0.51	0.50	0.53	0.54
Time (min) \downarrow	7.5	5.9	20.3	17.1	12.6	15.9	20.9
Time Gap	-	15.0	0.6	3.8	8.3	5.0	0.0

(3) *Threshold in \mathcal{L}_{ns} .* Our normal-based geometric alignment module incorporates a smoothing term \mathcal{L}_{ns} , with a threshold set to 0.01 by default. This threshold helps preserve sharp edges while reducing noise in low-curvature regions. We conduct ablations by testing alternative values, including 0.001 and 0.03. The results show that excessively small values (*e.g.*, 0.001) overly constrain normal variation, leading to surface oversmoothing, while large values (*e.g.*, 0.03) reduce the loss’s regularization effect. Our default setting (0.01) strikes a good balance and achieves the best performance.

The run-time cost of each loss. We provide the runtime of each loss ablation on the TNT dataset in Table 4. With all losses enabled, the training takes 20.9 minutes on RTX 4090. Removing \mathcal{L}_f (feature alignment) reduces training time to 15.9 minutes, yielding a ~ 5 -minute saving, while still preserving strong performance. By comparison, removing \mathcal{L}_p (photometric alignment) saves more time (8.3 min) but sacrifices more reconstruction quality. When only \mathcal{L}_I (image reconstruction) is used, our method has a lower time cost than the original 3DGS (5.9 min vs. 7.5 min) and provides better F1-score results than 3DGS (0.13 vs. 0.09).

4 More Discussion

More clarification of loss \mathcal{L}_f . We agree that \mathcal{L}_f brings modest improvements on the TNT dataset. \mathcal{L}_f improves general robustness and is beneficial in scenes with consistent lighting or texture. It improves robustness under low-texture or lighting-variant scenes, which is not fully reflected by the F1-score metric alone. Removing it often leads to over-smoothed meshes in challenging regions. Moreover, the contribution of the feature alignment loss \mathcal{L}_f is more significant on the DTU dataset: removing \mathcal{L}_f increases Chamfer distance from 0.49 to 0.52. We attribute this to dataset characteristics. DTU scenes benefit more from learned feature-level consistency due to cleaner lighting and smoother structure. TNT contains diverse indoor/outdoor scenes and severe lighting variation. The expressive power of off-the-shelf image features is limited, which may partially underutilize \mathcal{L}_f . We plan to explore stronger feature extractors in future work.

Clarification of performance gains and training time. (a) While our training time is longer than that of vanilla 3DGS, the added cost stems from the newly introduced geometry supervision and view-alignment mechanisms, which directly improve surface accuracy. These components are

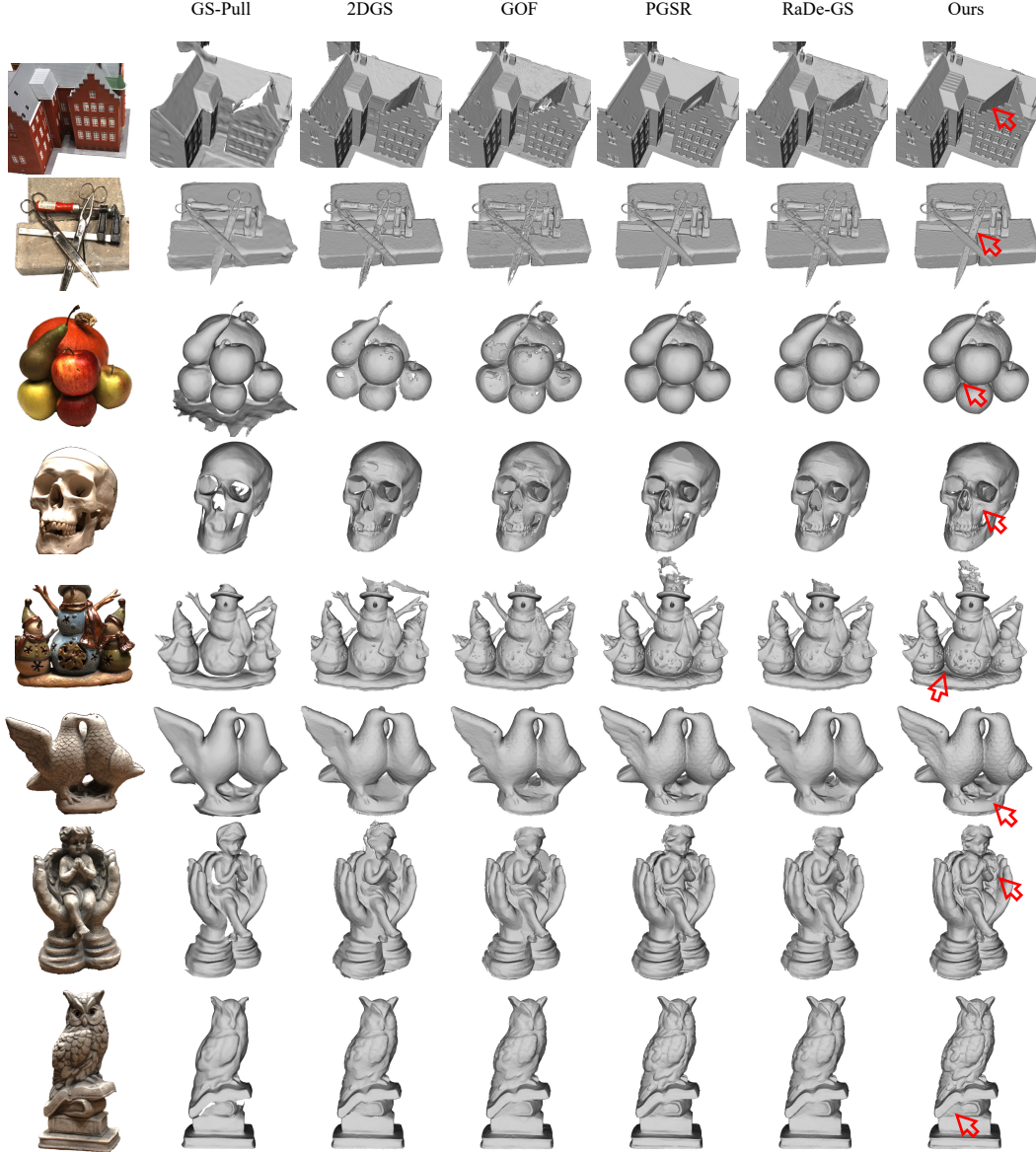


Figure 4: Visual comparison of surface reconstruction results on the DTU dataset. Our method demonstrates superior capability in handling reflective materials and recovering fine-grained surface geometry. Please zoom in on the digital version for detailed inspection.

essential for multi-view reconstruction, especially in challenging scenes. While the gains over the latest baselines may appear modest in some scenes, these differences represent noticeable geometric improvements. (b) Moreover, increased time is a common trade-off among geometry-enhanced 3DGS variants, where introducing geometric priors and view constraints improves reconstruction quality at the cost of runtime (see Tables 1 and 2 of main paper). The original 3DGS remains the fastest due to its lightweight constraints, but also exhibits the least accurate geometry. In future work, we plan to explore adaptive Gaussian pruning to further reduce training cost without sacrificing accuracy. (c) Our goal of this work is not to accelerate or compress 3DGS, but to enhance its geometric representation. On DTU, we reduce Chamfer distance from 1.96 to 0.49; on TNT, we raise F1-score from 0.09 to 0.54, a substantial improvement over 3DGS; on Mip-NeRF 360, we also observe consistent gains across all metrics. As shown in Table 4, when only using edge-aware image reconstruction loss \mathcal{L}_I on TNT, our method runs faster than 3DGS (5.9 min vs. 7.5 min) and achieves a higher F1-score (0.13 vs. 0.09). (d) In terms of practicality, our training time on TNT (~ 20 min) is production-viable for many offline applications. In contrast, neural implicit SDF methods often require 10+ hours of

training. We believe that accuracy is often prioritized over marginal runtime gains in such scenarios, with the rapid advancement of hardware.

More explanation of the ablations. In all ablation studies, we adopt a consistent and controlled protocol to ensure fair comparisons: (a) When ablating a loss term (*e.g.*, \mathcal{L}_p or \mathcal{L}_f), we drop the corresponding term entirely from the optimization objective. (b) The remaining loss weights are kept unchanged, and we do not re-normalize or re-scale other terms to preserve consistent training dynamics. (c) All experiments are conducted with the same number of training iterations and identical optimization settings (*e.g.*, learning rate, batch size, data split). (d) For the hyperparameter ablations, we similarly vary only the parameter under study (*e.g.*, source view count, threshold value), while keeping all other parameters and training settings fixed. This ensures that any performance change can be attributed directly to the presence or absence of the specific loss being tested.

The difference from PGSR. Firstly, we would like to clarify that our method is not a direct extension of PGSR with feature alignment added on top. Instead, our framework introduces a new combination of single-view alignment (including edge-aware image reconstruction and normal-based geometry supervision) and multi-view alignment (photometric and feature alignment), which are structurally and conceptually different from PGSR. As shown in the ablations of Table 4 of main paper, even without feature alignment, our method already achieves better performance than PGSR. This demonstrates that our core design, particularly single-view and multi-view photometric supervision, is effective. The additional feature alignment further improves the results, validating our design choices. Even when the improvement appears numerically small on the TNT dataset, reaching the same or slightly higher performance than PGSR still represents a state-of-the-art level. Moreover, on the DTU dataset, the contribution of the feature alignment loss \mathcal{L}_f is more significant. Removing \mathcal{L}_f increases Chamfer distance from 0.49 to 0.52, which is still better than PGSR’s reported 0.53. This shows that our method offers consistent geometric improvement across datasets. Finally, our method also outperforms PGSR on the Mip-NeRF 360 dataset for novel view synthesis, as shown in Table 3 of main paper. We achieve better scores on all metrics, indicating that our method is not merely comparable to PGSR but surpasses it in reconstruction/synthesis quality and generalizability across varied benchmarks.

Our Multi-View Photometric Alignment differs from PGSR’s in several key aspects: (a) Optimization complexity: PGSR couples its photometric consistency loss with geometric consistency regularization, which minimizes forward-backward reprojection error of neighboring views, resulting in more variables participating in network backpropagation and a more complex gradient computation. In contrast, our method simplifies the backpropagation path by aligning Gaussian orientations first and then applying image/feature-level constraints, which leads to better optimization efficiency. (b) Multi-view formulation: PGSR computes photometric consistency between one reference view and one source view. Our framework uses one reference view and multiple source views (three by default, varied in ablation), which introduces richer multi-view constraints and improves geometric supervision. (c) Gaussian flattening: PGSR flattens 3D Gaussians using scale regularization of the Gaussian ellipsoid to mimic planar surfaces. In our method, we found this strategy ineffective through ablation in Table 4 of main paper, and therefore designed our supervision differently. (d) Efficiency and performance: As shown in Table 2 of main paper, our method achieves better reconstruction quality across more scenes on real outdoor scenes of the TNT dataset while requiring less training time than PGSR. (e) Occlusion modeling: We explicitly define a visibility term and occlusion weight to mitigate the effect of outlier pixels caused by occlusion or misalignment. We give their motivation and clearly show the derivation of their formulas in the method section.

5 Limitations and Future Works

Despite achieving strong performance in both surface reconstruction and novel view synthesis, our method has several limitations that suggest promising directions for future work. First, our model assumes known camera poses as input, which may not always be available in real-world scenarios, especially when the number of input views is extremely sparse. Removing this requirement by exploring pose-free approaches [12, 15] is a compelling future direction that would increase the applicability of our method in unconstrained settings. Moreover, our method assumes a moderate number of input views. For extremely sparse-view settings, the multi-view consistency constraints may become less effective, reducing reconstruction quality. Second, the number of Gaussians grows significantly with the number of input views, potentially limiting the scalability of our method in

large-scale or densely captured scenes. Designing more compact or adaptive Gaussian representations, or incorporating efficient pruning and regularization strategies, could help improve scalability without compromising reconstruction quality. Third, our method currently relies on a pre-trained feature extraction network to guide geometric alignment and ensure robustness under varying lighting and viewpoints. This dependence introduces limitations, particularly when the pre-trained features are suboptimal for specific domains or scenes. Exploring more reliable, possibly self-supervised, feature learning strategies or reducing reliance on external networks altogether would further improve robustness and generalization. Addressing these limitations can pave the way for more flexible, efficient, and generalizable Gaussian-based surface reconstruction frameworks.

6 Potential Negative Social Impacts

While our method advances the state-of-the-art in surface reconstruction and novel view synthesis using 3D Gaussian Splatting, it is important to consider its potential negative social impacts. First, high-fidelity 3D reconstruction from multi-view imagery could be misused to reconstruct environments or individuals without consent. For instance, when applied to personal photos or public surveillance footage, our method may enable unauthorized digital replication of private spaces or identities. We strongly advocate for the ethical and consensual collection and use of input data. Second, enhanced geometry reconstruction and photorealistic novel view synthesis could be integrated into pipelines for generating synthetic scenes or manipulating real-world data. This could potentially contribute to the creation of deceptive media, raising concerns about misinformation, impersonation, or forgery. Preventative measures such as watermarking and provenance tracking should be considered in downstream applications. Third, although more efficient than dense volumetric methods, our approach still requires significant GPU resources for training and inference. As the field moves toward real-time and scalable 3D vision, energy consumption and environmental sustainability must remain part of the design considerations.

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] Danpeng Chen, Hai Li, Weicai Ye, Yifan Wang, Weijian Xie, Shangjin Zhai, Nan Wang, Haomin Liu, Hujun Bao, and Guofeng Zhang. PGSR: Planar-based gaussian splatting for efficient and high-fidelity surface reconstruction. *IEEE Transactions on Visualization and Computer Graphics*, 2024.
- [3] Antoine Guédon and Vincent Lepetit. 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*, pages 5354–5363, 2024.
- [4] Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow. Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics (ToG)*, 37(6):1–15, 2018.
- [5] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2D gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 conference papers*, pages 1–11, 2024.
- [6] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engin Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 406–413, 2014.
- [7] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, 2023.
- [8] 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):1–13, 2017.
- [9] Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. Neuralangelo: High-fidelity neural surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8456–8465, 2023.
- [10] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

- [11] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *Advances in Neural Information Processing Systems*, 34:27171–27183, 2021.
- [12] Botao Ye, Sifei Liu, Haofei Xu, Xueting Li, Marc Pollefeys, Ming-Hsuan Yang, and Songyou Peng. No pose, no problem: Surprisingly simple 3d gaussian splats from sparse unposed images. *arXiv preprint arXiv:2410.24207*, 2024.
- [13] Zehao Yu, Torsten Sattler, and Andreas Geiger. Gaussian opacity fields: Efficient adaptive surface reconstruction in unbounded scenes. *ACM Transactions on Graphics (TOG)*, 43(6):1–13, 2024.
- [14] Baowen Zhang, Chuan Fang, Rakesh Shrestha, Yixun Liang, Xiaoxiao Long, and Ping Tan. RaDe-GS: Rasterizing depth in gaussian splatting. *arXiv preprint arXiv:2406.01467*, 2024.
- [15] Shangzhan Zhang, Jianyuan Wang, Yinghao Xu, Nan Xue, Christian Rupprecht, Xiaowei Zhou, Yujun Shen, and Gordon Wetzstein. Flare: Feed-forward geometry, appearance and camera estimation from uncalibrated sparse views. *arXiv preprint arXiv:2502.12138*, 2025.
- [16] Wenyuan Zhang, Yu-Shen Liu, and Zhizhong Han. Neural signed distance function inference through splatting 3D gaussians pulled on zero-level set. *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.