

PlanarGS: High-Fidelity Indoor 3D Gaussian Splatting Guided by Vision-Language Planar Priors

Xirui Jin^{1*} Renbiao Jin^{1*} Boying Li²✉ Danping Zou¹✉ Wenxian Yu¹

¹Shanghai Jiao Tong University & ²Monash University

[NeurIPS2025]



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



MONASH
University

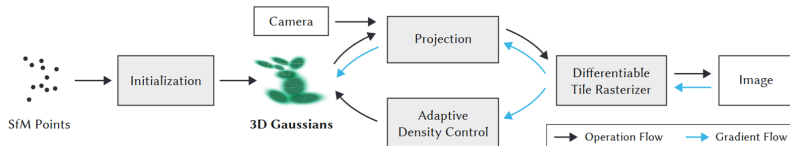
*Equal Contribution

✉boyong.li@monash.edu

✉dpzou@sjtu.edu.cn

Motivation

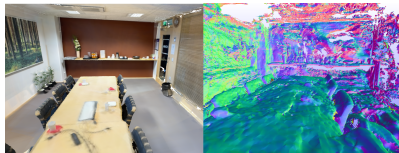
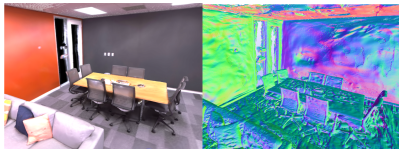
- 3D Gaussian Splatting (3DGS)



Impressive visual quality for novel-view synthesis

Motivation

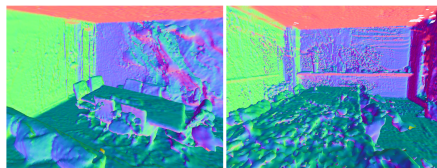
- **Challenge with 3DGS:** Photometric loss alone yields ambiguous geometry in texture-less areas



- **Existing solutions** still produce curved surfaces

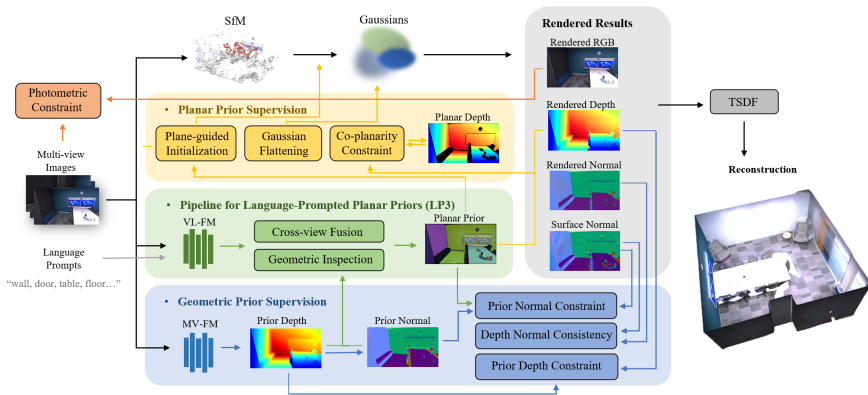


PGSR: multi-view consistency



DN-Splatter: depth & normal priors

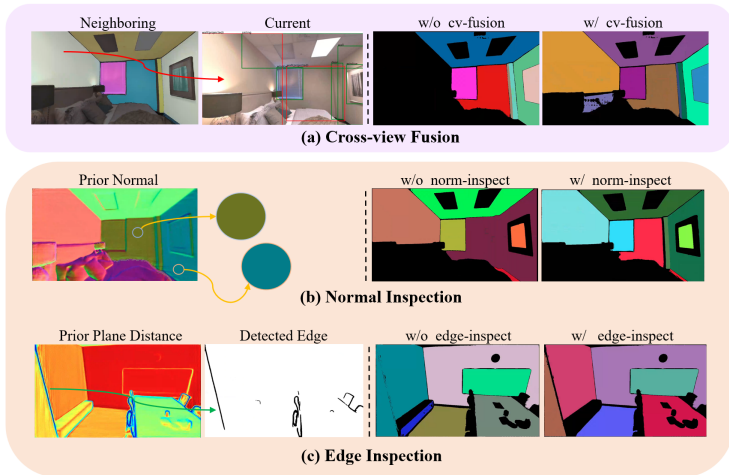
Method Overview



- We introduce planar priors for holistic supervision via the **LP3** pipeline.

Language-Prompted Planar Priors (LP3)

- **Flexible prompts** to the foundation model: (GroundedSAM)
"wall, floor, door, screen, window, ceiling, table"



- LP3 pipeline: (a) to supplement detection & (b)(c) to refine segmentation

Planar Prior Supervision

- **Plane-guided initialization:** Supplement SfM points in planar regions via depth priors.

$$P_s = D_r(p_s) \cdot K^{-1} \tilde{p}_s \quad p_s \in pr \quad (1)$$

- **Gaussian flattening:** Enable Gaussians to conform to the local surface.

$$L_s = ||\min(s_1, s_2, s_3)||_1 \quad (2)$$

- **Co-planarity constraint:** Enforce Gaussians in one planar region to stay on the same plane.

$$A_m = (Q_n^T Q_n + \epsilon E)^{-1} Q_n^T Y_m \quad (3)$$

$$D_p(p) = (A_m^T K^{-1} \tilde{p})^{-1} \quad (4)$$

$$L_p = \frac{1}{N_p} \sum_{p \in pr} \|D_p(p) - \hat{D}(p)\| \quad (5)$$

The pr represents the planar regions according to the planar priors.

Geometric Prior Supervision

- **Prior depth constraint:** Scale-corrected prior depth from multi-view foundation model(DUSt3R)

$$L_{rd} = \frac{1}{N_{lt}} \sum_{p \in lt} M_{cof}(p) \cdot \|D_r(p) - \hat{D}(p)\|^2 \quad (6)$$

- **Prior normal constraint:** Normal prior generated from the depth prior

$$L_{rn} = \|N_{dr} - \hat{N}_d\|_1 + (1 - N_{dr} \cdot \hat{N}_d) \quad (p \in pr) \quad (7)$$

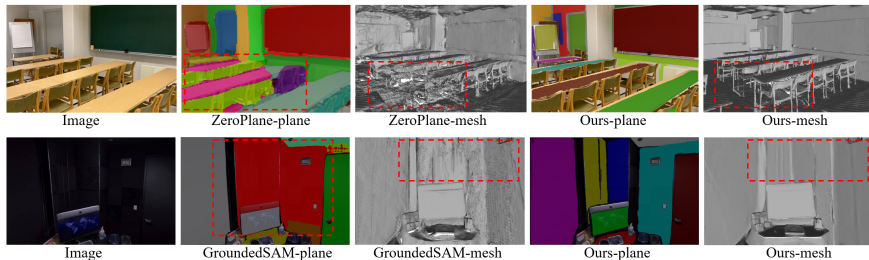
- **Depth-normal consistency:** Between the rendered normal map and the normal map generated from the rendered depth map

$$L_{dn} = \frac{1}{N_{lt}} \sum_{p \in lt} \cdot \|\hat{N}_d(p) - \hat{N}(p)\|_1 \quad (8)$$

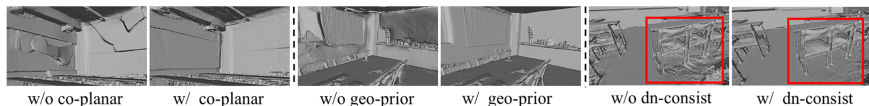
The lt represents the low-texture regions generated from the Canny edge detector on RGB.

Ablation Study

- **LP3 pipeline:** Inaccurate planar priors lead to worse reconstruction results.



- **Constraints:** Each contributes significantly to the final high-quality outcome.



Ablation Study

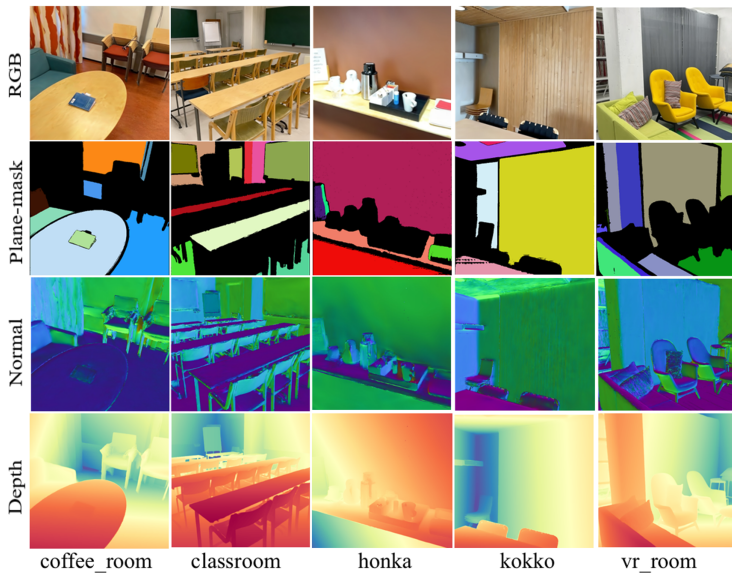
- Quantitative results of **module effectiveness**

Pre-proc. Planar Prior	Opt. Constraints			Metrics				
	Co-planar.	Geom. Prior	DN Consist.	Acc↓	Comp↓	CD↓	F1↑	NC↑
ZeroPlane	✓	✓	✓	4.76	5.73	5.25	62.49	82.41
GroundedSAM	✓	✓	✓	2.95	3.92	3.43	80.22	84.66
LP3 pipeline	-	✓	✓	2.83	3.53	3.18	85.62	84.63
LP3 pipeline	✓	-	✓	4.63	5.01	4.82	73.96	80.62
LP3 pipeline	-	-	✓	6.93	6.26	6.59	64.11	74.42
LP3 pipeline	✓	✓	-	2.57	3.17	2.87	88.19	81.95
LP3 pipeline	✓	✓	✓	2.55	3.40	2.98	87.32	85.28

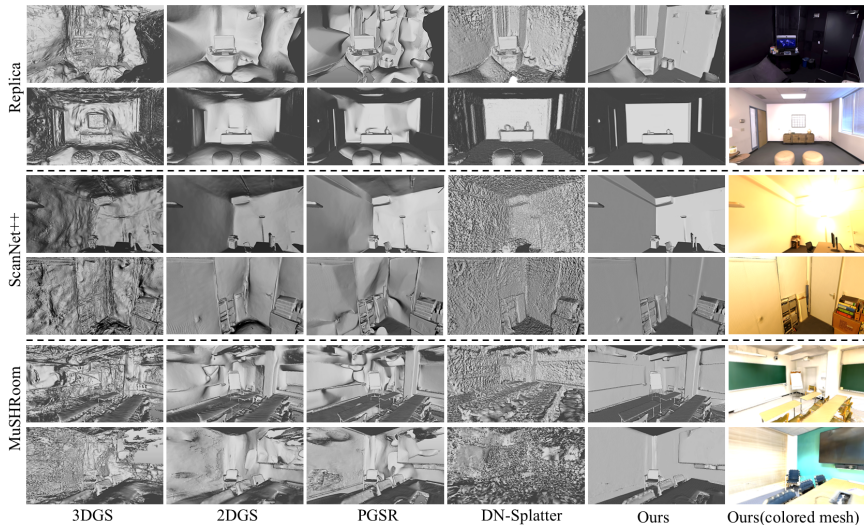
- Robustness** to the foundation models and prompts used

Pre-processing			Metrics				
MV-FM	VL-FM	Prompts	Acc↓	Comp↓	CD↓	F1↑	NC↑
VGGT	GroundedSAM	regular	3.96	6.32	5.14	75.41	88.85
DUST3R	YoloWorld+SAM	regular	2.92	5.56	4.24	80.96	89.04
DUST3R	GroundedSAM	more prompts	2.88	5.50	4.19	81.43	89.32
DUST3R	GroundedSAM	regular	2.80	5.45	4.13	81.90	89.88

Results



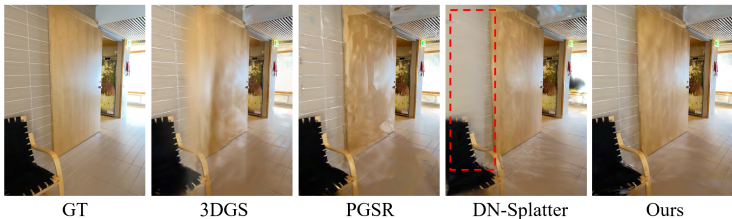
Results



Results

- Up to 36.8% relative improvements on the **MuSHRoom** dataset.
- Simultaneously ensure high-quality novel view synthesis.

	Surface Reconstruction					NVS		
	Acc↓	Comp↓	CD↓	F1↑	NC↑	PSNR↑	SSIM↑	LPIPS↓
3DGS	12.01	11.85	11.92	38.53	62.00	25.79	0.8775	0.2059
2DGS	9.16	10.27	9.71	51.50	73.65	25.67	0.8744	0.2265
GOF	15.65	8.50	12.07	42.57	61.47	24.76	0.8646	0.2189
PGSR	7.52	13.50	10.51	59.11	73.27	25.73	0.8761	0.2250
QGS	8.53	10.02	9.28	55.15	73.66	24.37	0.8624	0.1994
DN-Splatter	6.25	5.29	5.77	61.86	77.13	24.80	0.8549	0.2595
Ours	3.95	5.02	4.49	77.14	83.35	26.42	0.8874	0.2141



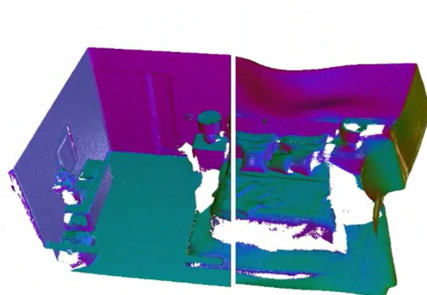
Result

- Compared with previous **SOTA** and the **feed-forward model**, PlanarGS achieves up to 43.4% and 62.8% relative improvements in surface reconstruction accuracy on the Replica dataset.

	ScanNet++					Replica				
	Acc↓	Comp↓	CD↓	F1↑	NC↑	Acc↓	Comp↓	CD↓	F1↑	NC↑
DUST3R	9.70	6.64	8.17	38.17	75.27	7.52	7.18	7.35	44.89	68.19
3DGS	11.02	10.12	10.57	31.78	77.65	12.40	12.75	12.57	41.64	69.83
2DGS	6.89	6.52	6.7	53.46	88.20	9.02	12.05	10.54	52.91	82.28
GOF	7.76	6.42	7.09	57.99	74.75	9.77	8.31	9.04	54.08	74.34
PGSR	7.32	7.13	7.22	53.73	87.30	7.32	9.79	8.56	62.98	83.30
DN-Splatter	4.88	3.44	4.16	75.86	82.06	4.95	6.25	5.60	68.12	80.80
Ours	3.86	3.46	3.66	82.78	90.52	2.80	5.45	4.13	81.90	89.88

Results

Project page: <https://planargs.github.io/> for more information.



PlanarGS(Ours)

2DGS

[video link](#)

References I

-  Chen, Danpeng et al. (2024). “Pgsr: Planar-based gaussian splatting for efficient and high-fidelity surface reconstruction”. In: *IEEE Transactions on Visualization and Computer Graphics*.
-  Cheng, Tianheng et al. (2024). “Yolo-world: Real-time open-vocabulary object detection”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16901–16911.
-  Huang, Binbin et al. (2024). “2d gaussian splatting for geometrically accurate radiance fields”. In: *ACM SIGGRAPH 2024 conference papers*, pp. 1–11.
-  Kerbl, Bernhard et al. (2023). “3d gaussian splatting for real-time radiance field rendering.”. In: *ACM Transactions on Graphics (TOG)* 42.4, pp. 139–1.
-  Liu, Jiachen et al. (2025). “Towards In-the-wild 3D Plane Reconstruction from a Single Image”. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 27027–27037.
-  Ren, Tianhe et al. (2024a). “Grounded SAM: Asembling Open-World Models for Diverse Visual Tasks”. In: *arXiv preprint arXiv:2401.14159*.

References II

-  Ren, Xuqian et al. (2024b). “MuSHRoom: Multi-Sensor Hybrid Room Dataset for Joint 3D Reconstruction and Novel View Synthesis”. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 4508–4517.
-  Straub, Julian et al. (2019). “The replica dataset: A digital replica of indoor spaces”. In: *arXiv preprint arXiv:1906.05797*.
-  Turkulainen, Matias et al. (2025). “Dn-splatter: Depth and normal priors for gaussian splatting and meshing”. In: *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE, pp. 2421–2431.
-  Wang, Jianyuan et al. (2025). “VGGT: Visual Geometry Grounded Transformer”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
-  Wang, Shuzhe et al. (2024). “Dust3r: Geometric 3d vision made easy”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20697–20709.
-  Yeshwanth, Chandan et al. (2023). “Scannet++: A high-fidelity dataset of 3d indoor scenes”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12–22.

References III

-  Yu, Zehao, Torsten Sattler, and Andreas Geiger (2024). “Gaussian opacity fields: Efficient adaptive surface reconstruction in unbounded scenes”. In: *ACM Transactions on Graphics (ToG)* 43.6, pp. 1–13.
-  Zhang, Ziyu et al. (2024). “Quadratic Gaussian Splatting for Efficient and Detailed Surface Reconstruction”. In: *arXiv preprint arXiv:2411.16392*.

Thank you!
Q&A