PanoSplatt3R: Leveraging Perspective Pretraining for Generalized Unposed Wide-Baseline Panorama Reconstruction

Supplementary Material

1. Experimental Details

Training Details.

Our model is trained on random pairs of panoramic images sampled from rendered HM3D video, with frame intervals evenly sampled between 75 and 100. The frame used for supervision is randomly selected between the two frames.

We employ the AdamW optimizer [4] for training, with an initial learning rate of 2×10^{-5} in the first stage. In the second stage, the initial learning rate for the backbone is set to 2×10^{-5}, while the learning rate for the remaining Gaussian parameter head is set to 2×10^{-4}. A warm-up strategy is applied for both stages over 2k steps. The final learning rate will decay to $1/10$ of the original value.

Dataset Details.

Since the training and test sets of the HM3D dataset [5] used by Splatter-360 [3] are not publicly available, we follow their dataset generation process using AI-Habitat [6] to construct our own HM3D training and test sets. Specifically, we render videos along random camera trajectories and generate panoramic images by stitching six cube maps for each viewpoint. For other datasets, we directly use the available off-the-shelf data.

2. Additional Quantitative Comparisons

We conducted additional experiments to assess and compare the models' extrapolation capabilities. Following the testing procedure outlined in the main text, we fixed the input frame interval at 100 and randomly selected test frames from a 50-frame range beyond the two input frames. The results, presented in Table 1, show that despite all methods being trained with supervision on frames between the inputs, our model consistently outperforms others across all metrics—except for SSIM on the HM3D dataset. This strong performance across most metrics indicates that our model learns more robust spatial representations, enabling more accurate extrapolation beyond the training distribution.

3. Visualization Results

We provide visual comparisons of synthesized panoramic images on the HM3D [5] and Replica [7] datasets, showcasing the performance of different methods. As shown in Figure 1 and Figure 2, PanoSplatt3R produces the most visually consistent and realistic results, with sharper details, fewer artifacts, and improved structural coherence compared to existing methods.

Table 1. **Quantitative comparison in view extrapolation.** Methods are evaluated on the Replica and HM3D datasets.

Dataset	Metric	MVSplat	Splatter-360	PanoSplatt3R
Replica [7]	PSNR↑	27.188	26.975	**29.371**
	SSIM↑	0.895	0.904	**0.914**
	LPIPS↓	0.135	0.123	**0.107**
	Abs Rel↓	0.130	0.095	**0.059**
	RMSE↓	0.313	0.277	**0.176**
	$\delta < 1.25$↑	85.007	90.648	**95.414**
HM3D [5]	PSNR↑	25.728	24.986	**26.306**
	SSIM↑	0.827	**0.831**	0.822
	LPIPS↓	0.205	0.193	**0.191**
	Abs Rel↓	0.129	0.136	**0.098**
	RMSE↓	0.339	0.334	**0.229**
	$\delta < 1.25$↑	85.390	86.258	**92.526**

References

[1] Yuedong Chen, Haofei Xu, Chuanxia Zheng, Bohan Zhuang, Marc Pollefeys, Andreas Geiger, Tat-Jen Cham, and Jianfei Cai. Mvsplat: Efficient 3d gaussian splatting from sparse multi-view images. In *Eur. Conf. Comput. Vis.*, pages 370–386. Springer, 2024. 2, 3

[2] Zheng Chen, Yan-Pei Cao, Yuan-Chen Guo, Chen Wang, Ying Shan, and Song-Hai Zhang. Panogrf: generalizable spherical radiance fields for wide-baseline panoramas. In *Adv. Neural Inform. Process. Syst.*, pages 6961–6985, 2023. 2, 3

[3] Zheng Chen, Chenming Wu, Zhelun Shen, Chen Zhao, Weicai Ye, Haocheng Feng, Errui Ding, and Song-Hai Zhang. Splatter-360: Generalizable 360 gaussian splatting for wide-baseline panoramic images. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 21590–21599, 2025. 1, 2, 3

[4] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 1

[5] Santhosh K Ramakrishnan, Aaron Gokaslan, Erik Wijmans, Oleksandr Maksymets, Alex Clegg, John Turner, Eric Undersander, Wojciech Galuba, Andrew Westbury, Angel X Chang, et al. Habitat-matterport 3d dataset (hm3d): 1000 large-scale 3d environments for embodied ai. *arXiv preprint arXiv:2109.08238*, 2021. 1

[6] Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, et al. Habitat: A platform for embodied ai research. In *Int. Conf. Comput. Vis.*, pages 9339–9347, 2019. 1

[7] Julian Straub, Thomas Whelan, Lingni Ma, Yufan Chen, Erik Wijmans, Simon Green, Jakob J Engel, Raul Mur-Artal, Carl Ren, Shobhit Verma, et al. The replica dataset: A digital replica of indoor spaces. *arXiv preprint arXiv:1906.05797*, 2019. 1

Figure 1. Visual comparisons between PanoGRF, MVSplat, Splatter-360 and PanoSplatt3R(ours) on the HM3D dataset.

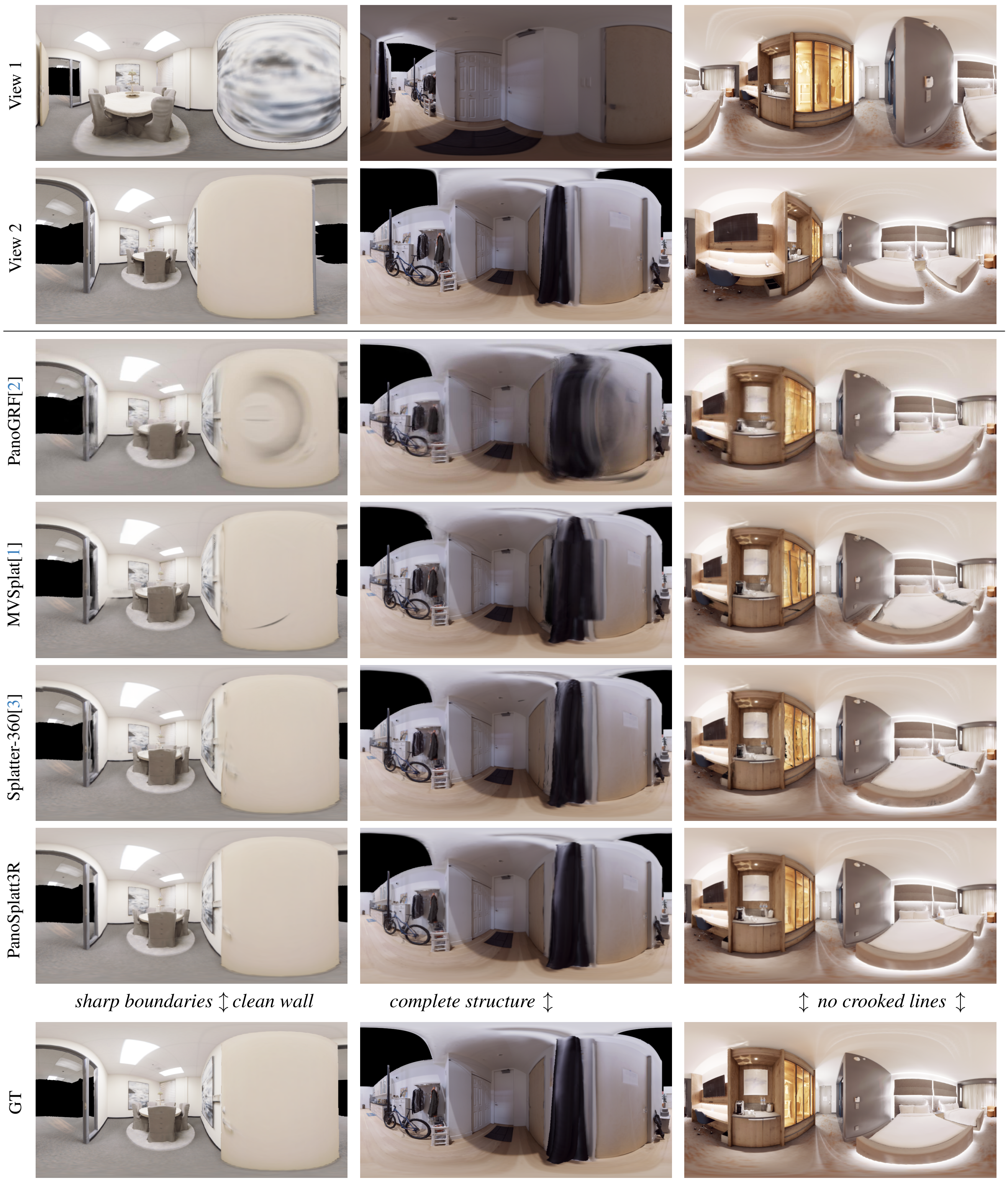

Figure 2. Visual comparisons between PanoGRF, MVSplat, Splatter-360 and PanoSplatt3R(ours) on the Replica dataset.