Long-LRM++: Preserving Fine Details in Feed-Forward Wide-Coverage Reconstruction

Supplementary Material

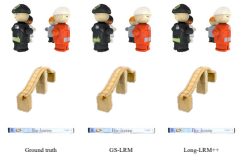

Figure 6. Qualitative comparison with GS-LRM for object reconstruction on the GSO dataset.

Figure 7. Qualitative comparison with Long-LRM on the Tanks&Temples dataset.

7. More implementation details

Due to its semi-explicit formulation, Long-LRM++ exhibits a stronger tendency to overfit to input frames when training on mixed sets of input and unseen target frames. This effect becomes more pronounced on datasets such as DL3DV, where neighboring frames have relatively large pose differences—that is, the effective frame density per unit scene coverage is lower. To mitigate this overfitting, we manually reduce the probability that input frames are selected as target frames during training. Concretely, during the random sampling (without replacement) of target frame indices, we decrease the selection weight of input frames to 0.1 while keeping all other frames at 1. For denser datasets such as ScanNetv2, this adjustment is unnecessary because the number of unseen frames significantly exceeds the number of input frames.

Method	PSNR↑
GS-LRM$_{\text{dim768}}$	31.39
GS-LRM	31.95
Long-LRM++	32.52

Table 7. Object reconstruction on GSO.

Method	PSNR↑
GS-LRM	28.10
Long-LRM	28.54
Long-LRM++	29.31

Table 8. Scene reconstruction on RE10K.

Method	PSNR↑
3D GS	18.10
Long-LRM	18.38
Long-LRM++	19.30

Table 9. Scene reconstruction on T&T.

8. More evaluation results

We evaluate Long-LRM++ on three additional datasets: object reconstruction on GSO (Table 7), scene reconstruction on RealEstate10K (Table 8), and zero-shot scene reconstruction on Tanks&Temples (Table 9). For GSO, we train Long-LRM++ on Objaverse for 80K steps and evaluate on GSO, comparing against GS-LRM and GS-LRM$_{\text{dim768}}$, which matches Long-LRM++'s backbone dimension. As shown in Fig. 6, Long-LRM++ achieves superior detail fidelity. For RealEstate10K, we train on the training split for 100K steps and evaluate on the test split, achieving state-of-the-art performance. For Tanks&Temples, we conduct zero-shot evaluation using a model trained on DL3DV, obtaining a +1 dB PSNR improvement over Long-LRM (see qualitative results in Fig. 7).

9. Comparison with Depth Anything 3 (DA3)

We compare Long-LRM++ (110M param) with the Gaussian prediction feature of DA3-GIANT [18] (1.15B param) on DL3DV using both COLMAP poses and DA3 poses. To obtain DA3 poses, we run the pose predictor of DA3 on all frames of a scene. During inference, we feed the obtained poses of selected input frames to the Gaussian prediction models. Quantitative comparison is shown in Table 10 and qualitative in Fig. 8. Both Long-LRM and Long-LRM++ show better rendering quality than DA3-GIANT.

Pose Source	Method	PSNR↑	SSIM↑	LPIPS↓
COLMAP	DA3-GIANT	17.52	0.562	0.382
	Long-LRM	24.10	0.783	0.254
	Long-LRM++	26.43	0.846	0.180
DA3	DA3-GIANT	17.27	0.540	0.395
	Long-LRM	22.98	0.731	0.277
	Long-LRM++	24.43	0.773	0.212

Table 10. Quantitative comparison with DA3 on DL3DV-140 (32 input views, 960×540 resolution).

10. Training stage configuration

Table 11 summarizes the detailed setup for each stage of training of Long-LRM++ on DL3DV, including training iterations, number of GPUs, and total GPU hours. Table 12 summarizes the stage setup for the ScanNetv2 training.

Stage	#Input	#Target	Resolution	Time/Step	#Step	Batch size	#GPU	GPU Hours
1	8	8	256×256	9.6sec	60K	256	16	2560
2	8	8	512×512	7.7sec	10K	64	16	342
3	8	8	960×540	17.6sec	10K	64	16	782
4	32	8	960×540	27.4sec	10K	64	64	4871

Table 11. Training stage configuration of Long-LRM++ for the DL3DV10K novel-view synthesis task. **GPU Hours** is calculated as **Time/Step** × **#Steps** × **#GPU**.

Ground truth DA3-GIANT Long-LRM Long-LRM++

Figure 8. Qualitative comparison with DA3 on DL3DV (32-input, 960×540-resolution) using DA3 poses.

Stage	#Input	#Target	Resolution	Time/Step	#Step	Batch size	#GPU	GPU Hours
1	8	8	256×256	6.7sec	20K	128	8	298
2	8	8	448×336	3.9sec	5K	128	32	173
3	128	8	448×336	30.6sec	2K	64	64	1088

Table 12. Training stage configuration of Long-LRM++ for the ScanNetv2 novel-view color+depth rendering task.