A Limitations and Discussions

Our *LaRa* demonstrates a remarkable efficiency feed-forward model that achieved high-fidelity all-around novel-view synthesis and surface reconstruction from sparse large-baseline images. However, our approach struggles to recover high-frequency geometry and texture details, mainly due to the low volume resolution. Enhancing our approach with techniques such as gradient checkpointing or mixed-precision training can potentially increase training batch size as well as volume resolution. We have also noticed that our method can yield inconsistent rendering results when the geometry is incorrectly estimated or when reconstructing multi-view inconsistent inputs, as demonstrated in the comparison video. This occurs because our method utilizes second-order Spherical Harmonic appearance modeling. Although such modeling can capture view-dependent effects, it also introduces a stronger ambiguity between geometry and appearance. We believe that incorporating our method with a physically-based rendering process can potentially address this issue. In addition, our work assumes posed inputs, but estimating precise camera poses for sparse views is a challenge in practice. Incorporating a pose estimation module [66] into the feed-forward setting is an orthogonal direction to our work.

B Acknowledgements

We thank Bozidar Antic for pointing out a bug that resulted in an improvement of about 1dB. Special thanks to BinBin Huang and Zehao Yu for their helpful discussion and suggestions. We would like to thank Bi Sai, Jiahao Li, Zexiang Xu for providing us with the testing examples of Instant3D, and Jiaxiang Tang for helping us to construct a comparison with LGM. This project was supported by the ERC Starting Grant LEGO-3D (850533) and the DFG EXC number 2064/1 - project number 390727645.

C More Visual Results

Finally, we report more qualitative results on the testing sets, as shown in Figure 8,Figure 9, and Figure 10.

Fig. 8: Reconstruction results on Gobjaverse testing set.

Fig. 9: Reconstruction results on Google Scanned Object dataset.

Fig. 10: Reconstruction results on Instant3D scenes.