Learning with Unreliability: Fast Few-shot Voxel Radiance Fields with Relative Geometric Consistency –*Supplementary Materials*–

Yingjie Xu[1,2*] Bangzhen Liu[2*] Hao Tang[1,3] Bailin Deng[4] Shengfeng He[1†]
[1]Singapore Management University [2]South China University of Technology
[3]Nanjing University of Science and Technology [4]Cardiff University

This supplementary material includes visualizations of more experimental results, further analysis on the selection of different warping angles, and extra implementation details. Besides, we also record a video demo, which gives an explanation of the overall framework and the visualizations.

1. Comparisons on More Scenes

1.1. Realistic Synthetic 360°

We provide more scene reconstruction results of our proposed ReVoRF together with recent advanced methods [1, 2, 5] on the Realistic Synthetic 360° dataset [4]. The visualization of the chair, ficus, and materials are shown in Fig. 1. With fewer artifacts and finer texture details, ReVoRF exhibits a superior capacity for reconstructing both geometry and appearance details in these scenes than the compared methods. We also conduct an additional evaluation for separating L_{rs} and L_{ds} in Realistic Synthetic 360°. The results are shown in Table 1. Note that the L_{rs} used here does not consider the unreliability. Our method outperforms these two variants.

	PSNR↑	SSIM↑	LPIPS↓
VGOS+L_{ds}	17.33	0.779	0.241
VGOS+L_{rs}	18.27	0.814	0.210
Ours	**20.72**	**0.848**	**0.179**

Table 1. Ablation on Realistic Synthetic 360° about L_{rs} and L_{ds}

1.2. LLFF

We provide the visualizations of rendering images and corresponding depth map of our ReVoRF and the state-of-the-art voxel-based few-shot nerf method VGOS on the LLFF dataset [3]. As shown in Fig. 2a and Fig. 2c, our method achieves better rendering quality, in terms of the clearer boundary of objects and less blurring. Compared to VGOS, our method also preserves better geometric consistency, which has finer reconstructed depth (Fig. 2b vs. Fig. 2d).

2. Ablations on the Choices of Warping Degree

We investigate the impact of warping angles γ on rendering quality by reconstructing a Lego scene from the Realistic Synthetic 360° dataset. For each selected γ, we randomly vary the values of pitch angle θ and azimuth angle ϕ in a range of $[\gamma - 5, \gamma]$. As illustrated in Fig. 3, when γ is either too large or too small, the rendering quality deteriorates. On the one hand, a small γ may not be able to provide sufficient multi-view information due to the slight variance of views. On the other hand, increasing the degree of γ could encounter more erroneously warped regions at the very beginning of training. Therefore, a moderate deformation angle yields the optimal rendering result.

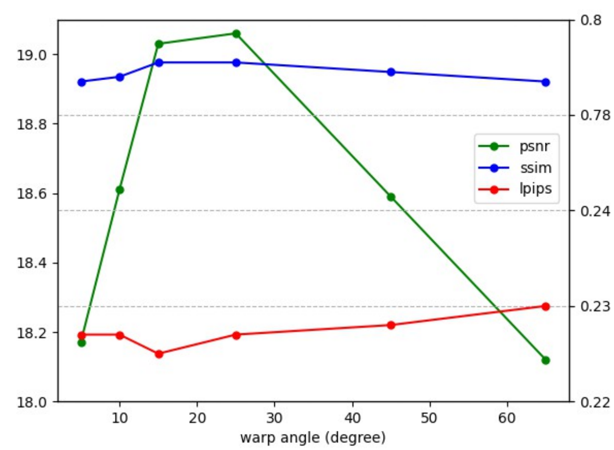


Figure 3. The ablation of warping angles γ on the Lego scene of Realistic Synthetic 360°. The horizontal axis in the graph represents the warping angles γ.

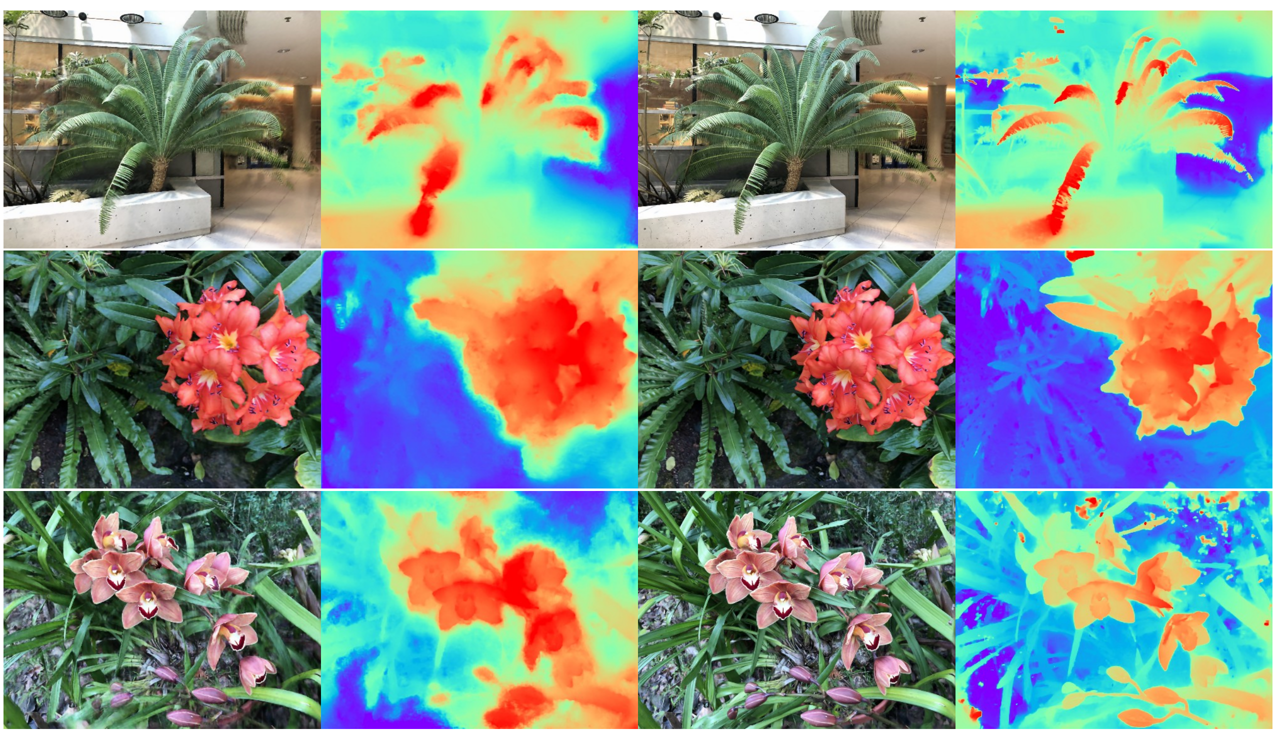
---

*The first two authors contributed equally.

†Corresponding author (*shengfenghe@smu.edu.sg*).

Diet-NeRF [1] InfoNeRF [2] VGOS [5] ReVoRF (Ours) GT

Figure 1. Comparisons on chair, ficus, and materials (from the top to the bottom) of the Realistic Synthetic 360° dataset [4] in 4-views setting. ReVoRF enables more consistent reconstruction with detailed appearance. Please zoom in for more details.



(a) VGOS RGB [1] (b) VGOS depth [2] (c) ReVoRF (Ours) RGB (d) ReVoRF (Ours) depth

Figure 2. Comparisons on LLFF dataset [3] in 3-views setting. ReVoRF exhibits greater clarity in depth and improved geometric shapes.

3. Additional Implementation Details in LLFF

In the LLFF dataset [3], we only use a fine optimization scheme to stabilize the training of ReVoRF and gradually improve the geometric details. During the whole training period, we set the values of λ_{rel} and λ_{unr} as 10^{-1} and 10^{-3}, respectively. The values of λ_d, λ_f, and λ_{ds} are set as $5{\cdot}10^{-5}$, $5{\cdot}10^{-6}$, and $5{\cdot}10^{-4}$ in the fine stage.

References

[1] Ajay Jain, Matthew Tancik, and Pieter Abbeel. Putting nerf on a diet: Semantically consistent few-shot view synthesis. In *ICCV*, pages 5885–5894, 2021. 1, 2

[2] Mijeong Kim, Seonguk Seo, and Bohyung Han. Infonerf: Ray entropy minimization for few-shot neural volume rendering. In *CVPR*, pages 12912–12921, 2022. 1, 2

[3] Ben Mildenhall, Pratul P. Srinivasan, Rodrigo Ortiz Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: practical view synthesis with prescriptive sampling guidelines. *ACM TOG*, 38(4):29:1–29:14, 2019. 1, 2, 3

[4] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, volume 12346, pages 405–421, 2020. 1, 2

[5] Jiakai Sun, Zhanjie Zhang, Jiafu Chen, Guangyuan Li, Boyan Ji, Lei Zhao, and Wei Xing. Vgos: Voxel grid optimization for view synthesis from sparse inputs. In *IJCAI*, pages 1414–1422, 8 2023. 1, 2