SiNeRF: Sinusoidal Neural Radiance Fields for Joint Pose Estimation and Scene Reconstruction*

Yitong Xia[1], **Hao Tang**[1], **Radu Timofte**[1,2], **Luc Van Gool**[1]

[1]Computer Vision Laboratory, D-ITET, ETH Zürich, Switzerland [2]Computer Vision Laboratory, CAIDAS, University of Würzburg, Germany

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision Lab

1 Overview

SiNeRF consists of two key modules: **SIREN-MLP** and **Mixed Region Sampling(MRS)**.

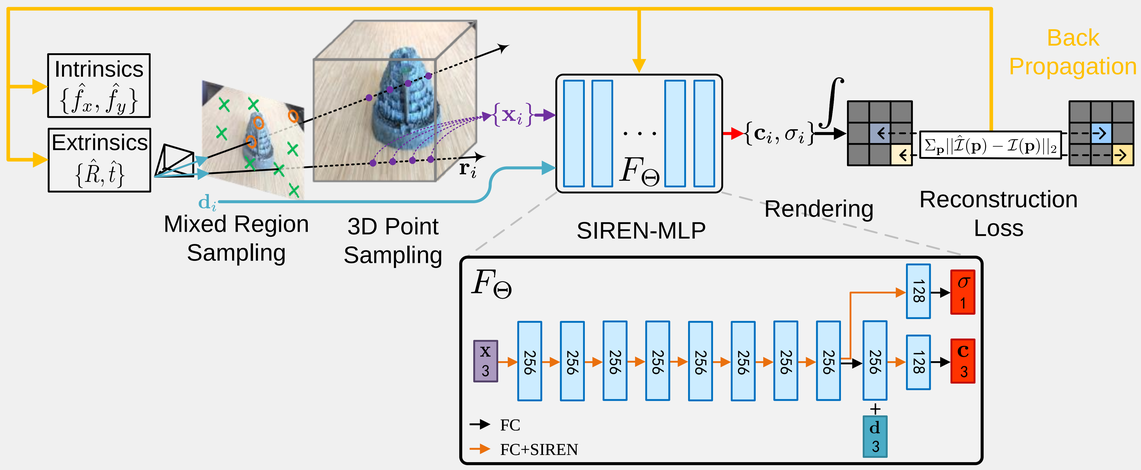

Figure 1: General overview of SiNeRF.

2 Brief introduction on paper

- **What can SiNeRF do?**
  SiNeRF is designed for improving the Joint Optimization (JO) [1], i.e. jointly learning camera poses and reconstructing scenes simultaneously.
- **Motivation?**
  JO is sensitive to scene contents and naively adopting a large network for radiance mapping does not always work. Thus, we believe there may exist a systematic sub-optimality in JO with multiple sources. We need non-trivial approaches to alleviate.
- **What we found?**
  - ★ **Radiance mapping network is not good enough.** Either increasing network size or adopting sinusoidal activations only improve performances on particular scenes.
  - ★ **Random ray sampling method is not good enough.** Random Sampling harms the performances on scenes with similar patterns. The pixels should not be treated equally.
  - ★ **Diversity matters for ray sampling.** Too-concentrated ray batch will lead to homogeneous rendering. Diversity is what we should learn from Random Sampling.

In our work, we discovered and verified the sources for the sub-optimality of JO and then improved them with SiNeRF.

⋆ Funded by: ETH General Fund, The Alexander von Humboldt Foundation, and Huawei.

3 Quantitative Results

Mean Pose Error					
Translation($\times 10^{-2}$) ↓			Rotation(°) ↓		
NeRFmm128	*NeRFmm256*	**SiNeRF**	*NeRFmm128*	*NeRFmm256*	**SiNeRF**
1.807	1.994	**1.519**	3.964	4.070	**2.578**

Table 1: Mean pose error comparisons on LLFF dataset. Please refer to the paper for full results.

Mean Image Quality								
PSNR ↑			SSIM ↑			LPIPS ↓		
NeRFmm128	*NeRFmm256*	**SiNeRF**	*NeRFmm128*	*NeRFmm256*	**SiNeRF**	*NeRFmm128*	*NeRFmm256*	**SiNeRF**
22.485	22.773	**23.554**	0.635	0.651	**0.689**	0.460	0.442	**0.407**

Table 2: Mean image quality comparisons on LLFF dataset. Please refer to the paper for full results.

4 Qualitative Results

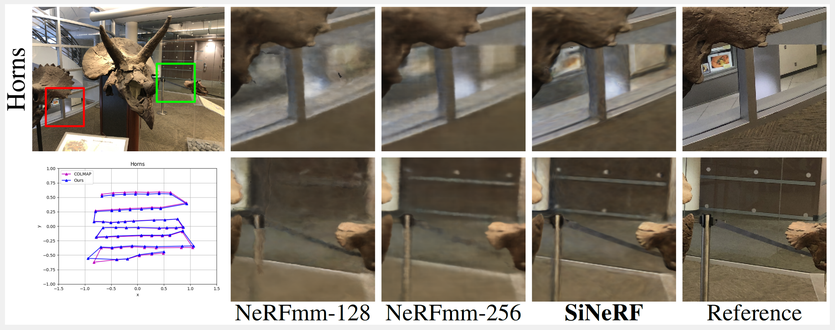

Figure 2: Part of qualitative results on LLFF dataset. Please refer to the paper for full results.

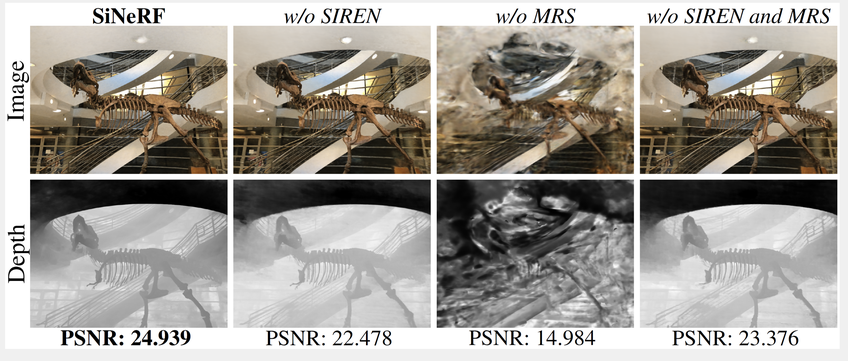

Figure 3: Qualitative results of ablation study.

References

[1] Zirui Wang, Shangzhe Wu, Weidi Xie, Min Chen, and Victor Adrian Prisacariu. Nerf–: Neural radiance fields without known camera parameters. *arXiv preprint arXiv:2102.07064*, 2021.