Supplementary Material for "SinNeRF: Training Neural Radiance Fields on Complex Scenes from a Single Image"

Dejia Xu[1*], Yifan Jiang[1*], Peihao Wang[1], Zhiwen Fan[1]
Humphrey Shi[2,3,4], Zhangyang Wang[1]

[1]The University of Texas at Austin, [2]UIUC,
[3]University of Oregon, [4]Picsart AI Research

1 Introduction

In the Supplementary Material, we will first illustrate the network architectures, then the implementation details of several baseline methods. Furthermore, we show more visual comparisons of our method against previous approaches.

2 Network Architecture

NeRF Architecture We adopt the same MLP architecture as the original NeRF [4]. We use 8 full-connected layers to obtain density σ with channel equals to 256. The last layer uses a shifted softplus $\log(1 + \exp(x - 1))$ as the activation while the rest layers use ReLU activation functions. The additional branch for RGB color prediction contains a full-connected layer using a "widened" sigmoid activation $(1 + 2\epsilon)/(1 + \exp(-x)) - \epsilon$ and 128 channels. Similar to Mip-NeRF [1], the ϵ is set to 0.001. Such design avoids failure modes when MLP emits negative values everywhere and the vanishing gradient issue of sigmoid.

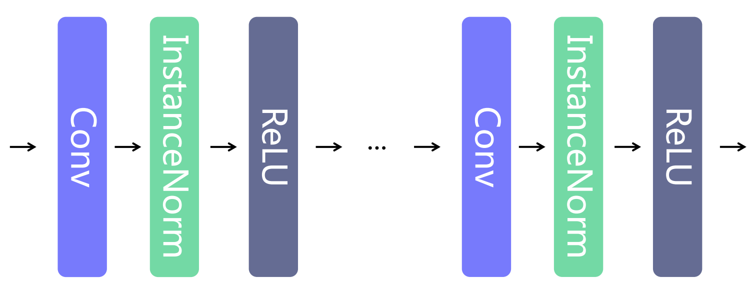

Fig. 1: Architecture of our discriminator.

Discriminator Architecture We adopt a sequential architecture for our discriminator. As shown in 1, we utilize several groups of layers, with each group consisting of a convolution, an instance norm and a ReLU layer. Note that the convolution layers are implemented with spectral norm. This enforces instance normalization on the features as well as spectral normalization on the convolution weights. In all our experiments, we use 10 groups for the proposed discriminator.

3 Implementation Details of Compared Methods

For comparison methods, we use the authors' original open-source implementation and pre-trained models (if available) for fair comparison. Since DS-NeRF [2] only released pretrained models on LLFF dataset, we trained new models from scratch on NeRF synthetic dataset. We also train DietNeRF [3] from scratch on each scene separately. For PixelNeRF [5], we utilize their pre-trained checkpoints on DTU dataset. Since their method is not a test-time optimization method, we fine-tune their pre-trained checkpoints on each new scene for fair comparison.

4 More Comparison Results

We provide more comparisons on the NeRF synthetic dataset, the Local Light Field Fusion(LLFF) dataset and the DTU dataset. **Note that** for each scene, we evaluate different methods using the same camera pose sequence. Each row corresponds to the same camera pose. Objects not being at center indicates that the geometry of the radiance field is mistaken.

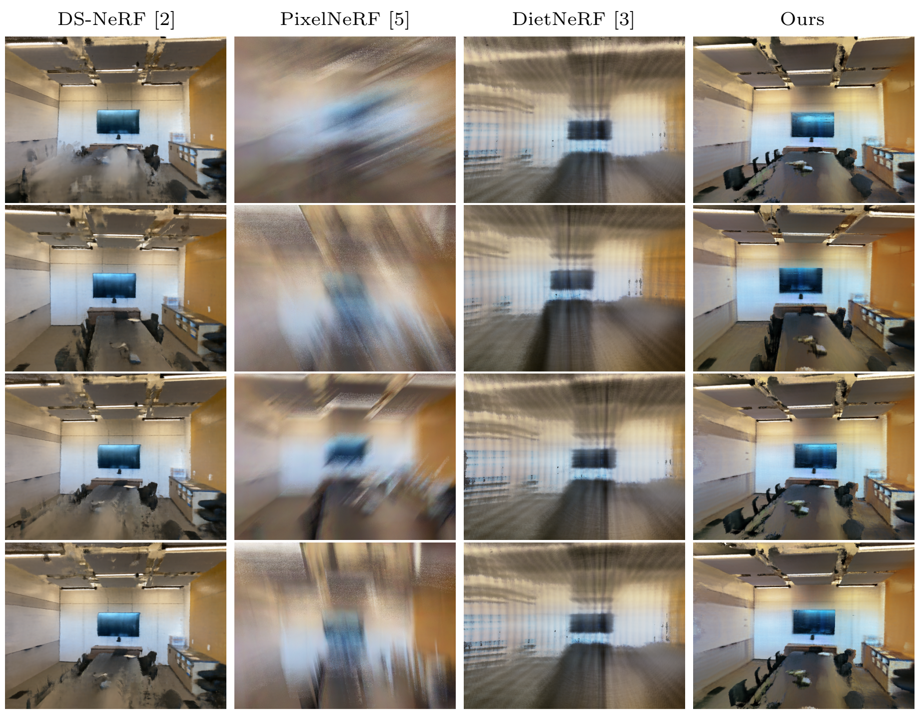

Fig. 2: The novel view synthesis performance of different methods on room scene of LLFF Dataset.

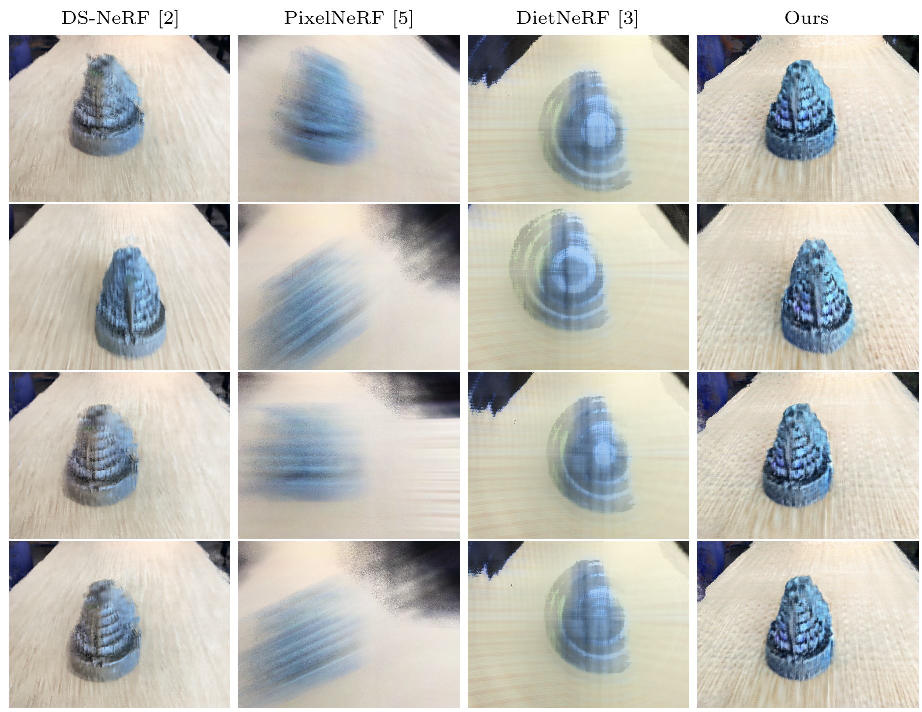

Fig. 3: The novel view synthesis performance of different methods on fortress scene of LLFF Dataset.

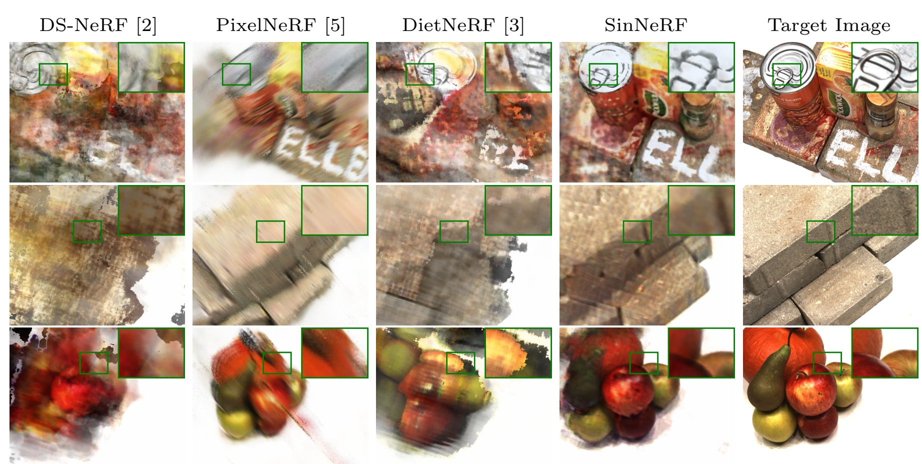

Fig. 4: Novel view synthesis results of different methods on DTU dataset.

References

1. Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P.: Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5855–5864 (2021)
2. Deng, K., Liu, A., Zhu, J.Y., Ramanan, D.: Depth-supervised nerf: Fewer views and faster training for free. arXiv preprint arXiv:2107.02791 (2021)
3. Jain, A., Tancik, M., Abbeel, P.: Putting nerf on a diet: Semantically consistent few-shot view synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5885–5894 (2021)
4. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: European conference on computer vision. pp. 405–421. Springer (2020)
5. Yu, A., Ye, V., Tancik, M., Kanazawa, A.: pixelnerf: Neural radiance fields from one or few images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4578–4587 (2021)