NeRF-SR: High Quality Neural Radiance Fields using Supersampling: Supplementary Materials

Besides this document, we also encourage the readers to watch the video to see multi-view renderings.

A EXPERIMENTAL DETAILS

Refinement Network For the refinement network, during training, we first apply the same random perspective transformation to both the reference image $\mathcal{I}_{\text{REF}}$ and synthesized SR image $\mathcal{I}_{\text{SR}}$ at the same pose to generate additional 199 pairs to form the set $\{(\mathcal{I}'_{\text{REF}}, \mathcal{I}'_{\text{SR}})_i\}_{i=1}^{200}$ for data augmentation. We use the *RandomPerspective* transformation in Pytorch to augment a patch during training, which transforms an image with perspective transformation and simulates the depth warping in the testing phase. Perspective transformation may not work well if the depths of the pixels in a patch have a large distinction, which rarely happens since the patch sizes are small. In each iteration, we first randomly select one pair $(\mathcal{I}'_{\text{REF}}, \mathcal{I}'_{\text{SR}})$ and crop the input patch and ground truth patch both at position p. The reference patches are randomly sampled from the square centered at p with side length 128 at $\mathcal{I}_{\text{REF}}$. For upscale of ×2 and ×4, the refinement network is trained for 3 and 8 epochs, each contains 500,000 patches.

B ADDITIONAL RESULTS

Figure 1: An example of refinement comparison for magnification of 2 for flower. Although the PSNR is decreased from 29.23 **dB to** 29.01 **dB after refinement, see how it still improves visual quality.**

Average Kernel We present additional comparison between "average" kernel on blender dataset in Figure 2. It is clear that symmetric "average" kernel produces better results than asymmetric kernel.

Figure 2: Visual results when both the downscale method and supervision signal are "average" (Input resolution is 100×100 **and upscaled by 4). Compared to asymmetric operation, symmetric averaged downscale and upscale produce more detailed super-resolution novel view synthesis.**

Additional Renderings Figure 4 and Figure 5 shows additional static results on blender and LLFF dataset.

Refinement We mentioned in the main text that the improvement of PSNR is not significant for magnification of 2 on LLFF dataset. We think it might because that supersampling already learns a

Figure 3: Mip-NeRF can synthesis reasonable novel views on trained resolution (left) but failed to generalize to higher resolutions (right).

reasonable geometry from multi-view images and the refinement only improves more subtle details. See Figure 1 for an example.

Mip-NeRF We have experimented on Mip-NeRF [1], a variant of NeRF that replaces the positional encoding (PE) in NeRF with integrated positional encoding (IPE). However, our experiments shows that IPE fails to generalize to resolutions higher that that of input images and only produce poor results (See Figure 3). Explanations from the author can be found at Mip-NeRF Github Repo.

C ADVERSARIAL TRAINING

The goal of the refinement process is to learn high-frequency details of reference images. Therefore, we have also considered using adversarial training [2] which has been employed in image-to-image translation [3], image super-resolution [4] for its ability to align one distribution to another. We treat NeRF as a generator and add an additional patch discriminator that takes as input patches from generator and reference images, expecting NeRF to generate images indistinguishable from HR references. However, the size of input patches are restricted by memory limitations and GAN struggles to provide meaningful guidance while introducing unignorable artifacts. The attempt to train a separate discriminator and freeze the weights when training with NeRF is also not successful.

REFERENCES

[1] Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. 2021. Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*. IEEE, 5835–5844. https://doi.org/10.1109/ICCV48922.2021.00580

[2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).

[3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1125–1134.

[4] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. 2019. Singan: Learning a generative model from a single natural image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4570–4580.

Figure 4: Additional Renderings on blender dataset for magnification of 4.

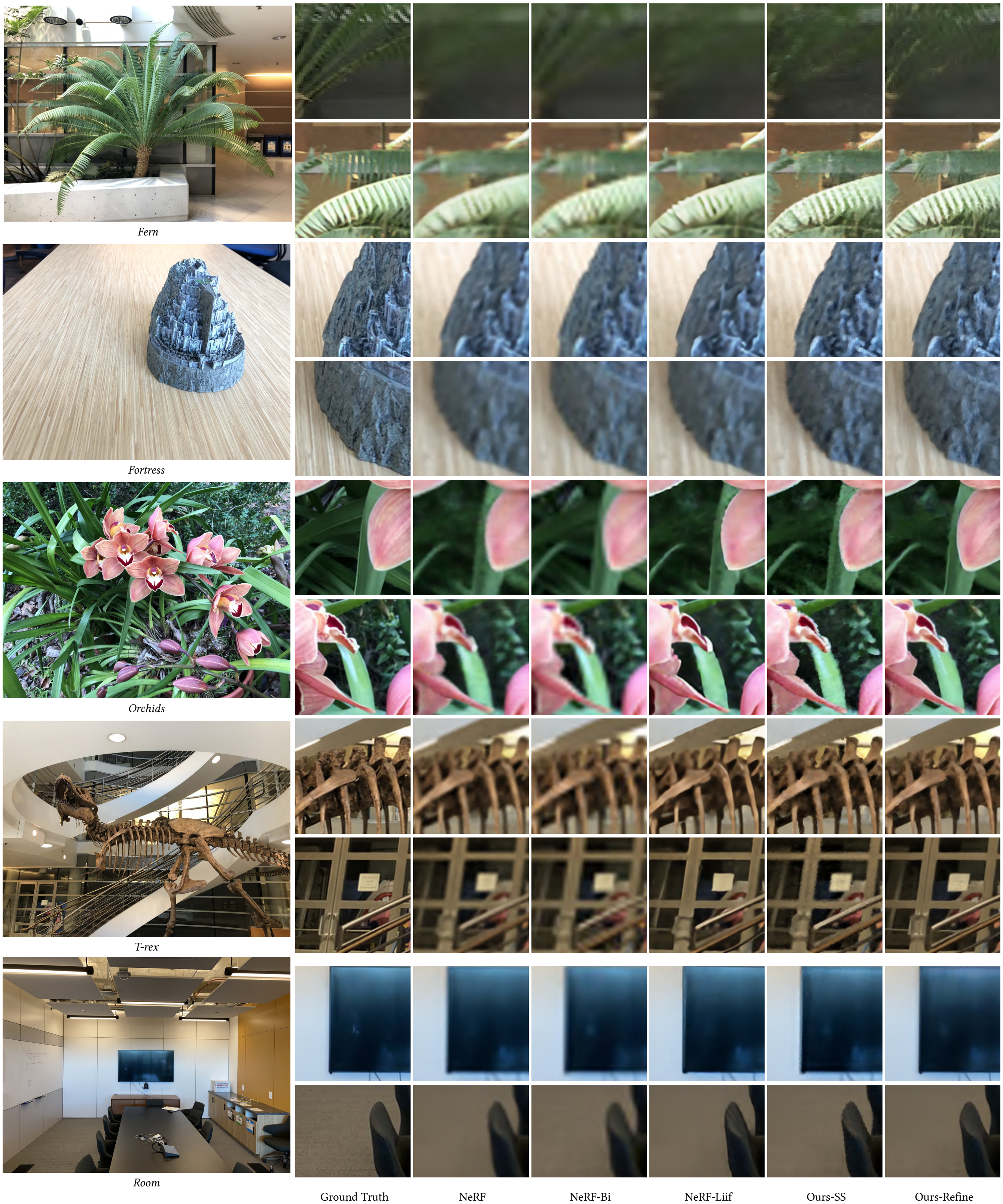

Figure 5: Additional Renderings on LLFF dataset for magnification of 4.