NeLF-Pro: Neural Light Field Probes for Multi-Scale Novel View Synthesis

Supplementary Material

In this **supplementary document**, we present visualizations of probe distribution, provide detailed analyses of the ablations along with observations, discuss the limitations and future work, and offer more quantitative and qualitative results.

A. Light Field Probes Distribution.

We propose distributing light field feature probes near the camera trajectory, utilizing the Farthest Point Sampling (FPS) algorithm to selectively determine $\hat{L}$ positions. The distribution of these feature probes within the scene and their corresponding content is depicted in Fig. 1. To enhance the visualization of the probe distribution within target scenes, we additionally present the calibration point cloud. These calibration point clouds are not used in training and are only for visualization purposes.

B. Ablation Details.

We conduct ablations on key components, including the point query scheme, probe distributions, period factor, and factor aggregation module. These ablations are performed on a small indoor scene and a larger scene encompassing distant and close viewpoints.

Our observations are as follows:

- Core factor modeling is critical. Disabling either factor $\bar{\mathbf{V}}$ or $\bar{\mathbf{M}}$ hampers reconstruction capabilities. Factor $\bar{\mathbf{M}}$ significantly influences overall quality, while factor $\bar{\mathbf{V}}$ improves fine-grained detail reconstruction qualitatively.
- Selecting factors close to the target camera yields substantial improvements, particularly for the objects with multi-scale appearances, where the camera neighborhood strategy effectively simulates a mipmap representation.
- Smaller scenes show less sensitivity to probe distribution than larger scenes. Farthest point position sampling (FPS) provides a more uniformly distributed set of probes, facilitating stable reconstruction across different areas.
- Coordinate scaling notably improves performance, especially in scenes captured from a small angle range. For instance, in the Free dataset, where most images are captured from a single viewing direction (as depicted in the middle of Fig. 1). The scale factor enables a more efficient use of the model capabilities.
- Factor aggregation delivers significant performance boosts than naive feature concatenation, benefiting from the order-invariant weighted fusion design.

In Fig. 2, we qualitatively illustrate the performance differences resulting from various design choices. Our full model provides high-quality reconstruction and better stability for diverse natural scenes that vary in extent and spatial granularity.

C. Limitations and Future Work.

Our *NeLF-Pro* demonstrates a remarkable ability to achieve high-fidelity novel view synthesis across a wide range of natural scenes, with various levels of spatial granularity. However, our approach currently cannot handle distractions, such as exposure changes and moving objects, and our method tends to produce 'floaters' in these regions. Combining our method with techniques like feature space regularization [1], appearance embedding [4] or pixel-wise loss reweighting [7] could be beneficial. In addition, our work cannot support fast rendering, it still takes about 2.6 seconds to render a 960×540 image. Incorporating fast rendering techniques, such as post-baking [6] is an orthogonal direction to our work.

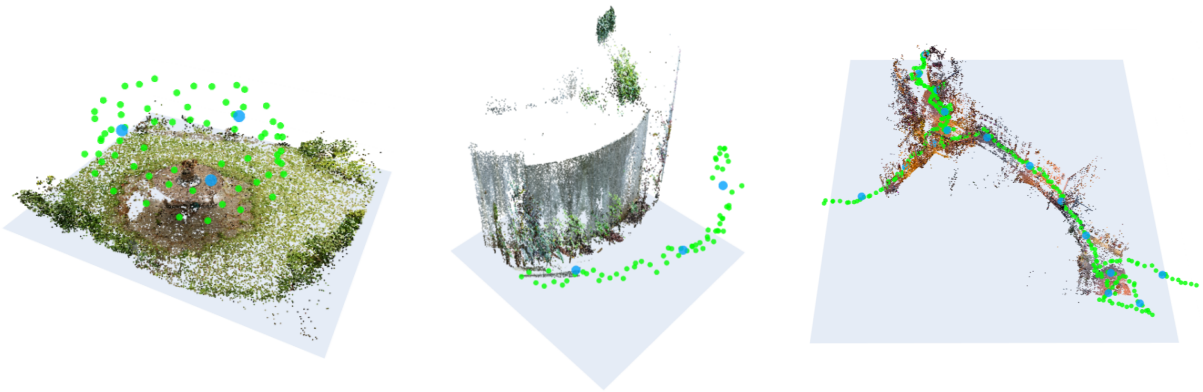
Future Work. To further enhance the per-scene reconstruction quality, a possible solution is to integrate our representation with advanced multi-level modeling [5] or cone tracing point sampler [2, 3]. We leave this combination as future work. Furthermore, while this paper demonstrates success in per-scene optimization, another interesting direction for future work involves the exploration and learning of general core and basis factors across various scenes. By analyzing a large-scale dataset, we can leverage data priors not only to improve quality but also to potentially enable new applications, such as in the development of generative models. In the ablation study, a notable finding is that matrix factors alone, without the vector core factor, can yield satisfactory reconstructions in small-scale scenes. This aspect is particularly advantageous for convolutional neural networks.

D. More Visual Results.

We show more qualitative results in Fig. 3, Fig. 4, and Fig. 5. Our *NeLF-Pro* is able to better preserve detail, sharpness, and thin structures more effectively than the baselines.

E. Per-scene Breakdown.

In Table 1 and Table 2, we provide breakdowns of the quantity metrics for the Free dataset [9] and the mip-NeRF360 dataset [2]. Our *NeLF-Pro* achieves consistently better rendering quality compared to previous grid-based approaches.



(a) **Garden** (b) **Grass** (c) **Scuol**

Figure 1. **The distribution of light field feature probes.** The green spheres denote basis factors and the blue dots are core factors. We densely distribute the basis factors in the scene and sparsely distribute the core factors, only using 3 for the small scenes, and 16 for the Scuol scene shown on the right side.

Figure 2. **Qualitative results for ablation study.** We conduct the ablations using one indoor-small-scene and outdoor-large-scene.

Figure 3. **More Qualitative Results on the Free Dataset [9].**

Method	Hydrant	Lab	Pillar	Road	Sky	Stair	Grass	Avg
NeRF++ [10]	22.21	21.82	25.73	23.29	23.91	26.08	21.26	23.47
mip-NeRF360 [2]	**25.03**	**26.57**	**29.22**	**27.07**	**26.99**	**29.79**	**24.39**	**27.01**
Plenoxels [1]	19.82	18.12	18.74	21.31	18.22	21.41	16.28	19.13
DVGO [8]	22.10	23.78	26.22	23.53	24.26	26.65	20.75	23.90
Instant-NGP [5]	22.30	23.21	25.88	24.24	25.80	27.79	21.82	24.43
F^2-NeRF [9]	24.34	25.92	28.76	26.76	26.41	29.19	22.87	26.32
NeLF-Pro (ours)	**24.92**	**26.39**	**29.56**	**27.65**	**27.06**	**29.55**	**24.00**	**27.02**
Our SSIM	0.770	0.834	0.818	0.834	0.873	0.853	0.629	**0.802**
Our LPIPS	0.260	0.251	0.233	0.231	0.217	0.203	0.398	**0.256**

Table 1. **Per-Scene breakdown on the Free dataset.** The baseline method scores are sourced from F^2-NeRF; however, F^2-NeRF does not furnish detailed per-scene breakdowns for SSIM and LPIPS metrics.

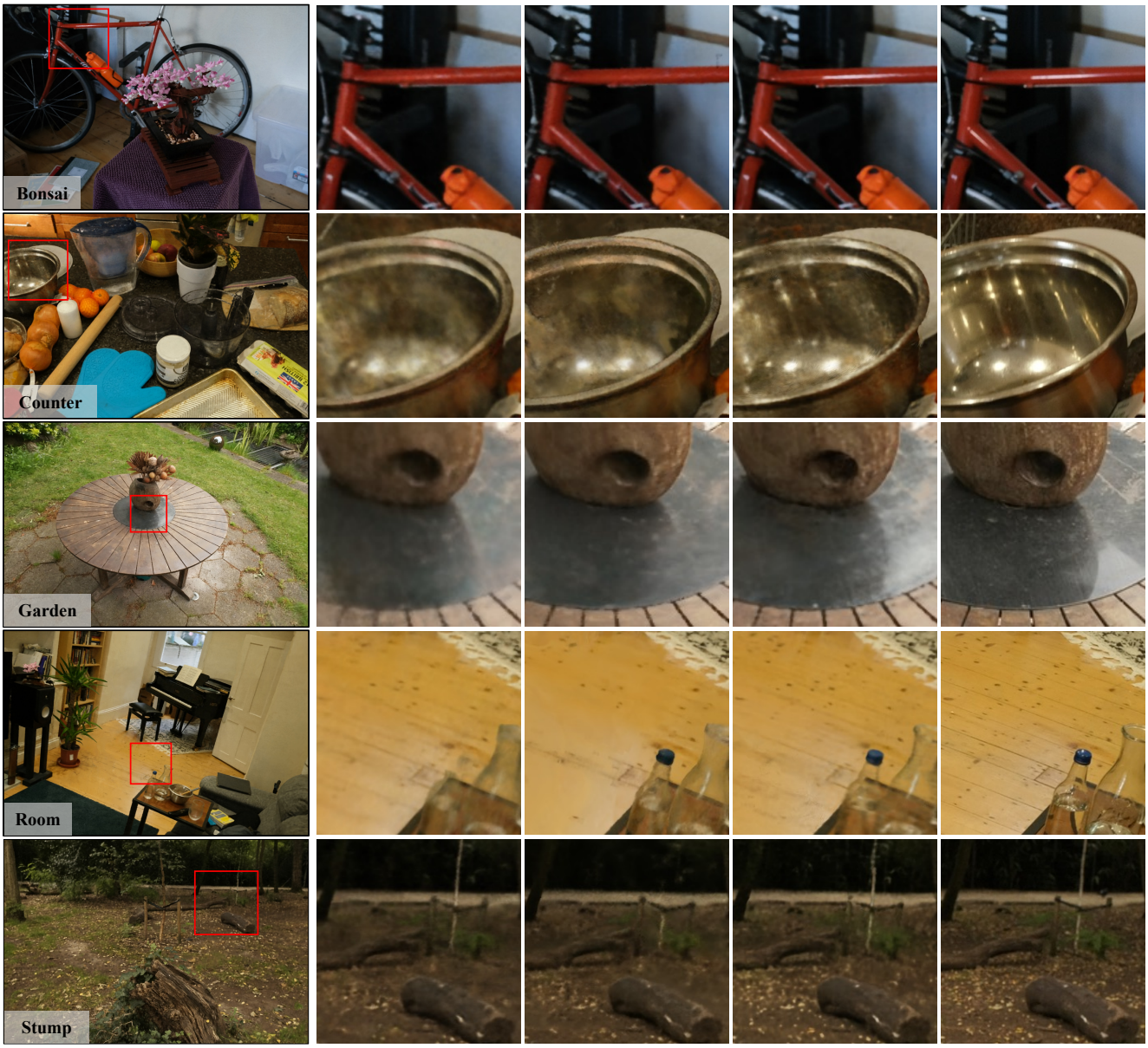


Figure 4. **More novel view synthesis results on mip-NeRF360 Dataset [2].**

Method	Bicycle	Bonsai	Counter	Garden	Kitchen	Room	Stump	Avg
NeRF++ [10]	22.64	29.15	26.38	24.32	27.80	28.87	24.34	26.21
mip-NeRF360 [2]	**23.99**	**33.06**	**29.51**	**26.10**	**32.13**	**31.53**	**26.27**	**28.94**
Plenoxels [1]	21.39	23.65	25.23	22.71	24.00	26.38	20.08	23.35
DVGO [8]	22.12	27.80	25.76	24.34	26.00	28.33	23.59	25.42
Instant-NGP [5]	22.08	29.86	26.37	24.26	28.27	28.90	23.93	26.24
F^2-NeRF [9]	22.11	29.65	25.36	24.76	28.97	29.30	**24.60**	26.39
NeLF-Pro (ours)	**22.42**	**31.28**	**27.52**	**25.08**	**29.79**	**30.10**	24.58	**27.27**
Our SSIM	0.496	0.907	0.823	0.691	0.868	0.871	0.615	**0.753**
Our LPIPS	0.500	0.289	0.366	0.331	0.240	0.367	0.427	**0.360**

Table 2. **Per-Scene breakdown on the mip-NeRF360 dataset.** The baseline method scores are sourced from F^2-NeRF; however, F^2-NeRF does not furnish detailed per-scene breakdowns for SSIM and LPIPS metrics.

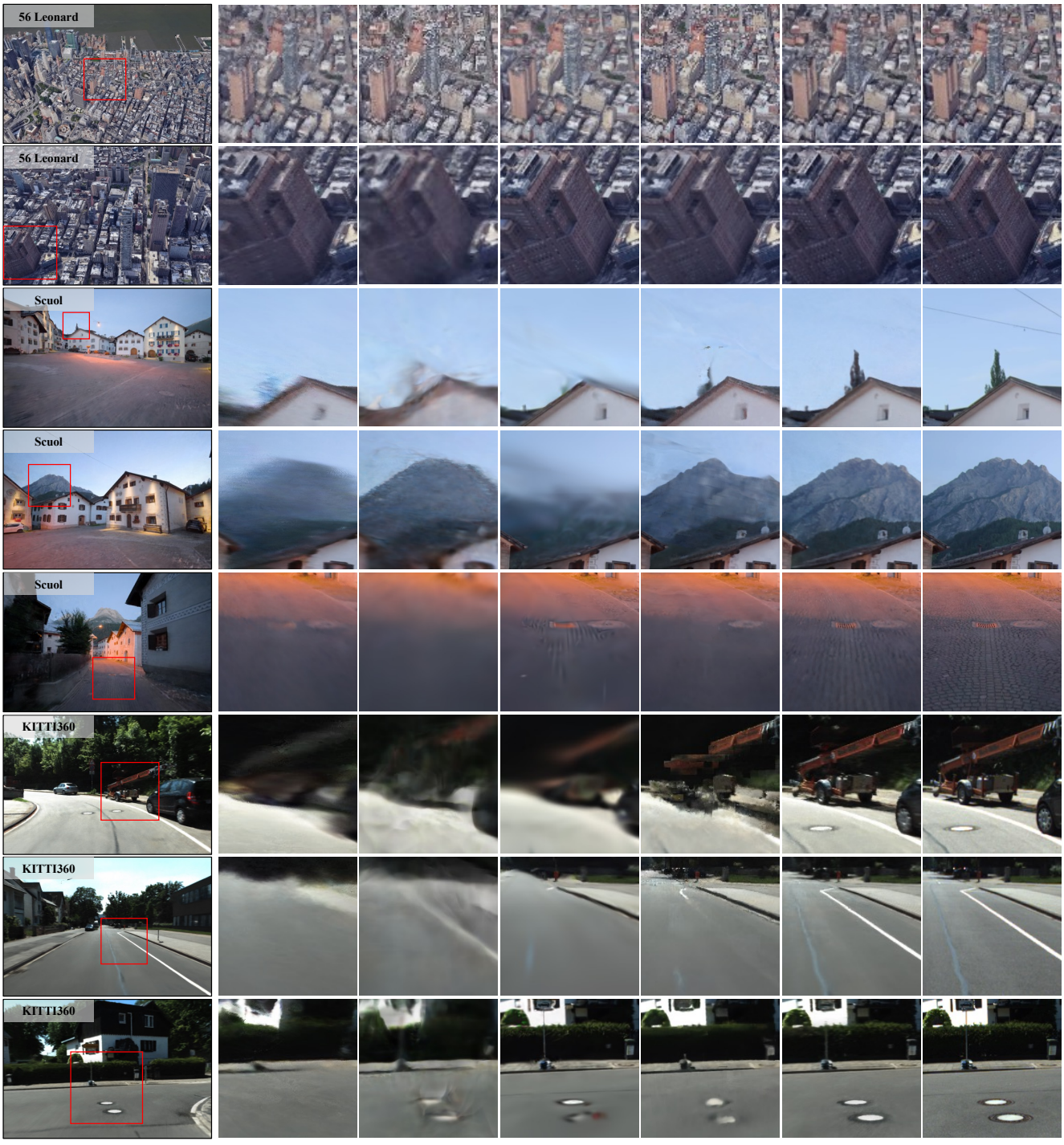


Figure 5. **More novel view synthesis results on the Large Scale scenes.**

References

[1] Alex Yu and Sara Fridovich-Keil, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. *CVPR*, 2022. 1, 3, 4

[2] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. 1, 3, 4

[3] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Zip-nerf: Anti-aliased grid-based neural radiance fields. *ICCV*, 2023. 1

[4] Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In *CVPR*, 2021. 1

[5] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. on Graphics*, 2022. 1, 3, 4

[6] Marie-Julie Rakotosaona, Fabian Manhardt, Diego Martín Arroyo, Michael Niemeyer, Abhijit Kundu, and Federico Tombari. Nerfmeshing: Distilling neural radiance fields into geometrically-accurate 3d meshes. *arXiv.org*, 2023. 1

[7] Sara Sabour, Suhani Vora, Daniel Duckworth, Ivan Krasin, David J. Fleet, and Andrea Tagliasacchi. Robustnerf: Ignoring distractors with robust losses. In *CVPR*, 2023. 1

[8] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. *CVPR*, 2022. 3, 4

[9] Peng Wang, Yuan Liu, Zhaoxi Chen, Lingjie Liu, Ziwei Liu, Taku Komura, Christian Theobalt, and Wenping Wang. F2-nerf: Fast neural radiance field training with free camera trajectories. 2023. 1, 3, 4

[10] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv.org*, 2010.07492, 2020. 3, 4