Supplementary Material for "Multiscale Representation for Real-Time Anti-Aliasing Neural Rendering"

1. Per-scene Quantitative Results

To perform the comparison on the individual scenes, we report the average error metrics across four scales for each individual scene on **Multiscale-NeRF** in Tab. 1. Our method performs superior rendering quality compared to all the baselines across all the scenes.

Method	Average PSNR↑							
	chair	drums	ficus	hotdog	lego	materials	mic	ship
SNeRG [2]	31.692	24.916	26.838	31.402	29.724	27.024	30.467	26.009
MobileNeRF [1]	29.002	23.363	23.821	29.871	25.914	24.337	26.400	24.900
MobileNeRF [1] w/o SS	27.890	22.209	22.660	28.926	24.627	22.844	24.621	23.875
Ours	**32.571**	**25.567**	**29.702**	**35.096**	**32.038**	**27.511**	**31.672**	**29.226**

Method	Average SSIM↑							
	chair	drums	ficus	hotdog	lego	materials	mic	ship
SNeRG [2]	0.960	0.922	0.951	0.961	0.949	0.940	0.970	0.840
MobileNeRF [1]	0.935	0.892	0.898	0.944	0.897	0.894	0.948	0.800
MobileNeRF [1] w/o SS	0.925	0.873	0.882	0.935	0.881	0.874	0.935	0.782
Ours	**0.971**	**0.932**	**0.969**	**0.978**	**0.972**	**0.945**	**0.978**	**0.889**

Method	Average LPIPS↓							
	chair	drums	ficus	hotdog	lego	materials	mic	ship
SNeRG [2]	0.045	0.078	0.047	0.049	0.052	0.063	0.045	0.159
MobileNeRF [1]	0.058	0.095	0.098	0.067	0.086	0.109	0.064	0.186
MobileNeRF [1] w/o SS	0.073	0.121	0.115	0.085	0.106	0.146	0.095	0.205
Ours	**0.033**	**0.068**	**0.041**	**0.035**	**0.030**	**0.050**	**0.040**	**0.113**

Table 1: **Per-scene quantitative results on Multiscale-NeRF.** For quantitative comparison of models trained and evaluated on multiscale dataset. All the metrics of the scene are averaged across four scales.

2. Additional Qualitative Results

We first visualize the rendering results evaluated on **Multiscale-NeRF** in Fig. 1. One can observe that training with images at multiple resolution yields a significant challenge to the baseline methods, as the results of SNeRG and MobileNeRF are excessively blurry. In contrast, our method learns the high frequencies in high-resolution images, and render low-frequencies when the resolution is decreased. We also provide the calculated LOD in Fig. 2, for better visibility we scale the low-resolution images to the size of full resolution. We can see that with the decreasing of frame resolution, the LOD is increasing, as the Mip-VoG represents a relatively large region when the sample comes from a relatively small resolution image.

We then provide the rendering results that trained on **Synthetic-NeRF** and evaluated on **Multiscale-NeRF** in Fig. 3 for illustration of anti-aliasing. We can see that MobileNeRF provides the best anti-aliasing effect thanks to supersampling, as in the low-resolution images low frequencies are retained. Our method achieve an admirable anti-aliasing performance, comparing with all the methods without using supersampling. The results suggest that our method provide an effective multiscale representation for real-time application, since it improves both multiscale training and anti-aliasing rendering.

Figure 1: **Multiscale-NeRF.** We demonstrate Mip-VoG rendering results compared to other baselines on the test set from eight scenes, trained and evaluated on multiscale dataset. We visualize a crop region (shown in red box) on a same image at 4 different scales as an image pyramid. MobileNeRF yields over smooth results on all scales, while SNeRG lost high frequencies in high-resolution images and product aliasing in low-resolution frames. Our method surpass the baselines by a large margin as the rendering quality is significantly better.

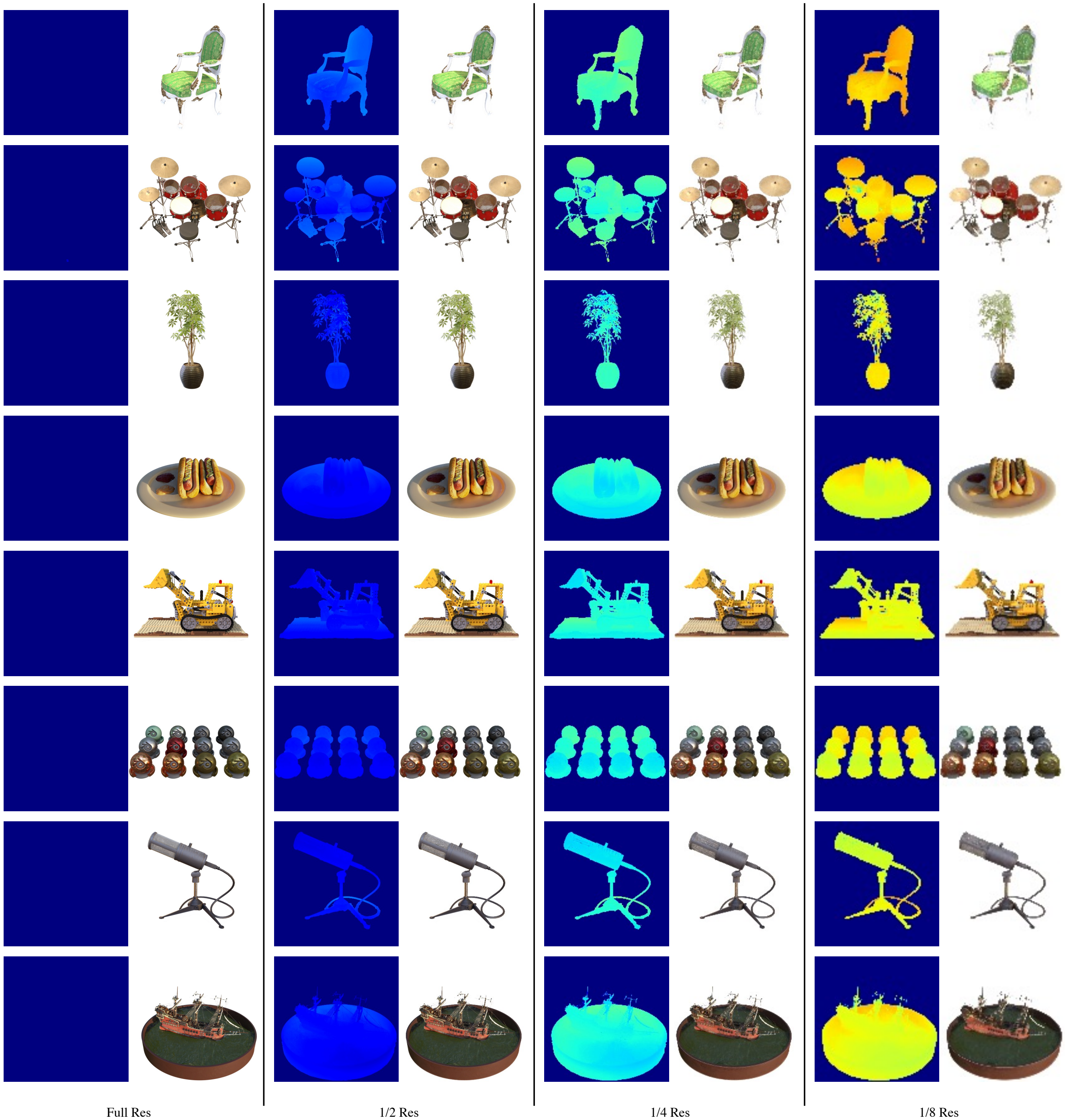

Figure 2: **Visualization of LOD.** We visualize the pixel-wise LOD at four different scales. The value is computed using volume rendering integral of the points' LOD along the ray. The object situated on the left-hand side represents the LOD, while the one positioned on the right-hand side corresponds to our predicted frame. We resized the low-resolution rendered images to match the dimensions of the full resolution images, enhancing visibility and facilitating direct comparison. Brighter color indicates higher value.

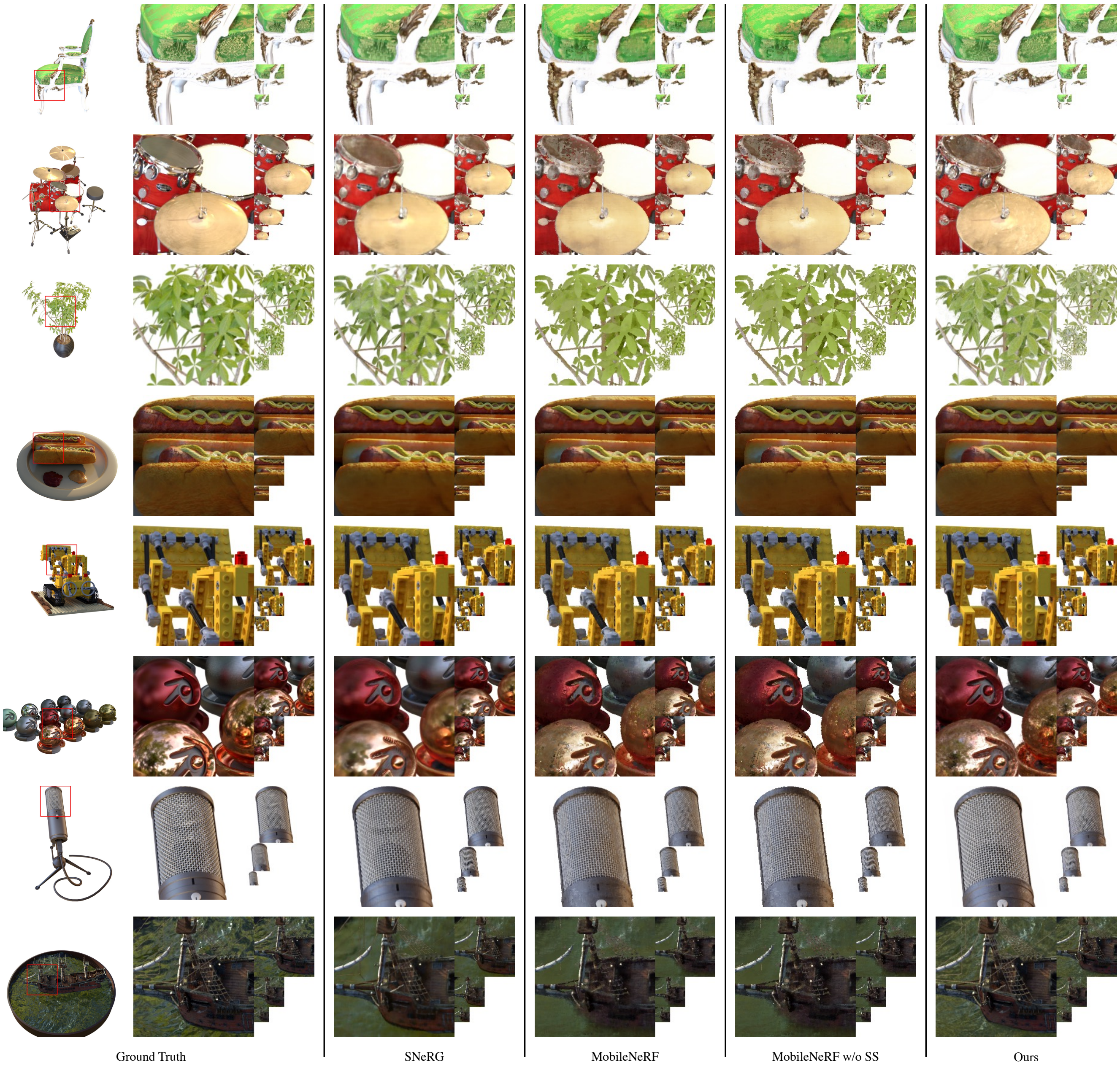

Figure 3: **Synthetic-NeRF.** We demonstrate Mip-VoG rendering results compared to other baselines on the multiscale test set from eight scenes of Multiscale-NeRF, trained on single-scale dataset. We visualize a crop region (shown in red box) on a same image at 4 different scales as an image pyramid. Our method outputs the sharp details on high-resolution images and smooth results on low-resolution images, with regards to the ground truth.

References

[1] Zhiqin Chen, Thomas Funkhouser, Peter Hedman, and Andrea Tagliasacchi. Mobilenerf: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures. In *The Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1

[2] Peter Hedman, Pratul P Srinivasan, Ben Mildenhall, Jonathan T Barron, and Paul Debevec. Baking neural radiance fields for real-time view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5875–5884, 2021. 1