

Supplementary Material for Fourier PlenOctrees for Dynamic Radiance Field Rendering in Real-time

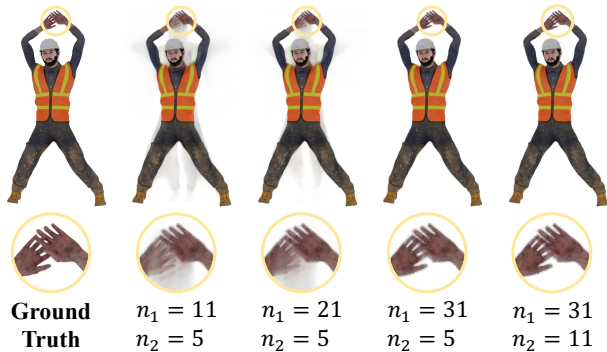


Figure 1. **Qualitative evaluation on Fourier dimensions.** The setting with $n_1 = 31$, $n_2 = 5$ achieves the satisfactory rendering quality while higher Fourier dimension does not result in a significant improvement.

A. Implementation details.

We use IBRNet [2] as generalized NeRF in our paper and we finetune it on the Twindom dataset [1]. The generalized NeRF is important to generate our FPO in 2 hours as it can give a good initialization of FPO, directly train Fourier NeRF-SH can achieve the same performance, but it takes 1-2 days for training it. We render our FPO real-time and online. The generation and fine-tuning are offline. We use Real DFT and Real IDFT to recover density and SH coefficients.

B. More ablation study.

Fourier dimensions. We carried out an experiment to find the best choice of Fourier dimension with both realistic rendering performance and acceptable memory usage. As is shown in Fig. 1 and Tab. 1, the results with $n_1 = 31$, $n_2 = 5$ have a better appearance than those using smaller Fourier dimensions and have less storage cost and faster rendering than using higher dimensions. Our model keeps an outstanding balance.

Components ablation study. To better evaluate the components in our pipeline, we do another additional analysis of different modules as shown in Fig. 2 and Tab. 2. **w/o DFT** means that we do not use DFT coefficients in our

Fourier dimensions	best	second-best	
	PSNR↑	FPS ↑	Storage (GB)↓
$n_1 = 11$ $n_2 = 5$	31.56	118.47	6.421
$n_1 = 21$ $n_2 = 5$	33.31	118.14	6.861
$n_1 = 31$ $n_2 = 5$ (ours)	36.21	117.87	7.251
$n_1 = 31$ $n_2 = 11$	36.40	109.95	14.91

Table 1. **Quantitative evaluation on Fourier dimensions.** Compared with other choices, the setting with $n_1 = 31$, $n_2 = 5$ achieves the best balance among rendering accuracy, time and storage.

	best	second-best		
	PSNR↑	SSIM↑	MAE↓	LPIPS↓
w/o DFT	21.33	0.9558	0.0206	0.0805
w/o generalized NeRF	27.26	0.9652	0.0094	0.0517
w/o running average	27.97	0.9780	0.0088	0.0376
w/o SH	27.93	0.9656	0.0088	0.0493
Ours	32.35	0.9799	0.0059	0.0289

Table 2. **Quantitative evaluation on the components of our pipeline.** Our full model achieves the best performance in PSNR, SSIM, LPIPS and MAE metrics.

model. Each leaf of octree only contains one density and SH coefficients to represent the whole 60 frames. It loses the ability to model the variation of leaf values over time, so it suffers from severe blurred effects. **w/o generalized NeRF** stands for that we do not use our coarse-to-fine scheme that utilizes generalizable NeRF for constructing FPO. Instead, we train Fourier NeRF-SH (describe in Sec. 4.1) and then proceed with Fourier PlenOctree fine-tuning for 2 hours in total as a baseline. It shows that our coarse-to-fine scheme can greatly save training time and provides a good initialization for Fourier PlenOctree fine-tuning. For **w/o running average**, we directly average all the dense views in our fine stage PlenOctree generation, which leads to blurry artifacts. In **w/o SH**, each leaf only stores the Fourier Transform coefficients for density and RGB. It produces voxel artifacts due to not considering the view-dependent effects in real data. Our full model (**Ours**) outperforms other methods both qualitatively and quantitatively.

! "#\$%&' (")\$* - .#' / 0 (- .#'12%2"34562&' - .#' "\$%5%1 - .#'UJ + \$",
7280 392"312

Figure 2. **Qualitative evaluation on our components in our pipeline.** Our full pipeline achieves more realistic rendering results.

References

- [1] Twindom dataset. <https://web.twindom.com/>. 1
- [2] Qianqian Wang, Zhicheng Wang, Kyle Genova, Pratul Srinivasan, Howard Zhou, Jonathan T Barron, Ricardo Martin-Brualla, Noah Snavely, and Thomas Funkhouser. IBRNet: Learning multi-view image-based rendering. In *CVPR*, 2021. 1