Appendix

A. Voronoi vs naïve decomposition – Table 3

To test that our learned decomposition actually contributes to an improvement in accuracy, we test against a version of DeRF with a naïve decomposition – where cells are fixed into a grid. We find that the learned decomposition results in improvement in all metrics.

B. Additional experiments on real data

Per-scene results – Figures 10–17. In this section we provide further comprehensive experiments on the real capture scenes showing how DeRFs with up to 16 heads perform in terms of reconstruction quality and inference cost.

We find that DeRF models provide the best quality-computation trade-off, for all cases in terms of LPIPS, and almost all cases for PSNR and SSIM. Moreover, for the strong majority of scenes, highly decomposed DeRF models give the best results, especially for the perceptual and structural metrics. We also find the advantage of DeRF is most consistent for scenes where the absolute error is higher, i.e. where gains in performance are needed the most. All experiments in this section use the same sample counts as [13].

Variance Test – Table 4. Additionally, to help explain the increasingly chaotic nature of the graphs as quality increases, we perform an experiment in which we train eight identical DeRF models on the scene with the highest reconstruction accuracy (the "room" scene). In this test we find a significant difference in results depending only on initialization, especially for PSNR. This suggests that, while the current experiments already provide enough evidence for the efficacy of DeRF, extracting robust statistics via multiple runs may deliver a more conclusive answer at the expense of compute. We note however, our results already demonstrate that decomposing the scene is almost always better, especially in terms of perceptual metrics, and when less computation is used for inference.

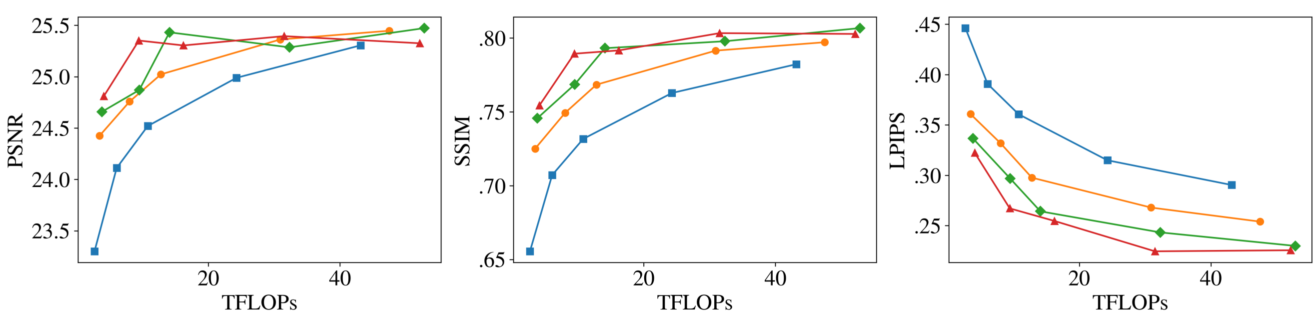
Video supplementary. We further direct the interested readers to the video supplementary that demonstrates the rendering quality and the decomposition in 3D.

	PSNR↑	SSIM↑	LPIPS↓
DeRF	**28.55**	**0.89**	**0.25**
DeRF-Grid	28.07	0.88	0.28

Table 3. **Voronoi vs grid decomposition –** Quantitative results for two equal-capacity 64-head, 32-unit models trained on the "room" scene. The DeRF model uses our Voronoi decomposition and the DeRF-Grid uses an untrained decomposition where the cells are arranged in a 4×4×4 grid. Our Voronoi decomposition brings significant benefit in terms of rendering quality.

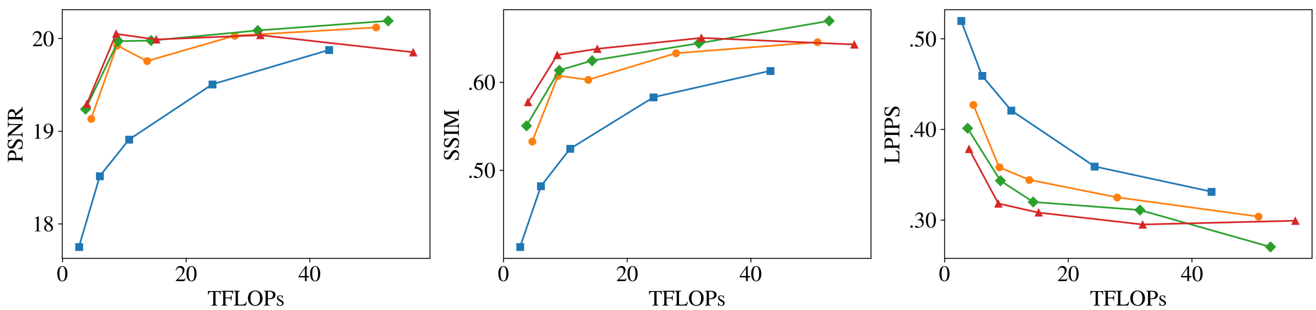
	PSNR↑	SSIM↑	LPIPS↓
Minimum	29.15	0.92	**0.15**
Maximum	**29.86**	**0.93**	0.16

Table 4. **Variation across initialization –** We report the variation in rendering quality resulting from different initializations by showing the minimum and maximum metrics, achieved by a set of 16-head, 128-unit DeRF models trained on the "room" scene.



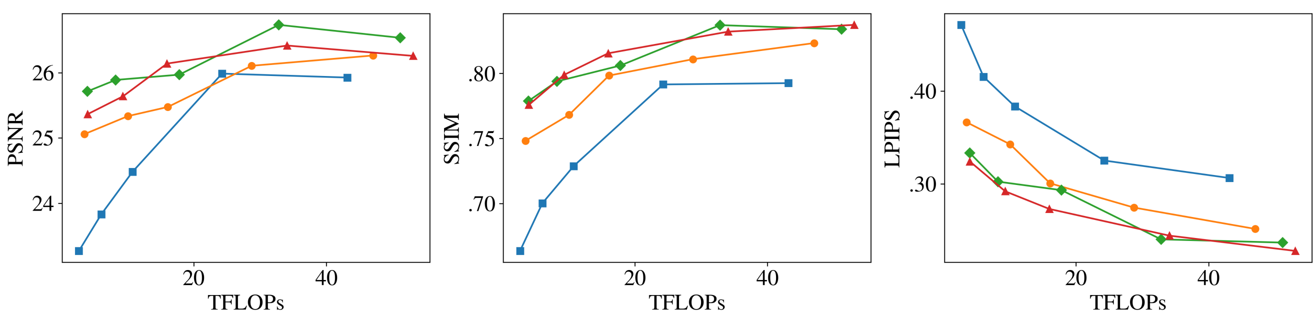
Units	1 Head (□)			4 Heads (○)			8 Heads (◇)			16 Heads (△)		
	PSNR ↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓
64	23.30	0.66	0.45	24.43	0.73	0.36	24.66	0.75	0.34	24.81	0.75	0.32
96	24.11	0.71	0.39	24.76	0.75	0.33	24.87	0.77	0.30	25.35	0.79	0.27
128	24.52	0.73	0.36	25.02	0.77	0.30	25.43	0.79	0.26	25.30	0.79	0.25
192	24.99	0.76	0.31	25.36	0.79	0.27	25.29	0.80	0.24	25.39	0.80	**0.22**
256	25.31	0.78	0.29	25.45	0.80	0.25	**25.47**	**0.81**	0.23	25.33	0.80	0.23

Figure 10. Quantitative results for the "fern" scene.



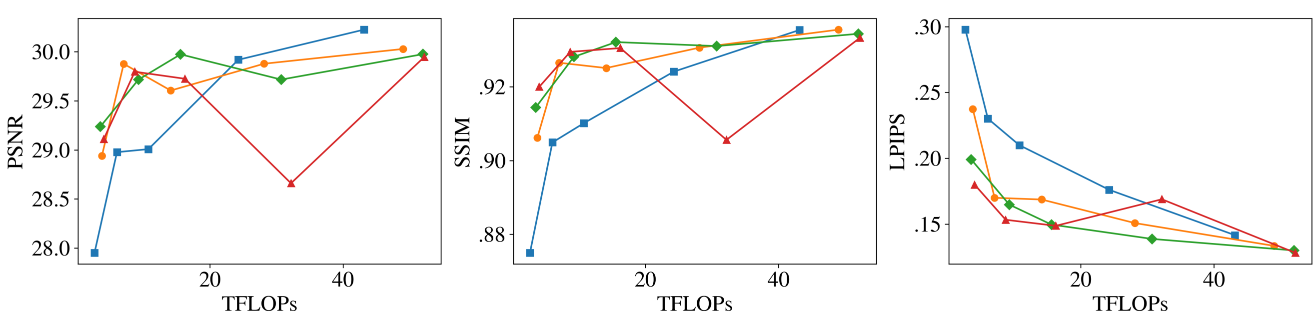
Units	1 Head (□)			4 Heads (○)			8 Heads (◇)			16 Heads (△)		
	PSNR ↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓
64	17.75	0.41	0.52	19.13	0.53	0.43	19.24	0.55	0.40	19.29	0.58	0.38
96	18.51	0.48	0.46	19.92	0.61	0.36	19.97	0.61	0.34	20.05	0.63	0.32
128	18.91	0.52	0.42	19.76	0.60	0.34	19.98	0.62	0.32	19.99	0.64	0.31
192	19.51	0.58	0.36	20.03	0.63	0.33	20.09	0.64	0.31	20.04	0.65	0.30
256	19.88	0.61	0.33	20.12	0.65	0.30	**20.19**	**0.67**	**0.27**	19.85	0.64	0.30

Figure 11. Quantitative results for the “orchids” scene.



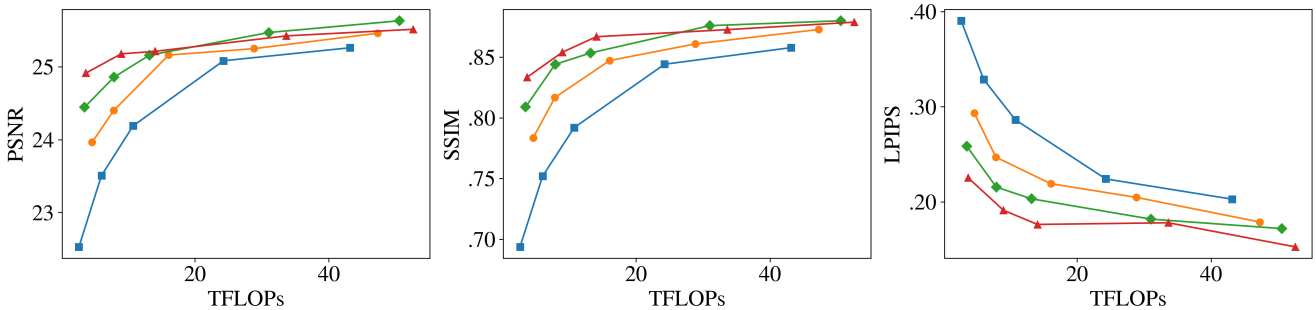
Units	1 Head (□)			4 Heads (○)			8 Heads (◇)			16 Heads (△)		
	PSNR ↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓
64	23.27	0.66	0.47	25.06	0.75	0.37	25.71	0.78	0.33	25.36	0.78	0.32
96	23.83	0.70	0.42	25.33	0.77	0.34	25.89	0.79	0.30	25.64	0.80	0.29
128	24.48	0.73	0.38	25.48	0.80	0.30	25.97	0.81	0.29	26.14	0.82	0.27
192	25.99	0.79	0.33	26.11	0.81	0.27	**26.73**	**0.84**	0.24	26.42	0.83	0.24
256	25.93	0.79	0.31	26.26	0.82	0.25	26.54	0.83	0.24	26.26	**0.84**	**0.23**

Figure 12. Quantitative results for the “horns” scene.



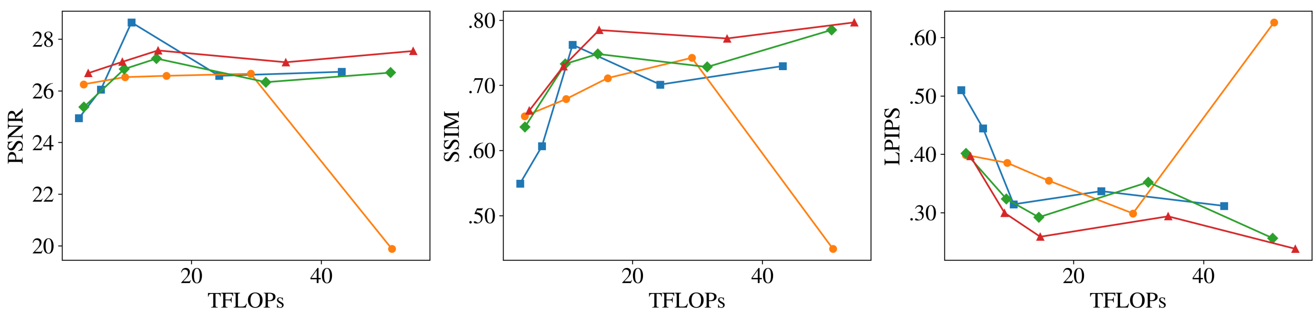
Units	1 Head (□)			4 Heads (○)			8 Heads (◇)			16 Heads (△)		
	PSNR ↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓
64	27.95	0.87	0.30	28.94	0.91	0.24	29.24	0.91	0.20	29.11	0.92	0.18
96	28.98	0.90	0.23	29.88	0.93	0.17	29.72	0.93	0.16	29.80	0.93	0.15
128	29.01	0.91	0.21	29.60	0.93	0.17	29.97	0.93	0.15	29.72	0.93	0.15
192	29.92	0.92	0.18	29.88	0.93	0.15	29.72	0.93	0.14	28.66	0.91	0.17
256	**30.22**	**0.94**	0.14	30.03	**0.94**	**0.13**	29.97	0.93	**0.13**	29.95	0.93	**0.13**

Figure 13. Quantitative results for the “room” scene.



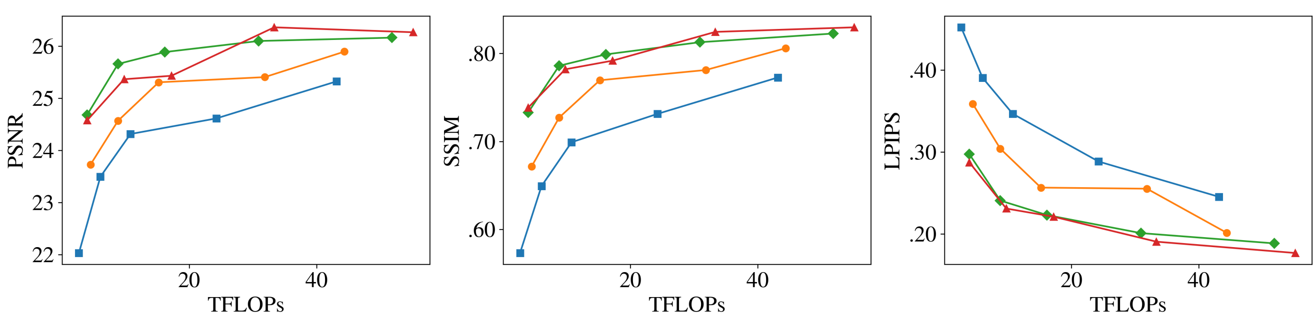
Units	1 Head (□)			4 Heads (○)			8 Heads (◇)			16 Heads (△)		
	PSNR ↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓
64	22.53	0.69	0.39	23.96	0.78	0.29	24.44	0.81	0.26	24.91	0.83	0.23
96	23.51	0.75	0.33	24.40	0.82	0.25	24.86	0.84	0.22	25.18	0.85	0.19
128	24.19	0.79	0.29	25.16	0.85	0.22	25.16	0.85	0.20	25.21	0.87	0.18
192	25.08	0.84	0.22	25.25	0.86	0.20	25.47	**0.88**	0.18	25.42	0.87	0.18
256	25.26	0.86	0.20	25.46	0.87	0.18	**25.63**	**0.88**	0.17	25.51	**0.88**	**0.15**

Figure 14. Quantitative results for the “trex” scene.



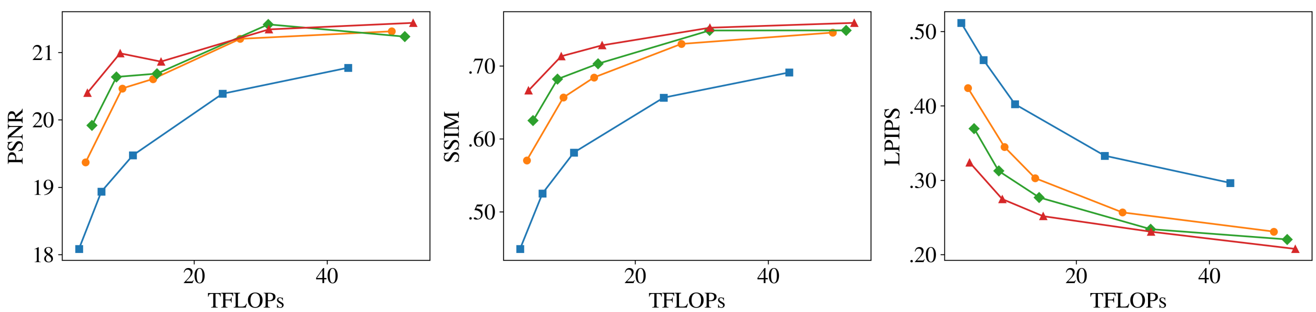
Units	1 Head (□)			4 Heads (○)			8 Heads (◇)			16 Heads (△)		
	PSNR ↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓
64	24.94	0.55	0.51	26.25	0.65	0.40	25.37	0.64	0.40	26.68	0.66	0.40
96	26.05	0.61	0.44	26.53	0.68	0.39	26.84	0.73	0.32	27.12	0.73	0.30
128	**28.65**	0.76	0.31	26.58	0.71	0.35	27.25	0.75	0.29	27.56	0.78	0.26
192	26.58	0.70	0.34	26.66	0.74	0.30	26.33	0.73	0.35	27.10	0.77	0.29
256	26.74	0.73	0.31	19.88	0.45	0.63	26.70	0.79	0.26	27.54	**0.80**	**0.24**

Figure 15. Quantitative results for the “fortress” scene.



Units	1 Head (□)			4 Heads (○)			8 Heads (◇)			16 Heads (△)		
	PSNR ↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓
64	22.03	0.57	0.45	23.73	0.67	0.36	24.68	0.73	0.30	24.57	0.74	0.29
96	23.49	0.65	0.39	24.57	0.73	0.30	25.66	0.79	0.24	25.36	0.78	0.23
128	24.31	0.70	0.35	25.30	0.77	0.26	25.88	0.80	0.22	25.43	0.79	0.22
192	24.61	0.73	0.29	25.40	0.78	0.26	26.10	0.81	0.20	**26.36**	0.82	0.19
256	25.32	0.77	0.25	25.89	0.81	0.20	26.16	0.82	0.19	26.26	**0.83**	**0.18**

Figure 16. Quantitative results for the “flower” scene.



Units	1 Head (□)			4 Heads (○)			8 Heads (◇)			16 Heads (△)		
	PSNR ↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓	PSNR↑	SSIM↑	LPIPS ↓
64	18.08	0.45	0.51	19.37	0.57	0.42	19.92	0.63	0.37	20.40	0.67	0.32
96	18.93	0.53	0.46	20.47	0.66	0.34	20.64	0.68	0.31	20.99	0.71	0.27
128	19.47	0.58	0.40	20.61	0.68	0.30	20.69	0.70	0.28	20.86	0.73	0.25
192	20.39	0.66	0.33	21.20	0.73	0.26	21.42	0.75	0.23	21.34	0.75	0.23
256	20.77	0.69	0.30	21.31	0.75	0.23	21.23	0.75	0.22	**21.44**	**0.76**	**0.21**

Figure 17. Quantitative results for the “leaves” scene.

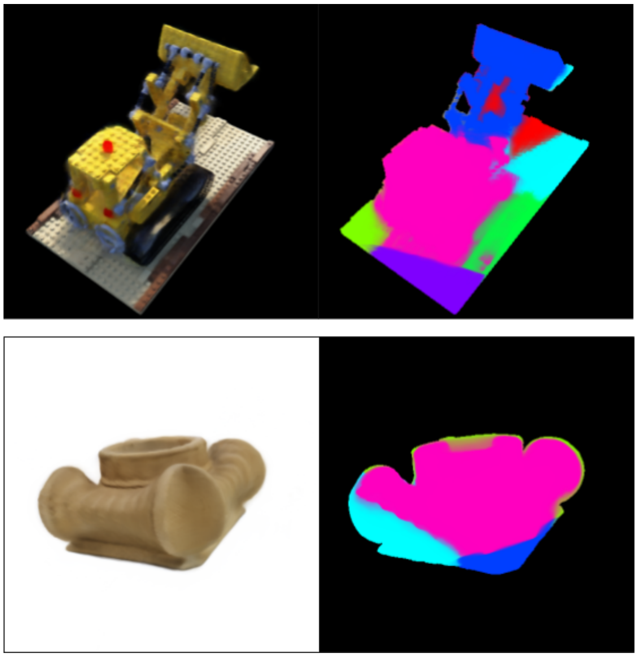


Figure 18. **Synthetic Data** – Renders and decompositions for the NeRF [13] "lego" scene (top) and DeepVoxels [18] "greek" scene (bottom).

C. Synthetic data – Figure 18

While we are mainly interested in photorealistic scenes, our method can also learn decomposed models from synthetic data. To show this we provide renders and visualize the decompositions for DeRF models trained on two synthetic scenes. Each model uses 192 units and 8 heads.