Supplementary Material for "AdaNeRF: Adaptive Sampling for Real-time Rendering of Neural Radiance Fields"

Andreas Kurz[1⋆], Thomas Neff[1⋆], Zhaoyang Lv[2], Michael Zollhöfer[2], and Markus Steinberger[1]

[1] Graz University of Technology, Austria
[2] Reality Labs Research, USA

1 Training speed

AdaNeRF requires $\approx 4-6$ hours to train the sampling network, and $\approx 1-3$ hours to train the shading network (less when using half precision). Similarly, DONeRF (assuming ground truth depth is available) requires ≈ 6 hours to train its oracle, and $\approx 1-3$ hours to train its shading network. However, this assumes that a depth estimate exists—we used a full NeRF to estimate and export depth for DONeRF ($\approx 24-48$ hours). TermiNeRF* requires even more dense training targets for its sampling network, which results in even slower training times compared to DONeRF, depending on the size of the dataset. Finally, the fast CUDA/C++ implementation of Instant-NGP trains in $\approx 10-30$ minutes. Similar concepts (hash table encoding, tiny efficient shading networks) could also potentially be applied to AdaNeRF in the future.

2 Large-Scale Scene Subdivision and Rendering

In the supplementary video on our project page, we show our approach is able to scale to larger scenes, with a new dataset based on the *Pavillon* scene of the DONeRF dataset, consisting of 15 slightly overlapping view cells, a complex scene with reflections and close up shots of intricate objects. For each view cell, we train a separate AdaNeRF. During rendering, we interleave the outputs of two overlapping AdaNeRF, thus evaluating only one AdaNeRF per pixel, which does not increase the computational cost. To this extent, we define a blue noise sequence that blends the transition between cells to show that AdaNeRF is consistent across different view cells in space. We greedily select the two closest view cells that are oriented in the camera viewing direction for blending, ensuring at most two networks in memory at any time.

[⋆] Authors contributed equally to this work.

3 Per-Scene Hyperparameters

We provide the major per-scene hyperparameters that we used to train AdaNeRF in Table 1 and Table 2. The positional encoding frequencies are given in the format X-Y, where X denotes the number of positional frequencies, and Y denotes the number of directional frequencies that are added to the input of our networks. The loss weights directly refer to λ_0 and λ_1 in Equ.2 in the main paper.

4 Per-Scene Quantitative Results

We present per-scene results for the DONeRF dataset in Table 3, and for the LLFF dataset in Table 4. Additionally, we present per-scene results for the comparison against AutoInt on the lower-resolution LLFF dataset in Table 5.

5 Adaptive Sampling and Estimated Depth Results

We present estimated depth and adaptive sample count visualizations for the *Classroom* and *T-Rex* scenes in Figure 1 and Figure 2.

6 Additional Qualitative Results

We further present a selection of additional example outputs in Figure 3, Figure 3, Figure 5 and Figure 6.

Table 1: Positional encoding parameters and loss weights used for the evaluation of the *DONeRF* dataset. Please refer to Equ.2 (and Section 3.2) in the main paper for explanation of the loss weights $\boldsymbol{\lambda}$.

	Positional Encoding		Loss Weights $\boldsymbol{\lambda}$	
	Sampling Network	Shading Network	Sparsity Loss (λ_0)	MSE Loss (λ_1)
Barbershop	10-4	10-4	0.001	1.0
Bulldozer	10-4	10-4	0.005	1.0
Classroom	10-4	10-4	0.001	1.0
Forest	10-4	10-4	0.001	1.0
Pavillon	4-8	14-4	0.0025	1.0
San Miguel	14-4	14-4	0.0025	1.0

Table 2: Positional encoding parameters and loss weights used for the evaluation of the *LLFF* dataset. Please refer to Equ.2 (and Section 3.2) in the main paper for explanation of the loss weights $\boldsymbol{\lambda}$.

	Positional Encoding		Loss Weights $\boldsymbol{\lambda}$	
	Sampling Network	Shading Network	Sparsity Loss (λ_0)	MSE Loss (λ_1)
Fern	2-2	10-4	0.005	1.0
Flower	2-2	10-4	0.005	1.0
Fortress	2-2	10-4	0.005	1.0
Horns	2-2	10-4	0.01	1.0
Leaves	2-2	10-4	0.025	1.0
Orchids	2-2	10-4	0.025	1.0
Room	2-2	10-4	0.0025	1.0
T-Rex	2-2	10-4	0.0075	1.0

Table 3: Image quality comparison on the DONeRF dataset. Best results are displayed as **Top 1**, **Top 2** and **Top 3** per category.

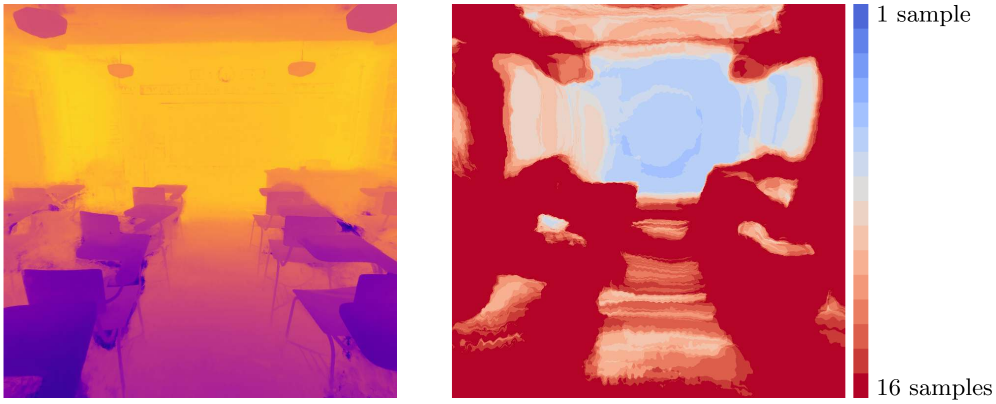
Method	Samples per Pixel	Time	Memory	Barbershop	Bulldozer	Classroom	Forest	Pavillon	San Miguel
		[ms]↓	[MB]↓	PSNR↑	PSNR↑	PSNR↑	PSNR↑	PSNR↑	PSNR↑
AdaNeRF	1.93	**31.32**	**4.14**	26.36	26.85	26.42	24.59	20.04	24.38
AdaNeRF	3.74	**48.20**	**4.14**	27.39	28.17	29.86	27.12	26.91	25.36
AdaNeRF	7.03	78.88	**4.14**	28.54	30.48	31.63	**28.85**	30.96	26.39
AdaNeRF	12.58	130.65	**4.14**	29.82	32.85	**33.26**	**29.57**	**31.96**	**27.58**
DONeRF	2.00	**51.30**	**4.14**	28.14	27.94	29.90	27.14	29.24	24.86
DONeRF	4.00	86.29	**4.14**	29.22	29.86	30.94	27.75	29.81	25.06
DONeRF	8.00	156.27	**4.14**	30.37	32.11	32.18	28.15	30.40	25.67
DONeRF	16.00	296.23	**4.14**	**31.29**	33.92	**33.30**	**29.55**	30.89	26.35
TermiNeRF*	2.00	**51.30**	**4.14**	27.41	24.93	29.57	26.52	29.94	24.83
TermiNeRF*	4.00	86.29	**4.14**	28.69	26.57	30.87	27.26	30.56	25.55
TermiNeRF*	8.00	156.27	**4.14**	29.61	27.94	31.81	28.04	**31.27**	26.32
TermiNeRF*	16.00	296.23	**4.14**	30.44	28.90	32.61	28.36	**31.66**	27.04
NeRF	256.00	2360.67	**3.79**	**32.84**	**36.14**	**34.62**	24.62	30.53	26.82
Plenoxels	-	**47.93**	212.75	29.78	29.70	31.65	20.25	26.67	24.45
Plenoxels-MSI	-	**47.52**	892.89	30.41	28.26	32.84	28.55	29.98	**27.79**
Instant-NGP-2^{14}	-	100.67	**2.00**	30.17	**34.04**	31.21	25.46	29.77	25.92
Instant-NGP-2^{19}	-	137.00	64.00	**34.17**	**38.42**	**36.54**	**29.09**	30.89	**29.36**

Table 4: Image quality comparison on the LLFF dataset. Best results are displayed as Top 1, Top 2 and Top 3 per category.

Method	Samples per Pixel	Time	Memory	Fern	Flower	Fortress	Horns	Leaves	Orchids	Room	T-Rex
		[ms]↓	[MB]↓	PSNR↑	PSNR↑	PSNR↑	PSNR↑	PSNR↑	PSNR↑	PSNR↑	PSNR↑
AdaNeRF	1.99	37.96	4.14	19.83	24.09	26.27	22.69	17.86	16.37	26.85	21.97
AdaNeRF	3.88	58.95	4.14	21.39	25.68	29.45	25.23	19.35	17.58	29.02	24.41
AdaNeRF	6.92	92.71	4.14	22.31	26.98	30.27	26.54	20.31	18.72	30.73	25.52
AdaNeRF	10.24	129.58	4.14	23.39	27.57	30.44	26.86	20.53	19.64	31.23	26.03
DONeRF	2.00	61.08	4.14	20.16	22.14	24.67	21.73	16.89	16.37	24.19	20.97
DONeRF	4.00	102.75	4.14	20.64	22.65	26.03	22.47	17.14	16.91	25.23	21.90
DONeRF	8.00	186.07	4.14	21.16	23.21	27.12	23.37	17.39	17.56	26.22	22.71
DONeRF	16.00	352.72	4.14	21.62	23.78	27.46	24.08	17.68	18.04	27.06	23.52
TermiNeRF*	2.00	61.08	4.14	21.35	22.97	25.33	23.02	17.40	16.11	24.98	22.29
TermiNeRF*	4.00	102.75	4.14	22.02	23.59	25.82	23.83	17.58	17.01	25.66	23.29
TermiNeRF*	8.00	186.07	4.14	22.68	24.09	26.60	24.53	17.96	17.33	26.63	24.18
TermiNeRF*	16.00	352.72	4.14	23.24	24.27	27.58	25.04	17.89	17.87	27.70	24.81
NeRF	256.00	2810.85	3.79	25.17	27.40	31.16	27.45	20.92	20.36	32.70	26.80
Plenoxels	-	51.30	184.58	23.40	26.73	27.80	24.67	20.15	20.57	27.23	23.46
Plenoxels-Large	-	110.06	3629.85	25.46	27.83	31.09	27.58	21.41	20.24	30.22	26.48
Instant-NGP-2^{14}	-	102.12	2.00	24.56	26.45	29.23	25.31	18.25	20.02	30.33	23.99
Instant-NGP-2^{19}	-	161.75	64.00	25.87	26.52	27.96	26.60	18.93	20.31	31.96	26.50

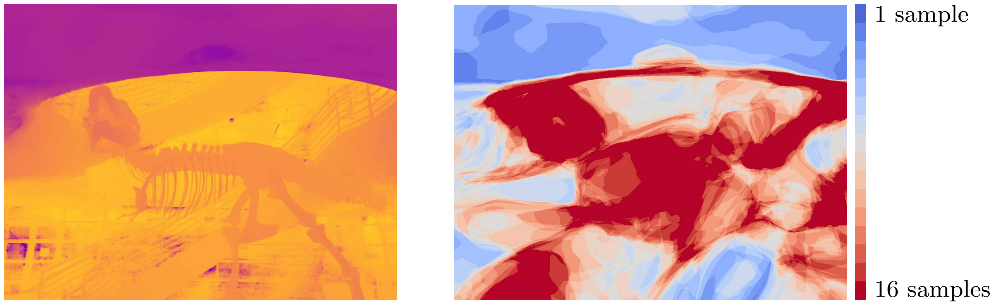
Table 5: Image quality comparison on a lower-resolution version (504 × 378) LLFF dataset. Best results are displayed as Top 1, Top 2 and Top 3 per category.

Method	Samples per Pixel	Time	Memory	Fern	Flower	Fortress	Horns	Leaves	Orchids	Room	T-Rex
		[ms]↓	[MB]↓	PSNR↑	PSNR↑	PSNR↑	PSNR↑	PSNR↑	PSNR↑	PSNR↑	PSNR↑
AdaNeRF	2.00	9.92	4.14	20.89	24.07	28.24	20.96	17.69	15.31	26.19	21.17
AdaNeRF	3.90	15.25	4.14	22.35	25.04	29.21	22.58	20.10	16.25	28.00	22.89
AdaNeRF	7.00	24.38	4.14	23.70	26.56	30.67	25.30	21.28	17.77	29.97	25.58
AdaNeRF	10.61	36.13	4.14	24.39	28.02	31.21	26.99	21.60	19.68	30.72	27.31
AutoInt (N = 8)	16.00	44.61	4.14	22.11	26.65	28.63	26.01	19.61	16.85	28.33	24.90
AutoInt (N = 16)	32.00	88.53	4.14	23.29	27.60	29.53	26.72	18.78	17.71	29.97	25.58
AutoInt (N = 32)	64.00	176.36	4.14	23.51	28.11	28.95	27.64	20.84	17.30	30.72	27.18



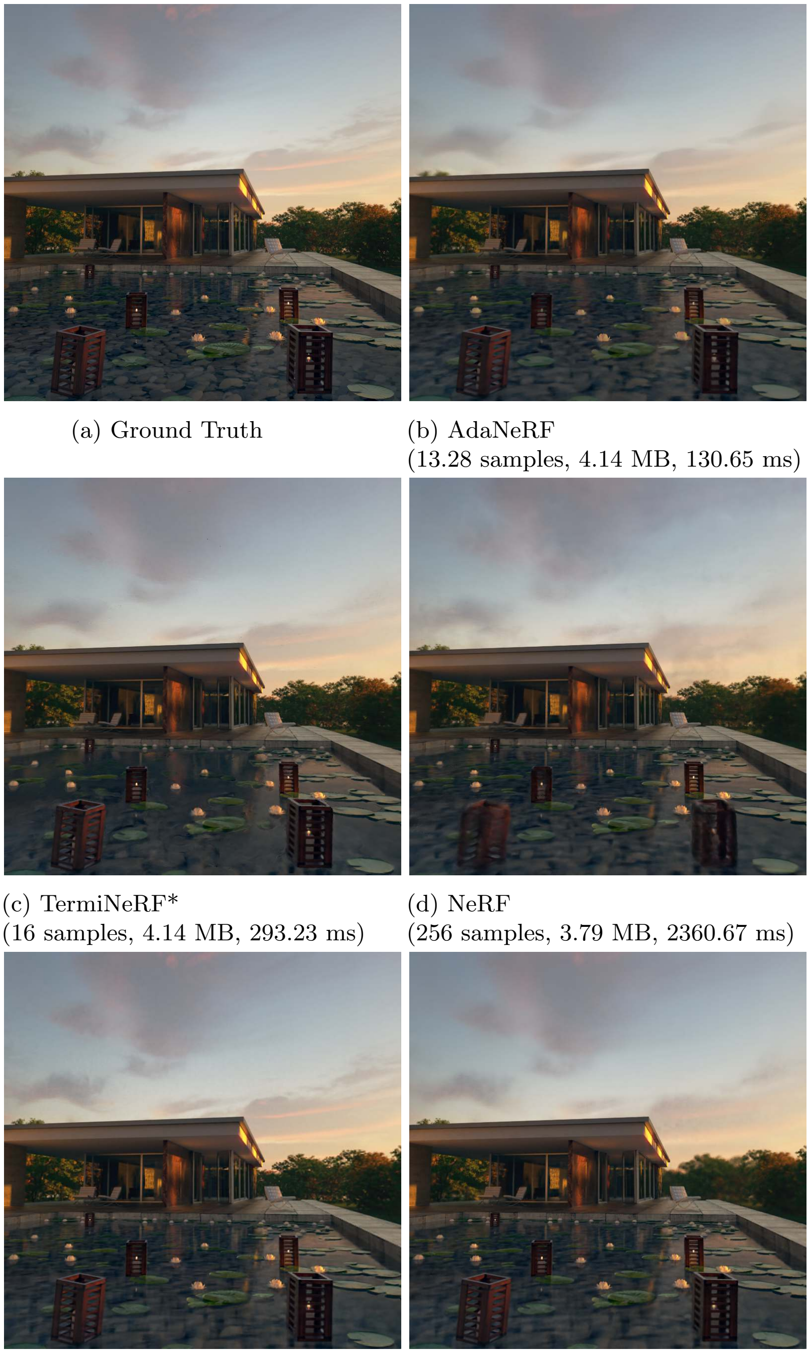
(a) Estimated depth (b) Sample count per pixel

Fig. 1: Estimated depth and adaptive sample counts using AdaNeRF (max. 16 samples per ray) per ray (red = 16 samples per ray, blue = 1 sample per ray) for the *Classroom* scene of the *DONeRF* dataset. Complex geometric objects (such as the chairs) require more samples, while the background mostly requires few samples to achieve good quality.



(a) Estimated depth (b) Sample count per pixel

Fig. 2: Estimated depth and adaptive sample counts using AdaNeRF (max. 16 samples per ray) per ray (red = 16 samples per ray, blue = 1 sample per ray) for the *T-Rex* scene of the *LLFF* dataset. The T-Rex requires many samples to correctly resolve depth and visibility, whereas the background can be resolved in 1-4 samples only.



(a) Ground Truth

(b) AdaNeRF
(13.28 samples, 4.14 MB, 130.65 ms)

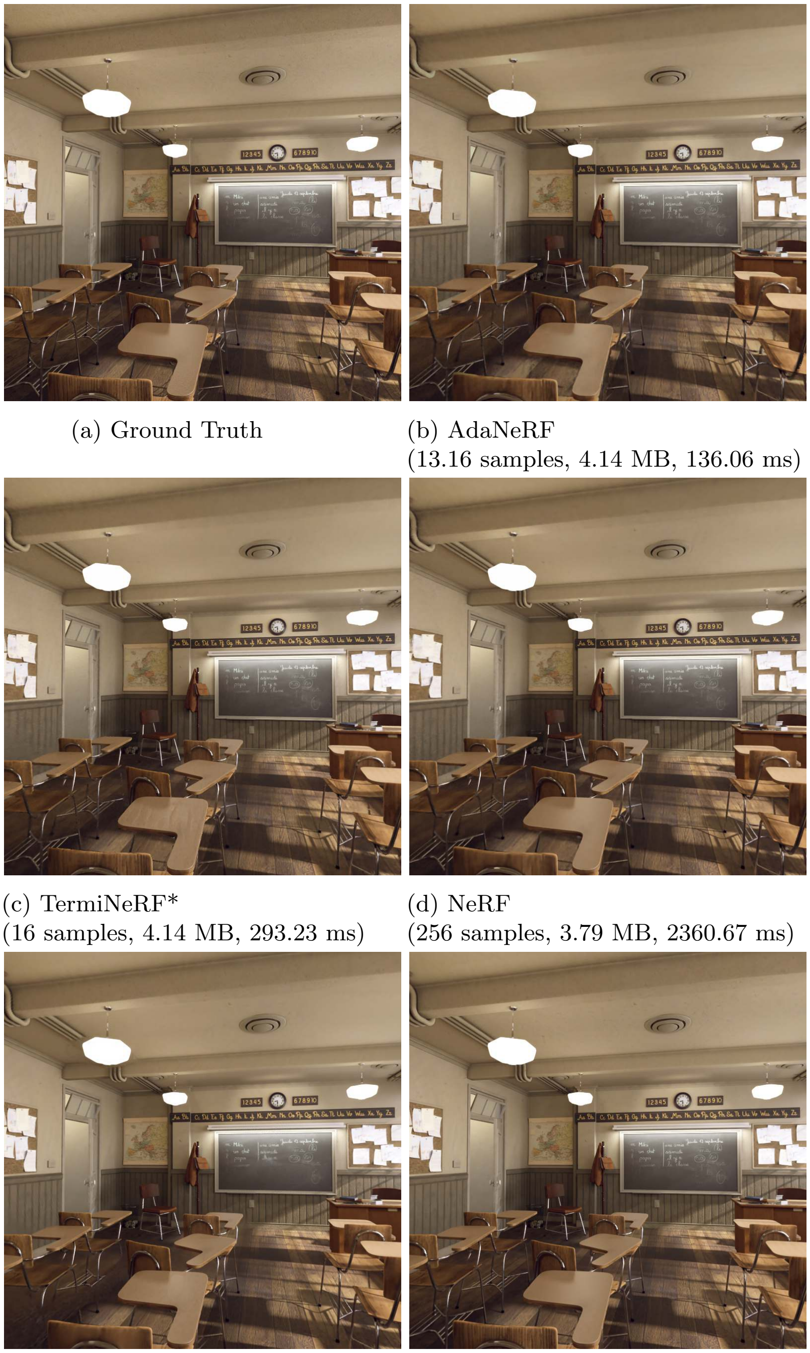
(c) TermiNeRF*
(16 samples, 4.14 MB, 293.23 ms)

(d) NeRF
(256 samples, 3.79 MB, 2360.67 ms)

(e) Plenoxels
(553.95 MB, 33.99 ms)

(f) Instant-NGP-2^{19}
(64.0 MB, 287.0 ms)

Fig. 3: Example images from the *Pavillon* scene of the DONeRF dataset.



(a) Ground Truth

(b) AdaNeRF
(13.16 samples, 4.14 MB, 136.06 ms)

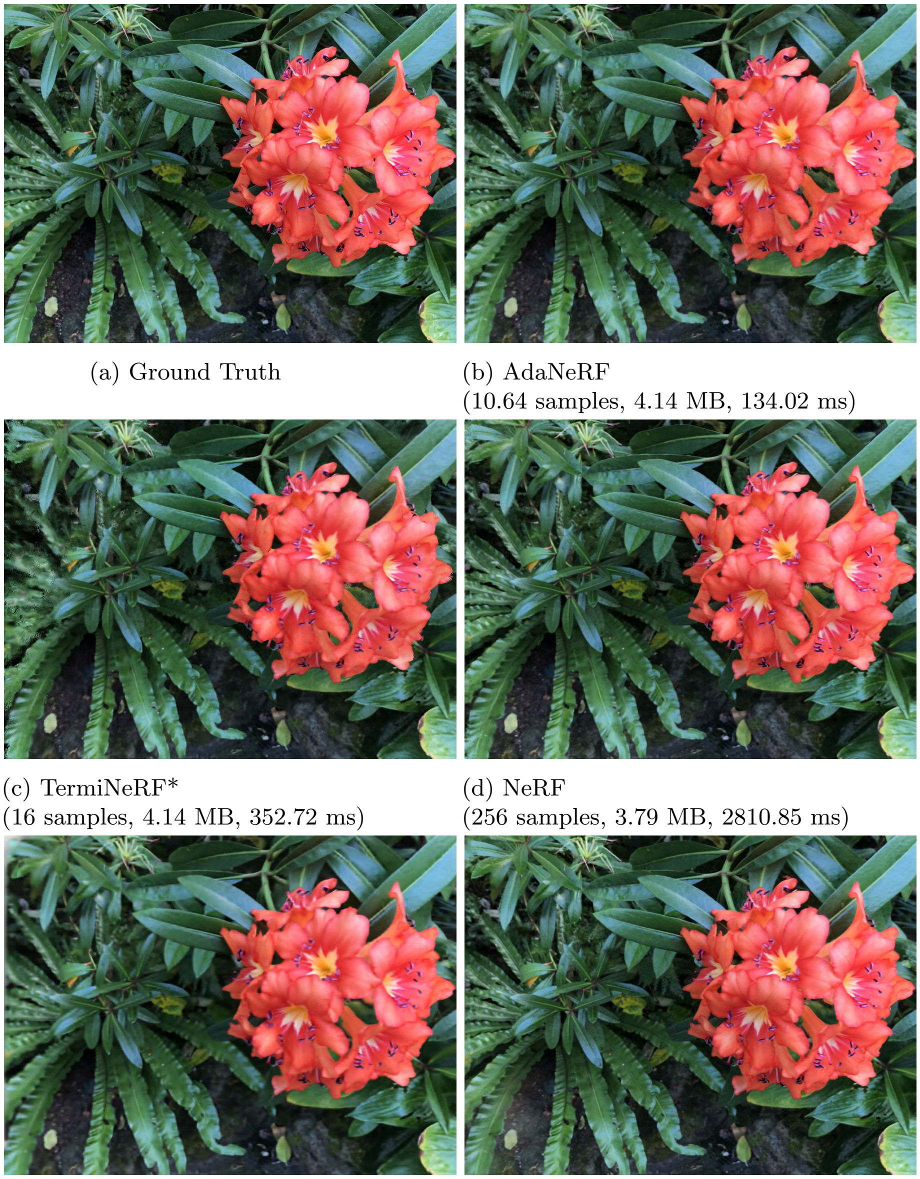
(c) TermiNeRF*
(16 samples, 4.14 MB, 293.23 ms)

(d) NeRF
(256 samples, 3.79 MB, 2360.67 ms)

(e) Plenoxels
(367.31 MB, 71.24 ms)

(f) Instant-NGP-2^{19}
(64.0 MB, 88.0 ms)

Fig. 4: Example images from the *Classroom* scene of the DONeRF dataset.



(a) Ground Truth

(b) AdaNeRF
(10.64 samples, 4.14 MB, 134.02 ms)

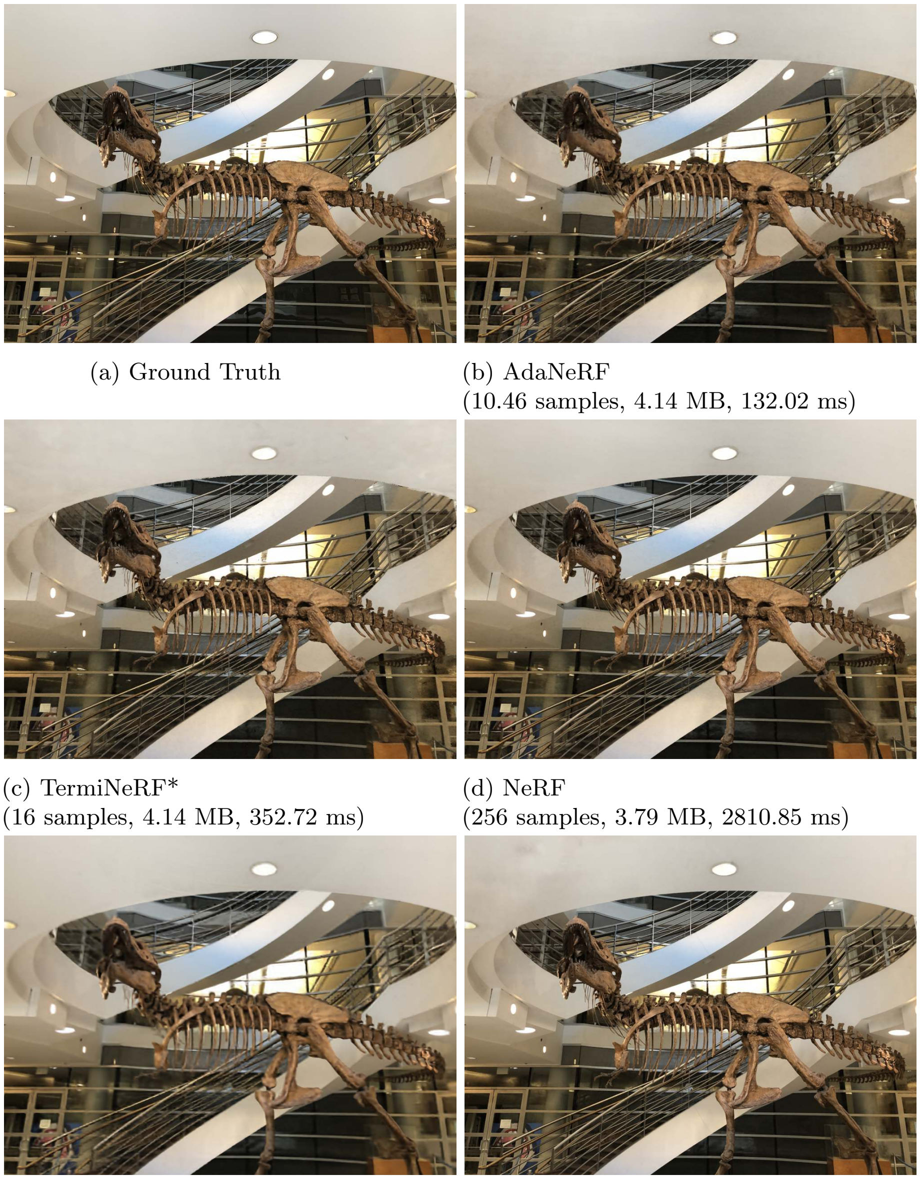
(c) TermiNeRF*
(16 samples, 4.14 MB, 352.72 ms)

(d) NeRF
(256 samples, 3.79 MB, 2810.85 ms)

(e) Plenoxels
(125.07 MB, 51.07 ms)

(f) Instant-NGP-2^{19}
(64.0 MB, 152.0 ms)

Fig. 5: Example images from the *Flower* scene of the LLFF dataset.



(a) Ground Truth

(b) AdaNeRF
(10.46 samples, 4.14 MB, 132.02 ms)

(c) TermiNeRF*
(16 samples, 4.14 MB, 352.72 ms)

(d) NeRF
(256 samples, 3.79 MB, 2810.85 ms)

(e) Plenoxels
(178.66 MB, 47.46 ms)

(f) Instant-NGP-2^{19}
(64.0 MB, 119.0 ms)

Fig. 6: Example images from the *T-Rex* scene of the LLFF dataset.