How Far Can We Compress Instant-NGP-Based NeRF?

Supplementary Material

In this supplementary material, we first report more implementation details in Sec. A, then exhibit the efficient backward of our bit estimator function in Sec. B, and we also present a notation table in Sec. C for clarity in understanding our paper. Additionally, more quantitative and qualitative results are included at the end of the document.

A. More Implementation Details

Context Fuser. *Context Fuser* is able to aggregate contexts from previous L_c levels. For 3D embeddings, it is a 3-layer MLP. It has an input channel of $F \times L_c + 1$, a hidden channel of 32 and an output channel of F with LeakyReLU activation, where "+1" is for f_G. For 2D embeddings, it is a 1-layer linear module. It has an input channel of $F \times L_c + 1 + F$ and an output channel of F, where "$+F$" is for dimension-wise context. Note that different levels of the same L_cs share one *Context Fuser* to save storage space.

Learning Rate. Our initial learning rate is 0.01 and the total iteration is $20K$. For the first $1K$ iterations, we adopt a linear warm-up stage to stabilize the training process. For the rest iterations, the learning rate is decayed by a factor of 0.33 at $9K, 12K, 15K, 17K$ and $19K$ iterations.

Sampling Strategy. During training, feeding all embeddings to the context models in a single iteration will lead to out-of-memory (OOM). To address this, we randomly sample $150K$ feature vectors θs for training 3D embeddings under $F = 8$ in each iteration, and $200K$ under $F = 1, 2, 4$. For 2D embeddings, we do not employ this strategy but feed them all together in one iteration.

Quantization of the Rendering MLP. We utilize 13 bits to quantize the rendering MLP. For each parameter $\omega_i \in \Omega$,

$$\omega_{qi} = \lfloor (\omega_i - \mathrm{MIN}(\Omega)) \frac{2^{\mathrm{D}} - 1}{\mathrm{MAX}(\Omega) - \mathrm{MIN}(\Omega)} \rfloor \quad (5)$$

where Ω is the parameter collection of the rendering MLP and ω_{qi} is the quantized parameter. $D = 13$ represents the number of digits for quantization. MIN and MAX represent operations to calculate minimum and maximum elements, respectively.

Inverse Hash Mapping. While the hash function [29] provides only a unidirectional mapping of $n \to \theta$, we are in need of its inverse mapping of $\theta \to n$. To accomplish this, during the initialization stage, we traverse all ns in voxels using the hash function and store their corresponding θs, which takes a GPU memory of 5 GB. Consequently, we can retrieve all associated vertices $\{n_i^k | k = 1, \ldots, K\}$ of a random vector θ_i by querying this recorded information during training.

B. Efficient Backward of Bit Estimator

In this paper, we estimate the bit consumption of a θ_i with its probability p_i in a differentiable formula, as shown below (same as Eq. 2):

$$bit(p_i|\theta_i) = -(\frac{1+\theta_i}{2}\log_2(p_i) + \frac{1-\theta_i}{2}\log_2(1-p_i)) \quad (6)$$

We discover this estimator is better than the below one:

$$bit(p_i|\theta_i) = -\log_2(\frac{1-\theta_i}{2} + \theta_i p_i) \quad (7)$$

Although these two forms of estimator functions produce the same results of bit consumption in their forward pass, they exhibit quite different behavior for backpropagation. For Eq. 6,

$$\begin{cases} \frac{\partial bit}{\partial \theta_i} = \frac{1}{2}\log_2(\frac{1}{p_i} - 1) \\ \frac{\partial bit}{\partial p_i} = \begin{cases} -\frac{1}{p_i \ln 2} & \theta_i = +1 \\ -\frac{1}{(p_i - 1)\ln 2} & \theta_i = -1 \end{cases} \end{cases} \quad (8)$$

However, there exists a different derivative formula of θ_i for Eq. 7,

$$\begin{cases} \frac{\partial bit}{\partial \theta_i} = \begin{cases} -\frac{p_i - 0.5}{p_i \ln 2} & \theta_i = +1 \\ -\frac{p_i - 0.5}{(1-p_i)\ln 2} & \theta_i = -1 \end{cases} \\ \frac{\partial bit}{\partial p_i} = \begin{cases} -\frac{1}{p_i \ln 2} & \theta_i = +1 \\ -\frac{1}{(p_i - 1)\ln 2} & \theta_i = -1 \end{cases} \end{cases} \quad (9)$$

This inherent difference of backward propagation of Eq. 7 results in a more intricate gradient, significantly slowing down the training speed to 68 minutes. Additionally, it exacerbates the RD performance, leading to undesirable BD-rate increases of +49.3%/+89.5%.

C. Notation Table and More Results

In this subsection, we first provide a notation table, which is necessary for readers to understand our paper. Subsequently, we showcase additional quantitative and qualitative results, offering more detailed data for thorough understanding.

Notation	Definition
$\mathbf{x}$	An input coordinate for rendering
$\mathbf{d}$	Viewing direction of the input coordinate $\mathbf{x}$
$\mathbf{o}$	Camera center to observe the input coordinate $\mathbf{x}$
$\mathbf{r}$	A ray for rendering
v	Index of a sampled point along the ray $\mathbf{r}$
σ	Density of the sampled point v
$\mathbf{c}$	Color of the sampled point v
T	Transmittance to the sampled point v along the ray $\mathbf{r}$
$\hat{C}$	The rendered pixel color of the ray $\mathbf{r}$
$\mathbf{f}$	The interpolated input feature for positional encoding
L	Total resolution level number of embeddings
l	A level out of L
Θ	Collection of feature embeddings in one level
$\boldsymbol{\theta}$	A vector element of embeddings Θ
θ	A scalar of $\boldsymbol{\theta}$, which can be either -1 or $+1$
i	Index of a randomly sampled θ
T	Size of embeddings Θ
f_G	Occurrence frequency of $+1$ in embeddings Θ
n	Associated vertex of θ in the voxel
p	Estimated probability for entropy modeling
L_c	Number of previous levels for context
L_d	Level from which context models are disabled
F	Dimension of feature vectors $\boldsymbol{\theta}$
$\boldsymbol{C_p}$	*Context Fusor* to aggregate contexts
$\boldsymbol{E_p}$	*Bit Estimator* to calculate bit consumption
K	Hash collision number of θ
k	A collided vertex out of K
AOE	Area of effect of the vertex n
PVF	Projected voxel feature for dimension-wise context of 3D to 2D
w	Normalized weights of vertices for hash fusion
L_{mse}	Mean Squared Error (MSE) loss, which measures fidelity
$L_{entropy}$	Entropy loss, which measures embedding size
λ	Tradeoff parameter to balance fidelity and size
Ω	Parameter collection of the rendering MLP
ω	A parameter of the collection Ω
ω_q	The quantized parameter of ω
D	Number of digits for quantizing the rendering MLP
M	Number of θs in the embeddings

Table A. Notation Table

Method	*chair*	*drums*	*ficus*	*hotdog*	*lego*	*materials*	*mic*	*ship*	*Avg.*
				PSNR↑					
Instant-NGP [29]	35.91	25.18	33.76	37.48	35.86	29.65	36.98	30.93	33.22
SHACIRA [9]	31.88	24.52	30.65	34.22	31.79	27.50	32.00	24.12	29.59
MaskDWT($1e-10$) [32]	34.14	25.53	32.87	35.93	34.93	29.54	33.48	29.15	31.94
MaskDWT($5e-11$) [32]	34.52	25.66	33.03	36.20	35.16	29.58	33.68	29.19	32.13
MaskDWT($2.5e-11$) [32]	34.68	25.56	33.17	36.37	35.50	29.56	33.74	29.34	32.24
BiRF-F1 [36]	33.38	25.07	32.26	35.78	33.52	28.74	34.42	29.04	31.53
BiRF-F2 [36]	34.75	25.59	33.91	36.59	35.06	29.49	36.01	29.74	32.64
BiRF-F4 [36]	35.66	25.84	34.42	37.13	36.02	29.80	36.91	30.30	33.26
BiRF-F8 [36]	36.17	26.05	34.71	37.51	36.48	30.09	37.44	30.27	33.59
Ours($F=8, \lambda=4e-3$)	34.76	26.11	34.15	36.96	35.38	30.53	36.64	31.00	33.19
Ours($F=8, \lambda=2e-3$)	35.13	26.08	34.35	37.28	35.76	30.63	37.00	31.46	33.46
Ours($F=8, \lambda=1e-3$)	35.37	26.08	34.46	37.46	35.98	30.75	37.31	31.72	33.64
Ours($F=8, \lambda=0.7e-3$)	35.51	26.18	34.43	37.42	36.16	30.72	37.28	31.83	33.69
				SSIM↑					
Instant-NGP [29]	0.986	0.933	0.983	0.983	0.981	0.950	0.992	0.896	0.963
SHACIRA [9]	0.967	0.929	0.969	0.974	0.966	0.936	0.980	0.847	0.946
BiRF-F1 [36]	0.973	0.921	0.974	0.973	0.965	0.934	0.985	0.877	0.950
BiRF-F2 [36]	0.980	0.930	0.981	0.978	0.976	0.943	0.989	0.888	0.958
BiRF-F4 [36]	0.984	0.934	0.983	0.980	0.980	0.948	0.911	0.895	0.962
BiRF-F8 [36]	0.986	0.937	0.984	0.981	0.982	0.951	0.992	0.897	0.964
Ours($F=8, \lambda=4e-3$)	0.980	0.941	0.983	0.978	0.978	0.958	0.991	0.901	0.964
Ours($F=8, \lambda=2e-3$)	0.982	0.942	0.984	0.980	0.980	0.959	0.992	0.909	0.966
Ours($F=8, \lambda=1e-3$)	0.984	0.941	0.984	0.982	0.981	0.960	0.993	0.913	0.967
Ours($F=8, \lambda=0.7e-3$)	0.984	0.942	0.984	0.982	0.982	0.960	0.993	0.915	0.968
				LPIPS↓					
Instant-NGP [29]	0.021	0.092	0.024	0.034	0.022	0.069	0.014	0.138	0.052
SHACIRA [9]	0.045	0.090	0.043	0.049	0.045	0.083	0.032	0.203	0.074
BiRF-F1 [36]	0.037	0.086	0.034	0.045	0.043	0.078	0.022	0.141	0.061
BiRF-F2 [36]	0.024	0.073	0.024	0.036	0.025	0.064	0.016	0.127	0.049
BiRF-F4 [36]	0.019	0.066	0.020	0.032	0.017	0.057	0.012	0.117	0.043
BiRF-F8 [36]	0.016	0.063	0.018	0.028	0.015	0.051	0.009	0.112	0.039
Ours($F=8, \lambda=4e-3$)	0.028	0.071	0.023	0.043	0.027	0.057	0.015	0.140	0.050
Ours($F=8, \lambda=2e-3$)	0.024	0.070	0.022	0.038	0.024	0.055	0.012	0.130	0.047
Ours($F=8, \lambda=1e-3$)	0.022	0.071	0.020	0.035	0.022	0.054	0.011	0.124	0.045
Ours($F=8, \lambda=0.7e-3$)	0.021	0.069	0.020	0.034	0.021	0.053	0.011	0.121	0.044
				SIZE(MB)↓					
Instant-NGP [29]	45.56	45.56	45.56	45.56	45.56	45.56	45.56	45.56	45.56
SHACIRA [9]	1.477	1.527	1.329	1.739	1.820	1.766	1.174	2.162	1.624
MaskDWT($1e-10$) [32]	0.985	0.988	1.011	0.529	0.787	0.988	0.555	0.766	0.826
MaskDWT($5e-11$) [32]	1.384	1.404	1.394	0.750	1.114	1.401	0.759	1.090	1.162
MaskDWT($2.5e-11$) [32]	1.988	1.858	1.968	1.118	1.647	2.000	1.208	1.712	1.687
BiRF-F1 [36]	0.7	0.7	0.8	0.7	0.7	0.8	0.6	0.8	0.7
BiRF-F2 [36]	1.3	1.5	1.4	1.4	1.4	1.4	1.3	1.5	1.4
BiRF-F4 [36]	2.7	2.9	2.8	2.8	2.8	2.8	2.7	3.0	2.8
BiRF-F8 [36]	5.6	5.7	5.8	5.8	5.8	5.7	5.6	6.0	5.8
Ours($F=8, \lambda=4e-3$)	0.406	0.488	0.365	0.332	0.377	0.485	0.332	0.560	0.418
Ours($F=8, \lambda=2e-3$)	0.511	0.649	0.444	0.367	0.454	0.610	0.366	0.717	0.515
Ours($F=8, \lambda=1e-3$)	0.618	0.852	0.534	0.420	0.554	0.727	0.442	0.915	0.633
Ours($F=8, \lambda=0.7e-3$)	0.689	1.003	0.588	0.470	0.602	0.851	0.471	1.106	0.722

Table B. Detailed quantitative results of storage size against fidelity quality (PSNR, SSIM, LPIPS) of each scene on NeRF-Synthetic dataset. We focus on NeRF compression approaches, along with our base model Instant-NGP. For quantitative results of other approaches, please refer to BiRF [36] paper, as we do not duplicate them here.

Method	*Barn*	*Caterpillar*	*Family*	*Ignatius*	*Truck*	*Avg.*
			PSNR↑			
Instant-NGP [29]	28.19	25.94	34.32	28.17	27.03	28.73
MaskDWT($1e-10$) [32]	26.49	25.50	32.57	28.06	26.21	27.77
BiRF-F1 [36]	27.11	25.48	33.21	27.71	26.80	28.06
BiRF-F2 [36]	27.65	25.87	33.86	27.78	27.31	28.49
BiRF-F4 [36]	27.74	25.97	34.33	27.92	27.46	28.68
BiRF-F8 [36]	27.69	26.00	34.45	27.92	27.54	28.72
Ours($F=8, \lambda=8e-3$)	28.15	26.22	33.23	27.91	27.53	28.61
Ours($F=8, \lambda=4e-3$)	28.32	26.18	33.60	28.08	27.57	28.75
Ours($F=8, \lambda=2e-3$)	28.51	26.36	33.80	28.02	27.48	28.83
Ours($F=8, \lambda=0.7e-3$)	28.76	26.44	34.12	27.93	27.62	28.97
			SSIM↑			
Instant-NGP [29]	0.881	0.915	0.968	0.948	0.918	0.926
BiRF-F1 [36]	0.851	0.894	0.955	0.940	0.894	0.907
BiRF-F2 [36]	0.869	0.904	0.963	0.944	0.907	0.917
BiRF-F4 [36]	0.877	0.909	0.966	0.946	0.914	0.922
BiRF-F8 [36]	0.882	0.910	0.968	0.947	0.917	0.925
Ours($F=8, \lambda=8e-3$)	0.866	0.911	0.955	0.941	0.910	0.917
Ours($F=8, \lambda=4e-3$)	0.872	0.914	0.959	0.944	0.914	0.921
Ours($F=8, \lambda=2e-3$)	0.879	0.917	0.961	0.946	0.917	0.924
Ours($F=8, \lambda=0.7e-3$)	0.884	0.920	0.965	0.947	0.921	0.927
			LPIPS↓			
Instant-NGP [29]	0.233	0.161	0.057	0.087	0.151	0.138
BiRF-F1 [36]	0.223	0.159	0.063	0.080	0.159	0.137
BiRF-F2 [36]	0.198	0.144	0.052	0.075	0.139	0.122
BiRF-F4 [36]	0.187	0.136	0.046	0.072	0.128	0.114
BiRF-F8 [36]	0.180	0.133	0.043	0.072	0.121	0.109
Ours($F=8, \lambda=8e-3$)	0.243	0.159	0.081	0.087	0.154	0.145
Ours($F=8, \lambda=4e-3$)	0.234	0.154	0.075	0.084	0.147	0.139
Ours($F=8, \lambda=2e-3$)	0.222	0.149	0.071	0.083	0.143	0.134
Ours($F=8, \lambda=0.7e-3$)	0.212	0.145	0.065	0.080	0.139	0.128
			SIZE(MB)↓			
Instant-NGP [29]	45.56	45.56	45.56	45.56	45.56	45.56
MaskDWT($1e-10$) [32]	0.886	1.219	0.666	0.769	1.038	0.916
BiRF-F1 [36]	0.8	0.8	0.8	0.8	0.8	0.8
BiRF-F2 [36]	1.6	1.6	1.5	1.6	1.6	1.6
BiRF-F4 [36]	3.1	3.1	2.9	3.2	3.1	3.1
BiRF-F8 [36]	6.1	6.0	5.8	6.3	6.0	6.0
Ours($F=8, \lambda=8e-3$)	0.546	0.579	0.384	0.432	0.511	0.490
Ours($F=8, \lambda=4e-3$)	0.726	0.824	0.455	0.559	0.708	0.654
Ours($F=8, \lambda=2e-3$)	0.976	1.067	0.543	0.721	0.992	0.860
Ours($F=8, \lambda=0.7e-3$)	1.465	1.652	0.710	1.146	1.539	1.302

Table C. Detailed quantitative results of storage size against fidelity quality (PSNR, SSIM, LPIPS) of each scene on Tanks and Temples dataset. We focus on NeRF compression approaches, along with our base model Instant-NGP. For quantitative results of other approaches, please refer to BiRF [36] paper, as we do not duplicate them here.

Setting of F	*chair*	*drums*	*ficus*	*hotdog*	*lego*	*materials*	*mic*	*ship*	*Avg.*
				PSNR↑					
$F = 1$	33.17	25.31	32.38	36.38	33.57	29.79	33.60	29.86	31.74
$F = 2$	34.41	25.77	33.57	36.94	35.17	30.20	35.65	30.94	32.83
$F = 4$	35.32	25.99	34.32	37.43	36.01	30.62	36.90	31.74	33.54
$F = 8$	35.64	26.06	34.52	37.48	36.49	30.76	37.25	31.81	33.75
				SSIM↑					
$F = 1$	0.973	0.929	0.976	0.977	0.968	0.950	0.984	0.883	0.955
$F = 2$	0.980	0.936	0.981	0.980	0.977	0.955	0.990	0.902	0.963
$F = 4$	0.984	0.941	0.984	0.982	0.982	0.959	0.992	0.914	0.967
$F = 8$	0.985	0.942	0.985	0.983	0.983	0.961	0.993	0.914	0.968
				LPIPS↓					
$F = 1$	0.045	0.094	0.037	0.044	0.048	0.071	0.027	0.159	0.066
$F = 2$	0.031	0.080	0.027	0.037	0.030	0.063	0.016	0.135	0.052
$F = 4$	0.022	0.071	0.021	0.032	0.021	0.054	0.011	0.122	0.044
$F = 8$	0.020	0.070	0.019	0.032	0.019	0.052	0.010	0.121	0.043
				SIZE(MB)↓					
$F = 1$	0.827	0.816	0.806	0.922	0.838	0.822	0.802	0.901	0.842
$F = 2$	1.445	1.434	1.425	1.530	1.456	1.442	1.421	1.531	1.460
$F = 4$	2.699	2.697	2.680	2.820	2.710	2.697	2.676	2.773	2.719
$F = 8$	5.210	5.202	5.191	5.334	5.222	5.208	5.187	5.281	5.229

Table D. Detailed quantitative results of upper bounds (*i.e.* $\lambda = 0$) of our CNC model on NeRF-Synthetic dataset. In this case, no entropy constraint is applied to the embeddings, thus their size is equal to the amount of θs as each parameter consumes 1 bit. The rendering MLP is not quantized but retained in float32. Context models are excluded.

Setting of F	*Barn*	*Caterpillar*	*Family*	*Ignatius*	*Truck*	*Avg.*
			PSNR↑			
$F = 8$	28.68	26.37	34.33	27.91	27.48	28.95
			SSIM↑			
$F = 8$	0.886	0.920	0.968	0.948	0.921	0.928
			LPIPS↓			
$F = 8$	0.209	0.146	0.059	0.080	0.138	0.126
			SIZE(MB)↓			
$F = 8$	5.326	5.277	5.315	5.362	5.263	5.309

Table E. Detailed quantitative results of upper bounds (*i.e.* $\lambda = 0$) of our CNC model on Tanks and Temples dataset. In this case, no entropy constraint is applied to the embeddings, thus their size is equal to the amount of θs as each parameter consumes 1 bit. The rendering MLP is not quantized but retained in float32. Context models are excluded.

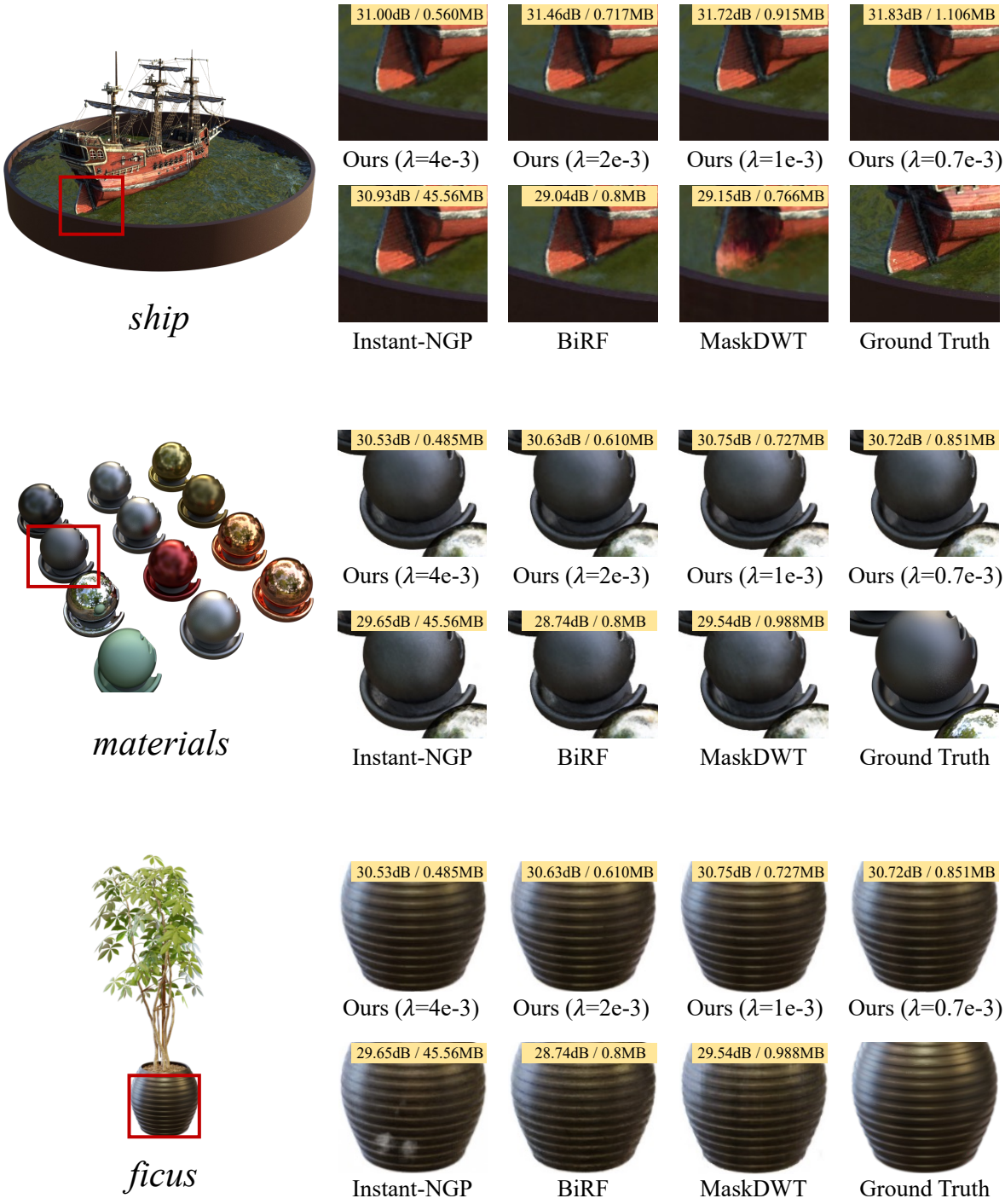

Figure A. Qualitative quality comparison of Synthetic-NeRF dataset. Quantitative results of PSNR/size are shown in the upper right.

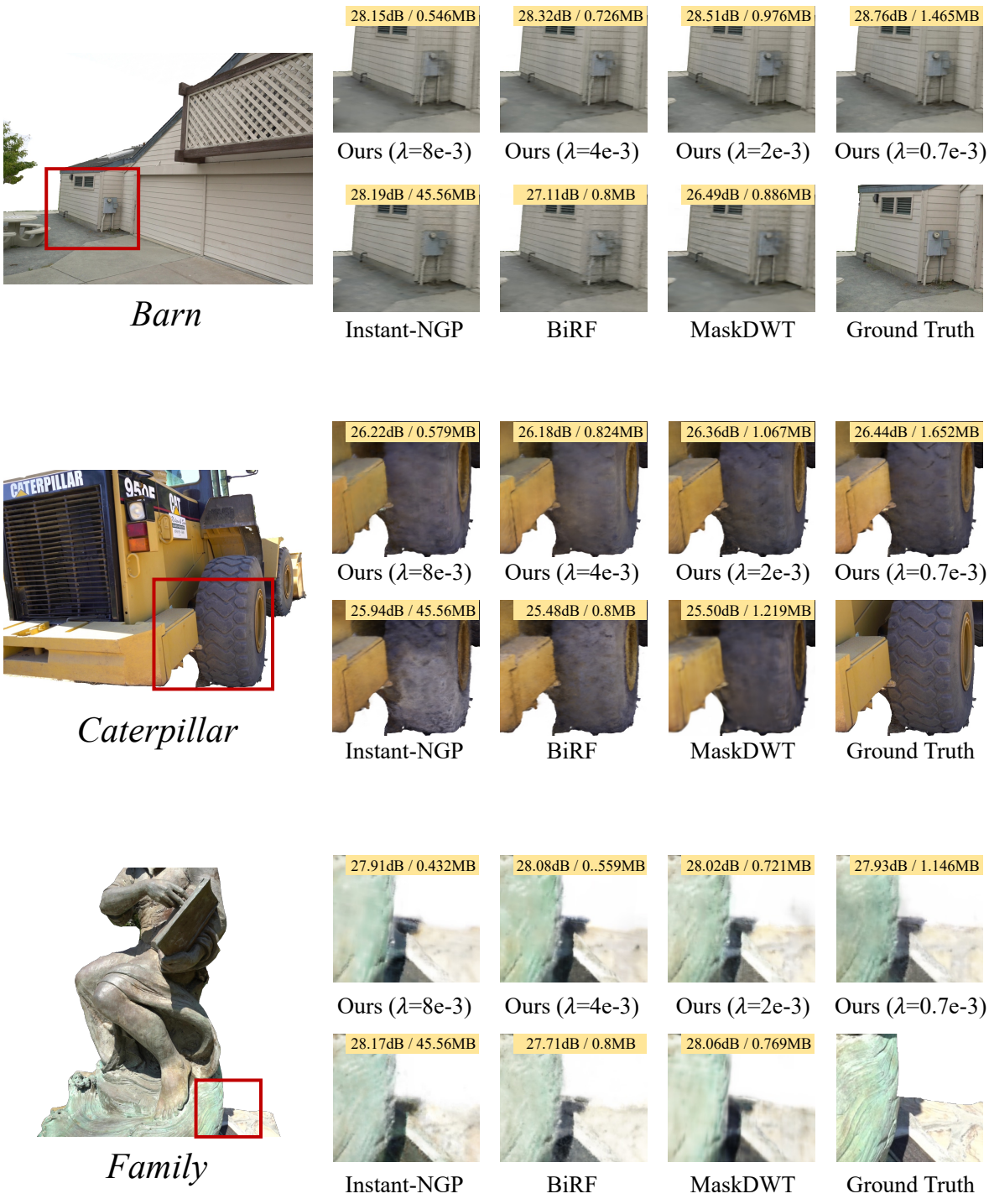

Figure B. Qualitative quality comparison of Tanks and Temples dataset. Quantitative results of PSNR/size are shown in the upper right.