Binary Radiance Fields
Supplement

A   Experimental Setting of Baselines

We have followed the experimental setting of the configurations released in the official code with minor modifications. We use reported scores in the original paper for Re:NeRF [1] and VQRF [2].

DVGO   DVGO [3, 4] consists of coarse and fine grids of density and features with a shallow MLP for color prediction. We set the number of voxels in the coarse grids to 100^3 and in the fine grids to 160^3 for both the density and feature grids. The dimension of features stored in each voxel of the coarse and fine density grid is 1, while in the coarse and fine feature grid, it is 3 and 12, respectively. The shallow MLP has two 128-channel hidden layers with ReLU activation.

Plenoxels   Plenoxels [5] consists of sparse voxel grids having density values and spherical harmonics coefficients. We set the number of voxels from 256^3 to 512^3 during training. The dimension of features for density voxel is 1, and for spherical harmonics, the number of coefficients is 27 since we use spherical harmonics of degree 3.

TensoRF   TensoRF-VM [6] consists of vector and matrix components with a shallow MLP for color prediction. We set the number of voxels from 128^3 to 300^3 during training. The dimension of features in each cell of vector and matrix components is 48 for density and 144 for color. Similarly, TensoRF-CP [6] consists of only vector components with a shallow MLP for color prediction. We also set the number of voxels from 128^3 to 300^3 during training. The dimension of features in each cell of vector components is 96 for density and 288 for color. The shallow MLP has two 128-channel hidden layers with ReLU activation.

CCNeRF   CCNeRF-HY [7] uses a hybrid representation that combines CANDECOMP/PARAFAC (CP) decomposition and Triple Plane (TP) decomposition. We set the number of voxels from 128^3 to 300^3 during training. CCNeRF-HY has 96 vector components for expressing density, and it has 96 vector components and 64 matrix components for expressing color. Due to the rank-residual learning strategy, we can efficiently reduce the model size by adopting rank truncation that excludes the matrix components for color. Since this model contains only vector components for CP decomposition, we denote this truncated CCNeRF-HY model as CCNeRF-CP in this paper. Note that it differs from the experimental setting of CCNeRF-CP in the original paper, which exclusively employs vector components for optimization like TensoRF-CP.

Instant-NGP   Instant-NGP [8] consists of a multi-resolution hash grid with two shallow MLPs each for density and color prediction. We set the 16 resolutions of the feature grid from 16 to 2048, the size of hash table 2^{19}, and the feature dimension 2. The density MLP has one 64-channel hidden layer with ReLU activation, and the color MLP has two 64-channel hidden layers with ReLU activation.

K-Planes   K-Planes [9] consists of multi-resolution planes with resolutions of 128, 256, and 512. Each cell in the planes has a 32-dim feature. K-Planes-hybrid model decodes the features using two shallow MLPs, similar to Instant-NGP [8]. The density MLP has one 64-channel hidden layer with ReLU activation, and the color MLP has two 64-channel hidden layers with ReLU activation. In the K-Planes-explicit model, the features are decoded using an explicit linear decoder.

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

NeRF NeRF [10] consists of a coarse implicit network and a fine implicit network. Each network has eight 256-channel hidden layers with ReLU activation, one additional 256-channel layer with ReLU activation, and one 128-channel layer with ReLU activation.

Mip-NeRF Mip-NeRF [11] consists of an implicit network. The network has eight 256-channel hidden layers with shifted Softplus activation, one additional 256-channel layer with shifted Softplus activation, and one 128-channel layer with shifted Softplus activation.

B Additional Results

B.1 Quantitative Results

Table 1 summarizes the average scores of quantitative results on the Synthetic-NeRF, Sythetic-NSVF, and Tanks and Temples datasets. Our model successfully outperforms data-structure- and compression-based models for all datasets with highly compact storage sizes. We provide scores of DWT-TensoRF [12] on the Tanks and Temples dataset reported in the original paper since several scenes require large memory of more than 48 GB. The failure of *Lifestyle* scene causes the noticeably worse performance of K-Planes [9] on the Synthetic-NSVF dataset; see Table. 8 & 9.

Table 1: Quantitative results of model size and reconstruction quality on the Synthetic-NeRF dataset, Synthetic-NSVF dataset, and Tanks and Temples dataset. We divide the baselines into data-structure-based models (upper rows) and compression-based (middle rows) models. † denotes that we provide scores reported in the original paper. We denote our small model as Ours-S and our base model as Ours-B. All our models are scaled by the feature dimension F; for instance, Ours-B2 denotes a base variant with feature dimension $F = 2$. We highlight the best score and second-best score .

Method	Synthetic-NeRF			Synthetic-NSVF			Tanks and Temples		
	Size↓ (MB)	PSNR↑	SSIM↑	Size↓ (MB)	PSNR↑	SSIM↑	Size↓ (MB)	PSNR↑	SSIM↑
DVGO [3, 4]	655.3	31.96	0.956	653.7	35.01	0.975	652.8	28.36	0.911
Plenoxels [5]	778.2	31.72	0.958	803.6	34.15	0.978	892.8	26.87	0.912
TensoRF-CP [6]	3.9	31.66	0.950	4.1	34.57	0.971	4.0	27.93	0.901
TensoRF-VM [6]	72.4	33.24	0.963	71.6	36.73	0.982	72.5	28.65	0.922
CCNeRF-CP [7]	1.5	29.30	0.937	1.6	31.99	0.950	1.6	26.25	0.873
CCNeRF-HY [7]	72.1	31.22	0.947	71.7	34.57	0.974	73.7	27.62	0.901
Instant-NGP [8]	24.6	32.66	0.958	24.6	33.01	0.973	24.5	27.72	0.926
K-Planes-explicit [9]	384.7	32.19	0.960	383.9	31.20	0.952	383.9	28.01	0.921
K-Planes-hybrid [9]	383.9	32.31	0.961	383.9	31.18	0.951	384.7	28.01	0.921
Re:DVGO-High† [1]	2.0	31.08	0.944	2.5	34.90	0.969	1.6	27.90	0.894
Re:Plenoxels-High† [1]	54.7	30.97	0.944	-	-	-	85.5	27.34	0.896
Re:TensoRF-High† [1]	7.9	32.81	0.956	8.5	36.14	0.978	6.7	28.24	0.907
VQ-DVGO† [2]	1.4	31.77	0.954	1.3	34.72	0.974	1.4	28.26	0.909
VQ-Plenoxels† [2]	13.7	31.53	0.956	11.9	33.91	0.976	14.3	26.73	0.908
VQ-TensoRF† [2]	3.6	32.86	0.960	4.1	36.16	0.980	3.3	28.20	0.913
DWT-TensoRF-S [12]	0.9	31.80	0.953	0.9	34.60	0.972	0.9	27.77	-
DWT-TensoRF-M [12]	1.2	31.91	0.954	1.3	34.88	0.974	1.3	27.83	-
DWT-TensoRF-L [12]	1.7	32.05	0.955	2.0	35.02	0.975	1.9	27.92	-
NeRF [10]	4.6	31.69	0.951	4.6	34.46	0.967	4.6	27.58	0.902
Mip-NeRF [11]	2.3	32.20	0.960	2.3	35.33	0.971	2.3	27.77	0.901
Ours-S2	0.5	32.03	0.954	0.5	34.48	0.972	0.5	28.20	0.910
Ours-S4	0.9	32.99	0.960	0.9	35.59	0.978	1.0	28.44	0.914
Ours-S8	1.7	33.46	0.963	1.8	36.28	0.981	1.9	28.61	0.920
Ours-B2	1.4	32.64	0.958	1.5	35.70	0.978	1.6	28.49	0.917
Ours-B4	2.8	33.26	0.962	2.9	36.58	0.982	3.1	28.68	0.922
Ours-B8	5.8	33.59	0.964	5.9	37.09	0.984	6.0	28.72	0.925

B.2 Training Time

We also evaluate the training time for each method. Fig. 1 shows that our method has a comparable convergence speed to SOTA fast radiance field models.

2
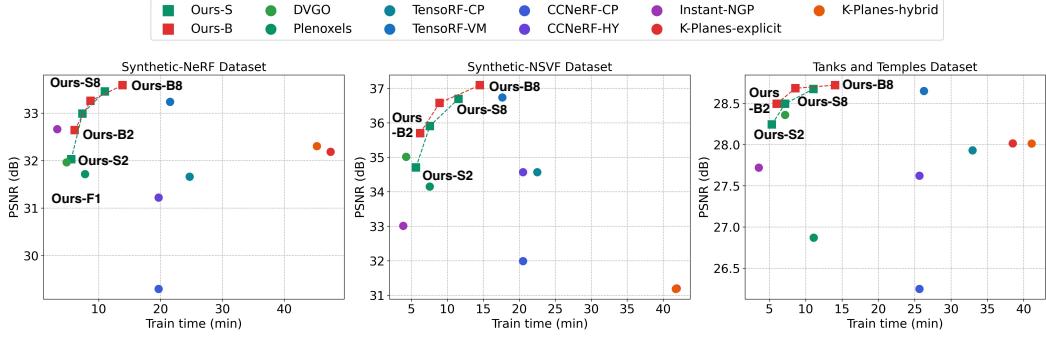
Figure 1: Illustration of the training time and reconstruction quality (PSNR) of each method on the Synthetic-NeRF dataset, Synthetic-NSVF dataset, and Tanks and Temples dataset.

B.3 Inference Speed

Additionally, we compare the inference speed of our model with baselines. Fig. 2 demonstrates that our method has a comparable rendering speed to SOTA fast radiance field models.
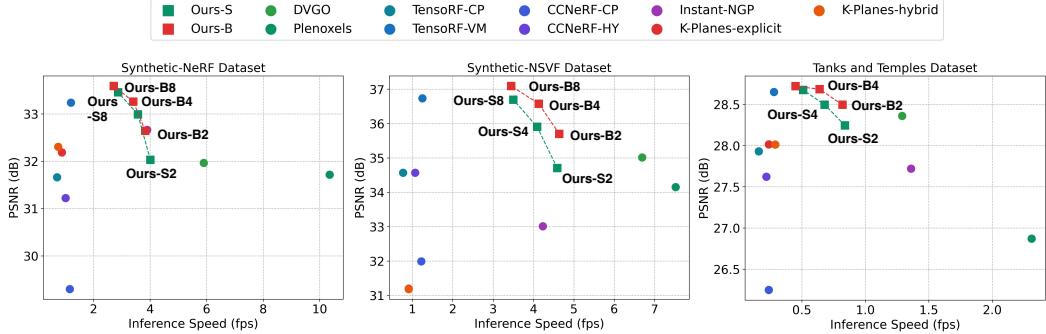
Figure 2: Illustration of the inference speed (fps) and reconstruction quality (PSNR) of each method on the Synthetic-NeRF dataset, Synthetic-NSVF dataset, and Tanks and Temples dataset.

B.4 Ablation Study

Feature grid design Table 2 shows the scores of the ablation study on the feature grid design. With a similar number of parameters, our 2D-3D hybrid feature grid design has superior performance on various datasets compared to tri-plane and voxel representation.

Table 2: Ablation study on the feature grid design. The experiments are conducted on the Synthetic-NeRF datatset, Synthetic-NSVF datatset, and Tanks and Temples datatset.

Design	# Params	Synthetic-NeRF		Synthetic-NSVF		Tanks and Temples	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Tri-plane (only 2D)	13.7 M	31.44	0.949	34.33	0.971	27.97	0.904
Voxel (only 3D)	13.6 M	32.35	0.956	35.10	0.975	28.35	0.912
Hybrid (2D + 3D)	13.2 M	**32.64**	**0.958**	**35.70**	**0.978**	**28.49**	**0.917**

Hash table size Table 3 reports the scores of the ablation study on the hash table size. We can further reduce the storage size of our model by using a smaller size of the hash table, but it might lead to a decline in the reconstruction quality. Also, we can achieve higher performance by increasing the hash table size, leading to a larger storage size. In this work, we use $\{2^{15}, 2^{17}\}$ and $\{2^{17}, 2^{19}\}$ size of hash tables as our small and base configuration, respectively.

3

Table 3: Ablation study on the hash table size. The experiments are conducted on the Synthetic-NeRF datatset.

$\{T_{2D}, T_{3D}\}$	$F = 1$		$F = 2$		$F = 4$		$F = 8$	
	Size (MB)↓	PSNR↑	Size (MB)↓	PSNR↑	Size (MB)↓	PSNR↑	Size (MB)↓	PSNR↑
$\{2^{13}, 2^{15}\}$	0.12	29.20	0.18	30.66	0.31	32.00	0.58	32.90
$\{2^{15}, 2^{17}\}$	0.27	30.66	0.46	32.03	0.87	32.99	1.73	33.46
$\{2^{17}, 2^{19}\}$	0.72	31.53	1.41	32.64	2.83	33.26	5.76	33.59
$\{2^{19}, 2^{21}\}$	1.94	31.61	3.99	32.71	7.84	33.20	16.61	33.52
$\{2^{21}, 2^{23}\}$	7.05	31.93	14.70	32.89	29.65	33.31	61.35	33.55

B.5 Dynamic Scene Reconstruction

Table 4 and 5 present full scores and Fig. 3 and 4 show the qualitative results of the dynamic scene reconstruction. We apply binary feature encoding to TiNeuVox [13] and compare our model with the original model.
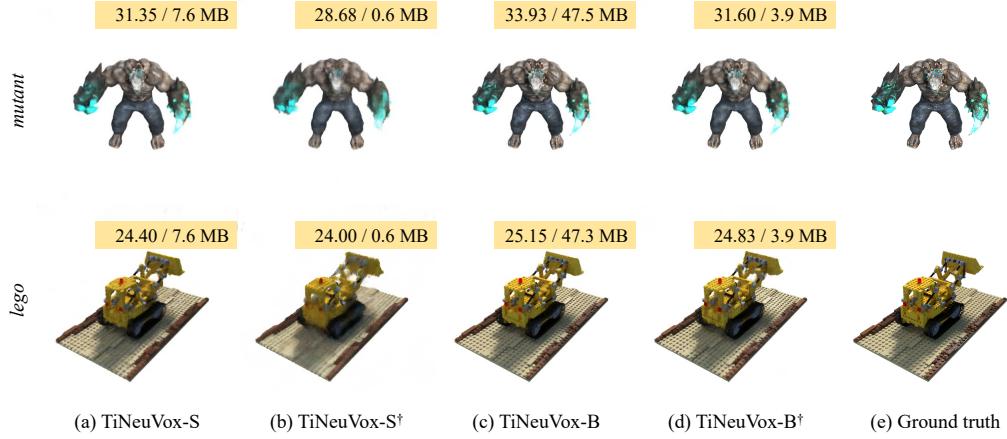
Figure 3: Qualitative results on each scene of the D-NeRF dataset. † denotes that binary feature encoding is applied. PSNR and storage size are shown on the upper right side in each subfigure.
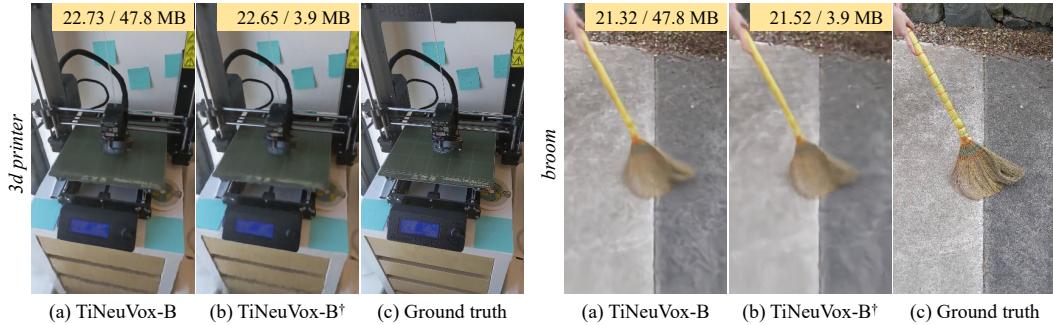
Figure 4: Qualitative results on each scene of the HyperNeRF dataset. † denotes that binary feature encoding is applied. PSNR and storage size are shown on the upper right side in each subfigure.

B.6 Qualitative Results

Fig. 5, 6, and 7 demonstrate the qualitative results on each scene of the Synthetic-NeRF dataset, Sythetic-NSVF dataset, and Tanks and Temples dataset, respectively. We report the rendering images of our method with the cropped area.

4

Table 4: Quantitative results of storage size and reconstruction quality (PSNR, SSIM, LPIPS) on each scene of the D-NeRF dataset. † denotes that binary feature encoding is applied.

Method	bouncingballs	hellwarrior	hook	jumpingjacks	lego	mutant	standup	trex	Avg.
PSNR↑									
TiNeuVox-S	39.34	27.07	29.60	32.95	24.40	31.35	33.50	30.04	31.03
TiNeuVox-S†	36.26	24.54	26.87	30.08	24.00	28.68	30.36	28.62	28.68
TiNeuVox-B	41.09	28.27	31.86	34.80	25.15	33.93	36.12	32.97	33.02
TiNeuVox-B†	39.18	26.86	29.47	33.24	24.83	31.60	33.79	30.86	31.23
SSIM↑									
TiNeuVox-S	0.988	0.954	0.954	0.975	0.886	0.961	0.977	0.958	0.957
TiNeuVox-S†	0.981	0.935	0.928	0.960	0.853	0.942	0.963	0.946	0.938
TiNeuVox-B	0.992	0.965	0.972	0.983	0.921	0.977	0.986	0.979	0.972
TiNeuVox-B†	0.989	0.955	0.954	0.976	0.906	0.964	0.978	0.964	0.961
LPIPS↓									
TiNeuVox-S	0.054	0.080	0.068	0.042	0.117	0.049	0.030	0.059	0.062
TiNeuVox-S†	0.092	0.114	0.097	0.068	0.164	0.069	0.052	0.076	0.091
TiNeuVox-B	0.040	0.066	0.045	0.033	0.070	0.030	0.020	0.031	0.042
TiNeuVox-B†	0.054	0.084	0.071	0.046	0.098	0.047	0.033	0.052	0.061
Size (MB)↓									
TiNeuVox-S	7.6	7.6	7.6	7.6	7.7	7.6	7.6	7.6	7.6
TiNeuVox-S†	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
TiNeuVox-B	47.5	47.5	47.5	47.5	47.3	47.5	47.5	47.5	47.5
TiNeuVox-B†	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9

Table 5: Quantitative results of storage size and reconstruction quality (PSNR, MS-SSIM) on each scene of the HyperNeRF dataset. † denotes that binary feature encoding is applied.

Method	3d printer	broom	chicken	peel-banana	Avg.
PSNR↑					
TiNeuVox-B	22.73	21.32	28.28	24.47	24.20
TiNeuVox-B†	22.65	21.52	27.75	23.54	23.87
MS-SSIM↑					
TiNeuVox-B	0.835	0.694	0.938	0.873	0.835
TiNeuVox-B†	0.835	0.694	0.938	0.836	0.826
Size (MB)↓					
TiNeuVox-B	47.8	47.8	47.8	47.8	47.8
TiNeuVox-B†	3.9	3.9	3.9	3.9	3.9

B.7 Scene-wise Results

Table 6, 7, 8, 9, 10, and 11 demonstrate the quantitative results on each scene of the Synthetic-NeRF dataset, Sythetic-NSVF dataset, and Tanks and Temples dataset, respectively. We measured the storage size of each method and evaluated the reconstruction quality (PSNR, SSIM, LPIPS) on various datasets.

B.8 Comparison with 3D Gaussian Splatting

Recently, 3D-GS [14] has demonstrated impressive performance in 3D reconstruction with real-time rendering speed (>100 fps). Although our model also has a fast rendering speed, it does not achieve real-time levels like 3D-GS. Nevertheless, 3D-GS requires a considerable storage size (>500 MB) to represent the 3D scene; even its compact representation [15] also needs tens of megabytes (>20 MB). In contrast, our approach can represent the 3D scene within only 1 MB, while 3D-GS still requires a substantial storage size.

5

Figure 5: Qualitative results on each scene of the Synthetic-NeRF dataset. PSNR and storage size are shown on the upper right side in each subfigure.

6
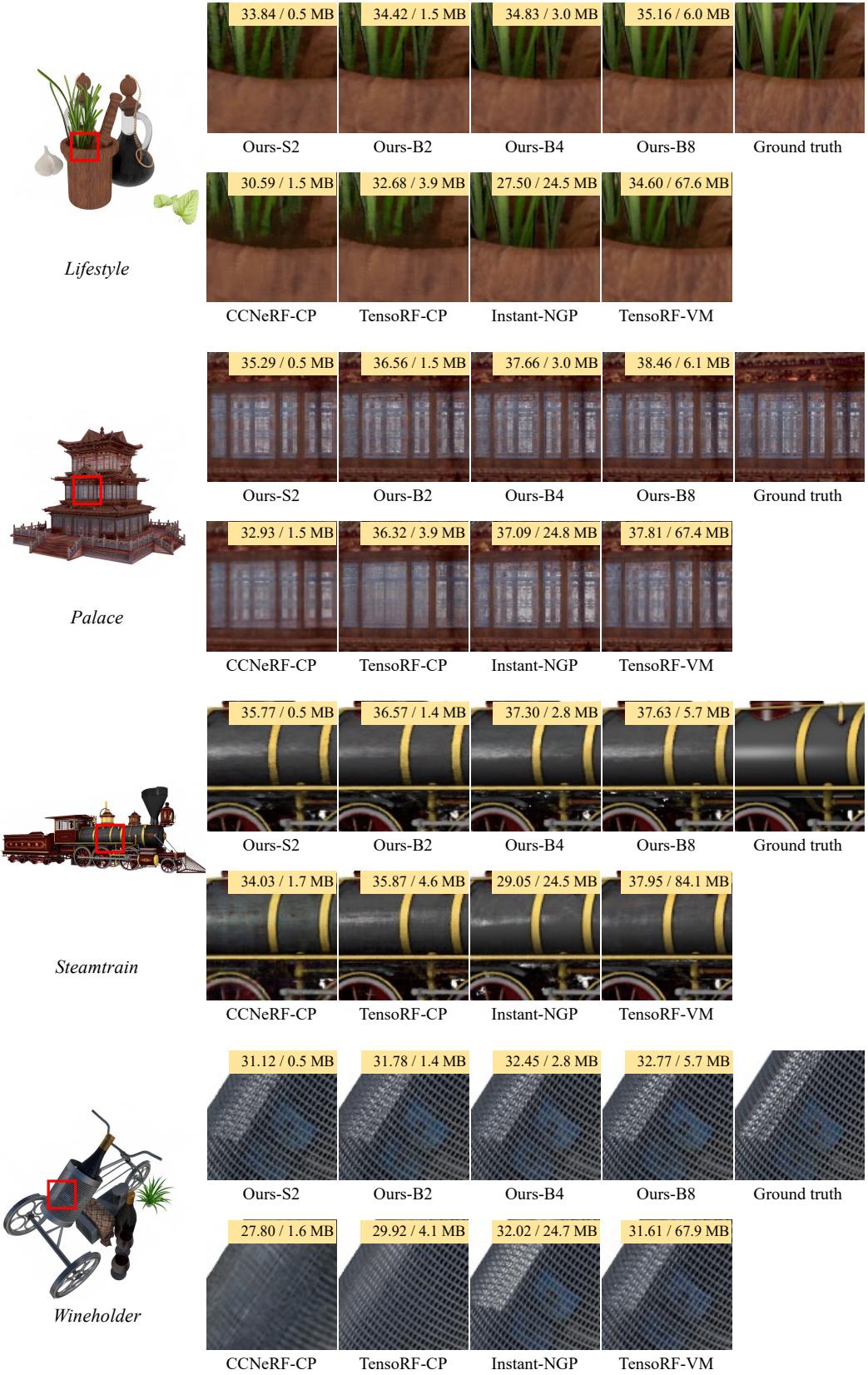
Figure 6: Qualitative results on each scene of the Synthetic-NSVF dataset. PSNR and storage size are shown on the upper right side in each subfigure.

7
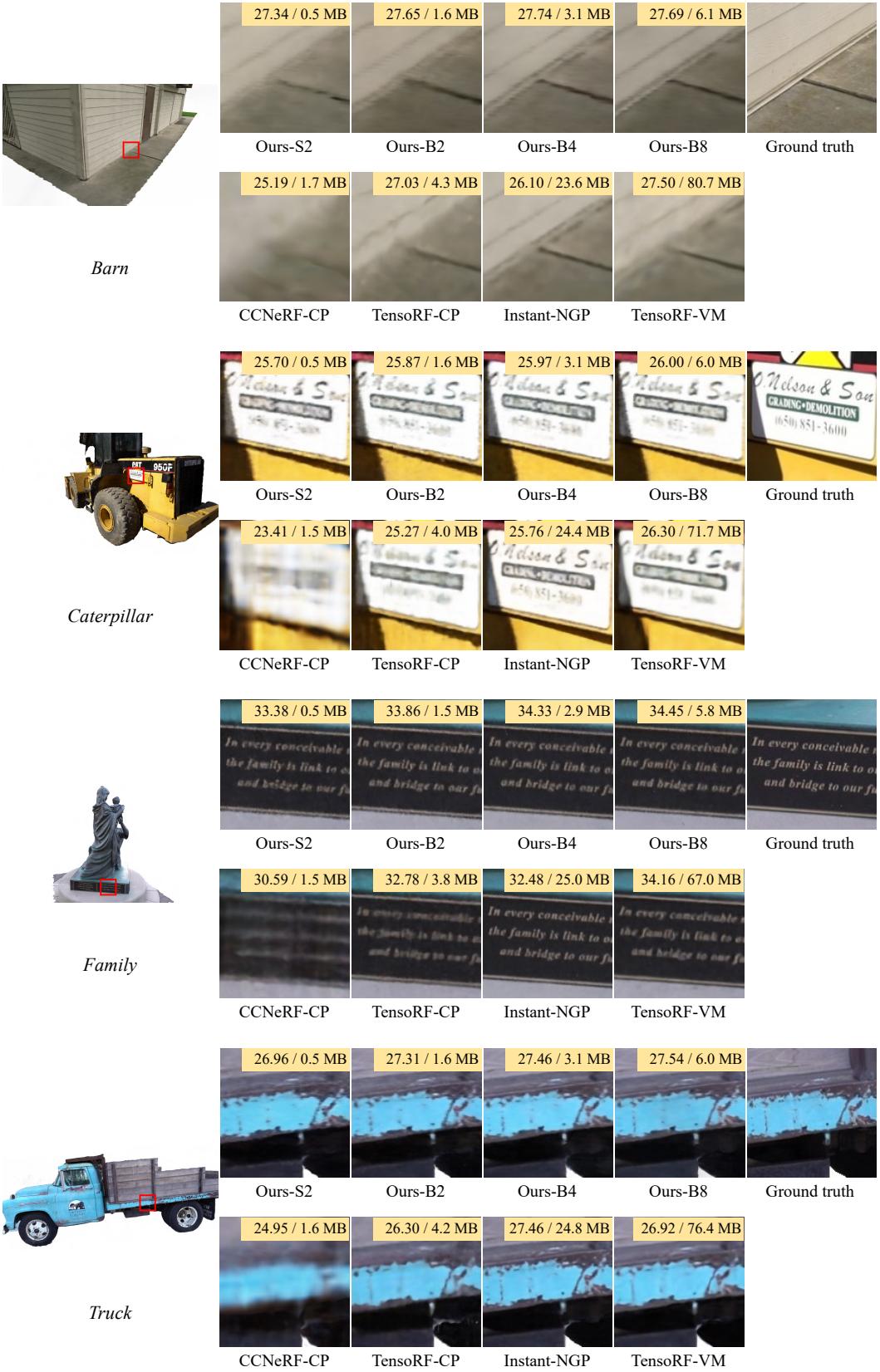
Figure 7: Qualitative results on each scene of the Tanks and Temples dataset. PSNR and storage size are shown on the upper right side in each subfigure.

8

Table 6: Quantitative results of storage size and reconstruction quality (PSNR) on each scene of the Synthetic-NeRF dataset. We highlight the best score and second-best score.

Method	chair	drums	ficus	hotdog	lego	materials	mic	ship	Avg.
				Size (MB)↓					
DVGO [3, 4]	4.3	4.1	4.6	5.2	4.7	5.2	4.0	6.2	4.8
Plenoxels [5]	7.1	6.9	6.1	8.5	8.0	7.5	5.9	12.3	7.8
TensoRF-CP [6]	22.6	21.1	26.5	23.3	23.9	33.7	21.4	25.1	24.7
TensoRF-VM [6]	24.2	13.4	25.8	18.0	25.4	22.8	23.1	19.6	21.5
CCNeRF-CP [7]	18.3	17.5	20.3	19.9	18.9	23.7	19.0	19.9	19.7
CCNeRF-HY [7]	18.3	17.5	20.3	19.9	18.9	23.7	19.0	19.9	19.7
Instant-NGP [8]	3.6	3.1	2.8	3.5	3.7	2.9	3.3	3.2	3.3
K-Planes-explicit [9]	47.2	48.0	48.0	46.8	47.0	47.9	47.7	47.4	47.5
K-Planes-hybrid [9]	45.5	45.6	45.9	44.3	44.3	45.9	45.4	45.4	45.3
NeRF [10]	4.6	4.6	4.6	4.6	4.6	4.6	4.6	4.6	4.6
Mip-NeRF [11]	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3
Ours-S2	0.4	0.5	0.4	0.5	0.5	0.5	0.4	0.5	0.5
Ours-S4	0.8	0.9	0.8	0.9	0.9	0.9	0.8	1.0	0.9
Ours-S8	1.7	1.7	1.7	1.8	1.8	1.8	1.6	1.8	1.7
Ours-B2	1.3	1.5	1.4	1.4	1.4	1.4	1.3	1.5	1.4
Ours-B4	2.7	2.9	2.8	2.8	2.8	2.8	2.7	3.0	2.8
Ours-B8	5.6	5.7	5.8	5.8	5.8	5.7	5.6	6.0	5.8
				PSNR↑					
DVGO [3, 4]	34.26	25.45	32.60	36.78	34.75	29.58	33.22	29.07	31.96
Plenoxels [5]	34.00	25.36	31.83	36.43	34.11	29.13	33.23	29.63	31.72
TensoRF-CP [6]	34.02	25.26	30.70	36.25	34.28	30.13	33.81	28.85	31.66
TensoRF-VM [6]	35.84	25.95	34.06	37.62	36.68	30.11	34.92	30.76	33.24
CCNeRF-CP [7]	30.41	24.13	29.08	33.71	31.04	27.74	31.30	26.96	29.30
CCNeRF-HY [7]	34.28	24.84	30.07	36.10	33.64	28.82	33.58	28.45	31.22
Instant-NGP [8]	34.90	25.78	32.35	36.82	35.57	29.56	35.77	30.55	32.66
K-Planes-explicit [9]	34.85	25.64	31.20	36.66	35.21	29.46	34.02	30.44	32.19
K-Planes-hybrid [9]	34.95	25.67	31.37	36.71	35.69	29.31	34.03	30.71	32.31
NeRF [10]	33.73	25.29	31.26	36.67	33.40	29.95	33.91	29.28	31.69
Mip-NeRF [11]	33.92	25.55	32.55	36.79	34.37	30.19	34.56	29.69	32.20
Ours-S2	33.73	25.77	33.13	36.15	33.91	29.15	35.06	29.33	32.03
Ours-S4	35.02	26.07	34.19	37.03	35.31	29.69	36.65	29.95	32.99
Ours-S8	35.83	26.27	34.64	37.30	36.20	29.93	37.20	30.29	33.46
Ours-B2	34.75	25.59	33.91	36.59	35.06	29.49	36.01	29.74	32.64
Ours-B4	35.66	25.84	34.42	37.13	36.02	29.80	36.91	30.30	33.26
Ours-B8	36.17	26.05	34.71	37.51	36.48	30.09	37.44	30.27	33.59

9

Table 7: Quantitative results of reconstruction quality (SSIM, LPIPS) on each scene of the Synthetic-NeRF dataset. We highlight the best score and second-best score .

Method	chair	drums	ficus	hotdog	lego	materials	mic	ship	Avg.
				SSIM↑					
DVGO [3, 4]	0.977	0.930	0.978	0.980	0.977	0.950	0.983	0.878	0.956
Plenoxels [5]	0.977	0.933	0.976	0.980	0.975	0.949	0.985	0.890	0.958
TensoRF-CP [6]	0.976	0.922	0.965	0.975	0.972	0.950	0.983	0.856	0.950
TensoRF-VM [6]	0.985	0.936	0.983	0.983	0.983	0.952	0.988	0.895	0.963
CCNeRF-CP [7]	0.941	0.902	0.952	0.956	0.944	0.927	0.969	0.826	0.927
CCNeRF-HY [7]	0.977	0.920	0.962	0.975	0.969	0.934	0.984	0.855	0.947
Instant-NGP [8]	0.979	0.931	0.968	0.984	0.979	0.946	0.984	0.895	0.958
K-Planes-explicit [9]	0.981	0.936	0.974	0.981	0.979	0.949	0.988	0.892	0.960
K-Planes-hybrid [9]	0.983	0.938	0.975	0.981	0.982	0.948	0.988	0.896	0.961
NeRF [10]	0.972	0.928	0.970	0.977	0.965	0.952	0.984	0.864	0.951
Mip-NeRF [11]	0.973	0.931	0.976	0.978	0.970	0.955	0.986	0.870	0.955
Ours-S2	0.974	0.930	0.977	0.975	0.968	0.939	0.987	0.880	0.954
Ours-S4	0.981	0.936	0.982	0.979	0.977	0.946	0.990	0.890	0.960
Ours-S8	0.984	0.939	0.984	0.980	0.981	0.950	0.992	0.896	0.963
Ours-B2	0.980	0.930	0.981	0.978	0.976	0.943	0.989	0.888	0.958
Ours-B4	0.984	0.934	0.983	0.980	0.980	0.948	0.991	0.895	0.962
Ours-B8	0.986	0.937	0.984	0.981	0.982	0.951	0.992	0.897	0.964
				LPIPS↓					
DVGO [3, 4]	0.026	0.079	0.025	0.033	0.026	0.058	0.017	0.160	0.053
Plenoxels [5]	0.030	0.067	0.026	0.038	0.028	0.057	0.015	0.134	0.050
TensoRF-CP [6]	0.041	0.107	0.059	0.052	0.036	0.069	0.033	0.194	0.074
TensoRF-VM [6]	0.022	0.073	0.022	0.030	0.018	0.058	0.014	0.138	0.047
CCNeRF-CP [7]	0.081	0.135	0.069	0.090	0.075	0.088	0.052	0.239	0.104
CCNeRF-HY [7]	0.035	0.105	0.054	0.055	0.038	0.080	0.031	0.189	0.073
Instant-NGP [8]	0.018	0.068	0.029	0.027	0.017	0.053	0.015	0.123	0.044
K-Planes-explicit [9]	0.020	0.057	0.028	0.027	0.019	0.051	0.010	0.132	0.043
K-Planes-hybrid [9]	0.019	0.058	0.029	0.026	0.019	0.051	0.012	0.123	0.042
NeRF [10]	0.041	0.087	0.035	0.044	0.046	0.060	0.022	0.184	0.065
Mip-NeRF [11]	0.040	0.085	0.030	0.043	0.041	0.057	0.020	0.177	0.062
Ours-S2	0.034	0.076	0.029	0.041	0.036	0.071	0.019	0.141	0.056
Ours-S4	0.022	0.066	0.022	0.034	0.022	0.061	0.013	0.127	0.046
Ours-S8	0.018	0.061	0.019	0.031	0.017	0.055	0.010	0.118	0.041
Ours-B2	0.024	0.073	0.024	0.036	0.025	0.064	0.016	0.127	0.049
Ours-B4	0.019	0.066	0.020	0.032	0.017	0.057	0.012	0.117	0.043
Ours-B8	0.016	0.063	0.018	0.028	0.015	0.051	0.009	0.112	0.039

10

Table 8: Quantitative results of storage size and reconstruction quality (PSNR) on each scene of the Synthetic-NSVF dataset. We highlight the best score and second-best score .

Method	Bike	Lifestyle	Palace	Robot	Spaceship	Steamtrain	Toad	Wineholder	Avg.
Size (MB)↓									
DVGO [3, 4]	4.0	4.1	4.7	4.0	4.9	4.3	4.2	4.0	4.3
Plenoxels [5]	6.1	8.1	10.0	5.8	6.2	6.5	11.1	6.6	7.6
TensoRF-CP [6]	22.2	21.5	21.8	18.5	29.3	22.9	24.7	19.0	22.5
TensoRF-VM [6]	18.6	15.1	18.5	13.5	23.7	18.3	19.2	14.3	17.6
CCNeRF-CP [7]	18.9	17.2	19.2	18.9	22.8	22.5	23.0	21.7	20.5
CCNeRF-HY [7]	18.9	17.2	19.2	18.9	22.8	22.5	23.0	21.7	20.5
Instant-NGP [8]	3.2	3.1	3.8	3.8	3.0	3.5	7.4	3.0	3.9
K-Planes-explicit [9]	40.7	43.6	42.2	41.5	41.2	41.6	41.5	42.5	41.9
K-Planes-hybrid [9]	40.2	43.2	41.9	41.9	41.2	41.8	41.2	42.7	41.8
NeRF [10]	4.6	4.6	4.6	4.6	4.6	4.6	4.6	4.6	4.6
Mip-NeRF [11]	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3
Ours-S2	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Ours-S4	0.9	0.9	1.0	0.9	1.0	0.9	0.9	0.9	0.9
Ours-S8	1.8	1.8	1.9	1.8	1.9	1.8	1.8	1.7	1.8
Ours-B2	1.5	1.5	1.5	1.5	1.5	1.4	1.5	1.4	1.5
Ours-B4	3.0	3.0	3.0	2.9	3.1	2.8	2.9	2.8	2.9
Ours-B8	5.9	6.0	6.1	5.8	6.1	5.7	5.8	5.7	5.9
PSNR↑									
DVGO [3, 4]	38.15	33.80	34.39	36.31	37.52	36.54	33.10	30.28	35.01
Plenoxels [5]	37.83	31.05	35.31	35.92	34.41	34.29	34.35	30.02	34.15
TensoRF-CP [6]	36.98	32.68	36.32	36.09	37.20	35.87	31.47	29.92	34.57
TensoRF-VM [6]	39.45	34.60	37.81	38.57	38.76	37.95	35.11	31.61	36.73
CCNeRF-CP [7]	34.96	30.59	32.93	32.79	33.96	34.03	28.86	27.80	31.99
CCNeRF-HY [7]	37.25	32.42	36.23	36.30	35.95	35.63	33.69	29.08	34.57
Instant-NGP [8]	37.98	27.50	37.09	37.04	27.12	29.05	36.29	32.02	33.01
K-Planes-explicit [9]	36.96	18.30	37.91	35.78	20.06	35.42	34.69	30.45	31.20
K-Planes-hybrid [9]	37.08	18.07	37.91	35.80	20.06	35.41	34.70	30.42	31.18
NeRF [10]	35.72	33.45	34.74	33.52	36.52	40.89	31.60	29.23	34.46
Mip-NeRF [11]	37.33	34.08	35.03	35.16	37.21	41.59	32.20	30.04	35.33
Ours-S2	37.12	33.84	35.29	35.37	36.93	35.77	32.18	31.12	34.70
Ours-S4	38.35	34.47	36.73	37.06	37.94	36.96	33.79	31.94	35.91
Ours-S8	39.14	34.89	37.80	38.16	38.48	37.53	35.09	32.47	36.69
Ours-B2	38.11	34.42	36.56	36.77	37.42	36.57	33.95	31.78	35.70
Ours-B4	38.92	34.83	37.66	37.96	38.19	37.30	35.29	32.45	36.58
Ours-B8	39.39	35.16	38.46	38.52	38.56	37.63	36.22	32.77	37.09

11

Table 9: Quantitative results of and reconstruction quality (SSIM, LPIPS) on each scene of the Synthetic-NSVF dataset. We highlight the best score and second-best score .

Method	Bike	Lifestyle	Palace	Robot	Spaceship	Steamtrain	Toad	Wineholder	Avg.
				SSIM↑					
DVGO [3, 4]	0.991	0.965	0.962	0.992	0.987	0.989	0.966	0.950	0.975
Plenoxels [5]	0.992	0.967	0.974	0.991	0.982	0.983	0.976	0.959	0.978
TensoRF-CP [6]	0.988	0.953	0.972	0.990	0.984	0.986	0.952	0.947	0.971
TensoRF-VM [6]	0.993	0.969	0.981	0.994	0.988	0.991	0.979	0.963	0.982
CCNeRF-CP [7]	0.980	0.936	0.938	0.978	0.973	0.975	0.904	0.920	0.950
CCNeRF-HY [7]	0.988	0.956	0.973	0.990	0.981	0.984	0.974	0.946	0.974
Instant-NGP [8]	0.986	0.955	0.976	0.990	0.966	0.976	0.980	0.959	0.973
K-Planes-explicit [9]	0.990	0.852	0.982	0.991	0.873	0.987	0.978	0.960	0.952
K-Planes-hybrid [9]	0.990	0.850	0.982	0.991	0.873	0.987	0.978	0.960	0.951
NeRF [10]	0.986	0.958	0.958	0.984	0.986	0.984	0.945	0.936	0.967
Mip-NeRF [11]	0.988	0.961	0.962	0.988	0.987	0.987	0.954	0.946	0.971
Ours-S2	0.988	0.960	0.964	0.989	0.984	0.984	0.958	0.955	0.973
Ours-S4	0.991	0.965	0.974	0.993	0.987	0.988	0.970	0.963	0.979
Ours-S8	0.992	0.968	0.979	0.994	0.988	0.990	0.978	0.967	0.982
Ours-B2	0.990	0.965	0.973	0.992	0.985	0.987	0.972	0.961	0.978
Ours-B4	0.992	0.968	0.979	0.994	0.987	0.989	0.979	0.966	0.982
Ours-B8	0.992	0.970	0.983	0.994	0.988	0.990	0.983	0.969	0.984
				LPIPS↓					
DVGO [3, 4]	0.011	0.053	0.043	0.013	0.020	0.018	0.045	0.055	0.032
Plenoxels [5]	0.011	0.047	0.026	0.013	0.025	0.028	0.031	0.046	0.028
TensoRF-CP [6]	0.022	0.080	0.029	0.016	0.028	0.031	0.065	0.080	0.044
TensoRF-VM [6]	0.010	0.046	0.020	0.010	0.020	0.017	0.028	0.047	0.025
CCNeRF-CP [7]	0.032	0.101	0.069	0.032	0.039	0.049	0.101	0.105	0.066
CCNeRF-HY [7]	0.020	0.073	0.029	0.015	0.031	0.031	0.037	0.075	0.039
Instant-NGP [8]	0.013	0.051	0.018	0.009	0.039	0.026	0.020	0.038	0.027
K-Planes-explicit [9]	0.011	0.223	0.021	0.012	0.175	0.017	0.027	0.041	0.066
K-Planes-hybrid [9]	0.012	0.219	0.022	0.012	0.177	0.016	0.027	0.041	0.066
NeRF [10]	0.021	0.067	0.045	0.023	0.023	0.031	0.066	0.075	0.044
Mip-NeRF [11]	0.019	0.063	0.041	0.018	0.022	0.027	0.059	0.066	0.039
Ours-S2	0.016	0.053	0.030	0.014	0.022	0.025	0.045	0.049	0.032
Ours-S4	0.012	0.045	0.020	0.008	0.018	0.017	0.031	0.040	0.024
Ours-S8	0.010	0.040	0.015	0.007	0.016	0.014	0.023	0.035	0.020
Ours-B2	0.013	0.044	0.020	0.010	0.021	0.020	0.029	0.043	0.025
Ours-B4	0.010	0.038	0.015	0.007	0.018	0.015	0.021	0.036	0.020
Ours-B8	0.009	0.035	0.012	0.006	0.016	0.013	0.017	0.032	0.017

12

Table 10: Quantitative results of storage size and reconstruction quality (PSNR) on each scene of the Tanks and Temples dataset. We highlight the best score and second-best score .

Method	Barn	Caterpillar	Family	Ignatius	Truck	Avg.
			Size (MB)↓			
DVGO [3, 4]	9.2	7.6	6.3	5.5	7.3	7.2
Plenoxels [5]	14.7	10.7	9.3	8.5	12.3	11.1
TensoRF-CP [6]	37.7	33.0	28.5	31.8	33.8	33.0
TensoRF-VM [6]	29.1	26.6	23.7	23.4	28.5	26.3
CCNeRF-CP [7]	33.8	29.0	23.5	18.7	23.1	25.6
CCNeRF-HY [7]	33.8	29.0	23.5	18.7	23.1	25.6
Instant-NGP [8]	3.7	3.6	3.4	3.3	3.7	3.5
K-Planes-explicit [9]	40.2	39.8	35.9	37.3	39.2	38.5
K-Planes-hybrid [9]	43.7	42.3	38.3	39.7	41.5	41.1
NeRF [10]	4.6	4.6	4.6	4.6	4.6	4.6
Mip-NeRF [11]	2.3	2.3	2.3	2.3	2.3	2.3
Ours-S2	0.5	0.5	0.5	0.5	0.5	0.5
Ours-S4	1.0	1.0	0.9	1.0	1.0	1.0
Ours-S8	1.9	1.9	1.8	1.9	1.8	1.9
Ours-B2	1.6	1.6	1.5	1.6	1.6	1.6
Ours-B4	3.1	3.1	2.9	3.2	3.1	3.1
Ours-B8	6.1	6.0	5.8	6.3	6.0	6.0
			PSNR↑			
DVGO [3, 4]	26.94	25.76	33.73	28.23	27.14	28.36
Plenoxels [5]	24.50	25.27	30.01	27.87	26.69	26.87
TensoRF-CP [6]	27.03	25.27	32.78	28.27	26.30	27.93
TensoRF-VM [6]	27.50	26.30	34.16	28.38	26.92	28.65
CCNeRF-CP [7]	25.19	23.41	30.59	27.09	24.95	26.25
CCNeRF-HY [7]	26.41	24.78	32.75	27.81	26.37	27.62
Instant-NGP [8]	26.10	25.76	32.48	26.80	27.46	27.72
K-Planes-explicit [9]	27.39	25.37	33.38	26.99	26.94	28.01
K-Planes-hybrid [9]	27.32	25.52	33.39	26.98	26.84	28.01
NeRF [10]	27.27	25.32	31.56	27.29	26.47	27.58
Mip-NeRF [11]	27.45	25.41	32.38	27.57	26.05	27.77
Ours-S2	27.35	25.70	33.38	27.83	26.96	28.24
Ours-S4	27.61	25.86	33.90	27.82	27.29	28.50
Ours-S8	27.69	25.95	34.33	27.93	27.47	28.67
Ours-B2	27.65	25.87	33.86	27.78	27.31	28.49
Ours-B4	27.74	25.97	34.33	27.92	27.46	28.68
Ours-B8	27.69	26.00	34.45	27.92	27.54	28.72

13

Table 11: Quantitative results of reconstruction quality (SSIM, LPIPS) on each scene of the Tanks and Temples dataset. We highlight the best score and second-best score .

Method	Barn	Caterpillar	Family	Ignatius	Truck	Avg.
SSIM↑						
DVGO [3, 4]	0.839	0.905	0.962	0.943	0.906	0.911
Plenoxels [5]	0.842	0.905	0.959	0.942	0.910	0.912
TensoRF-CP [6]	0.845	0.884	0.951	0.939	0.887	0.901
TensoRF-VM [6]	0.866	0.913	0.967	0.949	0.913	0.922
CCNeRF-CP [7]	0.794	0.859	0.931	0.924	0.858	0.873
CCNeRF-HY [7]	0.828	0.887	0.957	0.938	0.895	0.901
Instant-NGP [8]	0.862	0.920	0.968	0.953	0.929	0.926
K-Planes-explicit [9]	0.876	0.909	0.963	0.938	0.917	0.921
K-Planes-hybrid [9]	0.872	0.912	0.963	0.940	0.917	0.921
NeRF [10]	0.845	0.893	0.943	0.935	0.893	0.902
Mip-NeRF [11]	0.846	0.888	0.947	0.933	0.890	0.901
Ours-S2	0.853	0.898	0.956	0.941	0.897	0.909
Ours-S4	0.864	0.904	0.962	0.944	0.908	0.916
Ours-S8	0.872	0.907	0.965	0.945	0.913	0.920
Ours-B2	0.869	0.904	0.963	0.944	0.907	0.917
Ours-B4	0.877	0.909	0.966	0.946	0.914	0.922
Ours-B8	0.882	0.910	0.968	0.947	0.917	0.925
LPIPS↓						
DVGO [3, 4]	0.291	0.168	0.070	0.086	0.159	0.155
Plenoxels [5]	0.276	0.162	0.076	0.094	0.151	0.152
TensoRF-CP [6]	0.275	0.212	0.081	0.092	0.199	0.172
TensoRF-VM [6]	0.248	0.157	0.059	0.077	0.147	0.138
CCNeRF-CP [7]	0.371	0.263	0.108	0.117	0.237	0.219
CCNeRF-HY [7]	0.306	0.214	0.076	0.100	0.184	0.176
Instant-NGP [8]	0.210	0.118	0.040	0.056	0.106	0.106
K-Planes-explicit [9]	0.196	0.138	0.065	0.107	0.126	0.126
K-Planes-hybrid [9]	0.199	0.133	0.059	0.103	0.125	0.124
NeRF [10]	0.282	0.189	0.100	0.098	0.185	0.171
Mip-NeRF [11]	0.281	0.194	0.093	0.101	0.184	0.171
Ours-S2	0.220	0.154	0.062	0.079	0.153	0.134
Ours-S4	0.204	0.145	0.054	0.075	0.138	0.123
Ours-S8	0.192	0.138	0.048	0.074	0.128	0.116
Ours-B2	0.198	0.144	0.052	0.075	0.139	0.122
Ours-B4	0.187	0.136	0.046	0.072	0.128	0.114
Ours-B8	0.180	0.133	0.043	0.072	0.121	0.109

14

References

[1] Chenxi Lola Deng and Enzo Tartaglione. Compressing explicit voxel grid representations: fast nerfs become also small. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1236–1245, 2023.

[2] Lingzhi Li, Zhen Shen, Zhongshu Wang, Li Shen, and Liefeng Bo. Compressing volumetric radiance fields to 1 mb. *arXiv preprint arXiv:2211.16386*, 2022.

[3] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5459–5469, 2022.

[4] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Improved direct voxel grid optimization for radiance fields reconstruction. *arXiv preprint arXiv:2206.05085*, 2022.

[5] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5501–5510, 2022.

[6] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII*, pages 333–350. Springer, 2022.

[7] Jiaxiang Tang, Xiaokang Chen, Jingbo Wang, and Gang Zeng. Compressible-composable nerf via rank-residual decomposition. *arXiv preprint arXiv:2205.14870*, 2022.

[8] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022.

[9] Sara Fridovich-Keil, Giacomo Meanti, Frederik Warburg, Benjamin Recht, and Angjoo Kanazawa. K-planes: Explicit radiance fields in space, time, and appearance. *arXiv preprint arXiv:2301.10241*, 2023.

[10] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

[11] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5855–5864, 2021.

[12] Daniel Rho, Byeonghyeon Lee, Seungtae Nam, Joo Chan Lee, Jong Hwan Ko, and Eunbyung Park. Masked wavelet representation for compact neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20680–20690, 2023.

[13] Jiemin Fang, Taoran Yi, Xinggang Wang, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Matthias Nießner, and Qi Tian. Fast dynamic radiance fields with time-aware neural voxels. In *SIGGRAPH Asia 2022 Conference Papers*, pages 1–9, 2022.

[14] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023.

[15] Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, and Eunbyung Park. Compact 3d gaussian representation for radiance field. *arXiv preprint arXiv:2311.13681*, 2023.

15