Compressing Volumetric Radiance Fields to 1 MB

Supplementary Material

A. Detailed Results

We present detailed results for each scene on the datasets used in the main paper, including synthetic-NeRF [29] in Table 8, LLFF [28] in Table 9, synthetic-NSVF [23] in Table 10 and Tanks&Temples [19] in Table 11, respectively. The detailed comparison validate that our method can achieve comparable rendering quality with significant advantage on model size overhead compared to original DVGO [37], Plenoxels [44] and TensoRF [2], demonstrating the effectiveness and generalizability of our method.

B. More Qualitative Comparison

We further show rendering a randomly selected view of example scenes on the datasets in Figure 9, 10, 11 and 12, respectively. The rendering quality of the model after compression via our method can be effectively preserved compared to the model before compression, even on fine details, meanwhile we can achieve much lower storage cost via our compression framework.

C. Comparison with VBNF

We compare our approach with VBNF [38] alone and present the results in Table 6, as the experimental configuration introduced in its paper (and code) is quite different from ours and other baselines. It needs depth as input, and is less suitable to evaluate on widely-used NeRF datasets as other baselines, e.g., it fails to reconstruct "Lego" without depth. To ensure a fair comparison, we used the official code and configuration of VBNF (at 400×400 resolution) available on the GitHub repo of kaolin-wisp. Table 6 shows the comparison result.

Method	PSNR↑ dB	Size↓ MB	Train↓ min	Test↓ ms	Need depth?
VBNF	28.81	0.705	24	20	✓
DVGO	28.97	45.9	2.6	40	✗
VQ-DVGO	28.84	0.633	2.6+1.4	47	✗

Table 6. Comparison with VBNF on test set of V8 scene.

D. Ablation of voxel-pruning strategies.

Table 7 provides the ablation by replacing our pruning strategies with several alternatives and performing VQ and post-processing. The results demonstrate that, compared to simply choosing a fixed threshold, our proposed pruning metric achieves a more desirable balance between model size and rendering quality.

Pruning Strategy	Without Pruning	Density Threshold			Ours
		0	2.5	5.0	
PSNR(dB)↑	31.60	31.34	30.96	30.58	31.77
Size(MB)↓	3.06	1.58	1.46	1.26	1.43

Table 7. Ablation of voxel-pruning strategies with DVGO.

E. Composition of final storage.

Figure 8 demonstrates the proportion of the three different part after applying our compression pipeline to volumetric radiance fields. Since metadata and MLP weights take a fairly small amount of storage, we combine their size with codebook size for better visualization.

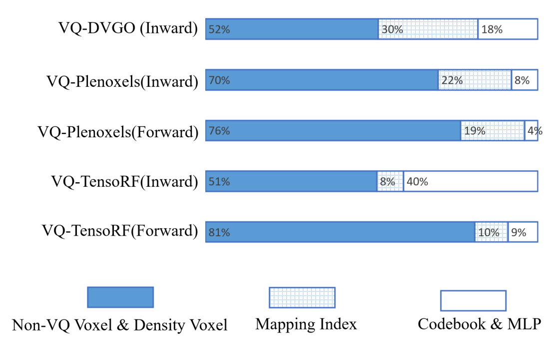

Figure 8. Visualization of model size proportion after compression. "Inward" represent the average results of synthetic-NeRF, synthetic-NSVF, and Tanks&Temples datasets." Forward" represent results on LLFF datasets.

F. Implementation Details

VQ-DVGO. The model of DVGO consists of density grid, feature grid and a shallow MLP for color estimation. As mentioned in the main paper, we first prune off less important voxels according to the quantile of the cumulative score rate. We only use the voxels that need to be vector quantized for codebook initialization. After codebook initialization, we start joint finetuning with the learning rate 8e-2 for density grid, 1.6e-3 for both the feature grid and the shallow MLP. We adopt a exponential learning rate decay schedule following its original setup, downscaling the learning rates by 0.3 for every 10k iterations. The finetuning stage takes 10000 iterations with 8192 rays per iteration. The overall compressing time takes about 50% of the original training time across different scenes.

VQ-Plenoxels. Plenoxels include a density grid and a feature grid that are filled with spherical harmonic coefficients. During the joint finetuning phase for Plenoxels, only the vector-quantized voxel grid is tuned since this method does not involve neural networks. The fine-tuning process consists of 25600 iterations, which is equivalent to two epochs in the original Plenoxels. The learning rate scheduler is reset to the sixth epoch in the original setting. The batch size for this process is 5000, and it incurs a time cost of about 20% of the original training time.

VQ-TensoRF. We choose TensoRF-VM-192 as the default model. As TensoRF utilizes a triplane structure of feature volumes, we assign three different codebooks to the three planes for more extensive model capacity, and also apply three codebooks for vector quantization on density planes. In practice, we leverage a virtual grid and compute importance scores for all the grid points. Then the points are projected onto triplanes by aggregating them along each axis. The joint finetuning stage for TensoRF takes 10000 iterations, with the learning rate 5e-3 for all the density triplane, feature triplane and the shallow MLP. We adopt an exponential learning rate decay for TensoRF, downscaling the learning rate by 0.3 for every 30k iterations.

Figure 9. **Synthetic-NeRF scenes.** We show a random view for each scene in the dataset, comparing ground truth with our VQ-DVGO, VQ-Plenoxels, VQ-TensoRF.

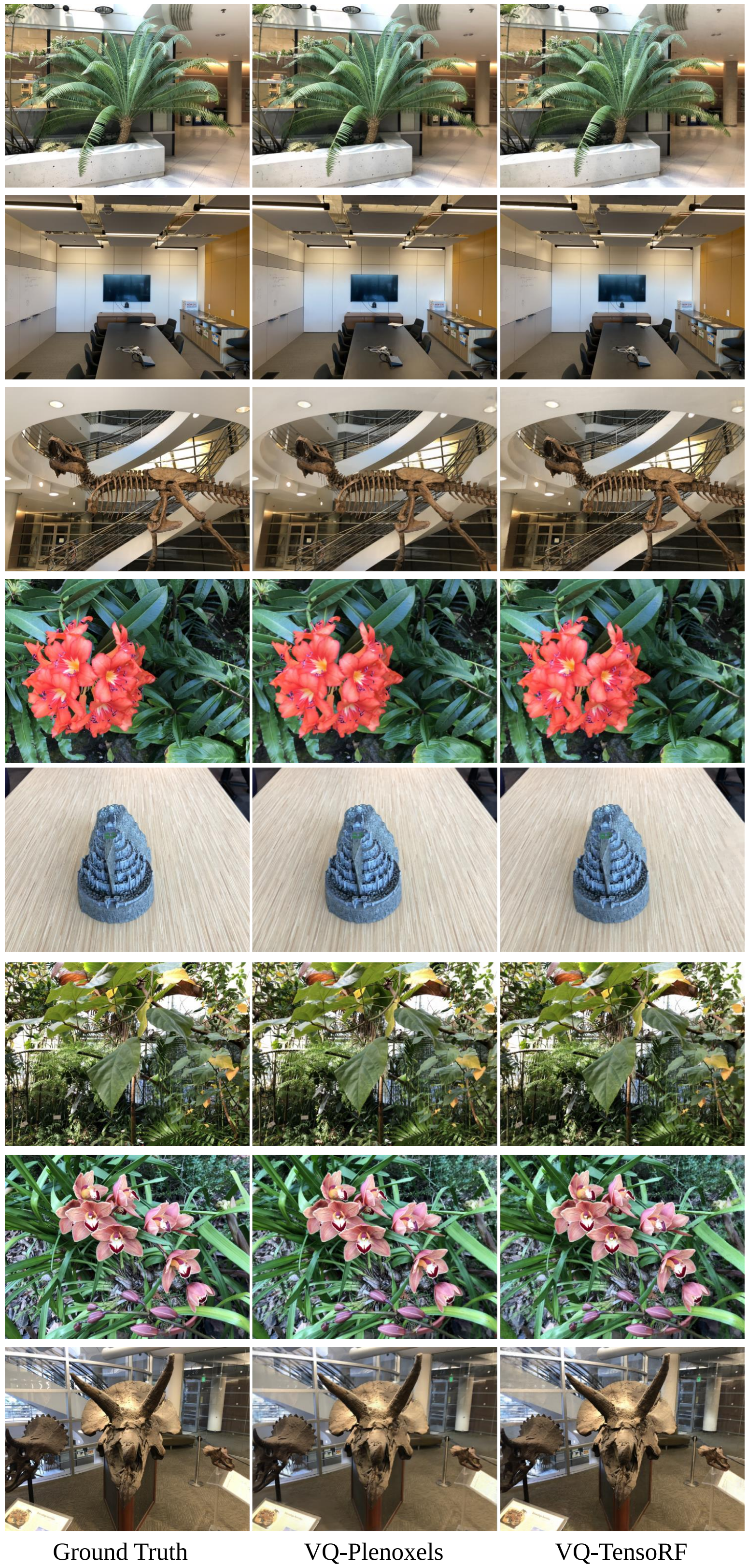

Figure 10. **LLFF scenes.** We show a random view for each scene in the dataset, comparing ground truth with our VQ-Plenoxels, VQ-TensoRF.

Figure 11. **Synthetic-NSVF scenes.** We show a random view for each scene in the dataset, comparing ground truth with our VQ-DVGO, VQ-Plenoxels, VQ-TensoRF.

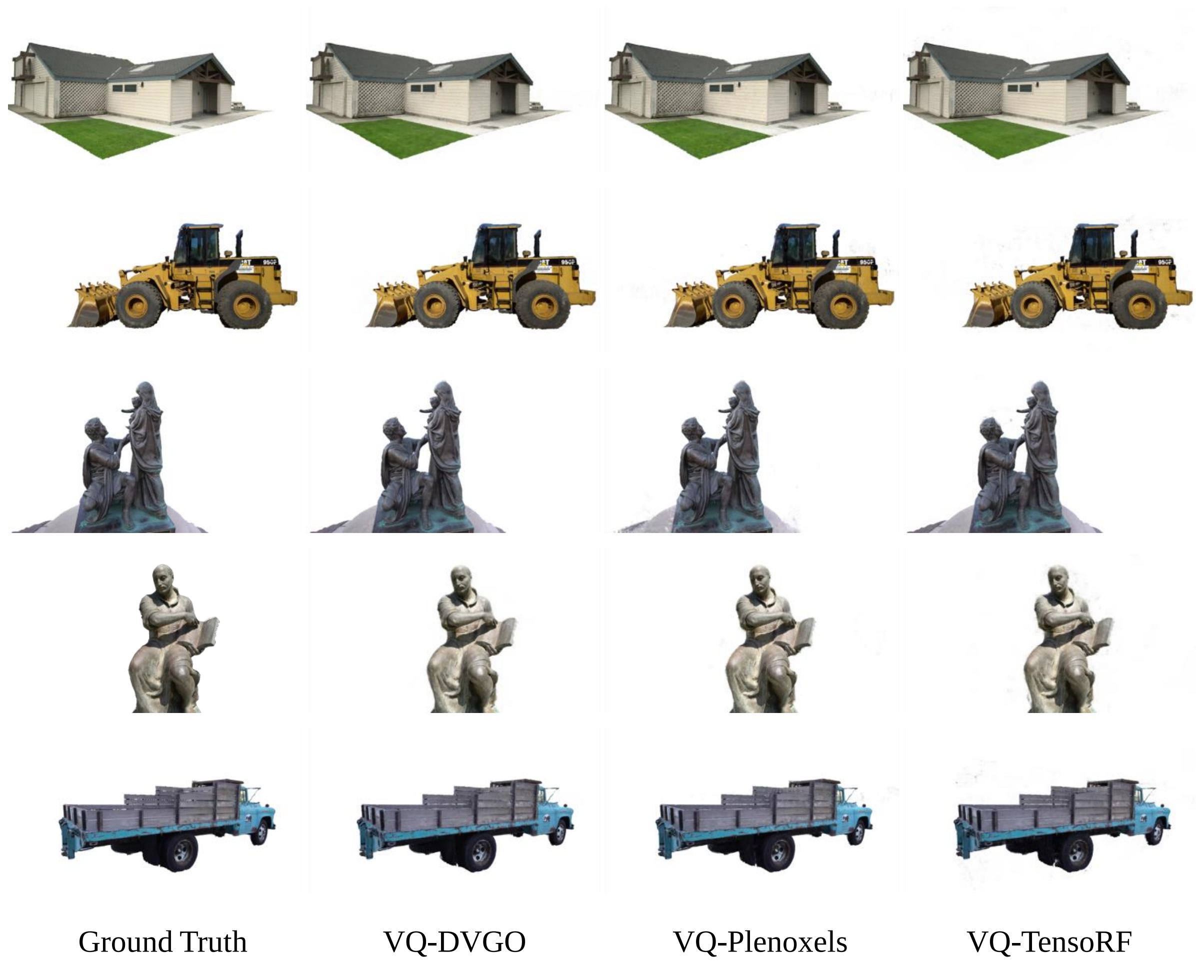

Figure 12. **Tanks&Temples scenes.** We show a random view for each scene in the dataset, comparing ground truth with our VQ-DVGO, VQ-Plenoxels, VQ-TensoRF.

	Synthetic-NeRF									
	Method	Chair	Drums	Ficus	Hotdog	Lego	Materials	Mic	Ship	**Avg.**
Size(MB) (↓)	DVGO	99.86	90.64	103.98	124.73	118.56	163.77	47.03	98.79	105.92
	VQ-DVGO	0.99	0.89	1.06	1.46	1.48	3.11	0.41	2.05	1.43
	Plenoxels	186.17	161.51	108.46	292.43	294.53	196.51	82.27	756.37	259.78
	VQ-Plenoxels	10.12	7.17	4.31	17.45	13.56	4.09	12.35	40.41	13.68
	TensoRF	62.99	63.20	66.00	78.06	63.88	78.51	62.09	65.74	67.56
	VQ-TensoRF	3.47	3.31	4.12	3.52	3.57	4.27	2.44	3.76	3.56
PSNR(dB) (↑)	DVGO	34.09	25.47	32.66	36.67	34.59	29.51	33.11	29.11	31.90
	VQ-DVGO	33.80	25.38	32.67	36.47	34.27	29.28	33.11	29.24	31.77
	Plenoxels	33.99	25.35	31.83	36.42	34.10	29.14	33.27	29.62	31.71
	VQ-Plenoxels	33.82	25.30	31.87	36.01	33.66	28.89	33.24	29.45	31.53
	TensoRF	35.61	25.98	33.95	37.40	36.36	30.03	34.82	30.57	33.09
	VQ-TensoRF	35.10	25.97	33.85	36.98	36.03	30.07	34.45	30.38	32.86
SSIM(↑)	DVGO	0.976	0.930	0.978	0.980	0.976	0.950	0.983	0.878	0.956
	VQ-DVGO	0.974	0.928	0.977	0.978	0.973	0.945	0.982	0.877	0.954
	Plenoxels	0.977	0.933	0.976	0.980	0.975	0.949	0.985	0.890	0.958
	VQ-Plenoxels	0.975	0.931	0.975	0.979	0.972	0.945	0.984	0.889	0.956
	TensoRF	0.984	0.937	0.982	0.982	0.983	0.952	0.988	0.892	0.963
	VQ-TensoRF	0.981	0.932	0.982	0.980	0.981	0.950	0.986	0.887	0.960
LPIPS_{ALEX}(↓)	DVGO	0.017	0.060	0.015	0.018	0.013	0.027	0.014	0.117	0.035
	VQ-DVGO	0.018	0.061	0.017	0.018	0.013	0.033	0.014	0.112	0.036
	Plenoxels	0.019	0.055	0.015	0.018	0.016	0.012	0.026	0.083	0.031
	VQ-Plenoxels	0.020	0.057	0.016	0.021	0.018	0.031	0.013	0.089	0.033
	TensoRF	0.010	0.051	0.013	0.014	0.007	0.027	0.008	0.087	0.027
	VQ-TensoRF	0.016	0.063	0.014	0.017	0.009	0.030	0.013	0.094	0.032
LPIPS_{VGG}(↓)	DVGO	0.028	0.078	0.025	0.034	0.027	0.059	0.018	0.160	0.054
	VQ-DVGO	0.032	0.082	0.028	0.039	0.030	0.066	0.020	0.160	0.057
	Plenoxels	0.031	0.067	0.026	0.038	0.028	0.057	0.015	0.134	0.050
	VQ-Plenoxels	0.033	0.072	0.028	0.041	0.033	0.064	0.017	0.139	0.053
	TensoRF	0.022	0.072	0.023	0.032	0.018	0.060	0.015	0.141	0.048
	VQ-TensoRF	0.035	0.099	0.028	0.040	0.024	0.064	0.025	0.149	0.058

Table 8. Per-scene results on Synthetic-NeRF [29].

	Method	Fern	Flower	Room	Leaves	Horns	Trex	Fortress	Orchids	**Avg.**
					LLFF					
Size(MB) (↓)	Plenoxels	1842.22	1719.30	1613.74	2061.28	1959.84	1819.81	1685.97	3347.22	2006.17
	VQ-Plenoxels	38.94	38.92	31.40	41.71	38.34	34.64	38.82	57.40	40.02
	TensoRF	179.92	179.81	179.87	179.70	179.81	179.85	179.87	179.90	179.84
	VQ-TensoRF	8.61	8.97	8.02	9.14	8.40	8.11	9.32	9.20	8.72
PSNR(dB) (↑)	Plenoxels	25.51	28.16	30.29	21.58	27.68	26.51	31.10	20.65	26.44
	VQ-Plenoxels	25.46	27.91	30.18	21.50	27.52	26.11	30.93	20.53	26.27
	TensoRF	25.03	28.10	32.16	21.12	28.31	27.56	31.44	19.85	26.70
	VQ-TensoRF	24.82	27.82	31.89	21.00	27.96	27.30	31.14	19.75	26.46
SSIM(↑)	Plenoxels	0.835	0.866	0.938	0.764	0.859	0.891	0.886	0.698	0.842
	VQ-Plenoxels	0.833	0.861	0.936	0.761	0.856	0.888	0.884	0.691	0.839
	TensoRF	0.801	0.857	0.952	0.744	0.883	0.910	0.898	0.644	0.836
	VQ-TensoRF	0.791	0.843	0.947	0.727	0.866	0.902	0.881	0.636	0.824
LPIPS$_{ALEX}$(↓)	Plenoxels	0.150	0.122	0.128	0.153	0.178	0.132	0.108	0.187	0.145
	VQ-Plenoxels	0.146	0.119	0.126	0.146	0.176	0.131	0.105	0.184	0.142
	TensoRF	0.157	0.103	0.076	0.144	0.103	0.080	0.067	0.192	0.115
	VQ-TensoRF	0.166	0.115	0.084	0.156	0.125	0.089	0.098	0.202	0.129
LPIPS$_{VGG}$(↓)	Plenoxels	0.835	0.866	0.938	0.764	0.859	0.891	0.886	0.698	0.842
	VQ-Plenoxels	0.220	0.180	0.199	0.200	0.232	0.238	0.178	0.241	0.211
	TensoRF	0.249	0.178	0.162	0.221	0.182	0.201	0.143	0.281	0.202
	VQ-TensoRF	0.263	0.200	0.173	0.249	0.214	0.218	0.183	0.294	0.224

Table 9. Per-scene Results on LLFF [28].

	Synthetic-NSVF									
	Method	Bike	Lifestyle	Palace	Robot	Spaceship	Steamtrain	Toad	Wineholder	**Avg.**
Size(MB) (↓)	DVGO	114.91	103.68	109.25	102.15	132.30	156.25	133.37	106.40	119.79
	VQ-DVGO	1.08	1.08	1.50	0.99	1.67	1.98	0.85	0.94	1.26
	Plenoxels	89.25	361.27	629.13	137.08	183.12	91.79	610.44	164.62	283.34
	VQ-Plenoxels	4.14	15.88	27.54	6.21	6.80	4.17	22.72	7.57	11.88
	TensoRF	73.53	67.56	67.37	70.73	70.60	83.71	71.44	68.18	71.64
	VQ-TensoRF	3.75	4.47	3.70	3.69	4.24	4.49	5.18	3.88	4.17
PSNR(dB) (↑)	DVGO	38.14	33.74	34.46	36.38	37.53	35.43	32.99	30.26	34.87
	VQ-DVGO	37.89	33.65	34.42	36.06	37.51	35.32	32.68	30.25	34.72
	Plenoxels	37.83	31.04	35.30	35.91	34.36	34.21	34.34	30.01	34.12
	VQ-Plenoxels	37.47	30.93	34.94	35.61	34.25	34.00	34.16	29.91	33.91
	TensoRF	39.39	34.64	37.84	38.55	38.74	37.99	35.10	31.49	36.72
	VQ-TensoRF	38.67	34.46	37.42	37.95	38.36	37.60	33.54	31.36	36.17
SSIM(↑)	DVGO	0.991	0.965	0.962	0.992	0.987	0.987	0.965	0.950	0.975
	VQ-DVGO	0.991	0.964	0.961	0.991	0.987	0.987	0.963	0.950	0.974
	Plenoxels	0.992	0.967	0.974	0.991	0.981	0.983	0.976	0.959	0.978
	VQ-Plenoxels	0.991	0.965	0.972	0.990	0.981	0.981	0.974	0.957	0.976
	TensoRF	0.993	0.969	0.981	0.995	0.989	0.991	0.979	0.962	0.982
	VQ-TensoRF	0.992	0.967	0.978	0.994	0.988	0.989	0.970	0.960	0.980
LPIPS_{ALEX}(↓)	DVGO	0.004	0.026	0.027	0.005	0.009	0.011	0.029	0.036	0.018
	VQ-DVGO	0.004	0.026	0.025	0.005	0.010	0.011	0.029	0.035	0.018
	Plenoxels	0.004	0.030	0.016	0.006	0.017	0.017	0.019	0.027	0.017
	VQ-Plenoxels	0.005	0.032	0.018	0.006	0.018	0.018	0.021	0.029	0.018
	TensoRF	0.003	0.020	0.010	0.003	0.009	0.006	0.014	0.022	0.011
	VQ-TensoRF	0.003	0.022	0.011	0.003	0.010	0.007	0.022	0.024	0.013
LPIPS_{VGG}(↓)	DVGO	0.011	0.053	0.043	0.013	0.020	0.022	0.046	0.054	0.033
	VQ-DVGO	0.013	0.056	0.043	0.013	0.022	0.023	0.047	0.056	0.034
	Plenoxels	0.011	0.047	0.026	0.013	0.025	0.030	0.031	0.046	0.029
	VQ-Plenoxels	0.013	0.050	0.029	0.014	0.026	0.033	0.032	0.050	0.031
	TensoRF	0.010	0.046	0.021	0.010	0.020	0.017	0.028	0.048	0.025
	VQ-TensoRF	0.013	0.051	0.024	0.011	0.022	0.023	0.044	0.053	0.030

Table 10. Per-scene Results on Synthetic-NSVF [23].

	Tanks&Temples						
	Method	Barn	Caterpillar	Family	Ignatius	Truck	**Avg.**
Size(MB) (↓)	DVGO	137.69	116.77	97.71	102.08	112.73	113.40
	VQ-DVGO	1.82	1.46	1.20	1.12	1.42	1.40
	Plenoxels	373.69	337.52	527.12	341.43	258.76	367.71
	VQ-Plenoxels	11.97	13.40	20.90	15.63	9.43	14.27
	TensoRF	80.82	72.07	67.11	67.23	75.95	72.64
	VQ-TensoRF	3.04	3.47	2.79	3.46	3.51	3.25
PSNR(dB) (↑)	DVGO	26.80	25.67	33.74	28.20	27.08	28.30
	VQ-DVGO	26.76	25.66	33.66	28.23	27.00	28.26
	Plenoxels	24.57	25.18	30.03	27.86	26.55	26.84
	VQ-Plenoxels	24.53	24.99	29.93	27.76	26.43	26.73
	TensoRF	27.48	25.92	34.06	28.38	26.89	28.54
	VQ-TensoRF	27.11	25.59	33.43	28.27	26.59	28.20
SSIM (↑)	DVGO	0.837	0.903	0.962	0.943	0.905	0.910
	VQ-DVGO	0.837	0.901	0.961	0.942	0.903	0.909
	Plenoxels	0.842	0.904	0.959	0.942	0.909	0.911
	VQ-Plenoxels	0.839	0.899	0.957	0.939	0.906	0.908
	TensoRF	0.866	0.910	0.966	0.949	0.913	0.921
	VQ-TensoRF	0.857	0.902	0.960	0.944	0.903	0.913
LPIPS$_{ALEX}$(↓)	DVGO	0.292	0.152	0.063	0.092	0.146	0.149
	VQ-DVGO	0.287	0.150	0.060	0.089	0.143	0.146
	Plenoxels	0.277	0.164	0.075	0.094	0.152	0.153
	VQ-Plenoxels	0.291	0.162	0.075	0.102	0.152	0.156
	TensoRF	0.208	0.135	0.053	0.076	0.126	0.120
	VQ-TensoRF	0.231	0.167	0.063	0.084	0.155	0.140
LPIPS$_{VGG}$ (↓)	DVGO	0.294	0.170	0.070	0.087	0.161	0.156
	VQ-DVGO	0.296	0.173	0.070	0.087	0.163	0.158
	Plenoxels	0.277	0.164	0.075	0.094	0.152	0.153
	VQ-Plenoxels	0.284	0.179	0.080	0.103	0.160	0.161
	TensoRF	0.248	0.160	0.060	0.077	0.148	0.139
	VQ-TensoRF	0.275	0.193	0.075	0.088	0.182	0.163

Table 11. Per-scene Results on Tanks&Temples [19].