Appendix

A. Overview

The algorithm of the DOT pipeline is illustrated in Alg. 1. We provide further qualitative comparisons with the baseline PlenOctree[4] and our models with recursive pruning in Fig. 2. We display more qualitative results of our methods on Fig. 4 and Fig. 5. We also report the per-scene evaluation against PlenOctree in Tab. 1 and Tab. 2. Additionally, to examine the assumption that the neighboring features can be propagated, we carry out the control experiment in Appendix C.1. Then we verify DOT's generalization ability with the case study described in Appendix C.2. Finally, more experiment details along the video demos are presented in Appendix C.3.

B. Additional Results

B.1. Qualitative Comparison

In Fig. 2, we compare our methods, including DOT denoted *Ours*, DOT with recursive pruning *Ours(R)*, *PlenOctree* with the ground truth *GT*. As observed in the rendering results, our methods show a similar visual quality to PlenOctree, while shrinking over half the memory sizes. Noticeably, compared with the ground truth images, both DOT and PlenOctree can provide satisfactory results regarding scenes such as *chair* in the first row. However, for more challenging scenes *e.g. materials*, and *drums* in the rest of the two rows, they cannot render the high-frequency regions *e.g.* the reflection on the cymbal of the drums pleasingly. The limitation may be addressed by increasing the sampling density on those regions using sampling operation like we have discussed in the main paper, splitting the scenes into the transmitted and reflected components[1], or using other physical-based rendering approximation[3].

B.2. Qualitative Evaluation

We provide more qualitative results on both the NeRF-Synthetic(See Fig. 4) and the Tank & Temples(See Fig. 5). We sample every scene in the datasets, using the poses picked randomly. All the samples are generated using the pre-trained DOT model *Ours*. It turns out that the DOT is capable of rendering photo-realistic results.

B.3. Per-scene Results

We also provide the per-scene evaluation metrics on the NeRF-synthetic dataset (See Tab. 1) and Tank&Template dataset (See Tab. 2).

Algorithm 1 DOT Model Training

Declaration:
Rays $rays$, ground truth GT and prediction $pred$
DOT functions: $render()$, $merge()$ and $sample()$
Use recursive pruning rec, and the epoch number i

procedure TRAINING($\tau, \gamma, rays, GT, i$)
    $\mathbf{Q} \leftarrow$ **optimization**($rays, GT$)
    if $i \mod T = 0$ **then**
        pruning($\mathbf{Q}, \tau$)
        sampling($\mathbf{Q}, \gamma$)

procedure OPTIMIZATION($rays, GT$)
    $rays \leftarrow permutate(rays)$
    $\mathbf{Q}, pred \leftarrow DOT.render(rays)$
    $loss \leftarrow MSE(pred, GT)$
    return $\mathbf{Q}$

procedure PRUNING($\mathbf{Q}, \tau, rec$):
    while $True$ **do**
        $sel \leftarrow \mathbf{Q} \leq \tau$
        $merge(sel)$
        if $not\ rec$ **or** $sel\ is\ empty$ **then**
            break;

procedure SAMPLING(($\mathbf{Q}, \gamma$):
    $sel \leftarrow topk(\gamma, Q)$
    $sample(sel)$

Feature Fusion **No Feature Fusion**

Figure 1: **The ablation study on the feature fusion.** We train the models on the *ficus* scene of the synthetic dataset for 100 epochs by feature fusion and initializing properties to learn from scratch.

C. Technical Details

C.1. Assumption Verification

To verify the neighboring assumption in Sec 3.2 of the main paper, we conduct a complementary experiment by

disabling fusion (see Fig. 1) in the training process. The loss of globally consistent features in the highlighted box confirms that the features are shared across octrees.

C.2. Generalization Ability Verification

Although we target the **same** scenes as POT, the generalization ability for other NeRF datasets such as BlendedMVS[2] is also tested. As shown in Fig. 3, the DOT outperforms POT with $+1.4$ PSNR, -15.3 MB memory, and a more compact octree structure. The zoomed boxes reveal that DOT provides much more intrinsic details. As we have included in our conclusion in the main paper, DOT can't reduce the excessive training time for its precursor NeRF-SH, which is required for both POT and DOT, from which they resample and cache the learned properties. We welcome interested readers to further the study on the generalization ability for more detests such as 360 data or the LLFF dataset.

C.3. Experiment Details

C.4. Device Information

MX150 embedded in the laptop has 384 CUDA cores with a clock rate of 1468 MHz and a memory data rate of 6.01 Gbps. Its bandwidth is 48.06 GB per second, and its shared system memory is 4038 MB.

D. Videos Details

The video attached to our supplementary materials consists of the following sections:

1) The general introduction about the octree representation.

2) The DOT's sampling and pruning operation demos.

3) The overview of DOT's pipeline.

4) The Real machine testing on the laptop with MX150.

5) The visual comparison with PlenOctree.

6) The visual comparison between the model *our*, *our(R)* and their compressed models.

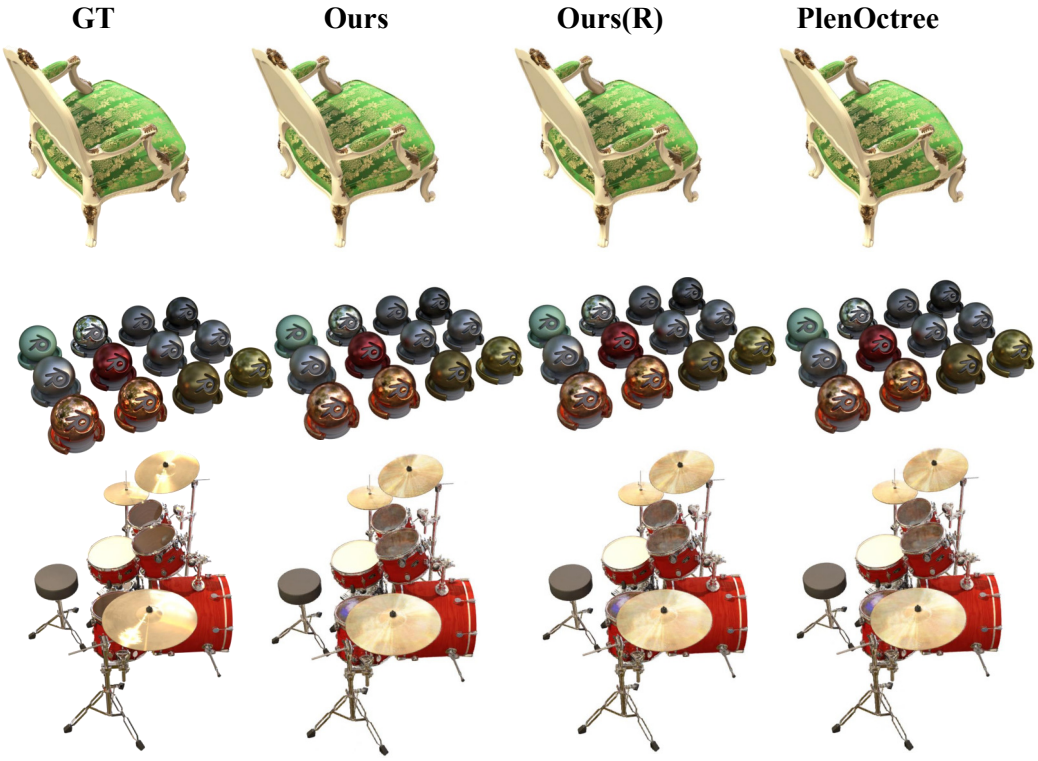

Figure 2: **Comparison on a list of models on NeRF-synthetic.**

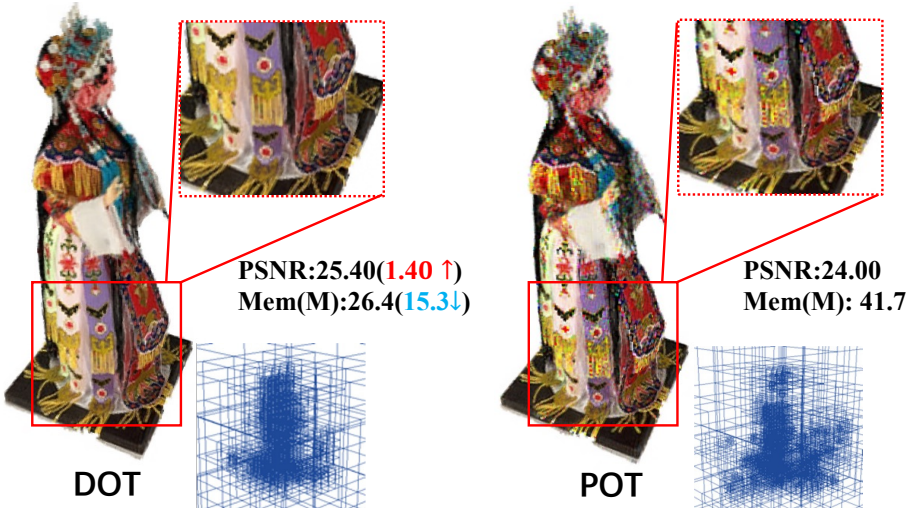

Figure 3: **The comparison between DOT and POT on Character scene of BlendedMVS.**

Figure 4: **More qualitative results of our DOT on NeRF-synthetic.**

Figure 5: **More qualitative results of our DOT on Tanks&Temples.**

PSNR↑

Methods	chair	drums	ficus	hotdog	lego	materials	mic	ship	Avg
PlenOctree	34.66	25.31	30.79	36.79	32.95	29.76	33.97	29.42	31.71
ours(R)	34.74	26.23	31.16	36.67	33.67	29.89	34.23	29.61	32.00
ours	34.82	26.25	31.16	36.76	33.82	30.24	34.24	29.61	32.11

SSMI↑

Methods	chair	drums	ficus	hotdog	lego	materials	mic	ship	Avg
PlenOctree	0.9809	0.9330	0.9705	0.9822	0.9714	0.9549	0.9872	0.8841	0.9580
ours(R)	0.9807	0.9327	0.9716	0.9818	0.9705	0.9498	0.9875	0.8867	0.9577
ours	0.9810	0.9329	0.9718	0.9823	0.9711	0.9547	0.9876	0.8868	0.9585

LPIPS↓

LPIPS	chair	drums	ficus	hotdog	lego	materials	mic	ship	Avg
PlenOctree	0.0223	0.0764	0.0378	0.0319	0.0337	0.0593	0.0168	0.1441	0.0528
ours(R)	0.0225	0.0762	0.0397	0.0338	0.0348	0.0683	0.0171	0.1374	0.0537
ours	0.0221	0.0764	0.0393	0.0317	0.0343	0.0621	0.0169	0.1372	0.0525

FPS↑

Methods	chair	drums	ficus	hotdog	lego	materials	mic	ship	Avg
A100 PlenOctree	593.5	327.9	188.9	256.3	358.1	147.1	577.3	160.4	326.2
A100 ours(R)	755.6	519.9	494.8	348.8	585.0	291.9	925.2	223.9	518.1
A100 ours	753.0	490.0	414.0	327.0	553.8	276.2	870.9	213.0	487.2
3090 PlenOctree	550.2	302.4	117.4	78.8	279.1	59.8	561.4	51.7	250.1
3090 ours(R)	744.3	468.4	536.4	196.6	571.7	249.1	929.5	97.4	474.2
3090 ours	722.8	462.0	488.9	189.7	531.6	249.5	879.6	92.5	452.1
MX150 PlenOctree	15.38	7.59	✗	✗	✗	✗	12.81	✗	✗
MX150 ours(R)	19.72	11.56	11.19	7.15	13.56	5.95	19.74	3.90	11.60
MX150 ours	18.87	10.67	9.92	6.70	12.43	5.75	18.19	3.70	10.78
A100 PlenOctree*	675.1	344.0	166.7	204.1	347.5	145.0	610.7	141.0	329.3
A100 ours(R)*	927.5	587.9	486.5	373.0	690.2	318.8	908.0	231.1	565.4
A100 ours*	894.4	550.0	442.7	351.6	638.8	302.4	854.0	215.2	531.1
3090 PlenOctree*	606.9	340.6	167.4	146.7	356.7	72.5	615.8	71.9	288.8
3090 ours(R)*	765.4	501.3	545.1	211.9	593.0	252.7	967.6	101.5	492.3
3090 ours*	733.8	477.2	502.5	207.5	557.1	252.4	909.0	97.7	467.2
MX150 PlenOctree*	15.90	7.85	✗	✗	✗	✗	13.62	✗	✗
MX150 ours(R)*	18.8	11.5	11.2	7.3	13.7	5.8	19.6	3.9	11.5
Mx150 ours*	19.6	11.0	10.5	7.3	13.1	5.7	18.8	3.9	11.2

Checkpoint/Memory (GB)↓

Methods	chair	drums	ficus	hotdog	lego	materials	mic	ship	Avg
PlenOctree	0.81	1.3	1.8	2.7	2.1	3.7	0.43	2.7	1.94
ours(R)	0.56	0.60	0.41	1.29	0.81	0.91	0.26	1.53	0.80
ours	0.58	0.65	0.47	1.38	0.91	1.01	0.28	1.70	0.87
PlenOctree*	0.19	0.28	0.43	0.43	0.10	0.26	0.88	0.56	0.39
ours(R)*	0.15	0.15	0.11	0.30	0.21	0.23	0.07	0.38	0.20
ours*	0.16	0.16	0.12	0.31	0.23	0.26	0.08	0.41	0.22

Table 1: Per-scene quantitive results on NeRF-synthetic dataset.* denotes the model is compressed

PSNR↑

Methods	Truck	Barn	Caterpillar	Family	Ignatius	Avg
PlenOctree	26.84	26.80	25.29	32.85	28.20	28.00
ours(R)	27.05	27.45	25.58	32.94	28.22	28.25
ours	27.08	27.43	25.59	33.05	28.25	28.28

SSMI↑

Methods	Truck	Barn	Caterpillar	Family	Ignatius	Avg
PlenOctree	0.8559	0.9067	0.9622	0.9139	0.9480	0.9173
ours(R)	0.8689	0.9099	0.9633	0.9180	0.9494	0.9219
ours	0.8691	0.9099	0.9644	0.9184	0.9488	0.9221

LPIPS↓

Methods	Truck	Barn	Caterpillar	Family	Ignatius	Avg
PlenOctree	0.2259	0.1477	0.0691	0.1296	0.0802	0.1305
ours(R)	0.2067	0.1416	0.0640	0.1203	0.0750	0.1215
ours	0.2070	0.1415	0.0612	0.1194	0.0761	0.1210

FPS↑

Methods	Barn	Caterpillar	Family	Truck	Ignatius	Avg
A100 PlenOctree	94.44	113.47	77.12	129.90	40.41	91.07
A100 ours(R)	180.9	176.50	258.70	234.20	159.50	202.00
A100 ours	159.93	167.09	225.55	222.33	108.59	176.70
3090 PlenOctree	85.55	99.79	39.64	17.39	127.42	73.96
3090 ours(R)	180.45	187.08	302.13	162.79	248.00	216.09
3090 ours	167.08	175.87	248.39	113.00	226.66	186.20
MX150 PlenOctree	✗	✗	✗	✗	✗	✗
MX150 ours(R)	✗	✗	✗	✗	✗	✗
MX150 ours	✗	✗	✗	✗	✗	✗
A100 PlenOctree*	131.4	129.5	87.70	145.50	68.40	112.50
A100 ours(R)*	177.97	178.90	281.52	245.00	156.20	207.92
A100 ours*	170.40	178.33	235.26	234.23	116.00	183.00
3090 PlenOctree*	85.5	99.8	39.6	17.4	127.4	73.9
3090 ours(R)*	348.0	379.0	150.0	275.0	196.0	269.6
3090 ours*	322.0	349.0	117.0	239.0	166.0	238.6
MX150 PlenOctree*	✗	✗	✗	✗	✗	✗
MX150 ours(R)*	2.92	2.85	4.60	2.55	3.89	3.36
MX150 ours*	2.85	2.68	3.87	1.90	3.61	2.98

Checkpoint/Memory (GB)↓

Methods	Truck	Barn	Caterpillar	Family	Ignatius	Avg
PlenOctree	2.5	2.4	2.9	2.3	3	2.62
ours(R)	1.3	1.4	0.4	0.87	0.64	0.92
ours	1.5	1.5	0.53	1.1	0.85	1.10
PlenOctree*	0.60	0.67	0.84	0.53	0.63	0.65
ours(R)*	0.25	0.34	0.37	0.12	0.17	0.25
ours*	0.29	0.37	0.40	0.16	0.21	0.29

Table 2: Per-scene quantitive results on Tanks&Temples dataset.* denotes the model is compressed

References

[1] Yuan-Chen Guo, Di Kang, Linchao Bao, Yu He, and Song-Hai Zhang. Nerfren: Neural radiance fields with reflections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18409–18418, June 2022. 1

[2] Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. Neural sparse voxel fields. *NeurIPS*, 2020. 2

[3] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd E. Zickler, Jonathan T. Barron, and Pratul P. Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *CVPR*, pages 5481–5490. IEEE, 2022. 1

[4] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. Plenoctrees for real-time rendering of neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5752–5761, 2021. 1