A. More Results on NeRF-Synthetic and OpenIllumination

Here we show more complete results including metrics and visualization views on all scenes in NeRF-Synthetic and OpenIllumination. See Tab. 5, 6, 7, 8 and Fig. 13, 14.

B. Comparisons on DTU

We include DTU for the sake of completeness though it is a forward-facing dataset and falls outside our focus of interest. There are different considerations in sparse view reconstruction for forward-facing and 360 – for forward-facing scenes and objects, as the back side is undefined, the features are also largely undefined. In this case, ZeroRF still performs better than or on-par with the state-of-the-art methods (Tab. 9), but does not show a significant margin.

C. Architecture Implementation

The SD Decoder generator (final generator for ZeroRF) architecture consists of ResNet convolutional blocks and upsampling modules. More hyperparameters are listed in Tab. 10. The input noise resolutions for NeRF-Synthetic, OpenIllumination and DTU are 20 while it is 7 for generation and editing tasks. It is about $1/40$ of the image resolution. The network has only 7M parameters, and the computation is negligible compared to per-point decoding and ray integral. The decoder architecture is illustrated in Fig. 11, which is a direct implementation of Eq. (5, 6) in the main paper.

D. Grid Resolution Influence

We experiment on the NeRF-Synthetic 6-view setting at various resolutions and find that resolution has little effect on model performance, as shown in Tab. 11. Comparing resolutions of 160, 240, 320, and 400, we observe that the PSNR varies by less than 0.15 dB.

E. Limitations and Future Work

We discuss more about the limitations and future work of ZeroRF in this section. We found in our experiments that ZeroRF has a chance to magnify the weakness in the underlying representations. For example, it is known that TensoRF exhibits axis-aligned artifacts under SO(3) rotations [19]. Under certain circumstances, ZeroRF (on TensoRF) will bias towards axis-aligned geometries (see the edges of the hat in Fig. 5 of main paper, as well as the pumpkins in Fig. 14). Applying ZeroRF to DiF does not have this issue, but minor floaters in unseen areas may occur.

Another future work for ZeroRF, as mentioned in the main paper, is to apply it for unbounded scenes. Grid representations usually perform a non-linear contraction in space to represent unbounded scenes, which leads to features being distorted, especially for the background areas. The features are thus hardly perceivable as a natural image, as shown in Fig. 12. Consequently, extra work would be needed to apply our technique to unbounded scenes.

Table 5. Comparison of per-scene metrics of NeRF-Synthetic 6 view settings.

		chair	drums	ficus	hotdog	lego	materials	mic	ship	mean
RegNeRF	PSNR	13.16	9.44	12.34	9.61	9.07	7.40	11.28	6.19	9.81
	SSIM	0.580	0.280	0.512	0.470	0.413	0.258	0.463	0.285	0.407
	LPIPS	0.510	0.609	0.543	0.569	0.600	0.634	0.523	0.651	0.580
FlipNeRF	PSNR	25.31	18.66	19.81	27.73	21.29	20.79	20.38	19.84	21.73
	SSIM	0.887	0.815	0.844	0.925	0.820	0.839	0.887	0.746	0.845
	LPIPS	0.080	0.239	0.144	0.173	0.207	0.242	0.172	0.361	0.202
DietNeRF	PSNR	25.21	19.81	20.18	9.37	20.31	7.78	25.85	6.79	16.91
	SSIM	0.887	0.838	0.852	0.644	0.805	0.447	0.941	0.405	0.727
	LPIPS	0.112	0.133	0.127	0.442	0.174	0.493	0.072	0.582	0.267
InfoNeRF	PSNR	24.87	18.39	20.59	23.61	21.92	20.42	20.84	20.86	21.44
	SSIM	0.892	0.824	0.859	0.903	0.854	0.838	0.904	0.758	0.854
	LPIPS	0.111	0.190	0.138	0.133	0.150	0.159	0.119	0.274	0.159
FreeNeRF	PSNR	26.57	18.16	18.46	27.18	24.32	21.63	25.64	20.23	22.77
	SSIM	0.916	0.827	0.840	0.929	0.887	0.853	0.942	0.729	0.865
	LPIPS	0.071	0.176	0.161	0.096	0.132	0.202	0.066	0.290	0.149
Ours	PSNR	27.62	20.88	22.21	29.93	26.26	21.41	27.40	22.13	24.73
	SSIM	0.926	0.869	0.898	0.949	0.913	0.849	0.954	0.756	0.889
	LPIPS	0.074	0.131	0.100	0.075	0.085	0.132	0.050	0.256	0.113

Table 6. Comparison of per-scene metrics of NeRF-Synthetic 4 view settings.

|		chair	drums	ficus	hotdog	lego	materials	mic	ship	mean
RegNeRF	PSNR	13.12	9.75	11.78	9.16	8.64	7.91	13.10	5.98	9.93
	SSIM	0.581	0.304	0.422	0.475	0.364	0.254	0.696	0.258	0.419
	LPIPS	0.507	0.615	0.594	0.565	0.639	0.621	0.353	0.683	0.572
FlipNeRF	PSNR	19.89	16.53	18.76	26.26	19.96	20.71	17.99	18.15	19.78
	SSIM	0.828	0.771	0.836	0.918	0.801	0.844	0.858	0.715	0.822
	LPIPS	0.130	0.281	0.151	0.170	0.209	0.200	0.182	0.374	0.212
DietNeRF	PSNR	17.47	12.96	9.50	12.33	7.87	6.19	14.81	6.21	10.92
	SSIM	0.775	0.650	0.451	0.658	0.397	0.363	0.773	0.389	0.557
	LPIPS	0.264	0.333	0.518	0.418	0.587	0.545	0.286	0.616	0.446
InfoNeRF	PSNR	20.02	12.13	19.47	18.92	17.77	20.38	15.79	18.18	17.83
	SSIM	0.841	0.686	0.849	0.864	0.770	0.850	0.845	0.713	0.802
	LPIPS	0.164	0.344	0.153	0.179	0.221	0.142	0.184	0.310	0.212
FreeNeRF	PSNR	20.22	14.99	17.35	23.58	20.43	21.36	15.05	17.52	18.81
	SSIM	0.843	0.746	0.809	0.899	0.818	0.857	0.802	0.687	0.808
	LPIPS	0.109	0.280	0.144	0.108	0.156	0.174	0.218	0.318	0.188
Ours	PSNR	23.04	16.91	20.12	29.11	22.11	20.50	24.76	19.01	21.94
	SSIM	0.880	0.791	0.866	0.944	0.868	0.848	0.944	0.707	0.856
	LPIPS	0.107	0.206	0.120	0.088	0.122	0.129	0.056	0.283	0.139
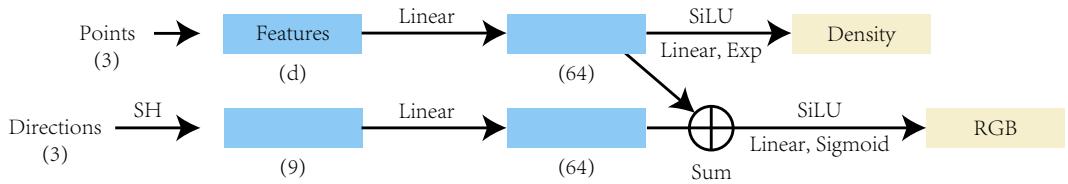
Figure 11. Decoder architecture.

Table 7. Comparison of per-scene metrics of OpenIllumination 6 view settings. We employ early-stopping by error on a validation view.

		stone	pumpkin	toy	potato	pine	shroom	cow	cake	mean
RegNeRF	PSNR	13.80	13.58	13.54	13.92	11.87	13.22	13.07	19.66	14.08
	SSIM	0.848	0.848	0.884	0.854	0.807	0.863	0.807	0.958	0.859
	LPIPS	0.288	0.350	0.237	0.348	0.337	0.329	0.405	0.128	0.303
DietNeRF	PSNR	24.87	24.80	25.37	25.63	18.16	23.71	21.50	29.58	24.20
	SSIM	0.921	0.966	0.944	0.955	0.902	0.930	0.930	0.973	0.940
	LPIPS	0.085	0.073	0.086	0.087	0.119	0.119	0.133	0.059	0.095
InfoNeRF	PSNR	14.37	26.02	25.91	25.55	21.71	22.99	22.04	19.60	22.27
	SSIM	0.910	0.960	0.952	0.946	0.917	0.914	0.915	0.962	0.935
	LPIPS	0.106	0.057	0.058	0.078	0.086	0.116	0.126	0.094	0.090
FreeNeRF	PSNR	11.62	11.71	10.65	11.35	8.85	10.12	11.09	16.33	11.46
	SSIM	0.791	0.864	0.814	0.832	0.753	0.764	0.784	0.900	0.813
	LPIPS	0.236	0.293	0.346	0.397	0.328	0.505	0.442	0.265	0.351
Ours	PSNR	26.30	27.87	27.28	27.26	22.26	26.34	23.74	31.00	26.49
	SSIM	0.929	0.966	0.950	0.951	0.918	0.928	0.921	0.969	0.941
	LPIPS	0.063	0.064	0.062	0.084	0.088	0.106	0.118	0.052	0.080

Table 8. Comparison of per-scene metrics of OpenIllumination 4 view settings. We employ early-stopping by error on a validation view.

		stone	pumpkin	toy	potato	pine	shroom	cow	cake	mean
RegNeRF	PSNR	10.26	11.74	10.04	11.63	9.37	10.66	11.99	17.21	11.62
	SSIM	0.602	0.749	0.637	0.719	0.571	0.658	0.748	0.868	0.694
	LPIPS	0.483	0.465	0.476	0.505	0.486	0.551	0.460	0.359	0.473
DietNeRF	PSNR	24.05	26.54	24.98	23.00	20.94	19.91	16.30	28.97	23.09
	SSIM	0.921	0.970	0.949	0.949	0.924	0.911	0.894	0.971	0.936
	LPIPS	0.085	0.060	0.079	0.103	0.093	0.166	0.207	0.060	0.107
InfoNeRF	PSNR	24.29	26.11	23.84	22.89	20.06	18.33	13.63	19.60	21.09
	SSIM	0.923	0.961	0.944	0.937	0.897	0.877	0.905	0.962	0.926
	LPIPS	0.069	0.059	0.073	0.092	0.117	0.161	0.181	0.094	0.106
FreeNeRF	PSNR	12.91	11.54	10.79	11.70	10.17	11.46	11.18	17.95	12.21
	SSIM	0.779	0.827	0.786	0.796	0.791	0.751	0.746	0.899	0.797
	LPIPS	0.210	0.312	0.351	0.461	0.220	0.554	0.458	0.299	0.358
Ours	PSNR	25.07	26.07	23.72	26.27	20.68	23.14	21.91	29.44	24.42
	SSIM	0.918	0.961	0.936	0.946	0.903	0.912	0.905	0.965	0.930
	LPIPS	0.072	0.075	0.089	0.096	0.116	0.134	0.139	0.058	0.098

Table 9. Comparison of per-scene metrics of DTU 3 view settings.

	Scan	24	37	40	55	63	65	69	83	97	105	106	110	114	118	122	mean
DietNeRF	PSNR	10.37	13.06	12.69	12.92	20.24	17.99	17.91	18.85	13.47	14.83	19.52	18.04	18.09	22.98	23.65	16.97
	SSIM	0.245	0.525	0.296	0.322	0.810	0.801	0.433	0.702	0.333	0.417	0.693	0.520	0.630	0.771	0.786	0.552
	LPIPS	0.615	0.372	0.541	0.411	0.219	0.198	0.413	0.214	0.432	0.383	0.313	0.320	0.301	0.228	0.199	0.344
InfoNeRF	PSNR	10.32	8.34	9.25	14.86	5.08	12.30	12.32	9.01	9.11	9.49	15.64	14.58	16.80	16.44	17.72	12.09
	SSIM	0.446	0.342	0.407	0.384	0.300	0.399	0.109	0.417	0.303	0.322	0.359	0.186	0.495	0.407	0.497	0.358
	LPIPS	0.564	0.505	0.556	0.567	0.605	0.549	0.561	0.571	0.569	0.568	0.481	0.495	0.451	0.460	0.451	0.530
FlipNeRF	PSNR	10.81	17.41	11.93	16.72	22.02	20.37	17.02	28.11	18.60	22.13	21.65	20.44	21.53	23.56	26.16	19.90
	SSIM	0.475	0.700	0.523	0.686	0.880	0.865	0.640	0.943	0.775	0.843	0.800	0.819	0.795	0.828	0.869	0.763
	LPIPS	0.462	0.186	0.452	0.268	0.174	0.160	0.332	0.104	0.257	0.196	0.294	0.209	0.267	0.267	0.181	0.254
SPARF	PSNR	16.80	14.55	17.05	20.75	16.02	18.82	19.12	12.17	16.66	15.83	23.16	19.45	23.69	24.10	21.50	18.64
	SSIM	0.679	0.633	0.688	0.705	0.767	0.727	0.622	0.617	0.663	0.713	0.767	0.687	0.794	0.799	0.701	0.704
	LPIPS	0.440	0.373	0.341	0.305	0.238	0.356	0.372	0.503	0.437	0.239	0.412	0.373	0.359	0.373	0.363	0.366
Ours	PSNR	14.43	15.46	17.66	19.06	21.19	17.26	16.25	23.65	20.14	20.08	18.07	20.43	19.69	21.14	23.72	19.21
	SSIM	0.532	0.677	0.590	0.731	0.882	0.822	0.730	0.922	0.784	0.835	0.767	0.857	0.762	0.812	0.884	0.772
	LPIPS	0.370	0.188	0.399	0.239	0.135	0.217	0.309	0.118	0.209	0.201	0.307	0.202	0.251	0.246	0.160	0.237

Table 10. Generator architecture listing.

Item	Configuration
Input noise channels	8
Output feature channels	16
Block resolutions	$1\times, 2\times, 4\times, 8\times, 16\times, 16\times$
ResNet basic blocks per block	2, 4, 4, 4, 4, 4
# Parameters	7.0 M

Table 11. Influence of gird resolution to the model performance on NeRF-Synthetic dataset.

| Resolution	400	320	240	160
PSNR	24.85	24.73	24.87	24.74
SSIM	0.891	0.889	0.891	0.887
LPIPS	0.117	0.113	0.119	0.124
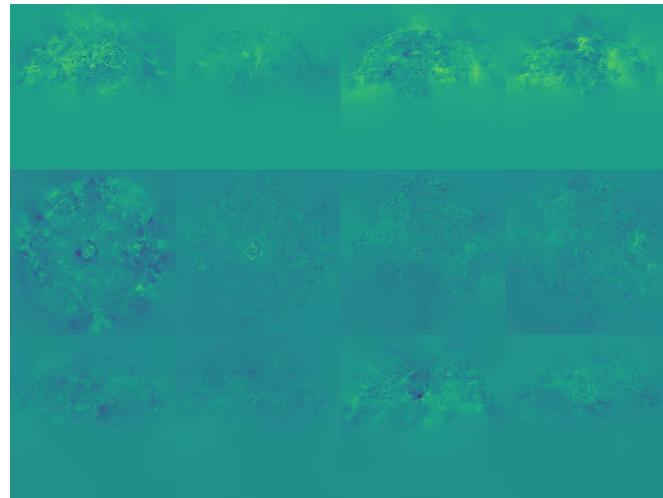
Figure 12. Visualization of features from dense-view TensoRF on the Bonsai scene from the mip-NeRF 360 dataset.

| RegNeRF | DietNeRF | InfoNeRF | FreeNeRF | FlipNeRF | Ours (Full) | GT |
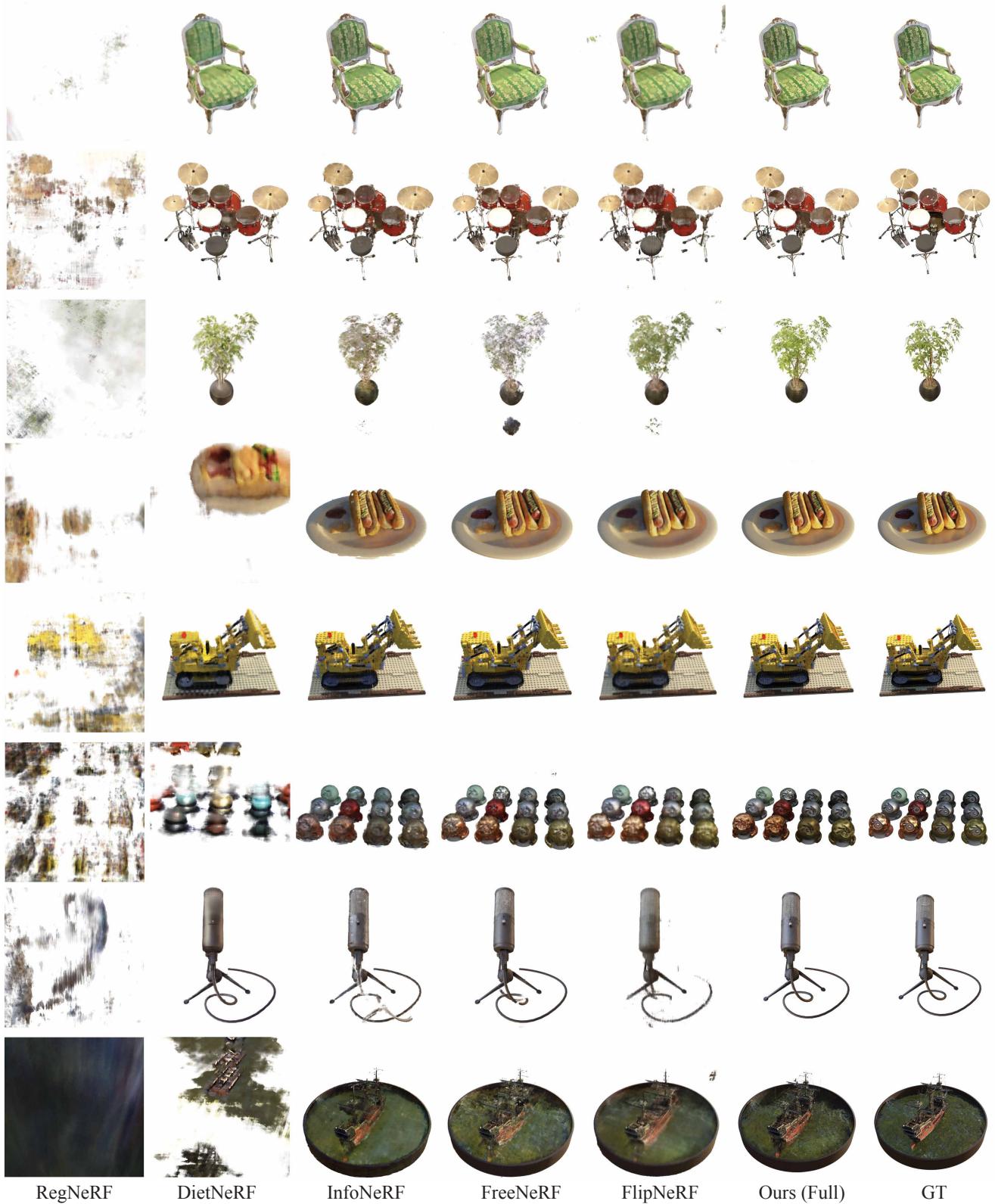
Figure 13. Per-scene qualitative comparisons of NeRF-Synthetic 6 view settings.

| RegNeRF | DietNeRF | InfoNeRF | FreeNeRF | Ours (Full) | GT |

Figure 14. Per-scene qualitative comparisons of OpenIllumination 6 view settings.