HDR-NeRF: High Dynamic Range Neural Radiance Fields (Supplementary Material)

Xin Huang[1]*, Qi Zhang[2], Ying Feng[2], Hongdong Li[3], Xuan Wang[2], Qing Wang[1]

[1] School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China
[2] Tencent AI Lab [3] Australian National University

xinhuang@mail.nwpu.edu.cn {nwpuqzhang, yfeng.von, xwang.cv}@gmail.com
HONGDONG.LI@anu.edu.au qwang@nwpu.edu.cn

1. Overview

The supplementary material shows the additional implementation details of our method, baselines, and our collected HDR dataset. Additional results are also presented to further demonstrate the superior performance of our method. We **strongly** encourage the reader to see our video supplementary, in which we present our results on the test scenes and comparisons with baselines.

2. Additional Implementation Details

Our code is built upon the PyTorch implementation of NeRF (https://github.com/yenchenlin/nerf-pytorch). During the training and testing, the rays are mapped from camera space to the normalized device coordinate (NDC) space [6]. The inference code and one model are provided in our supplementary materials.

To evaluate our estimated CRFs on synthetic scenes, we build a simple global tone-mapping function based on the classical Reinhard tone-mapping [7]. Using this function, the HDR views rendered by Blender are tone-mapped into LDR views. We then take the LDR views as our inputs. The simple tone-mapping function is defined as:

$$M(E) = \left(\frac{E}{E+1}\right)^{\frac{1}{2.2}}, \tag{1}$$

where E is the HDR pixel value. To generate LDR views with different exposure, we use the *exposure value* EV to scale the HDR pixel value E (*i.e.* $2^{EV}E$). We introduce the exposure value EV into Eq. (1):

$$M(E, EV) = \left(\frac{2^{EV}E}{2^{EV}E+1}\right)^{\frac{1}{2.2}}, \tag{2}$$

where 2^{EV} is also can be considered as the exposure time in our paper, that's $\Delta t = 2^{EV}$.

*Work done during an internship at Tencent AI Lab.

3. Baseline Methods Implementation Details

The import parameters of baseline methods, such as number of samples per ray, position encoding, and batch size, are all set as same as these of us for a fair comparison. All the models are trained with Adam about $200,000$ iterations.

NeRF: We use the PyTorch implementation of NeRF code open-source at https://github.com/yenchenlin/nerf-pytorch.

NeRF-W: The code of NeRF-W is provided at https://github.com/kwea123/nerf_pl/tree/nerfw, which is an unofficial implementation of NeRF-W using PyTorch (PyTorch-lightning).

NeRF-GT: The NeRF-GT is a version of NeRF that is directly trained from LDR views with consistent exposures or HDR views, which can be considered as the upper bound of our method. When we train the NeRF model from HDR views, the predicted HDR pixel values are tone-mapped into LDR pixel values and then compared to the tone-mapped ground truth. However, we find that it is difficult to ensure all the areas of a scene are encoded well by NeRF model, due to the high dynamic range of the scenes, even though we use the tone-mapped predicted color to calculate the loss.

4. HDR Dataset Details

Since no dataset is appropriate for the task of novel HDR views synthesis, we collect a new dataset for the evaluation of our method. Most 3D models used in our dataset are provided at https://sketchfab.com/feed. All the licenses of 3D models will be attached, when we release our dataset. The HDR views for each scene are rendered with Blender's Cycles path-tracer [1]. For real-world scenes, the LDR views with different exposures are captured by a Nikon D90 camera. We set the ISO gain to 200 and aperture to $f/6.7$. We calibrate a set of LDR images using an open-source software package COLMAP [8]. The calibration setting of COLMAP follows the one of LLFF [5]. We also capture 10

(a) Two of training HDR views (b) Novel HDR view by HDR imaging + vanilla NeRF (c) Novel HDR view by our method

Figure 1. The comparisons with HDR imaging + vanilla NeRF. All the HDR images are tone-mapped with same hyperparameters.

images with different exposures for each scene to calibrate the CRF of the Nikon D90 camera. The CRFs are calibrated with the classical method by Debevec and Malik [3].

5. Additional Results

The additional quantitative comparisons with baseline methods on the other synthetic scenes are shown in Tab. 1. Table 3 includes a breakdown of the quantitative results on real scenes presented in the main paper into per-scene metrics. The quantitative results further validate that our method outperforms the baseline methods. Figures 2, 3 and 4 show the qualitative results of our method and baselines. It can be seen that our method can accurately control the exposure of rendered LDR views compared NeRF-W, and the results by our method are reasonable close to those of NeRF-GT (the upper bound). On the other hand, our method can better reconstruct the small textures on rendering HDR views, as shown in Fig. 4. Finally, all the CRFs estimated by our method are exhibited in Fig. 5, which demonstrates that our method correctly models the tone-mapping operation of the camera. We have also tried to concatenate the $\ln e$ and $\ln \Delta t$ then feed them into the tone-mapper. Our method produces similar LDR and HDR results (PSNR: ± 0.1, SSIM:± 0.02, LPIPS:± 0.01).

We have tried to reconstruct HDR views using an HDR imaging method [9] then train vanilla NeRF, where LDR views ($\{t_1, t_3, t_5\}$) with small disparity are used to reconstruct HDR views. Some results reconstructed by the image-wise HDR imaging method are view inconsistent (Fig. 1 (a)) since the radiance scale of each view is different, which leads NeRF or IBR to render the novel views with artifacts (Fig. 1 (b)). HDR imaging + vanilla NeRF is straightforward. The HDR views captured by off-the-shelf cameras are also view-dependent and the radiance scales vary with the poses. Therefore, we propose a novel method for recovering radiance fields from LDR views. Compared with HDR imaging + vanilla NeRF, our method is an **end-to-end framework with fewer inputs and better performance.** The auto-exposure scenes can also be handled by modeling more camera settings, such as ISO and aperture. Our method can recover the radiance field and render novel HDR views from an auto-exposure video. Moreover, the exposure can also be learned, just like appearance vectors in NeRF-W.

References

[1] Blender. https://www.blender.org/.

[2] Photomatix Pro 6. https://www.hdrsoft.com/.

[3] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH*, page 369–378, 1997.

[4] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. NeRF in the wild: Neural radiance fields for unconstrained photo collections. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7210–7219, 2021.

[5] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Trans. Graph.*, 38(4):1–14, 2019.

[6] 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. In *Eur. Conf. Comput. Vis.*, pages 405–421. Springer, 2020.

[7] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. In *SIGGRAPH*, pages 267–276, 2002.

[8] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4104–4113, 2016.

[9] Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. Deep high dynamic range imaging with large foreground motions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 117–132, 2018.

Table 1. Quantitative comparisons with baseline methods on four synthetic scenes. LDR-OE denotes the average LDR results with exposure t_1, t_3, and t_5. LDR-NE denotes the average LDR results with exposure t_2, and t_4. HDR denotes the HDR results. We color code each column as best and second best .

		Diningroom			*Sponza*			*Bathroom*			*Desk*		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF [6]	LDR-OE	12.50	0.378	0.600	16.39	0.664	0.219	14.59	0.429	0.424	15.29	0.645	0.249
	LDR-NE	—	—	—	—	—	—	—	—	—	—	—	—
	HDR	—	—	—	—	—	—	—	—	—	—	—	—
NeRF-W[1][4]	LDR-OE	32.25	0.979	0.016	24.50	0.908	0.037	29.64	0.900	0.055	30.21	0.958	0.030
	LDR-NE	32.53	0.972	0.019	24.32	0.904	0.042	26.98	0.881	0.066	29.60	0.950	0.034
	HDR	—	—	—	—	—	—	—	—	—	—	—	—
Ours	LDR-OE	41.23	0.986	0.010	34.49	0.958	0.034	36.26	0.949	0.037	37.84	0.972	0.023
	LDR-NE	37.99	0.979	0.013	33.41	0.950	0.038	33.44	0.926	0.046	35.26	0.960	0.029
	HDR	38.57	0.981	0.015	32.33	0.939	0.049	33.97	0.925	0.048	43.38	0.993	0.007
NeRF-GT[2][6]	LDR-OE	43.66	0.991	0.007	37.25	0.973	0.020	38.51	0.964	0.027	39.22	0.978	0.017
	LDR-NE	41.14	0.989	0.007	34.55	0.958	0.031	35.42	0.949	0.030	37.46	0.973	0.020
	HDR	42.49	0.989	0.002	32.66	0.913	0.012	30.72	0.798	0.039	41.15	0.975	0.015

[1] The exposures of input views for NeRF-W are randomly selected from all five exposures to learn five appearance vectors for testing.
[2] A version of NeRF (as the upper bound of our method) that is trained from LDR images with consistent exposures or HDR images.

Table 2. Quantitative comparisons with baseline methods on four synthetic scenes. LDR-OE denotes the average LDR results with exposure t_1, t_3, and t_5. LDR-NE denotes the average LDR results with exposure t_2, and t_4. HDR denotes the HDR results. We color code each column as best and second best .

		Dog			*Sofa*			*Bear*			*Chair*		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF [6]	LDR-OE	13.69	0.619	0.279	15.06	0.718	0.229	11.97	0.560	0.515	12.23	0.422	0.492
	LDR-NE	—	—	—	—	—	—	—	—	—	—	—	—
	HDR	—	—	—	—	—	—	—	—	—	—	—	—
NeRF-W[1][4]	LDR-OE	31.01	0.967	0.022	30.76	0.955	0.029	32.24	0.978	0.021	28.01	0.840	0.161
	LDR-NE	30.41	0.964	0.026	30.31	0.952	0.031	32.67	0.976	0.022	26.96	0.815	0.157
	HDR	—	—	—	—	—	—	—	—	—	—	—	—
Ours	LDR-OE	37.77	0.981	0.016	38.29	0.977	0.014	42.91	0.990	0.010	32.45	0.905	0.081
	LDR-NE	36.52	0.976	0.018	38.35	0.976	0.014	41.19	0.987	0.012	30.78	0.886	0.083
	HDR	37.72	0.980	0.016	39.05	0.976	0.017	43.22	0.991	0.008	34.14	0.924	0.069
NeRF-GT[2][6]	LDR-OE	38.43	0.981	0.017	37.91	0.975	0.046	43.84	0.991	0.009	33.79	0.926	0.070
	LDR-NE	37.86	0.980	0.016	38.67	0.978	0.014	42.95	0.990	0.008	32.17	0.912	0.070
	HDR	35.66	0.967	0.007	36.38	0.955	0.044	38.43	0.971	0.014	33.72	0.922	0.010

[1] The exposures of input views for NeRF-W are randomly selected from all five exposures to learn five appearance vectors for testing.
[2] A version of NeRF (as the upper bound of our method) that is trained from LDR images with consistent exposures or HDR images.

Table 3. Quantitative comparisons with baseline methods on real scenes. LDR-OE denotes the average LDR results with exposure t_1, t_3, and t_5. LDR-NE denotes the average LDR results with exposure t_2, and t_4. We color code each column as best and second best.

		Computer			*Flower*			*Luckycat*			*Box*		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF [6]	LDR-OE	14.68	0.697	0.281	14.60	0.504	0.524	13.67	0.706	0.262	17.06	0.770	0.233
	LDR-NE	—	—	—	—	—	—	—	—	—	—	—	—
NeRF-W[1][4]	LDR-OE	28.91	0.919	0.112	26.23	0.933	0.094	30.00	0.927	0.076	29.21	0.927	0.097
	LDR-NE	27.54	0.892	0.136	26.84	0.939	0.078	30.78	0.940	0.058	29.59	0.923	0.104
Ours†	LDR-OE	31.41	0.944	0.086	27.84	0.943	0.078	31.82	0.937	0.067	30.59	0.952	0.070
	LDR-NE	29.01	0.923	0.112	26.82	0.939	0.072	31.40	0.944	0.059	30.45	0.945	0.079
Ours	LDR-OE	32.42	0.950	0.077	29.81	0.948	0.069	32.85	0.938	0.062	31.54	0.953	0.068
	LDR-NE	31.21	0.931	0.098	30.05	0.949	0.058	33.13	0.948	0.051	31.40	0.944	0.079
NeRF-GT[2][6]	LDR-OE	34.34	0.955	0.075	32.84	0.957	0.057	34.56	0.951	0.049	36.55	0.968	0.050
	LDR-NE	32.73	0.940	0.090	33.38	0.957	0.048	36.42	0.962	0.035	35.97	0.965	0.044

[1] The exposures of input views for NeRF-W are randomly selected from all five exposures to learn five appearance vectors for testing.
[2] A version of NeRF (as the upper bound of our method) that is trained from LDR images with consistent exposures.
† An ablation study of our method that models the tone-mapping operations of RGB channels with a single MLP.

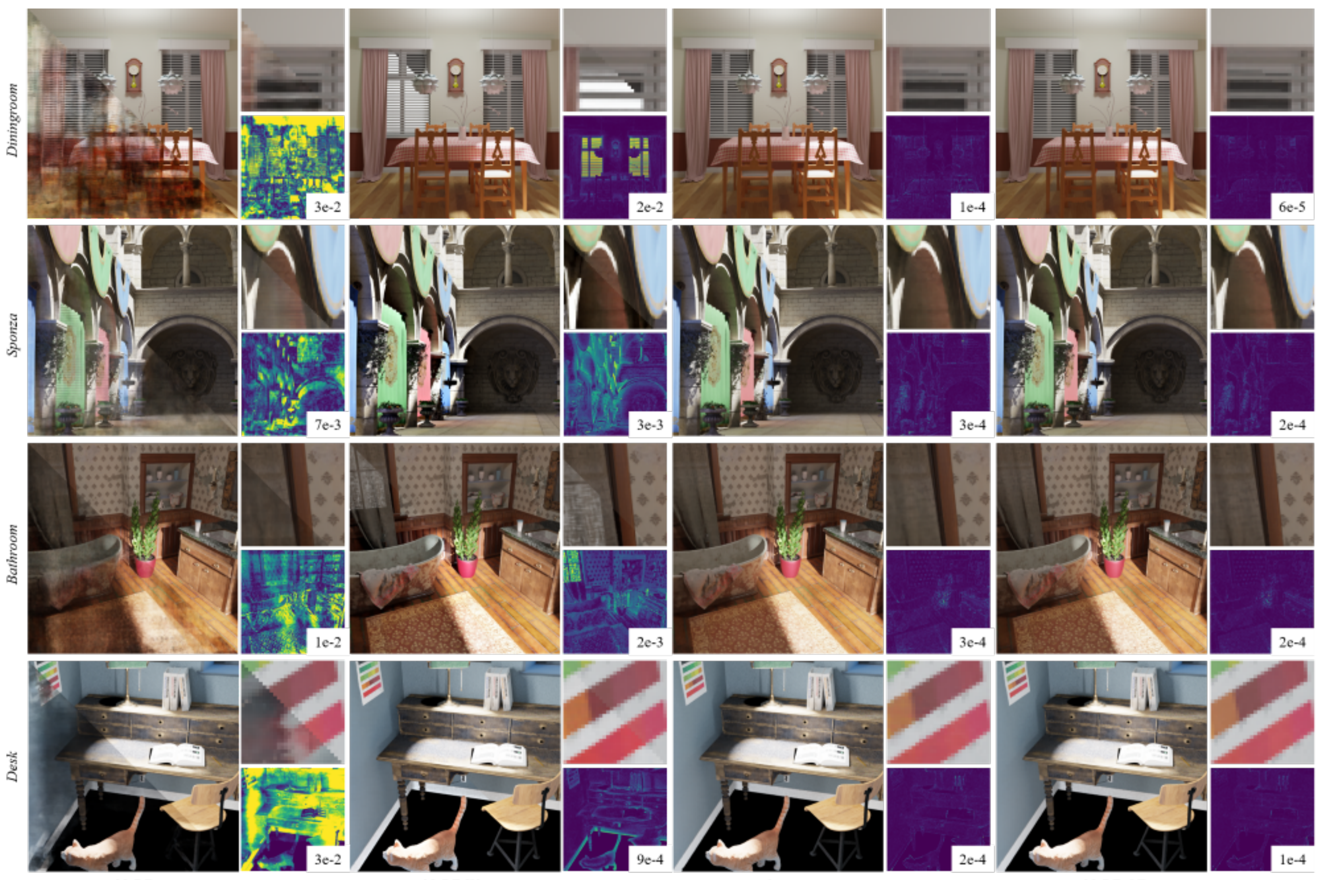
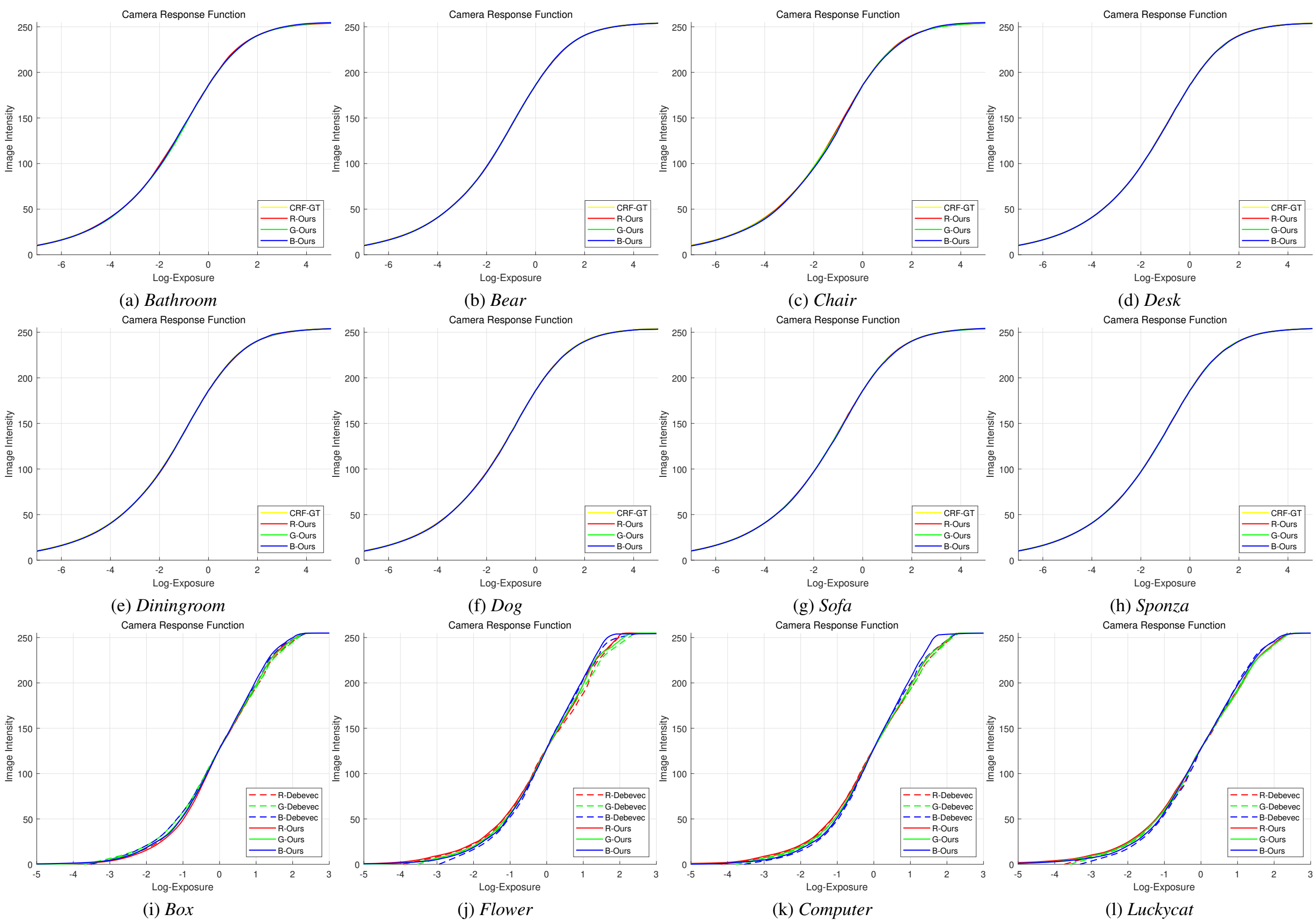


Figure 2. Qualitative comparisons of novel LDR views with novel exposures. The upper triangular images are the ground truth and the lower triangular images are the rendered views. Zoom-in insets and error maps are given on the right. MSE values are on the bottom right of error maps.

Figure 3. Qualitative results of our novel views on real scenes. (a) Our tone-mapped HDR views using Photomatix [2]. (b) Our novel LDR views with novel exposures. (c) Ground truth LDR views.

Figure 4. Qualitative results of our novel HDR views on synthetic scenes. All the HDR views are tone-mapped using Photomatix [2]. (a) Our novel HDR views. (b) The novel HDR views by NeRF-GT that a NeRF model is tarined from HDR views. (c) The ground truth HDR views.



(a) *Bathroom* (b) *Bear* (c) *Chair* (d) *Desk*

(e) *Diningroom* (f) *Dog* (g) *Sofa* (h) *Sponza*

(i) *Box* (j) *Flower* (k) *Computer* (l) *Luckycat*

Figure 5. All the discrete CRFs estimated by our method on (a–h) synthetic scenes and (i–l) real-world scenes. On real-world scenes, we calibrate the CRF of digital camera using the method by Debevec and Malik [3].