
Supplementary Material for Learning Neural Exposure Fields for View Synthesis

Michael Niemeyer* Fabian Manhardt Marie-Julie Rakotosaona Michael Oechsle
Christina Tsalicoglou Keisuke Tateno Jonathan T. Barron Federico Tombari
Google

1 Additional Experimental Results

1.1 Additional Results on HDRNeRF Dataset

In Fig. 1 and Fig. 2, we show qualitative results on the HDRNeRF synthetic dataset [3]. We observe that our model not only outperforms top-performing methods quantitatively (see main paper), but also qualitatively our approach is able to accurately produce high-quality view synthesis for provided input exposure. Our method accurately models the input exposure, even for challenging details such as the blinds in the diningroom scene.

Robustness against Noisy Exposure Input. To quantify the robustness of our method toward noise in the input exposure, we ran an ablation study on the HDRNeRF dog scene. We augment the input exposure with Gaussian noise with different standard deviation levels varying from 10^{-5} to 10^{-2} (while exposure is normalized to be in the range $(0, 1]$). We report results in Tab. 1. We observe that our model is robust against noise levels with a standard deviation of up to 10^{-4} . For a noise level of 10^{-3} , results slightly worsen while still exhibiting high SSIM. For a standard deviation of 10^{-2} , the input exposure is too drastically corrupted leading to drop in performance.

Table 1: **Robustness.** We find that our model is fairly robust against noise on the input exposure, producing high-quality results for noise levels with a standard deviation up to 10^{-4} .

Noise Std.	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
0.0	41.25	0.993	0.0008
10^{-5}	41.17	0.993	0.009
10^{-4}	40.71	0.992	0.011
10^{-3}	36.98	0.976	0.029
10^{-2}	15.70	0.513	0.537

Out-of-Distribution Exposure. To evaluate even more extreme out-of-distribution (OOD) lighting conditions as already done in the main paper, we trained our method on the HDRNeRF dog scene where input training images were only sampled with exposure values t_2 and t_4 , and we report evaluation on the test set images for all exposure values between t_1 and t_5 , where $t_1 < t_2 < t_3 < t_4 < t_5$ (see also Section 4.1 of the main paper). Different from the evaluation in Table 1 of the main paper, the exposures t_1 and t_5 are not only not observed during training but even outside of the observed exposure range $[t_2, t_4]$ which depicts an even more extreme case of OOD data.

In Tab. 2 we find that our model produces high-quality images for all sets, including the extreme out-of-distribution (OOD) scenario. Quantitatively, this can be seen in particular in the high SSIM and low LPIPS metrics for all exposure sets. The drop in PSNR, in particular for the extreme OOD scenario, can be explained by the fact that the PSNR metric is directly comparing per-pixel RGB

*Corresponding author: mniemeyer@google.com

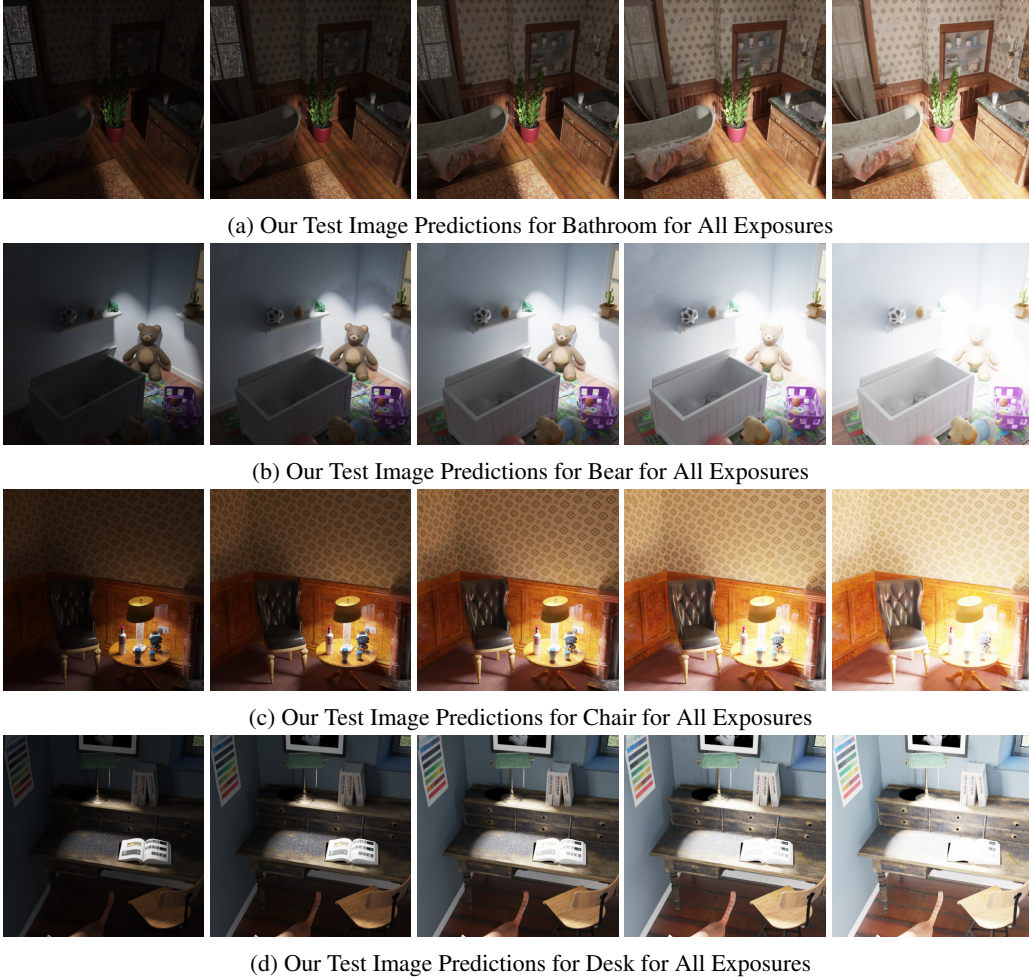


Figure 1: **Qualitative Results on HDRNeRF.** We show additional qualitative results for our method on the test set of the HDRNeRF dataset for all exposure values on all scenes. We find that our method produces high-quality view synthesis results and faithfully reconstructs the test images for different exposure inputs.

values while our model never observes any reference image for such exposure values as they are even outside of the training range.

Table 2: **Out-Of-Distribution Exposure.** We train our model with exposure values t_2 and t_4 and evaluate for in distribution ($\{t_2, t_4\}$), out-of-distribution ($\{t_3\}$), and extreme out-of-distribution exposures ($\{t_1, t_5\}$), finding that for all scenarios, our model produces high-quality results.

Exposure Set	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
ID: $\{t_2, t_4\}$	41.62	0.995	0.006
OOD: $\{t_3\}$	36.17	0.989	0.018
Extreme OOD: $\{t_1, t_5\}$	30.83	0.977	0.033

Architectural Comparison to HDRNeRF. In Tab. 3 we report number of parameters and render time on the HDRNeRF dataset next to the training time (that is also reported in Table 1 of the main paper) for HDR-NeRF [3] and our method. We find that our model trains 49x faster and renders 48x faster. We achieve this large training and test time speed up by incorporating a multi-resolution hash grid backbone [1, 5]. Compared to HDR-NeRF’s MLP architecture, this increases the number of parameters by two orders of magnitude. In most practical use cases, training and in particular test



Figure 2: **Qualitative Results on HDRNeRF.** We show additional qualitative results for our method on the test set of the HDRNeRF dataset for all exposure values on all scenes. We find that our method produces high-quality view synthesis results and faithfully reconstructs the test images for different exposure inputs.

time speed is significantly more important than the parameter count which can also be seen by the large adoption of grid-based architectures since the introduction in InstantNGP [5].

Table 3: **Architectural Comparison to HDRNeRF.** We find that our model trains 49x faster and renders 48x faster compared to HDR-NeRF [3]. We achieve this large training and test time speed up by incorporating a multi-resolution hash grid backbone [1, 5].

	Train Time (min) ↓	Params (mio) ↓	Render Time (sec) ↓
HDR-NeRF [3]	542	1.19	3.39
Ours	11	154.62	0.07

1.2 Additional Results on the Eyeful Tower Dataset

Additional Quantitative Results. In Tab. 4, Tab. 5, and Tab. 6, we report per-scene PSNR, SSIM, and LPIPS for the Eyeful Tower v2 dataset [10]. We observe that for all scenes our method performs best in all metrics except for LPIPS on office view1 where our method still performs better in PSNR and SSIM.

Additional Qualitative Results. In Fig. 3, we show additional qualitative results. Also quantitatively, we can observe that our method achieves the best results.

Additional Comparison to 2D Enhancement Pipeline. In Tab. 7 and Fig. 4, we report results for the following alternative pipeline: We first run a 2D exposure enhancement model [2] on the input images to perform exposure correction. In a subsequent step, we perform 3D reconstruction with the same ZipNeRF [1] backbone as our model uses (see main paper). We find that this baseline produces floating artifacts and overall leads to severe scene degradation compared to our method as the 2D enhanced images are not multi-view consistent. Our approach to instead learn a 3D consistent neural exposure field leads to high-quality scene reconstruction, also with significantly better qualitative (see Fig. 4) and quantitative results (see Tab. 7).

Table 4: **Per-Scene PSNR on Eyeful Tower.**

	apartment	kitchen	office1b	office_view1	office_view2	raf_furnishedroom	riverview	Mean
Vanilla NeRF [1]	16.80	19.17	15.91	14.90	15.07	14.85	20.35	16.72
Affine GLO [1]	23.37	21.45	19.11	20.61	18.07	19.11	19.18	20.13
GLO [4]	24.43	21.14	18.45	22.74	19.65	19.96	22.01	21.20
HDRNeRF* [3]	24.55	22.42	21.86	23.45	21.01	21.81	24.65	22.82
Ours	26.57	27.76	27.97	25.21	23.80	27.55	26.46	26.48

Table 5: **Per-Scene SSIM on Eyeful Tower.**

	apartment	kitchen	office1b	office_view1	office_view2	raf_furnishedroom	riverview	Mean
Vanilla NeRF [1]	0.654	0.745	0.739	0.628	0.653	0.578	0.780	0.682
Affine GLO [1]	0.845	0.854	0.849	0.812	0.803	0.744	0.797	0.815
GLO [4]	0.857	0.850	0.822	0.833	0.823	0.745	0.834	0.824
HDRNeRF* [3]	0.856	0.869	0.887	0.830	0.803	0.763	0.845	0.836
Ours	0.875	0.907	0.923	0.846	0.844	0.845	0.889	0.876

Table 6: **Per-Scene LPIPS on Eyeful Tower.**

	apartment	kitchen	office1b	office_view1	office_view2	raf_furnishedroom	riverview	Mean
Vanilla NeRF [1]	0.439	0.387	0.452	0.510	0.488	0.514	0.321	0.444
Affine GLO [1]	0.239	0.266	0.295	0.298	0.222	0.321	0.200	0.263
GLO [4]	0.249	0.284	0.409	0.306	0.246	0.370	0.223	0.298
HDRNeRF* [3]	0.273	0.295	0.313	0.345	0.314	0.368	0.268	0.311
Ours	0.235	0.252	0.251	0.305	0.211	0.229	0.156	0.234

1.3 Additional Results on Bilarf Dataset

In our main benchmarks, we focus on reporting results for state-of-the-art methods that are comparable to our model. Bilarf [9], as discussed in the main paper, requires, in contrast to our method and baselines, a target reference image produced with HDR software as input, and hence is not directly comparable. We nevertheless ran our method on the statue scene of the Bilarf dataset and find in Tab. 8 that our method overall performs slightly better than Bilarf on its own benchmark. Further, it is worth noting that Bilarf optimizes per-image 3D grids and requires two stage optimization, leading to significant memory requirements for larger captures and longer optimization times.

2 Limitations

While our method demonstrates significant improvements in view synthesis for scenes with varying exposure, it is not without limitations.

Extreme Lighting Conditions. Our method assumes that most areas in the 3D scene show well-exposed colors in some views. In scenarios with extremely challenging lighting conditions, such as very dark scenes with small, extremely bright light sources, or scenes with dramatic and complex cast shadows, the method’s performance might degrade. In this context, it might be forth to investigate our method in combination with recent restoration approaches [8].

Reflective and Transparent Objects. The current implementation does not explicitly address highly reflective or transparent objects. These phenomena can introduce complex lighting effects and view-dependent appearances that may not be fully captured by our model. Future research can involve to

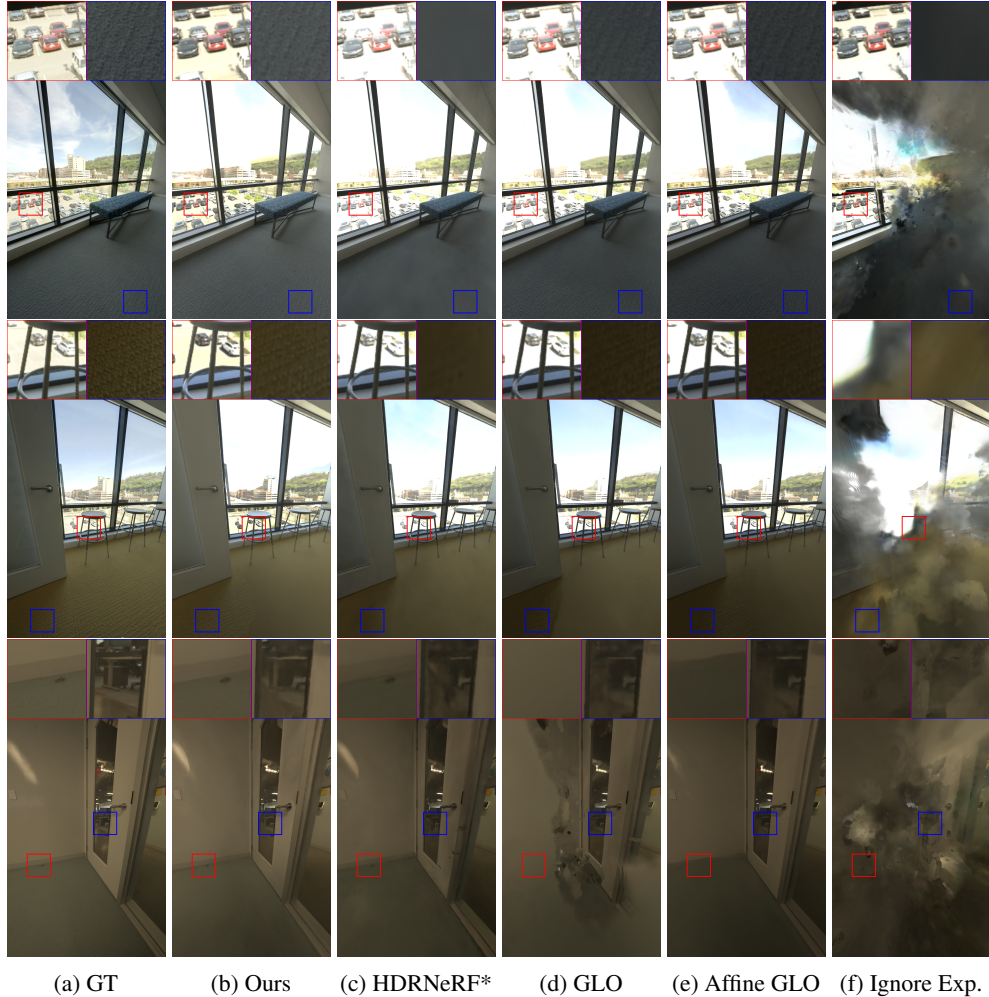


Figure 3: **Additional Qualitative Results on Eyeful Tower.**

Table 7: **Comparison to 2D Enhancement on Eyeful Tower.** We find that first running a 2D exposure enhancement model [2] and subsequently performing 3D reconstruction with the same backbone ZipNeRF [1] as our model leads to degraded performance. The enhanced 2D images are not multi-view consistent leading to floating artifacts and overall degradation of the scene reconstruction.

	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
2D Enhancement [2] + NeRF	13.53	0.656	0.447
Ours	26.48	0.876	0.234

investigate our proposed exposure field and latent conditioning together with radiance field methods that better model highly-reflective scenes [6, 7].

3 Potential Societal Impact

3.1 Potential Positive Societal Impact

Enhanced Realism in VR/AR. Our method significantly advances the quality of reconstructed 3D scenes from real-world captures, especially in challenging conditions with varying exposure. This capability is crucial for creating highly immersive and realistic experiences in virtual reality (VR) and augmented reality (AR) applications. Current neural scene representations often degrade when dealing with per-image variations like strong exposure changes, which are common in real-world



(a) 2D Enhancement [2] + NeRF



(b) Ours

Figure 4: **Qualitative Results for 2D Enhancement Baseline on Eyeful Tower.**

Table 8: **Comparison on Statue of Bilarf Dataset.**

	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Bilarf [9]	23.33	0.826	0.137
Ours	24.23	0.828	0.152

environments such as rooms with windows or indoor/outdoor transitions. Our model can generate virtual environments that faithfully represent real spaces, regardless of complex lighting. This leads to more realistic virtual tours, accurate digital twins for architectural visualization or urban planning, and richer, more engaging interactive experiences in entertainment and training simulations.

Democratizing High-Quality 3D Content Creation. A key advantage of Neural Exposure Fields (NExF) is its ability to produce high-quality, 3D-consistent view synthesis without the need for specialized post-processing steps, multi-exposure captures, or professionally created HDR images. This removes significant barriers to entry for creating advanced 3D content. Traditional methods for handling exposure variations often rely on commercial 2D tonemapping software or require artists to manually adjust HDR representations. NExF, in contrast, optimizes exposure in 3D, bypassing these complex and costly steps. This democratization of high-quality 3D content creation means that filmmakers, game developers, independent artists, and even casual users can achieve professional-grade visual effects and scene reconstructions with more accessible input data (standard RGB images with exposure information).

Advancing Robust Perception for Robotics and Autonomous Systems. The capability of Neural Exposure Fields to reconstruct 3D scenes with robust quality and 3D-consistent appearance in high dynamic range scenarios directly benefits fields like robotics and autonomous driving. Robotic systems, such as autonomous vehicles or industrial robots, rely heavily on accurate 3D perception of their environment for navigation, object detection, and interaction. Real-world operational environments

frequently present challenging lighting conditions, including direct sunlight, strong shadows, and transitions between indoor and outdoor areas, which can cause conventional vision systems to struggle with over- or underexposure. By producing well-exposed, 3D-consistent scene representations, NExF can provide more reliable visual input for these systems.

3.2 Potential Negative Societal Impact

While the primary focus of this research work are significant improvements in view synthesis for real-world captures with varying exposure, it is important to also consider unintended consequences and potential negative societal impact.

Creation of Misleading or Manipulated Content. The improved ability to generate realistic and consistent 3D scenes could potentially be misused to create highly convincing fake imagery or videos, i.e. deep fakes, for malicious purposes.

Privacy Concerns. As the technology enables reconstructing detailed 3D environments from common image data, there is a potential for privacy infringements if applied to sensitive locations or for surveillance purposes without consent.

Computational Resources. Although our method trains faster several prior works, the overall computational resources required for training and deploying such models, especially for large-scale scenes, can still be significant. This could lead to environmental concerns related to energy consumption and may limit access for individuals or organizations with fewer resources.

References

- [1] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields. *Int. Conf. Comput. Vis.*, 2023.
- [2] Xiaojie Guo, Yu Li, and Haibin Ling. LIME: low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.*, 2017.
- [3] Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang. HDR-NeRF: High Dynamic Range Neural Radiance Fields. *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022.
- [4] Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. *IEEE Conf. Comput. Vis. Pattern Recog.*, 2021.
- [5] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *SIGGRAPH*, 2022.
- [6] Dor Verbin, Pratul P. Srinivasan, Peter Hedman, Ben Mildenhall, Benjamin Attal, Richard Szeliski, and Jonathan T. Barron. Nerf-casting: Improved view-dependent appearance with consistent reflections. In *SIGGRAPH Asia*, 2024.
- [7] Fangjinhua Wang, Marie-Julie Rakotosaona, Michael Niemeyer, Richard Szeliski, Marc Pollefeys, and Federico Tombari. Unisdf: Unifying neural representations for high-fidelity 3d reconstruction of complex scenes with reflections. 2024.
- [8] Min Wang, Xin Huang, Guoqing Zhou, Qifeng Guo, and Qing Wang. Bright-NeRF: Brightening Neural Radiance Field with Color Restoration from Low-light Raw Images. *arXiv*, 2024.
- [9] Yuehao Wang, Chaoyi Wang, Bingchen Gong, and Tianfan Xue. Bilateral Guided Radiance Field Processing. *SIGGRAPH*, 2024.
- [10] Linning Xu, Vasu Agrawal, William Laney, Tony Garcia, Aayush Bansal, Changil Kim, Samuel Rota Bulò, Lorenzo Porzi, Peter Kotschieder, Aljaž Božič, Dahua Lin, Michael Zollhöfer, and Christian Richardt. VR-NeRF: High-Fidelity Virtualized Walkable Spaces. *SIGGRAPH Asia*, 2023.