HDR-NeRF--: Learning High Dynamic Range View Synthesis With Unknown Exposure Settings

Nam Nguyen Edward Du Jonathan Ventura
California Polytechnic State University
San Luis Obispo, CA, USA
{nnguy185, eddu, jventu09}@calpoly.edu

Abstract

We propose HDR-NeRF--, a method of learning high dynamic range (HDR) view synthesis from a set of low dynamic range (LDR) views with unknown and varying exposure and white balance. Our method not only renders LDR views that match the ground-truth exposure and white balance, but also renders novel HDR views without ground-truth supervision. HDR-NeRF-- extends HDR-NeRF method in two different ways. First, the exposure parameters are learnable and are optimized separately for each color band. Secondly, we constrain the tone mapping function to be monotonically increasing – meaning that a higher logarithmic radiance corresponds to a brighter pixel value. A quantitative evaluation on benchmark datasets shows that our method outperforms both HDR-NeRF and HDR-Plenoxels in LDR rendering quality.

1. Introduction

Recent studies in rendering radiance fields using deep neural networks – termed neural radiance field (NeRF) – have produced high-quality novel views limited to a low dynamic range [12]. Real-world scenes exceed the dynamic range of the camera, so it has been desirable to reconstruct HDR scenes from the LDR views. An extension to NeRF called HDR-NeRF is able to recover the high dynamic range neural radiance field from a set of varying exposure multi-view LDR images [5]. However, this method requires prior knowledge of the exposure time and consistency in the white balance between training viewpoints, which is not always attainable outside of experimental settings. An HDR radiance fields method called HDR-Plenoxels solves this problem by modeling the physical camera pipeline with explicit radiometric functions in a learnable tone-mapping function [6]. Extended from a well-optimized voxel-based radiance field method called Plenoxels [3], HDR-Plenoxels achieves HDR rendering with fast convergence speed.

We propose a novel method called HDR-NeRF-- to recover the HDR neural radiance field with unknown exposures and varying from LDR images. Since our primary focus is the fidelity of HDR view rendering, we extend our method from the original NeRF architecture, rather than applying a voxel-based approach like Plenoxels [3, 6]. Unlike HDR-NeRF, where the exposure times are read from EXIF metadata, we learn the exposure settings from scratch, and model the camera physical pipeline with a more comprehensive tone-mapper which handles varying exposure time, aperture, ISO and white balance.

In our initial experiments with learnable exposure parameters, we found that sometimes the radiance map would be inverted at the end of training. To address this issue, we enforce a monotonic constraint on the tone-mapping function. The HDR-NeRF method assumes that the tone-mapping function is monotonic, without enforcing this constraint in their architecture design [5].

To evaluate our method, we use datasets from HDR-NeRF [5] and HDR-Plenoxels [6]. We compare our method against HDR-NeRF on scenes from HDR-NeRF datasets, and we compare our method against HDR-Plenoxels on scenes from HDR-Plenoxels datasets. We provide quantitative and qualitative results to justify our main technical contributions. In LDR view rendering, our method outperformed baseline methods on average. In HDR view rendering, our method can reveal the under- and over-exposed regions in the scenes. Our contributions are summarized as follows:

1. We propose an extension to HDR-NeRF to learn HDR neural radiance fields from LDR images of unknown and varying exposure and white balance.

2. We introduce a method to ensure the learned tone-mapping function is monotonically increasing.

3. We outperform previous methods in LDR rendering accuracy.

2. Related work

Neural Radiance Fields. NeRF [12] represents a 3D scene by an implicit continuous function that maps a 3D position and 2D ray direction to a color and density. A pixel is synthesized by integrating over samples along the ray. NeRF-W [9] uses a per-image embedding vector to represent the changes in scene appearance, effectively handling variation in both exposure settings and illumination, a technique has been adopted in many subsequent methods (Cf. [11, 15]).

High Dynamic Range Neural Radiance Fields. HDR-NeRF [5] attempts to capture high dynamic ranges in the real world by explicitly modeling the camera processing pipeline. Specifically, HDR-NeRF adds a learnable tone-mapper that models the camera response function (CRF). HDR-NeRF uses ground-truth information about exposure time to render HDR radiance maps [5]. However, the exposure times of input viewpoints are not always available, such as when using photos from the Internet, or frames extracted from a video. In addition, HDR-NeRF cannot handle varying white balance, which is often observed in casually-captured images and videos with auto-white balance enabled.

High Dynamic Range Plenoxels. HDR-Plenoxels [6] uses a highly-optimized voxel-based volume rendering pipeline introduced in Plenoxels [3] method to recovers HDR neural radiance fields. HDR-Plenoxels learns the CRF and per-image exposure parameters during training. Similar to our method, HDR-Plenoxels does not require ground-truth information about the camera settings, and they handle varying exposure and white balance. However, their method is designed more for fast training and rendering rather than high-quality view synthesis, as they use explicit representations for both the radiance field and the tone mapping function.

3. HDR Neural Radiance Fields without known exposures

In this section, we explain our method HDR-NeRF-- for reconstructing high dynamic range neural radiance fields from multi-view images of unknown and varying exposure and white balance.

3.1. Scene representation

Similar to NeRF, our scenes are represented as a radiance field F within bounded 3D volumes. However, while NeRF's radiance field F outputs the colors and densities, our radiance field F outputs the radiance $\mathbf{e}$ and density σ of the given ray $\mathbf{r}$:

$$(\mathbf{e}(\mathbf{r}), \sigma(\mathbf{r})) = F(\mathbf{r}) \tag{1}$$

3.2. Learned tone-mapping and exposures

We use a multi-layer MLP g to estimate the camera response function (CRF) of a camera. While HDR-NeRF obtains the exposure time from the EXIF metadata, our method represents exposure parameters as a learnable vector of three coefficients corresponding to the three color channels R, G, and B. Specifically, we assume that the exposure time, aperture, ISO gain, and white-balance of a view can all be modeled by a per-channel multiplier to the radiance $\mathbf{e}$. Let $\boldsymbol{\lambda} \in \mathbb{R}^3$ be the exposure parameters of a view. The tone-mapping function f maps the radiance $\mathbf{e}$ of ray $\mathbf{r}$ to into the colors $\mathbf{c}$ given $\boldsymbol{\lambda}$:

$$\mathbf{c}(\mathbf{r}, \boldsymbol{\lambda}) = f(\text{diag}(\boldsymbol{\lambda})\mathbf{e}(\mathbf{r})) \tag{2}$$

Following the CRF calibration method by Debevec and Malik [2], we optimize our tone-mapping function in the logarithmic radiance domain:

$$\mathbf{c}(\mathbf{r}, \boldsymbol{\lambda}) = g(\ln \mathbf{e}(\mathbf{r}) + \ln \boldsymbol{\lambda}) \tag{3}$$

To ensure that the learned function g is monotonic and invertible, we replace the MLP from HDR-NeRF with a monotonic MLP. To constrain an MLP to be monotonic, it is sufficient to use strictly positive weights and strictly monotonic activation functions [4].

3.3. Neural rendering

Similar to HDR-NeRF, we use a conventional volume rendering technique [7] to render the color of each ray. To render HDR views, we skip the tone-mapping operation after obtaining the logarithmic radiance.

3.4. Optimization

Color reconstruction loss. Similar to NeRF, we minimize the mean squared error (MSE) between rendered LDR views to ground-truth LDR views on both the coarse model and the fine model:

$$L_c = \sum_{\mathbf{r} \in R} ||\hat{\mathbf{c}}_c(\mathbf{r}, \boldsymbol{\lambda}) - \mathbf{c}(\mathbf{r}, \boldsymbol{\lambda})||_2^2 + ||\hat{\mathbf{c}}_f(\mathbf{r}, \boldsymbol{\lambda}) - \mathbf{c}(\mathbf{r}, \boldsymbol{\lambda})||_2^2 \tag{4}$$

where $\mathbf{c}$ is the ground-truth color of each ray, $\hat{\mathbf{c}}_c$ and $\hat{\mathbf{c}}_f$ are the predicted colors from the coarse and fine models, respectively.

Unit exposure loss. Similar to HDR-NeRF, our method also fixes the scale factor α to which the radiance e is recovered. We fix the value of $g(0)$ to be 0.5, the midpoint of the normalized pixel value on real-world scenes. We define our exposure loss to be:

$$L_u = ||g(0) - 0.5||_2^2 \tag{5}$$

Finally, our loss function is the combination of the color reconstruction loss and the unit exposure loss:

$$L = L_c + w_u L_u \tag{6}$$

where w_u is the weight of unit exposure loss. We choose 0.5 to be a default value of w_u

4. Experiments

4.1. Experimental settings

We use a similar architecture to HDR-NeRF [5]. We use an MLP with eight layers and 256 channels to predict the radiance and the density at points in the volume. Our tone-mapper consists of a one-layer MLP of width 128 for each channel. To enforce the monotonic constraint to the tone-mapper, we take the absolute value of the tone-mapper's weights and use the ReLU activation function[1]. We use the Adam optimizer [8] with a learning rate that decays exponentially from 5×10^{-4} to 5×10^{-5}. We optimize a model for 200K iterations on a single NVIDIA Tesla V100 GPU, which runs for approximately 16 hours.

4.2. Evaluation metrics and datasets

Datasets. We use four real scenes from the HDR-NeRF [5] to evaluate our method's ability to learn the HDR radiance fields from LDR images of varying exposure. These scenes were captured using a digital SLR camera, using exposure bracketing with five different exposure times. White balance is kept fixed in this dataset.

We use datasets from the HDR-Plenoxels [6] to evaluate our method's ability to learn HDR radiance fields from LDR images of varying exposure and white balance. The HDR-Plenoxel datasets consists of five synthetic scenes generated from Blender and four real scenes captured from a digital SLR camera using exposure and white balance bracketing.

Metrics. We employ three metrics for our quantitative comparison between LDR synthesized views and the ground-truth views: PSNR, SSIM, and LPIPS [16]. Higher PSNR and SSIM values are better, and a lower LPIPS value is better. However, our metrics are not calculated on the entire test view. Following the evaluation methodology of HDR-Plenoxels [6], since we cannot predict the exposure parameters on test views, we use the left half of the test image for training and learning the exposure parameters. Then, we evaluate the performance on the unseen right half. Our tone-mapping operator for HDR qualitative results is μ-law, a simple and established tone-mapping operator used by HDR-NeRF [5] and other works [1, 10, 13]. This tone-mapping operator is $M(E) = \log(1 + \mu E)/\log(1 + \mu)$ where E is the HDR pixel value normalized to the range $[0, 1]$, and μ is the compression factor, which is set to 1.0 for best-looking results.

[1]ReLU is not strictly monotonic but in practice, we found it produces the best results.

4.3. Evaluation results

HDR-NeRF real dataset. Our baseline method for comparison is HDR-NeRF [5], which outperformed NeRF [12] and NeRF-W [9] methods in LDR rendering. The quantitative results of LDR novel view synthesis in Tab. 1 showed that our method outperformed HDR-NeRF on flower and luckycat scenes, and performed comparably on the others. On average, our method outperforms HDR-NeRF in PSNR (35.96 versus 35.34), and is comparable to HDR-NeRF in SSIM (0.952 versus 0.956) and LPIPS (0.072 versus 0.068). Note that HDR-NeRF uses the exposure times given in the EXIF metadata, while our method learns the exposure settings from scratch. Sample LDR and HDR views are shown in Fig. 1 and in the supplemental material.

HDR-Plenoxels real and synthetic datasets. Our baseline method for comparison is HDR-Plenoxels, which outperforms the original Plenoxels method [3] and Approximate Differentiable One-Pixel Point Rendering (ADOP) [14] in LDR rendering. The quantitative results in Tab. 2 and Tab. 3 shows that our method outperforms HDR-Plenoxels on most of the scenes. On average, our method achieved higher metrics than HDR-Plenoxels (PSNR: 30.14 versus 28.73, SSIM: 0.896 versus 0.891, LPIPS: 0.139 versus 0.294). Sample renderings are shown in Fig. 2 and in the supplemental material.

5. Conclusions and Future Work

We present a novel method of learning HDR view synthesis from LDR views with unknown and varying exposure and white balance. Besides rendering novel HDR views without ground-truth HDR views, our method can render LDR views that match the ground-truth exposure and white balance. Our method outperformed previous methods in view synthesis quality. The code and models will be publicly available.

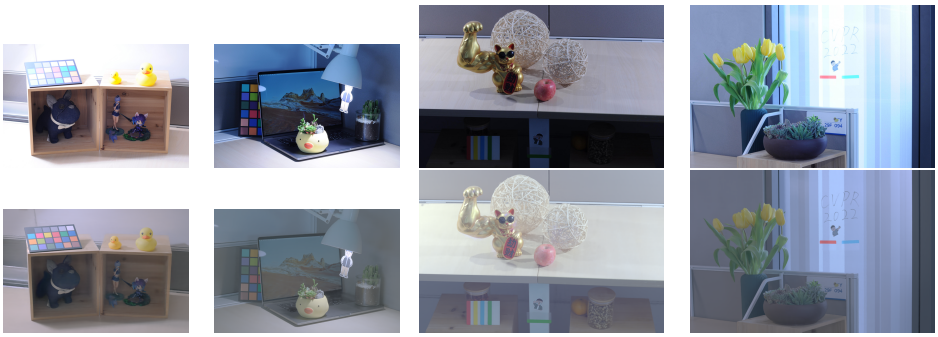
Modeling even more components of the camera pipeline, such as the quantization and compression steps, could further improve the reconstruction. Future work could consider how to handle more heavily post-processed images which have a spatially-varying tone mapping function (e.g. tone mapping faces differently from the background). Future work also lies in analyzing how well an HDR scene can be recovered from autoexposure video which doesn't systematically sample the dynamic range of the scene.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 21448822.

Table 1. Results of quantitative evaluation on **real** scenes from **HDR-NeRF**. Values are the average of the metrics for the test data of each scene. The best value of each metric are shown in **bold**.

Type Method	Computer			Flower			Luckycat			Box		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
HDR-NeRF	35.17	**0.945**	**0.093**	34.98	0.964	0.050	35.31	0.948	0.072	**35.90**	**0.964**	**0.056**
Ours	**35.34**	0.942	0.096	**36.81**	**0.971**	**0.041**	**36.04**	**0.950**	**0.069**	35.66	0.945	0.098



(a) box (b) computer (c) luckycat (d) flower

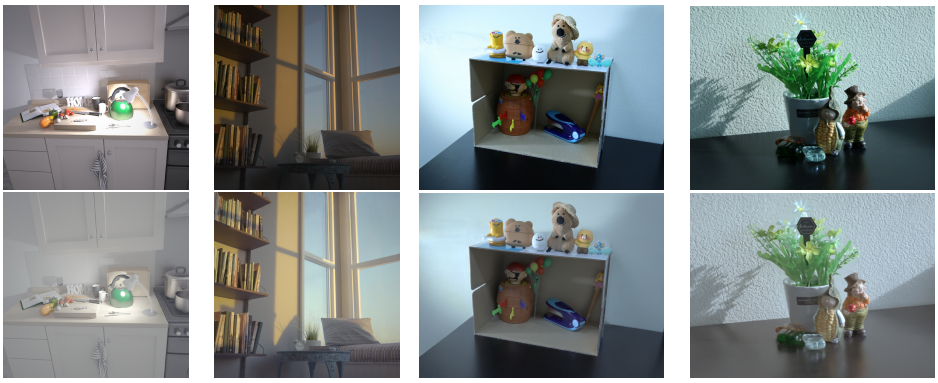
Figure 1. Ground-truth LDR views (top row) and our tone-mapped HDR views (bottom row) from the **HDR-NeRF** dataset. Our tone-mapped HDR views reveal the under-exposed and over-exposed regions of the scenes.

Table 2. Results of quantitative evaluation on **synthetic** scenes from **HDR-Plenoxels.**

Type Method	Book			Classroom			Monk			Room			Kitchen		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
HDR-Plenoxels	27.49	0.837	0.292	**29.87**	**0.908**	0.284	**28.27**	**0.852**	0.297	28.70	0.900	0.291	31.53	0.936	0.156
Ours	**28.62**	**0.849**	**0.201**	28.61	0.877	**0.169**	27.36	0.812	**0.232**	**33.21**	**0.940**	**0.079**	**35.30**	**0.961**	**0.068**

Table 3. Results of quantitative evaluation on **real** scenes from **HDR-Plenoxels.**

Type Method	Character			Desk			Plant			Coffee		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
HDR-Plenoxels	33.14	**0.960**	0.343	**28.32**	**0.907**	0.312	24.27	0.790	0.369	27.40	0.928	0.269
Ours	**33.47**	0.956	**0.092**	27.67	0.870	**0.160**	**28.28**	**0.865**	**0.166**	**29.25**	**0.930**	**0.096**



(a) kitchen - syn (b) room - syn (c) character - real (d) plant - real

Figure 2. Ground truth LDR views (top row) and our tone-mapped HDR views (bottom row) from the **HDR-Plenoxels** dataset.

References

[1] Prashant Chaudhari, Franziska Schirrmacher, Andreas Maier, Christian Riess, and Thomas Köhler. Merging-isp: Multi-exposure high dynamic range image signal processing. In *Pattern Recognition: 43rd DAGM German Conference, DAGM GCPR 2021, Bonn, Germany, September 28–October 1, 2021, Proceedings*, pages 328–342. Springer, 2022. 3

[2] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, page 369–378, USA, 1997. ACM Press/Addison-Wesley Publishing Co. 2

[3] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5501–5510, 2022. 1, 2, 3

[4] Chin-Wei Huang, David Krueger, Alexandre Lacoste, and Aaron Courville. Neural autoregressive flows. In *International Conference on Machine Learning*, pages 2078–2087. PMLR, 2018. 2

[5] Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang. Hdr-nerf: High dynamic range neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18398–18408, 2022. 1, 2, 3

[6] Kim Jun-Seong, Kim Yu-Ji, Moon Ye-Bin, and Tae-Hyun Oh. Hdr-plenoxels: Self-calibrating high dynamic range radiance fields. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII*, pages 384–401, 2022. 1, 2, 3

[7] James T. Kajiya and Brian P Von Herzen. Ray tracing volume densities. In *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '84, page 165–174, New York, NY, USA, 1984. Association for Computing Machinery. 2

[8] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 3

[9] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7210–7219, 2021. 2, 3

[10] Nico Messikommer, Stamatios Georgoulis, Daniel Gehrig, Stepan Tulyakov, Julius Erbach, Alfredo Bochicchio, Yuanyou Li, and Davide Scaramuzza. Multi-bracket high dynamic range imaging with event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 547–557, 2022. 3

[11] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16190–16199, 2022. 2

[12] 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 *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I*, pages 405–421, 2020. 1, 2, 3

[13] K Ram Prabhakar, Susmit Agrawal, Durgesh Kumar Singh, Balraj Ashwath, and R Venkatesh Babu. Towards practical and efficient high-resolution hdr deghosting with cnn. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*, pages 497–513. Springer, 2020. 3

[14] Darius Rückert, Linus Franke, and Marc Stamminger. Adop: Approximate differentiable one-pixel point rendering. *ACM Transactions on Graphics (TOG)*, 41(4):1–14, 2022. 3

[15] Matthew Tancik, Vincent Casser, Xinchen Yan, Sabeek Pradhan, Ben Mildenhall, Pratul P Srinivasan, Jonathan T Barron, and Henrik Kretzschmar. Block-nerf: Scalable large scene neural view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8248–8258, 2022. 2

[16] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 3