HDR-NeRF: High Dynamic Range Neural Radiance Fields

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

(a) Input views (b) Novel LDR views (c) One of the novel HDR views

Figure 1. We recover a high dynamic range neural radiance field from (a) multiple LDR views with different exposures. Our system is able to render (b) novel LDR views with arbitrary exposures and (c) novel HDR views.

Abstract

We present High Dynamic Range Neural Radiance Fields (HDR-NeRF) to recover an HDR radiance field from a set of low dynamic range (LDR) views with different exposures. Using the HDR-NeRF, we are able to generate both novel HDR views and novel LDR views under different exposures. The key to our method is to model the simplified physical imaging process, which dictates that the radiance of a scene point transforms to a pixel value in the LDR image with two implicit functions: a radiance field and a tone mapper. The radiance field encodes the scene radiance (values vary from 0 *to* $+\infty$*), which outputs the density and radiance of a ray by giving corresponding ray origin and ray direction. The tone mapper models the mapping process that a ray hitting on the camera sensor becomes a pixel value. The color of the ray is predicted by feeding the radiance and the corresponding exposure time into the tone mapper. We use the classic volume rendering technique to project the output radiance, colors and densities into HDR and LDR images, while only the input LDR images are used as the supervision. We collect a new forward-facing HDR dataset to evaluate the proposed method. Experimental results on synthetic and real-world scenes validate that our method can not only accurately control the exposures of synthesized views but also render views with a high dynamic range.*

1. Introduction

Novel view synthesis is one of the most pursued topics in computer graphics and computer vision. Limited by the dynamic range of camera sensors and input views, rendered novel views are often with a low dynamic range, while human eyes are able to perceive a much higher dynamic range than what is possible by a regular camera. It is therefore highly desirable to render novel HDR views to improve the overall visual experience.

Recently, a series of works have been focused on recovering the radiance field of a scene to render photorealistic novel views using deep neural networks [3,37,42,62]. They implicitly encode volumetric densities and colors using a multi-layer perceptron (MLP), which is termed *neural radiance field* (NeRF). These methods produce high-quality novel views, yet the dynamic range of the obtained radiance in NeRF is limited to a low dynamic range (between

*Work done during an internship at Tencent AI Lab.

0 and 255), while the radiance in the physical world scene often covers a much broader (higher) dynamics range (*e.g.* from 0 to $+\infty$). We also notice that 'NeRF in the Dark' tries to recover radiance field from raw images with noise [40], while it's different from our method.

High Dynamic Range (HDR) imaging is the set of techniques that recover HDR images from multiple LDR images with different exposures [55]. The most common way to reconstruct HDR images is to take a series of LDR images with different exposures at a fixed camera pose and then merge those LDR images into an HDR image [12, 39, 48]. These methods produce compelling results for tripod-mounted cameras but may lead to ghost artifacts when the camera is hand-held. To overcome the limitations of conventional multi-exposure stack-based HDR synthesis, some deep learning methods have been proposed to solve this problem via a two-stage approach [26, 59]: 1) aligning the input LDR images using optical flow or removing plausible motion regions, 2) merging the processed images into an HDR image. However, in cases with large motion, their approach typically introduces artifacts in the final results. Most critically, these HDR imaging methods are unable to render novel views and the learning-based methods require HDR images as training supervision. To render novel views, some methods try to merge image-based rendering and HDR imaging techniques. [30, 35, 49, 51]. However, the image-based methods struggle from preserving view consistency.

In this paper, we propose a method HDR-NeRF to recover the high dynamic range neural radiance field from a set of LDR images (Fig. 1a) with various *exposures* (the exposure is defined as the product of exposure time and radiance). To the best of our knowledge, this is the first end-to-end neural rendering system that can render novel HDR views (Fig. 1c) and control the exposure of novel LDR views (Fig. 1b). Building upon NeRF, we introduce a differentiable tone mapper to model the process that radiance in the scene becomes pixel values in the image. We use an MLP to model the tone-mapping operation. Overall, HDR-NeRF can be represented by two continuous implicit neural functions: a *radiance field* for density and scene radiance and a *tone mapper* for color, as shown in Fig. 2. Our pipeline enables joint learning of the two implicit functions, which is critical to recovering the HDR radiance field from such sparse sampled LDR images. We use the classical volume rendering technique [25] to accumulate radiance, colors, and densities into HDR and LDR images, but we only use LDR ground truth as supervision.

To evaluate our method, we collect a new HDR dataset that contains synthetic scenes and real-world scenes. We compare our method with original NeRF [42], NeRF-W (NeRF in the wild) [37], as well as NeRF-GT (a version of NeRF that is trained from LDR images with consistent exposures or HDR images). We provide quantitative and qualitative results and ablation studies to justify our main technical contributions. Our method achieves similar scores across all major metrics on this dataset compared with NeRF-GT. Besides, compared to the recent state-of-the-art NeRF and NeRF-W, our method can render LDR novel views with arbitrary exposures and spectacular novel HDR views. The main contributions of this paper can be summarized as follows:

1. An end-to-end method HDR-NeRF is proposed to recover the high dynamic range neural radiance field from multiple LDR views with different amounts of exposure.

2. The camera response function is modeled, both HDR views and LDR views with varying exposures are rendered from the radiance field.

3. A new HDR dataset including synthetic and real-world scenes is collected. Compared with SOTAs, our method achieves the best performance on this dataset. The dataset and code will be released for further research purposes in this community.

2. Related Work

Novel View Synthesis. Novel view synthesis aims to generate novel images from a new viewpoint using a set of input views. It is a typical application of image-based rendering technique [52], such as rendering novel views using depth [6, 7, 43, 67, 68] or explicit geometry information [13, 22, 23, 63]. Many classic IBR methods estimate radiance of input images using HDR imaging methods to render novel HDR views [30, 35, 49, 51]. The estimated radiance using HDR imaging methods is always image-wise. It may be hard to preserve the view consistency in challenging scenes. On the other hand, light field rendering methods interpolate views based on implicit soft geometry estimates derived from densely sampled images [5, 11, 19, 31, 38].

In recent years, deep learning techniques have been applied to novel view synthesis to get high-quality photorealistic views. These learning-based approaches can be classified into three categories according to scene representation models. The first category aims to combine the convolutional neural network (CNN) with traditional voxel grid representation [9, 34, 53], such that Sitzmann *et al*. [53] use a CNN to compensate the discretization artifacts from low resolution voxel grids. Lombardi *et al*. [34] control the predicted voxel grids based on the input time of dynamic scene. Inspired by the layered depth images [50], other learning-based methods focus on training a CNN to predict a multi-plane images representation from a set of input images and render novel views using alpha-compositing [10, 17, 41, 66]. These methods predict multi-planes images

to synthesize views for specific applications, such as light-field rendering [41] and baseline magnification [66]. The third category is the NeRF family which represents a scene with a neural radiance field [3, 4, 8, 33, 36, 37, 42, 62]. Although these recent methods achieved high-quality of rendered novel views, none of them has tackled the task of synthesizing a novel view with *high dynamic range*.

Neural Implicit Representation. Recently, there has been a surge in representing 3D scenes in implicit functions via a neural network. Compared to traditional explicit representations, such as point cloud [47], voxels [18] and octrees [57], neural implicit representations have shown high-quality view synthesis results such as continuous and high-fidelity. We focus on the neural radiance fields representation that implicitly models the volume densities and colors of the scenes with MLPs [42]. NeRF approximates a continuous 3D function by mapping from an input 5D location to scene properties. Recently, NeRF has been explored for novel view relighting [4, 54], view synthesis for dynamic scenes [14, 32, 33, 44, 46, 58], scene editing [21, 37, 61, 64]. Particularly, Martin-Brualla *et al*. [37] propose NeRF-W to build NeRF from internet photo collections with different photometric variations and occlusions. They learn a per-image latent embedding to capture photometric appearance variations in training images, which enable them to modify the lighting and appearance of a rendering. Although various extensions have been explored to NeRF, which enables them to effectively represent the scene radiance captured by cameras. However, all the NeRF based methods ignore the physics process from radiance to pixel values, which hinders them from representing the radiance in the real world.

High Dynamic Range Imaging. Traditional multiple exposures-based HDR imaging methods reconstruct HDR images by calibrating the CRF from an exposure stack that a series of LDR images under different exposures with a same pose [12] or directly merge the LDR images into an HDR image [39]. To overcome the limitations of traditional methods, such as ghosting in the HDR results when LDR images are captured by a hand-held camera or on a dynamic scene, some methods are proposed to detect the motion regions in the LDR images and then remove these regions in the fusion [20, 24]. In contrast, alignment-based methods align the input multiple LDR images by estimating optical flow then merge the aligned images [26, 56, 60]. Depending on the great potential of deep learning, some methods try to reconstruct an HDR image from a single LDR image [16, 27, 28]. However, most HDR imaging methods require the given LDR images with a fixed or quasi-fixed camera pose. Besides, these methods can only synthesize HDR images with original poses and require ground truth HDR images to supervise.

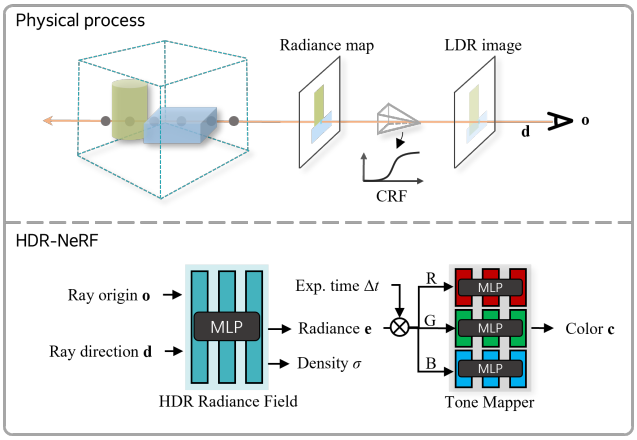


Figure 2. The pipeline of HDR-NeRF modeling the simplified physical process. Our method is consisted of two modules: an HDR radiance field models the scene for radiance and densities and a tone mapper models the CRF for colors.

3. Background

3.1. Neural Radiance Fields

NeRF [42] represents a scene using an implicit neural function, which maps a ray origin $\mathbf{o} = (x, y, z)$ and ray direction $\mathbf{d} = (\theta, \phi)$ into a color $\mathbf{c} = (r, g, b)$ and density σ, that is $(\mathbf{o}, \mathbf{d}) \rightarrow (\mathbf{c}, \sigma)$. Specifically, suppose a camera ray $\mathbf{r}$ is emitted from camera center $\mathbf{o}$ with direction $\mathbf{d}$, *i.e.* $\mathbf{r}(s) = \mathbf{o} + s\mathbf{d}$ where s denotes a position along the ray. The expected color $\widehat{C}(\mathbf{r})$ of $\mathbf{r}(s)$ is defined as:

$$\widehat{C}(\mathbf{r}) = \int_{s_n}^{s_f} T(s)\sigma(\mathbf{r}(s))\mathbf{c}(\mathbf{r}(s), \mathbf{d})\, ds, \tag{1}$$

$$T(s) = \exp\left(-\int_{s_n}^{s} \sigma(\mathbf{r}(p))\, dp\right), \tag{2}$$

where s_n and s_f denote the near and far boundary of the ray respectively, and $T(s)$ denotes an accumulated transmittance. The predicted pixel value is then compared to the ground truth $C(\mathbf{r})$ for optimization. For all the camera rays of the target view with a pose $\mathbf{P}$, the color reconstruction loss is thus defined by

$$\mathcal{L} = \sum_{\mathbf{r} \in \mathcal{R}(\mathbf{P})} \|\widehat{C}(\mathbf{r}) - C(\mathbf{r})\|^2, \tag{3}$$

where $\mathcal{R}(\mathbf{P})$ is a set of camera rays at target position $\mathbf{P}$.

In practice, naively feeding 5D coordinates into the MLP results in renderings that struggle from representing high-frequency variation in color and geometry. To tackle this problem, a positional encoding strategy is adopted in NeRF. Besides, NeRF simultaneously optimizes two models, where the densities predicted by the coarse model are used to bias the sample of a ray in the fine model.

3.2. Camera Response Functions

In most imaging devices, the incoming irradiance is mapped into pixel values and stored in images by a series of linear and nonlinear image processing (*e.g.* white balance). In general, all the image processing can be combined in a single function f called *camera response function* (CRF) [15]. It's hard to know the CRFs of cameras beforehand, because they are intentionally designed by the camera manufacturers. Taking ISO gain and aperture as implicit factors, without loss of generality, the nonlinear mapping can be modeled as [55]:

$$Z = f(H\Delta t), \tag{4}$$

where H is irradiance, the total amount of light incident on a camera sensor, Z denotes the pixel value, and Δt denotes exposure time which is decided by the shutter speed. Note that, in the neural radiance field, the integration of scene radiance over the lens aperture is ignored and the irradiance is considered as radiance [15].

4. HDR Neural Radiance Fields

In this section, we introduce our method HDR-NeRF for recovering high dynamic range neural radiance fields. As shown in Fig. 2, our method consists of two main modules to be described in this section. Our goal is to recover the real radiance field in which the radiance is between 0 and $+\infty$ by using the LDR images with different exposures as supervision. The main challenge is how to efficiently aggregate information in the LDR images to get an HDR radiance field.

4.1. Scene Representation

To render novel HDR views, we represent the scene as an HDR radiance field within a bounded 3D volume. An MLP F called *radiance field* is used to model the HDR scene radiance, which is similar to NeRF. For a given ray origin $\mathbf{o}$ and ray direction $\mathbf{d}$, the *radiance field* F outputs the radiance $\mathbf{e}$ and density σ of the ray $\mathbf{r}(s) = \mathbf{o} + s\mathbf{d}$, which is formulated as:

$$(\mathbf{e}(\mathbf{r}), \sigma(\mathbf{r})) = F(\mathbf{r}). \tag{5}$$

Note that, the outputs of implicit function in NeRF are colors and densities, while our outputs are radiance and densities.

4.2. Learned Tone-mapping

Representing a scene with an HDR radiance field, the key is how to ensure *radiance field* outputs the radiance of ray without the HDR ground truth as supervision. Inspired by the CRF calibration that the process of determining the mapping between the digital value of a pixel and the corresponding irradiance (up to a scale factor), a *tone mapper* is introduced to model the nonlinear mapping of HDR rays to LDR rays. Specifically, we use an MLP f to estimate the CRF of a camera and map our predicted radiance into colors. According to Eq. (4), our predicted radiance $\mathbf{e}$ by Eq. (5) is then tone-mapped into color $\mathbf{c}$. We formulate the differentiable tone-mapping operation as:

$$\mathbf{c}(\mathbf{r}, \Delta t) = f(\mathbf{e}(\mathbf{r})\Delta t(\mathbf{r})), \tag{6}$$

where $\Delta t(\mathbf{r})$ denotes the exposure time of a camera for capturing the ray $\mathbf{r}$. We can easily read exposure time from the EXIF files that contain metadata about photos, such as exposure time, focal length, f-number, *etc*. In practice, the RGB channels of images are tone-mapped with different CRFs, hence three MLPs are used in our method to process each channel independently.

Following the classical nonparametric CRF calibration method by Debevec and Malik [12], we transform all the images into a logarithm radiance domain to optimize the network. Specifically, we assume the *tone mapper* f is monotonic and invertible, so we can rewrite Eq. (6) as:

$$\ln f^{-1}\left(\mathbf{c}(\mathbf{r}, \Delta t)\right) = \ln \mathbf{e}(\mathbf{r}) + \ln \Delta t(\mathbf{r}). \tag{7}$$

We then present the inverse function of $\ln f^{-1}$ as g, thus:

$$\mathbf{c}(\mathbf{r}, \Delta t) = g\left(\ln \mathbf{e}(\mathbf{r}) + \ln \Delta t(\mathbf{r})\right), \tag{8}$$

where $g = (\ln f^{-1})^{-1}$. As a result, our *tone mapper* function is transformed to function g with a logarithm radiance domain.

4.3. Neural Rendering

We use the conventional volume rendering technique [25] to render the color of each ray passing through the scene. Combining the *radiance field* module and *tone mapper* module, we substitute Eq. (8) into Eq. (1). The expected color $\widehat{C}(\mathbf{r}, \Delta t)$ of ray $\mathbf{r}(s)$ with near and far bounds s_n and s_f is given by:

$$\widehat{C}(\mathbf{r}, \Delta t) = \int_{s_n}^{s_f} T(s)\sigma(\mathbf{r}(s))g\left(\ln \mathbf{e}(\mathbf{r}(s)) + \ln \Delta t(\mathbf{r})\right)\, ds, \tag{9}$$

where $T(s)$ is defined in Eq. (2). To render HDR views, the tone-mapping operation is removed. Similarly, an HDR pixel value is approximated as:

$$\widehat{E}(\mathbf{r}) = \int_{s_n}^{s_f} T(s)\sigma(\mathbf{r}(s))\mathbf{e}(\mathbf{r}(s))\, ds. \tag{10}$$

4.4. Optimization

Color reconstruction loss. To optimize the two implicit functions F and g from input LDR images, we minimize the mean squared error (MSE) between the LDR views rendered by HDR-NeRF and the ground truth LDR views. Similar to NeRF, we simultaneously optimize a coarse

model and a fine model. The color reconstruction loss is formulated as:

$$\mathcal{L}_c = \sum_{\mathbf{r}\in\mathcal{R}(\mathbf{P})} \|\widehat{C}_c(\mathbf{r},\Delta t) - C(\mathbf{r},\Delta t)\|_2^2 + \|\widehat{C}_f(\mathbf{r},\Delta t) - C(\mathbf{r},\Delta t)\|_2^2, \tag{11}$$

where C is the ground-truth color of each pixel, and $\widehat{C}_c$ and $\widehat{C}_f$ are the color predicted by the coarse model and fine model respectively.

Unit exposure loss. Our method recovers radiance $\mathbf{e}$ up to an unknown scale factor α (*i.e.*, $\alpha\mathbf{e}$) via the color reconstruction loss. It is equivalent to add a shift $\ln\alpha$ to the independent variable of function g, according to Eq. (8), as shown in Fig. 7d. As a consequence, we need to add an additional constraint to fix the scale factor α. Specifically, we fix the value of $g(0)$ to C_0, and the unit exposure loss is defined as:

$$\mathcal{L}_u = \|g(0) - C_0\|_2^2. \tag{12}$$

The meaning of this constraint is that the pixels with the value C_0 are assumed to have a unit exposure. However, C_0 is usually unknown in practice. We generally set the C_0 as the midway of the pixel value on real-world scenes.

Finally, our HDR-NeRF is end-to-end optimized using the following loss:

$$\mathcal{L} = \mathcal{L}_c + \lambda_u \mathcal{L}_u, \tag{13}$$

where λ_u denotes the weight of unit exposure loss.

5. Experiments

5.1. Implementation Details

In training and testing phases, an eight-layer MLP with 256 channels is used to predict radiance $\mathbf{e}$ and densities σ, and three one-layer MLPs with 128 channels to predict RGB values of color c respectively. We sample 64 points along each ray in the coarse model and 128 (64) points in the fine model on synthetic (real) dataset. The batch size of rays is set to 1024. As with NeRF, positional encoding [42] is applied for ray origins and ray directions. We fix the loss weight $\lambda_u = 0.5$ throughout the paper. The high parameter C_0 is 0.5 on real scenes. To compare with ground truth HDR views, we set $C_0 = C_0^{GT}$ on synthetic scenes, where C_0^{GT} denotes the pixel value of ground truth CRF when input logarithm radiance is 0. We use Adam optimizer [29] (default values $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-7}$) with a learning rate 5×10^{-4} that decays exponentially to 5×10^{-5} over the course of optimization. We optimize a single model for 200K iterations on a single NVIDIA V100 GPU (about one day).

5.2. Evaluation Dataset and Metrics.

Dataset. We evaluate the proposed method on our collected HDR dataset that contains 8 synthetic scenes rendered with Blender [1] and 4 real scenes captured by a digital camera. Images are collected at 35 different poses in the real dataset, with 5 different exposure time $\{t_1, t_2, t_3, t_4, t_5\}$ at each pose. For the synthetic dataset, we render 35 HDR views for each scene and build a tone-mapping function to map these HDR views into LDR images as our inputs (described in supplementary material). The pre-defined tone-mapping function can also be used to evaluate the discrete CRFs estimated by our *tone mapper*. We select 18 views with different poses as the training dataset. The exposure time of each input view is randomly selected from $\{t_1, t_3, t_5\}$. 34 views with exposure time t_3 or t_4 at the other 17 poses are chosen as our test dataset. Besides, the HDR views are also used for test. The resolution of each view is 400×400 pixels for synthetic scenes and 804×534 pixels for real scenes.

Metrics. We report quantitative performance using PSNR (higher is better) and SSIM (higher is better) metrics, as well as the state-of-the-art LPIPS [65] (lower is better) perceptual metric, which is based on a weighted combination of neural network activations tuned to match human judgments of image similarity [41]. Since HDR images are usually displayed after a tone-mapping operation, we quantitatively evaluate our HDR views in the tone-mapped domain via the μ-law, *i.e.* a simple and canonical operator that is widely used for benchmarking in HDR imaging [26,45,59]. The tone-mapping operation is:

$$M(E) = \frac{\log(1+\mu E)}{\log(1+\mu)}, \tag{14}$$

where μ defines the amount of compression and is always set to 5000, and E denotes an HDR pixel value which is always scaled to the range $[0, 1]$. To properly show the details in each HDR image for qualitative evaluations, all the HDR results are tone-mapped with Photomatix [2].

5.3. Evaluation

Baselines. We compare our method against the following baseline methods. 1) NeRF [42]: the original NeRF method. 2) NeRF-W [37]: unofficial implementation of NeRF in the wild with PyTorch. NeRF-W controls the appearance of rendered views by linearly interpolating their learned appearance vectors, which means that we can not render views by giving the novel exposure time we expect. To facilitate the comparison, the exposure time of input views for NeRF-W are chosen randomly from all the five exposure settings in order to learn five appearance vectors for testing. 3) NeRF-GT (the upper bound of our method): NeRF model trained from LDR views with a consistent exposure or HDR views. 4) Ours† (an ablation study): our method that models the tone-mapping operations of RGB channels with a single MLP.

Comparisons. The quantitative results of rendered novel views on our dataset are shown in Tab. 1. Our method outperforms NeRF and NeRF-W on both synthetic and real

Table 1. Quantitative comparisons with baseline methods on synthetic and real scenes. Metrics are averaged over the scenes from our dataset (per-scene metrics are shown in supplementary material). LDR-OE denotes the LDR results with exposure t_1, t_3, and t_5. LDR-NE denotes the LDR results with exposure t_2, and t_4. HDR denotes the HDR results. We color code each column as best and second best.

		LDR-OE (t_1, t_3, t_5)			LDR-NE (t_2, t_4)			HDR		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF [42]	Syn.	13.97	0.555	0.376	—	—	—	—	—	—
	Real	14.95	0.661	0.308	—	—	—	—	—	—
NeRF-W[1][37]	Syn.	29.83	0.936	0.047	29.22	0.927	0.050	—	—	—
	Real	28.55	0.927	0.094	28.64	0.923	0.089	—	—	—
NeRF-GT[2][42]	Syn.	37.66	0.965	0.028	35.87	0.955	0.032	37.80	0.964	0.029
	Real	34.55	0.958	0.057	34.59	0.956	0.051	—	—	—
Ours†	Syn.	—	—	—	—	—	—	—	—	—
	Real	30.37	0.944	0.075	29.37	0.938	0.078	—	—	—
Ours	Syn.	39.07	0.973	0.026	37.53	0.966	0.024	36.40	0.936	0.018
	Real	31.63	0.948	0.069	31.43	0.943	0.069	—	—	—

[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.
† An ablation study of our method that models the tone-mapping operations of RGB channels with a single MLP.

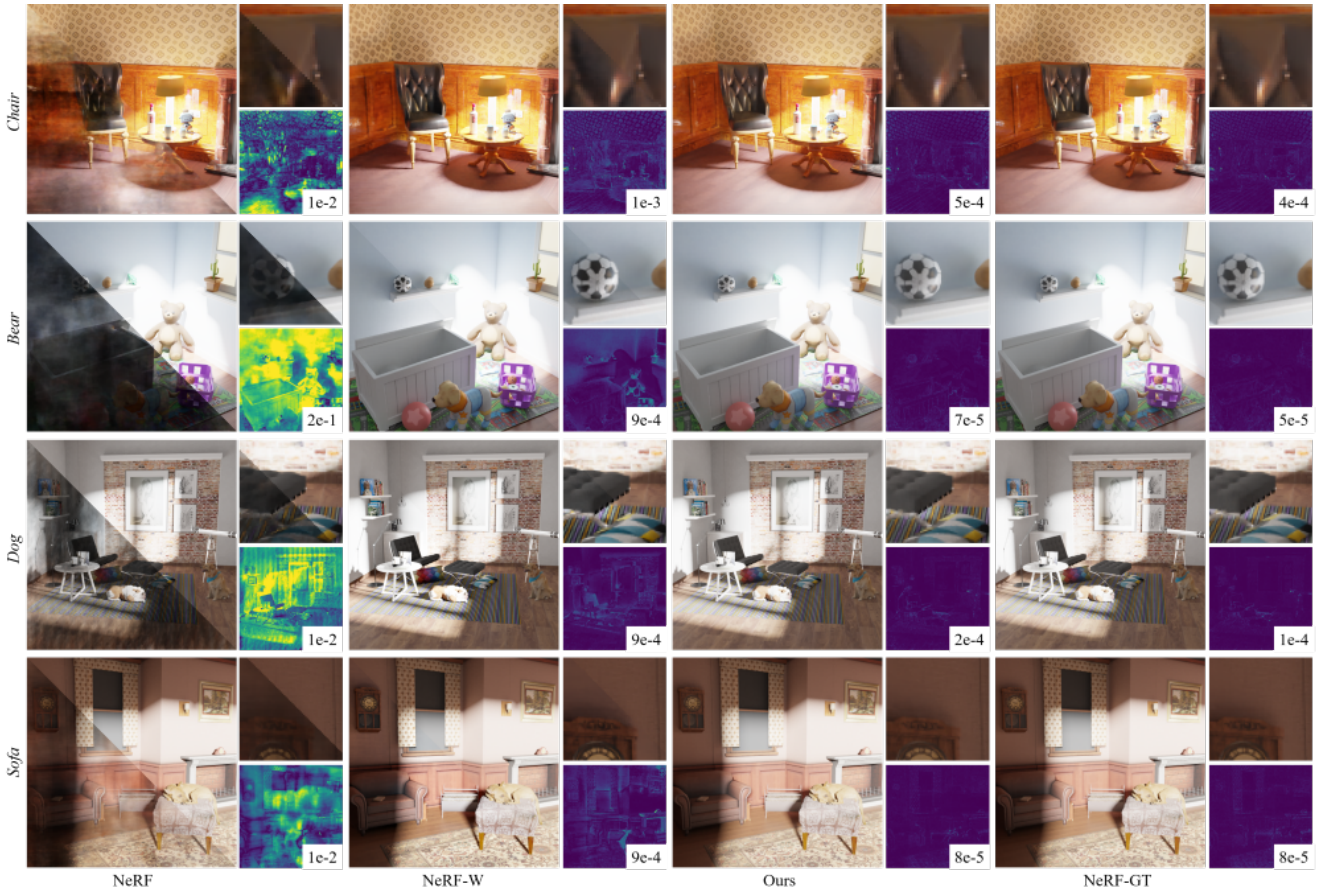


Figure 3. Qualitative comparison of rendered novel LDR view with a novel exposure. The upper triangular images are the ground truth and the lower triangular images are the rendered views. Zoom-in insets and error maps are given on the right. MSE values are on the bottom right of error maps.

datasets. Note that only our method can output both LDR and HDR views. Compared with NeRF-GT, our method achieves similar performance for rendering LDR views on the synthetic dataset, while our LDR views have a lower

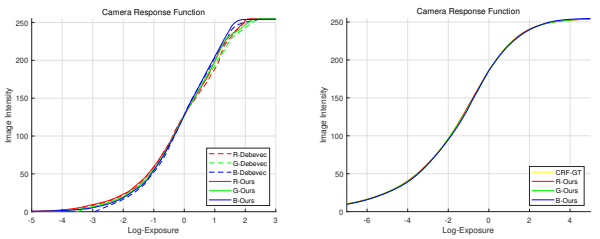
(a) (b) (c) (d) (e) (f)

Figure 4. Qualitative results of our novel LDR views and HDR views on synthetic scenes. (a–c) Our LDR views under different exposures. (d) Our tone-mapped HDR views and (e) ground truth tone-mapped HDR views. (f) Histograms of our novel HDR view (the upper one) and ground truth (the lower one). **Better viewed on screen with zoom in.**

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

Figure 5. Qualitative results of our novel HDR views on real scenes. Compared with the ground truth LDR views (the first row), our tone-mapped HDR views (the second row) reveal the details of over-exposure and under-exposure areas.

PSNR on real scenes. We notice that our estimated CRF of the blue channel has a bias due to the noise of training views, as seen in Fig. 6a, which results in the lower PSNR. As for rendering HDR views, our method is even comparable to NeRF-GT, and we find that directly training the NeRF model from HDR views is hard to produce the expected results, especially on the scene with a larger dynamic range. In addition, we qualitatively compare our method with baselines on rendering novel LDR views with a novel exposure in Fig. 3. One can see that the LDR views rendered by our method and NeRF-GT are close to ground truth, but the results of NeRF show serious artifacts because of the varying exposures between input views. The novel views synthesized by NeRF-W appear to be acceptable, yet exhibit inconsistent color with ground truth, as shown in zoom-in insets of Fig. 3. Moreover, our novel LDR views with dif-



(a) (b)

Figure 6. Discrete CRFs estimated by our method on (a) real *flower* scene and (b) synthetic *chair* scene. On the real scene, we calibrate the CRF of digital camera using the method by Debevec and Malik [12].

Table 2. A comparison of our method with 2 exposures $\{t_1, t_5\}$, 3 exposures $\{t_1, t_3, t_5\}$, or 5 exposures: $\{t_1, t_2, t_3, t_4, t_5\}$. Metrics (PSNR/SSIM/LPIPS) are averaged over synthetic scenes.

	LDR-OE	LDR-NE	HDR
2	32.39/0.954/0.040	32.76/0.950/0.036	33.00/0.949/0.040
3	37.52/0.964/0.022	35.73/0.954/0.025	37.60/0.963/0.021
5	37.73/0.968/0.020	36.26/0.960/0.022	37.86/0.969/0.019

ferent exposures are shown in Fig. 4. It validates that our method can control the exposure of rendered views by giving a specified exposure time.

The novel HDR views are presented in Fig. 4 and Fig. 5. It can be seen that the HDR results by our approach (Fig. 4d) are reasonably close to ground truth HDR images (Fig. 4e). Furthermore, compared with LDR views, our tone-mapped HDR views reveal the details of over-exposure and under-exposure areas. We also present the histograms of our and ground truth HDR views in Fig. 4f. The distributions of our histograms are similar to those of ground truth. Besides, discrete CRFs estimated by our method are shown in Fig. 6, which validates that our *tone mapper* can accurately model the response functions of cameras.

Ablation Studies. 1) Theoretically, recovering a camera response curve requires a minimum of two exposures [12]. We investigate the influence of the number of exposures in Tab. 2, where the number is set to $\{2, 3, 5\}$ respectively. We can see that the performance of the proposed method improves with the number of exposures. The results are close when the number is set to 3 or 5, and both significantly outperform the results of 2 exposures. Thereby, using 3 exposures is a reasonable choice. 2) The ablation study of unit exposure loss $\mathcal{L}_u$ is presented in Tab. 3 and Fig. 7. Table 3 shows that our method produces better quantitative results with the unit exposure loss, especially on rendering HDR views. The HDR images rendered by the approach without unit exposure loss suffer from severe chromatic aberration (Fig. 7b) due to the different shifts of three estimated CRF curves (Fig. 7d). 3) Since the RGB channels have the same

Table 3. Quantitative results with/without unit exposure loss $\mathcal{L}_u$. Metrics are averaged over synthetic scenes.

	with $\mathcal{L}_u$			w/o $\mathcal{L}_u$		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
LDR-OE	37.52	0.964	0.022	36.48	0.957	0.030
LDR-NE	35.73	0.954	0.025	34.77	0.947	0.035
HDR	37.60	0.963	0.021	13.35	0.765	0.163

(a) with $\mathcal{L}_u$ (b) w/o $\mathcal{L}_u$ (c) GT (d) CRFs

Figure 7. Qualitative results with/without unit exposure loss $\mathcal{L}_u$. (a–c) The tone-mapped HDR views. (d) The estimated CRFs. **Better viewed on screen with zoom in.**

CRF in synthetic scenes, we evaluate the efficiency of modeling the CRF with three MLPs on real scenes. As shown in Tab. 1, when three channels are processed independently, our method achieves superior results.

Limitations. Recovering an HDR radiance field from a series of LDR images with different exposures is challenging. Similar to the classic HDR radiance map recovering method [12], our recovered HDR radiance field is relative. There are three unknown scaling factors (for RGB channels) that relate the recovered radiance to absolute radiance. Consequently, different choices of these factors will recover HDR radiance fields with different white balances. Besides, our *tone mapper* models the camera coarsely without considering the effect of ISO gain and aperture for exposures.

6. Conclusion

We have proposed a novel method to recover the high dynamic range neural radiance field from a set of LDR views with different exposures. Our method not only renders novel HDR views without ground-truth HDR supervision, but also produces high-fidelity LDR views with specified exposures. The core of the method is modeling the process that captures scene radiance and maps them into pixel values. Compared with prior works, our method performs better in rendering LDR views. Importantly, to our knowledge our method is the first neural rendering method that synthesizes novel views with high dynamic range. Code and models will be made available to the research community to facilitate reproducible research.

Acknowledgements. The work was supported by NSFC under Grant 62031023. The authors thank Li Ma and Xiaoyu Li for their instructive and useful advice.

References

[1] Blender. https://www.blender.org/. 5

[2] Photomatix Pro 6. https://www.hdrsoft.com/. 5

[3] Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Int. Conf. Comput. Vis.*, pages 5855–5864, October 2021. 1, 3

[4] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. NeRD: Neural reflectance decomposition from image collections. In *Int. Conf. Comput. Vis.*, pages 12684–12694, 2021. 3

[5] Chris Buehler, Michael Bosse, Leonard McMillan, Steven Gortler, and Michael Cohen. Unstructured lumigraph rendering. In *SIGGRAPH*, pages 425–432, 2001. 2

[6] Rodrigo Ortiz Cayon, Abdelaziz Djelouah, and George Drettakis. A bayesian approach for selective image-based rendering using superpixels. In *Int. Conf. 3D Vis.*, pages 469–477, 2015. 2

[7] Gaurav Chaurasia, Sylvain Duchene, Olga Sorkine-Hornung, and George Drettakis. Depth synthesis and local warps for plausible image-based navigation. *ACM Trans. Graph.*, 32(3):1–12, 2013. 2

[8] Xingyu Chen, Qi Zhang, Xiaoyu Li, Yue Chen, Feng Ying, Xuan Wang, and Jue Wang. Hallucinated neural radiance fields in the wild. *arXiv preprint arXiv:2111.15246*, 2021. 3

[9] Zhang Chen, Anpei Chen, Guli Zhang, Chengyuan Wang, Yu Ji, Kiriakos N Kutulakos, and Jingyi Yu. A neural rendering framework for free-viewpoint relighting. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5599–5610, 2020. 2

[10] Inchang Choi, Orazio Gallo, Alejandro Troccoli, Min H Kim, and Jan Kautz. Extreme view synthesis. In *Int. Conf. Comput. Vis.*, pages 7781–7790, 2019. 2

[11] Abe Davis, Marc Levoy, and Fredo Durand. Unstructured light fields. In *Computer Graphics Forum*, volume 31, pages 305–314, 2012. 2

[12] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH*, page 369–378, 1997. 2, 3, 4, 8

[13] Paul E Debevec, Camillo J Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In *SIGGRAPH*, pages 11–20, 1996. 2

[14] Yilun Du, Yinan Zhang, Hong-Xing Yu, Joshua B Tenenbaum, and Jiajun Wu. Neural radiance flow for 4D view synthesis and video processing. In *Int. Conf. Comput. Vis.*, pages 14324–14334, 2021. 3

[15] Frédéric Dufaux, Patrick Le Callet, Rafal Mantiuk, and Marta Mrak. *High dynamic range video: from acquisition, to display and applications*. Academic Press, 2016. 4

[16] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K Mantiuk, and Jonas Unger. HDR image reconstruction from a single exposure using deep CNNs. *ACM Trans. Graph.*, 36(6):1–15, 2017. 3

[17] John Flynn, Michael Broxton, Paul Debevec, Matthew DuVall, Graham Fyffe, Ryan Overbeck, Noah Snavely, and Richard Tucker. Deepview: View synthesis with learned gradient descent. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2367–2376, 2019. 2

[18] Rohit Girdhar, David F Fouhey, Mikel Rodriguez, and Abhinav Gupta. Learning a predictable and generative vector representation for objects. In *Eur. Conf. Comput. Vis.*, pages 484–499. Springer, 2016. 3

[19] Steven J Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F Cohen. The lumigraph. In *SIGGRAPH*, pages 43–54, 1996. 2

[20] Thorsten Grosch et al. Fast and robust high dynamic range image generation with camera and object movement. *Vision, Modeling and Visualization, RWTH Aachen*, 277284, 2006. 3

[21] Michelle Guo, Alireza Fathi, Jiajun Wu, and Thomas Funkhouser. Object-centric neural scene rendering. *arXiv preprint arXiv:2012.08503*, 2020. 3

[22] Peter Hedman, Suhib Alsisan, Richard Szeliski, and Johannes Kopf. Casual 3D photography. *ACM Trans. Graph.*, 36(6):1–15, 2017. 2

[23] Peter Hedman and Johannes Kopf. Instant 3D photography. *ACM Trans. Graph.*, 37(4):1–12, 2018. 2

[24] Katrien Jacobs, Celine Loscos, and Greg Ward. Automatic high-dynamic range image generation for dynamic scenes. *IEEE Computer Graphics and Applications*, 28(2):84–93, 2008. 3

[25] James T Kajiya and Brian P Von Herzen. Ray tracing volume densities. *ACM SIGGRAPH computer graphics*, 18(3):165–174, 1984. 2, 4

[26] Nima Khademi Kalantari, Ravi Ramamoorthi, et al. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph.*, 36(4):144–1, 2017. 2, 3, 5

[27] Zeeshan Khan, Mukul Khanna, and Shanmuganathan Raman. FHDR: HDR image reconstruction from a single LDR image using feedback network. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1–5. IEEE, 2019. 3

[28] Junghee Kim, Siyeong Lee, and Suk-Ju Kang. End-to-end differentiable learning to HDR image synthesis for multi-exposure images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 1780–1788, 2021. 3

[29] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Int. Conf. Learn. Represent.*, 2015. 5

[30] Loubna Lechlek, Daniel Meneveaux, Mickael Ribardiere, Romuald Perrot, and Mohamed Chaouki Babahenini. Interactive hdr image-based rendering from unstructured ldr photographs. *Computers & Graphics*, 84:1–12, 2019. 2

[31] Marc Levoy and Pat Hanrahan. Light field rendering. In *SIGGRAPH*, pages 31–42, 1996. 2

[32] Tianye Li, Mira Slavcheva, Michael Zollhoefer, Simon Green, Christoph Lassner, Changil Kim, Tanner Schmidt, Steven Lovegrove, Michael Goesele, and Zhaoyang Lv. Neural 3D video synthesis. *arXiv preprint arXiv:2103.02597*, 2021. 3

[33] Zhengqi Li, Simon Niklaus, Noah Snavely, and Oliver Wang. Neural scene flow fields for space-time view synthesis of dy-

namic scenes. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 6498–6508, 2021. 3

[34] Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. Neural volumes: Learning dynamic renderable volumes from images. *ACM Trans. Graph.*, 38(4), 2019. 2

[35] Feng Lu, Xiangyang Ji, Qionghai Dai, and Guihua Er. Multi-view stereo reconstruction with high dynamic range texture. In *Asian Conference on Computer Vision*, pages 412–425. Springer, 2010. 2

[36] Li Ma, Xiaoyu Li, Jing Liao, Qi Zhang, Xuan Wang, Jue Wang, and Pedro V Sander. Deblur-nerf: Neural radiance fields from blurry images. *arXiv preprint arXiv:2111.14292*, 2021. 3

[37] 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. 1, 2, 3, 5, 6

[38] Leonard McMillan and Gary Bishop. Plenoptic modeling: An image-based rendering system. In *SIGGRAPH*, pages 39–46, 1995. 2

[39] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *15th Pacific Conference on Computer Graphics and Applications (PG'07)*, pages 382–390. IEEE, 2007. 2, 3

[40] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul Srinivasan, and Jonathan T Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. *arXiv preprint arXiv:2111.13679*, 2021. 2

[41] 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. 2, 3, 5

[42] 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. 1, 2, 3, 5, 6

[43] Ryan S Overbeck, Daniel Erickson, Daniel Evangelakos, Matt Pharr, and Paul Debevec. A system for acquiring, processing, and rendering panoramic light field stills for virtual reality. *ACM Trans. Graph.*, 37(6):1–15, 2018. 2

[44] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Int. Conf. Comput. Vis.*, pages 5865–5874, 2021. 3

[45] 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 *Eur. Conf. Comput. Vis.*, pages 497–513. Springer, 2020. 5

[46] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-NeRF: Neural radiance fields for dynamic scenes. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 10318–10327, 2021. 3

[47] Albert Pumarola, Stefan Popov, Francesc Moreno-Noguer, and Vittorio Ferrari. C-flow: Conditional generative flow models for images and 3d point clouds. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7949–7958, 2020. 3

[48] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010. 2

[49] Darius Rückert, Linus Franke, and Marc Stamminger. Adop: Approximate differentiable one-pixel point rendering. *arXiv preprint arXiv:2110.06635*, 2021. 2

[50] Jonathan Shade, Steven Gortler, Li-wei He, and Richard Szeliski. Layered depth images. In *SIGGRAPH*, pages 231–242, 1998. 2

[51] Mansi Sharma, Santanu Chaudhury, and Brejesh Lall. Parameterized variety for multi-view multi-exposure image synthesis and high dynamic range stereo reconstruction. In *2012 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, pages 1–4. IEEE, 2012. 2

[52] Harry Shum and Sing Bing Kang. Review of image-based rendering techniques. In *Visual Communications and Image Processing*, volume 4067, pages 2–13, 2000. 2

[53] Vincent Sitzmann, Justus Thies, Felix Heide, Matthias Nießner, Gordon Wetzstein, and Michael Zollhofer. Deepvoxels: Learning persistent 3D feature embeddings. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2437–2446, 2019. 2

[54] Pratul P Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T Barron. NeRV: Neural reflectance and visibility fields for relighting and view synthesis. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7495–7504, 2021. 3

[55] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010. 2, 4

[56] Okan Tarhan Tursun, Ahmet Oğuz Akyüz, Aykut Erdem, and Erkut Erdem. The state of the art in HDR deghosting: a survey and evaluation. In *Computer Graphics Forum*, volume 34, pages 683–707. Wiley Online Library, 2015. 3

[57] Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Chun-Yu Sun, and Xin Tong. O-CNN: Octree-based convolutional neural networks for 3D shape analysis. *ACM Trans. Graph.*, 36(4):1–11, 2017. 3

[58] Wenqi Xian, Jia-Bin Huang, Johannes Kopf, and Changil Kim. Space-time neural irradiance fields for free-viewpoint video. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 9421–9431, 2021. 3

[59] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. Attention-guided network for ghost-free high dynamic range imaging. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1751–1760, 2019. 2, 5

[60] Qingsen Yan, Yu Zhu, and Yanning Zhang. Robust artifact-free high dynamic range imaging of dynamic scenes. *Multimedia Tools and Applications*, 78(9):11487–11505, 2019. 3

[61] Bangbang Yang, Yinda Zhang, Yinghao Xu, Yijin Li, Han Zhou, Hujun Bao, Guofeng Zhang, and Zhaopeng Cui.

Learning object-compositional neural radiance field for editable scene rendering. In *Int. Conf. Comput. Vis.*, pages 13779–13788, October 2021. 3

[62] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelNeRF: Neural radiance fields from one or few images. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4578–4587, 2021. 1, 3

[63] Fisher Yu and David Gallup. 3D reconstruction from accidental motion. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3986–3993, 2014. 2

[64] Jiakai Zhang, Xinhang Liu, Xinyi Ye, Fuqiang Zhao, Yanshun Zhang, Minye Wu, Yingliang Zhang, Lan Xu, and Jingyi Yu. Editable free-viewpoint video using a layered neural representation. *ACM Trans. Graph.*, 40(4):1–18, 2021. 3

[65] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 586–595, 2018. 5

[66] Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, and Noah Snavely. Stereo magnification: learning view synthesis using multiplane images. *ACM Trans. Graph.*, 37(4):1–12, 2018. 2, 3

[67] Zihan Zhou, Hailin Jin, and Yi Ma. Plane-based content preserving warps for video stabilization. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2299–2306, 2013. 2

[68] C Lawrence Zitnick, Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High-quality video view interpolation using a layered representation. *ACM Trans. Graph.*, 23(3):600–608, 2004. 2