Physical-aware Neural Radiance Fields for Efficient Exposure Correction

Kai Xu, Mingwen Shao*, Yuanjian Qiao, Yan Wang

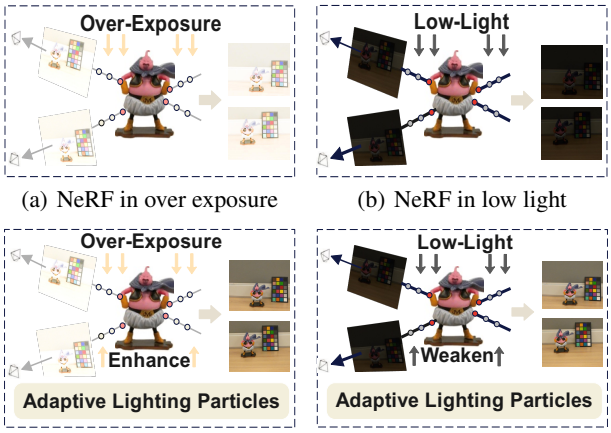
Qingdao Institute of Software, College of Computer Science and Technology, China University of Petroleum (East China)
kxu11489@gmail.com, smw278@126.com, yjqiao@s.upc.edu.cn, wangyanstu@163.com

Abstract

Neural Radiance Fields (NeRF) has achieved remarkable success in synthesizing impressive novel views. However, existing methods usually fail to handle scenes with adverse lighting conditions caused by external time variations and different camera settings, leading to poor visual quality. To address this challenge, we propose a physical-aware NeRF for efficient exposure correction, named PHY-NeRF. Specifically, we design Adaptive Lighting Particles inspired by the theory of light scattering and absorption, which can adjust the illumination intensity during volume rendering. Subsequently, we can handle scenes with different lighting conditions by jointly optimizing camera parameters and these lighting particles. Moreover, to promote natural brightness transitions, we devise a global illumination consistency module to control the lighting intensity across views at the feature level while completing more details. Benefiting from the above designs, our PHY-NeRF can tackle arbitrary low-light or overexposed scenes in an unsupervised manner. Extensive experiments show that our PHY-NeRF achieves state-of-the-art results in addressing adverse lighting problems while ensuring high rendering efficiency.

Introduction

3D scene reconstruction and rendering are long-standing problems in computer vision. Over the past few years, the emergence of Neural Radiance Fields (NeRF) (Mildenhall et al. 2020) has promoted the development of this field, which can synthesize high-quality novel views from only a set of 2D images. However, a notable shortcoming of NeRF is that the input images are taken under normal lighting conditions, i.e., there should not be regions that are excessively dark or bright. Unfortunately, variations in the capture time, e.g., high light in the morning and low light in the evening, as well as differences in camera exposure settings, e.g., longer exposures produce brighter images, result in the appearance of low light or overexposed images. These abnormal images obscure scene details and distort the original color due to lower or higher photons, making it impossible for NeRF to generate high-quality images (see Fig. 1 (a) (b)).



(a) NeRF in over exposure (b) NeRF in low light

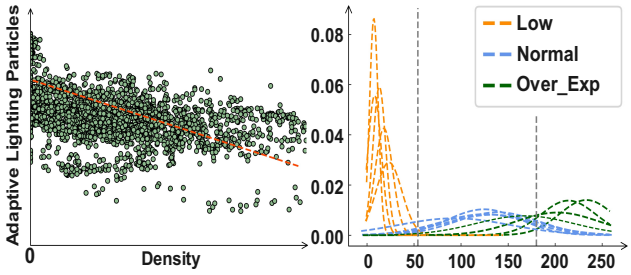
(c) PHY-NeRF in over exposure (d) PHY-NeRF in low light

Figure 1: Due to the defects of volume rendering, NeRF cannot handle scenes with abnormal lighting (a) (b). By adopting adaptive lighting particles, PHY-NeRF can take low-light and overexposed multi-view images as input and synthesize novel views under adverse light (c) (d).

Recently, several methods have been proposed to solve the problem of training NeRF in abnormal lighting conditions. The initial way is to directly preprocess the input image or post-process the NeRF-rendered image using exposure correction methods (Nsampi, Hu, and Wang 2021; Cui et al. 2022; Huang et al. 2023a). Although these methods can correct brightness, they tend to learn specific brightness mappings and cannot deal with multi-view inputs, which makes them incapable of being directly applied to real 3D scenes. As a potential solution, Aleth-NeRF (Cui et al. 2024) deals with challenging lighting scenes by proposing the assumption of concealing fields. It exploits local and global concealing fields to adapt dark scenes, achieving favorable results in challenging lighting scenes but has low overall rendering efficiency.

To address the above challenges, we analyze the physical properties of light transmission based on the theory of light scattering and absorption. As shown in Fig. 2 (a), the brightness or darkness of the scene is achieved through the interaction between light and objects, and the ability to adjust

*Corresponding author: Mingwen Shao



(a) Relationship (b) Pixel Distribution

Figure 2: Adaptive Lighting Particles have a negative correlation with density (a). Pixel distribution of low-light, normal, and overexposed images in five scenes (b).

light becomes weaker as the density of the object increases, e.g., light is reflected in low-density areas and converted into other forms of energy in high-density areas. We also visualized the pixel distribution of images with different illumination as shown in Fig. 2 (b). The pixel distribution of normal illumination is mostly based on the middle range, while the pixels of abnormal illumination tend to be either too low or too high. Therefore, we aim to exploit the physical properties to change the brightness level during NeRF training to adaptively deal with any lighting scenes.

In this paper, we propose the physical-aware NeRF (PHY-NeRF) to effectively handle low-light and overexposed scenes. Based on the above analysis, we design Adaptive Lighting Particles in the volume rendering process of NeRF, which can weaken the illumination intensity in low-light scenes and enhance it in overexposed scenes to address adverse lighting problems (see Fig. 1 (c) (d)). In addition, we devise a feature-based global consistency illumination module to promote natural lighting transitions. This module utilizes the feature-level illumination consistency loss to constrain cross-view brightness transfer while preventing learning its geometry into external illumination. Experimental results show that our PHY-NeRF can generate high-quality normal lighting scenes on challenging illumination datasets with high overall rendering efficiency. Our contributions can be summarized as follows:

- We propose a PHY-NeRF that can efficiently handle low-light and overexposed scenes while synthesizing novel views with normal lighting.
- We design Adaptive Lighting Particles in the volume rendering process to adjust the brightness of spatial points, enabling PHY-NeRF to handle scenes with different lighting conditions.
- We devise a feature-based global consistency illumination module to promote natural transitions of brightness while complementing local details.
- Extensive experiments have shown that our PHY-NeRF is superior to existing methods in terms of effectiveness and efficiency in dealing with abnormal lighting scenes.

Related Works

Novel View Synthesis with NeRF

With the development of neural representation (Tewari et al. 2020) and volume rendering (Lombardi et al. 2019), Neural Radiation Fields (NeRF) (Mildenhall et al. 2020) has been proposed to represent scenes by synthesizing new views based only on a set of 2D datasets and the corresponding camera parameters, without the need to model the explicit shape of each object. Due to its simplicity and effectiveness, many subsequent works have attempted to explore the architecture and applications of NeRF. Most of the former works address the shortcomings of NeRF itself, such as slow training speed (Fridovich-Keil et al. 2022; Müller et al. 2022) and poor generalization (Wang et al. 2021; Suhail et al. 2022). Some of the latter works consider how to apply NeRF to more challenging scenes, including blur (Wang et al. 2023b), occlusion (Zhu et al. 2023; Ren et al. 2024), reflection (Zeng et al. 2023), super-resolution (Huang et al. 2023b), etc.

There are also some works dealing with image illumination. Appearance editing (Martin-Brualla et al. 2021; Chen et al. 2022; Yang et al. 2023) achieves style transfer from different views by introducing appearance embedding vectors, but is usually unable to deal with adverse lighting scenes. Early approaches to deal with adverse lighting, such as RAW-NeRF (Mildenhall et al. 2022) and HDR-NeRF (Huang et al. 2022), can handle low-light, noisy camera raw images, but have difficulty processing sRGB images of real scenes. Subsequently, LLNeRF (Wang et al. 2023a) utilizes Retinex theory to decompose the radiation field for sRGB low-light image enhancement, but it cannot be applied to overexposed images. Furthermore, Aleth-NeRF (Cui et al. 2024) can handle abnormal lighting scenes by constructing concealing fields to change the transmission function, but it performs poorly in overexposed scenes due to problems with concealing field modeling. In contrast, our PHY-NeRF can handle both underexposed and overexposed scenes with remarkable results by adopting adaptive lighting particles and a feature-based global light consistency module.

Image Enhancement and Exposure Correction

Low light and overexposed images make processing image work extremely difficult due to color distortion. Several low-light enhancement and exposure correction methods have been proposed to cope with these adverse lighting conditions. Specifically, low-light enhancement aims to improve the visibility of captured images in low-light conditions. Early methods are based on Retinex theory (Land 1986) or Histogram Equalization (Tomasi and Manduchi 1998), but with the rapid development of deep learning, many methods based on CNN (Guo et al. 2020; Afifi et al. 2021), Transformer (Cai et al. 2023) and even generative models (Zhou, Yang, and Yang 2023; Shang et al. 2024). These methods have now become mainstream. Similar to low-light enhancement, exposure correction aims to simultaneously deal with bright and faded areas caused by overexposure, or dark areas resulted from underexposure. Early methods are also based on Histograms (Nayar and Branzoi 2003) to adjust the brightness of images. With the addition of training models,

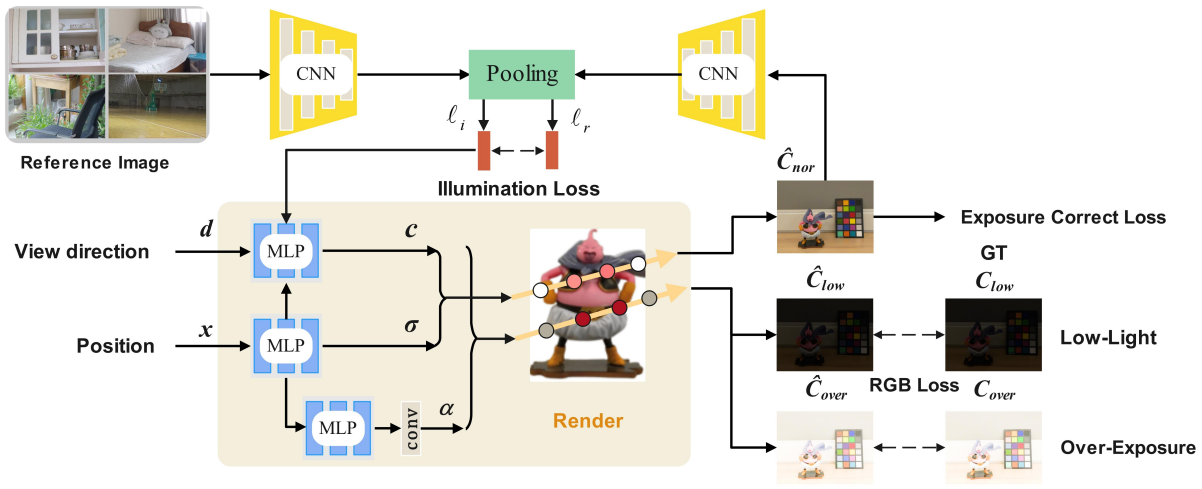


Figure 3: Pipeline of our PHY-NeRF. We handle different lighting scenes by proposing Adaptive Lighting Particles α. In addition, to promote a more natural transition of lighting, we adopt the illumination vector ℓ_i generated by the feature-level lighting consistency module. Benefiting from the above designs, we achieve the synthesis of normal illumination views under adverse lighting conditions.

many subsequent works (Nsampi, Hu, and Wang 2021; Cui et al. 2022; Huang et al. 2023a) have also been proposed to remove exposure errors.

Most low-light enhancement and exposure correction methods are for 2D single images and are usually not directly transferable to 3D scenes. The main reason is that these methods usually fail to utilize the geometry of scenes, which destroys 3D consistency and leads to blurring and artifacts in the rendered views. Compared to these methods, our method synthesizes high-quality views while maintaining multi-view consistency.

Methods

Given a collection of photos with low light and overexposure, we aim to reconstruct normal-lighting scenes and synthesize corresponding novel views. We first briefly review the vanilla-NeRF and then describe how the adaptive lighting particles adjust the light intensity to deal with different lighting scenes. Subsequently, we explain how to achieve reasonable transitions in brightness by designing a feature-based global illumination consistency module. Fig. 3 illustrates the proposed PHY-NeRF pipeline. Finally, we show how to train our model utilizing an effective unsupervised approach.

Preliminary

Neural Radiance Fields implicitly represents the scene given a set of images $\{I_i\}_{i=1}^{N}$ and camera parameters using a continuous volume function F_θ. NeRF learns to render the color of each pixel $\mathbf{C}(\mathbf{r})$ for a ray $\mathbf{r}$. It maps 3D points $\mathbf{x} = (x, y, z)$ and viewing directions $\mathbf{d} = (d_x, d_y, d_z)$ to volume densities σ and colors $\mathbf{c} = (r, g, b)$, modeled as two multilayer perceptrons (MLP). Over the entire process, this is given by

$$(\sigma, \mathbf{z}(t)) = F_{\theta_1}(\gamma_{\mathbf{x}}(\mathbf{x})), \tag{1}$$

$$\mathbf{c} = F_{\theta_2}(\gamma_{\mathbf{d}}(\mathbf{d}), \mathbf{z}(t)), \tag{2}$$

where θ_1, θ_2 are the MLP parameters, $\gamma_{\mathbf{x}}(\cdot), \gamma_{\mathbf{d}}(\cdot)$ are the fixed encoding functions for $\mathbf{x}$ and $\mathbf{d}$ respectively. To render the color of each pixel in the image, NeRF uses discrete quadrature to approximate the volume rendering integral, letting the camera ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ accumulate all the radiance along the ray from the given camera origin $\mathbf{o}$ and the direction $\mathbf{d}$ to render its corresponding pixel value $\hat{\mathbf{C}}(\mathbf{r})$ as:

$$\hat{\mathbf{C}}(\mathbf{r}) = \sum_{k=1}^{K} T_k(1 - \exp(-\sigma_k \delta_k))\mathbf{c}_k, \tag{3}$$

$$T_k = \exp(-\sum_{l=1}^{k-1} \sigma_l \delta_l), \tag{4}$$

where $\mathbf{c}_k$ and σ_k are the color and density at the point t_k of the ray, $\delta_k = t_{k+1} - t_k$ is the distance of adjacent points between the near and far planes of the ray sampling. Function T_k corresponds to the accumulative transmittance of the ray from the near plane to point t_k.

NeRF employs coarse and fine models for hierarchical sampling to improve the sampling efficiency. All parameters are optimized by minimizing the photometric loss given the true color $\mathbf{C}(\mathbf{r})$ as follows:

$$\mathcal{L}_{rgb} = \sum \left\| \hat{\mathbf{C}}^c(\mathbf{r}) - \mathbf{C}(\mathbf{r}) \right\|_2^2 + \left\| \hat{\mathbf{C}}^f(\mathbf{r}) - \mathbf{C}(\mathbf{r}) \right\|_2^2, \tag{5}$$

where $\hat{\mathbf{C}}^c(\mathbf{r})$ and $\hat{\mathbf{C}}^f(\mathbf{r})$ are the coarse and fine models, respectively.

Adaptive Exposure Control

The rendering performance of NeRF is degraded due to its problem in modeling adverse brightness scenes. To address

this issue, we consider modifying the vanilla-NeRF volume rendering way to handle different brightness scenes.

We analyze in physical properties that space objects adjust their external brightness by interacting with light. Based on this analysis, we propose the Adaptive Lighting Particles to solve the problem of adverse brightness. These particles define the ability of the 3D spatial point at $\mathbf{r}(t)$ to adjust the external lighting intensity, denoted as $\alpha(\mathbf{r}(t))$. As shown in Fig. 2 (a), the particles at each spatial point are negatively correlated with their density (the red dashed line). This verifies that the ability of an object to adjust light is independent of its density, so these particles rarely participate in scene reconstruction and only adjust the intensity of external lighting. Spatial points with high adjustment ability are usually located in areas with lower density, i.e., in the air outside the object.

Fig. 3 shows our training strategy for designing adaptive lighting particles. $\alpha(\mathbf{r}(t))$ is fed into a separate MLP conditioned on the intermediate representation $\mathbf{z}$ by the density MLP and learned separately for each 3D position. In the NeRF rendering process, we consider external lighting to be the same for every point. However, the ability of each point to adjust the light is diminished because the spatial points in front of each point occlude the points behind it. Therefore, we devise a convolutional layer to establish spatial information between different 3D points and enrich the lighting information, which can also effectively suppress noise and help achieve more reasonable rendering results. The formula is as follows:

$$\alpha(\mathbf{r}(t)) = conv\left(F_{\theta_3}(\mathbf{z}(t))\right). \tag{6}$$

We similarly discover that the pixels of abnormal exposure images are generally concentrated on the sides compared to normal levels (see Fig. 2 (b)). Following this discovery, our PHY-NeRF applies adaptive lighting particles to simulate light suppression or enhancement process, modeling these abnormal lighting scenes. The lighting particle α for each 3D point lies in the interval [0,1], multiplying it corresponds to weakening the brightness while dividing increases brightness. Applying it to Eq. (3), we get the following equations:

$$\hat{\mathbf{C}}_{low}(\mathbf{r}) = \sum_{k=1}^{K} \alpha_k T_k (1 - \exp(-\sigma_k \delta_k))\mathbf{c}, \tag{7}$$

$$\hat{\mathbf{C}}_{over}(\mathbf{r}) = \sum_{k=1}^{K} \frac{T_k}{\alpha_k} (1 - \exp(-\sigma_k \delta_k))\mathbf{c}, \tag{8}$$

where $\hat{\mathbf{C}}_{low}(\cdot)$, $\hat{\mathbf{C}}_{over}(\cdot)$ correspond to rendered views of underexposed and overexposed scenes. To prevent the pixels from being out of range when applying these operations, we introduce the tanh function for normalization. Later experiments justify our design to handle scenes with different brightness.

Global Illumination Consistency

We can deal with scenes of different brightness by jointly optimizing NeRF and adaptive lighting particles. However, directly applying exposure correction on images may make lighting transitions unnatural and lose local details. To address these issues, we design a global consistency illumination module by utilizing a latent encoding method at the feature level to learn external illumination. The reference image I_R passes through a CNN-based encoder E_θ to obtain a feature map, which is guided to an average pooling layer to generate the corresponding illumination embedding vectors ℓ_i for each image:

$$\ell_i = AvgPooling(E_\theta(I_R)), \tag{9}$$

we replace the color $\mathbf{c}$ in Eq. (2) with the color $\mathbf{c}^{\ell_i}$ related to the external illumination:

$$\mathbf{c}^{\ell_i} = F_{\theta_2}\left(\gamma_{\mathbf{d}}(\mathbf{d}), \mathbf{z}(t), \ell_i\right). \tag{10}$$

However, implementing illumination migration on mismatched image pairs is inherently inappropriate and we must add additional constraints. Inspired by generating potential optimization methods (Bojanowski et al. 2017), we attempt to map images to potential spaces and propose a feature-based loss of global illumination consistency:

$$\mathcal{L}_{\mathrm{f}} = \left\| AugPooling(E_\theta\left(\hat{\mathcal{I}}_{nor}\right)) - \ell_i \right\|_1, \tag{11}$$

where $\hat{\mathcal{I}}_{nor}$ is a normal lighting rendered image. We embed the illumination vector ℓ_i extracted by the illumination encoder E_θ and try to adopt it in different views. Since the illumination information is a global representation across different reference images, we can compensate for the missing details due to illumination changes. In addition, to prevent encoding image geometry into the lighting vector, we devise a regularization constraint inspired by the work of MSGAN (Mao et al. 2019):

$$\mathcal{L}_{\mathrm{ms}} = \max\left(\frac{\hat{\mathcal{I}}_{nor} - \hat{\mathcal{I}}_{deg}}{\hat{\ell}_{nor} - \hat{\ell}_{deg}}\right), \tag{12}$$

where $(\cdot)_{nor}, (\cdot)_{deg}$ represent generation under normal and adverse lighting conditions, respectively. The total loss is as follows:

$$\mathcal{L}_{\mathrm{g}} = \mathcal{L}_{\mathrm{f}} + \lambda_{ms}\mathcal{L}_{\mathrm{ms}}, \tag{13}$$

where the first term is balanced by the second term, and λ_{ms} is a hyperparameter set to $1e^{-3}$.

Optimization

To implement PHY-NeRF, we propose an unsupervised loss. We slightly modified the photometric consistency loss of the NeRF as follows:

$$\mathcal{L}_{rgb} = \sum \left\| \hat{\mathbf{C}}_d^c(\mathbf{r}) - \mathbf{C}_d(\mathbf{r}) \right\|_2^2 + \left\| \hat{\mathbf{C}}_d^f(\mathbf{r}) - \mathbf{C}_d(\mathbf{r}) \right\|_2^2, \tag{14}$$

where $\hat{\mathbf{C}}_d(\cdot)$ represents rendered low-light or overexposed pixel and $\mathbf{C}_d(\cdot)$ represents GT adverse pixel.

Inspired by Zero-DCE (Guo et al. 2020), we utilize image-level brightness consistency loss, spatial structure loss, and color constancy loss. These three losses can be expressed as:

$$\mathcal{L}_{\mathrm{bc}} = \left\| Mean(\hat{\mathbf{C}}_n(\mathbf{r})) - \mathbf{a} \right\|_2^2, \tag{15}$$

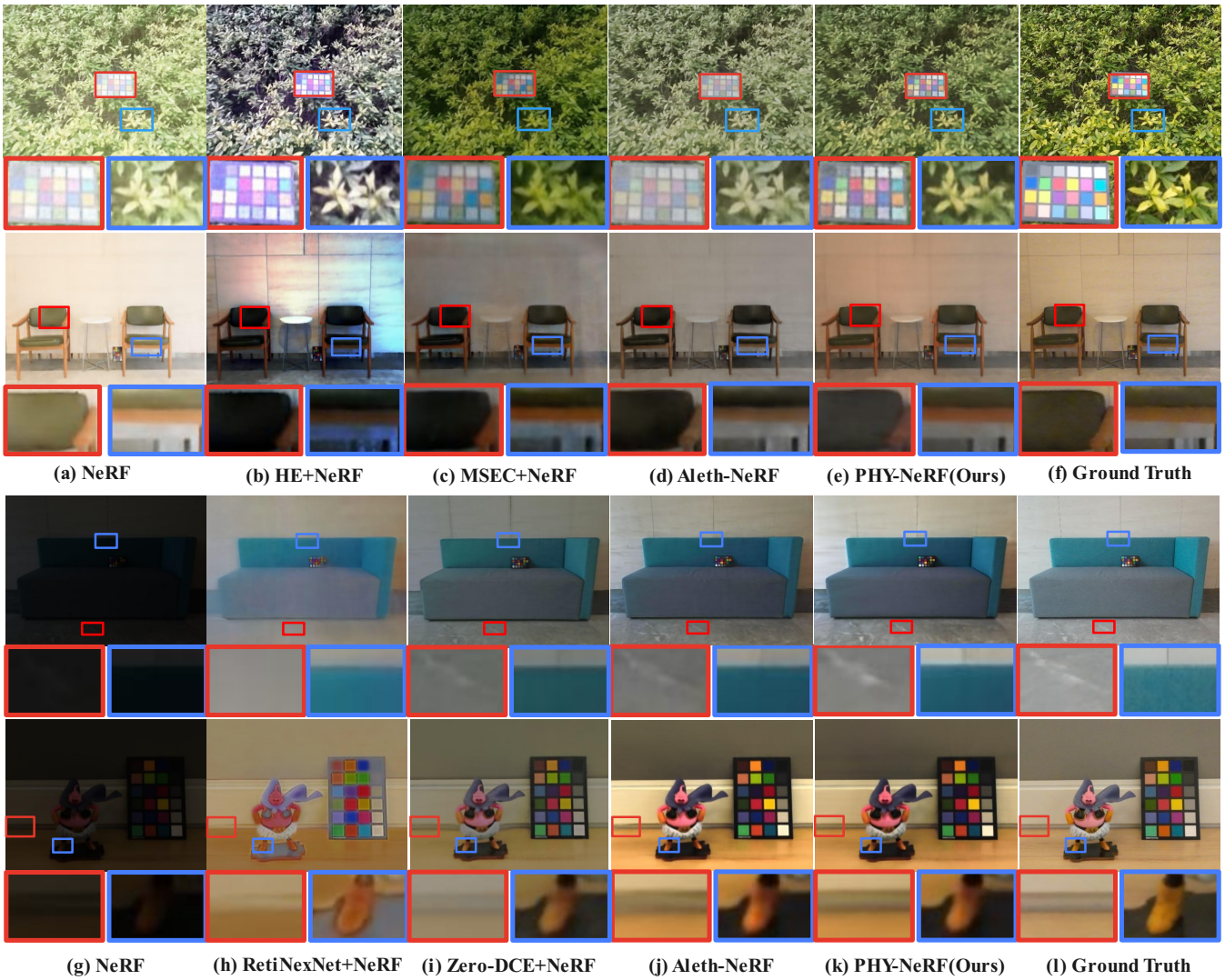


Figure 4: Qualitative results of different methods on the overexposed scene dataset.

and

$$\mathcal{L}_{ss} = \sum_{\mathbf{r}' \in \Omega(\mathbf{r})} \left(\hat{\mathbf{C}}_n(\mathbf{r}) - \hat{\mathbf{C}}_n(\mathbf{r}') \right) - \mathbf{c} \left(\mathbf{C}_d(\mathbf{r}) - \mathbf{C}_d(\mathbf{r}') \right), \tag{16}$$

and

$$\mathcal{L}_{cc} = \sum_{(p,q) \in \varepsilon} \left(\hat{\mathbf{C}}_n(\mathbf{r})^p - \hat{\mathbf{C}}_n(\mathbf{r})^q \right)^2, \tag{17}$$

where $\hat{\mathbf{C}}_n(\cdot)$ represents rendered normal lighting pixel. $\mathbf{a}$ and $\mathbf{c}$ represent the exposure intensity and contrast, respectively. $\Omega(\mathbf{r})$ are the four adjacent pixels (above, below, left, and right) centered at pixel $\mathbf{r}$, and $\varepsilon = \{(R,G),(R,B),(G,B)\}$ denotes the color channel pair.

To summarize, the total unsupervised loss of our training process is expressed as:

$$\mathcal{L}_{total} = \mathcal{L}_{rgb} + \lambda_g \mathcal{L}_g + \lambda_{bc} \mathcal{L}_{bc} + \lambda_{ss} \mathcal{L}_{ss} + \lambda_{cc} \mathcal{L}_{cc}, \tag{18}$$

where $\lambda_g, \lambda_{bc}, \lambda_{ss}, \lambda_{cc}$ are the hyperparameters for balancing the total loss weights, which are set to $1e^{-3}$, $1e^{-3}$, $1e^{-3}$, $1e^{-8}$, respectively.

Experiments

Experimental Settings

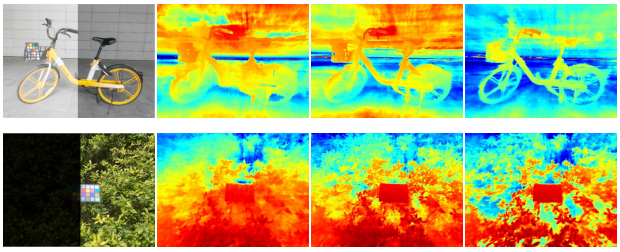
Datasets. We use the previously collected dataset of Aleth-NeRF (Cui et al. 2024), including low-light, normal-light, and over-exposure multi-view images. In this dataset, a total of 5 scenes in the real world ("***buu***", "***chair***", "***sofa***", "***bike***", "***shrub***") are collected. Each scene ranges from 25 to 48 images, and the resolution of each view image is 375 $\times$ 500. In each scene, we select 3 $\sim$ 5 images as the test set, 1 image as the validation set, and the other images as the training set. We also use the normal light images from the LOw Light paired dataset (LOL) (Wei et al. 2018) as the reference images, containing 500 low light and normal light image pairs, each with a 400 $\times$ 600 resolution.

Method	*buu*	*chair*	*sofa*	*bike*	*shrub*	*mean*
	PSNR/ SSIM/ LPIPS	PSNR/ SSIM/ LPIPS	PSNR/ SSIM/ LPIPS	PSNR/ SSIM/ LPIPS	PSNR/ SSIM/ LPIPS	PSNR/ SSIM/ LPIPS
① Over-Exposure Scene						
NeRF	7.12/ 0.674/ 0.499	11.05/ 0.741/ 0.418	10.22/ 0.783/ 0.475	9.65/ 0.698/ 0.416	9.96/ 0.405/ 0.480	9.60/ 0.660/ 0.457
NeRF + HE	14.34/ 0.613/ 0.673	15.37/ 0.661/ 0.590	16.69/ 0.733/ 0.558	15.32/ 0.627/ 0.458	11.97/ 0.468/ 0.556	14.74/ 0.620/ 0.567
NeRF + IAT	14.11/ 0.780/ 0.433	19.24/ 0.810/ 0.491	16.60/ 0.837/ 0.459	17.73/ 0.760/ 0.394	14.05/ 0.381/ 0.499	16.35/ 0.714/ 0.455
NeRF + MSEC	16.13/ 0.800/ 0.427	15.60/ 0.786/ 0.472	16.56/ 0.807/ 0.495	12.60/ 0.716/ 0.465	13.66/ 0.332/ 0.509	14.91/ 0.688/ 0.473
HE + NeRF	14.65/ 0.743/ 0.519	15.55/ 0.736/ 0.497	17.11/ 0.781/ 0.477	15.77/ 0.692/ **0.367**	12.61/ **0.506**/ 0.622	15.13/ 0.692/ 0.496
IAT + NeRF	16.22/ 0.815/ 0.486	18.98/ 0.799/ 0.503	18.45/ 0.849/ 0.478	19.63/ 0.776/ 0.408	15.63/ 0.434/ **0.477**	17.78/ 0.734/ 0.470
MSEC + NeRF	15.53/ 0.817/ 0.499	16.95/ 0.758/ 0.580	**19.60**/ 0.817/ 0.498	18.90/ 0.725/ 0.483	15.48/ 0.400/ 0.499	17.29/ 0.703/ 0.512
Aleth-NeRF	16.78/ 0.805/ 0.611	20.08/ 0.820/ 0.499	17.85/ 0.852/ 0.458	19.85/ 0.773/ 0.392	15.91/ 0.477/ 0.483	18.09/ 0.745/ 0.488
PHY-NeRF (Ours)	**22.51/ 0.858/ 0.407**	**25.48/ 0.839/ 0.413**	19.32/ **0.876/ 0.434**	**21.64/ 0.799**/ 0.393	**17.97**/ 0.447/ 0.527	**21.38/ 0.764/ 0.435**
② Under-Exposure Scene						
NeRF	7.51/ 0.291/ 0.448	6.04/ 0.147/ 0.594	6.28/ 0.210/ 0.568	6.35/ 0.072/ 0.623	8.03/ 0.031/ 0.680	6.84/ 0.150/ 0.582
NeRF + RetiNexNet	11.64/ 0.646/ 0.395	11.04/ 0.631/ 0.561	12.85/ 0.738/ 0.527	17.79/ 0.653/ 0.550	11.85/ 0.211/ 0.583	13.03/ 0.576/ 0.523
NeRF + Zero-DCE	17.81/ 0.833/ 0.357	12.44/ 0.684/ 0.547	14.43/ 0.787/ 0.539	10.16/ 0.468/ 0.557	12.58/ 0.282/ 0.540	13.48/ 0.610/ 0.488
NeRF + SCI	7.84/ 0.660/ 0.562	12.07/ 0.699/ 0.584	10.25/ 0.737/ 0.626	18.84/ 0.637/ 0.565	12.38/ 0.358/ 0.587	12.27/ 0.618/ 0.585
NeRF + IAT	14.03/ 0.656/ 0.453	19.08/ 0.800/ 0.565	10.49/ 0.528/ 0.678	17.16/ 0.657/ 0.534	16.15/ 0.344/ 0.573	15.38/ 0.597/ 0.561
RetiNexNet + NeRF	16.19/ 0.780/ 0.396	16.89/ 0.756/ 0.543	16.98/ 0.807/ 0.577	18.00/ 0.707/ 0.482	14.86/ 0.284/ 0.518	16.58/ 0.667/ 0.503
Zero-DCE + NeRF	17.90/ 0.858/ 0.376	12.58/ 0.721/ **0.460**	14.45/ 0.831/ 0.419	10.39/ 0.518/ 0.464	12.32/ 0.308/ 0.481	13.53/ 0.649/ 0.432
SCI + NeRF	7.76/ 0.692/ 0.525	19.77/ 0.802/ 0.674	10.08/ 0.772/ 0.520	13.44/ 0.658/ **0.435**	18.16/ 0.503/ 0.475	13.84/ 0.689/ 0.510
IAT + NeRF	14.46/ 0.705/ 0.386	18.70/ 0.780/ 0.665	17.88/ 0.829/ 0.547	13.65/ 0.616/ 0.528	13.87/ 0.317/ 0.536	15.71/ 0.649/ 0.532
Aleth-NeRF	20.22/ **0.859/ 0.315**	20.93/ 0.818/ 0.468	19.52/ 0.857/ **0.354**	20.46/ 0.727/ 0.499	**18.24**/ 0.511/ 0.448	19.87/ 0.754/ **0.417**
PHY-NeRF (Ours)	**20.29**/ 0.855/ 0.325	**21.19/ 0.822**/ 0.479	**20.45/ 0.865**/ 0.408	**22.03/ 0.729**/ 0.505	18.15/ **0.524/ 0.443**	**20.52/ 0.759**/ 0.432

Table 1: Quantitative results for five overexposed and underexposed scenes, where the bold and the underline indicate the best and second-best results, respectively. We evaluate PSNR↑, SSIM↑ and LPIPS↓.

Implementation Details. We implement PHY-NeRF using Pytorch and train our network using the Adam optimizer. The training batch size is 1024, with 12,500 iterations per epoch. We train different scenes for 4 $\sim$ 6 epochs on a single RTX3090 GPU, taking about 2 hours per scene. The illumination encoder E_θ is implemented by 4 downsampled convolutional layers. Adaptive Lighting Particles are implemented by 3 fully-connected layers with 128 channels and 1 convolutional layer. Other hyperparameter settings are detailed in the supplementary material.

Metrics. We use three metrics for evaluation: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS).



(a) Views (b) NeRF (c) Aleth-NeRF (d) PHY-NeRF

Figure 5: Depth results of our PHY-NeRF (d) compared with NeRF (b) and Aleth-NeRF (c) in partially over-exposure and under-exposure scenes.

Evaluation

Baselines. To evaluate the superiority of our PHY-NeRF, we compare with existing methods on the above challenging illumination dataset: NeRF (Mildenhall et al. 2020) as the baseline, enhancement or correction methods post-process NeRF (denoted as "NeRF + *"), enhancement or correction methods pre-process NeRF (denoted as "* + NeRF"), and Aleth-NeRF (Cui et al. 2024).

Comparisons of Overexposed Scenes. The qualitative results are shown in Fig. 4. Due to defects in the volume rendering process, vanilla-NeRF generates overbright images, which makes post-processing NeRF renderings using exposure correction methods ineffective. We then preprocess the overexposed images to utilize NeRF training on these corrected images, breaking multi-view consistency and causing artifacts in the rendered views. Table. 1 demonstrates the quantitative results. The metrics of these methods are generally low, but the pre-processed results are better than the post-processed ones.

Aleth-NeRF improves performance by utilizing concealing fields to suppress overbright scenes (see Table. 1). How-

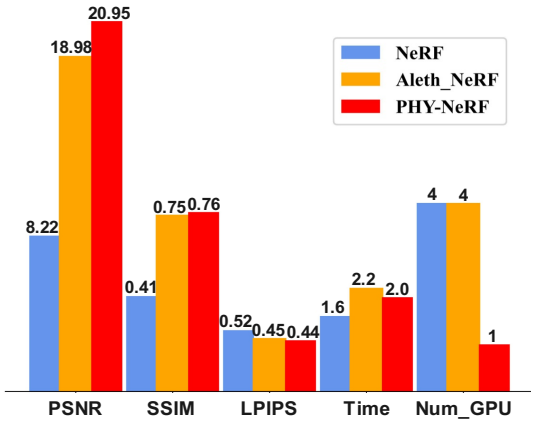


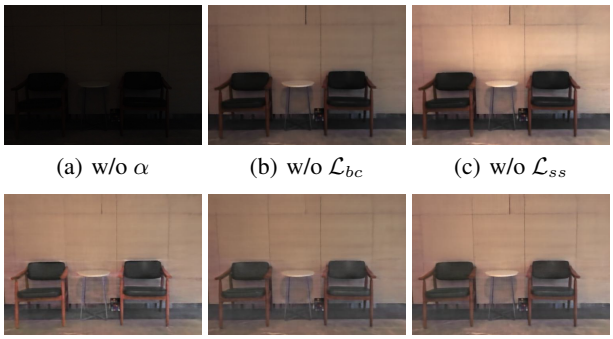
Figure 6: Comparison of PHY-NeRF with other end-to-end framework NeRF and Aleth-NeRF in terms of performance and efficiency.

ever, since the concealing fields cannot model the overbright scenes, training Aleth-NeRF directly on overexposed images results in brightness bias and color loss (see Fig. 4 (d)). In contrast, our PHY-NeRF achieves significant results in metrics by designing adaptive lighting particles and the global illumination consistency module as shown in Table. 1. Meanwhile, the qualitative results are shown in Fig. 4 (e), PHY-NeRF adjusts image brightness closer to normal while making color richer. In addition, it also captures more detailed information (see Fig. 5 (d)), saving a lot of data occupation and training time (see Fig. 6). More visualization results are detailed in the supplementary material.

Comparisons of Underexposed Scenes. Similar to the overexposed scenes, as shown in Fig. 4, the defects of NeRF make its rendered images in dark scenes with poor quality, leading to the performance degradation of these enhancement methods when processing NeRF rendered results. Moreover, applying these methods to preprocess NeRF destroys the consistency of multiple views, leading to artifacts. In addition, Aleth-NeRF adapts dark scenes by modeling the concealing fields, which achieves favorable results but loses some details. Compared to previous methods, our PHY-NeRF performs best in most scenes and average metrics (see Table .1), reconstructing normal lighting and color information. Meanwhile, it maintains multi-view consistency and highlights local details shown in Fig. 5. We utilize the same number of samples as in the overexposed scene, improving the rendering efficiency.

Ablation Study

As shown in Fig. 7, we conduct multiple ablation studies to demonstrate the effectiveness of each component. Adaptive Lighting Particles are a core component of our method, and exposure correction is not possible without these particles. In addition, the lack of a feature-based global illumination consistency module results in some dark shadows, unnatural brightness transitions, and loss of local details. When the image-level brightness consistency loss is discarded, the



(a) w/o α (b) w/o $\mathcal{L}_{bc}$ (c) w/o $\mathcal{L}_{ss}$

(d) w/o $\mathcal{L}_{cc}$ (e) w/o E_θ (f) full (Ours)

Figure 7: Visualization of ablation experiments to verify the effectiveness of the components, taking the "chair" scene as an example.

	chair	***sofa***
	PSNR/ SSIM/ LPIPS	PSNR/ SSIM/ LPIPS
w/o α	6.05/ 0.167/ 0.615	6.28/ 0.209/ 0.575
w/o $\mathcal{L}_{bc}$	11.97/ 0.653/ 0.512	15.46/ 0.801/ 0.408
w/o $\mathcal{L}_{ss}$	19.51/ 0.787/ <u>0.484</u>	18.62/ 0.845/ **0.406**
w/o $\mathcal{L}_{cc}$	<u>20.95</u>/ <u>0.817</u>/ 0.489	<u>19.49</u>/ <u>0.857</u>/ <u>0.407</u>
w/o E_θ	16.89/ 0.796/ 0.492	16.81/ 0.843/ 0.408
full (Ours)	**21.19/ 0.822/ 0.479**	**20.45/ 0.865**/ 0.408

Table 2: Ablation study on two low-light scenes, "***chair***" and "***sofa***", where the bold and the underline indicate the best and second-best results, respectively.

overall brightness becomes darker than the full result. Removing the spatial structure loss causes too much difference between neighboring pixels in the rendered image, leading to incorrect contrast (such as the wall and the chair). We remove the color constancy loss, which ignores the relationship between the three channels, causing severe color casts in some areas (such as the chair frame). As quantitatively shown in Table .2, our full components show the best results.

Conclusion

In this paper, we propose PHY-NeRF, a novel and efficient method to handle abnormal exposure scenes. Unlike existing methods based on Retinex theory or Concealing Fields assumption that perform poorly in overexposed scenes, we introduce physical knowledge to NeRF to cope with arbitrary lighting situations. Specifically, the proposed PHY-NeRF design Adaptive Lighting Particles that can simulate light suppression or enhancement process by adjusting the brightness of spatial points to model adverse lighting scenes and synthesize high-quality normal brightness views. Furthermore, the proposed PHY-NeRF also devises a global illumination consistency module to control the brightness intensity across views at the feature level, achieving natural lighting transitions. Experiments show that our PHY-NeRF achieves state-of-the-art performance on low-light and overexposed datasets while improving rendering efficiency.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (2021YFA1000102), National Natural Science Foundation of China (Nos. 62376285, 62272375, and 61673396), and Natural Science Foundation of Shandong Province, China (No. ZR2022MF260).

References

Afifi, M.; Derpanis, K. G.; Ommer, B.; and Brown, M. S. 2021. Learning multi-scale photo exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9157–9167.

Barron, J. T.; Mildenhall, B.; Tancik, M.; Hedman, P.; Martin-Brualla, R.; and Srinivasan, P. P. 2021. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5855–5864.

Bojanowski, P.; Joulin, A.; Lopez-Paz, D.; and Szlam, A. 2017. Optimizing the latent space of generative networks. *arXiv preprint arXiv:1707.05776*.

Buchsbaum, G. 1980. A spatial processor model for object colour perception. *Journal of The Franklin Institute-engineering and Applied Mathematics*, 310: 1–26.

Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; and Zhang, Y. 2023. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12504–12513.

Chen, X.; Zhang, Q.; Li, X.; Chen, Y.; Feng, Y.; Wang, X.; and Wang, J. 2022. Hallucinated neural radiance fields in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12943–12952.

Cui, Z.; Gu, L.; Sun, X.; Ma, X.; Qiao, Y.; and Harada, T. 2024. Aleth-nerf: Illumination adaptive nerf with concealing field assumption. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 1435–1444.

Cui, Z.; Li, K.; Gu, L.; Su, S.; Gao, P.; Jiang, Z.; Qiao, Y.; and Harada, T. 2022. You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction. *arXiv preprint arXiv:2205.14871*.

Fridovich-Keil, S.; Yu, A.; Tancik, M.; Chen, Q.; Recht, B.; and Kanazawa, A. 2022. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5501–5510.

Guo, C.; Li, C.; Guo, J.; Loy, C. C.; Hou, J.; Kwong, S.; and Cong, R. 2020. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1780–1789.

Guo, Y.-C.; Kang, D.; Bao, L.; He, Y.; and Zhang, S.-H. 2022. Nerfren: Neural radiance fields with reflections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18409–18418.

Huang, J.; Zhao, F.; Zhou, M.; Xiao, J.; Zheng, N.; Zheng, K.; and Xiong, Z. 2023a. Learning sample relationship for exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9904–9913.

Huang, X.; Li, W.; Hu, J.; Chen, H.; and Wang, Y. 2023b. Refsr-nerf: Towards high fidelity and super resolution view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8244–8253.

Huang, X.; Liu, M.-Y.; Belongie, S.; and Kautz, J. 2018. Multimodal unsupervised image-to-image translation. In *Proceedings of the European Conference on Computer Vision*, 172–189.

Huang, X.; Zhang, Q.; Feng, Y.; Li, H.; Wang, X.; and Wang, Q. 2022. Hdr-nerf: High dynamic range neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18398–18408.

Jiang, Y.; Ji, D.; Han, Z.; and Zwicker, M. 2020. Sdfdiff: Differentiable rendering of signed distance fields for 3d shape optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1251–1261.

Land, E. H. 1986. An alternative technique for the computation of the designator in the retinex theory of color vision. *Proceedings of the National Academy of Sciences*, 83(10): 3078–3080.

Lee, H.-Y.; Tseng, H.-Y.; Huang, J.-B.; Singh, M.; and Yang, M.-H. 2018. Diverse image-to-image translation via disentangled representations. In *Proceedings of the European Conference on Computer Vision*, 35–51.

Liu, H.-T. D.; Tao, M.; and Jacobson, A. 2018. Paparazzi: surface editing by way of multi-view image processing. *ACM Transactions on Graphics*, 37(6): 221.

Lombardi, S.; Simon, T.; Saragih, J.; Schwartz, G.; Lehrmann, A.; and Sheikh, Y. 2019. Neural volumes: Learning dynamic renderable volumes from images. *arXiv preprint arXiv:1906.07751*.

Ma, L.; Li, X.; Liao, J.; Zhang, Q.; Wang, X.; Wang, J.; and Sander, P. V. 2022a. Deblur-nerf: Neural radiance fields from blurry images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12861–12870.

Ma, L.; Ma, T.; Liu, R.; Fan, X.; and Luo, Z. 2022b. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5637–5646.

Mao, Q.; Lee, H.-Y.; Tseng, H.-Y.; Ma, S.; and Yang, M.-H. 2019. Mode seeking generative adversarial networks for diverse image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1429–1437.

Martin-Brualla, R.; Radwan, N.; Sajjadi, M. S.; Barron, J. T.; Dosovitskiy, A.; and Duckworth, D. 2021. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7210–7219.

Mildenhall, B.; Hedman, P.; Martin-Brualla, R.; Srinivasan, P. P.; and Barron, J. T. 2022. 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*, 16190–16199.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2020. Nerf: Representing scenes as neural radiance fields for view synthesis. In *Proceedings of the European Conference on Computer Vision*, 405–421.

Müller, T.; Evans, A.; Schied, C.; and Keller, A. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4): 1–15.

Nayar; and Branzoi. 2003. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1168–1175.

Nsampi, N. E.; Hu, Z.; and Wang, Q. 2021. Learning Exposure Correction Via Consistency Modeling. In *British Machine Vision Conference*.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32.

Ren, W.; Zhu, Z.; Sun, B.; Chen, J.; Pollefeys, M.; and Peng, S. 2024. NeRF On-the-go: Exploiting Uncertainty for Distractor-free NeRFs in the Wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8931–8940.

Roveri, R.; Öztireli, A. C.; Pandele, I.; and Gross, M. 2018. Pointpronets: Consolidation of point clouds with convolutional neural networks. In *Computer Graphics Forum*, volume 37, 87–99.

Schönberger, J. L.; and Frahm, J.-M. 2016. Structure-from-Motion Revisited. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4104–4113.

Shang, K.; Shao, M.; Wang, C.; Cheng, Y.; and Wang, S. 2024. Multi-Domain Multi-Scale Diffusion Model for Low-Light Image Enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 4722–4730.

Suhail, M.; Esteves, C.; Sigal, L.; and Makadia, A. 2022. Generalizable patch-based neural rendering. In *Proceedings of the European Conference on Computer Vision*, 156–174.

Tewari, A.; Fried, O.; Thies, J.; Sitzmann, V.; Lombardi, S.; Sunkavalli, K.; Martin-Brualla, R.; Simon, T.; Saragih, J.; Nießner, M.; et al. 2020. State of the art on neural rendering. In *Computer Graphics Forum*, volume 39, 701–727.

Tomasi, C.; and Manduchi, R. 1998. Bilateral filtering for gray and color images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 839–846.

Wang, H.; Xu, X.; Xu, K.; and Lau, R. W. 2023a. Lighting up nerf via unsupervised decomposition and enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12632–12641.

Wang, P.; Zhao, L.; Ma, R.; and Liu, P. 2023b. Bad-nerf: Bundle adjusted deblur neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4170–4179.

Wang, Q.; Wang, Z.; Genova, K.; Srinivasan, P. P.; Zhou, H.; Barron, J. T.; Martin-Brualla, R.; Snavely, N.; and Funkhouser, T. 2021. Ibrnet: Learning multi-view image-based rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4690–4699.

Wang, Z.; Bovik, A.; Sheikh, H.; and Simoncelli, E. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612.

Wang, Z.; Li, L.; Shen, Z.; Shen, L.; and Bo, L. 2022. 4k-nerf: High fidelity neural radiance fields at ultra high resolutions. *arXiv preprint arXiv:2212.04701*.

Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep Retinex Decomposition for Low-Light Enhancement. *arXiv preprint arXiv:1808.04560*.

Xiao, J.; Gu, S.; and Zhang, L. 2020. Multi-domain learning for accurate and few-shot color constancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3258–3267.

Yang, Y.; Zhang, S.; Huang, Z.; Zhang, Y.; and Tan, M. 2023. Cross-ray neural radiance fields for novel-view synthesis from unconstrained image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 15901–15911.

Zeng, J.; Bao, C.; Chen, R.; Dong, Z.; Zhang, G.; Bao, H.; and Cui, Z. 2023. Mirror-nerf: Learning neural radiance fields for mirrors with whitted-style ray tracing. In *Proceedings of the ACM International Conference on Multimedia*, 4606–4615.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 586–595.

Zhou, D.; Yang, Z.; and Yang, Y. 2023. Pyramid diffusion models for low-light image enhancement. *arXiv preprint arXiv:2305.10028*.

Zhu, C.; Wan, R.; Tang, Y.; and Shi, B. 2023. Occlusion-free scene recovery via neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20722–20731.