

Bright-NeRF: Brightening Neural Radiance Field with Color Restoration from Low-Light RAW Images

Min Wang¹, Xin Huang¹, Guoqing Zhou¹, Qifeng Guo², Qing Wang^{1*}

¹School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China

²Shenzhen Shenzi Weilai Co., Ltd., Shenzhen 518000, China

(wangmin_phd, xinquang)@mail.nwpu.edu.cn, (zhouguoqing, qwang)@nwpu.edu.cn, guoqifeng001@gmail.com

Abstract

Neural Radiance Fields (NeRFs) have demonstrated prominent performance in novel view synthesis. However, their input heavily relies on image acquisition under normal light conditions, making it challenging to learn accurate scene representation in low-light environments where images typically exhibit significant noise and severe color distortion. To address these challenges, we propose a novel approach, **Bright-NeRF**, which learns enhanced and high-quality radiance fields from multi-view low-light raw images in an unsupervised manner. Our method simultaneously achieves color restoration, denoising, and enhanced novel view synthesis. Specifically, we leverage a physically-inspired model of the sensor's response to illumination and introduce a chromatic adaptation loss to constrain the learning of response, enabling consistent color perception of objects regardless of lighting conditions. We further utilize the raw data's properties to expose the scene's intensity automatically. Additionally, we have collected a multi-view low-light raw image dataset to advance research in this field. Experimental results demonstrate that our proposed method significantly outperforms existing 2D and 3D approaches. Our code and dataset will be made publicly available.

Introduction

Neural Radiance Fields (NeRFs) have revolutionized the field of novel view synthesis by learning implicit scene representations from multi-view images, enabling the generation of high-quality views. Recently, NeRF has shown great potential in various applications such as relighting (Zhang et al. 2021; Rudnev et al. 2022), editing (Yuan et al. 2022; Song et al. 2023), autonomous driving (Tonderski et al. 2024), robot navigation (Adamkiewicz et al. 2022) and object detection (Xu et al. 2023; Hu et al. 2023), demonstrating its versatility and robustness. However, the NeRF framework predominantly relies on multi-view images captured under well-lit conditions. In low-light environments, NeRF struggles to reconstruct scenes accurately due to inherent visual quality deficiencies such as low visibility, poor contrast, and significant noise, which severely impair its performance. Addressing these challenges is crucial for extending NeRF's applicability to diverse real-world scenarios.

*Corresponding Author

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

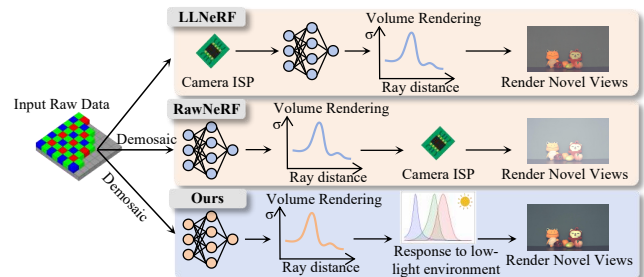


Figure 1: Comparison of LLNeRF, RawNeRF, and our Bright-NeRF. LLNeRF is trained on ISP-finished sRGB images, struggling to recover accurate color in extremely dark conditions via the low bit-depth sRGB data. Utilizing high bit-depth linear raw data, RawNeRF relies on a pre-calibrated ISP for post-processing the rendered images, which involves a complex and costly calibration process. Using a generic ISP offers a compromise, but it fails to deliver high-quality results across different cameras and scenarios. In contrast, Bright-NeRF eliminates these limitations by estimating the adaptive sensor's response to low-photon lighting conditions, ensuring consistent color perception and producing vivid color.

Several NeRF variants have attempted to reconstruct 3D scenes from degraded inputs such as reconstructing HDR scene representation from LDR views (Huang et al. 2022), motion blur (Ma et al. 2022a; Lee et al. 2023), large illumination variations (Martin-Brualla et al. 2021), and aliasing (Barron et al. 2021, 2022). Despite the prevalence of these conditions in real-world scenarios, novel view synthesis in low-light environments has not yet been adequately addressed by these methods. To achieve normal-light novel view synthesis from images captured in low-light conditions, an intuitive approach involves initially enhancing multi-view images using 2D low-light enhancement methods (Guo, Li, and Ling 2016; Chen et al. 2018; Ma et al. 2022b; Dong et al. 2022; Jin et al. 2023), followed by learning a NeRF from the processed images. However, these 2D methods tend to overfit to the training data processed by a fixed ISP, leading to poor generalization across the images captured by different cameras. Most importantly, they typically process each image independently and do not ensure

consistency across views during rendering (as shown in Fig. 5), resulting in bias-guided NeRF and inconsistent enhancement results.

Recently, some low-light image enhancement (LLIE) NeRF methods have been proposed. LLNeRF (Wang et al. 2023) focuses on recovering scene representations under normal-light conditions from a set of low-light sRGB images by decomposing colors into lighting-related and reciprocal components. However, it has difficulty in accurately decomposing colors in extremely low-light conditions (as shown in Fig 6). Additionally, limited by the bit-depth of sRGB images, LLNeRF struggles to recover fine color details, leading to unnatural color and lighting in the rendered images. Compared to sRGB data, raw data’s unique high bit-depth characteristic is advantageous for low-light enhancement tasks. Under extremely low-light conditions, the linear properties and high bit-depth of raw images facilitate the recovery of normal-light images with improved clarity and reduced noise from degraded low-light images. Utilizing this observation, RawNeRF (Mildenhall et al. 2022) learns normally exposed scene representations from multi-view low-light raw images and renders novel views by incorporating a known camera internal ISP. However, as shown in Fig. 1, its effectiveness is substantially constrained by its reliance on the camera’s pre-calibrated internal ISP, since calibrating the ISP of a specific camera is a tedious and complex process. Using a generic ISP offers a compromise, but this approach fails to achieve high-quality results across various cameras and scenarios, leading to suboptimal presentation and limited applicability.

In this paper, we propose a novel unsupervised method called **Bright-NeRF** that enables the rendering of normal-light novel views from a set of low-light raw images. Our method directly reconstructs an enhanced normal-light radiance field using only low-light images for supervision and avoids the need for a pre-calibrated ISP for post-processing. Specifically, inspired by physically-based imaging theory, we model the sensor’s raw-RGB response to scene illumination through collaborative learning with implicit scene representation, effectively compensating for the adverse effects of weak environmental illumination on image quality. Concurrently, we achieve effective noise suppression by aggregating information from multi-views. Leveraging the linear characteristics of raw images, we implement automatic exposure adjustment for low-light images. To address the significant performance degradation that can occur when transferring models trained on synthetic datasets to real-world scenarios (a challenge known as the “sim-to-real gap”), we have collected a **Low-light Multi-view RAW** dataset (**LMRAW**). Our contributions are as follows:

- We propose an unsupervised framework, **Bright-NeRF**, that enables novel view synthesis under normal lighting conditions from a set of low-light raw images. Our approach effectively achieves noise suppression, color distortion correction, and brightness enhancement without the supervision of normal-light images.
- We develop a model to characterize the sensor’s raw-RGB responses to scene illumination based on

physically-inspired imaging theory, which mitigates color distortion caused by low-light conditions.

- We have collected an **LMRAW** dataset. Extensive experiments and comparisons with various 2D image enhancement methods and NeRF-based approaches demonstrate that our method achieves state-of-the-art performance and maintains multi-view consistency on this dataset.

Related Work

Neural Radiance Field and Extensions

NeRF (Mildenhall et al. 2021) has drawn widespread attention due to its powerful representation of a scene using a set of posed images as input, they optimize a scene function via rendering the pixel color by volume rendering scheme along the corresponding ray, thus enabling novel views synthesis.

A primary limitation of NeRF lies in its assumption of ideal input views, which are captured under normal light without large illumination variations and do not contain blur. To address this constraint, researchers have proposed various innovative methods tailored to different types of scene degradation. Deblur-NeRF (Ma et al. 2022a) introduces a deformable sparse kernel module to explicitly simulate the physical blurring process. By synthesizing blurred images to match the input, it enables the recovery of clear scene representations from blurry images. NeRF-W (Martin-Brualla et al. 2021) focuses on modeling in-the-wild images characterized by illumination variations and transient objects by incorporating appearance embedding and transient embedding, implicitly integrating illumination changes and transient objects into the neural implicit representation. There is another line of work exploring the use of NeRF in low-light conditions. RawNeRF (Mildenhall et al. 2022) proposes to train NeRF in the raw data domain and employ Image Signal Processing (ISP) to post-process the rendered images. LLNeRF (Wang et al. 2023) decomposes the color of sampling points into lighting-related and reciprocal components, enabling the recovery of normal-light images from a set of low-light sRGB images. Despite the advancements in LLIE NeRF methods, they either depend on a pre-calibrated ISP for post-processing the renderings or struggle to recover accurate color details under extremely low-light conditions.

Low-light Image Enhancement

Low-light image enhancement aims to improve the visibility of images taken in insufficient illumination conditions while simultaneously suppressing noise, correcting color distortions, and preserving details. Deep learning-based image enhancement methods have demonstrated significant efficacy in this domain. LIME (Guo, Li, and Ling 2016) proposes imposing structural priors to optimizer the illumination map, enabling to increase the naturalness of the enhancement. Retinex-Net (Wei et al. 2018) introduces a dual-module structure comprising a reflectance module and an illumination adjustment module, combined with BM3D (Dabov et al. 2006) for noise suppression.

Recently some approaches have been proposed to directly retouch raw data from camera sensors into high-quality output images. SID (Chen et al. 2018) pioneers this research

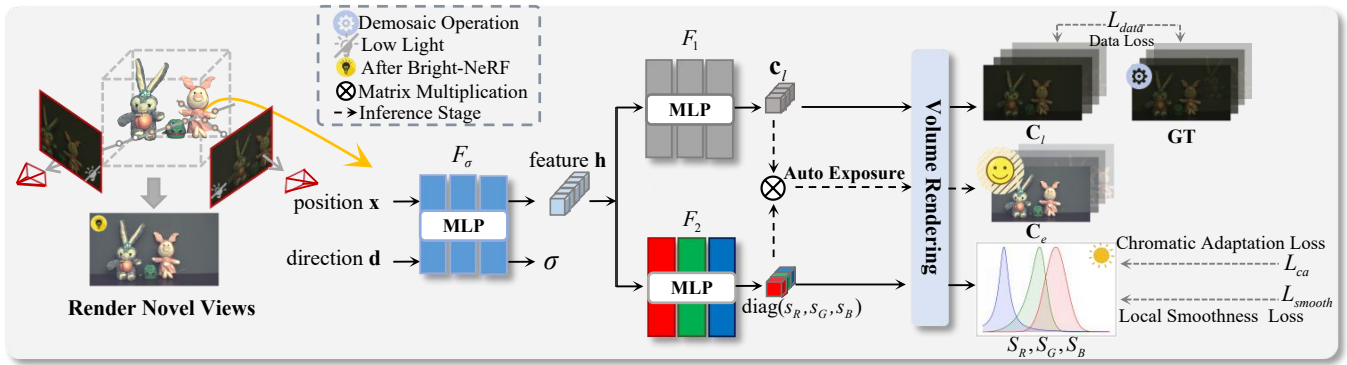


Figure 2: **Method overview.** In the training stage, we firstly sample a series of points \mathbf{x} from each ray and feed them into F_σ along with the direction \mathbf{d} to obtain the view-dependent features \mathbf{h} and density σ for each point. The low-light color \mathbf{c}_l is then calculated by an MLP F_1 . Another MLP F_2 is designed to estimate the sensor’s adaptive response to low-photon lighting conditions $\text{diag}(s_R, s_G, s_B)$. The chromatic adaptation loss L_{ca} optimizes the estimations regarding the sensor’s response to address color distortion in low-light conditions, while the local smooth loss L_{smooth} ensures smoothness in the learned response map $S_k, k \in \{R, G, B\}$. The data loss optimizes the modeling for low-light images. In the inference stage, the estimated sensor’s adaptive response $\text{diag}(s_R, s_G, s_B)$ is further applied to \mathbf{c}_l to obtain distortion-corrected color and automatic exposure adjustment is achieved by utilizing the linear characteristics of raw data to increase the brightness. Through the joint learning of the low-light color field and the color restoration field, along with implicit noise smoothing and automatic exposure adjustment, a bright NeRF is eventually obtained.

direction by proposing a fully convolutional neural network directly training on raw images. SGN (Gu et al. 2019) adopts a top-down self-supervised structure to effectively exploit multi-scale image information. (Lamba and Mitra 2021) proposes an amplifier module capable of directly estimating the amplification factor from input images, eliminating the need for ground truth exposure values to estimate the pre-amplification factor. DNF (Jin et al. 2023) leverages the characteristics of raw and sRGB data domains to decouple domain-specific subtasks, avoiding domain ambiguity issues between raw and sRGB domains. However, these 2D methods are prone to overfitting the training images processed by a fixed ISP, leading to poor generalization across different cameras and scenarios.

Preliminaries

Neural Radiance Field

NeRF (Mildenhall et al. 2021) represents a scene by learning a continuous function in 3D space, mapping a ray origin \mathbf{o} and view direction \mathbf{d} into volume density σ and color \mathbf{c} . This process is defined as:

$$\begin{cases} (\mathbf{h}, \sigma) = F_\sigma(\mathbf{o}, \mathbf{d}; \Theta_{F_\sigma}), \\ \mathbf{c} = F_1(\mathbf{h}; \Theta_{F_1}), \end{cases} \quad (1)$$

where \mathbf{h} is the view-dependent features learned by the neural network F_σ and F_1 is the color field to output the view-dependent color.

More specifically, suppose a camera ray \mathbf{r} is emitted from camera center \mathbf{o} with direction \mathbf{d} , *i.e.*, $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ where t denotes the distance from the origin \mathbf{o} . The expected color $\mathbf{C}(\mathbf{r})$ of \mathbf{r} is defined as:

$$\mathbf{C}(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})dt, \quad (2)$$

where t_n and t_f denote the near and far boundary of the ray respectively.

Method

In this section, we introduce our method Bright-NeRF for recovering enhanced neural radiance fields. The overall pipeline of Bright-NeRF is summarized in Fig. 2. Our method aims to address the challenges inherent in low-light scenes, focusing on novel view synthesis with color restoration, noise reduction, and brightness enhancement. We begin by detailing the architecture of Bright-NeRF, followed by a discussion on how we tackle these specific challenges. Finally, we outline our optimization strategies to improve the performance and quality of our method.

Overall Framework

Given a set of raw images captured under low-light conditions, our objective is to generate novel views as they would appear under normal lighting conditions. We assume that the density field of the scene representation under low-light and normal light conditions remains constant while the color field differs due to the influence of scene illumination. Specifically, the color field in low-light conditions exhibits lower brightness, stronger noise, and distorted color compared to well-lit conditions. Inspired by this observation, we propose learning a color restoration field from the view-dependent features \mathbf{h} , aiming to recover the adaptive sensor’s response to low-photon conditions in the scene. This approach enables the rendering of vivid colors, noise-free, and properly illuminated scene representations.

Specifically, the low-light pixel values $\mathbf{C}_l(\mathbf{r})$ and the en-

hanced pixel values $\mathbf{C}_e(\mathbf{r})$ can be computed as:

$$\mathbf{C}_l(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}_l(\mathbf{r}(t), \mathbf{d})dt, \quad (3)$$

$$\mathbf{C}_e(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}_e(\mathbf{r}(t), \mathbf{d})dt, \quad (4)$$

where \mathbf{c}_l and \mathbf{c}_e represent the color at point $\mathbf{r}(t)$ along the ray \mathbf{r} before and after enhancement, respectively. We then express the relationship between \mathbf{c}_l and \mathbf{c}_e as:

$$\mathbf{c}_e = \phi(\mathbf{c}_l), \quad (5)$$

where ϕ is a comprehensive enhancement function capable of performing three key operations: denoising, color restoration, and brightness improvement. By decomposing the enhancement of the scene representation into these subtasks, our model can address each challenge more effectively in an unsupervised manner.

Unsupervised Enhancement

Color Restoration. In low-light conditions, recovering an object’s true color function is a complex task, as the reflected color depends on the object’s intrinsic color and environmental illumination. Mathematically, to form a color image, consider a spatial point \mathbf{x} along a ray \mathbf{r} corresponding to a specific pixel, the color at this point can be expressed as:

$$c_k(\mathbf{x}) = \int_{\omega} c(\mathbf{x}, \lambda) s_k(\lambda) d\lambda, \quad (6)$$

where $c(\mathbf{x}, \lambda)$ is the measurement at position \mathbf{x} with respect to the wavelength λ , $s_k(\lambda)$ denotes the sensor characteristics as a function of wavelength λ , over the visible spectrum ω . The subscript k indicates the sensor’s response in the k channel ($k \in \{R, G, B\}$). In low-light conditions, the unknown sensor characteristics of the low-photon environmental lighting make it challenging to obtain an illumination-invariant image descriptor. To address this issue, our model implicitly estimates the sensor’s response to a low-light environment to eliminate the color distortion, enabling us to render the scene’s tone independently of its unsatisfactory environmental illumination.

We adopt the von Kries adaptation hypothesis (Brainard and Wandell 1992), which suggests that the three color sensors in the human visual system each operate with independent gain control. Based on this principle, we implicitly estimate a diagonal matrix $\text{diag}(s_R, s_G, s_B)$ as follows:

$$\text{diag}(s_R, s_G, s_B) = F_2(\mathbf{h}), \quad (7)$$

where \mathbf{h} is the view-dependent feature derived by Eq. 1. F_2 is an MLP used to model the adaptive response of the visual system to low-photon lighting conditions. By applying the estimated sensor response to the predicted color $\mathbf{c}_l(\mathbf{x}, \mathbf{d})$ of a spatial point \mathbf{x} , we can recover the original color:

$$\mathbf{c}_s(\mathbf{x}, \mathbf{d}) = \mathbf{c}_l(\mathbf{x}, \mathbf{d}) \times \text{diag}(s_R, s_G, s_B). \quad (8)$$

Note that our estimated sensor’s response can adaptively learn to accurately correct colors regardless of how dark the scene is, which is further analyzed in the experimental section.

Brightness Improvement. There is a linear relationship exists between input brightness and output pixel values in raw signals:

$$\mathbf{c}_e(\mathbf{r}) = \alpha \cdot \mathbf{c}_s(\mathbf{r}), \quad (9)$$

where α is the proportionality coefficient. Inspired by automatic exposure techniques, we dynamically adjust the transition curve slope α based on the average scene intensities to adapt to brightness levels.

Denoising. We empirically adopt a heteroscedastic Gaussian process model in raw data: (Mohsen, Tompsett, and Sèquin 1975; Liu, Tanaka, and Okutomi 2014):

$$n(\mathbf{r}) \sim N(0, \beta^2 \mathbf{c}_e(\mathbf{r}) + \delta^2), \quad (10)$$

where β is a sensor-specific scaling factor of the signal and δ^2 is the variance of the distribution.

Consider a spatial point \mathbf{x} with a large density in the scene, which has multiple projections $P_{\mathbf{x}} = \{\tilde{\mathbf{c}}_l(\mathbf{r})\}$ in the training images. For a color $\tilde{\mathbf{c}}_l(\mathbf{r})$ and its underlying noise-free color $\bar{\mathbf{c}}_e(\mathbf{r})$, their relationship can be written as:

$$\tilde{\mathbf{c}}_l(\mathbf{r}) = \bar{\mathbf{c}}_e(\mathbf{r}) + n(\mathbf{r}), \quad (11)$$

where $n(\mathbf{r})$ represents the noise corrupting $\bar{\mathbf{c}}_e(\mathbf{r})$. During training, the predicted color $\mathbf{c}_e(\mathbf{r})$ is supervised by all pixels in $P_{\mathbf{x}}$. Since the loss function for all rays corresponding to pixels in $P_{\mathbf{x}}$ is an unweighted average, the neural network tends to learn parameters that minimize the overall average deviation. Consequently, the learned $\mathbf{c}_e(\mathbf{r})$ would converge towards the expectation of $\tilde{\mathbf{c}}_l(\mathbf{r})$, *i.e.*,

$$\mathbf{c}_e(\mathbf{r}) \approx \mathbb{E}\{\tilde{\mathbf{c}}_l(\mathbf{r})\} = \bar{\mathbf{c}}_e(\mathbf{r}) + \mathbb{E}\{n\}. \quad (12)$$

As the noise in raw signal is zero-mean, *i.e.*, $\mathbb{E}\{n\} = 0$, it indicates that the multi-view optimization inherent in neural radiance fields can effectively smooth images and reduce the noise.

Optimization

In this section, we present the loss functions that guide the unsupervised training of Bright-NeRF: chromatic adaptation loss directs the optimization of the enhancement process, ensuring that the color adaptation is performed effectively and consistently in low-light conditions, and data loss is responsible for optimizing the radiance field. The combination of these loss functions enables Bright-NeRF to learn effectively without the need for paired low-light and normal-light image data, which is often difficult or expensive to obtain.

Chromatic Adaptation Loss. To enhance the accuracy of our estimated sensor’s response to low photon illumination in the scene, we introduce a chromatic adaptation loss, designed to constrain the learning of $\text{diag}(s_R, s_G, s_B)$, promoting to produce more realistic images. The chromatic adaptation loss, denoted as L_{ca} , is formulated as:

$$L_{ca} = \frac{1}{3} \sum_{\mathbf{r} \in \mathbb{R}} \sum_{k \in \{R, G, B\}} \left\| \frac{K_{avg}}{\bar{C}_k} - S_k(\mathbf{r}) \right\|^2, \quad (13)$$

where $K_{avg} = \frac{1}{3}(\bar{C}_R + \bar{C}_G + \bar{C}_B)$ and $\bar{C}_k = \frac{1}{N} \sum_{\mathbf{r} \in \mathbb{R}} C_k(\mathbf{r})$, N is the number of sampled rays. And $S_k(\mathbf{r}) =$

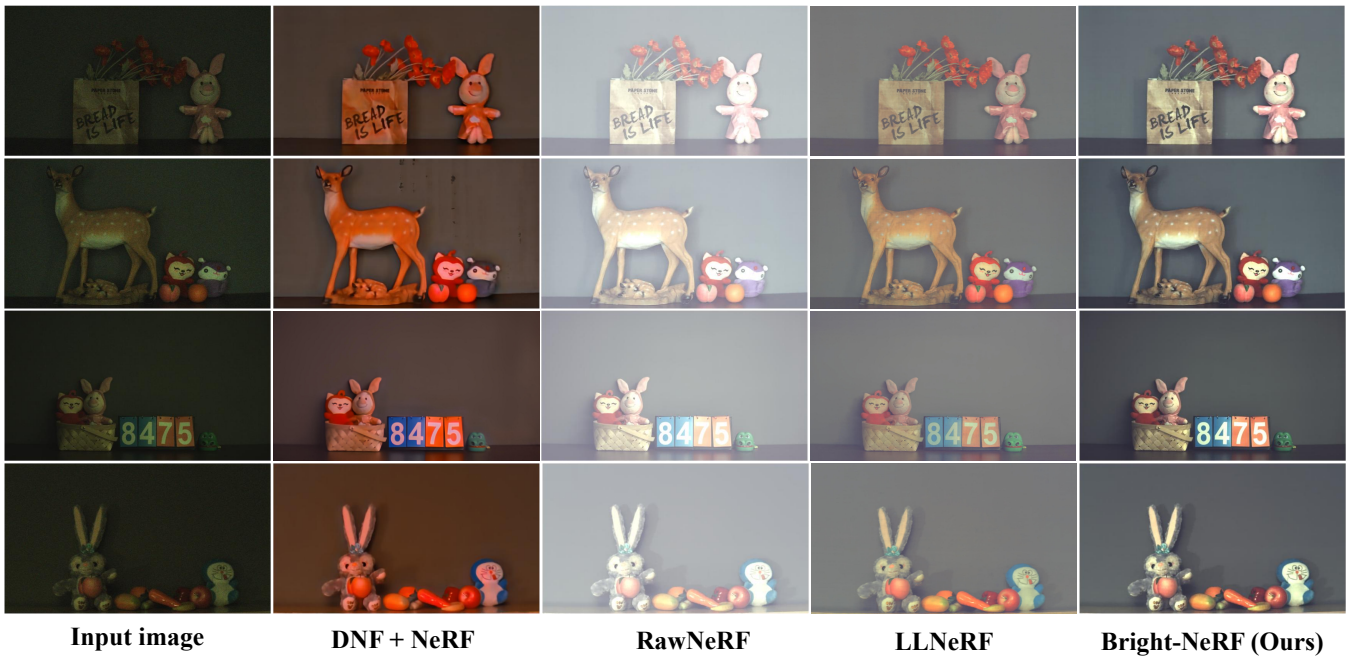


Figure 3: **Qualitative comparison with state-of-the-art LLIE NeRF methods.** Leveraging the sensor’s adaptive responses in the color restoration field, the results of our Bright-NeRF exhibit more natural color and more accurate tone restoration.

$\int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))s_k(\mathbf{r}(t), \mathbf{d})dt$, which denotes the integrated sensor’s response to the low-light scene along the ray \mathbf{r} for channel k .

Local Smoothness Loss. To ensure that the estimated sensor’s response to illumination maintains piece-wise smoothness, we impose the following constraint:

$$L_{smooth} = \sum_{\mathbf{r} \in \mathbb{R}} \left\| \frac{\gamma_1 * \mathbf{S}_v^2(\mathbf{r})}{\mathbf{C}_h^2(\mathbf{r}) + \epsilon} \right\| + \left\| \frac{\gamma_2 * \mathbf{S}_h^2(\mathbf{r})}{\mathbf{C}_v^2(\mathbf{r}) + \epsilon} \right\|, \quad (14)$$

where \mathbf{S}_v , \mathbf{S}_h and \mathbf{C}_v , \mathbf{C}_h represent the differences in the integral of sensor responses and colors between adjacent rays in the vertical and horizontal directions, respectively. γ_1 and γ_2 are weighting factors and ϵ is a small positive constant to prevent division by zero and ensure numerical stability.

Data Loss. Given that the pixels in our training data are predominantly low intensity due to the low-light conditions, we adopt the color loss function proposed in (Mildenhall et al. 2022), which amplifies the error in dark regions, making the optimization process more sensitive to subtle differences in low-light areas. The data loss is formulated as:

$$L_{data} = \sum_{\mathbf{r} \in \mathbb{R}} \|\psi(\mathbf{C}_l(\mathbf{r})) - \psi(\mathbf{C}_{GT}(\mathbf{r}))\|^2, \quad (15)$$

where \mathbf{C}_l represents the rendered low-light color and \mathbf{C}_{GT} denotes the low-light ground truth color. $\psi(y) = \log(y + \epsilon)$ is the linearized tone mapping function.

Above all, our loss function L comprises three components: data loss L_{data} , along with two unsupervised losses L_{ca} and L_{smooth} .

$$L = \lambda_1 L_{data} + \lambda_2 L_{ca} + \lambda_3 L_{smooth}, \quad (16)$$

where λ_1 , λ_2 , and λ_3 are three positive weights, which are set to 1.0, 0.1, and 0.1 respectively.

Train and Test Schemes. At the training phase, Bright-NeRF simultaneously renders low-light images \mathbf{C}_l and incorporates the estimated sensor’s response $\text{diag}(s_R, s_G, s_B)$ to render the sensor’s response map \mathbf{S} at the volume rendering stage. At the testing stage, we leverage the comprehensive enhancement function in Eq. 5 and perform volume rendering in Eq. 4 to directly produce enhanced normal-light images \mathbf{C}_e .

Experiments

Dataset

To facilitate the training and evaluation of our model, we collect a multi-view low-light raw image dataset LMRAW. While previous work (Mildenhall et al. 2022) includes some low-light scenarios, their focus is primarily on raw denoising rather than low-light enhancement. The absence of ground truth for low-light conditions in their test scenes makes it difficult to assess the enhancement performance for novel view synthesis.

Our dataset comprises scenes captured under challenging low-light conditions with illumination levels ranging from 0.01 to 0.05 lux. Each scene consists of 18 to 30 images at 1920×1080 resolution. To capture multi-view images, we move and rotate the camera tripod. Since COLMAP (Schonberger and Frahm 2016) does not support processing raw images directly, we use full-resolution post-processed JPEG images for pose estimation.

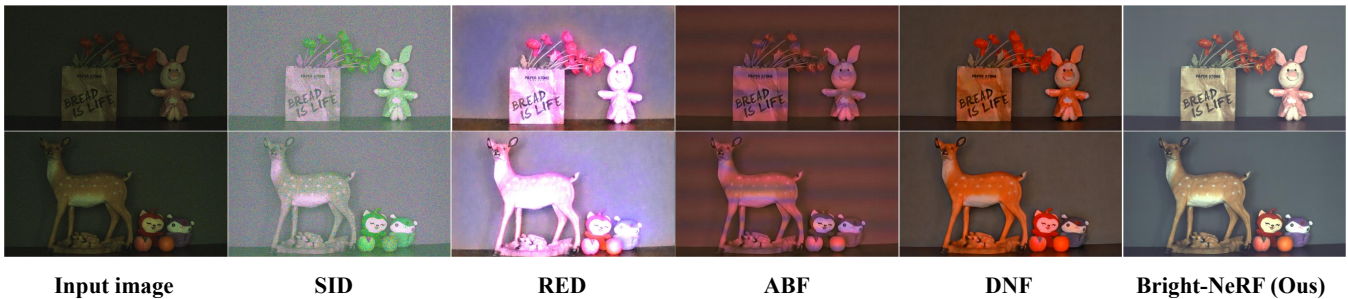


Figure 4: **Qualitative comparison with state-of-the-art 2D LLIE methods.** We show that our method effectively recovers detailed geometry and accurate colors while other methods often struggle with denoising and tend to produce incorrect colors. For these 2D methods, we use publicly available code and pre-trained model weights.

Method	Method Type	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	LIQE \uparrow
SID (Chen et al. 2018)	2D	13.37	0.233	0.786	1.004
RED (Lamba and Mitra 2021)		11.04	0.654	0.731	1.003
ABF (Dong et al. 2022)		18.20	0.826	0.675	1.010
DNF (Jin et al. 2023)		18.60	0.841	0.636	1.211
RawNeRF (Mildenhall et al. 2022)	3D	8.95	0.694	0.524	1.626
DNF (Jin et al. 2023)+NeRF		18.20	0.829	0.663	1.151
LLNeRF (Wang et al. 2023)		16.64	0.811	0.494	1.332
Bright-NeRF (Ours)		22.64	0.874	0.476	1.659

Table 1: **Quantitative comparison with existing LLIE methods.** The calculated metrics are the average values across nine scenes. The best results are marked in **bold**.

Results

Novel View Synthesis. To ensure a fair comparison, we train our Bright-NeRF, the baseline model, RawNeRF (Mildenhall et al. 2022) using the same images, and train LLNeRF (Wang et al. 2023) using post-processed RGB data as it is a RGB-based method. The comparison of novel view synthesis results is illustrated in Fig. 3. For the baseline, we choose to pre-enhance the low-light training images using DNF (Jin et al. 2023) as it provides better geometric information and subsequently trains a NeRF model with these enhanced images. We can see that the baseline model’s results show noticeable discrepancies in color reproduction compared to the expected natural appearance of the scenes. RawNeRF (Mildenhall et al. 2022) struggles with color distortion in extremely dark scenes without a calibrated ISP which is an intricate process, and a generic ISP yields suboptimal performance. Their results exhibit a general washout effect, resulting in hazy appearance and unnatural color reproduction. The rendered results produced by LLNeRF (Wang et al. 2023) demonstrate inadequate restoration of tone and saturation. In contrast, our model exhibits superior performance in producing more natural and vivid colors thanks to our color restoration strategy .

Beyond analyzing the performance of the LLIE NeRF methods, we further investigate the consistency of 2D image enhancement methods during the multi-view enhancement process. As illustrated in Fig. 5, when images from two different views are input into ABF (Dong et al. 2022) and DNF (Jin et al. 2023), they tend to produce inconsistent artifacts

as they enhance each view independently and do not consider the consistency between views. Compared to 2D enhancement methods, our approach guarantee the multi-view consistency in enhancement, ensuring high-quality rendering results.

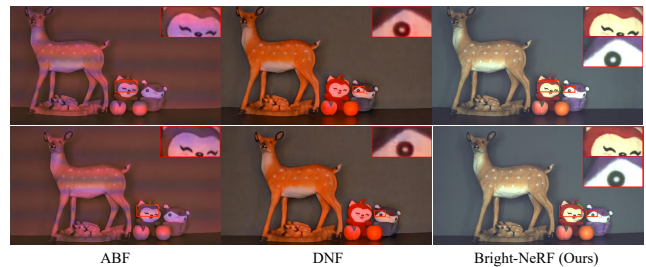


Figure 5: **Visualization of multi-view consistency in 2D LLIE methods and our method.** We present cropped regions of enhancement results from the ABF (Dong et al. 2022) and DNF (Jin et al. 2023) methods at different views. Both methods exhibit enhancement inconsistency such as the color inconsistency in the doll’s “forehead” in the first column and “eye” in the second column. In contrast, our Bright-NeRF ensures high-quality enhancement while maintaining multi-view consistency.

Low-light Enhancement. We conducted experimental comparisons with state-of-the-art 2D raw image low-light enhancement methods, as shown in Fig. 4. We use the pub-

licly available code and pre-trained model weights to enhance the images. Since SID (Chen et al. 2018) and RED (Lamba and Mitra 2021) are trained on the dataset composed of relatively brighter images, they struggle to handle significant noise in darker scenes. While ABF (Dong et al. 2022) and DNF (Jin et al. 2023) can reduce the noise, they tend to overfit to the data processed by the ISP used for supervision in the training dataset, failing to recover accurate and natural colors in diverse camera and scenarios. In contrast, our model effectively improves image brightness while simultaneously reducing noise, recovering natural colors from the color distortions present in low-light conditions.

Adaptability and Stability across Different Darkness Degrees. To investigate the adaptability and stability of our method across different brightness levels, we simulate scenes with different darkness degrees. The simulated exposure ratio is defined as the ratio of the exposure time compared to the original raw image. Leveraging the linear characteristics of raw data (Zhu et al. 2020), we generate 8 sets of scene darkness by simulating the exposure ratio and make a comparison with LLNeRF (Wang et al. 2023) which achieves the second best performance in existing LLIE NeRF methods, as shown in Fig. 6. As the scene becomes darker, LLNeRF suffers from increasingly severe color bias. We attribute this to the difficulty LLNeRF faces in accurately decomposing colors into lighting-related and reciprocal components in extremely dark scenes, as well as the limitations of low bit-depth RGB data in recovering color details. Despite these challenges, our method makes use of the high-bit characteristic of raw data and uses the color restoration field to produce accurate color details, even under extremely low exposure ratios such as 0.3 and 0.4. Our method adaptively renders results with stable brightness and natural color regardless of the scene’s darkness degree. We refer to the implementation of simulated exposure and more comparisons in the supplementary material.

Quantitative Evaluation. We evaluate our model’s performance across nine scenes, as shown in Tab. 1. Our method outperforms existing 2D and LLIE NeRF approaches across PSNR, SSIM, and LPIPS, which indicates that our method performs better in relative sharpness, structure, and perceptual quality.

To further evaluate the rendering and perceptual quality of our method, we leverage the latest non-reference image quality assessment metrics, LIQE (Zhang et al. 2023), which integrates multi-modality into low-level vision perception, ensuring robustness in image quality assessment across diverse scenarios. The comparison on the LIQE metric indicates that the proposed method delivers superior performance in terms of perceptual image quality without any reference information.

Ablation Study. As shown in Tab. 2, we analyze the impact of the color restoration field by removing the branch responsible for estimating the sensor’s response to scene illumination and its related chromatic adaptation loss. This results in the color field being affected by color distortion, leading to incorrect rendering and performance degrada-

tion. With the proposed adaptive color restoration field, we achieve accurate perception of environmental low-photon conditions, enabling correct color rendering. Additionally, by relaxing the constraints of the loss functions, we observe a decline in the quality of results when any item is removed.

$CR + L_{ca}$	L_{smooth}	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	LIQE \uparrow
\times	\times	21.37	0.868	0.737	1.497
\checkmark	\times	22.61	0.873	0.478	1.614
\checkmark	\checkmark	22.64	0.874	0.476	1.659

Table 2: **Ablation study results.** The quality of results is degraded as we remove any item. CR denotes “Color Restoration”.

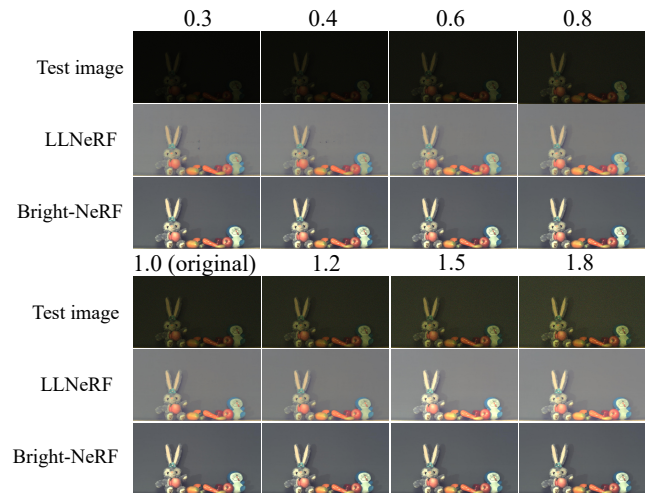


Figure 6: **Performance comparison on scenes with different darkness degrees.** As the simulated exposure ratio decreases (darker scene), LLNeRF’s results exhibit more severe color bias while our Bright-NeRF demonstrates adaptability and stability to different degrees of darkness, producing natural and stable color even in extremely low-light conditions. The simulated exposure ratio is defined as the ratio of the exposure time compared to the original raw image.

Conclusion

In this paper, we have proposed a novel unsupervised method for synthesizing novel views from a set of low-light raw images, simultaneously addressing the challenges of denoising, color restoration, and brightness improvement. Specifically, our method estimates the sensor’s response to low-light scenes jointly with the scene’s radiance field, enabling vivid color restoration and effective noise reduction regardless of the scene’s darkness degree. Our method innovatively enhances scene representation without relying on well-lit supervisions and achieves satisfactory rendering results. We have conducted extensive experimental evaluations to demonstrate its effectiveness in comparison to existing state-of-the-art approaches.

Acknowledgments

This work was supported by NSFC under Grant 62031023.

References

- Adamkiewicz, M.; Chen, T.; Caccavale, A.; Gardner, R.; Culbertson, P.; Bohg, J.; and Schwager, M. 2022. Vision-only robot navigation in a neural radiance world. *IEEE Robotics and Automation Letters*, 7(2): 4606–4613.
- 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.
- Barron, J. T.; Mildenhall, B.; Verbin, D.; Srinivasan, P. P.; and Hedman, P. 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5470–5479.
- Brainard, D. H.; and Wandell, B. A. 1992. Asymmetric color matching: how color appearance depends on the illuminant. *Journal of the Optical Society of America. A, Optics and image science*, 9 9: 1433–48.
- Chen, C.; Chen, Q.; Xu, J.; and Koltun, V. 2018. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3291–3300.
- Dabov, K.; Foi, A.; Katkovnik, V.; and Egiazarian, K. 2006. Image denoising with block-matching and 3D filtering. In *Image processing: algorithms and systems, neural networks, and machine learning*, volume 6064, 354–365. SPIE.
- Dong, X.; Xu, W.; Miao, Z.; Ma, L.; Zhang, C.; Yang, J.; Jin, Z.; Teoh, A. B. J.; and Shen, J. 2022. Abandoning the bayer-filter to see in the dark. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 17431–17440.
- Gu, S.; Li, Y.; Gool, L. V.; and Timofte, R. 2019. Self-guided network for fast image denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2511–2520.
- Guo, X.; Li, Y.; and Ling, H. 2016. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2): 982–993.
- Hu, B.; Huang, J.; Liu, Y.; Tai, Y.-W.; and Tang, C.-K. 2023. Nerf-rpn: A general framework for object detection in nerfs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 23528–23538.
- 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.
- Jin, X.; Han, L.-H.; Li, Z.; Guo, C.-L.; Chai, Z.; and Li, C. 2023. Dnf: Decouple and feedback network for seeing in the dark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18135–18144.
- Lamba, M.; and Mitra, K. 2021. Restoring extremely dark images in real time. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3487–3497.
- Lee, D.; Lee, M.; Shin, C.; and Lee, S. 2023. Dp-nerf: De-blurred neural radiance field with physical scene priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12386–12396.
- Liu, X.; Tanaka, M.; and Okutomi, M. 2014. Practical signal-dependent noise parameter estimation from a single noisy image. *IEEE Transactions on Image Processing*, 23(10): 4361–4371.
- 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.
- 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. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.
- Mohsen, A. M.; Tompsett, M. F.; and Sèquin, C. H. 1975. Noise measurements in charge-coupled devices. *IEEE Transactions on Electron Devices*, 22(5): 209–218.
- Rudnev, V.; Elgharib, M.; Smith, W.; Liu, L.; Golyanik, V.; and Theobalt, C. 2022. Nerf for outdoor scene relighting. In *European Conference on Computer Vision*, 615–631. Springer.
- Schonberger, J. L.; and Frahm, J.-M. 2016. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4104–4113.
- Song, H.; Choi, S.; Do, H.; Lee, C.; and Kim, T. 2023. Blending-nerf: Text-driven localized editing in neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 14383–14393.
- Tonderski, A.; Lindström, C.; Hess, G.; Ljungbergh, W.; Svensson, L.; and Petersson, C. 2024. Neurad: Neural rendering for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14895–14904.
- Wang, H.; Xu, X.; Xu, K.; and Lau, R. W. 2023. Lighting up nerf via unsupervised decomposition and enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12632–12641.
- Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*.

Xu, C.; Wu, B.; Hou, J.; Tsai, S.; Li, R.; Wang, J.; Zhan, W.; He, Z.; Vajda, P.; Keutzer, K.; et al. 2023. Nerf-det: Learning geometry-aware volumetric representation for multi-view 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 23320–23330.

Yuan, Y.-J.; Sun, Y.-T.; Lai, Y.-K.; Ma, Y.; Jia, R.; and Gao, L. 2022. Nerf-editing: geometry editing of neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18353–18364.

Zhang, W.; Zhai, G.; Wei, Y.; Yang, X.; and Ma, K. 2023. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14071–14081.

Zhang, X.; Srinivasan, P. P.; Deng, B.; Debevec, P.; Freeman, W. T.; and Barron, J. T. 2021. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG)*, 40(6): 1–18.

Zhu, M.; Pan, P.; Chen, W.; and Yang, Y. 2020. Eemefn: Low-light image enhancement via edge-enhanced multi-exposure fusion network. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 13106–13113.