Lighting up NeRF via Unsupervised Decomposition and Enhancement

Haoyuan Wang[1], Xiaogang Xu[2,3], Ke Xu[1*], Rynson W.H. Lau[1*]
[1]Department of Computer Science, City University of Hong Kong
[2] Zhejiang Lab, [3] Zhejiang University
https://onpix.github.io/llnerf

Abstract

Neural Radiance Field (NeRF) is a promising approach for synthesizing novel views, given a set of images and the corresponding camera poses of a scene. However, images photographed from a low-light scene can hardly be used to train a NeRF model to produce high-quality results, due to their low pixel intensities, heavy noise, and color distortion. Combining existing low-light image enhancement methods with NeRF methods also does not work well due to the view inconsistency caused by the individual 2D enhancement process. In this paper, we propose a novel approach, called Low-Light NeRF (or LLNeRF), to enhance the scene representation and synthesize normal-light novel views directly from sRGB low-light images in an unsupervised manner. The core of our approach is a decomposition of radiance field learning, which allows us to enhance the illumination, reduce noise and correct the distorted colors jointly with the NeRF optimization process. Our method is able to produce novel view images with proper lighting and vivid colors and details, given a collection of camera-finished low dynamic range (8-bits/channel) images from a low-light scene. Experiments demonstrate that our method outperforms existing low-light enhancement methods and NeRF methods.

1. Introduction

Neural Radiance Field (NeRF) [22] is a powerful approach to render novel view images through learning scene representations as implicit functions. These implicit functions are parameterized by multi-layer perceptrons (MLPs) and optimized by measuring the colorimetric errors of the input views. Consequently, high-quality input images are the precondition for the high-quality results of NeRF. In other words, training NeRF models typically requires the input images to have high visibility, and almost all the pixels to faithfully represent the scene illumination and object colors. However, when taking photos under low-light conditions, the quality of the images is not guaranteed. Low-light images typically have low visibility. Noise from the camera is also relatively amplified due to the low photons, which further buries the scene details and distorts object colors. Such characteristics of low-light images fail existing NeRF models in producing high-quality novel view images.

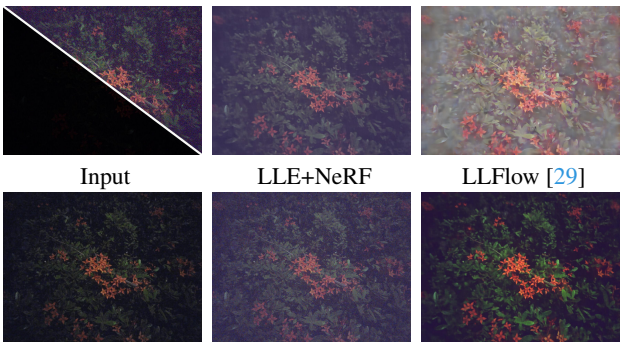


Figure 1. A comparison of the baseline model (LLE+NeRF), SOTA low light enhancement models, and our model.

We note that recently there are some methods proposed to train NeRF models from degraded inputs [21, 32, 18]. Ma *et al*. [18] present a method to synthesize novel view images from blurry inputs taken in normal-light scenes. Mildenhall *et al*. [21] show that when training with high dynamic range RAW data, NeRF can be robust to zero-mean noise of low-light input images. Huang *et al*. [32] propose HDR-NeRF, which produces high dynamic range (HDR) novel views from a set of low dynamic range (LDR) input images taken at different known exposure levels. The latter two methods take advantages of HDR information and metadata (*i.e*., exposure levels) recorded in the RAW images to enhance the scene representations. However, these methods do not work on camera-finished sRGB images (8-bits/channel) taken in low-light scenes. Unlike RAW data, sRGB images are produced by the camera ISP process. They are of low dynamic range and low signal-to-noise ratio.

A straightforward solution to this problem is to first enhance the low-light input images and then use the enhanced

* Joint corresponding authors.

results to train a NeRF model. However, while this may be able to improve the brightness, existing low-light enhancement models do not consider how to maintain consistency across multi-view images. Besides, these learning-based enhancement methods tend to learn specific mappings of brightness from their own training data, which may not generalize well to in-the-wild scenes. These two reasons cause NeRF to learn biased information across different views due to the view-dependent optimization of NeRF, resulting in unrealistic novel images. See examples in Fig. 1.

In this paper, we propose a new approach for rendering novel normal-light images from a set of 8-bit low-light sRGB images without the supervision of ground truth. Our key solution to this problem is that: the colors of 3D points can be decoupled into view-dependent and view-independent components within the NeRF optimization, and the view-dependent component is dominated by the effect of lighting. So the manipulations of the lighting-related view-independent components are able to enhance the brightness, correct the colors, and reduce the noise while keeping the texture and structure of the scene. Experiments demonstrate that the proposed method outperforms the state-of-the-art NeRF models and the baselines (*i.e*., combining NeRF with state-of-the-art enhancement methods).

In summary, we propose the first method to reconstruct a NeRF model of proper lighting from a collection of LDR low-light images. Our main contributions includes:

1. We propose to decompose NeRF into view-dependent and -independent color components for enhancement. The decomposition does not require ground truth.

2. We formulate an unsupervised method to enhance the lighting and correct the colors while rendering noise-free novel view images.

3. We collect a real-world dataset, and conduct extensive experiments to analyze our method and demonstrate its effectiveness in real-world scenes.

2. Related Work

Neural Radiance Field represents 3D scenes via parameterized implicit functions and allows to render high-quality novel view images. However, NeRF is sensitive to the input images as it relies on the colorimetric optimization of the input images. Some methods focus on improving the robustness of NeRF to dynamic scenes in the wild by using, *e.g*., time-of-flight data [4], latent appearance modelling [20], camera self-calibration [16], depth estimation [30, 11], and semantic labels [39].

Some other methods [21, 32, 18] propose to train NeRF models from degraded inputs. Ma *et al*. [18] propose a deformable sparse kernel module for deblurring while synthesizing novel view images from blurry inputs. Mildenhall *et al*. [21] propose to train NeRF directly on camera raw images for handling the low visibility and noise of low-light scenes. Huang *et al*. [32] proposes the HDR-NeRF to synthesize novel view HDR images from a collection of LDR images of different exposure levels, which implicitly handles the exposure fusion using a tone mapper. Unlike the above methods, in this paper, we aim to address the problem of training NeRF using a group of low-light sRGB images, which is more challenging due to the low visibility, low dynamic ranges, large noise, and high color distortions.

Low-light Enhancement aims to improve the content visibility of images taken from low-light scenes. A line of deep enhancement methods learns specific mappings from low-light images to expert-retouched images or images captured with high-end cameras. These methods propose different priors and techniques aiming to enhance the capacity of neural networks for learning such mappings, *e.g*., using HDR information [12, 35, 27], generative adversarial learning [15, 10, 17, 25], deep parametric filters [23], and reinforcement learning [24, 36]. Some methods propose to decompose the images into illumination and detail layers [7], layers of different frequency components [33], and regions of different exposures [3, 14] for enhancement. Recently, Xu *et al*. [34] propose to combine transformer and CNNs to model long-range correlations for low-light enhancement.

Our work is closer in spirit to the Retinex-based enhancement methods [37, 6, 28, 38, 26, 31]. These methods first decompose the input image into the illumination and reflectance layers and then enhance the illumination layer of the image. While these methods learn such decomposition from 2D images, which typically lack geometry information, our method works in the radiance field, resulting in a more realistic decomposition and enhancement.

3. Preliminary Knowledge and Analysis

We first summarize how neural radiance field (NeRF) works under normal-light scenes and then explain the challenges for NeRF to handle low-light scenes.

3.1. NeRF Preliminary

Given a set of posed training images, NeRF [22] learns to render the color of every single pixel $\mathbf{c}_r$ for a ray r, which could be uniquely identified by the camera index and the 2D pixel coordinates. NeRF represents a scene by a radiance field, which takes as input an arbitrary single ray cast $r(t) = \mathbf{o}+t\mathbf{d}$, where $\mathbf{o}, \mathbf{d}, t$ are the ray origin, ray direction, and the distance along the ray, respectively. The rendering process has three steps: (1) NeRF samples n points along the ray $r(t)$, *i.e*., $t_i \in \mathbf{t}$ where $\mathbf{t}$ is a n-D vector, between the near and far image planes using the hierarchical sampling strategy; (2) NeRF applies an optional transform function

$\psi(\cdot)$ to the sampled coordinate vector $\mathbf{t}$ along the ray; and (3) NeRF uses the MLPs $F_{\text{density}}, F_{\text{color}}$ to learn the volume density and the color along the rays, denoted by σ and $\mathbf{c}$, from $\mathbf{t}$ and the view direction $\mathbf{d}$ as:

$$\begin{cases} (\tau, \sigma) = F_{\text{density}}(\psi(r(t_i)), \mathbf{d}; \Theta_{F_{\text{density}}}), \quad t_i \in \mathbf{t} \\ \mathbf{c} = F_{\text{color}}(\tau, \mathbf{d}; \Theta_{F_{\text{color}}}), \end{cases} \tag{1}$$

where τ is the intermediate features learned by the neural network. Different NeRF implementations may have different versions of the transform function $\psi(\cdot)$. The original NeRF implementation [22] uses the frequency positional encoding function as $\psi(\cdot)$, while in Mip-NeRF [5], $\psi(\cdot)$ is implemented as interval splitting and integrated positional encoding. In this paper, we use the implementation of Mip-NeRF [5], and the pixel colors are rendered as:

$$\mathbf{c}_r = \sum_i w_i \mathbf{c}_i = \sum_i \left(1 - e^{-\sigma\delta_i}\right) e^{-\sum_{j<i} \sigma_j \delta_j} \mathbf{c}_i, \tag{2}$$

where $\delta_i = t_{i+1} - t_i$. $\mathbf{c}_r$ is the final rendered 3-channel pixel color of the corresponding ray $r(t)$. NeRF is then optimized under the supervision of the ground-truth pixel colors $\tilde{\mathbf{c}}_r$ of the training images.

3.2. Challenges

Since the NeRF model directly optimizes its implicit radiance field according to the 2D projected images, training a NeRF model using low-light sRGB images has two challenges. First, NeRF cannot handle the low pixel intensity of low-light images, and can only produce dark images as novel views. Second, although [21] shows that NeRF is robust to zero-mean noise in the raw domain due to its essential integration process, the signal-to-noise ratio of the camera-finished sRGB images is much lower than that of the raw images. In addition, the camera ISP process changes the linearity property of raw images and blends scene radiance with noise together in the camera-finished sRGB images. Hence, NeRF is not able to handle noise and color distortion when training on low-light sRGB images.

To obtain a normal-light NeRF, combining low-light enhancement methods with NeRF (LLE+NeRF) may be a possible solution. However, as existing low-light enhancement methods mainly learn a mapping from low light to normal light based on specific training data. This mapping may not generalize well to new scenes that are out of the distributions of the training data. Hence, using images enhanced by these existing methods to train a NeRF model may produce low-quality novel view images. On the other hand, taking multi-view images of both low-light and normal light at the same time as training data is not practical.

In this work, we aim to develop a method to produce high-quality novel view images from low-light scenes in an unsupervised manner.

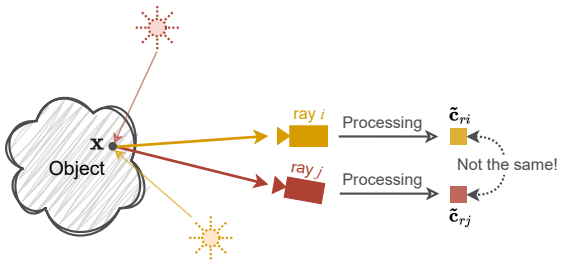


Figure 2. The 2D projection $\tilde{\mathbf{c}}_{ri}, \tilde{\mathbf{c}}_{rj}$ of the same spatial point $\mathbf{x}$ is not exactly identical but in the same color spectrum. The variance of color across views, *i.e.*, the view-dependent component of the observed color, is dominated by the effect of lighting.

4. Our Unsupervised Approach

The main idea of our work is to decompose the implicit radiance field of NeRF and then leverage priors to enhance the lighting, reduce noise and correct the colors of the novel-view images. Fig. 3(c) shows the pipeline of our method.

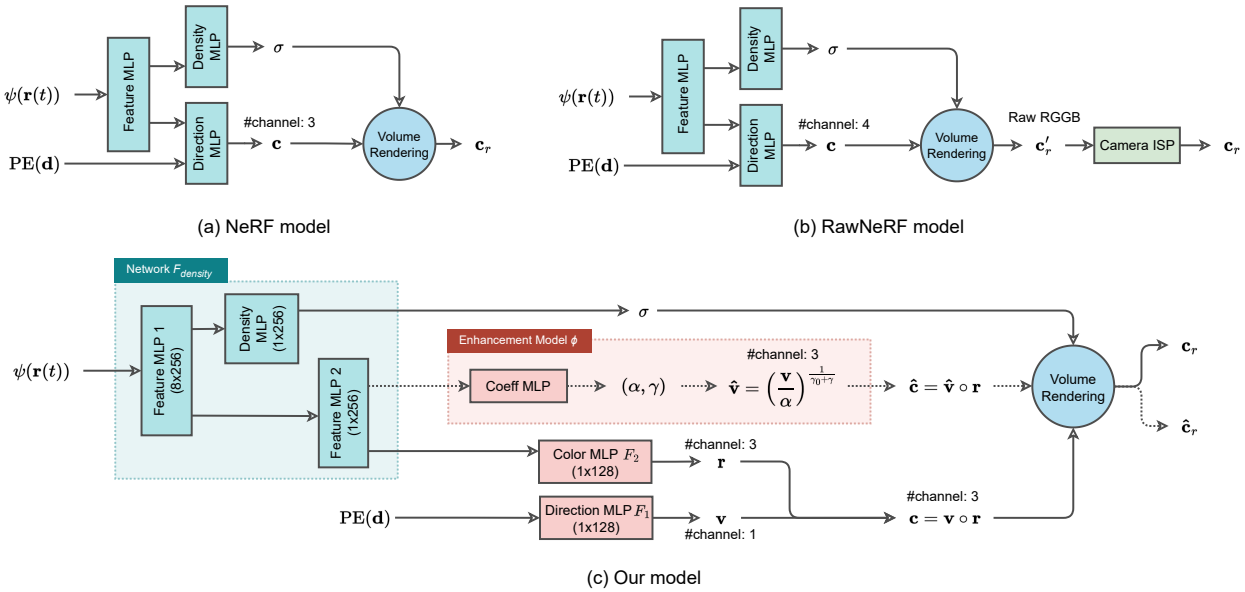
4.1. Neural Radiance Field Decomposition

As shown in Fig. 2, when one 3D point $\mathbf{x}$ in a static scene is projected to two pixels ($\tilde{\mathbf{c}}_{ri}$ and $\tilde{\mathbf{c}}_{rj}$) of two views, the colors of two pixels may appear differently, as the object surface may not be isotropic and the lighting is not uniform. However, the colors of these two pixels are still in the same range of the color spectrum. This suggests that the color of one 3D point $\mathbf{x}$ can be decomposed into a view-independent basis component and a view-dependent component. The view-independent basis component represents the intrinsic color, which determines the spectrum range of the color of $\mathbf{x}$. The view-dependent component accounts for factors that may cause color differences across views (in most situations lighting is the dominant factor, which varies depending on the position and color of the light sources and the orientation of the surface at $\mathbf{x}$).

Inspired by this, we propose to decompose the color $\mathbf{c}$ into the product of view-dependent component $\mathbf{v}$ that captures the lighting-related component and its reciprocal component $\mathbf{r}$ that represents the color basis. We leverage NeRF to constrain $\mathbf{v}$ to be view-dependent and further formulate it to be a single channel representation that focuses on the manipulation of lighting intensity.

Consider the rendering of a pixel $\mathbf{c}_r$ of image $\mathbf{I}$ in Eq. (2). Since an arbitrary image pixel $\mathbf{c}_r$ is the weighted accumulation of the view-dependent color of all $\{\mathbf{c}_i\}_{i=1}^n$ along the ray, we decompose each $\mathbf{c}$ along the ray into $\mathbf{v}$ and $\mathbf{r}$, and learn to enhance the color as:

$$\begin{cases} \mathbf{v} = F_1(\tau, \mathbf{d}; \Theta_{F_1}) \quad \text{and} \quad \mathbf{r} = F_2(\tau; \Theta_{F_2}), \\ \mathbf{c} = \mathbf{v} \circ \mathbf{r} \quad \text{and} \quad \hat{\mathbf{c}} = \phi(\mathbf{v}) \circ \mathbf{r}, \end{cases} \tag{3}$$



(a) NeRF model (b) RawNeRF model (c) Our model

Figure 3. The illustration of the NeRF [22] model (a), RawNeRF [21] model (b), and our proposed model (c). The data flow of our unsupervised enhancement is shown inside the dashed line. Our model jointly learns the novel view images and enhances the output of all samples along the ray. Each final enhanced pixel is rendered using the volume rendering equation as shown in Eq. (2).

where $\hat{\mathbf{c}}$ is the enhanced color, ϕ is an enhancement function parameterized by a neural network, and $\circ$ denotes the pixel-wise multiplication. F_1 and F_2 are two MLPs. Thus, the enhanced image $\mathbf{I}_e$ can be obtained as:

$$\mathbf{I}_e = \{\hat{\mathbf{c}}_r\}, \text{where } \hat{\mathbf{c}}_r = \sum_i w_i \phi(\mathbf{v}_i) \circ \mathbf{r}_i. \tag{4}$$

Such a method enables the model to learn a reasonable decomposition, which has a simple form but with strong constraints when the unenhanced colors $\mathbf{c}$ are supervised across views. We further demonstrate the effectiveness of the decomposition design in Sec. 5.2.

Differences to the Image-based Decomposition. Image-based low-light enhancement methods [28, 9, 31, 38] typically leverage the Retinex theory to decompose an image $\mathbf{I}$ into the illumination map $\mathbf{L}$ and reflectance map $\mathbf{R}$ as:

$$\mathbf{I} = \mathbf{L} \circ \mathbf{R}, \tag{5}$$

where $\mathbf{R}$ is invariant to the lighting condition, affected by the material and intrinsic color of objects in an image, and $\mathbf{L}$ is the response of the illumination. Their decomposition is typically guided with the normal-light ground truth images during training. The enhanced image is obtained by:

$$\mathbf{I}_e = \phi(\mathbf{L}) \circ \mathbf{R}, \tag{6}$$

where $\mathbf{I}_e$ is the enhanced image, and ϕ is the enhancement function (*e.g.*, the tone-mapping curve or a deep CNN), which is also supervised by GT.

In contrast, our method is unsupervised without ground truth for training. It works in the 3D neural radiance field with geometry information, and leverages reasonable prior (Fig. 2) to constrain the decomposition process. We compare the decomposition results of 2D-based method and ours in Fig. 6.

4.2. Unsupervised Enhancement

In addition to the unsupervised decomposition, we propose an unsupervised enhancement method to enhance light up NeRF model.

4.2.1 Denoising

Let $\mathbf{x}$ be a spatial point with a large density (*i.e.*, the color of $\mathbf{x}$ is dominant in the pixels) in a scene. It has multiple projections $C_{\mathbf{x}} = \{\tilde{\mathbf{c}}_r\}$ in the training images. We have $\tilde{\mathbf{c}}_r = \bar{\mathbf{c}}_r + \mathbf{n}$, where $\bar{\mathbf{c}}_r$ is the actual color and $\mathbf{n}$ is a small permutation noise sampled from an unknown distribution. During the training, the predicted color at $\mathbf{x}$, *i.e.*, $\mathbf{c}_{\mathbf{x}}$, is supervised by all pixels in $C_{\mathbf{x}}$ and the gradients are propagated from different rays.

As the loss function of different rays is an unweighted average, the model tends to learn the smallest average deviation from the observations in $C_{\mathbf{x}}$, and the learned $\mathbf{c}_r$ would converge to the expectation of $\tilde{\mathbf{c}}_r$, *i.e.*,

$$\mathbf{c}_r \approx \mathbb{E}\{\tilde{\mathbf{c}}_r\} = \bar{\mathbf{c}}_r + \mathbb{E}\{\mathbf{n}\}. \tag{7}$$

In RAW images, we could empirically assume the noise in each training image is zero-mean [21], *i.e.*, $\mathbb{E}\{\mathbf{n}\} = 0$. However, the non-linear processes applied to RAW images change the distribution of the raw noise, such that $\mathbf{c}_r$ is converged to a biased value $\bar{\mathbf{c}}_r + \mathbb{E}\{\mathbf{n}\}$. Accordingly, the predicted colors along the ray $\mathbf{c}$ would converge to $\bar{\mathbf{c}}+\mathbf{b}$, where

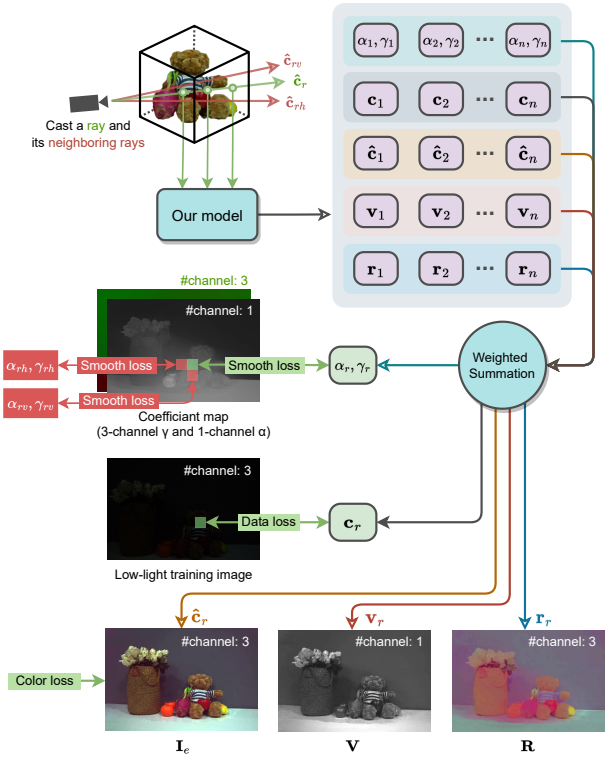


Figure 4. Illustration of our training pipeline and the proposed loss functions. The pixels are denoted as small blocks in green and red.

$\bar{\mathbf{c}}$ is the ideal predicted color, and $\mathbf{b}$ is the bias introduced by the noise.

This indicates that the multi-view optimization of the implicit neural radiance field can still smooth the image and reduce the noise in our problem. However, applying this denoising scheme is not sufficient, as the converged pixel values may be biased, leading to color distortions. We introduce our color correction and enhancement method next.

4.2.2 The Enhancement of v

We use Eq. 4 to enhance the $\mathbf{v}$ along the ray for each spatial coordinate and view direction, *i.e.*, $\hat{\mathbf{v}} = \phi(\mathbf{v})$. We propose to enhance $\mathbf{v}$ using a dynamic gamma correction under the constraint of the rendered RGB value $\hat{\mathbf{c}}_r$, as:

$$\hat{\mathbf{v}} = \phi(\mathbf{v}) = \left(\frac{\mathbf{v}}{\alpha}\right)^{\frac{1}{\gamma_0+\gamma}}, \tag{8}$$

where α is a scalar and γ is a 3D vector. Both the two coefficients are the output of the enhancement network ϕ . γ_0 is a fixed value to initialize the non-linear transform. α is defined to be a scalar to adjust the lighting gain globally, and γ is defined as a three-dimensional vector for color distortion correction by applying a small permutation to $\mathbf{v}$ in three color channels, under the constraint of the prior loss functions.

By applying Eq. 8, $\mathbf{v}$ along the ray is enhanced while $\mathbf{r}$ is not changed. Hence, our model can adjust the lighting and the color of the scene while preserving its geometry information. Although our model allows more complicated transformation functions to be applied, we find through experiments that Eq. 8 works well with a good trade-off between performance and computational cost.

4.3. Optimization Strategy

We train our model in an end-to-end manner, as shown in Fig. 4. While iteratively optimizing our model across the rays of the training dataset, three kinds of supervision signals are provided: gray-world prior-based colorimetric supervision and smooth prior-based supervision are used to optimize the enhancement network, and data supervision is used to optimize the radiance field.

Gray-world Prior-based Colormetric Supervision. To correct the bias mentioned in Sec. 4.2.1, we formulate a simple but effective gray-world prior-based loss L_c to constrain the learning of the enhancement network ϕ to produce realistic images, as:

$$L_c = \mathbb{E}[(\hat{\mathbf{c}}_r - e)^2] + \lambda_1 \mathbb{E}\left[\frac{\text{var}_c(\hat{\mathbf{c}}_r)}{\beta_1 + \text{var}_c(\mathbf{r}_r)}\right] + \lambda_2 ||\gamma||_2, \tag{9}$$

where $e, \beta_1, \lambda_1, \lambda_2$ are hyper-parameters and var_c denotes the channel-wise variance. The first term of Eq. 9 is to improve the brightness of the pixels (where $e = 0.55$). The second term is to correct colors based on the gray world prior, which pushes the distorted colors to the natural distribution by reducing the variance across three channels. To prevent the rendered pixels from converging to gray, we further add a dynamic weight based on the color of the weighted color basis $\mathbf{r}$ along the rays to relax the constraint for highly saturated colors. The third term is the regularization term to prevent overfitting.

Smoothness Prior-based Supervision. To preserve the color and structure of the scene in the enhanced radiance field and constrain the learning of the coefficients (α and γ), we expect the integrated coefficients to produce locally smoothed maps. Hence, we constrain the gradient of the weighted sum of these two coefficients with respect to the integrated $\mathbf{v}_r$ in the image space, as:

$$L_s = \underbrace{\mathbb{E}\left[\left(\frac{\partial \alpha_r}{\partial \mathbf{v}_r}\right)^2\right]}_{L_{sa}} + \underbrace{\mathbb{E}\left[\left(\frac{\partial \gamma_r}{\partial \mathbf{v}_r}\right)^2\right]}_{L_{sg}}. \tag{10}$$

Since it is difficult to obtain the desired gradient information directly from Eq. 10 due to the randomly sampled rays in training, we formulate a discrete approximation L_{sa}

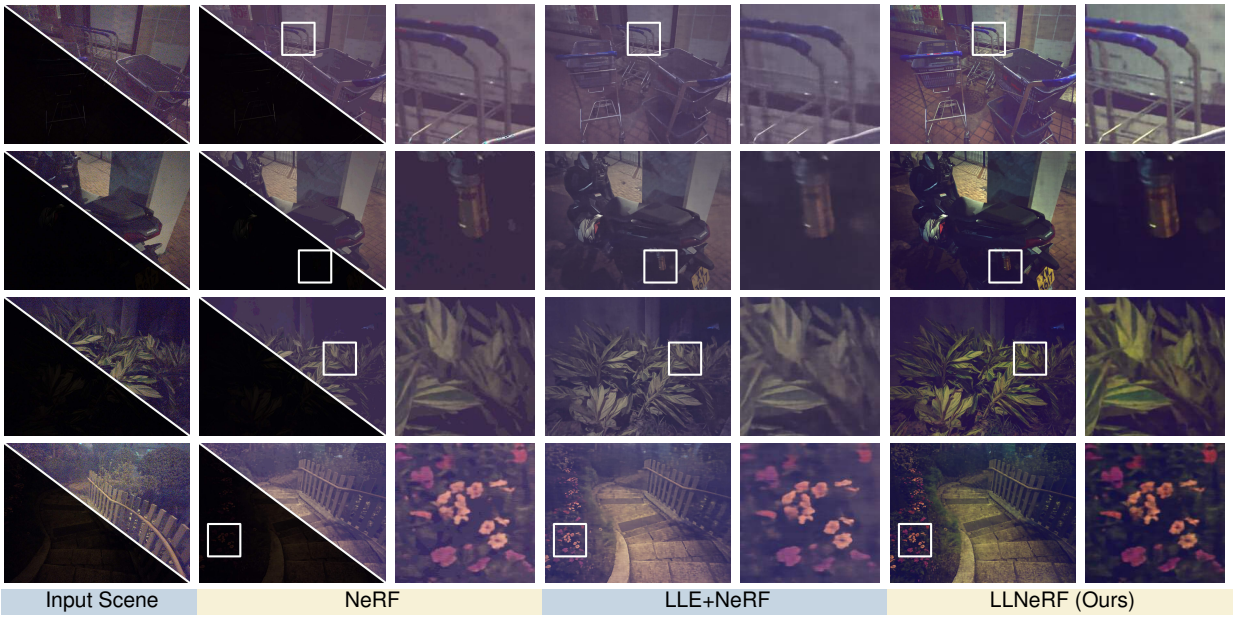


Figure 5. Visual comparison of novel view synthesis results of our model, NeRF, and the baseline model (LLE + NeRF). Note that the input scene image and the NeRF result are brightened for a better view. Our results have the best quality, with realistic color and fine details.

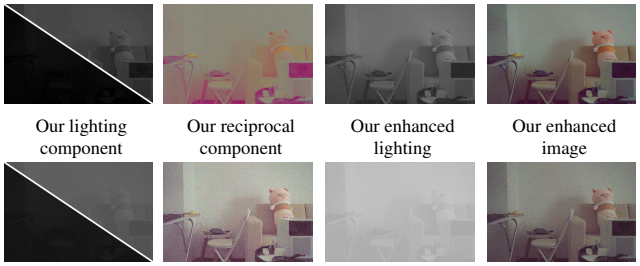


Figure 6. Visualization comparison on the decomposition of our model and the 2D-based method (URetinexNet [31]). Dark images are brightened for a better view.

of Eq. 10 as:

$$L_{sa} = \frac{1}{2}\left[\frac{(\alpha_r - \alpha_{rh})^2}{(\mathbf{v}_r - \mathbf{v}_{rh})^2 + \epsilon_1} + \frac{(\alpha_r - \alpha_{rv})^2}{(\mathbf{v}_r - \mathbf{v}_{rv})^2 + \epsilon_1}\right], \tag{11}$$

where $\alpha_{rh}, \alpha_{rv}, \mathbf{v}_{rh}, \mathbf{v}_{rv}$ are the integrated α and $\mathbf{v}$ of neighboring rays in the horizontal and vertical directions in the image space. To leverage the smoothness constraint, we sample rays with their neighboring rays in each optimization step, as shown in Fig. 4. L_{sg} is obtained in a similar way to L_{sa}.

Data Supervision. To learn the scene geometry, we apply the data loss in [21], which is the linearization of $\mathbb{E}\left[\eta(\tilde{\mathbf{c}}_r) - \eta(\mathbf{c}_r)\right]$, where $\eta(y) = \log(y + \epsilon_2)$. Since the majority of pixels in our training images have low intensity, the tone mapping function η is used to amplify the errors in the dark regions to facilitate the learning process.

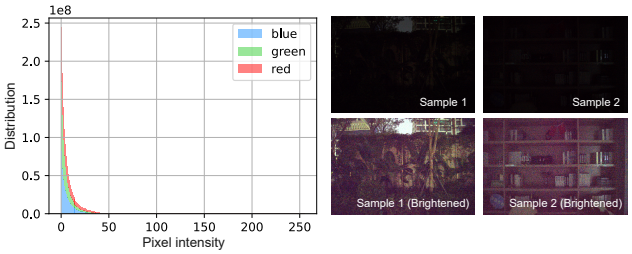


Figure 7. Intensity distribution and sample images of our dataset. We collect low-light images from both indoor and outdoor scenes. These images typically have low pixel intensity, obvious color distortion, and heavy noise.

5. Experiments

5.1. Our Dataset

We collect a real-world dataset as a benchmark for model learning and evaluation. To obtain real low-illumination images with real noise distributions, we take photos at nighttime outdoor scenes or low-light indoor scenes containing diverse objects. Since the ISP operations are device-dependent and the noise distributions across devices are also different, we collect our data using a mobile phone camera and a DSLR camera to enrich the diversity of our dataset. We show some samples and statistics of our dataset in Fig. 7. As illustrated, the average brightness of our dataset is extremely low (most pixels' intensities are below 50 out of 255). In addition, the noise and color distortion in these images are of a very high level, making our task extremely challenging.

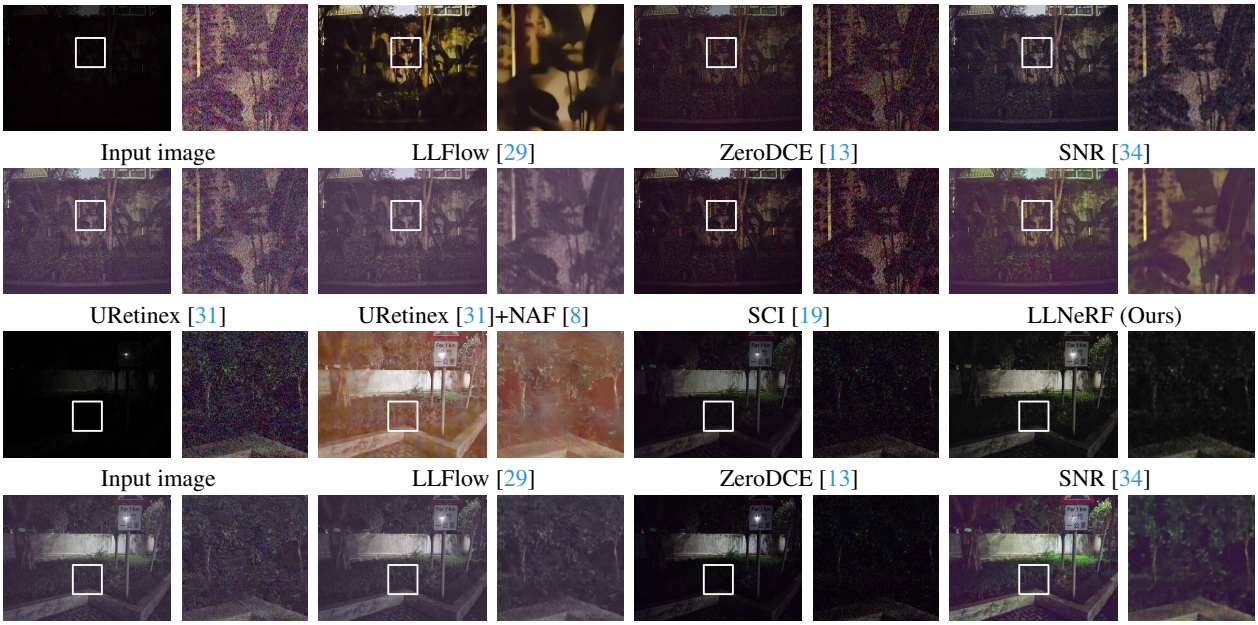


Figure 8. The visual comparison of the results of our model and the existing low-light enhancement methods. Our results have the best quality, with realistic color and fine details.

5.2. Results

We evaluate our model in three aspects. First, we evaluate the neural radiance field decomposition of our model by comparing to the Retinex-based state-of-the-art method URetinexNet [31]. Second, we evaluate the novel view synthesis performance of our model by comparing it to the baseline model (LLE + NeRF). Note that RawNeRF degrades to NeRF when RawNeRF is applied to handle sRGB images. Third, we evaluate the low-light enhancement performance by comparing our model to existing state-of-the-art LLE methods.

Visualization of V and R. We render $\mathbf{v}$ and $\mathbf{r}$ via volume rendering to obtain $\mathbf{V}$ and $\mathbf{R}$ for visualization, as shown in Fig. 4. Fig. 6 compares our decomposition to that of URetinex [31]. We can see that the reflectance map of URetinex tends to preserve all photometric information while its illumination map tends to be over-smoothed, as it is agnostic to the physical imaging process and 3D geometry information. In contrast, our model produces a more reasonable lighting-related component, and the view-independent color basis component has few shadows and lighting information. This demonstrates the effectiveness of our decomposition design in Sec. 4.1.

Novel View Synthesis. For a fair comparison, we train our model, NeRF, and the baseline model (LLE + NeRF) using the same images and compare the novel view results, as shown in Fig. 5. We choose URetinexNet as the LLE model in the baseline as it tends to produce better enhancement results compared to other enhancement methods. We can see that the results of NeRF are still low-light as there is no enhancement process inside it. Although the results of the baseline model are brightened, the image appears unrealistic as the distorted color is not corrected. In contrast, our model generates better details and natural colors.

Low-Light Enhancement. We further compare the results of our model with state-of-the-art low-light enhancement models. The comparison is shown in Fig. 8. It shows that some methods (*i.e.*, URetinexNet, SCI, ZeroDCE, SNR) cannot handle the noise. While LLFlow brightens the input and removes the noise, the visual quality is still low. We also combine the URetinexNet and a denoising model (NAFNet [8] trained on SIDD [1] dataset) for comparison. While this strategy can produce images with good details, the color is still distorted. In contrast, our model can enhance these images with better cleaner details and more natural colors. Refer to the videos in the Supplemental for more comparisons.

User Study. Due to the absence of ground truths for our low-light dataset in real-world scenarios, we employ a user study to assess the visual quality of the results of different methods. We invite 80 participants to conduct a user study to evaluate the perceptual quality of our results against those of existing approaches. Specifically, we randomly chose 10 images from the test set for comparison with LLE methods and compare the enhanced results using an AB test. For each test image, our produced result is "A" whereas the result from one of the baselines is "B". Each participant would

LLE Method	LLFlow [29]	SNR [34]	SCI [19]	URetinex [31]	ZeroDCE [13]	Ours
PSNR/SSIM	16.46/0.702	17.04/0.575	12.67/0.122	19.18/0.289	13.38/0.110	**20.50/0.758**
NVS Method	LLFlow+NeRF	SNR+NeRF	SCI+NeRF	URetinex+NeRF	ZzeroDCE+NeRF	Ours
PSNR/SSIM	16.44/0.702	17.02/0.687	13.08/0.505	19.93/0.746	14.17/0.612	**20.50/0.758**

Table 1. The quantitative comparison results between ours and existing methods on test scenes with paired normal-light images. We compare novel view synthesis results (top row) and low-light enhancement results (bottom row). The best results are marked in **bold**.

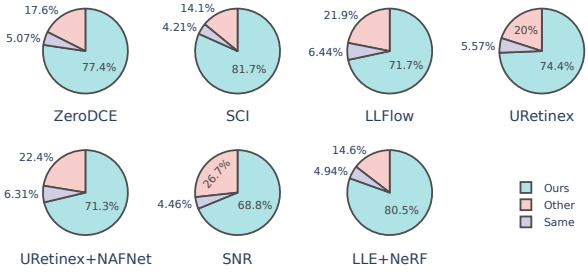


Figure 9. "Ours" is the ratio of test cases, in which the participant selected our results as better; "Other" is the percentage that another method was selected to be better; and "Same" is the percentage that the user has no preference.

simultaneously see A and B (we avoid the bias by randomizing the left-right presentation order when displaying A and B in each AB-test task) and select one from: "A is better", "B is better", and "I cannot decide". We ask the participants to make decisions based on natural brightness, rich details, distinct contrasts, vivid colors, and noise removal effects.

The comparison between ours and the baseline model, *i.e.*, LLE + NeRF, is conducted similarly, where "A" and "B" refers to the rendered videos. For each participant, the number of tasks is 7 methods $\times$ 10 questions, 70 in total. It takes on average around 30 minutes for each participant to complete the user study.

Fig. 9 summarizes the user study results, which shows that our results are more preferred by the participants than all other competing methods.

Quantitative Evaluation. We additionally evaluate three scenes quantitatively with normal-light images of long exposures. As shown in Tab. 1, our method performs better than existing methods on both PSNR and SSIM. It also shows that NeRF helps enhance image structures (better SSIM), due to the implicit 3D information of its radiance field optimization process.

Ablation Study. To investigate the effectiveness of our training strategy, we conduct the ablation study of our loss functions. By relaxing the constraints of loss functions, we compare the visual results produced by different settings of loss functions. Fig. 10 shows that removing terms from the proposed loss function generally results in the degradation of the results produced by our model.

Scene Editing. Our model allows different manipulations of the scene's illumination while producing realistic novel view images, *e.g.*, the scene's color temperature can be edited, as shown in Fig. 11. As a comparison, the existing deep-learning-based color temperature editing method [2] produces relatively unnatural editing results with artifacts in the highlight regions.

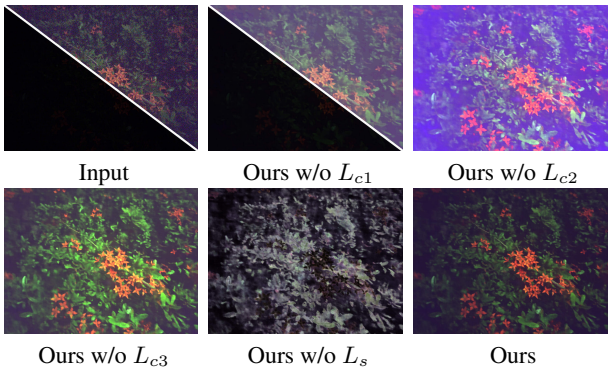


Figure 10. Ablation study results. L_{c1}, L_{c2}, L_{c3} are three items in L_c respectively. The quality of results is degraded as we remove any item. The dark images are brightened for a better view.

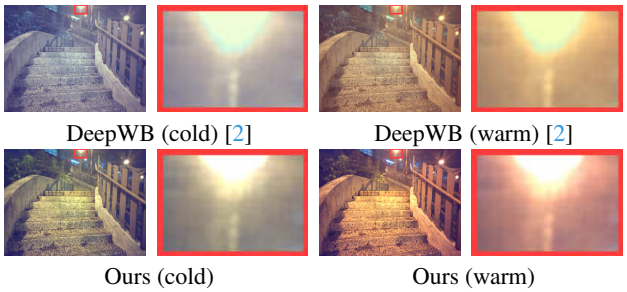


Figure 11. A possible application of our model besides the low-light enhancement. By modifying $\mathbf{v}$ along the rays, our model is able to produce realistic scenes with varying color temperatures.

6. Conclusion

In this paper, we propose a novel method to train a NeRF model from low-light sRGB images to produce novel view images of high visibility, vivid colors, and details. Based on the observation of the imaging process, our model decomposes the neural radiance field to the lighting-related view-dependent component and view-independent color basis components in an unsupervised manner. Our model enhances the lighting without reference images under the supervision of prior-based loss functions. We conduct extensive experiments to analyze the properties of our method and demonstrate its effectiveness against existing methods.

7. Acknowledgements

The work described in this paper was partially supported by a GRF grant from the Research Grants Council of Hong Kong (Project No. CityU 11205620).

References

[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S. Brown. A high-quality denoising dataset for smartphone cameras. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 7

[2] Mahmoud Afifi and Michael S Brown. Deep white-balance editing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 8

[3] Mahmoud Afifi, Konstantinos G Derpanis, Bjorn Ommer, and Michael S Brown. Learning multi-scale photo exposure correction. In *CVPR*, 2021. 2

[4] Benjamin Attal, Eliot Laidlaw, Aaron Gokaslan, Changil Kim, Christian Richardt, James Tompkin, and Matthew O'Toole. Törf: Time-of-flight radiance fields for dynamic scene view synthesis. In *NeurIPS*, 2021. 2

[5] 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 *ICCV*, 2021. 3

[6] Bolun Cai, Xianming Xu, Kailing Guo, Kui Jia, Bin Hu, and Dacheng Tao. A joint intrinsic-extrinsic prior model for retinex. In *ICCV*, 2017. 2

[7] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE TIP*, 2018. 2

[8] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. *arXiv preprint arXiv:2204.04676*, 2022. 7

[9] Wei Chen, Wang Wenjing, Yang Wenhan, and Liu Jiaying. Deep retinex decomposition for low-light enhancement. In *BMVC*, 2018. 4

[10] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *CVPR*, 2018. 2

[11] Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. Depth-supervised nerf: Fewer views and faster training for free. In *CVPR*, 2022. 2

[12] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *ACM TOG*, 2017. 2

[13] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *CVPR*, 2020. 7, 8

[14] Rynson Lau Haoyuan Wang, Ke Xu. Local color distributions prior for image enhancement. In *ECCV*, 2022. 2

[15] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. DSLR-quality photos on mobile devices with deep convolutional networks. In *ICCV*, 2017. 2

[16] Yoonwoo Jeong, Seokjun Ahn, Christopher Choy, Anima Anandkumar, Minsu Cho, and Jaesik Park. Self-calibrating neural radiance fields. In *ICCV*, 2021. 2

[17] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE TIP*, 2021. 2

[18] Li Ma, Xiaoyu Li, Jing Liao, Qi Zhang, Xuan Wang, Jue Wang, and Pedro V. Sander. Deblur-nerf: Neural radiance fields from blurry images. In *CVPR*, 2022. 1, 2

[19] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5637–5646, 2022. 7, 8

[20] 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 *CVPR*, 2021. 2

[21] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P. Srinivasan, and Jonathan T. Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In *CVPR*, 2022. 1, 2, 3, 4, 6

[22] 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 *ECCV*, 2020. 1, 2, 3, 4

[23] Sean Moran, Pierre Marza, Steven McDonagh, Sarah Parisot, and Gregory G. Slabaugh. Deeplpf: Deep local parametric filters for image enhancement. In *CVPR*, 2020. 2

[24] Jongchan Park, Joon-Young Lee, Donggeun Yoo, and In So Kweon. Distort-and-recover: Color enhancement using deep reinforcement learning. In *CVPR*, 2018. 2

[25] Wenqi Ren, Sifei Liu, Lin Ma, Qianqian Xu, Xiangyu Xu, Xiaochun Cao, Junping Du, and Minghsuan Yang. Low-light image enhancement via a deep hybrid network. *IEEE TIP*, 2019. 2

[26] Liu Risheng, Ma Long, Zhang Jiaao, Fan Xin, and Luo Zhongxuan. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *CVPR*, 2021. 2

[27] Aashish Sharma and Robby T Tan. Nighttime visibility enhancement by increasing the dynamic range and suppression of light effects. In *CVPR*, 2021. 2

[28] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *CVPR*, 2019. 2, 4

[29] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex C Kot. Low-light image enhancement with normalizing flow. *arXiv preprint arXiv:2109.05923*, 2021. 1, 7, 8

[30] Yi Wei, Shaohui Liu, Yongming Rao, Wang Zhao, Jiwen Lu, and Jie Zhou. Nerfingmvs: Guided optimization of neural radiance fields for indoor multi-view stereo. In *ICCV*, 2021. 2

[31] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *CVPR*, 2022. 1, 2, 4, 6, 7, 8

[32] Huang Xin, Zhang Qi, Feng Ying, Li Hongdong, Wang Xuan, and Wang Qing. Hdr-nerf: High dynamic range neural radiance fields. In *CVPR*, 2022. 1, 2

[33] Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In *CVPR*, 2020. 2

[34] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *CVPR*, 2022. 1, 2, 7, 8

[35] Xin Yang, Ke Xu, Yibing Song, Qiang Zhang, Xiaopeng Wei, and Rynson Lau. Image correction via deep reciprocating HDR transformation. In *CVPR*, 2018. 2

[36] Runsheng Yu, Wenyu Liu, Yasen Zhang, Zhi Qu, Deli Zhao, and Bo Zhang. Deepexposure: Learning to expose photos with asynchronously reinforced adversarial learning. In *NeurIPS*, 2018. 2

[37] Qing Zhang, Ganzhao Yuan, Chunxia Xiao, Lei Zhu, and Wei-Shi Zheng. High-quality exposure correction of underexposed photos. In *ACM MM*, 2018. 2

[38] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *ACM MM*, 2019. 2, 4

[39] Shuaifeng Zhi, Tristan Laidlow, Stefan Leutenegger, and Andrew J Davison. In-place scene labelling and understanding with implicit scene representation. In *ICCV*, 2021. 2