Enhancing Neural Radiance Fields with Adaptive Multi-Exposure Fusion: A Bilevel Optimization Approach for Novel View Synthesis

Yang Zou[1*], **Xingyuan Li**[2*], **Zhiying Jiang**[2], **Jinyuan Liu**[2†]

[1]School of Computer Science, The University of Sydney
[2]School of Software Technology, Dalian University of Technology
yzou9073@uni.sydney.edu.au, xingyuan_lxy@163.com, zyjiang0630@gmail.com, atlantis918@hotmail.com

Abstract

Neural Radiance Fields (NeRF) have made significant strides in the modeling and rendering of 3D scenes. However, due to the complexity of luminance information, existing NeRF methods often struggle to produce satisfactory renderings when dealing with high and low exposure images. To address this issue, we propose an innovative approach capable of effectively modeling and rendering images under multiple exposure conditions. Our method adaptively learns the characteristics of images under different exposure conditions through an unsupervised evaluator-simulator structure for HDR (High Dynamic Range) fusion. This approach enhances NeRF's comprehension and handling of light variations, leading to the generation of images with appropriate brightness. Simultaneously, we present a bilevel optimization method tailored for novel view synthesis, aiming to harmonize the luminance information of input images while preserving their structural and content consistency. This approach facilitates the concurrent optimization of multi-exposure correction and novel view synthesis, in an unsupervised manner. Through comprehensive experiments conducted on the LOM and LOL datasets, our approach surpasses existing methods, markedly enhancing the task of novel view synthesis for multi-exposure environments and attaining state-of-the-art results. The source code can be found at https://github.com/Archer-204/AME-NeRF.

Introduction

Deep learning has revolutionized numerous fields, with computer vision being one of the most prominent beneficiaries. Techniques such as Neural Radiance Fields (NeRF)(Mildenhall et al. 2020) have shown remarkable promise in synthesizing novel views of complex scenes from sparse 2D observations. However, similar to most computer vision algorithms such as semantic segmentation, object detection, and etc., the performance of these techniques is often compromised under varying exposure conditions(Zhang and Ma 2023; Xu, Haochen, and Ma 2023; Zhang et al. 2023), a common occurrence in real-world image capture.

The challenge lies in the fact that images captured under different exposure conditions can exhibit significant variations in color, brightness, and global detail(Wu et al. 2022; Wu, Chen, and Ma 2022; Han et al. 2022). This variation poses a significant problem for NeRF, which heavily rely on consistent and high-quality input data to generate accurate and realistic outputs. The limitations of the NeRF stem from its viewer-centric approach, which calculates the light emission from a specific location to the viewer, neglecting the interplay between illumination and the scene itself. This emitted light is actually a result of environmental light reflecting off the scene. This reflected light then undergoes further refraction and absorption, leading to attenuation as it travels through the environment once more(Srinivasan et al. 2021). As a result, the NeRF algorithm perceives a dimly lit scene as a result of inadequate radiation from the 3D particles that depict objects within the scene.

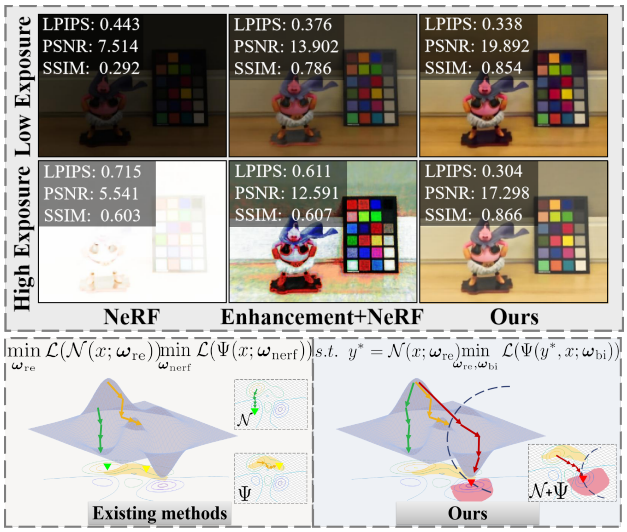

Figure 1: Visual comparison of our method with SOTA enhancement technique. Existing approaches primarily focus on either image enhancement(indicated by the green line) or novel view synthesis refinement(denoted by the yellow line). In contrast, our method concurrently optimizes both the NeRF network(highlighted by the blue line) and image enhancement(represented by the red line), leading to a significantly reduced loss.

*These authors contributed equally.
†Corresponding author.

Existing solutions to this problem have primarily focused on pre-processing techniques such as image enhancement(Mildenhall et al. 2022; Liu et al. 2023b; Ma et al. 2023; Liu et al. 2022b), brightness adjustment(Jiang et al. 2021), exposure correction(Nguyen et al. 2023) and etc. Although these techniques have demonstrated some effectiveness, enhancing dark or overexposure images using 2D enhancement methods does not ensure precise NeRF estimation. This is because the independent and inconsistent enhancement of 2D images across multiple views could disrupt the consistency of 3D geometry. Another approach, Aleth-NeRF(Cui et al. 2023), seeks to restructure the volume rendering pipeline. It learns from dark images to comprehend the volumetric object representation and concealing field under priors in an unsupervised manner. However, Aleth-NeRF is only applicable to dark environments and not suitable for overexposure conditions due to the design of the concealing field (Cui et al. 2023).

In this paper, we propose a proactive approach to this problem. We present a novel method for NeRF that is designed to operate under multiple exposures in an unsupervised manner, effectively mitigating the impact of exposure variation at the source. Our method employs a evaluator and simulator structure for High Dynamic Range (HDR)(Le et al. 2023) fusion, enabling the NeRF network to perceive and adapt to a wider range of exposure conditions. Furthermore, we introduce a bilevel optimization method(Li, Gu, and Huang 2022) in our framework which simultaneously optimizes multi-exposure correction and the synthesis of new perspectives in NeRF, leading to superior performance of our method compared to traditional NeRF models. Experiments compared with the state-of-the-art methods show the superior performance of our proposed methods. Ablation studies also prove our hypothesis.

Our main contributions are as follows:

- We propose a unsupervised enhancement framework, train on sRGB images under multiple exposure conditions, effectively mitigating the impact of exposure variation. Our method not only enhances image quality but also ensures consistent visual fidelity.
- We introduce a bilevel optimization model that builds a relationship between multi-exposure correction and the synthesis of new perspectives. Consequently, the two tasks mutually reinforce each other, outperforming traditional methods.
- Through extensive experiments, we demonstrate that our approach significantly outperforms state-of-the-art methods on the LOM(Cui et al. 2023) and LOL(Wei et al. 2018) datasets, achieving outstanding results by substantial margins.

Related Work

Low-Light Image Enhancement. Low-Light Image Enhancement (LLIE) is aiming to improve the visibility of images captured under poorly lit conditions. Traditional LLIE methods, such as histogram equalization(Patel, Maravi, and Sharma 2013) and gamma correction(Cao and Bermak 2011), are simple and computationally efficient.

With the advent of deep learning, learning-based LLIE methods have emerged(Wang et al. 2022b; Huang et al. 2022). A series of supervised methods, such as the LLNet(Lore, Akintayo, and Sarkar 2017), MBLLEN(Lv et al. 2018), LPNet(Li et al. 2021), and etc. leverage convolutional neural networks to model the complex mapping from low-light images to normal-light images(Yang et al. 2021a) and have shown promising results(Yang et al. 2021b).

Overexposure Correction. High Dynamic Range imaging(Mertens, Kautz, and Van Reeth 2007) and exposure fusion(Liu et al. 2023c,a, 2022a; Li et al. 2023) combines multiple images taken at different exposure levels and can effectively recover details in both underexposed and overexposed regions. However, this methods is frequently constrained because of the dearth of the overall and local complex gray distribution within an image. Recently, learning-based methods like SID(Chen et al. 2018) enhanced images from raw data.

View Synthesis with NeRF Neural Radiance Fields (NeRF)(Mildenhall et al. 2022; Hong et al. 2022; Wang et al. 2022a) offers a unique approach to generate new views of intricate 3D scenes from limited 2D images. Utilizing a multilayer perceptron (MLP), NeRF directly translates spatial location $\mathbf{x}$ and view direction $\mathbf{d}$ into RGB color $\mathbf{c}$ and density σ. The term $\sigma(\mathbf{x})$ represents the minuscule likelihood of a ray concluding at a particle at $\mathbf{x}$.

In the process of rendering an image with a neural radiance field, a camera ray $\mathbf{r}(t) = \mathbf{o} + t \cdot \mathbf{d}$ (where $\mathbf{r} \in \mathbf{R}$) is projected from the camera's center of projection $\mathbf{o}$ in the direction $\mathbf{d}$, denoted as $\mathbf{C}(\mathbf{r})$. The expected color $\mathbf{C}(\mathbf{r})$ constrained by near and far bounds t_n and t_f is typically executed using the volume rendering function. Formally,

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})dt, \tag{1}$$

where

$$T(t) = \exp\left(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))ds\right). \tag{2}$$

$T(t)$ represents the accumulated transmittance along the ray from t_n to t. The network is fine-tuned by minimizing:

$$\mathcal{L} = \sum_{\mathbf{r}\in\mathcal{R}} \left[\left\| \hat{C}_c(\mathbf{r}) - C(\mathbf{r}) \right\|_2^2 + \left\| \hat{C}_f(\mathbf{r}) - C(\mathbf{r}) \right\|_2^2 \right] \tag{3}$$

Where $\mathcal{R}$ is the set of rays in each batch. The $C(\mathbf{r})$, $\hat{C}_c(\mathbf{r})$, and $\hat{C}_f(\mathbf{r})$ represent the ground truth, coarse volume predicted, and fine volume predicted RGB colors for the ray $\mathbf{r}$, respectively.

Methods

We initiate our discussion by outlining the architecture of the multi-exposure correction process. Subsequently, we provide a detailed explanation of our innovative bilevel optimization model. Building on this foundation, we illustrate how this model is incorporated into the Neural Radiance Fields framework.

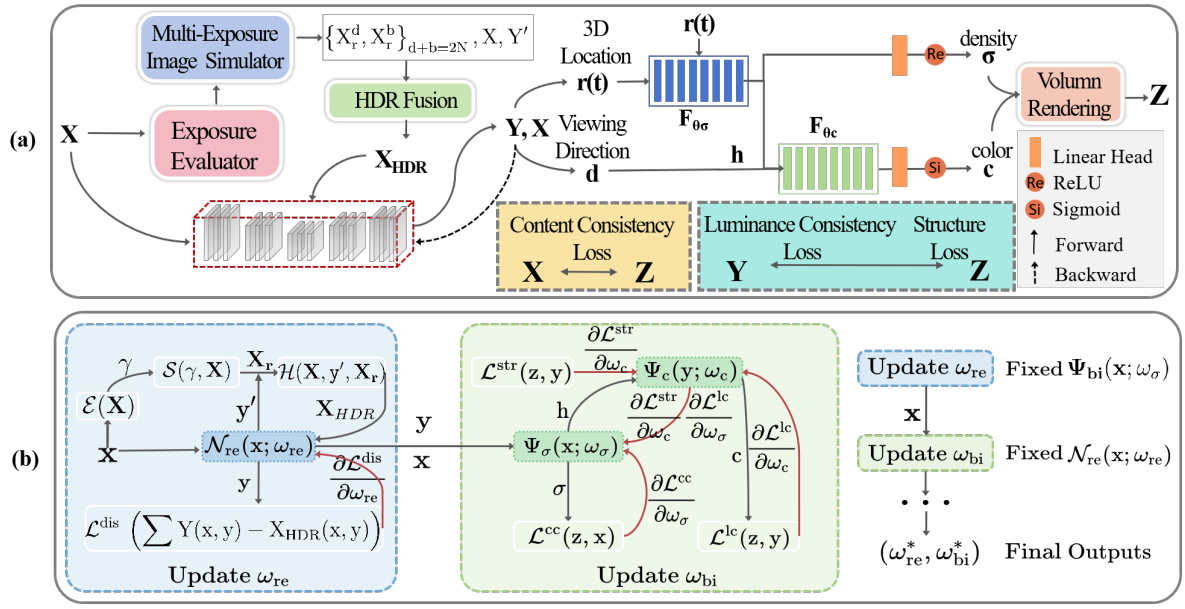

Figure 2: Overview of the architecture of our method. The general pipeline of our model is visualized in (a). The detailed optimization process is demonstrated in (b).

Dynamic Multi-Exposure Correction

The exposure evaluator discriminates the severity of the exposure deviation of the input image. The simulator then utilizes gamma mapping to adjust the exposure level. Subsequently, we employ HDR to fuse a well-exposed image.

Exposure Evaluator. The first step in our exposure evaluator is to analyze the histogram of the image. We compute the histogram $H(i)$ of the grayscale image as:

$$H(i) = \sum[\text{gray}(x, y) = i] \quad \text{for } i \in [0, 255], \tag{4}$$

where $H(i)$ represents the number of pixels at brightness level i, and $\text{gray}(x, y)$ is the pixel value at location (x, y) in the grayscale input image $\mathbf{X}$. The peak of the histogram P represents the most common brightness level, and the skewness S indicates the asymmetry of the brightness levels. They are computed as:

$$P = \text{argmax}\, H(i), \tag{5}$$

$$S = \frac{1}{N} \sum \left[\frac{(i - \mu)^3}{\sigma^3} \right], \tag{6}$$

where μ is the mean brightness, σ is the standard deviation of brightness, and i ranges over the set $\{0, 1, \ldots, 255\}$. Gamma correction is a nonlinear operation used to correct the exposure of an image. Based on the histogram's peak and skewness, we dynamically select the gamma value γ as follows:

$$\gamma = \frac{1 + e^{-(b \cdot P + c \cdot S + d)}}{a}, \tag{7}$$

where the parameters are carefully chosen to reflect the exposure characteristics of the image. Specifically, parameter a modulates the overall amplitude of the gamma correction, while b and c allow for fine-tuning the influence of the histogram's peak and skewness on the gamma value, respectively. Parameter d offers the flexibility to shift the entire gamma value selection curve either upward or downward. In our implementation, we select the weights within specific ranges:$a \sim U(2, 3)$, $b \sim U(-0.2, -0.1)$, $c \sim U(0.01, 0.1)$, and $d \sim U(0.5, 2)$.

This exposure evaluator dynamically selects the gamma value according to the exposure condition of the image, and can be adapted to different needs and scenarios by adjusting the parameters.

Dynamic Multi-Exposure Image Simulator. Based on the gamma value γ selected above and input image $\mathbf{X}$, the simulator employs gamma mapping to emulate images at varying exposure levels. The gamma mapping function M is defined as:

$$M(\mathbf{X}, \gamma) = \mathbf{X}^\gamma, \tag{8}$$

where γ is the gamma value selected by the exposure evaluator. For each input image $\mathbf{X}$, the exposure evaluator computes two gamma values, γ_1 and γ_2, based on the exposure condition detected from the histogram analysis (as described in Equation. 7). These gamma values are then used to create two images X1, X2 with different exposure levels.

Exposure Correction with HDR. The HDR Image Fusion is for synthesizing a well-exposed image from a set of differently exposed images. The input consists of the original network input image $\mathbf{X}$, denoted as $\mathbf{X}_0$, two differently exposed images generated by the multi-exposure image simulator $\mathbf{X}_1$ and $\mathbf{X}_2$(described in Equation. 9), and the output from the previous epoch $\mathbf{Y}'$, represented as $\mathbf{X}_3$.

The weights for each input image, signifying the contribution of individual images to the final HDR image, are computed based on their brightness level. The weight calculation is expressed as:

$$w_i = \exp(-\alpha \cdot (\mathbf{X}_i - \mu)^2), \tag{9}$$

where $\mathbf{X}_i$ represents the average brightness of the i-th input image, and μ is the mean brightness across all input images. The parameter α is a tunable factor that controls the distribution of the weights. In our experiment, we have set α to 0.8, allowing for a balanced contribution from images with varying exposure levels.

We then apply the fusion of all input images into a single high dynamic range image. The fusion is performed using a weighted sum of the input images:

$$\mathbf{X}_{HDR} = \frac{\sum_{i=0}^{n} w_i \cdot \mathbf{X}_i}{\sum_{i=0}^{n} w_i}, \tag{10}$$

where w_i is the weight of the i-th image. The output of the HDR generator is the well-exposed image $\mathbf{X}_{HDR}$.

Our refinement network is designed with a foundation in the U-Net architecture(Ronneberger, Fischer, and Brox 2015). To generate the exposure corrected image $\mathbf{Y}$, the predicted gamma map γ and the original input image $\mathbf{X}$ are used as inputs:

$$\mathbf{Y} = \mathbf{X}^{\gamma}. \tag{11}$$

The model is optimized by minimizing a loss function that measures the discrepancy between the network's predicted image $\mathbf{Y}$ and the HDR fusion image $\mathbf{X}_{\text{HDR}}$. The loss function is defined as:

$$\mathcal{L}^{dis} = \frac{1}{W \times H} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} \left(\mathbf{Y}(x,y) - \mathbf{X}_{\text{HDR}}(x,y)\right)^2, \tag{12}$$

where W and H represent the width and height of the images, respectively. $\mathbf{Y}(x,y)$ denotes the pixel value at location (x,y) in the network's predicted image, and $\mathbf{X}_{\text{HDR}}(x,y)$ is the corresponding pixel value in the HDR image.

NeRF with Bilevel Optimizatoin

We introduce a bilevel optimization model which is designed to flexibly incorporate implicit constraints. The bilevel optimization consists of two interconnected subproblems. The upper subproblem is a fundamental component of Neural Radiance Fields (NeRF), responsible for synthesizing an information-rich novel view image that encapsulates the essential visual details. The lower subproblem, on the other hand, focuses on adjusting the exposure of the input image. It ensures the preservation of the source image's inherent characteristics, thereby providing a refined feasibility domain for the upper subproblem's solution.

Let the input image, exposure-corrected image, and the synthesized novel view image be of size $m \times n$, denoted as $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathbb{R}^{m \times n}$, where m and n correspond to the height and width of the image, respectively. The bilevel optimization function is formulated as:

$$\begin{aligned} &\min \mathcal{F}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \\ &\text{s.t. } \mathbf{Z} \in \underset{\mathbf{Z}}{\operatorname{argmin}}(\mathcal{A}(\mathbf{Z}, \mathbf{X}), \mathcal{B}(\mathbf{Z}, \mathbf{Y})). \end{aligned} \tag{13}$$

Here, $\mathcal{F}(\cdot)$ denotes the data fidelity term, encapsulating the relationship between the synthesized novel view image $\mathbf{Z}$, the input image $\mathbf{X}$, and the exposure-corrected image $\mathbf{Y}$. The terms $\mathcal{A}(\cdot)$ and $\mathcal{B}(\cdot)$ act as feasibility constraints for upper subproblem and lower subproblem, representing prior functions that are oriented towards the intrinsic features of modalities $\mathbf{X}$ and $\mathbf{Y}$, respectively. The content consistency constraint $\mathcal{A}(\cdot)$ is formally expressed as:

$$\mathcal{A}(\mathbf{Z}, \mathbf{X}) = \sum_{\mathbf{r} \in \mathcal{R}} \left[\|\mathbf{Z}_c(\mathbf{r}) - \mathbf{X}(\mathbf{r})\|_2^2 + \|\mathbf{Z}_f(\mathbf{r}) - \mathbf{X}(\mathbf{r})\|_2^2 \right], \tag{14}$$

where $\mathcal{R}$ is the set of rays in each batch, and $\mathbf{Z}_c$ and $\mathbf{Z}_f$ correspond to the coarse and fine renderings, respectively. This constraint ensures that the synthesized image maintains content consistency with the target viewpoint, preserving both geometric and optical properties.

The feasibility constraint $\mathcal{B}(\cdot)$ is composed of a luminance consistency loss L^{lc} and a structure loss L^{str}, weighted by λ_1 and λ_2 respectively, the function can be expressed as:

$$\mathcal{B}(\mathbf{Z}, \mathbf{Y}) = \lambda_1 \cdot \mathcal{L}^{lc}(\mathbf{Z}, \mathbf{Y}) + \lambda_2 \cdot \mathcal{L}^{str}(\mathbf{Z}, \mathbf{Y}), \tag{15}$$

where

$$\mathcal{L}^{lc}(\mathbf{Z}, \mathbf{Y}) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} (L(\mathbf{Z}(i,j)) - L(\mathbf{Y}(i,j)))^2, \tag{16}$$

and

$$\mathcal{L}^{str}(\mathbf{Z}, \mathbf{Y}) = \sum_{l} \frac{1}{N_l} \|\phi_l(Z) - \phi_l(Y)\|_2^2, \tag{17}$$

where $\phi_l(\cdot)$ represents the feature map extracted from the l-th layer of a pre-trained deep neural network (in our experiment, VGG16), and N_l is the number of elements in the l-th feature map. $L(\cdot)$ represent the luminance values at pixel location in the image. Since pixels taken under extreme exposure conditions would lose some luminance information, L^{lc} is purposed to regularize the luminance of the predicted normal light images with generated novel view images. To maintain the structure information of the generated novel view images, we add a structure loss L^{str} to captures higher-level visual features, ensuring that the generated image appears natural and realistic to the human visual system.

To fully integrate the information from different exposure of images, we employs the Sequential Average Method(SAM)(Sabach and Shtern 2017). Specifically, the upper and lower subproblems are solved separately at first. Then, a balancing parameter α is introduced to harmonize these two subproblems, expressed as:

$$\mathbf{Z} = \alpha \odot \mathbf{U} + (1 - \alpha) \odot \mathbf{L}. \tag{18}$$

Where $\odot$ denotes element-wise multiplication. $\mathbf{U}$ and $\mathbf{L}$ are the solutions to the upper subproblem $\mathcal{A}(\cdot)$ and lower subproblem $\mathcal{B}(\cdot)$, respectively. α is a sequence of real numbers chosen from the interval $[0, 1]$.

Experiments

We commence by detailing our training environment, the parameters employed, and the dataset utilized for our experiments. Subsequently, we exhibit the comparative performance of our approach against methods that amalgamate

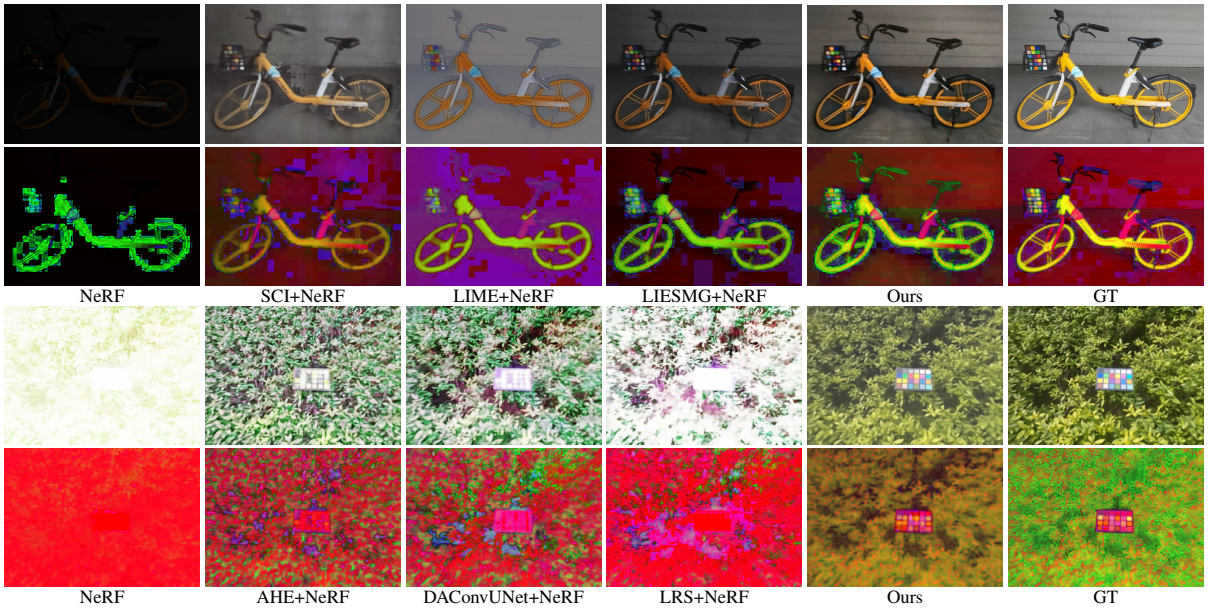

Figure 3: Example of the enhancement results on LOM(Cui et al. 2023) dataset under extreme exposure conditions. Both sRGB and HSV images are compared to show the content consistency. The results are from methods of NeRF, SCI(Ma et al. 2022), LIME(Guo, Li, and Ling 2017), LIESMG(Xu, Wang, and Lu 2023), AHE, LRS(Srinivasan and Balram 2006), DAConv(Wang et al. 2023) as well as the ground truth.

Figure 4: Comparison of the enhancement results with NeRF, LIME, RetinexNet(Wei et al. 2018), and EnGAN(Jiang et al. 2021) on the LOL dataset(Wei et al. 2018).

low-light enhancement with NeRF and those that integrate exposure adjustment with NeRF. Lastly, we present the findings of our ablation study, which effectively underscore the superior benefits of our proposed method.

Implementation Details

Our approach is evaluated on the LOM dataset(Cui et al. 2023), which encompasses five real-world scenes. Each scene comprises between 25 to 65 pairs of low-light and normal-light conditions. We specifically made a overexposure version of LOM dataset to simulate the overexposure conditions. The single image low-light enhancement experiments are done on the LOL dataset(Wei et al. 2018). The training of our network is conducted on a NVIDIA GeForce RTX 3080Ti GPU using the Adam optimizer. The model undergoes a training regimen of 100 epochs with a batch size of 1024 for 62500 iters training. The learning rate is $5e^{-4}$ and is reduced every 2500 iters with the cosine learning rate decay strategy. For detail, the coefficient of the number of coarse and fine samples are configured to 64 and 128.

Generation Quality Assessment

In this section, we showcase the results of multi-view rendering for both low-light and overexposure scenarios, utilizing the LOM and LOL datasets. Our comparison includes the original NeRF and five low-light enhancement methods: HE(Patel, Maravi, and Sharma 2013), LIME(Guo, Li, and Ling 2017), LIESMG(Xu, Wang, and Lu 2023), SCI(Ma et al. 2022), and IAT(Cui et al. 2022). Among these, HE and LIME represent traditional enhancement methods, while IAT, SCI and LIESMG are state-of-the-art network-based 2D enhancement methods that have emerged recently. We then compare with the vanilla NeRF and four exposure correction methods: DAConv based UNet(Wang et al. 2023),

Data	Method	"bike"			"buu"			"chair"			"shrub"			"sofa"		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Low Exposure	NeRF	6.36	0.072	0.633	7.51	0.292	0.443	6.04	0.147	0.603	8.01	0.028	0.716	6.27	0.209	0.557
	HE + NeRF	15.29	0.693	0.441	15.52	0.781	0.517	15.41	0.747	0.554	14.74	0.441	0.567	17.87	0.811	0.508
	IAT + NeRF	13.49	0.607	0.541	14.49	0.705	0.401	**18.79**	**0.781**	0.671	13.81	0.286	0.565	17.61	0.829	0.545
	LIESMG + NeRF	18.02	0.708	0.479	16.21	0.781	0.392	16.86	0.759	0.526	14.83	0.281	**0.517**	16.81	0.808	0.565
	LIME + NeRF	11.31	0.572	0.471	13.91	0.786	0.316	11.27	0.677	0.533	13.88	0.357	0.521	12.21	0.755	0.445
	SCI + NeRF	13.56	0.651	0.459	7.78	0.693	0.528	11.71	0.741	0.595	**17.63**	0.441	0.523	10.08	0.765	0.518
	Ours	**18.14**	**0.732**	**0.437**	**19.89**	**0.854**	**0.312**	17.05	0.751	**0.381**	15.23	**0.462**	0.518	**17.93**	**0.847**	**0.378**
High Exposure	NeRF	5.61	0.501	0.725	5.54	0.603	0.715	6.11	0.592	0.713	4.14	0.092	0.753	6.26	0.673	0.694
	DAConvUNet + NeRF	12.62	0.641	0.449	12.59	0.606	0.611	13.23	0.627	0.607	11.31	0.399	0.601	13.27	0.714	0.587
	LRS + NeRF	8.44	0.573	0.541	7.67	0.654	0.655	8.82	0.659	0.651	6.03	0.211	0.714	8.45	0.667	0.621
	CLAHE + NeRF	7.41	0.573	0.596	7.64	0.662	0.592	8.37	0.652	0.602	8.42	0.287	0.616	7.71	0.711	0.634
	AHE + NeRF	10.94	0.552	0.468	9.69	0.395	0.674	11.77	0.499	0.606	11.21	0.399	0.604	9.76	0.603	0.605
	Ours	**20.79**	**0.761**	**0.432**	**17.29**	**0.865**	**0.304**	**24.88**	**0.846**	**0.376**	**16.87**	**0.404**	**0.534**	**23.72**	**0.894**	**0.395**

Table 1: Quantitative comparison of various methods on LOM dataset under low and high exposure conditions.

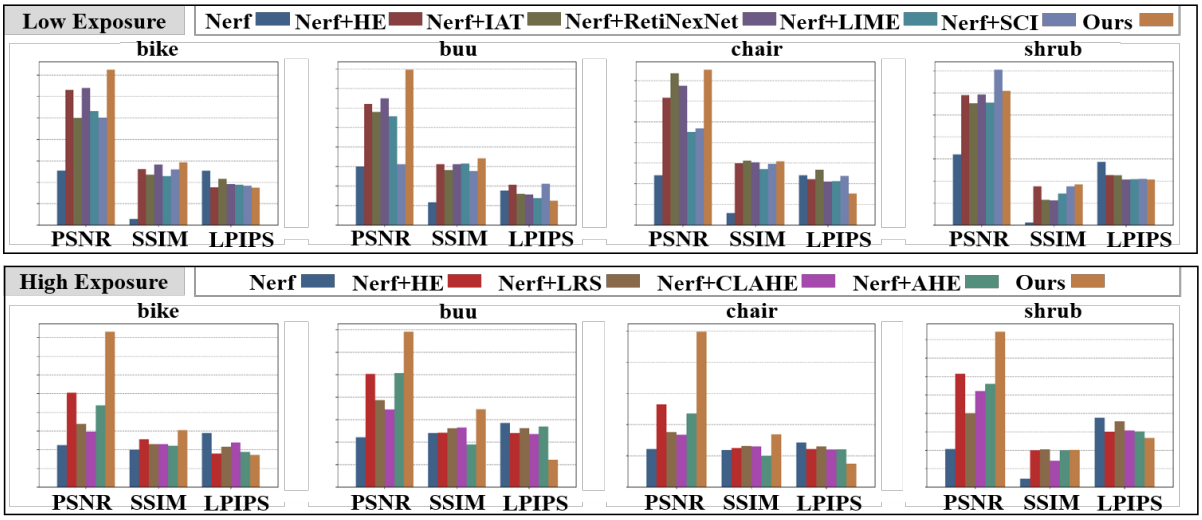

Figure 5: Comparison of enhancement results using different methods on the LOM dataset.

Figure 6: Example of using solely HDR images in optimization process (row one) and combined use of both multi-exposure inputs and HDR images (row two).

AHE, CLAHE, and LRS(Srinivasan and Balram 2006). Partial results are shown in Fig. 3, the upper part is the result comparing with methods of low-light enhancement, the lower part is the comparison with exposure correction methods. The full comparisons are shown in Table. 1. The results presented in Fig. 4 demonstrate the superior performance of our method compared to LIME, RetinexNet(Wei et al. 2018), and EnGAN(Jiang et al. 2021) on the LOL dataset. By using our approach, we effectively restore both color information and texture while preserving the overall structural consistency and content consistency of the images. We also evaluated various enhancement techniques as part of the post-processing stage. This involved rendering multi-exposure scenes using NeRF, followed by the application of 2D enhancement methods to post-process these novel views, a process we refer to as "NeRF + *". Conversely, the notation "* + NeRF" indicates the use of various pre-processing methods, followed by the synthesis of the novel view utilizing NeRF. As demonstrated in Fig. 5, our methods significantly outperform these techniques, showcasing remarkable performance improvements.

Ablation Study

Exposure Evaluator Analysis. As shown in Fig. 8 and Table. 2, we first analyze the influence of our exposure evaluator by replacing the gamma γ (Introduced in Equation.

Model	Components		”bike”			”buu”			”chair”			”shrub”			”sofa”		
	HDR	Eval	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
M1	✗	✗	5.99	0.287	0.679	6.53	0.448	0.579	6.08	0.370	0.658	6.08	0.061	0.735	6.27	0.441	0.626
M2	✗	✓	10.02	0.532	0.582	11.01	0.638	0.455	11.36	0.584	0.517	10.41	0.283	0.685	12.33	0.601	0.472
M3	✓	✗	16.37	0.659	0.477	15.36	0.729	0.396	16.23	0.672	0.436	13.23	0.372	0.544	17.61	0.732	0.395
M4	✓	✓	**19.47**	**0.747**	**0.435**	**18.59**	**0.860**	**0.308**	**20.97**	**0.799**	**0.379**	**16.05**	**0.433**	**0.526**	**20.83**	**0.871**	**0.387**

Table 2: Ablation study result on LOM dataset. Here ”HDR” denotes using HDR in optimization process, ”Eval” means using exposure evaluator for gamma mapping. The presented data represent the average across both low-light and overexposure conditions.

Method	Exp.	”buu”			”chair”		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Direct	Low	17.11	0.723	0.433	15.32	0.677	0.479
	High	15.74	0.654	0.479	20.93	0.721	0.441
Bilevel	Low	19.89	0.854	0.312	17.05	0.751	0.381
	High	17.29	0.865	0.304	24.88	0.846	0.376

Table 3: Results of the ablation study for the bilevel optimization model on the LOM dataset under both low and high exposure conditions.

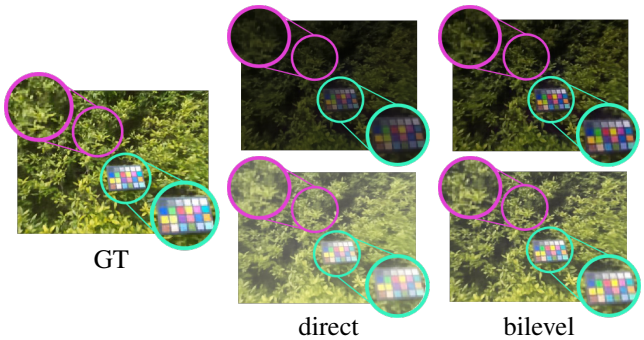

Figure 7: Comparison between the application (column 3) and non-application (column 2) of the proposed bilevel optimization under both low and high exposure conditions.

7) with randomly selected values and replacing the HDR fusion with evaluator selected gamma mapping. Using our proposed exposure evaluator with HDR fusion significantly improves the overall performance. Either using HDR fusion without the proposed evaluator or employing gamma mapping without incorporating multi-exposure image HDR fusion, can adversely affect the performance.

Experiments on Different Inputs. We further demonstrate the importance of using multi-exposure images as input by comparing our framework, which utilizes both multi-exposure and HDR fusion images, to an alternative approach where our framework solely relies on HDR images, as illustrated in Fig. 6. The result shows that utilization of multi-exposure input is crucial for preserving the content and structure consistency.

Bilevel Optimization or Joint Training. Fig. 7 showcases the enhancements achieved using our bilevel optimization approach relative to the direct joint training with the

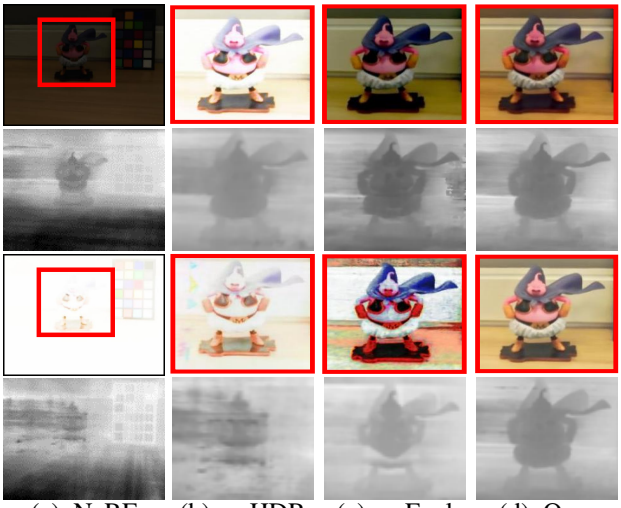

Figure 8: Visual comparison of our method(d) in contrast to the vanilla NeRF(a), our framework utilizing randomly selected gamma generated HDR(b), and our framework employing gamma mapping without the integration of multi-exposure image HDR fusion(c).

NeRF network. Our strategy not only facilitates the synthesis of novel view images but also sustains superior image quality under severe exposure conditions, as detailed in Table. 3. It effectively preserves and models both proximity and texture information.

Conclusion

We have introduced a innovative self-enhancement network for novel view synthesis that is robust to a variety of severe extreme lighting conditions, including under-exposure and over-exposure. In particular, we have developed a bilevel optimization method that enables the concurrent optimization of multi-exposure correction and the synthesis of new perspectives within NeRF in an unsupervised manner. The results demonstrate that compared to existing methods, our approach significantly improves the task of novel view synthesis for multi-exposure images, achieving state-of-the-art performance. These results justify the importance of our network in improving machine perception and visual understanding.

Acknowledgments

This work was partially supported by China Postdoctoral Science Foundation (2023M730741), and the National Natural Science Foundation of China (No.62302078).

References

Cao, Y.; and Bermak, A. 2011. An analog gamma correction method for high dynamic range applications. In *Proceedings of the IEEE International SOC Conference*, 318–322.

Chen, C.; Chen, Q.; Xu, J.; and Koltun, V. 2018. Learning to See in the Dark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3291–3300.

Cui, Z.; Gu, L.; Sun, X.; Qiao, Y.; and Harada, T. 2023. Aleth-NeRF: Low-light Condition View Synthesis with Concealing Fields. *arXiv preprint arXiv:2303.05807*.

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. In *Proceedings of the British Machine Vision Conference*, 238.

Guo, X.; Li, Y.; and Ling, H. 2017. LIME: Low-Light Image Enhancement via Illumination Map Estimation. *IEEE Transactions on Image Processing*, 26(2): 982–993.

Han, D.; Li, L.; Guo, X.; and Ma, J. 2022. Multi-exposure image fusion via deep perceptual enhancement. *Information Fusion*, 79: 248–262.

Hong, Y.; Peng, B.; Xiao, H.; Liu, L.; and Zhang, J. 2022. Headnerf: A real-time nerf-based parametric head model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20374–20384.

Huang, H.; Yang, W.; Hu, Y.; Liu, J.; and Duan, L.-Y. 2022. Towards low light enhancement with raw images. *IEEE Transactions on Image Processing*, 31: 1391–1405.

Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; and Wang, Z. 2021. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. *IEEE transactions on image processing*, 30: 2340–2349.

Le, P.-H.; Le, Q.; Nguyen, R.; and Hua, B.-S. 2023. Single-image hdr reconstruction by multi-exposure generation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 4063–4072.

Li, G.; Liu, J.; Ma, L.; Jiang, Z.; Fan, X.; and Liu, R. 2023. Fearless Luminance Adaptation: A Macro-Micro-Hierarchical Transformer for Exposure Correction. In *Proceedings of the 31st ACM International Conference on Multimedia*, 7304–7313.

Li, J.; Gu, B.; and Huang, H. 2022. A fully single loop algorithm for bilevel optimization without hessian inverse. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 7426–7434.

Li, J.; Li, J.; Fang, F.; Li, F.; and Zhang, G. 2021. Luminance-Aware Pyramid Network for Low-Light Image Enhancement. *IEEE Transactions on Multimedia*, 23: 3153–3165.

Liu, J.; Shang, J.; Liu, R.; and Fan, X. 2022a. Attention-guided global-local adversarial learning for detail-preserving multi-exposure image fusion. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(8): 5026–5040.

Liu, J.; Wu, G.; Luan, J.; Jiang, Z.; Liu, R.; and Fan, X. 2023a. HoLoCo: Holistic and local contrastive learning network for multi-exposure image fusion. *Information Fusion*, 95: 237–249.

Liu, R.; Ma, L.; Ma, T.; Fan, X.; and Luo, Z. 2022b. Learning with nested scene modeling and cooperative architecture search for low-light vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5): 5953–5969.

Liu, Y.; Liu, Z.; Ma, L.; Liu, J.; Fan, X.; Luo, Z.; and Liu, R. 2023b. Bilevel Generative Learning for Low-Light Vision. In *Proceedings of the 31st ACM International Conference on Multimedia*, 7758–7766.

Liu, Z.; Liu, J.; Wu, G.; Fan, X.; and Liu, R. 2023c. Embracing Compact and Robust Architectures for Multi-Exposure Image Fusion. *arXiv preprint arXiv:2305.12236*.

Lore, K. G.; Akintayo, A.; and Sarkar, S. 2017. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61: 127–141.

Lv, F.; Lu, F.; Wu, J.; and Lim, C. 2018. MBLLEN: Low-Light Image/Video Enhancement Using CNNs. In *Proceedings of the British Machine Vision Conference*, volume 220, 4.

Ma, L.; Jin, D.; An, N.; Liu, J.; Fan, X.; Luo, Z.; and Liu, R. 2023. Bilevel fast scene adaptation for low-light image enhancement. *International Journal of Computer Vision*, 1–19.

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

Mertens, T.; Kautz, J.; and Van Reeth, F. 2007. Exposure Fusion. In *Proceedings of the Pacific Conference on Computer Graphics and Applications*, 382–390.

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.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; Ng, R.; and Martin-Brualla, R. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Proceedings of the European Conference on Computer Vision*, 405–421.

Nguyen, H.; Tran, D.; Nguyen, K.; and Nguyen, R. 2023. PSENet: Progressive Self-Enhancement Network for Unsupervised Extreme-Light Image Enhancement. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1756–1765.

Patel, O.; Maravi, Y. P.; and Sharma, S. 2013. A comparative study of histogram equalization based image enhancement techniques for brightness preservation and contrast enhancement. *arXiv preprint arXiv:1311.4033*.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention*, 234–241.

Sabach, S.; and Shtern, S. 2017. A First Order Method for Solving Convex Bi-Level Optimization Problems. *arXiv preprint arXiv:1702.03999*.

Srinivasan, P. P.; Deng, B.; Zhang, X.; Tancik, M.; Mildenhall, B.; and Barron, J. T. 2021. NeRV: Neural Reflectance and Visibility Fields for Relighting and View Synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7495–7504.

Srinivasan, S.; and Balram, N. 2006. Adaptive contrast enhancement using local region stretching. In *Proceedings of the 9th Asian symposium on information display*, 152–155.

Wang, C.; Chai, M.; He, M.; Chen, D.; and Liao, J. 2022a. Clip-nerf: Text-and-image driven manipulation of neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3835–3844.

Wang, W.; Xu, Z.; Huang, H.; and Liu, J. 2022b. Self-aligned concave curve: Illumination enhancement for unsupervised adaptation. In *Proceedings of the 30th ACM International Conference on Multimedia*, 2617–2626.

Wang, Y.; Peng, L.; Li, L.; Cao, Y.; and Zha, Z.-J. 2023. Decoupling-and-Aggregating for Image Exposure Correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18115–18124.

Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*.

Wu, K.; Chen, J.; and Ma, J. 2022. DMEF: Multi-exposure image fusion based on a novel deep decomposition method. *IEEE Transactions on Multimedia*.

Wu, K.; Chen, J.; Yu, Y.; and Ma, J. 2022. ACE-MEF: adaptive clarity evaluation-guided network with illumination correction for multi-exposure image fusion. *IEEE Transactions on Multimedia*.

Xu, H.; Haochen, L.; and Ma, J. 2023. Unsupervised multi-exposure image fusion breaking exposure limits via contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 3010–3017.

Xu, X.; Wang, R.; and Lu, J. 2023. Low-Light Image Enhancement via Structure Modeling and Guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9893–9903.

Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; and Liu, J. 2021a. Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality. *IEEE Transactions on Image Processing*, 30: 3461–3473.

Yang, W.; Wang, W.; Huang, H.; Wang, S.; and Liu, J. 2021b. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30: 2072–2086.

Zhang, H.; and Ma, J. 2023. IID-MEF: A multi-exposure fusion network based on intrinsic image decomposition. *Information Fusion*, 95: 326–340.

Zhang, J.; Luo, Y.; Huang, J.; Liu, Y.; and Ma, J. 2023. Multi-exposure image fusion via perception enhanced structural patch decomposition. *Information Fusion*, 101895.