Enhancing Low-Light Images: A Synthetic Data Perspective on Practical and Generalizable Solutions

Yu Long[1,3], Qinghua Lin[2,3], Zhihua Wang[3*], Kai Zhang[4], Jianguo Zhang[5, 6], Yuming Fang[7]

[1]School of Computer Science, Beijing Institute of Technology
[2]School of Computer Science, Guangdong University of Technology
[3]Department of Engineering, Shenzhen MSU-BIT University
[4]School of Intelligence Science and Technology, Nanjing University, Suzhou
[5]Department of Computer Science and Engineering, Southern University of Science and Technology
[6]Pengcheng Laboratory
[7]School of Information Management, Jiangxi University of Finance and Economics
longyu2001@bit.edu.cn, 2112405034@mail2.gdut.edu.cn, zhihua.wang@my.cityu.edu.hk, kaizhang@nju.edu.cn, zhangjg@sustech.edu.cn, fa0001ng@e.ntu.edu.sg

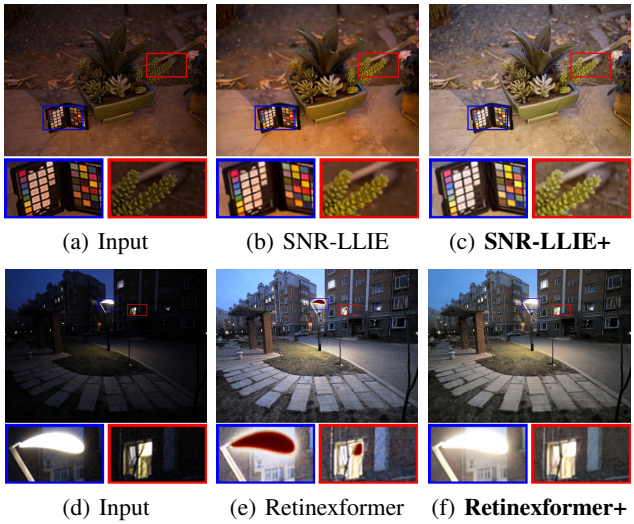
Abstract

Recently, deep neural networks (DNNs) have emerged as the leading approach for low-light image enhancement (LLIE). However, training these models requires large-scale paired datasets, which are challenging to obtain due to the labor-intensive nature of real-world data collection. To alleviate this issue, synthetic data are often combined with real-captured data for training. However, most existing low-light image synthesis methods are simply performed in the sRGB domain using Gamma correction or manual adjustments, which fail to incorporate the physical imaging prior through the image signal processing (ISP) pipeline and thus result in limited dataset size and degradation space. Consequently, LLIE methods trained on such data often exhibit some drawbacks in the results, such as inaccurate white balance and abnormal enhancement artifacts, which limit their practicality and generalizability. In this paper, we propose a practical low-light image synthesis pipeline capable of generating unlimited paired training data. Our pipeline starts with a reverse ISP model that converts sRGB images back to the unprocessed RAW domain, where we then simulate low-light degradation, noise degradation, and white balance adjustments. Finally, the degraded RAW images are processed through a forward ISP model to produce low-light sRGB images. The pipeline further employs multiple tone mapping curves and color correction matrices (CCMs) to expand the degradation space. Hence, trained with our proposed synthetic data, existing state-of-the-art (SOTA) LLIE deep models are expected to improve their performance. Extensive experiments across various datasets demonstrate that our synthetic data can indeed effectively enhance existing LLIE deep models, improving both their practicality and generalizability.

code — https://github.com/LongYu-LY/SynLLIE

Introduction

Digital images captured in insufficiently illuminated environments often suffer from degraded image quality, such as poor visibility, low contrast, and unexpected noise (Zhang et al. 2023). These issues negatively impact both human visual experience and high-level vision tasks, such as object detection (Loh and Chan 2019), semantic segmentation (Wang et al. 2022a), and depth estimation (Lamba, Rachavarapu, and Mitra 2020). These problems can be addressed with hardware-based methods. For example, extending the exposure time can capture more photons. Additionally, using a flash for light compensation frequently results in unwanted highlights and uneven lighting, with objects



(a) Input (b) SNR-LLIE (c) **SNR-LLIE+**

(d) Input (e) Retinexformer (f) **Retinexformer+**

Figure 1: Visual comparisons of recent SOTA LLIE methods, e.g. SNR-LLIE (Xu et al. 2022) and Retinexformer (Cai et al. 2023), trained on LOL-v2 (Yang, Nie, and Liu 2019) versus ours synthetic data (indicated by the "+" symbol). Methods only trained on LOL-v2 exhibit issues such as incorrect white balance (top) and abnormal enhancement (bottom). In contrast, the **SNR-LLIE+** and **Retinexformer+** can produce more visually pleasant results.

*Corresponding author.

closer to the camera being brightened more than those farther away (Petschnigg et al. 2004). In contrast to upgrading cameras, which may require additional hardware, Low-Light Image Enhancement (LLIE) techniques are designed to autonomously enhance the visibility of images captured in low-light conditions. Generally, LLIE is not solely a problem of light adjustment; it also involves addressing issues such as noise bursts and color distortion, which are often concealed in darkness due to the limited capabilities of photographic devices (Guo and Hu 2023).

Before the advent of deep learning, traditional LLIE primarily relied on straightforward illumination amplification techniques such as Gamma correction (Guo, Li, and Ling 2017; Wang et al. 2023) and histogram equalization (HE) (Kim 1997). Although these traditional techniques are grounded in robust theoretical principles, their practical applications often yield unsatisfactory results. Recent years have witnessed a significant surge in efforts to address the challenges of LLIE through deep learning-based methods (Ying, Li, and Gao 2017; Zhang, Zhang, and Guo 2019). Leveraging the advanced representational capabilities of deep neural networks (DNNs) and the availability of extensive datasets, these modern techniques have set new performance benchmarks for LLIE, achieving unprecedented levels of quality and accuracy (Cai et al. 2023; Bai, Yin, and He 2024). These approaches primarily conceptualize LLIE as either an image-to-image translation problem (Cai, Gu, and Zhang 2018; Fu et al. 2023; Shen et al. 2017) or reflectance decomposition inspired by Retinex theory (Wei et al. 2018; Cai et al. 2023; Bai, Yin, and He 2024). Typically, these methods employ convolutional neural networks (CNNs) (Li, Cheng, and Zhang 2020), Transformers (Cai et al. 2023), and Mamba networks (Bai, Yin, and He 2024). While these DNN-based LLIE methods demonstrate considerable promise, they often exhibit poor practicality and generalizability (see Figure 1).

These limitations in practicality and generalizability primarily arise from the scarcity of large-scale paired datasets for training, as acquiring such data in real-world settings is both labor-intensive and time-consuming (Lv, Li, and Lu 2021). To address this data shortage, a combination of real-captured and synthetic paired data is frequently used to enlarge the dataset (Wei et al. 2018; Yang et al. 2021b). Low-light images differ from normal images mainly due to two key characteristics: reduced visibility, which results in a loss of detail and clarity, and increased noise (Lv, Li, and Lu 2021). Existing low-light synthesis pipelines generally simulate only a subset of these degradations in the sRGB domain, often using a combination of linear and gamma transformations, or by making manual adjustments via Adobe Lightroom (Wei et al. 2018; Yang, Nie, and Liu 2019), to approximate underexposure (Lv, Li, and Lu 2021). However, these manual adjustments restrict the dataset size, and working exclusively in the sRGB domain fails to capture the complexities of processing low-light images through the image signal processing (ISP) pipeline. This limitation is particularly significant for learning-based techniques, as the mismatch between training data and real-world scenarios can hinder generalization.

In this paper, we introduce a practical low-light synthesis pipeline that emulates the image signal processing (ISP) pipeline from the RAW domain to the sRGB domain. This pipeline integrates several standard camera modules, including exposure reduction, noise simulation, white balance adjustment, and color correction (Brooks et al. 2019). Our pipeline ensures that the generated data covers a broad range of degradations and supports scalable dataset sizes. Overall, our contributions can be summarized as follows: (1) We introduce a meticulously designed synthetic pipeline that simulates real-world low-light image degradations, accurately replicating the data acquisition process from the RAW domain to the sRGB domain. This pipeline enables the generation of unlimited paired data with a broad degradation space for low-light conditions; (2) Leveraging the proposed synthetic pipeline, we refine the learning framework for existing state-of-the-art (SOTA) LLIE methods. Specifically, we train models with our generated data and fine-tune them to adapt to target image styles, thereby enhancing performance with current network architectures; (3) We conduct extensive experiments to validate the effectiveness of our synthetic pipeline. The results show that our synthetic data significantly improve LLIE methods in both quantitative and qualitative evaluations, leading to substantial gains in practicality and generalizability.

Related Work

Datasets for LLIE

Paired Datasets: Supervised learning methods for LLIE rely heavily on paired training datasets (Wang et al. 2019; Yang et al. 2021b). Typically, these datasets are obtained by capturing images in real-world settings with varying exposure times and ISO settings (Chen et al. 2018). This process is labor-intensive and results in relatively small datasets. Challenges such as camera shake, object movement, and fluctuating lighting conditions can lead to misalignment between paired images. Additionally, hardware constraints complicate the capturing process, particularly when aiming to cover a wide range of diverse scenes in controlled environments. To address these limitations, synthetic images are often used to meet the requirements for supervised learning (Yang et al. 2021b). These synthetic image generation methods usually involve basic techniques such as gamma correction or exposure adjustments with software like Lightroom. However, these approaches often fall short in accurately representing the complex illumination conditions found in real-world environments. Consequently, generating synthetic images that effectively capture the diversity and complexity of actual scenes remains a significant challenge.

Unpaired Datasets: Several publicly available datasets primarily provide under-exposed images intended for testing low-light image enhancement methods. Notable examples include NPE (Wang et al. 2013), LIME (Guo, Li, and Ling 2017), MEF (Ma, Zeng, and Wang 2015), and DICM (Lee, Lee, and Kim 2013), which contain 8, 10, 17, and 64 real low-light images, respectively. These datasets encompass a range of scenes such as indoor items, outdoor buildings, streetscapes, and natural landscapes. In addition, some un-

paired datasets are designed to evaluate performance in high-level vision tasks. For instance, Hwang et al.(Hwang et al. 2015) introduced the KAIST Dataset, which features nighttime traffic sequences for pedestrian detection. Loh and Chan (Loh and Chan 2019) developed the Exclusively Dark dataset (ExDark), providing annotations with 12 object classes for image-level categories and local object bounding boxes. Sakaridis et al.(Sakaridis, Dai, and Van Gool 2021) created the ACDC dataset for nighttime driving scene understanding, while Yang et al.(Yang et al. 2020b) proposed DarkFace, which includes 10,000 real low-light images captured in nighttime streetscapes.

Low-light Image Enhancement Methods

Traditional Methods: HE-based methods spread out the frequent intensity values of an image to enhance its global contrast. The basic HE approach (Pizer et al. 1987) focuses on global adjustment, which can result in poor local illumination and amplified degradation, such as noise, blur, and artifacts. To mitigate these issues, Pizer et al.(Pizer 1990) performed HE on partitioned regions with local contrast constraints to suppress noise. Retinex-based methods are grounded in the Retinex theory of color vision (Land and McCann 1971), which suggests that an image can be decomposed into a reflectance map and an illumination map. These Retinex-based models heavily rely on handcrafted priors to achieve satisfactory image enhancement. Other approaches, such as Dehazing-based and Statistical methods, have also been employed for LLIE.

Deep Learning-Based Methods: Most DNN-based LLIE methods primarily employ supervised learning techniques, which depend on synthesized or real-captured paired images (Zhang et al. 2021). However, collecting real-world paired data or developing an effective pipeline to synthesize images that capture the complexity of real-world low-light scenes is both labor-intensive and costly. To mitigate this issue, unsupervised learning (Jiang et al. 2021), semi-supervised learning (Yang et al. 2021a; Malik and Soundararajan 2023), and zero-shot learning (Zheng and Gupta 2022; Zhang et al. 2024) techniques are employed to reduce the reliance on paired datasets. Regarding network architecture, U-shape network structures (Zhang, Zhang, and Guo 2019) are the most prevalent in LLIE due to their ability to preserve both high-resolution detail features and low-resolution semantic features. Recently, transformer-based methods (Cai et al. 2023) and Mamba-based approaches (Bai, Yin, and He 2024) have gained prominence in the LLIE field. Due to the limited variation in scenes and lighting conditions, the trained models often fail to generalize effectively.

Proposed Method

Camera Simulation Pipeline

We present a sophisticated camera pipeline specifically designed to synthesize low-light images from readily available high-quality normal-light images (Brooks et al. 2019; Zhang et al. 2023). This pipeline begins by unprocessing[1] sRGB images to generate their RAW counterparts. This process involves several key steps to convert sRGB images back to RAW format, including inverse gamma correction and tone mapping to restore linear sensor data, inverse color interpolation to transform linear RGB into the XYZ color space, inverse white balance, exposure compensation, and applying a mosaic pattern to replicate RAW sensor data. In the RAW domain, we synthesize low-light degradations by adding shot and read noise and adjusting exposure values to mimic how digital sensors capture images under reduced lighting. The noise-affected and exposure-adjusted RAW images are then reprocessed through critical stages of the ISP, such as white balance, color matrix transformation, and tone mapping, to produce their final sRGB form. Compared to existing low-light synthesis pipelines (Lv, Li, and Lu 2021), our method incorporates the physical characteristics of digital sensors and key components of the imaging pipeline, resulting in synthetic low-light images that closely replicate real-world conditions.

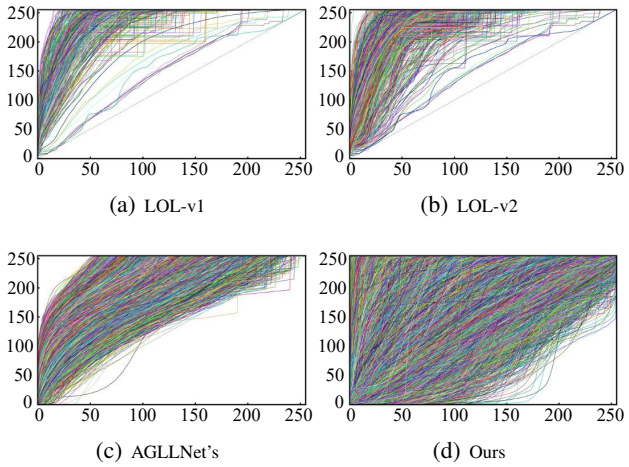


Figure 2: Comparison of the low-light coverage across different datasets using the exposure adjustment curves, which map the luminance histogram of the low-light images to that of their corresponding normal-light ground truth. A steeper curve suggests a higher degree of underexposure.

Noise Degradation Synthesis

We begin by synthesizing the low-light noise inherent in RAW images. Practically, the noise in the RAW domain is primarily attributed to two key factors: the randomness in photon arrival, termed as "shot" noise, and the imperfections in the readout electronics, known as "read" noise (Brooks et al. 2019). Shot noise follows a Poisson distribution, with its mean representing the actual light intensity, expressed in photoelectrons. Conversely, read noise is generally modeled by a Gaussian distribution with a mean of zero and a constant variance. As suggested by Brooks et al. (Brooks et al.

[1]"Unprocessing" refers to the reverse ISP process that converts sRGB images to their corresponding RAW format.

2019), the Poisson-Gaussian noise model can be approximated using a single heteroscedastic Gaussian, defined as

$$\mathbf{y} \sim \mathcal{N}\left(\mu = \mathbf{x}, \sigma^2 = \lambda_r + \lambda_s \mathbf{x}\right), \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^{3N}$ of N pixels is a raw clear image, λ_r and λ_s represent the sensor's analog and digital gains, respectively. In previous low-light synthesis pipelines (Yang, Nie, and Liu 2019; Lv, Li, and Lu 2021), noise is commonly simulated in the sRGB domain. This approach often overlooks the variations in noise distribution that arise from the sensor data processing through various ISP stages. Our low-light noise synthesis pipeline addresses this gap by accounting for the comprehensive redistributions caused by the varying configurations of ISP parameters. This ensures a more accurate representation of real-world noise characteristics in the synthesized low-light images.

Low-light Synthesis

Digital camera sensors can capture images in low-light conditions with long exposure times, allowing for much greater detail recording as light accumulates on the sensor over an extended period. Generally, exposure time dictates how long the camera collects light from the target object, with trade-offs between exposure time, image brightness, and phototoxicity. Therefore, it is feasible to linearly adjust the exposure values to synthesize low-light images effectively (Hu et al. 2018). This process mimics the natural accumulation of light, thereby creating realistic low-light conditions in the synthesized images. In our pipeline, the low-light effect is synthesized by multiplying an input image by exposure scaling parameters $\mathbf{s}$ applied across the three color channels, which are defined as

$$\mathbf{s} = 2^{-g}, \quad g \sim \mathcal{U}(0, k), \tag{2}$$

where $\mathcal{U}(0, k)$ determines the range of exposure reduction.

Color Correction

Images captured by digital cameras often exhibit color inaccuracies. To mitigate these discrepancies, color correction methods such as white balance and color matrix transformations are applied.

White Balance Variation. Digital cameras often encounter difficulties with auto white balance settings in different low-light conditions, leading to unwanted blue, orange, or green color casts in processed images. To improve the white balance capabilities of LLIE methods, we introduce variability in the white balance gains, generating a diverse set of images with intentional white balance inaccuracies. These white balance gains are randomly sampled from a predefined empirical range, which is defined as

$$\mathbf{w} = \left[\frac{1.0}{(1.2 + 2g_1)}, 1.0, \frac{1.0}{(1.2 + g_2)}\right], \quad g_1, g_2 \sim \mathcal{U}(0, 1), \tag{3}$$

where we only adjust the red and blue channels. To prevent the dimming of saturated pixels, we apply a highlight-preserving transformation that smoothly adjusts gains near white (Brooks et al. 2019). This approach enhances the effectiveness of LLIE methods in correcting various white balance inaccuracies, ensuring that the final images accurately reflect the true colors.

Color Matrix Transformation. The color transformation process involves applying a Color Correction Matrix (CCM) to adjust the red, green, and blue pixel data for each pixel. This process accounts for the interdependence between the color channels, correcting color shifts caused by the light source during image acquisition and compensating for imperfections in the sensor's color generation process. Camera profiles, which store the CCM, manage how colors are interpreted by specific camera sensors. The transformation is performed by premultiplying a 3×1 matrix of R, G and B pixel values by a 3×3 matrix containing the color transformation values. In our pipeline, we utilize 11 predefined camera profiles from popular brands such as Canon, Huawei, Nikon, and Olympus to control color and tonality rendering, generating a diverse set of color-rendered images.

Tone Mapping Curves

In the ISP pipeline, tone mapping is utilized to adjust the intensity of input pixels, enhancing contrast and color tones in the final image to improve its aesthetics. It is well known that tone mapping curves vary across different cameras, aiming to match the perception of the tone-mapped image with real-world perception. Rather than employing a fixed tone curve (Brooks et al. 2019), our low-light ISP utilizes 200 distinct tone curves for each camera to adjust the intensity of the input pixels, ensuring a diverse color representation.

Dataset Characteristics

Compared to existing low-light synthetic pipelines, our proposed method offers several key advantages. First, while previous approaches often introduce noise directly in the sRGB domain using AWGN and Poisson noise, they do not account for the alterations in noise characteristics caused by the ISP (Wei et al. 2018; Lv, Li, and Lu 2021). Our method synthesizes low-light noise in the RAW domain and then processes it through the ISP pipeline, providing a more accurate representation of real-world noise processing. Second, our approach simulates a more extensive low-light degradation space and supports unlimited data generation. Additionally, our pipeline further incorporates diverse configurations for white balance, gamma correction, and tone mapping, accurately simulating the color variations produced by different camera ISPs. Figure 2 compares the low-light range between our synthetic dataset and existing low-light datasets, including LOL-v1 (Wei et al. 2018), LOL-v2 (Yang, Nie, and Liu 2019), and the synthetic dataset used to train AGLLNet (Lv, Li, and Lu 2021) (denoted as AGLLNet's). The shape of these curves reflects the degree of underexposure, while their coverage indicates the diversity of lighting conditions. Our synthetic dataset exhibits a wide range of underexposure levels, which is beneficial to enhance the practicality and generalizability of existing SOTA LLIE methods.

Dataset	LOL-v1						LOL-v2					
Metric	PSNR↑	SSIM↑	LPIPS ↓	MUSIQ↑	CLIPIQA↑	Q-Align↑	PSNR↑	SSIM↑	LPIPS ↓	MUSIQ↑	CLIPIQA↑	Q-Align↑
SNR-LLIE	22.593	0.832	0.156	64.979	0.500	61.903	21.229	0.850	0.156	63.873	0.488	62.208
Retinexformer	23.622	0.828	0.148	62.910	0.523	57.413	22.236	0.849	0.161	61.209	0.454	59.319
RetinexMamba	23.527	0.829	0.143	63.568	0.527	58.933	22.010	0.841	0.159	61.402	0.489	58.086
SNR-LLIE+	23.193	0.846	0.144	68.140	0.537	68.934	21.645	0.856	0.160	64.303	0.496	65.418
Retinexformer+	23.999	0.854	0.100	72.287	0.622	77.490	22.514	0.874	0.106	69.030	0.590	77.940
RetinexMamba+	24.133	0.855	0.100	72.239	0.631	77.940	22.345	0.870	0.113	68.446	0.592	73.572

Table 1: Quantitative comparisons of LLIE models trained on our synthetic data and paired datasets: LOL-v1 (Wei et al. 2018) and LOL-v2 (Yang, Nie, and Liu 2019). The "+" symbol denotes models retrained using our synthetic pipeline.

(a) Input/GT (b) SNR-LLIE (c) **SNR-LLIE+** (d) Retinexformer (e) **Retinexformer+** (f) RetinexMamba (g) **RetinexMamba+**

Figure 3: Qualitative comparison of LLIE methods trained solely on the LOL-v1 (top) and LOL-v2 (bottom) datasets versus those pre-trained on our synthetic data and subsequently fine-tuned on the LOL-v1 (top) and LOL-v2 (bottom) datasets, as indicated by the "+" symbol. The images in the top and bottom rows are sampled from the LOL-v1 and LOL-v2 datasets, respectively. For optimal visual assessment, please zoom in.

Experiment and Results

Experimental Setups

Datasets. We utilize a combination of ImageNet (Deng et al. 2009) and COCO (Lin et al. 2014), comprising more than one million images, as normal-light images to synthesize low-light images. The performance assessment is carried out using three well-established paired datasets: LOL-v1 (Wei et al. 2018), LOL-v2 (Yang, Nie, and Liu 2019), and the MIT-Adobe FiveK Dataset (Bychkovsky et al. 2011), as well as four real-captured unpaired benchmarks: i.e., NPE (Wang et al. 2013), DICM (Lee, Lee, and Kim 2013), LIME (Guo, Li, and Ling 2017) and VV datasets.

Implementation Details. We employ three SOTA LLIE models—SNR-LLIE (Xu et al. 2022), Retinexformer (Cai et al. 2023), and RetinexMamba (Bai, Yin, and He 2024)—as benchmarks in our experiments. The models are trained using a weighted combination of losses, including L1 loss, SSIM (Wang et al. 2004), VGG perceptual loss (Johnson, Alahi, and Fei-Fei 2016), and UNetGAN loss (Wang et al. 2022b). We then randomly crop image patches of size $156 \times 156 \times 3$ pixels and simulate low-light conditions to create paired low-/high-quality image patches. To minimize any biases introduced by the pre-trained datasets, we adopt a two-stage training strategy: an initial pre-training phase followed by fine-tuning on the target sets. All models are trained on two NVIDIA RTX 4090 GPUs.

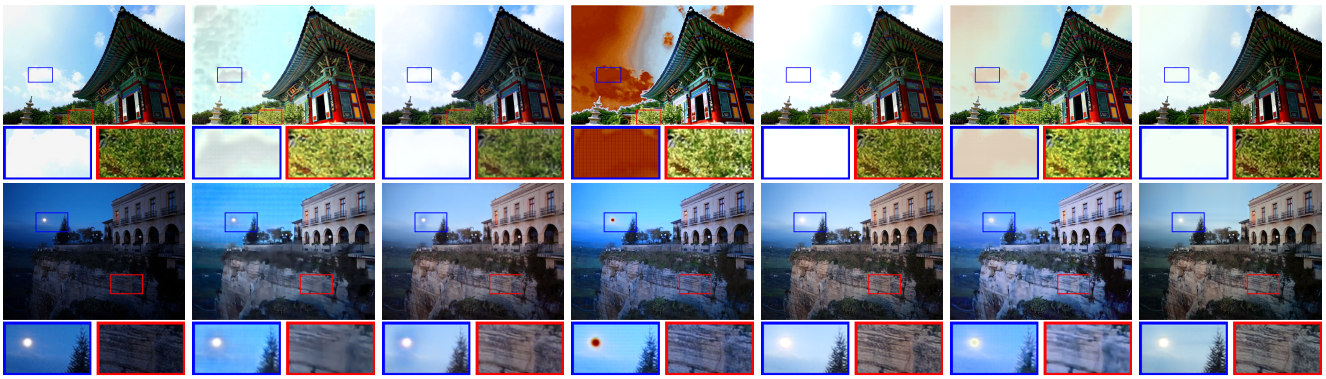
Evaluation Metrics. We employ three widely used full-reference IQA metrics—PSNR, Structural Similarity Index (SSIM) (Wang et al. 2004), and LPIPS (Zhang et al. 2018)—to assess the similarity between enhanced images and their corresponding ground-truth images. Besides, we also incorporate several SOTA blind IQA metrics (Liu et al. 2021), including MUSIQ (Ke et al. 2021), CLIPIQA (Wang, Chan, and Loy 2023) and Q-Align (Wu et al. 2023), due to their high performance in assessing authentic distortions.

Experimental Results

Quantitative Comparison. We begin by comparing models trained on our synthetic data with those trained on paired datasets, specifically LOL-v1 and LOL-v2. Table 1 presents the quantitative results across various metrics. Notably, LLIE models trained on our synthetic data, followed by fine-tuning to match the target image style, demonstrate superior performance compared to models trained exclusively on the LOL-v1 and LOL-v2 datasets. This highlights the effectiveness of our low-light synthetic pipeline in significantly enhancing model performance. To further demonstrate the generalizability of our approach, we evaluate its performance on the paired MIT-Adobe FiveK dataset in addition to four unpaired, real-world benchmarks: NPE, DICM, LIME, and VV. We employ CLIPIQA and Q-Align as evaluation metrics due to their superior effectiveness in assessing authentic distortions. In this study, we compare LLIE models trained

Dataset	MIT-Adobe FiveK					NPE		DICM		LIME		VV	
Metric	PSNR↑	SSIM↑	LPIPS↓	CLIPIQA↑	Q-Align↑	CLIPIQA↑	Q-Align↑	CLIPIQA↑	Q-Align↑	CLIPIQA↑	Q-Align↑	CLIPIQA↑	Q-Align↑
SNR-LLIE	17.841	0.783	0.162	0.561	63.309	0.556	58.256	0.522	54.918	0.387	53.471	0.491	54.546
Retinexformer	18.273	0.751	0.224	0.476	47.744	0.475	52.877	0.463	53.634	0.490	57.614	0.414	48.141
RetinexMamba	18.287	0.772	0.194	0.518	48.386	0.563	61.258	0.521	57.610	0.517	57.160	0.413	59.193
SNR-LLIE+	19.446	0.806	0.156	0.617	66.510	0.637	63.996	0.542	57.147	0.409	55.092	0.523	58.947
Retinexformer+	19.436	0.804	0.153	0.620	64.362	0.680	71.957	0.610	70.379	0.617	68.456	0.514	64.309
RetinexMamba+	20.262	0.815	0.153	0.636	66.164	0.668	71.364	0.602	69.392	0.617	66.592	0.498	64.080

Table 2: Quantitative comparisons conducted on the paired MIT-Adobe FiveK dataset (Bychkovsky et al. 2011) and the unpaired NPE (Wang et al. 2013), DICM (Lee, Lee, and Kim 2013), LIME (Guo, Li, and Ling 2017) and VV datasets to evaluate the generalizability of LLIE models trained using our proposed low-light image synthesis pipeline. The "+" indicates that the model are retrained using the low light dataset generated by our proposed synthetic pipeline.



(a) Input (b) SNR-LLIE (c) **SNR-LLIE+** (d) Retinexformer (e) **Retinexformer+** (f) RetinexMamba (g) **RetinexMamba+**

Figure 4: Generalization comparison of LLIE methods trained on the LOL-v2 dataset versus those trained on our synthetic data indicated by the "+" symbol, where the top and bottom images are sampled from the DICM and LIME datasets, respectively. For optimal viewing, please zoom in.

(a) Input (b) SNR-LLIE (c) **SNR-LLIE+** (d) Retinexformer (e) **Retinexformer+** (f) RetinexMamba (g) **RetinexMamba+**

Figure 5: Visual comparison of white balance correction. The input images are sourced from the Rendered WB dataset (Afifi et al. 2019).

on the LOL-v2 dataset with those trained using our synthetic data. As depicted in Table 2, models trained with our synthetic datasets show significant performance improvements across all benchmarks. These results indicate that our synthetic data not only generate visually pleasing images but also generalize robustly to a variety of scenes. Consequently, this enhances both the practicality and the generalizability of LLIE methods.

Qualitative Comparison. We present visual comparisons of our method against competitive approaches on paired datasets in Figure 3. The images in the top and bottom rows are selected from the LOL-v1 and LOL-v2 test sets, respectively. As shown, previous methods tend to produce results with color distortion or noise amplification. In contrast, our method effectively enhances both global and local contrast, reconstructs sharper details, and suppresses noise, resulting in more visually pleasing outcomes. Additionally, we provide results on unpaired benchmarks in Figure 4, with images in the top and bottom rows selected from the DICM and LIME datasets, respectively. Methods trained on LOL-v2 introduce artifacts around light sources or produce abnormal enhancement results. In contrast, models trained on our synthetic dataset achieve correct exposure and vibrant colors, demonstrating superior generalizability.

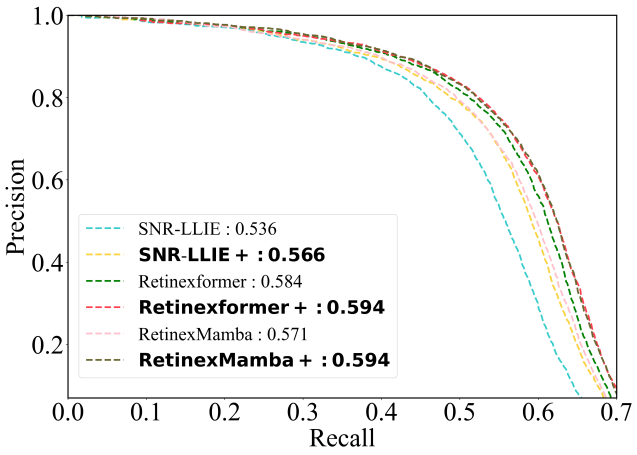


Figure 6: Precision-Recall curves and Average Precision (AP) values on the DARK FACE dataset.

Further Testing

White Balance Correction. Low-light conditions often result in improper color balance, producing images with unnatural hues. White balance correction, which adjusts the color temperature, is crucial for mitigating the impact of ambient light on color accuracy. In Figure 5, we present a visual comparison of LLIE methods trained exclusively on the LOL-v2 dataset versus those trained on our synthetic datasets, with a particular emphasis on white balance correction. The results clearly demonstrate that LLIE methods trained on our synthetic datasets produce more visually pleasing and color-accurate images.

Low-light Face Detection. We utilize the DARK FACE dataset (Yang et al. 2020a) to evaluate the high-level benefits of LLIE methods trained on LOL-v2 and our synthetic data. In our experiment, we use the YOLO-v5s as the detection model to generate precision-recall (P-R) curves and calculate the average precision (AP). As shown in Figure 6, the results indicate that the model, when trained with our synthetic data, exhibits enhanced practicality and performance.

Ablation Study

Our ablation studies are conducted using Retinexformer (Cai et al. 2023) as the baseline model, with performance evaluated on the LOL-v2 dataset (Yang, Nie, and Liu 2019).

Comparison of Low-light Synthesis Pipeline. Several prior LLIE methods have relied on synthetic low-light data for training. RetinexNet (Wei et al. 2018) was trained on synthetic datasets manually adjusted using Adobe Lightroom. MBLLEN (Lv et al. 2018) utilized gamma correction to modify pixel intensities in normal-light images and added Poisson noise to simulate low-light conditions. Similarly, GladNet (Wang et al. 2018) employed synthetic data generated from raw images by adjusting exposure, vibrance, and contrast in Adobe Lightroom. AGLLNet (Lv, Li, and Lu 2021) was trained on synthetic datasets created through linear and Gamma transformations, incorporating Gaussian-Poisson mixed noise to more accurately represent real-world noise characteristics. Table 3 presents a performance comparison of different low-light synthetic methods used to train the Retinexformer. The results demonstrate that the low-light data generated by our synthetic pipeline significantly outperforms other synthetic methods, yielding superior results across both full-reference and blind IQA metrics.

Metric	PSNR↑	SSIM↑	LPIPS↓	MUSIQ↑	CLIPIQA↑	Q-Align↑
RetinexNet's	18.103	0.652	0.354	52.907	0.449	42.538
MBLLEN's	19.183	0.761	0.241	52.921	0.473	43.972
GladNet's	20.831	0.660	0.304	53.361	0.543	49.036
AGLLNet's	21.236	0.796	0.224	52.615	0.451	44.360
Ours	**22.168**	**0.841**	**0.155**	**66.021**	**0.578**	**65.038**

Table 3: Quantitative comparisons of different low-light synthetic datasets.

Metric	PSNR↑	SSIM↑	LPIPS↓	MUSIQ↑	CLIPIQA↑	Q-Align↑
Without Noise	21.449	0.757	0.265	58.051	**0.580**	51.666
sRGB domain	21.830	0.800	0.217	59.475	0.539	49.932
RAW domain	**22.168**	**0.841**	**0.155**	**66.021**	0.578	**65.038**

Table 4: The ablation studies on noise synthesis, e.g., without noise or synthesizing noise in the sRGB domain or RAW domain. The default setting is marked with an underline.

Noise Synthesis in the sRGB vs. RAW Domain. Most synthetic low-light datasets add noise directly in the sRGB domain (Wei et al. 2018), which does not accurately capture the sensor-specific noise characteristics, especially after being processed by various steps in the ISP pipeline. Table 4 presents a comparison of noise synthesis approaches, including scenarios with no added noise, noise synthesized in the sRGB domain, and noise synthesized in the RAW domain. The results clearly demonstrate that noise synthesized in the RAW domain leads to superior improvements across various quantitative metrics, thus validating the effectiveness of our proposed practical low-light synthetic pipeline.

Conclusion

This paper introduces a novel approach to synthetic low-light data generation, designed to address the challenges associated with the limited availability of paired datasets crucial for training effective and generalizable DNN-based LLIE methods. Our proposed pipeline integrates practical noise synthesis and a wide range of illumination levels by emulating the ISP pipeline, capturing a variety of often-overlooked degradations that occur during different stages of camera processing. Additionally, we employ multiple tone-mapping curves and CCMs to further expand the degradation space. Our pipeline enables scalable dataset generation without the need for manual intervention, significantly improving the practicality and flexibility of synthetic data creation. Experimental results across various datasets demonstrate that our synthetic data substantially enhance the performance, practicality, and generalizability of existing SOTA LLIE methods, both in terms of quantitative metrics and visual quality.

Acknowledgements

This work was supported in part by the National Key Research and Development Program of China (2023YFE0210700), the National Natural Science Foundation of China (12326604, 62301323), the Shenzhen Natural Science Foundation (20231128191435002), the Key Technologies Special Program of Higher Education Institutions in Guangdong Province (2023ZDZX1035), and the Suzhou Key Technologies Project (Grant No. SYG2024136).

References

Afifi, M.; Price, B.; Cohen, S.; and Brown, M. S. 2019. When color constancy goes wrong: Correcting improperly white-balanced images. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1535–1544.

Bai, J.; Yin, Y.; and He, Q. 2024. Retinexmamba: Retinex-based mamba for low-light image enhancement. *arXiv preprint arXiv:2405.03349*.

Brooks, T.; Mildenhall, B.; Xue, T.; Chen, J.; Sharlet, D.; and Barron, J. T. 2019. Unprocessing images for learned raw denoising. In *IEEE Conference on Computer Vision and Pattern Recognition*, 11036–11045.

Bychkovsky, V.; Paris, S.; Chan, E.; and Durand, F. 2011. Learning photographic global tonal adjustment with a database of input/output image pairs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 97–104.

Cai, J.; Gu, S.; and Zhang, L. 2018. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4): 2049–2062.

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

Chen, C.; Chen, Q.; Xu, J.; and Koltun, V. 2018. Learning to see in the dark. In *IEEE Conference on Computer Vision and Pattern Recognition*, 3291–3300.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.

Fu, Z.; Yang, Y.; Tu, X.; Huang, Y.; Ding, X.; and Ma, K.-K. 2023. Learning a simple low-light image enhancer from paired low-light instances. In *IEEE Conference on Computer Vision and Pattern Recognition*, 22252–22261.

Guo, X.; and Hu, Q. 2023. Low-light image enhancement via breaking down the darkness. *International Journal of Computer Vision*, 131(1): 48–66.

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.

Hu, Y.; He, H.; Xu, C.; Wang, B.; and Lin, S. 2018. Exposure: A white-box photo post-processing framework. *ACM Transactions on Graphics*, 37(2): 1–17.

Hwang, S.; Park, J.; Kim, N.; Choi, Y.; and So Kweon, I. 2015. Multispectral pedestrian detection: Benchmark dataset and baseline. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1037–1045.

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.

Johnson, J.; Alahi, A.; and Fei-Fei, L. 2016. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, 694–711.

Ke, J.; Wang, Q.; Wang, Y.; Milanfar, P.; and Yang, F. 2021. MUSIQ: Multi-scale image quality transformer. In *IEEE International Conference on Computer Vision*, 5148–5157.

Kim, Y.-T. 1997. Contrast enhancement using brightness preserving bi-histogram equalization. *IEEE Transactions on Consumer Electronics*, 43(1): 1–8.

Lamba, M.; Rachavarapu, K. K.; and Mitra, K. 2020. Harnessing multi-view perspective of light fields for low-light imaging. *IEEE Transactions on Image Processing*, 30: 1501–1513.

Land, E. H.; and McCann, J. J. 1971. Lightness and retinex theory. *JOSA*, 61(1): 1–11.

Lee, C.; Lee, C.; and Kim, C.-S. 2013. Contrast enhancement based on layered difference representation of 2D histograms. *IEEE Transactions on Image Processing*, 22(12): 5372–5384.

Li, S.; Cheng, Q. S.; and Zhang, J. 2020. Deep multi-path low-light image enhancement. In *IEEE International Conference on Multimedia Information Processing and Retrieval*, 91–96.

Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft COCO: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 740–755. Springer.

Liu, J.; Xu, D.; Yang, W.; Fan, M.; and Huang, H. 2021. Benchmarking low-light image enhancement and beyond. *International Journal of Computer Vision*, 129: 1153–1184.

Loh, Y. P.; and Chan, C. S. 2019. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178: 30–42.

Lv, F.; Li, Y.; and Lu, F. 2021. Attention guided low-light image enhancement with a large scale low-light simulation dataset. *International Journal of Computer Vision*, 129(7): 2175–2193.

Lv, F.; Lu, F.; Wu, J.; and Lim, C. 2018. MBLLEN: Low-light image/video enhancement using CNNs. In *British Machine Vision Conference*, 220.

Ma, K.; Zeng, K.; and Wang, Z. 2015. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11): 3345–3356.

Malik, S.; and Soundararajan, R. 2023. Semi-supervised learning for low-light image restoration through quality assisted pseudo-labeling. In *IEEE Winter Conference on Applications of Computer Vision*, 4105–4114.

Petschnigg, G.; Szeliski, R.; Agrawala, M.; Cohen, M.; Hoppe, H.; and Toyama, K. 2004. Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics*, 23(3): 664–672.

Pizer, S. M. 1990. Contrast-limited adaptive histogram equalization: Speed and effectiveness. In *Conference on Visualization in Biomedical Computing*, 337–345.

Pizer, S. M.; Amburn, E. P.; Austin, J. D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J. B.; and Zuiderveld, K. 1987. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3): 355–368.

Sakaridis, C.; Dai, D.; and Van Gool, L. 2021. ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *IEEE International Conference on Computer Vision*, 10765–10775.

Shen, L.; Yue, Z.; Feng, F.; Chen, Q.; Liu, S.; and Ma, J. 2017. MSR-net: Low-light image enhancement using deep convolutional network. *arXiv preprint arXiv:1711.02488*.

Wang, H.; Chen, Y.; Cai, Y.; Chen, L.; Li, Y.; Sotelo, M. A.; and Li, Z. 2022a. SFNet-N: An improved SFNet algorithm for semantic segmentation of low-light autonomous driving road scenes. *IEEE Transactions on Intelligent Transportation Systems*, 23: 21405–21417.

Wang, J.; Chan, K. C.; and Loy, C. C. 2023. Exploring CLIP for assessing the look and feel of images. In *AAAI Conference on Artificial Intelligence*, 2555–2563.

Wang, R.; Zhang, Q.; Fu, C.-W.; Shen, X.; Zheng, W.-S.; and Jia, J. 2019. Underexposed photo enhancement using deep illumination estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 6849–6857.

Wang, S.; Zheng, J.; Hu, H.-M.; and Li, B. 2013. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9): 3538–3548.

Wang, W.; Wei, C.; Yang, W.; and Liu, J. 2018. GLADNet: Low-light enhancement network with global awareness. In *IEEE International Conference on Automatic Face & Gesture Recognition*, 751–755.

Wang, X.; Xie, L.; Dong, C.; and Shan, Y. 2022b. Training real-world blind super-resolution with pure synthetic data. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 14071–14081.

Wang, Y.; Liu, Z.; Liu, J.; Xu, S.; and Liu, S. 2023. Low-light image enhancement with illumination-aware gamma correction and complete image modelling network. In *IEEE International Conference on Computer Vision*, 13128–13137.

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

Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*.

Wu, H.; Zhang, Z.; Zhang, W.; Chen, C.; Liao, L.; Li, C.; Gao, Y.; Wang, A.; Zhang, E.; Sun, W.; et al. 2023. Q-Align: Teaching LMMs for visual scoring via discrete text-defined levels. In *International Conference on Machine Learning*.

Xu, X.; Wang, R.; Fu, C.-W.; and Jia, J. 2022. SNR-aware low-light image enhancement. In *IEEE Conference on Computer Vision and Pattern Recognition*, 17714–17724.

Yang, M.; Nie, X.; and Liu, R. W. 2019. Coarse-to-fine luminance estimation for low-light image enhancement in maritime video surveillance. In *IEEE Intelligent Transportation Systems Conference*, 299–304.

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.

Yang, W.; Yuan, Y.; Ren, W.; Liu, J.; Scheirer, W. J.; Wang, Z.; and Zhang. 2020a. Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Transactions on Image Processing*, 29: 5737–5752.

Yang, W.; Yuan, Y.; Ren, W.; Liu, J.; Scheirer, W. J.; Wang, Z.; Zhang, T.; Zhong, Q.; Xie, D.; Pu, S.; et al. 2020b. Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Transactions on Image Processing*, 29: 5737–5752.

Ying, Z.; Li, G.; and Gao, W. 2017. A bio-inspired multi-exposure fusion framework for low-light image enhancement. *arXiv preprint arXiv:1711.00591*.

Zhang, F.; Tu, Z.; Hao, W.; Chen, Y.; Li, F.; and Ye, M. 2024. Zero-shot parameter learning network for low-light image enhancement in permanently shadowed regions. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–16.

Zhang, K.; Li, Y.; Liang, J.; Cao, J.; Zhang, Y.; Tang, H.; Fan, D.-P.; Timofte, R.; and Gool, L. V. 2023. Practical blind image denoising via Swin-Conv-UNet and data synthesis. *Machine Intelligence Research*, 20(6): 822–836.

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

Zhang, Y.; Guo, X.; Ma, J.; Liu, W.; and Zhang, J. 2021. Beyond brightening low-light images. *International Journal of Computer Vision*, 129: 1013–1037.

Zhang, Y.; Zhang, J.; and Guo, X. 2019. Kindling the darkness: A practical low-light image enhancer. In *ACM International Conference on Multimedia*, 1632–1640.

Zheng, S.; and Gupta, G. 2022. Semantic-guided zero-shot learning for low-light image/video enhancement. In *IEEE Winter Conference on Applications of Computer Vision*, 581–590.