This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

ZERO-IG: Zero-Shot Illumination-Guided Joint Denoising and Adaptive Enhancement for Low-Light Images

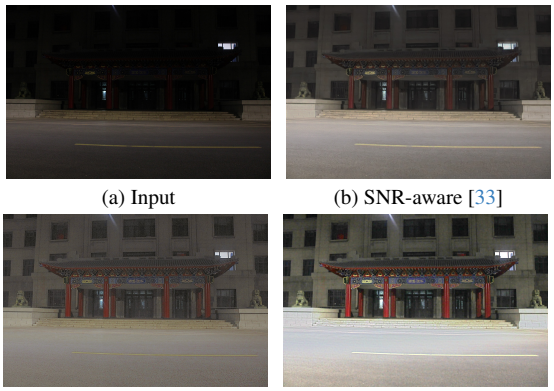
Yiqi Shi [1]*, Duo Liu [1]*, Liguo Zhang [1]†, Ye Tian [2], Xuezhi Xia [1,3], Xiaojing Fu [1]
[1] School of Computer Science and Technology, Harbin Engineering University
[2] Hangzhou Institute of Technology, Xidian University
[3] Wuhan Digital Engineering Research Institute
{shiyiqi, liu_duo, zhangliguo, fuxiaojing }@hrbeu.edu.cn,
tianye@xidian.edu.cn, xiaxuezhi709@gmail.com

Abstract

This paper presents a novel zero-shot method for jointly denoising and enhancing real-word low-light images. The proposed method is independent of training data and noise distribution. Guided by illumination, we integrate denoising and enhancing processes seamlessly, enabling end-to-end training. Pairs of downsampled images are extracted from a single original low-light image and processed to preliminarily reduce noise. Based on the smoothness of illumination, near-authentic illumination can be estimated from the denoised low-light image. Specifically, the illumination is constrained by the denoised image's brightness, uniformly amplifying pixels to raise overall brightness to normal-light level. We simultaneously restrict the illumination by scaling each pixel of the denoised image based on its intensity, controlling the enhancement amplitude for different pixels. Applying the illumination to the original low-light image yields an adaptively enhanced reflection. This prevents under-enhancement and localized overexposure. Notably, we concatenate the reflection with the illumination, preserving their computational relationship, to ultimately remove noise from the original low-light image in the form of reflection. This provides sufficient image information for the denoising procedure without changing the noise characteristics. Extensive experiments demonstrate that our method outperforms other state-of-the-art methods. The source code is available at https://github.com/Doyle59217/ZeroIG.

1. Introduction

Images taken in real low-light conditions typically exhibit a low signal-to-noise ratio (SNR) and contain underexposed regions. These images provide inadequate information and can impair computer vision tasks. Achieving high-quality low-light image enhancement (LLIE) requires improving brightness and contrast as well as effectively reducing noise. In the past decades, traditional methods like histogram equalization [14, 19] and gamma correction [9, 25] have been used to enhance low-light images. Additionally, various traditional methods [5, 17] are based on Retinex theory [12]. Retinex theory suggests that low-light observation can be decomposed into illuminated and reflected components. Illumination, influenced by ambient light, leads to low-light images formation. Reflection represents the image's intrinsic properties and is typically viewed as the target for enhancement. However, these hand-crafted constraints/priors are not adaptive enough and their results may suffer from under- and over- enhancement, in addition, they may present intensive noises.



(a) Input (b) SNR-aware [33] (c) PairLIE [6] (d) ZERO-IG

Figure 1. Visual comparison on a real night-time low-light image.

Recently, deep learning methods have known significant advancements [4, 18, 26, 30, 33, 38]. However, the effectiveness of these LLIE methods [21, 24, 27, 29, 31, 34, 36, 37] heavily rely on carefully selected paired/unpaired training data. Still, the low-light environments are highly

*Equal Contribution.
†Corresponding Author.

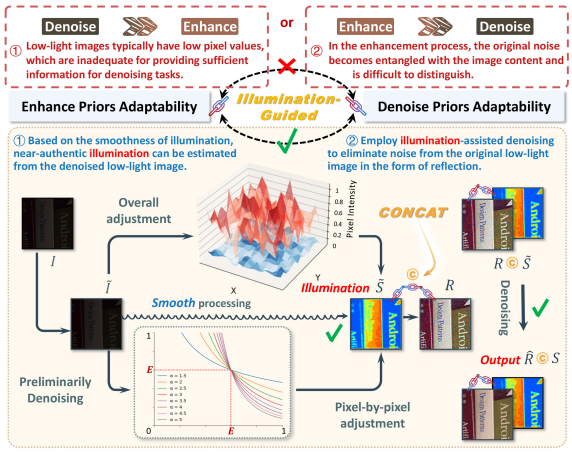


Figure 2. Illumination-guided: integrating denoising and enhancing. In the overall adjustment, X and Y represent pixel coordinates. Blue represents pixel intensity of the low-light image, while red represents adjusted intensity. Pixel-by-pixel adjustment shows curves with varying brightness coefficients. The horizontal axis signifies pre-adjustment pixel intensity, and the vertical axis post-adjustment intensity. Refer to Section 2 for symbol representation.

complex, with brightness levels varying significantly across images. Moreover, a single low-light image also exhibits uneven brightness distribution. Additionally, a substantial amount of sensor-specific noise, varying across different devices due to dark current and electronics shot in camera imaging, is introduced. Consequently, collecting training data to encompass all scenarios is impractical. This results in a performance drop for dataset-based methods when applied to test images from different datasets, especially in enhancing practical noisy low-light images. Meanwhile, some dataset-independent methods [6, 16, 20, 22, 35] tend to overlook noise reduction, leading to undesirable results.

In this paper, we propose a novel Retinex-based deep learning method for LLIE, termed **ZERO**-shot **I**llumination-**G**uided joint denoising and adaptive enhancement for low-light images (**ZERO-IG**). Leveraging intrinsic information of the real-world low-light image and constraining illumination, ZERO-IG achieves high-quality LLIE without requiring training data. Specifically, our network learns from each image individually and does not require pre-training. In contrast to cascade methods [36] that separate denoising from enhancement, our method uses illumination guidance to jointly denoise and enhance images, as shown in Figure 2. The integration of denoising and enhancement is based on the smoothness of illumination. Furthermore, we utilize illumination to achieve enhancing and aid denoising. Initially, we apply a preliminary denoising to the original low-light image. Inspired by ZS-N2N [23], we derive downsampled image pairs from the original and map them to mitigate noise. This allows ZERO-IG to be independent of noise distribution knowledge, making it applicable to diverse noise types. However, direct recovery of the enhanced image from the denoised version is not feasible. This limitation arises from the inherently low pixel values in low-light images, which yield insufficient detail for denoising. Consequently, the low-light image procured at this phase is not completely free of noise.

The denoised low-light image is suitable for estimating illumination. Reduced noise helps in accurately capturing light's propagation and reflection, resulting in a more precise estimation of illumination's statistical distribution. Additionally, illumination is characteristically smooth, with changes typically occurring continuously and gradually, rather than abruptly. Therefore, minor inaccuracies in estimated illumination can be naturally compensated by the inherent continuity of light, making them insignificant. In other words, the illumination derived from the preliminarily denoised low-light image can closely approximate that of a fully clean low-light image, enabling further enhancement.

We reference the average brightness of normal-light images, and proportionally scale up pixels in the denoised low-light images by constraining the illumination to boost overall brightness. However, such an overall adjustment may lead to under-enhancement in extremely dark regions and overexposure in already bright regions. Thus, we simultaneously constrain the illumination by scaling the denoised low-light image pixel-by-pixel based on pixel intensity. These result in an adaptive illumination that can control the enhancement amplitude of each pixel, enabling more enhancement in darker areas and less in brighter ones, while increasing the overall brightness, as shown in Figure 2. Adjusting the original low-light image using this illumination process yields the reflection with pixel-level enhancement. As a result, ZERO-IG can effectively enhance low-light images with varying brightness levels and uneven brightness distributions.

Still, the reflection contains noise. Direct denoising of the reflection, as other methods [36], would yield suboptimal result. This occurs as computational processes alter noise characteristics, causing them to become entwined with the image content. Such entanglement complicates the distinction between noise and actual image content. To address this problem, we innovatively use illumination to assist denoising. Specifically, we concatenate the reflection with the illumination, as shown in Figure 2. The key of this process is maintaining constant illumination before and after denoising. Given that the reflection is derived from the low-light image and the illumination, their computational relationship remains fixed. Keeping constant illumination can also be seen as eliminating noise from the original low-light image. Consequently, ZERO-IG supplies sufficient information for denoising via the reflection (brightened image), preserving the original noise characteristics and distribution, yielding the clean enhanced image. Additionally, we create a new dataset, VILNC, captured under real low-

light conditions. Figure 1 shows an example of the visual comparison on our VILNC dataset. See the Supplemental Materials for more details.

Our contributions are summarized as follows:

- We present a novel zero-shot method, ZERO-IG, for jointly denoising and enhancing real-world low-light images. Guided by illumination, we fully consider the coupled relationship between denoising and enhancing. Utilizing the inherent information of the low-light image, ZERO-IG attains high-quality LLIE without the necessity of any training data or knowledge of noise distribution.
- We employ illumination to adjust each pixel in the low-light image, achieving pixel-level adaptive enhancement. This prevents both under-enhancement and overexposure.
- We utilize illumination to assist in denoising. By concatenating the reflection with the illumination, we effectively remove noise from the original low-light image in the form of reflection.
- We create a new real-world dataset VILNC. Extensive experiments illustrate our superiority against other state-of-the-art methods.

2. Background

2.1. Retinex Theory

The Retinex theory [12], inspired by the Human Vision System, is an effective low-light image enhancement algorithm and can simulate human color perception. It decomposes observed images into reflected and illuminated components:

$$\hat{I} = \hat{R} \circ \hat{S} \tag{1}$$

The classical Retinex model overlooks the impact of noise on images. Therefore, $\hat{I}$ is regarded as a noise-free low-light image. $\hat{R}$ and $\hat{S}$ denote the reflection (also the enhanced image) and the illumination, both immune to noise. $\circ$ symbolizes element-wise multiplication.

However, real-world low-light images often have significant noise due to inadequate lighting and flaws in the imaging system. M. Li et al. [17] introduced a noise term N into the classic Retinex model, resulting in $I = \hat{I} + N = \hat{R} \circ \hat{S} + N$, where I represents a noisy low-light image. But the typically low pixel values in low-light images make accurately separating the exact noise N a challenge.

Since illumination determines the image's dynamic range, remaining uncontaminated by noise. The reflection, representing the image's inherent properties, frequently contains noise during the imaging process. The equation can be rewritten as $I = (\hat{R}+n) \circ \hat{S} = R \circ \hat{S}$, where R is the reflection tainted with noise n. While the reflection offers valuable information for denoising, its calculation can alter the noise's characteristics. The noise is more intertwined with the image content and harder to remove.

We adopt a different strategy by integrating the aforementioned equations:

$$I = \hat{R} \circ \hat{S} + N = R \circ \hat{S} \tag{2}$$

Our objective is to estimate the illumination $\hat{S}$. Next, using the reflection R to remove noise N from the original low-light image I, producing the enhanced image $\hat{R}$. This provides ample image information for denoising while maintaining the original noise's characteristics. Additionally, it creates a coupled relationship between denoising and enhancing. The details will be discussed in Section 3.

2.2. Noise2Noise and ZS-N2N

Noise2Noise [15] is a denoising method that does not require clean ground truth images. It only needs two independently noisy images of the same scene. With two independent noisy observations I_1 and I_2 of the same ground truth $\hat{I}$, Noise2Noise suggests that mapping I_1 to I_2 is analogous to supervised mapping to a clean image $\hat{I}$.

ZS-N2N [23] extends Noise2Noise [15], enabling training with only one noisy image. It decomposes a noisy image I into two downsampled images $G_1(I)$ and $G_2(I)$, considering them as equivalents for two noisy observations of the same scene. And training a denoising network by mapping $G_1(I)$ to $G_2(I)$.

Inspired by ZS-N2N [23], we also demonstrate our method's independence from noise distribution knowledge. Contrasting with methods rely on camera-specific datasets, our method is effective in scenarios with unknown noise distribution or levels, making it suitable for situations with scarce data.

3. Proposed Method

As shown in Figure 3, ZERO-IG comprises three subnetworks: the low-light image denoising network (LD-Net), the illumination estimation network (IE-Net), and the reflection denoising network (RD-Net). Our method employs illumination to sequentially connect these subnetworks. The IE-Net receives input from the LD-Net. The estimated illumination is initially used to restore the noise-affected reflection, which is then fed into the RD-Net. And the entire network is trained end-to-end, integrating denoising and enhancing effectively.

3.1. Low-light Image Denoising Network

Although illumination is not contaminated by noise, the presence of noise can significantly affect its statistical distribution. Directly using the original noisy low-light image I for illumination estimation may lead to inaccuracies. Consequently, we first introduce the LD-Net to preliminarily denoise I. This results in a more suitable low-light image $\tilde{I}$ to estimate illumination.

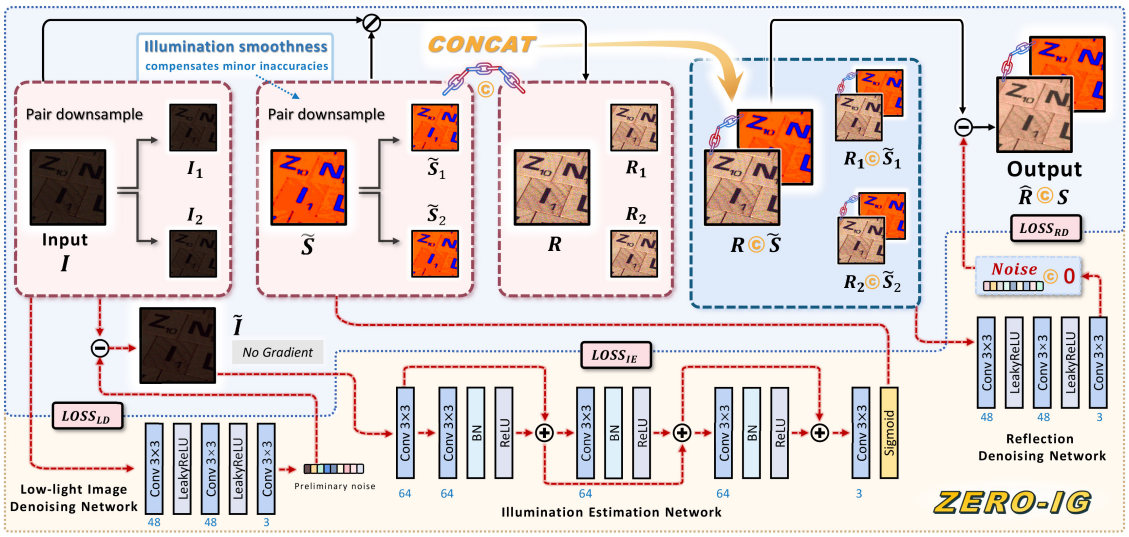


Figure 3. The framework of ZERO-IG. It comprises three subnetworks: the LD-Net, the IE-Net, and the RD-Net. The original low-light image I is initially input into the LD-Net. The IE-Net estimate the illumination $\tilde{S}$ using the preliminarily denoised low-light image $\tilde{I}$. $\tilde{S}$ is applied to I to calculate the reflection R. The RD-Net denoises the concatenated R and $\tilde{S}$, producing the final enhanced image $\hat{R}$. Notably, the RD-Net directly outputs a concatenation of noise and an all-zero pixel array. The noise matches R in size and channel number, while the all-zero array aligns with $\tilde{S}$ in the same dimensions. All downsampled images are used exclusively for training. The entire network is a straightforward CNN of eleven convolutional layers, each layer consists of 48 or 64 convolutional kernels of size 3×3.

Inspired by ZS-N2N [23], we downsample I of size H×W×C through operation $G = (G_1, G_2)$. Creating images $I_1 = G_1(I)$ and $I_2 = G_2(I)$ each of size H/2×W/2×C. The operation divides I into non-overlapping 2×2 patches, averaging diagonal pixels for I_1 and anti-diagonal pixels for I_2. Subsequently, we train the LD-Net using I_1 and I_2.

We express residual loss $\mathcal{L}_{\text{res}}(\boldsymbol{\theta})$ through symmetric loss:

$$\mathcal{L}_{\text{res}}(\boldsymbol{\theta}) = \|I_1 - f_{\boldsymbol{\theta}}(I_1) - I_2\|_2^2 + \|I_2 - f_{\boldsymbol{\theta}}(I_2) - I_1\|_2^2 \quad (3)$$

where $f_{\boldsymbol{\theta}}()$ represents the noise. We also express the consistency loss $\mathcal{L}_{\text{cons}}(\boldsymbol{\theta})$ using symmetric loss:

$$\begin{aligned}\mathcal{L}_{\text{cons}}(\boldsymbol{\theta}) = &\|I_1 - f_{\boldsymbol{\theta}}(I_1) - G_1(I - f_{\boldsymbol{\theta}}(I))\|_2^2 + \\ &\|I_2 - f_{\boldsymbol{\theta}}(I_2) - G_2(I - f_{\boldsymbol{\theta}}(I))\|_2^2\end{aligned} \quad (4)$$

The loss function for the LD-Net is defined as $\mathcal{L}_{\text{LD}} = \mathcal{L}_{\text{res}}(\boldsymbol{\theta}) + \mathcal{L}_{\text{cons}}(\boldsymbol{\theta})$. It is noteworthy that due to the low SNR, I lacks sufficient information for effective denoising. Although the output $\tilde{I}$ of the LD-Net has considerably less noise, it cannot be directly used to restore the enhanced image $\hat{R}$. As $\tilde{I}$ is not the completely clean low-light image $\hat{I}$ depicted in Eq. 1.

3.2. Illumination Estimation Network

In the IE-Net, we take $\tilde{I}$ as input to generate the illumination $\tilde{S}$. Given that the smoothness of illumination can compensate minor errors, $\tilde{S}$ can approximate the noise-unaffected $\hat{S}$ estimated from $\hat{I}$. Thus, $\tilde{S}$ can be used for enhancement.

The main problem with low-light images is the low brightness and poor visibility. It is essential to increase the overall brightness of the low-light image to match normal-light levels. Specifically, we constrain the illumination by image brightness to proportionally upscale all pixels in $\tilde{I}$. The overall adjustment loss $\mathcal{L}_{\text{over}}$ can be expressed as:

$$\mathcal{L}_{\text{over}} = \|\tilde{S} - \alpha^{-1}\|_2^2 \quad (5)$$

where the brightness coefficient $\alpha = Y_H {Y_L}^{-1}$. Y_H represents the mean value of the luminance plane Y of normal-light images. We statistically set Y_H to 0.5, see the Supplemental Materials for details. Y_L indicates the mean value of the luminance plane Y of $\tilde{I}$.

The current adjusted image can be expressed as $\tilde{I} \circ \tilde{S}^{-1} = \alpha \tilde{I}$. As our method is designed for low-light images, $\alpha > 1$. This process increases overall brightness and contrast of $\tilde{I}$. However, low-light images contain regions with varying brightness, each needing a different degree of enhancement. Uniformly enhancing all pixels may result in under-enhancement in darker areas or overexposure in brighter areas (light sources).

We additionally constrain the illumination by the intensity of each pixel in $\tilde{I}$. This achieves different amplitude enhancements for pixels with differing intensities, further enhancing extremely dark regions while preventing overexposure in brighter areas. Inspired by Gamma correction [3], we establish a relationship between $\tilde{S}$ and $\tilde{I}$. The pixel-by-pixel adjustment loss $\mathcal{L}_{\text{pix}}$ can be expressed as:

$$\mathcal{L}_{\text{pix}} = \|\tilde{S} - \beta(\alpha\tilde{I})^{\alpha}\|_2^2 \quad (6)$$

where α (as in Eq. 5) facilitates pixel-level scaling according to the brightness of $\tilde{I}$. We additionally incorporate a scaling factor β that controls the scaling degree and dictates the level of contrast enhancement.

The image, post per-pixel adjustment, can be represented as $\tilde{I}(x,y) \circ \tilde{S}(x,y)^{-1} = \tilde{I}(x,y) \circ (\beta(\alpha \tilde{I}(x,y))^{\alpha})^{-1}$, where (x,y) denotes pixel coordinates. Pixels with well-exposedness level E after overall adjustment should remain unaltered following per-pixel adjustment. Specifically, it implies $E = (\alpha^{-1}E) \circ (\beta E^{\alpha})^{-1}$, therefore $\beta = \alpha^{-1}E^{-\alpha}$. We empirically set $E = 0.7$, see the Supplemental Materials, resulting in $\beta = \alpha^{-1}0.7^{-\alpha}$. This indicates that both α and β depend on the brightness of $\tilde{I}$. Thus, our method achieves pixel-level adaptive enhancement while improving the brightness and contrast of low-light images based on their brightness and pixel intensity.

Illumination has smoothness properties. Variations in illumination in natural images are typically continuous and smooth, able to offset minor errors. To balance global and local changes, we introduce the smoothness loss $\mathcal{L}_{\text{smooth}}$:

$$\mathcal{L}_{\text{smooth}} = \sum_c (|\nabla_x \tilde{S}_c| + |\nabla_y \tilde{S}_c|)^2 + \sum_{i=1}^{N} \sum_j w_{i,j} |\tilde{S}_i - \tilde{S}_j| \quad (7)$$

where $c \in \{R, G, B\}$, with horizontal and vertical gradient operations denoted by ∇_x and ∇_y respectively. N is the total number of pixels, with i signifying the i^{th} pixel. $j \in \mathcal{N}(i)$ indicates the neighboring pixels of i within a 5×5 window. The weight $w_{i,j}$ is determined using a Gaussian kernel function, which is based on the difference between pixels in the YUV color space.

The loss function for the IE-Net is defined as $\mathcal{L}_{\text{IE}} = \mathcal{L}_{\text{over}} + \mathcal{L}_{\text{pix}} + \mathcal{L}_{\text{smooth}}$. $\mathcal{L}_{\text{IE}}$ enables our IE-Net to estimate pixel-level adaptive illumination for low-light images with varying or uneven brightness. And can obtain $\tilde{S}$ to approximate noise-unaffected illumination $\hat{S}$. Substituting $\tilde{S}$ and the original low-light image I into Eq. 2 results in the adaptively enhanced reflection R, which contains noise.

3.3. Reflection Denoising Network

In contrast to the LD-Net, the RD-Net utilizes the estimated illumination $\tilde{S}$ to aid denoising. We concatenate the reflection R with $\tilde{S}$ and feed them into the RD-Net to get the final noise-free enhanced image $\hat{R}$.

Initially, the downsampling operation $G = (G_1, G_2)$ is applied to $\tilde{S}$, yielding $\tilde{S}_1 = G_1(\tilde{S})$ and $\tilde{S}_2 = G_2(\tilde{S})$. Substituting I_1, I_2 and $\tilde{S}_1$, $\tilde{S}_2$ into Eq. 2 to calculate the corresponding noisy reflections R_1 and R_2. We concatenate R_1 with $\tilde{S}_1$ and R_2 with $\tilde{S}_2$ to train the RD-Net.

The residual loss $\mathcal{L}_{\text{res}}(\hat{\boldsymbol{\theta}})$ is expressed as:

$$\begin{aligned} \mathcal{L}_{\text{res}}(\hat{\boldsymbol{\theta}}) = &\| R_1 \copyright \tilde{S}_1 - f_{\hat{\boldsymbol{\theta}}}(R_1 \copyright \tilde{S}_1) - R_2 \copyright \tilde{S}_2 \|_2^2 + \\ &\| R_2 \copyright \tilde{S}_2 - f_{\hat{\boldsymbol{\theta}}}(R_2 \copyright \tilde{S}_2) - R_1 \copyright \tilde{S}_1 \|_2^2 \end{aligned} \quad (8)$$

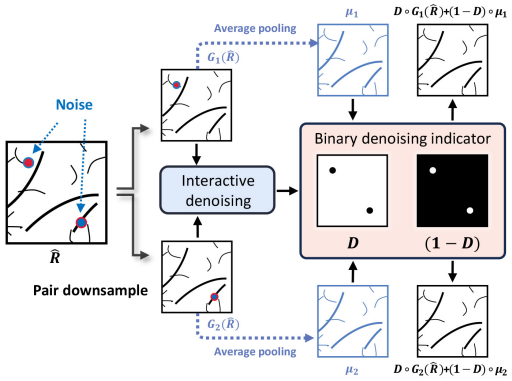


Figure 4. Depiction of the interactive denoising loss $\mathcal{L}_{\text{inter}}$. D acts as a binary denoising indicator, ranging from 0 to 1. Higher D values indicate smaller differences between $G_1(\hat{R})$ and $G_2(\hat{R})$. Values above the 0.975 threshold are classified as 1 (indicated in white), while others are 0 (indicated in black).

where © symbolizes the concatenation operation. $f_{\hat{\boldsymbol{\theta}}}()$ denotes the concatenation of noise and illumination.

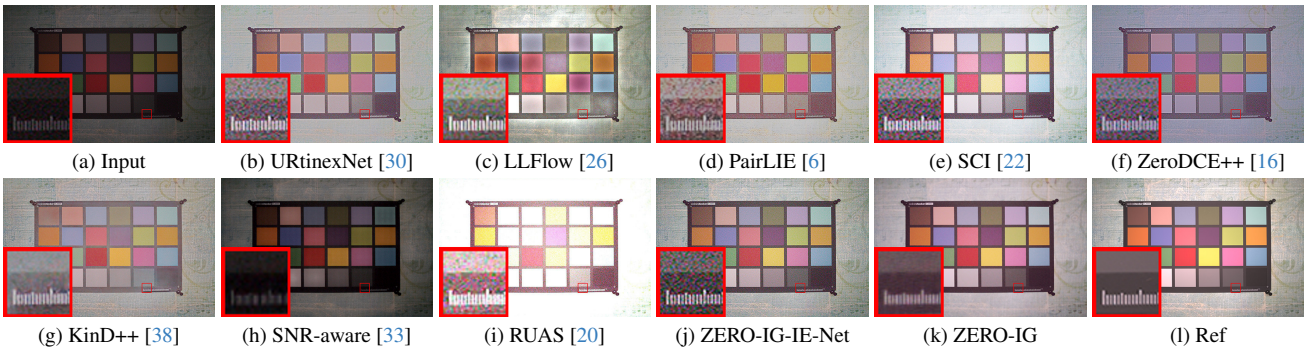
We also apply constraints to the final enhanced image $\hat{R}$. Employing the same downsampling to derive $G_1(\hat{R})$ and $G_2(\hat{R})$. $\hat{R}_1$ and $\hat{R}_2$ denote the denoising results for R_1 and R_2. The consistency loss $\mathcal{L}_{\text{cons}}$ is formulated as:

$$\mathcal{L}_{\text{cons}} = \|G_1(\hat{R}) - \hat{R}_1\|_2^2 + \|G_2(\hat{R}) - \hat{R}_2\|_2^2 \quad (9)$$

The key of the concatenation operation is the imposed illumination consistency loss $\mathcal{L}_{\text{ill}}$. It is applied to the final output illumination S concatenated with $\hat{R}$, as shown in Figure 3, and the illumination $\tilde{S}$ concatenated with R:

$$\mathcal{L}_{\text{ill}} = \|S - \tilde{S}\|_2^2 \quad (10)$$

Why does concatenation work? To train the RD-Net, we map one noisy concatenation to another (see Eq. 8). In simpler terms, one output is constrained to be equal to another input. This maintains the computational relationship between the reflection and the illumination from input to output. As outlined in Section 3.2, $R = I \circ \tilde{S}^{-1}$. The computational relationship linking R with $\tilde{S}$ mirrors that between $\hat{R}$ and S. Consequently, $\hat{R}$ can also be expressed as $\hat{R} = \bar{I} \circ S^{-1}$. Consider $\bar{I}$ for now as a component of $\hat{R}$. Based on Eq. 1, we can also derive $\hat{R} = \hat{I} \circ \hat{S}^{-1} = \hat{I} \circ \tilde{S}^{-1}$. Crucially, the illumination is constrained to keep constant, i.e., $S = \tilde{S}$ (as Eq. 10). This leads to $\bar{I} = \hat{I}$. It implies that $R \copyright \tilde{S} \rightarrow \hat{R} \copyright S$ can be viewed as $(I \circ \tilde{S}^{-1}) \copyright \tilde{S} \rightarrow (\hat{I} \circ S^{-1}) \copyright S$. Consequently, the RD-Net converts the original low-light image I into the clean version $\hat{I}$, effectively eliminating noise N (as Eq. 2). The actual input of the RD-Net is the reflection R, and the final output is the enhanced image $\hat{R}$. Essentially, through the concatenation operation, the RD-Net harnesses the form of the bright image R to provide sufficient information for denoising. While preserving



(a) Input (b) URtinexNet [30] (c) LLFlow [26] (d) PairLIE [6] (e) SCI [22] (f) ZeroDCE++ [16]

(g) KinD++ [38] (h) SNR-aware [33] (i) RUAS [20] (j) ZERO-IG-IE-Net (k) ZERO-IG (l) Ref

Figure 5. Visual comparison on the real-world low-light image from the SIDD [1] dataset.

the characteristics and distribution of the original noise N, achieving effective noise reduction. This further illustrates that our method uses illumination as guidance, establishing the connection between denoising and enhancing.

As shown in Figure 4, to further augment the RD-Net's ability of denoising, the interactive denoising loss $\mathcal{L}_{\text{inter}}$ is presented as:

$$\mathcal{L}_{\text{inter}} = \sum_{i=1,2} \|G_i(\hat{R}) - (D \circ G_i(\hat{R}) + (1-D) \circ \mu_i)\|_2^2 \quad (11)$$

where μ_i represents the average of all pixel values within a 5×5 window at the respective position in $G_i(\hat{R})$. D signifies the disparity in luminance channels between $G_1(\hat{R})$ and $G_2(\hat{R})$. Inspired by DeSRA [32], we define $D = 2\sigma_1\sigma_2(\sigma_1^2 + \sigma_2^2 + C)^{-1}$. σ_i represents the standard deviation of the pixels within a 5×5 window at the corresponding positions in $G_i(\hat{R})$'s luminance channel. A constant C is added to stabilize the division with a weak denominator.

We assume that for a sufficiently clean image, its two downsampled versions should be nearly identical. Consequently, D is employed to pinpoint discrepancies between $G_1(\hat{R})$ and $G_2(\hat{R})$. Points with significant differences are likely indicative of noise. Since adjacent pixels in a clean image are highly correlated and typically have similar values, we adjust these points to approach the average value of their 5×5 window.

For the ideal clean image, a small local area can be considered as a constant, with its variance nearly zero. The noisy image's variance equals the sum of variances of the noise and the clean image. To further purifies the clean image, the local variance loss $\mathcal{L}_{\text{var}}$ is formulated as follows:

$$\mathcal{L}_{\text{var}} = \frac{1}{M} \| \sum_{j=1}^{M} \sigma_j^2 - \sum_{j=1}^{M} \hat{\sigma}_j^2 \|_2^2 \quad (12)$$

where M denotes the number of 5×5 windows. σ_j^2 and $\hat{\sigma}_j^2$ represent variances of the j^{th} window in R and n, where $n = R - \hat{R}$.

Finally, inspired by DSLR [10], we introduce an extra color loss $\mathcal{L}_{\text{color}}$:

$$\mathcal{L}_{\text{color}} = \|R_b - \hat{R}_b\|_2^2 \quad (13)$$

where R_b and $\hat{R}_b$ represent the blurred versions of R and $\hat{R}$. $R_b(i,j) = \sum_{k,l} R(i+k, j+l) \cdot G(k,l)$ and $G(k,l)$ is a 2D Gaussian blur operator.

The loss function for the RD-Net is defined as $\mathcal{L}_{\text{RD}} = \mathcal{L}_{\text{res}}(\hat{\boldsymbol{\theta}}) + \mathcal{L}_{\text{cons}} + \mathcal{L}_{\text{ill}} + \mathcal{L}_{\text{inter}} + \mathcal{L}_{\text{var}} + \mathcal{L}_{\text{color}}$. Since our method is trained end-to-end, the total loss function is defined as $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{LD}} + \mathcal{L}_{\text{IE}} + \mathcal{L}_{\text{RD}}$.

4. Experiments

4.1. Implementation Details

Parameter Settings. All experiments are implemented with PyTorch on a single Nvidia Titan X Pascal GPU. The ADAM optimizer [11] is employed, with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The batch size is set to 1. The learning rate is fixed to 10^{-3}. The training epoch number is set to 2000.

Compared Methods. We compare our method with several state-of-the-art LLIE methods: four supervised learning methods (URetinexNet [30], LLFlow [26], SNR-aware [33], and KinD++ [38]), and four unsupervised learning methods (SCI [22], ZeroDCE++ [16], PairLIE [6], and RUAS [20]). The results are reproduced using publicly available source codes with recommended parameters.

Benchmarks Description and Metrics. For testing, we use 15 validation images from the LOL [28] dataset, 50 validation images from the LSRW [8] dataset (30 by Huawei, 20 by Nikon), and 22 randomly sampled images with the attribute of "Lowlight" from SIDD-Small [1] dataset. Notably, we adjust the SIDD-Small dataset's image resolution to 1280x720. We employ two commonly-used metrics, PSNR and SSIM. In addition, for visual comparisons, we use the LIME [7] and DICM [13] datasets.

4.2. Benchmark Evaluations

Performance Evaluation. Figure 5 displays a visual comparison of images taken in real low-light conditions. Our

Table 1. Quantitative comparisons in terms of PSNR and SSIM. The best and the second best results are highlighted in **red** and **blue**.

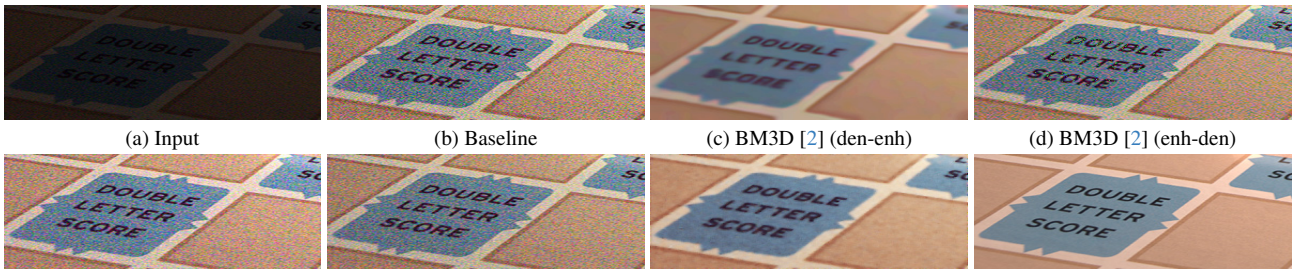
Datasets	Type	Metrics	*Supervised Learning Methods* URtinexNet (cvpr2022)	LLFlow (AAAI2022)	SNR-aware (cvpr2022)	KinD++ (IJCV2021)	*Unsupervised Learning Methods* SCI (cvpr2022)	ZeroDCE++ (TPAMI2021)	PairLIE (cvpr2023)	RUAS (cvpr2021)	ZERO-IG IE-Net	ZERO-IG
SIDD	-	PSNR ↑	16.2600	14.6107	14.8907	16.5524	15.5257	15.6757	**17.0254**	12.5997	15.1154	**18.9849**
		SSIM ↑	0.4247	0.4401	**0.6010**	0.5799	0.3537	0.3561	0.5266	0.4287	0.2865	**0.6253**
	Followed by ZS-N2N	PSNR ↑	16.5879	**17.2651**	14.8641	16.6056	15.7053	16.4329	17.1241	12.7437	16.0232	**18.9849**
		SSIM ↑	0.4715	0.5097	**0.5994**	0.5982	0.3610	0.45883	0.5497	0.4459	0.4577	**0.6253**
LOL	-	PSNR ↑	20.1405	**24.0641**	**24.6977**	17.6476	14.7839	15.1416	18.4684	16.5976	17.6255	22.1751
		SSIM ↑	0.8221	**0.8601**	**0.8494**	0.7714	0.52544	0.5657	0.7426	0.6559	0.4566	0.7719
LSRW-Huawei	-	PSNR ↑	18.1566	**19.2005**	17.6209	17.0251	15.7003	16.3821	18.9887	15.7422	17.6842	**19.8414**
		SSIM ↑	0.5464	0.5419	**0.5781**	0.4993	0.4279	0.4696	0.5502	0.4976	0.4101	**0.5944**
LSRW-Nikon	-	PSNR ↑	**15.9870**	15.3675	15.9362	15.4796	14.5542	15.2770	15.5214	12.2104	15.3901	**16.6157**
		SSIM ↑	0.4425	0.4491	**0.4691**	0.4411	0.4065	0.4129	0.4271	0.4394	0.3818	**0.4706**

(a) Input (b) SNR-aware [33] (c) RUAS [20] (d) SCI [22] (e) PairLIE [6] (f) LLFlow [26] (g) ZERO-IG

Figure 6. Visual comparisons on low-light images with uneven brightness from LIME [7], DICM [13], LOL [28] and LSRW [8] datasets.

method excels over other methods in image brightness, contrast, color fidelity, and noise reduction. Although to be a zero-shot method, we get more natural visual quality and closer to the reference. Even when using only our IE-Net for enhancement, we yield better result than others. This is because they heavily rely on training data, typically using images with negligible noise. Real noise in low-light images pollutes their enhanced results and degrades their performance, leading to inferior visual effects. We also perform comparisons on uneven brightness low-light images. Figure 6 demonstrates that our method effectively enhances darker regions without overexposing the brighter areas, contrasting with other methods that fail in this regard.

In quantitative comparisons, our method outperforms others on both SIDD [1] and LSRW [8] datasets, as indicated in Table 1. In particular, the versatility of our method is demonstrated by its strong performance across various shooting devices, as shown on the LSRW [8] dataset. On the LOL [28] dataset, we surpass all unsupervised methods and are comparable to supervised methods. Due to space



(a) Input (b) Baseline (c) BM3D [2] (den-enh) (d) BM3D [2] (enh-den)

(e) ZS-N2N [23] (den-enh) (f) ZS-N2N [23] (enh-den) (g) ZERO-IG (h) Ref

Figure 7. Visual comparison with cascaded methods on the SIDD [1] dataset.

Table 2. Ablation study of the contribution of the three subnetworks and concatenation operation. The best and the second best results are highlighted in **red** and **blue**.

Subnetworks				PSNR ↑	SSIM ↑
The IE-Net	The LD-Net	The RD-Net	Concat-enation		
✓	✗	✗	✗	18.2091	0.4134
✓	✓	✗	✗	18.2123	0.4143
✓	✗	✓	✓	18.1160	0.5580
✓	✓	✓	✗	**19.4880**	**0.6284**
✓	✓	✓	✓	**23.2349**	**0.6703**

Table 3. Ablation study of the contribution of loss terms in the RD-Net. The best and the second best results are highlighted in **red** and **blue**.

The RD-Net Losses				PSNR ↑	SSIM ↑
$\mathcal{L}_{\text{ill}}$	$\mathcal{L}_{\text{inter}}$	$\mathcal{L}_{\text{var}}$	$\mathcal{L}_{\text{color}}$		
✗	✓	✓	✓	19.1476	0.5442
✓	✗	✓	✓	18.0733	0.5000
✓	✓	✗	✓	**19.3852**	0.5785
✓	✓	✓	✗	19.3372	**0.6131**
✓	✓	✓	✓	**23.2349**	**0.6703**

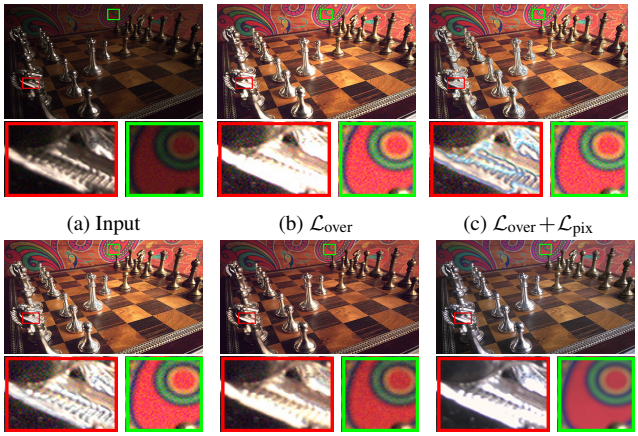
limitations, more comparisons are available in the Supplemental Materials.

Evaluation of Joint Enhancement and Denoising. To highlight the significance of joint denoising and enhancing, we initially denoise the enhanced images of the mentioned methods using ZS-N2N [23]. Then we compare the results with ours. Table 1 shows that, while these methods exhibit slight improvement post-denoising on the SIDD [1] dataset, they are still significantly less effective than ours. To further compare ZERO-IG with cascade methods, we incorporate ZS-N2N [23] and BM3D [2] for denoising. Specifically, we explore two scenarios: denoising then enhancing (den-enh) and enhancing then denoising (enh-den). The enhancement result of our IE-Net is used as the baseline. Figure 7 indicates that cascade networks either failed in denoising or over-smoothed the image details. Conversely, our joint method successfully preserves details and effectively suppresses noise.

4.3. Ablation Study

Table 2 shows the results of various combinations of the three subnetworks. Solely using the IE-Net, or introducing the LD-Net/RD-Net before/after enhancement, produces results inferior to the integrated of all three subnetworks. This confirms the importance of our proposed joint denoising and enhancing approach. Moreover, removing the concatenation operation leads to a substantial decrease in metrics.

Subsequently, we assess the results of training the IE-Net with different loss combinations. As shown in Figure 8, the reconstruction loss $\mathcal{L}_{\text{over}}$ increases the brightness of the low-light image. While leading to localized overexposure, evident in the enlarged area within the red frame. Adding the magnitude control loss $\mathcal{L}_{\text{pix}}$ enables adaptive enhancement, addressing the overexposure problem but resulting in overly sharp edges. The inclusion of smoothness loss $\mathcal{L}_{\text{smooth}}$ further improves the texture and details, yielding a more natural visual appearance.



(a) Input (b) $\mathcal{L}_{\text{over}}$ (c) $\mathcal{L}_{\text{over}}+\mathcal{L}_{\text{pix}}$

(d) $\mathcal{L}_{\text{over}}+\mathcal{L}_{\text{pix}}+\mathcal{L}_{\text{smooth}}$ (e) ZERO-IG (f) Ref

Figure 8. Ablation study of the contribution of loss terms in the IE-Net.

Finally, we evaluate the losses in the denoising networks. Figure 8(e) illustrates that incorporating all denoising losses effectively reduce noise, particularly in the areas magnified in the green and red frames. This yields an enhanced image closer to the reference. Table 2 confirms the effectiveness of residual loss $\mathcal{L}_{\text{res}}$ and consistency loss $\mathcal{L}_{\text{cons}}$ in the LD-Net. Therefore, we focus solely on evaluating other loss terms in the RD-Net. As shown in Table 3, omitting any loss term results in reduced performance metrics. The most significant decline is observed when $\mathcal{L}_{\text{inter}}$ is excluded, underscoring its critical role. Additionally, eliminating $\mathcal{L}_{\text{ill}}$ also leads to a notable decrease in performance, reaffirming the significance of the concatenation operation. More ablation experiments are included in the Supplemental Materials.

5. Conclusion

This paper proposes ZERO-IG, a method that jointly denoises and enhances real-world low-light images. ZERO-IG is guided by illumination, requiring only a single low-light image for training without needing prior knowledge of noise distribution. We employ the form of pixel-level adaptive enhanced reflection to remove original noise from the low-light image. This avoids under-enhancement and over-exposure while preserving the noise's original characteristics for effective enhancement. We also create a new dataset VILNC captured in real low-light conditions. Both qualitative and quantitative experiments demonstrate the superiority of ZERO-IG against existing advanced methods.

Acknowledgments: This work was supported by the National Key R&D Program of China (2021YFC3320302) and the National Sponsored Postdoctoral Researcher Program (GZC20232046).

References

[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1692–1700, 2018. 6, 7, 8

[2] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 7, 8

[3] Hany Farid. Blind inverse gamma correction. *IEEE transactions on image processing*, 10(10):1428–1433, 2001. 4

[4] Huiyuan Fu, Wenkai Zheng, Xiangyu Meng, Xin Wang, Chuanming Wang, and Huadong Ma. You do not need additional priors or regularizers in retinex-based low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18125–18134, 2023. 1

[5] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2782–2790, 2016. 1

[6] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22252–22261, 2023. 1, 2, 6, 7

[7] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 6, 7

[8] Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, and Songchen Han. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation*, 90: 103712, 2023. 6, 7

[9] Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE transactions on image processing*, 22(3):1032–1041, 2012. 1

[10] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 3277–3285, 2017. 6

[11] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[12] Edwin H Land and John J McCann. Lightness and retinex theory. *Josa*, 61(1):1–11, 1971. 1, 3

[13] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, 22(12): 5372–5384, 2013. 6, 7

[14] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, 22(12): 5372–5384, 2013. 1

[15] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018. 3

[16] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4225–4238, 2021. 2, 6

[17] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018. 1, 3

[18] Xiwen Liang, Xiaoyan Chen, Keying Ren, Xia Miao, Zhihui Chen, and Yutao Jin. Low-light image enhancement via adaptive frequency decomposition network. *Scientific Reports*, 13(1):14107, 2023. 1

[19] Jiaying Liu, Dejia Xu, Wenhan Yang, Minhao Fan, and Haofeng Huang. Benchmarking low-light image enhancement and beyond. *International Journal of Computer Vision*, 129:1153–1184, 2021. 1

[20] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10561–10570, 2021. 2, 6, 7

[21] Yucheng Lu and Seung-Won Jung. Progressive joint low-light enhancement and noise removal for raw images. *IEEE Transactions on Image Processing*, 31:2390–2404, 2022. 1

[22] 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. 2, 6, 7

[23] Youssef Mansour and Reinhard Heckel. Zero-shot noise2noise: Efficient image denoising without any data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14018–14027, 2023. 2, 3, 4, 7, 8

[24] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6849–6857, 2019. 1

[25] Wei Wang, Na Sun, and Michael K Ng. A variational gamma correction model for image contrast enhancement. *Inverse Problems and Imaging*, 13(3):461–478, 2019. 1

[26] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI conference on artificial intelligence*, pages 2604–2612, 2022. 1, 6, 7

[27] Yinglong Wang, Zhen Liu, Jianzhuang Liu, Songcen Xu, and Shuaicheng Liu. Low-light image enhancement with illumination-aware gamma correction and complete image modelling network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13128–13137, 2023. 1

[28] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 6, 7

[29] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 1

[30] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5901–5910, 2022. 1, 6

[31] Yuhui Wu, Chen Pan, Guoqing Wang, Yang Yang, Jiwei Wei, Chongyi Li, and Heng Tao Shen. Learning semantic-aware knowledge guidance for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1662–1671, 2023. 1

[32] Liangbin Xie, Xintao Wang, Xiangyu Chen, Gen Li, Ying Shan, Jiantao Zhou, and Chao Dong. Desra: Detect and delete the artifacts of gan-based real-world super-resolution models. 2023. 6

[33] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17714–17724, 2022. 1, 6, 7

[34] Xiaogang Xu, Ruixing Wang, and Jiangbo Lu. Low-light image enhancement via structure modeling and guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9893–9903, 2023. 1

[35] Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. Implicit neural representation for cooperative low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12918–12927, 2023. 2

[36] Feng Zhang, Yuanjie Shao, Yishi Sun, Kai Zhu, Changxin Gao, and Nong Sang. Unsupervised low-light image enhancement via histogram equalization prior. *arXiv preprint arXiv:2112.01766*, 2021. 1, 2

[37] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019. 1

[38] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129:1013–1037, 2021. 1, 6