# Unsupervised Low-light Image Enhancement with Generated Low-light Image Pairs

Yuping Xia [1,2][0009-0004-2585-8350], Fan Ji[1,2][0009−0007−3068−4158], Xiongxin Tang[1,*][0009−0000−7037−3097], and Fanjiang Xu[1]

[1] Institute of Software, Chinese Academy of Sciences, Beijing, China
[2] University of Chinese Academy of Sciences, Beijing, China
`{xiayuping2022, jifan2022, xiongxin, fanjiang}@iscas.ac.cn`

**Abstract.** Low-light image enhancement aims to improve the perception of images captured under low-light conditions. Many previous unsupervised methods rely solely on information from a single image and multiple priors for enhancement. However, the information within a single image is limited, and designing suitable priors could be challenging. Although some existing methods have addressed this using paired low-light images, obtaining such images is also intricate. To tackle this problem, we first generate paired low-light images with consistent content, noise independence, and slightly different illumination from a single low-light image. Second, we propose an unsupervised low-light enhancement network based on the paired images. Leveraging consistent image content, we establish mutual constraints between the two images to achieve identical enhancement results. To accomplish this, the Retinex theory is employed to decompose the images into illumination and reflectance components, ensuring consistency in the reflectance components of the two images, which facilitates the preservation of image content and avoids artifacts and color deviations. Moreover, as the images exhibit independent noise, we adopt the Noise2Noise for noise removal, ensuring comprehensive denoising without affecting image details, which further improves the quality of the enhanced images. Extensive experiments demonstrate that our approach has superior denoising capability while ensuring enhancement performance, and achieves results comparable to state-of-the-art methods.

**Keywords:** Low-light Image Enhancement, Unsupervised, Generated Low-light Image Pairs, Retinex Theory, Noise2Noise.

## 1 Introduction

When capturing photos in low-light conditions, the resulting images typically exhibit poor quality due to environmental and equipment limitations. These images usually exhibit low contrast and severe noise, potentially affecting the performance of subsequent visual tasks. Low-light image enhancement is demanded to improve the visual quality of the images and the accuracy of subsequent tasks.

Among existing low-light enhancement methods, supervised methods utilize paired low-light and normal-light images, yet collecting such paired images is challenging. To eliminate the problem, unsupervised learning methods are proposed, and many

---

* Corresponding author.

unsupervised methods [3, 12, 32, 33] learn enhancement from only a single image, relying on handcrafted priors. However, the information contained in a single image is limited, and the handcrafted priors are difficult to set. To solve this problem, PairLIE [2] utilizes paired low-light images and fewer handcrafted priors. This method does not require any normal-light reference image but requires multiple low-light images of the same scene. However, it is also difficult to obtain such images in reality, especially for moving scenes.



**Fig. 1.** The restoration results of our method, with the low-light images input above and the enhanced images below.

To tackle the challenges of insufficient information within a single image and the difficulty in designing suitable priors, along with the difficulty in obtaining paired low-light images in unsupervised learning methods. We propose a novel low-light enhancement method based on unsupervised learning, which does not require a special collection of datasets. It uses a single low-light image to generate a pair of images with consistent content, independent noise, and slightly different illumination. This allows us to use consistent reflectance and Noise2Noise [9], which well makes up for the shortcomings of relying on many handcrafted priors.

We first generate paired low-light images by downsampling the original image to create two sub-images with consistent content and independent noise. We then apply a linear transformation to one image, varying coefficient values from 1 to 1.5, introducing a slight difference in illumination components between the sub-images. Second, based on the paired images, we propose an unsupervised low-light enhancement network architecture. Considering the consistent contents of the two images, we constrain them to each other and maintain the intrinsic properties of the objects in the images during the enhancement process, which avoids issues like color deviations and artifacts often encountered in single-image enhancement methods. To achieve this, we apply Retinex theory, decomposing the images into illumination and reflectance components. The decomposed illumination represents brightness levels, whereas the reflectance embodies the intrinsic properties of the objects. Thus, the reflectance in both images should align. We utilized reflectance consistency to constrain the reflectance components of the two images, facilitating the restoration of object properties in the images. Moreover, low-light images often contain noise due to a low signal-to-noise ratio. Previous unsupervised methods, lacking reference images, mainly rely on TV Loss for denoising and

frequently lead to incomplete noise reduction or excessive smoothing, resulting in the loss of fine details. Considering the independence of noise in the two images, we employ the Noise2Noise [9] for noise removal, which efficiently reduces noise while preserving original details, consequently further enhancing the image quality. Fig. 1 shows that our method can effectively improve contrast, restore details and remove noise.

Our contributions are summarized as follows:

– We generate a pair of images with consistent content, independent noise, and slightly different illumination from a single low-light image, which not only retains the original image information but also meets the training requirements.

– To ensure that the image content does not migrate during the enhancement process, we use the Retinex theory to decompose the generated image and maintain the reflectance consistent.

– To reduce noise in low-light images, we take advantage of the independent noise characteristics of the images and utilize Noise2Noise.

## 2 Related Work

### 2.1 Low-light image enhancement

Low-light image enhancement is an area of great interest, and a variety of solutions have emerged. Next, we will introduce related research from two aspects: traditional methods [5, 6, 11, 19–21] and deep learning-based methods [2, 3, 8, 12–16, 23, 25, 30–33].

Traditional methods include histogram equalization methods [19, 20], which achieve the purpose of image enhancement by stretching the dynamic range of the image to 0-255, and methods based on Retinex theory [5, 6, 11, 21], which decomposes the original low-light image into illumination and reflectance. Generally, the reflectance is directly used as the result of enhancement.

Deep learning-based methods mainly include methods based on supervised learning [13–15,23,25,30,31] and methods based on unsupervised learning [2,3,8, 12, 16, 32, 33].

Supervised learning-based methods have strong learning capabilities, but require paired low-light and normal-light images and have poor generalization capabilities. LLNet [13] is the first deep learning-based low-light enhancement method. MBLLEN [15] is an end-to-end multi-branch enhancement network that implements enhancement through feature extraction, enhancement, and fusion modules. TBEFN [14] is a multiple-exposure fusion network. Retinex-Net [23] includes decomposition and enhancement modules for decomposing images and adjusting illumination. KinD [31] is composed of three sub-networks for decomposition, reflectance recovery, and illumination adjustment. The improved KinD is called KinD++ [30]. URetinex-Net [25] uses a retinex-based deep unfolding network to improve adaptability.

Unsupervised learning methods do not require paired low-light and normal-light images. EnlightenGAN [8] uses U-Net and global-local discriminator to ensure the enhancement results are realistic. RRDNet [33] decomposes the input image into

illumination, reflectance, and noise and designs specialized denoising losses. RetinexDIP [32] uses random noise to generate illumination and reflectance components. Zero-DCE [3] is a deep curve estimation network that utilizes high-order curves to perform pixel-level adjustments to dynamic range. RUAS [12] is a Retinex-inspired unrolling method that uses cooperative architecture search to discover prior architectures of basic blocks. SCI [16] is a self-calibrating illumination learning framework. PairLIE [2] uses pairs of low-light images and denoising before decomposition. Our network has a similar idea to PairLIE. The difference is that it does not require a paired dataset. Also, PairLIE uses single-image for denoising, but we use paired downsampling images for denoising.

### 2.2    Image denoising

Image denoising is also a hot issue. Traditional denoising algorithms like BM3D [1] and Anscombe [17] are generally designed for specific types of noise. Recently, deep learning-based denoising methods are becoming more and more popular. Supervised learning-based methods [4, 24, 27, 28] utilize noisy-clean image pairs for training. However, acquiring the datasets is challenging. Self-supervised learning-based methods do not require paired noisy-clean images. A typical method is Noise2Noise [9], which uses two noisy images of the same scene to restrict each other. The model can achieve the same effect as supervised learning when the amount of data is sufficient, but it is also very difficult to collect paired noise images in the same scene. Noise2Fast [10], Neighbor2Neighbor [7], and Zero-Shot Noise2Noise [18] are improved versions of Noise-2-Noise, which do not require paired images. These methods downsample two images as noise-noise pairs from a single original image. We follow the idea in Zero-Shot Noise2Noise [18] for noise removal.

## 3    Method

In this section, we will introduce the generation of low-light image pairs, network architecture, and loss functions.

### 3.1    Generation of low-light image pairs

We introduce a downsampler to generate low-light image pairs from a single low-light image for training. As depicted in Fig. 2, the original low-light image I of size [H, W, C]is convolved independently with two different convolution kernels:

$$k_1 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}, \tag{1}$$

and

$$k_2 = \begin{bmatrix} 0 & 0.5 \\ 0.5 & 0 \end{bmatrix}. \tag{2}$$

This process results in generating two downsampled images, each with dimensions $[H/2, W/2, C]$. Since adjacent pixel values have similar content but independent noise, the two images have the same content but independent noise [18]. To facilitate subsequent Retinex decomposition, we introduce a slight difference in illumination between the two images, specifically, we use linear transformation to slightly enhance one of the images:

$$I_2 = \alpha \cdot I_2^{tmp}, \ \alpha \epsilon [1, 1.5], \tag{3}$$

where $\alpha$ is the transformation coefficient, ranging from 1 to 1.5. As shown in Fig. 2, the enhancement is slight and will not cause color deviations.
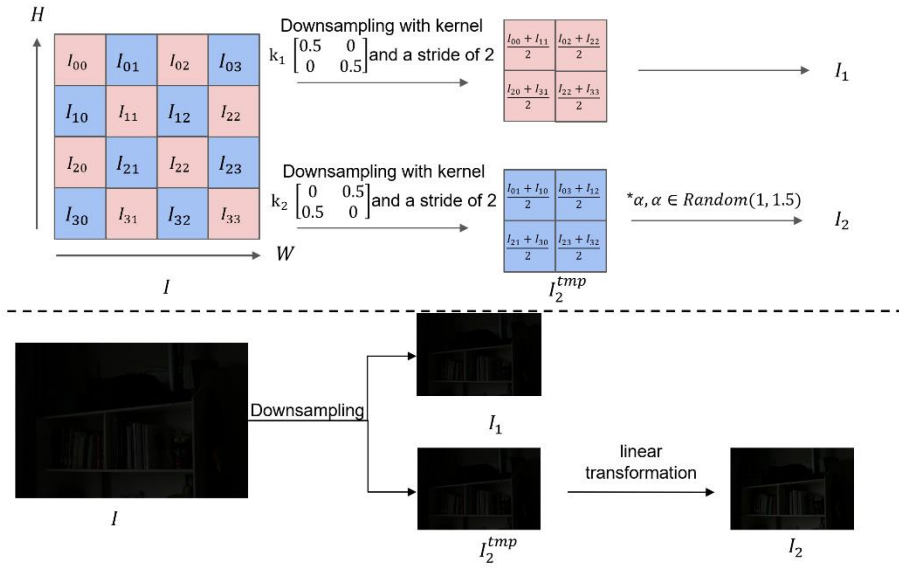


**Fig. 2.** An overview of the generation of low-light image pairs. First, the original image is downsampled to obtain two sub-images with consistent content and independent noise, and then a slight linear transformation is performed on one of them.

### 3.2   Network structure

We use the paired images generated in the previous part with independent noise, consistent content, and slightly different illumination to complete the training of the enhancement model. In this way, we can not only take advantage of its consistent reflectance characteristics, but also facilitate self-supervised denoising, and it removes noise well while enhancing.

Considering that the illumination component is generally smooth and noise often exists in the reflectance component, we use:

$$I = L \odot (R + N), \tag{4}$$
$$I_1 = L_1 \odot (R_1 + N_1), \tag{5}$$
$$I_2 = L_2 \odot (R_2 + N_2), \tag{6}$$

where represents pixel-wise multiplication, $I$, $I_1$, and $I_2$ are input image and its two sub-images, $L$, $L_1$, and $L_2$ represent illumination components, $R$, $R_1$, and $R_2$ represent reflectance components, $N$, $N_1$, and $N_2$ represent noise.

As illustrated in Fig. 3, In the training stage, $I$, $I_1$, and $I_2$ are first input to the illumination network L−Net and the reflectance network R−Net respectively for decomposition, and get the smooth Illumination components $L$, $L_1$, and $L_2$ and noisy reflectance components $R^n$, $R_1^n$, and $R_2^n$. L−Net and R−Net are both CNNs with five convolutional layers. Each layer consists of several convolution kernels with size 3×3 and stride 1, followed by ReLU activation function. After the last convolutional layer is the Sigmoid activation function. The difference is that the output result of L− Net is a single-channel illumination map, while the output result of R−Net is a three-channel reflectance map.
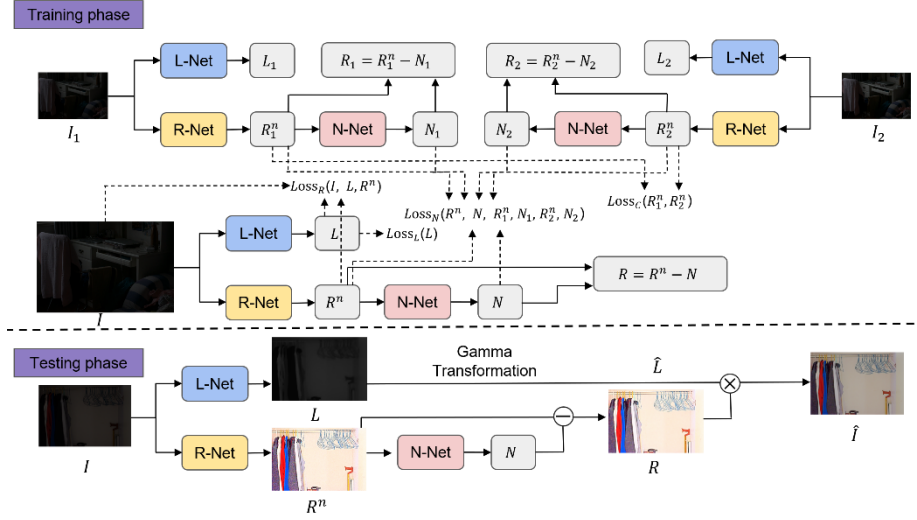


**Fig. 3.** The proposed framework involves a training phase and a testing phase. In the training phase, the original image and its two sub-images are decomposed by L-Net and R-Net to obtain illumination and noisy reflectance. The decomposition process is constrained using the illumination loss LossL, reconstruction loss $Loss_L$, and reflectance consistency loss $Loss_C$ . N-Net estimates the noise in the noisy reflectance and uses noise loss $Loss_N$ to constrain the noise estimation process. In the testing phase, a low-light image is input, and L-Net is used to obtain the illumination. R-Net and N-Net are used to obtain the denoised reflectance. The final result is obtained by multiplying the denoised reflectance and the illumination enhanced by gamma transformation.

The noise within the noisy reflectance components is estimated by inputting them into the N−Net for noise assessment, which generates $N$, $N_1$, and $N_2$. Subtracting this noise from the noisy reflectance components $R^n$, $R_1^n$, and $R_2^n$ results in the clean reflectance components $R$, $R_1$, and $R_2$. N−Net is a CNN with five convolutional layers. Each layer consists of several convolution kernels with size $3 \times 3$ and stride 1, followed by ReLU activation function. The last convolutional layer is followed by the Tanh activation function.

In the testing phase, a low-light image $I$ is input, and after L−Net, a smooth illumination $L$ is obtained. After R−Net and N−Net, a clean reflectance $R$ after denoising is finally obtained, then perform gamma transformation on the illumination L to obtain enhanced illumination $\hat{L}$:

$$\hat{L} = L^{\gamma}, \tag{7}$$

where $\gamma$ represents the illumination correction factor.

The results of image decomposition, denoising, and enhancement are shown in Fig. 4. It can be seen that it can denoise well while ensuring enhancement.



| (a) Input | (b) Illumination | (c) Noisy reflectance | (d) Reflectance | (e) Result |

**Fig. 4.** An example of our decomposition, denoising, and enhancement. Please zoom in for details.

### 3.3 Loss Function

**Illumination loss** We constrain the illumination map by illumination loss, which consists of illumination consistency loss and illumination smoothness loss:

$$Loss_L(L) = \lambda_1 ||L - L_0||_2 + \triangle (L), \tag{8}$$

where $\lambda_1$ denotes the weight, $L_0$ represents the initial illumination of $I$, $\triangle (L)$ represents the gradient operation of $L$. $L_0$ usually takes the maximum value of the three channels of $I$ :

$$L_0(x) = \max_{c \in \{R,G,B\}} I^c (x), \tag{9}$$

where $x$ represents each pixel, $c$ represents the channel.

**Reconstruction loss** The input image $I$ is decomposed into smooth illumination and noisy reflectance. The original image is reconstructed by multiplying the two. In addition, image reconstruction can also be considered from the perspective of reflectance reconstruction. The aim is to make the noisy reflectance consistent with the input image divided by the smooth illumination:

$$Loss_R(I, L, R^n) = ||I - L \odot R^n||_2 + \left\|R^n - \frac{I}{stopgrad(L)}\right\|_2. \tag{10}$$

When constraining reflectance, the illumination is fixed and reverse propagation is stopped to ensure stable training.

**Reflectance consistency loss** We generate low-light paired images with consistent content and slightly different illumination. We use the reflectance components of the two low-light images to refer to each other and introduce a reflectance consistency loss:

$$Loss_C(R_1^n, R_2^n) = ||R_1^n - R_2^n||_2. \tag{11}$$

**Noise loss** The images we generate are content-consistent and noise-independent, one image approximates another by subtracting estimated noise from noisy reflectance to achieve the purpose of denoising. To make full use of the image, a symmetric loss is used:

$$Loss_{N\_sym}(R_1^n, N_1, R_2^n, N_2) = \|(R_1^n - N_1) - R_2^n\|_2 + \|(R_2^n - N_2) - R_1^n\|_2. \quad (12)$$

To better ensure consistency, we fully utilize the image pre-sampling by directly inputting it into the network. Subsequently, we ensure consistency between the results obtained by sampling first and then decomposing, denoising, or decomposing, denoising first and then sampling, by using a consistency loss:

$$Loss_{N\_con}(R^n, N, R_1^n, N_1, R_2^n, N_2)$$
$$= \|D_1(R^n - N) - (R_1^n - N_1)\|_2 + \|D_2(R^n - N) - R_2^n - N_2\|_2, \quad (13)$$

among them, $D_1(R^n - N)$ and $D_2(R^n - N)$ represent the reflectance of two sub-images obtained by downsampling the denoised reflectance of the original input image.

The noise loss consists of $Loss_{N\_sym}$ and $Loss_{N\_con}$:

$$Loss_N = Loss_{N\_sym} + Loss_{N\_con}. \quad (14)$$

**Total loss** The total loss is a combination of each loss:

$$Loss_{total} = \lambda_2 Loss_L + \lambda_3 Loss_R + Loss_C + \lambda_4 Loss_N, \quad (15)$$

where $\lambda_2$, $\lambda_3$, and $\lambda_4$ denote the weights.

**Table 1.** Comparison of our and other methods on LOL-v1 [23] and LOL-v2-real [26] datasets. "S", and "U" represent "Supervised", and "Unsupervised" methods. The optimal and suboptimal values of the supervised and unsupervised methods are marked in red and blue respectively.

| Method | Type | LOL-v1 | | | LOL-v2-real | | |
|---|---|---|---|---|---|---|---|
| | | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ |
| MBLLEN [15] | S | 17.562 | 0.729 | 0.173 | 17.296 | 0.680 | 0.221 |
| TBEFN [14] | S | 17.350 | 0.777 | 0.209 | 20.283 | 0.830 | 0.171 |
| Retinex-Net [23] | S | 16.774 | 0.424 | 0.473 | 16.097 | 0.407 | 0.542 |
| KinD [31] | S | 17.647 | 0.771 | 0.174 | 20.588 | 0.817 | 0.143 |
| KinD++ [30] | S | 17.751 | 0.758 | 0.197 | 17.660 | 0.760 | 0.216 |
| URetinex-Net [25] | S | 21.328 | 0.832 | 0.120 | 19.780 | 0.842 | 0.107 |
| RRDNet [33] | U | 11.311 | 0.456 | 0.362 | 13.909 | 0.485 | 0.317 |
| RetinexDIP [32] | U | 9.094 | 0.325 | 0.449 | 11.462 | 0.361 | 0.404 |
| Zero-DCE [3] | U | 14.860 | 0.562 | 0.335 | 18.058 | 0.579 | 0.312 |
| RUAS [12] | U | 16.405 | 0.503 | 0.270 | 15.325 | 0.493 | 0.309 |
| EnlightenGAN [8] | U | 17.483 | 0.651 | 0.322 | 18.639 | 0.676 | 0.308 |
| SCI [16] | U | 14.783 | 0.525 | 0.339 | 17.303 | 0.539 | 0.307 |
| PairLIE($\lambda = 0.2$) [2] | U | 18.468 | 0.742 | 0.243 | 19.884 | 0.773 | 0.234 |
| Ours | U | 18.809 | 0.783 | 0.189 | 19.181 | 0.780 | 0.212 |

# 4    Experiments

In this section, we first introduce the experimental details, and datasets, and then conduct quantitative and qualitative comparisons with the latest models. Finally, ablation experiments are performed in various settings.

## 4.1    Experimental details

We use the Pytorch framework to complete experiments on NVIDIA GeForce RTX 3080 GPU, setting $\lambda_1 = 10$, $\lambda_2 = 5$, $\lambda_3 = 5$, $\lambda_4 = 0.5$, $\gamma = 0.2$. The number of convolution kernels of the five-layer R-Net and N-Net is [64,64,64,64,3], and the number of convolution kernels of I-Net is [64,64,64,64,1]. We trained the model for 400 epochs with a learning rate of 0.0001 and decayed it by half every 100 epochs.

Our approach is compared quantitatively and qualitatively with supervised learning-based methods such as MBLLEN [15], TBEFN [14], Retinex-Net [23], KinD [31], KinD++ [30], URetinex-Net [25], as well as unsupervised learning based methods include EnlightenGAN [8], SCI [16], PairLIE [2], Zero-DCE [3], RRDNet [33], RetinexDIP [32], RUAS [12].

**Table 2.** Comparison of our and other methods on LOL-v2-synthetic [26] dataset. "S", and "U" represent "Supervised", and "Unsupervised" methods. The optimal and suboptimal values of the supervised and unsupervised methods are marked in red and blue respectively.

| Method | Type | LOL-v2-synthetic | | |
| --- | --- | --- | --- | --- |
| | | PSNR↑ | SSIM↑ | LPIPS↓ |
| MBLLEN [15] | S | 17.822 | 0.778 | 0.134 |
| TBEFN [14] | S | 18.260 | 0.842 | 0.175 |
| Retinex-Net [23] | S | 17.136 | 0.756 | 0.255 |
| KinD [31] | S | 17.275 | 0.757 | 0.252 |
| KinD++ [30] | S | 17.477 | 0.785 | 0.231 |
| URetinex-Net [25] | S | 18.770 | 0.822 | 0.190 |
| RRDNet [33] | U | 14.838 | 0.654 | 0.246 |
| RetinexDIP [32] | U | 15.976 | 0.762 | 0.210 |
| Zero-DCE [3] | U | 17.756 | 0.813 | 0.168 |
| RUAS [12] | U | 13.404 | 0.639 | 0.363 |
| EnlightenGAN [8] | U | 16.572 | 0.771 | 0.211 |
| SCI [16] | U | 15.427 | 0.744 | 0.232 |
| PairLIE($\lambda = 0.2$) [2] | U | 19.074 | 0.794 | 0.229 |
| Ours | U | 19.741 | 0.829 | 0.185 |

## 4.2    Datasets and Metrics

We evaluate our method on LOL-v1 [23] and LOL-v2 [26]. LOL-v2 is further divided into LOL-v2-real and LOL-v2-synthetic, with LOL-v2-synthetic showing significant content differences from LOL-v1 and LOL-v2-real. The training and testing set proportions for LOL-v1, LOL-v2-real, and LOL-v2-synthetic are 485:15, 689:100, and 900:100, respectively. We only use low-light images from the LOL-v1 training dataset for training, and LOL-v1, LOL-v2-real, and LOL-v2-synthetic testing datasets for validation. As these datasets provide reference images, we assess our method's performance using the metrics PSNR, SSIM [22], and LPIPS [29]. Higher PSNR and SSIM values indicate better results, while the opposite holds for LPIPS.
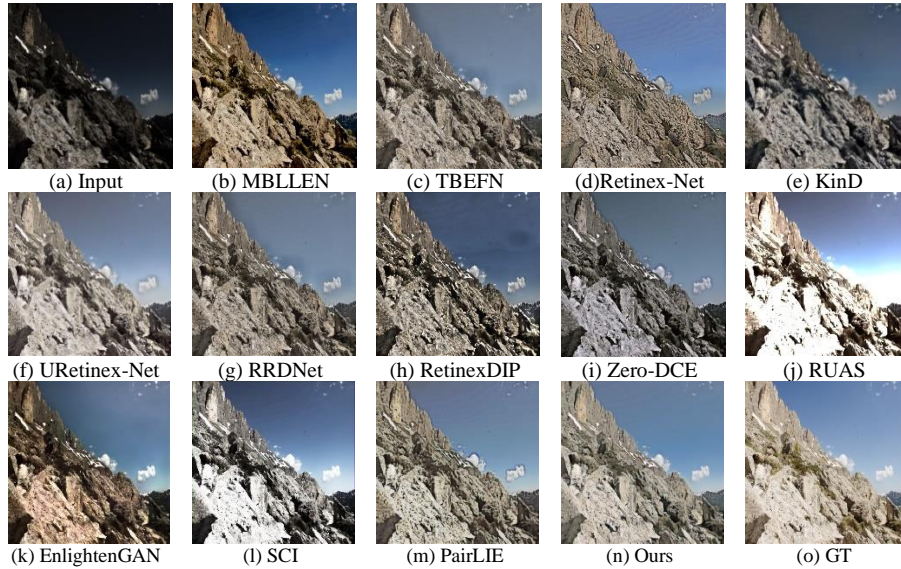


| | | | | |
|---|---|---|---|---|
| (a) Input | (b) MBLLEN | (c) TBEFN | (d)Retinex-Net | (e) KinD |
| (f) URetinex-Net | (g) RRDNet | (h) RetinexDIP | (i) Zero-DCE | (j) RUAS |
| (k) EnlightenGAN | (l) SCI | (m) PairLIE | (n) Ours | (o) GT |

**Fig. 5.** Comparison of the enhancement performances between our method and others. Please zoom in for details.

## 4.3    Quantitative Comparison

Table. 1 and Table. 2 present a quantitative comparison of our method and other methods on LOL-v1 [23] and LOL-v2 [26] datasets, where the illumination adjustment factor λ of PairLIE [2] is set to the default value of 0.2. It can be observed that our method outperforms all unsupervised methods, particularly in SSIM [22] and LPIPS [29], which are designed to align with human-perceptive image quality. This improvement is attributed to our method of generating paired low-light images, avoiding color deviations and artifacts through consistent reflectance, and effectively denoising using Noise2Noise. PairLIE also performs well, but PairLIE requires paired low-light images for training. In addition, our method is also competitive compared with most supervised algorithms. It can be seen that our method performs best on LOL-v2-synthetic [26]

because this dataset is significantly different from LOL-v1 [23] and LOL-v2-real [26], which means that our method has better generalization.

## 4.4    Qualitative Comparison

Fig.5 illustrates a visual comparison between our method and others. It can be observed that our method produces an image with better brightness and contrast and fewer artifacts, resulting in a more natural appearance, and other methods exhibit issues such as underexposure, overexposure, severe color deviations, and artifacts. Furthermore, Fig.6 shows the visual comparison of the denoising capabilities of our method and others. It is evident that our method thoroughly denoises the image, while other methods such as PairLIE [2], Retinex-Net [23], and SCI [16] still have noise in their results. It shows that the denoising module we designed based on Nose2Noise is effective. Additionally, other methods have problems such as over-enhancement, insufficient enhancement, and color deviation.



|  |  |  |  |  |
|---|---|---|---|---|
| (a) Input | (b) MBLLEN | (c) TBEFN | (d)Retinex-Net | (e) KinD |
| (f) URetinex-Net | (g) RRDNet | (h) RetinexDIP | (i) Zero-DCE | (j) RUAS |
| (k) EnlightenGAN | (l) SCI | (m) PairLIE | (n) Ours | (o) GT |

**Fig. 6.** Comparison of the denoising performances between our method and others. Please zoom in for details.

## 4.5    Ablation Study

We conduct an ablation study to examine the impact of each component in our model across different settings. We try the following settings: 1) without slight linear transformation(LT) when generating images. 2) without $Loss_L$. 3) without $Loss_R$. 4) without $Loss_C$. 5) without $Loss_N$.

As indicated in Table. 3, the setting we select is the best and every component is effective. Additionally, the illumination loss, reconstruction loss, and noise loss have a more significant impact on the results.

Fig. 7 illustrates visual comparisons under different settings. In Fig.7 (c), the absence of illumination loss $Loss_L$ prevents the restored image from reaching a normal illumination level. In Fig.7 (d), if the reconstruction loss $Loss_R$ is missing, the result will lose color and detail information. Fig.7 (f) illustrates that the absence of noise loss $Loss_N$ results in severe noise in the output. Fig.7 (e) demonstrates that the missing reflectance consistency loss $Loss_C$ has minimal impact on the resulting image, primarily because $Loss_N$ also encompasses the consistency constraints of reflectance. In Fig. 7 (b), not applying a linear transformation during image generation may not have a noticeable effect on the resulting image visually. However, quantitatively, the operation still works.

**Table 3.** Comparison of different settings on LOL-v1 [23] dataset, the best results are in bold.

| Setting | PSNR↑ | SSIM↑ | LPIPS↓ |
|---|---|---|---|
| w/o LT | 18.788 | 0.779 | 0.191 |
| w/o $Loss_L$ | 8.901 | 0.333 | 0.433 |
| w/o $Loss_R$ | 12.221 | 0.508 | 0.798 |
| w/o $Loss_C$ | 18.709 | 0.779 | 0.191 |
| w/o $Loss_N$ | 17.659 | 0.591 | 0.366 |
| Ours | **18.809** | **0.783** | **0.189** |



(a) Input  (b) w/o LT  (c) w/o $Loss_L$  (d) w/o $Loss_R$

(e) w/o $Loss_C$  (f) w/o $Loss_N$  (g) Ours  (h) GT
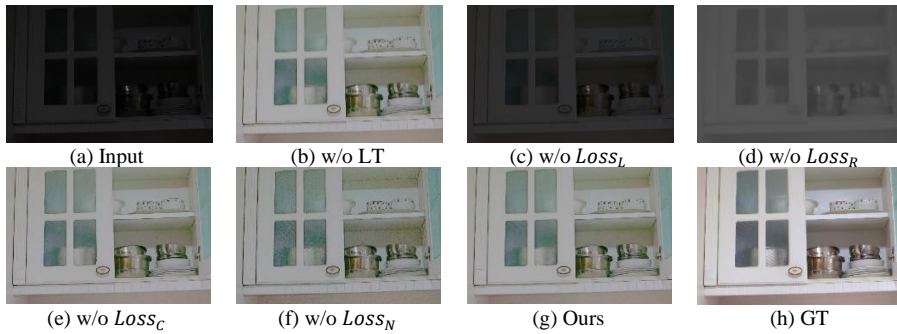
**Fig. 7.** Comparison of different settings. Please zoom in for details.

## 5    Conclusion

In this paper, we propose an unsupervised low-light image enhancement method that exclusively relies on low-light images for training. First, we construct paired images with consistent content, noise independence, and slightly different illumination. Second, we leverage reflectance consistency to facilitate the restoration of image content. Moreover, we use Noise2Noise to remove noise in low-light images. Extensive

experiments on public datasets have validated the efficacy of our method in enhancing images and removing noise. Our future endeavors will focus on addressing the challenge of adaptive illumination adjustment.

# References

1. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.O.: Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Trans. Image Process. 16(8), 2080–2095 (2007)
2. Fu, Z., Yang, Y., Tu, X., Huang, Y., Ding, X., Ma, K.: Learning a simple low-light image enhancer from paired low-light instances. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023. pp. 22252–22261. IEEE (2023)
3. Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020. pp. 1777–1786. Computer Vision Foundation / IEEE (2020)
4. Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L.: Toward convolutional blind denoising of real photographs. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019. pp. 1712–1722. Computer Vision Foundation / IEEE (2019)
5. Guo, X., Li, Y., Ling, H.: LIME: low-light image enhancement via illumination map estimation. IEEE Trans. Image Process. 26(2), 982–993 (2017)
6. Hao, S., Han, X., Guo, Y., Xu, X., Wang, M.: Low-light image enhancement with semi-decoupled decomposition. IEEE Trans. Multim. 22(12), 3025–3038 (2020)
7. Huang, T., Li, S., Jia, X., Lu, H., Liu, J.: Neighbor2neighbor: Self-supervised denoising from single noisy images. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021. pp. 14781–14790. Computer Vision Foundation / IEEE (2021)
8. Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z.: Enlightengan: Deep light enhancement without paired supervision. IEEE Trans. Image Process. 30, 2340–2349 (2021)
9. Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., Aila, T.: Noise2noise: Learning image restoration without clean data. In: Dy, J.G., Krause, A. (eds.) Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018. Proceedings of Machine Learning Research, vol. 80, pp. 2971–2980. PMLR (2018)
10. Lequyer, J., Philip, R., Sharma, A., Hsu, W., Pelletier, L.: A fast blind zero-shot denoiser. Nat. Mac. Intell. 4(11), 953–963 (2022)
11. Li, M., Liu, J., Yang, W., Sun, X., Guo, Z.: Structure-revealing low-light image enhancement via robust retinex model. IEEE Trans. Image Process. 27(6), 2828– 2841 (2018)
12. Liu, R., Ma, L., Zhang, J., Fan, X., Luo, Z.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021. pp. 10561–10570. Computer Vision Foundation / IEEE (2021)
13. Lore, K.G., Akintayo, A., Sarkar, S.: Llnet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognit. 61, 650–662 (2017)

14. Lu, K., Zhang, L.: TBEFN: A two-branch exposure-fusion network for low-light image enhancement. IEEE Trans. Multim. 23, 4093–4105 (2021)
15. Lv, F., Lu, F., Wu, J., Lim, C.: MBLLEN: low-light image/video enhancement using cnns. In: British Machine Vision Conference 2018, BMVC 2018, Newcastle, UK, September 3-6, 2018. p. 220. BMVA Press (2018)
16. Ma, L., Ma, T., Liu, R., Fan, X., Luo, Z.: Toward fast, flexible, and robust low-light image enhancement. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022. pp. 5627–5636. IEEE (2022)
17. Mäkitalo, M.J., Foi, A.: Optimal inversion of the anscombe transformation in lowcount poisson image denoising. IEEE Trans. Image Process. 20(1), 99–109 (2011)
18. Mansour, Y., Heckel, R.: Zero-shot noise2noise: Efficient image denoising without any data. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023. pp. 14018–14027. IEEE (2023)
19. Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Zuiderveld, K.: Adaptive histogram equalization and its variations. Computer Vision Graphics & Image Processing 39(3), 355–368 (1987)
20. Reza, A.M.: Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. J. VLSI Signal Process. 38(1), 35–44 (2004)
21. Wang, S., Zheng, J., Hu, H., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Trans. Image Process. 22(9), 3538–3548 (2013)
22. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13(4), 600–612 (2004)
23. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: British Machine Vision Conference 2018, BMVC 2018, Newcastle, UK, September 3-6, 2018. p. 155. BMVA Press (2018)
24. Wei, K., Fu, Y., Yang, J., Huang, H.: A physics-based noise formation model for extreme low-light raw denoising. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020. pp. 2755–2764. Computer Vision Foundation / IEEE (2020)
25. Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., Jiang, J.: Uretinex-net: Retinexbased deep unfolding network for low-light image enhancement. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022. pp. 5891–5900. IEEE (2022)
26. Yang, W., Wang, W., Huang, H., Wang, S., Liu, J.: Sparse gradient regularized deep retinex network for robust low-light image enhancement. IEEE Trans. Image Process. 30, 2072–2086 (2021)
27. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Trans. Image Process. 26(7), 3142–3155 (2017)
28. Zhang, K., Zuo, W., Zhang, L.: Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. IEEE Trans. Image Process. 27(9), 4608–4622 (2018)
29. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018. pp. 586–595. Computer Vision Foundation / IEEE Computer Society (2018)
30. Zhang, Y., Guo, X., Ma, J., Liu, W., Zhang, J.: Beyond brightening low-light images. Int. J. Comput. Vis. 129(4), 1013–1037 (2021)
31. Zhang, Y., Zhang, J., Guo, X.: Kindling the darkness: A practical low-light image enhancer. In: Amsaleg, L., Huet, B., Larson, M.A., Gravier, G., Hung, H., Ngo, C., Ooi, W.T. (eds.)

Proceedings of the 27th ACM International Conference on Multimedia, MM 2019, Nice, France, October 21-25, 2019. pp. 1632–1640. ACM (2019)

32. Zhao, Z., Xiong, B., Wang, L., Ou, Q., Yu, L., Kuang, F.: Retinexdip: A unified deep framework for low-light image enhancement. IEEE Trans. Circuits Syst. Video Technol. 32(3), 1076–1088 (2022)

33. Zhu, A., Zhang, L., Shen, Y., Ma, Y., Zhao, S., Zhou, Y.: Zero-shot restoration of underexposed images via robust retinex decomposition. In: IEEE International Conference on Multimedia and Expo, ICME 2020, London, UK, July 6-10, 2020. pp. 1–6. IEEE (2020)