International Journal of Computer Vision
https://doi.org/10.1007/s11263-022-01667-9

Low-light Image Enhancement via Breaking Down the Darkness

Xiaojie Guo[1] · **Qiming Hu**[1]

Received: 31 January 2022 / Accepted: 3 August 2022

Abstract
Images captured in low-light environments often suffer from complex degradation. Simply adjusting light would inevitably result in burst of hidden noise and color distortion. To seek results with satisfied lighting, cleanliness, and realism from degraded inputs, this paper presents a novel framework inspired by the divide-and-rule principle, greatly alleviating the degradation entanglement. Assuming that an image can be decomposed into texture (with possible noise) and color components, one can specifically execute noise removal and color correction along with light adjustment. For this purpose, we propose to convert an image from the RGB colorspace into a luminance-chrominance one. An adjustable noise suppression network is designed to eliminate noise in the brightened luminance, having the illumination map estimated to indicate noise amplification levels. The enhanced luminance further serves as guidance for the chrominance mapper to generate realistic colors. Extensive experiments are conducted to reveal the effectiveness of our design, and demonstrate its superiority over state-of-the-art alternatives both quantitatively and qualitatively on several benchmark datasets. Our code has been made publicly available at https://github.com/mingcv/Bread.

Keywords Low-light image enhancement · Image decomposition · Divide and rule

1 Introduction

Capturing high-quality images under less controlled conditions is challenging especially using mobile devices. Often, people take pictures in unsatisfactory light environments. For instance, we might take a photo against the light source (please see Fig. 1); or a surveillance camera may be monitoring a place in the nighttime. In such cases, the images will suffer from poor visibility. To obtain high-quality images, a few solutions can be applied, like one can expand exposure time to receive more photons, but if the target scene is dynamic, the blur effect likely happens; another possible way is to set a flash for light compensation, which however frequently introduces unexpected highlights and unbalanced lighting into photos. Hence, instead of upgrading hardware, developing effective low-light enhancement techniques is highly desired for practical use.

Low-light image enhancement is not a solo problem of light adjustment, which also has troubles of noise burst and color distortion concealed in the darkness because of the limited capability of photographing devices. A number of methods follow the Retinex theory (Land, 1977) through decomposing an image I into its reflectance map R and illumination map L in the following form:

$$I = R \circ L, \tag{1}$$

where $\circ$ designates the Hadamard product operator. Because the reflectance component is defined as the intrinsic property of material, it is constant against variant illuminations. Ideally, once the illumination is estimated or given, the reflectance can be immediately obtained and vice versa. One can adjust illumination according to different demands. However, due to the limited quality of sensors, the degradation E is always an annoying resident in images, thus the model turns out to be:

$$I = R \circ L + E. \tag{2}$$

Communicated by Yu Li.

✉ Xiaojie Guo
xj.max.guo@gmail.com

Qiming Hu
huqiming@tju.edu.cn

[1] College of Intelligence and Computing, Tianjin University, Tianjin 300350, China

Published online: 02 October 2022

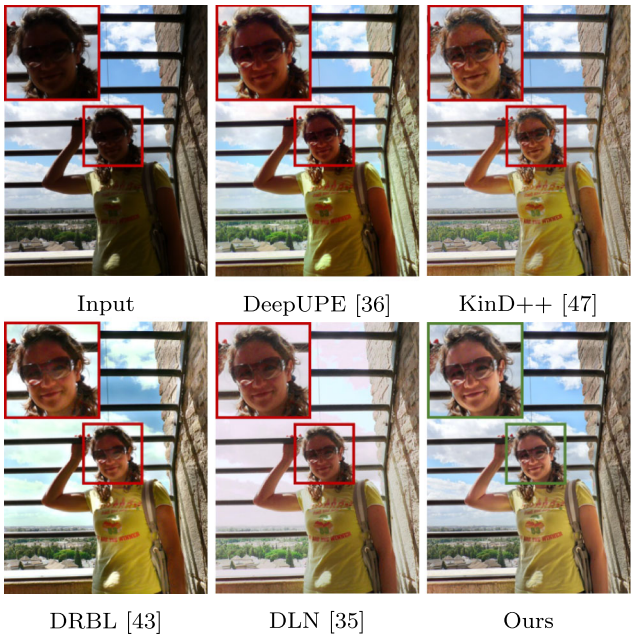

Fig. 1 Visual comparison on a sample from the VV dataset. Our method obtains striking improvement over the other competitors, *e.g.*, the sky tone and the realism of human skin

A simple algebraic transformation leads the above to

$$I = (R + \tilde{E}) \circ L, \tag{3}$$

where $\tilde{E}$ is defined as E/L with element-wise division. We can see from Eq. (3) that the degradation will be also amplified along with light enhancement, which becomes spatially-correlated with L as discussed in Zhang et al. (2019). Further, in low-light conditions, sensors (either CCD or CMOS) are sensitive and non-linear to insufficient photons of different light spectrum, resulting in color distortion even when the illumination and noise have been properly handled. Therefore, for the sake of producing high-quality results from degraded inputs, a qualified algorithm should remedy the highly entangled illness of dim light, noise, and color distortion.

This work studies, besides separating the illumination from the input, how to handle the complex degradation in the darkness from a divide-and-rule perspective. Suppose that an image can be disassembled into the texture (together with the main body of noise) and color components, the operations including noise removal and color correction along with light adjustment could be executed specifically. In other words, by the texture-color separation, we are possible to treat the degradation in the texture and color components individually, *i.e.*, focusing on the noise issue in the texture and the color distortion in the counterpart. Please notice that, although several methods (Yang et al., 2020; Zhang et al., 2021) have been recently proposed to take care of the noise and color shift issues as a whole without distinction in the recovered reflectance, they barely consider decomposing the degradation from a texture-color point of view to further ease the problem. It is worth mentioning that, by such a divide-and-rule principle, the original searching space for training deep networks would be largely (exponentially) reduced into smaller sub-ones, which can be trained in an individual and parallel way, thus relieving the burden of tuning hyper-parameters jointly.

Motivated by the above consideration, this paper develops an effective low-light image enhancement framework via breaking down the darkness (Bread for short). The major contributions of this paper are as follows:

1. To the best of our knowledge, this is a pioneering attempt to decouple the entanglement of noise and color distortion, further mitigating the difficulty of low-light enhancement with complex degradation.
2. We present an effective noise synthesis strategy under the guidance of illumination, significantly improving the quality of suppressing amplified and spatially correlated noise in the texture.
3. To tackle the color distortion issue left in light-enhanced images, we design a novel color adaption network, which can properly deal with the color according to given textures.
4. Extensive comparisons together with ablation studies are provided to verify the efficacy of our method, and reveal its advance over other state-of-the-art methods both qualitatively and quantitatively.

2 Related Work

Many low-light image enhancement methods have been proposed over last decades, which can be roughly grouped into traditional and deep learning-based methods.

2.1 Traditional Methods

The simplest and most intuitive way is to linearly stretch the value range of or execute a non-linear Gamma correction on inputs. To be more rational, global and local histogram-based methods (Pisano et al., 1998; Cheng & Shi, 2004; Abdullah-Al-Wadud et al., 2007; Celik & Tjahjadi, 2011; Lee et al., 2013) are introduced to dynamically expand the range of images. In spite of their ease of use, the enhancement quality is hardly guaranteed, due to the content-blindness. Derived from the Retinex theory (Land, 1977), Single-scale Retinex (SSR) (Jobson et al., 1997b) uses the Gaussian blurred input as the illumination map, which is then removed from the input as its final result. Multi-scale Retinex (MSR) (Jobson et al., 1997a) extends SSR by fusing the results of multiple

Gaussian blur functions with different variances. Besides the above mentioned attempts, NPE (Wang et al., 2013) takes the local maxima assumption to predict the illumination. The illumination is manipulated by a mapper as the enhanced version, and then merged with the reflectance. LIME (Xiaojie Guo & Li, and Haibin Ling., 2016) proposes to refine the initial illumination obtained by the Max-RGB assumption and the structure preserving constraint. Though these methods can somewhat brighten low-light images, they rarely take hidden degradation, like noise and color distortion, into account. To alleviate the noise effect, SRIE (Fu et al., 2016) further imposes the sparsity on the reflectance. Similarly, RRM (Li et al., 2018) integrates noise estimation into the optimization to produce desired reflectances. But, the applicability of these optimization-based methods is limited by the expensive computation, sensitivity to hyper-parameters, and unsatisfied enhancement quality.

2.2 Deep Learning-Based Methods

Recently, methods based on deep learning have dominated the target task. For instance, MSR-Net (Shen et al., 2017) integrates the MSR mechanism into a neural network and uses the BM3D (Dabov et al., 2007) for denoising. It gains improvement in visual quality, but still suffers from the drawbacks inherited from the traditional MSR. LLNet (Lore et al., 2017) synthesizes image pairs by randomly applying Gamma adjustment on clean images and adding noise to the adjusted images. An auto-encoder network is then constructed to learn the mapping function. LightenNet (Li et al., 2018) estimates the illumination, which is directly separated from the input. Ren et al. (2019) design an end-to-end network combining an encoder-decoder structure for content enhancement and a recurrent network for edge enhancement, while Fan et al. (2020) unite a semantic segmentation with the Retinex-based model, aiming to obtain semantic clues for the illumination adjustment. Moreover, DRBL (Yang et al., 2020) develops a deep recursive band network for semi-supervised low-light enhancement. DLN (Wang et al., 2020) introduces lighten-darken trade-off and feature aggregation blocks to ameliorate results. LPNet (Li et al., 2020) leverages a pyramid network with a carefully designed illumination loss. DSLR (Lim & Kim, 2020) models the Laplacian pyramid with a deep neural network to handle the darkness. Despite reasonable results, the aforementioned methods cannot effectively alleviate the noise and color distortion issues. Progressive Retinex (Wang et al., 2019) recurrently estimates illumination maps and noise levels to address the light and noise problems, but the color distortion is omitted. KinD and KinD++ (Zhang et al., 2021) resort to the layer decomposition strategy for better illumination adjustment and reflectance refinement, which further suppress noise in the reflectance layer. Xu et al. (2020) decouple an image into low-frequency and high-frequency layers for reducing noise and adding details, respectively. TBEFN (Kun & Zhang, 2020) fuses the enhanced and denoised results from two branches. Nevertheless, the relationship between the change in illumination and degradation is barely considered by most of the mentioned methods. Therefore, the noise residue and over-smoothing problems remain.

In unsupervised settings, EnlightenGAN (Jiang et al., 2021) takes advantage of larger-scale unpaired training data through employing the GAN mechanism, while Zero-DCE (Guo et al., 2020) alternatively learns a set of non-reference loss functions to enhance low-light images. ExCNet (Zhang et al., 2019) utilizes a block-based loss function, motivated by the Markov Random Field (Liu et al., 2017). It learns the mapping relationship online between images and their best "S-curve" parameters. Although relieving the requirement of paired data, they mainly focus on the light factor, and thus have insufficient abilities to address other defects.

Furthermore, several works have been devoted to process RAW images. SID (Chen, 2018), as a representative, introduces a dataset of raw image pairs of short-exposure (low-light) images and their long-exposure references. An end-to-end network for enhancing low-light images is then trained on the paired data. Schwartz et al. (2018) build a RAW-to-RGB pipeline based on the dataset captured by smartphone cameras. Ignatov et al. (2020) build another dataset with RAW-RGB image pairs captured by a camera phone and a professional photographing device, respectively. Considering that a group of images with multiple exposures and identical content can ease the problem, Cai et al. (2018) exploit to construct references from multi-exposure sequences for single image enhancement. However, its performance is upper-bounded by involved multi-exposure fusion methods. Zhu et al. (2020) synthesize multi-exposure images through changing the exposure ratio of a low-light input and adopting an edge detector to enhance the edge information of fused results. Despite a progress made toward addressing the problem, effective and efficient designs for handling complex and multi-entangled degradation are desirable for practical use.

3 Problem Analysis and Motivation

Low-light image enhancement typically encounters complex degradation mixed up by dim light, noise, and color distortion. Most previous methods ignore the relationship between the amplification of degradation and the illumination adjustment (enhancement). In other words, the defects hidden in the darkness will burst in final or intermediate results, which in nature correlate to the spatially-variant illumination. This fact results both end-to-end networks like (Wang et al., 2018) and post-processing methods like (Shen et al., 2017) in artifacts

left in texture and/or color. The problem would be simplified if images can be split into different components, each of which is responsible for one type of degradation.

Suppose we are already able to decompose an image I into a texture factor F_{tex} and a color factor F_{col} by a function $\mathcal{D}(\cdot)$, *i.e.*, $(F_{tex}, F_{col}) = \mathcal{D}(I)$. Inversely, there is a corresponding composition function $\mathcal{C}(\cdot)$ that can map F_{tex} and F_{col} back to the original I, say $I = \mathcal{C}(F_{tex}, F_{col})$. Concretely, for the texture factor F_{tex}, it reflects the image content, which plays a similar role as the reflectance in the Retinex model. It is of stronger contrast in images captured under properly-illuminated conditions than that in low-light environments. According to Eqs. (1), (2), and (3), we reach the following:

$$F_{tex}^{low} = F_{tex}^{high} \circ L^{low} + E_N = (F_{tex}^{high} + \tilde{E}_N) \circ L^{low}, \quad (4)$$

where F_{tex}^{low} is the texture of the low-light image I^{low} and F_{tex}^{high} represents the texture of the corresponding high-quality reference I^{high}. In addition, L^{low} denotes the illumination of I^{low}, E_N designates the noise term, while $\tilde{E}_N = E_N / L^{low}$ means the spatially-variant noise. Once F_{tex}^{high} and L^{low} are recovered from I^{low}, we can obtain various clean versions of texture F_{tex}^{adj} with arbitrary levels of illumination via:

$$F_{tex}^{adj} = F_{tex}^{high} \circ L^{adj}. \quad (5)$$

The illumination L^{adj} is generated by turning L^{low} up or down. As for the color factor F_{col}, it should vary with respect to the change in illumination, which can be formally written as follows:

$$F_{col}^{adj} = \Theta_{CA}(F_{col}^{low}, F_{tex}^{low}, F_{tex}^{adj}), \quad (6)$$

where $\Theta_{CA}(\cdot)$ stands for a trainable network (function) for color adjustment. Having F_{tex}^{adj} and F_{col}^{adj}, the enhanced result I^{adj} can be immediately formed by the composition function $\mathcal{C}(F_{tex}^{adj}, F_{col}^{adj})$.

Now, we begin with the first issue, *i.e.*, how to decompose an image into the texture and color factors. One intuitive thought is to build a network to accomplish the task. But, no well-defined ground-truth data available hinders training such a network. Alternatively, we resort to convert images from the RGB colorspace into a luminance-chrominance (similar to F_{tex}-F_{col}) one. Off-the-shell colorspace converters, such as RGB2HSV and RGB2YCbCr, are simple and efficient, which can perform as $\mathcal{D}(\cdot)$. Their inverse functions (act as $\mathcal{C}(\cdot)$) are also available. A question may be asked that the mentioned colorspace converters just simply transform the RGB information, why one can work better than another? To answer the question, we visualize how the noise distributes in different colorspaces and make statistics on the effect of single channels in restoration on the LOL training dataset

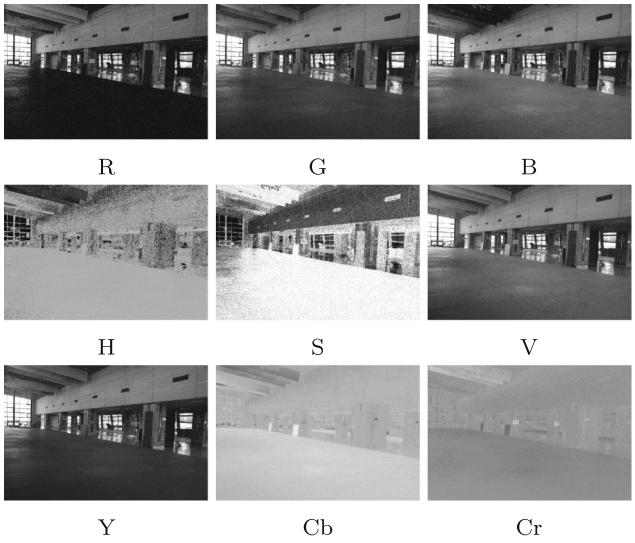

Fig. 2 An image from the LOL dataset, which is low-light before being linearly enhanced. As can be observed in its RGB, HSV and YCbCr channels, the chrominance components of YCbCr are obviously less affected by degradation than the others

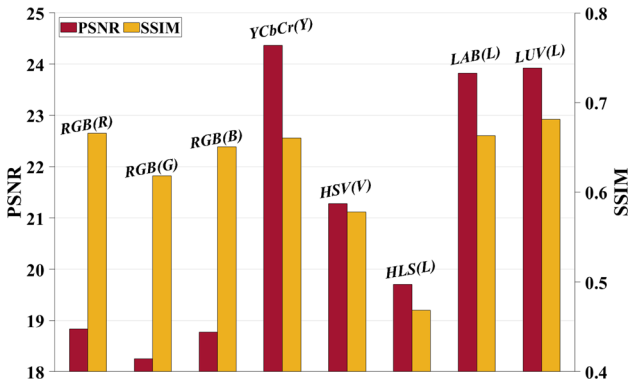

Fig. 3 The illuminations of low-light images in the LOL training dataset are first aligned with their references. We then replace one of the channels in different colorspaces with the corresponding ground-truth to see how much information a single channel statistically contains using the averaged PSNR and SSIM

(Wei et al., 2018), in Figs. 2 and 3. From the evidence, we can see that, although the colorspace transformations are linear, the information flows into different channels. The results suggest that the YCbCr space seems to be a "good" candidate to do the texture-color decomposition. Please notice that, other options might be also adopted. The better the decomposition, the easier the subsequent learning.

Next step is to handle the dim light and noise in the luminance (texture). For ease of exposition, we follow the notation F_{tex} to express the luminance (*i.e.*, Y in YCbCr). According to Eqs. (4) and (5), we acquire the texture with desired light by:

$$F_{tex}^{adj} = \frac{(F_{tex}^{low} - E_N) \circ L^{adj}}{L^{low}} = \frac{F_{tex}^{low}}{\hat{L}} - \frac{E_N}{\hat{L}}, \quad (7)$$

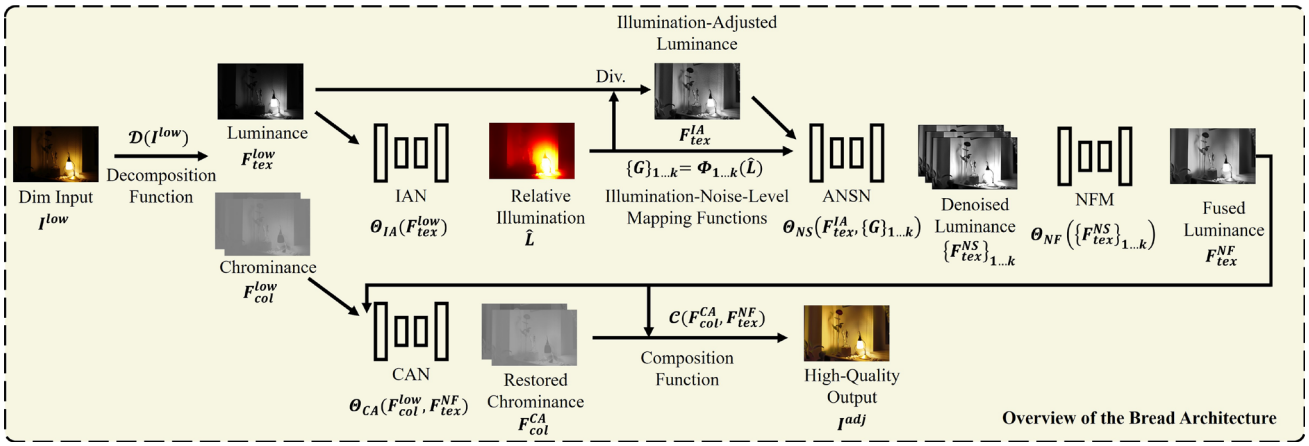

Fig. 4 The overall network architecture of our proposed Bread framework

where $\hat{L} = L^{low}/L^{adj}$ represents the relative illumination between the low-light and desired-light images with the element-wise division. To seek $F_{tex}^{IA} = F_{tex}^{low}/\hat{L}$, we construct an illumination adjustment network as a step forward towards achieving F_{tex}^{adj}. Further, we estimate the noise term $\tilde{E}_N = E_N/\hat{L}$ in order to eliminate it from F_{tex}^{IA}. However, even if E_N is simple, $\tilde{E}_N$ will become much more complicated due to its correlation with the spatially-variant illumination. Thus, it is natural to adopt $\hat{L}$ as an indicator for denoising. Barely previous methods take into account this property. KinD (Zhang et al., 2019) refines the reflectance by a simple concatenation of the estimated illumination and rough reflectance. This practice may not sufficiently exploit the guidance information from the illumination. And all the degradation needs to be simultaneously handled by a single restoration network. Alternatively, to be robust against the spatially-variant noise, we synthesize noisy images that have the same illumination with references but are corrupted with stimulated amplified noise guided by $G = \Phi(\hat{L})$, where $\Phi(\cdot)$ is a function that reflects the relationship between illuminations and noise levels ($\exp(-\hat{L})$ in this work). We will further discuss and compare the different forms of $\Phi(\cdot)$ and noise synthesis methods in Sec. 4.3. By this means, we can obtain a noise-suppression solver constrained by G. Though the relative noise level map is determined, its overall scale is inaccessible. To be adaptive, we further fuse the denoised luminances F_{tex}^{NS} under different suppression strengths to obtain the expected luminance map F_{tex}^{NF}. Having F_{tex}^{NF} estimated, it is reasonable to employ it as guidance for mapping chrominance (Cb and Cr in the YCbCr colorspace, corresponding to F_{col}) from the input to an adjusted/target light level, which is achieved by our color adaption network.

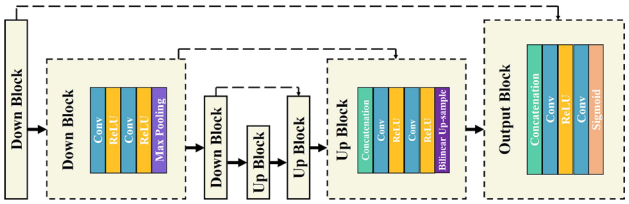

Fig. 5 The U-shaped architecture of our sub-networks, including IAN, ANSN, NFM, and CAN

4 Methodology

4.1 Overall Network

Figure 4 gives the overall architecture of our Bread, which mainly comprises an illumination adjustment net (IAN, Θ_{IA}), an adaptive noise suppression net (ANSN, Θ_{NS}), and a color adaption net (CAN, Θ_{CA}). As can be seen from Fig. 4, we firstly convert the input from the RGB to the luminance-chrominance colorspace, obtaining the luminance F_{tex}^{low} and chrominance F_{col}^{low} components. Then, F_{tex}^{low} is fed into the IAN to predict the relative illumination map $\hat{L}$, yielding the adjusted F_{tex}^{IA}. The ANSN takes in F_{tex}^{IA} and $\{G\}_{1...k} = \Phi_{1...k}(\hat{L})$ to produce $\{F_{tex}^{NS}\}_{1...k}$, where k is the number of different suppression strengths. The denoised luminance maps $\{F_{tex}^{NS}\}_{1...k}$ are then fused by an extra noise fusion module (NFM, Θ_{NF}), which produce F_{tex}^{NF}. The obtained luminance F_{tex}^{NF} is consequently utilized as the guidance of the chrominance enhancement. By taking the chrominance component F_{col}^{low} of the low-light image and F_{tex}^{NF}, the CAN outputs the adjusted color F_{col}^{CA}. Finally, we convert F_{tex}^{NF} and F_{col}^{CA} back to the RGB colorspace via the composition function, expressed as $I^{adj} = \mathcal{C}(F_{tex}^{NF}, F_{col}^{CA})$.

To largely exclude influence from other factors and verify the main claim, all the sub-networks follow the simple U-shaped architecture as depicted in Fig. 5. There are three down-sampling layers and three up-sampling layers in each sub-network. Two convolutional layers with ReLU activations are inserted before each scaling layer. The inner channels of each block are doubled after down-sampling and halved after up-sampling, the amount of which changes between 32 and 128. All the convolution layers employ 3×3 kernel size, stride = 1 and padding = 1. All the sub-networks are equipped with Sigmoid activation at their tails, except for the ANSN that is merely the convolution. *Note that our sub-networks are trained individually.*

4.2 IAN: Illumination Adjustment Network

IAN is proposed to estimate the relative illumination between the low-light luminance F_{tex}^{low} and the high-quality counterpart F_{tex}^{high}. The loss function used by IAN is as follows:

$$\begin{aligned}\mathcal{L}_{IA} = &\left\| F_{tex}^{low}/(\hat{L}+\epsilon) - F_{tex}^{high} \right\|_2^2 \\ &+ \alpha \left\| W \circ \nabla \hat{L} \right\|_1 + \beta \left\| \nabla \hat{L} - \nabla F_{tex}^{low} \right\|_1,\end{aligned} \tag{8}$$

where $\|\cdot\|_1$ and $\|\cdot\|_2$ respectively stand for the ℓ_1 and ℓ_2 norms, and ϵ is a small constant for avoiding zero denominator. In addition, α and β are two coefficients to balance different terms. $\hat{L} = \Theta_{IA}(F_{tex}^{low})$ is the predicted relative difference of illumination between F_{tex}^{low} and F_{tex}^{high}. The second term is to constrain the illumination to be piece-wise smooth in cooperation with the weight map $W = 1/(\nabla F_{tex}^{low} + \epsilon)$. The last term is to preserve the similarity between F_{tex}^{low} and $\hat{L}$ in the gradient domain for reducing halo artifacts. ∇ designates the first-order derivative filter. Once $\hat{L}$ is acquired, it is used to generate the output of the first stage through $F_{tex}^{IA} = F_{tex}^{low}/(\hat{L}+\epsilon)$.

4.3 ANSN: Adaptive Noise Suppression Network

According to Eq. (7), the obtained F_{tex}^{IA} actually combines F_{tex}^{adj} and $\tilde{E}_N$, in which the noise is also amplified. Moreover, possible errors in the estimated illumination should be prevented from flowing to the subsequent process. Therefore, we simulate amplified noise $\tilde{E}_N$ on references F_{tex}^{high} without changing their illuminations. Regions originally under darker light shall have more intense noise than those under brighter conditions. It is reasonable to adopt the relative illumination map $\hat{L}$ estimated previously as the indicator of noise level. The choices of noise pattern are investigated in LLNet (Lore et al., 2017), which suggests that Gaussian and Poisson noises are competent. In this work, we take the simple Gaussian model for noise synthesis. Modified from the traditional AWGN model (Zhang et al., 2017), we have the following for noise synthesis:

$$F_{tex}^{noisy} = F_{tex}^{high} + \mathcal{N}(0, G), \tag{9}$$

where G can be viewed as an attention/guidance map in inverse proportion to $\hat{L}$. Moreover, a simple loss function is used for ANSN as below:

$$\mathcal{L}_{NS} = \left\| \Theta_{NS}(F_{tex}^{noisy}, G) - \mathcal{N}(0, G) \right\|_2^2. \tag{10}$$

With the trained ANSN, the amplified noise in F_{tex}^{IA} can be removed through $F_{tex}^{NS} = F_{tex}^{IA} - \Theta_{NS}(F_{tex}^{IA}, G)$.

Although ANSN can produce noise-free enhancement results for most cases captured from the same device, the noise strength varies for different photographing devices in practice, which leaves a variable to be determined during processing different images. Moreover, the estimation of G might be imperfect, as no ground-truth information is available for supervision. To robustly remove the noise, we develop a noise fusion module (NFM) to merge the luminance maps $\{F_{tex}^{NS}\}_{1...k}$ at k different denoising strengths of $\{G\}_{1...k}$. The fusion operator is expected to further remedy errors caused by different cameras and/or left by the previous stages via minimizing the following objective:

$$\mathcal{L}_{NF} = \left\| F_{tex}^{NF} - F_{tex}^{high} \right\|_2^2 - \mathrm{SSIM}(F_{tex}^{NF}, F_{tex}^{high}), \tag{11}$$

where $F_{tex}^{NF} = \Theta_{NF}(\{F_{tex}^{NS}\}_{1...k})$, and $\mathrm{SSIM}(\cdot,\cdot)$ is the structural similarity loss.

Moreover, in the testing phase, the ground-truth illuminations of inputs are unavailable and thus there is no need to fix the errors in the illumination estimation. The training strategy of our NFM is accordingly modified for these non-reference images. In detail, the training data of the NFM is replaced by synthesized pairs $(F_{tex}^{noisy}, F_{tex}^{high})$. The objective for training NFM on them is expressed as:

$$\mathcal{L}_{NF(T)} = \left\| F_{tex}^{NF(T)} - F_{tex}^{high} \right\|_2^2 - \mathrm{SSIM}(F_{tex}^{NF(T)}, F_{tex}^{high}), \tag{12}$$

where $F_{tex}^{NF(T)} = \Theta_{NF}(\{F_{tex}^{NS(T)}\}_{1...k})$ and each $F_{tex}^{NS(T)}$ is obtained by $F_{tex}^{NS(T)} = F_{tex}^{noisy} - \Theta_{NS}(F_{tex}^{noisy}, G)$. Notably, $\mathcal{N}(0, A) = F_{tex}^{noisy} - F_{tex}^{high}$ is the synthesized noise map, while $F_{tex}^{IA} - F_{tex}^{high}$ contains both real noise and illumination errors. Therefore, the NFM trained on the synthesized pairs, namely NFM-T, disregards illumination errors that are specific to the training dataset to generalize better to testing cases. As shown in Table 2, we design Bread-T equipped with NFM-T for non-reference images. In addition, we visualize the respective textures produced by NFM and NFM-T

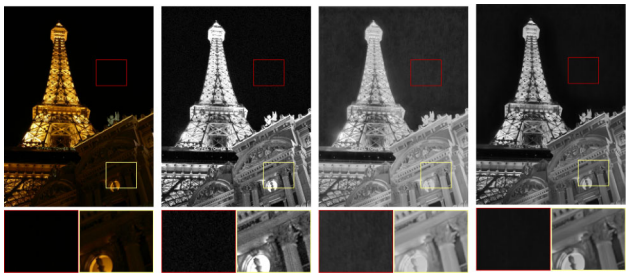

Fig. 6 Visual comparison between Bread (F_{tex}^{NF}) and Bread-T ($F_{tex}^{NF(T)}$) on a testing case

in Fig. 6. It can be seen that the night sky region in F_{tex}^{NF} is mistakenly brightened, making the image unrealistic, while for $F_{tex}^{NF(T)}$ only noise is suppressed, producing cleaner texture.

Figure 7 visualizes our noise synthesis results. To illustrate their relationship with real data, we provide an example of illumination adjustment to multi-exposure images. The first row in the left exhibits the synthesized multi-exposure inputs through adjusting the exposure of the left-most texture. The corresponding relative illuminations predicted by IAN are displayed in the second row. For training ANSN, we simulate amplified noise $\tilde{E}_N$ on references F_{tex}^{high} as shown in the third row. The simulated noise is of zero mean and variance determined by the noise level map G. In this work, we employ the scaled negative exponential function $G = \gamma \cdot \exp(-\hat{L})$ due to its simplicity, which plays a similar role to attention. Other manners can be investigated and testified. A sequence of multi-exposure images are given in the first row in the right. The images $I_1 \sim I_4$ are captured by increasing shutter time, hence their relative illumination maps $\hat{L}$ go brighter. After dividing $\hat{L}$ from the textures of their corresponding images, we obtain the results as depicted in the last row. As can be seen, the adjusted textures contain noise of different levels according to their respective illumination. A brighter illumination usually corresponds to a cleaner texture, and vice versa. Please notice the results of the last row, their responses to light changes are consistent. Moreover, we adopt G as the attention for the noise suppression network. As a result, we can control the denoising strength of ANSN by adjusting the relative scale of G and fuse the denoised results under different strengths to acquire more promising results.

4.4 CAN: Color Adaption Network

Having obtained F_{tex}^{NF}, the color components are expected to be adapted accordingly. Given the luminance and chrominance components of the original input (F_{tex}^{low}, F_{col}^{low}) and the reference (F_{tex}^{high}, F_{col}^{high}), the loss function for CAN is designed as follows:

$$\mathcal{L}_{CA} = \left\| F_{col}^{CA} - F_{col}^{high} \right\|_2^2 \tag{13}$$

where $F_{col}^{CA} = \Theta_{CA}(F_{col}^{low}, F_{tex}^{low}, F_{tex}^{high})$. Note that, following the divide-and-rule principle, we use F_{tex}^{high} rather than F_{tex}^{NF} during training to avoid the influence from possible errors left in previous stages. Once the network is trained, F_{tex}^{high} is replaced with F_{tex}^{NF} for testing.

Considering the low-saturation problem of the references in the LOL (Wei et al., 2018) training set as shown in Fig. 8, we alternatively introduce multi-exposure data to train the CAN by $F_{col}^{ME} = \Theta_{CA}(F_{col}^{e1}, F_{tex}^{e1}, F_{tex}^{e2})$, for the sake of covering a wide range of exposure and aplenty color patterns. Image pairs under different exposures, denoted as (F_{tex}^{e1}, F_{col}^{e1})

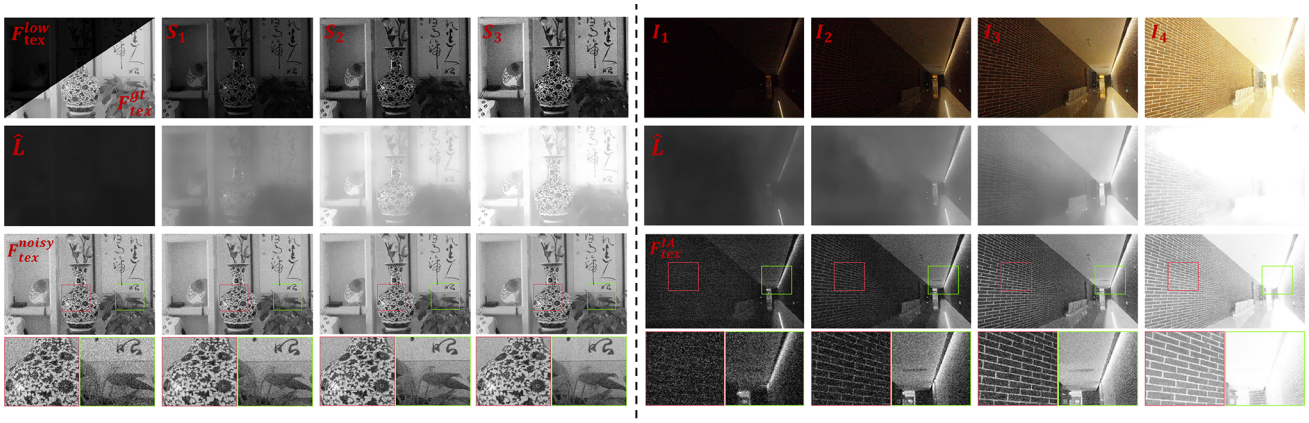

Fig. 7 Left: We adjust the exposure rate of F_{tex}^{low} to produce images $S_1 \sim S_3$. Their relative illumination maps are shown in the second row. The third row exhibits the synthesized noisy images guided by $\hat{L}$. Right: The first row displays a sequence of multi-exposure images, captured by increasing shutter time from I_1 to I_4. The second row visualizes the relative illumination predicted by a trained IAN. The illumination-adjusted textures of $I_1 \sim I_4$ are shown in the last row

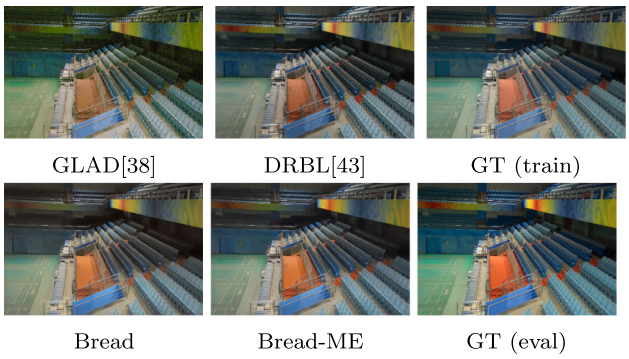

Fig. 8 The ground truths of training data in the LOL dataset are typically of less vivid color than those images in the evaluation set

and (F_{tex}^{e2}, F_{col}^{e2}), are randomly selected from the sequences of multi-exposure photographs. The objective function of this module shares the same form with Eq. (13). It forces the enhanced chrominance components to fit the exposure condition with respect to the luminance given by real multi-exposure data. After pretraining the CAN on multi-exposure data, we finetune it with the pairs from the LOL dataset for the task of low-light image enhancement. The low/high-light image pairs (F_{tex}^{low}, F_{col}^{low}) and (F_{tex}^{high}, F_{col}^{high}) can be seen as the multi-exposure image pairs during finetuning. Our approach also answers a key question that arises in the enhancement – to what extent the enhanced color should be. In most, if not all, of the cases, we expect it to be natural, under a certain exposure. The framework with multi-exposure data introduced is denoted as Bread-ME. We will show results of both Bread and Bread-ME in experiments for fair comparison.

5 Experiments

5.1 Implementation Details

All of our models are implemented in PyTorch and optimized with Adam optimizer, the parameters of which are set as $\beta_1 = 0.9$, and $\beta_2 = 0.999$. All the learning rates are fixed to 10^{-3}, except for the finetuning of the color adaption network based on multi-exposure data, which is set to 10^{-4}. α and β in Eq. (8) are set to 4.0 and 0.5, respectively. We adopt $G = \gamma \cdot \exp(-\hat{L})$ for the ANSN, while employing $k = 3$ and $\gamma_{1,2,3} = \{0, 0.05, 0.1\}$ for the NFM. We employ the LOL dataset as the training data, which contains 485 low/normal light pairs for training and 15 pairs for evaluation. To imitate various exposures in real-world photographs, we synthesize 8 images with different exposures for each low-light image. The magnitudes of exposure are uniformly distributed from 1 to the value with respect to 25% pixels being over-exposed at most. For training the color adaption network on multi-exposure data, we randomly select two images from each sequence of multi-exposure photographs from the SICE dataset (Cai et al., 2018). Note that the color adaption is bidirectional, *i.e.*, it can be either from under-exposed to over-exposed or vice versa. Consequently, we append the 485 pairs from the LOL dataset into the training data.

5.2 Performance Evaluation

To verify the effectiveness of our proposed method, several public datasets are used for evaluation, including LOL (Wei et al., 2018), DICM (Lee et al., 2013), NPE (Wang et al., 2013) and VV.[1] Representative state-of-the-art methods, including LIME (Xiaojie Guo & Li, and Haibin Ling., 2016), SIRE (Fu et al., 2016), NPE (Wang et al., 2013), RRM (Li et al., 2018), MSRNet (Shen et al., 2017), DUPE (Wang et al., 2019), GLAD (Wang et al., 2018), DRD (Wei et al., 2018), DRBL (Yang et al., 2020), KinD (Zhang et al., 2019), EG (Jiang et al., 2021) and ZDCE (Guo et al., 2020) are involved in comparisons. Image quality assessment metrics, including PSNR, SSIM, NIQE (Mittal et al., 2013), DeltaE (Sharma et al., 2005), are employed to measure performance.

5.2.1 Comparison on the LOL Evaluation Dataset

We report the quantitative comparison between Bread(-ME) and the other competitors on the LOL evaluation dataset in Table 1. Our method outperforms other state-of-the-art alternatives by a noticeable margin in both of the reference and no-reference metrics, demonstrating the efficacy of our proposed Bread framework. Because the absolute brightness is inaccessible during evaluation, it may be unfair to those non-data-driven methods and may interfere the assessment of fidelity in detail. For comparison fairness, we additionally align the predicted luminance to its reference image by simple Gamma correction. Visual comparison on several samples from the LOL evaluation is depicted in Fig. 9. The results by our method achieve remarkably higher performance with noise well-suppressed and less artifacts, while irregular illumination, noise residual, and texture/color distortion exist in the results of the other competitors. Moreover, we can see from Fig. 9 that the color adapter trained on multi-exposure data rectifies the greenish hue, producing more realistic tone.

5.2.2 Comparison on Other Datasets

We offer the non-reference results on five commonly-used testing data-sets in Table 2. As can be seen, our Bread-T

[1] https://sites.google.com/site/vonikakis/datasets.

Table 1 Quantitative results on LOL evaluation dataset of different methods in terms of PSNR, SSIM, NIQE and DeltaE. The subscript C indicates Gamma correction on the luminance towards references is conducted before evaluation. The best results are indicated in bold, the second-best results in italic and the third in bolditalic

Metrics	LIME Xiaojie Guo and Li, and Haibin Ling. (2016)	SRIE Fu et al. (2016)	NPE Wang et al. (2013)	RRM Li et al. (2018)	EG Jiang et al. (2021)	ZDCE Guo et al. (2020)	MSRNet Shen et al. (2017)
PSNR↑	16.76	11.86	16.97	13.88	17.48	14.86	13.17
SSIM↑	0.444	0.494	0.482	0.670	0.654	0.562	0.460
NIQE↓	9.779	8.073	9.788	4.234	5.238	8.811	9.261
DeltaE↓	21.43	32.62	21.77	26.18	19.31	24.56	30.17
$PSNR_C$ ↑	19.14	20.97	20.59	20.25	22.48	21.88	16.71
$SSIM_C$ ↑	0.471	0.656	0.513	0.774	0.710	0.640	0.461
$NIQE_C$ ↓	8.954	7.321	8.890	3.944	4.837	7.889	9.074
$DeltaE_C$ ↓	19.06	14.32	17.60	14.03	13.28	14.01	22.72

Metrics	DUPE Wang et al. (2019)	GLAD Wang et al. (2018)	DRD Wei et al. (2018)	DRBL Yang et al. (2020)	KinD++ Zhang et al. (2021)	Bread	Bread-ME
PSNR↑	14.77	19.72	16.77	18.80	***21.80***	*22.92*	**22.96**
SSIM↑	0.470	0.685	0.428	***0.831***	*0.836*	*0.836*	**0.838**
NIQE↓	9.079	7.283	10.424	***4.103***	4.290	*3.950*	**3.946**
DeltaE↓	26.19	16.54	23.65	15.59	*11.52*	***11.54***	**11.19**
$PSNR_C$ ↑	22.36	23.72	18.73	22.48	***23.91***	*25.98*	**26.06**
$SSIM_C$ ↑	0.594	0.724	0.448	***0.851***	0.847	*0.851*	**0.857**
$NIQE_C$ ↓	8.110	6.754	9.644	***3.753***	3.901	*3.649*	**3.614**
$DeltaE_C$ ↓	14.77	12.54	21.49	11.75	***10.08***	*9.41*	**9.06**

The best results are indicated in bold, the second-best results in italic and the third in bolditalic

Fig. 9 Visual comparison on samples from the LOL dataset. Zoom in for details

Table 2 Quantitative comparison (NIQE ↓) on the DICM, NPE, MEF, NPE and VV datasets

Datasets	LIME	SRIE	NPE	RRM	EG	ZDCE	MSRNet
DICM (44)	3.6642	3.0951	3.4304	3.3186	*2.7586*	3.3182	3.3937
LIME (10)	4.7748	***3.8053***	4.1523	4.2052	*3.6339*	4.0989	4.0872
MEF (16)	3.8765	3.2192	3.5884	3.9814	**2.8870**	3.3264	3.4077
NPE (8)	3.8422	**3.1788**	3.4455	3.9466	***3.3363***	3.6269	3.7255
VV (24)	4.4784	3.7925	4.2132	***3.3751***	*3.0363*	4.1915	4.1313
Datasets	DUPE	GLAD	DRD	DRBL	KinD++	Bread-ME	Bread-T
DICM (44)	3.1628	3.0875	4.7120	3.2784	***2.8584***	3.0869	**2.7302**
LIME (10)	3.8324	4.3366	5.2472	4.3352	5.1111	4.3993	**3.6145**
MEF (16)	***3.1025***	3.1994	5.0047	3.4632	3.3040	3.6931	*2.9748*
NPE (8)	*3.3327*	3.6143	4.0676	3.5843	3.9101	3.4596	3.5501
VV (24)	4.5100	4.4112	4.1508	3.3770	3.8084	3.6710	**2.8734**

reaches the state-of-the-art NIQE values on most of the testing datasets. Further, we supply additional visual results in Figs. 10 and 11 to reveal our effectiveness of suppressing noise and tuning color. It shows that the results provided by Bread-T are of lower noise levels, better visibility and fewer artifacts over the alternatives.

5.2.3 Comparison in Running Time

To meet real-time requirements, we design a lighter model following the Bread framework, namely Bread-ME-S, with only 0.21M parameters. In addition, previous real-time works (Tan et al., 2020; Lin et al., 2021; Ge et al., 2021) make use of FP16 precision for higher inference speed. We adopt this setting for our framework in this part. The comparisons in the number of parameters, inference time and quantitative performance between different low-light image enhancement models are provided by Table 3. It shows that our proposed Bread-ME-S keeps fewer parameters, faster inference speed and higher numerical performance compared with former supervised methods, which highlights the efficiency of our proposed framework. It is worth noting that our Bread-ME-S spends only 28.5 ms (35 FPS) to process a 480p color image on an RTX2080Ti GPU with FP16 precision. Moreover, we visualize the results of recent supervised models in Fig. 12. It can be seen that though our proposed Bread-ME-S has less parameters, it can still conquer the emerging noise. See also the results provided by the alternatives. Those models commonly have trouble handling the amplified noise or leave artifacts with even more parameters.

5.2.4 Comparison in Illumination and Reflectance Decomposition

We show the comparison in the illumination and reflectance decomposition between ours and Retinex-based alternatives in Fig. 13. As can be viewed from the visual results, the reflectance maps divided by our estimated relative illuminations are much more natural and accurate than those of DRD (Wei et al., 2018) and KinD (Zhang et al., 2019). In the framework of KinD, an extra illumination adjustment network is introduced to learn a map to counteract the irregular noise and halo existing in reflectances. Moreover, DRD shows poor overall performance due to the lack of such a mechanism.

5.3 Ablation Study

To validate the necessity of the usage of multi-exposure data, please compare the results of Bread and Bread-ME in Tables 1 and 2. The visual comparisons are offered in Figs. 8, 9 and 14. To evaluate the effectiveness of different settings in our framework, we conduct ablation studies including the choice of colorspace, the hyper-parameters of IAN, the choice of noise synthesis pattern, and the necessity of NFM. Please note that, the sub-networks are trained by 256×256 patches in all experiments of our ablation study for the sake of training speed. The overall performance may be inferior to the setting of full-resolution training in Sec. 5.2.

5.3.1 The Choice of Colorspace

This work aims to handle the mixed-up degradation in low-light images in a divide-and-conquer way. The principle is to deal with the low light problem and the noise issue in the texture component, and then use the improved texture as the guidance to process the color component. To this end, we have studied the existing colorspace transformations as shown in Fig. 3. We also evaluate the performance of Bread under different colorspaces (*i.e.*, RGB, LAB, YUV and YCbCr) in Table 4. In detail, to determine which colorspace stores the largest part of noise in their texture component, we

Fig. 10 Visual comparison between different methods on the DICM testing dataset. Please zoom in for details

first use Gamma correction to align the illumination of low light images with their respective references. Accordingly, the noise hiding in low-light regions is amplified. Then, we replace the texture components of these aligned images with the texture of their references in different colorspaces, respectively. After the above operations, we transform the images with replaced textures back to the RGB colorspace. It can be observed that the results in the YCbCr colorspace attain the highest PSNR in Fig. 3. Considering that the illumination maps of images are aligned with their references by Gamma correction, a higher PSNR of a colorspace means a larger amount of information is restored in its texture component. See also the best performance in numerical metrics achieved by the YCbCr in Table 4, which further supports our claim.

5.3.2 The Combination of Hyper-Parameters in IAN

The loss function for training IAN is defined by Eq. (8). The hyper-parameters (*i.e.*, α and β) of the last two terms control the relative smoothness of the predicted relative illumination and its gradient fidelity with the input texture. There should be a trade-off among the involved terms to preserve valid

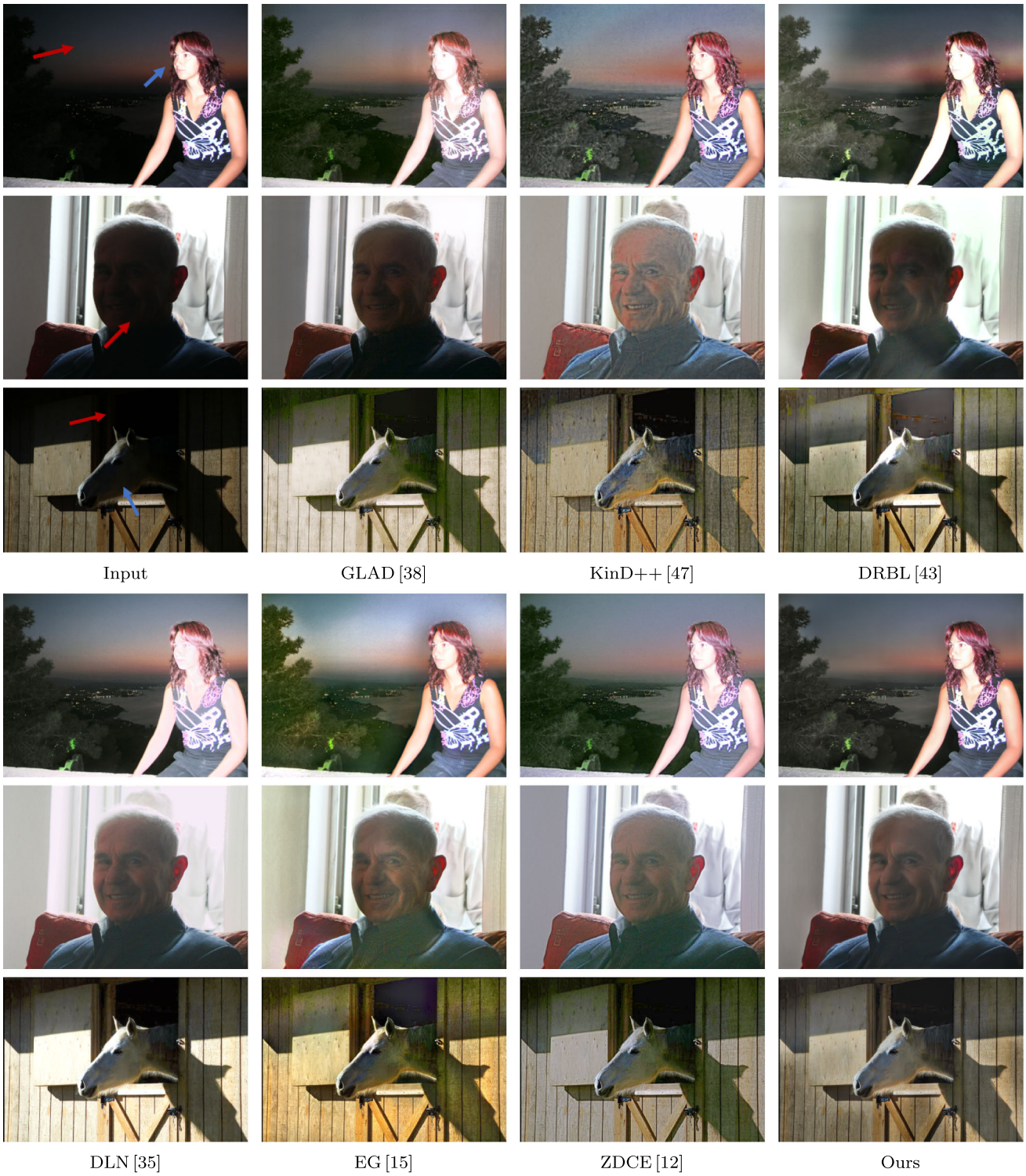

Fig. 11 Visual comparison between different methods on the VV (top two rows) and MEF-DS (Fang et al., 2020) (the last row). Please zoom in for details

Table 3 Comparison in the number of parameters (#params.), inference time (Infer. Time) and quantitative performance between supervised models. The inference times are taken by averaging the feed forward time of a 480p image for 100 times on an RTX2080Ti GPU. *-FP16 means the framework is conducted with FP16 precision. The best results are indicated in bold, the second-best results in italic and the third in bolditalic

Model Name	#params. (M) ↓	Infer. Time (ms)↓	PSNR↑	SSIM↑	$PSNR_C$ ↑	$SSIM_C$ ↑
DRD	*0.42*	***44.9***	16.77	0.428	18.73	0.448
GLAD	***0.89***	48.5	19.72	0.685	23.72	0.724
DRBL	5.27	63.9	18.80	***0.831***	22.48	*0.851*
KinD++	7.89	510.8	21.80	*0.836*	23.91	***0.847***
Bread-ME	2.02	62.7	*22.96*	**0.838**	**26.06**	**0.857**
Bread-ME-FP16	2.02	46.3	**22.97**	0.822	*26.00*	0.833
Bread-ME-S	**0.21**	*33.6*	***22.93***	0.827	***25.16***	0.842
Bread-ME-S-FP16	**0.21**	**28.5**	***22.93***	0.809	25.15	0.820

image gradients for the illumination adjustment and smooth out the undesired gradients. Thus, we conduct an experiment as reported in Table 5, searching for a better combination of the two hyper-parameters. From the numerical results, ($\alpha = 4$ and $\beta = 0.5$) seems to be the best choice among different configurations, hence we adopt it as the default setting for all the experiments.

5.3.3 The Usage of NFM

The noise fusion module (NFM) is introduced to fuse the intermediate results denoised by a set of strengths, which aims at preserving more natural details and less noise in the fused texture. Table 6 reveals how the performance changes with respect to the usage of NFM and the number of intermediate denoised images. We can observe that without the NFM, the performance degrades noticeably, because the possible errors in IAN and ANSN are no longer corrected properly. Moreover, one may wonder how many levels of denoised images the NFM should fuse. Table 6 tells that $k = 3$ is a better choice than the alternatives, by considering both the effectiveness and the efficiency.

5.3.4 The Choice of Noise Synthesis Pattern

The pattern of our noise synthesis is represented as $\mathcal{N}(0, \Phi(\hat{L}))$ in general, where $\hat{L}$ is the relative illumination predicted by IAN. The scaled negative exponential function is employed for $\Phi(\cdot)$ in this work. To validate the effectiveness of our strategy, we conduct a comparison between ours and several different patterns. As reported in Table 7, the performance sharply decreases when disabling the noise suppression (denoted as *w/o* NS). Further, an end-to-end network for texture restoration (E2E) is built to emphasize the necessity of noise synthesis. The network is trained on the pairs of F_{tex}^{IA} and F_{tex}^{high}. However, the F_{tex}^{IA} recovered by E2E holds a gap to F_{tex}^{high} in illumination. The results of E2E implies that such an illumination gap is hard to be bridged by a simple denoising network. The gap is shrunk by training ANSN on the data synthesized in our proposed manner. Next, we investigate which noise synthesis pattern is more suitable to real cases. One candidate is the Poisson noise model (PN) (Lore et al., 2017). In detail, the illuminations of normal-light textures are first adjusted according to their low-light counterparts. Then, the dark regions are contaminated through the Poisson process. Finally, the illumination maps are divided from the corrupted textures to simulate the textures with amplified noise. As can be seen in Table 7, the PN setting outperforms the E2E, which confirms the benefit from the noise synthesis. Another option is the Additive White Gaussian Noise (AWGN), which has been widely used for the task of noise synthesis (Zhang et al., 2017; Lore et al., 2017). To compare this pattern with our proposed one, we design the following ablation study. Given an AWGN model with a fixed intensity in the testing phase, we train it at three levels ($\sigma^2 = 15, 25, 50$), respectively. The textures denoised by them are fused by the NFM. Notably, the number of network parameters in the AWGN model is three times as many as the other noise synthesis patterns have. This may be a reason why it outperforms the PN setting. However, its denoising strength is almost the same at every location in an image without consideration of the spatial change in illumination. We also involve the noise model of Brooks et al. (2019) (UPC) for the comparison. It is worth mentioning that the hyper-parameters of the UPC noise model depend on the statistical results of a specific dataset. Since the noise distribution is inaccessible in many cases, the distribution discrepancy often weakens the performance of UPC on a different dataset (*e.g.*, the LOL dataset in this paper).

5.3.5 Other Issues

Some previous methods (Lore et al., 2017; Yang et al., 2016) tend to exploit the Gamma correction to synthesize multi-

Fig. 12 Visual comparison between the models shown in Table 3

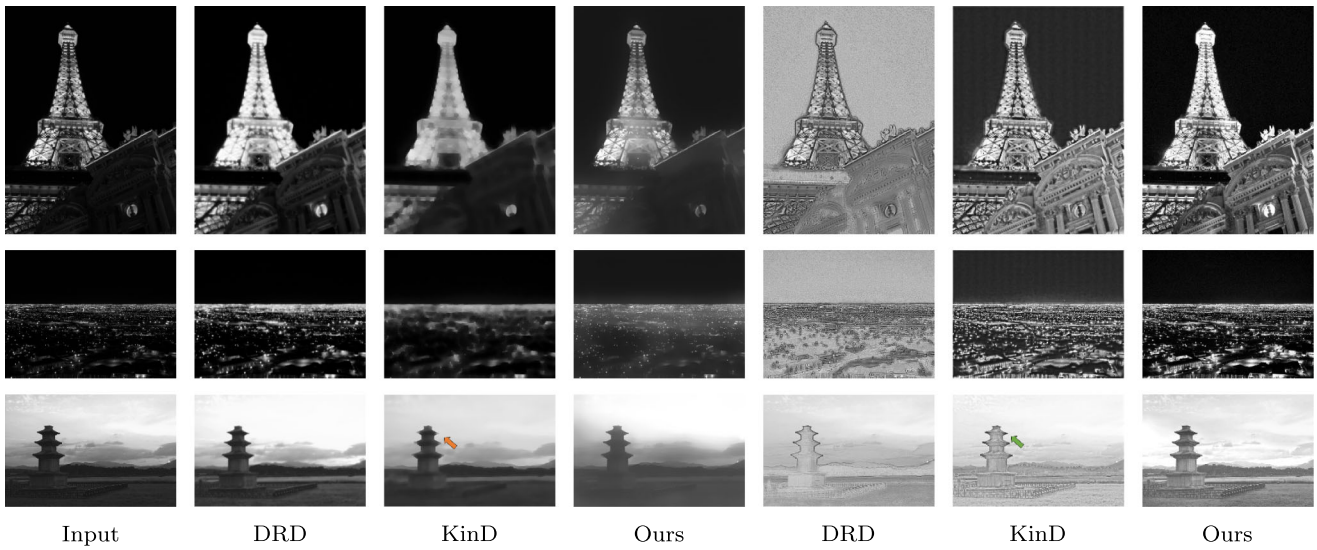

Fig. 13 Visual comparison on illumination (columns 2-4) and reflectance (columns 5-7) decomposition between ours and other Retinex-based methods recently proposed. Reflectances of other methods are converted to grayscale for the convenience of comparison

Fig. 14 Visual comparison between color adaption with and without multi-exposure data. The introduction of ME data contributes to more vibrant color

Table 4 Ablation study on the choice of colorspace. RGB (R) means using channel R as the texture channel (the other options are analogous)

	PSNR↑	SSIM↑	$PSNR_C$ ↑	$SSIM_C$ ↑
YCbCr (Y)	**22.05**	**0.834**	**25.57**	**0.856**
YUV (Y)	21.87	0.830	25.51	0.854
LAB (L)	21.72	0.823	24.83	0.855
RGB (R)	21.08	0.820	24.80	0.839
RGB (G)	20.33	0.833	24.04	0.852
RGB (B)	21.70	0.821	24.67	0.842

The best results are indicated in bold

Table 5 Ablation study on the choice of hyper-parameters of IAN

	PSNR↑	SSIM↑	$PSNR_C$ ↑	$SSIM_C$ ↑
$\alpha = 2, \beta = 0.50$	22.01	0.834	25.58	0.856
$\alpha = 4, \beta = 0.50$	**22.05**	**0.834**	**25.57**	0.856
$\alpha = 6, \beta = 0.50$	21.86	0.832	25.33	0.852
$\alpha = 8, \beta = 0.50$	21.95	0.832	25.48	0.853
$\alpha = 4, \beta = 0.25$	21.89	0.831	25.18	0.853
$\alpha = 4, \beta = 1.00$	21.98	0.830	25.60	0.851
$\alpha = 4, \beta = 1.50$	22.03	0.833	25.56	0.855
$\alpha = 4, \beta = 2.00$	21.93	0.833	25.44	**0.857**

The best results are indicated in bold

exposure data. But, as shown in Fig. 15, this manner is not suitable for our framework. In our supervised setting of training the IAN, $F_{tex}^{low}/F_{tex}^{high}$ are utilized as reference. Although the Gamma correction is capable of adjusting the illumination, it may hinder the IAN from learning rational illumination. Please pay attention to the second row in Fig. 15. We show the division of the multi-exposure textures with their references. The contrasts of the relative illumination maps are low. Moreover, the Gamma correction tends to over-enhance dark regions in images. This makes the results after

Table 6 Ablation study on the usage and the number of noise levels of the NFM. We test different configurations, including $k = 1$ ($\gamma = 0.1$), $k = 2$ ($\gamma = \{0, 0.1\}$), $k = 3$ ($\gamma = \{0, 0.05, 0.1\}$), $k = 4$ ($\gamma = \{0, 0.05, 0.1, 0.15\}$), $k = 5$ ($\gamma = \{0, 0.05, 0.1, 0.15, 0.2\}$) and $k = 6$ ($\gamma = \{0, 0.05, 0.1, 0.15, 0.2, 0.25\}$)

	PSNR↑	SSIM↑	$PSNR_C$ ↑	$SSIM_C$ ↑
w/o NFM	18.45	0.733	19.72	0.741
$k = 1$	19.44	0.789	24.38	0.821
$k = 2$	21.21	0.825	25.43	0.849
$k = 3$	22.05	**0.834**	**25.57**	**0.856**
$k = 4$	21.90	0.829	25.48	0.853
$k = 5$	**22.06**	0.831	25.73	0.855
$k = 6$	21.58	0.833	25.40	**0.856**

The best results are indicated in bold

Table 7 Ablation study on different patterns of noise synthesis, including no noise suppression (*w/o* NS), end-to-end network for restoration (E2E), using the patterns of Poisson Noise (PN), Additive White Gaussian Noise (AWGN), Brooks *et al.* (UPC) and ours

	PSNR↑	SSIM↑	$PSNR_C$ ↑	$SSIM_C$ ↑
w/o NS	16.91	0.586	23.33	0.623
E2E	18.69	0.785	24.58	0.829
PN	19.79	0.808	24.74	0.840
AWGN	21.36	0.832	25.25	0.851
UPC	21.82	0.825	25.42	0.845
Ours	**22.05**	**0.834**	**25.57**	**0.856**

The best results are indicated in bold

division much more noisy than those by exposure adjustment. The above factors make the Gamma correction fail to generate reasonable illumination maps. The results in Table 8 are the evidence supporting our analysis.

An extended version of LOL dataset with more low/ high image pairs has been constructed recently in Yang et al. (2021), namely LOL-v2. Considering that the performance

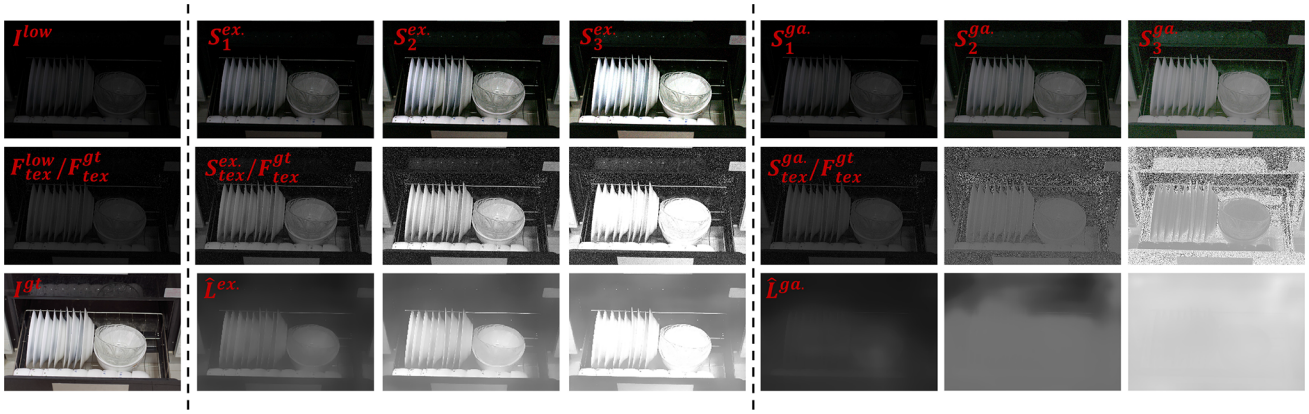

Fig. 15 Visual comparison on multi-exposure data synthesis between exposure adjustment and Gamma correction. Left: the low-light image I^{low}, its reference I^{high} and the division of their luminances. Middle: the examples of synthesized multi-exposure data produced by exposure adjustment of I^{low}. The division of their luminances and F_{tex}^{high} are visualized in the second row. The third row depicts the relative illumination $\hat{L}$ estimated by IAN that trained on synthesized sequences. Right: the results corresponding to Gamma correction

of our framework may be further improved by seeing more data, we retrain our model on the LOL-v2. For fair comparison, we keep the evaluation pairs the same, which are eliminated from the training dataset. As can be observed in Table 8, a larger dataset is somehow beneficial to our model in PSNR, but not in SSIM. This phenomenon reflects that our synthesis strategy likely reduces the reliance on the volume of data.

ANSN is proposed to cope with spatially variant noise, the spatial attention/guidance of which comes from the negative exponent of the estimated relative illumination $\hat{L}$ and a scaling factor γ. By adjusting γ, ANSN produces denoised results at different levels as visualized in Fig. 16. It is worth noticing that, guided by $\hat{L}$, the noise suppression concentrates more on darker regions, while brighter parts are less affected. To further remove residual noise while maintaining details, our proposed NFM assembles the denoised results in a fuse way. Please see the right-most picture in Fig. 16,

Table 8 Ablation study on settings of data synthesis and training datasets

	PSNR↑	SSIM↑	PSNR_C ↑	SSIM_C ↑
G.C.	21.74	0.827	24.85	0.849
LOL-v2	**23.29**	0.833	**26.06**	0.846
Ours	22.05	**0.834**	25.57	**0.856**

The best results are indicated in bold

Fig. 16 A sequence of denoised results of ANSN by adjusting the scaling factor of G and their fusion by NFM

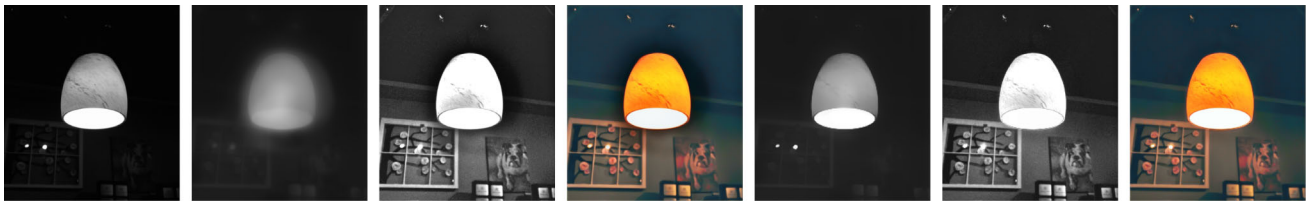

Fig. 17 Visual comparison on using loss term $\left\|\nabla\hat{L}-\nabla F_{tex}^{low}\right\|_1$ (denoted by $\mathcal{L}_g$) or not for training the IAN

in which the details are well-preserved with the noise sufficiently suppressed.

The last term of Eq. (8) confines the gradients of predicted relative illumination to those of the input texture. We introduce this term into the loss function of IAN training for reducing halo artifacts as shown in Fig. 17. We can see that, if discarding the term $\mathcal{L}_g$, $\hat{L}$ becomes blurry and its gradients are not consistent with those of the input texture.

6 Conclusion

This work discussed the mixture of multi-degradation in low-light images, which increases the training difficulty and limits the enhancement quality of previous methods. To disentangle the complex degradation, the colorspace is first converted from the RGB into a luminance-chrominance one, *i.e.*, YCbCr, from a texture-color perspective. By doing so, the main pressure of image brightening and denoising goes to the luminance component F_{tex}, while the chrominance component F_{col} responds to color correction, having the enhanced F_{tex} as guidance. Regarding different specific illnesses, the sub-networks including IAN, ANSN, and CAN were customized and trained individually, all of which follow the simple U-shaped net. Our designs make the training of each sub-network specific to one simple degradation, the effectiveness of which has been validated by the experiments. However, our framework still depends on the paired images to determine the target illuminations. This manner may lead to troubles when the framework generalizes to wider real-world datasets under various illumination conditions. At present, we make use of Bread-T to provide a better visual quality. However, in the further work, the unpaired version of the Bread may be proposed, which hopefully will have much stronger generalization ability. It is positive that such a divide-and-rule principle with texture-color decomposition can be applied to other enhancement and restoration tasks like dehazing and underwater image enhancement.

Acknowledgements The authors would like to thank the editors and reviewers for their effort in handling our submission, as well as the comments and suggestions that improve the quality of this paper. This work was supported by the National Natural Science Foundation of China under Grant no. 62072327.

References

Abdullah-Al-Wadud, M., Kabir, M. H., Dewan, M. A. A., & Chae, O. (2007). A dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics, 53*(2), 593–600.

Brooks, T., Mildenhall, B., Xue, T., Chen, J., Sharlet, D. & Barron, J. T. (2019) Unprocessing images for learned raw denoising. In CVPR, pages 11036–11045

Cai, J., Shuhang, G., & Zhang, L. (2018). Learning a deep single image contrast enhancer from multi-exposure images. *IEEE TIP, 27*(4), 2049–2062.

Celik, T., & Tjahjadi, T. (2011). Contextual and variational contrast enhancement. *IEEE TIP, 20*(12), 3431–3441.

Chen, C., Chen, Q., Xu, J., & Koltun, V.. (2018) Learning to see in the dark. In: CVPR, pages 3291–3300

Cheng, H.-D., & Shi, X. J. (2004). A simple and effective histogram equalization approach to image enhancement. *Digital signal processing, 14*(2), 158–170.

Dabov, K., Foi, A., Katkovnik, V., & Egiazarian, K. O. (2007). Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE TIP, 16*(8), 2080–2095.

Fan, M., Wang, W., Yang, W., & Liu, J. (2020) Integrating semantic segmentation and retinex model for low-light image enhancement. In ACM MM, pages 2317–2325

Fang, Y., Zhu, H., Ma, K., Wang, Z., & Li, S. (2020). Perceptual evaluation for multi-exposure image fusion of dynamic scenes. *IEEE TIP, 29,* 1127–1138.

Fu, X., Zeng, D., Huang, Y., Zhang, X.-P. & Ding, X. (2016) A weighted variational model for simultaneous reflectance and illumination estimation. In CVPR, pages 2782–2790

Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021) Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430

Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., & Cong, R. (2020) Zero-reference deep curve estimation for low-light image enhancement. In CVPR, pages 1780–1789

Guo, X., Li, Y., & Ling, H. (2016). Lime: Low-light image enhancement via illumination map estimation. *IEEE TIP, 26*(2), 982–993.

Ignatov, A., Van Gool, L., & Timofte, R.. (2020) Replacing mobile camera isp with a single deep learning model. In CVPR, pages 536–537.

Jiang, Y., Gong, X., Ding Liu, Yu., Cheng, C. F., Shen, X., Yang, J., et al. (2021). Enlightengan: Deep light enhancement without paired supervision. *IEEE TIP, 30*(2), 2340–2349.

Jobson, D. J., Rahman, Z., & Woodell, G. A. (1997). A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE TIP, 6*(7), 965–976.

Jobson, D. J., Rahman, Z., & Woodell, G. A. (1997). Properties and performance of a center/surround retinex. *IEEE TIP, 6*(3), 451–462.

Kun, L., & Zhang, L. (2020). Tbefn: A two-branch exposure-fusion network for low-light image enhancement. *IEEE TMM, 23*(2), 4093–4105.

Land, E. H. (1977). The retinex theory of color vision. *Scientific american, 237*(6), 108–129.

Lee, C., Lee, C., & Kim, C.-S. (2013). Contrast enhancement based on layered difference representation of 2d histograms. *IEEE TIP, 22*(12), 5372–5384.

Li, C., Guo, J., Porikli, F., & Pang, Y. (2018). Lightennet: A convolutional neural network for weakly illuminated image enhancement. *Pattern recognition letters, 104*(2), 15–22.

Li, J., Li, J., Fang, F., Li, F., & Zhang, G. (2020). Luminance-aware pyramid network for low-light image enhancement. *IEEE TMM, 23*(2), 3153–3165.

Li, M., Liu, J., Yang, W., Sun, X., & Guo, Z. (2018). Structure-revealing low-light image enhancement via robust retinex model. *IEEE TIP, 27*(6), 2828–2841.

Lim, S., & Kim, W. (2020). Dslr: Deep stacked laplacian restorer for low-light image enhancement. *IEEE TMM, 23*(2), 4272–4284.

Lin, S., Ryabtsev, A., Sengupta, S., Curless, B. L., Seitz, S. M. & Kemelmacher-Shlizerman, I. (2021) Real-time high-resolution background matting. In CVPR, pages 8762–8771

Liu, Z., Li, X., Luo, P., Loy, C. C., & Tang, X. (2017). Deep learning markov random field for semantic segmentation. *IEEE TPAMI, 40*(8), 1814–1828.

Lore, K. G., Akintayo, A., & Sarkar, S. (2017). Llnet: A deep autoencoder approach to natural low-light image enhancement. PR, 61(2):650–662

Mittal, A., Soundararajan, R., & Bovik, A. C. (2013). Making a completely blind image quality analyzer. *IEEE Signal Processing Letters, 20*(2), 209–212.

Pisano, E. D., Zong, S., Hemminger, B. M., Marla DeLuca, R., Johnston, E., Keith Muller, M., et al. (1998). Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital imaging, 11*(4), 193.

Ren, W., Liu, S., Ma, L., Qianqian, X., Xiangyu, X., Cao, X., et al. (2019). Low-light image enhancement via a deep hybrid network. *IEEE TIP, 28*(9), 4364–4375.

Schwartz, E., Giryes, R., & Bronstein, A. M. (2018). Deepisp: Toward learning an end-to-end image processing pipeline. *IEEE TIP, 28*(2), 912–923.

Sharma, G., Wencheng, W., & Dalal, E. N. (2005). The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research and Application, 30*(2), 21–30.

Shen, L., Yue, Z., Feng, F., Chen, Q., Liu, S., & Ma, J. (2017) Msr-net: Low-light image enhancement using deep convolutional network. arXiv preprint arXiv:1711.02488

Tan, M., Pang, R., & Le, Q. V. (2020) Efficientdet: Scalable and efficient object detection. In CVPR, pages 10781–10790

Wang, Y., Cao, Y., Zha, Z.-J., Zhang, J., Xiong, Z., Zhang, W., & Wu, F. (2019) Progressive retinex: Mutually reinforced illumination-noise perception network for low-light image enhancement. In ACM MM, pages 2015–2023

Wang, W., Wei, C., Yang, W., & Liu, Jiaying. (2018) Gladnet: Low-light enhancement network with global awareness. In IEEE International Conference on Automatic Face & Gesture Recognition, pages 751–755

Wang, R., Zhang, Q., Fu, C.-W., Shen, X., Zheng, W.-S. & Jia J. (2019) Underexposed photo enhancement using deep illumination estimation. In CVPR, pages 6849–6857

Wang, L.-W., Liu, Z.-S., Siu, W.-C., & Lun, D. P. K. (2020). Lightening network for low-light image enhancement. *IEEE TIP, 29*(2), 7984–7996.

Wang, S., Zheng, J., Hai-Miao, H., & Li, B. (2013). Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE TIP, 22*(9), 3538–3548.

Wei, C., Wang, W., Yang, W. & Liu, J. (2018) Deep retinex decomposition for low-light enhancement. In BMVC, page 155

Xu, Ke, Yang, Xin, Yin, Baocai, & Lau, Rynson WH. (2020) Learning to restore low-light images via decomposition-and-enhancement. In CVPR, pages 2281–2290

Yang, J., Jiang, X., Pan, C., & Liu, C. L. (2016) Enhancement of low light level images with coupled dictionary learning. In ICPR, pages 751–756

Yang, W., Wang, S., Fang, Y., Wang, Y., & Liu, J. (2020) From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In CVPR, pages 3063–3072

Yang, W., Wang, W., Huang, H., Wang, S., & Liu, J. (2021). Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE TIP, 30*(2), 2072–2086.

Zhang, Y., Zhang, J. & Guo, X. (2019).Kindling the darkness: A practical low-light image enhancer. In ACM MM, pages 1632–1640

Zhang, L., Zhang, L., Liu, X., Shen, Y., Zhang, S., & Zhao, S. (2019). Zero-shot restoration of back-lit images using deep internal learning. In ACM MM, pages 1623–1631

Zhang, Y., Guo, X., Ma, J., Liu, W., & Zhang, J. (2021). *Beyond brightening low-light images. IJCV, 129*(4), 1013–1037.

Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE TIP, 26*(7), 3142–3155.

Zhu, M., Pan, P., Chen, W. & Yang, Y. (2020). Eemefn: Low-light image enhancement via edge-enhanced multi-exposure fusion network. In AAAI, pages 13106–13113