Low-light Image Enhancement via Breaking Down the Darkness

Qiming Hu, Xiaojie Guo*

College of Intelligence and Computing, Tianjin University, Tianjin 300350, China

huqiming@tju.edu.cn xj.max.guo@gmail.com

Abstract

Images captured in low-light environment 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. Towards this purpose, we propose to convert an image from the RGB space 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 boosting 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-ofthe-art alternatives both quantitatively and qualitatively on several benchmark datasets. Our code is publicly available at https://github.com/mingcv/Bread.

1. Introduction

Capturing high-quality images under less controlled conditions is challenging especially using mobile devices. Often, people are shooting images in unsatisfactory light environment. For instance, we might take a photo against 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 extend exposure time to receive more information, but if the target scene is dynamic, the blur effect likely appears in captured images; another possible way is to set a flash for light compensation, which however frequently introduces unexpected



Figure 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.

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 [13] through decomposing an image I into its reflectance R and illumination L in the form of $I = R \circ L$, where \circ designates the Hadamard product operator. Because the reflectance component reflects 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. One can adjust illumination according to different demands. However, due to limited quality of sensors, the **noise** factor N will be always an annoying resident in im-

^{*}Corresponding Author



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



Figure 3. The illuminations of low-light images in the LOL training dataset are first aligned with their references and converted to different color-spaces. We replace one of their channels with the ground-truth one, and then compare them with references, obtaining average metrics that indicate to what extent a single channel in different spaces represents the major information of an image.

ages, thus the model turns out to be $I = R \circ L + N$. A simple algebraic transformation leads to $I = (R + \tilde{N}) \circ L$, where $\tilde{N} = N/L$ with element-wise division. We can see from above that the noise will be also amplified along with light enhancement, which becomes spatially-correlated with Las discussed in [30]. Furthermore, in low-light conditions, sensors (either CCD or CMOS) are sensitive and non-linear to insufficient photons of different light spectrum, which brings **color distortion** even with illumination and noise being 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, simultaneously.

This work studies how to handle the complex degradation in the darkness from a dividing and ruling perspective. Assuming that an image can be disassembled into 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 Figs. 2 and 3, we visualize how the noise distributes in different color-spaces and make statistics on the effect of the single-channel restoration on data pairs from the LOL training dataset [26]. The results suggest that the YCbCr space is a "good" candidate to do the job of texture-color decomposition. As aforementioned, the principle behind our design is that "mind your own business". In other words, a solver is customized for coping with one type of degradation intently. By the luminancechrominance separation, we are thereby possible to treat the degradation in the texture and color components individually, say focusing on the noise issue in the texture and the color distortion in the other part. Please notice that, although several methods [27, 29] have been recently proposed to take care of noise and color shift issues along with light enhancement without distinction on the recovered reflectance component, they barely consider decomposing the degradation from a texture-color point of view to further ease the problem.

This paper develops an effective low-light image enhancement framework via breaking down the darkness (Bread for short) based on the above analysis. The major contributions of this paper are summarized as follows:

- 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.
- We present an effective noise synthesis strategy under the guidance of illumination, significantly improving the suppression quality of amplified and spatiallycorrelated noise in the luminance.
- To tackle the color distortion issue left in lightenhanced images, we design a novel color adaption network, which can properly deal with the chrominance according to given luminance.
- 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

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

Traditional Methods. The simplest and most intuitive way is to linearly adjust the value range or execute a nonlinear Gamma correction on inputs. Further, global and local histogram-based methods [1, 3, 4, 14, 18] are introduced to expand the dynamic 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 [13], Single-scale Retinex (SSR) [12] first uses the Gaussian blurred input as its illumination map, and then removes the estimated illumination from the input as its final result. Multi-scale Retinex (MSR) [11] extends SSR by fusing the results of multiple Gaussian blur functions with different variances. Besides the above mentioned attempts, NPE [23] takes local maxima assumption to predict the illumination, which is manipulated by a mapper to act as the enhanced illumination and merged with the reflectance component. LIME [9] proposes to refine the initial illumination obtained by the Max-RGB assumption and structure preserving constraint. Although these methods can somewhat brighten low-light images, they barely take other hidden degradation, like noise and color distortion, into consideration. To suppress the noise effect, SRIE [7] further imposes the sparsity on the recovered reflectance. Similarly, RRM [15] integrates noise estimation into the main driving optimization to eliminate the noise in reflectance. However, the applicability of these optimization-based methods is limited because of the expensive computation, sensitivity to hyper-parameters, and unsatisfied enhancement quality.

Deep Learning-based Methods. Recently, methods based on deep learning have dominated the target task. For instance, MSR-Net [20] integrates the MSR mechanism into a neural network and uses the BM3D [5] for denoising. It gains improvement in visual quality, but still suffers from the drawbacks like over-enhancement as the traditional MSR and lacks an embedded denoising functionality. LLNet [16] synthesizes training pairs by randomly applying Gamma adjustment and adding synthetic noise to clean images. An auto-encoder network is then employed to learn the mapping function. However, the relationship between real-world illumination and noise is not touched, thus residual noise and over-smoothing problems show up. Cai et al. [2] exploit to construct references from multiexposure sequences for single image enhancement. However, its performance is upper-bounded by involved MEF methods. DUPE [22] and GLAD [25] learn the illumination map for image retouching. Despite reasonable results, they cannot effectively alleviate noise and color distortion issues. DRD [26] and KinD++ [29] resort to the layer decomposition strategy that is beneficial to both illumination adjustment and reflectance refinement. DRBL [27] develops a deep recursive band network for semi-supervised low-light enhancement. DLN [21] introduces lighten-darken tradeoff and feature aggregation blocks to ameliorate results. However, they frequently have troubles in over-exposure and color distortion. In unsupervised settings, Enlighten-GAN [10] attempts to take advantage of larger-scale unpaired training data through employing the GAN mechanism, while Zero-DCE [8] alternatively learns a set of nonreference loss functions. Although relieving the requirement of paired data, they mainly focus on the light factor, and thus have insufficient abilities to remedy other defects. *Despite a progress made toward addressing the problem, effective and efficient designs for handling complex and multientangled degradation are desirable for practical use.*

3. Methodology

3.1. Problem Analysis & Motivation

As previously discussed, low-light image enhancement encounters complex degradation mixed up by dim light, noise, and color distortion. Most previous methods tend to directly restore the reflectance from illumination-adjusted input under the guidance of ground-truth data, which in nature ignores the noise-amplification process by the illumination adjustment operation. In other words, the defects hidden in the darkness will burst in the RGB channels of the reflectance map, which are correlated to the spatially-variant illumination. This fact is tricky for both end-to-end networks like [25] and post-processing methods like [20], resulting in artifacts in texture and/or color. Motivated by the principle of divide-and-rule, we propose to convert images from the RGB colorspace into a luminance-chrominance one for decoupling the light and noise factors from the color degradation. Among various colorspaces, YCbCr seems to be a good choice supported by the evidence given in Fig. 3. To handle the dim light and noise in the luminance, we denote the low-light luminance as $Y_{low} = Y \circ L_{low} + N$, and the normal-light reference as $Y_{high} = Y \circ L_{high}$. The connection between Y_{low} and Y_{high} can be found as follows:

$$Y_{high} = \frac{(Y_{low} - N) \circ L_{high}}{L_{low}} = \frac{Y_{low}}{\hat{L}} - \frac{N}{\hat{L}}, \quad (1)$$

where $\hat{L} = L_{low}/L_{high}$ represents the relative difference between illuminations of low and normal-light images, and the division is element-wise. We propose a solver, *i.e.* illumination adjustment network, to seek $Y_{IA} = Y_{low}/\hat{L}$. From Eq. (1), we can see that the noise term $\tilde{N} = N/\hat{L}$ from Y_{low} remains in Y_{IA} . Even if N is simple, \tilde{N} will become much more complicated due to the correlation with spatially-variant illumination. Thus, it is natural to adopt L as an indicator for the denoising. Barely previous methods take into account this property. KinD [30] is limited to simply concatenate L_{low} and R together to restore the reflectance, which may not sufficiently exploit the guidance information from the illumination, and the single restoration network has to deal with all the degradation simultaneously. Inspired by the above, we alternatively synthesize noisy images that have the same illumination with references but are corrupted with stimulated amplified noise guided by $\mathcal{A} = \Phi(\hat{L})$, where $\Phi(\cdot)$ is a function that reflects the relationship between illuminations and noise levels. By



Figure 4. An overview of our proposed Bread architecture and the basic structures of the involved sub-networks. The estimated relative illumination map are shown by heat map for a better view.

this means, we obtain a noise-suppression solver, the solution of which are constrained by \mathcal{A} . Though the relative noise level map is determined, its overall scale is inaccessible due to the differences in photographing devices. To robustly remove the noise, we further fuse the denoised luminance Y_{NS} under different suppression strengths to obtain the expected luminance map Y_{NF} . Having Y_{NF} estimated, it is reasonable to employ it as guidance for mapping chrominance (color correction) from the input to an adjusted/target light levels, which is achieved by our color adaption network.

3.2. Overall Network Design

Figure 4 shows the overall architecture of our method, which comprises an illumination adjustment net (IAN), an adaptive noise suppression net (ANSN), and a color adaption net (CAN). As can be seen from Fig. 4 A, we firstly convert the input from the RGB to the YCbCr colorspace, obtaining the luminance Y_{low} and chrominance $C_{b_{low}}, C_{r_{low}}$ components. Then, Y_{low} is fed into the IAN to predict the relative illumination difference map L, yielding the adjusted Y_{IA} . The ANSN takes in Y_{IA} and $\Phi(L)$ (denoted as A) to produce Y_{NS} . As discussed in 3.1, we introduce a noise fusion module (NFM) into the noise suppression stage. Relative noise level maps $\{A_1, ..., A_k\}$ with k different suppression strengths are used to generate k denoised luminance maps $\{Y_{NS_1}, ..., Y_{NS_k}\}$, which are then fused as Y_{NF} . The obtained luminance Y_{NF} is consequently utilized as the guidance of the chrominance enhancement. Given the chrominance components of the lowlight image, and Y_{NF} , the CAN outputs $C_{b_{CA}}$ and $C_{r_{CA}}$. Finally, we combine Y_{NF} , $C_{b_{CA}}$ and $C_{r_{CA}}$, and convert them back to the RGB colorspace.

To largely exclude influence from other factors and verify the main claim, all the sub-networks follow the shape shown in Fig. 4 B. There are three down-sampling layers and three up-sampling layers in each sub-network, and two convolutional layers with ReLU activations are inserted before each scaling layer. The inner channels of each block are doubled after each down-sampling block and halved after each up-sampling layer, changing between 32 and 128. All the convolution layers employ 3×3 kernel size, stride = 1 and padding = 1. All the sub-networks are with a Sigmoid activation at their tails, except for the ANSN which is merely the convolution. *Note that our sub-networks are optimized individually*.

3.3. IAN: Illumination Adjustment Network

IAN is proposed to estimate the relative difference of illumination maps between low-light luminance Y_{low} and the normal-light Y_{high} counterpart. The difference map is adopted to adjust the low-light input into the normal-light and further contributes to the noise suppression. The following loss function is used for IAN:

$$\mathcal{L}_{IA} = \left\| Y_{low} / (\hat{L} + \epsilon) - Y_{high} \right\|_{2}^{2} + \alpha \left\| W \circ \nabla \hat{L} \right\|_{1} + \beta \left\| \nabla \hat{L} - \nabla Y_{low} \right\|_{1},$$
⁽²⁾

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 the importance of different terms. $\hat{L} = \theta_{IAN}(Y_{low}) \rightarrow L_{low}/L_{high}$ is the predicted relative difference of illumination between Y_{low} and Y_{high} . The second term is to constrain the illumination to be piece-wise smooth with $W = 1/(\nabla Y_{low} + \epsilon)$ as the weight map. The last term is to preserve the similarity between Y_{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, following $Y_{IA} = Y_{low}/(\hat{L} + \epsilon)$.

3.4. ANSN: Adaptive Noise Suppression Network

According to Eq. (1), we obtain the adjusted luminance $Y_{IA} = Y_{high} + \tilde{N}$, in which the noise in Y_{low} is also amplified. Moreover, possible errors in the estimated illumination should not flow to the subsequent process. Therefore,



Figure 5. Visual comparison of state-of-the-art methods and ours on samples from the LOL dataset.

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 red, the second-best results in blue and the third in green.

Metrics	LIME [9]	SIRE [7]	NPE [23]	RRM [15]	EG [10]	Zero-DCE [8]	MSRNet [20]
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
$\mathrm{PSNR}_C\uparrow$	19.14	20.97	20.59	20.25	22.48	21.88	16.71
$\mathrm{SSIM}_C\uparrow$	0.471	0.656	0.513	0.774	0.710	0.640	0.461
$\operatorname{NIQE}_C \downarrow$	8.954	7.321	8.890	3.944	4.837	7.889	9.074
$\text{DeltaE}_C \downarrow$	19.06	14.32	17.60	14.03	13.28	14.01	22.72
Metrics	DUPE [22]	GLAD [25]	DRD [26]	DRBL [27]	KinD++ [29]	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
$\mathrm{PSNR}_C\uparrow$	22.36	23.72	18.73	22.48	23.91	25.98	26.06
$\mathrm{SSIM}_C\uparrow$	0.594	0.724	0.448	0.851	0.847	0.851	0.857
$\operatorname{NIQE}_C \downarrow$	8.110	6.754	9.644	3.753	3.901	3.649	3.614
$\text{DeltaE}_C \downarrow$	14.77	12.54	21.49	11.75	10.08	9.41	9.06

we simulate amplified noise \tilde{N} on normal-light references Y_{high} without changing their illumination. Regions originally under darker light should have more intense noise than those under brighter conditions, thus relative illumination map \hat{L} estimated previously is natural to be adopted to in-

dicate noise levels. The choices of noise pattern are investigated in LLNet [16], which says that Gaussian and Poisson noises are competent. We take the simple Gaussian model for noise synthesis. Modified from the traditional AWGN model [28], we reach the following for noise synthesis:

$$Y_N = Y_{high} + \mathcal{N}(0, \mathcal{A}),\tag{3}$$

where A can be viewed as an attention map in inverse proportion to \hat{L} . Moreover, the following simple loss function is used for ANSN:

$$\mathcal{L}_{NS} = \left\| \theta_{NS}(Y_N, \mathcal{A}) - \mathcal{N}(0, \mathcal{A}) \right\|_2^2.$$
(4)

With the trained ANSN, the amplified noise in Y_{IA} can be removed through $Y_{NS} = Y_{IA} - \theta_{NS}(Y_{IA}, \mathcal{A})$.

Though ANSN is already able to produce noise-free enhancement results in most cases, the solo usage of it may still leads to over-smoothed luminance for some samples. The reason is that the overall scale of the noise level map is inaccessible due to the differences in photographing devices. To robustly remove the noise, we develop a noise fusion module (NFM) to merge the luminance maps $\{Y_{NS_1}, Y_{NS_2}, ..., Y_{NS_k}\}$ with k different denoising strengths of $\{A_1, A_2, ..., A_k\}$. Moreover, the fusion process is expected to further remedy errors caused by the previous stages. To be concluded, the following learning problem is introduced:

$$\mathcal{L}_{NF} = \left\| Y_{NF} - Y_{high} \right\|_{2}^{2} + \text{SSIM}(Y_{NF}, Y_{high}) \tag{5}$$

where $Y_{NF} = \theta_{NF}(Y_{NS_{1...k}}, \mathcal{A}_{1...k})$, and SSIM (\cdot, \cdot) is the structural similarity loss.

3.5. CAN: Color Adaption Network

Having obtained the Y_{NS} , the color components are expected to be adapted accordingly. Given the luminance and chrominance components of the original input $(Y_{low}, C_{b_{low}}, C_{r_{low}})$ and the reference $(Y_{high}, C_{b_{high}}, C_{r_{high}})$, the loss function for CAN is designed as follows:

$$\mathcal{L}_{CA} = \left\| C_{b_{CA}} - C_{b_{high}} \right\|_{2}^{2} + \left\| C_{r_{CA}} - C_{r_{high}} \right\|_{2}^{2}, \quad (6)$$

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

Considering the low-saturation problem of the references in the LOL [26] training data as shown in Fig. 6, we can alternatively introduce multi-exposure data for the training of CAN, which often covers a wide range of exposure and aplenty color patterns. The following learning problem is desired:

$$(C_{b_{ME}}, C_{r_{ME}}) = \theta_{ME}(Y_{e1}, C_{b_{e1}}, C_{r_{e1}}, Y_{e2}).$$
(7)

Image pairs under different exposures, denoted as $(Y_{e1}, C_{b_{e1}}, C_{r_{e1}})$ and $(Y_{e2}, C_{b_{e2}}, C_{r_{e2}})$, are randomly selected



Figure 6. The ground truth of training data in the LOL dataset typically has less vivid color than that of evaluation.

from the sequences of multi-exposure photographs. The objective function of this module shares the same form with Eqn. (6). It forces the enhanced chrominance components to fit the exposure condition with respect to a certain luminance guided by real multi-exposure data. 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 show results of both Bread and Bread-ME in quantitative experiments for fair comparisons.

4. Experiments

4.1. 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 as 10^{-3} , except for the finetuning of the color adaption network based on multi-exposure Data, which is set as 10^{-4} . α and β in Eqn. (2) are set as 4.0 and 0.5, respectively. We set $\mathcal{A} = \exp(-\hat{L})$, k = 3, $\mathcal{A}_{1,2,3} =$ $\{0, 0.05\mathcal{A}, 0.1\mathcal{A}\}$ for the ANSN and NFM.

We adopt the LOL dataset as the training data, which contains 485 low/normal light pairs for training and 15 for evaluation. To imitate various exposures in real-world photographs, we synthesize 8 images with different exposures for each low-light image. The magnifications of the exposures are uniformly distributed from one to making 25% pixels over-exposed at most. For the training of the color adaption network based on multi-exposure data, we first select the SICE dataset [2] for training, then append the original 485 pairs of LOL dataset into the training data for fine-tuning.



Figure 8. Visual comparison between different methods on the VV (top two rows) and MEF-DS [6] (the last row). Please zoom in for details.

4.2. Performance Evaluation

To verify the effectiveness of our proposed method, several public datasets are used for evaluation, including LOL [26], DICM [14], NPE [23] and VV¹. Representative stateof-the-art methods, including LIME [9], SIRE [7], NPE [23], RRM [15], MSRNet [20], DUPE [22], GLAD [25], DRD [26], DRBL [27], KinD [30], EG [10] and Zero-DCE [8] are adopted for comparisons. Image quality assessment metrics, including PSNR, SSIM, NIQE [17], DeltaE [19] and LOE [24], are employed to measure these methods.

The quantitative comparison between Bread(-ME) and other competitors on the LOL evaluation dataset is reported in Table 1. Our method outperforms other state-of-the-art alternatives by a noticeable margin, on both of the refer-

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

ence and no-reference metrics, demonstrating the efficacy of our proposed Bread framework. Note that the absolute brightness is inaccessible during evaluation, which may be unfair to those non-data-driven methods and may interfere the assessment of the fidelity in details, thus we also provide a version of metrics besides the original, the predicted luminance of which is aligned to its ground truth by simple Gamma correction. Visual comparisons on several samples from the LOL evaluation and testing datasets are depicted in Figs. 5 and 7, respectively. The results by our method achieve remarkably higher quality with noise wellsuppressed and less artifacts, while irregular illumination, noise residual, and texture/color distortion exist in the results of other methods. Moreover, we can see from Fig. 5 that multi-exposure-data-based color adapter rectifies the greenish hue, producing more realistic tone. Also, we pro-

Table 2. Quantitative comparison on the DICM, NPE and VV datasets. All the competitors are trained on the LOL dataset.

Datasets	Metrics	GLAD	DRD	DRBL	KinD++	Bread	Bread-ME
DICM (44)	NIQE↓	3.0875	4.7120	3.2784	2.8584	3.0893	3.0869
	LOE↓	240.71	608.70	660.10	535.34	400.07	388.85
NIDE (0)	NIQE↓	3.6143	4.0676	3.5843	3.9101	3.5103	3.4596
NFE (0)	LOE↓	211.27	741.78	814.82	494.38	398.57	392.72
VII (24)	NIQE↓	4.4112	4.1508	3.3770	3.8084	3.6770	3.6710
V V (24)	LOE↓	199.57	522.04	648.67	300.57	339.56	302.26
-	Size(MB)↓	10.90	9.05	20.10	34.9	8.21	8.21



Figure 9. Visual comparison between color adaption without and with multi-exposure data used. It is obvious that the introduction of such data contributes to more vibrant color.

vide non-reference results in Table 2. For fair comparison, the methods list in the table are all data-driven and trained on the LOL dataset. Our framework occupies the leading position in terms of both two non-reference metrics on three benchmark datasets. Please note that, NIQE does not monotonously change following the quantity of visual noises, which means that a more visual-pleasing noise-free texture can also cause a high NIQE value. We will continue this discussion in the supplementary materials. Additionally, we provide more visual results to further illustrate the effectiveness of the noise suppression and the fidelity of the color adaption in Fig. 8.

4.3. Ablation Study

As shown in Table 3, to evaluate the effectiveness of different settings in our framework, we conduct ablation studies, including removing the denoising process, the noise fusion module and the IA-NS separation setting, respectively. We further evaluate different noise synthesis models, including fixed Gaussian noise (noise level of 25, denoted as FGN), Poisson noise (PN) and our spatially-correlated Gaussian noise (Bread). To validate the necessity of the usage of multi-exposure data, please compare the results of Bread and Bread-ME. The visual comparison is also shown in Fig. 9, which illustrates that the color fidelity meets significant improvement from the data. Several key questions are highlighted as follows:

The necessity of NFM: We can see that without the NFM, the framework encounters an obvious degradation for the non-blind metrics, because the errors arising in IAN and

Table 3. Ablation study on different configurations.

	PSNR↑	SSIM↑	$\mathrm{PSNR}_C\uparrow$	$\mathrm{SSIM}_C\uparrow$
w/o DN	16.91	0.586	23.33	0.623
w/o NFM	17.08	0.721	23.72	0.816
w/o IA-NS Sep.	18.69	0.785	24.58	0.829
w/ FGN	21.40	0.799	25.26	0.821
w/ PN	22.44	0.831	25.84	0.849
Bread	22.92	0.836	25.98	0.851
Bread-ME	22.96	0.838	26.06	0.857

ANSN are left untreated.

The necessity of IA-NS separation: For directly learning the restoration without IA-NS separation, we use Y_{IA} and Y_{high} pairs and keep the same architecture of the noise suppression network. The gap between this setting and the noise synthesis approaches emphasizes the importance of the separation strategy.

The choice of noise synthesis settings: We synthesize noisy images polluted by Poisson noise through applying the estimated low illumination map to the normal-light images and then contaminate them through the Poisson process. We then remove its illumination from the corrupted lowlight images to simulate the adjusted images with amplified noise. For the fixed Gaussian results, we simply apply the AWGN on normal-light images. Unfortunately, the strength of Poisson and fixed-Gaussian denoiser is not adjustable, which means further fusion is futile, thus we adopt the spatial-correlated Gaussian as our first option. More details will be discussed in supplementary.

5. 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 of an image is first converted from the RGB into a luminancechrominance 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 Y, while the chrominance components C_b and C_r respond to color correction, having the enhanced Y as guidance. Regarding different specific illnesses, the sub-networks including IAN, ANSN, and CAN are customized and trained individually, all of which follow the simple U-shaped net. Our designs make the training environment for each sub-network specific to one simple degradation, the effect of which has been validated by the experiments. 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.

References

- [1] Mohammad Abdullah-Al-Wadud, Md Hasanul Kabir, M Ali Akber Dewan, and Oksam Chae. A dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(2):593– 600, 2007. 2
- [2] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE TIP*, 27(4):2049–2062, 2018. 3, 6
- [3] Turgay Celik and Tardi Tjahjadi. Contextual and variational contrast enhancement. *IEEE TIP*, 20(12):3431–3441, 2011.
 2
- [4] Heng-Da Cheng and XJ Shi. A simple and effective histogram equalization approach to image enhancement. *Digital signal processing*, 14(2):158–170, 2004. 2
- [5] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen O. Egiazarian. Image denoising by sparse 3d transform-domain collaborative filtering. *IEEE TIP*, 16:2080–2095, 2007. 3
- [6] Yuming Fang, Hanwei Zhu, Kede Ma, Zhou Wang, and Shutao Li. Perceptual evaluation for multi-exposure image fusion of dynamic scenes. *IEEE TIP*, 29:1127–1138, 2020.
 7
- [7] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *CVPR*, pages 2782–2790, 2016. 3, 5, 7
- [8] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *CVPR*, pages 1780–1789, 2020. 3, 5, 7
- [9] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE TIP*, 26(2):982–993, 2016. 3, 5, 7
- [10] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE TIP*, 30:2340–2349, 2021. 3, 5, 7
- [11] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE TIP*, 6(7):965– 976, 1997. 3
- [12] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell.
 Properties and performance of a center/surround retinex. *IEEE TIP*, 6(3):451–462, 1997. 3
- [13] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977. 1, 3
- [14] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE TIP*, 22(12):5372–5384, 2013. 2, 7
- [15] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE TIP*, 27(6):2828– 2841, 2018. 3, 5, 7

- [16] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 3, 5
- [17] Anish Mittal, Rajiv Soundararajan, and Alan Conrad Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20:209–212, 2013. 7
- [18] Etta D Pisano, Shuquan Zong, Bradley M Hemminger, Marla DeLuca, R Eugene Johnston, Keith Muller, M Patricia Braeuning, and Stephen M Pizer. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital imaging*, 11(4):193, 1998. 2
- [19] Gaurav Sharma, Wencheng Wu, and Edul N. Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research and Application*, 30:21–30, 2005. 7
- [20] Liang Shen, Zihan Yue, Fan Feng, Quan Chen, Shihao Liu, and Jie Ma. Msr-net: Low-light image enhancement using deep convolutional network. arXiv preprint arXiv:1711.02488, 2017. 3, 5, 7
- [21] Li-Wen Wang, Zhi-Song Liu, Wan-Chi Siu, and Daniel PK Lun. Lightening network for low-light image enhancement. *IEEE TIP*, 29:7984–7996, 2020. 1, 3
- [22] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *CVPR*, pages 6849–6857, 2019. 1, 3, 5, 7
- [23] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE TIP*, 22(9):3538–3548, 2013. 3, 5, 7
- [24] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE TIP*, 22:3538–3548, 2013. 7
- [25] Wenjing Wang, Chen Wei, Wenhan Yang, and Jiaying Liu. Gladnet: Low-light enhancement network with global awareness. In *IEEE International Conference on Automatic Face & Gesture Recognition*, pages 751–755, 2018. 3, 5, 6, 7
- [26] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *BMVC*, page 155, 2018. 2, 3, 5, 6, 7
- [27] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semisupervised approach for low-light image enhancement. In *CVPR*, pages 3063–3072, 2020. 1, 2, 3, 5, 6, 7
- [28] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE TIP*, 26(7):3142–3155, 2017. 5
- [29] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *IJCV*, 129(4):1013–1037, 2021. 1, 2, 3, 5, 7
- [30] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In ACM MM, pages 1632–1640, 2019. 2, 3, 5, 7