

Kindling the Darkness: A Practical Low-light Image Enhancer

Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo

Abstract—Images captured under low-light conditions often suffer from (partially) poor visibility. Besides unsatisfactory lightings, multiple types of degradations, such as noise and color distortion due to the limited quality of cameras, hide in the dark. In other words, solely turning up the brightness of dark regions will inevitably amplify hidden artifacts. This work builds a simple yet effective network for **Kindling the Darkness** (denoted as KinD), which, inspired by Retinex theory, decomposes images into two components. One component (illumination) is responsible for light adjustment, while the other (reflectance) for degradation removal. In such a way, the original space is decoupled into two smaller subspaces, expecting to be better regularized/learned. It is worth to note that our network is trained with paired images shot under different exposure conditions, instead of using any ground-truth reflectance and illumination information. Extensive experiments are conducted to demonstrate the efficacy of our design and its superiority over state-of-the-art alternatives. Our KinD is robust against severe visual defects, and user-friendly to arbitrarily adjust light levels. In addition, our model spends less than 50ms to process an image in VGA resolution on a 2080Ti GPU. All the above merits make our KinD attractive for practical use.

Index Terms—Low light enhancement, image decomposition, image restoration

1 INTRODUCTION

VERY often, capturing high-quality images in dim light conditions is challenging. Though a few operations, such as setting high ISO, long exposure, and flash, can be applied under the circumstances, they suffer from different drawbacks. For instance, high ISO increases the sensitivity of an image sensor to light, but the noise is also amplified, thus leading to the low (signal-to-noise ratio) SNR. Long exposure is limited to shoot static scenes, otherwise it highly likely gets in trouble of blurry results. Using flash can somehow brighten the environment, which however frequently introduces unexpected highlights and unbalanced lighting into photos, making them visually unpleasant. In practice, typical users may even not have the above options with limited photographing tools, *e.g.* cameras embedded in portable devices. Although the low-light image enhancement has been a long-standing problem in the community with a great progress made over the past years, *developing a practical low-light image enhancer remains challenging, since flexibly lightening the darkness, effectively removing the degradations, and being efficient should all be concerned.*

Figure 1 provides three natural images captured under challenging light conditions. Concretely, the first case is with extremely low light. Severe noise and color distortion are hidden in the dark. By simply amplifying the intensity of the image, the degradations show up as given on the top-right corner. The second image is photographed at sunset (weak ambient light), most objects in which suffer from backlighting. Imaging at noon facing to the light source (the sun) also hardly gets rid of the issue like the second case



Fig. 1: Left column: three natural images captured under different light conditions. Right column: our enhanced results. Notice that the first image is with extremely low light, we show its x20 version on the top-right corner.

exhibits, although the ambient light is stronger and the scene is more visible. Note that those relatively bright regions of the last two photos will be saturated by direct amplification.

Deep learning-based methods have revealed their superior performance in numerical low-level vision tasks, such as denoising and super-resolution, most of which need the

• Y. Zhang (zhangyonghua@tju.edu.cn), J. Zhang (jwzhang@tju.edu.cn), and X. Guo (xj.max.guo@gmail.com) are with the College of Intelligence and Computing, Tianjin University, Tianjin 300350, China.
Corresponding author: X. Guo (xj.max.guo@gmail.com)

training data with ground truth. For the target problem, say low-light image enhancement, *no ground-truth real data exists*, although the order of light intensity can be determined. Because, from the viewpoint of users, the favorite light levels for different people/requirements could be much diverse. In other words, one cannot say what light condition is the best/ground-truth. Therefore, it is not so felicitous to map an image only to a version with a specific level of light.

Based on the above analysis, we summarize challenges in low-light image enhancement as follows:

- How to effectively estimate the illumination component from a single image, and flexibly adjust light levels?
- How to remove the degradations like noise and color distortion previously hidden in the darkness after lightening up dark regions?
- How to train a model without well-defined ground-truth light conditions for low-light image enhancement by only looking at two/several different examples?

In this paper, we propose a deep neural network to take the above concerns into account simultaneously.

1.1 Previous Arts

A large number of low-light image enhancement schemes have been proposed. In what follows, we briefly review classic and contemporary works closely related to ours.

Plain Methods. Intuitively, for an image with the globally low light, the visibility can be enhanced by directly amplifying it. But, as shown in the first case of Figure 1, the visual defects including noise and color distortion show up along the details. For images containing bright regions, *e.g.* the last two pictures in Figure 1, this operation easily results in (partial) saturation/over-exposure. One technical line, with histogram equalization (HE) [1], [2], [3] and its follow-ups [4], [5] as representatives, tries to map the value range into $[0, 1]$ and balance the histogram of outputs for avoiding the truncation problem. These methods *de facto* aim to increase the contrast of image. Another mapping manner is gamma correction (GC), which is carried out on each pixel individually in a non-linear way. Although GC can promote the brightness especially of dark pixels, it does not consider the relationship of a certain pixel with its neighbors. *The main drawback of the plain approaches is that they barely consider real illumination factors, usually making enhanced results visually vulnerable and inconsistent with real scenes.*

Traditional Illumination-based Methods. Different from the plain methods, strategies in this category are aware of the concept of illumination. The key assumption, inspired by Retinex theory [6], is that the (color) image can be decomposed into two components, *i.e.* reflectance and illumination. Early attempts include single-scale Retinex (SSR) [7] and multi-scale Retinex (MSR) [8]. Limited to the manner of producing the final result, the output often looks unnatural and somewhere over-enhanced. Wang *et al.* proposed a method called NPE [9], which jointly enhances contrast and preserves naturalness of illumination. Fu *et al.* developed a method [10], which adjusts the illumination through fusing multiple derivations of the initially estimated illumination map. However, this method sometimes sacrifices the realism of those regions containing rich textures. Guo *et al.* focused

on estimating the structured illumination map from an initial one [11]. *These methods generally assume that the images are noise- and color distortion-free, and do not explicitly consider the degradations.* In [12], a weighted variational model for simultaneous reflectance and illumination estimation (SRIE) was designed to obtain better reflectance and illumination layers, then the target image is generated by manipulating the illumination. Following [11], Li *et al.* further introduced an extra term to host noise [13]. *Although both [12] and [13] can reject slight noise in images, they are short of abilities in handling color distortion and heavy noise.*

Deep Learning-based Methods. With the emergence of deep learning, a number of low-level vision tasks have been benefited from deep models, such as [14], [15] for denoising, [16] for super-resolution, [17] for compression artifact removal and [18] for dehazing. Regarding the target mission of this paper, the low-light net (LLNet) proposed in [19] builds a deep network that performs as a simultaneous contrast enhancement and denoising module. Shen *et al.* deemed that multi-scale Retinex is equivalent to a feed-forward convolutional neural network with different Gaussian convolution kernels. Motivated by this, they constructed a convolutional neural network (MSR-net) [20] to learn an end-to-end mapping between dark and bright images. Wei *et al.* designed a deep network, called Retinex-Net [21], that integrates image decomposition and illumination mapping. Please notice that Retinex-Net additionally employs an off-the-shelf denoising tool (BM3D [22]) to clean the reflectance component. *These strategies all assume that there exist images with “ground-truth” lights, without considering that the noise differently affects regions with various lights.* Simply speaking, after extracting the illumination factor, the noise level of dark regions is (much) higher than that of bright ones in the reflectance. In such a situation, adopting/training a denoiser with a uniform ability over an image (reflectance) is no longer suitable. In addition, the above methods do not explicitly cope with the degradation of color distortion, which is not uncommon in real images. More recently, Chen *et al.* proposed a pipeline for processing low-light images based on end-to-end training of a fully convolutional network [23], which can jointly deal with noise and color distortion. However, this work is specific to data in RAW format, limiting its applicable scenarios. As stated in [23], if modifying the network to accept data in JPEG format, the performance significantly drops.

Most existing methods manipulate the illumination by gamma correction, appointing a level existing in carefully constructed training data, or fusion. For gamma correction, it may be unable to reflect the relationship between different light (exposure) levels. As for the second manner, it is heavily restricted to whether the appointed level is contained in the training data. While for the last one, it even does not provide a manipulation option. Therefore, *it is desired to learn a mapping function to arbitrarily convert one light (exposure) level to another for offering users the flexibility of adjustment.*

Image Denoising Methods. In the fields of image processing, multimedia, and computer vision, image denoising has been a hot topic for a long time, with numerous techniques proposed over past decades. Classic ones model/regularize the problem by utilizing some specific priors of natural clean images, like non-local self-similarity,

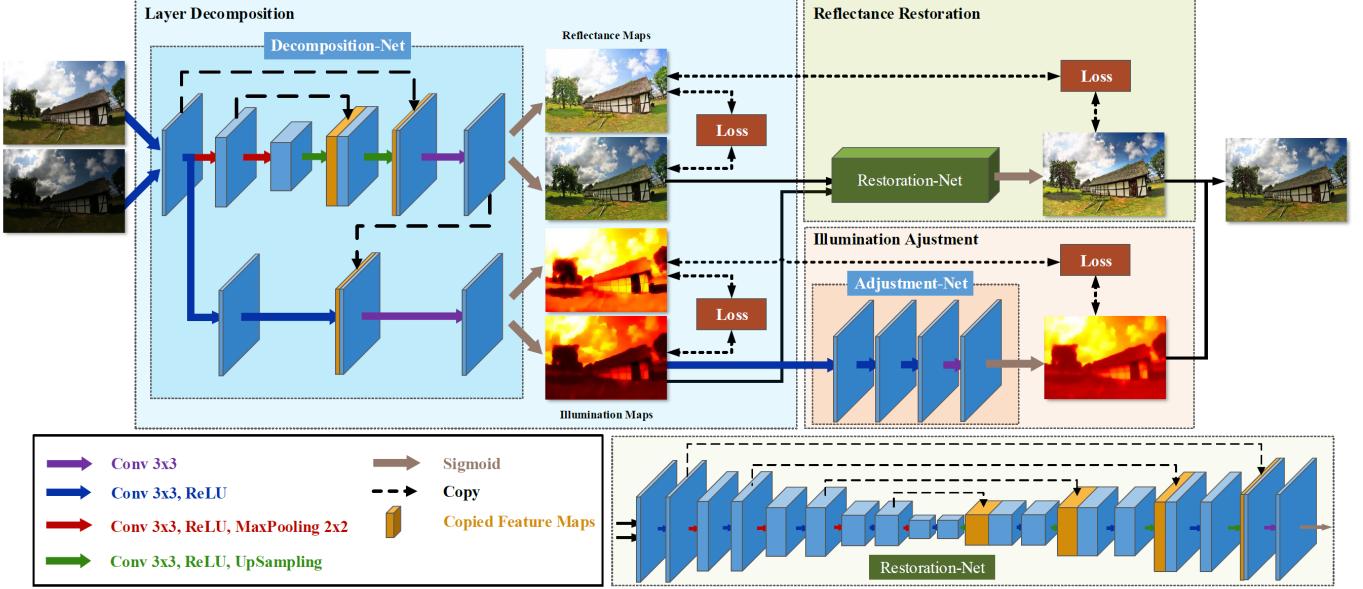


Fig. 2: The architecture of our KinD network. Two branches correspond to the reflectance and illumination, respectively. From the perspective of functionality, it also can be divided into three modules, including layer decomposition, reflectance restoration and illumination adjustment.

piecewise smoothness, signal (representation) sparsity, etc. The most popular schemes arguably go to BM3D [22] and WNNM [24]. Due to the high complexity of optimization procedure in the testing, and the large searching space of proper parameters, these traditional methods often show the unsatisfactory performance in real situations. Lately, deep learning based denoisers exhibit the superiority on the task. The representative works, such as SSDA using stacked sparse denoising auto-encoders [25], [26], TNRD by trainable non-linear reaction diffusion [27], DnCNN with residual learning and batch normalization [15], can save computational expense thanks to only feed-forward convolution operations involved in the testing phase. However, these deep models still have the difficulty for blind image denoising. One may train multiple models for varied levels or one model with a large number of parameters, which is obviously inflexible in practice. By taking the recurrent thought into the task, this issue is mitigated [28]. But, none of the mentioned approaches considers that different regions of a light-enhanced image host different levels of noise. Same problem happens to color distortion.

1.2 Our Contributions

This study presents a deep network for practically solving the low-light enhancement problem. The main contributions of this work can be summarized in the following aspects.

- Inspired by Retinex theory, the proposed network decomposes images into two components, i.e. reflectance and illumination, which decouples the original space into two smaller ones.
- The network is trained with paired images captured under different light/exposure conditions, instead of using any ground-truth reflectance and illumination information.

- Our designed model provides a mapping function for flexibly adjusting light levels according to different demands from users.
- The proposed network also contains a module, which is capable to effectively remove visual defects amplified through lightening dark regions.
- Extensive experiments are conducted to demonstrate the efficacy of our design and its superiority over state-of-the-art alternatives.

2 METHODOLOGY

A desired low-light image enhancer should be capable to effectively remove the degradations hidden in the darkness, and flexibly adjust light/exposure conditions. We build a deep network, denoted as KinD, to achieve the goal. As schematically illustrated in Figure 2, the network is composed of two branches for handling the reflectance and illumination components, respectively. From the perspective of functionality, it also can be divided into three modules, including layer decomposition, reflectance restoration, and illumination adjustment. In the next subsections, we shall explain the details about the network.

2.1 Consideration & Motivation

2.1.1 Layer Decomposition

As discussed in Sec. 1.1, the main drawback of plain methods comes from the blindness of illumination. Thus, it is key to obtain the illumination information. If having the illumination well-extracted from the input, the rest hosts the details and possible degradations, where the restoration (or degradation removal) can be executed on. In Retinex theory, an image \mathbf{I} can be viewed as a composition of two components, i.e. reflectance \mathbf{R} and illumination \mathbf{L} , in the fashion

of $\mathbf{I} = \mathbf{R} \circ \mathbf{L}$, where \circ designates the element-wise product. Further, decomposing images in the Retinex manner consequently decouples the space of mapping a degraded low-light image to a desired one into two smaller subspaces, expecting to be better and easier regularized/learned. Moreover, the illumination map is core to flexibly adjusting light/exposure conditions. Based on the above, *the Retinex-based layer decomposition is suitable and necessary for the target task.*

2.1.2 Data Usage & Priors

There is no well-defined ground-truth for light conditions. Furthermore, no/few ground-truth reflectance and illumination maps for real images are available. The layer decomposition problem is in nature under-determined, thus additional priors/regularizers matter. Suppose that the images are degradation-free, *different shots of a certain scene should share the same reflectance*. While the illumination maps, though could be intensively varied, *are of simple and mutually consistent structure*. In real situations, the degradations embodied in low-light images are often worse than those in brighter ones, which will be diverted into the reflectance component. This inspires us that the reflectance from the image in bright light can perform as the reference (ground-truth) for that from the degraded low-light one to learn restorers. One may ask that *why not use synthetic data?* Because it is hard to synthesize. The degradations are not in a simple form, and change with respect to different sensors. *Please notice that the usage of reflectance (well-defined) totally differs from using images in (relatively) bright light as the reference of low light ones.*

2.1.3 Illumination Guided Reflectance Restoration

In the decomposed reflectance, the pollution of regions corresponding to darker illumination is heavier than that to brighter one. Mathematically, a degraded low-light image can be naturally modeled as $\mathbf{I} = \mathbf{R} \circ \mathbf{L} + \mathbf{E}$, where \mathbf{E} designates the pollution component. By taking simple algebra steps, we have:

$$\mathbf{I} = \mathbf{R} \circ \mathbf{L} + \mathbf{E} = \tilde{\mathbf{R}} \circ \mathbf{L} = (\mathbf{R} + \tilde{\mathbf{E}}) \circ \mathbf{L} = \mathbf{R} \circ \mathbf{L} + \tilde{\mathbf{E}} \circ \mathbf{L}, \quad (1)$$

where $\tilde{\mathbf{R}}$ stands for the polluted reflectance, and $\tilde{\mathbf{E}}$ is the degradation having the illumination decoupled. The relationship $\mathbf{E} = \tilde{\mathbf{E}} \circ \mathbf{L}$ holds. Taking the additive white Gaussian noise $\mathbf{E} \sim \mathcal{N}(0, \sigma^2)$ for an example, the distribution of $\tilde{\mathbf{E}}$ becomes much more complex and strongly relates to \mathbf{L} , i.e. $\frac{\sigma^2}{\mathbf{L}_i}$ for each position i . This is to say, the reflectance restoration cannot be uniformly processed over an entire image, and the illumination map can be a good guider. One may wonder *what if directly removing \mathbf{E} from the input \mathbf{I} ?* For one thing, the unbalance issue still remains. By viewing from another point, the intrinsic details will be unequally confounded with the noise. For another thing, different from the reflectance, we no longer have proper references for degradation removal in this manner, since \mathbf{L} varies. Analogous analysis serves other types of degradation, like color-distortion.

2.1.4 Arbitrary Illumination Manipulation

The favorite illumination strengths of different persons/applications may be pretty diverse. Therefore, a



Fig. 3: Left column: Lower light input and its decomposed illumination and (degraded) reflectance maps. Right column: Brighter input and its corresponding maps. Three rows respectively correspond to inputs, illumination maps, and reflectance maps. These are testing images.

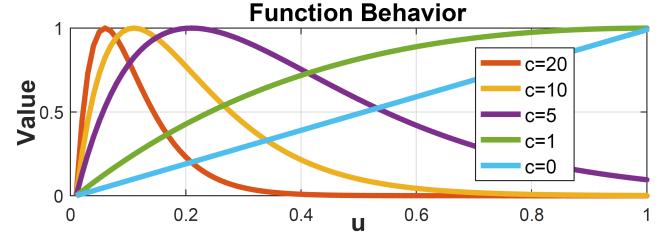


Fig. 4: The behavior of function $v = u \cdot \exp(-c \cdot u)$. The parameter c controls the shape of function.

practical system needs to provide an interface for arbitrary illumination manipulation. In the literature, three main ways for enhancing light conditions are fusion, light level appointment, and gamma correction. The fusion-based methods, due to the fixed fusion mode, lack in the functionality of light adjustment. If adopting the second option, the training dataset has to contain images with target levels, limiting its flexibility. For gamma correction, although it can achieve the goal by setting different γ values, it may be unable to reflect the relationship between different light (exposure) levels. *This paper advocates to learn a flexible mapping function from real data, which accepts users to appoint arbitrary levels of light/exposure.*

2.2 KinD Network

Inspired by the consideration and motivation, we build a deep neural network, denoted as KinD, for kindling the darkness. Below, we describe the three subnets in details from the functional perspective.

Inputs	Operator	Kernel	Output Channels	Stride	Output Name
RGB	Conv&ReLU	3×3	32	1	Decom_conv1
Decom_conv1	Max Pooling	2×2	32	2	Decom_pool1
Decom_pool1	Conv&ReLU	3×3	64	1	Decom_conv2
Decom_conv2	Max Pooling	2×2	64	2	Decom_pool2
Decom_pool2	Conv&ReLU	3×3	128	1	Decom_conv3
Decom_conv3	Deconv	2×2	64	2	Decom_up1
Decom_up1, Decom_conv2	Concat	-	128	-	Decom_concat1
Decom_concat1	Conv&ReLU	3×3	64	1	Decom_conv4
Decom_conv4	Deconv	2×2	32	2	Decom_up2
Decom_up2, Decom_conv1	Concat	-	64	-	Decom_concat2
Decom_concat2	Conv&ReLU	3×3	32	1	Decom_conv5
Decom_conv5	Conv	3×3	3	1	Decom_conv6
Decom_conv6	Sigmoid	-	3	-	Decom_Reflectance
Decom_i_conv1	Conv&ReLU	3×3	32	1	Decom_i_conv1
Decom_i_conv1, Decom_conv5	Concat	-	64	-	Decom_i_conv2
Decom_i_conv2	Conv	3×3	1	1	Decom_i_conv3
Decom_i_conv3	Sigmoid	-	1	-	Decom_Illumination

TABLE 1: Layer decomposition network

Inputs	Operator	Kernel	Output Channels	Stride	Output Name
Decom_i_conv3, Decom_conv5	Concat	-	33	-	RE_concat1
RE_concat1	Conv&ReLU	3×3	32	1	RE_conv1_1
RE_conv1_1	Conv&ReLU	3×3	32	1	RE_conv1_2
RE_conv1_2	Max Pooling	2×2	32	2	RE_pool1
RE_pool1	Conv&ReLU	3×3	64	1	RE_conv2_1
RE_conv2_1	Conv&ReLU	3×3	64	1	RE_conv2_2
RE_conv2_2	Max Pooling	2×2	64	2	RE_pool2
RE_pool2	Conv&ReLU	3×3	128	1	RE_conv3_1
RE_conv3_1	Conv&ReLU	3×3	128	1	RE_conv3_2
RE_conv3_2	Max Pooling	2×2	128	2	RE_pool3
RE_pool3	Conv&ReLU	3×3	256	1	RE_conv4_1
RE_conv4_1	Conv&ReLU	3×3	256	1	RE_conv4_2
RE_conv4_2	Max Pooling	2×2	256	2	RE_pool4
RE_pool4	Conv&ReLU	3×3	512	1	RE_conv5_1
RE_conv5_1	Conv&ReLU	3×3	512	1	RE_conv5_2
RE_conv5_2	Deconv	2×2	256	2	RE_up1
RE_up1, RE_conv4_2	Concat	-	512	-	RE_concat2
RE_concat2	Conv&ReLU	3×3	256	1	RE_conv6_1
RE_conv6_1	Conv&ReLU	3×3	256	1	RE_conv6_2
RE_conv6_2	Deconv	2×2	128	2	RE_up2
RE_up2, RE_conv3_2	Concat	-	128	-	RE_concat3
RE_concat3	Conv&ReLU	3×3	128	1	RE_conv7_1
RE_conv7_1	Conv&ReLU	3×3	128	1	RE_conv7_2
RE_conv7_2	Deconv	2×2	64	2	RE_up3
RE_up3, RE_conv2_2	Concat	-	128	-	RE_concat4
RE_concat4	Conv&ReLU	3×3	64	1	RE_conv8_1
RE_conv8_1	Conv&ReLU	3×3	64	1	RE_conv8_2
RE_conv8_2	Deconv	2×2	32	2	RE_up4
RE_up4, RE_conv1_2	Concat	-	64	-	RE_concat5
RE_concat5	Conv&ReLU	3×3	32	1	RE_conv9_1
RE_conv9_1	Conv&ReLU	3×3	256	1	RE_conv9_2
RE_conv9_2	Conv	3×3	3	1	RE_conv10
RE_conv10	Sigmoid	-	3	-	RE_reflectance

TABLE 2: Reflectance restoration network

2.2.1 Layer Decomposition Net

Recovering two components from one image is a highly ill-posed problem. Having no ground-truth information guided, a loss with well-designed constraints is important. Fortunately, we have paired images with different light/exposure configurations [$\mathbf{I}_l, \mathbf{I}_h$]. Recall that the reflectance of a certain scene should be shared across different images, we regularize the decomposed reflectance pair [$\mathbf{R}_l, \mathbf{R}_h$] to be close (ideally the same if degradation-free). Furthermore, the illumination maps [$\mathbf{L}_l, \mathbf{L}_h$] should be piece-wise smooth and mutually consistent. The following terms are adopted. We simply use $\mathcal{L}_{rs}^{LD} := \|\mathbf{R}_l - \mathbf{R}_h\|_2^2$ to regularize the *reflectance similarity*, where $\|\cdot\|_2$ means the ℓ^2 norm (MSE). The *illumination smoothness* is constrained by $\mathcal{L}_{is}^{LD} := \|\frac{\nabla \mathbf{L}_l}{\max(|\nabla \mathbf{L}_l|, \epsilon)}\|_1 + \|\frac{\nabla \mathbf{L}_h}{\max(|\nabla \mathbf{L}_h|, \epsilon)}\|_1$, where ∇ stands for the first order derivative operator containing ∇_x (horizontal) and ∇_y (vertical) directions, and $\|\cdot\|_1$ means the ℓ^1 norm. In addition, ϵ is a small positive constant (0.01 in this work) for avoiding zero denominator, and $|\cdot|$ means the absolute value operator. This smoothness term measures the relative structure of the illumination with respect to the input. For a location on an edge in \mathbf{I} , the penalty on \mathbf{L} is small; while for a location in a flat region in \mathbf{I} , the penalty turns to be large. As for the *mutual consistency*, we employ $\mathcal{L}_{mc}^{LD} := \|\mathbf{M} \circ \exp(-c \cdot \mathbf{M})\|_1$ with $\mathbf{M} := |\nabla \mathbf{L}_l| + |\nabla \mathbf{L}_h|$. Figure 4 depicts the function behavior of $u \cdot \circ \exp(-c \cdot u)$, where c is the parameter controlling the shape of function. As can be seen from Figure 4, the penalty first goes up but then drops towards 0 as u increases. This characteristic



Fig. 5: The polluted reflectance maps (top), and their results by BM3D (middle) and our reflectance restoration net (bottom). The right column corresponds to a heavier degradation (a lower light) level than the left. These are testing images.

well fits the mutual consistency, *i.e.* strong mutual edges should be preserved while weak ones depressed. We notice that setting $c = 0$ leads to a simple ℓ^1 loss on \mathbf{M} . Besides, the decomposed two layers should reproduce the input, which is constrained by the *reconstruction error*, say $\mathcal{L}_{rec}^{LD} := \|\mathbf{I}_l - \mathbf{R}_l \circ \mathbf{L}_l\|_1 + \|\mathbf{I}_h - \mathbf{R}_h \circ \mathbf{L}_h\|_1$. As a result, the loss function of layer decomposition net is as follows:

$$\mathcal{L}^{LD} := \mathcal{L}_{rec}^{LD} + 0.01\mathcal{L}_{rs}^{LD} + 0.08\mathcal{L}_{is}^{LD} + 0.1\mathcal{L}_{mc}^{LD}. \quad (2)$$

The layer decomposition network contains two branches corresponding to the reflectance and illumination, respectively. The reflectance branch adopts a typical 5-layer U-Net [29], followed by a convolutional (conv) layer and a Sigmoid layer. While the illumination branch is composed of two conv+ReLU layers and a conv layer on concatenated feature maps from the reflectance branch (for possibly excluding textures from the illumination), finally followed by a Sigmoid layer. The detailed layer decomposition network configuration is provided in Table 1.

2.2.2 Reflectance Restoration Net

The reflectance maps from low-light images, as shown in Figures 3 and 5, are more interfered by degradations than those from bright-light ones. Employing the clearer reflectance to act as the reference (informal ground-truth) for the messy one is our principle. For seeking a restoration function, the objective turns to be simple as follows:

$$\mathcal{L}^{RR} := \|\hat{\mathbf{R}} - \mathbf{R}_h\|_2^2 - \text{SSIM}(\hat{\mathbf{R}}, \mathbf{R}_h) + \|\nabla \hat{\mathbf{R}} - \nabla \mathbf{R}_h\|_2^2, \quad (3)$$

where $\text{SSIM}(\cdot, \cdot)$ is the structural similarity measurement, and $\hat{\mathbf{R}}$ corresponds to the restored reflectance. The third

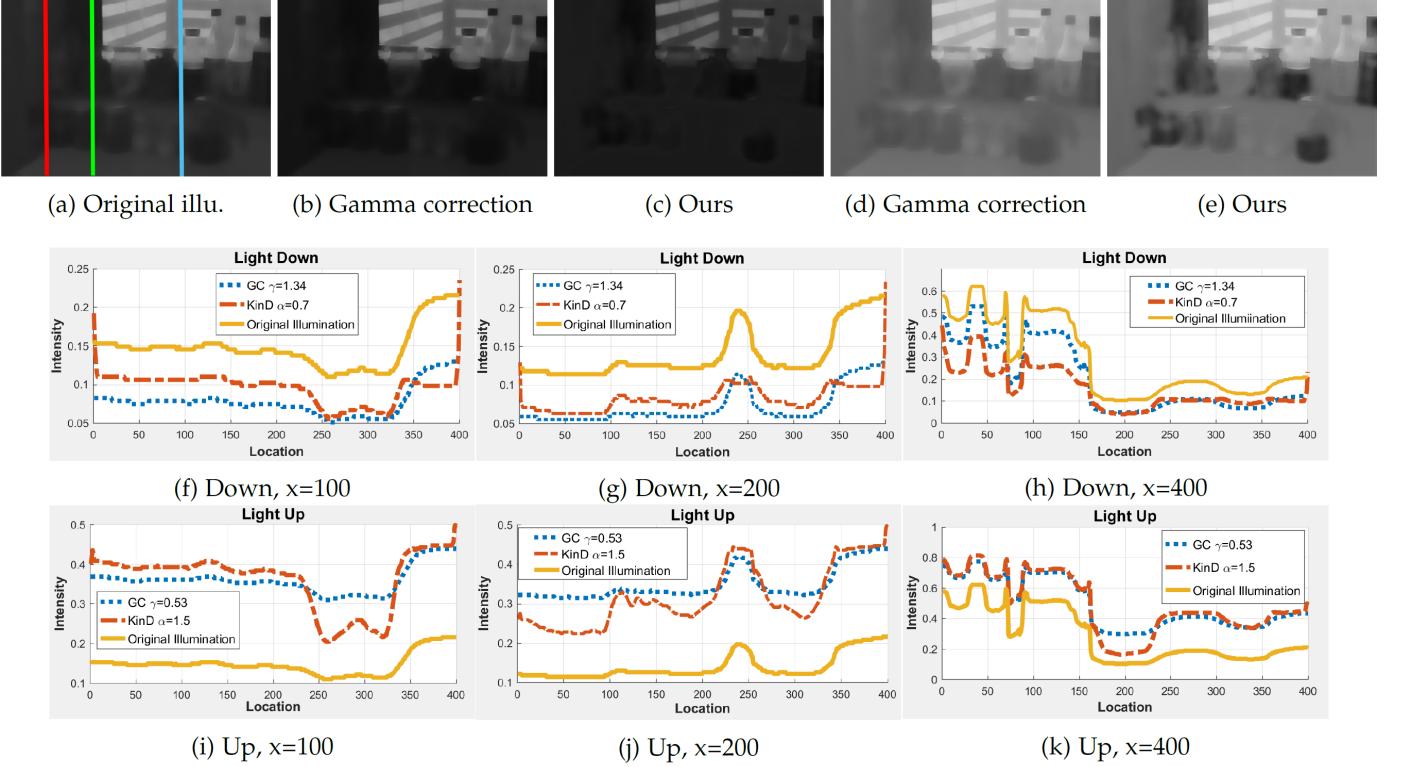


Fig. 6: Comparison between Gamma correction and our illumination adjustment manner. (a) shows the original/source illumination map. Two cases, including 1) turning the light down with $\gamma = 1.34$ (b) and $\alpha = 0.7$ (c), and 2) turning the light up with $\gamma = 0.53$ (d) and $\alpha = 1.5$ (e), are provided. (f)-(k) give the 1D curves at $x = 100, 200, 400$ corresponding to the red, green, and blue lines in (a), respectively.

Inputs	Operator	Kernel	Output Channels	Stride	Output Name
Decom_illumination, Ratio	Concat	-	2	-	Adjust_concat1
Adjust_conv1	Conv&ReLU	3×3	32	1	Adjust_conv1
Adjust_conv2	Conv&ReLU	3×3	32	1	Adjust_conv2
Adjust_conv3	Conv&ReLU	3×3	32	1	Adjust_conv3
Adjust_conv4	Conv	3×3	1	1	Adjust_conv4
	Sigmoid	-	1	-	Adjust_illumination

TABLE 3: Illumination adjustment network

term concentrates on the closeness in terms of textures. This subnet is similar to the reflectance branch in the layer decomposition subnet, but deeper. The schematic configuration is given in Figure 2 and detailed in Appendix. We recall that *the degradation distributes in the reflectance complexly, which strongly depends on the illumination distribution*. Thus, we bring the illumination information into the restoration net together with the degraded reflectance. The effectiveness of this operation can be observed in Figure 5. In the two reflectance maps with different degradation (light) levels, the results by BM3D can fairly remove noise (without regarding the color distortion in nature). The blur effect exists almost everywhere. In our results, the textures (the dust/water-based stains for example) of the window region, which is originally bright and barely polluted, keeps clear and sharp, while the degradations in the dark region get largely removed with details (*e.g.* the characters on the bottles) very well maintained. Besides, the color distortion is also cured by our method. The detailed reflectance restoration network configuration is provided in Table 2.

2.2.3 Illumination Adjustment Net

There does not exist a ground-truth light level for images. Therefore, for fulfilling diverse requirements, we need a mechanism to flexibly convert one light condition to another. We have paired illumination maps. Even though without knowing the exact relationship between the paired illuminations, we can roughly calculate their ratio of strength, *i.e.* α by $\text{mean}(\mathbf{L}_t / \mathbf{L}_s)$ where the division is element-wise. This ratio can be used as an indicator to train an adjustment function from a source light \mathbf{L}_s to a target one \mathbf{L}_t . If adjusting a lower level of light to a higher one, $\alpha > 1$, otherwise $\alpha \leq 1$. In the testing phase, α can be specified by users. The network is lightweight, containing 3 conv layers (two conv+ReLU, and one conv) and 1 Sigmoid layer. We notice that *the indicator α is expanded to a feature map, acting as a part of input for the net*. The following is the loss for illumination adjustment net:

$$\mathcal{L}^{IA} := \|\hat{\mathbf{L}} - \mathbf{L}_t\|_2^2 + \|\nabla \hat{\mathbf{L}} - \nabla \mathbf{L}_t\|_2^2, \quad (4)$$

where \mathbf{L}_t can be \mathbf{L}_h or \mathbf{L}_l , and $\hat{\mathbf{L}}$ is the adjusted illumination map from the source light (\mathbf{L}_h or \mathbf{L}_l) towards the target one. Figure 6 shows the difference between our learned adjustment function and gamma correction. For comparison fairness, we tune the parameter γ for gamma correction to reach a similar overall light strength with ours via $\gamma = \frac{\|\log(\hat{\mathbf{L}})\|_1}{\|\log(\mathbf{L}_s)\|_1}$. We consider two adjustments without loss of generality, including one light down and one light up. Figure 6 (a) depicts the source illumination, (b) and (d) are

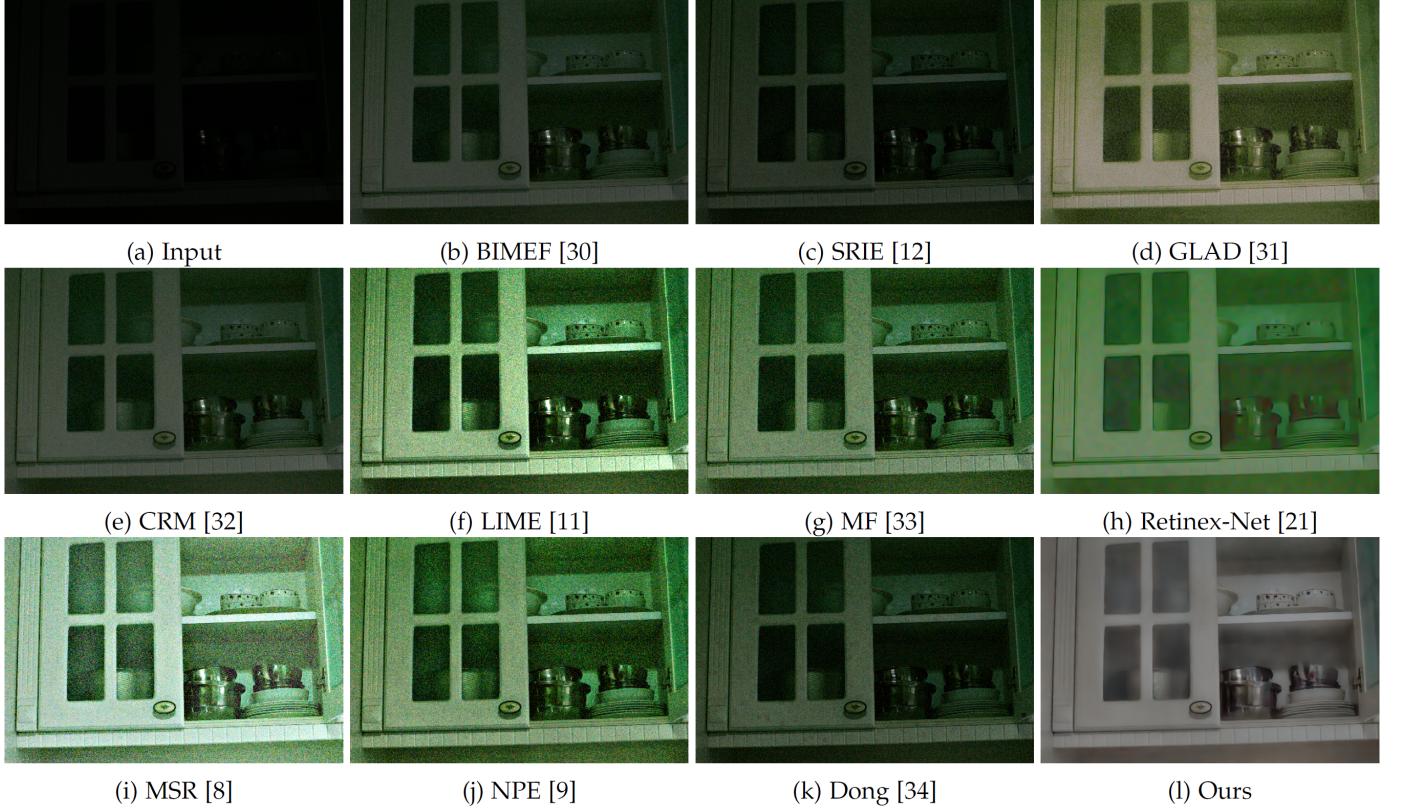


Fig. 7: Visual comparison with state-of-the-art low-light image enhancement methods.



Fig. 8: Visual comparison with state-of-the-art low-light image enhancement methods.

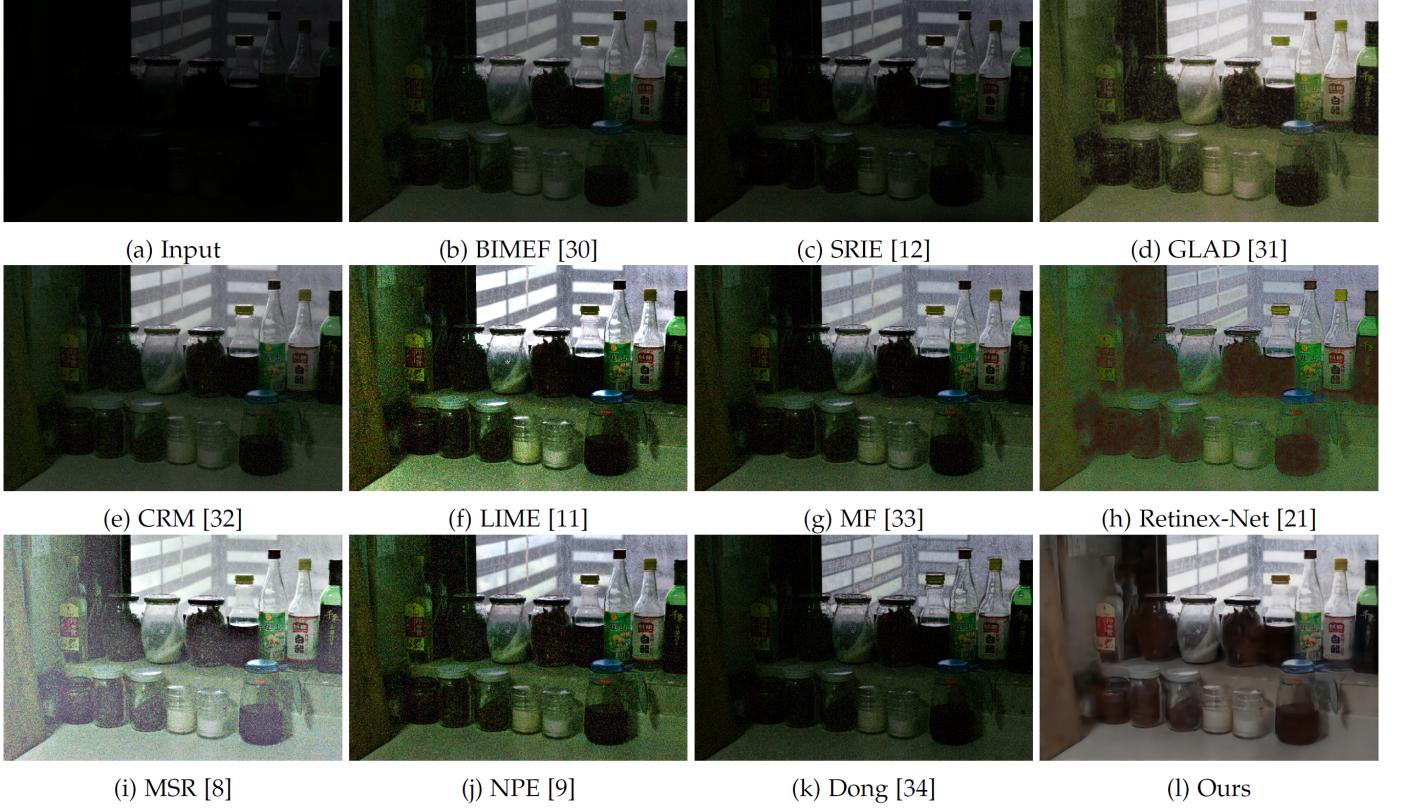


Fig. 9: Visual Comparison with state-of-the-art low-light image enhancement methods.

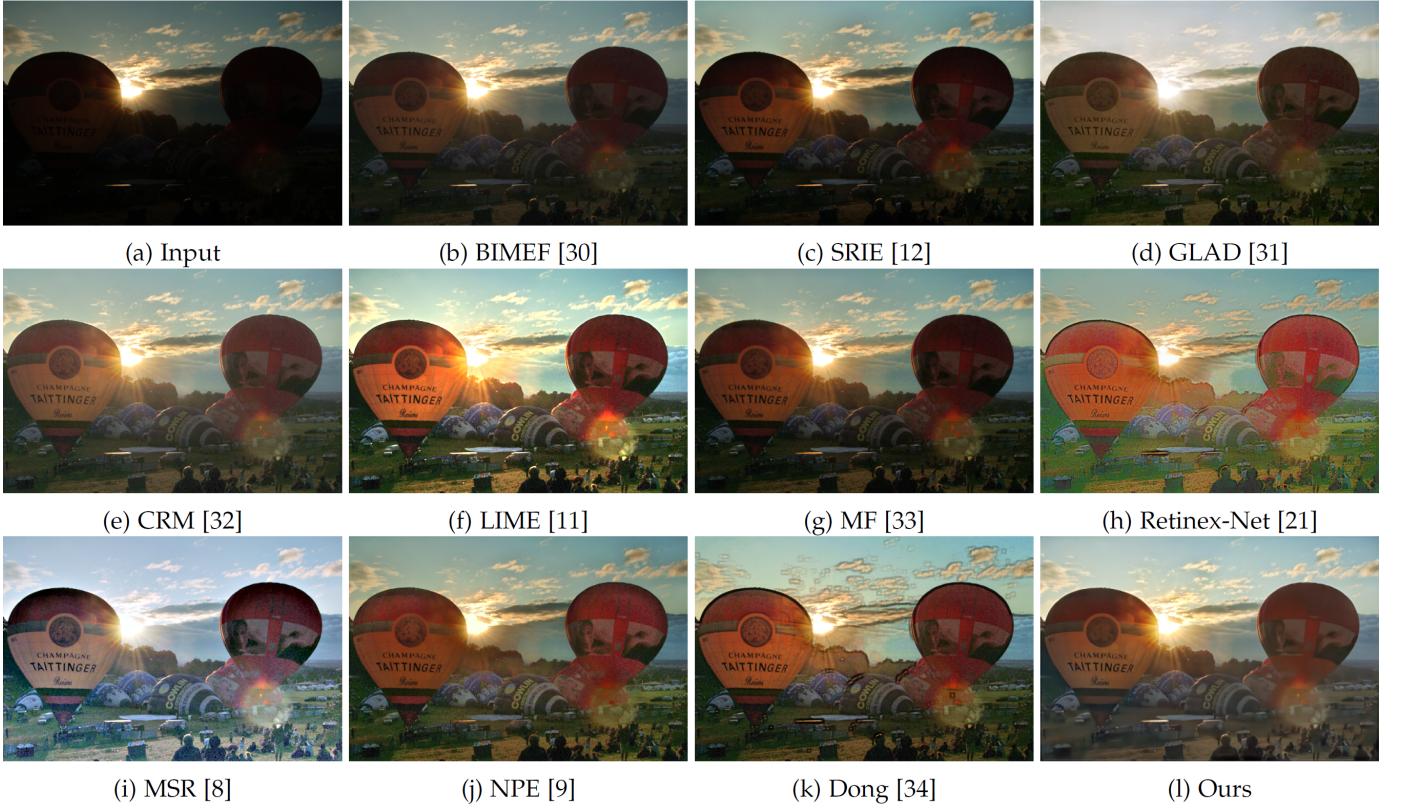


Fig. 10: Visual Comparison with state-of-the-art low-light image enhancement methods.



Fig. 11: Visual Comparison with state-of-the-art low-light image enhancement methods.

the adjusted results by gamma correction, while (c) and (e) are ours. To more clearly show the difference, we plot the 1D intensity curves at $x = 100, 200, 400$. As for the light-down case, our learned manner decreases more than gamma correction in intensity on relatively bright regions, while less or about the same on dark regions. Regarding the light-up case, the opposite trend appears. In other words, our method increases less the light on relatively dark regions, while more or about the same on bright regions. The learned manner is more corroborative with actual situations. Furthermore, the α fashion is more convenient than the γ way for users to manipulate. For instance, setting α to 2 means turns the light 2x up. The detailed illumination adjustment network configuration is provided in Table 3.

3 EXPERIMENTAL VALIDATION

3.1 Implementation Details

We use the LOL dataset as the training dataset, which includes 500 low/normal-light image pairs. In the training, we merely employ 450 image pairs, and no synthetic images are used. For the layer decomposition net, batch size is set to be 10 and patch-size to be 48x48. While for the reflectance restoration net and illumination adjustment net, batch size is set to be 4 and patch-size to be 384x384. We use the stochastic gradient descent (SGD) technique for optimization. The entire network is trained on a Nvidia GTX 2080Ti GPU and Intel Core i7-8700 3.20GHz CPU using the Tensorflow framework.

3.2 Performance Evaluation

We evaluate our method on widely-adopted datasets, including LOL [21], LIME [11], NPE [9], and MEF [30]. Four metrics are adopted for quantitative comparison, which are PSNR, SSIM, LOE [9], and NIQE [31]. A higher value in terms of PSNR and SSIM indicates better quality, while, in LOE and NIQE, the lower the better. The state-of-the-art methods of BIMEF [32], SRIE [12], CRM [33], Dong [34], LIME [11], MF [35], RRM [13], Retinex-Net [21], GLAD [36], MSR [8] and NPE [9] are involved as the competitors.

Table 4 reports the numerical results among the competitors on LOL dataset. For each testing low-light image, there is a “normal”-light correspondence. Thus, the correspondence can be taken as the reference to measure PSNR and SSIM. From the numbers, we see that our KinD significantly outperforms all the other methods. In terms of the non-reference metric NIQE, our KinD also takes the first place by a large margin. But, in LOE, our method seems falling behind many methods. As the authors of [11] stated, using the low-light input itself to compute LOE is problematic. One should choose a reliable reference. Similar to computing PSNR and SSIM, we again employ the correspondence image as the reference (denoted as LOE_{ref}). In this way, our KinD comes up to the 3rd place, slightly inferior to CRM (977.3 vs. 926.1). Regarding the LIME, NPE, and MEF datasets, no reference images are available. Thus, we only adopt the NIQE to evaluate the performance difference among the involved methods. In this comparison, as given in Tab. 5, our KinD shows its clear advantage against the others. Specifically, KinD outperforms all the competitors



Fig. 12: Visual Comparison with state-of-the-art low-light image enhancement methods.

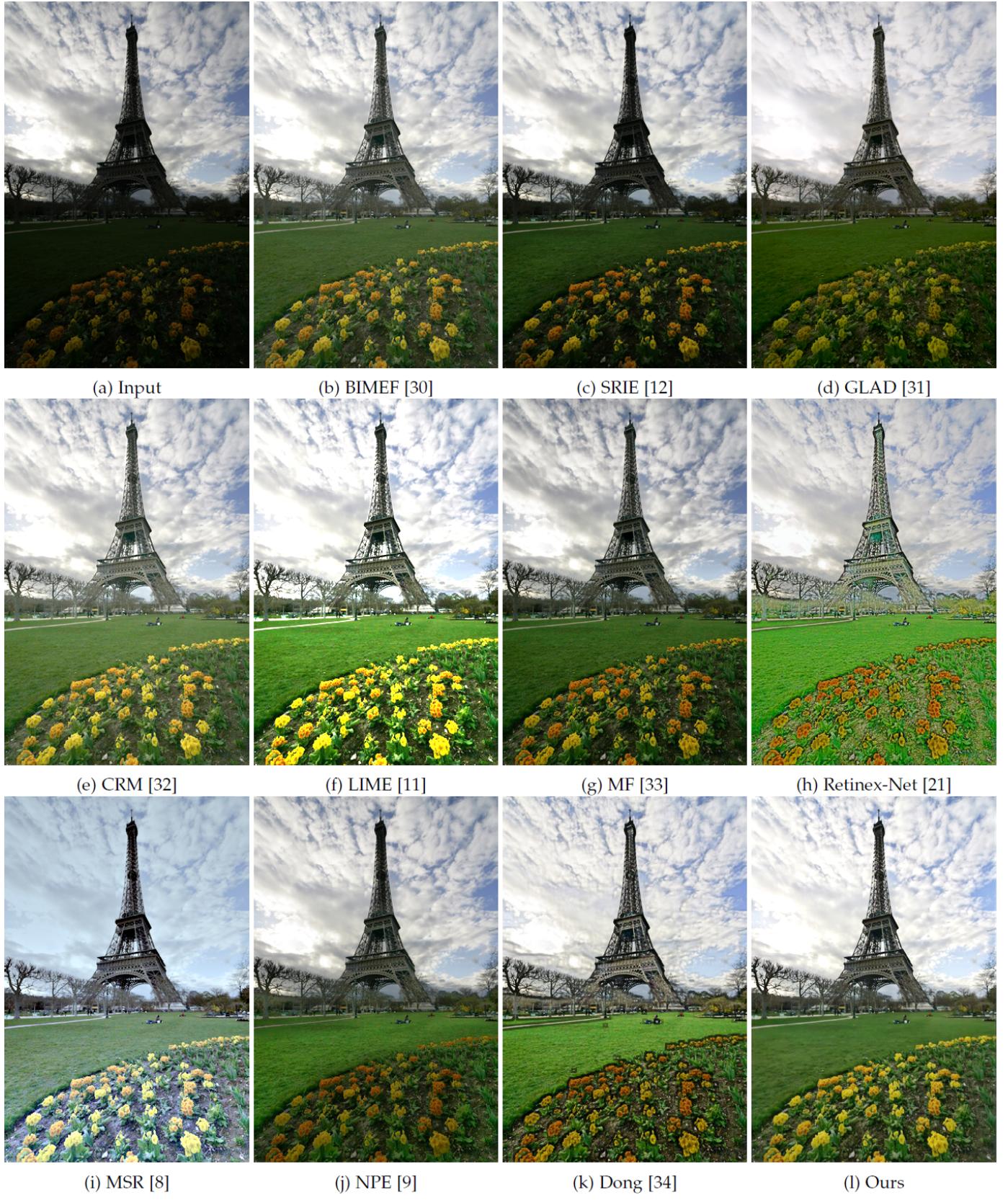


Fig. 13: Visual Comparison with state-of-the-art low-light image enhancement methods.

Metrics	BIMEF [32]	CRM [33]	Dong [34]	LIME [11]	MF [35]	RRM [13]
PSNR	13.8753	17.2033	16.7165	16.7586	18.7916	13.8765
SSIM	0.5771	0.6442	0.5824	0.5644	0.6422	0.6577
LOE	1456.1	1757.7	1283.2	1909.5	2051.7	2025.5
LOE _{ref}	985.9	926.1	1391.5	1342.4	1042.1	958.7
NIQE	7.5150	7.6865	8.3157	8.3777	8.8770	5.8101
Metrics	SRIE [12]	Retinex-Net [21]	MSR [8]	NPE [9]	GLAD [36]	KinD
PSNR	11.8552	16.7740	13.1728	16.9697	19.7182	20.8665
SSIM	0.4979	0.5594	0.4787	0.5894	0.7035	0.8022
LOE	1745.4	2449.3	2589.4	2076.3	1795.5	2012.2
LOE _{ref}	1199.8	2201.7	2084.8	1643.1	1017.1	<u>977.3</u>
NIQE	7.2869	8.8785	8.1136	8.4390	6.4755	5.1461

TABLE 4: Quantitative comparison on LOL dataset in terms of PSNR, SSIM, LOE, LOE_{ref}, and NIQE. The best results are highlighted in bold.

Metric	NIQE		
Datasets	LIME-data	NPE-data	MEF-data
BIMEF [32]	3.8169	4.1963	3.4237
CRM [33]	3.8546	3.9220	3.2708
Dong [34]	4.0516	4.1263	4.1094
LIME [11]	4.1549	4.2629	3.7159
MF [35]	4.0689	4.1096	3.4773
RRM [13]	4.6426	4.8452	4.1535
SRIE [12]	3.7863	3.9795	3.4577
Retinex [21]	4.5977	4.5674	4.4755
MSR [8]	3.7642	4.3663	3.6096
NPE [9]	3.9048	3.9520	3.5378
GLAD [36]	4.1280	3.9699	3.3435
KinD	3.7236	3.8826	<u>3.3429</u>

TABLE 5: Quantitative comparison on LIME, NPE, and MEF datasets in terms of NIQE. The best results are highlighted in bold.

on the LIME and NPE datasets. For the MEF data, it is only behind CRM by a small difference (3.34 vs. 3.27).

In addition, Figures 7-13 give a number of visual comparisons on the images with different light conditions. From the results, we can see that, although most of methods can somehow brighten the inputs, severe visual defects caused by unsatisfactory adjustment of light and/or obstinate noise and color distortion remain. Our KinD works well in these cases with the light properly adjusted and degradations clearly removed.

4 CONCLUSION

In this work, we have proposed a deep network, named KinD, for low-light enhancement. Inspired by Retinex theory, the proposed network decomposes images into the reflectance and illumination layers. The decomposition consequently decouples the original space into two smaller subspaces. As ground-truth reflectance and illumination information is in short, the network is alternatively trained using paired images captured under different light/exposure conditions. To remove the degradations previously hidden in the darkness, the proposed KinD builds a restoration module. A mapping function has also been learned in KinD, which better fits the actual situations than the traditional gamma correction, and flexibly adjusts light levels. Extensive experiments demonstrated the clear advantages of our design over the state-of-the-art alternatives. In the current version, KinD takes less than 50ms to handle an image in VGA resolution on a Nvidia 2080Ti GPU. By applying

techniques like MobileNet or quantization, our KinD can be further accelerated.

REFERENCES

- [1] E. Pisano, S. Zong, B. Hemminger, M. Deluca, R. Johnston, K. Muller, M. Braeuning, and S. Pizer, "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital Imaging*, vol. 11, no. 4, pp. 193–200, 1998.
- [2] H. D. Cheng and X. J. Shi, "A simple and effective histogram equalization approach to image enhancement," *Digital Signal Processing*, vol. 14, no. 2, pp. 158–170, 2004.
- [3] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Trans. on Consum. Electron*, vol. 53, pp. 593–600, May 2007.
- [4] C. Turgay and T. Tardi, "Contextual and variational contrast enhancement," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3431–3441, 2011.
- [5] C. Lee, C. Lee, and C. S. Kim, "Contrast enhancement based on layered difference representation of 2d histograms," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5372–5384, 2013.
- [6] E. H. Land, "The retinex theory of color vision," *Scientific American*, vol. 237, no. 6, pp. 108–128, 1977.
- [7] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 451–62, 1997.
- [8] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965–976, 2002.
- [9] S. Wang, J. Zheng, H. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, 2013.
- [10] X. Fu, D. Zeng, H. Yue, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, pp. 82–96, 2016.
- [11] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Trans Image Process*, vol. 26, no. 2, pp. 982–993, 2017.
- [12] X. Fu, D. Zeng, Y. Huang, X. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2782–2790, 2016.
- [13] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2828–2841, 2018.
- [14] J. Xie, L. Xu, E. Chen, J. Xie, and L. Xu, "Image denoising and inpainting with deep neural networks," in *NeurIPS*, pp. 341–349, 2012.
- [15] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2016.

- [16] C. Dong, C. L. Chen, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE TPAMI*, vol. 38, no. 2, pp. 295–307, 2016.
- [17] C. Dong, Y. Deng, C. C. Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *ICCV*, pp. 576–584, 2015.
- [18] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE TIP*, vol. 25, no. 11, pp. 5187–5198, 2016.
- [19] K. G. Lore, A. Akintayo, and S. Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017.
- [20] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, "Msr-net:low-light image enhancement using deep convolutional network," p. arXiv, 11 2017.
- [21] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *British Machine Vision Conference*, 2018.
- [22] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.
- [23] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3291–3300, 2018.
- [24] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2862–2869, 2014.
- [25] F. Agostinelli, M. R. Anderson, and H. Lee, "Adaptive multi-column deep neural networks with application to robust image denoising," in *NeurIPS*, 2013.
- [26] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *NeurIPS*, 2012.
- [27] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1256–1272, 2017.
- [28] X. Zhang, Y. Lu, J. Liu, and B. Dong, "Dynamically unfolding recurrent restorer: A moving endpoint control method for image restoration," in *ICLR*, 2018.
- [29] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, pp. 234–241, 2015.
- [30] L. Chulwoo, L. Chul, L. Young-Yoon, and K. Chang-Su, "Power-constrained contrast enhancement for emissive displays based on histogram equalization," *IEEE Trans Image Process*, vol. 21, no. 1, pp. 80–93, 2012.
- [31] A. Mittal, R. Soundararajan, and A. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3.
- [32] Z. Ying, L. Ge, and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement," *arXiv*, 2017.
- [33] Z. Ying, L. Ge, Y. Ren, R. Wang, and W. Wang, "A new low-light image enhancement algorithm using camera response model," in *IEEE International Conference on Computer Vision Workshop*, 2018.
- [34] X. Dong, Y. Pang, and J. Wen, "Fast efficient algorithm for enhancement of low lighting video," in *IEEE ICME*, pp. 1–6, 2011.
- [35] X. Fu, D. Zeng, H. Yue, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, pp. 82–96, 2016.
- [36] W. Wang, W. Chen, W. Yang, and J. Liu, "Gladnet: Low-light enhancement network with global awareness," in *IEEE International Conference on Automatic Face & Gesture Recognition*, 2018.