International Journal of Computer Vision (2021) 129:1013–1037
https://doi.org/10.1007/s11263-020-01407-x

# Beyond Brightening Low-light Images

**Yonghua Zhang[1] · Xiaojie Guo[1] · Jiayi Ma[2] · Wei Liu[3] · Jiawan Zhang[1]**

Received: 17 April 2020 / Accepted: 13 November 2020 / Published online: 6 January 2021


**Abstract**
Images captured under low-light conditions often suffer from (partially) poor visibility. Besides unsatisfactory lightings, multiple types of degradation, 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 pollution. Thus, low-light image enhancement should not only brighten dark regions, but also remove hidden artifacts. To achieve the goal, this work builds a simple yet effective network, which, inspired by Retinex theory, decomposes images into two components. Following a divide-and-conquer principle, 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 for better regularization/learning. It is worth noticing 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 the state-of-the-art alternatives, especially in terms of the robustness against severe visual defects and the flexibility in adjusting light levels. Our code is made publicly available at https://github.com/zhangyhuaee/KinD_plus.

**Keywords** Low-light image enhancement · Image decomposition · Image restoration · Light manipulation

## 1 Introduction

Veryoften,capturinghigh-qualityimagesindim-lightconditions 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 a low (signal-to-noise ratio) SNR. Long exposure is limited to shootstaticscenes,otherwiseithighlylikelygetsintroubleof 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-lightimageenhancement hasbeenalong-standingproblem in the community with a great progress made over the past years, developing an effective low-light image enhancer for simultaneously

CommunicatedbyMichaelS.Brown.

Y.ZhangandX.Guocontributedequallytothisstudy

B Xiaojie Guo xj.max.guo@gmail.com
Yonghua Zhang
zhangyonghua@tju.edu.cn
Jiayi Ma
jyma2010@gmail.com
Wei Liu
wl2223@columbia.edu
Jiawan Zhang
jwzhang@tju.edu.cn

[1] College of Intelligence and Computing, Tianjin University, Tianjin 300350, China
[2] Electronic Information School, Wuhan University, Wuhan 430072, China
[3] Tencent AI Lab, Shenzhen 519000, China

lightening the darkness and removing the degradations still remains challenging.

Figure 1 provides three natural images captured under different 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 degradation factors show up as given on the top-right corner. The second image is photographed at sun-

**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, so we show its X20 version on the top-right corner

set (weak ambient light), in which most objects suffer from backlighting. Imaging at noon facing to the light source (the sun)alsohardlygetsridoftheissuelikewhatthesecondcase 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 numerous low-level vision tasks, such as denoising and super-resolution, most of which need training examples with references. For the target problem, namely, the low-light image enhancement, no unified best light conditions exist, although the order of light intensity can be determined. In other words, one cannot say what light condition is the best. Because, from the viewpoint of users, the favorite light levels for different people/requirements could be much diverse. 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 componentfrom a single image, and flexibly adjust the light level?
- How to remove the degradation like noise and color distortion previously hidden in the darkness after lightening up dark regions?
- How to train a model without well-defined best light con-ditions for low-light image enhancement by only looking
  at example pairs captured under different light conditions?

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

## 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.

### 1.1.1 Plain Methods

Intuitively, for an image with globally low light, the visibility can be enhanced by directly amplifying it. But, as shown in the first case of Fig. 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 Fig. 1, this operation easily results in (partial) saturation or overexposure. One technical line, with histogram equalization (HE) (Pisano et al. 1998; Cheng and Shi 2004; AbdullahAl-Wadud et al. 2007) and its follow-ups (Turgay and Tardi 2011; Lee et al. 2013) 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 contrasts of images. Another mapping manner is gamma correction (GC), which is carried out on each pixel individually in a non-linear way. Using a global parameters for each pixel often leads to over-enhancement or under-enhancement, and the selection of appropriate gamma parameter is often heuristic. Several adaptive gamma correction (AGC) algorithms have been proposed to relieve this problem (Huang et al. 2013; Rahman et al. 2016; Wang et al. 2009). Specifically, Huang et al. (2013) proposed an AGC with weighting distribution (AGCWD) for contrast enhancementbysettinggammaasafunctionofcompensated

cumulative distribution. Rahman et al. (2016) developed an AGC method that dynamically determines an intensity transformation function based on the statistical characteristics of the input image. 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.

#### 1.1.2 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 (Land 1977), is that the (color) image can be decomposed into two components, *i.e.*, reflectance and illumination. Early attempts include singlescale Retinex (SSR) (Jobson et al. 1997) and multi-scale Retinex (MSR) (Jobson et al. 2002). Limited to the manner of producing the final result, the output often looks unnatural and somewhere over-enhanced. Wang et al. (2013) proposed a method called NPE, which jointly enhances contrast and preserves naturalness of illumination. Ma et al. (2015) developed a method, which adjusts the illumination through fusing multiple derivations of the initially estimated illumination map. However, it sometimes sacrifices the realism of those regions containing rich textures. Guo et al. (2017)focusedonestimatingthestructuredilluminationmap from an initial one. These methods generally assume that the images are noise- and color distortion-free, and do not explicitly consider degradation factors. In Fu et al. (2016), a weighted variational model for simultaneous reflectance and illumination estimation (SRIE) was designed to obtain better reflectance and illumination layers, and then the target image is generated by manipulating the illumination. Following (Guo et al. 2017), Li et al. (2018) further introduced an extra term to host noise. Despite (Fu et al. 2016) and (Li et al. 2018) can reject slight noises in images, they are short of abilities in handling color distortion and heavy noise.

#### 1.1.3 Deep Learning-based Methods

With the emergence of deep learning, a number of low-level vision tasks have benefited from deep models, such as (Xie et al. 2012; Zhang et al. 2016) for denoising, (Dong et al. 2016) for super-resolution, (Dong et al. 2015) for compression artifact removal, and (Cai et al. 2016) for dehazing. Regarding the target mission of this paper, the low-light net (LLNet) proposed in Lore et al. (2017) builds a deep network that performs as a simultaneous contrast enhancement and denoising module. Shen et al. (2017) 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) to learn an end-to-end mapping between dark and bright images. Wei et al. (2018) designed a deep network, called Retinex-Net, that integrates image decomposition and illumination mapping. Please notice that Retinex-Net additionally employs an off-the-shelf denoising tool (BM3D (Dabov et al. 2007)) to clean the reflectance component. These strategies all assume that there exist images with "best" 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 common in real images. More recently, Chen et al. (2018) proposed a pipeline for processing low-light images based on end-to-end training of a fully convolutional network, which can jointly deal with noise and color distortion. But, this work is specific to data in RAW format, limiting its applicable scenarios. As stated in Chen et al. (2018), if modifyingthenetworktoacceptdatainJPEGformat,theperformancesignificantlydrops.Ignatovetal.(2018)introduced a weakly supervised photo enhancer (WESPE) to translate photostakenbycameraswithlimitedcapabilitiesintoDSLRqualityphotos.Wangetal.(2019)presentedaneuralnetwork (DUPE) to estimate illumination maps that are then used for enhancing underexposed photos. Chen et al. (2018) developed a photo enhancer (DPE) by using two-way generative adversarial networks (GANs).

Most existing methods manipulate the illumination by 1) gamma correction, 2) appointing a level existing in carefully constructed training data, or 3) fusion of different illumination maps. 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,itisdesiredtolearnamappingfunctiontoarbitrarily convert one light (exposure) level to another for offering users the flexibility of adjustment.

#### 1.1.4 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 the past decades. Classic ones model/regularize the problem by utilizing some specific priors of natural clean images, like non-local self-similarity, piecewise smoothness, signal (representation)sparsity,*etc.*Themostpopularschemesarguably go to BM3D (Dabov et al. 2007) and WNNM (Gu et al. 2014). Due to the high complexity of the 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 (Agostinelli et al. 2013; Xie et al. 2012),TNRDbytrainablenonlinearreactiondiffusion(Chen and Pock 2017), DnCNN with residual learning and batch normalization (Zhang et al. 2016), 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 (Zhang et al. 2018). But, none of the mentioned approaches considers that different regions of a light-enhancedimagehostdifferentlevelsofnoise.Thesame problem happens to color distortion. Recently, Zhang et al. (2018) proposed a denoising convolutional neural network named FFDNet, which can be used to remove spatially variant noise by specifying a non-uniform noise level map. For a noisy image, its noise level is typically fixed, which depends on the hardware. In this sense, the ability of FFDNet in spatial variance seems barely useful in practice. In this work, we will show that, by decomposing a typical image into its reflectance and illumination, the spatially-variant characteristic becomes active, since the reflectance map appears to be non-uniformly noisy, while the illumination *per se* can naturally perform as an indicator to reflect the noise level of each pixel/position.

### 1.2 Our Contributions

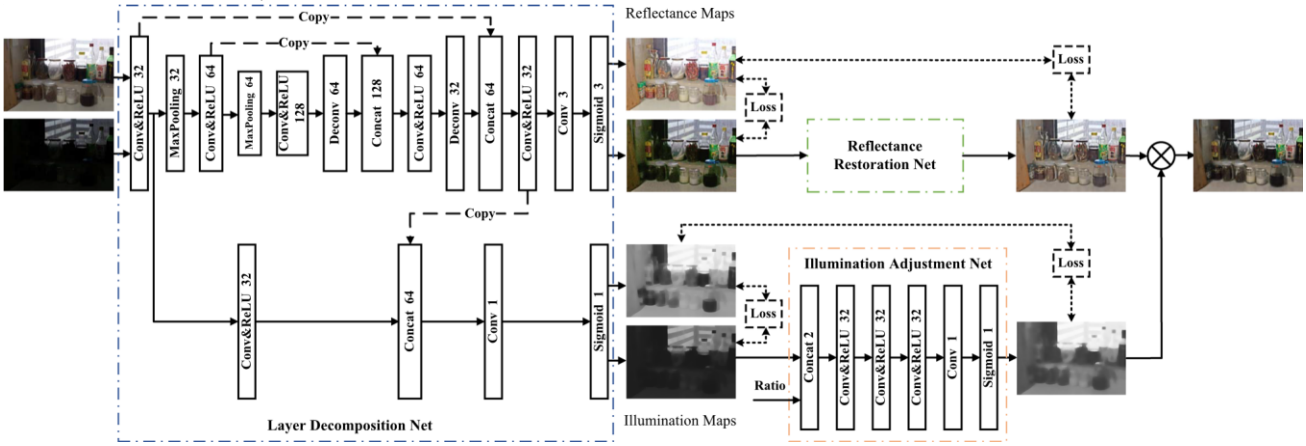
This study presents a deep network for practically solving the low-light enhancement problem. The main contributions of this work can be summarized in the following aspects.

- InspiredbyRetinextheory,theproposednetworkdecomposes images into two components, *i.e.*, reflectance and illumination,whichdecouples theoriginal spaceintotwo smaller ones.
- Thenetworkistrainedwithpairedimagescapturedunderdifferent light/exposure conditions, instead of using any ground-truth reflectance and illumination information.
- Our designed model provides a mapping function forflexibly adjusting light levels according to different demands from users.
- The proposed network also contains a module, which iscapable of effectively removing visual defects amplified through lightening dark regions.
- Extensive experiments are conducted to demonstrate theefficacy of our design and its superiority over state-ofthe-art alternatives.

A preliminary version of this manuscript appears in Zhang et al. (2019). Compared with (Zhang et al. 2019), this journal version presents a novel *multi-scale illumination attention* module(MSIA),whichcanalleviatevisualdefects(*e.g.*,nonuniformspotsandover-smoothing)leftinZhangetal.(2019). It also gives deeper ablation studies to investigate the effectiveness of different possible network architectures and loss functions. More experimental comparisons are conducted to verify the advantages of our method, and more applicable scenarios are discussed. To allow more comparisons from the community, we release our code at https://github.com/ zhangyhuaee/KinD_plus.

## 2 Methodology

A desired low-light image enhancer should be capable of effectively removing the degradation hidden in the darkness, and flexibly adjusting light/exposure conditions. In Zhang et al. (2019), we have built a deep network, denoted as KinD, to achieve the goal. However, we observed that in some cases the reflectance restoration results of KinD still contain some artifacts, such as over-exposure and non-uniform light spots. To further improve the enhancement quality of KinD, we design a new multi-scale illumination attention module. For convenience, we name the new network as KinD++. As schematically illustrated in Fig. 2, the network is composed of two branches for handling the reflectance and illumination components, respectively. From the perspective of functionality, it can also be divided into three modules, *i.e.*, layer decomposition, reflectance restoration, and illumination adjustment. In the next subsections, we shall explain the details about the network.




## 2.1 Motivation and Consideration

This part describes the main principles/considerations from four aspects, which support our work.

### 2.1.1 Layer Decomposition and Divide-and-conquer

As discussed in Sect. 1.1, the main drawback of plain methods comes from the blindness of illumination. Thus, it is the 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 **I** can be viewed as a composition of two components, *i.e.*, reflectance **R** and illumination **L**, in the fashion of $\mathbf{I} = \mathbf{R} \otimes \mathbf{L}$, where $\otimes$ designates the element-wise product.

Further, decomposing images in the Retinex manner consequentlydecouples thespaceofmapping adegraded 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 analysis, the Retinex-based layer decomposition is suitable and necessary for the target task.

### 2.1.2 Data Usage and Priors

There is no well-defined best light condition for an image. Furthermore, no/few ground-truth reflectance and illumination maps for real images are available. The layer decomposition problem is in nature under-determined, so additionalpriors/regularizersmatter.Supposethattheimages 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 structures. In real situations, the degradation embodied in low-light images is 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 (pseudo 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.Pleasenoticethattheusageofreflectance(well-defined) totally differs from using images in (relatively) bright light as the reference of those dim-light ones.

**Fig. 2** The network architecture. Two branches correspond to the reflectanceandillumination,respectively.Fromtheperspectiveoffunction ality, it can also be divided into three modules, including layer decomposition, reflectance restoration, and illumination adjustment.means the element-wise multiplication. Digits are channel numbers $\otimes$

### 2.1.3 Illumination Guided Reflectance Restoration

Inthedecomposedreflectance,thepollutioninregionscorrespo ndingtodarkerilluminationisheavierthanthattobrighter one. Mathematically, a degraded low-light image can be naturally modeled as $\mathbf{I} = \mathbf{R} \otimes \mathbf{L} + \mathbf{E}$, where **E** designates the pollution component. By taking simple algebra steps, we have the following:

$$\begin{aligned} \mathbf{I} &= \mathbf{R} \otimes \mathbf{L} + \mathbf{E} = \tilde{\mathbf{R}} \otimes \mathbf{L} = (\mathbf{R} + \tilde{\mathbf{E}}) \otimes \mathbf{L} \\ &= \mathbf{R} \otimes \mathbf{L} + \tilde{\mathbf{E}} \otimes \mathbf{L}, \end{aligned} \tag{1}$$

where $\mathbf{R}$˜ stands for the polluted reflectance, and $\mathbf{E}$˜ is the degradation having the illumination decoupled. The relationship $\mathbf{E} = \mathbf{E}$˜ $\otimes\mathbf{L}$ holds. Taking AWGN $\mathbf{E} \sim N(0,\sigma^2)$ for an example, the distribution of $\mathbf{E}$˜ becomes much more complex[2] and strongly relates to $\mathbf{L}$, *i.e.*, $\sigma_{\mathbf{L}i}$ 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 or applications may be pretty diverse. Therefore, a practical system needs to provide an interface for arbitrary illumination manipulation. In the literature, three main ways for enhancing light conditions are (1) fusion, (2) light level appointment, and (3) 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 $\gamma$ 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 Network Design

Driven by the motivation, we build a deep neural network for simultaneously kindling the darkness, removing the degradation, and providing users a friendly light adjustment manner.

**Fig. 3** Upper row: Lower light input and its decomposed illumination and (degraded) reflectance maps. Lower row: Brighter input and its corresponding maps. Three columns respectively correspond to inputs, illumination maps, and reflectance maps. These are testing images

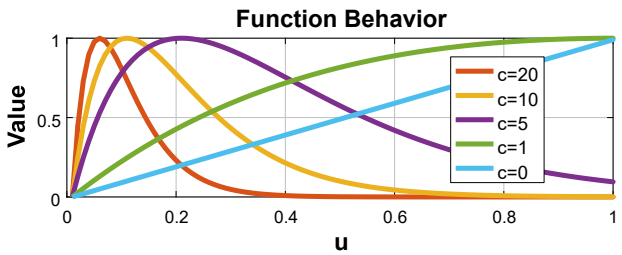



**Fig. 4** The behavior of function $v = u \cdot \exp(-c \cdot u)$. The parameter $c$ controls the shape of function

Below, we describe the three subnets in details from the functional perspective.

#### 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, so 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 $L_{rs}^D = \|\mathbf{R}_l - \mathbf{R}_h\|_1$ to regularize the

*reflectancesimilarity*,where· $_1$ means $\ell^1$ norm. The $_D$ $\nabla\mathbf{L}$ *illumination smoothness* is constrained by $\mathrm{L}_{is}$ $= \| \frac{l}{\max(\|\nabla\mathbf{I}_l\|,\epsilon)} \|_1 + \max\ \|\nabla\mathbf{I}(\ _h\ ,\epsilon)\ \|_{\nabla\mathbf{L}}$ , where $\nabla$ stands for the first order derivative operator containing $|\nabla_x$ (horizontal) and $\nabla_y$ (vertical) directions. In addition, 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.Foralocationononanedgein$\mathbf{I}$,thepenaltyon$\mathbf{L}$issmall; while for a location in a flat region in $\mathbf{I}$, the penalty turns to be large. Compared with the traditional total variation norm

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

that smooths the entire map equally , this relative structure takes the original input as reference to reduce the risk of over-smoothing on structural boundaries. As for the *mutual consistency*, we employ $\mathrm{L}_{mc}{}^{D} = \|\mathbf{M} \otimes \exp(-c \cdot \mathbf{M})\|_1$ with $\mathbf{M}$ defined by $|\nabla\mathbf{L}_l| + |\nabla\mathbf{L}_h|$. Figure 3 depicts the function behavior of $u \cdot \exp(-c \cdot u)$, where $c$ is the parameter controlling the shape of function. As can be seen from Fig. 3, the penalty first goes up and then drops towards 0 as $u$ increases. Taking $c = 10$ for example, the curve reaches the peak at $u$ (the sum of $|\nabla\mathbf{L}_h|$ and $|\nabla\mathbf{L}_l|$) around 0.1. It means that one of the magnitudes is or both of them are weak, which should be heavily punished, and thus be removed/smoothed out. As the value of $u$ increases (both are strong edges), the penalty decreases and the corresponding edge is better maintained. This characteristic 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 $^1$ loss on $\mathbf{M}$. In this work, adopting $c = 10$ performs reasonably well. Besides, the decomposed two layers should reproduce the input, which is constrained by the *reconstruction error*, *i.e.* $\mathrm{L}_{re}{}^{D} = \|\mathbf{I}_l - \mathbf{R}_l \otimes \mathbf{L}_l\|_1 + \|\mathbf{I}_h - \mathbf{R}_h \otimes \mathbf{L}_h\|_1$.

As a result, the loss function of layer decomposition net is as follows:

$$\mathcal{L}^D = \mathcal{L}^D_{re} + \omega_{rs}\mathcal{L}^D_{rs} + \omega_{mc}\mathcal{L}^D_{mc} + \omega_{is}\mathcal{L} \qquad {}^D \qquad {}_{is}. \tag{2}$$

In our experiments, setting $\omega_{rs} = 0.009$, $\omega_{mc} = 0.2$ and $\omega_{is} = 0.15$ performs sufficiently well. Equipped with the carefully designed loss terms, the layer decomposition network can be considerably simple, which contains two branches corresponding to the reflectance and illumination, respectively. The reflectance branch adopts a typical 5-layer U-Net(Ronnebergeretal.2015),followedbyaconvolutional (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.

### 2.2.2 Reflectance Restoration Net

The reflectance maps from low-light images, as shown in Figs. 4 and 5 , are more interfered by degradations than those from bright-light ones. Employing the clearer reflectance to act as the reference (pseudo ground-truth) for the messy one is our principle. For seeking a restoration function, two loss terms are used in the following way:

$$\mathcal{L}^R = \mathcal{L}^R_{mse} + \mathcal{L}^R_{dsim}. \tag{3}$$

The first term $\mathrm{L}_{mse}{}^{R}$ represents the *mean squared error* (MSE) between $\mathbf{R}_h$ and $\mathbf{R}\hat{}$ , *i.e.*, MSE($\mathbf{R}_h$,$\mathbf{R}\hat{}$ ), where $\mathbf{R}\hat{}$ corresponds to the restored reflectance. The second term $\mathrm{L}_{dsim}{}^{R} = 1 - \mathrm{SSIM}(\mathbf{R}_h,\mathbf{R}\hat{})$ evaluates the closeness in terms of structural similarity. Minimizing $\mathrm{L}^R$ expects high values in terms of PSNR and SSIM, two most commonly-used reference metrics for measuring image quality.

The degradation complexly distributes in the reflectance, which strongly depends on the illumination distribution. Thus, we bring the illumination information into the restoration net together with the degraded reflectance. A reflectance restoration net based on a *multi-scale illumination attention* (MSIA) module is designed to deal with degradation in decomposed reflectance maps, the architecture of which is given in Fig. 7a. The proposed net consists of 10 convolutionallayersand4MSIAmodules.AsdisplayedinFig.6 ,the MSIA module includes an illumination attention module and a multi-scale module. Notice that the scale numbers in MSIA can be varied as demanded. The effectiveness of this operation can be observed in Fig. 5. In the two reflectance maps

withdifferentdegradation(light)levels,theresultsbyBM3D can fairly remove noise (without regarding the color distortion in nature). The blur effect exists almost everywhere. In our results, the texture (the dust/water-based stains for example) of the window region, which is originally bright and barely polluted, keeps clear and sharp, while the degra-

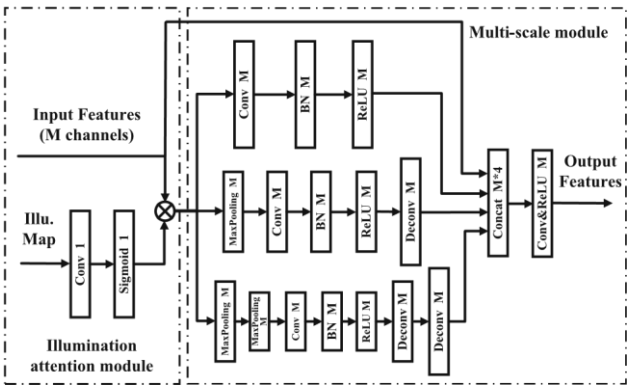


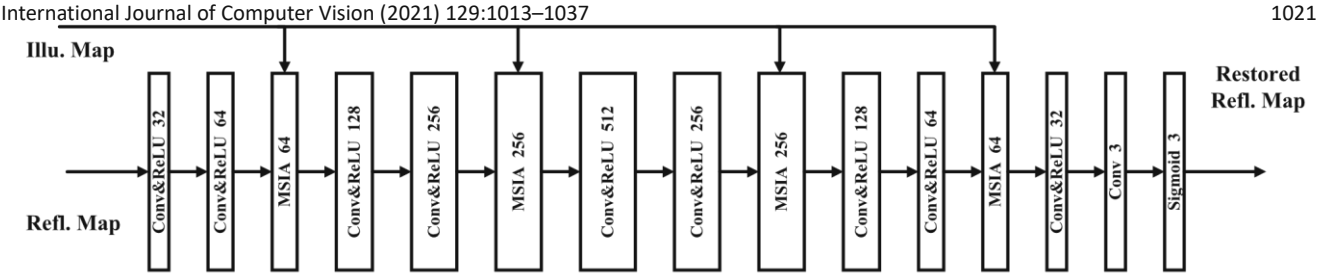

**Fig. 6** The proposed MSIA module includes an illumination attention module and a multi-scale module

dations in dark regions 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.

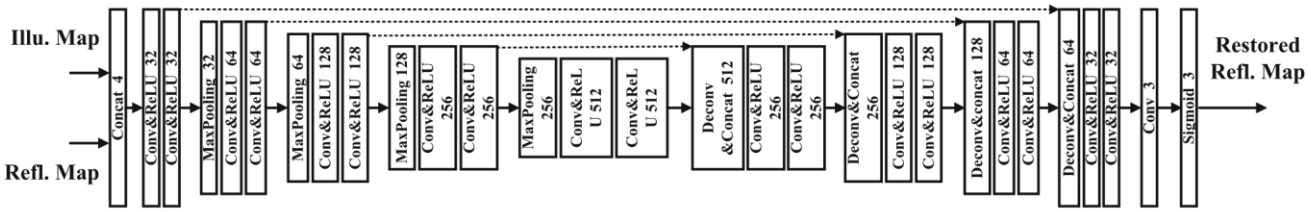
Please notice that, in the previous version (Zhang et al. 2019), the restoration net is in a U-net shape as shown in Fig. 7b. KinD is able to do noise removal and color correctionconsiderablywellthough,therestoredreflectancere sults in some cases may exist some defects such as over-exposure and halo artifacts. In fact, excessive pooling and upsampling operators in U-net easily generate halos. Using *fully convolutional neural networks* (FCN) like FFDNet (Zhang et al. 2018) can reduce such annoying troubles, however, merely using FCN cannot get satisfying restoration results. That is why we introduce the MSIA. The illumination module can better guide the net to pay more attention to the worse degraded regions and the multi-scale module can extract more abundant features to restore color and details. An evidence is given in Fig. 8, from which we can observe that the results by KinD clearly reject noise, but they suffer from halos/unbalanced lights and, over-exposing as well as over-smoothing effects in some regions. KinD++ with the MSIA disabled (KinD++ w/o MSIA) effectively eliminates the halo/light-unbalance issue, but the (heavy) noise cannot be removed thoroughly. As revealed in Fig. 8e, the complete KinD++exhibitsitssuperiorityindegradationremoval,lightba lance, and detail preservation. Although the complete KinD++ works well for most regions, we find the enhanced results exist some over-enhanced artifacts in regions that are originally black. Through re-analyzing the layer decomposition net and reflectance restoration net, we find that the illness comes from the unaligned $\mathbf{R}_h$ and $\mathbf{R}_l$. Because $\mathbf{R}_l$ is often interfered by heavy noise as shown in Fig. 9, the overall (mean) intensity of $\mathbf{R}_l$ is suppressed compared with that of $\mathbf{R}_h$. When feeding such unaligned reflectance pairs into the restoration net, the learned mapping inevitably increases the intensities of restored reflectances to match the overall intensity of $\mathbf{R}_h$, and thus brings the risk of over-enhancement. To remedy this problem, we propose to align the paired reflectances via adjusting $\mathbf{R}_h$ by $\mathbf{R}^-_h \leftarrow \mathbf{R}_h^{\beta}$, where $\beta \geq 1$ is to control the adjustment. Instead of the multiplication fashion of adjustment *i.e.* $\mathbf{R}^-_h \leftarrow \beta\mathbf{R}_h$, we choose the power one inspired by Stevens's power law (Stevens 1957), which is an empirical relationship in psychophysics between an increased intensity or strength in a physical stimulus and the perceived magnitude increase in the sensation created by the stimulus. The value of $\beta$ is simply determined by $\min(\text{mean}(\frac{\mathbf{R}_h}{}), \epsilon)_{\mathbf{R}^l}$, where the division is element-wise and $>1$ is the upper-bound of adjustment power. In our experiments, setting to 2 works sufficiently well. By doing so, the over-enhancement problem is largely mitigated. Figure 10 shows the comparison of enhanced results with and without the adjustment.

**Fig. 7** The network architectures of reflectance restoration used by our new KinD++ **(a)** and previous KinD (Zhang et al. 2019) **(b)**


**Fig. 8** **a** and **b** display two polluted reflectance maps and their corresponding illumination maps, respectively. **c**–**e** provide the restored results by

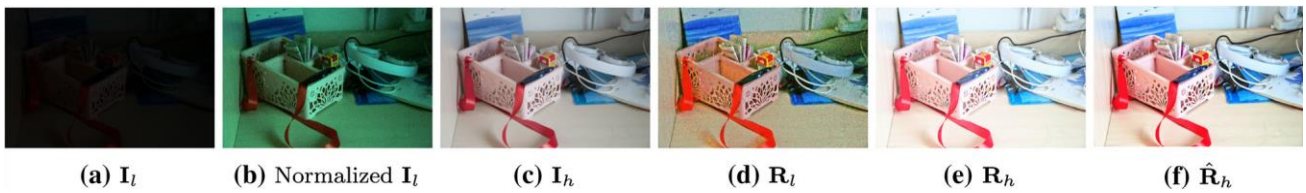



**(a)** Restoration network that KinD++ adopts.




**(b)** Restoration network that KinD [52] adopts.

KinD (Zhang et al. 2019) (Fig. 7b), KinD++ without MSIA module, and KinD++ (Fig. 7a), respectively



**(a)** $\mathbf{I}_l$ **(b)** Normalized $\mathbf{I}_l$ **(c)** $\mathbf{I}_h$ **(d)** $\mathbf{R}_l$ **(e)** $\mathbf{R}_h$ **(f)** $\hat{\mathbf{R}}_h$

**Fig. 9** Visual comparison on an unaligned reflectance pair, and the adjusted result of $\mathbf{R}_h$

**Fig. 10** The enhanced results without and with data alignment

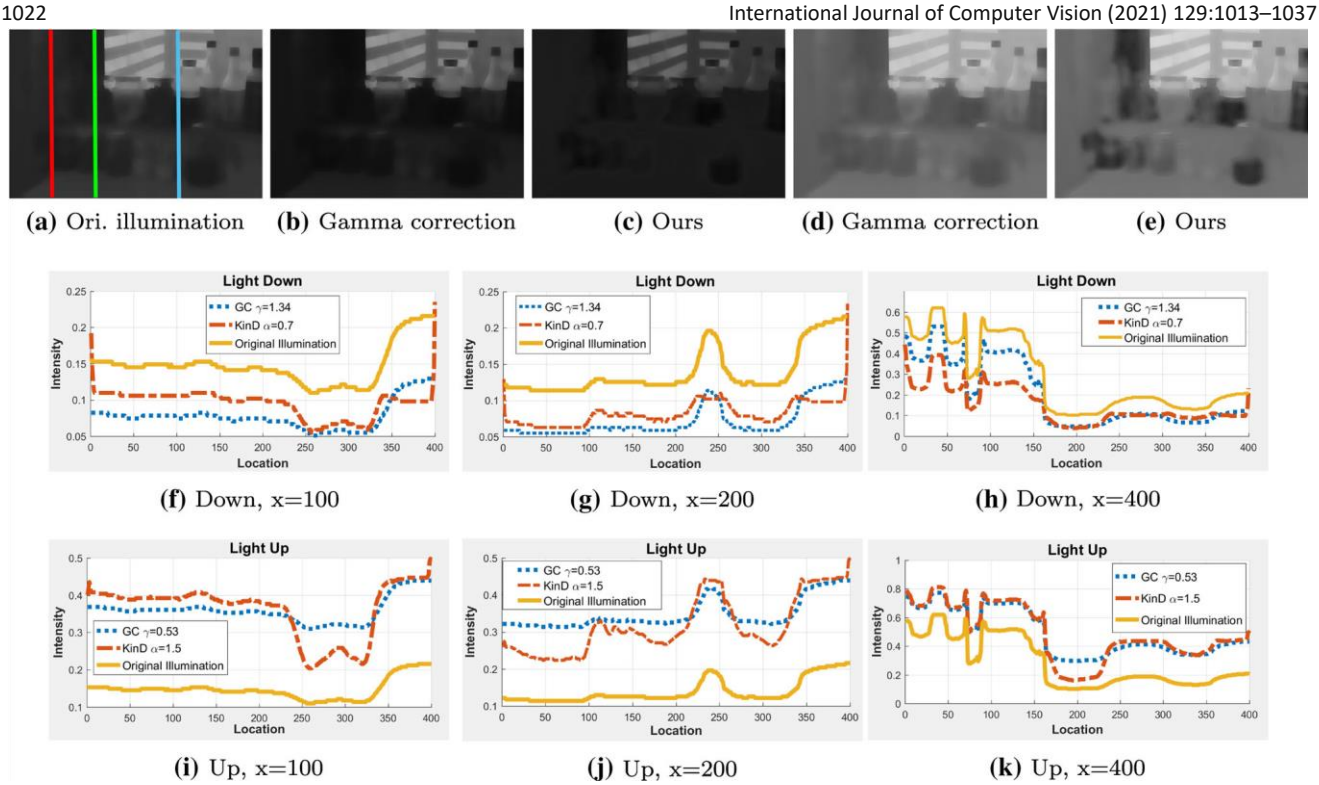



**Fig. 11** Comparison between Gamma correction and our illumination α = 0.7 **(c)**, and (2) turning the light up with γ = 0.53 **(d)** and α = 1.5 adjustment manner. **a** shows the original/source illumination map. Two **(e)**, are provided. **f**–**k** give the 1D curves at *x* = 100,200,400 correcases, including (1) turning the light down with γ = 1.34 **(b)** and sponding to the red, green, and blue lines in **(a)**, respectively

**Fig. 12** Two examples with different light levels manipulated by setting α = {0.2,0.7,1.0,2.0,5.0} and α = {0.5,1.0,3.0,5.0}, respectively

#### 2.2.3 Illumination Adjustment Net

There does not exist a unified best 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 strengths, *i.e.*, $\alpha$ by 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, $\alpha$ can be specified by users. The indicator $\alpha$ is first expanded to be a single channel feature map of the same size of $\mathbf{L}_s$, whose elements are all $\alpha$. Then the expanded feature map is concatenated with $\mathbf{L}_s$, acting as the input of illumination adjustment net. This network is lightweight, containing 4 conv layers (three conv+ReLu, and one conv) and 1 Sigmoid layer. The loss for the illumination adjustment net is simple as follows:

$$L_A = \mathrm{MSE}(\hat{\mathbf{L}},\mathbf{L}_t) + \mathrm{MSE}(\nabla\hat{\mathbf{L}},\nabla\mathbf{L}_t), \quad (4)$$

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 11 shows the difference between our learned adjustment function and gamma correction. For comparison fairness, we tune the parameter $\gamma$ 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. Figure11a depicts the source illumination, (b) and (d) are 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. Regarding 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 consistent with actual situations. Furthermore, the $\alpha$ fashion is more convenient than the $\gamma$ way for users to manipulate. For instance, setting $\alpha$ to 2 means turning the light 2X up. Figure 12 displays two examples containing results by setting different light levels.

## 3 Experimental Validation

In this section, we first report the proposed network's implementation details. Then, we qualitatively and quantitatively compare our method with several state-of-the-art methods.

### 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 460 image pairs, and 240 synthetic pairs[4]are used. For all of three nets, the batch size is set to 10 and the patch-size to 48×48. We employ Adam as the optimizer. The network is trained on a Nvidia GTX 2080Ti GPU and an Intel Core i7-8700 3.20GHz CPU under the Tensorflow framework.

### 3.2 Performance Evaluation

We evaluate our method on widely-adopted datasets, including LOL (Wei et al. 2018), LIME (Guo et al. 2017), NPE (Wang et al. 2013), MEF (Ma et al. 2015), DICM (Lee et al. 2013), MIT-Adobe FiveK and SICE (Cai et al. 2018). Five metrics are adopted for quantitative comparison, which are PSNR, SSIM, LOE (Wang et al. 2013), NIQE (Mittal et al. 2013) and DeltaE (Sharma et al. 2005). A higher value in terms of PSNR and SSIM indicates better quality, while, in LOE, NIQE and DeltaE, the lower the better. The state-ofthe-art methods of BIMEF (Ying et al. 2017), SRIE (Fu et al. 2016), CRM (Ying et al. 2018), Dong (Dong et al. 2011), LIME (Guo et al. 2017), MF (Fu et al. 2016), RRM (Li et al. 2018), Retinex-Net (Wei et al. 2018), DUPE (Wang et al. 2019), DPE (Chen et al. 2018), GLAD (Wang et al. 2018), and NPE (Wang et al. 2013)[5] are involved as the competitors.

Table 1 reports the numerical results among the competitors on the LOL dataset. For each testing low-light image, there is a "normal"-light correspondence. Thus, the correspondence can be taken as reference to measure PSNR and

[4] In the previous version, KinD (Zhang et al. 2019) is trained without using any synthetic pairs. In this version, for comparison fairness, we retrain KinD by embracing the synthetic data, and report new results accordingly.

[5] All the codes are from the authors' websites.

SSIM.Fromthenumbers,weseethatbothKinDandKinD++ significantly outperform all the other methods. In terms of the non-reference metric NIQE and DeltaE, our KinDs also show their superiority over the others by a large margin. With the new design, KinD++ further steps forward in comparison with KinD. But, in LOE, both KinD++ and KinD seem falling behind many methods. From the definition of LOE (Wang et al. 2013), we can see that the reference is crucial to quantitatively measuring the quality of enhancement. As pointed out by Guo et al. (2017), using the low-light input itself to compute LOE is problematic. Because, take an extreme case for example, the LOE reaches the lowest value 0 when no enhancement is performed. To more appropriately reflect the enhancement quality in terms of LOE, a suitable reference matters. Similarly to computing PSNR andSSIM,weagainemploythecorrespondence imageasthe reference (denoted as $LOE_{ref}$). For the sake of completeness and objectiveness, both the LOE and $LOE_{ref}$ are reported. In $LOE_{ref}$, our KinDs come up to the 1st (KinD++, 776.2) and 3rd (KinD, 946.3) places, while CRM (926.1) occupies the 2nd place.

For the LIME, NPE, MEF, and DICM datasets, there are no reference images available. Thus, we adopt NIQE to evaluate the performance difference among the competitors. As reported in Table 2, our KinDs show their clear advantages over the others. Specifically, KinD++ outperforms all the competitors on the LIME (with # 6 image excluded), NPE and DICM datasets. DUPE takes over the MEF dataset. The reason why excludes # 6 image is that the test case is not a natural-like image as shown in Fig. 13, however the NIQE

**Table 1** Quantitative comparison on the LOL dataset in terms of PSNR, SSIM, LOE, $LOE_{ref}$, NIQE, and DeltaE

| Metrics | BIMEF (Ying et al. 2017) | CRM (Ying et al. 2018) | Dong (Dong et al. 2011) | LIME (Guo et al. 2017) | MF (Fu et al. 2016) | RRM(Lietal. 2018) | DUPE (Wang et al. 2019) |
|---|---|---|---|---|---|---|---|
| PSNR ↑ | 13.8753 | 17.2033 | 16.7165 | 16.7586 | 16.9662 | 13.8765 | 16.7975 |
| SSIM ↑ | 0.5771 | 0.6442 | 0.5824 | 0.5644 | 0.6422 | 0.6577 | 0.5187 |
| LOE ↓ | *250.6* | **30.9** | 740.5 | 817.2 | 1060.1 | 924.3 | 398.9 |
| $LOE_{ref}$ ↓ | 985.9 | *926.1* | 1391.5 | 1342.4 | 1042.1 | 958.7 | 986.1 |
| NIQE ↓ | 7.6992 | 8.0182 | 9.1358 | 9.1272 | 9.7125 | 5.9416 | 8.4736 |
| DeltaE | 21.2383 | 15.7743 | 15.6163 | 14.9474 | 15.5635 | 20.7342 | 19.5868 |
| Metrics | SRIE (Fu et al. 2016) | Retinex-Net (Wei et al. 2018) | DPE (Chen et al. 2018) | NPE (Wang et al. 2013) | GLAD (Wang et al. 2018) | KinD (Zhang et al. 2019) | KinD++ |
| PSNR ↑ | 11.8552 | 16.7740 | 13.1728 | 16.9697 | 19.7182 | *20.7261* | **21.3003** |
| SSIM ↑ | 0.4979 | 0.5594 | 0.4787 | 0.5894 | 0.7035 | *0.8103* | **0.8226** |
| LOE ↓ | 599.4 | 1712.6 | 1735.0 | 1071.2 | 714.9 | 1056.4 | 849.6 |
| $LOE_{ref}$ ↓ | 1199.8 | 2084.8 | 999.6 | 1643.1 | 1017.1 | 946.3 | **776.2** |
| NIQE ↓ | 7.5349 | 9.7289 | 4.4931 | 9.1352 | 6.7972 | *4.1352* | **3.8807** |
| DeltaE ↓ | 25.2829 | 15.8936 | 12.2534 | 15.3318 | 12.2776 | *9.8632* | **8.7425** |

The best results are highlighted in bold, and the second best are in italic

**Table 2** Quantitative comparison on the LIME, NPE, MEF, and DICM datasets in terms of NIQE

| Metric | NIQE ↓ | | | | |
|---|---|---|---|---|---|
| Datasets | LIME-data w/o #6 | LIME-data #6 only | NPE-data | MEF-data | DICM-data |
| BIMEF (Ying et al. 2017) | 3.1681 | *6.9667* | 3.4975 | *3.1543* | 3.2659 |
| CRM (Ying et al. 2018) | 3.2688 | 7.9222 | 3.6800 | 3.1899 | 3.3624 |
| Dong (Dong et al. 2011) | 3.5429 | 10.5789 | 3.8562 | 4.5499 | 4.3412 |
| LIME (Guo et al. 2017) | 3.5181 | 11.8101 | 3.8422 | 3.8765 | 3.6642 |
| MF (Fu et al. 2016) | 3.3048 | 11.2817 | 3.6800 | 3.4256 | 3.4533 |
| RRM (Li et al. 2018) | 3.4095 | 8.4095 | 3.9466 | 3.9385 | 3.3186 |
| SRIE (Fu et al. 2016) | *2.9980* | 7.7079 | *3.1788* | 3.2192 | 3.0951 |
| Ret-Net (Wei et al. 2018) | 3.7644 | 15.1991 | 4.0676 | 5.0047 | 4.7120 |
| DPE (Chen et al. 2018) | 3.2090 | 9.2415 | 3.7953 | 3.9301 | 3.6346 |
| NPE (Wang et al. 2013) | 3.1806 | 9.7748 | 3.4455 | 3.5884 | 3.4304 |
| GLAD (Wang et al. 2018) | 3.1125 | 11.0162 | 3.2026 | 3.1994 | 3.0846 |
| DUPE (Wang et al. 2019) | 3.1624 | **6.8566** | 3.3327 | **3.1025** | 3.1628 |
| KinD (Zhang et al. 2019) | 3.1231 | 9.3538 | 3.3668 | 3.3274 | *3.0124* |
| KinD++ | **2.9807** | 9.4144 | **3.1466** | 3.2116 | **2.8768** |

The best results are highlighted in bold, and the second best are in italic

model is trained based on natural images. This makes the numbers in NIQE ineffective to reflect its visual quality. Please see the results given in Fig. 13, our KinDs provide more visually clear and striking pictures than the other methods, while the methods like SRIE and DUPE barely enhance thelow-lightregions.But,inNIQEasreportedinTable2,the numerical results by SRIE (7.7079) and DUPE (6.8566) are smaller than those by KinD (9.3538) and KinD++ (9.4144).

In addition, Figs. 14, 15, 16 and 17 give visual comparisons on several challenging images. 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. Both of KinD and KinD++ offer visually striking results in these cases with the light properly adjusted and degradations clearly removed. By picking out more and finer details from degradations, *e.g.*, the wallpaper in Fig. 14 and the regions indicated by boxes in Figs. 15, 16 and 17, KinD++ further pushes forward the performance of low-light image enhancement. Due to limited space, more visual comparisons together with related resources can be found at https://github.com/zhangyhuaee/KinD_plus.

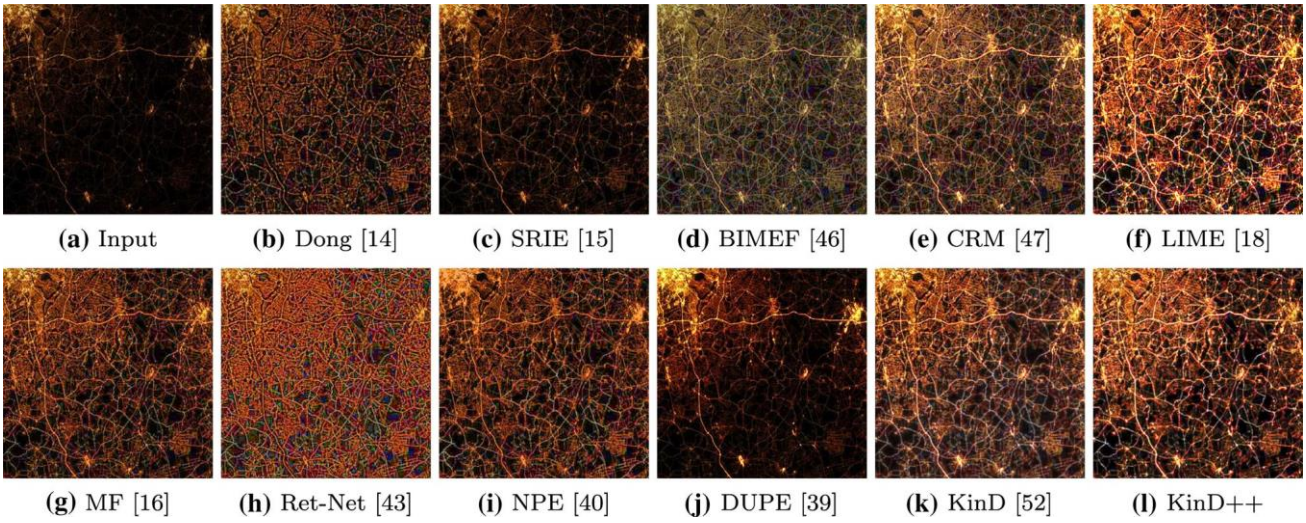



**Fig. 13** Visual comparison on # 6 image of the LIME dataset with state-of-the-art low-light image enhancement methods

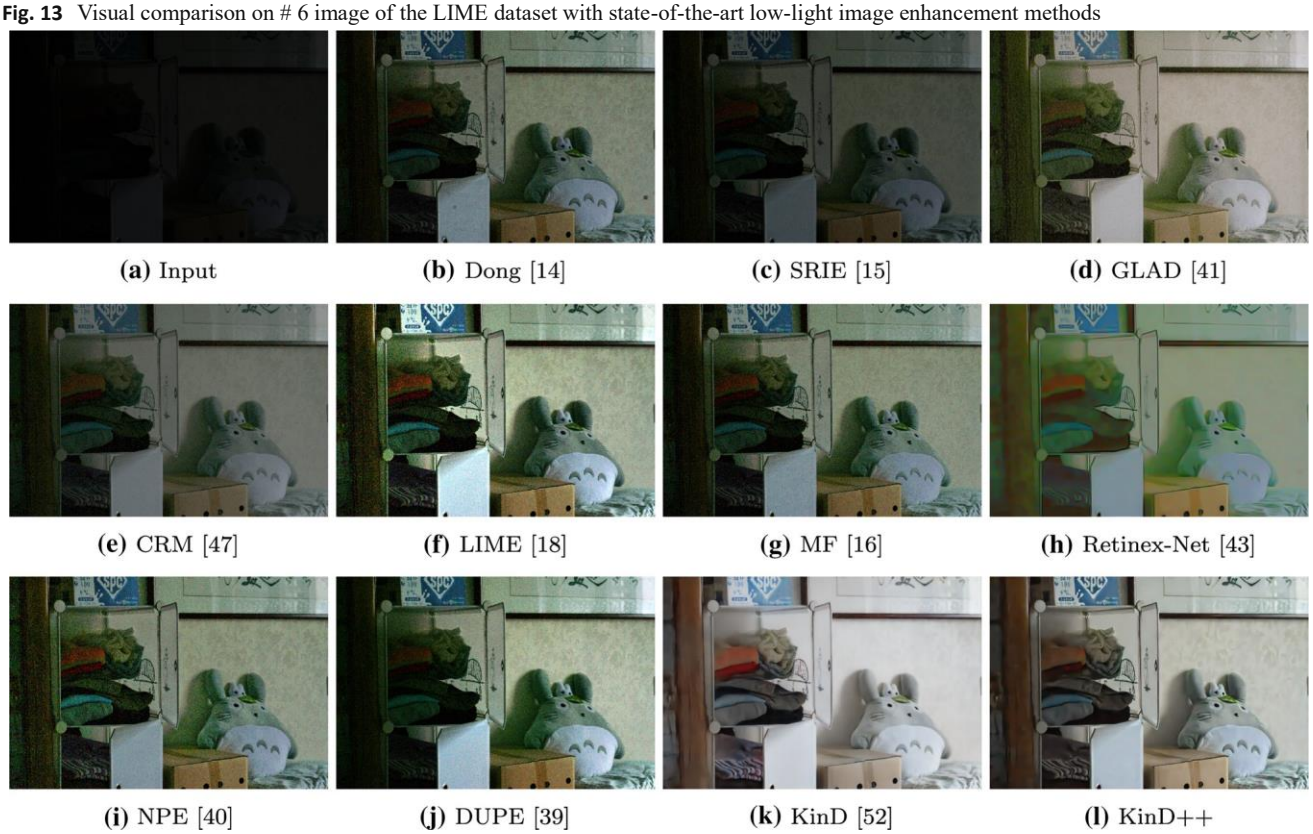



**Fig. 14** Visual comparison on an image from the LOL dataset with state-of-the-art methods

To make our evaluation more comprehensive, we have additionallyperformedapsychophysicalstatistic.Byfollowin g the Bradley-Terry method (Bradley and Terry 1952), the enhanced results of low-light images from the test datasets (LOL, DICM, LIME, NPE, and MEF) are conducted by different methods. For each pair of enhanced results, 50 human subjects are invited to independently vote for better ones


**Fig. 15** Visual comparison on an image from the DICM dataset with state-of-the-art methods


**Fig. 16** Visual comparison on an image from the DICM dataset with state-of-the-art methods

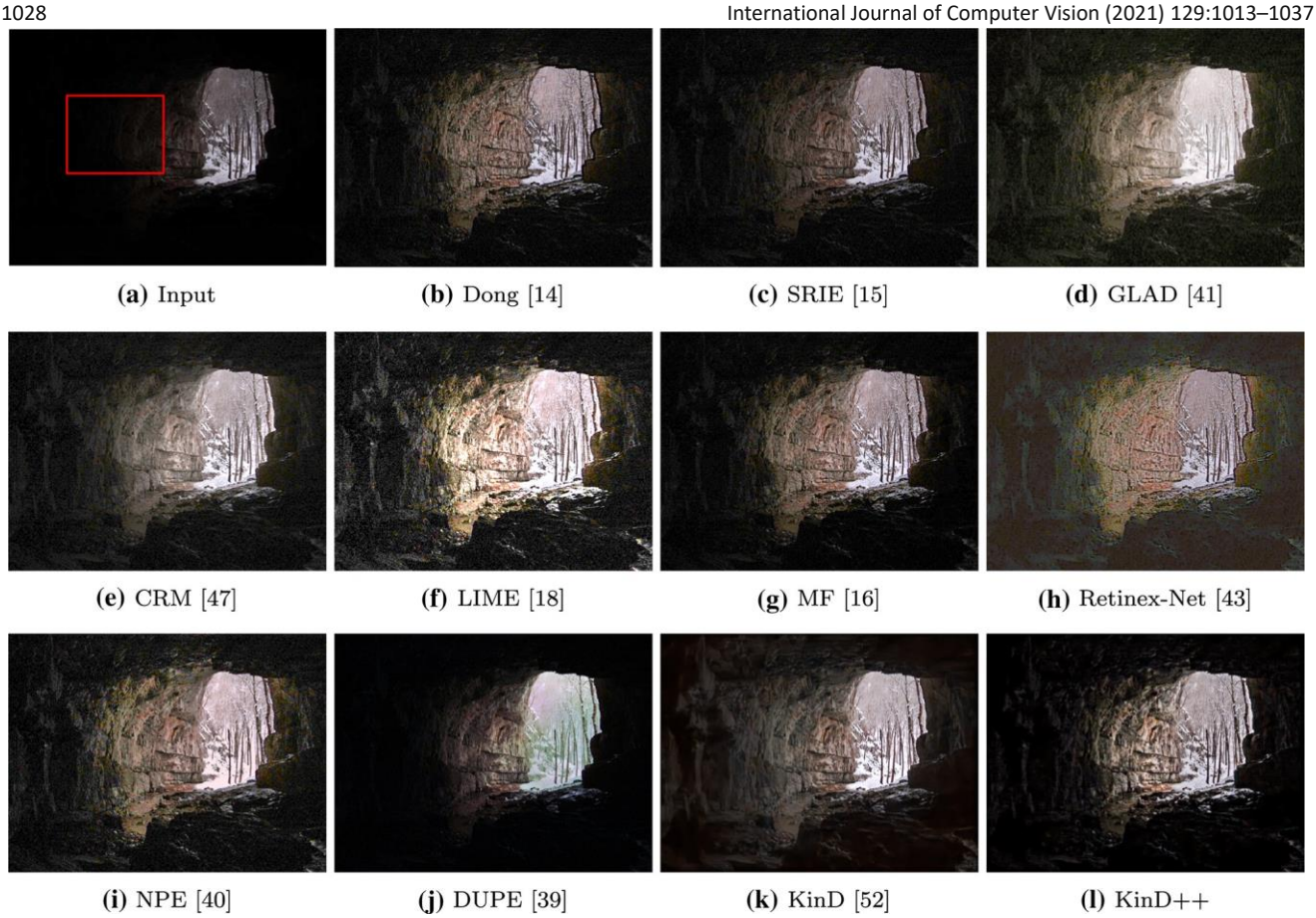



**Fig. 17** Visual comparison on an image from the MEF dataset with state-of-the-art methods

**Table 3** Psychophysical analysis of competing methods using the Bradley-Terry method (Bradley and Terry 1952)

| Method | Votes | Estimate | Std. Error | Z value | Pr(> \|z\|) | Rank |
|---|---|---|---|---|---|---|
| KinD++ | **714** | **0.64914** | 0.08016 | **8.098** | **5.61e-16** | **1** |
| CRM (Ying et al. 2018) | *669* | *0.50668* | 0.07968 | *6.359* | *2.04e-10* | *2* |
| KinD (Zhang et al. 2019) | 653 | 0.45656 | 0.07957 | 5.738 | 9.58e-09 | 3 |
| BIMEF (Ying et al. 2017) | 650 | 0.44718 | 0.07955 | 5.622 | 1.89e-08 | 4 |
| SRIE (Fu et al. 2016) | 647 | 0.43782 | 0.07953 | 5.505 | 3.69e-08 | 5 |
| DUPE (Wang et al. 2019) | 641 | 0.41909 | 0.07950 | 5.272 | 1.35e-07 | 6 |
| NPE (Wang et al. 2013) | 634 | 0.39104 | 0.07946 | 4.922 | 8.59e-07 | 7 |
| LIME (Guo et al. 2017) | 615 | 0.34436 | 0.07940 | 4.337 | 1.45e-05 | 8 |
| MF (Fu et al. 2016) | 611 | 0.32570 | *0.07939* | 4.102 | 4.09e-05 | 9 |
| GLAD (Wang et al. 2018) | 600 | 0.29149 | **0.07938** | 3.672 | 0.00024 | 10 |
| DPE (Chen et al. 2018) | 575 | 0.21369 | *0.07939* | 2.692 | 0.00711 | 11 |
| RRM (Li et al. 2018) | 507 | 0.00000 | – | – | – | 12 |
| Dong (Dong et al. 2011) | 478 | −0.09294 | 0.08008 | −1.161 | 0.24582 | 13 |
| Retinex-Net (Wei et al. 2018) | 469 | −0.12210 | 0.08020 | −1.522 | 0.12791 | 14 |

The best results are highlighted in bold, and the second best are in italic

**Table 4** Quantitative comparison on the FiveK dataset in terms of PSNR, SSIM, LOE, $LOE_{ref}$, NIQE and DelatE

| Metrics | BIMEF (Ying et al. 2017) | CRM (Ying et al. 2018) | Dong (Dong et al. 2011) | LIME (Guo et al. 2017) | MF (Fu et al. 2016) | NPE (Wang et al. 2013) | SRIE (Fu et al. 2016) | RRM(Lietal. 2018) | DUPE (Wang et al. 2019) |
|---|---|---|---|---|---|---|---|---|---|
| PSNR ↑ | 18.6679 | 13.7113 | 14.3570 | 11.2089 | 18.0777 | 18.9656 | 18.3009 | 13.9996 | 18.6190 |
| SSIM ↑ | 0.7736 | 0.6958 | 0.6634 | 0.6124 | 0.7501 | 0.7623 | 0.7888 | 0.6659 | 0.7376 |
| LOE ↓ | *305.2* | 782.1 | 1278.5 | 1342.4 | 721.3 | 593.7 | 548.6 | 1100.4 | 671.4 |
| $LOE_{ref}$ ↓ | *473.4* | 903.7 | 1359.1 | 1429.8 | 832.2 | 724.6 | 668.9 | 1198.3 | 796.8 |
| NIQE ↓ | **3.4242** | 3.5033 | 4.5046 | 3.9786 | 3.5868 | 3.5427 | *3.4640* | 4.2151 | 3.4653 |
| DeltaE ↓ | 11.4155 | 18.0197 | 16.7420 | 22.6701 | 12.0312 | 11.0845 | 10.6397 | 18.0770 | 15.5184 |

| Metrics | DPE (Chen et al. 2018) | Ret-Net (Wei et al. 2018) | GLAD (Wang et al. 2018) | KinD (Zhang et al. 2019) | KinD++ et al. | Ret-Net(R) (Wei et al. 2018) | GLAD(R) (Wang et al. 2018) | KinD(R) (Zhang et al. 2019) | KinD++(R) |
|---|---|---|---|---|---|---|---|---|---|
| PSNR ↑ | 19.9978 | 13.5068 | 19.1287 | 15.7489 | 20.4966 | 20.8069 | *21.1232* | 20.6758 | **21.9916** |
| SSIM ↑ | 0.7677 | 0.6515 | 0.7483 | 0.7198 | 0.7357 | 0.7658 | 0.7579 | *0.7970* | **0.8010** |
| LOE ↓ | **298.6** | 1661.4 | 451.2 | 720.1 | 659.2 | 461.9 | 525.7 | 957.6 | 562.3 |
| $LOE_{ref}$ ↓ | **426.2** | 1739.7 | 556.3 | 843.5 | 747.6 | 577.4 | 601.3 | 1015.5 | 641.2 |
| NIQE ↓ | 3.4663 | 4.4043 | 3.4679 | 3.6308 | 3.7934 | 5.1699 | 3.9426 | 3.6569 | 3.5117 |
| DeltaE ↓ | 9.8580 | 18.0197 | 11.4813 | 14.7590 | 10.1310 | 10.6496 | 11.2092 | *9.2204* | **8.6326** |

The retrained versions of related methods on the FiveK are marked by (R). The best results are highlighted in bold, and the second best are in italic

mainly based on several factors, including the degree of exposure, the naturalness, and the level of color deviation and noise. The winner and loser score 1 and 0, respectively. Then the averaged votes from 50 human subjects are analyzed by the standard Bradley-Terry method (Bradley and Terry 1952). The psychophysical statistic over various methods is listed in Table 3, without loss of generality, by setting theRRMasthebenchmark[6].AsshowninTable3,ourresults are most favored, followed by CRM and KinD. The methods like Dong and Retinex-Net obtain poor values and rankings because their results contain obvious over-smoothing, unnatural texture, and/or hue shift issues.

To demonstrate the generalization ability, we have further tested the proposed method on the MIT-Adobe FiveK dataset (Bychkovsky et al. 2011). We first enhance 496 input images(exceptfor4imagesduetoinconsistentsizesbetween the input and reference images) in the FiveK dataset (Pic# 4501-5000) with various methods. For the deep learning methods, including Retinex-Net (Wei et al. 2018), GLAD (Wang et al. 2018), KinD (Zhang et al. 2019) and KinD++, wehaveretrainedtheirmodelsbasedonthe4500pairimages from the FiveK (Pic# 0001-4500). We do not retrain the DPE (Chen et al. 2018) and DUPE (Wang et al. 2019) since they have already been trained on this dataset. Table 4 reports the quantitativeresultsontheFiveKdataset.AsshowninTable4, our KinD++ obtains the best PSNR and the $2^{nd}$ best DeltaE values among all the methods without retraining, which verifies the generalization ability of our method. After retraining on the FiveK dataset, the performance of four methods have all been improved on this dataset. Our KinD++(R) reaches the 1st place in PSNR, SSIM and DeltaE. From the results in Fig. 18, we can observe that the results by DPE and our KinDs are very close to the reference, while the results by the other methods suffer from (severe) color deviation.

To further demonstrate the illumination map adjustment ability, we also conduce a comparison of our enhancement with actual multi-exposure results from a real camera

[6] The rankings and votes will not change by setting any competitor as the benchmark.

pipeline. The multi-exposure images are from the SICE dataset (Cai et al. 2018), which contains 589 high-resolution multi-exposure sequences with 4413 images in total. Due to limited space, we only provide several comparisons in Fig. 19. Our results have similar visual perception with low exposure images. While for the high exposure images, the real images exist over-enhanced and color fading appearances in high light regions. Our method can enhance the low-light regions gradually and keep the normal light regions less over-enhanced. In addition, we conduct a user study to see the performance difference of our model to other commercial softwares that allow exposure/image enhancement, including Lightroom, Photoshop, iPhone XR and Samsung S20. The testing data includes 93 low-light images from the LOL, DICM, LIME, NPE, and MEF datasets, and extra 50 low-light images from a public dataset, *i.e.* SICE dataset (Cai et al. 2018). For each testing image and each software, a middle-level and a high-level enhanced results are produced. To make sure the comparison fair enough, different enhanced results are aligned to be approximate in bright-

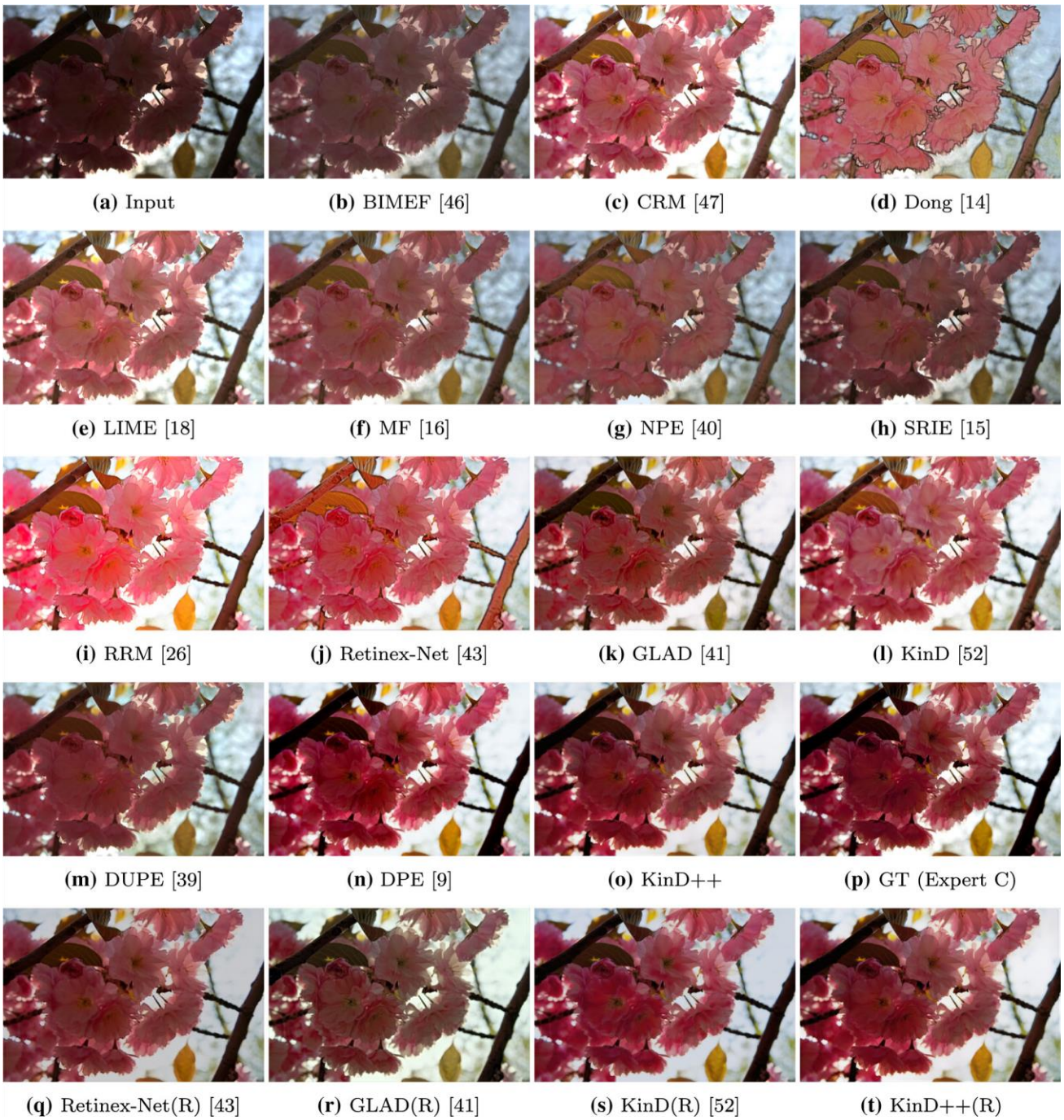



**Fig. 18** Visual comparison on a sample from the FiveK dataset

ness by adjusting the illumination. There are 50 participants invited to perform pairwise visual comparisons between our results and one of the competitors. Each participant gives an option from "A better", "B better", or "no preference". Figure 20 shows the statistics of user study. As can be seen, our results exhibit overwhelming superiority over those by the commercial tools. Some visual comparisons are given in Fig. 21.

## 4 Ablation Study

The performance of networks depends on both the architecture and the loss. This section evaluates the effectiveness of different architectures and multiple loss functions on the layer decomposition and the reflectance restoration subnets. Although the light adjustment is a critical function for users toflexiblymanipulateimages,anilluminationadjustmentnet

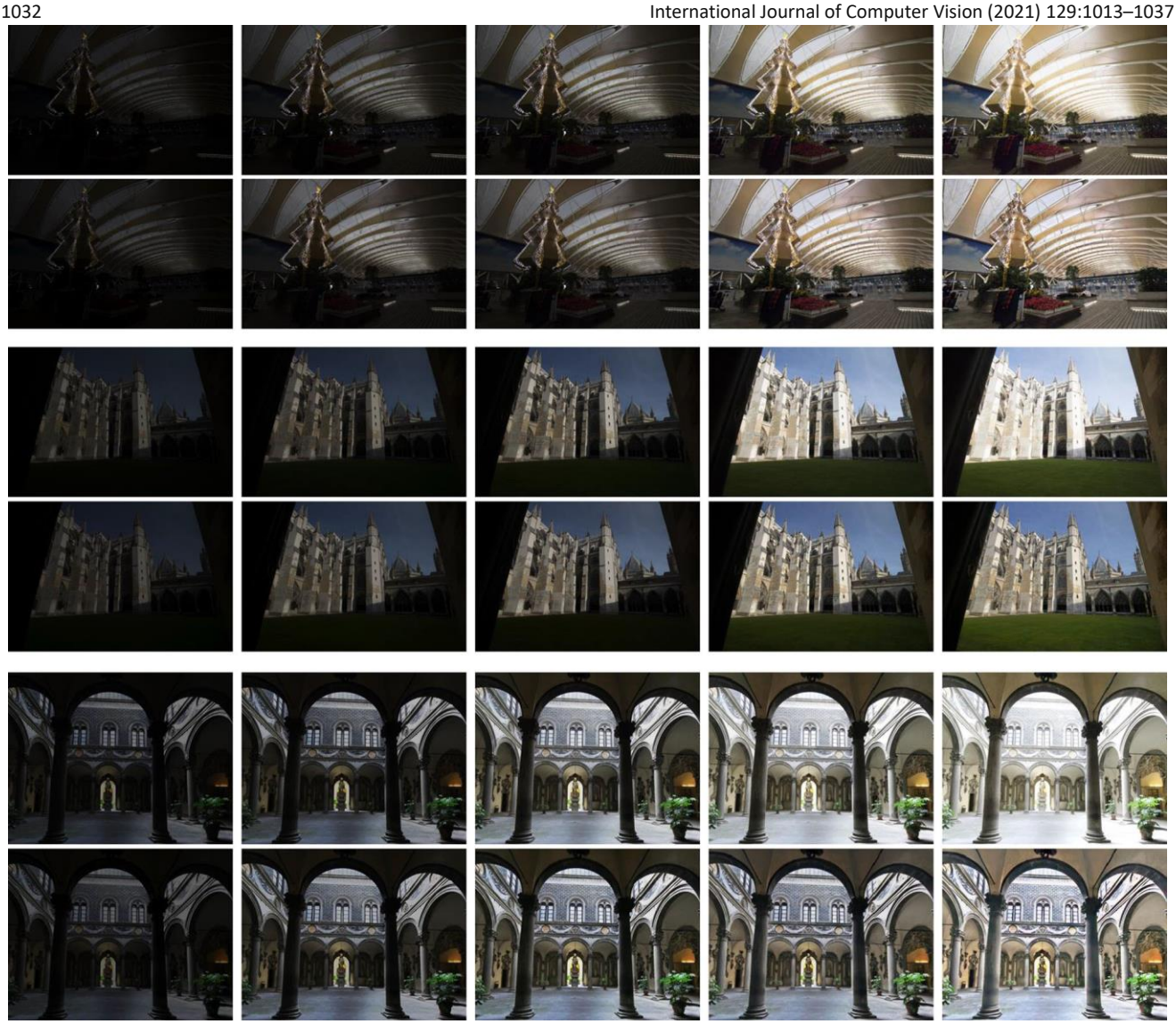


**Fig. 19** Visual comparisons with real multi-exposure images. The odd rows are the real multi-exposure images with different exposures. The even rows are our results with corresponding different adjustment ratios. The images in odd rows and the third columns are the input of our network

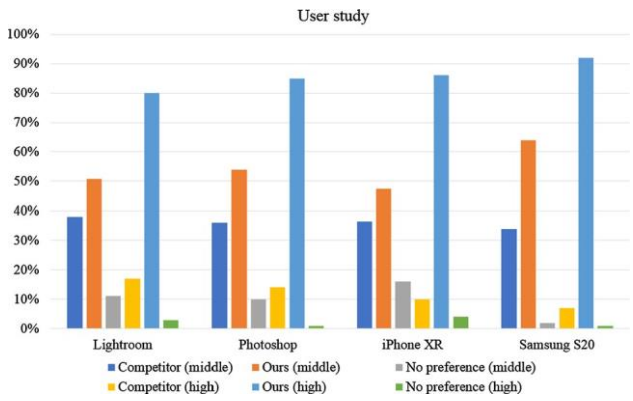



**Fig. 20** Subjective preference of KinD++ versus four commercial tools

with simple layers and loss terms, as used in this work, can achieve the goal with reasonably well results, so no further ablationanalysisontheadjustmentsubnetisgiven.Insteadof evaluating each architecture-loss combination, our strategy is to test one factor with the other fixed.

## 4.1 On the Layer Decomposition Net

### 4.1.1 Network Architecture

Equipped with the complete loss function $L^D$, three more net architectures for layer decomposition as shown in Fig. 22 together with the one adopted given in Figure 2 are tested.

Figure 22a displays a network. Its first 5 layers are operated by Conv+ReLU, followed by a Conv layer and a Sigmoid

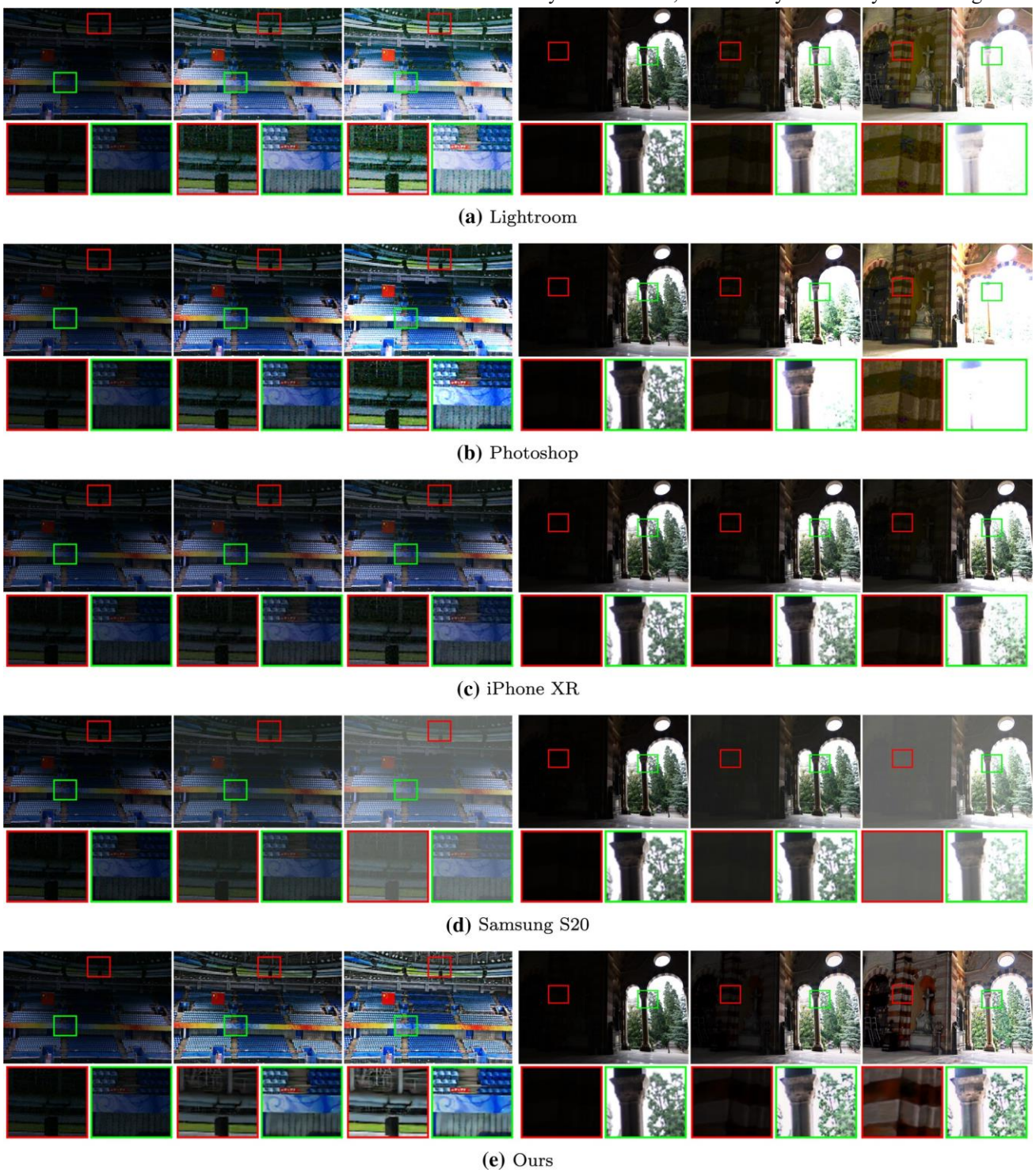



**Fig.21** Visualcomparisonswithcommercialsoftwares.Thefirstandfourthcolumnsareoriginalinputs,thesecondandfifthonesareofmiddle-level enhancement, while the third and sixth ones are high-level ones

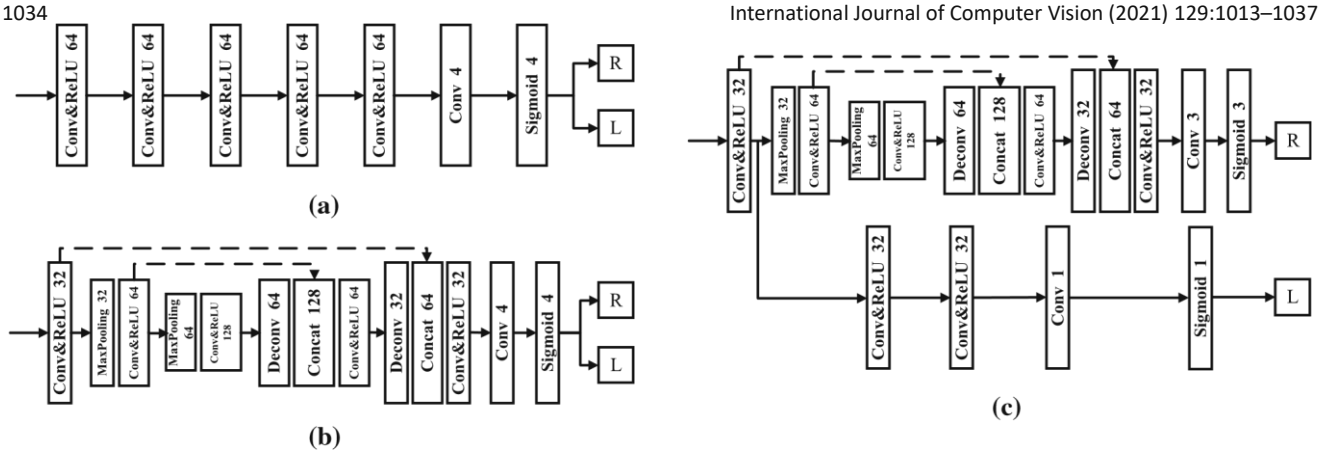



**Fig. 22** Different candidate architectures of the layer decomposition architecture while the illumination one is all-Conv. The letters 'R' and network to our proposed one as shown in Fig. 2. **a** an all-Conv archi- 'L' stand for the reflectance and illumination maps, respectively tecture, **b** a U-net architecture, and **c** the reflectance branch is a U-net

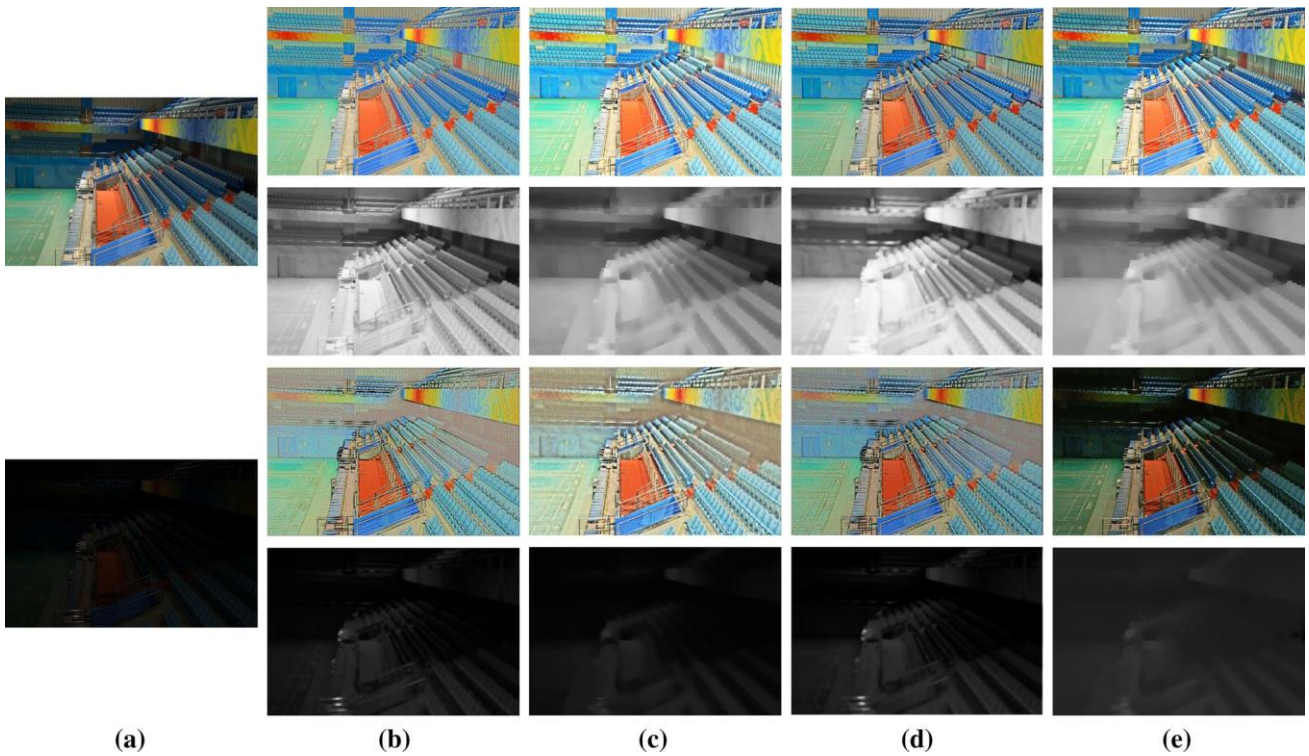



**Fig.23** Thedecompositionresultsofdifferentlayerdecompositionnetworkarchitectures.Exceptinputs**a**,theremainingcolorimagesarereflectance maps, while the gray images are illuminations. **b**–**d** depict the decomposed results by the networks given in Fig. 22a–c, while **e** corresponds to KinD(++)

layer. The kernel size of each Conv layer is 3x3. While (b) is a 5-layer U-net, and (c) is comprised of a U-net architecture for reflectance and a Conv branch for illumination. The layer decomposition net adopted by both KinD and KinD++ differs from Fig. 22c by only adding one connection from the reflectance branch to the illumination one. For fair comparison, we adjust experimentally the best possible parameters for each candidate network. Due to no references available for quantitatively measuring the performance, we showvisualresultsofdifferentnetarchitecturesinFig.23.As canbeseen,althoughthereflectancemapsofthenormal-light image (the 1st row) of four networks are slightly different, the illumination maps (the 2nd and 4th rows) provide stronger evidence. The illumination maps in (b) and (d) are insufficiently smooth. The reflectance maps of the low-light image (the 3rd row) show that the adopted network gets better results. Compared with the architecture in Fig. 7c, the (only) connection between the reflection and illumination branches exercises the exclusivity on the texture between the two decomposed components.

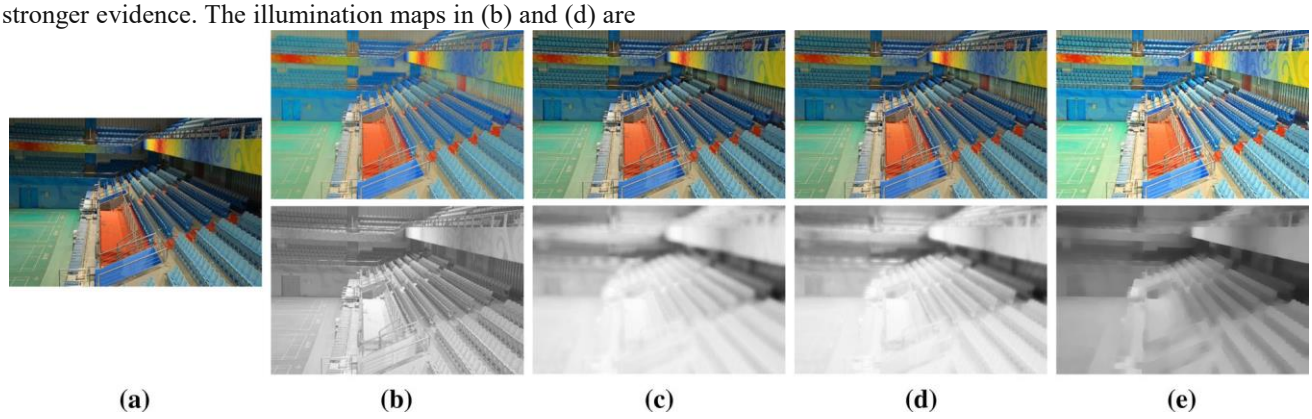



**Fig.24** Decompositionresultswithdifferentlossfunctions.**a**isthesameinputastheupperinFig.23a.**b**–**e**exhibittheresultsbyusing$L_{reD}+0.009L_{rsD}$, $L_{reD}+0.009L_{rsD}+0.2L_{mcD}|_{c=5}$, $L_{reD}+0.009L_{rsD}+0.2L_{mcD}|_{c=10}$, and complete $L_{reD}+0.009L_{rsD}+0.2L_{mcD}|_{c=10}+0.15L_{isD}$, respectively

#### 4.1.2 Loss Function

This part verifies the effectiveness of each loss term in $L^D$ with the net architecture fixed as given in Fig. 2. Because the illumination maps of low-light images are too dark, for better view, we show the results of normal-light images in Fig. 24. As shown in Fig. 24b, merely employing the reconstruction andreflectancesimilaritytermssuffersfromtheambiguityof decomposition, leading to the piece-wise smooth reflectance map and detailed illumination map, which is contrary to our expectation. With the mutual consistency joined, the decomposition ambiguity is greatly mitigated. However, the structure of the illumination map cannot effectively get rid of over-smoothing. Figure 24c and d reveal the behavior of the parameter $c$. For this paper, we empirically set $c = 10$. By further introducing the illumination smoothness constraint, the structure of the illumination map becomes sharper, and thus the reflectance map is more informative (Fig. 24e).

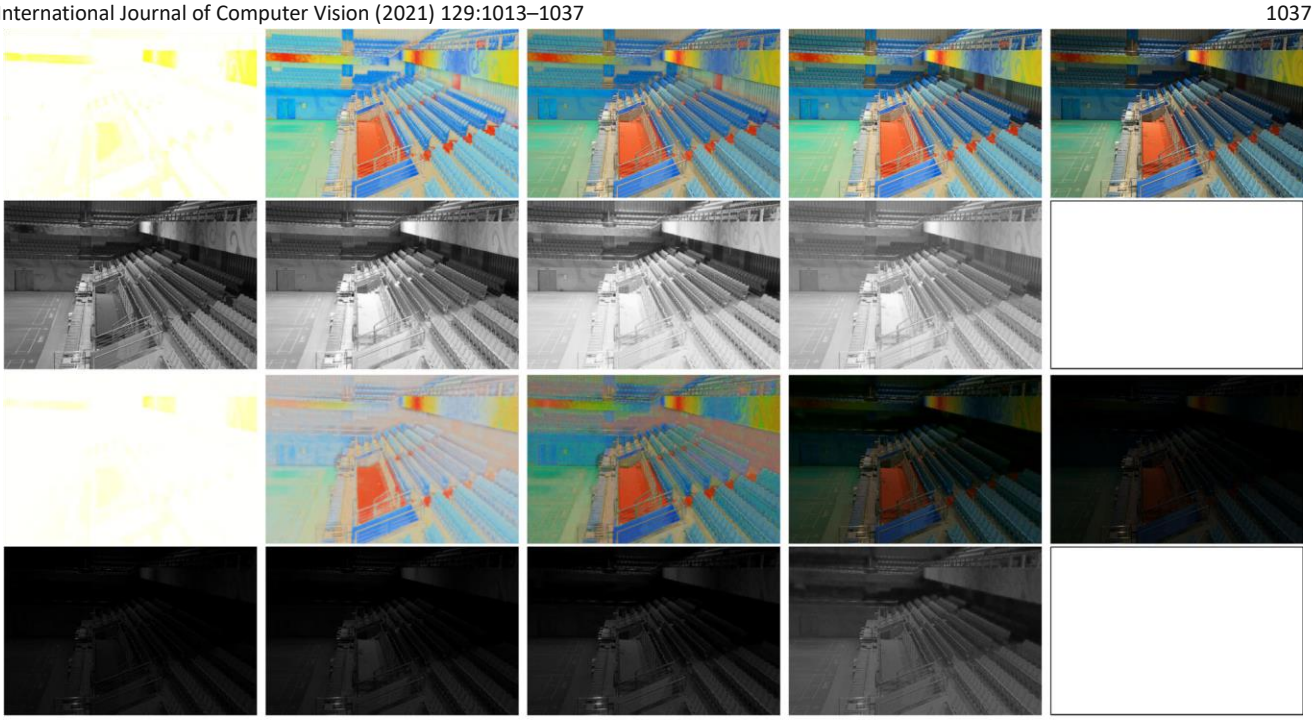
In what follows, we show more about the selection principle of involved weights in the layer decomposition loss function,*i.e.* $\{\omega_{rs},\omega_{mc},\omega_{is}\}$ in $L_{rec}^{LD}+\omega_{rs}\mathcal{L}_{rs}^{LD}+\omega_{mc}\mathcal{L}_{mc}^{LD}+\omega_{is}L_{is}^{LD}$. Due to the complex dependence of these weights, wetesttheminaprogressiveway.Onemaythinkthatenforcing $\mathbf{R}_l$ and $\mathbf{R}_h$ could produce satisfactory reflectances. This claim holds if the images are degradation-free. However, it is not the case in practice due to the existence of defects. As showninFig.25a,whenonly$L_{rec}^{LD}$ and$L_{rs}^{LD}$ termsareconsidered, the 1 : 1 weights generate a poor decomposition result. As decreasing the weight $\omega_{rs}$ of $L_{rs}^{LD}$, the results get better till being around 0.01 (Fig. 25b and c). This is because $\mathbf{R}_l$ usually contains amplified noise and color distortion [please refer to $\tilde{\mathbf{E}}$ in Eq. (1)]. The amplification leads to a smaller weight for balancing the two terms. If enforcing $\mathbf{R}_l$ and $\mathbf{R}_h$ to be strongly similar or same, most information including the degradation and intrinsic reflectance details would go to the other component, *i.e.* the illumination, to meet the reconstruction requirement. If we continue decreasing the weight $\omega_{rs}$, the ability of $L_{rs}^{LD}$ becomes trivial, as shown in Fig. 25d. Merely considering

$L_{rec}^{LD}$ and $L_{rs}^{LD}$ cannot decompose the layers sufficiently well. As can be seen from Fig. 25a–d, the illumination map contains rich textures. We further add the $L_{mc}^{LD}$ term for imposing the piece-wise smoothness on illumination maps. We fix $\omega_{rs}$ to 0.01 and test the effect of $L_{mc}^{LD}$. The visual results are shown in Fig. 25e–g with varying $\omega_{mc}$, from which it is easy to tell that the smoothing extent is proportional to the value of $\omega_{mc}$. Similarly, we repeat the above procedure to further verify the efficacy of $L_{is}^{LD}$ by fixing $\omega_{rs}$ and $\omega_{mc}$ to 0.01 and 0.1, respectively. Figure 25h–j depict the results corresponding to $\omega_{mc}$ = 1,0.1,0.01, respectively, from which we are able to obtain a similar conclusion with $\omega_{mc}$, *i.e.* the larger the value of $\omega_{mc}$ is, the stronger the smoothing effect appears. We here notice that, both $L_{mc}^{LD}$ and $L_{is}^{LD}$ are introduced to reduce textures in illumination maps, but from two different ways. Concretely, $L_{mc}^{LD}$ desires toextractmutualstructurefromtwoilluminationmaps,while $L_{is}^{LD}$ expects to enforce each individual illumination map to be piece-wise smooth.

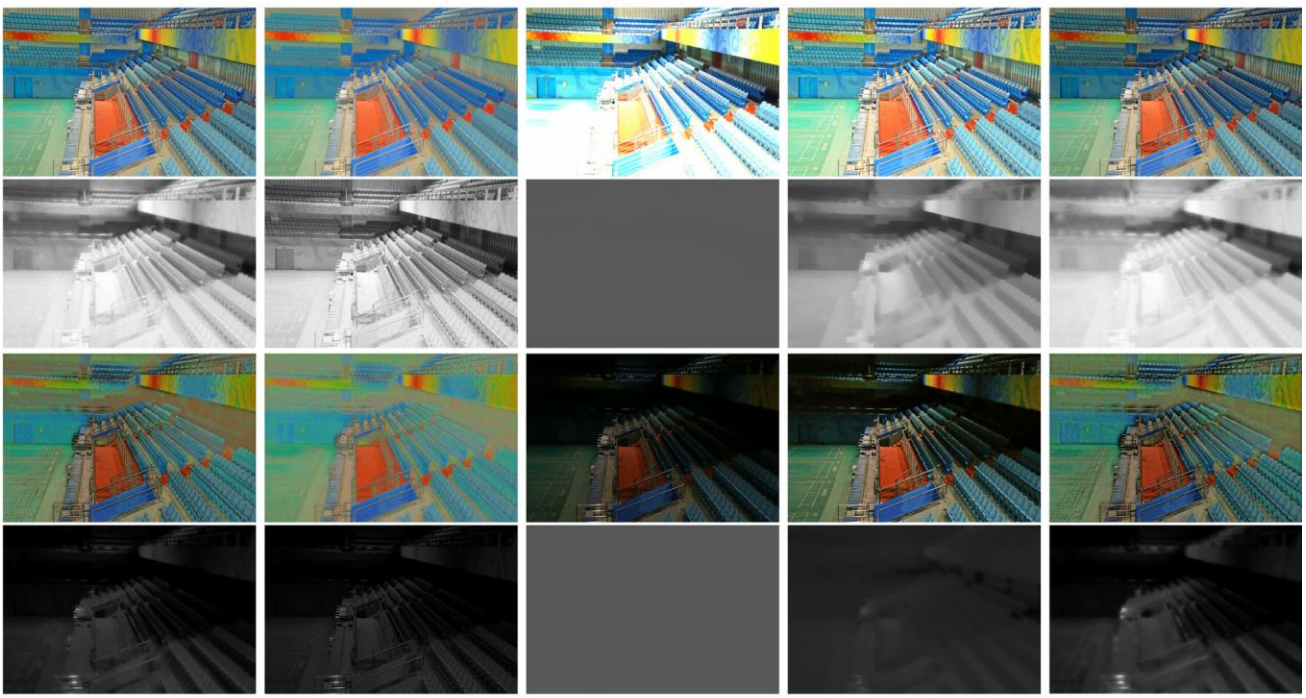
## 4.2 On the Reflectance Restoration Net

### 4.2.1 Network Architecture

This part compares five manners, including a traditional denoising tool BM3D (Dabov et al. 2007), a deep method FFDNet (Zhang et al. 2018) that can handle spatially variant degradations, KinD (Fig. 7b) (Zhang et al. 2019), KinD++ (Fig. 7a), and KinD++ with MSIA disabled, for reflectance restoration. The BM3D is one of the most classic and representativedenoisingmethods,whichinnatureisapatch-group



**(a)** [1, 0, 0] **(b)** [0.1, 0, 0] **(c)** [0.01, 0, 0] **(d)** [0.001, 0, 0] **(e)** [0.01, 1, 0]



**(f)** [0.01, 0.1, 0] **(g)** [0.01, 0.01, 0] **(h)** [0.01, 0.1, 1] **(i)** [0.01, 0.1, 0.1] **(j)** [0.01, 0.1, 0.01]

**Fig. 25** Visual results by different weights of loss terms. We fix the weight of $L_{rec}^{LD}$ as 1 and denote the other three weights as [$\omega_{rs}$, $\omega_{mc}$, $\omega_{is}$]

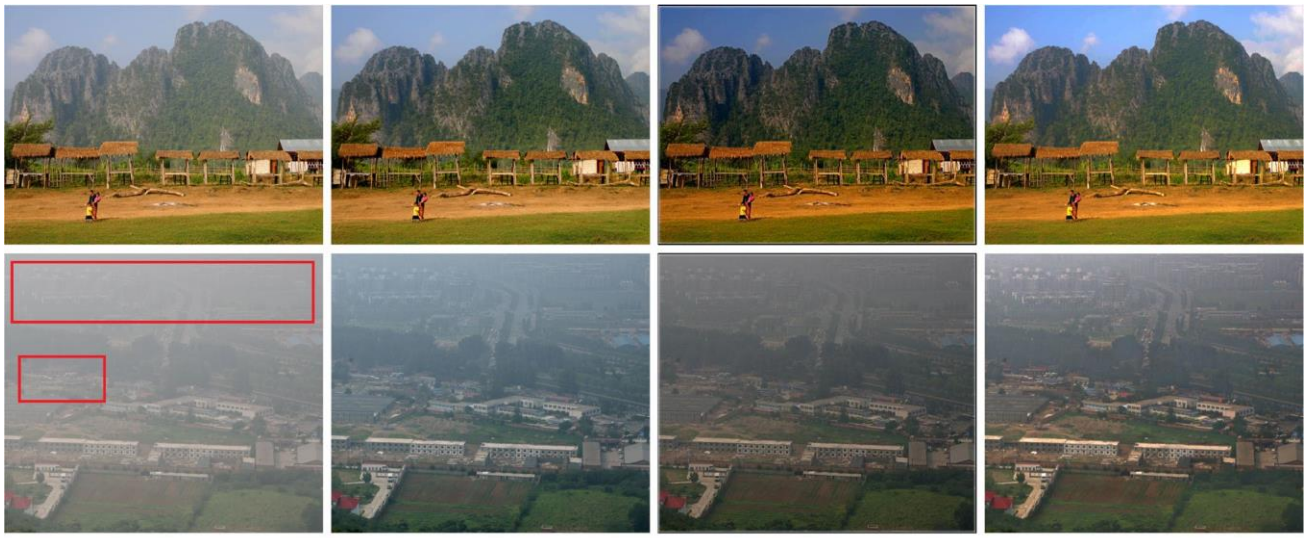
**(a)** BM3D [11] **(b)** FFDNet [49] **(c)** KinD [52] **(d)** w/o MSIA **(e)** KinD++

**Fig. 26** Reflectance restoration results using different architectures. The input is the 3rd picture of Fig. 23e

**Table 5** Quantitative comparison of the reflectance restoration net with different loss functions and net architectures on the LOL dataset in terms of PSNR, SSIM and DeltaE

| Metrics | BM3D | FFDNet | KinD | w/o MSIA | KinD++ | $^1$ loss | $^2$ (MSE) loss | $L^R$ loss |
|---|---|---|---|---|---|---|---|---|
| PSNR ↑ | 16.5076 | 19.4972 | *20.7261* | 20.1253 | **21.3003** | 19.4772 | 19.8645 | **21.3003** |
| SSIM ↑ | 0.6217 | 0.7749 | *0.8103* | 0.7943 | **0.8226** | 0.7647 | 0.7951 | **0.8226** |
| DeltaE ↓ | 13.2618 | 11.4758 | *9.8632* | 10.2477 | **8.7425** | 11.7846 | 11.4437 | **8.7425** |

The best results are highlighted in bold, and the second best are in italic



**(a)** Input **(b)** DehazeNet [5] **(c)** AODNet [25] **(d)** Ours

**Fig. 27** Visual comparison with state-of-the-art image dehazing methods

collaborative filtering strategy without training involved. Its performance guarantee comes from the self-similarity based on an enhanced sparse representation in transform domain. Due to its nature, the BM3D hardly deals with spatiallyvariant noise and color distortion. The FFDNet is originally designedforsolvingspatially-variantnoise,whichasksusers to input noise level maps as guidance. To increase the robustness against other types of degradations and relieve the requirement of manual intervention, we retrain the FFDNet, in the same way as our KinDs do, on reflectance pairs, and use illumination maps as the indicator. As can be observed from Table 5, FFDNet improves BM3D by a large margin in terms of PSNR and SSIM. KinD outperforms FFDNet by about 1dB in PSNR, 0.04 in SSIM, and 1.6 in DeltaE.

The new design of KinD++ proves its effectiveness by the best results in this competition. Please notice that, the

performance of KinD++ with MSIA disarmed heavily drops, which again confirms the effectiveness of MSIA. Figure 26 depicts a visual comparison, which corroborates the numerical result. The result by KinD++ contains clearer details and more vivid colors than those by the others.

#### 4.2.2 Loss Function

Having the complete KinD++ chosen, we now validate the effectiveness of our loss design. In this part, thanks to the paired data, the restoration quality is (pseudo-)referenced. Thus, the loss terms are simple. As reported in Table 5, the results by using the $^1$ loss only are numerically close but slightly behind those using the $^2$ loss. Via combining the $^2$ (MSE) and the structural dissimilarity (similarity) terms, the performance is greatly boosted.

## 5 Conclusion and Discussion

In this work, we have proposed a deep network for low-light enhancement. Inspired by the Retinex theory, the proposed network decomposes images into the reflectance and illumination layers. Following the divide-and-conquer principle,
thedecompositionconsequentlydecouplestheoriginalspace
into two smaller subspaces. As ground-truth reflectance and illumination information is 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 scheme builds a restoration module. A mapping function has also been learned, which better fits the actual situations than the traditional gamma correction, and flexibly adjusts light levels. The extensive experiments demonstrate the clear advantages of our design over the state-of-the-arts.

The proposed KinD(++) can also be applied to the dehazing problem. To verify this, we show a visual comparison in Fig. 27. As can be seen in this case, without

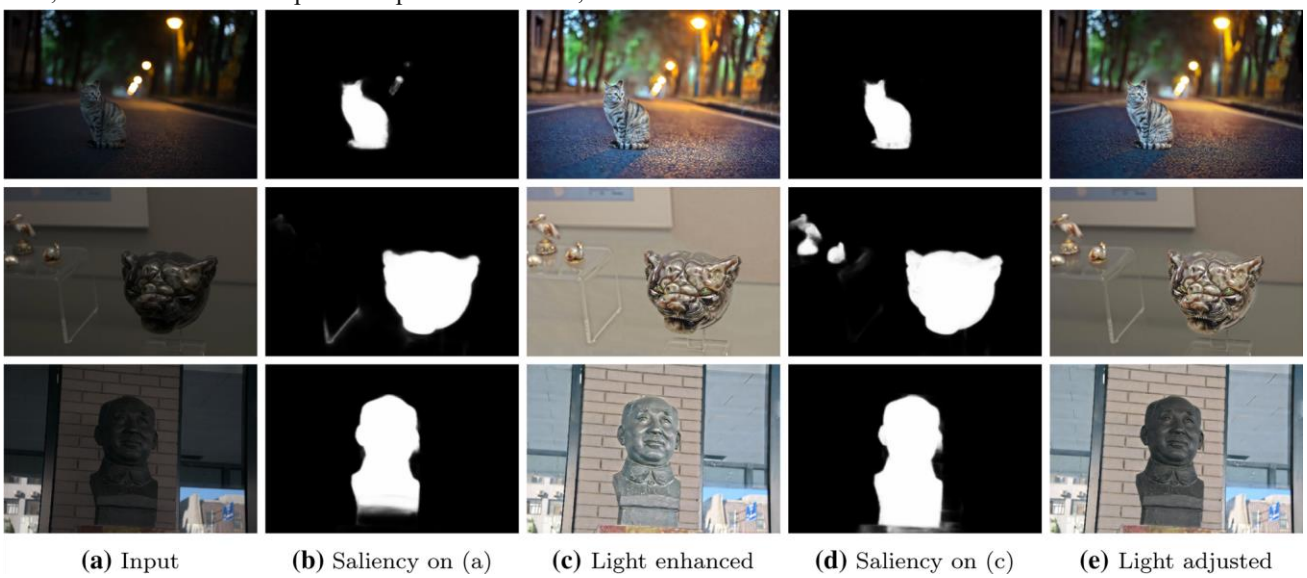



**Fig. 28** Flexible illumination adjustment for region of interest. **c** shows the overall enhanced results by our method on **(a)**. **b** and **d** display the salient regions detected by Zhang et al. (2018) on **a** and **c**, respectively. **e** gives the results differently adjusted with respect to salient or not salient regions


**Fig. 29** A failure case. Although the methods all poorly enhance extremely dark regions, our KinD++ possesses more details of higher contrast and looks more natural than the others

any modification on the net architecture, our method can do the job, with competitive or even better visual quality than two methods specific to dehazing (Cai et al. 2016) and (Li et al. 2017). The flexibility of light manipulation should be further promoted. Onemaydesiretoprocessdifferentregions/objects withinan image by different light operations. Figure 28 exhibits such examples that use saliency detection to distinguish regions. Other manners like semantic/instance segmentation are also optional. Besides, we observe from Fig. 28 that, under different light conditions, the saliency regions are changed. This phenomenon is related to the task of visual attention retargeting (Mateescu and Bajic 2016; Mechrez et al. 2018). Developing a light-invariant saliency detection, *i.e.* robust visual attention, may be an interesting idea. The limitation of our method, also of all the others, is the poor ability of enhancing extremely dark regions, because the information is almost lost. A failure case is provided in Fig. 29. A possible way to mitigate this issue is to employ a GAN thought for inpainting/generating some reasonable details.

**Acknowledgements** This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61772512 and 62072327, and in part by the National Key Research and Development Program of China under Grant No. 2019YFC1521200.

## References

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

Agostinelli, F., Anderson, M.R., & Lee, H. (2013). Adaptive multicolumn deep neural networks with application to robust image denoising. in: *NeurIPS*, pp. 1493–1501.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I, the method of paired comparisons. *Biometrika*, *39*, 324.

Bychkovsky, V., Paris, S., Chan, E., & Durand, F. (2011). Learning photographic global tonal adjustment with a database of input / output image pairs. in: *CVPR*, pp. 97–104.

Cai, B., Xu, X., Jia, K., Qing, C., & Tao, D. (2016). DehazeNet: An endto-end system for single image haze removal. *IEEE TIP*, *25*(11), 5187–5198.

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

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

Chen, Y., & Pock, T. (2017). Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE TPAMI*, *39*(6), 1256–1272.

Chen, Y., Wang, Y., Kao, M., & Chuang, Y. (2018). Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs. in: *CVPR*, pp. 6306–6314.

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. (2007). Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE TIP*, *16*(8), 2080–2095.

Dong, C., Chen, C. L., He, K., & Tang, X. (2016). Image superresolutionusingdeepconvolutionalnetworks.*IEEETPAMI*,*38*(2), 295–307.

Dong,C.,Deng,Y.,Loy,C.C.,&Tang,X.(2015).Compressionartifacts reductionbyadeepconvolutionalnetwork.in:*ICCV*,pp.576–584.

Dong, X., Pang, Y., & Wen, J. (2011). Fast efficient algorithm for enhancement of low lighting video. in: *ICME*, pp. 1–6.

Fu, X., Zeng, D., Huang, Y., Zhang, X., & Ding, X. (2016). A weighted variational model for simultaneous reflectance and illumination estimation. in: *CVPR*, pp. 2782–2790.

Fu, X., Zeng, D., Yue, H., Liao, Y., Ding, X., & Paisley, J. (2016). A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, *129*, 82–96.

Gu, S., Zhang, L., Zuo, W., & Feng, X. (2014). Weighted nuclear norm minimization with application to image denoising. in: *CVPR*, pp. 2862–2869.

Guo, X., Li, Y., & Ling, H. (2017). LIME: Low-light image enhancement via illumination map estimation. *IEEE TIP*, *26*(2), 982–993.

Huang,S.,Cheng,F.,&Chiu,Y.(2013).Efficientcontrastenhancement using adaptive gamma correction with weighting distribution. *IEEE TIP*, *22*(3), 1032–1041.

Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., & Van Gool, L. (2018). WESPE: Weakly supervised photo enhancer for digital cameras. in: *CVPRW*, pp. 691–700.

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

Jobson, D. J., Rahman, Z., & Woodell, G. A. (2002). A multiscale Retinex for bridging the gap between color images and the human observation of scenes. *IEEE TIP*, *6*(7), 965–976.

Land, E. H. (1977). The Retinex theory of color vision. *ScientificAmerican*, *237*(6), 108–128.

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, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2017). AOD-Net: Allin-one dehazing network. in: *ICCV*, pp. 4780–4788.

Li, M., Liu, J., Yang, W., Sun, X., & Guo, Z. (2018). Structure-revealing low-lightimageenhancementviarobustRetinexmodel.*IEEETIP*, *27*(6), 2828–2841.

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

Ma, K., Zeng, K., & Wang, Z. (2015). Perceptual quality assessment for multi-exposure image fusion. *IEEE TIP*, *24*(11), 3345–3356.

Mateescu, V., & Bajic, I. V. (2016). Visual attention retargeting. *IEEE Multimedia*, *23*(1),

Mechrez, R., Shechtman, E., & Zelnik-Manor, L. (2018). Saliency driven image manipulation. *WACV*, *30*, 189–202.

Mittal,A.,Soundararajan,R.,&Bovik,A.(2013).Makingacompletely blind image quality analyzer. *IEEE SPL*, *20*(3), 209–212.

Pisano, E., Zong, S., Hemminger, B., Deluca, M., Johnston, R., Muller, K., 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), 193–200.

Rahman, S., Rahman, M. M., Abdullah-Al-Wadud, M., Al-Quaderi, G. D., & Shoyaib, M. (2016). An adaptive gamma correction for image enhancement. *Eurasip Journal on Image & Video Processing*, *2016*(1), 35.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. in: *MICCAI*, pp. 234–241.

Sharma, G., Wu, W., & Dalal, E. N. (2005). The CIEDE2000 colordifference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research and Application*, *30*(1), 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: 1711.02488.

Stevens, S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153–181.

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

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

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

Wang, W., Chen, W., Yang, W., & Liu, J. (2018). GLADNet: Low-light enhancement network with global awareness. in: *FG*.

Wang,Z.G.,Liang,Z.H.,&Liu,C.(2009).Areal-timeimageprocessor with combining dynamic contrast ratio enhancement and inverse gamma correction for PDP. *Displays*, *30*, 133–139.

Wei, C., Wang, W., Yang, W., & Liu, J. (2018). Deep Retinex decomposition for low-light enhancement. in: *BMVC*.

Xie, J., Xu, L., & Chen, E. (2012). Image denoising and inpainting with deep neural networks. in: *NeurIPS*, pp. 341–349.

Xie, J., Xu, L., Chen, E., Xie, J., & Xu, L. (2012). Image denoising and inpainting with deep neural networks. in: *NeurIPS*, pp. 341–349.

Ying, Z., Ge, L., & Gao, W. (2017). A bio-inspired multi-exposure fusion framework for low-light image enhancement. arXiv: 1711.00591.

Ying, Z., Ge, L., Ren, Y., Wang, R., & Wang, W. (2018). A new lowlight image enhancement algorithm using camera response model. in: *ICCVW*, pp. 3015–3022.

Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2016). Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE TIP*, *26*(7), 3142–3155.

Zhang,K.,Zuo,W.,&Zhang,L.(2018).FFDNet:Towardafastandflexible solution for CNN-based image denoising. *IEEE TIP*, *27*(9), 4608–4622.

Zhang, L., Dai, J., Lu, H., He, Y., & Wang, G. (2018). A bi-directional message passing model for salient object detection. in: *CVPR*, pp. 1741–1750.

Zhang, X., Lu, Y., Liu, J., & Dong, B. (2018). Dynamically unfolding recurrent restorer: A moving endpoint control method for image restoration. in: *ICLR*.

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