An efficient framework for deep learning-based light-defect image enhancement

Chengxu Ma | Daihui Li | Shangyou Zeng | Junbo Zhao | Hongyang Chen

College of Electronic Engineering, Guangxi Normal University, Guilin, Guangxi, People’s Republic of China (Email: mcx950202@163.com)

Correspondence
Shangyou Zeng, College of Electronic Engineering, Guangxi Normal University, Guilin, 541004, Guangxi, People’s Republic of China.
Email: zsy@mailbox.gxnu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 61976063

Abstract
The enhancement of light-defect images such as extremely low-light, low-light and dim-light has always been a research hotspot. Most of the existing methods are excellent in specific illuminations, and there is much room for improvement in processing light-defect images with different illuminations. Therefore, this study proposes an efficient framework based on deep learning to enhance various light-defect images. The proposed framework estimates the reflectance component and illumination component. Next, we propose a generator guided by an attention mechanism in the reflectance part to repair the light-defect in the dark. In addition, we design a colour loss function for the problem of colour distortion in the enhanced images. Finally, the illumination map of the light-defect images is adjusted adaptively. Extensive experiments are conducted to demonstrate that our method can not only deal with the images with different illuminations but also enhance the images with clearer details and richer colours. At the same time, we prove its superiority by comparing it with state-of-the-art methods under both visual quality comparison and quantitative comparison of various datasets and real-world images.

1 | INTRODUCTION

High-visibility images have vivid details and rich colours, which not only meets human sensory needs but also plays a vital role in computer vision, such as image enhancement, feature matching [1], object detection [2], object recognition [3], object tracking [4], and so forth. However, due to the poor shooting environment, the limitation of the performance of the shooting equipment and the differences in shooting technology, the visibility of the images will be reduced. For this kind of low-light images, deep learning-based image enhancement has shown significantly good performance in improving image quality [5].

The Retinex theory [6] assumes that the image is divided into illumination component and reflectance component. It is a model of human visual brightness and colour perception. It concentrates on local enhancement, which solves the problems of insufficient brightness enhancement of local areas of the images by the global enhancement algorithm to a certain degree; and it can adaptively enhance different types of images. Early attempts based on Retinex, such as single-scale Retinex [7] and multi-scale Retinex [8], treat the reflectance as the final enhanced result, which often looks unnatural and appears to be over-enhanced frequently [9]. Guo et al. [9] estimated the illumination of each pixel individually and refined the initial illumination map by imposing a structure prior to it as the final illumination map. However, due to the blindness of light structure, they might lose the realism of regions with rich textures. Wang et al. [10] proposed a global illumination perception and detail preserving network (GLADNet). GLADNet calculated global illumination estimation for the low-light input, then adjusted the illumination under the guidance of the estimation and supplemented the details using a concatenation with the original input. This method was prone to distortions such as colour and background in low-light image enhancement. Chen et al. [11] learnt to decompose the observation image into reflectance image and illumination image in a data-driven way without considering the ground truth of the reflectance and illumination decomposition and then enhanced the obtained illumination image. This method can well maintain the colour saturation of the images during the enhancement process, but it needs to be improved in the processing of noise and details. On the basis of the classic Retinex model, Li et al. [12] proposed a robust...
image enhancement. As schematically illustrated in Figure 1, the RLDNet is composed of three modules, including image decomposition, reflectance restoration and illumination adjustment. In this study, the light-defect images are decomposed into illumination and reflectance images. On the reflectance images, the UNet++ generator guided by attention is used to restore the reflectance images and remove the noise to repair the dark defects. Furthermore, we add a colour loss function to ensure that the original colours of the images are not lost. The illumination part adopts adaptive adjustment, which is the core of the flexible adjustment of image brightness. Next, this study will introduce the three modules of RLDNet in detail as shown in Figure 1.

2.1 Image decomposition

The image decomposition module contains two branches corresponding to the illumination and the reflectance, respectively. We adopt UNet++, followed by convolutional and sigmoid layers to get the reflectance branch; while the illumination branch is composed of two convolutional layers following rectified linear unit (ReLU) layers and a convolutional layer on concatenated feature maps from the reflectance branch, followed by a sigmoid layer finally. We use UNet++ instead of U-Net used in KinD. UNet++ is equivalent to filling the original hollow U-Net, which can capture different levels of features and integrate them through feature superposition. Different sizes of receptive fields have different sensitivities to target objects of different sizes. It is easy to recognise large objects because of the large receptive field. In the actual decomposition process, the edge information of large objects and the small objects themselves are easy to be lost due to the down-sampling and up-sampling operations of the deep network, so we need the features of the small receptive field to help us. Using UNet++ can decompose more subtle features, which is helpful for subsequent operations such as restoration and denoising of reflectance images. The specific details of image decomposition are shown in Figure 2.

In the image decomposition process, we obtain a pair of light-defect/normal image each time and learn the decomposition of
The framework is composed of three modules, including image decomposition, reflectance restoration and illumination adjustment. Input_low refers to the input light-defect image. Input_norm refers to the normal image corresponding to the light-defect image. Ratio represents the enhancement ratio between $I_{norm}$ and $I_{low}$, and extends to the feature map. $R_{low}$ is a reflectance of the low-light image obtained by decomposition network. $R_{norm}$ is a reflectance of the normal image. $I_{low}$ represents the illumination map of the low-light image obtained by decomposition network. $I_{norm}$ represents the illumination map of the normal image. $R_{output}$ represents the enhanced reflectance map, and $I_{output}$ represents the enhanced illumination map. Enhanced image is obtained by multiplying the enhanced illumination map and the reflectance map.

The specific details of image decomposition. Input_low refers to the input light-defect image. Input_norm refers to the normal image corresponding to the light-defect image. $R_{low}$ is a reflectance of the low-light image obtained by decomposition network. $R_{norm}$ is a reflectance of the normal image. $I_{low}$ represents the illumination map of the low-light image obtained by decomposition network. $I_{norm}$ represents the illumination map of the normal image.

The input of light-defect image is denoted as $S_{low}$, and the normal image is denoted as $S_{norm}$. In this study, the reflectance components of the light-defect/normal images obtained by decomposition network are denoted as $R_{low}$ and $R_{norm}$, respectively, and the illumination components are $I_{low}$ and $I_{norm}$. Generally speaking, reflectivity is an inherent property of the object and is not affected by illumination. In the condition of no image degradation, the reflectivity should be the same. Furthermore, the illumination maps should be piecewise smooth and mutually consistent. Therefore, in the image decomposition part, we set two
constraints: (1) Light-defect and normal images share the reflectivity. (2) The illumination mapping is smooth and consistent with each other.

The decomposed two layers should reproduce the input, which is constrained by the reconstruction error. The reconstruction loss function is given by the following formula where $|.|_1$ means $L_1$ norm:

$$L_{rec} = ||S_{low} - R_{low}oI_{low}||_1 + ||S_{norm} - R_{norm}oI_{norm}||_1$$ (2)

In order to verify the correctness of the image decomposition, we usually need to ensure that the decomposed reflectance and the illumination maps are multiplied and the resulting image is consistent with the light-defect image.

Based on the principle of reflectance component consistency, we use the reflectance invariant loss function to regularize the reflectance similarity as shown in formula (3):

$$L_r = ||R_{low} - R_{norm}||_1$$ (3)

As for the mutual consistency, we adopt consistency loss function $L_{mc}$ [13]

$$L_{mc} = ||\omega \exp \left(-c \cdot m \right)||_1$$ (4)

where $c$ is the parameter controlling the shape of the function. In this study, $c$ is set to 10. In formula (4), $m$ represents the sum of the gradient of the illumination maps about low-light and the normal images. The specific expression of $m$ can refer to the following formula (5) in which $\nabla$ is a first-order derivative operator with horizontal and vertical directions.

$$m = |\nabla I_{low}| + |\nabla I_{norm}|$$ (5)

In general, the illumination changes greatly in the strong edge areas of the input images, while in the weak edge areas, the illumination distribution should be smooth. The illumination smoothness loss function is shown in formula (6) where $t$ is a small positive constant. This study sets $t$ to 0.01 in order to avoid the denominator being zero.

$$L_s = ||\frac{\nabla I_{low}}{\max(|\nabla S_{low}|, t)}||_1 + ||\frac{\nabla I_{norm}}{\max(|\nabla S_{norm}|, t)}||_1$$ (6)

Using the smoothness loss function can reduce over-fitting, improve the generalisation ability of the network, and enhance image contrast. This smoothness term measures the relative structure of the illumination with respect to the input. The penalty on illumination is small for a location on an edge in light-defect images, while for a location in a flat region in light-defect images, the penalty turns to be large.

To sum up, the expression of the loss function $L_{adv}$ of the image decomposition part is shown in formula (7):

$$L_{adv} = \eta L_{rec} + \lambda L_r + \mu L_{mc} + \nu L_s$$ (7)

In this study, $\eta$ is set to 1, $\lambda$ is set to 0.01, $\mu$ is set to 0.15, and $\nu$ is set to 0.1.

### 2.2 Reflectance restoration

The degradation of the light-defect image is more serious than that of the normal image, and the distribution on the reflectance is more complicated. When the estimated illumination map is smooth, all details are preserved in reflectance images, including the enhanced noise. Therefore, we can denoise on the reflectance images. For light-defect images, the distribution of noise is often uneven, and most of the noise is concentrated in the low-light area. Attention mechanism can make the network more focused on the part that needs attention, that is, we add attention to solve the problem of uneven distribution of noise. At the same time, the texture, colour and other details of the reflectance part will be reduced or even lost due to various reasons. The role of the attention mechanism is crucial. U-Net has achieved huge success on semantic segmentation, image restoration and enhancement [19, 20]. U-Net adopts superposition operation, and the most classic idea is encoder-decoder. U-Net retains rich information in texture and synthesises high-quality images using multi-scale context information. UNet++ is based on U-Net, which integrates long and short connections. UNet++ can capture features of different levels, integrate them by feature superposition, and add a shallower U-Net structure to make scale differences of the feature maps smaller during fusion, which is conducive to the extraction of more details. Therefore, this study proposes an attention-guided UNet++ generator in the reflectance part. Reflectance restoration cannot be processed uniformly on the whole image, so illumination map can be used as a guide. The original input is considered to contain more details, so we consider concatenating the original image with the illumination map. Then, we use five convolutional layers at $3 \times 3$ with ReLU as the activation function for feature extraction and use one convolutional layer at $1 \times 1$ to facilitate the information fusion between the feature maps. Finally, we get the attention feature map called $\alpha$. Among them, the number of channels is 32. The value of $\alpha$ is given by the following formula where $\sigma$ denotes the sigmoid function and $f^{3x3}$ represents convolution operation with the filter size $3 \times 3$.

$$\alpha = \sigma \left(f^{3x3} \left(S_{low}, I_{low} \right) \right)$$ (8)

Finally, the result of combining the attention and reflectance maps is used as the input of attention-guided UNet++ generator. Among them, the input of attention-guided UNet++ generator can be expressed as formula (9):

$$input_{UNet++} = \alpha R_{low} + R_{norm}$$ (9)

The specific details of reflectance restoration are shown in Figure 3.

Our principle is to use a clear reflectance as a reference for a clutter reflectance. The reference used for denoising in this
The specific details of reflectance restoration. Input_low refers to the input light-defect image. R_low is a reflectance of the low-light image obtained by decomposition network. R_norm is a reflectance of the normal image. I_low represents the illumination map of the low-light image obtained by decomposition network. R_output represents the enhanced reflectance map.

study is a reflectance image decomposed from a normal image. \( L_{sim} \) is a structural similarity loss function, which takes into account the brightness, contrast and structural indicators. It is more conducive to and close to human visual perception. Generally, the results produced by \( L_{sim} \) are more detailed than those produced by \( L_1 \) and \( L_2 \) loss function and do not make the images too smooth. It can be expressed as formula (10) where \( R \) corresponds to the restored reflectance:

\[
L_{sim} = SSIM(R, R_{norm})
\]

(10)

\( L_2 \) norm can better remove noise and ring artefacts in the enhanced results [21]. \( L_2 \) and the gradient loss function \( L_{grad} \) are shown in formulas (11) and (12):

\[
L_2 = || \dot{R} - R_{norm} ||^2_2
\]

(11)

\[
L_{grad} = || \nabla \dot{R} - \nabla R_{norm} ||^2_2
\]

(12)

In order to avoid insufficient colour saturation and distortion during the restoration process, we design a colour loss function \( L_{rgb} \) to ensure that the colour is as close as possible to the colour of the normal images. Its expression is shown as the following formula where \( \epsilon \) is a small positive constant:

\[
L_{rgb} = || \frac{R_{max} - \dot{R}}{R_{max} + \epsilon} - \frac{R_{norm_{max}} - \dot{R}_{norm}}{R_{norm_{max}} + \epsilon} ||^2_2
\]

(13)

This study sets \( \epsilon \) to 0.1 in order to avoid the denominator being zero.

To sum up, the expression of the loss function \( L_{re} \) of the reflectance restoration network is shown in formula (14):

\[
L_{re} = \eta L_{sim} + \lambda L_2 + \mu L_{grad} + \omega L_{rgb}
\]

(14)

In this study, \( \eta, \lambda, \mu \) are set to 1, and \( \omega \) is set to 4.2.

2.3 | Illumination adjustment

In practical applications, we will encounter light-defect images with different illuminations.

In order to achieve various light-defect image enhancement tasks with high efficiency, we need a mechanism to flexibly convert one light condition to another. Although we do not know the exact relationship between the target illumination source \( I_{norm} \) and the incident illumination source \( I_{low} \), we can know its enhancement ratio, which is expressed as formula (15):

\[
\beta = \frac{I_{norm}}{I_{low}}
\]

(15)

The target illumination sources in this study are higher than the incident illumination sources of the light-defect images. So the ratio should be greater than 1. \( \beta \) can be specified by users in the testing phase. \( \beta \) is extended to a characteristic graph with \( I_{low} \) as the input of the illumination adjustment network. The illumination adjustment module contains 11 convolutional layers (10 convolutional layers at \( 3 \times 3 \) ReLU and one convolutional layer at \( 1 \times 1 \)) and one sigmoid layer. The channels are 64. In addition, the convolutional layers 2, 3, 5, 6, 8, and 9 are all added with the features obtained by the previous layer of convolution, which can make it possible to repair the dark light defect while obtaining clearer details and richer colours. We call the 11 convolution layers the merge module in this study. The advantages of this module, compared to KinD, can be verified in the ablation experiment in the next section. The details of the illumination adjustment network are shown in Figure 4.

The loss for illumination adjustment network is given by the following formula where \( \dot{I} \) corresponds to the adjusted illumination:

\[
L_{ia} = || \dot{I} - I_{norm} ||^2_2 + || \nabla \dot{I} - \nabla I_{norm} ||^2_2
\]

(16)
FIGURE 4  The details of the illumination adjustment network $I_{low}$ represents the illumination map of the low-light image obtained by decomposition network. Ratio represents the enhancement ratio between $I_{norm}$ and $I_{low}$ and extends to the feature map. $I_{output}$ represents the enhanced illumination map.

FIGURE 5  An example of four datasets: LOL, Fivek, DICM, MEF (a) extreme low-light images in LOL dataset, (b) normal images in LOL dataset, (c) low-light images in Fivek dataset, (d) normal images in Fivek dataset, (e) dim-light images in DICM dataset, (f) dim-light images in MEF dataset.

3  EXPERIMENTS

3.1  Datasets and experimental details

In this study, we train on the public paired dataset LOL dataset [11]. The LOL dataset is a dataset of low/normal-light image pairs taken from real scenes by changing exposure time and ISO. LOL dataset contains 500 light-defect images and 500 normal images corresponding to the light-defect images. The scenes in this dataset are diverse, most of which are extreme low-light images, and the image resolution is $600 \times 400$. This study randomly selects 500 pairs of low-illumination images and image pairs corresponding to normal images in another public MIT-Adobe Fivek [22] dataset, which contains 5000 original images. Most of the images in this dataset are low illumination. And the image resolution is $512 \times 512$. In practical applications, there are no corresponding normal images as references for light-defect images, and in order to test that our method is not only applicable to extreme low-light, low-light, but also dim-light images. We also choose public datasets DICM [23] and MEF [24] without image pairs to test. An example of four datasets is shown in Figure 5.

For the image decomposition network and illumination adjustment network, the batch size is set to 10 and patch size to be $48 \times 48$. While for the reflectance restoration network, the batch size is set to be four and patch size to be $384 \times 384$. The entire network is trained on an NVIDIA Geforce GTX1070 GPU, RAM of 64GB, AMD Ryzen R9 3900 12-Core Processor 3.80 GHz CPU using the TensorFlow framework.

It is worth noting that KinD [13] has published a better pre-trained model on its GitHub. One of the methods compared in this study, KinD, can get better test results both visually
and quantitatively than the results published in the study by KinD authors. In other words, our method is compared with the improved version of KinD.

3.2 Ablation study

To demonstrate the effectiveness of each component proposed in our framework, we conduct several ablation experiments. Specifically, we design three experiments: (a) Removing UNet++ and Lrgb and the merge module on RLDNet. (b) Removing Lrgb and the merge module on RLDNet. (c) Removing the merge module on RLDNet; and RLDNet: Using a UNet++ generator guided by attention in the reflectance images, and a loss function is used to adjust the colour distribution, and a merge module is designed in the illumination part.

As shown in Figure 6, the first row shows the input images. The second row shows the images produced by KinD. The third, fourth and fifth rows represent three ablation experiments. The last row is produced by our proposed version of RLDNet.

Method (a) simply adds attention, it can be seen from the third column that the details such as the seat become obvious, but in the second column of the frame display, the detailed texture of the wool is very poor. In comparison, due to the role of UNet++, the details are relatively better in (b) and (c). Compared with KinD, (a), (b), a colour loss function is added in method (c). The purple blanket in the first column in Figure 6
can well prove the indispensability of colour loss. In contrast, our method RLDNet is accompanied by colour loss function under the action of the UNet++ generator guided by attention. At the same time, the illumination adjustment part uses more convolutional layers for feature extraction. It is excellent in terms of detail clarity, contrast or colour distribution. In summary, the effectiveness of the component designed in the framework can be verified.

3.3 Comparison with state-of-the-art methods on the datasets

In this section, we compare the performance of RLDNet with current state-of-the-art methods on the public datasets as mentioned in Section 3.1. We conduct a list of experiments including visual quality comparison, image quality assessments, which are elaborated in the next section.

3.3.1 Visual quality comparison

In order to evaluate the performance of our method, we compare our method with low-light image enhancement via illumination map estimation (LIME) [9], a global illumination-aware and detail-preserving network (GLAD) [10], Retinex-Net [11], Robust Retinex [12], kindling the darkness (KinD) [13] and Zero-DCE [14] for visual quality comparison. The experimental comparison images are shown in Figures 7–11.

To enhance the light-defect images, we must ensure that during the enhancement process, the main bodies of the images are enhanced while the backgrounds are not distorted. Combining colour, detail definition, denoising performance, and so forth, we analyse the enhanced images.

In Figure 7, the colour distortion of GLAD (c), KinD (f) can be seen clearly from the red frame. The colour of Retinex-Net (d) is very bright, but there is too much noise. The enhanced brightness of LIME (b), Robust Retinex (e) and Zero-DCE (g)
are insufficient. We can see that our method outperforms other methods in terms of noise and colour, which can be clearly seen from the place marked by the red frame.

In Figure 8, the colour distortion of GLAD (c) can be seen clearly. The colours of Retinex-Net (d) on stands, floors and other objects are very bright, but there is too much noise. The details of LIME (b), Robust Retinex (e), KinD (f) and Zero-DCE (g) are not as clear as our methods. Finally, we can see that our method outperforms other methods in terms of noise, colour and detailed clarity, which can be clearly seen from the place marked by the red frame.

In Figure 9, from the sky point of view, the brightness of the sky in method Robust Retinex (e) is too bright, which leads to the distortion of clouds compared with the input image (a), and the colour of the sky in method Zero-DCE (g) is too different from the input image (a). From the perspective of building colour and details, the colour of Retinex-Net(d) is so bright that it is different from the input image (a). The overall brightness of enhancement results of LIME (b) and GLAD (c) is not enough, which leads to the poor visual effect. KinD (f) is inferior to our method in terms of building colour and details of the window of arched doors. On the whole, our method RLDNet is excellent in detail, colour and overall effect.

In Figure 10, the sky brightness of GLAD (c) is over-enhanced, resulting in the loss of details such as white clouds. Zero-DCE (g) has distortion in sky colour. On the premise that the enhancement results are not distorted, the overall brightness of LIME (b) is not as good as other methods. Robust Retinex (e) does not perform as well as KinD (f) and RLDNet in mountain details. However, the colour of KinD on the green field is not as good as RLDNet. As a result, RLDNet performs best, no matter from which perspective.

In Figure 11, the colour distortion of GLAD (c), Retinex-Net (d), KinD (f) and Zero-DCE (g) can be seen clearly. Owing to
FIGURE 11  The visual quality comparison of proposed and current state-of-the-art methods on the Fivek dataset (a) input image, (b) LIME [9], (c) GLAD [10], (d) Retinex-Net [11], (e) Robust Retinex [12], (f) KinD [13], (g) Zero-DCE [14], (h) RLDNet

the lack of constraint on the reflectance, LIME (b) can easily cause a lack of naturalness in the enhanced images [25]. Because Figure 11 comes from a dataset with image pairs, our method is closer to the normal image in the overall color distribution of the image. The overall brightness of Li et al. (e) is too high, causing the color of the leaves to appear distorted.

Combined with a large number of experimental results, we can see that the proposed method is better than other methods in color, detailed clarity, and denoising performance. Furthermore, RLDNet can obtain good enhancement effects on extreme low-light, low-light, and dim-light images. It can be seen that the framework in this study has a wide range of application and strong practicability.

3.3.2  Quantitative comparison

The differences in user preferences will make the visual quality comparison not comprehensive enough. In order to further verify the advantages of the framework, this study uses structural similarity (SSIM) [26] and peak signal-to-noise ratio (PSNR) to evaluate the image quality of reference images. In this study, 15 images in LOL dataset are selected for testing. In order to ensure the fairness of the comparison, the test images are excluded where the SSIM and PSNR are extremely high or low. We use the average of 15 test images as the test value of SSIM and PSNR. Quantitative comparisons on LOL dataset in terms of SSIM and PSNR are summarised in Table 1.

TABLE 1  Quantitative comparison on LOL dataset in terms of peak signal-to-noise ratio (PSNR), structural similarity (SSIM)

| Methods         | Year | PSNR  | SSIM  |
|-----------------|------|-------|-------|
| LIME [9]        | 2017 | 14.91 | 0.686 |
| GLAD [10]       | 2018 | 21.06 | 0.733 |
| Retinex-Net [11]| 2018 | 18.06 | 0.539 |
| Robust Retinex [12]| 2018 | 14.25 | 0.715 |
| KinD [13]       | 2019 | 21.28 | 0.876 |
| Zero-DCE [14]   | 2020 | 15.43 | 0.666 |
| RLDNet          | —    | 22.91 | 0.885 |

To comprehensively evaluate various images. We select some images from four datasets separately to calculate the average value of their NIQE. Quantitative comparison on low-light dataset (LOL) [11], Fivek [22], DICM [23] and multi-exposure image fusion (MEF) [24] datasets in terms of NIQE are summarised in Table 2.

TABLE 2  Quantitative comparison on LOL, Fivek, DICM and MEF datasets in terms of NIQE

| Methods         | LOL  | Fivek | DICM  | MEF   |
|-----------------|------|-------|-------|-------|
| LIME [9]        | 7.51 | 7.08  | 4.14  | 3.83  |
| GLAD [10]       | 6.47 | 6.77  | 3.64  | 3.71  |
| Retinex-Net [11]| 8.88 | 7.32  | 5.12  | 5.25  |
| Robust Retinex [12]| 5.81 | 7.78  | 5.26  | 5.04  |
| KinD [13]       | 5.35 | 6.78  | 4.06  | 3.82  |
| Zero-DCE [14]   | 7.77 | 7.42  | 3.43  | 3.85  |
| RLDNet          | 5.01 | 6.64  | 3.05  | 3.13  |
The higher the value of SSIM and PSNR, the closer the enhanced images are to the normal images. On the contrary, a lower NIQE value indicates better visual quality. As can be seen from Tables 1 and 2, although Zero-DCE [14] is currently one of the latest methods for low-light image enhancement, it is mainly for non-reference images. For reference datasets LOL and Fivek, it is at a disadvantage from an objective perspective. In addition, due to excessive noise, colour distortion, different enhancement effects of Zero-DCE [14] appear on different non-reference datasets. Even though the method in this study uses a reference dataset for training in the training phase, the processing of noise, detailed texture and colour make RLD-Net still perform well on the non-reference datasets DICM and MEF.

3.4 Comparison with state-of-the-art methods on the real-world images

From the perspective of the application, we designed an effective framework for light-defect images, which not only has an excellent performance in various datasets but also can adapt to real-life scenes. We use the camera Nikon D5200 and Huawei mate9 as the shooting equipment and select the surrounding areas, urban areas and schools as our shooting sites. We select three representative images as the next comparison object.

3.4.1 Visual quality comparison

Figures 12–14 show several comparisons of our method RLD-Net with LIME [9], GLAD [10], Retinex-Net [11], Robust Retinex [12], KinD [13] and Zero-DCE [14] on real-world images.

LIME [9] adds a structure-aware prior to the illumination map to adjust the illumination of light-defect images. It leads to insufficient enhancement of LIME for low-light areas of light-defect images, such as curtains and pink plush phone case in Figure 12(b), the vase in Figure 13(b) and the texture of the tree in Figure 14(b).
GLAD [10] has shortcomings in the enhancement of brightness in extreme low-light areas, and many textures cannot be reflected, such as curtains and pink plush phone case in Figure 12(c), the vase in Figure 13(c). In addition, distortion is prone to appear in places with high brightness. For example, in the ground area marked by the red frame in Figure 14(c), its brightness is so bright that its colour and the texture of the ground are distorted.

Retinex-Net [11] performs very well for colour during the enhancement process, but its noise is too large, which can be seen in Figures 12—14(d).

Compared with Retinex-net, the method of Robust Retinex [12] has a great improvement in image noise processing, but there is still a lot of room for improvement in detail processing. For example, the texture of the vase of Figure 13(e), and the texture of the tree in Figure 14(e) are severely lost. In addition, the text part in the middle of the pink plush phone case in Figure 12(c) and the surrounding border are blurred.

KinD [13] is effective in noise processing, but the colour of the enhanced object will show a certain degree of distortion, such as the colour of the curtains and pink plush phone case in Figure 12(f). It can be seen from Figure 13(f) that the blue of the vase itself and the details of the veins in the middle of the vase are slightly worse than RLDNet. Due to the overall brightness, the details of the enhanced objects are lost. For example, the ground in Figure 14(f) looks smooth and has no texture, and the effect of leaves is not as good as our method.

Zero-DCE [14] is trained end-to-end with zero reference images by devising a set of differentiable non-reference losses. However, this method is prone to noise, such as the curtains and pink plush phone case in Figure 12(g). Owing to the lack of constraint on the reflectance map, the brightness of the enhancement result is prone to be too bright, resulting in the lack of texture and colour distortion of the object itself, such as the colour of the vase in Figure 13(g), the texture of the ground and its colour in Figure 14(g).

As can be seen from the above, our method RLDNet can enhance the brightness of the low-light area as much as possible during the enhancement process while preserving the background area without distortion. At the same time, our method is more excellent in denoising, colour and texture preservation.

### 3.4.2 Quantitative comparison

The real-world images we select represent different scenes and brightness. We select the indoor extreme low-light image (Figure 12), indoor dim-light image (Figure 13), and the image of the street at night with extreme low light (Figure 14). We use NIQE [27] to evaluate the quality of Figures 12—14. The quantitative measurement results are shown in Table 3.

It can be seen from Table 3 that the enhancement effect of our method RLDNet in this study is the same as the visual comparison result, which is the best. The order of the methods in
the above table is based on the publication year of the studies. Although zero-DCE [14] is one of the most advanced methods in the field of low-light image enhancement, it will also result in low-quality enhanced images due to excessive noise and texture loss caused by excessive brightness. Although our method is trained on a reference dataset, it benefits from the generator guided by the attention mechanism of the reflectance map and the design of the colour saturation loss function. For reference and non-reference images, it can get a better enhancement effect and can deal with the problems of light-defect images with different degrees of illumination.

Combining the visual quality comparison and quantitative comparison results on various datasets and real-world images, the framework in this study is feasible and universal.

4 CONCLUSION

In this study, we design a framework RLDNet based on deep learning for light-defect image enhancement. The RLDNet is composed of three modules, including image decomposition, reflectance restoration and illumination adjustment. Inspired by the Retinex theory, we decompose the light-defect image into the reflectance and illumination components. To remove the degradations previously hidden in the darkness, we choose to perform denoising and other operations on the amount of reflectance. We design a UNet++ generator guided by attention in the reflectance restoration part. In addition, a colour loss function is proposed in the reflectance restoration part to ensure that the colours of the images are as close as possible to the normal images. In the illumination part, we design an adaptive adjustment network to convert one illumination condition into another. Combining the ablation experiments, visual and quantitative comparisons in this study can all prove the effectiveness of the framework.

Compared with the state-of-the-art methods, our method can process the low-light images under various illumination conditions well. RLDNet enables the light-defect images to retain their original features in the background areas during the enhancement process, and the results of the enhancement of the low-light areas have clearer details and richer colours. The RLDNet in this study can be used in visual appreciation and also can be used in computer vision technology, which has high application value. Our future work is to make a dataset that combines low-light, normal, and exposed images to expand the application range of light-defect image enhancement network. In addition, we intend to lightweight the method or add timing processing so that our method can be better used in video post-processing.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (grant no. 61976063). Chengxu Ma, Dairui Li, Shangyou Zeng, Junbo Zhao and Hongyang Chen (College of Electronic Engineering, Guangxi Normal University, Guillin 541004, Guangxi, People’s Republic of China)

ORCID

Chengxu Ma https://orcid.org/0000-0003-4329-5606

REFERENCES

1. Wu, Q., et al.: Feature matching and deformation for texture synthesis. In: ACM SIGGRAPH 2004, Los Angeles, California, pp. 364–367 (2004).
2. Oueata, D., et al.: Spatio-temporal object detection proposals. In: Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, pp. 737–752 (2014).
3. Salve, S.G., et al.: Shape matching and object recognition using shape contexts. In: 2010 3rd International Conference on Computer Science and Information Technology, Chengdu, China, pp. 471–474 (2010).
4. Zhang, K., et al.: Real-time compressive tracking. In: Proceedings of the 12th European Conference on Computer Vision(ECCV 2012), Florence, pp. 864–877 (2012).
5. Du, J., et al.: RGB-IR cross input and sub-pixel upsampling network for infrared image super-resolution. Sensors 20(1), 281–295 (2020).
6. Land, E.H., et al.: The Retinex theory of color vision. Sci. Am. 237(6), 108–128 (1977).
7. Jobson, D.J., et al.: Properties and performance of a center/surround Retinex. IEEE Trans. Image Process. 6(3), 451–462 (1997).
8. Jobson, D.J., et al.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Trans. Image Process. 6(7), 965–976 (1997).
9. Guo, X., et al.: LIME: Low-light image enhancement via illumination map estimation. IEEE Trans. Image Process. 26(2), 982–993 (2017).
10. Wang, W., et al.: GLADNet: Low-light enhancement network with global awareness. In: Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi’an, China, pp. 751–755 (2018).
11. Chen, W., et al.: Deep retinex decomposition for low-light enhancement. In: Proceedings of the British Machine Vision Conference, Newcastle, UK (2018).
12. Li, M., et al.: Structure-revealing low-light image enhancement via robust retinex model. IEEE Trans. Image Process. 27(6), 2828–2841 (2018).
13. Zhang, Y., et al.: Kindling the darkness: A practical low-light image enhancer. In: The 27th ACM International Conference on Multimedia, Nice, France, pp. 1632–1640 (2019).
14. Guo, C., et al.: Zero-reference deep curve estimation for low-light image enhancement. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, pp. 1777–1786 (2020).
15. Zhou, Z., et al.: UNet++: A nested U-Net architecture for medical image segmentation. In: MICCAI 2018, Granada, Spain, pp. 3–11 (2018).
16. Burt, P.J., et al.: Attention mechanisms for vision in a dynamic world. In: [1988 Proceedings] 9th International Conference on Pattern Recognition, Rome, Italy, pp. 977–987 (1988).
17. Park, S., et al.: Dual autoencoder network for retinex-based low-light image enhancement. IEEE Access 6, 22084–22093 (2018).

| Images      | Figure 12 | Figure 13 | Figure 14 |
|-------------|-----------|-----------|-----------|
| LIME [9]    | 4.2522    | 4.9425    | 2.6165    |
| GLAD [10]   | 4.2480    | 4.7793    | 2.7108    |
| Retinex-Net [11] | 6.0610 | 5.7372    | 3.5363    |
| Robust Retinex [12] | 3.9682 | 4.4621    | 6.1050    |
| KinD [13]   | 3.2980    | 3.8287    | 2.8938    |
| Zero-DCE [14] | 4.7515 | 5.0283    | 2.8121    |
| RLDNet      | 3.1798    | 3.8218    | 2.4071    |
18. Ronneberger, O., et al.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI 2015, Munich, Germany, pp. 234–241 (2015)
19. Ding, L., et al.: When image denoising meets high-level vision tasks: A deep learning approach. In: IJCAI, Palo Alto, California, USA, pp. 842–848 (2017)
20. Jiang, Y., et al.: EnlightenGAN: Deep light enhancement without paired supervision. arXiv:1906.06972. (2019)
21. Zhao, H., et al.: Loss functions for neural networks for image processing. IEEE Trans. Comput. Imaging 3(1), 47–57 (2016)
22. Bychkovsky, V., et al.: Learning photographic global tonal adjustment with a database of input/output image pairs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Rhode Island, USA, pp. 97–104 (2011)
23. Lee, C., et al.: Contrast enhancement based on layered difference representation. In: 2012 19th IEEE International Conference on Image Processing, Orlando, Florida, USA, pp. 965–968 (2012)
24. Ma, K., et al.: Perceptual quality assessment for multi-exposure image fusion. IEEE Trans. Image Process. 24(11), 3345–3356 (2015)
25. Wu, Y., et al.: Low light image enhancement based on non-uniform illumination prior model. IET Image Proc. 13(13), 2448–2456 (2019)
26. Wang, Z., et al.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. 13(4), 600–612 (2004)
27. Mittal, A., et al.: Making a “completely blind” image quality analyzer. IEEE Signal Process Lett. 20(3), 209–212 (2013)

How to cite this article: Ma C, Li D, Zeng S, Zhao J, Chen H. An efficient framework for deep learning-based light-defect image enhancement. IET Image Process. 2021;15:1553–1566.
https://doi.org/10.1049/ipr2.12125