Deep Color Consistent Network for Low-Light Image Enhancement

Zhao Zhang[1,2], Huan Zheng[1,2], Richang Hong[1,2], Mingliang Xu[3], Shuicheng Yan[4], Meng Wang[1,2]

[1]School of Computer Science and Information Engineering, Hefei University of Technology, China
[2]Key Laboratory of Knowledge Engineering with Big Data (Ministry of Education) & Intelligent
Interconnected Systems Laboratory of Anhui Province, Hefei University of Technology, China
[3]School of Information Engineering, Zhengzhou University, China
[4]Sea AI Lab (SAIL), Singapore

Abstract

Low-light image enhancement (LLIE) explores how to refine the illumination and obtain natural normal-light images. Current LLIE methods mainly focus on improving the illumination, but do not consider the color consistency by reasonably incorporating color information into the LLIE process. As a result, color difference usually exists between the enhanced image and ground-truth. To address this issue, we propose a new deep color consistent network termed DCC-Net to retain the color consistency for LLIE. A new "divide and conquer" collaborative strategy is presented, which can jointly preserve color information and enhance the illumination. Specifically, the decoupling strategy of our DCC-Net decouples each color image into two main components, i.e., gray image plus color histogram. Gray image is used to generate reasonable structures and textures, and the color histogram is beneficial for preserving the color consistency. That is, they both are utilized to complete the LLIE task collaboratively. To match the color and content features, and reduce the color consistency gap between enhanced image and ground-truth, we also design a new pyramid color embedding (PCE) module, which can better embed color information into the LLIE process. Extensive experiments on six real datasets show that the enhanced images of our DCC-Net are more natural and colorful, and perform favorably against the state-of-the-art methods.

1. Introduction

Low-light image enhancement (LLIE) is a task of refining the illumination to obtain natural norm-light images, which aims at improving the perception and visual quality of low-light images captured in poor illumination environment. Low-light images are rather common in reality, e.g., images captured in outdoor or indoor scenes with poor lighting conditions, which suffers from the unclear contents and

RetinexNet (15.49/0.29) Zero-DCE++ (18.71/0.37) KinD++ (21.84/0.75)

DCC-Net (27.54/0.79) EnlightenGAN (13.66/0.50) Ground-Truth

Figure 1. Comparison of our DCC-Net and other deep LLIE methods in terms of PSNR/SSIM metrics. We clearly see that there is a large color gap between the enhanced images of RetinexNet, Zero-DCE++, Kind++ and EnlightenGAN, and the ground-truth image. In contrast, our DCC-Net can retain the color consistency effectively, and the enhanced image is more natural and colorful.

textures, low contrast and noises. These degradations will not only have negative effect on human perception, but also will be not conducive to the subsequent multimedia computing and computer vision tasks designed for high-quality images, for instance face recognition [3], object detection [25] and semantic segmentation [4].

Traditional LLIE methods aim to build a model to refine the illumination and obtain the enhanced image, which can be roughly categorized into histogram equalization (HE)-based and retinex-based methods. Traditional methods are relatively simple and easy, but they usually cannot restore the consistent colors and detailed textures.

With the impressive performance of deep neural networks (DNN) in diverse high-level and low-level vision tasks [5, 10, 31, 36], deep LLIE methods also achieve great improvement [2, 11, 30]. Deep LLIE methods usually design

a deep neural network equipping with different modules to reverse the degradation process. Compared to traditional methods that usually produce undesirable illumination and noises, deep LLIE methods can obtain better results due to the strong ability of DNN. However, these methods tend to generate inconsistent colors, which can be seen in Figure 1. There is obvious color difference between the generated images of RetinexNet, Zero-DCE++, Kind++ and EnlightenGAN and the ground-truth. While the result of our DCC-Net is more natural and conforms to the real color. We ask: *what makes the enhanced images lose color consistency*? We attempt to answer this question from two respects:

1) **Different architectures**. There are two popular ways to handle the LLIE task in current studies: 1) end-to-end deep frameworks that directly handle the low-light image to obtain normal-light images; 2) retinex-based frameworks that decompose the image into reflectance and illumination for further processing. Both the two modes focus on refining illumination, while ignoring the color consistency and naturalness. Thus, there will be color gap in the enhanced images.

2) **Information mismatch**. Color histogram describes color information globally, which does not contain any spatial information. As a result, we cannot find suitable color information for specific contents of images. The connection between the color and content features is therefore unable to be directly built. This kind of information mismatch will make the enhanced images unnatural and contain inconsistent colors.

We therefore propose a new "divide and conquer" collaborative strategy, which can jointly retain the color consistency and enhance the illumination. Generally, the main contributions of this paper are summarized as follows:

- Technically, we introduce a new strategy to retain the color consistency for LLIE, and also propose a deep color consistent network termed DCC-Net to reduce the color difference between the enhanced image and ground-truth. To the best of our knowledge, this is the first work to enhance the illumination of low-light image by directly exploring the color consistency. Extensive experiments show that our DCC-Net can better enhance the illumination, and the enhanced images are more natural and consistent in color.

- To jointly retain the color consistency and enhance the illumination, DCC-Net designs a decoupling strategy to decouple a color image into a gray image and a color histogram that complete the LLIE task collaboratively. We design three sub-nets for DCC-Net, i.e., G-Net, C-Net and R-Net, as shown in Figure 2. G-Net aims at recovering the gray image that can offer rich structure and texture information. C-Net aims to learn the color

distributions, which will be conductive to the color coherence. R-Net combines the gray image and color information to restore the normal-light image.

- To better overcoming the weakness of lacking spatial information for color histogram, we also design a pyramid color embedding (PCE) module that consists of six color embedding (CE) sub-modules with pyramid structure. CE can match the color and content features, according to the affinities between them, so that the color information can be dynamically incorporated, which can further reduce the color gap between the enhanced image and ground-truth image.

2. Related work

In this section, a brief review on both traditional and deep LLIE methods will be presented.

2.1. Traditional LLIE Methods

HE-based methods. Based on various image priors, HE-based LLIE methods [17, 24] focus on changing the dynamic range of the image to improve the contrast, such as [1, 17]. However, HE-based methods pay attention to enhancing the contrast, rather than directly refining the illumination. Thus, the enhanced results may suffer from the under-enhancement or over-enhancement.

Retinex-based methods. This kind of methods [7, 8, 20] decompose image into the pixel-wise product of reflectance and illumination, inspired by the retinex theory [15]:

$$S = R \odot I, \tag{1}$$

where S denotes an image, R and I denote the corresponding reflectance and illumination respectively. By further processing the reflectance and illumination, enhanced results can be obtained [7, 8, 12, 13, 20, 27]. Since Retinex-based methods aim at estimating the illumination, which is hand-crafted and depends on intensive parameters tuning, the final result often contain inconsistent colors and noises.

2.2. Deep Learning-based LLIE Methods

Deep LLIE methods can usually outperform traditional methods due the strong learning ability of DNNs. According to whether paired data are used, current deep LLIE methods can be divided into three categories, that is, supervised, unsupervised and semi-supervised methods.

Supervised deep LLIE methods. For supervised methods, all training data are paired. We can further divide this kind of methods into retinex-based methods and end-to-end methods. Retinex-based deep LLIE methods similarly uses deep learning to compose an image into reflectance and illumination. For example, Wei el al. [30] proposed
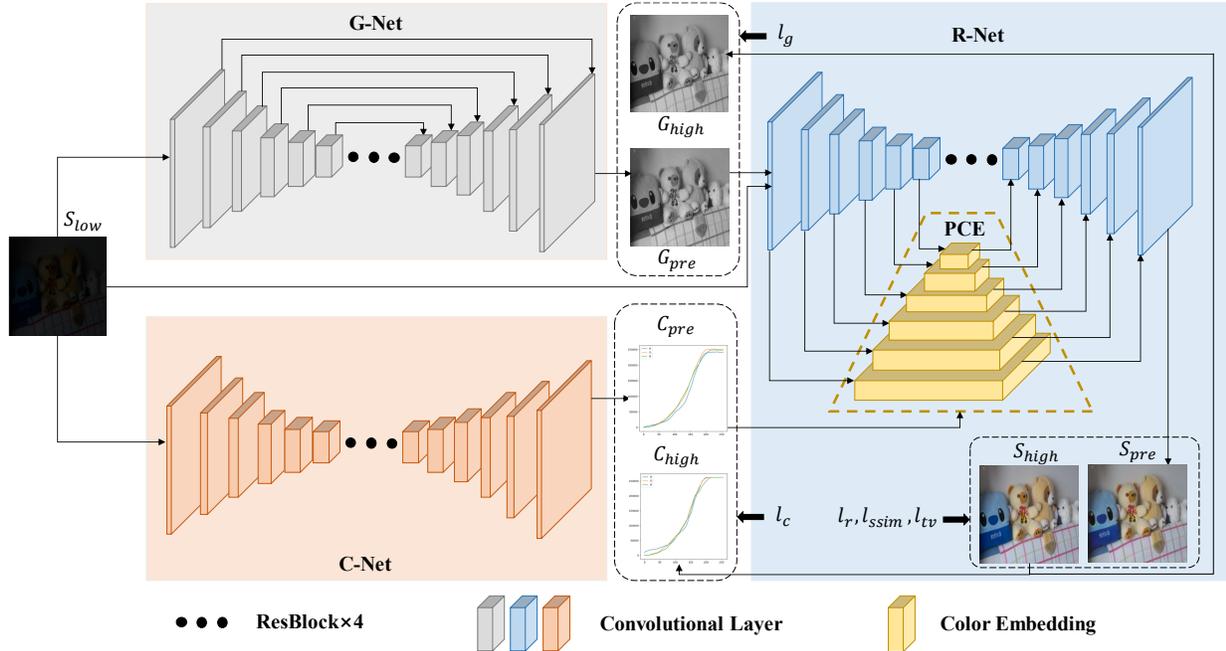
Figure 2. The overall framework of our DCC-Net. As can be seen, there are three sub-nets: G-Net, C-Net and R-Net, where G-Net aims at recovering the gray images with rich content information, C-Net focuses on learning the color distributions, and R-Net combines the gray image and color information to restore the natural and color-consistent normal-light images.

a RetinexNet with two stages, where the first stage decomposes an image into reflectance and illumination, and the second one adjusts the illumination map. Zhang et al. also presented two RetinexNet based improved models, called KinD [35] and KinD++ [34]. Compared with RetinexNet, there are three sub-networks for KinD and KinD++, which are decomposition-net, restoration-net and adjustment-net.

End-to-end deep LLIE methods directly handle the low-light image, rather than decomposing the image. For example, LLNet [21] is a pioneering work using deep auto-encoder approach for deep LLIE. Due to the strong representation ability of Convolutional Neural Networks (CNN), Li et al. [19] further presented a CNN-based deep LLIE model. By estimating the global contents of low-light image, a deep hybrid network was also developed to preserve the structure details in the enhanced image [26]. Supervised methods can achieve better weakly illuminated image enhancement, but the color consistency between enhanced image and ground-truth is still a difficult issue.

Unsupervised/semi-supervised deep LLIE methods. In reality, it is challenging or even impractical to simultaneously obtain the paired data, i.e., degraded and ground-truth images, of the same scene. Therefore, unsupervised/semi-supervised LLIE methods are studied to alleviate this problem. For example, Yang et al. [32] designs a deep recursive band network using paired and unpaired low/normal-light images to obtain a linear band representation of an enhanced normal-light image. Jiang et al. [11] present an unsuper-

vised LLIE method that employs the generative adversarial network (GAN) as main framework. There are also several zero-shot methods whose input only contains the low-light images [6, 18, 33, 38]. By training these zero-shot models with carefully-designed loss functions, the illumination of input low-light images also can be enhanced. Though these methods solves the LLIE problem without or with partial paired data, the enhancement quality is usually limited.

3. Proposed Method

In this section, we introduce the framework (see Figure 2) and details of DCC-Net, which aims at preserving the color consistency and naturalness in obtaining normal-light images. DCC-Net has three sub-nets (i.e., G-Net, C-Net, R-Net) and one pyramid color embedding (PCE) module.

3.1. Network Structure

G-Net. Given an input low-light image, the target of G-Net is to predict the gray image of normal-light image, which contains rich structure and texture information, without color information. This process is formulated as

$$G_{pre} = GNet(S_{low}), \qquad (2)$$

where G_{pre} denotes the predicted gray image, S_{low} denotes the input low-light image and $GNet$ denotes the transformation of G-Net. Specifically, G-Net employs the encoder-decoder pipeline, which is similar with classic U-Net [28].

For G-Net, we use l_1 loss to reconstruct the gray image:

$$l_g = \frac{1}{H \times W} \|G_{pre} - G_{high}\|_1, \qquad (3)$$

where l_g denotes the gray image reconstruction loss, G_{high} denotes the gray image of normal-light image, H and W denote the height and width of the gray image G_{high}. Hence, G-Net does not consider color information and devotes to recovering the textures and structures.

C-Net. Color histogram is a kind of color features, which is widely used in image retrieval systems [9]. Color histogram mainly describes the proportion of different colors in the entire image, while not caring the spatial position of color. In this paper, we calculate the color histogram in RGB color space. In particular, the color histogram of an image should be a matrix with size of $N \times 256$, where $N = 3$ correspond to the three color channels (i.e., R, G and B), and 256 is consistent with the range of pixel values.

C-Net is designed based on the color histogram for color feature learning. The goal of C-Net is to obtain consistent color features with the normal-light image (see Figure 2). We also utilize the encoder-decoder pipeline for C-Net, which transforms the input low-light image to the predicted color histogram by the following formula:

$$C_{pre} = CNet(S_{low}), \qquad (4)$$

where C_{pre} denotes the obtained color histogram and $CNet$ denotes the calculation process of C-Net. To better reconstruct the color histogram, we also apply the l_1 loss to constrain C-Net, which can be described as follows:

$$l_c = \frac{1}{N \times 256} \|C_{pre} - C_{high}\|_1, \qquad (5)$$

where l_c is color histogram reconstruction loss and C_{high} is the real color histogram of normal-light image. Note that color histogram cannot describe the contents and details in images. That is, C-Net pays all attention to learning consistent color features, which is beneficial to enhancement.

R-Net. Based on the gray image and color histogram obtained by G-Net and C-Net, R-Net combines them to restore normal-light image collaboratively. The input low-light image, predicted gray image and color histogram are transformed into the normal-light image by R-Net as follows:

$$S_{pre} = RNet(S_{low}, G_{pre}, C_{pre}), \qquad (6)$$

where S_{pre} denotes the enhanced image. To reconstruct the normal-light image in pixel level, we use color image reconstruction loss l_r, which is defined as follows:

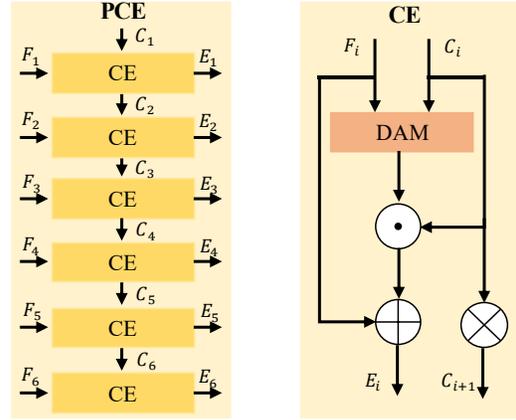$$l_r = \frac{1}{N \times H \times W} \|S_{pre} - S_{high}\|_1, \qquad (7)$$

Figure 3. The detailed structures of the pyramid color embedding (PCE) and color embedding (CE) modules, where $\odot$ denotes element-wise multiplication, $\oplus$ denotes element-wise addition and $\otimes$ denotes upsampling operation.

where N, H and W denote the channel number, height and width of the normal-light image S_{high}. In terms of structure level, we employ the ssim loss as the constraint:

$$l_{ssim} = 1 - SSIM(S_{pre}, S_{high}). \qquad (8)$$

where the similarity function $SSIM(\cdot)$ is described as

$$SSIM(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \cdot \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}, \qquad (9)$$

where $x, y \in \mathbb{R}^{H \times W \times 3}$ denote two images to be measured, $\mu_x, \mu_y \in \mathbb{R}$ represent the mean values of the two images, $\sigma_x, \sigma_y \in \mathbb{R}$ are the corresponding variances of the two images, c_1 and c_2 are two constant parameters which can prevent the denominator from being zero. In addition, the total variation loss l_{tv} is also employed as a regularization term to retain the smoothness for the enhanced image.

3.2. Pyramid Color Embedding (PCE)

The PCE module is designed to well embed the color information into R-Net, as shown in Figure 3. Clearly, PCE has six color embedding (CE) modules with pyramid structure. Specifically, CE achieves the dynamic embedding of color features. The main component of CE is dual affinity matrix (DMA) that solves the information mismatch issue.

Dual affinity matrix. From G-Net and C-Net, we can obtain the corresponding gray image and color histogram, which can provide rich structure and texture details, and color information respectively. R-Net applies both of them to achieve better enhancement. Since color histogram does not contain spatial information, simply concatenating them will cause inaccurate illumination in the enhanced image. Besides, the simple concatenation will also result in mismatch between color information and contents, which may produce color difference in the enhanced images.

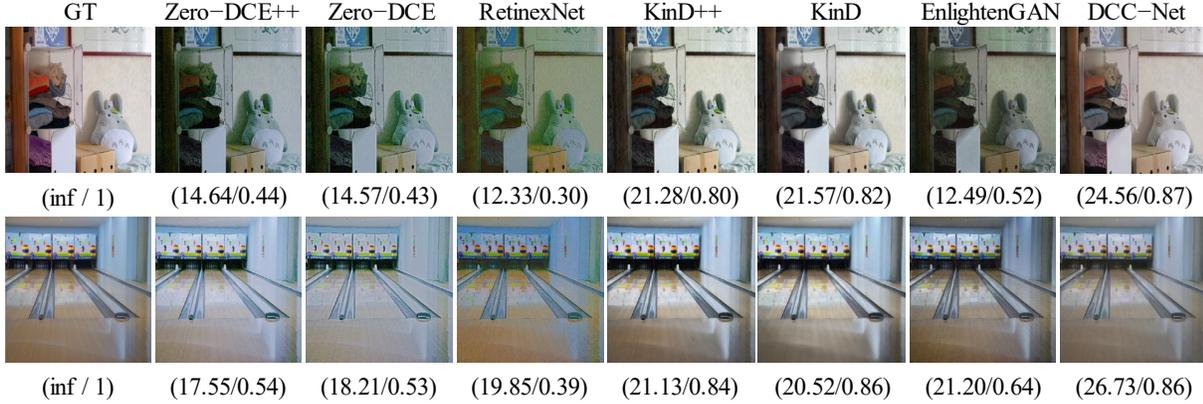| GT	Zero−DCE++	Zero−DCE	RetinexNet	KinD++	KinD	EnlightenGAN	DCC−Net
(inf / 1)	(14.64/0.44)	(14.57/0.43)	(12.33/0.30)	(21.28/0.80)	(21.57/0.82)	(12.49/0.52)	(24.56/0.87)
(inf / 1)	(17.55/0.54)	(18.21/0.53)	(19.85/0.39)	(21.13/0.84)	(20.52/0.86)	(21.20/0.64)	(26.73/0.86)

Figure 4. Visual comparison of deep LLIE methods on LOL dataset.

To solve the information mismatch issue and obtain better color information embedding, we present a new color embedding (CE) module, which can dynamically incorporate color features into R-Net according to the affinity between color and content features. The proposed dual affinity matrix (DAM) aims at computing the affinity matrix to match color and content features, and further prevent the enhanced image from producing inconsistent colors. Specifically, given color features C and content feature F with size of $N \times H \times W$, DAM first computes the Manhattan distance and inner product between C and F for each position, which are formulated as follows:

$$M(x,y) = -\|F(x,y) - C(x,y)\|_1, \qquad (10)$$

$$P(x,y) = F(x,y) \cdot C(x,y), \qquad (11)$$

where $F(x,y), C(x,y) \in \mathbb{R}^N$ denote the vectors of F and C for position (x,y), $M, P \in \mathbb{R}^{H \times W}$ are Manhattan distance matrix and inner product matrix. Then, the dual affinity matrix A can be calculated as follows:

$$A = 2 \times sigmoid(M) \odot tanh(P), \qquad (12)$$

where $tanh(\cdot)$ and $sigmoid(\cdot)$ are tanh function and sigmoid function respectively. Note that $M(x,y) \leq 0$ for each position (x,y), so that $sigmoid(M) \in [0, 0.5]$. Hence, we use $2 \times sigmoid(M)$ to ensure the range of $A \in [0, 1]$.

Color embedding. CE obtains the dynamic embedding of color information, whose structure is given in Figure 3. After obtaining the dual affinity matrix A, CE computes the element-wise multiplication of A and color features C. The weighted color features are summed with the content feature F to obtain the color information embedded features:

$$E = A \odot C + F, \qquad (13)$$

where E denotes the output features used in the decoder of R-Net. There is also an upsampling operation for color features C to change its resolution, and then further feed into the next CE as original color features.

Pyramid structure. Given color features, we can use them to guide the enhancement process to obtain consistent colors. To fully explore the color information, we present PCE including six CEs with pyramid structure (see Figure 3). Given the color features C_i and content feature F_i from i-th CEs, the features obtained by PCE in each layer from shallow to deep are described as follows:

$$E_i, C_{i+1} = CE(F_i, C_i), i = 1, 2, \cdots, 6, \qquad (14)$$

where E_i denotes the output features, $CE(\cdot)$ denotes the transformation of CE. C_i is computed by the $(i-1)$-th CE. In contrast, F_i is copied from the corresponding layer in the encoder of R-Net. The pyramid structure embeds color features into six layers. In other words, the progressive design can make full use of the color information. As a result, the enhanced image will be more consistent in colors.

3.3. Objective Function

The objective function of our DCC-Net is described as

$$l_{total} = \lambda_g l_g + \lambda_c l_c + \lambda_r l_r + \lambda_{ssim} l_{ssim} + \lambda_{tv} l_{tv} \quad (15)$$

where $\lambda_g, \lambda_c, \lambda_r, \lambda_{ssim}, \lambda_{tv}$ are several trade-off parameters. Specifically, l_g and l_c are used to recover the gray image and color histogram, respectively. l_r and l_{ssim} are utilized to reconstruct the norm-light image in pixel and structure level. l_{tv} can be regarded as an regularization term to prevent over-fitting and preserve smoothness.

4. Experiments

In this section, we evaluate the LLIE performance of our DCC-Net on several datasets, and describe the comparison results with some related deep LLIE methods.

4.1. Experimental Settings

Evaluated datasets. For training, we use LOL synthetic dataset and LOL real dataset [30]. Specifically, LOL synthetic dataset contains 1,000 paired synthetic low/normal

Table 1. Evaluation results in terms of PSNR, SSIM, MAE, CSE and inference time on the LOL dataset, where the red denotes best performance and the blue denotes the second best.

Method	PSNR	SSIM	MAE(%)	CSE (ratio)	Time(s)
RetinexNet	16.82	0.43	14.93	2.51	0.0390
KinD	20.42	0.82	9.82	2.30	0.0650
Zero-DCE	16.02	0.51	15.98	5.98	0.0026
EnlightenGAN	18.32	0.64	13.71	1.66	0.0150
Zero-DCE++	16.11	0.53	15.89	9.59	0.0012
KinD++	20.92	0.80	8.83	1.15	0.0320
DCC-Net	22.72	0.81	8.72	1.00	0.0260

light images. The LOL real dataset includes 485 paired real low/normal light images. For testing, we use LOL testing dataset [30] (15 paired images), DICM [16] (64 images), LIME [7] (10 images), MEF [22] (17 images), NPE [29] (85 images) and VV[1] (24 images) datasets.

Evaluation metrics. To evaluate the performance of different LLIE methods, we use both full-reference and non-reference image quality evaluation metrics. For LOL testing data with paired data, peak signal-to-noise ratio (PSNR), structural similarity (SSIM), mean absolute error (MAE), inference time and color-sensitive error (CSE) [37] are employed. For the DICM, LIME, MEF, NPE and VV datasets without paired data, only naturalness image quality evaluator (NIQE) is used, as there is no ground-truth. Note that CSE is a new metric that can measure the color difference of two images [37] . To be specific, the greater PSNR and SSIM, the better enhancement. In contrast, the smaller CSE, MAE and NIQE, the more realistic the refined images.

Compared methods. Since we mainly focus on deep LLIE methods, DCC-Net is compared with six state-of-the-art deep neural network-based LLIE methods:

- **RetinexNet** (BMVC'18) [30]: RetinexNet is a deep LLIE network incorporating the Retinex theory.
- **EnlightenGAN** (TIP'21) [11]: EnlightenGAN is an GAN-based unsupervised deep LLIE method.
- **KinD** (MM'19) [35]: KinD is a classic supervised deep LLIE method based on Retinex theory.
- **KinD++** (IJCV'21) [34]: KinD++ extends KinD by equipping with multi-scale illumination attention.
- **Zero-DCE** (CVPR'20) [6]: Zero-DEC develops a new zero-shot learning framework for deep LLIE.
- **Zero-DCE++** (TPAMI'21) [18]: Zero-DEC++ is a light version of Zero-DEC, which replaces the convolution layer with depth-wise separable convolution.

Implementation details. We conduct all experiments by the Pytorch [23] platform on Python environment with

---

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

two NVIDIA GeForce RTX 2080i GPUs. All training and testing images are resized into 512×512 pixels. Adam optimizer [14] is utilized with a batch size of 6. We train our DCC-Net for 400 epochs, where the learning rate is 0.0001 during the first 200 epochs and 0.00001 for the next 200 epochs. For the hyper-parameters of our DCC-Net, we empirically set $\lambda_g = 1$, $\lambda_c = 2$, $\lambda_r = 2$, $\lambda_{ssim} = 2$, $\lambda_{tv} = 0.1$.

4.2. Quantitative Enhancement Results

LOL dataset with paired data. We first evaluate each deep LLIE model on LOL dataset. The numerical results are described in Table 1. We can see that: (1) our DCC-Net obtains the greatest PSNR value and smallest MAE value, i.e., the enhanced results are more close to the ground-truth among all the compared methods; (2) for the SSIM metric, our DCC-Net is comparable to KinD, and superior to other methods, i.e., our DCC-Net can better restore the structures for LLIE; (3) compared with supervised methods, unsupervised methods are still weak in retaining the quality of enhanced images; (4) our DCC-Net obtains obvious improvement on the CSE metric compared with other deep LLIE methods, where "raio" indicates the ratio of other method's results to that of our DCC-Net. Since CSE can directly measure the color difference between two images, so the results can express that our DCC-Net is effective to preserve the color consistency; (5) The inference time of our DCC-Net is comparable to other methods. As such, in terms of enhancement performance and inference time, DCC-Net obtains the best results with relatively short inference time.

Datasets (DICM, LIME, MEF, NPEE and VV) without paired data. We also conduct experiments on unpaired and real low-light images. Table 2 displays the quantitative image quality of LLIE results in terms of NIQE metric. In general, DCC-Net obtains better NIQEs results among all methods. To be specific, KinD and KinD++ achieve preferable performance on MEF dataset, which is slightly better than ours. For other datasets, our DCC-Net is the best one.

4.3. Visual Image Analysis and Evaluations

LOL dataset with paired data. Figure 4 shows several enhanced images of LOL datset. It is clear that our DCC-Net can prevent the enhanced image from inaccurate colors. In most cases, there are obvious color difference with the ground-truth image in the enhanced images of compared methods. For the light refined images of KinD, KinD++, the results are usually over-enhanced. The resulted images of EnlightenGAN and our DCC-Net looks better. In terms of numerical PSNR/SSIM metrics, our proposed DCC-Net method achieves the best enhancement results.

Datasets without paired data. We further exhibit the visual enhancement results on the DICM, LIME, MEF, NPE and VV datasets in Figure 5-7. We can find that: (1) Zero-DCE++ and Zero-DCE tend to generate over-enhancement

Low   Zero−DCE++   Zero−DCE   RetinexNet   KinD++   KinD   EnlightenGAN   DCC−Net

Figure 5. Visual comparison of deep LLIE methods on DICM and LIME datasets, where the results in the first row are based on the LIME dataset, and the results in the second row are based on the DICM dataset.
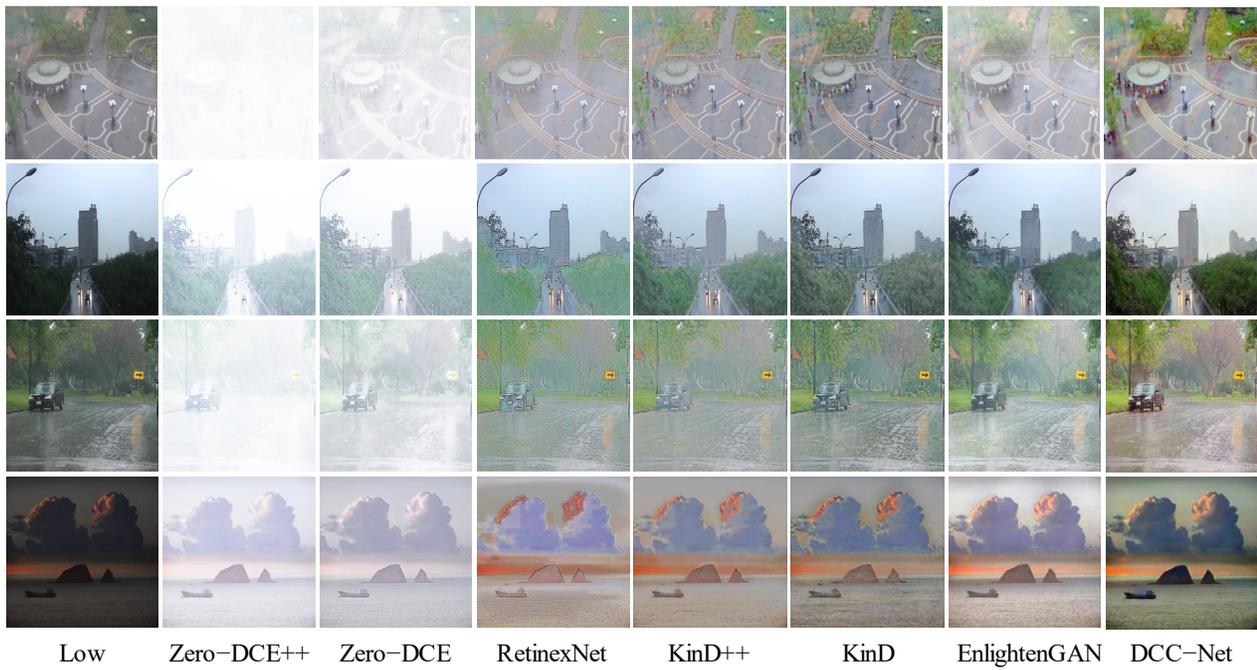

Low   Zero−DCE++   Zero−DCE   RetinexNet   KinD++   KinD   EnlightenGAN   DCC−Net

Figure 6. Visual comparison of deep LLIE methods on NPE dataset.

Low   Zero−DCE++   Zero−DCE   RetinexNet   KinD++   KinD   EnlightenGAN   DCC−Net

Figure 7. Visual comparison of deep LLIE methods on MEF and VV datasets, where the results in the first row are based on the MEF dataset, and the results in the second row are based on the VV dataset.

Table 2. Evaluation results in terms of NIQE on DICM, LIME, MEF, NPE and VV dataset, where the red color denotes best performance and the blue color denotes the second best.

Method	DICM	LIME	MEF	NPE	VV
RetinexNet	4.33	5.75	4.93	4.95	4.32
KinD	3.95	4.42	4.45	3.92	3.72
Zero-DCE	4.58	5.82	4.93	4.53	4.81
EnlightenGAN	4.06	4.59	4.70	3.99	4.04
Zero-DCE++	4.89	5.66	5.10	4.74	5.10
KinD++	3.89	4.90	4.55	3.91	3.82
DCC-Net	3.70	4.42	4.59	3.70	3.28

Table 3. LLIE results of our DCC-Net with different structures on LOL dataset, where the bold denotes the best.

Model	W/o G-Net	W/o C-Net	W/o PCE	DCC-Net
PSNR	21.51	21.01	21.14	**22.72**
SSIM	0.79	0.79	0.79	**0.81**
MAE(%)	10.27	10.13	10.43	**8.72**

images, which are full of white pixels and lose many details; (2) there is obvious color gap for the enhance results of RentinexNet, which makes them seem unreal; (3) for KinD, Kind++ and EnlightenGAN, the illumination improved images lack of naturalness; (4) in contrast, the enhanced images of our DCC-Net are more natural and colorful.

4.4. Ablation Study

We evaluate the effect of the network structure and PCE module on the performance of our DCC-Net.

Effectiveness of network structure. To demonstrate the effectiveness of the sub-networks G-Net and C-Net, we conduct the LLIE task on LOL dataset with and without them. Figure 8 displays the LLIE results of different models, where **W/o G-Net** and **W/o C-Net** represent our DCC-Net without G-Net and C-Net respectively. We find that unreasonable colors are produced by **W/o G-Net** and **W/o C-Net**. Table 3 describes the quantitative results. We see that there is obvious performance decline without G-Net or C-Net, which demonstrates the rationality and validity of the proposed "divide and conquer" collaborative strategy.

Effectiveness of PCE. As can be seen from Table 3, when PCE is removed from DCC-Net, denoted as **W/o PCE**, the values of PSNR and SSIM are smaller than DCC-Net, i.e., PCE is important to ensure the performance. Similarly, the value of MAE is greater than **W/o PCE**, which suggests that PCE is effective to enhance the illumination. Since PCE can effectively match the color and content features layer by layer, which can take full advantage of the color information. From the third row of Figure 8, we see that the enhanced image contains undesired yellow color,
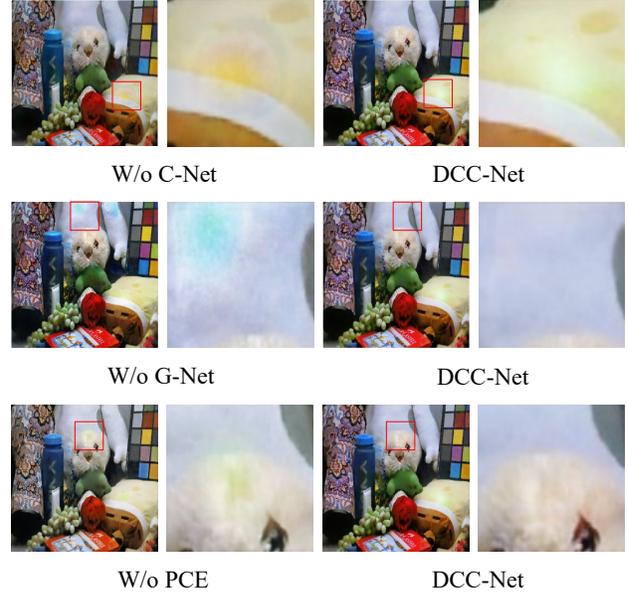

Figure 8. Comparison of different LLIE models on LOL dataset.

which is obviously inconsistent with peripheral regions.

5. Conclusion

We have discussed the issues of retaining the color consistency and naturalness for LLIE task. Technically, we proposed a new "divide and conquer" collaborative strategy to retain color information and naturalness, and developed a deep color consistent network called DCC-Net. To be specific, two sub-nets are designed to learn gray image and color histogram from a low-light image, where the gray image offers rich content information and the color histogram provides color information. Since color histogram does not consider spatial position, a new module PCE is further designed to match color and content features, and progressively embed color information. By the collaborative strategy, DCC-Net can jointly preserve color information and refine illumination. Extensive experiments show the superiority and effectiveness of DCC-Net for obtaining more natural and colorful normal-light images. In future, we will investigate more effective networks to further improve the naturalness and color consistency for LLIE. Besides, how to quantitatively assess the naturalness and image quality in terms of contents and color difference still remains an open problem, which is also an interesting future work.

6. Acknowledgments

References

[1] Turgay Celik and Tardi Tjahjadi. Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing*, 20(12):3431–3441, 2011. 2

[2] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018. 1

[3] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019. 1

[4] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3146–3154, 2019. 1

[5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 1

[6] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1780–1789, 2020. 3, 6

[7] Xiaojie Guo. Lime: a method for low-light image enhancement. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 87–91, 2016. 2, 6

[8] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 2

[9] Ju Han and Kai-Kuang Ma. Fuzzy color histogram and its use in color image retrieval. *IEEE Transactions on image Processing*, 11(8):944–952, 2002. 4

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[11] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021. 1, 3, 6

[12] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997. 2

[13] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *IEEE transactions on Image Processing*, 6(3):451–462, 1997. 2

[14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[15] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977. 2

[16] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation. In *2012 19th IEEE International Conference on Image Processing*, pages 965–968. IEEE, 2012. 6

[17] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, 22(12):5372–5384, 2013. 2

[18] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *arXiv preprint arXiv:2103.00860*, 2021. 3, 6

[19] Chongyi Li, Jichang Guo, Fatih Porikli, and Yanwei Pang. Lightennet: A convolutional neural network for weakly illuminated image enhancement. *Pattern recognition letters*, 104:15–22, 2018. 3

[20] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018. 2

[21] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 3

[22] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015. 6

[23] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32:8026–8037, 2019. 6

[24] Stephen M Pizer. Contrast-limited adaptive histogram equalization: Speed and effectiveness stephen m. pizer, r. eugene johnston, james p. ericksen, bonnie c. yankaskas, keith e. muller medical image display research group. In *Proceedings of the First Conference on Visualization in Biomedical Computing, Atlanta, Georgia*, volume 337, 1990. 2

[25] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28:91–99, 2015. 1

[26] Wenqi Ren, Sifei Liu, Lin Ma, Qianqian Xu, Xiangyu Xu, Xiaochun Cao, Junping Du, and Ming-Hsuan Yang. Low-light image enhancement via a deep hybrid network. *IEEE Transactions on Image Processing*, 28(9):4364–4375, 2019. 3

[27] Xutong Ren, Wenhan Yang, Wen-Huang Cheng, and Jiaying Liu. Lr3m: Robust low-light enhancement via low-rank regularized retinex model. *IEEE Transactions on Image Processing*, 29:5862–5876, 2020. 2

[28] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. Unet: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 3

[29] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013. 6

[30] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 1, 2, 5, 6

[31] Yanyan Wei, Zhao Zhang, Yang Wang, Mingliang Xu, Yi Yang, Shuicheng Yan, and Meng Wang. Deraincyclegan: Rain attentive cyclegan for single image deraining and rainmaking. *IEEE Transactions on Image Processing*, 30:4788–4801, 2021. 1

[32] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3063–3072, 2020. 3

[33] Lin Zhang, Lijun Zhang, Xiao Liu, Ying Shen, Shaoming Zhang, and Shengjie Zhao. Zero-shot restoration of back-lit images using deep internal learning. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1623–1631, 2019. 3

[34] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129(4):1013–1037, 2021. 3, 6

[35] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019. 3, 6

[36] Zhao Zhang, Yanyan Wei, Haijun Zhang, Yi Yang, Shuicheng Yan, and Meng Wang. Data-driven single image deraining: A comprehensive review and new perspectives. *TechRxiv Preprint*, pages 1–20, 2021. 1

[37] Suiyi Zhao, Zhao Zhang, Richang Hong, Mingliang Xu, Haijun Zhang, Meng Wang, and Shuicheng Yan. Unsupervised color retention network and new quantization metric for blind motion deblurring. *TechRxiv Preprint*, 2021. 6

[38] Anqi Zhu, Lin Zhang, Ying Shen, Yong Ma, Shengjie Zhao, and Yicong Zhou. Zero-shot restoration of underexposed images via robust retinex decomposition. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2020. 3