

Local Color Distributions Prior for Image Enhancement

Haoyuan Wang, Ke Xu*, and Rynson W.H. Lau*

Department of Computer Science, City University of Hong Kong



Fig. 1: Given an input image (a) with both over-exposure (background windows) and under-exposure (foreground persons), existing methods fail to handle both problems well. While (b) performs better on the background, the foreground is only slightly brightened. Although (c) performs better on the foreground, the background is still over-exposed. (d) slightly brightens the foreground but further over-exposed the background. In contrast, our method (e), which is based on learning local color distributions, can handle both problems well. The textures of the window curtains and the patterns of the clothes can both be seen clearly.

Abstract. Existing image enhancement methods are typically designed to address either the over- or under-exposure problem in the input image. When the illumination of the input image contains both over- and under-exposure problems, these existing methods may not work well. We observe from the image statistics that the local color distributions (LCDs) of an image suffering from both problems tend to vary across different regions of the image, depending on the local illuminations. Based on this observation, we propose in this paper to exploit these LCDs as a prior for locating and enhancing the two types of regions (*i.e.*, over-/under-exposed regions). First, we leverage the LCDs to represent these regions,

* Joint corresponding authors. This work was led by Rynson Lau.

and propose a novel local color distribution embedded (LCDE) module to formulate LCDs in multi-scales to model the correlations across different regions. Second, we propose a dual-illumination learning mechanism to enhance the two types of regions. Third, we construct a new dataset to facilitate the learning process, by following the camera image signal processing (ISP) pipeline to render standard RGB images with both under-/over-exposures from raw data. Extensive experiments demonstrate that the proposed method outperforms existing state-of-the-art methods quantitatively and qualitatively. Codes and dataset are in <https://hywang99.github.io/lcdpnet/>.

1 Introduction

When taking photos, the illumination condition of a scene may not always be ideal, and the photos may suffer from under-exposure (due to low-light/back-light) or over-exposure (due to some intense lights inside the image region). Often, both over- and under-exposures may occur together in the same image due to unbalanced lighting conditions. The illumination may change significantly, burying the local image contents in *both* over- and under-exposed regions, as shown in Fig. 1(a). While photography experts may leverage high-end DSLR cameras and carefully tune them (*e.g.*, the aperture, ISO, and special filters) to alleviate the problem, it requires photography expertise and expensive equipment.

Many methods have been proposed to enhance the quality of images that are captured with poor illumination conditions. A line of methods focus on enhancing the under-exposed images captured in low-light scenes via the Retinex based approach [33], bilateral learning [13], generative adversarial learning [11], deep parametric filtering [23], and self-supervised [14] or semi-supervised learning [39]. Other works [4,1] try to enhance the over-exposed or under-exposed images in one network. All these methods typically assume the scene illumination to be generally uniform, such that an improper exposure would result in either over- or under-exposure. Hence, they tend to adjust the image intensity globally, *i.e.*, either increasing or decreasing the intensity. However, if the illumination of a scene is non-uniform, causing the input image to suffer from both over- and under-exposures as shown in Fig. 1(a), existing methods may not work well. For example, ZeroDCE [14] (c) and RUAS [29] (d) worsen the over-exposure problem at the background regions as they try to enhance the under-exposed foreground persons. On the other hand, while MSEC [1] (b) enhances the foreground slightly, it produces some color distortions around the background windows.

In this paper, we aim to address both over- and under-exposure problems appearing in a single image. The key challenge is how to effectively separate these two types of regions and recover their local illuminations accordingly. We observe that the local color distributions (LCDs), which consist of regional local histogram vectors, can be a reliable prior to address this challenge for two reasons. First, we note that these LCDs in the over- and under-exposed regions show significant divergences and deviate from those in the properly exposed regions. Hence, they can be used to help identify and separate different types of regions.

Second, while it may not be reliable to directly infer the true scene lighting from the image with both over- and under-exposures, colors are important cues that are readily available in the image. Modeling the LCDs essentially help estimate the proper local illumination and recover the buried contents. Based on our observation, we propose a novel neural network to jointly tackle the two problems (*i.e.*, both over- and under-exposure problems) with two novel modules: (1) the local color distribution embedded (LCDE) module to formulate LCDs in multi-scales to learn the representations of over- and under-exposed regions as well as their correlations, and (2) the dual-illumination learning mechanism to constrain the learning of the LCDE module, estimate and combine an over-illumination map and an under-illumination map for enhancing the image. In addition, as existing datasets (*e.g.*, [1,10,33,5]) contain mostly images with either over- or under-exposure, we further construct a new image dataset containing $\sim 1,700$ diverse scenes with both over- and under-exposures, to facilitate training and evaluation. We follow the camera ISP pipeline by applying linear transform functions and clipping on the pixel intensity of the input raw images to render sRGB images. As shown in Fig. 1(e), our proposed method based on the LCD prior brightens the under-exposed regions while darkening the over-exposed regions, allowing the textures of the over-exposed window curtains to be revealed as well as the patterns of the under-exposed sweaters that the two persons are wearing to become visible.

Our main contributions of this work can be summarized as follows:

- We propose to exploit the local color distributions (LCDs) to jointly address both over- and under-exposure problems in the input image, and a neural network to leverage the LCDs for locating and enhancing over-/under-exposed regions of the image.
- We propose the LCDE module to formulate multi-scale LCDs in order to learn the representations of over- and under-exposed regions as well as their correlations to the global illumination. We also propose a dual-illumination estimator to combine both over- and under-illumination maps to enhance the input image.
- We construct a new paired dataset consisting of over 1700 images of diverse, non-uniformly illuminated scenes to facilitate the learning process.
- Extensive experiments demonstrate that the proposed method outperforms state-of-the-art methods qualitatively and quantitatively on the popular MSEC [1] and our datasets.

2 Related Work

Image-to-Image Translation-Based Methods. A line of methods enhance under-exposed images by learning different image-to-image translation mappings. Histogram equalization [26] and gamma correction are the most representative methods. Some methods propose to combine global and local contrast enhancement operators with semantic region detection (*e.g.*, face, building and

sky) [20], regional templates [16] or contrast statistics along image boundaries and in textured regions [30].

Recent deep-learning based methods typically learn the mapping functions using high-quality retouched images or images taken using high-end cameras, with bilateral learning [13], intermediate HDR supervision [40], multi-stage restoration [42,9], generative adversarial learning [17,11,19,27], or reinforcement learning [24,41]. Cai *et al.* [4] propose to enhance an under-exposed image by separately modeling the illuminance and detail layers from multiple exposure images. Moran *et al.* [32] propose to learn a set of piece-wise linear scaling curves and apply them in different color spaces for under-exposed image enhancement. Xu *et al.* [37] propose to enhance under-exposed images based on frequency decomposition. In [23], different kinds of local parametric filters are learned for image enhancement. Mahmoud *et al.* [1] propose a coarse-to-fine network to learn color and detail enhancement for addressing either over- or under-exposure.

Retinex-Based Methods. Another line of works are Retinex-based image enhancement methods [15,12,44,3,33,45]. They first decompose the input image into illumination and reflectance layers, and then enhance the illumination layer of the image. Conventional optimization based methods [15,12,3] propose different hand-craft priors for constraining the illumination or reflectance layers. Deep learning based methods [10,33,45,29] learn such intrinsic decomposition from a large amount of data. For example, DeepUPE [33] propose to directly estimate the proper illumination layer from the input under-exposed image via bilateral learning [6]. Most recently, Liu *et al.* [29] propose an architecture search based method to leverage cooperative priors for under-exposed image enhancement.

Limitation of Existing Works. All aforementioned methods typically assume that only either over- or under-exposure problem would appear in a single image. Hence, they tend to brighten the under-exposed images or darken the over-exposed images, but lack the capability to tackle images with both over- and under-exposures. In this paper, we aim to tackle this common shortfall. We first construct a new dataset, and propose a LCD prior guided deep-learning approach for enhancing images containing both over- and under-exposures.

3 Proposed Dataset

To study our problem, we first revisit a recently published large-scale image enhancement dataset, the Multi-Scale Exposure Correction (MSEC) dataset [1]. Although it contains images with different levels of over- or under-exposures, each of the images is either over- or under-exposed. We visualize the luminance intensity mapping between input and reference images in Fig. 2. Fig. 2(a) shows that the shapes of the luminance mappings for individual images from the MSEC dataset [1] are either concave or convex. These luminance mappings do not show any non-uniform illumination exhibited from images that suffer from both over- and under-exposures. Statistics shown in Fig. 2(d) demonstrate that the MSEC model [1] trained on this dataset either brightens or darkens the input image

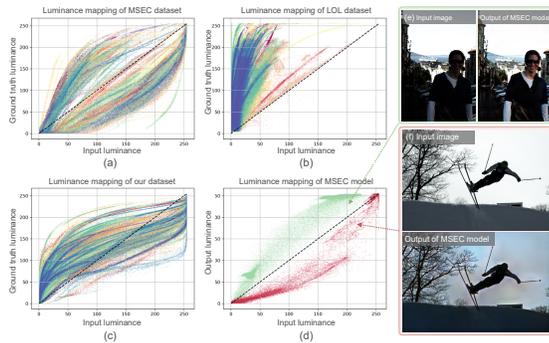


Fig. 2: The input-ground truth luminance mapping curve of (a) MSEC dataset [1], (b) LOL dataset [10], and (c) our dataset. Each cluster represents an image. For a single image, both (a) and (b) contains a single mapping of either brightening or darkening. (d) is the input-output mapping learned by the model trained on the MSEC dataset when given (e, f) as the input.

(see Fig. 2(e,f) for illustration). Fig. 2(b) shows the statistics of the luminance mappings for the images from the LOL dataset [10], which is a popular low-light image enhancement dataset. We can see that it only allows the learning of under-exposure enhancement. This means that methods trained on the LOL dataset [10] would typically brighten the input images. This demonstrates that a new dataset containing both over- and under-exposures in individual images is desired.

We construct our dataset from the raw images in the MIT Adobe5k dataset [2], which contains 5000 raw and expert-retouched sRGB image pairs for learning the tone mapping process. Since raw images have higher dynamic ranges to preserve scene information than sRGB images and their intensities are linearly proportional to the scene radiance, we generate our input images of both over- and under-exposures from the raw images. We use the expert-retouched sRGB images as our ground truth images. However, we note that not all raw images in this dataset are suitable for generating our images with both over- and under-exposures (*e.g.*, synthesizing over-exposure from a very dark image would likely produce an unrealistic image), and many expert-retouched images contain very dark/bright regions that cannot be used to form our ground truth images.

In order to generate high-quality learning pairs, we formulate the dataset generation pipeline in three steps as: (1) We manually go through all the image pairs in Adobe5k, and remove those pairs whose expert-retouched images contain very dark or very bright regions. (2) For each candidate raw image from (1), we follow the camera ISP pipeline to render an sRGB image with both over- and under-exposures, by adjusting the exposure level with a linear transformation function. (3) For each rendered sRGB image from (2), we ask a volunteer (who is a photographer) to help assess its quality. If the volunteer points out that the

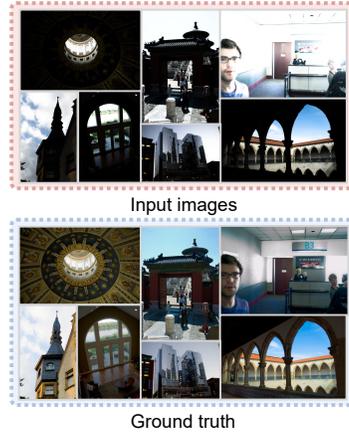


Fig. 3: Some input-ground truth pairs in our dataset. Our dataset contains images of over 1700 scenes.

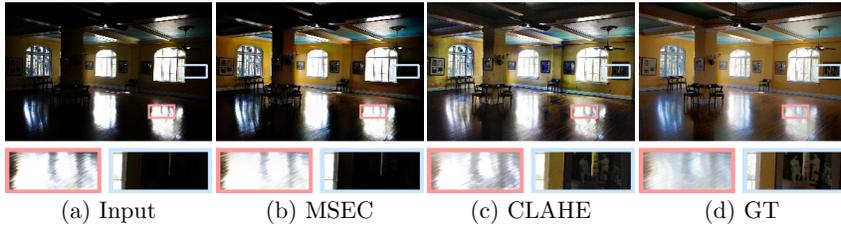


Fig. 4: MSEC [1] (b) enhances the input image (a) with a better overall visual quality, while CLAHE [28] (c) produces more faithful details in the local regions.

rendered image is not realistic or has obvious artifacts (*e.g.*, color bleeding), we feed it back to step (2) to produce another sRGB image for re-assessment. This iterates until we have a good quality image, or after five iterations and we simply discard this pair.

The linear mapping function used in step (2) is:

$$I'(i, j) = \phi[k(I(i, j) - 0.5) + 0.5], \quad (1)$$

where I is the input raw image whose pixel values range between $[0, 1]$. i, j are the spatial position in the image, and $\phi[\cdot]$ is the clipping operator to drop the overflow value with the upper/lower bounds being $1/0$. k is the slope value that represents the scaling factor of the exposure level. We manipulate k to generate images with over- and under-exposures. We apply Eq. 1 on all three channels of the input raw image equally to obtain the final rendered sRGB image. In total, we generate 1,733 pairs of images, which are split into 1,415 for training, 100 for validation, and 218 for testing. Some sample pairs are shown in Fig. 3.

4 Proposed Method

The over- and under-exposed image enhancement task could be formulated as seeking a mapping function F , which maps an 8-bit per channel sRGB image I_x to an enhanced image I_y such that $I_y = F(I_x)$. Instead of directly learning the image-to-image translation model or Retinex-based image-to-illumination mapping model, we propose to learn a *region-aware illumination mapping model* by learning to exploit the multi-scale LCDs and dual-illumination estimation.

4.1 Local Color Distribution (LCD) Pyramid

Directly inferring the true scene lighting from a single image is challenging, especially when the input image contains both over- and under exposures. Instead, as the color is a key component of scene lighting preserved in the image, we can leverage the local color information to help build representations of regional illuminations. Fig. 4(c) shows an example, where the color histogram based method CLAHE [28] recovers more details than the latest deep method MSEC [1] in both

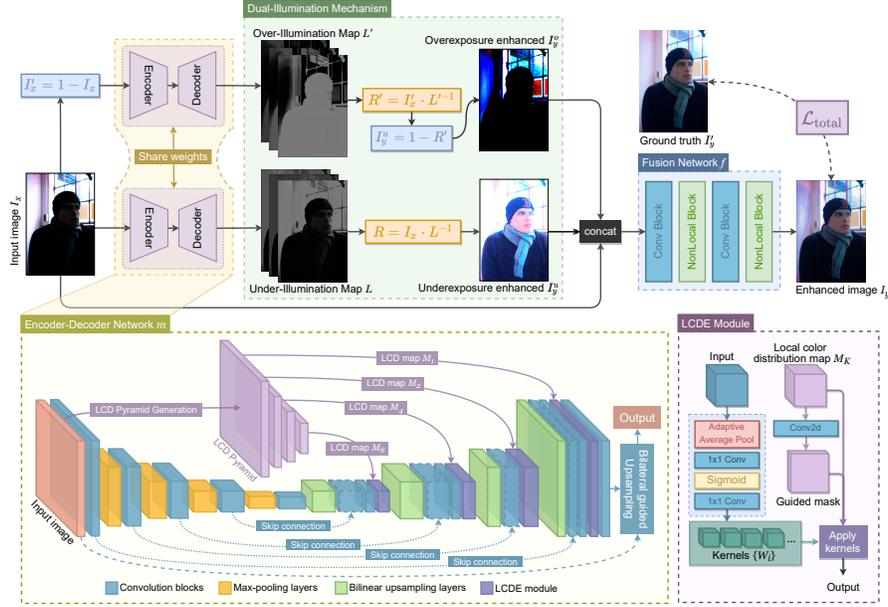


Fig. 5: Overview of our proposed network. It leverages the LCD pyramid with an encoder-decoder architecture for detecting the regions with problematic exposures implicitly, and the for enhancement of the over- and under-exposed regions.

over- and under-exposed regions. However, CLAHE [28] tends to produce images of inconsistent colors, due to its lack of global information. We model LCDs based on CLAHE [28], and extend it to be a multi-scale pyramid in our neural network, in order to tackle the inconsistency by learning the local-to-global illumination correlations.

Building the LCD Pyramid. Given an input image I_x of size $h \times w$, whose pixel values range between $[0, 1]$, we split I_x into $N = \lceil h/K \rceil \cdot \lceil w/K \rceil$ patches, where $\lceil \cdot \rceil$ is the closest-integer operator. We define a LCD as the color histogram within a local patch of size $K \times K$. We then use the 4D LCD map M_K to represent the distribution of scale K . We first build a $h \times w$ bilateral grid Γ [7] by splatting the pixel histogram voting along the range dimension, and then we compute M_k using:

$$M_k\left(\left\lceil \frac{i}{K} \right\rceil, \left\lceil \frac{j}{K} \right\rceil, c, b\right) = \frac{1}{K^2} \sum_{p, q \in \Omega_K(i, j)} \Gamma(p, q, c, b), \quad (2)$$

where i, j, c are the horizontal, vertical, and channel indices of the image, respectively. b is the index to the histogram bins, which can be computed by $b = \lfloor I_x(i, j, c) \cdot B \rfloor$. $\Omega_K(i, j)$ returns the indices of the pixels in the $K \times K$ patch that pixel (i, j) belongs to.

By assigning K with different values, we obtain LCD maps at different scales, e.g., M_1 is the pixel-wise color distribution, where the local histogram vector in M_1 is a sparse one-hot vector. When K increases, the locality representation of

M_K grows correspondingly. Let K take its value from $\{2^l, l \in \mathbb{N}^+\}$. $\mathbf{M} = \{M_K\}$ is then a multi-scale LCD pyramid with different levels of color distribution maps. A LCD pyramid contains regional illumination distributions in multi-scale, as visualized in Fig. 6. It can help differentiate over-/under-exposed regions.

4.2 Proposed Network

Fig. 5 shows the overview of the proposed network. It has an encoder-decoder [31] architecture, incorporating the LCDE module to leverage the LCD pyramid for learning the representations of over- and under-exposed regions and the dual-illumination mechanism for adaptive enhancement.

LCDE module. In order to learn adaptive representations for over- and under-exposed regions, we propose the LCDE module to predict the adaptive convolutional kernels with the guidance of the LCD pyramid. Our design of the LCDE module is based on DRconv [8] (originally proposed for high-level image classification and detection tasks). We choose DRconv [8] to exploit its idea of producing region-wise kernels. Unlike DRconv [8] that learn different kernels according to the local semantics, which are not reliable in the over- and under-exposed regions, we learn different kernels according to the local illuminations guided by the local color distributions.

Specifically, the LCDE module has two branches, *i.e.*, the convolution-kernels-generation branch and the guided-mask-prediction branch (bottom right part of Fig. 5). The first branch takes as input the precursor feature map and produces the parameters of n kernels $\{W_1, W_2, \dots, W_n\}$. The guided-mask-prediction branch inputs the LCD map M_K , and uses it to guide the prediction of a multi-value mask for dividing the spatial feature maps into n regions in order to apply different convolution kernels on different regions. By applying a different convolution kernel on a different region, the multi-scale LCD pyramid guides the network to differentiate regions of different exposures and enhances them separately.

To handle high resolution inputs, we construct this module upon the bilateral upsampling method [6,7] to achieve a fast inference speed.

Dual-illumination Estimation. To constrain the model to learn the exposure-aware masks, we exploit the Retinex theory in our model. The Retinex based methods (*e.g.*, [33]) typically decompose the input image I_x into an illumination map L and a reflectance map R . By considering the reflectance map R as the enhanced image, DeepUPE [33] generates the under-exposure enhanced result I_y^u with $I_y^u = R = I_x \cdot L^{-1}$. Since the values in L fall within $[0, 1]$, the pixel values in the result image are always larger than the those of the input image. Hence, these methods cannot suppress over-exposure. To address this problem, we propose to extend the illumination prediction mechanism of [33] to dual-illumination prediction mechanism by incorporating dual-path learning in [43]. The main idea here is that over-exposures in I_x could be regarded as under-exposures in the reverse image of I_x . By first computing the reverse image I_x' via $I_x' = 1 - I_x$, we can then compute the over-illumination map L' of the input image I_x in addition to the under-illumination map L . Thus, the over-exposure enhanced images I_y^o could be obtained by $I_y^o = 1 - R' = 1 - I_x' \cdot L'^{-1}$, where L' is

the over-illumination map of I_x . L and L' are estimated via the encoder-decoder network with the LCDE module. The two enhanced components are then fused by our fusion network to generate the final result.

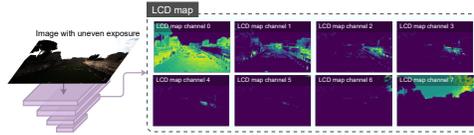


Fig. 6: The visualization of a LCD pyramid layer, taking an over-/under-exposed image as input. Over-/under-exposed regions are implicitly separated along the channel dimension.

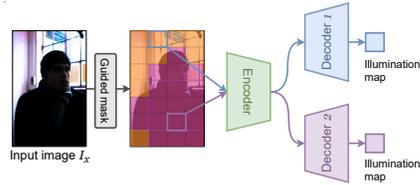


Fig. 7: Inferring the illumination maps helps constrain the learning of the guided mask. Our model implicitly assigns one decoder to each region type for adaptive enhancement.

Fusion Network. The fusion network f takes the two separately enhanced images, I_y^u and I_y^o , and the original image I_x as inputs to regress the final enhanced image. It contains two convolution layers and two non-local [34] blocks. With the non-local blocks, the network is able to capture long-range spatial correlations across pixels. The fusion network fuses the three inputs by predicting a 3-channel weight map to generate the enhanced result I_y .

Given the input image I_x , the whole process of our method to produce I_y is:

$$I_y = f \left(\frac{I_x}{m(I_x)}, 1 - \frac{1 - I_x}{m(1 - I_x)}, I_x \right), \quad (3)$$

where m represents the encoder-decoder network and f is the fusion network.

Why would our model work? Since the LCDE module applies n different kernels on n regions, each region can be regarded as being assigned with an individual decoder to learn the luminance mapping but without introducing extra computational cost, while all regions share the same encoder for feature extraction (see Fig. 7). In addition, by inferring the illumination maps, it helps constrain the learning of the guided mask in the LCDE module to focus on the exposure levels of these regions.

4.3 Loss Function

We adopt four loss terms to train our model. We apply the widely used MSE term \mathcal{L}_{mse} to measure the intensity reconstruction errors. To correct the color distortions in the over- and under-exposed regions, we apply the cosine similarity term \mathcal{L}_{cos} [33,35,36], which measures the color similarity of the reconstructed image and its ground truth in the sRGB color space. In addition, in order to guide the network to estimate the illumination maps, we apply the local smoothness term [33,38] to our dual-illumination estimation process, denoted as \mathcal{L}_{tv1}



Fig. 8: Visual comparison of over-/under-exposed images from our dataset. Our model reconstructs the details in the over-exposed regions (sky and tower) as well as the under-exposed regions (wall and door).

and \mathcal{L}_{tv2} , respectively. The local smoothness term aims to preserve the local smoothness characteristics of image illuminations by minimizing their gradient variations. The overall function can be written as:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{mse} + \lambda_2 \mathcal{L}_{cos} + \lambda_3 \mathcal{L}_{tv1} + \lambda_4 \mathcal{L}_{tv2}, \quad (4)$$

where λ_1 , λ_2 , λ_3 and λ_4 are the balancing hyper-parameters. Refer to the Supplemental for more details on the loss functions.

5 Experiments

5.1 Implementation Details

We implement our model using PyTorch [25]. All our experiments are conducted on a single NVIDIA GTX3080 GPU. The parameters of the network are optimized by the ADAM optimizer [21] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and the learning rate is $1e^{-4}$. The desired region number is set to $n = 2$, indicating over- and under-exposed regions. The weights for the terms in the loss function in Eq. 4

are $\lambda_1 = 1.0, \lambda_2 = 0.5, \lambda_3 = \lambda_4 = 0.01$. We utilize a 4-scale LCD pyramid and a 4-scale encoder-decoder network in our experiments. We implement the splatting operation in I in Eq. 2 using the soft-histogram-voting in [22]. During training, we resize the input images to 512×512 and apply random horizontal and vertical flippings to augment the input data. The model reported in the experiments is implemented based on a plain encoder-decoder network with 218K parameters.

Method	PSNR \uparrow	SSIM \uparrow
HE [26]	16.525	0.696
CLAHE [28]	15.383	0.599
DSLR [18] (Sony)	18.020	0.683
DSLR [18] (BlackBerry)	17.606	0.653
DSLR [18] (iPhone)	15.907	0.622
RetinexNet [10]	11.135	0.605
DeepUPE [33]	13.689	0.632
ZeroDCE [14]	12.058	0.544
MSEC [1]	20.205	0.769
Ours	22.295	0.855

Table 1: Quantitative comparison on the MSEC [1] test set. Best performances are marked in **bold**.

Method	PSNR \uparrow	SSIM \uparrow	Method	PSNR \uparrow	SSIM \uparrow
HE[26]	15.975	0.684	DSLR _{Sony} [18]	16.991	0.672
CIAHE[28]	16.327	0.642	DSLR _{BlackBerry} [18]	17.215	0.693
LIME[15]	17.335	0.686	DSLR _{iPhone} [18]	18.560	0.712
RetinexNet [10]	16.200	0.630	DSLR*[18]	20.856	0.758
RetinexNet*[10]	19.250	0.704	DeepUPE*[33]	20.970	0.818
MSEC*[1]	20.377	0.779	RUAS[29]	13.927	0.634
MSEC[1]	17.066	0.642	RUAS*[29]	13.757	0.606
ZeroDCE*[14]	12.587	0.653	HDRnet*[13]	21.834	0.818
Ours	23.239	0.842			

Table 2: Quantitative comparison on the proposed test set. * indicates that the model is re-trained on our proposed training set. Best performances are marked in **bold**.

5.2 Comparisons with State-of-the-art Methods

To verify the effectiveness of our method, we compare our model with the existing exposure correction and image enhancement methods. We select three conventional enhancement methods, including histogram equalization (HE), CLAHE [28] and LIME [15], and seven deep-learning-based methods: ZeroDCE [14], RetinexNet [10], MSEC [1], DSLR [18], HDRnet [13], DeepUPE [33] and RUAS [29]. We adopt the commonly used Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) as our evaluation metrics.

Quantitative Comparison. Table 1 reports the performance evaluation on the MSEC [1] test set. Both MSEC [1] and ours are trained on the MSEC [1] training set, while other numbers are copied from MSEC [1]. From the table, we can see that our method outperforms the second best method (*i.e.*, MSEC [1]) by a large margin. Table 2 further shows the comparison on our proposed test dataset. For a fair comparison, we report the performances of existing deep learning based methods using their pre-trained models as well as the models retrained on our training set. The results show that our method outperforms all existing methods on both PSNR and SSIM metrics.

Visual Comparisons. We visually compare the results of our method with those of the existing methods. Note that all models used in this experiment are retrained on our dataset for a fair comparison. Fig. 8 shows two examples containing both over- and under-exposed regions from our dataset, and results of existing methods and ours. We can see that our method can correct both



Fig. 9: Visual comparison of an over-exposed image from our dataset. Our result has the best visual quality.

Method	Ours _{plain}	Ours _{single}	Ours _{DRconv}	Ours _{mse}	Ours _{mse+tv}	Ours
PSNR \uparrow	21.878	22.198	21.001	21.261	22.421	23.239
SSIM \uparrow	0.783	0.840	0.785	0.793	0.815	0.842

Table 3: Ablation study.

over- and under-exposed regions, and produce more visually pleasing details and colors. We further show comparisons on images with large amount of over-exposed pixels (Fig. 9) or under-exposed pixels (Fig. 10) from our dataset, an image from the MSEC [1] dataset (Fig. 11), and two images from the Internet (Fig. 12). These comparisons generally demonstrate that our method generalizes well to different exposure levels and varying illumination conditions. Refer to the supplemental for more visual comparisons.

5.3 Internal Analysis

Ablation study. In order to analyze the effectiveness of our proposed module and pipeline, we perform ablation experiments on the network structure. Specifically, we train four different models: **(1)** A plain encoder-decoder (Ours_{plain}), **(2)** Add LCDE modules to the decoder of (1) (Ours_{single}), **(3)** Add the DRconv blocks [8] to the decoder of (1) (Ours_{DRconv}), **(4)** Add dual-illumination estimation to (2) (Ours).

As shown in Table 3, the dual-illumination estimation or the LCDE module significantly increase the performance, which verifies the effectiveness of the dual-illumination learning and the local color distribution prior. The comparison between (2) and (3) shows that the performance gain of our model is mainly

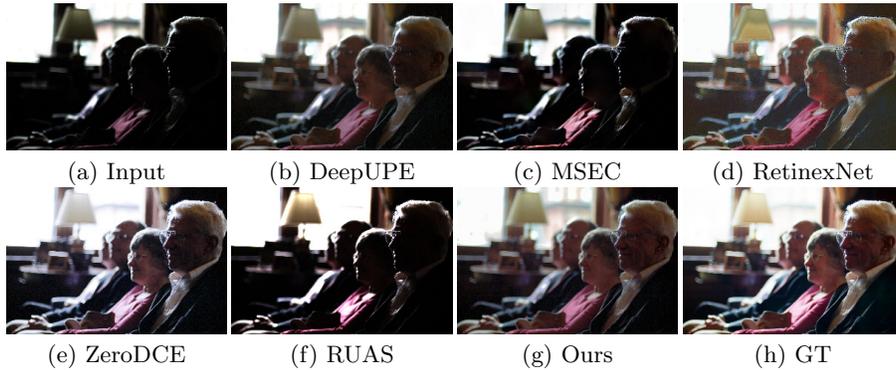


Fig. 10: Visual comparison of an under-exposed image from our dataset. Our result has the best visual quality.

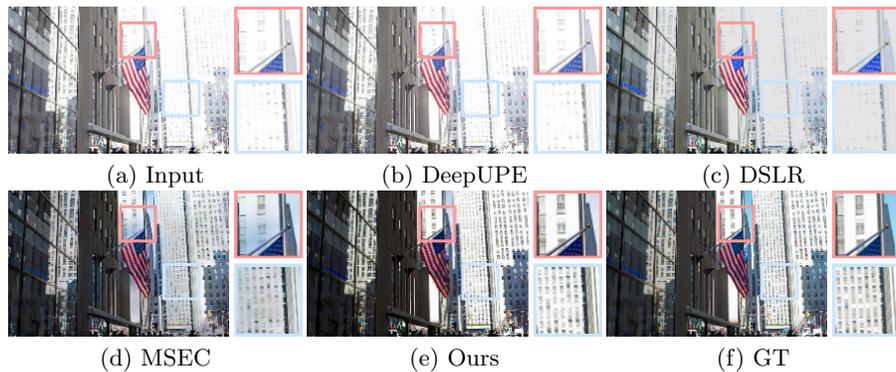


Fig. 11: Visual comparison of an over-exposed image from the MSEC dataset [1]. Our result has the best visual quality and details.

due to the LCD pyramid prior instead of regional dynamic convolution. We also perform the ablation study of loss function terms in Table 3. By comparing the last three columns, we can see that gradually incorporating the local smoothness term \mathcal{L}_{tv} and the cosine similarity term \mathcal{L}_{cos} consistently improves the enhancement performance.

Visualization and interpretation. We visualize the intermediate results of our network to examine whether the network has the region-aware capability. Fig. 13 shows the predicted multi-scale guided mask features by the LCDE module. We expect that under the guidance of LCD pyramid, the input images and feature maps could be divided into different regions by the guided mask according to the exposure level. As shown in Fig. 13, the intermediate guided masks of the LCDE module generally divide the pixels into the bright region and dark region when $n = 2$. This indicates that under the guidance of the LCDE module, the network does learn the region-aware adaptive enhancement for input images.

Limitations. Fig. 14 shows two challenging cases that our method may fail to enhance. If an image contains a large under-exposed region (Fig. 14(a)) or a large

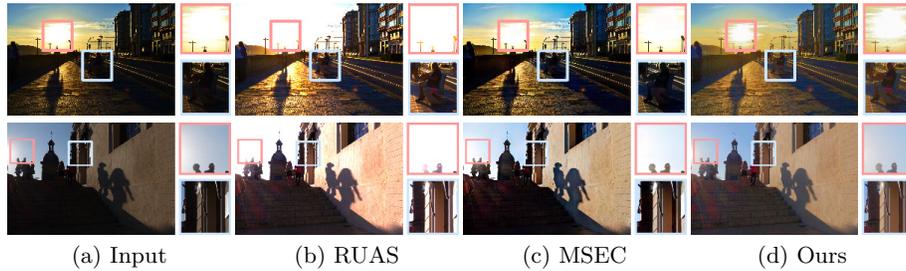


Fig. 12: Visual comparison on over-/under-exposed images from the Internet. Our result has the best visual quality.

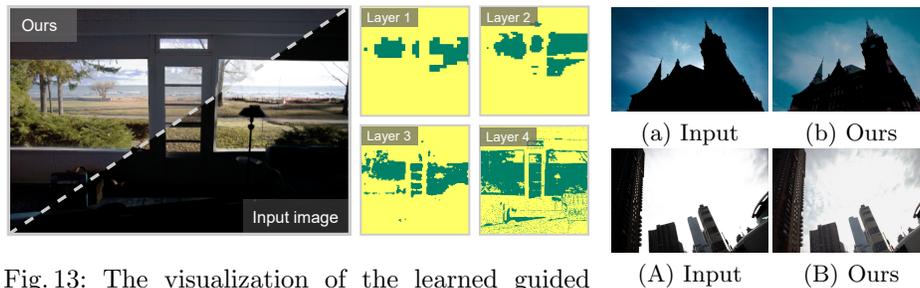


Fig. 13: The visualization of the learned guided mask in multi-scales intermediate LCDE module layers. With the guidance of the LCD pyramid and the constraint of the dual-illumination map, the model learns the guided mask adaptively.

Fig. 14: Failure cases. Our method may fail to enhance images with a large region of under-exposed pixels (building in (b)) or over-exposed pixels (sky in (B)).

over-exposed region (Fig. 14(A)), it may be difficult for our model to enhance the region. As a future work, we would like to explore semantic scene layouts as well as image inpainting techniques to handle this situation.

6 Conclusion

In this paper, we have tackled the image enhancement problem of correcting images with both over- and under-exposed regions. We have proposed a new dataset and designed a new end-to-end model to address the problem. We propose the LCDE module to detect over- and under-exposed regions under the guidance of the local color distributions. We extend the Retinex theory based illumination by proposing the dual-illumination estimator for better detail reconstruction. Extensive experiments show that our method performs favorably against state-of-the-art methods.

7 Acknowledgments

This project is in part supported by a General Research Fund from RGC of Hong Kong (RGC Ref.: 11205620).

References

1. Affi, M., Derpanis, K.G., Ommer, B., Brown, M.S.: Learning multi-scale photo exposure correction. In: CVPR (2021)
2. Bychkovsky, V., Paris, S., Chan, E., Durand, F.: Learning photographic global tonal adjustment with a database of input / output image pairs. In: CVPR (2011)
3. Cai, B., Xu, X., Guo, K., Jia, K., Hu, B., Tao, D.: A joint intrinsic-extrinsic prior model for retinex. In: ICCV (2017)
4. Cai, J., Gu, S., Zhang, L.: Learning a deep single image contrast enhancer from multi-exposure images. IEEE TIP (2018)
5. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: CVPR (2018)
6. Chen, J., Adams, A., Wadhwa, N., Hasinoff, S.: Bilateral guided upsampling. ACM TOG (2016)
7. Chen, J., Paris, S., Durand, F.: Real-time edge-aware image processing with the bilateral grid. ACM TOG (2007)
8. Chen, J., Wang, X., Guo, Z., Zhang, X., Sun, J.: Dynamic region-aware convolution. In: CVPR (2021)
9. Chen, L., Lu, X., Zhang, J., Chu, X., Chen, C.: Hinet: Half instance normalization network for image restoration. In: CVPR Workshops (2021)
10. Chen, W., Wenjing, W., Wenhan, Y., Jiaying, L.: Deep retinex decomposition for low-light enhancement. In: BMVC (2018)
11. Chen, Y., Wang, Y., Kao, M., Chuang, Y.: Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In: CVPR (2018)
12. Fu, X., Zeng, D., Huang, Y., Zhang, X., Ding, X.: A weighted variational model for simultaneous reflectance and illumination estimation. In: CVPR (2016)
13. Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F.: Deep bilateral learning for real-time image enhancement. ACM TOG (2017)
14. Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: CVPR (2020)
15. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. IEEE TIP (2017)
16. Hwang, S.J., Kapoor, A., Kang, S.B.: Context-based automatic local image enhancement. In: ECCV (2012)
17. Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., Van Gool, L.: DSLR-quality photos on mobile devices with deep convolutional networks. In: ICCV (2017)
18. Ignatov, A., Kobyshev, N., Timofte, R., Vanhoey, K., Van Gool, L.: Dslr-quality photos on mobile devices with deep convolutional networks. In: ICCV (2017)
19. Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z.: Enlightengan: Deep light enhancement without paired supervision. IEEE TIP (2021)
20. Kaufman, L., Lischinski, D., Werman, M.: Content-aware automatic photo enhancement. Computer Graphics Forum (2012)
21. Kingma, P., Ba, J.: Adam: A method for stochastic optimization. arXiv:1412.6980 (2014)
22. Liu, Y.L., Lai, W.S., Chen, Y.S., Kao, Y.L., Yang, M.H., Chuang, Y.Y., Huang, J.B.: Single-image hdr reconstruction by learning to reverse the camera pipeline. In: CVPR (2020)
23. Moran, S., Marza, P., McDonagh, S., Parisot, S., Slabaugh, G.G.: Deeplpf: Deep local parametric filters for image enhancement. In: CVPR (2020)

24. Park, J., Lee, J.Y., Yoo, D., So Kweon, I.: Distort-and-recover: Color enhancement using deep reinforcement learning. In: CVPR (2018)
25. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch. In: NeurIPS Workshops (2017)
26. Pizer, S., Amburn, E.P., Austin, J., Cromartie, R., Geselowitz, A., Greer, T., Romeny, B.T.H., Zimmerman, J.: Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing* (1987)
27. Ren, W., Liu, S., Ma, L., Xu, Q., Xu, X., Cao, X., Du, J., Yang, M.: Low-light image enhancement via a deep hybrid network. *IEEE TIP* (2019)
28. Reza, A.M.: Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology* (2004)
29. Risheng, L., Long, M., Jiaao, Z., Xin, F., Zhongxuan, L.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: CVPR (2021)
30. Rivera, R., Ryu, B., Chae, O.: Content-aware dark image enhancement through channel division. *IEEE TIP* (2012)
31. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
32. Sean Moran, Ales Leonardis, G.S.S.M.: Curl: Neural curve layers for global image enhancement. *arXiv:1911.13175* (2019)
33. Wang, R., Zhang, Q., Fu, C., Shen, X., Zheng, W., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: CVPR (2019)
34. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: CVPR (2018)
35. Wei, H., Yifeng, Z., Rui, H.: Low light image enhancement network with attention mechanism and retinex model. *IEEE Access* (2020)
36. Xu, K., Tian, X., Yang, X., Yin, B., Lau, R.W.H.: Intensity-aware single-image deraining with semantic and color regularization. *IEEE TIP* (2021)
37. Xu, K., Yang, X., Yin, B., Lau, R.W.: Learning to restore low-light images via decomposition-and-enhancement. In: CVPR (2020)
38. Xu, L., Yan, Q., Xia, Y., Jia, J.: Structure extraction from texture via natural variation measure. *ACM TOG* (2012)
39. Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J.: From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In: CVPR (2020)
40. Yang, X., Xu, K., Song, Y., Zhang, Q., Wei, X., Lau, R.: Image correction via deep reciprocating HDR transformation. In: CVPR (2018)
41. Yu, R., Liu, W., Zhang, Y., Qu, Z., Zhao, D., Zhang, B.: Deepexposure: Learning to expose photos with asynchronously reinforced adversarial learning. In: NeurIPS (2018)
42. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: CVPR (2021)
43. Zhang, Q., Nie, Y., Zheng, W.: Dual illumination estimation for robust exposure correction. *Computer Graphics Forum* (2019)
44. Zhang, Q., Yuan, G., Xiao, C., Zhu, L., Zheng, W.S.: High-quality exposure correction of underexposed photos. In: ACM MM (2018)
45. Zhang, Y., Zhang, J., Guo, X.: Kindling the darkness: A practical low-light image enhancer. In: ACM MM (2019)