Color Shift Estimation-and-Correction for Image Enhancement

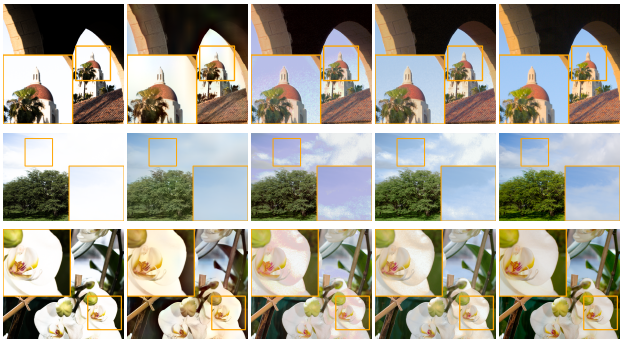
Yiyu Li Ke Xu Gerhard Petrus Hancke Rynson W.H. Lau
City University of Hong Kong
yiyuli.cs@my.cityu.edu.hk, kkangwing@gmail.com, {gp.hancke, rynson.lau}@cityu.edu.hk

Abstract

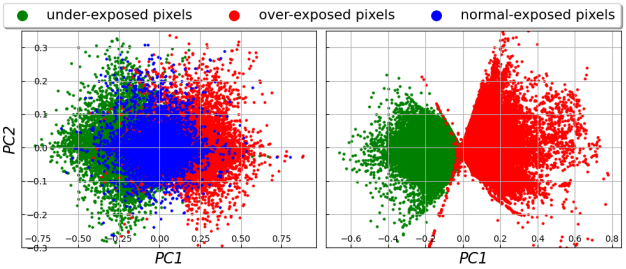
Images captured under sub-optimal illumination conditions may contain both over- and under-exposures. Current approaches mainly focus on adjusting image brightness, which may exacerbate color tone distortion in under-exposed areas and fail to restore accurate colors in over-exposed regions. We observe that over- and over-exposed regions display opposite color tone distribution shifts, which may not be easily normalized in joint modeling as they usually do not have "normal-exposed" regions/pixels as reference. In this paper, we propose a novel method to enhance images with both over- and under-exposures by learning to estimate and correct such color shifts. Specifically, we first derive the color feature maps of the brightened and darkened versions of the input image via a UNet-based network, followed by a pseudo-normal feature generator to produce pseudo-normal color feature maps. We then propose a novel COlor Shift Estimation (COSE) module to estimate the color shifts between the derived brightened (or darkened) color feature maps and the pseudo-normal color feature maps. The COSE module corrects the estimated color shifts of the over- and under-exposed regions separately. We further propose a novel COlor MOdulation (COMO) module to modulate the separately corrected colors in the over- and under-exposed regions to produce the enhanced image. Comprehensive experiments show that our method outperforms existing approaches. Project webpage: https://github.com/yiyulics/CSEC.

1. Introduction

Real-world scenarios often involve a wide range of illumination, which poses a significant challenge for photographing. Uneven scene illumination may easily produce both over- and under-exposures. Although cameras have an automatic exposure mode to determine an "ideal" exposure setting based on scene brightness, uniformly adjusting the exposure across the entire image may still result in excessively bright and excessively dark regions. Such under- and over-exposed regions can exhibit obvious color tone distortions. The relatively high noise level in under-exposed regions alters the data distribution to cause shifts in color tone, while the saturated over-exposed regions lose the original colors. Hence, enhancing such images typically involves brightness adjustment and color tone shift correction.



(a) Input (b) MSEC [1] (c) LCDP [30] (d) Ours (e) GT



(f) Sampled pixels in MSEC [1] (g) Sampled pixels in LCDP [30]

Figure 1. Given images with both over- and under-exposures (a), SOTA methods [1, 30] may still fail to correct color distortion (b,c). We show PCA results of pixels randomly sampled from MSEC [1] (images with either over- or under-exposed) (f) and LCDP [30] (images with both over- and under-exposures) (g) datasets, and make two observations. First, we observe that under-exposed pixels (greenish dots) tend to have a reverse distribution shift compared to over-exposed pixels (reddish dots) in both datasets. Second, unlike MSEC [1] that contains normal-exposed pixels (bluish dots) in their 0 EV input images, images with both over- and under-exposures do not have such "reference pixels" as guidance. These two observations inspire us to estimate and correct such color shifts conditioned on the created pseudo-normal exposed features. Our method can properly adjust images with both over- and under-exposures (d).

In recent years, numerous endeavors have been made to

enhance images that are improperly exposed. They can be broadly grouped into two categories. The first category focuses on enhancing either over-exposed or under-exposed images. Some methods [17, 24] propose to learn an exposure invariant representation space, where different exposure levels can be mapped to a standardized and invariant representation. Other methods [18, 36] propose to integrate frequency information with spatial information, which helps model the inherent image structural characteristics to enhance the image brightness and structure distortions. However, as the above methods typically assume that the over- or under-exposure happens to the whole image, they do not perform well for images with both over- and under-exposures (*e.g.*, Figure 1(b)). Recently, Wang *et al*. [30] propose the second category of work, which aims to enhance images with both over- and under-exposures. They utilize the local color distributions as a prior to guide the enhancement process. However, despite their pyramidal design of local color distributions prior, their method still tends to produce results with significant color shifts, particularly in large homogeneous regions (*e.g.*, Figure 1(c)).

In this paper, we aim to correct the image brightness and color distortions for images with both over- and under-exposures. To approach this problem, we first visualize the PCA results of pixels randomly sampled from two relevant datasets, MSEC [1] (each scene has five input images of different exposure values (EV)) and LCDP [30] (each scene has only one input image with both over- and under-exposures), in Figure 1(f) and 1(g), respectively. We have two observations from this initial study. First, under-exposed pixels (green dots) tend to have an opposite distribution shift compared to over-exposed pixels (red dots) in both datasets. Second, unlike MSEC [1] that contains 0 EV input images, which can serve as reference images for the exposure normalization process, images of LCDP [30] do not contain such "normal-exposed" pixels. The first observation inspires us to consider estimating and correcting such color shifts, while the second observation inspires us to create pseudo-normal exposed feature maps to function as a reference for color shift estimation and rectification.

To this end, we propose a novel method to adjust the image brightness and correct the color tone distortion jointly. Our method first uses a UNet-based network to derive color feature maps of the over- and under-exposed regions from the brightened and darkened versions of the input image. A pseudo-normal feature generator then creates the pseudo-normal color feature maps based on the derived color feature maps. A novel COlor Shift Estimation (COSE) module is then proposed to estimate and correct the color shifts between the derived brightened (or darkened) color feature maps and the created pseudo-normal color feature maps separately. We implement the COSE module by extending the deformable convolution in the spatial domain to the color feature domain. We further propose a novel COlor MOdulation (COMO) module to modulate the separately corrected colors of the over- and under-exposed regions to produce the enhanced image. We implement this COMO module via a tailored cross-attention mechanism performed on the input image and estimated darken/brighten color offsets. Figure 1(d) shows that our method can produce visually pleasing images.

Our main contributions can be summarized as follows:

1. We propose a novel neural approach to enhance images with both over- and under-exposures via modeling the color distribution shifts.
2. We propose a novel neural network with two novel modules, a novel COlor Shift Estimation (COSE) module for the estimation and correction of colors in over- and under-exposed regions separately, and a novel COlor MOdulation (COMO) module for modulating the corrected colors to produce the enhanced images.
3. Extensive experiments show that our network is lightweight and outperforms existing image enhancement methods on popular benchmarks.

2. Related Works

Exposure Correction. When photographing scenes with complex illuminations, the captured images may have over- and/or under-exposure problems. Exposure correction methods [16, 46, 48–50] aim to enhance such images to recover image details buried in over-/under-exposed regions.

A group of methods focus on under-exposed image enhancement. Some of these methods are Retinex-based, which employ either handcrafted features [5, 11, 15] or deep features [23, 32, 38, 51, 52], to decompose the image into illumination and reflectance components for their enhancement. Recently, Fu *et al*. [10] propose to combine contrastive learning and self-knowledge distillation to learn the decomposition process. Fu *et al*. [12] propose an unsupervised approach to learn the decomposition process, which uses paired images as input to guide the decomposition of each other with the consistency regularization between the two decomposed reflectance components of each image in the pair. Other methods directly learn the mappings from under-exposed images to normal-exposed images. These methods focus on designing various priors, *e.g.*, frequency information [40], lagrange multiplier [53], de-bayer filter [8], 3D lookup table [44], normalizing flow [35], and signal-to-noise ratio [42]. Most recently, Wu *et al*. [39] propose to use semantic segmentation maps to help maintain color consistency. Xu *et al*. [43] leverage the edge detection for structural modeling, which helps enhance the appearance. Wang *et al*. [31] propose to reconstruct an implicit neural radiance field (NeRF) using low-light images and then enhance the NeRF to produce normal-light images.

Recently, Afifi *et al*. [1] construct a large-scale dataset

that contains over-, normal- and under-exposed images, and propose a Laplacian pyramid-based network for exposure correction. Huang *et al*. [17] propose to learn an exposure invariant space to bridge the gap between over-exposure and under-exposure features in image-level. Later, Huang *et al*. [18] further leverage Fourier transform to combine spatial and frequency information to enhance image brightness and structures. Most recently, Huang *et al*. [19] model the relationship between over-exposed and under-exposed samples in mini-batches to maintain a stable optimization of exposure correction. Wang *et al*. [36] propose to decouple the high and low frequencies of images via specially designed convolution kernels to improve the image structural modeling during exposure correction. Yang *et al*. [45] propose a unified frequency decomposition method for multiple illumination-related tasks, in which the exposure information is adjusted in their low-frequency component. Baek *et al*. [2] propose a luminance-aware method for multi-exposure correction.

The aforementioned methods focus on modeling the exposure correction problem at the image-level, assuming that each image is either over- or under-exposed. They are unable to handle images with both over- and under-exposures. In contrast, Wang *et al*. [30] propose to leverage local color distributions to guide the network to locate and enhance over- and under-exposed regions. Despite the success, their local color distributions prior often fails to correct colors, especially in large over- and under-exposed regions. In contrast, in this work, we propose a novel approach to address images with both over- and under-exposures through adaptive brightness adjustment and color shift correction.

Deformable Convolutions. Convolutional neural networks [22] have an inherent limitation in modeling the geometric transformations due to the fixed kernel configuration. To facilitate the transformation modeling capability of CNNs, Dai *et al*. [7] propose the deformable convolution operation, which adds additional spatial offsets and learns the spatial offsets from the target tasks. Later, Zhu *et al*. [55] propose an improved version of deformable convolution, which incorporates modulation scalar to measure the importance of different locations. Wang *et al*. [33] propose to improve the deformable convolution by introducing a multi-group mechanism and sharing weights among convolutional neurons to make the model computationally efficient. Numerous works have adopted deformable convolutions in recent years for different high-level vision tasks, *e.g*., image classification [33], object detection [7], video object detection [3], semantic segmentation [7], and human pose estimation [28]. Meanwhile, deformable convolution has also been adopted for low-level vision tasks, *e.g*., for aligning multi-frame information with a video [9, 29, 34].

While these methods typically use deformable convolutions to aggregate spatial contextual info, we propose in this work to leverage deformable convolutions to estimate color shifts of over-/under-exposed regions for enhancement.

3. Proposed Method

Our method is inspired by two observations. First, the over-exposed pixels tend to have an inverse distribution shift compared to the under-exposed pixels. This suggests that it is necessary to capture and correct such color shifts separately. Second, since majority (if not all) of pixels are affected by over-/under-exposures, it is necessary to create pseudo-normal exposure information as guidance for the estimation of the color shifts of over-/under-exposed pixels. Inspired by these two observations, we propose a novel network (Sec. 3.1) with two novel modules: a novel COlor Shift Estimation (COSE) module (Sec. 3.2) and a novel COlor MOdulation (COMO) module (Sec. 3.3), for the enhancement of images with both over-/under-exposures.

3.1. Network Overview

Given an input image $I_x \in \mathcal{R}^{3\times H\times W}$ with both over- and under-exposures, our network aims to produce an enhanced image $I_y \in \mathcal{R}^{3\times H\times W}$ with rectified image brightness and recovered image details and colors. The overview of our model is shown in Figure 2. Given the input image I_x, we first compute its inverse version of $\hat{I}_x$ by $1 - I_x$, both of which are fed into a UNet-based network to extract two illumination maps $F_L^U \in \mathcal{R}^{1\times H\times W}$ and $F_L^O \in \mathcal{R}^{1\times H\times W}$. These two maps (*i.e*., F_L^U and F_L^O) indicate the regions affected by under-exposure and over-exposure, respectively. We then compute the darkened (F_D) and brightened (F_B) feature maps as:

$$F_B = \frac{I_x}{F_L^U} = \frac{I_x}{f(I_x)}, \tag{1}$$

$$F_D = 1 - \frac{1 - I_x}{F_L^O} = 1 - \frac{1 - I_x}{f(1 - I_x)}, \tag{2}$$

where $f(\cdot)$ represents the UNet-based feature extractor. We model the color shifts based on the brightened and darkened feature maps $F_B, F_D \in \mathcal{R}^{3\times H\times W}$.

Given F_B and F_D, we first apply a Pseudo-Normal Feature Generator to fuse them with the input image I_x into a pseudo-normal feature map F_N, as:

$$F_N = g(F_B, F_D, I_x), \tag{3}$$

where $g(\cdot)$ denotes the pseudo normal-exposure generator. F_N can then serve as a reference to guide the estimation of color shifts between F_B and F_N (and F_D and F_N), separately, via two proposed COSE modules. The darken offset O_D and the brighten offset O_B produced by the two COSE modules model both brightness and color shifts w.r.t. the input image I_x, and therefore O_D, O_B, and I_x are fed into the proposed COMO module to adjust the image brightness and correct color shifts, to produce the final image I_y.

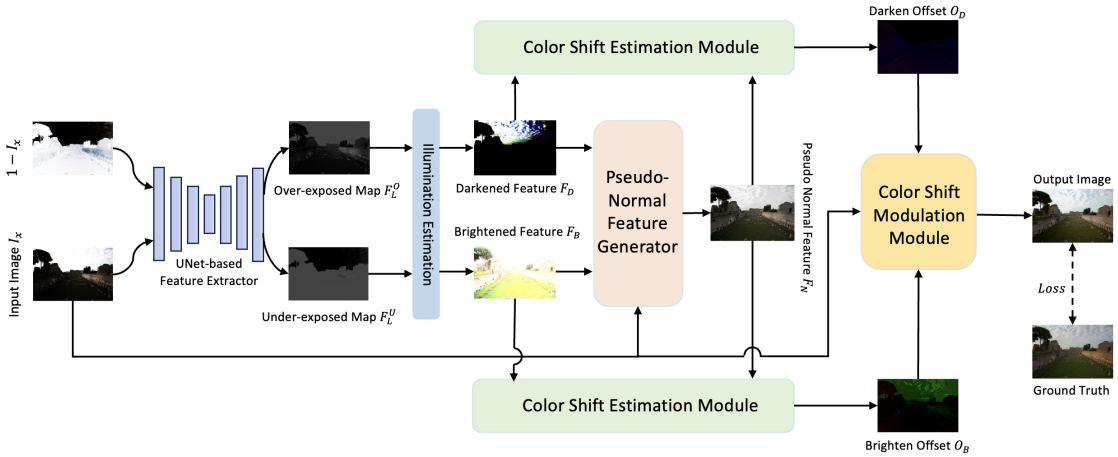


Figure 2. Overview of our proposed model. We first generate darkened features F_D and brightened features F_B using the UNet-based feature extractor. We then derive a pseudo-normal feature map F_N using the generated brightened/darkened feature maps and the input image I_x. We further estimate the color shifts between the brightened/darkened color features F_B/F_D and the created pseudo-normal feature map F_N using the proposed Color Shift Estimation (COSE) module to obtain two individual offset maps O_B and O_D. Finally, we modulate the image brightness and colors using the proposed Color Modulation (COMO) module, to produce the final output image.

3.2. Color Shift Estimation (COSE) Module

Unlike brightness adjustment, color shift correction is more challenging as it essentially requires the network to model the pixel directions in RGB color space, instead of the magnitudes of pixel intensity. Although there are some works [30, 32, 41] that use the cosine similarity regularization to help maintain the image colors during training, such a strategy often fails in large under-/over-exposed regions where pixels of small/high-intensity values are expected to have different colors.

We propose the COSE module to address this problem based on the deformable convolution techniques. Deformable convolution (DConv) extends vanilla convolution by incorporating spatial offsets Δp_n to adaptively perform convolution at any location of any $N \times N$ pixels, where $N \times N$ denotes the kernel size. The modulation Δm_n is proposed to assign different weights to different kernel positions, making the convolution operator focusing on important pixels. While DConv can predict the offsets with respect to the basis, it is possible for DConv to capture the color distribution shift. However, since previous methods [7, 33, 55] only apply DConv in the pixel spatial domain, we propose to extend DConv to be performed in both spatial domain and color space, for modeling the brightness changes and color shifts jointly. Specifically, as depicted in Figure 3, our COSE module first concatenates the pseudo normal feature map F_N and the brightened/darkened feature map F_B/F_D along the channel dimension. It then employs three separate 3×3 convolutions to extract positional offsets $\Delta p_n \in \mathcal{R}^{B\times 2N\times H\times W}$, color offsets $\Delta c_n \in \mathcal{R}^{B\times 3N\times H\times W}$, and modulation $\Delta m_n \in \mathcal{R}^{B\times N\times H\times W}$. Positional offsets Δp_n and modulation Δm_n are the same as those in [55]. They are performed in the spatial domain to aggregate spatial contextual information within the deformed irregular receptive fields of the convolution operation. Additionally, the color offsets Δc_n are introduced, to represent the color offsets for each channel at each kernel position. The learned color offsets Δc_n are formulated to have $3N$ channels to model the color shifts of the input sRGB image with 3 channels.

Formally, the computation of deformable convolution in both spatial domain and color space can be written as:

$$y = \sum_{p_n \in \mathcal{R}} (w_n \cdot x(p_0 + p_n + \Delta p_n) + \Delta c_n) \cdot \Delta m_n, \quad (4)$$

where x denotes the input features to the convolution operation, while p_0, p_n, and Δp_n are 2-dimensional variables indicating the spatial location, and y (or $y(p_0)$) de-

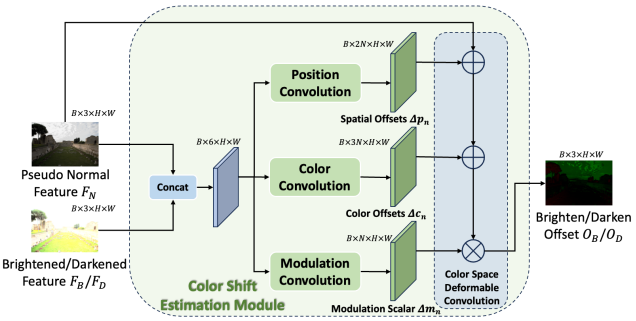


Figure 3. The proposed Color Shift Estimation (COSE) module aims to model the color distribution shifts in over- and under-exposed regions. We use three different convolutions to estimate the positional offsets, color shift offsets, and modulation scalars. We then utilize the color space deformable convolution in Eq. 4 to compute the color offsets and produce the offset feature map for brightened and darkened features.

notes the output of color space deformable convolution for each pixel p_0 within the input image. The set $\mathcal{R} = \{(-1,-1),(-1,0),\ldots,(1,1)\}$ represents the grids of a regular 3×3 kernel. n is the enumerator for the elements in $\mathcal{R}$ indicating the n-th position, and N is the length of $\mathcal{R}$ ($N = 9$ for a regular 3×3 kernel). Since the displacement Δp_n may assume fractional values in practice, we apply bilinear interpolation for computation, which aligns with the spatial deformable convolution [7].

3.3. Color Modulation (COMO) Module

We propose the COMO module to modulate the brightness and color of the input image and produce the final output image I_y, based on the learned offsets O_B/O_D between brightened features F_B/darkened features F_D and pseudo normal features F_N. Since it is crucial to aggregate global information in order to produce the corrected image with harmonious colors, we draw inspiration from the non-local context modeling in [54] and formulate our COMO module by extending the self-affinity computation in [54] into the cross-affinity computation so that COMO can enhance the input image by querying both O_B and O_D.

As shown in Figure 4, we assign three branches to process the input image I_x, darken offset O_D, and brighten offset O_B, separately, while each branch contains three 1×1 convolution layers (denoted as $Conv\psi$, $Conv\phi$, and $ConvZ$). We then compute the self-affinity matrix A_i in each branch, as:

$$A_i = \psi_i \otimes \phi_i, \ for\ i \in \{I, B, D\}, \tag{5}$$

where $\otimes$ is the matrix multiplication, ψ_i and ϕ_i are the feature maps obtained by $Conv\psi$ and $Conv\phi$, respectively. A_i is then symmetrized and normalized to ensure the existence of real eigenvalues and the stabilization of backward propagation. Each row of A_i serves as a spatial attention map, and Z_i (obtained by $ConvZ$) serves as weights for the attention maps. Next, we model the correlations between I_x and O_B/O_B via matrix multiplication and add them together with the self-affined features, as:

$$f_j = w_1 A_j \otimes Z_j + w_2 A_j \otimes Z_I, \tag{6}$$

where $j \in \{B, D\}$ is the index of the affinity matrix A_j and feature map Z_j from the brighten or darkened branch. w_1 and w_2 are weight matrices generated by 1×1 convolutions. While the first term in Eq. 6 is to discover the significant color offset regions in O_B and O_D learned by COSE, and the second term aims to use the learned weights of input Z_I to attend to attention maps of O_B and O_D to discover what the offsets are like in the significant regions of input.

Finally, we combine f_B, f_D and input image I_x to fuse the explored color offsets as guidance for input image. We then produce the final result I_y, as:

$$I_y = w_4(BN(f_B) + BN(f_D) + w_3 A_I \otimes Z_I) + I_x, \tag{7}$$

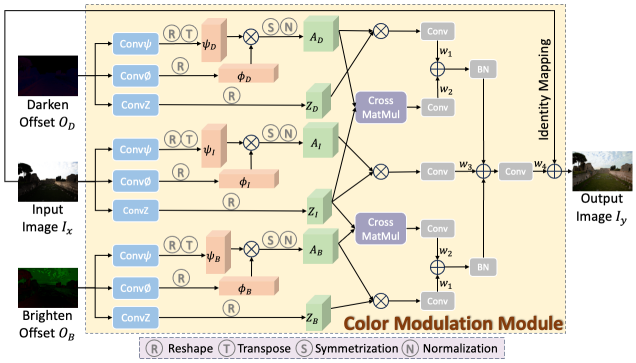


Figure 4. The proposed Color Modulation (COMO) module aims to adjust the brightness and correct the color shifts of the input image I_x, guided by the color offsets from our COSE modules. The COMO module takes the image I_x and brighten/darken offsets O_B/O_D as input and produces the enhanced image via a tailored cross-attention mechanism.

where $BN(\cdot)$ represents Batch Normalization, and w_3, w_4 are weight matrices generated by 1×1 convolutions.

3.4. Loss Function

Two loss functions, $\mathcal{L}_{pesudo}$ and $\mathcal{L}_{output}$ are used to train our network. As we need to generate a pseudo-normal feature map to help identify the color shifts, we use $\mathcal{L}_{pesudo}$ to provide intermediate supervision for the generation process:

$$\mathcal{L}_{pesudo} = ||F_N - GT||_1. \tag{8}$$

The $\mathcal{L}_{output}$ contains four terms to supervise the network to produce the enhanced images, which are $L1$ loss, Cosine similarity $\mathcal{L}_{cos}$, SSIM loss $\mathcal{L}_{ssim}$ [37], and VGG loss $\mathcal{L}_{vgg}$ [27]. The $\mathcal{L}_{output}$ can be formulated as:

$$\mathcal{L}_{output} = \lambda_1 \mathcal{L}_{L1} + \lambda_2 \mathcal{L}_{cos} + \lambda_3 \mathcal{L}_{ssim} + \lambda_4 \mathcal{L}_{vgg}, \tag{9}$$

where λ_1, λ_2, λ_3 and λ_4 are four balancing hyper-parameters. The overall loss function is then:

$$\mathcal{L} = \lambda_p \mathcal{L}_{pesudo} + \lambda_o \mathcal{L}_{output}, \tag{10}$$

where λ_p and λ_o are two balancing hyper-parameters.

4. Experiments

4.1. Experimental Settings

Datasets. We evaluate our proposed method on two datasets: LCDP [30], and MSEC [1]., both of which are derived from the MIT-Adobe FiveK dataset [4]. Specifically, LCDP contains images with both over-exposure and under-exposure regions, and we use it to test the ability of our method to handle both over- and under-exposure conditions. The LCDP dataset consists of 1,733 images, which

are split into 1,415 for training, 100 for validation, and 218 for testing. MSEC contains images with either over- or under-exposure, while their exposure values are in the collection of {-1.5EV, -1EV, 0EV, +1EV, +1.5EV}, and the expert-retouched images are used as the ground truths. The MSEC dataset comprises 17,675 training images, 750 validation images, and 5,905 testing images. We conduct experiments on this dataset in order to test the generalization ability of the proposed method.

Evaluation Metrics. To evaluate the performance of different image enhancement methods and our proposed method, we adopt the commonly used metrics Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [37] as numerical evaluation metrics. Additionally, to gauge the accuracy of color correction, we utilize the Root Mean Squared Error (RMSE) on LAB color space as a color-related metric.

4.2. Comparisons with State-of-the-art Methods

Quantitative Comparisons. To verify the effectiveness of our proposed method, we compare our method with one conventional method: Histogram Equalization [26], and 13 deep learning based SOTA methods: DSLR [20], HDRNet [13], RetinexNet [38], DeepUPE [32], ZeroDCE [14], MSEC [1], RUAS [23], SNRNet [42], LCDPNet [30], FecNet [18], LANet [45], SMG [43], and RetinexFormer [6].

We first evaluate the effectiveness of our model in enhancing images with both over- and under-exposures by comparing our model to existing methods on the LCDP [30] dataset. Table 1 reports the comparison results, where all methods are re-trained for a fair comparison. We can see that our method outperforms the second-best method, the RetinexFormer [6] in terms of the PSNR and SSIM metrics while using only around 20% network parameters compared to it. Our method outperforms the LCDPNet [30] by a large margin (ΔPSNR: $+0.70$, ΔSSIM: $+0.038$) while we have approximately similar network parameters (Ours/LCDPNet: 0.30M/0.28M).

Method	PSNR ↑	SSIM ↑	# params
HE'87 [26]	16.215	0.669	-
DSLR'17 [20]	20.856	0.758	0.39M
HDRNet'17 [13]	21.834	0.818	0.48M
RetinexNet'18 [38]	20.199	0.709	0.84M
DeepUPE'19 [32]	20.970	0.818	1.02M
ZeroDCE'20 [14]	12.861	0.668	0.08M
MSEC'21 [1]	20.377	0.779	7.04M
RUAS'21 [23]	13.757	0.606	0.003M
SNRNet'22 [42]	20.829	0.711	4.01M
LCDPNet'22 [30]	22.931	0.817	0.28M
FECNet'22 [18]	23.333	0.823	0.15M
LANet'23 [45]	21.444	0.691	0.57M
SMG'23 [43]	22.427	0.786	17.90M
RetinexFormer'23 [6]	<u>23.360</u>	<u>0.850</u>	1.61M
Ours	**23.627**	**0.855**	0.30M

Table 1. Quantitative comparison between the proposed method and SOTA methods on the LCDP [30] test set. All methods are re-trained on the LCDP training set. Best results are marked in **bold** and second best results are <u>underlined</u>.

Method	HE [26]	Retinex [38]	ZeroDCE [14]
RMSE ↓	7.525	6.945	7.960
Method	MSEC [1]	SNR [42]	LCDP [30]
RMSE ↓	7.378	6.715	6.608
Method	FEC [18]	RetinexF [6]	Ours
RMSE ↓	6.384	<u>6.148</u>	**6.105**

Table 2. Quantitative comparison between the proposed method and SOTA methods using RMSE on the LCDP [30] test set. All methods are re-trained on the LCDP training set. Best performances are marked in **bold** and second best results are <u>underlined</u>.

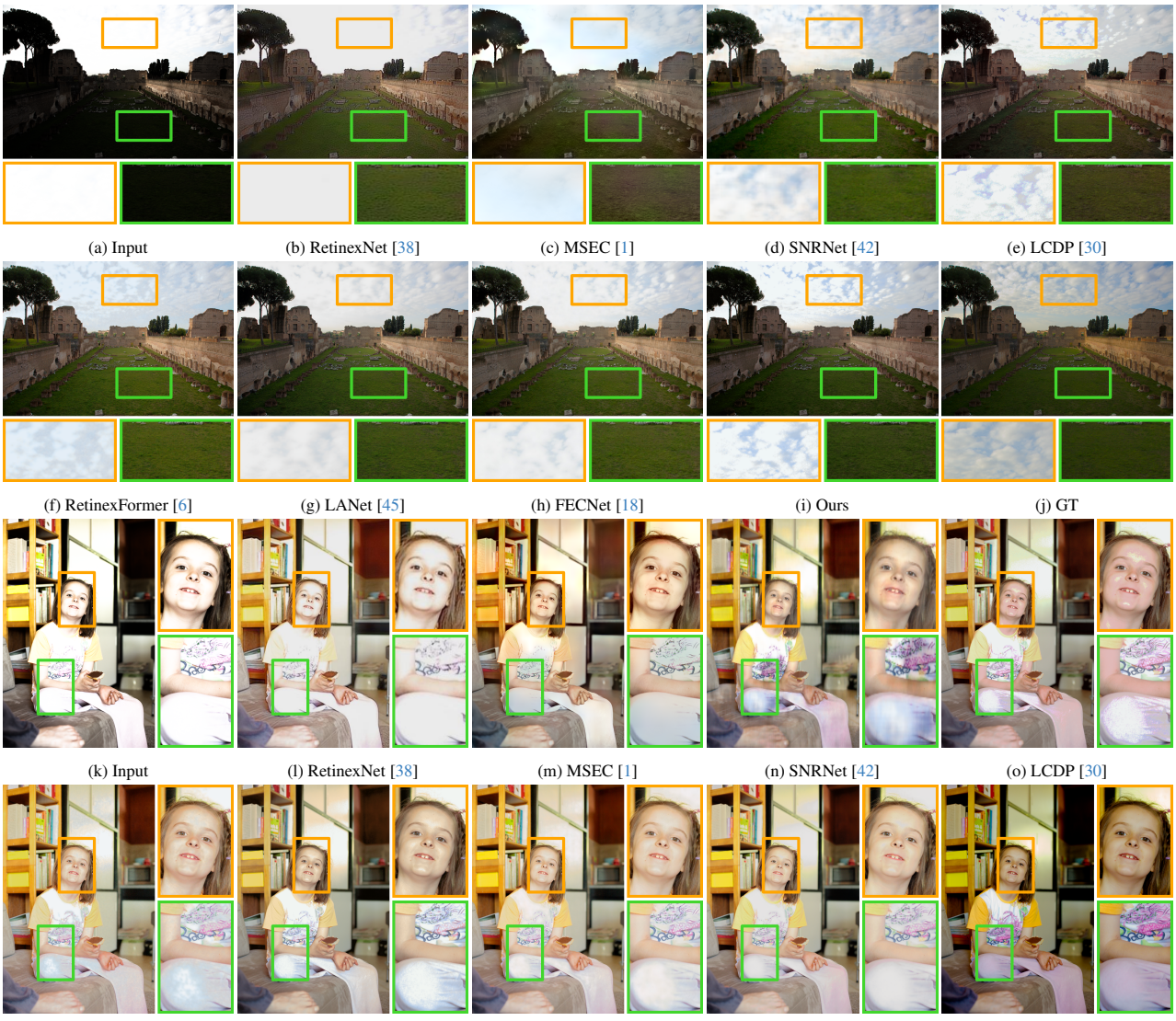
Method	PSNR ↑	SSIM ↑
HE'87 [26]	15.873	0.731
MSEC'21 [1]	20.482	0.825
LCDPNet'22 [30]	22.224	<u>0.849</u>
FECNet'22 [18]	<u>22.519</u>	0.848
SMG'23 [43]	22.075	0.781
RetinexFormer'23 [6]	21.809	0.846
Ours	**22.728**	**0.863**

Table 3. Quantitative comparison between the proposed method and SOTA methods on the MSEC [1] test set. All methods are re-trained on the MSEC training set. Best performances are marked in **bold** and second best performances are <u>underlined</u>.

To evaluate the color correction performance of our method, we also compare our model with SOTA methods using color-related metrics, which is the root mean squared error (RMSE) on LAB color space. Table 2 shows the results, where our method achieves the best performance.

We further compare our method with LCDPNet [30], FECNet [18], SMG [43] and RetinexFormer [6], which are the best-four performing existing methods according to Table 1, on the MSEC [1] dataset. We also report the performance of MSEC [1] and HE [26] for reference. As shown in Table 3, our method also outperforms these existing methods, which demonstrates that our color shift estimation and correction method can handle either over- or under-exposure at image level well.

Qualitative Comparisons. We further visually compare the results of the proposed method and state-of-the-art methods in Figure 5, where input images are from the LCDP [30] dataset. We can see that methods proposed for handling either over- or under-exposures (*i.e.*, MSEC [1], LANet [45], and FECNet [18]) cannot rectify the brightness of over- and under-exposed regions at the same time. For example, they either fail to restore the greenish color of under-exposed grasses, or fail to restore the bluish color of the over-exposed sky (see Figure 5 (c, g, and h)). Methods originally proposed for low-light image enhancement (*i.e.*, RetinexNet [38], SNRNet [42], and RetinexFormer [6]) tend to increase the overall image brightness, which further buries the details of clouds (Figure 5 (b, d, and f)). LCDP [30] is designed to deal with the coexistence of over-



(a) Input (b) RetinexNet [38] (c) MSEC [1] (d) SNRNet [42] (e) LCDP [30]

(f) RetinexFormer [6] (g) LANet [45] (h) FECNet [18] (i) Ours (j) GT

(k) Input (l) RetinexNet [38] (m) MSEC [1] (n) SNRNet [42] (o) LCDP [30]

(p) RetinexFormer [6] (q) LANet [45] (r) FECNet [18] (s) Ours (t) GT

Figure 5. Visual comparison between our method and state-of-the-art methods on LCDP [30] dataset, which has images with both over-exposure and under-exposure.

and under-exposure; however, the sky regions after the enhancement tend to have purplish colors (e). Similar problems can be found in the second case. In contrast, our method can enhance these images with better details and colors (Figure 5 (i and s)).

Figure 6 further shows visual results of real-world images from the Flickr 30K [47] dataset. While existing methods tend to produce obvious color artifacts after enhancements, our method produces results of better visual quality.

4.3. Ablation Study

We now conduct ablation studies on the proposed network designs in Table 4. First, we replace the proposed color space deformable convolution with three different types of spatial deformable convolutions (*i.e.*, deformable convolution with only spatial offsets $Ours_{Spatial}$, deformable convolution with spatial offsets and modulation scalars $Ours_{Spatial+Modul}$, and deformable convolution with spatial offsets and proposed color offsets $Ours_{Spatial+Color}$), and with the standard convolution (denoted as $Ours_{Conv}$), to test the effectiveness of the proposed Color Shift Estimation module. Second, we replace the proposed Color Modulation module with the original non-local block and concatenate O_B, O_D, and I_x as input (denoted as $Ours_{NL}$) to test the effectiveness of Color Modulation module. The first three rows in Table 4 show that replacing either the proposed COSE or COMO modules results in the degradation of the enhancement performance.

We also investigate different pipeline variants. First, we change the over-exposed illumination map F_L^O from

Options	PSNR ↑	SSIM ↑
$Ours_{Conv}$	23.443	0.834
$Ours_{Spatial}$	22.623	0.827
$Ours_{Spatial+Modul}$	23.569	0.840
$Ours_{Spatial+Color}$	23.415	0.836
$Ours_{NL}$	23.201	0.844
$Ours_{opposition}$	21.829	0.799
$Ours_{3channel}$	23.016	0.830
$Ours_{NoShare}$	23.513	0.844
$Ours_{output}$	23.588	0.836
$Ours_{w/oVGG}$	23.593	0.846
$Ours_{w/oVGGSSIM}$	23.512	0.837
$Ours$	**23.627**	**0.855**

Table 4. Ablation study. Best results are marked in **bold**.

$f(1 - I_x)$ to $1 - f(I_X)$ (where $f(\cdot)$ denotes the UNet-based feature extractor), to make a strict opposition between over-exposed map and under-exposed map (denoted as $Ours_{opposition}$). Second, we change two illumination maps from having 1-channel to 3-channels (which is adopted in LCDP [30] and DeepUPE [32]) (denoted as $Ours_{3channel}$). We also use an independent pseudo-normal feature generator that does not share parameters with the color modulation module (denoted as $Ours_{NoShare}$). The 4th to 6th rows in Table 4 show that using other pipeline variants may significantly degrade the performance, which verifies the effectiveness of our designs.

Last, we perform ablation studies on the loss function. Specifically, we first remove $\mathcal{L}_{pseudo}$ and only supervise the final result (denoted as $Ours_{output}$). Second, we remove the VGG loss (denoted as $Ours_{w/oVGG}$). Last, we remove both SSIM loss and VGG loss (denoted as $Ours_{w/oVGGSSIM}$). The 7th to 9th rows shows that adding these losses help improve the performance of our network.

5. Conclusion

In this paper, we have studied the problem of enhancing images with both over- and under-exposures. We have observed opposite color tone distribution shifts between over- and under-exposed pixels, and a lack of "normal-exposed" regions/pixels as reference resulting in the unsatisfactory performance of existing methods. We have proposed a novel method to enhance images with both over- and under-exposures by learning to estimate and correct the color tone shifts. Our method has a novel color shift estimation module to estimate the color shifts based on our created pseudo-normal color feature maps and correct the estimated color shifts of the over- and under-exposed regions separately. A novel color modulation module is proposed to modulate the separately corrected colors in the over- and under-exposed regions to produce the enhanced image. We have conducted extensive experiments to show that our method outperforms existing image enhancement approaches.

Our method does have limitations. Since our method relies on the generation of pseudo-normal feature maps to help estimate and correct the brightness and color shifts, it may fail in scenarios where the over-exposed pixels are completely saturated (*e.g.*, the forehead and cheeks in Figure 7(a)), resulting in an unsatisfactory enhancement result (Figure 7(b)). Incorporating generative models into our pseudo-normal feature generator may help address such limitations and can be interesting for future work.

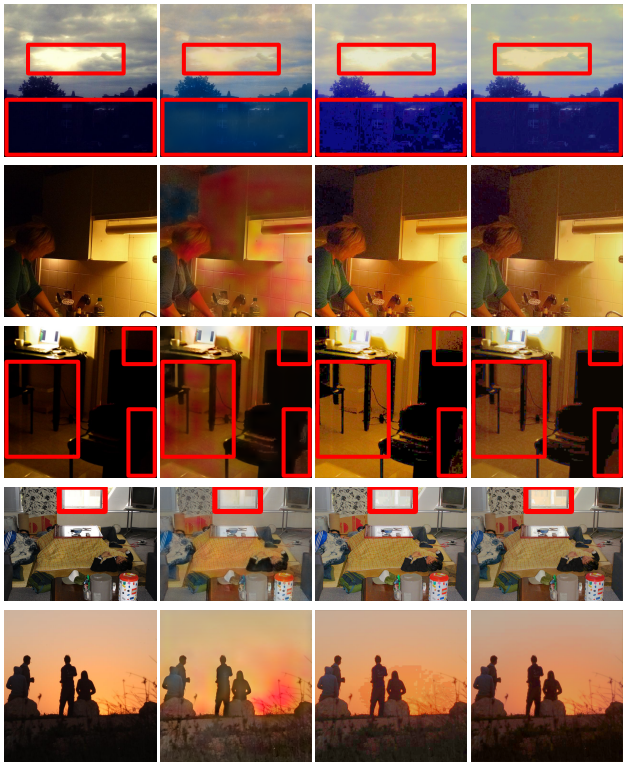


Figure 6. Visual results of real-world examples from the Flickr 30K [47] dataset. Our method produces better visual results.

Acknowledgements. This work is partly supported by an ITF grant from the Innovation and Technology Commission of Hong Kong SAR (ITC Ref.: PRP/003/22FX).



(a) Input (b) Ours

Figure 7. Our method may fail when pixels are completely saturated in over-exposed regions (*e.g.*, the forehead and cheeks) (a), where there is insufficient contextual information for our model to estimate color shifts and restore the original colors (b).

References

[1] Mahmoud Afifi, Konstantinos G Derpanis, Björn Ommer, and Michael S Brown. Learning multi-scale photo exposure correction. In *CVPR*, 2021. 1, 2, 5, 6, 7, 8, 3, 4

[2] Jong-Hyeon Baek, DaeHyun Kim, Su-Min Choi, Hyo-jun Lee, Hanul Kim, and Yeong Jun Koh. Luminance-aware color transform for multiple exposure correction. In *ICCV*, 2023. 3

[3] Gedas Bertasius, Lorenzo Torresani, and Jianbo Shi. Object detection in video with spatiotemporal sampling networks. In *ECCV*, 2018. 3

[4] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR*, 2011. 5

[5] Bolun Cai, Xianming Xu, Kailing Guo, Kui Jia, Bin Hu, and Dacheng Tao. A joint intrinsic-extrinsic prior model for retinex. In *ICCV*, 2017. 2

[6] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *ICCV*, 2023. 6, 7, 1, 2, 3

[7] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *ICCV*, 2017. 3, 4, 5

[8] Xingbo Dong, Wanyan Xu, Zhihui Miao, Lan Ma, Chao Zhang, Jiewen Yang, Zhe Jin, Andrew Beng Jin Teoh, and Jiajun Shen. Abandoning the bayer-filter to see in the dark. In *CVPR*, 2022. 2

[9] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burst image restoration and enhancement. In *CVPR*, 2022. 3

[10] Huiyuan Fu, Wenkai Zheng, Xiangyu Meng, Xin Wang, Chuanming Wang, and Huadong Ma. You do not need additional priors or regularizers in retinex-based low-light image enhancement. In *CVPR*, 2023. 2

[11] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *CVPR*, 2016. 2

[12] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In *CVPR*, 2023. 2

[13] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. In *ACM TOG*, 2017. 6

[14] 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 *CVPR*, 2020. 6

[15] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. In *TIP*, 2016. 2

[16] Yuanming Hu, Hao He, Chenxi Xu, Baoyuan Wang, and Stephen Lin. Exposure: A white-box photo post-processing framework. In *SIGGRAPH*, 2018. 2

[17] Jie Huang, Yajing Liu, Xueyang Fu, Man Zhou, Yang Wang, Feng Zhao, and Zhiwei Xiong. Exposure normalization and compensation for multiple-exposure correction. In *CVPR*, 2022. 2, 3

[18] Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction with spatial-frequency interaction. In *ECCV*, 2022. 2, 3, 6, 7, 1

[19] Jie Huang, Feng Zhao, Man Zhou, Jie Xiao, Naishan Zheng, Kaiwen Zheng, and Zhiwei Xiong. Learning sample relationship for exposure correction. In *CVPR*, 2023. 3

[20] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *ICCV*, 2017. 6

[21] D Kinga, Jimmy Ba Adam, et al. A method for stochastic optimization. In *ICLR*, 2015. 1

[22] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. In *Proc. IEEE*, 1998. 3

[23] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *CVPR*, 2021. 2, 6

[24] Ntumba Elie Nsampi, Zhongyun Hu, and Qing Wang. Learning exposure correction via consistency modeling. In *BMVC*, 2018. 2

[25] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NeurIPS Workshops*, 2017. 1

[26] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. In *Computer Vision, Graphics, and Image Processing*, 1987. 6

[27] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015. 5

[28] Xiao Sun, Bin Xiao, Fangyin Wei, Shuang Liang, and Yichen Wei. Integral human pose regression. In *ECCV*, 2018. 3

[29] Yapeng Tian, Yulun Zhang, Yun Fu, and Chenliang Xu. Tdan: Temporally-deformable alignment network for video super-resolution. In *CVPR*, 2020. 3

[30] Haoyuan Wang, Ke Xu, and Rynson W.H. Lau. Local color distributions prior for image enhancement. In *ECCV*, 2022. 1, 2, 3, 4, 5, 6, 7, 8

[31] Haoyuan Wang, Xiaogang Xu, Ke Xu, and Rynson W.H. Lau. Lighting up nerf via unsupervised decomposition and enhancement. In *ICCV*, 2023. 2

[32] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *CVPR*, 2019. 2, 4, 6, 8

[33] Wenhai Wang, Jifeng Dai, Zhe Chen, Zhenhang Huang, Zhiqi Li, Xizhou Zhu, Xiaowei Hu, Tong Lu, Lewei Lu, Hongsheng Li, Xiaogang Wang, and Yu Qiao. Internimage: Exploring large-scale vision foundation models with deformable convolutions. In *CVPR*, 2023. 3, 4

[34] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *CVPRW*, 2019. 3

[35] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *AAAI*, 2022. 2

[36] Yang Wang, Long Peng, Liang Li, Yang Cao, and Zheng-Jun Zha. Decoupling-and-aggregating for image exposure correction. In *CVPR*, 2023. 2, 3

[37] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. In *TIP*, 2004. 5, 6

[38] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *BMVC*, 2018. 2, 6, 7, 1, 3

[39] Yuhui Wu, Chen Pan, Guoqing Wang, Yang Yang, Jiwei Wei, Chongyi Li, and Heng Tao Shen. Learning semantic-aware knowledge guidance for low-light image enhancement. In *CVPR*, 2023. 2

[40] Ke Xu, Xin Yang, Baocai Yin, and Rynson W.H. Lau. Learning to restore low-light images via decomposition-and-enhancement. In *CVPR*, 2020. 2

[41] Ke Xu, Xin Tian, Xin Yang, Baocai Yin, and Rynson W. H. Lau. Intensity-aware single-image deraining with semantic and color regularization. In *TIP*, 2021. 4

[42] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *CVPR*, 2022. 2, 6, 7, 1, 3

[43] Xiaogang Xu, Ruixing Wang, and Jiangbo Lu. Low-light image enhancement via structure modeling and guidance. In *CVPR*, 2023. 2, 6

[44] Canqian Yang, Meiguang Jin, Xu Jia, Yi Xu, and Ying Chen. Adaint: Learning adaptive intervals for 3d lookup tables on real-time image enhancement. In *CVPR*, 2022. 2

[45] Kai-Fu Yang, Cheng Cheng, Shi-Xuan Zhao, Hong-Mei Yan, Xian-Shi Zhang, and Yong-Jie Li. Learning to adapt to light. In *IJCV*, 2023. 3, 6, 7, 1, 2

[46] Xin Yang, Ke Xu, Yibing Song, Qiang Zhang, Xiaopeng Wei, and Rynson W.H. Lau. Image correction via deep reciprocating HDR transformation. In *CVPR*, 2018. 2

[47] Peter Young, Alice Lai, Micah Hodosh, and Julia Hockenmaier. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. In *TACL*, 2014. 7, 8, 1, 4

[48] Runsheng Yu, Wenyu Liu, Yasen Zhang, Zhi Qu, Deli Zhao, and Bo Zhang. Deepexposure: Learning to expose photos with asynchronously reinforced adversarial learning. In *NeurIPS*, 2018. 2

[49] Lu Yuan and Sun Jian. Automatic exposure correction of consumer photographs. In *ECCV*, 2012.

[50] Qing Zhang, Yongwei Nie, and Wei-Shi Zheng. Dual illumination estimation for robust exposure correction. In *CGF*, 2019. 2

[51] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *ACM MM*, 2019. 2

[52] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. In *IJCV*, 2021. 2

[53] Chuanjun Zheng, Daming Shi, and Wentian Shi. Adaptive unfolding total variation network for low-light image enhancement. In *CVPR*, 2021. 2

[54] Lei Zhu, Qi She, Duo Li, Yanye Lu, Xuejing Kang, Jie Hu, and Changhu Wang. Unifying nonlocal blocks for neural networks. In *ICCV*, 2021. 5

[55] Xizhou Zhu, Han Hu, Stephen Lin, and Jifeng Dai. Deformable convnets v2: More deformable, better results. In *CVPR*, 2019. 3, 4

Color Shift Estimation-and-Correction for Image Enhancement

Supplementary Material

Yiyu Li Ke Xu Gerhard Petrus Hancke Rynson W.H. Lau
City University of Hong Kong
yiyuli.cs@my.cityu.edu.hk, kkangwing@gmail.com, {gp.hancke, rynson.lau}@cityu.edu.hk

This supplementary material provides implementation details and more visual comparisons between the proposed method and seven state-of-the-art image enhancement methods, *i.e.*, RetinexNet [38], MSEC [1], SNRNet [42], LCDP [30], Retinex-Former [6], LANet [45], and FECNet [18], in Figure 8 and Figure 9.

Implementation Details. We implement our model using PyTorch [25]. All our experiments are conducted on a single NVIDIA V100 GPU. The network parameters are optimized using Adam optimizer [21] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate is set to $1e^{-4}$. During training, we resize the image to 512×512 and set the batch size to 6. We implement our pseudo-normal feature generator with the same structure as our color shift modulation module and share their weights. For the loss functions, we empirically set the balancing hyper-parameters in Eq. 9 and Eq. 10 to $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $\lambda_3 = 1.0$, $\lambda_4 = 0.01$, $\lambda_p = 0.5$, and $\lambda_o = 1.0$. The UNet-based feature extractor is the same as LCDPNet [30], which is a UNet with histogram information in different scales. The number of output channels of the 1×1 convolution layers in COMO block is set to 16.

More Results. We can see that methods proposed for handling either over- or under-exposures (*i.e.*, MSEC [1], LANet [45], and FECNet [18]) cannot rectify the brightness of over- and under-exposed regions at the same time. For example, they either fail to restore the greenish color of under-exposed trees and grasses, or fail to restore the bluish color of the over-exposed sky (see Figure 8 and Figure 9 (c, g, and h)). Methods originally proposed for low-light image enhancement (*i.e.*, RetinexNet [38], SNRNet [42], and RetinexFormer [6]) tend to increase the overall image brightness, which further buries the details of clouds (Figure 8 and Figure 9 (b, d, and f)). LCDP [30] is designed to deal with the coexistence of over- and under-exposure; however, the sky regions after the enhancement tend to have purplish colors (Figure 8 and Figure 9 (e)). Similar problems can be found in the other case. In contrast, our method can enhance these images with better details and colors (Figure 8 and Figure 9 (i, s, and C)).

Figure 10 further shows our visual results of real-world images from the Flickr 30K [47] dataset. While existing methods tend to produce obvious color artifacts after their enhancements, our method can produce results of better visual quality.

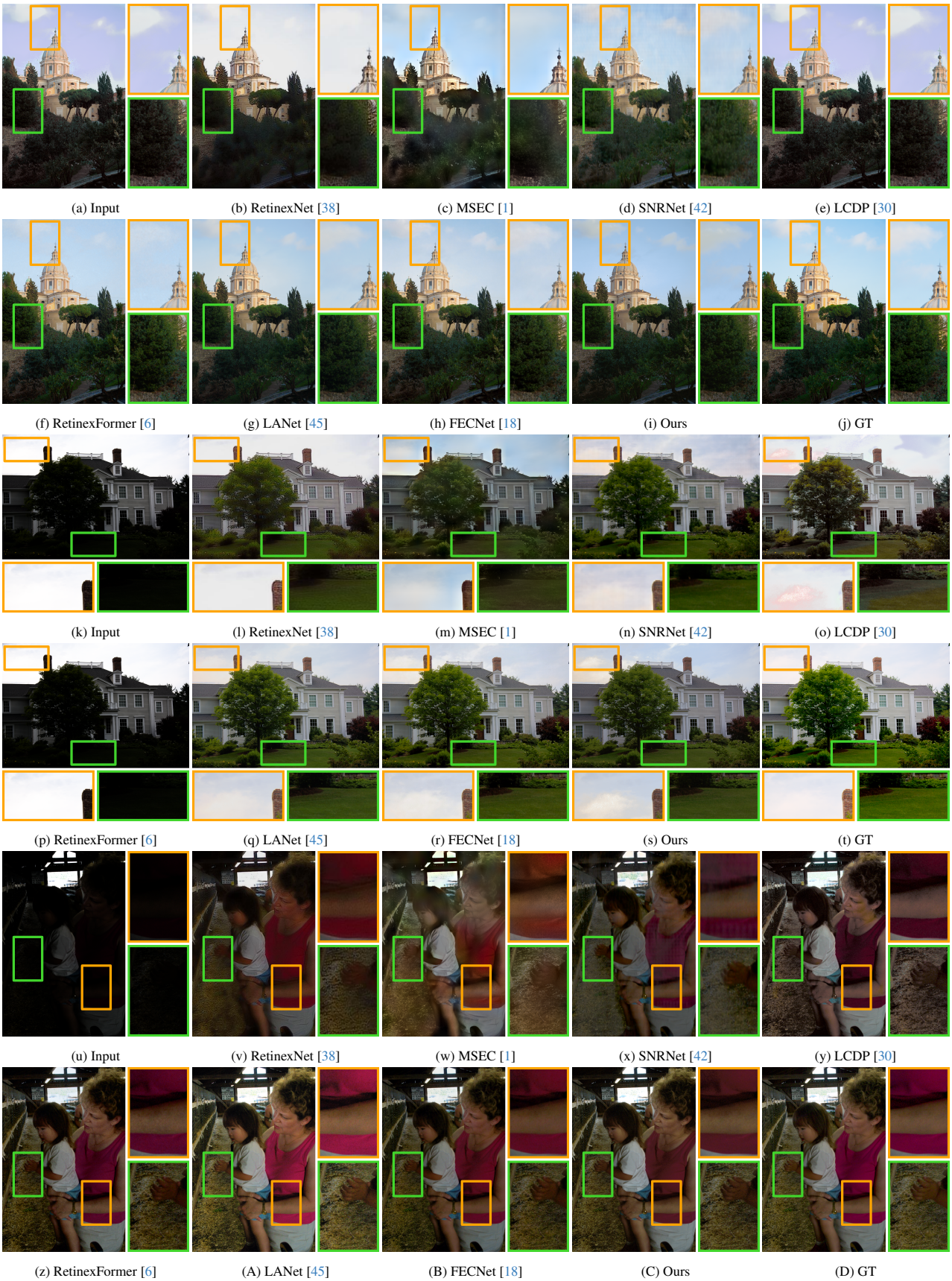
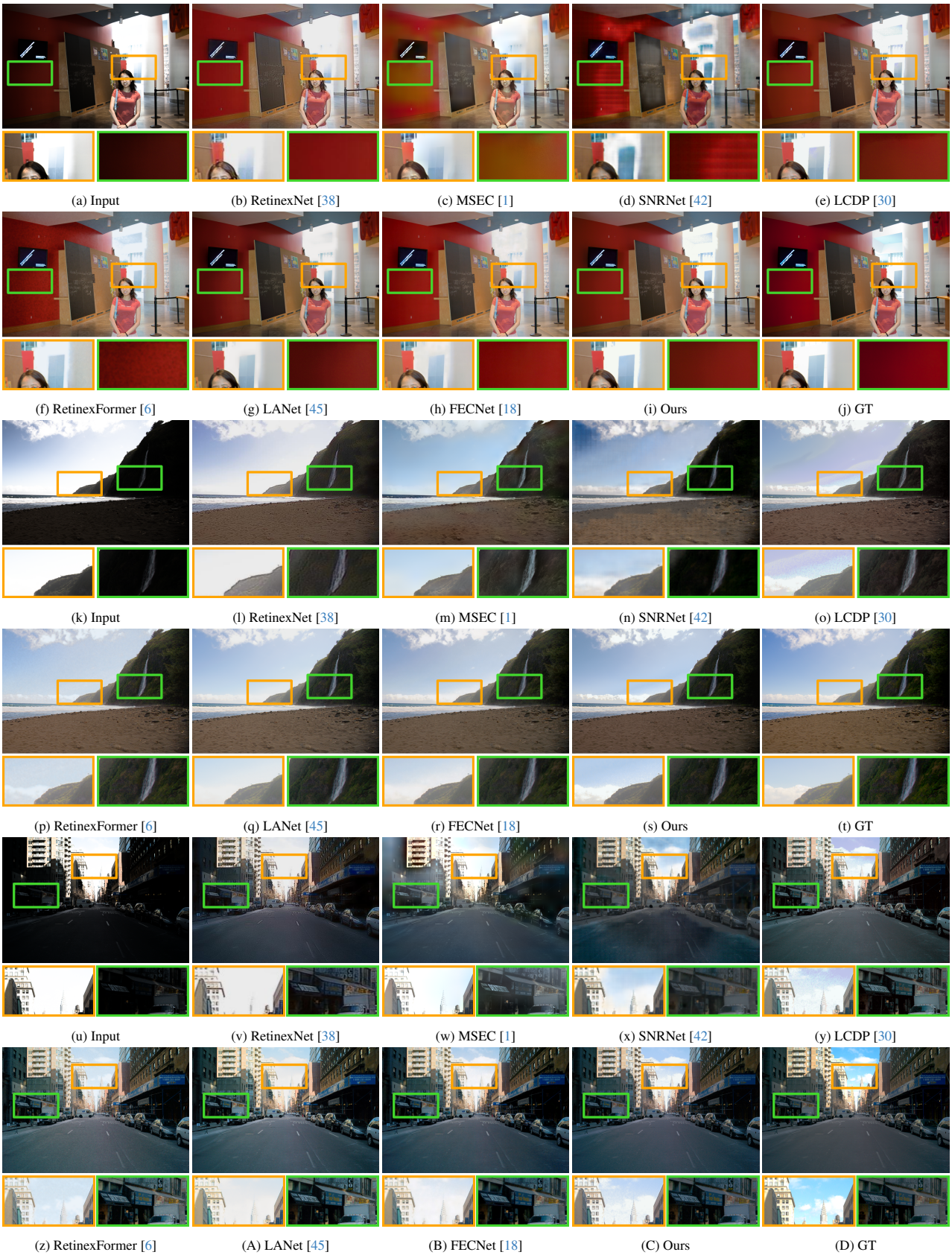


Figure 8. Visual comparison between our method and state-of-the-art methods on LCDP [30] dataset, which has images with both over-exposure and under-exposure.



(a) Input (b) RetinexNet [38] (c) MSEC [1] (d) SNRNet [42] (e) LCDP [30]

(f) RetinexFormer [6] (g) LANet [45] (h) FECNet [18] (i) Ours (j) GT

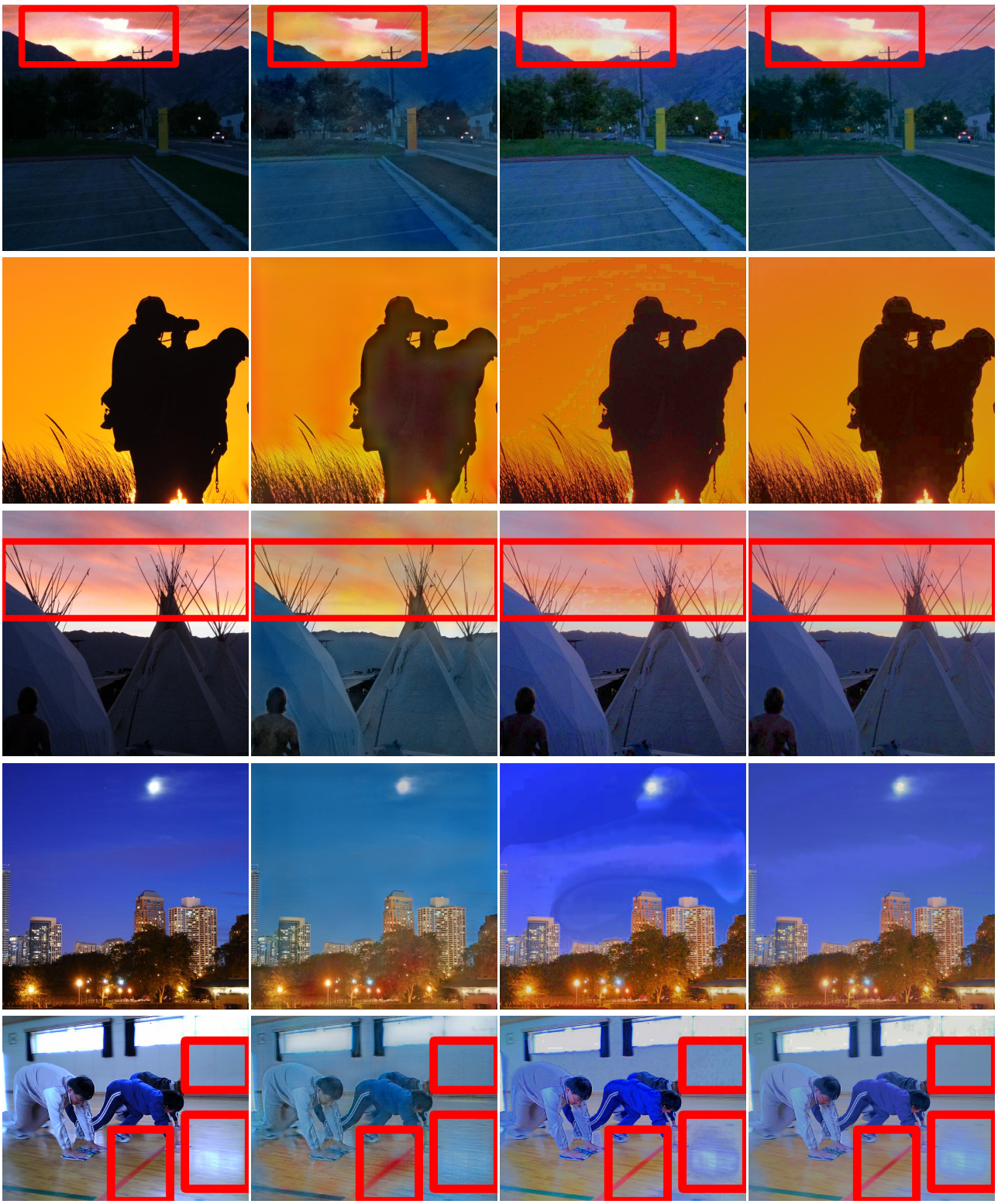
(k) Input (l) RetinexNet [38] (m) MSEC [1] (n) SNRNet [42] (o) LCDP [30]

(p) RetinexFormer [6] (q) LANet [45] (r) FECNet [18] (s) Ours (t) GT

(u) Input (v) RetinexNet [38] (w) MSEC [1] (x) SNRNet [42] (y) LCDP [30]

(z) RetinexFormer [6] (A) LANet [45] (B) FECNet [18] (C) Ours (D) GT

Figure 9. Visual comparison between our method and state-of-the-art methods on LCDP [30] dataset, which has images with both over-exposure and under-exposure.



Input MSEC [1] LCDP [30] Ours

Figure 10. Visual results of real-world examples from the Flickr 30K [47] dataset. Our method produces better visual results.