Luminance-aware Color Transform for Multiple Exposure Correction

Jong-Hyeon Baek[1†]   DaeHyun Kim[1†]   Su-Min Choi[2]   Hyo-jun Lee[1]   Hanul Kim[3]   Yeong Jun Koh[1*]

[1]Chungnam National University   [2]42dot Inc.

[3]Seoul National University of Science and Technology

whdgusdl97@gmail.com, seven776484@gmail, sumin.choi@42dot.ai,

gywns6287@gmail.com, hukim@seoultech.ac.kr, yjkoh@cnu.ac.kr

Abstract

Images captured with irregular exposures inevitably present unsatisfactory visual effects, such as distorted hue and color tone. However, most recent studies mainly focus on underexposure correction, which limits their applicability to real-world scenarios where exposure levels vary. Furthermore, some works to tackle multiple exposure rely on the encoder-decoder architecture, resulting in losses of details in input images during down-sampling and up-sampling processes. With this regard, a novel correction algorithm for multiple exposure, called luminance-aware color transform (LACT), is proposed in this study. First, we reason the relative exposure condition between images to obtain luminance features based on a luminance comparison module. Next, we encode the set of transformation functions from the luminance features, which enable complex color transformations for both overexposure and underexposure images. Finally, we project the transformed representation onto RGB color space to produce exposure correction results. Extensive experiments demonstrate that the proposed LACT yields new state-of-the-arts on two multiple exposure datasets. Code is available at `https://github.com/whdgusdl48/LACT`.

1. Introduction

Exposure is a crucial component in image formation. So, images captured with improper exposures usually have unsatisfactory visual effects, such as distorted hue and brightness. Figure 1 shows examples of images captured with overexposure and underexposure, which can cause overly bright or darkened regions, leading to significant degradation of image quality. Many methods [3, 21, 1, 22, 8, 23, 41, 20] have been proposed to correct exposure-related issues. However, most of them have focused on correcting

---

*Corresponding author

†These authors contributed equally to this work.

either overexposure or underexposure, limiting their applicability to real-world scenarios with various exposure levels. Some studies [2, 14, 35, 15] have attempted to address both overexposure and underexposure cases, leveraging the U-Net [31] like architecture. While these methods have shown improved performance, they still suffer from limitations such as the down-sampling and up-sampling processes in their approaches, resulting in the loss of image details.

The transformation function-based approach has gained much attention in image enhancement [10, 20, 19, 39]. This approach estimates transformation functions for color channels and performs color transformation based on the predicted functions. These methods are free from the down-sampling and up-sampling processes, thereby preserving the image details. However, they are still vulnerable to multiple exposures, since they do not exploit exposure information explicitly. Also, as shown in Figure 1, a single transformation function may not be sufficient to address both overexposed or underexposed images, particularly those containing saturated regions.

In this work, we present a novel approach, named luminance-aware color transform (LACT), to tackle the problem of multiple exposure correction. The proposed method leverages two key insights - the critical role of luminance for correcting images with multiple exposure levels, and the need to overcome the limitation of single intensity transformation function. To this end, the proposed method extracts luminance features, which encode exposure levels in input images, by exploiting the luminance order relationship between two randomly sampled input images. The luminance features are used in the luminance-aware color transform to estimate the sets of transformation functions from the local structure, enabling complex color transform for both overexposure and underexposure images. Then, the proposed method generates a multi-channel representation for exposure correction using the set of transformation functions. Finally, a post-processing module projects this multi-channel representation onto the RGB color space to produce correction results. Extensive experiments on ME [2] and

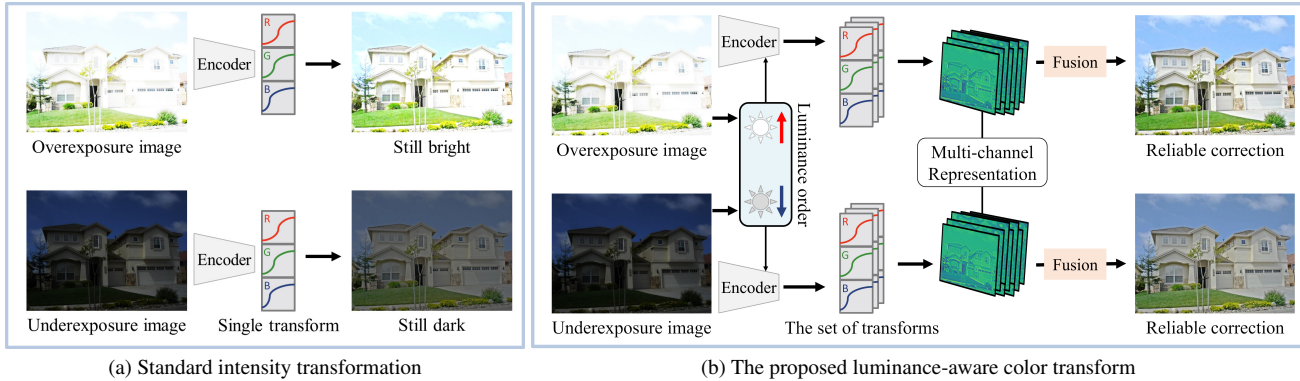|	
(a) Standard intensity transformation	(b) The proposed luminance-aware color transform

Figure 1. Examples of multiple exposure correction with the standard intensity transformation and the proposed luminance-aware color transformation.

SICE [6] datasets demonstrate that the proposed method significantly outperforms state-of-the-arts [14, 15] in multiple exposure correction.

The main contributions of our work are three folds:

- We propose the luminance comparison module that leverages the luminance order relationship to extract the effective luminance feature.

- We propose the luminance-aware color transform based on the luminance feature to obtain reliable correction results for both underexposure and overexposure images.

- We experimentally show the effectiveness of our approach in multiple exposure correction. Our method surpasses recent state-of-the-arts on the ME [2] and SICE [6] datasets.

2. Related Work

2.1. Exposure correction

Early studies in exposure correction mainly aimed at enhancing the global contrast of an input image. Histogram equalization [28] and its modifications [1, 30, 27] are a well-known technique that modifies an image's histogram to stretch its limited contrast. Another popular method is the power-law (gamma) transformation [9, 16, 28], which maps input pixel values to output pixel values using predefined transformations. Retinex theory [21]-based methods [11, 18, 25, 36] decompose exposure correction into illumination enhancement and reflectance regularization.

Recently, deep learning-based methods [7, 41, 40, 38] have been proposed for exposure correction. For instance, SID [7] trains a fully-convolutional network to process low-light sensing images. In [38], frequency decomposition is adopted for underexposure correction. DBRN [40] decomposes image features into different bands and recursively recomposes them to enhance underexposed images. STAR [41] introduces a structure-aware lightweight

Transformer [34] to extract long-short range contexts of low light images. Another approach to exposure correction [19, 20, 10] enhances irregular colors based on an intensity transformation function. Zero-DCE [10] formulates exposure correction as a task of image-specific curve estimation with deep neural networks. GEN-LEN [20] applies global intensity transformation functions and then locally revises the global correction with spatial filtering. RCTN [19] implements color transformation to enhance low-light images based on the similarity between the input image and its representative colors. However, most of these methods focus on underexposure correction, which limits their applications for various exposure levels.

2.2. Multiple exposure correction

Afifi *et al*. [2] introduced the ME [2] dataset, which includes over 20,000 images with multiple exposure scenes. This has led to lots of attempts [26, 2, 5, 15, 35] to correct exposed images in multiple exposure levels. Wang *et al*. [35] proposed the local color distribution embedded module to detect overexposed and underexposed regions based on the guidance of the local color distributions. ENC [14] extents the previous underexposure correction [7, 40] to the multiple exposures one by adding a normalization module that embeds multiple exposure features to the exposure-invariant space. FECNet [15] converts the different exposure images into frequency properties, which provide detailed information about lightness and structure for the exposure correction networks. CMEC [26] applies stacked self-attention layers [34] to model interactions between pixels with a diverse range of exposures. These methods [26, 2, 5, 15, 35], leverage U-Net [31] like architecture and have shown promising performance. However, the down-sampling and up-sampling processes in their approaches may result in the loss of image details. In contrast, the proposed LACT aims to estimate transformation functions that are free from these processes, preserving image details while correcting unnatural exposures.
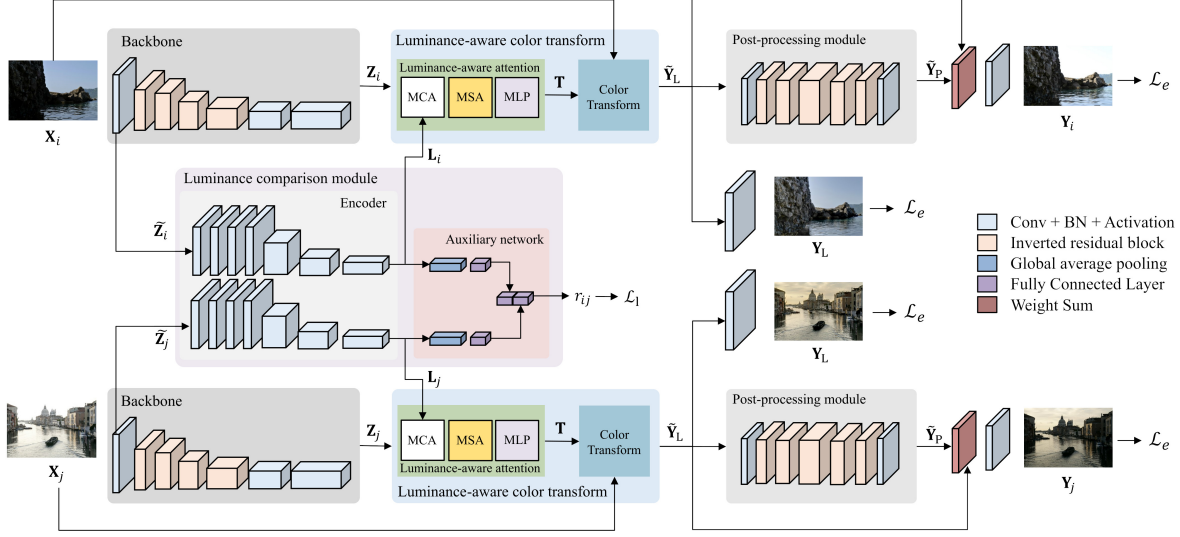
Figure 2. Overview of the training process of the proposed luminance-aware multiple exposure correction network.

3. Proposed Method

Let $\mathbf{X} \in \mathbb{R}^{H \times W \times 3}$ be an input RGB image captured with either underexposure or overexposure settings, where H and W denote the image height and width, respectively. The aim of multiple exposure correction is to restore the high-quality image $\mathbf{Y} \in \mathbb{R}^{H \times W \times 3}$ by compensating for exposure errors in the input image $\mathbf{X}$. However, due to the distinct nature of underexposure and overexposure, different correction procedures are necessary for each exposure case. Therefore, we present a multiple exposure correction network based on the luminance-aware color transform (LACT), which adaptively enhances the input image by considering its exposure condition explicitly.

Figure 2 illustrates the overall training process of the luminance-aware multiple exposure correction network, which consists of four main components: backbone, luminance comparison module, luminance-aware color transform, and post-processing module. The backbone network encodes the input image into image features. The luminance comparison module takes early features of the backbone and extracts luminance features to exploit the exposure conditions of the input image. The luminance-aware color transform combines the luminance features and the image features to estimate the set of transformation functions for each color channel. Then, it performs multi-channel color transformation based on the transformation functions to correct exposure in the input image. Finally, the post-processing module further enhances the image to yield visually pleasing results.

3.1. Backbone

Table 1 presents the detailed architecture of the backbone, which consists of three convolution blocks and four

Stage	Operations	Outputs
0	Resize	$h \quad \times w \quad \times 3$
1	Conv-block, 5×5	$h/2 \quad \times w/2 \quad \times 16$
2	IR-block, 5×5	$h/4 \quad \times w/4 \quad \times 40$
3	IR-block, 5×5	$h/8 \quad \times w/8 \quad \times 40$
4	IR-block, 5×5	$h/16 \times w/16 \times 80$
5	IR-block, 5×5	$h/32 \times w/32 \times 112$
6	Conv-block, 3×3	$h/64 \times w/64 \times 128$
7	Conv-block, 1×1	$h/64 \times w/64 \times 256$

Table 1. Specification of the backbone architecture. Conv-block contains a convolution operation, a batch normalization [17], and a swish activation [29]. IR-block is an inverted residual block [13].

inverted residual blocks [13]. Each convolution block sequentially includes a convolution operation, batch normalization [17], and a swish activation [29]. Given the input image $\mathbf{X} \in \mathbb{R}^{H \times W}$, the backbone first resizes it to have the fixed spatial resolution of $h \times w$, where h and w the resized height and width. The backbone then feeds the resized input through a single convolution block to extract an intermediate image feature $\tilde{\mathbf{Z}} \in \mathbb{R}^{\frac{h}{2} \times \frac{w}{2} \times \tilde{c}}$, where $\tilde{c}$ is the feature dimension. Next, the backbone transforms the intermediate features into an image feature $\mathbf{Z} \in \mathbb{R}^{\frac{h}{64} \times \frac{w}{64} \times c}$ using four inverted residual blocks (IR-block) and two convolution blocks, where c is the feature dimension.

3.2. Luminance comparison module

For effective multiple exposure correction, it needs to analyze the exposure levels of input images. However, it is hard to directly identify exposure levels from arbitrary images. In contrast, luminance information, which is easily derived from images, can be an important cue for repre-

senting the various exposure levels of images. Thus, we design the proposed network to extract effective luminance features from input images. Inspired by order learning [24], we compare two images to estimate which image is brighter or darker through the luminance comparison module. The proposed luminance comparison module contains the encoder, which is composed of seven convolutional blocks, to extract the luminance feature $\mathbf{L} \in \mathbb{R}^{\frac{h}{16} \times \frac{w}{16} \times c}$.

Figure 2 illustrates the proposed training strategy based on the luminance comparison module to relate the luminance feature to exposure levels. Specifically, two different training images, $\mathbf{X}_i$ and $\mathbf{X}_j$, are randomly sampled and their features $\tilde{\mathbf{Z}}_i$ and $\tilde{\mathbf{Z}}_j$ are fed into the encoder in the luminance comparison module to obtain the luminance features $\mathbf{L}_i$ and $\mathbf{L}_j$, respectively. We then employ an auxiliary network to estimate the order relationship between $\mathbf{X}_i$ and $\mathbf{X}_j$. The auxiliary network takes $\mathbf{L}_i$ and $\mathbf{L}_j$, and sequentially applies a global average pooling, a fully connected layer, and concatenation operations to them. Finally, from the concatenated feature, the auxiliary network predicts the order relationship $r_{ij} \in [0, 1]$ through a fully connected layer with sigmoid activation, where r_{ij} indicates which image is brighter or darker. In this work, the ground-truth order relationship $\hat{r}_{ij}$ is set to 1 when $\mathbf{X}_i$ is brighter than $\mathbf{X}_j$, while $\hat{r}_{ij}$ is set to 0 when $\mathbf{X}_i$ is darker than $\mathbf{X}_j$. The luminance comparison module is trained to minimize the cross-entropy loss betweem r_{ij} and $\hat{r}_{ij}$. Note that the auxiliary network is used in training only, which requires two input images. In contrast, only one input image is used during inference, since its luminance feature can be extracted from the encoder in the luminance comparison module.

3.3. Luminance-aware color transform

To obtain reliable correction results for both underexposure and overexposure images, the LACT module includes a luminance-aware attention block to effectively convey the luminance-related information to the image. The luminance-aware attention block explicitly interweaves the luminance feature $\mathbf{L}$ into the image feature $\mathbf{Z}$ using the multi-head cross-attention (MCA) mechanism and then estimates transformation functions based on the standard multi-head self-attention (MSA) [34]. MCA is similar to MSA, but it takes the queries and key-value pairs from different sources. Here, we linearly transform the image feature $\mathbf{Z}$ into the query and the luminance feature $\mathbf{L}$ into the key and value.

Let $n = \frac{h}{64} \cdot \frac{w}{64}$ be the spatial dimension of the image feature. Given $\mathbf{Z}$ and $\mathbf{L}$, we define the luminance-aware image feature $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \cdots \mathbf{f}_n] \in \mathbb{R}^{n \times c}$ as

$$\mathbf{F} = \mathsf{LN}(\mathbf{Z}^* + \mathsf{MCA}(\mathbf{Z}^*, \mathbf{L}^*)) \tag{1}$$

where $\mathsf{LN}(\cdot)$ refers to a layer normalization [4], and $\mathbf{Z}^* \in \mathbb{R}^{n \times c}$ and $\mathbf{L}^* \in \mathbb{R}^{(\frac{h}{16} \cdot \frac{w}{16}) \times c}$ are the reshaped image and

luminance features, respectively.

It is noteworthy that we take into consideration the local structure of the input image since $\mathbf{f}_m$ relates to the m-th position of the image feature. Underexposed or overexposed images may contain locally saturated regions, where most pixel intensities in the specific color channel are distributed around 0 or 255. These regions may require different transformation functions to enhance narrow ranges of input intensities. Thus, it is essential to consider these local structures, which spatially adjust image regions, for more accurate multiple exposure correction.

Next, we estimate the multi-channel color transformation set $\mathbf{T} \in \mathbb{R}^{n \times 256}$ by

$$\tilde{\mathbf{T}} = \mathsf{LN}(\mathbf{F} + \mathsf{MSA}(\mathbf{F})) \tag{2}$$
$$\mathbf{T} = \mathsf{MLP}(\tilde{\mathbf{T}}). \tag{3}$$

Here, $\mathsf{MLP}(\cdot)$ includes three fully connected layers with GELU non-linearity [12]. The estimated set is decomposed into three groups of channel-wise intensity transformation functions:

$$\mathbf{T} = [\mathbf{T}_{\mathrm{r}}, \mathbf{T}_{\mathrm{g}}, \mathbf{T}_{\mathrm{b}}] \tag{4}$$

where $\mathbf{T}_{\mathrm{r}} \in \mathbb{R}^{256 \times \frac{n}{3}}$, $\mathbf{T}_{\mathrm{g}} \in \mathbb{R}^{256 \times \frac{n}{3}}$, $\mathbf{T}_{\mathrm{b}} \in \mathbb{R}^{256 \times \frac{n}{3}}$ are the set of transformation functions for red, green, blue channels, respectively. Thus, there are $\frac{n}{3}$ transformation functions for each color channel.

We perform channel-wise intensity transform to correct the exposure of input image $\mathbf{X}$ using $\mathbf{T}$. For red channel, for example, each column vector $\mathbf{t}_{\mathrm{r}} = [t_0, t_1, \ldots, t_{255}]^T$ in $\mathbf{T}_{\mathrm{r}}$ is regarded as a transformation function, which maps k in $\mathbf{X}$ to t_k in the output image. We perform channel-wise intensity transform for all transformation functions in $\mathbf{T}_{\mathrm{r}}$, and then stack the outputs to obtain a multi-channel enhanced feature $\tilde{\mathbf{Y}}_{\mathrm{r}} \in \mathbb{R}^{H \times W \times \frac{n}{3}}$ for red channel. After repeating this process for all color channels, the luminance-aware color transform module provides the multi-channel enhanced feature, which is given by

$$\tilde{\mathbf{Y}}_{\mathrm{L}} = [\tilde{\mathbf{Y}}_{\mathrm{r}}, \tilde{\mathbf{Y}}_{\mathrm{g}}, \tilde{\mathbf{Y}}_{\mathrm{b}}]. \tag{5}$$

3.4. Post-processing module

The multi-channel representation $\tilde{\mathbf{Y}}$ is obtained by exploiting various transform functions from the local structure of the input image. However, it should be projected onto the RGB color space to restore a high-quality image. As depicted in Figure 2, the post-processing module is designed at the end of the network, which yields a refined correction feature $\tilde{\mathbf{Y}}_{\mathrm{P}}$ using sequential stages. Finally, it produces a final correction image $\mathbf{Y}$ by a weighted sum of $\tilde{\mathbf{Y}}_{\mathrm{L}}$ and $\tilde{\mathbf{Y}}_{\mathrm{P}}$, which is given by

$$\mathbf{Y} = \mathsf{Conv}(\alpha_1 \tilde{\mathbf{Y}}_{\mathrm{L}} + \alpha_2 \tilde{\mathbf{Y}}_{\mathrm{P}}) \tag{6}$$

Stage	Operations	Outputs
0	Input	$H \times W \times n$
1	Conv-block, 5×5	$H \times W \times 16$
2	IR-block, 5×5	$H \times W \times 24$
3	IR-block, 5×5	$H \times W \times 40$
4	IR-block, 5×5	$H \times W \times 80$
5	IR-block, 5×5	$H \times W \times 24$
6	IR-block, 5×5	$H \times W \times 16$
7	Conv-block, 5×5	$H \times W \times n$
8	Weighted sum	$H \times W \times n$
9	Conv-block, 1×1	$H \times W \times 3$

Table 2. Specification of the post-processing module architecture.

where α_1 and α_2 are learnable weights, and Conv represents the convolution block. Table 2 specifies the detailed architecture of the post-processing module.

3.5. Loss functions

Luminance ordering loss. We use the luminance ordering loss $\mathcal{L}_l$ to train the luminance comparison module. To define the luminance ordering loss, we first convert each input RGB image $\mathbf{X}_i$ into a grayscale image $\mathbf{G}_i$:

$$\mathbf{G}_i = 0.299\mathbf{X}_{i,r} + 0.587\mathbf{X}_{i,g} + 0.114\mathbf{X}_{i,b} \qquad (7)$$

where $\mathbf{X}_{i,r}$, $\mathbf{X}_{i,g}$, $\mathbf{X}_{i,b}$ denote intensity images for red, green, and blue channels, respectively. Next, we randomly sample two images $\mathbf{X}_i$ and $\mathbf{X}_j$ and set the ground-truth order relationship $\hat{r}_{ij}$ between $\mathbf{X}_i$ and $\mathbf{X}_j$ as

$$\hat{r}_{ij} = \begin{cases} 1 & \text{if } \bar{\mathbf{G}}_i - \bar{\mathbf{G}}_j \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad (8)$$

where $\bar{\mathbf{G}}_i$ is an average of gray values in $\mathbf{G}_i$. Then, we define the luminance ordering loss $\mathcal{L}_l$:

$$\mathcal{L}_l = \hat{r}_{ij} \cdot \log(r_{ij}) + (1 - \hat{r}_{ij}) \cdot \log(1 - r_{ij}) \qquad (9)$$

as the binary cross-entropy between the predicted order relationship r_{ij} and the ground-truth $\hat{r}_{ij}$.

Exposure correction loss. We design the exposure correction loss $\mathcal{L}_e$ to penalize discrepancies between the predicted exposure correction image $\mathbf{Y}$ and the ground-truth image $\hat{\mathbf{Y}}$ in various aspects. $\mathcal{L}_e$ is defined as a weighted sum of three loss terms: reconstruction loss $\mathcal{L}_c$, frequency loss $\mathcal{L}_f$, and perceptual loss $\mathcal{L}_p$:

$$\mathcal{L}_e = \mathcal{L}_c + \lambda_f \mathcal{L}_f + \lambda_p \mathcal{L}_p. \qquad (10)$$

Here, we set the balancing parameters λ_f and λ_p to 0.5 and 0.05 to scale the relative importance of each term. The reconstruction loss $\mathcal{L}_c$ measures the l_1 distance between

the predicted image $\mathbf{Y}$ and its ground-truth $\hat{\mathbf{Y}}$ in the color space:

$$\mathcal{L}_c = \|\mathbf{Y} - \hat{\mathbf{Y}}\|_1. \qquad (11)$$

The frequency loss $\mathcal{L}_f$, on the other hand, quantifies the difference between the predicted and ground-truth ones in the frequency space using the 2D fast Fourier transform $\mathcal{F}(\cdot)$:

$$\mathcal{L}_f = \|\mathcal{F}(\mathbf{Y}) - \mathcal{F}(\hat{\mathbf{Y}})\|_1. \qquad (12)$$

$\mathcal{L}_f$ is crucial for preserving image details. Finally, the perceptual loss $\mathcal{L}_p$ measures the dissimilarity between the predicted and ground-truth enhanced images in the feature space defined by the embedding function $\phi^k(\cdot)$, which is the output of the kth layer in VGG-16 network [33] pre-trained on the ImageNet dataset [32]:

$$\mathcal{L}_p = \sum_{k=2,4,6} \|\phi^k(\mathbf{Y}) - \phi^k(\hat{\mathbf{Y}})\|_1 \qquad (13)$$

In this work, the informative multi-channel enhanced image $\tilde{\mathbf{Y}}_L$ is essential to achieve high-quality results. To ensure this, we apply the exposure correction loss $\mathcal{L}_e$ not only to the final estimation $\mathbf{Y}$, but also to $\tilde{\mathbf{Y}}_L$. For this purpose, we convert $\tilde{\mathbf{Y}}_L$ into the image domain, $i.e.$ $\mathbf{Y}_L \in \mathbb{R}^{H \times W \times 3}$, using one auxiliary 1×1 convolution. Then, we compute the exposure correction loss between $\mathbf{Y}_L$ and $\hat{\mathbf{Y}}$.

4. Experiments

4.1. Setting

Datasets. For evaluation of the proposed multi-exposure image error correction network, we use two representative multi-exposure datasets: ME [2] and SICE [6]. The ME dataset [2] consists of multi-exposure scenes, where each scene contains 5 exposure-level images and their one ground-truth image. ME is decomposed into 3,535 and 1,181 scenes for training and test sets, respectively. We use the training scenes for training the proposed network and the test scenes for evaluation. The SICE dataset [6] contains 589 scenes with various exposures and corresponding ground truth. For the evaluation of SICE [6], we sample 5 exposure subsets based on the middle exposure subset for each scene. Also, we divide 589 scenes into 472 training and 117 test scenes. We use Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) for the quantitative evaluation in all experiments.

Implementation details. We resize input images to 384×384 and train the proposed network with a batch size of 2 using a single NVIDIA RTX A6000 GPU. We set 20 epochs for the ME dataset and 55 epochs for the SICE dataset. During training, we optimize the networks using the RMSProp optimizer and WarmUpCosine learning rate schedule with an initial learning rate of 3e-4.

Methods	Under		Over		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
RetinexNet [37]	12.13	0.620	10.47	0.595	11.14	0.604
Zero-DCE [10]	14.55	0.588	10.40	0.514	12.06	0.544
GEN-LEN [20]	21.52	0.774	21.28	0.769	21.17	0.763
RCTN [19]	21.73	0.774	21.10	0.766	21.33	0.763
MSEC [2]	20.52	0.812	19.79	0.815	20.35	0.821
CMEC [26]	21.12	0.839	21.88	0.866	21.58	0.855
DRBN-ENC [14]	22.72	0.854	22.11	0.852	22.35	0.853
SID-ENC [14]	22.59	0.842	22.36	0.851	22.45	0.843
FECNet [15]	<u>22.96</u>	<u>0.859</u>	<u>23.22</u>	**0.874**	<u>23.12</u>	<u>0.868</u>
Ours	**23.49**	**0.862**	**23.68**	<u>0.872</u>	**23.57**	**0.869**

Table 3. Quantitative comparison on the ME [2] dataset. The best results are boldfaced, and the second ones are underlined.

(a) Overexposure input	(b) DBRN-ENC	(c) FECNet	(d) Ours	(e) Ground-truth

(f) Underexposure input	(g) DBRN-ENC	(h) FECNet	(i) Ours	(j) Ground-truth

Figure 3. Qualitative comparison on the ME [2] dataset.

Methods	Under		Over		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
GEN-LEN [20]	21.52	<u>0.774</u>	<u>21.28</u>	<u>0.769</u>	21.17	0.763
RCTN [19]	<u>21.73</u>	<u>0.774</u>	21.10	0.766	<u>21.33</u>	<u>0.763</u>
DRBN-ENC [14]	21.39	0.768	20.49	0.758	20.87	0.757
FECNet [15]	21.46	**0.788**	20.33	0.740	21.09	**0.772**
Ours	**22.34**	0.773	**21.76**	**0.771**	**22.02**	**0.772**

Table 4. Quantitative comparison on the SICE [6] dataset. The best results are boldfaced, and the second ones are underlined.

4.2. Comparison with State-of-the-arts

ME dataset. Table 3 compares the proposed method with the existing exposure correction methods [37, 10, 20, 19, 2, 26, 14, 15] on the ME dataset. Following MSEC, Table 3 shows the PSNR and SSIM scores for underexposure and overexposure scenes and their average scores. For comparison, we obtain the results of the existing methods [37, 10, 2, 26, 14, 15] are from [2, 15, 26]. Also, we obtain the results of methods [20, 19] using their published source codes and default settings. RetinexNet [37]

and Zero-DCE [10] provide lower performance since they cannot consider multi-exposure examples. In contrast, the proposed network provides remarkable performance on the ME dataset. As compared with color transform-based methods [20, 19], the proposed method significantly improves the performance by exploiting luminance features, extracted from the luminance comparison module. Also, the proposed method outperforms recent state-of-the-art methods ENC [14] and FECNet [15], which are designed for multi-exposure scenes, for example, by a margin of 0.45 in terms
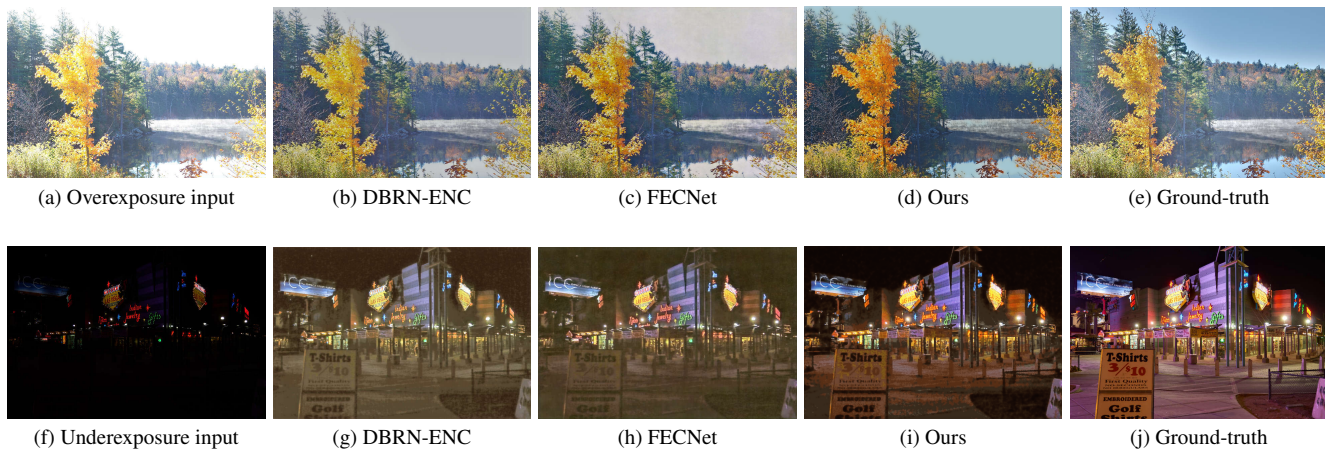
| (a) Overexposure input | (b) DBRN-ENC | (c) FECNet | (d) Ours | (e) Ground-truth |
| (f) Underexposure input | (g) DBRN-ENC | (h) FECNet | (i) Ours | (j) Ground-truth |

Figure 4. Qualitative comparison on the SICE [6] dataset.

Methods	Under PSNR	Under SSIM	Over PSNR	Over SSIM	Average PSNR	Average SSIM
w/o LCM	22.97	0.852	23.41	0.868	23.21	0.864
w/o $\mathcal{L}_l$	23.15	0.855	23.58	0.871	23.40	0.866
w/o OR	23.29	0.857	23.15	0.869	23.21	0.864
Ours	**23.49**	**0.862**	**23.68**	**0.872**	**23.57**	**0.869**

Table 5. Results on the ME [2] dataset with different LCM settings. The best results are boldfaced.

# of IR	Under PSNR	Under SSIM	Over PSNR	Over SSIM	Average PSNR	Average SSIM
0	22.86	0.855	23.39	0.867	23.16	0.861
1	23.21	0.856	23.38	0.867	23.32	0.862
3	23.38	0.861	23.52	0.871	23.42	0.864
5 (Ours)	**23.49**	**0.862**	**23.68**	**0.872**	**23.57**	**0.869**

Table 6. Results on the ME [2] dataset with different numbers of IR blocks in PPM. The best results are boldfaced.

of average PSNR against FECNet [15]. Figure 3 qualitatively compares the proposed method with ENC and FEC-Net on the ME dataset. We see that the proposed method corrects both overexposure and underexposure scenes more faithfully than ENC and FECNet.

SICE dataset. Table 4 shows PSNR and SSIM scores for the proposed method and the existing methods [20, 19, 14, 15] on the SICE dataset. For comparison, we obtain the results of [20, 19, 14, 15] using their published source codes and default settings. The proposed method surpasses both color transform-based methods [20, 19] and multi-exposure correction methods [14, 15]. Figure 4 illustrates exposure correction results of ENC, FECNet, and the proposed method. In these examples, the proposed method enhances both overexposure and underexposure images to have more similar color tones to their ground truth than ENC and FEC-Net.

4.3. Ablation Study

We present a comprehensive analysis of the efficacy of our proposed components: luminance comparison module (LCM); post-processing module (PPM); the number of transformation functions (n_t); loss functions. All ablation studies are performed on the ME [2] dataset.

Luminance comparison module. We conduct experiments in three ways to analyze the contribution of LCM: First,

we remove LCM from the proposed network (w/o LCM). Second, we use layers in LCM, but exclude the luminance loss $\mathcal{L}_l$ (w/o $\mathcal{L}_l$). In other words, layers in LCM are trained with the exposure correction loss $\mathcal{L}_e$ only. Third, we train LCM to predict the brightness difference between the input image and its corresponding ground-truth, instead of minimizing the luminance loss to estimate the order relationship (w/o OR). Table 5 shows that the proposed method achieves the best scores. Additionally, Figure 5 shows the effectiveness of LCM on overexposure and underexposure images. Without LCM, the proposed method cannot achieve reasonable performance in both exposure levels as shown in Figure 5b and Figure 5f. This can support our main idea that the awareness of luminance is critical to multiple exposure correction.

Post-processing module. Table 6 shows results on the ME dataset by varying the numbers of IR blocks in PPM. We see that the deeper PPM yields the better result. Also, the results of the proposed method without PPM (zero IR block) are the same as $\mathbf{Y}_L$, which is the output LACT, and they are higher than the conventional state-of-art method in Table 3. This validates the effectiveness of the proposed LACT.

Loss functions. Table 7 shows the performance evaluation of various combinations of loss terms in the exposure correction loss. The comparison of PSNR and SSIM scores indicates that solely utilizing color loss yields the lowest

| (a) Overexposure input | (b) w/o LCM | (c) Ours | (d) Ground-truth |

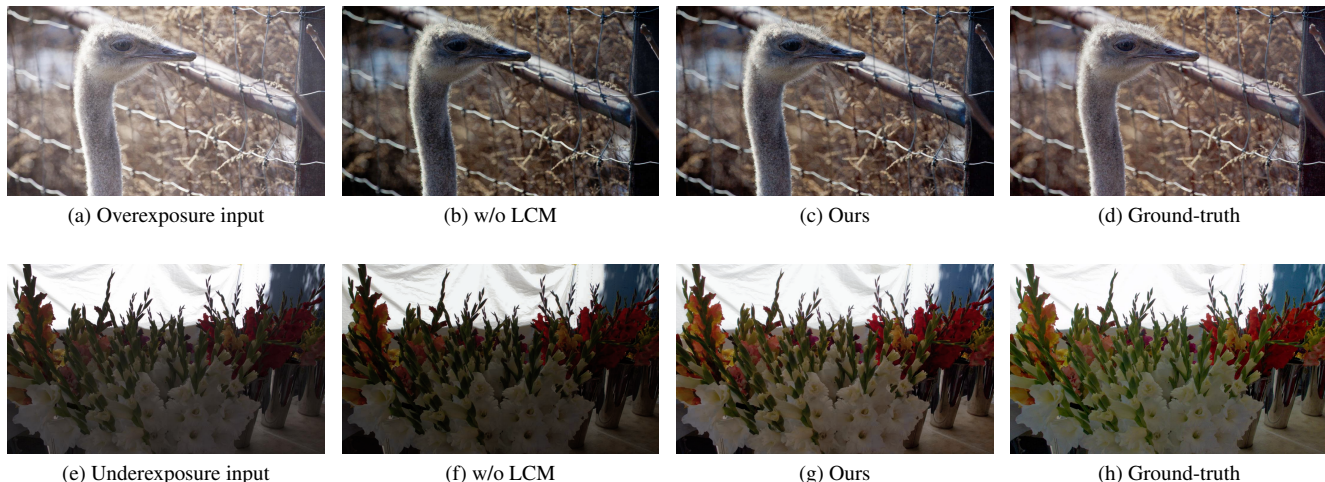| (e) Underexposure input | (f) w/o LCM | (g) Ours | (h) Ground-truth |

Figure 5. Qualitative results on overexposure and underexposure images with different LCM settings.

Methods	Under		Over		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
$\mathcal{L}_c$	23.11	0.848	23.03	0.855	23.07	0.852
$\mathcal{L}_c + \mathcal{L}_p$	23.30	0.849	23.35	0.865	23.33	0.859
$\mathcal{L}_c + \mathcal{L}_p + \mathcal{L}_f$	**23.49**	**0.862**	**23.68**	**0.872**	**23.57**	**0.869**

Table 7. Results on the ME [2] dataset with different loss functions. The best results are boldfaced

Methods	Under		Over		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
w/o aux	23.41	0.858	23.60	0.869	23.49	0.865
All (Ours)	**23.49**	**0.862**	**23.68**	**0.872**	**23.57**	**0.869**

Table 8. Results on the ME [2] dataset depending on additional exposure correction loss. The best results are boldfaced.

n	Under		Over		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
3	23.48	0.861	23.39	0.872	23.44	0.867
36 (Ours)	**23.49**	**0.862**	**23.68**	**0.872**	**23.57**	**0.869**

Table 9. Results for ME [2] dataset with different numbers of intensity transformation functions. The best results are boldfaced.

performance, while the highest performance is obtained by incorporating all loss terms. Notably, the integration of frequency loss term significantly increases the SSIM score by preserving image details. Furthermore, Table 8 demonstrates the impact of an additional exposure correction loss for $\mathbf{Y}_L$. We see that the proposed method outperforms the other one by producing informative multi-channel correction images.

Transformation functions. Table 9 presents an evaluation of the effectiveness of our color transformation method based on multi-channel representation. A comparison is made between our method and an alternative approach that uses a single intensity transformation function for each color channel. The results demonstrate that our method achieves better scores. Moreover, our method exhibits an advantage in correcting overexposure by leveraging the local structures of the input image and estimating appropriate transformation functions for these structures.

5. Conclusion

We proposed the multiple exposure correction network to perform LACT. The key insights of LACT are the critical role of luminance in correcting images with multiple exposures and the need to overcome the limitation of a single intensity transformation function. We developed the luminance comparison module to extract luminance features, which encodes the exposure information. Also, we proposed LACT, which estimates the various transformation functions from the local structure based on the luminance features. LACT converts the input image into multi-channel correction representation and the post-processing module yields the exposure correction results. Experiments on ME and SCIE validated that the proposed method significantly outperforms the existing multiple exposure correction methods.

References

[1] Mohammad Abdullah-Al-Wadud, Md Hasanul Kabir, M Ali Akber Dewan, and Oksam Chae. A dynamic histogram equalization for image contrast enhancement. *IEEE Trans. Consum. Electron.*, 53(2):593–600, 2007.

[2] Mahmoud Afifi, Konstantinos G Derpanis, Bjorn Ommer, and Michael S Brown. Learning multi-scale photo exposure correction. In *CVPR*, pages 9157–9167, 2021.

[3] Tarik Arici, Salih Dikbas, and Yucel Altunbasak. A histogram modification framework and its application for image contrast enhancement. *TIP*, 18(9):1921–1935, 2009.

[4] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

[5] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR*, pages 97–104, 2011.

[6] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *TIP*, 27(4):2049–2062, 2018.

[7] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *CVPR*, pages 3291–3300, 2018.

[8] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *CVPR*, pages 6306–6314, 2018.

[9] Xu Guan, Su Jian, Pan Hongda, Zhang Zhiguo, and Gong Haibin. An image enhancement method based on gamma correction. In *2009 Second international symposium on computational intelligence and design*, volume 1, pages 60–63. IEEE, 2009.

[10] 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*, pages 1780–1789, 2020.

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

[12] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

[13] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *ICCV*, pages 1314–1324, 2019.

[14] 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*, pages 6043–6052, 2022.

[15] Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction network with spatial-frequency interaction. In *ECCV*, pages 163–180. Springer, 2022.

[16] Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *TIP*, 22(3):1032–1041, 2012.

[17] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015.

[18] 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. *TIP*, 6(7):965–976, 1997.

[19] Hanul Kim, Su-Min Choi, Chang-Su Kim, and Yeong Jun Koh. Representative color transform for image enhancement. In *ICCV*, pages 4459–4468, 2021.

[20] Han-Ul Kim, Young Jun Koh, and Chang-Su Kim. Global and local enhancement networks for paired and unpaired image enhancement. In *ECCV*, pages 339–354, 2020.

[21] Edwin H Land. The retinex theory of color vision. *Sci. Am.*, 237(6):108–129, 1977.

[22] Shutao Li, Xudong Kang, and Jianwen Hu. Image fusion with guided filtering. *TIP*, 22(7):2864–2875, 2013.

[23] Jie Liang, Hui Zeng, and Lei Zhang. High-resolution photorealistic image translation in real-time: A laplacian pyramid translation network. In *CVPR*, pages 9392–9400, 2021.

[24] Kyungsun Lim, Nyeong-Ho Shin, Young-Yoon Lee, and Chang-Su Kim. Order learning and its application to age estimation. In *ICLR*, 2020.

[25] Laurence Meylan and Sabine Susstrunk. High dynamic range image rendering with a retinex-based adaptive filter. *TIP*, 15(9):2820–2830, 2006.

[26] NE Nsamp, Zhongyun Hu, and Qing Wang. Learning exposure correction via consistency modeling. In *BMVC*, pages 1–12, 2021.

[27] 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. *Comput. Vis., Graph., Image Process.*, 39(3):355–368, 1987.

[28] Richard E. Woods Rafael C. Gonzalez. Digital image processing, pearson, 2018.

[29] Prajit Ramachandran, Barret Zoph, and Quoc V Le. Searching for activation functions. In *ICLRW*, 2017.

[30] Ali M Reza. 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*, 38:35–44, 2004.

[31] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241, 2015.

[32] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *IJCV*, 115:211–252, 2015.

[33] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2014.

[34] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS*, 2017.

[35] Haoyuan Wang, Ke Xu, and Rynson WH Lau. Local color distributions prior for image enhancement. In *ECCV*, pages 343–359, 2022.

[36] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *TIP*, 22(9):3538–3548, 2013.

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

[38] Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In *CVPR*, pages 2281–2290, 2020.

[39] Canqian Yang, Meiguang Jin, Yi Xu, Rui Zhang, Ying Chen, and Huaida Liu. Seplut: Separable image-adaptive lookup tables for real-time image enhancement. In *ECCV*, pages 201–217. Springer, 2022.

[40] 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 *CVPR*, pages 3063–3072, 2020.

[41] Zhaoyang Zhang, Yitong Jiang, Jun Jiang, Xiaogang Wang, Ping Luo, and Jinwei Gu. Star: A structure-aware lightweight transformer for real-time image enhancement. In *ICCV*, pages 4106–4115, 2021.