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

ECCV 2022

[1]Wang H, Xu K, Lau R W H. Local color distributions prior for image enhancement[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 343-359.

Motivations

(a) Input image (b) MSEC [1] (c) ZeroDCE [14]

(d) RUAS [29] (e) Ours (f) Ground truth

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.

Image enhancement issues:
Photos taken under unfavorable lighting conditions may have over-exposure or under-exposure issues. Often, both over- and under-exposures may occur together in the same image due to unbalanced lighting conditions.

Limitations of existing methods:
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 underexposure problems, these existing methods may not work well.

Contributions

- Exploit the local color distributions (LCDs) to jointly address both over- and under-exposure problems in the input image. 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.
- Propose a dual-illumination estimator to combine both over- and under-illumination maps to enhance the input image.
- Construct a new paired dataset consisting of over 1700 images of diverse, non-uniformly illuminated scenes to facilitate the learning process.

Proposed Dataset

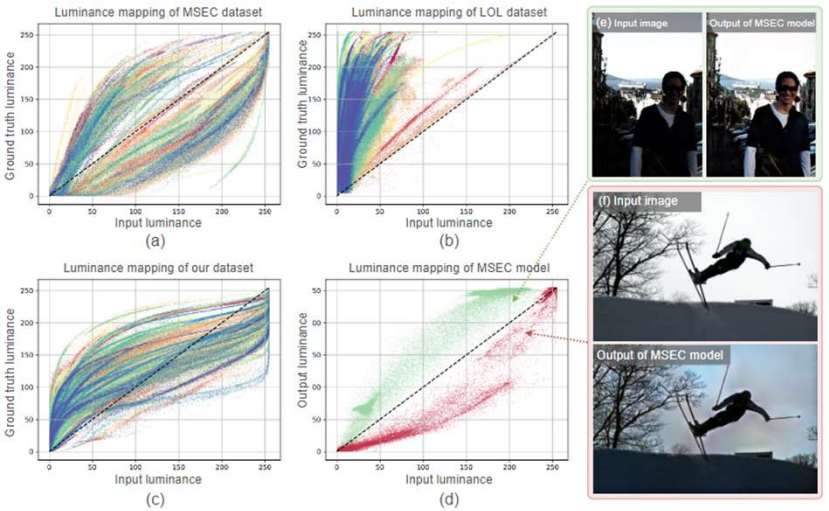


Fig. 2: The input-ground truth luminance mapping curve of (a) MSEC dataset [1], (b) LOL dataset [10], and (c) our dataset. Each cluster represents an image. For a single image, both (a) and (b) contains a single mapping of either brightening or darkening. (d) is the input-output mapping learned by the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

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

$$I'(i,j) = \phi\left[k\left(I(i,j) - 0.5\right) + 0.5\right], \tag{1}$$

Proposed Method--Framework

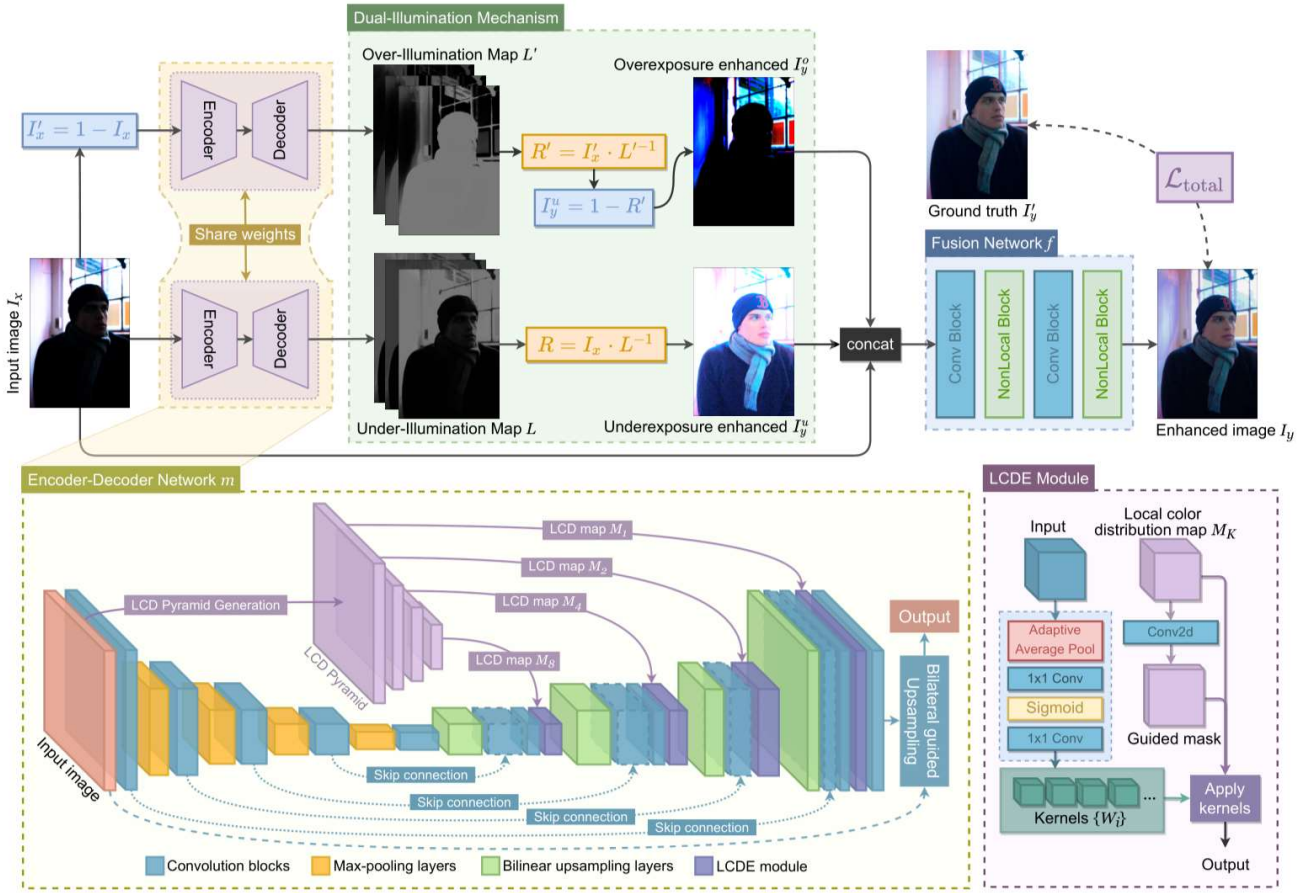
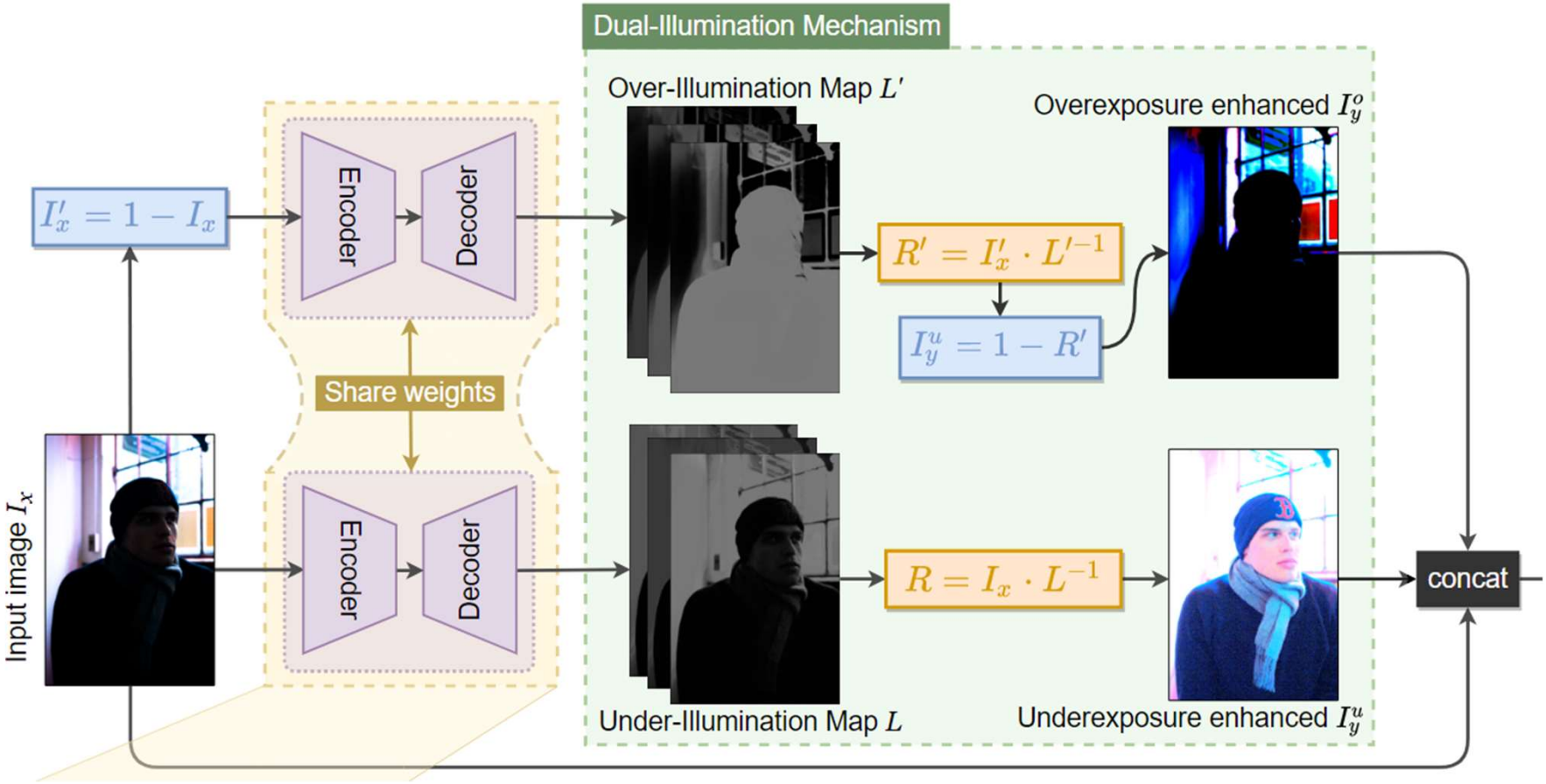
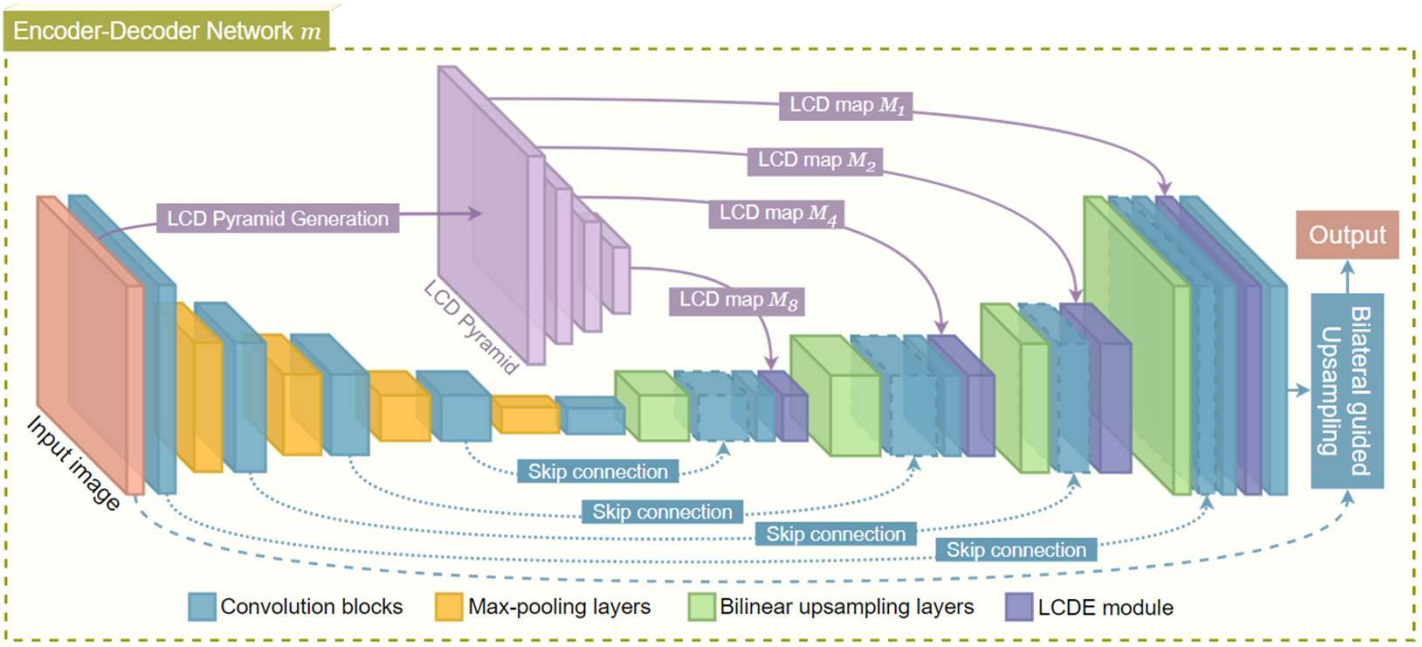


Fig. 5: Overview of our proposed network. It leverages the LCD pyramid with an encoder-decoder architecture for detecting the regions with problematic exposures implicitly, and the for enhancement of the over- and under-exposed regions.

Proposed Method--Dual-illumination Estimation



Proposed Method-- Encoder-Decoder Network & LCD Pyramid Generation



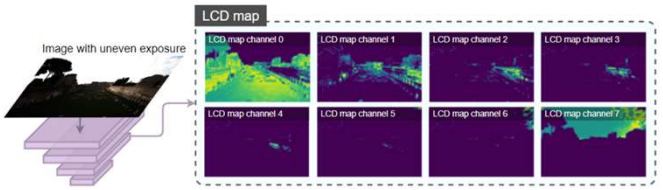
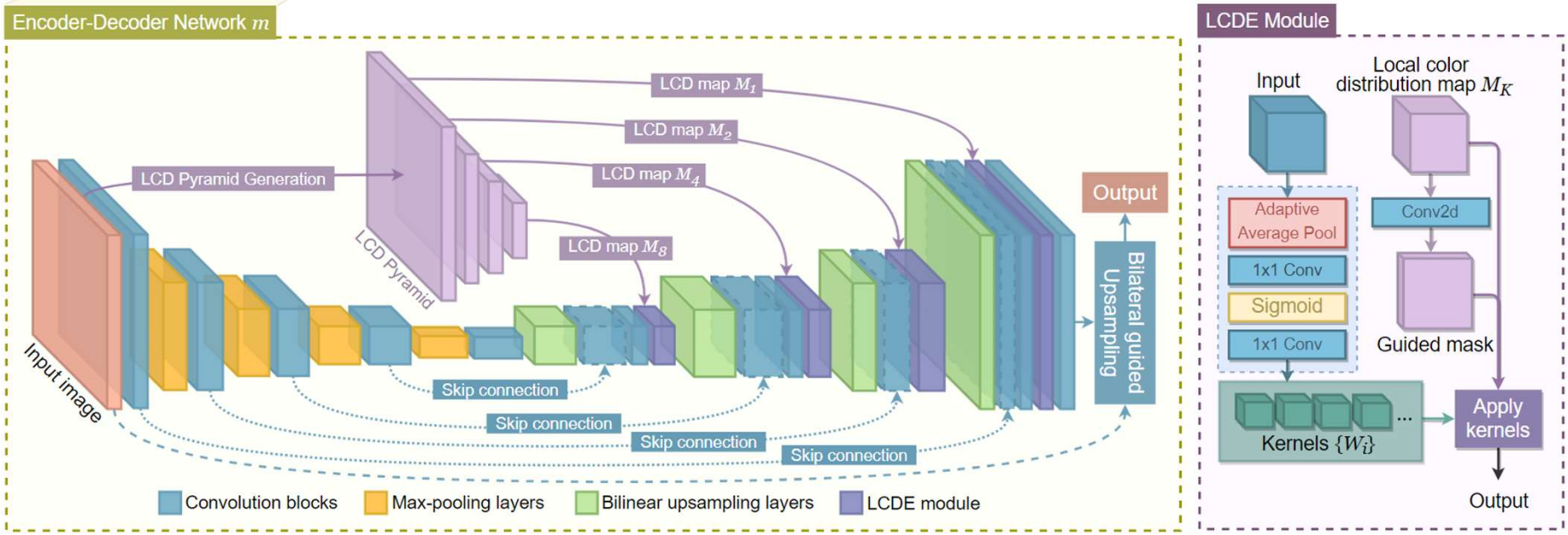


Fig. 6: The visualization of a LCD pyramid layer, taking an over-/under-exposed image as input. Over-/under-exposed regions are implicitly separated along the channel dimension.

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

Proposed Method-- Encoder-Decoder Network & LCDE module



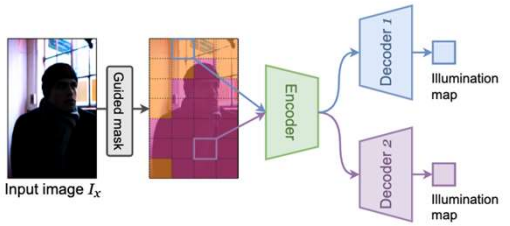
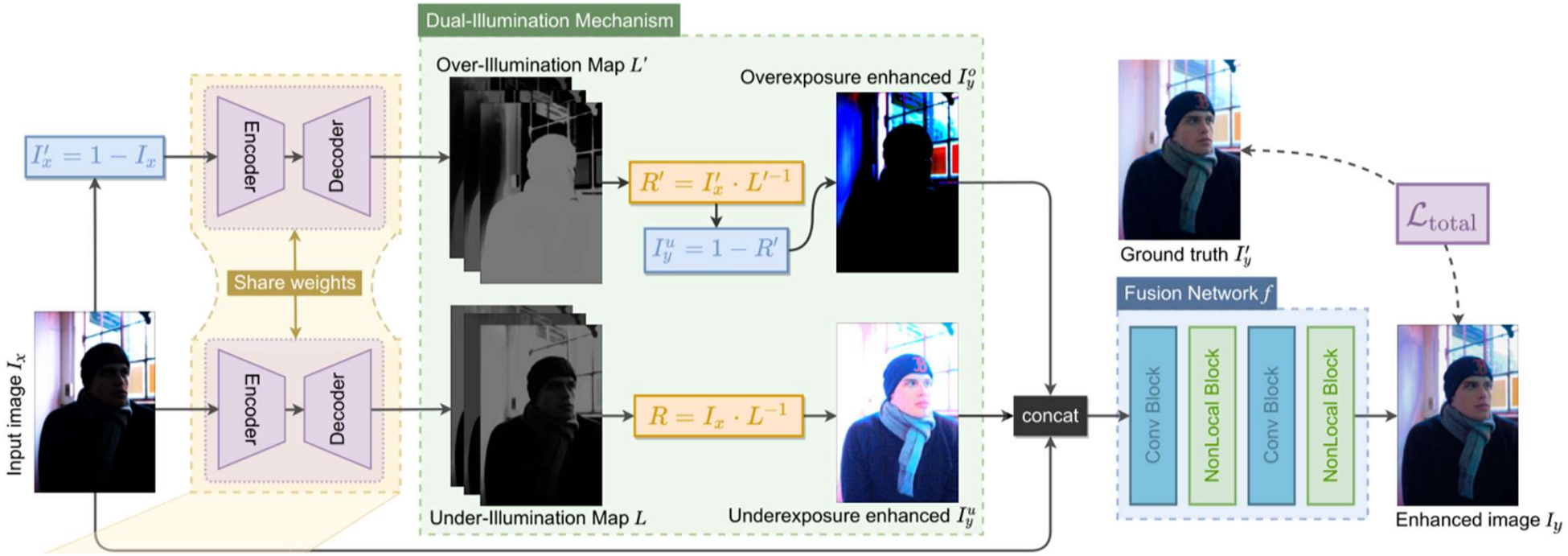


Fig. 7: Inferring the illumination maps helps constrain the learning of the guided mask. Our model implicitly assigns one decoder to each region type for adaptive enhancement.

Proposed Method--Fusion Network



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

Loss Functions

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

$$\mathcal{L}_{\text{mse}} = \frac{1}{hw} ||I_y - I'_y||_2^2$$

$$\mathcal{L}_{\cos} = \frac{1}{hw} \sum_p s \left(1 - \frac{I_{y_p} \cdot I'_{y_p}}{||I_{y_p}||_2 \cdot ||I'_{y_p}||_2} \right)$$

$$\begin{aligned} \mathcal{L}_{\text{tv1}} &= \mathcal{L}_{\text{tv}}(L, I_x) \\ &= \sum_p \left(\frac{|\nabla_h L_p|^2}{|\nabla_h \log(I_{x_p})|^\alpha} + \frac{|\nabla_v L_p|^2}{|\nabla_h \log(I_{x_p})|^\alpha} \right) \end{aligned}$$

Experiments--Visual Comparisons

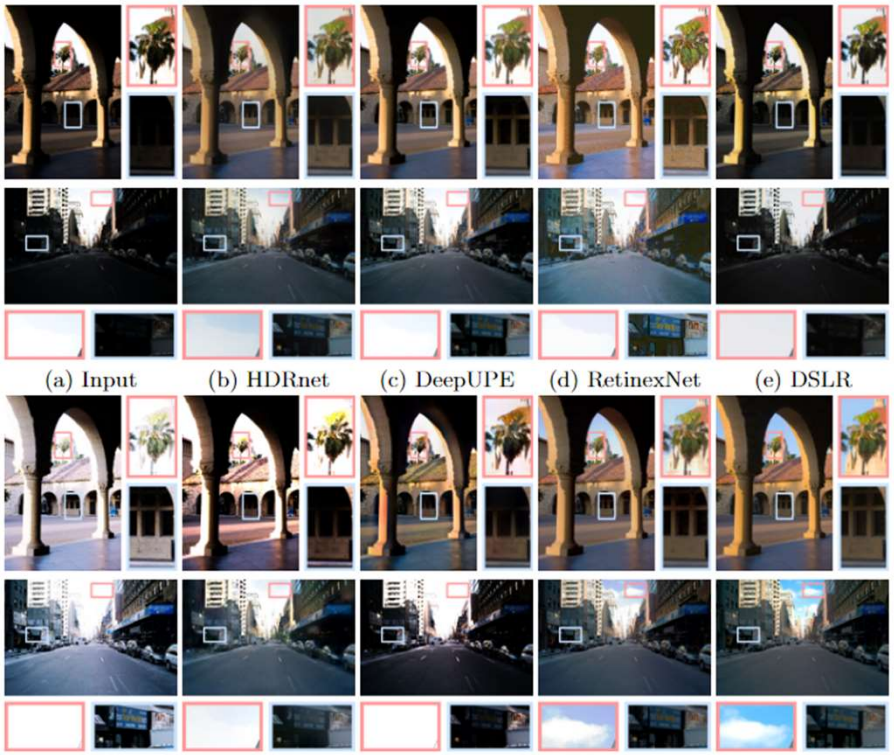


Fig. 8: Visual comparison of over-/under-exposed images from our dataset. Our model reconstructs the details in the over-exposed regions (sky and tower) as well as the under-exposed regions (wall and door).

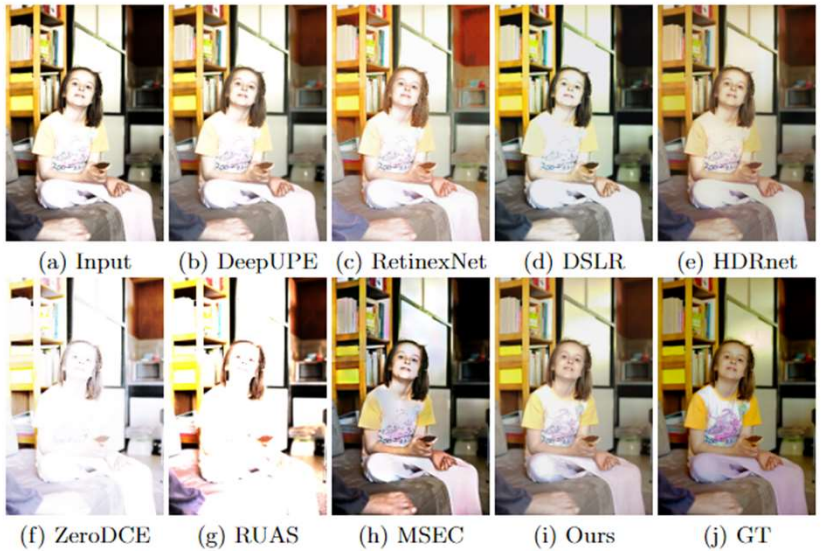


Fig. 9: Visual comparison of an over-exposed image from our dataset. Our result has the best visual quality.

Experiments--Quantitative Comparison

Method	PSNR↑	SSIM↑
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**.

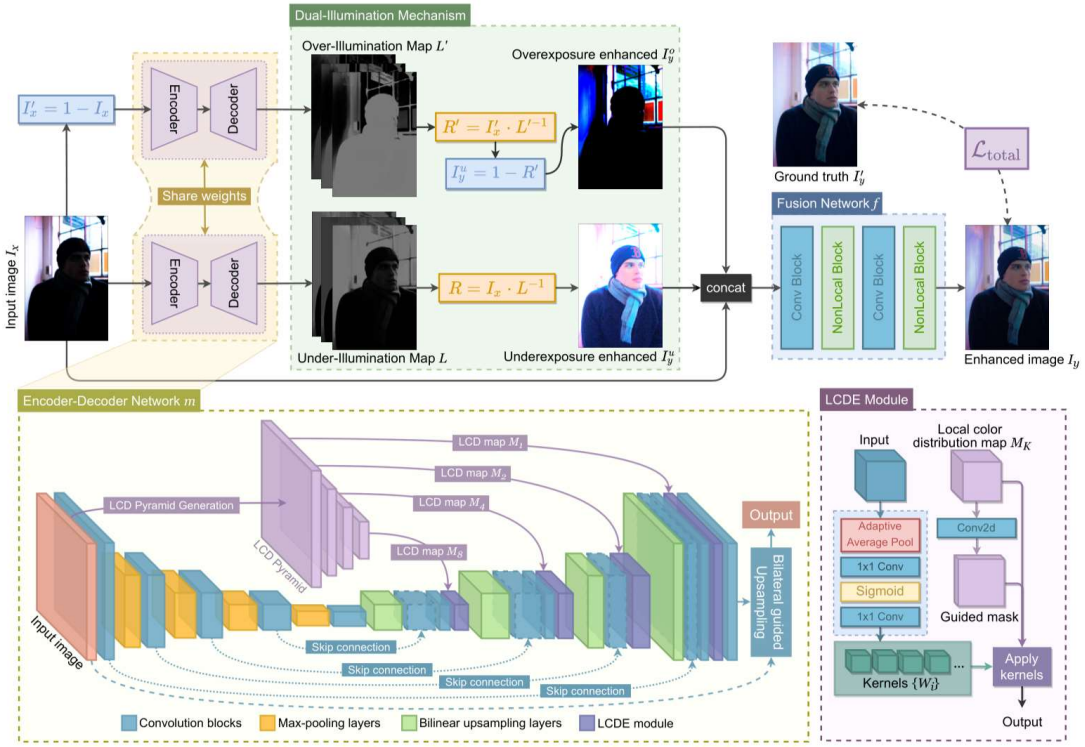
Method	PSNR↑	SSIM↑	Method	PSNR↑	SSIM↑
HE[26]	15.975	0.684	$DSLR_{Sony}$[18]	16.991	0.672
ClAHE[28]	16.327	0.642	$DSLR_{BlackBerry}$[18]	17.215	0.693
LIME[15]	17.335	0.686	$DSLR_{iPhone}$[18]	18.560	0.712
RetinexNet [10]	16.200	0.630	DSLR*[18]	20.856	0.758
RetinexNet*[10]	19.250	0.704	DeepUPE*[33]	20.970	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 retrained on our proposed training set. Best performances are marked in **bold**.

Experiments—Ablation study

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

Table 3: Ablation study.



Visualization

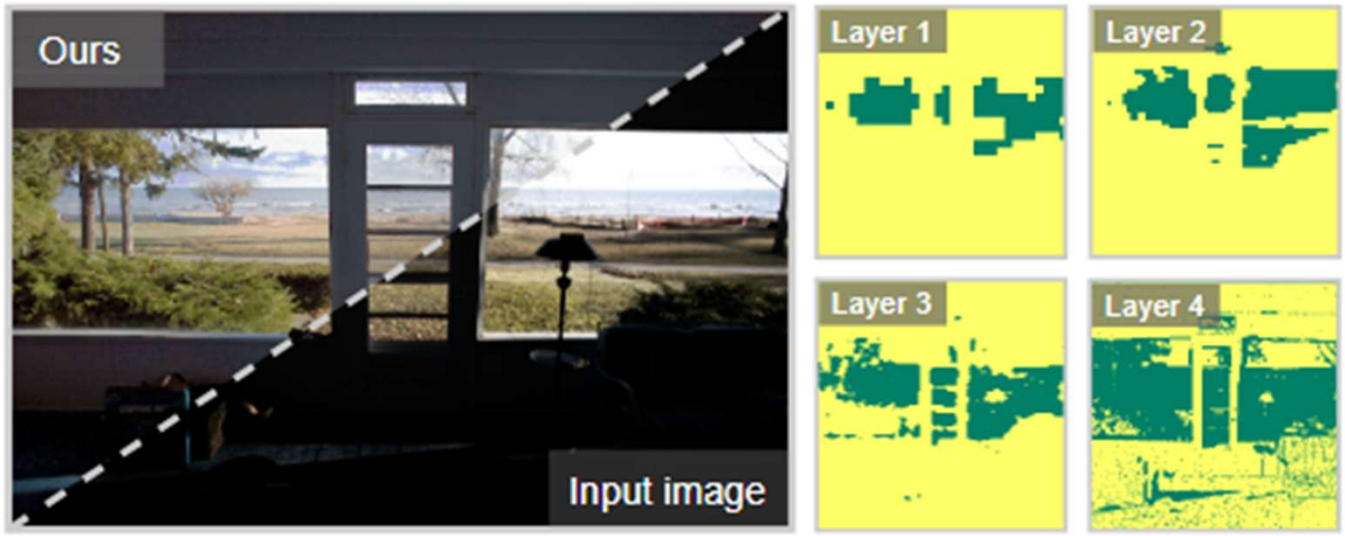


Fig. 13: The visualization of the learned guided mask in multi-scales intermediate LCDE module layers. With the guidance of the LCD pyramid and the constraint of the dual-illumination map, the model learns the guided mask adaptively.

Limitations

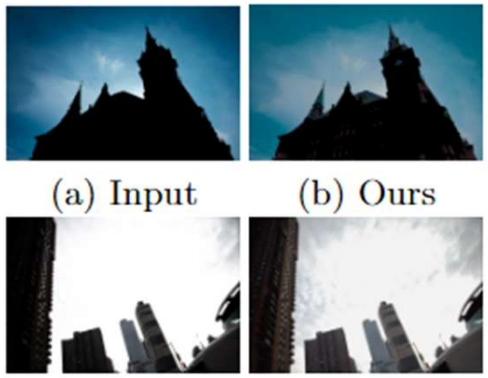


Fig. 14: Failure cases. Our method may fail to enhance images with a large region of under-exposed pixels (building in (b)) or over-exposed pixels (sky in (B)).

Thanks!