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, Hong Kong, People's Republic of China

Introduction

Given an input image (a) with both over-exposure (background windows) and under-exposure (foreground persons)

- Assume the scene illumination to be generally uniform, such that an improper exposure would result in either overor under-exposure.However, if the illumination of a scene is non-uniform, causing the input image to suffer from both over- and under-exposures
- On the other hand, while MSEC (b) enhances the foreground slightly, it produces some color distortions around the background windows.

(a) Input image (b) MSEC [1] (c) ZeroDCE [14] (d) RUAS [29] (e) Ours (f) Ground truth

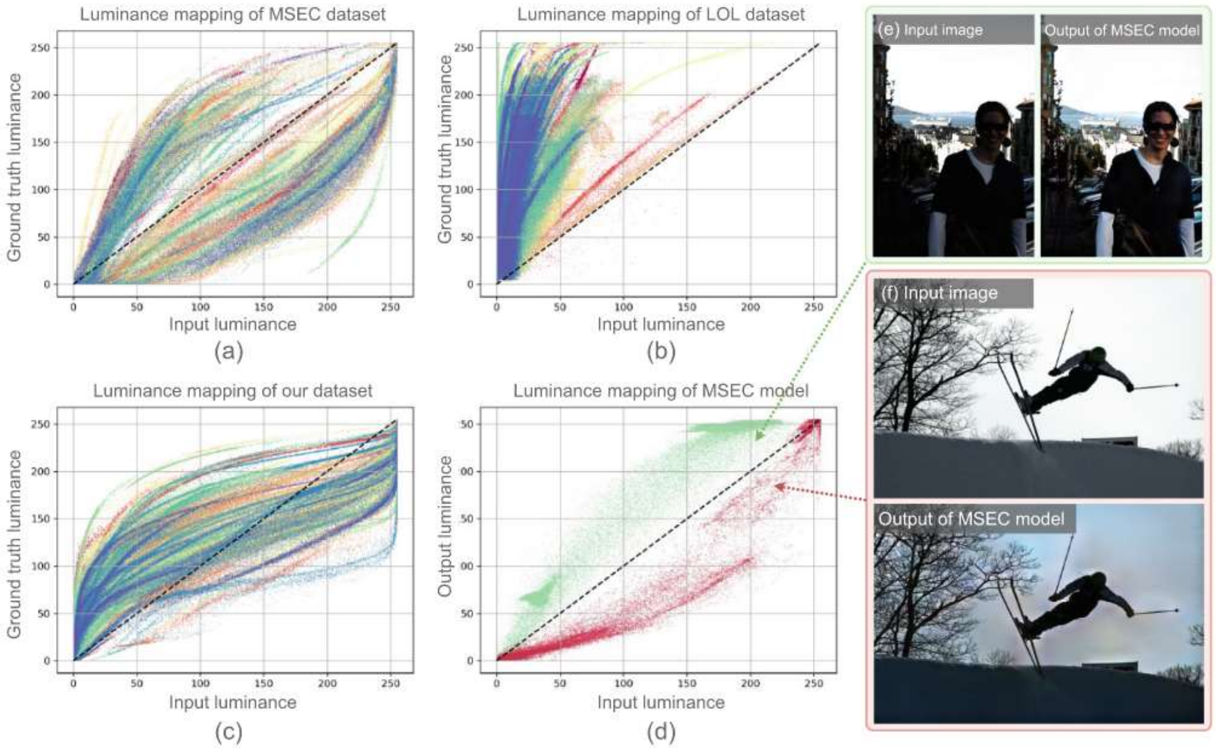
The key challenge is how to effectively separate these two types of regions and recover their local illuminations accordingly.

Introduction

Contribution:

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

Proposed DataSet



MSEC: Although it contains images with different levels of over- or under-exposures, each of the images is either over- or under-exposed.

LOL: It only allows the learning of under-exposure enhancement. This means that methods trained on the LOL dataset would typically brighten the input images.

——Need a new dataset containing both over- and under-exposures in individual images is desired.

(d) is the input-output mapping learned by the model trained on the MSEC dataset when given (e, f) as the input.

Proposed DataSet

We construct our dataset from the raw images in the MIT Adobe5k dataset, which contains 5000 raw and expert-retouched sRGB image pairs for learning the tone mapping process.

Method:

- We manually go through all the image pairs in Adobe5k, and remove those pairs whose expert-retouched images contain very dark or very bright regions.

- For each candidate raw image from (1), we follow the camera ISP pipeline to render an sRGB image with both over- and under-exposures, by adjusting the exposure level with a linear transformation function.

- For each rendered sRGB image from (2), we ask a volunteer (who is a photographer) to help assess its quality.

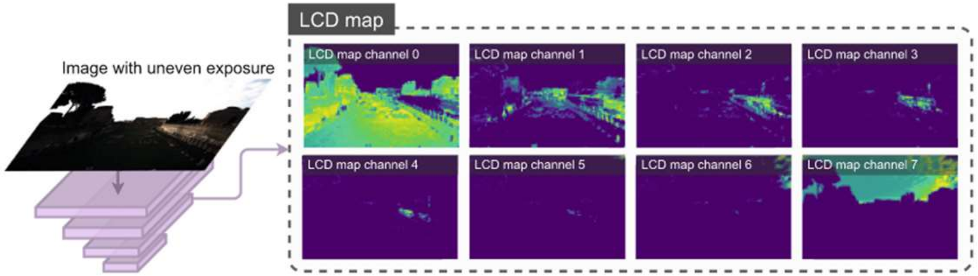
In total, we generate 1,733 pairs of images, which are split into 1,415 for training, 100 for validation, and 218 for testing.

Input images

Ground truth

Local Color Distribution (LCD) Pyramid

- Given an input image Ix of size h × w, whose pixel values range between [0, 1], we split Ix into N = [h/K] · [w/K] patches



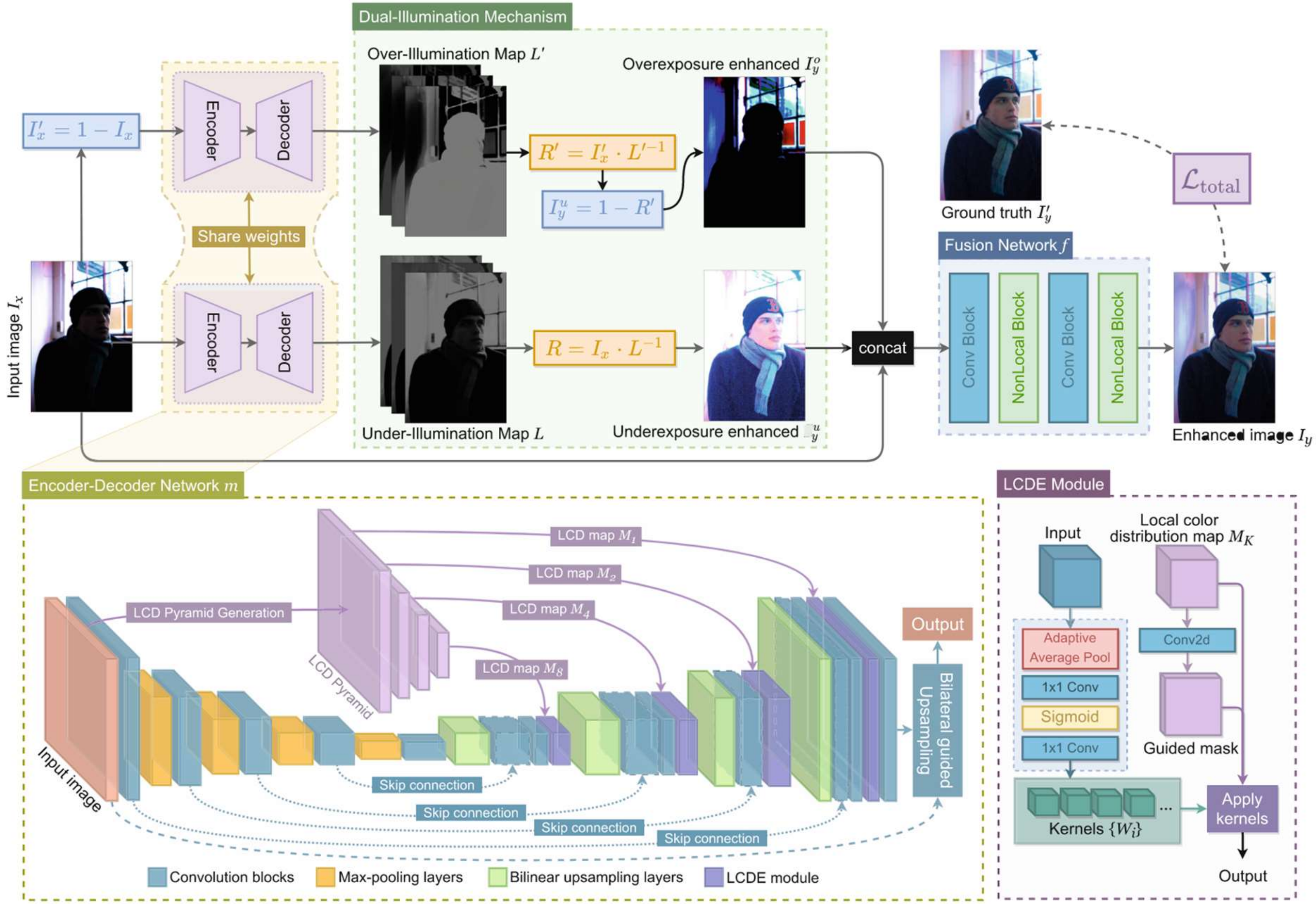
Method

- Define a LCD as the color histogram within a local patch of size K × K.
- Use the 4D LCD map Mk to represent the distribution of scale K.
- Build a h × w bilateral grid Γ by splatting the pixel histogram voting along the range dimension, then compute Mk

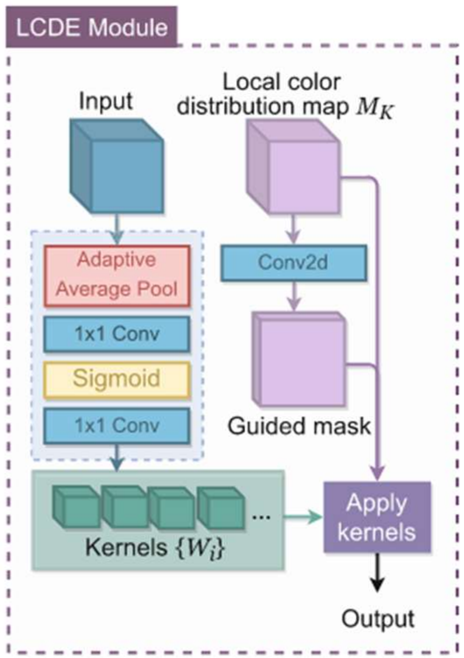
$$M_k(\left[\frac{i}{K}\right], \left[\frac{j}{K}\right], c, b) = \frac{1}{K^2} \sum_{p,q \in \Omega_K(i,j)} \Gamma(p, q, c, b)$$

K： {2^l, l ∈ N+}
b = [Ix(i, j, c) · B]， the index to the histogram bins

Proposed Network



Proposed Network

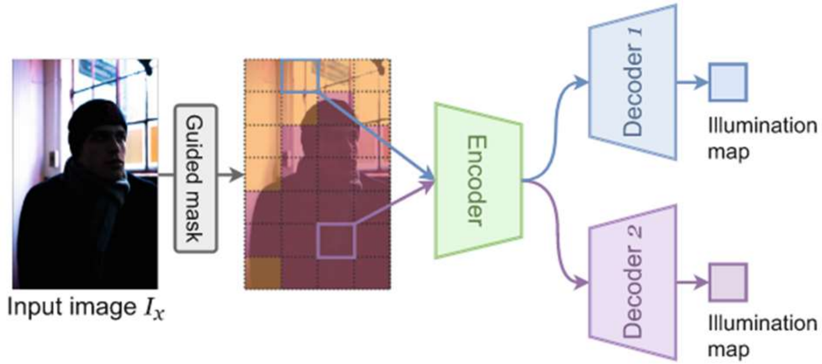


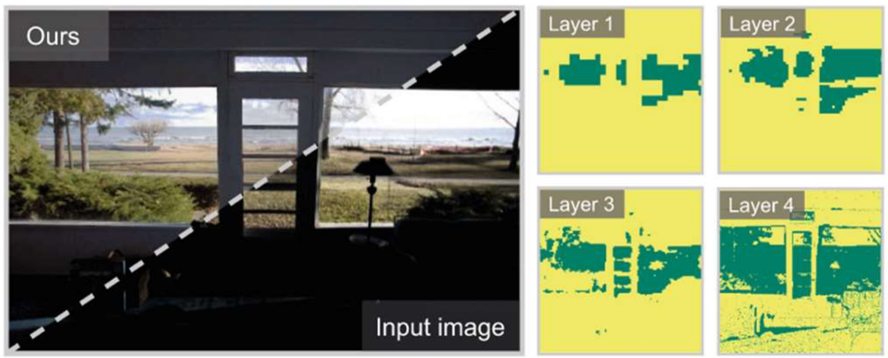
LCDE module

Design of the LCDE module is based on DRconv：
——Learn different kernels according to the local illuminations guided by the local color distributions.

Two branches：

- the convolutionkernels-generation branch
  produces the parameters of n kernels {W1, W2, ...Wn}
- the guided-mask-prediction branch
  inputs the LCD map MK, and uses it to guide the prediction of a multivalue mask for dividing the spatial feature maps into n regions





Proposed Network

Dual-Illumination Estimation

Exploit the Retinex theory in our model
Decompose the input image Ix into an illumination map L and a reflectance map R

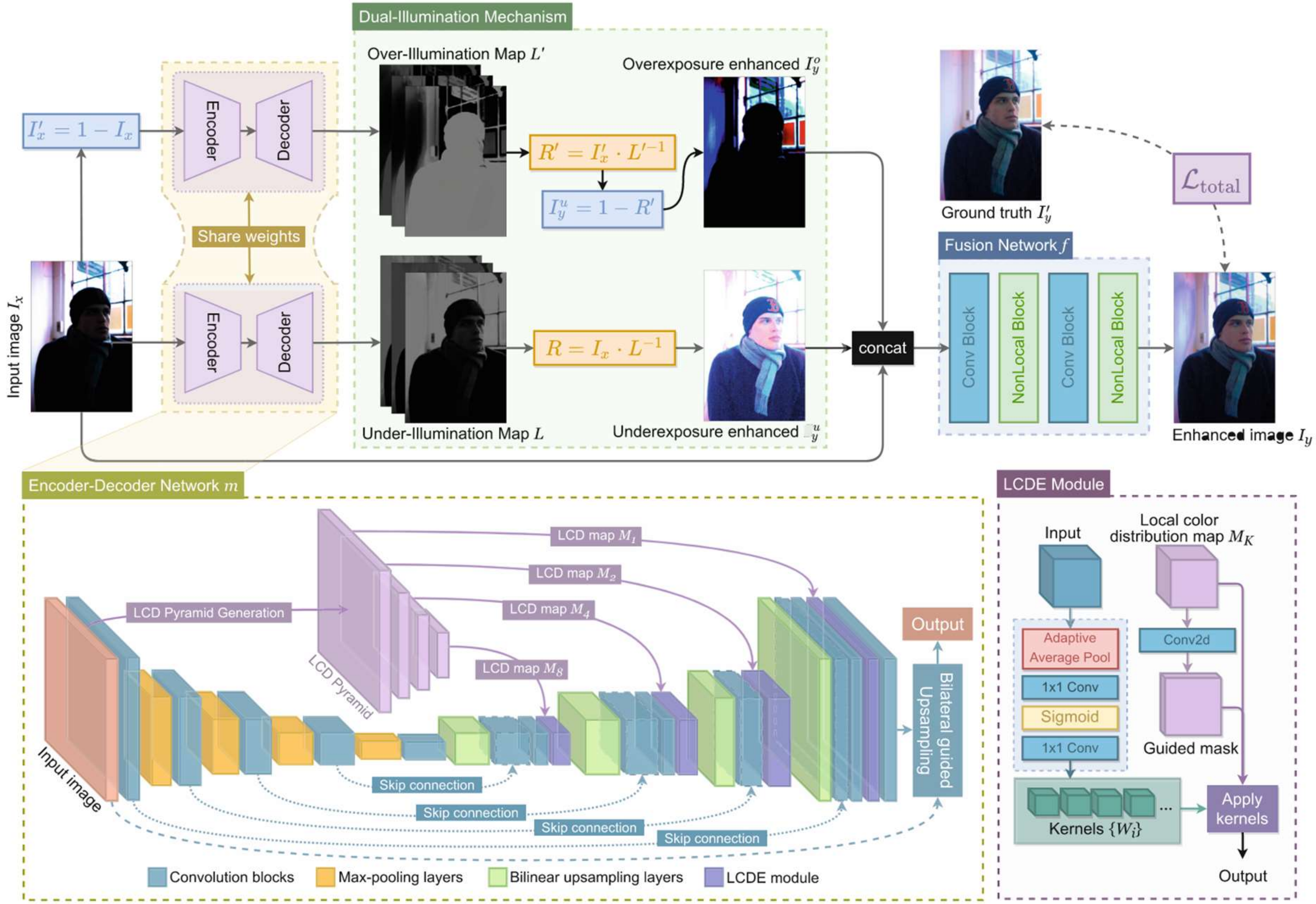
The main idea：
over-exposures in Ix could be regarded as underexposures in the reverse image of Ix

$$I_y^u = R = I_x \cdot L^{-1}$$

$$I_y^o = 1 - R' = 1 - I_x' \cdot L'^{-1}$$

L and L′ are estimated via the encoder-decoder network with the LCDE module.

Proposed Network



Loss Function

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

- Lmse： measure the intensity reconstruction errors
- Lcos： measures the color similarity of the reconstructed image and its ground truth in the sRGB color space
- Ltv1, Ltv2： preserve the local smoothness characteristics of image illuminations by minimizing their gradient variations.

Experiments

Quantitative Comparisons

Method	PSNR↑	SSIM↑
HE [25]	16.525	0.696
CLAHE [27]	15.383	0.599
DSLR [17] (Sony)	18.020	0.683
DSLR [17] (BlackBerry)	17.606	0.653
DSLR [17] (iPhone)	15.907	0.622
RetinexNet [10]	11.135	0.605
DeepUPE [32]	13.689	0.632
ZeroDCE [14]	12.058	0.544
MSEC [1]	20.205	0.769
Ours	**22.295**	**0.855**

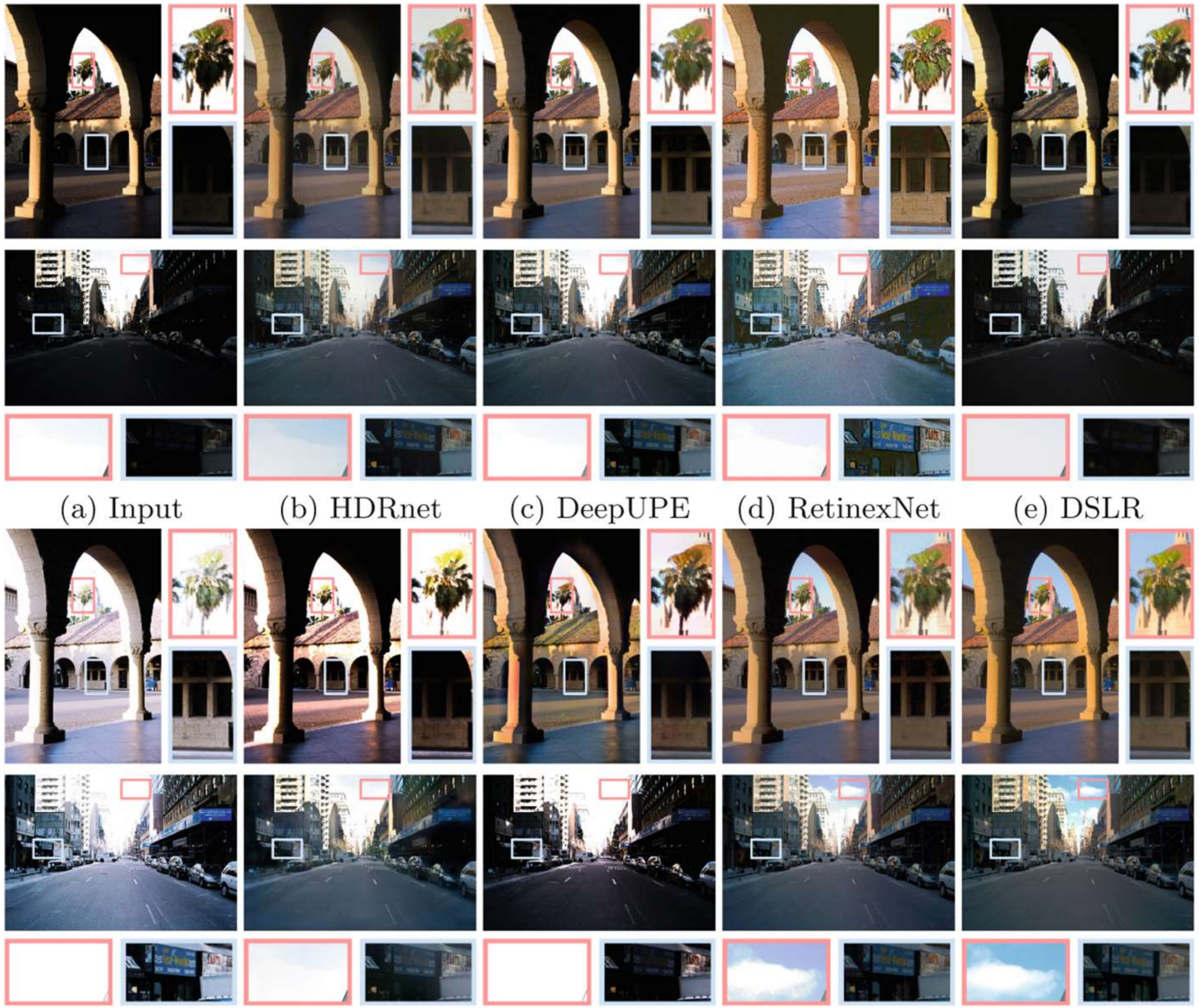
on the MSEC test set

Method	PSNR↑	SSIM↑	Method	PSNR↑	SSIM↑
HE [25]	15.975	0.684	$\text{DSLR}_{\text{Sony}}$ [17]	16.991	0.672
ClAHE [27]	16.327	0.642	$\text{DSLR}_{\text{BlackBerry}}$ [17]	17.215	0.693
LIME [15]	17.335	0.686	$\text{DSLR}_{\text{iPhone}}$ [17]	18.560	0.712
RetinexNet [10]	16.200	0.630	DSLR* [17]	20.856	0.758
RetinexNet* [10]	19.250	0.704	DeepUPE* [32]	20.970	0.818
MSEC* [1]	20.377	0.779	RUAS [28]	13.927	0.634
MSEC [1]	17.066	0.642	RUAS* [28]	13.757	0.606
ZeroDCE* [14]	12.587	0.653	HDRnet* [13]	21.834	0.818
Ours	**23.239**	**0.842**			

on the proposed test set
(* indicates that the model is retrained on our proposed training set)

Experiments

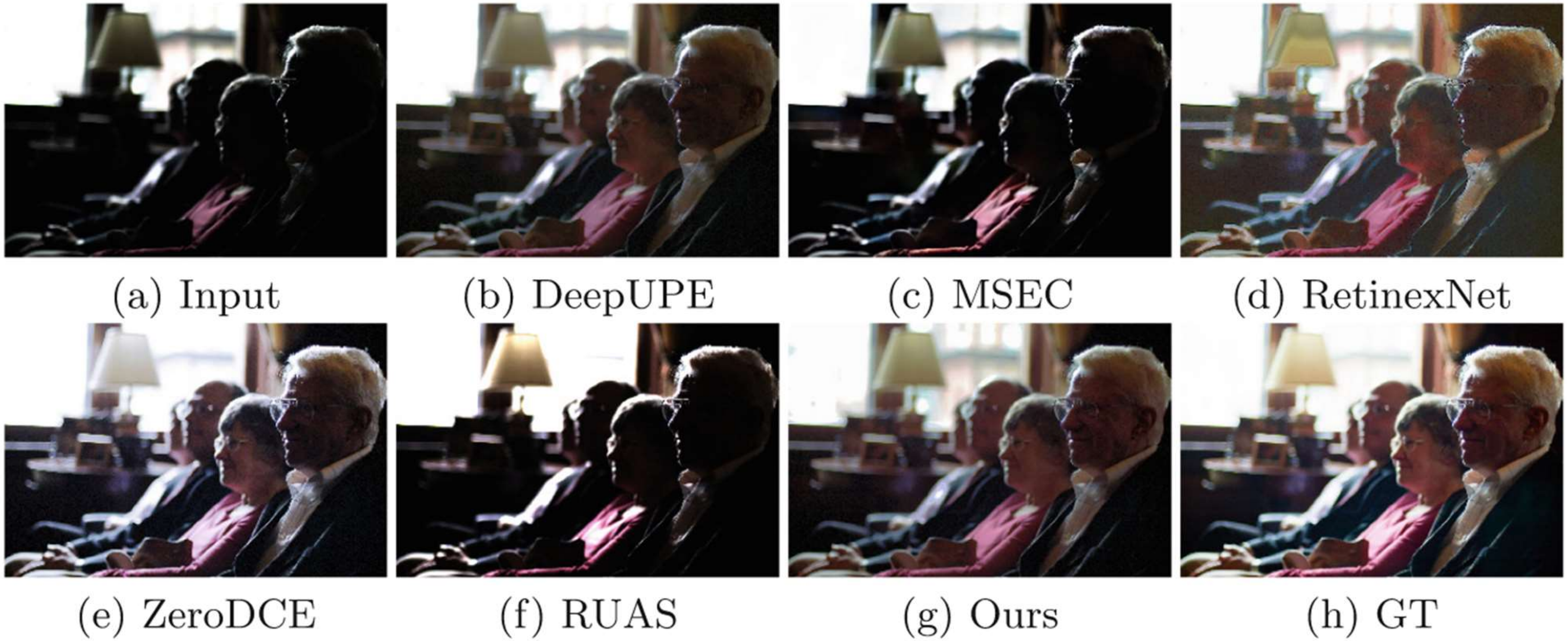
Qualitative Comparisons



(a) Input (b) HDRnet (c) DeepUPE (d) RetinexNet (e) DSLR

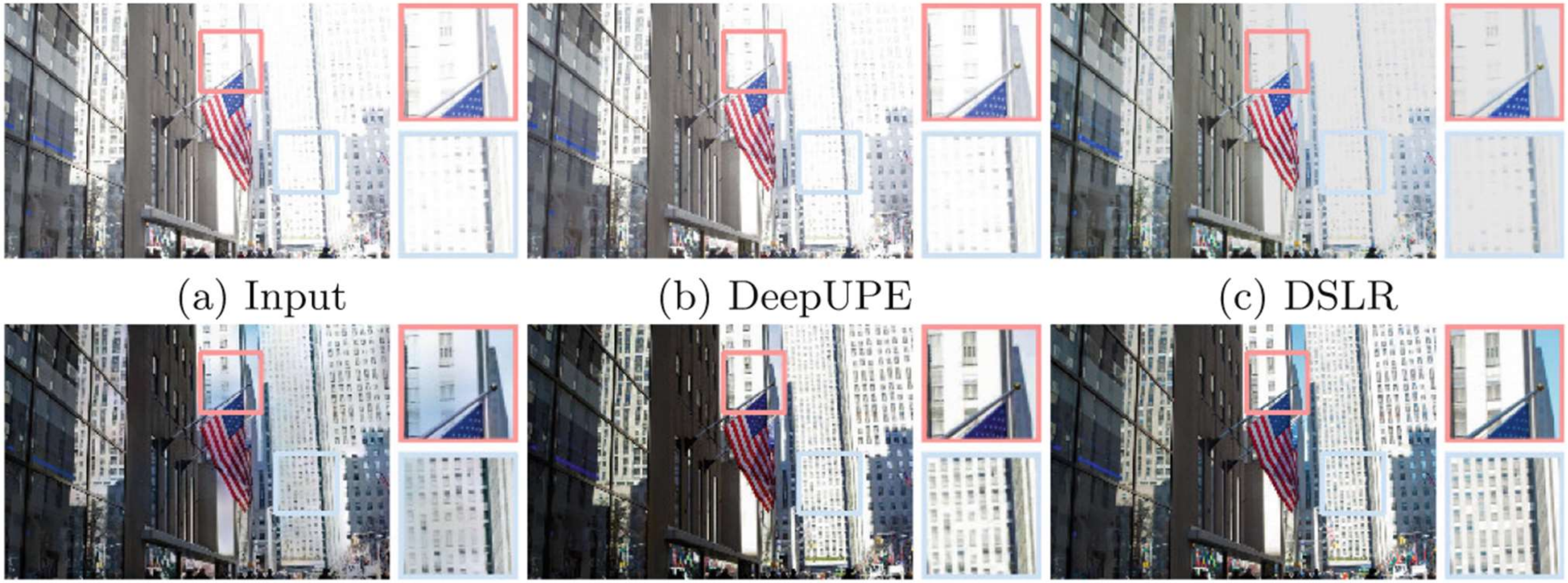
(f) ZeroDCE (g) RUAS (h) MSEC (i) Ours (j) GT

Experiments



Visual comparison of an under-exposed image from our dataset.

Experiments



(a) Input (b) DeepUPE (c) DSLR (d) MSEC (e) Ours (f) GT

Visual comparison of an over-exposed image from the MSEC dataset

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

1. Oursplain: A plain encoder-decoder
2. Ourssingle: Add LCDE modules to the decoder of (1)
3. OursDRconv: Add the DRconv blocks to the decoder of (1)

Thanks