Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement

CVPR 2020 - Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, Runmin Chong
BIIT Lab, Tianjin University, City University of Hong Kong, Nanyang Technological University, Beijing Jiaotong University

Ivan Surya H - 0860812
M. Farhan Tandia - 0860814
Martin Dominikus - 0856643

What?

Zero-DCE is a novel method to do light enhancement in an image. We could obtain do light enhancement in image while keep maintain the detail and preserves quality of the image

Table of Contents

01

Introduction

Introduction and related work

To help enhance low-light photos without losing its quality or any information. Enhancement may also recover the object detection or recognition in the low-light area

What unique?

"It does not require any paired or unpaired data in the training process as in existing CNN-Based because It using non-reference loss functions.

Zero-Reference

Contributions

Independent Network

Independent of paired and unpaired training data, thus avoiding the risk of overfitting

Image-Specific Curve

This paper design an image-specific curve that is able to approximate pixel-wise and higher-order curves by iteratively applying itself

Non-Reference Loss Function

Show the potential of training a deep image enhancement model in the absence of reference images through non-reference loss function that indirectly evaluate enhancement quality

Zero-DCE supersedes State-of-the-Arts. It is capable of processing images in real-time (about 500 FPS for images of size 640x480x3 on GPU)

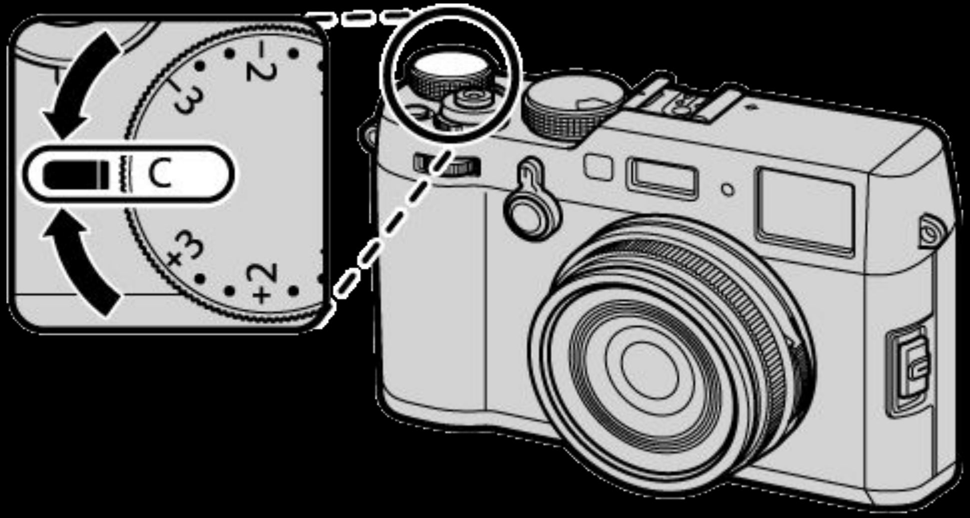
Related Works

Conventional Method

Histogram Enhancement methods perform light enhancement through expanding the dynamic range of an image or using Retinex Theory

Data-Driven Method

Data-Driven method is method that using neural network, largely categorized into CNN-Based and GAN-Based



Conventional Method

	Method	Weakness	Zero-DCE
Histogram Enhancement	Expanding Dynamic Range of image using histogram adjustment	Potentially inaccurate physical model	Using Image-Specific curve mapping that has better result
Retinex Theory	Decomposes an image into reflectance and illumination	Potentially produce unrealistic enhancement	Able to enhance image without creating unrealistic enhancement
S-Shaped Curve Method	Estimate S-Shaped curve of given image using global optimization algorithm.	Less robustness and narrower dynamic range adjustment	Better robustness and wider image dynamic range adjustment

Data-Driven Method

	Method	Weakness	Zero-DCE
Underexposed Photo Enhancement Network (CNN)	Estimating the Illumination Map using CNN	Based on paired data, produce artifacts and color-cast.	Does not need paired data, more practical, more natural
Unsupervised GAN-Method	Same as CNN methods but does not need paired data (using unpaired normal light data)	Require careful selection of unpaired training data	Using zero reference hence eliminate the requirement of paired/unpaired data, highly efficient, and cost effective.

02

Methodology

Method of Zero-DCE in detail

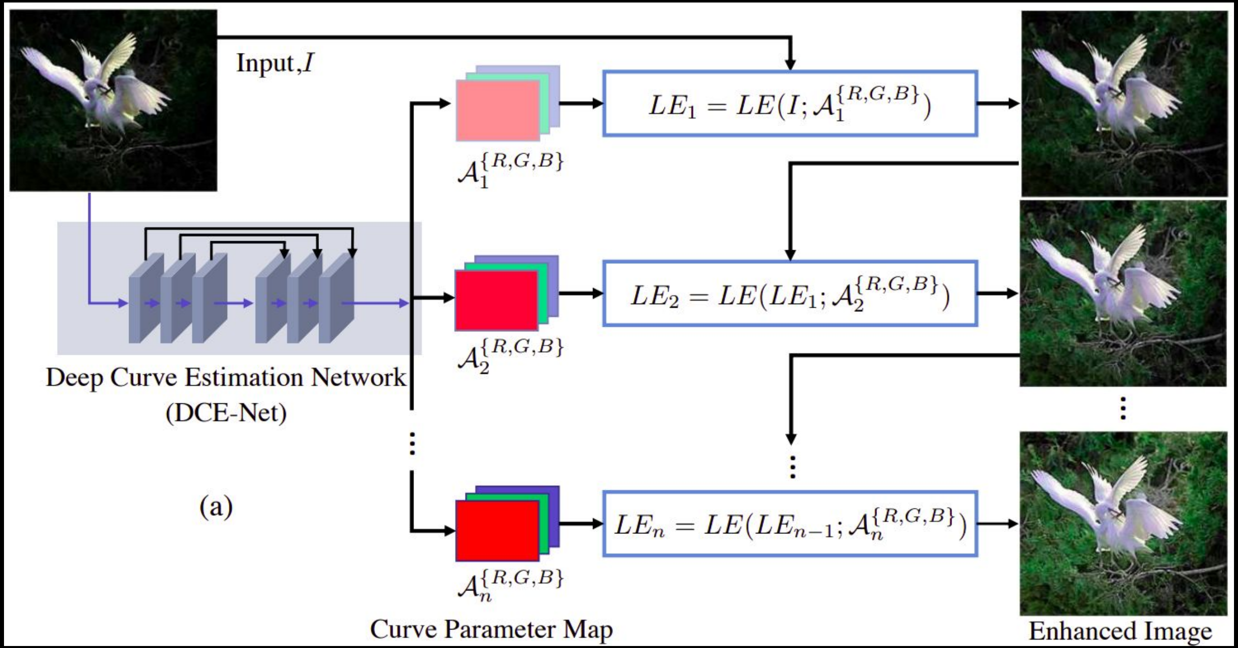
Algorithm Components

1. Light-Enhancement Curve
2. DCE-Net
3. Non-Reference Loss Functions
   a. Spatial Consistency Loss
   b. Exposure Control Loss
   c. Color Constancy Loss
   d. Illumination Smoothness Loss

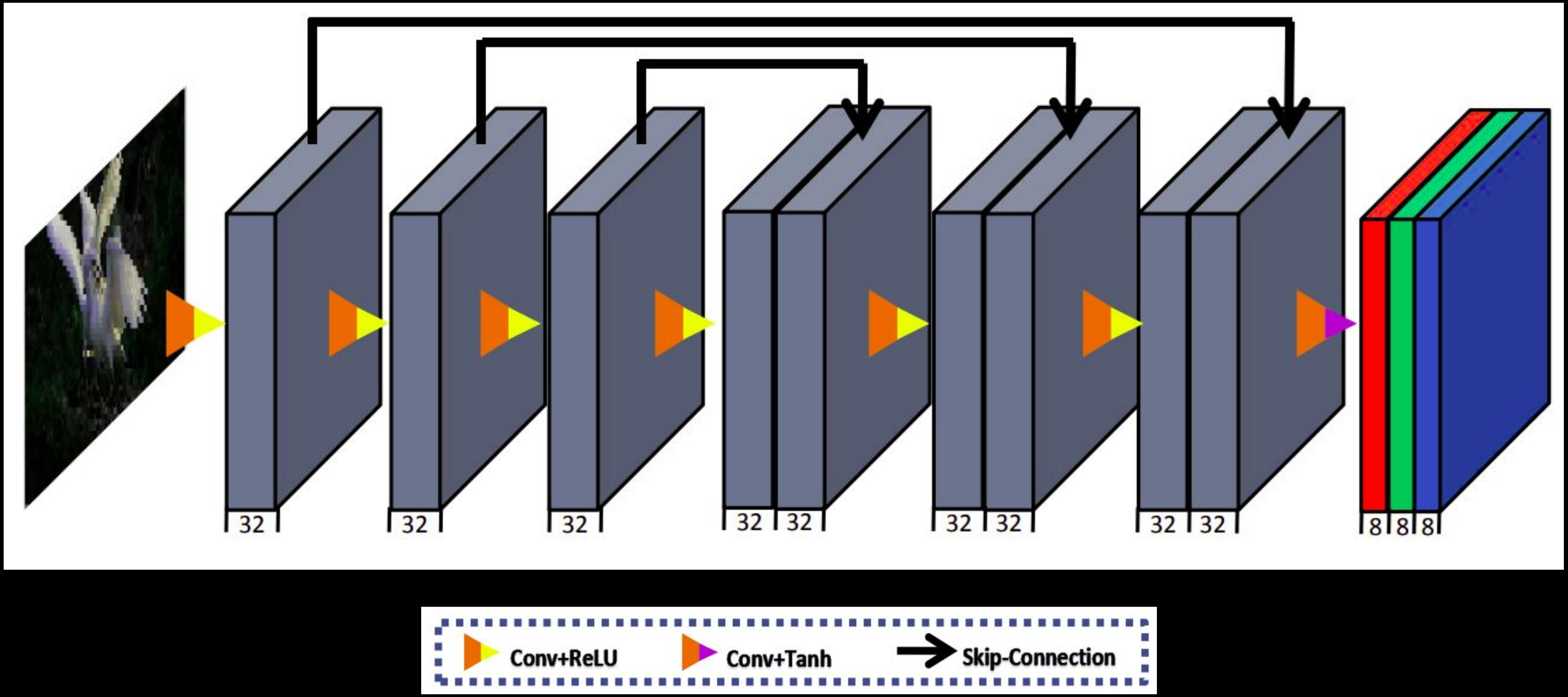
LE-Curve

1. Has the same size as input image
2. Each element is trainable parameters
3. Pixel-wise operation → $LE(I(\mathbf{x}); \alpha) = I(\mathbf{x}) + \alpha I(\mathbf{x})(1\text{-}I(\mathbf{x}))$
   a. Each pixel $I(\mathbf{x})$ is normalized to [0,1]
   b. $\alpha \in [-1,1]$; trainable parameters
4. Each channel (R, G, B) has its own curve
5. Work in iteration

Zero-DCE Framework



DCE-Net



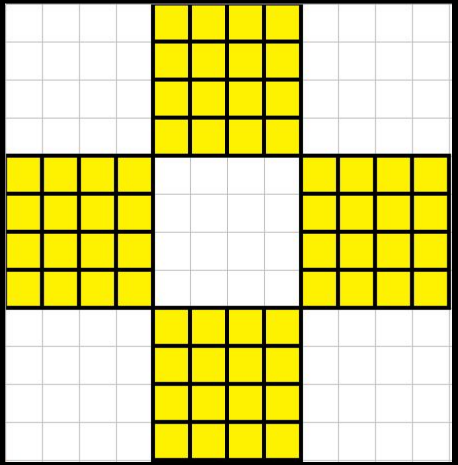
32
32
32
32 32
32 32
32 32
8 8 8
Conv+ReLU
Conv+Tanh
Skip-Connection

Loss Functions: Spatial Constancy Loss

$$L_{spa} = \frac{1}{K}\sum_{i=1}^{K}\sum_{j\in\Omega(i)}(|(Y_i - Y_j)| - |(I_i - I_j)|)^2$$

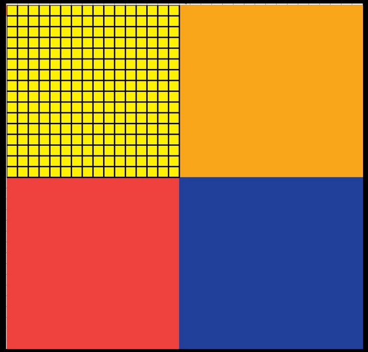
where,

1. K : # local regions
2. Ω(i): one of four neighboring regions
3. Y : average of intensity value of local region in *enhanced image*
4. I : Same as Y but now in *input image*



Loss Functions: Exposure Control Loss

$$L_{exp} = \frac{1}{M} \sum_{k=1}^{M} |Y_k - E|,$$



where,

1. M : 16x16 non overlapping region
2. Y : average intensity value of local region in the enhanced image
3. E : scalar, gray level in the RGB color space

Loss Functions: Color Constancy Loss

$$L_{col} = \sum_{\forall(p,q)\in\varepsilon}(J^p - J^q)^2, \varepsilon = \{(R,G),(R,B),(G,B)\}$$

where,

1. J^P : average intensity value of *p* channel in enhanced image
2. J^q : same thing but now it's value of *q* channel
3. (p,q) : pair of channel, defined by set ε

Loss Functions: Illumination Smoothness Loss

$$L_{tv_{\mathcal{A}}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{c\in\xi}(|\nabla_x\mathcal{A}_n^c| + \nabla_y\mathcal{A}_n^c|)^2, \xi = \{R, G, B\}$$

where,

1. N : # iteration
2. ∇x : image gradient, horizontal direction
3. ∇y : image gradient, vertical direction
4. A : LE-curve

03

Experiment and Results

Implementation and Benchmark Evaluation

Training Data

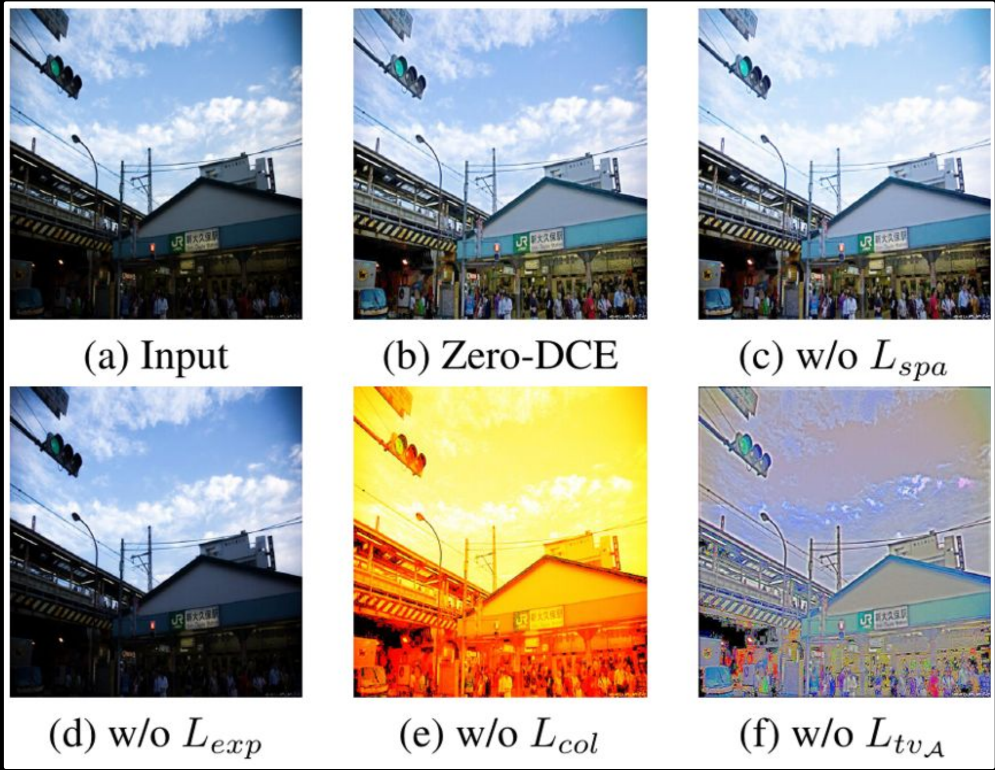
SICE dataset (Single Image Contrast Enhancer)

Part 1 of SICE
Total 3,022 (360 images sequences) with different exposures levels (512x512)

Random Split
80% for Training (2422 images)
20% for validation

Ablation Study

Contribution of Each Loss



By Removing

- **Spatial consistency loss (*Lspa*)**
  - Relatively low contrast
- **Exposure control loss (*Lexp*)**
  - Fails to recover low-light region
- **Color constancy loss (*Lcol*)**
  - Ignores the relations of 3 channels when curve mapping is applied
- **Illumination smoothness loss (*LtvA*)**
  - Leading to obvious artifacts

Ablation Study

Impact of Training Data

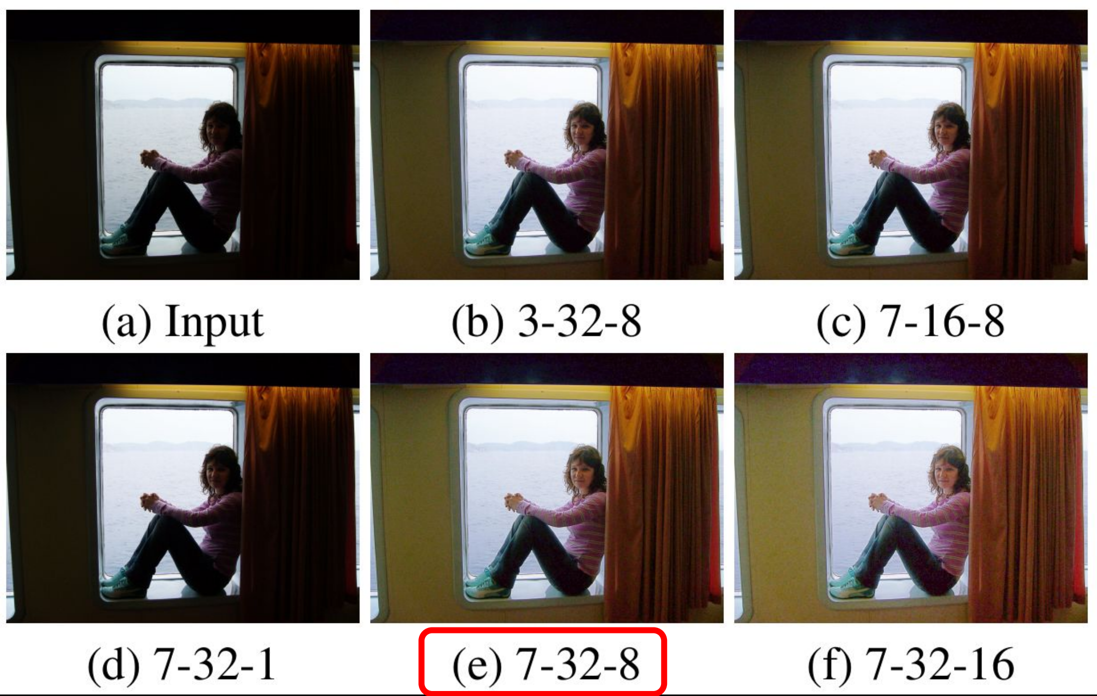
(a) Input | (b) Zero-DCE | (c) Zero-DCE$_{Low}$ | (d) Zero-DCE$_{LargeL}$ | (e) Zero-DCE$_{LargeLH}$

- (b) Zero-DCE: Part 1 SICE Dataset (2422 multi exposure images)
- (c) Zero-DCE$_{Low}$: Part 1 SICE Dataset (900 low light images)
- (d) Zero-DCE$_{LargeL}$: DARK FACE dataset (9000 low light images)
- (e) Zero-DCE$_{LargeLH}$: Part 1 and Part 2 SICE Dataset (4800 multi exposure images)

BEST RESULTS

Ablation Study

Effect of Parameter Settings



(a) Input (b) 3-32-8 (c) 7-16-8
(d) 7-32-1 (e) 7-32-8 (f) 7-32-16

$l - f - n$

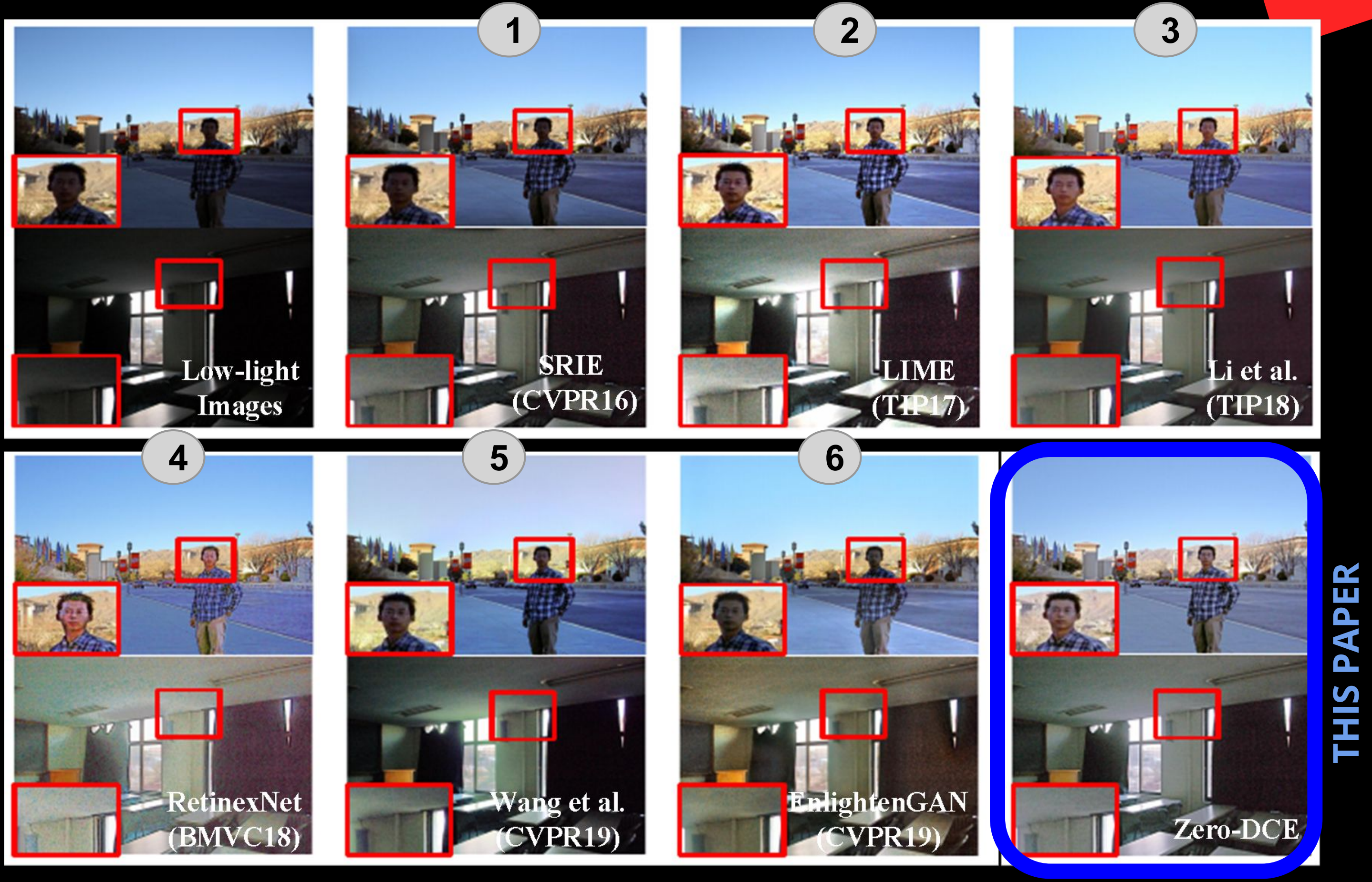
(conv layer - filter - iteration)

Why they choose (e) ?

(b), (c), (d) suffer from low light

(f) good result but prone to overfit /not generalize well

Visual and Perceptual Comparisons



User Studies & Quantitative Comparisons

1

Method	NPE	LIME	MEF	DICM	VV	Average
SRIE [8]	3.65/2.79	3.50/2.76	3.22/2.61	3.42/3.17	2.80/3.37	3.32/2.94
LIME [9]	3.78/3.05	3.95/3.00	3.71/2.78	3.31/3.35	3.21/3.03	3.59/3.04
Li *et al.* [19]	3.80/3.09	3.78/3.02	2.93/3.61	3.47/3.43	2.87/3.37	3.37/3.72
RetinexNet [32]	3.30/3.18	2.32/3.08	2.80/2.86	2.88/3.24	1.96/2.95	2.58/3.06
Wang *et al.* [28]	3.83/2.83	3.82/2.90	3.13/2.72	3.44/3.20	2.95/3.42	3.43/3.01
EnlightenGAN [12]	3.90/2.96	3.84/2.83	3.75/2.45	3.50/3.13	3.17/4.71	3.63/3.22
Zero-DCE	3.81/2.84	3.80/2.76	4.13/2.43	3.52/3.04	3.24/3.33	3.70/2.88

User study (US)↑/
Perceptual index (PI)↓
Score range (1-5)

15 people independently score
the visual quality of enhanced image

2

Method	PSNR↑	SSIM↑	MAE↓
SRIE [8]	14.41	0.54	127.08
LIME [9]	16.17	0.57	108.12
Li *et al.* [19]	15.19	0.54	114.21
RetinexNet [32]	15.99	0.53	104.81
Wang *et al.* [28]	13.52	0.49	142.01
EnlightenGAN [12]	16.21	0.59	102.78
Zero-DCE	16.57	0.59	98.78

1) The result contain **over/under enhanced regions and exposed artifacts**
2) The results **color deviation**
3) The results have **unnatural texture and obvious noise**

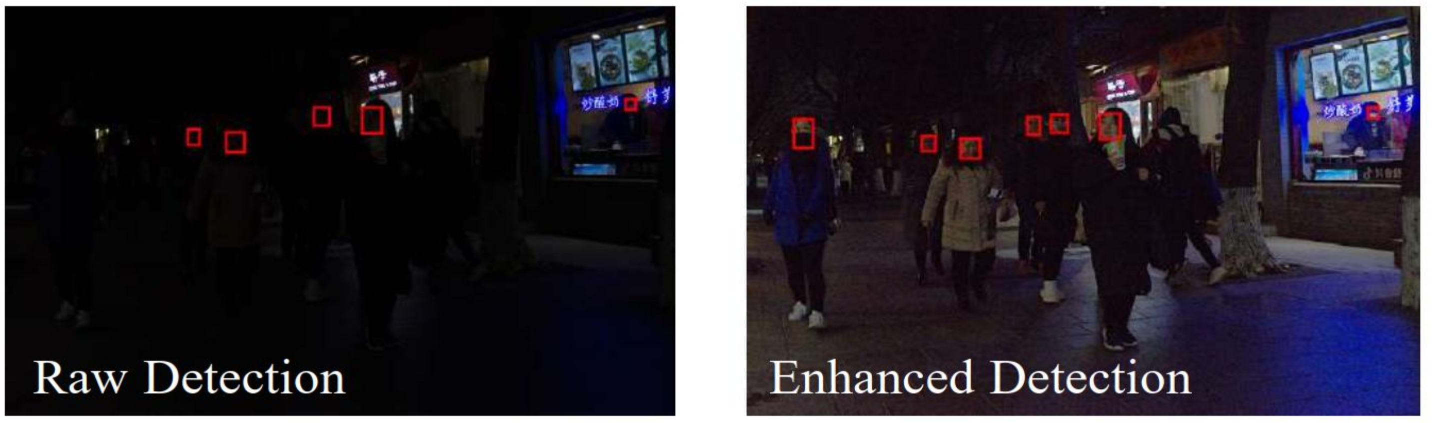
PNSR, dB : Peak Signal to Noise Ratio ↑
SSIM : Structural Similarity ↑
MAE : Mean Absolute Error ↓

Runtime Performance & Face Detection

- Zero Reference Model Test in DARK FACE dataset (10,000 images)
- Then using Dual Shot Face Detection [18] to detect the faces.

Method	RT	Platform
SRIE [8]	12.1865	MATLAB (CPU)
LIME [9]	0.4914	MATLAB (CPU)
Li *et al.* [19]	90.7859	MATLAB (CPU)
RetinexNet [32]	0.1200	TensorFlow (GPU)
Wang *et al.* [28]	0.0210	TensorFlow (GPU)
EnlightenGAN [12]	0.0078	PyTorch (GPU)
Zero-DCE	0.0025	PyTorch (GPU)

= 2.5 ms



[18] Jian Li, Yabiao Wang, Changan Wang, Ying Tai, Jianjun Qian, Jian Yang, Chengjie Wang, Jilin Li, and Feiyuen Huang. Dsfd: Dual shot face detector. In CVPR, 2019.

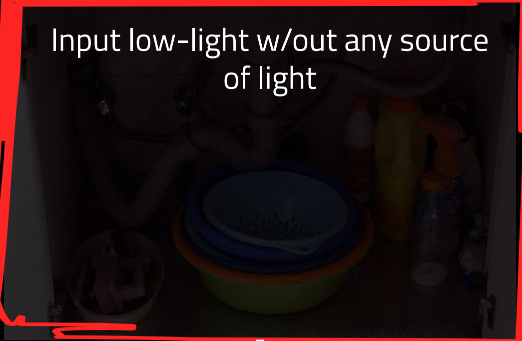
Our Testing Result

https://github.com/Li-Chongyi/Zero-DCE

before

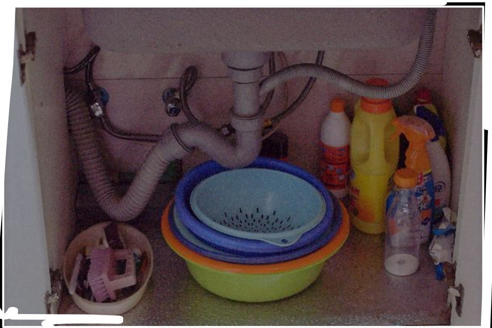
1

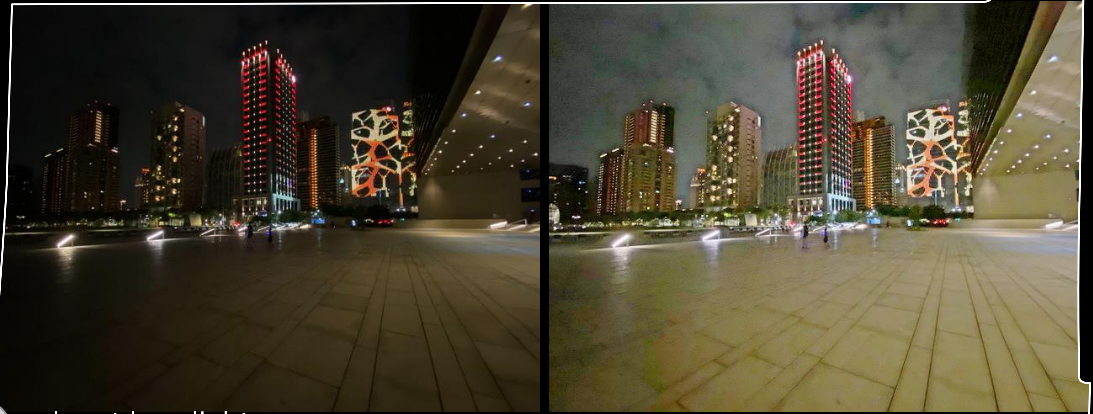
2



3

after





4

Input low-light
With noise

5

Input low-light
without noise

04

Final Demo Goals

Improvement and Implementation

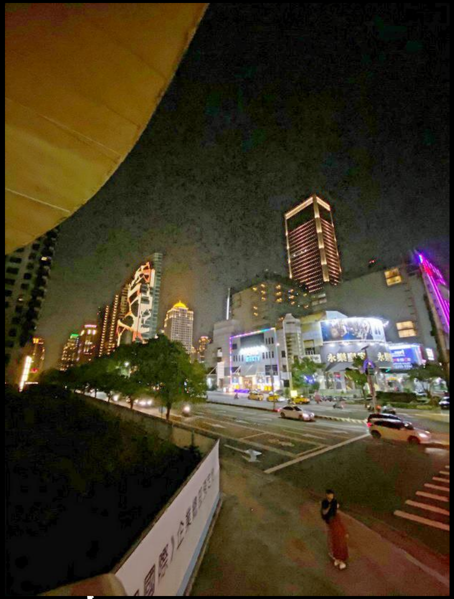
Improvement

Color Correction

fixes color issues and makes footage appear as naturalistic as possible (post-processing)

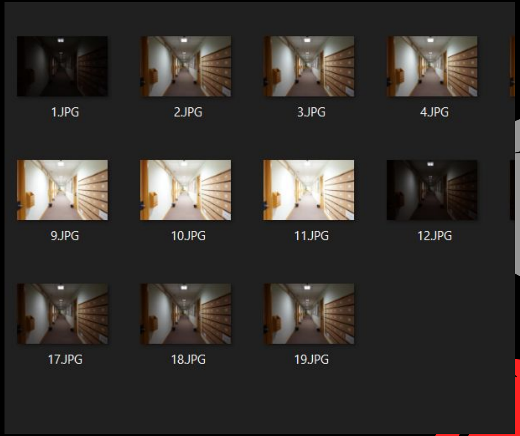
Image Denoising

Removing noise from an image (pre-processing)



Training Data

Use a larger amount of training data



Implementation

Real-time video demo

Capture real-time Image using DSLR Camera

Capture from DSLR using gPhoto2 and Node-RED (using the lowest ISO)

Image Denoising and DCE-Zero Network

Denoising every frames from DSLR Camera then send to DCE-Zero Network to improve DCE-Zero result

Post Processing (Color Correction)

Do the color correction from Image output of DCE-Zero Network to enhance the color from DCE-Zero result.

Result

Real time light enhancement image from DSLR Camera. This is really useful for entry-level camera to perform well in low light without high-spec sensor capability.

Conclusion

1

The proposed deep network for low-light image enhancement can be trained end to end with zero reference images (Zero-DCE)

2

Zero-DCE method can generate a superior result in terms of image quality among previous state-of-the-art.

3

Zero-DCE method achieved the fastest runtime performance so that can be used in real-time application such as face detection and etc.

Thank You!