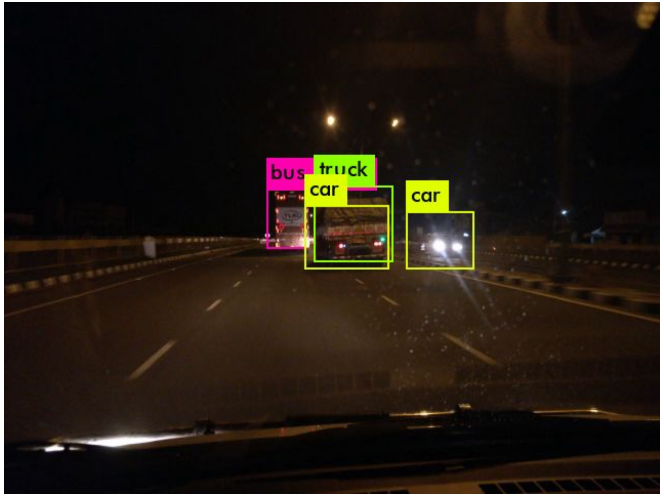
The University of Texas at Austin
Electrical and Computer Engineering
Cockrell School of Engineering

Learning to Enhance Low-light Images without Paired Supervision

Yifan Jiang

The University of Texas at Austin

Motivation



Autonomous Driving

Night-time Photography

Background

- Using deep neural network to replace traditional ISO processing in digital camera

Background

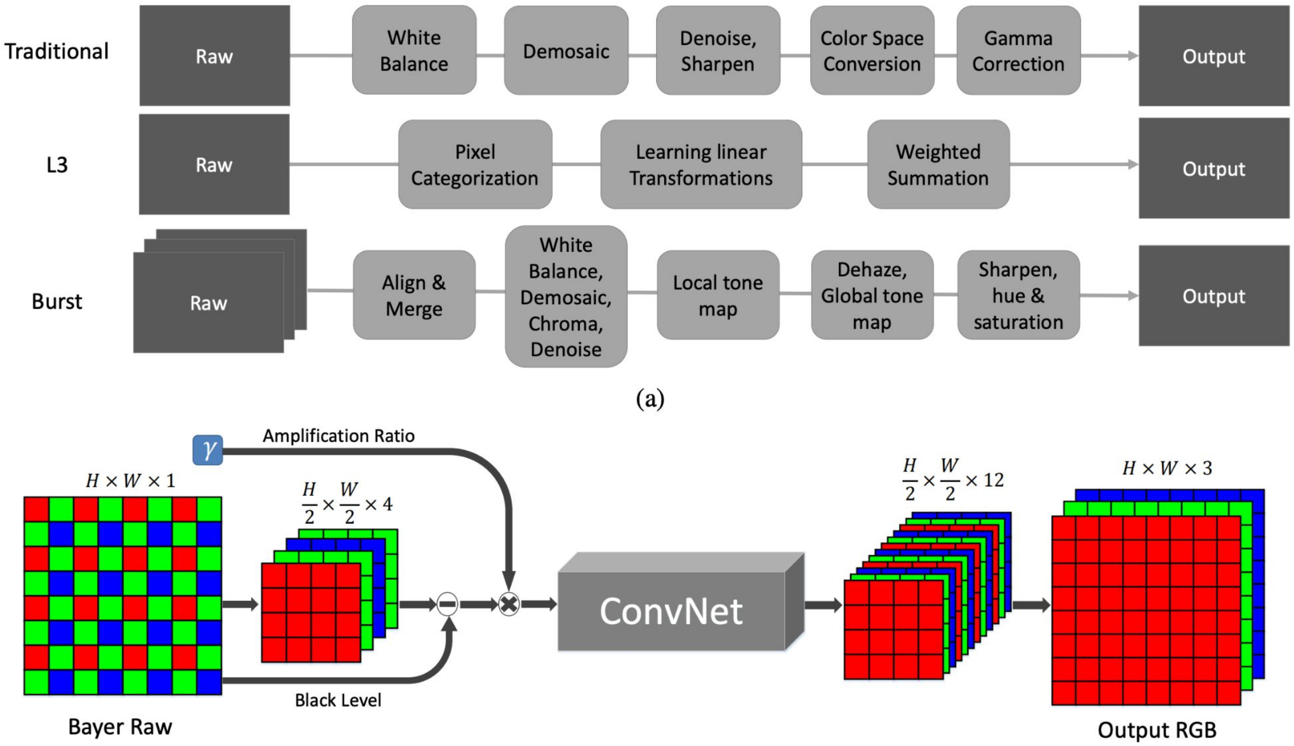
- Using deep neural network to replace traditional ISO processing in digital camera

- Leveraging motion estimation approach to process burst images

Background

- Using deep neural network to replace traditional ISO processing in digital camera
- Leveraging motion estimation approach to process burst images
- Adopting learning-based method for image post-processing

Neural ISO Processing



Learning to See in the Dark [Chen et. al.]

Neural ISO Processing

(a) Camera output with ISO 8,000 (b) Camera output with ISO 409,600 (c) Our result from the raw data of (a)

Figure 1. Extreme low-light imaging with a convolutional network. Dark indoor environment. The illuminance at the camera is < 0.1 lux. The Sony α7S II sensor is exposed for 1/30 second. (a) Image produced by the camera with ISO 8,000. (b) Image produced by the camera with ISO 409,600. The image suffers from noise and color bias. (c) Image produced by our convolutional network applied to the raw sensor data from (a).

Learning to See in the Dark [Chen et. al.]

Burst Image Processing

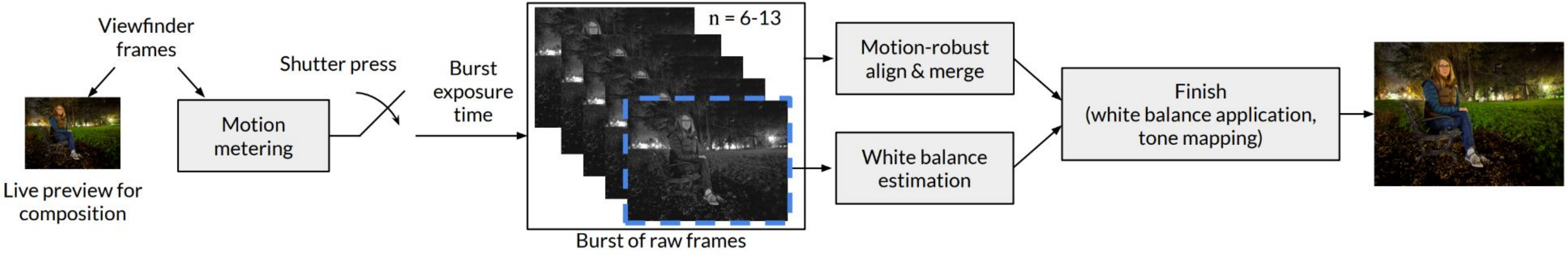


Fig. 2. An overview of our processing pipeline, showing how we extend (Hasinoff et al. 2016). Viewfinder frames are used for live preview for composition and for motion metering, which determines a per-frame exposure time that provides a good noise vs. motion blur tradeoff in the final result (Section 2). Based on this exposure time, we capture and merge a burst of 6–13 frames (Section 3). The reference frame is down-sampled for the computation of the white-balance gains (Section 4). White balance and tone mapping (Section 5) are applied at the "Finish" stage, as well as demosaicing, spatial and chroma denoising and sharpening. Total capture time is between 1 and 6 seconds after the shutter press and processing time is under 2 seconds.

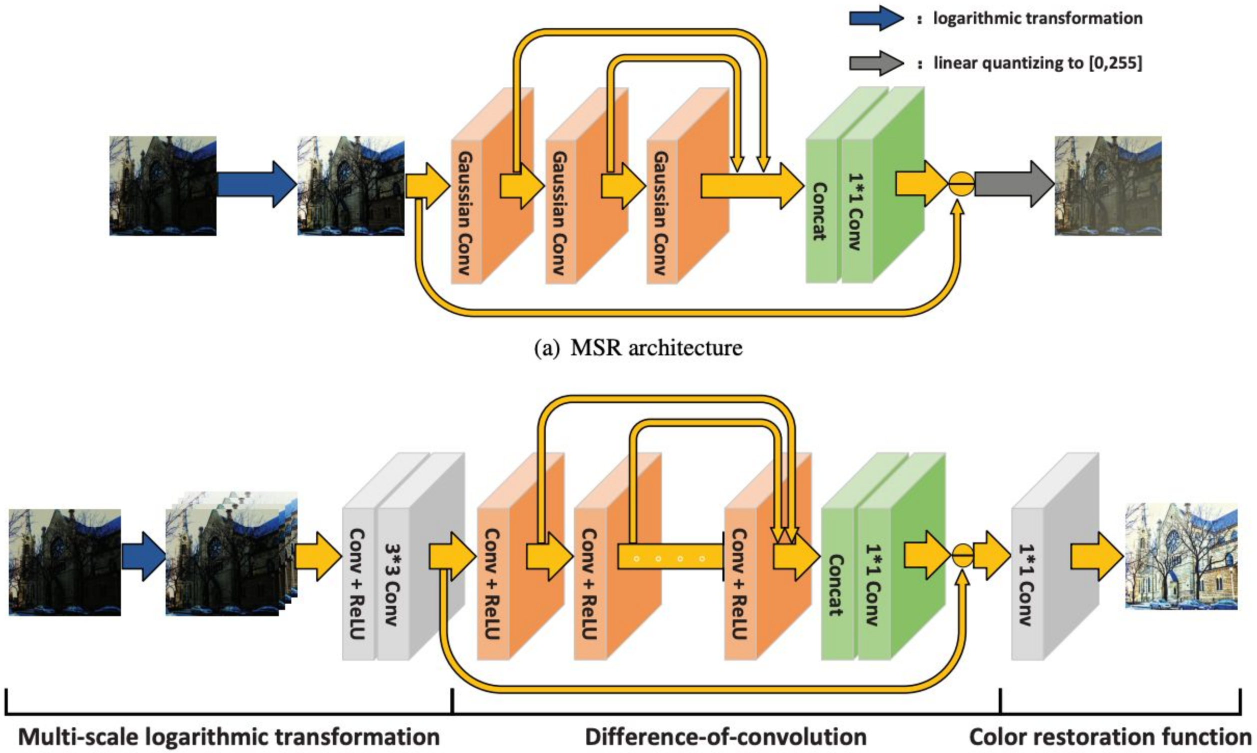
Handheld Mobile Photography in Very Low Light [Liba et. al.]

Burst Image Processing

(a) Previously described result (b) Previously described result, gained (c) Our result

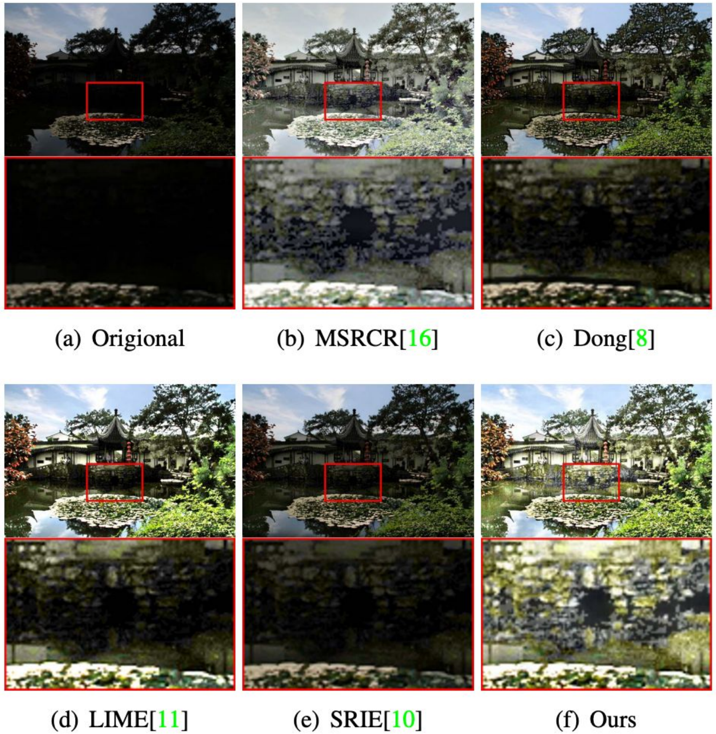
Handheld Mobile Photography in Very Low Light [Liba et. al.]

Learning-based Image Processing



MSR-net:Low-light Image Enhancement Using Deep Convolutional Network [Shen et. al.]

Learning-based Image Processing



MSR-net:Low-light Image Enhancement Using Deep Convolutional Network [Shen et. al.]

Issues of Learning-based Approaches

- Although globally adjusting the brightness of a natural image can create synthetic low-/normal-light data pairs, it can not simulate real-world scenarios where brightness is non-linear and spatially-varying

Issues of Learning-based Approaches

- Although globally adjusting the brightness of a natural image can create synthetic low-/normal-light data pairs, it can not simulate real-world scenarios where lighting is non-linear

- it is very difficult or even impractical to simultaneously capture corrupted and ground truth images of the same visual scene (e.g., low-light and normal-light image pairs at the same time)

Issues of Learning-based Approaches

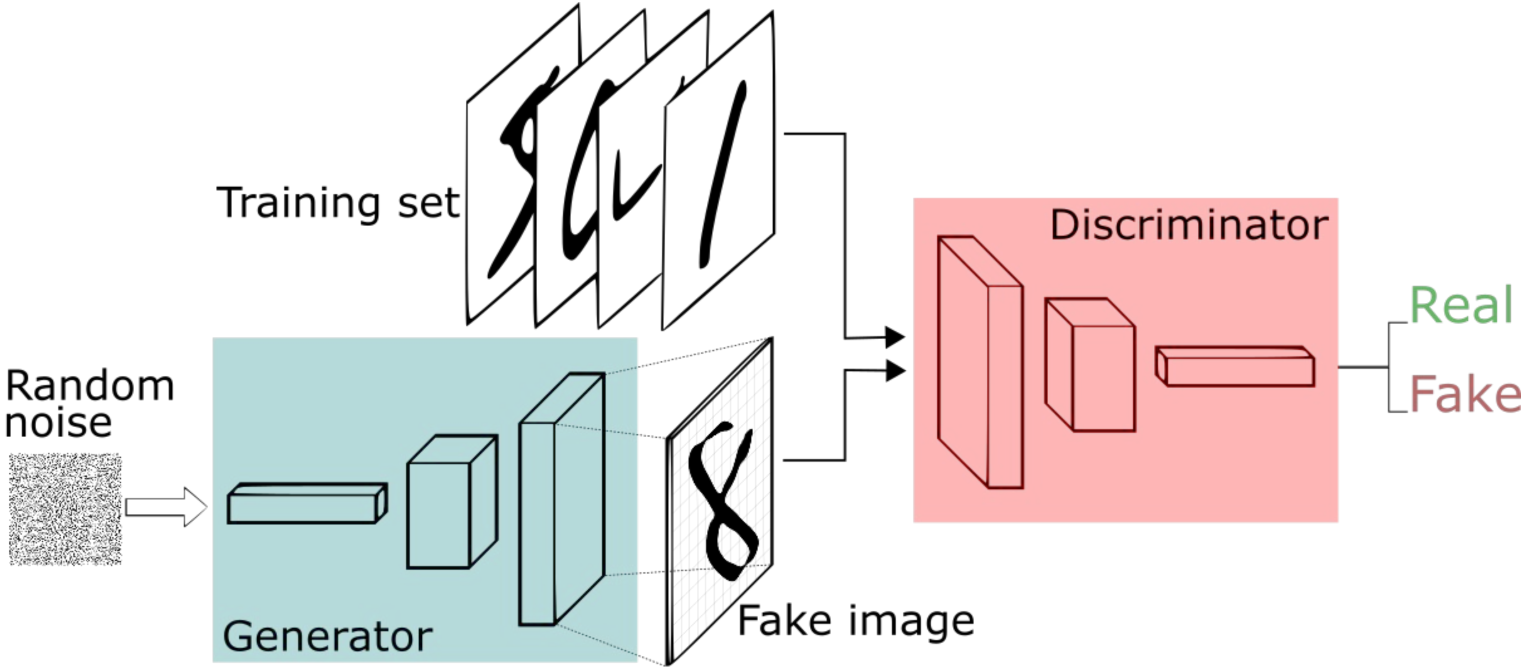
- Although globally adjusting the brightness of a natural image can create synthetic low-/normal-light data pairs, it can not simulate real-world scenarios where lighting is non-linear

- it is very difficult or even impractical to simultaneously capture corrupted and ground truth images of the same visual scene (e.g., low-light and normal-light image pairs at the same time)

Solution:

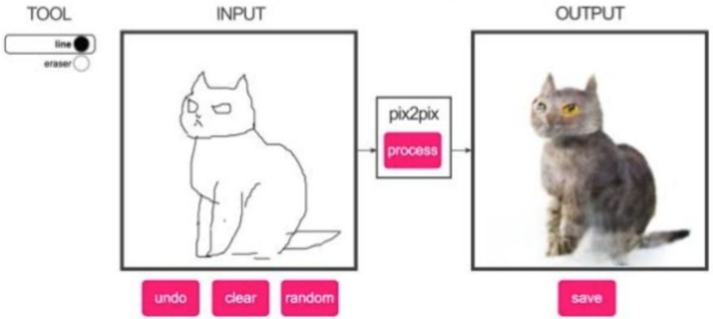
Instead of training a neural network with supervised signals, we seek for a way to learn it in an unsupervised manner

Insight from Generative Adversarial Networks



Generative Adversarial Networks

Karras et. al. 2018



Zhu et. al. 2016

Unsupervised Application from CycleGAN

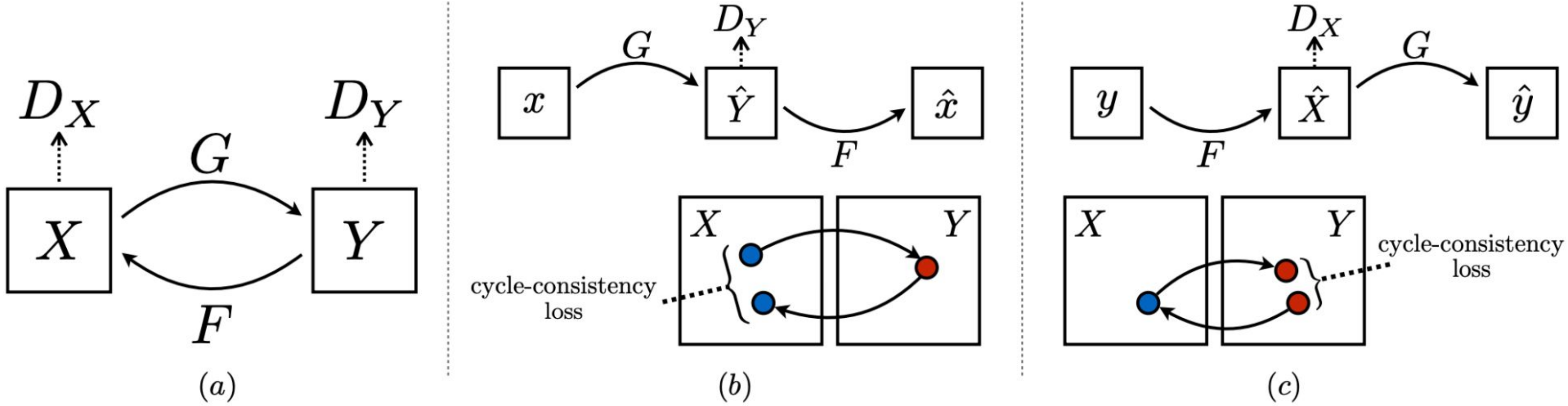
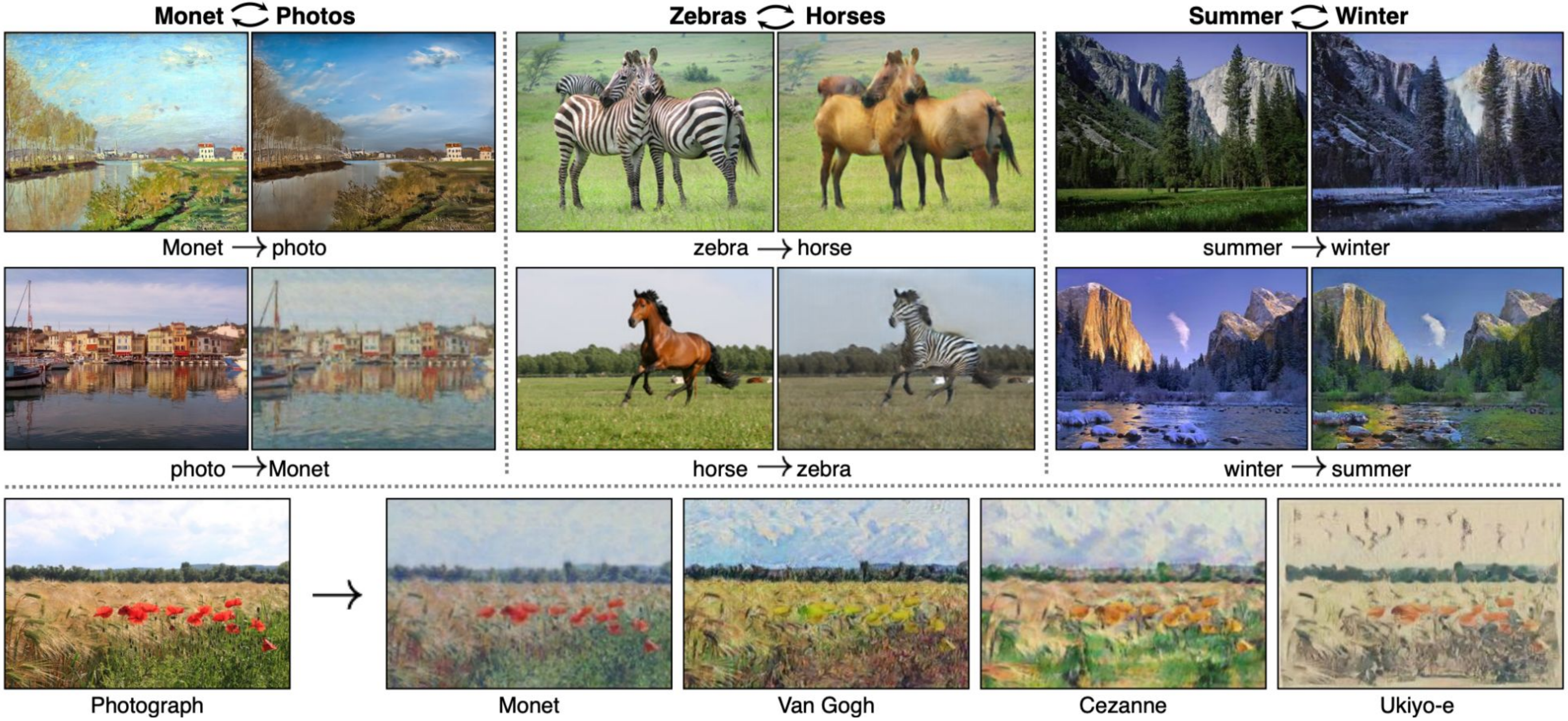


Figure 3: (a) Our model contains two mapping functions $G : X \rightarrow Y$ and $F : Y \rightarrow X$, and associated adversarial discriminators D_Y and D_X. D_Y encourages G to translate X into outputs indistinguishable from domain Y, and vice versa for D_X and F. To further regularize the mappings, we introduce two *cycle consistency losses* that capture the intuition that if we translate from one domain to the other and back again we should arrive at where we started: (b) forward cycle-consistency loss: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$, and (c) backward cycle-consistency loss: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$

Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks [Zhu et. al. 2017]

Unsupervised Application from CycleGAN

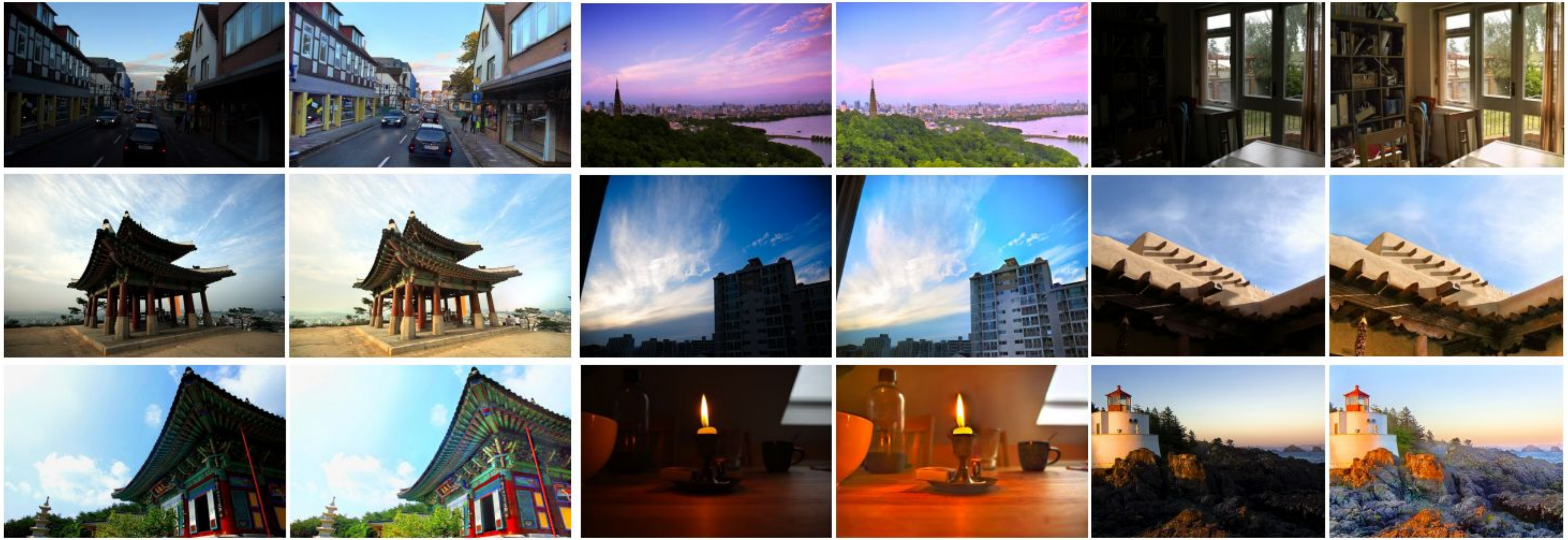


Issue with Cycle-consistency

Cycle-consistency solves image editing task well, however, does not perform well on image processing task, due to the lack of content constraint

Meanwhile, its dual-path strategy further increase the training time

Representative Examples Produced by EnlightenGAN

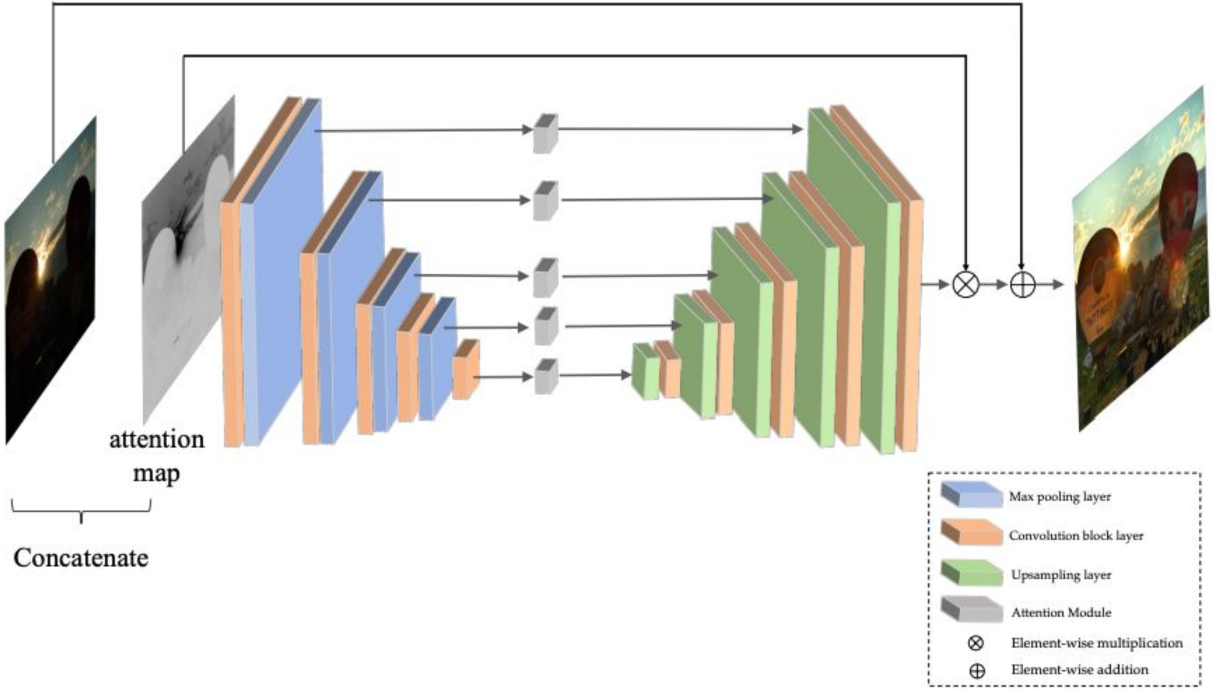


Dataset Preparation

- Training: We assemble a mixture of 914 low light and 1016 normal light images from several different datasets. Without the need to keep any pair. Manual inspection and selection are performed to remove images of medium brightness. All these photos are converted to PNG format and resized to 600 × 400 pixels.

- Evaluation: We choose those standard benchmark used in previous works (NPE , LIME , MEF , DICM , and VV.).

Main Pipeline



Loss Design: Global and Local Collaboration

Global:

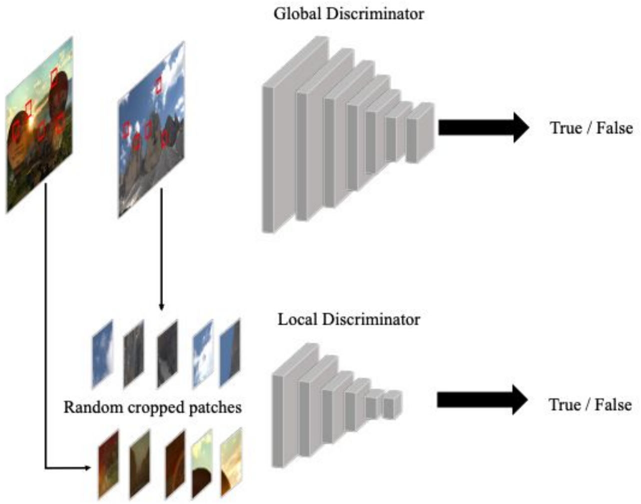
$$\mathcal{L}_D^{Global} = \mathbb{E}_{x_r \sim \mathbb{P}_{\text{real}}}[(D_{Ra}(x_r, x_f) - 1)^2] + \mathbb{E}_{x_f \sim \mathbb{P}_{\text{fake}}}[D_{Ra}(x_f, x_r)^2],$$

$$\mathcal{L}_G^{Global} = \mathbb{E}_{x_f \sim \mathbb{P}_{\text{fake}}}[(D_{Ra}(x_f, x_r) - 1)^2] + \mathbb{E}_{x_r \sim \mathbb{P}_{\text{real}}}[D_{Ra}(x_r, x_f)^2],$$

Local:

$$\mathcal{L}_D^{Local} = \mathbb{E}_{x_r \sim \mathbb{P}_{\text{real-patches}}}[(D(x_r) - 1)^2] + \mathbb{E}_{x_f \sim \mathbb{P}_{\text{fake-patches}}}[(D(x_f) - 0)^2],$$

$$\mathcal{L}_G^{Local} = \mathbb{E}_{x_r \sim \mathbb{P}_{\text{fake-patches}}}[(D(x_f) - 1)^2],$$



Loss Design: Self-Feature Preserving Loss

$$\mathcal{L}_{SFP}(I^L) = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^L) - \phi_{i,j}(G(I^L)))^2,$$

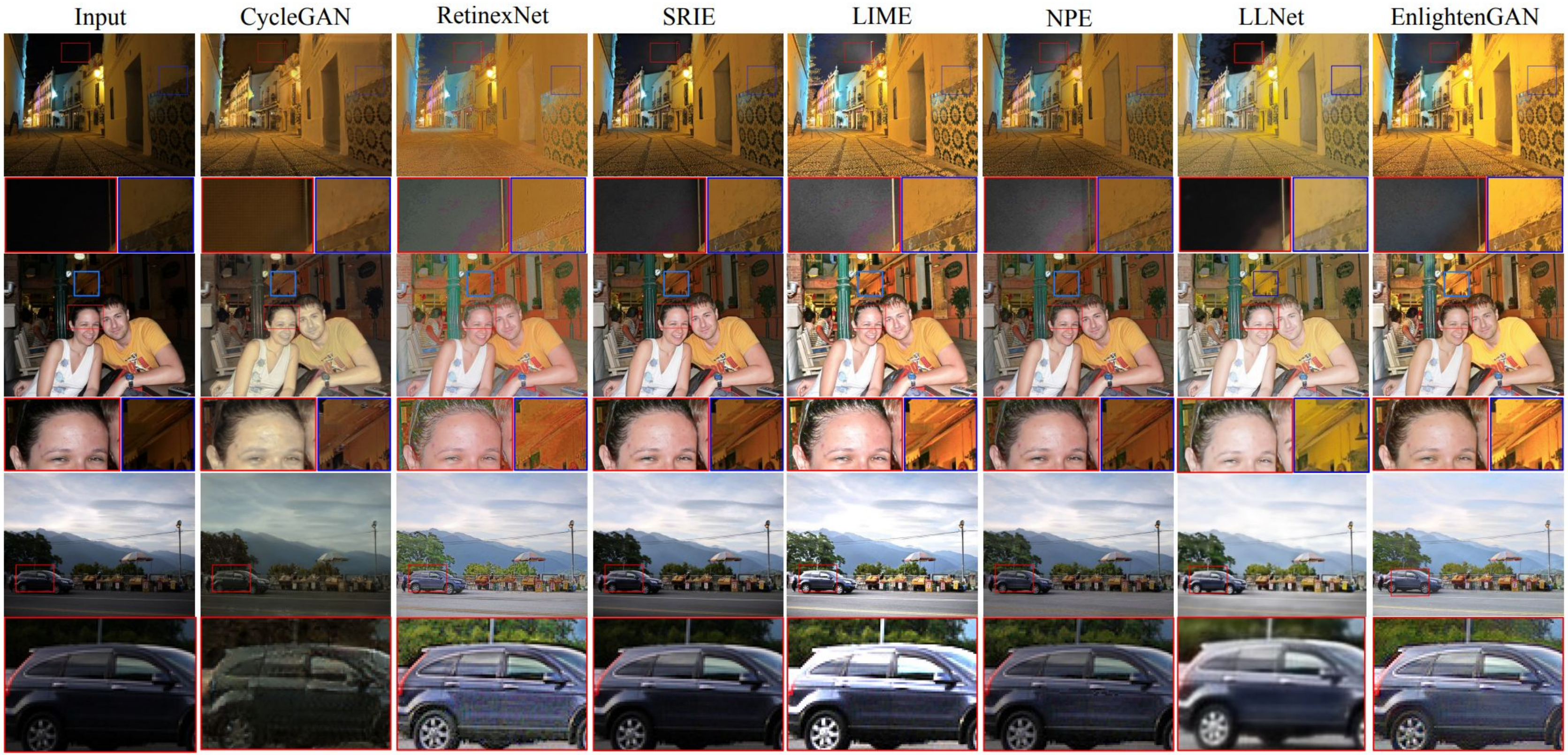
Loss Design: Self-feature Preserving Loss

$$\mathcal{L}_{SFP}(I^L) = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^L) - \phi_{i,j}(G(I^L)))^2,$$

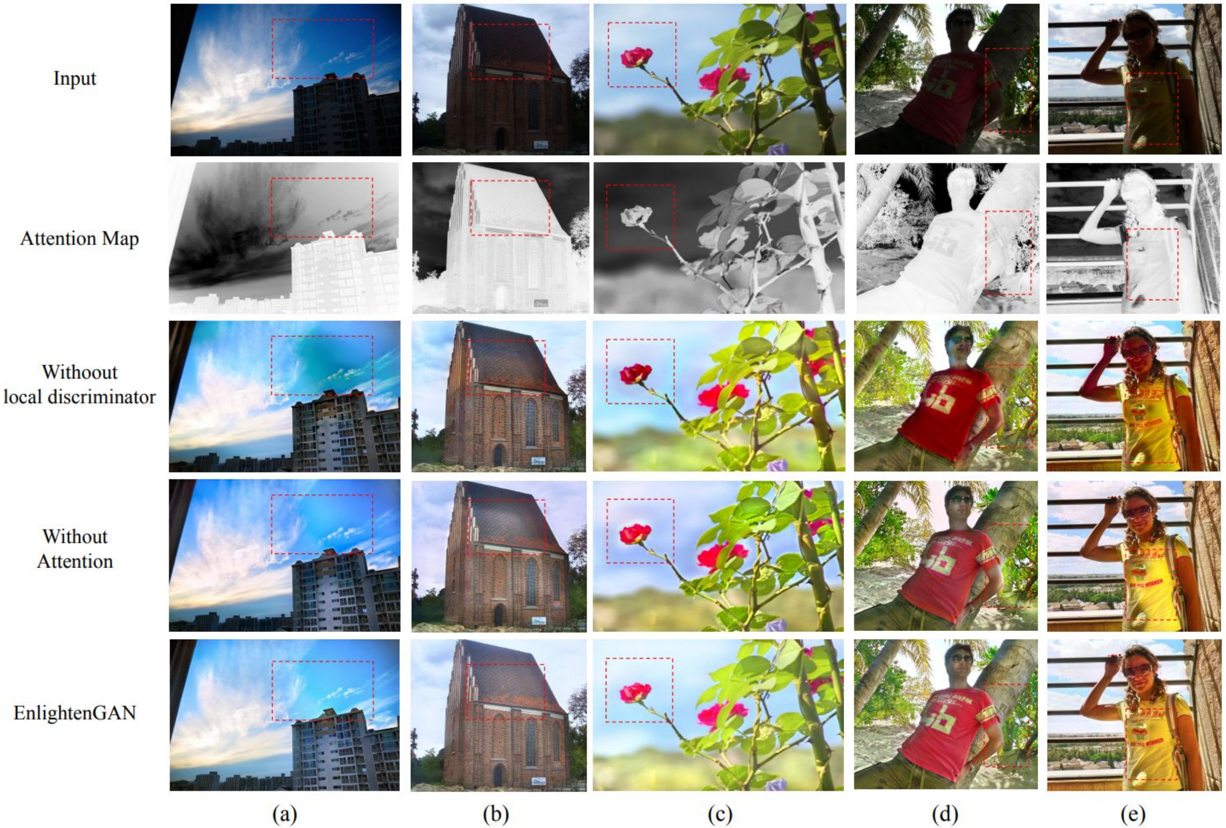
Final Loss Function:

$$Loss = \mathcal{L}_{SFP}^{Global} + \mathcal{L}_{SFP}^{Local} + \mathcal{L}_{G}^{Global} + \mathcal{L}_{G}^{Local}$$

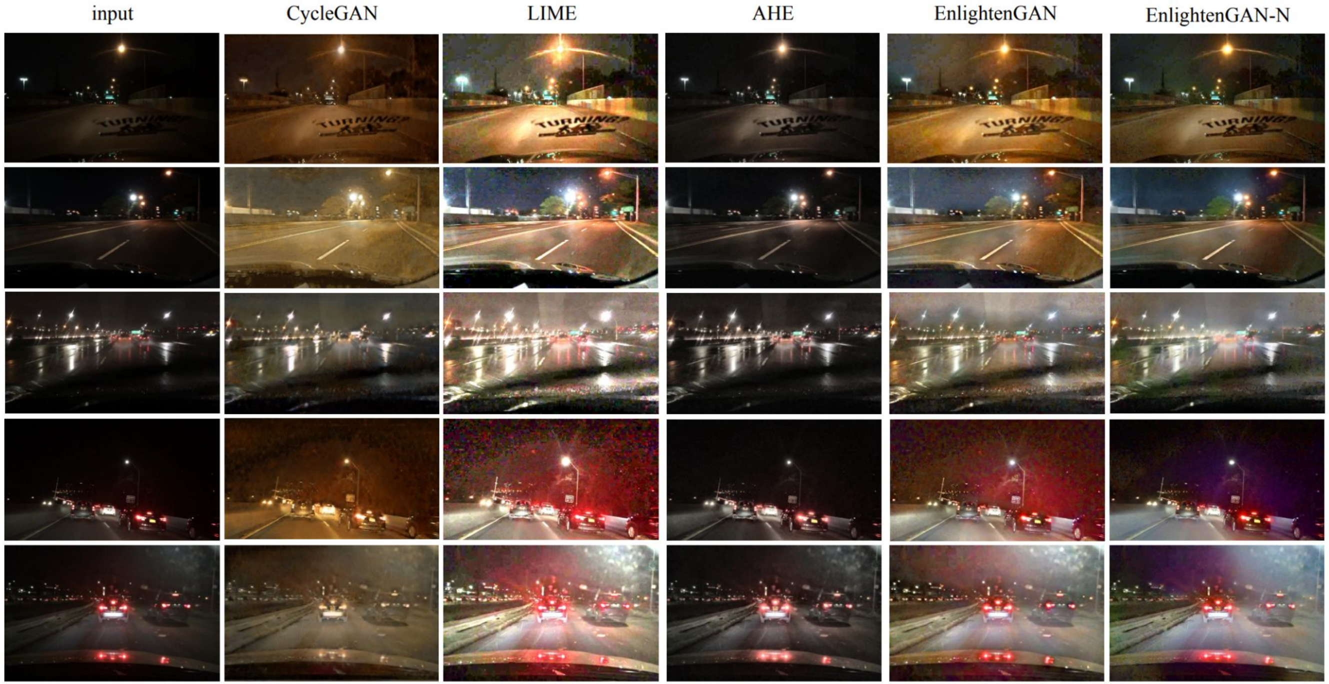
Visual Comparison



Ablation Study



Real-world Generalization

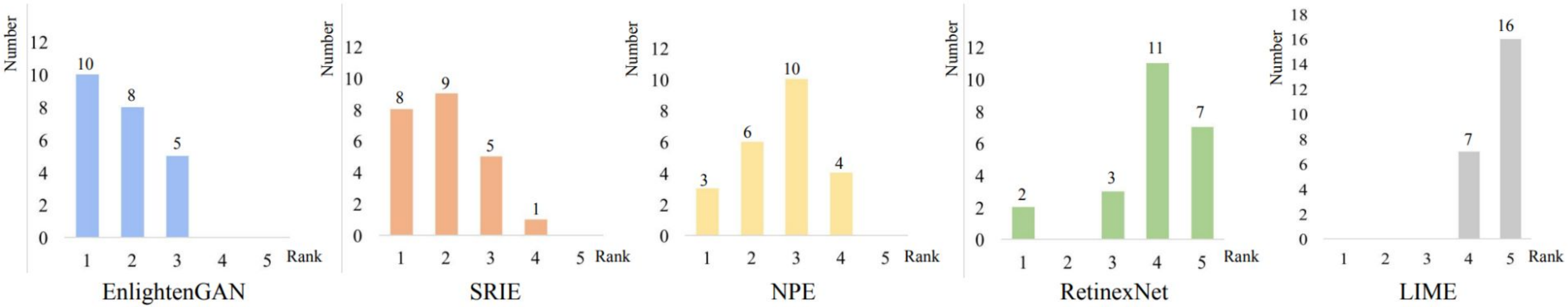


Quantitative Evaluation

TABLE I: NIQE scores on the whole testing set (All) and each subset (MEF, LIME, NPE, VV, DICM) respectively. Smaller NIQE indicates more perceptually favored quality.

Image set	MEF	LIME	NPE	VV	DICM	All
Input	4.265	4.438	4.319	3.525	4.255	4.134
LLNet	4.845	4.940	4.78	4.446	4.809	4.751
CycleGAN	3.782	3.276	4.036	3.343	3.560	3.554
RetinexNet	4.149	4.420	4.485	2.602	4.200	3.920
LIME	3.720	4.155	4.268	**2.489**	3.846	3.629
SRIE	3.475	3.788	3.986	2.850	3.899	3.650
NPE	3.524	3.905	**3.953**	2.524	3.760	3.525
EnlightenGAN	**3.232**	**3.719**	4.113	2.581	**3.570**	**3.385**

Human Study


Number
Rank
10
8
5
EnlightenGAN
8
9
5
1
SRIE
3
6
10
4
NPE
2
3
11
7
RetinexNet
7
16
LIME

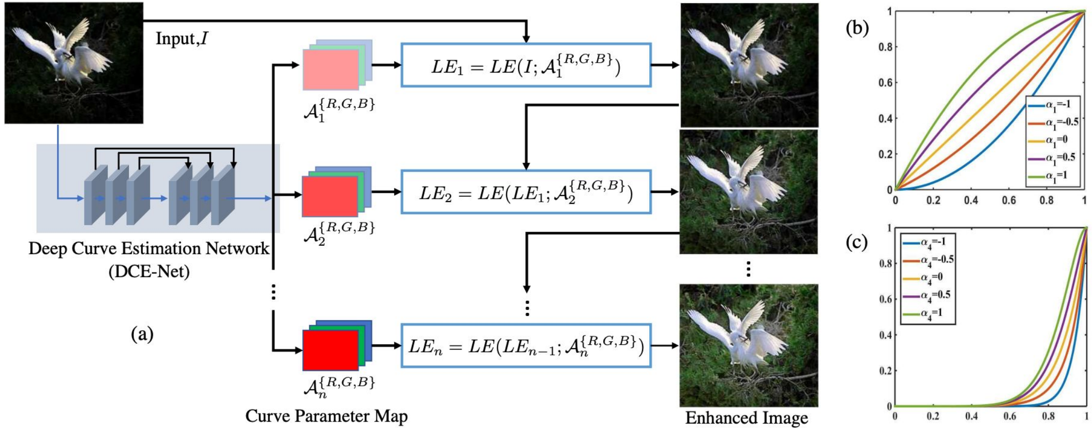
Summarization

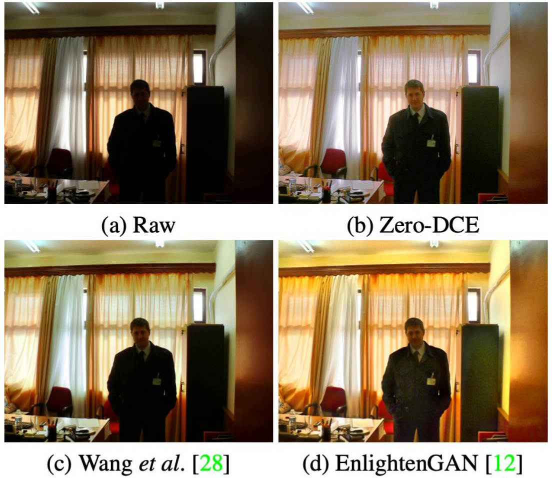
- EnlightenGAN is the <u>first</u> work that successfully introduces unpaired training to low-light image enhancement. Such a training strategy removes the dependency on paired training data and enables us to train with larger varieties of images from different domains. It also avoids overfitting any specific data generation protocol or imaging device, hence leading to notably improved real-world generalization

- EnlightenGAN gains remarkable performance by imposing (i) a global-local discriminator structure that handles spatially-varying light conditions in the input image; (ii) the idea of self feature preserving loss and attention mechanism.

- EnlightenGAN is compared with several state-of-the-art methods via comprehensive experiments. The results are measured in terms of visual quality, no-referenced image quality assessment, and human subjective survey. All results consistently endorse the superiority of EnlightenGAN.

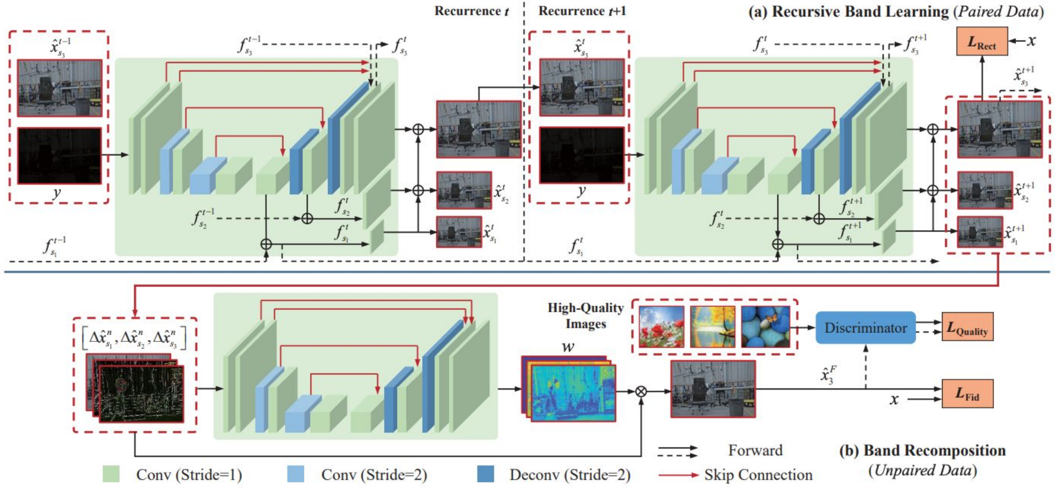
Followers’ Works Motivated by EnlightenGAN

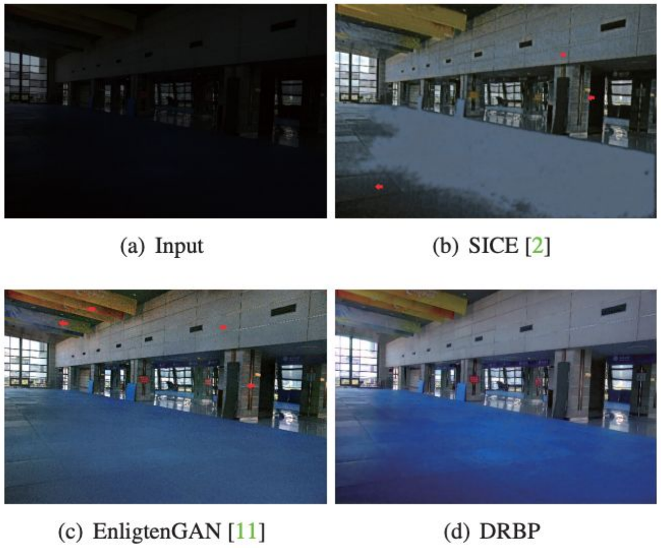




Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. [Li et. al. 2020]

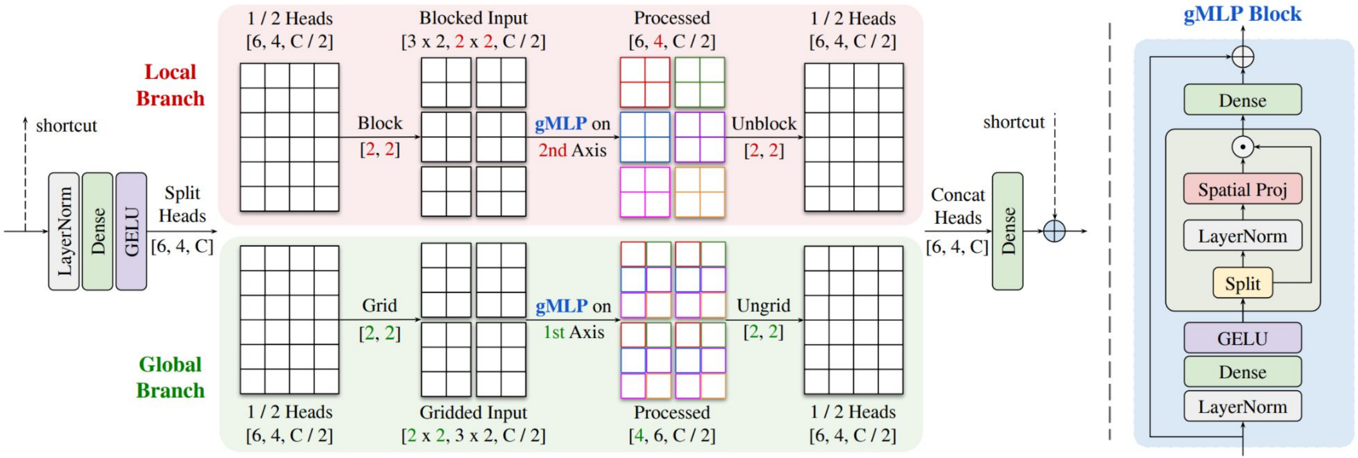
Followers’ Works Motivated by EnlightenGAN

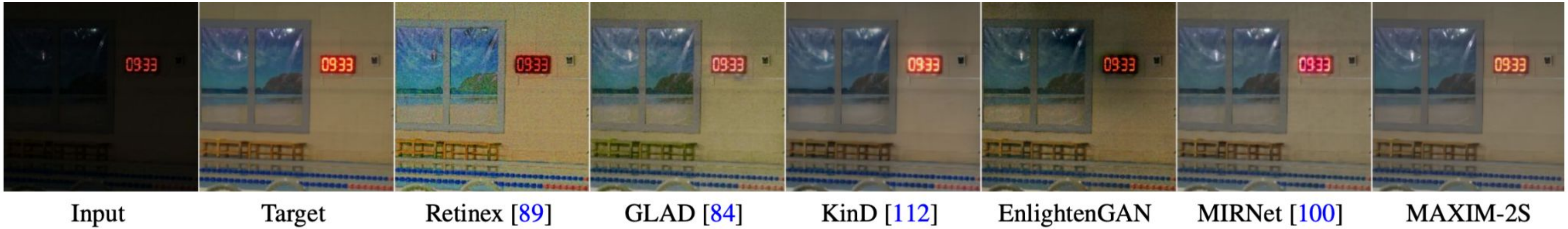




From Fidelity to Perceptual Quality: A Semi-Supervised Approach for Low-Light Image Enhancement. [Yang et. al.]

Followers’ Works Motivated by EnlightenGAN





MAXIM: Multi-Axis MLP for Image Processing. [Tu et. al. 2022]

Thanks!