IVUL Group Meeting

Ali K. Thabet

27/09/2018

What will we cover?

- 2 papers from ECCV ‘18 on Computational Photography:
    1. Deep Recursive HDRI: Inverse Tone Mapping using Generative Adversarial Networks
    2. Deep High Dynamic Range Imaging with Large Foreground Motions

Computational Photography

"*Digital image capture and processing techniques that use digital computation instead of optical processes*" Wikipedia

Examples: in camera panoramas, HDRI, light field cameras, etc...

Deep Recursive HDRI: Inverse Tone Mapping using Generative Adversarial Networks

Siyeong Lee, Gwon Hwan An, Suk-Ju Kang
Sogang University

HDRI

- HDRI: High Dynamic Range Image
- Low Dynamic Range (LDR) images characterized by exposure value (EV):

$$EV = 2log_2 F - log_2 S + log_2 \frac{ISO}{100}$$

- LDR images have under/over-exposed areas
- Traditionally: combine several LDRs to get HDRI

HDRI - LDRs at different EVs

Images from https://www.digitaltrends.com/photography/what-is-hdr-photography/

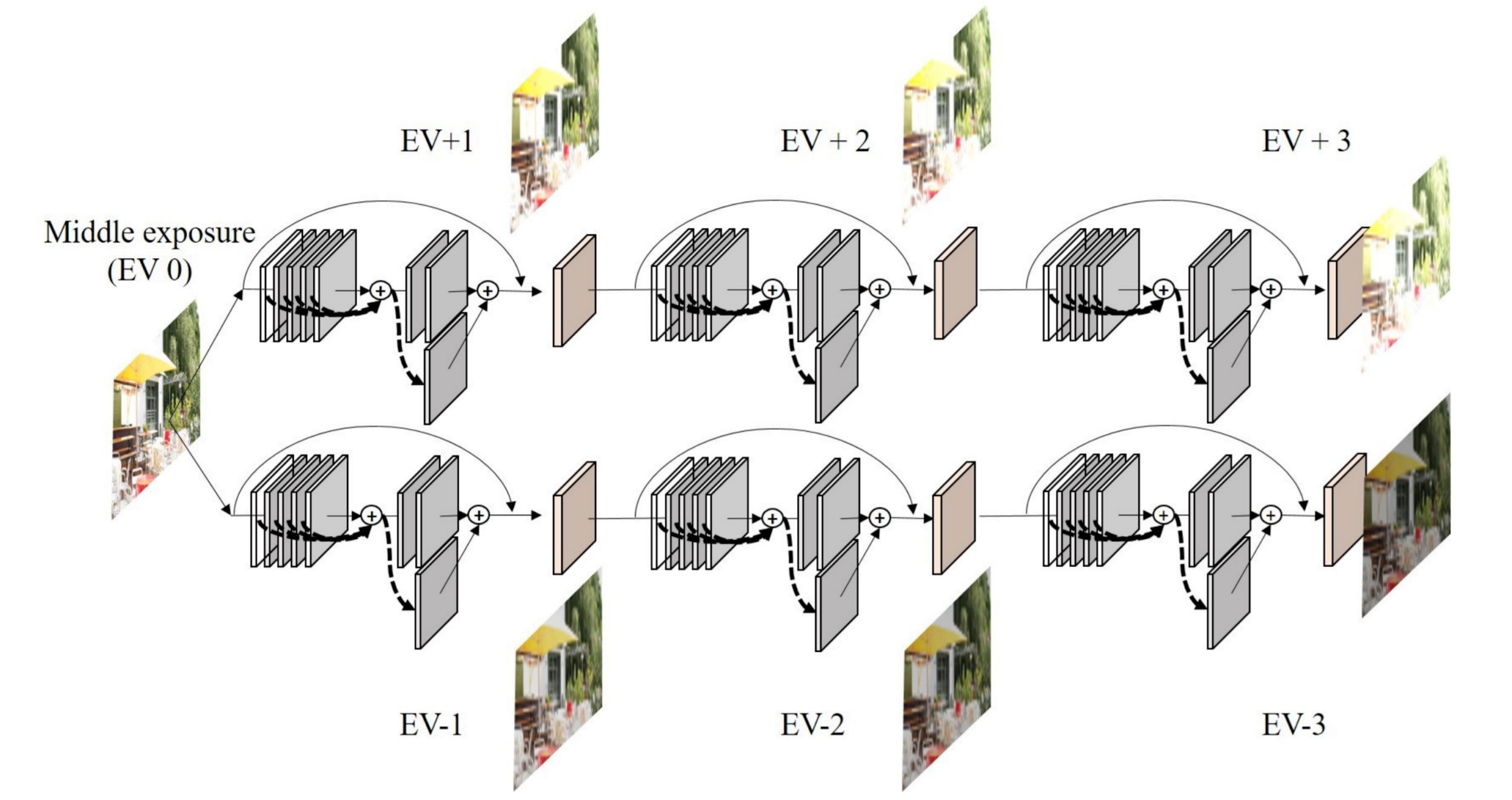
HDRI - Final DHRI

Images from https://www.digitaltrends.com/photography/what-is-hdr-photography/

Inverse Tone Mapping

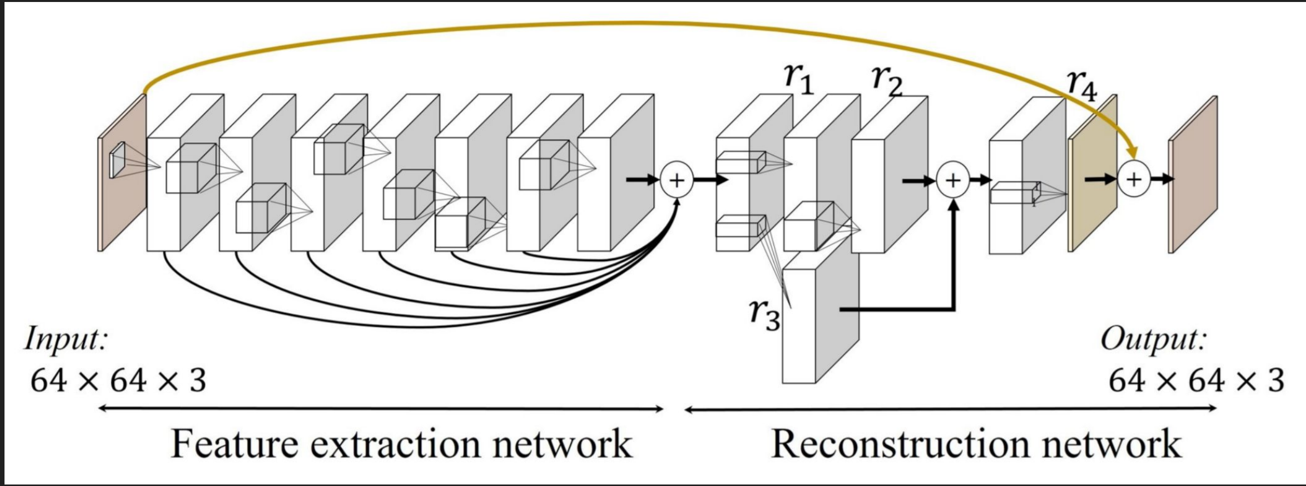
- Generating an HDR image with one LDR image
- Ill-posed problem: restore missing signal not appearing on input
- Deep Learning as a solution

Deep Chain HDRI



Deep Chain HDRI

- Estimate several exposures around "0" EV image
- Central EV image determined as one with uniform histogram
- Combine estimated LDR images to get HDRI
- Each EV image computed through subnetwork



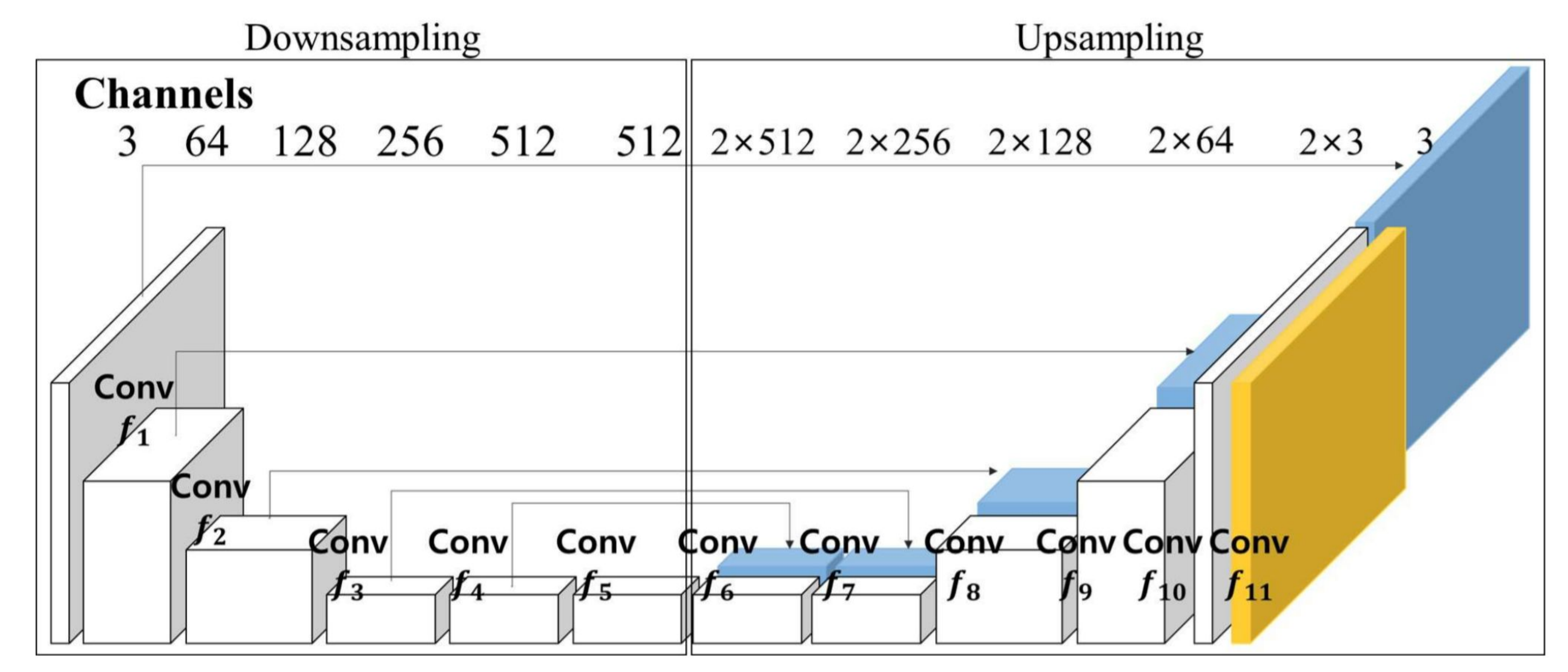
Deep Recursive HDRI

- Instead of multiple subnetworks, have 2 networks, + and -
- Each network estimates one EV above (below) an image
- Networks are Conditional GANs (G^+, D^+) and (G^-, D^-)
- (G^+, D^+) and (G^-, D^-) solve min-max problem:

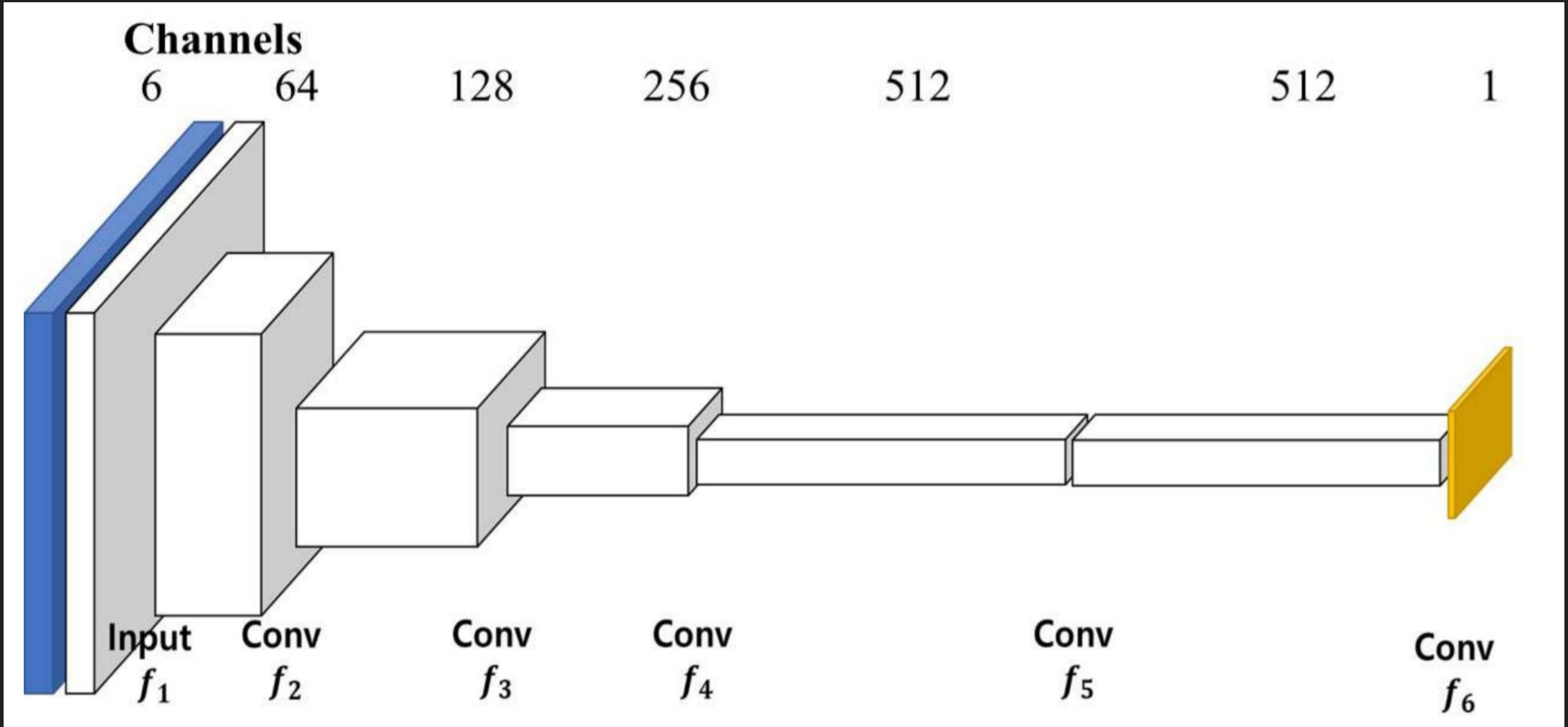
$$\begin{aligned} G^+, D^+ = \min_G \max_D \{ & E_{I^{EV_{i+1}}, I^{EV_i}} [\log D(I^{EV_{i+1}}, I^{EV_i})] \\ & + E_{I^{EV_i}, z} [1 - \log D(G(I^{EV_i}, z), I^{EV_i})] \} \end{aligned}$$

$$\begin{aligned} G^-, D^- = \min_G \max_D \{ & E_{I^{EV_{i-1}}, I^{EV_i}} [\log D(I^{EV_{i-1}}, I^{EV_i})] \\ & + E_{I^{EV_i}, z} [1 - \log D(G(I^{EV_i}, z), I^{EV_i})] \} \end{aligned}$$

Generator: U-Net



Discriminator: PatchGAN



Loss Functions

$$G^{plus} = \arg\min_G L_{LSGAN}(G) + \lambda L_{L1}(G) \text{ for training pairs } (I^{EV1}, I) \text{ and}$$

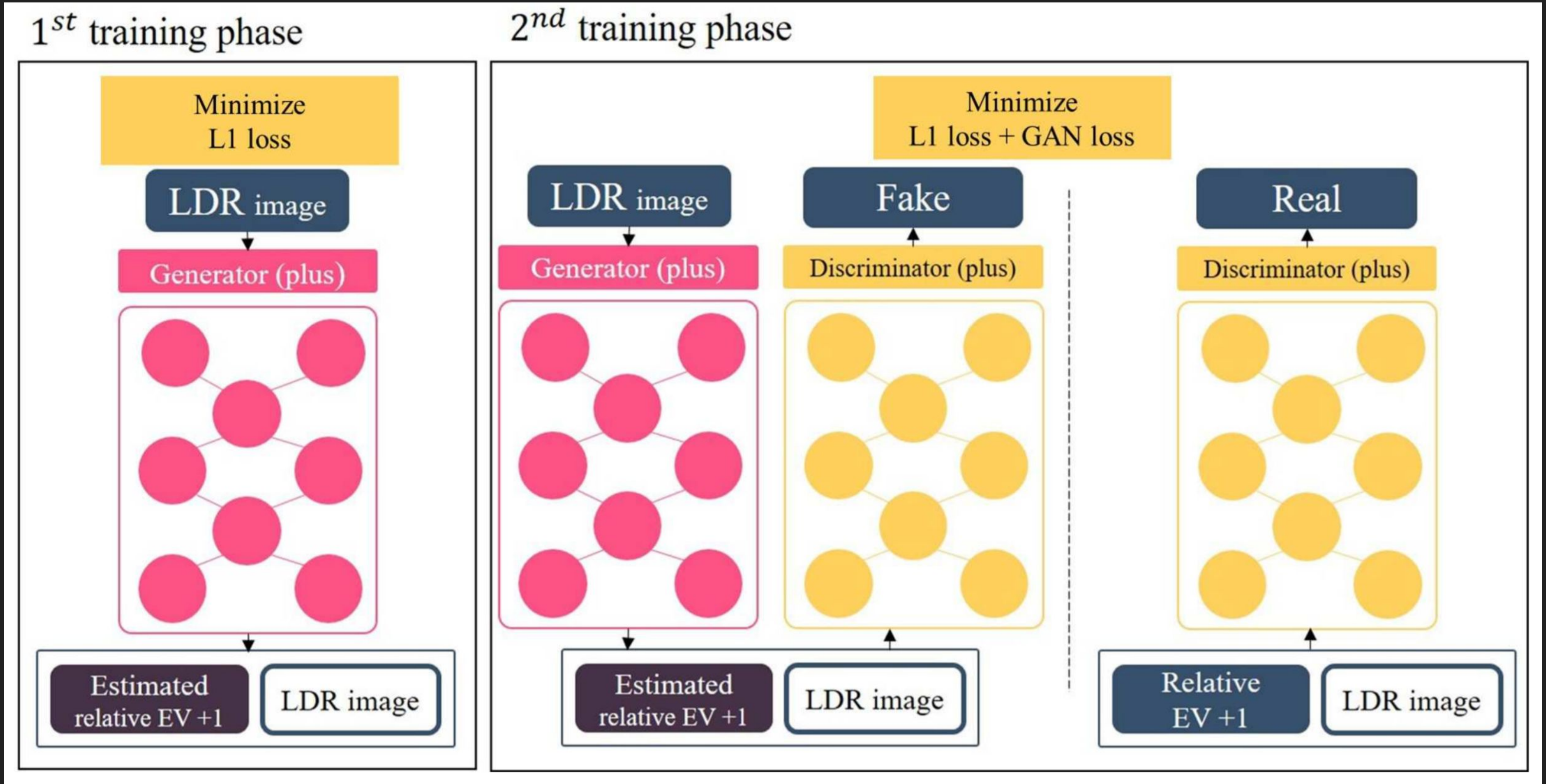
$$G^{minus} = \arg\min_G L_{LSGAN}(G) + \lambda L_{L1}(G) \text{ for training pairs } (I^{EV-1}, I),$$

$$L_{LSGAN}(D) = \frac{1}{2}\mathbb{E}_{x,y}[(D(y,x) - 1)^2] + \frac{1}{2}\mathbb{E}_{x,z}[(D(G(x,z),x))^2],$$

$$L_{LSGAN}(G) = \mathbb{E}_{x,z}[(D(G(x,z),x) - 1)^2],$$

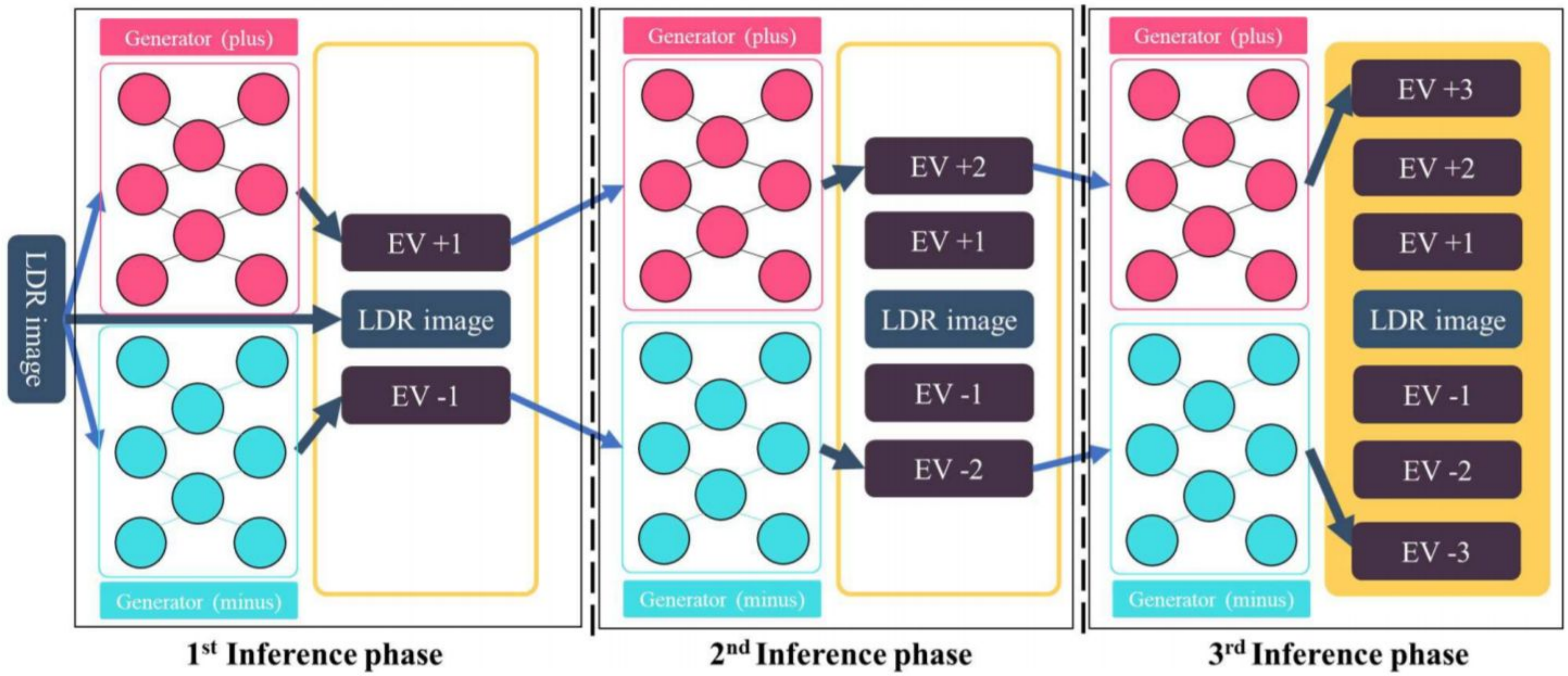
$$L_{L1}(G) = \mathbb{E}_{x,y,z}[||y - G(x,z)||_1].$$

Training



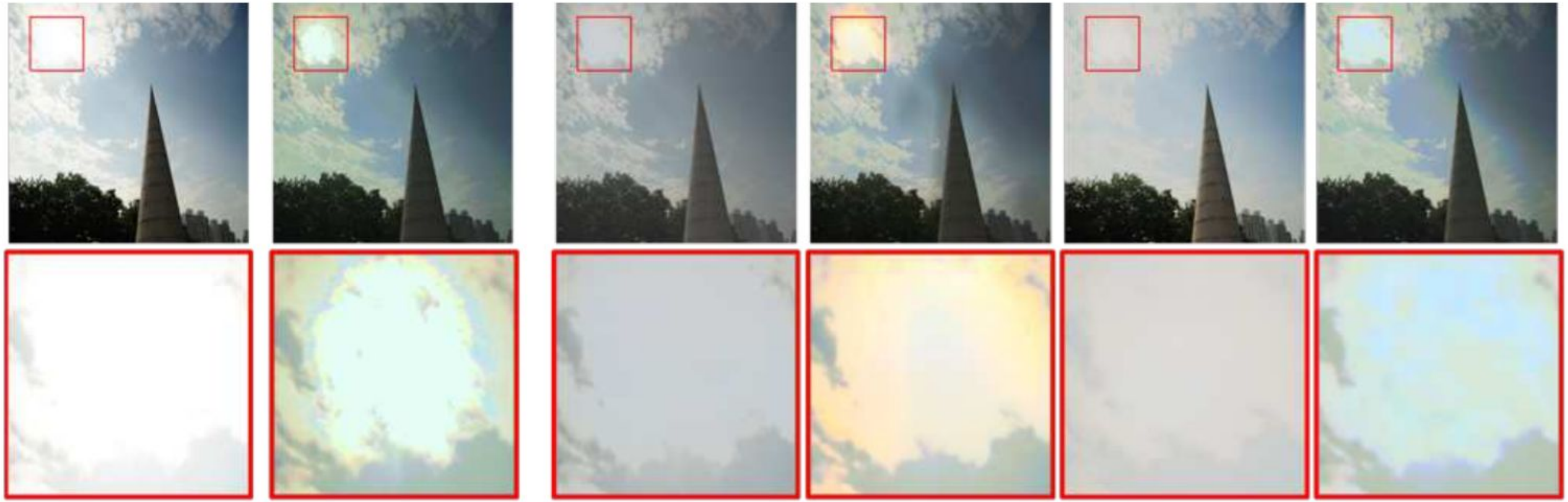
1st training phase
Minimize
L1 loss
LDR image
Generator (plus)
Estimated relative EV +1
LDR image
2nd training phase
Minimize
L1 loss + GAN loss
LDR image
Generator (plus)
Fake
Discriminator (plus)
Real
Discriminator (plus)
Estimated relative EV +1
LDR image
Relative EV +1
LDR image

Inference



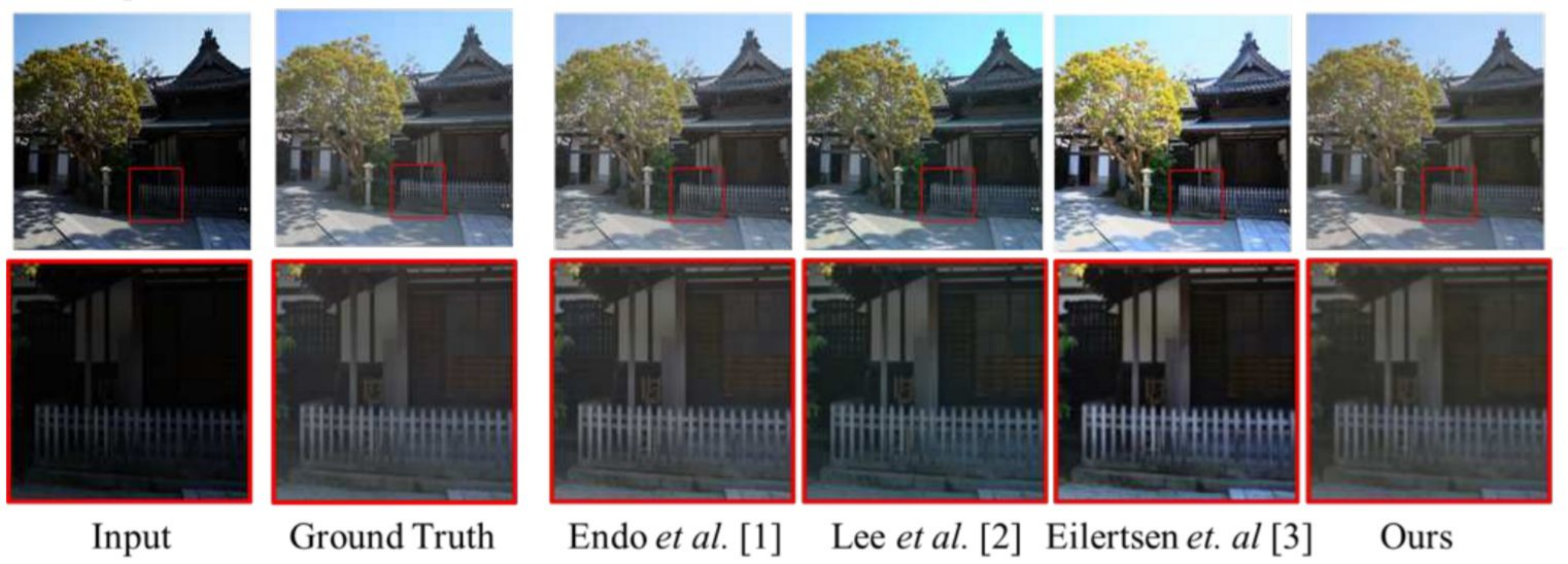
LDR image
Generator (plus)
Generator (minus)
EV +1
LDR image
EV -1
Generator (plus)
Generator (minus)
EV +2
EV +1
LDR image
EV -1
EV -2
Generator (plus)
Generator (minus)
EV +3
EV +2
EV +1
LDR image
EV -1
EV -2
EV -3
1st Inference phase
2nd Inference phase
3rd Inference phase

Results



Input Ground Truth Endo *et al.* [1] Lee *et al.* [2] Eilertsen *et al.* [3] Ours

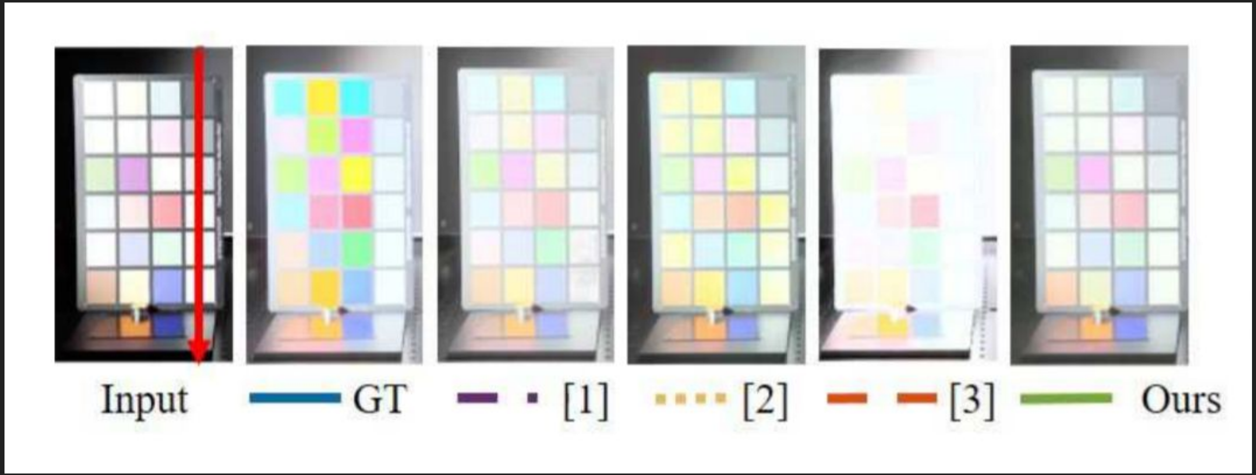
Results


Input
Ground Truth
Endo *et al.* [1]
Lee *et al.* [2]
Eilertsen *et. al* [3]
Ours

Results

Input Ground Truth Endo *et al.* [1] Lee *et al.* [2] Eilertsen *et al.* [3] Ours

Results



Deep High Dynamic Range Imaging with Large Foreground Motions

Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang
The Hong Kong University of Science and Technology
Tencent Youtu
University of Oxford

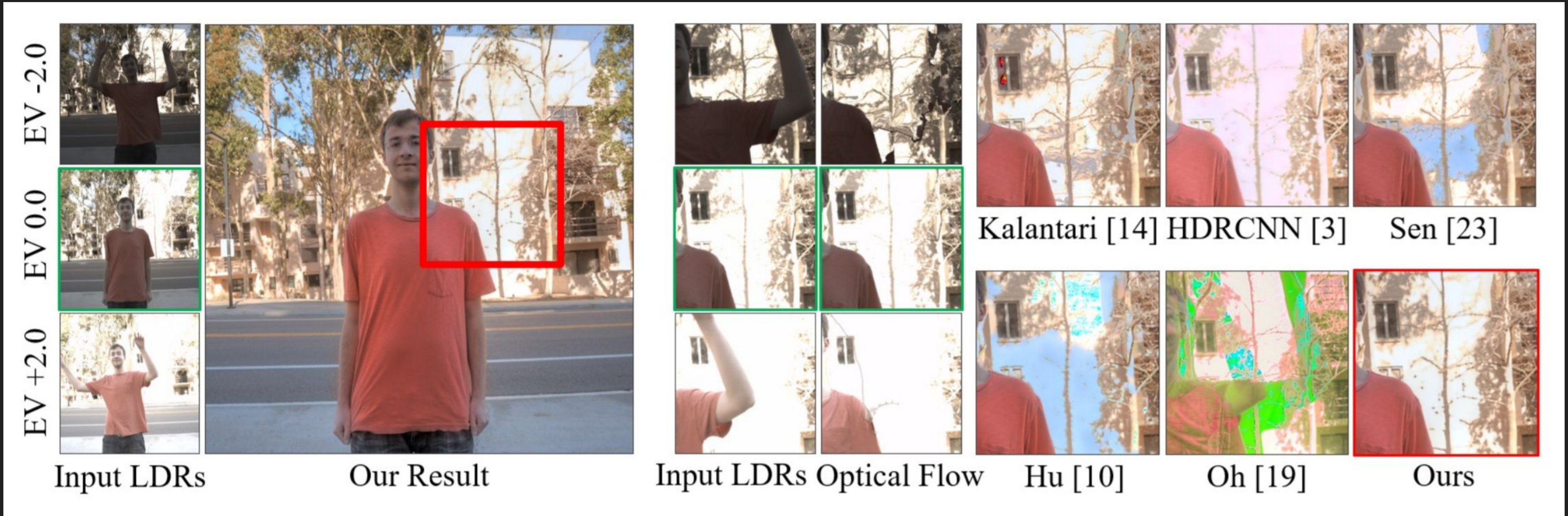
Motivation

- Aligning LDRs requires no motion or image alignment
- Large foreground motion causes artifacts in HDRI synthesis
- Current methods use optical flow to align images
- Flow methods can fail if images captured with different EVs

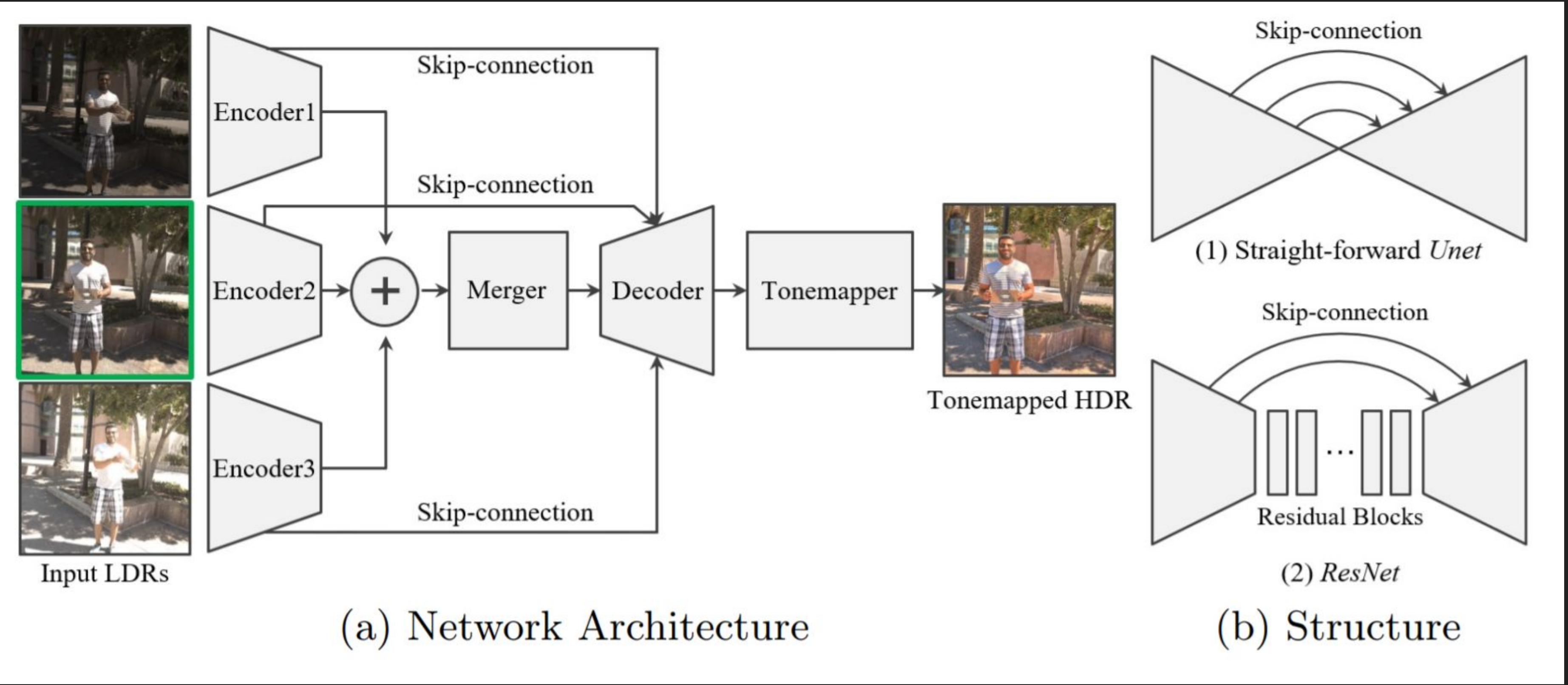
Solution:

Simple end-to-end network that can learn to translate multiple LDR images into a ghost-free HDR image even in the presence of large foreground motions

Motivation


EV -2.0
EV 0.0
EV +2.0
Input LDRs
Our Result
Input LDRs
Optical Flow
Kalantari [14]
HDRCNN [3]
Sen [23]
Hu [10]
Oh [19]
Ours

Methodology



Processing Pipeline

- Input is series of LDRs I = {I_1, I_2, I_3}
- LDRs mapped to HDR domain using:

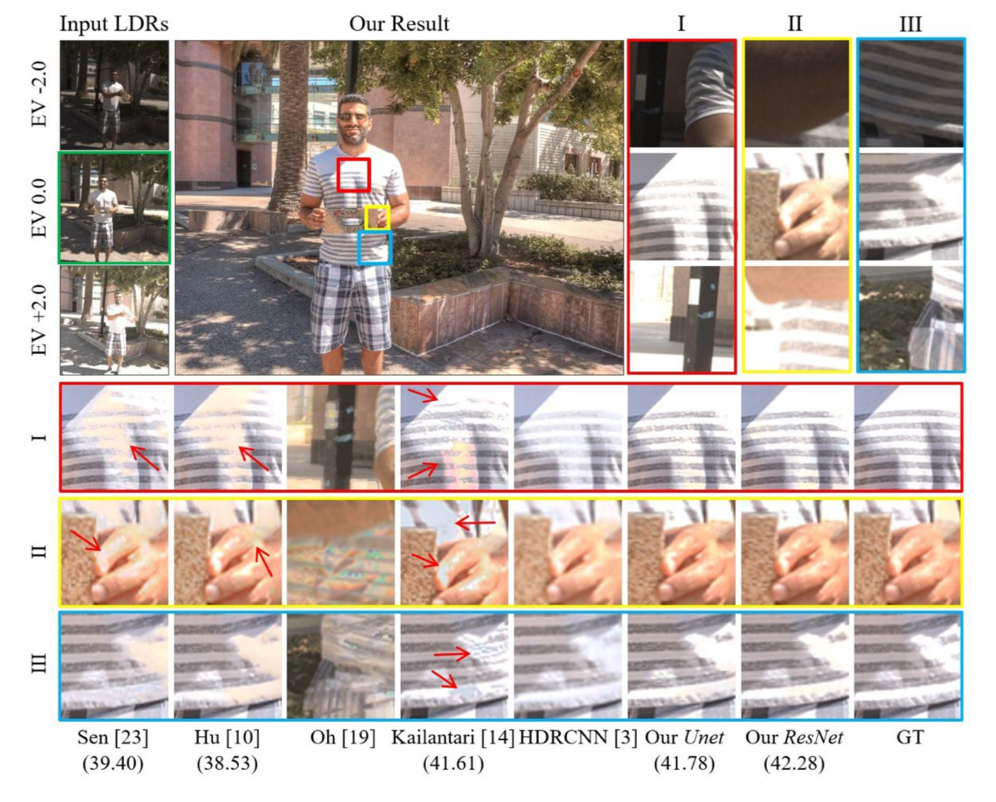
$$H_i = \frac{I_i}{t_i}$$

- I_i and H_i concatenated into 6-channel input
- Network estimates final HDR image, but loss computed on tonemapper:

$$\tau(H) = \frac{\log(1+\mu H)}{\log(1+\mu)}$$

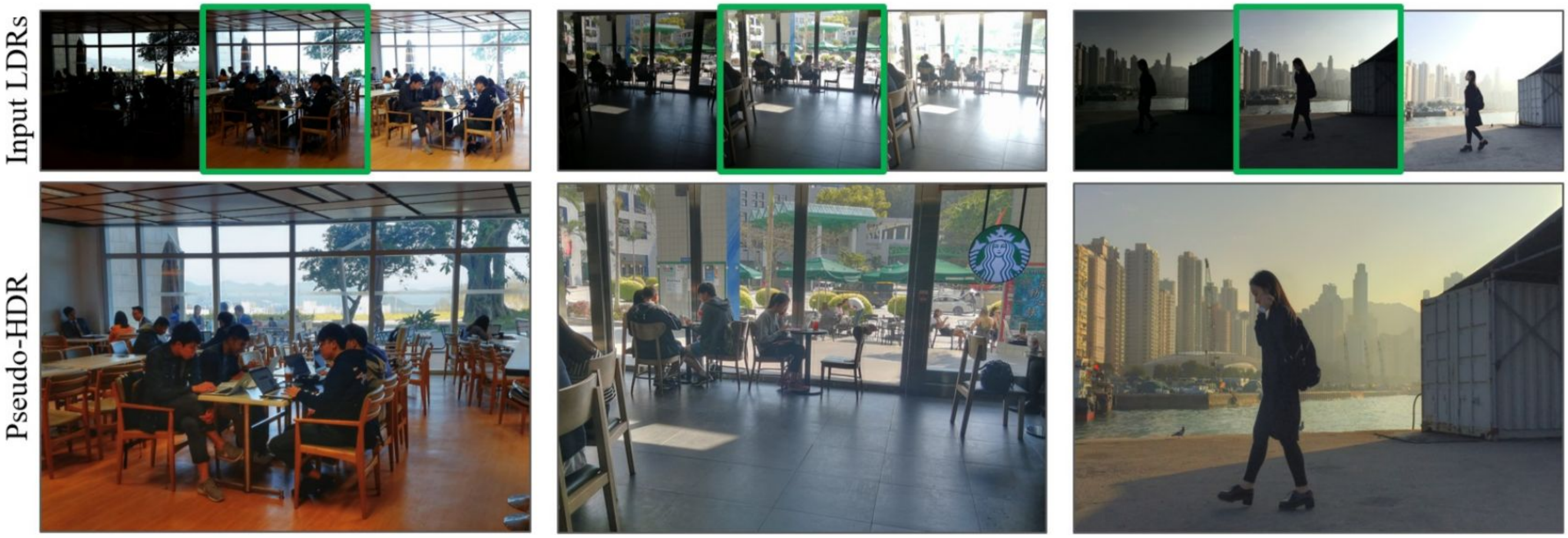
$$L_{Unet} = ||\tau(\hat{H}) - \tau(H)||_2$$

Results



Input LDRs
Our Result
I
II
III
EV -2.0
EV 0.0
EV +2.0
I
II
III
Sen [23] (39.40)
Hu [10] (38.53)
Oh [19]
Kailantari [14] (41.61)
HDRCNN [3]
Our *Unet* (41.78)
Our *ResNet* (42.28)
GT

Results



(a) Samsung Galaxy S5 (b) Huawei Mate 9 (c) iPhone 6s