北京大學
PEKING UNIVERSITY

Low-Light Image Enhancement for Intelligent Analytics

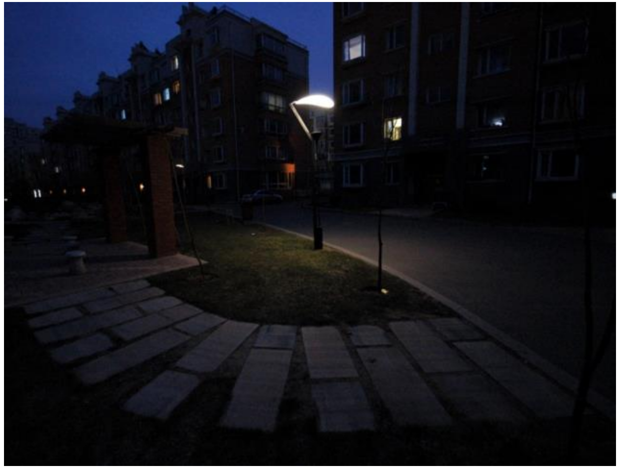
Jiaying Liu
Wangxuan Institute of Computer Technology
Peking University

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

MIPI Workshop in conjunction with CVPR 2023

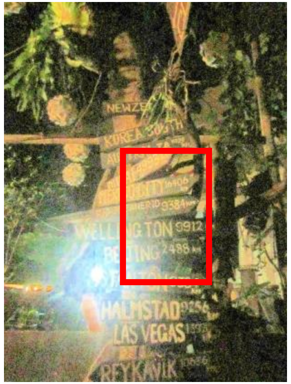
Low-Light Degradation

- **Low Visibility**
  - Details are buried due to *degraded contrast* and *low illumination*



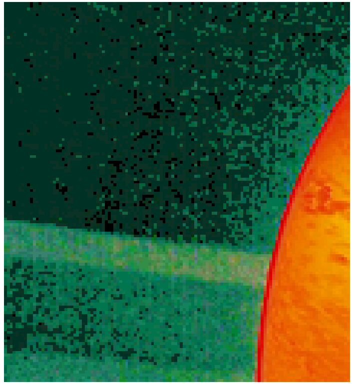
Low-Light Degradation

- **Intensive Noises**
  - After simple operations, *e.g. histogram equalization*, noises become noticeable!



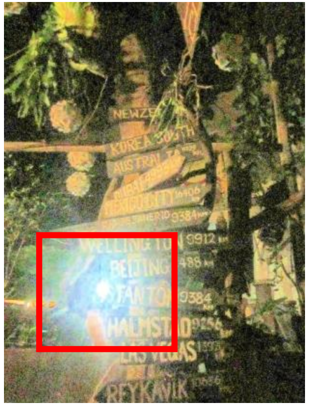
Low-Light Degradation

- **Intensive Noises**
  - After simple operations, *e.g. histogram equalization*, noises become noticeable!



Low-Light Degradation

- **Non-Uniform Illumination**
  - Under-exposures
  - Over-exposures



Low-Light Degradation

■ **Problem:** High-level vision in low-light scenarios

Low light degrades not only human vision but also machine vision

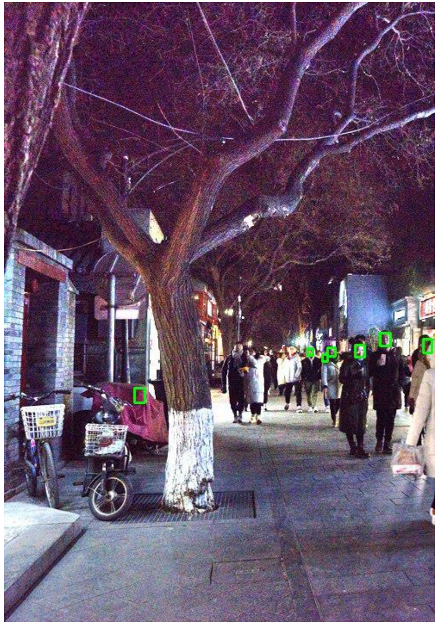
- Nighttime autonomous-driving
- Surveillance video analysis
- Low-light face detection

…

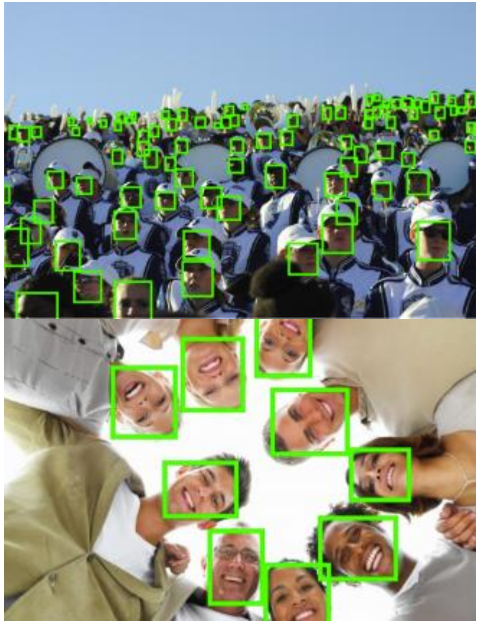
Low-Light Degradation

- **Domain Gaps** Between Low and Normal Light Images

Low-Light



LIME[TIP17] processed

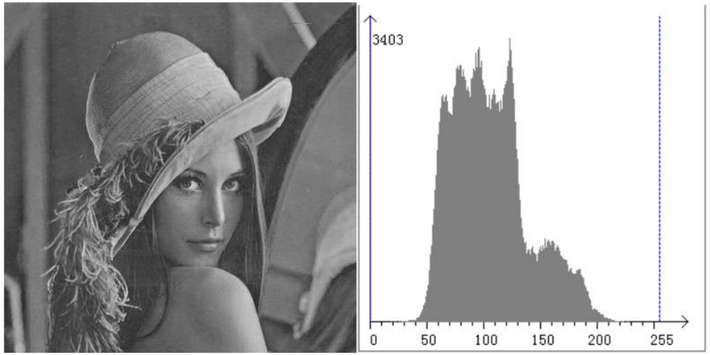


Normal Light

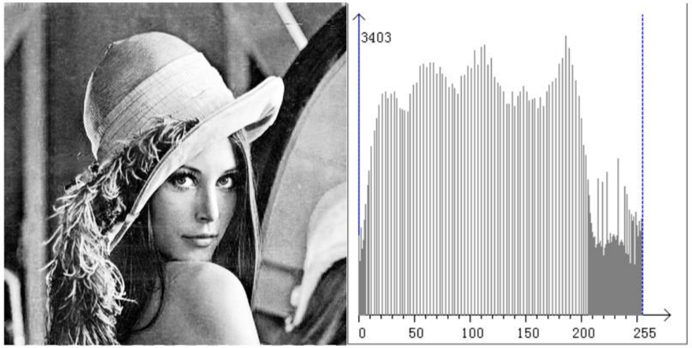
Representative Work

Histogram Equalization

- Enhance the contrast
- Over-enhancement / under-enhancement
- Amplify the noise

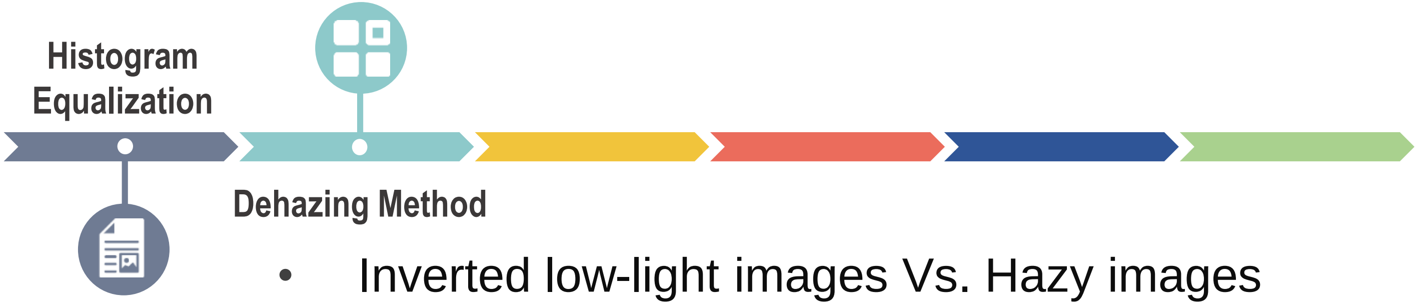


Before HE



After HE

Representative Work

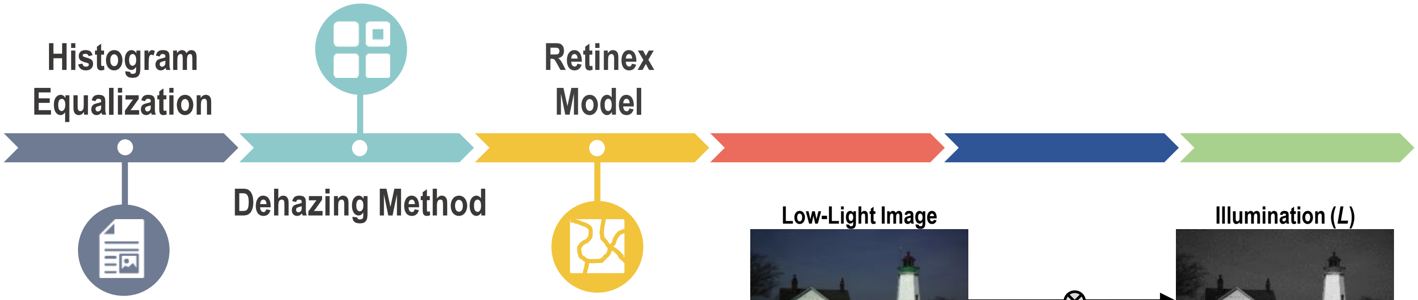


- Inverted low-light images Vs. Hazy images

Low-Light Inversion Dehazing Result

- Invert → dehaze → invert again
- Require an additional denoising process

Representative Work



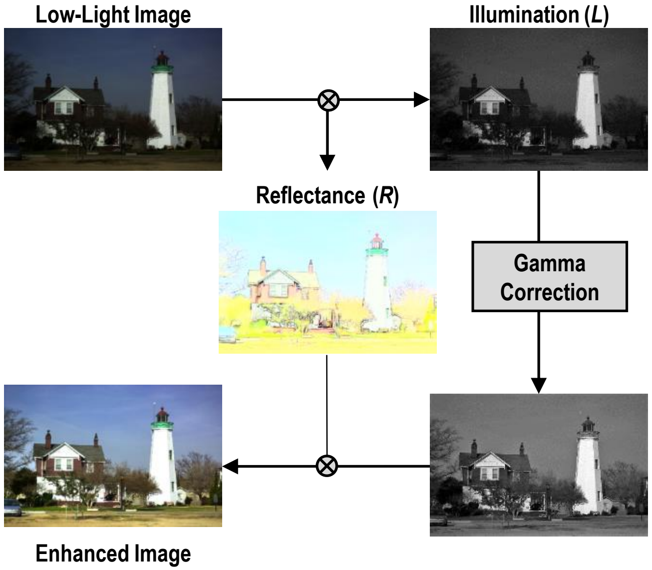
Retinex-Based Methods

- Retinex decomposition

$$S = R \cdot L$$

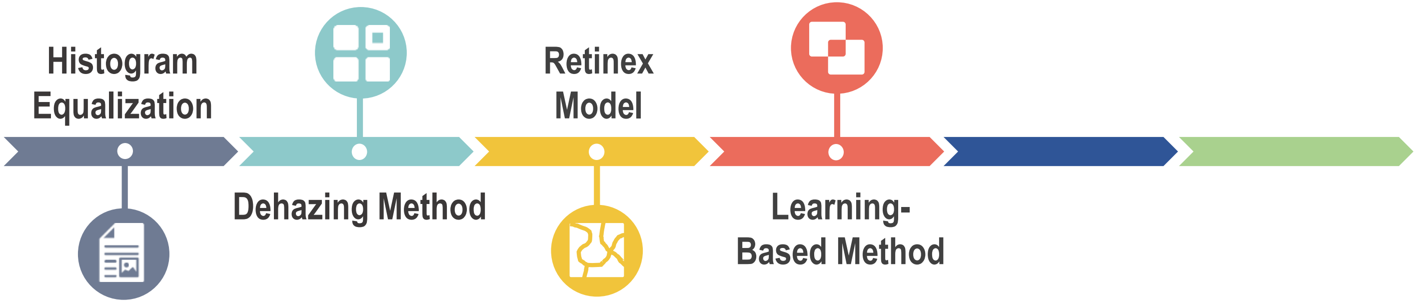
- Generate results

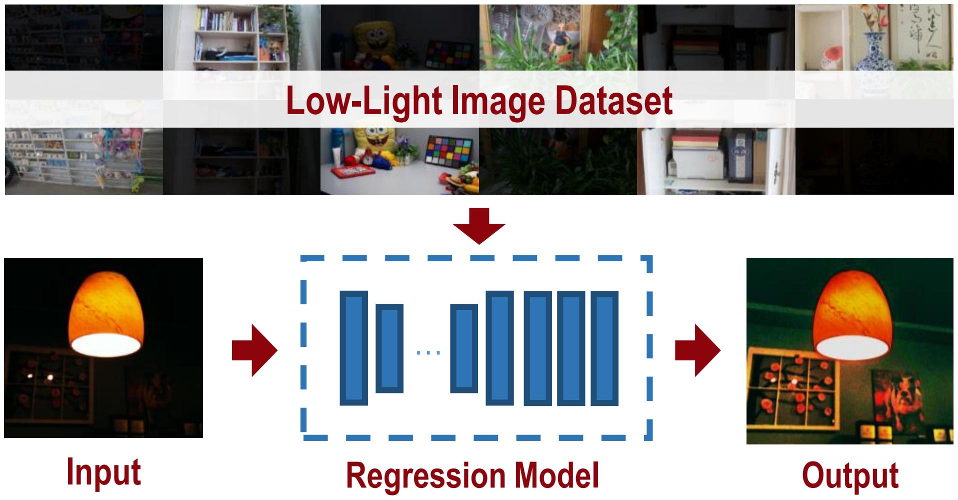
$$S_{\text{enhance}} = R \cdot L^{\frac{1}{\gamma}}$$



Xiaojie Guo, Yu Li, Haibin Ling. "LIME: Low-light image enhancement via illumination map estimation", *IEEE Trans. on image processing*, 2017.

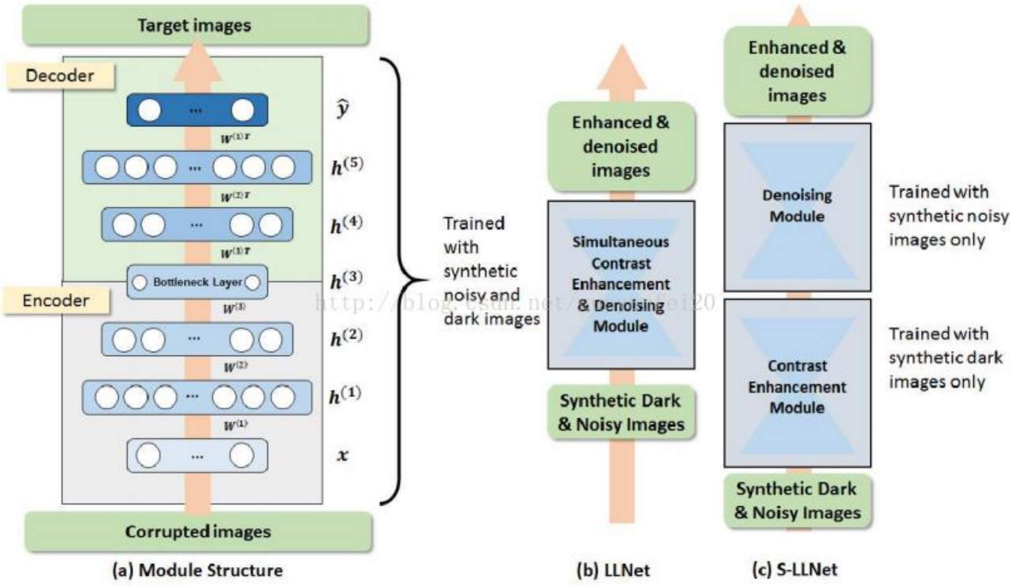
Representative Work





LLNet [PR17]

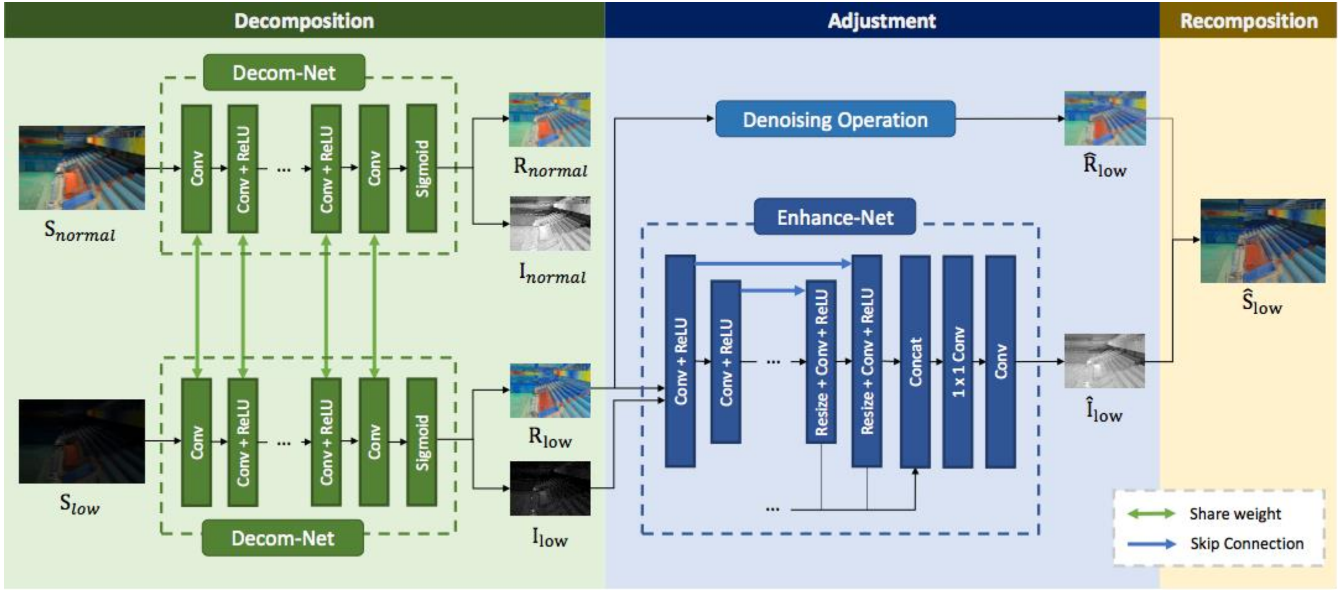
- Deep autoencoder



Kin Gwn Lore, Adedotun Akintayo, Soumik Sarkar, "LLNet: A Deep Autoencoder Approach to Natural Low-light Image Enhancement", *Pattern Recognition*, 2017.

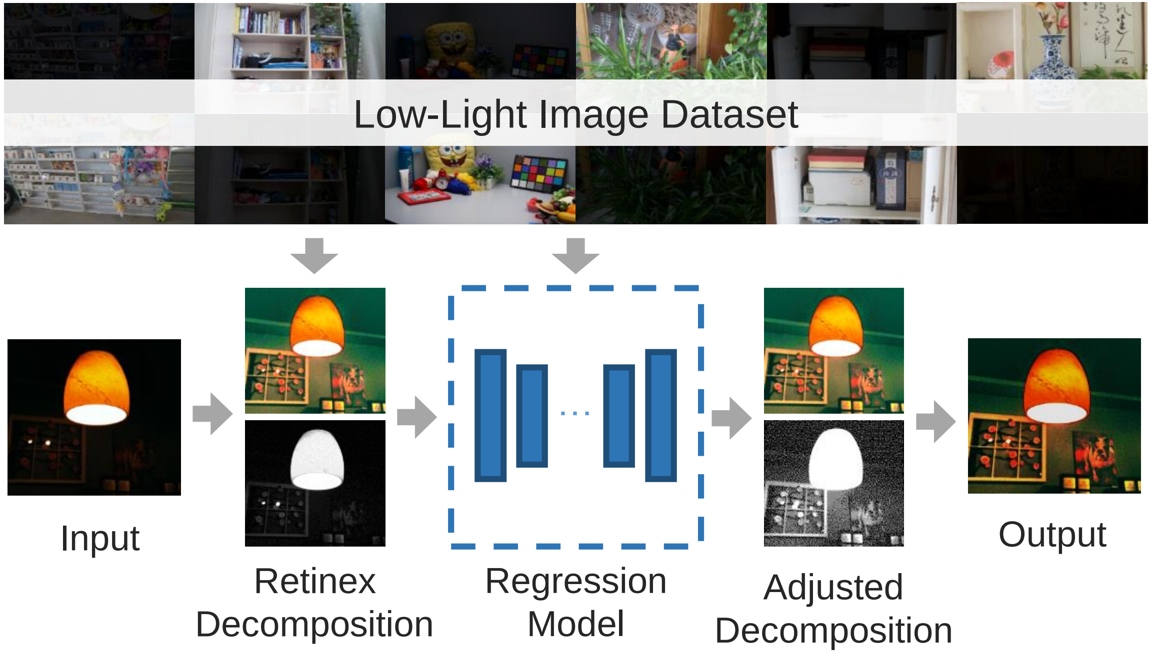
Deep Retinex Decomposition for Low-Light Enhancement

Chen Wei, Wenjing Wang, Wenhan Yang, Jiaying Liu **BMVC 2018**



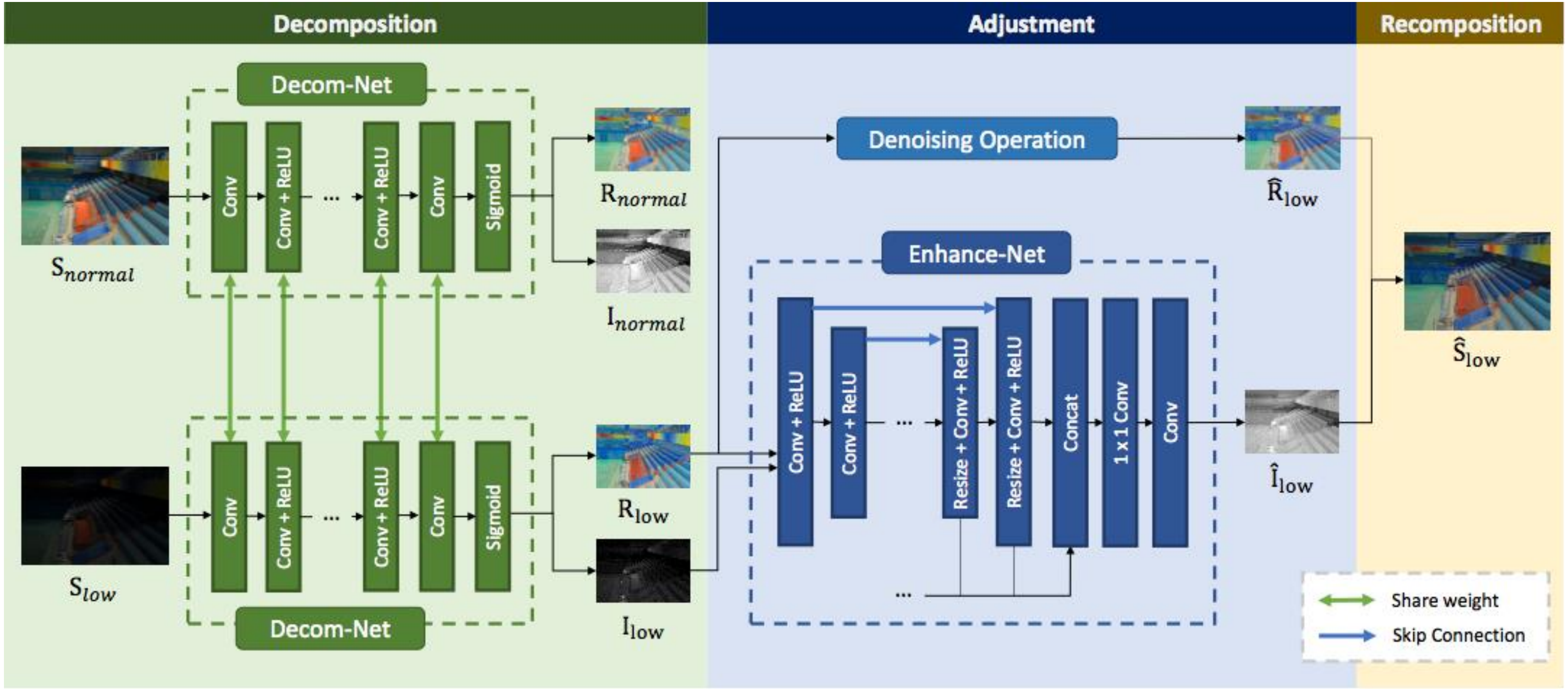
Deep Retinex Decomposition

- **Retinex Model + Deep Learning**



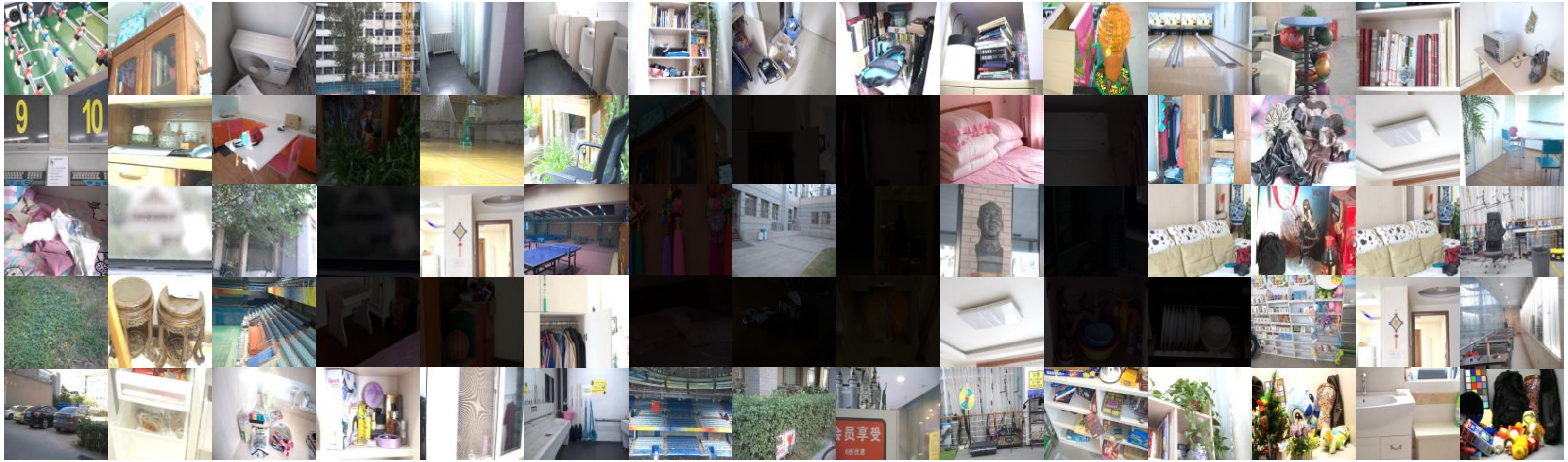
Deep Retinex Decomposition

- **Retinex Model + Deep Learning**



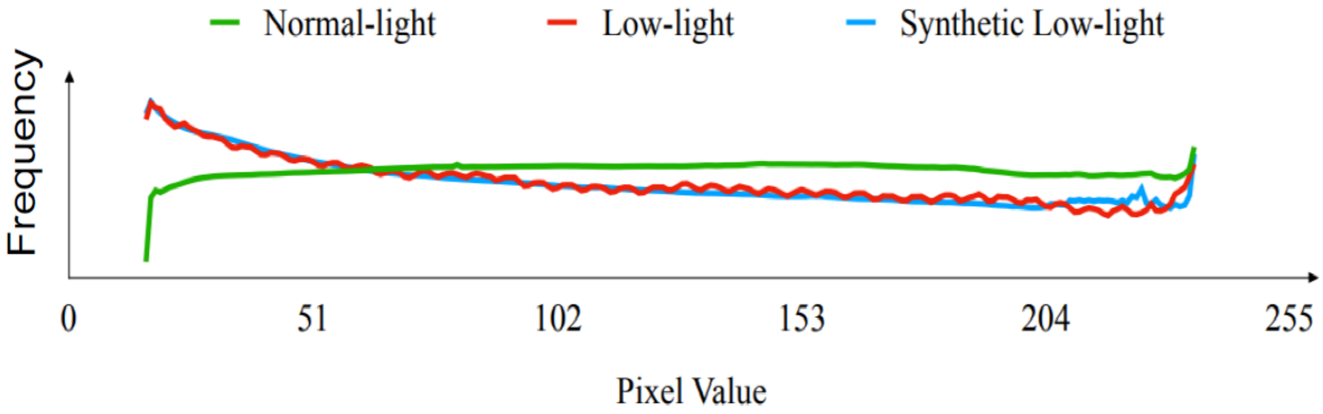
Deep Retinex Decomposition

- Real Photography Pairs: **LOw Light (LOL) Paired Dataset**
  - 1000 low/normal-light image pairs
  - 500 are collected by changing only exposure time and ISO
  - Various scenes, *e.g*. houses, clubs, streets.



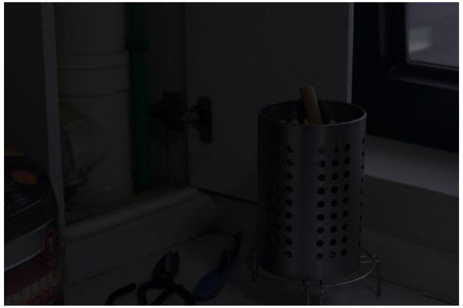
Deep Retinex Decomposition

- Synthetic Pairs from Raw Images
  - 1000 raw images from RAISE [Dang-Nguyen 2015]
  - Fitting the histogram of Y channel in YCbCr to real low-light images
  - Online available: https://daooshee.github.io/BMVC2018website/



Deep Retinex Decomposition

- **Decomposition**



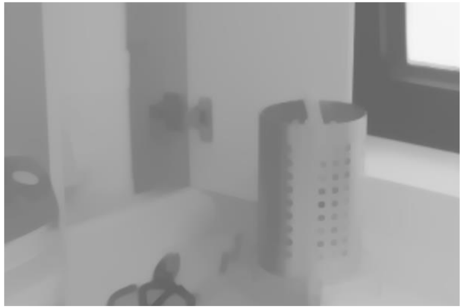
Low-Light Image

R by LIME

I by LIME

Normal-Light Image

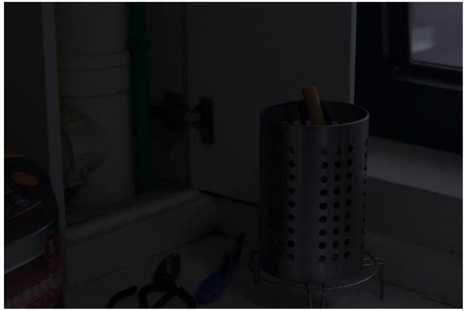
R by LIME



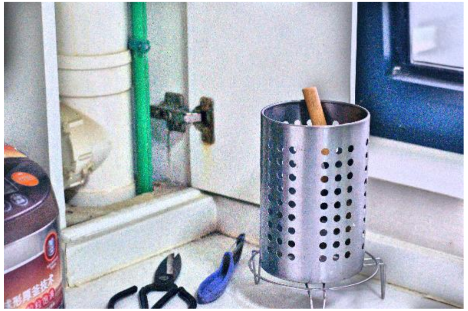
I by LIME

Deep Retinex Decomposition

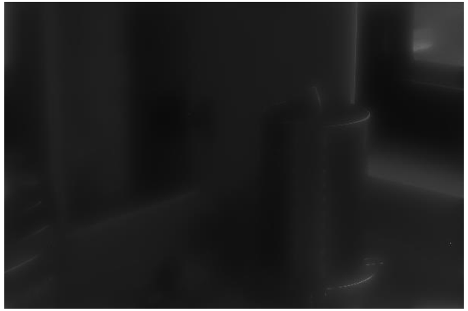
- **Decomposition**



Low-Light Image



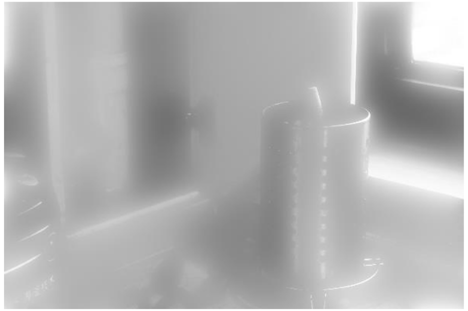
R by NPE



I by NPE

Normal-Light Image

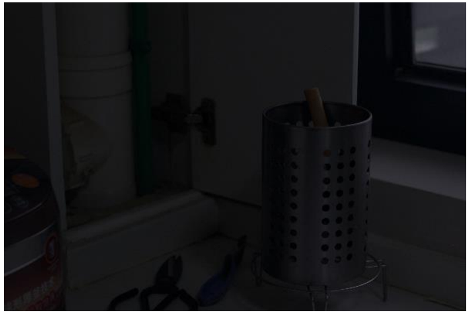
R by NPE



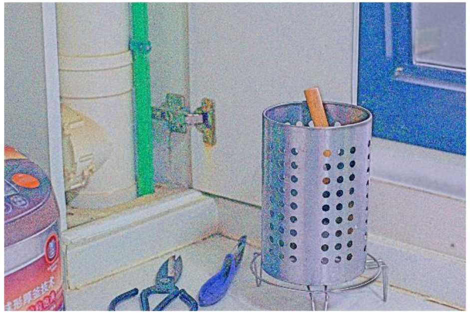
I by NPE

Deep Retinex Decomposition

- **Decomposition**



Low-Light Image



R by Retinex-Net

I by Retinex-Net

Normal-Light Image

R by Retinex-Net

I by Retinex-Net

Deep Retinex Decomposition

- **Experiments:**

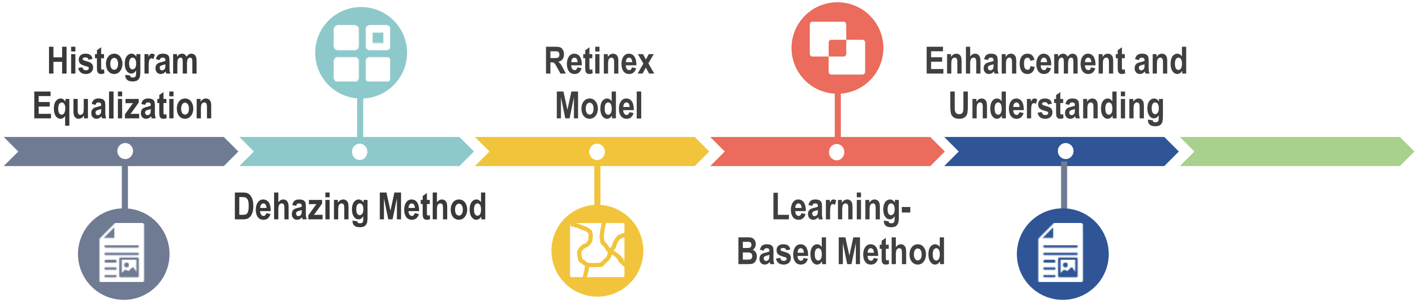
  Low-Light Enhancement

  - Visual Results

Low-Light Input DeHz NPE

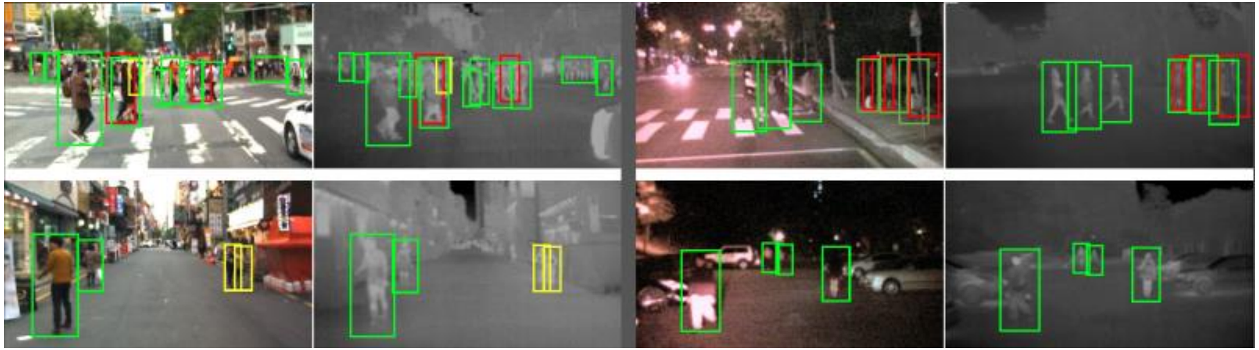
LIME SRIE Retinex-Net

Representative Work



Low-Light Datasets for High-Level Tasks

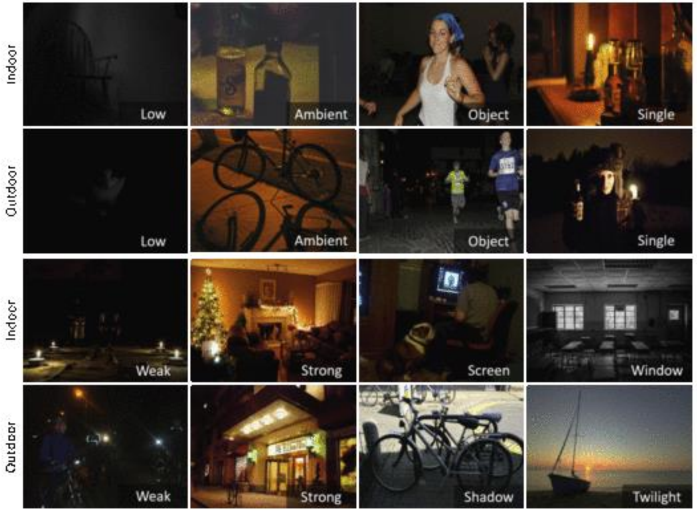
- **KAIST** [CVPR15]
  - Multispectral Pedestrian Detection
  - 44,871 night time annotations
  - 1,182 pedestrians in all (day and night)

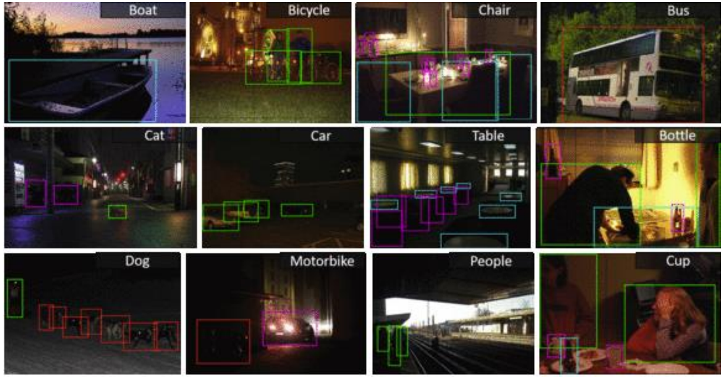


S. Hwang, J. Park, N. Kim, Y. Choi and I. S. Kweon, "Multispectral pedestrian detection: Benchmark dataset and baseline," *Proc. of Conference on Computer Vision and Pattern Recognition* (*CVPR*), 2015.

Low-Light Datasets for High-Level Tasks

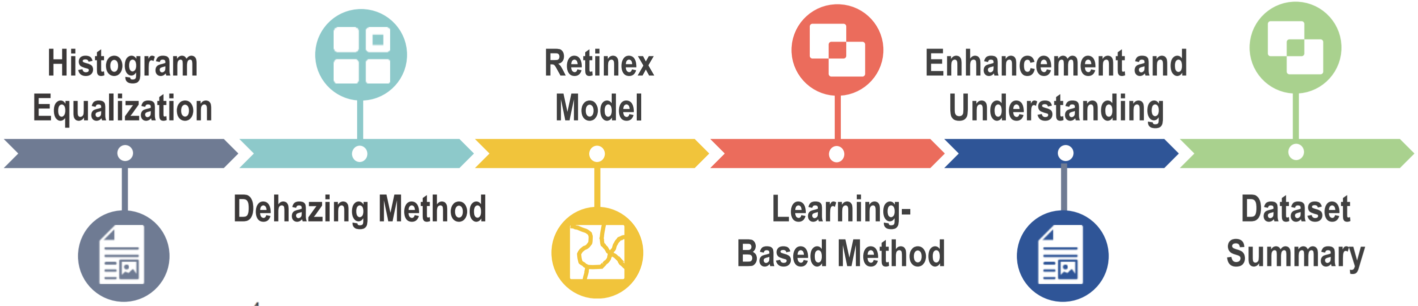
- **Exclusively Dark** [CVIU19]
  - 10 light conditions for Object Detection
  - 7,363 low-light images, 12 classes

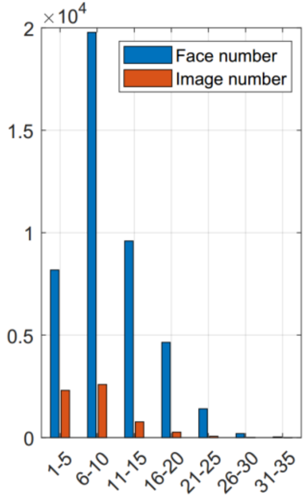


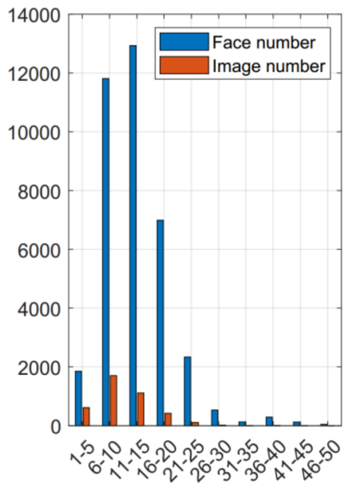


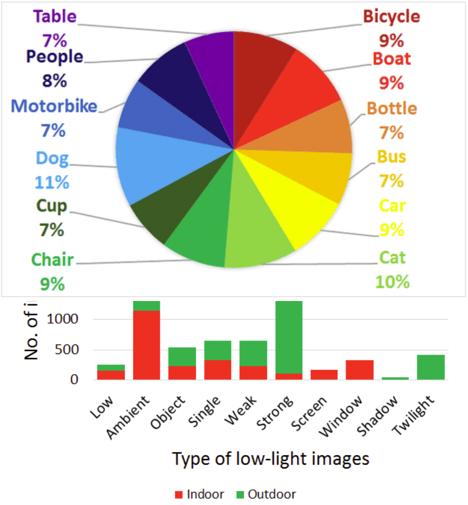
Yuen Peng Loh and Chee Seng Chan, "Getting to Know Low-light Images with The Exclusively Dark Dataset," *Computer Vision and Image Understanding* (*CVIU*), 2019.

Representative Work





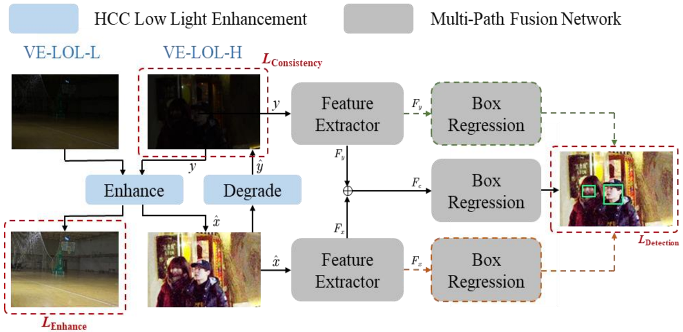




Our Dataset and Benchmark

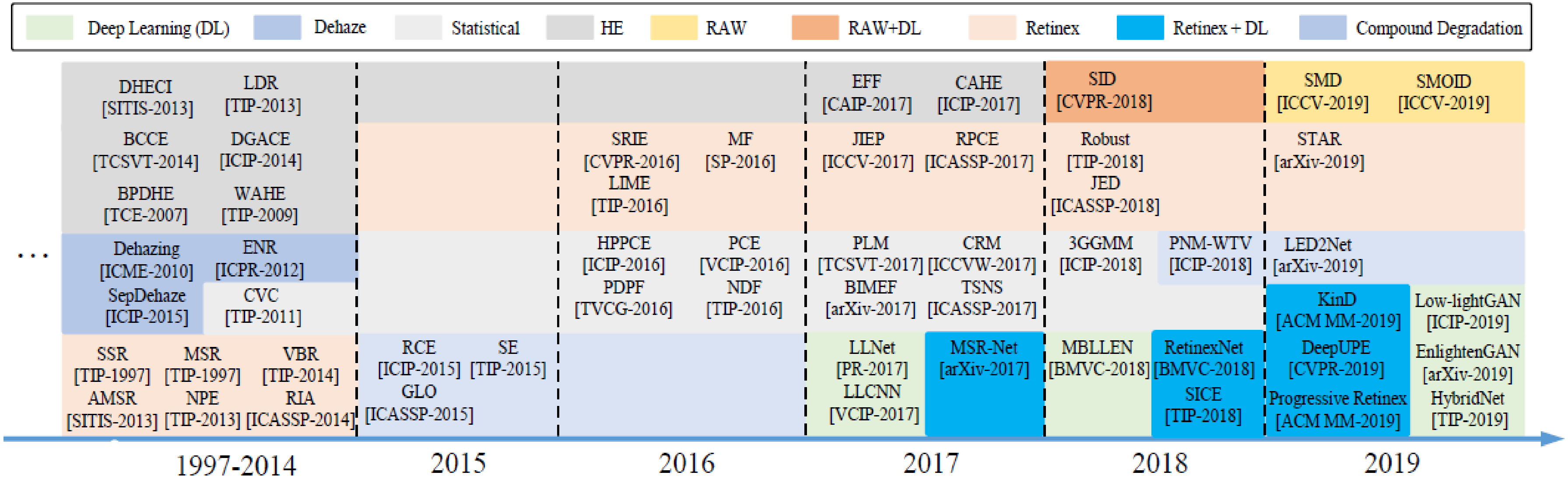
Benchmarking Low-Light Image Enhancement and Beyond

Jiaying Liu, Dejia Xu, Wenhan Yang, Minhao Fan, Haofeng Huang **IJCV 2021**



Our Dataset and Benchmark

- **Comprehensive Survey**
  - Chronological order and classification
  - Dehaze, Statistical, HE, RAW, Retinex, Compound degradation, Deep Learning (DL) based, RAW+DL, Retinex + DL



Our Dataset and Benchmark

- **Comprehensive Survey**

Table 3 An overview of low-light enhancement methods (Part 1): including histogram equalization (HE), dehaze, Retinex, statistical model-based methods

Category	Methods	Variables/models	Main idea	Publication
HE	CLAHE	Contrast; local histogram; partitioned regions	The method divides the input into regions and performs the adaptive histogram equalization locally with contrast limitation, which reduces noise by partially suppressing local histogram equalization	Pizer et al. (1990) [CVBC]
	BPDHE	Smoothed histogram; Histogram partition	The mean intensity of the output image is kept to be almost the same to that of the input to prevent visual deterioration	Ibrahim and Pik Kong (2007) [TCE]
	WAHE	Contrast adjustment; Noise robustness; White/black stretching; Mean-brightness preservation	A general framework based on histogram equalization is presented to integrate contrast adjustment, noise robustness, white/black stretching and mean-brightness preservation	Arici et al. (2009) [TIP]
	BCCE	Brightness compensation distortion; Backlight -scaled image contrast	The work formulates an objective function consisting of contrast enhancement and a newly proposed distortion model to adjust the backlight-scaled image contrast	Lee et al. (2014a) [TCSVT]
	LDR	2D histogram; Tree -like gray-level differences	The image contrast is enhanced by amplifying the gray-level differences between adjacent pixels	Lee et al. (2013a) [TIP]
	DHECI	Intensity histogram; Saturation histogram	A differential gray-levels histogram equalization is designed for color images with two differential gray-level histograms, i.e. intensity gray-levels histogram and saturation gray-levels histogram	Nakai et al. (2013) [SITIS]
	DGACE	Depth; 2D histograms; Adaptive space-variant transform function	A novel contrast enhancement method utilizes 2D histograms to transform pixel values adaptively based on the depth information	Lee et al. (2014b) [ICIP]
	EFF	Weighting matrix; Camera response model; Best exposure ratio	The weighting matrix and camera response model are introduced to synthesize multi-exposure images with the best exposure ratio	Ying et al. (2017b) [CAIP]
	CAHE	Visual importance; Dark-pass filtered gradients	The method adaptively controls the contrast gain based on the potential visual importance of intensities and pixels	Wu et al. (2017) [ICIP]
Dehaze	Dehazing	Inverted video	The method first inverts an input video and then applies a dehazing approach on the inverted video	Dong et al. (2011) [ICME]
	ENR	Inverted image; Filter weighting	After enhancement, the joint bilateral filter is introduced to suppress noise	Li et al. (2015) [ICPR]
	SepDehaze	Base layer; Detail layer; Superpixel	The input image is decomposed into base layer and detail layer and then enhance them adaptively	Zhang et al. (2012) [ICIP]
	SSR	Single-scale Retinex; Chromaticity coordinates; Color restoration function	It defines a practical implementation of Retinex center and surround Retinex, and treats the reflectance as the final enhanced result	Jobson et al. (1997b) [TIP]
	MSR	Multi-scale Retinex; Chromaticity coordinates; Color restoration function	It creates the enhanced results by fusing different single-scale Retinex outputs	Jobson et al. (1997a) [TIP]

Table 3 continued

Category	Methods	Variables/models	Main idea
Statistical model	CVC	2-D interpixel relationship histogram;	The work enhances the cont interpixel contextual infor mutual relationship betwe pixels
	HPPCE	Local contrast measure; Discrete total variation	A variational model is intro contrast measure, preserve inhibition. The control of
	PDPF	Multi-exposed results	The video frame is adjusted Guided by some visual pe exposed regions are integr manner
	PCE	Gradient map; Just noticeable difference	The textural coefficient is in just noticeable difference, the optimal contrast tone
	NDF	Nonlinear diffusion filtering; Texture suppression;	The illumination is estimate Surround suppression is e function to enhance the di the image
	PLM	Environmental light; Light-scattering attenuation	The initial environmental li surrounding function. The light-scattering attenuatio information loss constrain
	BIMEF	Weighting matrix; Camera response model; Best exposure ratio	The weighting matrix is ext Then, camera response m multi-exposure images an for each region
	CRM	Camera response model; Exposure ratio map	The method uses the inferre the pixel intensity to the d estimated exposure ratio
	TSNS	Noise level function; Just noticeable difference	The method first performs using noise level function difference model to suppr
	3GGMM	Generalized Gaussian mixture model	A three-component generali used to fit the histogram o probabilistically character overexposures

- **Our Dataset (VE-LOL)**
  - **Versatility:** Evaluation of **low/high-level** visions
  - **Authenticity:** Contain **real-captured** paired low/normal-light images
  - **Diversity:** Contain synthesized images with diversified backgrounds/objects
  - **Large-Scale:** VE-LOL-H (10,940 images) is comparable to WIDER-FACE (32,203 images) → Enables model training

Subset	#Image	Real/Synthetic	Paired	Annotations
VE-LOL-L-Syn	1,000	Synthetic	Yes	No
VE-LOL-L-Cap	1,500	Real	Yes	No
VE-LOL-H	10,940	Real	No	Yes

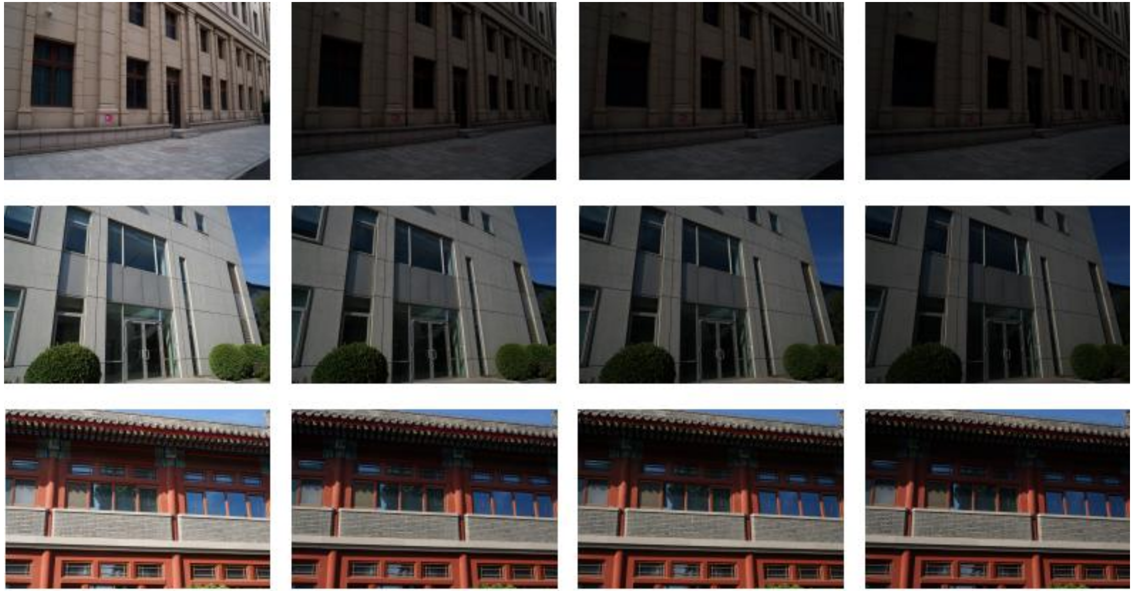
Our Dataset and Benchmark

- **Our Dataset (VE-LOL)**
  - Comparison against face detection datasets and detection datasets in degraded conditions

Dataset	#Image	#Object(Face)	#Train/Test	Conditions
ExDark[CVIU19]	7,363	23,710	4,800/2,563	Low Light
UFDD[Arxiv18]	6,424	10,895	0/6,424	Complex
MALF[FG15]	5,250	11,931	250/5,000	Normal
WIDER Face[CVPR16]	**32,303**	393,703	12,921/16,152	Normal
VE-LOL-H	**10,940**	83,885	6,940/4,000	Low Light

Our Dataset and Benchmark

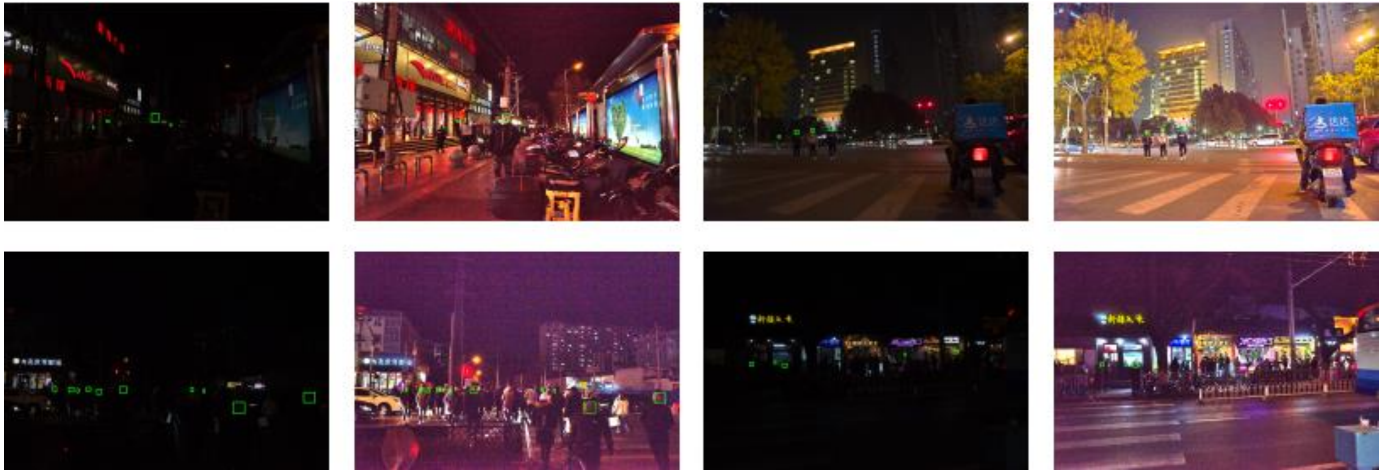
- **Our Dataset (VE-LOL)**
  - Example images of **VE-LOL-L-Cap**



- **Our Dataset (VE-LOL)**
  - Example images of **VE-LOL-L-Syn**

Our Dataset and Benchmark

- **Our Dataset (VE-LOL)**
  - Example images of **VE-LOL-H**



Low-Light | LIME[TIP17] | Low-Light | LIME[TIP17]

Our Dataset and Benchmark

- **Our Dataset (VE-LOL)**
  - Face detection (DSFD[CVPR19]) results on VE-LOL-H

Low-Light | LIME[TIP17] | Low-Light | LIME[TIP17]

Our Dataset and Benchmark

- **Our Dataset (VE-LOL)**
  - Face detection (DSFD[CVPR19]) results on enhanced VE-LOL-H

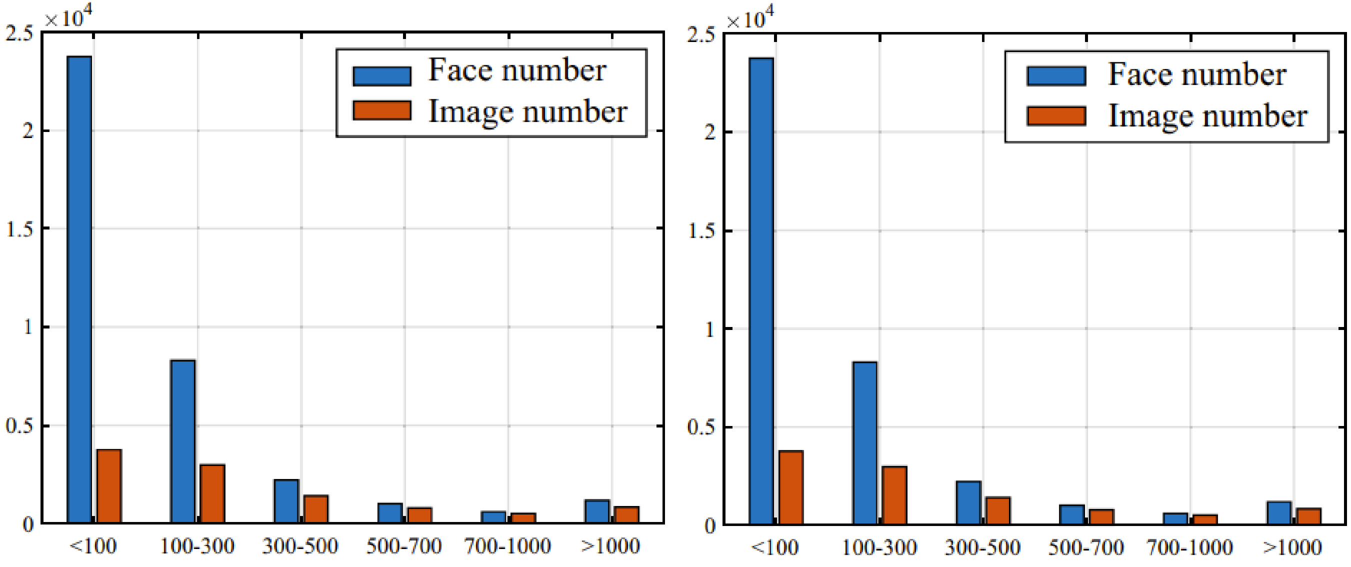
Low-Light LIME[TIP17] MSR[TIP97] MF[SP16]

Our Dataset and Benchmark

- **Our Dataset (VE-LOL)**
  - Diversity in scale, pose, occlusion, appearance and illumination

Our Dataset and Benchmark

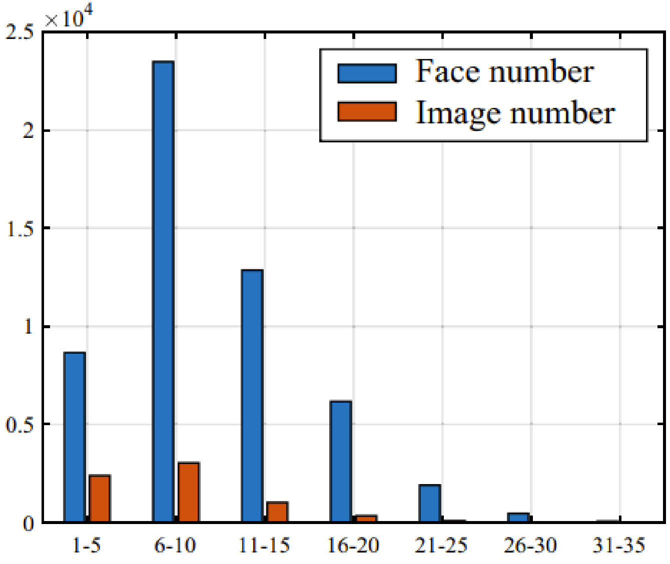
- **Our Dataset (VE-LOL)**
  - Face resolution (FR), face number (FN) distributions



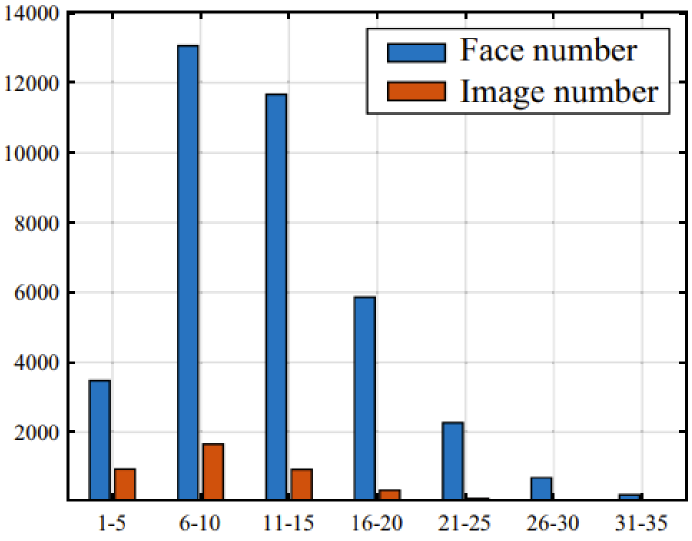
(a) FR in Train (b) FR in Test

Our Dataset and Benchmark

- **Our Dataset (VE-LOL)**
  - Face resolution (FR), face number (FN) distributions



(c) FN in Train



(d) FN in Test

Our Dataset and Benchmark

- **Benchmark** of low-light enhancement result
  - **Synthetic images**

Metrics	Larger	input	AMSR	MSR	Dehazing	NPE	LIME	MF	SRIE	BIMEF	BPDHE	LLNET
PSNR	Larger	10.24	11.79	11.95	15.38	15.38	14.07	16.26	13.66	15.95	12.75	17.57
SSIM	Larger	0.2941	0.4027	0.5493	0.5471	0.567	0.5274	0.5998	0.5469	0.6386	0.4651	0.7388
VIF	Larger	0.2937	0.2711	0.4525	0.3772	0.4502	0.4821	0.4378	0.4351	0.4377	0.3802	0.3347
Angular Error	Smaller	25.32	45.41	17.90	20.33	19.73	19.79	18.58	19.83	16.07	25.69	13.20
LOE	Smaller	0	1546.58	1245.56	200.62	445.46	889.51	186.53	140.11	142.41	18.11	452.5
NIQE	Smaller	24.62	82347.96	28.38	30.70	29.66	30.96	30.63	27.70	27.83	27.57	18.97
BRISQUE	Smaller	21.39	95.15	38.26	42.69	43.40	46.87	44.54	33.56	34.74	41.10	20.58
ENIQA	Smaller	0.1999	0.2191	0.2405	0.1703	0.2287	0.1748	0.1962	0.1951	0.1708	0.0600	0.2116
ILNIQE	Smaller	52.88	83.69	37.72	46.07	47.79	50.85	47.75	47.06	45.98	44.48	32.76
HOSA	Larger	37.18	56.53	54.70	44.95	47.12	47.01	47.58	38.31	42.80	40.09	38.18
SSEQ	Smaller	18.69	40.32	35.00	34.16	34.34	36.53	34.40	27.42	29.14	30.23	30.79
BLIINDS-II	Larger	44.52	65.55	282.30	133.15	315.17	217.35	251.26	372.72	31.38	64.63	133.55
Perceptual 1	Smaller	20522	23648	25595	15392	16877	28213	13695	13213	11781	18514	11138
Perceptual 4	Smaller	3482	4397	4124	3904	3618	4745	3308	3111	2972	3899	3161

Our Dataset and Benchmark

- **Benchmark** of low-light enhancement result
  - **Real images**

Metrics	Larger	JED	RetinexNet	CVC	DHECI	HE	LDR	Robust	SICE	WAHE	KinD	DeepUPE
PSNR	Larger	16.73	14.68	13.01	14.24	13.26	15.11	15.78	18.06	15.07	18.42	13.19
SSIM	Larger	0.6817	0.5252	0.4469	0.5312	0.5238	0.6114	0.6378	0.7094	0.6309	0.7658	0.4902
VIF	Larger	0.3744	0.3482	0.3501	0.4299	0.4388	0.4681	0.375	0.3747	0.4377	0.4381	0.4222
Angular Error	Smaller	13.02	21.32	28.83	19.58	17.53	19.33	16.06	12.42	17.08	11.67	22.7
LOE	Smaller	405.38	808.58	243.59	15.60	303.77	231.21	466.72	439.61	200.02	363.29	262.05
NIQE	Smaller	23.07	31.52	25.11	30.58	29.53	30.36	24.89	24.36	27.75	21.38	27.68
BRISQUE	Smaller	28.51	55.43	34.08	50.23	45.98	40.48	41.99	30.06	39.49	23.30	29.70
ENIQA	Smaller	0.1293	0.4049	0.0659	0.0699	0.2316	0.0941	0.1837	0.147	0.0549	0.1118	0.1906
ILNIQE	Smaller	35.53	47.27	36.08	51.63	18.32	36.42	46.32	33.85	36.13	29.01	48.99
HOSA	Larger	36.53	55.47	37.99	47.11	44.86	44.02	43.22	30.57	43.18	32.98	34.88
SSEQ	Smaller	18.39	38.88	23.41	37.29	35.38	24.39	26.58	26.36	22.29	23.19	25.45
BLIINDS-II	Larger	184.65	43.53	78.82	97.16	163.22	109.89	341.5	130.84	161.86	44.52	89.04
Perceptual 1	Smaller	11028	20333	26335	25581	25664	14901	13211	9871	13333	9735	14108
Perceptual 4	Smaller	2998	4341	4752	4410	4423	3290	3201	2838	3180	2434	3184

Our Dataset and Benchmark

- **Benchmark** of low-light enhancement result
  - **Running Time**

Method	MSR	Dehazing	BPDHE	NPE	LIME	MF	SRIE	BIMEF	JED	AMSR
Running Time (s)	1.4160	0.9574	0.7506	8.1812	1.2454	1.5136	6.7943	0.1761	1.9646	0.7592
Method	**LLNet**	**RetinexNet**	**CVC**	**DHECI**	**HE**	**LDR**	**Robust**	**SICE**	**WAHE**	**KinD**
Running Time (s)	4.0210	0.4690	1.2660	25.336	0.2166	0.3602	44.6750	0.8075	1.4023	3.0031

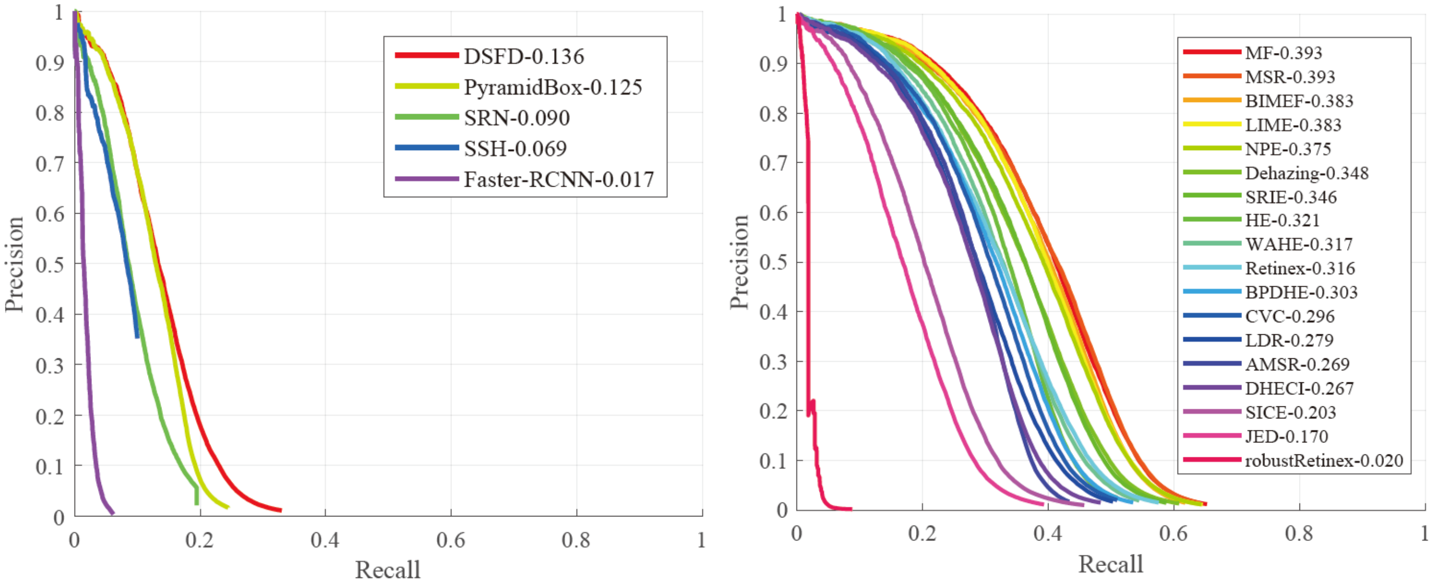
Our Dataset and Benchmark

- Example of enhanced results
  - Real image from **VE-LOL-L-Real**

(a) Input (b) BPDHE (c) BIMEF (d) LIME (m) SRIE (n) Robust
(e) GT (f) JED (g) WAHE (h) SICE (q) NPE (r) DHECI
(i) MF (j) MSR (k) LLNet (l) RetinexNet (s) KinD (t) DeepUPE

Our Dataset and Benchmark

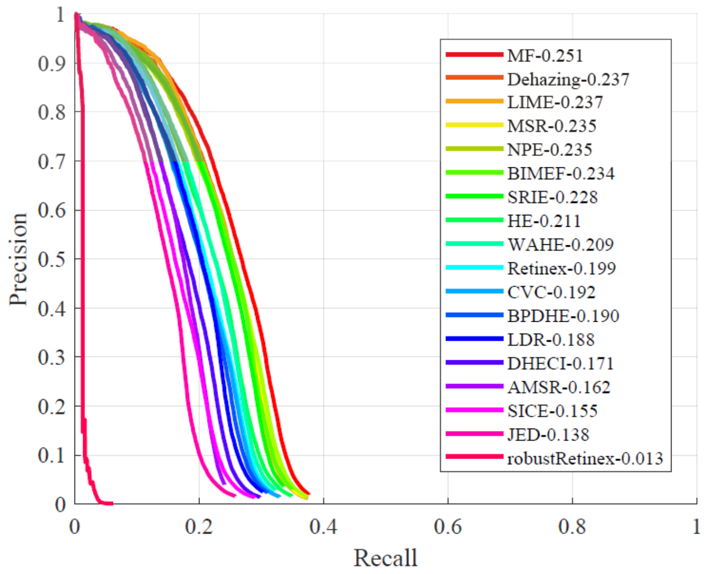
- **Evaluation results** of pretrained baseline on original and enhanced images of the proposed VE-LOL-H dataset



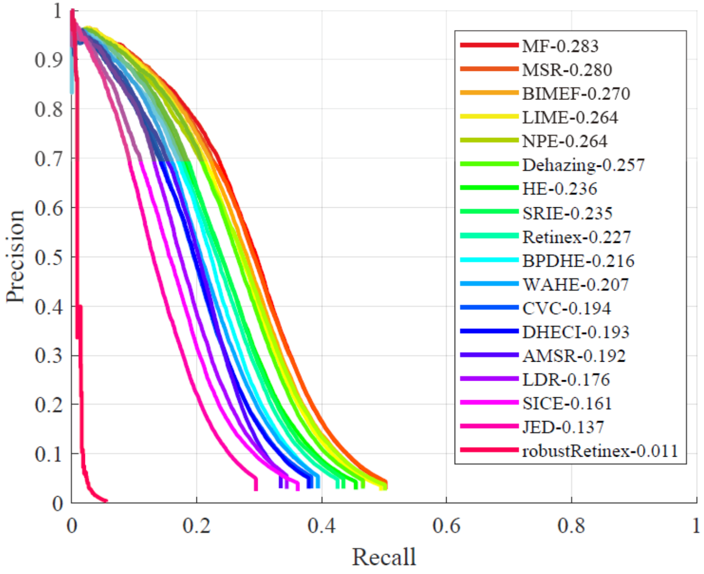
(a) Original (b) DSFD (Li et al., 2019)

Our Dataset and Benchmark

- **Evaluation results** of pretrained baseline on original and enhanced images of the proposed VE-LOL-H dataset



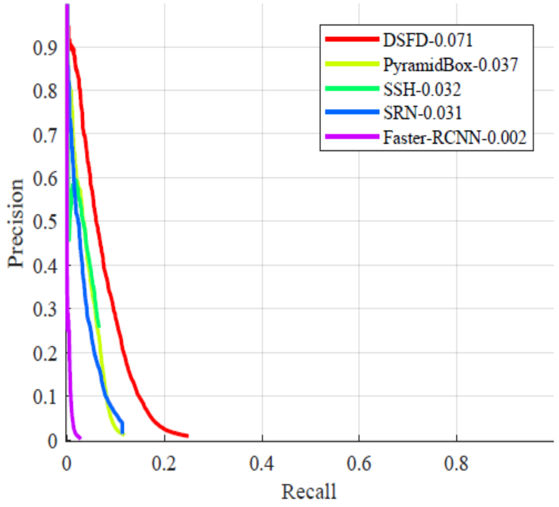
(c) PyramidBox (Tang et al., 2018)



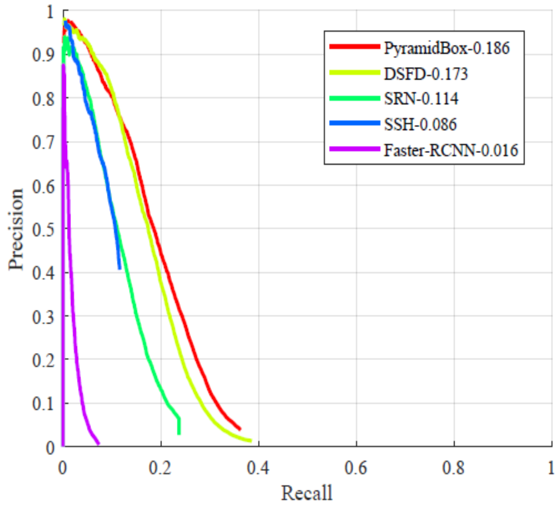
(d) SRN (Chi et al., 2018)

Our Dataset and Benchmark

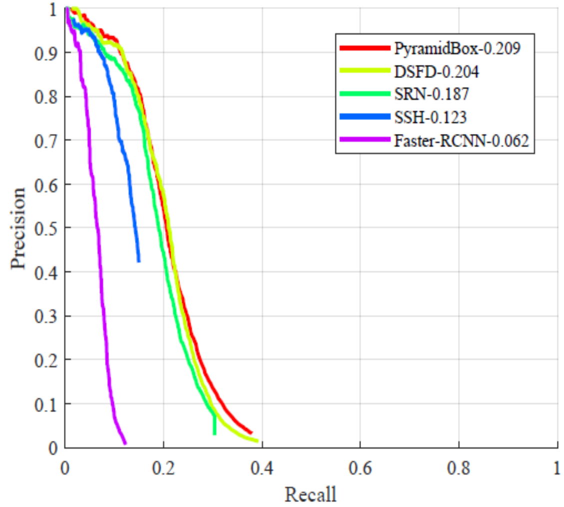
- Comparison of detection accuracies for different face scales



Small Face



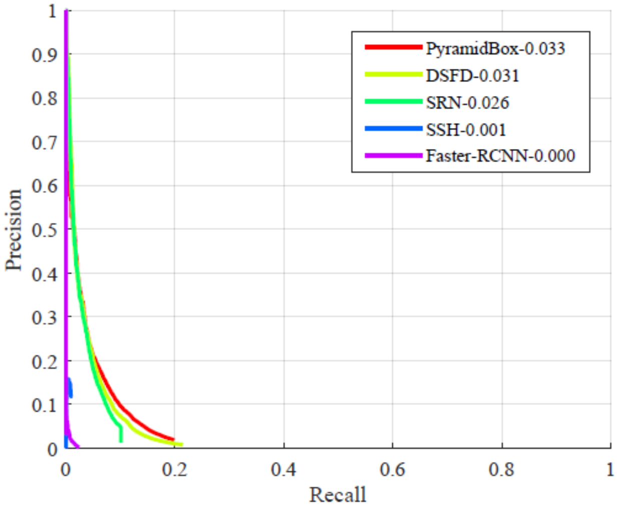
Medium Face



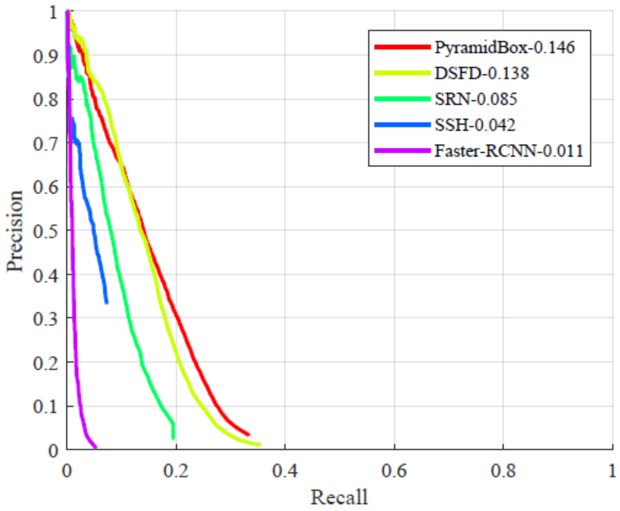
Large Face

Our Dataset and Benchmark

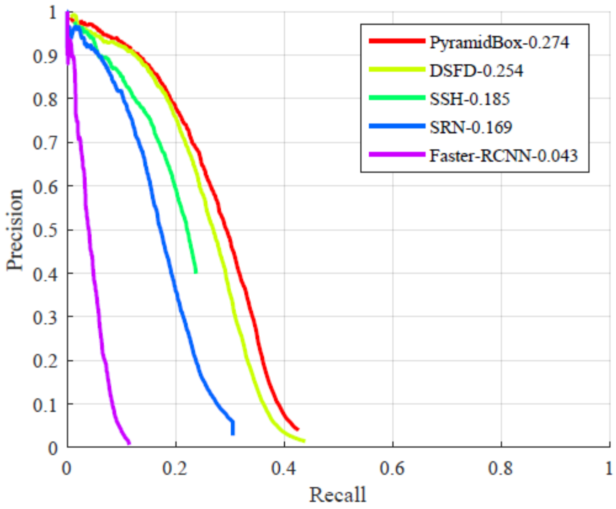
- Comparison of detection accuracies for different face scales



Low illumination of face



Medium illumination of face



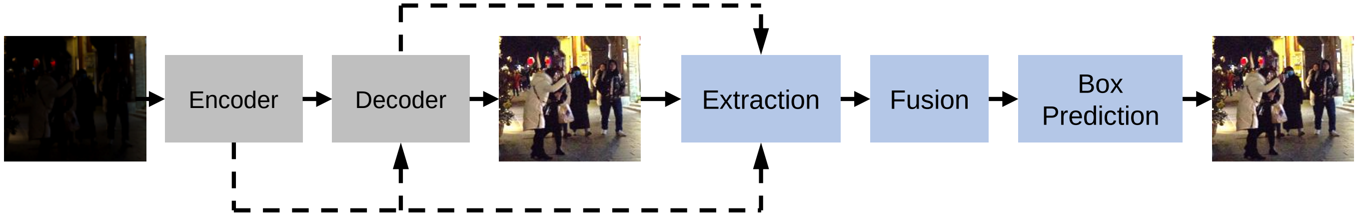
High illumination of face

Joint Enhancement and Face Detection

Enhancement and Detection Pipeline

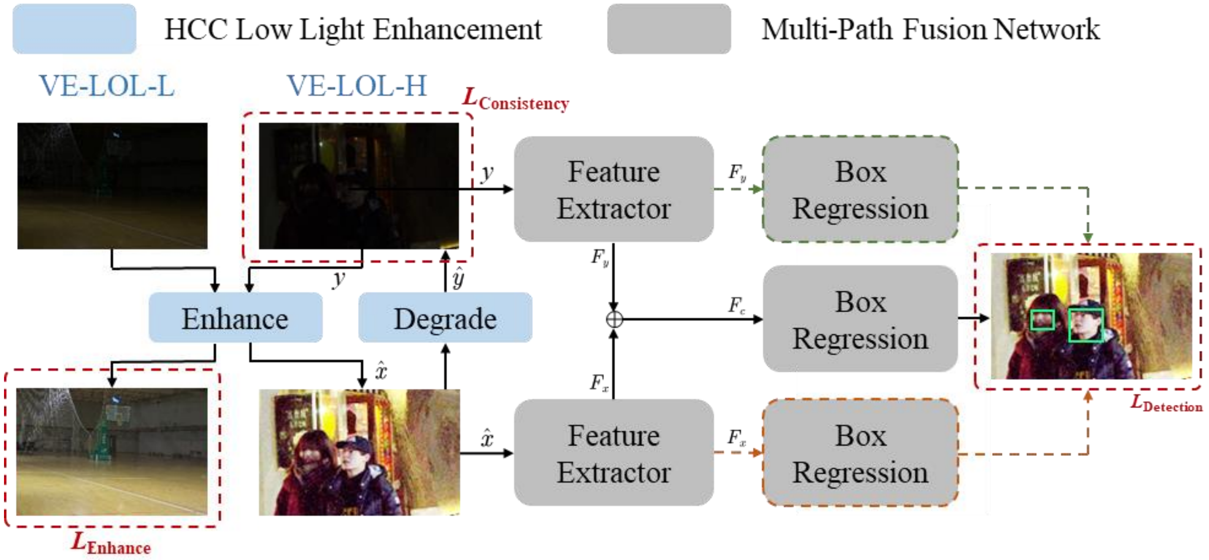
- IDEA 1: Prior Modeling + Multiple Exposure Fusion

- IDEA 2: Share Information between Two Stages



Joint Enhancement and Face Detection

- Prior Modeling + Multiple Exposure Fusion
  - Half Cyclic Constrained (HCC) Enhancement
  - Multi-Path Fusion Network



Joint Enhancement and Face Detection

- Prior Modeling + Multiple Exposure Fusion
    - Enhancement

$$L_{\text{Enhance}} = \gamma\left(\|\hat{x} - x\| - \alpha \text{SSIM}\left(\hat{x}, x\right)\right) + L_{\text{Adv}}(\hat{x}, x)$$

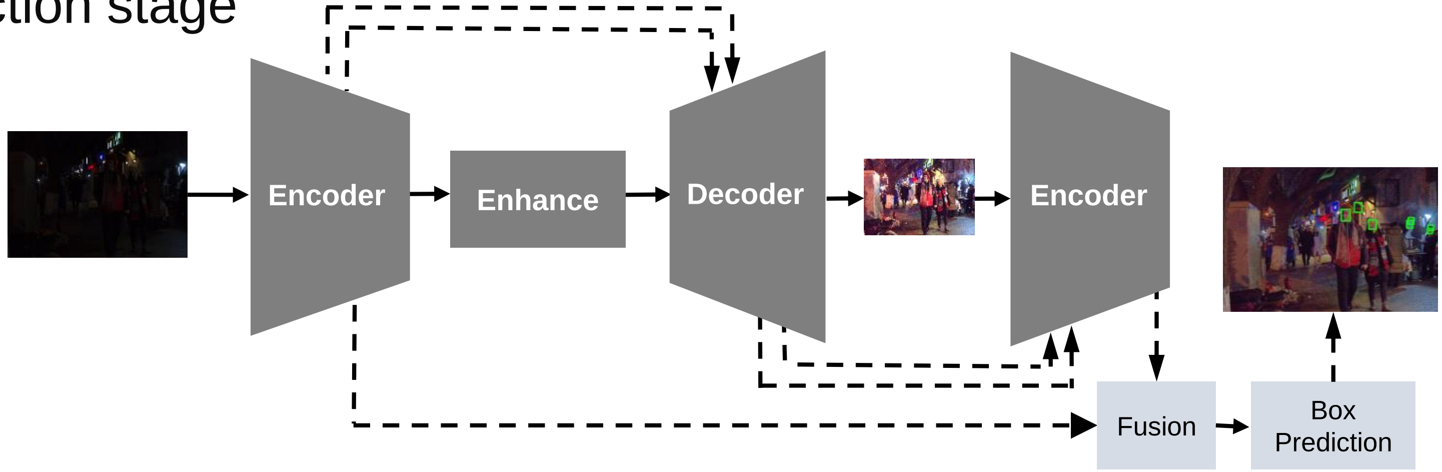
    - Cycle Consistency with a learned degradation model

$$L_{\text{Consistency}} = \|\hat{y} - y\|$$

    - Multi-Path Fusion: concat features and feed them into a box regression network

Joint Enhancement and Face Detection

- Share Information between Two Stages
  - FishNet aggregates multi-scale context information
  - Skip Connections: use features from enhancement stage to guide detection stage



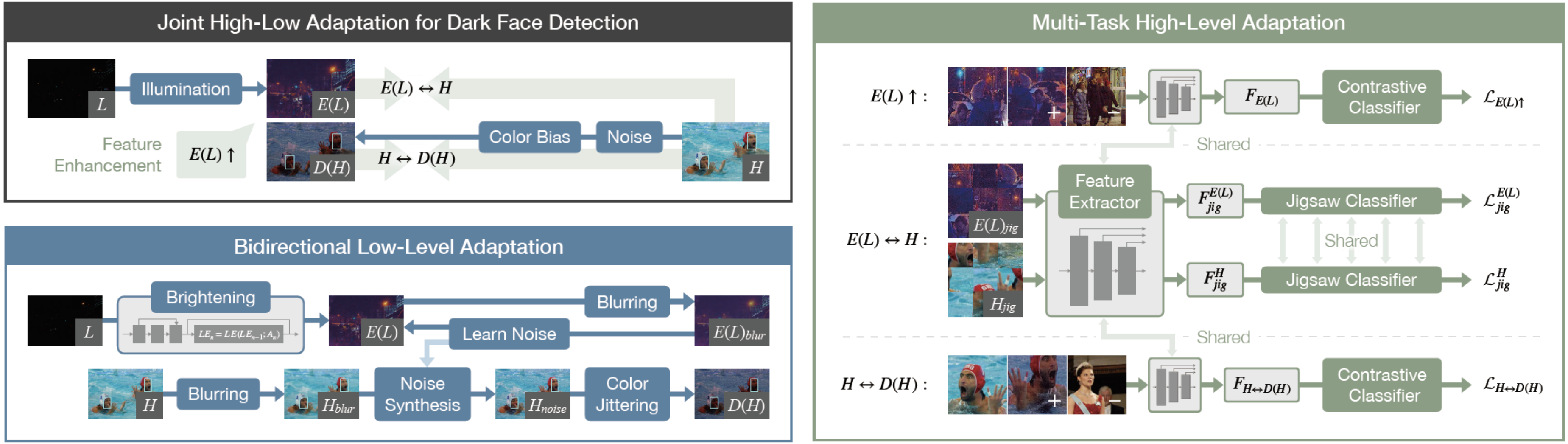
Experiments

- The mAP scores of different methods

Method	mean Average Precision
Pretrained DSFD	13.6
Finetuned DSFD	44.3
MF + Pretrained DSFD	39.3
MF + Finetuned DSFD	46.8
Proposed w/o Multiple Detection Loss	48.0
Proposed	48.9

HLA-Face: Joint High-Low Adaptation for Low Light Face Detection

Wenjing Wang, Wenhan Yang, Jiaying Liu **CVPR 2021**



Unsupervised Low-Light Face Detection

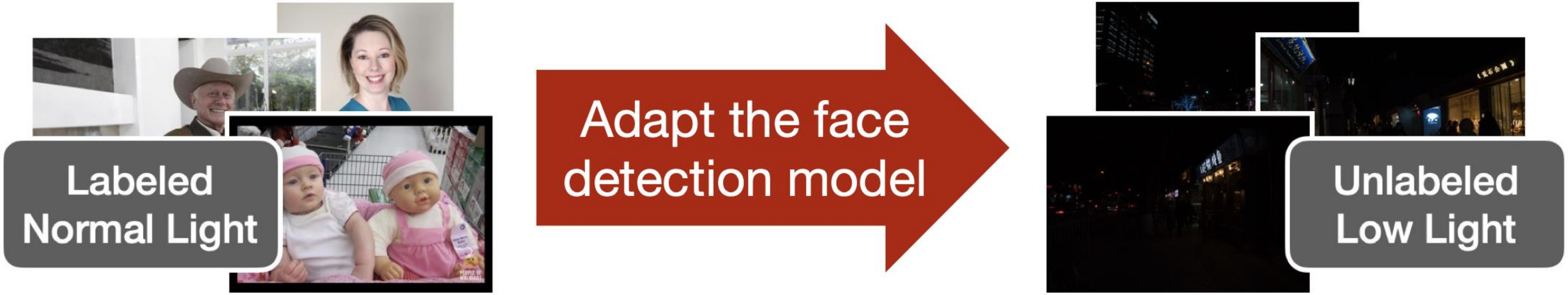
Face detection **under low light circumstance**

- **Naive solution:**
  - Construct a low light face detection dataset & train a corresponding new model
- **Drawbacks:**
  - Cost of human and financial resources
  - Poor robustness and scalability

Unsupervised Low-Light Face Detection

Face detection **under low light circumstance**

Our Method: Adapt the model from normal light to low light

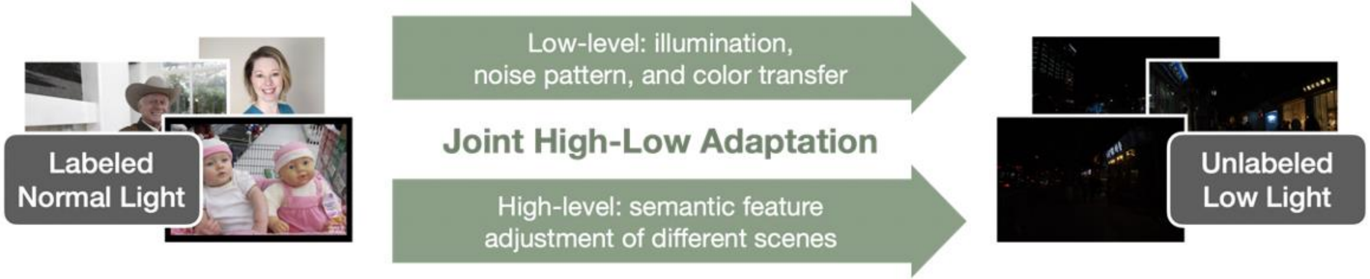


Unsupervised Low-Light Face Detection

- **Gaps between normal light and low light**
  - Pixel-level appearances **(Low-level gap)**
    - E.g. illumination, noise pattern, and color cast
  - Object-level semantics **(High-level gap)**
    - E.g. the existence of street lights, vehicle headlights, and advertisement boards

Unsupervised Low-Light Face Detection

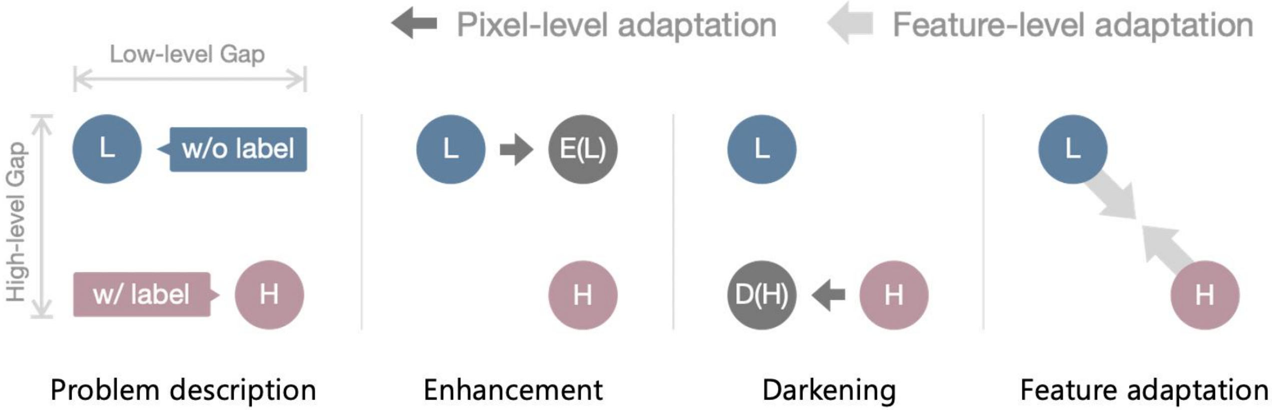
- **Our solution:** Joint low-level and high-level adaptation



DSFD LIME + DSFD Our HLA-Face

Unsupervised Low-Light Face Detection

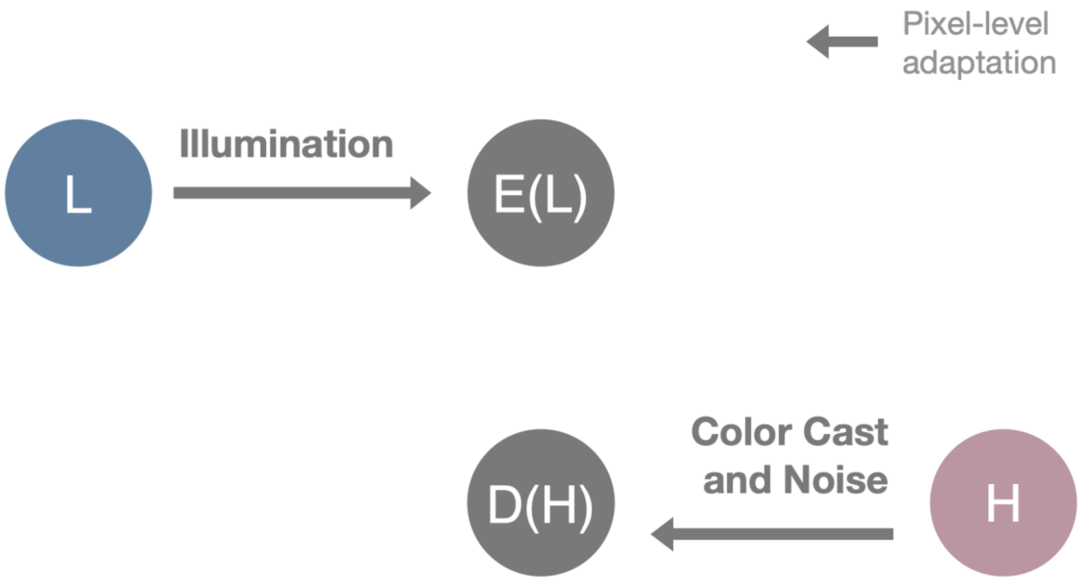
- Reviewing adaptive low light detection techniques



- Enhancement and darkening only consider the pixel-level gap
- Feature adaptation methods try to fill the whole gap in one step

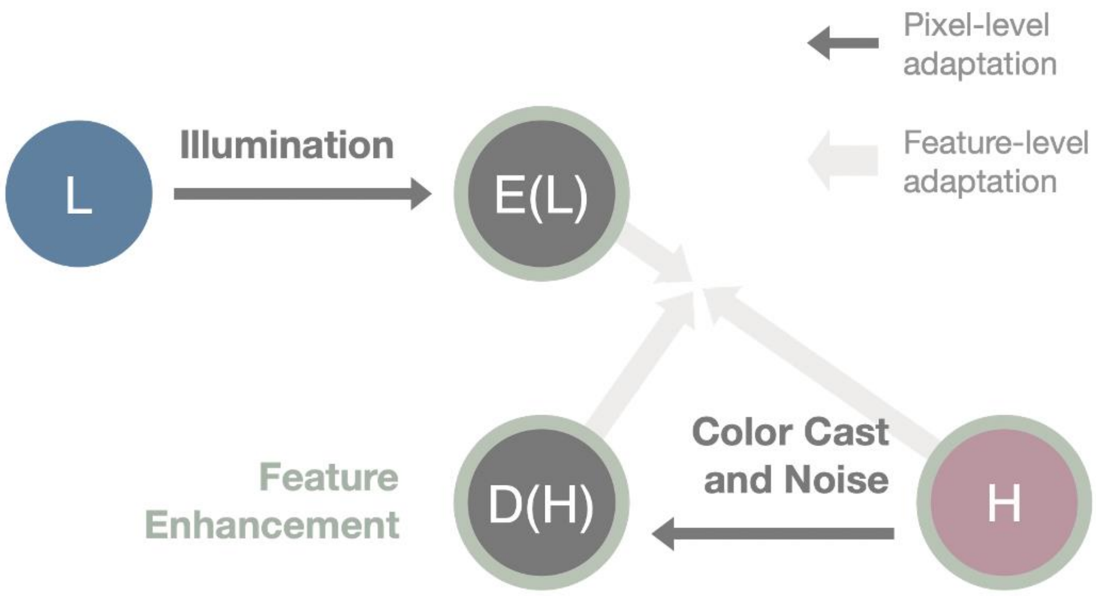
Unsupervised Low-Light Face Detection

- We instead consider joint low-level and high-level adaptation



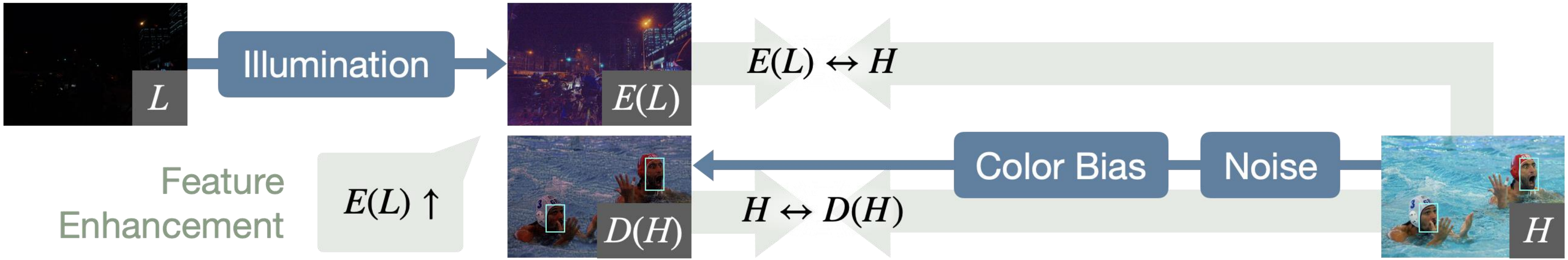
Unsupervised Low-Light Face Detection

- We instead consider joint low-level and high-level adaptation



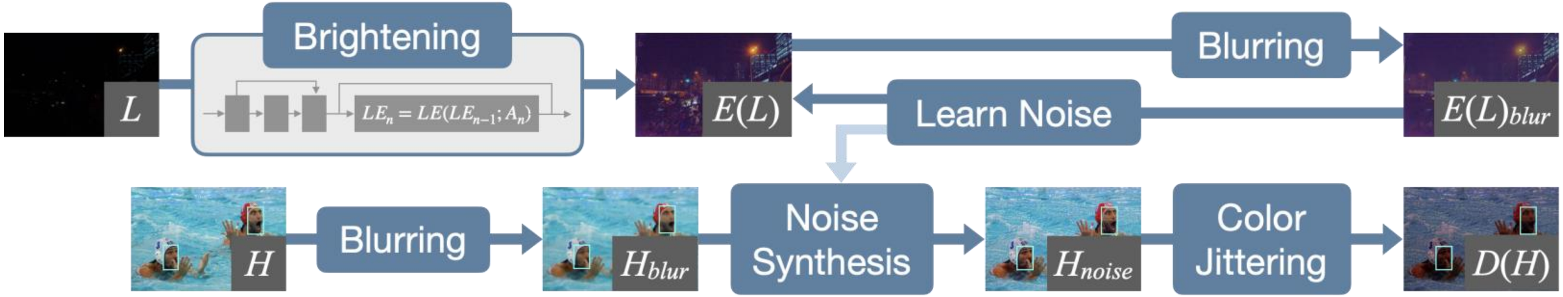
Unsupervised Low-Light Face Detection

- **Overall network architecture**



Unsupervised Low-Light Face Detection

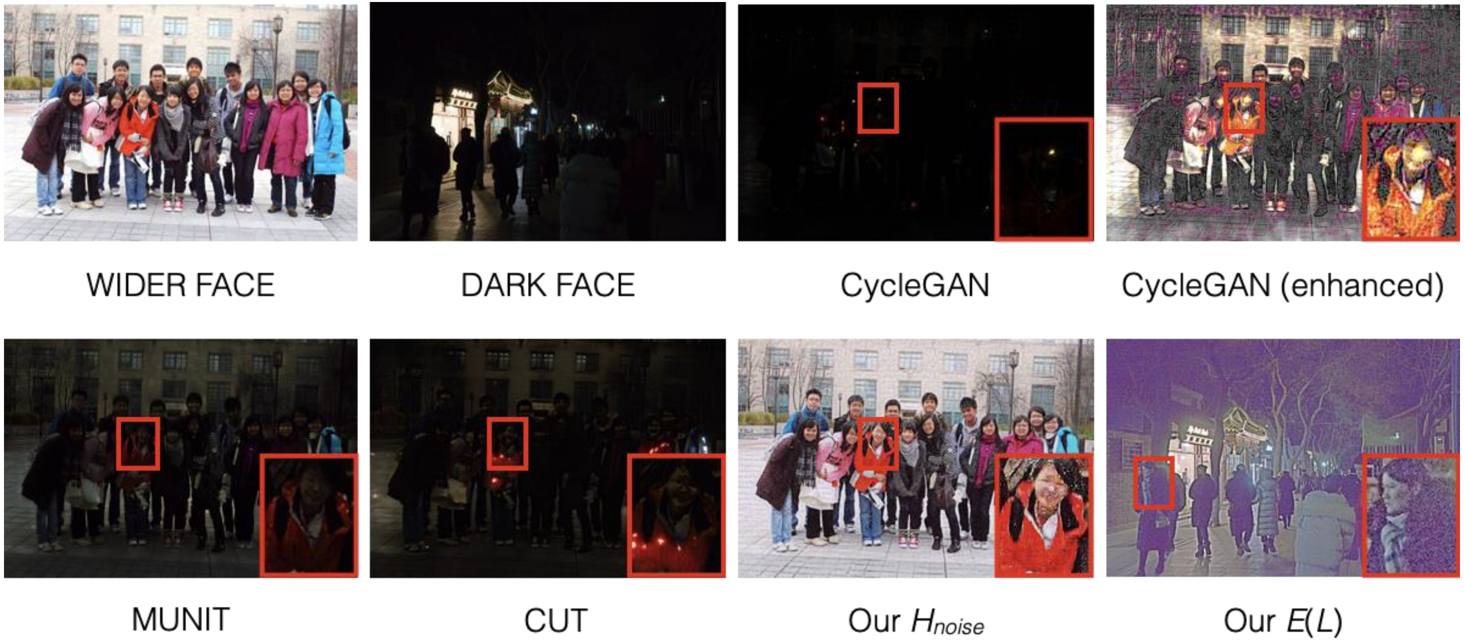
- **Bidirectional low-level adaptation**



- Brightening: nonlinear curve mapping
- Noise Synthesis: supervised adversarial learning
- Color Jittering

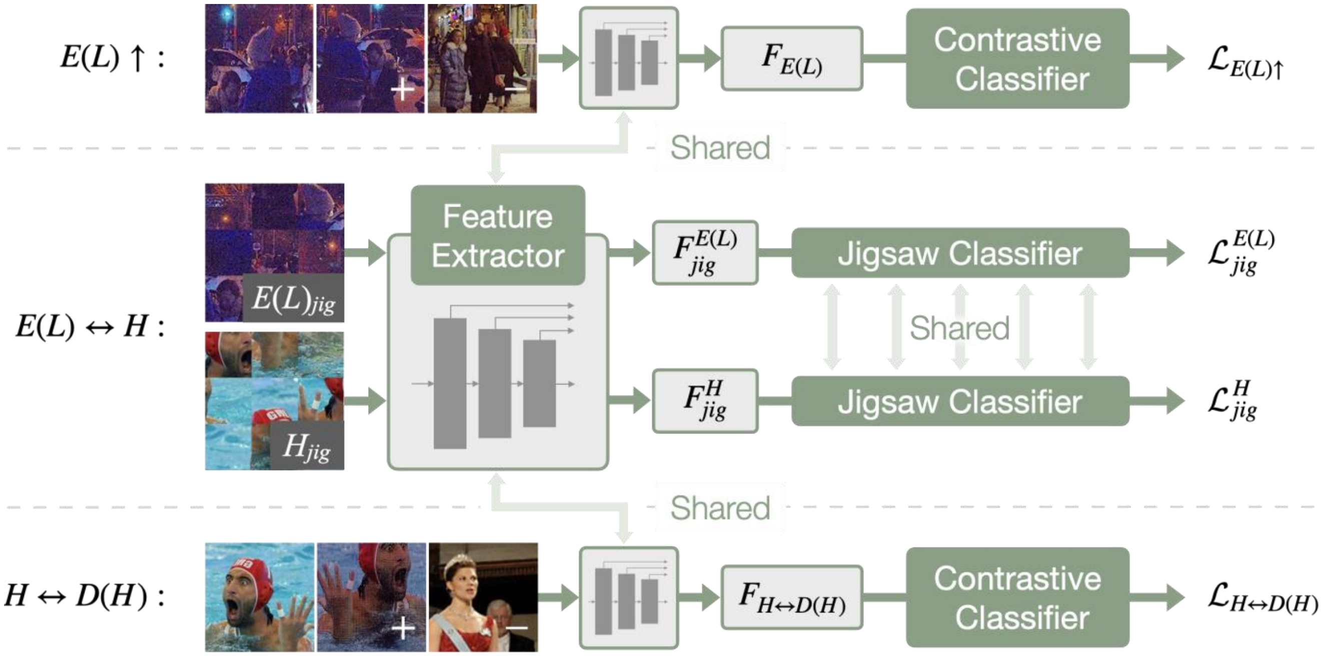
Unsupervised Low-Light Face Detection

- **Bidirectional low-level adaptation**
  - Comparison results of pixel-level transferring



Unsupervised Low-Light Face Detection

- **Multi-task high-level adaptation**



Unsupervised Low-Light Face Detection

- Benchmarking state-of-the-art methods

Category	Method	mAP (%)
Face Detection	Faster-RCNN	1.7
	SSH	6.9
	RetinaFace	8.6
	SRN	9.0
	SFA	9.3
	PyramidBox	12.5
	Small Hard Face	**16.1**
	DSFD	**16.1**

Unsupervised Low-Light Face Detection

- Benchmarking state-of-the-art methods

Category	Method	mAP (%)
Enhancement (with Small Hard Face)	Zero-DCE	37.7
	MF	38.3

Category	Method	mAP (%)
Enhancement (with DSFD)	EnlightenGAN †	20.8
	Zero-DCE †	37.3

† denotes retrained with DARKFACE

Category	Method	mAP (%)
Enhancement (with DSFD)	SICE	4.7
	RetinexNet	12.0
	KinD	15.8
	EnlightenGAN	31.3
	LIME	40.7
	Zero-DCE	41.3
	MF	**41.4**

Unsupervised Low-Light Face Detection

- Benchmarking state-of-the-art methods

Category	Method	mAP (%)
Darkening (with DSFD)	MUNIT	29.7
	CycleGAN	31.9
	CUT	**32.7**

Category	Method	mAP (%)
Unsupervised DA (with DSFD)	OSHOT	25.4
	Progressive DA	28.5
	Pseudo Labeling	**35.1**

Unsupervised Low-Light Face Detection

- Best results from each category

Category	Method	mAP (%)
Face Detection	DSFD	16.1
Enhancement	DSFD + MF	41.4
Darkening	DSFD + CUT	32.7
Unsupervised DA	DSFD + Pseudo Labeling	35.1

- Our performance

Category	Method	mAP (%)
	Ours	**44.4**
Fully Supervised	Fine-tuned DSFD	46.0

- **Subjective Results**

(a) Ground truth (b) SICE (c) RetinexNet (d) KinD (e) Ours

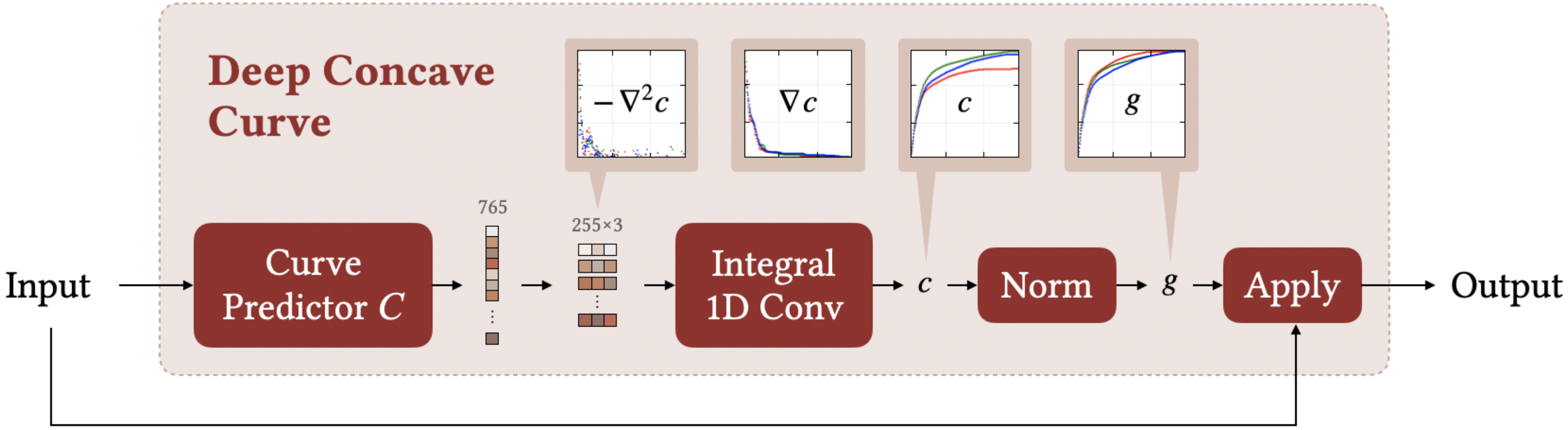
Unsupervised Low-Light Face Detection

- **Applications**
  - Improve fully supervised model
    - DSFD with labels and our adaptation, mAP 0.460 → 0.486
  - Transfer from COCO to ExDark

	AP (%)	AP 50 (%)	AP 75 (%)
original	29.261	59.784	24.538
w/ Ours	**30.056**	**60.746**	**25.835**

Self-Aligned Concave Curve: Illumination Enhancement for Unsupervised Adaptation

Wenjing Wang, Zhengbo Xu, Haofeng Huang, Jiaying Liu **ACM MM 2022**



Aim and Challenge

- **Existing works:**

Categories	**Illumination Adjustment**	**High-Level Vision**
Low-light enhancement *e.g.* Zero-DCE (CVPR-20), RUAS (CVPR-21)	√	×
Domain adaptation *e.g.* HLA-Face (CVPR-21), CIConv (ICCV-21)	×	√
Our Target	√	√

Our Method: **First** to propose a learnable **pure** illumination enhancement model for high-level vision

Aim and Challenge

- **Challenge:** How to restore underexposed images/videos from the perspective of machine vision?

Our approach consists of two aspects:

- **Network:** an illumination enhancement model which can maximize the model's abilities while being easy to learn
- **Training Strategy:** guide the model to adjust illumination from the perspective of machine vision

Method

- **Network:** Deep Concave Curve

Idea: find a function $\boldsymbol{g}$ and use it to enhance the low-light input I_L

$$\hat{I}_L = g(I_L)$$

We assume that $g(\cdot)$ should:

- Pass (0,0) and (1,1)
- Increase monotonically
- Be spatially shared
- Be concave

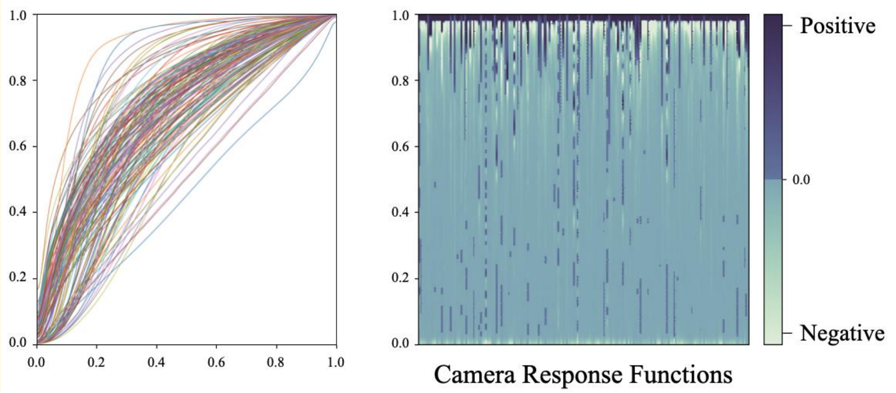
Ignoring spatial variations like lens fall-off and signal-dependent noise

Method

- **Network:** Deep Concave Curve

Idea: find a function $\boldsymbol{g}$ and use it to enhance the low-light input I_L

Most camera response functions are concave



Left figure: real camera CRFs from the DoRF dataset.

Right figure: the heat map of second-order derivatives in DoRF.

We assume that $g(\cdot)$ should:

- Pass (0,0) and (1,1)
- Increase monotonically
- Be spatially shared
- Be concave

- **Network:** Deep Concave Curve

Idea: find a function g and use it to enhance the low-light input I_L

$$\hat{I}_L = g(I_L)$$

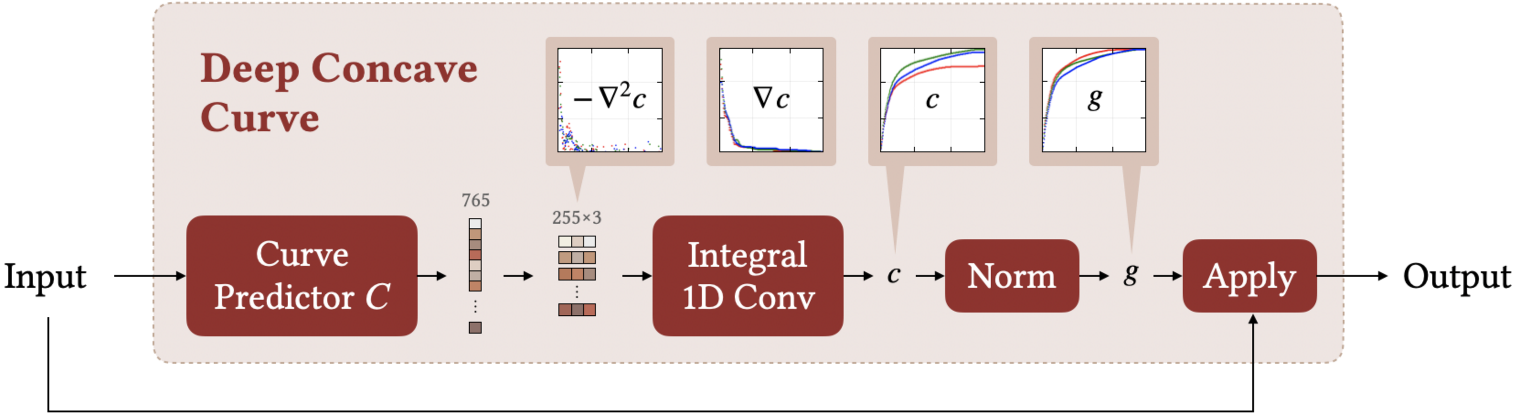
We assume that $g(\cdot)$ should:

- Pass (0,0) and (1,1)
- Increase monotonically
- Be spatially shared
- Be concave

g should be a Concave Curve

How to design a neural network that can follow these constraints?

- **Network:** Deep Concave Curve



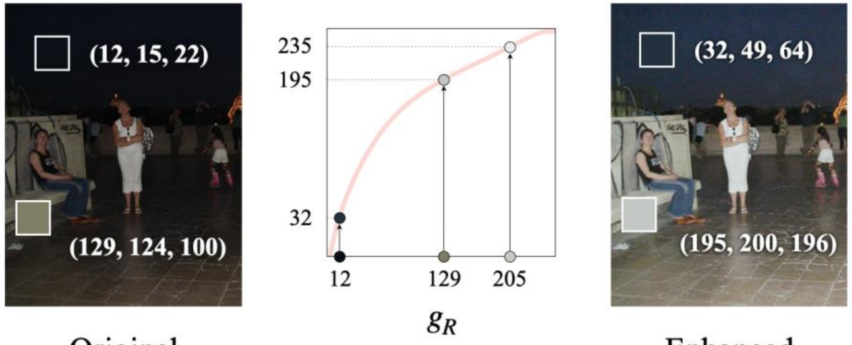
Step 1. Predicts a non-negative minus second derivative $-\nabla^2 c$

Step 2. Integrates and normalizes $-\nabla^2 c$ into a concave curve g

- **Network:** Deep Concave Curve

Idea: find a function g and use it to enhance the low-light input I_L

Let g be a mapping from original pixel values to new pixel values



Original Enhanced

We assume that $g(\cdot)$ should:

- Pass (0,0) and (1,1)
- Increase monotonically
- Be spatially shared
- Be concave

Method

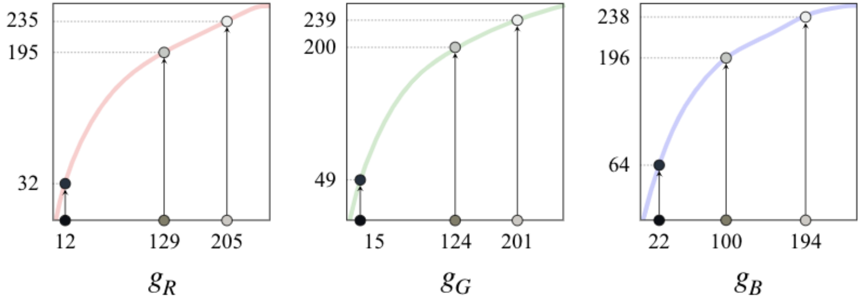
- **Network:** Deep Concave Curve

Idea: find a function g and use it to enhance the low-light input I_L

We assume that $g(\cdot)$ should:

- Pass (0,0) and (1,1)
- Increase monotonically
- Be spatially shared
- Be concave

Independent g for each color channel, *i.e.*, g_R, g_G, g_B



- **Training Strategy:** Asymmetric Self-supervised Alignment

Idea: employ high-level vision models as guidance

Existing strategies:

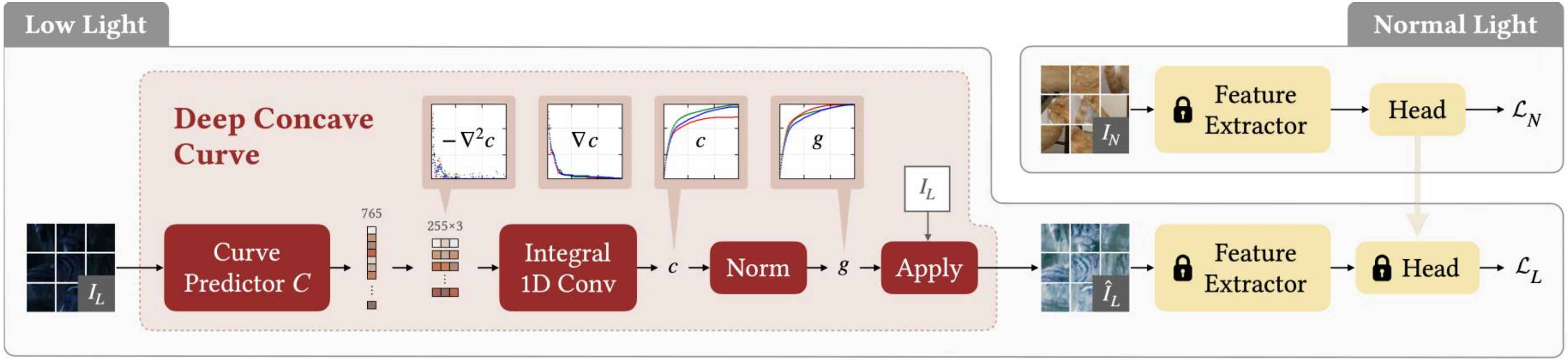
- Discrepancy metrics
- Adversarial learning

Bring extra semantic supervision, which can mislead our model and make training hard to converge

We use: **Cross-domain self-supervised pretext tasks**

Method

- **Training Strategy:** Asymmetric Self-supervised Alignment



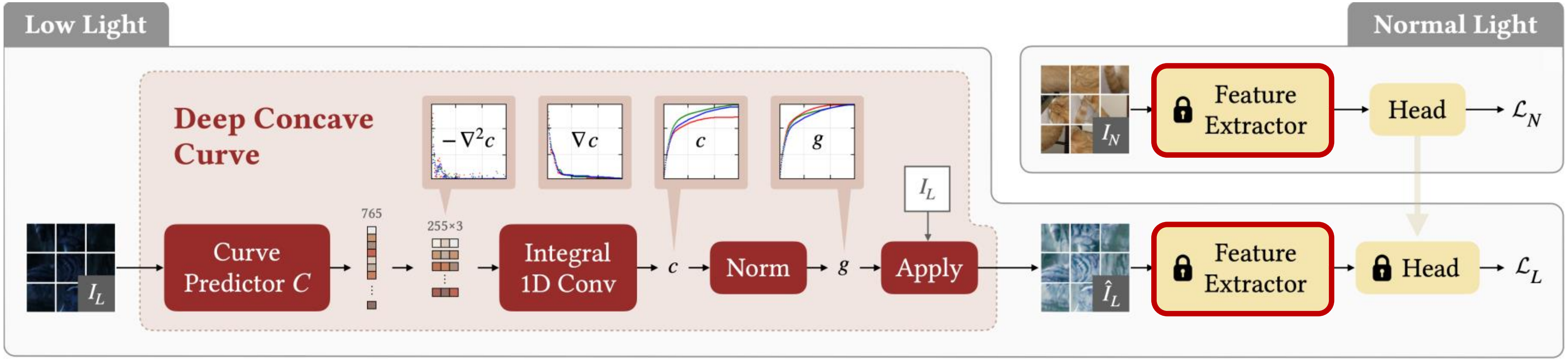
Pretrained and fixed-weight feature extractor

Step 1. Train a pretext head on normal light images

Step 2. Train deep concave curve with fixed-weight head on low-light images

Method

- **Training Strategy:** Asymmetric Self-supervised Alignment

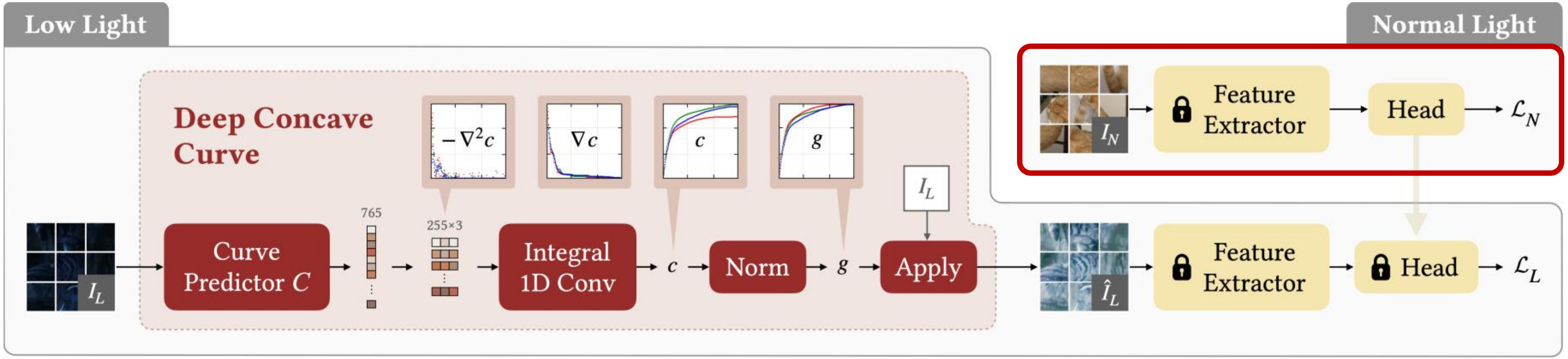


Pretrained and fixed-weight feature extractor

Step 1. Train a pretext head on normal light images

Step 2. Train deep concave curve with fixed-weight head on low-light images

- **Training Strategy:** Asymmetric Self-supervised Alignment

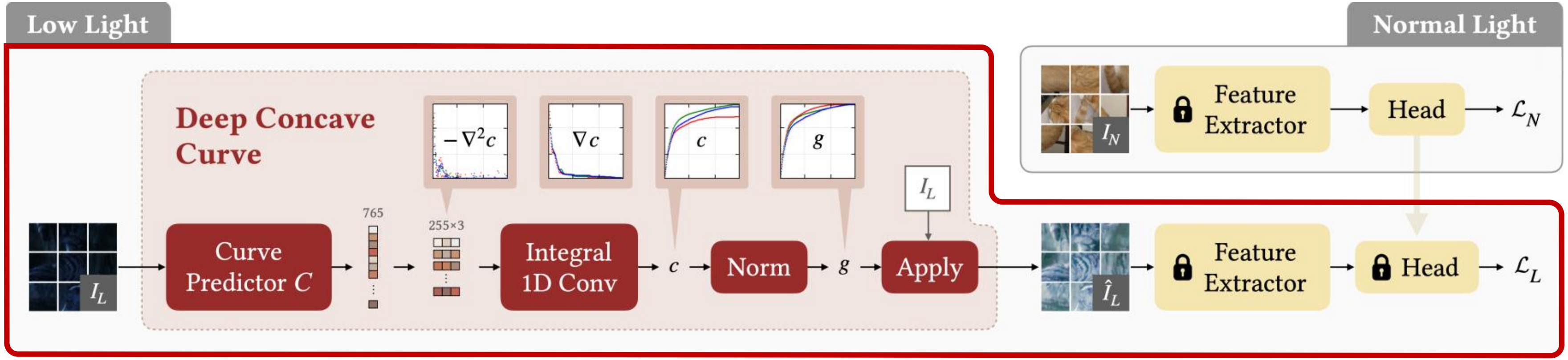


Pretrained and fixed-weight feature extractor

Step 1. Train a pretext head on normal light images

Step 2. Train deep concave curve with fixed-weight head on low-light images

- **Training Strategy:** Asymmetric Self-supervised Alignment

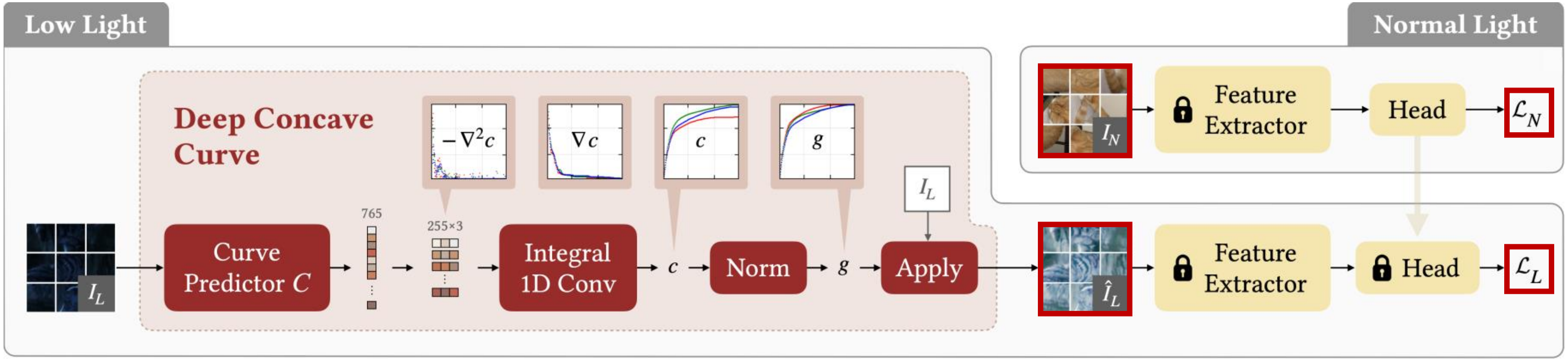


Pretrained and fixed-weight feature extractor

Step 1. Train a pretext head on normal light images

Step 2. Train deep concave curve with fixed-weight head on low-light images

- **Training Strategy:** Asymmetric Self-supervised Alignment



Pretext task: rotated jigsaw puzzles

First rotate the input image by random angles, then apply 3×3 jigsaw shuffling and ask the network to recognize the permutation

Method

- **Network + Training Strategy:** Self-Aligned Concave Curve (SACC)

Train: given a pretrained normal-light downstream model, use its feature backbone to train our deep concave curve

Test: enhance the input image and apply the downstream model

Properties:

- Blind to both normal and low-light annotations
- Does not need to adjust the downstream model

- **Network + Training Strategy:** Self-Aligned Concave Curve (SACC)

Can SACC solve color bias?

Yes, SACC has independent curves for each color channel.

How to solve camera noise?

Improving the robustness to noise is much easier than noise removal.

We adopt pseudo labeling to adapt the downstream model.

This advanced version is called SACC+.

- **Analysis: Why Self-Supervised Alignment?**

Table 1: Effects of different learning strategies for guiding our illumination enhancement model.

Category	Strategy	mAP (%)
Baseline	-	16.09
Discrepancy Metrics	CMD [68]	28.90
	MMD [2]	33.76
Adversarial Learning	LSGAN [43]	38.05
Self-Supervised Learning	MoCo [16]	38.46
	Rotation [12]	39.59
	Jigsaw [46]	41.01
	Rotated Jigsaw (proposed)	**41.31**

- **Analysis: Why Deep Concave Curve?**

Table 2: Effects of different low-light enhancement backbones under our asymmetric self-supervised strategy.

Category	Method	mAP (%)
Baseline	-	16.09
Other	EnlightenGAN [27]	18.42
	Zero-DCE [14]	25.36
	Gamma Correction x^{γ}	38.25
Ours	No Constraint	12.44
	$\nabla g \geq 0$	34.40
	$\nabla g \geq 0$ and $\nabla^2 g \leq 0$ (proposed)	**41.31**
	$\nabla g \geq 0$, $\nabla^2 g \leq 0$, and $\nabla^3 g \geq 0$	40.74

- **Analysis: Why Deep Concave Curve?**

Table 2: Effects of different low-
bones under our asymmetric self-

Category	Method
Baseline	-
Other	EnlightenGAN [27]
	Zero-DCE [14]
	Gamma Correction x
Ours	No Constraint
	$\nabla g \geq 0$
	$\nabla g \geq 0$ and $\nabla^2 g \leq 0$
	$\nabla g \geq 0$, $\nabla^2 g \leq 0$, and

Figure 3: Effects of asymmetric self-supervised learning with different low-light enhancement model backbones.

- **Analysis: Why Deep Conc**

Table 2: Effects of different l
bones under our asymmetric

Category	Method	
Baseline	-	
Other	EnlightenGAN	
	Zero-DCE [14]	
	Gamma Correct	
Ours	No Constraint	
	$\nabla g \geq 0$	34.40
	$\nabla g \geq 0$ and $\nabla^2 g \leq 0$ (proposed)	**41.31**
	$\nabla g \geq 0$, $\nabla^2 g \leq 0$, and $\nabla^3 g \geq 0$	40.74

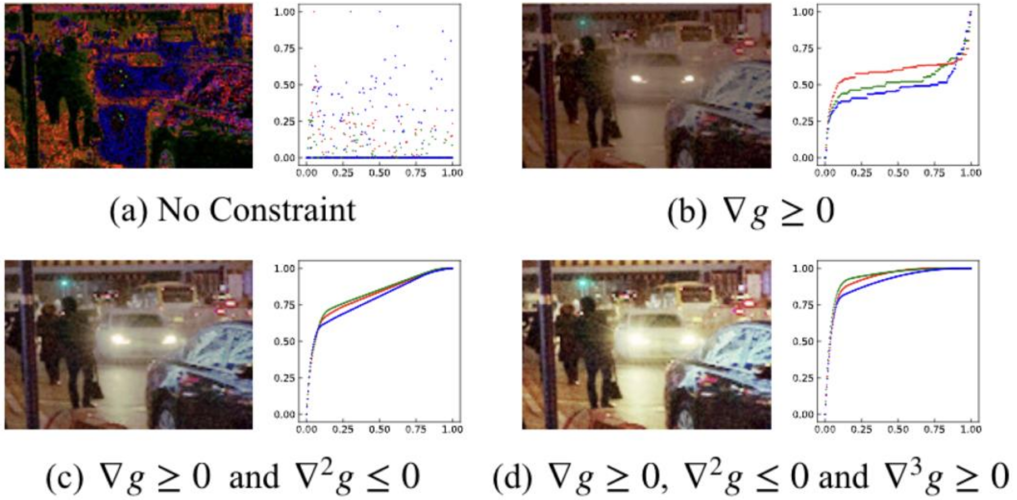


Figure 4: Low-light enhancement results (left) and curve shapes (right) under different curve form constraints.

- **Analysis: Why Deep Concave Curve?**

Table 3: Comparison with other low-light enhancement networks under different training strategies.

Method	Version	mAP (%)
Baseline	-	16.09
EnlightenGAN [27]	Original	28.85
	Retrained	20.08
	Original + Eq. (3)	32.88
Zero-DCE [14]	Original	38.32
	Retrained	32.53
	Original + Eq. (3)	40.54
SACC (Ours)	-	**41.31**

- **Analysis: Why Deep Concave Curve?**

Table 4: Effects of denoising enhanced low-light images and the proposed advanced version SACC+.

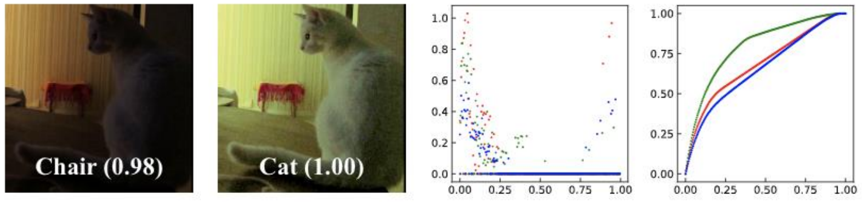
Method	Version	mAP (%)
Baseline	SACC	41.31
Denoising	SACC + BM3D [7]	24.51
	SACC + Neighbor2Neighbor [21]	40.54
Proposed	SACC+	**45.51**

Experiments

- **Settings**
  - Low-light Object Classification
  - Dark Face Detection
  - Low-light Action Recognition
  - Optical Flow Estimation in the Dark
  - Subjective Human Vision

Experiments

- Low-light Object Classification



(a) Original (b) SACC (c) $-\nabla^2 c$ (d) g

Figure 5: Example result for classification. The original image is wrongly categorized as “Chair” with a confidence of 0.98, after enhancement, it can be correctly recognized as “Cat”.

† denotes that the low-light enhancement model is retrained.

Category	Method	Top-1 (%)
Baseline	ResNet-18 [17]	55.04
Fully Supervised	Finetuned ResNet-18 [17]	71.52
Low-Light Enhancement	RetinexNet [60]	44.72
	Zero-DCE† [14]	50.80
	EnlightenGAN [27]	57.76
	Zero-DCE++ [34]	58.56
	MF [10]	59.20
	KinD [70]	59.28
	LIME [15]	59.44
	Zero-DCE [14]	59.44
	RUAS† [39]	59.84
	EnlightenGAN† [27]	60.24
	LLFlow [59]	60.72
	SACC (Ours)	**61.44**
Unsupervised Domain Adaptation	CMD [68]	55.92
	AdaBN [36]	59.68
	DANN [11]	59.76
	CIConv [32]	60.96
	SACC+ (Ours)	**63.92**

- Dark Face Detection

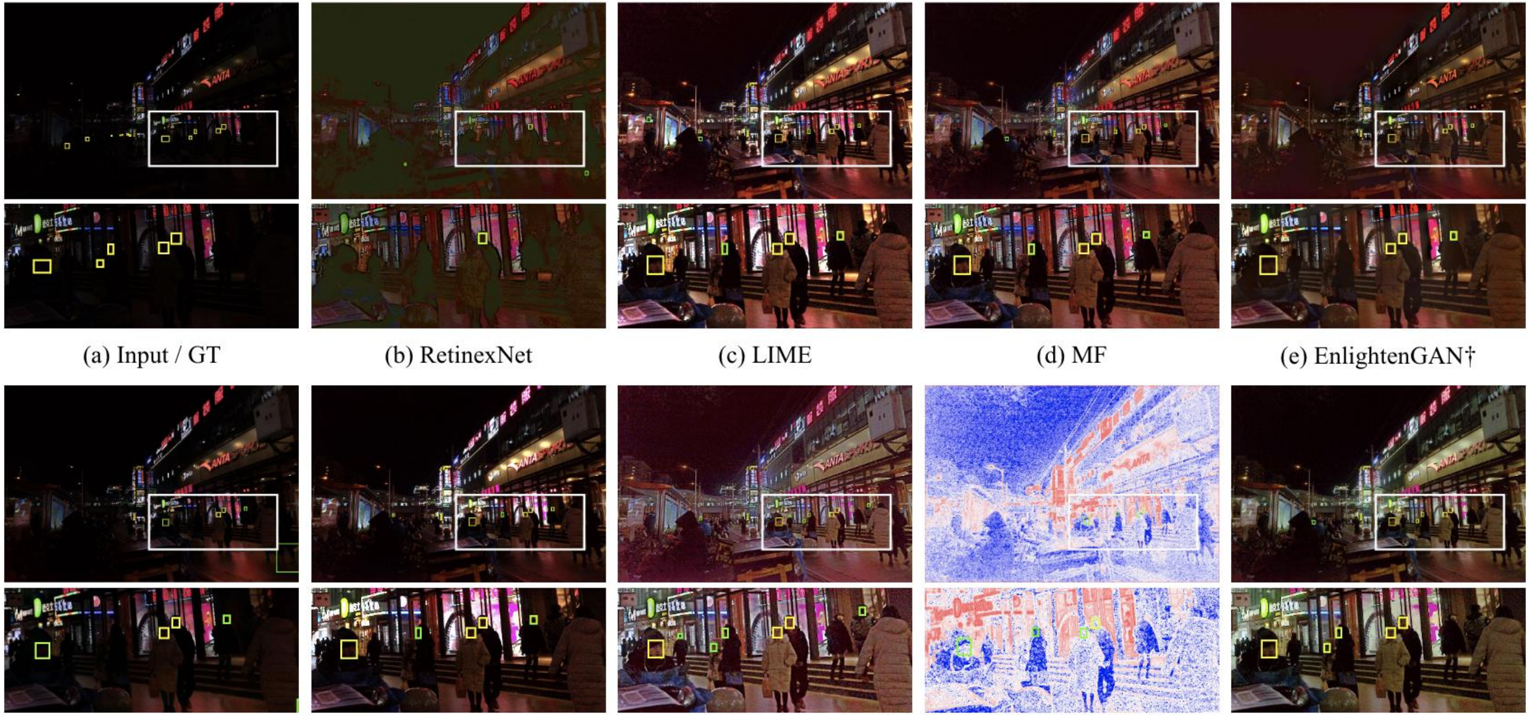
† denotes that the low-light enhancement model is retrained.

Category	Method	mAP (%)
Baseline	DSFD [35]	16.09
Fully Supervised	Finetuned DSFD [35]	45.99
Low-Light Enhancement	RetinexNet [60]	12.04
	KinD [70]	15.84
	EnlightenGAN† [27]	20.77
	EnlightenGAN [27]	31.31
	Zero-DCE † [14]	37.30
	LLFlow [59]	37.41
	RUAS† [39]	38.36
	LIME [15]	40.71
	Zero-DCE++ [34]	40.90
	Zero-DCE [14]	41.27
	MF [10]	41.43
	SACC (Ours)	**44.57**

Category	Method	mAP (%)
Unsupervised Domain Adaptation	CIConv [32]	4.40
	OSHOT [8]	25.38
	Progressive DA [18]	28.47
	Pseudo Labeling [23]	35.07
	HLA-Face [58]	44.44
	SACC+ (Ours)	**46.91**

Experiments

- Dark Face Detection



(a) Input / GT (b) RetinexNet (c) LIME (d) MF (e) EnlightenGAN†

(f) EnlightenGAN (g) RUAS† (h) HLA-Face (i) CIConv (j) SACC (Ours)

Experiments

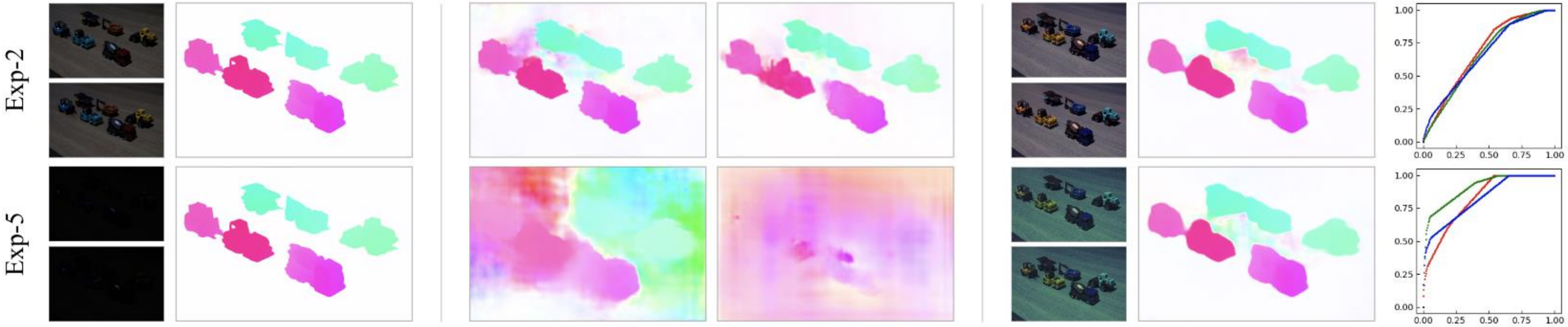
- Low-Light Action Recognition

Original	StableLLVE [69]	SMOID [25]	**SACC (Ours)**
47.14%	49.70%	50.18%	**52.13%**

Experiments

- Low-Light Action Recognition

	Exp-2	Exp-3	Exp-4	Exp-5	AVG
Original	10.21	11.90	14.36	17.82	13.57
Zheng *et al.* [71]	9.04	8.81	8.78	9.31	8.99
SACC (Ours)	**6.70**	**7.03**	**7.57**	**8.47**	**7.44**



(a) Input (b) Ground Truth (c) Original (d) Zheng *et al.* (e) SACC (Ours) (f) Our g

Take Home Messages

- Low-light condition: low visibility, low contrast, intensive noise
- Challenges: visual unpleasure (**Humans**), system failure (**Machines**)
- **Human Vision**: Robust Enhancement
  - Structural deep prior → deep Retinex
  - Dataset and benchmarking
- **Machine Vision**: Joint Enhancement and Downstream Tasks
  - Joint high-level and low-level adaptation
  - Self-aligned Concave Curve

北京大學
PEKING UNIVERSITY

THANKS

Jiaying Liu

Peking University

liujiaying@pku.edu.cn