恶劣场景下视觉感知与理解

Visual perception and understanding in degraded scenarios

报告人：刘家瑛

Visual Degradation

Data Acquisition

Degradation before Data Acquisition

- Heavy Rain/Snow
- Underwater
- Low Light
- Haze/Sandstorm

Degradation in Data Acquisition

- Downsample
- Motion Blur
- System Noise
- Optical Distortion

Degradation after Data Acquisition

- Scratches
- Watermark
- Package Loss
- Compression Loss

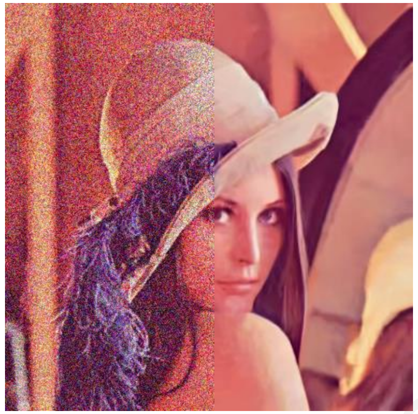
Visual Restoration and Enhancement

Underwater Enhancement

Dehazing

Sand Dust Removal

Rain Streak Removal



Denoising

Rain Drop Removal

Low Light Enhancement

Visual Restoration and Enhancement

- **Human vision** perspective
- The visual information in images and videos

Smartphone Camera

Digital Camera

Surveillance Camera

Visual Restoration and Enhancement

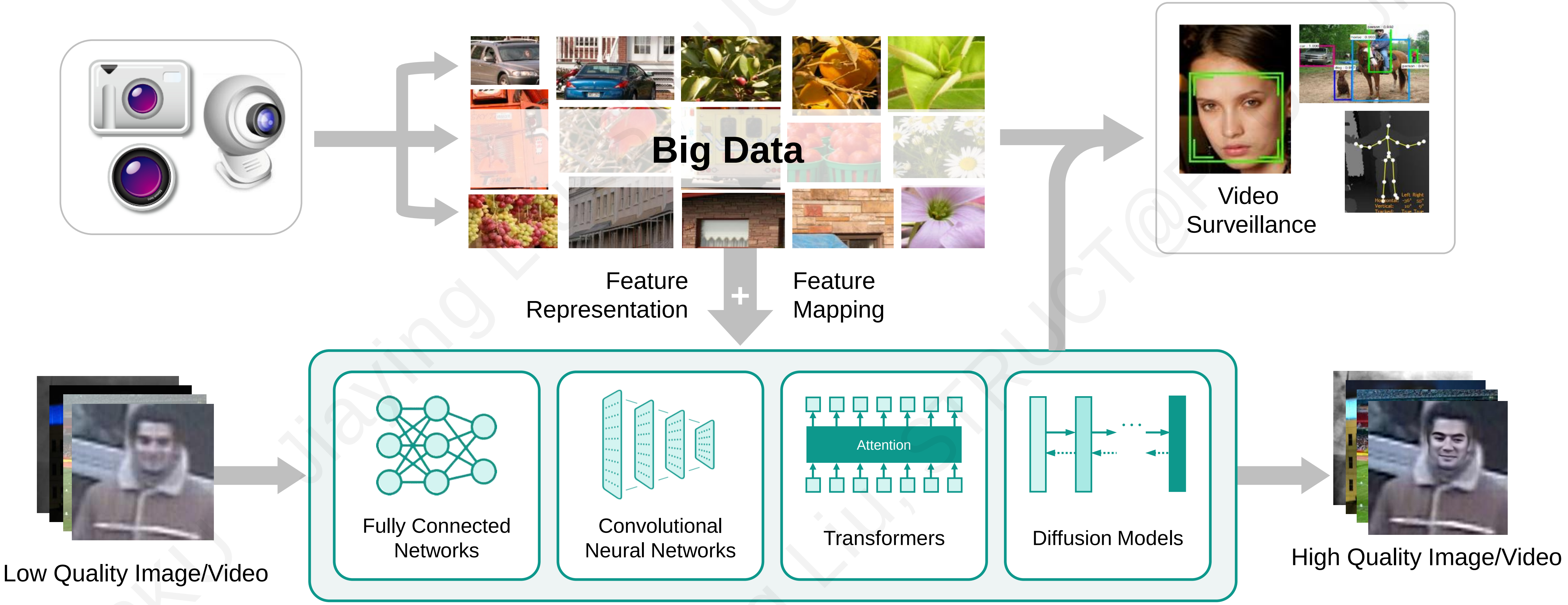
- **Machine vision** perspective
- Performance of downstream visual analysis

Autonomous driving

Smart security

Search engine

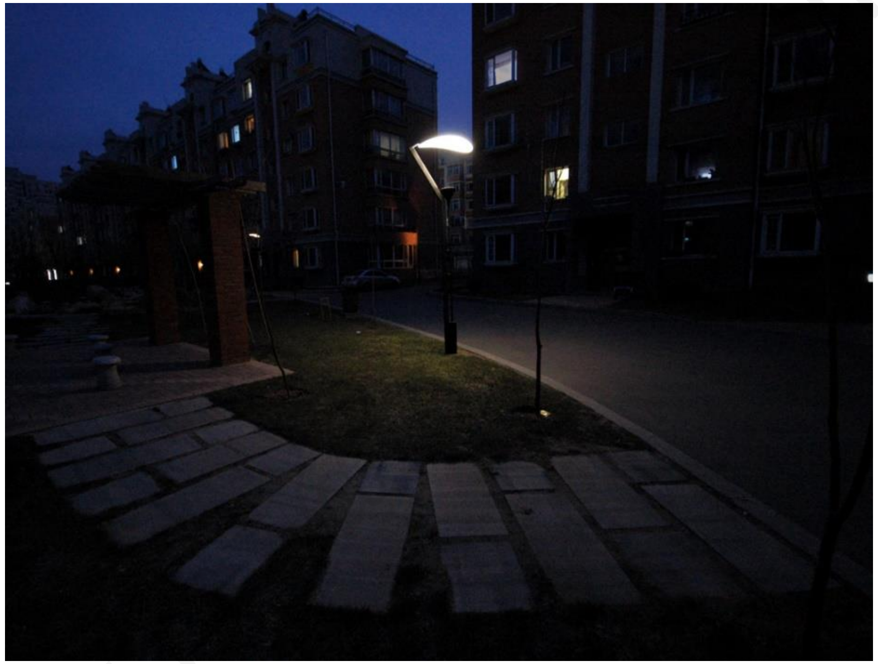
Visual Restoration and Enhancement



Data-Driven Solution

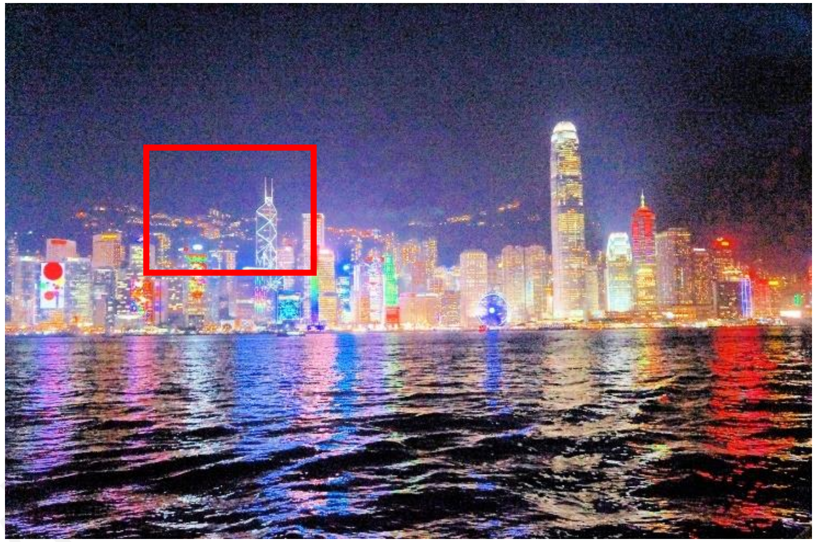
Low-light Degradation

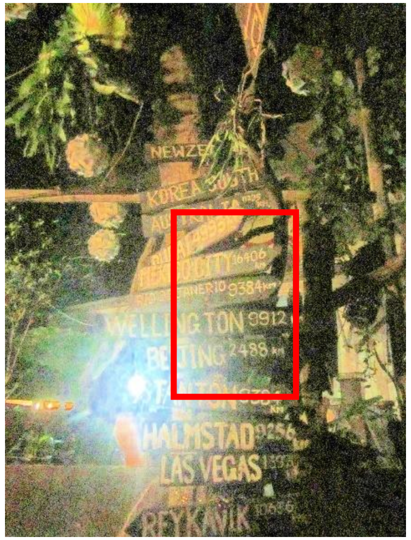
- **Low visibility**
  - Details are buried due to degraded contrast and low illumination



Low-light Degradation

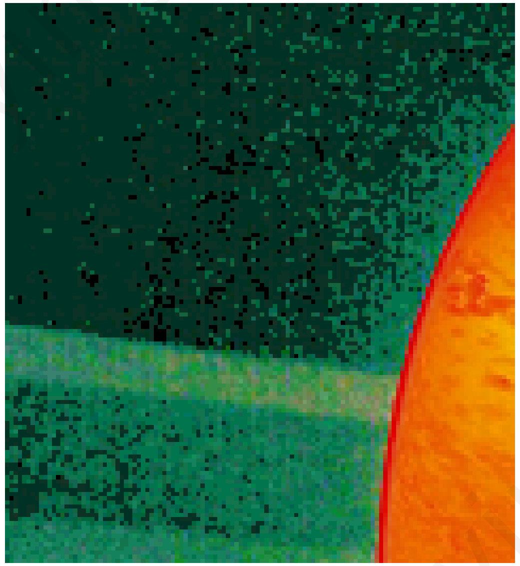
- **Intensive noises**
  - After simple operations, e.g. histogram equalization, noises become visible





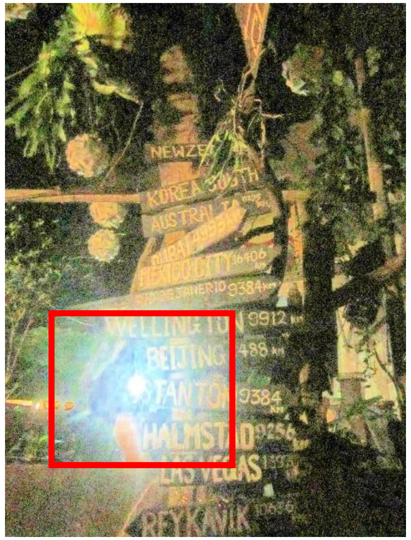
Low-light Degradation

- **Intensive noises**
  - After simple operations, e.g. histogram equalization, noises become visible



Low-light Degradation

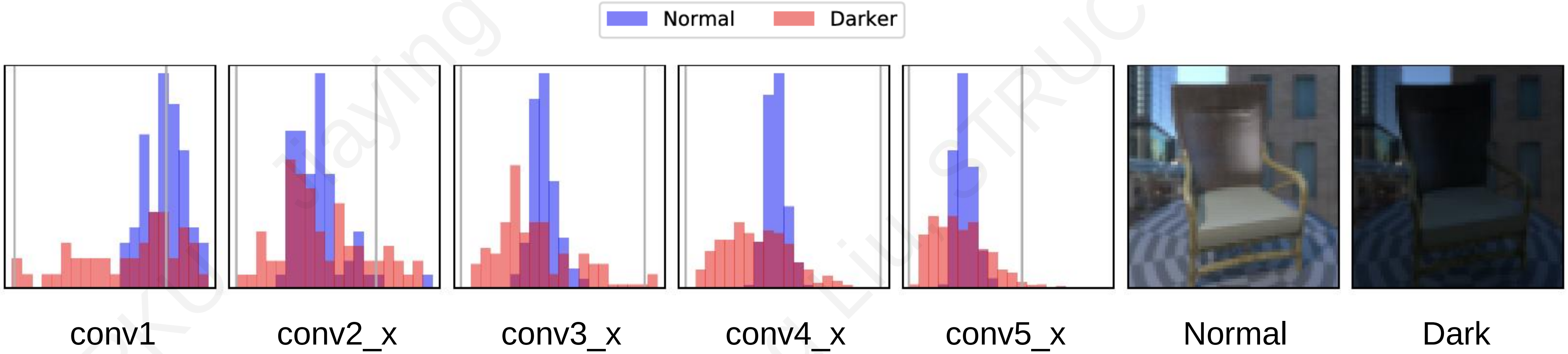
- **Non-uniform illumination**
  - Under-exposures
  - Over-exposures



Low-light Degradation

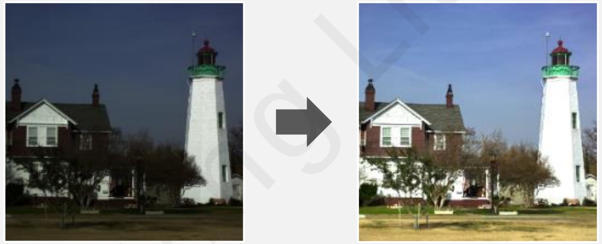
- **Visual feature distribution shift**
  - Models pretrained in general scenarios give incorrect predictions

Feature map activations in various layers of a baseline ResNet-18



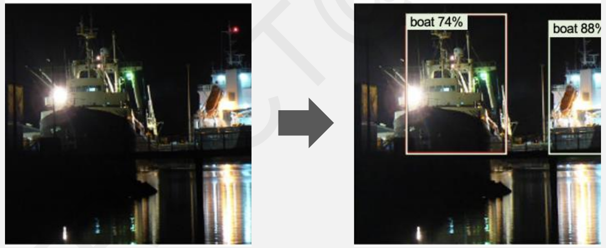
Visual perception and understanding

Human vision



Visual perception

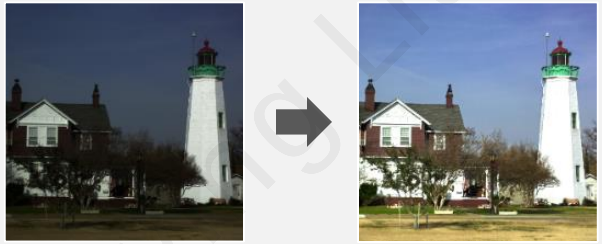
Machine vision



Visual understanding

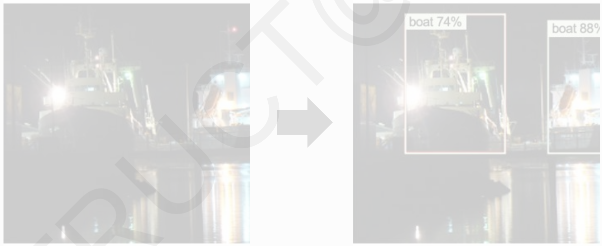
Visual perception and understanding

Human vision



Visual perception

Machine vision



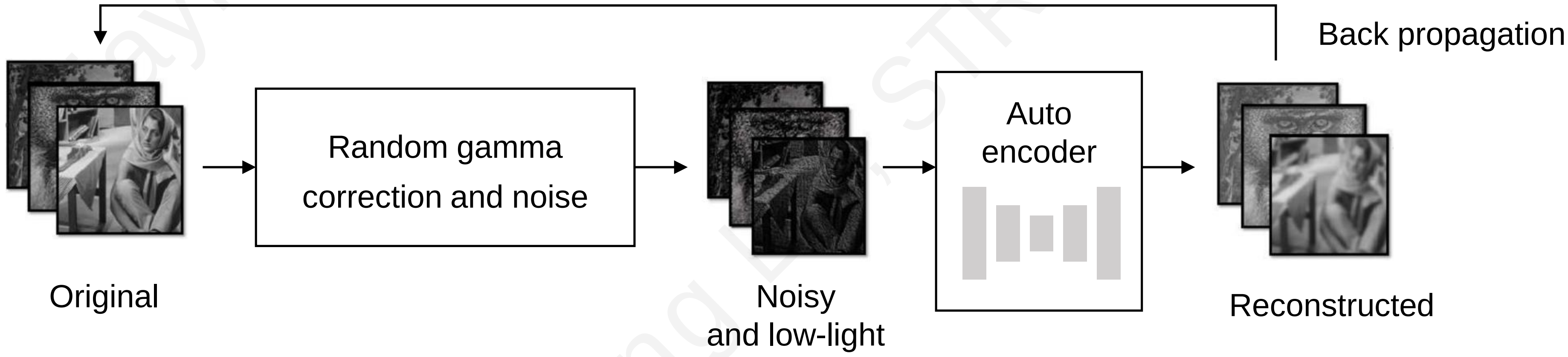
Visual understanding

Data-Driven Low-Light Enhancement

- Supervised
- Reinforcement
- Semi-supervised
- Unsupervised
- Zero-Reference

- The first deep-based method for low-light enhancement
- End-to-end, synthetic training data

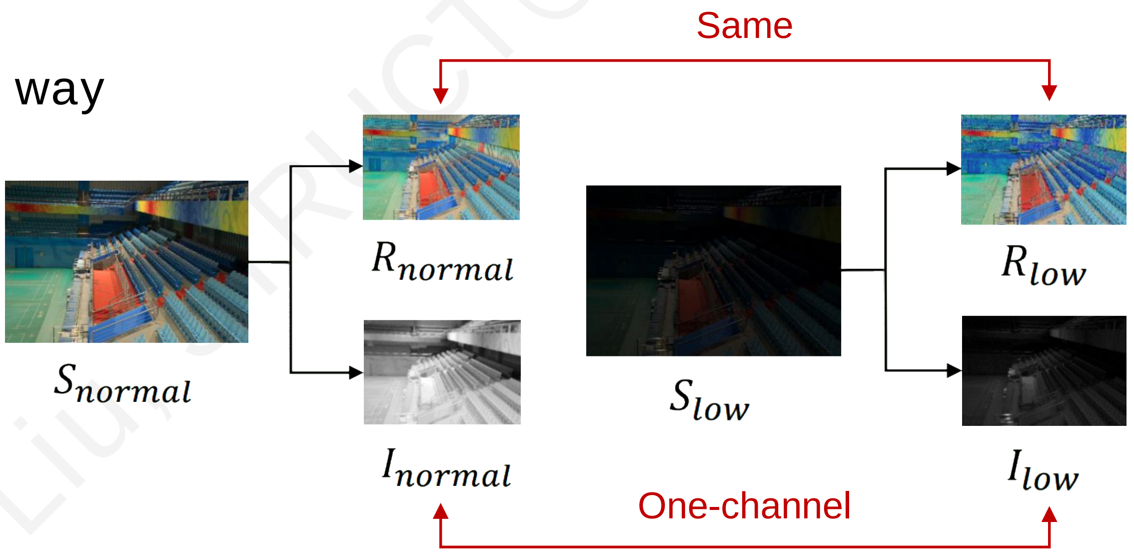
Kin Gwn Lore, Adedotun Akintayo, Soumik Sarkar: LLNet: A deep auto-encoder approach to natural low-light image enhancement. PR 2017



Data-Driven Low-Light Enhancement

- Supervised
- Reinforcement
- Semi-supervised
- Unsupervised
- Zero-Reference

- Introduce Retinex into deep models
- Learn Retinex decomposition in an unsupervised way
  - R is the same
  - L is different
- Real-captured dataset
  - 500 image pairs
  - Support training and evaluation



Chen Wei, Wenjing Wang, Wenhan Yang, Jiaying Liu: Deep Retinex Decomposition for Low-Light Enhancement. BMVC 2018

Data-Driven Low-Light Enhancement

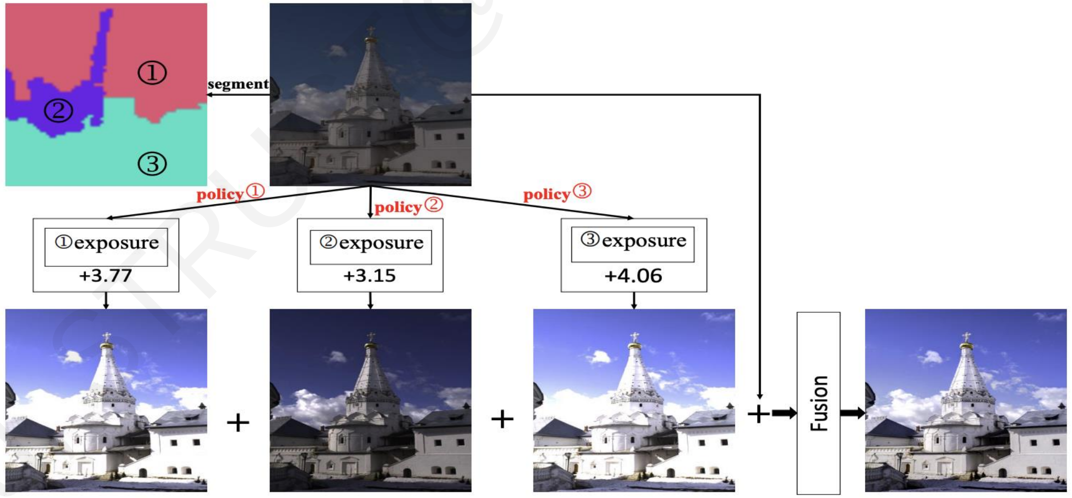
- Supervised
- Reinforcement
- Semi-supervised
- Unsupervised
- Zero-Reference

2017 PR
LLNet

2018 BMVC
Retinex-Net

2018 NeurIPS
DeepExposure

- **Reinforcement Learning + adversarial learning**
- Segment an image into sub-images
  - Learn a local exposure for each sub-image
  - Reward: a balance of overall exposures
- Aesthetic evaluation by a discriminator



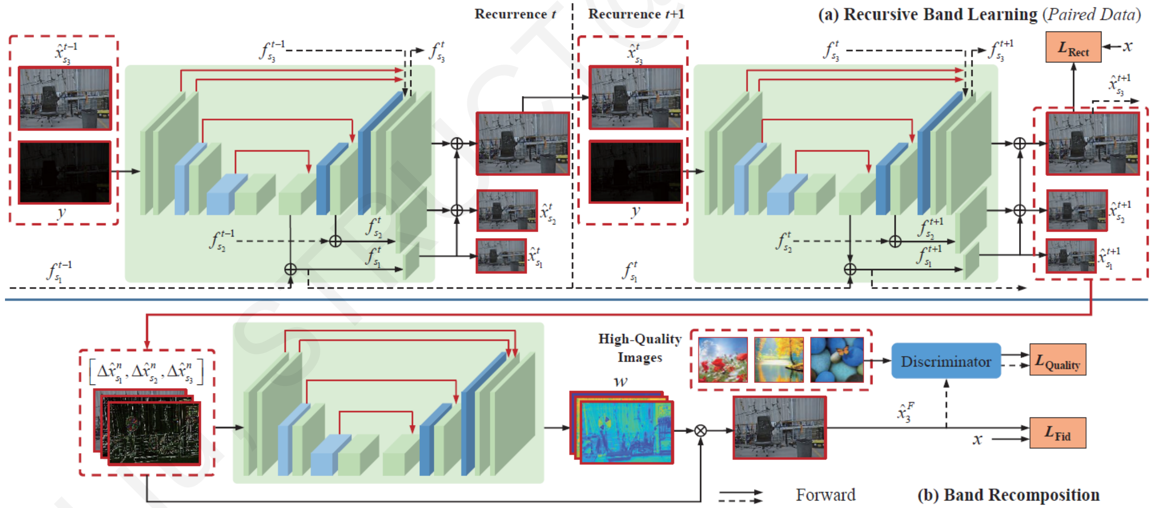
Runsheng Yu, Wenyu Liu, Yasen Zhang, Zhi Qu, Deli Zhao, Bo Zhang: DeepExposure: Learning to Expose Photos with Asynchronously Reinforced Adversarial Learning. NeurIPS 2018

Data-Driven Low-Light Enhancement

- Supervised
- Reinforcement
- Semi-supervised
- Unsupervised
- Zero-Reference

2017 PR	2018 BMVC	2018 NeurIPS	2020 CVPR
LLNet	Retinex-Net	DeepExposure	DRBN

- Semi-supervised learning
  - **Supervised**: recover a linear band representation of an enhanced normal-light image
  - **Unsupervised**: recompose the given bands via another learnable linear transformation



Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, Jiaying Liu: From Fidelity to Perceptual Quality: A Semi-Supervised Approach for Low-Light Image Enhancement. CVPR 2020

Data-Driven Low-Light Enhancement

Supervised
Reinforcement
Semi-supervised
Unsupervised
Zero-Reference

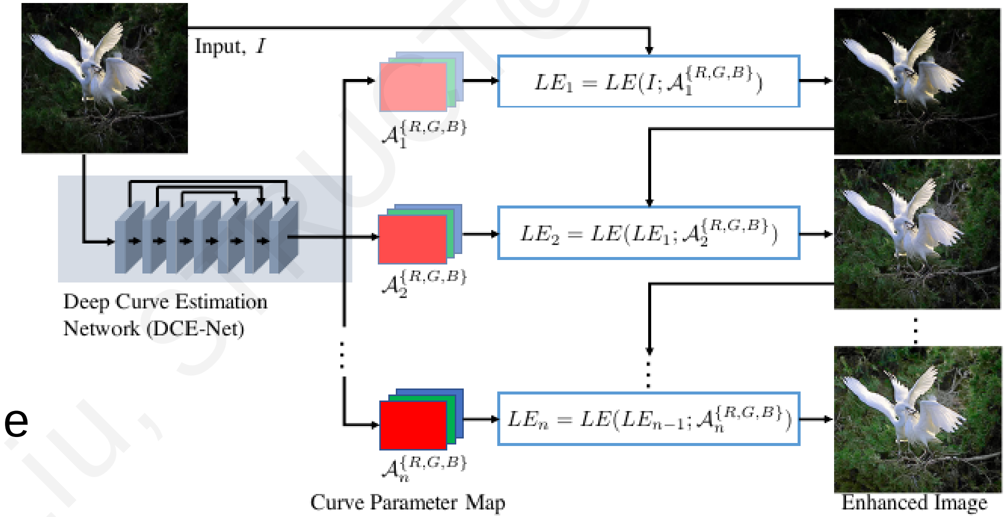
2017 PR LLNet | 2018 BMVC Retinex-Net | 2018 NeurIPS DeepExposure | 2020 CVPR DRBN | 2020 CVPR Zero-DCE

- **Nonlinear curve mapping**
  - Pixel-wise and high-order curves

$$LE(x, A) = x + Ax(1 - x)$$

$$LE_n = LE(LE_{n-1}; A_n)$$

- No need for paired normal/low light data
  - **Non-reference loss functions** for exposure control, color constancy, *etc.*



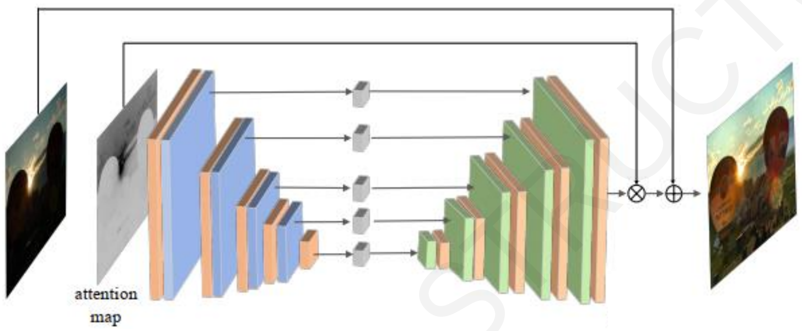
Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, Runmin Cong: Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. CVPR 2020

Data-Driven Low-Light Enhancement

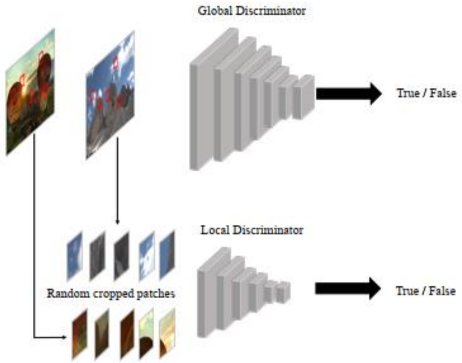
- Supervised
- Reinforcement
- Semi-supervised
- Unsupervised
- Zero-Reference

2017 PR	2018 BMVC	2018 NeurIPS	2020 CVPR	2020 CVPR	2021 IEEE TIP
LLNet	Retinex-Net	DeepExposure	DRBN	Zero-DCE	EnlightenGAN

- **Unpaired** normal/low light data
- **Adversarial learning**
  - Generator
    - Low-light enhancement
  - Discriminator



U-Net Generator
Guided with Self-Regularized Attention



Global-local discriminator

Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, Zhangyang Wang: EnlightenGAN: Deep Light Enhancement Without Paired Supervision. IEEE TIP 2021

Data-Driven Low-Light Enhancement

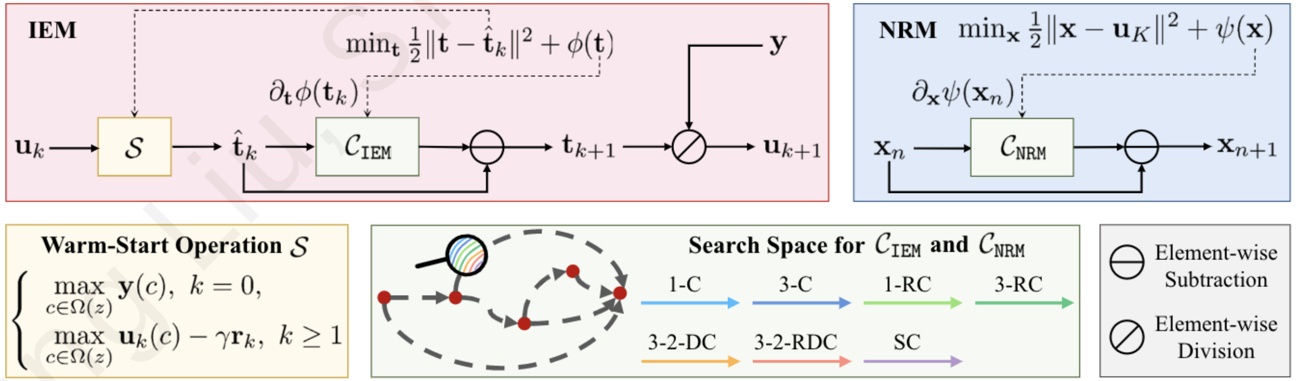
Supervised
Reinforcement
Semi-supervised
Unsupervised
Zero-Reference

2017 PR	2018 BMVC	2018 NeurIPS	2020 CVPR	2020 CVPR	2021 IEEE TIP	2021 CVPR
LLNet	Retinex-Net	DeepExposure	DRBN	Zero-DCE	EnlightenGAN	RUAS

Retinex-inspired Unrolling with **Architecture Search**

- Unroll the optimization of Retinex-inspired models
  - Illumination Estimation Module (IEM)
  - Noise Removal Module (NRM)
- Cooperative Architecture Search
  - For both illumination estimation and noise removal

Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, Zhongxuan Luo: Retinex-Inspired Unrolling With Cooperative Prior Architecture Search for Low-Light Image Enhancement. CVPR 2021

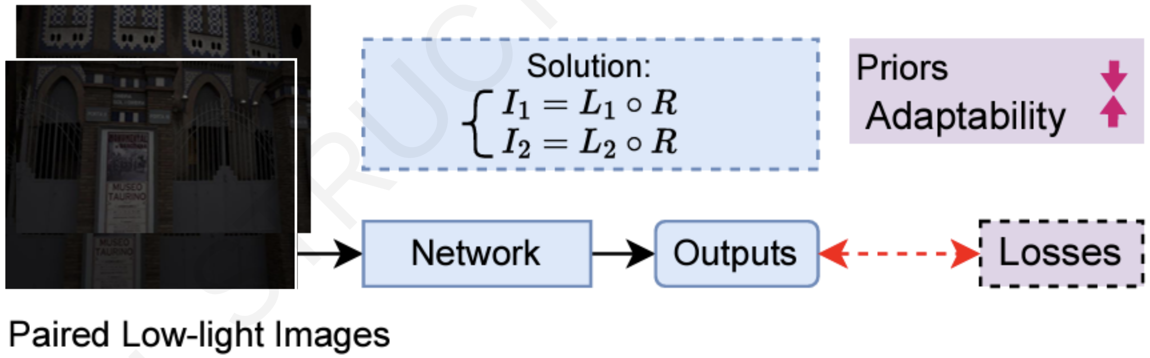


Data-Driven Low-Light Enhancement

- Supervised
- Reinforcement
- Semi-supervised
- Unsupervised
- Zero-Reference

2017 PR LLNet → 2018 BMVC Retinex-Net → 2018 NeurIPS DeepExposure → 2020 CVPR DRBN → 2020 CVPR Zero-DCE → 2021 IEEE TIP EnlightenGAN → 2021 CVPR RUAS → 2023 CVPR PairLIE

- Learn from **paired low-light** images
  - Given paired low-light inputs, the model should produce the same results
  - **Retinex theory**, two reflectance components are consistent
- Remove inappropriate features in the raw image with a simple self-supervised mechanism



Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, Kai-Kuang Ma: Learning a Simple Low-Light Image Enhancer from Paired Low-Light Instances. CVPR 2023

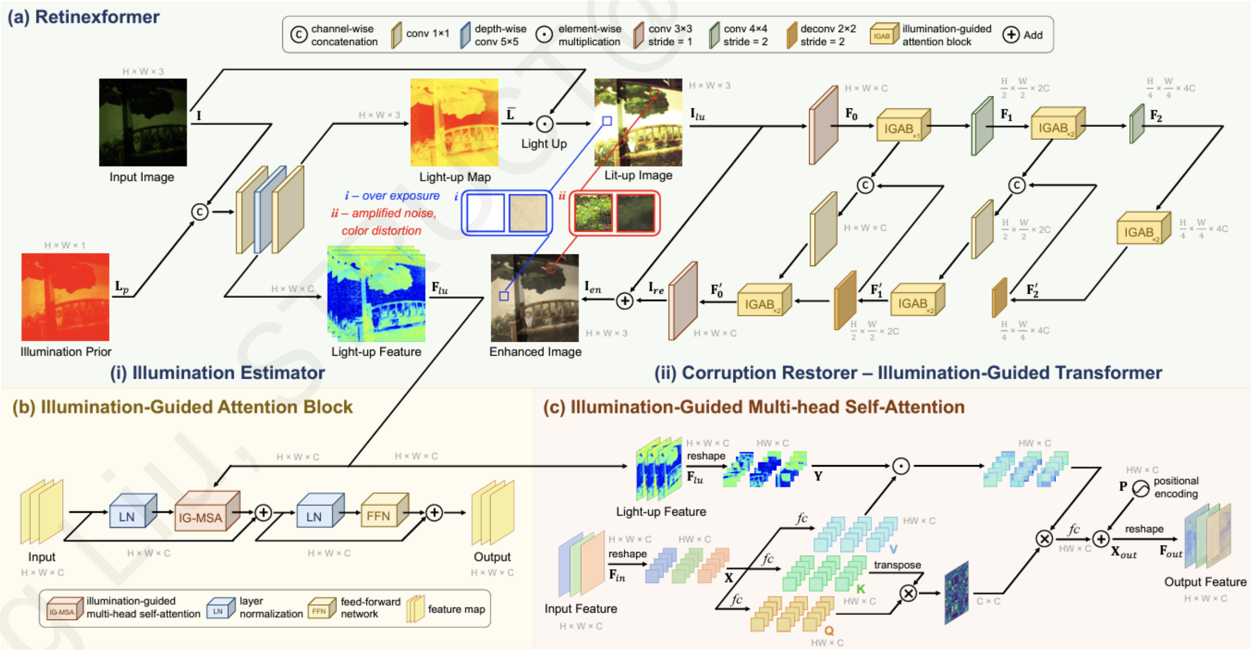
Data-Driven Low-Light Enhancement

- Supervised
- Reinforcement
- Semi-supervised
- Unsupervised
- Zero-Reference

2017 PR	2018 BMVC	2018 NeurIPS	2020 CVPR	2020 CVPR	2021 IEEE TIP	2021 CVPR	2023 CVPR	2023 ICCV
LLNet	Retinex-Net	DeepExposure	DRBN	Zero-DCE	EnlightenGAN	RUAS	PairLIE	Retinexformer

One-stage Retinex-based **Transformer**

- One-stage Retinex-based Framework
  - Estimate the illumination information
  - Restore the corruption
- Illumination-Guided Transformer
  - Model non-local interactions of regions with different lighting conditions



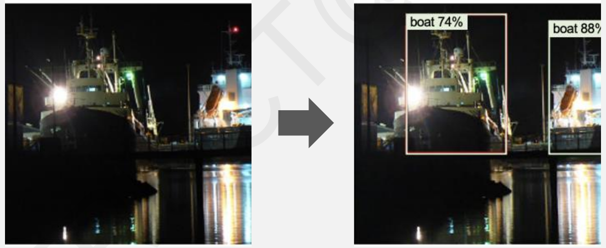
Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, Yulun Zhang: Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement. ICCV 2023

Visual perception and understanding

Human vision

Visual perception

Machine vision

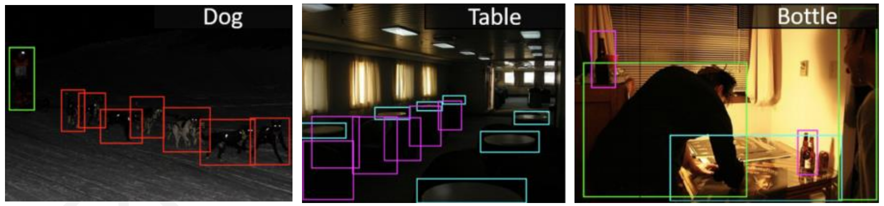


Visual understanding

Visual understanding

- Visual understanding datasets in low-light scenarios

CODaN: Image Classification [1]



ExDark: Object Detection [2]

Dark Zurich: Semantic Segmentation [3]

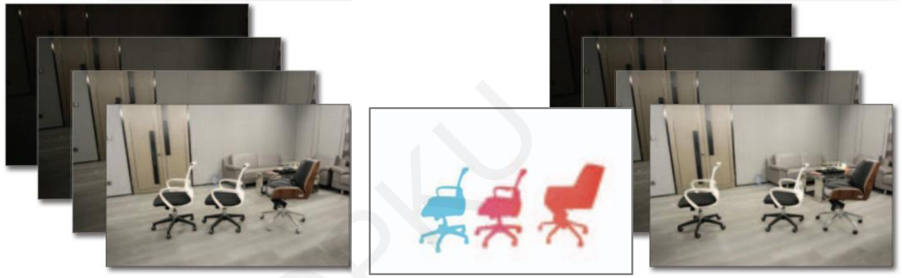
[1] Attila Lengyel, *et al.*: Zero-Shot Day-Night Domain Adaptation with a Physics Prior. ICCV 2021

[2] Yuen Peng Loh, Chee Seng Chan: Getting to know low-light images with the Exclusively Dark dataset. CVIU 2019

[3] Christos Sakaridis, *et al.*: Guided Curriculum Model Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation. ICCV 2019

Visual understanding

- Visual understanding datasets in low-light scenarios



VBOF: Optical Flow Estimation [2]

DARK FACE: Face Detection [1]

ARID: Video Action Recognition [3]

[1] Wenhan Yang, *et al.*: Advancing Image Understanding in Poor Visibility Environments: A Collective Benchmark Study. IEEE TIP 2020

[2] Yinqiang Zheng, *et al.*: Optical Flow in the Dark. CVPR 2020

[3] Yuecong Xu, *et al.*: ARID: A New Dataset for Recognizing Action in the Dark. DL-HAR 2021

Visual understanding

Normal light vs. low light vision datasets

Task	Illumination	Dataset	Samples	Classes
Image Classification	Normal Light	ImageNet	14M	22K
	Low Light	CODaN	2.5K	10
Object Detection	Normal Light	COCO	123K	80
	Low Light	ExDark	7K	12
Semantic Segmentation	Normal Light	Cityscapes	25K	19
	Low Light	Dark Zurich	5K	19
Face Detection	Normal Light	WIDER FACE	32K	1
	Low Light	DARK FACE	10K	1

Visual understanding

Simple strategy: Build a low-light dataset and train the model

- Shortcoming
  - Building new datasets takes effort and resources
  - Lack of robustness and scalability

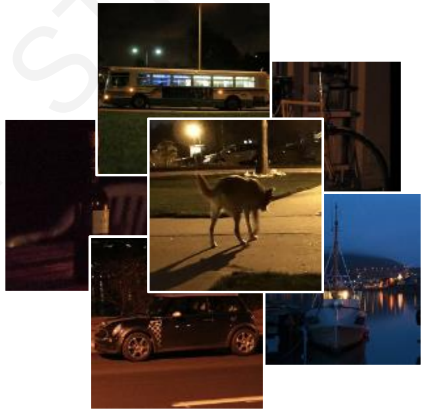
Visual understanding

- Unsupervised domain adaptation

Source domain with labels → Model adaptation → Target domain without labels

- Normal-to-low light unsupervised domain adaptation

Model adaptation



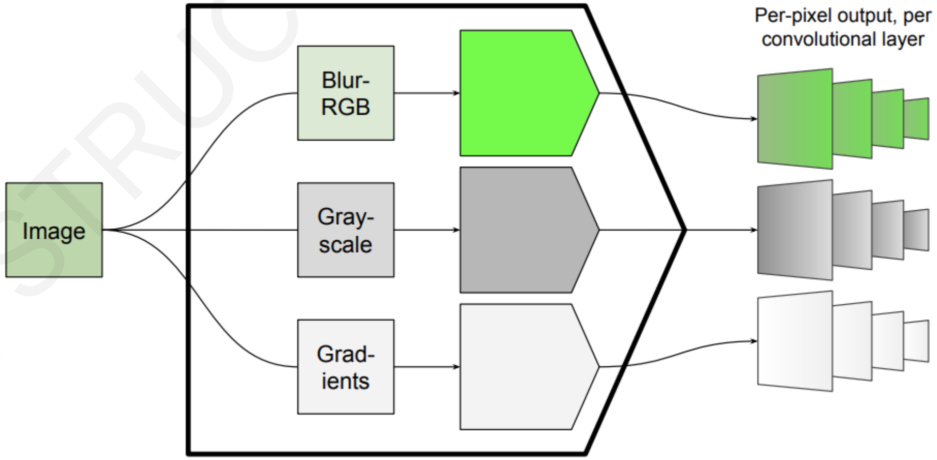
Low Light Unsupervised Domain Adaptation

Detection
Others
Segmentation
Multiple

2019 IEEE ICRA
ToDayGAN

Task: image retrieval across different illumination conditions

- **Night-to-day image-translation**
  - Unpaired image-to-image translation
  - Three discriminators：
    Blur-RGB, Gray-scale, Gradients
- Provide enhanced images to downstream models



Asha Anoosheh, Torsten Sattler, Radu Timofte, Marc Pollefeys, Luc Van Gool: Night-to-Day Image Translation for Retrieval-based Localization. ICRA 2019

Low Light Unsupervised Domain Adaptation

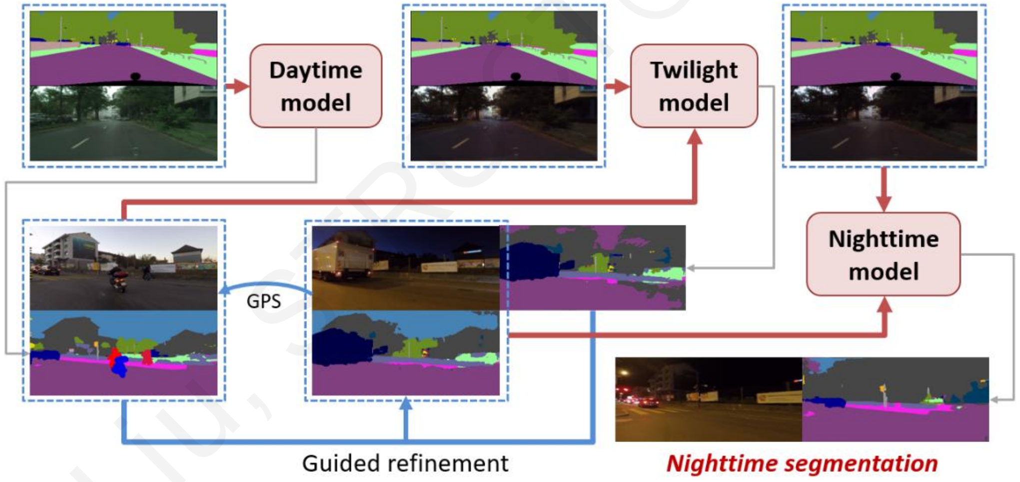
Detection | Others | Segmentation | Multiple

2019 IEEE ICRA
ToDayGAN

2019 ICCV
GCMA

Task: nighttime semantic segmentation

- **Gradual adaptation**
  - Source domain: Daytime
  - Middle domain: Twilight time
  - Target domain: Nighttime
- GPS-based nearest neighbor assignment



Red: model training

Gray: model prediction

Christos Sakaridis, Dengxin Dai, Luc Van Gool: Guided Curriculum Model Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation. ICCV 2019

Low Light Unsupervised Domain Adaptation

Detection Others
Segmentation Multiple

2019 IEEE ICRA
ToDayGAN

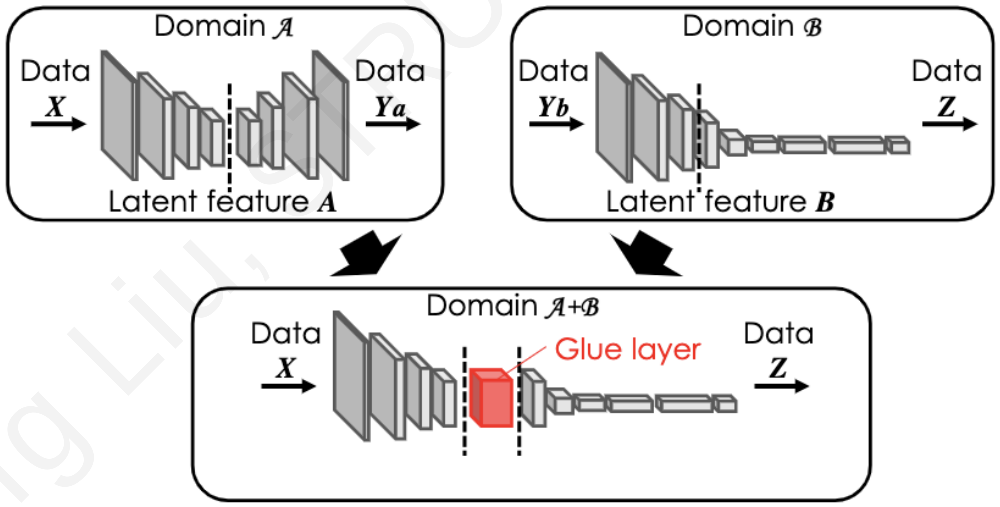
2019 ICCV
GCMA

2020 ECCV
YOLO-in-the-Dark

Task: low-light object detection

- **Merge two models trained in different domains**
  - Glue layer
  - Knowledge distillation

Domain 𝒜 : train by RAW image with annotation based on RGB image
Domain ℬ : train by RGB image with annotation (object label, bounding boxes)



Yukihiro Sasagawa, Hajime Nagahara: YOLO in the Dark - Domain Adaptation Method for Merging Multiple Models. ECCV 2020

Low Light Unsupervised Domain Adaptation

Detection
Others
Segmentation
Multiple

2019 IEEE ICRA
ToDayGAN

2019 ICCV
GCMA

2020 ECCV
YOLO-in-the-Dark

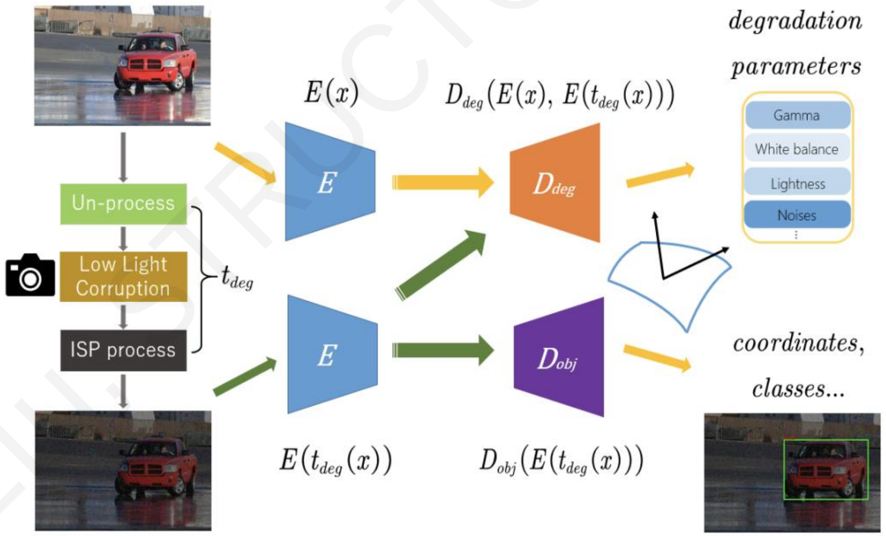
2021 ICCV
MAET

Task: low-light object detection

- **Multi-Task auto-encoding transformations**

$$\hat{t} = D[E(x), E(t(x))]$$

  - Decode low-light degradation
  - Decode object position and categories



Ziteng Cui, Guo-Jun Qi, Lin Gu, Shaodi You, Zenghui Zhang, Tatsuya Harada: Multitask AET with Orthogonal Tangent Regularity for Dark Object Detection. ICCV 2021

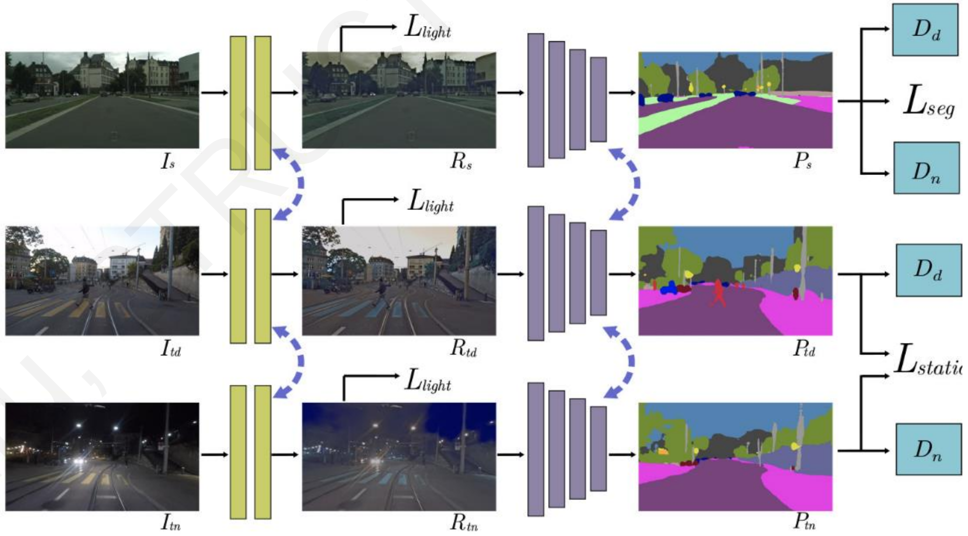
Low Light Unsupervised Domain Adaptation

Detection
Others
Segmentation
Multiple

2019 IEEE ICRA	2019 ICCV	2020 ECCV	2021 ICCV	2021 CVPR
ToDayGAN	GCMA	YOLO-in-the-Dark	MAET	DANNet

Task: low-light semantic segmentation

- **Gradual adaptation**
  - Source domain: Daytime
  - Middle domain: Twilight time
  - Target domain: Nighttime
- **Adversarial training** with labeled daytime and unlabeled coarsely aligned day-night pairs.



Xinyi Wu, *et al.*: DANNet: A One-Stage Domain Adaptation Network for Unsupervised Nighttime Semantic Segmentation. CVPR 2021

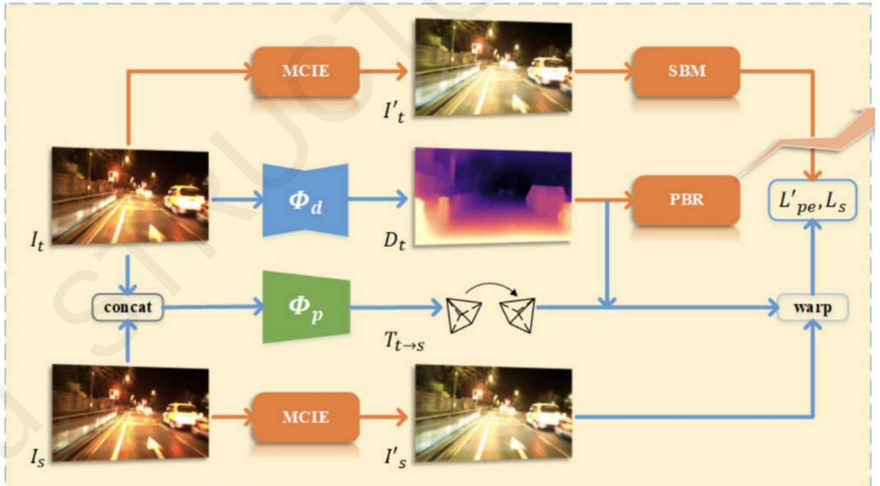
Low Light Unsupervised Domain Adaptation

Detection · Others · Segmentation · Multiple

2019 IEEE ICRA ToDayGAN	2019 ICCV GCMA	2020 ECCV YOLO-in-the-Dark	2021 ICCV MAET	2021 CVPR DANNet	2021 ICCV RNW

Task: depth estimation in the dark

- Low-light enhancement based on CLHE
- Priors-Based Regularization
  - Learn from unpaired references in an adversarial manner
- Statistics-based masking strategy



Nighttime Self-Supervised Depth Estimation Framework

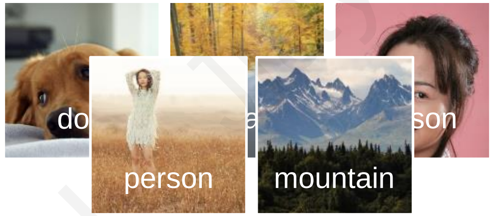
Kun Wang, Zhenyu Zhang, Zhiqiang Yan, Xiang Li, Baobei Xu, Jun Li, Jian Yang: Regularizing Nighttime Weirdness: Efficient Self-supervised Monocular Depth Estimation in the Dark. ICCV 2021

Low Light Unsupervised Domain Adaptation

- Detection
- Segmentation
- Others
- Multiple

2019 IEEE ICRA ToDayGAN | 2019 ICCV GCMA | 2020 ECCV YOLO-in-the-Dark | 2021 ICCV MAET | 2021 CVPR DANNet | 2021 ICCV RNW | 2021 ICCV CIConv

- Unsupervised adaptation



Labeled daytime data

+

Unlabeled nighttime data

adapt

Nighttime conditions

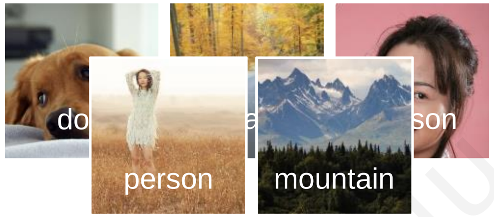
Attila Lengyel, *et al*.: Zero-Shot Day-Night Domain Adaptation with a Physics Prior. ICCV 2021

Low Light Unsupervised Domain Adaptation

- Detection
- Others
- Segmentation
- Multiple

2019 IEEE ICRA	2019 ICCV	2020 ECCV	2021 ICCV	2021 CVPR	2021 ICCV	2021 ICCV
ToDayGAN	GCMA	YOLO-in-the-Dark	MAET	DANNet	RNW	CIConv

- Unsupervised adaptation
- **Zero-shot adaptation**



Labeled daytime data

adapt

Nighttime conditions

Attila Lengyel, *et al*.: Zero-Shot Day-Night Domain Adaptation with a Physics Prior. ICCV 2021

Low Light Unsupervised Domain Adaptation

- Detection
- Others
- Segmentation
- Multiple

2019 IEEE ICRA ToDayGAN | 2019 ICCV GCMA | 2020 ECCV YOLO-in-the-Dark | 2021 ICCV MAET | 2021 CVPR DANNet | 2021 ICCV RNW | 2021 ICCV CIConv

Task: various vision tasks, **zero-shot**

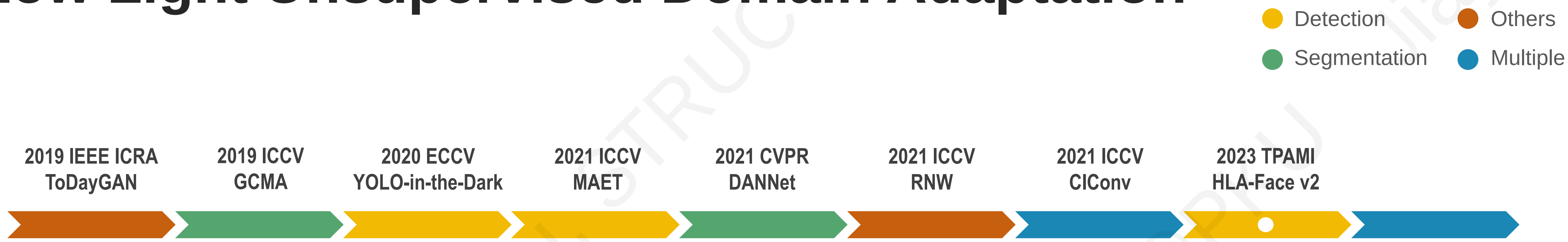
- Physics-based visual inductive prior
  - Illumination invariant
- CIConv layer, learn illumination invariant information

$$\mathrm{CIConv}(x,y)=\frac{\log\left(\mathrm{CI}^2(x,y,\sigma=2^s)+\epsilon\right)-\mu_S}{\sigma_S}$$

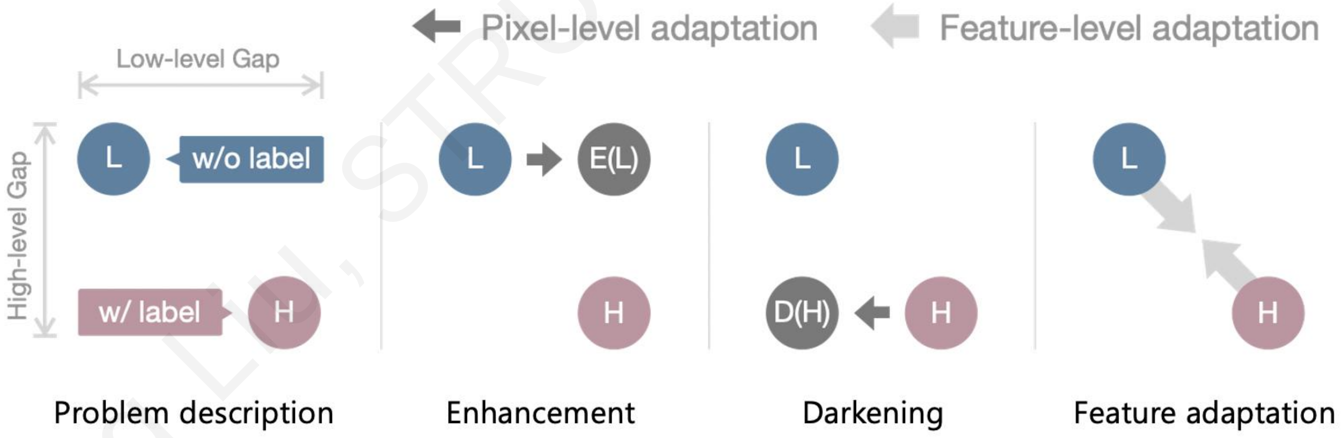
Input $\quad\sigma=0.50$

Attila Lengyel, *et al*.: Zero-Shot Day-Night Domain Adaptation with a Physics Prior. ICCV 2021

Low Light Unsupervised Domain Adaptation



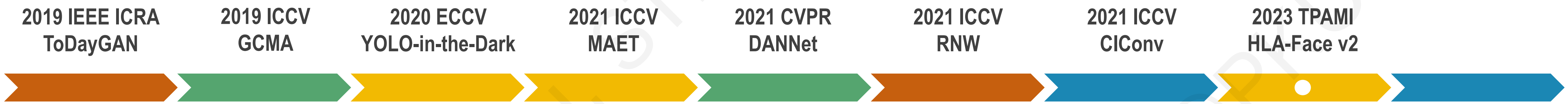
- Categorize adaptive low light techniques into three types
  - **Enhancement** and **darkening** only consider the pixel-level gap
  - **Feature adaptation** methods try to fill the whole gap in one step



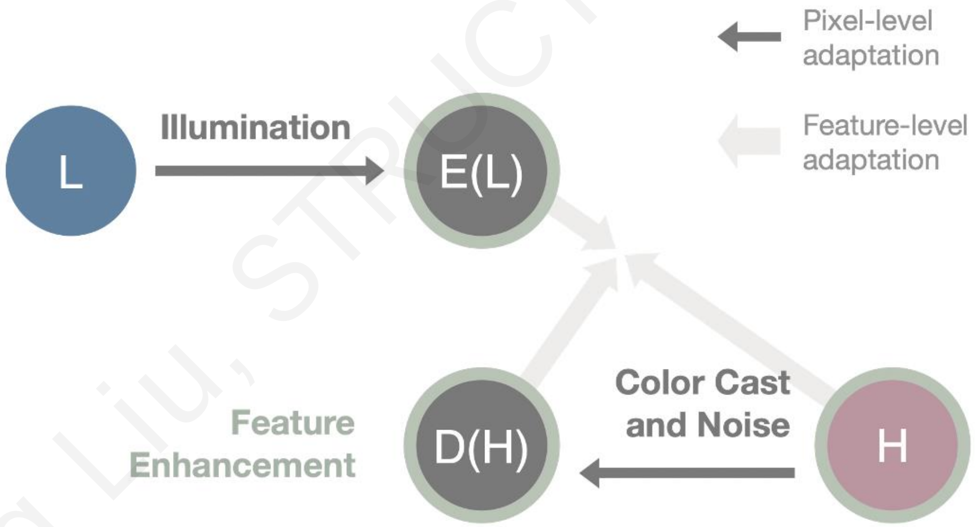
Wenjing Wang, *et al.*: Unsupervised Face Detection in the Dark. TPAMI 2023

Low Light Unsupervised Domain Adaptation

Detection
Others
Segmentation
Multiple



- Joint low-level and high-level adaptation
  - Bidirectional low-level adaptation
  - Multi-task high-level adaptation



Wenjing Wang, *et al.*: Unsupervised Face Detection in the Dark. TPAMI 2023

Low Light Unsupervised Domain Adaptation

Detection | Others | Segmentation | Multiple

2019 IEEE ICRA ToDayGAN | 2019 ICCV GCMA | 2020 ECCV YOLO-in-the-Dark | 2021 ICCV MAET | 2021 CVPR DANNet | 2021 ICCV RNW | 2021 ICCV CIConv | 2023 TPAMI HLA-Face v2

- Joint low-level and high-level adaptation

(g) EnlightenGAN † (h) EnlightenGAN (i) Zero-DCE † (j) Zero-DCE (k) MF (l) Ours

Wenjing Wang, *et al.*: Unsupervised Face Detection in the Dark. TPAMI 2023

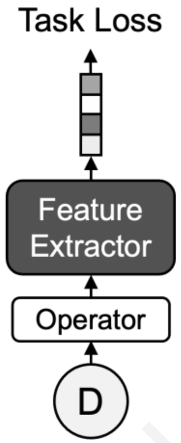
Low Light Unsupervised Domain Adaptation

Detection · Others · Segmentation · Multiple

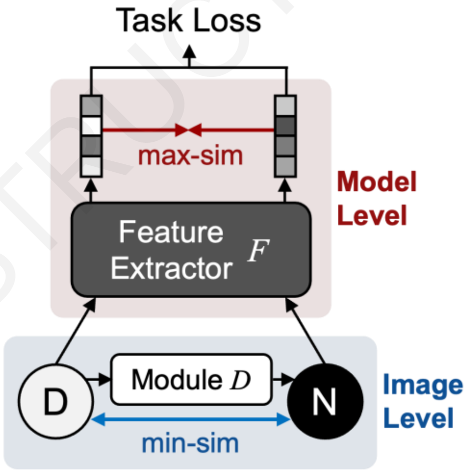
2019 IEEE ICRA ToDayGAN	2019 ICCV GCMA	2020 ECCV YOLO-in-the-Dark	2021 ICCV MAET	2021 CVPR DANNet	2021 ICCV RNW	2021 ICCV CIConv	2023 TPAMI HLA-Face v2	2023 ICCV Sim Min-Max

Task: various vision tasks, **zero-shot**

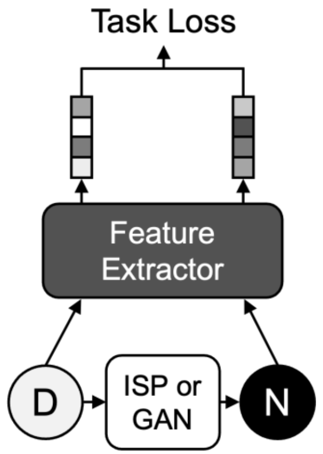
- Feature-level: maximize similarity
- Image-level: minimize similarity



Operator-based
Example: CIConv



Sim Min-Max



Darkening-based
Example: MAET

Rundong Luo, *et al*.: Similarity Min-Max: Zero-Shot Day-Night Domain Adaptation. ICCV 2023

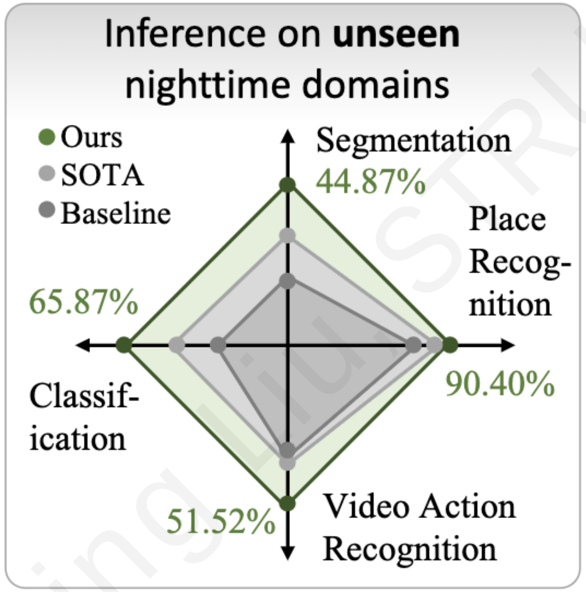
Low Light Unsupervised Domain Adaptation

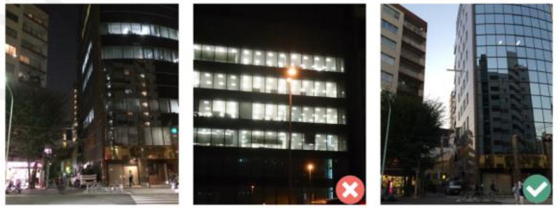
Detection · Others · Segmentation · Multiple

2019 IEEE ICRA ToDayGAN	2019 ICCV GCMA	2020 ECCV YOLO-in-the-Dark	2021 ICCV MAET	2021 CVPR DANNet	2021 ICCV RNW	2021 ICCV CIConv	2023 TPAMI HLA-Face v2	2023 ICCV Sim Min-Max

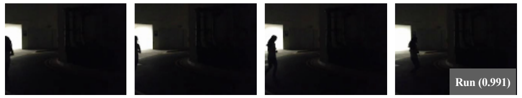
Task: various vision tasks, **zero-shot**

- Feature-level: maximize similarity
- Image-level: minimize similarity

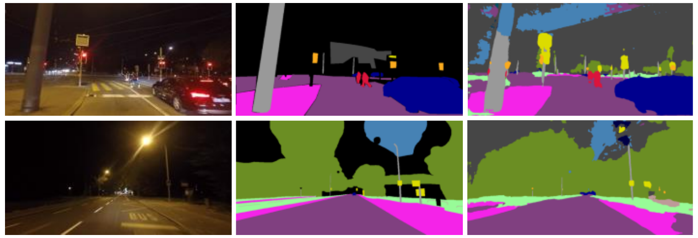




(a) Query (b) GeM (c) Ours



Ours

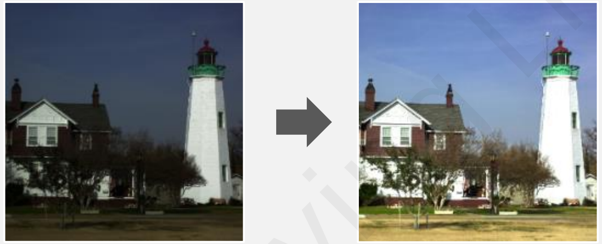


(a) Input (b) Ground truth (c) Ours

Rundong Luo, *et al*.: Similarity Min-Max: Zero-Shot Day-Night Domain Adaptation. ICCV 2023

Visual perception and understanding

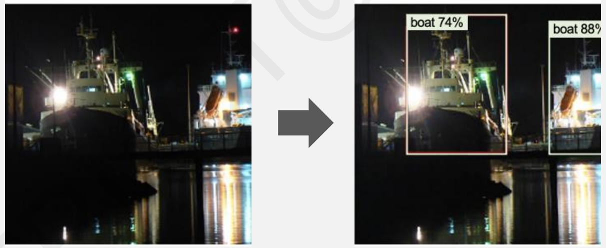
Human vision



Visual perception

Joint

Machine vision



Visual understanding

Joint visual perception and understanding

Unsupervised Illumination Adaptation for Low-Light Vision

Wenjing Wang, Rundong Luo*, Wenhan Yang, and Jiaying Liu*

***IEEE TPAMI* 2024**

Categories	Illumination Adjustment	High-Level Vision
Low-light enhancement *e.g.* RUAS (CVPR-21), Retinexformer (CVPR-23)	√	×
Domain adaptation *e.g.* CIConv (ICCV-21), Sim Min-Max (ICCV-23)	×	√
Our Target	√	√

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

Joint visual perception and understanding

- **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

Joint visual perception and understanding

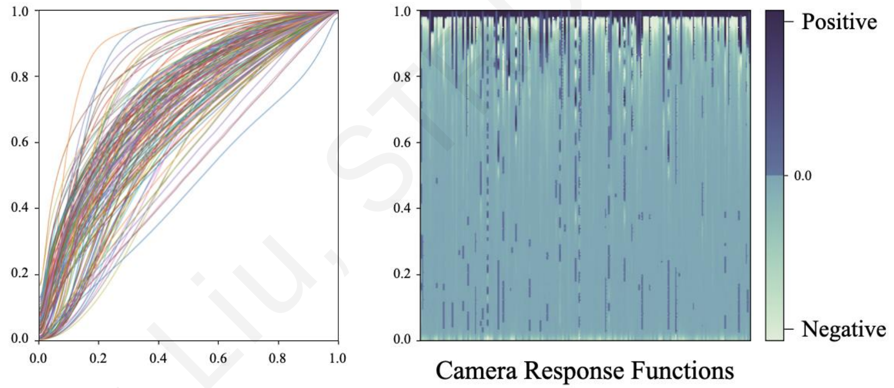
- **Network:** Deep Concave Curve

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

Most camera response functions are concave

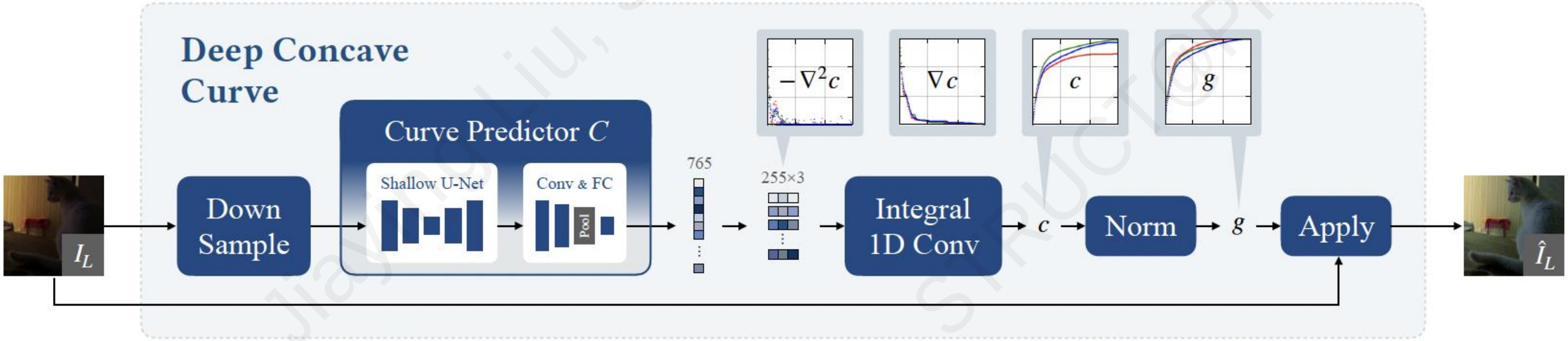


Left figure: real camera CRFs from the DoRF dataset.

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

Joint visual perception and understanding

- **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

Joint visual perception and understanding

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

Idea: employ high-level vision models as guidance

Existing strategies:

- Discrepancy metrics
- Adversarial learning

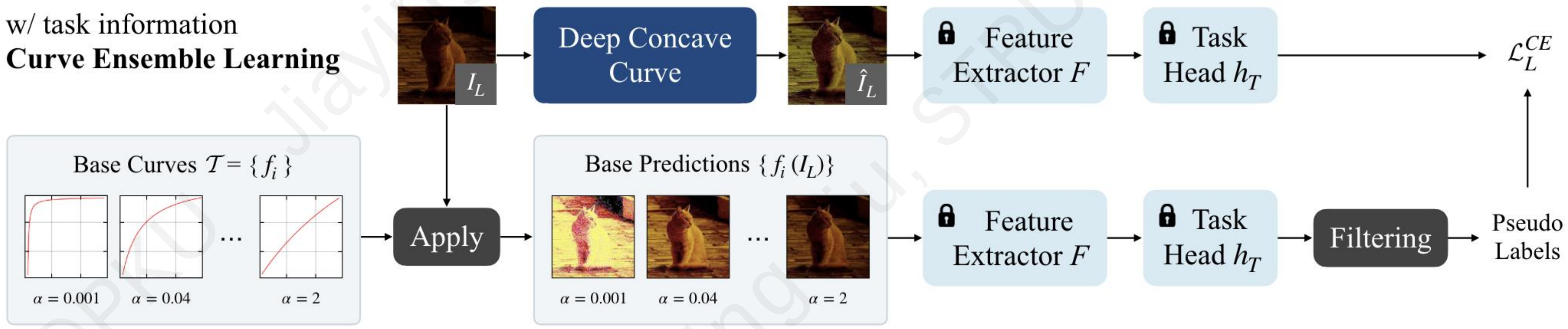
Bring extra semantic supervision, which can mislead our model and make training hard to converge

We use: **base enhancement curves or pretext tasks**

Joint visual perception and understanding

- **Training Strategy**

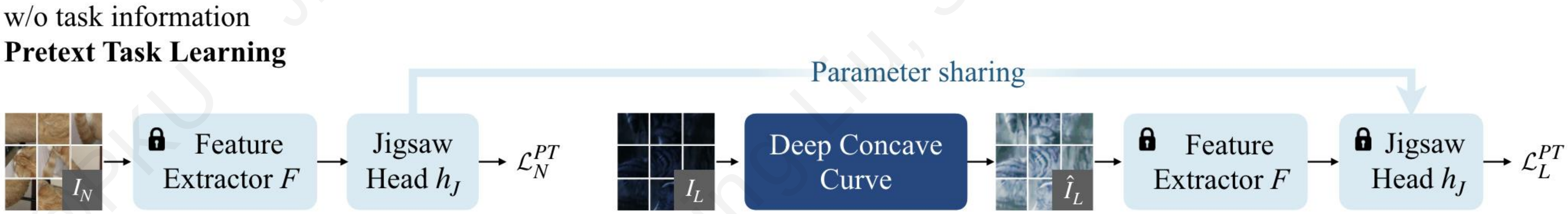
- With task information: Curve Ensemble Learning
  - Assembling weak models into a more robust model for supervision



Joint visual perception and understanding

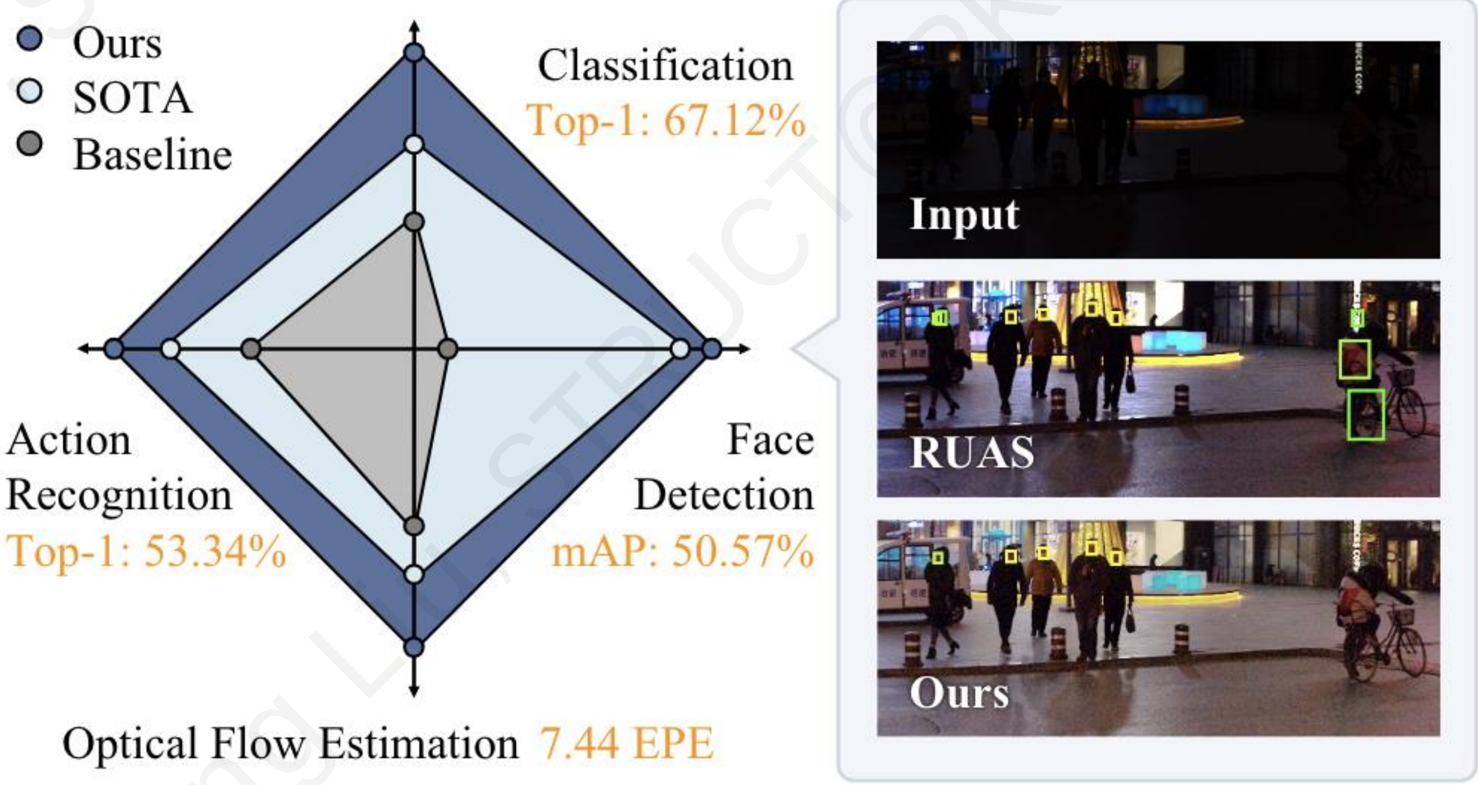
- **Training Strategy**

- With task information: Curve Ensemble Learning
- Without task information: Pretext Task Learning
  - Train a pretext head on normal light images
  - Guide enhancement model on low-light images



Joint visual perception and understanding

- **Experimental Results**
- Outperforming state-of-the-art on multiple downstream tasks



Joint visual perception and understanding

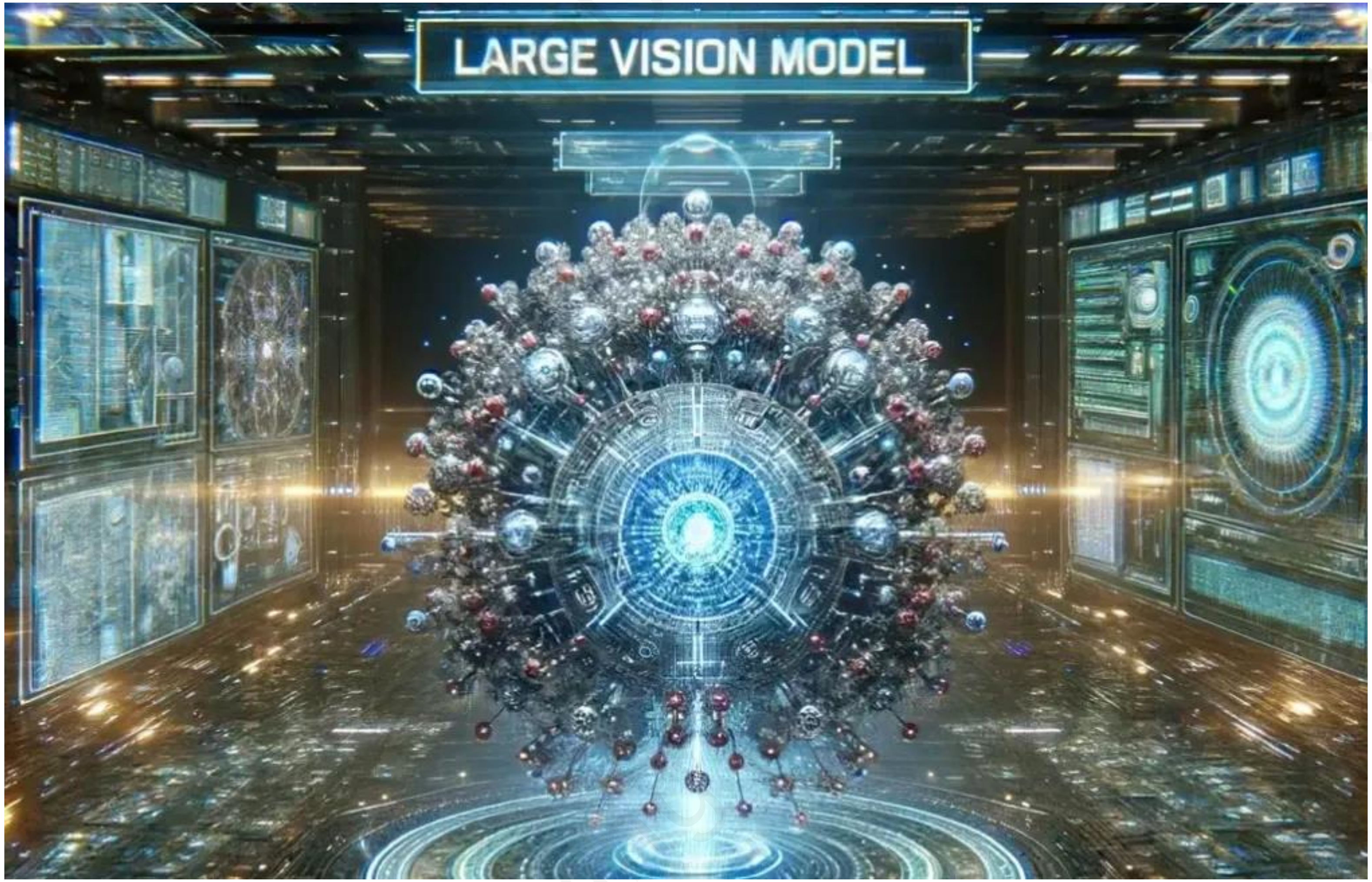
- **Experimental Results**

- Subjective Human Vision & Downstream Machine Vision

	Detection mAP (%)				NR-IQA	
	PyramidBox [80]	DSFD [64]	MogFace [81]	Average	NIQE [82]	SSEQ [83]
Original	13.99	16.09	16.36	15.48	8.5	24.1
RetinexNet [13]	11.42	12.04	14.56	12.67	7.7	21.9
KinD [14]	15.61	15.84	21.27	17.57	9.8	42.2
EnlightenGAN† [16]	19.54	20.77	24.02	21.44	9.0	22.6
EnlightenGAN [16]	28.45	31.31	35.79	31.85	9.7	14.1
Zero-DCE† [18]	33.41	37.30	37.75	36.15	6.7	18.4
LLFlow [24]	32.84	37.41	41.08	37.11	8.4	10.3
RUAS† [2]	32.77	38.36	40.71	37.28	6.2	4.4
LIME [73]	35.69	40.71	42.82	39.74	6.5	5.7
Zero-DCE++ [70]	35.56	40.90	43.45	39.97	6.4	6.7
Zero-DCE [18]	35.95	41.27	43.62	40.28	6.4	7.9
MF [72]	37.49	41.43	43.87	40.93	6.5	8.2
SACC-PT (Ours)	39.20	44.57	46.45	43.41	6.3	3.2

The correlation coefficient of NIQE/SSEQ and the average mAP of all detectors is-0.69/-0.84.

Opportunities and challenges of large vision models



Diffusion for Image Restoration

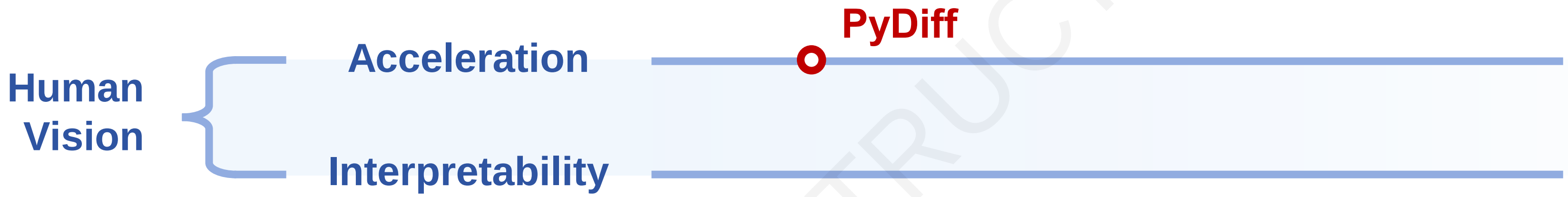
Diffusion which creates visual details with high quality can play a role.

Text:	A gray sketch on paper of a Ferrari car, full car, pencil art

Image synthesis

[1] https://prog.world/sherudim-under-the-hood-of-stable-diffusion/

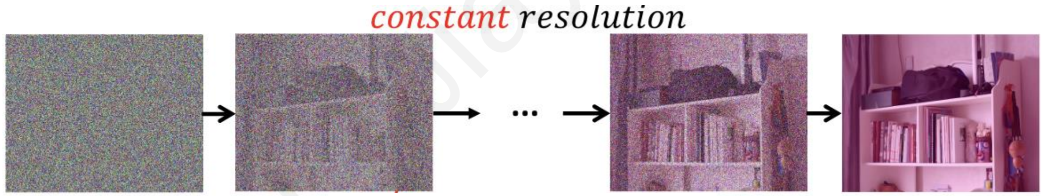
Diffusion for Low-Light Enhancement



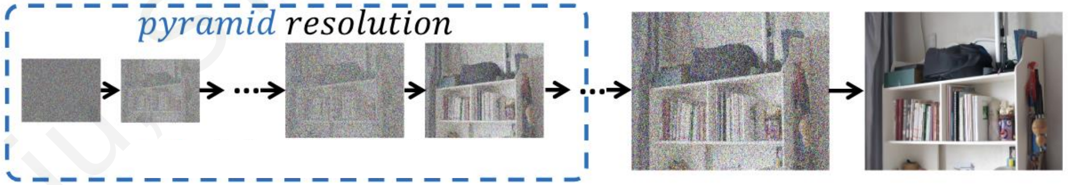
■ Pyramid Diffusion Model (PyDiff)

Problems of diffusion-based low-light enhancement

- Slow inference due to high resolution → Sampling in a **pyramid** resolution style



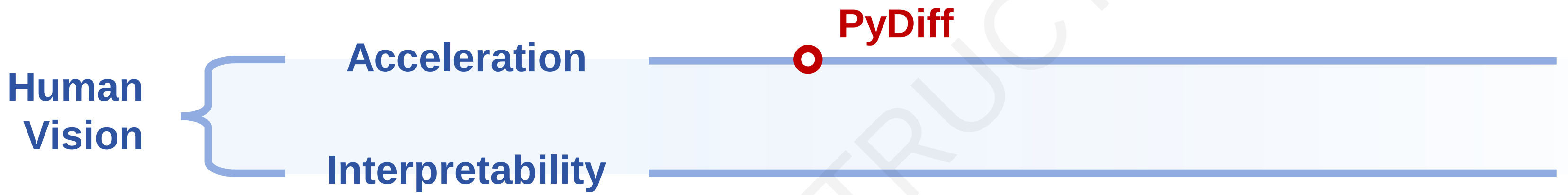
Vanilla diffusion



PyDiff

[1] Dewei Zhou, et al. Pyramid Diffusion Models For Low-light Image Enhancement, IJCAI 2023.

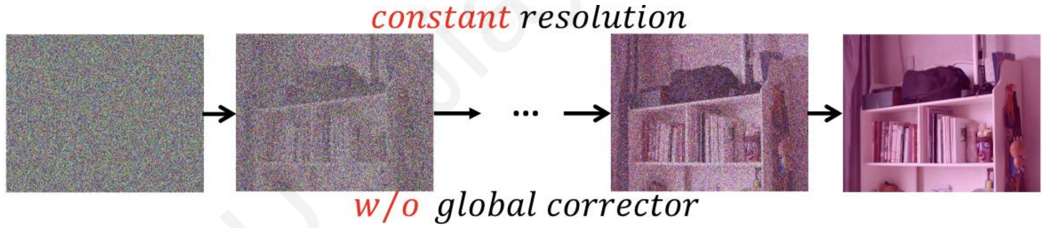
Diffusion for Low-Light Enhancement



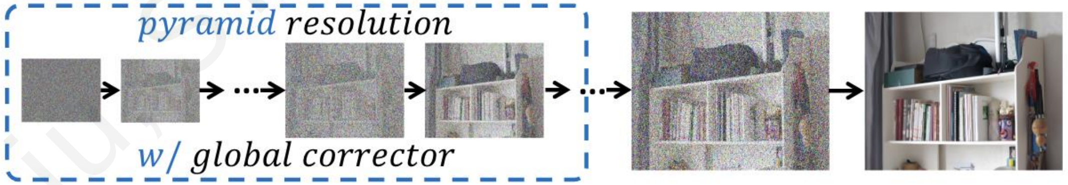
■ Pyramid Diffusion Model (PyDiff)

Problems of diffusion-based low-light enhancement

- Slow inference due to high resolution → Sampling in a **pyramid** resolution style
- Global degradation (*e.g.*, RGB shift) → **A global corrector**



Vanilla diffusion



PyDiff

[1] Dewei Zhou, et al. Pyramid Diffusion Models For Low-light Image Enhancement, IJCAI 2023.

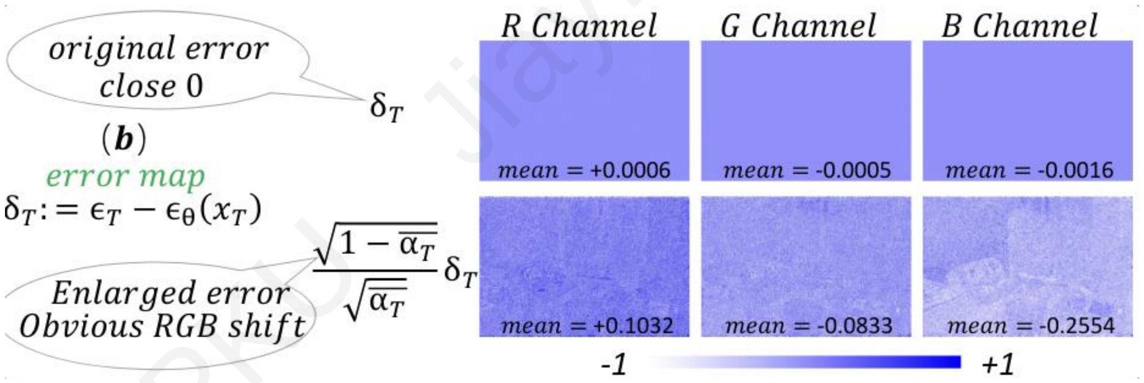
Diffusion for Low-Light Enhancement

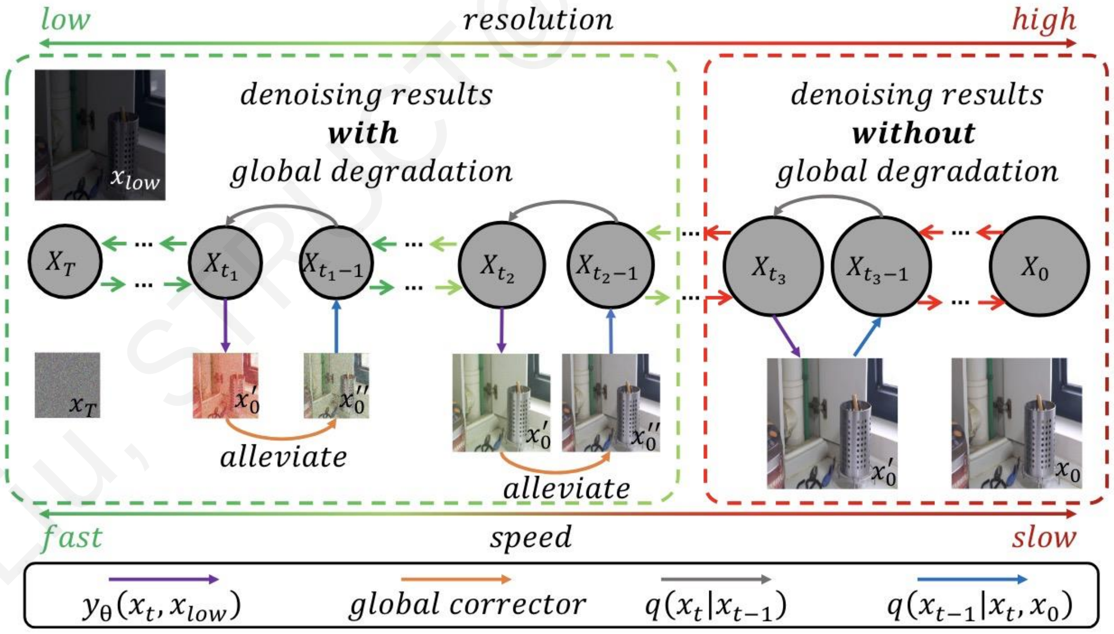
Human Vision
- Acceleration — PyDiff
- Interpretability

■ Pyramid Diffusion Model (PyDiff)

Review the forward process of diffusion: $$\mathbf{y}_\theta(\mathbf{x}_t) := \mathbf{x}_0 + \frac{\sqrt{1-\bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}}\boldsymbol{\delta}_t,$$

Caused RGB shift





[1] Dewei Zhou, et al. Pyramid Diffusion Models For Low-light Image Enhancement, IJCAI 2023.

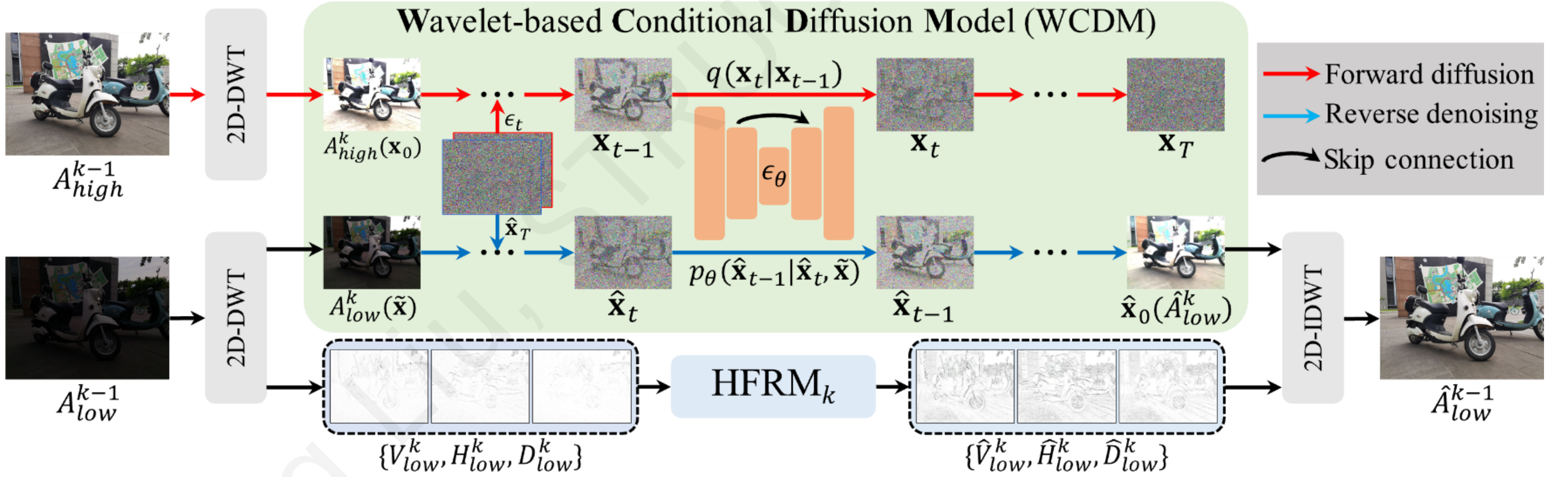
Diffusion for Low-Light Enhancement

Human Vision
- Acceleration: PyDiff, **DiffLL**
- Interpretability

■ DiffLL

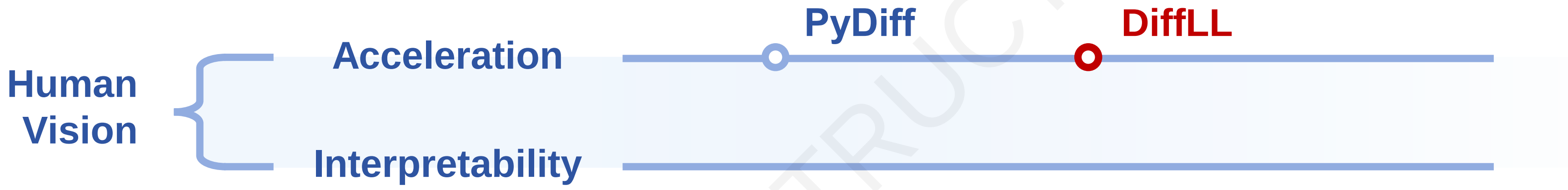
Wavelet transformation: accelerate inference

- *Diffusion-based* WCDM restores **the average coefficient** A_{low}^{K}
- *Conv-based* HFRM restores **high-frequency coefficients** $\{V_{low}^{K}, H_{low}^{K}, D_{low}^{K}\}$

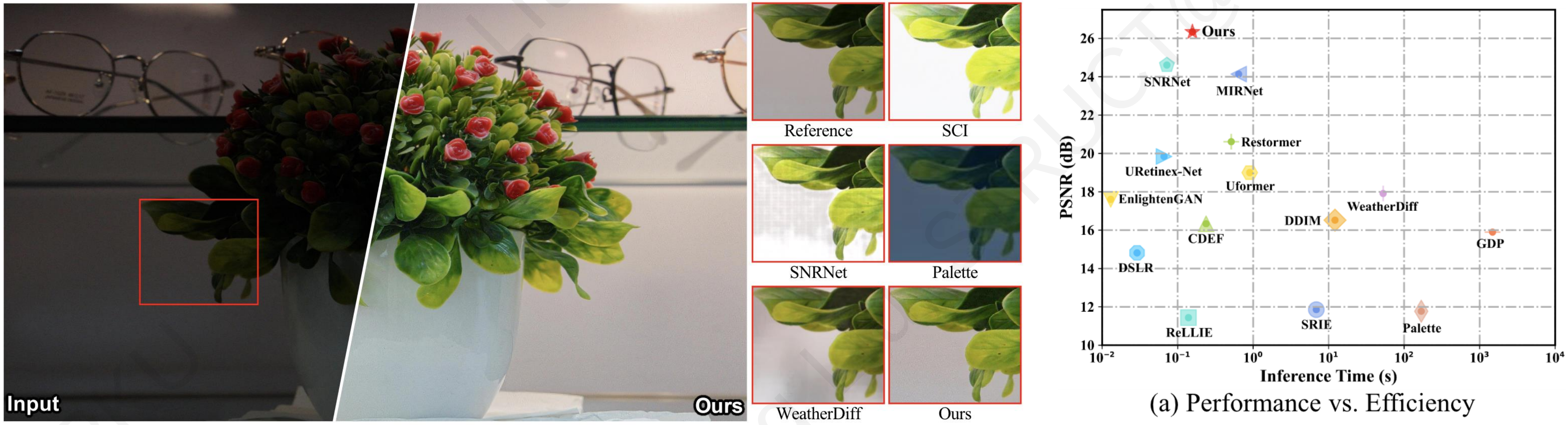


[1] Hai Jiang, et al. Low-light Image Enhancement with Wavelet-based Diffusion Models, TOG 2023.

Diffusion for Low-Light Enhancement



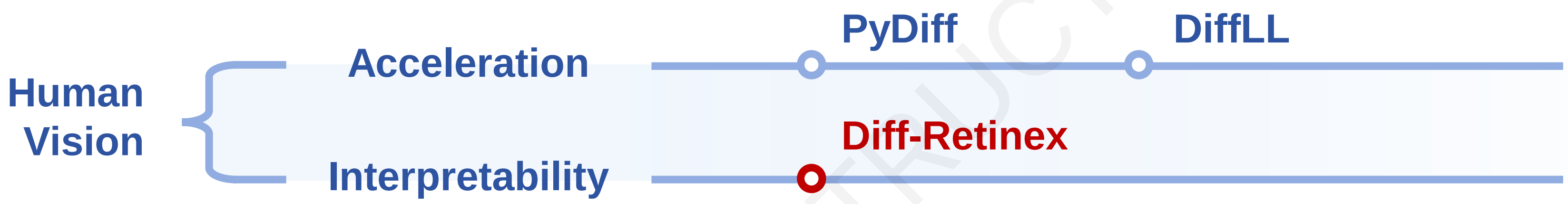
- DiffLL



(a) Performance vs. Efficiency

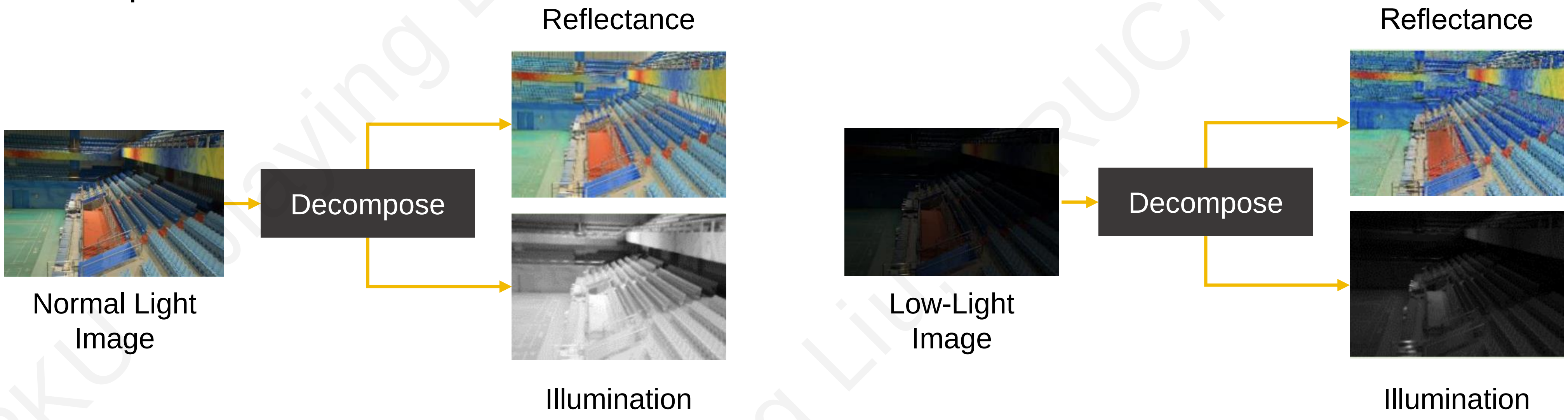
[1] Hai Jiang, et al. Low-light Image Enhancement with Wavelet-based Diffusion Models, TOG 2023.

Diffusion for Low-Light Enhancement



■ Diff-Retinex

Retinex decomposition



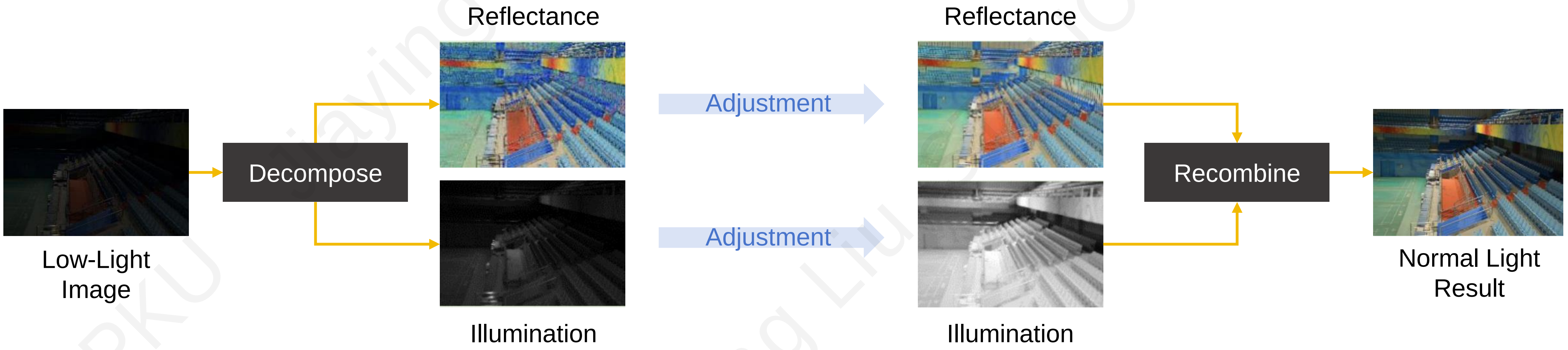
[1] Xunpeng Yi, et al. Diff-Retinex: Rethinking Low-light Image Enhancement with A Generative Diffusion Model, ICCV 2023.

Diffusion for Low-Light Enhancement

Human Vision
- **Acceleration** — **PyDiff**, **DiffLL**
- **Interpretability** — **Diff-Retinex**

■ Diff-Retinex

Retinex-based low-light enhancement



[1] Xunpeng Yi, et al. Diff-Retinex: Rethinking Low-light Image Enhancement with A Generative Diffusion Model, ICCV 2023.

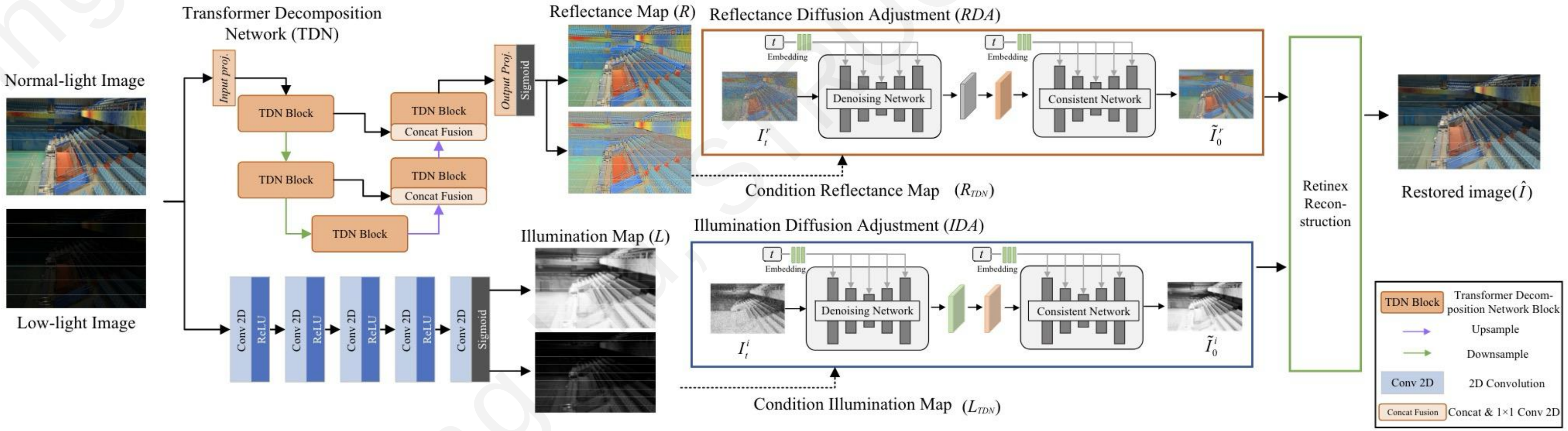
Diffusion for Low-Light Enhancement

Human Vision
- Acceleration: PyDiff, DiffLL
- Interpretability: **Diff-Retinex**

■ Diff-Retinex

Physically explainable (**Retinex decomposition**) + Diffusion

- Diffusion for adjusting reflectance and illumination



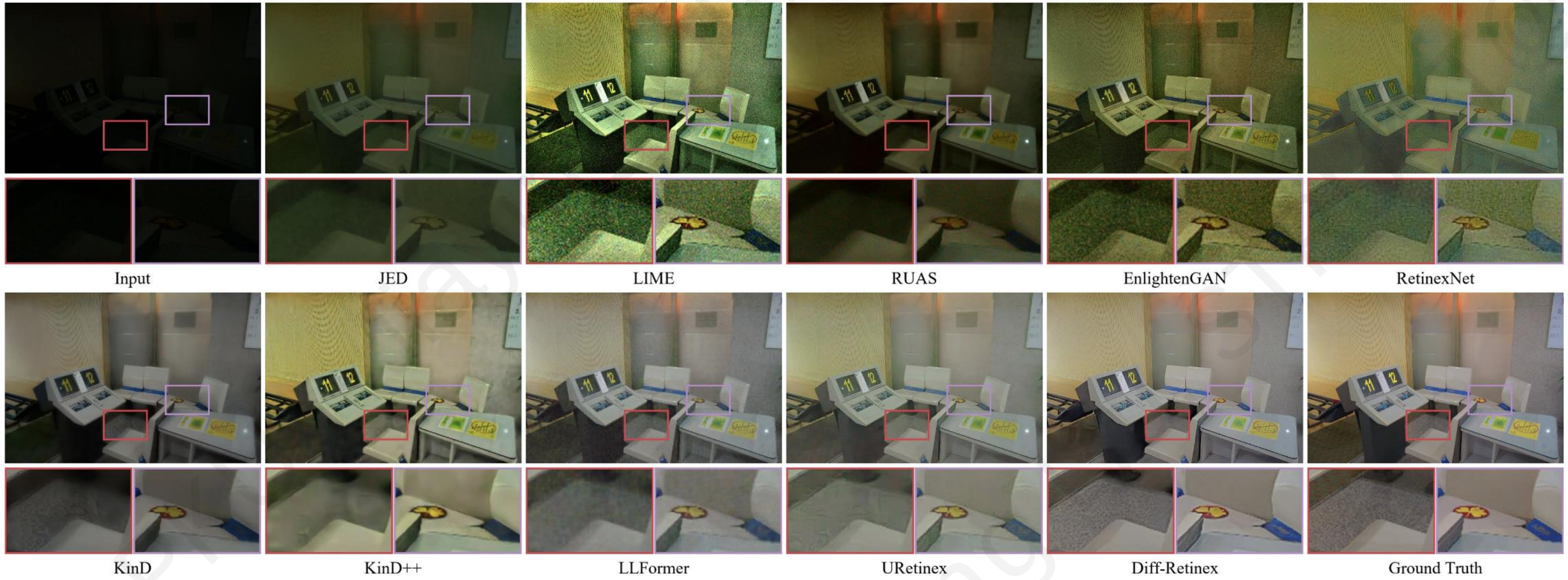
[1] Xunpeng Yi, et al. Diff-Retinex: Rethinking Low-light Image Enhancement with A Generative Diffusion Model, ICCV 2023.

Diffusion for Low-Light Enhancement

Human Vision

- Acceleration: PyDiff, DiffLL
- Interpretability: Diff-Retinex

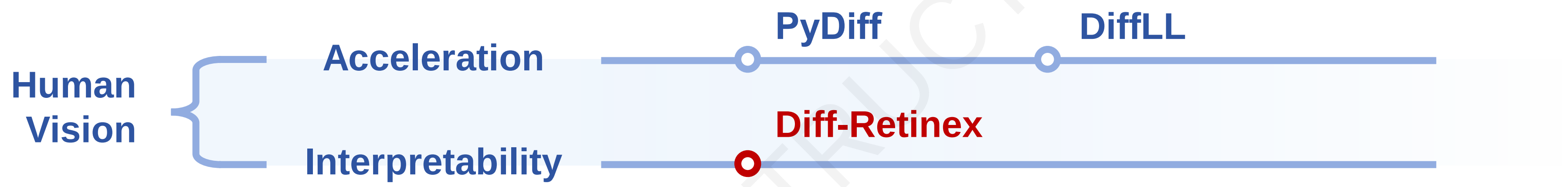
Diff-Retinex



Diff-Retinex

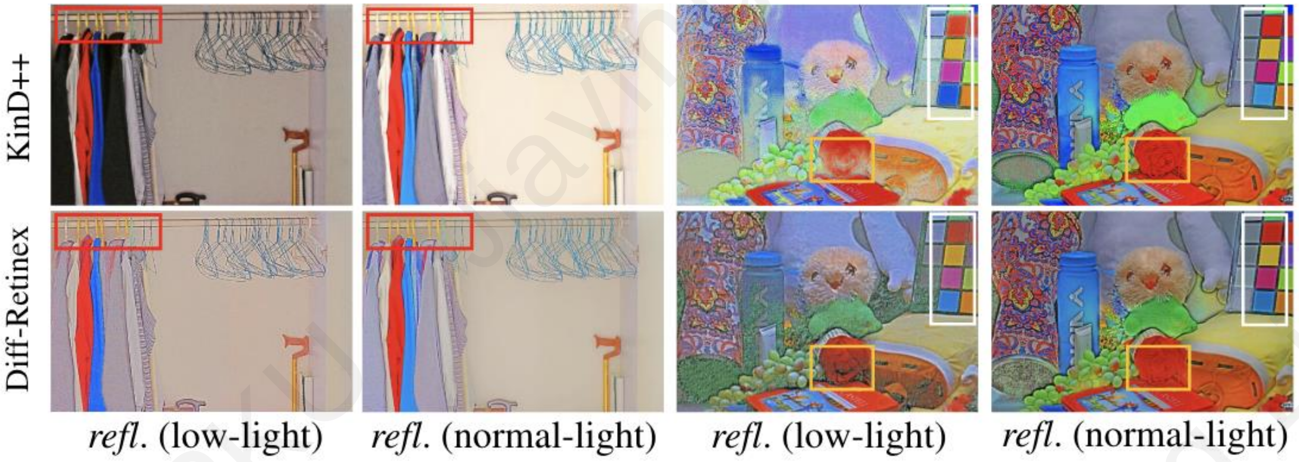
[1] Xunpeng Yi, et al. Diff-Retinex: Rethinking Low-light Image Enhancement with A Generative Diffusion Model, ICCV 2023.

Diffusion for Low-Light Enhancement

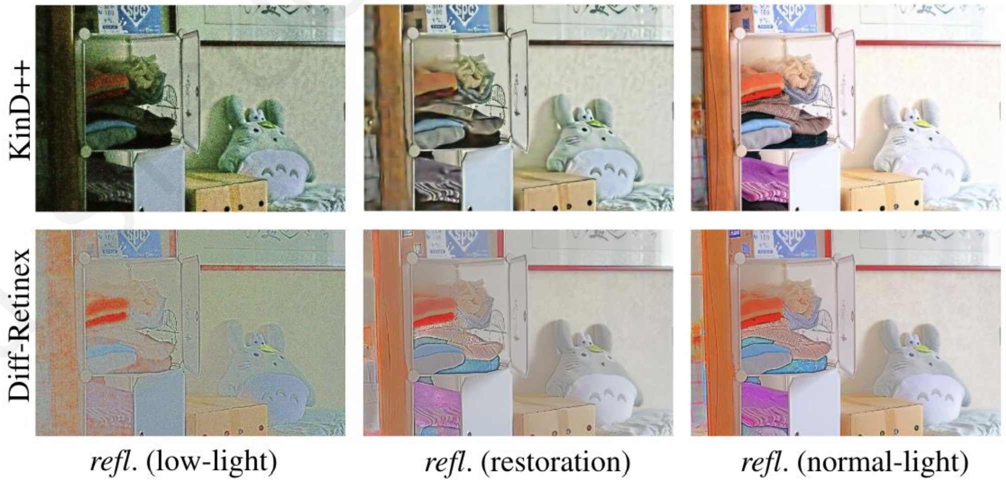


- Diff-Retinex

Results of Retinex-decomposition



Visualization of restoring the reflectance map



[1] Xunpeng Yi, et al. Diff-Retinex: Rethinking Low-light Image Enhancement with A Generative Diffusion Model, ICCV 2023.

Diffusion for Low-Light Enhancement

Human Vision

Acceleration: PyDiff, DiffLL

Interpretability: Diff-Retinex, ExposureDiffusion

■ ExposureDiffusion

Physically explainable (**exposure process**) + Diffusion

- A raw image can be formulated as

$$X_t = \underbrace{\lambda_t K I + K N_p}_{} + N_{ind}$$

$$(\lambda_t I + N_p) \sim \mathcal{P}(\lambda_t I)$$

Poisson distribution

- λ_t exposure time
- K overall system gain
- I rate of the photoelectrons
- N_p photon shot noise
- N_{ind} signal independent noise

	Conditional diffusion	Exposure diffusion
Objective	maximizing $p_\Theta(X\|Y)$	minimizing KL divergence with the real exposure process
Initial state X_T	$X_T \sim \mathcal{N}(0,1)$	$X_T \sim q(X_T)$
Assumption	$q(X_t\|X_{t-1}) := \mathcal{N}(X_t; \sqrt{1-\beta_t} X_{t-1}, \beta_t \mathbf{I})$	$q(X_{t-1}\|X_t, X_{ref}) := \mathcal{P}(\frac{X_{t-1}-X_t}{K}; \frac{(\lambda_{t-1}-\lambda_t)X_{ref}}{\lambda_{ref} K})$
Reverse process	$p_\Theta(X_{t-1}\|X_t, Y) := \mathcal{N}(X_{t-1}; \mu_\Theta(X_t, Y, t), \sigma_t^2 \mathbf{I})$	$p_\Theta(X_{t-1}\|X_t) := \mathcal{P}(\frac{X_{t-1}-X_t}{K}; \frac{(\lambda_{t-1}-\lambda_t)F_\Theta(X_t)}{\lambda_{ref} K})$
Training	The expectation over $q(X_t\|X_0)$	The expectation over $p_\Theta(X_t\|X_T)$

[1] Yufei Wang, et al. ExposureDiffusion: Learning to Expose for Low-light Image Enhancement, ICCV 2023.

Diffusion for Low-Light Enhancement

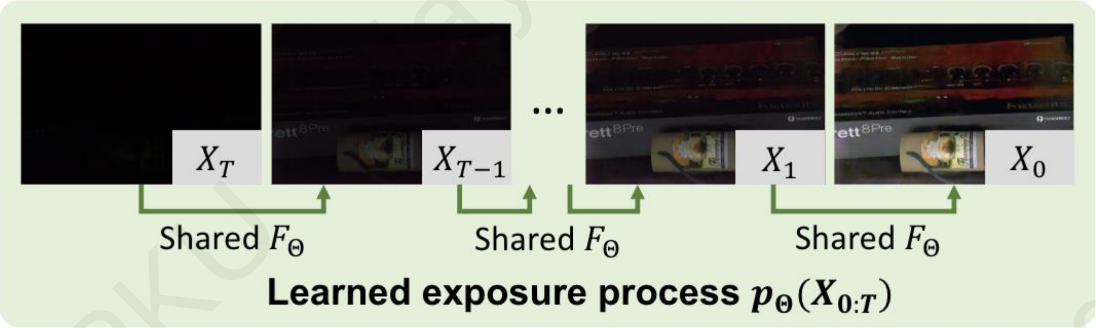
Human Vision
- Acceleration: PyDiff, DiffLL
- Interpretability: Diff-Retinex, ExposureDiffusion

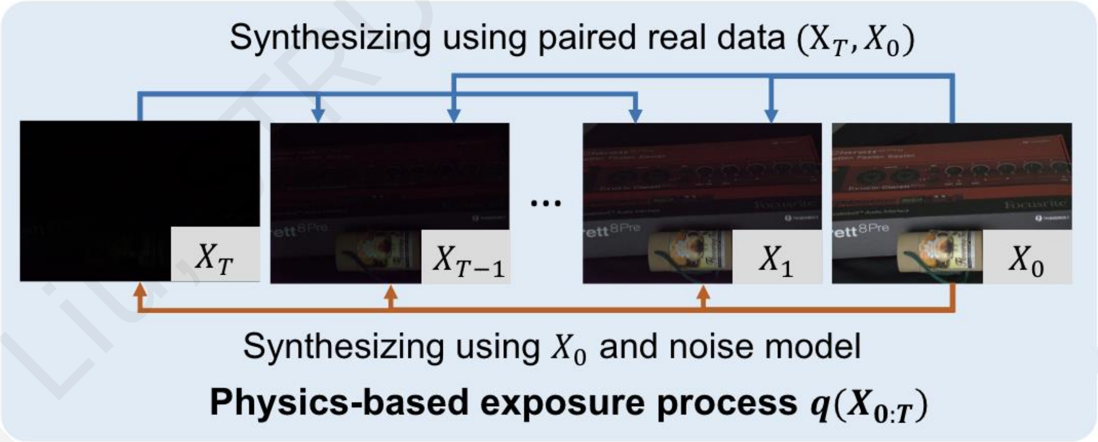
ExposureDiffusion

Physically explainable (**exposure process**) + Diffusion

- Simulate using a **shared** neural network
- In a progressive manner

Applicable to both real-captured and synthetic noise models





[1] Yufei Wang, et al. ExposureDiffusion: Learning to Expose for Low-light Image Enhancement, ICCV 2023.

Diffusion for Low-Light Enhancement

Human Vision
- Acceleration: PyDiff, DiffLL
- Interpretability: Diff-Retinex, ExposureDiffusion

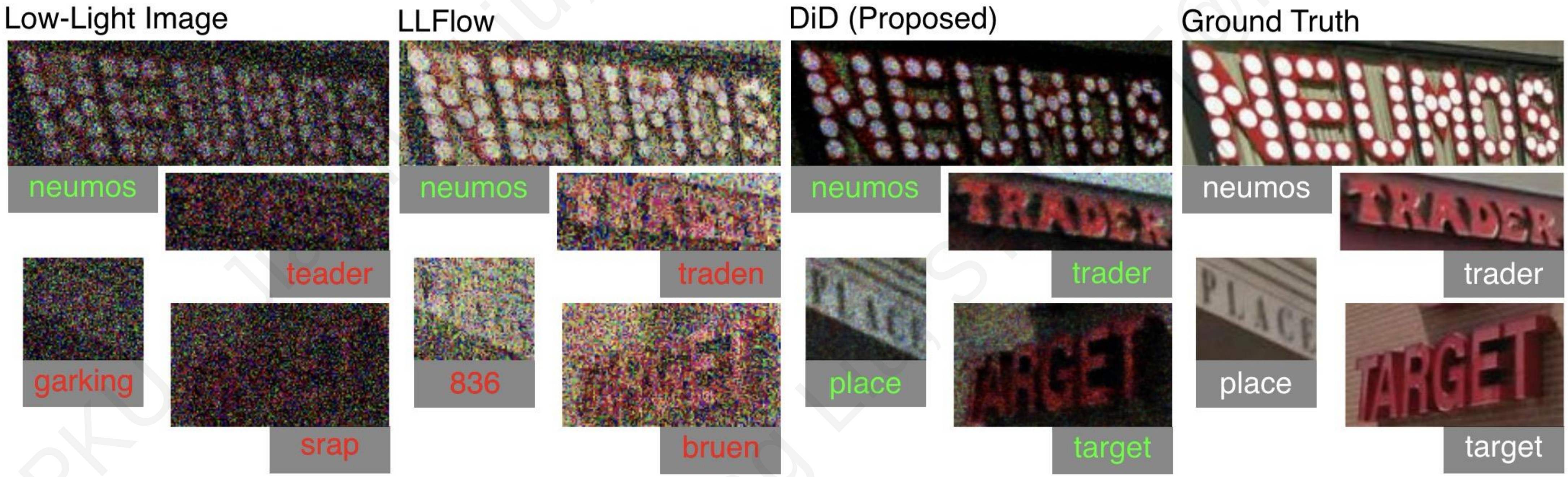
■ ExposureDiffusion

Input | BM3D | N2N | P+G | Ours | Reference

[1] Yufei Wang, et al. ExposureDiffusion: Learning to Expose for Low-light Image Enhancement, ICCV 2023.

Diffusion for Low-Light Enhancement

- Diffusion in the Dark (DiD) for low-light text recognition

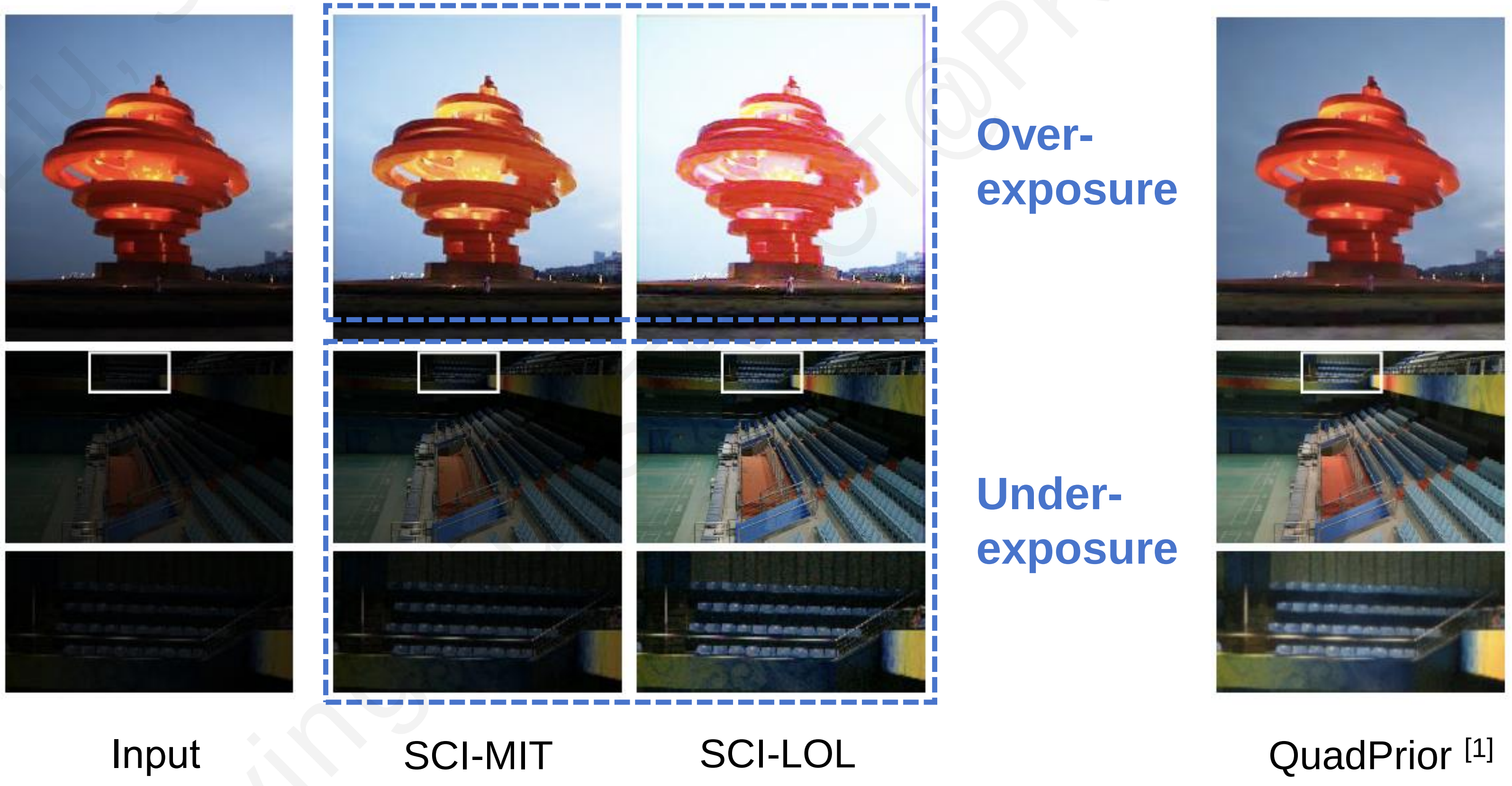


[1] Cindy M. Nguyen, et al. Diffusion in the Dark: A Diffusion Model for Low-Light Text Recognition, WACV 2024.

Diffusion for Low-Light Enhancement

Zero-Reference — QuadPrior

- Physical
  Quadruple
  Priors



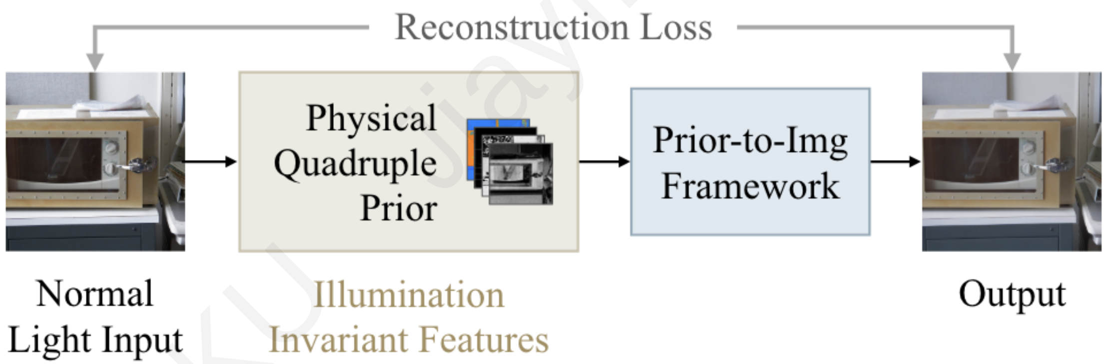
[1] Wenjing Wang, et al. Zero-Reference Low-Light Enhancement via Physical Quadruple Priors. CVPR 2024.

Diffusion for Low-Light Enhancement

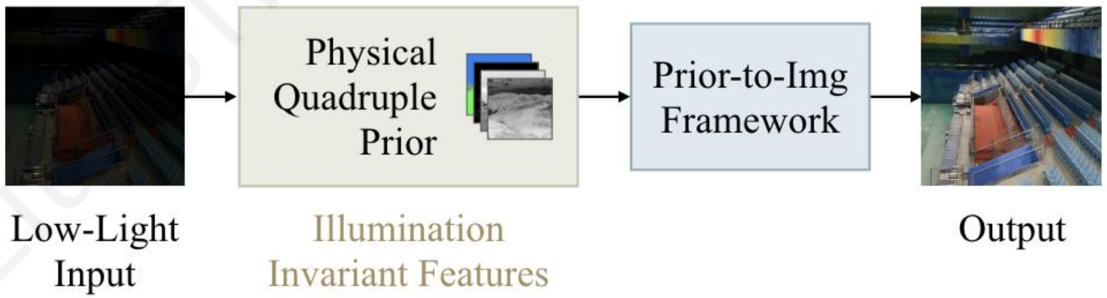
Zero-Reference QuadPrior

■ Physical Quadruple Priors

- Develop an illumination-invariant prior (Physical Quadruple Prior)
- Employ it as an intermediary between low-light and normal light images



Training on normal light images



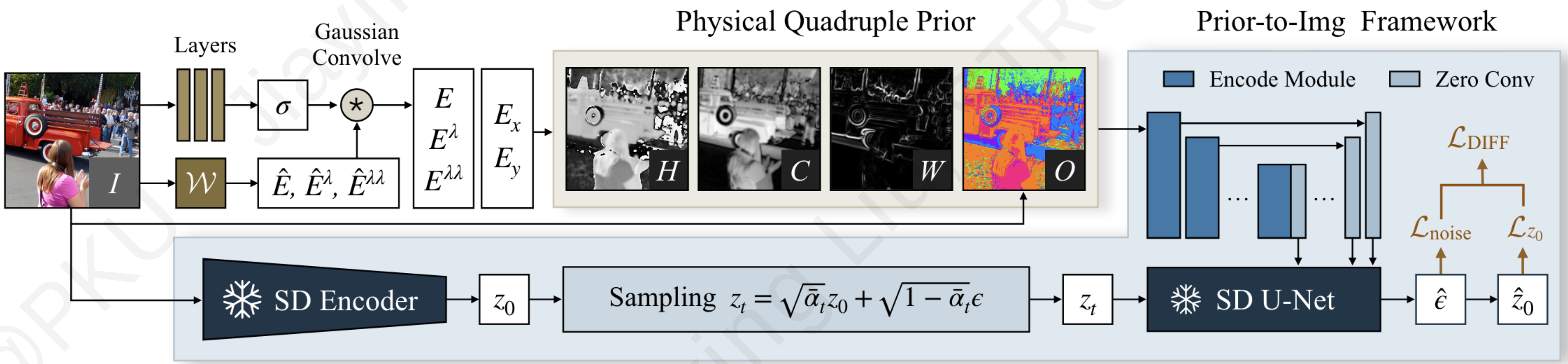
Inference on low-light images

Diffusion for Low-Light Enhancement

Zero-Reference

QuadPrior

- Physical Quadruple Priors
  - Leverage the knowledge of a pretrained image generative model
  - Use **Physical Quadruple Prior** as the condition of generation

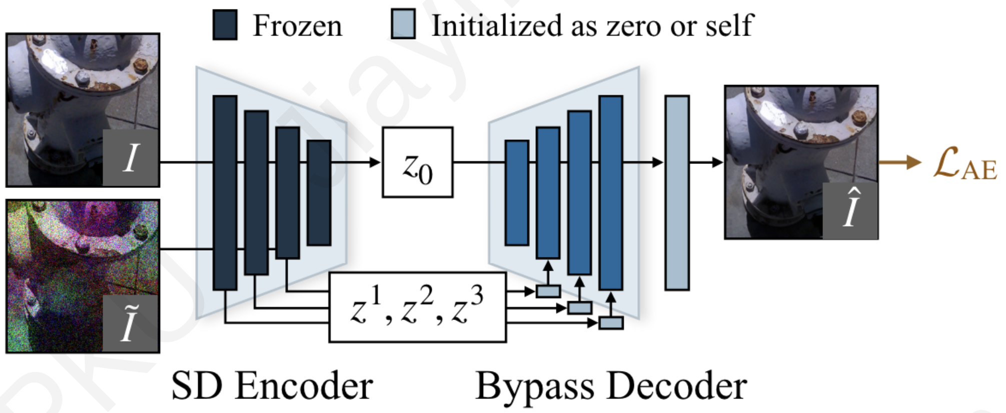


Diffusion for Low-Light Enhancement

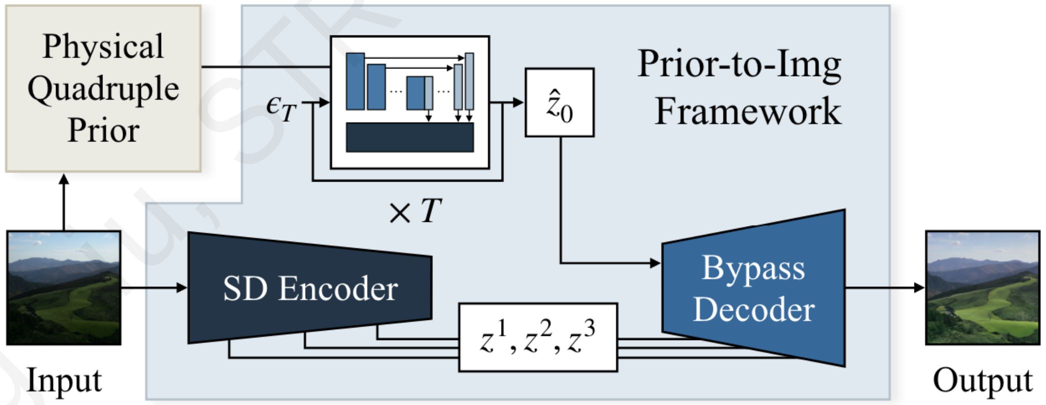
Zero-Reference — QuadPrior

■ Physical Quadruple Priors

- Improvements to make it more suitable for image restoration task
- Bypass decoder: handle detail distortion through a new deco



Training

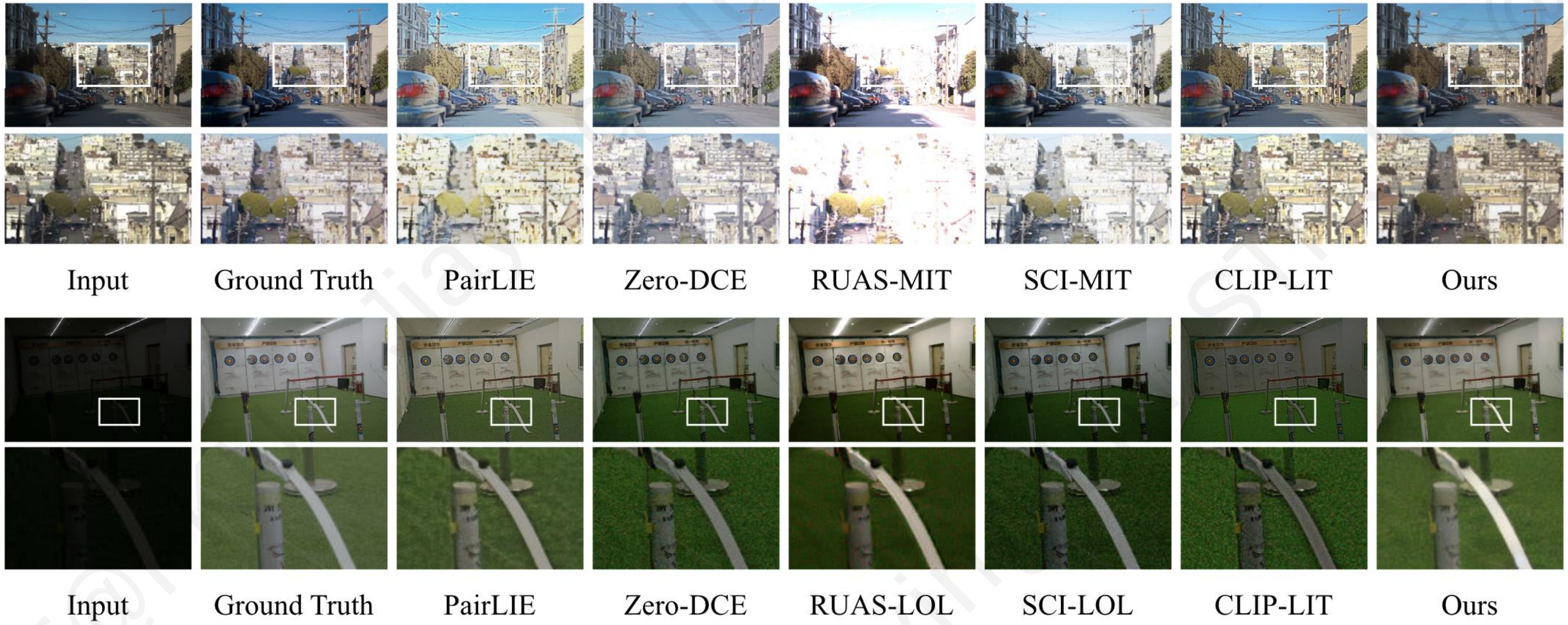


Inference

Diffusion for Low-Light Enhancement

Zero-Reference — QuadPrior

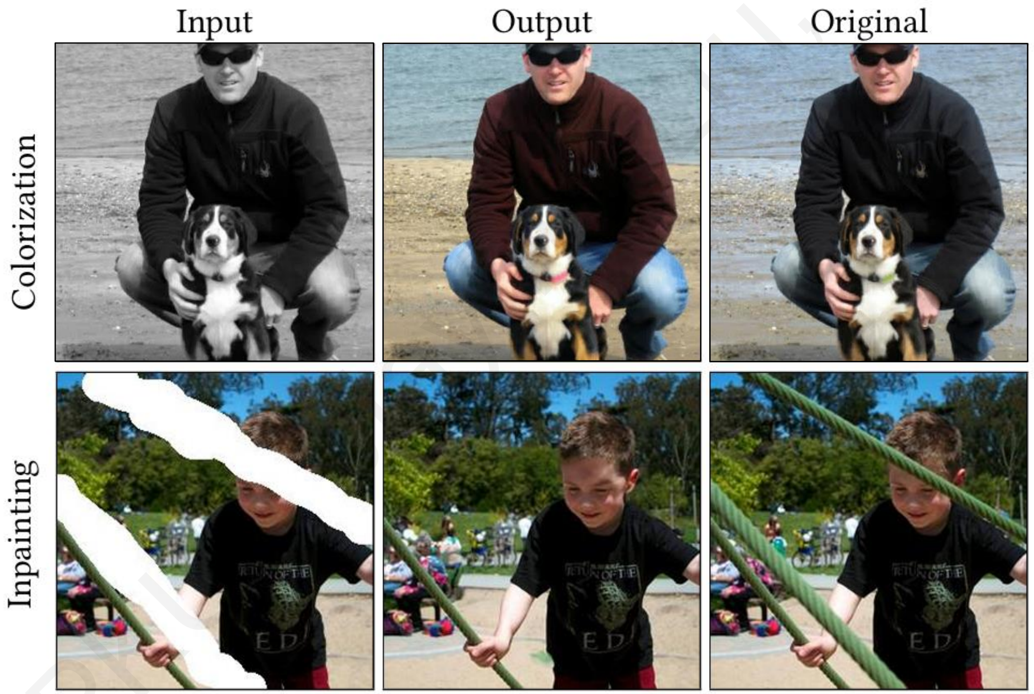
■ Physical Quadruple Priors

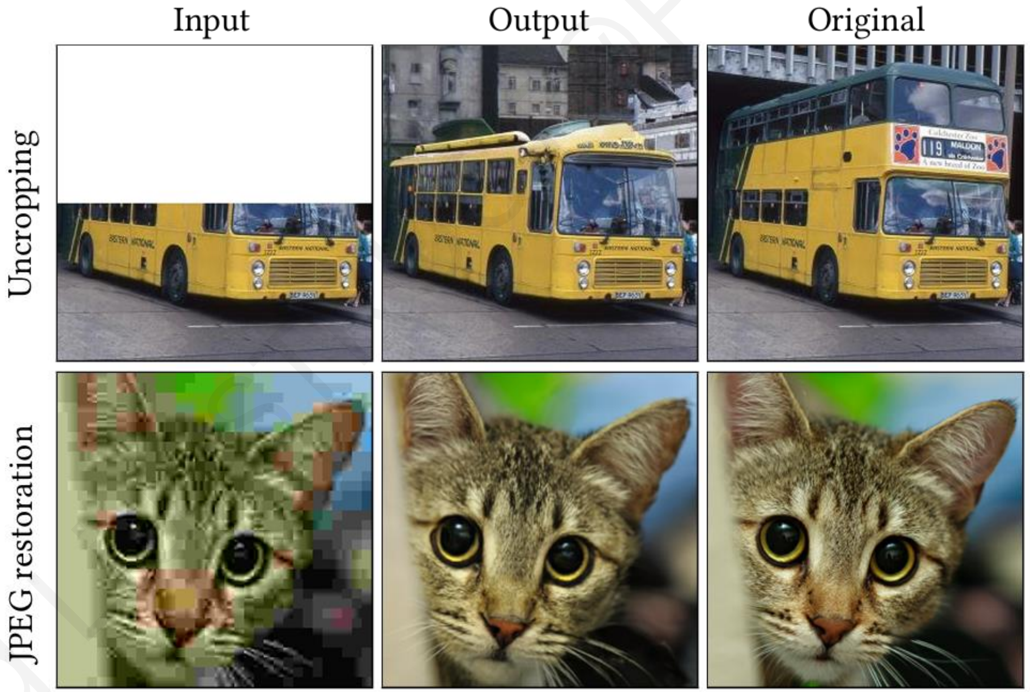


[1] Wenjing Wang, et al. Zero-Reference Low-Light Enhancement via Physical Quadruple Priors. CVPR 2024.

Unified Image Restoration and Enhancement

- Palette: Image-to-image diffusion models
  - Based on the SR3 structure: U-Net architecture
  - Different type of conditional images

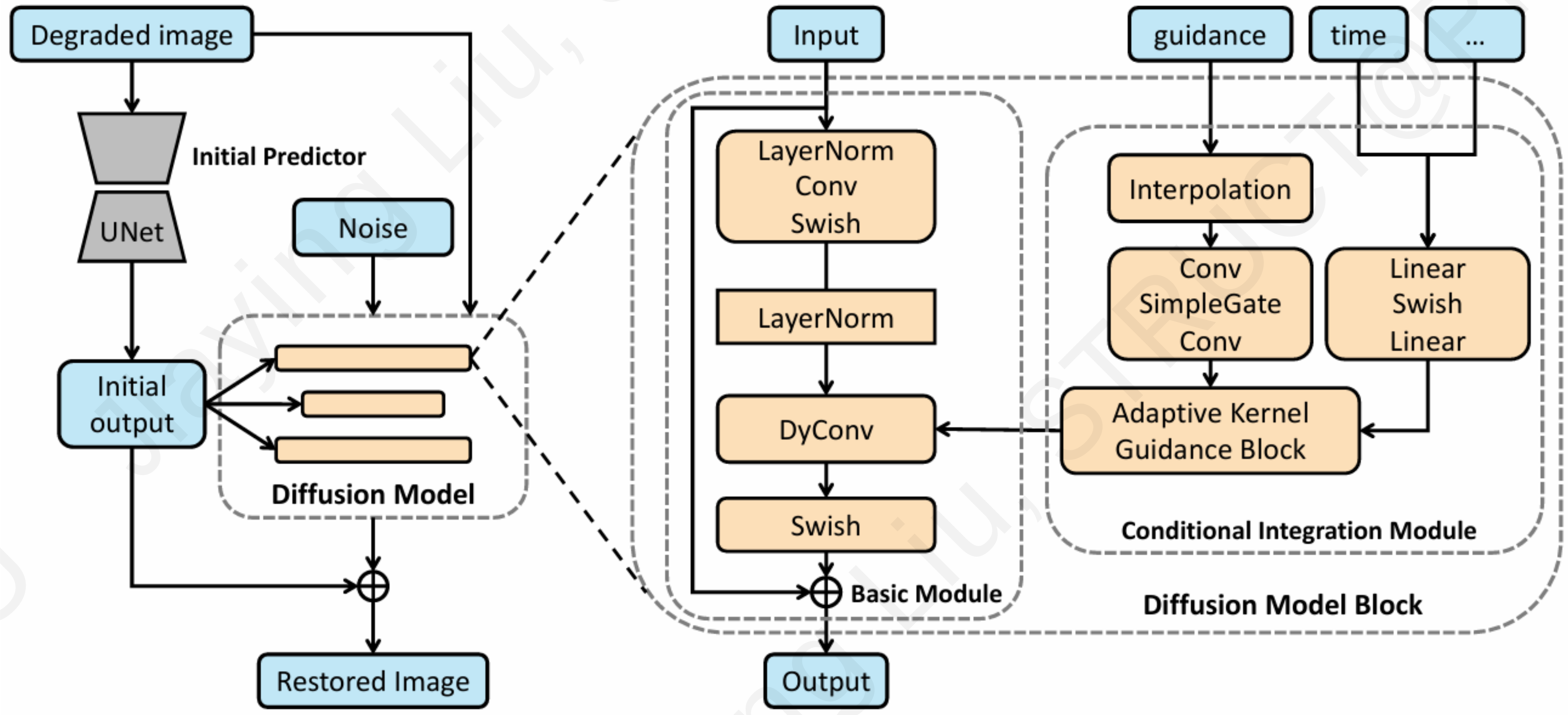




Chitwan Saharia, et al. Palette: Image-to-image diffusion models. ACM SIGGRAPH, 2022.

Unified Image Restoration and Enhancement

- A Unified Conditional Framework for Diffusion-based
  - Same model structure and training hyper-parameters
  - Take different conditional guidance and train on different datasets



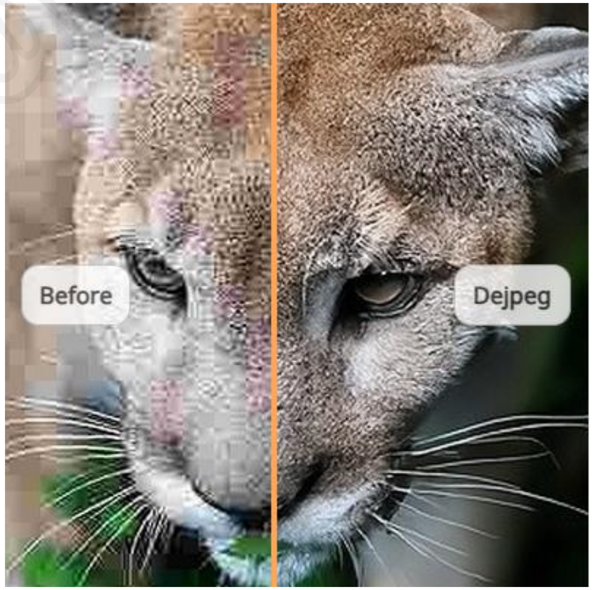
Yi Zhang, et al. A unified conditional framework for diffusion-based image restoration. NIPS, 2023

Unified Image Restoration and Enhancement

- A Unified Conditional Framework for Diffusion-based
  - Same model structure and training hyper-parameters
  - Take different conditional guidance and train on different datasets

Low-light denoise

Deblur

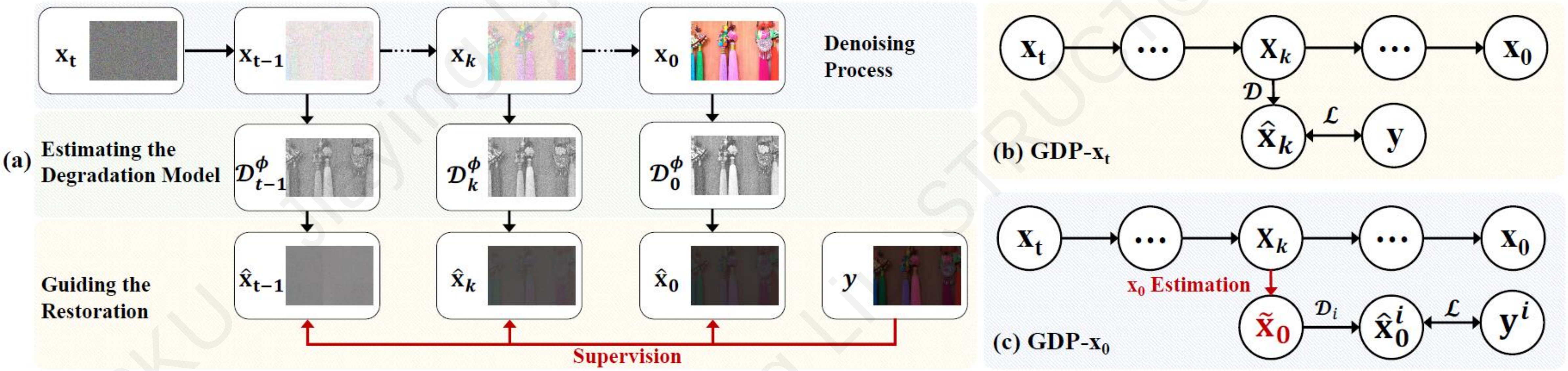


JPEG restoration

Yi Zhang, et al. A unified conditional framework for diffusion-based image restoration. NIPS, 2023

Unified Image Restoration and Enhancement

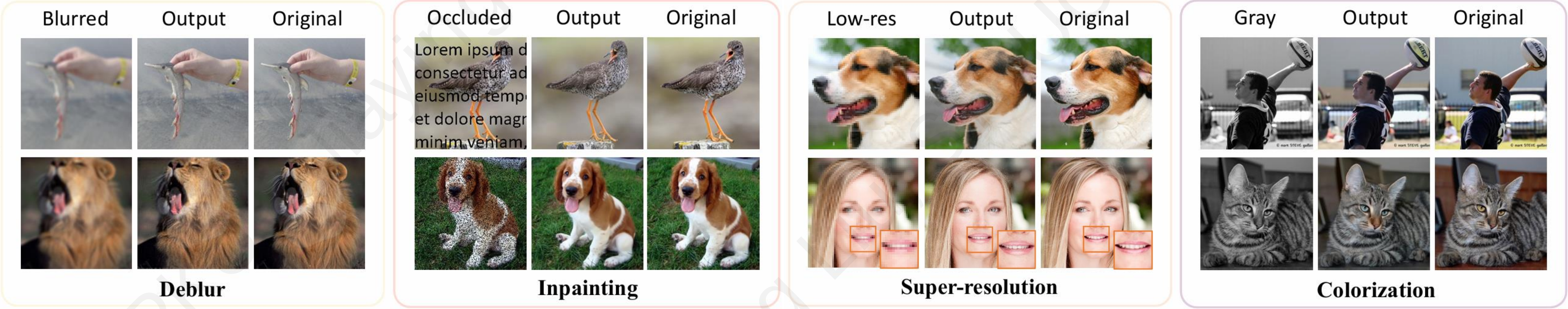
- Generative Diffusion Prior (GDP)
  - A unified framework for multiple restoration and enhancement tasks.
  - Use a pretrained unconditional image synthesis diffusion model as prior.



Ben Fei, et al. Generative Diffusion Prior for Unified Image Restoration and Enhancement, CVPR, 2023

Unified Image Restoration and Enhancement

- Generative Diffusion Prior (GDP)
  - A unified framework for multiple restoration and enhancement tasks.
  - Use a pretrained unconditional image synthesis diffusion model as prior.
  - Different degradation models learned during the sampling process.



Ben Fei, et al. Generative Diffusion Prior for Unified Image Restoration and Enhancement, CVPR, 2023

Unified Image Restoration and Enhancement

- Generative Diffusion Prior (GDP)
  - A unified framework for multiple restoration and enhancement tasks.

Low-light Image Enhancement   Brightness Control   HDR Image Recovery

Ben Fei, et al. Generative Diffusion Prior for Unified Image Restoration and Enhancement, CVPR, 2023

Summary

- Visual enhancement in degraded scenarios
- Challenges:
  - Visual unpleasure (**Humans**)
  - System failure (**Machines**)
  - Jointly considering human and machine vision

谢　谢！

CSIG MM专委公众号

PKU STRUCT小组网页