Dark-ISP: Enhancing RAW Image Processing for Low-Light Object Detection

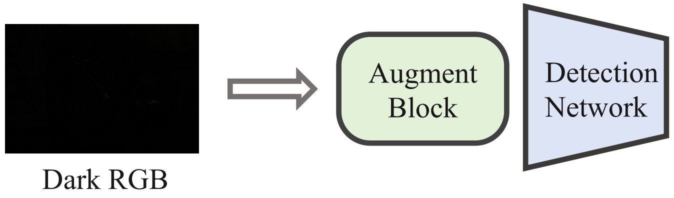
Jiasheng Guo, Xin Gao, Yuxiang Yan, Guanghao Li, and Jian Pu

Institute of Science and Technology for Brain-inspired Intelligence, Fudan University

ICCV HONOLULU HAWAII OCT 19-23, 2025

Background

Traditional RGB Methods



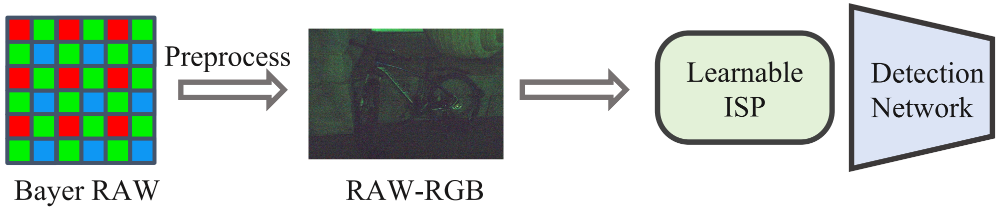
Characteristic:

- Low bit-depth RGB images from compression

Deficiency:

- Limited information for low bit-depth
- Noise introduced from ISP

RGB-format RAW Methods



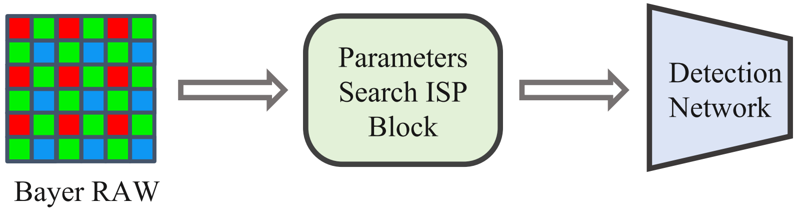
Characteristic:

- Quantized three-channel RGB format RAW
- Learnable enhancement module

Deficiency:

- Reduced bit-depth leads to limited benefits of RAW

Parameters Serach Methods



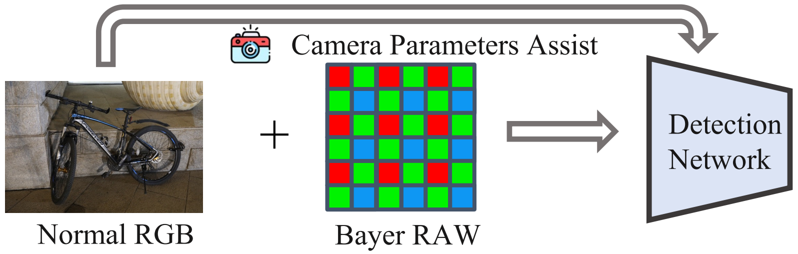
Characteristic:

- Parameters searching algorithm for ISP block

Deficiency:

- Computationally expensive cost
- Impractical for real-world deployment

Auxiliary Information Methods



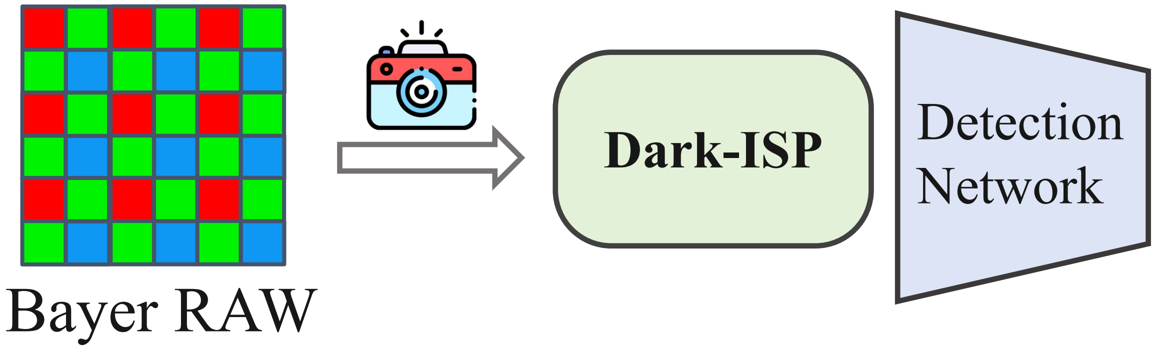
Characteristic:

- Rely on auxiliary materials to assist training

Deficiency:

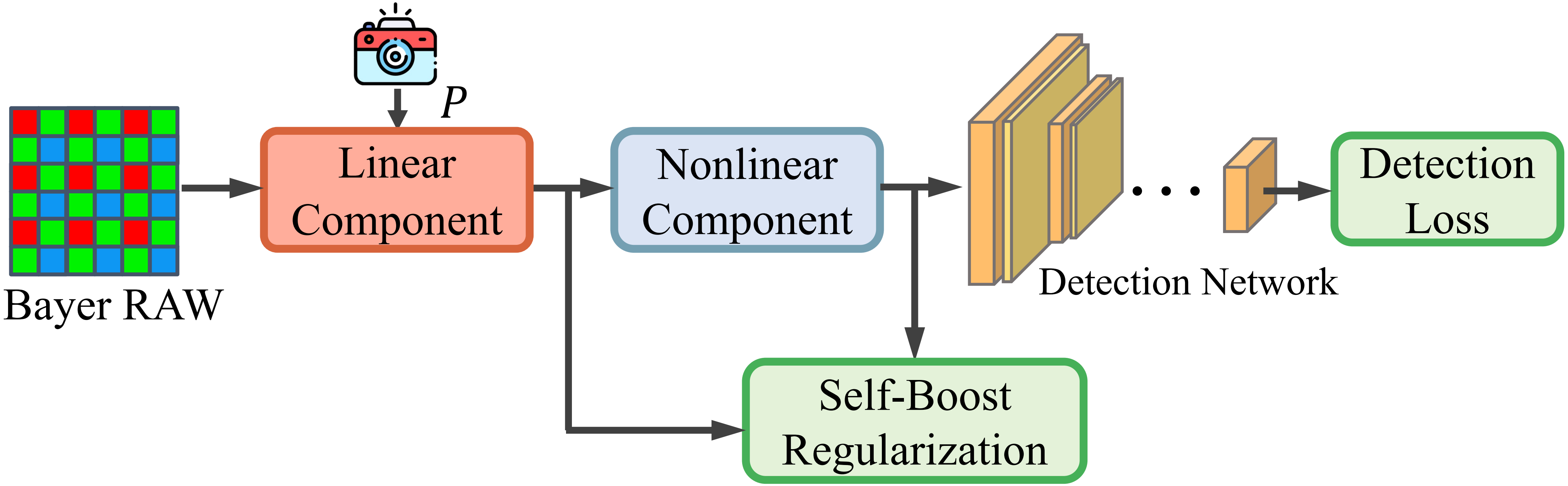
- Additional data requirements
- Overcomplicated training framework

Our Method



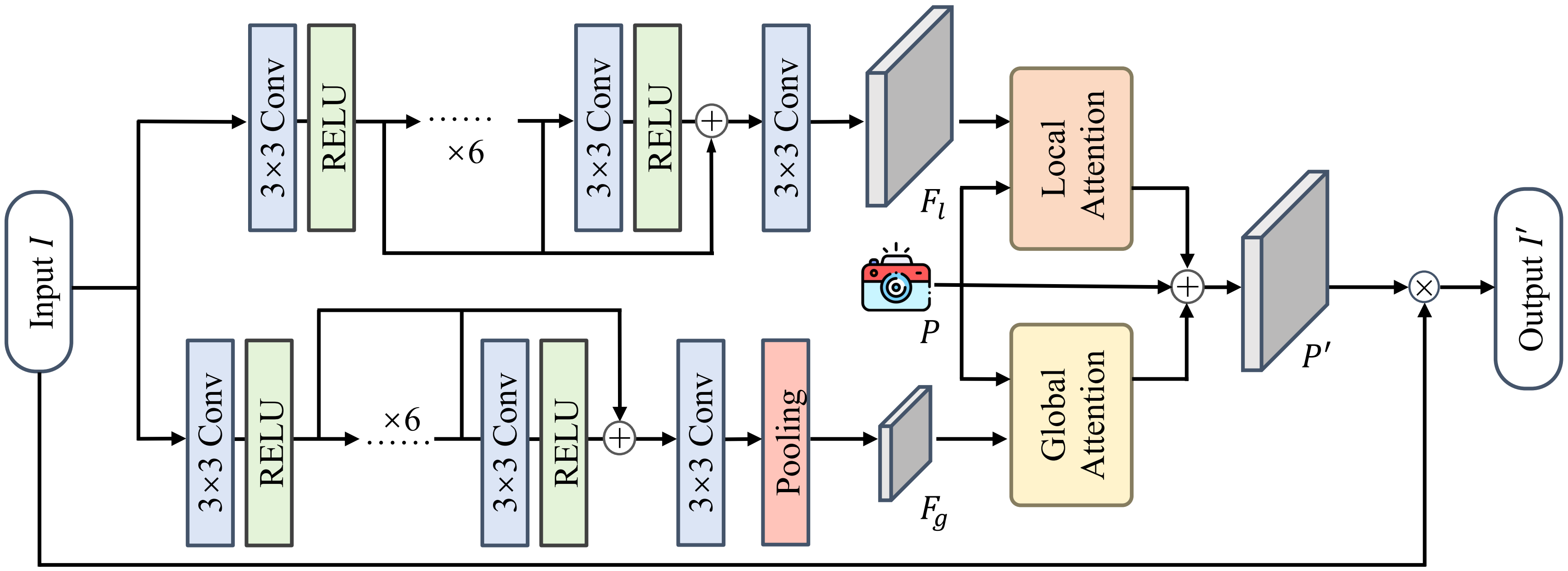
- Propose an end-to-end training framework from sensor data to the downstream task
- Design a plug-and-play lightweight Image Signal Processing plugin
- No cost for obtaining additional auxiliary data

Framework



- **Linear Component:** Incorporate camera parameters with global/local image features
- **Nonlinear Component:** Weighting physics-interpretable non-convex polynomial basis functions for image processing
- **Self-Boost Regularization:** Enhance inter-module collaboration

Linear Component



- **Joint parameter matrix** $P = C \cdot B \cdot W$ and RAW image I are used as input.
- **Global and Local Attention** mechanism integrate image information:

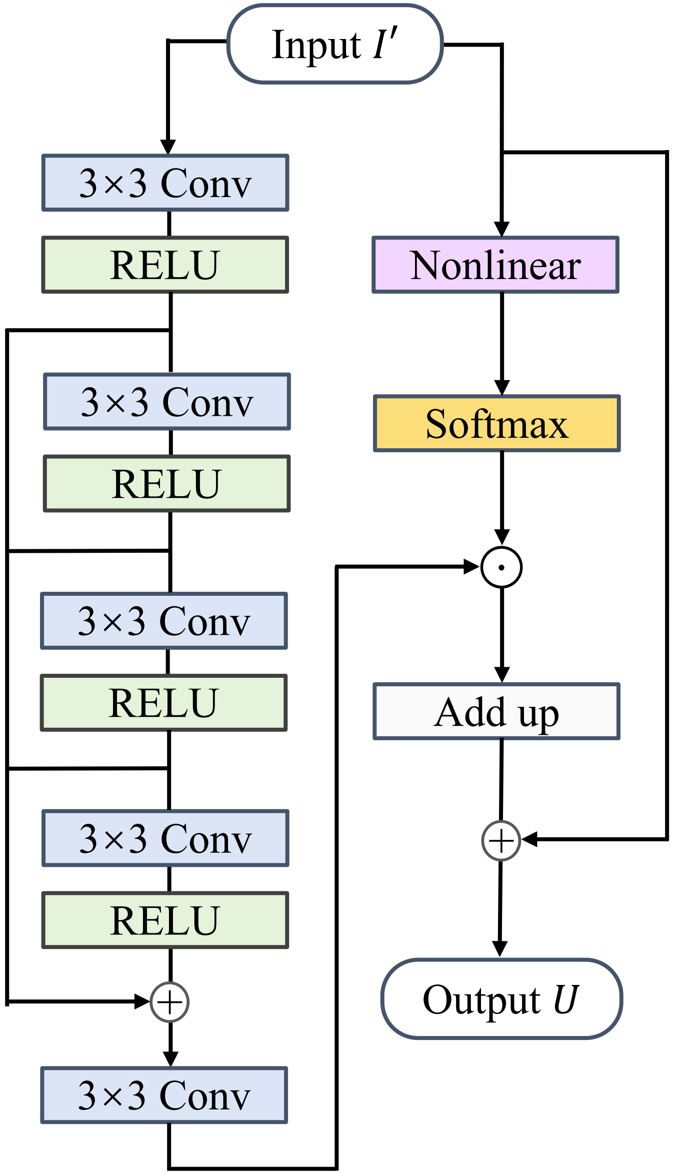
$$P_l = LocalAttn(Q = F_l, K = P, V = P),$$

$$P_g = GlobalAttn\left(Q = P, K = F_g, V = F_g\right).$$

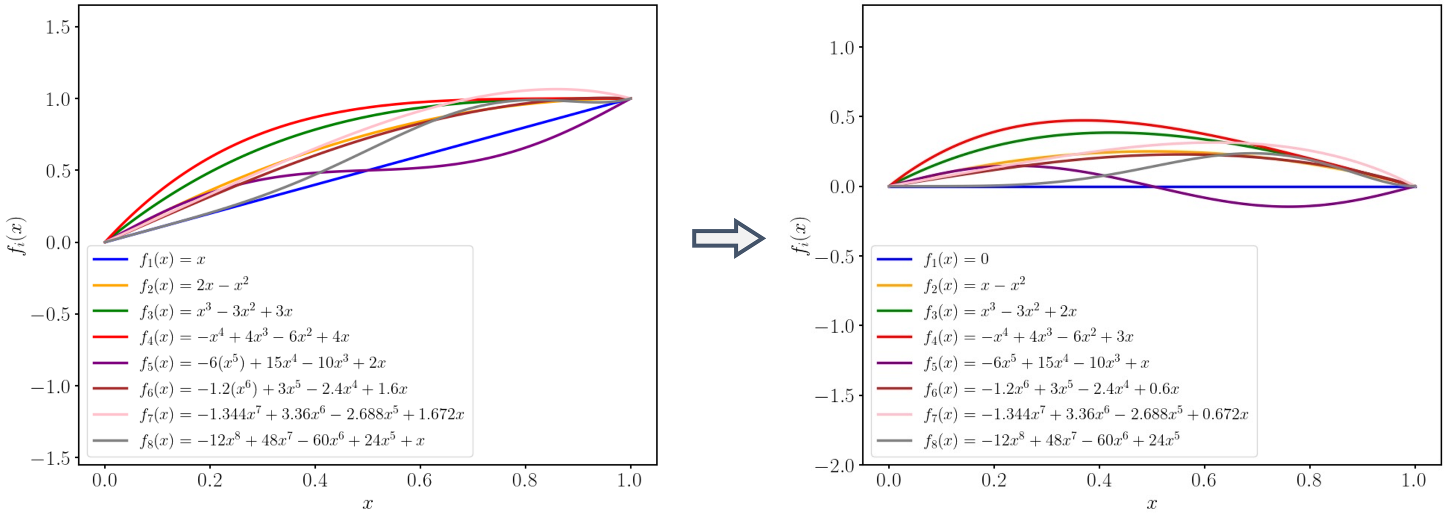
- **Enhanced parameter matrix** P' is used to process I to produce I':

$$I' = \left(P_g + P_l + P\right) \cdot I = P' \cdot I.$$

Nonlinear Component



Nonlinear Mapping Functions



- The shape of the $f_k(x)$ needs to be non-convex to stretch the dark regions while compressing the bright regions.
- Each $f_k(x)$ is a polynomial of order k that passes through the points (0, 0) and (1, 1) .

The network learns appropriate coefficients to weight the functions based on the output image I' of Linear Component.

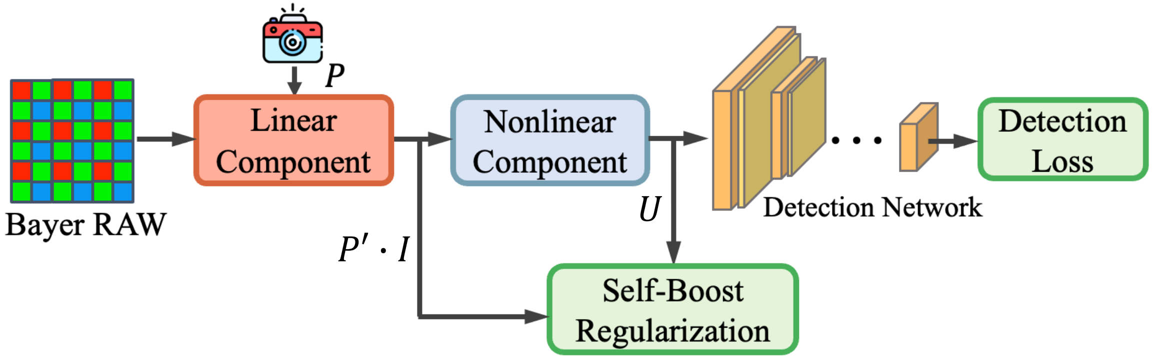
Self-Boost Regularization

Ideally, given an oracle sRGB image $U^* \in \mathbb{R}^{3\times \mathrm{H}\times W}$:

$$\min \|U^* - P' \cdot I\|$$

We get the closed-form least squares solution:

$$P^* = U^* \cdot I^T \cdot (I \cdot I^T)^{-1}$$



In order to avoid incurring additional data acquisition costs, we use the output U of the Nonlinear Component as a substitute for U^*, and define the resulting approximate solution mapping as $\tilde{P}$:

$$\tilde{P} := Q \cdot I^T \cdot (I \cdot I^T)^{-1},$$

which serves as a pseudo-target to facilitate the learning of the Linear Component. Since U is a function of P', obtaining $\tilde{P}$ through least squares lacks rigor and is continuously updated during training. Thus, we encourage the directional consistency for each channel transformation and activate it after a few epochs at the start of training.

$$\mathrm{P} = (p_1, p_2, p_3)^T$$

$$\mathcal{L}_{sb} = \sum_i \|1 - cos(p_i', \tilde{p}_i)\|$$

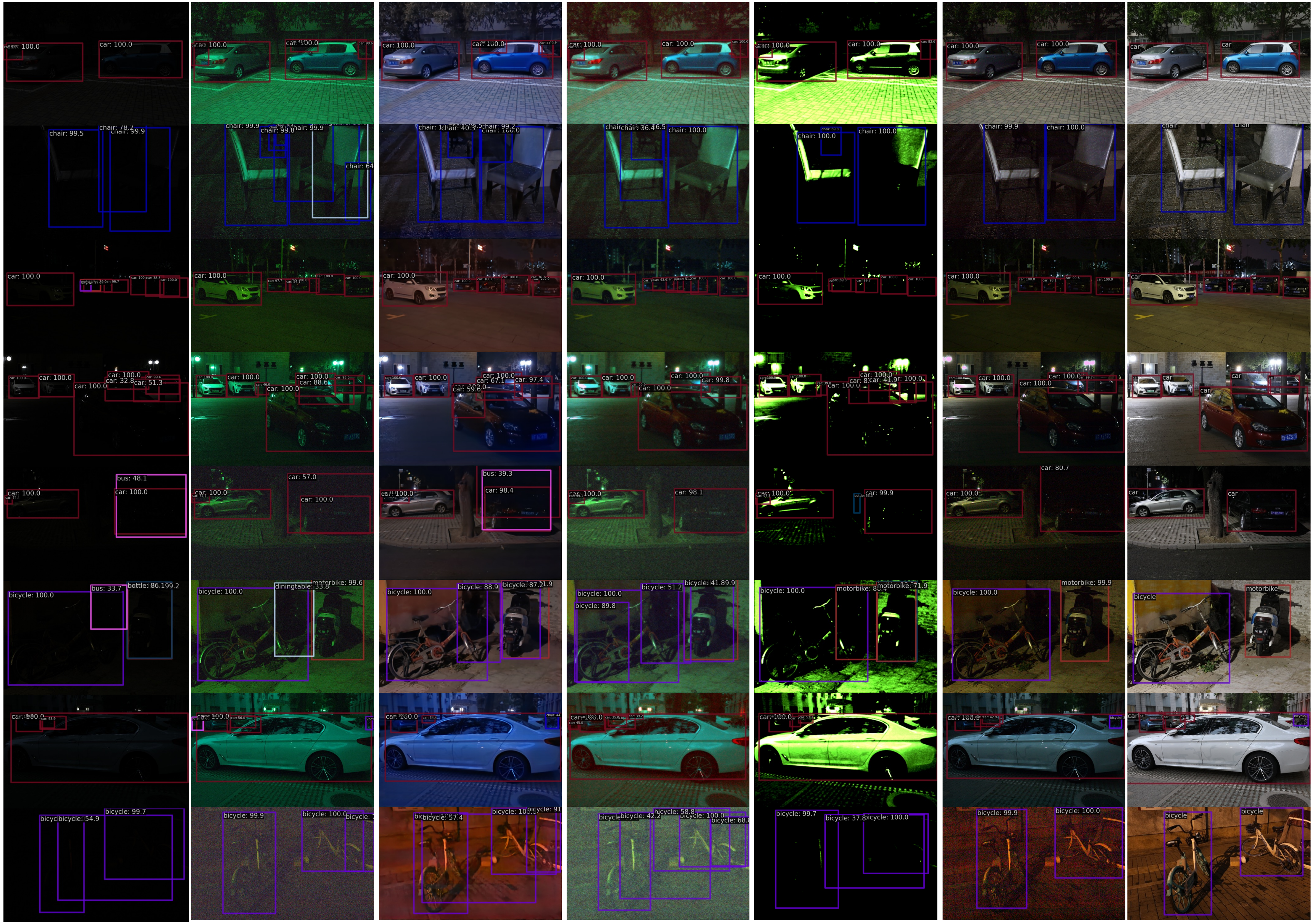
$\Longrightarrow$

$$\mathcal{L} = \mathcal{L}_{det} + \lambda \cdot \mathcal{L}_{sb}$$

Results

Table 1. Detection performance comparison on the real word LOD dataset. **Bold** indicates the best result, and underline indicates the second best. $*$ denotes a two-stage training method.

Image Format	Method	RetinaNet18 mAP	RetinaNet50 mAP	SpCNN50 mAP
sRGB or RAW-RGB	Default ISP	55.2	59.1	57.3
	Demosaic	52.6	61.3	61.2
	LIS	50.5	60.8	58.5
	FeatEnhancer	60.8	64.3	64.0
	RAW-Adapter	55.4	61.1	58.7
RAW (Bayer)	Default ISP	63.6	67.3	66.3
	SID*	61.2	64.7	68.3
	Demosaic	59.7	65.1	65.7
	LIS	58.4	67.9	63.3
	FeatEnhancer	63.4	67.0	64.8
	RAW-Adapter	59.9	66.2	64.2
	Ours	**64.9**	**70.4**	**68.8**



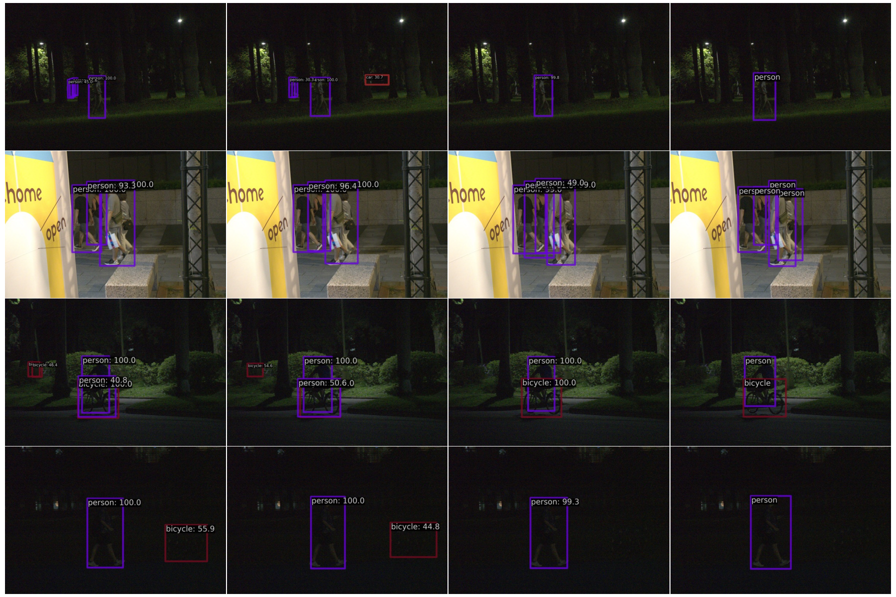
Default ISP | LIS | SID | RAW-Adapter | FeatEnHancer | **Dark-ISP** | Ground Truth

Results

Table 2. Detection performance comparison on the NOD dataset using RAW images captured in low-light conditions with two cameras.

Camera	Method	mAP	mAP_{50}	mAP_{75}
	Default ISP	28.3	51.0	29.2
	Demosaic	19.7	40.0	18.4
	SID	27.4	47.2	28.2
Sony	LIS	28.3	48.3	29.7
	FeatEnhancer	30.3	52.1	31.5
	RAW-Adapter	28.9	50.7	29.6
	Ours	**31.5**	**53.4**	**32.2**
	Default ISP	27.4	47.5	28.5
	Demosaic	27.6	48.6	27.8
	SID	22.2	42.0	20.5
Nikon	LIS	26.5	46.4	27.4
	FeatEnhancer	28.8	48.9	30.8
	RAW-Adapter	28.3	48.2	28.7
	Ours	**29.9**	**50.9**	**30.7**

FeatEnHancer RAW-Adapter **Dark-ISP** Ground Truth



FeatEnHancer RAW-Adapter **Dark-ISP** Ground Truth

Summary

We propose Dark-ISP, a novel approach to RAW image processing for low-light object detection, which is lightweight, effective and convenient to deploy.

Thank You very much for listening!