NEURAL INFORMATION PROCESSING SYSTEMS

You Only Look Around: Learning Illumination Invariant Feature for Low-light Object Detection

Mingbo Hong[1], Shen Cheng[1], Haibin Huang[2], Haoqiang Fan[1], Shuaicheng Liu[3],

https://github.com/MingboHong/YOLA

[1]Megvii Technology [2]Kuaishou Technology [3]University of Electronic Science and Technology of China

Method

You Only Look Around: Learning Illumination Invariant Feature for Low-light Object Detection

Motivation:
Leveraged **illumination-invariant** features to mitigate the effect of illumination.

How to acquire :

① Based on Dichromatic Reflection Model:

$$C_{p_i} = m(\vec{n}_{p_i}, \vec{l}_{p_i}) e^{C_{p_i}}(\lambda) \boxed{\rho^{C_{p_i}}(\lambda)}$$ **Preserved term**

$\vec{n}_{p_i}$: *Surface normal* $\vec{l}_{p_i}$: *Light direction*

$e^{C_{pi}}$: *Spectral power distribution of the illuminant*

$\rho^{C_{pi}}$: *Intrinsic property of the object*

② Cross color ratio

$$M_{rb} = \frac{R_{p_1} B_{p_2}}{R_{p_2} B_{p_1}}.$$

③ $log\ M_{rb}$

$$\begin{aligned} log(M_{rb}) = & \boxed{log(m(\vec{n_{p_1}}, \vec{l_{p_1}})) - log(m(\vec{n_{p_1}}, \vec{l_{p_1}}))} \\ & \boxed{+log(e^{R_{p1}}(\lambda)) - log(e^{R_{p2}}(\lambda))} \\ & +log(\rho^{R_{p1}}(\lambda)) - log(\rho^{R_{p2}}(\lambda)) \\ & \boxed{+log(m(\vec{n_{p_2}}, \vec{l_{p_2}})) - log(m(\vec{n_{p_2}}, \vec{l_{p_2}}))} \\ & \boxed{+log(e^{B_{p2}}(\lambda)) - log(e^{B_{p1}}(\lambda))} \\ & +log(\rho^{B_{p2}}(\lambda)) - log(\rho^{B_{p1}}(\lambda)). \end{aligned}$$

④ Illumination assumption

$$e^{C_{p1}} \approx e^{C_{p2}}$$

⑤ Simplified equation

$$\begin{aligned} log(M_{rb}) = & log(\rho^{R_{p1}}(\lambda)) - log(\rho^{R_{p2}}(\lambda)) \\ & +log(\rho^{B_{p2}}(\lambda)) - log(\rho^{B_{p1}}(\lambda)) \end{aligned}$$

Method

You Only Look Around: Learning Illumination Invariant Feature for Low-light Object Detection

Cross color ratio

$$M_{rb} = \frac{R_{p_1} B_{p_2}}{R_{p_2} B_{p_1}}.$$

Convolution operation

$$M_{rb} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} \circledast I_r - \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} \circledast I_b$$

③ $log\ M_{rb}$

$$\begin{aligned} log(M_{rb}) = & log(m(\vec{n_{p_1}}, \vec{l_{p_1}})) - log(m(\vec{n_{p_1}}, \vec{l_{p_1}})) \\ & + log(e^{R_{p1}}(\lambda)) - log(e^{R_{p2}}(\lambda)) \\ & + log(\rho^{R_{p1}}(\lambda)) - log(\rho^{R_{p2}}(\lambda)) \\ & + log(m(\vec{n_{p_2}}, \vec{l_{p_2}})) - log(m(\vec{n_{p_2}}, \vec{l_{p_2}})) \\ & + log(e^{B_{p2}}(\lambda)) - log(e^{B_{p1}}(\lambda)) \\ & + log(\rho^{B_{p2}}(\lambda)) - log(\rho^{B_{p1}}(\lambda)). \end{aligned}$$

Subtraction

- **Same channel**: Eliminate the illumination term
- **Cross-channel** : Eliminate surface normal and light direction terms

Method

You Only Look Around: Learning Illumination Invariant Feature for Low-light Object Detection

Cross color ratio

$$M_{rb} = \frac{R_{p_1} B_{p_2}}{R_{p_2} B_{p_1}}.$$

Convolution operation

$$M_{rb} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} \circledast I_r - \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} \circledast I_b$$

Subtraction

- **Same channel**: Eliminate the illumination term
- **Cross-channel** : Eliminate surface normal and light direction terms

(a) Result Baseline (b) IIM-Edge (c) Result IIM-Edge (d) IIM (e) Result IIM

Why learnable kernel:

Produce task-specific illumination invariant features for downstream tasks.

How to build a learnable kernel

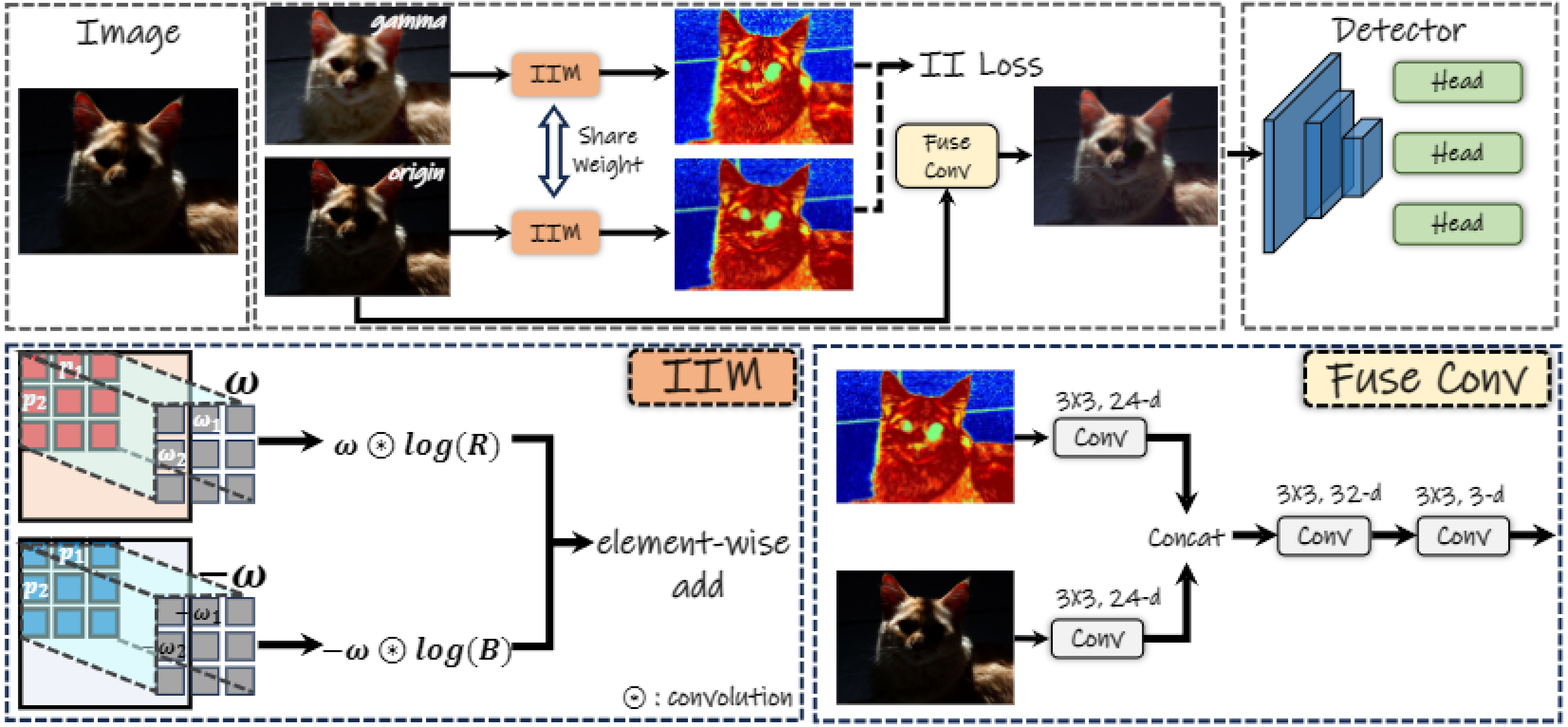
$$f_{\mathcal{W}_i}(I) = \begin{bmatrix} \mathcal{W}_i \circledast log(R) + (-\mathcal{W}_i) \circledast log(B) \\ \mathcal{W}_i \circledast log(R) + (-\mathcal{W}_i) \circledast log(G) \\ \mathcal{W}_i \circledast log(G) + (-\mathcal{W}_i) \circledast log(B) \end{bmatrix}$$

s.t. $$\overline{\mathcal{W}_n} = \frac{1}{k^2} \sum_{i=1}^{k^2} w_i = 0$$ (Zero mean constraint)

Method

You Only Look Around: Learning Illumination Invariant Feature for Low-light Object Detection

Pipeline



Detector	mAP	Size(M)	FPS
Baseline	72.5	32.044	57.7
IAT[7]	73.0	32.135	50.9
IAYOLO[8]	65.0	32.209	52.5
GDIP[9]	72.8	167.00	54.0
DENet[10]	73.5	32.089	55.7
PEYOLO[11]	67.8	32.135	38.8
Ours	75.2	32.052	56.6

Only Need 0.008M

Method

You Only Look Around: Learning Illumination Invariant Feature for Low-light Object Detection

Uneven lighting Condition:

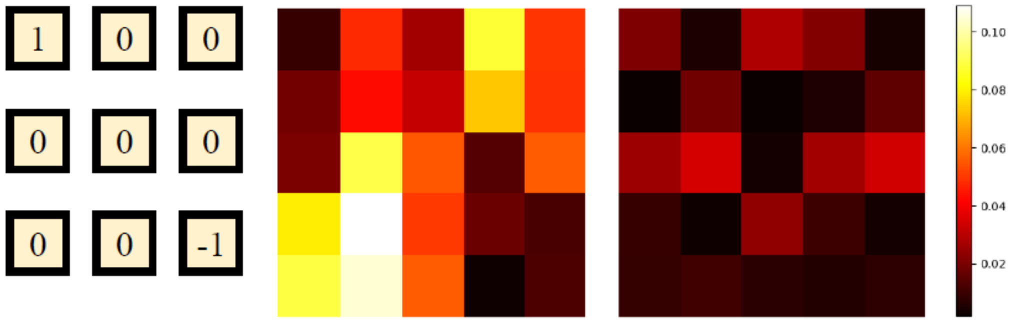
$$e^{C_{p1}} \approx e^{C_{p2}}$$

Illumination Invariant Loss

$$L = \begin{cases} \frac{1}{2}(f_{\mathcal{W}_i}(I) - f_{\mathcal{W}_i}(\sigma(I)))^2 & |f_{\mathcal{W}_i}(I) - f_{\mathcal{W}_i}(\sigma(I))| \leq \beta \\ |f_{\mathcal{W}_i}(I) - f_{\mathcal{W}_i}(\sigma(I))| - \frac{1}{2}\beta, & \text{otherwise.} \end{cases}$$

II Loss is proposed to encourage consistency of outputs from IIM across images with different illuminations, preventing trivial solutions within the kernel implicitly.

local means $\longrightarrow 0$



(a) Trivial Weight (b) w/o II loss (c) w/ II loss

Visualization of performing 3 × 3 mean filtering on the kernel weights

1)	Dataset	IIM	II-Loss	$\mathcal{K}_s$	YOLOv3	TOOD
2)	Exdark			3	71.0	72.5
3)		✓		3	71.1	74.8
4)		✓	✓	3	**72.7**	75.0
5)		✓		5	71.5	75.0
6)		✓	✓	5	**72.7**	**75.2**
7)	DarkFace			3	60.0	62.1
8)		✓		3	61.0	66.9
9)		✓	✓	3	**61.5**	**67.4**
10)		✓		5	60.2	65.8
11)		✓	✓	5	60.7	67.1

Experiment

You Only Look Around: Learning Illumination Invariant Feature for Low-light Object Detection

Methods	YOLOv3		TOOD	
	recall	mAP_{50}	recall	mAP_{50}
Baseline	84.6	71.0	91.9	72.5
KIND [53]	83.3	69.4	92.1	72.6
SMG [46]	82.3	68.5	91.8	71.5
NeRCo [47]	83.4	68.5	91.8	71.8
DENet [36]	84.2	71.3	92.6	73.5
GDIP [53]	84.8	72.4	92.2	72.8
IAT [53]	85.0	72.6	92.9	73.0
MAET [7]	85.1	72.5	92.5	74.3
YOLA-Naive	84.8	71.6	91.8	71.6
YOLA	**86.1**	**72.7**	**93.8**	**75.2**

Table 1: Quantitative comparisons of the ExDark dataset based on YOLOv3 and TOOD detectors.

Methods	YOLOv3		TOOD	
	recall	mAP_{50}	recall	mAP_{50}
Baseline	77.9	60.0	81.5	62.1
KIND [53]	76.0	58.4	82.4	63.8
SMG [46]	69.3	48.9	77.1	55.8
NeRCo [47]	68.9	49.1	76.8	55.6
DENet [36]	77.7	60.0	84.1	66.2
GDIP [53]	77.8	60.4	82.1	62.9
IAT [53]	77.6	59.8	82.1	62.0
MAET [7]	77.9	59.9	83.6	64.8
YOLA-Naive	76.6	59.2	82.8	64.6
YOLA	**79.1**	**61.5**	**84.9**	**67.4**

Table 2: Quantitative comparisons of the UG^2+DARK FACE dataset based on YOLOv3 and TOOD detectors.

Dataset	Method	AP_{50}	AP_{75}	mAP
well-lit	TOOD	59.0	45.3	41.7
	+ YOLA	**59.4**	**46.0**	**42.3**
over-light	TOOD	57.4	43.8	40.5
	+ YOLA	**58.3**	**44.6**	**41.2**

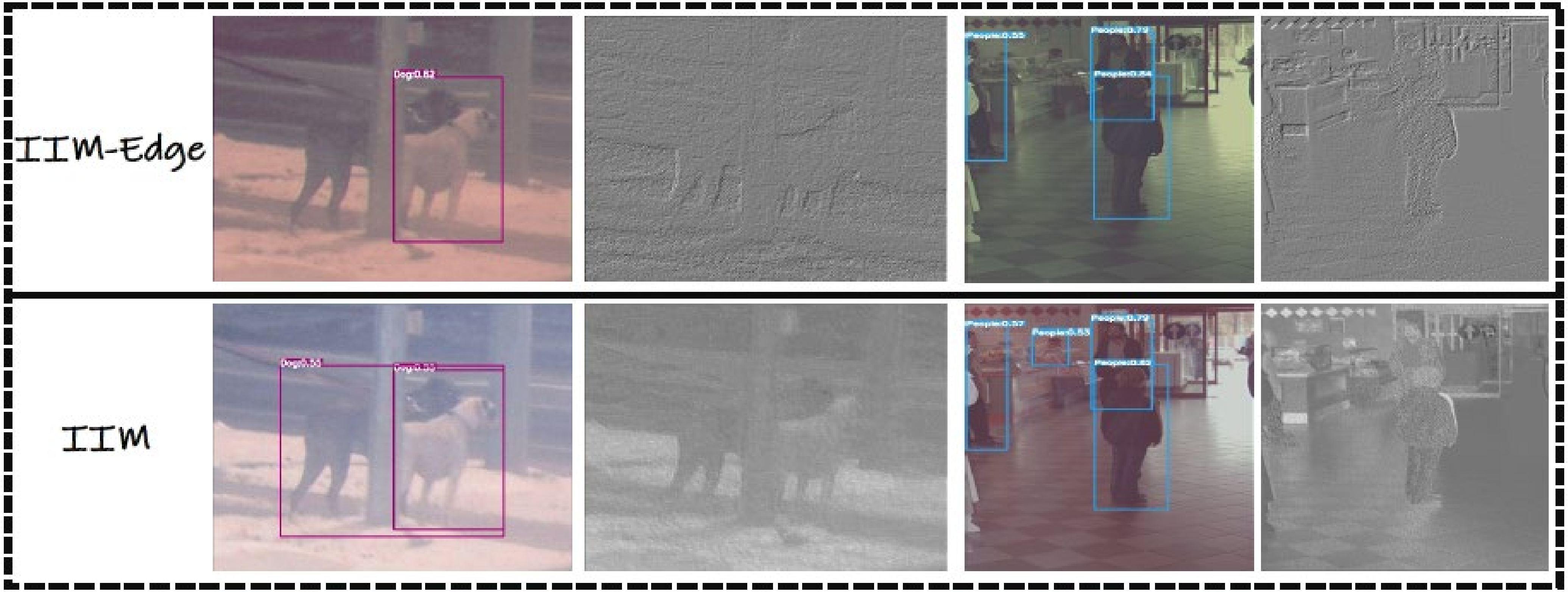
Table 4: Ablation study for YOLA on COCO 2017val.

Method	Kind	SMG	NeRco	DENet	MAET	Ours
Size(M)	8.21	17.90	23.30	0.04	40	**0.008**

Table 5: Model size of different methods.

Experiment

You Only Look Around: Learning Illumination Invariant Feature for Low-light Object Detection

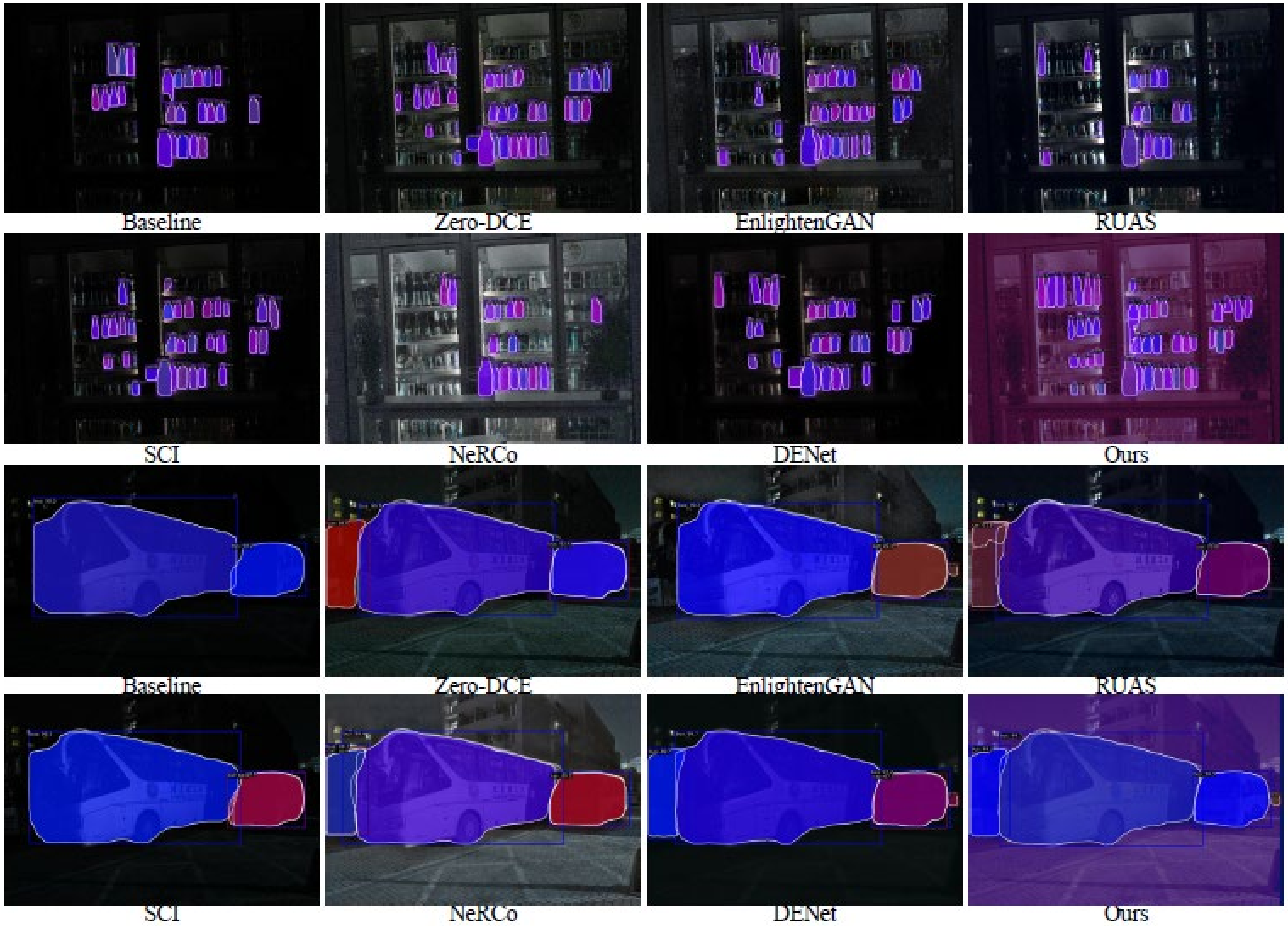


Experiment

You Only Look Around: Learning Illumination Invariant Feature for Low-light Object Detection

Experiment

You Only Look Around: Learning Illumination Invariant Feature for Low-light Object Detection



Conclusion

You Only Look Around: Learning Illumination Invariant Feature for Low-light Object Detection

- We introduce YOLA, a novel framework for object detection in low-light conditions by leveraging **illumination-invariant features**.
- We design a novel Illumination-Invariant Module to extract illumination-invariant features without requiring additional paired datasets, and can be **seamlessly integrated** into existing object detection methods.
- We provide an in-depth analysis of the extracted illumination-invariant paradigm and propose a learning illumination-invariant paradigm.
- Our experiments show YOLA can significantly improve the detection accuracy of existing methods when dealing with low-light images.