HUAWEI

中国科学技术大学

University of Science and Technology of China

JUNE 18-22, 2023

CVPR

VANCOUVER, CANADA

PAPER TAG: WED-PM-098

Toward RAW Object Detection: A New Benchmark and A New Model

Ruikang Xu[1*] Chang Chen[2*] Jingyang Peng[2*] Cheng Li[2]
Yibin Huang[2] Fenglong Song[2] Youliang Yan[2] Zhiwei Xiong[1†]

[1]University of Science and Technology of China [2]Huawei Noah's Ark Lab

VIDAR

Visual Information Discovery And Recovery

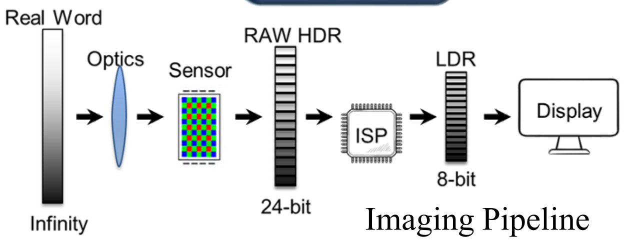
NEL-BITA

National Engineering Laboratory for Brain-Inspired Intelligence Technology and Application

Toward RAW Object Detection: A New Benchmark and A New Model

——OVERVIEW——

Background



Imaging Pipeline

SDR vs. RAW for Detection

Analysis on RAW Detection

Impact of the dynamic range for object detection.

Date Type	10-bit dataset			12-bit dataset			24-bit dataset			Params
	AP	AP50	AP75	AP	AP50	AP75	AP	AP50	AP75	
SDR	43.8	65.6	47.4	67.3	93.6	78.4	52.1	74.6	56.8	0.90M
RAW	43.3	64.3	47.3	65.3	92.9	75.8	34.6	54.7	35.4	
SDR	48.1	69.4	53.2	70.9	94.9	84.0	63.3	88.4	69.6	2.27M
RAW	47.8	69.0	51.6	68.4	93.9	81.5	43.9	66.8	46.1	
SDR	51.8	73.2	56.5	72.8	95.5	86.2	69.7	91.3	76.7	8.92M
RAW	51.2	72.6	56.1	70.5	94.7	84.2	47.5	67.1	52.7	

Ablation of the software ISP pipeline.

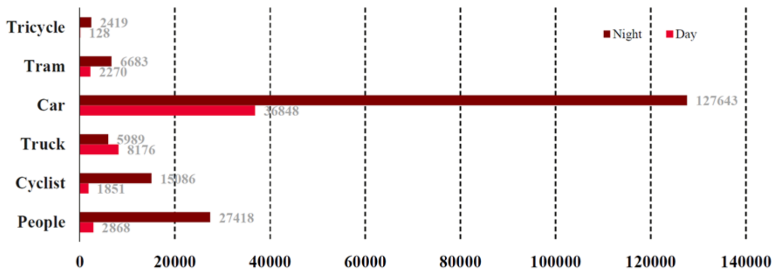
Data Type	AP	AP50	AP75
RAW	32.3	53.7	32.9
RAW (DM+AWB)	34.6	54.7	35.4
RAW (DM+AWB+DRA)	52.1	74.4	56.9
RAW (ISP Pipeline)	53.3	76.8	58.6
RAW (DRA)	51.7	76.1	56.2
RAW (ISP Pipeline w/o DRA)	35.2	56.9	35.7

(Dynamic Range Adjustment is Important for Detection.)

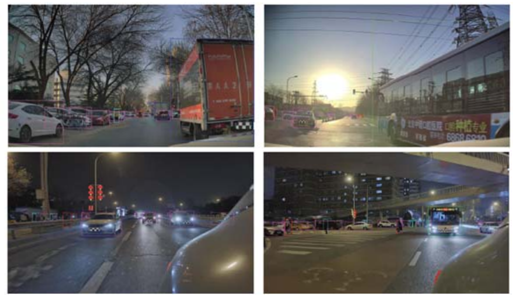
RAW Object Detection Dataset

Dataset	Sensor	Dynamic Range	Images	Category	Instance	Scenario
PASCALRAW [27]	Nikon D3200 DSLR	12-bit	4,259	3 classes	6,550	Day
LOD [39]	Canon EOS 5D Mark IV	14-bit	2,230	8 classes	9,726	Low-light
Ours	Sony IMX490	24-bit	25,207	6 classes	237,379	Day & Night

Comparison between the PASCALRAW, LOD, and our RAW sensor datasets.

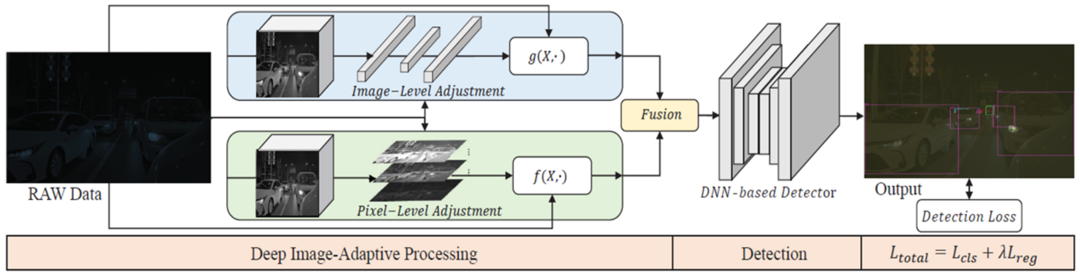


Number of instances per category for our ROD dataset.



Example scenes in our ROD dataset.
(Show SDR data for better visualization)

HDR RAW Detection Pipeline



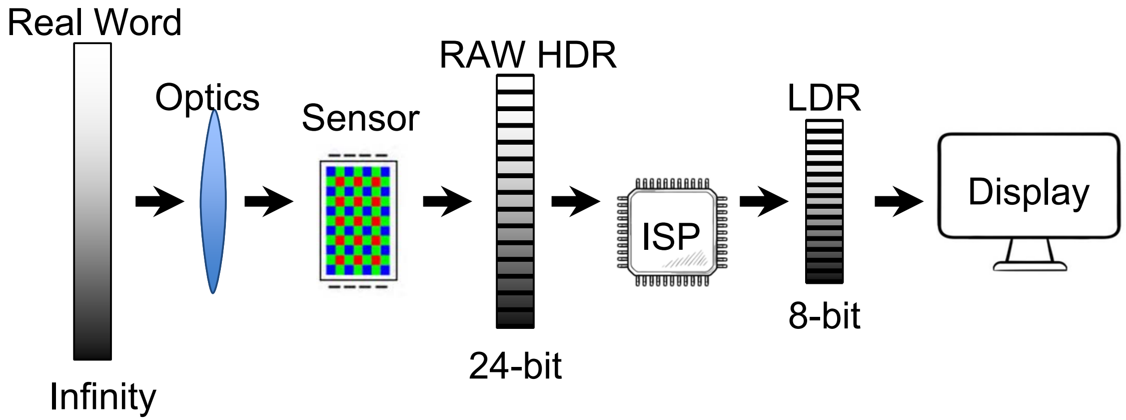
- Image-level adjustment module

$$Y_I = g(X, \mathcal{F}_I(X_{lr}; \vartheta_I)) = X^{\gamma}$$

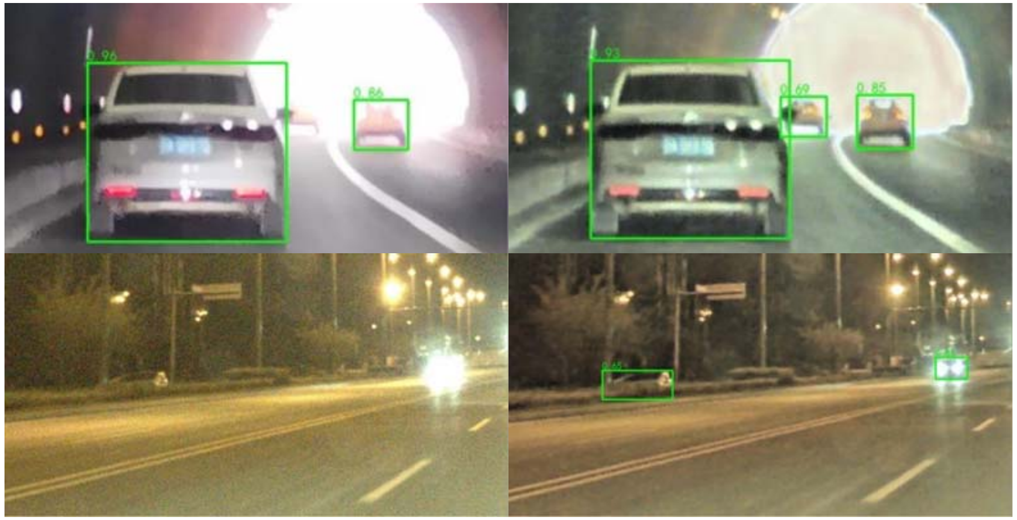
- Pixel-level adjustment module

$$Y_P = f(X, \mathcal{F}_P(X_{lr}; \vartheta_P)) = \sum_{k=0}^{K-1} m_k \delta_k(X).$$

RAW Sensor Data for Detection



(a) Imaging Pipeline



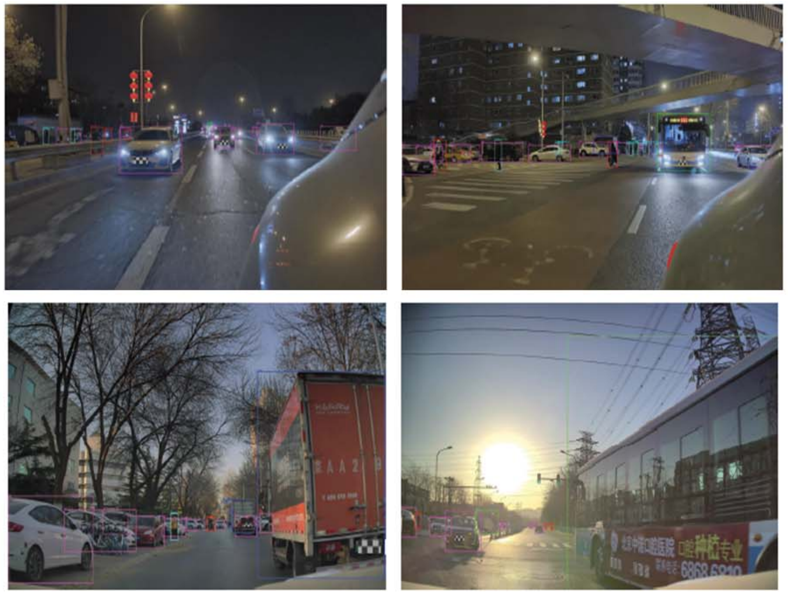
(b) LDR vs. RAW for Detection

- ***RAW sensor data preserve all the from imaging sensor and naturally has a high dynamic range.***
- ***RAW sensor data preserving all information and reducing latency if directly used as input.***

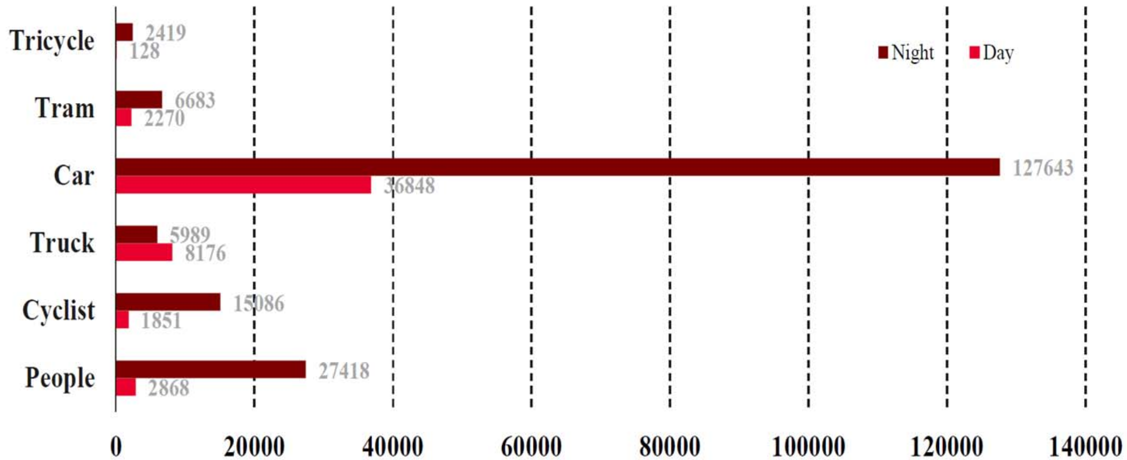
Dataset for RAW Object Detection (ROD)

Dataset	Sensor	Dynamic Range	Images	Category	Instance	Scenario
PASCALRAW	Nikon D3200 DSLR	12-bit	4,259	3 classes	6,550	Day
LOD	Canon EOS 5D Mark IV	14-bit	2,230	8 classes	9,726	Low-light
Ours	Sony IMX490	24-bit	25,207	6 classes	237,379	Day & Night

Comparison between the PASCALRAW, LOD, and our RAW sensor datasets.



Example scenes in our ROD dataset.
(Show SDR data for better visualization)



Number of instances per category for our ROD dataset.

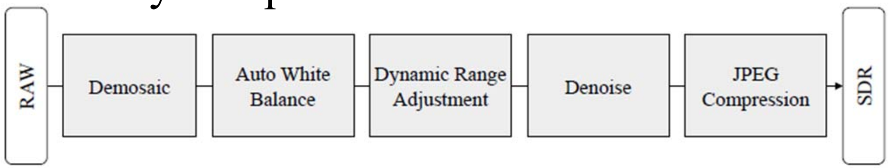
HUAWEI

Analysis on RAW Detection

Impact of the dynamic range for detection

Date Type	10-bit			12-bit			24-bit			Params
	AP	AP50	AP75	AP	AP50	AP75	AP	AP50	AP75	
SDR	43.8	65.6	47.4	67.3	93.6	78.4	52.1	74.6	56.8	0.90M
RAW	43.3	64.3	47.3	65.3	92.9	75.8	34.6	54.7	35.4	
SDR	48.1	69.4	53.2	70.9	94.9	84.0	63.3	88.4	69.6	2.27M
RAW	47.8	69.0	51.6	68.4	93.9	81.5	43.9	66.8	46.1	
SDR	51.8	73.2	56.5	72.8	95.5	86.2	69.7	91.3	76.7	8.92M
RAW	51.2	72.6	56.1	70.5	94.7	84.2	47.5	67.1	52.7	

Key components of the software ISP

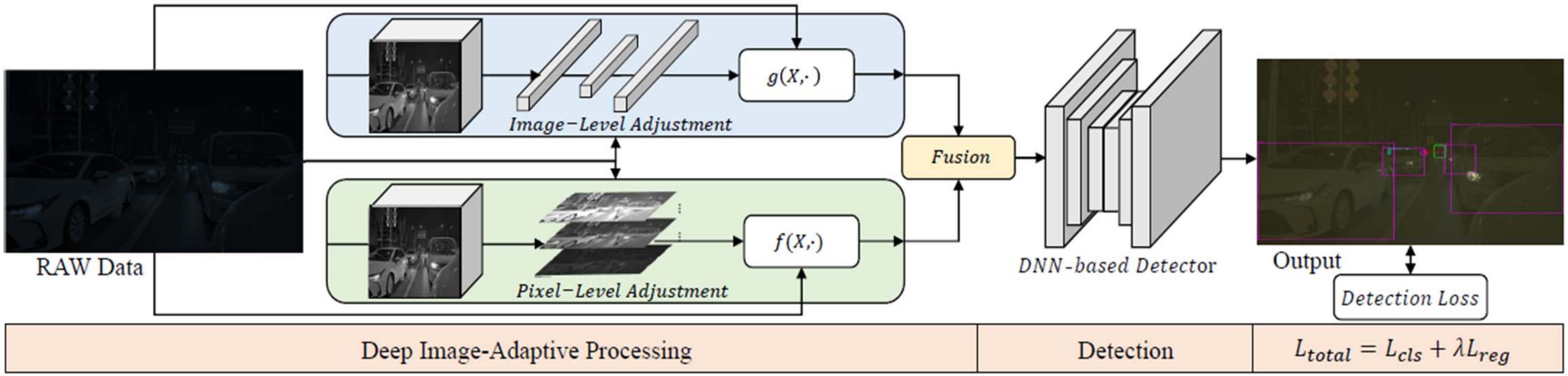


Ablation on the software ISP

Data Type	AP	AP50	AP75
RAW	32.3	53.7	32.9
RAW (DM+AWB)	34.6	54.7	35.4
RAW (DM+AWB+DRA)	52.1	74.4	56.9
RAW (ISP Pipeline)	53.3	76.8	58.6
RAW (DRA)	51.7	76.1	56.2
RAW (ISP Pipeline w/o DRA)	35.2	56.9	35.7

- ***DNNs-based detection algorithms cannot handle HDR data, and the performance degradation gets worse when dynamic range increases; ISP system is important for DNNs-based detection.***
- ***Dynamic range adjustment is inevitable to detection on the HDR RAW sensor data, since the higher the dynamic range, the more difficult it is to extract information by DNNs.***

HDR RAW Detection Pipeline



- Image-level Adjustment Module

$$Y_I = g(X, \mathcal{F}_I(X_{lr}; \vartheta_I)) = X^{\gamma}$$

- Pixel-level Adjustment Module

$$Y_P = f(X, \mathcal{F}_P(X_{lr}; \vartheta_P)) = \sum_{k=0}^{K-1} m_k \delta_k(X).$$

- Loss Function

$$L_{total} = L_{cls} + \lambda L_{reg}$$

- ***Jointly optimized with the downstream detection network in an End-to-End scheme.***
- ***Trained together with the detector from scratch only using detection loss functions.***
- ***Respectively explore the image-level and pixel-level information; Light-weight and computationally efficient.***

Qualitative Evaluation

Method	Day				Night				Params (M)	Flops (G)
	AP	AR	AP50	AP75	AP	AR	AP50	AP75		
SDR	52.1	57.2	74.6	56.8	50.3	59.7	80.0	53.7	-	-
RAW	34.6	40.6	54.7	35.4	1.7	5.1	4.5	0.9	-	-
Gamma [12]	52.1	57.3	74.4	56.9	50.8	60.2	80.8	55.0	-	-
Mu-Log [2]	51.5	56.5	74.0	55.6	49.8	57.7	79.3	52.3	-	-
IA-Gamma [12]	53.5	59.7	78.2	57.1	51.8	59.8	81.4	55.8	0.02	0.97
IA-Mu-Log [2]	23.0	31.3	41.9	21.9	50.2	58.9	80.1	54.4	0.02	0.97
GTM [16]	45.1	51.4	68.4	47.6	1.7	4.2	4.1	1.1	0.02	0.97
GTM-DI [16]	45.7	51.9	69.8	48.6	4.8	11.0	10.5	3.7	0.02	0.97
MW-ISPNet [17]	43.4	50.4	67.2	45.8	33.6	44.6	59.1	33.3	9.14	1690.54
Lite-ISPNet [17]	46.8	52.4	68.9	47.3	37.5	47.1	61.9	36.6	5.94	2860.12
IA-ISPNet [23]	54.6	61.2	81.9	59.3	52.7	60.9	81.9	56.8	0.26	0.91
Ours	**58.7**	**63.9**	**85.3**	**61.3**	**54.2**	**61.7**	**83.0**	**58.2**	0.08	0.64

Quantitative comparison with YOLOX (0.90M) on the day and night scenarios of ROD in terms of AP, AR, AP50, and AP75.

- ***DNNs-based detector is ineffective on night scenario RAW sensor data.***
- ***Surpasses SDR data with improvements of 6.6% and 3.9% on the day and night scenarios.***
- ***Boosts the performance with only 0.08 (M) parameters and 0.64 (G) FLOPs.***

Quantitative Evaluation

(a) Detection on the day scenario (b) Detection on the night scenario

Visualizing results with confidence scores over 0.4 in the day and night scenarios of ROD.

- ***Day scenario, our method can deal with strong glare of sunlight and severe lighting variance.***
- ***Night scenario, our method can handle lowlight condition and accurately recognize objects.***

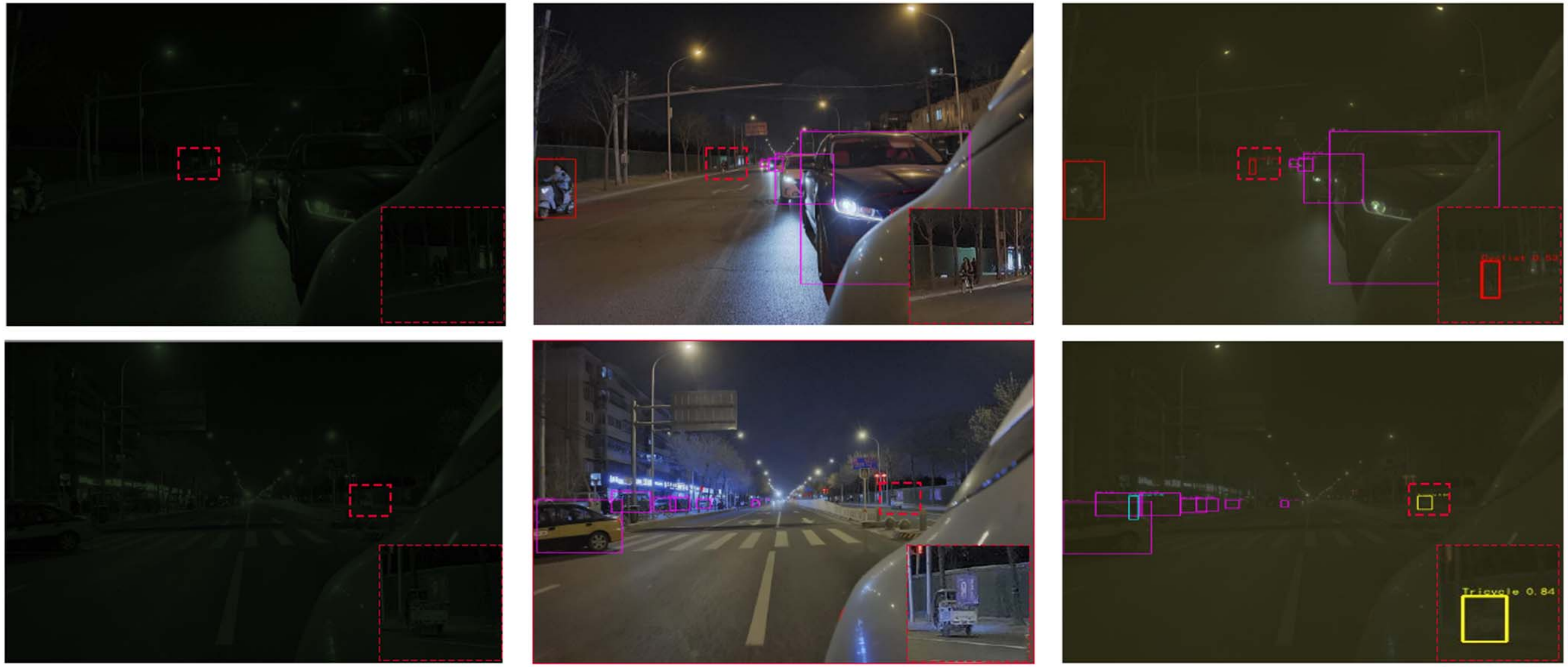
Quantitative Evaluation

——MoreResults——

(a) RAW (b) SDR (c) Ours

Visualizing results with confidence scores over 0.4 in the day scenarios of ROD.

Quantitative Evaluation
——MoreResults——



(a) RAW (b) SDR (c) Ours

Visualizing results with confidence scores over 0.4 in the night scenarios of ROD.

Ablation Studies

Quantitative comparison with Sparse R-CNN (104.54M) on the day scenario of the ROD dataset.

Method	AP	AR	AP50	AP75
SDR	73.5	80.8	91.8	84.0
RAW	66.3	73.6	88.1	78.9
Gamma [12]	73.7	82.0	92.2	83.1
Mu-Log [2]	72.7	81.4	91.0	84.0
IA-Gamma [12]	75.1	82.4	92.4	85.7
IA-Mu-Log [2]	74.2	80.2	91.2	84.6
GTM [16]	71.4	78.5	89.4	82.2
GTM-DI [16]	72.6	79.2	89.6	82.5
MW-ISPNet [17]	71.6	77.8	91.4	84.2
MW-ISPNet [17]	72.7	79.2	91.9	85.2
IA-ISPNet [23]	75.6	81.2	91.6	85.1
Ours	**77.4**	**83.6**	**93.2**	**87.3**

Quantitative comparison with YOLOX (8.92M) on the day scenario of the ROD dataset.

Method	AP	AR	AP50	AP75
SDR	69.3	72.4	91.3	76.7
RAW	47.5	52.2	67.1	52.7
Gamma [12]	71.2	74.7	94.2	82.4
Mu-Log [2]	69.1	72.8	93.9	78.1
IA-Gamma [12]	72.4	75.6	94.4	82.3
IA-Mu-Log [2]	42.7	64.6	46.9	48.0
GTM [16]	66.0	70.3	88.9	76
GTM-DI [16]	66.4	71.9	90.3	72.7
MW-ISPNet [17]	51.3	66.4	83.3	71.2
Lite-ISPNet [17]	54.6	68.8	85.3	77.2
IA-ISPNet [23]	73.1	76.7	94.5	83.1
Ours	**75.5**	**78.6**	**94.9**	**83.9**

- ***Proposal-based detector on the day scenario of the ROD dataset. Our method outperforms SDR and IA-ISPNet.***
- ***Increasing the parameters number of YOLO-X to 8.92(M). Our method outperforms SDR and IA-ISPNet.***

Ablation Studies

Quantitative comparison with YOLOX (0.90M) on different dynamic range datasets.

Method	10-bit Dataset		12-bit Dataset		24-bit dataset		Params
	AP	AP50	AP	AP50	AP	AP50	
SDR	43.8	65.6	67.3	93.6	52.1	74.6	
RAW	43.3	64.3	65.3	92.9	34.6	54.7	0.90(M)
Ours	44.8	66.2	68.7	94.2	58.7	85.3	
SDR	48.1	69.4	70.9	94.9	63.3	88.4	
RAW	47.8	69.0	68.4	93.9	43.9	66.8	2.27(M)
Ours	50.6	71.2	70.2	94.9	67.8	92.2	
SDR	51.8	73.2	72.8	95.5	69.7	91.3	
RAW	51.2	72.6	70.5	94.7	47.5	67.1	8.92(M)
Ours	54.1	74.9	72.5	95.2	75.5	94.9	

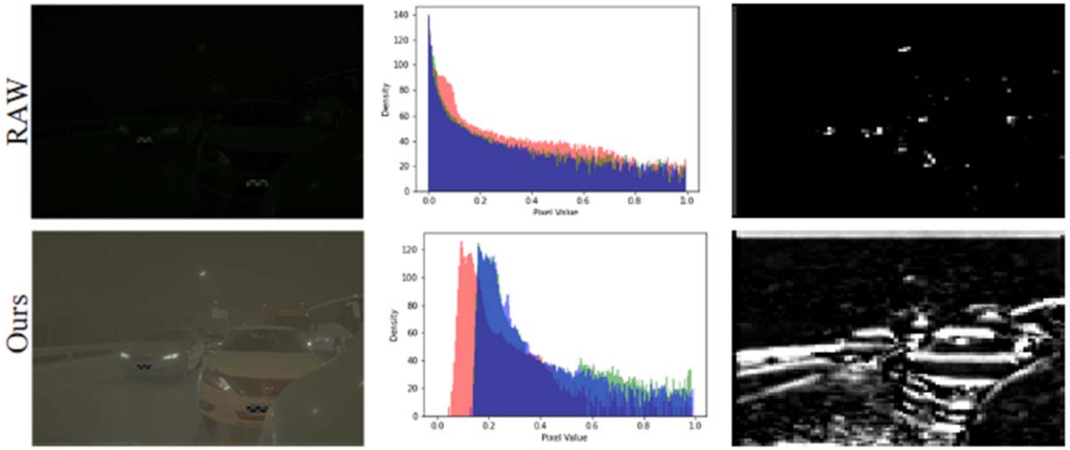
- ***Our method outperforms the SDR data on different dynamic range datasets.***

Ablation on the Image-Level Adjustment (ILA) and Pixel-Level Adjustment (PLA) modules.

Method	Day			Night		
	AP	AR	AP50	AP	AR	AP50
SDR	52.1	57.2	74.6	50.3	59.7	80.0
RAW	34.6	40.6	54.7	1.7	5.1	4.5
Ours w/o PLA	56.5	63.8	82.3	53.0	60.9	82.1
Ours w/o ILA	50.0	55.9	73.5	52.0	60.4	81.4
Ours	58.7	63.9	85.3	54.2	61.7	83.0

- ***Proposed modules are effective for the detection on RAW sensor data.***

Analysis for Performance Drop on RAW



(a) Data (b) Distribution (c) Feature visualization

- ***Pixels of RAW sensor data are distributed in the low-value area resulting in a lack of texture information.***
- ***The case of imaging a strong glare in an extremely dark scene, which means several close-to-one values inside a nearly zero-value background***

Impact of texture information on the performance of detection with YOLOX on the day and night scenario of the ROD dataset.

Method	Skew	Entropy of GLCM	AP
RAW	8.1742	11.1691	34.6
GTM-DI [16]	2.3311	20.9431	45.1
Gamma [12]	0.8873	24.0634	52.1
IA-Gamma [12]	0.6098	24.1645	53.5
Ours	0.9719	24.5954	58.7

Method	Skew	Entropy of GLCM	AP
RAW	136.4161	0.0876	1.7
GTM-DI [4]	123.4059	0.1865	1.8
Gamma [2]	2.8328	15.6294	50.8
IA-Gamma [2]	2.2821	15.7872	51.8
Ours	1.6242	16.3371	54.2

- ***Dynamic range adjustment method is effective to boost texture information.***
- ***Performance of detection on RWA sensor data is positively associated with the entropy of GLCM.***

Conclusion

- *Novel RAW sensor dataset for object detection on HDR RAW sensor data.*
- *Simple and effective adjustment method for detection on RAW sensor data.*
- *Extensive experiments demonstrate that object detection on HDR RAW sensor data significantly outperforms that on SDR data in different situations.*

HUAWEI

中国科学技术大学

University of Science and Technology of China

JUNE 18-22, 2023

CVPR

VANCOUVER, CANADA

Thanks!

VIDAR

Visual Information Discovery And Recovery

NEL-BITA

National Engineering Laboratory for Brain-Inspired Intelligence Technology and Application