Toward RAW Object Detection: A New Benchmark and A New Model
——Supplementary Material——

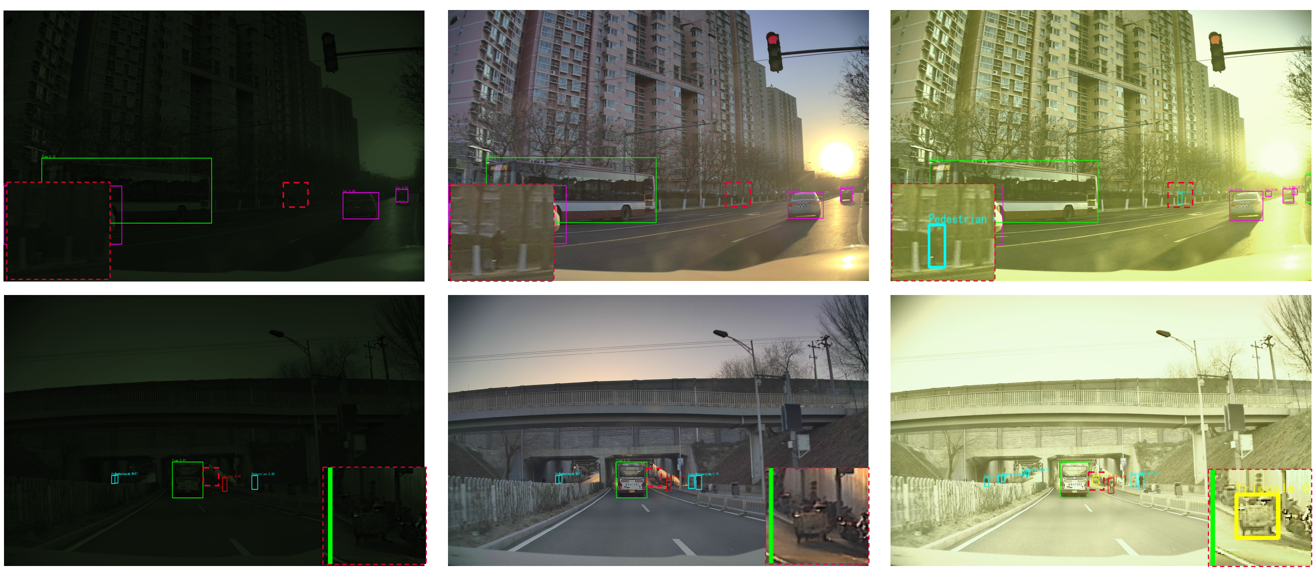
Ruikang Xu[1*] Chang Chen[2*] Jingyang Peng[2*] Cheng Li[2]
Yibin Huang[2] Fenglong Song[2] Youliang Yan[2] Zhiwei Xiong[1†]
[1]University of Science and Technology of China [2]Huawei Noah's Ark Lab
xurk@mail.ustc.edu.cn, zwxiong@ustc.edu.cn,
{chenchang25, pengjingyang1, licheng89, huangyibin1, songfenglong, yanyouliang}@huawei.com

The content of this supplementary material is organized as:

- More results of qualitative comparison are in Section 1.
- More results of ablation study in Section 2.
- Analysis of the impact of texture information on the performance of DNNs-based detector is in Section 3.

1. More Visual Results

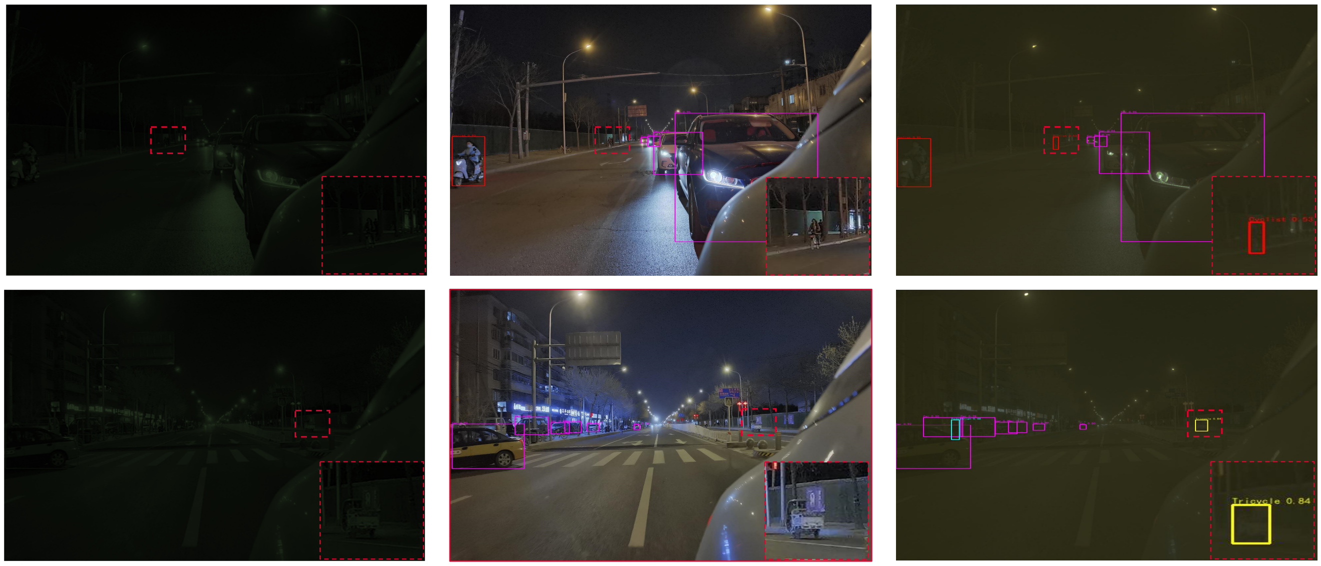
We show more visual results in Figure 1 and Figure 2. Specifically, we visualize detection results of original RAW data, SDR data, and our method with confidence scores over 0.4 in the day and night scenarios of the ROD dataset. For the day scenario, object detection on the HDR RAW data with our method can effectively handle the over-exposed regions caused by the glaring sunlight. For the night scenario, objective detection on the HDR RAW data with our method can accurately recognize small objects in the low-light condition.



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

Figure 1. Visual examples of object detection on the day scene of the ROD dataset. (a), (b) and (c) are results of RAW data, SDR data and our method, respectively. Our method significantly outperforms the SDR data. Please zoom in for confidence scores and class predictions.

*Equal contribution. †Corresponding author. This work was done when Ruikang Xu was a research intern at Huawei Noah's Ark Lab.



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

Figure 2. Visual examples of object detection on the night scene of the ROD dataset. (a), (b) and (c) are results of RAW data, SDR data and our method, respectively. Our method significantly outperforms the SDR data. Please zoom in for confidence scores and class predictions.

2. Ablation Study

2.1. Ablation Study on Model Size

To evaluate the impact of the model size for detection on the RAW sensor data, we also increase the number of channels of all detector convolutions by a factor of 2, increasing the number of model parameters to 2.27(M). The experiment results are shown in Table 1. We can see that the proposed model outperforms the SDR data by 4.5% and 2.9% in terms of AP and AR. The results demonstrate that detection on the RAW sensor data outperforms the SDR data with different model sizes.

Table 1. Quantitative comparison with YOLOX (2.27M) on the day scenario of the ROD dataset in terms of AP, AR, AP50, and AP75. The best results are highlighted with bold fonts. Detection on the RAW sensor data outperforms the SDR data with various model sizes.

Method	AP	AR	AP50	AP75	Params(M)	FLOPs(G)
SDR	63.3	69.1	88.4	69.6	-	-
RAW	43.9	50.1	66.8	46.1	-	-
Gamma [2]	64.2	68.8	90.4	68.2	-	-
Mu-Log [1]	62.7	67.4	88.9	66.8	-	-
IA-Gamma [2]	65.8	70.7	90.9	69.8	0.02	0.97
IA-Mu-Log [1]	42.7	48.0	64.6	46.9	0.02	0.97
GTM [4]	56.8	64.6	82.4	61.1	0.02	0.97
GTM-DI [4]	60.7	66.0	86.9	67.2	0.02	0.97
MW-ISPNet [5]	46.3	52.6	69.3	49.4	9.14	1690.54
Lite-ISPNet [7]	48.7	55.1	70.9	49.4	5.94	2860.12
IA-ISPNet [6]	65.9	72.3	90.9	71.1	0.26	0.91
Ours	**67.8**	**72.0**	**92.2**	**76.0**	0.08	0.64

2.2. Ablation Study on Proposed Modules

To demonstrate the effectiveness of the proposed modules, we retrain our method by including the modules step by step. The evaluation results of each component are listed in Table 2, which demonstrate that our proposed modules are effective for the detection on RAW sensor data.

Table 2. 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

3. Analysis

To further analyze the impact of the texture information of RAW sensor data on the performance of DNNs-based detection methods, we perform the experiments on the night scenario of the ROD dataset as shown in Table 3. We employ the entropy of the gray-level co-occurrence matrix (GLCM) [3] as the metric to evaluate the necessity of dynamic range adjustment methods and the YOLOX (0.90M) as the detection network for analysis. We can see that the performance of detection is positively associated with the entropy of GLCM, which shows the importance of dynamic range adjustment algorithms for object detection on HDR RAW sensor data.

Table 3. Impact of texture information on the performance of detection with YOLOX on the night scenario of the ROD dataset.

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

References

[1] Helder Araujo and Jorge M Dias. An introduction to the log-polar mapping [image sampling]. In *Proceedings II Workshop on Cybernetic Vision*, pages 139–144. IEEE, 1996. 2

[2] Hongwei Guo, Haitao He, and Mingyi Chen. Gamma correction for digital fringe projection profilometry. *Applied optics*, 43(14):2906–2914, 2004. 2, 3

[3] Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, (6):610–621, 1973. 3

[4] Yuanming Hu, Hao He, Chenxi Xu, Baoyuan Wang, and Stephen Lin. Exposure: A white-box photo post-processing framework. *ACM Transactions on Graphics*, 37(2):1–17, 2018. 2, 3

[5] Andrey Ignatov, Radu Timofte, Zhilu Zhang, Ming Liu, Haolin Wang, Wangmeng Zuo, Jiawei Zhang, Ruimao Zhang, Zhanglin Peng, Sijie Ren, et al. Aim 2020 challenge on learned image signal processing pipeline. In *ECCVW*, 2020. 2

[6] Wenyu Liu, Gaofeng Ren, Runsheng Yu, Shi Guo, Jianke Zhu, and Lei Zhang. Image-adaptive yolo for object detection in adverse weather conditions. *arXiv preprint arXiv:2112.08088*, 2021. 2

[7] Zhilu Zhang, Haolin Wang, Ming Liu, Ruohao Wang, Jiawei Zhang, and Wangmeng Zuo. Learning raw-to-srgb mappings with inaccurately aligned supervision. In *ICCV*, 2021. 2