Rawgment: Noise-Accounted RAW Augmentation Enables Recognition in a Wide Variety of Environments (Supplementary Material)

Masakazu Yoshimura Junji Otsuka Atsushi Irie Takeshi Ohashi
Sony Group Corporation
{masakazu.yoshimura, junji.otsuka, atsushi.irie, takeshi.a.ohashi}@sony.com

1. Visualization Results

In Fig. 1, the detection results are drawn on the output of the corresponding ISPs. The proposed method shows a significant improvement in accuracy under the condition that the simplest gamma tone mapping is used as an ISP. In addition, the accuracy of the proposed method is the best despite the use of the simple ISP with limited visibility against a rich black-box ISP because of the effective noise-accounted RAW augmentation.

2. Additional Experiments

2.1. Versatility to Different Detectors

TTFNet [3] is used as a detector in the main paper because of the low training cost. In this section, the versatility of the proposed noise-accounted RAW augmentation is checked. As a different type of detector, we choose DeformableDETR [4] as a detector. Also, we change the backbone to ResNet50 [2] to check the proposed method's effectiveness with a larger model. Furthermore, the backbone is pre-trained with ImageNet [1] to compare with the best accuracy. Other experimental setups are the same as those with TTFNet.

The result is shown in Fig. 1. Because a larger detector with the pre-trained backbone is used, all methods have improved accuracy, but there is still a great improvement from the conventional augmentation after ISP setup to the proposed noise-accounted RAW augmentation when the simplest ISP is used. Moreover, if parameterized gamma tone mapping and the proposed augmentation are used, the accuracy is even improved from the result with the elaborated black-box ISP, which should benefit most from the pre-training with sRGB images.

The future work is to check the effectiveness of the combination of the black-box ISP and the proposed augmentation by implementing the black-box ISP as software that works on a computer.

Table 1. Evaluation with DeformableDETR [4] whose backbone is ResNet50 pre-trained with ImageNet.

augmentation		noise	mAP@0.5:0.95 [%] black-box ISP	simple ISP simplest	simple ISP parameterized
Color	after	-	51.6	40.2	-
+	before	-	-	46.8	47.5
Blur	(ours)	ours	-	**51.5**	**52.0**

3. Acknowledgements

We would like to thank Aji Widya and Iheb Begalcem for their helpful comments to this manuscript.

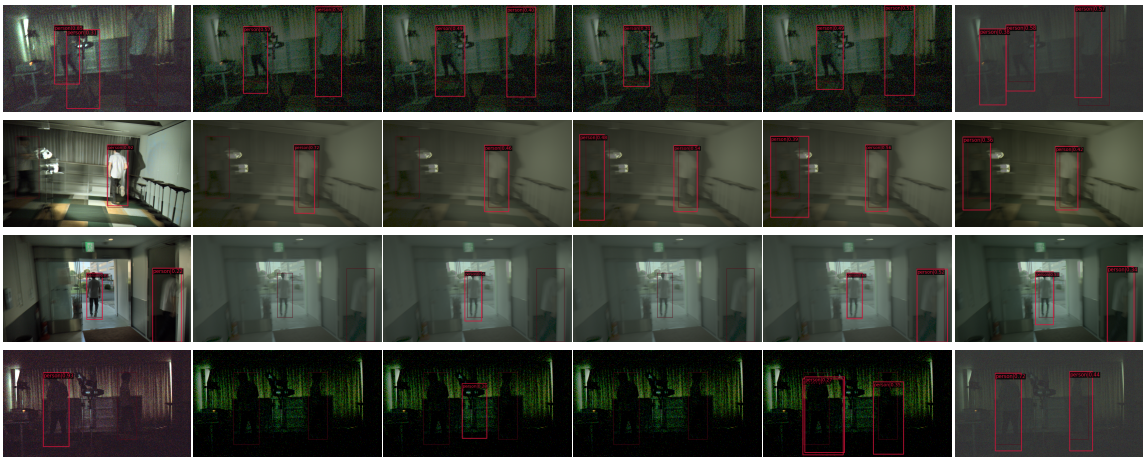
References

[1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 2.1

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2.1

[3] Zili Liu, Tu Zheng, Guodong Xu, Zheng Yang, Haifeng Liu, and Deng Cai. Training-time-friendly network for real-time object detection. In *proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 11685–11692, 2020. 2.1

[4] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*, 2020. 2.1, 1



ISP	black-box ISP	simplest gamma	simplest gamma	simplest gamma	simplest gamma	parameterized gamma
augmentation order	after ISP	after ISP	before ISP (ours)	before ISP (ours)	before ISP (ours)	before ISP (ours)
noise-accounted	w/o	w/o	w/o	w/o prior	w/ (ours)	w/ (ours)

Figure 1. The visualization of the detection results. To make a fair comparison, we set an adequate confidence threshold per model. Specifically, we adjuste the threshold to achieve a precision@0.5 value of 80%. The darker bounding boxes represent the ground truth, while the brighter ones represent the prediction result.