**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>
<thead>
<tr>
<th></th>
<th>augmentation</th>
<th>noise</th>
<th>black-box ISP mAP@0.5:0.95 [%]</th>
<th>simplest ISP mAP@0.5:0.95 [%]</th>
<th>parameterized ISP mAP@0.5:0.95 [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Color</td>
<td>after</td>
<td>-</td>
<td>51.6</td>
<td>40.2</td>
<td>-</td>
</tr>
<tr>
<td>+ Blur</td>
<td>before</td>
<td>-</td>
<td>-</td>
<td>46.8</td>
<td>47.5</td>
</tr>
<tr>
<td>Blur</td>
<td>(ours)</td>
<td>ours</td>
<td>-</td>
<td>51.5</td>
<td>52.0</td>
</tr>
</tbody>
</table>

3. Acknowledgements

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

**References**


Figure 1. The visualization of the detection results. To make a fair comparison, we set an adequate confidence threshold per model. Specifically, we adjust 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.