Guiding a Harsh-Environments Robust Detector via RAW Data Characteristic Mining

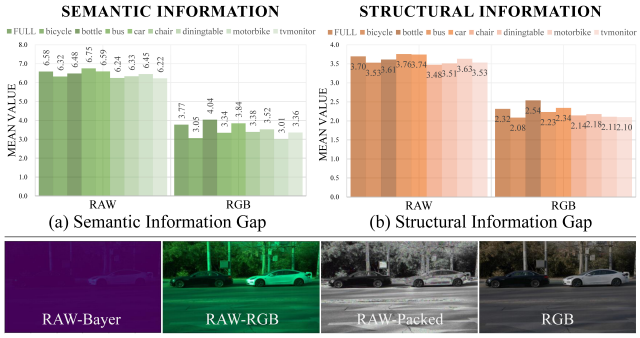
Hongyang Chen[1], **Hung-Shuo Tai**[2], **Kaisheng Ma**[3*]
[1]Xi'an Jiaotong University, Xi'an, China
[2]KargoBot.ai, Beijing, China
[3]Tsinghua University, Beijing, China
chenhy@stu.xjtu.edu.cn, hungshuotai@didiglobal.com, kaisheng@mail.tsinghua.edu.cn

Abstract

Consumer-grade cameras capture the RAW physical description of a scene and then process the image signals to obtain high-quality RGB images that are faithful to human visual perception. Conventionally, dense prediction scenes require high-precision recognition of objects in RGB images. However, predicting RGB data to exhibit the expected adaptability and robustness in harsh environments can be challenging. By capitalizing on the broader color gamut and higher bit depth offered by RAW data, in this paper, we demonstrate that RAW data can significantly improve the accuracy and robustness of object detectors in harsh environments. Firstly, we propose a general Pipeline for RAW Detection (*PRD*), along with a preprocessing strategy tailored to RAW data. Secondly, we design the RAW Corruption Benchmark (*RCB*) to address the dearth of benchmarks that reflect realistic scenarios in harsh environments. Thirdly, we demonstrate the significant improvement of RAW images in object detection for low-light and corrupt scenes. Specifically, our experiments indicate that *PRD* (using FCOS) outperforms RGB detection by 13.9mAP on LOD-Snow without generating restored images. Finally, we introduce a new nonlinear method called Functional Regularization (*FR*), which can effectively mine the unique characteristics of RAW data. The code is available at https://github.com/DreamerCCC/RawMining.

Introduction

Background The recent advancements in deep neural networks (DNNs) (Liu et al. 2021; Yao et al. 2023) and image signal processing (ISP) (Wu et al. 2019; Chen and Ma 2022) have led to significant progress in low-level vision imaging tasks, such as image enhancement and image deraining. However, while these techniques have made fundamental breakthroughs in the field, they do not always translate into valuable solutions for high-level visual recognition tasks (VidalMata et al. 2020; Al Sobbahi and Tekli 2022). Gradually, the research community and industry have reached a consensus that there is a discrepancy between human-driven and machine-driven imaging, whereby the visual perception and interpretation of images by humans differs from that of models. On the one hand, ISP is used to receive the raw signal RAW data (Wang et al. 2023) of the sensor for processing the whole camera pipeline. Although the image enhancement algorithm driven by the human eye sense can improve the visual quality, more is needed to solve the object recognition problem (Li et al. 2017; Banerjee et al. 2021). On the other hand, visual recognition tasks are represented by object detection (Zhang et al. 2022b; Zou et al. 2023) and semantic segmentation (Zhang et al. 2022a). To bridge the gap with computational photography in recent years, the community has tried to post-process the acquired RGB images in specific situations (Shyam et al. 2022; Xu et al. 2023), such as object detection in the haze (Pham et al. 2022; Wu et al. 2022).



(a) Semantic Information Gap (b) Structural Information Gap

Figure 1: (Upper panel) Statistical comparison of RAW and RGB data. (a) Semantic (Shannon Entropy): Image entropy measures information content and is used to assess image complexity. (b) Structural (CEIQ-Contrast (Yan, Li, and Fu 2019)): Based on the contrast assumption, high-contrast images are always more similar to contrast-enhanced images. Tested on the full dataset of LOD (Hong et al. 2021). (Lower panel) Visual comparison of RAW versus RGB data.

Challenges Deep learning (DL) based object detection methods (Zhou, Wang, and Krähenbühl 2019; Tian et al. 2019; Tan, Pang, and Le 2020) face challenges in achieving consistent high efficiency in harsh environments (Bi et al. 2022; Sun et al. 2022). These methods are limited in their ability to effectively handle the variability and complexity of environmental factors such as lighting conditions, weather, and occlusions, which can significantly affect the accuracy and robustness of detection algorithms. To overcome the interference of extreme weather conditions on the camera in

* Corresponding author.

autonomous driving, the system's adaptability to different situations is usually enhanced by combining technologies such as lidar and millimeter-wave radar. Tesla Vision focuses on a solely camera-based approach, enabling a new solution through multi-camera and multi-sensor fusion (Talpes et al. 2020; Ajitha and Nagra 2021). However, in monocular vision scenarios, object detectors have not had the opportunity to exhibit more robust performance in harsh environments: Firstly, inadequate lighting will lead to quality degradation problems such as low brightness and poor contrast of RGB images. Secondly, deep models always fall into different corruption scenarios when densely predicting images while facing the domain gap between the simulation of corruption scenarios and the authentic situation. Thirdly, a corruption benchmark that can reflect natural physical laws is critical for finding general improvement of deep models in harsh environments (Chen et al. 2021).

Rethinking: Can RAW be employed for object detection, and how to use it?

It is widely recognized that cameras are precise instruments for measuring light. In many applications, it is assumed that the radiance of an image is directly related to the radiance of the scene being captured (Huang et al. 2022). This assumption is based on the principle of radiometry (Shaw 2013), which states that the amount of light energy captured by a camera sensor is proportional to the amount of light energy emitted by the scene being imaged. The goal of ISP is to produce visually pleasing images that are faithful to the photographer's intent, rather than to provide a precise representation of the physical properties of the scene. Figure 1 demonstrates that RAW images contain richer semantic and structural information than RGB. RAW is the original data that the sensor converts the captured light source signal into a digital signal, including the original color information of the object, *etc*. Generally, the RAW data format used in cameras is based on the Bayer arrangement (Richter and Fößel 2019). Since improvements in RGB image quality do not always lead to better performance in high-level recognition tasks (VidalMata et al. 2020; Al Sobbahi and Tekli 2022; Banerjee et al. 2021; Li et al. 2017), and RAW data contain richer and more accurate information about the physical properties of the scene, making it an invaluable resource for developing robust object detection algorithms. In this paper, we investigate the potential of mining RAW data characteristics for improving the performance of object detectors, focusing on three key research questions:

(1) *Why is RAW data required?* We propose a general Pipeline for RAW Detection (*PRD*) to adapt to different off-the-shelf object detectors. Specifically, the network takes RAW Bayer as input, skips the traditional ISP, and directly produces the prediction result of the target task. By exploring the preprocessing of RAW data, as well as different formats of model input, normalization, and training strategies, we validate the performance of *PRD* at nighttime. We demonstrate superiority across different detectors and over other pipelines, where CenterNet improves nighttime RGB detection accuracy and inference speed by 3.0mAP and 3.02x, respectively, on the LOD (Hong et al. 2021).

(2) *A reliable RAW corruption benchmark?* The gap between methods for artificially synthesizing corruption scenes and the natural world reduces confidence in the results. The community has always needed a corruption benchmark to reflect natural physical laws in RAW data. We design the RAW Corruption Benchmark (*RCB*) to fill the gap where RAW data reflect realistic scenarios in harsh environments. We verify the considerable improvement of the RAW image on the corrupt scenes. On LOD-Snow, FCOS (Tian et al. 2019) on *PRD* is 13.9mAP higher than RGB detection.

(3) *How can we best utilize the information in RAW data?* The number of photons received by the camera is typically collected from multiple positions within the sensor, and usually, the RAW data maintains a linear relationship at different exposure levels. Unlike the 8-bit sRGB color space, RAW data records pixel-level bit depths typically ranging from 10-16 bits, allowing for a wider range of potential values to be captured and processed. We propose a novel nonlinear method called Functional Regularization (*FR*) to exploit the unique characteristics of RAW data further.

To the best of our knowledge, this paper is the first to comprehensively study the advantages of RAW data in harsh-environments object detection, as well as the feasible processing procedures and methods. By capitalizing on the broader color gamut and higher bit depth offered by RAW data, our results demonstrate that object detection performance can be significantly improved, especially in settings with challenging lighting conditions or other types of interference. The findings of this study have implications for the development of more effective and reliable object detection systems in a range of practical applications.

Preliminaries and Related Work

RAW Data Characteristics

Physical Analysis Compared with the processed sRGB image, the RAW file has the following good properties:

- Data linearity (Huang et al. 2022). The photon count collected from different locations in the sensor maintains a linear relationship at different exposure levels.
- Shooting parameters (Kroon-Batenburg et al. 2017). A RAW file usually contains sensor and metadata, such as different camera parameter configurations (camera-specific) and shooting parameters.
- Greater information content and higher bit depth (Bauer and Becker 2011). The RAW file contains a lot of resolution, density, and color information, while the RGB final output is only 8-bit level.

The above characteristics disappear when the RAW files are processed into final sRGB images. Most image processing systems serve human vision perceptual quality. Therefore, the subsequent adjustment stages in processing based on human vision are conducted, *e.g.* white balance, tone mapping, and gamma correction.

Prior Arts

RAW Data Processing RAW data is a kind of information that records the digital camera sensor and also records the meta-data generated by the camera, such as ISO, wide door speed, aperture value, *etc*. Research on RAW data in low-level vision has primarily focused on denoising (Zhang et al. 2021) and synthesis of RAW images (Xing, Qian, and Chen 2021; Zhou et al. 2021). Although there have been studies on the application of RAW data in downstream tasks (Ljungbergh et al. 2023; Morawski et al. 2022), the main focus has been on improving RAW detection performance in normal scenarios, often requiring additional parameters or branches in the pipeline. Additionally, the integration of camera-specific physical characteristics into object detection tasks also has garnered attention in the research community. The importance of dynamic range for RAW object detection was first demonstrated by (Xu et al. 2023), which introduced the HDR RAW Pipeline. Shyam et al. proposed an image restoration architecture suitable for sRGB and RAW images, showcasing improved detection performance under low-light conditions through a two-stage training process (Shyam et al. 2022). Rawgment (Yoshimura et al. 2022), on the other hand, is a data augmentation method designed explicitly for RAW images, combining color jittering and blur enhancement. The key focus of our work, however, is to explore the robustness of raw data in challenging object detection environments.

Low-light Object Detection Different from general object detection, low-light object detection has been regarded as a new topic rather than a special scene. With the emergence of ExDark (Loh and Chan 2019), LOD (Hong et al. 2021), and RAW-NOD (Morawski et al. 2022) datasets, methods based on sRGB data and RAW data have also been proposed. For sRGB data, they can be regarded as degraded images and supplemented with specific preprocessing modules (Jiang et al. 2022) or joint training strategies (Guo, Lu, and Wu 2021). For RAW data, researchers find that powerful smart ISP models can replace traditional ISPs to improve detection accuracy, and even RAW images can be directly used to train recognition models. Hong et al. proposed a detector based on RAW images with excellent performance under low light (Hong et al. 2021). Our research on RAW data is aimed at addressing both theoretical and practical gaps in the area of object detection under challenging conditions.

Benchmarks for Corruption Since Dodge and Karam studied the fragile performance of deep recognition models under noise and blur (Dodge and Karam 2016), a series of research on corruption robustness based on deep learning methods has been launched. Benchmarks on image classification (Hendrycks and Dietterich 2019), object detection (Michaelis et al. 2019), and person re-identification (Chen, Wang, and Zheng 2021) are proposed to evaluate the performance of models on common corruption. Studies have shown that in image classification (Azulay and Weiss 2018), object detection (Chen et al. 2021), and other tasks (Kamann and Rother 2020), the model always suffers considerable accuracy loss on corrupt images. However, in common dense prediction scenarios, it has been a challenge to obtain a model that balances corruption robustness and performance without complex process design and multi-task training (Fan et al. 2022). In this paper, we propose the corruption benchmark on RAW data for the first time and discuss RAW's superior performance on corruption robustness in general object detection.

Why is RAW Data Required? & A Reliable RAW Corruption Benchmark

Pipeline for RAW Detection (*PRD*)

Based on the analysis presented in Sec. *Introduction* and the statistical comparison shown in Figure 1, we have concluded that RAW data can offer a more precise and immediate representation of object information in low-light conditions while also capturing the physical features of the scene with accuracy. Recent studies investigating the use of RAW data for object detection have revealed that only a handful of operations within the traditional ISP pipeline actually improve the performance of high-level visual tasks (Ljungbergh et al. 2023). However, our experiments have demonstrated that preprocessing RAW data into 8-bit 3-channel data (RAW-RGB style) is suboptimal, primarily due to the excessive discarding of important RAW data features. As a result, we propose the RAW detection pipeline (*PRD*), as illustrated in Figure 2, which involves dynamic preprocessing of RAW data and integration with a general detector. Specifically, (a) the RAW preprocessing step involves packaging the RAW-Bayer input ($H \times W \times 1$) into RAW-Packed ($\frac{H}{2} \times \frac{W}{2} \times 4$) rather than interpolating it into RGB (RAW-RGB[1]: $H \times W \times 3$), as described in the following ablation comparison. (b) To adjust the model's ability to represent RAW images, we utilize a normalization range based on the number of bits, as higher pixel value ranges in RAW data correspond to richer semantics. For instance, the normalized range for a 14-bit RAW image is [0, 2^6). (c) In terms of the detector, the first layer of the model must be adapted to RAW-Packed by changing the number of channels in the first layer to 4 while retaining the same structure and supervision functions. Notably, *PRD* does not require metadata.

Normalization Range Adjusting the normalization range of the input is crucial not only to ensure the model's capability to represent high-bit-depth data but also to enhance the pretrained model's adaptability to the unique characteristics of RAW images. For instance, RGB images (8-bit) are typically normalized to the range [0, 1]. However, this range becomes inadequate for representing the rich information in 14-bit RAW images. Therefore, we base our normalization on the 8-bit range and incrementally expand the processing range of the data during preprocessing.

Datasets and Pre-training Settings To evaluate the real-world performance of low-light detection, we utilize the LOD (Hong et al. 2021) dataset, which consists of 2230 image pairs that are randomly split into a training set of 1830 pairs and a test set of 400 pairs. The RAW-NOD (Morawski et al. 2022) dataset contains 7K raw images captured in outdoor low-light conditions. PASCALRAW (Omid-Zohoor,

[1]The detector takes a demosaiced 3-channel RAW-RGB as input to ensure detector compatibility with sRGB.

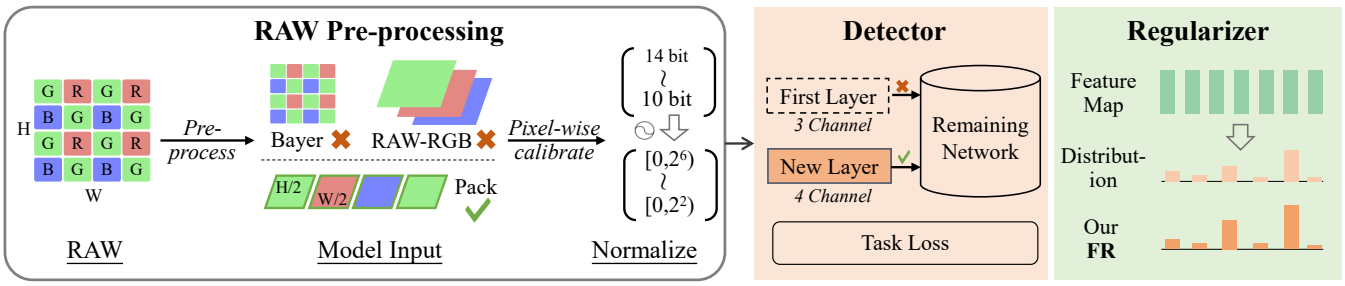


Figure 2: The pipeline of RAW Detection (*PRD*). (i) The preprocessing of RAW Bayer needs to go through "Pack" and "Normalize" to complete the packaging of $H \times W \times 1$ to $\frac{H}{2} \times \frac{W}{2} \times 4$ and normalize the range of pixels. The Bayer and RAW-RGB (Hong et al. 2021; Morawski et al. 2022; Ljungbergh et al. 2023) formats in related works are shown to be sub-optimal. (ii) Adjusting the object detector only requires modifying the number of channels in the first layer to match the input of RAW data. (iii) We propose Functional Regularization (*FR*) to exploit the unique characteristics of RAW data further.

Ta, and Murmann 2014) contains 4,259 annotated RAW images, with three annotated object classes (car, person, and bicycle), and is modeled after the PASCAL VOC database. The detector pre-training process adheres to the default configuration of MMDetection (Chen et al. 2019). For instance, when the "Pre-trained" is set to true for CenterNet, irrespective of the data type (RAW or RGB), the original pre-trained model is obtained from training on the MS COCO dataset, following the default configuration.

Data Preprocessing and Implementation Details All data annotation formats follow the COCO standard. The preprocessing process in *PRD* does not perform additional processing on the RAW image, such as using the white balance in the camera's saved parameters to prevent color cast. We implement our approach using the Open MMLab Detection Toolbox (Chen et al. 2019) and PyTorch, running on 8 RTX NVIDIA 2080Ti GPUs (12GB). We follow the official default settings of detectors, *e.g.* for CenterNet, use RandomCenterCropPad and RandomFlip as data augmentation.

RAW Corruption Benchmark (*RCB*)

Benchmarking RAW Corruption: Existing Challenges The discrepancy between synthetic and authentic scenarios often leads to less confidence in model predictions. Commonly, deep learning-based methods face difficulty maintaining high efficiency in harsh environments. This is primarily attributed to the domain gap between the training and test datasets. Additionally, deep models are typically designed to capture complex features and patterns in the data, which can limit their capability to identify and mitigate corruption outside the scope of the training data. As the opportunities to harness the potential of RAW data continue to increase, there is an urgent need for a reliable RAW corruption benchmark to address these challenges. Current corruption benchmarks (Hendrycks and Dietterich 2019; Chen, Wang, and Zheng 2021; Chen et al. 2021) typically consist of several types of blur, noise, weather, and digital. Each corruption type is parameterized with five severity levels. However, these corruption definitions are all based on the RGB level, making it impractical to apply them to RAW data directly.

RAW Corruption Deep models are susceptible to noise, and previous noise models have often been simplistic, involving evenly distorting images with Gaussian noise. However, real image noise differs from the noise generated by these simple models, as it is a combination of multiple types of noise, such as photon noise, kTC noise, and dark current noise. While RGB data is often subject to non-linear noise or digital corruption, the application scenarios of RAW data are typically oriented toward real interference. To this end, we have developed a RAW corruption benchmark that falls into two broad categories: (1) blur, including defocus, glass, motion, and zoom, and (2) weather, including snow, frost, fog, and bright.

(1) *RAW noise characteristics:* After passing through the ISP, the noise properties will become more complex and challenging to process. Specifically, first, lens shading correction will significantly enhance the noise at the edge of the image. Second, demosaic will turn image noise into structural noise, such as unique pattern noise. Third, Gamma correction leads to nonlinear changes, and AWB leads to linear shifts in noise. Fourth, CCM's color gamut space conversion enhances noise correlation and worsens the visual noise effect.

(2) *Dataset collection and preprocessing:* To ensure that the dataset is sufficiently large and representative of real-world scenarios, we combine the dataset: LOD (Hong et al. 2021), RAW-NOD (Morawski et al. 2022) and PASCALRAW (Omid-Zohoor, Ta, and Murmann 2014). The image preprocessing steps follow the settings outlined in our *PRD*.

(3) *Corruption transformation:* Previous benchmarks involve generating corrupted sRGB images. In *RCB*, we generate perturbations for RAW-RGB. Our methodology is inspired by previous works (Zhang et al. 2021) and (Michaelis et al. 2019). We adopt the severity settings outlined in (Hendrycks and Dietterich 2019).

Evaluation Metric To unify performance evaluation across diverse datasets in corruption scenarios, we propose an enhanced metric $\mathcal{Q}$:

$$\mathcal{Q} = \begin{cases} AP(\%) & \text{for LOD \& RAW-NOD,} \\ AP_{50}(\%) & \text{for PASCALRAW.} \end{cases} \quad (1)$$

Items	Data Type	Pre-trained	Normalize	Range	mAP	mAP@50	mAP@75	mAP@S	mAP@M	mAP@L
Pre Processing Format	RGB	✗	✗	[0,1)	**24.6**	**41.5**	**25.8**	0.0	**10.6**	**29.1**
	RAW-Bayer	✗	✗	[0,1)	17.8	32.2	18.3	0.0	4.7	22.0
	RAW-RGB	✗	✗	[0,1)	22.6	38.7	23.6	0.0	7.8	27.4
	RAW-Packed	✗	✗	[0,1)	23.4	40.6	24.4	0.7	6.1	28.6
	RGB	✗	✓	[0,1)	**25.0**	**40.7**	**26.9**	0.0	**12.9**	**29.8**
	RAW-Bayer	✗	✓	[0,1)	21.2	37.2	21.6	0.4	6.2	26.0
	RAW-RGB	✗	✓	[0,1)	22.8	39.7	23.5	0.6	7.9	27.5
	RAW-Packed	✗	✓	[0,1)	23.7	40.4	24.1	0.0	7.7	28.8
Higher Range	RAW-Bayer	✗	✓	[0,4)	24.7	43.1	25.1	0.0	7.3	29.8
	RAW Packed	✗	✓	[0,4)	**25.7**	**43.5**	**27.0**	0.0	**8.3**	**31.2**
	RAW-Bayer	✗	✓	[0,8)	24.1	42.0	24.8	0.0	6.5	29.3
	RAW Packed	✗	✓	[0,8)	**26.7**	**45.0**	**28.2**	0.0	**11.5**	**31.9**
	RAW-Bayer	✗	✓	[0,64)	24.9	43.7	25.4	0.7	7.8	30.0
	RAW Packed	✗	✓	[0,64)	**28.0**	**47.5**	**29.1**	0.1	**10.4**	**33.3**

Table 1: *PRD* ablation results on pre-processing for RAW data. The experiments are performed on CenterNet (Zhou, Wang, and Krähenbühl 2019), LOD (Hong et al. 2021). RAW-RGB represents the demosaic pre-processed RAW image (3 channels, 8-bit), RAW-Bayer, and RAW-Packed are from raw data (1 channel and 4 channels, 14-bit).

This approach maintains the standard measures of AP_{50} and AP, allowing for consistent evaluation across datasets. For a holistic assessment of corrupted scenarios, we introduce the composite corruption performance:

$$\mathcal{CCP} = \frac{1}{N_c \times N_s} \sum_{c=1}^{N_c} \sum_{s=1}^{N_s} \mathcal{Q}_{c,s}. \tag{2}$$

$\mathcal{Q}_{c,s}$ denotes the performance metric tailored to each dataset, assessed for corruption type c and its severity s. This encompasses 8 distinct corruption categories (*i.e.*, $N_c = 8$) and spans 5 levels of severity (*i.e.*, $N_s = 5$).

Where is RAW Better than RGB?

In most DNNs designed for RAW-to-RGB mapping, the Bayer mosaic pattern present in the RAW input image is typically removed by stacking each 2×2 block in the original image in four input channels. This approach ensures translational color invariance in each channel and biases the input interpretation toward color separation. One advantage of using the stacked method is that, for a fixed kernel size, the receptive field of the first layer is doubled compared to the flat input, as encountered in dilated convolutions. To address the issue of selecting an appropriate form of RAW input, Table 1 compares the effects of using RAW-Bayer, RAW-RGB, and RAW-Packed inputs. Additionally, to fully leverage the rich information provided by the high-bit characteristics of the RAW data, we also vary the normalization range during the preprocessing stage. *Pre-trained* is somewhat unfair to RAW-Packed (detector receives 3-channel sRGB for pre-training, but receives 4-channel RAW-Packed for fine-tuning), so we compare without it. Under the same normalization standard, RAW Packed (14-bit, 28.0mAP) surpasses RGB detection (8-bit, 25.0mAP) by 3.0mAP. Due to our inability to establish the accuracy of the RGB inverse transformation method, we refrain from drawing conclusions about the performance of RAW versus RGB detection from the unprocessed comparison experiments. We next present experimental results, including the detection performance of RAW images in low-light and corrupt scenes, as well as a comparison regarding inference speed and training overhead.

Night-time To validate the unique advantages of RAW data identified through our statistical and physical analyses, we conduct *PRD* experiments on the low-light LOD with various detectors. In CenterNet (Zhou, Wang, and Krähenbühl 2019), RAW detection achieves an accuracy improvement of 3.0mAP over RGB detection (from 25.0 to 28.0 mAP), and in EfficientNet-B3 (Tan and Le 2019), the improvement is 4.6mAP (from 24.5 to 29.1 mAP). This demonstrates that utilizing RAW images instead of sRGB images can enhance detection precision under low-light conditions. These results highlight the effectiveness of RAW-input design in enabling the detector to extract signals that would otherwise be lost or degraded when using sRGB inputs with low signal-to-noise ratios (SNR).

Corruption Scenarios Our evaluation using the *RCB* reveals the distinct advantages of *PRD* in handling corrupt scenes, as compared to standard RGB detection. As shown in Table 2, RAW data consistently outperforms RGB in resisting corruption (the reason for using RAW-RGB is that the performance of the two data in the same data format can be compared fairly). Notably, in terms of corruption robustness, RAW detection significantly surpasses RGB in both CNN-based and Transformer-based detectors, including FCOS and DETR, with improvements of 6.95mAP and 7.96mAP, respectively. This superiority holds even though Transformer models show greater robustness in RGB scenarios under varying weather conditions (Zhou et al. 2022). Please note that during the training period, the data used also includes corrupted scene data.

Domain Shift Our investigations into the inference capabilities of normally trained detectors across varied environ-

Scenarios	Corruption	Data Type	FCOS mAP	FCOS mAP@50	FCOS mAP@75	DETR mAP	DETR mAP@50	DETR mAP@75
Normal	Night Time	RAW	44.0	71.0	47.1	44.8	71.5	47.2
		RGB	44.4	69.9	47.1	46.5	72.4	50.4
Weather	Snow	RAW	38.7 (*-5.30*)	64.4	39.8	36.2 (*-8.60*)	60.3	39.1
		RGB	24.8 (*-19.2*)	45.8	23.9	20.0 (*-26.5*)	37.6	18.5
	Frost	RAW	38.3 (*-5.70*)	64.6	39.1	37.8 (*-7.00*)	62.5	40.0
		RGB	19.2 (*-25.2*)	38.1	17.1	12.7 (*-33.8*)	27.1	9.90
	Fog	RAW	42.9 (*-1.10*)	69.3	45.2	42.6 (*-2.20*)	68.2	47.1
		RGB	34.8 (*-9.60*)	57.6	36.0	30.9 (*-15.6*)	52.6	31.3
	Contrast	RAW	42.1 (*-1.90*)	68.4	44.6	42.0 (*-2.80*)	68.7	44.7
		RGB	39.2 (*-4.80*)	61.9	42.1	39.2 (*-7.30*)	61.9	42.1
Blur	Brightness	RAW	43.0 (*-1.00*)	69.3	46.5	42.4 (*-2.40*)	68.4	46.3
		RGB	39.7 (*-4.70*)	65.4	41.2	37.7 (*-8.80*)	64.2	38.4
	Defocus Blur	RAW	43.6 (*-0.40*)	69.3	47.3	41.8 (*-3.00*)	68.5	44.8
		RGB	41.2 (*-3.20*)	66.2	43.2	39.7 (*-6.80*)	66.6	41.4
	Motion Blur	RAW	43.9 (*-0.10*)	69.7	47.9	43.1 (*-1.70*)	70.2	45.6
		RGB	43.3 (*-1.10*)	68.0	45.8	42.5 (*-4.00*)	67.8	45.4
	Zoom Blur	RAW	31.3 (*-12.7*)	55.2	32.1	30.7 (*-14.1*)	54.8	31.3
		RGB	26.0 (*-18.4*)	49.8	24.5	30.2 (*-16.3*)	52.0	29.2

Table 2: *PRD* results under corruption scenarios, fine-tuned with the pre-trained FCOS (Tian et al. 2019) and DETR (Carion et al. 2020) on LOD (Hong et al. 2021). RAW here refers to RAW-RGB that has been demosaic preprocessed. The RAW samples in the corruption scenarios are obtained by *RCB*. Corruption is implemented with severity=2.

mental conditions revealed a significant performance degradation in RGB detection, starkly contrasting to RAW detection, which demonstrated relative resilience. Using FCOS, initial RAW and RGB performances are 44.0 and 44.4 mAP, respectively. In snowy conditions, RAW and RGB fell to 16.4 and 1.2 mAP; in frost, to 18.8 and 2.1 mAP; and in fog, to 28.4 and 20.1 mAP.

Inference Speed and Training Overhead Firstly, inference speed advantage. In *PRD*, when processing RAW-Packed data, the speed of detector inference samples has also been multiplied. For instance, under an RTX NVIDIA 2080Ti, CenterNet processes RAW and RGB at speeds of 18.4 and 55.6 images per second (img/s), respectively, achieving a 3.02-fold increase. Secondly, the memory usage during model training is basically unchanged. The training load does not escalate when employing detectors like FCOS and EfficientDet, among others.

Discussion on the Data Volume Required for Pre-training Concerning the volume of data required for pre-training, we contend that it largely hinges on the diversity and representativeness of the dataset rather than sheer quantity. Since RGB data benefits from ISP compression and optimization, it facilitates quicker learning of effective features, implying that RAW data may require a larger dataset for model pre-training.

Summary: *RAW data has shown significant promise in object detection, offering numerous advantages. PRD enables current detectors to accurately predict on RAW data, bypassing the need for additional RGB reconstruction or denoising branches. In low-light scenarios, RAW detection notably surpasses RGB, especially as PRD excludes traditional ISP modules. Furthermore, PRD maintains training efficiency comparable to RGB detection while tripling inference speed. Our benchmarking of RAW data's robustness reveals PRD's effectiveness in corrupt scenes, with RAW detection outperforming RGB by average margins of 6.95mAP and 7.96mAP for FCOS and DETR, respectively. These findings underscore RAW data's enhanced performance and PRD's potential to bolster object detection robustness in challenging real-world scenarios.*

How Can We Better Utilize The Information in RAW Data?

To improve the mapping and representation of RAW-specific data in neural networks, one feasible approach is to enhance the nonlinear activation function in the detector. However, an activation function tailored for RAW data requires careful consideration of several factors. Firstly, RAW data often undergoes nonlinear noise that can be challenging to model accurately. Secondly, RAW data has a high dynamic range, and designing an activation function that can effectively capture the full range of values in the data can be difficult. Thirdly, the specific task being addressed is another essential consideration in choosing an activation function for RAW data.

Previous Functions Across theoretical research into activation functions, those sharing properties similar to Swish (Ramachandran, Zoph, and Le 2017), which includes non-monotonicity, the ability to preserve small negative weights and a smooth profile. For instance, GELU (Hendrycks and Gimpel 2016) and Mish (Misra 2019). Among them, GELU introduces random regulariza-

Range	Activation	Accuracy		
		mAP	mAP@50	mAP@75
[0,4)	ReLU	25.7	43.5	27.0
	FR	$28.5_{+2.8\%}$	$47.2_{+3.7\%}$	$30.3_{+3.3\%}$
[0,8)	ReLU	26.7	45.0	28.2
	FR	$27.9_{+1.2\%}$	$46.9_{+1.9\%}$	$29.1_{+0.9\%}$
[0,64)	ReLU	28.0	47.5	29.1
	FR	$28.7_{+0.7\%}$	$48.2_{+0.7\%}$	$30.3_{+1.2\%}$

Table 3: Ablation of *FR* in the *PRD* (normalize range). The input format: RAW-Packed. Detector: CenterNet.

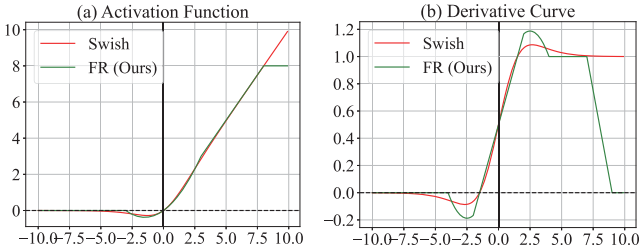


Figure 3: The activation functions of the Swish (Ramachandran, Zoph, and Le 2017) and the *FR* (Ours) (left panel), and the derivatives of these functions (right panel).

tion and retains the function's dependence on the input. Heuristically, we also want not to be affected too much by outlier noise when larger inputs are retained.

Functional Regularization (*FR*) To enable effective processing of high-dimensional and information-rich RAW data, we develop a new regularization function that combines ReLU6 (Krizhevsky and Hinton 2010) and H-Swish (Howard et al. 2019). Pixels with a higher bit range typically contain more semantic information, but they may also contain more noise in the form of RAW data. To address this issue during training, we have modified the activation function to mitigate the noise by using a bounded activation function with stronger regularization. This modification also helps to solve the problem of large negative inputs. Firstly, it should have no lower bounds to avoid gradient saturation, which can cause a sharp drop in training speed. Additionally, incorporating non-monotonicity can help keep negative values small, thereby stabilizing the gradient flow. Furthermore, due to the preservation of a small amount of negative information, *FR* eliminated the preconditions necessary for the Dying ReLU phenomenon by design.

FR, additionally, is non-monotonic, smooth, and preserves a small amount of negative weights. These properties contribute to the consistent performance and improvement observed when using *FR* in place of ReLU in networks (The 1st derivative of *FR* is shown in Figure 3):

$$FR : y = min\left(x \cdot \frac{\text{ReLU}\,6(x+3)}{6}, 8\right) \tag{3}$$

Inspired by (Howard et al. 2019), we add a maximum limit to the eigenvalue in *FR*, which was initially designed to measure the fixed-point value of the detector. However, with higher-bit raw values, we find that setting a limit of 8 for the eigenvalue can provide stronger regularization.

Format	Activation	Accuracy		
		mAP	mAP@50	mAP@75
RAW-Bayer	ReLU	21.2	37.2	21.6
	FR	$21.9_{+0.7\%}$	$37.2_{+0.0\%}$	$22.6_{+1.0\%}$
RAW-RGB	ReLU	22.8	39.7	23.5
	FR	$25.1_{+2.3\%}$	$42.4_{+2.7\%}$	$26.8_{+3.3\%}$
RAW-Packed	ReLU	23.7	40.4	24.1
	FR	$25.8_{+2.1\%}$	$43.3_{+2.9\%}$	$27.8_{+3.7\%}$

Table 4: Ablation of *FR* in the *PRD* (RAW format). The normalize range: [0,1). Detector: CenterNet.

Confirmatory Experiment The *FR* can produce a strong regularization effect by reducing the sensitivity of neurons to large values and minimizing differences in the upper and lower areas. The main experiment is CenterNet (Zhou, Wang, and Krähenbühl 2019) done by *PRD* on LOD. As shown in Table 3 and 4, without exploding the memory bandwidths, *FR* can effectively improve the performance of the detector in *PRD*. Compared with ReLU6 and H-Swish, *FR* brings CenterNet an improvement of 0.3 and 0.8 mAP under night-time LOD, respectively. We conduct an in-depth investigation into the optimal of *FR* to further enhance performance improvements. We have made the following observations: (1) Neither ReLU6 nor H-Swish outperforms ReLU in the head. (2) The appropriate component positions for ReLU6 and H-Swish differ. (3) Achieving optimal results involves replacing the activation function with our *FR* in the initial layers of both the backbone and neck.

Summary: *In contrast to RGB detection, PRD employs a dynamic amplification of input normalization from a 0-1 distribution. Additionally, to further enhance the performance of RAW detection, we develop a method called FR, which leverages the unique properties of RAW data to improve the detector's sensitivity to pixels and features.*

Conclusion

Object detectors in monocular vision scenes suffer from interference from harsh environments. Although RAW data is limited to a specific sensor, it includes more color gamut and a higher bit depth. Nevertheless, the main problems are (1) Proving the rationality and advantages of RAW application in object detection. (2) The performance of different methods varies significantly across evaluation criteria, especially in corruption scenarios. (3) Design appropriate methods to cope with the RAW characteristics. In this paper, we aim to tackle these challenges. Experimental results demonstrate the superiority of RAW in nighttime and corruption scenarios, and the regularization of the nonlinear method also improves the feature representation ability of the model.

Potential Negative Impact Model training requires extensive camera data for diverse scenes, lighting, and devices. However, this data collection process could violate image rights without stringent regulatory guidelines.

Acknowledgements

This work was supported by the National Key R&D Program of China (2022YFB2804103), the Key Research and Development Program of Shaanxi (2021ZDLGY01-05), the National Natural Science Foundation of China (20211710187, 31970972), Tsinghua University Dushi Program, Tsinghua University Talent Program, Institute for Interdisciplinary Information Core Technology (IIISCT) and Ant Group through CCF-Ant Research Fund.

References

Ajitha, P.; and Nagra, A. 2021. An Overview of Artificial Intelligence in Automobile Industry–A Case Study on Tesla Cars. *Solid State Technology*, 64(2): 503–512.

Al Sobbahi, R.; and Tekli, J. 2022. Comparing deep learning models for low-light natural scene image enhancement and their impact on object detection and classification: Overview, empirical evaluation, and challenges. *Signal Processing: Image Communication*, 116848.

Azulay, A.; and Weiss, Y. 2018. Why do deep convolutional networks generalize so poorly to small image transformations? *arXiv preprint arXiv:1805.12177*.

Banerjee, S.; VidalMata, R. G.; Wang, Z.; and Scheirer, W. J. 2021. Report on ugˆ2+ challenge track 1: Assessing algorithms to improve video object detection and classification from unconstrained mobility platforms. *Computer Vision and Image Understanding*, 213: 103297.

Bauer, S.; and Becker, C. 2011. Automated preservation: the case of digital raw photographs. In *Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation: 13th International Conference on Asia-Pacific Digital Libraries, ICADL 2011, Beijing, China, October 24-27, 2011. Proceedings 13*, 39–49. Springer.

Bi, X.; Gao, H.; Chen, H.; Wang, P.; and Ma, C. 2022. Evaluating the Robustness of Object Detection in Autonomous Driving System. In *2022 9th International Conference on Dependable Systems and Their Applications (DSA)*, 645–649. IEEE.

Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; and Zagoruyko, S. 2020. End-to-end object detection with transformers. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, 213–229. Springer.

Chen, H.; and Ma, K. 2022. LW-ISP: A Lightweight Model with ISP and Deep Learning. *arXiv preprint arXiv:2210.03904*.

Chen, K.; Wang, J.; Pang, J.; Cao, Y.; Xiong, Y.; Li, X.; Sun, S.; Feng, W.; Liu, Z.; Xu, J.; et al. 2019. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*.

Chen, M.; Wang, Z.; and Zheng, F. 2021. Benchmarks for Corruption Invariant Person Re-identification. *arXiv preprint arXiv:2111.00880*.

Chen, X.; Xie, C.; Tan, M.; Zhang, L.; Hsieh, C.-J.; and Gong, B. 2021. Robust and accurate object detection via adversarial learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16622–16631.

Dodge, S.; and Karam, L. 2016. Understanding how image quality affects deep neural networks. In *2016 eighth international conference on quality of multimedia experience (QoMEX)*, 1–6. IEEE.

Fan, Q.; Segu, M.; Tai, Y.-W.; Yu, F.; Tang, C.-K.; Schiele, B.; and Dai, D. 2022. Normalization Perturbation: A Simple Domain Generalization Method for Real-World Domain Shifts. *arXiv preprint arXiv:2211.04393*.

Guo, H.; Lu, T.; and Wu, Y. 2021. Dynamic Low-Light Image Enhancement for Object Detection via End-to-End Training. In *2020 25th International Conference on Pattern Recognition (ICPR)*, 5611–5618. IEEE.

Hendrycks, D.; and Dietterich, T. 2019. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*.

Hendrycks, D.; and Gimpel, K. 2016. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*.

Hong, Y.; Wei, K.; Chen, L.; and Fu, Y. 2021. Crafting object detection in very low light. In *BMVC*, volume 1, 3.

Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; Le, Q. V.; and Adam, H. 2019. h-swish: Hybrid Activation Function for Deep Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Huang, H.; Yang, W.; Hu, Y.; Liu, J.; and Duan, L.-Y. 2022. Towards low light enhancement with raw images. *IEEE Transactions on Image Processing*, 31: 1391–1405.

Jiang, K.; Wang, Z.; Wang, Z.; Chen, C.; Yi, P.; Lu, T.; and Lin, C.-W. 2022. Degrade is upgrade: Learning degradation for low-light image enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 1078–1086.

Kamann, C.; and Rother, C. 2020. Benchmarking the robustness of semantic segmentation models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8828–8838.

Krizhevsky, A.; and Hinton, G. 2010. Convolutional deep belief networks on cifar-10. *Unpublished manuscript*, 40(7): 1–9.

Kroon-Batenburg, L. M.; Helliwell, J. R.; McMahon, B.; and Terwilliger, T. C. 2017. Raw diffraction data preservation and reuse: overview, update on practicalities and metadata requirements. *IUCrJ*, 4(1): 87–99.

Li, B.; Peng, X.; Wang, Z.; Xu, J.; and Feng, D. 2017. An all-in-one network for dehazing and beyond. *arXiv preprint arXiv:1707.06543*.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022.

Ljungbergh, W.; Johnander, J.; Petersson, C.; and Felsberg, M. 2023. Raw or Cooked? Object Detection on RAW Images. In *Image Analysis: 23rd Scandinavian Conference, SCIA 2023, Sirkka, Finland, April 18–21, 2023, Proceedings, Part I*, 374–385. Springer.

Loh, Y. P.; and Chan, C. S. 2019. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178: 30–42.

Michaelis, C.; Mitzkus, B.; Geirhos, R.; Rusak, E.; Bringmann, O.; Ecker, A. S.; Bethge, M.; and Brendel, W. 2019. Benchmarking robustness in object detection: Autonomous driving when winter is coming. *arXiv preprint arXiv:1907.07484*.

Misra, D. 2019. Mish: A self regularized non-monotonic neural activation function. *arXiv preprint arXiv:1908.08681*, 4(2): 10–48550.

Morawski, I.; Chen, Y.-A.; Lin, Y.-S.; Dangi, S.; He, K.; and Hsu, W. H. 2022. GenISP: Neural ISP for Low-Light Machine Cognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 630–639.

Omid-Zohoor, A.; Ta, D.; and Murmann, B. 2014. PASCALRAW: raw image database for object detection.

Pham, L. H.; Jeon, H.-J.; Tran, D. N.-N.; Tran, T. H.-P.; Nguyen, H.-H.; Jeon, H.-M.; and Jeon, J. W. 2022. 5th UG2+ Challenge Track 1.1: Object Detection in the Hazy Condition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Ramachandran, P.; Zoph, B.; and Le, Q. V. 2017. Searching for activation functions. *arXiv preprint arXiv:1710.05941*.

Richter, T.; and Fößel, S. 2019. Bayer pattern compression with JPEG XS. In *2019 IEEE International Conference on Image Processing (ICIP)*, 3177–3181. IEEE.

Shaw, J. A. 2013. Radiometry and the Friis transmission equation. *American journal of physics*, 81(1): 33–37.

Shyam, P.; Sengar, S. S.; Yoon, K.-J.; and Kim, K.-S. 2022. Lightweight hdr camera isp for robust perception in dynamic illumination conditions via fourier adversarial networks. *arXiv preprint arXiv:2204.01795*.

Sun, S.; Ren, W.; Wang, T.; and Cao, X. 2022. Rethinking Image Restoration for Object Detection. *Advances in Neural Information Processing Systems*, 35: 4461–4474.

Talpes, E.; Sarma, D. D.; Venkataramanan, G.; Bannon, P.; McGee, B.; Floering, B.; Jalote, A.; Hsiong, C.; Arora, S.; Gorti, A.; et al. 2020. Compute solution for tesla's full self-driving computer. *IEEE Micro*, 40(2): 25–35.

Tan, M.; and Le, Q. V. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*.

Tan, M.; Pang, R.; and Le, Q. V. 2020. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10781–10790.

Tian, Z.; Shen, C.; Chen, H.; and He, T. 2019. Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, 9627–9636.

VidalMata, R. G.; Banerjee, S.; RichardWebster, B.; Albright, M.; Davalos, P.; McCloskey, S.; Miller, B.; Tambo, A.; Ghosh, S.; Nagesh, S.; et al. 2020. Bridging the gap between computational photography and visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 43(12): 4272–4290.

Wang, Y.; Yu, Y.; Yang, W.; Guo, L.; Chau, L.-P.; Kot, A.; and Wen, B. 2023. Raw image reconstruction with learned compact metadata. *arXiv preprint arXiv:2302.12995*.

Wu, C.-T.; Isikdogan, L. F.; Rao, S.; Nayak, B.; Gerasimow, T.; Sutic, A.; Ain-kedem, L.; and Michael, G. 2019. VisionISP: Repurposing the image signal processor for computer vision applications. In *2019 IEEE International Conference on Image Processing (ICIP)*, 4624–4628. IEEE.

Wu, W.; Chang, H.; Zheng, Y.; Li, Z.; Chen, Z.; and Zhang, Z. 2022. Contrastive Learning-Based Robust Object Detection Under Smoky Conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4295–4302.

Xing, Y.; Qian, Z.; and Chen, Q. 2021. Invertible image signal processing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6287–6296.

Xu, R.; Chen, C.; Peng, J.; Li, C.; Huang, Y.; Song, F.; Yan, Y.; and Xiong, Z. 2023. Toward RAW Object Detection: A New Benchmark and a New Model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13384–13393.

Yan, J.; Li, J.; and Fu, X. 2019. No-reference quality assessment of contrast-distorted images using contrast enhancement. *arXiv preprint arXiv:1904.08879*.

Yao, T.; Li, Y.; Pan, Y.; Wang, Y.; Zhang, X.-P.; and Mei, T. 2023. Dual vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Yoshimura, M.; Otsuka, J.; Irie, A.; and Ohashi, T. 2022. Rawgment: Noise-Accounted RAW Augmentation Enables Recognition in a Wide Variety of Environments. *arXiv preprint arXiv:2210.16046*.

Zhang, B.; Tian, Z.; Tang, Q.; Chu, X.; Wei, X.; Shen, C.; et al. 2022a. Segvit: Semantic segmentation with plain vision transformers. *Advances in Neural Information Processing Systems*, 35: 4971–4982.

Zhang, H.; Li, F.; Liu, S.; Zhang, L.; Su, H.; Zhu, J.; Ni, L. M.; and Shum, H.-Y. 2022b. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*.

Zhang, Y.; Qin, H.; Wang, X.; and Li, H. 2021. Rethinking noise synthesis and modeling in raw denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4593–4601.

Zhou, D.; Yu, Z.; Xie, E.; Xiao, C.; Anandkumar, A.; Feng, J.; and Alvarez, J. M. 2022. Understanding the robustness in vision transformers. In *International Conference on Machine Learning*, 27378–27394. PMLR.

Zhou, W.; Zhang, X.; Wang, H.; Gao, S.; and Lou, X. 2021. Raw Bayer Pattern Image Synthesis for Computer Vision-oriented Image Signal Processing Pipeline Design. *arXiv preprint arXiv:2110.12823*.

Zhou, X.; Wang, D.; and Krähenbühl, P. 2019. Objects as points. *arXiv preprint arXiv:1904.07850*.

Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; and Ye, J. 2023. Object detection in 20 years: A survey. *Proceedings of the IEEE*.