



# WiSE-OD: Benchmarking Robustness in Infrared Object Detection

Heitor R. Medeiros      Atif Belal      Masih Aminbeidokhti

Eric Granger      Marco Pedersoli

Laboratoire d'imagerie, de vision et d'intelligence artificielle (LIVIA)

International Laboratory on Learning Systems (ILLS)

Dept. of Systems Engineering, ETS Montreal, Canada

ÉCOLE DE  
TECHNOLOGIE  
SUPÉRIEURE

Université du Québec

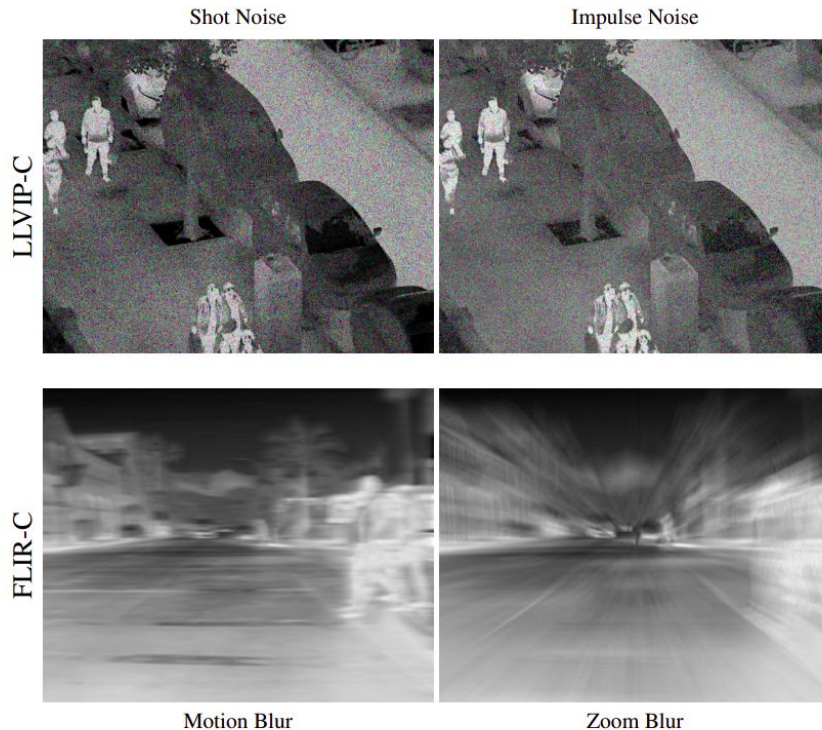


**LIVIA**  
LABORATOIRE  
D'IMAGERIE, DE VISION  
ET D'INTELLIGENCE  
ARTIFICIELLE

**DISTECH  
CONTROLS™**

# Introduction

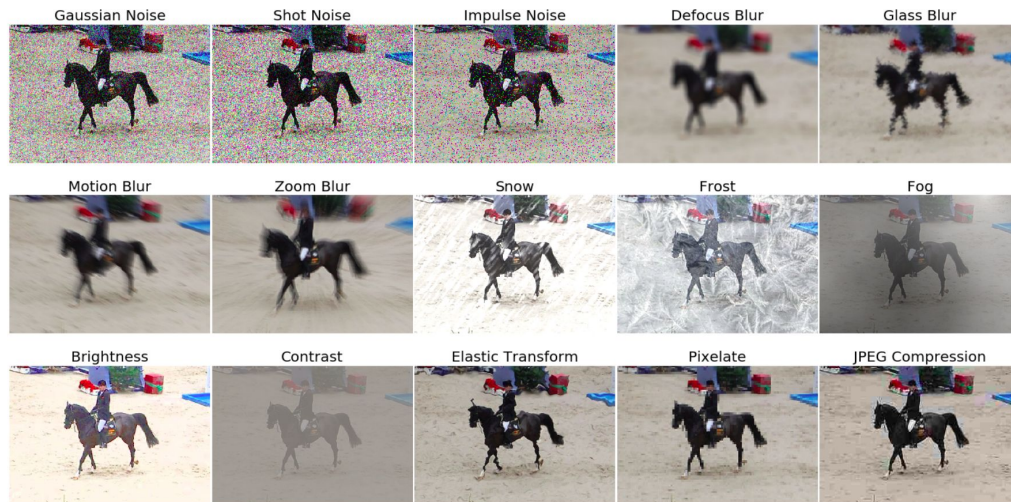
- Infrared object detection is key for low-light applications, but fine-tuning from RGB to IR often hurts robustness under corruptions and domain shifts.
- We introduce LLVIP-C and FLIR-C, the first corruption benchmarks for cross-modality IR object detection.
- Additionally, we propose WiSE-OD, a training-free weight-space ensemble that improves IR robustness with no additional inference-time cost.



**Figure 1. LLVIP-C and FLIR-C examples.**

# Robustness in OD

- Hendrycks & Dietterich [1] proposed diverse corruptions for robustness in image classification, resulting in ImageNet-C and CIFAR10-C.
- Michaelis et al [2]. extended the classification work to OD, proposing Pascal-C, COCO-C, and Cityscapes-C, with a study of corruption severity and detector performance.



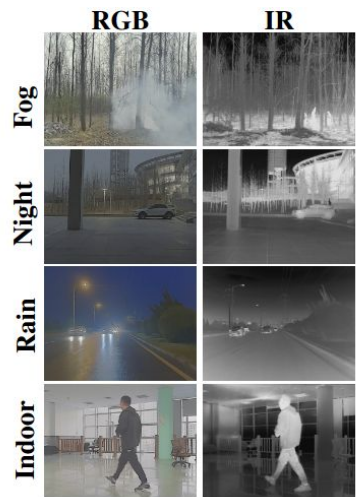
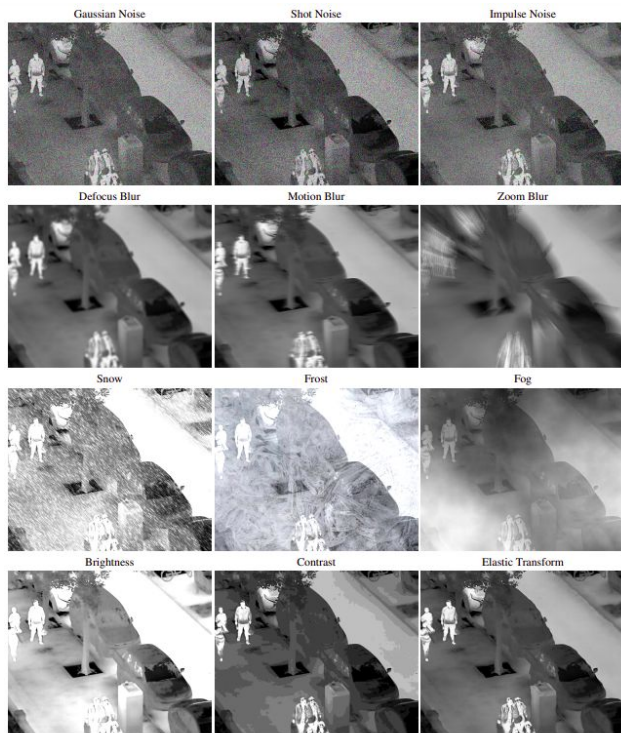
**Figure 2. Corruption types** from Hendrycks and Dietterich, adapted to corrupt arbitrary images. **Source:** Michaelis et al [2].

[1] Hendrycks, Dan, and Thomas Dietterich. "Benchmarking Neural Network Robustness to Common Corruptions and Perturbations." International Conference on Learning Representations.

[2] Michaelis, Claudio, et al. "Benchmarking robustness in object detection: Autonomous driving when winter is coming." arXiv preprint arXiv:1907.07484 (2019).

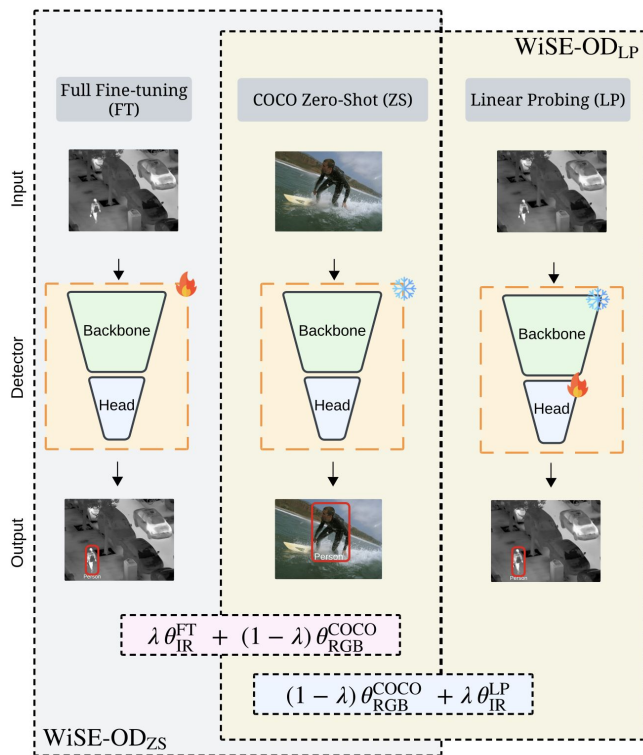
# Proposed Benchmark for Robustness in IR OD

- We construct corrupted IR benchmarks LLVIP-C and FLIR-C.
- We evaluate IR ODs over 15 synthetic corruptions, and natural shifts (M3FD) and report the average performance.
- We benchmark ZS, LP, LP-FT and WiSE-OD as baselines.



**Figure 3.** In the left examples of corruption on LLVIP-C. On the right, the natural shifts on IR for the M3FD dataset.

# Propose Method: WiSE-OD



**Figure 4.** Our proposed method: **WiSE-OD**. Our two variations can combine FT or LP weights to improve OOD performance.

- WiSE-OD starts from a pre-trained RGB model.
- Then, we fine-tune the pre-trained RGB model using IR data generating the IR weights.
- Next, we merge the pre-trained RGB weights and the IR weights by linear interpolation:

$$f_{\text{wod}}(\theta_{\text{RGB}}^{\text{COCO}}, \theta_{\text{IR}}^{\text{FT}}; \lambda) = (1 - \lambda) \theta_{\text{RGB}}^{\text{COCO}} + \lambda \theta_{\text{IR}}^{\text{FT}}.$$

- Finally, we run the final inference using the merged weights.

# Detection performance on Synthetic corruptions

Detector	LLVIP-C					
	Zero-Shot	FT	LP	LP-FT	WiSE-OD <sub>ZS</sub>	WiSE-OD <sub>LP</sub>
Faster R-CNN	40.96	56.40	70.96	70.02	75.08	<b>75.83</b>
FCOS	36.11	61.17	63.91	60.26	<b>76.50</b>	75.95
RetinaNet	37.50	61.13	60.37	61.11	73.69	<b>74.39</b>
Detector	FLIR-C					
	Zero-Shot	FT	LP	LP-FT	WiSE-OD <sub>ZS</sub>	WiSE-OD <sub>LP</sub>
Faster R-CNN	33.67	44.80	47.11	46.92	50.52	<b>51.59</b>
FCOS	28.85	41.07	38.92	38.14	<b>47.13</b>	46.76
RetinaNet	28.27	42.71	36.71	36.45	45.35	<b>47.53</b>

**Table 1.** Detection performance for the OD IR Robustness Benchmark for LLVIP-C with severity 5 and FLIR-C with severity level 2.

# Detection performance on Real Shifts

Method	Rain			Night			Fog			Indoor			Avg.		
	AP <sub>50</sub>	AP <sub>75</sub>	mAP	AP <sub>50</sub>	AP <sub>75</sub>	mAP	AP <sub>50</sub>	AP <sub>75</sub>	mAP	AP <sub>50</sub>	AP <sub>75</sub>	mAP	AP <sub>50</sub>	AP <sub>75</sub>	mAP
<b>Faster R-CNN</b>															
Zero-shot	13.1	6.1	6.7	14.3	8.3	8.2	87.4	81.7	68.4	5.8	5.7	4.4	30.2	25.4	22.0
Fine-tuning	28.0	11.8	13.8	<b>47.8</b>	<b>30.9</b>	<b>29.4</b>	93.0	77.5	66.2	<b>59.8</b>	32.8	33.2	57.2	38.3	35.7
WiSE-OD	<b>29.4</b>	<b>13.2</b>	<b>14.9</b>	43.3	30.3	27.9	<b>97.2</b>	<b>84.5</b>	<b>74.2</b>	59.2	<b>36.2</b>	<b>33.9</b>	<b>57.3</b>	<b>41.1</b>	<b>37.7</b>
<b>FCOS</b>															
Zero-shot	10.6	5.5	5.6	12.3	7.2	7.1	83.4	74.8	64.0	4.3	4.0	3.3	27.7	22.9	20.0
Fine-tuning	21.7	9.5	10.7	<b>42.3</b>	<b>28.3</b>	<b>27.9</b>	93.5	78.7	65.9	<b>61.6</b>	32.1	31.7	<b>54.8</b>	37.2	34.1
WiSE-OD	<b>25.5</b>	<b>11.7</b>	<b>13.0</b>	34.7	22.0	21.6	<b>96.0</b>	<b>86.8</b>	<b>73.0</b>	56.4	<b>35.2</b>	<b>33.7</b>	<b>53.2</b>	<b>38.9</b>	<b>35.3</b>
<b>RetinaNet</b>															
Zero-shot	10.6	4.9	5.4	12.9	7.7	7.4	91.4	84.6	69.6	9.0	6.9	6.3	31.0	26.0	22.2
Fine-tuning	<b>29.3</b>	11.5	13.5	<b>45.0</b>	<b>28.8</b>	<b>27.3</b>	96.5	82.8	67.7	55.3	<b>34.4</b>	<b>32.9</b>	<b>56.5</b>	39.4	35.3
WiSE-OD	27.2	<b>12.7</b>	<b>13.7</b>	40.1	27.1	25.8	<b>97.6</b>	<b>89.9</b>	<b>73.6</b>	<b>56.7</b>	33.6	32.8	<b>55.4</b>	<b>40.8</b>	<b>36.5</b>
<b>YOLOv8n</b>															
Zero-shot	28.4	22.4	20.5	37.3	34.6	31.8	88.9	84.9	73.8	22.1	22.1	19.3	44.2	41.0	36.4
Fine-tuning	29.2	23.1	21.7	49.1	37.7	36.6	92.3	83.6	70.3	<b>55.1</b>	<b>43.4</b>	<b>38.5</b>	56.4	46.9	41.8
WiSE-OD	<b>42.2</b>	<b>35.5</b>	<b>32.5</b>	<b>55.5</b>	<b>46.1</b>	<b>43.7</b>	<b>95.3</b>	<b>90.9</b>	<b>77.5</b>	48.8	43.6	38.7	<b>60.4</b>	<b>54.0</b>	<b>48.1</b>

**Table 2.** Detection performance on **M3FD** dataset. Results are reported for **Rain, Night, Fog, Indoor**, and **average** across conditions.

# Qualitative Results over Synthetic Corruptions

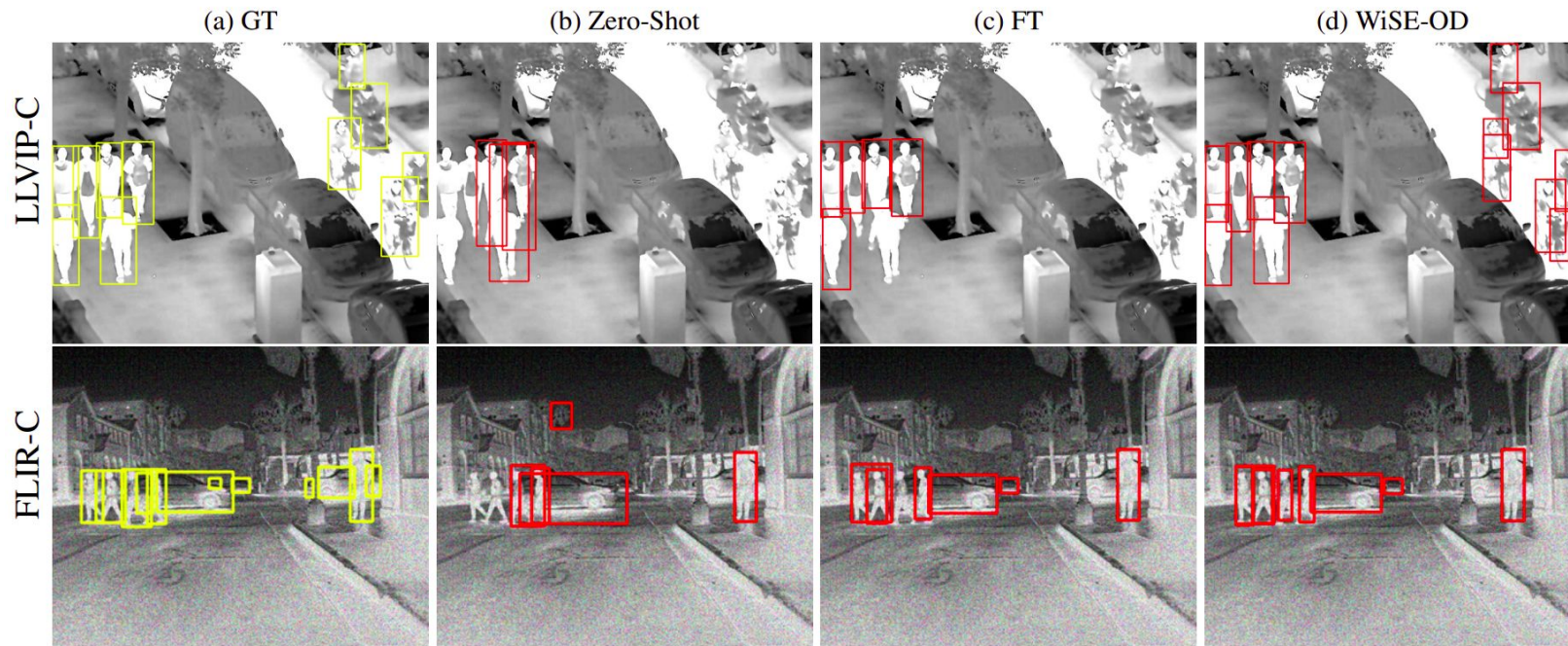
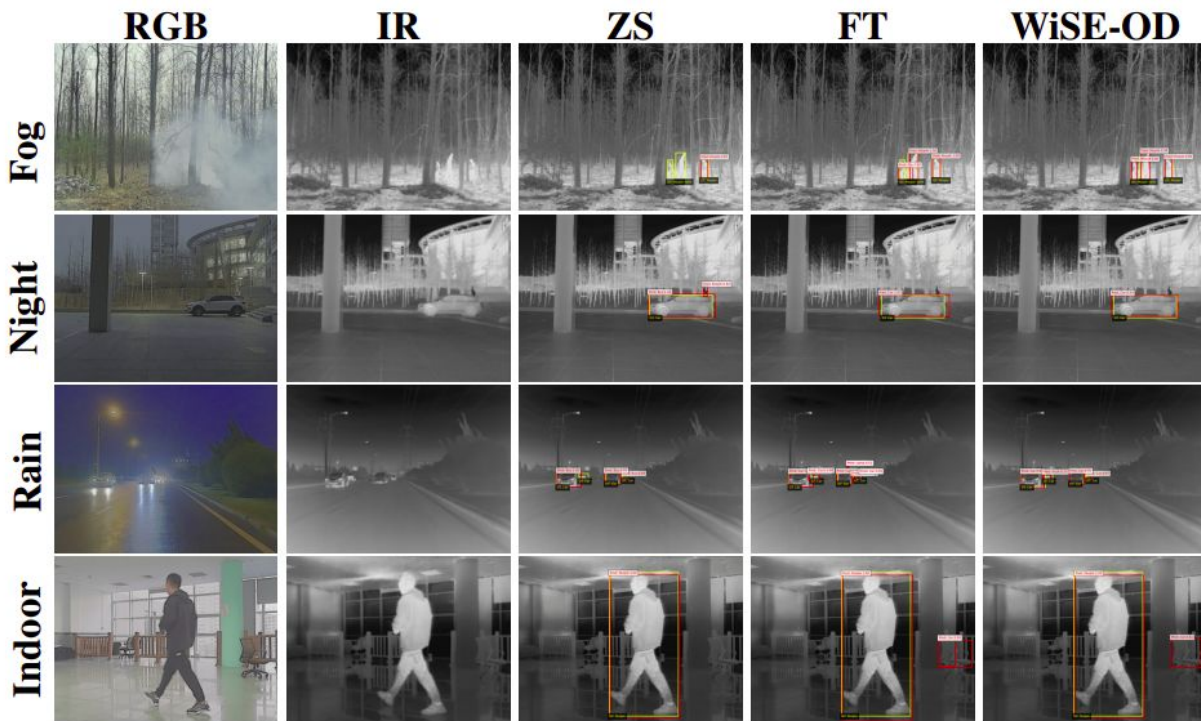


Figure 5. Robustness of Infrared Object Detection on LLVIP-C and FLIR-C datasets for **brightness** (first row), and **shot noise** with **Faster R-CNN**.

# Qualitative Results over Real Shifts



**Figure 6.** Qualitative comparison on **M3FD** under **adverse conditions**. **Rows:** fog, night, rain, indoor. **Columns:** RGB, IR, Zero-Shot (ZS), Fine-Tuning (FT), and WiSE-OD. in yellow, we have GTs and in red the detections.

# Conclusion

- WiSE-OD consistently improves robustness across synthetic corruptions and real-world IR shifts, while preserving in-domain accuracy.
- The method is simple and practical, requiring no additional training compared to fine-tuning, parameters, or inference overhead.
- Our benchmarks and results highlight the need for robustness-aware IR evaluation, and provide a strong baseline for future work.

# WiSE-OD: Benchmarking Robustness in Infrared Object Detection



Paper



Code



# WiSE-OD: Benchmarking Robustness in Infrared Object Detection

Heitor R. Medeiros      Atif Belal      Masih Aminbeidokhti

Eric Granger      Marco Pedersoli

Laboratoire d'imagerie, de vision et d'intelligence artificielle (LIVIA)

International Laboratory on Learning Systems (ILLS)

Dept. of Systems Engineering, ETS Montreal, Canada

ÉCOLE DE  
TECHNOLOGIE  
SUPÉRIEURE

Université du Québec



**LIVIA**  
LABORATOIRE  
D'IMAGERIE, DE VISION  
ET D'INTELLIGENCE  
ARTIFICIELLE

**DISTECH  
CONTROLS™**