Supplementary Material: 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

In this supplementary material, we provide additional information to reproduce our work. The source code is provided alongside the supplementary material, and we are going to provide the official repository. This supplementary material is divided into the following sections: Dataset Visualization (Section 1), Detection performance per corruption (Section 2), Activation map analysis (Section 3), WiSE-OD$_{ZS}$: Ablation study on λ (Section 4), Performance over different corruption levels (Section 5), Results on real-shit M3FD dataset (Section 6), Qualitative results different levels of severity (Section 7), and Additional Details Benchmark (Section 8).

1. Dataset Visualization

In this section, we provide additional visualization of each corruption for both datasets: LLVIP-C in Figure 1 and FLIR-C in Figure 2. Here, we wanted to highlight how strong the severity level of 5 is for FLIR-C, which can destroy the whole image, for instance, for the Frost corruption.

2. Detection performance per corruption

In this section, we expanded our evaluation of the benchmark per corruption. In the main manuscript, we provided the results per corruption for Faster R-CNN with a severity level of 5 for LLVIP-C. Here, we provide the additional results for the FCOS (Table 1) and RetinaNet (Table 2) for LLVIP-C, FCOS (Table 3) and RetinaNet (Table 4) for FLIR-C with severity level of 2. We further provide some plots for Faster R-CNN with FLIR-C with all severity levels from 1 to 5, as well as additional per-class evaluation.

3. Activation map analysis

In this section, we provide more plots with the grad-cam activations. Here, we divided into three figures due to space constraints: Figure 5, Figure 6, and Figure 7. In many cases, the WiSE-OD$_{ZS}$ had a person detected on average, which is shown by the red highlighted part of the images.

Table 1. **AP$_{50}$ performance over the perturbations for LLVIP-C with severity level 5 for FCOS.**

	LLVIP-C					
	Zero-Shot	FT	LP	LP-FT	WiSE-OD$_{ZS}$	WiSE-OD$_{LP}$
Gaussian Noise	55.03 ± 0.07	62.32 ± 5.96	73.63 ± 0.09	64.46 ± 0.42	83.68 ± 0.67	81.78 ± 0.08
Shot Noise	43.83 ± 0.36	58.75 ± 5.48	66.05 ± 1.98	59.34 ± 0.08	80.55 ± 0.81	80.00 ± 0.59
Impulse Noise	51.60 ± 0.17	65.26 ± 5.31	68.98 ± 0.04	65.06 ± 0.21	85.86 ± 0.53	85.21 ± 0.95
Defocus Blur	33.67 ± 0.00	80.30 ± 1.90	78.14 ± 0.00	76.47 ± 0.00	87.74 ± 0.68	88.77 ± 0.00
Motion Blur	15.09 ± 0.34	78.87 ± 1.09	70.67 ± 0.21	67.88 ± 0.25	84.32 ± 0.76	84.06 ± 0.01
Zoom Blur	01.28 ± 0.00	20.13 ± 2.95	18.75 ± 0.87	14.01 ± 0.00	31.47 ± 2.96	30.27 ± 0.00
Snow	37.37 ± 0.25	33.17 ± 2.74	45.72 ± 0.31	42.42 ± 0.25	69.72 ± 1.83	70.94 ± 0.38
Frost	35.33 ± 0.07	60.18 ± 1.17	53.68 ± 1.07	48.67 ± 0.37	76.33 ± 1.12	74.94 ± 1.47
Fog	55.52 ± 0.13	69.47 ± 4.86	85.08 ± 0.02	81.24 ± 0.19	89.49 ± 0.56	90.45 ± 0.05
Brightness	36.46 ± 0.00	55.69 ± 2.98	65.16 ± 0.00	65.88 ± 0.00	82.52 ± 0.90	82.84 ± 0.00
Contrast	42.04 ± 0.00	00.64 ± 0.74	56.61 ± 0.21	49.36 ± 0.00	23.45 ± 4.15	20.26 ± 0.00
Elastic transform	43.02 ± 0.09	92.96 ± 0.57	83.77 ± 0.15	83.65 ± 0.22	93.97 ± 0.38	93.67 ± 0.03
Pixelate	02.14 ± 0.00	88.12 ± 0.65	55.20 ± 0.00	54.17 ± 0.00	89.43 ± 0.90	88.50 ± 0.00
JPEG compression	53.22 ± 0.00	90.53 ± 1.18	73.40 ± 0.00	71.14 ± 0.00	92.59 ± 0.40	91.71 ± 0.00
mPC	36.11	61.17	63.91	60.26	**76.50**	75.95

Table 2. **AP$_{50}$ performance over the perturbations for LLVIP-C with severity level 5 for RetinaNet.**

	LLVIP-C					
	Zero-Shot	FT	LP	LP-FT	WiSE-OD$_{ZS}$	WiSE-OD$_{LP}$
Gaussian Noise	52.52 ± 0.21	70.55 ± 8.23	66.70 ± 0.19	66.32 ± 0.25	82.63 ± 3.59	88.10 ± 0.07
Shot Noise	44.70 ± 0.22	64.02 ± 10.87	57.85 ± 0.13	59.03 ± 0.17	78.94 ± 4.68	78.14 ± 1.53
Impulse Noise	48.17 ± 0.19	73.01 ± 6.68	66.97 ± 1.82	67.78 ± 0.08	84.19 ± 3.05	86.53 ± 1.95
Defocus Blur	43.13 ± 0.00	75.38 ± 8.71	71.92 ± 0.00	73.08 ± 0.20	84.31 ± 3.89	88.45 ± 0.00
Motion Blur	18.79 ± 0.07	71.69 ± 9.38	62.14 ± 0.15	63.62 ± 0.17	79.92 ± 4.78	78.59 ± 3.16
Zoom Blur	01.61 ± 0.00	12.73 ± 3.63	09.21 ± 0.00	09.93 ± 0.55	21.02 ± 5.04	16.92 ± 0.00
Snow	35.54 ± 0.34	36.48 ± 11.52	56.68 ± 0.14	55.19 ± 0.18	71.09 ± 5.52	72.05 ± 4.35
Frost	36.52 ± 0.06	57.94 ± 6.50	53.83 ± 0.26	53.47 ± 0.32	75.35 ± 3.81	78.60 ± 1.74
Fog	58.48 ± 0.21	65.31 ± 6.58	83.18 ± 0.27	82.28 ± 0.06	85.07 ± 2.92	84.11 ± 0.12
Brightness	39.03 ± 0.00	54.83 ± 8.34	69.76 ± 1.69	70.79 ± 0.00	82.41 ± 1.89	80.70 ± 0.00
Contrast	49.09 ± 0.00	00.99 ± 0.00	52.12 ± 0.00	54.61 ± 2.76	13.90 ± 1.79	13.64 ± 1.58
Elastic transform	37.37 ± 0.04	93.05 ± 1.27	76.89 ± 0.07	77.56 ± 1.20	94.52 ± 0.07	94.18 ± 0.09
Pixelate	03.86 ± 0.00	88.82 ± 1.97	48.35 ± 0.00	49.68 ± 0.00	85.68 ± 2.38	89.83 ± 0.00
JPEG compression	56.24 ± 0.00	91.09 ± 1.73	69.68 ± 0.00	72.32 ± 2.93	92.67 ± 1.14	91.71 ± 0.00
mPC	37.50	61.13	60.37	61.11	73.69	**74.39**

4. WiSE-OD$_{ZS}$: Ablation study on λ

In this section, we expanded our ablation study. In the Table 5, we have the full study for Faster R-CNN with LLVIP-C and in Table 6 for FLIR-C, as well as the FCOS in Table 7 and RetinaNet in Table 8. As mentioned in the main manuscript, the $\lambda = 0.5$ has a good performance under corruption without having to tune the λ over a validation set,

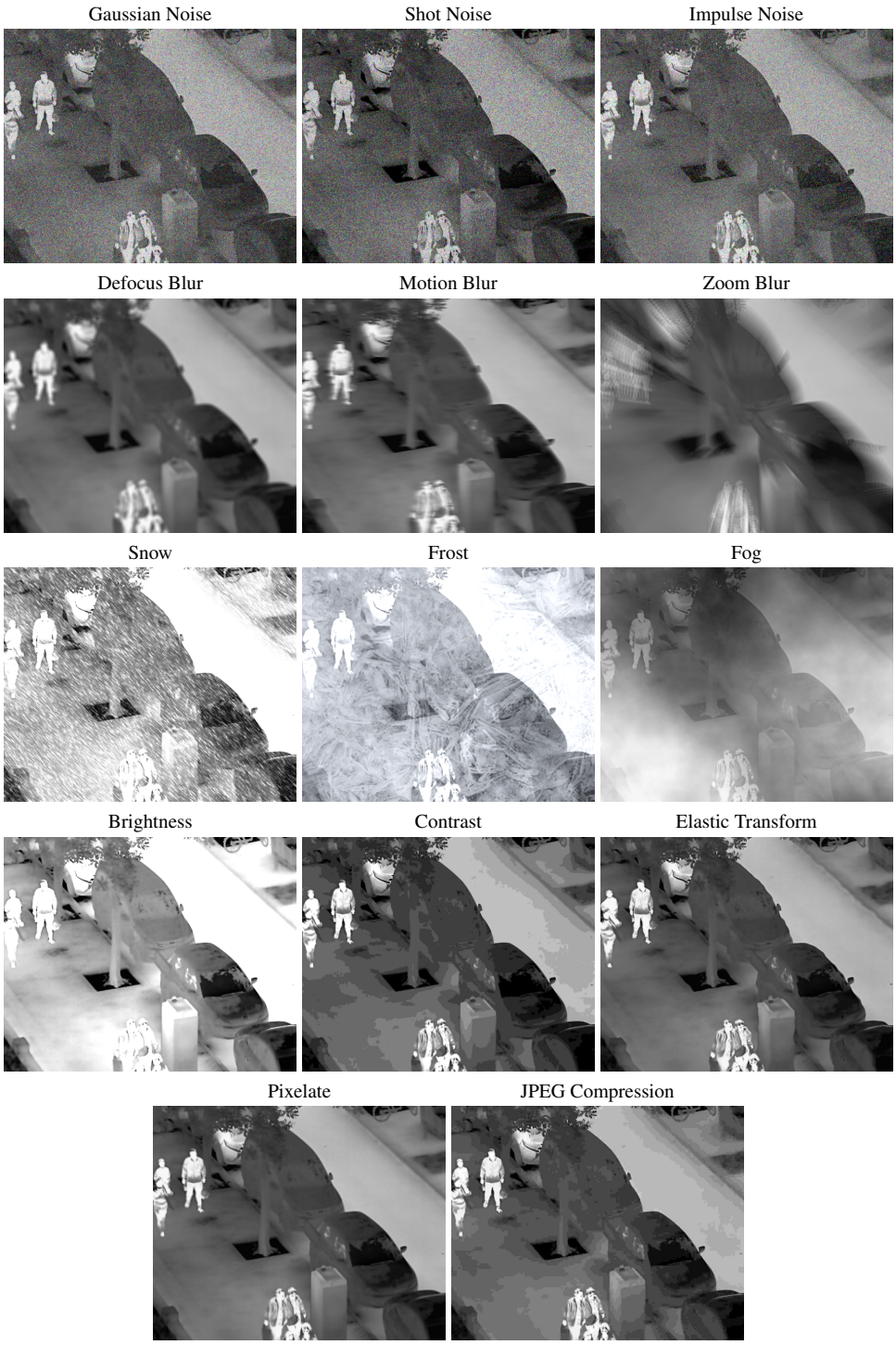

Figure 1. **All the 14 corruptions types from [1], adapted to our LLVIP-C benchmark with severity of 5.**

so it is a good choice for most of the cases; for being better in a specific corruption, we could tune the λ.

5. Performance over different corruption levels

In this section, we measured the per AP_{50} performance for Faster R-CNN, FCOS, and RetinaNet over different corruption severity levels for the benchmark. Here, we focus on

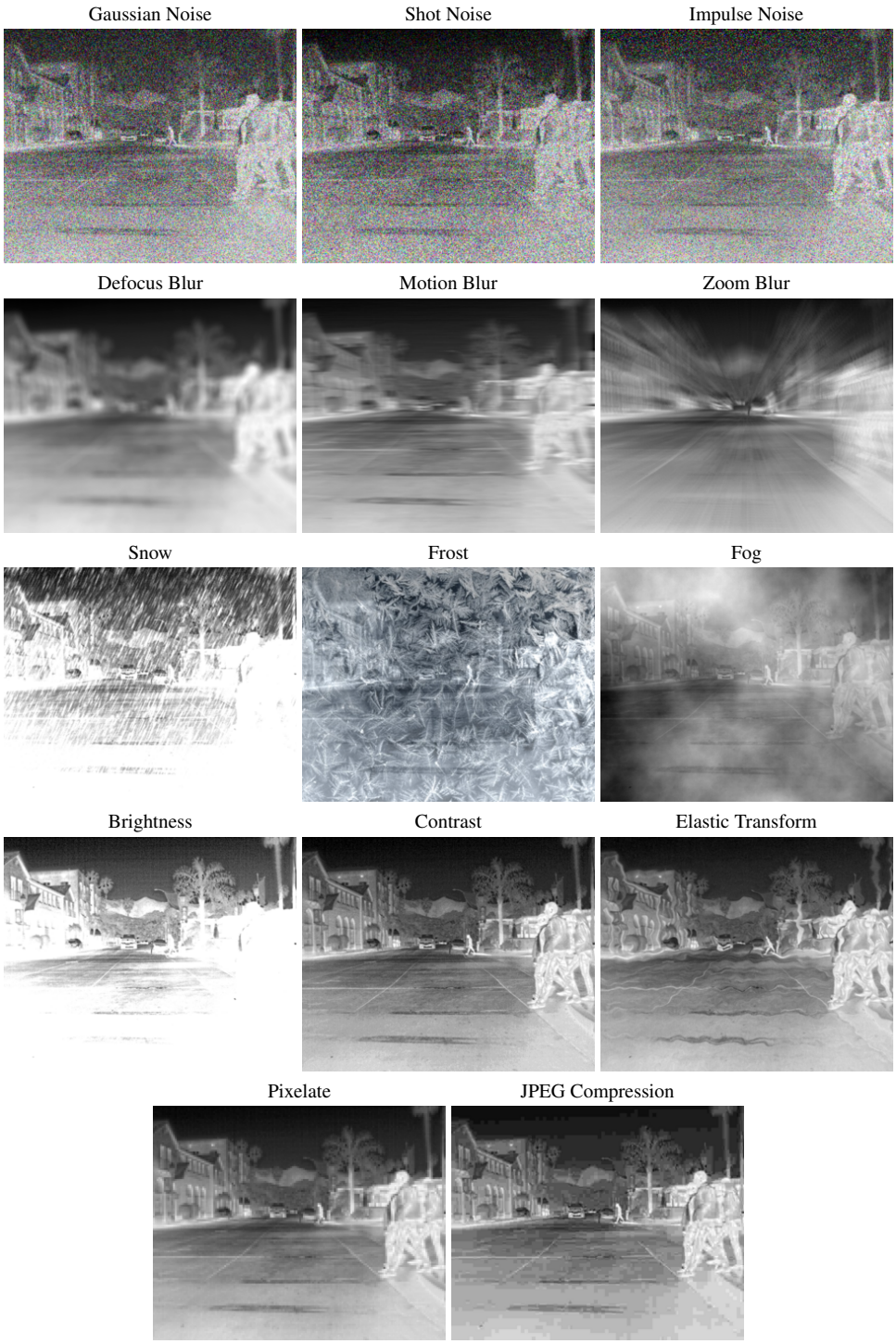

Figure 2. **All the 14 corruptions types from [1], adapted to our FLIR-C benchmark with severity of 5.**

Zero-Shot, FT and WiSE-OD$_{ZS}$ for most of the corruptions on LLVIP-C illustrated in Figure 8 and Figure 9, and for FLIR-C in Figure 10 and Figure 11.

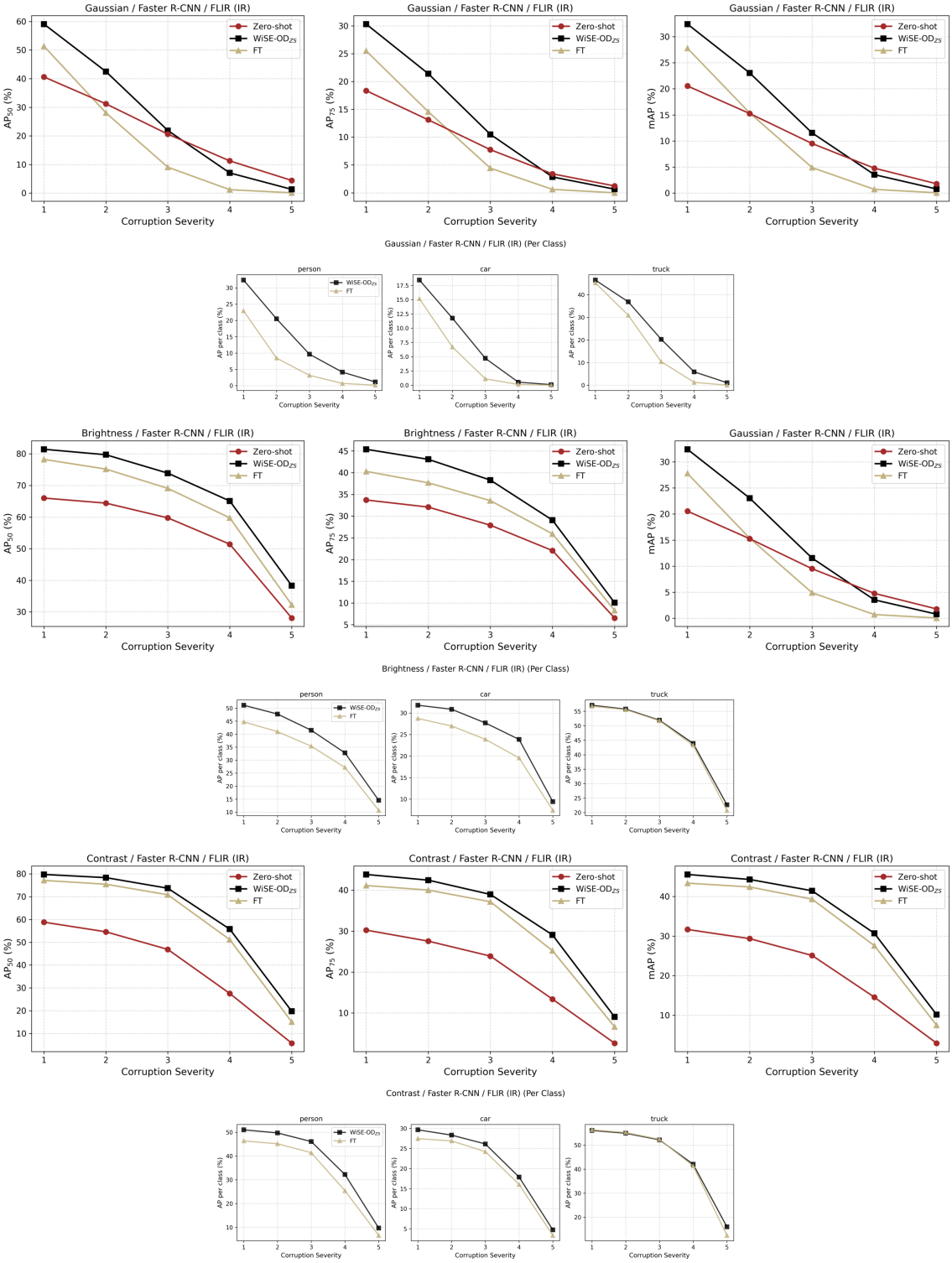

Figure 3. **Robustness analysis under Gaussian noise, brightness, and contrast corruptions on FLIR/Faster R-CNN. Top rows:** AP_{50}, AP_{75}, and mAP vs. corruption severity (1–5) for Zero-Shot (ZS), Fine-Tuning (FT), and WiSE-OD. **Bottom rows:** Per-class AP for *person*, *car*, and *truck* under each corruption type. Results show that WiSE-OD consistently reduces degradation compared to ZS and FT baselines.

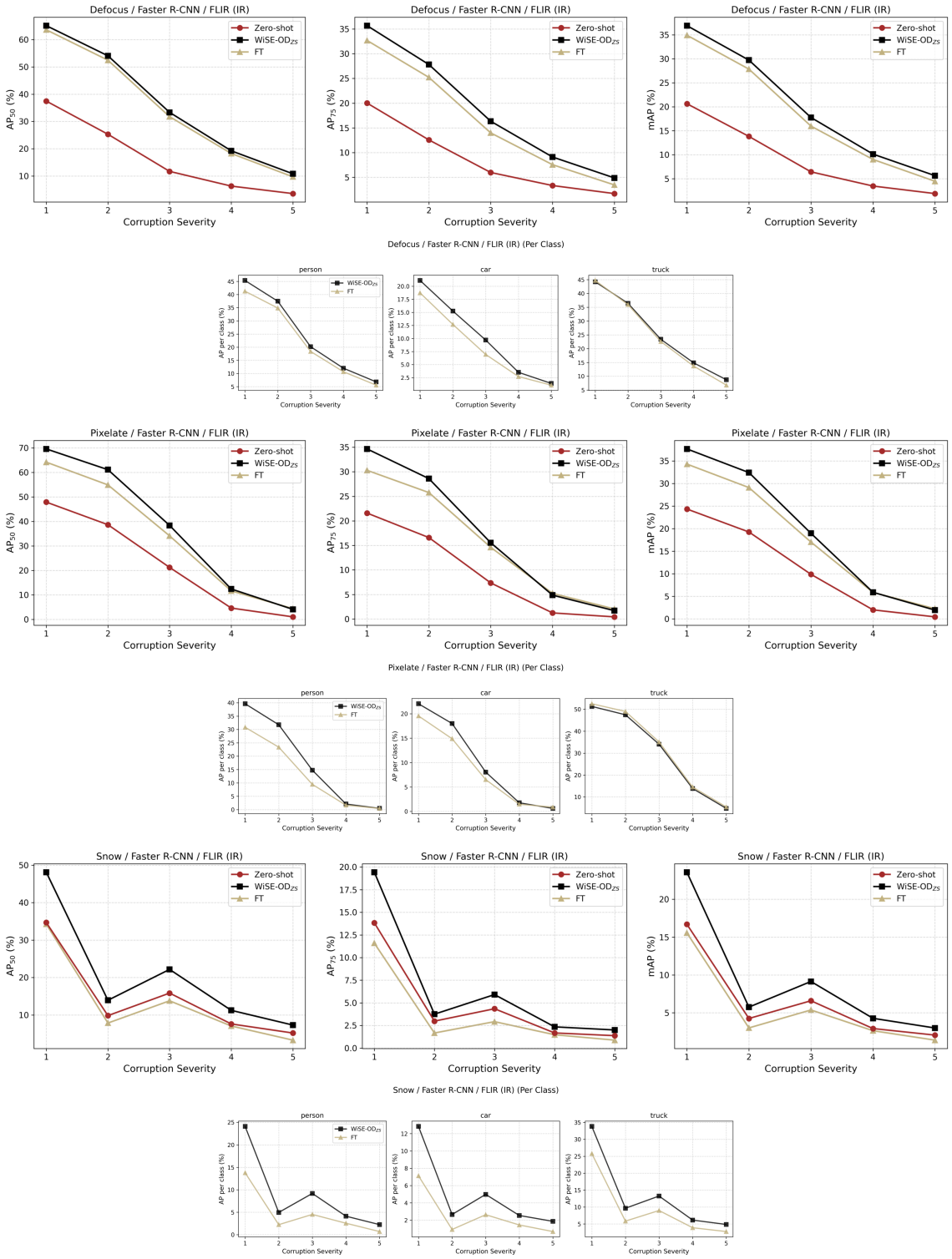

Figure 4. **Robustness analysis under defocus blur, frost, and snow corruptions on FLIR/Faster R-CNN. Top rows:** AP_{50}, AP_{75}, and mAP vs. corruption severity (1–5) for Zero-Shot (ZS), Fine-Tuning (FT), and WiSE-OD. **Bottom rows:** Per-class AP for *person*, *car*, and *truck* under each corruption type. Results show that WiSE-OD consistently reduces degradation compared to ZS and FT baselines.

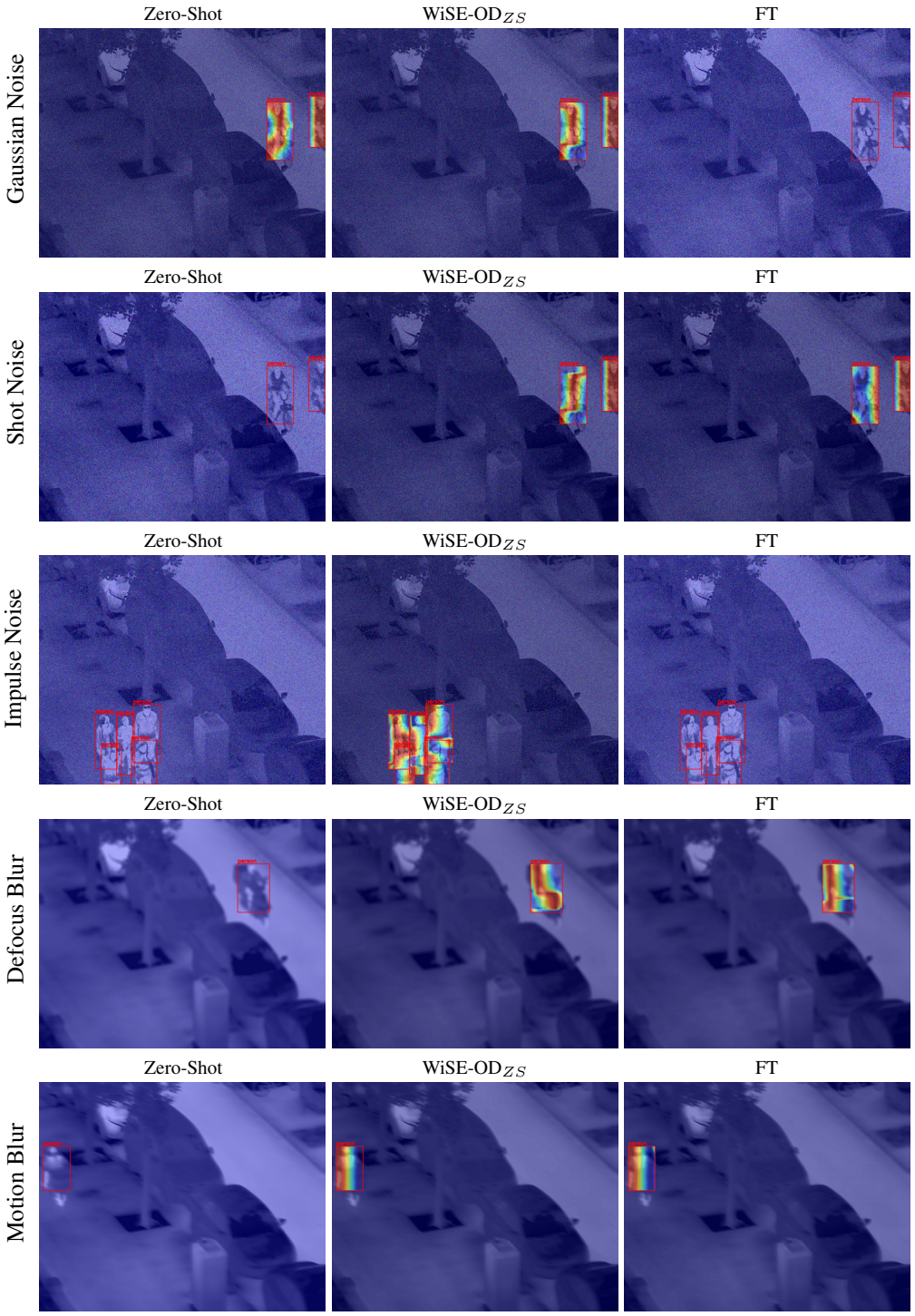

Figure 5. **Activation map analysis for zero-shot COCO pre-train Faster R-CNN detector, the WiSE-OD$_{ZS}$ and FT detector on IR for LLVIP-C dataset.** In red are the GTs, and for the WiSE-OD$_{ZS}$, the models are able to activate the features that represent a person for such corruptions (Part 1 with 5 of the 14 corruptions).

6. Qualitative comparison on M3FD under adverse conditions

Figure 12 illustrates qualitative results on the M3FD dataset under challenging conditions, including fog, night, rain, and indoor environments. Each row corresponds to a specific condition, while the columns compare RGB inputs, in-

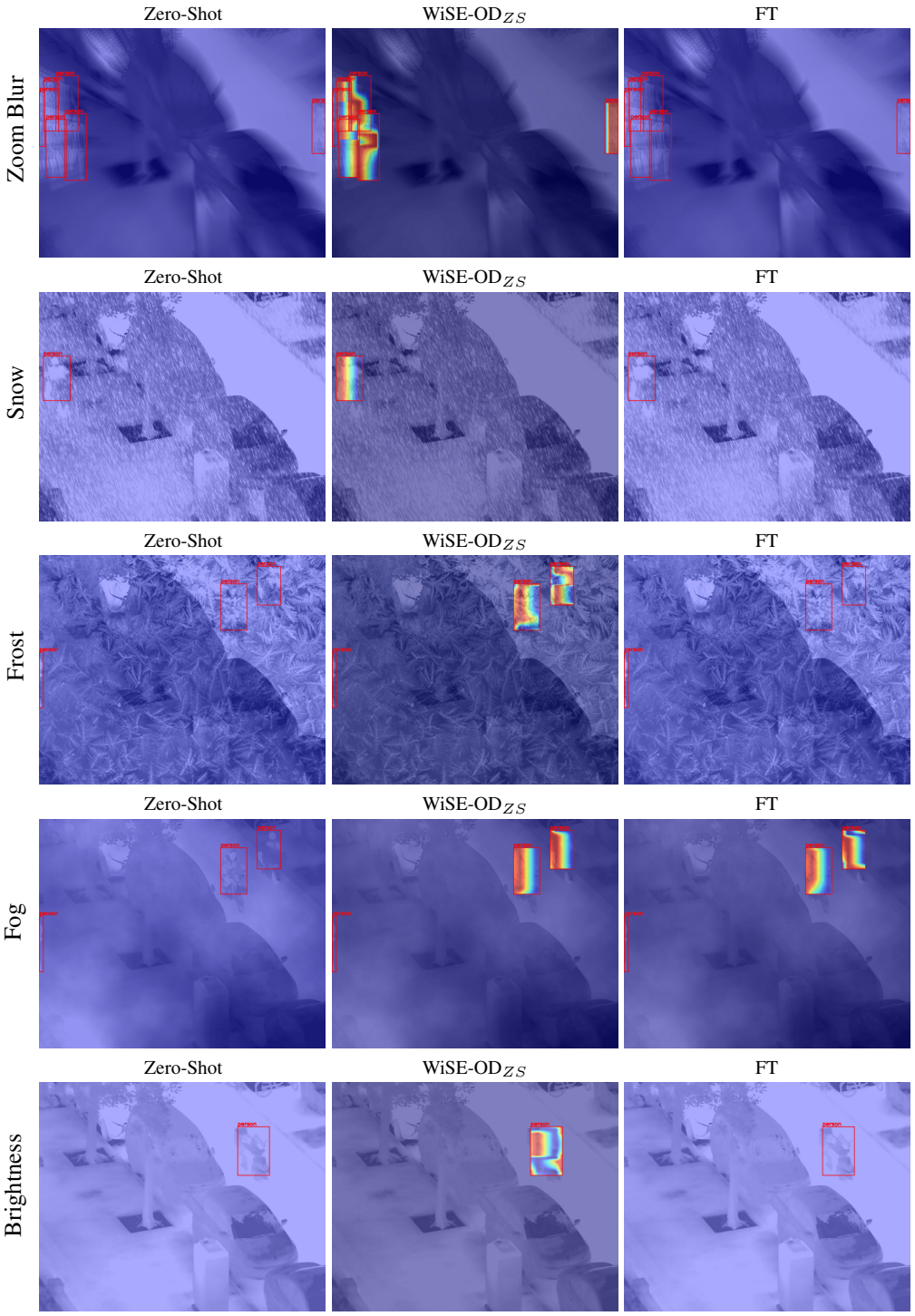

Figure 6. **Activation map analysis for zero-shot COCO pre-train Faster R-CNN detector, the WiSE-OD$_{ZS}$ and FT detector on IR for LLVIP-C dataset.** In red are the GTs, and for the WiSE-OD$_{ZS}$, the models are able to activate the features that represent a person for such corruptions (Part 2 with 5 of the 14 corruptions).

frared (IR) counterparts, and detection outputs from Zero-Shot (ZS), Fine-Tuning (FT), and WiSE-OD. As shown, WiSE-OD provides more reliable detections across all adverse settings, capturing objects that are often missed or

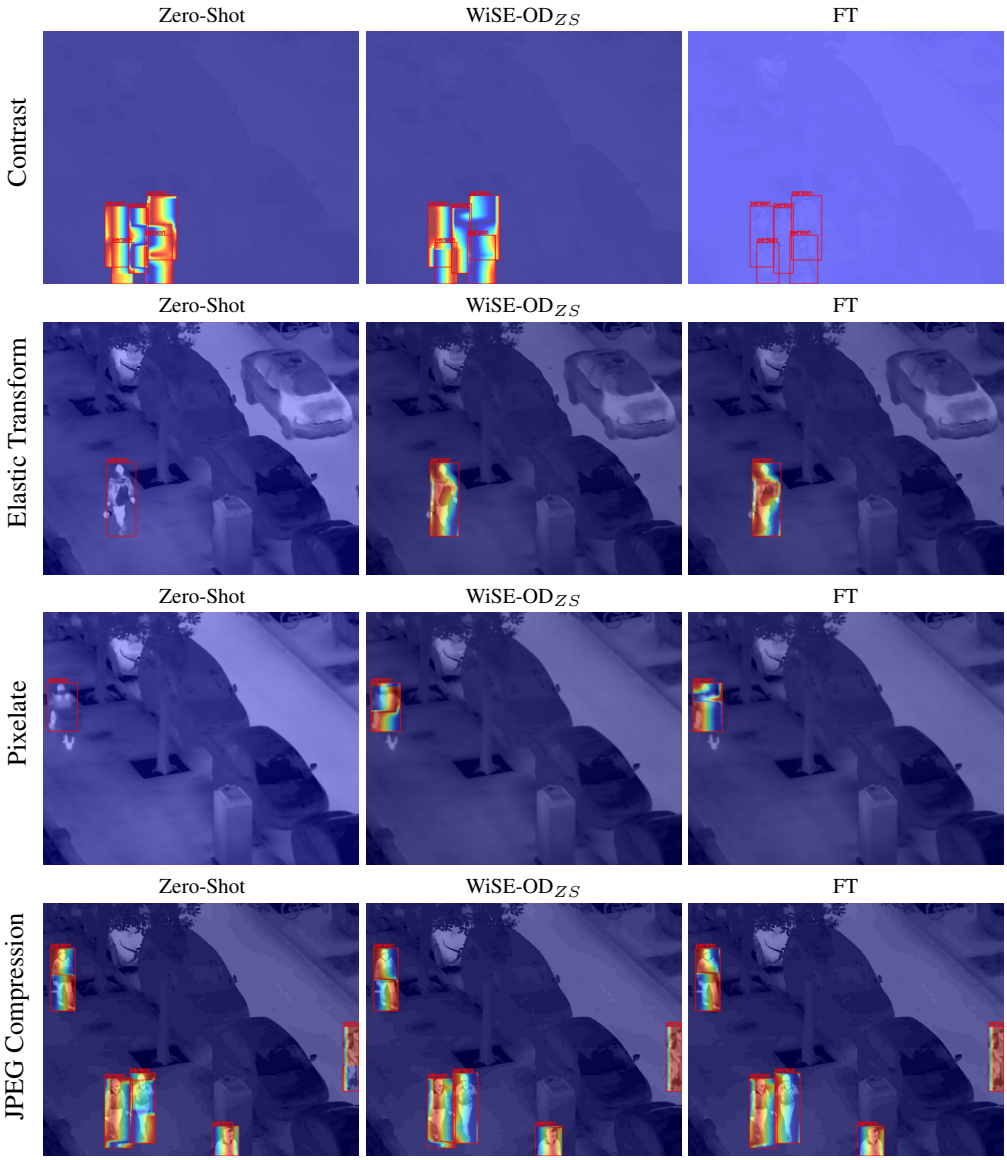

Figure 7. **Activation map analysis for zero-shot COCO pre-train Faster R-CNN detector, the WiSE-OD$_{ZS}$ and FT detector on IR for LLVIP-C dataset**. In red are the GTs, and for the WiSE-OD$_{ZS}$, the models are able to activate the features that represent a person for such corruptions (Part 3 with 4 of the 14 corruptions).

mislocalized by ZS and FT. This highlights the effectiveness of WiSE-OD in improving robustness and consistency under real-world degradations.

7. Qualitative results different levels of severity

Figure 13 presents visual examples of synthetic corruptions applied to LLVIP infrared images, specifically fog, brightness, and contrast perturbations. Each row corresponds to one corruption type, while columns represent increasing severity levels from 1 to 5. As the severity increases, the images exhibit progressively stronger degradations, such as reduced visibility in fog, oversaturation in brightness, and loss of detail in contrast. These examples illustrate the range of challenging conditions used to evaluate model robustness.

8. Additional Details Benchmark

For the our benchmark, follows the 14 corruption types and severities of the ImageNet-C [1]/COCO-C [2] protocol. Below, we detail the corruptions.

Gaussian Noise: This corruption adds Gaussian-distributed noise with zero mean and severity-dependent standard deviation. The severity levels correspond to σ =

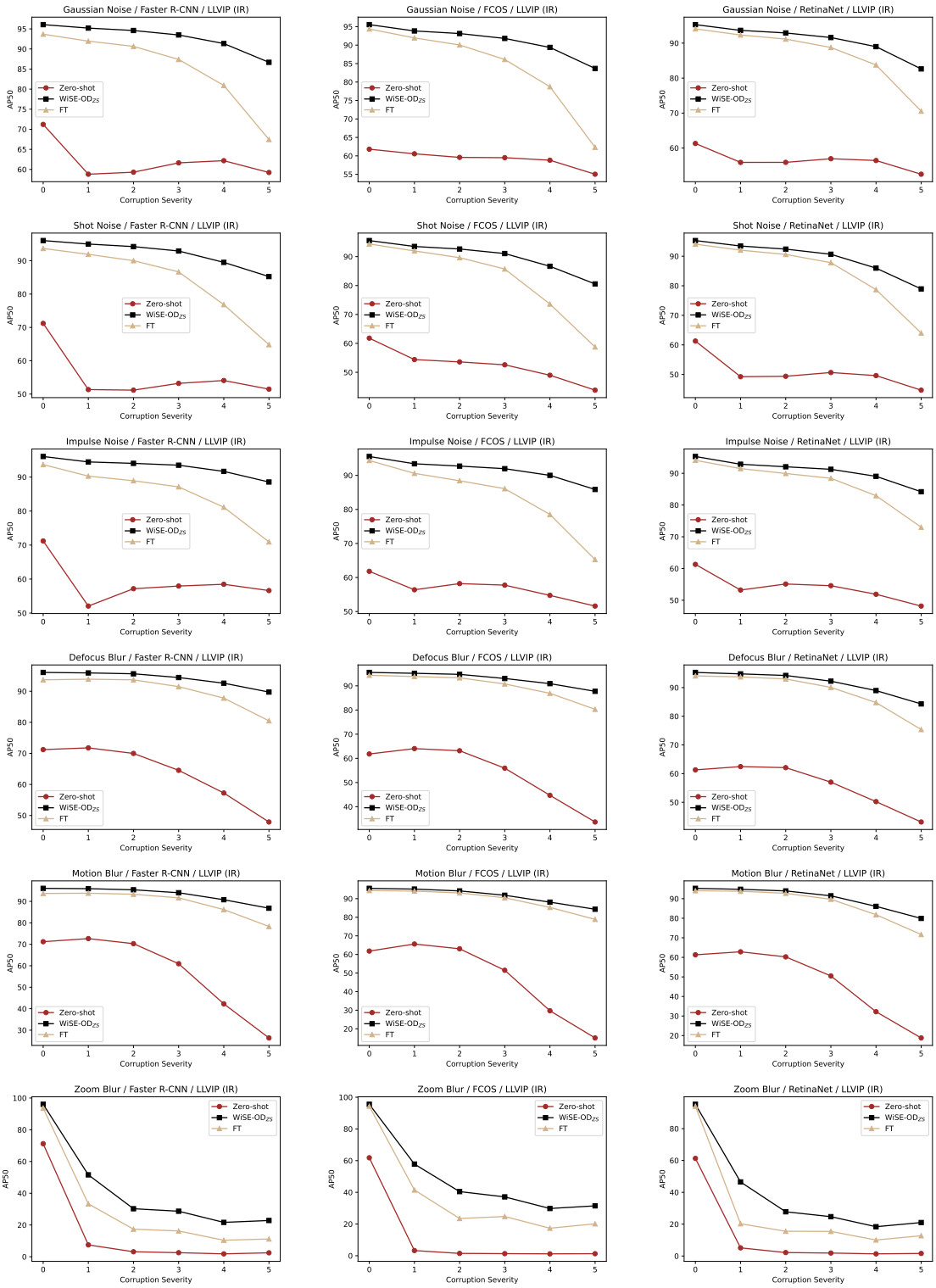

Figure 8. **AP$_{50}$ performance for all detectors over different corruption severity levels for Gaussian Blur, Shot Noise, Impulse Noise, Defocus Blur, Motion Blur and Zoom Blur.** For each perturbation, we evaluated different levels of corruption for the Zero-Shot, WiSE-OD$_{ZS}$, and FT models for LLVIP-C.

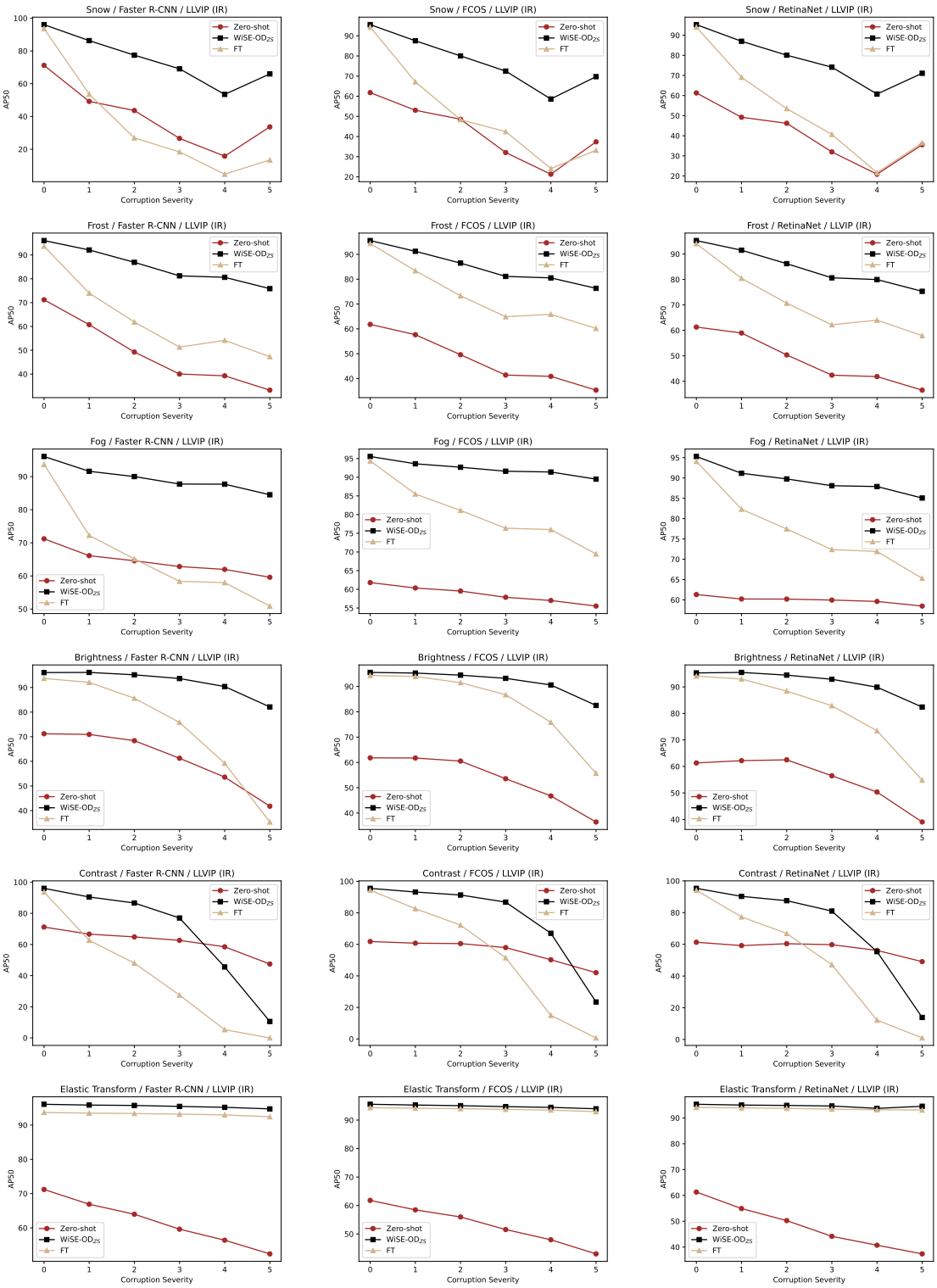

Figure 9. **AP$_{50}$ performance for all detectors over different corruption severity levels for Snow, Frost, Fog, Brightness, Contrast, Elastic Transform.** For each perturbation, we evaluated different levels of corruption for the Zero-Shot, WiSE-OD$_{ZS}$, and FT models for LLVIP-C.

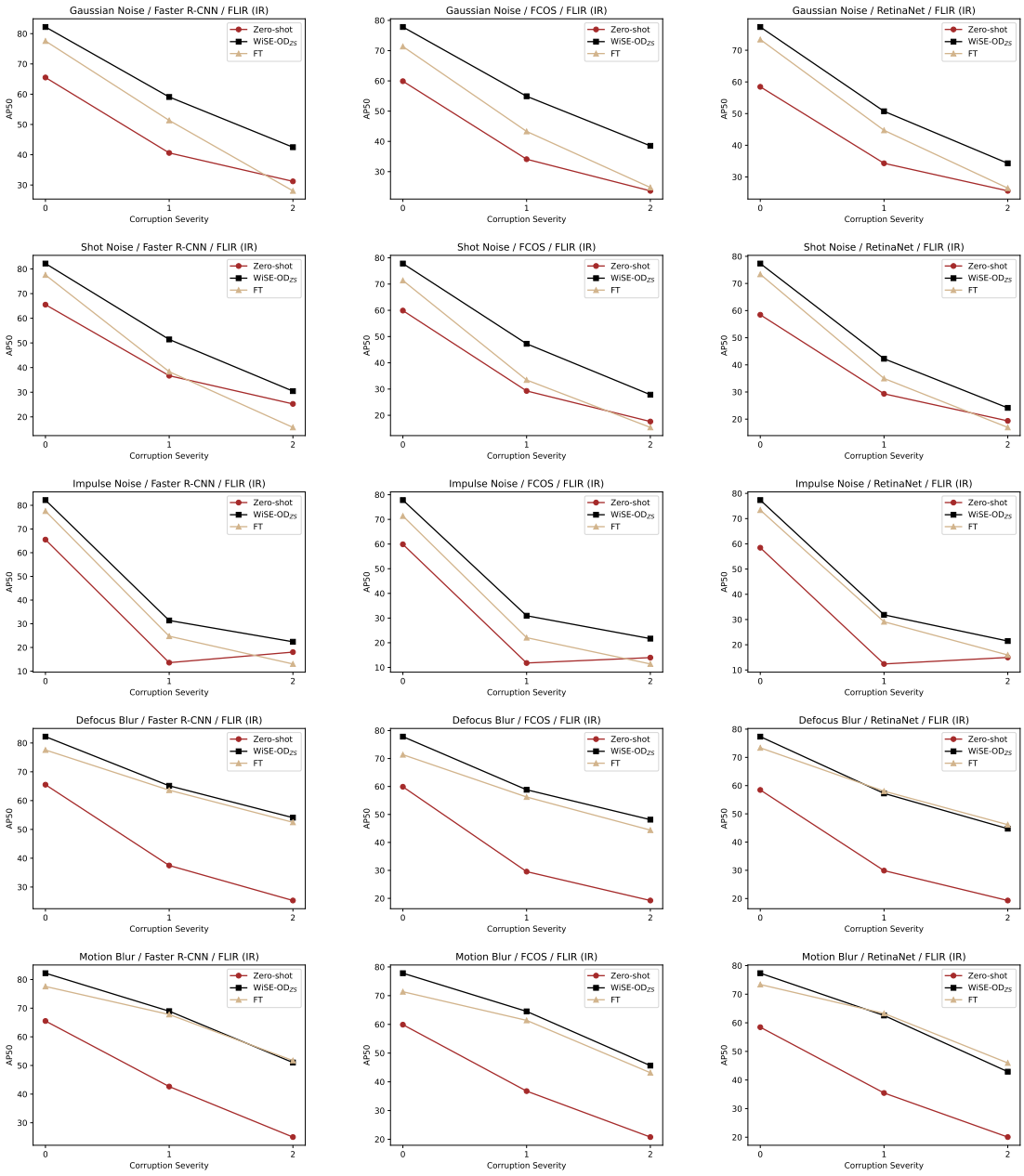

Figure 10. **AP$_{50}$ performance for all detectors over different corruption severity levels for Gaussian Noise, Shot Noise, Impulse Noise, Defocus Blur, Motion Blur.** For each perturbation, we evaluated different levels of corruption for the Zero-Shot, WiSE-OD$_{ZS}$, and FT models for FLIR-C.

$\{0.08, 0.12, 0.18, 0.26, 0.38\}$, where higher values introduce stronger random pixel fluctuations and increasingly degrade the image.

Shot Noise: This corruption simulates sensor-related Poisson noise, where severity controls the Poisson rate parameter. The values are $\lambda = \{60, 25, 12, 5, 3\}$, with lower λ producing higher variance and more pronounced noise. At higher severity, images appear grainier and more distorted.

Impulse Noise: This corruption applies salt-and-pepper noise by randomly flipping pixel values to black or white with a given probability. The severity levels are $p = \{0.03, 0.06, 0.09, 0.17, 0.27\}$, indicating the fraction of corrupted pixels. Increasing severity corresponds to a higher proportion of corrupted pixels, making the degradation progressively stronger.

Defocus Blur. This corruption mimics camera defocus

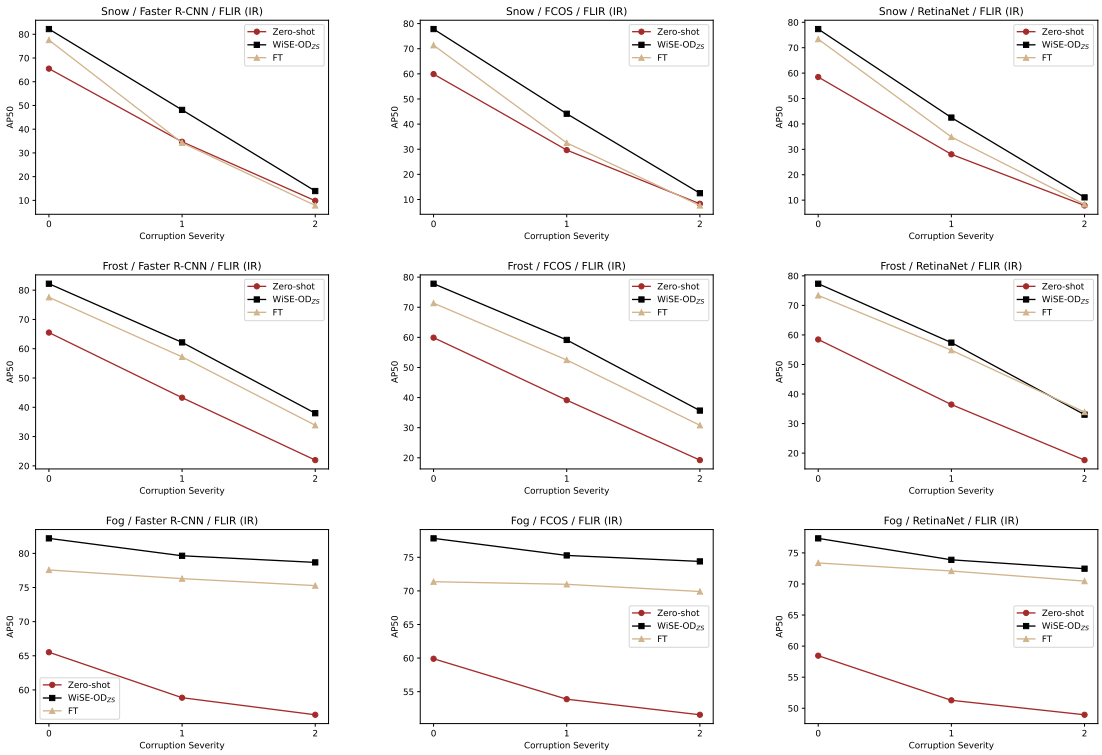

Figure 11. **AP$_{50}$ performance for all detectors over different corruption severity levels for Snow, Frost and Fog.** For each perturbation, we evaluated different levels of corruption for the Zero-Shot, WiSE-OD$_{ZS}$, and FT models for FLIR-C.

Table 3. **AP$_{50}$ performance over the perturbations for FLIR-C with severity level 2 for FCOS.**

	FLIR-C					
	Zero-Shot	FT	LP	LP-FT	WiSE-OD$_{ZS}$	WiSE-OD$_{LP}$
Gaussian Noise	23.65 ± 0.16	24.76 ± 2.44	32.89 ± 0.41	31.78 ± 1.07	38.53 ± 1.47	38.35 ± 1.20
Shot Noise	17.63 ± 0.26	15.39 ± 1.80	27.41 ± 0.17	25.10 ± 0.12	27.85 ± 1.06	30.10 ± 0.43
Impulse Noise	14.12 ± 0.23	11.34 ± 0.88	20.93 ± 0.53	18.22 ± 0.05	21.07 ± 1.25	23.10 ± 0.19
Defocus Blur	19.23 ± 0.00	44.40 ± 1.40	33.08 ± 0.00	32.27 ± 0.32	48.14 ± 1.00	48.75 ± 0.85
Motion Blur	20.77 ± 0.21	43.17 ± 1.44	33.95 ± 0.62	33.58 ± 0.31	45.65 ± 1.29	48.20 ± 0.50
Zoom Blur	6.73 ± 0.00	15.32 ± 0.14	11.81 ± 0.00	11.17 ± 0.00	15.44 ± 0.52	16.38 ± 0.54
Snow	8.29 ± 0.32	07.63 ± 1.31	12.32 ± 0.18	13.04 ± 0.22	12.50 ± 1.08	11.42 ± 0.14
Frost	19.23 ± 0.26	30.80 ± 0.69	28.14 ± 0.27	27.60 ± 0.35	35.68 ± 0.55	36.37 ± 0.69
Fog	51.56 ± 0.07	70.56 ± 0.41	63.60 ± 0.53	62.12 ± 0.14	74.39 ± 0.84	72.79 ± 0.26
Brightness	58.43 ± 0.00	69.69 ± 1.05	64.79 ± 0.00	64.78 ± 0.00	75.59 ± 0.94	73.02 ± 0.17
Contrast	50.28 ± 0.00	70.41 ± 0.96	62.97 ± 0.10	61.36 ± 0.00	73.99 ± 0.69	73.02 ± 0.33
Elastic transform	36.76 ± 0.48	64.18 ± 0.74	54.62 ± 0.82	52.40 ± 0.75	68.95 ± 0.52	66.47 ± 0.32
Pixelate	32.65 ± 0.00	51.74 ± 0.86	45.58 ± 0.00	47.96 ± 0.00	59.03 ± 0.09	57.42 ± 0.00
JPEG compression	44.71 ± 0.00	55.63 ± 0.96	52.90 ± 0.00	52.61 ± 0.00	63.03 ± 0.74	59.33 ± 0.00
mPC	28.85	41.07	38.92	38.14	**47.13**	46.76

Table 4. **AP$_{50}$ performance over the perturbations for FLIR-C with severity level 2 for RetinaNet.**

	FLIR-C					
	Zero-Shot	FT	LP	LP-FT	WiSE-OD$_{ZS}$	WiSE-OD$_{LP}$
Gaussian Noise	25.61 ± 0.20	26.47 ± 3.14	33.34 ± 0.38	31.44 ± 0.52	34.31 ± 2.83	33.09 ± 0.26
Shot Noise	19.36 ± 0.06	16.95 ± 2.60	25.49 ± 0.26	24.16 ± 0.26	24.15 ± 2.42	29.00 ± 0.85
Impulse Noise	14.82 ± 0.17	15.90 ± 1.98	18.58 ± 0.28	19.12 ± 0.28	21.77 ± 1.50	24.58 ± 0.37
Defocus Blur	19.30 ± 0.00	46.15 ± 1.86	31.28 ± 0.00	30.62 ± 0.00	44.74 ± 1.61	48.43 ± 0.63
Motion Blur	20.05 ± 0.16	45.96 ± 3.97	29.39 ± 0.11	30.28 ± 0.16	42.91 ± 2.90	50.99 ± 0.23
Zoom Blur	07.05 ± 0.00	15.99 ± 1.06	10.48 ± 0.00	10.55 ± 0.00	15.40 ± 0.32	16.03 ± 0.00
Snow	07.91 ± 0.16	08.34 ± 1.84	09.29 ± 0.08	09.54 ± 0.34	11.09 ± 1.44	12.80 ± 1.38
Frost	17.63 ± 0.32	33.90 ± 3.17	24.81 ± 0.34	24.31 ± 0.44	33.02 ± 2.41	36.32 ± 0.31
Fog	48.95 ± 0.13	70.48 ± 0.40	60.61 ± 0.11	60.32 ± 0.11	72.46 ± 0.50	74.75 ± 1.09
Brightness	56.74 ± 0.00	70.06 ± 0.30	61.90 ± 0.00	63.21 ± 0.00	74.80 ± 0.56	74.52 ± 1.18
Contrast	47.50 ± 0.00	70.75 ± 0.38	59.36 ± 0.00	59.94 ± 0.00	71.89 ± 0.43	74.44 ± 0.48
Elastic transform	34.27 ± 0.34	65.63 ± 0.54	50.84 ± 0.50	49.74 ± 0.42	68.19 ± 0.54	68.58 ± 1.09
Pixelate	32.52 ± 0.00	55.52 ± 1.21	45.84 ± 0.00	45.96 ± 0.00	57.66 ± 0.30	59.29 ± 0.00
JPEG compression	44.10 ± 0.00	55.89 ± 0.86	52.80 ± 0.00	51.23 ± 0.00	62.57 ± 1.02	62.71 ± 0.00
mPC	28.27	42.71	36.71	36.45	45.35	**47.53**

by convolving the image with a disk kernel of increasing radius. The severity levels are defined as $(r, \alpha) = \{(3, 0.1), (4, 0.5), (6, 0.5), (8, 0.5), (10, 0.5)\}$, where r is the kernel radius and α the aliasing blur factor. Larger radii correspond to stronger out-of-focus effects, leading to significant loss of image sharpness at higher severity.

Motion Blur. This corruption simulates the effect of camera or object motion during exposure by applying a blur kernel with random orientation. The severity levels are parameterized as $(r, \sigma) = \{(10, 3), (15, 5), (15, 8), (15, 12), (20, 15)\}$, where r controls the blur radius and σ the smoothing strength. Higher severity corresponds to longer and stronger motion streaks, significantly reducing image clarity.

Zoom Blur. This corruption simulates the effect of zooming a camera lens while capturing an image, producing radial streaks that blur the scene outward from the center. The severity levels are defined by different zoom factor ranges:

$$[1, 1.11],\ [1, 1.16],\ [1, 1.21],\ [1, 1.26],\ [1, 1.31],$$

with progressively larger step sizes. Higher severity introduces stronger zoom distortions, reducing image sharpness

Table 5. **Ablation of λ over LLVIP-C dataset for Faster R-CNN. Where $\lambda = 0.0$ represents the zero-shot model, $\lambda = 0.5$ represents default WiSE-OD$_{ZS}$ and $\lambda = 1.0$ represents the fine-tuning model. For LLVIP-C, the severity level is 5.**

	Test Set IR (Dataset: LLVIP-C)										
	$\theta(\lambda = 0.0)$	$\theta(\lambda = 0.1)$	$\theta(\lambda = 0.2)$	$\theta(\lambda = 0.3)$	$\theta(\lambda = 0.4)$	$\theta(\lambda = 0.5)$	$\theta(\lambda = 0.6)$	$\theta(\lambda = 0.7)$	$\theta(\lambda = 0.8)$	$\theta(\lambda = 0.9)$	$\theta(\lambda = 1.0)$
Original	71.21 ± 0.02	90.88 ± 0.12	93.88 ± 0.28	95.16 ± 0.10	95.73 ± 0.15	96.06 ± 0.22	96.03 ± 0.29	95.88 ± 0.33	95.41 ± 0.60	94.72 ± 0.53	93.68 ± 0.86
Gaussian Noise	59.24 ± 0.07	83.30 ± 0.26	86.52 ± 0.40	87.13 ± 0.50	87.34 ± 0.15	86.68 ± 0.44	85.04 ± 1.02	82.37 ± 1.94	78.47 ± 3.43	73.39 ± 5.44	67.46 ± 7.45
Shot Noise	51.48 ± 0.14	80.36 ± 0.15	83.86 ± 0.70	85.49 ± 0.60	86.00 ± 0.16	85.26 ± 0.50	83.47 ± 1.19	80.54 ± 2.08	76.42 ± 3.74	71.03 ± 5.67	64.83 ± 7.79
Impulse Noise	56.62 ± 0.07	82.52 ± 0.23	86.93 ± 0.65	88.47 ± 0.83	88.90 ± 0.32	88.54 ± 0.33	87.09 ± 0.89	84.64 ± 1.73	80.94 ± 2.70	76.78 ± 4.21	71.32 ± 6.33
Defocus Blur	47.90 ± 0.08	82.35 ± 0.29	88.41 ± 0.31	89.78 ± 0.61	90.22 ± 0.78	89.74 ± 0.98	88.89 ± 1.54	87.42 ± 2.19	85.85 ± 2.55	83.41 ± 3.10	80.48 ± 3.60
Motion Blur	26.39 ± 0.23	71.73 ± 0.89	81.10 ± 0.44	85.01 ± 0.23	86.54 ± 0.49	86.81 ± 0.71	86.30 ± 1.22	84.99 ± 1.56	83.24 ± 1.70	80.90 ± 2.21	78.32 ± 3.18
Zoom Blur	02.47 ± 0.02	20.59 ± 0.62	27.97 ± 0.62	28.10 ± 1.23	26.04 ± 2.00	22.83 ± 2.44	19.69 ± 2.18	17.03 ± 2.09	14.46 ± 1.82	12.63 ± 1.72	11.18 ± 1.56
Snow	33.65 ± 0.01	59.62 ± 0.70	67.67 ± 0.77	70.10 ± 0.36	69.55 ± 1.04	65.97 ± 1.90	59.54 ± 2.56	50.03 ± 3.09	38.50 ± 2.61	25.35 ± 3.58	13.46 ± 4.45
Frost	33.25 ± 0.38	63.68 ± 0.18	72.31 ± 0.23	75.45 ± 0.31	76.36 ± 0.33	75.87 ± 0.39	73.83 ± 0.27	70.41 ± 0.37	65.10 ± 0.57	57.51 ± 1.79	47.32 ± 3.45
Fog	59.60 ± 0.10	85.74 ± 0.04	89.79 ± 0.43	90.17 ± 0.96	88.01 ± 1.99	84.51 ± 3.80	78.78 ± 6.13	72.14 ± 8.00	64.47 ± 9.62	57.25 ± 10.1	50.90 ± 10.0
Brightness	41.77 ± 0.03	75.37 ± 0.17	82.38 ± 0.17	84.12 ± 0.20	84.07 ± 0.58	82.10 ± 1.20	78.20 ± 1.97	71.94 ± 2.81	62.81 ± 3.16	49.92 ± 4.49	35.36 ± 6.97
Contrast	47.48 ± 0.04	58.36 ± 1.99	50.59 ± 4.48	34.73 ± 5.82	20.25 ± 4.48	10.57 ± 3.82	04.99 ± 2.13	02.25 ± 1.17	00.77 ± 0.31	00.33 ± 0.47	00.00 ± 0.00
Elastic transform	52.42 ± 0.18	84.78 ± 0.42	89.92 ± 0.51	92.67 ± 0.27	94.10 ± 0.05	94.72 ± 0.07	94.94 ± 0.11	94.92 ± 0.17	94.32 ± 0.34	93.73 ± 0.36	92.41 ± 0.93
Pixelate	03.95 ± 0.01	42.41 ± 1.64	66.27 ± 3.14	76.01 ± 3.53	81.29 ± 3.70	85.06 ± 3.32	87.01 ± 2.91	88.18 ± 2.82	88.97 ± 2.35	88.75 ± 2.39	87.69 ± 2.67
JPEG compression	57.22 ± 0.02	82.36 ± 0.73	87.87 ± 0.89	90.49 ± 1.05	91.95 ± 1.11	92.59 ± 1.22	92.62 ± 1.25	92.42 ± 1.20	91.86 ± 1.38	90.53 ± 1.47	88.93 ± 1.69
mPC	40.96	69.51	75.82	76.98	76.47	75.08	72.88	69.94	66.15	61.53	56.40

Table 6. **Ablation of λ over FLIR-C dataset for Faster R-CNN. Where $\lambda = 0.0$ represents the zero-shot model, $\lambda = 0.5$ represents default WiSE-OD$_{ZS}$ and $\lambda = 1.0$ represents the fine-tuning model. For FLIR-C, the severity level is 2.**

	Test Set IR (Dataset: FLIR-C)										
	$\theta(\lambda = 0.0)$	$\theta(\lambda = 0.1)$	$\theta(\lambda = 0.2)$	$\theta(\lambda = 0.3)$	$\theta(\lambda = 0.4)$	$\theta(\lambda = 0.5)$	$\theta(\lambda = 0.6)$	$\theta(\lambda = 0.7)$	$\theta(\lambda = 0.8)$	$\theta(\lambda = 0.9)$	$\theta(\lambda = 1.0)$
Original	65.52 ± 0.07	73.51 ± 0.14	77.49 ± 0.10	79.74 ± 0.10	80.87 ± 0.09	82.20 ± 0.07	81.31 ± 0.17	80.95 ± 0.26	80.18 ± 0.11	79.03 ± 0.09	77.57 ± 0.24
Gaussian Noise	31.21 ± 0.29	38.94 ± 1.77	42.62 ± 2.92	44.05 ± 3.60	44.11 ± 4.31	42.49 ± 4.48	40.60 ± 4.35	37.88 ± 4.13	34.85 ± 3.85	31.43 ± 3.37	28.07 ± 2.91
Shot Noise	25.26 ± 0.12	31.53 ± 1.55	33.91 ± 2.98	33.86 ± 4.03	32.81 ± 4.12	30.45 ± 3.96	27.89 ± 3.63	24.81 ± 3.25	21.88 ± 2.93	18.79 ± 2.52	15.73 ± 2.05
Impulse Noise	17.69 ± 0.03	22.94 ± 1.52	24.85 ± 2.26	25.00 ± 2.70	24.15 ± 2.64	22.51 ± 2.78	21.18 ± 2.89	18.95 ± 2.49	16.96 ± 2.17	15.15 ± 2.33	13.22 ± 2.27
Defocus Blur	25.32 ± 0.22	37.37 ± 1.81	44.57 ± 2.40	49.45 ± 2.37	52.30 ± 2.01	54.08 ± 1.74	54.99 ± 1.38	55.33 ± 1.10	55.00 ± 0.98	54.01 ± 0.64	52.47 ± 0.99
Motion Blur	25.01 ± 0.25	34.24 ± 1.35	40.63 ± 2.17	45.17 ± 2.38	48.75 ± 2.17	51.03 ± 2.16	53.19 ± 2.13	53.96 ± 2.00	53.85 ± 2.19	53.29 ± 2.18	51.71 ± 2.12
Zoom Blur	08.98 ± 0.05	11.77 ± 0.49	13.72 ± 0.97	15.20 ± 1.01	16.10 ± 1.01	16.93 ± 1.00	17.69 ± 1.10	18.02 ± 1.11	18.32 ± 1.09	18.24 ± 0.94	17.97 ± 0.90
Snow	09.84 ± 0.14	13.39 ± 1.37	14.55 ± 2.07	15.19 ± 2.60	14.85 ± 2.69	13.94 ± 2.66	12.73 ± 2.68	11.47 ± 2.59	10.36 ± 2.48	08.99 ± 2.25	07.86 ± 2.01
Frost	21.96 ± 0.50	29.17 ± 1.92	33.37 ± 2.39	35.98 ± 2.87	37.34 ± 3.17	37.97 ± 3.63	37.71 ± 3.59	37.22 ± 3.82	36.47 ± 4.00	35.17 ± 4.45	33.87 ± 4.69
Fog	56.36 ± 0.28	67.20 ± 1.18	72.17 ± 0.86	75.52 ± 0.67	77.56 ± 0.97	78.68 ± 1.24	78.85 ± 1.00	78.71 ± 1.23	78.11 ± 1.49	76.92 ± 1.12	73.61 ± 0.06
Brightness	64.41 ± 0.26	71.92 ± 0.51	75.68 ± 0.19	75.60 ± 0.13	79.09 ± 0.49	79.72 ± 0.35	79.53 ± 0.96	78.56 ± 1.28	77.79 ± 1.24	76.54 ± 1.33	75.18 ± 0.99
Contrast	54.59 ± 0.04	66.11 ± 0.79	71.38 ± 0.95	74.78 ± 1.19	77.03 ± 0.90	78.36 ± 1.06	78.62 ± 1.12	78.36 ± 0.99	78.02 ± 1.09	76.87 ± 1.15	75.47 ± 1.29
Elastic transform	41.88 ± 0.24	55.93 ± 0.47	63.89 ± 0.37	68.62 ± 0.39	71.71 ± 0.79	73.39 ± 0.40	73.66 ± 0.41	73.37 ± 0.25	72.51 ± 0.58	71.51 ± 0.42	69.68 ± 1.15
Pixelate	38.67 ± 0.11	49.47 ± 1.66	55.23 ± 2.37	58.37 ± 2.52	60.02 ± 2.70	61.12 ± 3.13	60.93 ± 3.43	60.12 ± 4.13	58.87 ± 4.58	57.14 ± 5.60	54.91 ± 6.21
JPEG compression	50.24 ± 0.14	59.12 ± 0.61	63.21 ± 0.56	65.59 ± 0.63	66.55 ± 0.80	66.65 ± 0.89	66.12 ± 1.48	64.94 ± 1.84	63.27 ± 2.36	60.63 ± 2.39	57.55 ± 3.04
mPC	33.67	42.07	46.41	48.74	50.16	50.52	50.26	49.40	48.30	46.76	44.80

and creating the illusion of rapid inwards or outwards motion.

Fog. This corruption simulates the effect of atmospheric fog by blending the image with fractal noise patterns of varying intensity and smoothness. The severity levels are parameterized as $(\alpha, \beta) = \{(1.5, 2), (2, 2), (2.5, 1.7), (2.5, 1.5), (3.0, 1.4)\}$, where α controls the fog density and β the fractal decay rate. Higher severity produces denser fog, reducing scene contrast and obscuring objects, particularly in the background.

Brightness. This corruption modifies the overall intensity of an image by linearly adjusting pixel values. The severity levels correspond to additive shifts $c = \{0.1, 0.2, 0.3, 0.4, 0.5\}$, where larger values progressively brighten the image. At higher severities, the scene becomes overexposed, washing out details.

Saturate. This corruption changes the saturation of an image in HSV color space, either reducing it toward grayscale or amplifying color intensity. The severity levels are defined as $(\alpha, \beta) = \{(0.3, 0), (0.1, 0), (2, 0), (5, 0.1), (20, 0.2)\}$, where α controls saturation scaling and β adds variability. Low severity reduces colors, while high severity creates oversaturated, unnatural-looking images.

Pixelate. This corruption reduces image resolution and then upscales it back, producing blocky artifacts that mimic low-quality image compression. The severity levels are $c = \{0.6, 0.5, 0.4, 0.3, 0.25\}$, where smaller values correspond to lower effective resolution. Higher severity leads to coarser pixelation and significant loss of fine details.

Elastic Transform. This corruption applies random elastic deformations by perturbing pixel coordinates with smoothed displacement fields. The severity levels are controlled by $\alpha = \{0.05, 0.085, 0.1, 0.12\}$, which scales the deformation intensity relative to image size. At higher

Table 7. **Ablation of λ over FLIR-C dataset for FCOS. Where $\lambda = 0.0$ represents the zero-shot model, $\lambda = 0.5$ represents default WiSE-OD$_{ZS}$ and $\lambda = 1.0$ represents the fine-tuning model. For FLIR-C, the severity level is 2.**

	Test Set IR (Dataset: FLIR-C)										
	$\theta(\lambda = 0.0)$	$\theta(\lambda = 0.1)$	$\theta(\lambda = 0.2)$	$\theta(\lambda = 0.3)$	$\theta(\lambda = 0.4)$	$\theta(\lambda = 0.5)$	$\theta(\lambda = 0.6)$	$\theta(\lambda = 0.7)$	$\theta(\lambda = 0.8)$	$\theta(\lambda = 0.9)$	$\theta(\lambda = 1.0)$
Original	59.90 ± 0.00	69.17 ± 0.00	73.37 ± 0.70	75.92 ± 0.61	76.42 ± 0.00	77.82 ± 0.00	76.13 ± 0.70	75.64 ± 1.17	73.84 ± 0.00	73.18 ± 1.00	71.78 ± 1.27
Gaussian Noise	23.65 ± 0.16	35.87 ± 0.60	39.92 ± 0.34	40.94 ± 0.69	40.21 ± 1.22	38.53 ± 1.47	36.49 ± 1.58	33.92 ± 1.84	31.18 ± 1.93	27.93 ± 2.20	24.76 ± 2.44
Shot Noise	17.63 ± 0.26	28.36 ± 0.75	31.61 ± 0.30	31.49 ± 0.15	30.10 ± 0.62	27.85 ± 1.06	25.41 ± 1.18	22.97 ± 1.38	20.49 ± 1.46	17.95 ± 1.66	15.39 ± 1.80
Impulse Noise	14.12 ± 0.23	22.48 ± 0.30	24.23 ± 0.54	23.70 ± 0.82	22.94 ± 1.06	21.07 ± 1.25	19.60 ± 0.93	17.57 ± 0.84	15.50 ± 0.70	13.98 ± 0.97	11.34 ± 0.88
Defocus Blur	19.23 ± 0.00	32.58 ± 0.60	39.97 ± 0.61	44.29 ± 1.13	46.66 ± 1.10	48.14 ± 1.00	48.96 ± 1.24	48.92 ± 1.33	47.87 ± 1.58	46.38 ± 1.54	44.40 ± 1.40
Motion Blur	20.77 ± 0.21	31.73 ± 0.59	37.81 ± 0.69	41.68 ± 1.03	44.17 ± 1.43	45.65 ± 1.29	46.54 ± 0.90	46.60 ± 0.96	46.01 ± 1.07	44.90 ± 1.19	43.17 ± 1.44
Zoom Blur	06.73 ± 0.00	11.12 ± 0.35	13.23 ± 0.43	14.27 ± 0.52	14.94 ± 0.54	15.44 ± 0.52	15.70 ± 0.51	15.89 ± 0.18	16.06 ± 0.19	15.91 ± 0.15	15.32 ± 0.14
Snow	08.29 ± 0.32	11.98 ± 0.26	13.57 ± 0.41	13.62 ± 0.71	13.19 ± 1.05	12.50 ± 1.08	11.63 ± 1.20	10.85 ± 1.26	09.83 ± 1.24	08.73 ± 1.22	07.63 ± 1.31
Frost	19.23 ± 0.26	28.62 ± 0.65	32.77 ± 0.39	34.68 ± 0.25	35.33 ± 0.38	35.68 ± 0.55	35.44 ± 0.77	34.87 ± 0.89	33.96 ± 0.73	32.55 ± 0.68	30.80 ± 0.69
Fog	51.56 ± 0.07	63.39 ± 0.15	68.76 ± 0.18	72.17 ± 0.55	73.84 ± 0.71	74.39 ± 0.84	74.45 ± 0.85	73.36 ± 1.40	72.63 ± 1.28	71.42 ± 1.25	70.56 ± 0.41
Brightness	58.43 ± 0.00	67.75 ± 0.15	72.80 ± 0.42	75.20 ± 0.50	75.88 ± 0.88	75.59 ± 0.94	74.73 ± 0.90	73.78 ± 1.00	72.72 ± 1.04	71.21 ± 1.10	69.69 ± 1.05
Contrast	50.28 ± 0.00	62.55 ± 0.12	68.03 ± 0.30	71.68 ± 0.53	73.50 ± 0.67	73.99 ± 0.69	73.83 ± 1.02	73.28 ± 1.39	72.75 ± 1.12	71.68 ± 1.15	70.41 ± 0.96
Elastic transform	36.76 ± 0.48	53.13 ± 0.82	61.25 ± 0.88	65.65 ± 0.70	67.93 ± 0.34	68.95 ± 0.52	68.84 ± 0.48	68.30 ± 0.81	67.21 ± 1.01	65.96 ± 1.01	64.18 ± 0.74
Pixelate	32.65 ± 0.00	45.53 ± 0.92	52.09 ± 0.78	55.86 ± 0.74	57.68 ± 0.95	59.03 ± 0.09	57.56 ± 0.54	57.57 ± 0.67	56.31 ± 0.63	54.36 ± 0.82	51.74 ± 0.86
JPEG compression	44.71 ± 0.00	55.25 ± 0.38	59.69 ± 0.05	62.18 ± 0.29	63.13 ± 0.58	63.03 ± 0.74	62.49 ± 0.45	61.14 ± 0.36	59.44 ± 0.34	57.37 ± 0.64	55.63 ± 0.96
mPC	28.85	39.31	43.98	46.24	47.10	47.13	46.54	45.64	44.42	42.88	41.07

Table 8. **Ablation of λ over FLIR-C dataset for RetinaNet. Where $\lambda = 0.0$ represents the zero-shot model, $\lambda = 0.5$ represents default WiSE-OD$_{ZS}$ and $\lambda = 1.0$ represents the fine-tuning model. For FLIR-C, the severity level is 2.**

	Test Set IR (Dataset: FLIR-C)										
	$\theta(\lambda = 0.0)$	$\theta(\lambda = 0.1)$	$\theta(\lambda = 0.2)$	$\theta(\lambda = 0.3)$	$\theta(\lambda = 0.4)$	$\theta(\lambda = 0.5)$	$\theta(\lambda = 0.6)$	$\theta(\lambda = 0.7)$	$\theta(\lambda = 0.8)$	$\theta(\lambda = 0.9)$	$\theta(\lambda = 1.0)$
Original	58.46 ± 0.00	67.08 ± 0.22	71.71 ± 0.32	74.93 ± 0.15	76.38 ± 0.62	77.34 ± 0.05	77.00 ± 0.19	76.88 ± 0.01	76.08 ± 0.15	74.35 ± 0.30	73.38 ± 0.89
Gaussian Noise	25.61 ± 0.20	32.69 ± 0.58	35.32 ± 0.84	35.81 ± 1.74	35.33 ± 2.47	34.31 ± 2.83	33.28 ± 2.92	31.93 ± 3.02	30.34 ± 3.13	28.58 ± 3.11	26.47 ± 3.14
Shot Noise	19.36 ± 0.06	25.03 ± 0.61	26.67 ± 0.46	26.40 ± 1.27	25.48 ± 2.02	24.15 ± 2.42	22.78 ± 2.53	21.43 ± 2.41	19.98 ± 2.53	18.43 ± 2.63	16.95 ± 2.60
Impulse Noise	14.82 ± 0.17	20.28 ± 0.66	22.12 ± 0.18	22.72 ± 0.54	21.78 ± 1.37	21.77 ± 1.50	20.62 ± 1.79	19.08 ± 1.54	17.98 ± 1.93	16.96 ± 2.15	15.90 ± 1.98
Defocus Blur	19.30 ± 0.00	29.18 ± 0.55	35.62 ± 0.32	39.61 ± 0.79	42.52 ± 1.27	44.74 ± 1.61	46.50 ± 1.67	47.22 ± 1.77	47.46 ± 1.81	47.04 ± 1.84	46.15 ± 1.86
Motion Blur	20.05 ± 0.16	26.99 ± 0.15	32.06 ± 0.46	36.34 ± 1.21	39.83 ± 2.32	42.91 ± 2.90	45.13 ± 3.39	46.75 ± 3.63	47.25 ± 3.87	46.89 ± 3.94	45.96 ± 3.97
Zoom Blur	07.05 ± 0.00	10.04 ± 0.26	11.99 ± 0.40	13.25 ± 0.26	14.43 ± 0.28	15.40 ± 0.32	16.07 ± 0.52	16.43 ± 0.65	16.52 ± 0.81	16.37 ± 0.93	15.99 ± 1.06
Snow	07.91 ± 0.16	09.65 ± 0.25	10.62 ± 0.40	11.12 ± 0.85	11.29 ± 1.21	11.09 ± 1.44	10.83 ± 1.73	10.40 ± 1.91	09.79 ± 1.88	09.06 ± 1.91	08.34 ± 1.84
Frost	17.63 ± 0.32	23.01 ± 0.57	26.44 ± 0.96	29.27 ± 1.55	31.40 ± 1.95	33.02 ± 2.41	34.19 ± 2.59	34.95 ± 2.84	34.91 ± 3.13	34.53 ± 3.04	33.90 ± 3.17
Fog	48.95 ± 0.13	58.32 ± 0.38	64.25 ± 0.24	68.23 ± 0.25	70.96 ± 0.45	72.46 ± 0.50	73.24 ± 0.26	73.41 ± 0.45	72.88 ± 0.24	71.92 ± 0.28	70.48 ± 0.40
Brightness	56.74 ± 0.00	65.14 ± 0.27	69.99 ± 0.39	72.90 ± 0.28	74.43 ± 0.64	74.80 ± 0.56	74.84 ± 0.55	74.02 ± 0.40	73.09 ± 0.11	71.80 ± 0.16	70.06 ± 0.30
Contrast	47.50 ± 0.00	56.81 ± 0.50	62.87 ± 0.31	67.26 ± 0.07	70.04 ± 0.23	71.89 ± 0.43	72.81 ± 0.27	73.09 ± 0.20	72.80 ± 0.10	71.94 ± 0.12	70.75 ± 0.38
Elastic transform	34.27 ± 0.34	48.18 ± 0.94	57.07 ± 0.64	62.39 ± 0.47	65.90 ± 0.29	68.19 ± 0.54	69.25 ± 0.68	69.51 ± 0.44	68.86 ± 0.53	67.56 ± 0.53	65.63 ± 0.54
Pixelate	32.52 ± 0.00	41.93 ± 0.67	48.02 ± 0.74	52.27 ± 0.66	55.46 ± 0.47	57.66 ± 0.30	58.83 ± 0.15	59.10 ± 0.40	58.58 ± 0.58	57.48 ± 0.82	55.52 ± 1.21
JPEG compression	44.10 ± 0.00	52.89 ± 0.51	57.47 ± 0.47	60.21 ± 0.60	61.79 ± 0.86	62.57 ± 1.02	62.33 ± 0.77	61.31 ± 0.63	59.89 ± 0.65	58.10 ± 0.54	55.89 ± 0.86
mPC	28.27	35.72	40.03	42.69	44.33	45.35	45.76	45.61	45.02	44.04	42.71

severity, the images exhibit stronger local distortions, bending and warping objects in ways similar to elastic transformations seen in real-world imaging artifacts.

References

[1] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *International Conference on Learning Representations*, 2019. 2, 3, 8

[2] Claudio Michaelis, Benjamin Mitzkus, Robert Geirhos, Evgenia Rusak, Oliver Bringmann, Alexander S Ecker, Matthias Bethge, and Wieland Brendel. Benchmarking robustness in object detection: Autonomous driving when winter is coming. *arXiv preprint arXiv:1907.07484*, 2019. 8

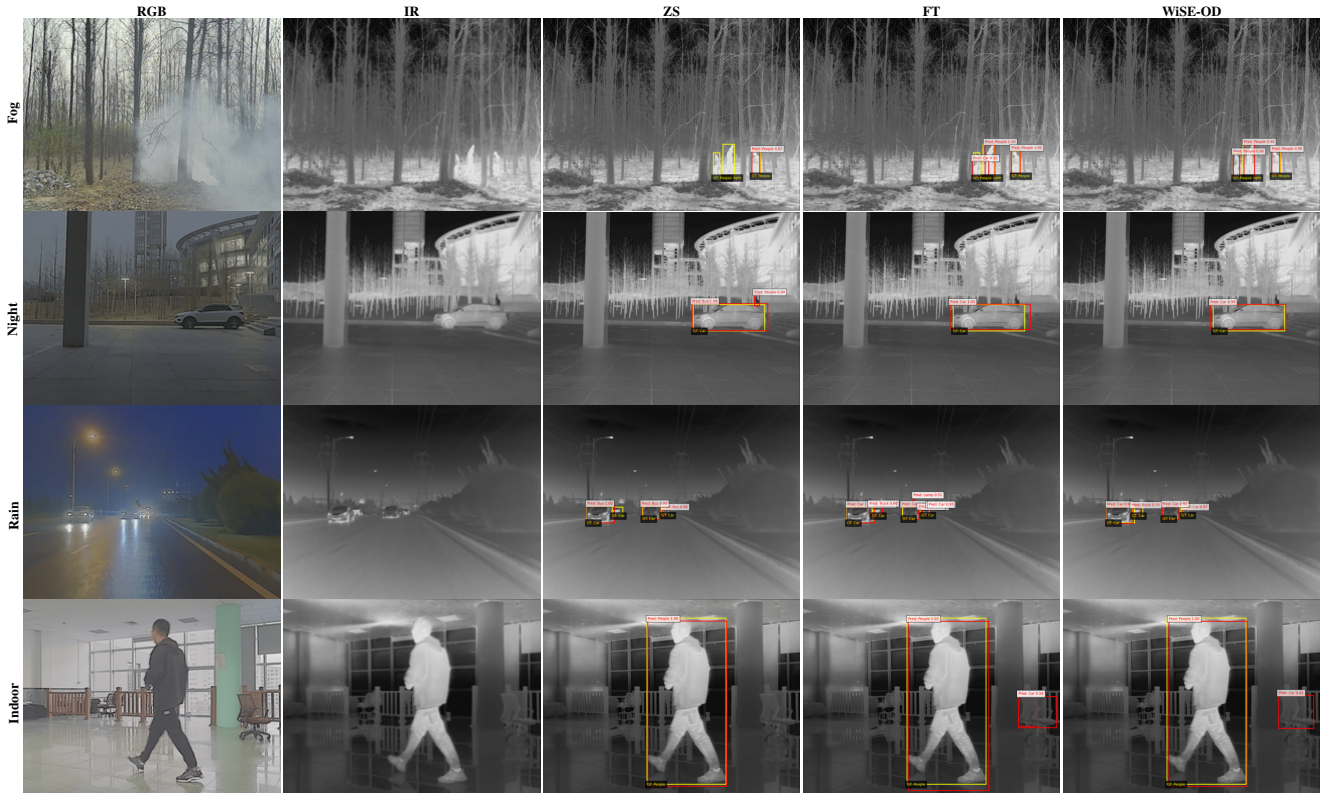

Figure 12. **Qualitative comparison on M3FD under adverse conditions.** Rows: fog, night, rain, indoor. Columns: RGB, IR, Zero-Shot (ZS), Fine-Tuning (FT), and WiSE-OD$_{ZS}$.

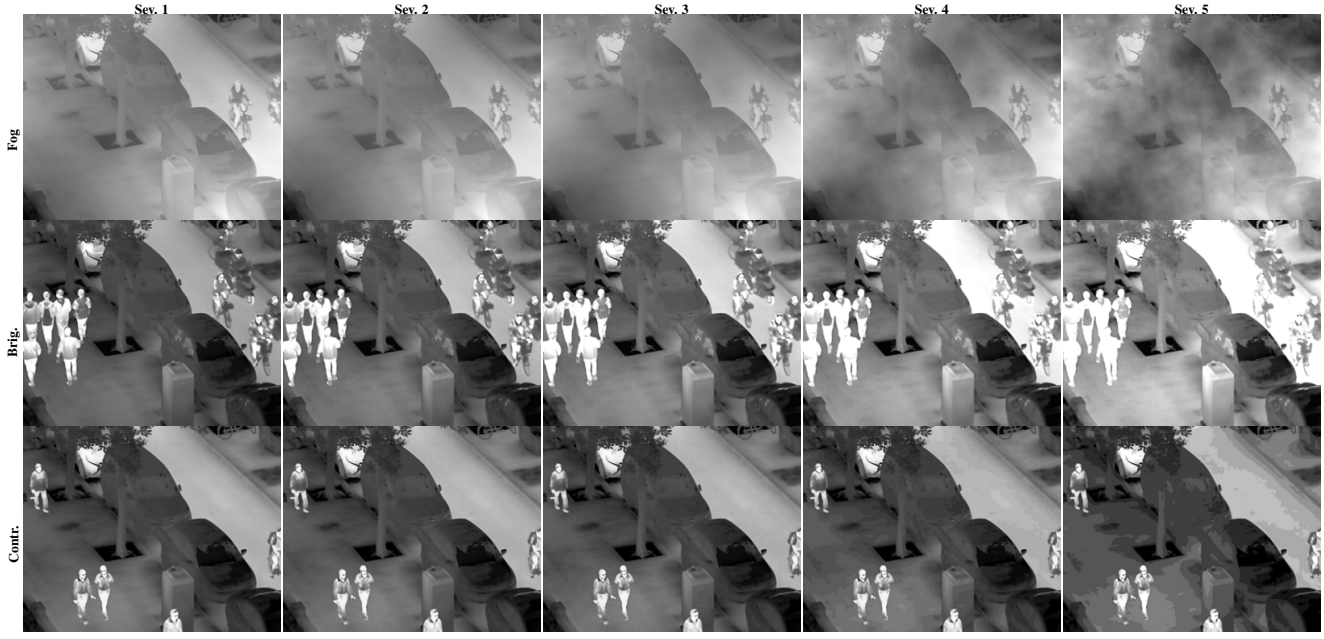

Figure 13. **Examples of fog, brightness, and contrast perturbations at different severity levels for LLVIP.** Each column shows the effect of increasing corruption severity (1–5) on infrared images. Rows: fog, brightness, and contrast. Higher severities introduce stronger degradations, simulating real-world challenging conditions.