Would SWIR modality help for detection and segmentation in harsh weather conditions? An experimental study.

Rohan Mehra [1, 2] Alexandre Riffard [1] Mathieu Labussière [1]
Pierre Duthon [3] Romuald Aufrère [1]

[1] Université Clermont Auvergne, Clermont Auvergne INP, CNRS, Institut Pascal, F-63000 Clermont-Ferrand, France
[2] Indian Institute of Science Education and Research (IISER) Bhopal, India
[3] Cerema, Research Team "Intelligent Transport Systems", F-63017 Clermont-Ferrand, France

Contact: romehraa@gmail.com, {firstname}.{lastname}@uca.fr, {firstname}.{lastname}@cerema.fr

ICCV 2025 MIRA Workshop


Introduction and Objectives

(a) Visible spectrum (b) SWIR spectrum

Figure : Comparison of detections in harsh weather conditions, here fog, in our experimental setup.

Short-Wave Infrared (SWIR), defined typically in the spectral range of 1.0 to 2.5 µm, offers distinctive advantages over other modalities in object detection under harsh weather conditions[a].

Objective: Comparative study to evaluate the performance (through deep learning models) of two SWIR technologies alongside a visible camera on controlled harsh weather conditions for driving applications.

[a] A. Bertozzi *et al.*, "Pedestrian detection in poor visibility conditions: would SWIR help?", *ICIAP*, 2013.

Presentation of Cameras

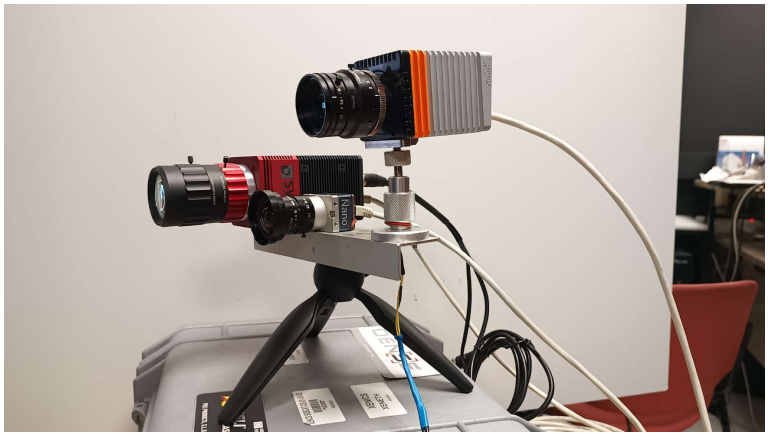

Figure : from L to R, SWIR (SVS), Visible & SWIR (Xenics) Cameras

Camera	**Type**	**Spectral Range**
Dalsa Genie Nano 1630	Visible Camera	380 nm–700 nm
Xenics Bobcat 320	InGaAs-based SWIR	900 nm–1700 nm
SVS Acuros CQD 1280	Quantum Dot-based SWIR	400 nm–1700 nm*
SWIR-11 25 mm lens M42	SWIR-coated lens	800 nm–1700 nm

Table : Information on the sensors used for analysis. * *The SVS camera is mounted with a SWIR-coated objective lens, using then an effective range of* 800 nm–1700 nm.

Dataset Acquisition Setup

Figure : Acquisition setup at Cerema, PAVIN Fog and Rain platform.

Dataset acquisition was done under controlled weather at Cerema Platform[1]. *Comparison is done per frame, 1 car and 1 person are considered in each frame.* Three weather conditions were evaluated:

1. **Clear Day (26 images):** High visibility conditions for reference.
2. **Foggy Day (87 images):** Measured in meteorological visibility (m).
3. **Rainy Day (125 images):** Measured in rainfall rate (mm/h).
   - Small sized droplets in first half - considered as light rain.
   - Droplet size increases in the second half - considered heavy rain.

[1] S. Liandrat *et al.*, "A review of Cerema PAVIN fog & rain platform: From past and back to the future," *ITS World Congress*, 2022.

(*On-the-shelf*) Deep Learning Algorithms

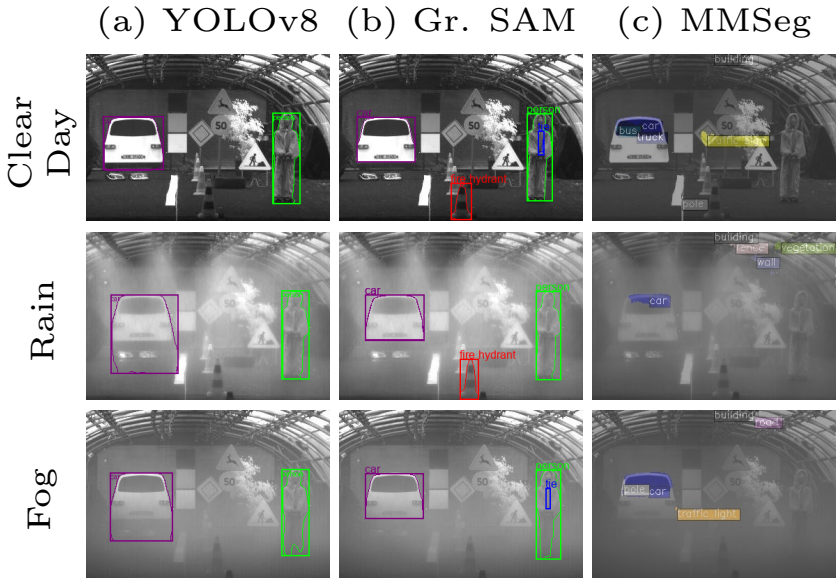

Figure : Results from detection/segmentation on-the-shelf algorithms on SWIR images.

Observations:

- YOLOv8x-seg and GroundedSAM visually give comparative performance on SWIR images.
- MMSegmentation (*default config.*) visually results in inconsistent detections. We chose to discard it from further study.

Algorithms	**Tasks**	**Avg Time/Image**
YOLOv8x-seg	Detect+Segment (COCO labels)	0.07 s to 0.1 s
YOLOv11x-seg	Detect+Segment (COCO labels)	0.07 s to 0.1 s
Grounded SAM	Detect+Segment (open set)	2 s to 4 s
MMSegmentation	Detect+Segment (Cityscapes)	0.4 s to 0.8 s

Table : List of detection and segmentation algorithms used. YOLO proves to be the fastest of the algorithms. YOLOv11x-seg was used solely for the comparison of models.

Evaluation Metrics

The objective is to obtain instance-level segmentation, and we evaluate using the following metrics:

$$\text{Segmentation Quality (SQ)} = \frac{\sum_{(p,g)\in \text{TP}} \text{IoU}(p,g)}{|\text{TP}|}$$

$$\text{Recognition Quality (RQ)} = \frac{|\text{TP}|}{|\text{TP}| + \frac{1}{2}|\text{FP}| + \frac{1}{2}|\text{FN}|}$$

$$\text{Panoptic Quality (PQ)} = \text{SQ} \times \text{RQ}$$

where p is a predicted mask, g is its corresponding ground truth mask, and (p, g) is a true positive (TP).

Note: SQ measures the segmentation quality of detected objects (TP). PQ provides an overall performance score across all classes.

Comparison of Sensing Technologies

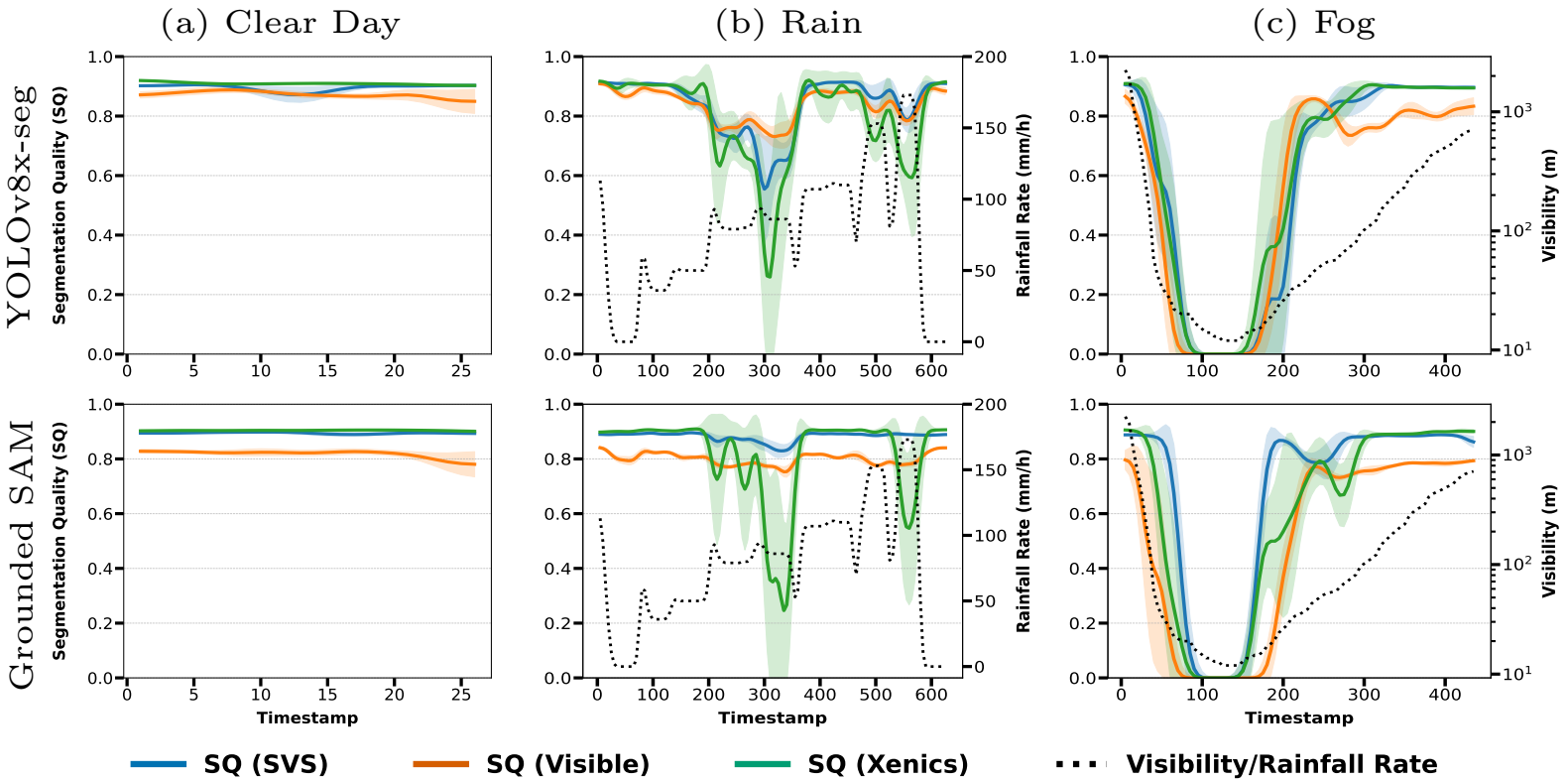

Figure : Qualitative analysis of the Segmentation Quality (SQ) as a function of time (with additional axis for visibility / rainfall rate) comparing YOLOv8x-seg and GroundedSAM.

Xenics exhibits an unstable behavior with changes in weather conditions, and the SVS and visible perform comparatively. SWIR proves to be operational for a longer duration in case of reduced visibility with fog.

Comparison of Models

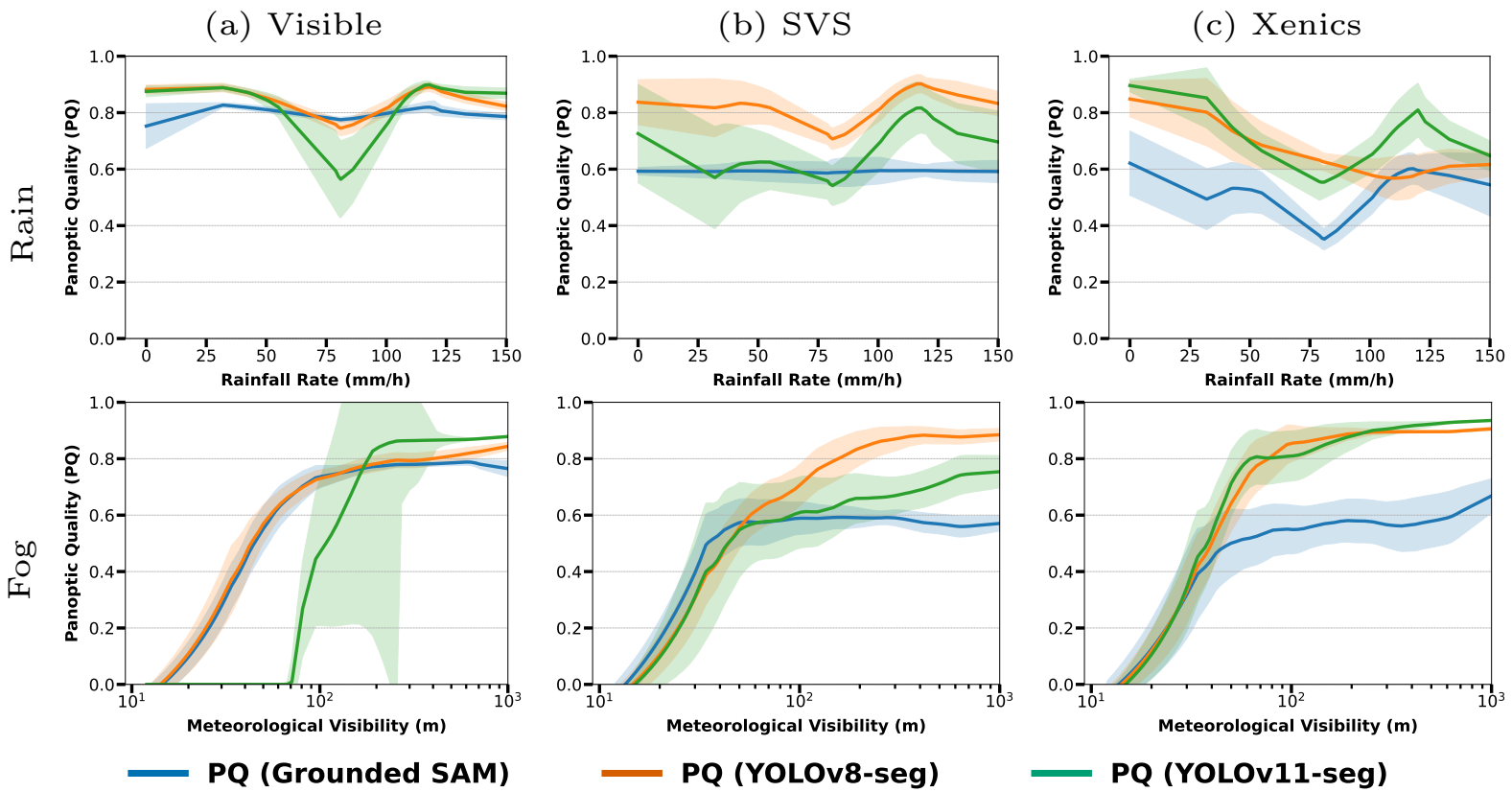

Figure : Qualitative analysis of the Panoptic Quality (PQ) as a function of visibility / rainfall rate for comparison of detection and segmentation models.

GroundedSAM yields the lowest PQ score in both conditions under high visibility due to the number of FPs detected. In fog, both the YOLO variants show comparative performance in both the SVS and Xenics modalities, whereas YOLOv11x-seg exhibits performance degradation when applied to visible camera images.

Takeaways

- SWIR cameras remain functional longer than visible cameras in dense fog and perform competitively in light/heavy rain, making them suitable for low-visibility conditions.
- Among SWIR modalities, SVS is generally more stable than Xenics.
- GroundedSAM and YOLOv8x-seg perform comparably in low-visibility; YOLOv11x-seg shows inconsistent performance.
- Considering SQ, PQ, and computational time, YOLOv8x-seg on SVS SWIR images is a good solution for perception in harsh weather.

Future Works

- Evaluating dynamic scenes using the newly published RASMD dataset[2] containing paired RGB-SWIR images.
- Fine-tune models on SWIR using domain adaptation or knowledge distillation to improve performance.
- Explore fusion of both modalities to combine visible semantics with SWIR resilience in harsh conditions.

[2] Jin et al., "RASMD: RGB and SWIR Multispectral Driving Dataset for Robust Perception in Adverse Conditions," 2025.

Questions

Thanks!

This work was supported by the International Research Center "Innovation Transportation and Production Systems" of the I-SITE CAP 20-25.

References I

[1] A. Bertozzi, R. I. Fedriga, A. Miron, and J.-L. Reverchon, "Pedestrian detection in poor visibility conditions: Would SWIR help?," in *Image Analysis and Processing – ICIAP 2013*, pp. 229–238, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.

[2] O. G. Driggers, V. Hodgkin, and R. Vollmerhausen, "What good is SWIR? Passive day comparison of VIS, NIR, and SWIR," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIV*, pp. 187–201, SPIE, 2013.

[3] M. P. Hansen and D. S. Malchow, "Overview of SWIR detectors, cameras, and applications," *Cameras, and Applications*, p. 69390I, 2013.

[4] M. S. Pavlović, P. D. Milanović, M. S. Stanković, D. B. Perić, I. V. Popadić, and M. V. Perić, "Deep learning based SWIR object detection in long-range surveillance systems: An automated cross-spectral approach," *Sensors*, vol. 22, no. 7, 2022.

[5] N. Pinchon, O. Cassignol, A. Nicolas, F. Bernardin, P. Leduc, J.-P. Tarel, R. Brémond, E. Bercier, and J. Brunet, "All-weather vision for automotive safety: which spectral band?," in *Proceedings of the 22nd International Forum on Advanced Microsystems for Automotive Applications (AMAA'18)*, pp. 3–15, Berlin, Germany, 2018. Available: `http://perso.lcpc.fr/tarel.jean-philippe/publis/amaa18.html`

[6] Y. F. Yucesoy and C. Sahin, "Object detection in infrared images with different spectra," in *2024 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pp. 1–6, 2024.

[7] Exosens, "BOBCAT/BOBCAT+," `https://www.exosens.com/products/bobcat`, last accessed: June 30, 2025.

[8] SWIR Vision Systems, "Acuros SWIR Cameras," `https://www.swirvisionsystems.com/acuros-swir-camera/`, last accessed: June 30, 2025.

[9] S. Liandrat, P. Duthon, F. Bernardin, A. Ben Daoued, and J.-L. Bicard, "A review of Cerema PAVIN fog & rain platform: From past and back to the future," presented at the ITS World Congress, 2022, `https://hal.archives-ouvertes.fr/hal-03844483`.

References II

[10] Teledyne Vision Solutions, "Genie Nano 1-GigE — G3-GC11-C1630," https://www.teledynevisionsolutions.com/products/genie-nano-1-gige/?model=G3-GC11-C1630, last accessed: June 30, 2025.

[11] Exosens, "BOBCAT/BOBCAT+," https://www.exosens.com/products/bobcat, last accessed: June 30, 2025.

[12] SWIR Vision Systems, "Acuros SWIR Cameras," https://www.swirvisionsystems.com/acuros-swir-camera/, last accessed: June 30, 2025.

[13] SWIR Vision Systems, "SWIR Lenses for Acuros Cameras," https://www.swirvisionsystems.com/swir-lenses/, last accessed: July 23, 2025.

[14] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. *Ultralytics YOLOv8*. 2023. Available at: https://github.com/ultralytics/ultralytics.

[15] Glenn Jocher and Jing Qiu. *Ultralytics YOLOv11*. 2024. Available at: https://github.com/ultralytics/ultralytics.

[16] Jianhe Ren, Shilong Liu, Ailing Zeng, Jing Lin, Kunchang Li, He Cao, Jiayu Chen, Xinyu Huang, Yukang Chen, Feng Yan, Zhaoyang Zeng, Hao Zhang, Feng Li, Jie Yang, Hongyang Li, Qing Jiang, and Lei Zhang. *Grounded SAM: Assembling Open-World Models for Diverse Visual Tasks*. 2024. Available at: https://github.com/IDEA-Research/Grounded-Segment-Anything.

[17] MMSegmentation Contributors. *MMSegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark*. 2020. Available at: https://github.com/open-mmlab/mmsegmentation. Last accessed: 2025-06-30.

[18] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[19] Y. Jin, M. Kovac, Y. Nalcakan, H. Ju, H. Song, S. Yeo, and S. Kim, "RASMD: RGB and SWIR Multispectral Driving Dataset for Robust Perception in Adverse Conditions," 2025.

Additional Material - Preprocessing Scheme

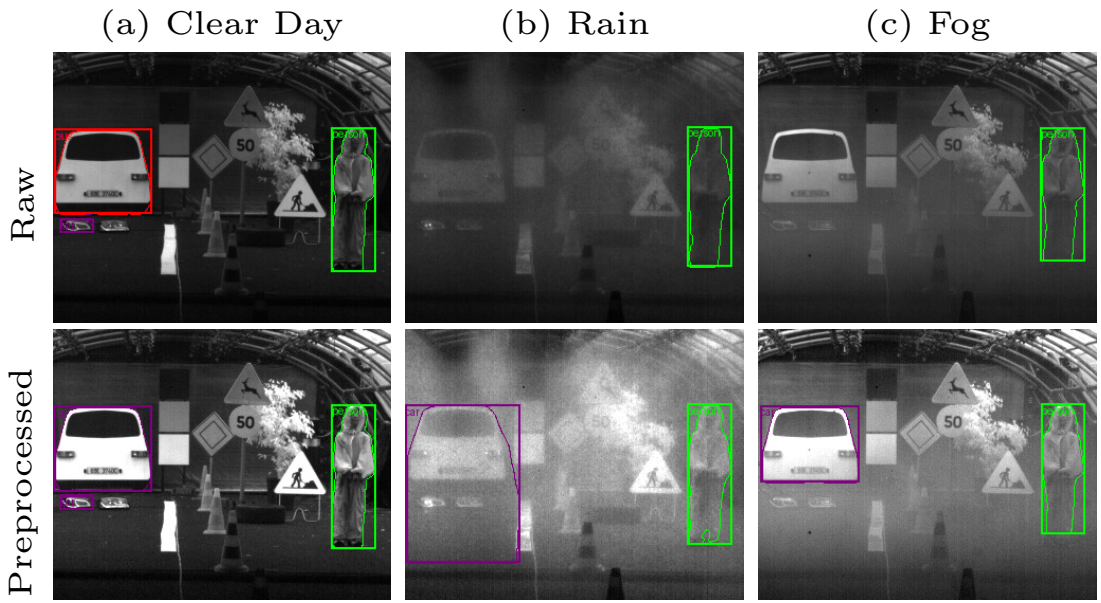

Figure : Visual comparison of preprocessed images against raw images acquired through the Xenics and detected through YOLOv8x-seg.

The following scheme was used to preprocess only SWIR images:

1. *Thresholding:* Pixel intensity values were clipped at a 99% confidence interval threshold to remove outliers that may distort overall contrast (e.g., saturated pixels).
2. *Stretching:* The histograms were stretched throughout the range to improve visual contrast.

Additional Material - Quantitative Evaluation

Sequence	Camera	Algorithm	Preprocessed	TP ↑	FP ↓	FN ↓	Precision ↑	Recall ↑	F1 Score ↑
Fog (87 images)	SVS	Grounded SAM	Yes	120	111	54	0.5195	**0.6897**	0.5926
			No	117	98	57	0.5442	0.6724	0.6015
		YOLOv8	Yes	101	12	73	0.8938	0.5805	0.7038
			No	100	25	74	0.8000	0.5747	0.6689
	Visible	Grounded SAM	No	99	4	75	0.9612	0.5690	*0.7148*
		YOLOv8	No	103	3	71	**0.9717**	0.5920	0.7357
	Xenics	Grounded SAM	Yes	101	80	73	0.5580	0.5805	0.5690
			No	75	60	99	0.5556	0.4310	0.4854
		YOLOv8	Yes	110	6	64	0.9483	*0.6322*	**0.7586**
			No	78	4	96	*0.9512*	0.4483	0.6094
Rain (125 images)	SVS	Grounded SAM	Yes	245	249	5	0.4960	0.9800	0.6586
			No	214	219	36	0.4942	0.8560	0.6266
		YOLOv8	Yes	244	28	6	0.8971	*0.9760*	*0.9349*
			No	225	33	25	0.8721	0.9000	0.8858
	Visible	Grounded SAM	No	250	27	0	*0.9025*	**1.0000**	0.9488
		YOLOv8	No	250	7	0	**0.9728**	**1.0000**	**0.9862**
	Xenics	Grounded SAM	Yes	186	211	64	0.4685	0.7440	0.5750
			No	159	123	91	0.5638	0.6360	0.5977
		YOLOv8	Yes	192	37	58	0.8384	0.7680	0.8017
			No	153	10	97	0.9384	0.6120	0.7409

Table : Quantitative evaluation across fog, rain and clear day sequences for preprocessed and non-preprocessed frames. Recall and F1 score increase significantly for SWIR images upon preprocessing in most of the cases.

Additional Material - Image formats fed to models

Camera	Channels	Image dimensions	Pixel Range
Xenics Bobcat 320 (SWIR)	1	320×256	$[0, 255]$
SVS Acuros CQD 1280 (SWIR)	1	350×240	$[0, 255]$
Dalsa Genie Nano 1630 (Vis)	3	370×250	$[0, 255]$

Table : Image formats of SWIR and visible images after preprocessing to be taken as input by the algorithms. Since the resolution is different for each sensor, we cropped images around the same part of the scene to allow a fair comparison. Note also that since the algorithms used 8-bits images, the SWIR images has been converted to that same format.

Additional Material - Implementation details of Deep Learning Algorithms

This study explores four popular on-the-shelf deep learning detection and segmentation models pre-trained on RGB images without fine-tuning:

- **YOLOv8x-seg** and **YOLOv11x-seg** pretrained on COCO dataset were used with default parameters (i.e., a confidence threshold of 0.25).
- **GroundedSAM** was executed with the SwinT backbone for GroundingDINO and ViT-H backbone for SAM, while the box threshold used was 0.35 and the text threshold was set to 0.25. The same COCO labels detected in our dataset by YOLO have been provided to GroundingDINO.
- **MMSegmentation** pretrained on Cityscapes was also used with the default parameters.

All models were executed on NVIDIA GeForce GTX 1080 GPU, CUDA 11.7, with 8 GB of VRAM.