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

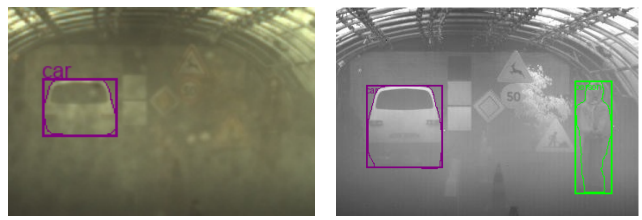
romehraa@gmail.com {firstname}.{lastname}@uca.fr {firstname}.{lastname}@cerema.fr

Abstract

In the context of road perception for autonomous vehicles, short-wave infrared (SWIR) has opened up new perspectives beyond the visible spectrum, which is prone to performance degradation in harsh weather conditions like fog, rain and dust. This paper aims to analyze the feasibility of using SWIR images to enhance object detection and segmentation in such weather conditions. In our experiments, we used data obtained from three different cameras – including two SWIR technologies and a conventional visible camera – in different weather conditions. The conditions include a clear day for reference, rain at different rainfall rates, and fog at different visibility ranges. We explored the performance of several deep learning algorithms, originally trained on images from visible domain applied directly to SWIR images. Quantitative and qualitative analyses for detection and segmentation were conducted. When applied to SWIR modalities, the algorithms prove to perform comparatively to visible in the reference case and to improve detection and segmentation in harsh weather cases.

1. Introduction

Short-Wave Infrared (SWIR), defined typically in the spectral range of 1.0 to 2.5 µm has opened new perspectives in context of road perception for autonomous vehicles. Currently used algorithms for autonomous driving mainly rely on visible spectrum. However, visible cameras are prone to performance degradation in adverse weather conditions such as fog and rain [3, 21]. The integration of modalities that capture information beyond visible light, such as long-wave infrared (LWIR) (8 – 12 µm) and near-infrared (NIR) (0.7 – 1.0 µm), has been explored as a way to close this gap [2, 10, 14]. LWIR or thermal cameras are used for low-light imaging for military and surveillance applications [7, 17, 23]. However, such cameras are usually of lower resolution, bulkier, more costly, and the appearance of objects varies according to ambient conditions. The SWIR spectrum offers distinctive advantages over these modalities in object detection under harsh weather conditions, as shown in [1, 5, 9, 22, 24, 31]. Figure 1 illustrates a visual comparison of the same scene in foggy conditions, captured using a visible and SWIR camera. When processed through the same perception algorithm, the SWIR image provides more correct detections than the visible image.



(a) Visible spectrum image. (b) SWIR spectrum image.

Figure 1. Comparison of detections in harsh weather conditions, here fog, for (a) visible and (b) SWIR images captured at the Cerema "Rain and Fog" platform [18]. In that case, SWIR allows to detect the person and the car where the same algorithm fails in the visible case. *Images have been cropped for visualization.*

Historically, the adoption of SWIR was limited to military use because of the high cost of sensors. However, recent advancements, particularly the development of quantum dot-based sensors leveraging CMOS technology has led to a significant reduction in cost, enabling broader use in industrial and civilian applications. Furthermore, quantum dot sensors provide improved sensitivity and resolution compared to traditional alternatives [8], making them suitable for perception tasks.

Contribution. To the best of our knowledge, this paper presents the first comparative study evaluating the performance of two SWIR technologies – InGaAs [6] and Colloidal Quantum Dot (CQD) [28] – alongside a visible-light camera with a specific focus on harsh weather conditions in controlled environments. The evaluation is conducted using detection and segmentation algorithms pre-trained on the visible (RGB) images, with focus on applications in autonomous driving. The results show that when applied to SWIR modalities (in particular the CQD modality), the algorithms prove to perform comparatively to visible in the reference case, and to improve detection and segmentation in harsh weather cases.

The remainder of the paper has been organized as follows: section 2 describes the experimental setup along with the pre-processing and evaluation metrics that we used. In section 3, we present the results comparing various sensing modalities, as well as different detection and segmentation algorithms under low visibility conditions. Limitations are also discussed. The section 4 presents a conclusion and directions for further research.

2. Methodology

In this study, we propose to evaluate on-the-shelf detection and segmentation algorithms on SWIR images acquired in different weather conditions to determine if the SWIR modality would help in these conditions. Quantitative and qualitative evaluations are performed using several sensor modalities, during various weather conditions in a controlled environment, and for several algorithms. This section contains information on the dataset and sensors used, the pre-processing step introduced to improve the detections, the models used for detection and segmentation, as well as the evaluation metrics used to compare the sensors and the models.

2.1. Dataset

For our analysis, we took inspiration from the experimental setup presented in [26]. We acquired data in controlled conditions at the Cerema "Rain and Fog" platform [18]. Our dataset includes three weather conditions (clear day, rain and fog) with three different cameras. The same static scene is used across the entire data set, whereas the dynamic is given by weather changes. *Dataset is available upon request.*

2.1.1. Cameras

The dataset provides three sensing modalities: two SWIR cameras – an InGaAs based [6] and CQD based [28] – and one visible (RGB) camera [30]. Cameras will be referred to as Xenics, SVS and visible, respectively, in the rest of the paper. Table 1 details the spectral ranges of the sensors. Note that even though the spectral range of the SVS is broader, the camera is mounted with a SWIR-coated objective lens [29], using then an effective range of 800 nm – 1700 nm. Both SWIR cameras produce a single-channel grayscale image, in contrast to the three-channel RGB images produced by the visible camera. The three sensors are mounted such that they see the same part of the scene. Note that the resolution is different for each sensor, and therefore cropped images aligned around the same part of the scene are used to allow a fair comparison (see subsection 2.3 and Table 3).

Camera	Type	Spectral Range
Dalsa Genie Nano 1630 [30]	Visible Camera	380 nm–700 nm
Xenics Bobcat 320 [6]	InGaAs-based SWIR	900 nm–1700 nm
SVS Acuros CQD 1280 [28]	Quantum Dot-based SWIR	400 nm–1700 nm*
SWIR-11 25 mm lens M42 [29]	SWIR-coated lens	800 nm–1700 nm

Table 1. Information on the sensors used for analysis. Two SWIR sensors as well as one visible has been used for comparison. **The SVS camera is mounted with a SWIR-coated objective lens, using then an effective range of* 800 nm–1700 nm.

2.1.2. Weather conditions

We divided the dataset into three sequences: clear day (26 images), foggy day (87 images), and rainy day (125 images). The clear day sequence shows good visibility conditions and acts as our reference (1 image every second). The foggy day sequence progressively increases the amount of fog, reducing the visibility conditions, and then dissipates progressively (1 image every 5 seconds). The rainy day sequence progressively increases the rainfall rate (1 image every 5 seconds). The first half of the sequence uses small size droplets, and thus is considered as light rain. In the second half (at around 375 seconds), the size of the droplets increases, and is thus considered heavy rain. Figure 3 gives an overview of the SWIR images in all three weather conditions. Note that as we focus the study on weather conditions, up to our knowledge, the dataset used is the only one providing controlled meteorological metrics (visibility (m), rainfall rate (mm/h), and droplet size (small or large).

2.2. Detection and Segmentation 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** [13] and **YOLOv11x-seg** [12] pretrained on COCO dataset [19] were used with default parameters (i.e., a confidence threshold of 0.25).
- **GroundedSAM** [25] was executed with the SwinT backbone for GroundingDINO [20] and ViT-H backbone for SAM [16], 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** [4] pretrained on Cityscapes was also used with the default parameters.

Algorithms	Tasks	Avg Time/Image
YOLOv8x-seg [13]	Detect+Segment (COCO labels)	0.07 s to 0.1 s
YOLOv11x-seg [12]	Detect+Segment (COCO labels)	0.07 s to 0.1 s
Grounded SAM [25]	Detect+Segment (open set)	2 s to 4 s
MMSegmentation [4]	Detect+Segment (Cityscrape)	0.4 s to 0.8 s

Table 2. List of detection and segmentation algorithms used to compare SWIR and visible images along with their precessing time per image. YOLO proves to be the fastest of the algorithms.

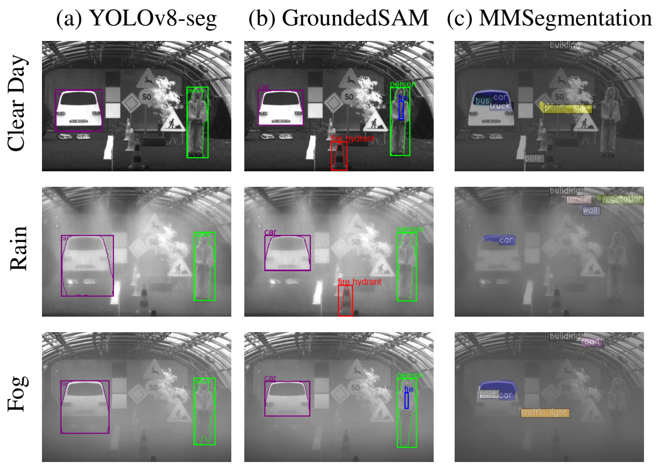


Figure 2. Outputs of different detections and segmentation algorithms on SWIR images acquired through the SVS [28]. (a) YOLOv8x-seg and (b) GroundedSAM give comparative performances, (c) MMSegmentation struggles to adapt to SWIR images, hence we discarded MMSegmentation from further analysis.

All models were executed on NVIDIA GeForce GTX 1080 GPU, CUDA 11.7, with 8 GB of VRAM. The models are summarized in Table 2 with their execution time. Note that the YOLO algorithms prove to be the fastest of the evaluated algorithms.

We focused only on the detection of one person and one car for this analysis, which are also the only two COCO labels [19] present in the frames. For ground truth, we therefore independently annotated each sensor for each weather condition using RoboFlow [27]. Figure 2 shows SWIR images annotated with the output of the selected models. We observed that while YOLOv8x-seg and GroundedSAM give comparative performance on SWIR images, MMSegmentation fails to adapt and results in inconsistent detections. Henceforth, we chose to discard MMSegmentation from further study. Moreover, YOLOv11x-seg was used solely for the comparison of the different models (see subsubsection 3.2.2).

2.3. Preprocessing

We introduced an initial preprocessing step that is carried out on images taken by both SWIR cameras. It allows

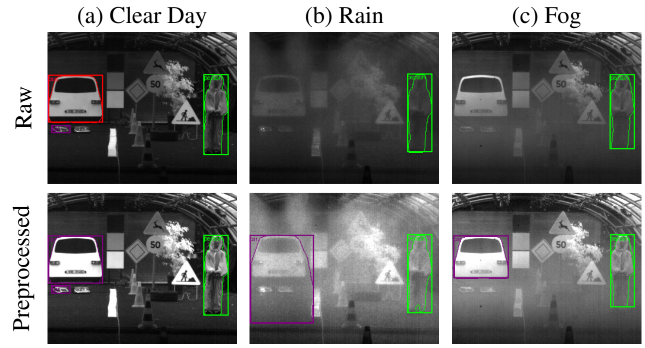


Figure 3. Visual comparison of preprocessed images against raw images acquired through the Xenics [6] and detected through YOLOv8x-seg [13]. (a) Clear Day: without preprocessing a truck is detected instead of a car; (b) Rain and (c) Fog: the car is not detected in raw images but effectively detected after preprocessing.

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 3. 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.

us to balance the contrast and to improve detections when run through the visible pre-trained models. Note that the images taken by the visible camera were used without any preprocessing. The following scheme was used to preprocess 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. In addition, images taken from the SVS camera and the visible camera were cropped and aligned to focus on the region of interest and match it with the focus of the Xenics. Table 3 shows the image formats of the three sensor outputs after preprocessing which are then fed into the models.

Figure 3 shows the visual improvements after the use of the preprocessing method which improves the detections through the YOLOv8x-seg model. Our preprocessing will be quantitatively validated, across all sensor modalities, in section 3.

2.4. Evaluation Metrics

The evaluation was carried out along two axes: 1) quantitative analyses for detection of objects across all images, per weather condition and model; and 2) qualitative analyses for detection and segmentation results in individual frames. To compute the metrics, we followed the same process as in [15] where the intersection-over-union (IoU) of each ground truth mask is calculated against each predicted mask with the same class label. The predicted mask is then classified as:

- **True Positive (TP):** If the predicted mask matches the ground truth mask with an IoU ≥ 0.3.
- **False Positive (FP):** If the prediction does not match any of the ground truth masks.
- **False Negative (FN):** If the ground truth segment does not match the prediction.

Note that in case of double detections for the same object in the ground truth, the prediction matched with higher IoU is considered as TP and the other one is considered as FP. We henceforth compute per-frame TP, FP and FN.

2.4.1. Quantitative metrics for detection

Precision, Recall and F1-score have been computed over all frames from each camera, for each weather condition, and for each algorithms. The metrics have been defined as:

$$\text{Precision} = \frac{\sum |\text{TP}|}{\sum |\text{TP}| + \sum |\text{FP}|}, \tag{1}$$

$$\text{Recall} = \frac{\sum |\text{TP}|}{\sum |\text{TP}| + \sum |\text{FN}|}, \tag{2}$$

$$\begin{aligned} \text{F1-score} &= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \\ &= \frac{\sum |\text{TP}|}{\sum |\text{TP}| + \frac{1}{2} \sum |\text{FP}| + \frac{1}{2} \sum |\text{FN}|}. \end{aligned} \tag{3}$$

For this comparison, the metrics have been aggregated over each sequence so that it provides a broader assessment over variable visibility / rainfall rate range. Precision measures the accuracy of positive predictions. Recall measures the model's ability to find all the positive instances. F1-score balances the two metrics into a single number.

2.4.2. Qualitative metrics for segmentation

Since the objective is to obtain a segmentation per instance of the scene, we based our evaluation on the Panoptic Quality (PQ) and Segmentation Quality (SQ) as defined in [15]. The comparison of segmentation quality has been done on a per frame basis. Panoptic Quality (PQ) is calculated for each frame from SQ and Recognition Quality (RQ) which are defined as:

$$\text{Segmentation Quality (SQ)} = \frac{\sum_{(p,g)\in \text{TP}} \text{IoU}(p,g)}{|\text{TP}|} \tag{4}$$

$$\text{Recognition Quality (RQ)} = \frac{|\text{TP}|}{|\text{TP}| + \frac{1}{2}|\text{FP}| + \frac{1}{2}|\text{FN}|} \tag{5}$$

where p denotes a predicted mask and g denotes its corresponding ground truth mask, with the pair (p, g) belonging to the set of true positives (TP). Finally, the PQ score is computed as

$$\text{Panoptic Quality (PQ)} = \text{SQ} \times \text{RQ}. \tag{6}$$

Note that RQ is similar in nature to F1-score (see Equation 3), and thus impacted by incorrect detections (FP), while SQ focuses solely on the quality of segmentation of the detected elements (TP). SQ is simply the average IoU of the TP, measuring then the performance of the segmentation task. PQ captures performance for all classes in an interpretable and unified manner.

3. Results and Discussions

The analyses of experimental results have been organized as follows: Quantitative analysis that validates our preprocessing step and globally compares models / cameras on object detection tasks, and qualitative analysis that provides a frame-by-frame comparison of preprocessed outputs from different sensing modalities and a comparison of different deep learning models based on the quality of detection and segmentation.

3.1. Quantitative Analysis

Quantitative metrics are reported in Table 4, for each sequence, each camera, for two algorithms, with and without our preprocessing step.

We observe that with YOLOv8x-seg, the recall and F1-score always increase when the data is preprocessed. The increase of FP and the subsequent decrease of precision is related to the double detections introduced likely due to another set of headlights near the car in the frames. These double detections could drastically be reduced by implementing a Non-Maximum Suppression (NMS) post-processing step. A similar trend is observed with GroundedSAM where the precision and F1-score either increase significantly or remain within a comparable range. However, GroundedSAM detects a 'fire extinguisher' and 'tie' (as also seen in Figure 2) in most of the high visibility frames which makes the FP count high and thereby the precision lower in some cases. This also results in a lower range of overall F1-scores of GroundedSAM compared to those of YOLOv8x-seg. Overall, our preprocessing pipeline increases the number of TP and improves the F1-scores, validating its efficiency in model performance.

Through all three sequences, we observe a higher F1-score on visible image data compared to SWIR images. It can be explained in part by the fact that the models are pretrained on visible images, which favors then the visible modality. Overall, the scores are similar given the modality, showing the applicability of the algorithms on

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
Clear Day (26 images)	SVS	Grounded SAM	Yes	52	60	0	0.4643	**1.0000**	0.6341
			No	52	60	0	0.4643	**1.0000**	0.6341
		YOLOv8	Yes	52	4	0	*0.9286*	**1.0000**	0.9630
			No	49	10	3	0.8305	0.9423	0.8829
	Visible	Grounded SAM	No	47	2	5	0.9592	0.9038	*0.9307*
		YOLOv8	No	52	0	0	**1.0000**	**1.0000**	**1.0000**
	Xenics	Grounded SAM	Yes	49	25	3	0.6622	0.9423	0.7778
			No	47	25	5	0.6528	0.9038	0.7581
		YOLOv8	Yes	51	15	1	0.7727	0.9808	0.8644
			No	50	14	2	0.7812	*0.9615*	0.8621

Table 4. 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. In clear and rain cases, the modalities perform comparatively, whereas SWIR shows a slight improvement in the fog case. **Bold** font indicates the best results, underline font indicates the second best, and *italic* font indicates the third best per sequence.

SWIR. In relation to the two SWIR modalities, the SVS provides results superior to those of the Xenics on the three sequences. Even if GroundedSAM provides globally more TP than YOLOv8, it also gives significantly more false positives (i.e., other objects detected). Note also that the metrics are aggregated over the whole sequence, which does not clearly highlight the differences during the harsher weather conditions of the sequences. However, we can see that in case of fog, the SWIR cameras show a higher recall than visible, highlighting the potential usefulness of this spectrum in this harsh condition.

3.2. Qualitative Analyses

Those qualitative analyses have been carried out for two purposes. First, a comparison of the sensing modalities where SQ is plotted as a function of time (with additional axis to show varying visibility / rainfall rate) for the same model and weather condition. Note that this metric does not take FP and FN into account, and we focus solely on the quality of segmentation of the detected elements for this analysis. Second, a comparison of models where PQ for each model is plotted as a function of meteorological visibility in fog conditions and as a function of rainfall rate in rainy conditions using data from the same camera. Since we focus on the comparison of the algorithms, we used the PQ score as we want to better highlight the overall capacity

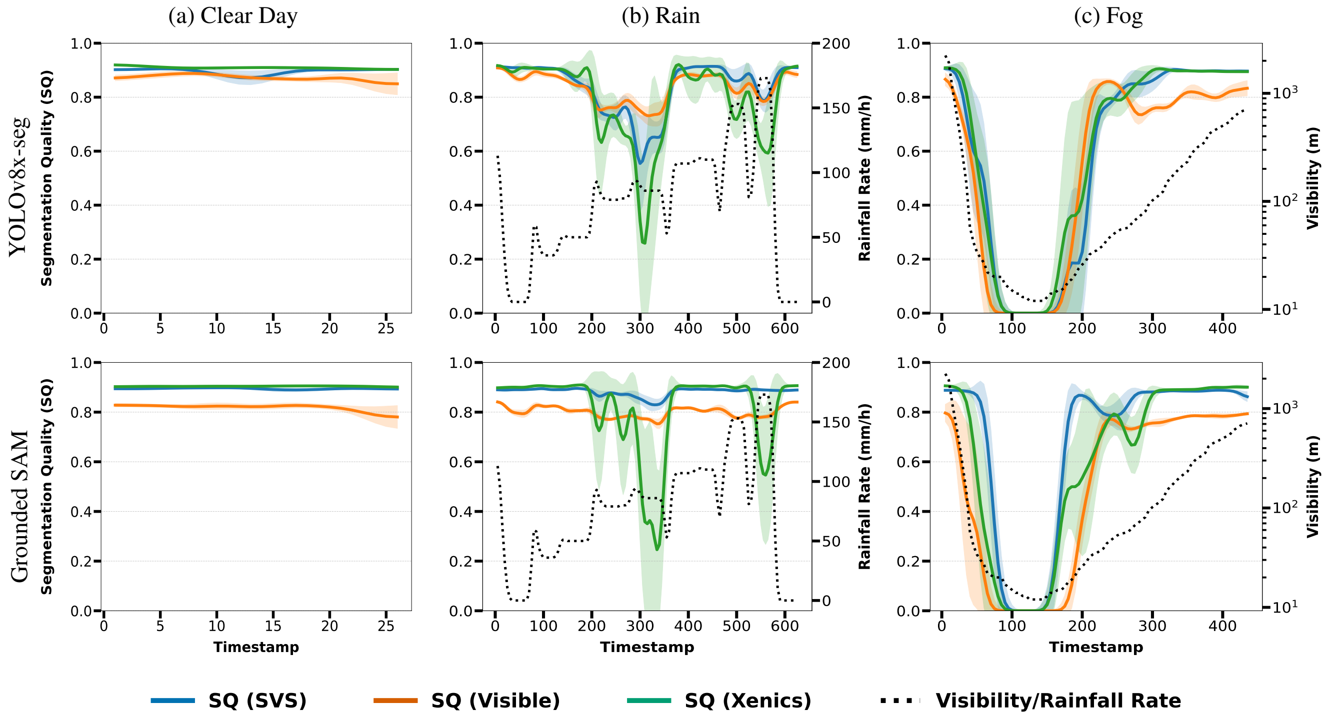


Figure 4. Qualitative analysis of the Segmentation Quality (SQ) as function of the time (with additional axis for visibility / rainfall rate) for comparison of the three sensing modalities, in different weather conditions. (a) Clear day for reference performance in high visibility conditions, (b) in rainy conditions, SWIR and visible cameras show competitive performances while SVS achieves a higher SQ when rainfall rate increases, (c) in foggy conditions, SWIR cameras stay functional longer in low visibility, marking superior performances.

of each algorithm for different weather scenarios.

Only preprocessed frames have been used for SWIR images. PQ graphs have been interpolated with LOWESS (with a span of 0.5) since several values can be acquired for the same meteorological visibility. SQ graphs have been smoothed using a Gaussian filter ($\sigma = 2$) to enhance the visual clarity of the trends.

3.2.1. Comparison of Sensors

Figure 4 presents the graphs for the comparison of three cameras, under each weather condition, and for two algorithms. Figure 4 (a) shows performance under clear weather condition with high visibility throughout the image sequence for reference, where all three cameras show comparable SQ in the same range for both models. Figure 4 (b) introduces rain in the scene, which is parameterized by rainfall rate (mm/h). When processed with YOLOv8x-seg, the SVS and visible camera give competitive results, with both exhibiting a drop in SQ during spikes in rainfall rate. However, with GroundedSAM, SVS demonstrates a superior performance maintaining a higher score even under increased rainfall spikes. Xenics struggles to adapt to sudden increases in rainfall, resulting in the lowest SQ in severe condition. To simulate heavy rain, at around 375 seconds into the sequence, the size of the droplets increases while the density of droplets decreases. This in turn results in increased visibility of the scene, which also explains the increased SQ across all three cameras despite the increasing rainfall rate after this point. Figure 4 (c) compares modalities in foggy conditions, where the meteorological visibility (m) has been considered as a parameter. With both GroundedSAM and YOLOv8x-seg, as the visibility decreases, SWIR cameras maintain a higher SQ score and remain functional for a longer duration compared to the visible camera. Under high-visibility conditions, the performance of all the cameras remains in a comparable range as also seen in the clear-day sequence.

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

3.2.2. Comparison of Models

Figure 5 presents the comparison of deep learning models using the PQ score for the detection and segmentation task of a person and a car.

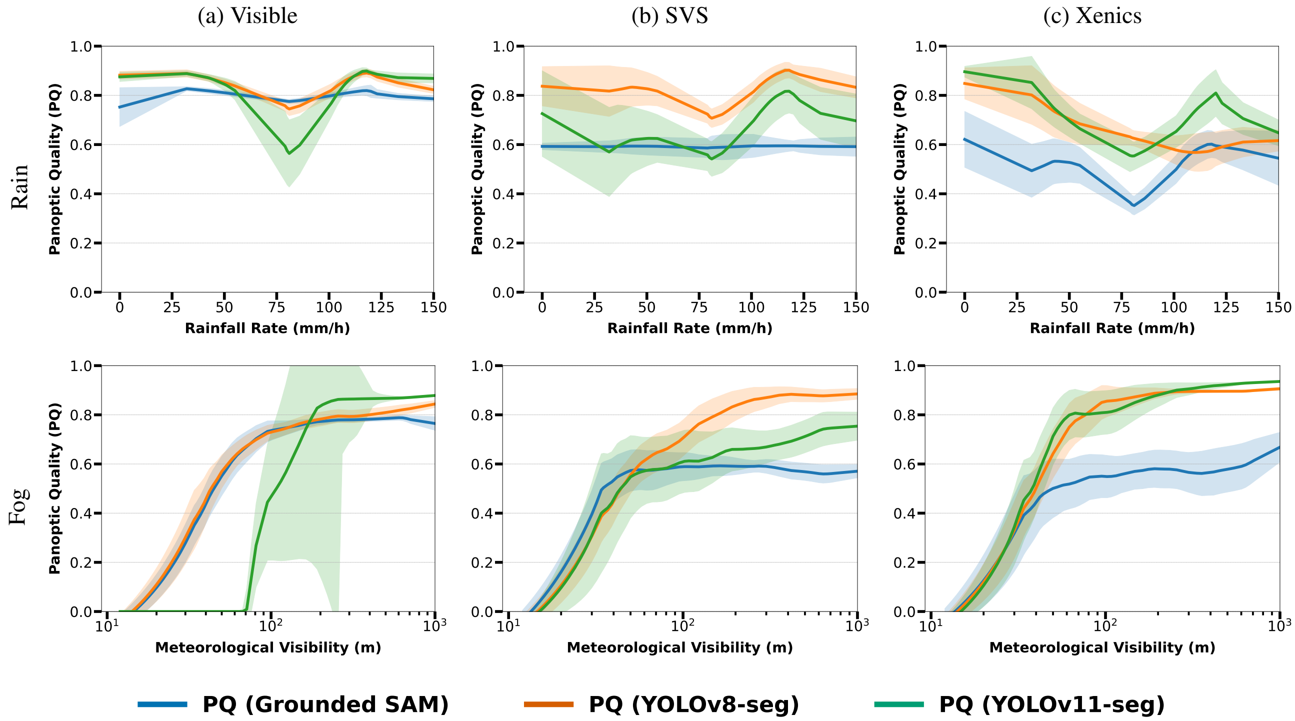


Figure 5. Qualitative analysis of the Panoptic Quality (PQ) as function of the visibility / rainfall rate for comparison of detection and segmentation models with (a) Visible, (b) SVS, and (c) Xenics cameras. YOLOv8x-seg and GroundedSAM give comparative performance in low-visibility conditions, while GroundedSAM results in suboptimal PQ score in high-visibility conditions. YOLOv11x-seg shows unstable behavior across different evaluations. YOLOv8x-seg can be a suitable choice to be applied to SWIR modality, which offers a balance between detection accuracy and inference speed.

In rainy conditions, all models show competitive performance with visible images. However, with both SWIR cameras, GroundedSAM consistently yields the lowest PQ across the sequence, which is primarily due to the number of FPs detected. In addition, a sharp edge is observed with all three cameras at a rainfall rate of 80 mm/h. This corresponds to the transition from light to heavy rain, associated with a change in droplet size. Beyond this, PQ improves as the visibility of the scene increases. Xenics demonstrates a higher sensitivity to rain, as evidenced by the lowest PQ score of all models compared to other sensing modalities. Note that we also computed the SQ score, and in the rain sequence, the scores are similar for all modalities, for all algorithms, and do not vary significantly as visibility decreases. In this condition, YOLOv8x-seg applied to visible and SVS yield the best performances.

In fog conditions, YOLOv11x-seg exhibits performance degradation when applied to visible camera images, with PQ dropping to zero at a relatively high visibility compared to other models. In contrast, GroundedSAM and YOLOv8x-seg remain competitive throughout. When applied on both the SWIR camera images, all models remain competitive during low visibility conditions. However, GroundedSAM yields low PQ is high-visibility conditions, again due to an increased number of FPs detected. Both YOLO variants show comparative performance in both the SVS and Xenics modalities.

Overall, considering the SQ and PQ scores of both algorithms, and the computational time reported in Table 2, YOLOv8x-seg applied to SWIR imaging with the SVS can be considered a good solution for perception in harsh weather conditions.

3.2.3. Limitations of the comparisons

Our comparisons present several limitations. All models used for this analysis have been pre-trained on visible (RGB) dataset, and may lack domain-specific adaptation to SWIR spectrum. This may result in suboptimal performance on SWIR data, particularly under harsh weather conditions. Moreover, quantitative metrics are computed across the entire sequences, which does not directly emphasize the potential advantages of the SWIR spectrum under severe weather cases. In contrast, we can see the influence of the meteorological conditions through the segmentation quality (SQ) score graphs. In addition, modalities are not perfectly registered (slightly different point of view, not the same focal length, different field of view, etc.) which might influence the evaluation results. Furthermore, no post-processing has been applied on detections which could have significantly reduced the number of false positives (FPs). Similarly, no tracking is considered across the sequences, which could have improved the number of correct detections.

4. Conclusion

This paper provided an analysis of SWIR and visible imaging systems for outdoor perception tasks, with potential applications in driving assistance systems (ADAS). Based on our results, we validated our preprocessing step for SWIR images through quantitative and qualitative analyses. In particular, our study focuses on the detection of people and cars in urban environments under weather conditions that included rain and fog using the SWIR spectrum. The experiments comparing two SWIR cameras with a visible camera, demonstrate that the SWIR cameras remain functional for longer periods during dense fog conditions and remain competitive to visible during light and heavy rain conditions. They are therefore suitable for operation in low-visibility conditions. With regard to comparison between the two SWIR modalities, the SVS is generally more stable than the Xenics with slightly better performances.

We also give a comparison of various detection and segmentation models pre-trained on visible images while being applied on SWIR images. GroundedSAM and YOLOv8x-seg exhibit comparable performance in terms of PQ score in low-visibility conditions, and GroundedSAM performs suboptimally in high-visibility conditions primarily due to a number of FP detections. YOLOv11x-seg demonstrates inconsistent behavior across different plots, indicating instability in its performance. Taking into account SQ, PQ and computational time, YOLOv8x-seg applied to SWIR imaging with the SVS can be considered a good solution for perception in harsh weather conditions.

In conclusion, given our results, when applied to SWIR modalities, the algorithms prove to perform comparatively to visible in the reference case, and to improve detection and segmentation in harsh weather cases. However, further work need to be done to unlock the full potential of the SWIR modality.

Future works. In parallel of our investigations, a new dataset named RASMD containing SWIR images has been published [11]. As our current experiments focus on a static scene to evaluate algorithms in different controlled weather conditions, we plan to evaluate the case of dynamic scenes with the help of the RASMD dataset. Future works will include fine-tuning the models on SWIR based on methods such as domain adaptation or knowledge distillation to improve performance in adverse weather. A potential direction can involve the fusion of both modalities to take advantage of both, such as semantics for visible and resilience to harsh conditions for SWIR.

Acknowledgments

This work was supported by the International Research Center "Innovation Transportation and Production Systems" of the I-SITE CAP 20-25. We also thank the Cerema team for their help during the acquisitions. The dataset used in this paper is available upon request.

References

[1] Massimo Bertozzi, Rean Isabella Fedriga, Alina Miron, and Jean-Luc Reverchon. Pedestrian detection in poor visibility conditions: Would swir help? In *Image Analysis and Processing – ICIAP 2013*, pages 229–238, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. 1

[2] Shubhadeep Bhowmick, Somenath Kuiry, Alaka Das, Nibaran Das, and Mita Nasipuri. Deep learning-based outdoor object detection using visible and near-infrared spectrum. *Multimedia Tools and Applications*, 81, 2022. 1

[3] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather, 2020. 1

[4] MMSegmentation Contributors. MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. https://github.com/open-mmlab/mmsegmentation, 2020. Last accessed: 2025-06-30. 2, 3

[5] Ronald G Driggers, Van Hodgkin, and Richard Vollmerhausen. What good is SWIR? passive day comparison of VIS, NIR, and SWIR. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIV*, pages 187–201. SPIE, 2013. 1

[6] Exosens. BOBCAT/BOBCAT+. https://www.exosens.com/products/bobcat. Last accessed: 2025-06-30. 2, 3

[7] Rikke Gade and Thomas B Moeslund. Thermal cameras and applications: a survey. *Machine vision and applications*, 25: 245–262, 2014. 1

[8] Christopher Gregory, Allan Hilton, Katherine Violette, and Ethan Klem. 66-3: Invited paper: Colloidal quantum dot photodetectors for large format nir, swir, and eswir imaging arrays. *SID Symposium Digest of Technical Papers*, 52: 982–986, 2021. 1

[9] MP Hansen and DS Malchow. Overview of SWIR detectors. *Cameras, and Applications*, page 69390I. 1

[10] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and Inso Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. 2015. 1

[11] Youngwan Jin, Michal Kovac, Yagiz Nalcakan, Hyeongjin Ju, Hanbin Song, Sanghyeop Yeo, and Shiho Kim. RASMD:

RGB and SWIR multispectral driving dataset for robust perception in adverse conditions, 2025. 8

[12] Glenn Jocher and Jing Qiu. Ultralytics yolo11, 2024. 2, 3

[13] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8, 2023. 2, 3

[14] Ayatollah Karimzadeh. Refractive dual band infrared imager optical design. *International Journal of Optics and Photonics*, 11:133–138, 2017. 1

[15] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic segmentation, 2019. 4

[16] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. *arXiv:2304.02643*, 2023. 2

[17] Mate Krišto, Marina Ivasic-Kos, and Miran Pobar. Thermal object detection in difficult weather conditions using yolo. *IEEE access*, 8:125459–125476, 2020. 1

[18] Sébastien Liandrat, Pierre Duthon, Frédéric Bernardin, Amine Ben Daoued, and Jean-Luc Bicard. A review of cerema PAVIN fog & rain platform: From past and back to the future. Presented at the ITS World Congress, 2022. hal-03844483. 1, 2

[19] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft COCO: Common objects in context, 2015. 2, 3

[20] Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Chunyuan Li, Jianwei Yang, Hang Su, Jun Zhu, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. *arXiv preprint arXiv:2303.05499*, 2023. 2

[21] Abdul Sajeed Mohammed, Ali Amamou, Follivi Kloutse Ayevide, Sousso Kelouwani, Kodjo Agbossou, and Nadjet Zioui. The perception system of intelligent ground vehicles in all weather conditions: A systematic literature review. *Sensors*, 20(22), 2020. 1

[22] Miloš S. Pavlović, Petar D. Milanović, Miloš S. Stanković, Dragana B. Perić, Ilija V. Popadić, and Miroslav V. Perić. Deep learning based swir object detection in long-range surveillance systems: An automated cross-spectral approach. *Sensors*, 22(7), 2022. 1

[23] Dragana Perić, Branko Livada, Miroslav Perić, and Saša Vujić. Thermal imager range: Predictions, expectations, and reality. *Sensors*, 19(15):3313, 2019. 1

[24] 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 22nd International Forum on Advanced Microsystems for Automotive Applications (AMAA'18)*, pages 3–15, Berlin, Germany, 2018. http://perso.lcpc.fr/tarel.jean-philippe/publis/amaa18.html. 1

[25] Tianhe 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. 2, 3

[26] Alexandre Riffard, Mathieu Labussière, Pierre Duthon, and Romuald Aufrère. Exploitation d'un capteur proche infrarouge (SWIR) pour la perception des robots mobiles en conditions météorologiques difficiles. In *Reconnaissance des Formes, Image, Apprentissage et Perception (RFIAP'24)*, Lille, France, 2024. 2

[27] Roboflow. Roboflow - train custom computer vision models. https://roboflow.com. Last accessed: 2025-04-05. 3

[28] SWIR Vision Systems. SWIR Lenses for Acuros Cameras. https://www.swirvisionsystems.com/swir-lenses/, . Last accessed: 2025-07-23. 2, 3

[29] SWIR Vision Systems. Acuros SWIR Cameras — SWIR Vision Systems. https://www.swirvisionsystems.com/acuros-swir-camera/, . Last accessed: 2025-06-30. 2

[30] 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: 2025-06-30. 2

[31] Yusuf Furkan Yucesoy and Cagri Sahin. Object detection in infrared images with different spectra. In *2024 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pages 1–6, 2024. 1