
Analysis and Use of Learning-Based Detection and Segmentation Algorithms on SWIR Images

Project Report

Submitted by

Rohan Mehra

BS Student

Department of Electrical Engineering and Computer Science
Indian Institute of Science Education and Research Bhopal

E-mail: rohan21@iiserb.ac.in

Under the guidance of

Dr. Mathieu Labussiere

Associate Professor



Institut Pascal
Université Clermont Auvergne
Clermont-Ferrand, France

May-July 2024

Abstract

This report presents a proof of concept for leveraging Short-Wave Infrared (SWIR) images to enhance object detection and segmentation in harsh weather conditions. The study explores the performance of deep learning algorithms, originally trained on RGB images, when applied to SWIR imagery, and compares the results with those obtained using visible-light images. Four imaging modalities were utilized: two infrared cameras (Xenics and SVS), a visible-light camera (Dalsa Genie Nano 1630), and a fused modality created through TarDAL fusion of the three individual sources.

To ensure consistency and enhance image quality across modalities, pre-processing techniques, including normalization, thresholding, were applied to all acquired images. Advanced detection and segmentation models (Grounded SAM, YOLOv8-seg, and YOLOv11-seg) were employed to evaluate the comparative effectiveness of these modalities for improved detection accuracy under challenging environmental conditions.

Acknowledgment

With immense gratitude and appreciation, I would like to acknowledge the incredible support and inspiration I received throughout the internship.

I extend my deepest thanks to my supervisor, Dr. Mathieu Labussière, whose unwavering guidance, encouragement, and profound expertise have been a constant source of motivation. His mentorship has not only shaped this work but also enriched my learning experience in countless ways.

A special note of appreciation goes to Alexandre Riffard, whose insightful advice, collaborative spirit, and tireless patience have been invaluable. Working alongside him has been both a privilege and a joy, and his dedication to research has been truly inspiring.

I am equally grateful to the exceptional members of the ComSee team at Institut Pascal. Their warmth, camaraderie, and constant readiness to support have made this experience memorable and fulfilling. Being part of such an encouraging and dynamic group has been nothing short of a blessing.

To each of these individuals and to the enriching environment at the Institut Pascal Lab, I owe the success of this endeavor. Thank you for making this journey so meaningful and rewarding.

Contents

1	Introduction	6
1.1	Background and Motivation	6
1.2	Objectives	6
2	Methodology	8
2.1	Experimental Setup and Data Acquisition	8
2.2	Preprocessing	9
2.3	Algorithms Used	10
2.4	Evaluation Metrics	11
2.4.1	Quantitative Evaluation	11
2.4.2	Qualitative Evaluation	11
2.5	Procedure	12
3	Results and Analysis	13
3.1	Quantitative Evaluation	13
3.2	Qualitative Evaluation	13
3.2.1	Comparison of Modalities for the Same Algorithm	13
3.2.2	Comparison of Algorithms for the Same Modality	16
3.3	Analysis and Deductions	17
3.3.1	Quantitative Analysis	21
3.3.2	Qualitative Analysis	21
4	Conclusion and Future Work	24
4.1	Conclusion	24
4.2	Future Work	25

List of Figures

1.1	Images acquired through Visible and SWIR Cameras	6
2.1	Experimental Setup for data acquisition	8
2.2	L to R - SVS Acuros CQD, Dalsa Genie Nano, Xenics Bobcat	9
2.3	Top row: Images. Bottom row: Corresponding histograms.	10
2.4	Sample detections by Semantic Segmentation algorithms.	11
3.1	Comparison of Modalities for Foggy Day Sequence with preprocessed and non-preprocessed frames.	15
3.2	Comparison of Modalities for Rainy Day Sequence with preprocessed and non-preprocessed frames.	15
3.3	Comparison of Modalities for Clear Day Sequence with preprocessed and non-preprocessed frames.	16
3.4	Comparison of Modalities with Fused Images for Foggy Day Sequence with preprocessed frames.	16
3.5	Comparison of Modalities with Fused Images for Rainy Day Sequence with preprocessed frames.	17
3.6	Modalities Comparison vs Visibility (Foggy Day Sequence)	17
3.7	Modalities Comparison vs Visibility (Rainy Day Sequence)	17
3.8	Comparison of Algorithms for Foggy Day Sequence with preprocessed and non-preprocessed frames.	18
3.9	Comparison of Algorithms for Rainy Day Sequence with preprocessed and non-preprocessed frames.	18
3.10	Comparison of Algorithms (including Fused Images and YOLOv11-seg) for Foggy Day Sequence with preprocessed frames.	19
3.11	Comparison of Algorithms (including Fused Images and YOLOv11-seg) for Rainy Day Sequence with preprocessed frames.	19
3.12	Algorithms Comparison vs Visibility (Foggy Day Sequences)	20
3.13	Algorithms Comparison vs Visibility (Rainy Day Sequences)	20

List of Tables

2.1	Specifications of Cameras Used for Data Acquisition	9
2.2	List of Algorithms used	10
3.1	Quantitative Evaluation across Different Sequences	14

Chapter 1

Introduction

1.1 Background and Motivation



(a) Visible-light image

(b) SWIR Image

Figure 1.1: Images acquired through Visible and SWIR Cameras

Object detection and segmentation are critical tasks in computer vision, with applications spanning from autonomous vehicles to security. SWIR imaging, defined as light in the $0.9\text{--}1.7\ \mu\text{m}$ wavelength range, is valuable in these applications due to its ability to penetrate adverse weather conditions like fog and rain, as well as to enhance contrast in low-light environments. Traditional RGB sensors struggle in such environments, but SWIR sensors, such as InGaAs-based cameras, can help improve vision by capturing additional details. Figure 1.1 shows the image output by RGB and SWIR sensors respectively. This project investigates using SWIR data alongside RGB-trained models to enhance segmentation accuracy.

1.2 Objectives

The primary objectives of this internship were:

- To acquire data under simulated rain and fog conditions at Cerema, Clermont-Ferrand, utilizing setups that recreate these adverse weather conditions.

- To preprocess the frames for infrared and visible modality to ensure data consistency and quality.
- To run the registered sequences from four imaging modalities through deep learning-based detection and segmentation algorithms and validate the pre-processing.
- To conduct quantitative and qualitative evaluations by manually labeling ground truth for one pedestrian and one car in these sequences.

Chapter 2

Methodology

2.1 Experimental Setup and Data Acquisition

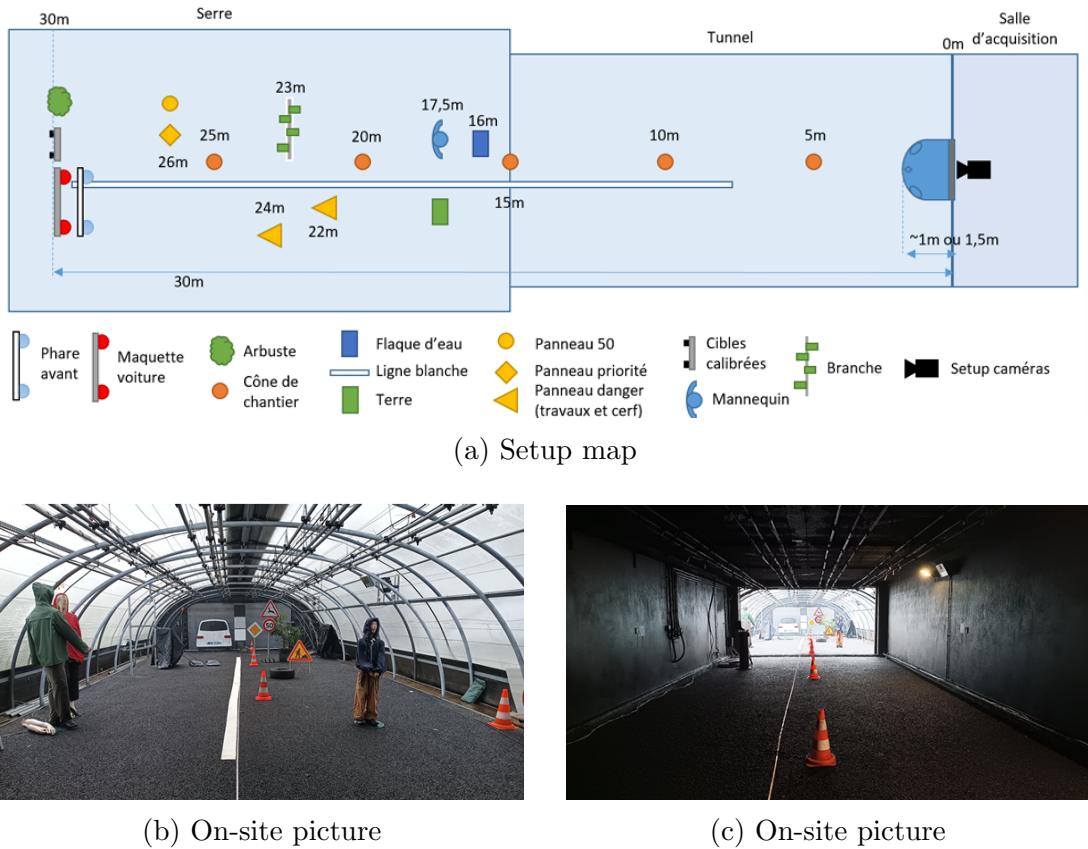


Figure 2.1: Experimental Setup for data acquisition

Data acquisition was conducted at Cerema, 8 Rue Bernard Palissy, Clermont-Ferrand [1], where controlled rain and fog environments allowed for realistic testing conditions. Various sequences, including different light conditions, were recorded. Figure 2.1a shows the map of the setup used. A total of four modalities

were used including one visible camera and two SWIR Cameras ([2], [3]). The fourth modality was created by fusing the Visible camera images with preprocessed images from both SWIR Cameras through TarDAL fusion [4]. Table 2.1 gives details about the cameras used.

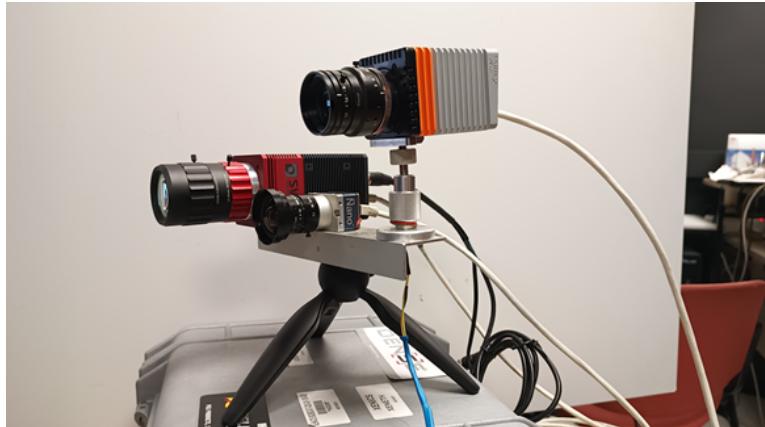


Figure 2.2: L to R - SVS Acuros CQD, Dalsa Genie Nano, Xenics Bobcat

Camera	Type	Range of Camera
Xenics Bobcat 320	InGaAs-based SWIR	900 nm–1700 nm
SVS Acuros CQD 1280	Quantum Dot-based SWIR	400 nm–1700 nm
Dalsa Genie Nano 1630	Visible Camera	380 nm–700 nm

Table 2.1: Specifications of Cameras Used for Data Acquisition

2.2 Preprocessing

Preprocessing is a critical step to ensure consistent and high-quality images across frames, especially when working with data collected under varying environmental conditions. By enhancing image clarity and reducing noise, preprocessing allows for more accurate analysis and interpretation in subsequent stages. To achieve this, several techniques such as standard histogram equalization, CLAHE, Thresholding, and Normalization were tried. Following this exploration, the most effective preprocessing steps were selected for the final pipeline:

- **Normalization and Thresholding:** Images were normalized, with thresholding applied at a 99% confidence interval. (Sample shown in Figure 2.3c and 2.3f)

Preprocessing steps were validated on older acquisitions from the Xenics camera in both night and rain conditions. Further validation was conducted using new acquisitions taken in night and fog conditions, confirming the effectiveness of the preprocessing pipeline.

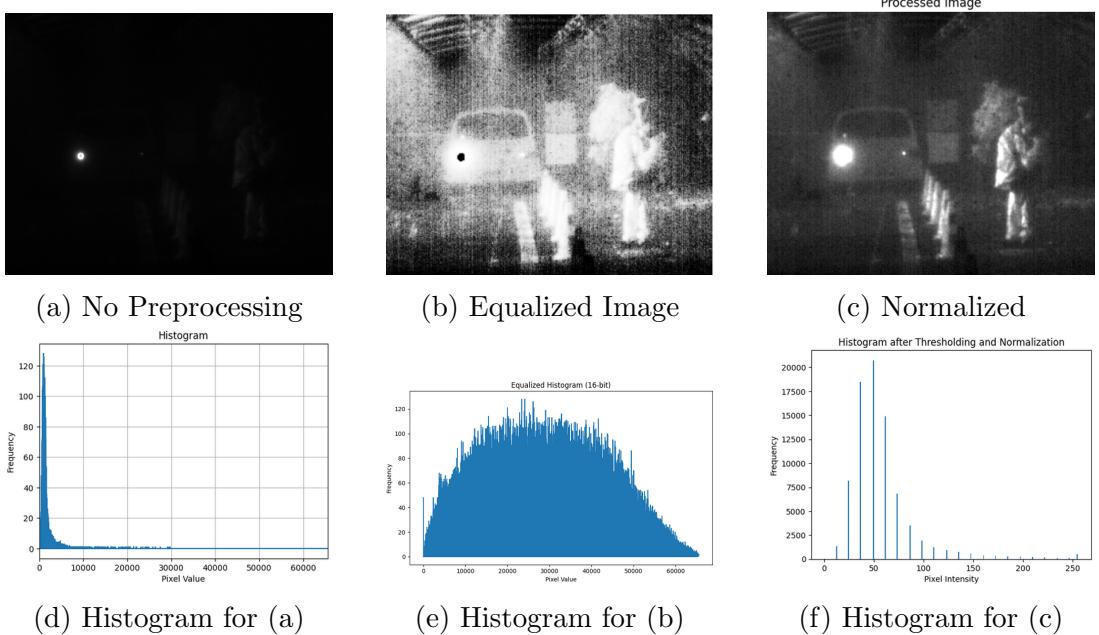


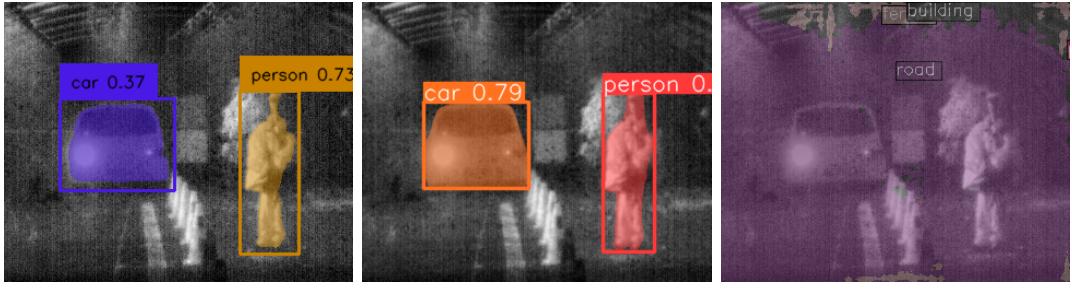
Figure 2.3: Top row: Images. Bottom row: Corresponding histograms.

2.3 Algorithms Used

Algorithms	Tasks	Avg Time/Image
Grounding DINO	Detection (open set)	10 s to 30 s
SAM	Segmentation (whole image)	30 s to 3 min
Grounded SAM	Detect+Segment (open set)	30 s to 2 min
YOLOv11-seg	Detect+Segment (COCO labels)	0.2 s to 0.4 s
YOLOv8-seg	Detect+Segment (COCO labels)	0.3 s to 0.5 s
YOLOv8-oiv	Detection (OpenImages)	1 s to 3 s
MMSegmentation	Detect+Segment (Cityscrape)	5 s to 10 s

Table 2.2: List of Algorithms used

The evaluation involved testing various detection and segmentation algorithms on the SWIR data as shown in Table 2.2. Among them, only Grounded SAM [5], YOLOv8-seg [6], and YOLOv11-seg [7] produced meaningful and valid detection and segmentation results. Other algorithms tested failed to adapt effectively to SWIR data, generating outputs that appeared random and highly inconsistent visually with expected results. Figure 2.4 provides a visual comparison of outputs from these algorithms, highlighting the superior performance of Grounded SAM and YOLO over others. the following sections of the report will focus solely on Grounded SAM, YOLOv8 and YOLOv11 algorithms for further analysis and discussion.



(a) Grounded SAM (b) YOLOv8-seg (c) MM Segmentation

ions by Semantic

entation algorithms.

Figure 2.4: Sample detections by Semantic Segmentation algorithms.

2.4 Evaluation Metrics

To evaluate the performance of the detection and segmentation algorithms, both quantitative and qualitative metrics were employed. These metrics provide a comprehensive framework for assessing the accuracy and consistency of the predictions compared to the ground truth.

2.4.1 Quantitative Evaluation

For quantitative evaluation, Precision, Recall, and F1 Score were used to measure the accuracy and reliability of the detection and segmentation outputs. These metrics are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Precision quantifies the proportion of true positive predictions out of all positive predictions made by the algorithm, indicating a low false positive rate.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall, also known as sensitivity, measures the proportion of true positive predictions out of all actual positive instances, showing how well the algorithm captures all relevant instances.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

The F1 Score is the harmonic mean of Precision and Recall, providing a balanced measure that considers both false positives and false negatives.

2.4.2 Qualitative Evaluation

For qualitative evaluation, Intersection over Union (IoU) was used. IoU provides a measure of overlap between the predicted segmentation and the ground truth segmentation, normalized by their union. It evaluates the spatial agreement of

the predicted segmentation and is particularly useful for visual comparisons. IoU is calculated as:

$$\text{IoU} = \frac{|\text{Intersection}|}{|\text{Union}|}$$

By combining these quantitative and qualitative evaluation metrics, the performance of the algorithms was assessed holistically, providing insights into both numerical accuracy and visual quality of the segmentation outputs.

2.5 Procedure

The methodology involved:

1. Data acquisition under rain and fog conditions.
2. Registration and synchronization of frames across modalities.
3. Manual annotation of ground truth for a pedestrian and a car in the sequences.
4. Running each model on the four modalities and analyzing segmentation outputs.
5. Quantitative evaluation through tables of IoU, precision, and recall.
6. Qualitative analysis through IoU vs. Visibility and IoU vs. Timestamp graphs, with Visibility plotted as an axis for comparative insights.

Chapter 3

Results and Analysis

This chapter presents the results of both quantitative and qualitative evaluations conducted on three distinct sequences: Foggy Day (also labeled as 51st Sequence), Rainy Day (also labeled as 23 et 24th Sequence), and Clear Day (also labeled as 21st Sequence). The evaluation involved assessing the performance of Grounded SAM, YOLOv8-seg, and YOLOv11-seg on individual modalities (SVS, Xenics, and Visible Cameras) and, in certain cases, the fused modality. Key insights and deductions are also discussed.

3.1 Quantitative Evaluation

For quantitative evaluation, Precision, Recall, and F1 Scores were computed for each sequence across the three modalities: SVS, Xenics, and Visible Cameras on preprocessed as well as non-preprocessed frames. Table 3.1 summarizes these metrics, providing a comparative assessment of performance OF Grounded SAM and YOLOv8.

3.2 Qualitative Evaluation

Qualitative evaluation involved analyzing segmentation performance through graphical comparisons for preprocessed images. The fused modality was included in comparisons but only for Foggy Day and Rainy Day sequences. No qualitative evaluation was performed for the fused modality itself, as it was used solely for comparison purposes.

3.2.1 Comparison of Modalities for the Same Algorithm

The performance of different imaging modalities (Visible, SVS, Xenics, and Fused) was evaluated under various environmental conditions using the same algorithm. The analysis includes comparisons for Foggy Day, Rainy Day, and Clear Day sequences, incorporating both preprocessed and non-preprocessed frames.

Sequence	Camera	Algorithm	Preprocessed	TP	FP	FN	Precision	Recall	F1 Score
51 (Foggy Day)	SVS	GroundedSAM	Yes	122	38	54	0.762	0.694	0.726
			No	116	30	60	0.794	0.659	0.720
		YOLOv8	Yes	128	36	48	0.780	0.727	0.753
			No	125	31	51	0.801	0.710	0.753
	Visible	GroundedSAM	Yes	103	1	73	0.990	0.585	0.736
			No	116	56	60	0.674	0.659	0.667
		YOLOv8	Yes	103	6	73	0.945	0.585	0.723
			No	104	3	72	0.972	0.591	0.735
23 et 24 (Rainy Day)	Xenics	GroundedSAM	Yes	83	1	93	0.989	0.472	0.638
			No	68	6	108	0.919	0.387	0.544
		YOLOv8	Yes	112	3	64	0.979	0.636	0.770
			No	80	0	96	1.000	0.455	0.625
	SVS	GroundedSAM	Yes	252	78	0	0.764	1.000	0.866
			No	252	85	0	0.748	1.000	0.856
		YOLOv8	Yes	252	61	0	0.805	1.000	0.892
			No	252	51	0	0.831	1.000	0.908
21 (Clear Day)	Visible	GroundedSAM	Yes	248	51	4	0.829	0.984	0.900
			No	251	28	1	0.900	0.996	0.945
		YOLOv8	Yes	251	3	1	0.988	0.996	0.992
			No	251	7	1	0.973	0.996	0.984
	Xenics	GroundedSAM	Yes	191	36	61	0.841	0.758	0.797
			No	202	53	50	0.793	0.802	0.797
		YOLOv8	Yes	193	8	59	0.960	0.766	0.852
			No	154	5	98	0.969	0.612	0.749

Table 3.1: Quantitative Evaluation across Different Sequences

For the initial analysis, IoU vs. Timestamp graphs with an added visibility axis were used to evaluate the temporal performance of the modalities under varying environmental conditions. Subsequently, visibility-based comparisons were conducted using IoU vs. Visibility graphs to directly assess the impact of visibility on detection and segmentation accuracy.

Figure 3.1 compares the performance of modalities for the Foggy Day sequence with preprocessed and non-preprocessed frames. Similarly, Figure 3.2 illustrates the comparison for the Rainy Day sequence, and Figure 3.3 shows results for the Clear Day sequence. The impact of preprocessing is highlighted, particularly in challenging conditions.

Fused images are also included in the analysis for Foggy Day and Rainy Day sequences, as shown in Figures 3.4 and 3.5. These results are complemented by visibility-based comparisons, presented in Figures 3.6 and 3.7.

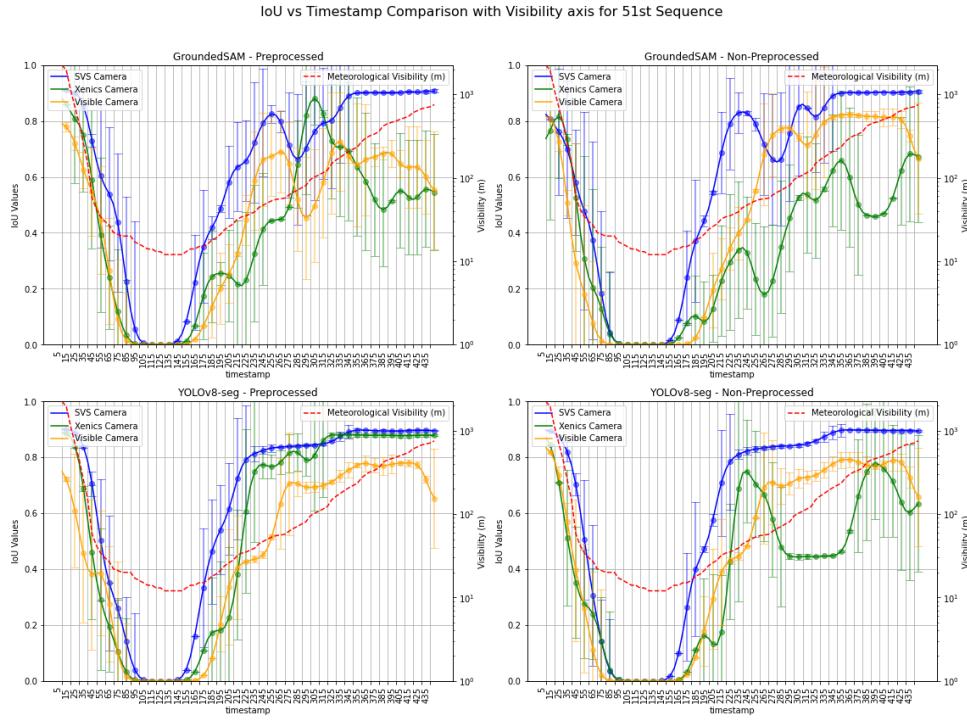


Figure 3.1: Comparison of Modalities for Foggy Day Sequence with preprocessed and non-preprocessed frames.

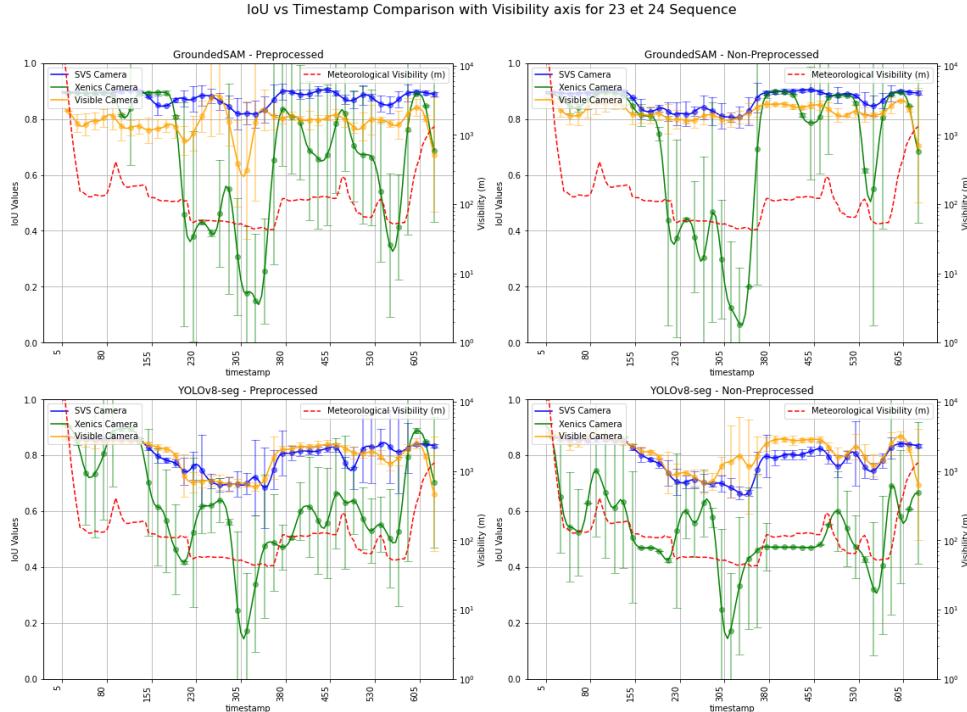


Figure 3.2: Comparison of Modalities for Rainy Day Sequence with preprocessed and non-preprocessed frames.

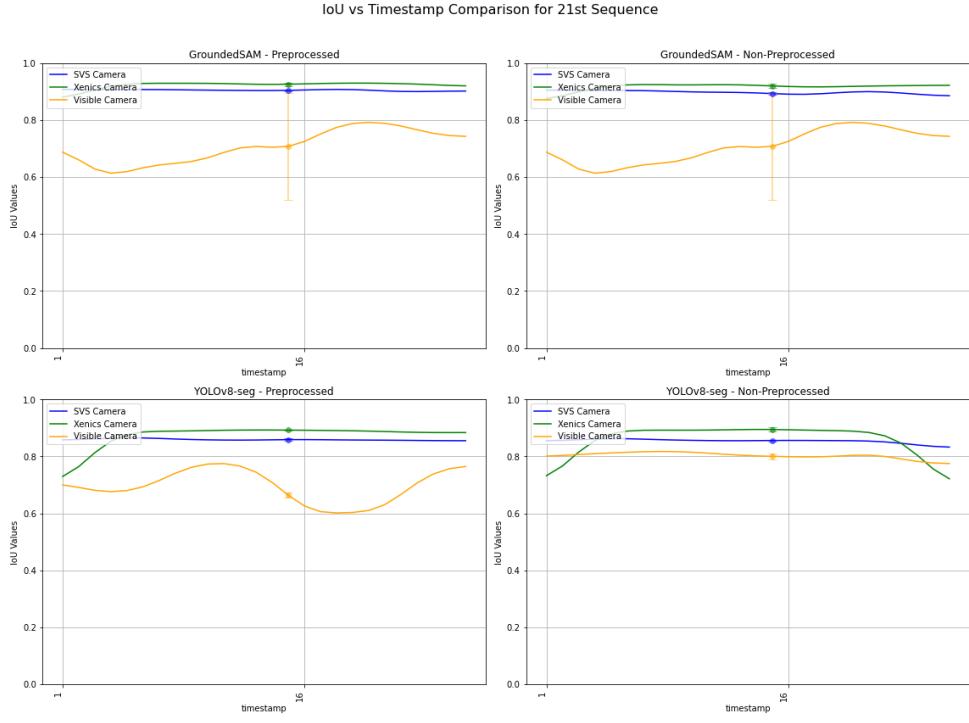


Figure 3.3: Comparison of Modalities for Clear Day Sequence with preprocessed and non-preprocessed frames.

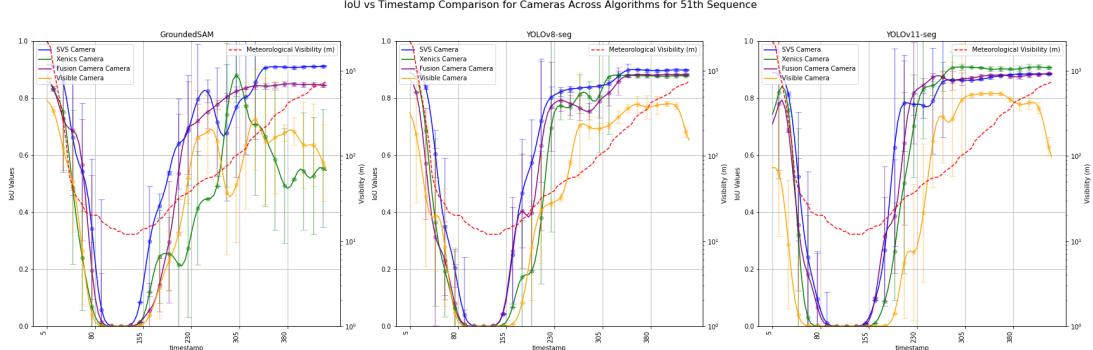


Figure 3.4: Comparison of Modalities with Fused Images for Foggy Day Sequence with preprocessed frames.

3.2.2 Comparison of Algorithms for the Same Modality

The performance of algorithms (YOLOv8-seg, YOLOv11-seg, and Grounded SAM) was analyzed for a single modality under Foggy Day, Rainy Day, and Clear Day sequences. The results include comparisons with preprocessed and non-preprocessed frames, as well as visibility-based analysis.

Figure 3.8 compares the performance of algorithms for the Foggy Day sequence. Similarly, Figure 3.9 shows the comparison for the Rainy Day sequence. The fused modality results are also presented in Figures 3.10 and 3.11 for Foggy Day and Rainy Day sequences, respectively. Visibility-based comparisons are

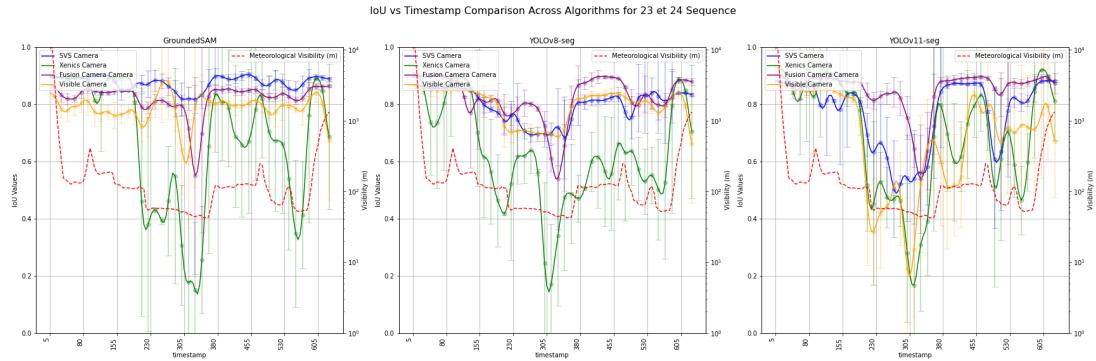


Figure 3.5: Comparison of Modalities with Fused Images for Rainy Day Sequence with preprocessed frames.

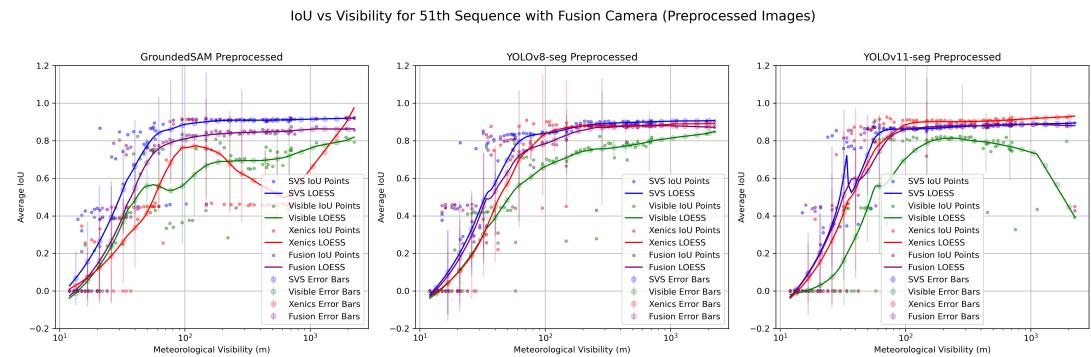


Figure 3.6: Modalities Comparison vs Visibility (Foggy Day Sequence)

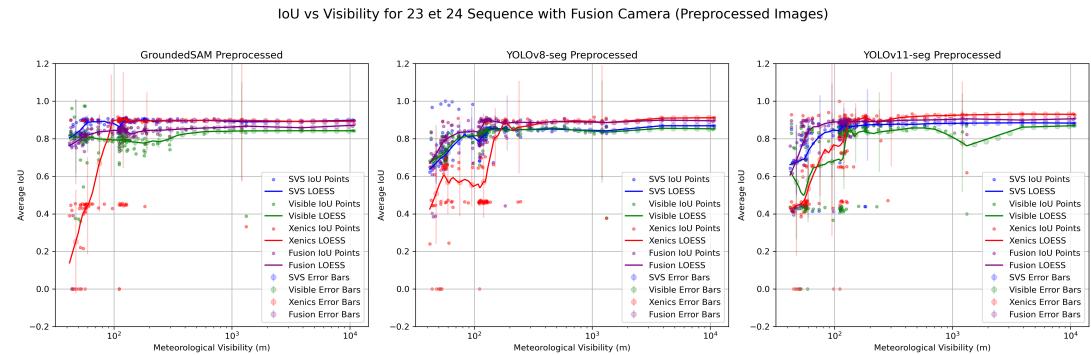


Figure 3.7: Modalities Comparison vs Visibility (Rainy Day Sequence)

provided in Figures 3.12 and 3.13.

3.3 Analysis and Deductions

Based on the results presented above, the following observations and deductions were made:

IoU Comparison vs Time for Algorithms for 51st Sequence

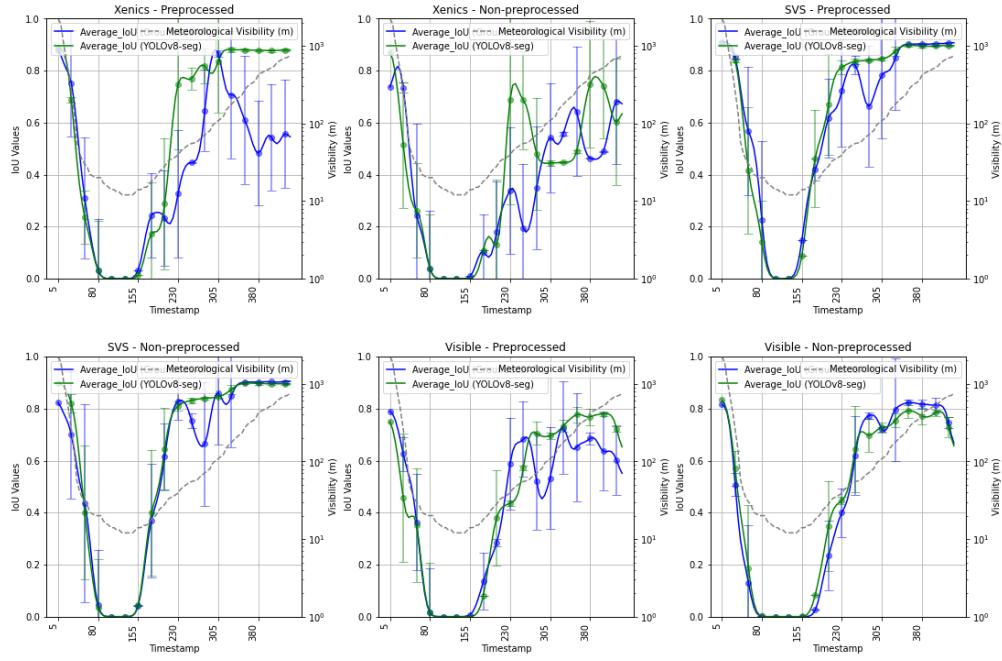


Figure 3.8: Comparison of Algorithms for Foggy Day Sequence with preprocessed and non-preprocessed frames.

IoU Comparison vs Time for Algorithms for 23 et 24 Sequence

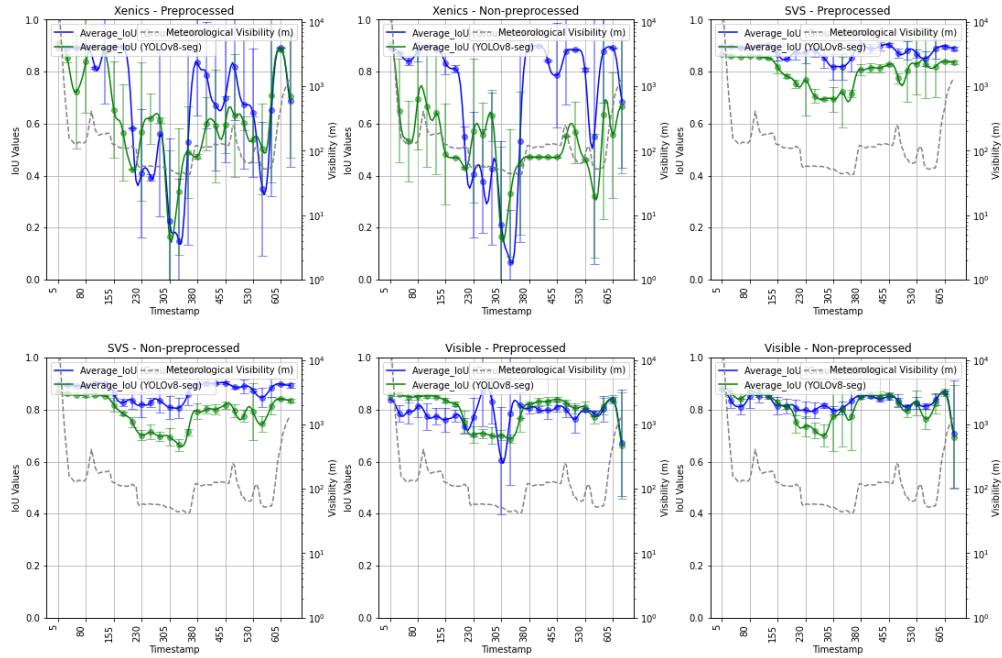


Figure 3.9: Comparison of Algorithms for Rainy Day Sequence with preprocessed and non-preprocessed frames.

Algorithm Comparison vs Time across Multiple Cameras for 51th Sequence

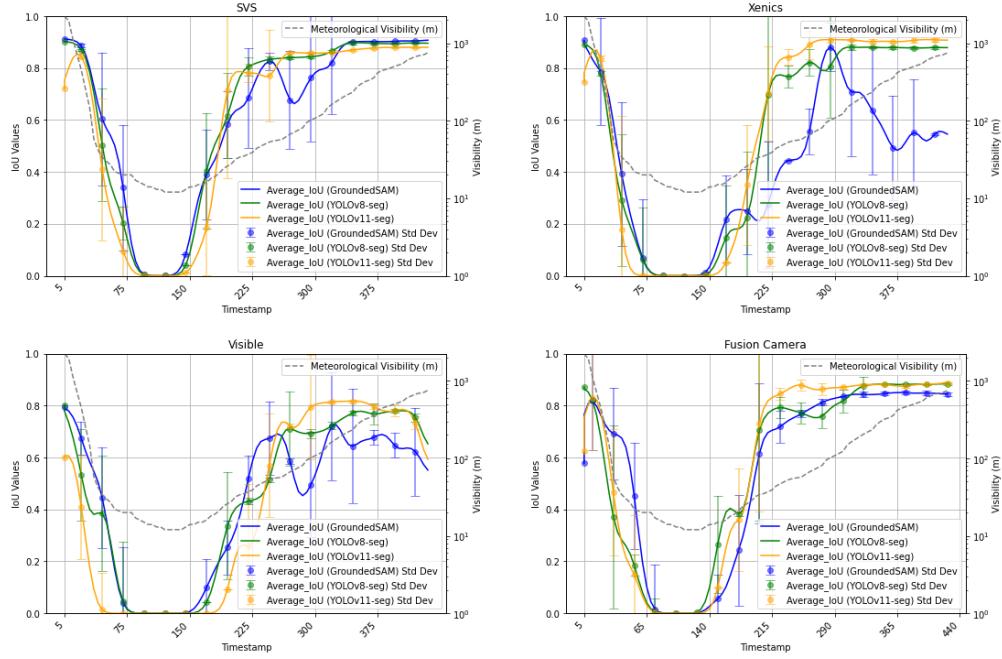


Figure 3.10: Comparison of Algorithms (including Fused Images and YOLOv11-seg) for Foggy Day Sequence with preprocessed frames.

Algorithm Comparison vs Time across Multiple Cameras for 23 et 24 Sequence

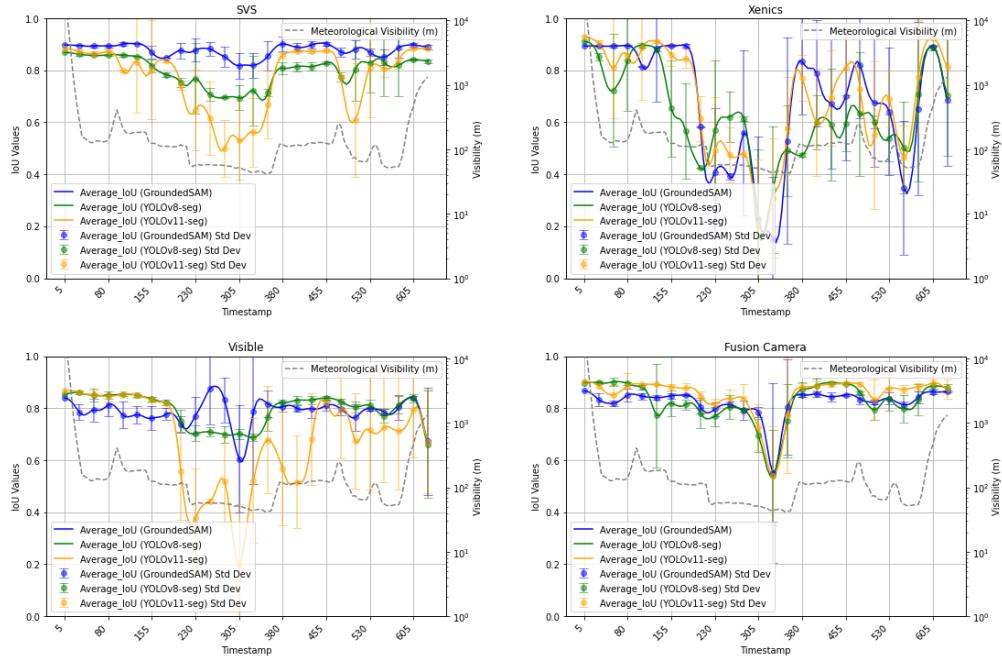


Figure 3.11: Comparison of Algorithms (including Fused Images and YOLOv11-seg) for Rainy Day Sequence with preprocessed frames.

IoU vs Visibility for GroundedSAM, YOLOv8-seg, and YOLOv11-seg for 51th Sequence (Preprocessed)

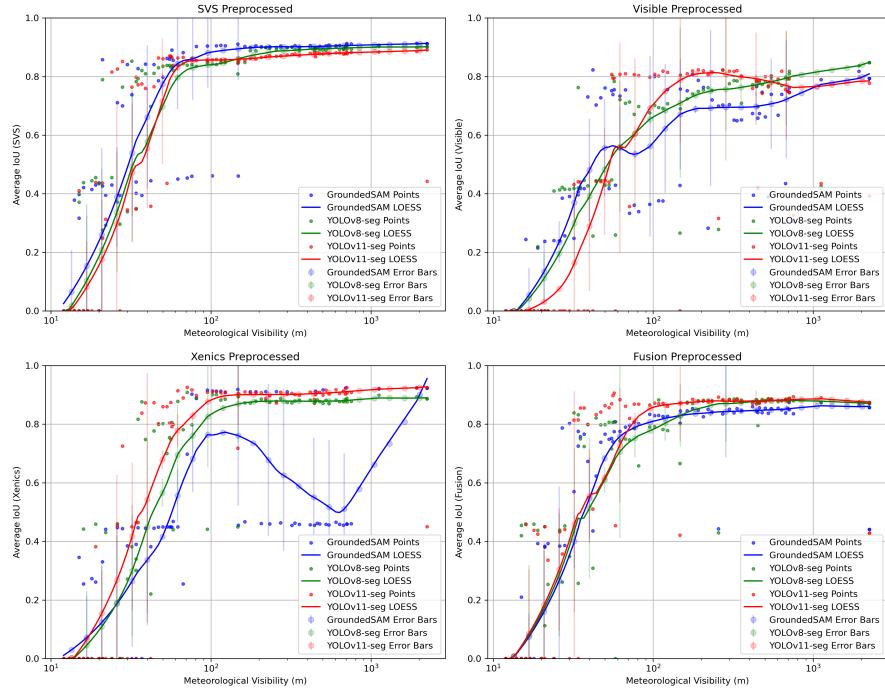


Figure 3.12: Algorithms Comparison vs Visibility (Foggy Day Sequences)

IoU vs Visibility for GroundedSAM, YOLOv8-seg, and YOLOv11-seg for 23 et 24 Sequence (Preprocessed)

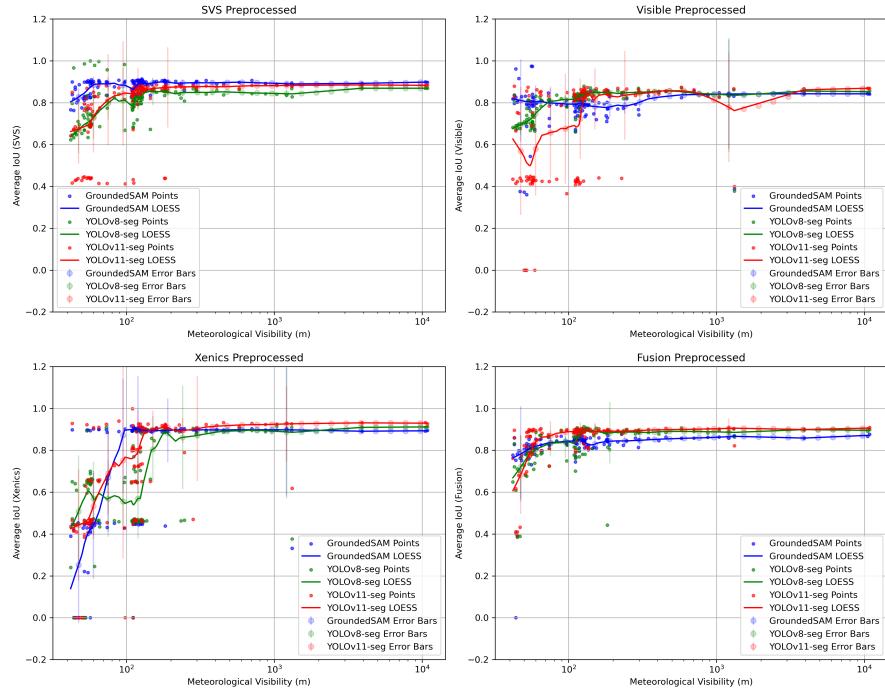


Figure 3.13: Algorithms Comparison vs Visibility (Rainy Day Sequences)

3.3.1 Quantitative Analysis

The quantitative evaluation provides insights into the performance of the algorithms (GroundedSAM and YOLOv8) across various environmental conditions and preprocessing states. Key observations include:

- **Impact of Preprocessing:** Preprocessing consistently improves performance for both algorithms, particularly in challenging conditions like *Foggy Day*. For example:
 - Xenics with YOLOv8 improved its F1 score from 0.625 (No Preprocessing) to 0.770 (With Preprocessing) in *Foggy Day*.
 - GroundedSAM also exhibited an improvement for SVS in *Rainy Day*, with its F1 score increasing from 0.856 (No Preprocessing) to 0.866 (With Preprocessing).
- **Performance by Algorithm:** YOLOv8 demonstrated higher precision across most cases, particularly for clearer conditions such as *Clear Day*, where it achieved an F1 score of 1.000 for the Visible camera (No Preprocessing). GroundedSAM, on the other hand, showed better performance in complex segmentation tasks, such as detecting objects in *Foggy Day*.
- **Environmental Impact:** Environmental conditions greatly influenced performance metrics. For instance:
 - In *Foggy Day*, all cameras experienced a significant drop in recall values, highlighting challenges in detecting objects under low visibility.
 - Conversely, in *Rainy Day*, precision and recall values remained high, especially for the Visible camera using YOLOv8, which achieved an F1 score of 0.992 (With Preprocessing).
- **Camera-Specific Trends:**
 - The SVS camera generally maintained stable F1 scores across conditions, showing robustness in diverse scenarios.
 - The Xenics camera's performance fluctuated more, particularly under foggy conditions, where preprocessing significantly enhanced its results.

3.3.2 Qualitative Analysis

IoU and Algorithm Performance Across Different Conditions

- **Figure 3.1: Comparison of Modalities for Foggy Day Sequence with Preprocessed and Non-Preprocessed Frames**
 - The SVS camera demonstrates remarkable stability in IoU values, particularly under low-visibility conditions, compared to other modalities.

- Preprocessed data consistently outperformed non-preprocessed data for both GroundedSAM and YOLOv8-seg algorithms, as observed in smoother curves and higher IoU values.
- Notably, the SVS modality achieved a baseline IoU even in areas with minimal visibility, highlighting its robustness.
- **Figure 3.2 Comparison of Modalities for Rainy Day Sequence with Preprocessed and Non-Preprocessed Frames**
 - The SVS camera once again exhibited more stable IoU values compared to other modalities.
 - The Xenics camera showed noticeable fluctuations in IoU values, even with preprocessing applied.
 - While preprocessing improved IoU values at certain points, the overall impact on segmentation was limited under rainy conditions.
- **Figure 3.3: Comparison of Modalities for Clear Day Sequence with Preprocessed and Non-Preprocessed Frames**
 - All modalities performed well in detecting and segmenting under clear day conditions.
 - The visible modality exhibited fluctuations due to the poor quality of the RGB camera. However, the error bars balanced out these inconsistencies, ensuring reasonable performance.
- **Figure 3.4: Comparison of Modalities with Fused Images for Foggy Day Sequence with Preprocessed Frames**
 - Fused images provided a stable IoU curve, effectively combining the strengths of individual modalities.
 - The SVS camera’s performance was comparable to fused images, offering reliable detections in low-visibility scenarios alongside stable IoU trends.
- **Figure 3.6 and Figure 3.7: Modalities Comparison vs. Visibility (Foggy and Rainy Day Sequences)**
 - These figures reinforce the above claims, showcasing the direct relationship between visibility and IoU values across all modalities.

Algorithmic Comparison Across Environmental Conditions

- **Figure 3.8: Comparison of Algorithms for Foggy Day Sequence with Preprocessed and Non-Preprocessed Frames**
 - YOLOv8 consistently delivered a more stable IoU curve, with segmentation and visibility performance comparable to GroundedSAM.

- **Figure 3.9: Comparison of Algorithms (Including Fused Images and YOLOv11-seg) for Foggy Day Sequence with Preprocessed Frames**
 - YOLOv11 struggled to provide superior segmentation in low-visibility areas.
 - YOLOv8 emerged as the most stable and reliable algorithm, outperforming GroundedSAM, which exhibited higher fluctuations.
- **Figure 3.11: Comparison of Algorithms (Including Fused Images and YOLOv11-seg) for Rainy Day Sequence with Preprocessed Frames**
 - The Xenics modality showed fluctuations across all three algorithms.
 - YOLOv8-seg again displayed superior stability, with IoU values comparable to those of GroundedSAM.
- **Figure 3.12 and Figure 3.13: Algorithms Comparison vs. Visibility (Foggy and Rainy Day Sequences)**
 - These figures provide an alternate perspective, affirming the consistency of YOLOv8-seg and highlighting GroundedSAM's fluctuations under varying visibility conditions.

Chapter 4

Conclusion and Future Work

4.1 Conclusion

This study demonstrates a proof of concept for leveraging Short-Wave Infrared (SWIR) imaging to enhance object detection and segmentation in harsh weather conditions. By analyzing the performance of deep learning algorithms trained on RGB images when applied to SWIR imagery, as well as comparing their results with visible-light images, several key insights were derived:

- **Preprocessing Validation:** The application of preprocessing techniques, including normalization and thresholding, was validated as a crucial step to improve image quality and segmentation performance across all modalities. Preprocessing showed substantial benefits, particularly under challenging conditions such as *Foggy Day*, where it significantly enhanced both IoU and F1 scores.
- **Superior Algorithm Performance:** Among the algorithms tested, YOLOv8-seg emerged as the most reliable model, consistently delivering superior detection and segmentation performance across all environmental conditions. Its ability to adapt to SWIR imagery and maintain stability under low-visibility scenarios positions it as the preferred algorithm for future applications. Grounded SAM also demonstrated promise in complex segmentation tasks, though it requires further optimization to address performance fluctuations.
- **SWIR Modality Effectiveness:** The SWIR imaging modalities (Xenics and SVS) proved their potential in harsh weather conditions. The SVS camera, in particular, exhibited remarkable stability and robust segmentation performance across diverse scenarios. The fused modality, created through TarDAL fusion, effectively combined the strengths of individual cameras, achieving enhanced IoU values and stable detection trends in challenging environments.
- **Environmental Resilience:** The study highlighted the importance of adaptive strategies to address the impact of environmental conditions on

detection performance. While severe weather conditions such as fog significantly impacted performance metrics, preprocessing and the utilization of SWIR modalities mitigated these challenges effectively.

In conclusion, this report validates the utility of SWIR imaging and preprocessing techniques for enhanced object detection and segmentation in adverse weather conditions. With YOLOv8-seg leading the way, this proof of concept lays a strong foundation for further research and development in leveraging SWIR technology for real-world applications.

4.2 Future Work

While this study establishes the efficacy of SWIR imaging for object detection and segmentation, several avenues for future research remain:

- **Transfer Learning:** Future work could involve fine-tuning deep learning models directly on SWIR datasets to improve their performance further.
- **Expanded Dataset:** Collecting and training on larger and more diverse SWIR datasets, encompassing various environmental conditions and object classes, will enhance model robustness.
- **Synthetic Dataset Creation:** Developing a synthetic SWIR dataset using generative models or simulation techniques could address the current lack of a full-fledged SWIR training dataset. This would enable better pretraining and benchmarking for SWIR-specific tasks.
- **Fusion Optimization:** Further exploration of fusion techniques could yield even better results by combining the strengths of individual modalities more effectively.
- **Algorithmic Advancements:** Improving stability and precision in segmentation algorithms, particularly for low-visibility conditions, remains a priority.

By addressing these areas, future research can build upon the foundation established by this work and extend the capabilities of SWIR imaging in practical scenarios.

References

- [1] S. Liandrat, P. Duthon, F. Bernardin, B. Daoued, and J.-L. Bicard, “A review of cerema pavin fog rain platform : from past and back to the future,” 2022.
- [2] SWIR Vision Systems, “Acuros SWIR Cameras — SWIR Vision Systems,” accessed: 2024-11-06. [Online]. Available: <https://www.swirvisionsystems.com/acuros-swir-camera/>
- [3] Exosens, “BOBCAT/BOBCAT+,” accessed: 2024-11-06. [Online]. Available: <https://www.exosens.com/products/bobcat->
- [4] J. Liu, X. Fan, Z. Huang, G. Wu, R. Liu, W. Zhong, and Z. Luo, “Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5792–5801.
- [5] T. Ren, S. Liu, A. Zeng, J. Lin, K. Li, H. Cao, J. Chen, X. Huang, Y. Chen, F. Yan, Z. Zeng, H. Zhang, F. Li, J. Yang, H. Li, Q. Jiang, and L. Zhang, “Grounded sam: Assembling open-world models for diverse visual tasks,” 2024. [Online]. Available: <https://arxiv.org/abs/2401.14159>
- [6] G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics yolov8,” 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [7] G. Jocher and J. Qiu, “Ultralytics yolo11,” 2024. [Online]. Available: <https://github.com/ultralytics/ultralytics>