

Robustness of Object Detection Models Under Simulated Weather Hazards: A Comparative
Study Using nuScenes

Sittipat TAPSUTAR, Thanatus SUWAN

Kasetsart University

Author Note

Department of Computer Science, Faculty of Science, Kasetsart University

E-mail: sittipat.te@ku.th, thanatus.s@ku.th

Abstract

This study investigates the robustness of object detection models under adverse weather conditions using the nuImages–nuScenes dataset. By simulating eight distinct weather hazards across three severity levels, we evaluate the performance of several models—SSD, YOLO, DETR, and Mask R-CNN—based on precision, recall, F1 score, and Intersection over Union (IoU). The findings highlight the trade-offs between detection accuracy and completeness, with Mask R-CNN demonstrating the most balanced performance.

Keywords: Object Detection, Adverse Weather Conditions, nuImages–nuScenes Dataset, SSD, YOLO, DETR, Mask R-CNN, Precision, Recall, F1 Score, Intersection over Union (IoU), Model Robustness, Detection Accuracy, Environmental Hazard

Introduction

Autonomous driving has become one of the most rapidly advancing fields in intelligent transportation systems, combining computer vision, deep learning, and sensor fusion to enable vehicles to perceive and understand their surroundings. A core component of any autonomous driving system is object detection, which allows the vehicle to identify and localize critical elements such as pedestrians, vehicles, traffic signs, and obstacles in real time. Accurate detection is essential not only for navigation and path planning but also for ensuring the overall safety and reliability of the driving process.

However, in real-world driving scenarios, the performance of these object detection models often degrades significantly under adverse environmental conditions. Factors such as fog, rain, snow, dust, heatwaves, and motion blur can distort the visual input captured by cameras and other sensors. These corruptions reduce image clarity, alter brightness and contrast, and introduce noise, making it difficult for deep learning models, especially those trained on clean datasets, to recognize objects accurately. As a result, autonomous vehicles may misclassify or fail to detect critical obstacles, leading to potential safety risks.



Fig.1.1 Detection on a clean image



Fig.1.2 Detection on a rain filtered image

Figure 1.1 illustrates the performance of the Mask R-CNN model on a clean image, where the red bounding box represents the model's detection and the green box indicates the ground truth label. The alignment between the two boxes demonstrates high model confidence and accurate object localization under ideal conditions. In contrast, Figure 1.2 presents the same image subjected to a synthetic rain filter, simulating adverse weather conditions. The detection results reveal a noticeable decline in model confidence, evidenced by misaligned bounding boxes and increased false positives. This comparison underscores the vulnerability of Mask R-CNN to visual distortions introduced by rain, highlighting the need for enhanced robustness in object detection models deployed in real-world autonomous driving scenarios.

Background

Autonomous driving systems rely heavily on accurate object detection to perceive vehicles, pedestrians, and obstacles in real-world environments. These systems depend on deep learning models such as Fast R-CNN, Faster R-CNN, Mask R-CNN, SSD, YOLO, and DETR, which achieve high accuracy on clean datasets but often fail under adverse weather conditions such as fog, rain, snow, sandstorms, heatwaves, blur, and low illumination. Such environmental degradation reduces image clarity and leads to significant drops in detection accuracy, making robustness a key challenge for real-world deployment.

To address robustness issues, researchers have explored approach [1]. The article addresses a critical vulnerability in autonomous driving systems: the lack of robustness in multi-sensor fusion models—particularly those combining camera and LiDAR data—when exposed to sensor corruptions such as adverse weather, misalignment, or hardware failures. These models typically rely on clean, complete sensor inputs, making them fragile in real-world conditions where disruptions are common. To tackle this, the authors introduce MSC-Bench, a comprehensive benchmark designed to evaluate and analyze the resilience of perception models under 16 types of sensor corruptions. By systematically testing six 3D object detection models and four HD map construction models across various corruption scenarios, MSC-Bench reveals significant performance degradation and highlights the need for more robust fusion strategies that can adapt to partial or degraded sensor data.

This study introduces a camera-based evaluation framework for object detection under controlled and reproducible weather corruptions. Six major models Fast R-CNN, Faster R-CNN, Mask

R-CNN, SSD, YOLO, and DETR are tested across eight weather types (dark, blur, motion blur, fog, snow, rain, sandstorm, and heatwave) and three severity levels (1–3). Performance is measured using Precision, Recall, and Mean IoU@0.50 (TP-only) for six object classes (car, bus, bicycle, motorcycle, person, truck). Unlike previous works that emphasize overall mAP, this study provides fine-grained, per-class and per-weather analysis, revealing which models perform best under specific environmental conditions. The results offer both theoretical and practical insight guiding the selection of suitable object detection models for different weather scenarios in autonomous driving.

Methodology

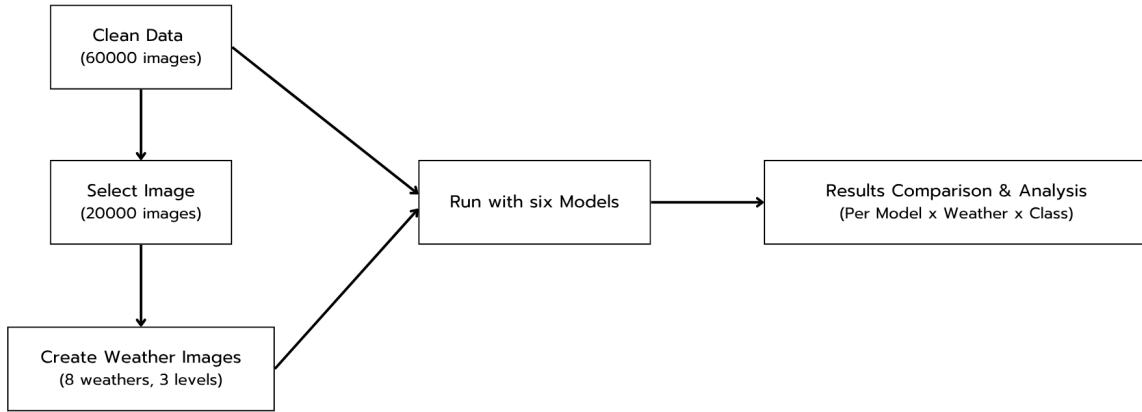


Fig. 3.1. Abstract workflow diagram

This study evaluates the robustness of six object detection models—FastRCNN, FasterRCNN, SSD, YOLO, DETR, and Mask R-CNN—under simulated adverse weather conditions using the nuImages–nuScenes dataset. The baseline dataset comprises 60,000 clean images. To simulate environmental challenges, eight types of synthetic weather hazards were applied: blur, darkness, fog, heatwave, motion blur, rain, sandstorm, and snow. 20,000 clean images were sampled and synthetically corrupted for each weather hazard type. Each hazard was simulated across three distinct severity levels, yielding a total of 60,000 corrupted samples per hazard category.

To assess model performance across these conditions, four key metrics were employed: Precision, Recall, F1 Score, and Intersection over Union (IoU). These metrics collectively measure both the accuracy of object classification and the spatial precision of object localization.

Performance Metrics

Precision measures the proportion of correctly identified objects among all detected objects. High precision indicates fewer false positives, which are calculated as follows:

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

where TP denotes an object correctly detected ($\text{IoU} > 0.5$) and correctly classified and FP denotes an object is detected but misclassified

Recall captures the proportion of actual objects that were successfully detected. High recall indicates fewer missed detections, which are calculated as follows:

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

where FN denotes an object that is present but not detected.

F1 Score represents the harmonic mean of precision and recall, balancing both metrics to provide a single measure of detection quality. A high F1 score indicates a good trade-off between precision and recall.

Intersection over Union (IoU) evaluates the spatial accuracy of object localization by comparing the overlap between predicted and ground truth bounding boxes. IoU ranges from 0 (no overlap) to 1 (perfect overlap). In this study, IoU is computed from true positive detections.

Resilience Score (RS) [1] introduced the relative robustness indicator for measuring how much accuracy a model can retain when evaluated on the corruption sets, which are calculated as follows:

$$RS_i = \frac{\sum_{l=1}^3 Acc_{i,l}}{3 \times Acc^{clean}}, \quad mRS_i = \frac{1}{N} \sum_{i=1}^N RS_i$$

where $\text{Acc}_{i,l}$ denotes the task-specific accuracy scores, with f1 score, on corruption type i at severity level l. N is the total number of corruption types, and $\text{Acc}^{\text{clean}}$ denotes the f1 score on the “clean” evaluation set.

Autoencoder Denoising

To further evaluate the resilience of object detection systems, we conducted experiments using the nuImages dataset augmented with hazardous weather conditions (darkness, rain, snow, and fog) with highest severity level. A total of 48,000 cropped training samples per hazard were generated from 1,000 base images, with corresponding clean-hazard pairs used to train hazard-specific autoencoders. Testing was performed on 200 randomly selected images subjected to the same hazards.

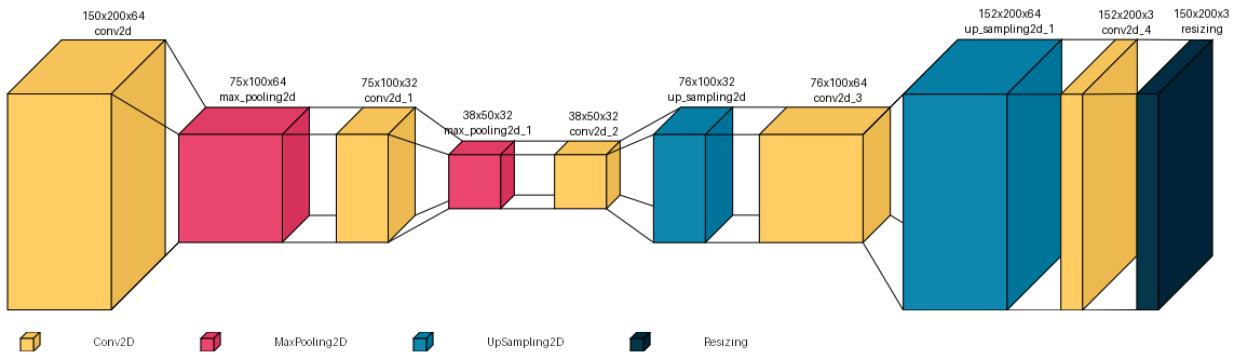


Fig. 3.2. Autoencoder architecture

Input: RGB crops of size $200 \times 150 \times 3$

Encoder: Conv2D(64) → MaxPool → Conv2D(32) → MaxPool

Decoder: Conv2D(32) → UpSample → Conv2D(64) → UpSample → Conv2D(3, sigmoid)

Optimization: Adam optimizer with MSE loss

Output: Denoised full images reconstructed by merging crops

This architecture allowed hazard-specific autoencoders to learn mappings from corrupted to clean image domains, producing denoised outputs that were subsequently evaluated with object detectors.



Fig. 3.3. Comparison of rain filtered image, original clean image and denoised image

Multi-Level and Cross-Autoencoder Approaches

Beyond single-hazard models, multi-level and cross-autoencoder strategies were explored. These approaches leverage pretrained weights for fog and rain denoising, demonstrating that autoencoders can generalize across multiple hazards. Results showed notable improvements in complex scenarios (e.g., fog + rain), highlighting the potential of layered denoising pipelines.

Layer 1: Parallel Pretrained Autoencoders

A copy of the input image is fed into two separate pretrained autoencoders, each specialized for a specific hazard domain (e.g., fog and rain). Each autoencoder independently reconstructs a denoised version of the corrupted input. The outputs from these two autoencoders are then concatenated along the channel dimension, resulting in a tensor of size **150×200×6** (two RGB outputs stacked together). Importantly, the **weights of these pretrained autoencoders are**

frozen during subsequent training, ensuring that their learned hazard-specific representations remain intact and stable.

Layer 2: Crossencoder Training

The concatenated feature map ($150 \times 200 \times 6$) serves as the input to a **new autoencoder**, **termed the crossencoder**. This crossencoder is trained to learn how to combine and refine the dual outputs into a single clean image representation. The architecture mirrors the standard encoder-decoder pipeline same as above autoencoder denoising architecture.

The training objective is to minimize the mean squared error (MSE) between the crossencoder output and the corresponding clean ground-truth image.



Fig. 3.4. Comparison of fog + rain filtered image, original clean image and denoised image

Limitations of These Studies

Limitation of this study lies in the computational demands required to run inference across large-scale, corrupted datasets. Autoencoder training and Object detection models such as Mask R-CNN and DETR are particularly resource-intensive, necessitating high-performance hardware and extended processing time to generate results.

Hazard augmentations were generated synthetically. While effective for controlled experiments, synthetic corruption may not perfectly replicate real-world weather conditions, leading to a domain gap that could affect deployment in practical applications.

Future Work

A natural extension of this study is the integration of a **hazard classifier** as a preprocessing stage. Instead of sending every image through multiple autoencoders, the classifier would first analyze the input and determine whether it belongs to a single hazard domain (such as fog, rain, snow, or darkness) or a multi-hazard scenario. Images identified as single-hazard could then be routed directly to the corresponding pretrained autoencoder, while multi-hazard images would be processed in parallel through multiple autoencoders before entering the crossencoder for refinement.

This approach offers several advantages. By introducing classification, the system can reduce unnecessary computational overhead, since only complex multi-hazard inputs would require the full multi-level pipeline. It also improves adaptability, as the classifier can guide dynamic resource allocation and ensure that the most appropriate denoising path is chosen for each image. Furthermore, the classifier-driven workflow better reflects real-world deployment, where autonomous systems must first recognize environmental hazards before applying corrective preprocessing.

Future research should also explore adaptive training strategies that link classifier confidence to crossencoder fine-tuning. For example, uncertain classifications could trigger joint optimization

of hazard-specific autoencoders and the crossencoder, improving robustness to ambiguous or overlapping conditions. Ultimately, this layered design—classification followed by hazard-aware denoising—would strengthen the pipeline’s scalability and bring it closer to practical application in safety-critical domains such as autonomous driving.

Experimental Results

To evaluate the robustness of object detection models under adverse weather conditions, we conducted extensive experiments using synthetically corrupted versions of the nuImages–nuScenes dataset. Each weather hazard—blur, darkness, fog, heatwave, motion blur, rain, sandstorm, and snow—was applied at three severity levels, resulting in 60,000 images per hazard. Six models were tested: FastRCNN, FasterRCNN, SSD, YOLO, DETR, and Mask R-CNN.

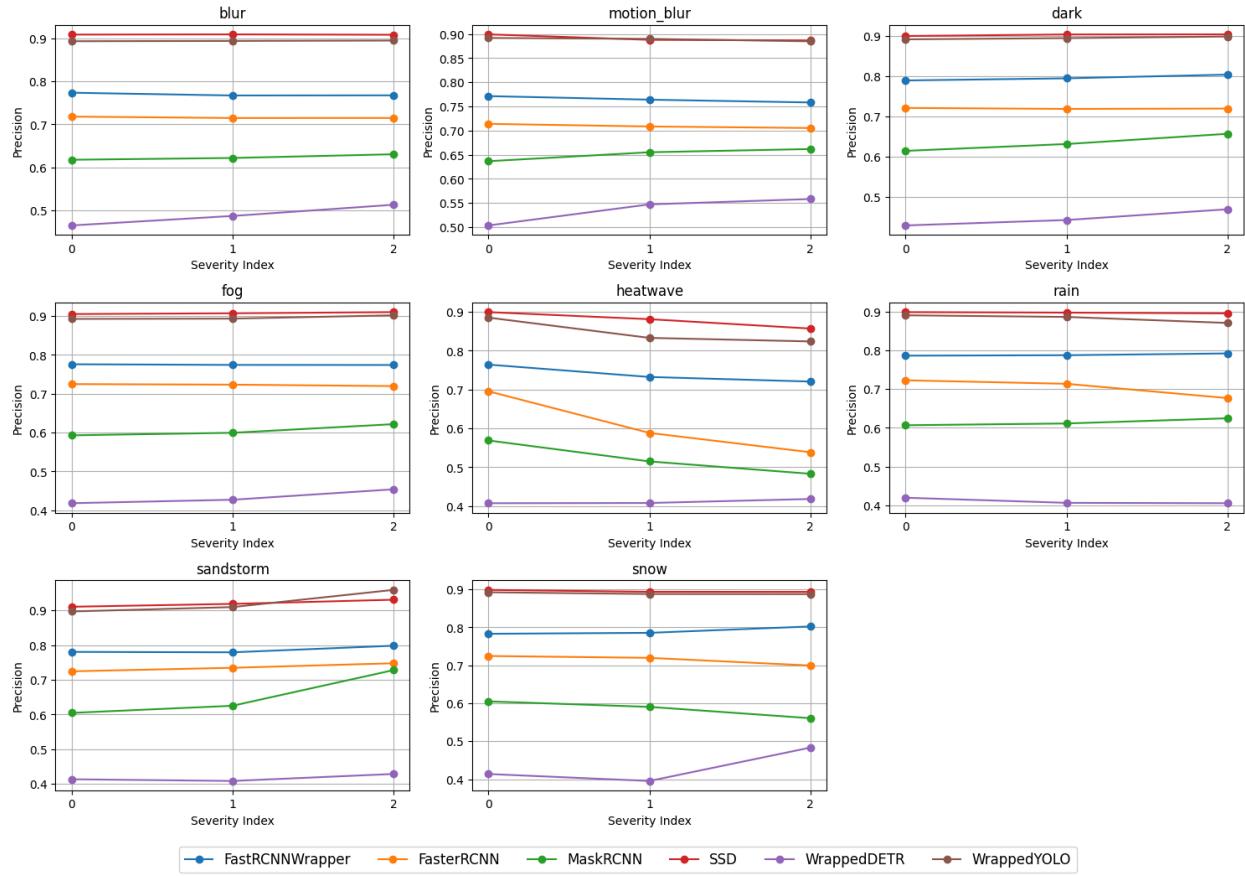


Fig. 4.1. Precision score drop by weather hazard type across models.

According to Fig. 4.1. SSD and YOLO consistently achieved the highest precision scores across all conditions, indicating strong object classification capabilities. Their performance

remained stable under most weather hazards, with only a slight decline observed during heatwaves.

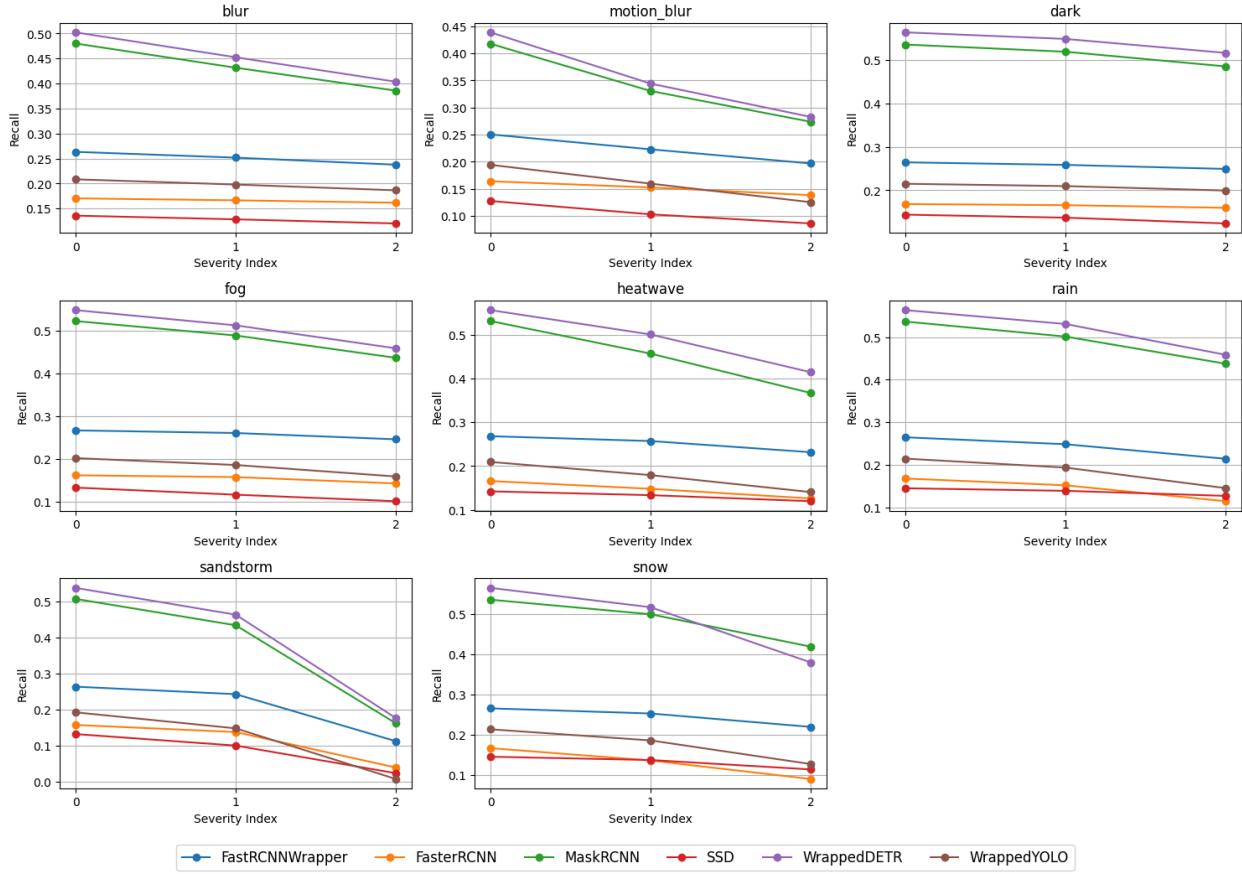


Fig. 4.2. Recall score drop by weather hazard type across models.

Despite high precision, SSD and YOLO exhibited low recall presented by Fig. 4.2, suggesting frequent missed detections. DETR, in contrast, achieved the highest recall but suffered from the lowest precision, indicating a tendency to detect more objects at the cost of increased false positives.

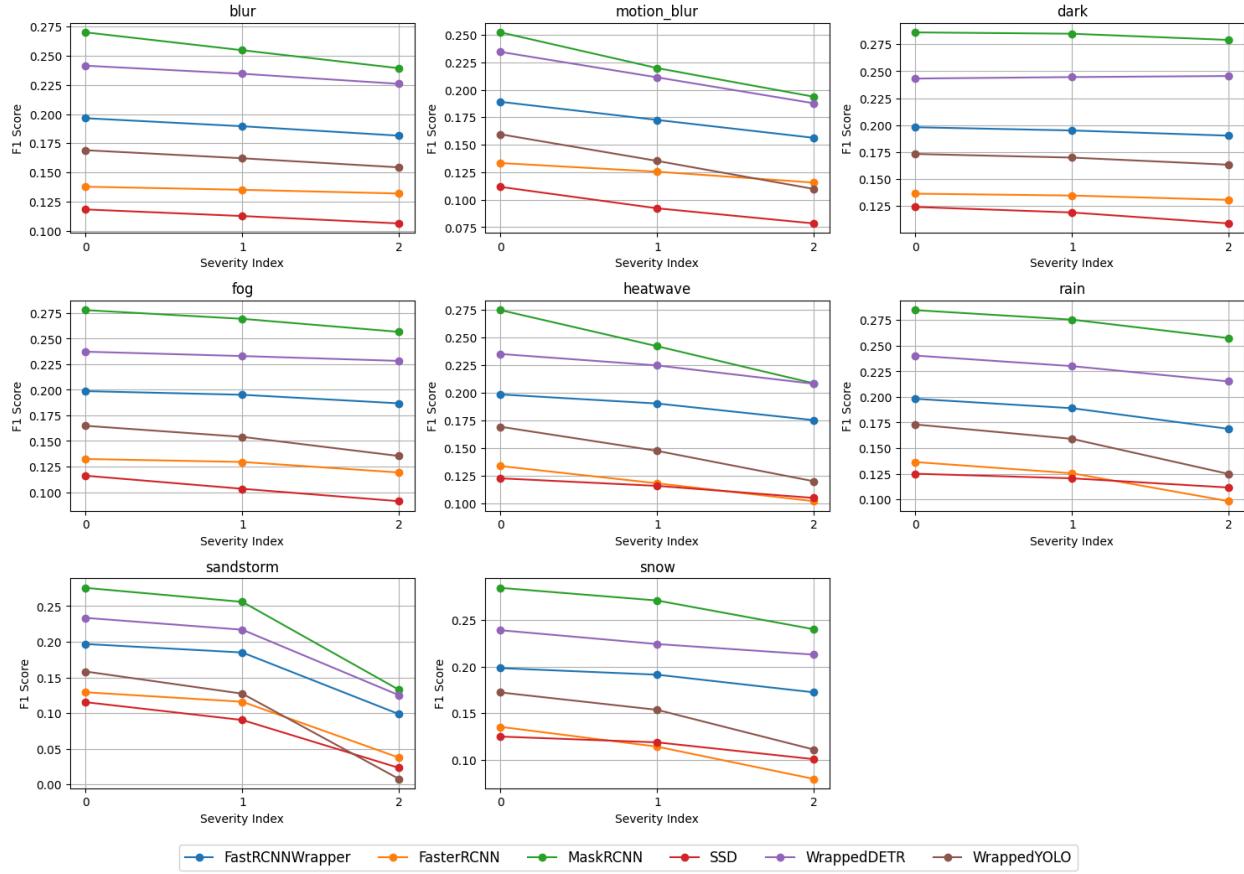


Fig. 4.3. F1 score drop by weather hazard type across models.

Mask R-CNN demonstrated the most balanced performance, achieving high F1 scores across all conditions. This reflects its ability to maintain a strong trade-off between precision and recall, making it effective in both classification and detection tasks.

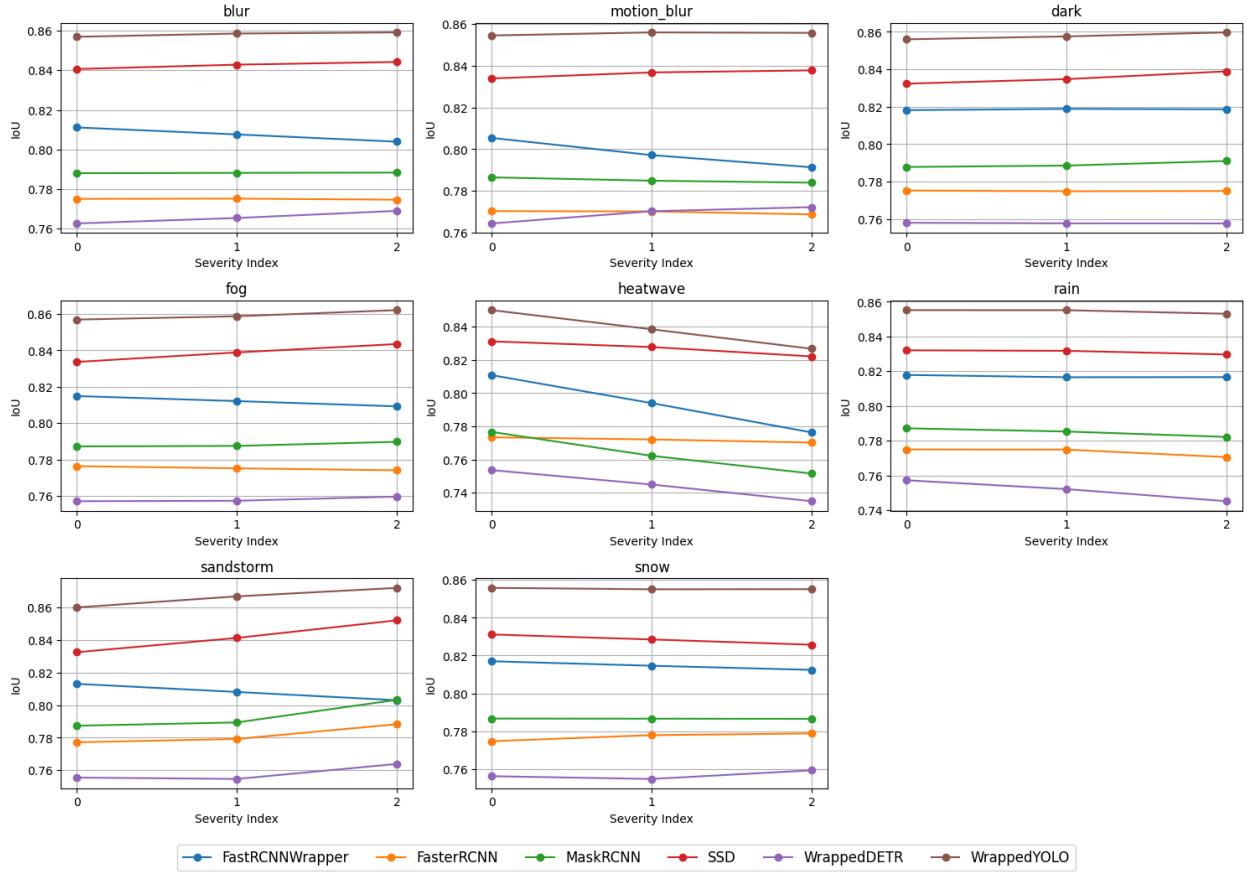


Fig. 4.4. IoU drop by weather hazard type across models.

YOLO and SSD recorded the highest IoU, confirming their superior localization accuracy. DETR had the lowest IoU, consistent with its lower precision. Among all weather conditions, heatwaves had the most detrimental effect on IoU, while other hazards showed minimal impact.

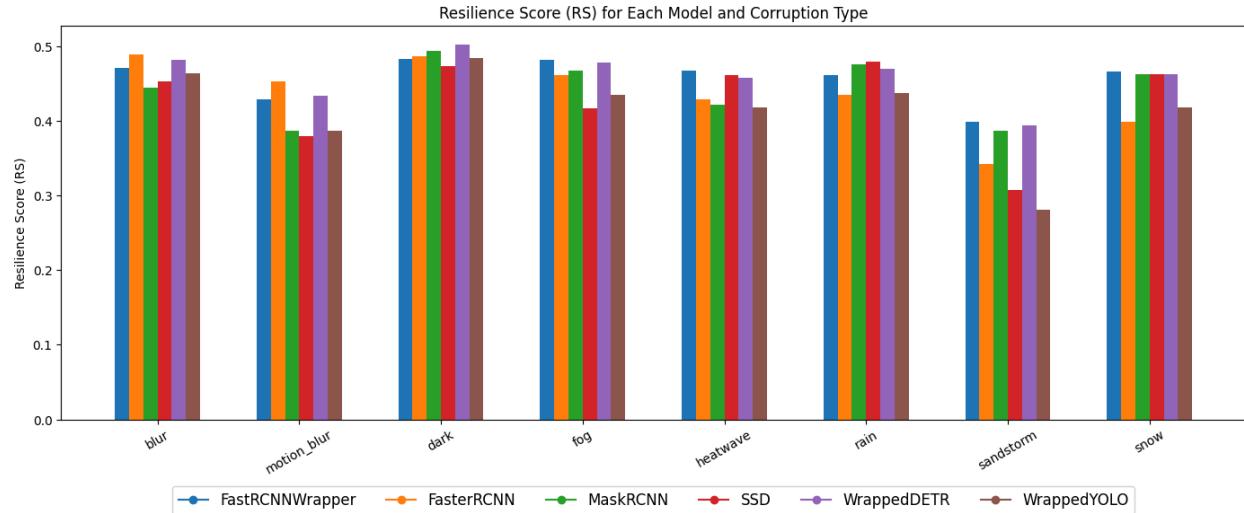


Fig. 4.5. Resilience score by weather hazard type across models.

According to the data presented in Figure 4.5, sandstorm and motion blur conditions exhibit the most pronounced negative impact on resilience scores across object detection models. This indicates that these two weather hazards significantly degrade model robustness and detection reliability. Although other environmental conditions also contribute to performance deterioration, sandstorm and motion blur consistently result in the steepest declines, underscoring their disruptive influence on visual perception systems.

Table 4.1. Average Precision of RCNN Models Across Object Classes

Average of Precision		Object Classes					
Models		bicycle	bus	car	motorcycle	person	truck
FasterRCNN		0.61	0.35	0.78	0.78	0.60	0.55
FastRCNNWrapper		0.58	0.57	0.84	0.78	0.75	0.60
MaskRCNN		0.42	0.45	0.64	0.68	0.60	0.53
SSD		0.82	0.74	0.95	0.95	0.75	0.74
WrappedDETR		0.38	0.39	0.50	0.57	0.41	0.29
WrappedYOLO		0.81	0.54	0.93	0.89	0.84	0.73
Average Precision		0.60	0.51	0.77	0.77	0.66	0.57

Table 4.1 presents the average precision scores of six object detection models evaluated across six key object classes—bicycle, bus, car, motorcycle, person, and truck. SSD and WrappedYOLO emerged as the top-performing models, with SSD achieving the highest precision in detecting cars (0.95) and motorcycles (0.95), while WrappedYOLO closely followed with strong scores for cars (0.93), motorcycles (0.89), and persons (0.84). FastRCNNWrapper also demonstrated competitive performance, particularly in detecting cars (0.84) and persons (0.75), outperforming FasterRCNN and MaskRCNN in most categories. In contrast, WrappedDETR consistently recorded the lowest precision across all object classes, with its weakest performance in truck detection (0.29). Overall, SSD maintained the highest average precision across all classes (0.79), indicating its superior capability in object classification tasks, especially for vehicles. These results underscore the importance of model selection based on object category, with SSD and WrappedYOLO offering robust detection across diverse urban entities.

Further details regarding the recall, Intersection over Union (IoU), and F1 score for each object class and detection model are comprehensively presented in the **Appendix** of this document. These metrics provide deeper insights into the models' detection completeness, localization accuracy, and overall performance balance under varying conditions.

Inference Time

Table 4.2. Inference Time for each weather condition and RCNN models

Processor AMD Ryzen 5 8645HS w/ Radeon 760M Graphics (4.30 GHz) - 16 Gb RAM						
NVIDIA GeForce RTX 4050 Laptop GPU (6 Gb)						
Using CUDA Version 12.9						
Inference Time for each weather condition type and models with 60,000 images (h:mm:ss)						
Weather Type	FastRCNNWrapper	FasterRCNN	MaskRCNN	SSD	WrappedDETR	WrappedYOLO
blur	2:52:44	0:32:00	2:07:02	0:53:15	2:43:06	1:00:43
motion_blur	2:50:11	0:40:56	2:01:54	0:56:21	2:46:17	1:07:00
dark	2:52:30	0:40:45	2:09:54	0:54:29	2:42:26	1:02:25
fog	2:52:16	0:44:51	2:11:30	0:56:15	2:43:51	1:06:04
heatwave	2:51:40	0:41:35	2:19:37	0:57:07	2:49:28	1:04:32
rain	2:49:21	0:42:01	2:11:07	0:58:00	2:47:24	1:08:16
sandstorm	2:48:31	0:40:42	2:07:13	0:56:22	2:48:37	1:07:35
snow	2:50:06	0:42:39	2:15:25	0:57:16	2:47:06	1:09:19
clean	2:50:50	0:32:27	2:14:55	0:56:22	2:44:13	1:03:19
Prediction rate for each weather condition and models (images/sec)						
Weather Type	FastRCNNWrapper	FasterRCNN	MaskRCNN	SSD	WrappedDETR	WrappedYOLO
blur	5.79	31.24	7.87	18.78	6.13	16.47
motion_blur	5.88	24.43	8.2	17.75	6.01	14.93
dark	5.8	24.54	7.7	18.36	6.16	16.02
fog	5.8	22.29	7.61	17.78	6.1	15.14
heatwave	5.83	24.05	7.16	17.51	5.9	15.49
rain	5.91	23.8	7.63	17.24	5.97	14.65
sandstorm	5.93	24.57	7.86	17.74	5.93	14.8
snow	5.88	23.45	7.38	17.46	5.98	14.43
clean	5.85	30.82	7.41	17.74	6.09	15.79

On table 4.2., the benchmark results obtained using an AMD Ryzen 5 8645HS processor paired with an NVIDIA GeForce RTX 4050 Laptop GPU and CUDA Version 12.9, the inference performance of six object detection models was evaluated across nine weather conditions using a dataset of 60,000 images. Among the models tested—FastRCNNWrapper, FasterRCNN,

MaskRCNN, SSD, WrappedDETR, and WrappedYOLO—FasterRCNN consistently demonstrated the fastest inference times, completing tasks in under 45 minutes across all conditions, with its best performance under blur (32:00) and clean (32:27) scenarios. In contrast, FastRCNNWrapper and MaskRCNN exhibited significantly longer processing durations, often exceeding two hours. Prediction rates further reinforced FasterRCNN’s efficiency, achieving up to 31.24 images/sec under blur conditions, while SSD and WrappedYOLO also maintained competitive throughput, averaging around 17–18 images/sec. These results highlight FasterRCNN’s superior speed and responsiveness, making it particularly suitable for real-time applications under diverse environmental conditions.

Model Complexity

Table 4.3. Inference Time for each weather condition and RCNN models

Models	Total Params	Trainable Params	~Size (MB)
FastRCNNWrapper	41,161,351	40,938,951	157.02
FasterRCNN	19,386,354	19,327,458	73.95
MaskRCNN	44,401,393	44,178,993	169.38
SSD	35,641,826	35,603,106	135.96
WrappedDETR	41,524,768	41,302,368	158.4
WrappedYOLO	3,151,904	0	12.02

Table 4.3. presents a comparative overview of model complexity and memory requirements for the six object detection architectures evaluated. MaskRCNN and WrappedDETR exhibit the highest parameter counts and memory footprints, exceeding 158 MB, which reflects their computational intensity and potential for higher representational capacity. In contrast, WrappedYOLO stands out for its lightweight design, with only 3.15 million parameters

and a compact size of 12.02 MB, though it lacks trainable components in this configuration. FasterRCNN offers a balanced trade-off between model size and trainability, making it suitable for resource-constrained environments. These metrics are critical when considering deployment scenarios, especially in edge computing or real-time inference settings where memory and processing budgets are limited.

Denoising Results

Table 4.4. presents the average performance scores of six object detection models

Condition	Average of Recall	Average of F1	Average of Precision
Clean	0.27	0.33	0.67
Dark	0.24	0.31	0.70
Dark_denoised	0.25	0.32	0.70
Fog	0.19	0.25	0.66
Fog_denoised	0.22	0.30	0.69
Rain	0.20	0.27	0.67
Rain_denoised	0.24	0.31	0.68
Snow	0.17	0.24	0.67
Snow_denoised	0.21	0.28	0.67
Fog&rain	0.08	0.12	0.57
Fog&rain_denoised	0.22	0.30	0.69
Average	0.21	0.28	0.67

Table 4.4 presents the average performance scores of six object detection models Fast R-CNN, Faster R-CNN, Mask R-CNN, SSD, YOLO v8, and DETR under various weather conditions. For each weather type, corrupted images were first passed through a denoising model (e.g., snow removal, rain removal, fog removal). The denoised images were then used as inputs for all detection models, and the final values represent the averaged results across all models.

The table shows the Average Recall, Average F1-score, and Average Precision for each condition. Overall, the denoised versions (e.g., *Dark_denoised*, *Fog_denoised*, *Rain_denoised*, *Snow_denoised*) consistently achieve slightly higher scores than their original corrupted counterparts.

However, even after denoising, the scores remain lower than the clean condition, showing that weather corruption still has a noticeable impact on detection accuracy. The final row (“Average”) summarizes the mean values across all weather types, highlighting the overall degradation caused by weather effects and the benefits of denoising preprocessing.

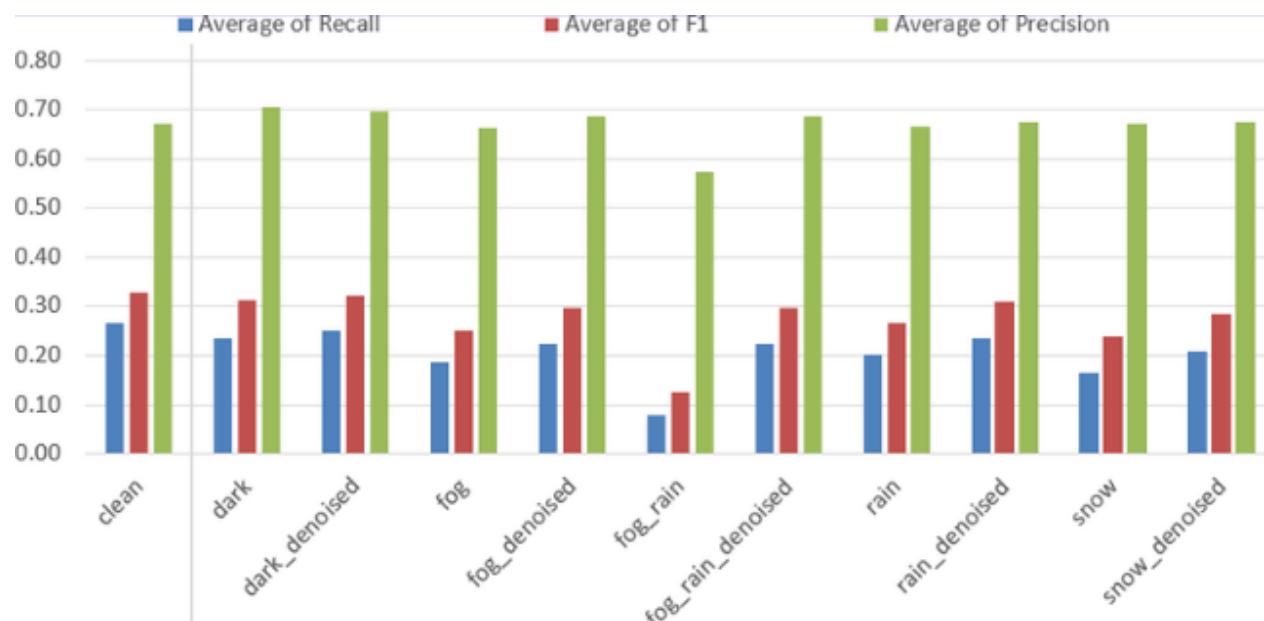


Fig. 4.6. The bar chart illustrates the average Recall, F1-score, and Precision of six object detection models under different weather conditions, both before and after denoising. The *clean* condition shows the highest overall performance, serving as a baseline. For corrupted conditions such as dark, fog, rain, and snow, performance drops noticeably—especially in Recall and F1.

Conclusion

This study evaluated object detection models across eight simulated weather hazards and three severity levels using the nuImages–nuScenes dataset. Mask R-CNN delivered the most balanced performance across precision, recall, F1 score, and IoU, demonstrating superior trade-offs between accurate classification and reliable localization. SSD and WrappedYOLO achieved the highest precision and IoU, making them strong choices when minimizing false positives and preserving localization accuracy is critical. DETR variants showed high recall but lower precision and IoU, indicating a tendency toward over-detection at the cost of spatial accuracy.

The analysis revealed that sandstorm and motion blur produced the largest drops in resilience across models, while heatwaves most strongly degraded IoU. Performance differences were object-class dependent, with vehicle categories generally retaining higher scores than pedestrians and bicycles. These nuanced patterns emphasize that robustness is conditional on both the model architecture and the specific environmental corruption.

Future work should prioritize real-world validation, adaptive model training with weather-aware augmentations, and multi-sensor fusion strategies to mitigate single-sensor weaknesses. Practical deployment will benefit from hybrid approaches that combine SSD- or YOLO-style localization strengths with Mask R-CNN’s balanced detection behavior to improve safety and reliability in adverse weather.

References

- [1] Xiaoshuai Hao, Guanqun Liu, Yuting Zhao, Yuheng Ji, Mengchuan Wei, Haimei Zhao, Lingdong Kong, Rong Yin, Yu Liu: *MSC-Bench: Benchmarking and Analyzing Multi-Sensor Corruption for Driving Perception*. CoRR abs/2501.01037 (2025)

Appendix

Table 4.1. Average Precision of RCNN Models Across Object Classes (From page #14)

Average of Precision		Object Classes				
Models	bicycle	bus	car	motorcycle	person	truck
FasterRCNN	0.61	0.35	0.78	0.78	0.60	0.55
FastRCNNWrapper	0.58	0.57	0.84	0.78	0.75	0.60
MaskRCNN	0.42	0.45	0.64	0.68	0.60	0.53
SSD	0.82	0.74	0.95	0.95	0.75	0.74
WrappedDETR	0.38	0.39	0.50	0.57	0.41	0.29
WrappedYOLO	0.81	0.54	0.93	0.89	0.84	0.73
Average Precision	0.60	0.51	0.77	0.77	0.66	0.57

Table 6.1. Average Recall of RCNN Models Across Object Classes

Average of Recall		Object Classes				
Models	bicycle	bus	car	motorcycle	person	truck
FasterRCNN	0.10	0.13	0.21	0.15	0.07	0.10
FastRCNNWrapper	0.26	0.16	0.32	0.26	0.15	0.19
MaskRCNN	0.36	0.25	0.56	0.35	0.36	0.28
SSD	0.05	0.10	0.18	0.08	0.05	0.08
WrappedDETR	0.38	0.28	0.58	0.36	0.38	0.35
WrappedYOLO	0.10	0.12	0.25	0.10	0.12	0.08
Average Recall	0.21	0.17	0.35	0.22	0.19	0.18

Table 6.2. Average IOU of RCNN Models Across Object Classes

Average of IoU		Object Classes					
Models		bicycle	bus	car	motorcycle	person	truck
FasterRCNN		0.74	0.86	0.79	0.71	0.72	0.81
FastRCNNWrapper		0.76	0.85	0.83	0.72	0.80	0.81
MaskRCNN		0.75	0.86	0.80	0.71	0.77	0.81
SSD		0.78	0.89	0.86	0.73	0.75	0.85
WrappedDETR		0.74	0.85	0.77	0.71	0.73	0.80
WrappedYOLO		0.82	0.88	0.87	0.71	0.82	0.90
Average IoU		0.76	0.87	0.82	0.71	0.76	0.83

Table 6.3. Average Precision of RCNN Models Across Weather Hazard Conditions Grouped by Object Class

Average of Precision			Weather Condition						
Object Class & Models	blur	clean	dark	fog	heatwave	motion_blur	rain	sandstorm	snow
bicycle	0.60	0.59	0.59	0.62	0.57	0.61	0.59	0.61	0.63
FasterRCNN	0.58	0.61	0.58	0.64	0.53	0.62	0.61	0.70	0.65
FastRCNNWrapper	0.57	0.58	0.58	0.60	0.54	0.57	0.57	0.59	0.59
MaskRCNN	0.43	0.40	0.41	0.43	0.36	0.51	0.40	0.43	0.43
SSD	0.86	0.82	0.81	0.83	0.82	0.74	0.84	0.81	0.87
WrappedDETR	0.35	0.32	0.36	0.41	0.35	0.39	0.35	0.40	0.41
WrappedYOLO	0.83	0.80	0.80	0.78	0.81	0.82	0.80	0.76	0.84
bus	0.55	0.55	0.59	0.49	0.56	0.45	0.46	0.42	0.50
FasterRCNN	0.40	0.40	0.42	0.35	0.34	0.34	0.30	0.30	0.30
FastRCNNWrapper	0.58	0.60	0.66	0.56	0.59	0.45	0.49	0.61	0.64
MaskRCNN	0.46	0.49	0.58	0.41	0.45	0.38	0.39	0.39	0.52
SSD	0.78	0.78	0.79	0.76	0.77	0.68	0.76	0.63	0.70
WrappedDETR	0.49	0.47	0.52	0.29	0.47	0.38	0.37	0.20	0.35
WrappedYOLO	0.58	0.59	0.60	0.55	0.73	0.46	0.47	0.39	0.52
car	0.78	0.76	0.78	0.76	0.73	0.79	0.78	0.78	0.78
FasterRCNN	0.76	0.78	0.78	0.78	0.75	0.76	0.80	0.82	0.81
FastRCNNWrapper	0.83	0.85	0.87	0.82	0.81	0.82	0.87	0.82	0.86
MaskRCNN	0.66	0.61	0.67	0.61	0.55	0.69	0.65	0.67	0.60
SSD	0.95	0.96	0.96	0.95	0.95	0.96	0.96	0.96	0.95

Average of Precision				Weather Condition					
Object Class & Models	blur	clean	dark	fog	heatwave	motion_blur	rain	sandstorm	snow
WrappedDETR	0.54	0.46	0.50	0.46	0.44	0.62	0.48	0.46	0.50
WrappedYOLO	0.93	0.93	0.93	0.92	0.91	0.92	0.93	0.95	0.93
motorcycle	0.78	0.79	0.80	0.81	0.75	0.73	0.79	0.77	0.75
FasterRCNN	0.79	0.79	0.82	0.81	0.67	0.71	0.81	0.85	0.75
FastRCNNWrapper	0.76	0.80	0.81	0.81	0.76	0.71	0.80	0.80	0.76
MaskRCNN	0.65	0.70	0.72	0.71	0.65	0.63	0.70	0.75	0.64
SSD	0.95	0.95	0.96	0.97	0.93	0.92	0.95	0.98	0.94
WrappedDETR	0.57	0.57	0.58	0.60	0.57	0.55	0.55	0.60	0.51
WrappedYOLO	0.93	0.94	0.94	0.94	0.92	0.86	0.94	0.63	0.92
person	0.68	0.68	0.68	0.69	0.55	0.65	0.67	0.69	0.67
FasterRCNN	0.64	0.65	0.64	0.65	0.37	0.61	0.64	0.63	0.64
FastRCNNWrapper	0.74	0.78	0.77	0.78	0.69	0.69	0.78	0.77	0.78
MaskRCNN	0.60	0.62	0.62	0.65	0.50	0.60	0.62	0.65	0.59
SSD	0.80	0.78	0.77	0.78	0.67	0.69	0.77	0.79	0.75
WrappedDETR	0.46	0.40	0.40	0.42	0.39	0.49	0.37	0.38	0.38
WrappedYOLO	0.86	0.87	0.87	0.87	0.68	0.84	0.86	0.90	0.87
truck	0.59	0.57	0.58	0.59	0.56	0.61	0.52	0.60	0.54
FasterRCNN	0.59	0.58	0.55	0.56	0.52	0.66	0.48	0.53	0.49
FastRCNNWrapper	0.61	0.61	0.61	0.61	0.56	0.63	0.57	0.64	0.58
MaskRCNN	0.56	0.52	0.54	0.56	0.48	0.59	0.48	0.57	0.50
SSD	0.77	0.72	0.74	0.75	0.76	0.71	0.71	0.74	0.74
WrappedDETR	0.32	0.30	0.31	0.31	0.29	0.28	0.24	0.29	0.27
WrappedYOLO	0.72	0.71	0.72	0.78	0.74	0.80	0.64	0.83	0.66
Average Precision	0.66	0.66	0.67	0.66	0.62	0.64	0.64	0.65	0.64

Table 6.4. Average Recall of RCNN Models Across Weather Hazard Conditions Grouped by Object Class

Average of Recall		Weather Condition								
Object Class & Models		blur	clean	dark	fog	heatwave	motion_blur	rain	sandstorm	snow
bicycle		0.20	0.28	0.27	0.24	0.21	0.12	0.24	0.17	0.21
FasterRCNN		0.11	0.13	0.13	0.12	0.08	0.06	0.11	0.08	0.09
FastRCNNWrapper		0.26	0.34	0.32	0.30	0.26	0.16	0.29	0.21	0.28
MaskRCNN		0.32	0.49	0.46	0.41	0.36	0.18	0.42	0.29	0.39
SSD		0.05	0.06	0.07	0.05	0.06	0.05	0.06	0.03	0.05
WrappedDETR		0.34	0.53	0.50	0.42	0.41	0.19	0.42	0.31	0.37
WrappedYOLO		0.11	0.13	0.14	0.11	0.09	0.06	0.12	0.08	0.09
bus		0.17	0.23	0.20	0.19	0.14	0.15	0.20	0.12	0.18
FasterRCNN		0.14	0.16	0.14	0.14	0.10	0.13	0.15	0.09	0.13
FastRCNNWrapper		0.16	0.22	0.18	0.18	0.12	0.13	0.20	0.11	0.15
MaskRCNN		0.25	0.34	0.28	0.29	0.21	0.20	0.29	0.18	0.25
SSD		0.11	0.13	0.12	0.09	0.09	0.10	0.12	0.06	0.11
WrappedDETR		0.25	0.37	0.32	0.33	0.23	0.22	0.32	0.22	0.31
WrappedYOLO		0.13	0.16	0.14	0.11	0.07	0.11	0.14	0.07	0.12
car		0.35	0.40	0.38	0.37	0.36	0.32	0.36	0.28	0.35
FasterRCNN		0.24	0.23	0.23	0.22	0.21	0.23	0.20	0.16	0.18
FastRCNNWrapper		0.33	0.34	0.33	0.34	0.33	0.31	0.31	0.28	0.32
MaskRCNN		0.54	0.65	0.61	0.60	0.58	0.48	0.59	0.46	0.59
SSD		0.19	0.21	0.20	0.18	0.20	0.16	0.20	0.13	0.20
WrappedDETR		0.56	0.67	0.64	0.62	0.61	0.48	0.61	0.49	0.58
WrappedYOLO		0.27	0.29	0.28	0.25	0.26	0.24	0.25	0.16	0.24
motorcycle		0.23	0.27	0.25	0.21	0.24	0.18	0.22	0.14	0.24
FasterRCNN		0.17	0.18	0.17	0.15	0.15	0.14	0.14	0.10	0.13
FastRCNNWrapper		0.27	0.31	0.29	0.27	0.29	0.21	0.26	0.19	0.29
MaskRCNN		0.36	0.45	0.40	0.35	0.38	0.27	0.36	0.23	0.40
SSD		0.09	0.10	0.09	0.07	0.09	0.07	0.09	0.04	0.09
WrappedDETR		0.37	0.46	0.44	0.35	0.41	0.29	0.37	0.26	0.40
WrappedYOLO		0.12	0.14	0.13	0.07	0.10	0.09	0.10	0.04	0.11

Average of Recall		Weather Condition							
Object Class & Models	blur	clean	dark	fog	heatwave	motion.blur	rain	sandstorm	snow
person	0.19	0.24	0.23	0.20	0.18	0.12	0.22	0.15	0.21
FasterRCNN	0.08	0.09	0.09	0.08	0.07	0.07	0.08	0.05	0.07
FastRCNNWrapper	0.16	0.16	0.15	0.15	0.16	0.13	0.15	0.12	0.15
MaskRCNN	0.35	0.48	0.45	0.39	0.34	0.21	0.43	0.30	0.41
SSD	0.05	0.07	0.06	0.05	0.06	0.04	0.07	0.03	0.06
WrappedDETR	0.35	0.49	0.46	0.41	0.37	0.22	0.45	0.31	0.42
WrappedYOLO	0.14	0.16	0.15	0.13	0.11	0.09	0.14	0.08	0.12
truck	0.18	0.23	0.22	0.19	0.18	0.13	0.20	0.13	0.19
FasterRCNN	0.11	0.12	0.12	0.10	0.10	0.07	0.10	0.08	0.09
FastRCNNWrapper	0.19	0.24	0.23	0.20	0.19	0.14	0.21	0.13	0.21
MaskRCNN	0.27	0.36	0.33	0.29	0.29	0.19	0.31	0.20	0.32
SSD	0.09	0.11	0.10	0.08	0.09	0.07	0.10	0.05	0.08
WrappedDETR	0.35	0.44	0.42	0.37	0.37	0.25	0.39	0.25	0.35
WrappedYOLO	0.10	0.11	0.10	0.07	0.06	0.05	0.09	0.05	0.09
Average Recall	0.22	0.28	0.26	0.23	0.22	0.17	0.24	0.17	0.23

Detailed per-class analysis revealed that performance varied significantly across object categories and weather types. For each condition, the highest scores were highlighted, with green indicating top-performing classes. Notably, the nuImages dataset lacked annotations for traffic lights, which were excluded from analysis.

Table. 6.5. Average Precision of RCNN Models Across Weather Hazard Conditions

Average Precision		Weather Condition								
Models		blur	clean	dark	fog	heatwave	motion_blur	rain	sandstorm	snow
FasterRCNN		0.63	0.63	0.63	0.63	0.53	0.61	0.61	0.64	0.61
FastRCNNWrapper		0.68	0.70	0.71	0.70	0.66	0.65	0.68	0.71	0.70
MaskRCNN		0.56	0.56	0.59	0.56	0.50	0.57	0.54	0.58	0.55
SSD		0.85	0.83	0.84	0.84	0.82	0.78	0.83	0.82	0.82
WrappedDETR		0.46	0.42	0.44	0.41	0.42	0.45	0.39	0.39	0.40
WrappedYOLO		0.81	0.80	0.81	0.81	0.80	0.79	0.77	0.74	0.79
Average Precision		0.66	0.66	0.67	0.66	0.62	0.64	0.64	0.65	0.64

In Table 4.3, the average precision of six RCNN-based object detection models was evaluated across nine simulated weather hazard conditions to assess their robustness under environmental perturbations. Among the models, SSD consistently achieved the highest average precision across all conditions, peaking at 0.85 under blur and maintaining strong performance even under challenging scenarios such as heatwave (0.82) and motion blur (0.78). WrappedYOLO also demonstrated high resilience, with precision values ranging from 0.74 (sandstorm) to 0.81 (blur and dark). In contrast, WrappedDETR exhibited the lowest performance, with average precision scores fluctuating between 0.39 and 0.46, indicating a pronounced sensitivity to visual distortions. FastRCNNWrapper outperformed both FasterRCNN and MaskRCNN, particularly under dark (0.71) and sandstorm (0.71) conditions. Notably, all models experienced a dip in performance during heatwave and motion blur scenarios, with MaskRCNN and FasterRCNN showing the most significant degradation. Overall, the results highlight SSD and WrappedYOLO as the most robust detectors under adverse weather, while

WrappedDETR and MaskRCNN lag behind, emphasizing the need for further optimization of transformer-based and segmentation-focused architectures in challenging visual environments.

Hazards

Blur



Dark



Fog



Heatwave



Motion Blur



Rain



Sandstorm



Snow



Experiment result sheets



https://docs.google.com/spreadsheets/d/1INV8FyaV7iyZEO8r7NvHRmL8Dj9bSHLr/edit?usp=s_haring&ouid=104928440836170381911&rtpof=true&sd=true