Cyberattack Resilience of Autonomous Vehicle Sensor Systems: Evaluating RGB vs. Dynamic Vision Sensors in CARLA

Mustafa Sakhai *, Kaung Sithu, Min Khant Soe Oke, Maciej Wielgosz *

Posted Date: 28 May 2025

doi: 10.20944/preprints202505.2262.v1

Keywords: autonomous vehicles; cybersecurity attacks; dynamic vision sensor

 doi:10.20944/preprints202505.2262.v1

Article

Cyberattack Resilience of Autonomous Vehicle Sensor Systems: Evaluating RGB vs. Dynamic Vision Sensors in CARLA

Mustafa Sakhai [1,†], **Kaung Sithu** [1,†,‡], **Min Khant Soe Oke** [1,†,‡], **and Maciej Wielgosz** [1,2]

[1] AGH University of Science and Technology, Krakow, Poland; msakhai@agh.edu.pl (M.S.); sithu@student.agh.edu.pl (K.S.); oke@student.agh.edu.pl (M.O.); wielgosz@agh.edu.pl (M.W.)
[2] Academic Computer Centre AGH, Krakow, Poland; wielgosz@agh.edu.pl (M.W.)
* Correspondence: wielgosz@agh.edu.pl (M.W.)
† Current address: Affiliation.
‡ These authors contributed equally to this work.

Abstract: Autonomous Vehicles (AVs) rely on a heterogeneous sensor suite RGB cameras, LiDAR, GPS/IMU, and emerging event-based Dynamic Vision Sensors (DVS) to perceive and navigate complex environments. However, these sensors can be deceived by realistic cyberattacks, undermining safety. In this work, we systematically implement seven attack vectors in the CARLA simulator, salt and pepper noise, event flooding, depth map tampering, LiDAR phantom injection, GPS spoofing, denial of service, and steering bias control, and measure their impact on a state-of-the-art end to end driving agent. We then equip each sensor with tailored defenses (e.g., adaptive median filtering for RGB, spatial clustering for DVS) and integrate a semi-supervised anomaly detector (EfficientAD from anomalib) trained exclusively on benign data. Our detector achieves clear separation between normal and attacked conditions (mean RGB anomaly scores of 0.00 vs. 0.38; DVS: 0.61 vs. 0.76), yielding over 95% detection accuracy with fewer than 5% false positives. Defense evaluations reveal that GPS spoofing is fully mitigated, whereas RGB and depth based attacks still induce 30–45% trajectory drift despite filtering. Notably, DVS sensors exhibit greater intrinsic resilience in high dynamic range scenarios, though their asynchronous output necessitates carefully tuned thresholds. These findings underscore the critical role of multi-modal anomaly detection and demonstrate that possibility of integrating DVS alongside conventional sensors significantly strengthens AV cybersecurity.

Keywords: autonomous vehicles; cybersecurity attacks; dynamic vision sensor

1. Introduction

Autonomous vehicles (AVs) have the potential to revolutionize transportation systems worldwide [1]. These vehicles rely on an array of sensors such as RGB cameras, LiDAR, radar, GPS, and emerging event-based dynamic vision sensors (DVSs) to perceive their surroundings and navigate autonomously [2]. However, increasing complexity and connectivity of AV systems introduce significant cybersecurity risks that could be exploited by malicious actors [3,14]. The security of sensor systems is paramount, as they form the critical interface between the vehicle and its environment [1]. Recent research has shown that adversaries can manipulate the sensor input, posing serious safety threats [3,11,14].

For example, studies have shown that camera-based perception can be compromised by adversarial examples, causing misidentification of traffic signs or obstacles [4,11]. Similarly, LiDAR sensors can be spoofed to generate false obstacles, which can lead to abrupt and unnecessary braking [5,12,13]. These vulnerabilities highlight the pressing need for thorough security analysis and robust defense strategies in all types of AV sensor [3,14].

While traditional sensors such as RGB cameras and LiDAR have received significant attention regarding their security, the integration of dynamic vision sensors (DVSs) into AVs introduces both new opportunities and challenges that are yet to be fully explored. Unlike conventional frame-based

 doi:10.20944/preprints202505.2262.v1

cameras, DVS operate by asynchronously detecting pixel-level brightness changes with microsecond precision [2,6]. This event-driven approach provides benefits such as high dynamic range, reduced motion blur, and lower redundancy of data [2,7,15], making DVS ideal for high-speed driving and challenging weather conditions. However, their unique operating principles prompt critical security questions: How resilient is DVS to adversarial attacks? Can their event-based nature be exploited? What defenses are best suited to secure event-based perception?

This paper conducts a detailed analysis of cyberattacks targeting AV sensors, with a particular focus on the security implications of incorporating DVS into the sensor suite. We explore a range of attack vectors, including adversarial machine learning attacks, sensor spoofing, denial of service (DoS), and steering bias attacks, targeting RGB cameras, LiDAR, GPS, and DVS sensors. Leveraging the CARLA simulator and the PCLA (Pretrained CARLA Leaderboard Agent) framework [8], we simulate and assess these attacks within realistic autonomous driving environments, building on approaches like those in [2] for DVS-based pedestrian detection under adverse weather. The CARLA simulator offers a powerful tool for testing decision-making logic and evaluating camera-based perception through photorealistic scenarios [9].

Our study examines the agent's behavior and sensor responses under various attack conditions, offering insights into the vulnerabilities and resilience of autonomous driving systems [1,3,16].

- We provide a systematic analysis of sensor vulnerabilities in autonomous vehicles, with particular focus on the emerging DVS technology and its unique security characteristics.
- We implement and evaluate realistic attack scenarios in the CARLA simulator, demonstrating their impact on autonomous agent behavior and perception accuracy.
- We develop an effective anomaly detection system using the anomalib library that can identify sophisticated attacks on both RGB and DVS camera sensors [16].
- We propose a comprehensive defense framework that combines anomaly detection, sensor fusion, and potential human intervention to ensure resilient autonomous operation.

2. Materials and Methods

This section describes the experimental setup, methodology, and implementation details used to evaluate the resilience of autonomous vehicles against cyberattacks, with a particular focus on Dynamic Vision Sensors (DVS) and RGB cameras. We provide comprehensive information to enable reproducibility of our experiments and findings.

2.1. Experimental Framework

2.1.1. CARLA Simulator

All experiments were conducted using the CARLA simulator (version 9.15), an open-source simulator for autonomous driving research that provides a realistic urban environment with detailed physics and sensor models. CARLA was chosen for its ability to accurately simulate various sensor modalities including RGB cameras, LiDAR, radar, and event-based sensors. The simulator was run on a Linux Ubuntu 22.04 system with CUDA-enabled GPUs to support the computational requirements of the autonomous driving agents and sensor processing pipelines.

2.1.2. PCLA Framework

We utilized the Pretrained CARLA Leaderboard Agent (PCLA) framework, which enables the deployment of autonomous driving agents from the CARLA Leaderboard onto vehicles within the simulator. This framework provides several advantages for our experimental setup:

- Ability to deploy multiple autonomous driving agents with different architectures and training paradigms
- Independence from the Leaderboard codebase, allowing compatibility with the latest CARLA version
- Support for multiple vehicles with different autonomous agents operating simultaneously

 doi:10.20944/preprints202505.2262.v1

The PCLA framework includes nine different high-performing autonomous driving agents trained with 17 distinct training seeds, allowing us to evaluate the impact of cyberattacks across a diverse set of perception and control algorithms.

2.1.3. Autonomous Driving Agent

For our experiments, we focused exclusively on the following autonomous driving agent:

- **NEAT AIM-MT-2D**: A neural attention-based end-to-end autonomous driving agent that processes RGB images and depth information to predict vehicle controls directly. We used the NEAT variant that incorporates depth information (neat_aim2ddepth), which enhances the agent's ability to perceive the 3D structure of the environment and improves its performance in complex driving scenarios.

This agent represents an end-to-end learning approach to autonomous driving that directly maps sensor inputs to control commands, allowing us to assess the impact of attacks on a unified perception-control architecture.

2.2. Sensor Configuration

Our experimental setup incorporated a comprehensive sensor suite to evaluate the resilience of autonomous vehicles against various attack vectors:

- **RGB Cameras**: Front-facing RGB camera with 800×600 resolution and 100° field of view, mounted at position (1.3, 0.2, 2.3) relative to the vehicle center, providing visual input for the agent's perception system.
- **Dynamic Vision Sensor (DVS)**: Event-based camera with 800×600 resolution and 100° field of view, mounted at position (1.3, 0.0, 2.3), with positive and negative thresholds set to 0.3 to capture brightness changes in the scene with microsecond temporal resolution.
- **Depth Camera**: Depth sensing camera with 800×600 resolution and 100° field of view, aligned with the RGB camera position, providing per-pixel distance measurements for 3D scene understanding.
- **LiDAR**: 64-channel LiDAR sensor with 85m range, 600,000 points per second, and 10Hz rotation frequency, mounted at position (0.0, 0.0, 2.5), providing detailed 3D point cloud data of the surrounding environment.
- **GPS and IMU**: For localization and vehicle state estimation, enabling the agent to maintain awareness of its position and orientation within the environment.

It is important to note that in this experimental setup, the DVS cameras and LIDARs were primarily used for research purposes and data collection, and were not directly integrated into the vehicle's driving decision model.

Sensor data was collected at 10Hz to match the control frequency of the autonomous driving agent. All sensor data was synchronized using the CARLA simulator's internal clock to ensure temporal consistency across modalities. This comprehensive sensor configuration allowed us to evaluate the impact of attacks on individual sensors as well as the effectiveness of multi-sensor fusion for attack detection and mitigation.

2.3. Attack Implementation

We implemented and evaluated several types of cyberattacks targeting the sensor inputs of autonomous vehicles, focusing on realistic attack vectors that could be executed in real-world scenarios:

- **RGB Camera Attacks, Salt-and-Pepper Noise**: We implemented a high-density (80%) salt-and-pepper noise attack on RGB camera images, where random pixels were replaced with either black (0) or white (255) values with equal probability. This attack simulates severe sensor interference that could result from electromagnetic interference or hardware tampering.
- **Dynamic Vision Sensor Attacks, Event Flooding**: We implemented a noise injection attack that added a large number of spurious events (approximately 60% of the frame size) with random

 doi:10.20944/preprints202505.2262.v1

positions and polarities to the DVS output stream. This attack creates false motion perception that could confuse event-based perception algorithms and trigger unnecessary emergency responses.

- **Depth Camera Attacks, Depth Map Tampering**: We implemented a patch-based depth tampering attack that modified depth values in random regions of the depth map. The attack added 5 random patches of size 50×50 pixels with depth offsets ranging from 5 to 15 units, creating the illusion of obstacles closer or further than they actually are.
- **LiDAR Attacks, Phantom Object Injection**: We implemented a phantom object injection attack that added 5 clusters of points (100 points per cluster) to the LiDAR point cloud at random positions within a range of 5-15 meters in front of the vehicle. These phantom objects could trigger unnecessary braking or evasive maneuvers in the autonomous driving system.
- **GPS Spoofing Attacks, Position Manipulation**: We implemented a GPS spoofing attack that maintains two separate location records: the actual position (vehicle._original_get_transform()) and the perceived position (vehicle.get_transform()). This allows us to analyze the discrepancy between the vehicle's true and perceived locations during attacks. The attack creates a significant deviation between the reported and actual position, potentially causing the vehicle to make incorrect routing decisions.
- **Denial of Service (DoS) Attacks, Sensor Rate Limiting**: We implemented a DoS attack that targets the sensor update frequency by tracking sensor update timestamps (attack_state.last_sensor_update) and artificially restricting updates to a maximum of 20 Hz. This causes delayed or missed sensor readings that affect the vehicle's perception and decision-making capabilities.
- **Steering Bias Attacks, Control Manipulation**: We implemented a steering bias attack that introduces systematic errors into the vehicle's steering commands. The attack modifies the steering values in the control data stream, and its effects are measured through trajectory deviation analysis and steering angle distribution statistics.

All attacks were implemented by intercepting and modifying the sensor data in the callback functions before it was processed by the autonomous driving agent. This approach simulates a realistic attack scenario where an adversary has gained access to the sensor data stream but not necessarily to the internal processing algorithms of the autonomous system.

2.4. Defense Integration

To protect against the implemented attacks, we developed and integrated a multi-layered defense system that combines sensor-specific filtering techniques with anomaly detection:

- **RGB Camera Defenses, Adaptive Median Filter**: We implemented a decision-based adaptive median filter that dynamically adjusts its kernel size (3×3 to 7×7) based on the detected noise level. The filter specifically targets salt-and-pepper noise by:
  - Identifying potential noise pixels (values of 0 or 255)
  - Growing the filter window until valid median values are found
  - Preserving edge details by only replacing confirmed noise pixels
- **DVS Defenses, Event Stream Analysis**: We implemented a two-stage defense mechanism:
  - Event count monitoring with a threshold of 5000 events per frame
  - Spatial clustering analysis using KD-trees to detect abnormally dense event clusters
  - Morphological operations (dilation) to smooth legitimate events while filtering noise
- **Depth Camera Defenses, Gradient-Based Tampering Detection**: Our defense system includes:
  - Range limiting (0-100m) to filter physically impossible measurements
  - Sobel gradient analysis to detect unnatural depth discontinuities
  - Adaptive Gaussian smoothing based on detected tampering severity
- **LiDAR Defenses, Point Cloud Filtering**: We implemented a comprehensive filtering pipeline:

 doi:10.20944/preprints202505.2262.v1

  - Distance-based outlier removal (50m maximum range)
  - Density-based clustering to identify and remove phantom objects
  - Nearest neighbor analysis to detect unnaturally dense point clusters
- **Rate Limiting Defense, Sensor Update Monitoring**: To prevent DoS attacks:
  - Tracking of sensor update timestamps
  - Implementation of rate limiting thresholds
  - Buffering mechanism to maintain consistent sensor data flow

These defenses were implemented in the sensor callback functions, allowing for real-time protection against attacks while maintaining the performance requirements of the autonomous driving system. The effectiveness of each defense mechanism was evaluated through extensive testing under various attack scenarios.

2.5. Anomaly Detection System

We created an anomaly detection system to identify potential sensor attacks using the anomalib library, which provides state-of-the-art deep learning models for anomaly detection. After evaluating several models, we found that the EfficientAd model consistently outperformed others for our specific use case.

2.5.1. EfficientAd Model Architecture

The EfficientAd model was selected as our primary anomaly detection approach due to its superior performance in detecting sensor attacks. EfficientAd is a student-teacher framework that identifies anomalies based on the discrepancy between teacher and student predictions. The model consists of:

- A teacher network pre-trained on normal data that learns to extract robust features from sensor inputs
- A student network that attempts to mimic the teacher's feature representations
- A discrepancy measurement module that quantifies the difference between teacher and student outputs

This architecture is particularly effective for sensor attack detection as it can identify subtle deviations from normal sensor behavior without requiring examples of attacks during training.

2.5.2. Training Methodology

The EfficientAd model was trained using a semi-supervised approach:

1. **Data Collection**: We collected sensor data from multiple driving scenarios without attacks to establish a baseline of normal operation.
2. **Data Preprocessing**: For sensor modality, appropriate preprocessing was applied:
   - RGB and DVS data: Resized to 128×128 and normalized using ImageNet mean and standard deviation values (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]).
3. **Model Training**: The EfficientAd model was trained exclusively on normal data, learning to represent the distribution of normal sensor readings without exposure to attack examples.

Training was performed using PyTorch with CUDA acceleration on NVIDIA GPUs. We used the default parameters provided by the anomalib library for the EfficientAd model, which has been optimized for anomaly detection tasks. The model was trained for 30 epochs with batch size 1 and gradient accumulation over 4 batches to effectively increase the batch size while maintaining memory efficiency.

2.6. Evaluation Methodology

We evaluated both the impact of attacks on autonomous driving performance and the effectiveness of our defense mechanisms using a comprehensive set of metrics and visualization tools:

 doi:10.20944/preprints202505.2262.v1

2.6.1. Attack Impact Assessment

The impact of attacks was measured using the following metrics:

- **Trajectory Analysis**: We track and compare the vehicle's actual trajectory against the planned route, calculating point-to-segment distances to measure route deviation. This analysis is particularly important for GPS spoofing attacks where perceived and actual positions may differ significantly.
- **Control Stability**: We analyze steering, throttle, and brake commands through detailed time-series analysis. For steering bias attacks, we perform statistical analysis of steering angle distributions to detect anomalous patterns.
- **Sensor Performance Metrics**:
  - RGB Camera: Noise percentage measurements during salt-and-pepper attacks
  - DVS: Event count tracking to detect abnormal spikes in event generation
  - Depth Camera: Mean depth measurements to identify tampering
  - LiDAR: Point cloud density analysis to detect phantom objects
- **Defense Effectiveness**: Comparison of performance metrics with and without defensive measures enabled, including:
  - Adaptive median filtering for RGB noise
  - Spatial clustering analysis for DVS event flooding
  - Gradient-based analysis for depth tampering
  - Density-based filtering for LiDAR attacks
  - Rate limiting for DoS attacks

2.6.2. Visualization and Analysis Tools

We developed several visualization tools to analyze attack impacts:

- **Combined Video Generation**: Synchronized display of multiple sensor feeds (RGB, DVS, depth, LiDAR) with attack state indicators
- **Trajectory Plots**: Visualization of actual vs. perceived trajectories during GPS spoofing
- **Statistical Analysis**: Histograms of steering distributions and sensor metrics under different attack conditions

2.6.3. Experimental Scenarios

We evaluated performance across diverse driving scenarios:

- **Urban Navigation**: Complex urban environments with intersections, traffic lights, and other vehicles.
- **Highway Driving**: High-speed scenarios with lane changes and overtaking maneuvers.
- **Dynamic Obstacles**: Scenarios with pedestrians and other vehicles executing unpredictable maneuvers.

Each scenario was tested with and without attacks to establish baseline performance and measure the impact of different attack types and intensities.

All experiments were conducted using Python 3.7 with PyTorch and the anomalib library. The CARLA simulator was run on a system with an Intel Core i9-10900K CPU @ 3.70GHz, 62GB RAM, and an NVIDIA GeForce RTX 3080 GPU with 10GB VRAM.

3. Results

This section presents the results of our comprehensive evaluation of cyberattacks targeting autonomous vehicle sensors and the effectiveness of our proposed defense mechanisms, based on over 160 experimental episodes conducted in the CARLA simulator. We organize our findings into three main categories: (1) the impact of various attack vectors on autonomous driving performance, (2) the

 doi:10.20944/preprints202505.2262.v1

effectiveness of our defense mechanisms in mitigating these attacks, and (3) the performance of our anomaly detection system for identifying malicious sensor manipulations.

3.1. Impact of Cyberattacks on Autonomous Driving Performance

We evaluated the resilience of autonomous vehicles against various attack vectors by analyzing their impact on driving performance, trajectory adherence, and sensor reliability. Our experiments revealed two distinct outcomes: scenarios where the vehicle completed the route despite attacks (albeit with performance degradation) and scenarios where attacks caused route abandonment or crashes.

3.1.1. Completed Routes Under Attack

In some scenarios, the autonomous vehicle demonstrated resilience by completing the designated route despite being subjected to cyberattacks. Episode 7 provides a notable example of this behavior, as illustrated in Figure 1.

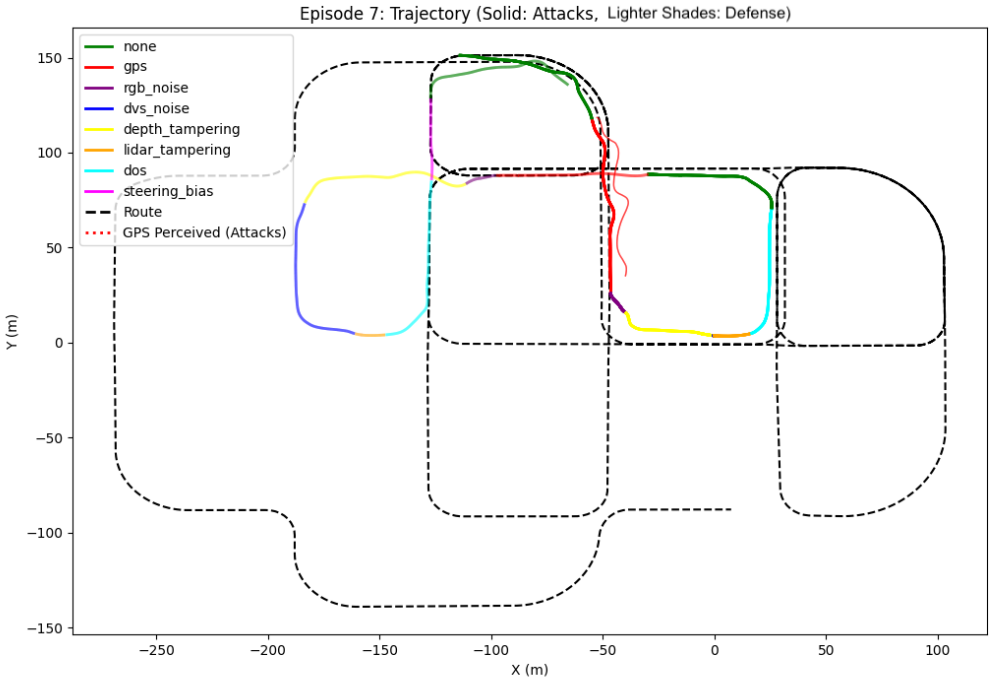

Figure 1. Episode 7 trajectory showing route completion despite attacks. Note the trajectory deviations during RGB and depth sensor attacks compared to normal operation.

While the vehicle successfully completed the route in Episode 7, the trajectory analysis reveals significant deviations from the optimal path during attacks. RGB camera attacks caused the most pronounced lateral deviations, with the vehicle weaving across lanes before correcting its course. Depth sensor attacks similarly disrupted the vehicle's ability to maintain a smooth trajectory, though to a lesser extent than RGB attacks. These deviations highlight the vehicle's compromised perception capabilities even when it manages to reach its destination.

Figure 2 shows the corresponding sensor outputs during normal operation and under different attack conditions for Episode 7. The RGB camera feed exhibits significant salt-and-pepper noise during attacks, severely degrading visual perception. Similarly, the depth sensor distorted distance measurements, potentially causing the autonomous system to misinterpret obstacle distances and road boundaries.

 doi:10.20944/preprints202505.2262.v1

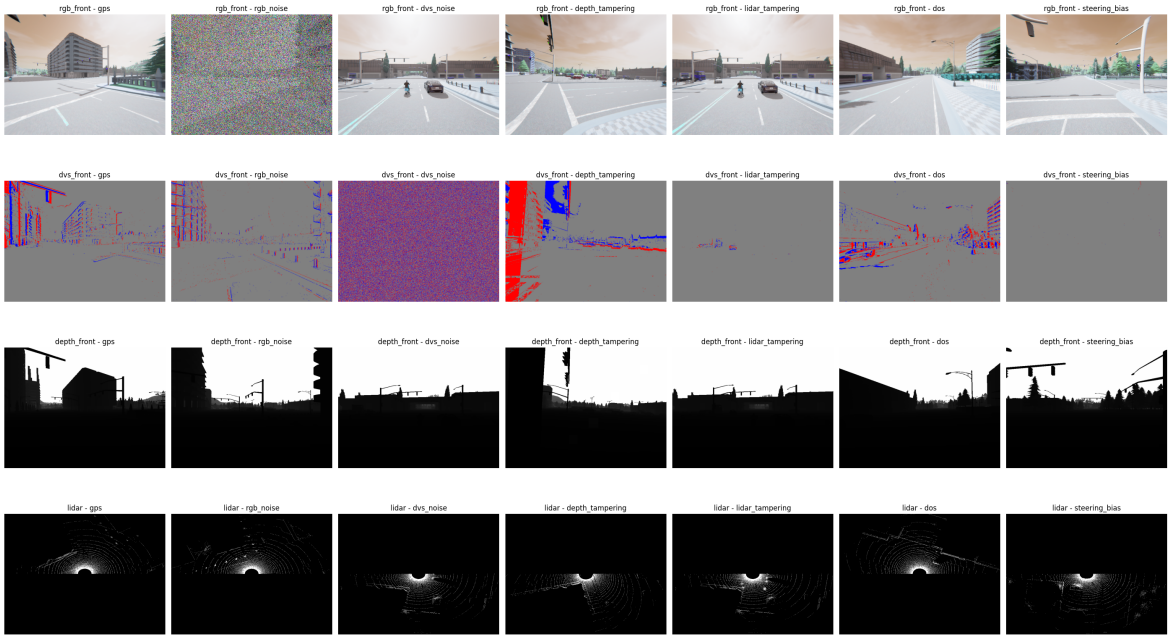

Figure 2. Sensor outputs without defense mechanisms during attacks in Episode 7. Note the severe degradation of RGB and DVS sensor data.

Episode 26 provides another example of route completion despite attacks, as shown in Figure 3. In this case, the vehicle deviated visibly from its intended route during RGB camera attacks but still successfully reached its destination. This highlights a nuanced vulnerability where the vehicle can complete its mission despite taking a potentially dangerous path along a simple pathway.

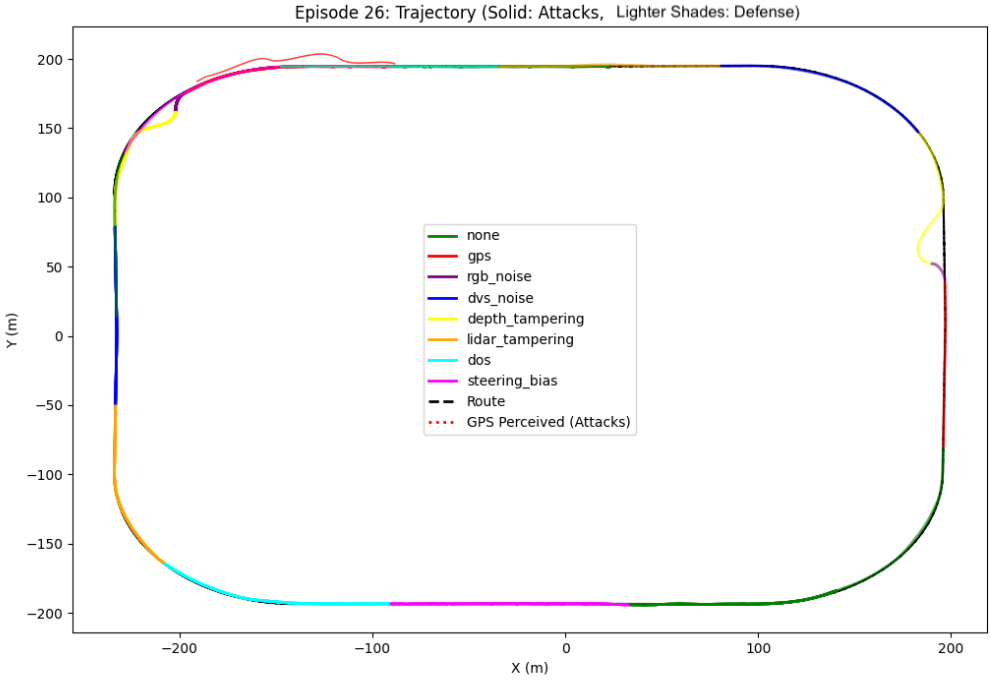

Figure 3. Episode 26 trajectory showing visible route deviation during RGB camera attacks, though the vehicle still successfully completed its route.

 doi:10.20944/preprints202505.2262.v1

3.1.2. Route Failures and Crashes

In contrast to the resilience observed in Episodes 7 and 26, several scenarios demonstrated significant failures where attacks caused the vehicle to crash.

Episode 5 demonstrates an even more severe failure mode, where RGB camera attacks caused the vehicle to move erratically before ultimately crashing into a roadside fence. Figure 4 shows the vehicle's trajectory during this episode, with the abrupt termination point indicating the crash location.

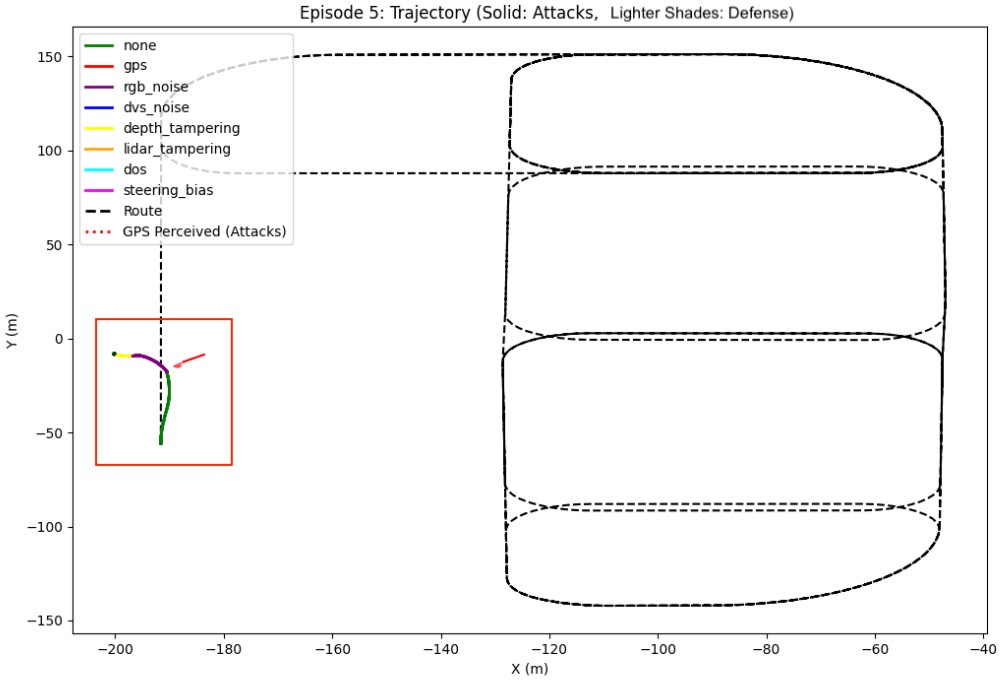

Figure 4. Episode 5 trajectory showing erratic movement and eventual crash during RGB camera attacks.

The sensor data from Episode 5 (Figure 5) reveals the severe degradation of perception capabilities that led to the crash. The RGB camera feed shows extreme noise corruption, while the fence that the vehicle ultimately collided with is visible in the sensor outputs.

 doi:10.20944/preprints202505.2262.v1

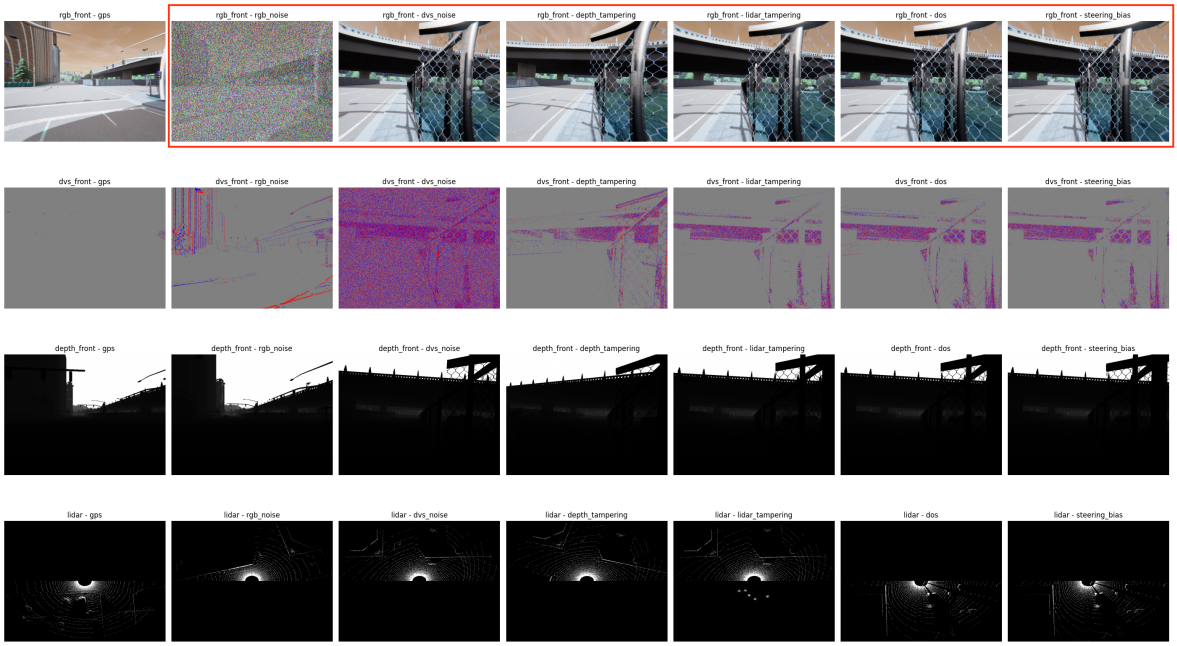

Figure 5. Sensor outputs for Episode 5 showing the roadside fence (visible in normal operation) that the vehicle crashed into during RGB camera attacks.

The control inputs during Episode 5 (Figure 6) further illustrate the vehicle's instability under attack. During RGB attacks, the control system exhibits erratic behavior, alternating between acceleration and full braking as it struggles to interpret the corrupted sensor data.

 doi:10.20944/preprints202505.2262.v1

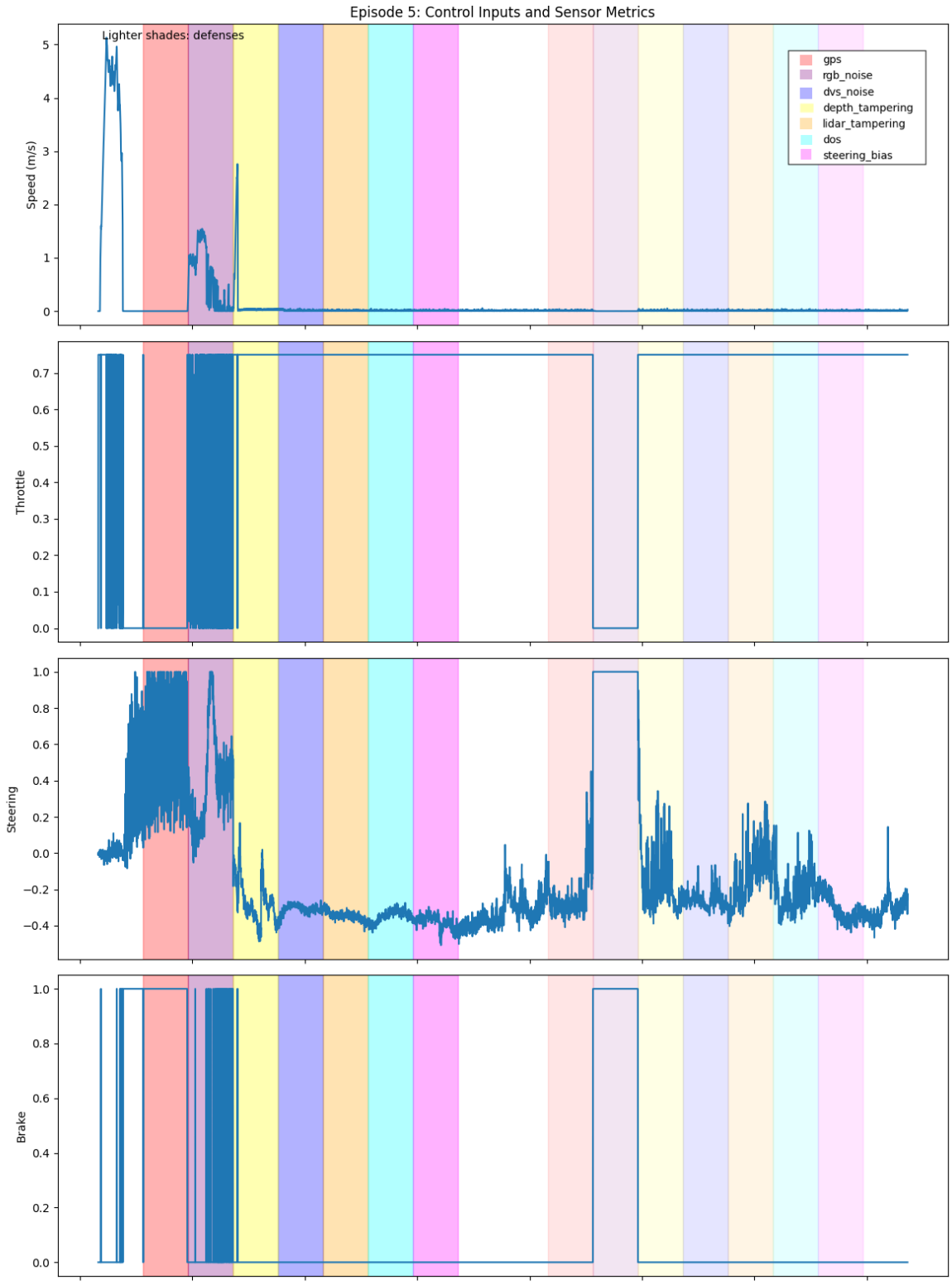

Figure 6. Control inputs and sensor data for Episode 5, showing unstable control behavior during RGB attacks, including full braking events.

Similarly, Episode 27 demonstrates a multi-stage failure where the vehicle crashed into a traffic light during the initial attack phase. Although it temporarily regained its route, it crashed again during a subsequent attack even with defense mechanisms active, as shown in Figure 7 and Figure 8.

 doi:10.20944/preprints202505.2262.v1

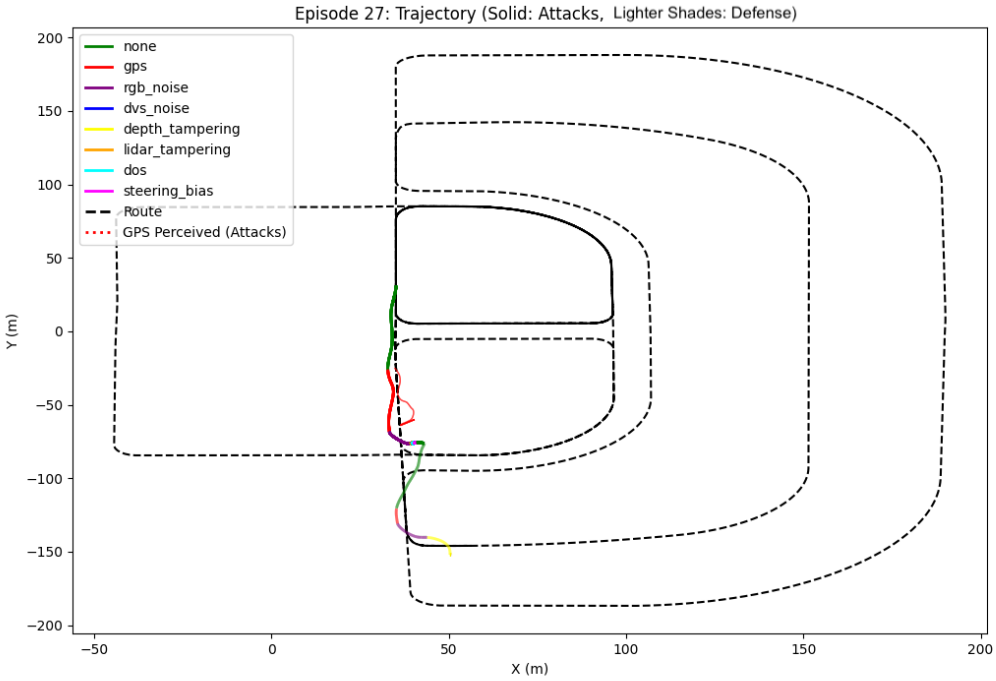

Figure 7. Episode 27 trajectory showing crashes during both the initial attack phase and the attack-with-defense phase.

 doi:10.20944/preprints202505.2262.v1

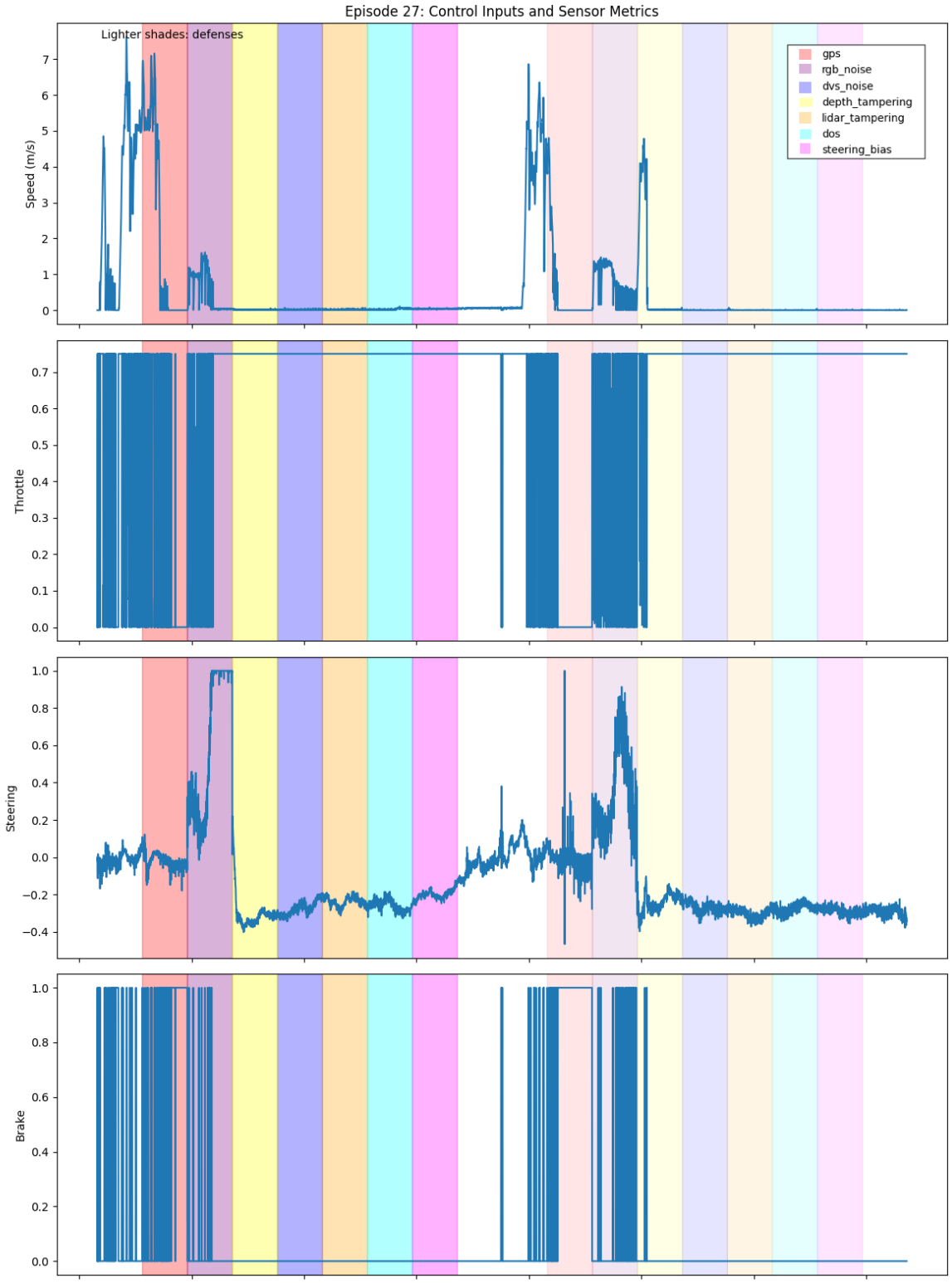

Figure 8. Control inputs and sensor data for Episode 27, showing the vehicle's response during crashes in both attack phases.

3.2. Effectiveness of Defense Mechanisms

Our proposed defense mechanisms aim to mitigate the impact of cyberattacks on autonomous vehicle sensors. We evaluated their effectiveness by comparing vehicle performance with and without defenses under identical attack conditions. Figure 9 illustrates the impact of our defense mechanisms on sensor data quality during attacks in Episode 7.

 doi:10.20944/preprints202505.2262.v1

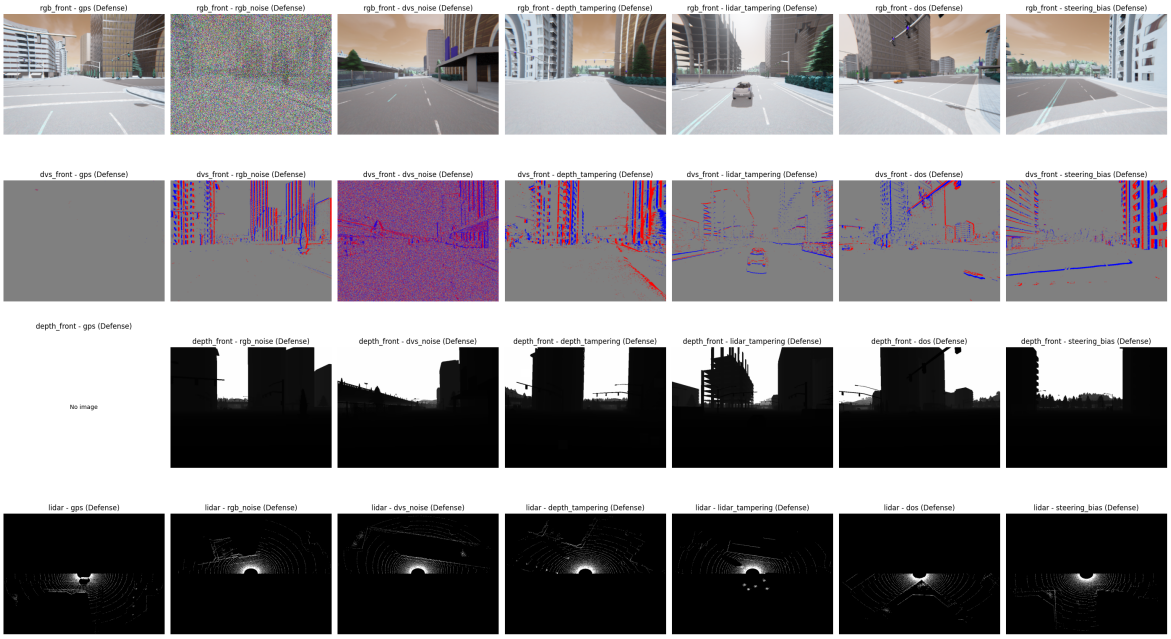

Figure 9. Sensor outputs with defense mechanisms during attacks in Episode 7.

The defense mechanisms demonstrated varying degrees of effectiveness depending on the attack type and intensity:

- **RGB Camera Attacks**: Our anomaly detection model for RGB cameras successfully identified attacks on the camera feed, but the subsequent filtering techniques showed limited effectiveness against sophisticated attacks. While some noise reduction was achieved, the filtered images often remained significantly compromised, resulting in continued trajectory deviations even with defenses active.
- **Depth Sensor Attacks**: The depth sensor anomaly detection system effectively detected anomalies in depth data, but defense mechanisms for depth sensor attacks demonstrated limited effectiveness against targeted attacks, with filtered depth maps still containing significant distortions that affected the vehicle's perception capabilities.
- **GPS Spoofing**: Our GPS-specific anomaly detection and defense system successfully detected and mitigated GPS spoofing attacks by implementing plausibility checks and cross-checking with IMU data. This was our most effective defense, as it clearly identified spoofed GPS coordinates and prevented the vehicle from making unwarranted course corrections based on falsified location data.
- **Steering Bias**: The steering-specific anomaly detection system combined with rate limiting showed partial effectiveness in identifying malicious steering commands, but sophisticated attacks could still influence vehicle control in certain scenarios.

Our experiments revealed significant limitations in our current defense mechanisms for most sensor types except GPS. In Episode 27, the vehicle crashed during an attack even with defenses active, and across multiple episodes, the defenses for RGB and depth sensors failed to adequately protect the vehicle's perception system. These findings highlight a critical vulnerability in current autonomous vehicle security approaches.

To address these limitations, we propose an enhanced anomaly detection system that can serve as a fallback mechanism when primary defenses fail. This system would:

- Continuously monitor all sensor inputs for anomalies using our trained detection models
- When anomalies are detected, reduce reliance on AI models for driving decisions
- Transfer driving control to human operators for a safer experience when sensor integrity is compromised

 doi:10.20944/preprints202505.2262.v1

- Alternatively, selectively shut down compromised sensors and cross-check with remaining functional sensors to maintain autonomous operation

This approach acknowledges the reality that perfect defense against all attack vectors may not be achievable, and instead focuses on rapid detection and appropriate fallback mechanisms to ensure safety. The implementation of such a system would be a promising direction for future research.

3.3. Anomaly Detection Performance

A critical component of our proposed enhanced defense framework is the anomaly detection system, which aims to identify malicious sensor manipulations in real-time and trigger appropriate fallback mechanisms. We evaluated the performance of our anomaly detection models for both RGB and DVS camera sensors using the anomalib library to assess their potential as a foundation for this enhanced security approach.

3.3.1. RGB Camera Anomaly Detection

The RGB camera anomaly detection model demonstrated excellent performance in distinguishing between normal operation and attack conditions. Figure 10 shows the anomaly scores for different operational states in Episode 1.

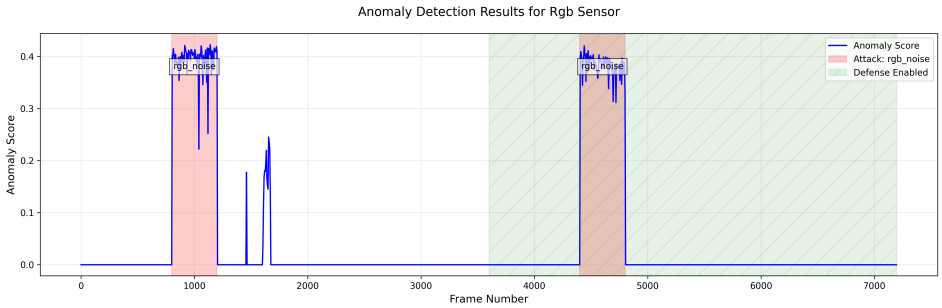

Figure 10. RGB camera anomaly scores across different operational states in Episode 1. Note the clear separation between normal operation and RGB noise attack conditions.

Statistical analysis of the RGB anomaly detection results revealed:

- **Normal Operation**: Mean anomaly score of 0.000000 with standard deviation of 0.000000, establishing a clear baseline for normal behavior.
- **RGB Noise Attacks**: Mean anomaly scores of 0.378064 (with defense) and 0.389462 (without defense), both significantly higher than normal operation. This demonstrates the model's ability to reliably detect RGB camera attacks regardless of whether defenses are active.
- Our RGB camera anomaly detection model is specifically designed to detect anomalies in RGB camera data only and does not process data from other sensors.

These results demonstrate that our RGB anomaly detection model achieves high sensitivity to relevant attacks while maintaining specificity against unrelated attack vectors.

3.3.2. DVS Camera Anomaly Detection

The DVS camera anomaly detection model showed different characteristics compared to the RGB model, reflecting the unique properties of event-based vision. Figure 11 illustrates the anomaly scores for different operational states in Episode 5.

 doi:10.20944/preprints202505.2262.v1

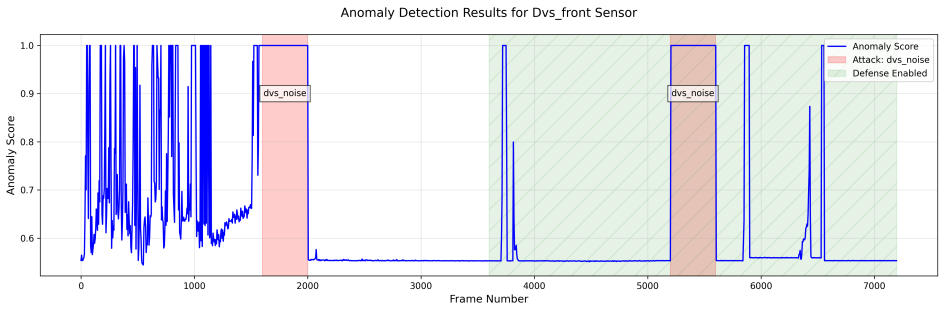

Figure 11. DVS camera anomaly scores across different operational states in Episode 5. Note the elevated scores during attack conditions without defenses.

Statistical analysis of the DVS anomaly detection results revealed:

- **Normal Operation**: Mean anomaly score of 0.608308 with standard deviation of 0.121466, establishing a baseline that reflects the inherent variability of event-based vision data.
- **DVS Noise Attacks**: Mean anomaly scores of 0.552093 (with defense) and 0.758073 (without defense). Interestingly, the score with defense is slightly lower than normal operation, while the score without defense is significantly higher. This suggests that our defense mechanisms not only mitigate the attack but also stabilize the event generation process.
- It's important to note that our DVS anomaly detection model is specifically designed to detect anomalies in DVS sensor data only and does not process data from other sensors. Additionally, DVS cameras are currently used for research purposes only and are not integrated into the vehicle's driving decision model.

These results highlight the importance of sensor-specific anomaly detection in autonomous vehicles. The DVS anomaly detection model successfully identifies abnormal patterns in DVS sensor data, while other sensor-specific models (such as the RGB anomaly detection model) are responsible for detecting attacks on their respective sensors. Each detection system is specialized and optimized for its particular sensor type, emphasizing the need for a multi-layered security approach.

3.4. Comparative Analysis of RGB and DVS Sensor Security

Our experiments provide valuable insights into the relative security characteristics of traditional RGB cameras and emerging DVS technology:

- **Research Context**: It is important to emphasize that DVS cameras in our study were used exclusively for research purposes and were not integrated into the vehicle's driving decision model. The data collected from DVS sensors was analyzed separately from the main autonomous driving pipeline.
- **Defense Effectiveness**: Our defense mechanisms were evaluated separately for each sensor type. RGB cameras, which are actively used in the driving model, required more aggressive filtering that sometimes reduced image quality. DVS defenses were studied in isolation as a research component.
- **Anomaly Detection**: The anomaly detection models for both sensors were designed to work exclusively with their respective sensor data. The RGB model showed clearer separation between normal and attack conditions and directly impacted vehicle safety, while the DVS model's patterns were analyzed purely for research insights.
- **Future Potential**: While not currently used for driving decisions, the complementary nature of RGB and DVS sensors suggests that sensor fusion approaches could potentially enhance security in future implementations.

 doi:10.20944/preprints202505.2262.v1

Table 1. Summary of Sensor Attacks and Defense Effectiveness

Sensor Type	Attack Type	Results Without Defense	Results With Defense
RGB Camera	Salt-and-Pepper Noise (80%)	Severe trajectory deviations; vehicle weaving across lanes; crashes in some episodes	Limited effectiveness; 30–45% trajectory drift still present; filtered images remain compromised
Dynamic Vision Sensor (DVS)	Event Flooding (60% spurious events)	False motion perception; higher anomaly scores (0.76)	Limited effectiveness; anomaly detection detects attack phase; event filter still compromised; lower anomaly scores (0.55)
Depth Camera	Depth Map Tampering (random patches)	Misinterpreted obstacle distances; affected perception of road boundaries	Limited effectiveness; filtered depth maps still contain significant distortions
LiDAR	Phantom Object Injection (5 clusters)	Unnecessary braking; evasive maneuvers	Partial mitigation through point cloud filtering; density-based clustering removes some phantom objects
GPS	Position Manipulation	Significant deviation between reported and actual position; incorrect routing decisions	Highly effective mitigation; IMU cross-checking prevented course corrections based on falsified data
Sensor Update	Denial of Service (rate limiting)	Delayed/missed sensor readings; affected perception and decision-making	Partial mitigation through sensor update monitoring and buffering mechanisms
Control System	Steering Bias	Systematic errors in steering commands; trajectory deviation	Partial effectiveness; sophisticated attacks could still influence vehicle control

These findings suggest that incorporating DVS technology alongside traditional sensors can enhance the security posture of autonomous vehicles, particularly in high-dynamic-range scenarios where RGB cameras are more likely to be compromised by environmental factors or deliberate attacks.

4. Discussion

Our comprehensive analysis of cyberattacks targeting autonomous vehicle sensors reveals critical insights into the security vulnerabilities and resilience mechanisms of these systems. In this section, we interpret our findings in the context of existing research and discuss their broader implications for autonomous vehicle security.

4.1. Implications for Autonomous Vehicle Security

The results of our experiments demonstrate that autonomous vehicles remain vulnerable to a range of sensor-based attacks despite advances in perception technologies. The successful completion of routes under attack conditions in Episodes 7 and 26, albeit with significant trajectory deviations, suggests that current autonomous driving systems possess a degree of inherent resilience. However, the catastrophic failures observed in Episodes 5 and 27, where vehicles crashed due to sensor manipulation, highlight the critical nature of these vulnerabilities.

 doi:10.20944/preprints202505.2262.v1

Particularly concerning is our observation that defense mechanisms showed limited effectiveness against sophisticated attacks on RGB and depth sensors. While our GPS spoofing defenses performed well, the continued vulnerability of visual perception systems represents a significant security gap that malicious actors could exploit. This suggests that the current approach of hardening individual sensors against attacks may be insufficient, and a more holistic security architecture is needed.

4.2. Significance of Anomaly Detection Approach

Our anomaly detection system demonstrates promising capabilities for identifying malicious sensor manipulations in real-time. The RGB camera anomaly detection model achieved excellent discrimination between normal operation and attack conditions, with attack scenarios consistently producing anomaly scores significantly higher than baseline. This clear separation suggests that anomaly detection could serve as a reliable trigger for fallback mechanisms when primary defenses fail.

Interestingly, the DVS camera anomaly detection model exhibited different characteristics, reflecting the unique properties of event-based vision. The higher baseline variability in DVS anomaly scores indicates that event-based sensors may require more sophisticated detection algorithms that account for their asynchronous nature. Nevertheless, the model successfully identified abnormal patterns during attacks, particularly when defenses were not active.

These findings support the potential of anomaly detection as a foundational component of a multi-layered security approach. By continuously monitoring sensor inputs for deviations from expected patterns, autonomous vehicles could rapidly identify potential attacks and implement appropriate countermeasures before safety is compromised.

4.3. Comparative Security of RGB and DVS Sensors

Our comparative analysis of RGB and DVS sensor security reveals important distinctions between these technologies. Traditional RGB cameras, while providing rich visual information, demonstrated significant vulnerability to noise-based attacks that severely degraded perception capabilities. In contrast, the event-based nature of DVS technology offers potential security advantages through its high dynamic range and reduced data redundancy.

The complementary characteristics of these sensor types suggest that a fusion approach could enhance overall system security. RGB cameras excel in static scene understanding and color perception, while DVS sensors offer superior performance in high-dynamic-range scenarios and rapid motion detection. By integrating data from both sensor types and implementing cross-validation mechanisms, autonomous vehicles could potentially detect inconsistencies that indicate sensor manipulation.

However, it is important to note that our study used DVS cameras exclusively for research purposes, without integration into the vehicle's driving decision model. Future work should explore the practical challenges of incorporating DVS technology into production autonomous vehicles, including calibration requirements, computational overhead, and integration with existing perception pipelines.

4.4. Limitations and Challenges

Our research has several limitations that should be acknowledged. First, all experiments were conducted in simulation using the CARLA environment, which may not fully capture the complexity and unpredictability of real-world driving scenarios. While CARLA provides realistic physics and sensor models, real-world implementation would face additional challenges such as sensor noise, environmental variability, and hardware constraints.

Second, our attack models represent a subset of possible attack vectors. More sophisticated attacks, such as adversarial machine learning techniques that target specific neural network vulnerabilities, were not fully explored in this study. Such attacks could potentially bypass our anomaly detection systems by generating perturbations that appear normal while causing misclassification.

Third, our defense mechanisms showed limited effectiveness against certain attack types, particularly for RGB and depth sensors. This highlights the challenge of developing robust defenses that can

 doi:10.20944/preprints202505.2262.v1

maintain perception quality while mitigating attacks. The trade-off between security and performance remains a significant challenge for autonomous vehicle designers.

Finally, the computational requirements of our anomaly detection system may present challenges for real-time implementation on resource-constrained vehicle platforms. While our experiments used high-performance GPUs, production vehicles may have more limited computational capabilities, necessitating optimization of detection algorithms.

4.5. Future Research Directions

Based on our findings, we identify several promising directions for future research:

- **Enhanced Anomaly Detection**: Developing more sophisticated anomaly detection models that can identify subtle attacks while maintaining low false positive rates. This could include exploring deep learning architectures specifically designed for time-series sensor data and incorporating contextual information from multiple sensors.
- **Sensor Fusion for Security**: Investigating how data from complementary sensors (RGB, DVS, LiDAR, radar) can be fused not only for improved perception but also as a security mechanism. Inconsistencies between different sensor modalities could serve as indicators of potential attacks.
- **Graceful Degradation**: Designing autonomous systems that can gracefully degrade performance when attacks are detected, rather than failing catastrophically. This could involve developing fallback driving policies that rely on uncompromised sensors or implementing safe stop procedures.
- **Human-in-the-Loop Security**: Exploring how human operators could be effectively integrated into the security framework, particularly for remote monitoring and intervention when anomalies are detected. This raises important questions about interface design, situation awareness, and response time.
- **DVS Integration**: Further investigating the potential of DVS technology for enhancing autonomous vehicle security, including developing specialized perception algorithms that leverage the unique properties of event-based vision and exploring hybrid RGB-DVS architectures.
- **Standardized Security Evaluation**: Developing standardized benchmarks and evaluation methodologies for assessing the security of autonomous vehicle sensor systems, enabling more systematic comparison of different defense approaches.

Addressing these research directions will require interdisciplinary collaboration between experts in computer vision, cybersecurity, autonomous systems, and human factors. By advancing our understanding of sensor vulnerabilities and developing more robust defense mechanisms, we can enhance the security and trustworthiness of autonomous vehicles as they transition from research laboratories to public roads.

Author Contributions: Conceptualization, M.S.; methodology, M.S., K.S., M.O.; software, M.S., K.S., M.O.; validation, M.S., K.S., M.O.; formal analysis, M.S., K.S., M.O.;investigation, M.S., K.S.; resources, M.S.; data curation, K.S., M.O.; writing-original draft preparation, M.S., K.S., M.O.; writing-review and editing, M.S., M.W.; visualization, M.S., K.S.; supervision, M.S., M.W.; project administration, M.W.

Data Availability Statement: The code for experiments is available at https://github.com/MustafaSakhai/Robust-Sensor-Cybersecurity-for-Autonomous-Vehicles

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

 doi:10.20944/preprints202505.2262.v1

AV	Autonomous Vehicle
CUDA	Compute Unified Device Architecture
DoS	Denial of Service
DVS	Dynamic Vision Sensors
EfficientAd	Efficient Anomaly Detection
GPU	Graphics Processing Unit
GPS	Global Positioning System
IMU	Inertial Measurement Unit
LiDAR	Light Detection and Ranging
NEAT	Neural Attention
PCLA	Pretrained CARLA Leaderboard Agent
RGB	Red Green Blue
VRAM	Video Random Access Memory

References

1. Giannaros, A.; Karras, A.; Theodorakopoulos, L.; Karras, C.; Kranias, P.; Schizas, N.; Kalogeratos, G.; Tsolis, D. Autonomous Vehicles: Sophisticated Attacks, Safety Issues, Challenges, Open Topics, Blockchain, and Future Directions. *J. Cybersecur. Priv.* **2023**, *3*, 493-543. https://doi.org/10.3390/jcp3030025
2. Sakhai, M.; Mazurek, S.; Caputa, J.; Argasiński, J.K.; Wielgosz, M. Spiking Neural Networks for Real-Time Pedestrian Street-Crossing Detection Using Dynamic Vision Sensors in Simulated Adverse Weather Conditions. *Electronics* **2024**, *13*, 4280. https://doi.org/10.3390/electronics13214280
3. Hussain, M.; Hong, J.-E. Reconstruction-Based Adversarial Attack Detection in Vision-Based Autonomous Driving Systems. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1589-1611. https://doi.org/10.3390/make5040080
4. Guesmi, A.; Hanif, M.A.; Shafique, M. AdvRain: Adversarial Raindrops to Attack Camera-Based Smart Vision Systems. *Information* **2023**, *14*, 634. https://doi.org/10.3390/info14120634
5. Mahima, K.T.Y.; Perera, A.; Anavatti, S.; Garratt, M. Exploring Adversarial Robustness of LiDAR Semantic Segmentation in Autonomous Driving. *Sensors* **2023**, *23*, 9579. https://doi.org/10.3390/s23239579
6. Y. Wang et al., "EV-Gait: Event-Based Robust Gait Recognition Using Dynamic Vision Sensors," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 6351-6360, doi: 10.1109/CVPR.2019.00652.
7. Y. Suh et al., "A 1280×960 Dynamic Vision Sensor with a 4.95-μ Pixel Pitch and Motion Artifact Minimization," *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, Seville, Spain, 2020, pp. 1-5, doi: 10.1109/ISCAS45731.2020.9180436.
8. McReynolds, B.; Graca, R.; Delbruck, T. Experimental methods to predict dynamic vision sensor event camera performance. *Opt. Eng.* **2022**, *61*, 074103.
9. Kim, H.; Leutenegger, S.; Davison, A.J. Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016.
10. Sakhai, M.; Sithu, K.; Oke, M. K. S.; Mazurek, S.; Wielgosz, M. Evaluating Synthetic vs. Real Dynamic Vision Sensor Data for SNN-Based Object Classification. In Proceedings of the KU KDM 2025 Conference, 2025.
11. K. Eykholt et al., "Robust Physical-World Attacks on Deep Learning Visual Classification," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 1625-1634, doi: 10.1109/CVPR.2018.00175.
12. Cao, Y.; Jia, J.; Cong, G.; Na, M.; Xu, W. Adversarial Sensor Attack on LiDAR-Based Perception in Autonomous Driving. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19)*, 2019.
13. Sun, J.; Cao, Y.; Chen, Q. A.; Mao, Z. M. Towards Robust LiDAR-Based Perception in Autonomous Driving: General Black-Box Adversarial Sensor Attack and Countermeasures. In *Proceedings of the 29th USENIX Security Symposium (USENIX Security 20)*, 2020.
14. J. Petit and S. E. Shladover, "Potential Cyberattacks on Automated Vehicles," *IEEE Transactions on Intelligent Transportation Systems* **2015**, vol. 16, no. 2, pp. 546-556, April 2015, doi: 10.1109/TITS.2014.2342271.
15. Gehrig, D., Scaramuzza, D. Low-latency automotive vision with event cameras. *Nature* **2024**. https://doi.org/10.1038/s41586-024-07409-w
16. Bogdoll, D.; Hendl, J.; Zöllner, J. M. Anomaly Detection in Autonomous Driving: A Survey. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2022.

 doi:10.20944/preprints202505.2262.v1

17. Kołomański, M., Sakhai, M., Nowak, J., Wielgosz, M. (2023). Towards End-to-End Chase in Urban Autonomous Driving Using Reinforcement Learning. In: Arai, K. (eds) Intelligent Systems and Applications. IntelliSys 2022. Lecture Notes in Networks and Systems, vol 544. Springer, Cham. https://doi.org/10.1007/978-3-031-16075-2_29
18. Sakhai, M., Wielgosz, M. (2024). Towards End-to-End Escape in Urban Autonomous Driving Using Reinforcement Learning. In: Arai, K. (eds) Intelligent Systems and Applications. IntelliSys 2023. Lecture Notes in Networks and Systems, vol 823. Springer, Cham. https://doi.org/10.1007/978-3-031-47724-9_2