Securing the Perception of Advanced Driving Assistance Systems Against Digital Epileptic Seizures Resulting from Emergency Vehicle Lighting

Elad Feldman, Jacob Shams, Satoru Koda, Yisroel Mirsky, Asaf Shabtai, Yuval Elovici, Ben Nassi

{eladfeld, jacobsh, yisroel, nassib}@post.bgu.ac.il, {shabtaia, elovici}@bgu.ac.il, koda.satoru@fujitsu.com
Ben-Gurion University of the Negev, Fujitsu Limited

Abstract

The safety of autonomous cars has come under scrutiny in recent years, especially after 16 documented incidents involving Teslas (with autopilot engaged) crashing into parked emergency vehicles (police cars, ambulances, and firetrucks). Despite widespread public interest, the technical factors behind these accidents have remained largely unexplored. In this research, we unveil *EpileptiCar*, a digital epileptic seizure phenomenon that causes an object detector's confidence score to fluctuate when exposed to activated emergency vehicle lighting. This vulnerability poses a significant risk, because not only does it cause autonomous vehicles to crash near emergency vehicles, but it can also be exploited by adversaries to cause such accidents. We delve into the underlying causes of the *EpileptiCar* phenomenon, assessing five commercial advanced driving assistance systems (HP, Pelsee, AZDOME, Imagebon, Rexing), four object detectors (YOLO, SSD, RetinaNet, Faster R-CNN), and 14 patterns of emergency vehicle lightnings to understand the influence of different factors. To mitigate this risk, we propose *Caracetamol*, a robust framework designed to enhance the resilience of object detectors against the effects of activated emergency vehicle lighting. Our evaluation shows that the implementation of *Caracetamol* improves the detection capabilities of three object detectors (YOLO, Faster R-CNN, and SSD) by 0.21 and reduces the confidence fluctuation range by 0.19, significantly enhancing safety in the presence of emergency vehicle lightning.

1 Introduction

The safety of autonomous cars has been questioned in the last few years following 16 documented incidents [1] in which Tesla cars with autopilot engaged collided with parked emergency response vehicles, such as fire engines [2], police cars [3], and ambulances [4]. These accidents resulted in significant injuries and fatality [5], causing the National Highway Transportation Safety Administration (NHTSA) to launch an in-depth investigation of the incidents to determine the factors underlying these accidents [6]. While driver distraction has been identified as a primary cause [5], the technical issues that led to these accidents remain a mystery [5].

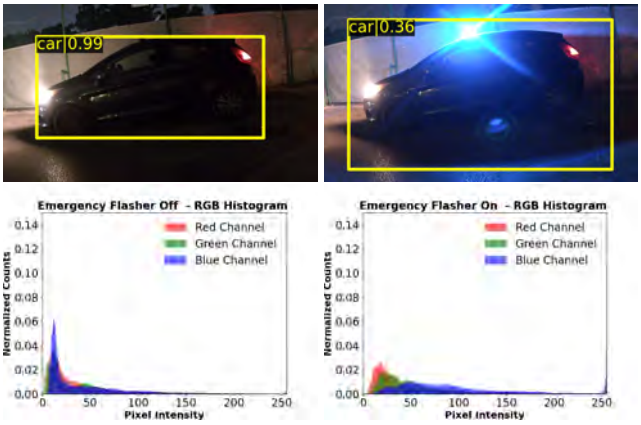

Figure 1: Top: A vehicle detected by Faster R-CNN with a confidence of 0.99 (left - emergency vehicle lighting off) and 0.36 (right - emergency vehicle lighting on) in pictures obtained by AZDOME ADAS. Bottom: The associated normalized tonal distribution of these images.

Despite public interest and the importance of the investigation, the automotive industry and NHTSA have not disclosed their findings, leaving the public with open questions regarding the technological aspects of these accidents. Shedding light on the technological aspects of this mystery is crucial, especially in light of the increased adoption of semi-autonomous cars and advanced driving assistance systems (ADASs), because (1) such accidents can result in injuries or the death of drivers, passengers, and pedestrians [5]. (2) There is an industry shift in the architectural design of new semi-autonomous cars that is increasingly leaning towards vision-based systems [7] without RADARs and ultrasonic sensors; Therefore, there is a need to understand the consequences of this shift in light of the technical causes of the 16 accidents. (3) Until this mystery is solved, the ability to develop methods capable of securing the perception of semi-autonomous cars and other perception-based systems (e.g., drones, humanoids) against

these accidents will be curtailed.

To address this lack of understanding, this study aims to explore why vision-based perception fails when encountering vehicles whose emergency vehicle lighting is activated. More specifically, we strive to identify the role of object detectors in this failure. This is not only important from a safety perspective, but also from a security perspective, since attackers can leverage this failure to target semi-autonomous cars and deliberately cause damage and endanger the lives of road users. First, we analyze video footage obtained from five commercial ADASs (HP [8], Pelsee [9], AZDOME [10], Imagebon [11], Rexing [12]) and show that four object detectors are vulnerable to a phenomenon we have termed *EpileptiCar*, a digital epileptic seizure that causes the confidence score of object detectors regarding a detected object to fluctuate due to the presence of an activated emergency vehicle flasher. This has dangerous implications, as the fluctuation in the confidence score may result in a value below the detection threshold, causing an object to go unnoticed (i.e., undetected) by semi-autonomous cars and ADASs.

In our investigation of the *EpileptiCar* phenomenon, we discover that the confidence score of an object detector is affected by the intensity of the light emitted from activated emergency vehicle lighting. An activated flasher's light intensity changes the tonal distribution of the car (in the frame captured by the video camera) over time (between different frames) according to the pattern and frequency of the activated emergency vehicle lighting. Consequently, lower confidence detection scores are produced by the object detector in different frames, affecting the object detector's behavior and performance and causing detection loss of objects that appear in the frame (see Fig. 1). We also investigate the factors affecting the *EpileptiCar* phenomenon and find that it appears in a variety of object detectors and commercial ADASs, creates unique detection behavior depending on the pattern of the emergency vehicle lighting, and mainly compromises object detectors' performance in dark conditions.

To address the *EpileptiCar* phenomenon, we introduce *Caracetamol*, a framework aimed at improving the performance of a given object detector in the presence of emergency vehicles. This is done using a multi-layer architecture consisting of (1) a classification layer, intended to reduce unnecessary runtime in cases when an emergency vehicle lighting does not appear in an image, and (2) a detectors layer which is composed of three replications of the original ADAS object detector, each refined on augmented data we created due to the lack of images of emergency vehicle lighting operated at night. The detections of the detectors layer and the original object detector are aggregated in (3) an ensemble layer which returns the final detections. We tested the performance of three implementations for *Caracetamol*, based on three different object detectors (YOLO, SSD, Faster R-CNN), using videos of emergency vehicles downloaded from YouTube and found that it improves the resiliency of an object detector to *EpileptiCar* significantly. On average, it improved the models' average confidence of car detection by 0.21, the lower confidence bound by 0.27, and reduced the fluctuation range by 0.19.

Contributions. This paper makes the following contributions: (1) we empirically investigate in five ADASs (HP [8], Pelsee [9], AZDOME [10], Imagebon [11], Rexing [12]) and popular open-source object detectors the reasons for the degradation in performance in the presence of activated emergency vehicle lighting; (2) in the absence of a relevant comprehensive dataset, we collect and augment a unique dataset that includes over 40,000 images of emergency response vehicles that we make available to the research community[1] to enable further research; and (3) we introduce a framework aimed at enhancing the robustness of object detectors against the *EpileptiCar* phenomenon, which improves their resilience significantly.

Disclaimer. Lacking access to the specific type of object detectors and video cameras deployed in Tesla, we perform our analysis based on commonly used object detectors with commercial ADASs available to purchase on Amazon. For this reason, there may be a discrepancy between our findings and the actual technological factors underlying the documented accidents of Tesla. We hope that this paper will encourage the automotive industry to validate our findings on their ADASs and semi autonomous cars with their object detectors.

Structure. The paper is structured as follows: We present the threat model in Section 2. In Section 3, we examine why object detectors' performance degrades in the presence of activated emergency vehicle lighting. In Section 4, we describe and evaluate *Caracetamol*, a framework aimed at securing an object detector against the *EpileptiCar* phenomenon. In Section 5 we discuss the limitations of our research and in Section 6, we review related work. We discuss our findings in Section 7.

Ethical Considerations, Responsible Disclosure & Open Science. We disclosed our findings to the NHTSA, Tesla, and the five manufacturers of the ADASs. We will add their responses to the paper when provided. We uploaded our code and dataset to GitHub[1] to allow reproducibility and replicability of our findings.

2 Threat Model

In this section, we review the threat model in two related yet distinct cases; unintentional and intentional threats.

Unintentional Threat Model (resulting from a safety issue). In this case, an ADAS encounters an activated flasher, causing the ADAS to unintentionally demonstrate undesired behavior. This is a safety issue caused by commonly occurring natural phenomena, activated emergency vehicle lighting. The involved actors are: (1) a car controlled by an ADAS,

[1] we will add the URL to the camera-ready version of the paper

i.e., a car with an autonomy level of 2+ to 5, with autopilot functionality engaged, and (2) an emergency vehicle (e.g., police car, ambulance, fire truck) with its emergency vehicle lighting activated. The scenario is as follows: a distracted driver is driving their car towards an emergency vehicle in dark conditions; the autopilot functionality (controlled by the ADAS) is engaged; the emergency lights of the emergency vehicle are flashing; and the emergency vehicle is parked in the path of the car. A distracted driver who was not focused on monitoring the car behavior would be unable to properly respond to the failed or late detection of the emergency vehicle [5]. This unfortunate scenario can result in an accident involving the car, the emergency vehicle, and nearby cars and pedestrians. We note that this was the case associated with 16 documented incidents in which Teslas collided with emergency vehicles. A video of one of the incidents was recently published in [13].

Intentional Threat Model (the result of an attack). In this case, an attacker exploits the safety issue stemming from object detectors' undesired perception and reaction to emergency vehicle lighting, in order to attack the ADAS and deliberately cause undesired behavior. Much like how Rowhammer [14] is a safety issue in DRAM memory, that was exploited by attackers to compromise the security of computers [15–19], *EpileptiCar* can be exploited by attackers to damage the performance and behavior of ADASs. By exploiting emergency vehicle lighting, attackers can cause undesired reactions of ADASs which could threaten the safety of road users. We categorize two types of attacks: *untargeted attack* and *targeted attack*. In an untargeted attack, the attacker utilizes an emergency vehicle lighting at random areas and random times of day to disrupt the functionality of ADASs that happen to cross the affected area. Such attackers have the desire to induce an undesired reaction from any ADAS and/or don't target a particular target ADAS. In a targeted attack, the attacker targets a specific car and operates the flasher when it is recognized nearby.

Potential Outcomes. The failure to detect emergency vehicles can lead to: (1) Safety Hazards: There is an increased risk of collisions with emergency vehicles, posing serious safety threats to both vehicle occupants and emergency first responders. Notably, this risk has already led to 16 documented accidents between Teslas and emergency vehicles, resulting in injuries. (2) System Reliability Issues: Detection failures, particularly those that necessitate manual intervention or overrides by the driver, can affect the overall reliability of ADASs, eroding confidence in their adoption.

3 Understanding the EplipetiCar Phenomenon

In this section, we describe the analysis we performed to understand the technical factors that led advanced driver-assistance system (ADAS) controlled cars with the autopilot engaged to crash into emergency vehicles. We note that driver distraction was already raised as a primary cause of these accidents [5]. The analysis described in this section focuses on the technical aspects of these accidents in an attempt to understand whether the object detector fails to detect emergency vehicles and what factors affect detection in such cases.

Table 1: The functionalities supported by the ADASs we used in this research.

ADAS Name	Lane Departure	Collision Warnings	Vehicle Detection	FPS	Resolution
HP	✓	✗	✗	30	1080p
Pelsee	✓	✓	✓	60	1440p
AZDOME	✓	✓	✗	24	720p
Imagebon	✓	✓	✓	60	1080p
Rexing	✓	✓	✓	30	1080p

Figure 2: The ADASs we used: Rexing is marked in Yellow, Pelsee is marked in Blue, HP is marked in Green, Imegebon is marked in Red, and AZDOME is marked in Orange.

We first explain the *EpileptiCar* phenomenon by analyzing video footage of a car with an activated emergency vehicle lighting using video footage obtained by five ADASs; HP [8], Pelsee [9], AZDOME [10], Imagebon [11], Rexing [12]. These ADASs were chosen due to them being off-the-shelf level one ADASs which are readily available for purchase on Amazon (see Fig. 2) and support functions of forward-collision warning, lane departure warning, and vehicle detection (see Table 1). We also obtained video using a Samsung Galaxy S22 Ultra.

The video footage was analyzed using four object detectors (YOLOv9 [20], Faster R-CNN [21], SSD [22], and RetinaNet [23]). We selected these object detectors because three of them (YOLO, SSD, and RetinaNet) are considered the *"most popular one-stage object detectors"*[2] used for real-time applications, and we selected Faster R-CNN because it is considered a time-efficient two-stage object detector [24]. For YOLOv9, we used its open-source pretrained implementation, and for the other three object detectors, we used the pretrained implementations provided by MMDetection [25]. All four object detectors were pretrained on the COCO dataset. Some of the experiments in this section were conducted with YOLOv3 instead of YOLOv9 (this is noted in the relevant experiments), because YOLOv3 is also implemented in MMDetection, which made the analysis easy. Later in this section, we demonstrate that the application of object trackers on top of object detectors does not mitigate *EpileptiCar*. Finally, we

[2] https://viso.ai/deep-learning/object-detection/

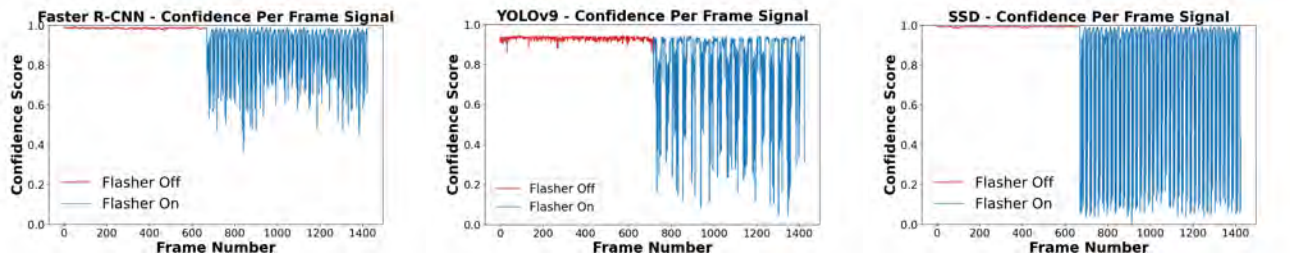

Figure 3: Comparison of the confidence score per frame for Faster R-CNN, YOLOv9, and SSD.

evaluate the influence of ambient light.

3.1 Analysis of Object Detector Behavior

In this section, we analyze the behavior of object detectors when detecting objects in an environment with and without activated emergency vehicle lighting.

3.1.1 Time-based Analysis

First, we evaluate four object detectors' confidence scores for the task of identifying an emergency vehicle, comparing a scenario in which the emergency vehicle lighting is on to a scenario when it is off.

Experimental Setup: An AZDOME ADAS was used to record a 60-second video recording (the FPS and resolution of the ADAS footage can be found in Table 1) of a grey Ford Fiesta equipped with a blue-red emergency vehicle lighting (purchased on AliExpress[3]). In the first 30 seconds of the video, the car's emergency vehicle lighting was off, and in the last 30 seconds, it was on (see Fig. 1). The video was segmented into frames, and the four object detectors were applied to each frame. The outputs were used to generate confidence score signals, a time series of the detector confidence score as a function of time/frame, for each detector regarding car detection.

Results: The results of YOLO, SSD and Faster R-CNN are presented in Fig. 3, the result of RetinaNet is presented in the appendix in 26. As can be seen from the results, in the first half of the video (when the flasher was off), the detectors' confidence remained stable when detecting a car with confidence higher than 0.9. In contrast, in the second half of the video (when the flasher was on), the scores of the object detector fluctuated, down to values below 0.4 in YOLOv9, Faster R-CNN, and SSD.

Insight 1: *The activation of emergency vehicle lighting creates a phenomenon, which we term the EpileptiCar phenomenon, that causes the confidence score of object detectors regarding a detected object to fluctuate, within a wide score range, with the score dipping below a reasonable detection threshold in some cases.*

[3] https://aliexpress.com/item/1005005852809515.html

As can be seen in Fig. 3, similar behavior was observed in all the examined object detectors, indicating a systemic issue in object detectors rather than an isolated case.

Insight 2: *The EpileptiCar phenomenon is consistent across different object detectors, but the score fluctuation ranges differ depending on the object detector.*

3.1.2 Camera Analysis

Next, we evaluate footage from five ADASs and a smartphone camera and their effect on the confidence scores of an object detector for the task of identifying an emergency vehicle, both when the emergency vehicle lighting is on and when it is off.

Experimental Setup: Five ADASs (HP, Pelsee, AZDOME, Imagebon, Rexing) and a smartphone camera (Samsung Galaxy S22 Ultra) were used individually to record a 60-second video recording (the FPS and resolution of the ADAS footage can be found in Table 1, and the Samsung Galaxy footage was recorded at 24 FPS at FHD resolution) of a grey Ford Fiesta equipped with a blue emergency vehicle lighting (purchased on Amazon[4]). The experimental setup is presented in Fig. 2. In the first 30 seconds of each recording, the car's emergency vehicle lighting was off, and in the last 30 seconds, it was on (see Fig. 1). Each recorded video was segmented into frames, and the four object detectors were applied to each frame. The outputs were used to generate confidence score signals, a time series of the detector confidence score as a function of time/frame, for each ADAS/smartphone and object detector regarding car detection.

The results for SSD are presented in Fig. 4. As can be seen from the results, in the first half of most of the recordings (when the flasher was off), the detectors' confidence remained more stable when detecting a car than when turning on the flasher afterwards. In contrast, in the second half of the recordings (when the flasher was on), the confidence score of the object detector fluctuated, down to values below 0.4 in the case of four ADAS and in the smartphone.

Insight 3: *The EpileptiCar phenomenon is consistent across different ADASs, with the score fluctuation ranges and stability when the flasher is off differing depending on the ADAS recording the video frames.*

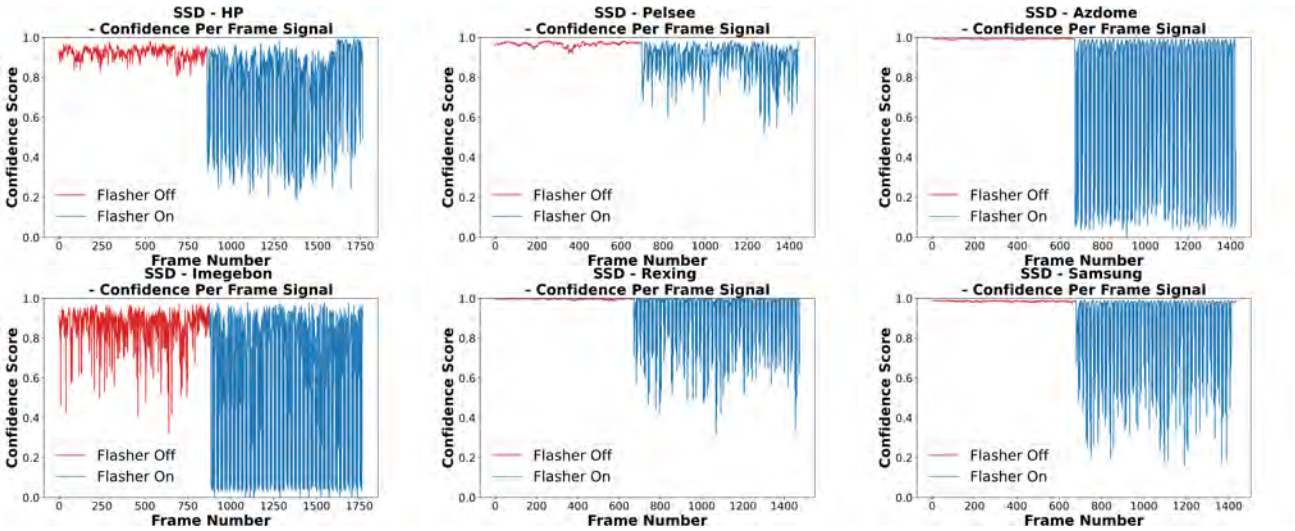

Figure 4: Comparison of the confidence score of SSD per frame for videos recorded by different ADASs and a smartphone.

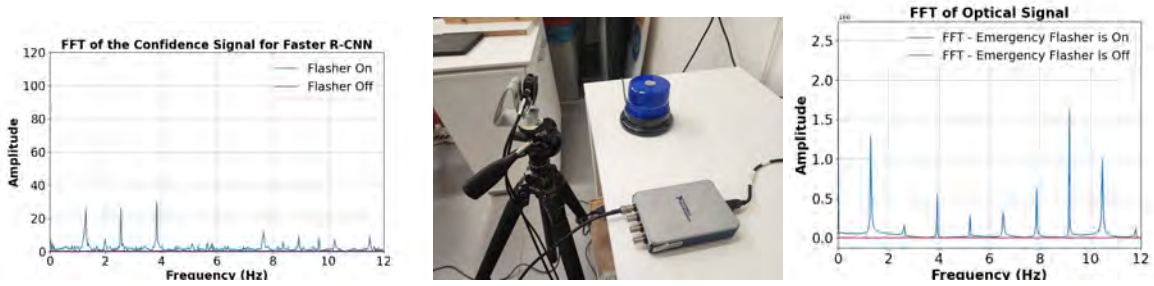

Figure 5: Left: Extracted FFT graph from the confidence signal of Faster R-CNN recorded by AZDOME when the emergency vehicle lighting is on. A peak around 1.3 Hz can be observed. Center: Experimental setup; a photodiode was directed at a blue emergency vehicle lighting in a lab setting. Right: An FFT graph of the recorded optical signal with a peak around 1.3 Hz.

3.1.3 RGB-Based Analysis

Next, we analyze the impact of an activated emergency vehicle lighting on a captured frame.

Experimental Setup: We extracted two frames from the previously recorded 60-second video recording that was obtained by AZDOME: (1) a random frame from the first 30 seconds of the video recording (when the emergency vehicle lighting was off) in which a car was detected by Faster R-CNN with a confidence of 0.99, and (2) one frame from the last 30 seconds (when the emergency vehicle lighting was on) in which a car was detected by Faster R-CNN with a confidence of 0.36.

Results: We computed the tonal distribution for the two extracted frames. Fig. 1 presents an analysis of the normalized RGB distribution of the colors of the car when the flasher is on and off. As can be seen, there is a clear difference in the color distributions of the two frames resulting from the blue light emitted from the emergency vehicle lighting.

Insight 4: *The light emitted from the emergency vehicle lighting changes the distribution of the colors of the car in the captured frame, which changes the confidence of the object detector regarding the car.*

3.1.4 Spectral Analysis

Next, we show that the presence of an activated emergency vehicle lighting determines the object detector behavior.

Experimental Setup: We used AZDOME to obtain a 60-second video recording from a blue emergency vehicle lighting (purchased on Amazon[4]). In the first 30 seconds of the video, the car's emergency vehicle lighting was off, and in the last 30 seconds of the video it was on. The video was segmented into frames, and the four object detectors were applied to each frame. The outputs were used to generate confidence score signals (*i.e.,* detector confidence as a function of time) for each detector regarding car detection. For each of the four signals extracted from the object detector's confidence, we computed two fast Fourier transforms (FFTs), one from the first 30 seconds of the signal (when the emergency vehicle lighting is off) and one from the last 30 seconds of the signal (when the emergency vehicle lighting is on).

Results: The results of the FFT graphs extracted from the detections of Faster R-CNN are presented in Fig. 5 left, and the results for YOLO, SSD, and RetinaNet are presented in Fig. 23 in the appendix. All FFT graphs show a peak around

[4] https://www.amazon.com.au/Xprite-Rotating-Revolving-Magnetic-Emergency/dp/B07B7YYL321

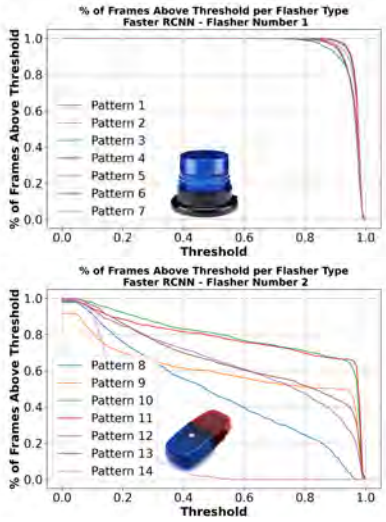

Figure 6: The percentage of frames in which a car was detected by Faster R-CNN with confidence higher than the threshold (in the presence of a flasher).

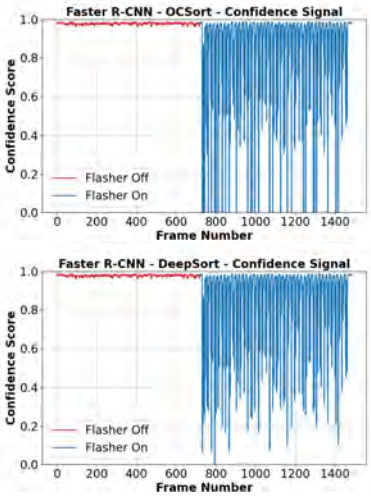

Figure 7: The confidence score signals of two object tracking models (OCSort and DeepSort) applied on the Faster R-CNN object detector

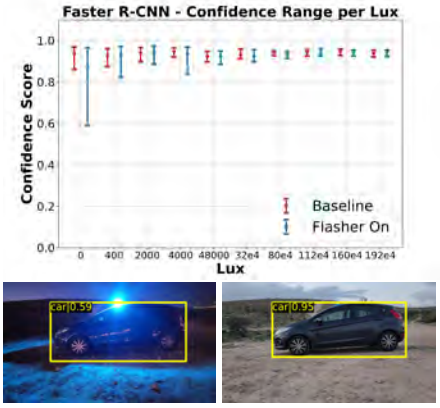

Figure 8: Top: Faster R-CNN confidence score signal range for each examined lux value. Bottom: There is a difference in the confidence score of a detected car of 0.36: in 0 lux a car is detected with a score of 0.59 (left) and in 160,000 lux a car is detected with a score of 0.95 (right).

1.3 Hz when the emergency vehicle lighting is activated which does not appear when the emergency vehicle lighting is off.

In an attempt to explain the behavior that appears in the four FFT signals, we conducted an additional experiment.

Experimental Setup: We directed a Thorlabs PDA100A2 photodiode (an optical sensor used to convert the intensity of captured light to electrical voltage) in a lab setup and captured optical measurements for 60 seconds: in the first 30 seconds of the experiment, the emergency vehicle lighting was off and in the last 30 seconds the emergency vehicle lighting was on. The voltage (optical measurements) was obtained from the photodiode using a 24-bit ADC NI-9234 card that was connected to a laptop and was processed in a LabVIEW script that we wrote. The experimental setup is presented in Fig. 5 middle.

Results. Fig. 5 right presents the application of FFT on the optical measurements. We can see a peak around 1.3 Hz in the FFT graph when the emergency vehicle lighting is activated (which does not appear when the emergency vehicle lighting is off), presenting the same behavior we observed when analyzing the confidence signal of the four object detectors.

Insight 5: *The frequency of the emergency vehicle lighting determines/dictates the behavior of the object detector's confidence in detecting the car over time.*

3.1.5 Analysis of Detection Loss

Next, we analyzed the detection loss (the rate at which the object detector failed to detect a car in the frames) due to an activated flasher.

Experimental Setup. We used two emergency lights: (1) the blue emergency vehicle lighting[4], and (2) a blue-red emergency vehicle lighting[3]. Each of the emergency lights has settings that enable seven different light patterns. We mounted each of the emergency lights to the roof of a car and used AZDOME to record 30-second videos for each of the 14 light patterns (seven from each flasher). We applied the four object detectors to the frames of each of the videos and calculated the percentage of frames (0-1.0) in which the car was detected beyond a given threshold (0-1.0).

Results. The results of Faster R-CNN for detecting cars in the two videos are presented in Fig. 6, while the results for YOLOv9, SSD, and RetinaNet are presented in Fig. 22 in the appendix. As can be seen from the results presented in Fig. 6, the two emergency lights yield different results for Faster R-CNN. For the blue emergency vehicle lighting, the detection loss of Faster R-CNN is marginal up to a threshold of 0.9, meaning that a car is detected in 90% of the frames at a detection threshold of 0.9. On the other hand, the detection loss of the red-blue emergency vehicle lighting is significant as 20% of the scores of the cars in the frames are detected with a confidence lower than 0.4.

Insight 6: *The EpileptiCar phenomenon presents a trade-off between a high true positive rate of an object detector (which could be satisfied with low detection thresholds) and the object detector's accuracy/low false positive rate (which could be satisfied with high detection thresholds).*

Insight 7: *The intensity of the light emitted from the emergency vehicle lighting changes the tonal distribution of the car in the frame captured by the video camera. As a result, the behavior and performance of an object detector are affected by the pattern and frequency of the emergency vehi-*

cle lighting. This causes detection losses in object detectors that are unique to the flasher's pattern/frequency.

3.2 Analysis of Object Tracker Behavior

We note that object trackers are employed one layer above the object detectors in some ADASs [26] and are used to identify the paths of detected objects. These paths are provided as input to the decision-making algorithms (which are employed one layer on top of the trackers). One might argue the analysis conducted so far lacks the full picture, because it does not consider the use of object trackers, which might compensate for an object detector's detection loss in the presence of an activated emergency vehicle lighting. In the next experiment, we analyze the behavior of object trackers in response to the *EpileptiCar* phenomenon.

Experimental Setup. We applied three object trackers (OCSORT [27], DeepSORT [28], and ByteTrack [29]) to the previously recorded 60-second video captured by AZDONE. Object trackers rely on a preliminary step of object detection, so we applied the three object trackers with the Faster R-CNN.

Results. The results of Faster R-CNN, with object trackers DeepSort and OCSort presented in Fig. 7 while the results of ByteTrack object tracker can be found in the appendix in Fig. 27. As can be seen, the object tracking algorithms do not compensate for Faster R-CNN's detection loss. The same behavior is seen in response to the activated emergency vehicle lighting regardless of whether an object tracker was used or not. Thus, the intended behavior of object trackers (associating new detections of objects with previous detections) does not mitigate the object detectors' detection loss in the presence of an activated emergency vehicle lighting.

Insight 8: *Object tracking fails to compensate for object detectors' detection loss in the presence of activated emergency vehicle lighting.*

3.3 Effect of Ambient Light

Here we examine how the level of ambient light affects object detectors' confidence in recognizing a car.

Experimental Setup: We positioned a Samsung Galaxy S22 Ultra five meters away from a grey Ford Fiesta with an emergency vehicle lighting mounted on its roof. We recorded 10 different 30-second videos of the car, with a 15-minute interval between the videos, allowing us to examine the effect of progressively diminishing light levels (transitioning from daylight to darkness). We used a professional lux meter (Extech HD450 [30]) to measure the lux level. We applied the four pretrained object detectors to the frames from each video and measured the car's detection confidence range as a function of the video's lux level. This experimental setup was repeated twice: first when the emergency vehicle lighting was off and a second time when it was on.

Results: The results of this analysis for the Faster R-CNN detector are presented in Fig. 8. The results for the other object detectors (YOLOv3, SSD, RetinaNet) are presented in Fig. 24 in the appendix. As can be seen, for all four object detectors, detection confidence remained high when either: (1) there was sufficient ambient light, or (2) the emergency vehicle lighting was off. The results also show that the confidence range changed depending on the environmental conditions, indicating less stability in the object detectors' confidence, particularly when: (1) ambient light levels were low, and (2) the emergency vehicle lighting was on. The frames associated with the lowest confidence scores in the videos recorded when the ambient light was 0 lux and 160,000 lux are presented in Fig. 8 and show that the emergency vehicle lighting significantly affects the image captured in darkness with respect to the image captured in light.

Insight 9: *The EpileptiCar phenomenon creates wider confidence score ranges in dark environments, leading to less stable object detector performance.*

In addition, we further investigate the effect of an ADAS' camera settings, the effect of various characteristics of the emergency vehicle (color, orientation, etc.), and the influence of distance between the camera and the emergency vehicle lighting, on the *EpileptiCar* phenomenon. These analyses can be found in the appendix.

4 The Caracetamol Framework

In this section, we introduce *Caracetamol*, a framework aimed at enhancing the resilience and accuracy of object detection in the presence of emergency vehicle flashers that takes the constraints of real-time driving into account. In the subsections that follow, we discuss the considerations that guided *Caracetamol*'s development, describe the datasets employed for training and evaluation, describe the framework's architecture, and evaluate its performance.

4.1 Considerations

Various factors must be considered when developing an effective and practical framework capable of mitigating the *EpileptiCar* phenomenon in real-world conditions. The following considerations, underlined by insights gained in the analysis described in Section 3, guided the design and implementation of the proposed framework.

(1) Real-time Performance. Although our analysis was performed based on controlled experiments involving stationary vehicles, it nonetheless highlights the importance of timely detection of emergency vehicles in high-speed autonomous driving scenarios when the conditions are less predictable

Figure 9: Samples taken from three authentic collected datasets; a sample from the Berkeley dataset (left), YouTube dataset (center), and DashCam dataset (right).

Figure 10: Samples from the 'Berkeley - Manual Flasher Augmentation' dataset (left) and the 'Berkeley - CycleGAN-based synthesis' dataset (right).

and the consequences can be fatal. Therefore, the required solution must be capable of making split-second decisions to ensure the safety of occupants and other road users. The framework's runtime must be minimized to ensure that objects are identified quickly by object detectors, allowing sufficient time for the advanced driver-assistance system (ADAS) to respond appropriately and avoid a collision. Methods that introduce significant delays may not be suitable for addressing the *EpileptiCar* phenomenon under real-time constraints.

(2) Safety First. Given the severe consequences of misdetection of an emergency vehicle, the framework must prioritize the detection of emergency vehicles (true positive rate) while maintaining a reasonable false positive rate.

(3) Preserving Original Performance. The object detectors in ADASs are responsible for detecting a wide range of objects in the driving environment under diverse driving conditions. Therefore, a solution aimed at mitigating the *EpileptiCar* phenomenon must not compromise object detectors' ability to detect other objects. The required framework must only be activated when the conditions associated with the *EpileptiCar* phenomenon arise, without introducing unnecessary overhead or compromising the original detection performance in detecting other objects.

(4) Limited Dataset Availability. When undertaking this research we found that there was a lack of comprehensive datasets containing images of emergency vehicles with their flashers on at night. To overcome this obstacle and enable the development of a suitable solution, we created augmented datasets which were used to train the models; the models' performance was evaluated with real images of cars obtained from various sources.

(5) Maximal Deployment. Due to the widespread occurrence of the *EpileptiCar* phenomenon, the framework must be easy and cost-effective to deploy and capable of deployment on both new and existing systems. Therefore, we aimed to develop a software-based framework; doing so eliminates the need to install costly dedicated hardware or update existing hardware (e.g., changing the camera lens).

4.2 Datasets

Various datasets were used to develop and evaluate *Caracetamol*. Before we describe the framework, we will first introduce the datasets used to train the machine learning models in it and to evaluate the system. We collected three authentic open-source image/video datasets to evaluate the system and created two augmented datasets to address the lack of datasets containing images that include emergency vehicles with their flashers on at night, and to enhance the framework's robustness and effectiveness in practical driving conditions. The augmented datasets are available online[1] to enable the research community to advance research in this area. Samples from the datasets can be seen in Figs. 9 and 10. We will now elaborate on these datasets.

4.2.1 Authentic Datasets

The datasets utilized in this work are as follows: **(1) YouTube Dataset.** This dataset consists of 243 videos that we downloaded from YouTube, comprising over 30,000 frames. It was utilized to both evaluate *Caracetamol* and facilitate the training of a CycleGAN model that was used to generate synthetic emergency flasher imagery (described below). **(2) DashCam Dataset.** We assembled this dataset from real-world recordings that we obtained by using a 70mai Dash Cam A400. This dataset, consisting of about 900 images, enriched the training data with authentic emergency flasher instances. **(3) BDD100K (Berkeley) Dataset.** The BDD100K dataset [31], also known as the Berkeley dataset, is the largest and most diverse driving video dataset available. The Berkeley dataset was used for fine-tuning the models in *Caracetamol* for car detection. This collection of 100,000 diverse car-related images with rich annotations lacks images of activated emergency flashers at night.

4.2.2 Augmented Datasets

Due to the limited number of relevant images in the Berkeley dataset [31], we created additional datasets containing images with emergency vehicles whose flashers are on at night: 'Berkeley - Manual Flasher Augmentation' and 'Berkeley - CycleGAN-Based Synthesis'. These datasets were created using the Berkeley dataset by performing a process aimed at integrating emergency flasher elements into the nighttime images. We note that these datasets were only used for training specific components of *Caracetamol* where explicitly specified in the training setup. These datasets were not used for testing/evaluating the performance of *Caracetamol* or any baseline object detector. To achieve this, we identified nighttime images in the Berkeley dataset and added the effect of

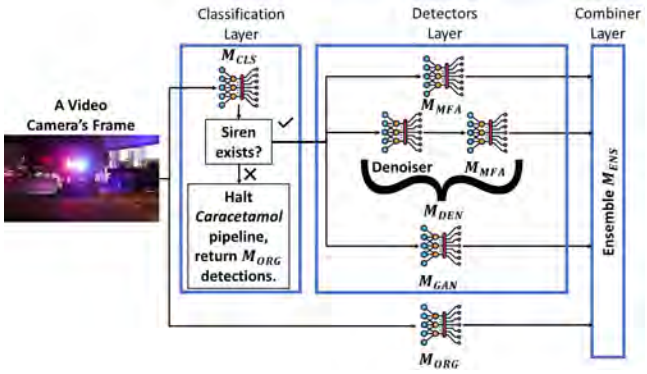

Figure 11: The *Caracetamol* framework, consisting of three layers (classification, detection, and combiner) which are applied to input frames in parallel to the original object detector M_{ORG}. The outputs of M_{ORG} and the detectors layer are aggregated in the combiner layer ensemble model M_{ENS}.

flashers to the identified images. For each image, we used two methods to add the flasher images: (1) a manual code-based process, and (2) a CycleGAN-based approach. Images modified by the first method were included in the 'Berkeley - Manual Flasher Augmentation' dataset and images modified by the second method were included in the 'Berkeley - CycleGAN-Based Synthesis' dataset. Each dataset consisted of: (1) unmodified images from the Berkeley dataset (*i.e.,* all the images not identified as nighttime images) and (2) modified images from the respective modification method (*i.e.,* all the images identified as nighttime images, with flasher augmentations).

Preprocessing. To identify nighttime images in the Berkeley dataset for targeted modification, we employed a method proposed in prior research [32] in which the average brightness of each pixel in the image is calculated to determine whether the image represents a daytime or nighttime scene. Images with an average brightness value above a predefined threshold of 60 are categorized as daytime, while those below the threshold are classified as nighttime. Using this method, we identified approximately 27,500 nighttime images in the Berkeley dataset, to which we added emergency flasher imagery.

Augmentation. Two methods were used to add emergency flasher elements to the nighttime images:

(1) Berkeley - MFA (Manual Flasher Augmentation). In this method, synthetic emergency flasher imagery was added to nighttime images utilizing random noise patterns resembling the lights from emergency flashers. This allowed us to simulate the presence of emergency vehicles in nighttime car-related images and generate more diverse training data (the Python script we used can be found in the Appendix in Listing 1). We used the script to add emergency flasher imagery on 27,500 images from the Berkeley dataset (one for each nighttime image) enriching the dataset with images "containing" an emergency flasher. The color (red or blue) of the flasher was randomly chosen for each image, and it was applied in a random area of the image. The application of the flasher imagery in random areas of the image was performed in order to simulate an attacker placing a siren in random locations, not necessarily on the roof of an emergency vehicle (e.g., in a tree, on the ground, etc.). By performing this modification, we created a new dataset called the 'Berkeley - MFA' dataset.

(2) Berkeley - CycleGAN-Based Synthesis. In this method, an AI flasher generator, which was created using the CycleGAN architecture [33], was used to add emergency flasher imagery to images in the Berkeley dataset. To train the model, two groups of images were used: Group A (without flashers), which consisted of approximately 1,000 images from the Berkeley dataset, and Group B (with flashers), which consisted of approximately 1,000 images from the YouTube dataset. The model was trained for 200 epochs on all the images. This training process produced two models/generators: one generator that adds imagery of emergency flashers (*i.e.,* streaks of light and color) to input images and a second generator, a denoiser, that filters the emergency flasher imagery from input images. We used the first generator to add emergency flasher imagery on 27,500 images (one for each nighttime image) from the Berkeley dataset. By performing this modification, we created a new dataset called the 'Berkeley - CycleGAN-Based Synthesis' dataset. Later in this section we discuss the use of the second generator, which we refer to as the denoiser.

4.3 The Framework

General Overview. *Caracetamol* is a software-based wrapper for existing object detection algorithms aimed at increasing their robustness in the presence of emergency flashers in nighttime settings (an illustration of *Caracetamol*'s architecture can be found in Fig. 11). The system works as follows on each frame processed by an ADAS: (1) the frame passes through the original object detection model M_{ORG} and the predictions are obtained, (2) the same frame is also passed to a lightweight classifier M_{CLS} which detects the presence of a flash. If M_{CLS} does not detect a flash, then the framework outputs the original predictions. Otherwise, the framework executes a series of models which are robust to flashes and revise the original predictions accordingly. This process not only ensures high performance in nominal settings, but also minimizes overhead as well. The robust model is an ensemble of models M_{ENS} which cover different aspects of the data and environment. We found that the use of M_{ENS} is necessary for reliable results (see section 4.4.2). We will now elaborate on each component.

Original Object Detector (M_{ORG}). Every input frame to *Caracetamol* is processed independently by the original ADAS' object detector, M_{ORG}, to preserve the original performance by ensuring that any detections made by M_{ORG} are also made when utilizing the *Caracetamol* framework. If the *Caracetamol* framework is activated for a frame with an emer-

gency flasher, the detections are combined with those from M_{ORG} in the final combiner layer (described below).

Classification Layer (M_{CLS}). The first layer consists of an emergency flasher classifier, M_{CLS} , tasked with classifying frames with emergency vehicles. This initial classification ensures that subsequent layers are only applied when emergency vehicles are detected in the frame. This minimizes runtime by selectively utilizing the other two layers. If M_{CLS} identifies an emergency flasher in an input frame, the frame is processed by both M_{ORG} and the additional *Caracetamol* layers. Otherwise, it is only processed by M_{ORG}.

Detectors Layer. This layer is applied if M_{CLS} identifies the presence of an emergency flasher in a frame. This layer consists of three distinct components which are applied in parallel, each independently processing the frame: **(1) MFA (Manual Flasher Augmentation) Model M_{MFA}:** This component consists of a copy of M_{ORG} fine-tuned on the Berkeley - MFA dataset. **(2) CycleGAN-Based Synthesis Model M_{GAN}:** This component consists of a copy of M_{ORG} fine-tuned on the Berkeley - CycleGAN-based synthesis dataset. **(3) Denoising Model M_{DEN}:** This component consists of a filter created by the second CycleGAN generator described in Section 4.2.2. This filter (the second generator) removes emergency flasher imagery from input images and is applied to denoise the input frame. The denoised frame is then fed to a subcomponent identical to M_{MFA}.

Combiner Layer. This layer consists of the four object detectors (M_{ORG}, as well as M_{MFA}, M_{GAN}, and M_{DEN}, if activated). The results of an object detector take the form of a set of bounding boxes and confidence scores. *Caracetamol*'s final layer aggregates these results using the NMS (non-maximum suppression) algorithm. This is the standard algorithm used to aggregate bounding boxes, however, here we apply it across all of the models' predictions. The algorithm merges co-occurrences of the same object by examining whether the intersection over union of the resulting bounding boxes exceeds a given threshold (we found a threshold of 0.7 yields good results). The final output of this layer is either (1) M_{ORG}'s bounding boxes if no flash was detected or (2) an aggregation of all the models' bounding boxes. Therefore, in the event of a flash, any objects which were missed by M_{ORG} are added by the additional detectors. The utilization of an ensemble framework is important because the different models in *Caracetamol* often return different classifications for the same input (see Fig. 14). These disagreements are handled by the ensemble, ensuring all potential detections are considered.

4.4 Evaluation

In this section we evaluate *Caracatmol*'s performance in detecting vehicles with emergency flashers and compare its performance when three different base object detectors (YOLO, SSD, Faster R-CNN) are used. We evaluate the performance of each layer of *Caracatmol*, as well as the added time overhead of the *Caracetamol* framework.

4.4.1 Classification Layer Evaluation

To evaluate the classification layer's performance, we evaluated three emergency flasher classifiers, each with the same architecture but trained on different datasets: (1) trained solely on the YouTube/Berkeley datasets, (2) trained solely on the DashCam dataset, and (3) trained on a combination of images from all three datasets.

Training. The classifier receives 200x200 images and outputs a decision as to whether or not the image contains an emergency flasher. The classifiers' architecture is as follows: (1) two convolutional layers, each followed by a max pooling layer; the extracted features are then flattened into a one-dimensional array, (2) two fully connected layers with ReLU activation and dropout regularization, and (3) a sigmoid-activated output layer for binary classification, determining the presence or absence of an emergency flasher.

The three classifiers had the same architecture (described above), while each was trained on one of the following datasets: (1) a dataset consisting of 833 images from the DashCam dataset (half with flashers and half without flashers), (2) a dataset consisting of 833 images, 335 of which contain flashers from the YouTube dataset and 498 of which do not contain flashers from the Berkeley dataset, and (3) a mixed dataset consisting of the 1,666 images from the previous two datasets (833 from DashCam and 833 from YouTube). In each case, the datasets were split into training/validation/test sets with a 70%/10%/20% split after randomly shuffling the dataset. Each model's training and validation sets were taken from their respective dataset (DashCam, YouTube, or mixed, respectively), while all three models were tested with the mixed test set. Each classifier was trained for 20 epochs with early stopping (*i.e.,* returning the configuration that yielded the best performance on the validation set), using the Adam optimizer with binary cross-entropy and loss.

Metrics. We calculated each classifier's: AUC score, accuracy (given a confidence threshold of 0.5), and false positive rate (FPR) according to a 'safety-first' policy based on a true positive rate (TPR) of 1.0. This policy is important to consider, because in a road safety context, each dangerous situation must be detected and addressed. False negative identifications (even one) can have life-threatening consequences for road users and should be minimized, even at the expense of an elevated FPR.

Results. The ROC curves and AUC scores are presented in Fig. 12, and the accuracy and FPR (according to a policy based on TPR = 1.0) are presented in Table 2. We can see that the mixed model achieves the highest AUC (0.98), the highest accuracy (0.90), and lowest FPR (0.35) when applying this policy, signifying its superior performance in distinguishing between true and false positives. Therefore, we utilize the mixed model in the *Caracatamol* pipeline.

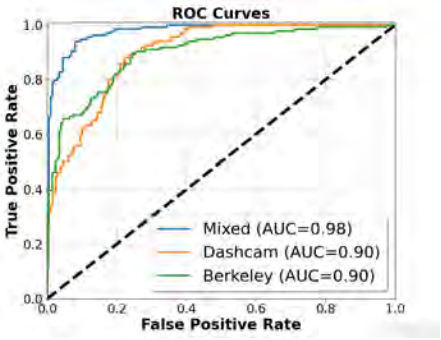

Figure 12: Object detection models' receiver operating characteristic (ROC) curves for three CNN-based object detection models (mixed, Berkeley, and DashCam) when tested on the task of emergency flasher detection. The legend also includes each model's area under the curve (AUC).

Table 2: Object detection models' performance on the examined metrics (accuracy, FPR given a TPR of 1, and TPR and FPR given a confidence threshold of 0.5), for three CNN-based object detection models (mixed, Berkeley, and DashCam) when tested on the task of emergency flasher detection.

Metric	**Mixed**	**DashCam**	**Berkeley**
Accuracy	0.90	0.82	0.81
FPR @ TPR = 1	0.345	0.475	0.975
TPR @ Threshold = 0.5	0.828	0.881	0.701
FPR @ Threshold = 0.5	0.035	0.225	0.12

4.4.2 Detector Layer Evaluation

Here we evaluate the detectors layer, providing an ablation study on the performance of each object detector in M_{ENS}.

Training Setup. We utilized three object detector architectures: YOLOv3, SSD, and Faster R-CNN. The YOLO model was fine-tuned for 50 epochs with early stopping, using the SGD optimizer with a learning rate of 0.1, momentum of 0.9, and weight decay of 0.0005. The SSD model was fine-tuned for 50 epochs with early stopping, using the SGD optimizer with a learning rate of 0.0001, momentum of 0.9, and weight decay of 0.0001. The Faster R-CNN model was fine-tuned for 50 epochs with early stopping, using the SGD optimizer with a learning rate of 0.01, momentum of 0.9, and weight decay of 0.0001. The fine-tuning data for the baseline models was the Berkeley dataset (100,000 images) with a 70/10/20 training/validation/test split. The fine-tuning data for the *Caracetamol* detectors layer models was the Berkeley - 'Manual Flasher Augmentation dataset' dataset and the 'Berkeley - CycleGAN-Based Synthesis dataset' (100,000 images each) with a 70/10/20 training/validation/test split. The fine-tuning of the models in the detectors layer is described below. Training for each model was stopped if 10 consecutive epochs failed to yield an improvement of at least 0.01 in the performance in terms of the average confidence score.

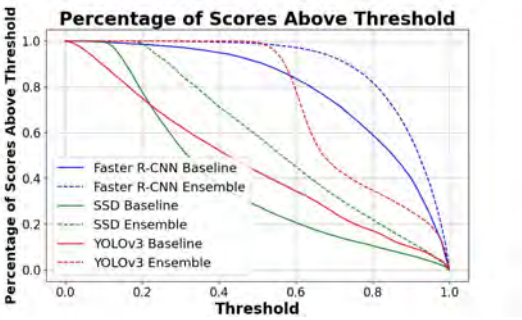

Figure 13: Comparison between the performance of the baseline object detectors and their respective implementation of the Caracatamol (the results of M_{ENS}).

Metrics. The metrics used in this evaluation are as follows: (1) average object detector confidence (average performance), (2) minimum confidence (lower bound of performance), (3) maximum confidence (upper bound of performance), (4) confidence range (stability of performance), and (5) percentage of detections with confidence scores above 0.5/0.6/0.7/0.8 (distribution of high confidence detections).

Performance Breakdown. For each architecture, we compare the performance on the following models: (1) the baseline object detector fine-tuned on the Berkeley dataset (M_{ORG}), (2) a version of M_{ORG} fine-tuned on the Berkeley - 'Manual Flasher Augmentation dataset' (M_{MFA}), (3) a version of M_{ORG} fine-tuned on the 'Berkeley - CycleGAN-Based Synthesis dataset' (M_{GAN}), (4) a version of M_{ORG} fine-tuned on the 'Berkeley - Manual Flasher Augmentation' dataset with a denoising preprocessor (M_{DEN}), and (5) an ensemble of the four models listed above M_{ENS}.

Results. We used the YouTube data test set to assess and compare the performance of three possible implementations of the *Caracetamol* detectors layer, with the three object detectors (YOLO, SSD, and Faster R-CNN) each serving as base object detectors (M_{ORG} in the framework). The results for YOLO, SSD, and Faster R-CNN are presented in Table 3 and in Fig. 13. As can be seen from the results presented in Table 3, each metric is improved by fine-tuning the original model on the artificial data. On average, M_{ENS} improved the average confidence by 0.21, the minimum confidence by 0.27, the maximum confidence by 0.07, and the percentage of detections with confidence scores above 0.5/0.6/0.7/0.8 by 0.33/0.34/0.31/0.19. The results indicate improved robustness with respect to the baseline (M_{ORG}) in the presence of an activated flasher. It also decreased the confidence range by 0.19, indicating improved stability with respect to the baseline.

Fig. 13 also shows an improvement in the overall detection rates of M_{ENS} with respect to the baseline (M_{ORG}) whose detection rates appear below the refined models in all three evaluated object detectors for all tested detection thresholds. This indicates an improved detection probability in time of emergency vehicles in the presence of an activated flasher/siren.

Table 3: Comparison between the performance of Caracatamol based on three object detectors (YOLO, SSD, and Faster R-CNN) for each model in the detectors layer separately(M_{ORG}, M_{MFA}, M_{GAN}, and M_{DEN}) as well as in various combinations (M_{MFA} + M_{GAN}, M_{MFA} + M_{DEN}, and M_{GEN} + M_{DEN}) and in a full ensemble M_{ENS}.

Yolo	M_{ORG}	M_{MFA}	M_{GAN}	M_{DEN}	M_{MFA} + M_{GAN}	M_{MFA} + M_{DEN}	M_{GAN} + M_{DEN}	M_{ENS}
Average Confidence	0.46	0.54	0.53	0.71	0.70	0.71	0.79	0.79
Min Confidence	0.16	0.14	0.20	0.59	0.38	0.60	0.64	0.64
Max Confidence	0.83	0.66	0.89	0.82	0.91	0.82	0.93	0.93
Range (Max Conf – Min Conf)	0.67	0.52	0.69	0.22	0.52	0.22	0.29	0.28
% of Data with Confidence Above 0.5	0.43	0.89	0.52	0.99	0.94	0.99	0.99	0.99
Above 0.6	0.35	0.38	0.46	0.94	0.68	0.95	0.96	0.97
Above 0.7	0.27	0.007	0.40	0.62	0.40	0.62	0.79	0.79
Above 0.8	0.20	0.0002	0.33	0.06	0.33	0.06	0.38	0.38

SSD	M_{ORG}	M_{MFA}	M_{GAN}	M_{DEN}	M_{MFA} + M_{GAN}	M_{MFA} + M_{DEN}	M_{GAN} + M_{DEN}	M_{ENS}
Average Confidence	0.40	0.39	0.31	0.56	0.42	0.56	0.53	0.58
Min Confidence	0.16	0.16	0.18	0.23	0.20	0.23	0.23	0.26
Max Confidence	0.83	0.83	0.70	0.93	0.85	0.93	0.93	0.94
Range (Max Conf - Min Conf)	0.67	0.67	0.52	0.70	0.65	0.69	0.70	0.68
% of Data with Confidence Above 0.5	0.27	0.26	0.13	0.56	0.30	0.56	0.50	0.59
Above 0.6	0.20	0.19	0.09	0.42	0.22	0.42	0.36	0.45
Above 0.7	0.15	0.15	0.07	0.31	0.16	0.30	0.25	0.33
Above 0.8	0.11	0.11	0.05	0.21	0.12	0.20	0.16	0.22

Faster R-CNN	M_{ORG}	M_{MFA}	M_{GAN}	M_{DEN}	M_{MFA} + M_{GAN}	M_{MFA} + M_{DEN}	M_{GAN} + M_{DEN}	M_{ENS}
Average Confidence	0.77	0.80	0.75	0.81	0.82	0.85	0.84	0.89
Min Confidence	0.44	0.47	0.41	0.45	0.51	0.55	0.55	0.66
Max Confidence	0.96	0.97	0.96	0.98	0.97	0.98	0.97	0.98
Range (Max Conf - Min Conf)	0.51	0.49	0.54	0.52	0.46	0.42	0.42	0.31
% of Data with Confidence Above 0.5	0.89	0.91	0.85	0.94	0.93	0.96	0.96	0.99
Above 0.6	0.81	0.85	0.76	0.89	0.87	0.93	0.92	0.97
Above 0.7	0.70	0.75	0.63	0.79	0.77	0.85	0.85	0.93
Above 0.8	0.56	0.61	0.49	0.61	0.63	0.71	0.71	0.83

Figure 14: Inputs (disagreements) which were detected with different confidence scores by our models in M_{ENS}.

4.4.3 Runtime Analysis

In order to be used in real-time ADASs, our solution must exhibit real-time performance. We evaluate the runtime of each stage of the *Caracetamol* pipeline, to ensure that the solution is efficient and capable of operating within the constraints of automotive hardware in real-world driving situations.

Experimental Setup: We fed 10,000 images from the Berkeley dataset to each layer/component of the *Caracetamol* framework using a baseline model M_{ORG} and detectors layer models derived from the YOLOv3 object detector. We loaded each component on a GPU and calculated the average runtime of each component on each image. We repeated this evaluation for: RTX-2080, RTX-3090, RTX-6000.

Table 4: Execution time and average confidence of YOLO-based *Caracetamol* components, loaded on various GPUs.

GPU Model	M_{ORG}	M_{DEN}	M_{GAN}	M_{MFA}	M_{ENS}	M_{CLS}+M_{ENS}
RTX-6000	16ms	97ms	16ms	16ms	129ms	129 ms
RTX-3090	21ms	113ms	21ms	21ms	154ms	171ms
RTX-2080	34ms	165ms	34ms	33ms	233ms	248ms
Avg Score	0.36	0.71	0.53	0.54	0.79	0.79

Results: As can be seen from the results in Table 4, on the RTX-6000 GPU, the runtime of the entire pipeline is 129 ms. This means that *Caracetamol* can process 7 frames per second (FPS) during a flash, while maintaining high performance (average detection of 0.79).

4.4.4 Robustness against Adversaries

In this section we discuss the robustness of *Caracetamol* against potential adversaries aiming to attack its functionality.

One potential attack vector would be to utilize adversarial perturbations, however due to the physical-world setting of *Caracetamol*, perturbations cannot be applied to emergency flashers. This prevents an attacker from introducing adversarial perturbations in *Caracetamol*'s input. In addition, the utilization of an ensemble framework in *Caracetamol* signifi-

cantly increases the difficulty of utilizing adversarial perturbations, since the framework is able to handle disagreements between the ensemble's models. In Fig. 14 we can see that the models that compose *Caracetamol* disagree on many inputs, which are resolved by the ensemble M_{ENS}.

An additional attack vector would be to perform a Denial-of-Service attack, which would result in delayed responses from the ADAS and prevent timely responses. This attack vector isn't practical since *Caracetamol*'s framework is designed to: (1) not waste computational resources when no flashers are detected, and (2) run at a frame-rate that doesn't burden computational resources when flashers are detected (as presented in Sec. 4.4.3). In both cases, *Caracetamol* doesn't impose unsustainable computational burdens which would negatively impact an ADAS' functionality.

5 Limitations

We performed this research independently, with no collaboration with the automotive industry. While this allowed us to avoid a conflict of interest preventing us from publishing our findings, some limitations with our work could have been overcome with the automotive industry's involvement.

Lack of Access to Reports on the Relevant Accidents. The reports on the 16 incidents that include the footage obtained by Tesla's video cameras before the accidents occurred and the behavior of the relevant object detectors are not available online. We performed the most comprehensive research possible given the lack of access to confidential footage, reports, and analysis, relying purely on the limited data and information that is publicly available via the Internet.

Difference in the Tested Components. We note that in our research, we used commercial ADASs to obtain video footage and four commonly used object detectors to conduct the analysis. The transferability of our findings to Tesla must be examined. We hope that this research will motivate the automotive industry to validate our findings on real advanced driver-assistance systems (ADASs).

Inability to Explore and Understand the Full Semi-Autonomous Cars Context. While our research revealed the limitation of object detectors in the presence of activated emergency flashers, other technical factors must also be analyzed to understand the full picture and explore the potential interplay of various factors in the context of ADASs. For example, the reason why the integrated RADAR did not detect the parked emergency vehicles should also be analyzed. While some explanations regarding this can be found in Volvo's manual [34], it is our hope that a detailed report covering all of the technological aspects will be published and shed light on the perfect storm that caused Teslas to crash into emergency vehicles.

Limited Availability of Datasets Containing Emergency Vehicles. One challenge we faced in designing our framework was the limited number of nighttime images of emergency vehicles with activated emergency flashers. While we overcame this limitation by augmenting the training data, we believe that improved detection confidence results could be obtained by fine-tuning the object detector using actual images of emergency vehicles with activated flashers at night.

6 Related Work

In recent years, researchers have investigated the robustness and limitations of advanced driver-assistance systems (ADASs). One line of research investigated the robustness and limitations of the various sensors incorporated in ADASs. This includes studies that tested the limitations of LiDAR [35–40] and video cameras [41–43]. A second avenue of research investigated the robustness and limitations of object detectors [44–48]. Other studies used simulators to investigate the implications of attacks [38, 49]. These studies above shed light on the limitations of object detection algorithms and sensors. However, a recent SoK [50] argued that the findings of the abovementioned studies might not apply to real ADASs, because *"AI component-level errors do not necessarily lead to system-level effects (e.g., vehicle collisions)"*.

A third area of research investigated the robustness and limitations of real ADASs (e.g., Tesla, Mobileye, HARMAN's ADAS, and OpenPilot). This research includes studies that examined the limitations of real ADASs, using attacks that triggered an undesired reaction from the controlled cars. A few studies emphasized the limitations of an ADAS' video camera to time domain attacks [51, 52], projections [51, 53], infrared light [48, 54], laser beams [35], and physical patches [55–60]. Other studies investigated the limitations of an ADAS' RADAR [61, 62], ultrasonic sensors [62], and GPS sensor [63].

7 Discussion

The objective of this research was to shed light on the technical factors underlying the documented incidents in which semi-autonomous cars crashed into emergency vehicles, and it was performed in response to the lack of empirical research on this topic and the limited information available on the Internet. By conducting experiments independently and in a controlled environment using video footage obtained by commercial ADAS, we found that activated emergency flashers degraded object detectors' performance and proposed a framework aimed at securing object detectors in the presence of activated emergency flashers. We hope our findings will motivate the automotive industry, NHTSA, and other stakeholders to validate our findings on additional ADASs and cars.

References

[1] NHTSA, "Odi resume," https://static.nhtsa.gov/odi/inv/2022/INOA-EA22002-3184.PDF.

[2] C. ATIYEH, "Tesla Driver Was on Autopilot Eating a Bagel When He Smashed into a Fire Truck," https://www.caranddriver.com/news/a28911259/tesla-crash-california-autopilot-driver-ntsb/, 2019.

[3] J. Klawans, "Tesla Autopilot Crashes Into Police Car, Hits Cop as Driver 'Watches Movie'," https://www.newsweek.com/tesla-autopilot-crashes-police-car-hits-cop-driver-watches-movie-1677814, 2022.

[4] J. Titcomb, "Investigation launched after Tesla cars crash into ambulances while on Autopilot," https://www.telegraph.co.uk/technology/2021/08/16/investigation-launched-tesla-cars-crash-ambulances-autopilot/, 2021.

[5] A. N. Angie Moreschi and L. Deal, "Mystery Accidents: Teslas in 'Autopilot' crashing into emergency vehicles," https://cbsaustin.com/news/spotlight-on-america/responders-at-risk-nhtsa-probes-driver-assistance-systems-after-a-series-of-crashes-involving-teslas-and-emergency-vehicles, 2023.

[6] B. Templeton, "Teslas Are Crashing Into Emergency Vehicles Too Much, So NHTSA Asks Other Car Companies About It," https://www.forbes.com/sites/bradtempleton/2021/09/20/teslas-are-crashing-into-emergency-vehicles-too-much-so-nhtsa-asks-other-car-companies-about-it/, 2021.

[7] "Tesla vision update: Replacing ultrasonic sensors with tesla vision," https://www.tesla.com/support/transitioning-tesla-vision.

[8] "Hp." [Online]. Available: https://www.amazon.com/dp/B081YDHHBR/

[9] "Pelsee." [Online]. Available: https://www.amazon.com/dp/B0BF4XB3VP/

[10] "Azdome." [Online]. Available: https://www.amazon.com/dp/B094YDVV7L/

[11] "Imagebon." [Online]. Available: https://www.amazon.com/dp/B0C86CV679/

[12] "Rexing." [Online]. Available: https://www.amazon.com/dp/B08N1KMSZ7/

[13] "Tesla dashcam footage suggests reasons for autopilot crashes." [Online]. Available: https://youtu.be/V2u3dcH2VGM

[14] Y. Kim, R. Daly, J. Kim, C. Fallin, J. H. Lee, D. Lee, C. Wilkerson, K. Lai, and O. Mutlu, "Flipping bits in memory without accessing them: An experimental study of dram disturbance errors," in *2014 ACM/IEEE 41st International Symposium on Computer Architecture (ISCA)*, 2014, pp. 361–372.

[15] K. Razavi, B. Gras, E. Bosman, B. Preneel, C. Giuffrida, and H. Bos, "Flip feng shui: Hammering a needle in the software stack," in *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, Aug. 2016, pp. 1–18. [Online]. Available: https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/razavi

[16] V. van der Veen, Y. Fratantonio, M. Lindorfer, D. Gruss, C. Maurice, G. Vigna, H. Bos, K. Razavi, and C. Giuffrida, "Drammer: Deterministic rowhammer attacks on mobile platforms," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1675–1689. [Online]. Available: https://doi.org/10.1145/2976749.2978406

[17] P. Frigo, C. Giuffrida, H. Bos, and K. Razavi, "Grand pwning unit: Accelerating microarchitectural attacks with the gpu," in *Proceedings - 2018 IEEE Symposium on Security and Privacy, SP 2018*. United States: Institute of Electrical and Electronics Engineers Inc., 2018, pp. 195–210, 39th IEEE Symposium on Security and Privacy, SP 2018 ; Conference date: 21-05-2018 Through 23-05-2018.

[18] D. Gruss, M. Lipp, M. Schwarz, D. Genkin, J. Juffinger, S. O'Connell, W. Schoechl, and Y. Yarom, "Another flip in the wall of rowhammer defenses," in *39th IEEE Symposium on Security and Privacy 2018*, Jan. 2018.

[19] P. Pessl, D. Gruss, C. Maurice, M. Schwarz, and S. Mangard, "DRAMA: Exploiting DRAM addressing for Cross-CPU attacks," in *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, Aug. 2016, pp. 565–581. [Online]. Available: https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/pessl

[20] C.-Y. Wang and H.-Y. M. Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," 2024.

[21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[22] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.

[23] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[24] J. Dey and S. Pasricha, "Robust perception architecture design for automotive cyber-physical systems," in *2022 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE, 2022, pp. 241–246.

[25] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu *et al.*, "Mmdetection: Open mmlab detection toolbox and benchmark," *arXiv preprint arXiv:1906.07155*, 2019.

[26] D. Gruyer, V. Magnier, K. Hamdi, L. Claussmann, O. Orfila, and A. Rakotonirainy, "Perception, information processing and modeling: Critical stages for autonomous driving applications," *Annual Reviews in Control*, vol. 44, pp. 323–341, 2017.

[27] J. Cao, X. Weng, R. Khirodkar, J. Pang, and K. Kitani, "Observation-centric sort: Rethinking sort for robust multi-object tracking," *arXiv preprint arXiv:2203.14360*, 2022.

[28] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 3464–3468.

[29] Y. Zhang, P. Sun, Y. Jiang, D. Yu, Z. Yuan, P. Luo, W. Liu, and X. Wang, "Bytetrack: Multi-object tracking by associating every detection box," 2021.

[30] "Extech." [Online]. Available: https://www.amazon.com/Extech-HD450-Datalogging-Heavy-Light/dp/B003N3UOCK

[31] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving dataset for heterogeneous multitask learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2636–2645.

[32] F. E. Sandnes, "Towards calibration-free geolocalization of stationary outdoor webcams," 2010.

[33] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.

[34] "Why tesla's autopilot can't see a stopped firetruck." [Online]. Available: https://www.wired.com/story/tesla-autopilot-why-crash-radar/

[35] J. Petit, B. Stottelaar, M. Feiri, and F. Kargl, "Remote attacks on automated vehicles sensors: Experiments on camera and lidar," *Black Hat Europe*, vol. 11, p. 2015, 2015.

[36] Y. Cao, C. Xiao, B. Cyr, Y. Zhou, W. Park, S. Rampazzi, Q. A. Chen, K. Fu, and Z. M. Mao, "Adversarial sensor attack on lidar-based perception in autonomous driving," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 2019, pp. 2267–2281.

[37] J. Sun, Y. Cao, Q. A. Chen, and Z. M. Mao, "Towards robust lidar-based perception in autonomous driving: General black-box adversarial sensor attack and countermeasures," in *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, Aug. 2020, pp. 877–894.

[38] Y. Cao, N. Wang, C. Xiao, D. Yang, J. Fang, R. Yang, Q. A. Chen, M. Liu, and B. Li, "Invisible for both camera and lidar: Security of multi-sensor fusion based perception in autonomous driving under physical-world attacks," in *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2021, pp. 176–194.

[39] Y. Cao, S. H. Bhupathiraju, P. Naghavi, T. Sugawara, Z. M. Mao, and S. Rampazzi, "You can't see me: Physical removal attacks on {LiDAR-based} autonomous vehicles driving frameworks," in *32nd USENIX Security Symposium (USENIX Security 23)*, 2023, pp. 2993–3010.

[40] T. Sato, Y. Hayakawa, R. Suzuki, Y. Shiiki, K. Yoshioka, and Q. A. Chen, "Lidar spoofing meets the newgen: Capability improvements, broken assumptions, and new attack strategies," *arXiv preprint arXiv:2303.10555*, 2023.

[41] C. Yan, Z. Xu, Z. Yin, X. Ji, and W. Xu, "Rolling colors: Adversarial laser exploits against traffic light recognition," *arXiv preprint arXiv:2204.02675*, 2022.

[42] G. Lovisotto, H. Turner, I. Sluganovic, M. Strohmeier, and I. Martinovic, "{SLAP}: Improving physical adversarial examples with {Short-Lived} adversarial perturbations," in *30th USENIX Security Symposium (USENIX Security 21)*, 2021, pp. 1865–1882.

[43] X. Ji, Y. Cheng, Y. Zhang, K. Wang, C. Yan, W. Xu, and K. Fu, "Poltergeist: Acoustic adversarial machine learning against cameras and computer vision," in *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2021, pp. 160–175.

[44] D. Song, K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, F. Tramer, A. Prakash, and T. Kohno, "Physical adversarial examples for object detectors," in *12th USENIX workshop on offensive technologies (WOOT 18)*, 2018.

[45] S.-T. Chen, C. Cornelius, J. Martin, and D. H. P. Chau, "Shapeshifter: Robust physical adversarial attack on faster r-cnn object detector," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2018, pp. 52–68.

[46] Y. Zhao, H. Zhu, R. Liang, Q. Shen, S. Zhang, and K. Chen, "Seeing isn't believing: Towards more robust adversarial attack against real world object detectors," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '19. New York, NY, USA: ACM, 2019, pp. 1989–2004.

[47] A. Zolfi, M. Kravchik, Y. Elovici, and A. Shabtai, "The translucent patch: A physical and universal attack on object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15 232–15 241.

[48] T. Sato, S. H. V. Bhupathiraju, M. Clifford, T. Sugawara, Q. A. Chen, and S. Rampazzi, "Invisible reflections: Leveraging infrared laser reflections to target traffic sign perception," *arXiv preprint arXiv:2401.03582*, 2024.

[49] J. Shen, J. Y. Won, Z. Chen, and Q. A. Chen, "Drift with devil: Security of {Multi-Sensor} fusion based localization in {High-Level} autonomous driving under {GPS} spoofing," in *29th USENIX security symposium (USENIX Security 20)*, 2020, pp. 931–948.

[50] J. Shen, N. Wang, Z. Wan, Y. Luo, T. Sato, Z. Hu, X. Zhang, S. Guo, Z. Zhong, K. Li *et al.*, "Sok: On the semantic ai security in autonomous driving," *arXiv preprint arXiv:2203.05314*, 2022.

[51] B. Nassi, Y. Mirsky, D. Nassi, R. Ben-Netanel, O. Drokin, and Y. Elovici, "Phantom of the adas: Securing advanced driver-assistance systems from split-second phantom attacks," in *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*, 2020, pp. 293–308.

[52] B. Nassi, Y. Mirsky, J. Shams, R. Ben-Netanel, D. Nassi, and Y. Elovici, "Protecting autonomous cars from phantom attacks," *Communications of the ACM*, vol. 66, no. 4, pp. 56–69, 2023.

[53] D. Nassi, R. Ben-Netanel, Y. Elovici, and B. Nassi, "Mobilbye: attacking adas with camera spoofing," *arXiv preprint arXiv:1906.09765*, 2019.

[54] W. Wang, Y. Yao, X. Liu, X. Li, P. Hao, and T. Zhu, "I can see the light: Attacks on autonomous vehicles using invisible lights," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021, pp. 1930–1944.

[55] T. Sato, J. Shen, N. Wang, Y. Jia, X. Lin, and Q. A. Chen, "Dirty road can attack: Security of deep learning based automated lane centering under {Physical-World} attack," in *30th USENIX Security Symposium (USENIX Security 21)*, 2021, pp. 3309–3326.

[56] keen labs, "Tencent keen security lab: Experimental security research of tesla autopilot," https://keenlab.tencent.com/en/2019/03/29/Tencent-Keen-Security-Lab-Experimental-Security-Research-of-Tesla-Autopilot/, 2019.

[57] N. Morgulis, A. Kreines, S. Mendelowitz, and Y. Weisglass, "Fooling a real car with adversarial traffic signs," *arXiv preprint arXiv:1907.00374*, 2019.

[58] B. by McAfee, "Model Hacking ADAS to Pave Safer Roads for Autonomous Vehicles," https://www.mcafee.com/blogs/other-blogs/mcafee-labs/model-hacking-adas-to-pave-safer-roads-for-autonomous-vehicles/, 2020.

[59] P. Jing, Q. Tang, Y. Du, L. Xue, X. Luo, T. Wang, S. Nie, and S. Wu, "Too good to be safe: Tricking lane detection in autonomous driving with crafted perturbations," in *30th USENIX Security Symposium (USENIX Security 21)*, 2021, pp. 3237–3254.

[60] B. Nassi, J. Shams, R. B. Netanel, and Y. Elovici, "badvertisement: Attacking advanced driver-assistance systems using print advertisements," in *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2022, pp. 376–383.

[61] D. Hunt, K. Angell, Z. Qi, T. Chen, and M. Pajic, "Madradar: A black-box physical layer attack framework on mmwave automotive fmcw radars," *arXiv preprint arXiv:2311.16024*, 2023.

[62] C. Yan, W. Xu, and J. Liu, "Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicle," *DEF CON*, vol. 24, 2016.

[63] Regulus, "Tesla model 3 spoofed off the highway - regulus navigation system hack causes car to turn on its own," https://www.regulus.com/blog/tesla-model-3-spoofed-off-the-highway-regulus-researches-hack-navigation-system-causing-car-to-steer-off-road/.

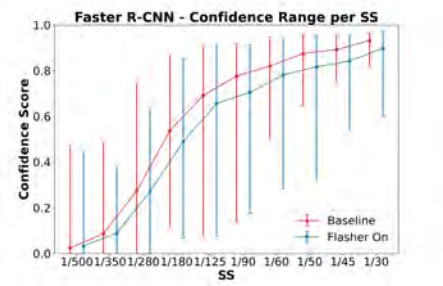

Figure 15: Faster R-CNN confidence score signal range for each examined shutter speed (SS). The results obtained when the emergency flasher was on are presented in blue, while the results obtained when the emergency flasher was off are in red; the points in each range indicate the average value for each signal.

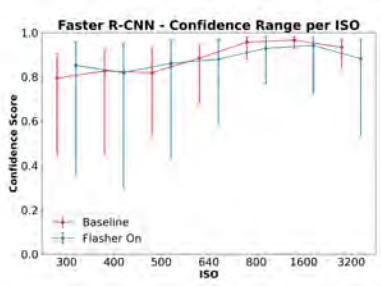

Figure 16: Faster R-CNN confidence score signal range for each examined ISO value. The results obtained when the emergency flasher was on are presented in blue, while the results obtained when the emergency flasher was off are in red; the points in each range indicate the average value for each signal.

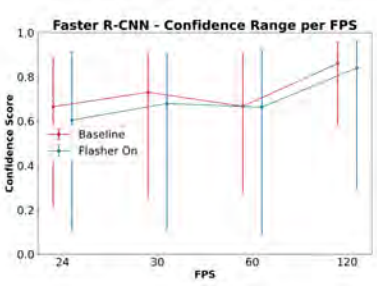

Figure 17: Faster R-CNN confidence score signal range for each examined FPS value. The results obtained when the emergency flasher was on are presented in blue, while the results obtained when the emergency flasher was off are in red; the points in each range indicate the average value for each signal.

8 Appendix A: Effect of the Camera Settings

Here we examine whether advanced driver-assistance system (ADAS) camera settings influence the *EpileptiCar* phenomenon, focusing on: (1) the shutter speed, (2) the ISO sensitivity, and (3) the frame rate of the vehicle's camera. Due to space limitations, we only present the analysis performed using Faster R-CNN.

8.1 Effect of the Camera's Shutter Speed

In this experiment, we investigate whether the shutter speed of an ADAS' camera affects the confidence score range of object detectors when detecting a vehicle whose emergency flasher is on and off.

Experimental Setup: A Samsung Galaxy S22 Ultra was positioned five meters away from a grey Ford Fiesta with an emergency flasher mounted on its roof in a dark setting. Ten 30-second videos of the car were recorded with the flasher on and with the flasher off, varying the camera's shutter speed (1/500, 1/350, 1/250, 1/180, 1/125, 1/90, 1/60, 1/50, 1/45, 1/30 seconds). Faster R-CNN was applied to the video frames, and for each video, we assessed the confidence score range of the object detectors when detecting the car. This experiment was conducted twice: first with the emergency flasher on and again when it was off.

Results: The results of this analysis for the Faster R-CNN detector are presented in Fig. 15. As can be seen, a decrease in the camera shutter speed leads to reduced stability in the detection confidence and an increase in the confidence score range.

Insight 10: *As the camera's shutter speed decreases, the effect of the EpileptiCar phenomenon increases.*

8.2 Effect of the Camera's ISO

Here we investigate whether the ISO setting of an ADAS' camera influences the confidence score range of object detectors when detecting emergency vehicles whose emergency flashers are on and off.

Experimental Setup: A Samsung Galaxy S22 Ultra was positioned five meters away from a grey Ford Fiesta with an emergency flasher mounted on its roof in a dark setting. Seven 30-second videos of the car were recorded with the flasher on and with the flasher off, varying the camera's ISO setting in each video (300, 400, 500, 640, 800, 1600, 3200). Faster R-CNN was applied, and for each video we assessed the confidence score range of the object detectors when detecting the car. This experiment was conducted twice: first with the emergency flasher on and again when it was off.

Results: The results of the Faster R-CNN object detector are presented in Fig. 16. As can be seen, higher camera ISO values correspond to increased stability in object detection confidence and a smaller confidence range.

Insight 11: *The EpileptiCar phenomenon's effect decreases as the camera's ISO value increases.*

8.3 Effect of the Camera's Frame Rate

In this experiment, we investigate whether the camera's frame rate has an impact on the confidence score range of object detectors when detecting a vehicle whose emergency flasher is on and off.

Experimental Setup: A Samsung Galaxy S22 Ultra was positioned five meters away from a Ford Fiesta with an emergency flasher mounted on its roof in a dark setting. Four 30-second videos of the car were recorded with the flasher on and with the flasher off, varying the camera's frame rate

in each video (24, 30, 60, 120 FPS). Faster R-CNN was applied to the video frames, and for each video we assessed the confidence range of the object detectors when detecting the car. This experiment was performed twice: first with the emergency flasher on and again when it was off.

Results: The results of this analysis for the Faster R-CNN detector are presented in Fig. 17. As can be seen, the confidence range of the detectors remains unchanged despite variations in the FPS value.

Insight 12: *he EpileptiCar phenomenon is not affected by the camera's frame rate.*

9 Appendix B: Effect of Emergency Vehicle's Characteristics

In this section we examine the influence of an emergency vehicle's characteristics on the *EpileptiCar* phenomenon, specifically investigating the effects of (1) the color of the emergency vehicle, (2) the color of the emergency flasher, and (3) the orientation of the vehicle requiring detection relative to the object detector/camera. Due to space limitations, we only present the analysis performed using Faster R-CNN.

9.1 Effect of Emergency Vehicle Color

Here we investigate whether the color of the emergency vehicle impacts the detection capabilities of object detectors, analyzing their confidence range when detecting a vehicle whose emergency flasher is on.

Experimental Setup: A Samsung Galaxy S22 Ultra was positioned five meters away from four vehicles (a white Toyota RAV4, a silver Toyota RAV4, a yellow Fiat Panda, and a red Peugeot 207), each with an emergency flasher mounted and activated on the roof. A 30-second video (24 FPS) of each car was recorded in a dark environment. Faster R-CNN was applied to the video frames, and an FFT (fast Fourier transform) graph was extracted from each detector's confidence score signal.

Results: The results for the Faster R-CNN detector are presented in Fig. 18. As can be seen, a consistent frequency peak of approximately 1.3 Hz was maintained for the confidence score signals of all car colors examined.

Insight 13: *The EpileptiCar phenomenon is present in cars with various colors.*

9.2 Effect of Emergency Flasher Color

Here we investigate whether the color of an activated emergency flasher impacts the detection capabilities of object detectors, analyzing their confidence range when detecting a vehicle whose emergency flasher is on.

Experimental Setup: A Samsung Galaxy S22 Ultra was positioned five meters away from a grey Ford Fiesta with an emergency flasher mounted on its roof. We recorded a 30-second video of the car at a rate of 24 FPS. Faster R-CNN was applied to the frames of each video, and the confidence range of each detector when recognizing the vehicle was measured for each video. This experiment was conducted four times: first with the emergency flasher turned off, and then three additional times with the emergency flasher activated, using different colored flashers (blue, yellow, red).

Results: The results of this analysis for Faster R-CNN are presented in Fig. 19. As can be seen, (1) detection confidence was higher when the emergency flasher was off, and (2) the confidence ranges were much larger when the emergency flashers were activated, regardless of color.

Insight 14: *The EpileptiCar phenomenon is not affected by the color of the emergency flasher.*

9.3 Effect of Emergency Vehicle Orientation

Here we investigate how an emergency vehicle's orientation relative to the object detector/camera impacts the performance of object detectors in the presence of an activated emergency flasher.

Experimental Setup: A Samsung Galaxy S22 Ultra was positioned five meters away from a grey Ford Fiesta with an emergency flasher mounted on its roof. Five 30-second videos were recorded of the car; in each video, the angle between the camera and the front of the car was different (90, 135, 180, 225, 270 degrees). Four pretrained object detectors (YOLOv3, SSD, RetinaNet, Faster R-CNN) were applied to the frames from each video, and the confidence range of each detector when recognizing the vehicle was measured for each video. This experiment was conducted twice: first with the emergency flasher on and again when it was off.

Results: The results of this analysis for the Faster R-CNN detector are presented in Fig. 20. As can be seen, the detection confidence was at or near its peak when there was a 225 degree angle between the camera and the car. The analysis showed that the confidence range tended to increase, indicating decreased stability in the object detector's confidence, when: (1) the camera and the car were positioned at angles other than 225 degrees, and (2) the emergency flasher was on.

Insight 15: *The EpileptiCar phenomenon is present in all analyzed orientations between the ADAS camera and the car requiring detection.*

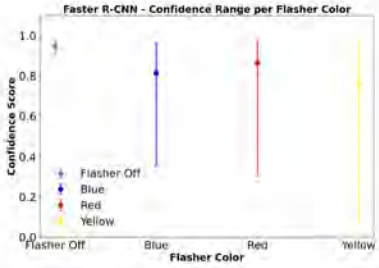

Figure 18: Faster R-CNN confidence score signal range for each examined vehicle color, as well as when the flasher was off. The points in each range indicate the average value for each signal.

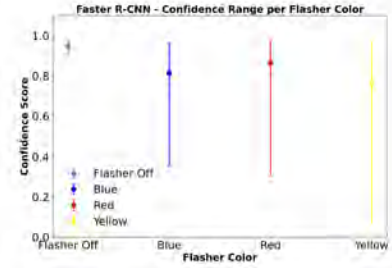

Figure 19: Faster R-CNN confidence score signal range for each examined flasher color, as well as when the flasher was off. The points in each range indicate the average value for each signal.

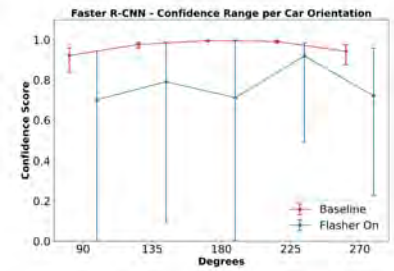

Figure 20: Faster R-CNN confidence score signal range for each examined orientation between an ADAS' camera and an observed vehicle. The results obtained when the emergency flasher was on and off are presented in blue and red, respectively; the points in each range indicate the average value for each signal.

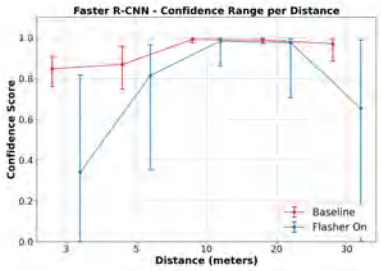

Figure 21: Left: Faster R-CNN confidence score signal range for each examined distance between an autonomous vehicle's camera and an observed vehicle. The results obtained when the emergency flasher was on are presented in blue, while results obtained when the emergency flasher was off are in red; the points in each range indicate the average value for each signal. A car is not detected in the frames associated with the lowest confidence scores of Faster R-CNN in the videos recorded from 3 meters (center) and 30 meters (right).

10 Appendix C: Effect of Camera Distance

Here we analyze the effect of the distance between the semi-autonomous car's camera and the emergency vehicle.

Experimental Setup: A Samsung Galaxy S22 Ultra was positioned in front of a grey Ford Fiesta with an emergency flasher mounted on its roof. Five experiments were conducted, each with the camera at a different distance from the car: 3 meters away (where the car occupies 74% of the frame), 5 meters away (where the car occupies 42% of the frame), 10 meters away (where the car occupies 7% of the frame), 20 meters away (where the car occupies 1.9% of the frame), and 30 meters away (where the car occupies 0.9% of the frame). In each experiment, we recorded a 60-second video in which the emergency flasher was off for the first 30 seconds and on for the last 30 seconds. Four object detectors were applied to the frames. The confidence score range for car detection was measured as a function of the distance.

Results: The results for the Faster R-CNN object detector are presented in Fig. 21. The results for YOLOv3, SSD, and RetinaNet are presented in Fig. 25 in the appendix. As can be seen from the results, the average of the confidence range behaves in a parabolic manner with a peak score at a distance of 10 meters and decreased performance between 3-10 meters (because the flasher significantly saturates the image at closer distances) and 10-30 meters (because the car cannot be seen in the picture due to the flasher's effect at greater distances). The frames associated with the lowest confidence scores in the ranges associated with 3 and 30 meters are presented in Fig. 21.

Insight 16: *The EpileptiCar phenomenon is dependent on the distance between the autonomous car's camera and the vehicle requiring detection.*

11 Appendix D: Additional Material

```
def make_light(x, y, img_shape, option):
    r_w, r_h = option_dict[option]
    channel = 0 if np.random.rand() >
        0.5 else 2
```

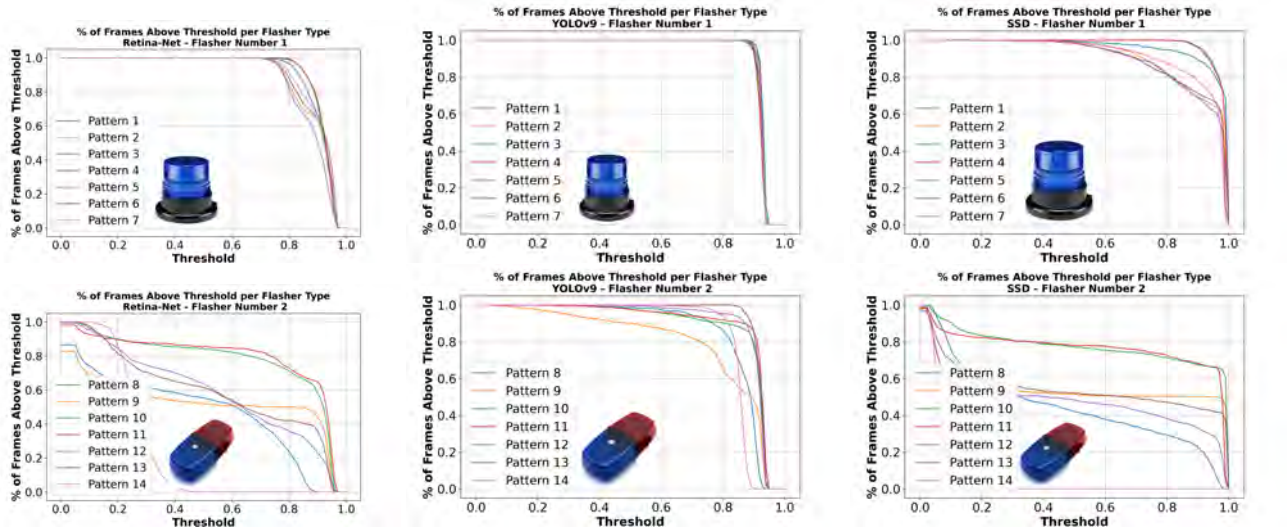

Figure 22: Comparison of percentage of frames with confidence higher than threshold for 14 emergency flasher patterns observed by RetinaNet, YOLOv3, and SSD.

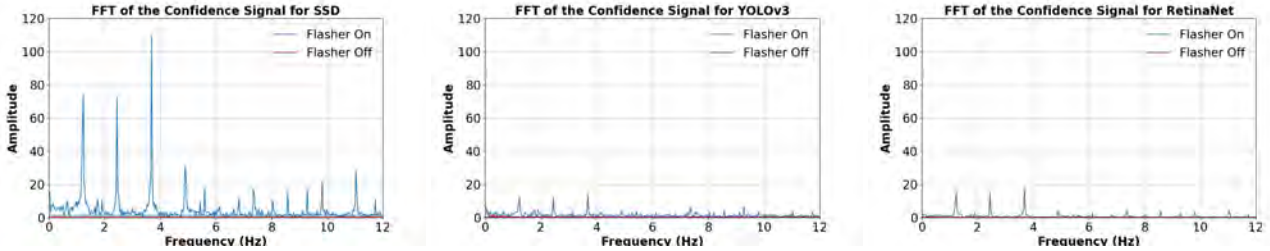

Figure 23: Extracted FFT graphs from the confidence signals of three object detectors when the emergency flasher is on. A peak around 1.3 Hz can be observed in both FFT graphs.

```
    layer_c = np.zeros(img_shape)
    layer_c[(y - r_h):(y + r_h), (x -
        r_w):(x + r_w), channel] = 255 *
         65
    layer_c[:, :, channel] =
        gaussian_filter(layer_c[:, :,
        channel], sigma=100, mode='
        constant')
    layer_w = np.zeros(img_shape)
    layer_w[(y - r_h):(y + r_h), (x -
        r_w):(x + r_w), :] = 255 * 900
    layer_w = gaussian_filter(layer_w,
        sigma=100, mode='constant')
    return np.clip(layer_c + layer_w, 0,
        255)
```

Listing 1: The script used to simulate the manual noise.

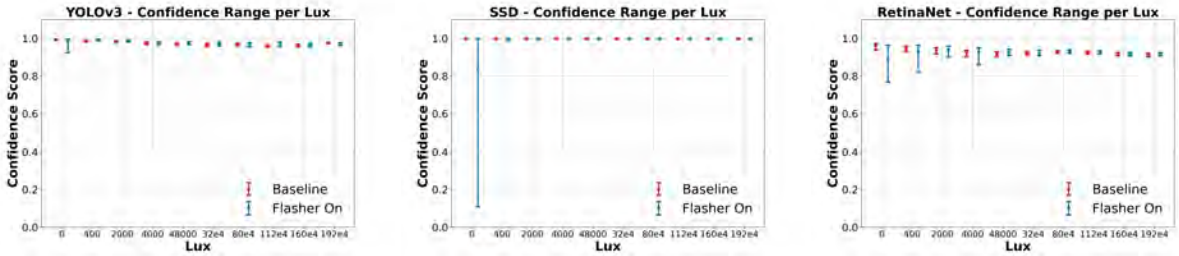

Figure 24: Confidence score signal ranges of various object detectors (YOLO, SSD, and RetinaNet) for each examined lux value. The results obtained when the emergency flasher was on are presented in blue, while the results obtained when the emergency flasher was off are in red; the points in each range indicate the average value for each signal.

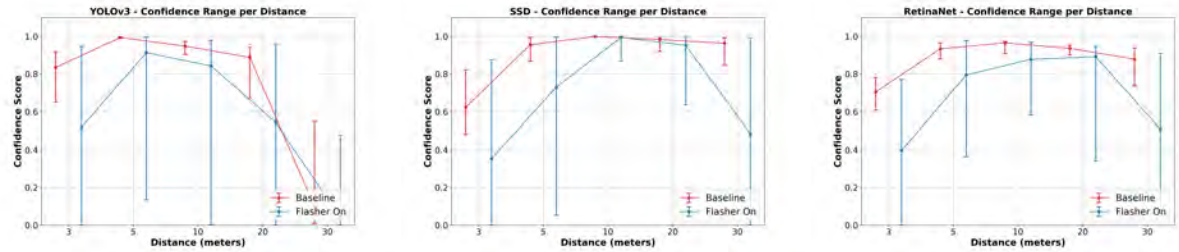

Figure 25: Confidence score signal ranges of various object detectors (YOLO, SSD, and RetinaNet) for each examined distance between an ADAS' camera and an observed vehicle. The results obtained when the emergency flasher was on are presented in blue, while the results obtained when the emergency flasher was off are in red; the points in each range indicate the average value for each signal.

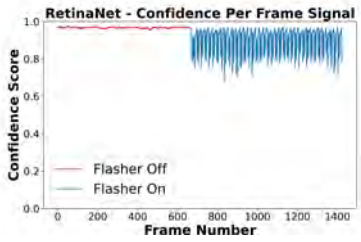

Figure 26: confidence score per frame for RetinaNet

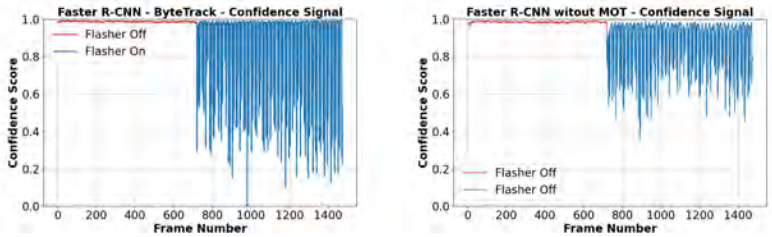

Figure 27: The confidence score signals of bytetrack object tracker model applied on the Faster R- CNN object detector and Faster R-CNN object without any multi object tracking model.