Artifact Evaluated
NDSS
Available
Functional
Reproduced

The Heat is On: Understanding and Mitigating Vulnerabilities of Thermal Image Perception in Autonomous Systems

Sri Hrushikesh Varma Bhupathiraju*, Shaoyuan Xie†, Michael Clifford‡,
Qi Alfred Chen†, Takeshi Sugawara§, Sara Rampazzi*
*University of Florida; †University of California, Irvine; ‡Toyota InfoTech Labs; §The University of Electro-Communications

***Abstract*—Thermal cameras are increasingly considered a viable solution in autonomous systems to ensure perception in low-visibility conditions. Specialized optics and advanced signal processing are integrated into thermal-based perception pipelines of self-driving cars, robots, and drones to capture relative temperature changes and allow the detection of living beings and objects where conventional visible-light cameras struggle, such as during nighttime, fog, or heavy rain. However, it remains unclear whether the security and trustworthiness of thermal-based perception systems are comparable to those of conventional cameras. Our research exposes and mitigates three novel vulnerabilities in thermal image processing, specifically within equalization, calibration, and lensing mechanisms, that are inherent to thermal cameras. These vulnerabilities can be triggered by heat sources naturally present or maliciously placed in the environment, altering the perceived relative temperature, or generating time-controlled artifacts that can undermine the correct functioning of obstacle avoidance.**

We systematically analyze vulnerabilities across three thermal cameras used in autonomous systems (FLIR Boson, InfiRay T2S, FPV XK-C130), assessing their impact on three fine-tuned thermal object detectors and two visible-thermal fusion models for autonomous driving. Our results show a mean average precision drop of 50% in pedestrian detection and 45% in fusion models, caused by flaws in the equalization process. Real-world driving tests at speeds up to 40 km/h show pedestrian misdetection rates up to 100% and the creation of false obstacles with a 91% success rate, persisting minutes after the attack ends. To address these issues, we propose and evaluate three novel threat-aware signal processing algorithms that dynamically detect and suppress attacker-induced artifacts. Our findings shed light on the reliability of thermal-based perception processes, to raise awareness of the limitations of such technology when used for obstacle avoidance.

I. Introduction

Thermal cameras are increasingly considered a potential solution for autonomous systems, such as Connected Autonomous Vehicles (CAVs) and drones, to enhance visibility in low lighting conditions, including under complete absence of illumination, such as during nighttime and severe weather [1], [2]. CAV companies, such as Zoox [3], Nuro [4], Waymo Via [5], Adastec [6], and ADAS providers [7], [8], [9] have incorporated thermal cameras into their vehicle stacks for safer navigation during night and fog conditions. Similarly, autonomous drones, such as DJI, Skydio [10], and Teledyne [11], use thermal cameras for obstacle avoidance and navigation in degraded visual environments.

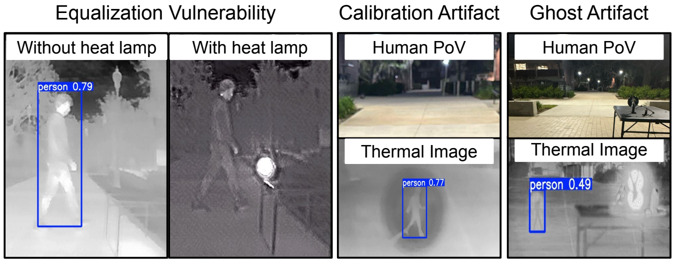


Fig. 1: Three discovered vulnerabilities in thermal image perception caused by invisible heat radiation: genuine pedestrian misdetection (left), calibration-induced obstacle creation (middle), and ghost artifacts (right).

Unlike normal RGB (red, green, and blue) cameras, which perceive visible light to capture images, thermal cameras capture infrared radiation from surrounding objects and living beings, building a heat map of the environment [12]. Due to this property, they have recently been integrated into obstacle avoidance frameworks to detect pedestrians, animals, and objects where RGB cameras underperform [2], [1].

Although the trustworthiness of these sensors is critical in such applications, their security aspects have been largely unexplored. For example, previous work has shown the possibility of conducting evasion attacks on surveillance systems based on thermal cameras [12], [13], [14], [15], [16], [17], [18]. The attackers in these studies wear materials with varying thermal properties, such as ice packs and hot patches, strategically placed on human bodies. This process generates adversarial examples specifically crafted to evade person detection in deep learning models used by security cameras. Another recent study demonstrated the susceptibility of thermal camera sensors to electromagnetic interference [19], which can induce numerical errors during sensor data transmission. All these attacks, however, target detection models in static surveillance settings and require direct manipulation of the attacker's appearance to induce misdetection. In other words, the security of thermal cameras in dynamic critical

Network and Distributed System Security (NDSS) Symposium 2026
23-27 February 2026, San Diego, CA, USA
ISBN 979-8-9919276-8-0
https://dx.doi.org/10.14722/ndss.2026.230330
www.ndss-symposium.org

scenarios, such as obstacle avoidance using in-vehicle thermal cameras, remains an open research problem.

This study aims to answer the following critical research questions: *Can thermal camera-based perception be used for critical tasks such as obstacle avoidance in autonomous systems? What are the limitations of such technology under adversarial manipulation, and how can these limitations be mitigated to ensure reliable detection?*

Our work explores vulnerabilities in thermal image acquisition and processing, fundamental to the trustworthiness of thermal camera operations. Thermal image processing typically requires specialized calibration and equalization, as illustrated in Figure 2 to handle thermal drift and high dynamic range, unlike RGB image processing, primarily optimized for visual clarity and color correction. These vulnerabilities can be exploited before the input reaches object detection and fusion models for autonomous systems' perception, causing them to falsely detect non-existent obstacles or misdetect genuine ones, as shown in Figure 1. Finally, we demonstrate how these vulnerabilities can be mitigated using our attack-aware signal processing techniques, which dynamically suppress the vulnerabilities' effects on thermal images before they are fed into machine learning models for obstacle avoidance. Our analysis aims to provide a comprehensive view of the challenges and potential solutions that autonomous systems manufacturers should consider when adopting thermal imaging technology in their perception systems. Our contributions can be summarized as follows:

Discovery of Three Vulnerabilities in Thermal Imaging.

1) **Vulnerability due to linearity in equalization.** Real-world scenarios involve dynamic variations in thermal maps, which can trigger a linear response in various equalization methods, such as plateau, CLAHE [20], and Bi-histogram equalization (BBHE) [21] algorithms used in thermal imaging equalization. Adversaries can exploit this naturally occurring condition to reduce the heat signature of genuine obstacles, causing misdetection.
2) **Flaws in thermal calibration processes.** Thermal cameras need a periodic heat intensity correction in their calibration processes, which exposes a new attack surface, enabling attackers to manipulate the resulting heat map. Attackers can exploit such vulnerability to induce delayed artifacts, which can appear in the resulting images several minutes after the attack termination, and trigger continuous detection of non-existent obstacles.
3) **Image acquisition weaknesses.** We demonstrate a vulnerability in the shutter assembly and lens design of thermal cameras, which, unlike RGB cameras [22], preserves the structure of heat signals. This allows adversaries to generate controllable flare patterns (e.g., ghost artifacts) that can appear in thermal images, triggering false object detection.

Vulnerability Characterizations. We characterize the cause of the vulnerabilities and conduct an extensive evaluation on three different thermal cameras used in autonomous systems applications (InfiRay T2S [23], FLIR Boson [24], FPV XK-C130 [25]). We focus our analysis and end-to-end evaluation on driving scenarios as a safety-critical application to quantify the impact of the threat on three state-of-the-art object detection models fine-tuned for thermal imaging processes (YOLOv5 [26], YOLOv8 [27], and Faster-RCNN [28]) and two RGB-thermal fusion models (DAMSDet [29] and Faster-RCNN [28]). Our results show the vulnerability in equalization due to linearity causes up to a 50% drop in mean average precision for pedestrian detection on the FLIR ADAS dataset [30], for all tested models. Flaws in calibration and image acquisition can induce the creation of fake pedestrian obstacles, with success rates of up to 100% in our real-world outdoor testing, which appear minutes after the attack is terminated. In realistic driving scenarios with vehicle speeds reaching up to 40 km/h, our experiments demonstrate a 100% misdetection rate and up to a 91% success rate in inducing false obstacle detections. Details and demo videos are available on our website at **https://sites.google.com/view/thermal-vuln-ad/**.

Defense Strategies. We propose three defense strategies that leverage changes in the behavior of signal processing algorithms when the underlying vulnerabilities are triggered. These strategies are designed to accurately identify false obstacles and suppress artifacts, with minimal degradation of the overall system performance. For instance, our strategy reduces the drop in the mAP score due to the equalization vulnerability from 50% to 4%, and suppresses the artifacts created by the calibration and image acquisition vulnerabilities with 100% accuracy in our real-world outdoor scenarios.

II. Background

A. Thermal Cameras vs Visible RGB Cameras

Thermal cameras operate by detecting infrared radiation emitted from objects and translating it into visual images [31]. Unlike conventional cameras, which rely on reflected visible light captured by photodiodes, thermal cameras utilize integrated bolometer arrays [32], [33] to detect infrared radiation, typically in the long wavelengths (8–14 μm) by measuring minute relative temperature variations and converting them into electrical signals to form an image. Thermal cameras generate a heat map by converting the above electrical signals into pixel values based on a predefined table, such as iron or ironbow color palette. Since conventional glass lenses are opaque to the long wavelengths of infrared thermal radiation, thermal cameras typically use lenses made from special materials (e.g., germanium or zinc selenide lenses [34]), instead of the optical glass used in visible RGB cameras. These lenses focus infrared radiation onto the bolometer, enabling the camera to form a clear thermal image by directing and concentrating heat signatures from the scene. To enable this, circular apertures are used to ensure uniform distribution of the infrared radiation across the sensor and to minimize optical aberrations [35].

Thermal Imaging Process. The temperature sensitivity of thermal cameras enables them to operate effectively in complete darkness or visually obscured environments, such

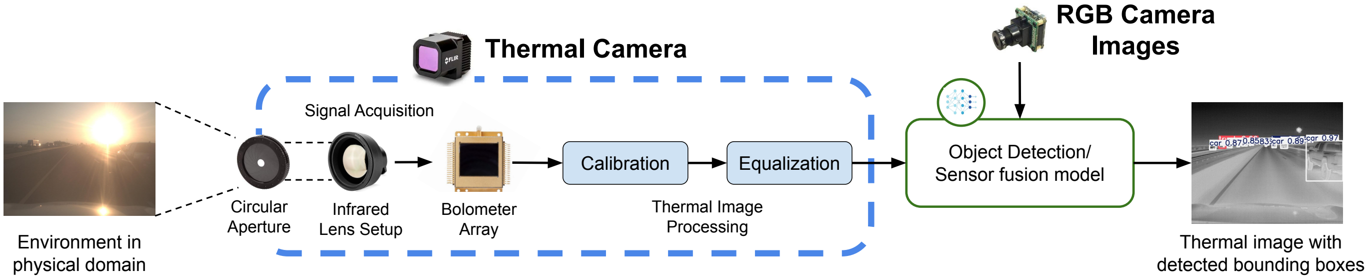


Fig. 2: Typical thermal camera-based perception in CAVs involves capturing thermal radiation, followed by calibration and equalization. These images are used as input for object detection and sensor fusion models to identify and annotate obstacles.

as fog-filled areas, where conventional optical cameras would struggle. Thus, thermal cameras have been increasingly used to complement visible imaging techniques in obstacle avoidance under harsh environments, such as during nighttime driving or under certain weather conditions [2], [36], [1].

However, operating in these dynamic environments with continuous shifts in scene temperature semantics and perceived temperature leads to challenges in contrast adjustment and susceptibility to thermal noise [33], [37]. To handle this, these sensors implement sophisticated image processing mechanisms that substantially differ from the ones in conventional RGB cameras, and consist on three main phases: acquisition, calibration, and equalization, as shown in Figure 2. While the image acquisition relies on the special lenses, the subsequent processing is based on non-linear image equalization and correction algorithms described below. Our work identifies three vulnerabilities in both the image acquisition and processing typical of thermal cameras employed in high dynamic scenarios such as autonomous driving.

Non-Uniform Correction. RGB cameras typically address sensor noise, distortion, and color balance issues using algorithms such as dark frame subtraction, flat field correction, and white balance adjustment [38], [39], [40]. In contrast, thermal cameras suffer from fixed pattern noise and pixel-to-pixel variability due to thermal drift [41], [42]. *Non-Uniform Correction (NUC)* algorithms address these issues in the calibration phase, ensuring consistent pixel intensity levels across thermal images over time [37]. *Shutter-less* NUC algorithms, in particular, work by periodically adjusting the gain and offset of each pixel to compensate for heat fluctuations, scene changes, and environmental conditions. This enhances the quality of thermal images, enabling uninterrupted image acquisition in dynamic applications, and serves as a critical component in the operation of modern commercial automotive thermal cameras [43], [44]. In contrast, shutter-based NUC methods are more commonly employed in static industrial or surveillance applications, such as equipment monitoring, because of their poor performance in dynamic settings [12]. These calibration processes differ from RGB cameras, which generally require a one-time calibration when the internal configuration or operational mode is modified, and require continuous periodic calibration, which usually repeats every 2-5 minutes, based on the thermal sensor sensitivity.

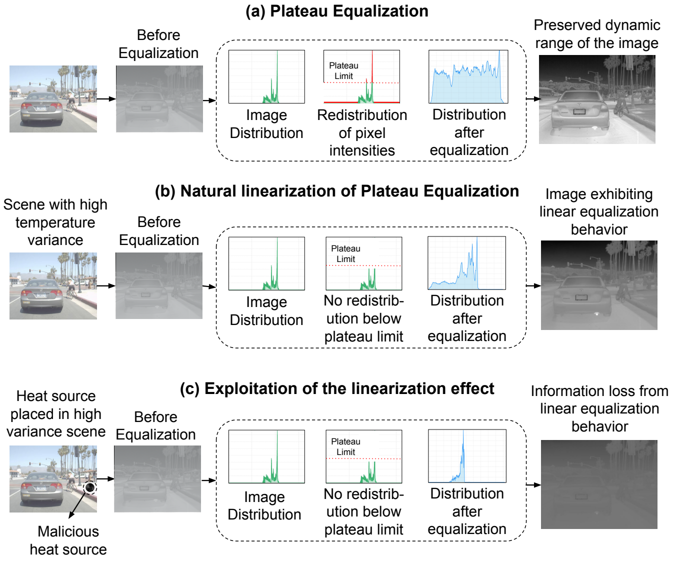

(a) Plateau Equalization
Before Equalization
Image Distribution
Plateau Limit
Redistribution of pixel intensities
Distribution after equalization
Preserved dynamic range of the image
(b) Natural linearization of Plateau Equalization
Scene with high temperature variance
Before Equalization
Image Distribution
Plateau Limit
No redistribution below plateau limit
Distribution after equalization
Image exhibiting linear equalization behavior
(c) Exploitation of the linearization effect
Heat source placed in high variance scene
Before Equalization
Malicious heat source
Image Distribution
Plateau Limit
No redistribution below plateau limit
Distribution after equalization
Information loss from linear equalization behavior

Fig. 3: (a) In natural driving scenarios, plateau equalization preserves dynamic range by redistributing pixel intensities. (b) In scenes with high temperature variance, the image distribution often falls below the plateau limit, preventing redistribution and causing linearization. (c) The linearization effect can be exploited by strategically placing heat sources, inducing information loss in equalized images.

Non-Linear Image Equalization Thermal cameras often produce images with a narrow dynamic range, where most pixel intensities cluster around certain temperature values, leading to poor contrast and loss of critical details when acquiring thermal data in diverse environments [31], [45]. Non-linear equalization algorithms, such as plateau [46], CLAHE [20], and BBHE [47], redistribute the intensities by enhancing underrepresented (low visibility) regions while limiting dominant intensity ranges, improving overall object visibility in complex thermal scenes. In contrast, linear equalization applies a uniform mapping across the image pixel distribution, often resulting in a loss of details.

Plateau equalization, particularly, is a widely used equalization method in advanced thermal cameras for object detection. It achieves non-linearity by clipping the peaks of the image histogram at a predefined or dynamic threshold, known as the plateau limit, and redistributing the clipped intensities across the remaining image [46]. This prevents over-enhancement of dominant pixel regions, improving overall contrast without amplifying noise, as shown in Figure 3 (a). Plateau equalization is adopted by major commercial automotive thermal cameras vendors (e.g., FLIR [30], BAE [48]), holding ∼42% of the current automotive thermal imaging market [49], due to its balance of performance and computational efficiency, and easy integration into embedded hardware compared to others equalization algorithms (e.g., CLAHE) [50].

B. Thermal Camera-based Perception Systems

Automotive thermal cameras typically produce 14-bit or 16-bit raw image resolutions, with pixel values ranging from 0–16,000 or up to 65,536, depending on the sensor. These images are monochromatic, representing temperature variations in a single intensity channel. Such images are typically converted to 8-bit format to facilitate visualization and processing for various applications. Object detection models, in particular, are pre-trained on RGB datasets such as COCO [51] and then fine-tuned on thermal images to leverage the rich visual representations learned in the RGB domain and enhance performance in the thermal domain [30]. In critical applications such as obstacle avoidance of CAVs, RGB-thermal fusion models combine feature representations extracted separately from RGB and thermal images, typically at early or middle stages of the neural network pipeline, to enhance robustness and accuracy in perception systems [29], [52].

Adversarial Examples on Thermal Cameras. Adversarial example attacks add noise to machine learning inputs to force a change in the output [53], [54]. Previous research has focused on generating such adversarial examples against thermal image-based object detectors used in static surveillance systems [13], [14], [15], [16], [17], [18]. For example, Zhu et al. [17] and Wei et al. [18] show how to create adversarial patches using infrared materials placed in the attacker's clothing, while Zhu et al. [16] use heat bulbs to achieve similar results. More recently, Wei et al. [15] and Hu et al. [13] use hot and cold patches to cause misdetection in a black-box setting. Separately from these works, Zhang et al [19] show how electromagnetic interference can interrupt data transmission from thermal imaging sensors, causing Denial-of-Service and data errors.

However, all these works overlook flaws in thermal cameras' imaging processes, leaving such attack surfaces unexplored. Here, we demonstrate how such vulnerabilities can undermine sophisticated dynamic perception systems, such as the one used for obstacle avoidance.

III. Threat Model and Vulnerabilities Overview

As described in Section II-A, thermal camera image processing consists of particular image acquisition, calibration, and equalization stages, encompassing capture, processing, and refinement of the thermal images. In this section, we describe the threat model and the discovered vulnerabilities in each of these stages, starting from equalization.

A. Threat Model

In this work, we investigate three vulnerabilities in the thermal imaging processing and their effect on object detectors and fusion models for obstacle avoidance.

Adversary Goal. We consider the adversary's goal of compromising the overall safety of autonomous systems (e.g., a self-driving car) by intentionally inducing edge cases in state-of-the-art thermal camera-based perception systems (e.g., FLIR ADAS object detection benchmark [30]), resulting in misdetections or the false detection of non-existent obstacles as shown in Figure 1, which in turn can trigger the activation of emergency brakes, freezing, or collision.

Previous Knowledge and Assumptions. We assume that the adversary has knowledge of the thermal image processing and acquisition of the victim thermal camera, such as the plateau configuration and calibration period. This information can be acquired from thermal camera manuals or data sheets available online [24], [55], [56]. Furthermore, an attacker can perform black-box analysis on a thermal camera similar to the one used in the victim CAV to study its sensitivity and precisely control the manipulation. For all three attack methodologies, the remote adversary does not require access to any hardware or firmware of the victim camera and autonomous system, including object detection and fusion models used in their perception (e.g., black-box setting). Finally, for false obstacle creation, we do assume the most effortless shapes that a non-expert attacker can use (e.g., human shapes) without resorting to sophisticated optimization techniques or adversarial machine learning. The adversary can also choose more complicated patterns or adversarial examples to achieve the same goal. We discuss such cases in Section IX.

Capabilities. The adversary can exploit the three vulnerabilities using a simple, commercially available heat lamp or emitters [57], [58] strategically placed in the expected victim camera field of view (FoV), such as along the roadside, near intersections, or on the back of a lead vehicle in front of the victim. Such heat sources (e.g., ceramic heat lamps) emit long-wave infrared radiations, which are detected by the thermal camera but remain invisible to the driver and nearby pedestrians, enabling stealthy manipulation of the thermal images. Moreover, due to its high diffusion, the heat signal dissipates within just a few feet (approximately 3–4), making it unnoticeable beyond that range. In detail, the adversary can use the setup to perform the following actions.

- **Exploiting Linearity in Plateau Equalization.** The adversary generates high-contrast thermal images in scenarios where the linear behavior of plateau equalization algorithms is naturally triggered. This can be achieved in common driving scenes containing objects with diverse thermal signatures, such as buildings, other cars, and trees. The adversary uses the heat source to drop the relative pixel intensity of the resulting thermal images, causing misdetection of genuine obstacles in the scene.
- **Exploiting Thermal Calibration.** The adversary induces persistent and delayed artifacts in the victim camera

thermal images, which appear minutes after the attack is concluded and are perceived as genuine obstacles by object detectors. The adversary achieves this by structuring the appearance of the heat source (e.g., resembling a human shape) and performing the attack during the periodic calibration as described in Section II-A. For example, the heat source can be placed within a lead vehicle or along the roadside during a traffic stop, leveraging prolonged idle periods (e.g., 60–120 seconds at intersections or traffic lights), as detailed in Section IV-B. Since the radiation is invisible to human eyes, the shape appears in the resulting thermal images after sufficient heat intensity accumulation (e.g., in minutes), triggering a delayed detection of false obstacles.

- **Exploiting Image Acquisition Weaknesses.** The adversary generates non-existent objects in the CAV front view by exploiting the special material lens and circular apertures of thermal cameras used in automotive [24], [55], [56], [23]. The setup allows for retaining the structural features and shapes of flares, which is not possible in RGB cameras [22], [59]. Thus, by altering the heat source's appearance, the adversary can generate ghost artifacts of arbitrary shapes, which are detected as genuine obstacles.

B. Vulnerabilities Overview

1) Plateau Equalization Vulnerability: Real-world scenarios exhibit significant temperature variance due to a combination of environmental conditions, weather, and the presence of objects with varying thermal signatures (e.g., buildings, cars, living beings) [30], [60]. Plateau equalization algorithms, as described in Section II-A, are employed in thermal cameras to enhance contrast by adjusting the gain in a nonlinear fashion based on a distribution threshold (plateau limit) to optimize the visibility of targets and contextual information. Generally, thermal cameras used for obstacle avoidance, such as the automotive FLIR Tau2 [56], require the configuration of a fixed plateau limit, which typically depends on sensor sensitivity and application. This value is expressed as the mean or median of the pixel intensity and typically ranges from 1000 to 4095 [61], [56] as lower values overly redistribute the pixel intensities, leading to loss of contrast and detail in the thermal images. However, the choice of such limit can trigger involuntary undesirable effects due to the discrepancy between the expected accuracy and the variability of driving scenario scenes. In other words, if the pixel intensity distribution in a scenario is below this value, the plateau equalization exhibits linear behavior, leading to information loss. For example, shadow regions in a highly bright image (e.g., on a sunny day) will be poorly enhanced, causing loss of relevant image details. This occurs because the algorithm does not perform the clipping and redistribution, thereby increasing the contrast of the image, as illustrated in Figure 3 (b).

To show the widespread nature of this phenomenon, we conduct an analysis on three popular thermal camera datasets used in automotive and drone perception: the FLIR ADAS [30], LTIR [62], and ASL [60] thermal datasets. Figure 4 shows that 19%, 35%, and 61% of scenes in the FLIR, ASL, and LTIR datasets, respectively, exhibit peak histogram intensity frequencies below a plateau limit of 2500, while 32%, 56%, and 70% fall below a plateau limit of 3500. These trends indicate that real-world scenarios frequently present natural conditions under which linear equalization can be triggered. Attackers can take advantage of such contexts to introduce heat sources and amplify the effect to cause extreme contrast and information loss, as shown in Figure 3 (c). This results in misdetection and a decline in perception performance, which we analyze in detail in Section V-A.

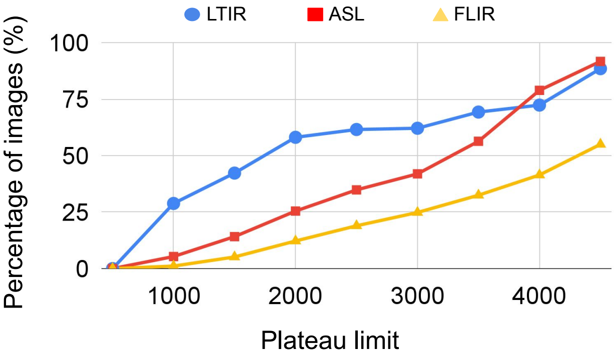


Fig. 4: Percentage of images in the three datasets below the corresponding plateau limits. Higher plateau limits trigger linearization.

2) Calibration Vulnerability: To suppress thermal drift, NUC algorithms in thermal cameras track changes in the sensor's noise characteristics in real time using frame averaging techniques and periodic calibration, as discussed in Section II-A. At the end of each calibration cycle, which usually lasts for a few minutes, they compute an offset based on the accumulated noise profile. This offset is then applied to all subsequent thermal images during the next cycle. When the next calibration cycle ends, a new offset is calculated using the updated noise profile, and the process repeats, ensuring continuous correction as the dynamic scene evolves.

Unaccounted hot objects in the thermal camera FoV, such as direct sunlight or heat sources, can elevate the estimated noise profile, resulting in an overestimation of the offset in the regions. Although this behavior is inherent to the calibration process, artifacts can appear if the hot object is removed from the FoV after the offset is updated. Since the offset remains fixed until the next calibration cycle, the overestimated correction continues to be applied to subsequent thermal images, assuming the hot object will remain in the scene, as shown in Figure 5. Consequently, if the heat source exits the FoV (e.g., the CAV moves far from the heat source), the algorithm suppresses not only noise but also part of the actual image signal, leading to a reduction in pixel intensity. This results in persistent artifacts in the affected region where the heat source was located, which remain visible until the next calibration cycle. An adversary can exploit this to induce controlled artifacts by exposing the camera to timed

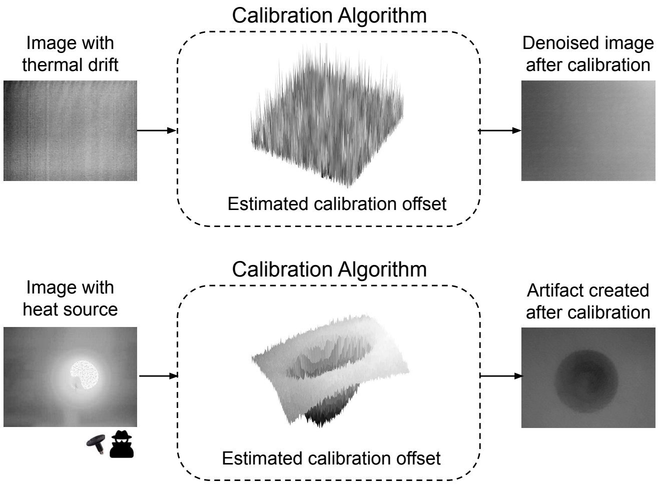

Calibration Algorithm
Image with thermal drift
Denoised image after calibration
Estimated calibration offset
Calibration Algorithm
Image with heat source
Artifact created after calibration
Estimated calibration offset

Fig. 5: Shutter-less NUC algorithms estimate pixel-wise offset, suppress thermal drift, and denoise images (top). Strategically timed heat sources manipulate the estimated pixel-wise offset, inducing artifacts in the image (bottom).

heat sources between periodic calibrations. We further study this vulnerability and characterize its effects in Section IV-B and evaluate its impact on thermal image and fusion-based detection models in Sections V-B and VI, respectively.

3) Acquisition Vulnerability: Bright light coming from a source (e.g., the sun) can reflect off the front or internal lens surfaces, producing undesirable artifacts appearing as a secondary image in RGB cameras. This phenomenon is referred to as the lens flare effect [63], [64]. The flares are typically formed as polygonal artifacts called *ghosts* on the image [22], [65], because their shape depends on the polygonal aperture used by the cameras [66], [67]. In contrast, thermal camera apertures are typically circular to allow uniform heat capture and the special material lenses minimize the diffusion of long-wave infrared radiation [34] as described in Section II-A. Such characteristics allow for the creation of arbitrarily shaped ghost artifacts. An adversary can leverage heat sources placed in the thermal camera FoV to craft and create ghosts with controlled shapes on the thermal image, and trigger detection of fake obstacles, as shown in Figure 1. We study and evaluate the impact of such structured ghost artifacts on thermal-image-based and fusion-based detection models in Sections V-C and VI, respectively.

IV. Vulnerability Characterization

To explore the vulnerabilities and the induced pixel-level changes, we first conduct proof-of-concept experiments in a controlled indoor lab scenario. We then verify our findings with three different thermal cameras for autonomous system applications (e.g., drones, CAVs) in real-world dynamic scenarios in VII. In all our evaluations in this work, the attacker setup consists of a single $20 commercial dimmable ceramic heat lamp used for terraria and animal care [68], which can reach a maximum temperature of 240°C. An adversary can also use more sophisticated setups, as discussed in Section IX.

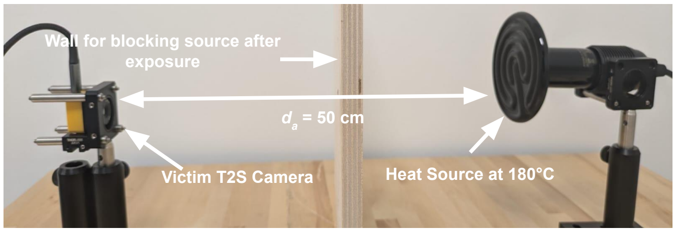


Fig. 6: Illustration of the experimental setup in our characterization analysis.

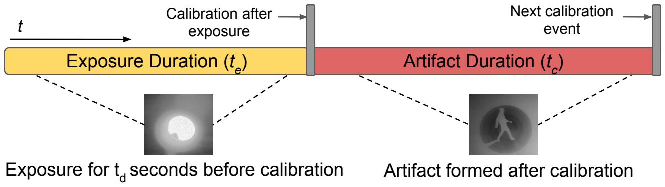

t
Calibration after exposure
Next calibration event
Exposure Duration (t_e)
Artifact Duration (t_c)
Exposure for t_d seconds before calibration
Artifact formed after calibration

Fig. 7: Calibration artifact generation after exposure to the heat source. Note that the artifact persists in the resulting images until the next calibration event.

The victim thermal camera is the InfiRay XTherm2 T2S [23] (referred to as T2S for the rest of the paper). The experimental setup is illustrated in Figure 6.

A. Analysis of the Linear Equalization Effect

As discussed in Section III-B, adversaries can amplify (under certain environmental conditions) the linear properties of plateau equalization algorithms to reduce the pixel intensity of perceived obstacles (I_{eq}). To characterize the relationship between the heat source presence and intensity changes in the output thermal images under linearization, we measure the T2S linear response in terms of relative pixel intensity drop (meaning the perceived relative temperature) in the resulting thermal images of a genuine obstacle (a pedestrian) under three different factors: temperature of the source T_a, distance from the victim camera d_a, and distance of the target pedestrian obstacle from the camera d_t.

Heat Source Temperature vs Pixel Intensity Drop. Placing the heat source at d_a at 3 m and the target pedestrian obstacle at a distance d_t of 5 m, to allow the entire figure in the camera FoV, we increase the temperature starting from 30°C to 240°C (maximum temperature of the heat source). The average relative drop in pixel intensity (average of 50 images) increases exponentially to 83%, leveling off at 91% for any further increase in temperature, up to 240°C as shown in Figure 8 (left). This result highlights the significant drop in measured relative temperature of the target obstacle after equalization, especially with increasing heat source temperature.

Heat Source Distance vs Pixel Intensity Drop. Maintaining $d_t = 5$ m and $T_a = 240$°C, we increase the distance of the heat source to the camera d_a from 2 up to 20 m at a 1 m increment for each iteration. As shown in Figure 8 (right), at $d_a = 2$ m, the relative intensity of pedestrian pixels drops by 91% and gradually decreases to 50% at $d_a = 20$ m. Based on these measurements, the linear equalization algorithm used by the victim

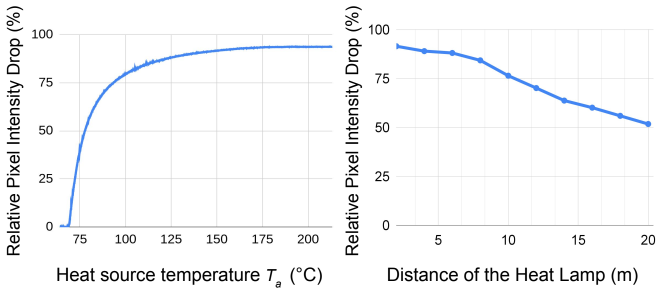


Fig. 8: Relative pixel intensity drop at increasing heat source temperature (left) and distance (right)

thermal camera is formalized as $I_{eq} = \alpha \cdot \frac{I_{xy} - I_{min}}{I_{max} - I_{min}} \cdot 255$, where I_{eq} and I_{xy} indicate the pixel intensity of the pedestrian obstacle before and after the equalization, respectively, I_{min} and I_{max} are the minimum and maximum pixel intensity in the entire image before equalization and α indicates the linear scaling factor. When $\alpha = 0.83$, the model best fits the observed results for the T2S camera. We then verify that our formalization applies to all our target thermal cameras under the same scenarios in Section VII.

Obstacle Position vs Pixel Intensity Drop. To verify if the linearization effect varies based on the target obstacle position, maintaining $d_a = 3$ m and $T_a = 240$°C, we increase the target pedestrian distance from $d_t = 5$ m to 15 m at 1 m intervals and their lateral position from 1 m to 5 m at 1 m intervals.

We observe that the drop in relative pixel intensity remains unchanged regardless of the obstacle position. This demonstrates that the measured relative temperature is uniform throughout the images. As a result, the adversary can place the heat source at any location in the FoV of the victim camera to trigger the same drop in pixel intensity.

B. Analysis of the Calibration Artifacts

To analyze the decrease in average pixel intensity (and correlated relative temperature variation) induced by calibration in response to heat source exposure, we consider the temperature of the heat source T_a and the duration of heat source exposure before (t_e) and after calibration (t_c), as shown in Figure 7. These factors directly influence the offset compensation estimated by the calibration algorithm as described in Section III-B.

Exposure Duration vs Intensity Offset. We expose the victim camera to the heat source at 50 cm distance ($d_a = 50$ cm), for 10, 20, and 30 sec durations before the calibration. Then we apply a physical block to emulate the heat source disappearance from the camera's FoV, as illustrated in Figure 6, collecting the resulting images for the subsequent two minutes until the next calibration occurs. The offset estimated by the calibration algorithm, as shown in Figure 9 (top-left), follows a weighted moving average trend, resulting in a gradual decrease in artifact intensity that is directly proportional to the increase in exposure time. The equivalent model based on the weighted moving average equation used to estimate the offset from the pixel intensity of the new image (I_{new}) is given by $I_t = \omega \cdot I_{new} + (1 - \omega) \cdot I_{t-1}$, where I_t and I_{t-1} refer to the pixel intensity after calibration in the current (at the instant t) and previous (at the instant $t - 1$) images, respectively, and ω indicates the weight of the moving average. Based on this model, the time required to converge to the maximum intensity drop t_{max} is obtained as $t_{max} \geq \ln(\frac{p}{100}) / \ln(1 - \omega)$, where p refers to the convergence percentage. Our characterization shows that the T2S camera uses a weight $\omega \approx 0.2$ and the maximum intensity drop in the artifact reaches $p = 0.1\%$ convergence percentage $t_c = 30.9$ seconds after the exposure. As in the case of the equalization analysis, we verify that this calibration model also applies to our tested thermal cameras, as shown in Figure 9 (top-right), converging to the same 0.1% at 2 minutes and 48 sec for the Boson and 41 sec for the XC-C130 camera, allowing the attacker to generate delayed attacks.

Heat Source Temperature vs Intensity Offset. We expose the heating source at 50 cm in front of the victim camera for 30 seconds before the calibration occurs. We then increase the temperature of the heat source from 20 to 240°C at 40°C intervals. We observe that a heating source of at least 60°C is required to induce a measurable drop in pixel intensity relative to the background. Beyond this point, the intensity of the artifact decreases linearly with temperature up to 100°C. After this, we observe an exponential drop in the average pixel intensity until 240°C. An adversary can select the heat source to control the resulting artifact intensity.

Calibration Timing. To induce calibration artifacts, an adversary typically needs to time the heat source exposure to align it with the calibration event. A sophisticated attacker can estimate the timing using physical cues, such as audible clicks emitted by the camera [24]. In addition, our experiments at $d_a = 50$ cm, shows that 30 second heat source exposure can still trigger a pixel intensity drop if the exposure duration ends 10 seconds before or after the calibration event, indicating that precise synchronization is not required. Particularly, we observe a relative pixel intensity drop of 18% compared to $\approx 21\%$ at the exact calibration time. Based on this observation, following our threat model, the adversary can place the heat source close to a traffic light stop and adjust the heat exposure time to improve the probability of spoofing a fake obstacle (P_{spoof}) as $P_{\text{spoof}} = (T_l / t_c) \cdot (t_e + k) / t_c$, where T_l is the traffic light duration, t_c is the calibration period (which can be derived from publicly available information such as sensor manuals), t_e is the duration of heat source exposure, and k is the duration before calibration event, when the heat source can be exposed (e.g., $k = 10$ seconds). For example, a thermal camera with $t_c = 60$ seconds and $T_l = 120$ seconds allows the attacker to succeed $\approx 83.3\%$ of the time. Based on this formalization, the adversary can adjust t_e based on the camera and the intersection setup to maximize the probability of inducing successful artifacts and spoofing a fake obstacle.

C. Analysis of Ghost Artifacts

As described in Section III-B, an adversary can create ghost artifacts in the resulting thermal images by leveraging the

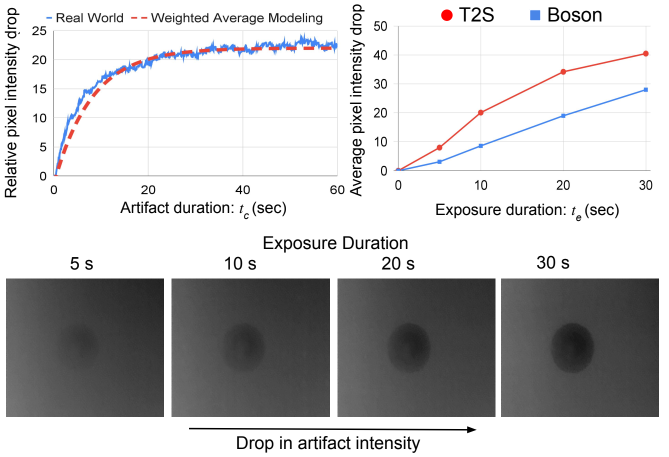


Fig. 9: The relative drop in average pixel intensity due to calibration with respect to artifact duration (t_c), after the calibration event (top-left). The pixel intensity drop due to increasing exposure (t_d) for Boson and T2S cameras (top-right) and the corresponding artifacts on the images captured from the T2S camera (bottom).

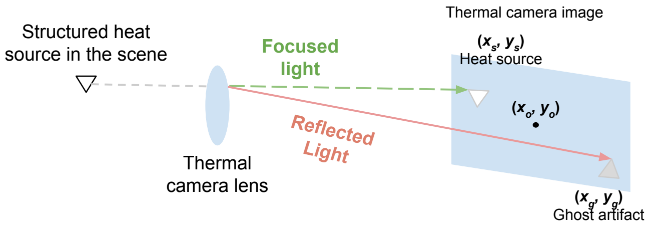


Fig. 10: Infrared light from a heat source reflecting within the lens system, resulting in an inverted ghost artifact at the geometrically opposite location in the thermal image.

aperture and special lenses of thermal cameras. As lens flair phenomena in RGB cameras, ghost images formed due to the heat source appear at a position spatially opposite to the actual heat source, relative to the optical center of the thermal camera. Precisely, the pixel coordinates of the ghost (x_g, y_g) can be estimated using the equation (x_g, y_g) = ($2x_o - x_s$, $2y_o - y_s$), where (x_o, y_o) and (x_s, y_s) are the pixel coordinates of the camera optical center and the heat source respectively, as illustrated in Figure 10. However, in contrast with RGB camera flares, the ghost artifacts preserve their structure, allowing the creation of arbitrary shapes as in Figure 11. Our experiments with the heat source placed at 1 m in front of the victim camera verify this hypothesis. Based on this formulation, an adversary can adjust the position of the heat source to create a ghost artifact in a target region of the camera FoV.

Relationship with Equalization Algorithms. As discussed in Section IV-A, the presence of heat sources in the FoV of the camera reduces the dynamic range of the image when linear equalization is triggered. This causes a drop in pixel intensity not only for genuine obstacles in the scene but also for ghost artifacts created by the lens flare effect. This introduces an inherent trade-off in thermal image acquisition, where stronger equalization algorithms (such as the plateau), while enhancing contrast, can inadvertently amplify lens flare effects, thereby increasing the visibility of ghost artifacts in the final image. On the contrary, thermal cameras with weaker equalization algorithms (such as linear equalization) increase the contrast of the image, inducing the loss of ghost artifact features, along with relevant scene information. For this reason, we use the automotive FLIR Boson camera [24] with plateau equalization to accurately characterize the lens flare effect. We manually set the plateau limit of the Boson camera below the image distribution to prevent triggering the linearization vulnerability and investigate the relationship between the heat source temperature and the artifact intensity.

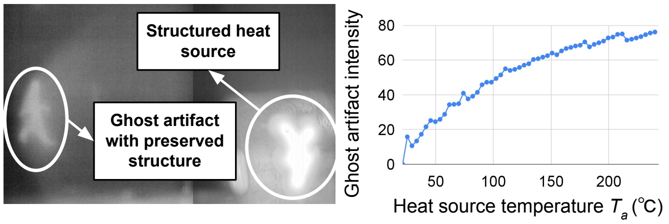


Fig. 11: Ghost artifact of human shape with preserved heat source structure in the resulting thermal image (left). The relation between ghost artifact intensity and the temperature of the heat source (right).

Heat Source Temperature vs Ghost Artifact Intensity. Based on the considerations above, we set $d_a = 0.5$ m and gradually increase T_a from 0 to 240°C. We calculate the average pixel intensity difference in the camera FoV region with and without the artifact. Figure 11 (right) illustrates the ghost artifact pixel intensity increase with increasing temperatures of the heating source, as captured by the Boson camera.

V. Evaluation on Object Detection Models

We use the three vulnerability characterizations and formalizations described in Sections IV to evaluate the extent of the threat on three object detection models. We achieve this by synthesizing the real-world results in our laboratory setting on the state-of-the-art FLIR ADAS dataset [30]. Details on the methodology are described in Appendix A. We further present the evaluation results on two RGB-thermal fusion models and real-world experiments in Section VI.

Experimental setup. Heat lamps can be shaped into arbitrary patterns by placing structured aluminum foil in front of the heating source, selectively blocking heat in certain regions to create distinct shapes. Leveraging this, we synthesize calibration-induced and ghost artifacts of different pedestrian structures and poses, as shown in Figure 13.

Our evaluation covers three state-of-the-art object detection models: (i) YOLOv5 [26], (ii) YOLOv8 [27], and (iii) Faster R-CNN [28]. While the YOLO models employ a single-stage architecture that directly predicts bounding boxes from feature maps, the Faster-RCNN model employs a two-stage architecture with a Region Proposal Network (RPN) that first generates candidate object regions, then classifies and refines those proposals. These models are fine-tuned on thermal

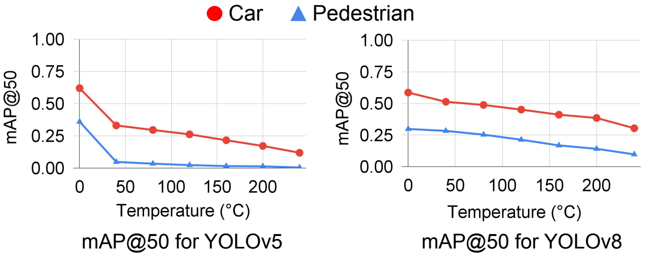


Fig. 12: The drop in mAP@50 for car and pedestrian obstacles for YOLOv5 (left) and YOLOv8 (right) at increasing heat source temperature, synthesized at 3 m distance.

images from the FLIR ADAS dataset and trained based on prior work methodology [13], [14], [15], [16], [17], [18]. We set the confidence threshold of YOLOv5 and v8 models to 0.25 and the FasterRCNN model to 0.7 (default values).

A. *Evaluation of Equalization Vulnerability*

16-bit raw images from the Boson camera are captured at heat source temperatures ranging from 0 to 240°C, in 40°C intervals. We employ the Boson camera for its image processing pipeline, similar to the Tau2 thermal camera used in the FLIR dataset. The heat source distances from the camera varied from 3 to 15 meters, in 2-meter intervals. Then we synthesize the heat source as described in Appendix A to emulate the linearization effect on the dataset.

We measure the mAP@50 score for the three models across pedestrian and car obstacle detection, which refers to the mean Average Precision computed at a 50% Intersection over Union (IoU) threshold, indicating how accurately predicted bounding boxes match the ground truth.

Results and Observations. The mAP@50 scores for pedestrian detection across the entire FLIR validation dataset drop below 0.1 (from 0.35, 0.3, and 0.73 for YOLOv5, YOLOv8, and Faster-RCNN models, respectively) when the heat source is set to 150°C, regardless of the distance, for all three models. Similarly, for car detection, the mAP@50 scores fall below 0.1 for YOLOv5, 0.15 for Faster-RCNN, and 0.25 for YOLOv8 when considering a heat source temperature of 240°C. These results suggest that scene-induced linear equalization effects can be exploited to substantially degrade the performance of object detection models, causing misdetection. Furthermore, Figure 12 shows the decline in mAP scores with increasing temperature follows a linear trend for YOLOv8, whereas it is exponential for both YOLOv5 and Faster R-CNN. We hypothesize that this difference arises from the data augmentation used during training. The baseline YOLOv8 model, pretrained on the COCO dataset, incorporates pixel intensity and contrast augmentations [27], making it more robust to thermal distortions. In contrast, the baseline YOLOv5 and Faster R-CNN models are trained using more basic augmentations, such as rotation, random cropping, and geometric distortions, resulting in greater sensitivity to intensity shifts.

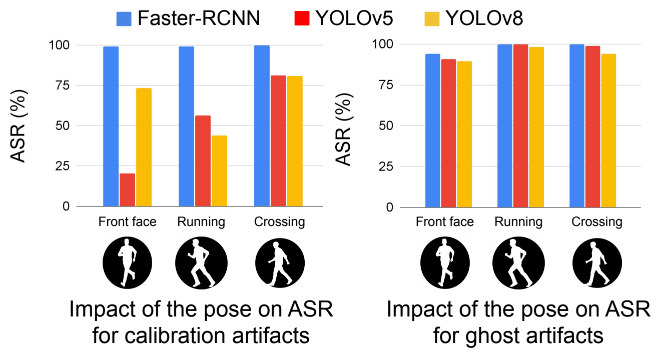


Fig. 13: Impact of artifact pose and structural characteristics on the attack success rate (ASR) across three models, for calibration-induced artifacts in the three thermal cameras (left) and ghost artifacts created with the Boson camera (right).

B. *Evaluation of Calibration Vulnerability*

Using the synthesis methodology in Appendix A, we generate human shape artifacts resulting from calibration offset manipulation on the test set of the FLIR dataset, which contains 3,749 images. The synthesized artifacts are generalizable across all three thermal cameras, as demonstrated through the modeling presented in Section IV-B. We select three distinct human shapes, each reflecting different poses of pedestrian target obstacles, as shown in Figure 13. These poses correspond to general pedestrian activities in driving scenarios, such as crossing the road or walking on the sidewalk. For each pose, we vary the pixel intensity drop within the synthesized artifacts corresponding to different durations of heat source exposure, as characterized in Section IV-B. The artifacts correspond to a 10 cm diameter heat source placed at 0.5–2.5 m distances from the camera, in 0.5 m increments, to reflect varying spatial placements during calibration exposure. We then evaluate the Attack Success Rate (ASR) over the test set, defined as the percentage of images with the synthesized artifact incorrectly detected as a pedestrian obstacle.

Results and Observations. We observe that artifacts generated due to calibration vulnerabilities reach up to 93.9% ASR on the FLIR dataset across all three evaluated models. Particularly, the artifacts achieve consistently 100% ASR on the Faster R-CNN model. This is potentially due to the model's higher sensitivity to pedestrian-like structures present in the artifacts. As illustrated in Figure 14, a given reduction in pixel intensity results in a higher ASR across all three models when the heat lamp is positioned 1.5–2 m from the thermal camera. We hypothesize that this is because, at this distance, the pedestrian structure on the 10 cm diameter heat lamp artifact closely approximates the average 118-pixel height of pedestrians in the FLIR training set. Moreover, as shown in Figure 13 (left), the pose depicting a pedestrian walking on the street consistently results in a higher ASR for a given distance and pixel intensity drop. We hypothesize that this is because the induced artifact primarily retains the structural features of the pedestrian, which are most clearly defined in the crossing pose. This pose retains the outlines of limbs, such

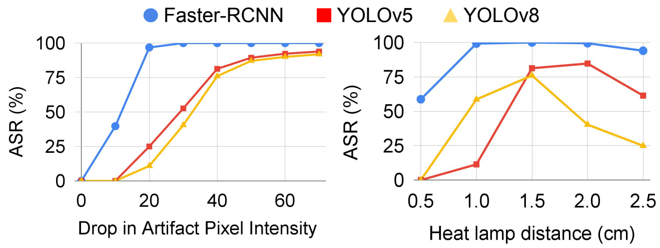


Fig. 14: Resulting attack success rate (ASR), at increasing drop in artifact pixel intensity (left) and heat source distance (right) for the pedestrian crossing pose.

as arms and legs, resulting in a more recognizable and detailed silhouette that is more effective at triggering model detections. Overall, we observe that a 40 pixel intensity drop, relative to the background, achieved with the heat lamp positioned at a distance of 1.5-2 m, can reliably trigger an ASR of ≥80% across all three models. These values correspond to exposure durations of only 30, 38, and 43 seconds for the T2S, XK-C130, and Boson cameras, respectively.

C. Evaluation of Ghost Artifacts

We use the same methodology of the calibration vulnerability to evaluate the ASR resulting from structured ghost artifacts created with the Boson camera.

Results and Observations. Figure 13 (right) shows that ghost artifacts generated by a heat source temperature of 80°C can achieve an ASR of $\geq$ 90% across all three tested models. Similar to the artifacts induced by the calibration vulnerability, these human-shaped ghost artifacts exhibit high detection when the heat lamp is positioned within the 1–2 m range. However, unlike calibration-based artifacts, where the pose significantly influences effectiveness, the ASR remains consistently high across all pedestrian poses due to their higher pixel intensity.

D. Other Equalization Algorithms

We determine the generality of the equalization vulnerability by investigating two other state-of-the-art algorithms, CLAHE [20] and BBHE [21], on the three object detection models (YOLOv5, YOLOv8, and Faster-RCNN). The analysis evaluates the detection rates of targeted pedestrian and car objects under the attack using the validation set of the FLIR dataset.

Evaluation on CLAHE. CLAHE enhances image contrast by applying histogram equalization in image regions while suppressing noise through contrast clipping. We simulate the presence of the heat source at $T_a = 240$°C, using the methodology described in Section V-A. Due to CLAHE's localized contrast enhancement, an adversary can strategically place the heat source in a specific region in the image space to target objects in that region, selectively increasing local contrast and triggering localized linear equalization behavior. We find that this targeted manipulation results in pedestrian misdetection of 66%, 69%, and 81% for YOLOv8, YOLOv5, and Faster-RCNN models, respectively, while 38%, 31%, and 46% for car obstacles. Consistent with the plateau results, we hypothesize that the disparity between objects arises because the models exhibit stronger baseline performance for car obstacle detection compared to pedestrians. Moreover, the large obstacle surface of cars typically contains multiple local regions which are impacted differently by the heat source. Nevertheless, the results indicate that the underlying vulnerability can be effectively exploited through local regions.

Evaluation on BBHE. BBHE enhances contrast by splitting the histogram at the mean intensity and applying histogram equalization separately to the lower and upper sub-histograms, thereby preserving overall brightness. We observe that malicious heat sources shift the histogram balancing point, altering contrast allocation between the sub-regions. The resulting imbalance increases linearly with heat source temperature T_a and inversely with distance d_a. Using the same FLIR validation set as CLAHE, we observe 72%, 78%, and 87% pedestrian misdetection rates on YOLOv5, YOLOv8, and Faster-RCNN models, and 51%, 54%, and 65% for car obstacles. Similar to plateau and CLAHE, we observe lower misdetection rates for cars due to better baseline performance in benign car detection. These results confirm the inherent vulnerability across state-of-the-art algorithms and the significant susceptibility of Faster-RCNN model in all our testing. For the rest of the work, we focus primarily on plateau equalization due to its widespread adoption in real-world autonomous systems [30], [48].

VI. Evaluation on Sensor Fusion Models

RGB-thermal fusion models extract feature representations from both RGB and thermal images, leveraging the complementary characteristics of each and improving the overall performance of object detection systems [29], [69], [70]. This can be achieved with two different approaches. The first, feature-level fusion, extracts feature maps from RGB and thermal images independently and then combines them [69], [70]. The second, image-level fusion combines the images at the pixel-level to learn and detect joint feature representations [36]. Our evaluation considers the state-of-the-art feature-level fusion model DAMSDet [29], which employs a transformer model to dynamically select basic salient modality feature representation for each object from both images. To evaluate image-level fusion instead, we fine-tune the Faster-RCNN detection model described in Section V, modifying its input layer to accommodate four-channel data (RGB + thermal). We chose this model due to its better performance on the FLIR dataset relative to the YOLO-based models. The DAMSDet and the fused Faster-RCNN models demonstrate mAP@50 scores $\geq$ 0.94 and $\geq$ 0.98 for pedestrian and car obstacles.

Experimental setup. To assess the impact of each vulnerability on the fusion-based models, we use the same methodology as described in Section V by synthesizing the artifacts collected from the Boson camera on thermal images of the FLIR dataset. Note that the RGB images remain unchanged for this analysis, as they are not affected by the vulnerabilities. For calibration and ghost artifacts, we consider the pose

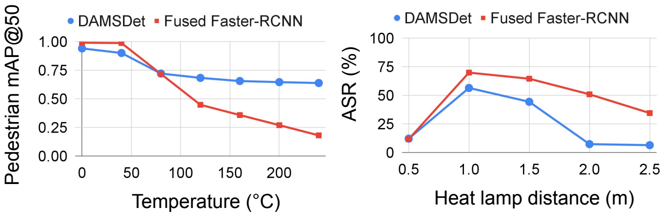


Fig. 15: Drop in pedestrian mAP@50 for fusion models (left). Attack success rate (ASR) due to the calibration-induced artifacts at increasing heat source distance (right).

corresponding to a pedestrian crossing the road, as it exhibited a higher success rate during evaluation against object detection models (worst-case scenario from the attacker's perspective).

A. Impact of Equalization Vulnerability

We observe that the mAP@50 for pedestrian obstacles gradually drops with an increase in temperature. As shown in Figure 15 (left), the mAP@50 score over the entire validation set drops by 0.80 and 0.31 for the Fused Faster-RCNN and DAMSDet models, respectively, when exposed to a 240°C heat source placed at a distance of 3 m. For car obstacles, the mAP@50 score drops by only 0.1 and 0.18 for Fused Faster-RCNN and DAMSDet, respectively. We hypothesize that the greater performance drop for pedestrian obstacles is due to their typically higher pixel intensities in thermal images compared to their RGB counterparts, which is no longer satisfied because of the intensity drop caused by linear equalization, thereby leading to misdetection.

B. Impact of Calibration Vulnerability

Similar to the results in Section V-B, the calibration artifacts demonstrate high ASR, with d_a = 1-2 m, as shown in Figure 15 (right). The maximum ASR of 56.4% and 69.7% occurs with d_a = 1-2 m and t_e = 36 s, where the average drop in pixel intensity, relative to the background, reaches $\approx$ 50 for a 240°C heat source. We observe that any higher drop in pixel intensities (from longer exposures) further decreases the ASR, consistent with the measurements in Section V-B. The lower ASR observed in comparison to detection models is expected, primarily due to the reduced thermal image intensity, as the model concurrently extracts features from RGB images. Nevertheless, the artifact can still induce false pedestrian detections, achieving more than 50% ASR with only 18 sec of exposure from a distance of 1.5 m. This demonstrates that RGB-thermal fusion models are susceptible to calibration-induced artifacts, even though they only affect thermal images.

C. Impact of Ghost Artifacts

The results indicate that the ghost artifacts can achieve up to 98.0% and 91.1% ASR in inducing false detections on DAMSDet and fused Faster-RCNN models, respectively, with T_a = 240°C and d_a = 1.5 m. Similar to the results in Section V-C, the ASR remains high with d_a = 1–2 m, achieving ASR $\geq$ 82.6% on both models with a lower T_a value of 100°C. This is because, at this distance range, the corresponding ghost artifacts fall under the size distributions that resemble those of pedestrian obstacles in the dataset. The higher success rate of ghost artifacts against sensor fusion models is due to the pixel intensity rise they introduce in the thermal images, which overwhelms the feature selection process. This is particularly true for the DAMSDet model, where thermal features are prioritized based on the assumption that the corresponding RGB images fail to capture complementary features under adverse lighting conditions.

VII. Evaluation in Outdoor Settings

We further conduct outdoor experiments to assess these vulnerabilities in real-world scenarios using three thermal cameras T2S, XK-C130, and Boson, located on top of our autonomous, rugged unmanned ground vehicle (UGV) Agile-X Hunter 2.0 [71] moving at speeds up to 10 km/h, and on a car moving at 40 km/h for equalization and calibration vulnerabilities. For this evaluation, we use the three thermal image-based object detection models described in Section V, and the ceramic heat lamp set to 240°C as our heat source, placed on the side of a road as in Figure 1. The experiments were conducted in nighttime conditions, with the average scene temperatures reaching 26-28°C and daytime with scene temperatures reaching 31–33°C.

A. Equalization Vulnerability

To investigate the extent to which an adversary can exploit equalization vulnerabilities to induce misdetections, we design and conduct experiments across two distinct real-world scenarios: a real pedestrian crossing at an intersection and a vehicle moving towards the pedestrian at different speeds.

Nighttime Scenario. We set d_a = 5 m and capture images from all three cameras as a pedestrian crosses in front of the camera at distances ranging from 5 to 12 m, in 1 m increments. We set the plateau limit of the Boson camera above the image pixel intensity distribution, emulating a highly diverse thermal scene and triggering linear equalization. In the absence of the heat lamp, all three detection models successfully detect the pedestrian at each distance with an accuracy of $\geq$96%.

Results and Observations. Across all pedestrian crossing distances, thermal images from the T2S and Boson cameras result in 100% misdetection across all three tested models. In contrast, images from the XC-C130 camera show a maximum detection rate of 46% for the Faster R-CNN model when the pedestrian crosses at 5 m, which gradually declines to 26% at 12 m. We hypothesize that the difference might arise from the proprietary non-linear equalization algorithm used in XC-C130, different from standard plateau algorithms. However, the results suggest that the algorithm's linear characteristics can be triggered, as seen with the drop in detection rates compared to the pedestrian crossing scenario without the presence of the heat lamp.

Daytime Scenario. We repeat the same experiments with pedestrian crossing in front of the camera at distance ranges from 5 to 12 m and d_a = 5 m in daytime conditions ($\approx$ 5000 lux ambient illumination). Consistent with nighttime experiments, we observe 100% misdetection across all tested models when using thermal images from the T2S and Boson cameras. For the Faster-RCNN model on the XC-C130 camera, the success rate of the attack drops from 51% when the pedestrian crossed at 12 m to 24% at a 5 m crossing distance for Faster-RCNN model, following the observation in night conditions. These results show that the equalization vulnerability does not appear to be influenced by environmental light changes.

Dynamic Scenario. In this scenario, we set d_a = 20 m and assume a static pedestrian adjacent to the heat lamp (at 20 m). We place the thermal cameras on our ground vehicle and move towards the pedestrian from 20 m away, until it crosses the heat lamp at speeds of both 5 and 10 km/h. We collect the thermal images from all three cameras and evaluate the misdetection rates across all three detection models.

Results and Observations. The YOLOv5 and YOLOv8 models exhibit 100% misdetection rates for images captured by both the T2S and Boson cameras at UGV speeds of 5 and 10 km/h. The Faster R-CNN model shows similarly high misdetection rates, with 97% and 94% for the T2S camera, and 93% and 90% for the Boson camera at 5 km/h and 10 km/h, respectively. In contrast, the XC-C130 camera demonstrates comparatively lower misdetection rates of 56% and 48% at 5 km/h and 10 km/h, respectively. These results indicate that linear properties can be triggered across non-linear algorithms, inducing misdetections in real-world conditions.

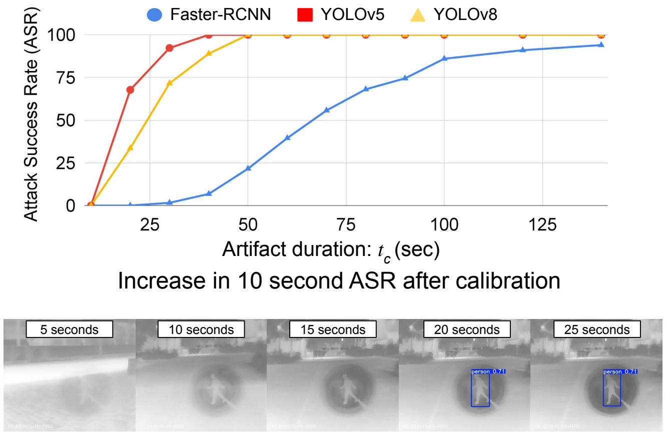
On-road Driving Scenario. We further demonstrate the attack with the thermal camera mounted on a car, repeating the dynamic experiment at 40 km/h ($d_a = 50$ m). Our results show a 100% misdetection rate across all three detection models with Boson camera images, and for the YOLOv5 and YOLOv8 models with T2S camera data. The Faster R-CNN model on T2S images exhibits a misdetection rate of 93%. For the XC-C130 camera, we observe a misdetection rate of at least 43% and up to 78% across the three models. These results indicate the effectiveness of the vulnerability at high speeds.

B. Calibration Vulnerability

Using the experimental methodology in Section IV-B, we estimate the exposure duration t_e required to trigger misdetections across the three thermal cameras. For this setup, the heat source distance d_a is fixed at 50 cm for both the T2S and XC-C130 cameras. The Boson camera's focal point is set to infinity by default, and a heat lamp positioned close to the camera (within ≈ 10 feet) is out of focus, resulting in blurred artifacts. For this reason, we set d_a = 3 m for the Boson camera and use a 15 cm diameter heat source to emulate the corresponding artifact pixel-size on the thermal image.

Nighttime Scenario. As described in Section IV-B, the artifacts emerge after a delay following calibration. To evaluate the delay, we capture images post-calibration to determine the time required for the misdetection to be induced.

Results and Observation. For the T2S camera, we observe that an exposure duration t_e = 5 sec is sufficient to induce misdetection. The artifact intensity progressively increases over time, ultimately leading to 100% misdetection approximately 25 seconds after the calibration. Figure 16 illustrates the gradual increase in ASR over 10 seconds average after the calibration. For the XC-C130 camera, an exposure duration (t_e) of 21 seconds results in a 100% ASR, with the artifact reaching full intensity drop within 50 seconds. For the Boson camera, we find that an exposure duration of 60 seconds is necessary to achieve a 90% ASR, approximately 2 minutes after calibration. The longer required t_e is attributed to the farther d_a used for the Boson camera, which results in smaller artifact projections within the image. As discussed in Section V, these smaller artifacts require a higher drop in pixel intensity to successfully trigger misdetections.



Increase in 10 second ASR after calibration

Artifact induced at different durations after calibration

Fig. 16: Trend showing the increase in attack success rate with respect to the time elapsed since the last calibration event (top). Induced artifact at increasing time intervals (bottom).

Daytime Scenario. Under daytime conditions (31–33°C), we observe that an exposure time of t_e = 12 seconds is required to induce misdetection on the T2S camera. In daytime scenarios, a 100% ASR is achieved with 25-second exposure after calibration. Similarly, for the XC-C130 camera, an exposure duration of 46 seconds is needed to reach 100% ASR, occurring 50 seconds post-calibration. For the Boson camera, 83 seconds of exposure are required to achieve 73% ASR, observed 2 minutes after calibration. These results indicate that daylight conditions require longer exposure to achieve comparable ASR, as higher temperatures of the surfaces in daytime conditions (upon which artifacts are formed) demand greater contrast in the induced artifact to trigger misdetection.

Dynamic Scenario. We collect the thermal images with the UGV moving at speeds of up to 10 km/h over a 120 m trajectory, to evaluate the robustness in dynamic conditions. In this scenario, we observe ASR $\geq$ 94.6% for T2S, $\geq$ 81% for the XC-C130, and $\geq$ 61% for the Boson camera across all three detection models. These results confirm that calibration artifacts induced due to unaccounted or malicious heat sources in the scene can trigger persistent fake obstacle detection.

On-road Driving Scenario. We collect thermal images from all three cameras mounted on a car moving at 40 kmph over a 500 m trajectory. We observe a ASR $\geq$ 91.8% for T2S, $\geq$ 83.6% for CX-C130, and $\geq$ 55.1% for the Boson camera across all detection models. These results indicate the consistency of the induced artifacts and the practicality of the attack in high-speed driving conditions.

C. Ghost Artifacts

This analysis is conducted on the Boson camera, which has amplified ghost effects due to its strong equalization and lens setup, as discussed in Section IV-C. We position a heat lamp with the pedestrian structure at distances of 1, 1.5, and 2 m in front of the camera. We set the heat lamp temperature to 240°C, and capture 120 images at each distance, containing the resulting ghost artifacts to assess the ASR across all three detection models.

Nighttime Scenario. We observe that the ghost artifact consistently triggers pedestrian detections on the Faster R-CNN model, achieving a 100% ASR across all tested heat lamp distances. For the YOLOv5 and YOLOv8 models, the attack achieves 100% ASR at a 2 m distance, but it drops to 76% and 61%, respectively, when the heat lamp is placed at 1 m. These results align with Section V-C, where larger ghost artifacts, typically produced at closer distances, exhibit lower ASR.

Daytime Scenario. Similar to the nighttime evaluation, the ghost attack achieves 100% ASR on Faster-RCNN and YOLOv5 models for all the tested distances. Similarly, the ASR for YOLOv8 drops from 100% at 2 m to 81% at 1 m heat lamp distance. As for equalization vulnerability, ghost artifacts appear not to be influenced by changes in environmental light.

Dynamic Scenario. We collect thermal images by driving the UGV toward the heat lamp from a distance of 2.5 m to approximately 1 m at a controlled speed of 2.5 km/h. The evaluation shows ASR of 76%, 28%, and 24% for Faster R-CNN, YOLOv5, and YOLOv8, respectively. As observed in the other vulnerabilities, Faster R-CNN exposes higher susceptibility compared to other models under motion.

In this case, the vehicle speed is intentionally limited to ensure safety, as creating ghost artifacts with our 15 cm diameter heat source requires proximity within 1–2 m, as seen in the nighttime scenario and simulation in Section V-C. Driving at higher speed at such close proximity poses risks of collision. Our capability analysis shows that an adversary can potentially increase the required distance for a successful attack by using a larger heat source. For example, spoofing artifacts of similar size can be induced as far as 4.5 m away with a heat source of 30 cm diameter. However, similar to previous ghost attacks [22], the heat source requires adjustment under movement to ensure correct lens reflection.

VIII. Defenses

While LiDARs and radars are commonly employed to enhance perception in autonomous systems, they can also serve as complementary modalities to validate the consistency of thermal imaging data. However, each sensor type has inherent limitations. LiDARs performance deteriorates under adverse weather conditions such as fog and rain due to scattering and absorption of laser signals, resulting in reduced range and accuracy compared to thermal cameras [72]. Radars, though more resilient to weather, lack the spatial resolution required for fine-grained scene understanding [73]. Dynamic plateau limits can reduce linearization in thermal image equalization [47], but they introduce variability in heat mapping based on scene content [74], [75]. Additionally, hot objects outside the dynamic range can still trigger linear behavior, undermining contrast enhancement.

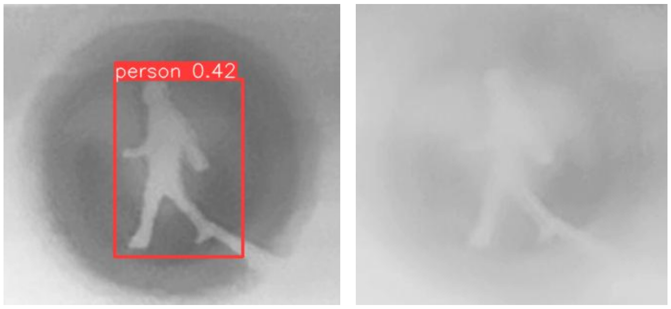
Adversarial detection techniques [76], [77] offer a potential defense against calibration-induced and ghost artifacts. However, distinguishing real-world artifacts (e.g., heat reflections) from malicious ones remains challenging. Finally, lens hoods can mitigate ghosting by limiting infrared reflections [78], but at the cost of reduced FoV. While these trade-offs may be acceptable for certain applications (e.g., surveillance systems), they quickly become inapplicable in autonomous systems with real-time reaction constraints. To address this, we propose three novel threat-aware signal processing techniques to mitigate equalization effects and suppress unwanted artifacts.

1) Attack Aware Equalization: We enhance the plateau equalization algorithm to address the linearization effect. Our method explicitly detects and excludes high-intensity regions caused by malicious heat sources, which manifest as sharp spikes in the image histogram. Excluding these spikes by filtering the histogram bins corresponding to abrupt intensity peaks allows the algorithm to adaptively mitigate the resulting linear effect and minimize the pixel-intensity drop.

We evaluate the effectiveness of this proposed algorithm under real-world conditions, using the methodology outlined in Section VII. Thermal images of a static pedestrian are captured at distances of 20, 15, 10, and 5 meters, with a heat source positioned adjacent to the pedestrian. At each distance, 100 image frames are collected. The proposed equalization method is applied to these frames, and detection performance is assessed using the Faster R-CNN model. Results show a notable improvement in detection accuracy, with average detection rates increasing to 96.3% and 98.1% at 20 and 15 meters, respectively, and reaching 100% at 10 and 5 meters. Next, we validate the ability to preserve scene integrity in the presence of natural heat sources. To do this, we select 565 test images from the FLIR dataset containing only benign heat sources (e.g., sunlight). We then apply plateau equalization using both a plateau limit set below the scene threshold and our proposed method. The resulting images yield a mean structural similarity index (SSIM) of 0.96 with a standard deviation of 0.002. These results indicate that the proposed method maintains the structural integrity of thermal images even under natural sources such as sunlight.

This technique can complement existing plateau equalization algorithms by effectively mitigating performance degradation when linear equalization is triggered in the scene.

2) Attack Aware Calibration: The calibration algorithms in automotive thermal cameras are designed with the assumption that the camera operates in a highly dynamic scene, where the estimated offset is averaged out, preventing extremely



(a) Artifact created after 10 seconds of heat source exposure

(b) Artifact suppressed using the proposed attack-aware calibration strategy

Fig. 17: Example of artifact created in real-world (left). Resulting image with the suppressed artifact by the proposed mitigation strategy (right).

high offset values. However, the exposure to natural heat sources (e.g., sunlight, vehicle tailpipes, or motorcycle exhaust cylinders) or malicious heat sources can induce the artifact, as demonstrated in Section V-B. We design a novel calibration algorithm that limits the incremental updates to the estimated offset, suppressing the artifacts formed on the thermal images.

Our real-world experiments in Section VII show that only 5 seconds of exposure to a heat lamp is sufficient to induce artifacts capable of triggering object detection. This duration corresponds to approximately 125 frames on the T2S camera and 600 frames on the Boson camera under continuous exposure. To mitigate this, we implement a pixel-wise offset counter that tracks the number of consecutive images during which the averaged offset at each pixel increases. The counter resets if the averaged offset decreases, indicating the removal of the heat source. Using this mechanism, we design a calibration algorithm that temporarily halts offset updates for pixels exhibiting continuous offset increases over a predefined threshold (e.g., <125 frames for T2S and <600 for Boson), preventing excessive drift and artifact formation. Offset updates resume once a decrease in the average offset is measured. This approach suppresses the escalation of offset value due to unaccounted heat sources while preserving the algorithm's ability to correct for genuine thermal drift.

We evaluate the proposed calibration algorithm using our real-world image sequences captured by the Boson camera. Following the experimental setup detailed in Section IV-B, we position a heat lamp at 240°C at a distance of 50 cm in front of the camera and expose it for 30 seconds prior to calibration. This prolonged exposure creates an average relative pixel intensity drop of 24. The resulting artifact is illustrated in Figure 17 in the Appendix. Using the image frames collected during the exposure period, we apply our attack-aware calibration algorithm, which restricts the offset accumulation once it has continuously increased for 600 images. As a result, the induced artifact is mitigated as shown in Figure 17. Furthermore, we evaluate the capability of our calibration algorithm not to suppress genuine obstacles by positioning a pedestrian 5 m in front of the camera for a duration of 5 minutes with the heat source. The algorithm pauses the offset update for only 3.6 seconds throughout the sequence, due to the subtle body movements of the pedestrian, introducing slight fluctuations in the averaged offset rather than the consistent increase observed with malicious heat signal injection. Faster-RCNN model maintains a high detection confidence of 92%, demonstrating that minor variations in pixel intensity do not significantly impact detection performance.

3) Mitigating Ghosts: Leveraging the geometric characteristics of ghost artifacts, we propose a mitigation strategy to suppress their formation. As demonstrated in Section V-C, inducing ghost artifacts capable of triggering false obstacle detection with high success rates ($\geq$ 90%) requires a heat source with temperatures reaching 80°C. This corresponds to an average pixel intensity of approximately 9500 in the raw thermal images. Building on this observation, our methodology identifies such high-intensity regions and suppresses the intensity of the corresponding geometrically opposite pixels, where ghost artifacts appear, based on the relationship characterized in Figure 10. Using the methodology outlined in Section VII-C, the proposed mitigation achieves a 100% success rate in suppressing ghost artifacts. The typical raw pixel intensity of human subjects in the scene ranges between 7000–7500 (corresponding to 30–36°C), which is insufficient to induce ghost artifacts capable of triggering false detections. As a result, the mitigation algorithm does not alter such regions, preserving the integrity of genuine detections.

IX. Discussion

CLAHE as Defense. While CLAHE can be better in preserving local contrast, compared to plateau equalization, implementing it in real-world automotive applications has been proven challenging due to its higher computational complexity, difficult hardware implementation, and limited real-time performance in dynamic scenarios, which is still an ongoing research problem [50], [79]. Our analysis in Section V-D shows how CLAHE is still affected by linearization, however, future research may focus on augmenting its resilience.

Limitations. Our experimental setup is constrained to controlled outdoor scenarios, focusing on close-proximity heat sources ranging from 0.5 to 5 meters. This is due to the relatively small heat lamp (10 cm in diameter) used as the heat source. In practical settings, larger heat-emitting surfaces (such as vehicle exhaust manifolds or engine cylinders) reaching comparable temperatures could induce artifacts of similar size from greater distances. Further, the perturbations demonstrated in this work manifest as circular artifacts surrounding the pedestrian silhouette, primarily due to the use of the circular shape of the heat lamp. While effective in inducing pedestrian detections, this design is relatively simple and easily identifiable. A more sophisticated adversary could enhance stealth by tailoring the heat source's structure or employing adversarial optimization techniques [15], [80], [81] to generate imperceptible or natural-looking artifacts, thereby making the injection harder to detect while maintaining high effectiveness, as shown in Figure 18 in the Appendix.

Heat Source Temperature. All experiments in this work were conducted using a cheap commercial reptile heater [68] which can reach 240°C. As demonstrated in the outdoor experiments (Section VII), 10 seconds or exposure is sufficient to induce artifacts detected as obstacles. To compensate for lower heat source temperatures, an adversary can increase the exposure duration. For example, our capability analysis (Section IV-B) shows that a 120°C source can produce similar artifact intensities with ≈ 42 seconds of exposure.

Safety Considerations. All experiments were conducted in controlled environments. While the surface temperature of the heat lamp reaches 240°C, its diffusive emission characteristics cause the temperature to drop to approximately 30°C at a distance of 50 cm, making it safe to operate at that range.

A. Related Work

Sensors, such as cameras, LiDAR, and radar, serve as the foundation of perception in autonomous systems by collecting environmental data [82], [83], [28], [84], [12]. However, extensive research has shown that these sensors are susceptible to spoofing and injection attacks. For instance, several works demonstrated that LiDAR sensors are vulnerable to laser injection [85], [86], [87], [88] and electromagnetic interference (EMI) [89], [90]. Similarly, cameras are vulnerable to laser [91], [92], [93], [65] and EMI injection [94] attacks. Beyond perception systems, prior research has also identified vulnerabilities in inertial measurement units (IMUs) [95], [96] and GPS [97], [98] used in autonomous platforms. In this work, we present a comprehensive investigation into vulnerabilities of thermal camera image processing, motivated by their growing adoption in autonomous driving systems. Cao et al. [88] investigate the automatic transformation and filtering mechanisms in LiDAR systems, demonstrating how adversaries can manipulate input data to trigger point removal before it reaches the perception model. Hunt et al. [99] expose vulnerabilities in radar signal processing pipelines, highlighting their susceptibility to adversarial interference. Similarly, Ji et al. [100] identify weaknesses in image stabilization mechanisms embedded in camera sensors; their findings show that acoustic signals can trigger unnecessary motion compensation, resulting in image blur and degraded perception performance.

In addition to vulnerabilities at the sensor and processing levels, machine learning models themselves pose significant security concerns. Deep Neural Networks (DNNs), despite their success in enabling accurate perception for autonomous vehicles, have been shown to be highly susceptible to adversarial perturbations [53]. This vulnerability has been extensively demonstrated across various perception tasks, including object detection [101], [102], [103], [104], [105], object tracking [106], [107], [108], semantic segmentation [109], automated lane centering [110], and traffic sign recognition [111], [103], [102]. Unlike these works that directly manipulate DNN inputs to trigger incorrect outputs, our study focuses on how upstream vulnerabilities of thermal cameras can indirectly lead to perception failures.

X. Conclusion

Our work identifies a new class of vulnerabilities against thermal cameras, causing perception failures such as misdetection or creation of fake obstacles. Our evaluation demonstrates the effectiveness of the induced pixel intensity drop or the creation of artifacts, which can undermine three thermal image-based object detection models and two RGB-thermal fusion-based models. We further provide feasibility evaluations of attacks in real-world driving conditions. While the safety of integrating thermal cameras into autonomous systems remains uncertain, our work aims to serve as an initial step towards understanding threats in thermal imaging processing designs, while proposing attack-aware signal processing techniques to effectively mitigate them.

Acknowledgements

We thank the anonymous shepherd, reviewers, and AE evaluators for their valuable comments. This research was supported in part by the NSF CNS-2145493, CNS-2413877, USDOT under Grant 69A3552348327 for the CARMEN+ University Transportation Center, JSPS KAKENHI Grant Number 24K14943, JST CREST JPMJCR23M4, SoCal Hub grant, and unrestricted research funds from Toyota InfoTech Labs. We also would like to thank Srajana Chandrashekhar and Vladimir A. Alekseev for their assistance with the outdoor experiments.

References

[1] Zoox, Inc., "Zoox Perception," 2023. [Online]. Available: https://zoox.com/journal/perception/

[2] PlusAI, Inc., "Plus Starts Development with Teledyne FLIR to Test Thermal Cameras for Autonomous Trucks," https://plus.ai/news-and-insights/plus-starts-development-with-teledyne-flir-to-test-thermal-cameras-for-autonomous-trucks, 2021.

[3] Teledyne FLIR, "FLIR to Provide Thermal Imaging Cameras for Zoox Robotaxi," 2020. [Online]. Available: https://www.flir.com.mx/news-center/camera-cores--components/flir-to-provide-thermal-imaging-cameras-for-zoox-robotaxi/

[4] Nuro, "Safety @ Nuro: Our Vehicles," 2022. [Online]. Available: https://www.nuro.ai/blog/safety-nuro-our-vehicles

[5] Waymo, "Designed to deliver: Bringing the benefits of our 5th generation hardware to trucking," 2021. [Online]. Available: https://waymo.com/blog/2021/12/designed-to-deliver

[6] ADASTEC, "ADASTEC selects Teledyne FLIR thermal sensors for flowride.ai Level-4 autonomous bus platform." 2025. [Online]. Available: https://www.adastec.com/news-coverage/mass-transit-article-adastec-selects-teledyne-flir

[7] Seek Thermal, "See The Unseen: The Future of Automotive Safety Starts with Thermal Imaging," 2025. [Online]. Available: https://www.thermal.com/automotive.html

[8] Veoneer, "Veoneer Starts Production of World's Most Advanced Automotive Thermal Sensing System on the All-New 2021 Cadillac Escalade," https://www.veoneer.com/en/press/veoneer-starts-production-worlds-most-advanced-automotive-thermal-sensing-system-all-new-2021, 2021.

[9] Owl Autonomous Imaging, "Owl AI Applications: Automotive ADAS & Autonomy," 2024. [Online]. Available: https://www.owlai.us/applications-auto/

[10] Skydio, "Skydio X10," 2025. [Online]. Available: https://www.skydio.com/x10

[11] Teledyne FLIR, "Military & Defense UAS Solutions," 2025. [Online]. Available: https://defense.flir.com/unmanned/unmanned-aerial-systems/

[12] R. Gade and T. B. Moeslund, "Thermal cameras and applications: a survey," *Machine vision and applications*, 2014.

[13] C. Hu, W. Shi, T. Jiang, W. Yao, L. Tian, X. Chen, J. Zhou, and W. Li, "Adversarial infrared blocks: A multi-view black-box attack to thermal infrared detectors in physical world," *Neural Networks*, vol. 175, 2024.

[14] X. Wang and W. Li, "Physical adversarial attacks for infrared object detection," in *4th International Conference on Consumer Electronics and Computer Engineering (ICCECE)*. IEEE, 2024.

[15] H. Wei, Z. Wang, X. Jia, Y. Zheng, H. Tang, S. Satoh, and Z. Wang, "Hotcold block: Fooling thermal infrared detectors with a novel wearable design," in *Proceedings of the AAAI conference on artificial intelligence*, 2023.

[16] X. Zhu, X. Li, J. Li, Z. Wang, and X. Hu, "Fooling thermal infrared pedestrian detectors in real world using small bulbs," in *Proceedings of the AAAI conference on artificial intelligence*, 2021.

[17] X. Zhu, Z. Hu, S. Huang, J. Li, and X. Hu, "Infrared invisible clothing: Hiding from infrared detectors at multiple angles in real world," in *IEEE/CVF CVPR*, 2022.

[18] X. Wei, J. Yu, and Y. Huang, "Physically adversarial infrared patches with learnable shapes and locations," in *IEEE/CVF CVPR*, 2023.

[19] H. Zhang, Q. Jiang, Y. Cheng, X. Ji, and W. Xu, "Intentional electromagnetic interference attack against infrared thermal imaging sensor," in *IEEE 6th Conference on Energy Internet and Energy System Integration (EI2)*, 2022.

[20] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *Journal of VLSI signal processing systems for signal, image and video technology*, 2004.

[21] M. Moniruzzaman, M. Shafuzzaman, and M. F. Hossain, "Brightness preserving bi-histogram equalization using edge pixels information," in *2013 international conference on electrical information and communication technology (EICT)*. IEEE, 2014, pp. 1–5.

[22] Y. Man, M. Li, and R. Gerdes, "GhostImage: Remote perception attacks against camera-based image classification systems," in *23rd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2020)*, 2020.

[23] Xinfrared, "InfiRay T2S Plus Thermal Camera," 2022. [Online]. Available: https://xinfraredx.com/products/infiray-t2s-plus-thermal-camera

[24] FLIR Systems Inc., "Datasheet: FLIR Boson Thermal Imaging Core," 2019.

[25] Xingkai Technology, "XK-C130 Thermal Imager Camera Module." [Online]. Available: https://xingkaitech.com/gimbal-camera/xk-c130/

[26] G. Jocher, A. Stoken, J. Borovec, L. Changyu, A. Hogan, L. Diaconu, J. Poznanski, L. Yu, P. Rai, R. Ferriday *et al.*, "ultralytics/yolov5: v3. 0," 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3983579

[27] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics yolov8," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[28] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, 2016.

[29] J. Guo, C. Gao, F. Liu, D. Meng, and X. Gao, "DAMSDet: Dynamic adaptive multispectral detection transformer with competitive query selection and adaptive feature fusion," in *ECCV*, 2024.

[30] Teledyne FLIR, "Teledyne FLIR Free ADAS Thermal Datasets v2," 2022. [Online]. Available: https://www.flir.com/oem/adas/adas-dataset-form/

[31] T. Sosnowski, G. Bieszczad, and H. Madura, "Image processing in thermal cameras," in *Advanced Technologies in Practical Applications for National Security*. Springer International Publishing, 2018, pp. 35–57.

[32] F. Niklaus, C. Vieider, and H. Jakobsen, "MEMS-based uncooled infrared bolometer arrays: a review," *MEMS/MOEMS technologies and applications III*, 2008.

[33] T. Williams, *Thermal imaging cameras: characteristics and performance*. CRC Press, 2009.

[34] D. Gibson, S. Bayya, V. Nguyen, J. Myers, E. Fleet, J. Sanghera, J. Vizgaitis, J. Deegan, and G. Beadie, "Diffusion-based gradient index optics for infrared imaging," *Optical Engineering*, 2020.

[35] B. Guenther, "Diffraction — Fresnel diffraction," in *Encyclopedia of Modern Optics*, R. D. Guenther, Ed. Oxford: Elsevier, 2005.

[36] M. Brenner, N. H. Reyes, T. Susnjak, and A. L. C. Barczak, "RGB-D and thermal sensor fusion: A systematic literature review," *IEEE Access*, vol. 11, pp. 82 410–82 442, 2023.

[37] O. Riou, S. Berrebi, and P. Bremond, "Nonuniformity correction and thermal drift compensation of thermal infrared camera," in *Thermosense XXVI*. SPIE, 2004.

[38] M. Afifi, M. A. Brubaker, and M. S. Brown, "Auto white-balance correction for mixed-illuminant scenes," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022.

[39] Y.-C. Liu, W.-H. Chan, and Y.-Q. Chen, "Automatic white balance for digital still camera," *IEEE Transactions on Consumer Electronics*, vol. 41, no. 3, pp. 460–466, 1995.

[40] S. Brutzer, B. Höferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *CVPR*, 2011.

[41] S. Brazane, O. Riou, F. Delaleux, L. Ibos, and J. F. Durastanti, "Management of thermal drift of bolometric infrared cameras: limits and recommendations," *Quantitative InfraRed Thermography Journal*, 2025.

[42] S. Brazane, O. Riou, F. Delaleux, L. Ibos, and J.-F. Durastanti, "Assessment of thermal drift of the FLIR A325sc camera: limits and recommendations," in *16th Quantitative InfraRed Thermography conference*, 2022.

[43] C. Liu, X. Sui, G. Gu, and Q. Chen, "Shutterless non-uniformity correction for the long-term stability of an uncooled long-wave infrared camera," *Measurement Science and Technology*, 2018.

[44] A. Averbuch, G. Liron, and B. Z. Bobrovsky, "Scene based non-uniformity correction in thermal images using Kalman filter," *Image and Vision Computing*, 2007.

[45] R. Dulski, P. Powalisz, M. Kastek, and P. Trzaskawka, "Enhancing image quality produced by IR cameras," in *Electro-Optical and Infrared Systems: Technology and Applications VII*. SPIE, 2010.

[46] V. E. Vickers, "Plateau equalization algorithm for real-time display of high-quality infrared imagery," *Optical engineering*, 1996.

[47] C. H. Ooi, N. S. P. Kong, and H. Ibrahim, "Bi-histogram equalization with a plateau limit for digital image enhancement," *IEEE Trans. Consumer Electron.*, vol. 55, no. 4, pp. 2072–2080, 2009.

[48] BAE Systems, "Thermal Imaging Sensors," 2025. [Online]. Available: https://www.baesystems.com/en-us/product/thermal-imaging-sensors

[49] PW Consulting Automotive and Research Center, "Automotive Infrared Cores Market," 2025. [Online]. Available: https://pmarketresearch.com/auto/automotives-infrared-cores-market/

[50] P. Härtinger and C. Steger, "Adaptive histogram equalization in constant time," *Journal of Real-Time Image Processing*, vol. 21, no. 3, p. 93, 2024.

[51] T. Lin, M. Maire, S. J. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: common objects in context," in *ECCV 2014*, ser. LNCS, vol. 8693. Springer, 2014, pp. 740–755.

[52] X. Zhang, P. Ye, H. Leung, K. Gong, and G. Xiao, "Object fusion tracking based on visible and infrared images: A comprehensive review," *Information Fusion*, 2020.

[53] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," *arXiv:1312.6199*, 2013.

[54] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *3rd International Conference on Learning Representations, ICLR*, 2015.

[55] Teledyne FLIR, "FLIR Lepton Series Datasheet," 2015. [Online]. Available: https://flir.netx.net/file/asset/56466/original/attachment

[56] FLIR Systems Inc., "FLIR Tau2 Product Specification," 2015. [Online]. Available: https://flir.custhelp.com/ci/fattach/get/107854/0/filename/102-PS242-40-Tau2ProductSpec_Rev141.pdf

[57] Philips, "Philips Heat Lamp," 2025. [Online]. Available: https://www.assets.signify.com/is/content/Signify/US.en_US.046677416744

[58] Thorlabs, "Metal Ceramic Heater Element, 24 W," 2023. [Online]. Available: https://www.thorlabs.com/drawings/5d27add3d04a0967-E2BFB763-CB7B-DEA3-26B5BF365534EC62/HT24S-SpecSheet.pdf

[59] Y. Kotp and M. Torki, "Toward flare-free images: A survey," *arXiv preprint arXiv:2310.14354*, 2023.

[60] Autonomous System Lab, "Thermal Infrared Dataset," 2014. [Online]. Available: https://projects.asl.ethz.ch/datasets/doku.php?id=ir:iricra2014

[61] FLIR, "FLIR Camera Controller GUI," 2014. [Online]. Available: https://grupoacre.com.pt/wp-content/uploads/downloads/flir_gui_user.pdf

[62] A. Berg, J. Ahlberg, and M. Felsberg, "A thermal object tracking benchmark," in *Advanced Video and Signal Based Surveillance (AVSS), 12th IEEE International Conference on*, 2015.

[63] M. Hullin, E. Eisemann, H.-P. Seidel, and S. Lee, "Physically-based real-time lens flare rendering," in *ACM SIGGRAPH papers*, 2011.

[64] D. Reddy and A. Veeraraghavan, “Lens flare and lens glare,” in *Computer Vision: A Reference Guide*. Springer International Publishing, 2021, pp. 741–744.

[65] C. Zhou, Q. Yan, Y. Shi, and L. Sun, “DoubleStar: Long-range attack towards depth estimation based obstacle avoidance in autonomous systems,” in *USENIX Security*, 2022.

[66] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, “Image and depth from a conventional camera with a coded aperture,” *ACM Trans. Graph.*, vol. 26, no. 3, p. 70, 2007.

[67] SNAPSHOT, “Lens FAQ: What Are Aperture Blades? How Do They Influence My Photos?” 2022. [Online]. Available: https://snapshot.canon-asia.com/article/eng/lens-faq-what-are-aperture-blades-how-do-they-influence-my-photos

[68] ExoTerra, “Ceramic Heater Operating Instructions,” 2012. [Online]. Available: https://exo-terra.com/wp-content/uploads/PT2044_45_46_47_48_Ceramic_Heater_EU.pdf

[69] J. Wang, K. Song, Y. Bao, L. Huang, and Y. Yan, “CGFNet: Cross-guided fusion network for RGB-T salient object detection,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[70] W. Zhou, Q. Guo, J. Lei, L. Yu, and J. Hwang, “ECFFNet: Effective and consistent feature fusion network for RGB-T salient object detection,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1224–1235, 2022.

[71] AgileX Robotics Team, “Hunter2.0 user manual,” 2023. [Online]. Available: https://cdn.shopify.com/s/files/1/0551/0630/6141/files/HUNTER2.0_USER_MANUAL2023.12_50805.pdf

[72] J. M. R. Velázquez, L. Khoudour, G. Saint Pierre, P. Duthon, S. Liandrat, F. Bernardin, S. Fiss, I. Ivanov, and R. Peleg, “Analysis of thermal imaging performance under extreme foggy conditions: Applications to autonomous driving,” *Journal of imaging*, vol. 8, no. 11, p. 306, 2022.

[73] A. S. Mohammed, A. Amamou, F. K. Ayevide, S. Kelouwani, K. Agbossou, and N. Zioui, “The perception system of intelligent ground vehicles in all weather conditions: A systematic literature review,” *Sensors*, vol. 20, no. 22, p. 6532, 2020.

[74] R. Khalid, S. Rehman, F. Riaz, and A. Hassan, “Enhanced dynamic quadrant histogram equalization plateau limit for image contrast enhancement,” in *Fifth International Conference on Digital Information and Communication Technology and its Applications (DICTAP)*. IEEE, 2015.

[75] C. H. Ooi and N. A. M. Isa, “Adaptive contrast enhancement methods with brightness preserving,” *IEEE Trans. Consumer Electron.*, vol. 56, no. 4, pp. 2543–2551, 2010.

[76] Z. Deng, X. Yang, S. Xu, H. Su, and J. Zhu, “Libre: A practical bayesian approach to adversarial detection,” in *CVPR*, 2021.

[77] Y. Xu, H. Nagahara, A. Shimada, and R.-i. Taniguchi, “Transcut: Transparent object segmentation from a light-field image,” in *ICCV*, 2015.

[78] E. Talvala, A. Adams, M. Horowitz, and M. Levoy, “Veiling glare in high dynamic range imaging,” *ACM Trans. Graph.*, vol. 26, no. 3, p. 37, 2007.

[79] G. F. C. Campos, S. M. Mastelini, G. J. Aguiar, R. G. Mantovani, L. F. d. Melo, and S. Barbon Jr, “Machine learning hyperparameter selection for contrast limited adaptive histogram equalization,” *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 1, p. 59, 2019.

[80] W. Xu, S. Szyller, C. Cornelius, L. M. Rojas, M. Arvinte, A. Velasquez, J. Martin, and N. Himayat, “Imperceptible adversarial examples in the physical world,” *arXiv preprint arXiv:2411.16622*, 2024.

[81] S.-M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard, “Universal adversarial perturbations,” in *CVPR*, 2017.

[82] Z. Liu, H. Tang, A. Amini, X. Yang, H. Mao, D. L. Rus, and S. Han, “BEVFusion: Multi-task multi-sensor fusion with unified bird’s-eye view representation,” in *IEEE International Conference on Robotics and Automation, ICRA*, 2023.

[83] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *CVPR*, 2016.

[84] F. Roos, J. Bechter, C. Knill, B. Schweizer, and C. Waldschmidt, “Radar sensors for autonomous driving: Modulation schemes and interference mitigation,” *IEEE Microwave Magazine*, 2019.

[85] 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 *ACM CCS*, 2019.

[86] T. Sato, R. Suzuki, Y. Hayakawa, K. Ikeda, O. Sako, R. Nagata, R. Yoshida, Q. A. Chen, and K. Yoshioka, “On the realism of LiDAR spoofing attacks against autonomous driving vehicle at high speed and long distance,” in *NDSS*, 2025.

[87] T. Sato, Y. Hayakawa, R. Suzuki, Y. Shiiki, K. Yoshioka, and Q. A. Chen, “LiDAR spoofing meets the new-gen: Capability improvements, broken assumptions, and new attack strategies,” in *NDSS*, 2024.

[88] 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 *USENIX Security*, 2023.

[89] S. H. V. Bhupathiraju, J. Sheldon, L. A. Bauer, V. Bindschaedler, T. Sugawara, and S. Rampazzi, “Emi-lidar: Uncovering vulnerabilities of lidar sensors in autonomous driving setting using electromagnetic interference,” in *Proceedings of the 16th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2023.

[90] Z. Jin, Q. Jiang, X. Lu, C. Yan, X. Ji, and W. Xu, “PhantomLiDAR: Cross-modality signal injection attacks against LiDAR,” in *NDSS*, 2025.

[91] D. Nassi, R. Ben-Netanel, Y. Elovici, and B. Nassi, “MobilBye: Attacking ADAS with camera spoofing,” *CoRR*, vol. abs/1906.09765, 2019.

[92] C. Yan, W. Xu, and J. Liu, “Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicle,” *Def Con*, 2016.

[93] 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,” in *NDSS*, 2024.

[94] S. Köhler, R. Baker, and I. Martinovic, “Signal injection attacks against CCD image sensors,” in *ASIA CCS*, 2022, pp. 294–308.

[95] Y. Son, H. Shin, D. Kim, Y. Park, J. Noh, K. Choi, J. Choi, and Y. Kim, “Rocking drones with intentional sound noise on gyroscopic sensors,” in *USENIX Security*, 2015.

[96] J. Jeong, D. Kim, J. Jang, J. Noh, C. Song, and Y. Kim, “Un-rocking drones: Foundations of acoustic injection attacks and recovery thereof,” in *NDSS*, 2023.

[97] 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 *USENIX Security*, 2020.

[98] S. P. Arteaga, L. A. M. Hernandez, G. Sanchez-Perez, A. L. S. Orozco, and L. J. García-Villalba, “Analysis of the GPS spoofing vulnerability in the drone 3DR Solo,” *IEEE Access*, vol. 7, pp. 51 782–51 789, 2019.

[99] D. Hunt, K. Angell, Z. Qi, T. Chen, and M. Pajic, “MadRadar: A black-box physical layer attack framework on mmWave automotive FMCW radars,” in *NDSS*, 2024.

[100] 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 *IEEE S&P*, 2021.

[101] 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 *IEEE S&P*, 2021.

[102] 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)*, 2018.

[103] 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 *ACM CCS*, 2019.

[104] J. Tu, M. Ren, S. Manivasagam, M. Liang, B. Yang, R. Du, F. Cheng, and R. Urtasun, “Physically realizable adversarial examples for lidar object detection,” in *IEEE/CVF CVPR*, 2020.

[105] S. Chen, C. Cornelius, J. Martin, and D. H. P. Chau, “Shapeshifter: Robust physical adversarial attack on faster R-CNN object detector,” in *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD*, 2018.

[106] C. Ma, N. Wang, Z. Zhao, Q. Wang, Q. A. Chen, and C. Shen, “ControlLoc: Physical-world hijacking attack on visual perception in autonomous driving,” *arXiv preprint arXiv:2406.05810*, 2024.

[107] R. Muller, Y. Man, Z. B. Celik, M. Li, and R. M. Gerdes, “Physical hijacking attacks against object trackers,” in *ACM CCS*, 2022.

[108] Y. Jia, Y. Lu, J. Shen, Q. A. Chen, H. Chan, Z. Zhong, and T. Wei, “Fooling detection alone is not enough: Adversarial attack against multiple object tracking,” in *8th International Conference on Learning Representations, ICLR 2020*. OpenReview.net, 2020. [Online]. Available: https://openreview.net/forum?id=rJl31TNYPr

[109] F. Nesti, G. Rossolini, S. Nair, A. Biondi, and G. C. Buttazzo, "Evaluating the robustness of semantic segmentation for autonomous driving against real-world adversarial patch attacks," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022.

[110] 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 *USENIX Security*, 2021.

[111] N. Wang, S. Xie, T. Sato, Y. Luo, K. Xu, and Q. A. Chen, "Revisiting physical-world adversarial attack on traffic sign recognition: A commercial systems perspective," *arXiv preprint arXiv:2409.09860*, 2024.

Appendix

A. Synthetization Methodology

This section outlines the methodology for synthesizing the three vulnerability effects on thermal images that we characterized and formalized in Section IV.

Linearization Effect. To synthesize the pixel intensity drop, we first collect real-world heat source traces to accurately replicate the intensity distribution patterns they produce in thermal images. We use the FLIR Boson camera for this purpose, as it provides access to raw pre-equalized thermal images necessary for accurately modeling the pixel intensity drop. Next, we synthesize the heat source traces onto the raw images of the FLIR ADAS dataset by converting them to 14-bit depth to match the dataset format and overlaying them onto the validation images. We then apply linear equalization to the augmented images, using the formalization described in Section IV-A. This results in the final 8-bit images used for object detection.

Calibration Artifacts. As detailed in Section IV-B, NUC algorithms can be manipulated to introduce an additional offset, resulting in unintended artifacts in the thermal image. Adversaries can exploit this mechanism to induce controlled artifacts in thermal images by using highly reflective materials, such as aluminum foil cut in specific patterns and shapes (e.g., human poses), and positioning them in front of a heat source. This approach selectively blocks thermal radiation, creating artifacts only at adversary-defined regions in the image. Based on this approach, we synthesize controlled artifacts by selectively reducing pixel intensities according to the desired pattern, emulating the effects of sustained exposure on the camera during calibration. We further modulate the pixel intensity dro based on I_t and t_{max} as discussed in Section IV-B and adjust the artifact scale to simulate heat sources of different distances and locations.

Ghost Artifacts. Using the same methodology for controlled calibration artifacts described above, an adversary can selectively create flares and structured ghosts in thermal images. These ghosts appear as overlays on the background pixel data captured by the thermal camera. To emulate this effect, we synthesize ghost artifacts in the FLIR dataset by increasing the pixel intensity within adversary-targeted regions, based on the heat source temperature as characterized in Section IV-C.

B. Ethics Considerations

In adherence to ethical guidelines, we follow a responsible disclosure process and have shared our findings with the vendors of the tested devices. At the time of writing, we are currently awaiting their response. This study utilizes publicly available datasets to ensure transparency and reproducibility. All real-world case scenarios involving pedestrians were conducted by the authors. No human studies were conducted as part of this research.

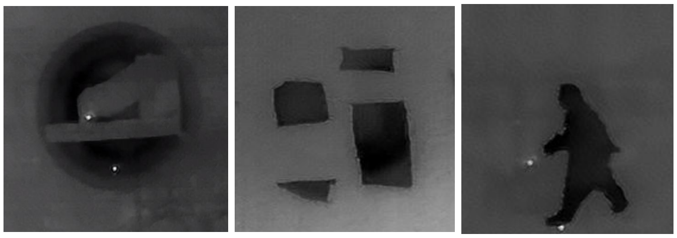


Fig. 18: Real-world arbitrary calibration-induced artifacts in the thermal image generated by applying reflective material such as aluminum foil over a heat source.

Appendix

Artifact Appendix

A. Description & Requirements

The artifact includes scripts and models used to evaluate the vulnerabilities presented in the paper, *The Heat is On: Understanding and Mitigating Vulnerabilities of Thermal Image Perception in Autonomous Systems*, https://dx.doi.org/10.14722/ndss.2026.230330. It also contains datasets for both simulated and real-world evaluations, along with detailed instructions to set up the models, run the evaluation tests, and extract results. Additionally, the artifact provides scripts to validate the defense mechanisms proposed in the paper.

1) How to access: Our research artifacts, including the simulated data, the models and scripts to evaluate them, along with the real-world experimental data for evaluating the vulnerabilities, and the scripts for testing the proposed defense methodologies are available at: https://zenodo.org/records/17051228, Artifact DOI: *10.5281/zenodo.17051228*.

We include a README file in the repository to guide the reviewers through the setup of the environment and the information regarding the provided data corresponding to the experiment in the paper.

2) Hardware dependencies: Our implementation does not support certain experiments on non-x86_64 systems as the required libraries are compiled for x86_64. While most experiments are architecture-independent, evaluating the equalization attack on FasterRCNN models requires an x86_64 CPU.

3) Software dependencies: To run the inference codes, users need Python 3.9+, a Conda environment with ultralytics, torch, and paddledetection. An optional Docker implementation is provided and tested on Ubuntu and Windows, running an Ubuntu container. It is not intended for macOS, as included dependencies (e.g., mmcv 2.1.0, PaddleDetection) are compiled for x86_64; building on macOS (Apple Silicon/ARM) causes architecture mismatches. macOS users should set up the environment natively outside Docker.

4) Benchmarks: We use the FLIR ADAS_v2 dataset for attack synthesis and evaluation. However, the synthesized images are included in the provided artifact, and no additional data download is required.

B. Experiment Workflow

The artifact provides comprehensive validation of the proposed attack and defense methods, including evaluation scripts and pre-trained weights to reproduce top attack results on three object detection models (YOLOv5, YOLOv8, Faster R-CNN) and two sensor fusion models (Faster R-CNN, DMASDet). It also includes real-world data from outdoor experiments to assess attack effectiveness and scripts to evaluate the proposed defenses, supporting analysis of the paper's results.

C. Major Claims

The artifact supports validation of the key evaluation claims in the paper. Focusing on the most significant results due to data and computational constraints, it includes all components needed to reliably replicate these core findings.

- (C1): Simulation: The key evaluation results for the three vulnerabilities across the tested object detection models are as follows:
  (i) *YOLOv5* : Equalization - pedestrian mAP@50 < 0.1, calibration Attack Success Rate (ASR) = 93%, ghost attack - ASR = 99%, showing the best attack results illustrated in Section V of the paper. This is proven by experiments (E1).
  (ii) *YOLOv8* : Equalization - pedestrian mAP@50 < 0.1, calibration ASR = 91%, ghost ASR = 99%. These are the best attack results illustrated in Section V of the paper. This is proven by experiments (E2).
  (ii) *FasterRCNN* : Equalization - pedestrian mAP@50 $<$ 0.1, calibration ASR = 100%, ghost ASR = 100%. These values are the best attack results illustrated in Section V of the paper. This is proven by experiments (E3).
  (iv) *FasterRCNN-Fusion* : Equalization - mAP@50 = 0.24, calibration ASR = 69%, ghost ASR = 91%. This is proven by experiments (E4).
  (v) *DAMSDet* : Equalization - mAP@50 = 0.55, calibration ASR = 56%, ghost ASR = 98%. This is proven by experiments (E5).
- (C2): Real World: The key evaluation results for the vulnerabilities in real world conditions are as follows:
  (i) Equalization : 100% ASR in both static and dynamic conditions, proven by experiments (E6).
  (ii) Calibration : 100% and 95% ASR in static and dynamic scenarios, proven in experiments (E6).
  (iii) Ghost : ASR of 100% is achieved in static scenarios across 10 frames. This is proven by experiments (E6).
- (C3): Defense: The key evaluation results for the proposed defenses against vulnerabilities in real world conditions are as follows:
  (i) Equalization : We achieve 100% success rate pedestrian obstacle is detected after mitigating the equalization vulnerability. This is proven by experiments (E7).
  (ii) Calibration : The proposed attack aware algorithm suppresses the calibration artifacts on the image with 100% success rate. This is proven by experiments (E7).
  (iii) Ghost : The proposed attack aware algorithm suppresses the ghost artifacts on the image with 100% success rate. This is proven by experiments (E7).

D. Evaluation

Here, we provide the steps to follow and the commands to run for experiments and validating the results claimed in our paper. The README attached with the artifact submission repeats the below process and details the setup instuctions.

1) Experiment (E1): [YOLOV5] [5 human-minutes + 1 compute-hour]: Demonstrates the results of the three attacks tested on the simulated FLIR dataset with YOLOv5 model.

[How to] The data and scripts required to run the evaluation are provided in the artifact:

[Preparation] Move into the cd Simulation/object_detection path to execute the commands

[Execution] Run the following commands as described in the README to execute

```
python yolov5_eval.py calibration_simulated results/
    v5_calibration_results
python check_iou.py results/v5_calibration_results
python yolov5_eval.py ghost_simulated results/
    v5_ghost_results
python check_iou.py results/v5_ghost_results
python yolov5_val.py
```

[Results] The equalization experiment result will show mAP@50 for pedestrian below 0.1. The ASR for calibration and ghost attacks is 93.9% and 99.9%, respectively.

2) Experiment (E2): [YOLOV8] [5 human-minutes + 1 compute-hour]: Demonstrates the results of the three attacks tested on the simulated FLIR dataset with YOLOv8 model.

[How to] The data and scripts required to run the evaluation are provided in the artifact:

[Preparation] Move into the cd Simulation/object_detection path to execute the commands

[Execution] Run the following commands as described in the README to execute

```
python yolov8_eval.py calibration_simulated results/
    v8_calibration_results
python check_iou.py results/v8_calibration_results
python yolov8_eval.py ghost_simulated results/
    v8_ghost_results
python check_iou.py results/v8_ghost_results
python yolov8_val.py
```

[Results] The equalization experiment result will show mAP@50 for pedestrian below 0.1. The ASR for calibration and ghost attacks is 91.3% and 99.7%, respectively.

3) Experiment (E3): [FasterRCNN] [5 human-minutes + 1 compute-hour]: Demonstrates the results of the three attacks tested on the simulated FLIR dataset with FasterRCNN model.

[How to] The data and scripts required to run the evaluation are provided in the artifact:

[Preparation] Move into the cd Simulation/ThermalAttack-master/mmdetection path to execute the commands

[Execution] Run the following commands as described in the README to execute

```
python tools/infer.py configs/faster_rcnn/faster-
    rcnn_r50_fpn_1x_coco_flir_finetune.py
    faster_rcnn_epoch_6.pth ./../../object_detection
    /calibration_simulated/ --out
    calibration_results
python check_iou.py calibration_results
python tools/infer.py configs/faster_rcnn/faster-
    rcnn_r50_fpn_1x_coco_flir_finetune.py
    faster_rcnn_epoch_6.pth ./../../object_detection
    /ghost_simulated/ --out ghost_results
python check_iou.py ghost_results
python tools/test.py configs/faster_rcnn/faster-
    rcnn_r50_fpn_1x_coco_flir_finetune.py
    faster_rcnn_epoch_6.pth
```

[Results] The equalization experiment result will show mAP@50 for pedestrian below 0.1. The ASR for calibration and ghost attacks is 100%.

4) Experiment (E4): [FasterRCNN-SF] [5 human-minutes + 1 compute-hour]: Demonstrates the results of the three attacks tested on the simulated FLIR dataset with FasterRCNN model with IR+RGB fusion.

[Preparation] Move into the cd Simulation/ThermalAttack-master/mmdetection path to execute the commands

[Execution] Run the following commands as described in the README to execute

```
python tools/infer_msf.py configs/faster_rcnn/faster
    -rcnn_r50_fpn_1x_align_msf.py faster-
    rcnn_r50_fpn_1x_align_msf.pth ./data/
    calibration_msf/ --out calibration_res_msf
python recheck_ASR.py calibration_res_msf/results/
    calib.txt
python tools/infer_msf.py configs/faster_rcnn/faster
    -rcnn_r50_fpn_1x_align_msf.py faster-
    rcnn_r50_fpn_1x_align_msf.pth ./data/ghost_msf/
    --out ghost_res_msf
python recheck_ASR.py ghost_res_msf/results/ ghost.
    txt
python tools/test.py configs/faster_rcnn/faster-
    rcnn_r50_fpn_1x_align_msf.py faster-
    rcnn_r50_fpn_1x_align_msf.pth
```

[Results] The equalization experiment result will show mAP@50 for pedestrian below 0.25. The ASR for calibration and ghost attacks is 69% and 91%, respectively.

5) Experiment (E5): [DAMSDet] [5 human-minutes + 5 compute-hour]: Demonstrates the results of the three attacks tested on the simulated FLIR dataset with DAMSDet model.

[Preparation] Move into the cd Simulation/DAMSDet/-DAMSDet path to execute the commands

[Execution] Run the following commands as described in the README to execute

```
python tools/multi_infer.py -c configs/damsdet/
    damsdet_r50vd_flir.yml --infer_vis_dir=dataset/
    coco_FLIR_align/val_imgs/vis_imgs --infer_ir_dir
    =dataset/coco_FLIR_align/val_imgs/calib --
    output_dir=calib_results -o weights=
    flir_best_model.pdparams
python recheck_ASR.py calib_results/ calib_ref.txt
python tools/multi_infer.py -c configs/damsdet/
    damsdet_r50vd_flir.yml --infer_vis_dir=dataset/
    coco_FLIR_align/val_imgs/vis_imgs --infer_ir_dir
    =dataset/coco_FLIR_align/val_imgs/ghosts --
    output_dir=ghost_results -o weights=
    flir_best_model.pdparams
python recheck_ASR.py ghost_results/ ghost_ref.txt
```

[Results] The equalization experiment result will show mAP@50 for pedestrian below 0.55. The ASR for calibration and ghost attacks is 43.6% and 98.0%, respectively.

6) Experiment (E6): [Real World] [5 human-minutes + 15 compute-minutes]: Demonstrates the results of the equalization, calibration, and ghost attacks on the real world data.

[Execution] Perform the evaluation of the FasterRCNN model by running the command *./real_world.sh*

[Results] The expected real world results are described in C2.

7) Experiment (E6): [Defense] [5 human-minutes + 15 compute-minutes]: Demonstrates the results of the proposed defense methodologies in real world data.

[Execution] Perform the evaluation of the FasterRCNN model by running the command *./defense.sh*

[Results] The expected results are described in C3.