I Can See the Light: Attacks on Autonomous Vehicles Using Invisible Lights

Wei Wang, Yao Yao, Xin Liu, Xiang Li[§], Pei Hao, Ting Zhu
University of Maryland Baltimore County, Cambricon SingGo[§]
{ax29092,yaoyaoumbc,xinliu1}@umbc.edu,lixiang@cambricon.com,{phao1,zt}@umbc.edu

ABSTRACT

The camera is one of the most essential sensors for an autonomous vehicle (AV) to perform Environment Perception and Simultaneous Localization and Mapping (SLAM). To secure the camera, current autonomous vehicles not only utilize the data gathered from multiple sensors for environment perception and SLAM but also require the human driver to always realize the driving situation, which can effectively defend against previous attack approaches (i.e., creating visible fake objects or introducing perturbations to the camera by using advanced deep learning techniques). Different from their work, in this paper, we in-depth investigate the features of Infrared light and introduce a new security challenge called I-Can-See-the-Light-Attack (ICSL Attack) that can alter environment perception results and introduce SLAM errors to the AV. Specifically, we found that the invisible infrared lights (IR light) can successfully trigger the image sensor while human eyes cannot perceive IR lights. Moreover, the IR light appears magenta color in the camera, which triggers different pixels from the ambient visible light and can be selected as key points during the AV's SLAM process. By leveraging these features, we explore to i) generate invisible traffic lights, ii) create fake invisible objects, iii) ruin the in-car user experience, and iv) introduce SLAM errors to the AV. We implement the ICSL Attack by using off-the-shelf IR light sources and conduct an extensive evaluation on Tesla Model 3 and an enterprise-level autonomous driving platform under various environments and settings. We demonstrate the effectiveness of the ICSL Attack and prove that current autonomous vehicle companies have not yet considered the ICSL Attack, which introduces severe security issues. To secure the AV, by exploring unique features of the IR light, we propose a software-based detection module to defend against the ICSL Attack.

CCS CONCEPTS

• **Security and privacy** → *Side-channel analysis and countermeasures.*

KEYWORDS

Security; Autonomous Vehicle

CCS '21, November 15–19, 2021, Virtual Event, Republic of Korea

ACM ISBN 978-1-4503-8454-4/21/11…$15.00
https://doi.org/10.1145/3460120.3484766

ACM Reference Format:
Wei Wang, Yao Yao, Xin Liu, Xiang Li[§], Pei Hao, Ting Zhu. 2021. I Can See the Light: Attacks on Autonomous Vehicles Using Invisible Lights. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS '21), November 15–19, 2021, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3460120.3484766

1 INTRODUCTION

Recent autonomous vehicle (AV) developments have shown great potential to improve driving quality and safety. With the support of governments and private investments, the number of autonomous vehicles on the road is estimated to reach 745,705 by the end of 2023 while the market size is projected to be valued at USD 24.73 Billion by 2027 [16, 17]. To enable self-driving, one of the most fundamental components is environment perception, which utilizes multiple sensors (e.g., Camera, Ultrasonic Sensor, Radar, or LiDAR) to sense the environment. Then, based on the data gathered from those sensors, a simultaneous localization and mapping (SLAM) process is conducted to make proper driving decisions. During these procedures, the camera is one of the most important sensors for the AV to understand its surroundings and current locations.

To avoid the severe consequences such as injuries or fatalities generated by the camera malfunctions, prior studies have been conducted to investigate the potential security risks of the camera. For example, researchers have shown that the camera can be easily blinded by the light with high intensities [57, 77]. Moreover, by applying deep neural networks, RP_2 is even able to create fake stop sign to mislead the real road sign recognition system [38]. Built on the top of RP_2, the attacker can introduce physical adversarial perturbations to fool the image-based Faster R-CNN object detectors [35]. However, these works either require the attacker to have advanced knowledge of the hardware or required advanced machine learning techniques. To overcome these challenges, researchers have demonstrated that an unskilled attacker can alter the perception results of the camera by creating fake traffic signs or objects using a projector [52]. Recently, according to the requirements of current autonomous vehicles, the human driver should be always aware of the driving conditions and be prepared to take control of the vehicle. Therefore, since the fake traffic signs or objects created by these attack methods can be perceived by the human eyes, the driver can easily detect the potential attacks and manually control the AV in this scenario. Despite the failure of the attack, the successful detection of the attack may also result in reporting to the police officers for further investigation, which will increase the possibility of exposing the identity of the attacker.

In this paper, we propose the work that in-depth investigates the possibility of attacking the camera using invisible lights. Specifically, we found that human eyes cannot perceive the lights with

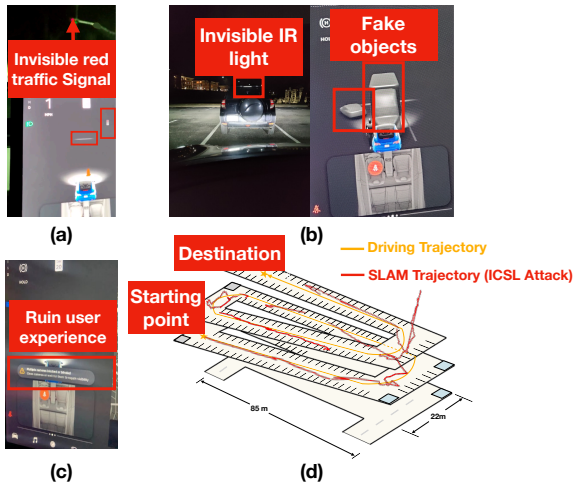


Figure 1: An example of ICSL Attack: (a) Generate invisible fake red traffic signal; (b) Create invisible fake objects; (c) Intentionally create system alters to ruin the in-car user experience; and (d) Introduce SLAM localization errors. (During the experiment, human cannot see the IR light.)

wavelengths larger than 740nm while Infrared lights (IR lights) can be detected by the cameras on the AV [9]. Moreover, the invisible IR lights appear the magenta color in the camera, which is exactly the same as the visible magenta light. In addition, since the IR light triggers different pixels from the ambient lights in the camera, the SLAM system in the autonomous vehicle tends to select the IR light source as the key point for AV localization and mapping. Based on these observations, we propose the I-Can-See-the-Light Attack (ICSL Attack) by using IR lights. Different from simply blinding the camera using visible lights [57, 77], we explore to leverage the features of IR lights to **i)** generate invisible traffic signals, **ii)** create fake invisible objects, **iii)** ruin the in-car user experience, and **iv)** introduce SLAM errors to the autonomous vehicle. We believe that by carefully selecting the attack scenarios and setting up the attack, the AV can be attacked without the notice from the human driver.

To do this, we first analyze the architecture of the autonomous vehicle. Then, we perform experiments to understand the features of IR light and conduct a survey of human eyes to select the suitable IR lights to attack the autonomous vehicle (in section 2). By leveraging the features of IR light, we demonstrate the effectiveness of the ICSL Attack by altering the environment perception results of a Tesla Model 3. A simple example is shown in Figure 1 (a), ICSL Attack successfully generates an invisible red traffic light by using IR light. Based on the experiment results, we also discuss the related parameters that may affect the attack successful rate (in section 4.1.1). As shown in Figure 1 (b) and Figure 1 (c), to show severer security risks introduced by ICSL Attack, we continue to demonstrate how attackers can utilize ICSL Attack to create invisible objects (in section 4.1.2) and intentionally ruin the user experience (in section 4.2). To demonstrate the harmful impacts on the SLAM system of an AV, we utilize an enterprise-level autonomous driving platform to show the SLAM trajectories under ICSL Attack (in section 5.2), which is shown in Figure 1 (d). At last, by analyzing the SLAM errors, we propose a SLAM attack model to show how attackers can control the SLAM trajectory of an AV by using the ICSL Attack.

To defend against the ICSL Attack, we propose a software-based detection module to secure the autonomous vehicle. Specifically, it is possible to implement IR filters on the camera to filter out the IR lights. However, since there are more than 500, 000 autonomous vehicles to be delivered each year and each vehicle has multiple cameras [1], this hardware-based defense strategy will significantly increase the cost for the autonomous vehicle company. On the other hand, since the IR light appears the same as the visible magenta color to the camera of an AV, it is difficult for the AV to distinguish the IR lights from visible lights. To overcome this challenge, we found that the IR light absorption rate is higher than visible light. In other words, the camera cannot detect the reflections of IR lights from other objects. By leveraging this unique feature, we propose a simple and effective module to defend against the ICSL Attack. The experiment results show the effectiveness of the proposed module (in section 6).

The contributions of this paper can be summarized as follows:

- To the best of our knowledge, this is the first work that in-depth analyzes the effect of invisible IR light on the autonomous vehicle. According to our analysis, we propose the I-Can-See-the-Light Attack (ICSL Attack). We leverage the features of IR lights and introduce related parameters and models to **i)** generate invisible traffic signals, **ii)** create invisible objects, **iii)** ruin the in-car user experience, and **iv)** introduce SLAM errors to the AV.
- We conduct extensive real-world experiments by using a Tesla Model 3 and an enterprise-level autonomous vehicle platform under various scenarios and settings. The experiment results demonstrate the harmful impacts of the ICSL Attack.
- To defend against the ICSL Attack, we explore a unique feature of IR lights and propose a novel software-based detection module to secure the autonomous vehicle. The experiment results show the effectiveness of the proposed approach.

2 BACKGROUND

2.1 Environment Perception and SLAM in AV

Figure 2 shows a system architecture of an autonomous vehicle, which consists of six parts: Environment Perception, Pre-processing, Localization and Mapping, Sensor Fusion, Trajectory Prediction, and Decision Making. The Environment Perception utilizes Camera, Radar, Ultrasonic Sensor or USS, LiDAR, IMU, Wheel Speed Sensor, and RTK GPS to sense the surroundings and understand the current driving conditions (i.e., acceleration, orientation, and location, etc). The data gathered from these sensors will be processed by the Sensor Hub for synchronization and rectification, which will then be categorized into different layers in the pre-processing step. Meanwhile, the vision data from multiple cameras will be fused in the localization and mapping step. To improve the environment perception accuracy, the autonomous vehicle will perform sensor fusion to better understand its surroundings. At last, in the trajectory prediction and decision making steps, the fused data will be used to predict objects' trajectories and the driving decisions are made based on the prediction results.

As we can observe from this architecture, Environment perception and SLAM in the Localization and Mapping are two important parts for the vehicle to perform autonomous driving. For the environment perception, it is the 'eyes' for the autonomous vehicle to understand its surroundings. For the SLAM, it is not only important

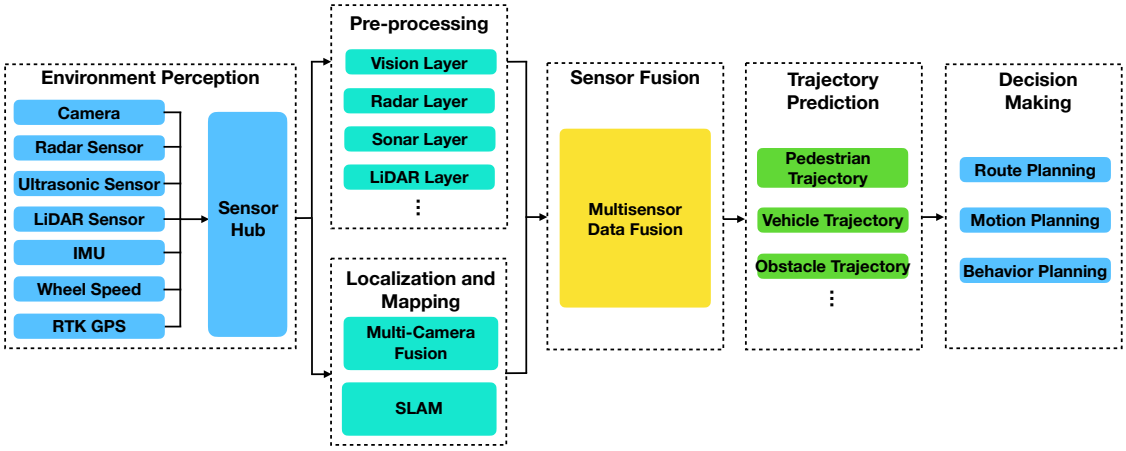
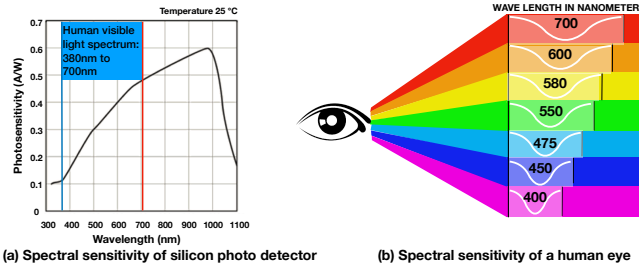


Figure 2: A system architecture of the autonomous vehicle



(a) Spectral sensitivity of silicon photo detector (b) Spectral sensitivity of a human eye

Figure 3: Spectral sensitivity of silicon photo detector and a human eye

(a) SONY IMX598 with 850nm IR light (b) SONY IMX689 with 850nm IR light

(c) SONY IMX598 with 940nm IR light (d) SONY IMX689 with 940nm IR light

Figure 4: The detected IR lights (850nm and 940nm) on SONY IMX598 and IMX689

for the AV to improve the environment perception accuracy and eliminate AV's blind spots but also essential for the AV to make driving decisions. Specifically, since the GPS sensor only provides a coarse location information in an outdoor scenario, it is insufficient for the AV to conduct navigation and perform self-driving. To overcome this challenge, the vision data gathered from multiple cameras is used to understand the location and build a local HD map to make driving decisions. For example, in an indoor parking lot, the GPS signal is weak. In this case, the AV analyzes each frame that captured by the camera and extracts the corresponding key points. Then, a matcher will be generated to find the same key points on consecutive frames. According to the position changes of the matched key points, the AV's location and trajectory is calculated. At last, the local HD map will be built according to the trajectory of the AV. By using this approach, the AV can understand its location and exit the parking lot according to the local map.

2.2 Basics of IR Light and Camera

Infrared light (IR light) is the electromagnetic radiation with wavelengths varies from $740nm$ to $3 \times 10^5 nm$, which has been widely used in industrial, scientific, military, commercial, and medical applications [4, 20]. For example, IR light can be utilized in security cameras and Night-vision devices to capture nighttime images. To do this, these devices need to detect the reflected lights from an object by using IR light detectors. According to the requirements of applications, the IR light detectors are made of different IR materials, such as Calcium fluoride, fused silica and sodium chloride [14]. Normally, these materials should be carefully selected according to their unique attributes in order to improve the IR light detection accuracy and reliability. However, we found that the industry image sensors that are not designed for IR light also can detect the IR light.

	Men	Women	Total
larger than $780nm$	3	4	7
larger than $900nm$	0	0	0

Table 1: A survey of the IR light visibility.

Specifically, current industry image sensors are either a Complementary Metal-Oxide Semiconductor (CMOS) or Charge-coupled Device (CCD), which uses arrays of silicon to convert the incident light (photons) into electronic charge (electrons). As shown in Figure 3 (a), since silicon has high sensitivity in both visible and invisible IR spectrum, the image sensor can be triggered to detect the invisible IR light [8]. However, human eyes are only able to detect lights with wavelengths from $380nm$ to $700nm$ [10], which is shown in Figure 3 (b). To prove the concept, we use $850nm$ and $940nm$ IR light LEDs (3W) to generate invisible light and test those lights on the most recent SONY IMX598 and IMX689 [22] image sensors. As we can observe in Figure 4, although different image sensors have different IR light sensitivities, both $850nm$ and $940nm$ IR lights still successfully trigger the image sensors. In addition, since the wavelengths of IR lights are close to the wavelength of red light, the invisible lights are detected as magenta color lights in the camera.

Ideally, the attacker can use any IR lights that close to $700nm$ to attack the AV without human driver's notice. As the wavelength

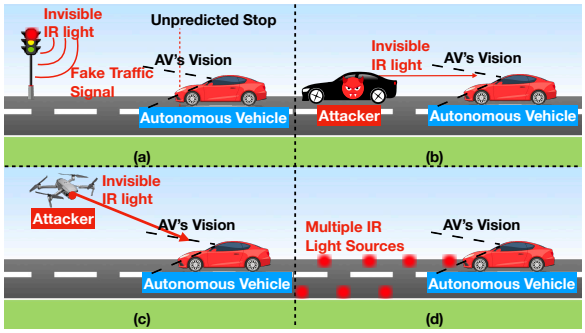


Figure 5: (a) The attacker creates fake invisible traffic signal; (b) The attacker uses IR light to alter AV's environment perception results; (c) The attacker can blind the AV's camera to create frequent system alters; (d) The attacker controls the IR sources on the road to alter AV's SLAM results.

of selected IR light close to 700nm, the detected light intensity by the camera will be high, which can increase the attack distance. However, we found that some people can perceive lights with wavelengths that slightly larger than 700nm. This is because IR light activates the human photoreceptors through a nonlinear optical process, which enables human to visualize the near IR light [54]. To successfully attack the AV without human notice, we conduct a survey of 100 men and 100 women with ages vary from 18 to 50 to see how many people can perceive IR lights. As shown in Table 1, only 3% of men and 4% of women can perceive IR lights with wavelengths larger than 780nm while no one can see the IR light with wavelengths larger than 900nm. Since the lower wavelengths IR lights can be detected by most of cameras with higher intensity and the color is close to red, in this paper, we mainly utilize 780nm – 850nm IR lights to implement the attack.

3 ATTACK OVERVIEW AND THREAT MODEL

In this section, we first present the overview of I-Can-See-the-Light Attack (ICSL Attack). Then, we introduce the threat model.

3.1 ICSL Attack Overview

Attack Goal. Our attack goal is to alter the environment perception and SLAM results of an autonomous vehicle embedded with different sensors (i.e., Camera, Radar, USS, LiDAR and GPS etc). To do this, we mainly attack the cameras on the AV by using invisible IR lights. As a result, the target vehicle will make unexpected harmful driving behaviors. In the worst case, the autonomous vehicle will wrongfully change its driving behavior, such as terminate the autopilot mode, reducing its driving speed or even make unexpected stop, etc. In order to make the human driver unaware of the ICSL attack and the presence of attackers, we mainly utilize IR light with wavelengths larger than 780nm to attack AV's cameras.
The Vulnerability. The AV is vulnerable to ICSL Attack for the following reasons:

I) Human cannot see IR lights. Previous attacks on AV's camera are mainly based on the visible light, such as creating visible objects [52] and toxic traffic signs [50, 70]. However, modern AVs require the human driver to always be aware of the driving conditions (i.e., Defense Driving Strategy.) Therefore, although these attacks can effectively attack the autonomous vehicle, they also can be detected by the human driver. Different from these attacks, ICSL mainly relies on the invisible light to attack the autonomous vehicle. By carefully setting up the attacks in proper driving scenarios, the human driver will not notice the IR light attacks.

II) The enterprise-level autonomous vehicle has to trust the data gathered from cameras. Normally, to increase the environment perception accuracy, the autonomous vehicle utilizes the data gathered from multiple sensors to perform environment perception. Then, during the sensor fusion process, an Extended Kalman Filter [3] is implemented to dynamically assign different gains (i.e., Kalman gains) to the sensors according to the current driving scenarios. For example, the Radar sensor suffers high wireless interference in the indoor parking lot, which will result in a relatively low Kalman gain. On the other hand, LiDAR will be assigned with a high Kalman gain to improve the environment perception accuracy. However, camera data is critical for the AV (e.g., perform object detection and recognition, localization and mapping, etc). It is important for an AV to understand the shape, color and texture of an object, which cannot be done by other sensors. As a result, the autonomous vehicle has to trust the camera to perform autonomous driving.

III) The Invisible IR light is detected as the visible magenta color light in the camera, which is difficult for an AV to verify if the light is reflected by a real object or generated by an IR light source. Moreover, since most people cannot see the IR light, the heavily relied on 'human eyes defense strategy' does not work properly. In other words, the driver is unaware of the ICSL Attack until the AV has already made unexpected harmful driving behaviors. To make the situation even worse, although multiple approaches have been proposed to defend against attacks on AV's cameras [44, 46, 52, 70], they are not designed to distinguish the IR lights from ambient lights. In addition, these approaches require complex deep learning algorithms and high computation resources, which is not suitable for an autonomous vehicle with strict cost restrictions.

IV) During the SLAM process, it is possible for the invisible IR source to be selected as the key points, which may vary the localization and mapping results. Specifically, in order to guarantee real-time processing, AVs mainly utilize ORB-SLAM-related architectures to extract key points from each frame to perform localization and mapping [33, 51]. The key points extraction process depends on the intensity weighted centroid of the patch in each frame. However, since IR lights are bright and significantly different from the background lights, the corresponding IR light pixels in the camera may be selected as the key points. Therefore, when the attacker moves the key points or turns off the IR lights, the SLAM results in an AV may suffer high variance.

3.2 Threat Model

We assume the attacker knows the positions of the cameras on the target autonomous vehicle. Since autonomous vehicle companies normally publicly announce their autonomous driving solutions and their hardware vendors for advertisement purposes, the attacker can easily get the camera specifications by browsing the internet. We also assume the attacker has the basic knowledge of IR lights. As shown in Figure 5, to perform the ICSL Attack, multiple IR light sources can either be deployed on attackers' vehicles, flying drones or on the road. In addition, we assume the attacker can set up the attack in the night or find proper attack scenarios to reduce the possibility of being detected by others. In addition, the attacker

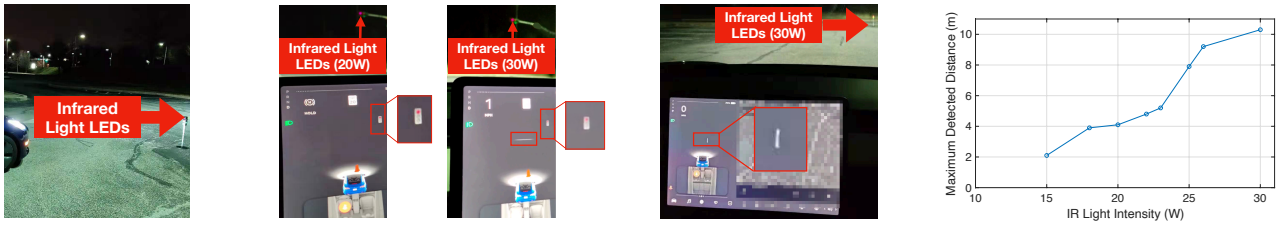


Figure 6: (a) Experiment setup: we implement the IR light LEDs in the traffic light to create fake invisible red signal. (b) Tesla detects the fake red light signal; (c) The height of the traffic light will affect the attack result; (d) The fake traffic signal detected distance vs. IR light intensity.

has an unlimited budget to purchase proper attack devices (i.e., drones and small IR LEDs). Specifically, the attacker can the create following harmful results by performing the following attack:

• **Alter Environment Perception Results.** As shown in Figure 5 (a), the attacker can deploy the IR light source on the traffic light to create fake traffic signals. Since the IR light is close to the red color in the AV's camera, the AV may make an unpredicted stop, which will result in a traffic jam or potential accident. In Figure 5 (b), an attacker can also drive the vehicle close to the target AV and uses IR light to alter AV's environment perception results.

• **Ruin the In-car User Experience.** In Figure 5 (c), the attacker can fly a drone with equipped IR light source to blind AV's target cameras. By doing this, the AV will frequently show system alert messages. Since human cannot see the IR light, the alert messages may be considered as system bugs and ruin the in-car user experience.

• **Introduce SLAM Errors.** In Figure 5 (d), by deploying multiple IR light sources on the road, the AV will select several detected IR light sources in each frame as the key points for SLAM propose. Then, by dynamically changing IR light sources' positions and light intensities, the AV will surfer relatively high SLAM errors.

4 ICSL ATTACKS ON TESLA

In this section, we mainly use ICSL Attack to **1)** alter environment perception results and **2)** ruin the in-car user experience. We use iPhone 12 pro to take photos in this paper. During the entire experiment, human eyes cannot see IR lights.

4.1 Alter Environment Perception Results

In this experiment, we show how ICSL Attack can alter environment perception results on Tesla Model 3. Then, we analyze the security insight and discuss the related parameters that the attacker should determine to improve the attack success rate.

4.1.1 Create Fake Invisible Traffic Light. We first create the fake invisible traffic signal to alter the environment perception results of Tesla, which is similar to the attack scenario in Figure 5 (a).

Experiment Setup. As shown in Figure 6 (a), since we cannot make modifications to a real traffic light on the road, we embed IR light LEDs on a smaller traffic light model ($31.19\times25.60\times13.69cm$) to create fake invisible red traffic signals. We believe that it is sufficient to prove the effectiveness of ICSL Attack. In practice, a real attacker can implement IR LEDs with color close to red on a real traffic light to conduct the attack.

Experiment Result. Figure 6 (b) shows the attack results when the IR light intensities are $20W$ and $30W$. Tesla detects the invisible IR light and considers it as the red traffic signal. Moreover, when the IR light intensity is $30W$, Tesla also shows the line to make the stop. On the contrary, the human driver cannot see the red traffic signal generated by IR light in this scenario.

Analysis. Since Tesla does not validate the size of the traffic light, even a smaller traffic light model is considered as a legitimate traffic light. Moreover, the color of the IR light in the camera is close to the red. As a result, Tesla detects the fake invisible IR light and considers it as a legitimate red traffic signal. According to this experiment, the attacker can either modify the existing legitimate traffic light on the road to create the invisible red traffic signal or even build a smaller invisible traffic light to change the driving behavior of an AV without human driver notice. Since the human cannot see the IR light, the AV will make unpredicted driving behavior before the driver has a chance to take control.

In Figure 6 (c) and (d), we analyze the parameters that will affect ICSL Attack, including the height of the traffic light, the light intensity, and the distance to the AV. In Figure 6 (c), we show that the height of the traffic light will affect the attack results. Since the height of the traffic light is normally fixed, the fake traffic light with a smaller height will be misclassified as an obstacle. According to our experiment, to avoid attack failure, the attacker should make sure that the height of the fake traffic light is at least larger than $2.45m$. As shown in Figure 6 (d), we show the maximum attack distances under different IR light intensities. When the IR light intensity is smaller than $15W$, Tesla cannot detect the fake traffic light regardless of the distance. As the IR light intensity increases, the distances for Tesla to detect the fake traffic signal also increase. When the light intensity reaches $30W$, the maximum attack distance is around $10m$. Therefore, the attacker should at least use the $15W$ IR light and make sure the height of the traffic light is higher than $2.45m$ in order to successfully attack Tesla.

4.1.2 Create Fake Objects. In this experiment, we assume the attacker is driving in front of the AV and uses a drone with equipped $850nm$ IR light sources to create fake objects, which is similar to the attack scenario in Figure 5 (b).

Experiment Setup. In this section, we show how to utilizes ICSL Attack to create fake objects. Since Tesla utilizes a triple forward camera and fuses the perception results to increase the detection accuracy, simply attacking a single camera will not alter the environment perception results. Therefore, we first utilized a DJI Robot Master S1 [13] equipped with a $850nm$ IR light source to blind the right main forward camera and the narrow forward camera, which is shown in Figure 7 (a). Then, we utilized a drone equipped with six

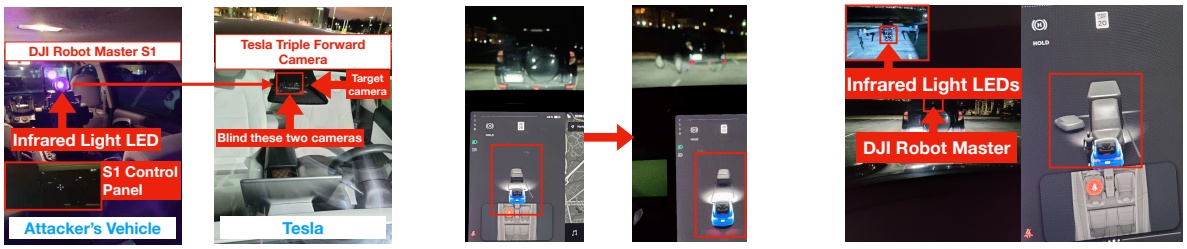


Figure 7: (a) Experiment setup: we use the IR light to blind the right main forward camera and the narrow forward camera of Tesla. (b) The environment perception results of Tesla is accurate (only one vehicle is in front of Tesla); (c) By implementing ICSL Attack on a drone and attacking the right main forward camera, Tesla detects multiple surrounding objects.

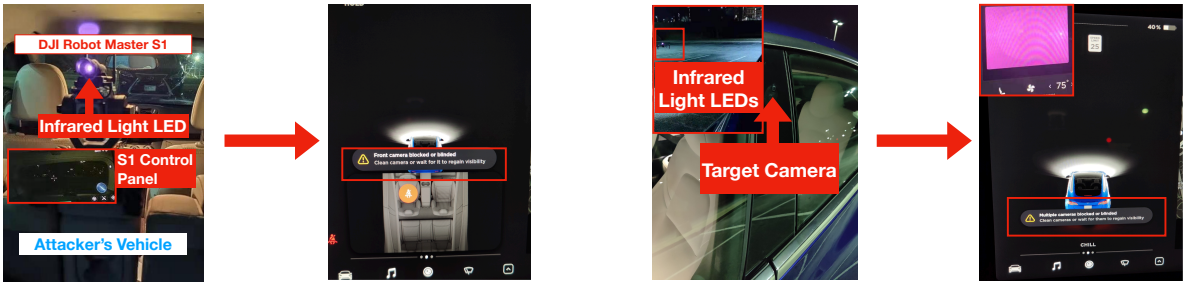


Figure 8: (a) Experiment setup: Blind the triple forward cameras of Tesla. (b) System Alert is triggered. (c) Experiment setup: Blind the left side camera. (d) System Alert is triggered.

850*nm* IR light LEDs (3W) to attack the left main forward camera, which is shown in Figure 7 (c).

Experiment Result. Figure 7 (b) shows the ground truth when Tesla is not under ICSL Attack. In this scenario, Tesla successfully detects the front vehicle even if the vehicle is making a left turn (only the left main forward camera and the narrow forward camera can 'see' the vehicle). However, as shown in Figure 7 (c), when we utilize drones equipped with IR light LEDs to attack the left main forward camera, the environment perception results of Tesla shows that there are two vehicles (one sedan and one truck) and a pedestrian is in front of the Tesla and the sedan is making a left turn.

Analysis. In this experiment, since the narrow forward camera and the right main forward camera are blinded by the IR light, only the left main camera is used to detect the front objects. Therefore, when this camera is under ICSL Attack, Tesla still has to believe the information provided by the left camera. Moreover, during the entire experiment, Tesla's system alert is not triggered. This is because we only blind two cameras while the left main camera can still detect the front objects. In addition, Tesla tends to consider the IR light sources equipped on the drone as the rear position lamps from a legitimate vehicle. As a result, multiple fake objects are detected by Tesla. During the experiment, since the human cannot see the IR light, the ICSL Attack is not detected by the human driver. However, we need to mention that in practice, it is possible for the passengers or the drivers in other vehicles (have different filed of views) to see the suspicious drone. In addition, the noise generated by drones is around 65dB, which may be heard by the human driver. Therefore, the attacker should carefully select the attack scenarios (i.e., night scenario with limited vehicles on the road) and use a better drone with lower noise to remain stealthy.

4.2 Ruin In-Car User Experience

In this experiment, we show how to create frequent system alerts to ruin the In-Car user experience, which is similar to the attack scenario in Figure 5 (c).

Experiment Setup. During this experiment, we assume the attacker can either drive in front of the target AV or utilizes drones to attack the AV. Specifically, as shown in Figure 8 (a), the attacker utilizes IR light to blind the Triple Front Camera (two main forward cameras and one narrow forward camera) of Tesla. In addition, the attacker also can utilize drones to blind the side camera of Tesla. For example, the left side camera is blinded in Figure 8 (c).

Experiment Result. As we can see from Figure 8 (b), Tesla generates system alerts to inform the drive that the front camera is blocked or blinded. In this scenario, since the IR light is invisible, the human driver will get confused and annoyed about system alerts. In 8 (d), by attacking the left side forward camera, Tesla triggers the system alert again. In this figure, we also show an example of the view from the blinded camera. However, we need to mention that the human driver cannot access the forward camera data while the vehicle is moving. Therefore, although the camera detects the invisible light, it is still difficult for the human driver to be aware of the presence of the ICSL Attack.

Analysis. In this experiment, we found that the system alerts will only be triggered when the **entire** triple front camera is blinded by the IR light, which introduces a severe security risk. This experiment also answers why the system alert is not triggered when the attacker is trying to create fake objects in the previous section 4.1.2. Since Tesla fuses the data gathered from multiple cameras to understand

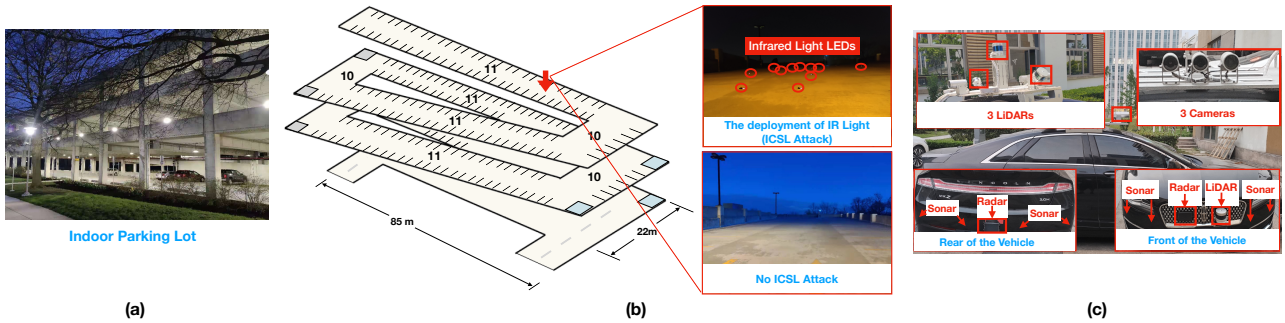


Figure 9: The experiment setup.

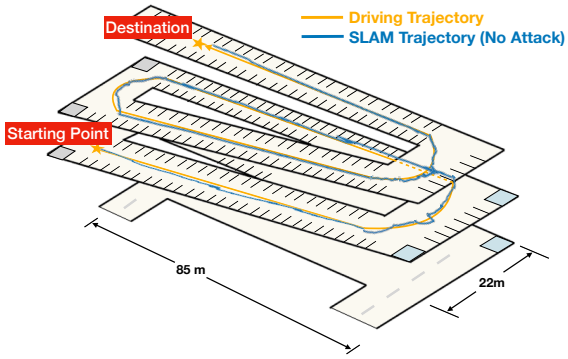


Figure 10: SLAM trajectory: no ICSL Attack.

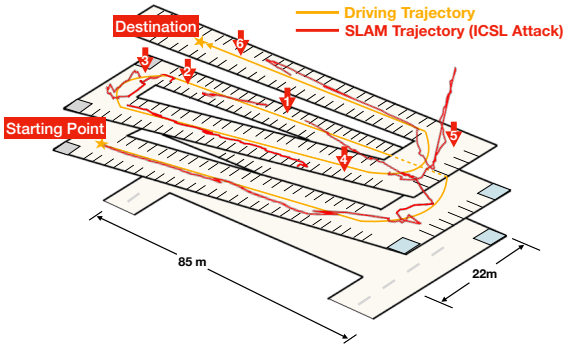


Figure 11: SLAM trajectory: under ICSL Attack.

the driving scenario, these cameras are complement to each other. Therefore, even if several cameras are interfered by IR lights, the driving performance and in-car user experience are not affected.

5 ICSL ATTACKS ON SLAM SYSTEM

To show that it is possible to introduce SLAM errors to the AV, we utilize an enterprise-level autonomous driving platform to analyze the ICSL Attack. Similar to section 4, all the photos are taken by iPhone 12 pro and human eyes cannot see any IR lights during the experiment. In this section, we first introduce the experiment setup and experiment results. Then, we analyze the security insights. At last, we introduce a SLAM attack model to show how to manipulate the SLAM trajectory.

5.1 Experiment Setup

We conduct the experiment in an indoor parking lot, which is shown in Figure 9 (a). There are 74 850nm IR light LEDs deployed in the parking lot. Each corner is deployed 10 IR light LEDs and each floor is deployed 11 IR light LEDs. These LEDs are blinking according to their own schedules, which is shown in Figure 9 (b). As shown in Figure 9 (c), we use an enterprise-level autonomous vehicles - 2018 Lincoln MKZ 2.0H (fully loaded) to evaluate ICSL Attack. The front of the autonomous vehicle is equipped with one triple forward camera, four Ultrasonic Sensors, one LiDAR, and one Mid-Range Radar. The rear of the vehicle is equipped with four Ultrasonic Sensors and one Mid-Range Radar. The roof of the autonomous vehicle is equipped with three LiDARs. The AV utilizes the most commonly used ORB-SLAM2 architecture [5, 51] to perform the SLAM process. Due to the safety concern, the steering speed of the vehicle is set to 4mph and the straight-line speed is set to 5mph. The frame rate is set to 30fps. During the visual odometry (VO) process, the autonomous vehicle extracts FAST key points [61] and uses an ORB descriptor [62] to find the matched key points every 6 frames. Then, according to the matched key points, the motions (rotation and translation matrices) of the camera (AV) are estimated. To improve the estimation accuracy, the AV will search the local map in its dataset in order to find the accurate position of the matched key points. At last, bundle adjustment (BA) will be used to better estimate the location of the autonomous vehicle.

During the experiment, we have the access to the private parking lot. However, in practice, the deployment process of IR light LEDs will increase the possibilities of being detected by others. Specifically, although the IR light cannot be perceived by human eyes and the LEDs are small, it is still possible for the employees in the parking lot (i.e., managers, sweepers etc.) to see the suspicious behavior of the attacker or find small LED devices on the floor or on the wall. Therefore, to remain stealthy, the attacker should carefully select the locations of IR LEDs and conduct the deployment when no one is in the parking lot.

5.2 Introduce SLAM Errors

Experiment Result. The experiment results are shown in Figure 10 and 11. When the autonomous vehicle is not under ICSL Attack, the calculated SLAM trajectory is smooth and close to the actual driving trajectory. In this scenario, even if the indoor parking lot does not have the GPS signal, the AV can navigate itself and find the entrance and exit. On the contrary, when the AV is under ICSL Attack in Figure 11, the calculated SLAM trajectory cannot provide any useful information. Moreover, as we can see from the red arrows 1, 2 ,3 ,4 and 6 in this figure, the VO process in SLAM is not working properly and the trajectory is discontinuous while arrow 5 shows that the vehicle is driving to the sky.

Analysis. In this section, we analyze why the calculated SLAM trajectory under ICSL Attack cannot provide any useful information

to navigate the AV. As shown in Figure 12, we plot the matching results during the VO process in SLAM between two consecutive frames. Since the IR light provides significantly different pixels from the ambient light, the AV considers the IR light LEDs as the key points and finds all the matches between two consecutive frames. Moreover, since all the IR light LEDs are the same and the background of the indoor parking lot cannot provide any useful information to distinguish between different IR lights, it is highly possible for the AV to mismatch those IR lights and gets the wrong trajectory.

Specifically, to calculate the trajectory of the AV, we first define the rotation matrix and translation matrix of the AV's camera as R and t, respectively. Since the rotation and translation matrices belong to the lie group SE(3), we have:

$$SE(3) = \{T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{4\times4} | R \in SO(3), t \in \mathbb{R}^3\} \tag{1}$$

Assume the homogeneous coordinate of the detected key point P is $P = (x, y, z, 1)^T$ and the corresponding projected point's coordinate on the camera is $Q = (u, v, 1)^T$. Then, we can represent the relationship between the key point and the projected point as:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \approx (K e^{\xi^\wedge} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix})_{1:3} \tag{2}$$

where s is the depth information of the detected key point and K is the camera matrix that will be provided in the specifications of the camera. $\xi^\wedge$ represents the skew-symmetric matrix of lie algebra for the homogeneous matrix T ($T \in SE(3)$) in equation 1. The left side of this equation is a three dimensional vector while the right side of this equation is a 4×1 vector. Therefore, in order to calculate the pose ξ of AV's camera, we only need to match the first three rows of the right side to the left side in order to make the above equation hold, which is denoted as $(1:3)$. Due to the noise introduced during the environment perception process, the following equation is leveraged to minimize calculated pose error:

$$\xi^{opt} = \arg\min_{\xi} \frac{1}{2} \sum_{i=1}^{n} ||Q_i - \frac{1}{s_i} K e^{\xi^\wedge} P_i||_2^2 \tag{3}$$

Therefore, we can find the optimized camera pose ξ by a linearization process. We first denote $(Q_i - \frac{1}{s_i} K e^{\xi^\wedge} P_i)$ as $e(\xi)$. Then, the corresponding linearization form can be represented as $e(\xi + \Delta\xi) = e(\xi) + J\Delta\xi$, where J is the Jacobian matrix of $e(\xi)$ and can be calculated by left multiply the perturbation δ: $J = \frac{\partial e}{\partial \delta \xi}$. Formally, the camera matrix K can be represented as:

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where f_x and f_y are the focal lengths while c_x and c_y are considered as the principal point of the camera. Finally, the Jacobian matrix J can be calculated as:

$$J = -\begin{bmatrix} \frac{f_x}{z} & 0 & -\frac{f_x x}{z^2} & -\frac{f_x xy}{z^2} & f_x + \frac{f_x x^2}{z^2} & -\frac{f_x y}{z} \\ 0 & \frac{f_y}{z} & -\frac{f_y y}{z^2} & -f_y - \frac{f_y y^2}{z^2} & \frac{f_y xy}{z^2} & \frac{f_y x}{z} \end{bmatrix} \tag{5}$$

As we can observe in this equation 5, since f_x and f_y are fixed according to the camera matrix, the three-dimensional rotation matrix R is affected by the first three columns while the three-dimensional translation matrix t is affected by the column from four to six. Therefore, when the ICSL Attack introduces mismatches between two consecutive frames, the optimization direction is wrong, which introduces a wrong R and t. As a result, the calculated trajectory is significantly affected by the ICSL Attack.

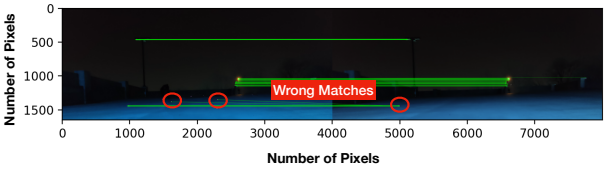


Figure 12: An example of key point matching errors between two consecutive frames.

SLAM attack model. According to the above analysis, in this section, we show how to manipulate the SLAM trajectory. Specifically, as shown in Figure 12, IR light LEDs introduce mismatches between two consecutive frames. Therefore, by changing the blinking IR light LEDs, the mismatches between two consecutive frames will also be changed, which will result in the change of the corresponding optimization directions J. Formally, we define the mismatched key point (with a different IR LED on) and its corresponding Jacobian matrices are $P' = (x', y', z', 1)^T$ and J', respectively. Then, in order to change the orientation of the AV from the original $R = (r_x, r_y, r_z)^T$ to $R' = (r'_x, r'_y, r'_z)^T$, the attacker should make sure that the first three columns of J' satisfy:

$$\begin{cases} \frac{f_x x' + f_y y'}{z'^2} \geq \frac{f_x x + f_y y + z^2 (r'_z - r_z)}{z^2}, \\ z' \geq \frac{z f_x}{f_x - z(r'_x - r_x)}, \end{cases} \tag{6}$$

Similarly, to alter the translation matrix from $t = (t_x, t_y, t_z)^T$ to $t' = (t'_x, t'_y, t'_z)^T$, the attacker should make sure the last three columns of J' satisfy:

$$\begin{cases} \frac{f_x y' + f_y y'^2}{z'^2} \geq \frac{f_x xy + f_y y^2 + z^2 (t'_x - t_x)}{z^2}, \\ \frac{f_x y' - f_y x'}{z'} \geq \frac{f_x y - f_y x - z(t'_z - t_z)}{z}, \end{cases} \tag{7}$$

In practice, the maximum detection range of the forward camera is around $80m$ and the angle of view is around $120°$ [24]. Assume the speed of the target AV is v and the frame rate of the camera is f_r. Then, in order to change the optimization direction, the IR light LEDs in the range $\frac{\pi}{3}(\frac{80v}{f_r})^2$ that follows the above equations 6 and 7 should start to blink with the frequencies higher than f_r.

6 POTENTIAL SOLUTION

In this section, we propose a lightweight lCSL Attack detection module to defend against ICSL Attack without requiring any hardware modifications by utilizing a unique feature of IR light.

6.1 Unique Features of IR Light

Since the IR light shows a legitimate magenta color in AV's camera, the camera cannot distinguish if the magenta color is from a legitimate visible light or from the IR light source, which makes it challenging to defend against ICSL Attack. Intuitively, it is possible to implement IR filters on the camera to defend against ICSL. Attack. However, IR filters will introduce severe disadvantages, including i) high implementation cost, and ii) filtering out useful information.

i) High implementation cost. Each AV uses at least 6 cameras (3 front cameras, 2 side cameras, and 1 rear camera) to perform autonomous driving. Since the price of each camera is around

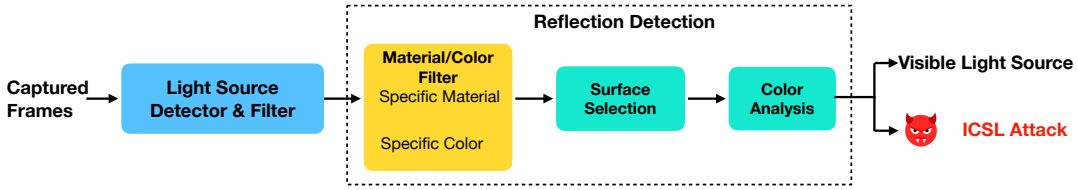


Figure 13: The overview of the detection module.

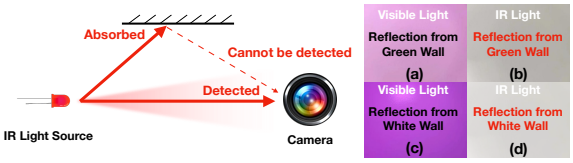


Figure 14: The camera (SONY IMX 689) cannot detect the 850*nm* IR light reflected from an object. (a) and (c): the camera detects the visible magenta color light reflected from the wall. (b) and (d): the camera cannot detect reflected IR light.

$1000 while the price of the IR filter is around $100 to $150 [18, 26], the hardware cost will be increased by 10% to 15%. For the AV companies, it is important to reduce the cost to win the ever-increasing competition in the autonomous driving field.

ii) Filtering out useful information. The autonomous driving systems rely on the IR light intensity, exposure value and tone curve from the IR light to make the decision, especially under severe lighting scenarios (i.e., strong sunlight, high ambient light noise, etc.). Moreover, the front and side cameras of the AV (i.e., Tesla Model 3 and Tesla Model S, etc.) are RCCB or RCCC cameras. Different from traditional RGGB cameras, these cameras are designed for computer vision and very sensitive to red and blue color. Since most of the light sources (i.e., moonlight, vehicle light, halogen lamp, etc.) also emit IR light energy. By detecting IR lights, RCCB or RCCC cameras will significantly reduce the exposure time, increase the camera dynamic range and the camera performance at night. However, filtering out the IR light will reduce the night performance of RCCC or RCCB cameras.

In this work, instead of requiring hardware modifications, we leverage a unique feature of IR light to defend against ICSL Attack. Specifically, since the IR light absorption rate is higher than the visible light [19, 55], when the IR light reaches a surface of an object, most of its energy will be absorbed while the reflected energy will be low. Therefore, **it is difficult for the camera to detect the IR light reflected from the surface of an object.** To prove the concept, we conduct experiments to compare the reflected visible magenta color lights and 850*nm* IR lights from walls with different colors. As shown in Figure 14 (a) and (c), when visible magenta color lights are transmitted to the wall, the camera can still detect the reflected lights. In addition, the detected light color is also changing according to the color of the wall. On the other hand, most of the 850*nm* IR lights are absorbed by the wall. The camera cannot detect the reflected IR lights from the wall and the wall shows its original color. According to this unique feature, instead of directly detecting the IR light source, the AV can detect ICSL Attack by detecting the light reflected from the surface from other objects.

6.2 Overview of the Detection Module

According to the unique feature of IR light, we introduce one possible way to detect ICSL Attack. As shown in Figure 13, the proposed detection module mainly consists of two parts: the Light Source Detector & Filter module and the Reflection Detection module.

Light Source Detector & Filter. To defend against ICSL Attack, the first step to is to recognize the light source. Since the wavelength of IR light is close to 700*nm*, the color of the light shown in the camera (magenta) is close to red. Therefore, according to the frames captured by the camera, the detection module should also validate the colors of the lights. If the detected colors are significantly different from red, the corresponding light sources will be ignored and considered as legitimate devices while the remaining light sources will be transmitted to the reflection detection module for further analysis.

Reflection Detection. The Reflection Detection module analyzes the remaining light sources that pass through the Light Source Filter. Since the IR light reflected from other objects is weak and cannot be detected by the commodity camera, the detection module can utilize this feature to distinguish the IR light from visible light. Specifically, the Reflection Detection module contains three parts to analyze the reflected light:

• Material/Color Filter. Since some specific materials or objects with specific colors have high IR light reflection factors, it is possible for the camera to detect the IR light reflected from those objects and misclassify those objects as light sources. For example, most of the car paints are made of acrylic enamel or urethane [2], which can reflect the IR light. In addition, since the bright silver color can slightly reflect the IR light, the camera can detect the reflected IR light at a close distance. Therefore, in this paper, the detection module will ignore the red lights that are reflected from the body of the vehicle or a bright color object.

• Surface Selection. Since different materials have different reflection factors [15], it is important that the detection module selects the right surface of the object to analyze the reflected light. If the reflected light is not detected, then the corresponding light source should be considered as the IR light source.

• Color Analysis. Based on the color and intensity of the reflected lights, the detection module should match reflected lights to the corresponding light sources. Since the color of the surface will affect the color of the reflected visible light, the detection module should also recognize the reflected visible light even if the reflected light color is not the same as the color of light source.

6.3 Experiment Setup

6.3.1 Light Source Detector & Filter. It is possible to recognize light sources using filters and thresholds. However, different light

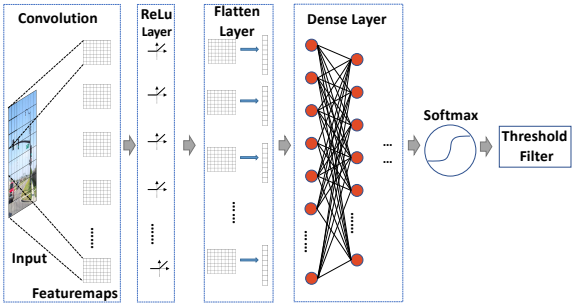


Figure 15: The Light Source Detector & Filter module.

sources may have different configurations. For example, a traffic light can be suspended or supported. In addition, the viewing angle will also differ with the position of the car and the elevation of the light. To get the best result, we applied a convolutional neural network for traffic light recognition. The structure of the network is shown in Figure 15.

We applied a two-dimensional convolution layer to the input, which consists of 32 filters. The output featuremap was fed into a maximum pooling layer to eliminate the difference made by location and rotation. Then, an activation layer with the ReLu function was adopted to process each of the pooled featuremaps. After the activation layer, a flatten layer reduced the dimension of the last output for the following dense layer. The dense layer converts the problem into a regular deeply connected neural network layer. Then, a softmax activation layer gives the light source detection output. At last, a color filter is implemented based on the color threshold in order to filter out the light source that is not close to the red color. During the experiment, we combined three datasets to train the model. The first dataset is the Bosch Night dataset [7] and the second is the Kitti dataset [28], which are famous and widely used in autonomous vehicle experiments. The third dataset is the dataset we self-recorded while driving the vehicle in a downtown area and in an indoor parking lot.

6.3.2 Reflection Detection. The first step of reflection detection is to filter out the potential materials or objects that may affect the detection results. To do this, we leverage one of the most commonly used deep learning-based object detection approaches–YOLOv5 [6, 60]. Similar to section 6.3.1, we use the Bosch Night dataset [7], the Kitti dataset [28] and the self-recorded dataset to train the detection model. In order to filter out the bright color objects, a threshold-based color filter is also implemented.

Then, the key step of reflection detection is to choose the proper reflection surface for detection. As stated before, we need to choose a surface that reflects visible red light instead of IR light. The most ideal surface would be the road since the concrete or pitch surface absorbs the IR light but reflects the visible light. In the meantime, we need to rule out the surface of other cars since their paint would reflect IR light.

The next step is color analysis. Matching the reflected light to the light source is challenging, since there could be multiple light sources that pass the Light Source Detector & Filter module. To overcome this challenge, we leverage the fact that although those lights sources are transmitting red lights, the colors and light intensities detected by the camera are different. This is because the

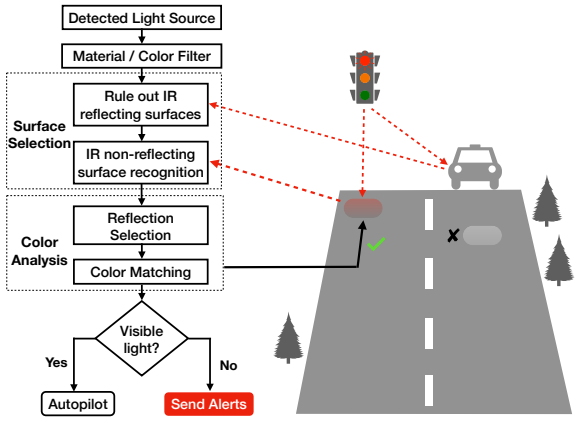


Figure 16: The workflow of reflection detection.

detected colors and light intensities are affected by the position, distance, and angle of view of the autonomous vehicle. Since the camera is much more sensitive than the human eyes, the detected light sources will show slightly different colors (i.e., crimson, magenta and maroon, etc) with different light intensities. Therefore, by matching the reflected light to the light source according to the colors and intensities, we can find the remaining light sources, which can be categorized as IR light sources.

The workflow of our reflection detection is shown in Figure 16. We first apply the Material/Color Filter object recognition to rule out the special materials and the objects with special colors. Then, to improve the detection accuracy, we also rule out the objects with smooth surfaces, such as road signs and mirrors. Then the IR non-reflecting surfaces are picked. During the experiment, the road surface is always the first choice, since the position, color, and area are easy to identify.

On the IR non-reflecting surfaces, we apply a threshold to pick out the area of reflection. The threshold is dynamically determined based on the position of the light source. Since the light source is already recognized, we can find out the reflection of light by using color matching. Finally, if we get a solid reflection, we take it as the legitimate light source. Otherwise, the detection module can send alerts to the driver.

6.4 Experiment Results

As shown in Figure 17, we study the proposed defense strategy by plotting Receiver Operating Characteristic (ROC) curves for the Light Source Detector & Filter, the Material/Color Filter, and the Combined Detection Module. We also calculate the corresponding Area Under the ROC Curves (AUC) to analyze the performance of the detection module. The block dot line in this figure serves as the reference line to represent the performance of random guessing (AUC = 0.5).

As we can observe from Figure 17, the Light Source Detector & Filter performs well during the experiment (AUC = 0.99). This module provides the reliable light source information for the detection module to recognize the light sources and filter out the light that is not close to the red color. The Material/Color Filter in the Reflection Detection module also performs well (AUC = 0.99), which can effectively recognize and filter out the special materials and objects with special colors. The overall AUC of the proposed

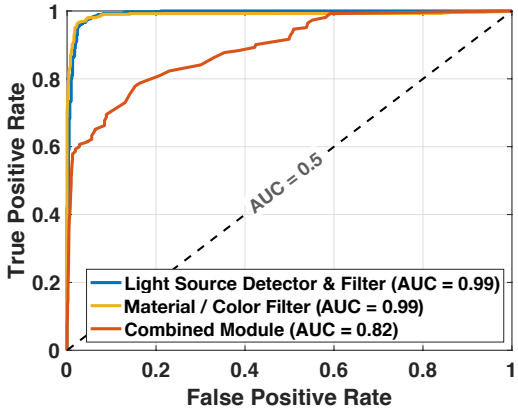


Figure 17: The Receiver Operating Characteristic (ROC) curves for the detection module.

ICSL Attack detection module is 0.82. This is because we set several thresholds to conduct the reflection surface selection. In real-world scenarios, especially at night, the light conditions are complex and hard to predict. In our experiment, although we utilize the road surface to analyze the reflected light, the light interference from the environment is severe, which reduces the accuracy of the final results. It is possible to utilize advanced deep learning algorithms to predict the potential reflection surface. However, this approach typically requires high computation resources and does not support real-time processing, which is not suitable for autonomous vehicles [36, 39]. According to our experiment, we believe that 0.82 AUC is sufficient so that the human driver to be aware of a potential attack.

7 DISCUSSIONS & ATTACK LIMITATIONS

In this section, we first discuss the attack generality and attack stealthiness. Second, we will analyze the precision requirements for the attacker to control the IR light source. Then, we will introduce several important issues that may affect the attack results. At last, we will discuss the trade-offs between different defense solutions.

7.1 Attack Generality

The proposed ICSL Attack mainly focuses on the camera embedded in the autonomous vehicle. To alter the environment perception results, the attacker can drive in front of the target AV or use a drone equipped with IR light LEDs to attack the AV's forward camera. Moreover, the attacker can also deploy the IR light LEDs on the road to attack the AV's SLAM system. We believe that the proposed attack has the following advantages, which introduce severe security risks to the autonomous vehicle.

IR light is invisible. As mentioned in section 2.2, human eyes cannot see the lights with wavelengths larger than $740nm$. Therefore, even if the current autonomous vehicle requires the human driver to always be aware of the surroundings, it does not provide any benefit against the ICSL Attack. On the contrary, the camera detects the IR light and simply considers the IR light as visible light to make driving decisions. The mismatches between human eyes and the camera introduce security risks to the AV. For example, in this paper, we have shown that the attacker can build an invisible red traffic light to change the driving behaviors of AV. In practice, we believe that more magenta objects (i.e., emergency light, fire alarm light, etc.) may be spoofed by the ICSL Attack.

Does not require any specific designed hardware. Instead of requiring specifically designed hardware, we mainly utilize the off-the-shelf IR light LEDs to implement the ICSL Attack, which is cheap (around $8/each) and can be purchased by anyone. As a result, the attacker can easily and effectively attack the autonomous vehicle with very little hardware or software knowledge.

7.2 Attack Stealthiness

The stealthiness of the proposed attack is guaranteed as long as the human cannot see the IR light. However, as mentioned in previous sections, the implementation process of ICSL Attack may increase the risks of exposing the identity of the attacker. Specifically, to create a fake traffic light, the attacker should implement the IR light sources on the traffic signal before the autonomous vehicle detects the signal. This suspicious process will significantly increase the possibilities of being detected by pedestrians or police officers. Therefore, it is important for the attacker to conduct the implementation process during the night to mitigate the potential risks of being seen by others. To introduce environment perception error, the attacker should drive the vehicle in front of the target AV and utilize drones equipped with IR lights to attack the target AV. According to our experiments, although it is difficult for the human driver to see or hear the drones during the night, it is still possible for passengers or other vehicles to see or detect the suspicious drone, which may increase the risk of attack failure. For the SLAM attack, the attacker should implement multiple small IR light sources in the parking lot, which may increase the risk of exposing the identity of the attacker.

7.3 Precision Requirements

As introduced in section 4.1.2 and section 4.2, the attacker should control the IR light source to aim the target cameras on the AVs. However, in practice, the speed of the AV and the distance between the target AV and the attacker will affect the attacker results. Specifically, as the speed increases, the attack success rate will be reduced. This is because the larger speed will introduce higher vibration (i.e., engine vibration and high speed wind vibration) to the target vehicle and the attacker's vehicle. Since the size of the forward camera for the autonomous vehicle (i.e., Tesla model 3) is limited (i.e., around 20cmx20cm), the attacker should make sure that the spot generated by the IR light source is in the correct area, which requires the attacker to precisely control the direction of IR light source. This requirement is relatively hard to achieve due to the non-linear properties introduced by the vibration and vehicle motions. Fortunately, since the IR light is continuous and travels at light speed, the attacker can conduct an extremely high number of attacks in a very short time with no delay. Therefore, it is relatively easy for the attacker to attack the autonomous vehicle.

When the distance between the attacker and the target vehicle increases, the attacker has to use a high power IR light source to attack the autonomous vehicle, which will significantly increase the cost of the hardware. Moreover, the larger distance will introduce larger IR light spot on the target vehicle, which will trigger the system alerts. Therefore, the attacker should carefully control the

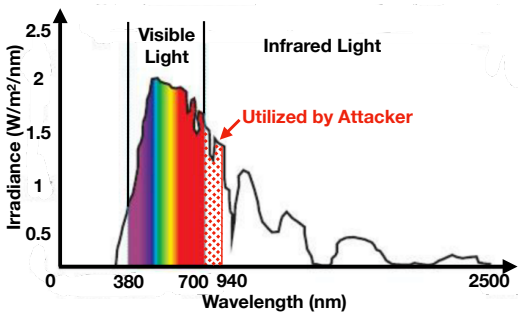


Figure 18: The spectrum of Solar Radiation and the ICSL Attack

mounted aspheric lens on the IR light source according to the attack scenarios.

The changing rate of the autonomous vehicle heading will also affect the attack results. Formally, we represent the vehicle heading as φ and the changing rate of the heading can be represented as $\dot{\varphi}$. Generally, the heading of the autonomous vehicle is determined by the steering angle δ, the wheelbase length (the length of the front axle to the center of the gravity l_f plus the length of the rear axle to the center of the gravity l_r) and the speed of the vehicle v [30, 42, 72]. Then, according to the vehicle dynamics, the changing rate of the vehicle heading can be represented as:

$$\dot{\varphi} = \frac{v}{l_f + l_r} tan(\delta) \tag{8}$$

As shown in this equation, despite the impact of high speed, the change of the steering angle will affect the heading of the AV. In practice, to maintain proper lane position, the AV should control its steering angle to adjust its lateral position. For example, the wheelbase of Tesla model 3 is 2875mm and the maximum steering angle is around 36 degrees. Since the weight distribution of Tesla model 3 is 50/50, the maximum changing rate of the vehicle heading for Tesla can be modeled as: $\dot{\varphi} = 0.289v$. Therefore, the attacker should make sure the minimal changing rate of the aiming direction should at least satisfy the above equation. Assume the distances between the attacker and the Tesla model 3 is between 5m to 15m, when the speed of the Tesla model 3 is 40mph, the approximate changing rate of the aiming direction should be in the range: $0.5°/s - 1.8°/s$.

7.4 Attack Environments & Limitations

Impact of Solar Radiation. Due to the power limitation of the IR light LEDs, in this paper, the ICSL Attack is mainly conducted during the night to prove the concept. This is because the solar radiation (i.e., electromagnetic energy from the sun) will interfere with the low-power IR light. As we can see from Figure 18, although most of the solar radiation is visible light, the lights from 700nm to 940nm (i.e., overlapped with the ICSL attack) still receives high radiation power. As a result, although the camera can still detect the IR light, the detection results are unpredictable. However, in practice, we argue that the attacker can always purchase high-power IR light sources to mitigate the impact of the solar radiation interference.

Impact of GPS. To alter the SLAM localization and mapping results, we mainly attack the AV in an indoor parking lot. In this scenario, since there is no GPS signal available, the AV has to rely on the data gathered from the camera. In the outdoor scenario, the autonomous vehicle may fuse the data gathered from GPS to correct its calculated SLAM trajectory, which may reduce the impact of ICSL Attack. However, since lots of autonomous vehicle companies mainly focus on the Automated Valet Parking in the indoor parking lot scenario and the key part of their solution is to conduct HD mapping based on SLAM [11, 12, 27], we believe that the ICSL Attack introduces security risks to the autonomous vehicle.

Impact of Cameras. Since the proposed attack relies on the camera to detect the IR light, it is important to analyze the types of shutter and the color filter arrays in the camera.

• **Types of shutter.** There are two important types of shutters in the camera: global shutter and rolling shutter. For the global shutter, it exposes all the pixels in a frame at the same time while for the rolling shutter, it exposes the pixels in a frame sequentially (line-by-line). However, these two types of shutter will not affect the attack results. For the environment perception attack, the attacker continuously transmit the IR light to the camera. Therefore, as long as the camera can detect the IR light, it will be affected by the attack. For the SLAM attack, the IR light is flashing. However, the flashing rate for the IR light is lower than 1Hz, which is much less than the minimum shutter speed for cameras (30 Hz). Therefore, it is highly possible for the camera to capture the IR light.

• **Color filters arrays.** Currently, the autonomous vehicles mainly utilize RCCC (Red Clear Clear Clear), RCCB (Red Clear Clear Blue) and RGGB (Red Green Green Blue) color filter arrays. Generally, RCCC and RCCB color filter arrays are designed for computer vision purposes. Although they cannot achieve high color accuracy, these types of filters have high dynamic range and good performance during the nighttime. Therefore, according to their advantages, these cameras are normally utilized as the forward cameras and side cameras in the AVs. To be more specific, for the RCCC filter, it is highly sensitive to red light and IR light. Therefore, the RCCC camera is vulnerable to ICSL attack. For the RCCB filter, it is sensitive to red and blue color. This type of filters are the filters utilized in the front and side cameras in the Tesla Model 3. As shown in this paper, the camera with RCCB filter is also vulnerable to ICSL Attack.

Different from RCCC and RCCB cameras, the RGGB color filter array is designed for human eyes. It has the highest color accuracy. Normally, the camera with RGGB color filter arrays is utilized as the rear camera of the AV. This type of camera also can detect the IR light. For ICSL Attack, since the RGGB camera has the highest color accuracy, it can detect the fake traffic signal based on the color of the red light. However, in practice, since the RGGB camera has low dynamic range and poor performance in the nighttime, it is not utilized as the forward camera and side camera. Therefore, the ICSL Attack remains unaffected.

7.5 Trade-offs Between Defense Solutions

The proposed ICSL attack can alter the environment perception and SLAM results of the AV. For the environment perception attack, it can introduce fake objects or fake traffic signals, which may result in potential car accidents. For the SLAM attack, it introduces high localization error to the AV, which will affect the performance of the AV or even result in navigation to a wrong destination.

In practice, to analyze the performance and the safety of the AV, miles per disengagement (MPD) is one of the most important

metrics [23], which measures the total miles that the AV can perform autonomous driving without detecting the system failure or requiring the human driver to disengage the autonomous mode due to the safe operation requirements. Normally, according to the different AV companies, the MPD varies from 0.49 to 29,944.69 [25]. However, as shown in Figure 11, the autonomous vehicle requires 6 interventions in around 0.27 miles in a parking lot. The corresponding MPD is around 0.045, which is much worse than the MPD of the AV without ICSL attack and the performance of the AV is hampered

As mentioned in section 6, to defend against the ICSL attack, the easiest way is to implement IR filters in the camera. In this case, since the IR light is filtered out, the AV is not vulnerable under ICSL attack. This approach is robust under different driving scenarios and does not require complex machine learning models.

However, this solution also introduces new challenges. First, since most of the light sources (i.e., the Moonlight, vehicle light, halogen lamp, etc.) emit IR light energy, the camera in the AV can utilize these lights to improve the quality of the captured frame at night [41]. By filtering out the IR light, the performance of the camera in the AV will be hampered. Second, after implementing the IR filter, the color information of the frames captured by the camera will be inaccurate or wrong [31, 71], which will affect the environment perception results.

To overcome the first challenge, the AV company can implement LiDAR (around $5000 [21]) on the AV and develop algorithms or architectures to fuse the environment perception results of LiDAR and camera together. To overcome the second challenge, color correction and calibration should be conducted, which requires engineers to not only build complicated models but also manually correct the color according to different scenarios and types of the cameras. This process is time-consuming and will increase the cost of the AV.

As introduced in section 6, another solution is to explore the features of IR lights and utilize machine learning techniques to distinguish IR lights and visible lights. This solution is a software-based approach and does not require additional hardware or redesign of the camera, which reduces the cost of the AV. However, since the driving scenarios of the autonomous vehicles are various, it requires the AV company to collect a huge amount of data to train the model, which is time-consuming. Moreover, different from the IR light filter solution, the performance of this software-based approach is highly sensitive to different light conditions, object materials and weather conditions. Therefore, the robustness of this approach is not as high as the IR light filter solution.

8 RELATED WORK

Attacks on Autonomous Vehicles. Extensive prior works have been conducted to explore the security risks of Autonomous Vehicle [32, 34, 47, 53, 58, 65–68, 75, 77]. In the field of sensor attack, by utilizing advanced hardware, researchers have shown that the Ultrasonic Sensor is vulnerable to jamming, Denial of Service (DoS) and the delay injection attack [37, 56]. For the LiDAR sensor, the one of the most common attack methods are LiDAR spoofing attacks [34, 58, 67]. The vulnerability of Mid-range Radar is also widely studied by multiple researchers [48, 77]. Instead of requiring specially designed hardware, the attack on AV's camera is relatively easy for the attacker to implement, which introduces severe security risks. For example, the attack can simply deploy toxic signs to misleading the autonomous vehicle [50, 70]. One of the most recent work Phantom Attack successfully changes the driving decisions of AV by projecting a phantom (i.e., a fake image or traffic sign) on the road by using a drone equipped with a projector [52]. According to these researches, current autonomous driving system requires the human driver to always ber aware of the driving conditions, which helps the AV to defend against the potential attacks.

Different from their works, in this paper, we propose ICSL Attack, which utilizes invisible IR light to alter the environment perception results of AV. Since AV's camera can detect the IR light, it considers the invisible lights as real objects without the human driver's notice. As a result, the human driver defense strategy is not working, which introduces severe security risks to current AV.

Potential Defense Strategies. According to the existing security risks, lots of work has been proposed to secure AV and defend against potential attacks [40, 43, 59, 69, 73]. For example, SAVIOR mainly leverage the physical invariants to validate the data gathered from GPS/IMU sensors [59]. PyCRA utilizes the random probes transmitted by the sensor and validate the received signal to detect the potential attacks [69] while PGFUZZ provides a framework to locate the bugs in robotic vehicle's control software. In order to defend against attacks on computer vision, the most common solution is to utilize machine learning (ML) techniques to identify the adversarial attacks [29, 49, 74]. The researcher can either utilize adversarial training [63], modifying existing ML networks [45, 64] or propose new models [52, 76] to defend against potential attacks. For example, the author proposes a novel detection module including context model, surface model and light model to validate the frame captured by the camera [52]. However, these approaches mainly focus on the visible light. Since IR light and visible magenta light appear the same to the camera, tradition approaches cannot work to defend against the ICSL Attack.

Different from their works, we leverage a unique feature of IR light to defend against ICSL Attack. By analyzing the reflections from objects, we can distinguish the IR light from visible light with high accuracy.

9 CONCLUSION

In this paper, we propose the first work that explores the threat of IR light to autonomous vehicles (AV) and introduce the I-Can-See-the-Light Attack (ICSL Attack), which can effectively i) generate invisible traffic lights, ii) create fake objects, iii) ruin the in-car user experience, and iv) introduce SLAM errors to the autonomous vehicle without human notice. To defend against the ICSL Attack, we explore the features of IR light and introduce a novel lightweight software-based detection module to secure the autonomous vehicle. We also believe that it is easy for the attacker to implement the IR light related attack, which requires the AV company to be aware of the threat of the IR light.

10 ACKNOWLEDGEMENT

This project is partially supported by NSF grants CNS-1652669 and CNS-1824491

REFERENCES

[1] 2021. https://cleantechnica.com/tesla-sales/.
[2] 2021. https://en.wikipedia.org/wiki/Automotive_paint.
[3] 2021. https://en.wikipedia.org/wiki/Extended_Kalman_filter.
[4] 2021. https://en.wikipedia.org/wiki/Infrared.
[5] 2021. https://github.com/raulmur/ORB_SLAM2.
[6] 2021. https://github.com/ultralytics/yolov5.
[7] 2021. https://hci.iwr.uni-heidelberg.de/content/bosch-small-traffic-lights-dataset.
[8] 2021. https://hub.hamamatsu.com/jp/en/technical-note/WITS-guide-detector-selection/index.html.
[9] 2021. https://science.nasa.gov/ems/09_visiblelight.
[10] 2021. https://science.nasa.gov/ems/09_visiblelight.
[11] 2021. https://www.bosch.com/stories/automated-valet-parking/.
[12] 2021. https://www.daimler.com/innovation/case/autonomous/driverless-parking.html.
[13] 2021. https://www.dji.com/robomaster-s1.
[14] 2021. https://www.edmundoptics.com/knowledge-center/application-notes/optics/the-correct-material-for-infrared-applications/.
[15] 2021. https://www.engineeringtoolbox.com/light-material-reflecting-factor-d_1842.html.
[16] 2021. https://www.gartner.com/en/newsroom/press-releases.
[17] 2021. https://www.globenewswire.com/news-release/2020/12/07/2140428/0/en/.
[18] 2021. https://www.mouser.com/ProductDetail/FRAMOS/Depth-Camera-D455e-Starter-Kit?qs=QNEnbhJQKvZvXjcsxT4qKw%3D%3D.
[19] 2021. https://www.perkinelmer.com/lab-solutions/resources/docs/TCH_reflection-Measurements.pdf.
[20] 2021. https://www.redsun.bg/en/infraredheating/infrared-heat-waves/.
[21] 2021. https://www.robotshop.com/en/m8-1-ultra-lidar-sensor.html.
[22] 2021. https://www.sony-semicon.co.jp/e/products/IS/camera/.
[23] 2021. https://www.statista.com/chart/17144/test-miles-and-reportable-miles-per-disengagement/.
[24] 2021. https://www.tesla.com/autopilot.
[25] 2021. https://www.therobotreport.com/cruise-waymo-lead-way-calif-autonomous-vehicle-tests/.
[26] 2021. https://www.uqgoptics.com/catalogue/filters/cut-off-blocking-filters/ir-cut-off-filters/.
[27] 2021. https://www.wipro.com/engineeringNXT/.
[28] 2021. http://www.cvlibs.net/datasets/kitti/.
[29] Naveed Akhtar and Ajmal Mian. 2018. Threat of adversarial attacks on deep learning in computer vision: A survey. Ieee Access 6 (2018), 14410–14430.
[30] Mithun Babu, Yash Oza, Arun Kumar Singh, K Madhava Krishna, and Shanti Medasani. 2018. Model predictive control for autonomous driving based on time scaled collision cone. In 2018 European Control Conference (ECC).
[31] Radhesh Bhat et al. 2019. Learning based demosaicing and color correction for RGB-IR patterned image sensors. Electronic Imaging (2019).
[32] Adith Boloor, Karthik Garimella, Xin He, Christopher Gill, Yevgeniy Vorobeychik, and Xuan Zhang. 2020. Attacking vision-based perception in end-to-end autonomous driving models. (2020), 101766.
[33] Carlos Campos, Richard Elvira, Juan J Gómez Rodríguez, José MM Montiel, and Juan D Tardós. 2020. ORB-SLAM3: An accurate open-source library for visual, visual-inertial and multi-map SLAM. arXiv preprint arXiv:2007.11898 (2020).
[34] Yulong Cao, Chaowei Xiao, Benjamin Cyr, Yimeng Zhou, Won Park, Sara Rampazzi, Qi Alfred Chen, Kevin Fu, and Z Morley Mao. 2019. Adversarial sensor attack on lidar-based perception in autonomous driving. In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security.
[35] Shang-Tse Chen, Cory Cornelius, Jason Martin, and Duen Horng Polo Chau. 2018. Shapeshifter: Robust physical adversarial attack on faster r-cnn object detector. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer.
[36] Valentin Deschaintre, Miika Aittala, Fredo Durand, George Drettakis, and Adrien Bousseau. 2018. Single-image svbrdf capture with a rendering-aware deep network. ACM Transactions on Graphics (ToG) (2018).
[37] Raj Gautam Dutta, Xiaolong Guo, Teng Zhang, Kevin Kwiat, Charles Kamhoua, Laurent Njilla, and Yier Jin. 2017. Estimation of safe sensor measurements of autonomous system under attack. In Proceedings of the 54th Annual Design Automation Conference 2017.
[38] Ivan Evtimov, Kevin Eykholt, Earlence Fernandes, Tadayoshi Kohno, Bo Li, Atul Prakash, Amir Rahmati, and Dawn Song. 2017. Robust physical-world attacks on machine learning models. arXiv preprint arXiv:1707.08945 (2017).
[39] Qingnan Fan, Jiaolong Yang, Gang Hua, Baoquan Chen, and David Wipf. 2017. A generic deep architecture for single image reflection removal and image smoothing. In Proceedings of the IEEE International Conference on Computer Vision. 3238–3247.
[40] Aidin Ferdowsi, Ursula Challita, Walid Saad, and Narayan B Mandayam. 2018. Robust deep reinforcement learning for security and safety in autonomous vehicle systems. In 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 307–312.
[41] Clément Fredembach and Sabine Süsstrunk. 2008. Colouring the near-infrared. In Color and Imaging Conference.
[42] Jianjun Hu, Songsong Xiong, Junlin Zha, and Chunyun Fu. 2020. Lane detection and trajectory tracking control of autonomous vehicle based on model predictive control. International journal of automotive technology (2020).
[43] Hyungsub Kim, Muslum Ozgur Ozmen, Antonio Bianchi, Z Berkay Celik, and Dongyan Xu. [n.d.]. PGFUZZ: Policy-Guided Fuzzing for Robotic Vehicles. ([n. d.]).
[44] Juncheng Li, Frank Schmidt, and Zico Kolter. 2019. Adversarial camera stickers: A physical camera-based attack on deep learning systems. In International Conference on Machine Learning.
[45] Chunchuan Lyu, Kaizhu Huang, and Hai-Ning Liang. 2015. A unified gradient regularization family for adversarial examples. In 2015 IEEE international conference on data mining. IEEE, 301–309.
[46] Francesco Marra, Diego Gragnaniello, and Luisa Verdoliva. 2018. On the vulnerability of deep learning to adversarial attacks for camera model identification. Signal Processing: Image Communication (2018).
[47] Mark J Mears. 2005. Cooperative electronic attack using unmanned air vehicles. In Proceedings of the 2005, American Control Conference, 2005.
[48] Noriyuki Miura, Tatsuya Machida, Kohei Matsuda, Makoto Nagata, Shoei Nashimoto, and Daisuke Suzuki. 2019. A Low-Cost Replica-Based Distance-Spoofing Attack on mmWave FMCW Radar. In Proceedings of the 3rd ACM Workshop on Attacks and Solutions in Hardware Security Workshop.
[49] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. 2016. Deepfool: a simple and accurate method to fool deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2574–2582.
[50] Nir Morgulis, Alexander Kreines, Shachar Mendelowitz, and Yuval Weisglass. 2019. Fooling a real car with adversarial traffic signs. arXiv preprint arXiv:1907.00374 (2019).
[51] Raul Mur-Artal and Juan D Tardós. 2017. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. IEEE Transactions on Robotics (2017).
[52] Ben Nassi, Dudi Nassi, Raz Ben-Netanel, Yisroel Mirsky, Oleg Drokin, and Yuval Elovici. 2020. Phantom of the ADAS: Phantom Attacks on Driver-Assistance Systems. IACR Cryptol. ePrint Arch. 2020 (2020).
[53] Sen Nie, Ling Liu, and Yuefeng Du. 2017. Free-fall: Hacking tesla from wireless to can bus. Briefing, Black Hat USA (2017).
[54] Grazyna Palczewska, Frans Vinberg, Patrycjusz Stremplewski, Martin P Bircher, David Salom, Katarzyna Komar, Jianye Zhang, Michele Cascella, Maciej Wojtkowski, Vladimir J Kefalov, et al. 2014. Human infrared vision is triggered by two-photon chromophore isomerization. Proceedings of the National Academy of Sciences.
[55] W Scott Pegau, Deric Gray, and J Ronald V Zaneveld. 1997. Absorption and attenuation of visible and near-infrared light in water: dependence on temperature and salinity. Applied optics (1997).
[56] Jonathan Petit and Steven E Shladover. 2014. Potential cyberattacks on automated vehicles. IEEE Transactions on Intelligent transportation systems (2014).
[57] J. Petit, Bas Stottelaar, and M. Feiri. 2015. Remote Attacks on Automated Vehicles Sensors : Experiments on Camera and LiDAR.
[58] Jonathan Petit, Bas Stottelaar, Michael Feiri, and Frank Kargl. 2015. Remote attacks on automated vehicles sensors: Experiments on camera and lidar. Black Hat Europe 11 (2015).
[59] Raul Quinonez, Jairo Giraldo, Luis Salazar, Erick Bauman, Alvaro Cardenas, and Zhiqiang Lin. 2020. {SAVIOR}: Securing Autonomous Vehicles with Robust Physical Invariants. In 29th {USENIX} Security Symposium ({USENIX} Security 20). 895–912.
[60] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition.
[61] Edward Rosten and Tom Drummond. 2006. Machine learning for high-speed corner detection. In European conference on computer vision. 430–443.
[62] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. 2011. ORB: An efficient alternative to SIFT or SURF. In 2011 International conference on computer vision. 2564–2571.
[63] Swami Sankaranarayanan, Arpit Jain, Rama Chellappa, and Ser Nam Lim. 2018. Regularizing deep networks using efficient layerwise adversarial training. In Proceedings of the AAAI Conference on Artificial Intelligence.
[64] Uri Shaham, Yutaro Yamada, and Sahand Negahban. 2015. Understanding adversarial training: Increasing local stability of neural nets through robust optimization. arXiv preprint arXiv:1511.05432 (2015).
[65] Prinkle Sharma, David Austin, and Hong Liu. 2019. Attacks on Machine Learning: Adversarial Examples in Connected and Autonomous Vehicles. In 2019 IEEE International Symposium on Technologies for Homeland Security (HST).
[66] Junjie Shen, Jun Yeon Won, Zeyuan Chen, and Qi Alfred Chen. 2020. Drift with Devil: Security of Multi-Sensor Fusion based Localization in High-Level Autonomous Driving under {GPS} Spoofing. In 29th {USENIX} Security

Symposium ({USENIX} Security 20). 931–948.
[67] Hocheol Shin, Dohyun Kim, Yujin Kwon, and Yongdae Kim. 2017. Illusion and dazzle: Adversarial optical channel exploits against lidars for automotive applications. In International Conference on Cryptographic Hardware and Embedded Systems. 445–467.
[68] Yasser Shoukry, Paul Martin, Paulo Tabuada, and Mani Srivastava. 2013. Non-invasive spoofing attacks for anti-lock braking systems. In International Workshop on Cryptographic Hardware and Embedded Systems.
[69] Yasser Shoukry, Paul Martin, Yair Yona, Suhas Diggavi, and Mani Srivastava. 2015. Pycra: Physical challenge-response authentication for active sensors under spoofing attacks. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. 1004–1015.
[70] Chawin Sitawarin, Arjun Nitin Bhagoji, Arsalan Mosenia, Mung Chiang, and Prateek Mittal. 2018. Darts: Deceiving autonomous cars with toxic signs. arXiv preprint arXiv:1802.06430.
[71] Huixuan Tang, Xiaopeng Zhang, Shaojie Zhuo, Feng Chen, Kiriakos N Kutulakos, and Liang Shen. 2015. High resolution photography with an RGB-infrared camera. In 2015 IEEE International Conference on Computational Photography (ICCP).
[72] Emma Thilén. 2017. Robust model predictive control for autonomous driving.
[73] Vrizlynn LL Thing and Jiaxi Wu. 2016. Autonomous vehicle security: A taxonomy of attacks and defenses. In 2016 ieee international conference on internet of things (ithings) and ieee green computing and communications (greencom) and ieee cyber, physical and social computing (cpscom) and ieee smart data (smartdata). IEEE.
[74] Qinglong Wang, Wenbo Guo, Alexander G Ororbia II, Xinyu Xing, Lin Lin, C Lee Giles, Xue Liu, Peng Liu, and Gang Xiong. 2016. Using non-invertible data transformations to build adversarial-robust neural networks. arXiv preprint arXiv:1610.01934 (2016).
[75] Shu Wang, Jiahao Cao, Kun Sun, and Qi Li. 2020. {SIEVE}: Secure In-Vehicle Automatic Speech Recognition Systems. In 23rd International Symposium on Research in Attacks, Intrusions and Defenses ({RAID} 2020). 365–379.
[76] Weilin Xu, David Evans, and Yanjun Qi. 2017. Feature squeezing: Detecting adversarial examples in deep neural networks. arXiv preprint arXiv:1704.01155 (2017).
[77] Chen Yan, Wenyuan Xu, and Jianhao Liu. 2016. Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicle. DEF CON 24 (2016).