Spoofing Mobileye 630's Video Camera Using a Projector

Ben Nassi, Dudi Nassi, Raz Ben-Netanel, Yuval Elovici
The Department of Software and Information Systems Engineering, Ben-Gurion University of the Negev
nassib@post.bgu.ac.il, nassid@post.bgu.ac.il, razx@post.bgu.ac.il, elovici@bgu.ac.il
Video Demonstration - https://youtu.be/C-JxNHKqgtk

***Abstract*—In this paper, we evaluate the robustness of Mobileye 630 PRO, the most popular off-the-shelf ADAS on the market today, to camera spoofing attacks applied using a projector. We show that Mobileye 630 issues false notifications about road signs projected in proximity to the car that the system is installed in. We assess how changes of the road signs (e.g., changes in color, shape, projection speed, diameter and ambient light) affect the outcome of an attack. We find that while Mobileye 630 PRO rejects fake projected road signs that consists of non-original shapes and objects, it accepts fake projected road signs that consists of non-original colors. We demonstrate how attackers can leverage these findings to apply a remote attack in a realistic scenario by using a drone that carries a portable projector which projects the spoofed traffic sign on a building located in proximity to a passing car equipped with Mobileye 630. Our experiments show that it is possible to fool Mobileye 630 PRO to issue false notification about a traffic sign projected from a drone.**

I. Introduction

Advanced driver assistance systems (ADASs) [1] are electronic systems that aid automobile drivers while they are driving. Such systems aim to help drivers by: 1) issuing alerts (e.g., collision avoidance) regarding potential threats and 2) recognizing upcoming traffic signs. ADASs have already become an integral part of the current car generation, and they will provide automated functionalities for the next generation of cars (autonomous vehicles).

Security of ADAS have recently attracted attention of researchers that has begun to investigate the their robustness to various attacks. Recent studies [2], [3], [4], [5] showed that ADAS alerts and notifications can be spoofed by applying adversarial machine learning techniques to traffic signs, allowing attackers to control the output of the ADAS for their benefit. The application of the methods suggested in these studies exposes drivers that respond to ADAS alerts and notifications and other nearby drivers and pedestrians to variety of risks that can cause accidents. However, the suggested attacks are complicated and require deep understanding of the ADAS used in order to manipulate it. In addition, the nature of the attacks necessitated that the attacker be located near the traffic sign in order to perform the attack. Because of this, the attacks suggested in the recent studies have not appeared in the wild. We wonder whether a complicated attack is really needed to manipulate an ADAS?

In this paper, we evaluate the robustness of Mobileye 630 PRO, the most popular off-the-shelf ADAS on the market today, to camera spoofing attacks applied using a projector. We show that attackers can use a projector in order to inject traffic signs into Mobileye 630 causing this system to issue false warnings. We performed various experiments and assessed the influence of color, shape, projection speed, diameter, and ambient light on the outcome of the attack by mounting a projector onto a drone and injecting traffic signs into the ADAS of a real driving car.

We make the following contributions: First, unlike other studies in this area that trained a classifier and found vulnerabilities to attack the ADAS, we evaluated the practicality of our attack against Mobileye, a real off-the-shelf ADAS. Second, our vector attack doesn't require the attacker to be in the attack location; we demonstrate a remote attack that can be executed by a drone.

II. Related Work

In this section, we describe related work on attacks against ADASs and provide an overview of adversarial attacks. Computer vision object detectors are integrated to ADAS and used to detect traffic signs from a video stream. Many of these detectors are trained using deep learning techniques. Several studies created adversarial instances to trick such deep learning classifiers and showed that this type of classifier is vulnerable to spoofing attacks. *Petrakieva et al.* [2] demonstrated how perturbations that are often too small to be perceptible to humans can fool deep learning models. *Sitawarin et al.* [6] showed that they could embed two traffic signs in one traffic sign with a dedicated array of lens that causes a different traffic sign to appear depending on the angle of view. *Eykholt et al.* [7] and *Lu et al.* [3] showed that physical artifacts (e.g., stickers, graffiti) misled computer vision classifiers. In the abovementioned studies, the researchers only trained dedicated models by themselves and identified instances that could exploit them using white-box techniques. Furthermore, the researchers did not show the effectiveness of the attack against an off-the-shelf ADAS. In contrast, we demonstrate our attack against

Workshop on Automotive and Autonomous Vehicle Security (AutoSec) 2021
25 February 2021, Virtual
ISBN 1-891562-68-1
https://dx.doi.org/10.14722/autosec.2021.23007
www.ndss-symposium.org

the Mobileye system and mislead it so it recognizes spoofed traffic signs using black-box techniques.

Attacks against ADAS are not, however, limited to misleading the classifier using an adversarial traffic sign. *Petit et al.* [4] presented two attack vectors against car's sensors such as LiDAR and cameras. They were able to show that: 1) a laser directed at the camera can destroy the optical sensor permanently, and 2) LiDAR's output can be spoofed using infrared light. *Yan et al.* [8] demonstrated various spoofing attacks against a camera, ultrasonic sensor, and radar that can cause Tesla's Model S to misperceive the distance of nearby obstacles. However, it is not possible to perform the suggested attacks [4], [8] on a driving car due to the complexity of the attacks because: 1) they require deploying devices at specific ranges from the attacked car, and 2) the attacker must connect the hardware directly to a driving autonomous car which can be a major challenge due to the driving speed. A recent study [5] misdirected an autopiloted vehicle, taking it in the wrong direction. The authors placed interference patches (small stickers) on the ground at two way route, causing the vehicle to turn in to the opposite lane. In this case, the attacker must physically put the stickers on the road; in contrast, our attack vector doesn't require the attacker to be on site, since the drone can be deployed remotely.

Other famous attacks against cars that are not related to our work are [9], [10], which were based on compromising the firmware of the car or an internal device. Our attack model is much lighter than these attacks, since it does not require us to hack to those systems.

III. Threat Model

We consider an attacker as any malicious entity with the aim of attacking a driving car with equipped with Mobileye 630 PRO. The attacker can inject spoofed traffic signs into Mobileye using a portable projector mounted on a drone. The attacker's goals can be to: 1) harm or manipulate the car of a specific victim, or 2) cause environmental chaos (e.g., harm multiple cars in a specific region such as a city, neighborhood, highway, etc.).

IV. Mobileye Analysis

Here we analyze the robustness of Mobileye 630 PRO against various projected road signs. Mobileye is an external ADAS that provides function-specific vehicle automation (Level 0). The Mobileye 630 PRO contains two main components. The first is a camera, which is installed on the windshield, under the rear view mirror, and the second is a small display which is placed in front of the driver and provides visual and audible alerts about the surroundings as needed. Mobileye has the following features: lane departure warning, pedestrian collision warning, forward collision warning, headway monitoring and warning, intelligent high beam control, and traffic sign recognition. In this section we focus on testing the robustness of its traffic sign recognition.

In the following subsections, we learn the effect of environmental factors (ambient light, distance) on the result of the attack. In addition, we test the robustness of the Mobileye for classifying traffic signs that do not exist.

A. Experimental Setup

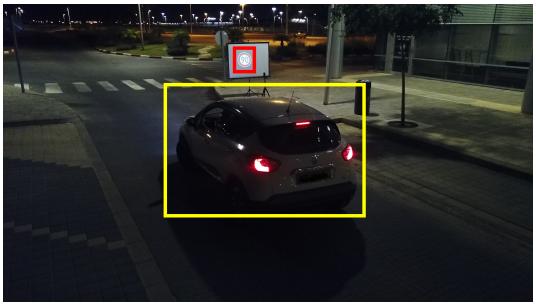


Fig. 1. Experimental setup: the projected sign is boxed in red, and the attacked vehicle is boxed in yellow

In this subsection, we describe the setup for the experiments performed. For convenience we used a white projector screen, in order as screen for the projected traffic sign. A portable projector was used to provide the sign's content. The injection method, as described in Section III, is comprised of the projector and screen. The portable projector was placed on a tripod about 2.5 meters from the screen and projected a traffic sign onto the center of the screen; while the sign was projected in this way we drove the car (a Renault Captur equipped with Mobileye 630 PRO) in a straight line at a speed of approximately 25 km/h. Figure 1 presents an illustration of our experimental setup.

B. Influence of the Projected Sign's Diameter

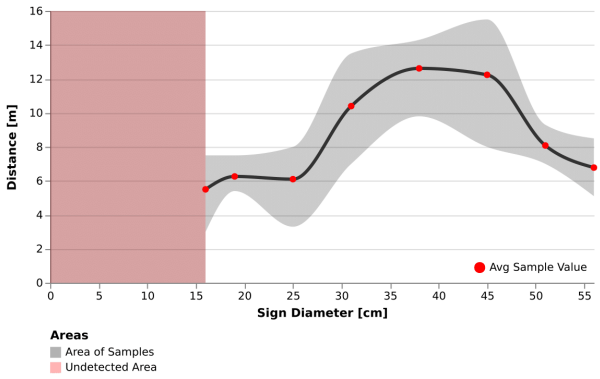


Fig. 2. Influence of the sign's diameter

1) Experimental Setup: In this case, we investigate whether the size of the projected sign influences the distance from which the Mobileye 630 PRO's sensor can detect the projected sign. We repeated the experiment five times, projecting a different sized sign each time, and calculated the average detection distance.

2) Results: Figure 2 presents the results of this experiment. As can be seen, if the sign is too small (less than 16 cm in diameter) the Mobileye 630 Pro sensor didn't detect it at all. The red dots in the graph symbolize the average distance at which we managed to mislead the sensor, and the grey area shows the range of the entire samples set.

3) Conclusion: The diameter range is wide and provides a lot of room for error when projecting a traffic sign. Based on our measurements, the distance can range from approximately 5-16 meters.

C. Influence of the Color of the Projected Sign

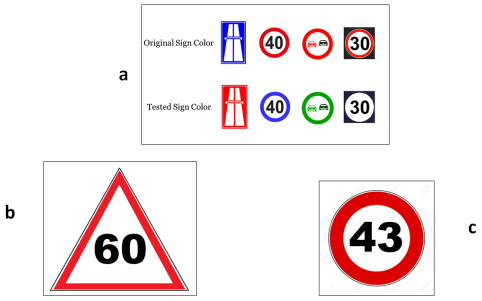


Fig. 3. (a) examples of different colored traffic signs, (b) an example of a different traffic sign shape, (c) an example of an incorrect or unrecognized speed limit value

1) Experimental Setup: Here we assess whether the Mobileye 630 PRO sensor is sensitive to the color of the sign. We tried various colors as seen in (Figure 3a). First we projected the sign with its true colors, and then we verified that the Mobileye 630 PRO sensor managed to recognize the sign. Next, we projected the same sign but this time with a color scheme which is different from the real one.

2) Results: We could see, quite quickly, that Mobileye is not sensitive to color, since all of the signs tested managed to mislead the sensor (even the black and white speed limit sign seen in, Figure 3a).

3) Conclusion: Based on these results we conclude that the Mobileye 630 PRO sensor only considers the shape of a sign when trying to classify the sign's content.

D. Influence of the Projected Sign's Shape

1) Experimental Setup: In this case, we evaluate whether the Mobileye 630 PRO sensor can detect signs which do not take the form of their original shape. For this experiment we simply took a speed limit sign and modified its shape (as seen in Figure 4b). This experiment was binary, i.e., we only wanted to know if the sign can be detected or not.

2) Results: We utilized a total of seven different shapes (a triangle, rectangle, pentagon, and hexagon, as well as three other more unusual shapes - a star, arrow, and random polygon). The Mobileye 630 PRO's sensor was unable to detect any of these shapes.

3) Conclusion: We can conclude the Mobileye system considers just the shape of the sign and isn't fooled by unknown shapes.

4) Experimental Setup: In this case, we tested the effect of ambient light, utilizing 20 samples (drives) from every hour of the day to check our injection success rate.

5) Results: Figure 4 presents the results of this experiment; success is considered a sample (drive) in which Mobileye recognized the projected sign.

6) Conclusion: Based on our analysis of these results we can conclude that is possible to inject a spoofed traffic sign at all hours of the day, but performance is best later in the day (in the evening and at night). One thing that should be considered with regard to ambient light is the equipment used, since the opacity of the projected sign depends on the ambient light as well as the projector used (a better success rate may be achieved with a better projector).

E. Influence of the Speed of the Projection Time

1) Experimental Setup: Here we assessed the speed of the projection time that is needed to fool the system. We conducted a few experiments that measured the amount of time required for injection.

2) Results: We discovered that a projection speed of 100 ms is sufficient for fooling the system. We were unable to fool the system with faster projection speeds probably due to the frame per second rate of the optical sensor of the Mobileye.

3) Conclusion: The fast projection speed causes the attack vector on the target to be very easy to inject and doesn't require staying for to long.

F. Influence of the Number on the Projected Sign

1) Experimental Setup: In this case, we investigate whether the Mobileye 630 PRO sensor can also detect speed limit signs with speed values that are not used in the real world (e.g. Figure 3.c).

2) Results: Table 1 presents the results of this experiment.

3) Conclusion: Based on these results, we can conclude that incorrect speed limit signs are effective at misleading the system. The system do not ignore them and classify them as other similar traffic signs.

V. Attacking a Car While Driving

In this section, we demonstrate how attackers can spoof Mobileye 630 PRO's video camera remotely using drone.

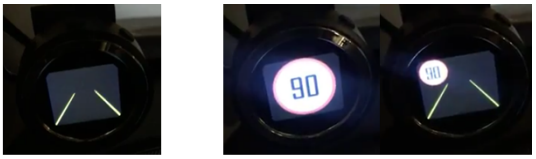
1) Experimental setup: We mounted a portable projector on a drone (DJI Matrice 600) (Figure 4). In this experimental setup our car (a Renault Captur equipped with Mobileye 630 PRO) was driven in an urban environment as the attacker operated a drone, positioning the drone so the spoofed speed limit sign can be injected into the Mobileye sensor. The attacker projected the incorrect speed limit sign (see Figure 4), managing to mislead the Mobileye sensor which recognized the sign as a 90 km/h speed limit sign (see Figure 5). The implemented attack vector can be seen in an uploaded video of the attack [1].

[1] https://youtu.be/C-JxNHKqgtk

Fig. 4. Left: Influence of Ambient Light. Middle: the drone with the projector used in our experiments, Right: the moment of the attack (the projected road sign is boxed in blue, the attacker's drone is boxed in purple, and the victim's car is boxed in red).

TABLE I
DETECTION OF INCORRECT SPEED LIMIT SIGNS,(LEFT: SPEED LIMIT ON THE PROJECTED SIGN. RIGHT: THE DETECTED SPEED LIMIT, AS SHOWN ON THE MOBILEYE DISPLAY, X MEANS NO DETECTION)

Projected Speed limit	Detected Speed limit
0	X
1	X
2	X
3	X
4	X
5	5
6	X
7	X
8	X
9	X
27	X
43	X
69	60
71	70
88	80
150	X
160	X
170	X
180	110
190	110
200	X



Before the attack After the attack

Fig. 5. Mobileye display before and during the attack.

2) Results: We managed to fool Mobileye so it classified the speed limit as 90 km/h when the speed limit for a city road is only 30 km/h.

VI. FUTURE WORK

As future work, we suggest to test whether camera spoofing can be applied by embedding a traffic sign to an advertisement presented on a digital billboard in an invisible manner (e.g., by flashing the traffic sign for a split second) [11]. We also suggest examining whether the attack can be applied using infrared projection, exploiting the fact that a narrow spectrum of frequencies, the near infrared, is also captured by some CMOS sensors (this fact was exploited to establish an optical covert channel [12] and to break the confidentiality of FPV channel of commercial drones [13], [14]). We also suggest to test whether attackers can project a fake lane in order to fool Mobileye 630's lane detection functionality [15].

REFERENCES

[1] Wikipedia contributors, "Advanced driver-assistance systems — Wikipedia, the free encyclopedia," 2019, [Online; accessed 27-May-2019]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Advanced_driver-assistance_systems&oldid=898181591

[2] N. Akhtar and A. Mian, "Threat of adversarial attacks on deep learning in computer vision: A survey," *IEEE Access*, vol. 6, pp. 14 410–14 430, 2018.

[3] J. Lu, H. Sibai, E. Fabry, and D. Forsyth, "Standard detectors aren't (currently) fooled by physical adversarial stop signs," *arXiv preprint arXiv:1710.03337*, 2017.

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

[5] T. K. S. Lab, "Experimental security research of tesla autopilot," https://keenlab.tencent.com/en/whitepapers/Experimental_Security_Research_of_Tesla_Autopilot.pdf, 2019.

[6] C. Sitawarin, A. N. Bhagoji, A. Mosenia, M. Chiang, and P. Mittal, "Darts: Deceiving autonomous cars with toxic signs," *arXiv preprint arXiv:1802.06430*, 2018.

[7] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno, and D. Song, "Robust physical-world attacks on deep learning models," *arXiv preprint arXiv:1707.08945*, 2017.

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

[9] A. Greenberg, "Hackers remotely kill a jeep on the highway-with me in it (2015)," *URL https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway*, 2017.

[10] S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, S. Savage, K. Koscher, A. Czeskis, F. Roesner, T. Kohno *et al.*, "Comprehensive experimental analyses of automotive attack surfaces." in *USENIX Security Symposium*, vol. 4. San Francisco, 2011, pp. 447–462.

[11] B. Nassi, Y. Mirsky, D. Nassi, R. Ben-Netanel, O. Drokin, and Y. Elovici, "Phantom of the adas: Securing advanced driver-assistance systems from split-second phantom attacks," in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 293âĂŞ308. [Online]. Available: https://doi.org/10.1145/3372297.3423359

[12] B. Nassi, A. Shamir, and Y. Elovici, "Xerox day vulnerability," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 2, pp. 415–430, 2018.

[13] B. Nassi, R. Ben-Netanel, A. Shamir, and Y. Elovici, "Drones' cryptanalysis-smashing cryptography with a flicker," in *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2019, pp. 1397–1414.

[14] R. Ben Netanel, B. Nassi, A. Shamir, and Y. Elovici, "Detecting spying drones," *IEEE Security Privacy*, vol. 19, no. 1, pp. 65–73, 2021.

[15] B. Nassi, D. Nassi, R. Ben-Netanel, Y. Mirsky, O. Drokin, and Y. Elovici, "Phantom of the adas: Phantom attacks on driver-assistance systems."