# Demo: Security of Multi-Sensor Fusion based Perception in AD under Physical-World Attacks

Yulong Cao[⋆,§] Ningfei Wang[⋆,†] Chaowei Xiao[⋆,‡‡] Dawei Yang[⋆,§] Jin Fang[‡] Ruigang Yang[††]
Qi Alfred Chen[†] Mingyan Liu[§] Bo Li[¶]
[†]University of California, Irvine [§]University of Michigan [‡‡]Arizona State University [††]Inceptio
[‡]Baidu Research, China [¶]University of Illinois at Urbana-Champaign [⋆]Co-first authors

***Abstract*—In autonomous driving (AD) vehicles, Multi-Sensor Fusion (MSF) is used to combine perception results from multiple sensors such as LiDARs (Light Detection And Ranging) and cameras for both accuracy and robustness. In this work, we design the first attack that fundamentally defeats MSF-based AD perception by generating 3D adversarial objects. This demonstration will include video and figure demonstrations for the generated 3D adversarial objects and the end-to-end consequences.**

## I. Introduction

Autonomous Driving (AD) vehicles are already providing services on public roads. They have adopted machine learning (ML) models and achieved promising performance. Recent studies show that ML models for AD perception are vulnerable to adversarial attacks. However, they focus on attacking models for individual sensor. Multi-Sensor Fusion (MSF) mechanisms have been applied and shown to help improve model robustness and accuracy, which thus have the potential to correct the attack effects from any individual sensors. Thus, in current stage, it is unclear whether adversarial attacks for MSF-based AD perception can be generated or not.

In this demo, we show our attack *MSF-ADV*, the first attack that can fundamentally defeat MSF-based AD perception by generating physical-world adversarial 3D objects that simultaneously fool both LiDAR- and camera-based perception. We will show the generated adversarial 3D objects, some physical-world attack demos, and end-to-end attack attack impact demo.

## II. Threat model and attack goal

**Threat model.** We assume that attacker has the full knowledge of the LiDAR- and camera-based AD perception in the victim AD system. We also assume that attacker can place an adversarial 3D object on the road, and can collect the required sensor data in the target road beforehand to facilitate the attack.

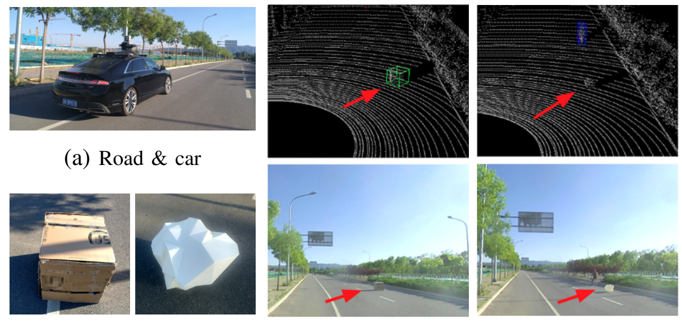
**Attack goal.** The attack goal is to cause the victim AD vehicles to fail in detecting the object and thus collide into it. This threatens the safety of the passengers in AD vehicles.

## III. Attack design

For LiDARs and cameras, DNN-based perception has the state-of-the-art performance and thus is used widely in practice today. In *MSF-ADV*, we combine the designs below into a single optimization problem to simultaneously attack two models by changing the shape of a 3D object.

**Attack design for LiDAR-based perception.** First, we use a differentiable ray-casting renderer to project shape changes of the 3D object to point cloud. Then, we design differentiable approximation functions to approximate non-differentiable preprocessing steps. After that, we design an objective function to reduce the detection confidence of the adversarial object.

Workshop on Automotive and Autonomous Vehicle Security (AutoSec) 2022
24 April 2022, San Diego, CA, USA
ISBN 1-891562-75-4
https://dx.doi.org/10.14722/autosec.2022.23038
www.ndss-symposium.org



(a) Road & car (b) Benign and adv. cubes (c) Benign case (d) Adversarial case

Fig. 1: Physical-world experiment with real vehicle setup. The adversarial cube is 3D-printed at 1:1 scale.

**Attack design for camera-based perception.** Similar to design for LiDARs, we use a physics-based renderer to project the shape changes to images, and model the pre-processing steps. We use objective function from previous work.

## IV. Demonstration plan

**Demonstration of the generated adversarial 3D object.** We will show benign objects and our generated adversarial 3D objects in miniature scale and their corresponding images.

**Demonstration of physical-world attacks.** We will demonstrate videos [1] of the physical-world attack on Baidu Apollo [2] under both miniature scale and real vehicle setup. Fig. 1 shows the physical-world experiment with real vehicle setup. We use a benign box and 3D print the adversarial one, which looks like a strange-looking rock. As the results, the benign cube can be detected, while the adversarial one cannot.

**Demonstration of end-to-end attack impact.** We will demonstrate videos [1] of the end-to-end attack impact on AD by launching attack to a Baidu Apollo [2] AD vehicle running in LGSVL [3], an production-grade AD simulator. Our demo will show that the AD vehicle fully stops before the benign traffic cone while it crashes into the adversarial traffic cone.

## Acknowldgments

This research was supported in part by the NSF under CNS-1850533, CNS-1932464, CNS-1929771, and CNS-2145493.

## References

[1] "Demo Videos," https://sites.google.com/view/msf-adv-ndss.
[2] "Baidu Apollo," https://apollo.auto/.
[3] "LGSVL Simulator," https://www.lgsvlsimulator.com/.