Poster: Invisible for both Camera and LiDAR: Security of Multi-Sensor Fusion based Perception in Autonomous Driving 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, {ningfei.wang, alfchen}@uci.edu
[§]University of Michigan, {yulongc, ydawei, mingyan}@umich.edu
[‖]NVIDIA Research [‡‡]Arizona State University [††]Inceptio
[‡]Baidu Research and National Engineering Laboratory of Deep Learning Technology and Application, China
[¶]University of Illinois at Urbana-Champaign, lbo@illinois.edu

Abstract

In Autonomous Driving (AD) systems, perception is both security and safety critical. Despite various prior studies on its security issues, *all* of them only consider attacks on camera- or LiDAR-based AD perception *alone*. However, production AD systems today predominantly adopt a Multi-Sensor Fusion (MSF) based design, which in principle can be more robust against these attacks under the assumption that not all fusion sources are (or can be) attacked at the same time. In this paper, we present the first study of security issues of MSF-based perception in AD systems. We directly challenge the basic MSF design assumption above by exploring the possibility of attacking *all* fusion sources simultaneously. This allows us for the first time to understand how much security guarantee MSF can fundamentally provide as a general defense strategy for AD perception.

We formulate the attack as an optimization problem to generate a physically-realizable, adversarial 3D-printed object that misleads an AD system to fail in detecting it and thus crash into it. To systematically generate such a physical-world attack, we propose a novel attack pipeline that addresses two main design challenges: (1) non-differentiable target camera and LiDAR sensing systems, and (2) non-differentiable cell-level aggregated features popularly used in LiDAR-based AD perception. We evaluate our attack on MSF algorithms included in representative open-source industry-grade AD systems in real-world driving scenarios. Our results show that the attack achieves over 90% success rate across different object types and MSF algorithms. Our attack is also found stealthy, robust to victim positions, transferable across MSF algorithms, and physical-world realizable after being 3D-printed and captured by LiDAR and camera devices. To concretely assess the end-to-end safety impact, we further perform simulation evaluation and show that it can cause a 100% vehicle collision rate for an industry-grade AD system. We also evaluate and discuss defense strategies.

I. Main content

This research [1] is recently published in IEEE S&P 2021 with DOI Bookmark: 10.1109/SP40001.2021.00076. The original abstract and author list are shown above. We post the paper links with conference version[1] and arXiv version[2].

References

[1] Y. Cao, N. Wang, C. Xiao, D. Yang, J. Fang, R. Yang, Q. A. Chen, M. Liu, and B. Li, "Invisible for both Camera and LiDAR: Security of Multi-Sensor Fusion based Perception in Autonomous Driving Under Physical World Attacks," in *Proceedings of the 42nd IEEE Symposium on Security and Privacy (IEEE S&P 2021)*, May 2021.

[1] https://www.computer.org/csdl/proceedings-article/sp/2021/893400b302/1t0x9btzenu
[2] https://arxiv.org/abs/2106.09249

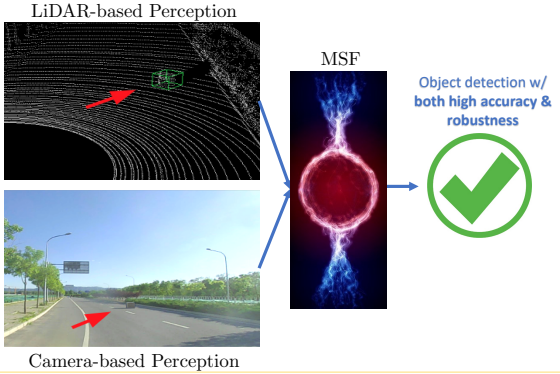
Poster: Invisible for both Camera and LiDAR: Security of Multi-Sensor Fusion based Perception in Autonomous Driving Under Physical-World Attacks

Published in IEEE S&P 2021

AS²Guard UCI U-M ASU Yulong Cao*, Ningfei Wang*, Chaowei Xiao*, Dawei Yang*, Jin Fang, Ruigang Yang, Qi Alfred Chen, Mingyan Liu, Bo Li (* Co-first authors)

Multi Sensor Fusion (MSF) based Perception in Autonomous Driving (AD)

- Prior works only consider attacking AD perception on **single sensor (e.g., LiDAR or camera)**
- Production high-level AD systems adopt **MSF-based perception**
  - To achieve higher accuracy and robustness
- Can improve security ***if not all perception sources are (or can be) attacked simultaneously***
  - If hold, theoretically always possible to rely on the unattacked source(s) to detect/prevent such attack
  - **Believed to hold in general**, thus widely recognized as a general defense strategy against existing attacks on AD perception



Research Question

- Can such basic security design assumption actually be broken, especially in practical AD settings?

Our Work

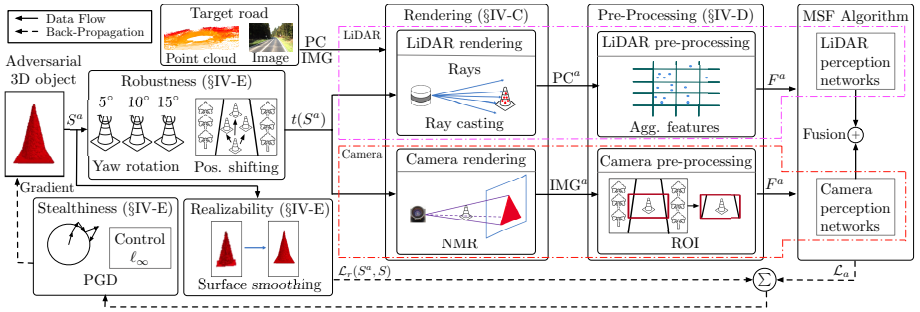
- First study on security of MSF-based AD perception
  - Challenging the basic security design assumption in practical AD settings
- Physically-realizable & stealthy attack vector: adversarial 3D object
- Design a novel attack method, **MSF-ADV**
  - Generate adversarial 3D objects that can **simultaneously** fool ***all*** perception sources used in MSF-based AD perception

Attack Goal

- Fool MSF-based AD perception in victim AD vehicles to fail in detecting a front obstacle & thus crash into it
  - Cause severe crash by filling dense materials (e.g., granite or metal)
  - Leverage sematic meaning of a certain road object (e.g., traffic cone)

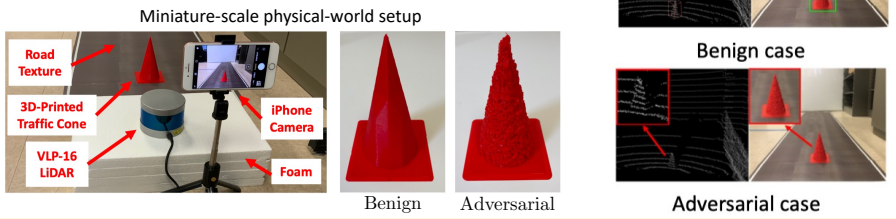
Our Approach: MSF-ADV

- Generate adversarial 3D object
  - For LiDAR, we generate malicious point cloud by simulating the physics of a LiDAR by **ray casting** and **differentiably** rendering synthetic object into the point cloud
    - Design **differentiable approximation functions** to approximate the non-differentiable pre-processing steps (e.g., point inclusion)
  - For camera, we obtain malicious image by calibrating the object position with LiDAR point cloud and **differentiably** rendering it in the middle of the road using NMR



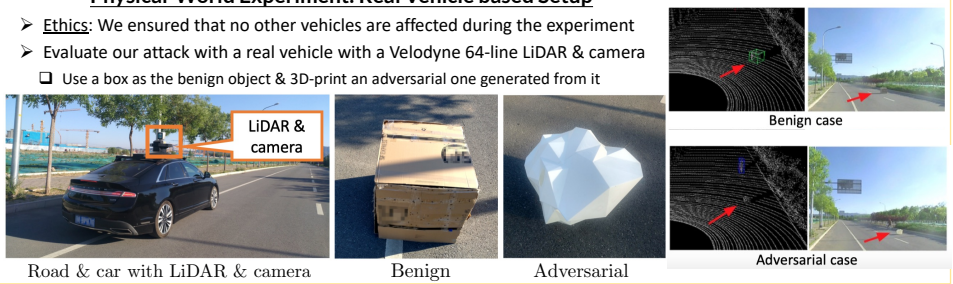
Physical-World Experiment: Miniature-Scale Setup

- Evaluate our attack in a miniature-scale physical-world setup with real camera, LiDAR, and 3D printed benign and adversarial traffic cones



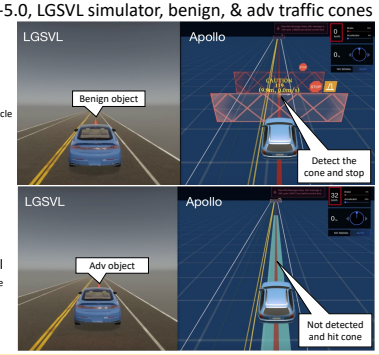
Physical-World Experiment: Real Vehicle based Setup

- Ethics: We ensured that no other vehicles are affected during the experiment
- Evaluate our attack with a real vehicle with a Velodyne 64-line LiDAR & camera
  - Use a box as the benign object & 3D-print an adversarial one generated from it



End-to-End Attack Simulation Evaluation

- Apollo-5.0, LGSVL simulator, benign, & adv traffic cones



Evaluation Highlight

- Setup: 4 MSF included in open-source full-stack AD systems, Apollo (industry-grade) & Autoware.AI
  - 3 object types & 100 scenarios from KITTI dataset
- Effectiveness: >=91% success rate
- Robustness: >95% average success rate
- Transferability: 75% success rate over different MSF
- Physical-world realizability: >=85% success rate
- End-to-end attack simulation
  - 100% collision rate across 100 runs

Defenses Experiments & Discussions

- DNN-level defense
  - Experimented against 6 existing defenses
  - Most effective one reduced attack success rate to 66% w/o harming benign performance
    - Not quite enough to render our attack practically unexploitable
- Fuse more perception sources
  - More cameras/LiDARs mounted at different positions or including RADAR
    - Cannot fundamentally defeat our attack, but may make it more difficult to generate

Responsible Vulnerability Disclosure

- As of 01/14/2022, informed 31 companies
  - 18 (~58%) has replied so far & have started investigation

Take a picture for more **details & related materials**

Contact: ningfei.wang@uci.edu