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

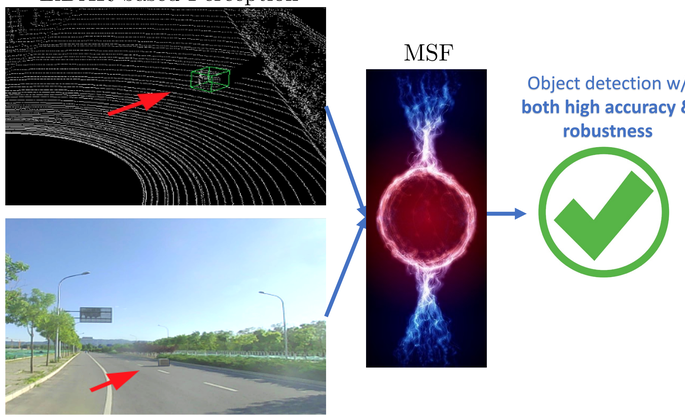
AS²Guard Autonomous & Smart Systems Guard Research Group · UCI · Michigan · ASU

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

Illinois · NVIDIA · Baidu 百度 · Inceptio Technology

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

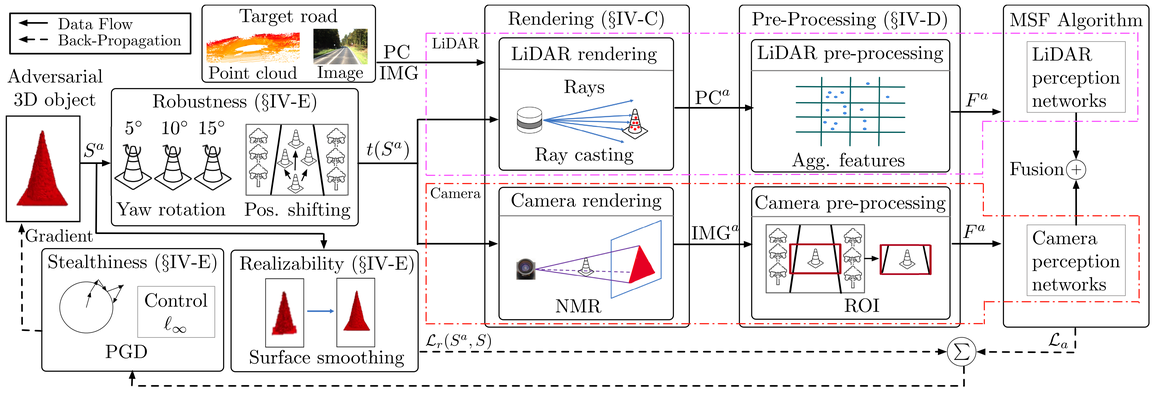
- First study on security of MSF-based AD perception
  - Challenging the basic security design assumption in practical AD settings
- Physically-realizable & stealthy attack vector: adversarial 3D object
- Design a novel attack method, **MSF-ADV**
  - Generate adversarial 3D objects that can **simultaneously**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

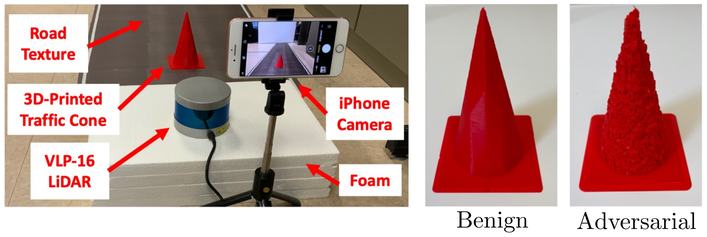
- Generate adversarial 3D object
  - For LiDAR, we generate malicious point cloud by simulating the physics of a LiDAR by **ray casting** and **differentiably** rendering synthetic object into the point cloud
    - Design **differentiable approximation functions** to approximate the non-differentiable 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

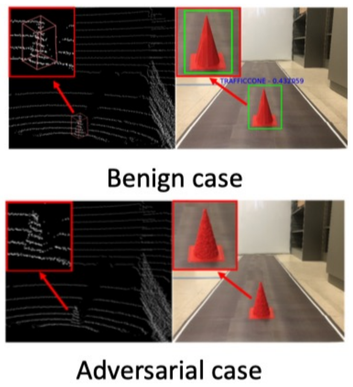


Physical-World Experiment: Miniature-Scale Setup

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

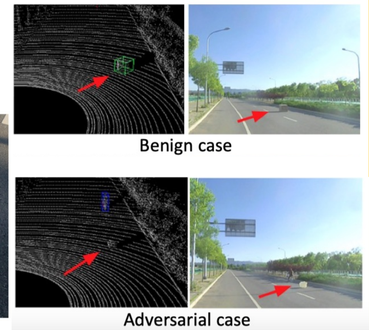
Miniature-scale physical-world setup





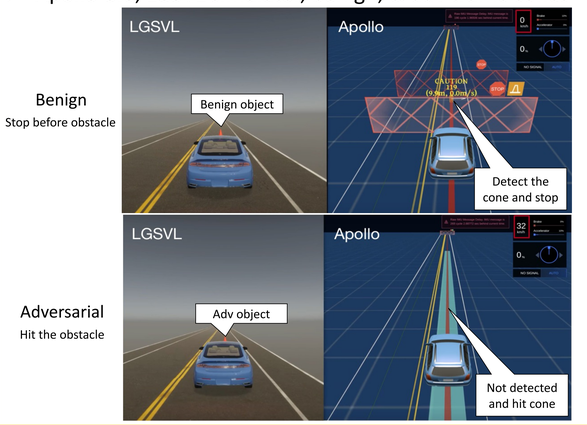
Physical-World Experiment: Real Vehicle based Setup

- 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 4/25/2021, informed 31 companies
  - 17 (~55%) has replied so far & have started investigation

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

Contact: ningfei.wang@uci.edu