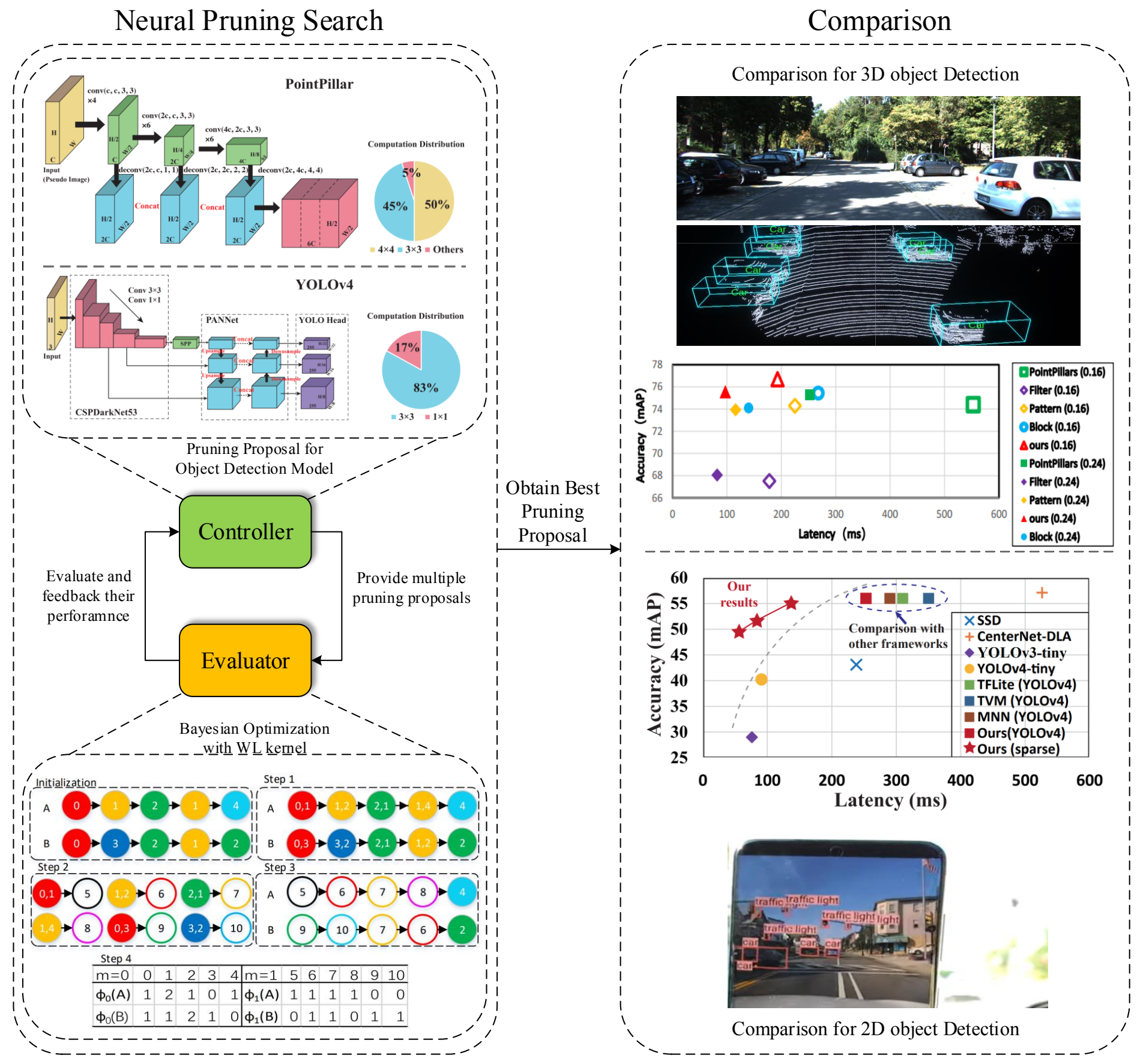
SecureNN: Design of Secured Autonomous Cyber-Physical Systems Against Adversarial Machine Learning Attacks

PI: Xue (Shelley) Lin, Northeastern University; PI: Qi (Alfred) Chen, University of California, Irvine
https://sites.google.com/view/securenn

Abstract: This project aims to enhance the security of autonomous cyber-physical systems such as self-driving cars by investigating new attack surfaces exposed through the multi-modal sensing, and deep learning-based perception and control systems. **Broader Impacts**: The project will advance technologies for autonomous systems, artificial intelligence, and security analysis and defense. The developed techniques will be able to applied to other application domains involving deep learning and ubiquitous sensing. The project provides training opportunities for graduate/undergraduate/K12 students, releases our findings to the community, and has industry outreach activities.

Towards Real-Time 2D/3D Object Detection for Autonomous Vehicles

- We propose neural pruning search to facilitate 2D/3D object detection in autonomous vehicles.
- The method automatically searches a best-suiting pruning scheme and pruning ratio.
- It also combines compiler optimization techniques to achieve accelerated execution.
- We focus on 2D detection for camera images and 3D detection for point clouds from LiDARs.
- For the first time, (close-to) real-time detection is achieved i.e., 55ms and 97ms inference time for 2D and 3D detection, respectively.

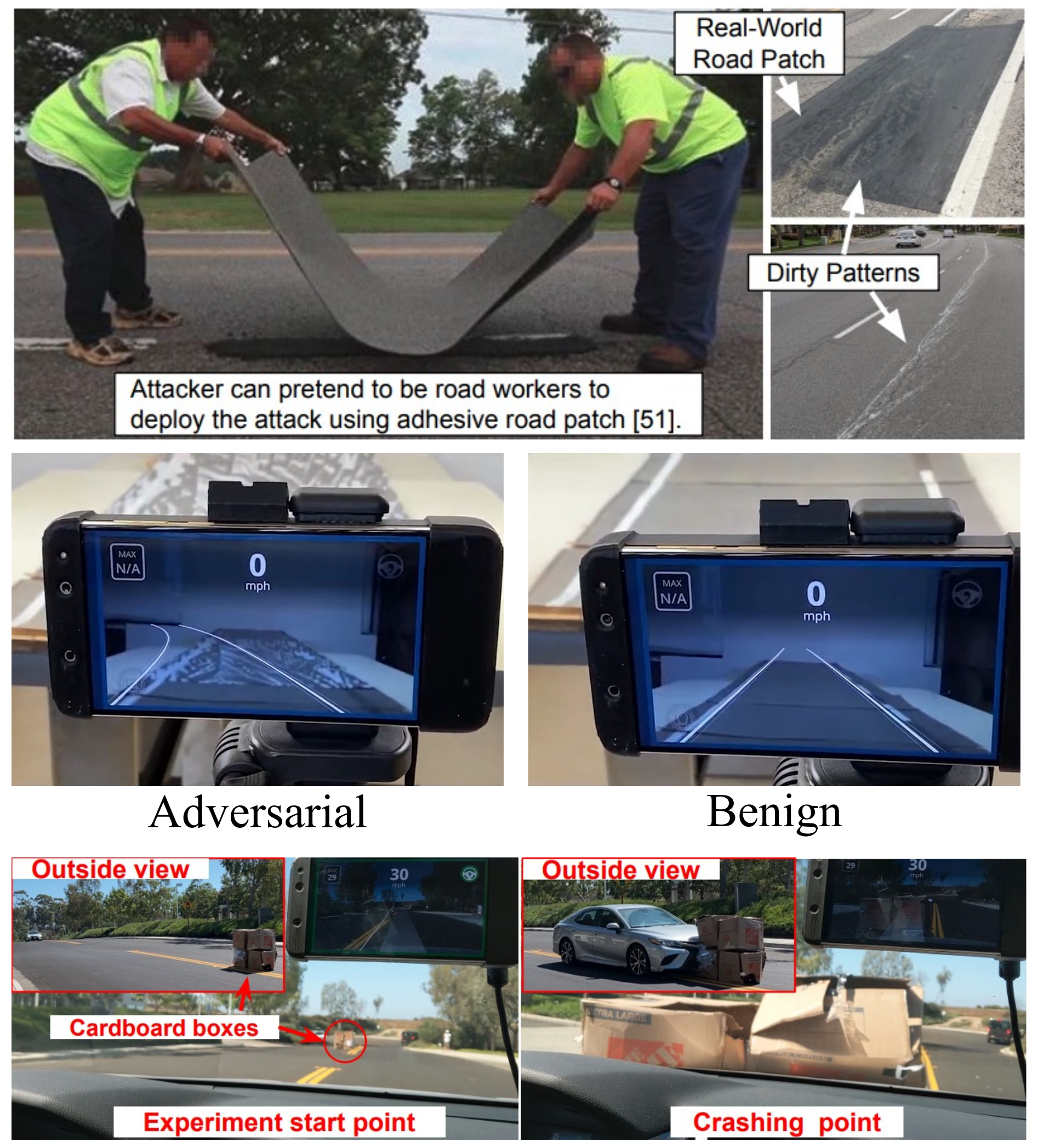

Neural Pruning Search
PointPillar
YOLOv4
Pruning Proposal for Object Detection Model
Controller
Evaluator
Evaluate and feedback their performance
Provide multiple pruning proposals
Bayesian Optimization with WL kernel
Obtain Best Pruning Proposal
Comparison
Comparison for 3D object Detection
Accuracy (mAP)
Latency (ms)
Our results
Comparison with other frameworks
Comparison for 2D object Detection

Adversarial T-shirt for Evading Person Detectors in a Physical World

- This work investigates adversarial examples deceiving DNN based decision makers by attaching adversarial patches to real objects.
- Most existing works in this area focus on static objects.
- We propose Adversarial T-shirts, a robust physical adversarial example for evading person detectors even if it could undergo non-rigid deformation due to a moving person's pose changes.
- It is the first work models the deformation effect for designing physical adversarial examples with respect to non-rigid objects.
- It achieves 74% and 57% attack success rates in the digital and physical worlds respectively against YOLOv2.

Dirty Road Can Attack: Security of Deep Learning based Automated Lane Centering (ALC) under Physical-World Adversarial Attack (https://sites.google.com/view/cav-sec/alc-adv-attack)

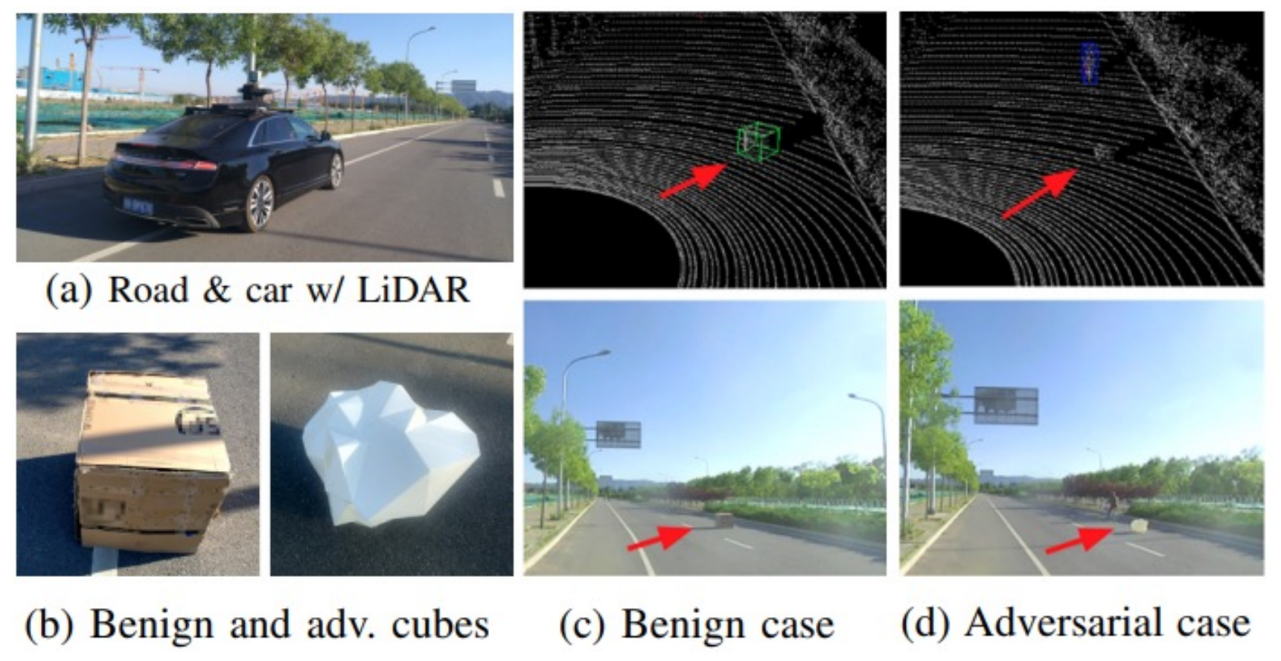
- First security analysis of deep learning based ALC under physical-world adversarial attacks.
- Identify a novel & domain-specific attack vector: Dirty Road Patches (DRP), adopt optimization based approach, overcomes various challenges.
- Evaluate on production ALC, showing high effectiveness (>97.5% success rate in <1 sec), robustness, transferability, stealthiness, physical-world realizability (after printed out), & end-to-end impact (100% collision rate)
- Evaluate & discuss defenses
- To appear in Usenix Security'21 (top-tier security conference)

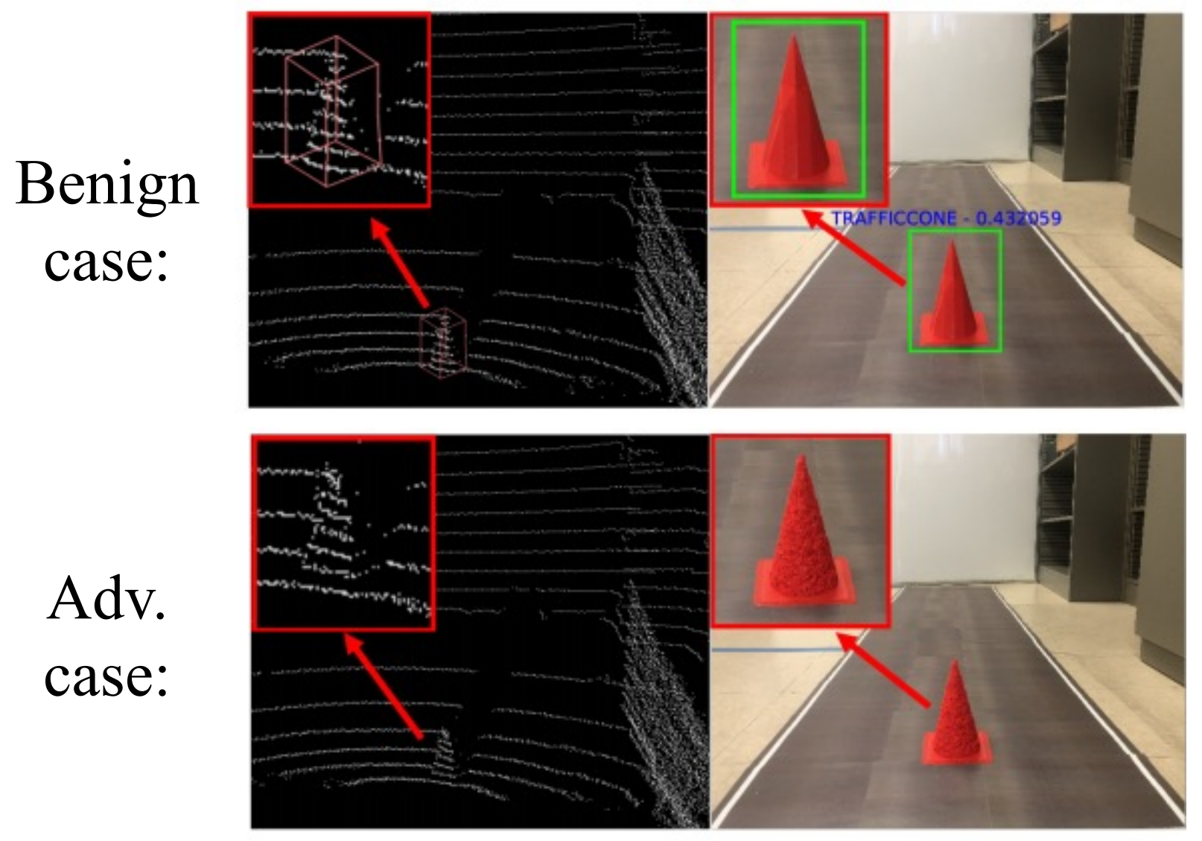


Invisible for both Camera & LiDAR: Security of Multi-Sensor Fusion (MSF) based Perception in Autonomous Driving (AD) under Physical-World Attacks

(https://sites.google.com/view/cav-sec/msf-adv)

- First security analysis of MSF-based AD perception, challenge basic security design assumption of MSF as a defense strategy in AD perception
- Identify advrsarial 3D object as physically-realizable & stealthy attack vector, adopt optimization-based approach that addresses various design challenges.
- Evaluate on industry-grade AD systems, showing high effectiveness (>91% success rate), stealthiness, robustness, transferability, physical realizability (after 3D-printed), & end-to-end impact (100% crash rate).
- Evaluate & discuss defense strategies.
- To appear in IEEE S&P'21 (top-tier security conf.)


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



2021 NSF Cyber-Physical Systems Principal Investigators' Meeting
June 2-4, 2021

Award ID#: CNS-1932351/1932464