Demo: Security of Deep Learning based Automated Lane Centering under Physical-World Attack

Takami Sato*, Junjie Shen*, Ningfei Wang, Yunhan Jack Jia†, Xue Lin‡, and Qi Alfred Chen
Univesity of California, Irvine; †ByteDance; ‡Northeastern University

***Abstract*—Automated Lane Centering (ALC) systems are convenient and widely deployed today, but also highly security and safety critical. Recently, Dirty Road Patch (DRP) attack is proposed as a state-of-the-art adversarial attack against ALC systems. We will show interactive demonstrations for the malicious road patch generation and the serious consequences.**

I. Introduction

Automated Lane Centering (ALC) is a Level-2 driving automation technology that automatically steers a vehicle to keep it centered in the traffic lane. So far, Deep Neural Network (DNN) based lane detection achieves the highest accuracy and is adopted in the most performant production ALC systems today such as Tesla Autopilot. Recent works show that DNNs are vulnerable to physical-world adversarial attacks.However, these methods cannot be directly applied to attack ALC systems due to three domain-specific design challenges: lack of legitimately-deployable attack vector, camera frame inter dependencies, lack of the optimization objective function designs for lane detection. To overcome the challenges, *Dirty Road Patch* (DRP) attack [1] is recently proposed as a domain-specific adversarial attack to ALC systems.

We will demonstrate the malicious road patch generation and the serious consequences of the DRP attack with a production-grade Autonomous Driving (AD) simulator LGSVL and a production ALC system in OpenPilot [2], which is reported to have close performance to Tesla Autopilot.

II. Threat Model and Attack Goal

Threat Model. We assume that the attacker can obtain a full (white-box) knowledge of the victim ALC.

Attack Goal. We consider an attack goal that directly breaks the design goal of ALC systems. The attack goals are to hit the concrete barrier or the truck under the average driver reaction time is 2.5 sec.

III. Attack Design: Dirty Road Patch Attack

We identify *dirty road patches* as a novel and domain-specific attack vector for physical-world adversarial attacks on ALC systems because road patches with dirt or white stains can appear to be legitimately deployed on the physical-world. We then design systematic malicious road patch generation with 2 major novel and domain specific designs: motion model based input generation, and lane-bending objective function. Motion model based input generation combines vehicle motion model and perspective transformation to dynamically synthesize camera frame updates according to attack-influenced vehicle control. Lane-bending objective function allow us to adopt the optimization-based attack generation, which has shown both high efficiency and effectiveness in previous works. More details are in [1].

*The first two authors contributed equally.

(a) Highway: hit concrete barrier. (b) Local: crash into a truck.

Fig. 1: Snapshots of the attack demo videos.

IV. Demonstration Plan

Demonstration of the malicious road patch generation. We will demonstrate how the road patch is generated by displaying intermediate and final patches, and camera inputs.

Demonstration of the attack in AD simulator. We will show two videos [1] of end-to-end attack scenarios for highway and local road settings respectively. The snapshots of the demo are shown in Fig 1. We deploy our malicious road patches in the production-grade the AD simulator, LGSVL. The target ALC system is the production ALC system OpenPilot [2]. For the highway scenario, we place a concrete barrier on the left, and for the local road scenarios, we place an NPC truck driving on an opposite direction lane. These videos demonstrate that our attacks cause the victim vehicle to hit the concrete barrier and the truck respectively.

Acknowledgements

This research was supported in part by the National Science Foundation under grants CNS-1850533, CNS-1929771, CNS-1932464, and USDOT UTC Grant 69A3552047138.

References

[1] T. Sato, J. Shen, N. Wang, Y. J. Jia, X. Lin, and Q. A. Chen, "Hold tight and never let go: Security of deep learning based automated lane centering under physical-world attack," *arXiv:2009.06701*, 2020.

[2] "OpenPilot," https://github.com/commaai/openpilot, 2018.

[1] https://youtu.be/y_5spn0Kxt8, https://youtu.be/TpipXnIvVhA

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.2021.23026
www.ndss-symposium.org