Demo: Does Physical Adversarial Example Really Matter to Autonomous Driving? Towards System-Level Effect of Adversarial Object Evasion Attack

Ningfei Wang Yunpeng Luo Takami Sato Kaidi Xu† Qi Alfred Chen
University of California, Irvine †Drexel University

***Abstract*—In autonomous driving (AD), accurate perception is indispensable to achieving safe and secure driving. Due to its safety-criticality, the security of AD perception has been widely studied. Among different attacks on AD perception, the physical adversarial object evasion attacks are especially severe. However, we find that all existing literature only evaluates their attack effect at the targeted AI component level but not *at the system level*, i.e., with the entire system semantics and context such as the full AD pipeline. Thereby, this raises a critical research question: can these existing researches effectively achieve system-level attack effects (e.g., traffic rule violations) in the real-world AD context? In this work, we conduct the first measurement study on whether and how effectively the existing designs can lead to system-level effects, especially for the STOP sign-evasion attacks due to their popularity and severity. Our evaluation results show that all the representative prior works cannot achieve any system-level effects. We observe two design limitations in the prior works: 1) physical model-inconsistent object size distribution in pixel sampling and 2) lack of vehicle plant model and AD system model consideration. Then, we propose SysAdv, a novel system-driven attack design in the AD context and our evaluation results show that the system-level effects can be significantly improved, i.e., the violation rate increases by around 70%.**

I. Introduction

In this Demo, we focus on the security of Autonomous Driving (AD) perception, particularly the vulnerability to physical adversarial object evasion attacks, such as evading detection of STOP signs and pedestrians, due to their safety critical nature in driving. Previous studies mainly assessed these attacks at the component level (such as per-frame object misdetection rates) without considering system-level implications, such as traffic rule violations and vehicle crashes.

To address this, we perform a comprehensive measurement study of existing attacks, assessing their impact on the AD system as a whole. The study uses a new framework to model the AD system's response to these attacks, particularly focusing on STOP sign evasion [1]. The findings reveal that, while some attacks show high success rates at the AI component level, they fail to cause significant system-level effects such as traffic rule violations in realistic scenarios.

Symposium on Vehicles Security and Privacy (VehicleSec) 2024
26 February 2024, San Diego, CA, USA
ISBN 979-8-9894372-7-6
https://dx.doi.org/10.14722/vehiclesec.2024.25014
www.ndss-symposium.org

We identifies two main reasons for this failure: 1) physical model-inconsistent object size distribution in pixel sampling and 2) lack of vehicle plant model and AD system model consideration. To overcome these limitations, the paper proposes a new system-driven attack design SysAdv, which, when integrated with existing methods, significantly improves system-level impact. This design was tested across various AD system settings and object types, showing notable improvements in both component- and system-level effectiveness.

II. Demonstration Plan

Demonstration of the generated adversarial object evasion attack. We will show 6 STOP signs for measurement study on prior works and 10 STOP signs for SysAdv [2]. We print the high-resolution STOP signs on multiple ledger-size papers and concatenate them together to form full-size real STOP signs. All the demo images are captured from real world. Some demos are currently available in our website https://sites.google.com/view/cav-sec/sysadv.

Demonstration of real-world videos for STOP sign detection. To enhance the perception fidelity of simulators, we model the perception results using a practical setup in the real world using an iPhone 12 Pro Max starting from 45 m to 4 m [2]. We will demonstrate the videos.

Demonstration of system-level effects in simulation. To measure system-level effects, we adopt a simulation-centric evaluation methodology with SVL and San Francisco map on a sunny day at noon. We will demonstrate the system-level attack effect, i.e., STOP sign traffic rule violation, on a representative AD system design [1], [2] with both prior attacks and our SysAdv showing improvements in system-level effectiveness.

III. Acknowledgments

This research was supported by the NSF under grants CNS-1932464, CNS-1929771, and CNS-2145493; and USDOT UTC Grant 69A3552348327.

References

[1] J. Shen, N. Wang, Z. Wan, Y. Luo, T. Sato, Z. Hu, X. Zhang, S. Guo, Z. Zhong, K. Li *et al.*, "SoK: On the Semantic AI Security in Autonomous Driving," *arXiv preprint arXiv:2203.05314*, 2022.

[2] N. Wang, Y. Luo, T. Sato, K. Xu, and Q. A. Chen, "Does Physical Adversarial Example Really Matter to Autonomous Driving? Towards System-Level Effect of Adversarial Object Evasion Attack," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023, pp. 4412–4423.