Shuai ZHAO, Boyuan ZHANG, Yucheng SHI, Yang ZHAI, Yahong HAN, Qinghua HU, 2025. A comprehensive survey of physical adversarial vulnerabilities in autonomous driving systems. *Frontiers of Information Technology & Electronic Engineering*, 26(4):510-533. https://doi.org/10.1631/FITEE.2300867

A comprehensive survey of physical adversarial vulnerabilities in autonomous driving systems

Front Inform Technol Electron Eng

Key words: Physical adversarial attacks; Physical adversarial defenses; Artificial intelligence safety; Deep learning; Autonomous driving system; Data-fusion; Adversarial vulnerability

Corresponding author: Yahong HAN
E-mail: yahong@tju.edu.cn

ORCID: https://orcid.org/0000-0003-2768-1398

Motivation

- Autonomous driving systems (ADSs) have attracted wide attention in the machine learning communities. With the help of deep neural networks (DNNs), ADSs have shown both satisfactory performance under significant uncertainties in the environment and the ability to compensate for system failures without external intervention.
- The growing interest in the development of ADSs has introduced new security challenges and vulnerabilities. Besides the usual cyber attacks, such as denial-of-service (DoS) attack, black-hole attack, and malware attack, the vulnerability of DNNs in ADSs needs to be investigated.

Front Inform Technol Electron Eng

Main idea

- We present a comprehensive survey of the current physical adversarial vulnerabilities in ADSs. We first divide the physical adversarial attack methods and defense methods by their restrictions of deployment into three scenarios: real-world, simulator-based, and digital-world scenarios.
- We consider the adversarial vulnerabilities that focus on various sensors in ADSs and separate them as camera-based, LiDAR-based, and multifusion-based attacks.
- We divide the attack tasks by traffic elements. For the physical defenses, we establish the taxonomy with reference to input image preprocessing, adversarial example detection, and model enhancement for the DNN models to achieve full coverage of the adversarial defenses.

Front Inform Technol Electron Eng

Method

We propose a novel taxonomy to map the ADSs under adversarial at tacks and defenses. We analyze the adversarial attacks of ADSs from three perspectives: attack scenarios, attack sensors, and attack tasks. After that, we investigate the robustness of ADSs and discuss the corresponding defense strategies in the following three steps: input image preprocessing, adversarial example detection, and model enhancement.

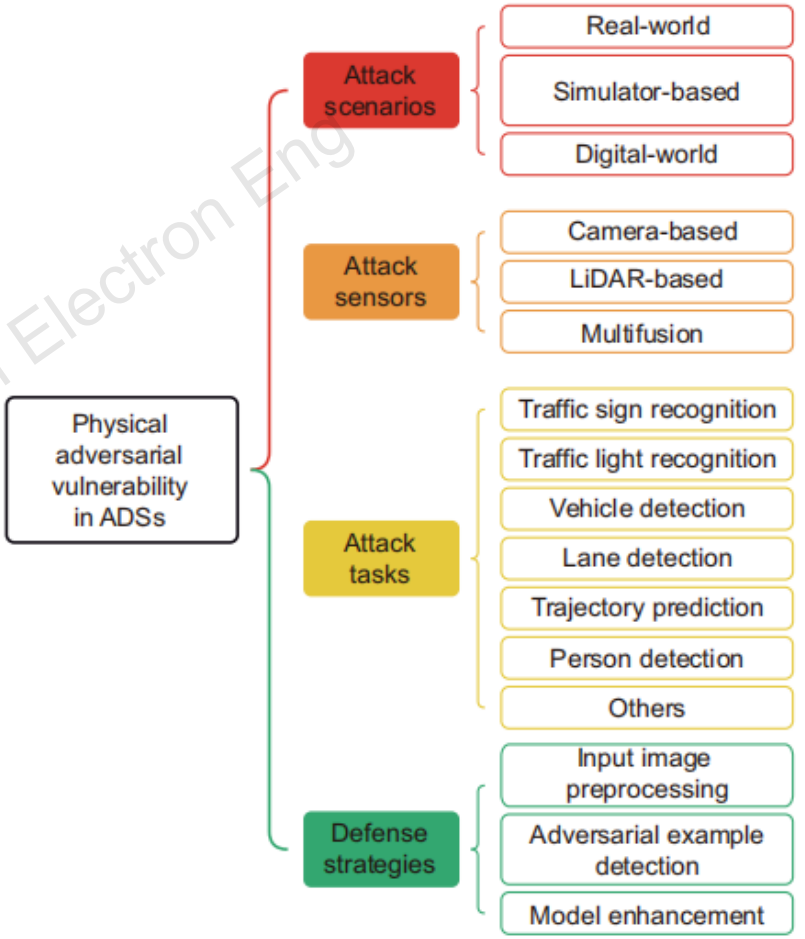

Fig. 2 Framework of this survey. We analyze the adversarial vulnerability from four perspectives: attack scenarios, attack sensors, attack tasks, and defense strategies

Front Inform Technol Electron Eng

Framework

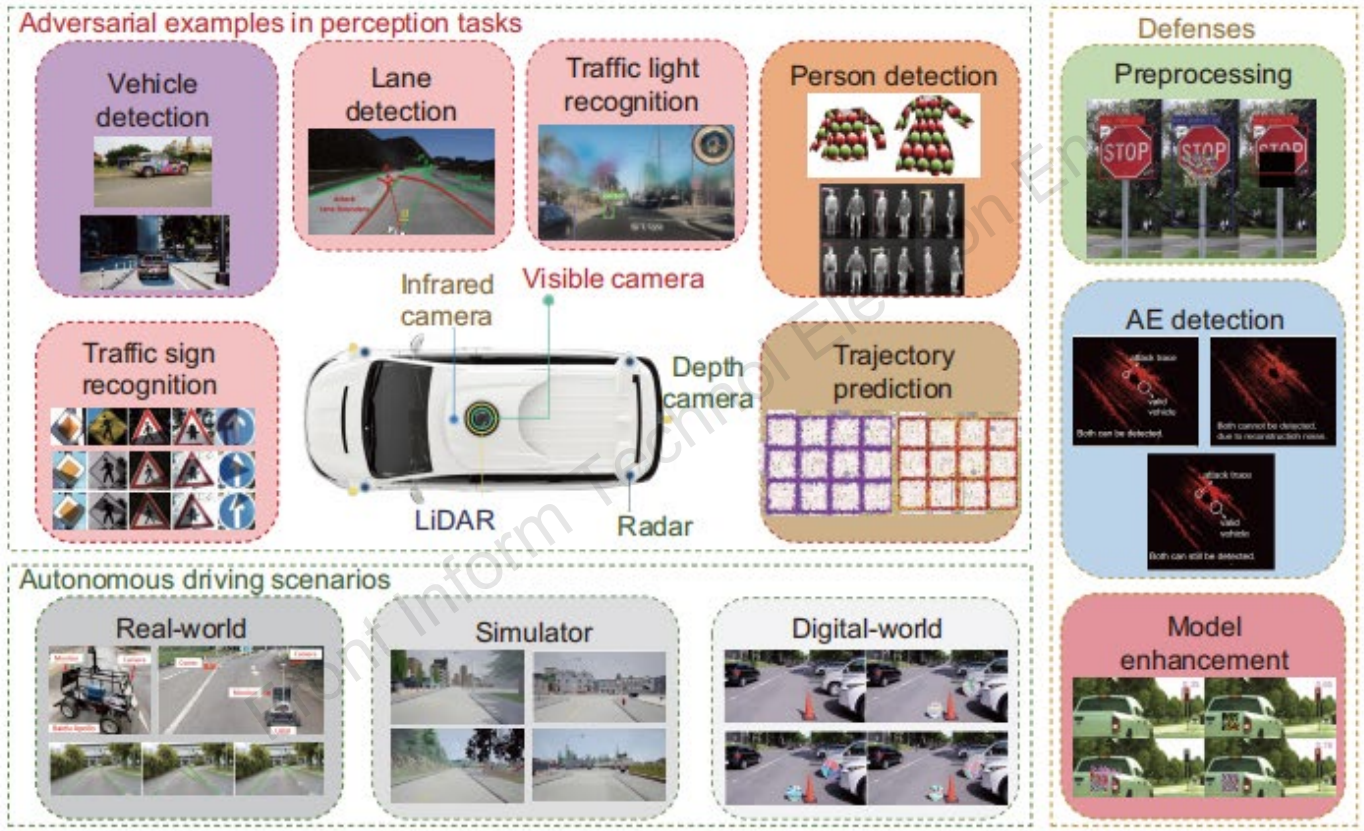

Fig. 3 Adversarial examples in autonomous driving systems and different perspectives to evaluate the vulnerability of ADSs. A few examples of recent research are presented to illustrate each subcategory

Discussion

- The adversarial vulnerability in ADSs lacks united platforms and comprehensive benchmarks.
- The adversarial transferability of physical adversarial attacks under ADS requires more investigation.
- The deployment and evaluation of query-based attacks in real-world scenarios present significant practical challenges.
- The cost and technical difficulty of evaluating ADS perception tasks in the real world are high, except for traffic sign recognition and person detection tasks.
- The stealthiness of physical adversarial examples in ADS needs to be enhanced.

Front Inform Technol Electron Eng

Conclusions

In this survey, we present a comprehensive analysis of the physical adversarial attacks and defenses in the ADS area. We categorize the physical adversarial attacks considering attack scenarios, sensors, and tasks. Subsequently, we divide the physical adversarial defenses in terms of input image preprocessing, adversarial example detection, and model enhancement. We discuss the current limits and future directions about the adversarial vulnerability in ADSs. We hope that this survey can bring inspiration to the autonomous driving community for future directions.

Front Inform Technol Electron Eng

Shuai ZHAO, currently pursuing an Engineering PhD at Tianjin University, is a Chief Expert of China Automotive Technology and Research Center Co., Ltd., and a Deputy General Manager of CATARC Intelligent and Connected Technology Co., Ltd. He is mainly engaged in intelligent vehicle scenario simulation research.

Yahong HAN received the PhD degree from Zhejiang University, Hangzhou, China, in 2012. He is currently a professor with the College of Intelligence and Computing, Tianjin University, Tianjin, China. From 2014 to 2015, he visited Prof. Bin YU's group with UC Berkeley as a visiting scholar. His research interests include multimedia analysis, computer vision, and machine learning.

Qinghua HU received the MS and PhD degrees from Harbin Institute of Technology, Harbin, China. He was a postdoctoral fellow with the Department of Computing at the Hong Kong Polytechnic University, Hong Kong, China. Currently, he is the director of the Tianjin Key Laboratory of Machine Learning and the chairperson of the Tianjin Branch of China Computer Federation and the Chinese Association of Artificial Intelligence. His research focuses on uncertainty modeling in big data, machine learning with multimodal data, and intelligent unmanned systems.

Front Inform Technol Electron Eng

Yang ZHAI, currently pursuing an Engineering PhD at Tianjin University, is the business director of CATARC Intelligent and Connected Technology Co., Ltd., a registered expert of National Technical Committee of Auto Standardization, and a deputy leader of AI Working Group for China Software Testing Center. She is engaged in long-term research on autonomous driving visual perception and adversarial machine learning.

Boyuan ZHANG received the MS degree from University of Leeds, Leeds, UK, in 2019. He is currently pursuing the PhD degree with the College of Intelligence and Computing, Tianjin University, Tianjin, China. His research interests include computer vision, adversarial machine learning, and domain adaptation.

Yucheng SHI received the PhD degree from Tianjin University, Tianjin, China, in 2023. He is currently a lecturer with the School of Computer Science and Artificial Intelligence, Zhengzhou University, Zhengzhou, China. His research interests include computer vision, adversarial machine learning, and federated learning.

Front Inform Technol Electron Eng