Security of Deep Learning based Lane Keeping Assistance System under Physical-World Adversarial Attack

PRESENTERS

Takami Sato
takamis@uci.edu

Junjie Shen
junjies1@uci.edu

RESEARCH PROBLEM & MOTIBATION

- **Lane-Keeping Assistance System (LKAS)**
  - ➢ Automatically steers vehicle to keep it in lane
  - ➢ Level-2 driving automation
  - ➢ Widely available in a variety of vehicle models
    - Honda Civic, Toyota Prius, Nissan Cima, Volvo XC90, Audi A4, and Tesla Model S
- When LKAS fails to keep in lane:
  - ➢ **Avg. driver reaction time 2.3 s not enough for**
    - Drive off-road to hit road curbs or fall down highway cliff
    - Crash into other cars or be crashed into
- **Our work:**
  - **First** systematic adversarial attack to LKAS
  - Target most performant design today: DNN-based LKAS

THREAT MODEL

- Attacker possesses the same LKAS as victim
- Has full knowledge of the LKAS
- Can collect road images before attack

ATTACK GOAL

- **Drive out of current lane boundary** within **2.3 seconds (avg. driver reaction time)**
- The required deviation is **0.745 m** on highway

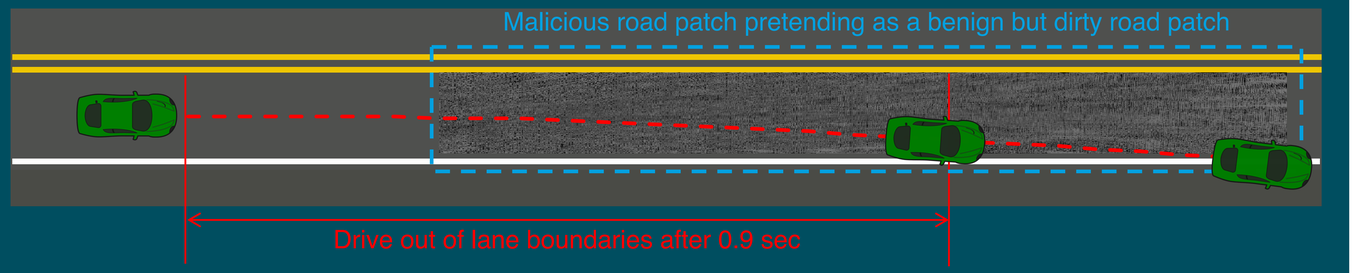
DESGIN CHALLENGES

- Most prior attack targets obj. detection, **not LKAS**
- Need to be **realizable** & **stealthy** in physical world
- Attack to consecutive frames are inter-dependent

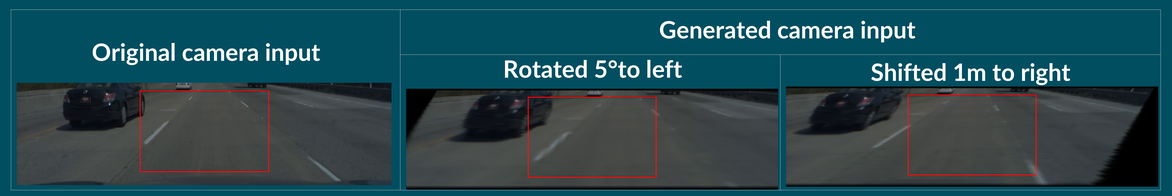
REALIZABLE & STEALTHY PHYSICAL-WORLD PERTURBATION DESIGN

- **New attack vector: malicious dirty road patch**
  - ➢ Realizable in the physical world
  - ➢ Can normally appear around traffic lane
  - ➢ Not cover the original lane lines
  - ➢ Only use gray-scale color to pretend to be benign but dirty

Our attack successfully deviates state-of-the-art LKAS to **drive off lane boundary** after **0.9 seconds**.



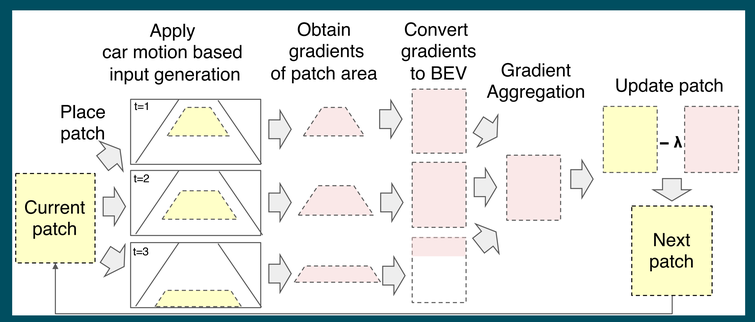
Car Motion Model based Input Generation



Multi-frame Path Bending Objective Function

$$f(X_1, ..., X_T, s_0) = \sum_{t=1}^{T} \sum_{d \in D} \nabla p_t(d; \{X_j | j \le t\}, s_0) + \lambda ||\Omega_t(X_t)||_2^2$$

Gradient Aggregation

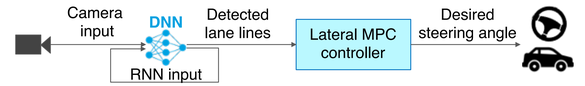


ATTACK METHODOLOGY

- **Car Motion Model based Input Generation**
  - ➢ Simulate **inter-dependency** by **bicycle model**
- **Multi-frame Path Bending Objective Function**
  - ➢ Design attack as optimization problem integrated **inter-dependency**
- **Gradient Aggregation**
  - ➢ Update malicious road path while keeping **inter-dependency , realizability,** and **stealthiness**

EVALUATION SETUP

- **SCENARIOS**
  - ➢ Comma2k19 dataset: real-world highway driving
  - ➢ LGSVL: industry-grade driving simulator
- **TARGET LKAS: OpenPilot**
  - Open source, on par with Tesla and GM Super Cruise



EARLY RESULTS

- Success criteria: car deviates **0.745 m** within **2.3 s**

Eval. Scenario	Avg. Speed	Attack Success Time	Patch Size
comma2k19-1	126 km/h (78 mph)	**0.9 s**	3.6m × 36m
comma2k19-2	105 km/h (65 mph)	**1.0 s**	3.6m × 36m
LGSVL-1	72 km/h (45 mph)	**1.3s**	3.6m × 36m

FUTURE PLANS

- More comprehensive evaluation
- Real-world experiment
- Design effective defenses

Takami Sato*, Junjie Shen*, Ningfei Wang, Yunhan Jack Jia, Xue Lin, Qi Alfred Chen

*Contributed equally

UC Irvine

ByteDance 字节跳动 Northeastern University

Take a picture to **download** the **poster abstract**