# Too Good to Be Safe: Tricking Lane Detection in Autonomous Driving with Crafted Perturbations

Pengfei Jing, Qiyi Tang, Yuefeng Du, Lei Xue, Xiapu Luo, Ting Wang, Sen Nie, Shi Wu

Department of Computing, The Hong Kong Polytechnic University
Keen Security Lab, Tencent
College of Information Sciences and Technology, Pennsylvania State University

Present by Deng Pan, November 3rd, 2022

# Motivation

- Autonomous Vehicle relays on multiple sensors and machine learning algorithms to detect and reconstruct the surrounding environment
- Lane detection is one of the major tasks
- misleading the lane detection module can lead to severe consequences

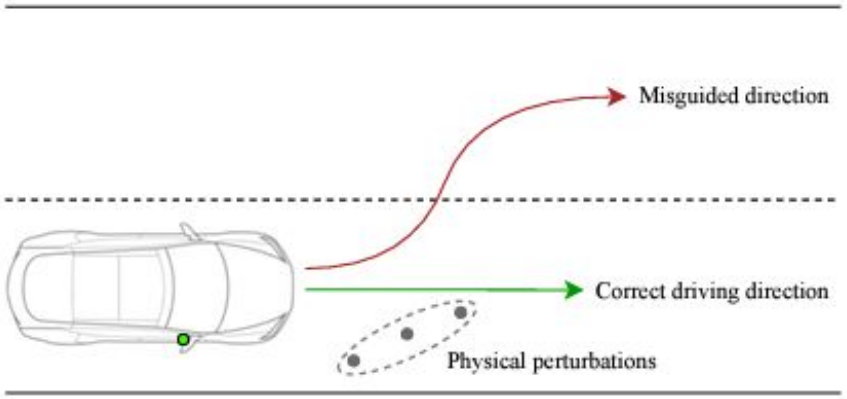
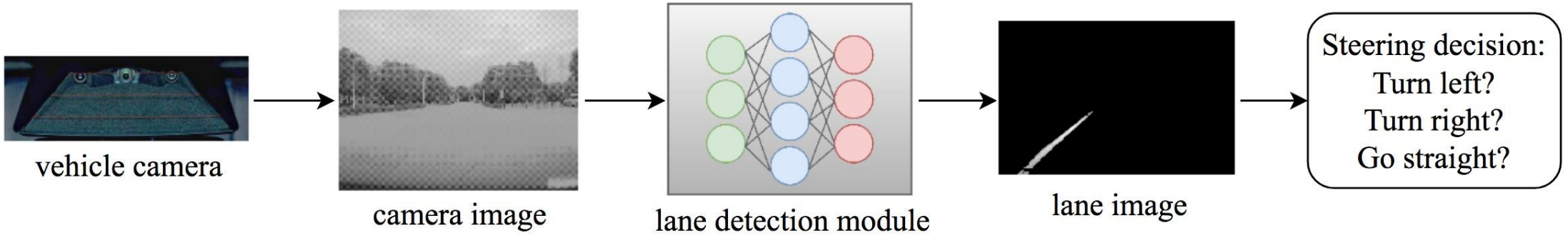



Figure 1: If the physical perturbations added by an adversary are recognized as a lane, the vehicle is likely to follow the fake lane and swerve into the wrong direction.

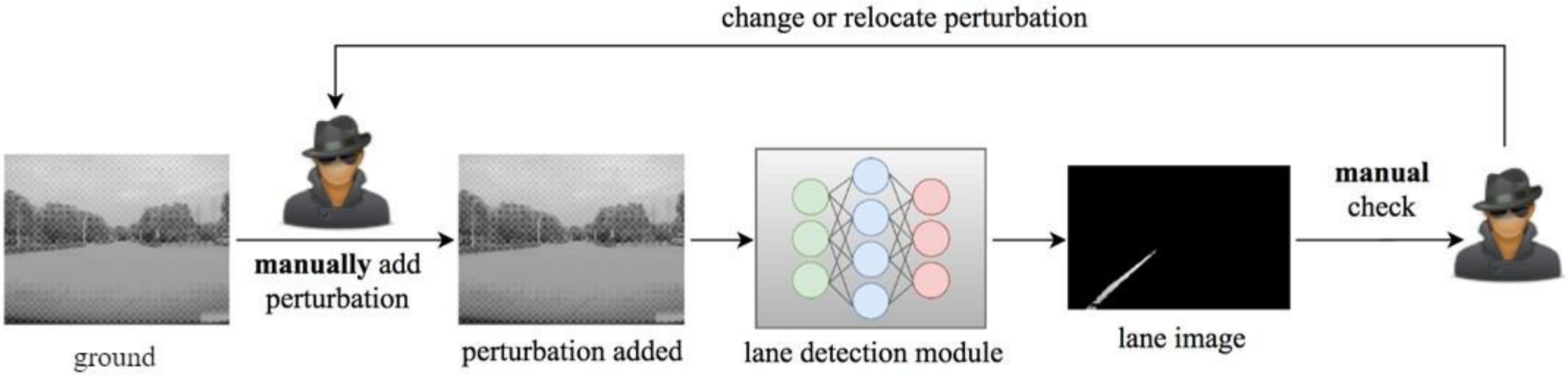
# Lane Detection




1. Images are collected by camera.
2. Based on the camera images, lane detection module generates the corresponding lanes.
3. Autonomous vehicle behaves based on the lane detection result.

- **Changing the lane detection result can affect the steering decision (e.g., exploiting its over-sensitivity to create a fake lane!).**

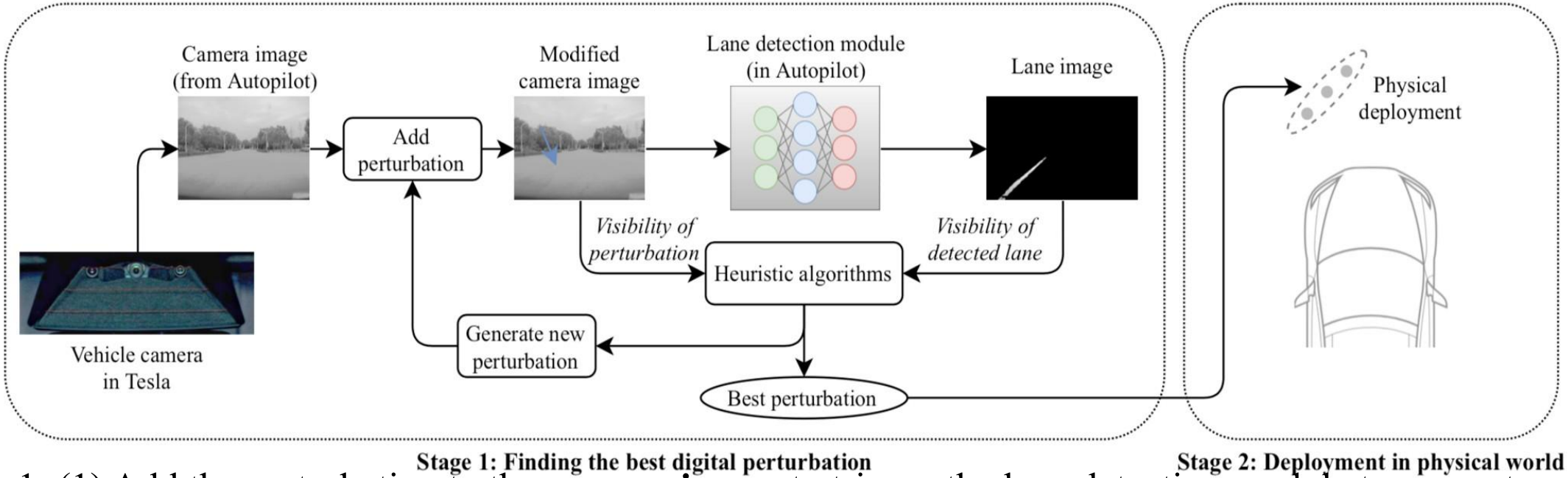
# Creating a Fake Lane - An Intuitive Approach




Add perturbations and check whether the module will be affected ***Manually***. If not, the perturbation should be changed or relocated.

Unfortunately, such an approach is very **labor-intensive** and **error-prone**.

# Two-stage Attack




- Stage 1: (1) Add the perturbation to the **camera image** to trigger the lane detection module to generate the corresponding **lane image**.

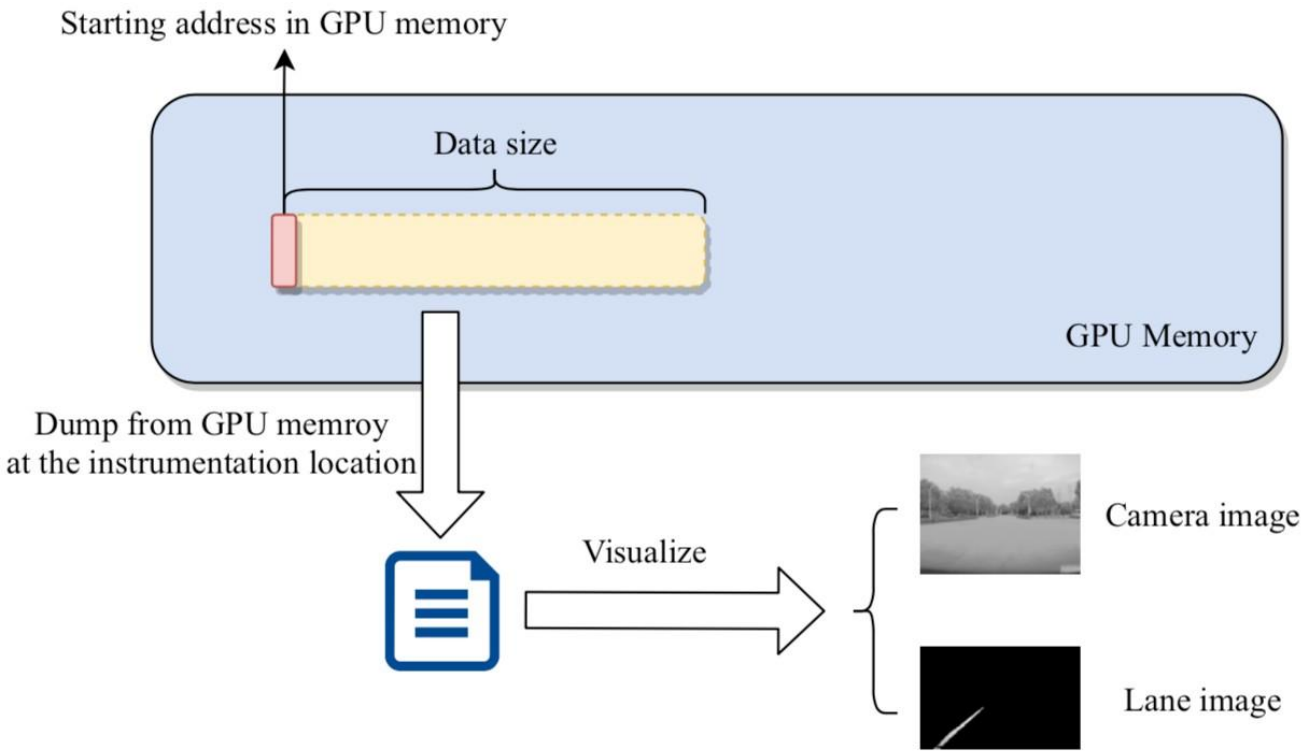
  (2) Formulate an optimization problem based on the **visibility** of perturbation and that of detected lane and adopt **heuristic algorithms** to find the best perturbation.

- Stage 2: We deploy the best perturbation in physical world for evaluation.

# Challenges and Solutions (I)

- *Challenge_1:* How to extract the data from the **real vehicle**, which is not exposed to users? (the closed-source firmware of Tesla Autopilot)

- *Solution_1:* Conduct static and dynamic analysis on the firmware responsible for lane detection to collect the data (camera image and lane image) from the vehicle.



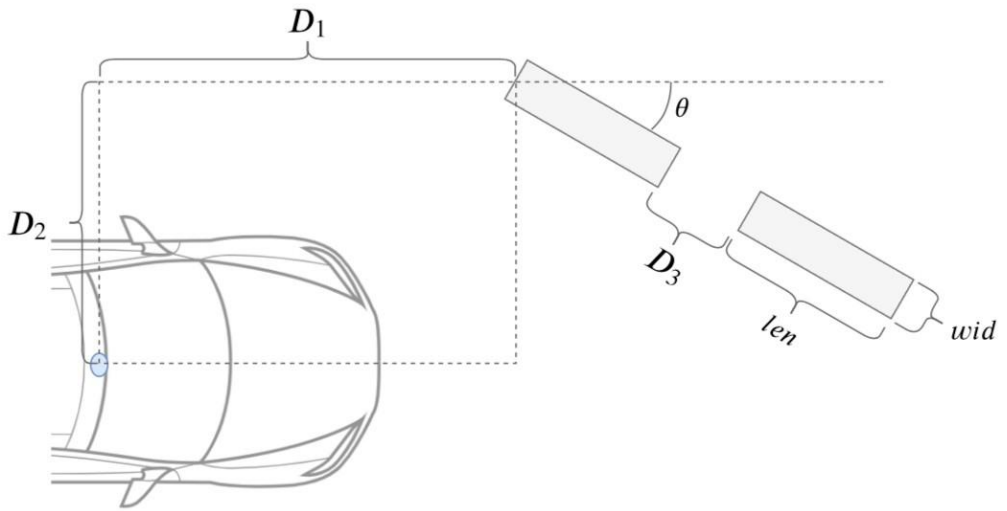

Dumping and visualizing the target data from the GPU on Autopilot

# Challenges and Solutions (II)

- *Challenge_2:* How to add perturbations to input camera image?
- *Solution_2:* For the ease of deployment, we use 8 parameters, which form a vector $x$, to represent the attributes of the perturbations. With pinhole camera model and undistortion techniques, these perturbations can be accurately mapped to **digital** images.

| Parameters | Explanation |
|---|---|
| $len$ | Length of a single perturbation |
| $wid$ | Width of a single perturbation |
| $D_1$ | Longitudinal distance from the vehicle camera to the edge of the first perturbation |
| $D_2$ | Lateral distance from the vehicle camera to the edge of the first perturbation |
| $D_3$ | Distance between adjacent perturbations |
| $\Delta G$ | Increment of grayscale value of the perturbed pixels |
| $\theta$ | Rotation angle of the perturbation |
| $n$ | Number of the perturbations |

Parameters determining the added perturbation




$$x = (len, wid, D_1, D_2, D_3, \Delta G, \theta, n) \in X$$

Illustration of the parameters

# Challenges and Solutions (III)

- *Challenge_3:* How to find the best perturbations?
- *Solution_3:* We formulate an **optimization problem** to find the best perturbations. Specifically, we quantify the quality by the visibility of lane and visibility of perturbation. The visibility of lane should be high (to make the attack effective), and the visibility of perturbation should be low (to make the perturbation unobtrusive).

$$V_{lane}(x) = \sum_{p \in lane_o(x)} G_p$$

$$V_{perturb}(x) = \sum_{p \in perturb_i(x)} \Delta G$$

$$S(x) = \frac{V_{lane}(x)}{V_{perturb}(x)}$$

| Parameters | Explanation |
|---|---|
| $p$ | One single pixel in the image |
| $lane_o(x)$ | Lane pixels in the output image |
| $perturb_i(x)$ | Pixels on the added perturbations |
| $G_p$ | Grayscale value of pixel $p$ |
| $V_{lane}(x)$ | Visibility of the fake lane created by $x$ |
| $V_{perturb}(x)$ | Visibility of the perturbations added by $x$ |
| $S(x)$ | Overall score of the parameter $x$ |

Explanations of parameters

*S(x)* represents the overall score, based on which we use heuristic algorithms to find the perturbation with the highest score:

$$x^* = \max_{x \in X} S(x)$$

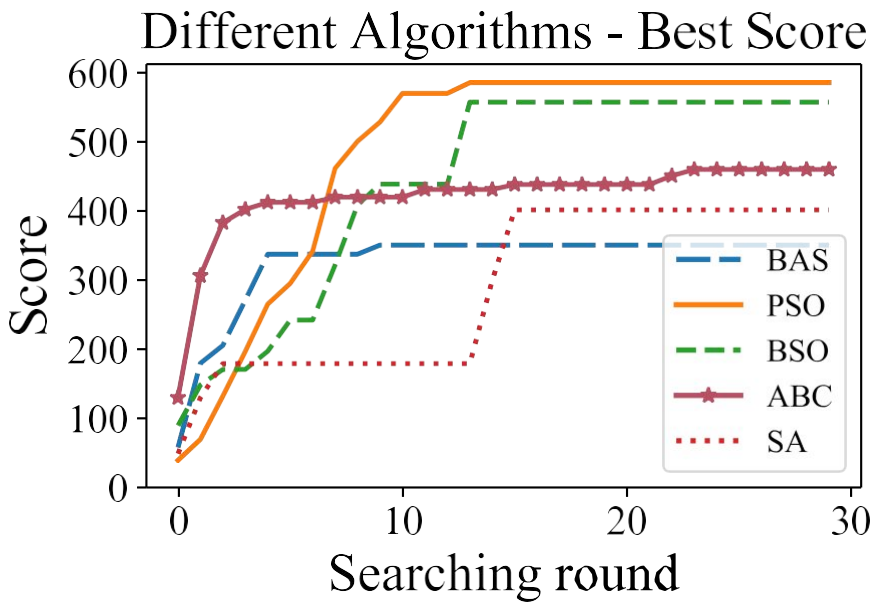
# Evaluation

**Q1:** How efficient are the heuristic algorithms to find the best perturbation?

**Approach:** Use 5 heuristic algorithms to find the best perturbations:

- *Beetle Antennae Search (BAS)*
- *Particle Swarm Optimization (PSO)*
- *Beetle Swarm Optimization (BSO)*
- *Artificial Bee Colony (ABC)*
- *Simulated Annealing (SA)*

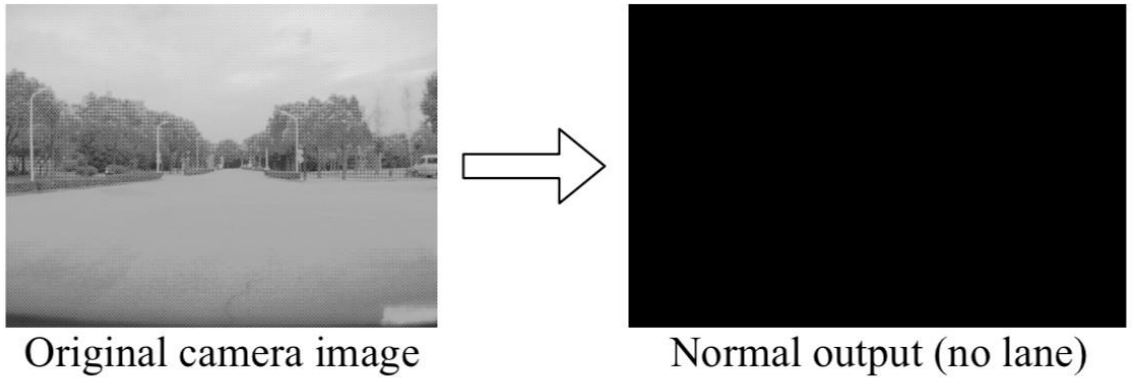
**Answer:** ***PSO*** is the most efficient one and thus we use it in other experiments.




Comparison between 5 algorithms

# Evaluation

A best perturbation is shown below. The added perturbation is only 1cm wide in physical world, but it causes the lane detection module to generate a fake lane.




Effect of a best perturbation

*Zoom in*

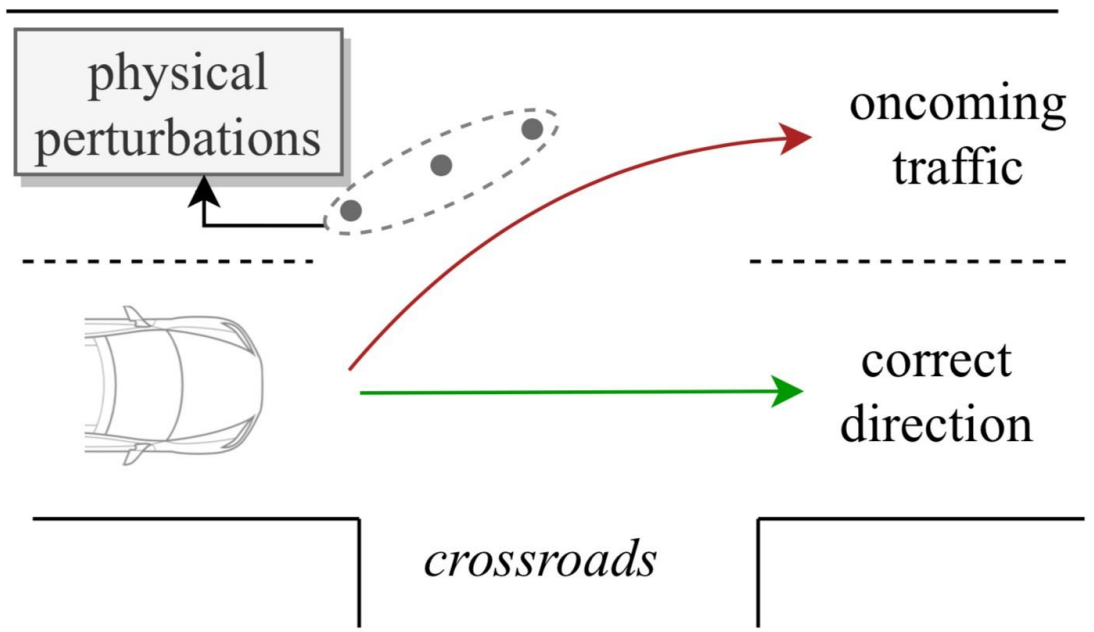
The perturbation can be hardly noticed

# Evaluation

**Q2:** Can we misguide the vehicle in physical world?

**Approach:** Deploy the perturbations in a crossroads scenario. Then we switch the vehicle to **auto-steer mode** and let it pass the crossroads.

**Answer:** the fake lane can successfully misguide the vehicle into **oncoming traffic**.




Misguiding the vehicle in a crossroads scenario

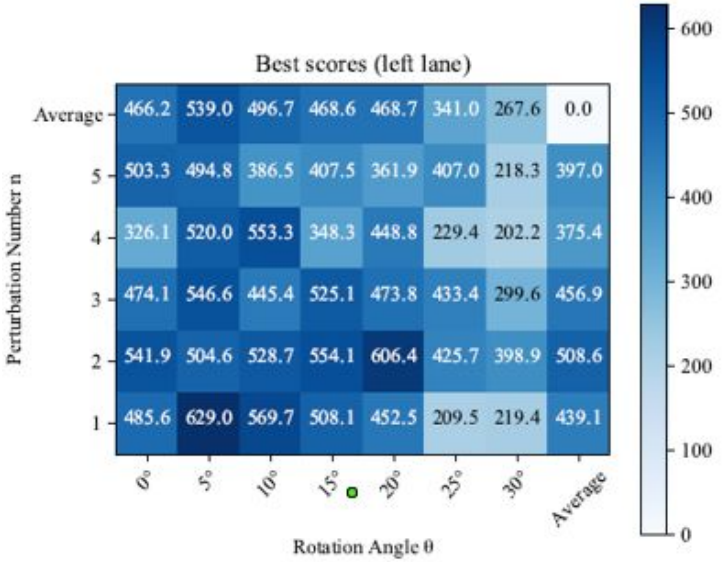

RQ6: Misguide the vehicle to the oncoming traffic
科恩实验室
KEEN SECURITY LAB
Play (k)
1:05 / 1:14

# Evaluation

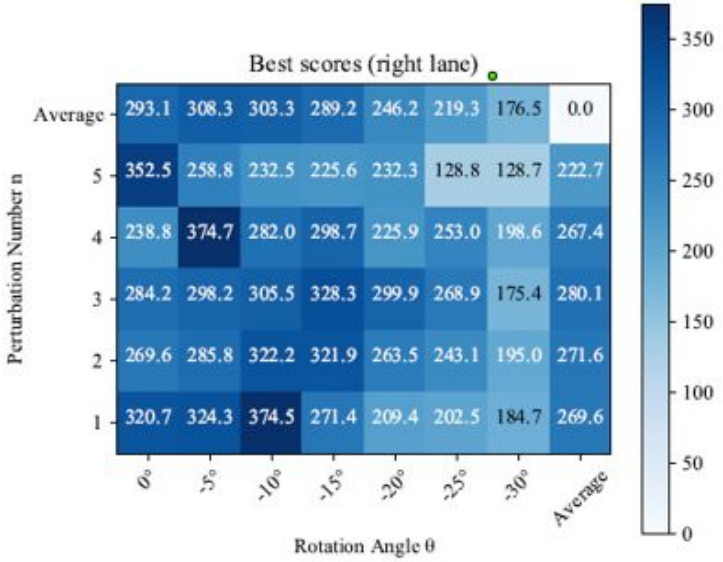
**Q3:** How do the perturbation number n and the rotation angle q affect the best perturbation?

**Approach: The** perturbation number is set from 1 to 5, and the absolute value of q is 0 to 30 degrees with an interval of 5 degrees (70 different settings).

**Answer:** Perturbation number n does not have significant effect on S(x). Rotation angle q reduces S(x) when it increases.




(a) Left lane




(b) Right lane

# Evaluation

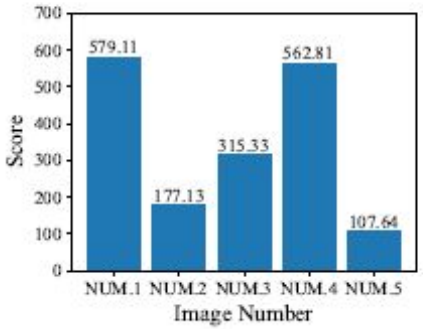
**Q4:** How is the performance of the approach given different input camera images?

**Approach:** Use four other images taken by the vehicle camera in different environments. NUM.1 is the original input image. NUM.2 and NUM.3 are in the same outdoor environment but under different light conditions. NUM.4 is taken in an underground garage, where the ground is clean and the light is dim. NUM.5 is a corner where the ground is dirty. Setting n = 1 and q = 0 and using S(x).

**Answer:**NUM.1 and NUM.4 lead to higher score than the others. Although the scores of NUM.2/3/5 are relatively low, the perturbations are unnoticeable to human eyes and the fake lane is valid and strong. The approach can successfully generate perturbations that can mislead the lane detection module without being noticed by the driver.

NUM.1 NUM.2 NUM.3 NUM.4 NUM.5

| Num | Environmental Features |
|---|---|
| 1 | Clean and bright ground, without other disturbing objects in view |
| 2 | Clean and bright ground, with disturbing objects in view |
| 3 | Clean and dark ground, with disturbing objects in view |
| 4 | Clean and dark ground, without other disturbing objects in view |
| 5 | Dirty and bright ground, with disturbing objects in view |



# Evaluation

**Q5:** What are the common characteristics of the best perturbations?

**Approach:** Analyze the parameters x of the best perturbations obtained in the five different scenarios for answering RQ4 and summarize the common characteristics.

**Answer**: 'Narrow but long' perturbations are more likely to create a fake lane. The required increment of grayscale value (DG) depends on the brightness and cleanliness of the ground.

| Num | Environmental Features |
|---|---|
| 1 | Clean and bright ground, without other disturbing objects in view |
| 2 | Clean and bright ground, with disturbing objects in view |
| 3 | Clean and dark ground, with disturbing objects in view |
| 4 | Clean and dark ground, without other disturbing objects in view |
| 5 | Dirty and bright ground, with disturbing objects in view |

| *x* / *Num* | *wid* | *len* | $D_1$ | $D_2$ | $\Delta G$ |
|---|---|---|---|---|---|
| NUM.1 | 1cm | 117cm | 15.30m | 2.23m | 12 |
| NUM.2 | 5cm | 59cm | 13.37m | 2.27m | 28 |
| NUM.3 | 3cm | 72cm | 12.53m | 1.51m | 12 |
| NUM.4 | 1cm | 133cm | 11.68m | 1.79m | 7 |
| NUM.5 | 1cm | 83cm | 10.14m | 2.38m | 25 |
| Average | 2cm | 93cm | 12.60m | 2.04m | 17 |

# Discussions

- The author finally indicates two possible solution, one is enhancing the lane detection module (Detecting abnormal lane's feature) and the other one is enhancing the control policy (e.g., combine with other visual elements). Any thinking about these two possible solution and any other ideas?
- Even if the driver notices the perturbations when the vehicle is going to the place where the crafted perturbations have been deployed, there may not be enough time for the driver to react. Any solutions?
- The author provides the demo and details of that attack to Tesla company and the company confirmed that this attack can change the target car's behavior when the vehicle is in auto-steer mode, but the company don't mention any fix plan. If you are a user that wants to buy a Tesla car, will you change mind if you know such information?

# Thinking

- Accepted.
- First investigation on the lane detection module in real vehicles.
- The safety of autonomous driving technology still needs to be improved.
- In the future, if the autonomous drive system is developed, Would user 100% believe in self-driving model?

# Thank you!