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

Pengfei Jing[12], Qiyi Tang[2], Yuefeng Du[2], Lei Xue[1], Xiapu Luo[1], Ting Wang[3], Sen Nie[2], Shi Wu[2]

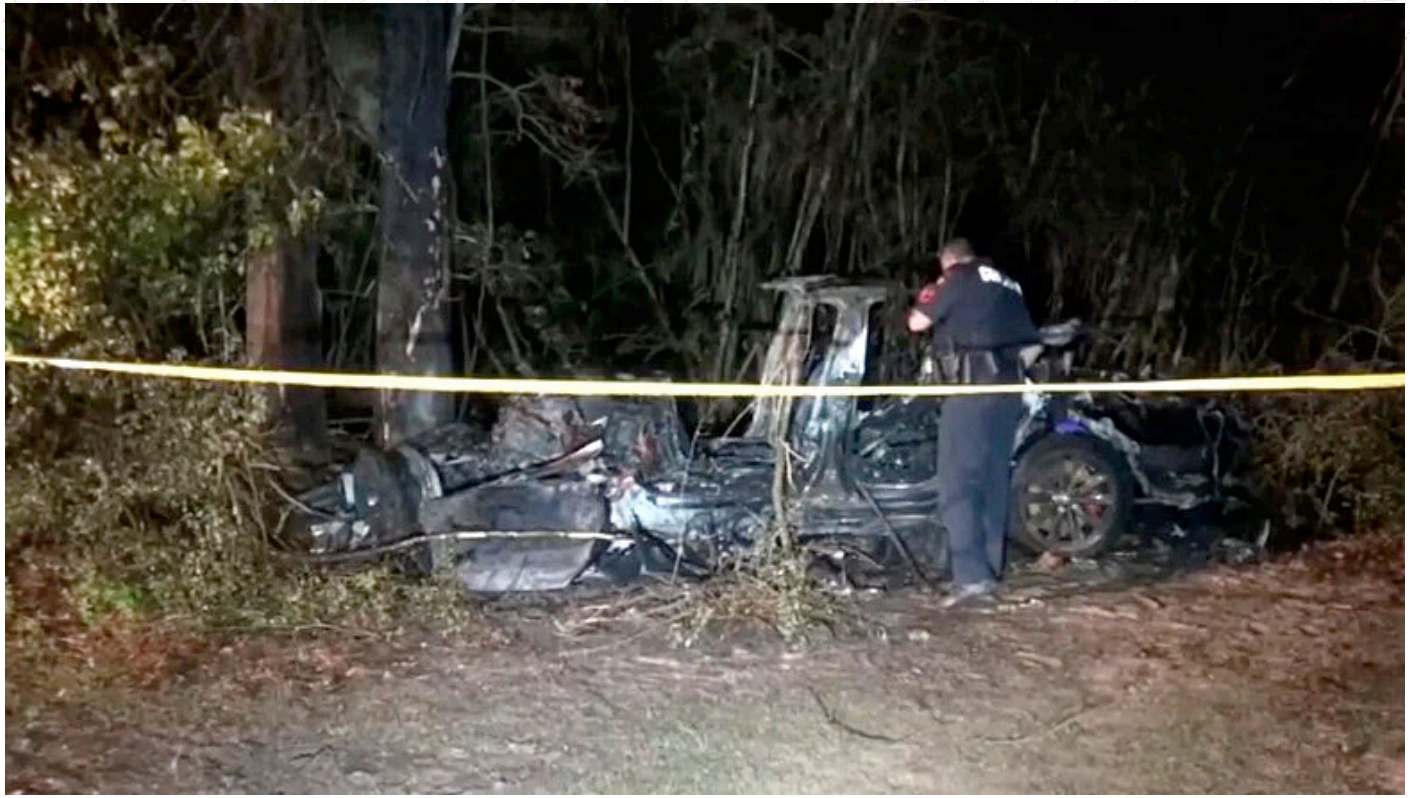
[1] Department of Computing, The Hong Kong Polytechnic University
[2] Keen Security Lab, Tencent
[3] College of Information Sciences and Technology, Pennsylvania State University

Autonomous driving system is SAFETY-CRITICAL!

May 2020: Tesla on Autopilot Crashes into Overturned Truck on Busy Highway in Taiwan

https://www.youtube.com/watch?v=X3hrKnv0dPQ

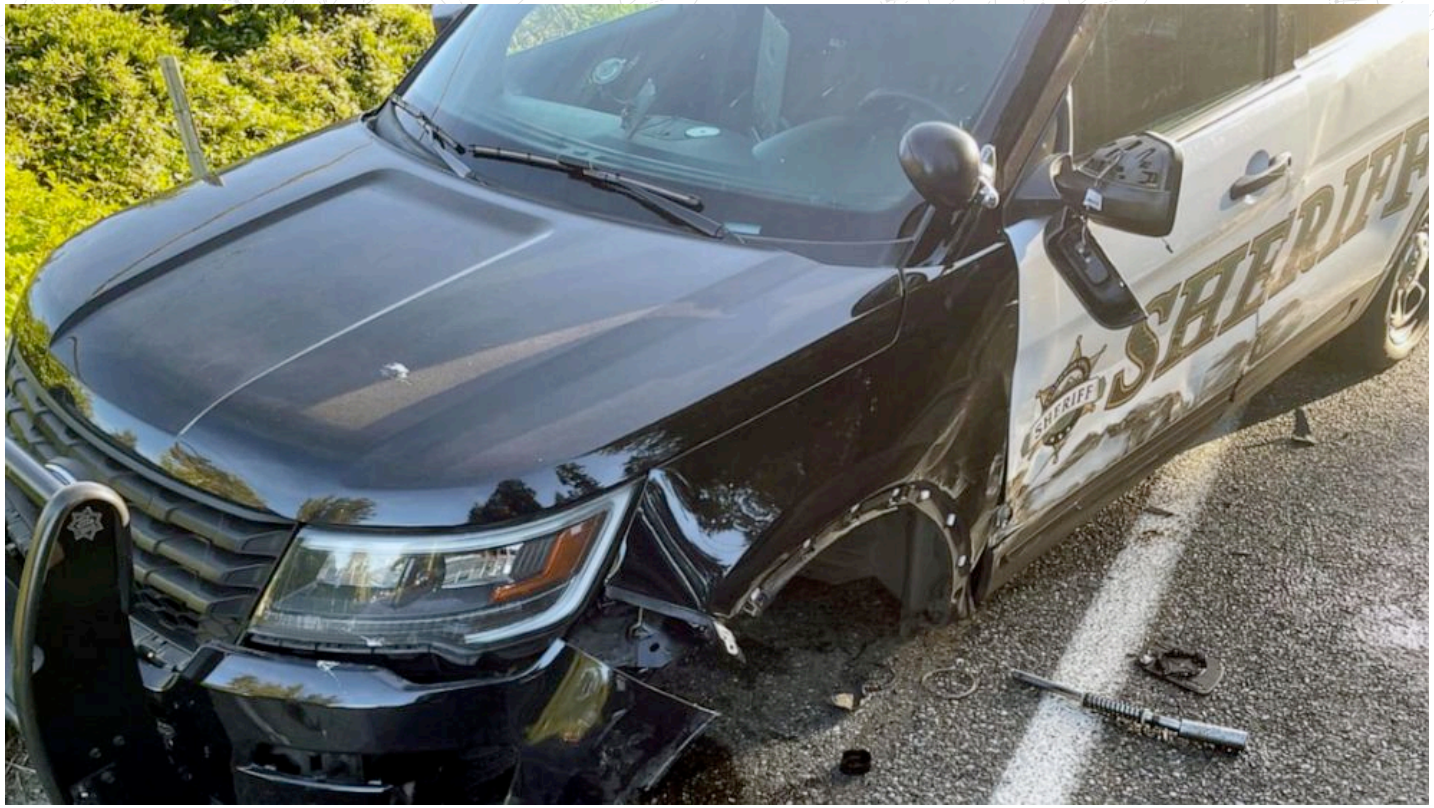
Autonomous driving system is SAFETY-CRITICAL!



April 2021: 2 Killed in Driverless Tesla Car Crash

https://www.nytimes.com/2021/04/18/business/tesla-fatal-crash-texas.html

Autonomous driving system is SAFETY-CRITICAL!



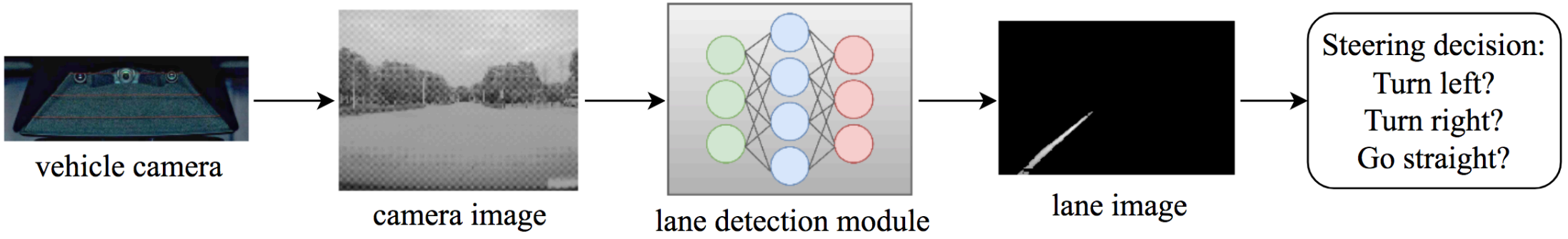
May 2021: Tesla in Autopilot mode crashes into parked police car

https://abcnews.go.com/Business/tesla-autopilot-mode-crashes-parked-police-car/story?id=77753735

Outline

- **Background**
- Two-stage attack
- Evaluation
- Conclusion

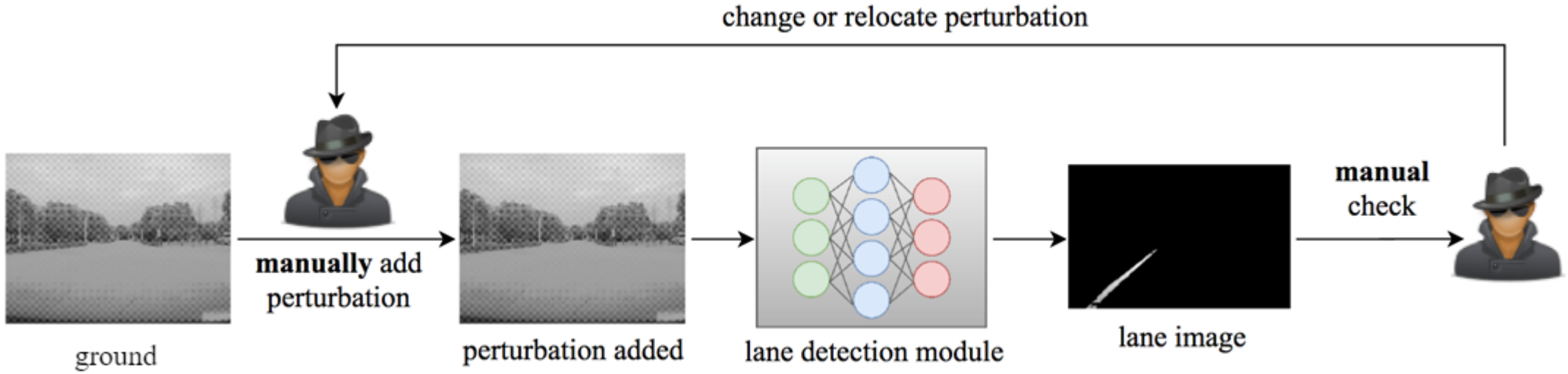
Lane Detection



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



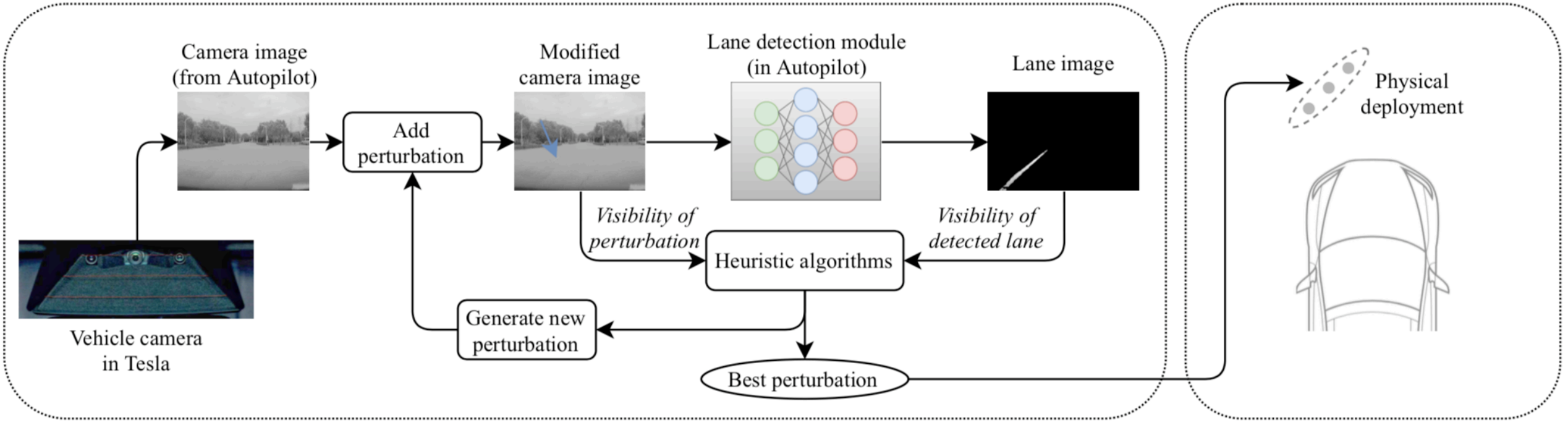
Add perturbations and check whether the module will be affected ***Manually***. If not, the perturbation should be changed or relocated.

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

Outline

- Backgrounds
- **Two-stage attack**
- Evaluation
- Conclusion

Two-stage Attack

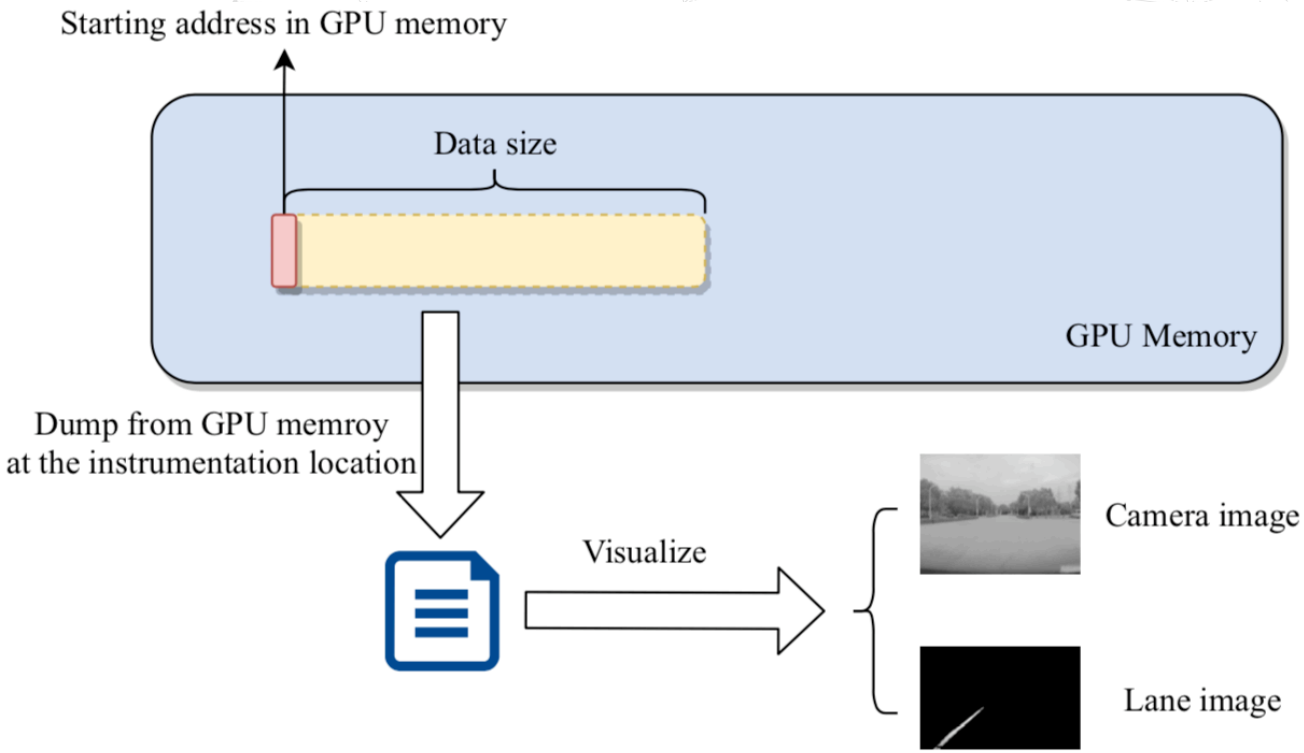


- Stage 1: (1) Add the perturbation to the **camera image** to trigger the lane detection module to generate the corresponding **lane image**.
  (2) Formulate an optimization problem based on the **visibility** of perturbation and that of detected lane and adopt **heuristic algorithms** to find the best 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?

- *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.



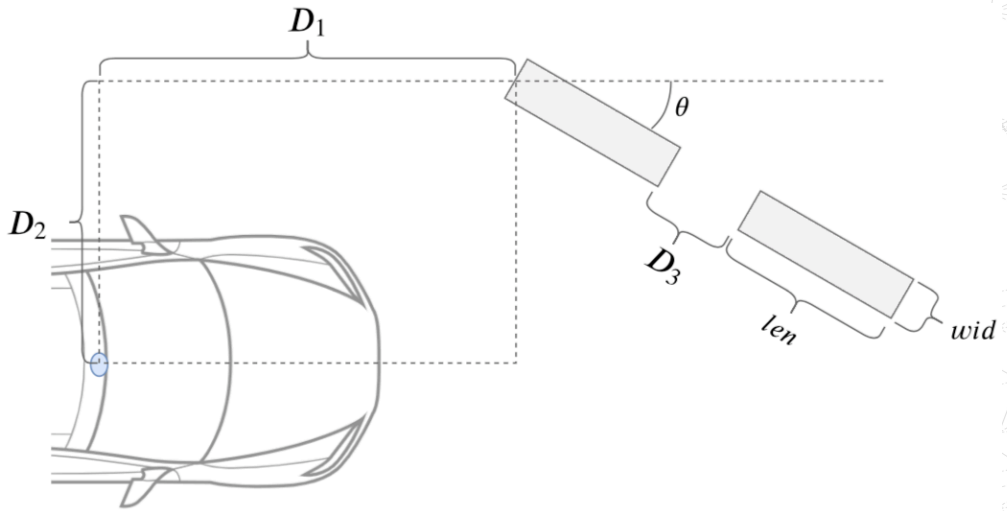
Dumping and visualizing the target data from the GPU on Autopilot

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
ΔG	Increment of grayscale value of the perturbed pixels
θ	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:

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

Outline

- Backgrounds
- Two-stage attack
- **Evaluation**
- Conclusion

Evaluation

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

Q2: Can we misguide the vehicle in physical world?

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

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

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

Q6: How effective is the attack in physical world?

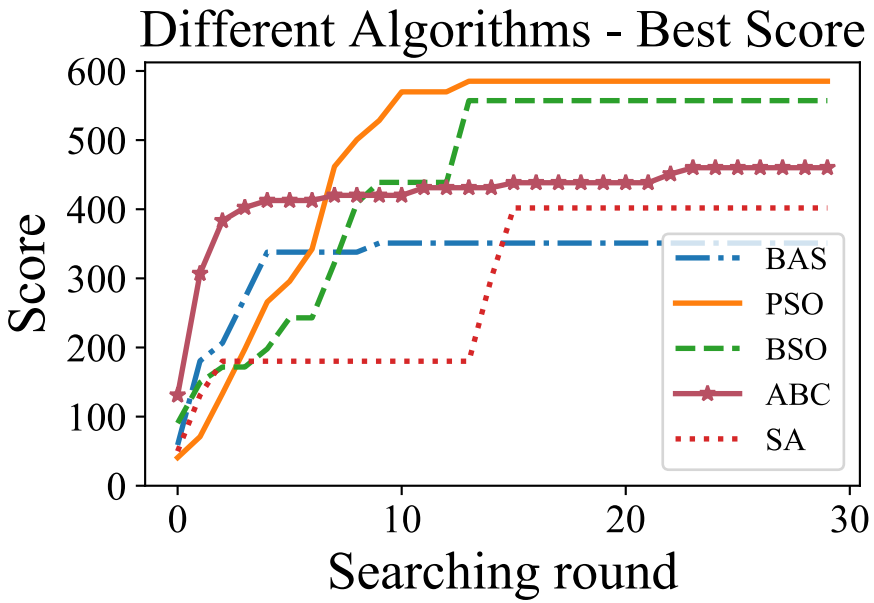
Evaluation – Q1

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

Approach: We 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)*

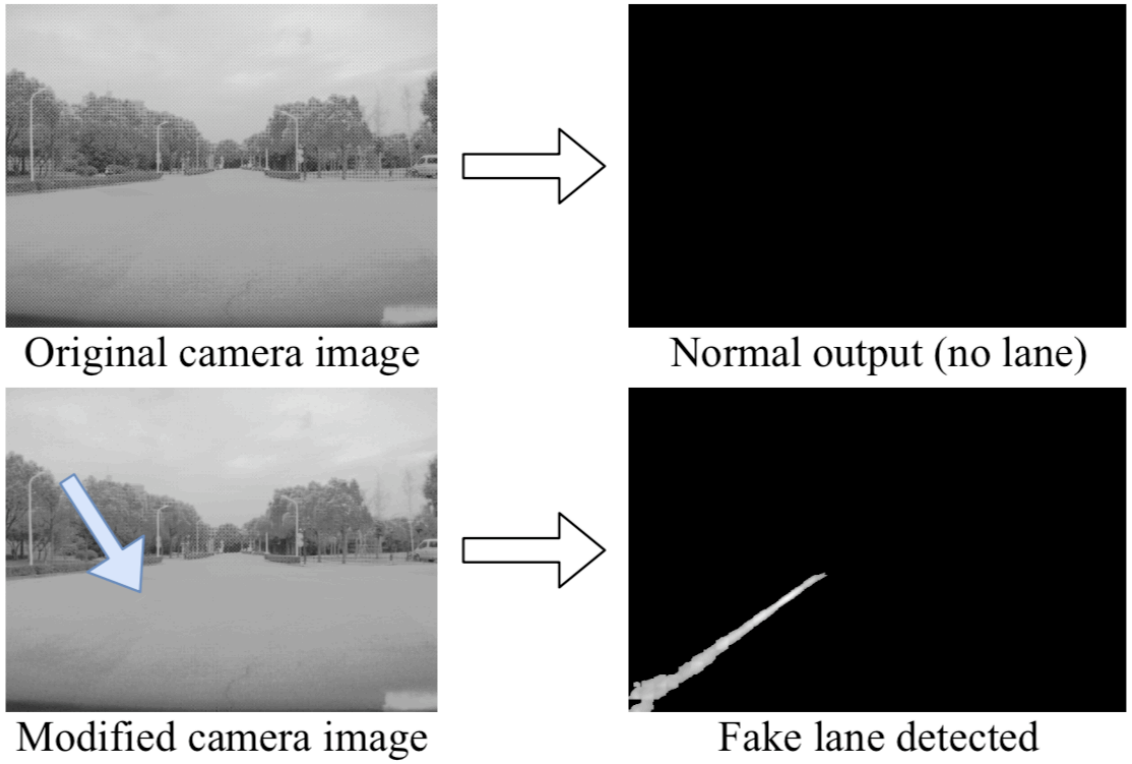
Answer: ***PSO*** is the most efficient one and thus we use it in other experiments.



Comparison between 5 algorithms

Evaluation – Q1

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

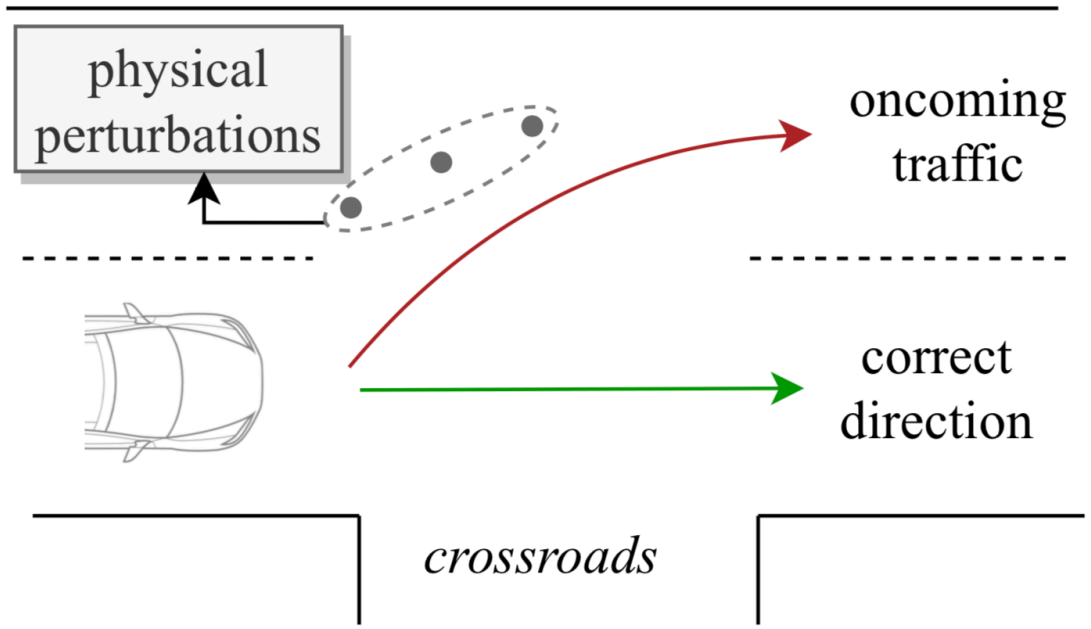
The perturbation can be hardly noticed

Evaluation – Q2

Q2: Can we misguide the vehicle in physical world?

Approach: We 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

Evaluation – Q2: Video

- Please note that the the vehicle is in **auto-steer** mode when recording.

Outline

- Backgrounds
- Two-stage attack
- Evaluation
- **Conclusion**

Conclusion

- Conduct the first investigation on the security of the lane detection module in a **real vehicle** and reveal that its **sensitivity** can be exploited to generate fake lanes and consequently **mislead** the vehicle.
- Propose a novel two-stage approach to generate the optimal perturbations against the lane detection module.
- Conduct extensive experiments on a Tesla vehicle to evaluate our approach. The experimental results show that the lane detection module in Tesla Autopilot is vulnerable to our attack.
- Our future works includes assessments on other autonomous driving systems (e.g., Apollo [1] and Openpilot [2]), and other attack methods (e.g., erasing the existing lane).

[1] Apollo autonomous driving. https://github.com/ ApolloAuto/apollo.
[2] Openpilot autonomous driving. https://github.com/ commaai/openpilot.

Thanks!