EE515 Security of Emerging Systems

Too Good to Be Safe:

Tricking lane detection in autonomous driving with crafted perturbations

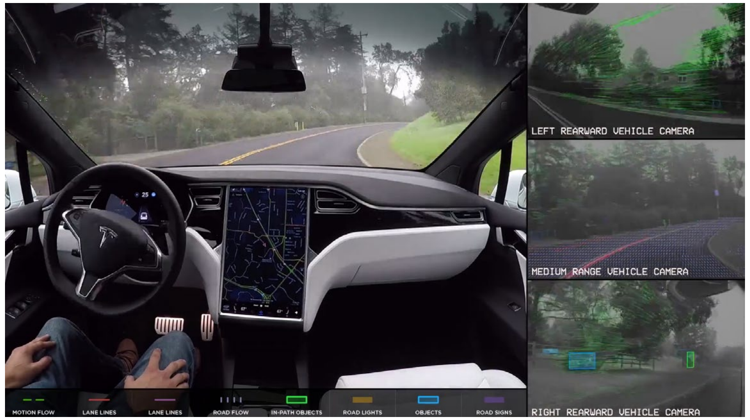
P. Jing, Q. Tang, Y. Du, L. Xue, X. Luo, T. Wang, S. Wu

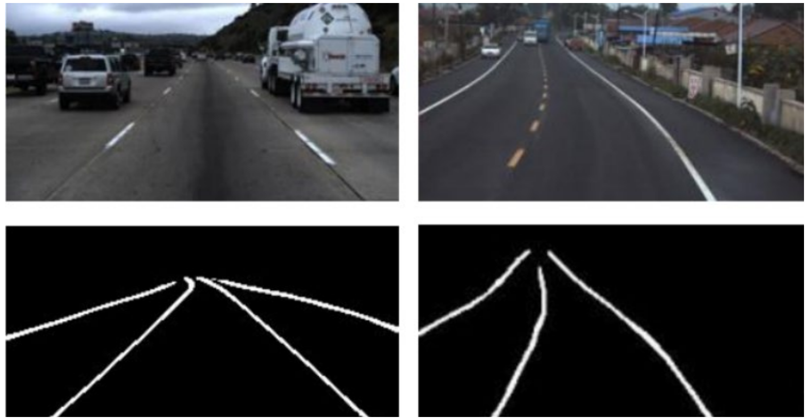
USENIX Security ’21

20225378 WEONJI CHOI

Introduction

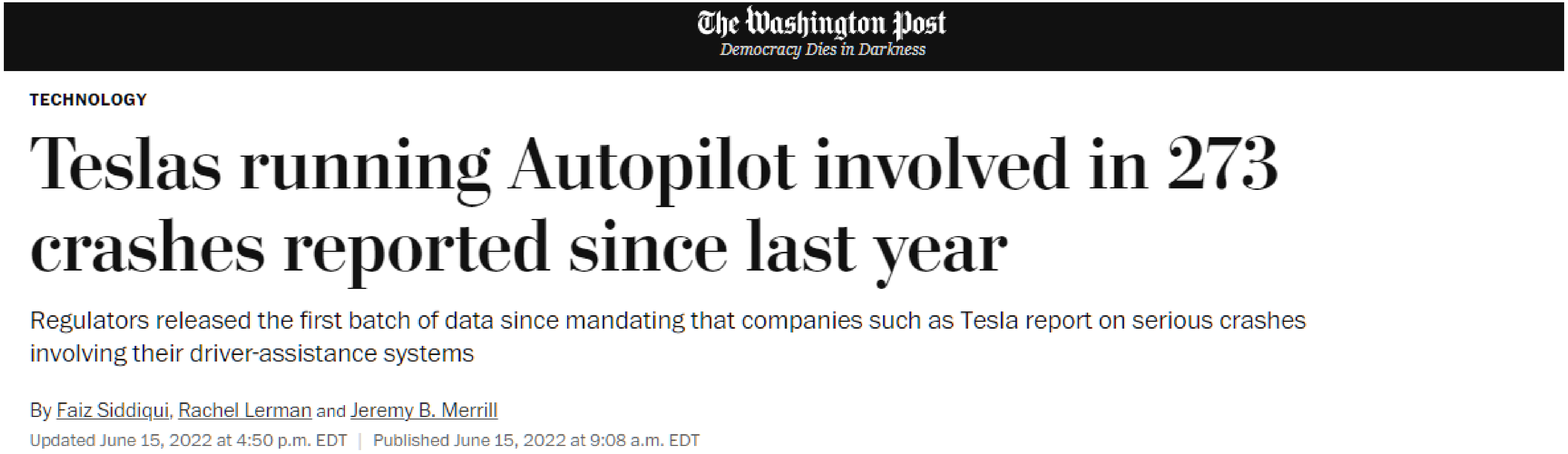
- **Goal : Changing the lane detection result to misguide the autonomous vehicle**
  - Target system & service: **Tesla autopilot's lane detection module** (in auto steering mode)
  - How:
    - **Reverse engineering on the firmware**
    - **Use a fake lane as a perturbation**


LEFT REARWARD VEHICLE CAMERA
MEDIUM RANGE VEHICLE CAMERA
RIGHT REARWARD VEHICLE CAMERA
MOTION FLOW
LANE LINES
LANE LINES
ROAD FLOW
IN-PATH OBJECTS
ROAD LIGHTS
OBJECTS
ROAD SIGNS



Background

- **Autonomous driving systems is SAFETY-CRITICAL!**
  - Tesla autonomous vehicle accidents



The Washington Post
Democracy Dies in Darkness

TECHNOLOGY

Teslas running Autopilot involved in 273 crashes reported since last year

Regulators released the first batch of data since mandating that companies such as Tesla report on serious crashes involving their driver-assistance systems

By Faiz Siddiqui, Rachel Lerman and Jeremy B. Merrill
Updated June 15, 2022 at 4:50 p.m. EDT | Published June 15, 2022 at 9:08 a.m. EDT

Background

- **Autonomous driving systems is SAFETY-CRITICAL!**
  - Tesla autonomous vehicle accidents

May 2020

April 2021

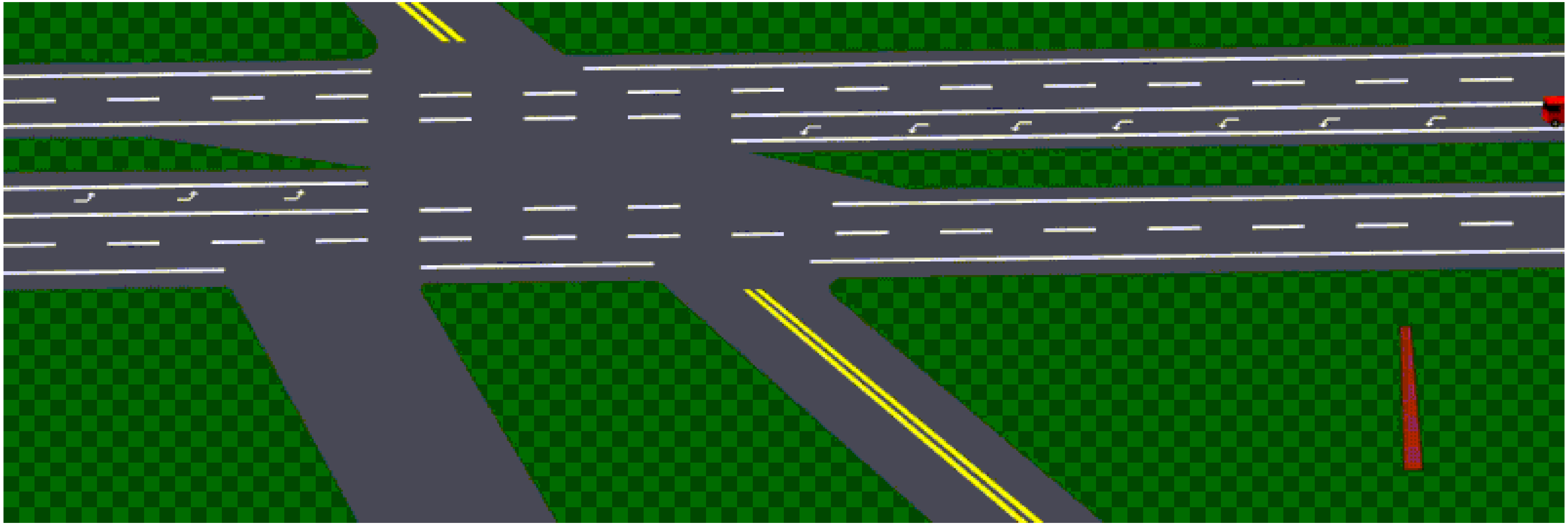
May 2021

More than 5 fatal accidents in 2023

The vehicle was allegedly driving in Autopilot mode when it hit the overturned truck.
odn

Background

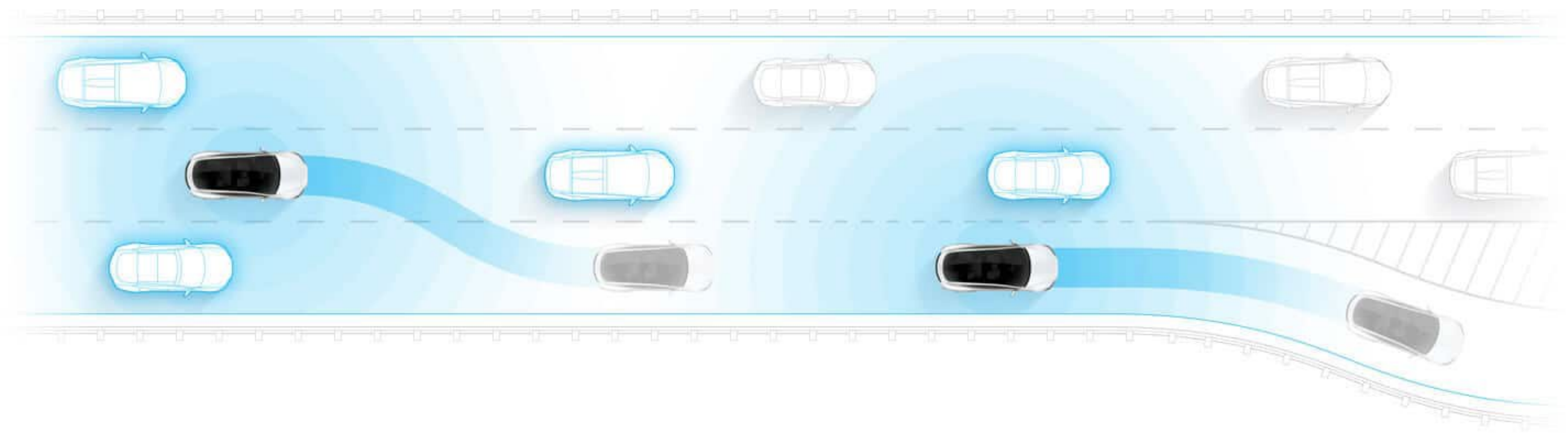
- **Autonomous driving systems is SAFETY-CRITICAL!**
  - Tesla autonomous vehicle accidents[1)]

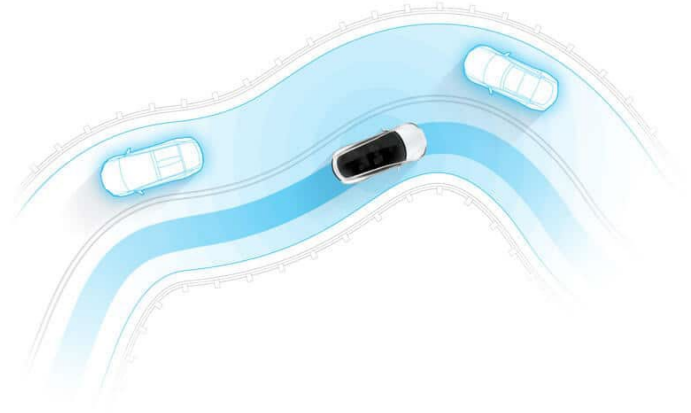


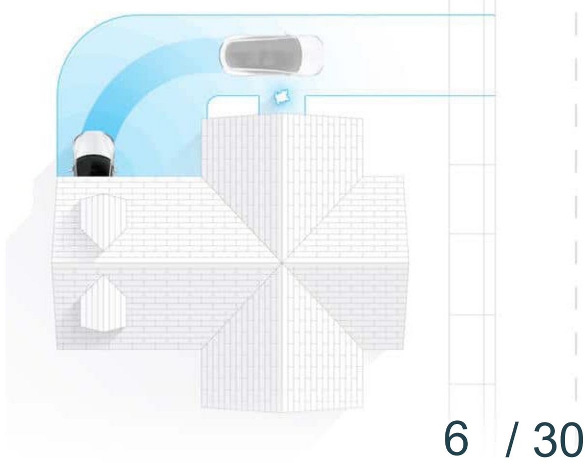
[1)] https://en.wikipedia.org/wiki/Tesla_Autopilot

Background

- **Autopilot** : A system used to control an vehicle
- **Tesla's autopilot for lane changing**
  - Lane changes to optimize the route, and make adjustments
  - Automatic steering

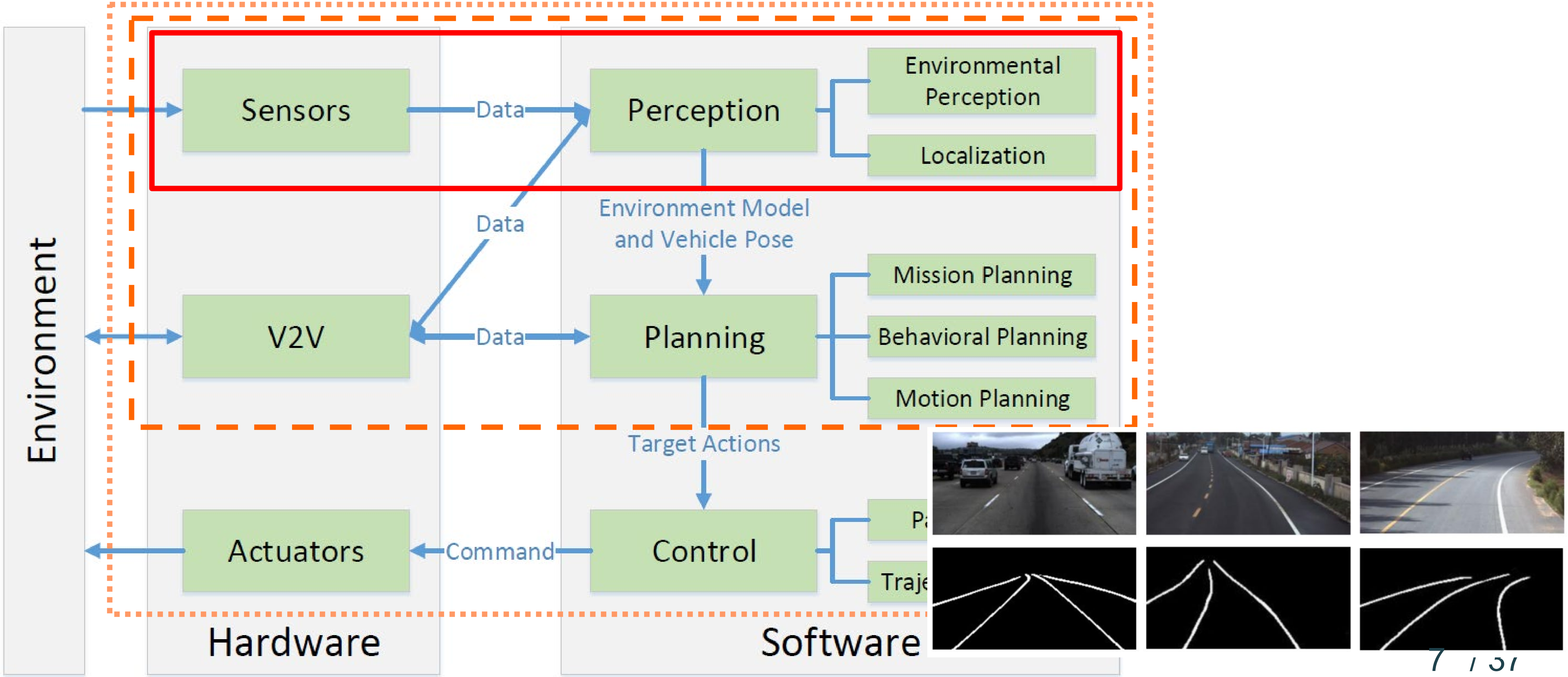






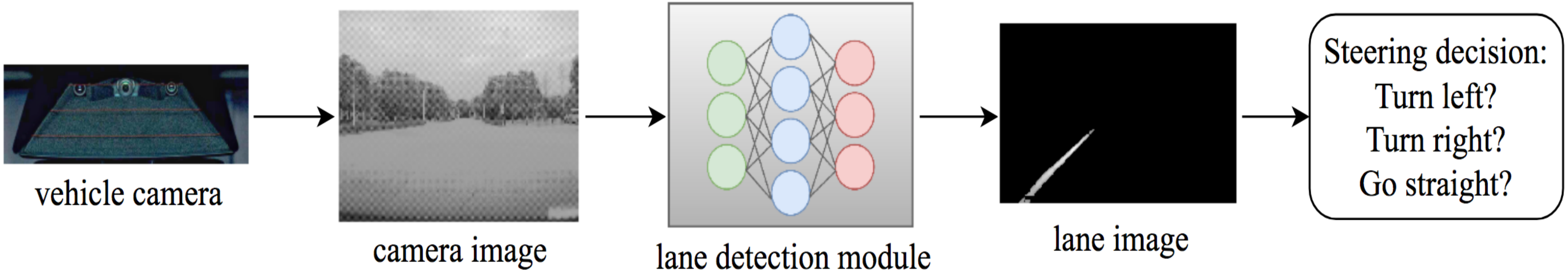
Background

- **Autonomous vehicle system overview**



Background

- **Lane Detection**

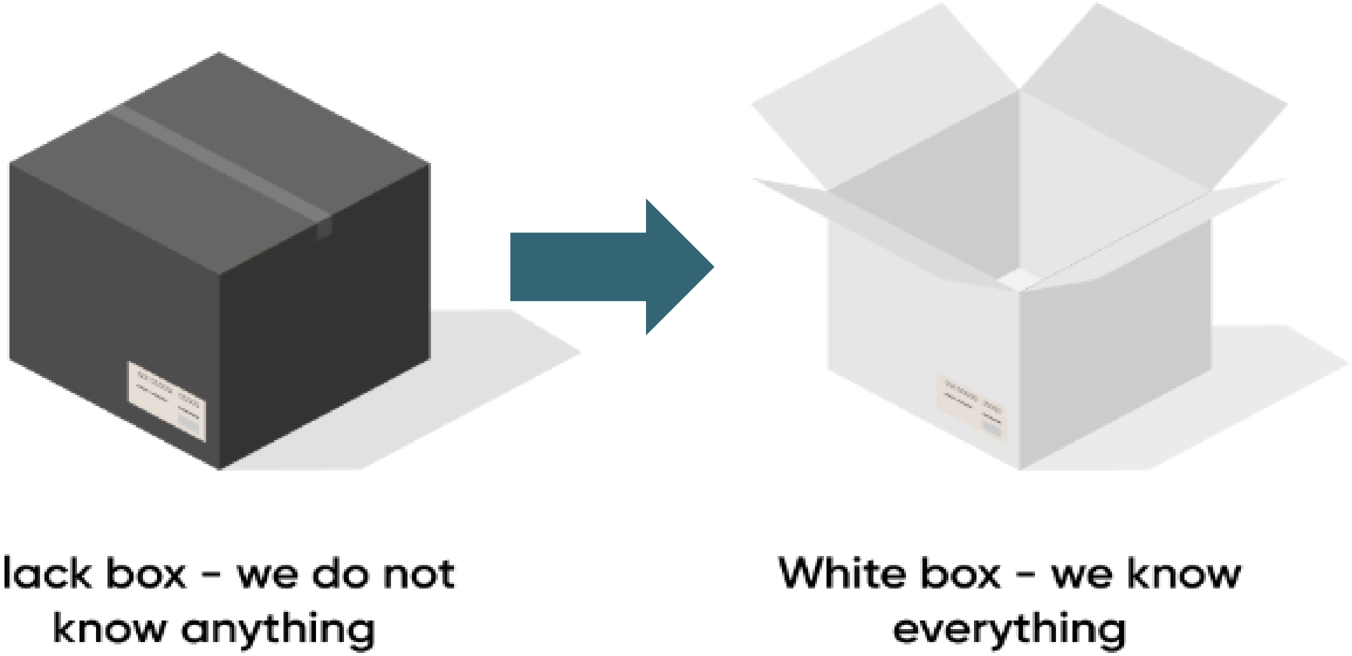


Changing the lane detection result can affect the steering decision.

(i.e., exploiting its over-sensitivity to create a fake lane!)

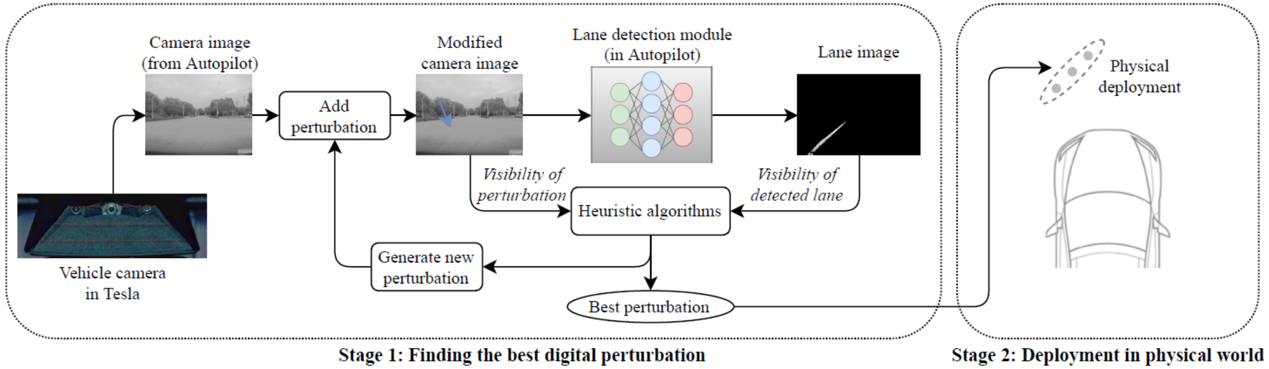
Background

- **Reverse engineering**
  - The process of opening up or dissecting a system to see how it works


Black box - we do not know anything
White box - we know everything

Contributions

- Reverse engineering on the firmware of Tesla Autopilot
- Two-stage approach to generate the optimal perturbations
- Extensive experiments on a Tesla vehicle (Tesla Model S)

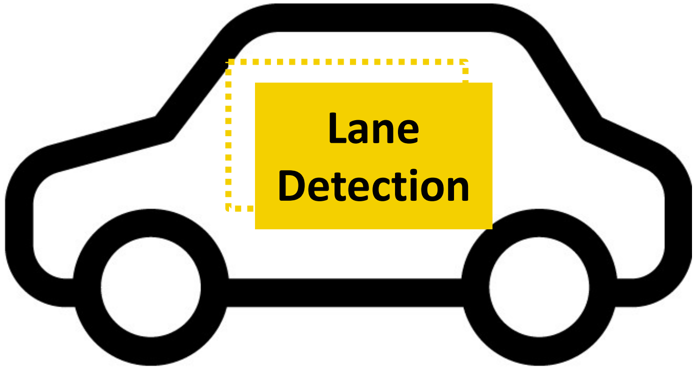


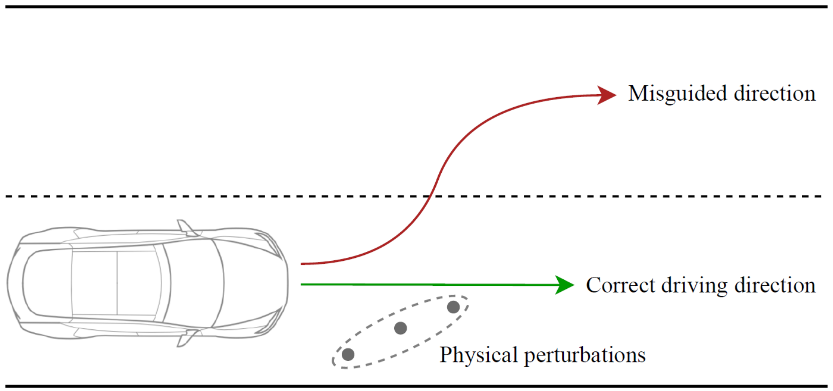
Threat Model

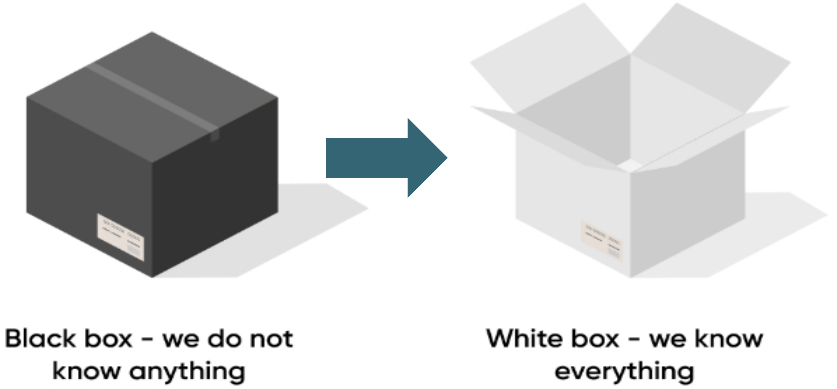
- Attacker has an autonomous vehicle with identical lane detection module.
- Attacker aims to add unobtrusive marking on the ground. (keep change the position and shape)

  -> very labor-intensive and error-prone

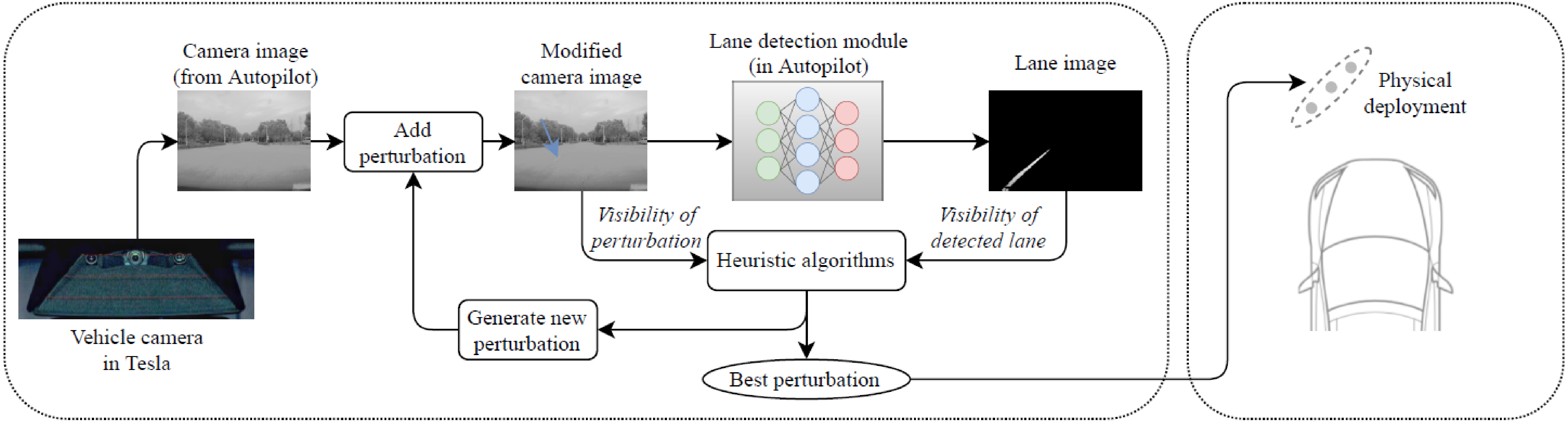
  -> better to be done in digital world <- **Reverse Engineering!**






Black box - we do not know anything
White box - we know everything

Two-stage attack



Stage 1: Finding the best digital perturbation

Stage 2: Deployment in physical world

Challenges

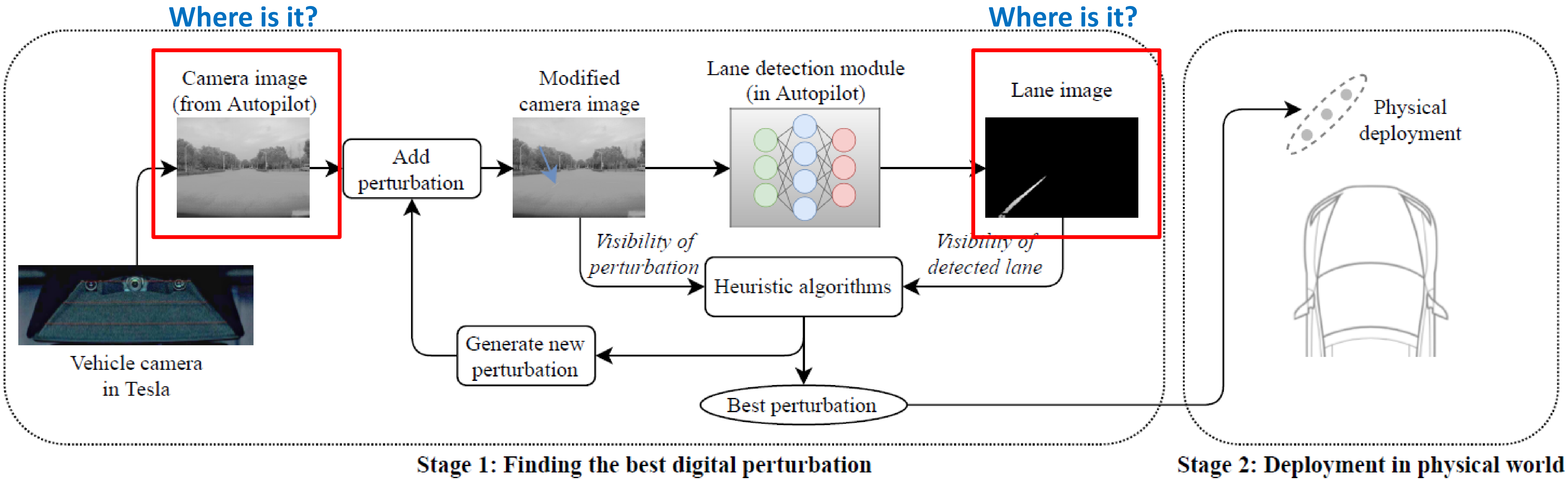
- C1. How to locate the input camera image and output lane image in the vehicle?

- C2. How to add perturbations?

- C3. How to find the best perturbations?

S1. Accessing Data in Tesla Autopilot

C1. How to locate the input camera image and the corresponding output lane image in the vehicle?



Where is it?
Where is it?
Camera image (from Autopilot)
Add perturbation
Modified camera image
Lane detection module (in Autopilot)
Lane image
Physical deployment
Vehicle camera in Tesla
Visibility of perturbation
Heuristic algorithms
Visibility of detected lane
Generate new perturbation
Best perturbation
Stage 1: Finding the best digital perturbation
Stage 2: Deployment in physical world

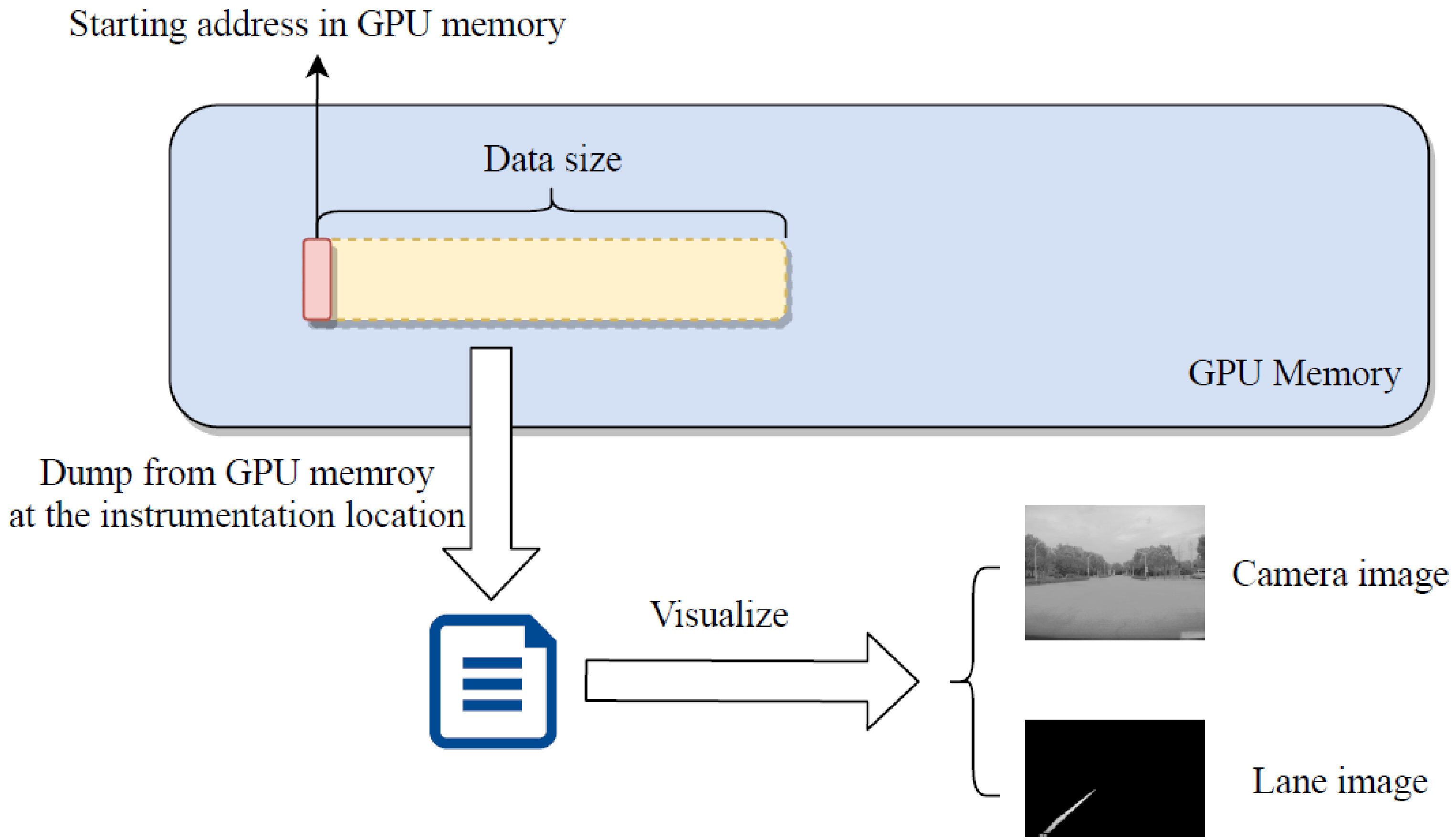
S1. Accessing Data in Tesla Autopilot

C1. How to locate the input camera image and the corresponding output lane image in the vehicle?

- **Firmware under examination**
  - Tesla Model S 75, with the Autopilot hardware version of 2.5 and software version of 2018.6.1.
  - Running an AArch64 Linux OS and uses NVDIA GPU for deep learning computation.
- **CUDA**
  - Memory management functions: cudaMalloc, cudaMemcpy, cudaConfigurecall
- **Static and dynamic analysis**
  - Find **(1) source address, (2) destination address, (3) data size, and (4) mode of transfer**

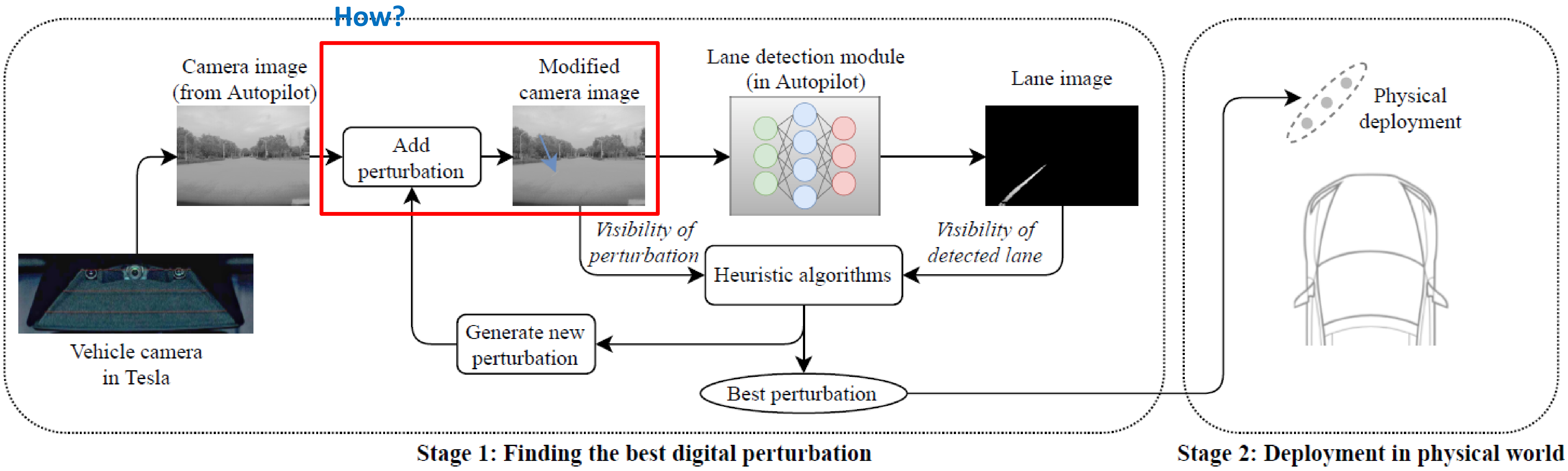
S1. Accessing Data in Tesla Autopilot

C1. How to locate the input camera image and the corresponding output lane image in the vehicle?



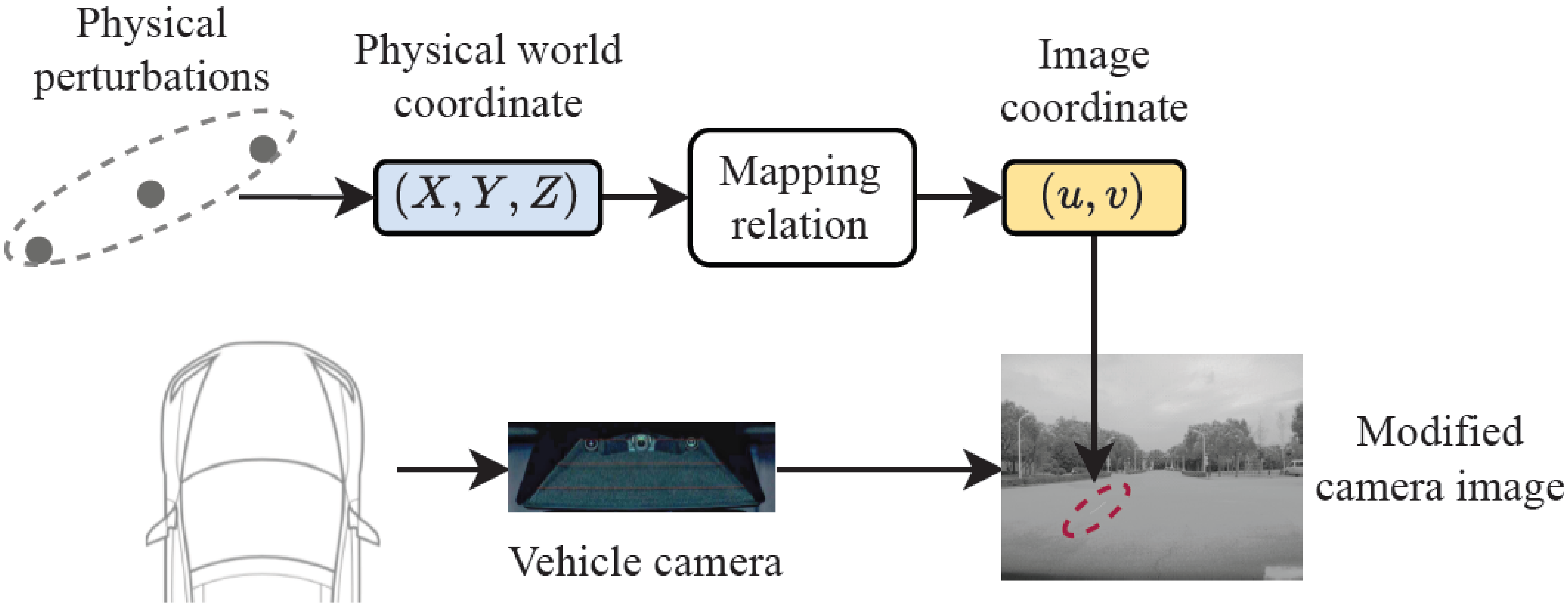
S2. Adding Digital Perturbations

C2. How to add perturbations to input camera image?



S2. Adding Digital Perturbations

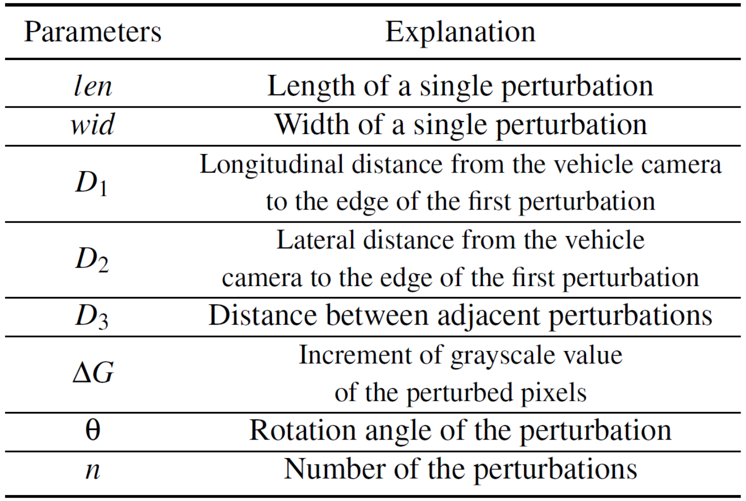
C2. How to add perturbations to input camera image?



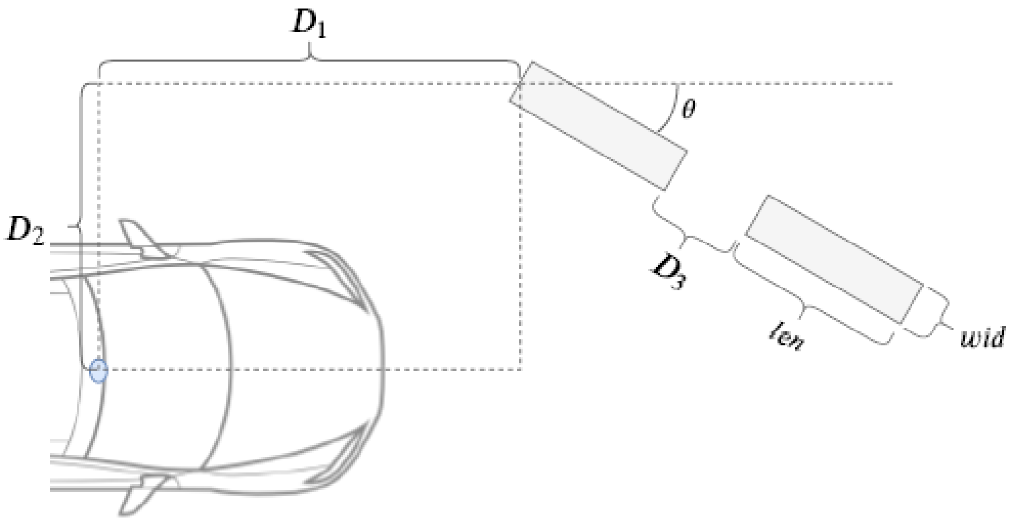
S2. Adding Digital Perturbations

C2. How to add perturbations to input camera image?

- **Project physical world markings**
  - Map a physical world coordinate (X, Y, Z) -> image coordinate *(u, v)*
  - Modifying the grayscale value of the corresponding pixels
- **Parameterized perturbations**

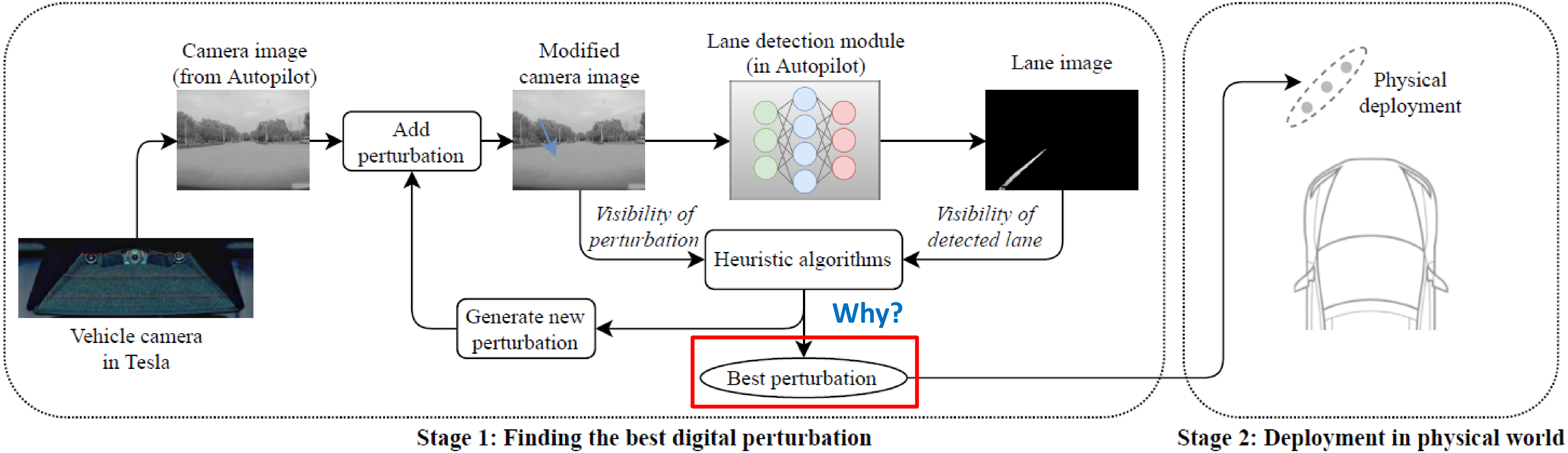


Parameters	Explanation
len	Length of a single perturbation
wid	Width of a single perturbation
D_1	Longitudinal distance from the vehicle camera to the edge of the first perturbation
D_2	Lateral distance from the vehicle camera to the edge of the first perturbation
D_3	Distance between adjacent perturbations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



S3. Finding the Best Perturbations

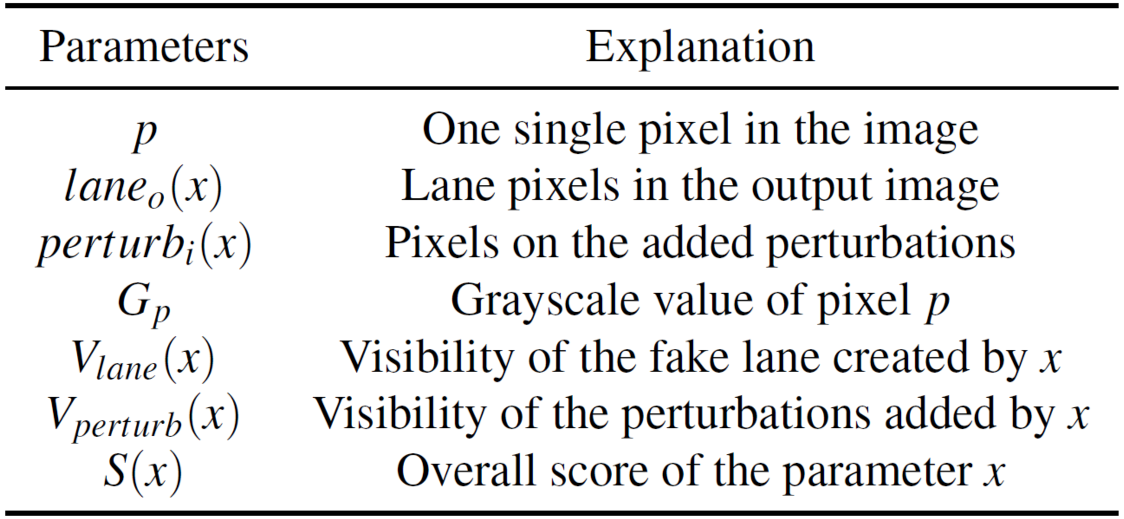
C3. How to find the best perturbations?



S3. Finding the Best Perturbations

C3. How to find the best perturbations?

- **Quality of Perturbations: Visibility of lane & Visibility of perturbation**



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

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

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

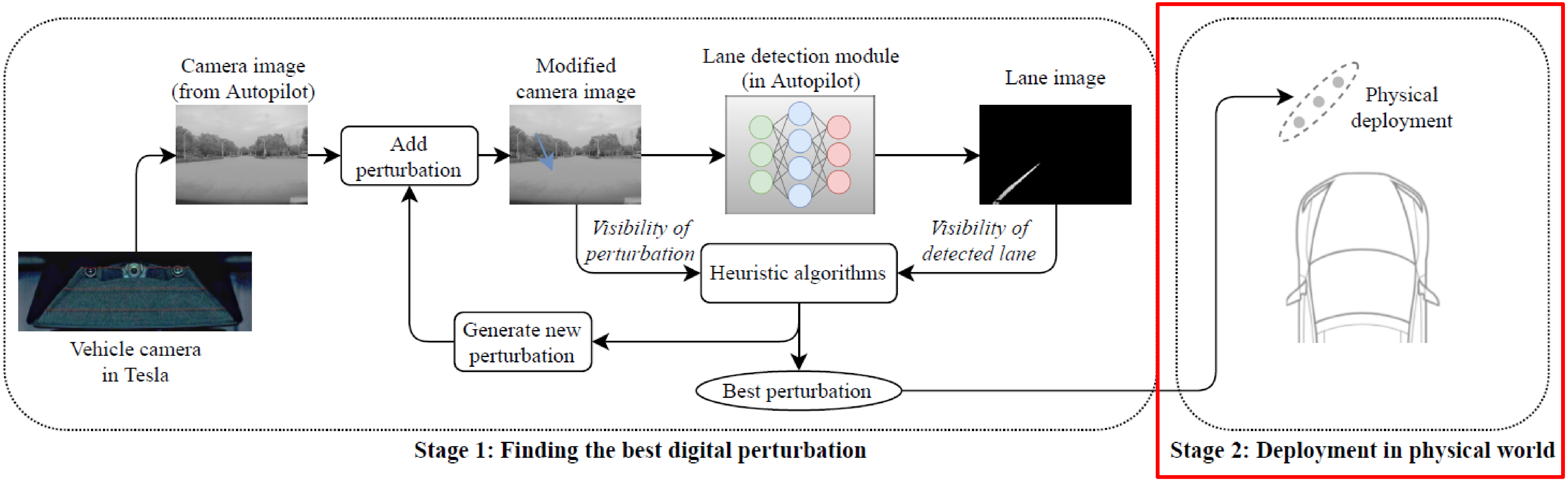
$$S(x) = \frac{V_{lane}(x)}{V_{perturb}(x)}$$

- **Optimization problem:**
    - Heuristic algorithm: BAS, **PSO**, BSO, ABC, SA

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

Evaluation

Let's start!



Evaluation

- RQ1:Efficiency of the heuristic algorithms to find the best perturbation
- RQ2: Effect of the perturbation number and the rotation angle θ
- RQ3: Performance with different camera images
- RQ4: Common characteristics of the best perturbations

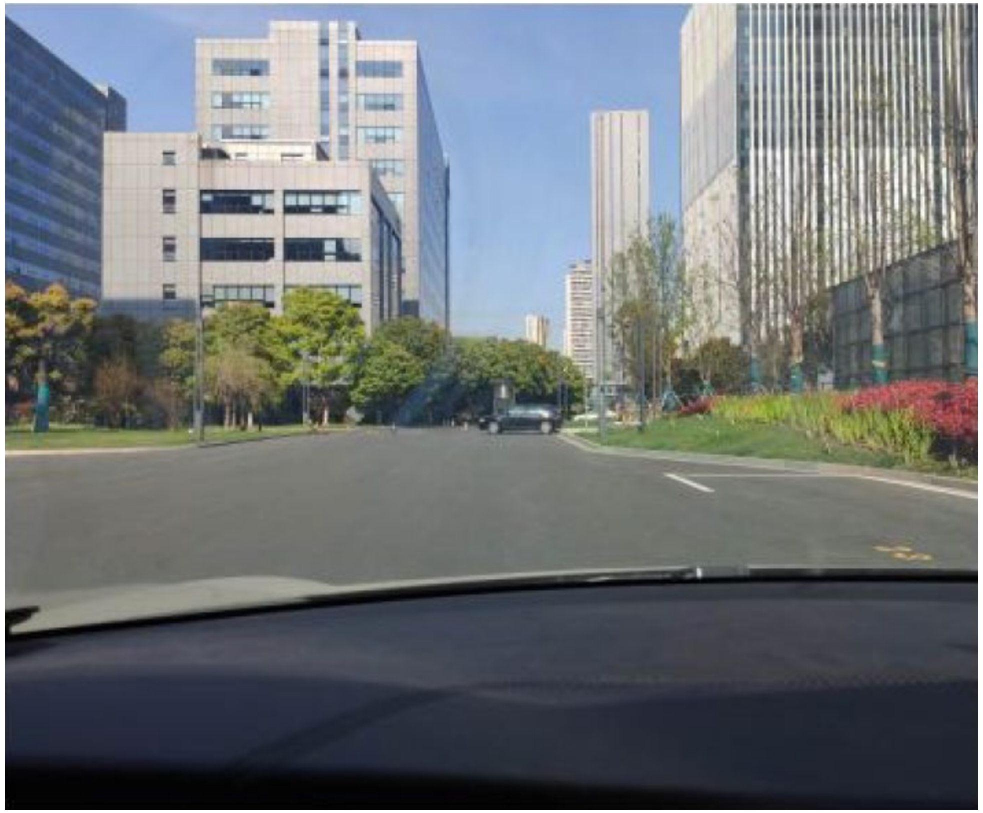
Digital World

- RQ5: Effectiveness of the attack in physical world
- RQ6: Feasibility of the attack in physical world

Physical World

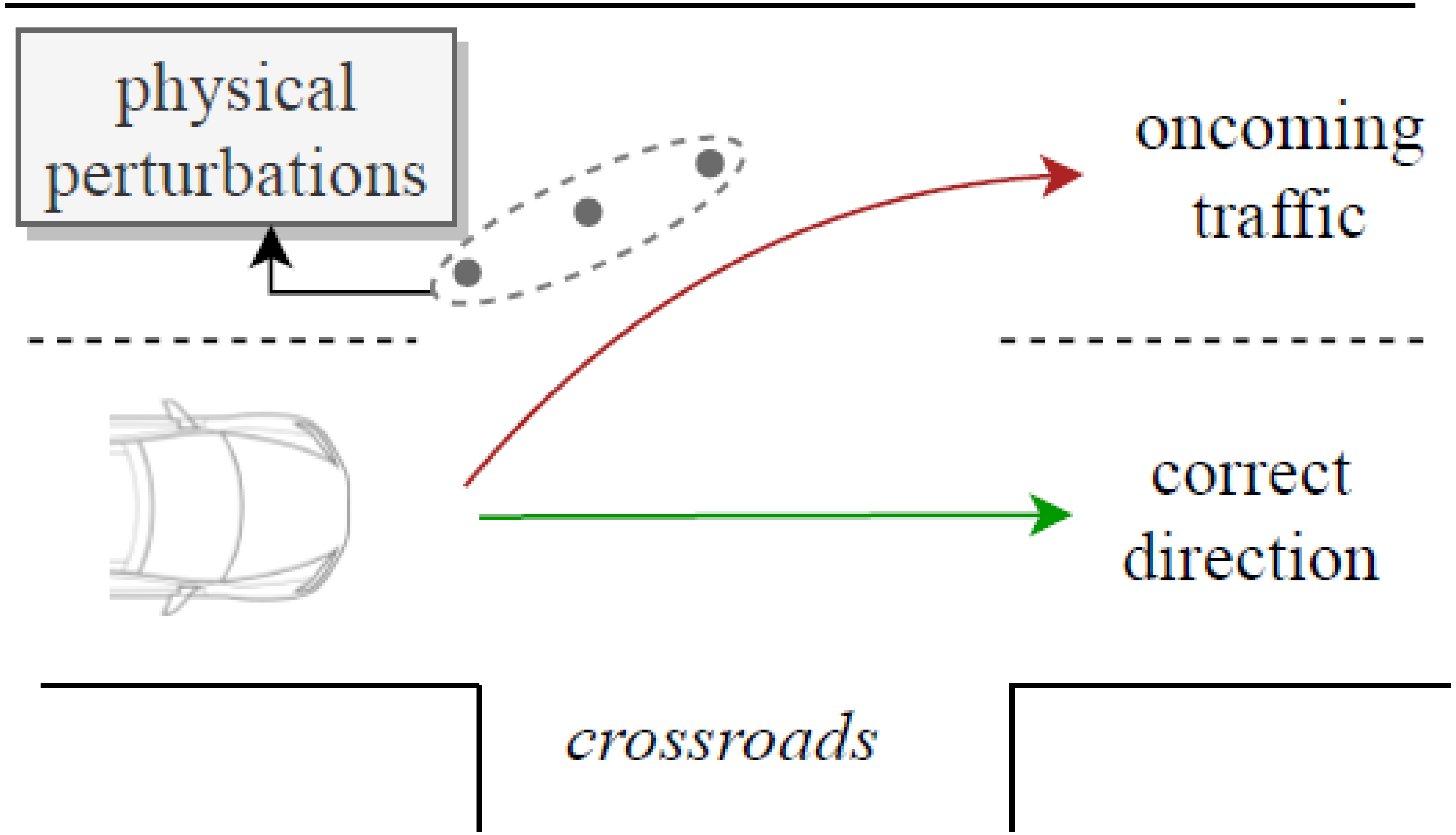
Evaluation

- **RQ5: How effective is the attack in physical world?**
  - Most effective with below conditions
    - Perturbation number: 1 is enough
    - Rotation angle: θ=0 (straight perturbations)
    - Light condition: doesn’t matter
    - Longitudinal Distance: from 15m to 3m



Evaluation

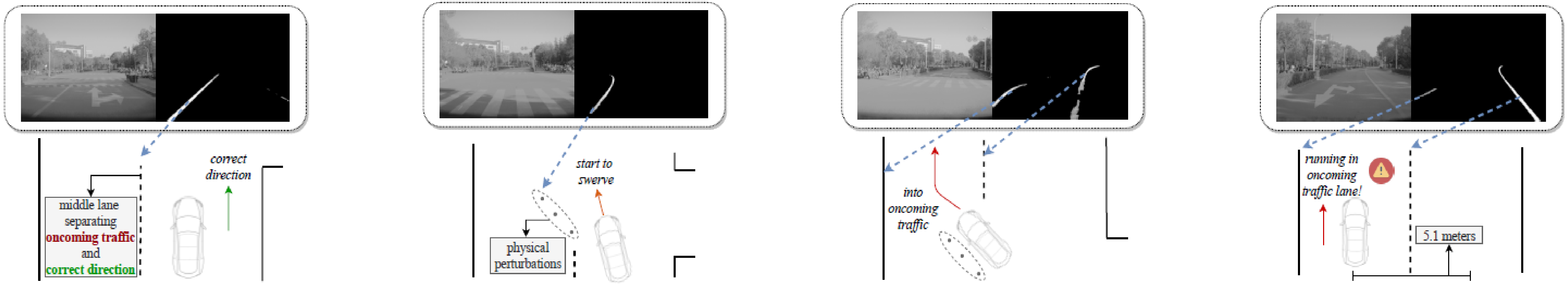
- **RQ6: Can we misguide the vehicle in physical world?**



Evaluation

(Demo Video: https://www.youtube.com/watch?v=aSe2MrjVs)

- **RQ6: Can we misguide the vehicle in physical world?**


correct direction
middle lane separating oncoming traffic and correct direction
start to swerve
physical perturbations
into oncoming traffic
running in oncoming traffic lane!
5.1 meters

(a) Vehicle is running on the correct direction.

(b) Fake lane is detected and vehicle starts to swerve.

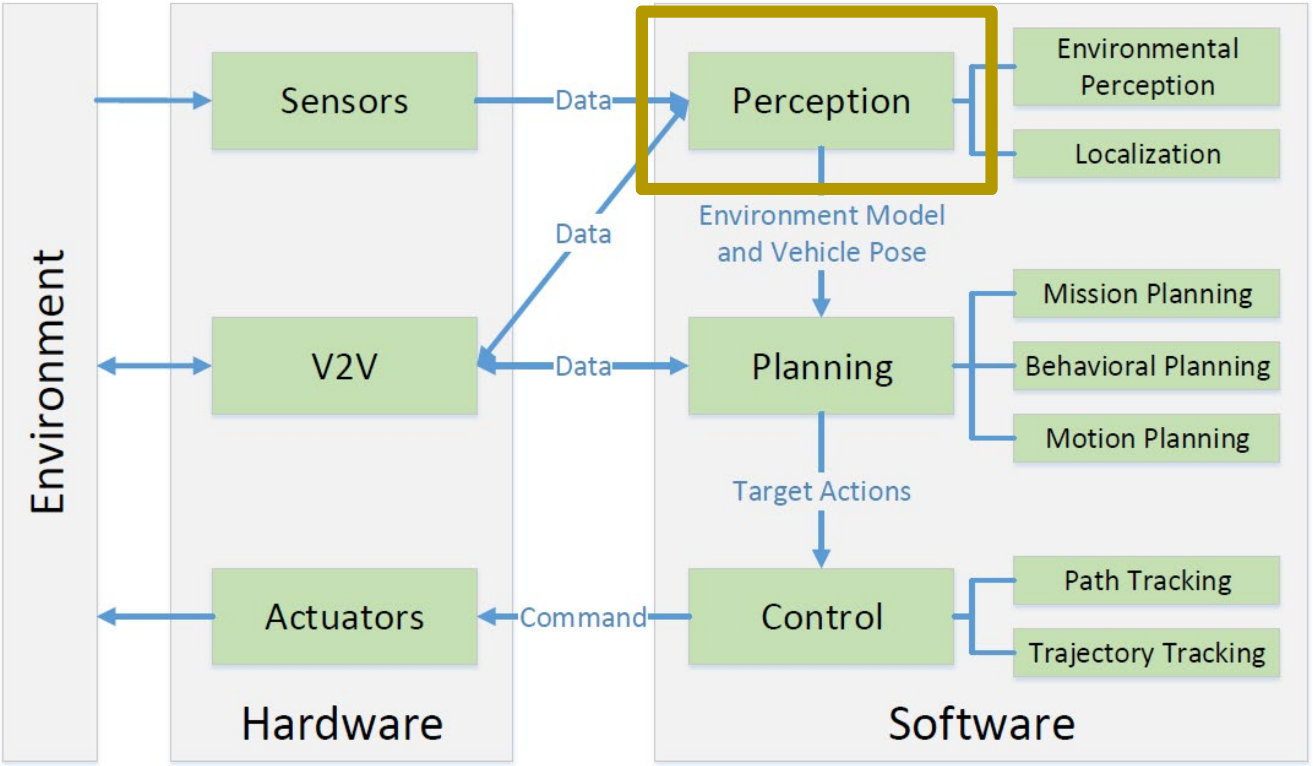
(c) Vehicle follows the fake lane into oncoming traffic.

(d) Vehicle finally runs in the oncoming traffic lane!

- **Demo Video**

Defense

- **Better lane detection module** to distinguish craft perturbations
- **Better control policy** : more considerable elements, multi-sensor fusion

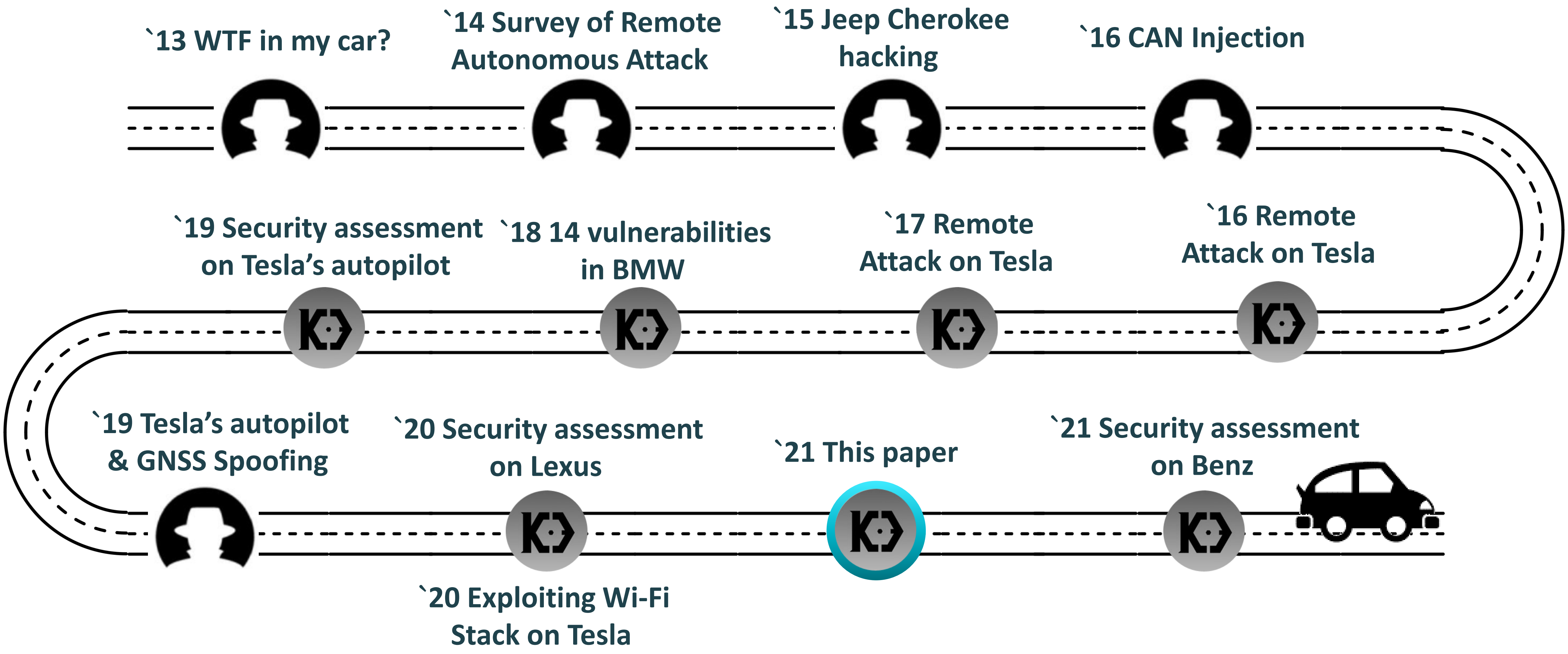


Limitations and Discussion

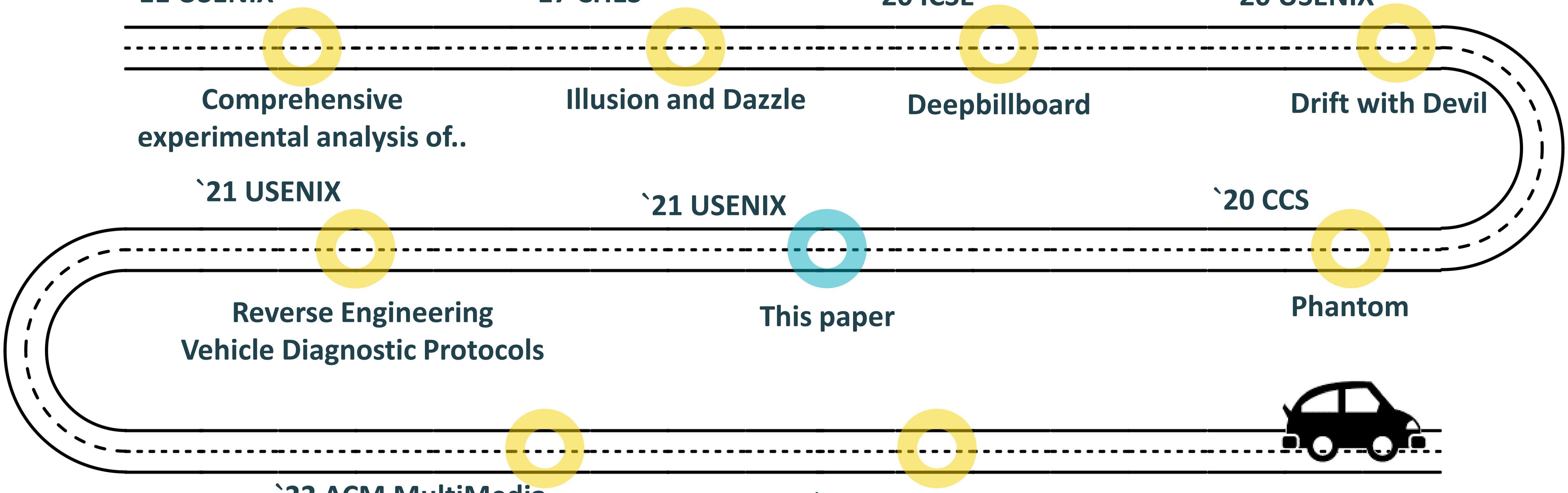
- **Limitation**
  - A physical set up process, and it must be installed at a specific point.
  - Cannot be completely invisible (a driver may notice)

- **Future works**
  - Same vulnerability in other autonomous driving systems (e.g., Apollo, Openpilot, etc.)
  - Launching attacks on real lanes (e.g., dark markings to cover, etc.)

Related Work - Hackings


`13 WTF in my car?
`14 Survey of Remote Autonomous Attack
`15 Jeep Cherokee hacking
`16 CAN Injection
`16 Remote Attack on Tesla
`17 Remote Attack on Tesla
`18 14 vulnerabilities in BMW
`19 Security assessment on Tesla's autopilot
`19 Tesla's autopilot & GNSS Spoofing
`20 Security assessment on Lexus
`20 Exploiting Wi-Fi Stack on Tesla
`21 This paper
`21 Security assessment on Benz

Related Work - Papers



`11 USENIX

Comprehensive experimental analysis of..

`17 CHES

Illusion and Dazzle

`20 ICSE

Deepbillboard

`20 USENIX

Drift with Devil

`20 CCS

Phantom

`21 USENIX

This paper

`21 USENIX

Reverse Engineering Vehicle Diagnostic Protocols

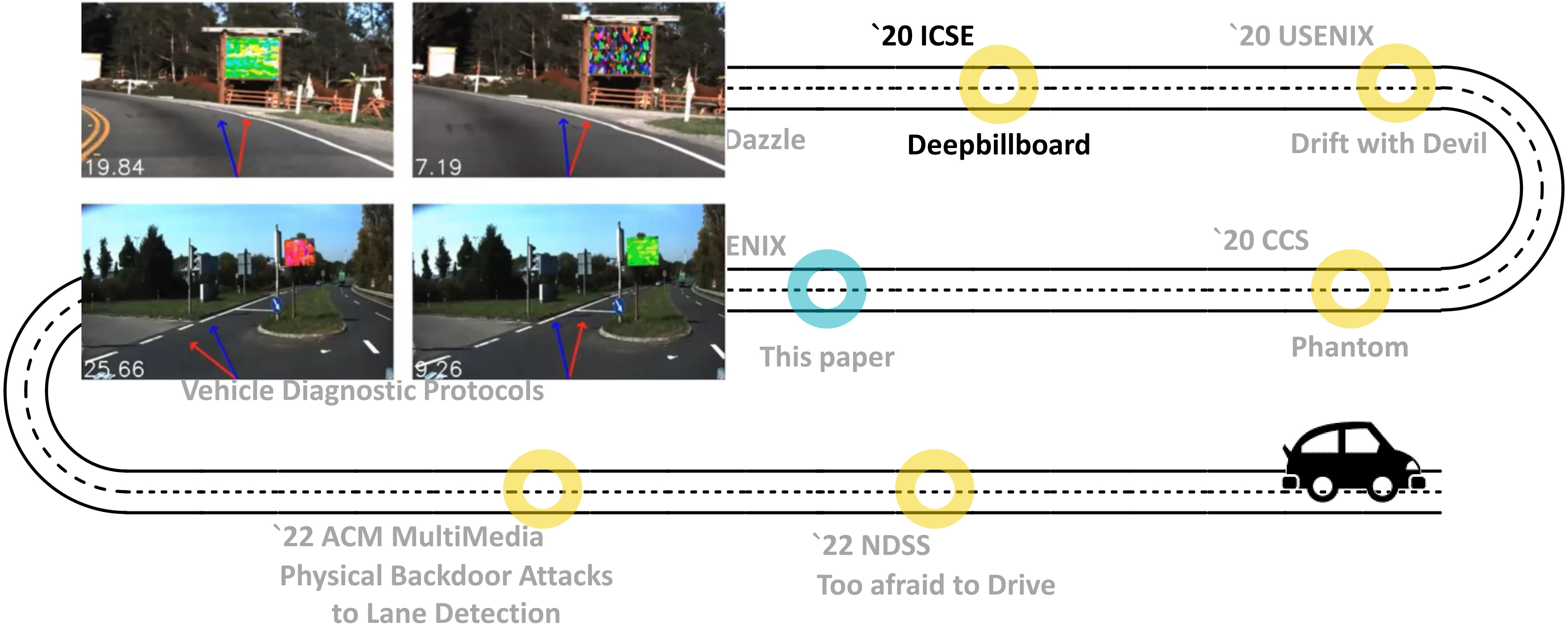
`22 ACM MultiMedia

Physical Backdoor Attacks to Lane Detection

`22 NDSS

Too afraid to Drive

Related Work - Papers



`20 ICSE

Deepbillboard

`20 USENIX

Drift with Devil

Dazzle

ENIX

This paper

`20 CCS

Phantom

Vehicle Diagnostic Protocols

`22 ACM MultiMedia

Physical Backdoor Attacks to Lane Detection

`22 NDSS

Too afraid to Drive

Related Work - Papers

`11 USENIX

Comprehensive experimental analysis of..

`17 CHES

Illusion and Dazzle

`20 ICSE

Deepbillboard

`20 USENIX

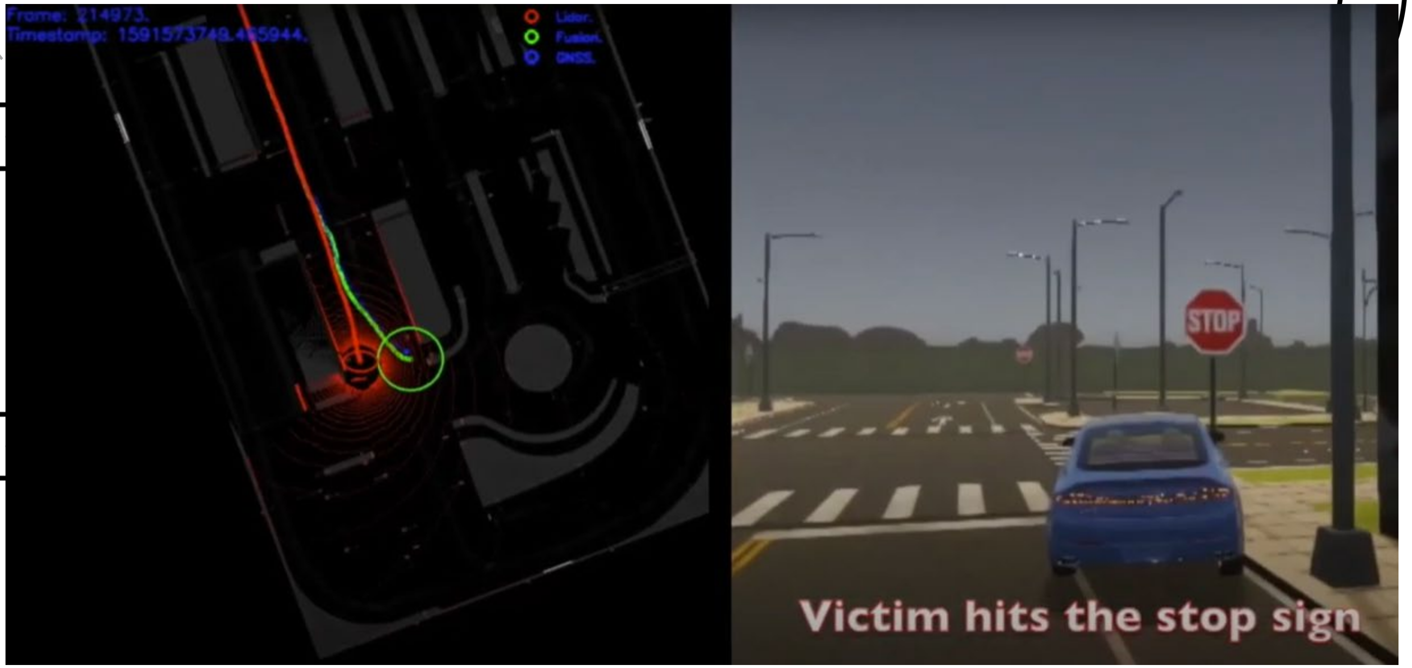
Drift with Devil

`21 USENIX

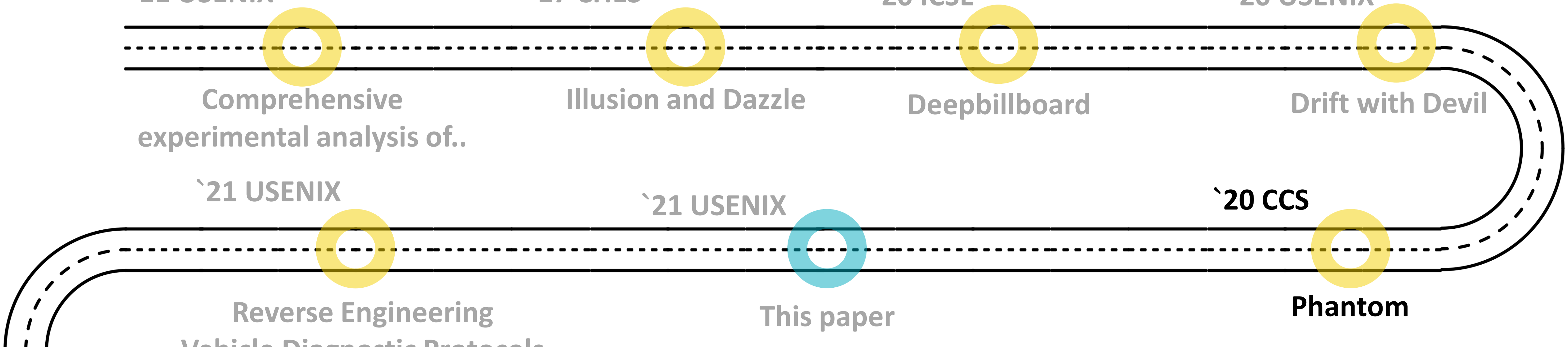
Reverse Engineering Vehicle Diagnostic Protocols

`22 ACM MultiMedia

Physical Backdoor Attacks to Lane Detection



Related Work - Papers



`11 USENIX

Comprehensive experimental analysis of..

`17 CHES

Illusion and Dazzle

`20 ICSE

Deepbillboard

`20 USENIX

Drift with Devil

`20 CCS

Phantom

`21 USENIX

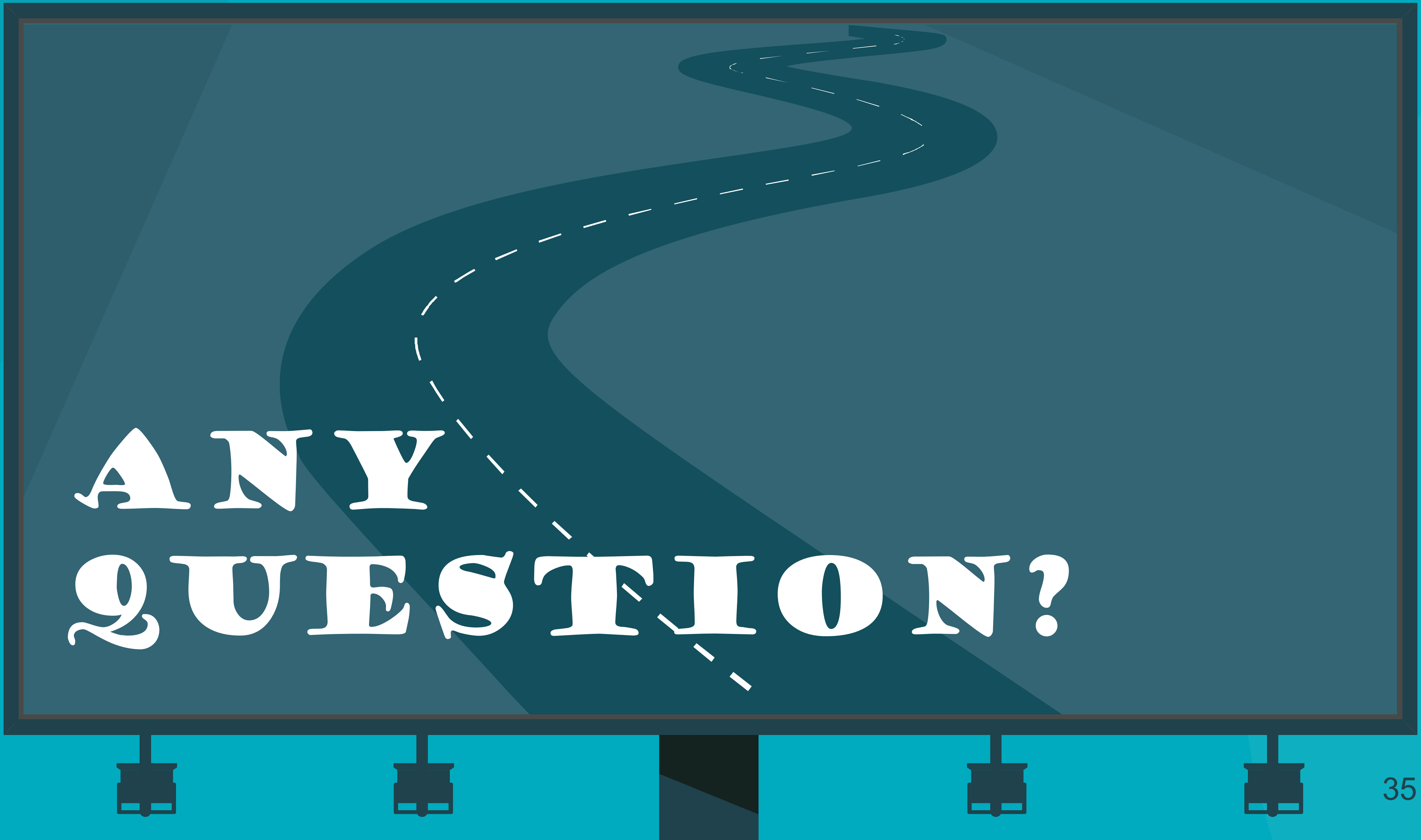
This paper

`21 USENIX

Reverse Engineering Vehicle Diagnostic Protocols

Conclusion

- **Two-stage approach to generate the optimal perturbations**
    - Reverse engineering to access data
    - Misguide the vehicle into oncoming lane
    - Extensive evaluation
- **Need more reliable self-driving system**
    - Safety critical system
    - Standards and policies



ANY QUESTION?