Exorcising “Wraith”: Protecting LiDAR-based Object Detector in Automated Driving System from Appearing Attacks

32ND USENIX
SECURITY SYMPOSIUM

Qifan Xiao, **Xudong Pan**, Yifan Lu, Mi Zhang*, Jiarun Dai, Min Yang*

System and Software Security Lab

School of Computer Science

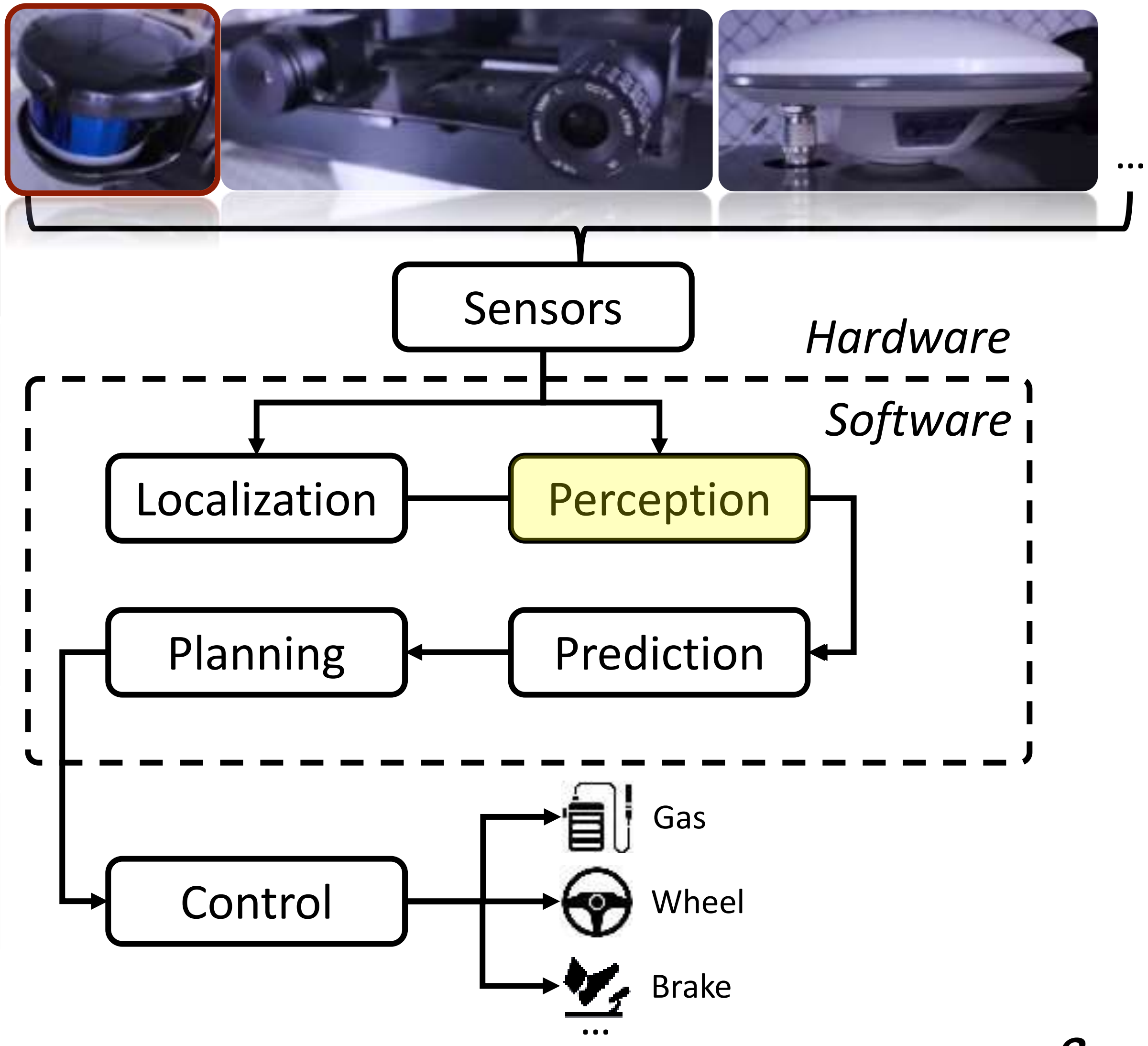
Fudan University

More Research
on AI Security

LiDARs in Automated Driving System

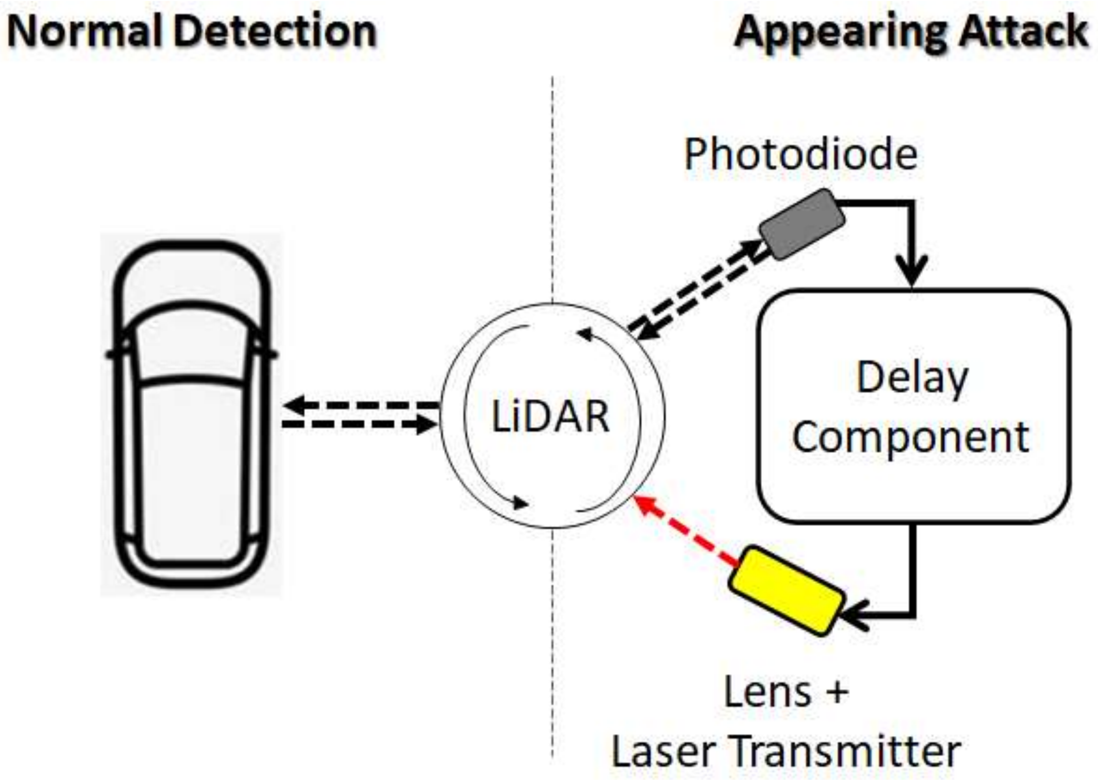
- Most ADS companies take LiDARs as main sensors

ADS Company	LiDAR Type	LiDAR as main sensor?	Open-Source?
apollo	Velodyne	**Y**	**Y**
WAYMO	unknown	**Y**	**N**
TESLA	/	**N**	**N**
Aurora	FirstLight	**Y**	**N**
pony.ai	IRIS	**Y**	**N**

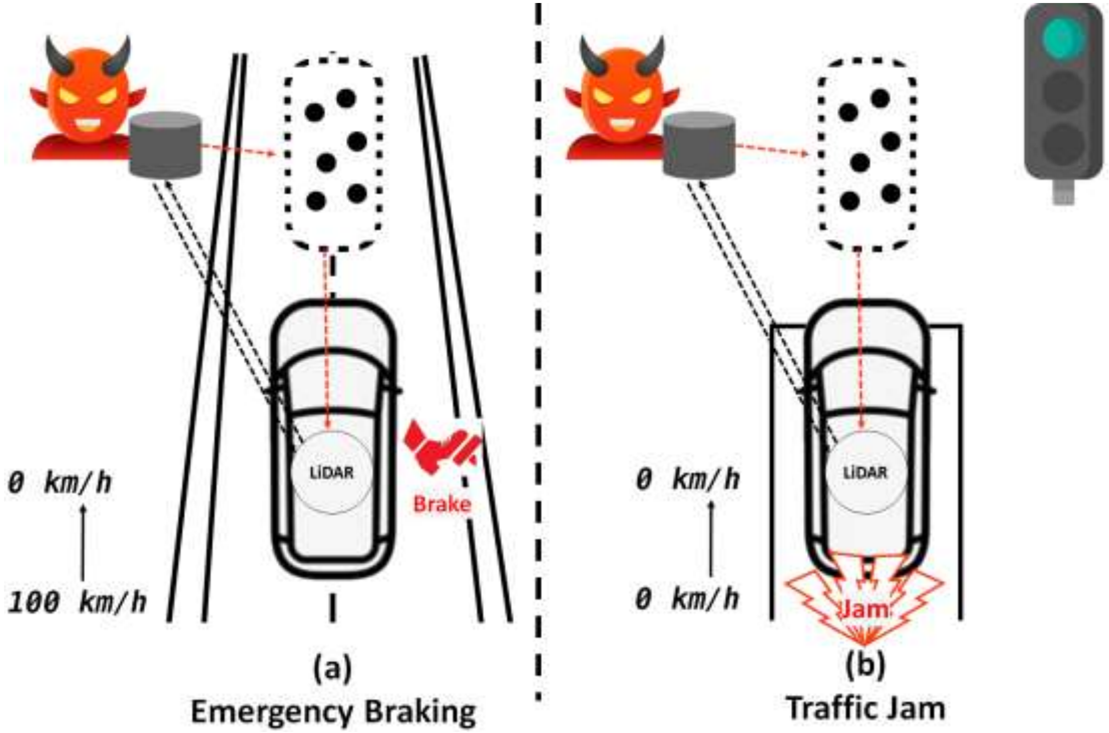


Threats of Appearing Attacks

- Injecting points into LiDAR point clouds
  1. Photodiode captures the lasers sent by LiDAR
  2. Laser transmitter sent back the fake reflected lasers



- Forging non-existent vehicles to pose threat
  1. Forcing the ADS vehicle to emergency brake
  2. Keeping the ADS vehicle immobile

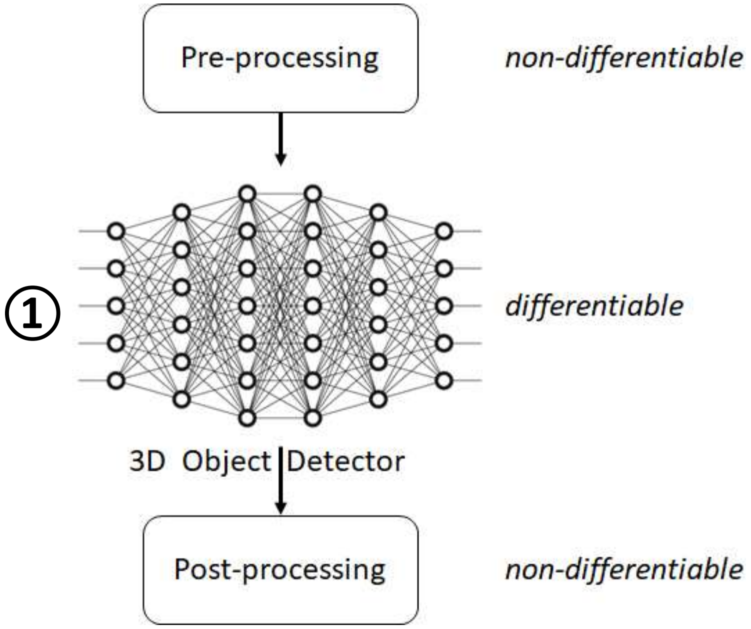


(a) Emergency Braking

(b) Traffic Jam

The Magic of Such Attacks

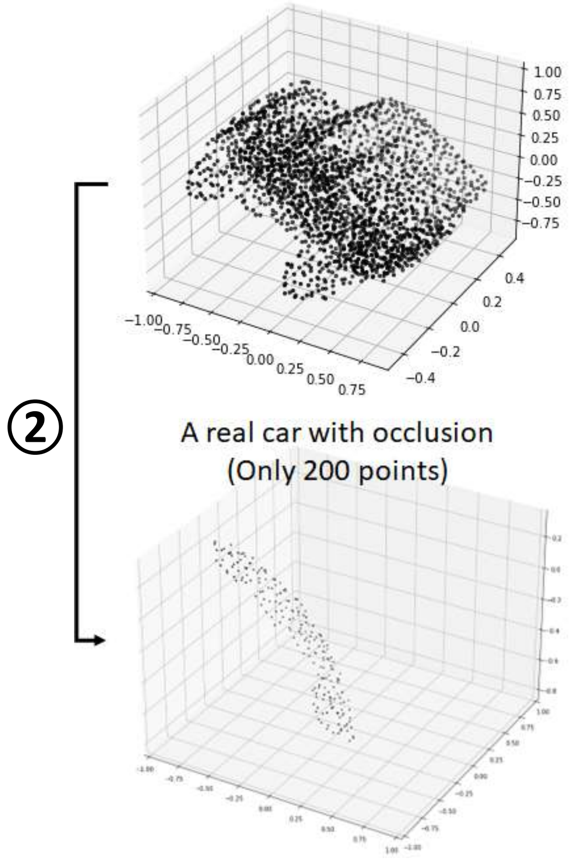
1. **Practicability**
   - Reusable traditional adversarial methods (FGSM, PGD, C&W…)
2. **Naturalness**
   - Difficult for human to distinguish
3. **Variability**
   - Various attack goals



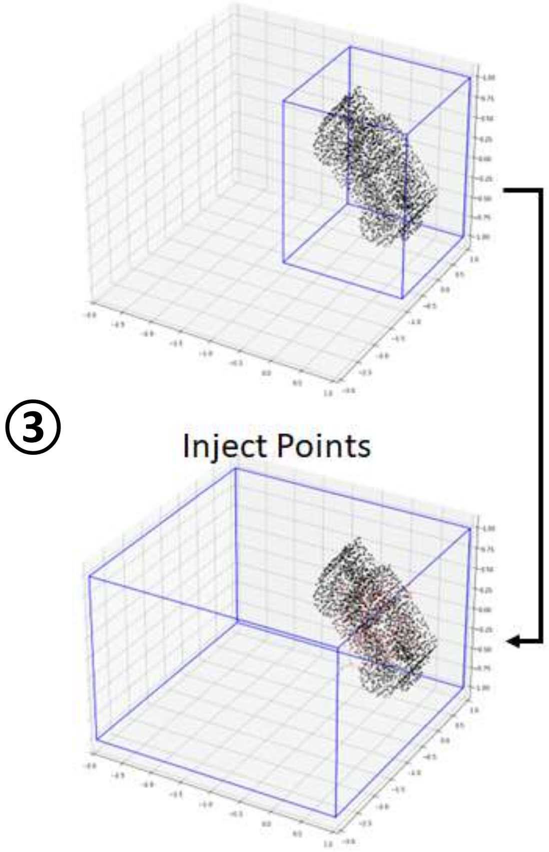
Pre-processing *non-differentiable*

① 3D Object Detector *differentiable*

Post-processing *non-differentiable*



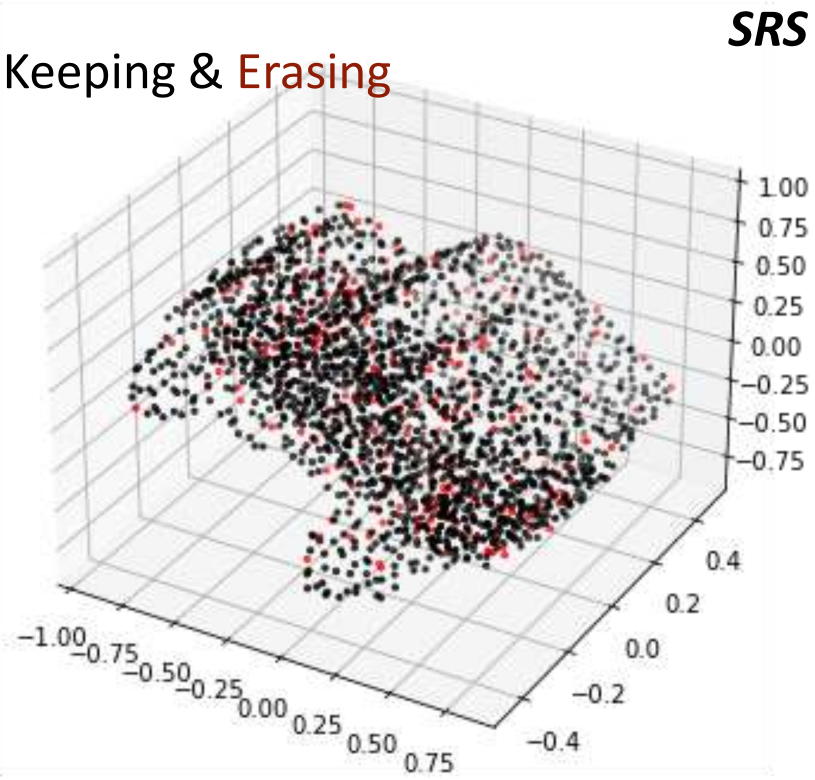
A real car with occlusion (Only 200 points)



Inject Points

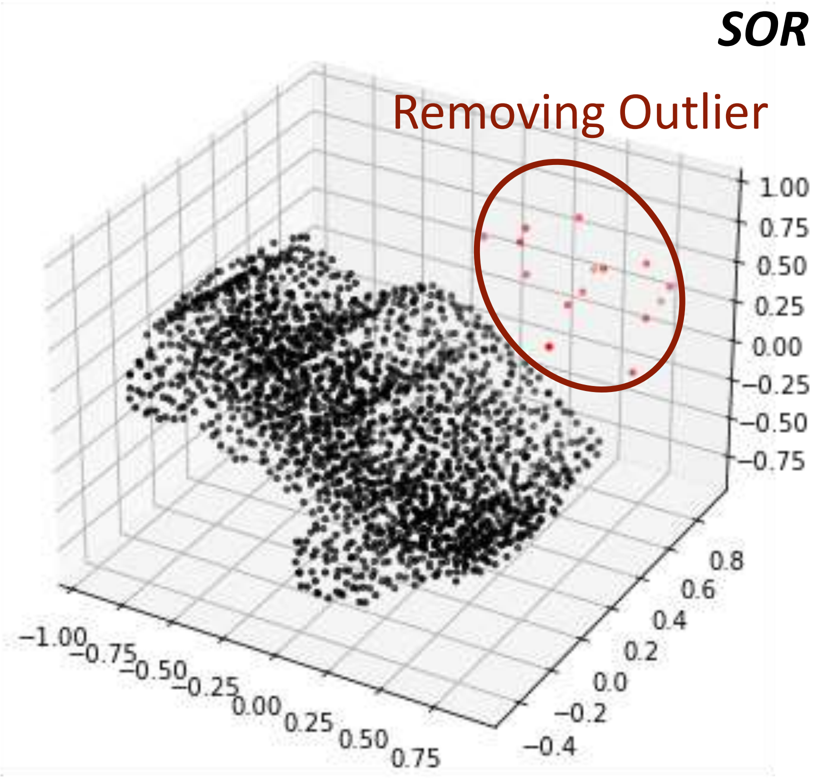
Existing Defense Methods

- **Universal Defenses**
    - Initial Motivation: Improving the robustness of PC models against noise
    - SRS and SOR



SRS

Keeping & Erasing

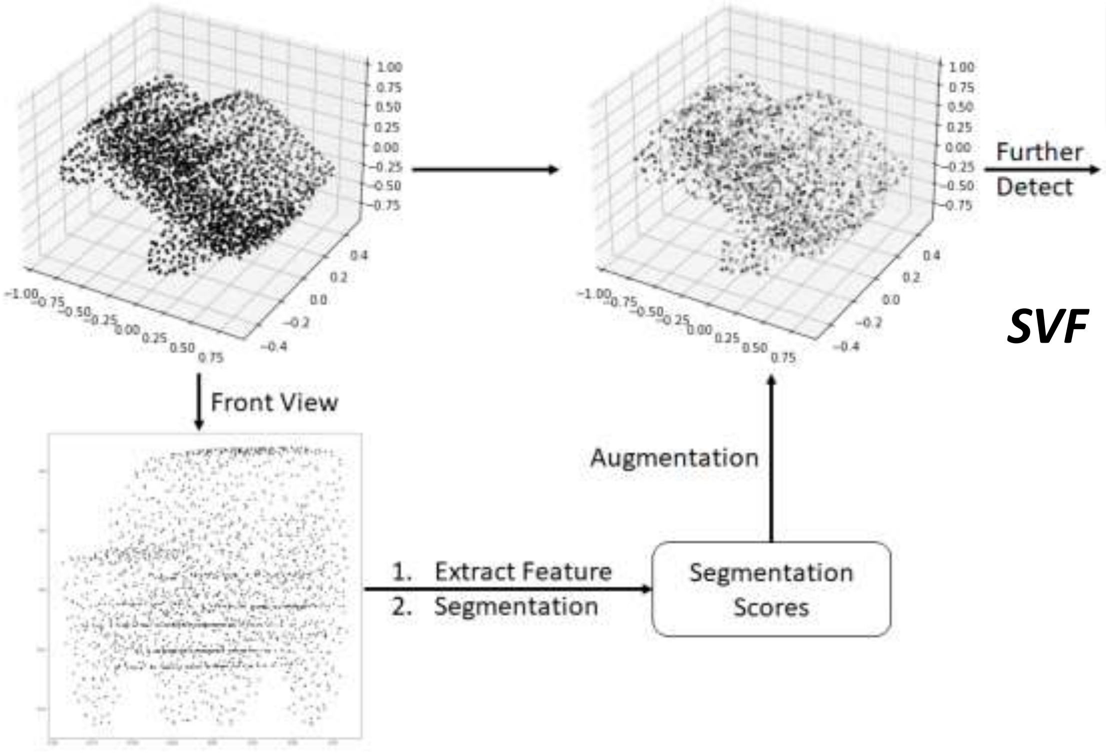


SOR

Removing Outlier

Existing Defense Methods

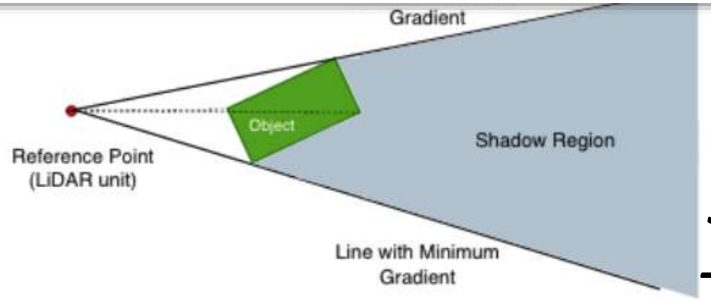
- **Specific Defenses**
  - Initial Motivation: Mitigating specific attack methods
  - SVF, CARLO and Shadow-Catcher



SVF

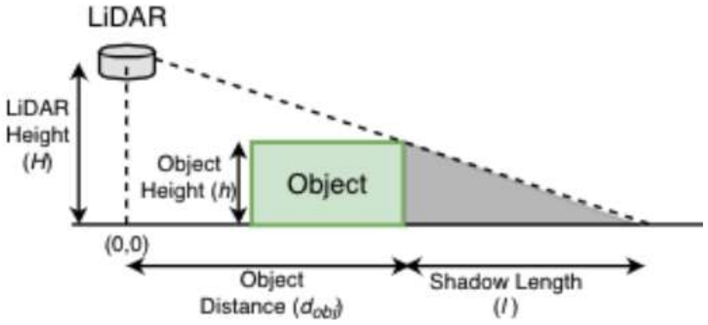
Refer from
https://sites.google.com/view/shadow-catcher

2D Shadow Region



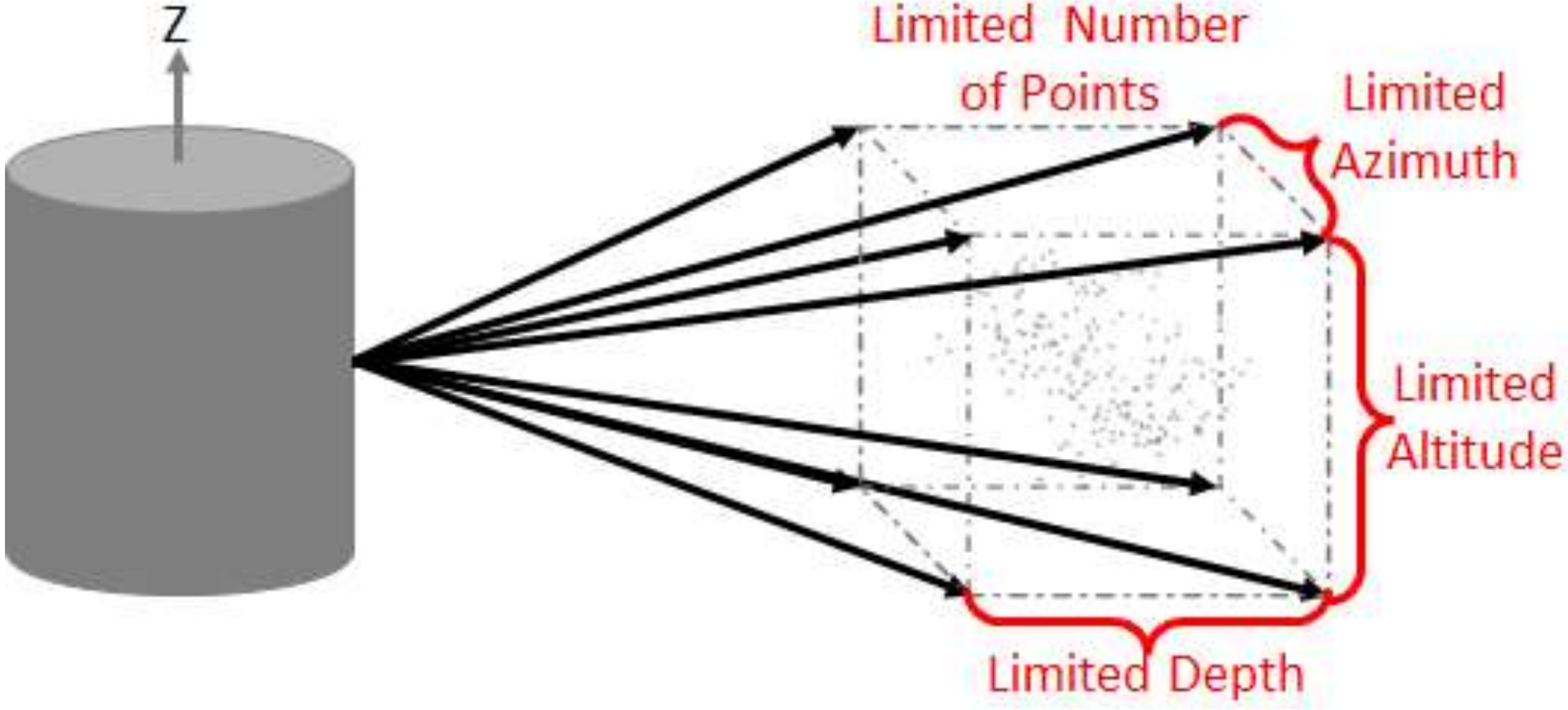
Shadow -Catcher

3D Shadow Region



Limitations of Existing Attacks

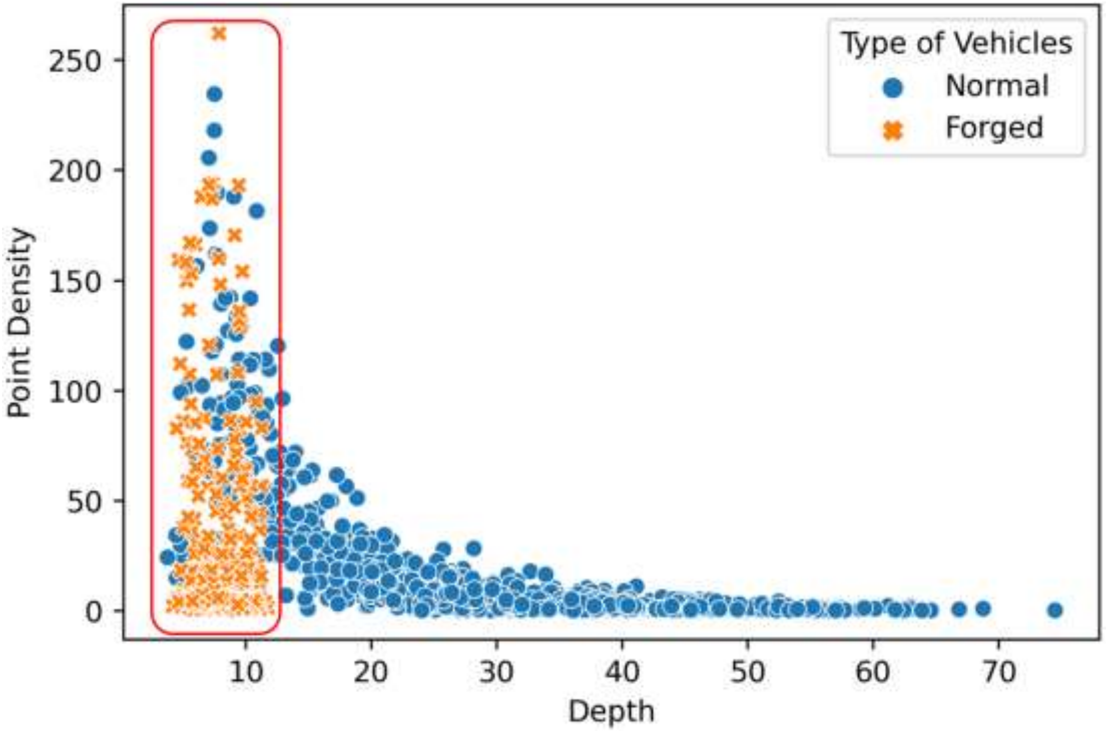
- **Two Common Limitations**
  1. Constrained by the **attack device** → the **position** and **number** of forged points
  2. Constrained by the **attack goal** → the **shape** of forged objects



Defense Insight

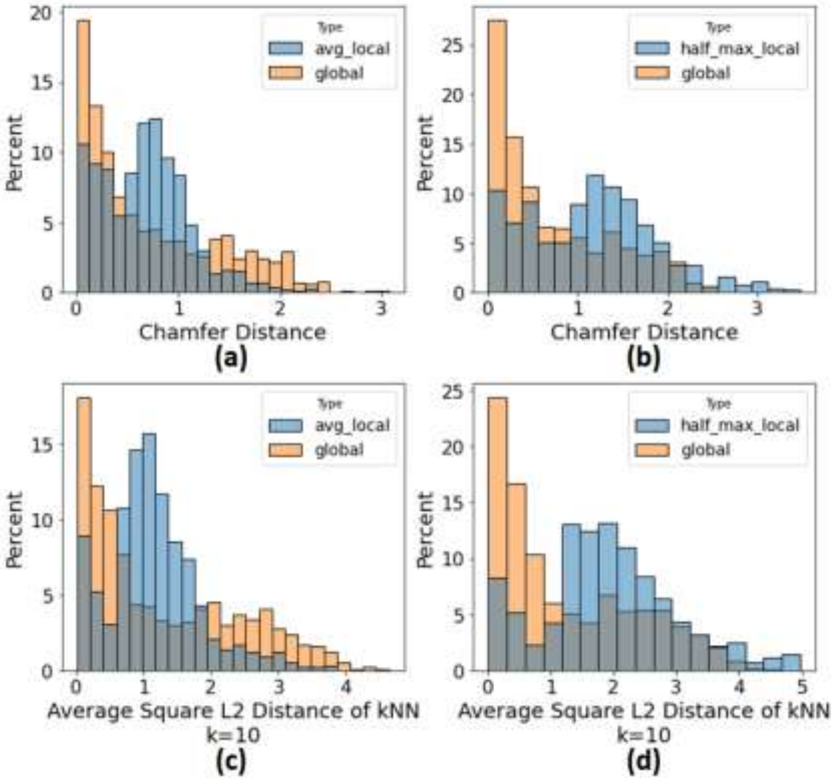
1. On the Position and Number

- the distributions of point density and depth are different



2. On the Shape

- the local difference is mostly larger than the global difference



LiDARs in Automated Driving System

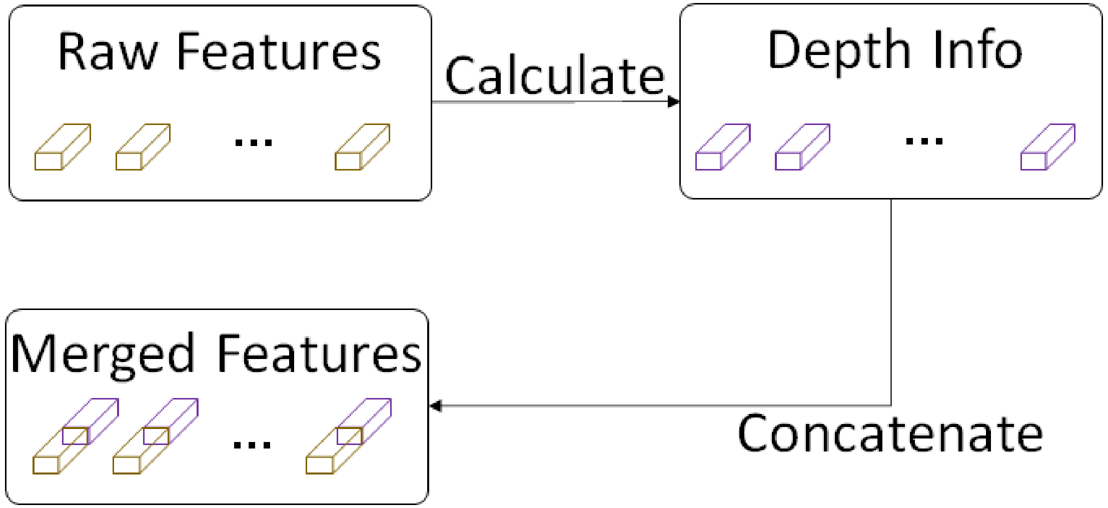
1. On the Position and Number

- Modeling the depth-density relation

2. On the Shape

- Deploying local detector + Voting

Explicit depth feature & Implicit density feature



Raw Features

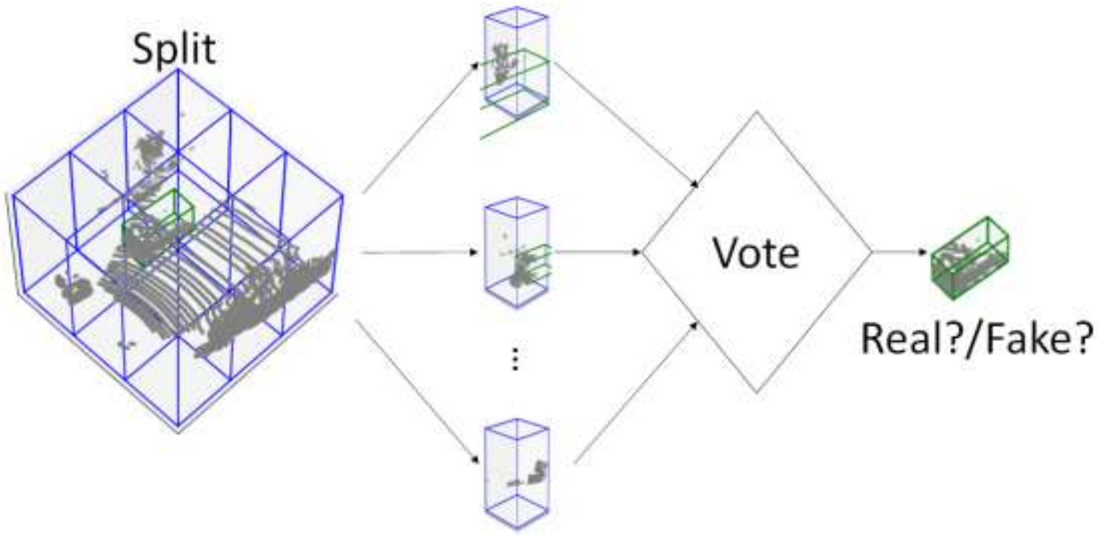
Calculate

Depth Info

Concatenate

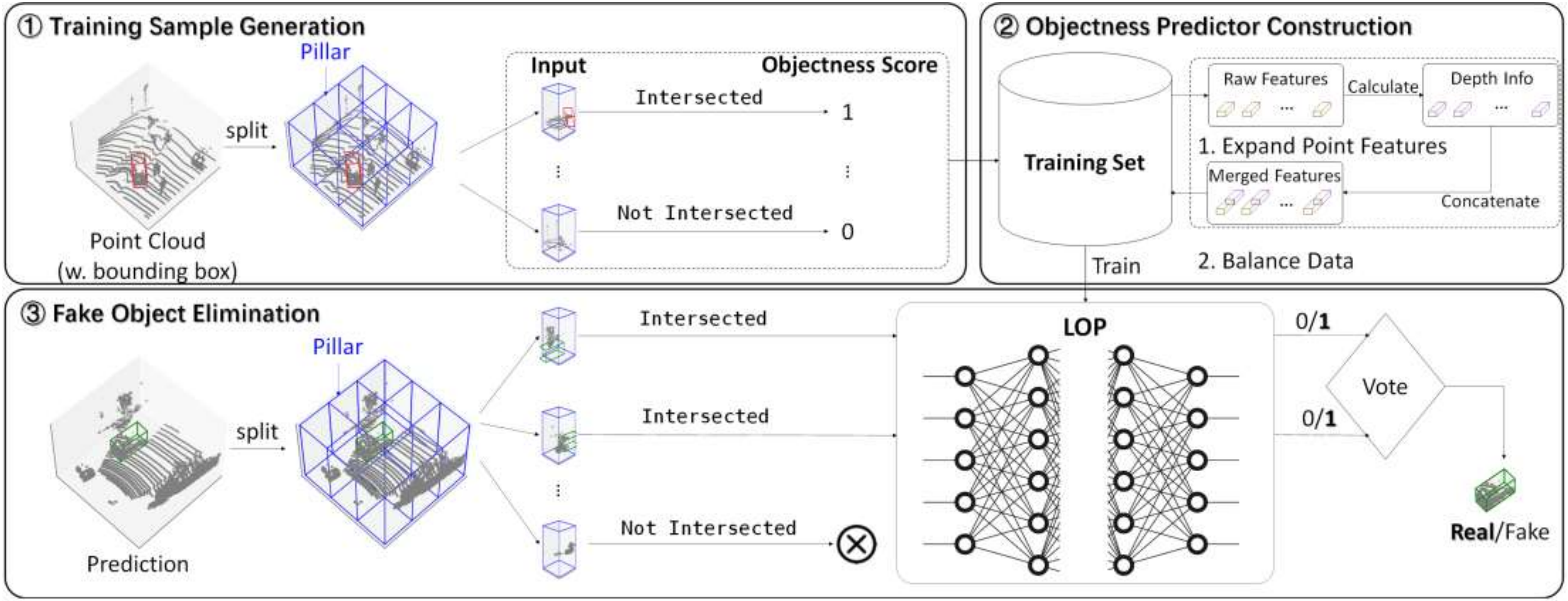
Merged Features

Split the prediction by splitting input space



Our Proposed Method

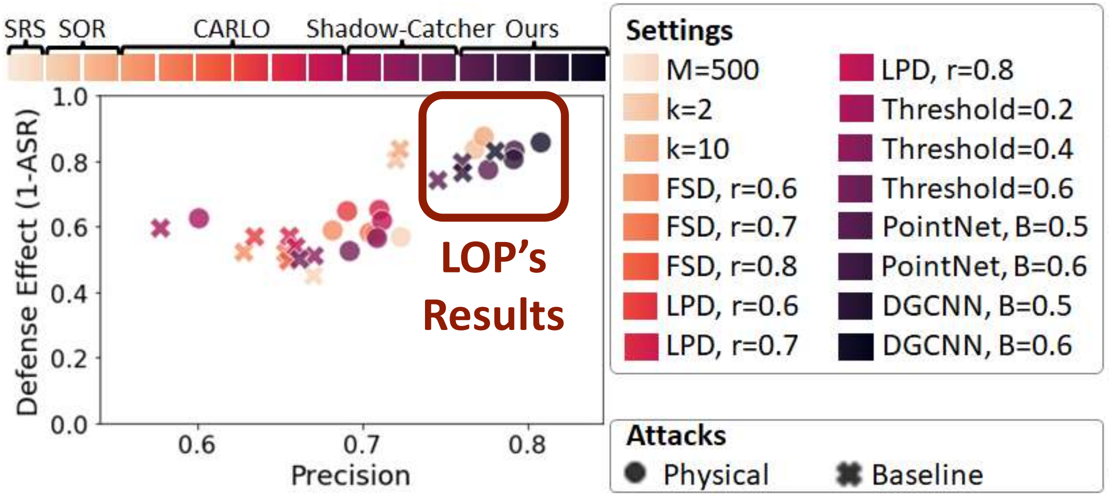
- **Local Objectness Predictor**
  - **Plug-and-Play Design** (**No need to retrain the whole detector*)



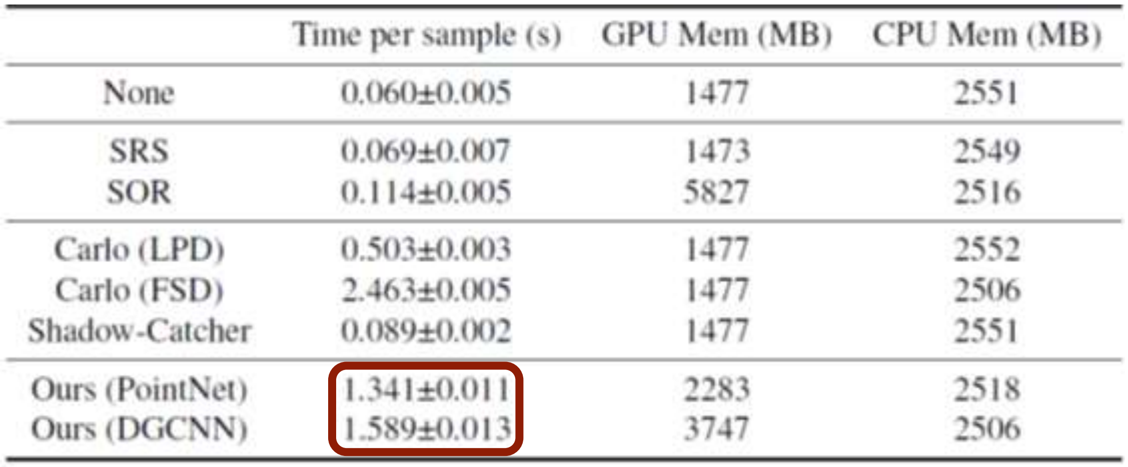
Defense Effectiveness

- More improvement on the performance and robustness of protected 3D object detectors

- Acceptable costs on memory and slightly larger costs on time



LOP's Results

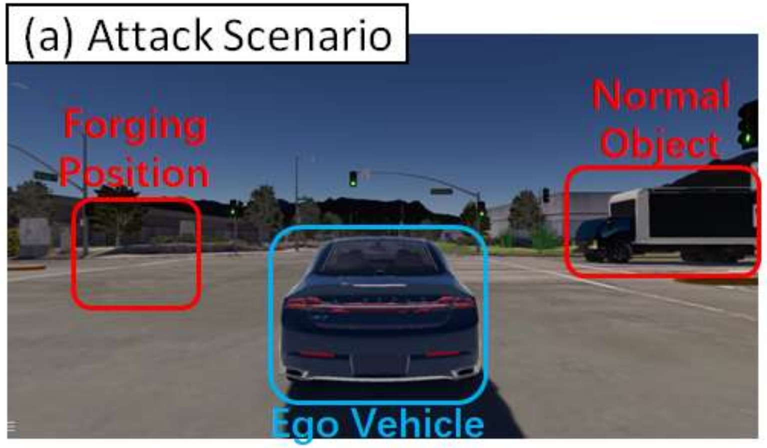


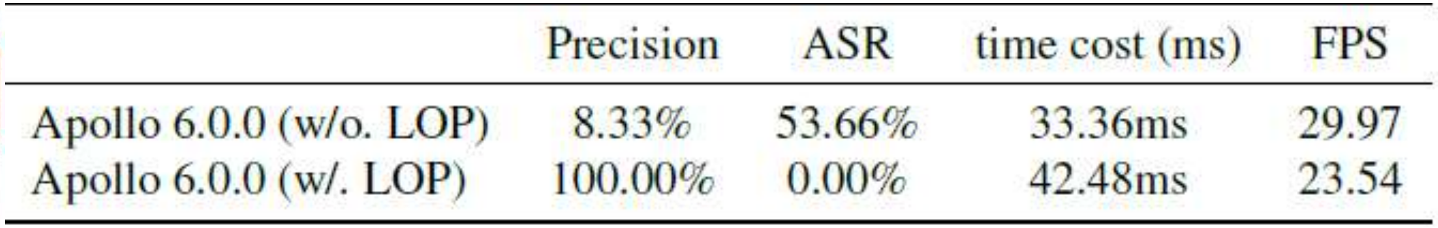
	Time per sample (s)	GPU Mem (MB)	CPU Mem (MB)
None	0.060±0.005	1477	2551
SRS	0.069±0.007	1473	2549
SOR	0.114±0.005	5827	2516
Carlo (LPD)	0.503±0.003	1477	2552
Carlo (FSD)	2.463±0.005	1477	2506
Shadow-Catcher	0.089±0.002	1477	2551
Ours (PointNet)	1.341±0.011	2283	2518
Ours (DGCNN)	1.589±0.013	3747	2506

can further reduce by multi-processing

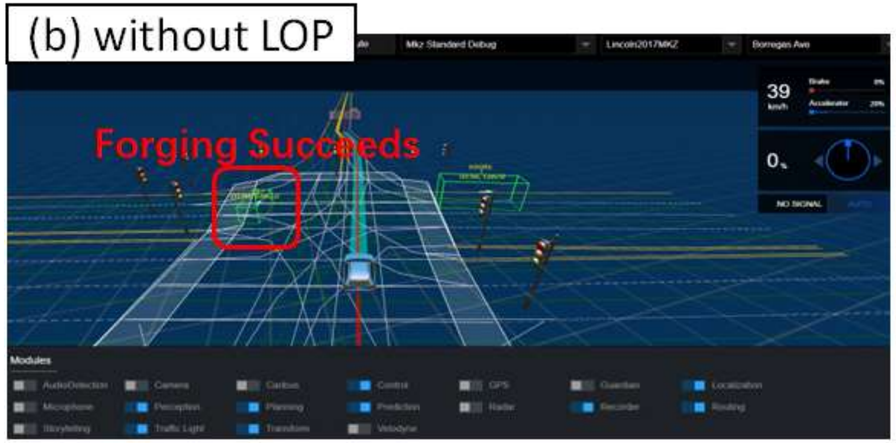
Simulation Experiments

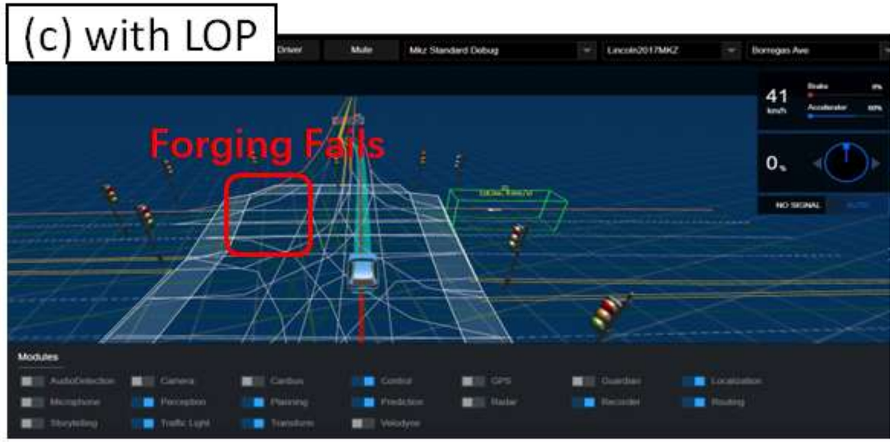
- The performance of Apollo 6.0.0 deployed with LOP, evaluated in LGSVL simulator





	Precision	ASR	time cost (ms)	FPS
Apollo 6.0.0 (w/o. LOP)	8.33%	53.66%	33.36ms	29.97
Apollo 6.0.0 (w/. LOP)	100.00%	0.00%	42.48ms	23.54

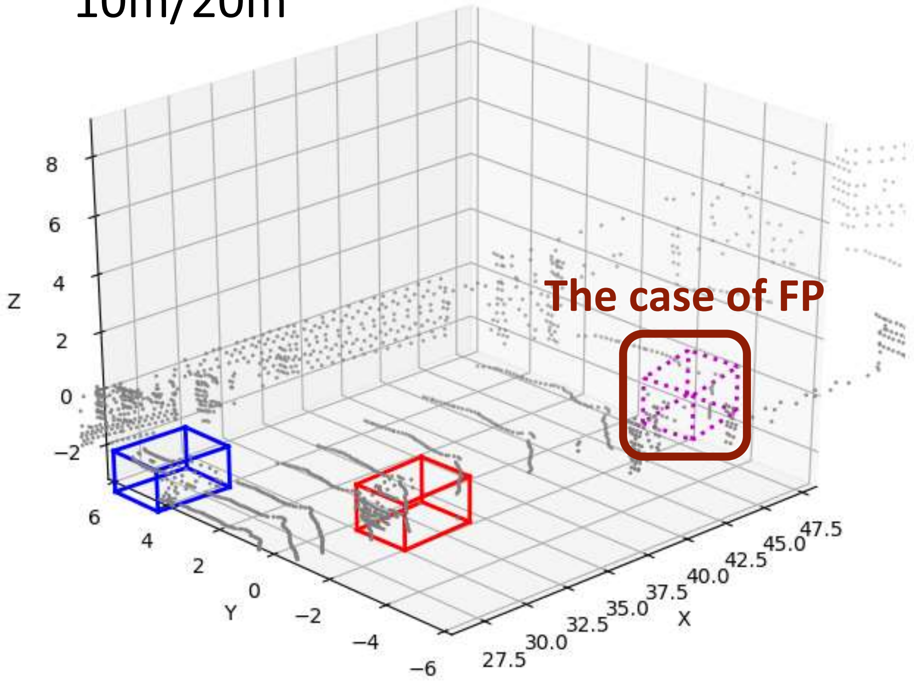




Future Directions

Direction 1
The Existence of False positives

- Farther objects are harder to detect
  - 12.95%/16.53% of FP with depth < 10m/20m

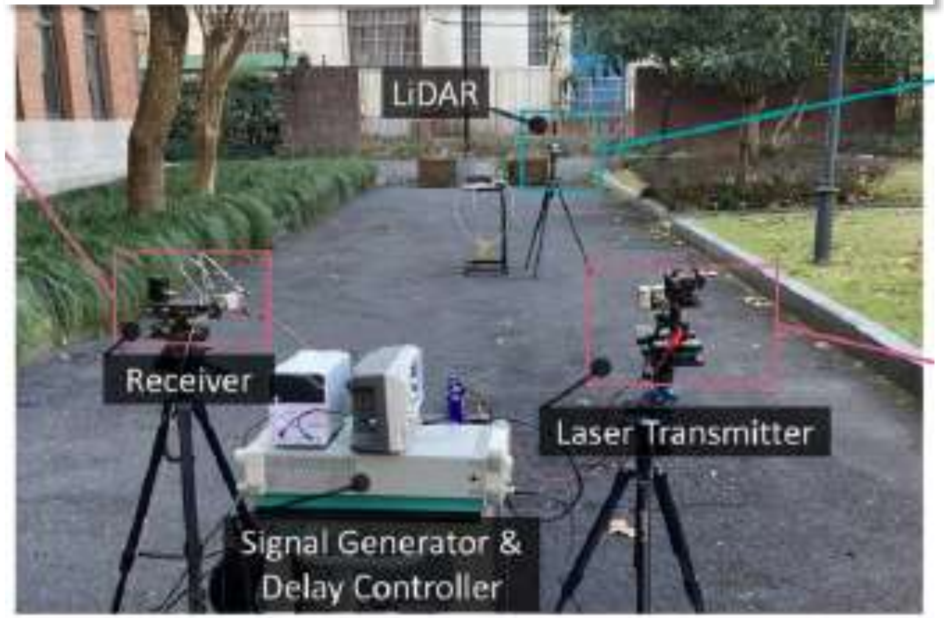


The case of FP

Direction 2
The Upgrade of Attack Device

- The maximum of forged points is already up to 2500

The Fig.7 in PLA-LiDAR (S&P 2023)



Take Away Message

1. We conclude the limitations of existing appearing attacks
2. We propose a plug-and-play defense method LOP
3. We prove the effectiveness of our LOP online and offline

↓ Full paper of our LOP

More Research on AI Security

Thank you for your Audience!

For more details, welcome to follow our paper