NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

School of Computer Science and Engineering
College of Engineering

Safety of Autonomous Vehicles

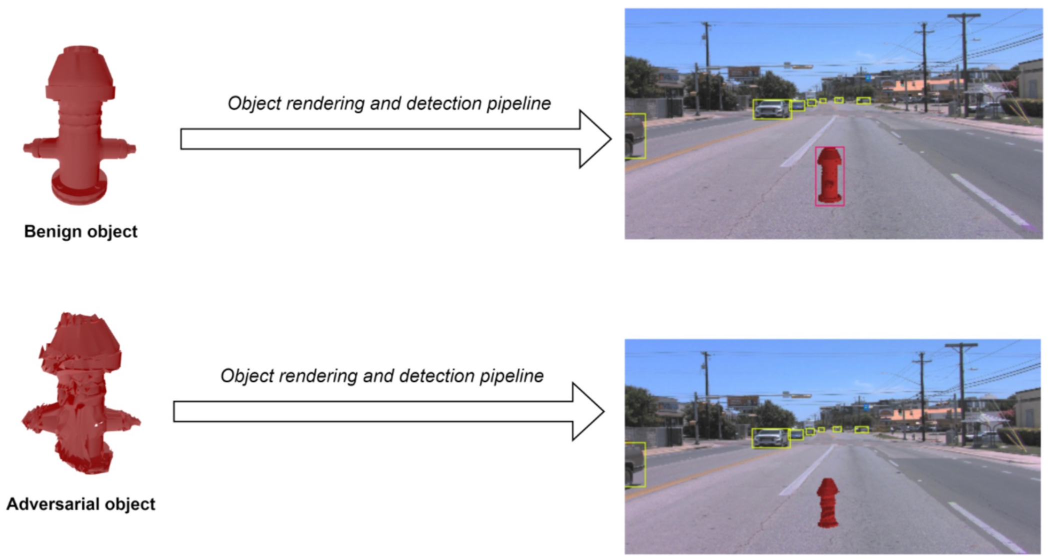
Understanding Adversarial Attacks on Autonomous Vehicle's Perception Modules

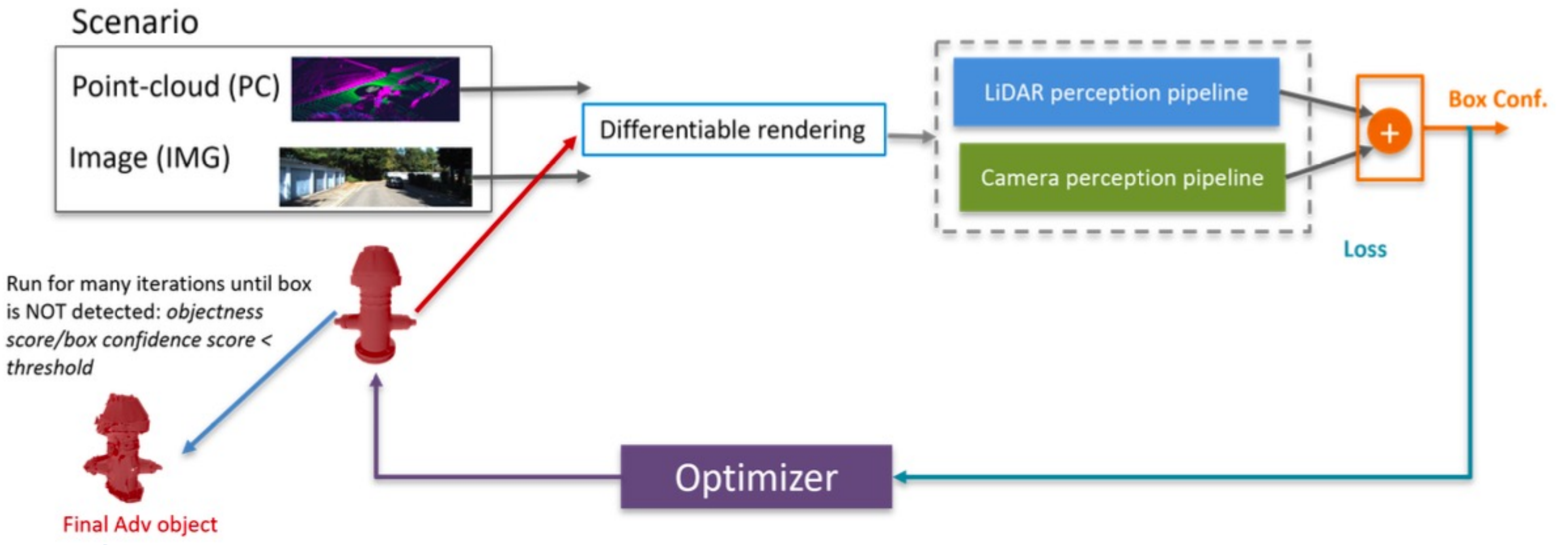
Student: Jonathan Chan Chew Meng   Supervisor: Associate Professor Anupam Chattopadhyay

Project Objectives:

This project aims to discuss adversarial attacks on Autonomous Vehicles (AVs), and the defence mechanisms that can be utilized to prevent such attacks. The project first focuses on spoofing multiple cameras with overlapping field of view, then moves on to discuss other various feature squeezing countermeasure techniques that can be used to protect AVs from these adversarial attacks. The paper includes experiments that evaluate the effectiveness of these countermeasures using different scenarios and datasets. The project also highlight potential future works, including exploring other types of adversarial attacks and implementing adversarial training of neural networks.

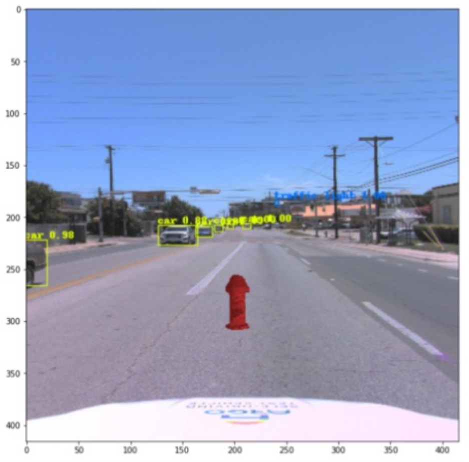
Multiple Camera with Overlapping FOV: Formulation of Attack Objective


Object rendering and detection pipeline
Benign object
Object rendering and detection pipeline
Adversarial object


Scenario
Point-cloud (PC)
Image (IMG)
Differentiable rendering
LiDAR perception pipeline
Camera perception pipeline
Box Conf.
Loss
Run for many iterations until box is NOT detected: *objectness score/box confidence score < threshold*
Final Adv object
Optimizer

Feature Squeezing Techniques

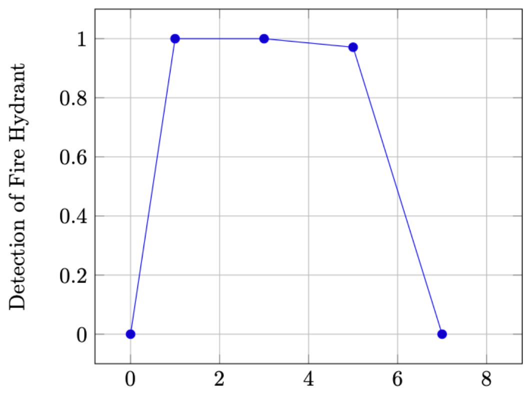
Image Compression



Colors	Fire Hydrant	Cars
256	0.999205	0.613719-0.998912
128	0.999241	0.618021-0.998865
64	0.981825	0.731816-0.998967
32	0.904038	0.908702-0.999631
16	0.940976	0.883776-0.999583
8	Undetected	0.965660-0.999893

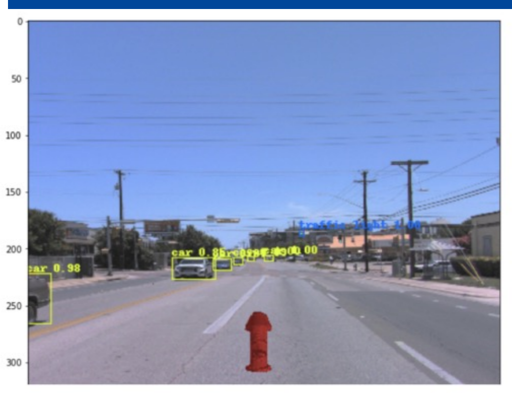
From the results, it can be seen that image compression technique was successful in reducing the threats significantly. At the highest level of compression, detection was successful for both Fire Hydrants and Cars with a precision of 99.9%, overcoming the adversarial attacks.

Median Smoothing



As seen from the results, suitable amount of median smoothing can help to defend against possible adversarial attacks by making the attacks more difficult.

Colour Depth Reduction



	Benign det. rate (%)		Attack success rate (%)	
Color depth	Single-cam	Multi-cam	Single-cam	Multi-cam
24-bit (orig.)	75.75	98.38	78	43.75
8-bit	51.88	84.38	39	17.6
24-bit, 8-bit	78.75	**99**	37.25	**13**

Due to the low effectiveness of dimension reduction, we combine colour depth reduction and multi-camera setup as one unified countermeasure. Our results show that such technique combined leads to a much more robust camera perception system: 99% benign detection rate and just 13% attack success rate in the adversarial case

Link to paper accepted in International Conference on Security, Privacy and Applied Cryptography Engineering 2022:
https://link.springer.com/chapter/10.1007/978-3-031-22829-2_14

https://www.ntu.edu.sg/scse