Unity is Strength?
Benchmarking the Robustness of Fusion-based 3D Object Detection against Physical Sensor Attack

Zizhi Jin, Xuancun Lu, Bo Yang, Yushi Cheng, Chen Yan, Xiaoyu Ji*, Wenyuan Xu

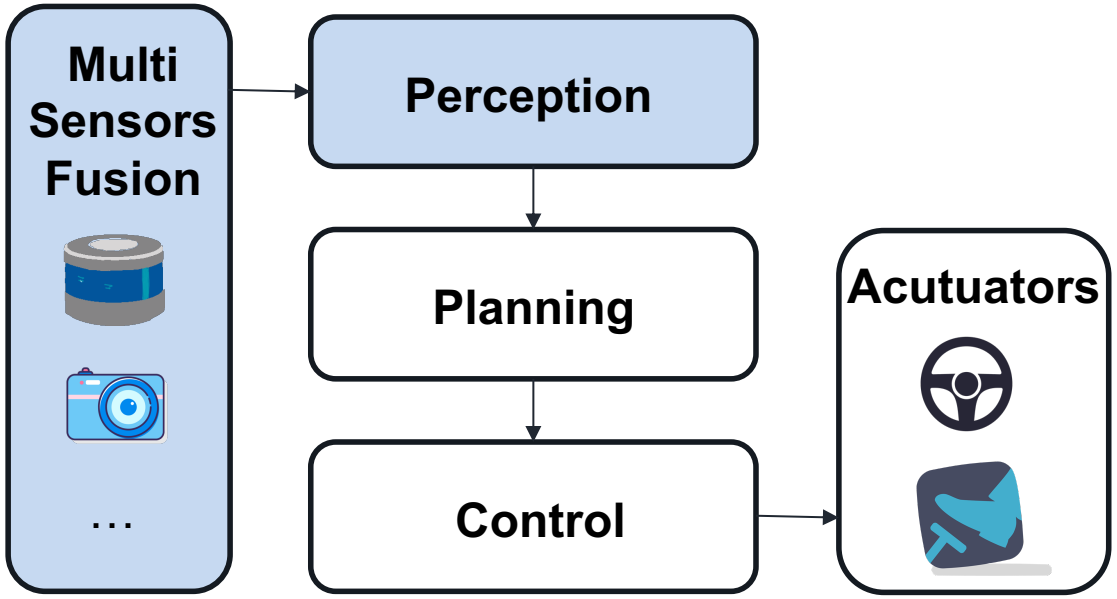
Zhejiang University

智能系统安全实验室
UBIQUITOUS SYSTEM SECURITY LAB.

浙江大学
ZHEJIANG UNIVERSITY

Background - Sensor Fusion in Autonomous Driving

- 3D object detection serves as the core basis of the perception stack.
- LiDAR and camera are broadly used in autonomous driving



Source: Cruise

Source: Waymo

Source: Zoox

Source: Cruise

Perception security is the prerequisite for safe driving

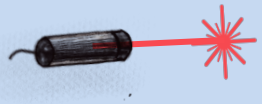
119+ cars to be released with LiDAR and Camera by OEMs in 2023[1]

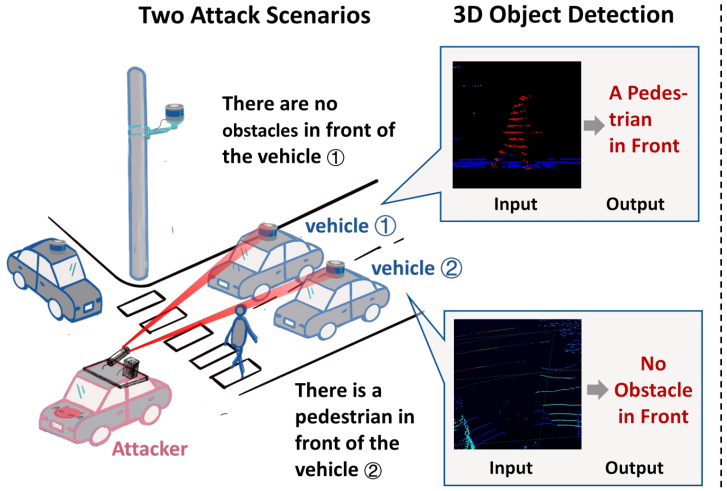
[1] Yole Intelligence. 2023. LiDAR for Automotive 2023.

Background – Physical Sensor Attack

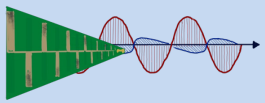
- In real world, LiDAR and camera systems can be compromised by **physical signals**, e.g., laser, electromagnetic interference (EMI) and ultrasound.

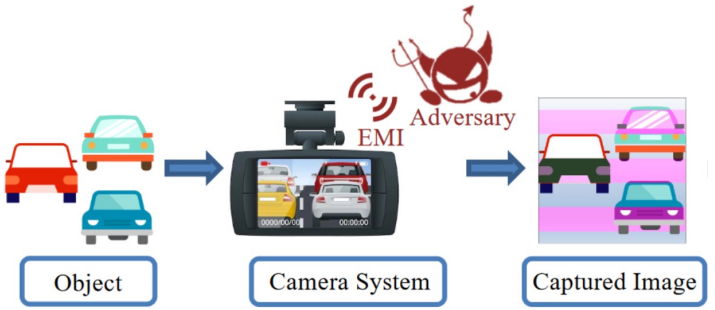
Laser Attack[1]



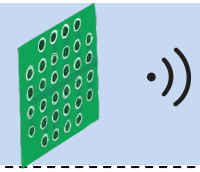


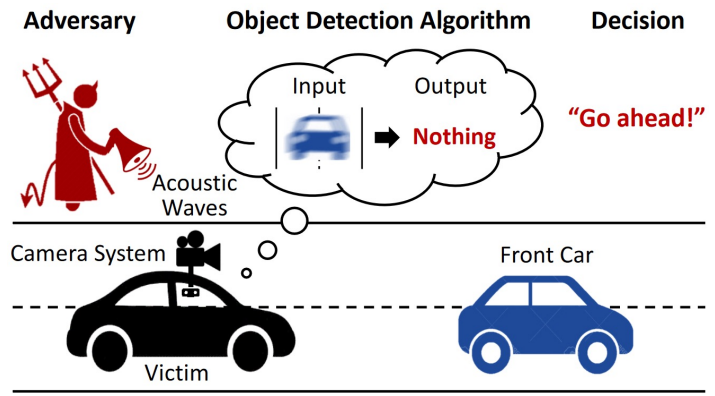
EMI Attack[2]





Ultrasound Attack[3]





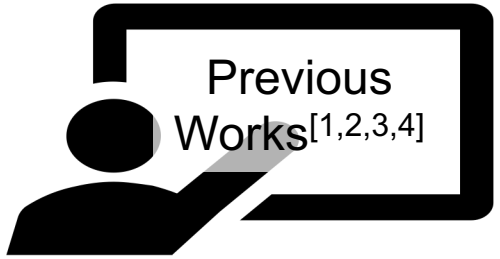
[1] Jin, Zizhi, et al. "Pla-lidar: Physical laser attacks against lidar-based 3d object detection in autonomous vehicle." 2023 IEEE Symposium on Security and Privacy (SP). IEEE, 2023
[2] Jiang, Qinhong, et al. "{GlitchHiker}: Uncovering Vulnerabilities of Image Signal Transmission with {IEMI}." 32nd USENIX Security Symposium (USENIX Security 23). 2023.
[3] Ji, Xiaoyu, et al. "Poltergeist: Acoustic adversarial machine learning against cameras and computer vision." 2021 IEEE Symposium on Security and Privacy (SP). IEEE, 2021.

… please refer to our paper for more reference

Motivation 1: Lack Experimental Exploring for Common Assumptions

Common Assumptions:
*Sensor fusion is a **potential** **defend method** against physical sensor attack*

Is it true?



10.3.2 *Sensor-Level Defenses.* Several defenses could be adopted against spoofing attacks on LiDAR sensors:

5) Sensor Fusion: Defense by sensor fusion enhances resiliency against transduction attacks by utilizing output from

Sensor Fusion Techniques. Another complementary defense approach is to exploit sensor fusion for decision making.

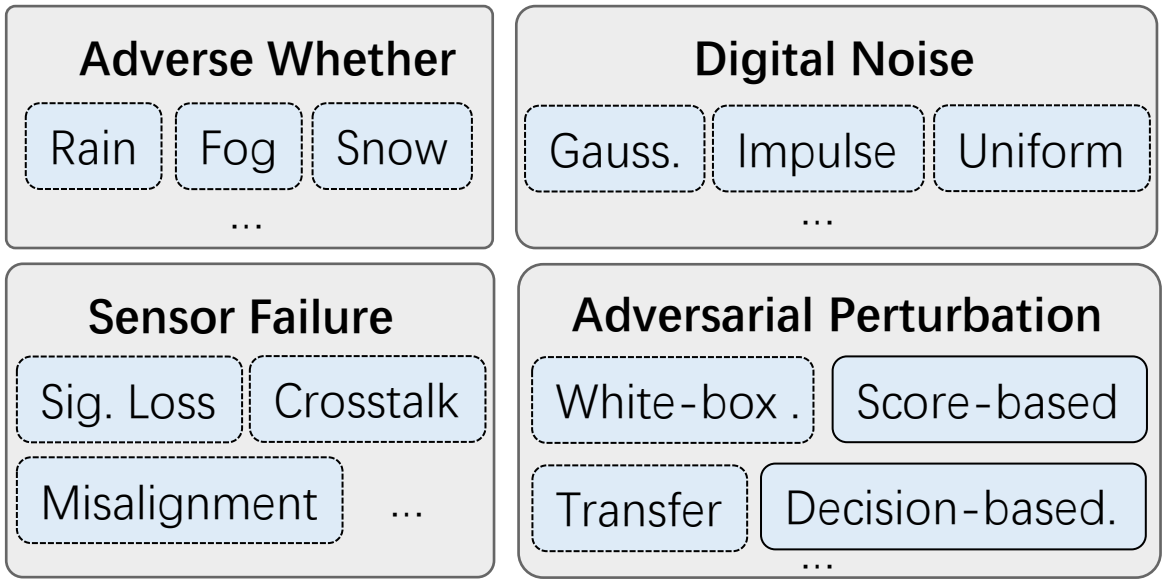
Multi-sensor Fusion and Security Redundancy. Another complementary defense approach is to exploit multi-sensor fusion for decision-making. Autonomous vehicles can employ multiple types of sensors, e.g., cameras, radars, ultrasonic sensors combined with LiDARs to perceive the environment. Such information fusion and redundancy may help further improve the security of autonomous vehicles.

Good Fusion ☺ Bad Fusion ☹

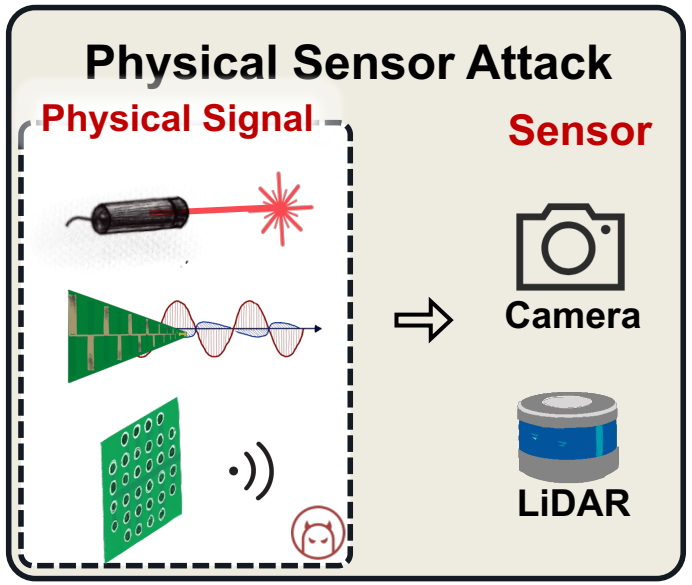
[1] Cao, Yulong, et al. "Adversarial sensor attack on lidar-based perception in autonomous driving." Proceedings of the 2019 ACM SIGSAC conference on computer and communications security. 2019.
[2] Yan, Chen, et al. "Sok: A minimalist approach to formalizing analog sensor security." 2020 IEEE Symposium on Security and Privacy (SP). IEEE, 2020.
[3] Ji, Xiaoyu, et al. "Poltergeist: Acoustic adversarial machine learning against cameras and computer vision." 2021 IEEE Symposium on Security and Privacy (SP). IEEE, 2021.
[4] Jin, Zizhi, et al. "Pla-lidar: Physical laser attacks against lidar-based 3d object detection in autonomous vehicle." 2023 IEEE Symposium on Security and Privacy (SP). IEEE, 2023.

Motivation 2: Lack of Corruptions Dataset for Sensor Attack

- One common practice for robustness analysis is to establish a benchmark.



The Corruptions in Previous Robustness Benchmark[1,2,···]
Unintentionally induced.

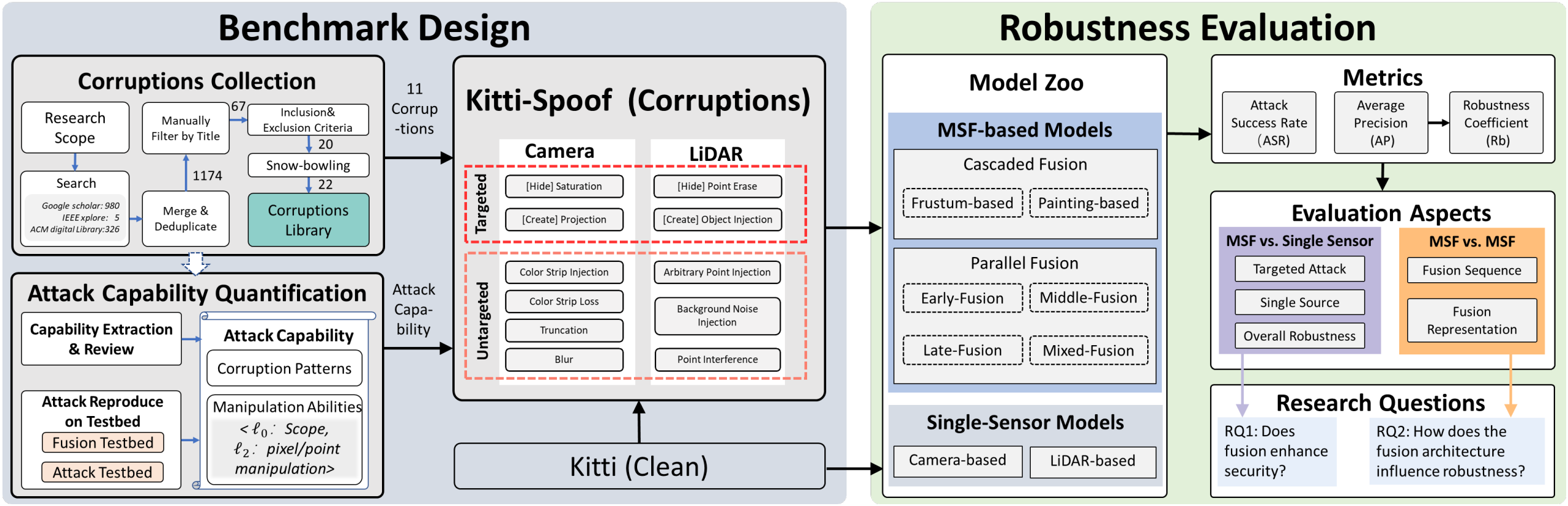


Ours.
Intentionally induced by Attackers

[1] Dong, Yinpeng, et al. "Benchmarking robustness of 3d object detection to common corruptions." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
[2] Gao, Xinyu, et al. "Benchmarking Robustness of AI-enabled Multi-sensor Fusion Systems: Challenges and Opportunities." Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. 2023.

… please refer to our paper for more reference

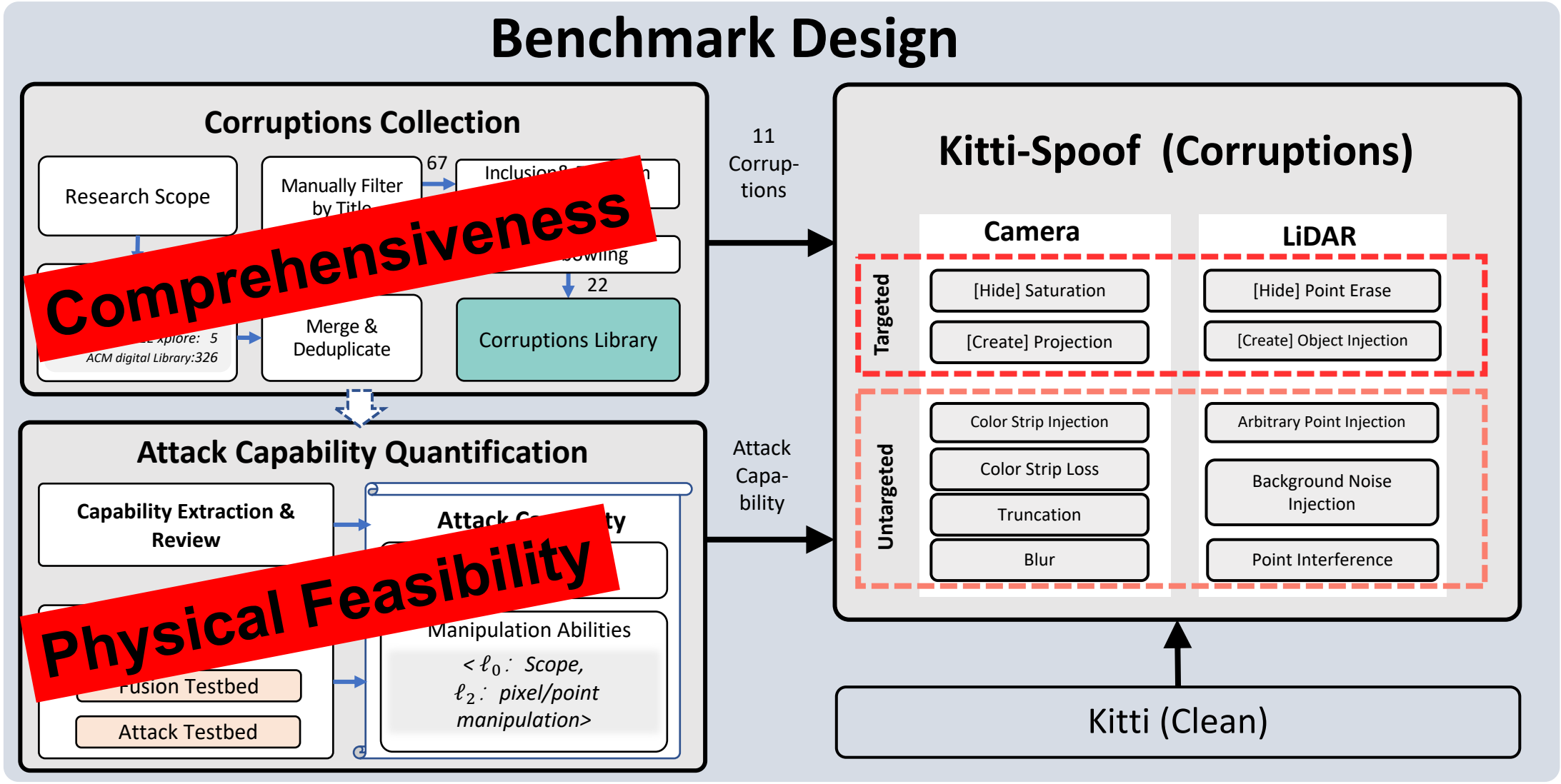
Overview



Research Question 1: Does fusion enhance robustness or not?

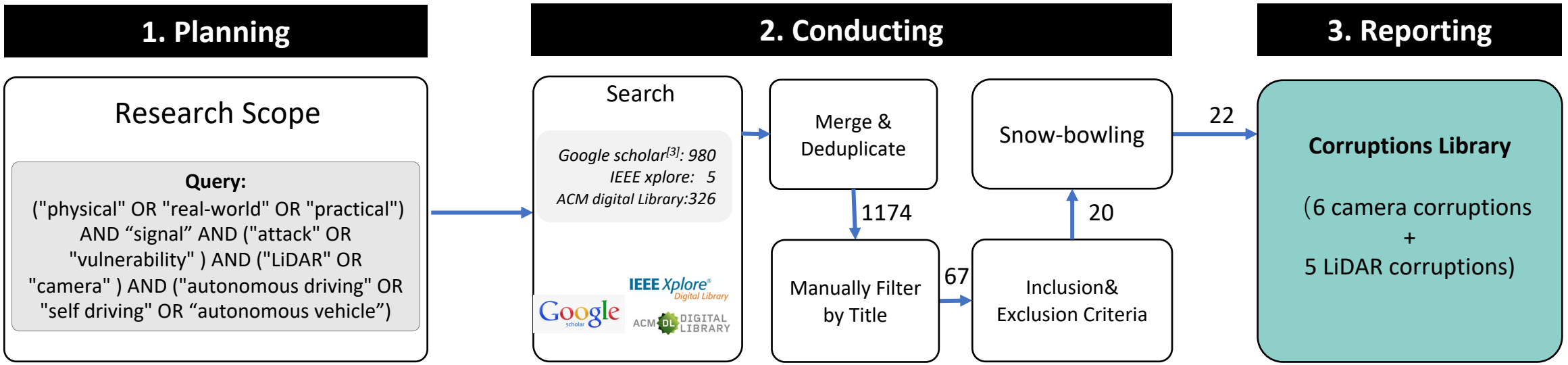
Research Question 2: How does the architecture of the fusion model influence robustness?

Benchmark Design



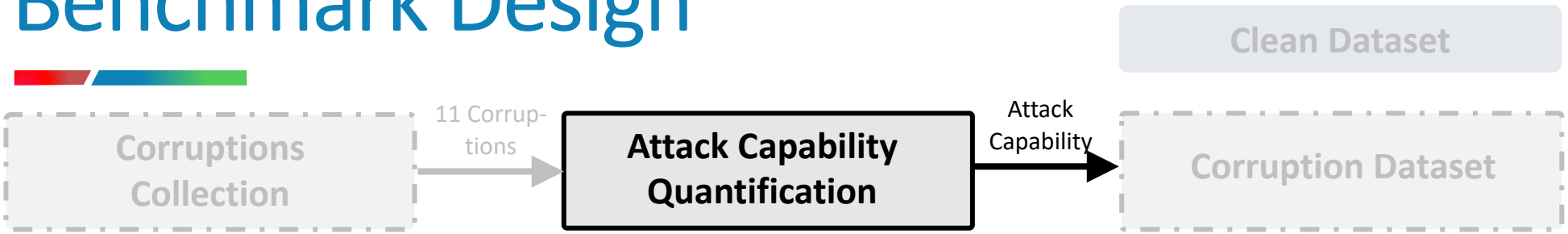
Benchmark Design

Collect corruptions with regard of physical sensor attacks through a Systematic Literature Review (SLR) process[1,2].



[1] Kitchenham, Barbara, et al. "Systematic literature reviews in software engineering–a systematic literature review." Information and software technology 51.1 (2009): 7-15.
[2] Ladisa, Piergiorgio, et al. "Sok: Taxonomy of attacks on open-source software supply chains." 2023 IEEE Symposium on Security and Privacy (SP). IEEE, 2023.
[3] A.W. Harzing. 2023. Publish or Perish. https://harzing.com/resources/publishor- perish

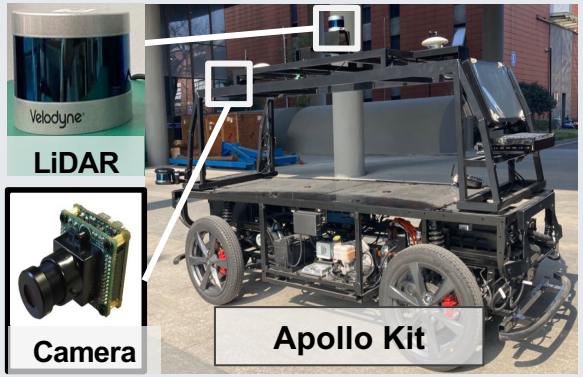
Benchmark Design



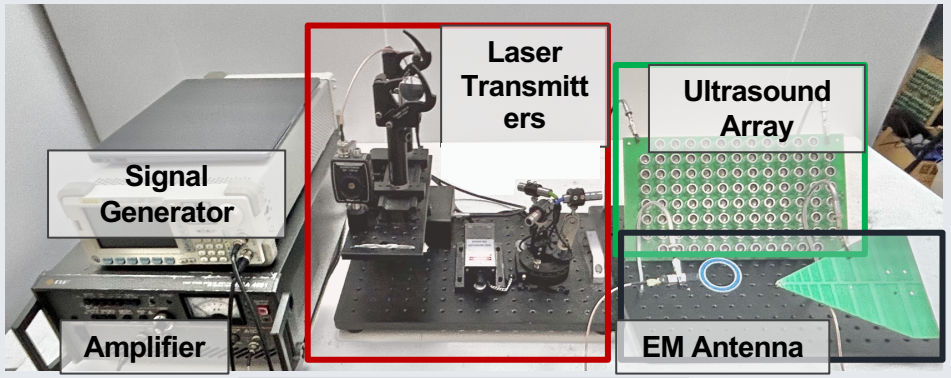
We quantify the capabilities of sensor attacks through reviewing source papers and reproducing the attacks.

Capability Extraction From Source Papers

Attack Reproduce on Testbed



Fusion Testbed



Attack Testbed

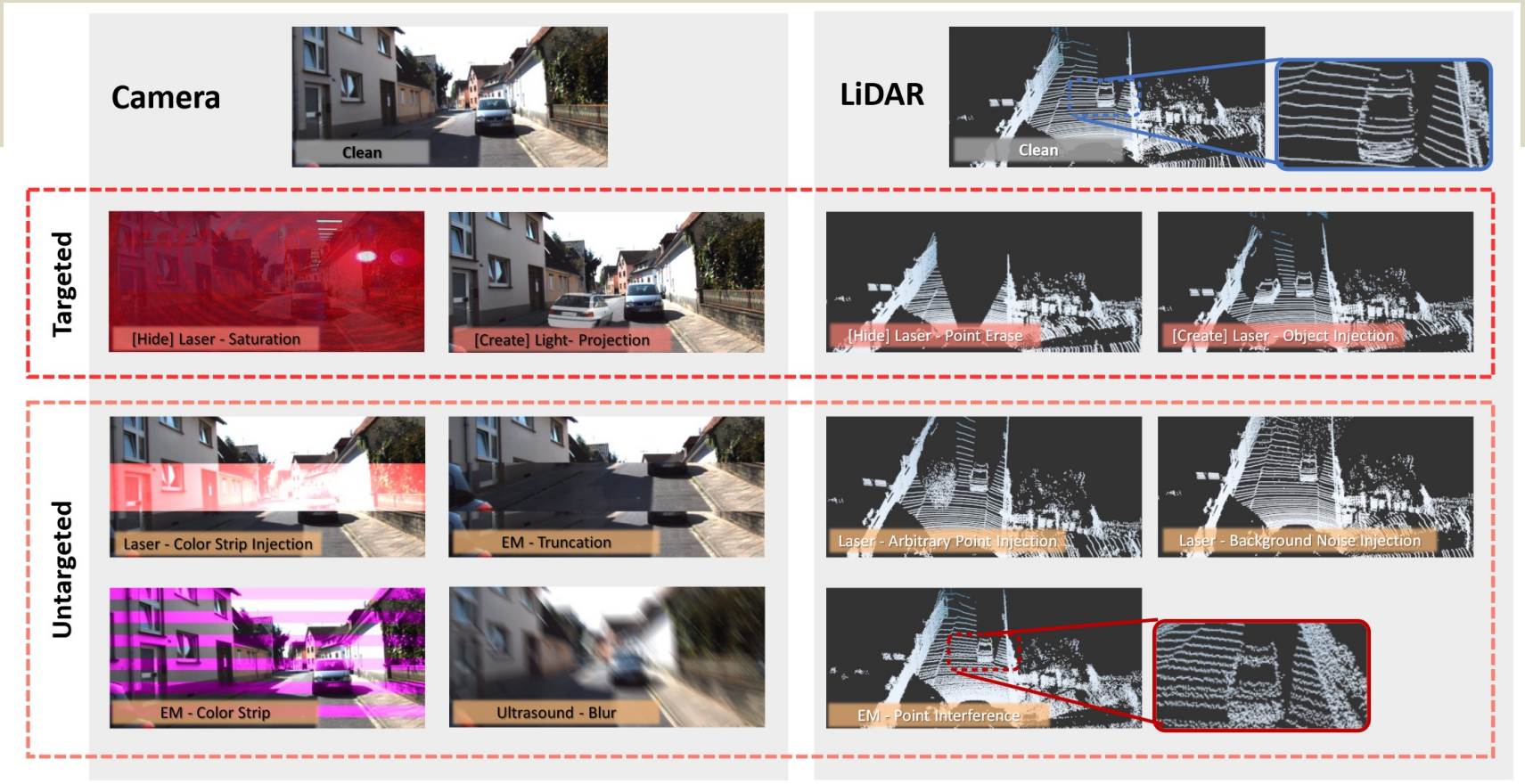
Attack Capability

Corruption Patterns

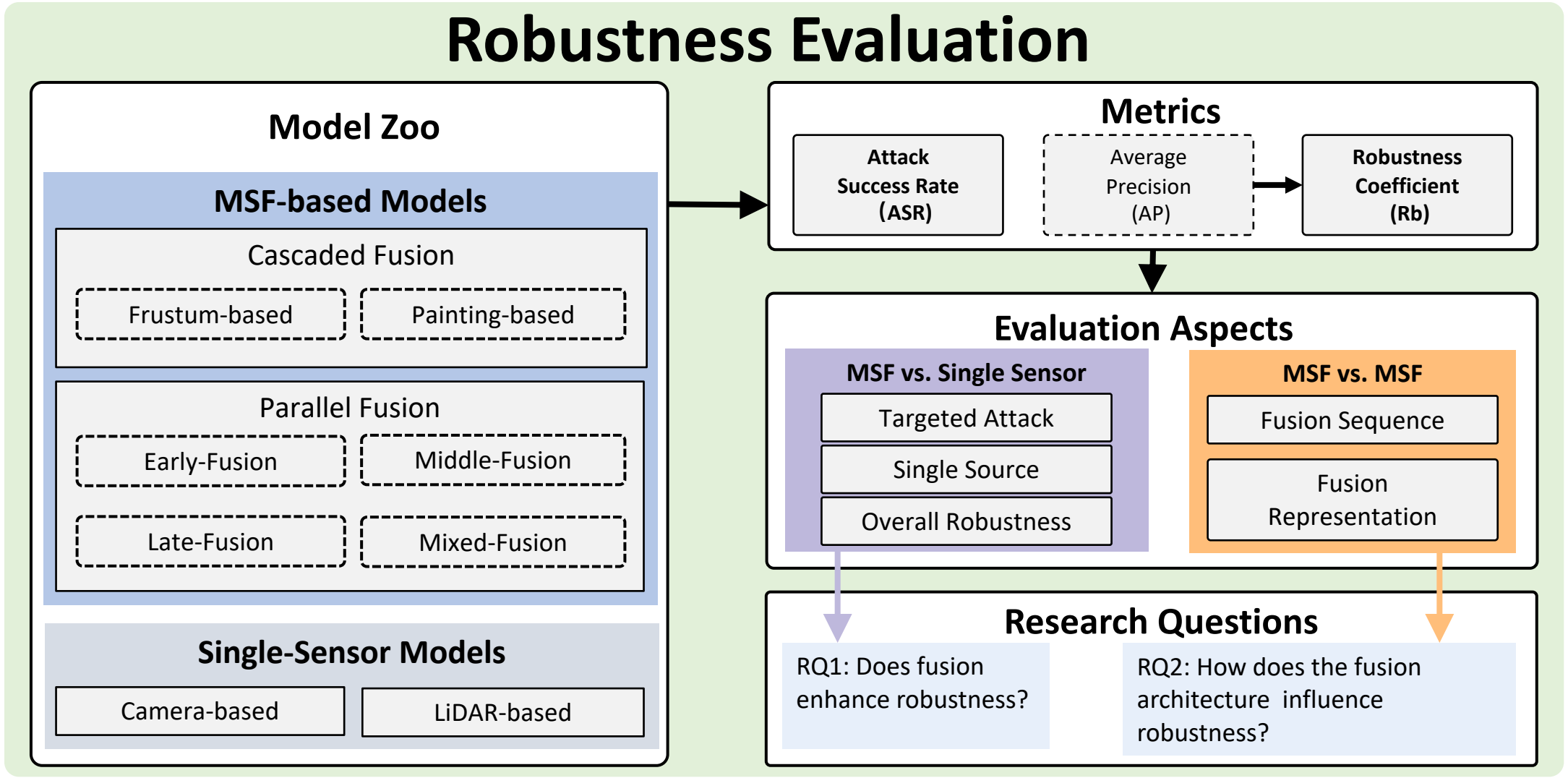
Manipulation Abilities

$< \ell_0$: *Scope*, ℓ_2: *pixel/point manipulation*>

Benchmark Design



Robustness Evaluation



Robustness Evaluation – Setup: Metrics

➢ Average Precision (AP):

- AP describes the overall performance of one model on a dataset:

➢ Robustness (*Rb*):

- We define the Robustness of one model on a corruption c as $\boldsymbol{Rb_c}$:

$$Rb_c = \frac{mAP_C}{mAP_{Clean}}$$

Robustness Evaluation – Overall Results

Table : The Robustness(Rb) of 5 single-modality and 7 MSF-based Detectors on Kitti-Spoof.

Target Sensor	Corruption	Camera-only		LiDAR-only			Fusion Model						
		ImVoxelNet	SMOKE	Second	PointPillar	3DSSD	F-PointNet	PointPainting	VirConv_L	VirConv_T	EPNet	AVOD	CLOCs
Camera	[Hide] Laser - Saturation	0.415	0.069	/	/	/	0.226	0.402	**0.999**	0.995	0.804	0.592	0.315
	[Create] Light- Projection	0.668	0.852	/	/	/	0.467	0.973	0.999	**1.000**	0.999	0.995	0.984
	Laser - Color Strip Injection	0.520	0.203	/	/	/	0.962	0.832	**0.999**	0.993	0.797	0.752	0.993
	EM - Color Strip	0.549	0.749	/	/	/	0.947	0.916	0.967	**0.985**	0.891	0.790	0.992
	EM - Truncation	0.010	0.000	/	/	/	0.080	0.330	**0.999**	0.933	0.782	0.404	0.320
	Ultrasound - Blur	0.001	0.000	/	/	/	0.386	0.330	**0.967**	0.958	0.790	0.411	0.636
LiDAR	[Hide] Laser - Point Erase	/	/	0.655	0.645	0.661	0.597	0.638	0.653	0.676	**0.683**	0.611	0.665
	[Create] Laser - Object Injection	/	/	0.781	0.778	0.767	0.775	**0.890**	0.793	0.796	0.707	0.830	0.889
	Laser - Arbitrary Point Injection	/	/	0.893	0.873	0.894	0.784	0.892	0.890	**0.910**	0.884	0.875	0.888
	Laser - Background Noise Injection	/	/	0.814	0.855	0.742	0.516	0.898	0.854	0.922	0.729	0.839	**0.959**
	EM - Point Interference	/	/	0.979	0.987	0.981	0.960	0.985	0.971	**0.994**	0.993	1.001	0.993
Mean Robustness on Camera Cor. (mRb^C)		0.359	0.312	/	/	/	0.511	0.630	**0.988**	0.977	0.844	0.657	0.707
Mean Robustness on LiDAR Cor. (mRb^L)		/	/	0.825	0.827	0.809	0.726	0.861	0.824	0.850	0.799	0.831	**0.879**
Mean Robustness on All Corruptions (mRb)		0.650	0.625	0.920	0.922	0.913	0.609	0.735	0.918	**0.923**	0.824	0.737	0.785

Robustness Evaluation - RQ1. Does fusion enhance robustness?

Target Sensor	Corruption	Camera-only		LiDAR-only			Fusion Model						
		ImVoxelNet	SMOKE	Second	PointPillar	3DSSD	F-PointNet	PointPainting	VirConv_L	VirConv_T	EPNet	AVOD	CLOCs
Mean Robustness on Camera Cor. (mRb^C)		0.359	0.312	/	/	/	0.511	0.630	**0.988**	0.977	0.844	0.657	0.707
Mean Robustness on LiDAR Cor. (mRb^L)		/	/	0.825	0.827	0.809	0.726	0.861	0.824	0.850	0.799	0.831	**0.879**
Mean Robustness on All Corruptions (mRb)		0.650	0.625	0.920	0.922	0.913	0.609	0.735	0.918	**0.923**	0.824	0.737	0.785

1. Single Source Robustness: average robustness of a model when facing attacks on a **single sensor**.

(RQ1) Observation1 : Regarding single source robustness, in comparison to the camera-based model, all MSF-based models (7/7) enhance robustness against camera attacks. Relative to LiDAR-based models, the majority of MSF-based models (5/7) demonstrate increased robustness against LiDAR attacks.

2. Overall Robustness: average robustness under **all corruptions** in this benchmark.

(RQ1) Observation 2: : Regarding single source robustness, in comparison to the camera-based model, 6/7 fusion models enhance robustness. Conversely, 6/7 fusion models **do not** enhance overall robustness when compared to LiDAR-based Models.

taxonomy

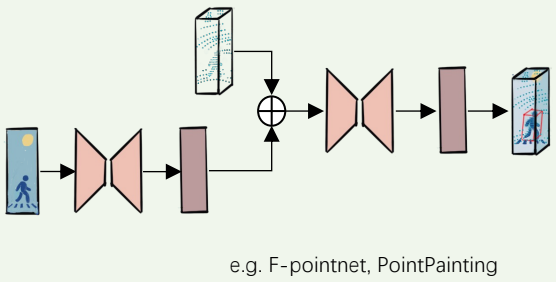
Robustness Evaluation

RQ2 How does the architecture of the fusion model influence robustness?

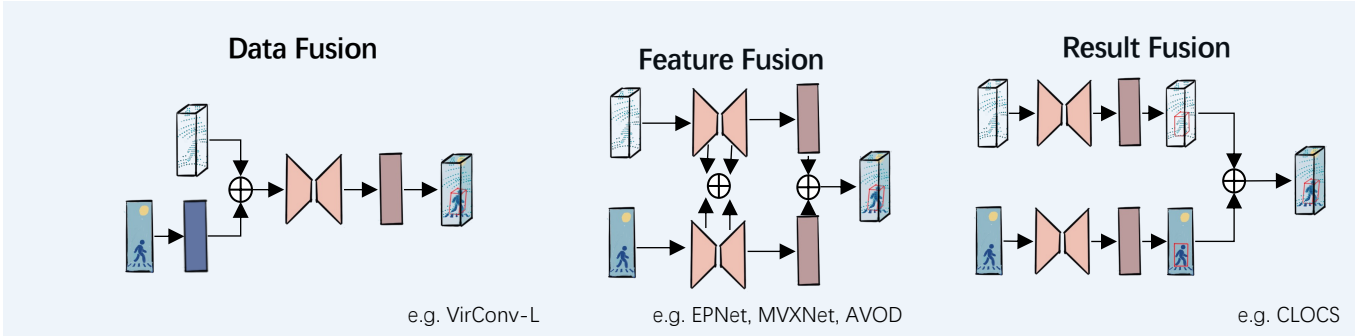
1. Model Architecture Analysis: Fusion Sequence Classification and Information Entropy Sorting

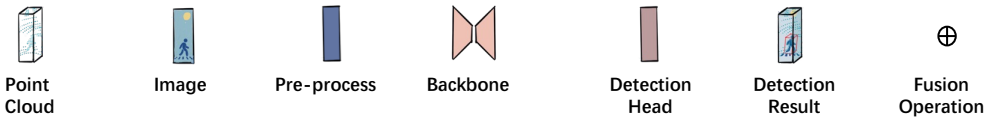
	F-Pointnet	Pointpainting	CLOCs	AVOD	EPNet	VirConv-L	VirConv-T
Fusion Sequence							
Fusion Rep. - Camera							
Fusion Rep. - LiDAR							

Cascaded Fusion



Parallel Fusion





taxonomy

Robustness Evaluation

RQ2 How does the architecture of the fusion model influence robustness?

1. Model Architecture Analysis: Fusion Sequence Classification and Information Entropy Sorting

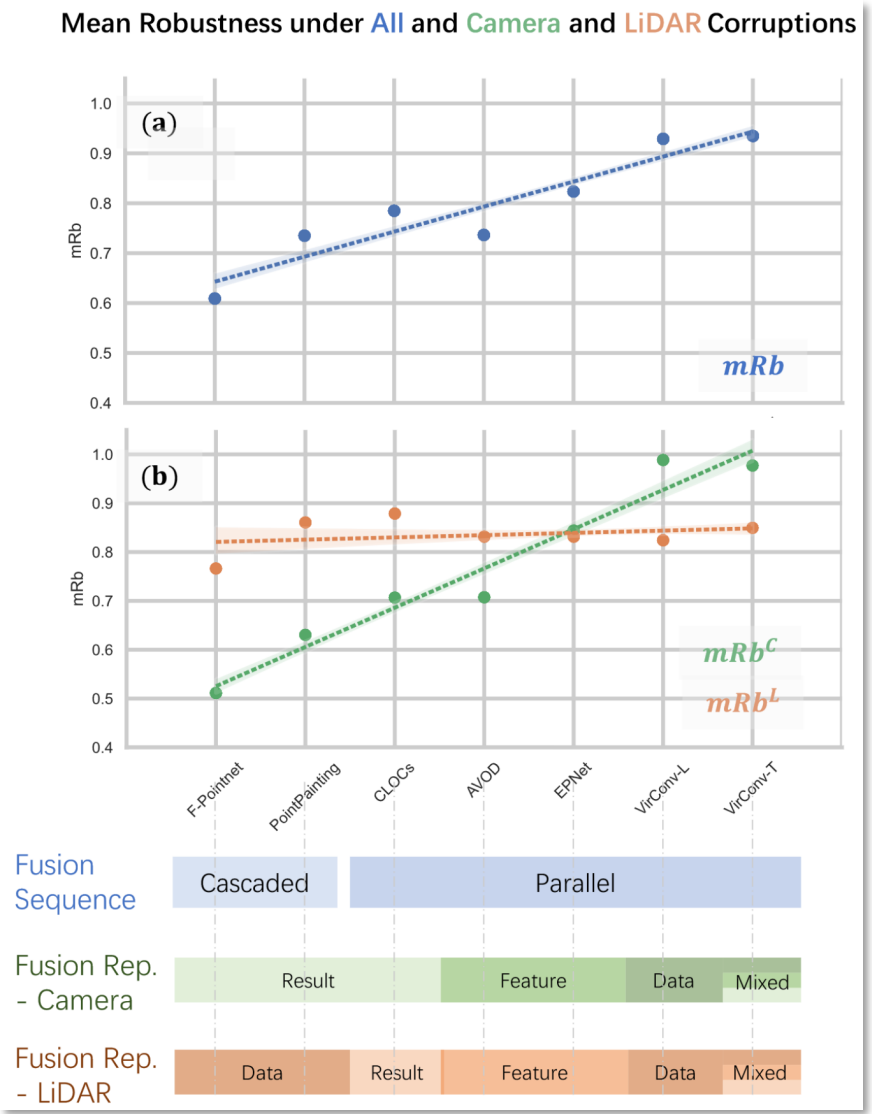
	F-Pointnet	Pointpainting	CLOCs	AVOD	EPNet	VirConv-L	VirConv-T
Fusion Sequence	Cascaded	Cascaded	Parallel	Parallel	Parallel	Parallel	Parallel
Fusion Rep. - Camera	Result (bounding box)	Result (Sementic)	Result (bounding box)	Feature	Feature	Data	Data,Feature, Result
Fusion Rep. - LiDAR	Data (partial)	Data	Result (bounding box)	Feature (BEV)	Feature	Data	Data,Feature, Result

- **Information Entropy[1] (H)** : the average level of information content
  - $H_{data} > H_{feature} > H_{result}$
- **Information Entropy of Fusion Representation ($H_{FR}(M)$)**
  - Cacaded Fusion: $H_{FR}(Pointpainting) > H_{FR}(F-Pointnet)$ ---------- **[1]**
  - Parallel Fusion: $H_{FR}(VirConv-T) > H_{FR}(VirConv-L) > H_{FR}(EPNet) > H_{FR}(AVOD)$, $H_{FR}(CLOCs)$----**[2]**

[1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. Deep learning. MIT press.

Robustness Evaluation

RQ2 How does the architecture of the fusion model influence robustness?



Answer to RQ2.

(RQ2) **Observation1** : From the perspective of the fusion sequence, parallel fusion exhibits better robustness than cascaded fusion.

(RQ2) **Observation2** : In general, given the same fusion sequence, the more comprehensive the information contained in the fused representation, the stronger the robustness. The comprehensiveness of information is ranked as data > feature > results..

(RQ2) **Observation3** : Different fusion architectures primarily influence the robustness to camera corruption

Contribution

- **Benchmark.** We present a large-scale robustness benchmark for MSF-based 3D object detection under physical sensor attack, namely Kitti-Spoof. The dataset contains 11 corruptions induced by laser, EMI, and acoustic.

- **Empirical Evaluation.** Based on the benchmark, we perform a large-scale (542,736 frames) empirical study to evaluate the sensor attack robustness on 7 MSF-based detectors and 5 single-modality detectors with different architectures.

- **Insights for Critical Research Question.** This paper systematically answers the fundamental and critical questions related to the robustness of MSF-based models. Additionally, we provide insights for enhancing MSF robustness.

Unity is Strength? Benchmarking the Robustness of Fusion-based 3D Object Detection against Physical Sensor Attack







Github:
https://github.com/Jinzizhisir/PSA-Fusion

Corresponding Authors:
Xiaoyu Ji: xji@zju.edu.cn

USSLAB Website: www.usslab.org

What We do?

☐ Key Research Questions:

- RQ 1: Does fusion enhance robustness or not?
- RQ 2: How does the architecture of the fusion model influence robustness?

☐ Efforts to Address the Research Questions :

- Build a Robustness Evaluation Benchmark (Datasets, Metrics, etc.)
- Conprehensive Experimental Evaluation