alundber@qualcomm

Adversarial Image Attacks Against Automotive Systems

Andreas Lundberg

Safety Lead
Qualcomm

Autoliv-Veoneer-Arriver-Qualcomm

- Automotive safety since 1953
- Vision systems since 2005
- Passenger cars up to L3
- Currently:
  - Delivering camera systems to MB, VCC, Geely, GAC, BYD, Chery, BTET, and GM
  - In development agreements with BMW, Renault, and others

Source sample text

Agenda

- What, when, why?
- Technology overview
- Risk management

Adversarial image attacks against automotive systems

Disrupting ADAS or AD vehicle operation

Fools vision perception

Manipulation of the environment

Not disruptive to humans

Source sample text

Scenarios

Adhesive markings on street signs

- Disrupts sign detections
- Misclassifying street signs

Painting road surfaces

- Misleading lane markings

Markings or designs on vehicles

- Disrupts vehicle detections
- Misclassifying vehicle

Patches placed near road

- Appears as street sign or traffic participant
- Disrupts detections of signs, traffic, or lanes

Imaged by Heritage Auctions, HA.com

Motivations and Actors

Economic

- Competitors

Political

- Terrorists
- State actors

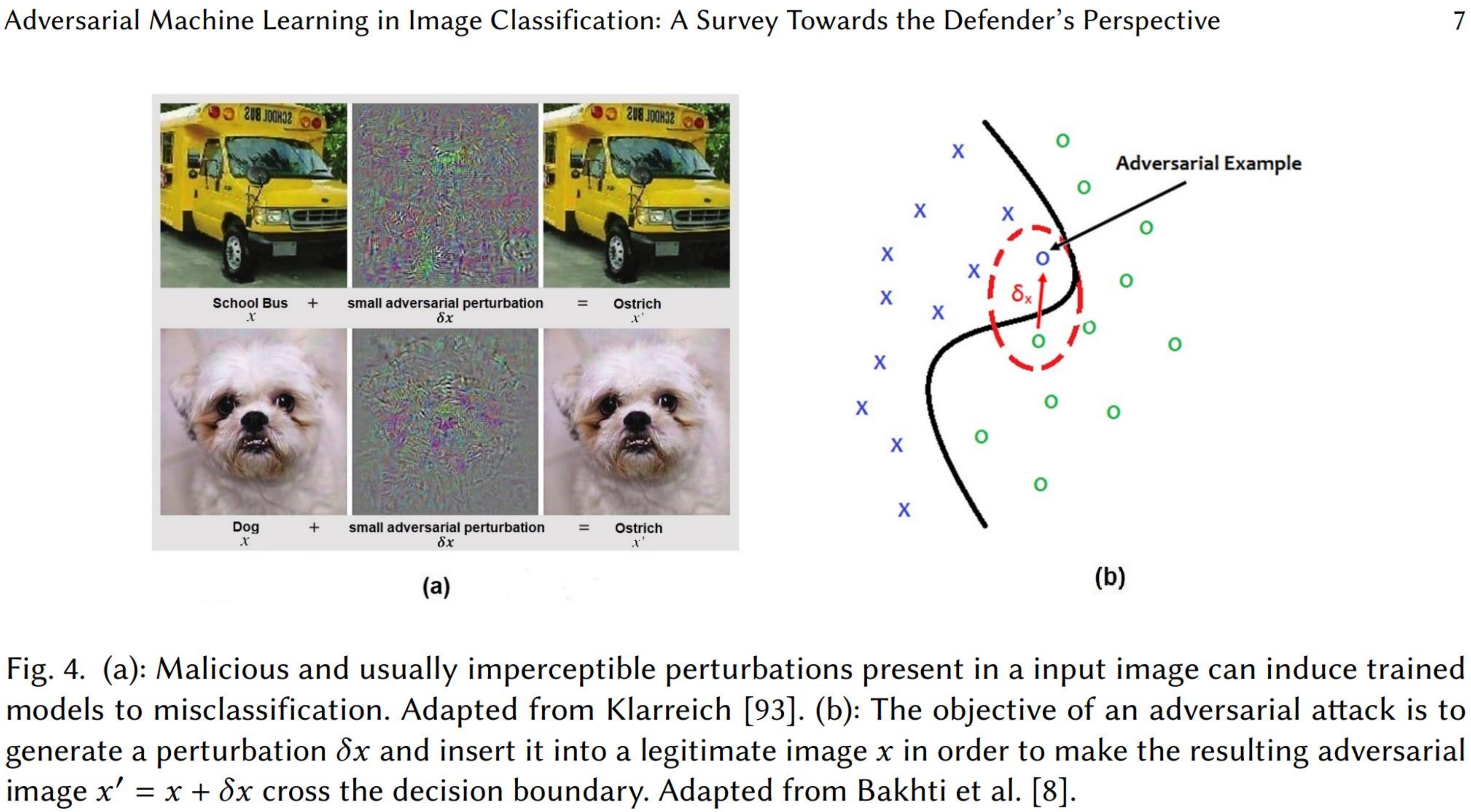
Environmental

- Activists

Source sample text

Technology

Misclassification



Adversarial Machine Learning in Image Classification: A Survey Towards the Defender's Perspective 7

Fig. 4. (a): Malicious and usually imperceptible perturbations present in a input image can induce trained models to misclassification. Adapted from Klarreich [93]. (b): The objective of an adversarial attack is to generate a perturbation δx and insert it into a legitimate image x in order to make the resulting adversarial image $x' = x + \delta x$ cross the decision boundary. Adapted from Bakhti et al. [8].

https://arxiv.org/pdf/2009.03728.pdf

Technology

Disrupting detections

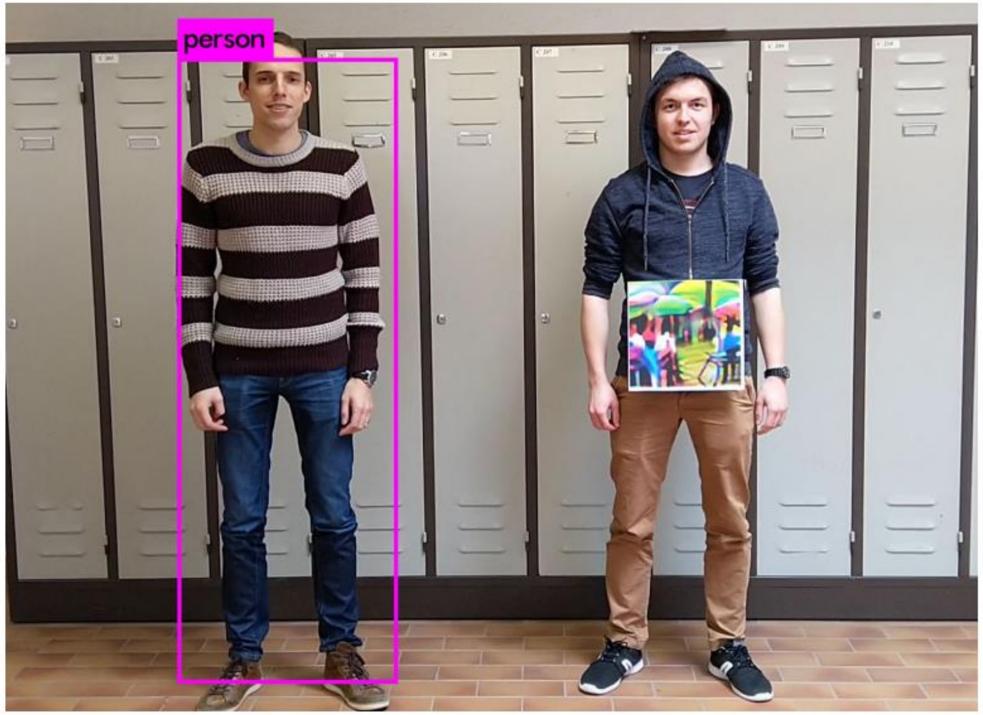


Figure 1: We create an adversarial patch that is successfully able to hide persons from a person detector. Left: The person without a patch is successfully detected. Right: The person holding the patch is ignored.

https://arxiv.org/pdf/1904.08653.pdf

Technology

Billboards

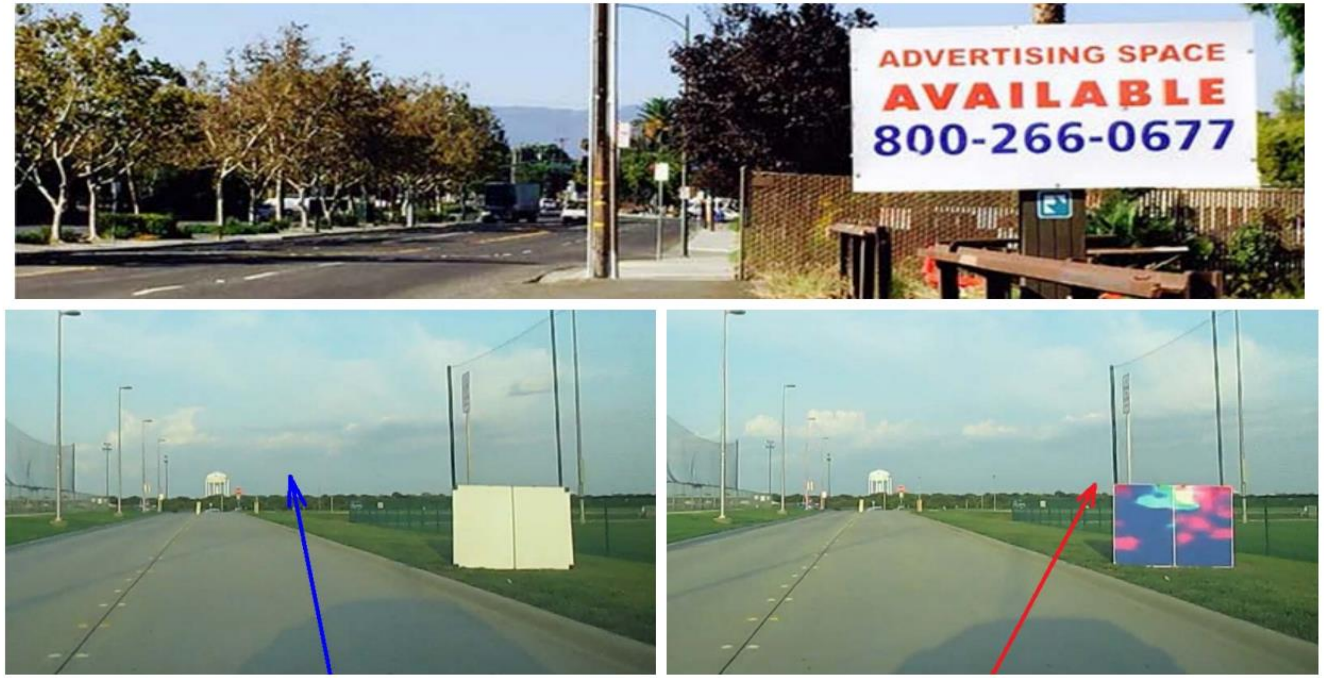
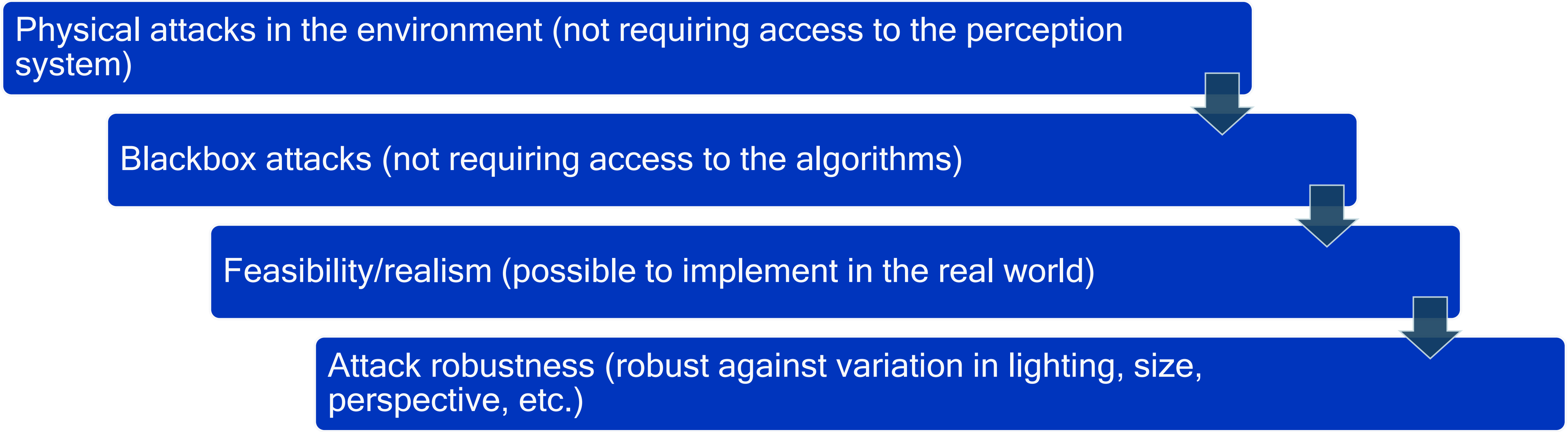


Fig. 1: The top subfigure shows an example customizable roadside billboard. The bottom two subfigures show an adversarial billboard example, where the Dave [3] steering model diverges under our proposed approach.

https://arxiv.org/pdf/1812.10812.pdf

Technology



Physical attacks in the environment (not requiring access to the perception system)

Blackbox attacks (not requiring access to the algorithms)

Feasibility/realism (possible to implement in the real world)

Attack robustness (robust against variation in lighting, size, perspective, etc.)

Source sample text

Focus on attack methods with these features

Risk Management

SOTIF
ISO 21488

SAFETY AND AI
ISO/PAS 8800

SAFETY FOR DRIVING AUTOMATION SYSTEMS
ISO/TS 5083

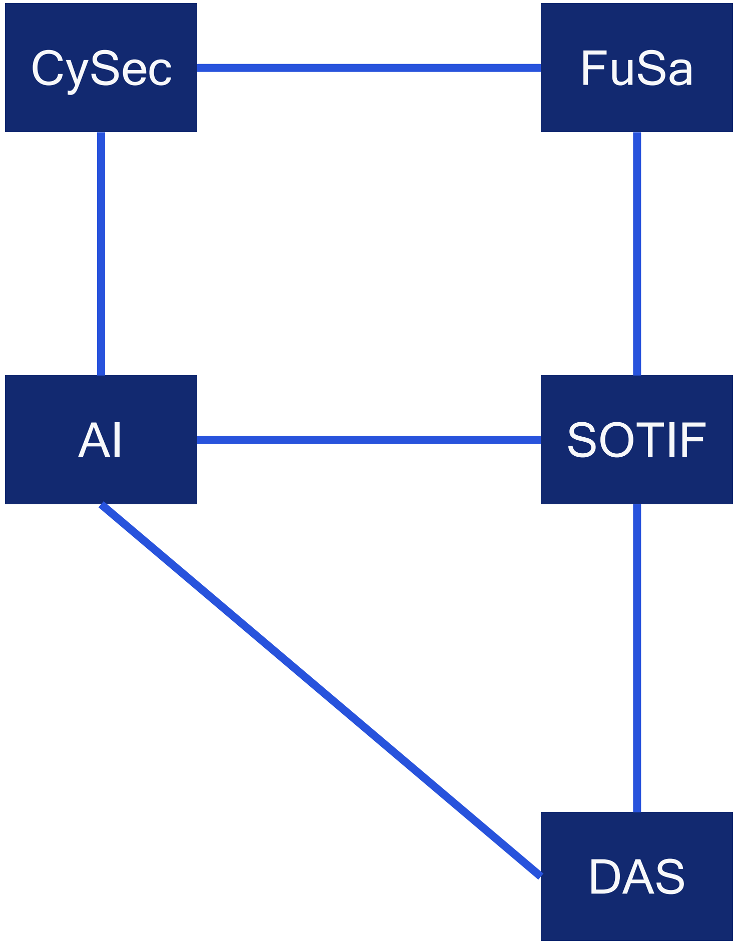
CYBERSECURITY
ISO 21434,
ISO/TR 4804

FUNCTIONAL SAFETY
ISO 26262

Source sample text

Which standards apply?

Standards Landscape



Threat management
Continual monitoring

CySec

FuSa

Safety management

Machine Learning

AI

SOTIF

Analysis of reasonably foreseeable misuse,
Including adversarial attacks

DAS

KPI’s at vehicle level, covering adversarial robustness

Source sample text

Risk Mitigation

Cybersecurity

- To prevent white-box attacks: Attack path analysis, vulnerability analysis, risk treatment
- Continual monitoring for both cybersecurity but also adversarial image attacks

ISO/PAS 8800 and Assurance of Machine Learning for use in Autonomous Systems (AMLAS)

- To prevent ground-truth attacks and poisoning attacks
- To prevent black-box attacks

ISO/TS 5083

- For robustness against black-box attacks

ISO 21488

- For resilience
- Analysis of hazards, triggering conditions, etc

See AMLAS - Assuring Autonomy International Programme, University of York

Defenses

Proactive

Adversarial training
Defensive distillation
Model ensemble
Network regulatization
Certified robustness
…

Reactive

Adversarial detection
Adversarial transformation
…

Source sample text

Risk Mitigation

Stay updated on the literature

- https://www.researchgate.net/profile/Pan-He-9/publication/321936593_Adversarial_Examples_Attacks_and_Defenses_for_Deep_Learning/links/5a5cc59e0f7e9b4f7839614f/Adversarial-Examples-Attacks-and-Defenses-for-Deep-Learning.pdf
- https://arxiv.org/pdf/2009.03728.pdf
- https://arxiv.org/pdf/2104.01789.pdf
- https://arxiv.org/pdf/1707.02476.pdf
- https://arxiv.org/pdf/1801.09344.pdf
- …

Qualcomm

Stay Safe

Follow us on: f 🐦 in ◎

For more information, visit us at:

www.qualcomm.com & www.qualcomm.com/blog

Nothing in these materials is an offer to sell any of the components or devices referenced herein.

References in this presentation to “Qualcomm” may mean Qualcomm Incorporated, Qualcomm Technologies, Inc., and/or other subsidiaries or business units within the Qualcomm corporate structure, as applicable. Qualcomm Incorporated includes Qualcomm’s licensing business, QTL, and the vast majority of its patent portfolio. Qualcomm Technologies, Inc.,
a wholly-owned subsidiary of Qualcomm Incorporated, operates, along with its subsidiaries, substantially all of Qualcomm’s engineering, research and development functions, and substantially all of its product and services businesses, including its semiconductor business, QCT.