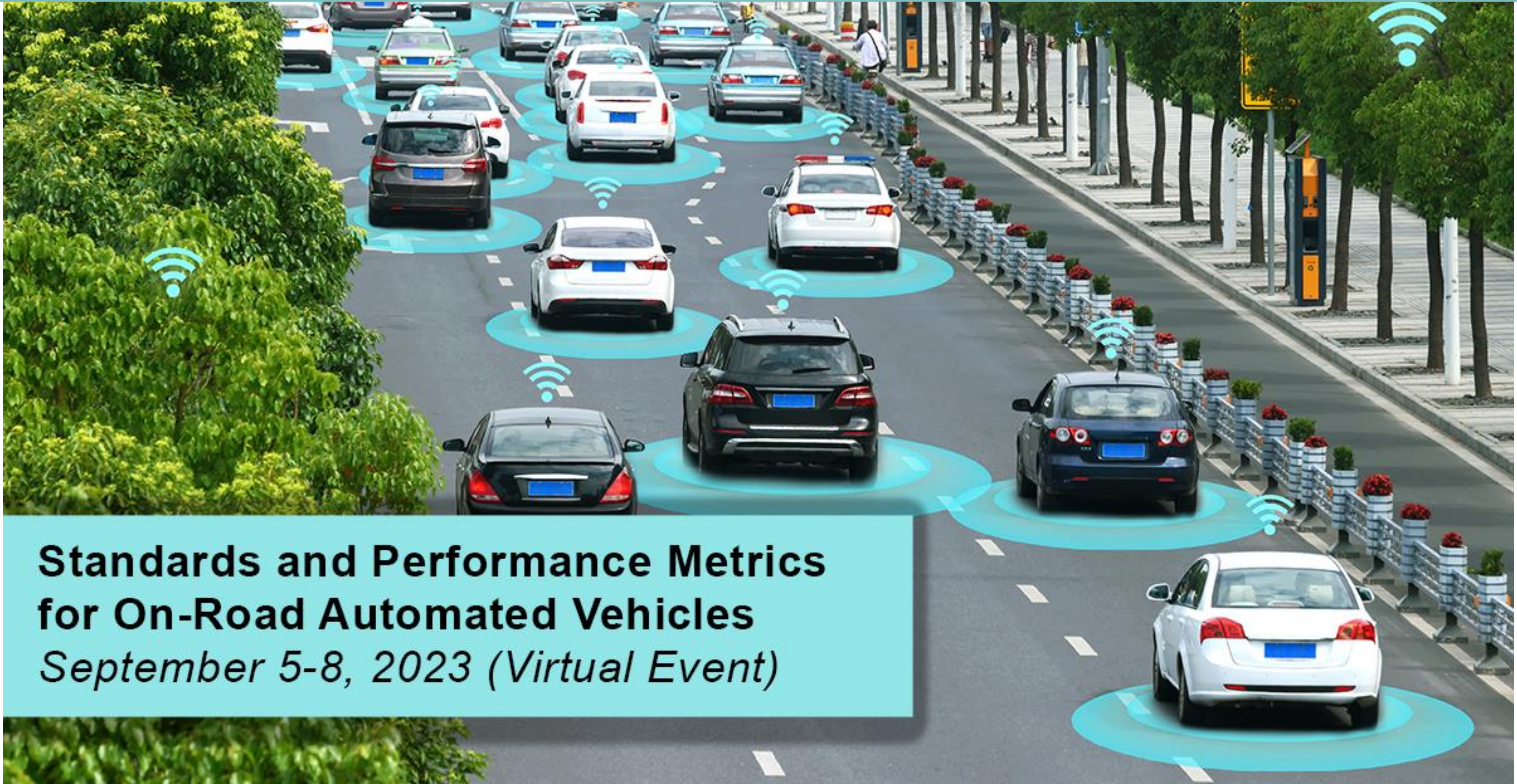
NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Day 3: Robust AI for ADS



Standards and Performance Metrics for On-Road Automated Vehicles

September 5-8, 2023 (Virtual Event)

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Robust AI for ADS

Objective: Improving the robustness and developing mechanisms for technical evaluation of object detection and classification in AI perception systems used in ADS

NIST AI AV Team

- Working together to promote U.S. innovation and industrial competitiveness by advancing **measurement science, standards,** and **technology** in ways that enhance economic security and improve our quality of life

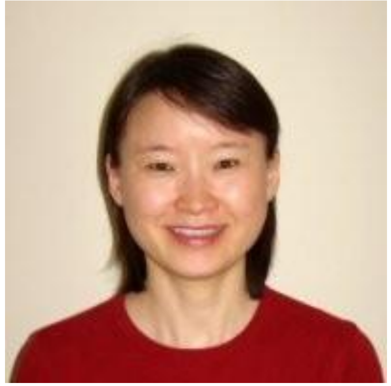

Apostol Vassilev


Edward Griffor


Alina Oprea


Munawar Hasan


Pavel Piliptchak


Honglan Jin

Industry Voices: What did stakeholders request from NIST?

* Within NIST scope and expertise/infrastructure is available
*Within NIST scope and expertise/infrastructure is lacking (NIST can support agencies)
*Not within NIST scope

Develop novel individual and fused sensor *measurement science* solutions for vehicles	*Measure* how different parts of an AV work together
Help define *testing* guidance for stakeholders to meet regulatory agency requirements	"Do you know that NIST cybersecurity framework? Just do that for autonomous vehicles."
Develop mitigation *standards* for adversarial AI	Define the data that should be *measured* before, during, and after operation of automated vehicles
Develop AV simulation-based *measurement science*	Provide *reference materials* for what infrastructure investment state and local governments should invest in
Advance *standards* with SAE, 3GPP, and Teleoperation Consortium	Collect *standardized* data from the DoT from accidents to develop representative testing environments
Develop *measurement science* for traffic infrastructure that can support AVs	Provide classification and levels for AV components
Develop *metrics* to identify what aspects of AVs should be measured to ensure safety	Create and enforce a baseline for AV safety systems testing
Create *test models* and *measurement science* for AV communications	Enforce sensor specs that should be used in Avs
Foster a community of stakeholders to agree on common *taxonomies and standards*	Create regulation on periodic testing and updating
Be a one-stop-shop for pointers to relevant autonomous vehicle *standards*	

ADVERSARIAL ML (AML)

A TAXONOMY OF ATTACKS and MITIGATIONS

A new standard NIST AI 100-2e2023 ipd (Initial Public Draft):

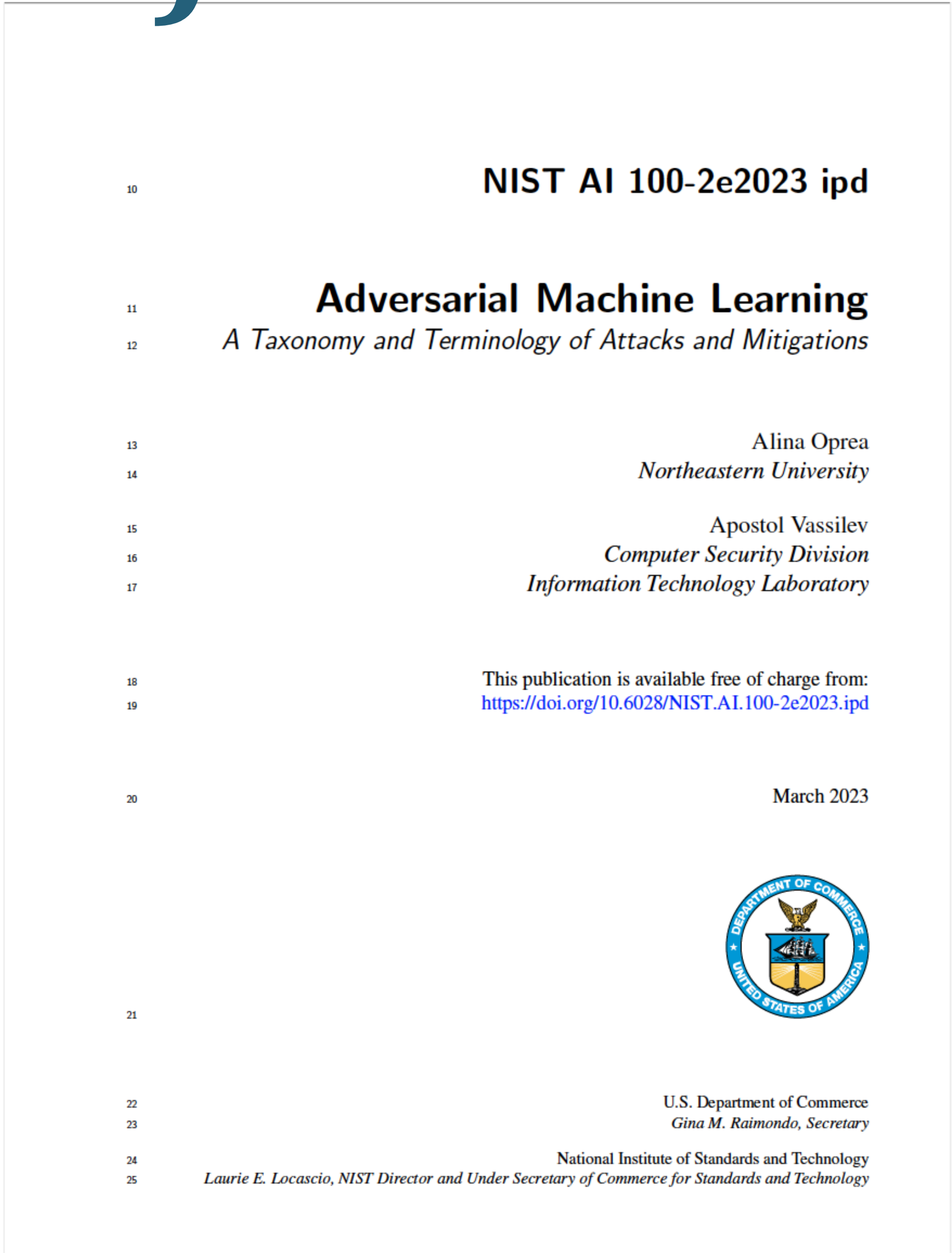
- *Published on March 8, 2023*
- *Comment period closes on September 30, 2023*
- *Will finalize as NIST AI 100-2e2023*

Maintained annually

- *NIST AI 100-2e2024 ipd*
- *NIST AI 100-2e2024*
- *etc.*

NIST is specifically interested in comments on and recommendations for the following topics

- *What are the latest attacks that threaten the existing landscape of AI models?*
- *What are the latest mitigations that are likely to withstand the test of time?*
- *What are the latest trends in AI technologies that promise to transform the industry/society? What potential vulnerabilities do they come with? What promising mitigations may be developed for them?*
- *Is there new terminology that needs standardization?*


NIST AI 100-2e2023 ipd

Adversarial Machine Learning

A Taxonomy and Terminology of Attacks and Mitigations

Alina Oprea
Northeastern University

Apostol Vassilev
Computer Security Division
Information Technology Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.AI.100-2e2023.ipd

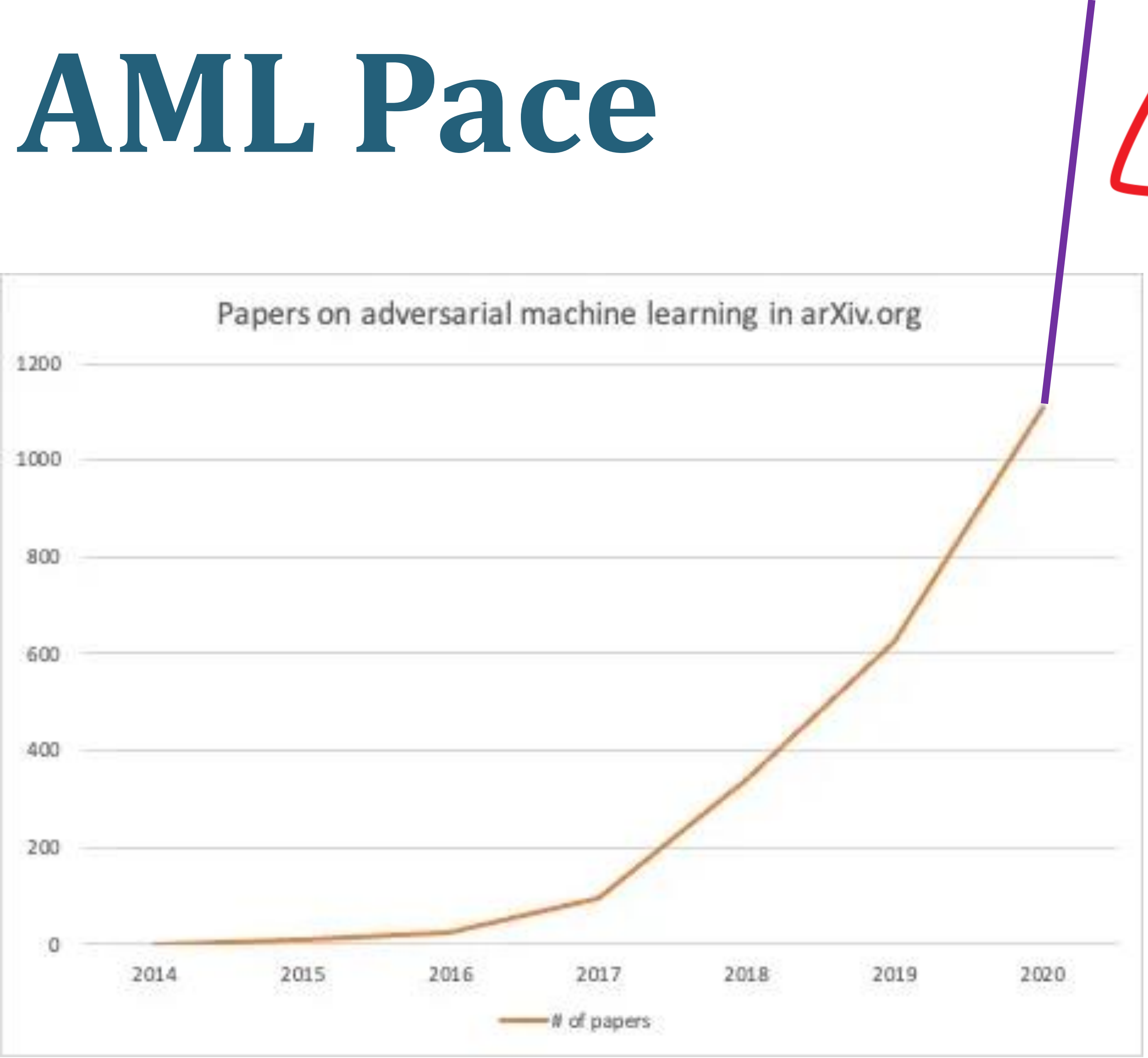
March 2023

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology

AML Pace

A search on arXiv for AML articles in
2021 and **2022**
yielded more than **5,000** references



The advent of generative AI into the public domain this year is driving an enormous growth in attacks against them with only partial mitigations available.

Credit: Ben Dickson
https://www.kdnuggets.com/2021/01/machine-learning-adversarial-attacks.html

Why robustness?

NIST

Omission or misclassification of road objects can lead to crashes or near misses

Image credit: Pavel Vinnik, Shuttershock, Portswigger LTD.

Not all objects in the vision of the car may be what or where they appear to be!

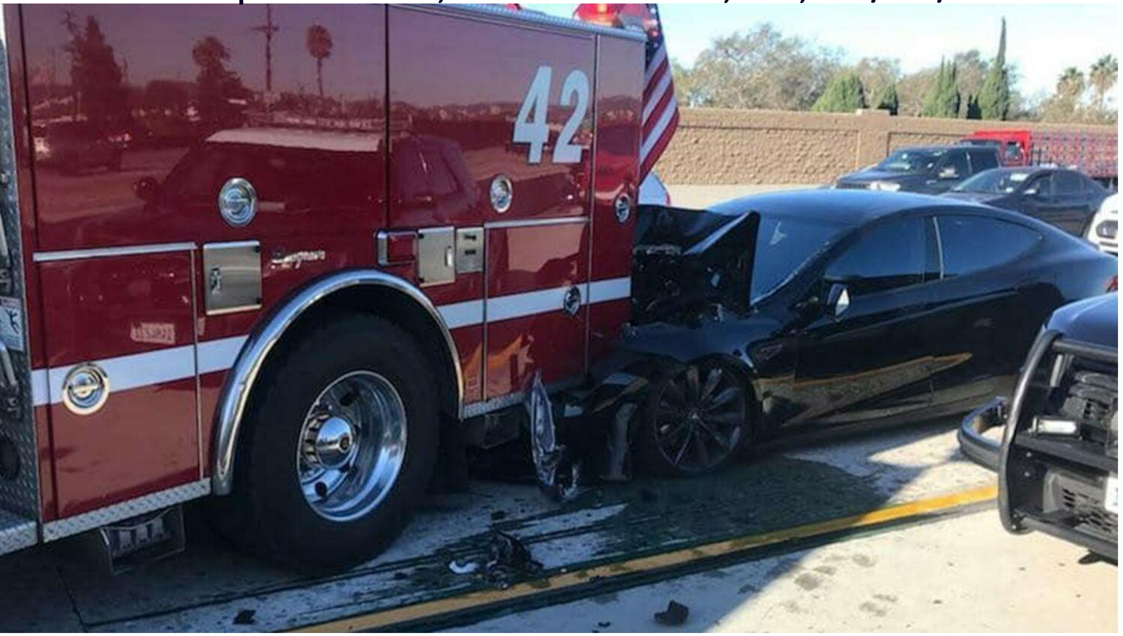
environmental factors: lighting conditions, weather

traffic conditions: road surfaces, object occlusion and object deformation.

malicious attacks: modification of road signs and markings

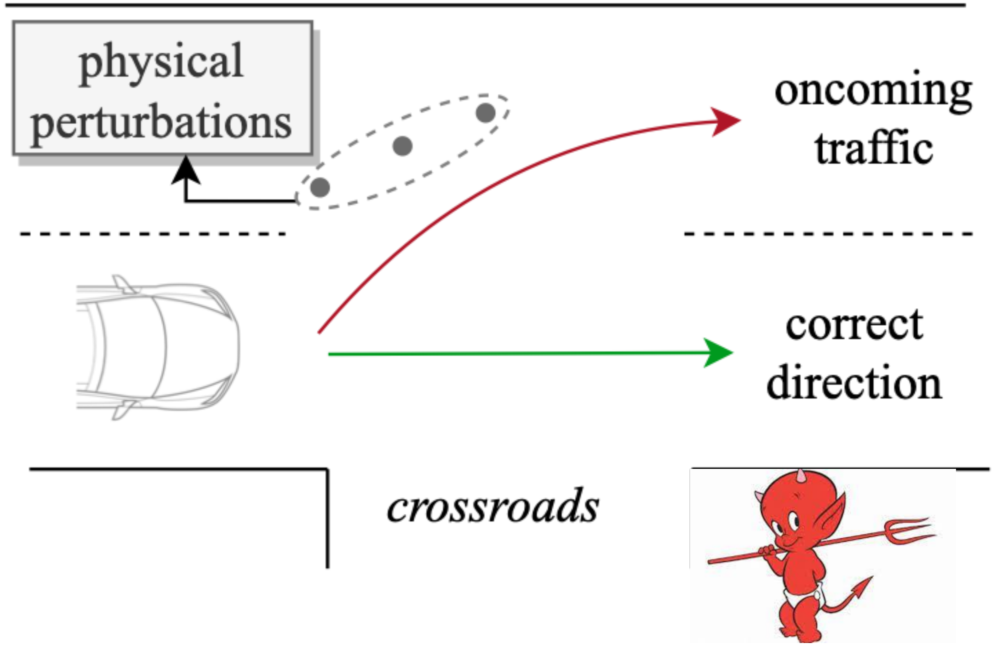
Is AV perception robust?

Autopilot crash, Walnut Creek, CA, 02/18/2023



Physical evasion attack

Credit: Jing et al., [“Too Good to Be Safe: Tricking Lane Detection in Autonomous Driving with Crafted Perturbations”]() USENIX 2021.



NHTSA report: ADS safety record is currently lagging human driver performance for the same number of traveled miles.

Why now?

NIST

WIRED BACKCHANNEL BUSINESS CULTURE GEAR IDEAS MORE SIGN IN SUBSCRIBE

AARIAN MARSHALL BUSINESS AUG 10, 2023 9:51 PM

Robotaxis Can Now Work the Streets of San Francisco 24/7

Robotaxis can offer paid rides in San Francisco around the clock after Alphabet's Waymo and GM's Cruise got approval from the California Public Utilities Commission.

PHOTOGRAPH: SHIIKO ALEXANDER/ALAMY

CNBC Search quotes, news & videos WATCHLIST

MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV INVESTING CLUB PRO MAKE IT SELECT

TECH

Cruise will reduce robotaxi fleet by 50% in San Francisco while California DMV investigates 'incidents'

PUBLISHED SAT, AUG 19 2023·12:36 PM EDT

Kif Leswing
@KIFLESWING

SHARE

Characterizing the robustness and developing mechanisms for technical evaluation of object detection and classification in AI perception systems for ADS is timely and critically important

Robustness improvement techniques

Adversarial Training (AT)

The most robust approach

- Due to Goodfellow et al. in 2015
- Substantially improved by Madry et al. in 2018

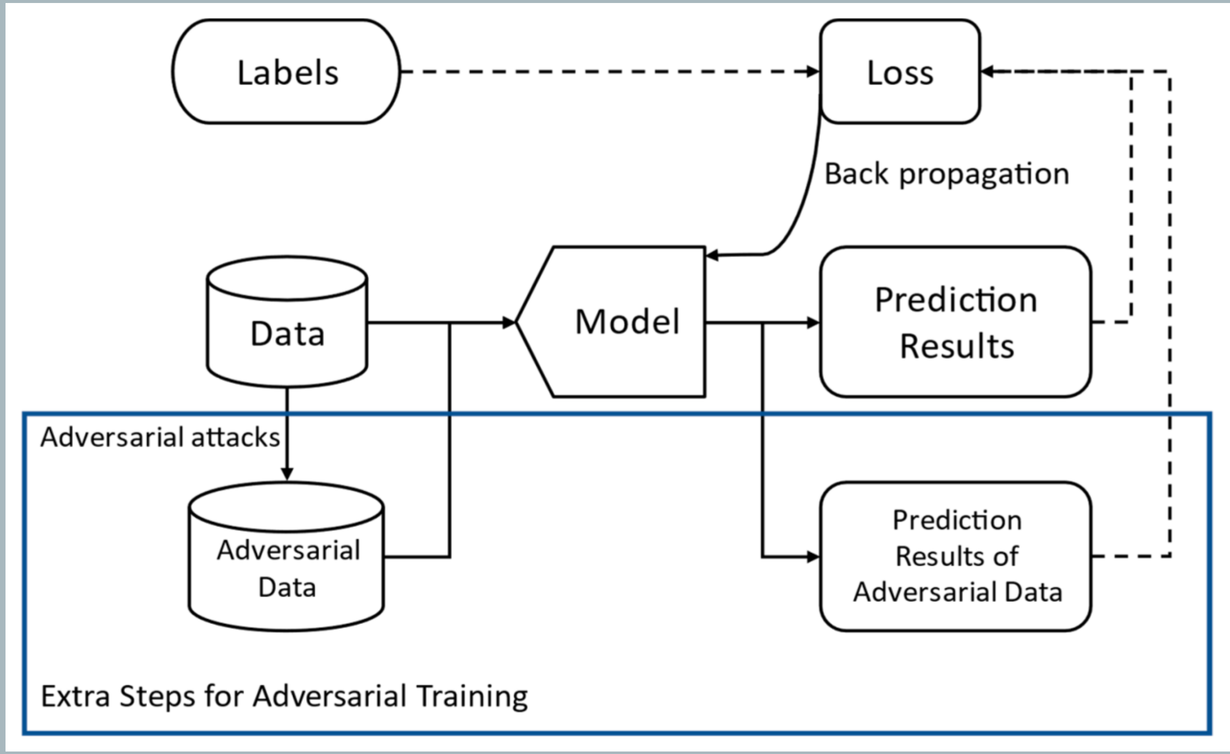


Image credit: Zhao et al., "Adversarial Training Methods for Deep Learning: A Systematic Review", MDPI, 2022.

But,

In automotive setting AT is <u>reactive</u> by construction:
 - not all road conditions leading to incidents are known in advance.

- actual accident data is fed into the training of the next AI model

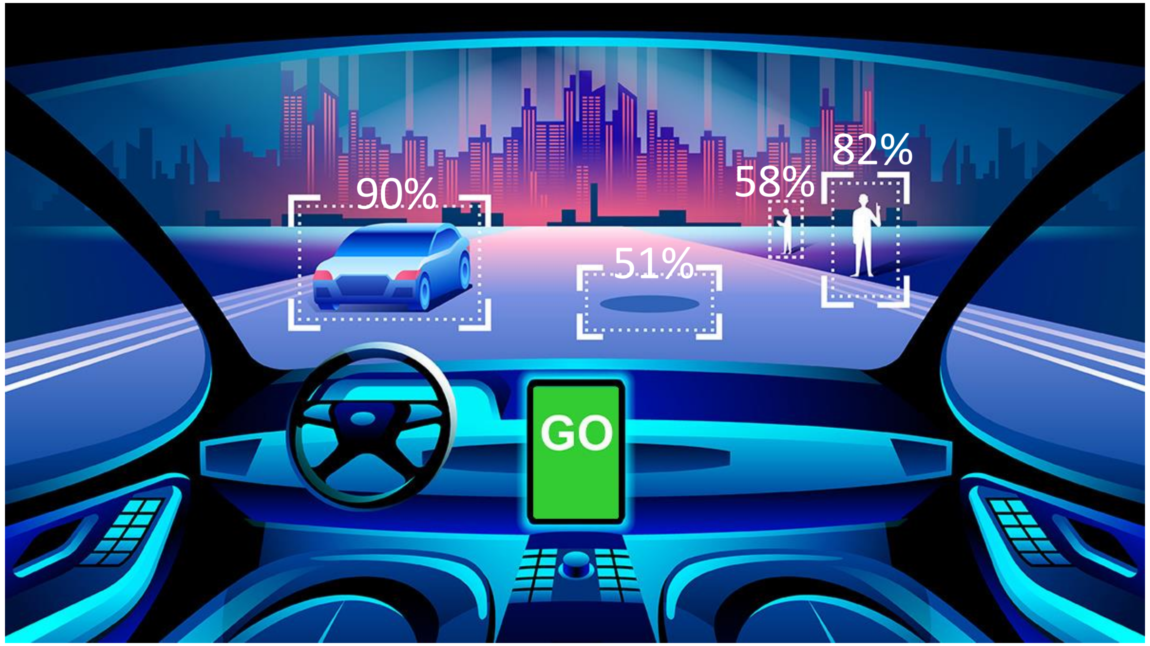
Cognitive task automation!

$\neq$

cognitive intelligence

Our focus – uncertainty estimation

Uncertainty estimates <u>proactively</u> help the car make safe driving decisions in real time



There are <u>two</u> types of uncertainties in machine learning for ADS

<u>**Aleatoric:**</u> a.k.a., statistical uncertainty, refers to refers to the variability in the outcome of an experiment which is due to inherently random effects.

E.g., the atmosphere is a chaotic system, thus atmospheric events impacting the road conditions where the AV operates are a source of aleatoric uncertainty.

E.g., sensor data is noisy

<u>**Epistemic:**</u> a.k.a. systematic uncertainty, refers to deficiencies by a lack of knowledge or information.

E.g., models produced by deep learning ML systems exhibit epistemic uncertainty in the parameters of the model.

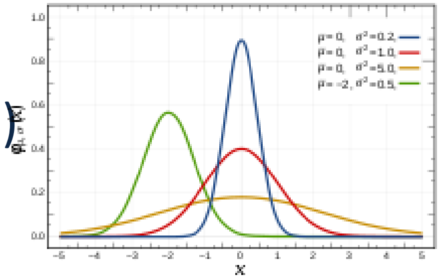
E.g., vandalized street signs – images of these can be fed into AT and the model learns.

Uncertainty estimation in Computer Vision

NIST

Three main approaches in the literature and practice:

1. Gaussian



- models the **bbox** coordinates of an object as Gaussian parameters (μ, σ^2)
- limited overall computational complexity of the algorithm
- improves robustness to noisy data

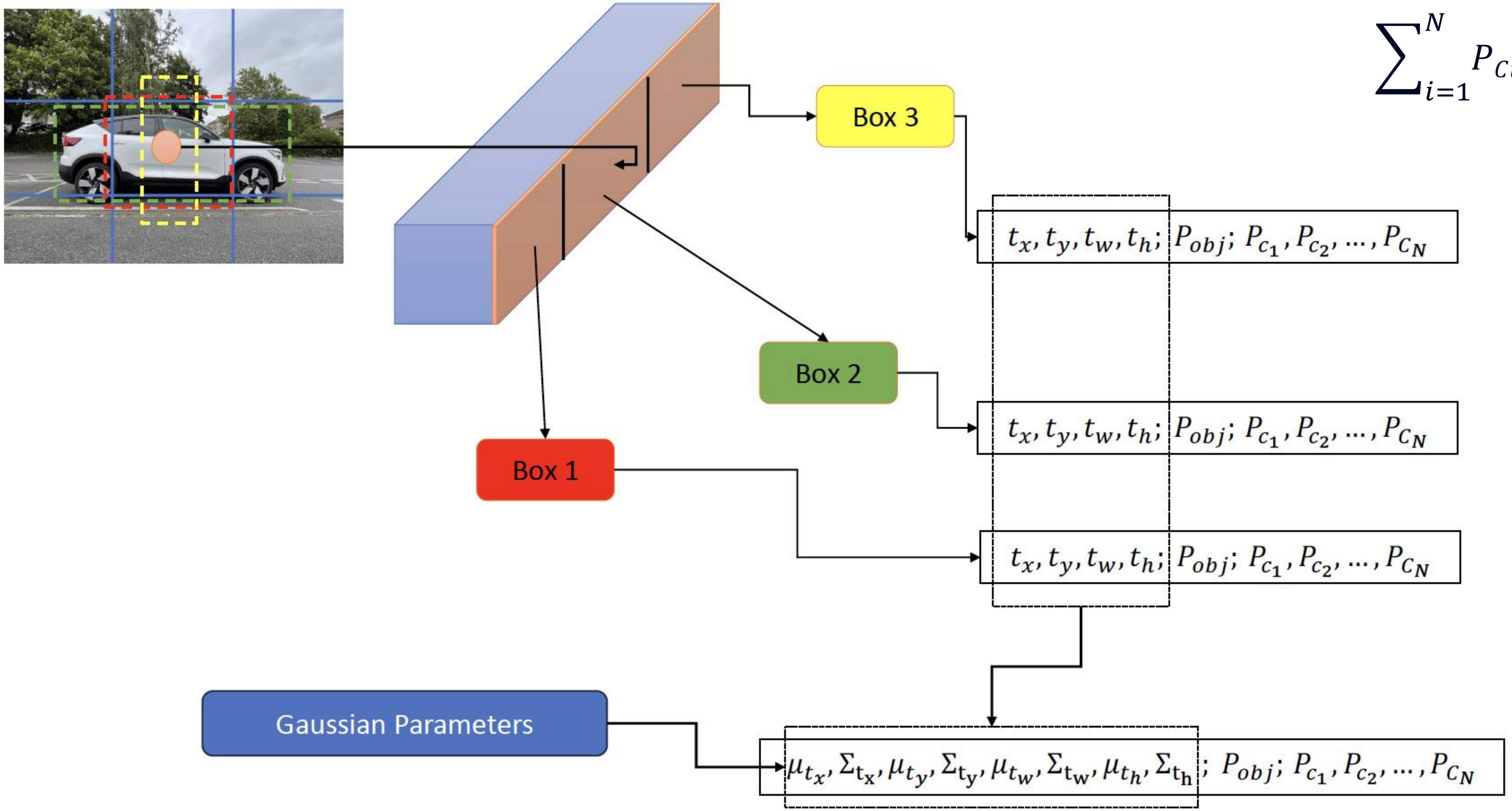
2. Bayesian NN

- learns mappings from input data to aleatoric uncertainty and composes these together with epistemic uncertainty approximations
- implementation with Monte-Carlo dropout in layers of the network (p=0.2, 50 samples)

3. Non-Bayesian ensembles

- random initialization of NN parameters combined with random shuffling of data
- mixture contains ~ 200 NN's
- ensemble prediction is a Gaussian determined by the (μ, σ^2) of the mixture

Gaussian uncertainty modeling



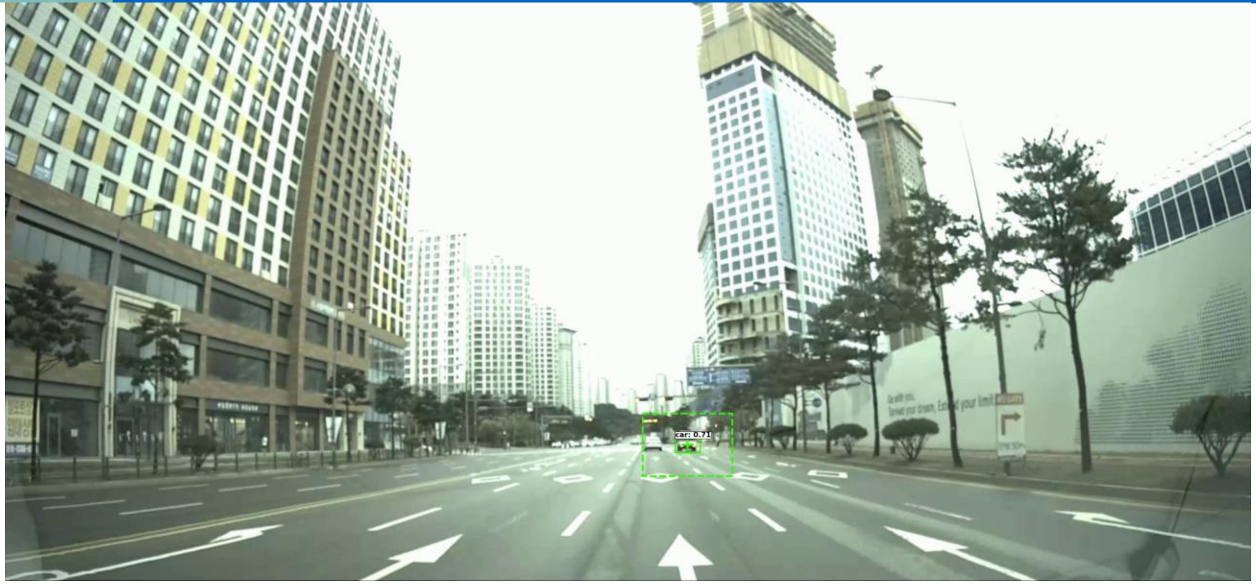
$$\sum_{i=1}^{N} P_{Ci} = 1$$

https://arxiv.org/pdf/1904.04620.pdf

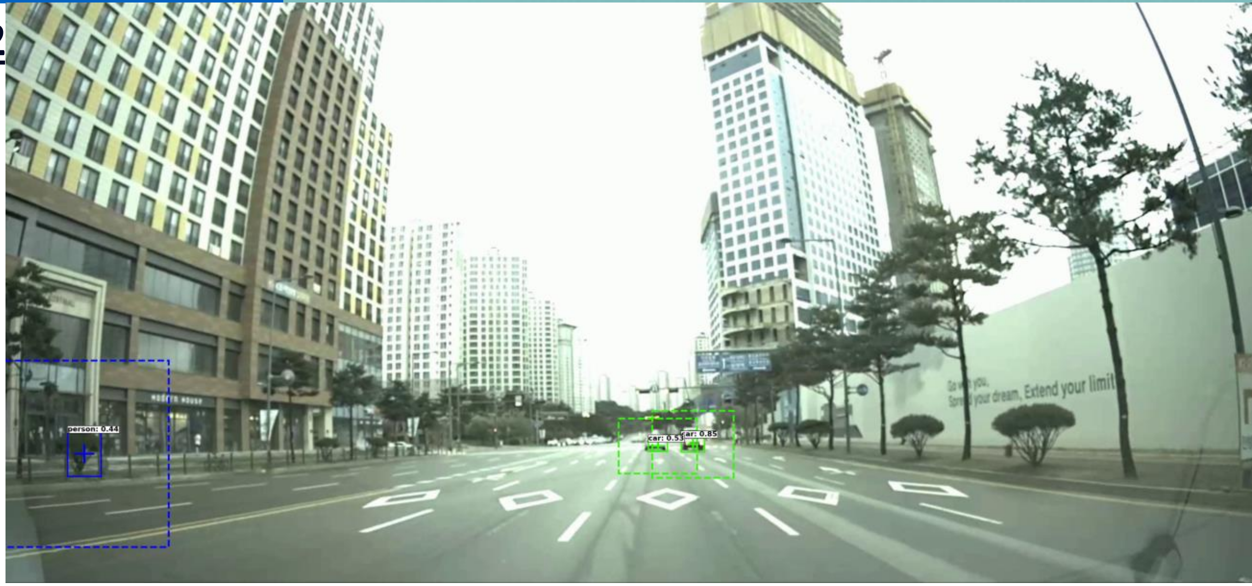
Example: uncertainty estimation w/ Gaussian YOLOv3
https://arxiv.org/pdf/1904.04620.pdf

NIST

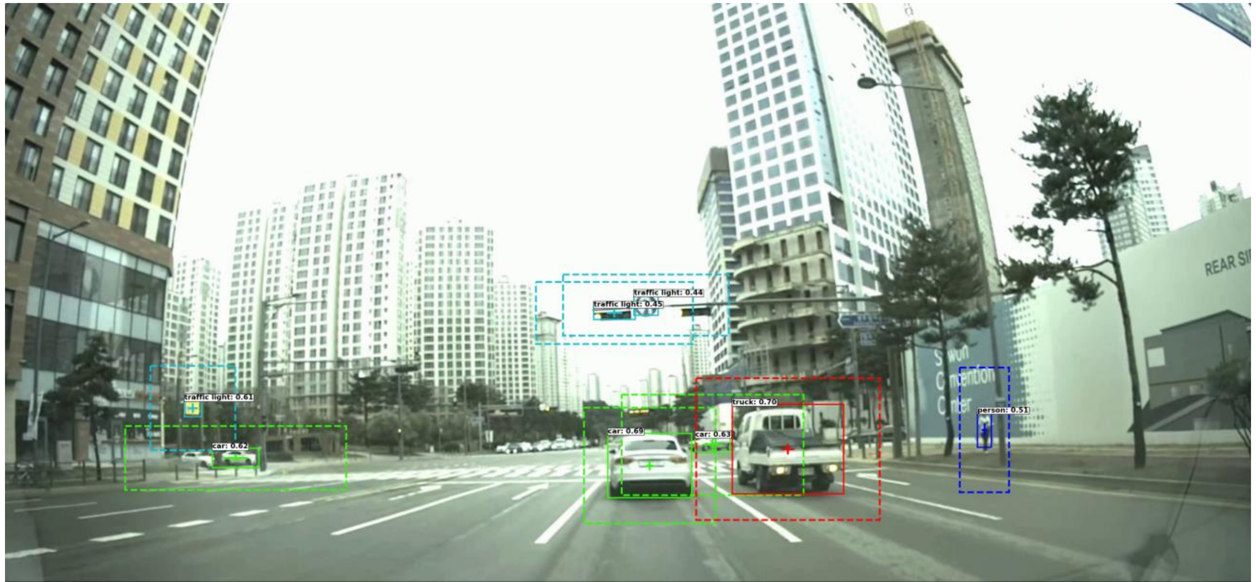
1



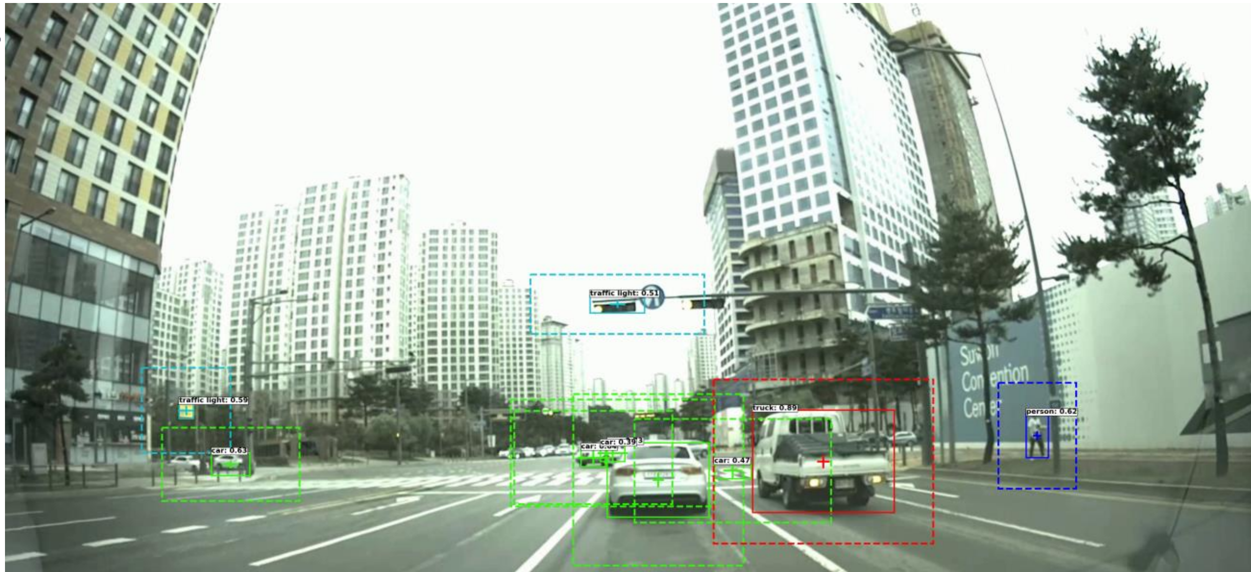
2



3

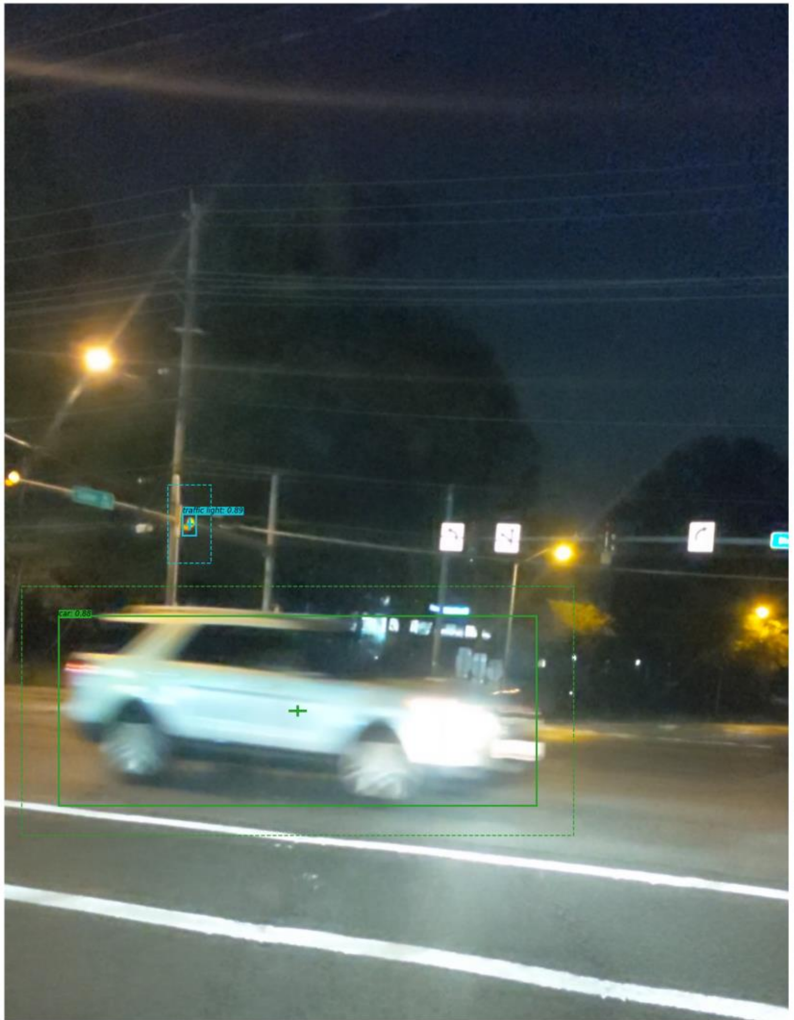


4

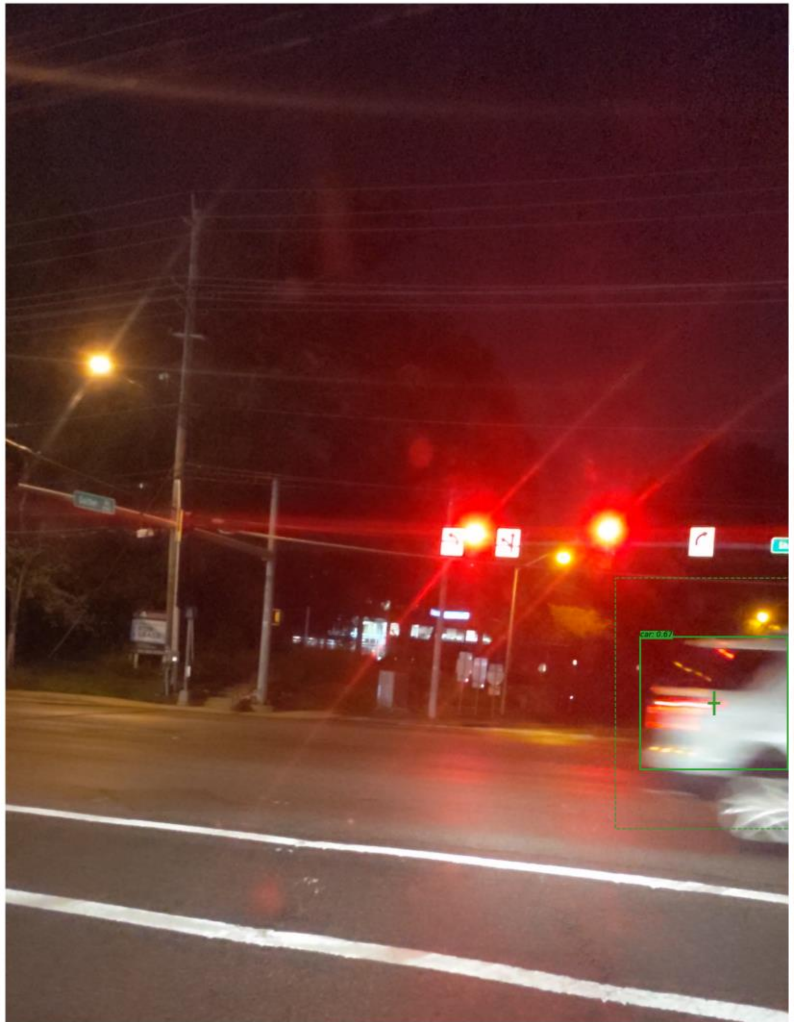


Example: uncertainty estimation with Gaussian YOLOv3, https://arxiv.org/pdf/1904.04620.pdf

running car with motion blur



partial view of car with motion blur

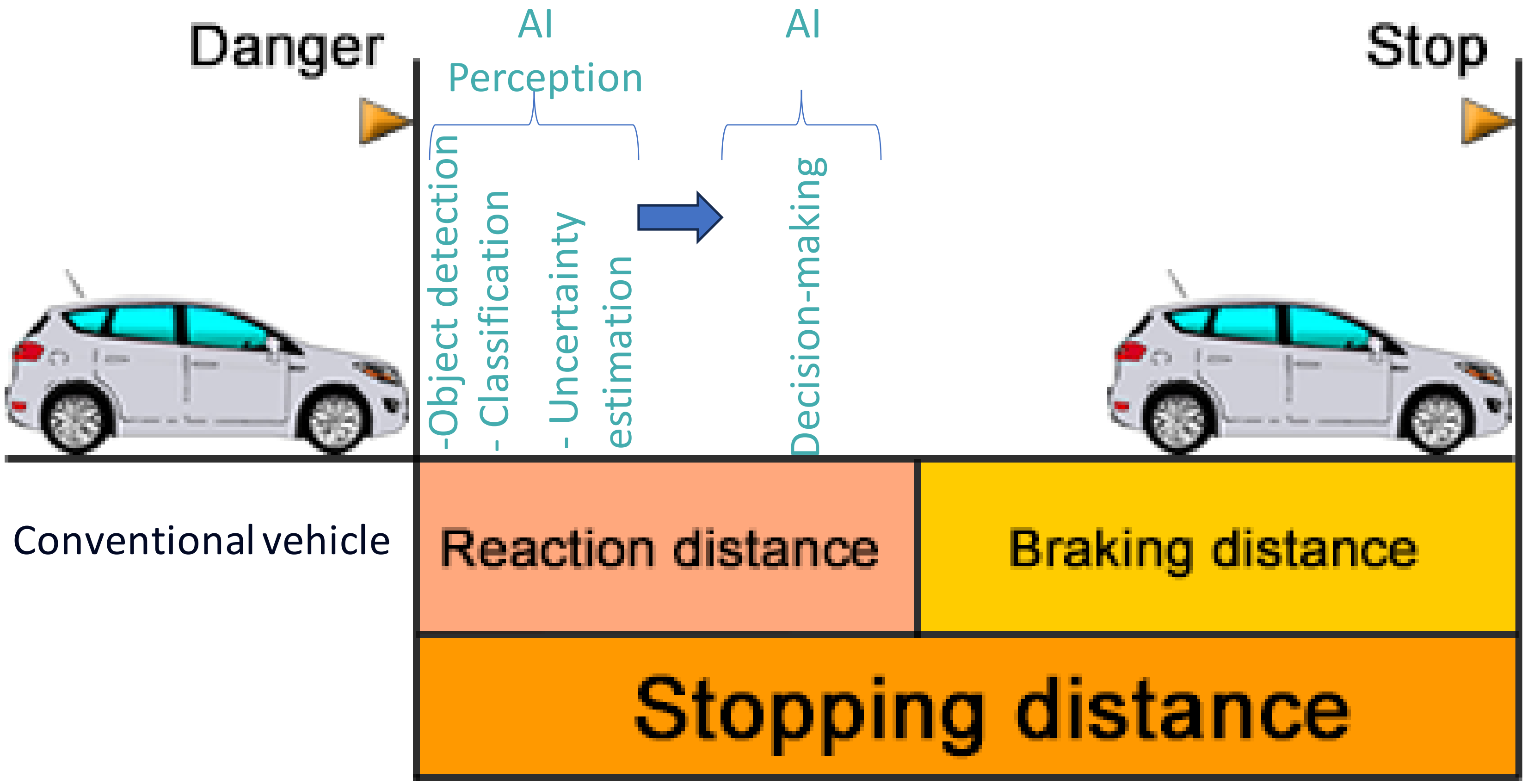


rain with motion blur

Challenges in uncertainty estimation for ADS

1. Compute resource for real time

the vehicle interacts with the environment in real time, small decision-making window



Challenges in uncertainty estimation for ADS

NIST

2. Data quality and quantity

- widely used methodology of data collection by vehicle-mounted cameras can lead to an excess of uneventful data
- edge cases (e.g. people emerging from manholes) are important but hard to find
- restricts generalization of the model during training

3. Theoretical limitations

- quantifying and associating an uncertainty with an outcome is a difficult task in ADS

theory
vs.
hypothesis

Who would care about this work?

Developing recommendations for robustness improvements and criteria for measurement and technical evaluation of AI perception performance in the form of measurement standards and supplemental code may benefit

- design engineers from the industry

- researchers in academia

- Federal agencies interested in AV

- U.S. and international standards bodies

Standards Development Organizations (SDOs)

- Process of developing a standard is typically facilitated by a Standards Development Organization (SDO)
- SDOs adhere to fair and equitable processes that ensure the highest quality outputs and reinforce the market relevance of standards.
- SDOs such as IEEE, International Electrotechnical Commission (IEC), International Organization for Standardization (ISO), and others offer time-tested platforms, rules, governance, methodologies, and services that objectively address the standards development lifecycle, and help facilitate the development, distribution and maintenance of standards.

IEEE

Work with us

We have a GPU cluster and have started w/ open-source models and public datasets

We are getting an automated test vehicle and will be working to validate the initial AI test methods

We are partnering with VTTI and their Smart Road infrastructure

We are looking for other partners from industry, government, and academia to share data and AI models for ADS and collaborate on these problems

STAY IN TOUCH

CONTACT US

ai-av@nist.gov

Program details

QUESTIONS ?