CDT-D^{2}AIR Course on

D^2 Robots and Autonomous Agents

Autonomous Vehicles – A Case Study

Subramanian Ramamoorthy
School of Informatics
University of Edinburgh

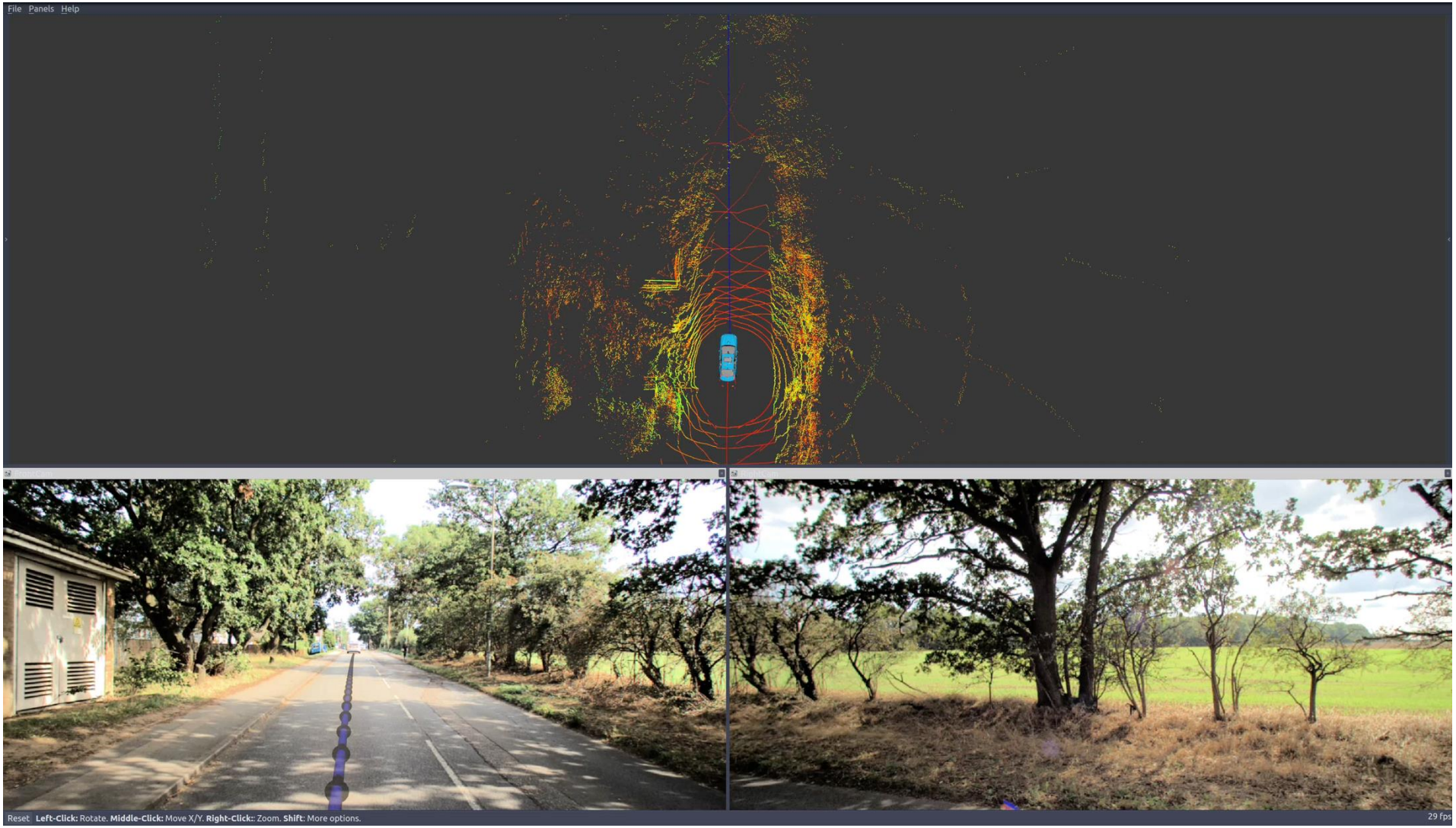
25 March 2026

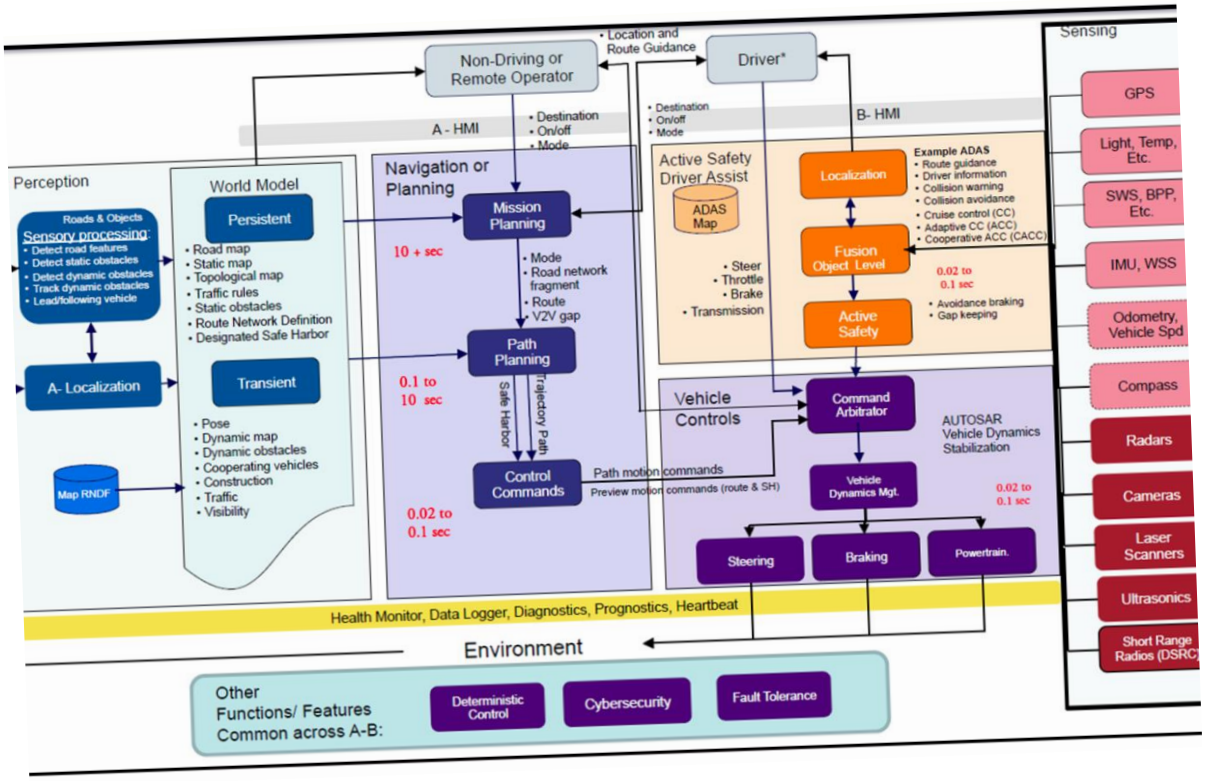
Course Logistics: Remit of Presentation and Report

Some clarifications following student questions:

- The specific focus of the presentation and report should be on *technical issues around evaluation and ensuring safe deployment (e.g. see ideas in today's lecture and discussion)*
- For completeness you need to describe your system architecture, but the new content in this assignment addresses tools and techniques for *analysing properties*
  - It would be fine to have the same architecture as in your masterclass, but the question of how you will evaluate/analyse is salient here
  - The report expected for this part is much shorter (3 pages) and the majority of it (e.g. 2 pages) should focus on above question – including how this question is handled in the literature

AVs: Mobile Robots in the Wild


File Panels Help
Reset Left-Click: Rotate. Middle-Click: Move X/Y. Right-Click:: Zoom. Shift: More options.
29 fps

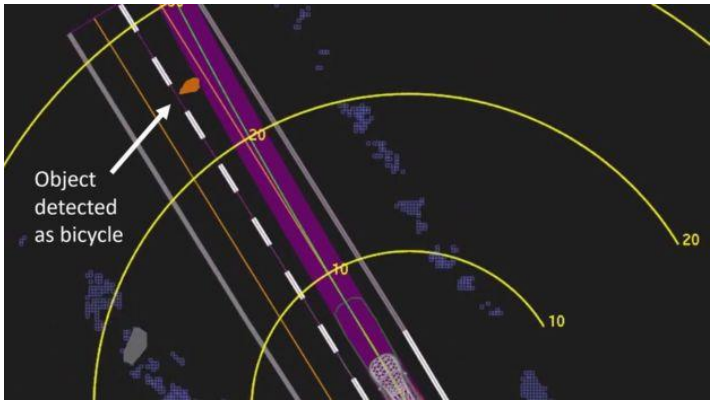


SAE International Autonomous Mode Functional Architecture Flow Diagram (Underwood, 2016)

Effect of Perception Errors on Action and Closed-Loop Behaviour

[Source: TfL CCTV]

[Source: bbc.com]



[Source: BBC/NTSB]

"Should I enter or not? When?"

… "how do others respond to me?"

Goal Recognition as Rational Inverse Planning

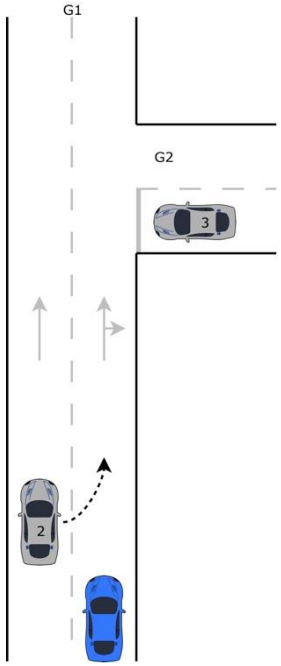
T_1: lane follow path

T_2: lane change path

T_3: slowdown path …

Best fit?

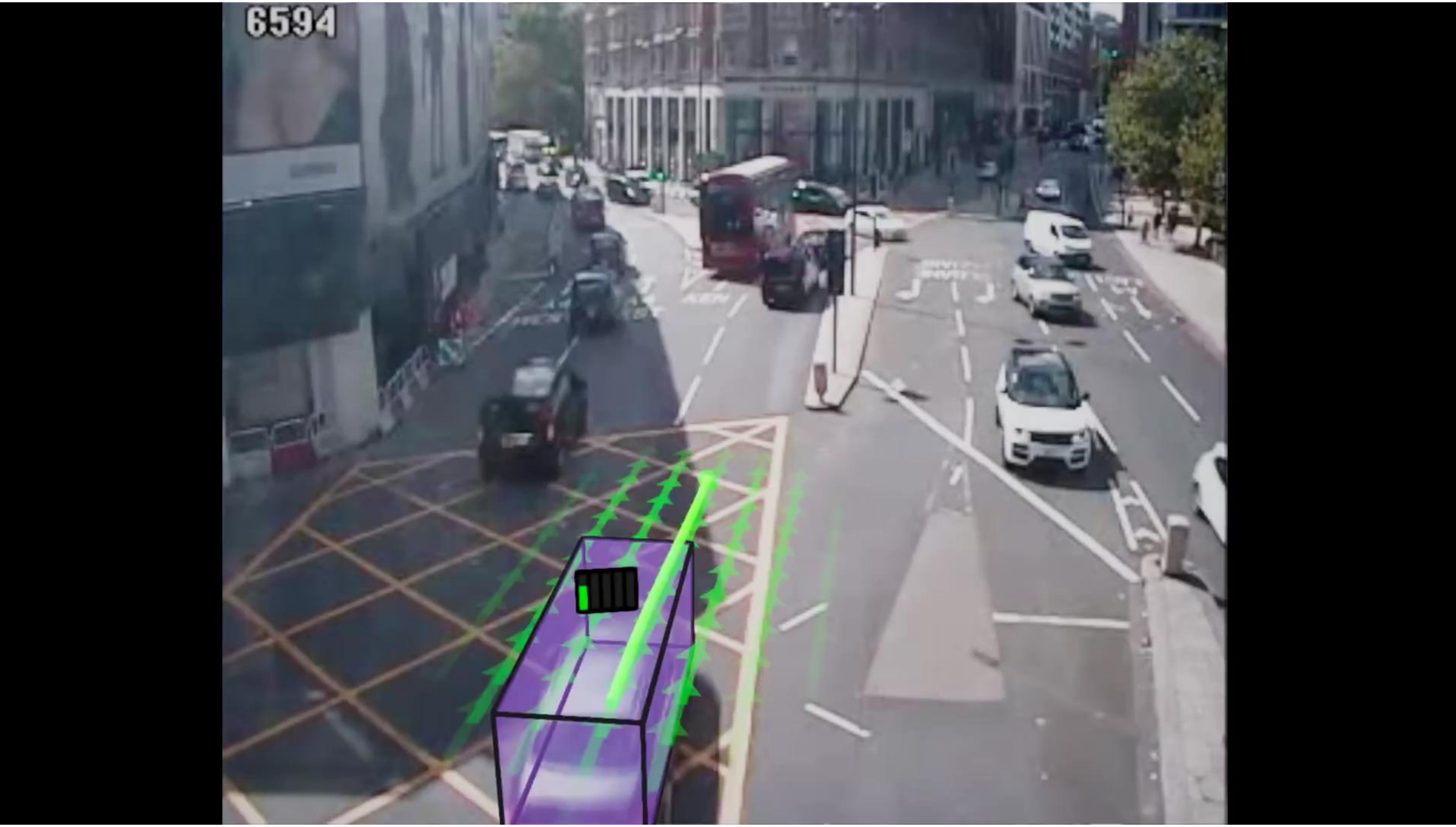
O: observed path



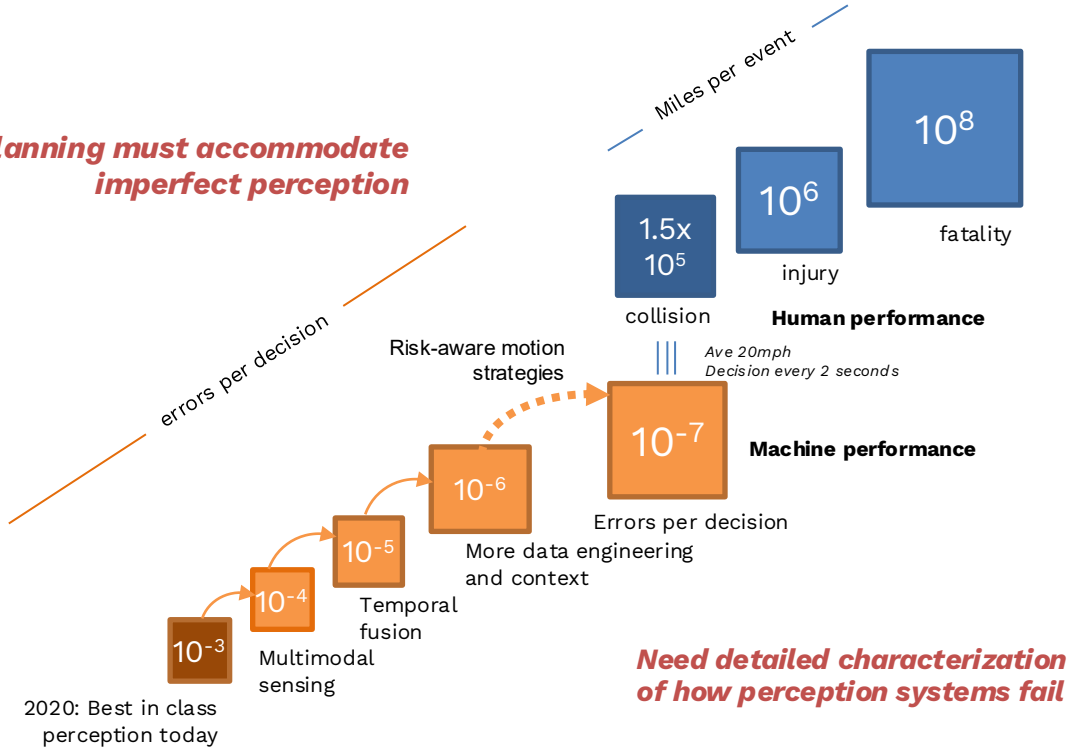
Where could car #2 be going?

- Landmarks can be extracted from maps
- Observation can be compared against, e.g., lane follow, lane change, cautious slowdown

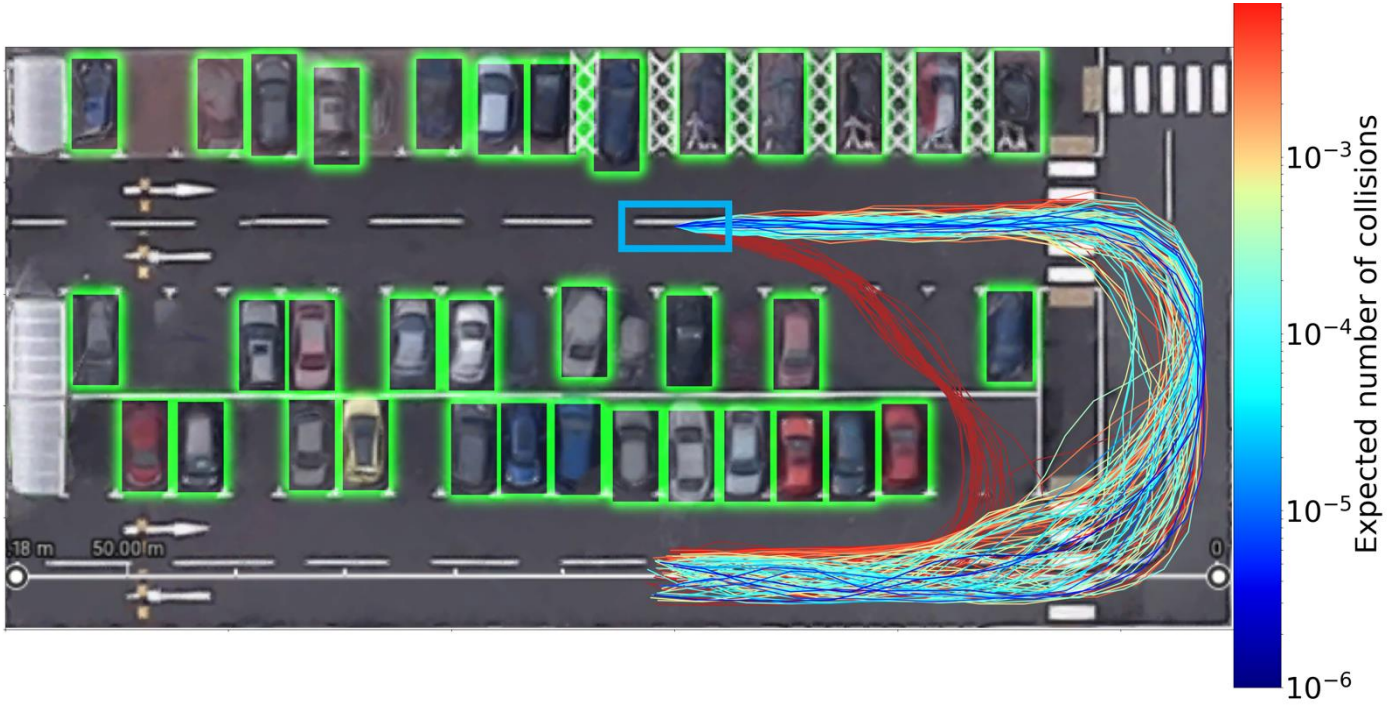
S.V. Albrecht, C. Brewitt, J. Wilhelm, F. Eiras, M. Dobre, S. Ramamoorthy, **Interpretable goal-based prediction and planning for autonomous driving**, In Proc. *IEEE International Conference on Robotics and Automation* (ICRA), 2021.



A Hierarchy of Errors



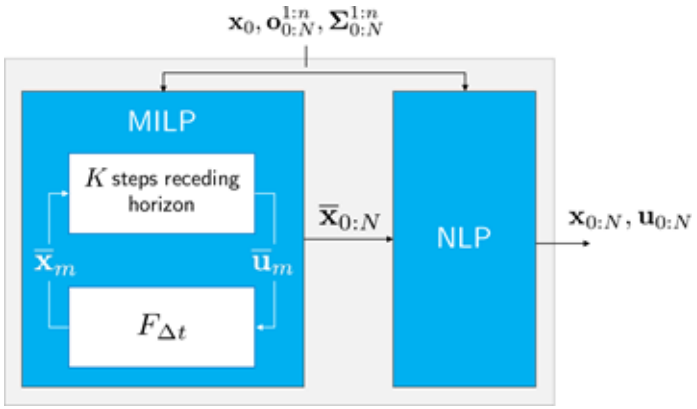
Planning with Imperfect Perception: Quantifying Uncertainty and Risk



A. Blake, A. Bordallo, K. Brestnichki, M. Hawasly, S.V. Penkov, S. Ramamoorthy, A. Silva, **FPR-Fast Path Risk algorithm to evaluate collision probability**, *IEEE Robotics & Automation Lett. 5(1): 1-7*, 2020.

An Approach to Safe Planning

1. Compute an approximate solution to the problem using a **Mixed Integer Linear Programming** receding-horizon formulation (encodes variety of rules/constraints)
2. Use it as an **initialization** to the non-linear trajectory optimization problem



F. Eiras, M. Hawasly, S.V. Albrecht, S. Ramamoorthy, **A two-stage optimization-based motion planner for safe urban driving**, *IEEE Transactions on Robotics* (T-RO), 2021.

MILP formulation

$$\overline{\mathbf{u}}^*_{m:m+K-1} = \underset{\overline{\mathbf{u}}_{m:m+K-1}}{\operatorname{argmin}} \sum_{k=m}^{m+K} \sum_{\Theta_\iota, \Omega_\iota \in \mathcal{C}} \Omega_\iota \Theta_\iota(\overline{\mathbf{x}}_k, \overline{\mathbf{u}}_k, \mathbf{o}_k^{1:n}, \mathbf{\Sigma}_k^{1:n})$$

Soft constraints: linearized progress, comfort, and risk reduction

$$s.t. \quad \forall k \in \{m, ..., m+K\}:$$

$$\overline{\mathbf{x}}_{k+1} = F_{\Delta t}(\overline{\mathbf{x}}_k, \overline{\mathbf{u}}_k)$$
$$a^x_{\min} \leq a^x_k \leq a^x_{\max}$$
$$a^y_{\min} \leq a^y_k \leq a^y_{\max}$$
$$|a^x_{k+1} - a^x_k| < \Delta a^x_{\max} \Delta t$$
$$|a^y_{k+1} - a^y_k| < \Delta a^y_{\max} \Delta t$$
$$v^x_{\min} \leq v^x_k \leq v^x_{\max}$$
$$v^y_{\min} \leq v^y_k \leq v^y_{\max}$$
$$v^x \leq \rho |v^y|$$

Hard constraints (could come from compositional logical specs):

Linear Vehicle dynamics constraints

$$d + b^{\mathcal{M}}_l(x_k) \leq y_k$$
$$y_k \leq b^{\mathcal{M}}_r(x_k) - d$$

Driveable surface constraints

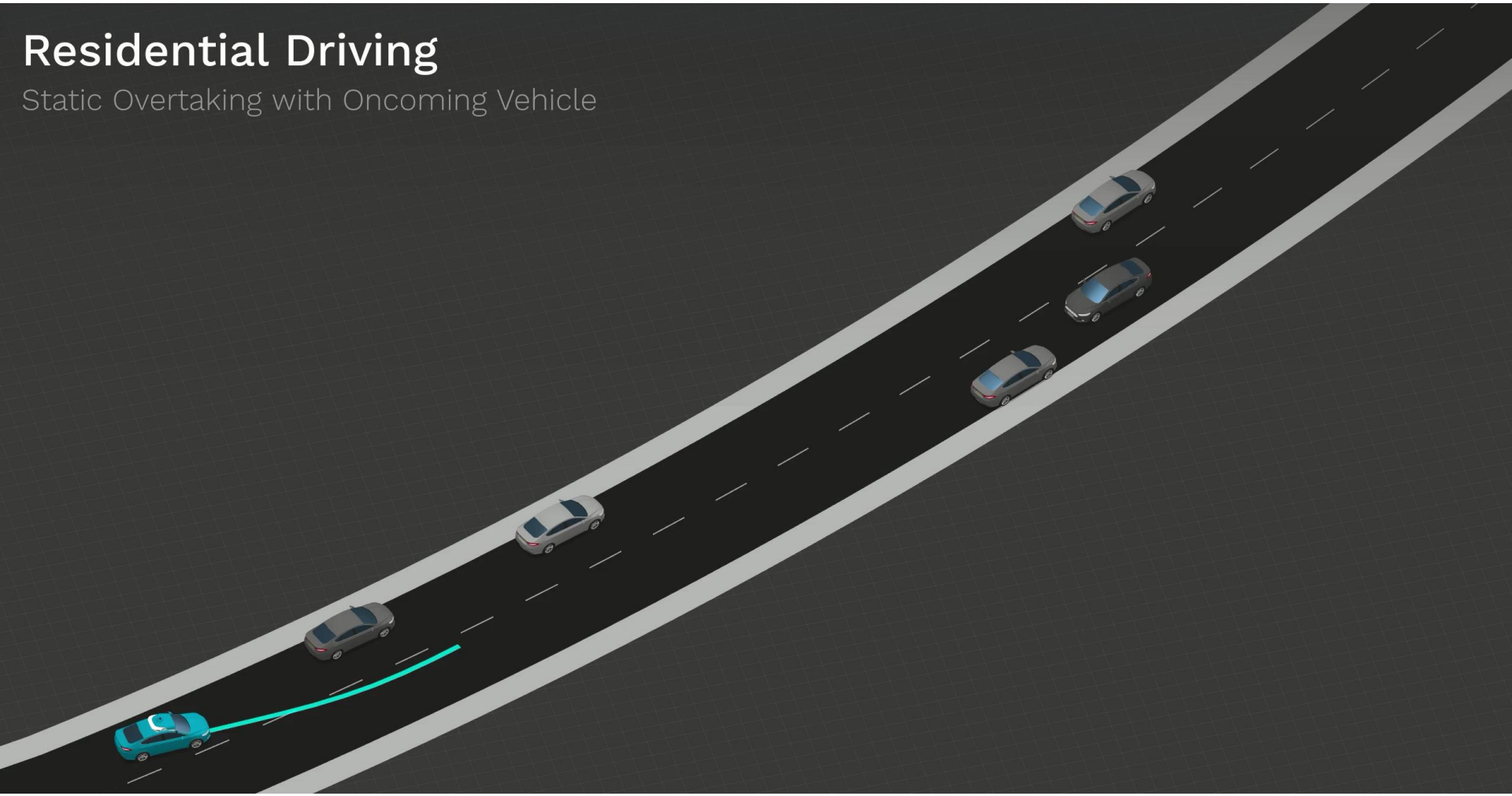
$$y^i_{k,\max} - M\mu^i_k \leq y_k, i \in \{1, ..., n\}$$

Collision avoidance up to certain level of uncertainty

Solve in a receding-horizon of length K

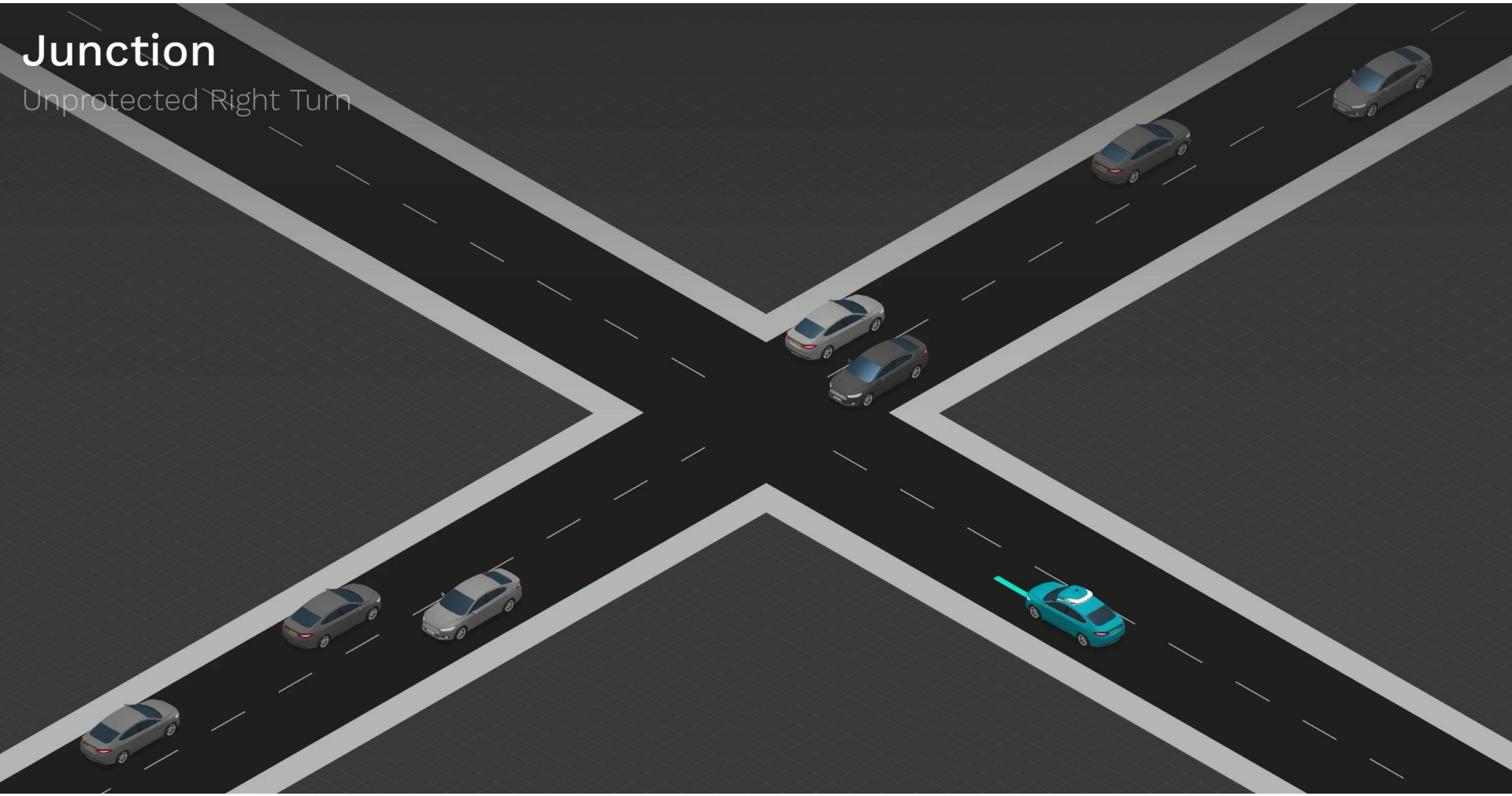
Residential Driving

Static Overtaking with Oncoming Vehicle

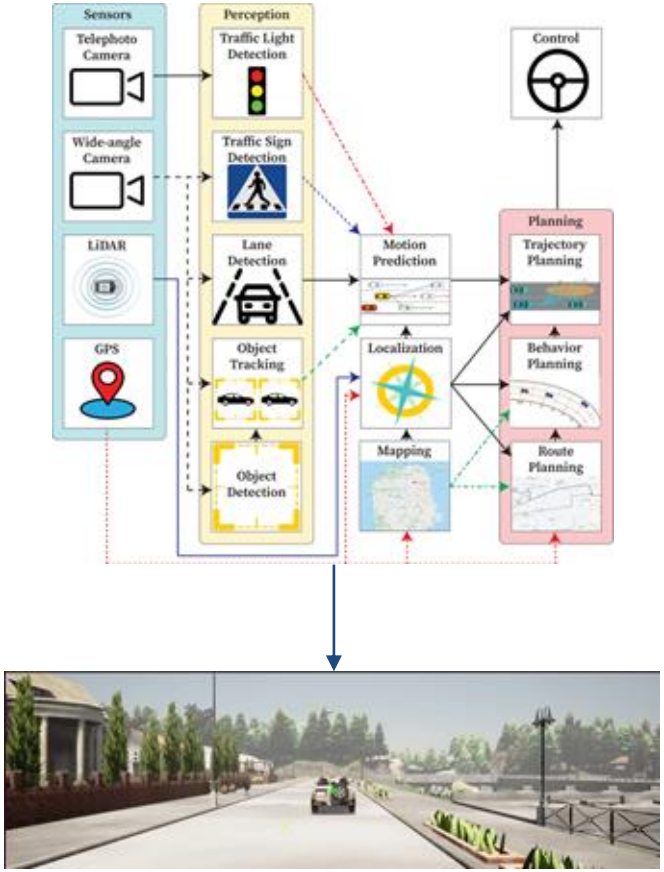


Junction

Unprotected Right Turn



What is the Probability the *System Fails*?



C. Innes, S. Ramamoorthy, **Testing rare downstream safety violations via upstream adaptive sampling of perception error models**. In Proc. *IEEE International Conference on Robotics and Automation* (ICRA), 2023.

Many Sources of Difficulty

1. Specifying Safety

"The ego vehicle is not allowed to cross a red traffic light. If the traffic light is yellow and the ego vehicle can come to a standstill in front of the intersection without falling below an acceleration threshold a_{pos}, the ego vehicle is not allowed to cross a yellow traffic light"

2. The Fidelity Gap

3. Rare Events

1e-22

Specifying Safety: Signal Temporal Logic

The first thing must remain true **until** the second thing becomes true

$$\varphi := p \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \mathcal{N}\varphi \mid \Box\varphi \mid \Diamond\varphi \mid \varphi_1 \mathcal{U} \varphi_2$$

This will **always** be true

Eventually, this will be true

"For the first T=100 seconds, the ego vehicle should always stay at least 2 metres from the other vehicle"

$$\Box_{[0,T]} \left(\| C_{ego}.pos - C_{other}.pos \| \geq 2.0 \right)$$

Signal Temporal Logic: Robustness Metrics

Definition 5 (Cumulative Robustness):

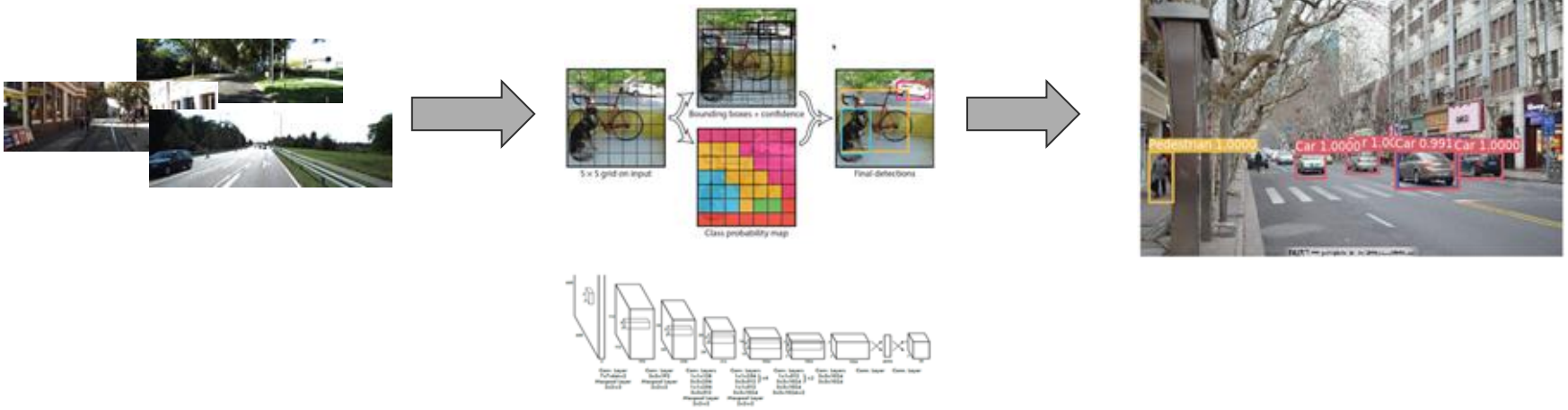
$$
\begin{aligned}
\rho^{+}(l(\sigma) \geq 0, \sigma, t_k) &:= \mathfrak{R}^{+}(l(\sigma[k])), \\
\rho^{-}(l(\sigma) \geq 0, \sigma, t_k) &:= \mathfrak{R}^{-}(l(\sigma[k])), \\
\rho^{+}(\neg\varphi, \sigma, t_k) &:= -\rho^{-}(\varphi, \sigma, t_k), \\
\rho^{-}(\neg\varphi, \sigma, t_k) &:= -\rho^{+}(\varphi, \sigma, t_k), \\
\rho^{+}(\psi \vee \varphi, \sigma, t_k) &:= \max\{\rho^{+}(\psi, \sigma, t_k), \rho^{+}(\varphi, \sigma, t_k)\}, \\
\rho^{-}(\psi \vee \varphi, \sigma, t_k) &:= \max\{\rho^{-}(\psi, \sigma, t_k), \rho^{-}(\varphi, \sigma, t_k)\}, \\
\rho^{+}(\psi \wedge \varphi, \sigma, t_k) &:= \min\{\rho^{+}(\psi, \sigma, t_k), \rho^{+}(\varphi, \sigma, t_k)\}, \\
\rho^{-}(\psi \wedge \varphi, \sigma, t_k) &:= \min\{\rho^{-}(\psi, \sigma, t_k), \rho^{-}(\varphi, \sigma, t_k)\}, \\
\rho^{+}(\mathbf{F}_I\varphi, \sigma, t_k) &:= \sum_{k' \in I} \rho^{+}(\varphi, \sigma, t_{k+k'}), \\
\rho^{-}(\mathbf{F}_I\varphi, \sigma, t_k) &:= \sum_{k' \in I} \rho^{-}(\varphi, \sigma, t_{k+k'}), \\
\rho^{+}(\mathbf{G}_I\varphi, \sigma, t_k) &:= \min_{k' \in I} \rho^{+}(\varphi, \sigma, t_{k+k'}), \\
\rho^{-}(\mathbf{G}_I\varphi, \sigma, t_k) &:= \min_{k' \in I} \rho^{-}(\varphi, \sigma, t_{k+k'}), \\
\rho^{+}(\psi \mathbf{U}_I\varphi, \sigma, t_k) &:= \sum_{k' \in I} (\min\{\rho^{+}(\varphi, \sigma, t_{k+k'}), \\
&\qquad \min_{k'' \in [k, k+k']} \rho^{+}(\psi, \sigma, t_{k''})\}), \\
\rho^{-}(\psi \mathbf{U}_I\varphi, \sigma, t_k) &:= \sum_{k' \in I} (\min\{\rho^{-}(\varphi, \sigma, t_{k+k'}), \\
&\qquad \min_{k'' \in [k, k+k']} \rho^{-}(\psi, \sigma, t_{k''})\}).
\end{aligned}
\tag{15}
$$

$$\rho(\varphi, \tau_{fail}) = -0.7$$ ❌

$$\rho(\varphi, \tau_{pass}) = 0.9$$ ✅

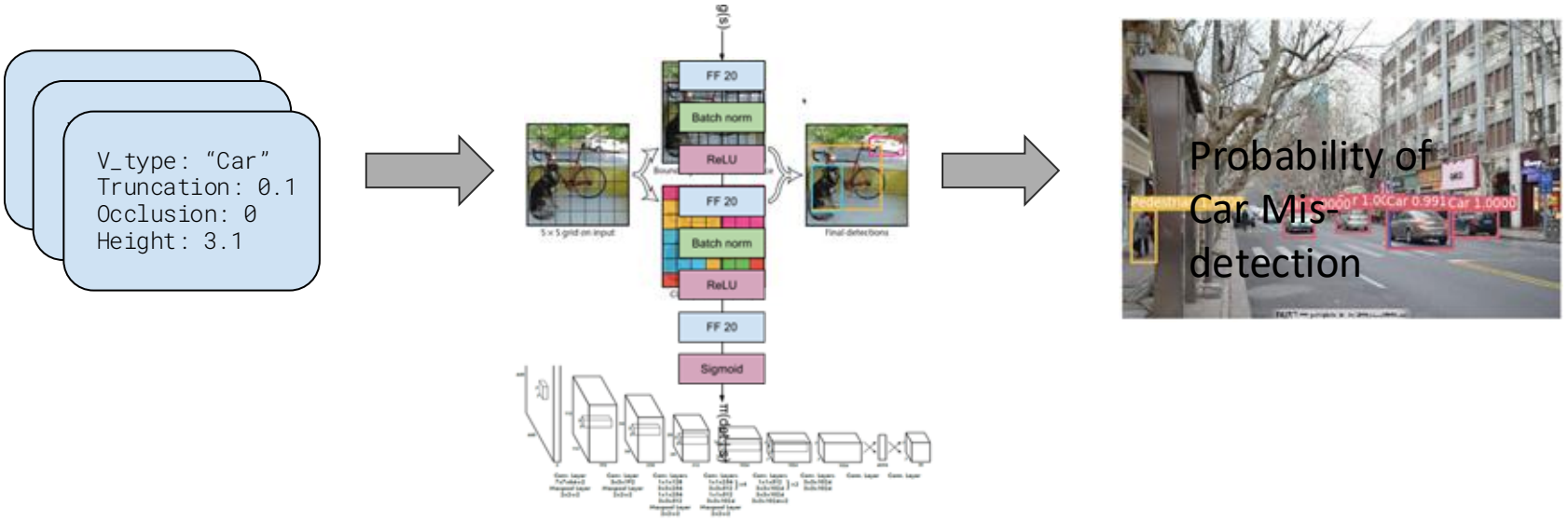
$$\rho(\varphi, \tau_{close}) = 0.0001$$ 🤔

Fidelity Gap: Obstacle Detection Example



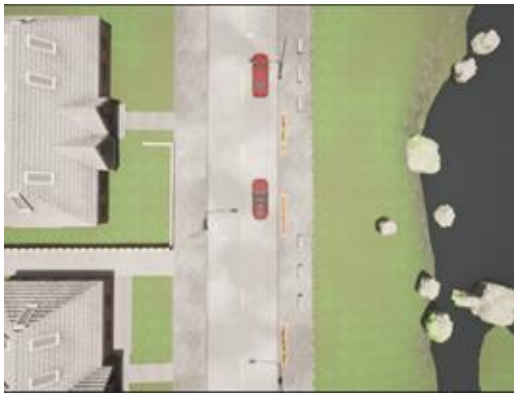
Perception Error Model Insight

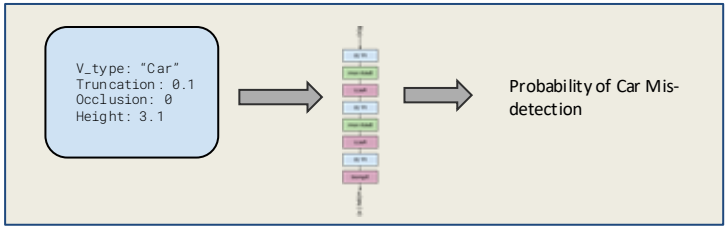
$$\forall s \in \mathcal{S}, f_\theta(s) = \hat{f}_{\hat{\theta}}(g(s))$$



e.g., "A Step Towards Efficient Evaluation of Complex Perception Tasks in Simulation." Sadeghi et.al., (2021).

Rarity: Sampling Detection Errors





$$\hat{\mu}_{\mathrm{MC}} = \frac{1}{N}\sum_{i=0}^{N} \mathbb{1}_{\{r(\tau_i,\varphi)\leq\gamma\}}$$

State-dependent Expectation:

$$\mathbb{E}_{\pi}\left[\mathbb{1}\left\{r(\tau,\varphi)\leq\gamma\right\}\right] = \int_{\tau} \mathbb{1}\left\{r(\tau,\varphi)\leq\gamma\right\} p(\tau)$$

Simulation Rollout

Specification

Safety Threshold

True Crash Probability: 1 in 10,000

Expected Simulations to get within 1% of true probability: 1,000,000,000,000

$$\mathrm{RE}_{\mathrm{MC}}(\mu) \approx \sqrt{\frac{1}{N\mu}}$$

Importance Sampling

$$\hat{\mu}_{\mathrm{IS}} = \frac{1}{N}\sum_{i=0}^{N} \mathbb{1}_{\{r(\tau_i,\varphi)\leq\gamma\}} \frac{p(\tau_i)}{q(\tau_i)}$$

Original "target" distribution

New "proposal" distribution

Ideal Proposal generates lots of failures:

$$q^*(\tau) = \left(\frac{p(\tau)}{\mu}\right) \mathbb{1}_{r(\tau,\varphi)<\gamma}$$

Learning a Proposal Distribution: e.g. Cross Entropy Method (CEM)

$$-\sum_{i=0}^{N} w_i \left(\sum_{t=0}^{T} \log q_\phi(h(s_{i,t})) \right) \mathbb{1}_{r(\tau_i,\varphi)<\gamma}$$

Projection Function

$$w_i = \prod_{t=0}^{T} \frac{\hat{f}_{\hat{\theta}}(g(s_{i,t}))}{q_\phi(h(s_{i,t}))}$$

State-dependent IS (weighting)

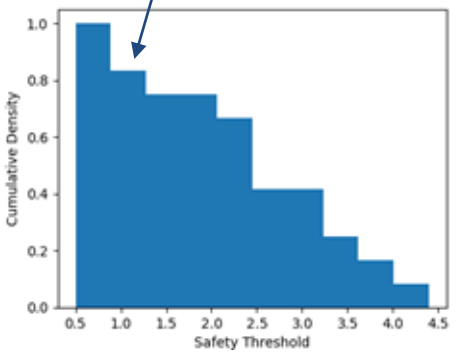
Other Ideas: Adaptive Thresholding (AT)

What if Failures are still Rare?

for $\kappa = 1$ to K **do**
  $\{\tau_0 \dots \tau_{N_\kappa}\} \leftarrow$ Sample N_κ rollouts with q_ϕ
  Sort $\{\tau_0 \dots \tau_{N_\kappa}\}$ by $r(\tau, \varphi)$
  $\gamma_\kappa \leftarrow max\left(\gamma, r\left(\varphi, \tau_{\lfloor \sigma N_\kappa \rfloor}\right)\right)$
  $\phi \leftarrow$ Min (18) with $\gamma_\kappa, q_\phi, \hat{f}_{\hat{\theta}}, \{\tau_0 \dots \tau_{N_\kappa}\}$

$$-\sum_{i=0}^{N} w_i \left(\sum_{t=0}^{T} \log q_\phi(h(s_{i,t}))\right) \mathbb{1}_{r(\tau_i,\varphi)<\gamma}$$

0.82th Quantile



Automated Braking: Experiment Results

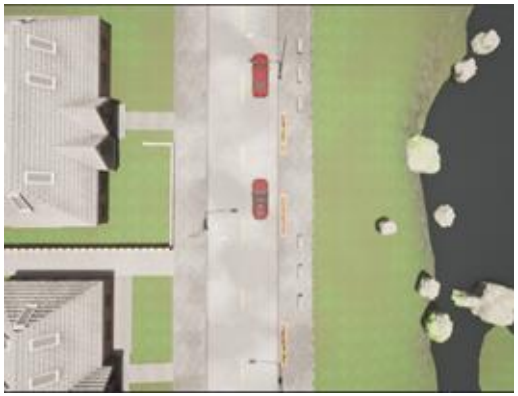
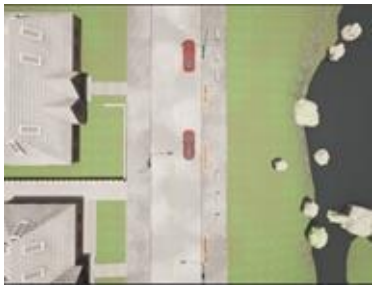


TABLE II: Sampling Strategies

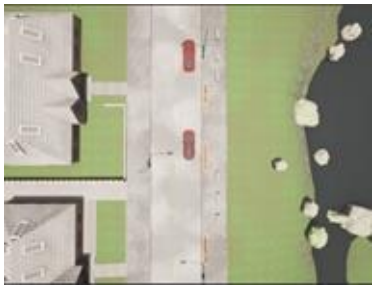
Method	$\hat{\mu}$	Failures	NLL
MC_{100}	0.0	0 / 100	-
MC_{10000}	0.0	0 / 10000	-
NAIVE-50_{100}	3.97×10^{-22}	77 / 100	162.54
NAIVE-50_{10000}	4.46×10^{-20}	7478 / 10000	162.53
$\text{ADAPTIVE-PEM}_{100}$	3.36×10^{-15}	52 / 100	**48.03**

Limitations: Is the Scenario Big Enough?



"The distance between the ego car and other car must never drop below 2 metres"

Limitations: Is the Scenario Big Enough?

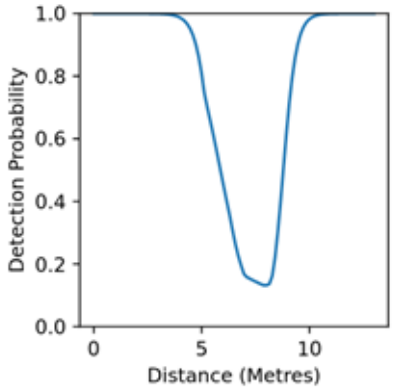


"The distance between the ego car and other car must never drop below 2 metres"

"The ego vehicle is not allowed to enter an intersection if there is another vehicle with the right of way that will be endangered by the ego vehicle.

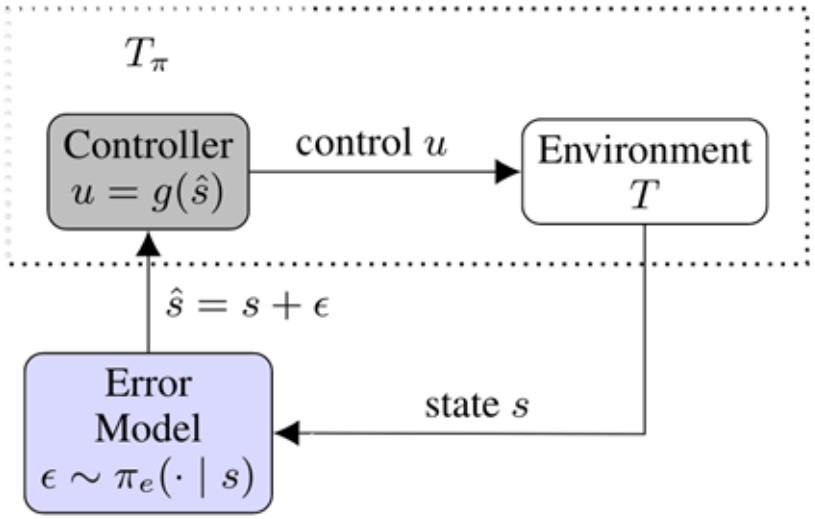
The left turning ego vehicle that has no priority (given by traffic signs) over an oncoming vehicle may only drive onto the oncoming lane if the ego vehicle does not endanger the other vehicle. The same applies if another vehicle turns right into the same road as the ego vehicle"

Risk-driven Design: Where are Perception Errors Most Risky?



A.L. Corso, S.M. Katz, C. Innes, X. Du, S. Ramamoorthy, M.J. Kochenderfer, **Risk-driven design of perception systems**. In Proc. *Conference on Neural Information Processing Systems* (NeurIPS), 2022.

Perceptual Error Risk as Policy Evaluation



Transition Function

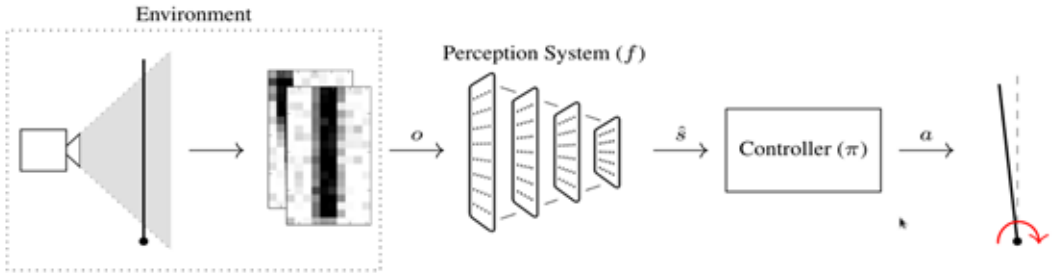
$$T_{\pi_e}(s' \mid s, \epsilon) = T(s' \mid s, g(s+\epsilon))$$

State-Action Value

$$Q^{\pi_e}(s, \epsilon) = R(s, g(s+\epsilon)) + \sum_{s'} T_\pi(s' \mid s, \epsilon) V^{\pi_e}(s')$$

- Evaluate risk of making perception in a particular state
- Evaluate long term consequence according to CVaR value function (evaluating on upper quantile of worst case outcomes)

Risk-aware Design



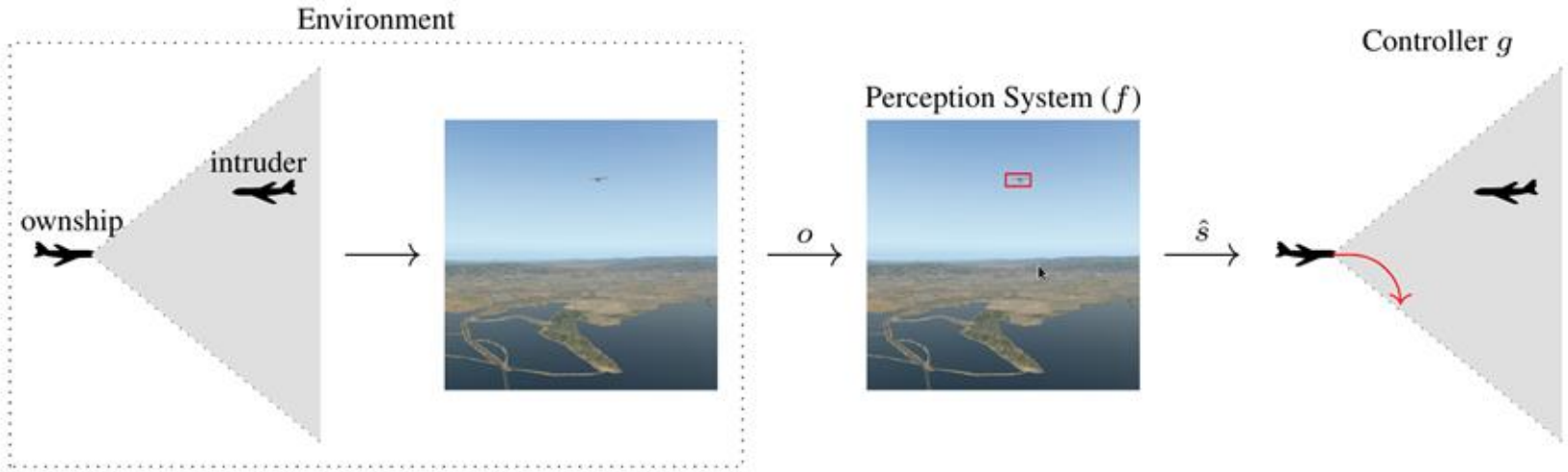
Risk-Aware Perceptual Loss

$$\mathcal{L}_R(s, \hat{s}) = \mathcal{L}(s, \hat{s}) + \lambda \rho_\alpha^{\pi_e}(s, \hat{s} - s)$$

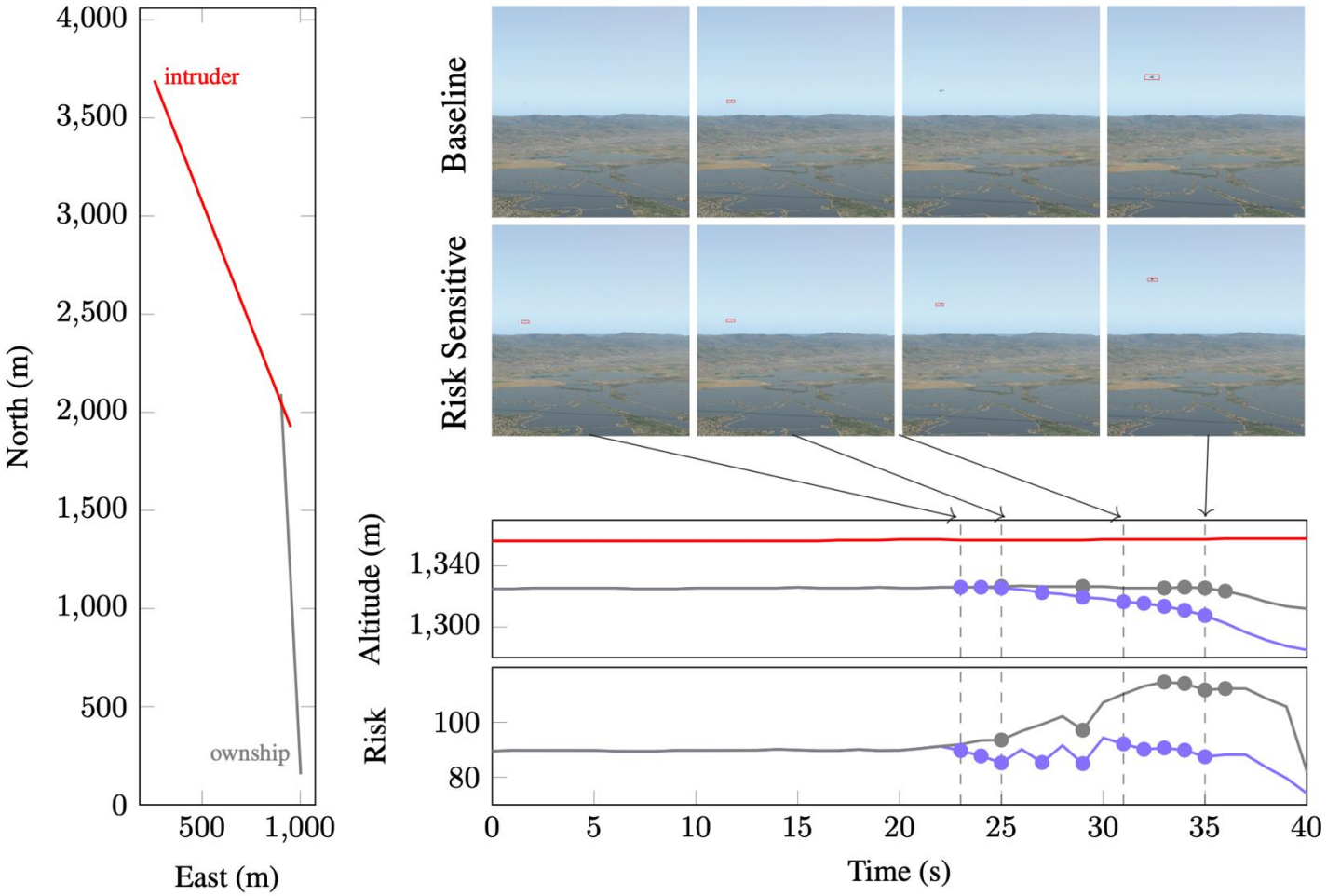
Risk-Aware Data Generation

$$w_\alpha(s) = \max_{\epsilon \in \mathcal{E}} \rho_\alpha^{\pi_e}(s, \epsilon) - \rho_\alpha^{\pi_e}(s, 0)$$

Case Study:
Aircraft Collision Avoidance



Case Study:
Aircraft Collision Avoidance



Use of Simulation in Development

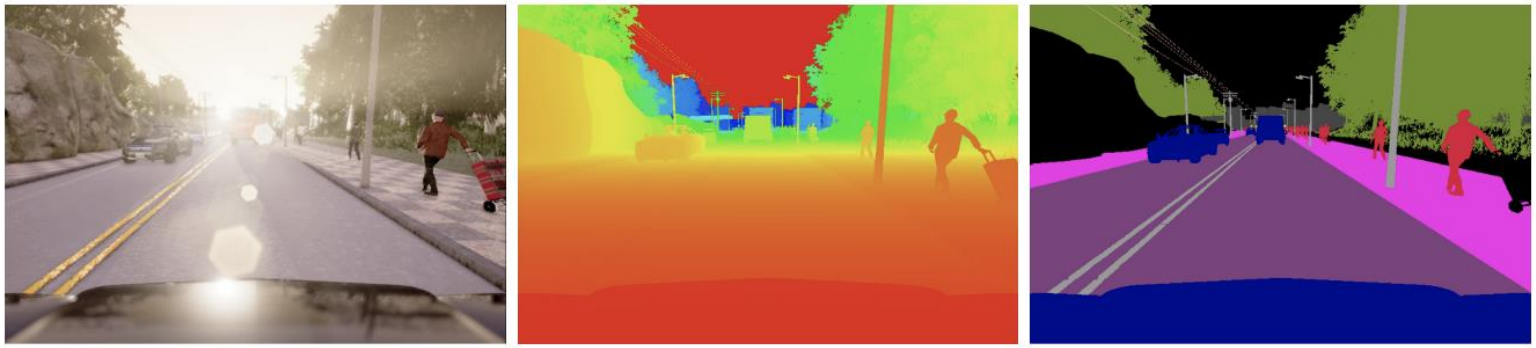
CARLA: An Open Urban Driving Simulator

Alexey Dosovitskiy[1], **German Ros**[2,3], **Felipe Codevilla**[1,3], **Antonio López**[3], **and Vladlen Koltun**[1]

[1]Intel Labs [2]Toyota Research Institute [3]Computer Vision Center, Barcelona

Abstract: We introduce CARLA, an open-source simulator for autonomous driving research. CARLA has been developed from the ground up to support development, training, and validation of autonomous urban driving systems. In addition to open-source code and protocols, CARLA provides open digital assets (urban layouts, buildings, vehicles) that were created for this purpose and can be used freely. The simulation platform supports flexible specification of sensor suites and environmental conditions. We use CARLA to study the performance of three approaches to autonomous driving: a classic modular pipeline, an end-to-end model trained via imitation learning, and an end-to-end model trained via reinforcement learning. The approaches are evaluated in controlled scenarios of increasing difficulty, and their performance is examined via metrics provided by CARLA, illustrating the platform's utility for autonomous driving research.

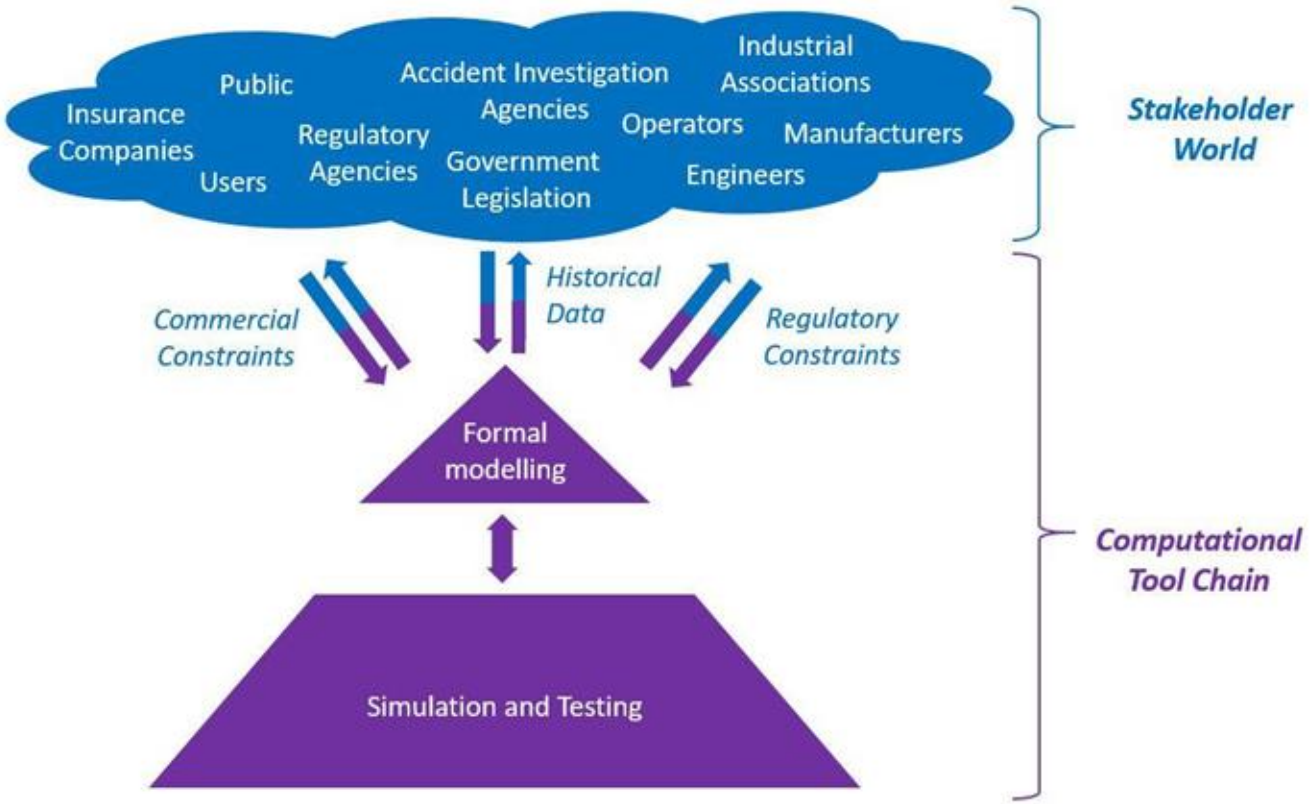
Keywords: Autonomous driving, sensorimotor control, simulation



Use of Simulation in Development

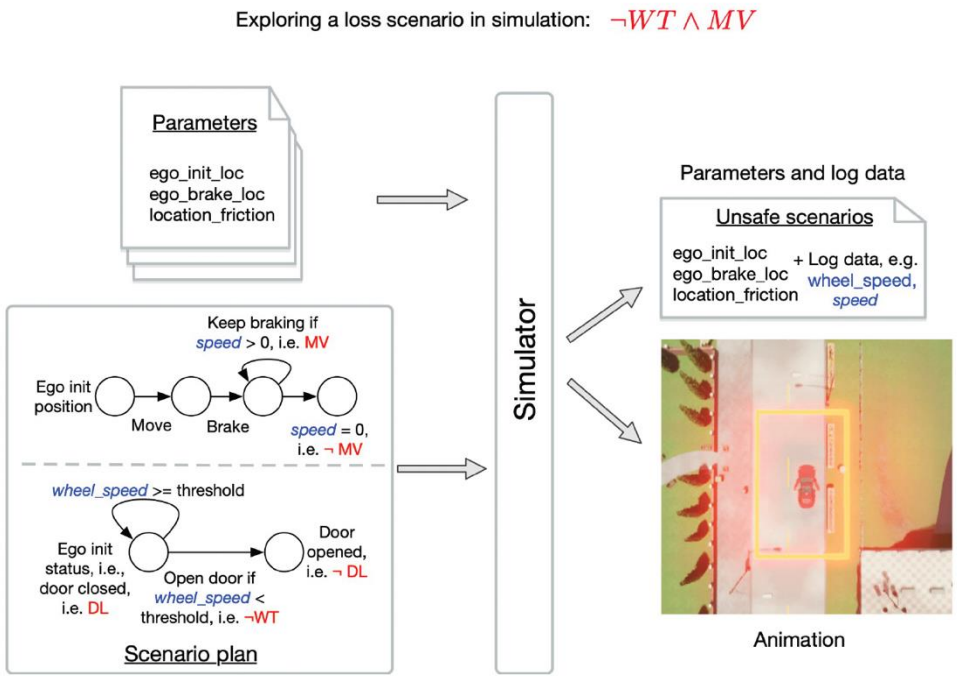
Browse examples of Applied Intuition toolchain:

https://www.appliedintuition.com/



C. Innes, A. Ireland, Y. Lin, S. Ramamoorthy, **Anticipating accidents through reasoned simulation**, Safety Critical Systems Symposium (SSS 2023).

Combining *Reasoning* and Simulation: e.g. exploring Loss Scenarios



The inputs above on the left include, i) setup parameters for the simulation, and ii) template scenarios that encode the abstract loss scenario identified by the formal modelling. Note that while a graphical notation is shown, the developer uses a Python API to specify the inputs. The output from the simulator takes the form of i) an animation of the loss scenario, and ii) a log of all the relevant parameters. Note that a demo of the case study simulation is available via [17], which includes the animations generated by CARLA.

C. Innes, A. Ireland, Y. Lin, S. Ramamoorthy, **Anticipating accidents through reasoned simulation**, Safety Critical Systems Symposium (SSS 2023).

Summary

- Establishing safety of learning enabled autonomous systems is a key requirement for broader adoption

- Layered optimization/decision making architectures represent one promising path, appropriately interleaving a hierarchy of concerns
- For this to work well, we need careful treatment of uncertainty flows and how component errors lead to system errors

- Towards this end, we discussed several technical methods including:
  - Adaptive importance sampling with logical specifications
  - Risk-driven design of perception systems
- Current and future work along all of these directions aims at expanding the scope and scale of such a methodology

Discussion Points: How Safe is Safe Enough?

RAND CORPORATION

Driving to Safety

How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?

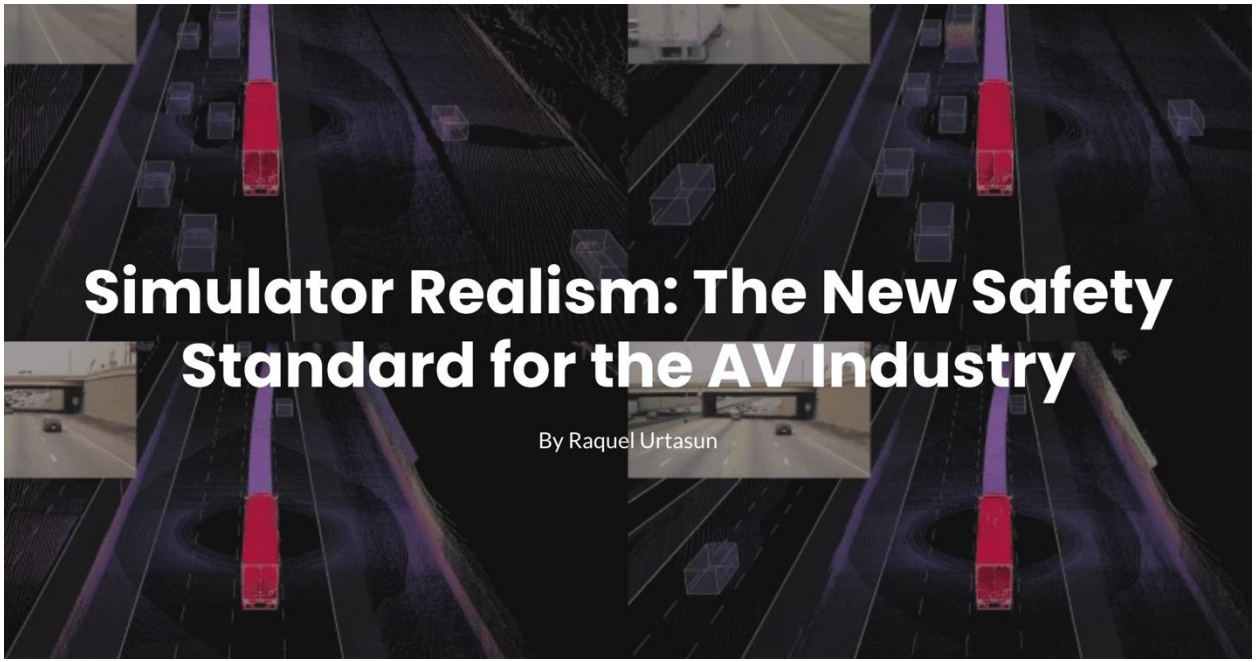
Nidhi Kalra, Susan M. Paddock

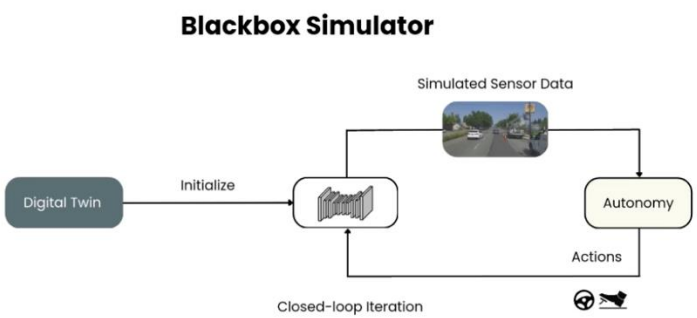
Key findings

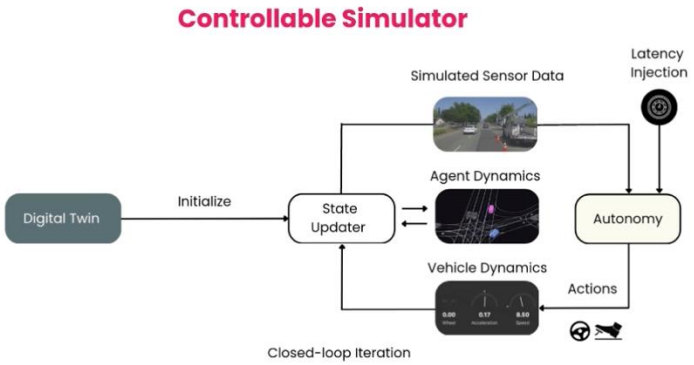
- Autonomous vehicles would have to be driven hundreds of millions of miles and sometimes hundreds of billions of miles to demonstrate their reliability in terms of fatalities and injuries.
- Under even aggressive testing assumptions, existing fleets would take tens and sometimes hundreds of years to drive these miles—an impossible proposition if the aim is to demonstrate their performance prior to releasing them on the roads for consumer use.
- Therefore, at least for fatalities and injuries, test-driving alone cannot provide sufficient evidence for demonstrating autonomous vehicle safety.
- Developers of this technology and third-party testers will need to develop innovative methods of demonstrating safety and reliability.
- Even with these methods, it may not be possible to establish with certainty the safety of autonomous vehicles. Uncertainty will persist.
- In parallel to creating new testing methods, it is imperative to develop adaptive regulations that are designed from the outset to evolve with the technology so that society can better harness the benefits and manage the risks of these rapidly evolving and potentially transformative technologies.

https://www.rand.org/content/dam/rand/pubs/research_reports/RR1400/RR1478/RAND_RR1478.pdf

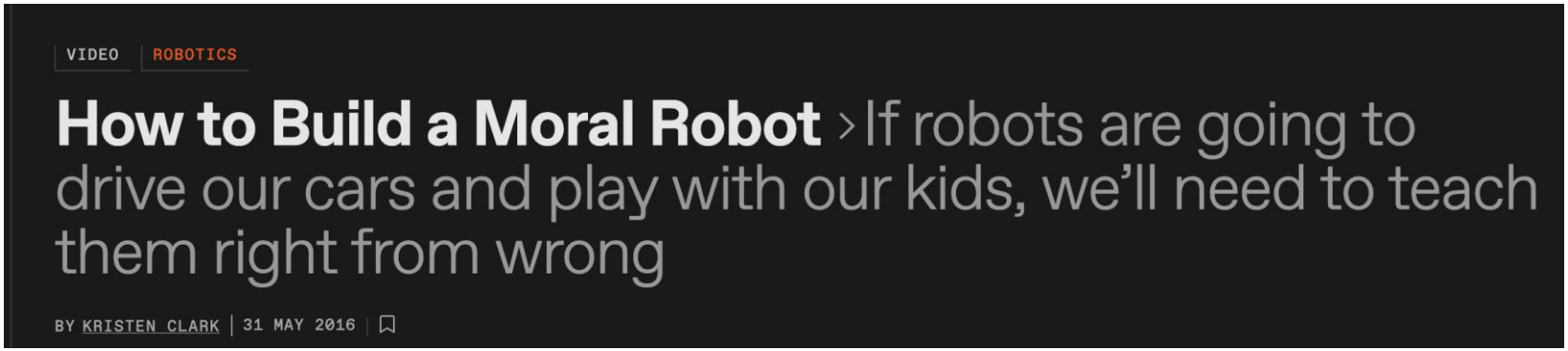
Discussion Points: Realism in Simulations







Discussion Points: Ethics and Morality



https://spectrum.ieee.org/how-to-build-a-moral-robot

https://spectrum.ieee.org/av-trolley-problem