WAYVE

The Next Frontier in Embodied AI: Autonomous Vehicles

Engineering IB Paper 8 - Autonomous Driving | April 2022

Dr Alex Kendall, Co-Founder & CEO

1. Driving Intelligence
2. Sensors
3. Offboard software
4. Safety

Part 1:
Driving Intelligence

2007 DARPA Urban Challenge

Commercial self driving car efforts

HD Maps

(brittle / slow to build / expensive to maintain)

LiDAR Sensors

(expensive / short lifespan)

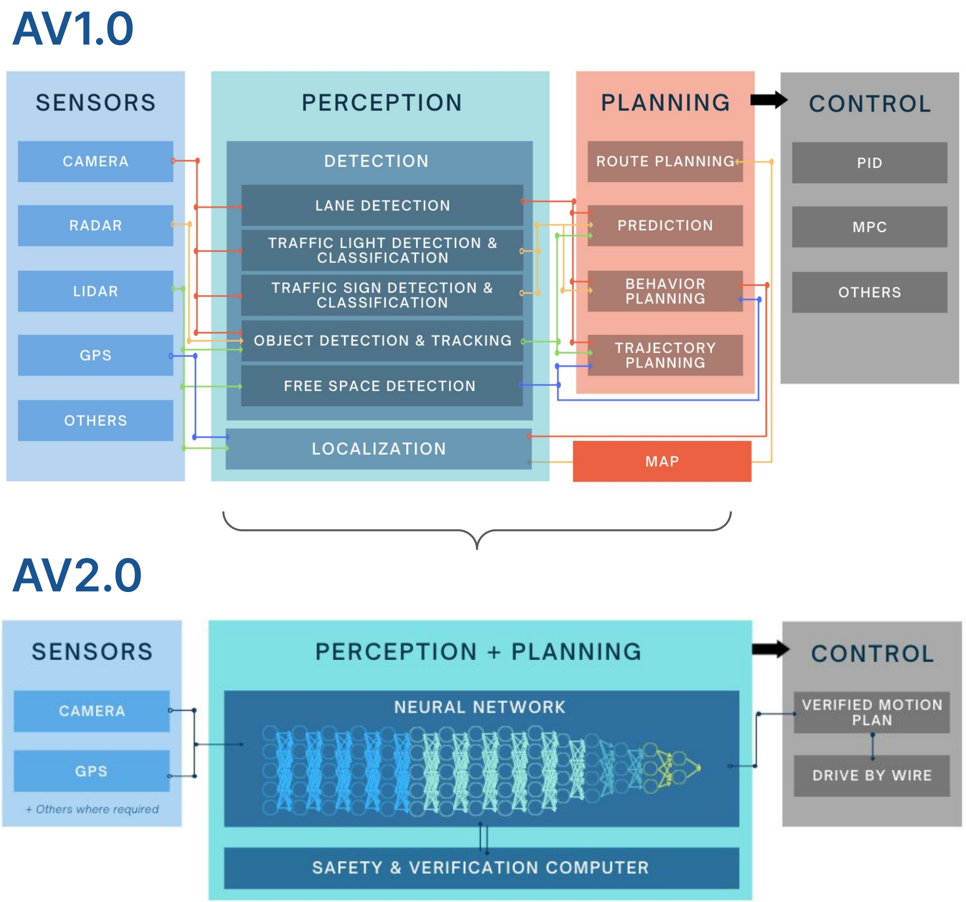
Hand-Designed Rules

(rigid / clunky)

A PARADIGM SHIFT

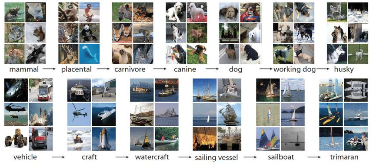
Pioneering the next-generation AV architecture

- Solving self-driving with data
- End-to-end deep learning
- Lean sensors & compute



Deep learning has achieved superhuman performance in comparably complex settings to autonomous driving which are more accessible

IMAGE RECOGNITION



ImageNet considered a solved problem in 2017

NATURAL LANGUAGE & VISION

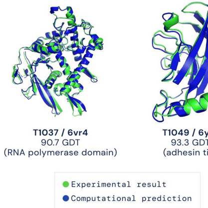
DALL-E creating images from text (OpenAI)

GAMES

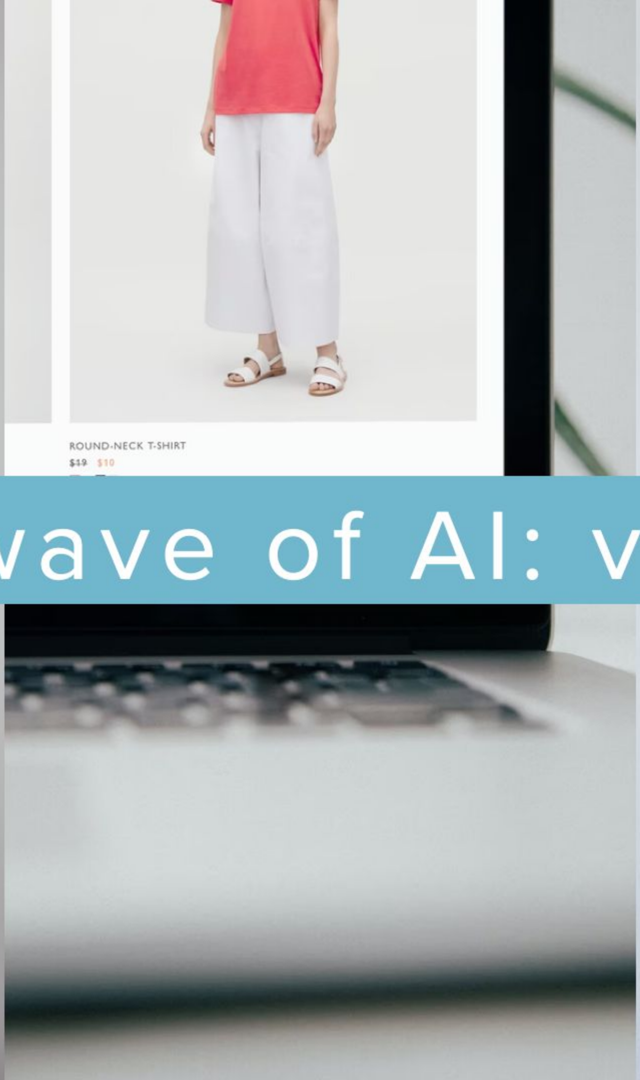
Alphastar winning Starcraft (Deepmind)

MuZero learning Go, Chess, Shogi, and Atari (Deepmind)

BIOCHEMISTRY



Alphafold solving protein folding (Deepmind)

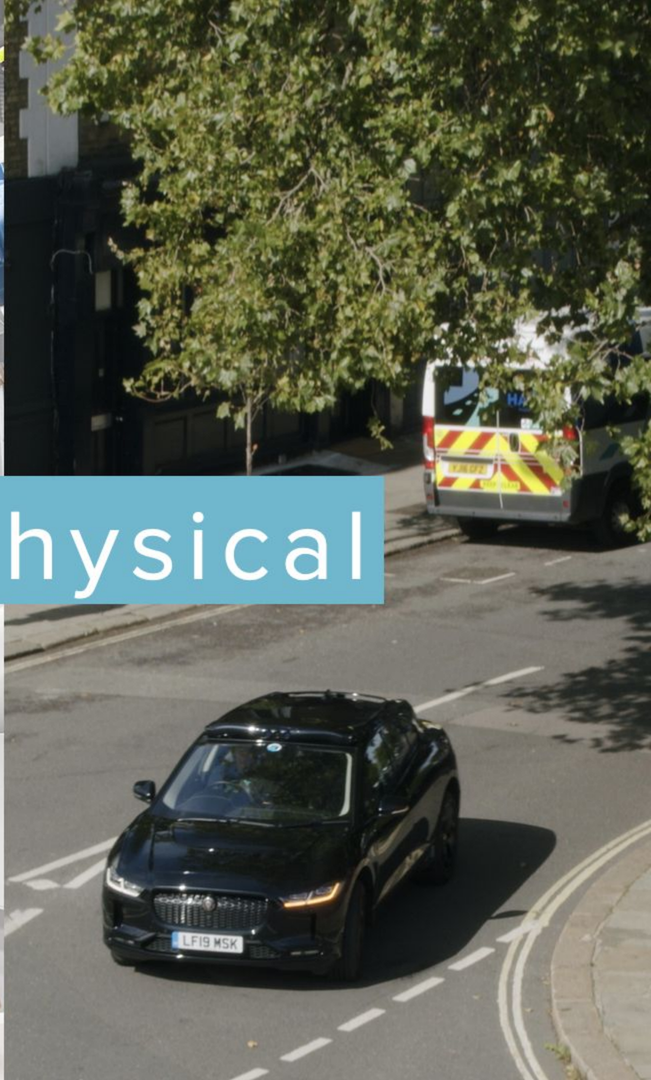


First wave of AI: virtual

Next wave of AI: physical

Photo by Skydio

Photo by Covariant



Autonomous driving is an embodied intelligence problem

Alex Kendall, Jeffrey Hawke, David Janz, Przemyslaw Mazur, Daniele Reda, John-Mark Allen, Vinh-Dieu Lam, Alex Bewley, Amar Shah. **Learning to Drive in a Day.** ICRA, 2019.

Learning to drive in London

Autonomous driving from monocular cameras and end-to-end deep learning.
No HD mapping, unnecessary sensors or hand-coded rules. Just pure intelligence.

AV 2.0: learning to drive with end-to-end deep learning and computer vision

Driving Input, 10^8 dimensions

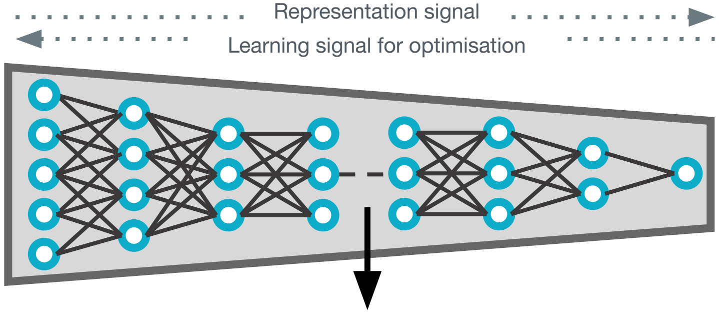
Cameras (6 @ 25 Hz)

GNSS

Goal conditioning from standard sat-nav Map

Vehicle State

+ others where required

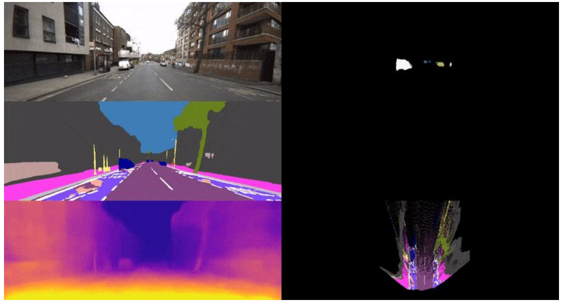


Driving Output, 10^1 dimensions

Motion Plan

Vehicle Controls

Decoded human-interpretable intermediate representations



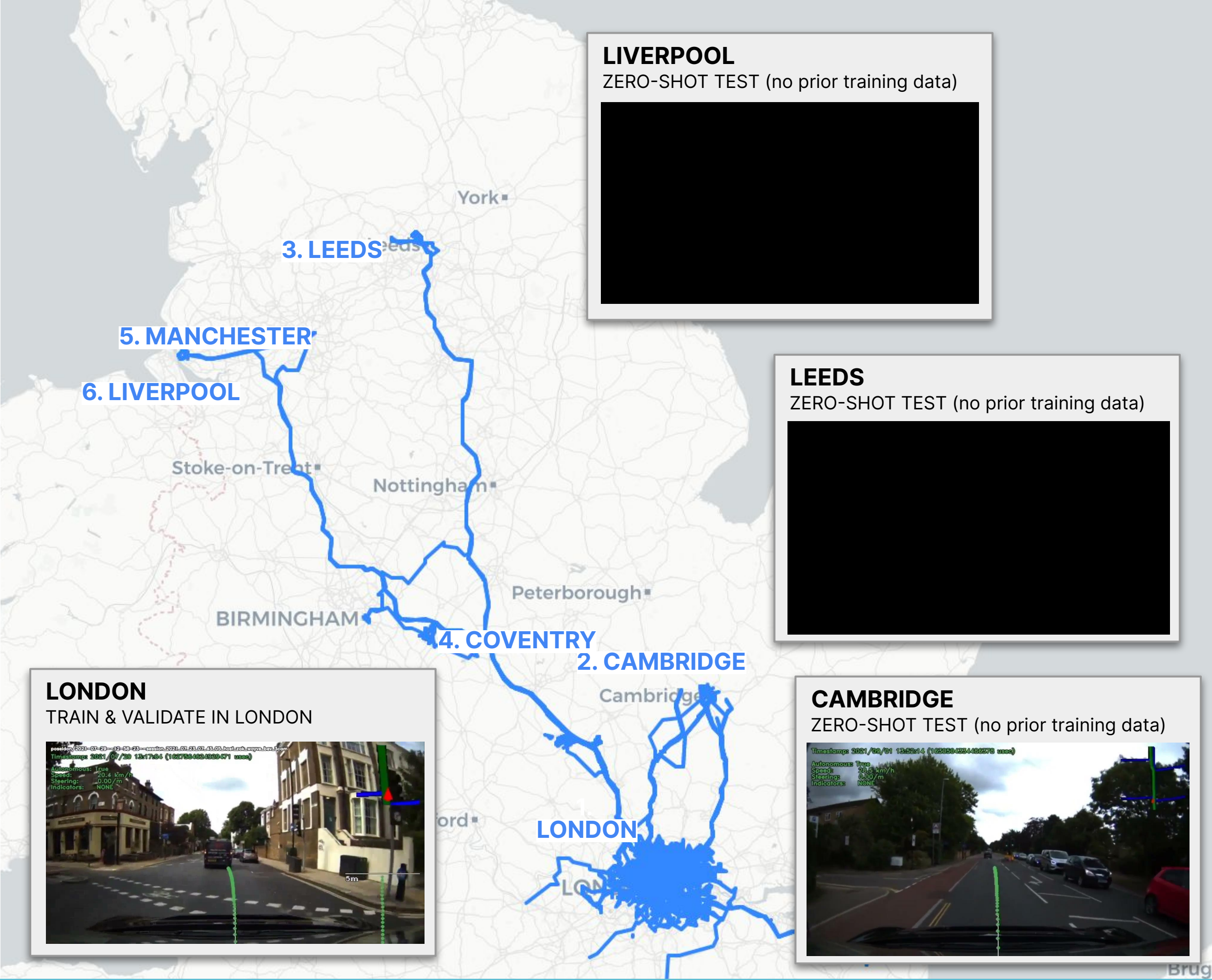
Semantics, geometry, motion prediction.

MULTI-CITY GENERALISATION TEST

Unlocking new markets faster

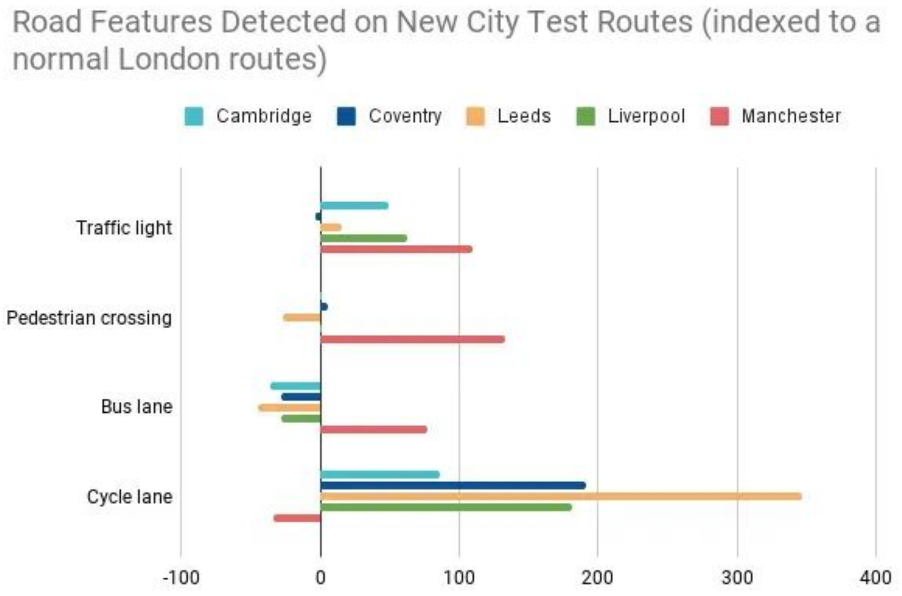
We first trained a driving model using only training data collected in London, UK.

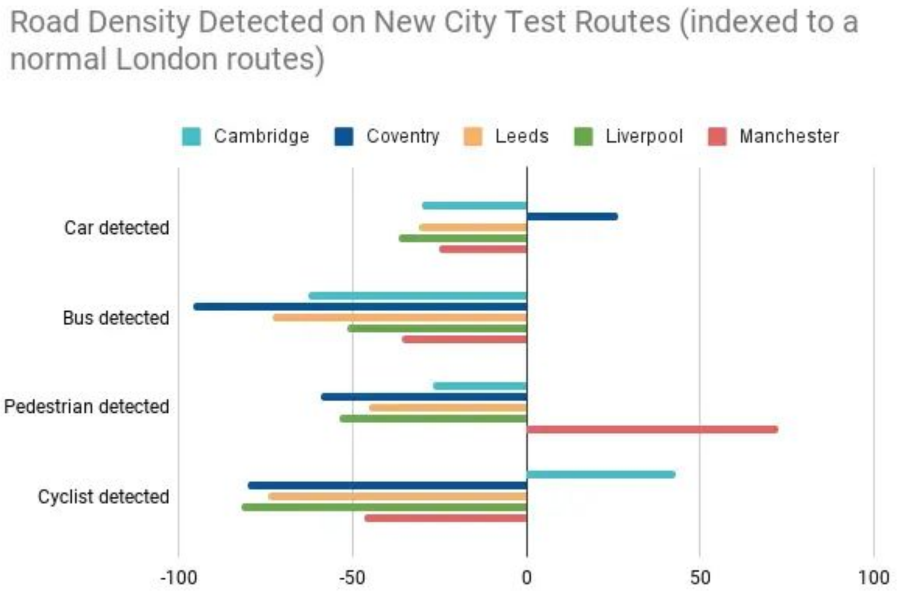
We then tested this model in five other UK cities, exposing it to diverse driving scenarios over a period of two months.



MULTI-CITY GENERALISATION TEST

Quantifying driving domain differences between cities






Learning to drive in London


6 (1630504316126224 usec)
First time driving in Cambridge


9 13:09:15 (1627564155645506 usec)
Learning to drive in London


12:12:09 (1631794329363656 usec)
First time driving in Coventry


05:15 (1627563915963480 usec)
20
Learning to drive in London


2:40:17 (1632400817448611 usec)
First time driving in Liverpool


/21 13:24:54 (1626873894685236 usec)
Learning to drive in London
13:46:14 (1631022374274544 usec)
First time driving in Leeds


10:28:51 (1627036131180629 usec)
Learning to drive in London
First time driving in Manchester

Part 2: Sensor Design on an Autonomous Vehicle

Sensing

The dominant sensor modalities used in robotics are:

Proprioceptive (internal state)

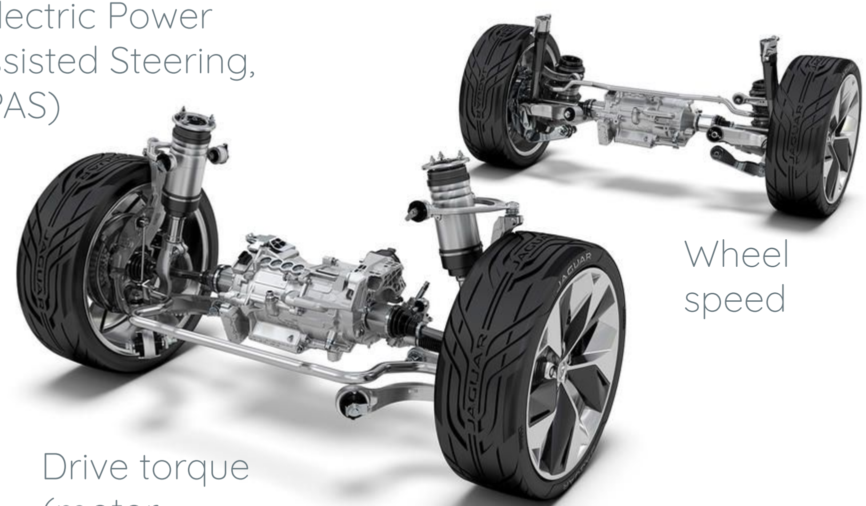
- Actuators (i.e., motor speed, position)
- Inertial Measurement Unit (IMU)

Exteroceptive (external state)

- Global Navigation Satellite System (GNSS)
- RADAR
- LiDAR (a.k.a. laser sensors)
- Cameras

Sensing: Actuators

Steering motor position (Electric Power Assisted Steering, EPAS)

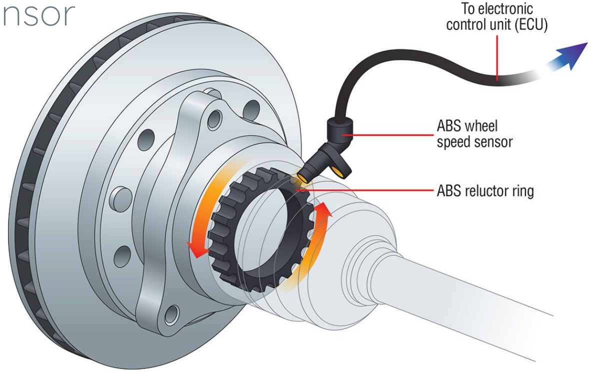


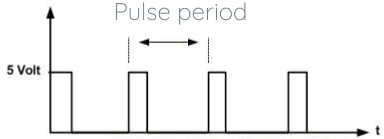
Wheel speed

Drive torque (motor current)

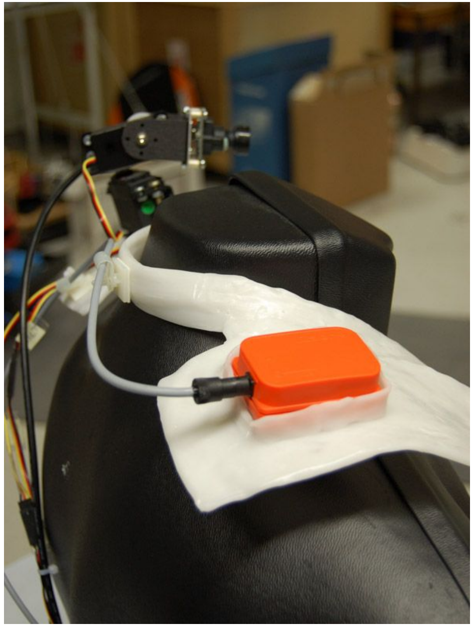
Wheel speed

Rotary speed/position encoders measure the motion of teeth past a sensor



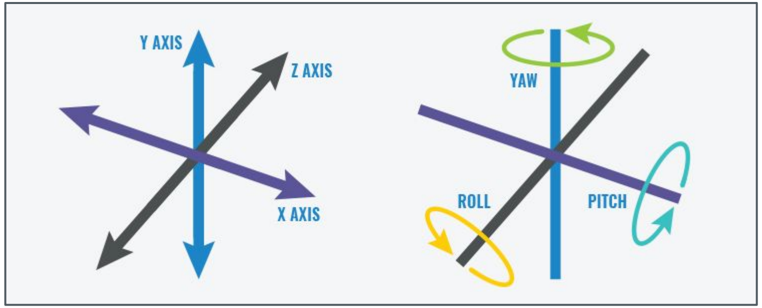


Sensing: Inertial Measurement



Microelectromechanical system (MEMS)

- Acceleration sensing (3D)
- Angular velocity sensing (3D)



IMUs are extremely useful, but suffer from drift over time

Sensing: Global Navigation Satellite System (GNSS)

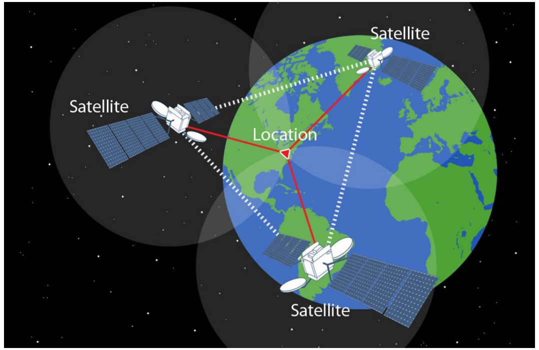
Pros:
Global 2.5D positioning: [x, y, θ]

Cons:

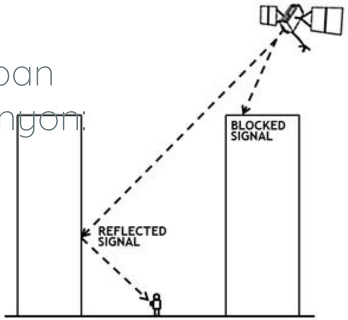
~1-10m accuracy
Consumer-grade limited to ~5Hz

Requires 4+ satellites for a fix

Urban canyons hugely degrade GNSS performance: multipath effects + blocked signal



Urban Canyon:



Sensing: Cameras

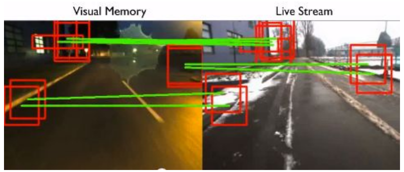
Typical camera:

~1-8MP, ~8-14 bit colour depth (Red, Green, Blue), 30-200Hz

Tradeoff: frame rate vs resolution – limited by serial data rate and compute

Monocular cameras

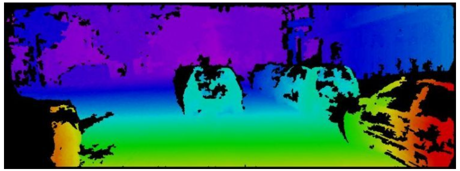
Optical flow, visual odometry, and localisation



Object detection, tracking, segmentation

Stereo Cameras

Depth sensing from a pair of images

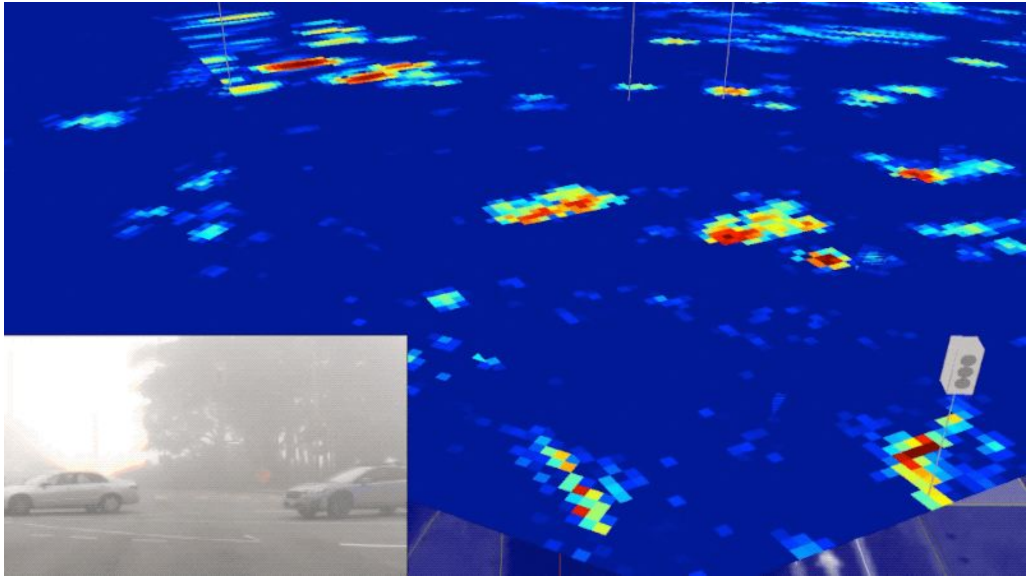


Sensing: Radar

Pros:
Depth sensing robust to weather, lighting conditions, ~200m+ range, can 'see through'

Cons:
Noisy, multipath effects

E.g., Continental ARS441



E.g., Waymo imaging radar visualisation

Sensing: LIDAR

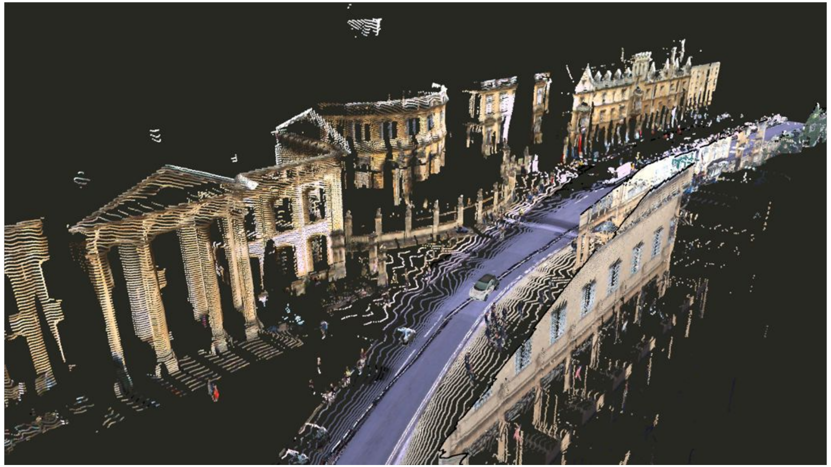
Pros:

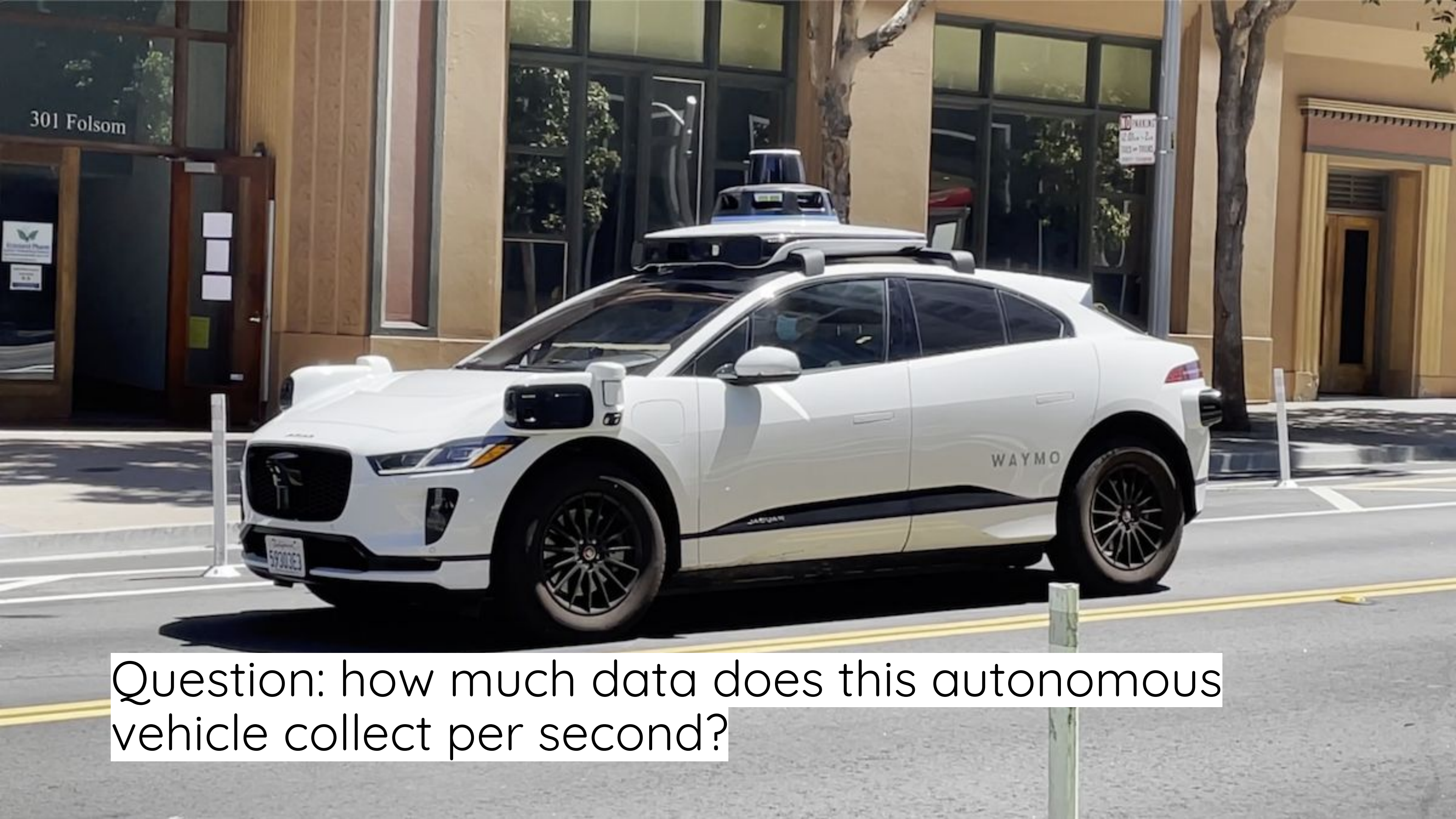
- Depth sensing robust to lighting conditions, very accurate with low noise
- 100-300m+ range pointcloud: [x, y, z]
- 0.3-10M points/second at 5-20Hz

Cons:

- Degraded by rain, snow
- Expensive (though improving)

E.g., Velodyne 3D lidar





Question: how much data does this autonomous vehicle collect per second?

Question: how much data does this autonomous vehicle collect per second?

6x forward facing cameras, 4x cameras per side: ~10MP images, 15Hz

= 150M pixels / sec / camera = 8,100 MB/s → 810 MB/s after compression

+

6× 3D LiDAR: ~128 vertical * 3600 horizontal * 10Hz = 5M points / sec / lidar = 332 MB/s

+

3x Radar: 256 channel * 512 depth * 30Hz = 4M points / sec / radar = 94 MB / s

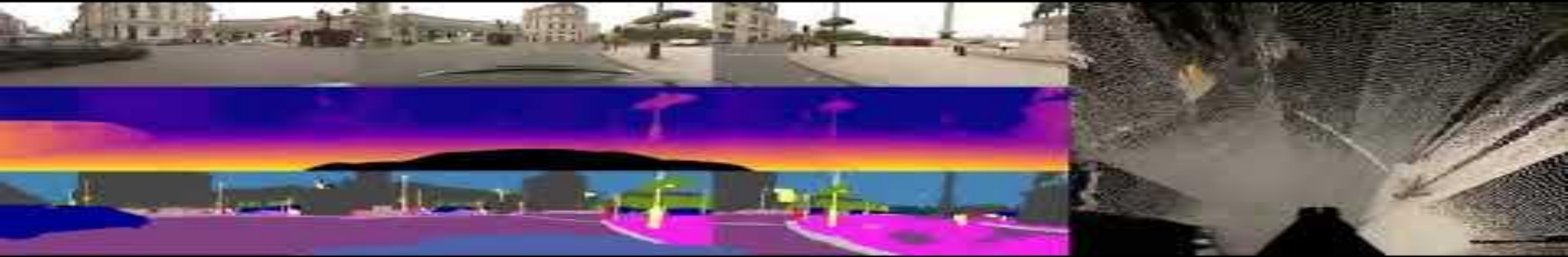
+

Vehicle state (speed, location, etc) = minimal

=

1.2 GB / s ! (or 30+ PB / day !)

Computer Vision



3D geometric and semantic perception from surround vehicle monocular cameras

Part 3: Offboard Software

A large part of the complexity in developing an AV is off board the vehicle. AVs require substantial software platforms and tools:

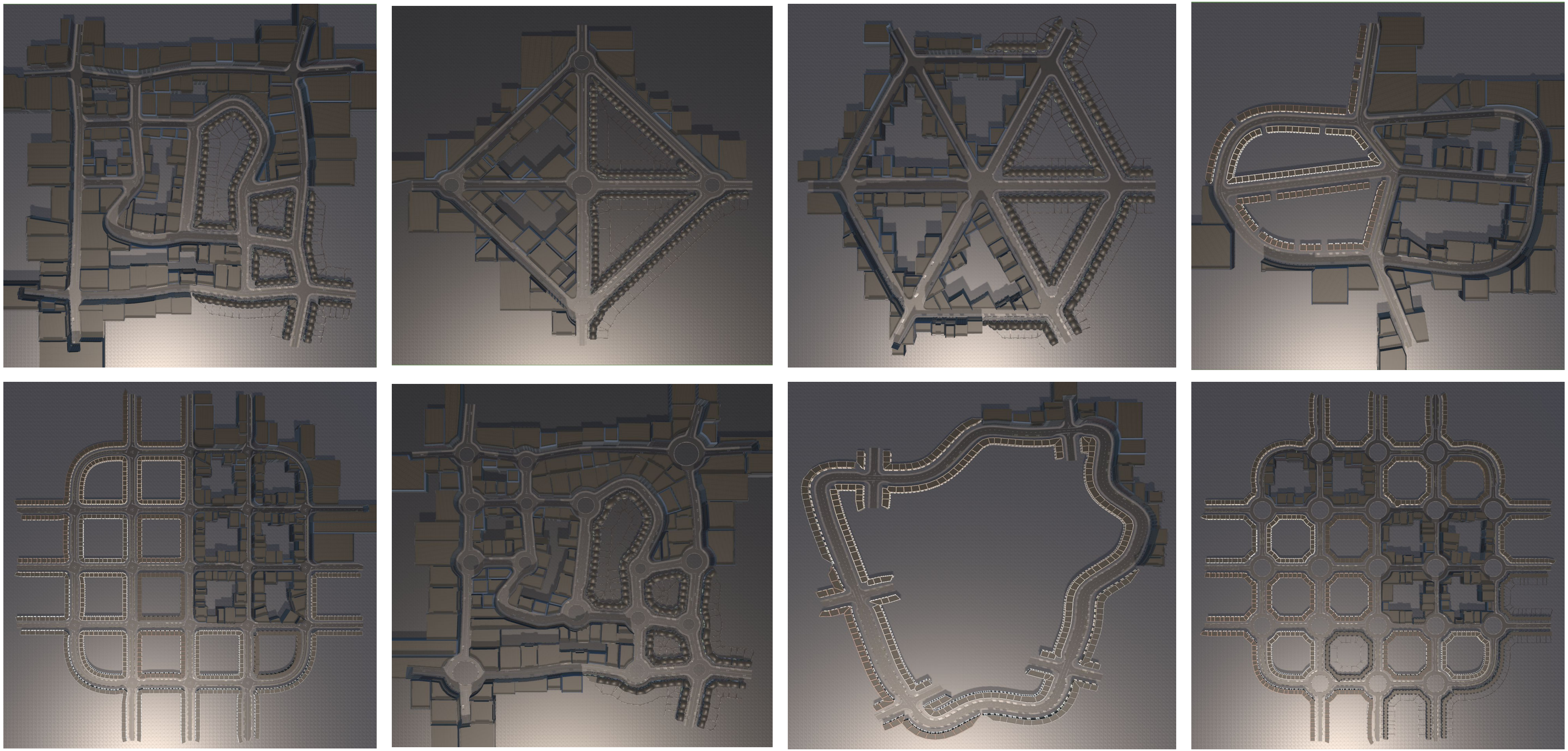
- Data storage and log replay
- machine learning training systems
- Validation and verification
- Fleet management
- Simulation


WAYVE

Simulation for Autonomous Driving

- Simulation is crucial to scale autonomy.
- Simulation offers limitless, realistic (visual, behavioural, and kinematic), unbiased, diverse training and evaluation samples.
- Simulation provides statistically significant insights at 1/10th the cost, 10x the speed, and 10x the repeatability compared to real-world testing.
- Simulation offers probably the only route to cost-effective deployment.

Simulation World Diversity



Visual Diversity

Part 4: Safety

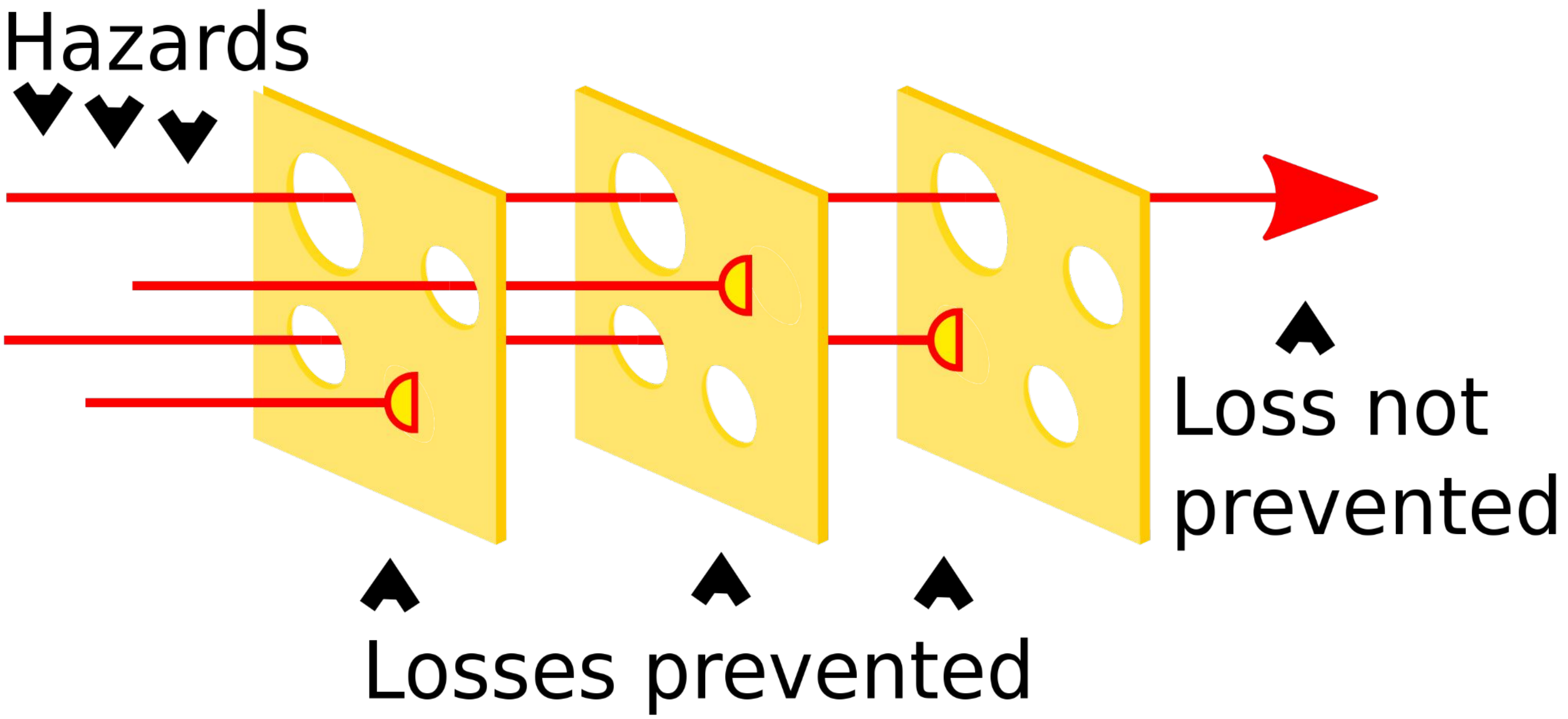
Important Safety Concepts

- **Operational Design Domain (ODD):** the operating environment within which the autonomous vehicle can perform safely, described by the static elements, dynamic elements and environmental elements of the scene
- **Safety Case:** a structured argument, supported by evidence, intended to justify that a system is acceptably safe for a specific application in a specific operational design domain.
- **Functional Safety:** Ensuring electronic failures will not cause unacceptable risks to human life
- **Safety of the Intended Functionality (SOTIF):** Ensuring autonomous vehicle behaviour is absent of unacceptable risk

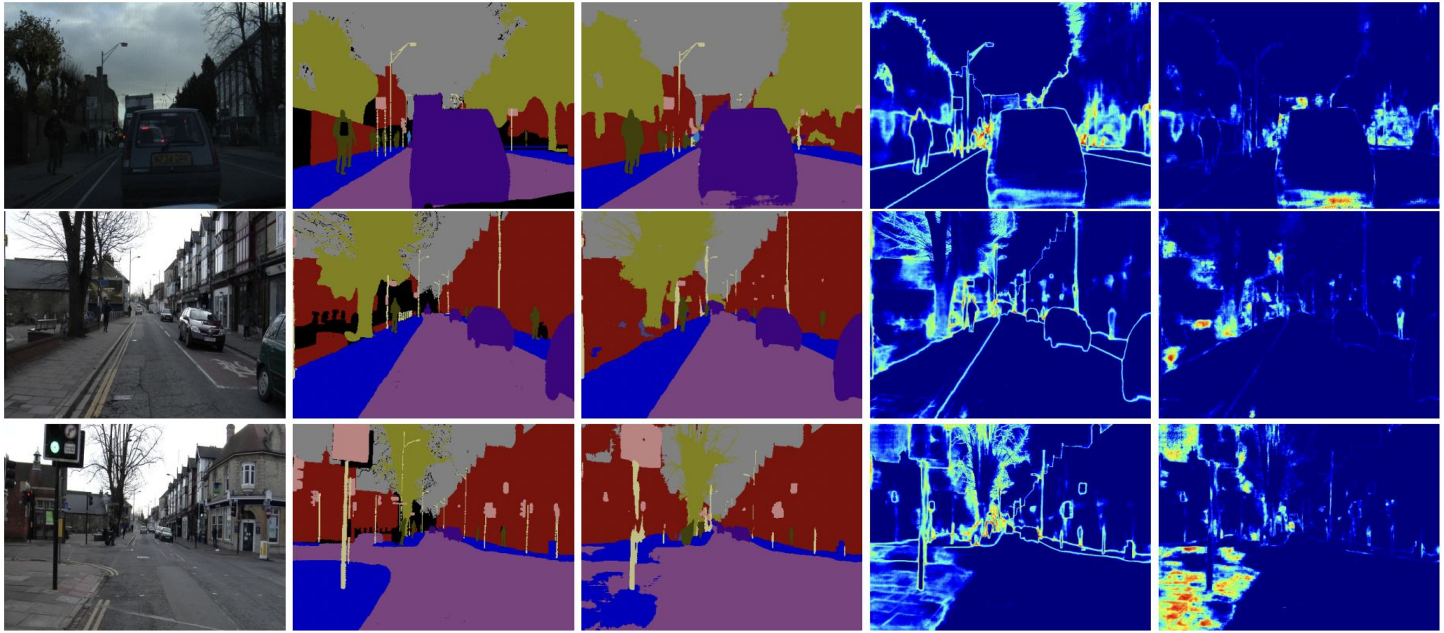
Safety != perfect: Case study on human performance

Metric	Method of calculation	UK Human-Level Performance	Notes / Method of estimation
Intervention rate (km / intervention)	total autonomous distance driven divided by number of valid interventions (ignoring platform-fault and end-of-run success interventions)	95,400 km / intervention	13.1% of vehicles make motor insurance claims per year (source, 2016). Vehicles travel on average 12,500 km per year (source, 2016). Therefore we can estimate distance per accident for UK drivers as 95,400 km / accident.
Speed Limit Compliance (%)	% of time the vehicle drives in excess of the speed limit provided by our map API	97.993 %	calculated from 100 hours of expert human driving data
Lane Following (km / int)	total autonomous distance driven divided by number of valid, non-junction interventions	221,400 km / intervention	43.09 % of total accidents occur during lane following. Assuming the distance driven by a car is negligible in junctions, this value is overall Intervention rate divided by 43.1% (source)
Unprotected Intersections (%)	% of unsignalised intersections successfully navigated without any valid intervention within 10m before or after the junction	99.999238 % (every 130,000 intersections)	From the information about roundabouts below, and an estimate that roundabouts are ~40% safer than junctions (source), we estimate that humans are 99.999238% successful at junctions.
Protected Intersection (%)	% of traffic lights successfully navigated without any valid intervention within 10m before or after the junction		
Roundabouts (%)	% of roundabouts successfully navigated without any valid intervention within 10m before or after the junction	99.999543 % (every 220,000 intersections)	There are ~25,000 roundabouts in the UK (source) and 397,025 kms of road network (source) in GB. Assuming the number of roundabout in the UK is approximately equal to GB (NI only has 2% of UK's roads (source)), there are 0.0629 roundabouts / km in UK. On average, vehicles encounter 2,533M roundabouts per year, considering 40,234M kms are driven per year (source). Therefore, humans have a 99.999543% success at roundabouts, considering total UK roundabout accidents is 11,571 per year (source).

Swiss cheese model of Safety



Understanding epistemic model uncertainty: When and what we don’t know



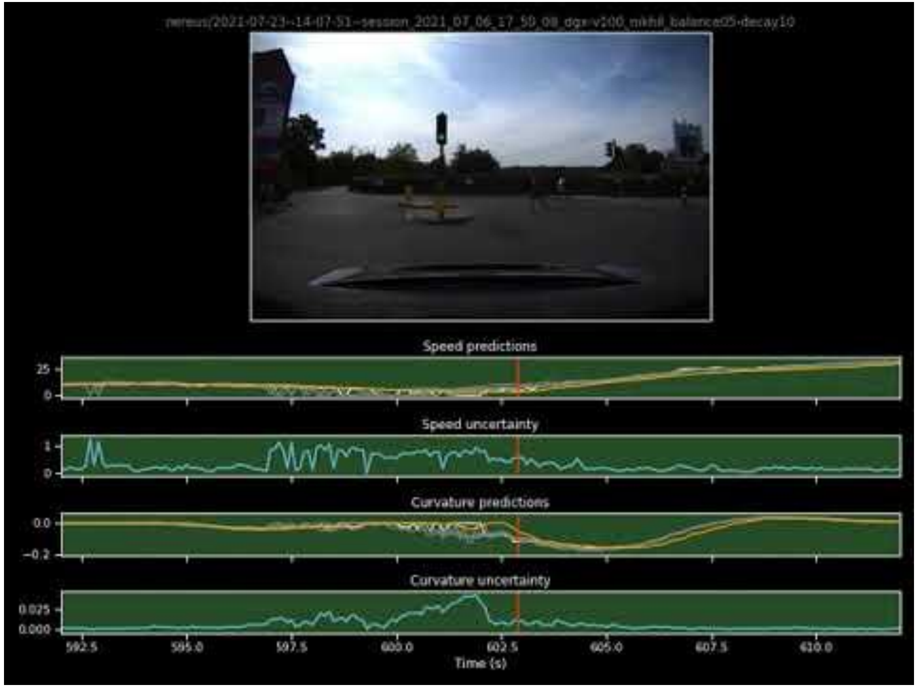
(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

1. *Alex Kendall and Yarin Gal.* ***What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?*** *NeurIPS, 2017.*
2. *Alex Kendall, Vijay Badrinarayanan and Roberto Cipolla.* ***Bayesian SegNet: Model Uncertainty for Scene Understanding.*** *BMVC, 2017.*

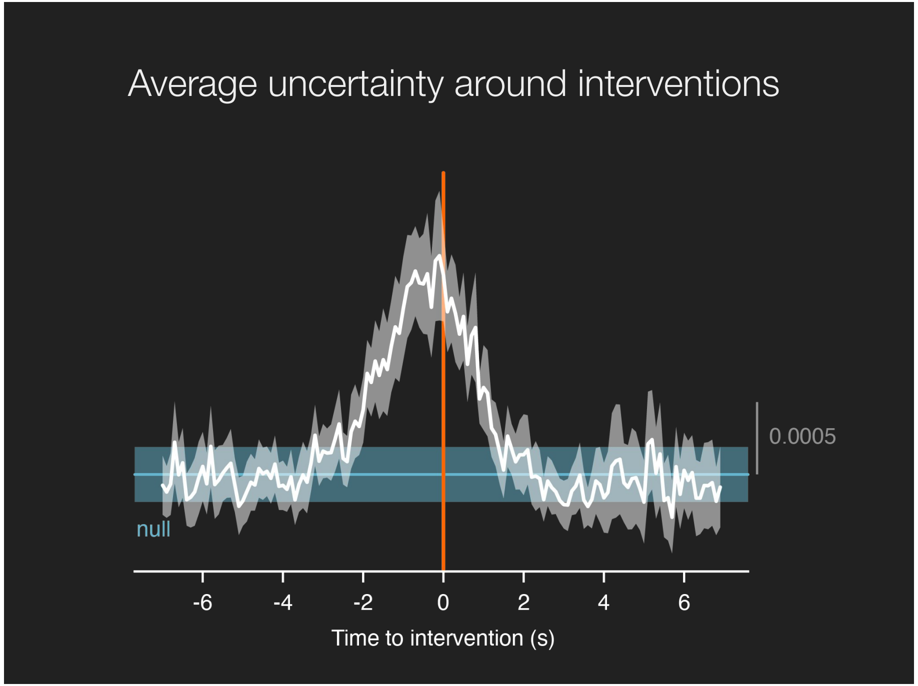
Measuring Uncertainty for Autonomous Driving

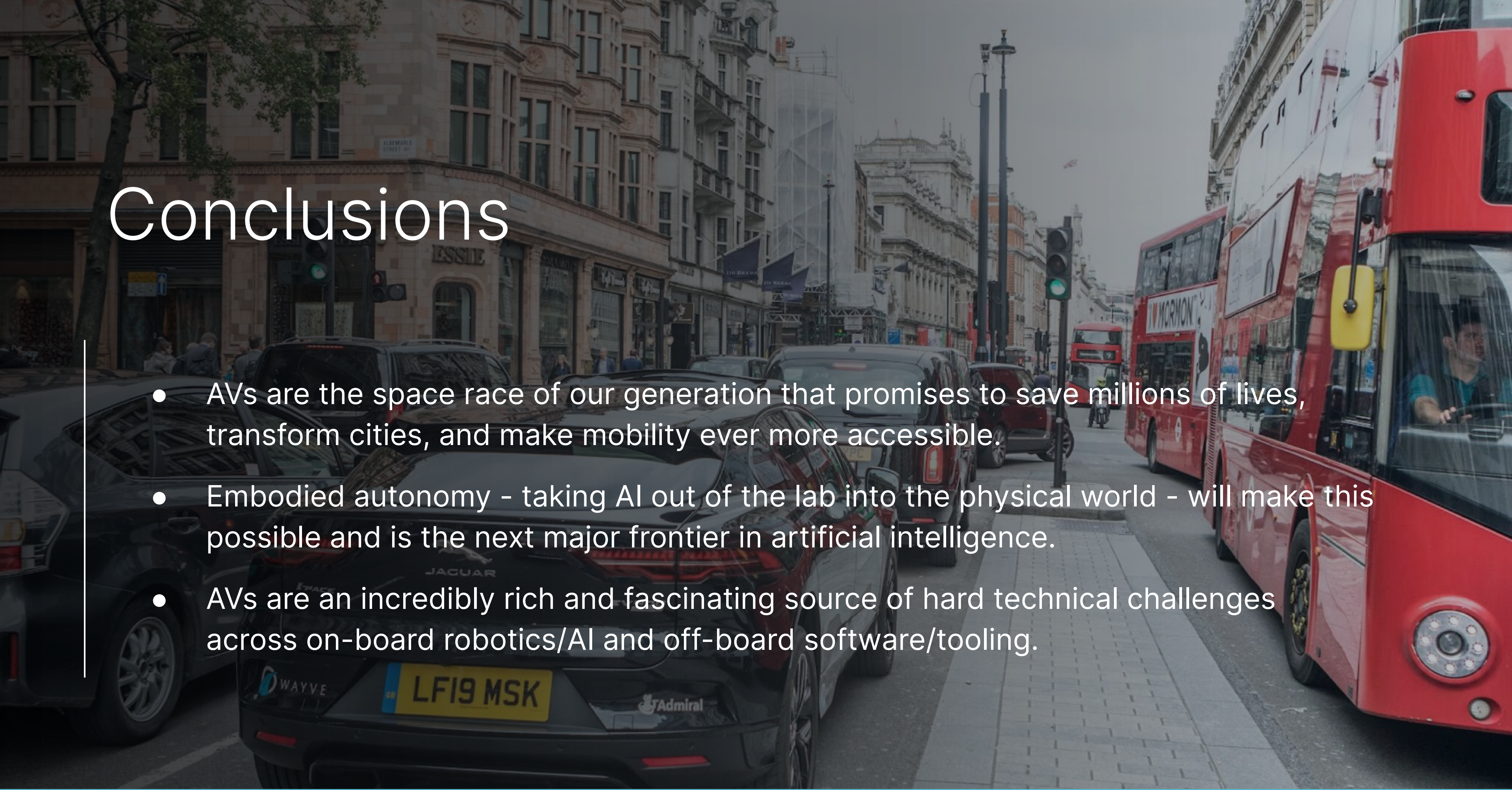
Real world closed-loop testing

High uncertainty, **no intervention**



Intervention during **high uncertainty**





Conclusions

- AVs are the space race of our generation that promises to save millions of lives, transform cities, and make mobility ever more accessible.
- Embodied autonomy - taking AI out of the lab into the physical world - will make this possible and is the next major frontier in artificial intelligence.
- AVs are an incredibly rich and fascinating source of hard technical challenges across on-board robotics/AI and off-board software/tooling.

Interested in tackling the
space race of our generation with Wayve?

careers@wayve.ai

Further reading:

1. **Learning to Drive in a Day,** ICRA 2019. Alex Kendall, Jeffrey Hawke, David Janz, Przemyslaw Mazur, Daniele Reda, John-Mark Allen, Vinh-Dieu Lam, Alex Bewley, Amar Shah.

2. **Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics**, CVPR 2018. Alex Kendall, Yarin Gal, Roberto Cipolla.

3. **Learning to Drive from Simulation without Real World Labels,** ICRA, 2019. Alex Bewley, Jessica Rigley, Yuxuan Liu, Jeffrey Hawke, Richard Shen, Vinh-Dieu Lam and Alex Kendall.

4. **Orthographic Feature Transform for Monocular 3D Object Detection**, BMVC 2019. Thomas Roddick, Alex Kendall, Roberto Cipolla.

5. **Learning a Spatio-Temporal Embedding for Video Instance Segmentation**, arXiv 2020. Anthony Hu, Alex Kendall and Roberto Cipolla.

6. **Urban Driving with Conditional Imitation Learning**, ICRA, 2020. Jeffrey Hawke et al.

7. **Probabilistic Future Prediction for Video Scene Understanding**, ECCV 2020. Anthony Hu, Fergal Cotter, Nikhil Mohan, Corina Gurau, Alex Kendall.

8. **FIERY: Future Instance Prediction in Bird's-Eye View from Surround Monocular Cameras,** ICCV 2021. Anthony Hu, Zak Murez, Nikhil Mohan, Sofía Dudas, Jeff Hawke, Vijay Badrinarayanan, Roberto Cipolla, Alex Kendall.

9. **Reimagining an autonomous vehicle**, arXiv 2021. Jeff Hawke, Haibo E, Vijay Badrinarayanan, Alex Kendall.