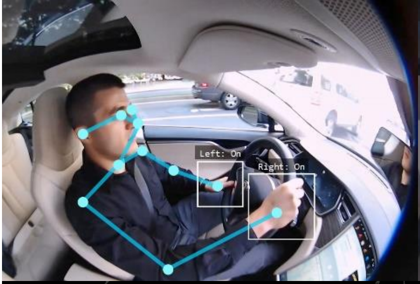
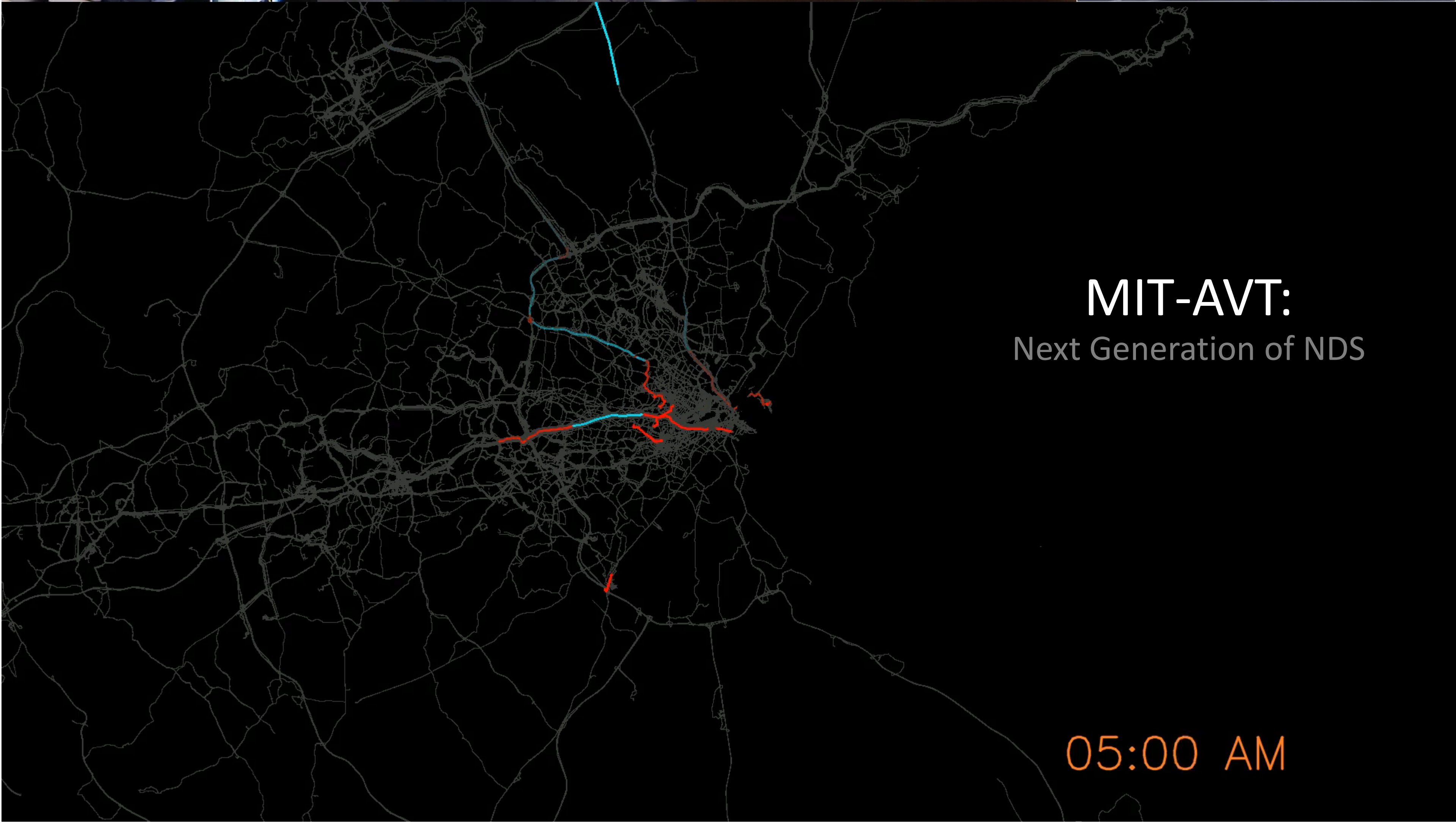




MIT-AVT:

Next Generation of NDS

Human-Centered Artificial Intelligence

https://hcai.mit.edu

MIT Autonomous Vehicle Technology Study

https://hcai.mit.edu

MIT Autonomous Vehicle Technology Study: Large-Scale Deep Learning Based Analysis of Driver Behavior and Interaction with Automation

Lex Fridman*, Daniel E. Brown, Michael Glazer, William Angell, Spencer Dodd, Benedikt Jenik, Jack Terwilliger, Julia Kindelsberger, Li Ding, Sean Seaman, Hillary Abraham, Alea Mehler, Andrew Sipperley, Anthony Pettinato, Bobbie Seppelt, Linda Angell, Bruce Mehler, Bryan Reimer*

[cs.CY] 19 Nov 2017

***Abstract*—Today, and possibly for a long time to come, the full driving task is too complex an activity to be fully formalized as a sensing-acting robotics system that can be explicitly solved through model-based and learning-based approaches in order to achieve full unconstrained vehicle autonomy. Localization, mapping, scene perception, vehicle control, trajectory optimization, and higher-level planning decisions associated with autonomous vehicle development remain full of open challenges. This is especially true for unconstrained, real-world operation where the margin of allowable error is extremely small and the number of edge-cases is extremely large. Until these problems are solved, human beings will remain an integral part of the driving task, monitoring the AI system as it performs anywhere from just over 0% to just under 100% of the driving. The governing objectives of the MIT Autonomous Vehicle Technology (MIT-AVT) study are to (1) undertake large-scale real-world driving data collection that includes high-definition video to fuel the development of deep learning based internal and external perception systems, (2) gain a holistic understanding of how human beings interact with vehicle automation technology by integrating video data with vehicle state data, driver characteristics, mental models, and self-reported experiences with technology, and (3) identify how technology and other factors related to automation adoption and use can be improved in ways that save lives. In pursuing these objectives, we have instrumented 21 Tesla Model S and Model X vehicles, 2 Volvo S90 vehicles, and 2 Range Rover Evoque vehicles for both long-term (over a year per driver) and medium term (one month per driver) naturalistic driving data collection. Furthermore, we are continually developing new methods for analysis of the massive-scale dataset collected from the instrumented vehicle fleet. The recorded data streams include IMU, GPS, CAN messages, and high-definition video streams of the driver face, the driver cabin, the forward roadway, and the instrument cluster (on select vehicles). The study is on-going and growing. To date, we have 78 participants, 7,146 days of participation, 275,589 miles, and 3.5 billion video frames. This paper presents the design of the study, the data collection hardware, the processing of the data, and the computer vision algorithms currently being used to extract actionable knowledge from the data.**

Overview

- **Approach**: Use computer vision (deep learning) to convert raw video data to knowledge in all data *before* considering epochs.
- **Challenges**
  - New algorithms
  - Compute resources to train neural network models
  - New annotation methods and tools…
    - to build supervised learning datasets for machines
    - to interpret and label highly subjective scenarios
  - Large-scale distributed compute for inference
  - Hot storage (a lot more read than write)
- **Deep Learning + NDS**
  - 300 Petabytes of data processed
  - 3 million hours of GPU-enabled, 16 core, 64-128gb RAM machines

Vehicles and Automation

Tesla Autopilot

Cadillac Super Cruise

Range Rover Lane Keep Assist

Volvo Pilot Assist II

MIT Autonomous Vehicle Technology Study

Study months to-date: **30**
Participant days: **11,846**
Drivers: **99**
Vehicles: **29**
Miles driven: **405,807**
Video frames: **5.5 billion**

Study data collection is ongoing.
Statistics updated on: ***Jul 20, 2018.***

Tesla Model S
39,320 miles
583 days in study

Tesla Model S
33,177 miles
861 days in study

Tesla Model X
31,600 miles
748 days in study

Tesla Model S
25,491 miles
572 days in study

Range Rover Evoque
22,957 miles
598 days in study

Range Rover Evoque
22,644 miles
763 days in study

Tesla Model X
21,915 miles
499 days in study

Tesla Model S
20,433 miles
647 days in study

Volvo S90
19,231 miles
634 days in study

Tesla Model X
17,035 miles
701 days in study

Tesla Model S
15,735 miles
322 days in study

Volvo S90
15,570 miles
672 days in study

Tesla Model S
15,256 miles
714 days in study

Tesla Model S
14,398 miles
371 days in study

Tesla Model S
13,010 miles
463 days in study

Tesla Model S
12,353 miles
321 days in study

Tesla Model S
10,149 miles
146 days in study

Tesla Model X
9,556 miles
378 days in study

Tesla Model S
9,076 miles
183 days in study

Tesla Model X
8,587 miles
316 days in study

Tesla Model S
8,484 miles
325 days in study

Tesla Model S
6,718 miles
194 days in study

Tesla Model X
4,587 miles
233 days in study

Tesla Model X
4,441 miles
416 days in study

Tesla Model S
2,925 miles
133 days in study

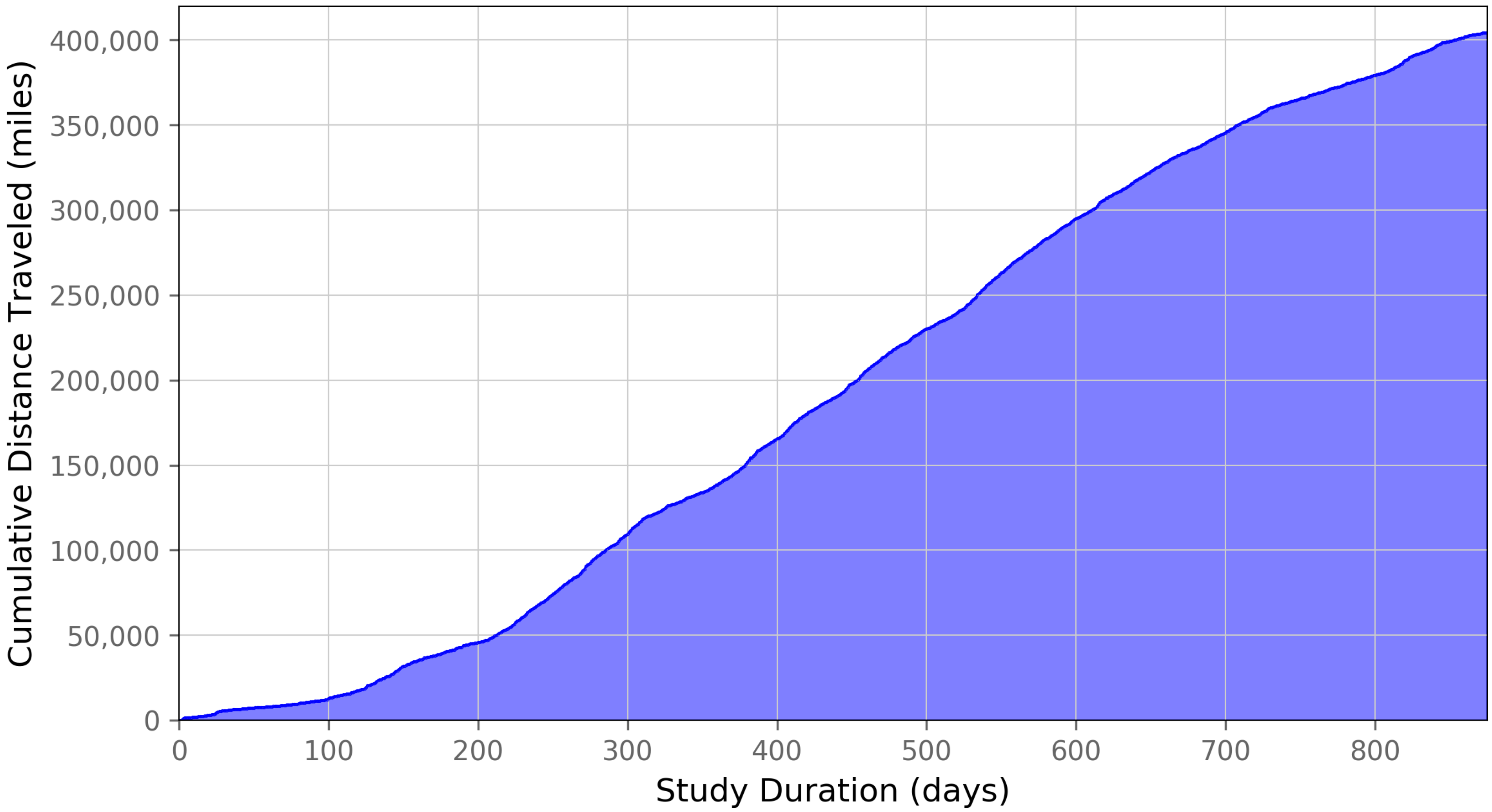
Cadillac CT6
1,161 miles
53 days in study

Tesla Model S
(Offload Pending)

Tesla Model S
(Offload Pending)

Cadillac CT6
(Offload Pending)

Dataset Growth

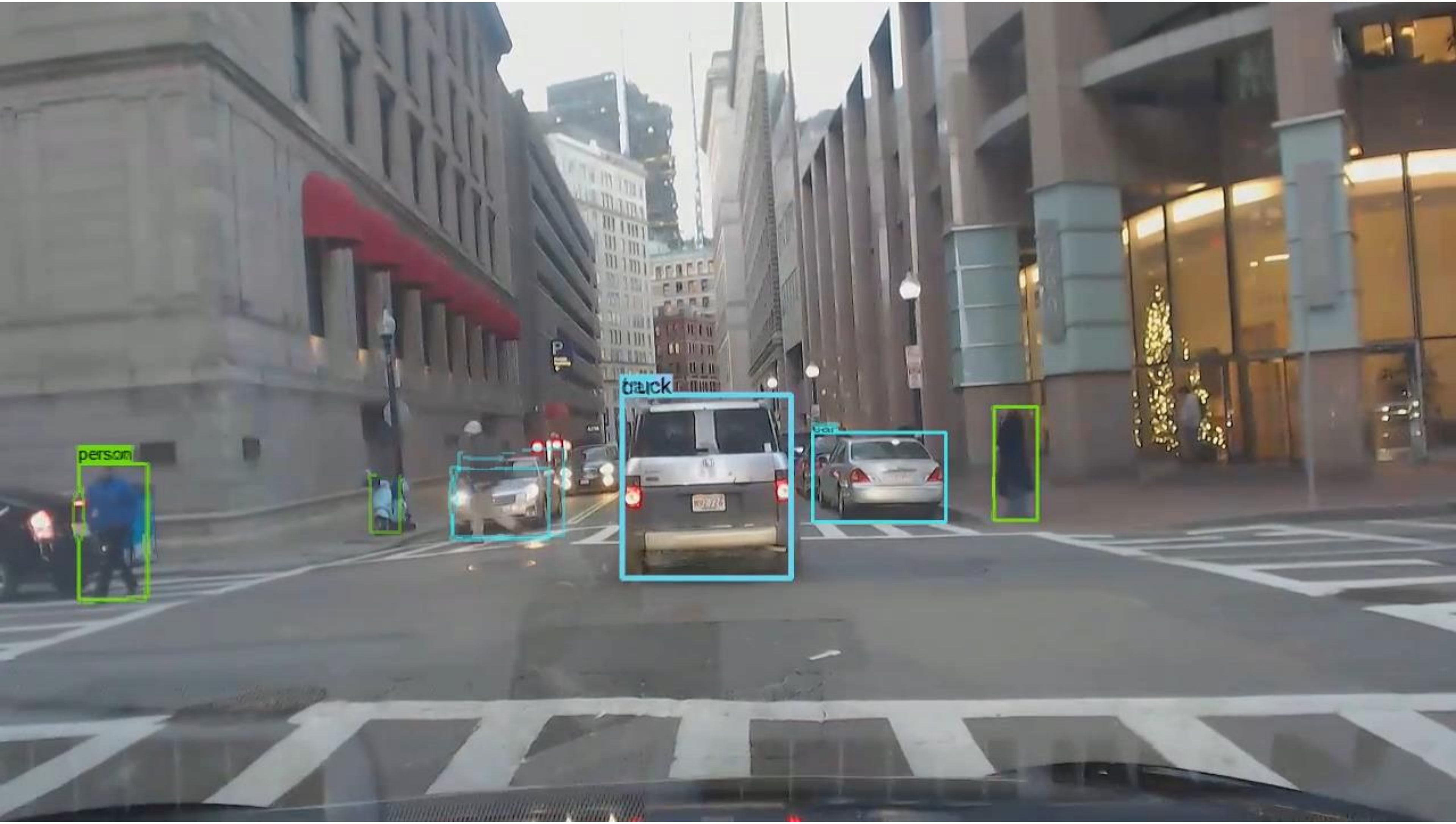


From Pixels to Knowledge:

Driving Scene Understanding

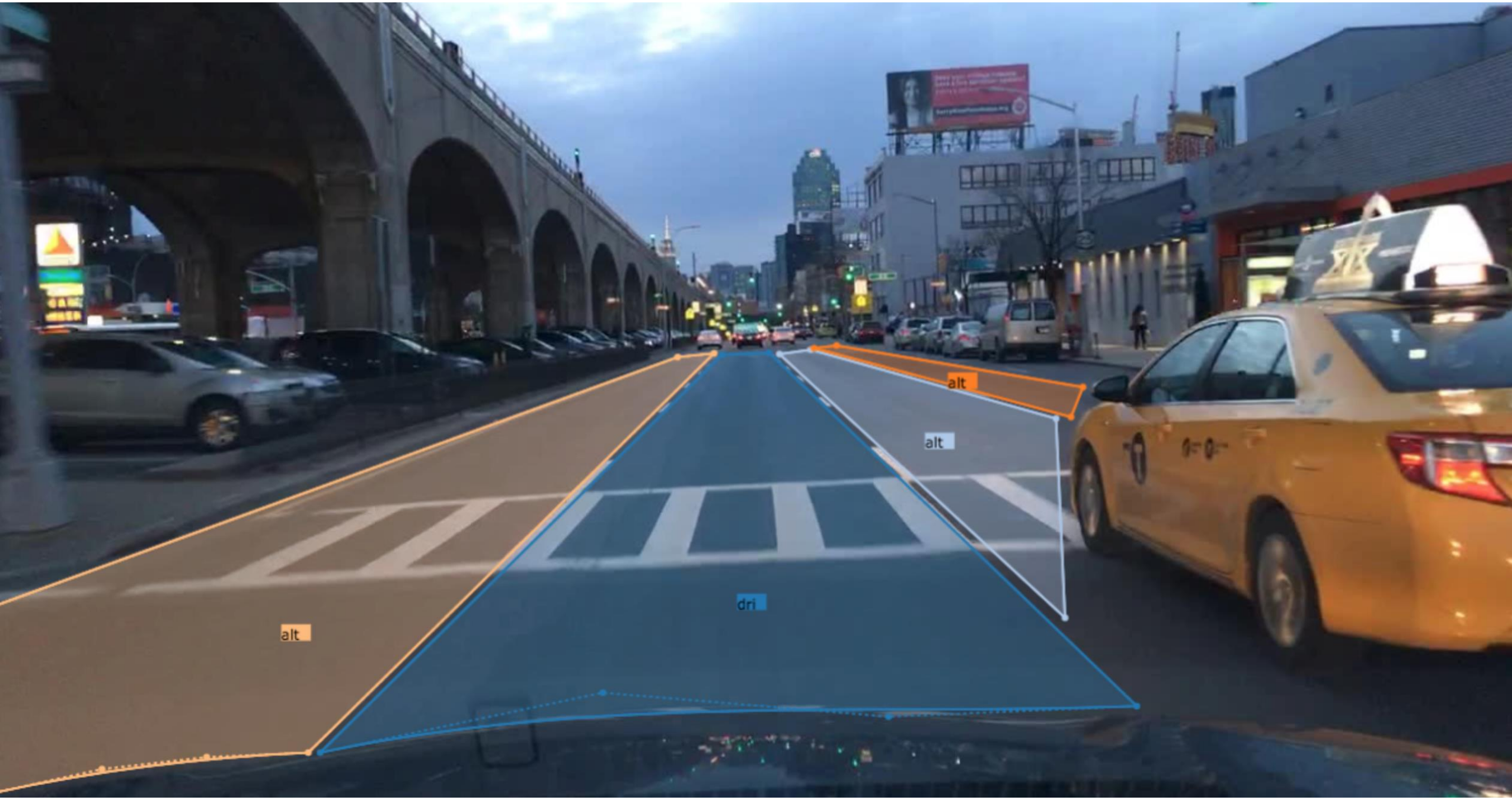
From Pixels to Knowledge:

Driving Scene Object Detection



From Pixels to Knowledge:

Lanes (“Drivable Area”)

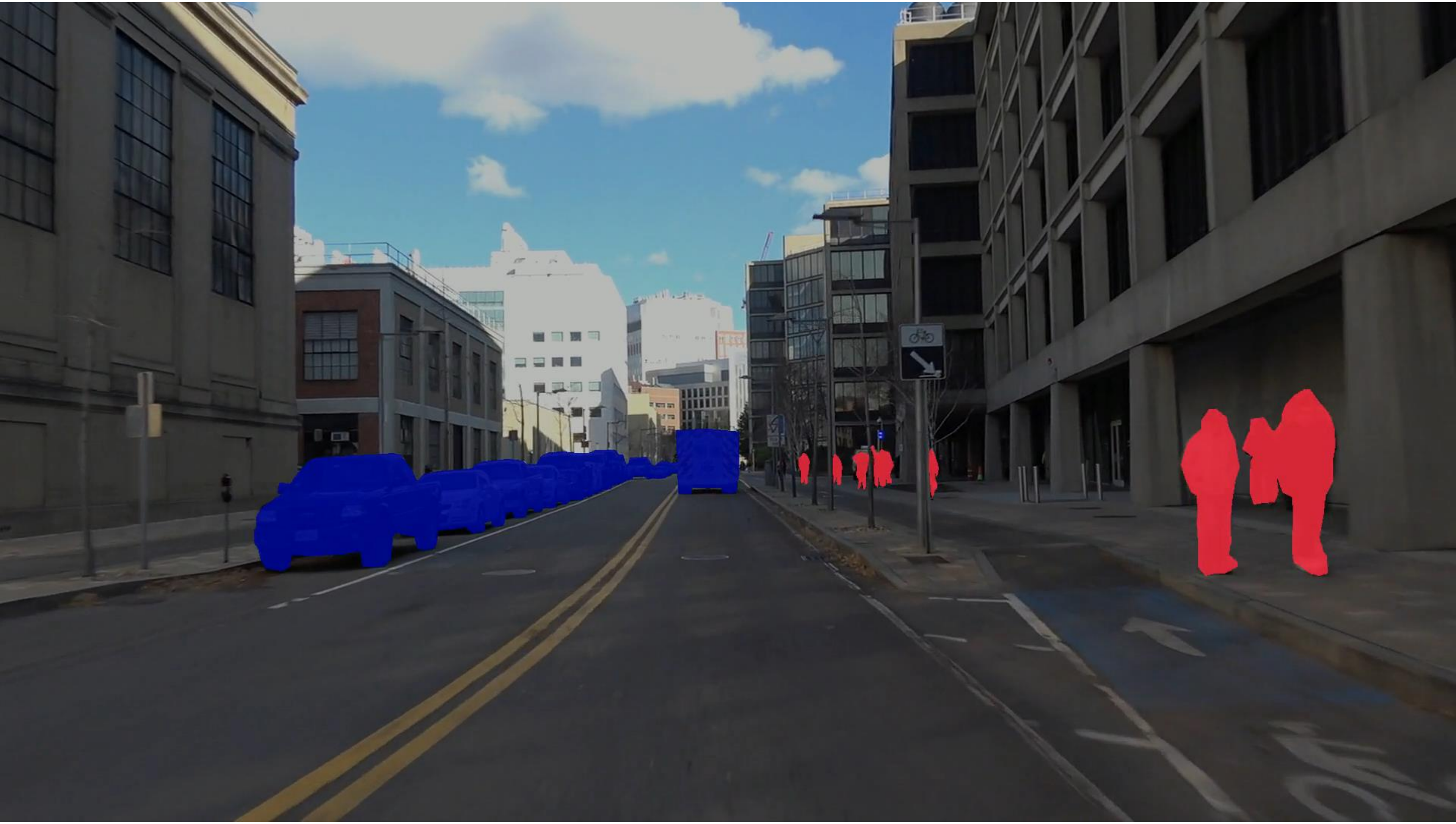


From Pixels to Knowledge:

Driving Scene Segmentation

From Pixels to Knowledge:

Driving Scene Segmentation

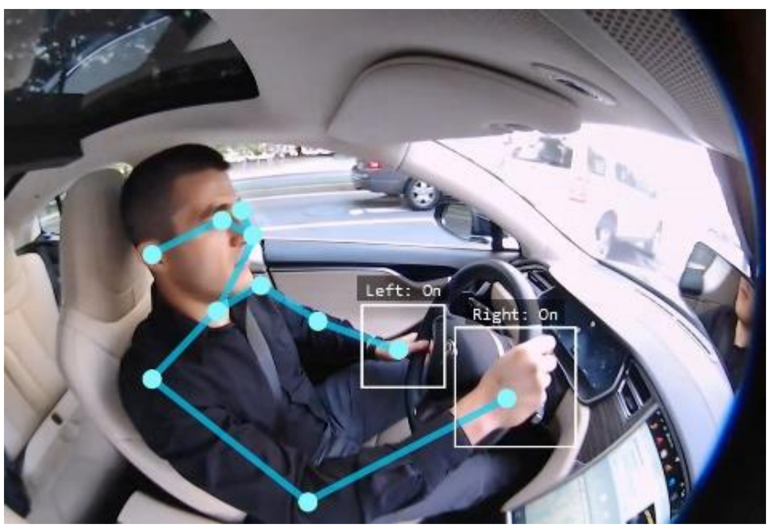


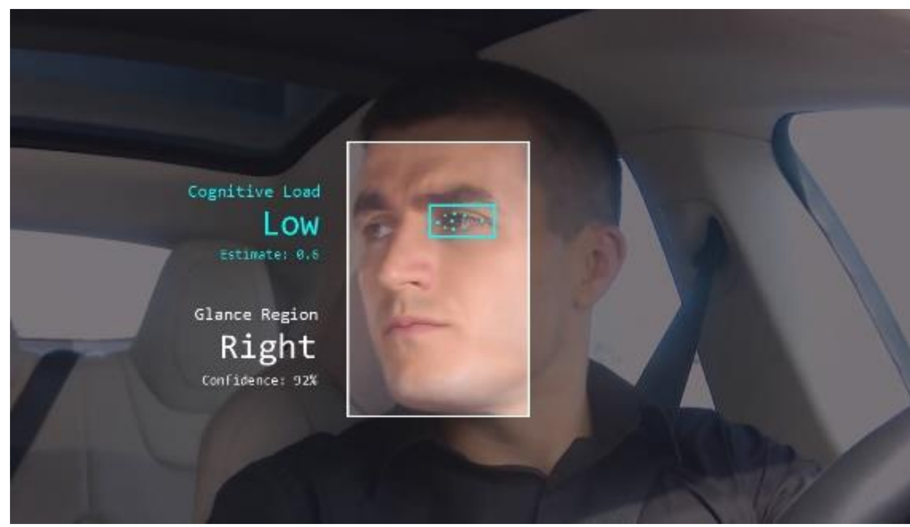
Driver State

Increasing level of detection resolution and **difficulty**

Pedestrian Detection | Body Pose | Head Pose | Blink Rate | Blink Duration | Eye Pose | Blink Dynamics | Pupil Diameter | Micro Saccades

Face Detection | Face Classification | Glance Classification | Drowsiness | Micro Glances | Cognitive Load





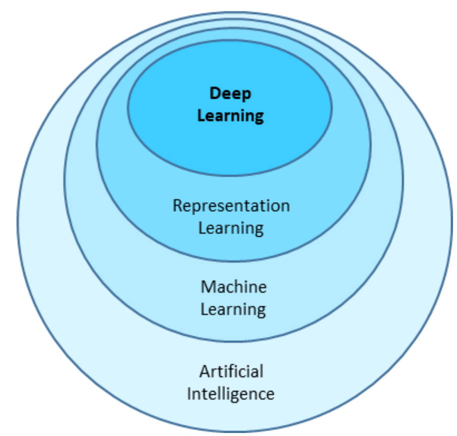
Deep Learning: Principles of Application

- Requirements for success (from more to less critical)
  - **Data:** A lot of real-world data (and algorithms that learn from data)
  - **Semi-supervised:** Human annotations of representative subsets of data
  - **Efficient annotation:** Specialized annotation tooling
  - **Hardware:** Large-scale distributed compute and storage
  - **Robustness:** Algorithms that don’t need calibration (learn the calibration)
  - **Temporal dynamics:** Algorithms that consider time
- Current importance relation for successful application of deep learning:

>

Good Algorithms*

* As long as they learn from data



Deep Learning for Driver State

What:

- Glance
  (CHI 2018)
- Cognitive Load
  (CHI 2017)
- Drowsiness
- Emotion
- Body
- Activity

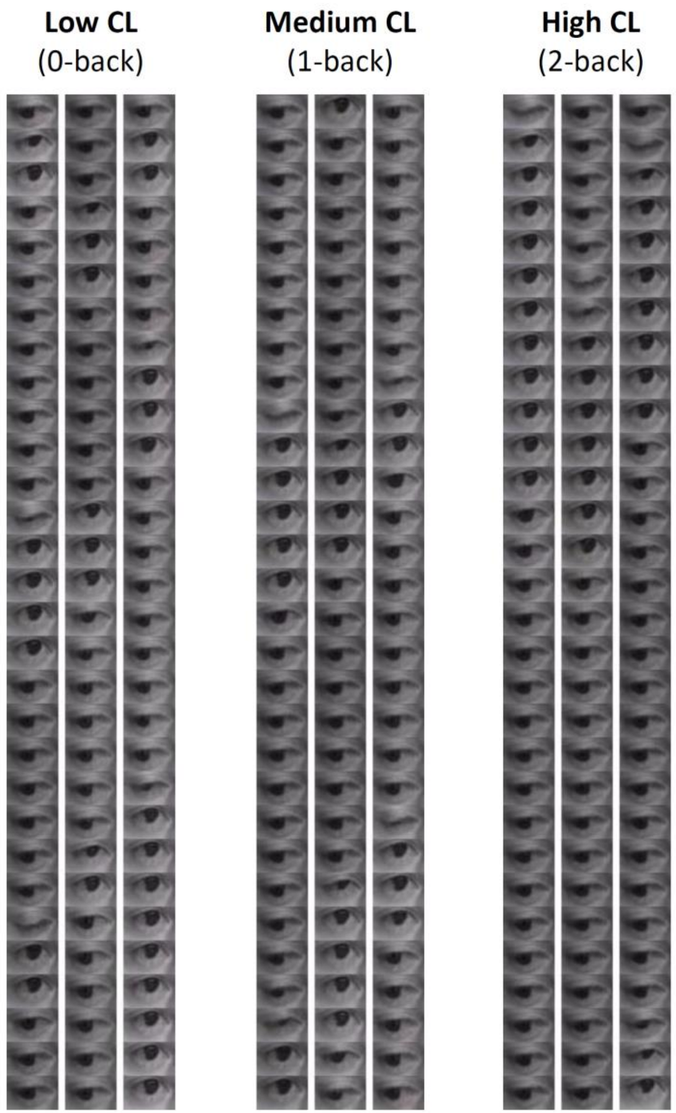
How:

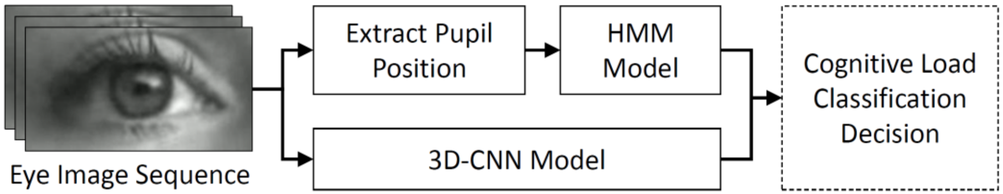
- Real-world data
- Formulation of the task such that it can be labeled and trained in a supervised way.
- Deep learning

Cognitive Load Estimation

- **What is it?**
  Algorithm to detect how hard you're "thinking" (accessing working memory) from camera
- **Where?**
  Real-world (aka anywhere)
- **Just driving?**
  No. Any activity (aka anywhere)
- **Why?**
  Attention is more than gaze. Lost in thought.
- **Why camera?**
  Cheap, data, deep learning.

Cognitive Load Estimation



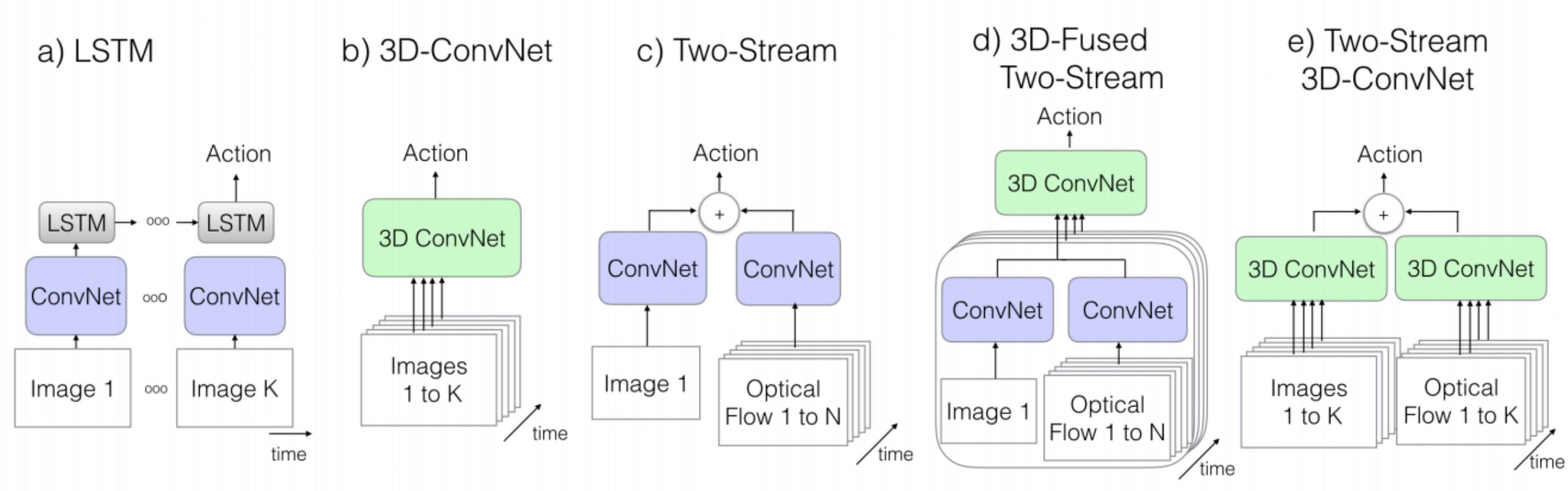


- 6 seconds, 15 fps, 90 images
- Two approaches: HMM and 3D-CNN

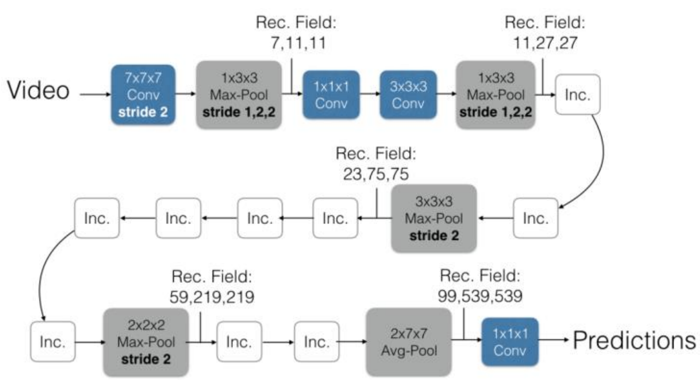
- **HMM:** Hidden Markov Model
  - **Input:** Sequence of pupil positions (normalized by intraocular segment)

- **3D-CNN:** Three Dimensional Convolutional Neural Network
  - **Input:** Sequence of raw images of eye region

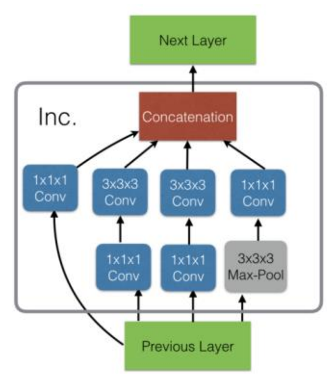
Two-Stream 3D Convolutional Neural Networks



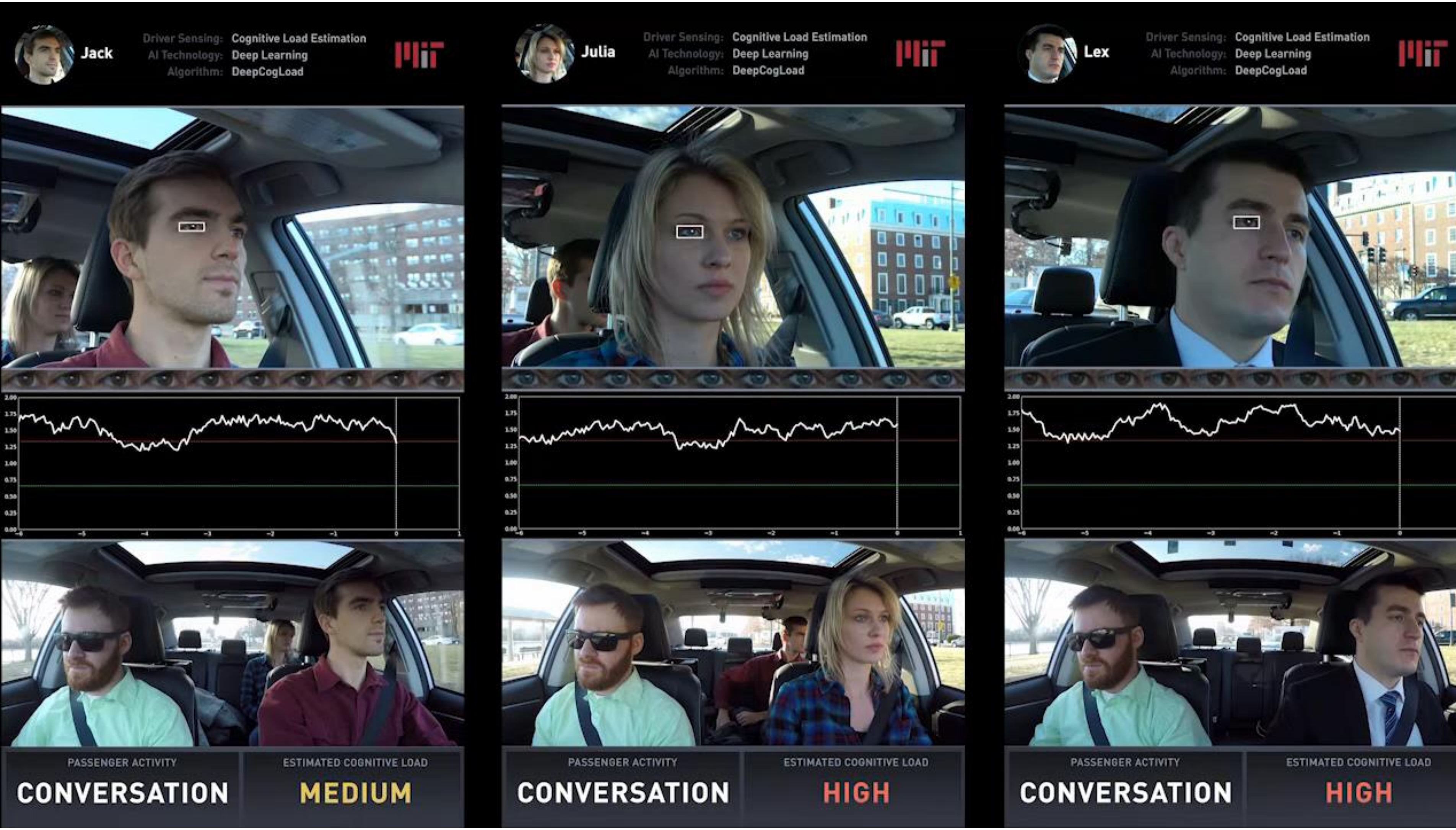
Inception Architecture:



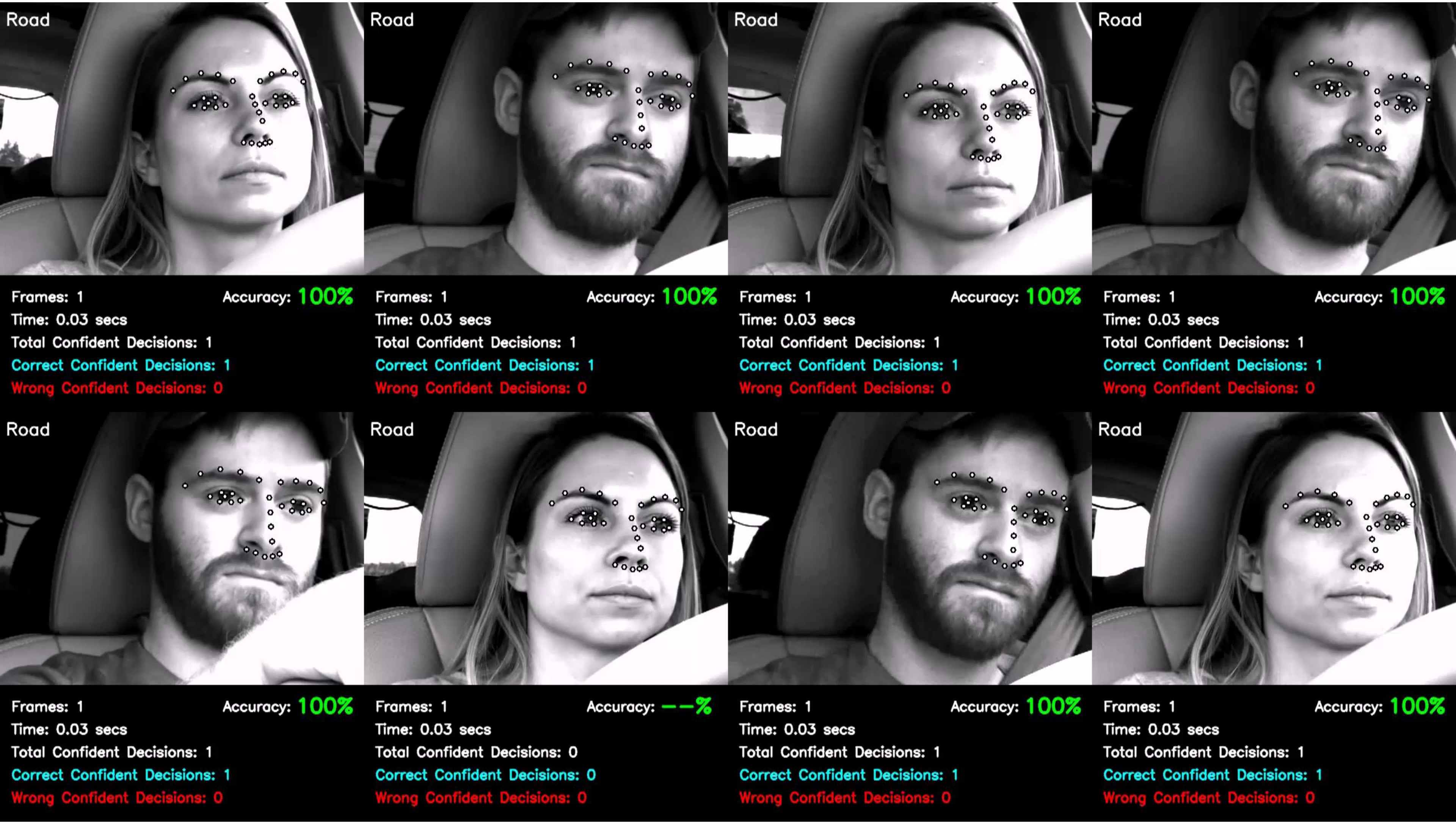
Inception Module:



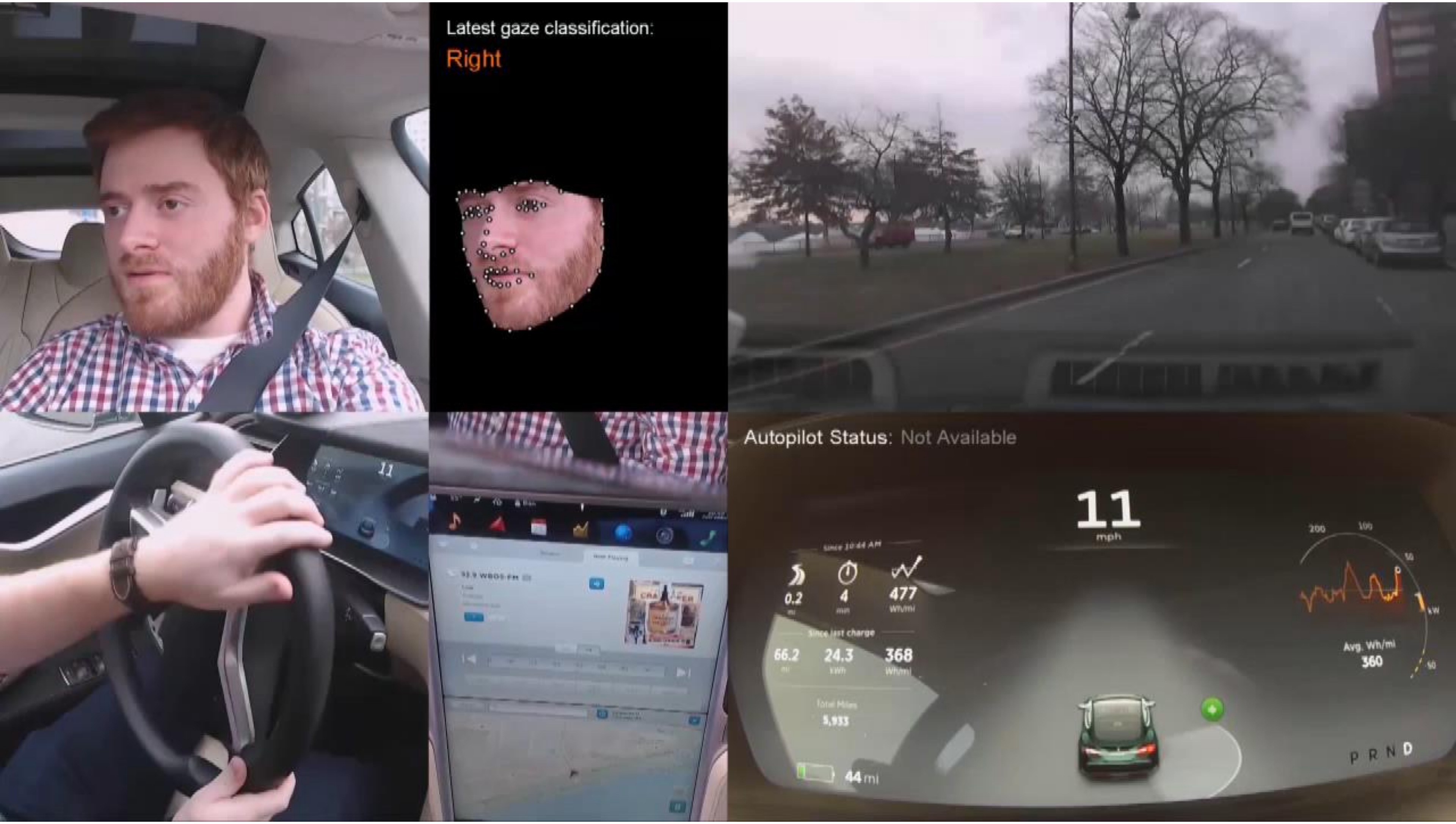
Real-Time Cognitive Load Estimation



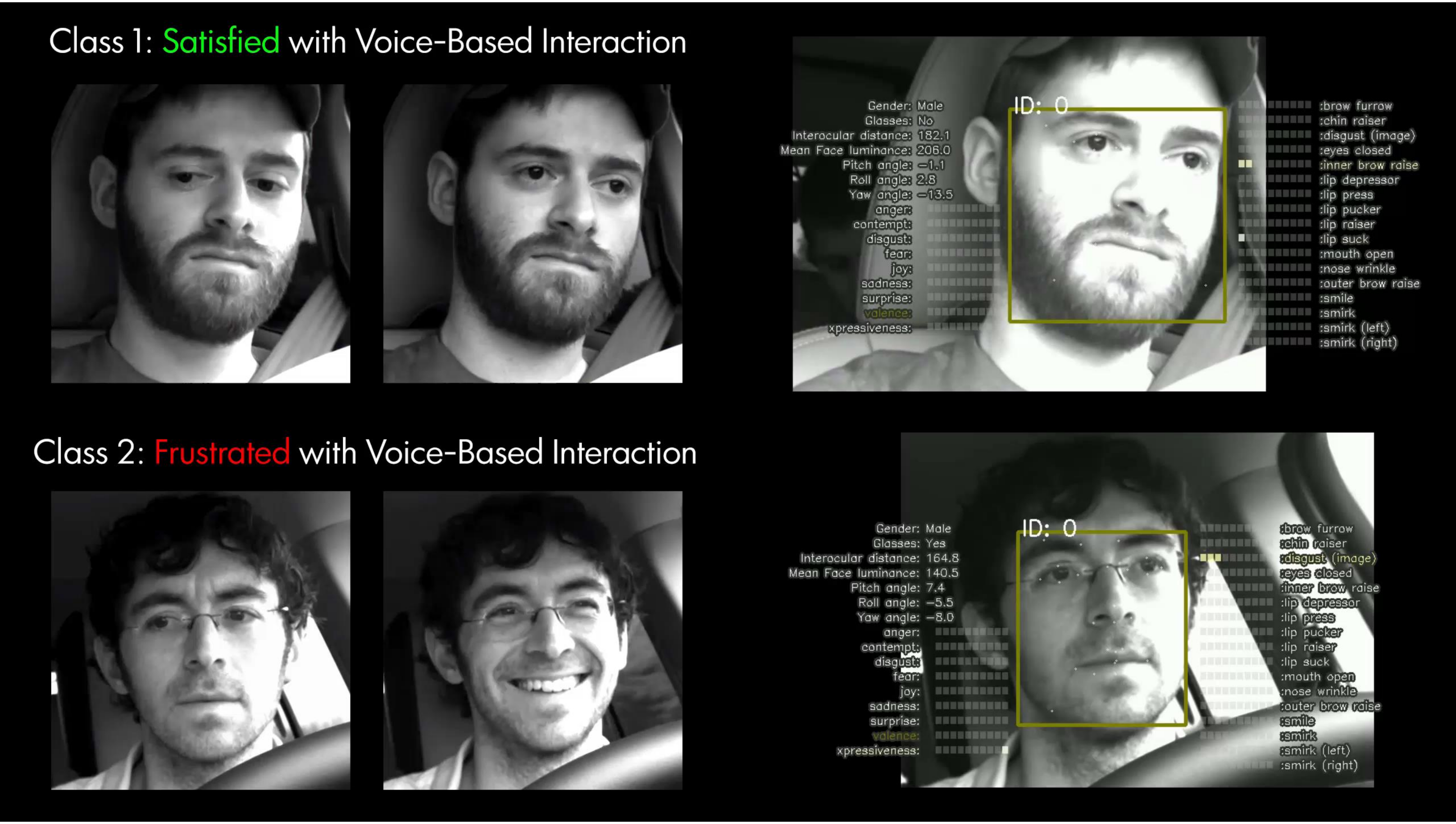
Glance Classification vs Gaze Estimation

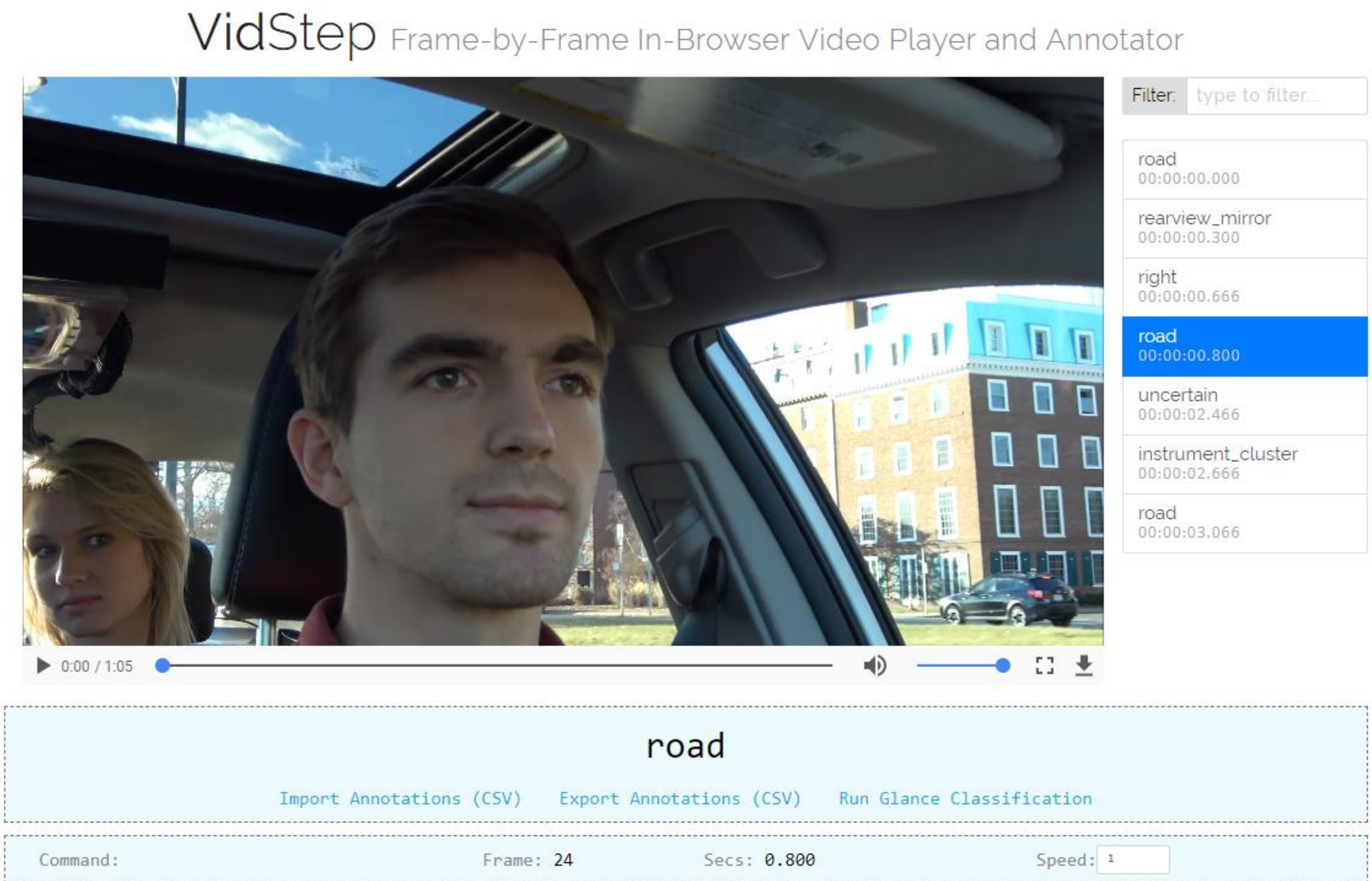


Glance Classification



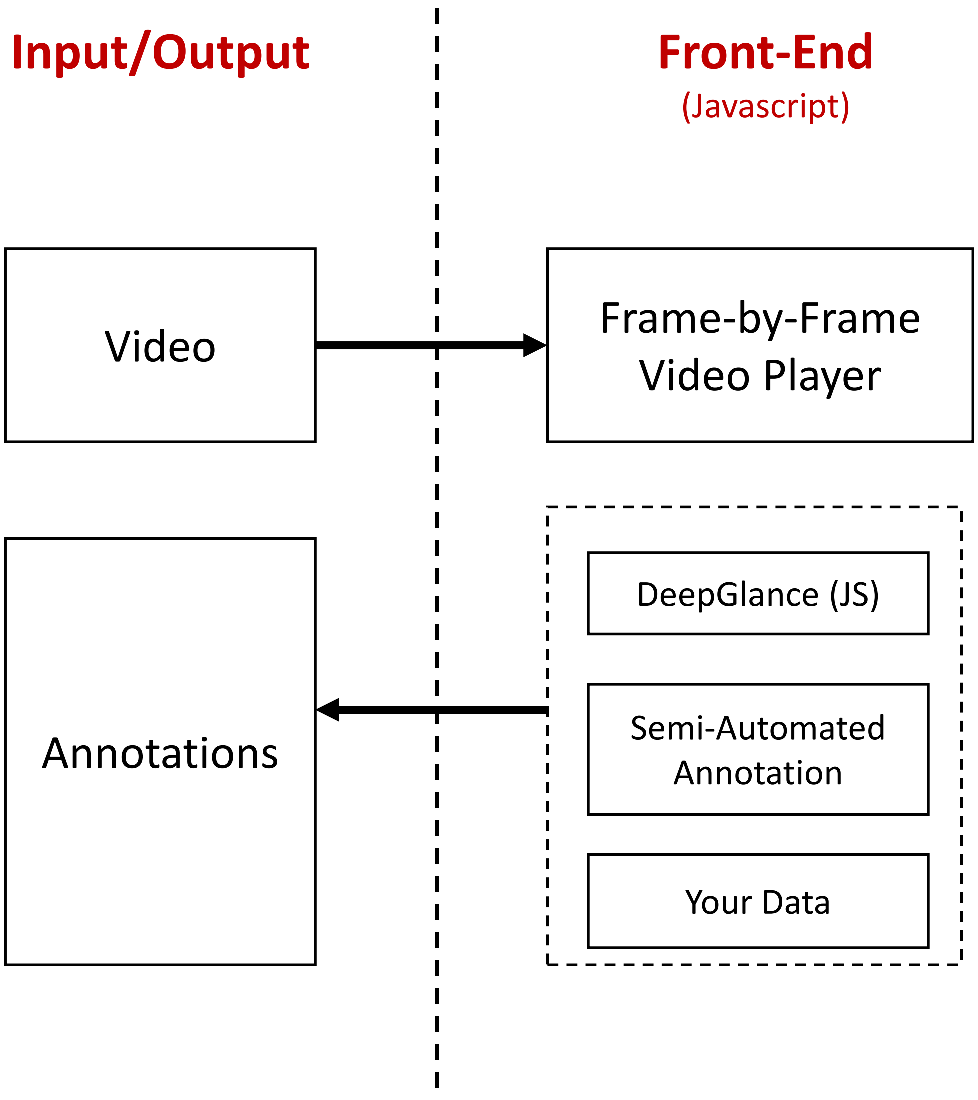
Application-Specific Emotion Recognition: Driver **Frustration**





https://vidstep.com

(Semi-Automated) Glance Annotator

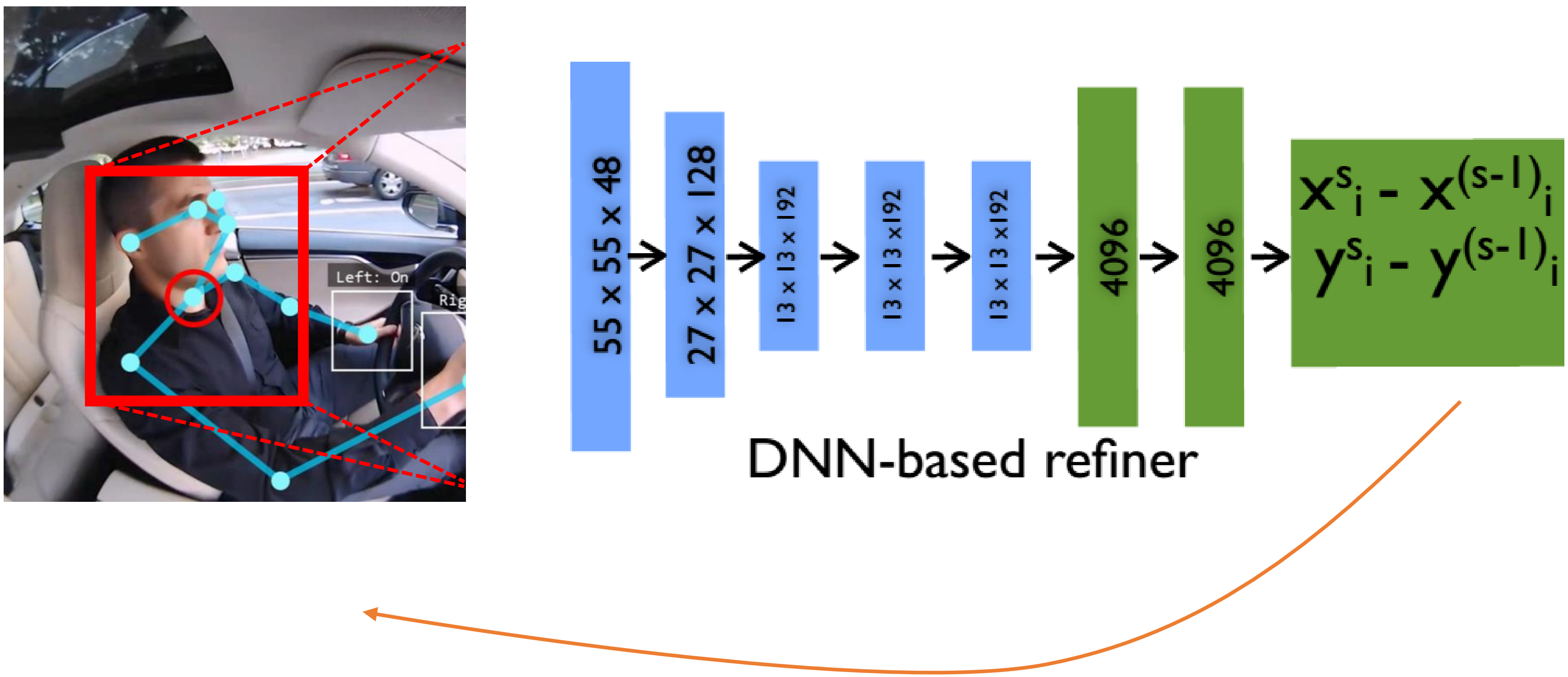


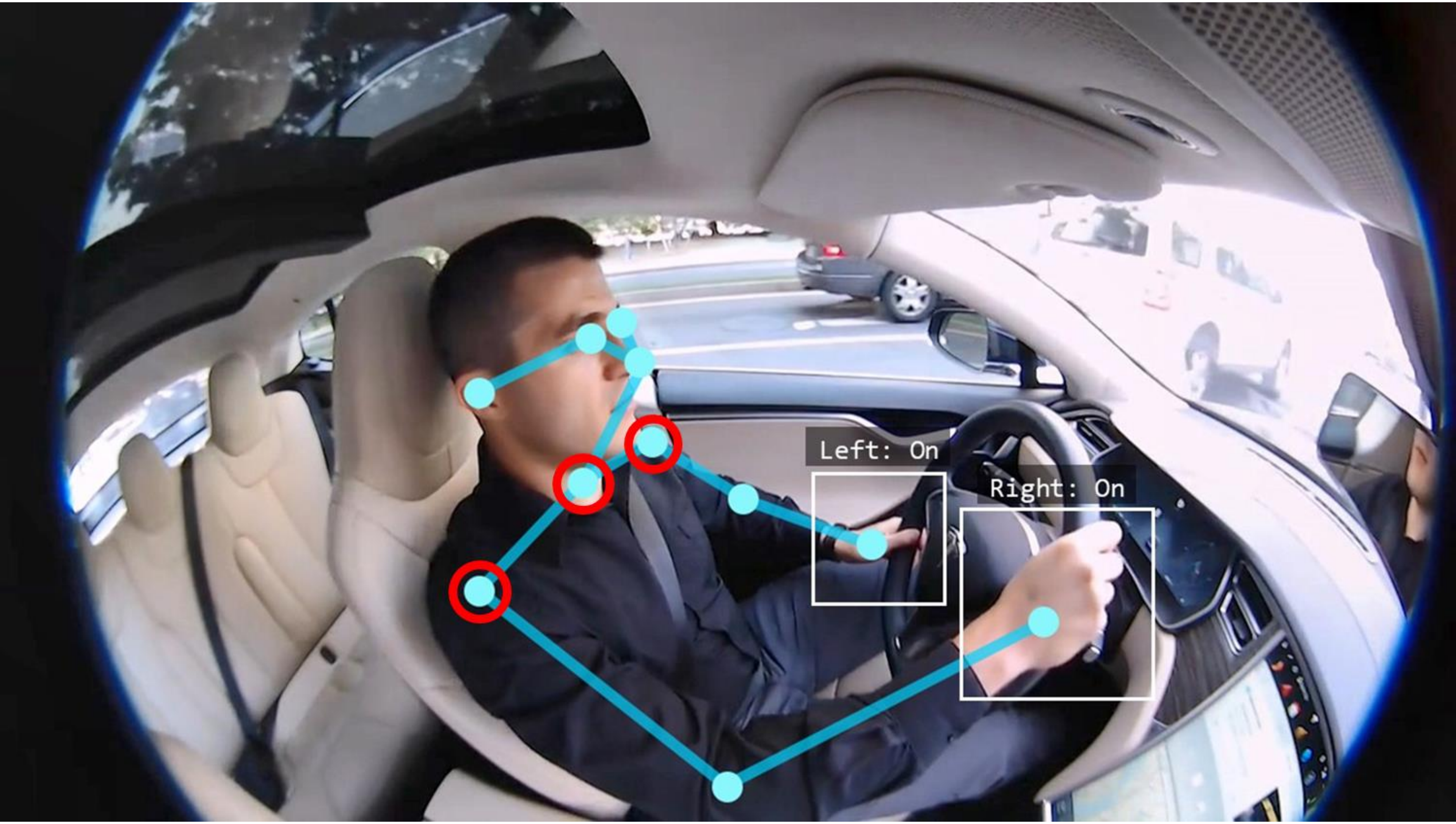
TensorFlow.js

A WebGL accelerated, browser based JavaScript library for training and deploying ML models.

- **TensorFlow.js**
- Keras.js
- WebDNN
- deeplearn.js
- convnet.js

Body Pose Estimation: Cascade of Pose Regressors

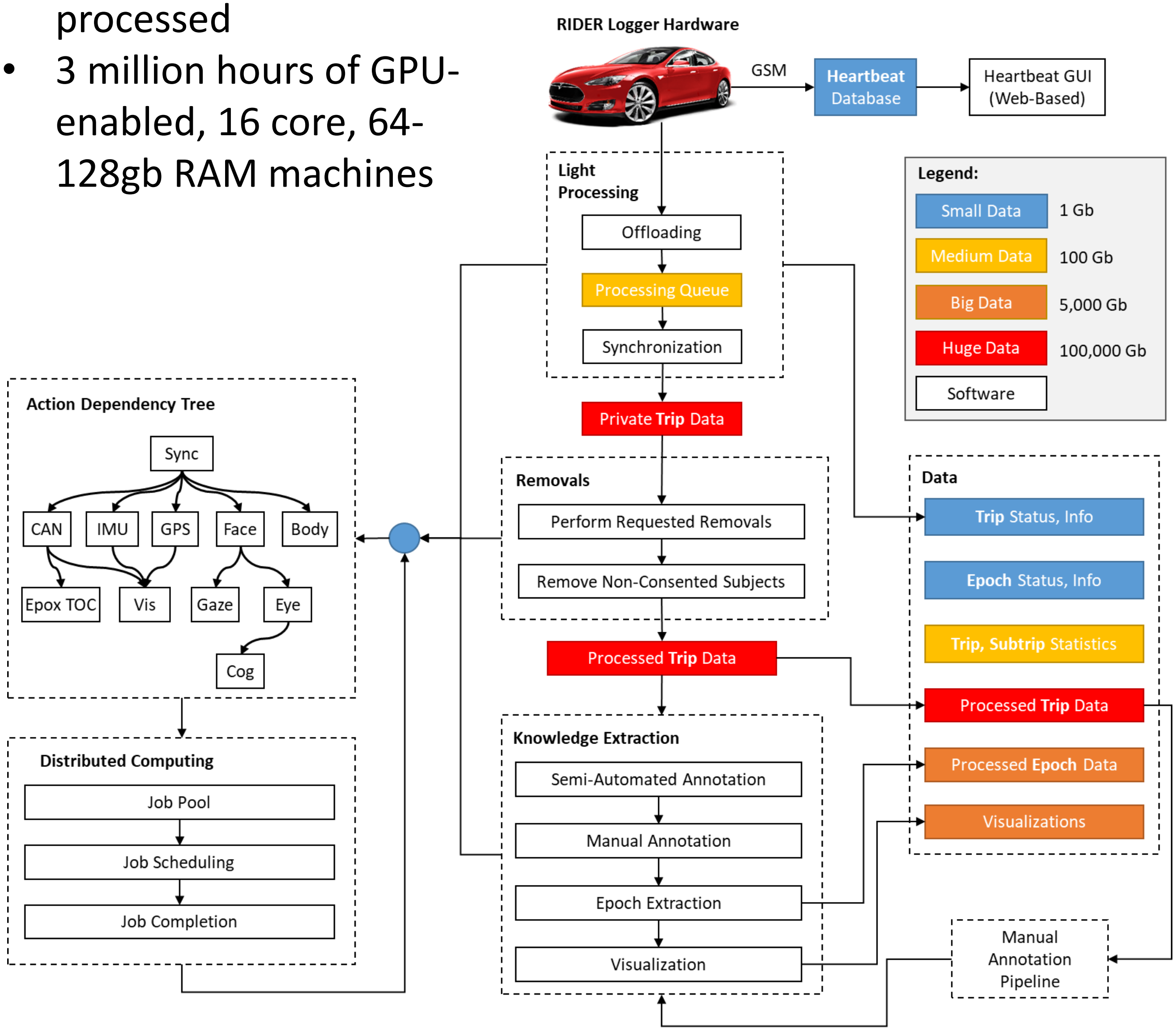


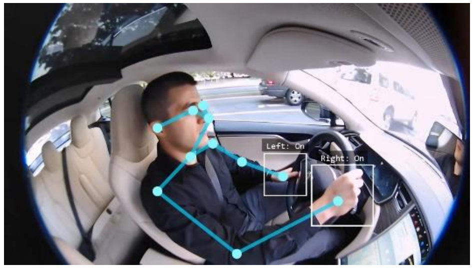

Left: On
Right: On

Costs of deep learning:

- 300 Petabytes of data processed
- 3 million hours of GPU-enabled, 16 core, 64-128gb RAM machines

MIT-AVT Data Pipeline





https://selfdrivingcars.mit.edu

Lecture 1

Deep Learning

[Slides] - [Lecture Video]

Lecture 2

Self-Driving Cars

[Slides] - [Lecture Video]

Lecture 3

Deep Reinforcement Learning

[Slides] - [Lecture Video]

Lecture 4

Computer Vision

[Slides] - [Lecture Video]

Lecture 5

Deep Learning for Human Sensing

[Slides] - [Lecture Video]

Guest Talk

Sacha Arnoud

Director of Engineering, Waymo

[Lecture Video]

Guest Talk

Emilio Frazzoli

CTO, nuTonomy. Previously: Professor, MIT.

[Lecture Video]

Guest Talk

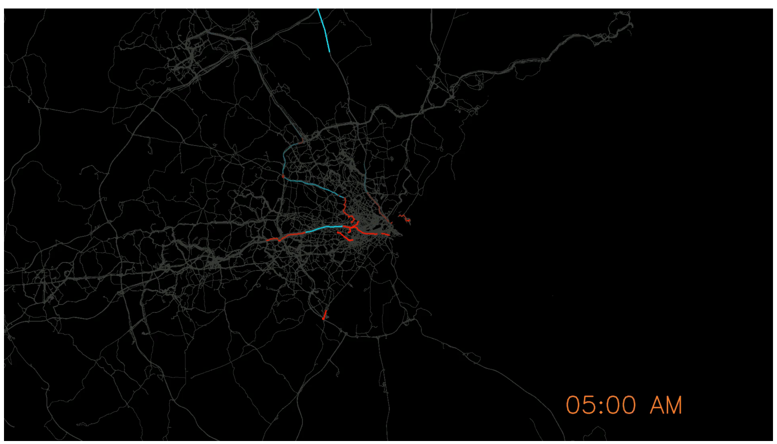
Sterling Anderson

Co-Founder, Aurora. Previously: Director, Tesla Autopilot.

[Lecture Video]

https://hcai.mit.edu

Thank you



https://lex.mit.edu

@lexfridman

Twitter

LinkedIn

Facebook

Instagram

YouTube