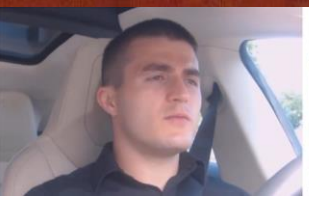
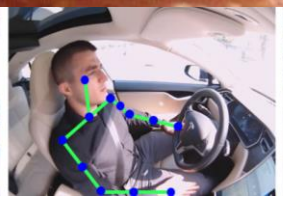




6.S094: Deep Learning for Self-Driving Cars

Deep Learning for Human-Centered Semi-Autonomous Vehicles

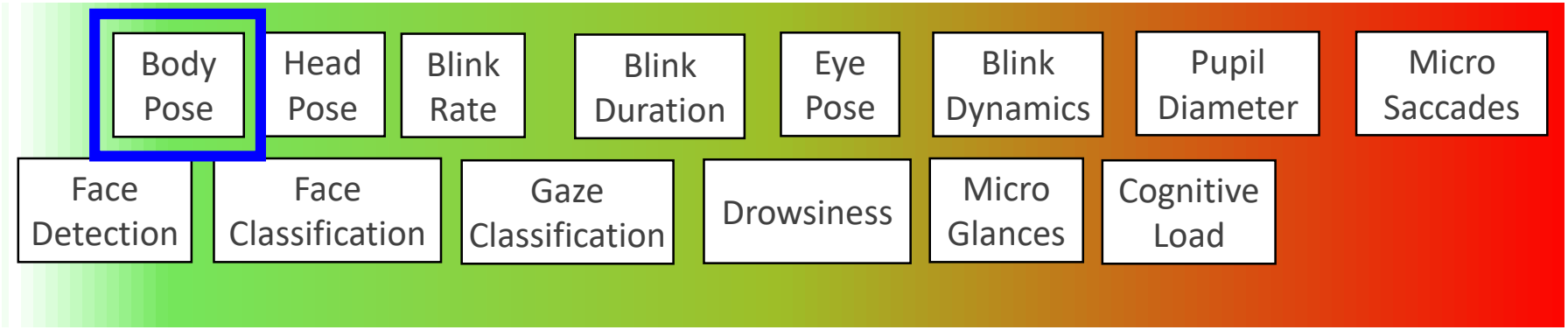
cars.mit.edu

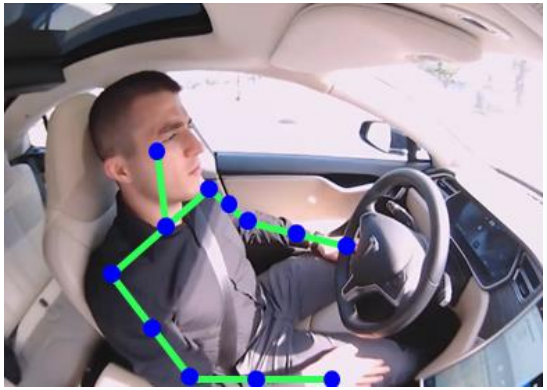
Body Pose | Head Pose | Blink Rate | Blink Duration | Eye Pose | Blink Dynamics | Pupil Diameter | Micro Saccades

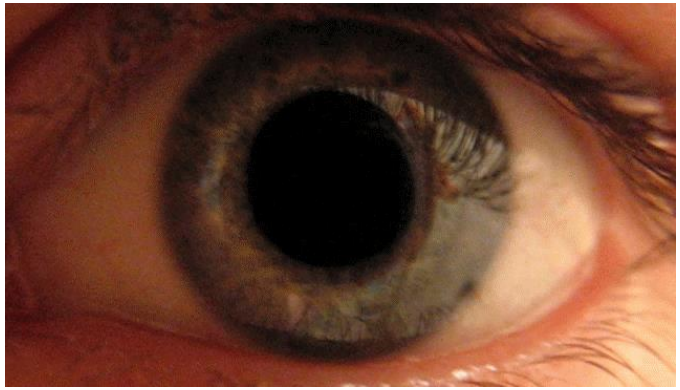
Face Detection | Face Classification | Gaze Classification | Drowsiness | Micro Glances | Cognitive Load

Drive State Detection:
A Computer Vision Perspective

Increasing level of detection resolution and **difficulty**







Why Body Pose?

Safety Systems: Seatbelt and Airbag Design

- How much time to we spend in the “optimal” crash test dummy position?

Physical Distraction and Incapacitation

“Physical” is a more dramatic manifestation of “mental” inattention

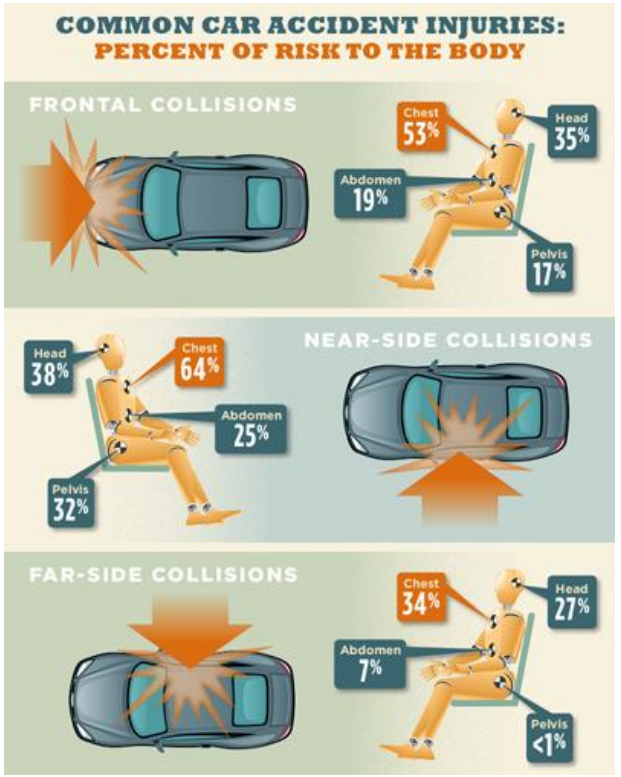
- **Manual:** How much time to be back in physical position to avoid crash?
- **Semi-automated control:** How much time to be back in physical position

Extended Gaze Classification

- How often and where does gaze classification break down in naturalistic driving when body is out of position for camera?

Effect of Seatbelts

- Decreased traumatic brain injury (10.4% to 4.1%)
- Decreased head, face, and neck injury (29.3% to 16.6%)
- Increased spine injury (17.9% to 35.5%)
  - But severity decreased (e.g., fracture from 22% to 4%)
- Seat belts saved 12,802 lives in 2014.



- Han, Guang-Ming, Ashley Newmyer, and Ming Qu. "Seat belt use to save face: impact on drivers' body region and nature of injury in motor vehicle crashes." Traffic injury prevention 16.6 (2015): 605-610.
- National Highway Traffic Safety Administration. Lives saved in 2014 by restraint use and minimum-drinking-age laws. Washington, DC: US Department of Transportation, National Highway Traffic Safety Administration; 2015. Publication no. DOT-HS-812-218. Available at http://www-nrd.nhtsa.dot.gov/Pubs/812218.pdf

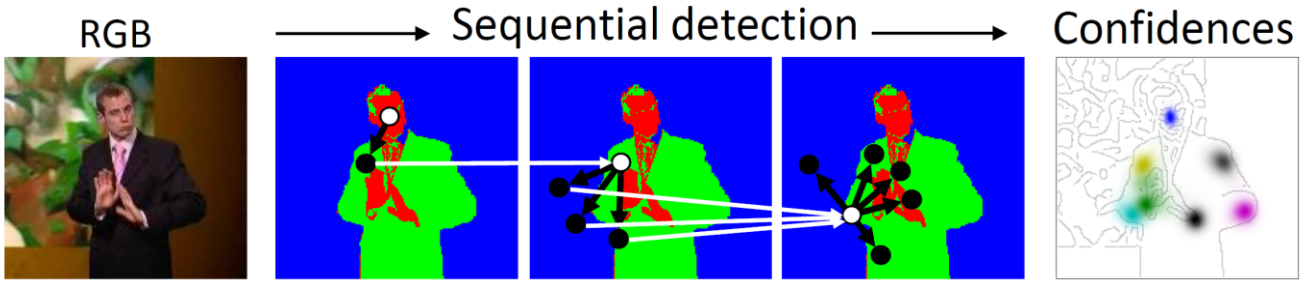
Crash Test Dummy Design: Hybrid III

Biofidelity: "It's important to have a dummy that acts like a human so that the restraints that you develop have a benefit for the human." - Jesse Buehler (Toyota)

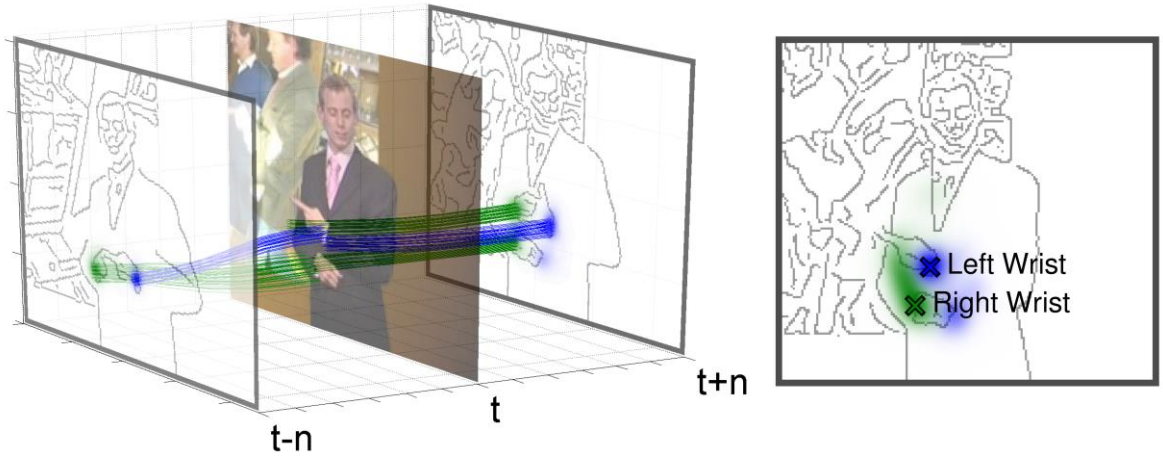
This helps address variation in body type, but does not help address temporal variation in body pose during naturalistic driving.

Sequential Detection Approach

Sequential Upper Body Pose Estimation:

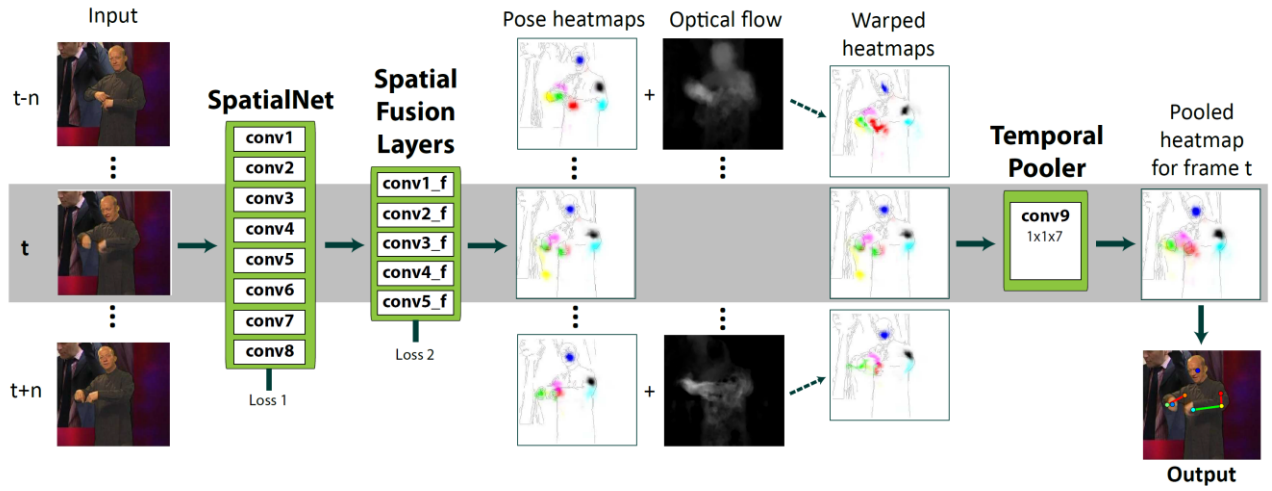


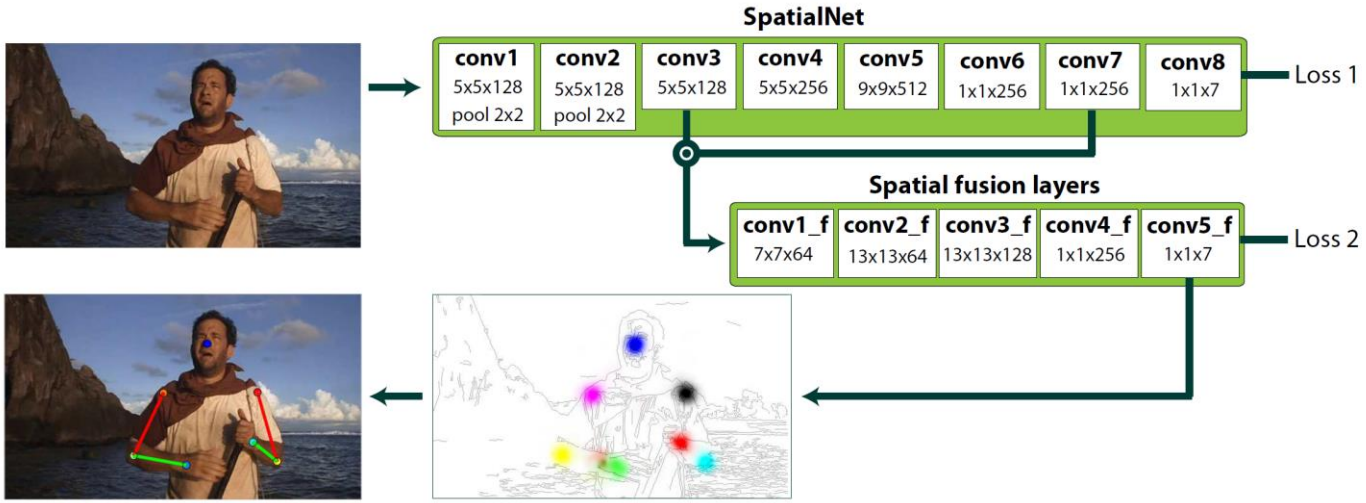
Temporal Fusion of Localized Confidences:



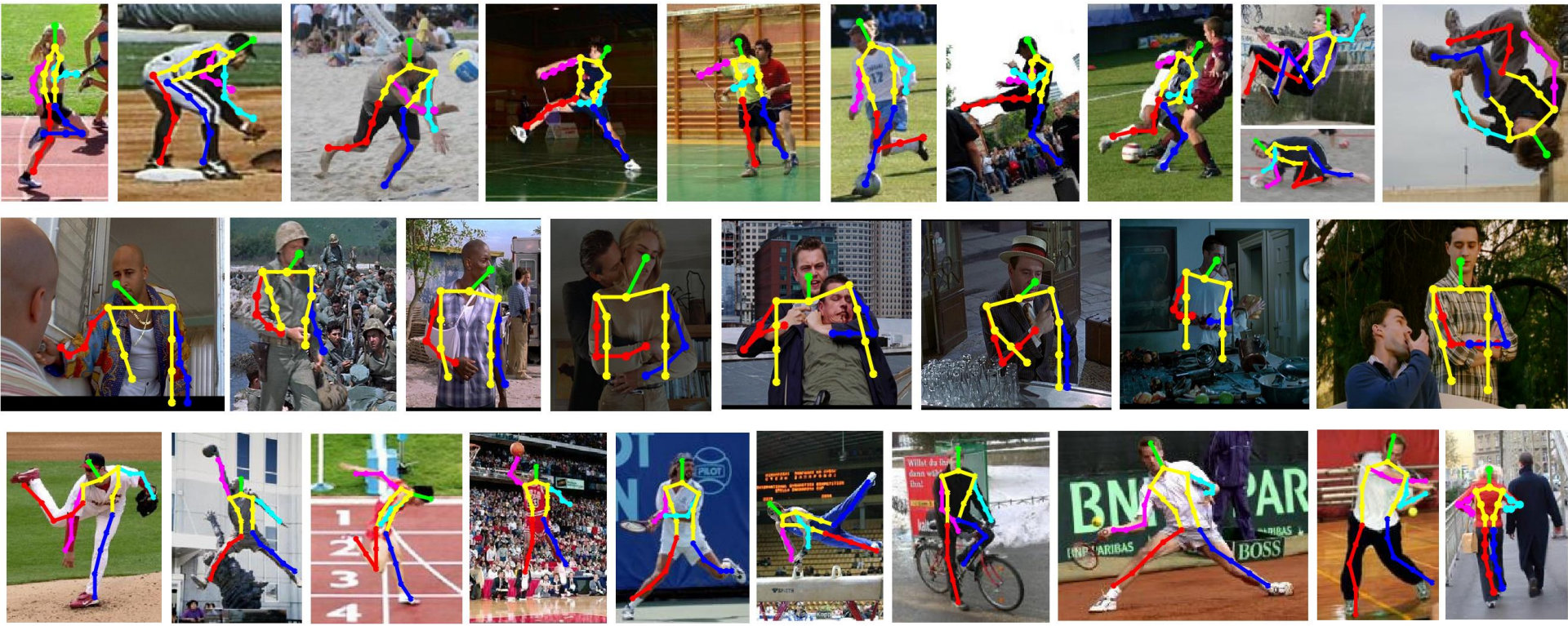
Charles, James, et al. "Upper body pose estimation with temporal sequential forests." *Proceedings of the British Machine Vision Conference 2014*. BMVA Press, 2014.

Temporal Convolutional Neural Networks





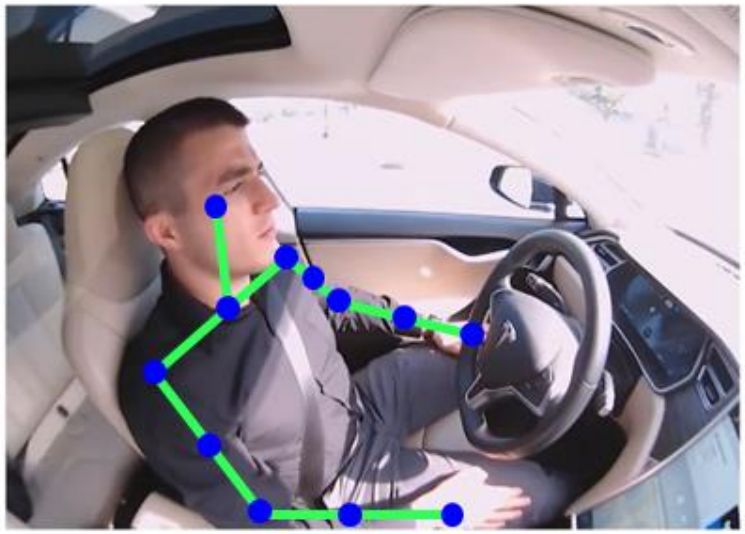
Pfister, Tomas, James Charles, and Andrew Zisserman. "Flowing convnets for human pose estimation in videos." *Proceedings of the IEEE International Conference on Computer Vision*. 2015.

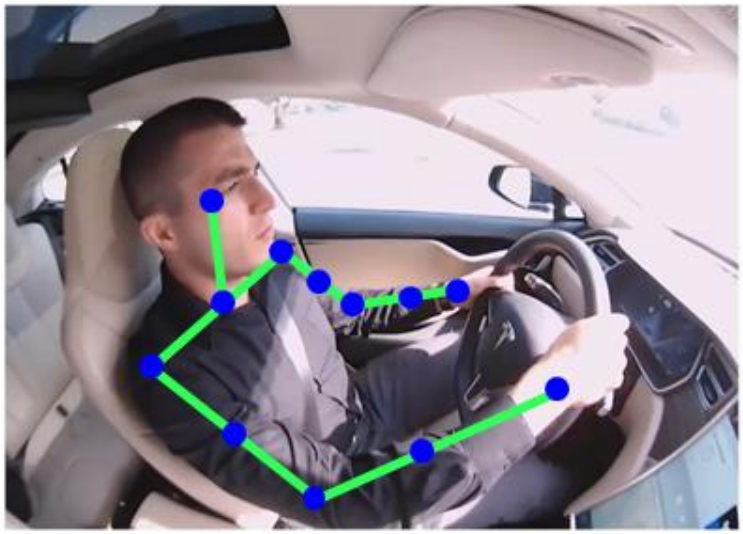


Datasets

1. LSP dataset *(Leeds Sports Dataset)*
2. FLIC dataset *(Frames Labeled in Cinema)*
3. PARSE dataset
4. Our own?

Open Question: What Do We Need to Detect?

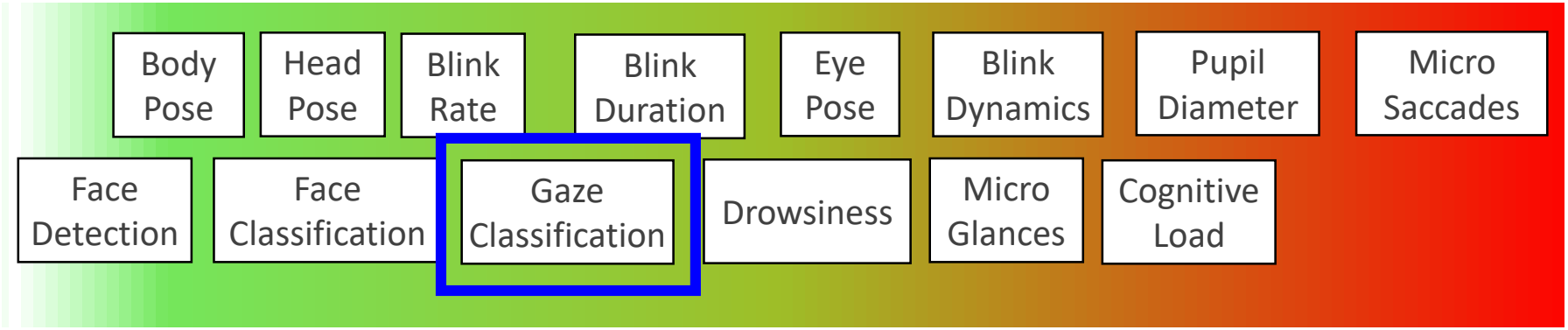


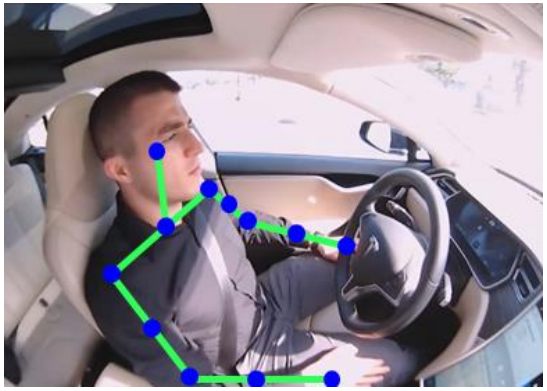


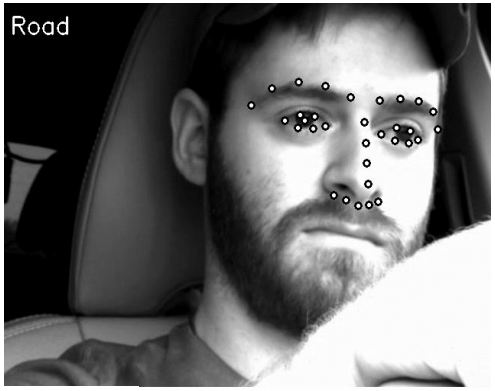
- **Detection task 1:** Hands on wheel
- **Detection task 2:** Position of head relative to head rest
- **Detection task 3:** Full upper body pose

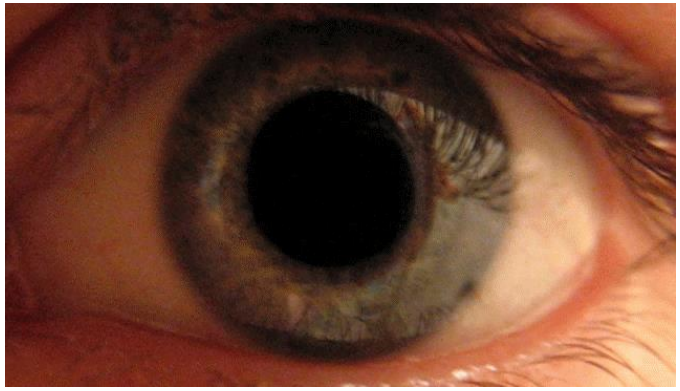
Drive State Detection:
A Computer Vision Perspective

Increasing level of detection resolution and **difficulty**

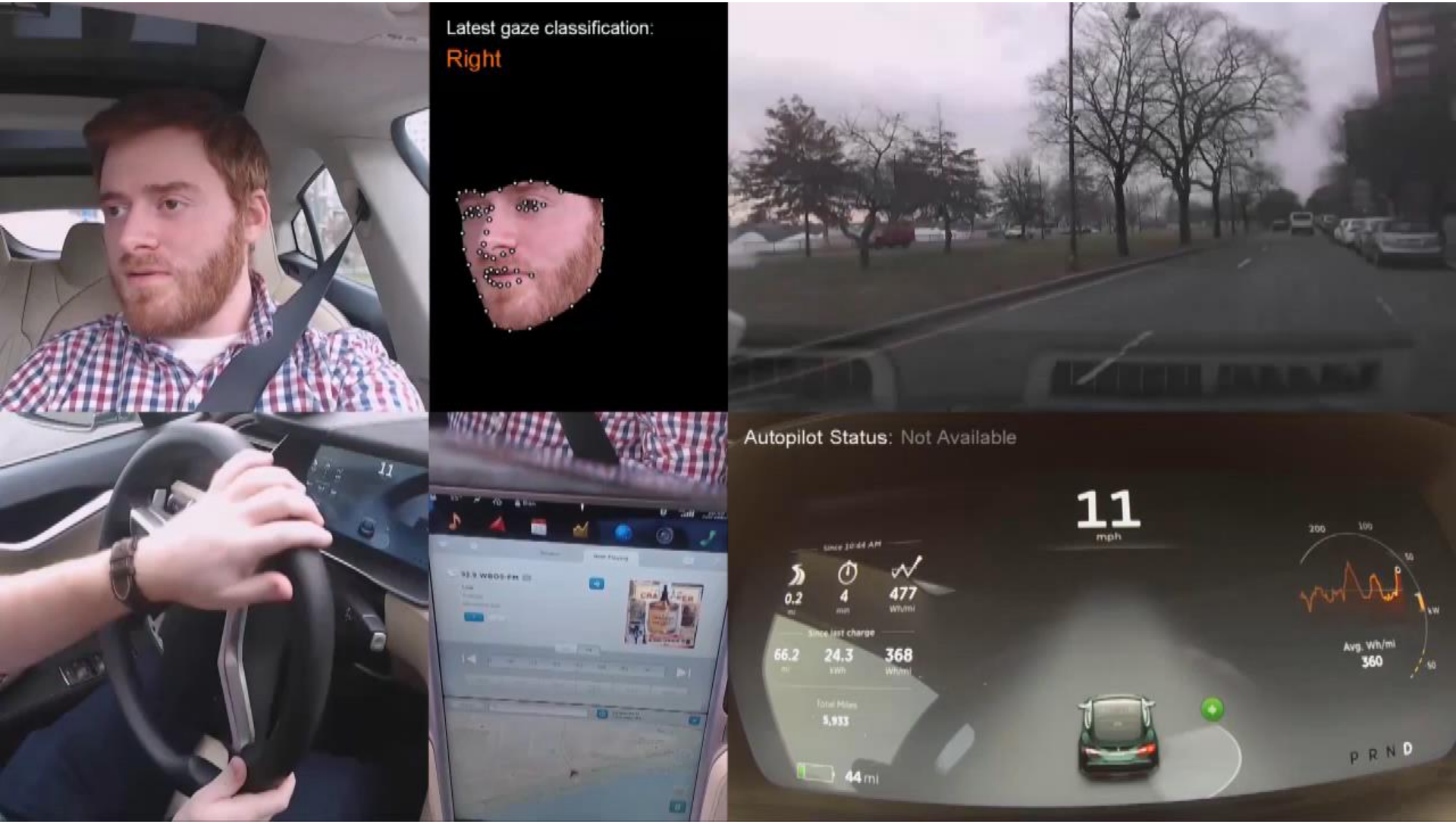








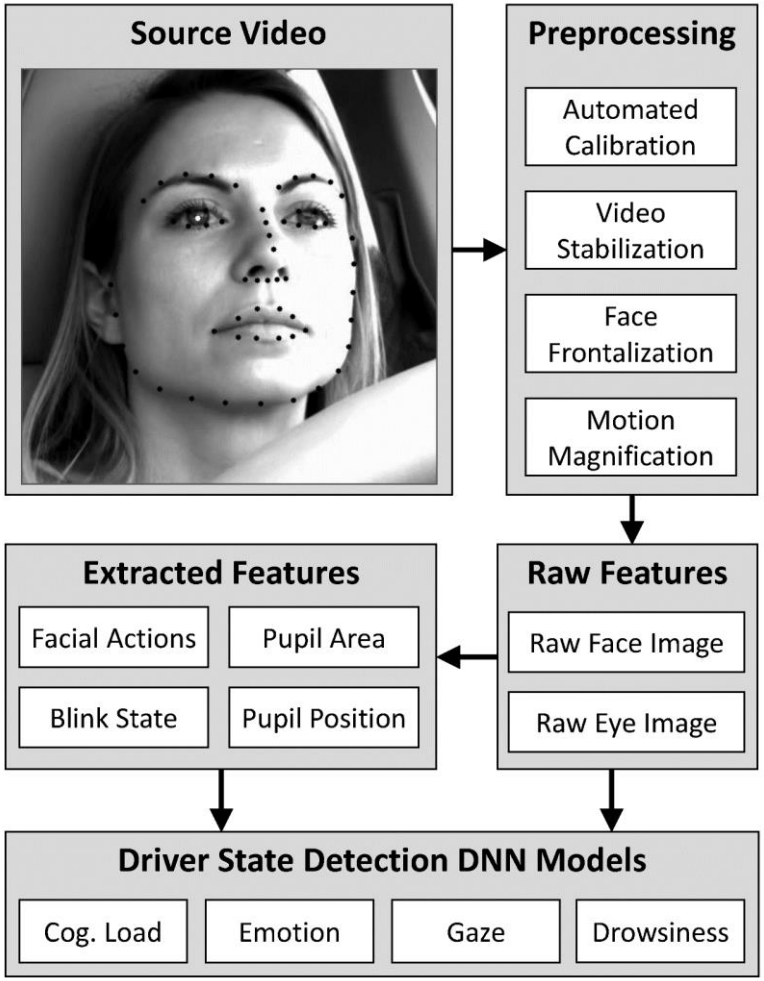
Gaze Classification vs Gaze Estimation



Drive State Detection

- **Challenge:** real-world data is “messy”, have to deal with:
  - Vibration
  - Lighting variation
  - Body, head, eye movement

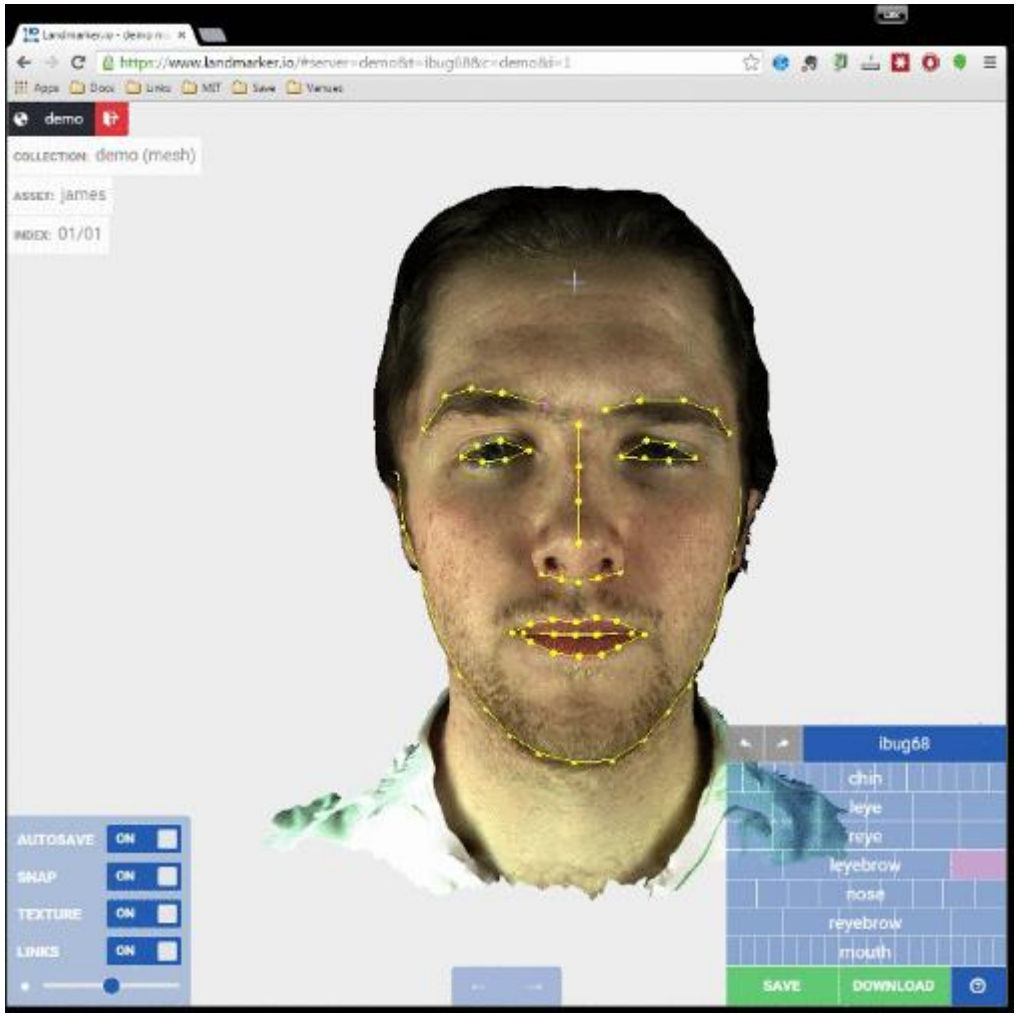
- **Solution:**
  - Automated calibration
  - Video stabilization (multi-resolutional)
  - Face part frontalization
  - Phase-based motion magnification
  - Use deep neural networks (DNN)
    - No feature engineering
    - Use raw data



Gaze Classification Pipeline

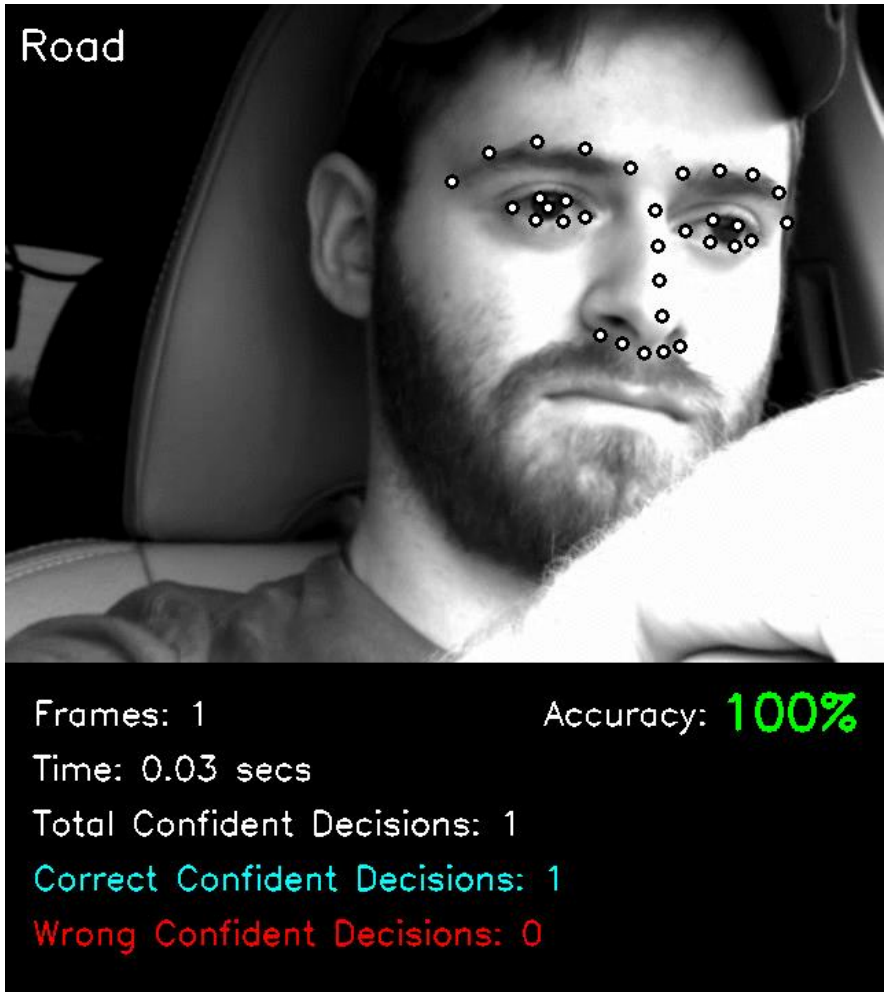
1. Face detection *(the only easy step)*
2. Face alignment *(active appearance models or deep nets)*
3. Eye/pupil detection *(are the eyes visible?)*
4. Head (and eye) pose estimation *(+ normalization)*
5. Classification *(supervised learning = improves from data)*
6. Decision pruning *(how confident is the prediction)*

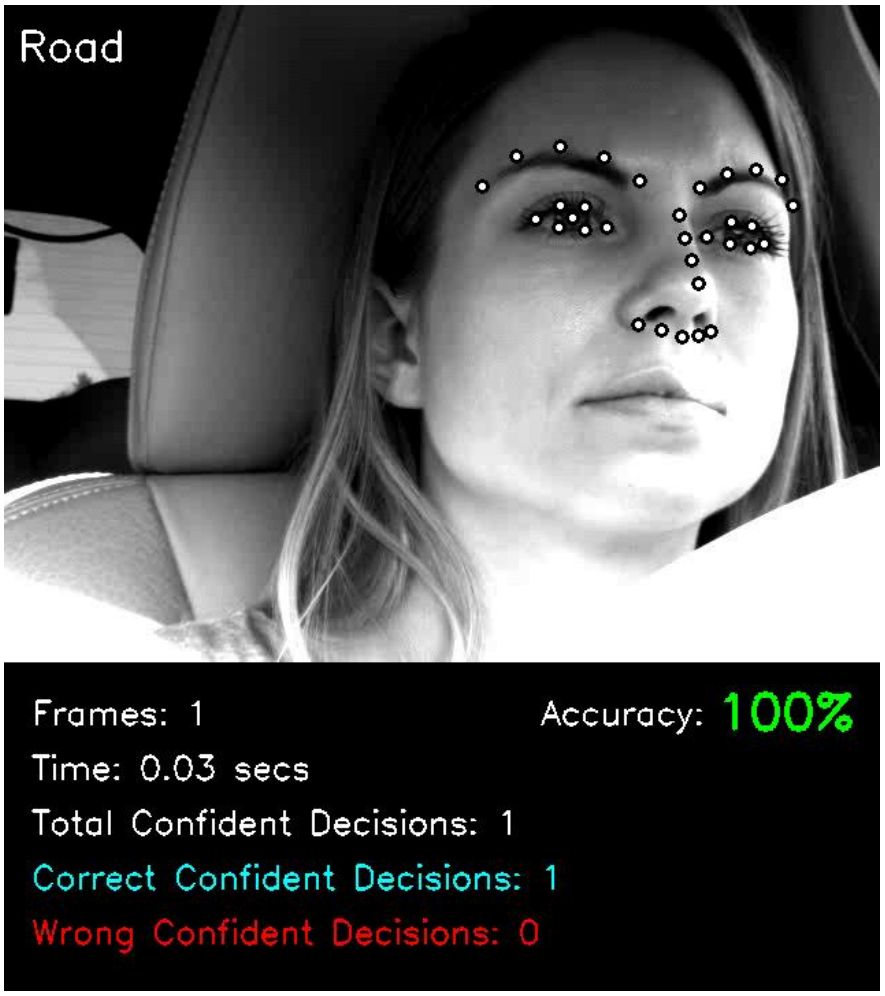
Face Alignment



- Landmarker.io
  - Imperial College London
- Face in the Wild Challenge
  - XM2VTS
  - FRGC Ver.2
  - LFPW
  - HELEN
  - AFW
  - IBUG
- New Datasets
  - MPIIGaze
  - Columbia Gaze
  - 300VW
  - MIT Driver Gaze
- Goal: **1,000,000 images**

Driver Gaze Classification





A General Framework for Semi-Automated Object State Annotation

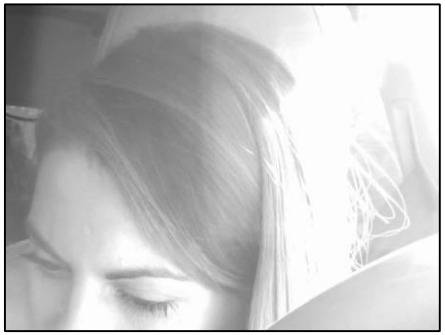
"Semi-automated":

Ask a human for help with annotation when the machine is not confident.

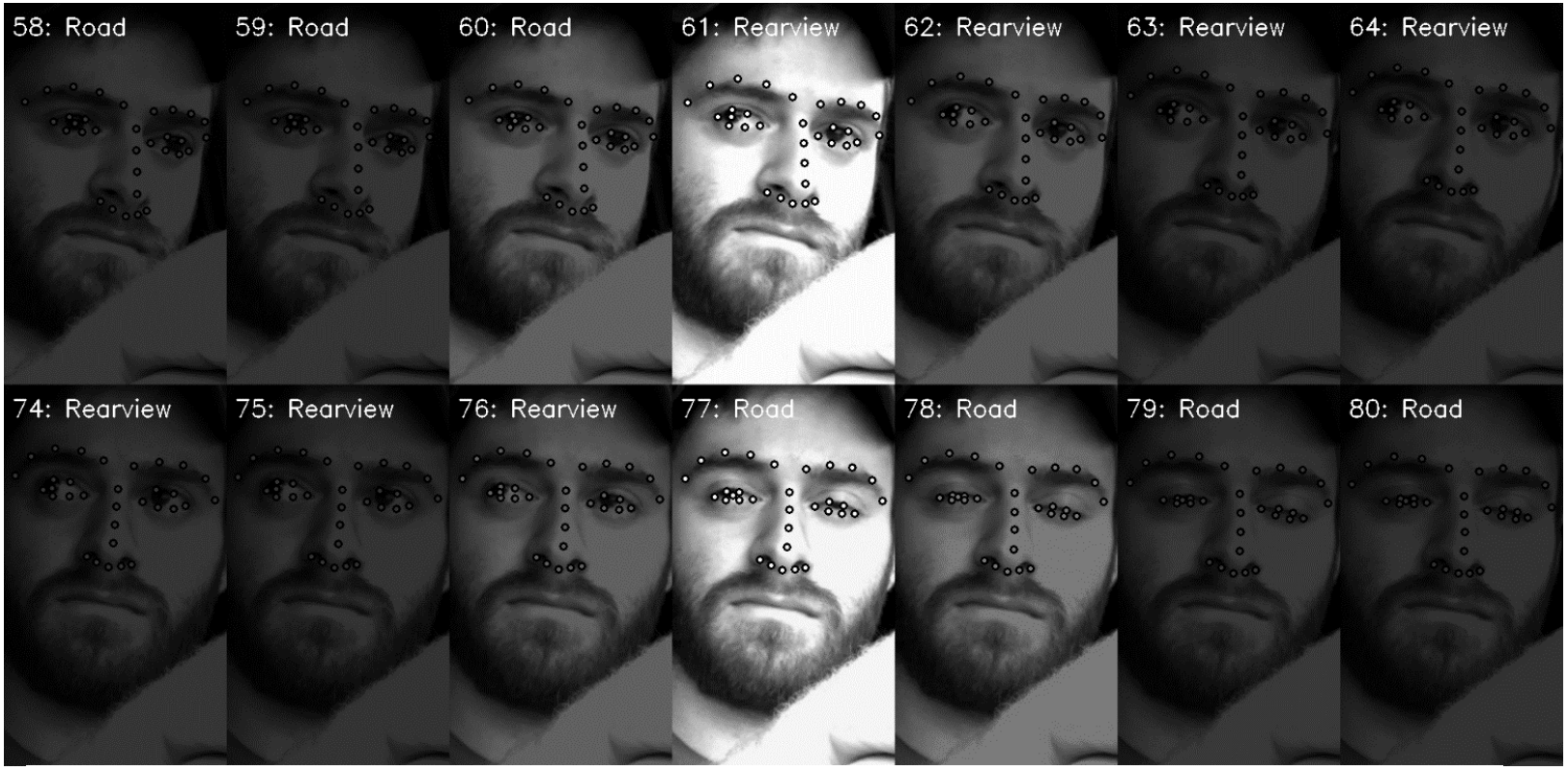
Partial light occlusion

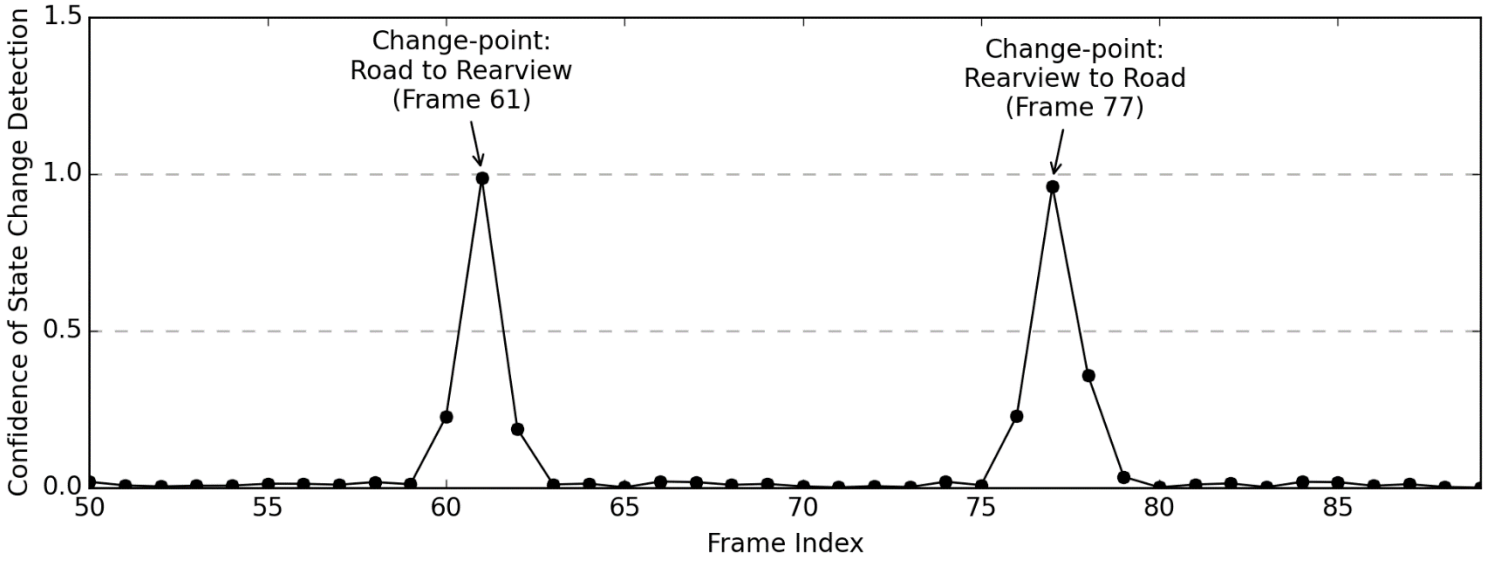
Full light occlusion

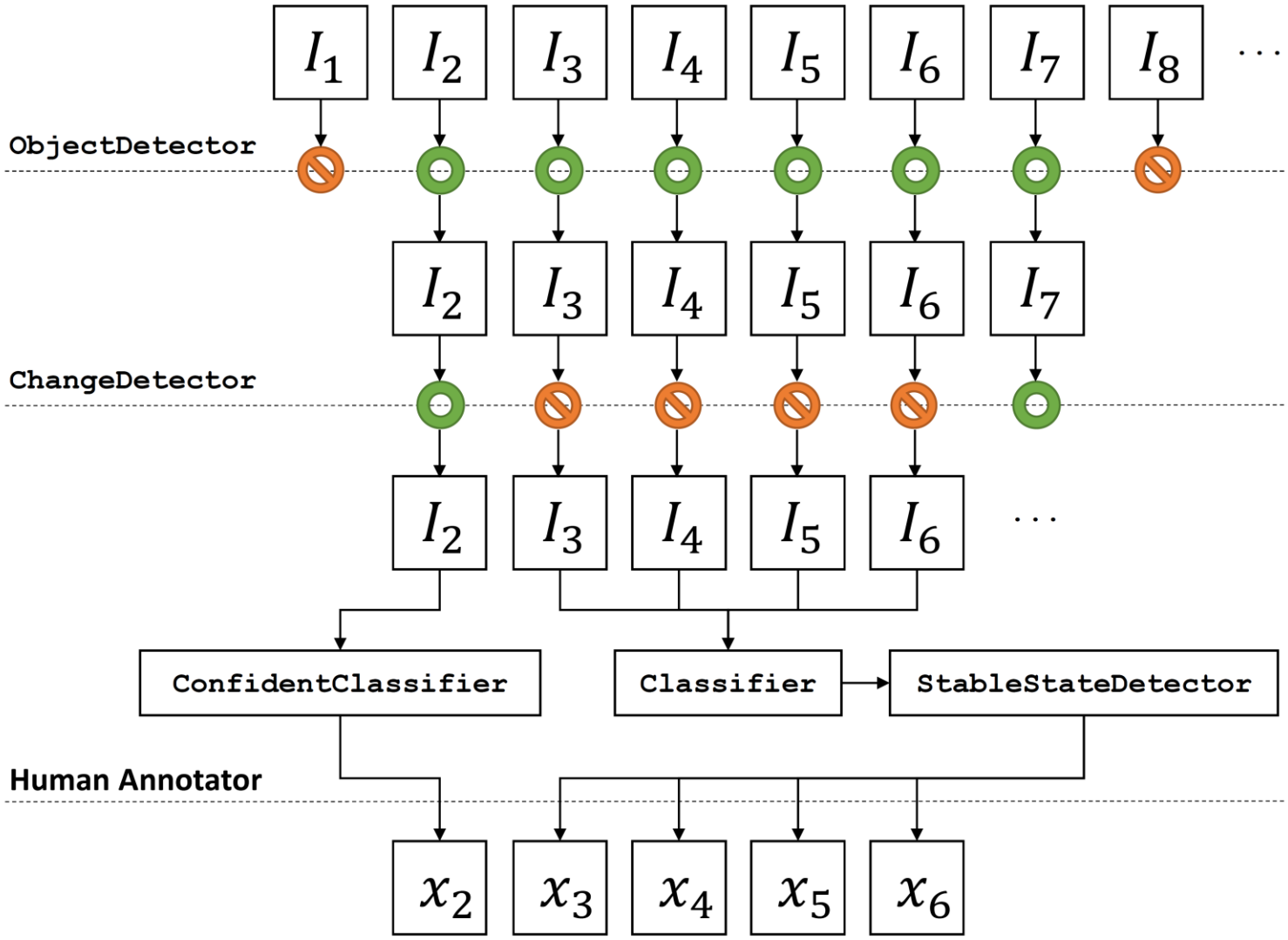
Move out of frame



Hand occlusion

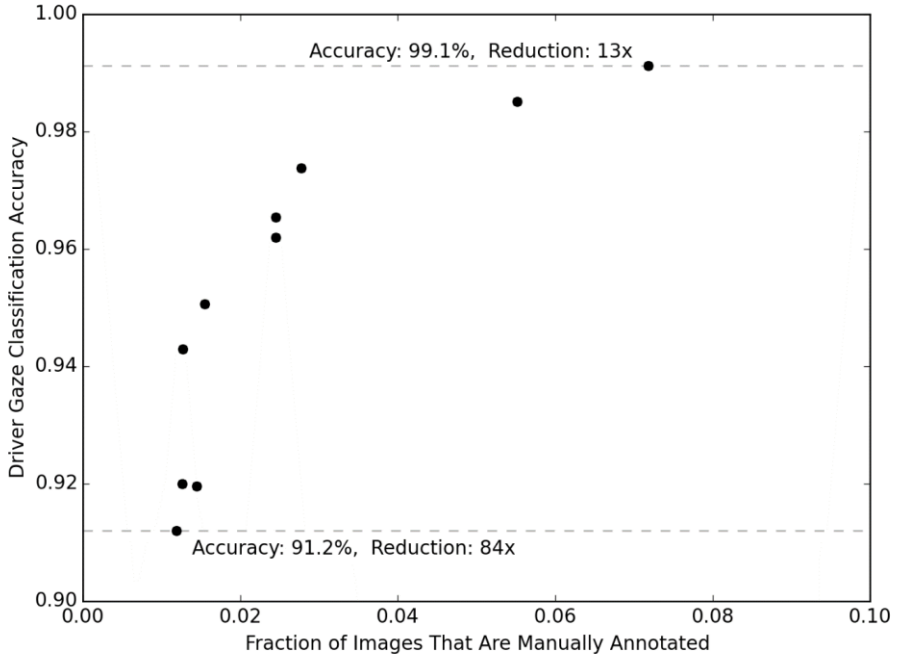

58: Road
59: Road
60: Road
61: Rearview
62: Rearview
63: Rearview
64: Rearview
74: Rearview
75: Rearview
76: Rearview
77: Road
78: Road
79: Road
80: Road


Change-point:
Road to Rearview
(Frame 61)
Change-point:
Rearview to Road
(Frame 77)
Confidence of State Change Detection
1.5
1.0
0.5
0.0
50
55
60
65
70
75
80
85
Frame Index


I_1 I_2 I_3 I_4 I_5 I_6 I_7 I_8 ...
ObjectDetector
I_2 I_3 I_4 I_5 I_6 I_7
ChangeDetector
I_2 I_3 I_4 I_5 I_6 ...
ConfidentClassifier
Classifier
StableStateDetector
Human Annotator
x_2 x_3 x_4 x_5 x_6

Large-Scale Solution Using Gaze as Example

- Semi-automated annotation
  - **Preliminary results:** 84x reduction in # of frames a human annotates.
  - **Goal results:** 1000x reduction in # of frames a human annotates



- Software **interface design** challenge
  - **In general:** 1000x frame reduction ≠ 1000x time reduction
  - **Our goal:** 1000x frame reduction ≈ 1000x time reduction

Semi-Automated Annotation Work Flow

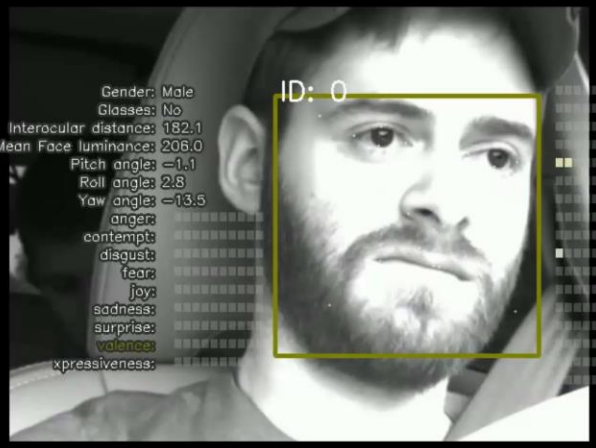
* Human in **red** and machine in **blue**

1. Select and load in video of driver face.
2. Detect face: have we seen this person before?
3. Localize camera: have we seen this angle before?
4. Provide tradeoff between accuracy and percent frames.
5. Select target accuracy: 95%, 99%, or 99.9%
6. Perform gaze classification on full video *(1 hour per 1 hour of video)*
7. Step through and annotate the frames machine did not classify.
8. (Optional) Re-run steps 6 and 7.
9. Enjoy fully annotated video!

Semi-Automated Annotation Result Example

Driver **Frustration**

Class 1: Satisfied with Voice-Based Interaction

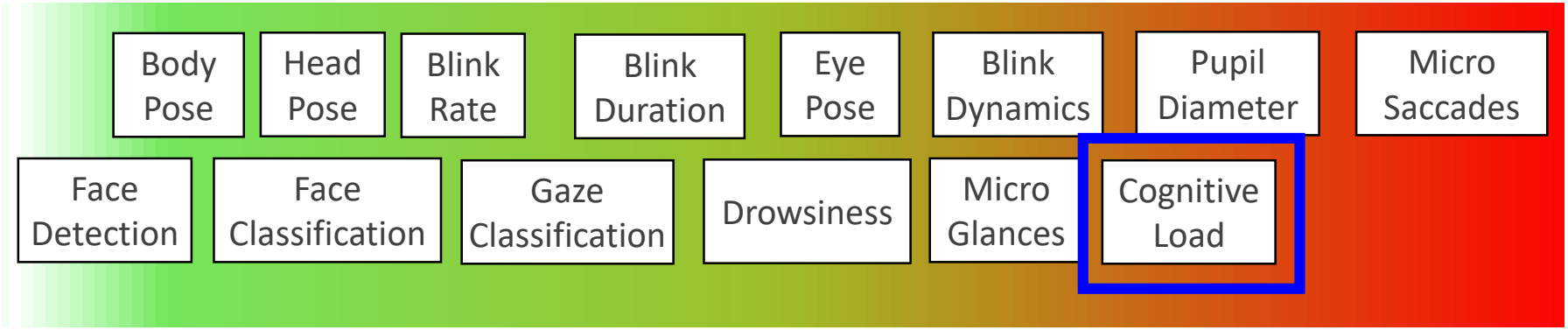


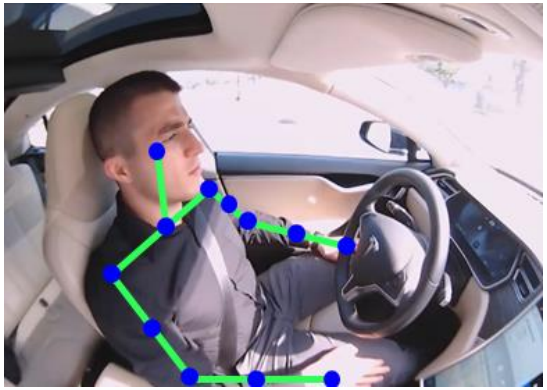
Class 2: Frustrated with Voice-Based Interaction

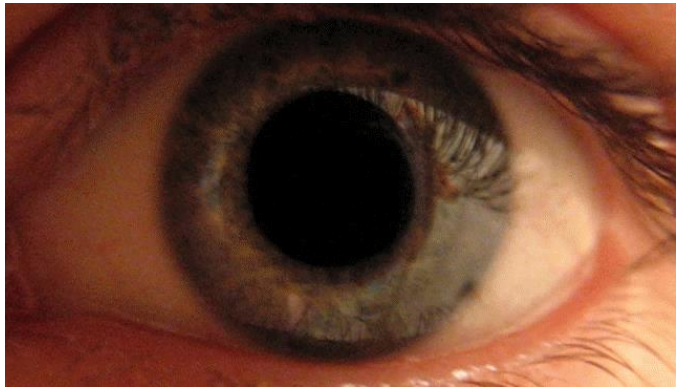
Drive State Detection:

A Computer Vision Perspective

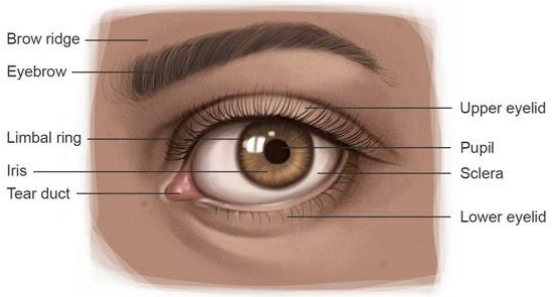
Increasing level of detection resolution and **difficulty**

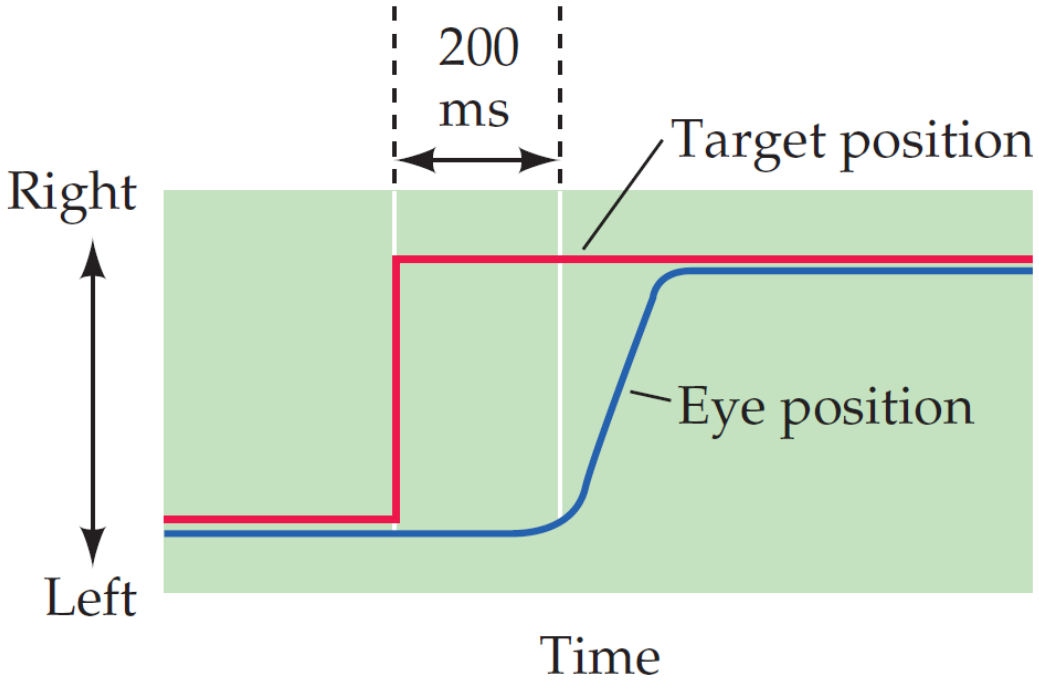






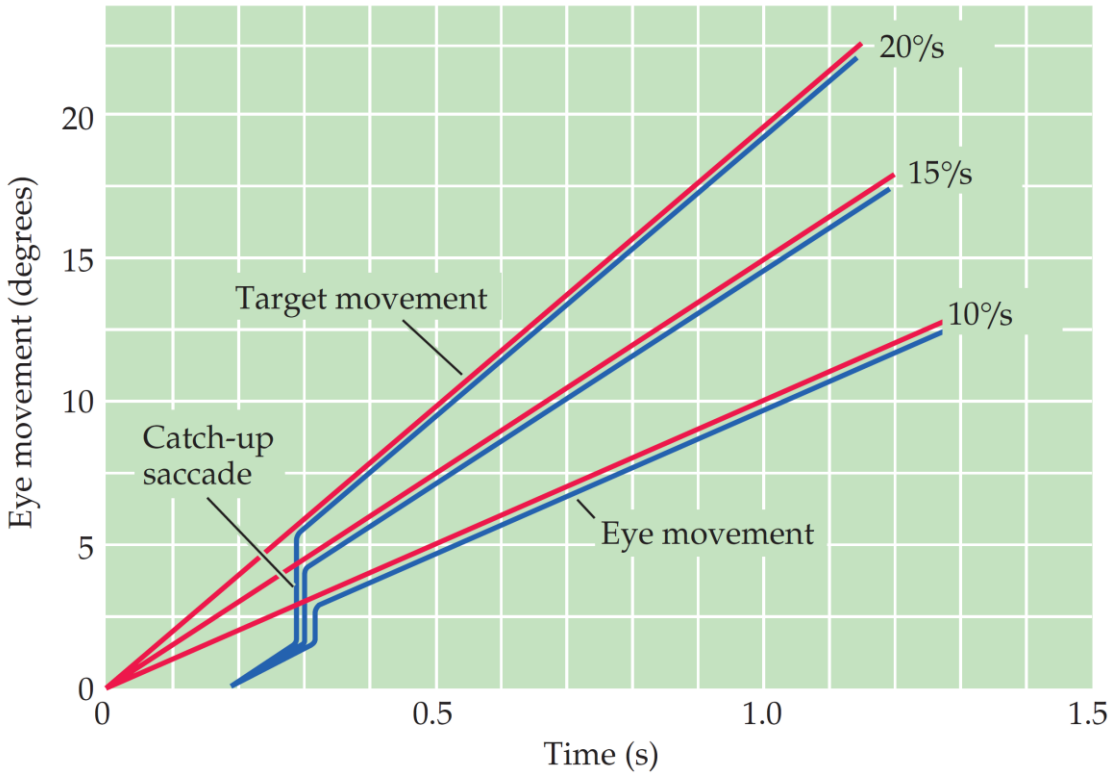
Eye in Motion: Saccades





- Ballistic movements
- Can be small or large
  (reading vs exploring the room)
- Can be voluntary or reflexive
- During 200ms period: compute the position of target with respect to fovea and convert to motor command
- The eye movement is 15-100 ms
- If target moves during eye movement, adjustments have to be made **after** movement is completed.

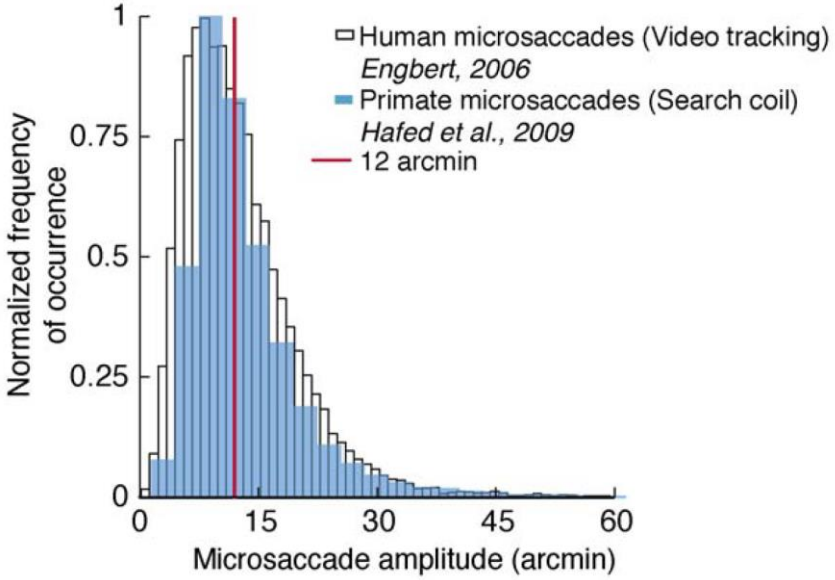
Eye in Motion: Smooth Pursuits



- Slower tracking movements that keep stimulus on the fovea
- Voluntary in that observer can choose whether or not to track moving stimulus
- Only highly trained observers can make a smooth pursuit movement in the absence of a moving target

Motion During Fixation

- **Drifts:**
  slow movements away from fixation point, 20 to 40 Hz
- **Flicks (microsaccades):**
  reposition the eye on target, 1 degree max
- **Ocular micro tremors:**
  150-2500nm, 40-100Hz



Cognitive Load Overview

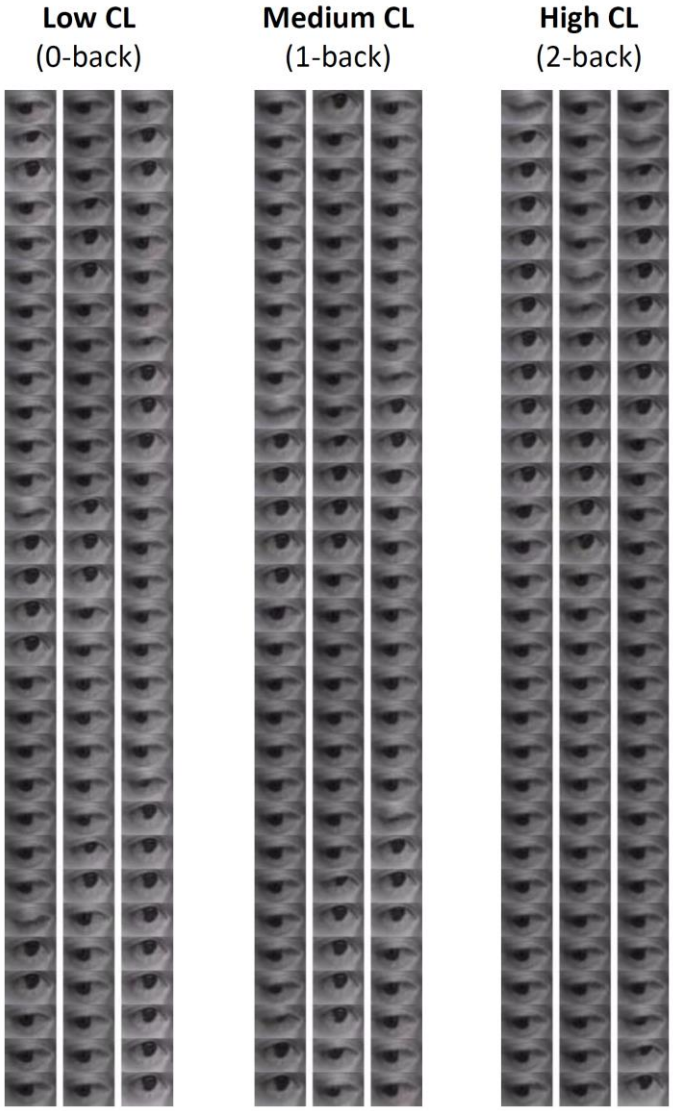
From the Perspective of Computer Vision

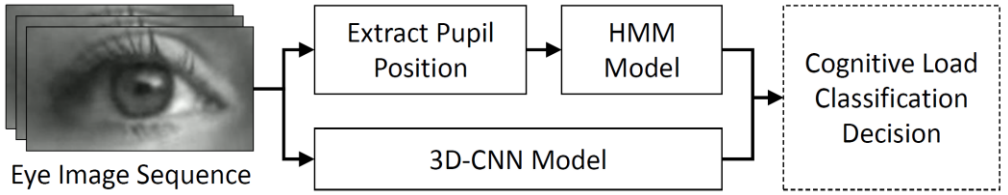
** Each of the following bullet points have several papers validating it.*

- Pupil equations:
  - Brighter **light** = smaller **pupil**
  - Higher **cognitive load** = larger **pupil**
- Blink equations
  - Higher **cognitive load** = slower **blink rate**
  - Higher **cognitive load** = shorter **blink duration**

- **Questions:**
  - Which of these metrics can be accurately extracted in real-world driving data?
  - Are there other metrics that may work better in such conditions?

Cognitive Load Estimation



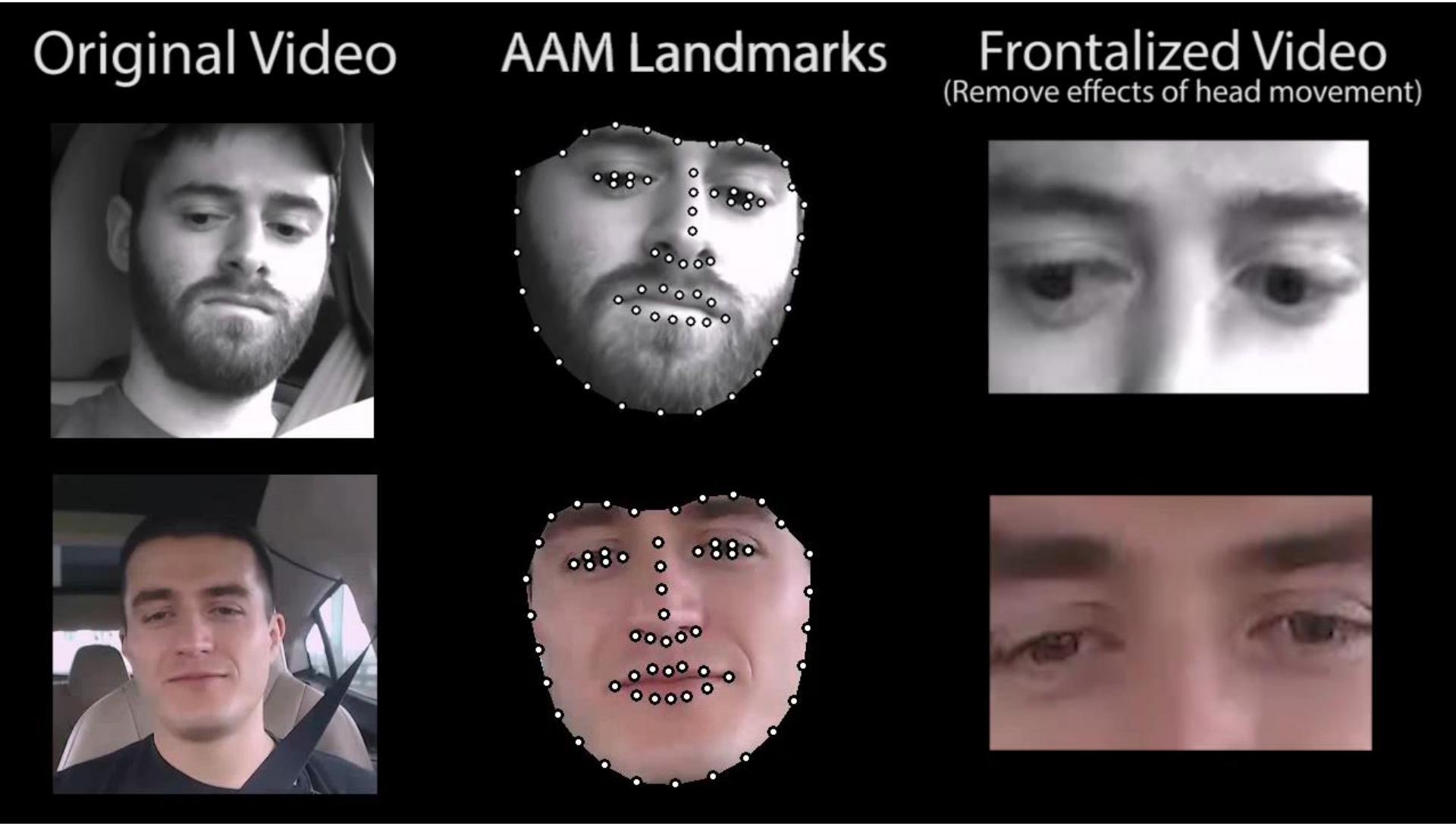


- 6 seconds, 16 fps, 90 images
- Two approaches: HMM and 3D-CNN

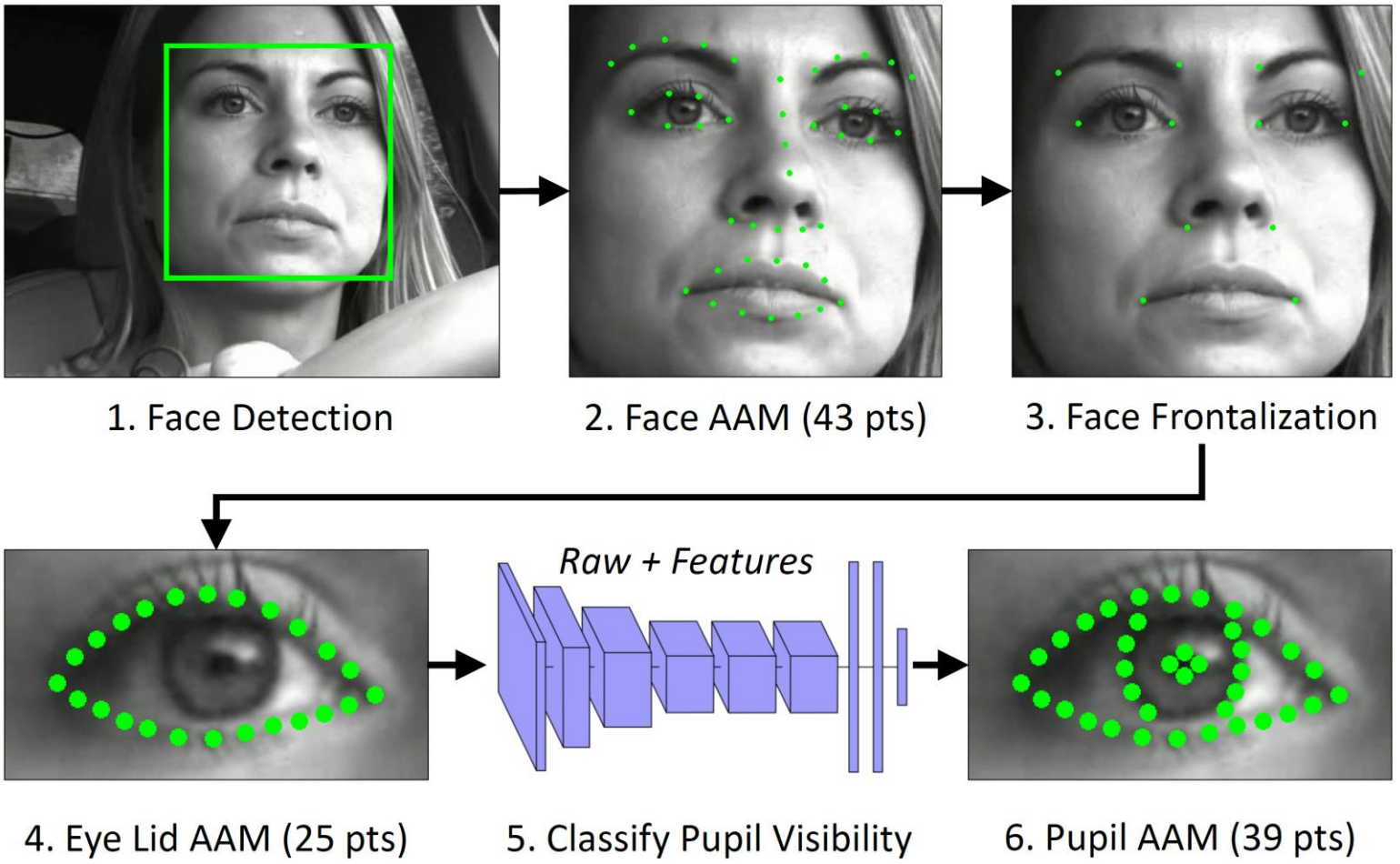
- **HMM:** Hidden Markov Model
  - **Input:** Sequence of pupil positions (normalized by intraocular segment)

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

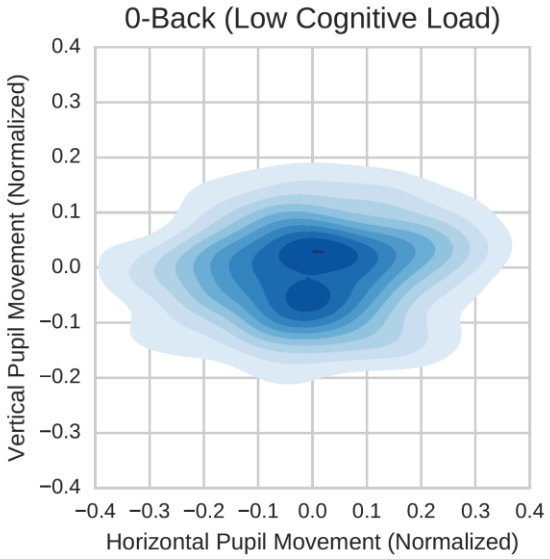
Dealing with Vibration and Movement

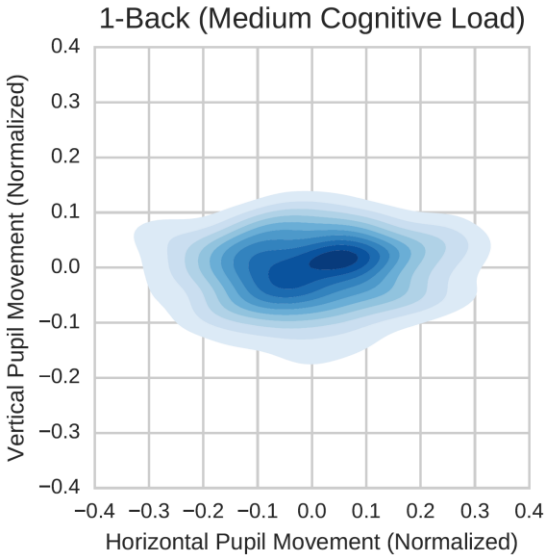


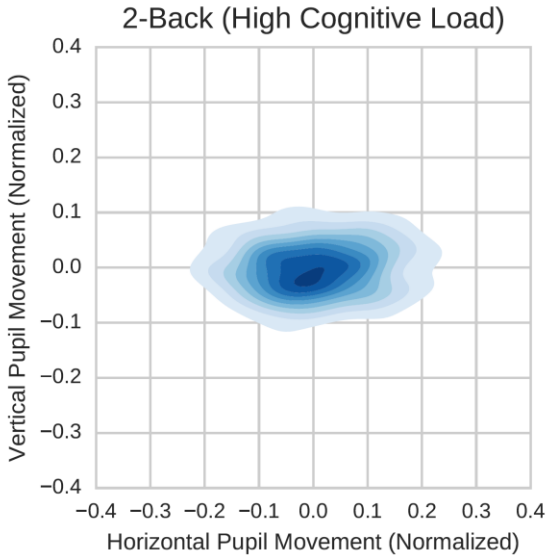
Preprocessing Pipeline



Visualizing the Dataset: Pupil Movement

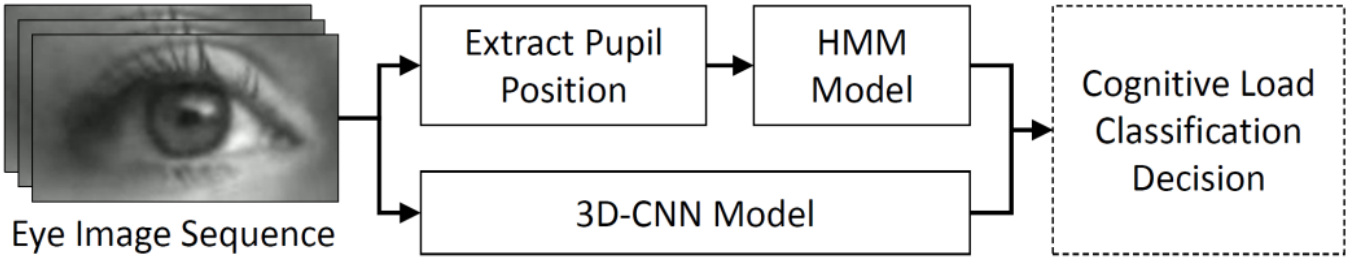


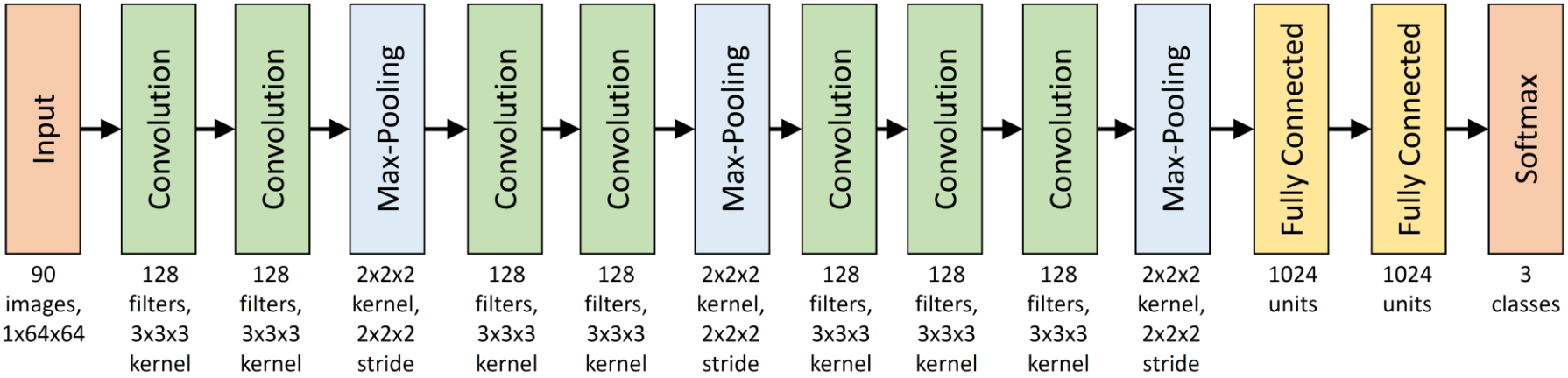




- **Metric:** Pupil position normalized by intraocular distance
- **Visualization:** Kernel density estimation (KDE)
- **Dataset size:** 92 subjects
- **Takeaway:** Observable aggregate differences between all 3 levels

Cognitive Load Estimation





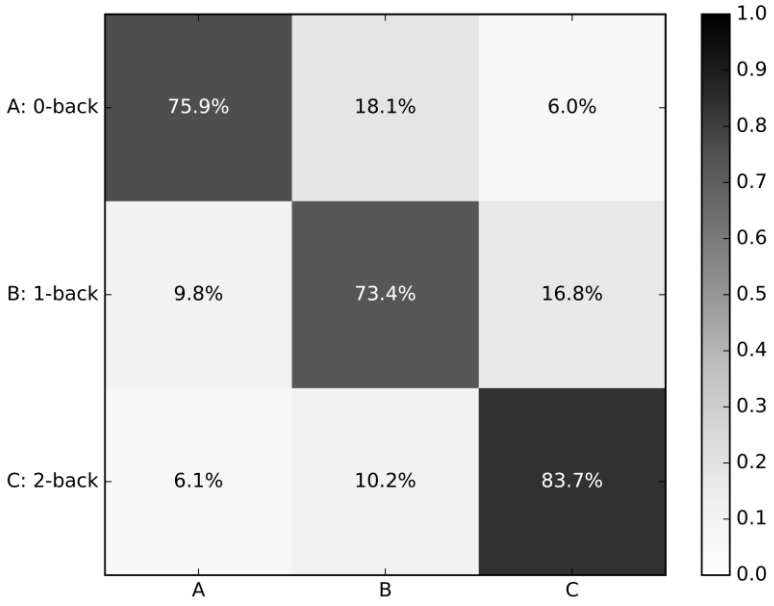
HMM: Hidden Markov Model

Input: Sequence of pupil positions (normalized by intraocular distance)

3D-CNN: Three Dimensional Convolutional Neural Network

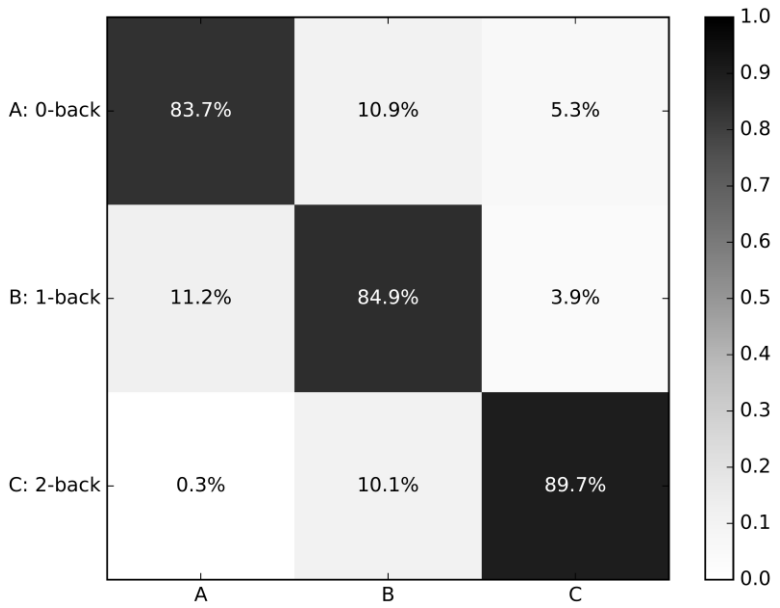
Input: Sequence of raw images of eye region

Driver Cognitive Load Estimation



HMM Approach

Average Accuracy: **77.7%**



3D-CNN Approach

Average Accuracy: **86.1%**

Cognitive Load Estimation:
Open Source = Open Innovation

Implication: Make driver cognitive load estimation accessible

Webcam

Video Stream

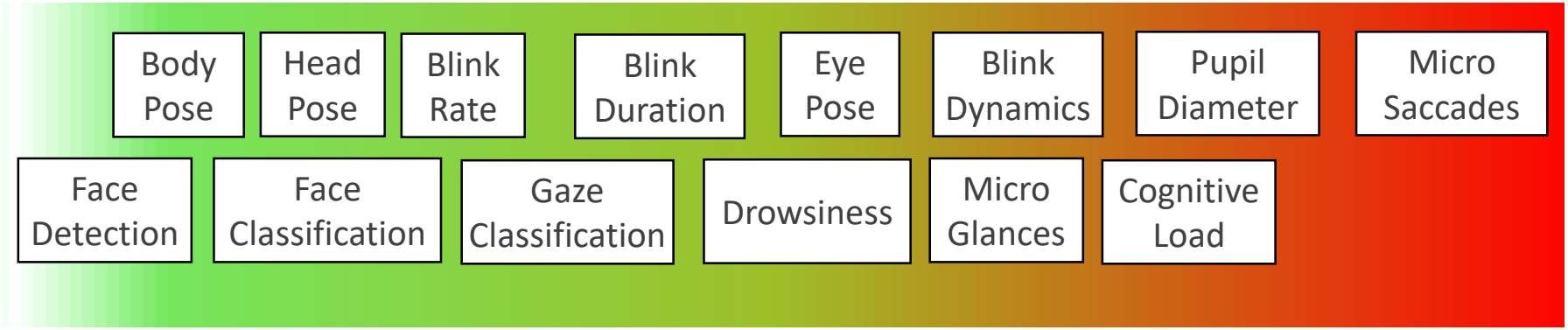
Low
Cognitive Load

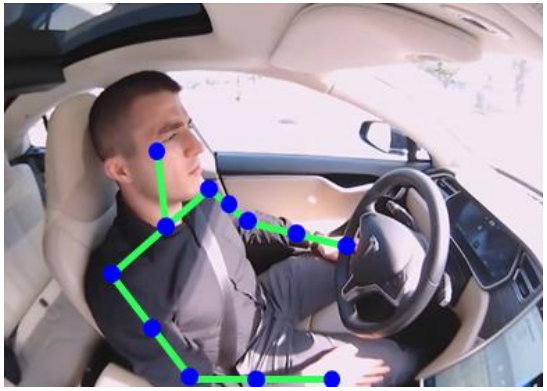
Medium
Cognitive Load

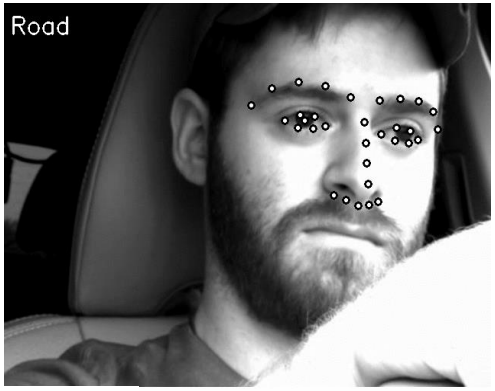
High
Cognitive Load

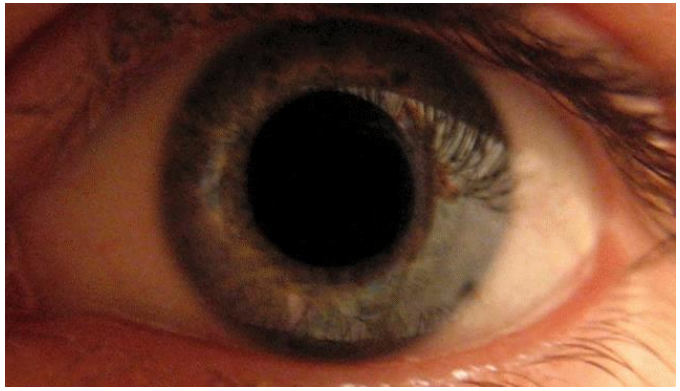
Drive State Detection:
A Computer Vision Perspective

Increasing level of detection resolution and **difficulty**









Human at the Center of Automation:
The Way to Full Autonomy Includes the Human

Fully Human Controlled

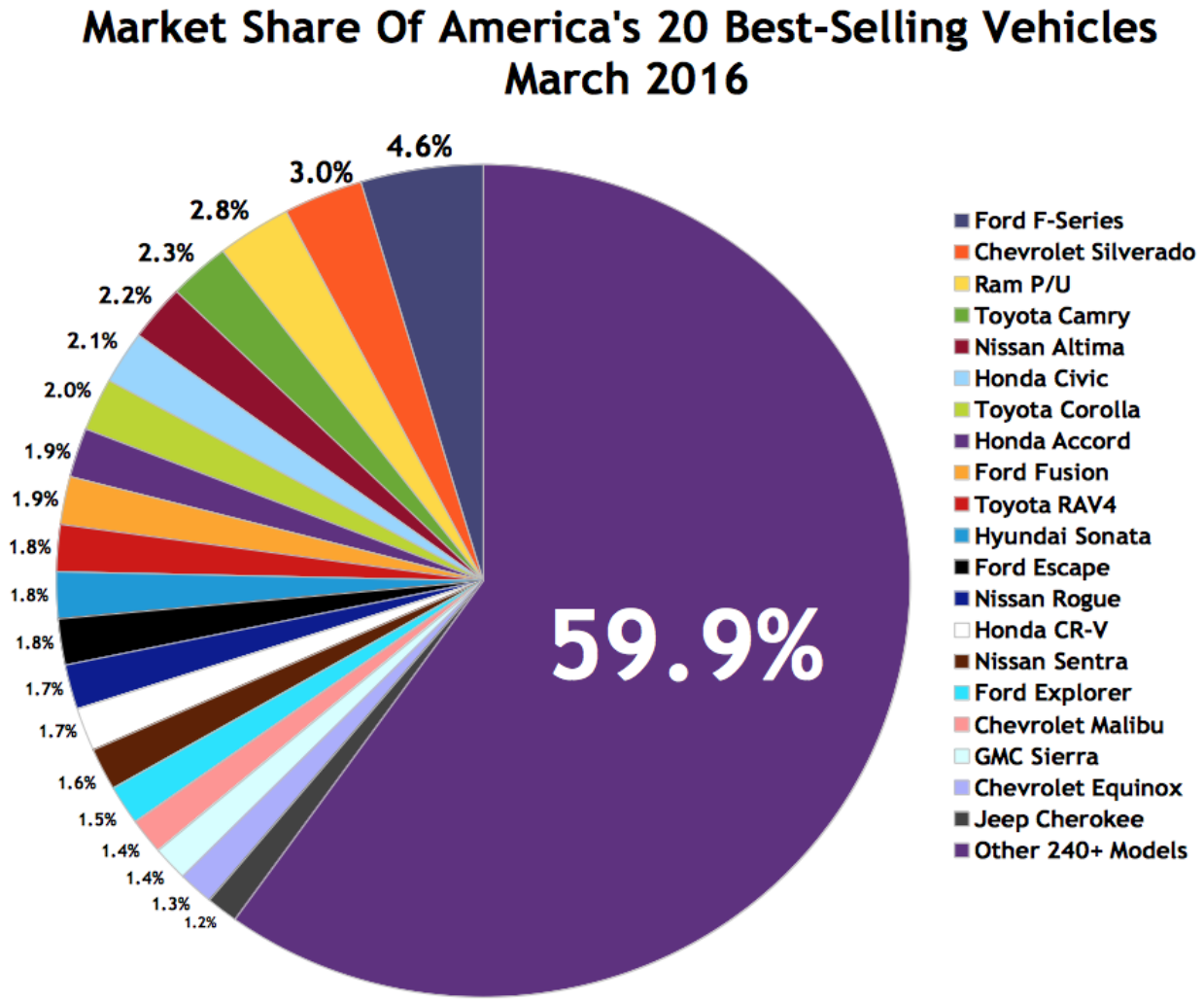
Fully Machine Controlled

Ford F150

Tesla Model S

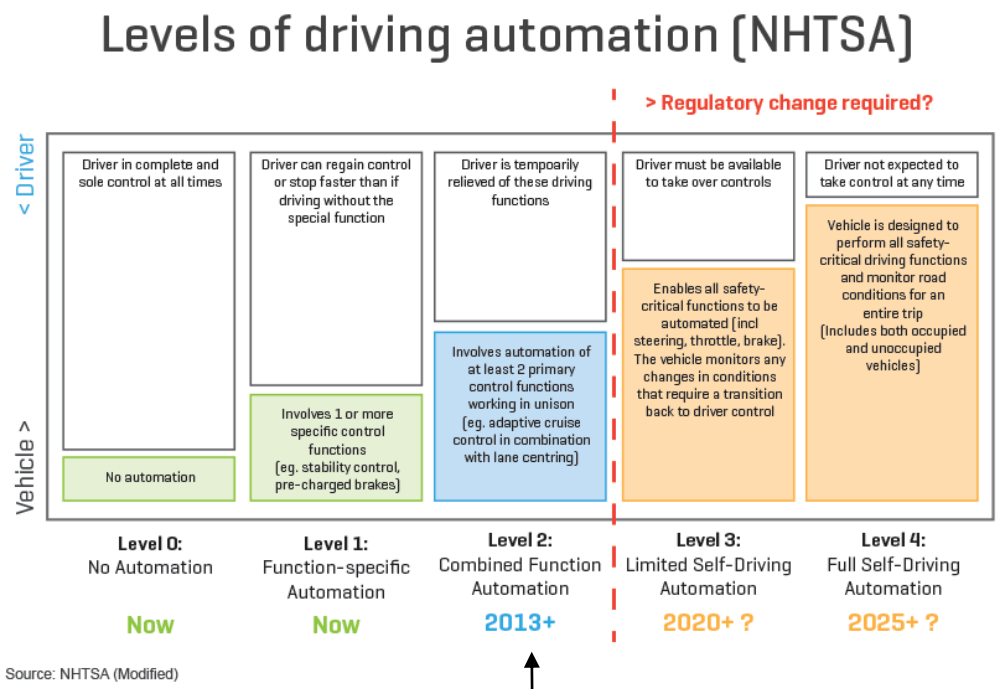
Google Self-Driving Car

Cars We Drive



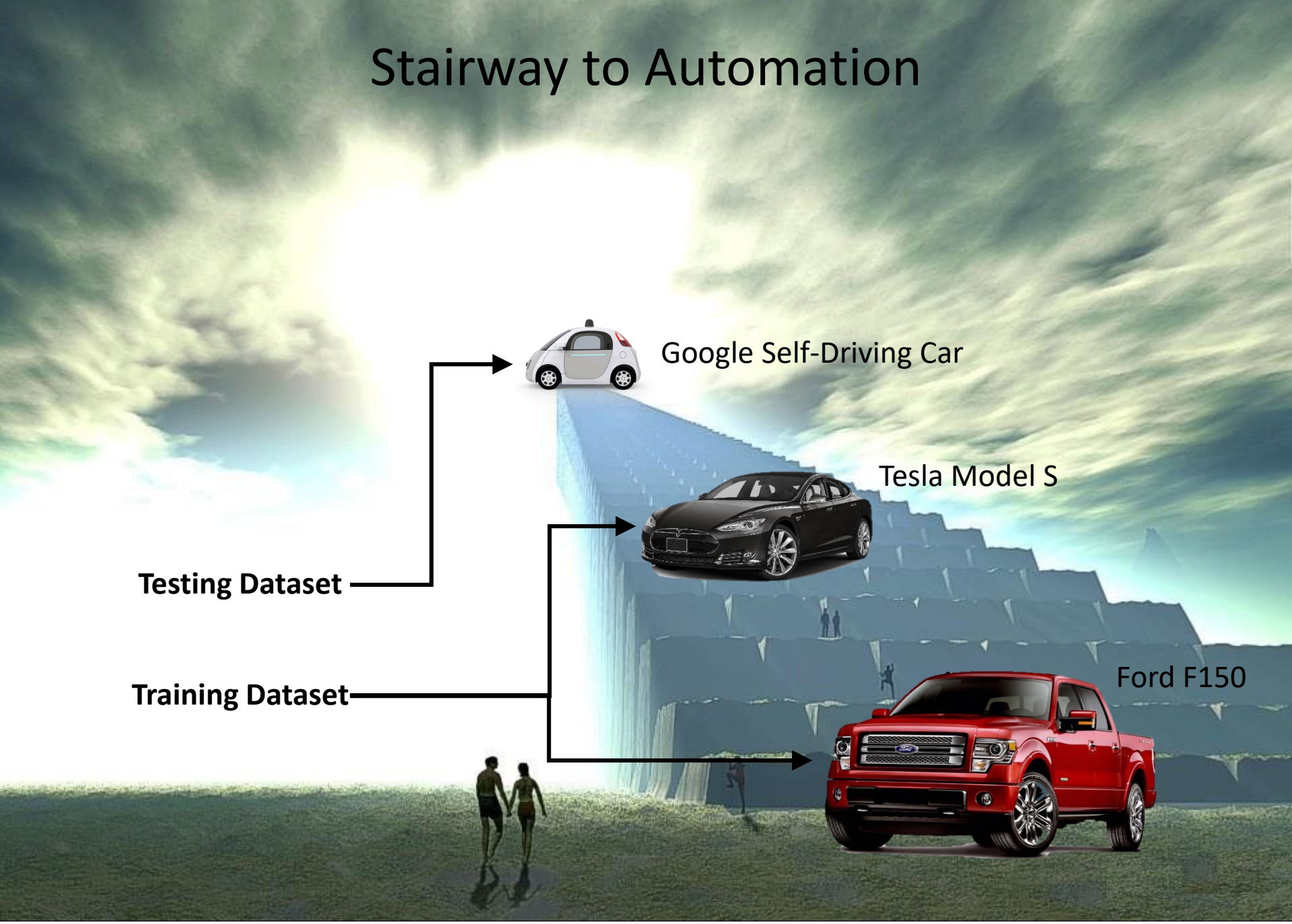
Human at the Center of Automation:
The Way to Full Autonomy Includes the Human

- **Emergency**
  - Automatic emergency breaking (AEB)
- **Warnings**
  - Lane departure warning (LDW)
  - Forward collision warning (FCW)
  - Blind spot detection
- **Longitudinal**
  - Adaptive cruise control (ACC)
- **Lateral**
  - Lane keep assist (LKA)
  - Automatic steering
- **Control and Planning**
  - Automatic lane change
  - Automatic parking



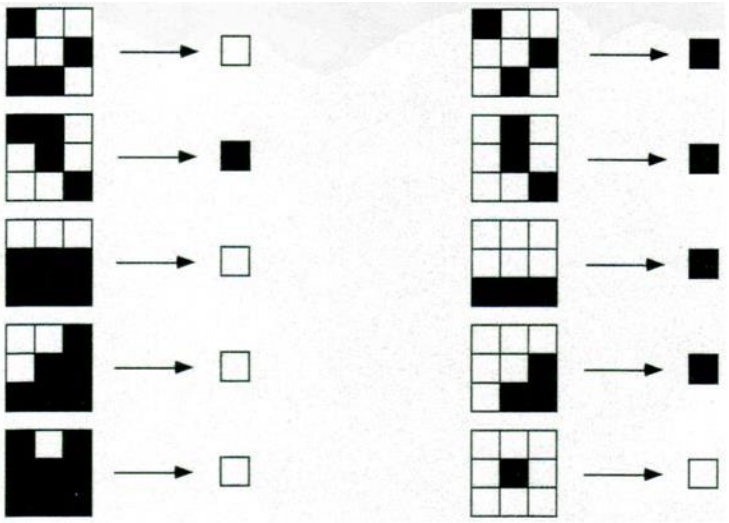
Tesla Autopilot

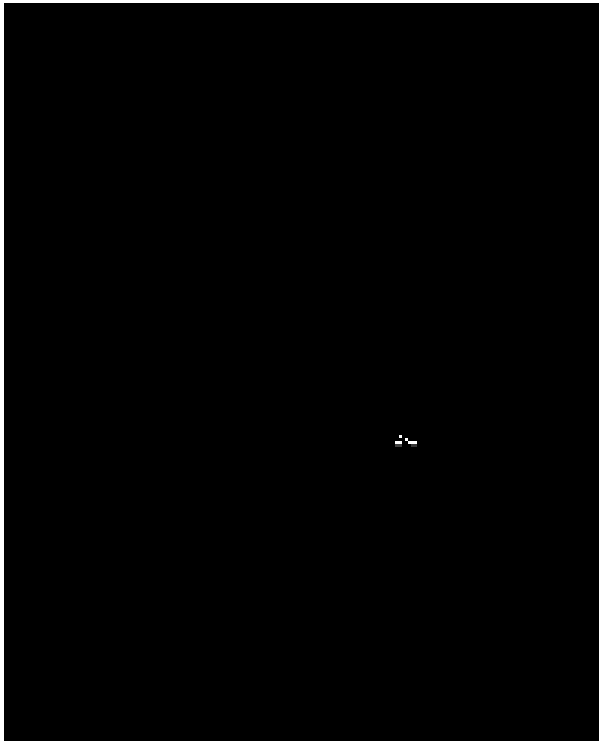
Stairway to Automation



We Still Need **Billions** of Miles of Data

…and Fundamental Breakthroughs in Deep Learning





What Next?

- Technical resources:
  - Deep Learning book: http://deeplearningbook.org
  - Foundational papers: http://deeplearning.net/reading-list
  - "Awesome" list: https://github.com/terryum/awesome-deep-learning-papers
- Fun/insightful blogs:
  - WildML (Denny Britz) blog: http://www.wildml.com
  - Andrej Karpathy blog: http://karpathy.github.io
  - Christopher Olah blog: http://colah.github.io
- Upcoming MIT Courses:
  - 6.S191: Introduction to Deep Learning: http://introtodeeplearning.com
  - 15.S14: Global Business of AI & Robotics: http://tiny.cc/gbair

- If you're interested in the application of deep learning in the automotive space, come do research with us: fridman@mit.edu

DeepTraffic Competition Winners

DeepTraffic Leaderboard

DeepTraffic - About DeepTraffic

Submissions for the competition and completion of the course will end 11 AM Friday (1/20/17)

Registered Students Top 3:

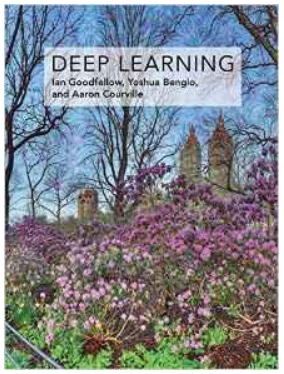
Rank	User	MPH
1	p_dolly	74.48
2	michael_gump	74.04
3	Jeffrey Hu	73.59

Top 3
($800 value)

UDACITY
Self-Driving Car Engineer Nanodegree

+

Top 1
(Priceless)



Purnawirman Purnawirman

Michael Gump

Jeffrey Hu