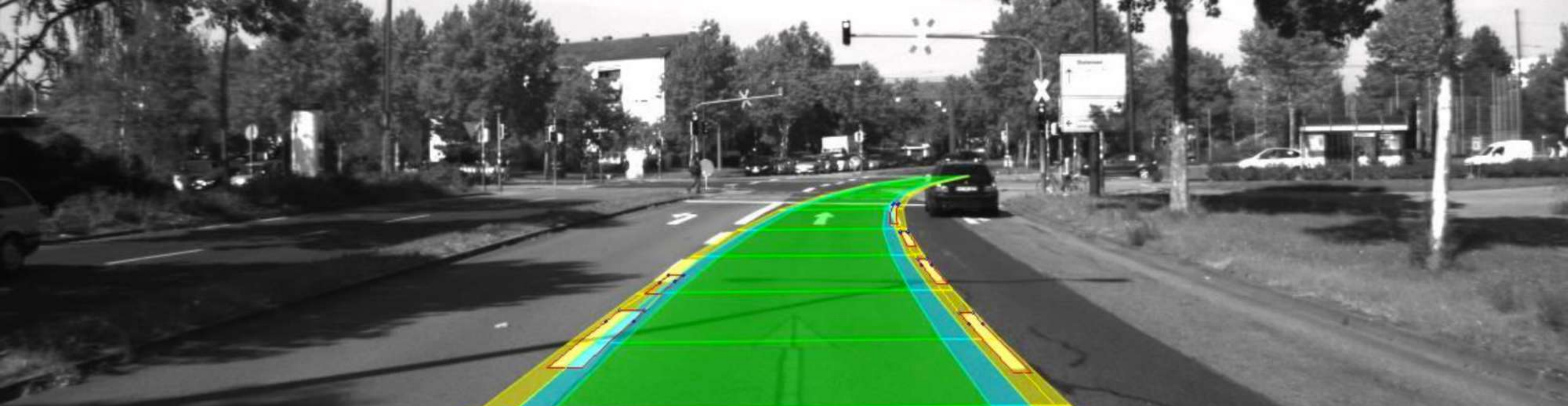


Module 3: Automotive CPS Data driven modeling

Principles of Modeling for Cyber-Physical Systems

Instructor: Madhur Behl

Slides credits:
- Urs Muller

Everything that moves will go autonomous




Cars

Trucks

Carts

Drones

Safe

Self Learning

Secure

Autonomous Systems

Trusted

Connected

Resilient

THE FUTURE OF TRANSPORTATION STACK

COMET LABS


WAYMO
7XJT313
Google
WAYMO

THE 6 LEVELS OF AUTONOMOUS DRIVING

LØ	L1	L2	L3	L4	L5
No Automation	Driver Assistance	Partial Automation	Conditional Automation	High Automation	Full Automation

SOCIETY OF AUTOMOTIVE ENGINEERS (SAE) AUTOMATION LEVELS

Full Automation

0	1	2	3	4	5
No Automation	**Driver Assistance**	**Partial Automation**	**Conditional Automation**	**High Automation**	**Full Automation**
Zero autonomy; the driver performs all driving tasks.	Vehicle is controlled by the driver, but some driving assist features may be included in the vehicle design.	Vehicle has combined automated functions, like acceleration and steering, but the driver must remain engaged with the driving task and monitor the environment at all times.	Driver is a necessity, but is not required to monitor the environment. The driver must be ready to take control of the vehicle at all times with notice.	The vehicle is capable of performing all driving functions under certain conditions. The driver may have the option to control the vehicle.	The vehicle is capable of performing all driving functions under all conditions. The driver may have the option to control the vehicle.

1.35 million deaths worldwide due to vehicle crashes

94% of crashes involve human choice or error in the US.

3 million

Americans age 40 and older are blind or have low vision

79%

of seniors age 65 and older living in car-dependent communities

42 hours

wasted in traffic each year per person

Localization and Mapping

Where am I ?

Scene Understanding

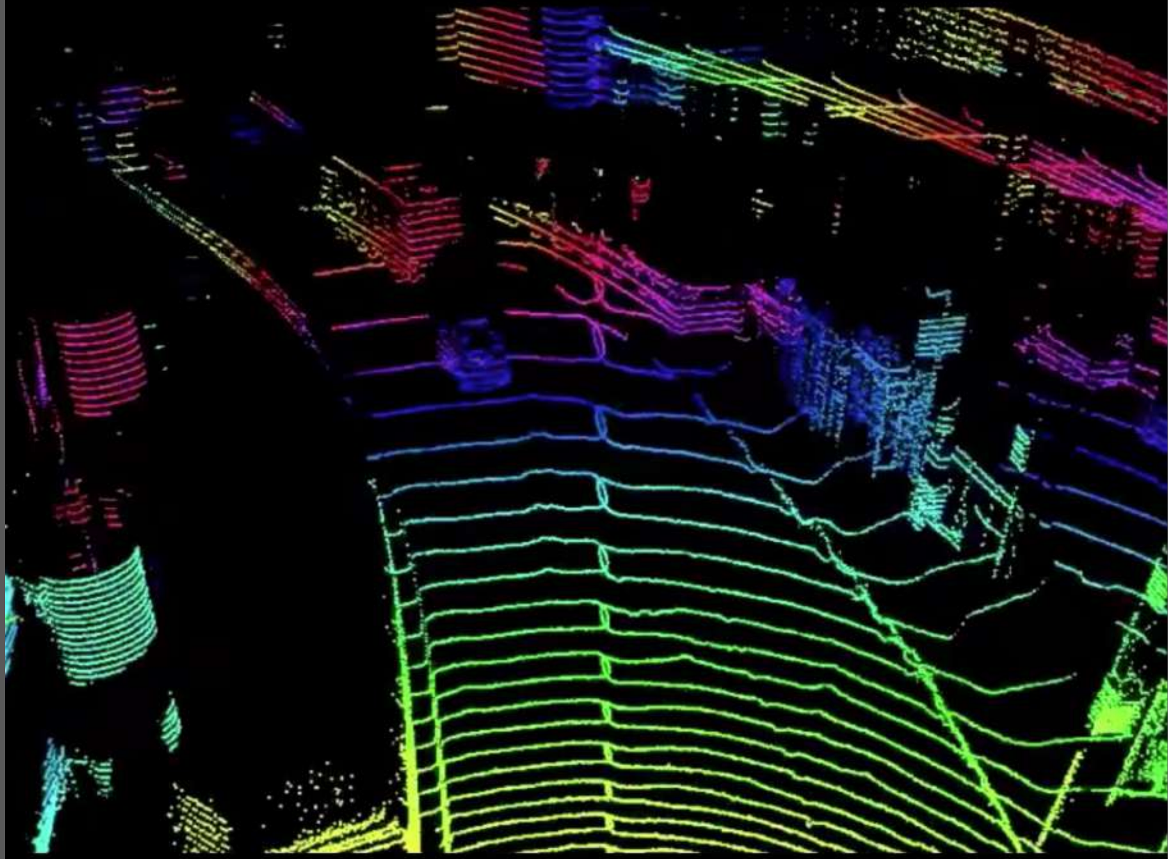
Where/who/what/why of everyone/everything else ?

Trajectory Planning and Control

Where should I go next ?
How do I steer and accelerate ?

Human Interaction

How do I convey my intent to the passenger and everyone else ?



Localization and Mapping

Where am I ?

Scene Understanding

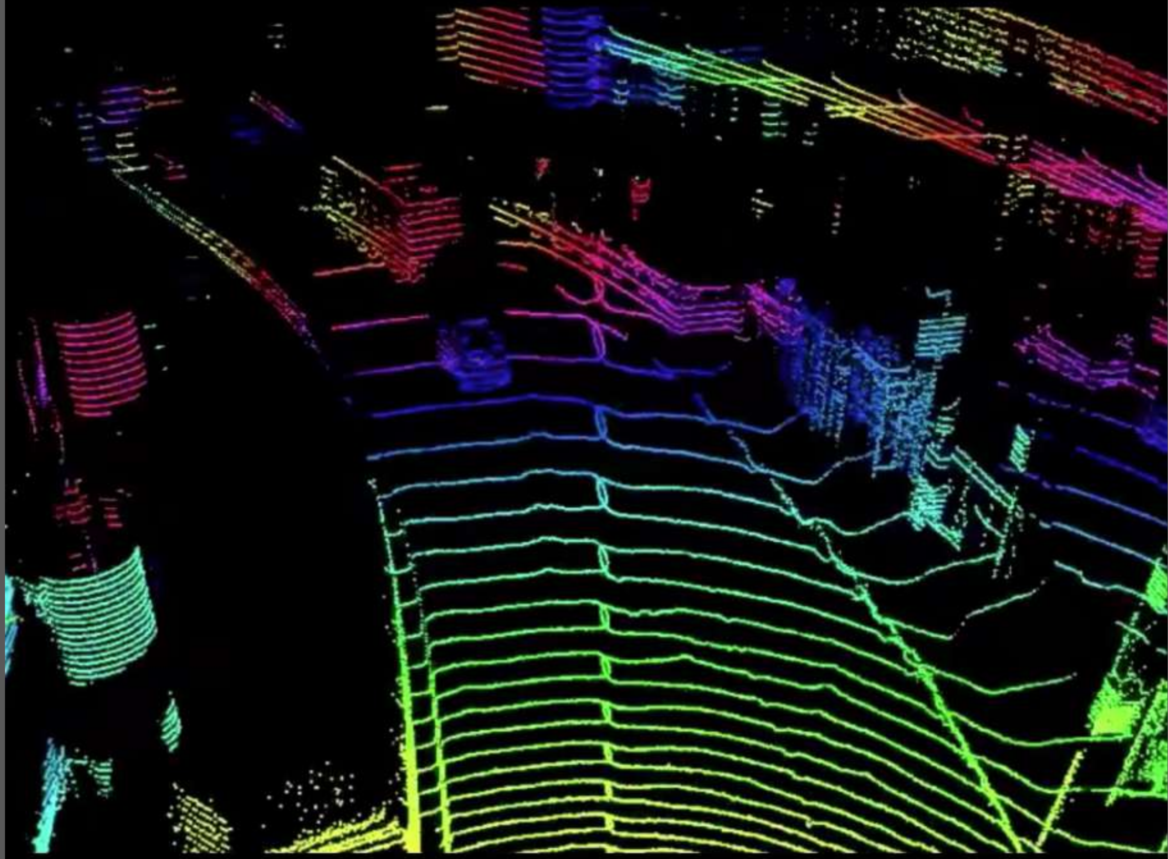
Where/who/what/why of everyone/everything else ?

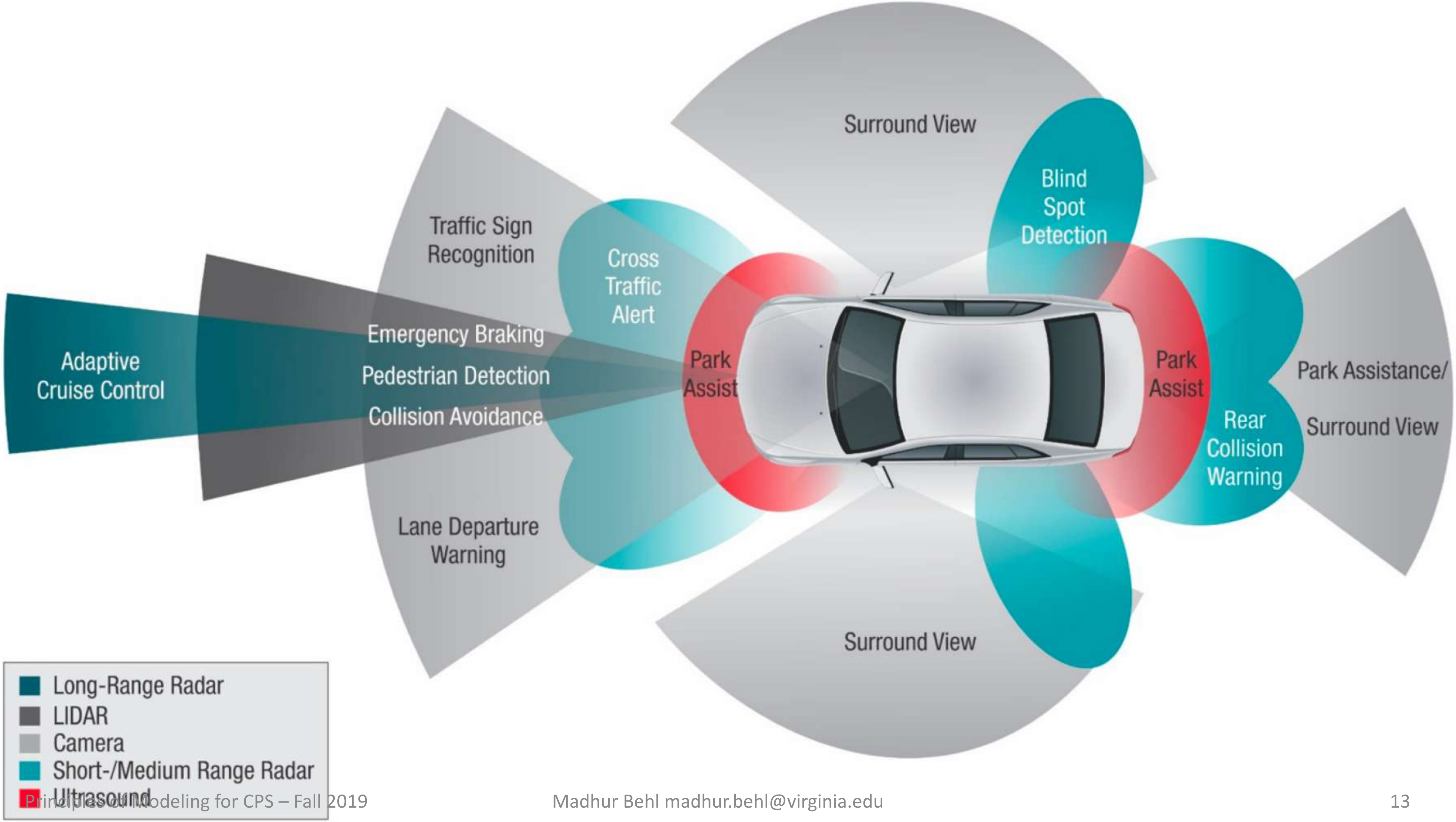
Trajectory Planning and Control

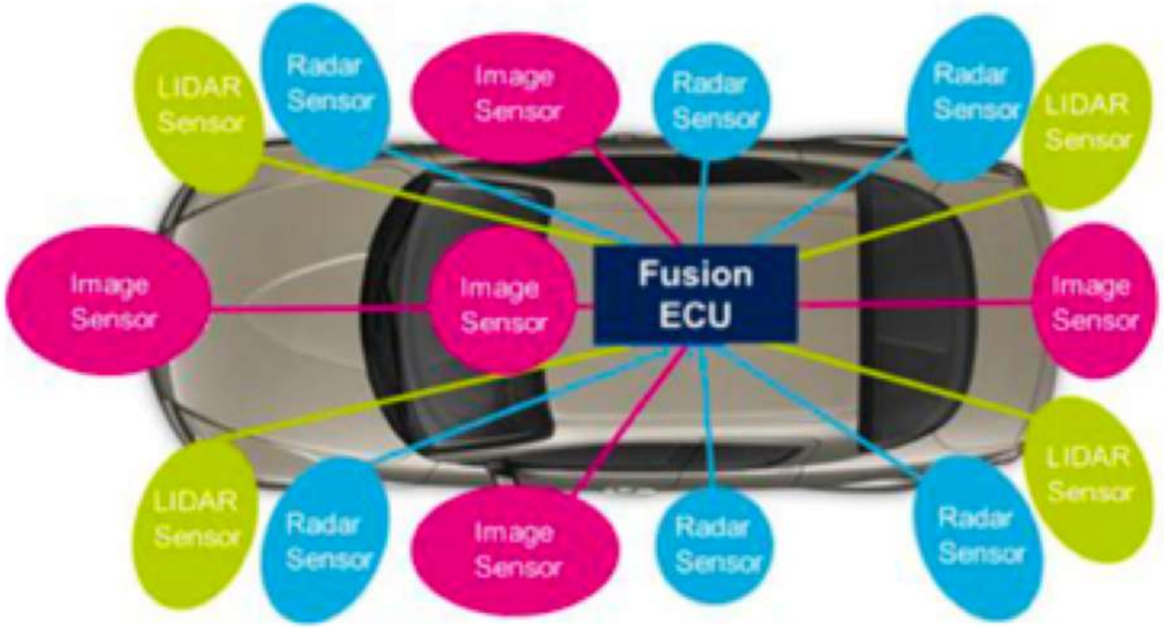
Where should I go next ?
How do I steer and accelerate ?

Human Interaction

How do I convey my intent to the passenger and everyone else ?



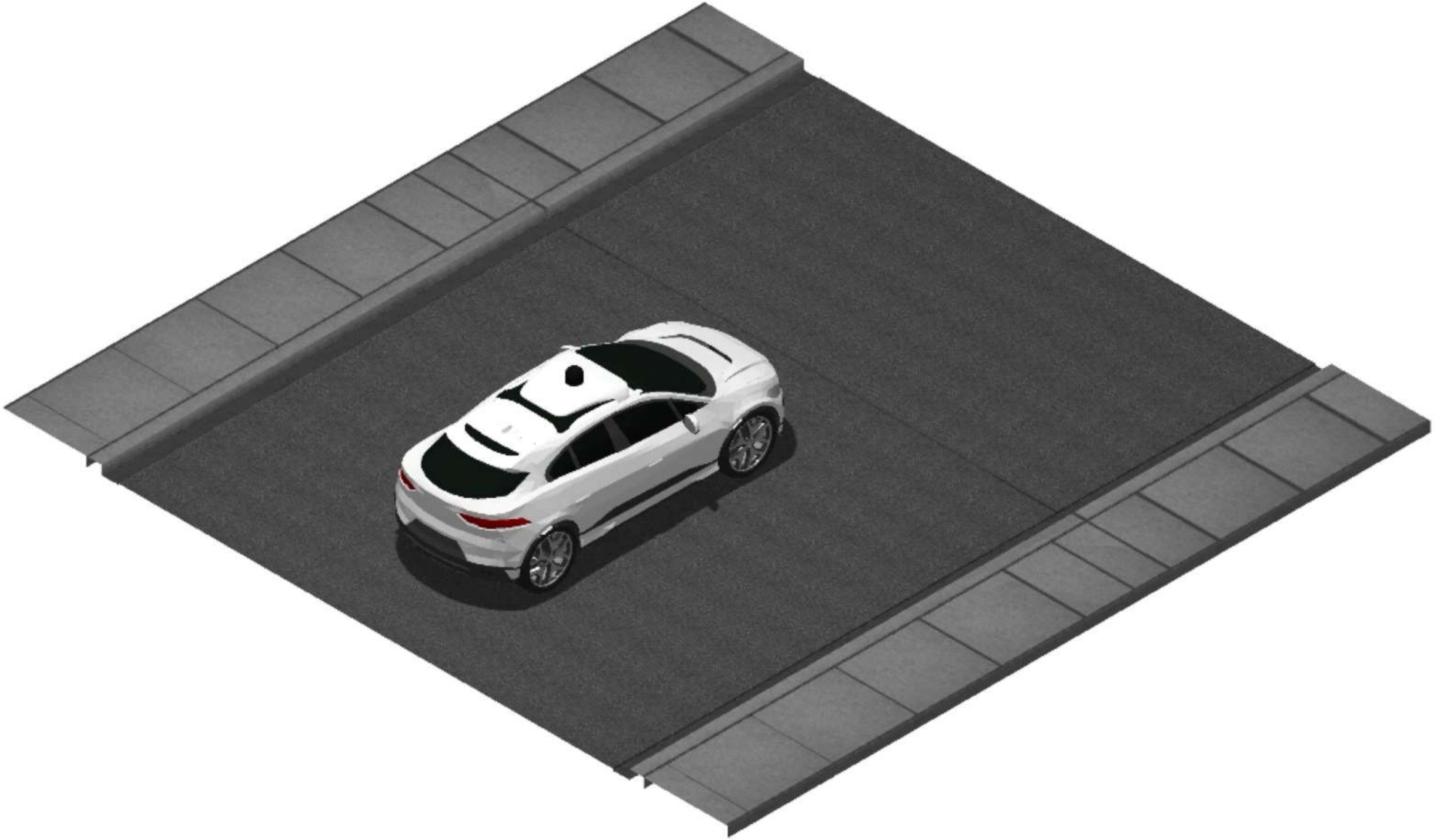

Surround View
Blind Spot Detection
Traffic Sign Recognition
Cross Traffic Alert
Adaptive Cruise Control
Emergency Braking
Pedestrian Detection
Collision Avoidance
Park Assist
Park Assist
Park Assistance/ Surround View
Rear Collision Warning
Lane Departure Warning
Surround View
Long-Range Radar
LIDAR
Camera
Short-/Medium Range Radar
Ultrasound



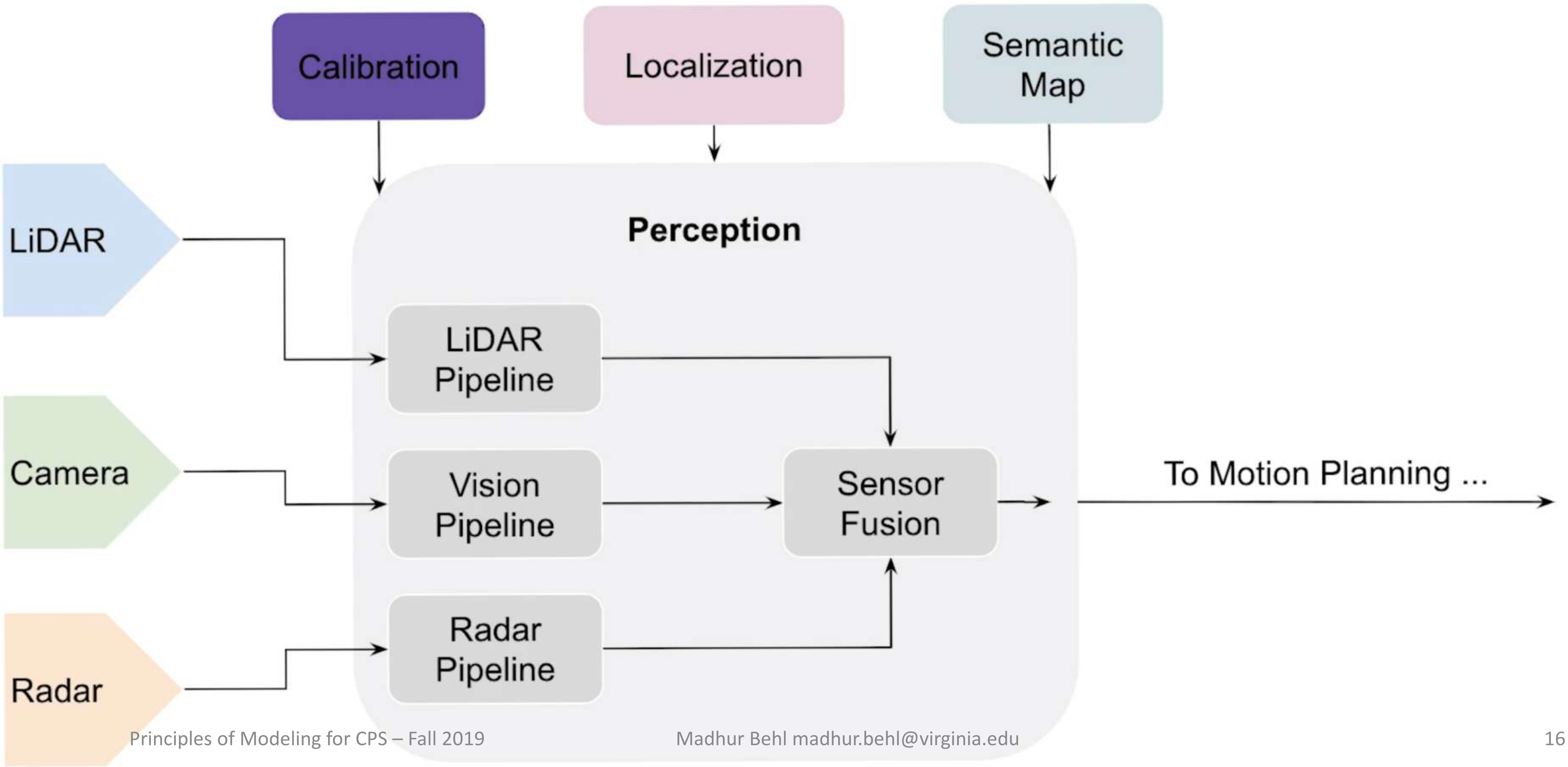
Camera

Radar

LIDAR

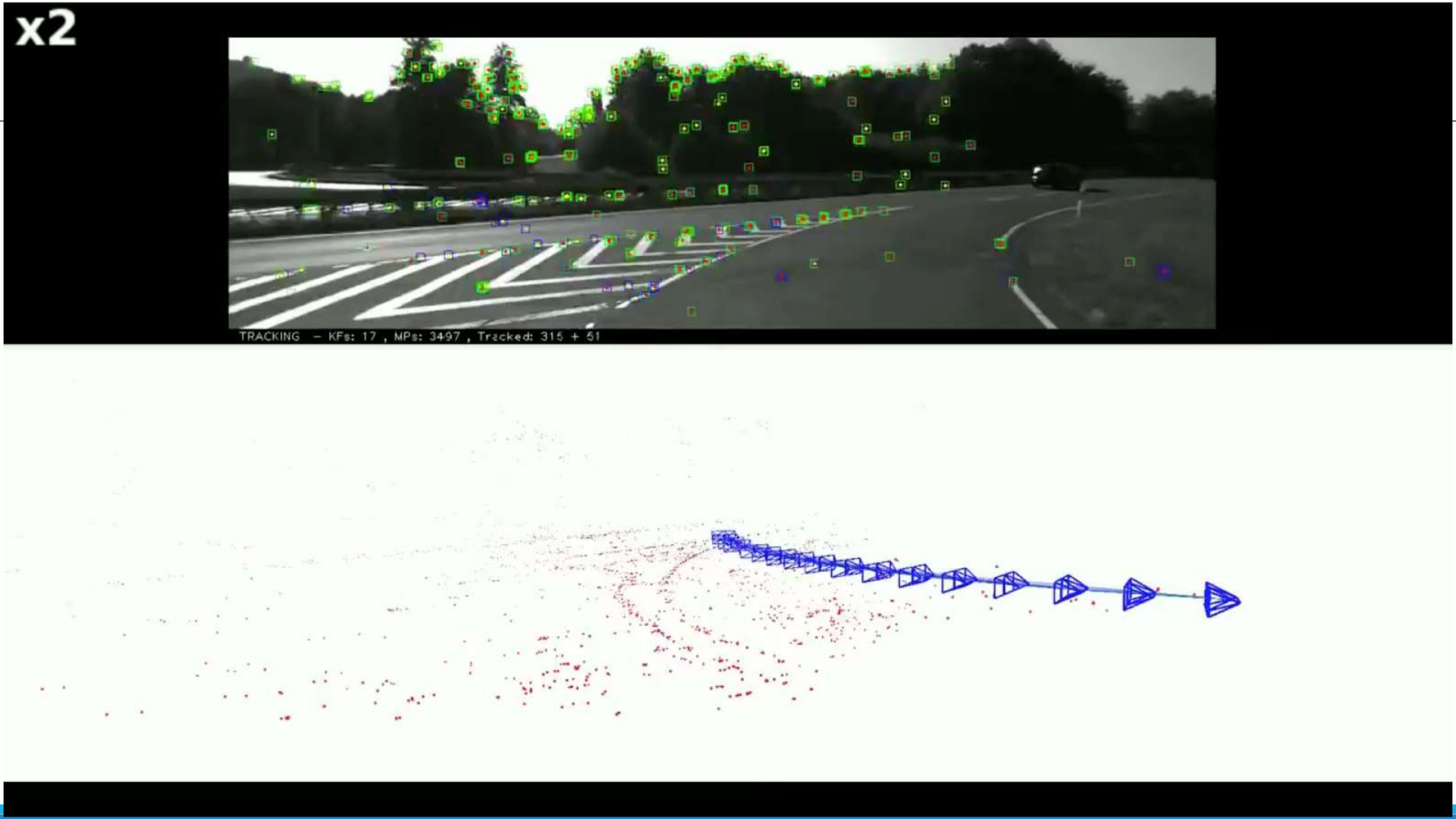


Perception in AV Stack



SLAM: Simultaneous Localization and Mapping

What works: SIFT and optical flow

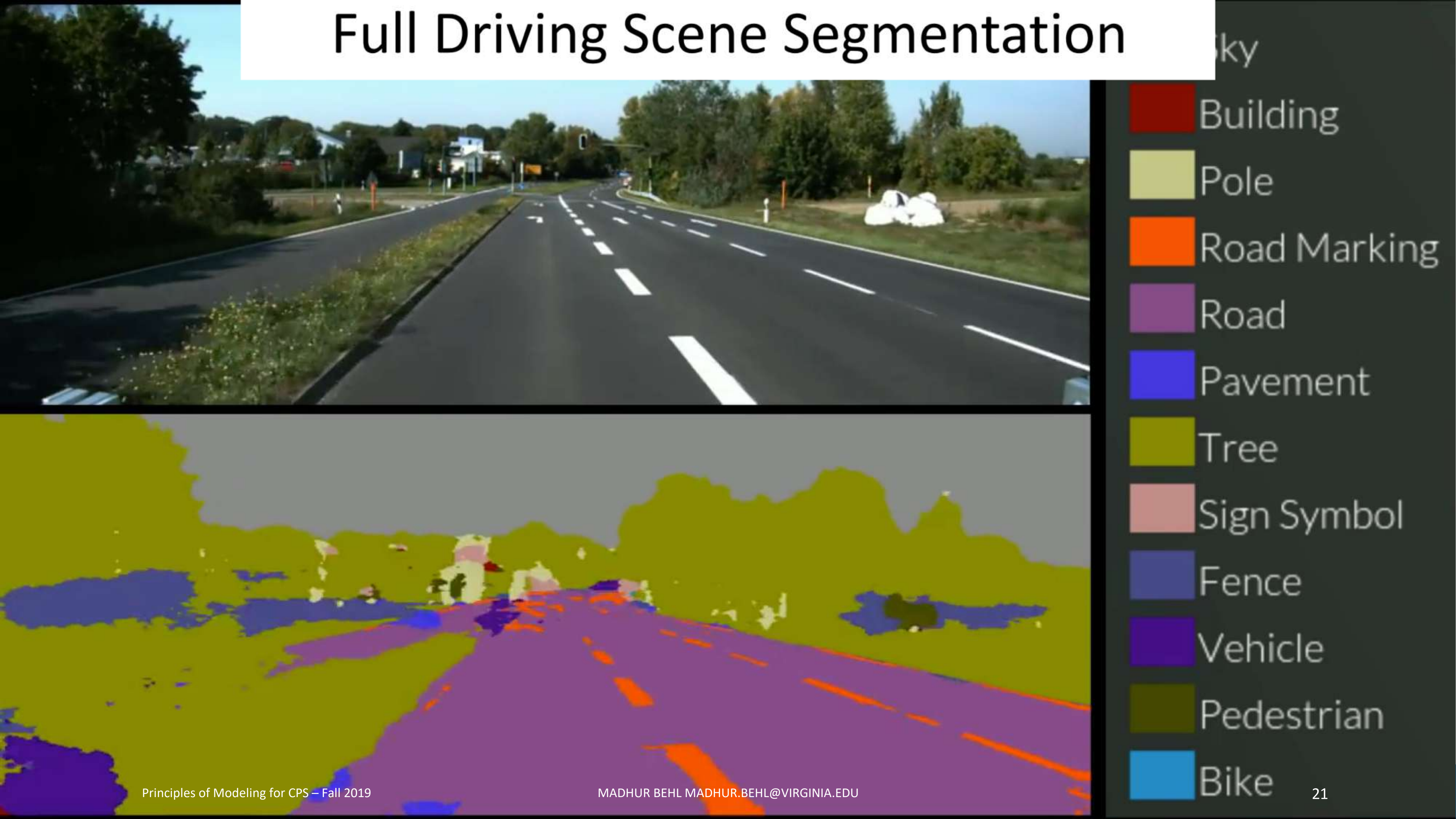


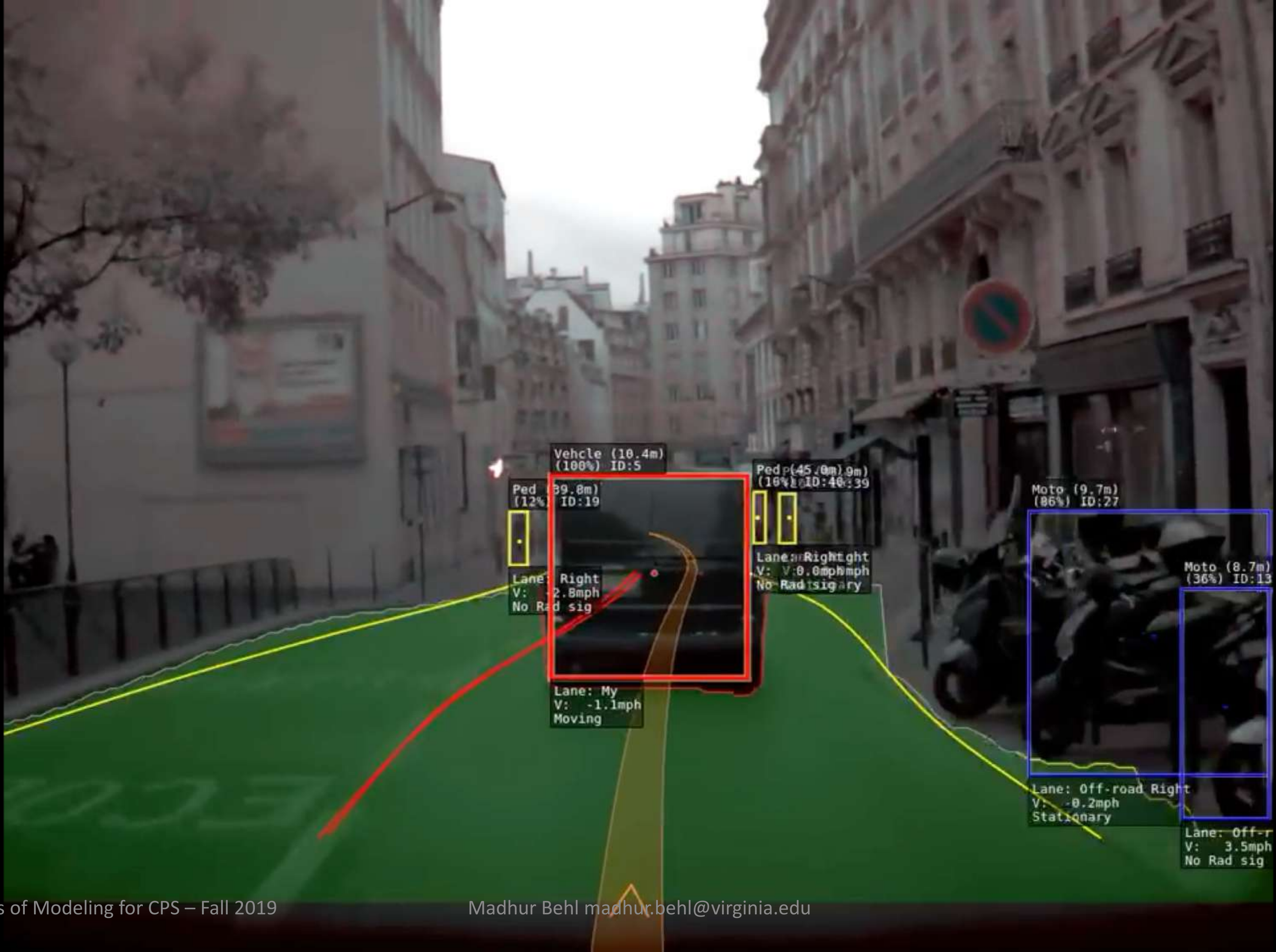
Object Detection

- Past approaches: cascades classifiers (Haar-like features)
- Where deep learning can help:
  recognition, classification, detection


Sensor Fusion
International
CES
TECH EVENTS
ection - NVDriveNet detection
suv-truck
car
suv-truck
suv-truck
suv-truck
suv-truck
Front:
car
car
suv-truck
suv-truck
Rear :

Full Driving Scene Segmentation




Vehcle (10.4m)
(100%) ID:5
Lane: My
V: -1.1mph
Moving
Moto (9.7m)
(86%) ID:27
Moto (8.7m)
(36%) ID:13
Lane: Off-road Right
V: -0.2mph
Stationary
V: 3.5mph
No Rad sig

Localization and Mapping

Where am I ?

Scene Understanding

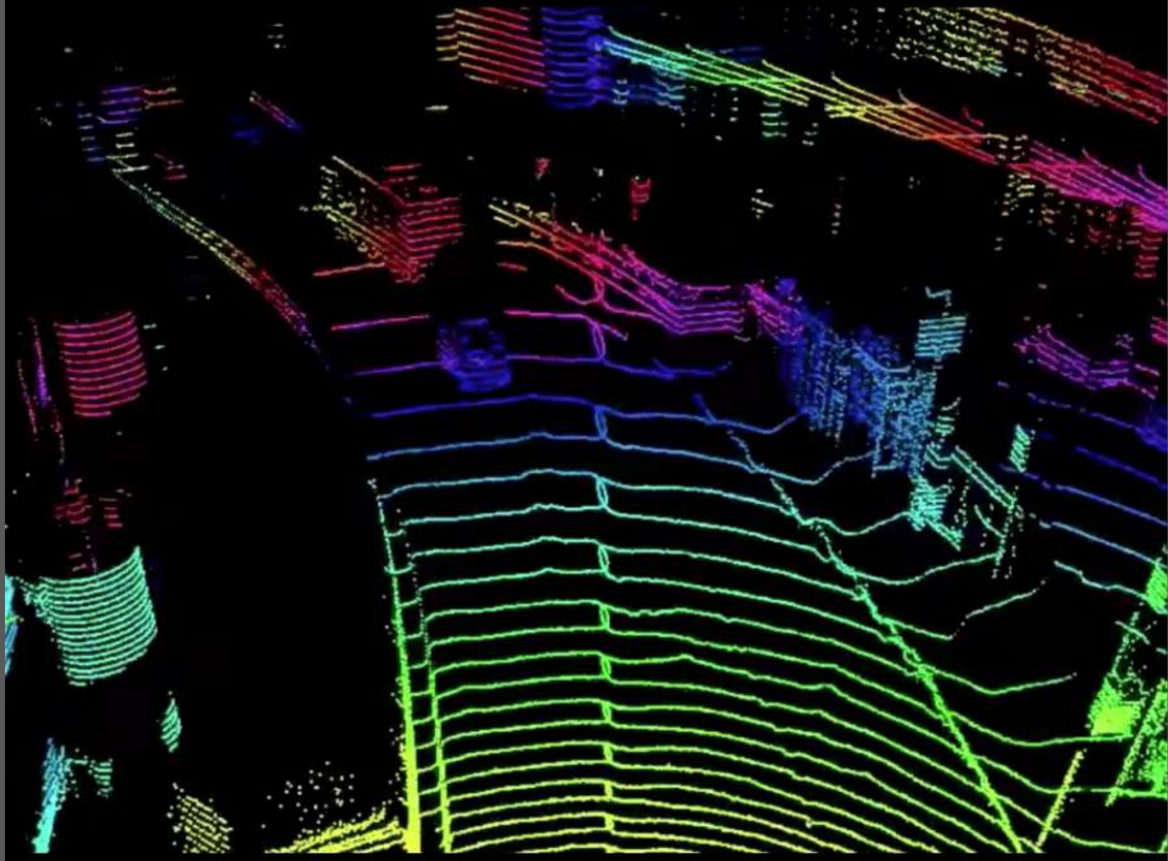
Where/who/what/why of everyone/everything else ?

Trajectory Planning and Control

Where should I go next ?
How do I steer and accelerate ?

Human Interaction

How do I convey my intent to the passenger and everyone else ?




Vehcle (28.9m)
(100%) ID:38
Lane: Overlp-Left
V: -1.2mph
Moving
Vehcle (44.3m)
(100%) ID:20
Lane: Imm-Right
V: 0.7mph
Moving
Moto (9.6m)
(100%) ID:34
Lane: Imm-Right
V: 2.7mph
Moving

Detailed three-dimensional maps that highlight information such as road profiles, curbs and sidewalks, lane markers, crosswalks, traffic lights, stop signs, and other road features.

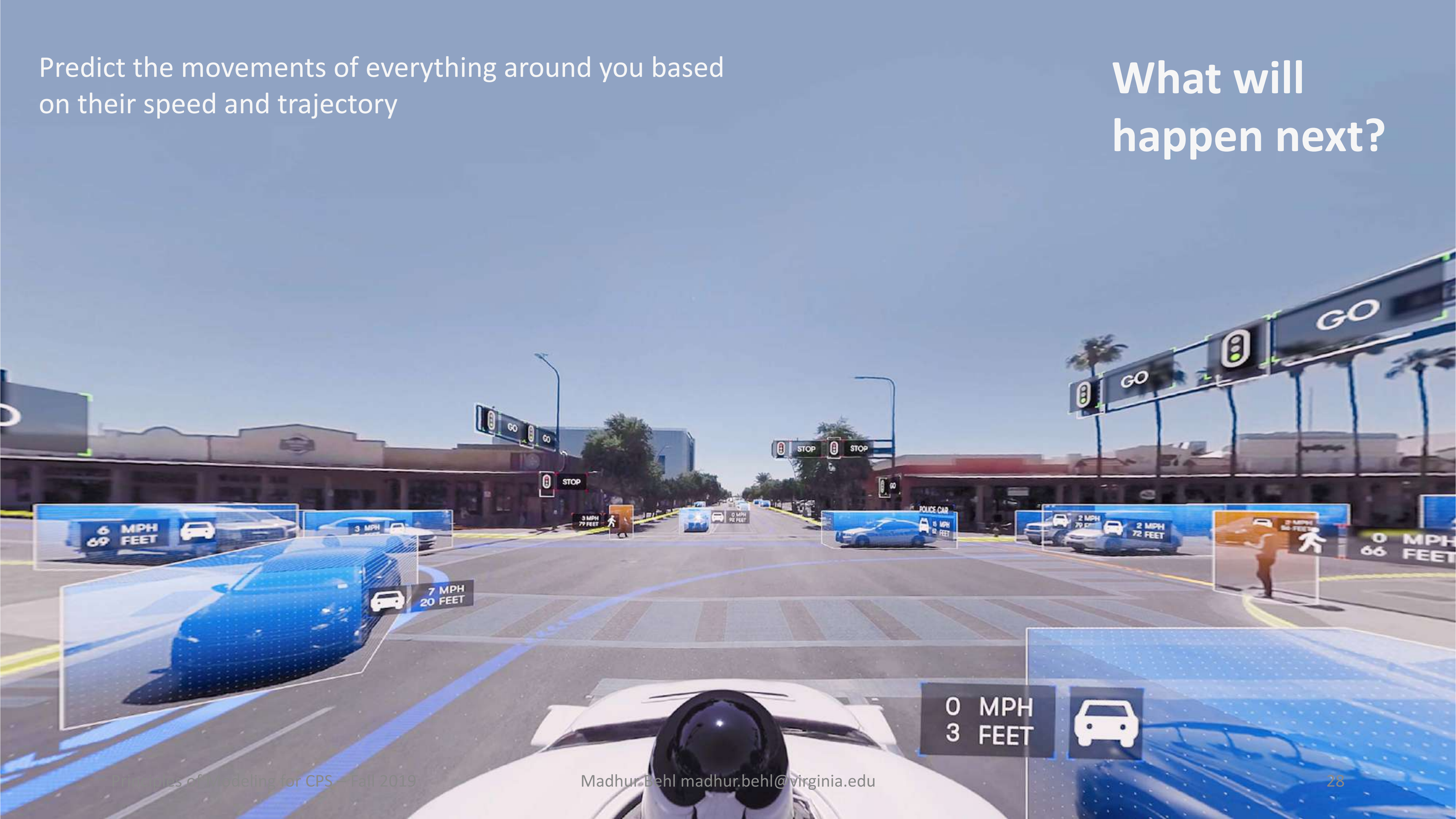
Where am I?

Scan constantly for objects around the vehicle—pedestrians, cyclists, vehicles, road work, obstructions—and continuously read traffic controls, from traffic light color and railroad crossing gates to temporary stop signs.

What's around me?

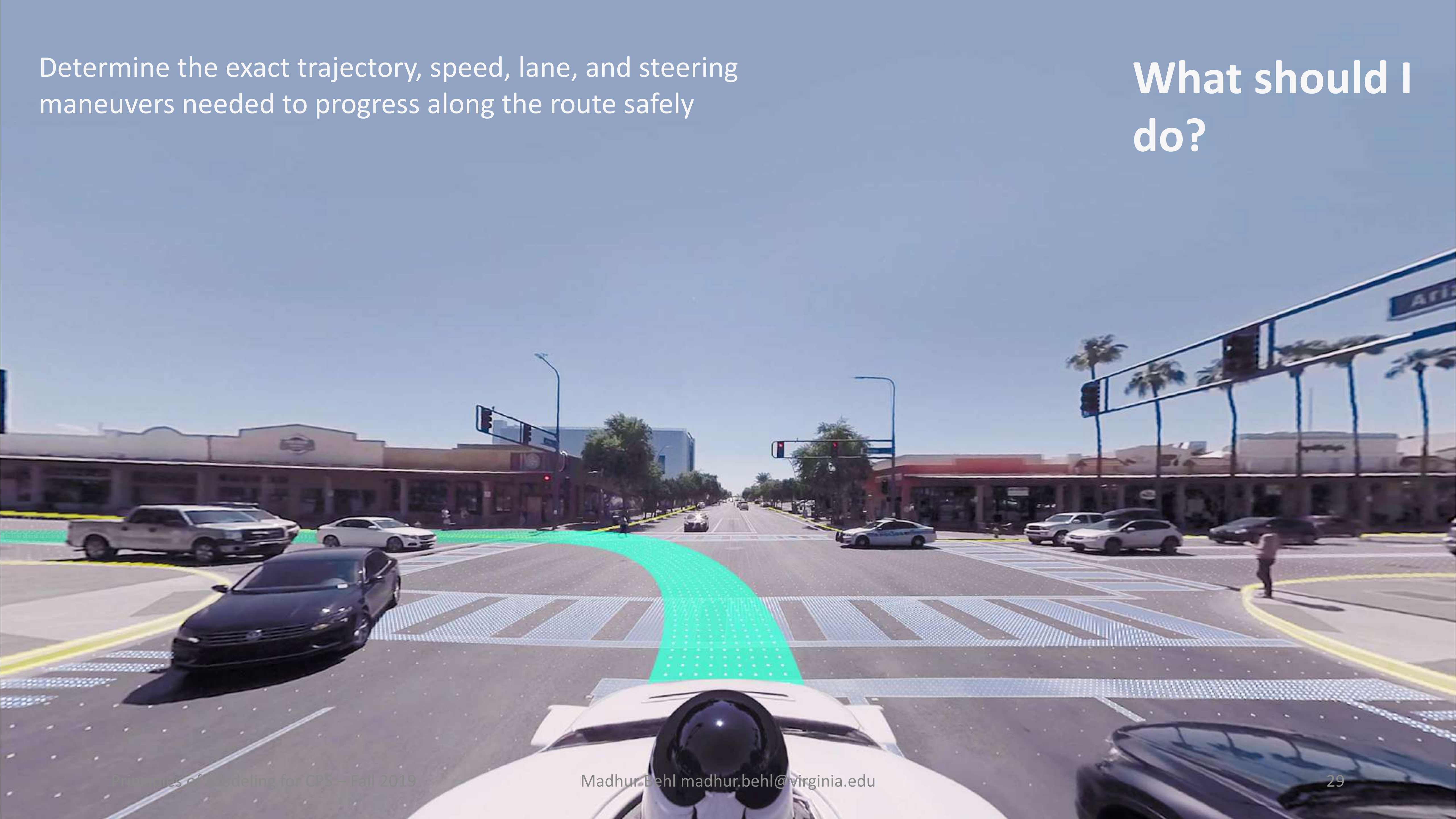
Predict the movements of everything around you based on their speed and trajectory

What will happen next?

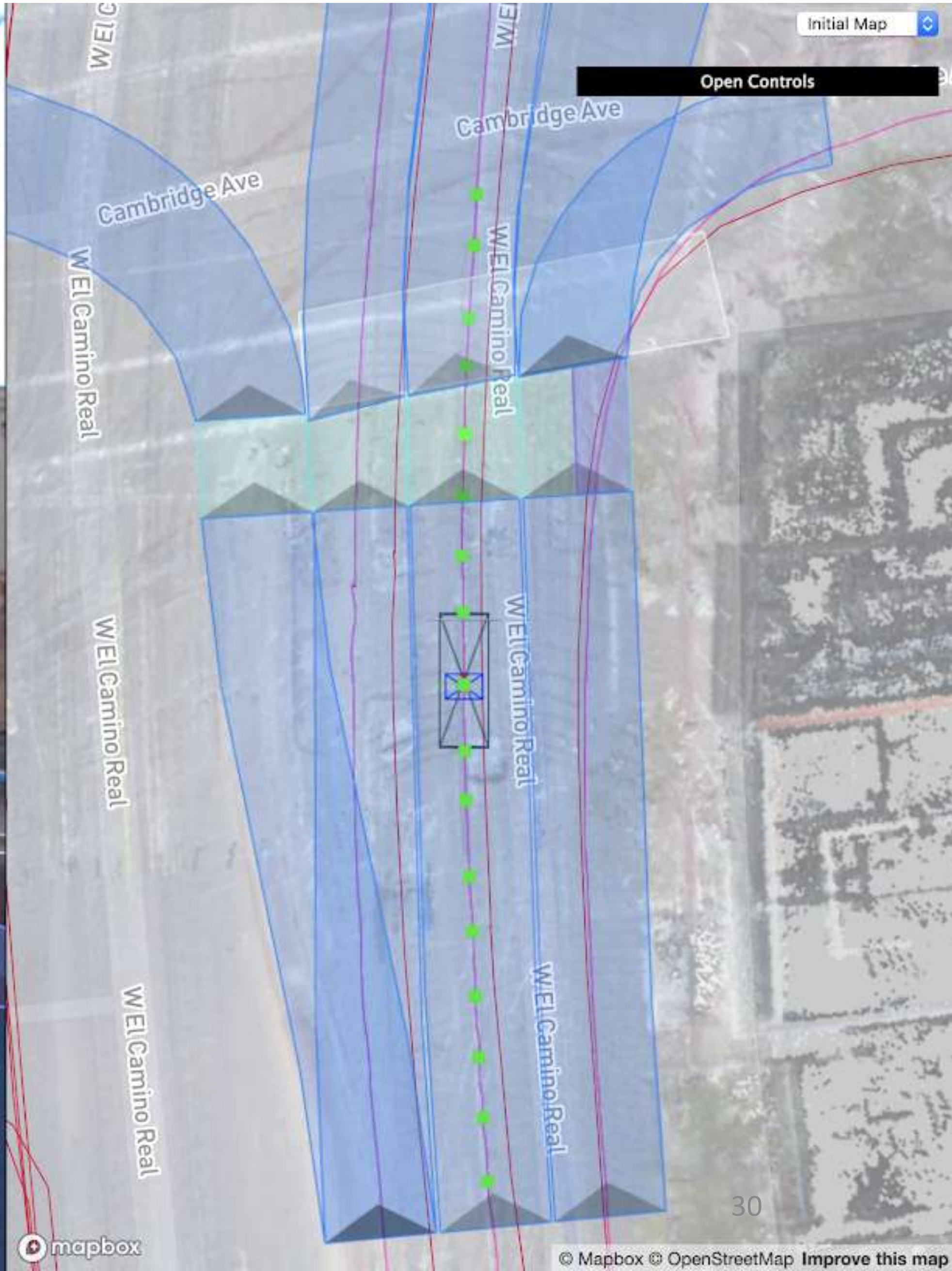


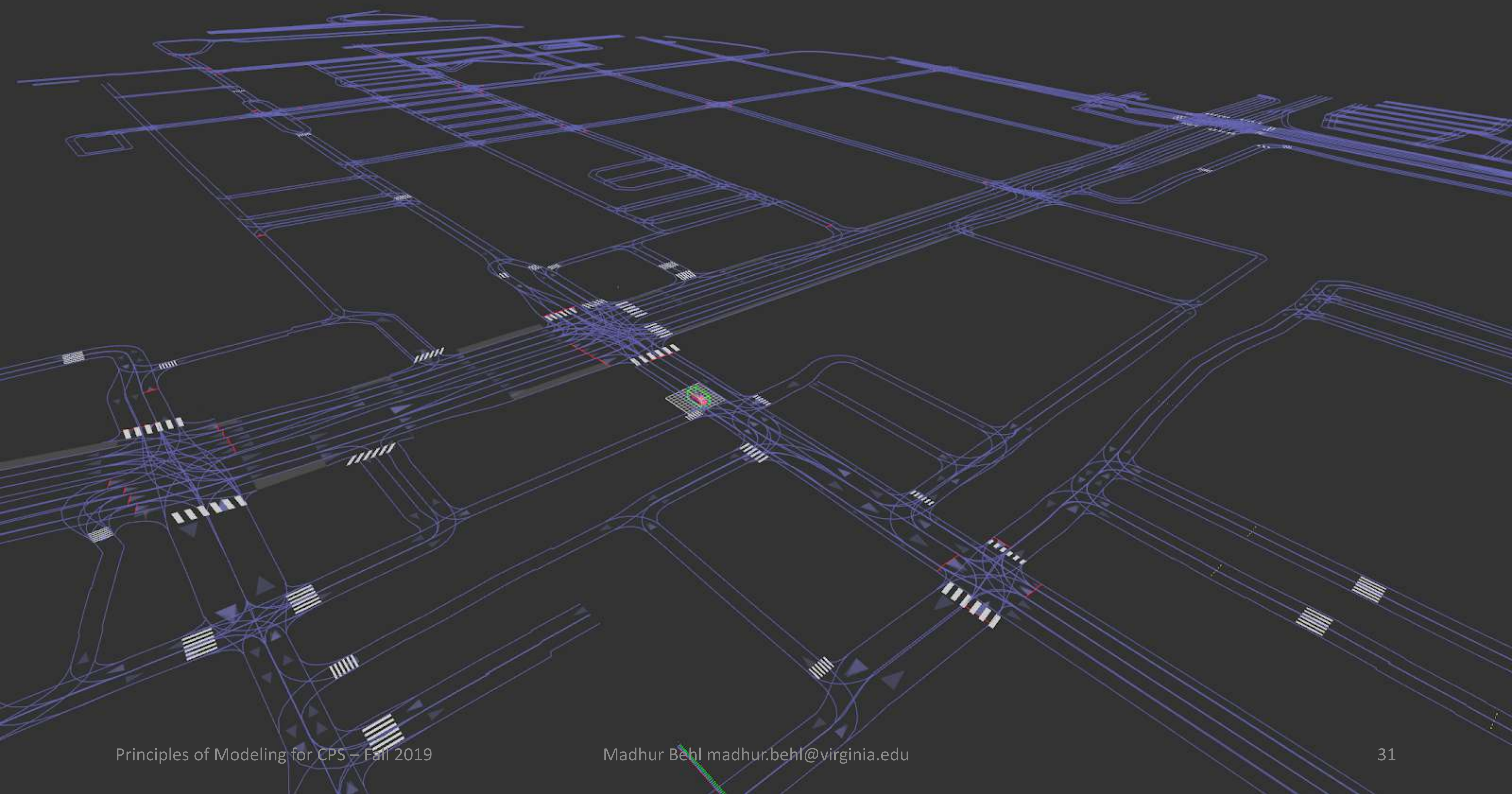
What should I do?

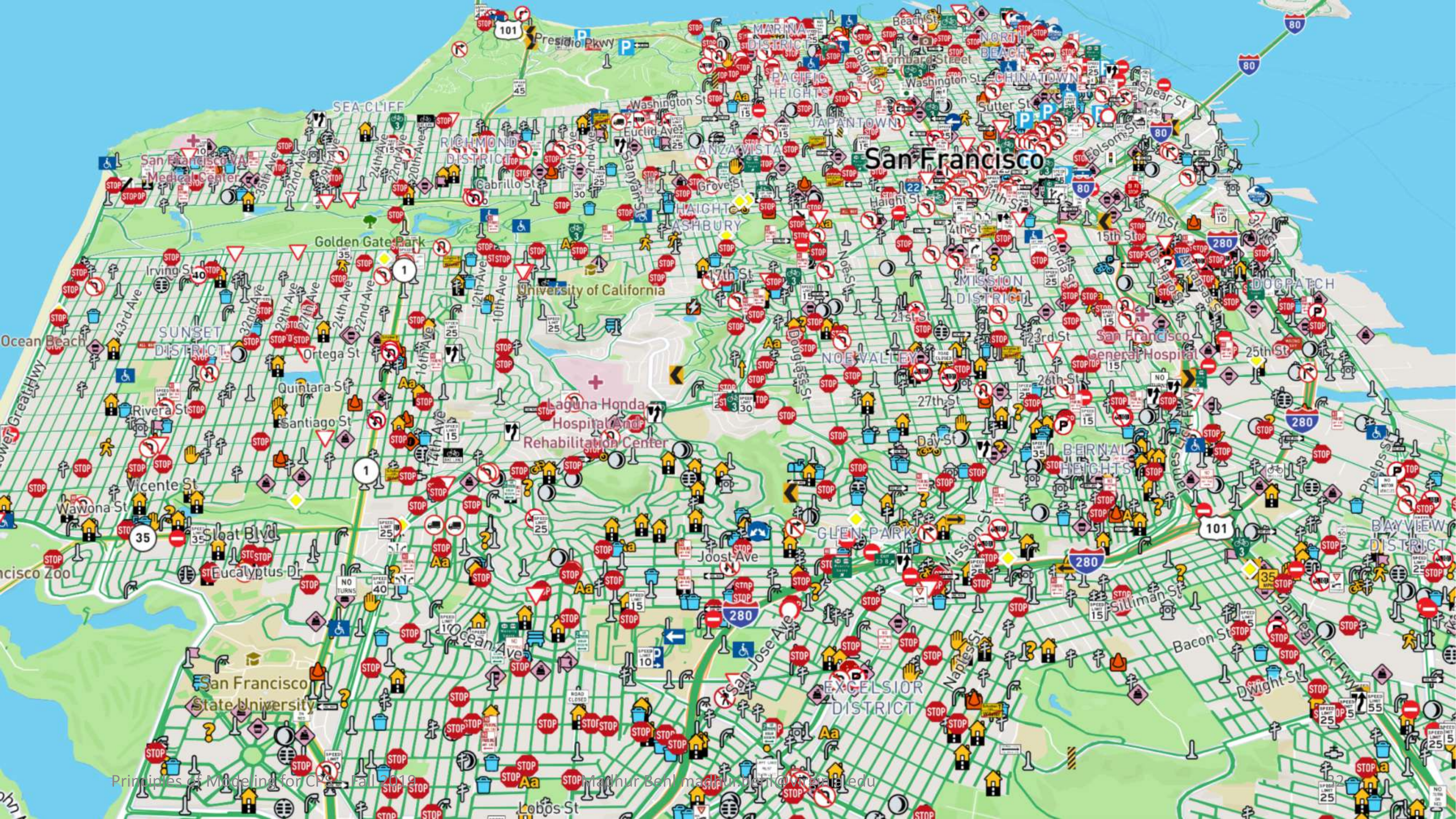
Determine the exact trajectory, speed, lane, and steering maneuvers needed to progress along the route safely



HD Maps: Localization






San Francisco
MARINA DISTRICT
NORTH BEACH
PACIFIC HEIGHTS
CHINATOWN
JAPANTOWN
ANZA VISTA
SEA CLIFF
RICHMOND DISTRICT
San Francisco VA Medical Center
HAIGHT-ASHBURY
Golden Gate Park
University of California
MISSION DISTRICT
DOGPATCH
SUNSET DISTRICT
Ocean Beach
NOE VALLEY
San Francisco General Hospital
Laguna Honda Hospital And Rehabilitation Center
BERNAL HEIGHTS
GLEN PARK
BAYVIEW DISTRICT
San Francisco State University
EXCELSIOR DISTRICT
Presidio Pkwy
Lombard Street
Washington St
Sutter St
Folsom St
Spear St
Euclid Ave
Cabrillo St
Grove St
Haight St
17th St
Irving St
Ortega St
Quintara St
Santiago St
Rivera St
Vicente St
Wawona St
Sloat Blvd
Eucalyptus Dr
Joost Ave
Ocean Ave
San Jose Ave
Mission St
Silliman St
Bacon St
Dwight St
Day St
27th St
26th St
25th St
23rd St
21st St
15th St
7th St
4th St
Lobos St
101
280
80
35
1

Localization: Scan Matching

Localization and Mapping

Where am I ?

Scene Understanding

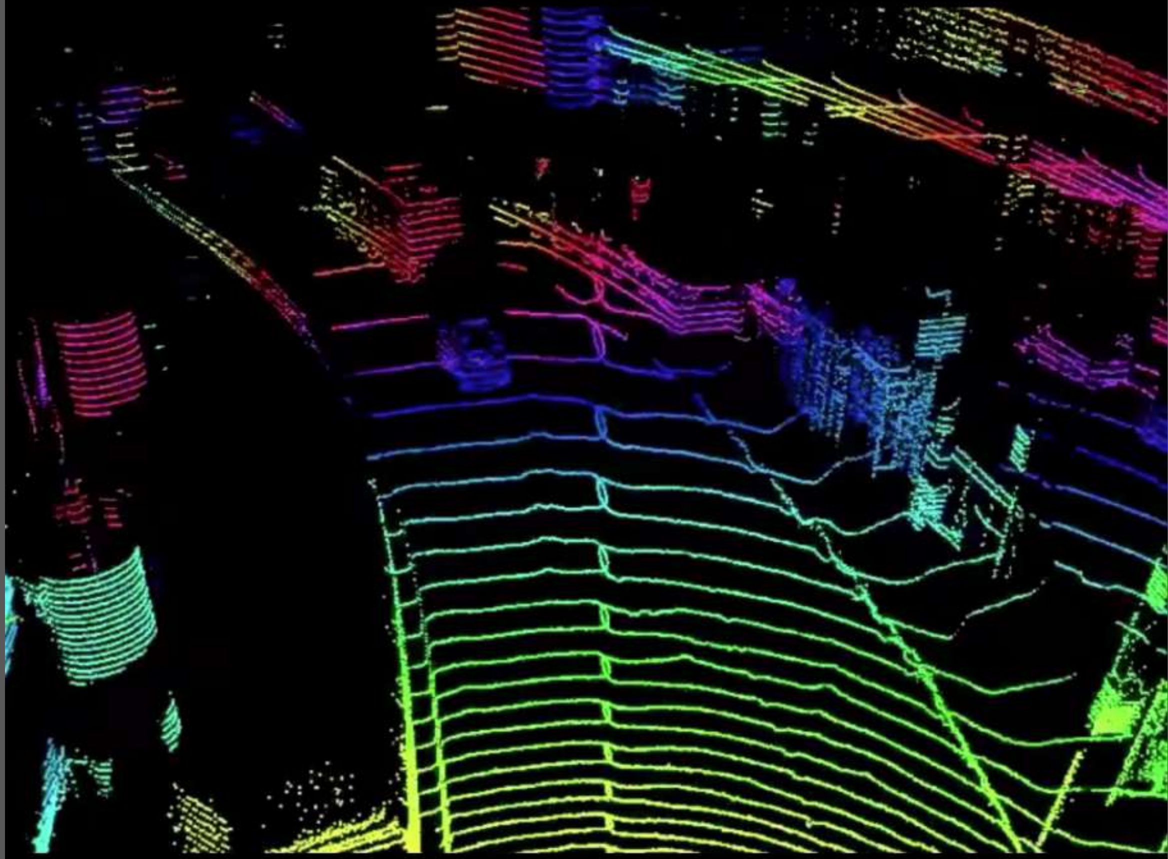
Where/who/what/why of everyone/everything else ?

Trajectory Planning and Control

Where should I go next ?
How do I steer and accelerate ?

Human Interaction

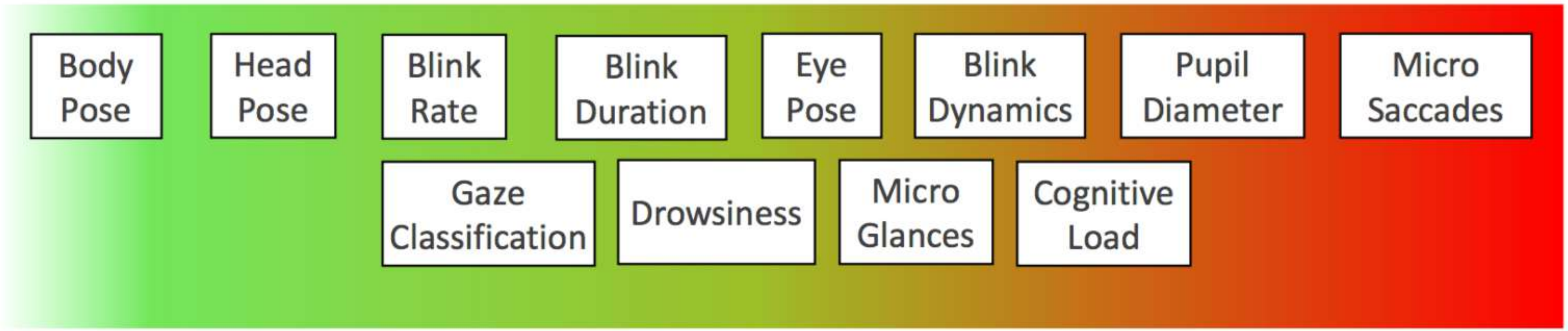
How do I convey my intent to the passenger and everyone else ?

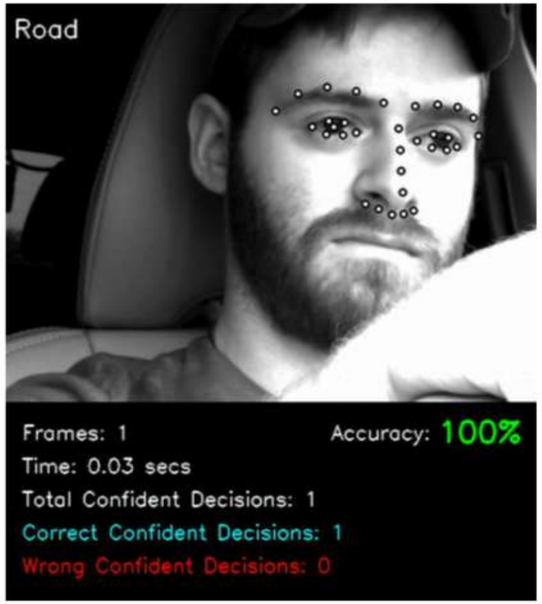


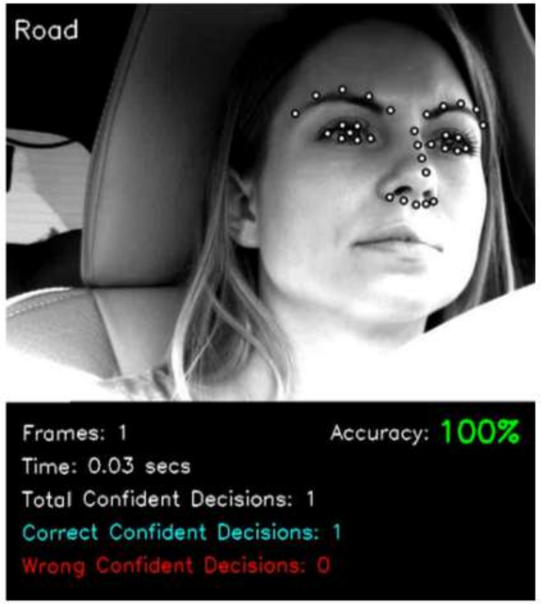
Drive State Detection:

A Multi-Resolutional View

Increasing level of detection resolution and difficulty







Localization and Mapping

Where am I ?

Scene Understanding

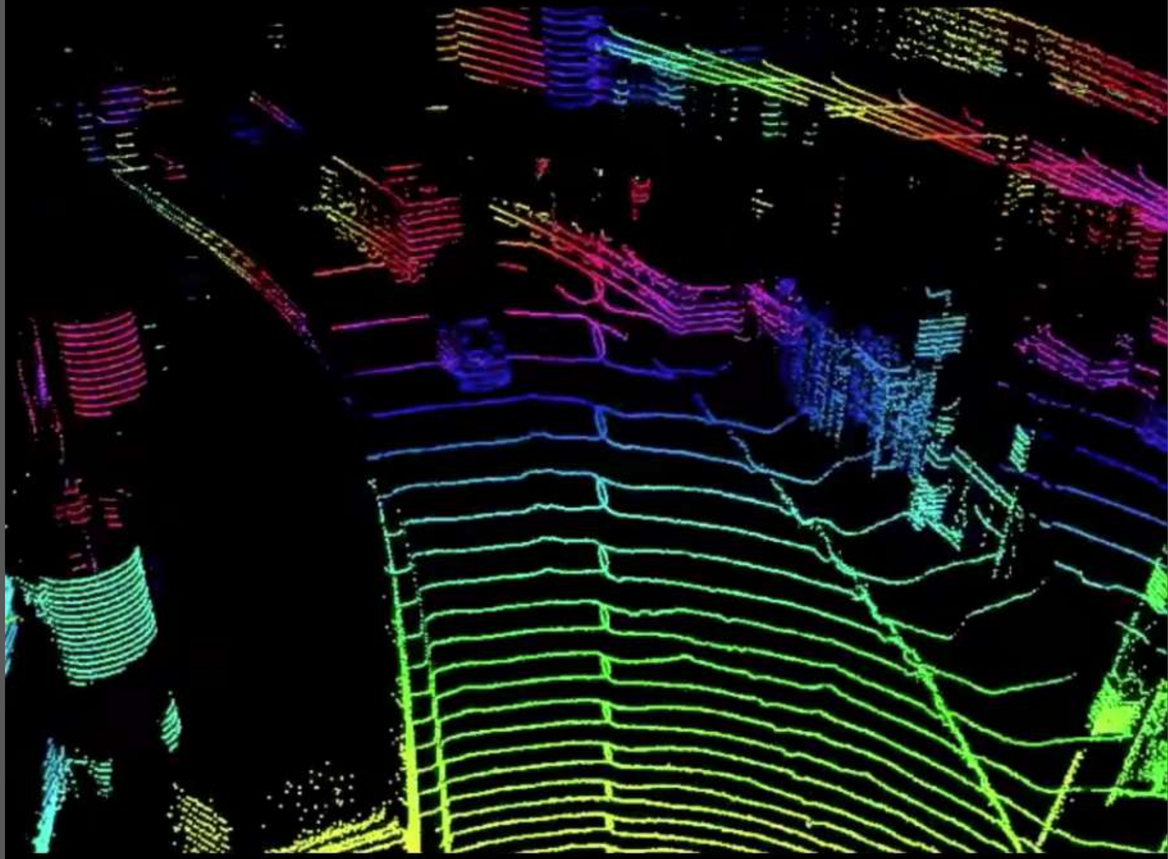
Where/who/what/why of everyone/everything else ?

Trajectory Planning and Control

Where should I go next ?
How do I steer and accelerate ?

Human Interaction

How do I convey my intent to the passenger and everyone else ?



Localization and Mapping

Where am I ?

Scene Understanding

Where/who/what/why of
everyone/everything else ?

Deep Neural Networks

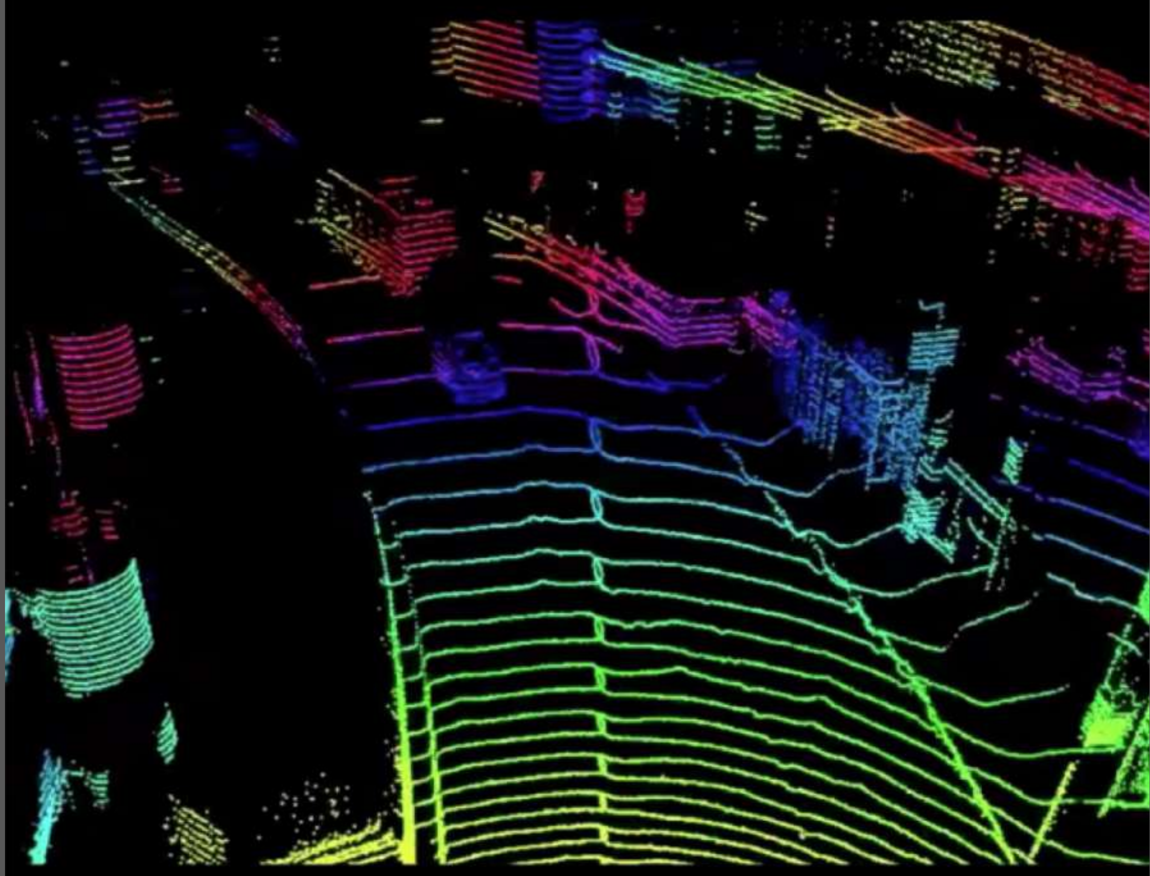
Trajectory Planning and Control

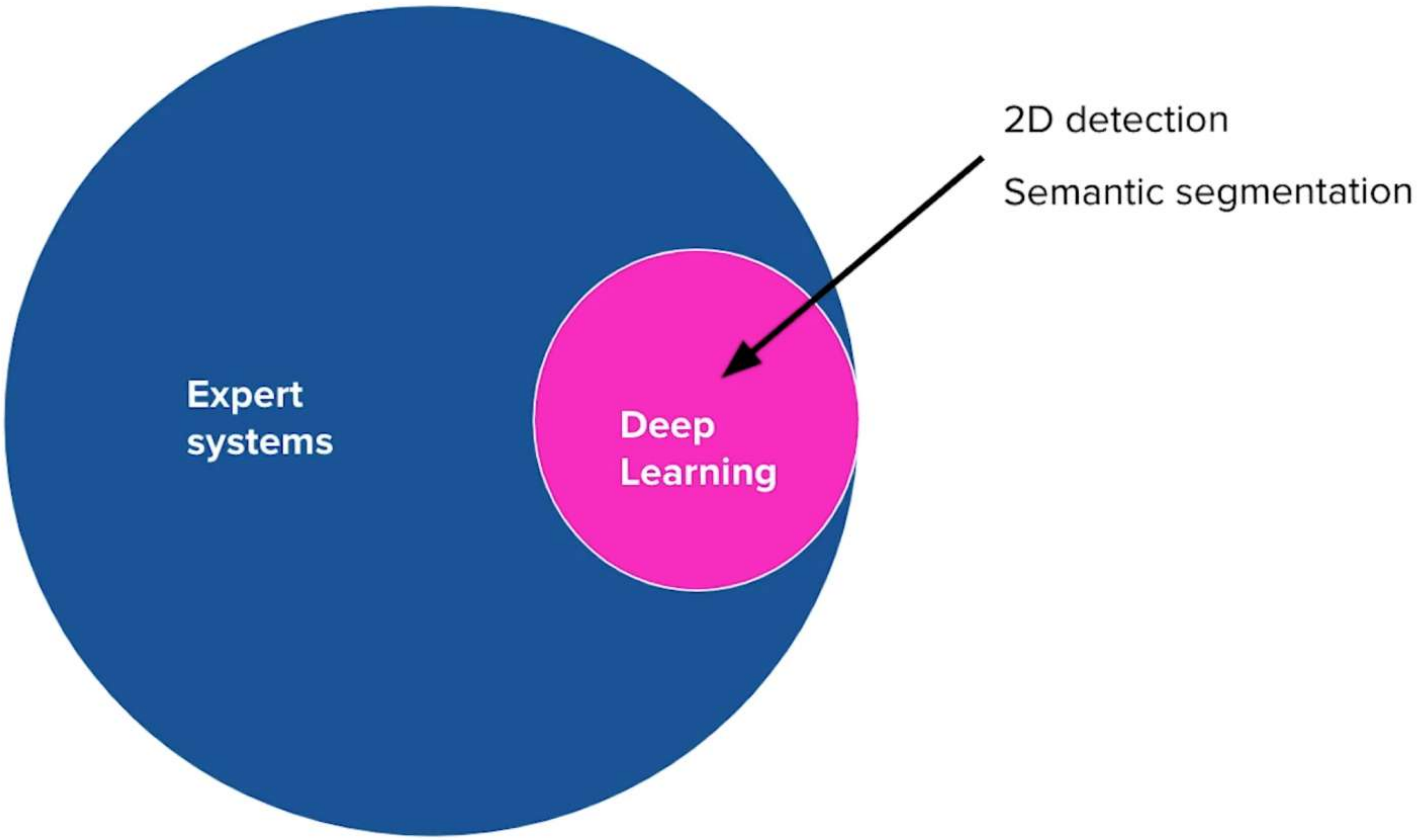
Where should I go next ?
How do I steer and accelerate ?

End-to-end learnable

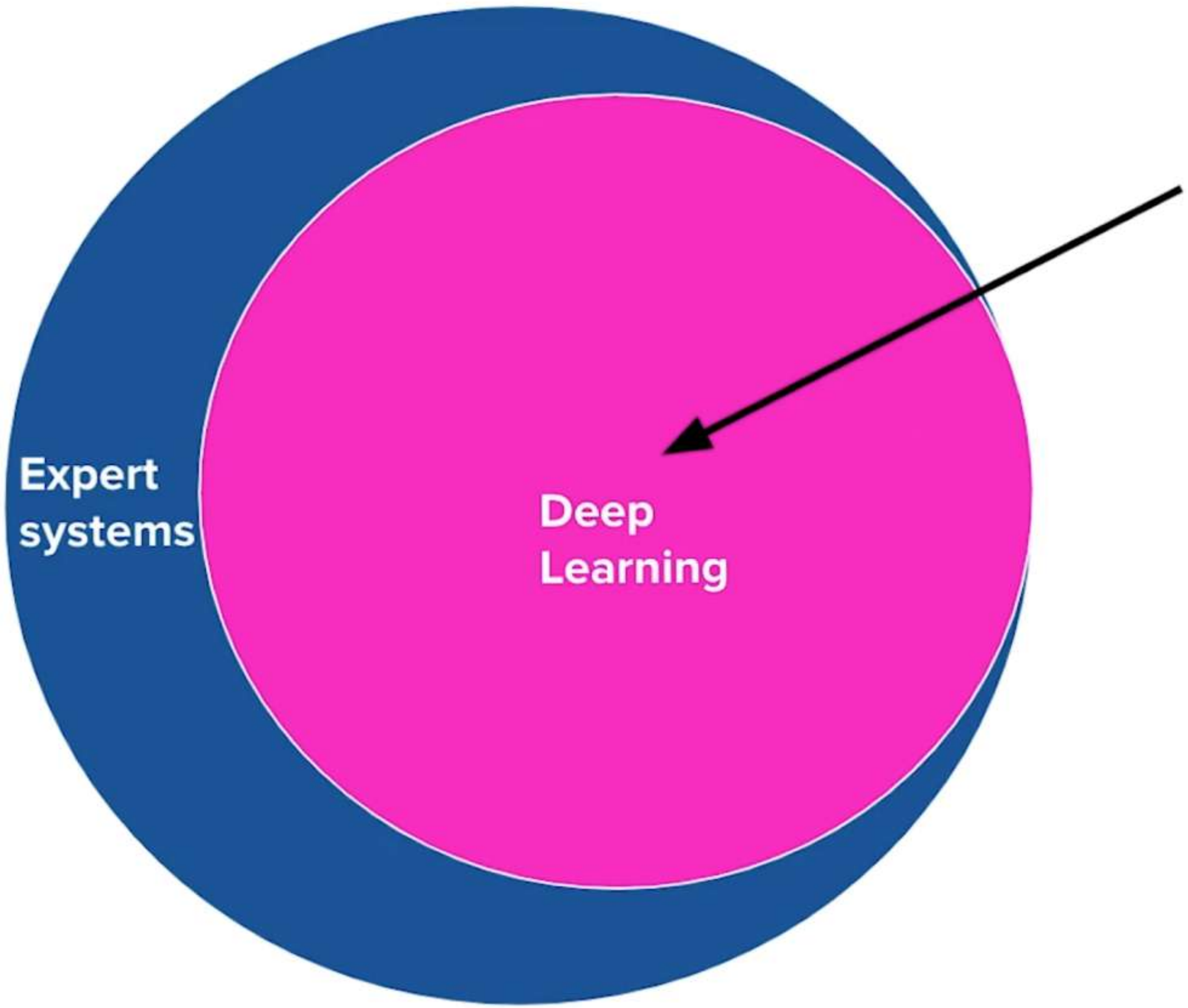
Human Interaction

How do I convey my intent to the
passenger and everyone else ?





AV Perception in 2015



2D detection

Instance segmentation

Point cloud segmentation

Tracking

Intent prediction

Monocular depth estimation

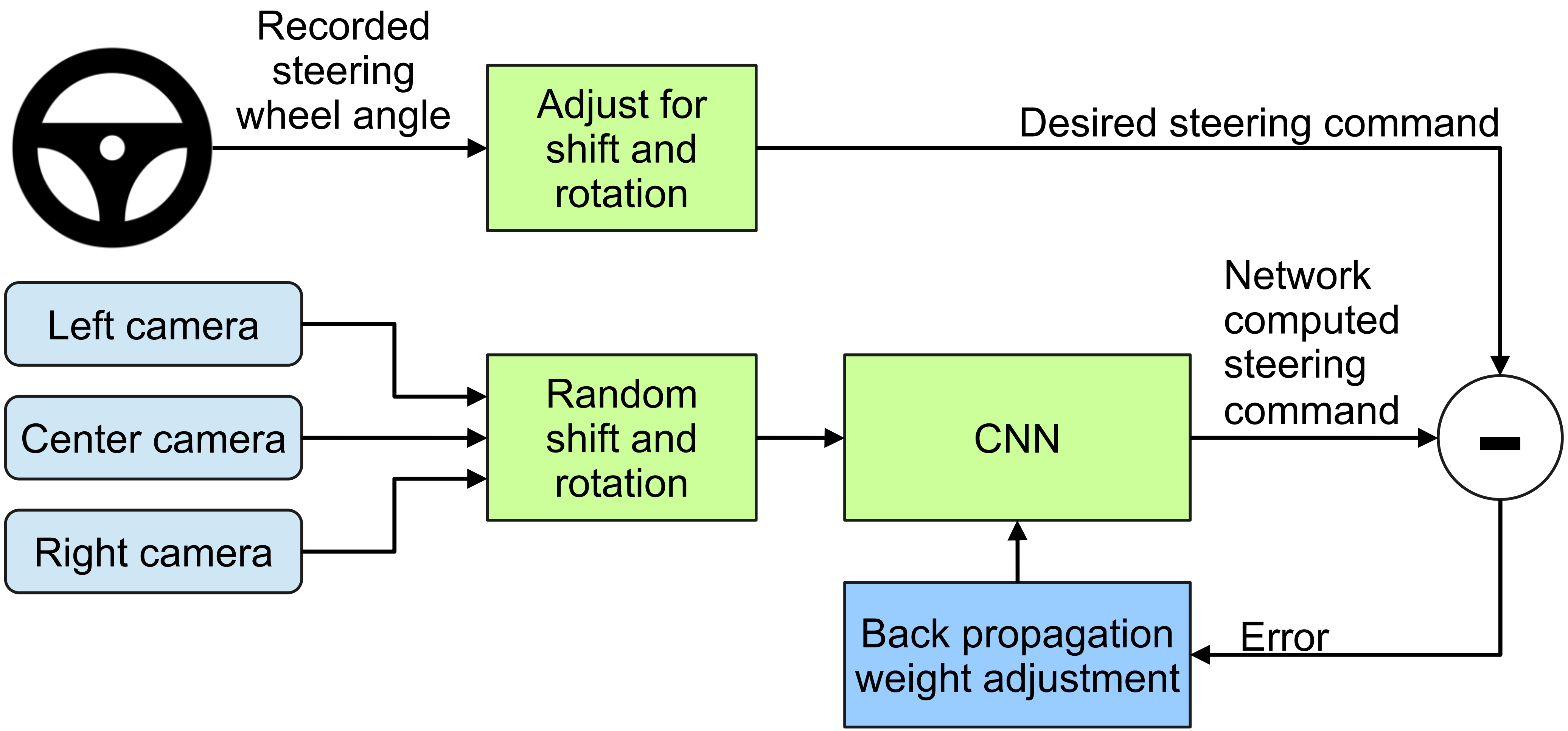
Scene flow

Anomaly detection

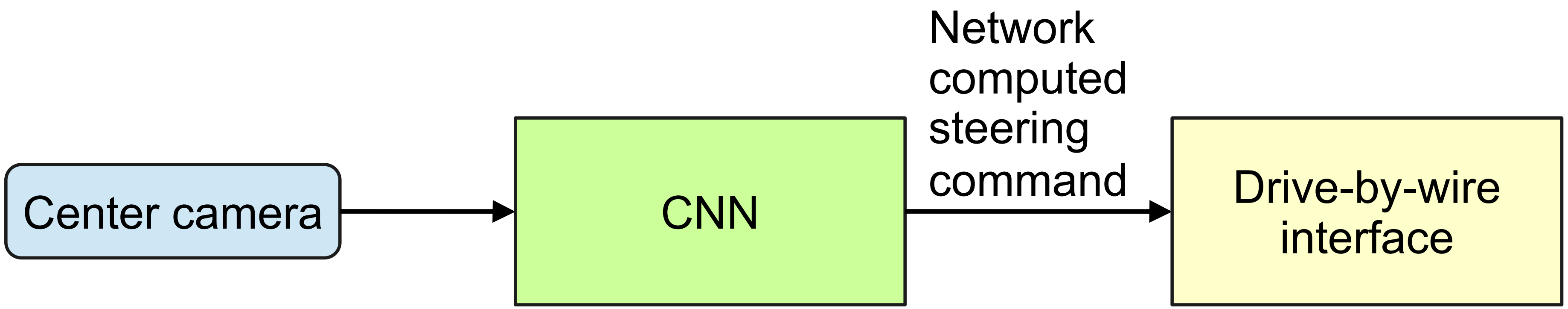
……

AV Perception today

End-to-End Driving: PilotNET

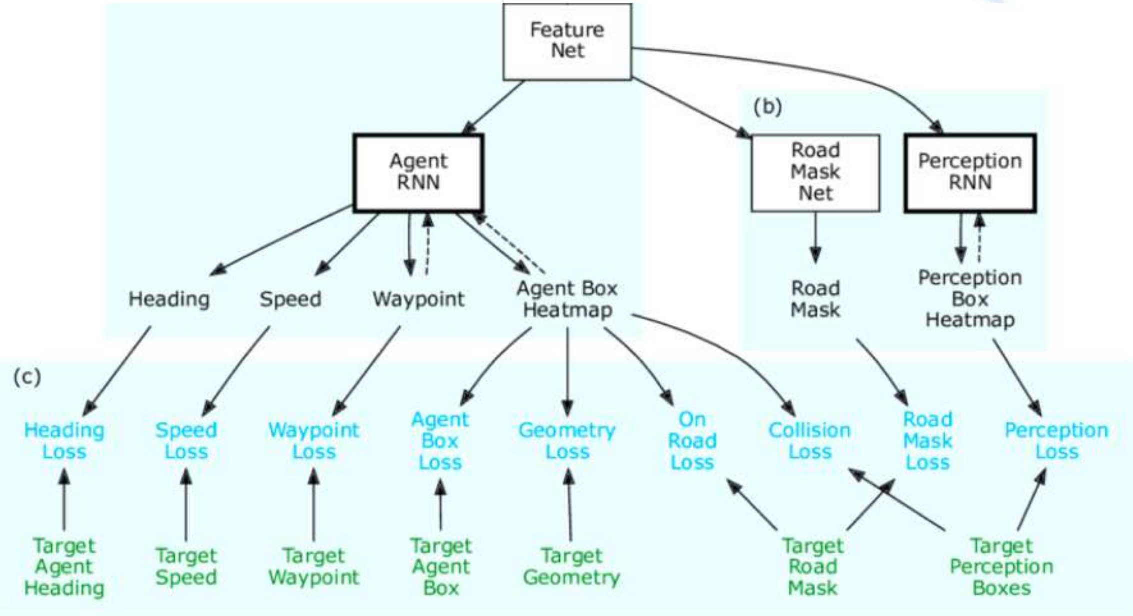


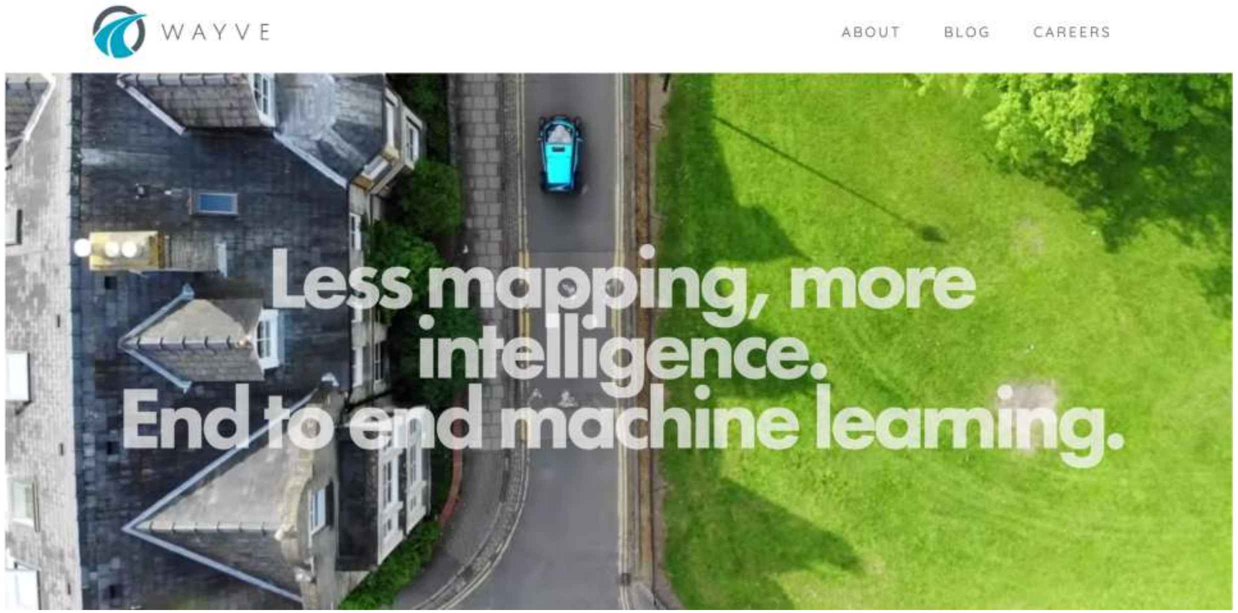
With a single front-facing camera



ChauffeurNet

A Deep Learned Driving Network





Machine intelligence is largely about training data.

When's a pedestrian not a pedestrian? When it's a decal.

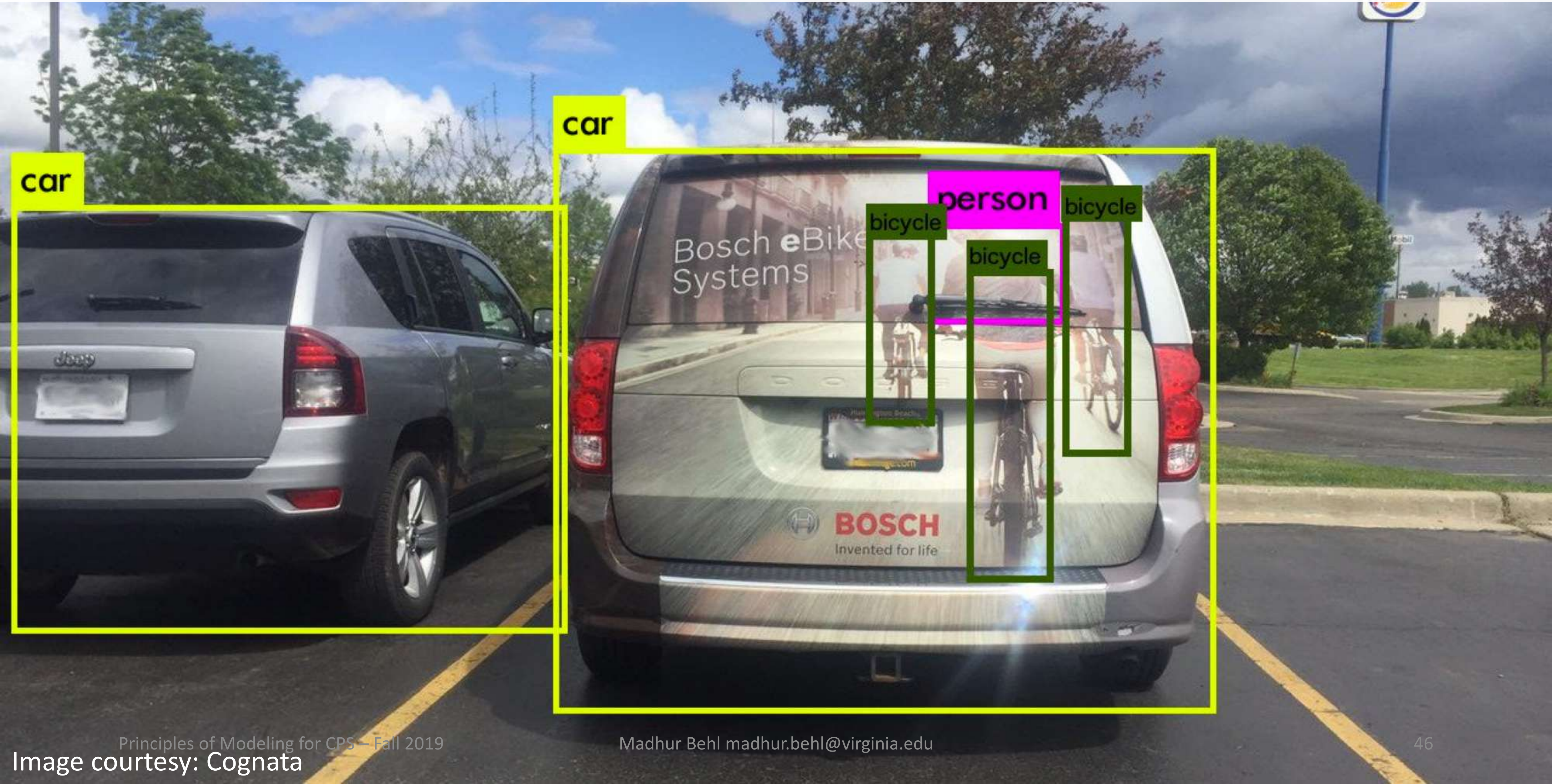


Image courtesy: Cognata

One car ? or Multiple cars ?

Ramen Noodle place or Do Not Enter Sign ?


下鳥羽城ノ越町
20m
MENU
AUDIO
今日
明日
16:10
12°C
77.5km
000160km
x1000 r/min
km/h


今日
明日
16:10
12/5
60
9/3
40
12℃
77.5km
000160km
50

There is a bus right next to you!!


EXIT 23 B
30 BY-PASS
West
Killingsworth St
EXIT ONLY
1
0MPH
20/01/2018 16:03:15

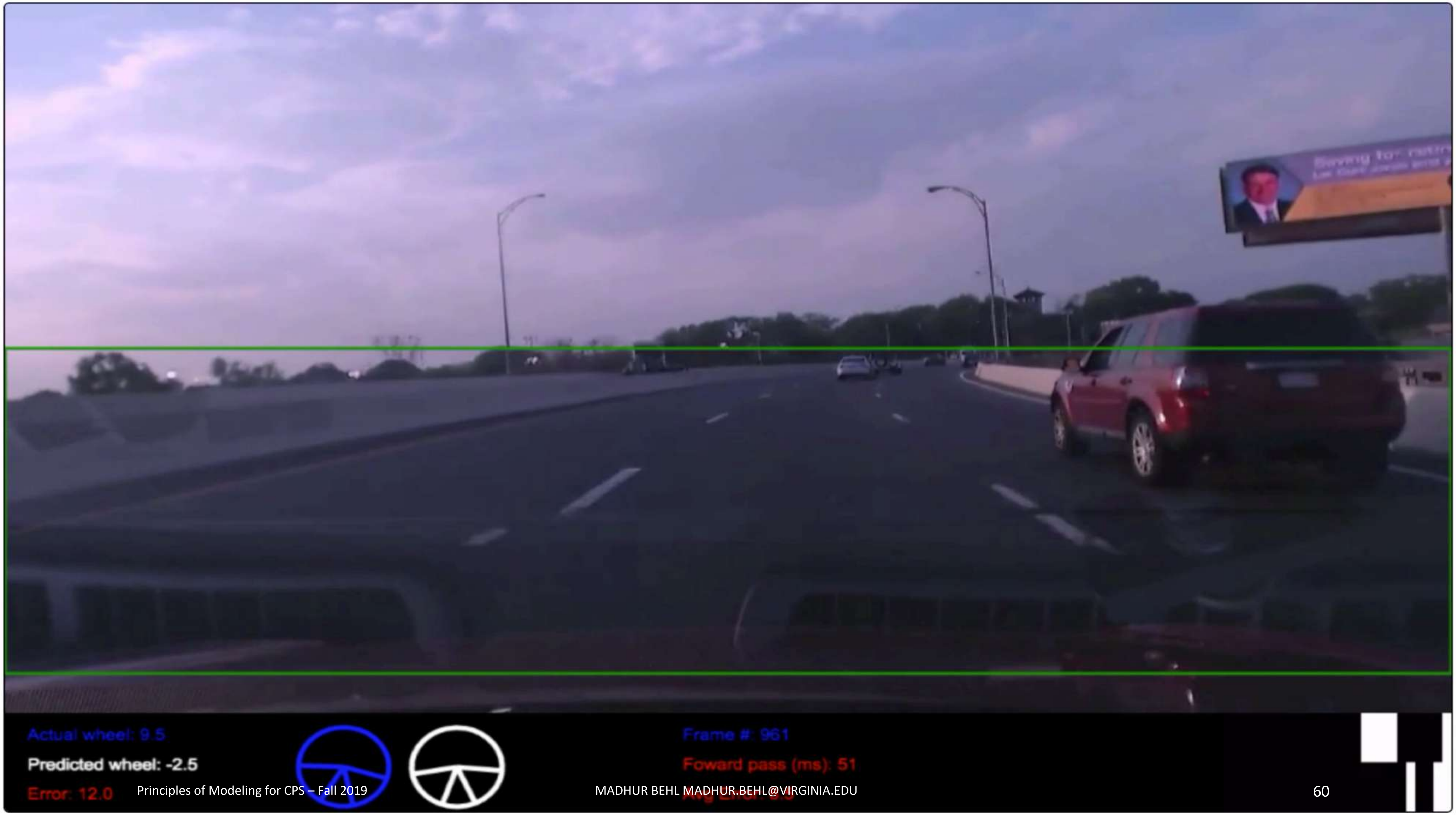

ENTERING
Richmond
NEXT 4 EXITS

How can we ensure that an autonomous vehicle drives safety upon encountering an unusual traffic situation ?

DeepRacing AI

UVA ENGINEERING LINK LAB

How can CNNs help us drive ?



Actual wheel: 9.5
Predicted wheel: -2.5
Error: 12.0
Frame #: 961
Foward pass (ms): 51

Autonomous Driving: End-to-End

End to End Learning for Self-Driving Cars

Mariusz Bojarski
NVIDIA Corporation
Holmdel, NJ 07735

Davide Del Testa
NVIDIA Corporation
Holmdel, NJ 07735

Daniel Dworakowski
NVIDIA Corporation
Holmdel, NJ 07735

Bernhard Firner
NVIDIA Corporation
Holmdel, NJ 07735

Beat Flepp
NVIDIA Corporation
Holmdel, NJ 07735

Prasoon Goyal
NVIDIA Corporation
Holmdel, NJ 07735

Lawrence D. Jackel
NVIDIA Corporation
Holmdel, NJ 07735

Mathew Monfort
NVIDIA Corporation
Holmdel, NJ 07735

Urs Muller
NVIDIA Corporation
Holmdel, NJ 07735

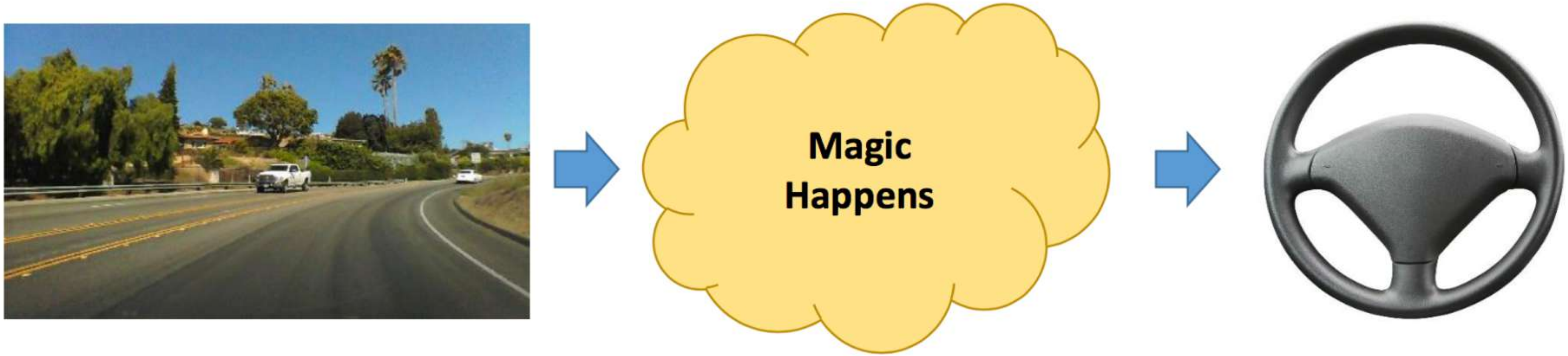
Jiakai Zhang
NVIDIA Corporation
Holmdel, NJ 07735

Xin Zhang
NVIDIA Corporation
Holmdel, NJ 07735

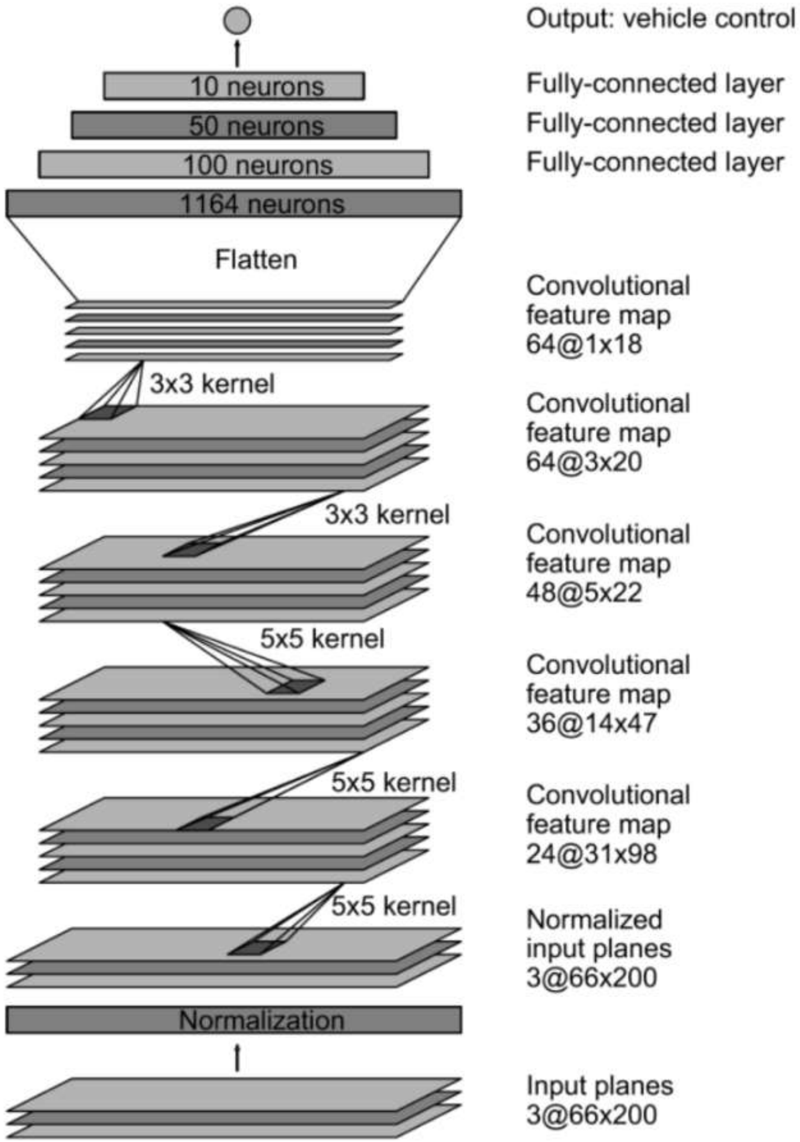
Jake Zhao
NVIDIA Corporation
Holmdel, NJ 07735

Karol Zieba
NVIDIA Corporation
Holmdel, NJ 07735

Autonomous Driving: End-to-End

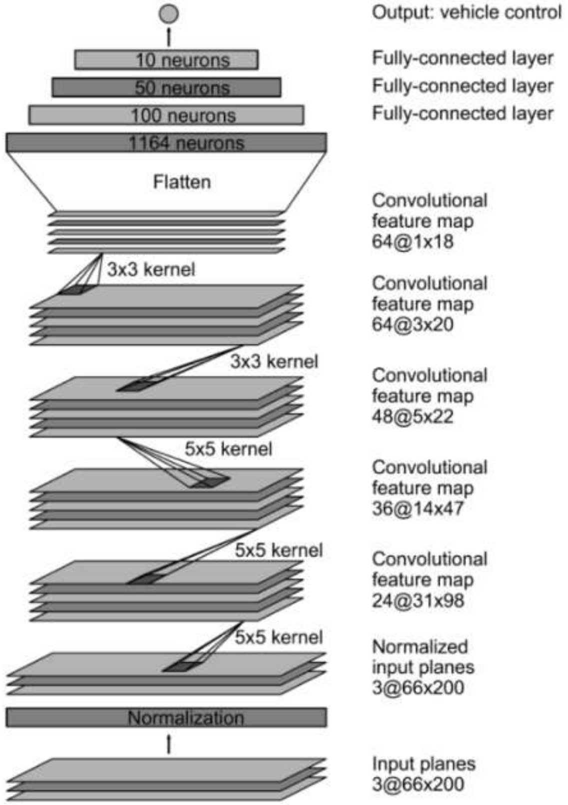


Autonomous Driving: End-to-End



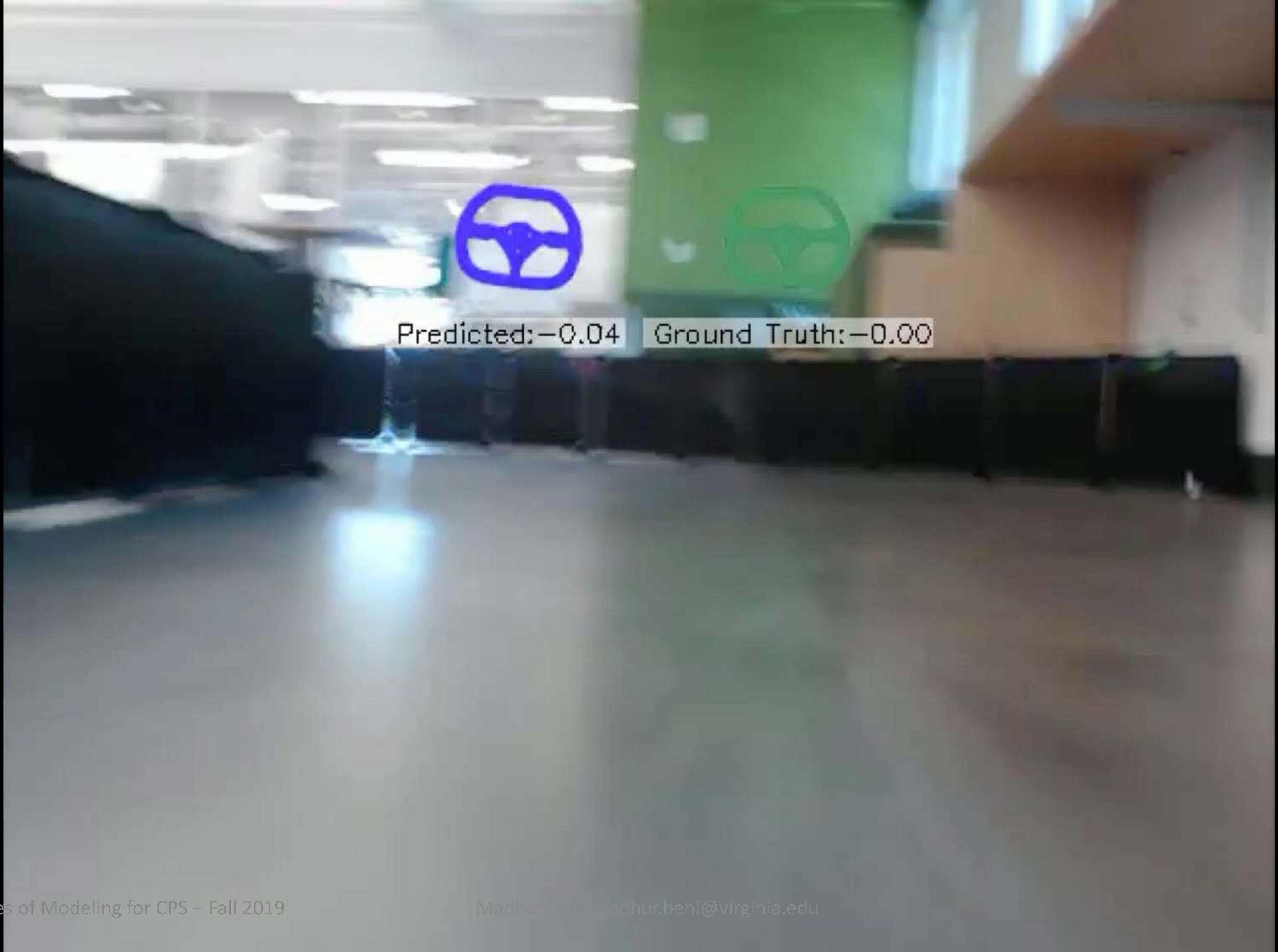
- 9 layers
  - 1 normalization layer
  - 5 convolutional layers
  - 3 fully connected layers
- 27 million connections
- 250 thousand parameters

Autonomous Driving: End-to-End



F1/10 FPV Driving


Autonomous racing in progress !
CAUTION

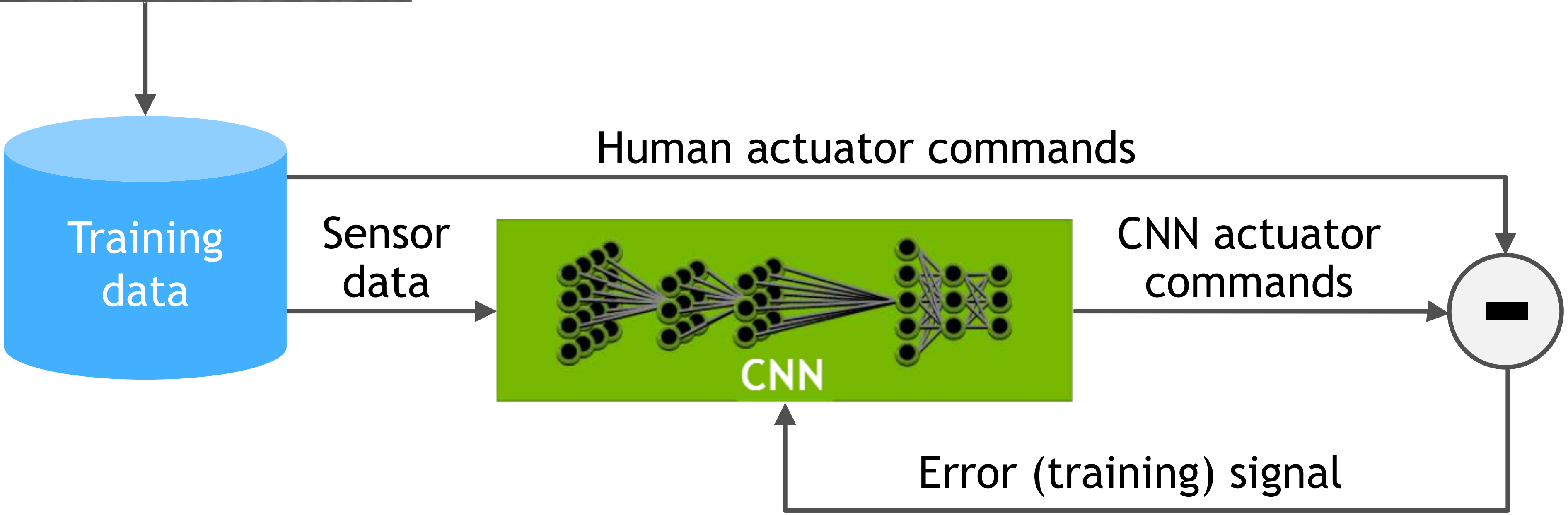

Predicted:-0.04
Ground Truth:-0.00

THE BASIC IDEA

Learn from human drivers

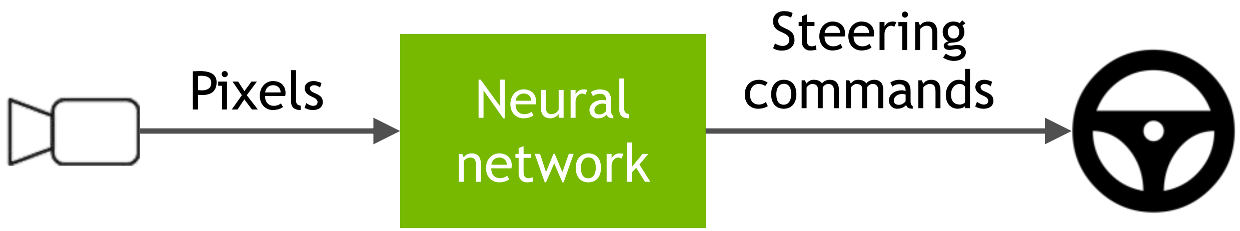
Record data from lots of humans driving their cars:

- Sensor data
- Actuator data



EARLY EXAMPLES

Of end-to-end learning



ALVINN, CMU, late 80es
(Pomerleau et Al.)

Lane following with a small 2-layer fully connected network and low-resolution video input

30x32 pixel

DAVE, Net-Scale/NYU, 2004
(LeCun et Al.)

Off-road obstacle avoidance using a convolutional network (ConvNet)

149x48 pixel

TRAINING EXAMPLES

Label: turn right

Label: go straight

Label: turn right

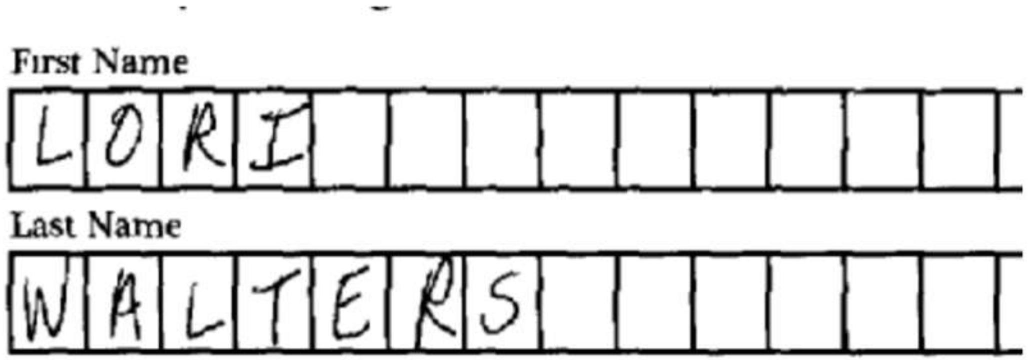
Label: turn left

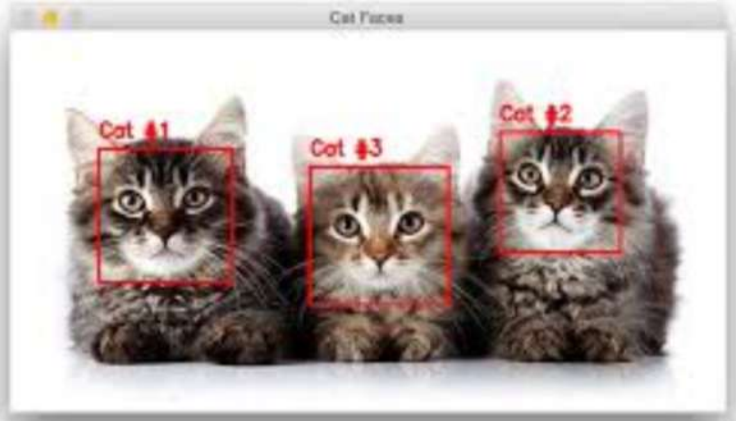
225K images

Machine Learning

– Machine Learning is the ability to teach a computer without explicitly programming it

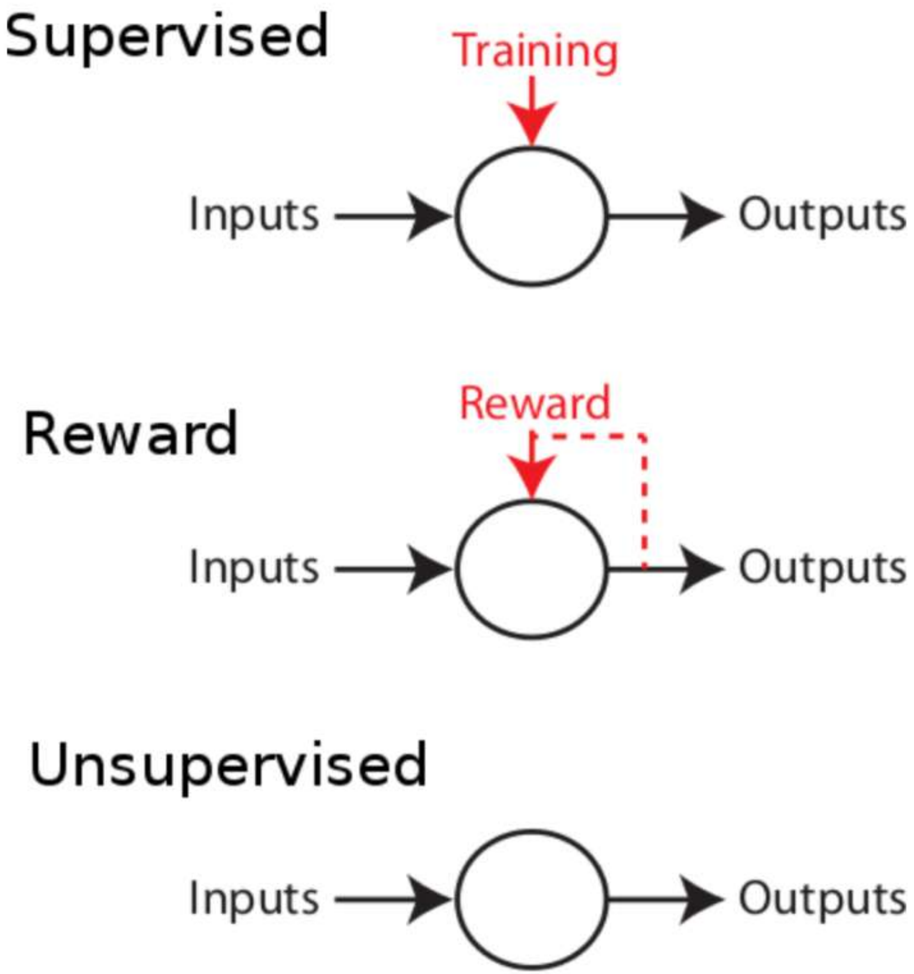
– Examples are used to train computers to perform tasks that would be difficult to program



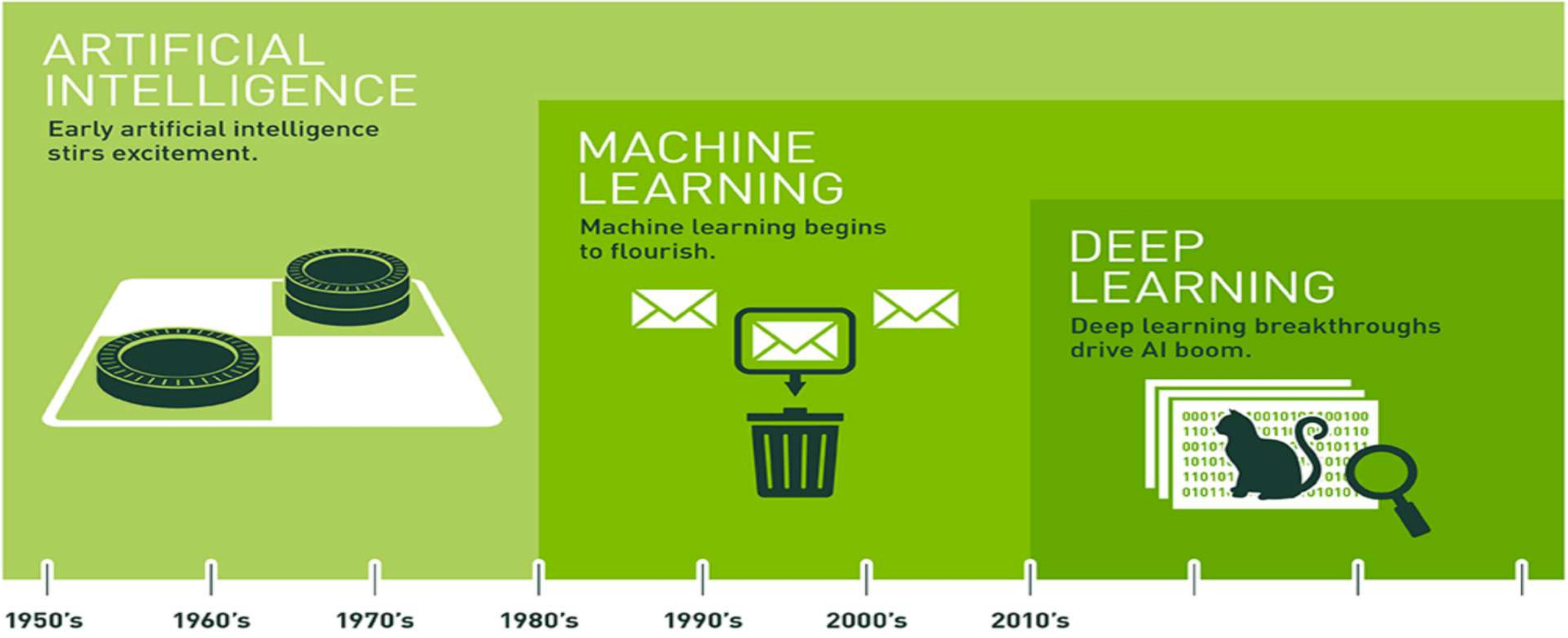


Types of machine Learning

- Supervised Learning
  - Training data is labeled
  - Goal is correctly label new data
- Reinforcement Learning
  - Training data is unlabeled
  - System receives feedback for its actions
  - Goal is to perform better actions
- Unsupervised Learning
  - Training data is unlabeled
  - Goal is to categorize the observations



Capability of Machine to imitate intelligent behavior



Supervised learning setup

Inputs (AKA features) - real-valued vectors of data

e.g. Image pixels, audio spectrograms, character sequences

Outputs (AKA labels) - real-valued or categorical “truth” vectors

e.g. class labels for images, audio transcription, sentiment

Training data - many samples of input-output pairs

Score function (AKA model)

A function that predicts the output given an input

Example: linear regression

Data

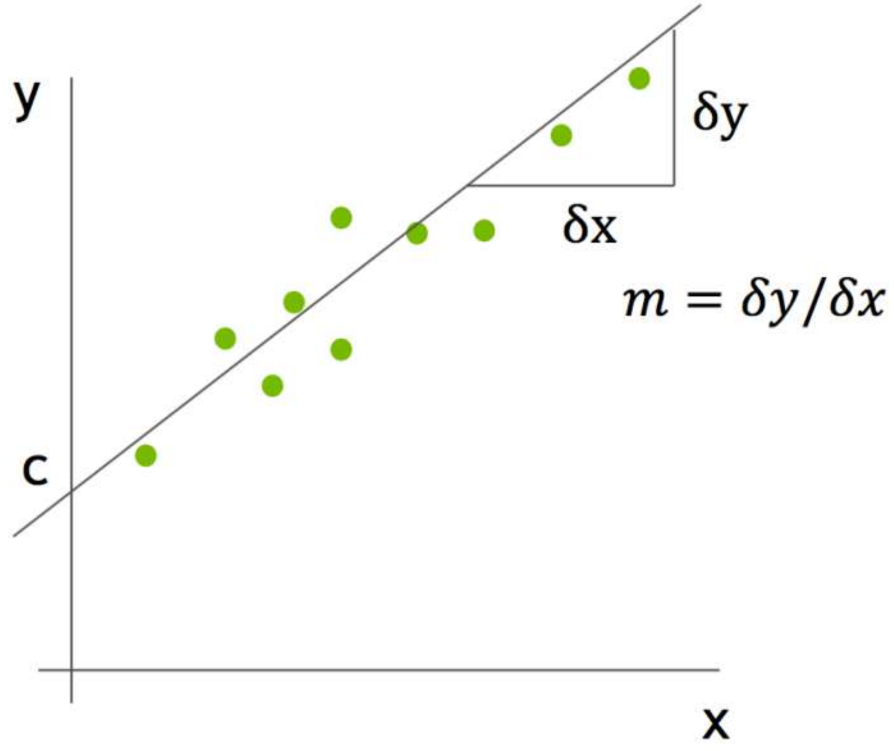
$$y_i = mx_i + c$$

Predicted output

Slope

Intercept

Together, m and c are called the **model parameters**



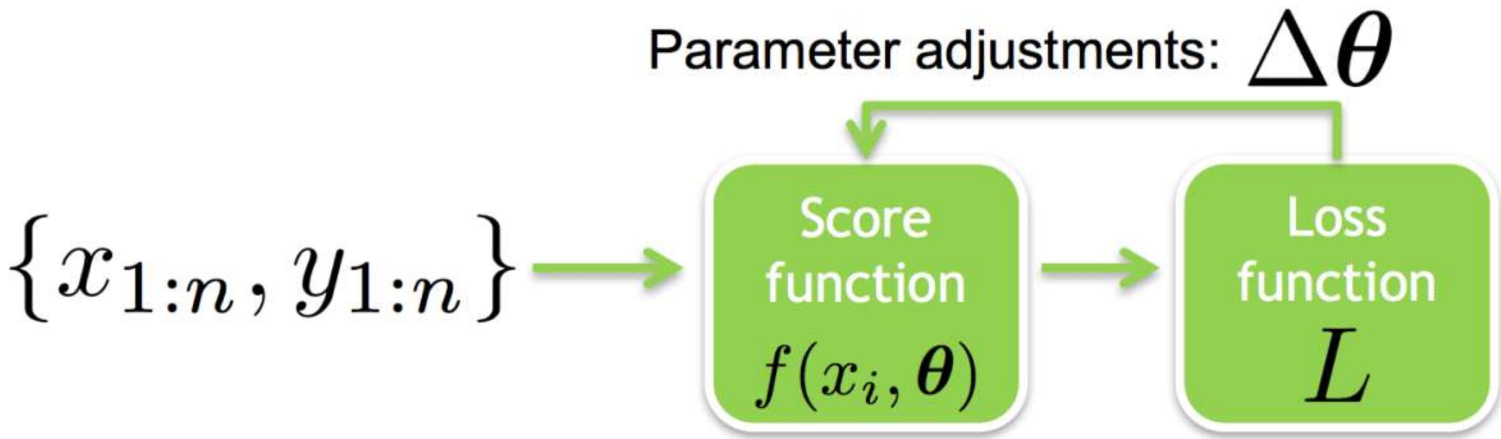
Supervised learning

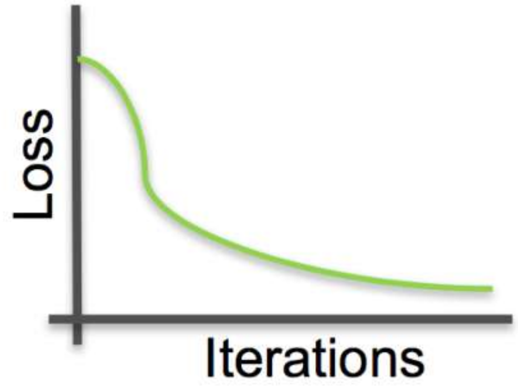
How do we do this?

Repeatedly feed training data into a learning algorithm

Iteratively modify the model parameters to optimize (e.g. minimize) the loss function

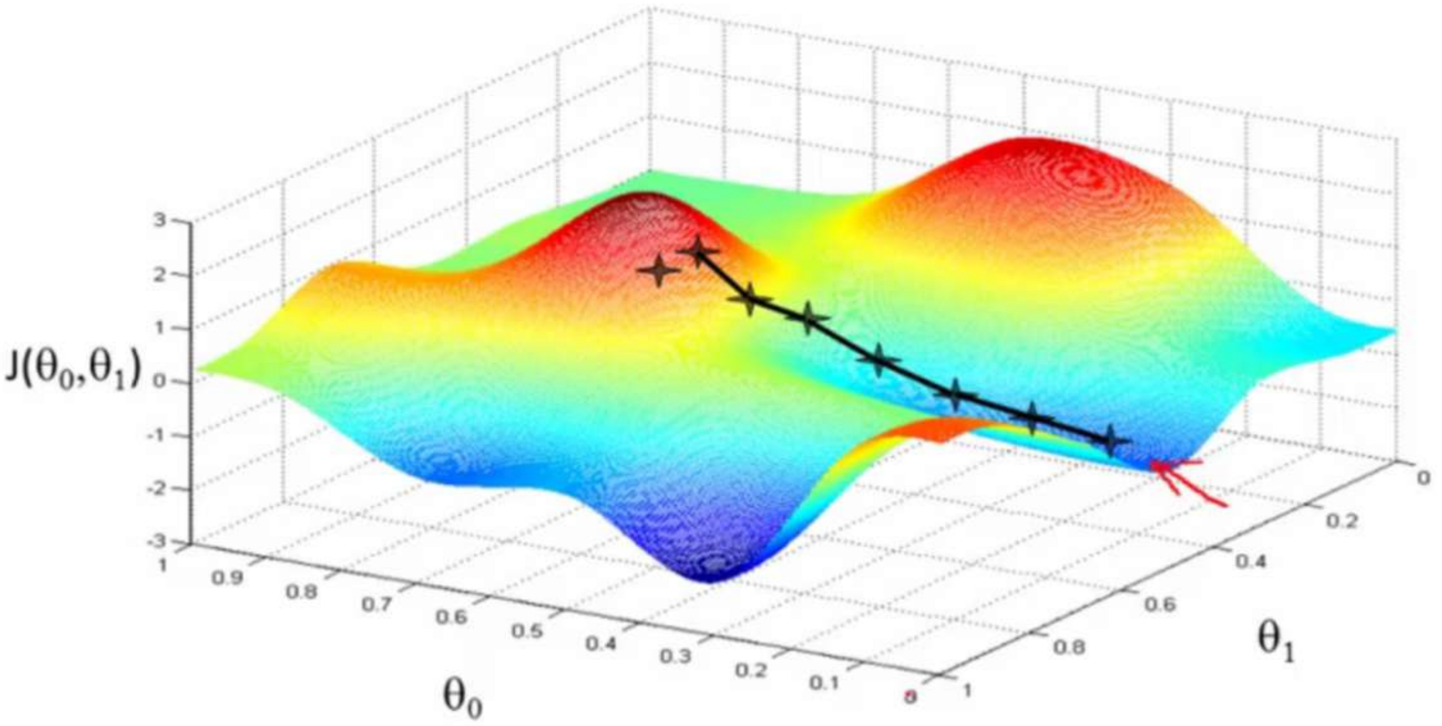
Repeat until the model is “good enough”

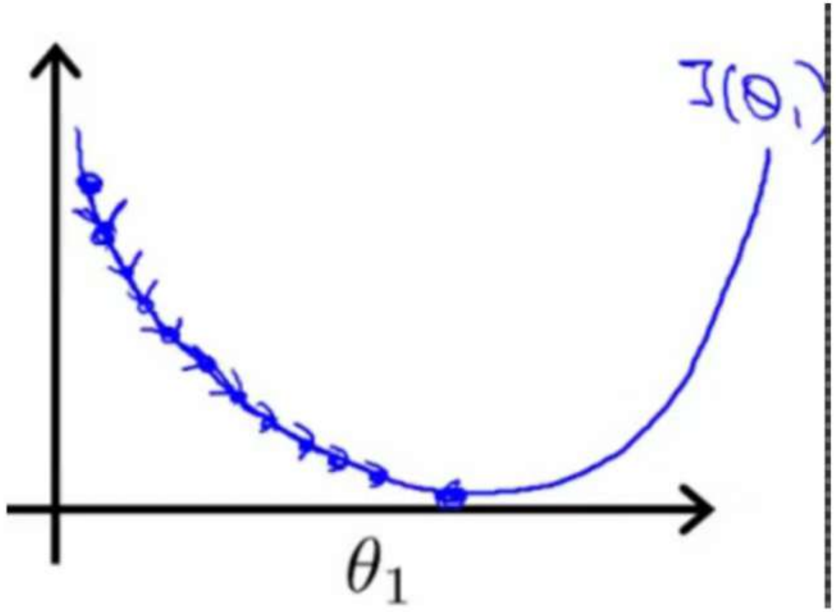




Gradient descent

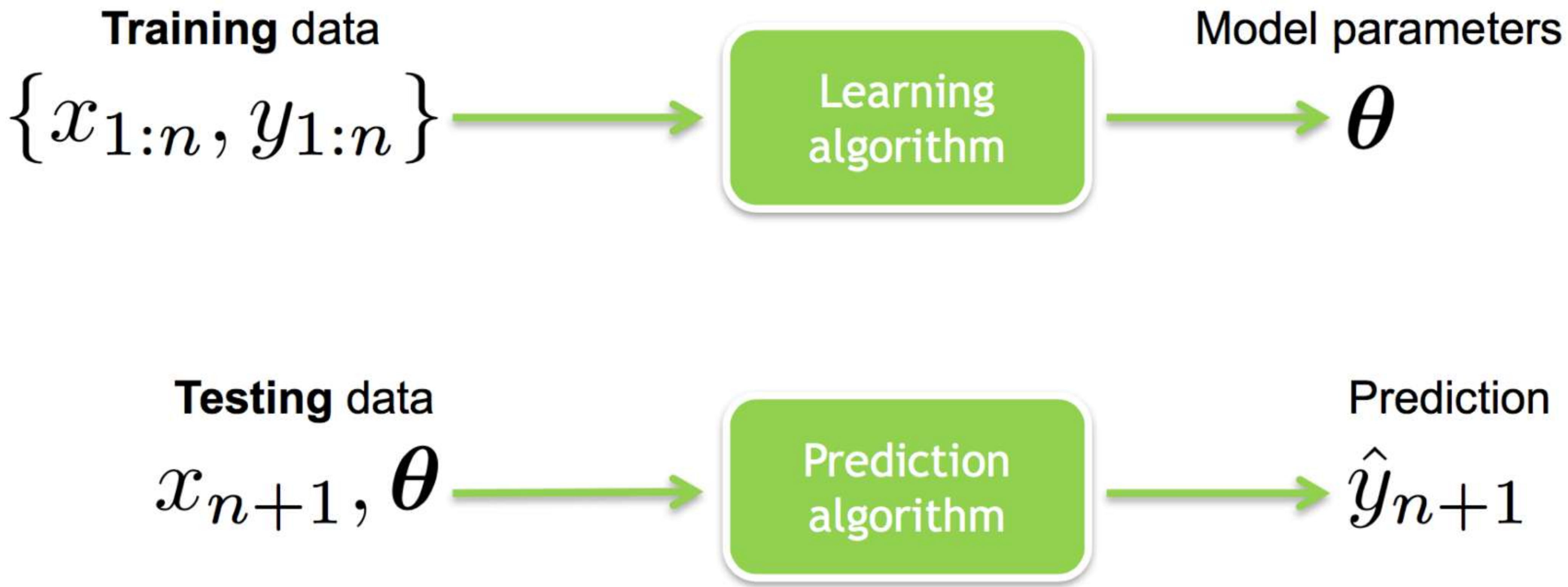
Finding the Optimal Parameters for our Hypothesis



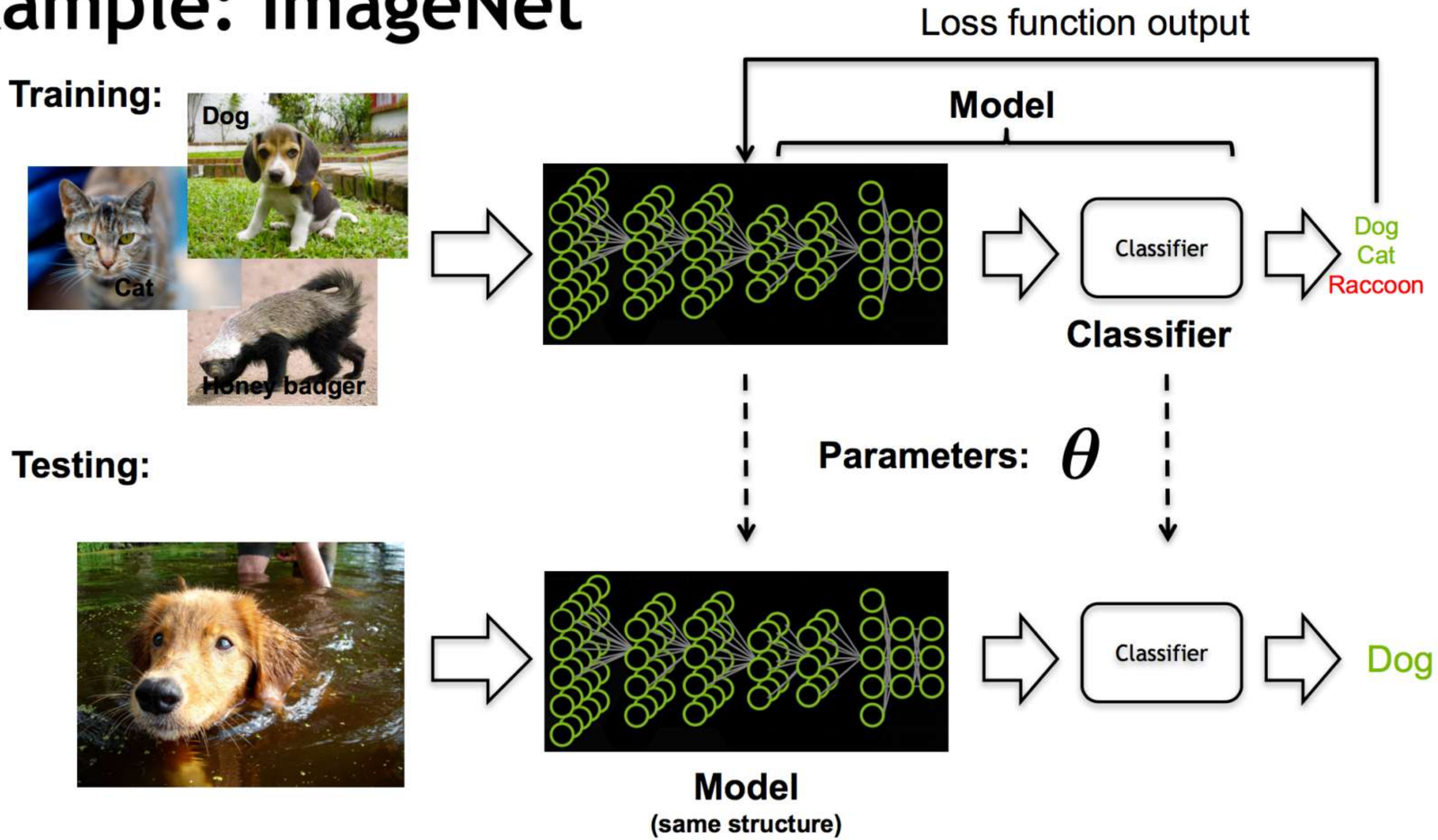


Supervised learning
Why do we do this?

Given the **model** we can take previously unseen inputs and predict the corresponding output. We call this **testing** or **deployment.**



Example: ImageNet



Deep Learning success

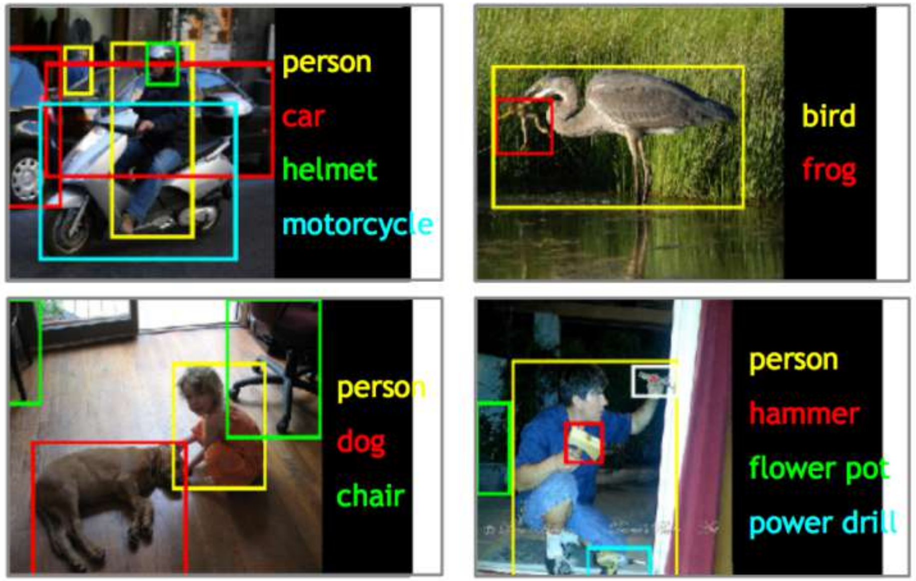
Object classification and localization in images

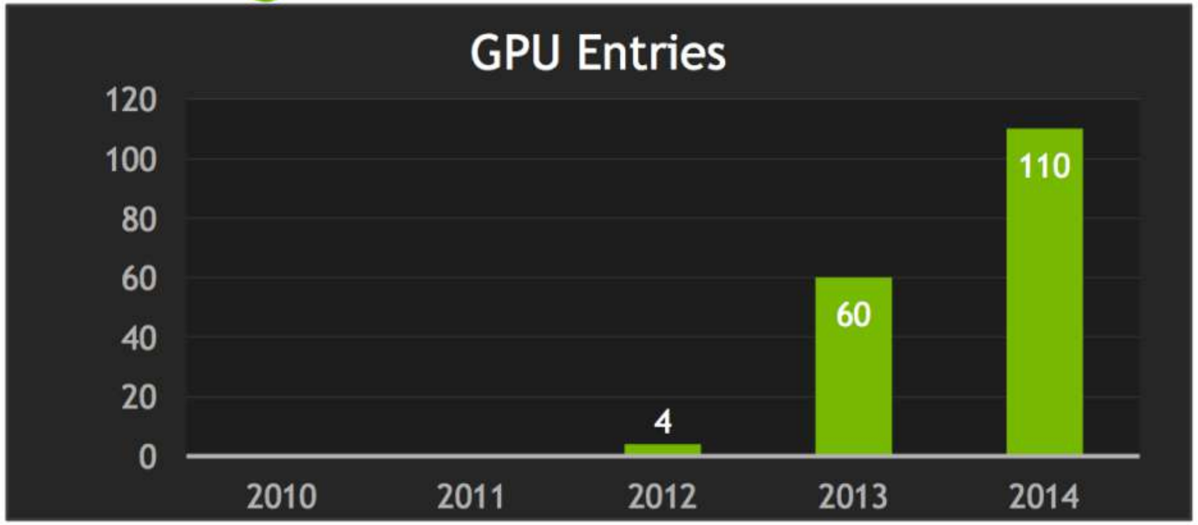
Image Recognition Challenge

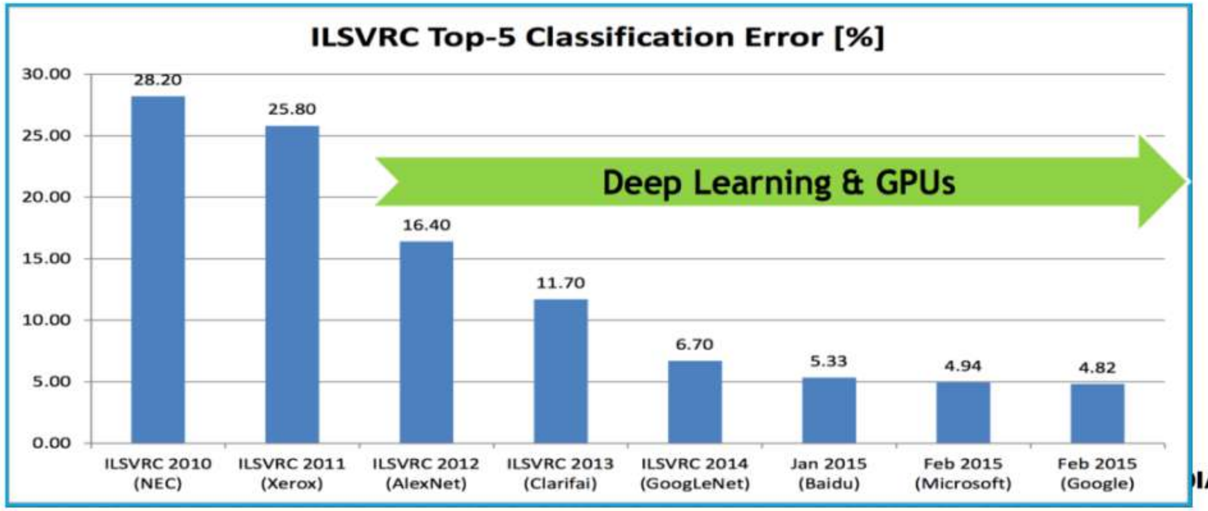
1.2M training images • 1000 object categories

Hosted by

IMAGENET







Training problems

Two major problems

Underfitting: model is bad at it's objective for all data

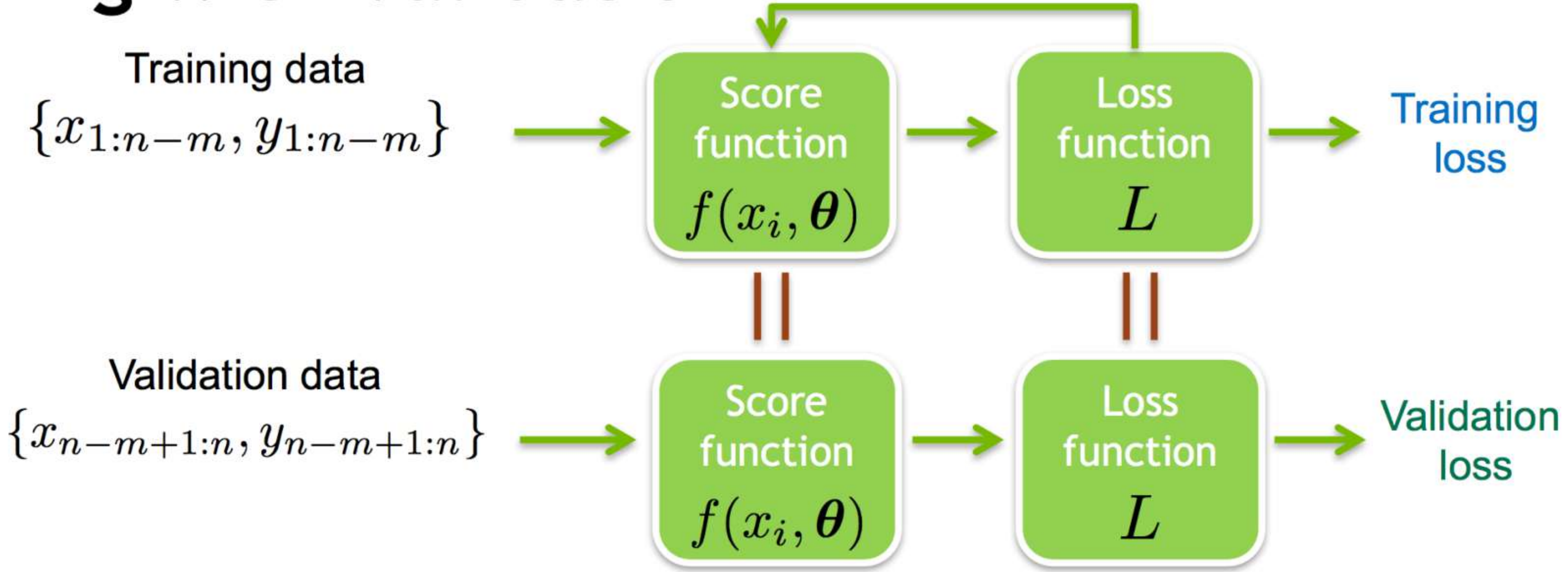
Overfitting: model is really good at the objective for the training data but bad on the testing data

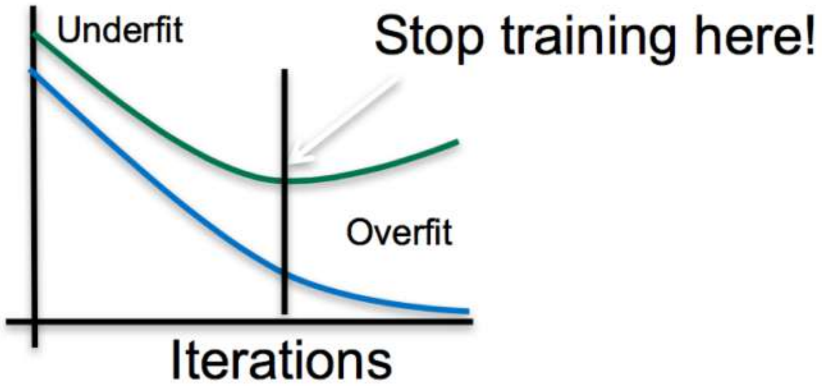
First line of defense:

Break off a **validation** dataset from the training data, e.g. 25%

Use it during training to check model performance on unseen data

Training with validation





Deep Learning

Learning Representation/Features

The traditional model of pattern recognition (since the late 50's)

Fixed/engineered features (or fixed kernel) + trainable classifier

Hand-crafted Feature Extractor → "Simple" Trainable Classifier →

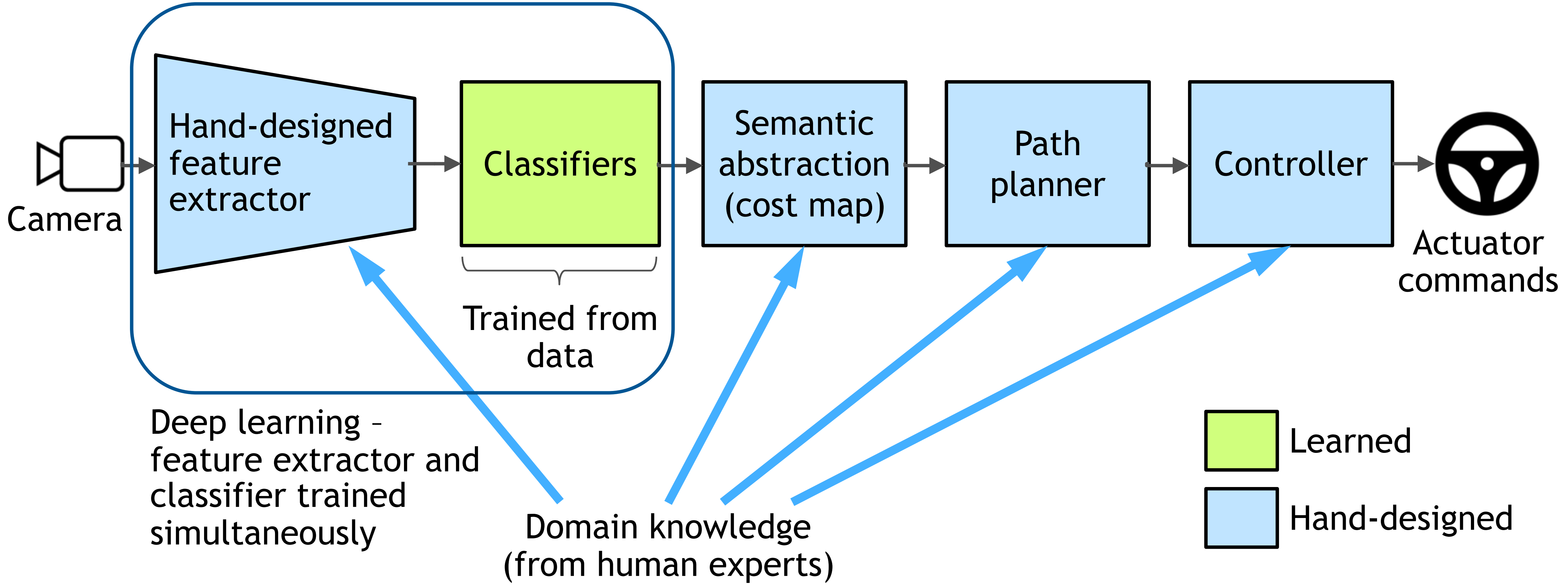
End-to-end learning / Feature learning / Deep learning

Trainable features (or kernel) + trainable classifier

Trainable Feature Extractor → Trainable Classifier →

TRADITIONAL DECOMPOSITION

Necessary approach when data and compute power are limited



EXAMPLE: ROAD FOLLOWING

Good quality lane markers, good driving conditions

Traditional lane detection-based systems expected to work well

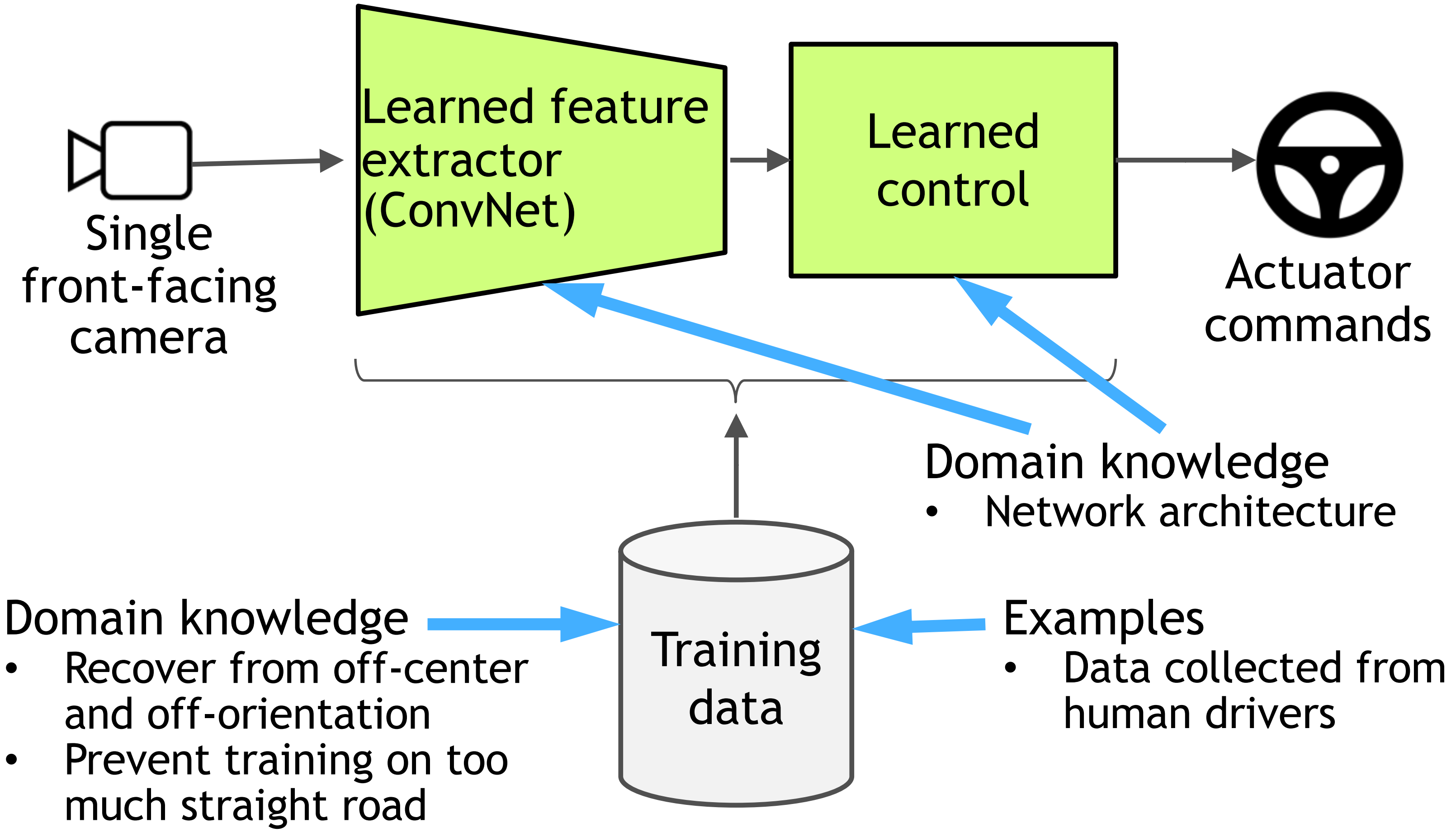
Poor quality lane markers

Lane detection-based systems struggle

End-to-end learning empowers the network to use additional cues

LEARNED ROAD FOLLOWING (PILOTNET)

Highway, local, residential – with or without lane markings



Both blocks trained simultaneously

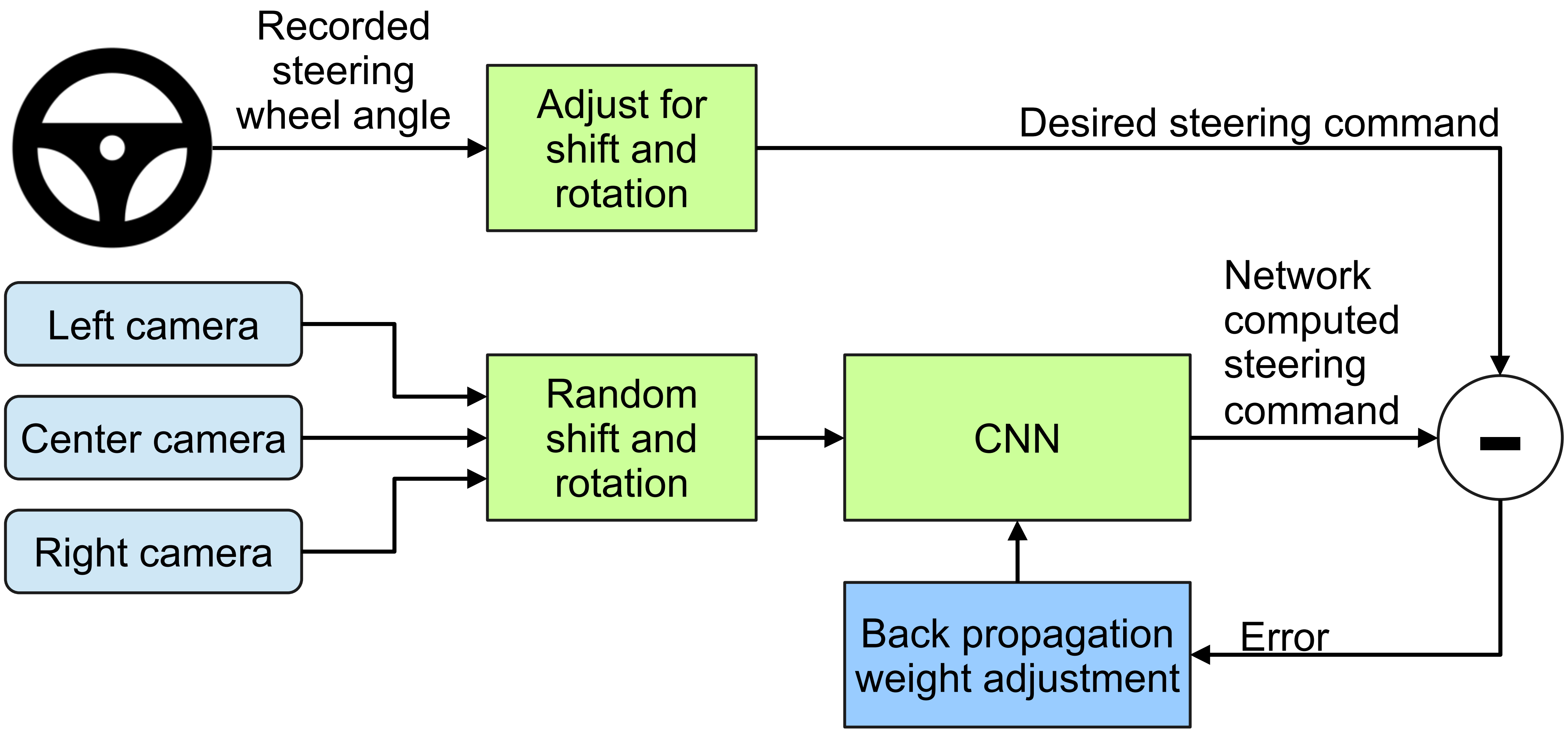
No explicit object detection nor path planning

→ Maps pixels directly to steering

Learned

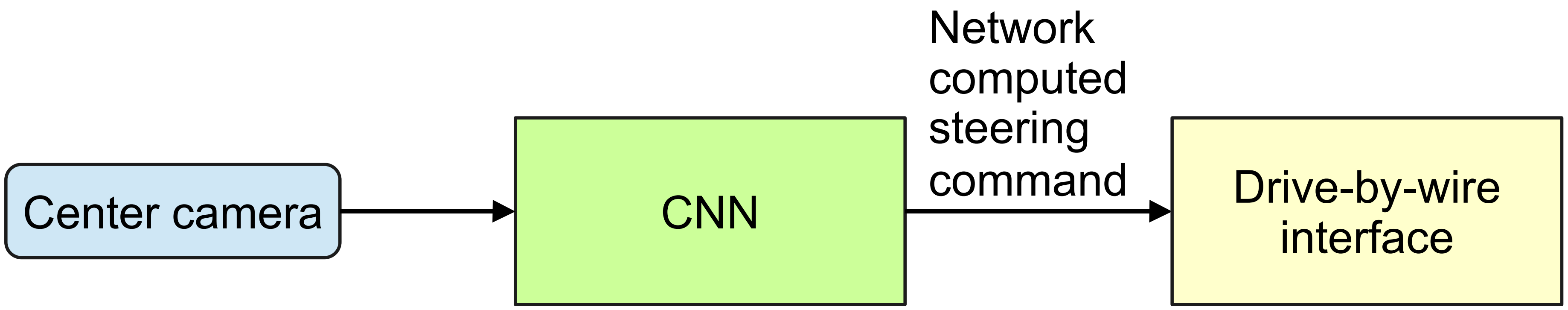
Hand-designed

TRAINING THE NEURAL NETWORK



DRIVING

With a single front-facing camera



VISUALIZATION

What the network pays attention to