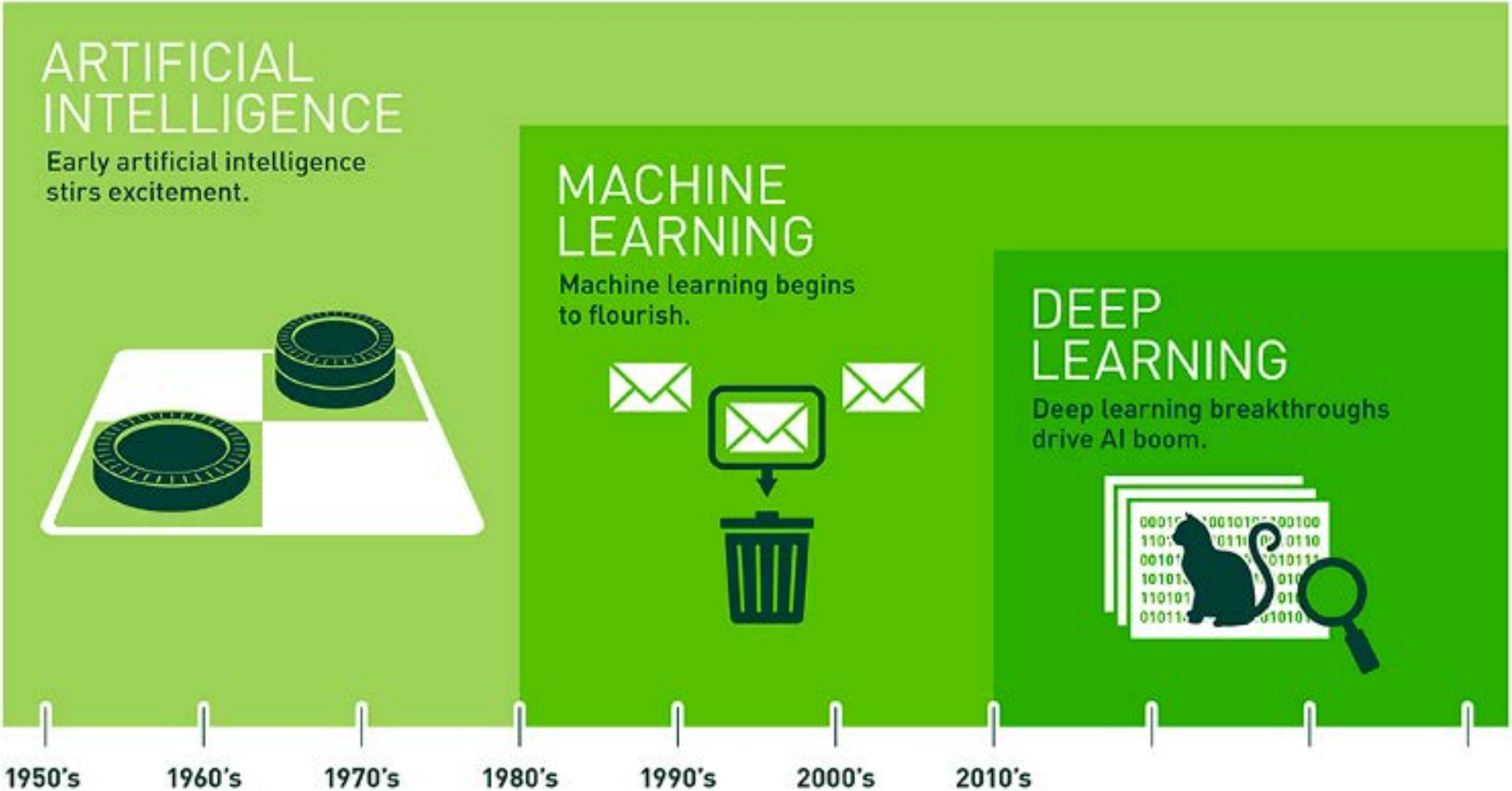
# Deep Learning & Autonomous Vehicles

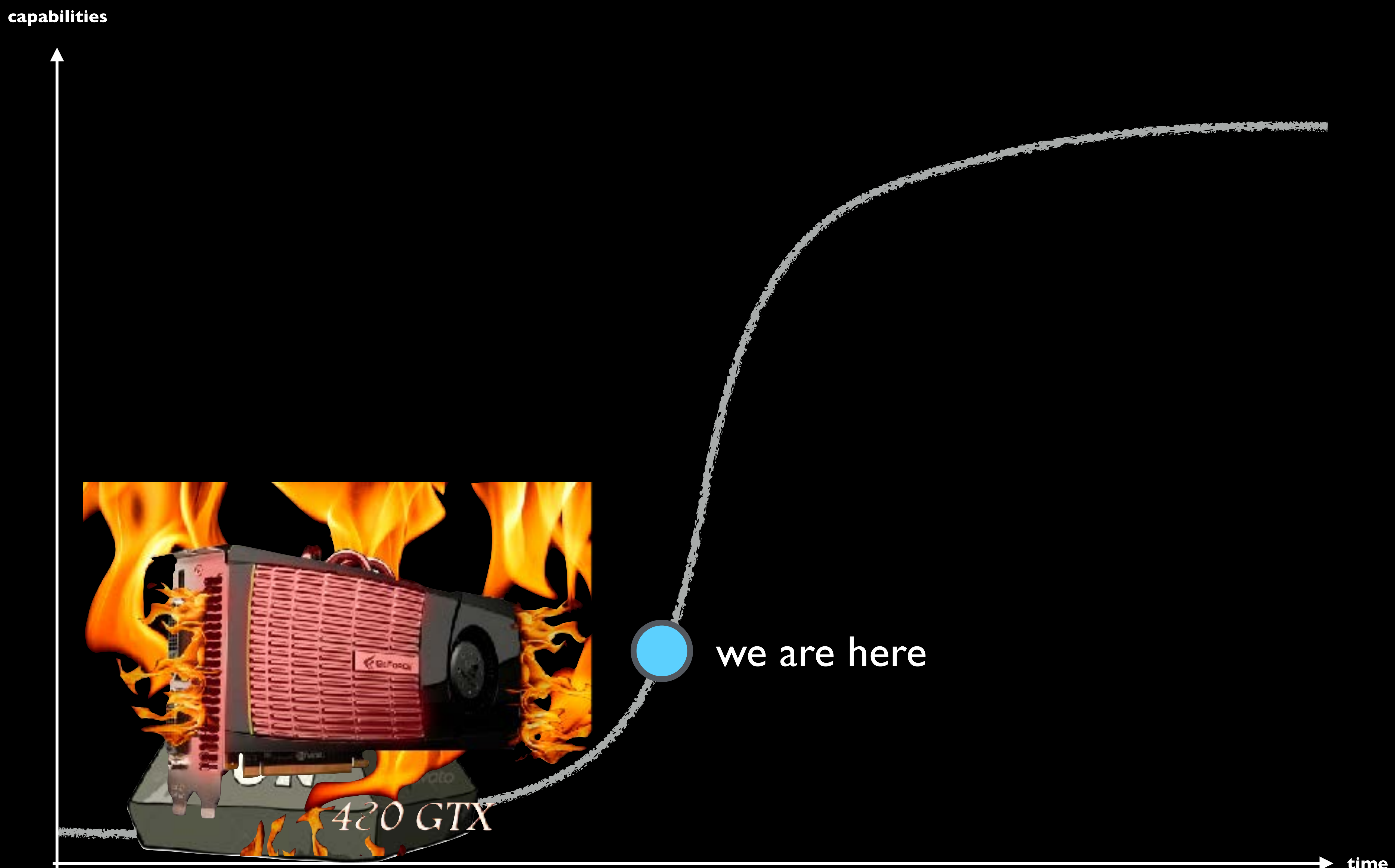
motus
ventures

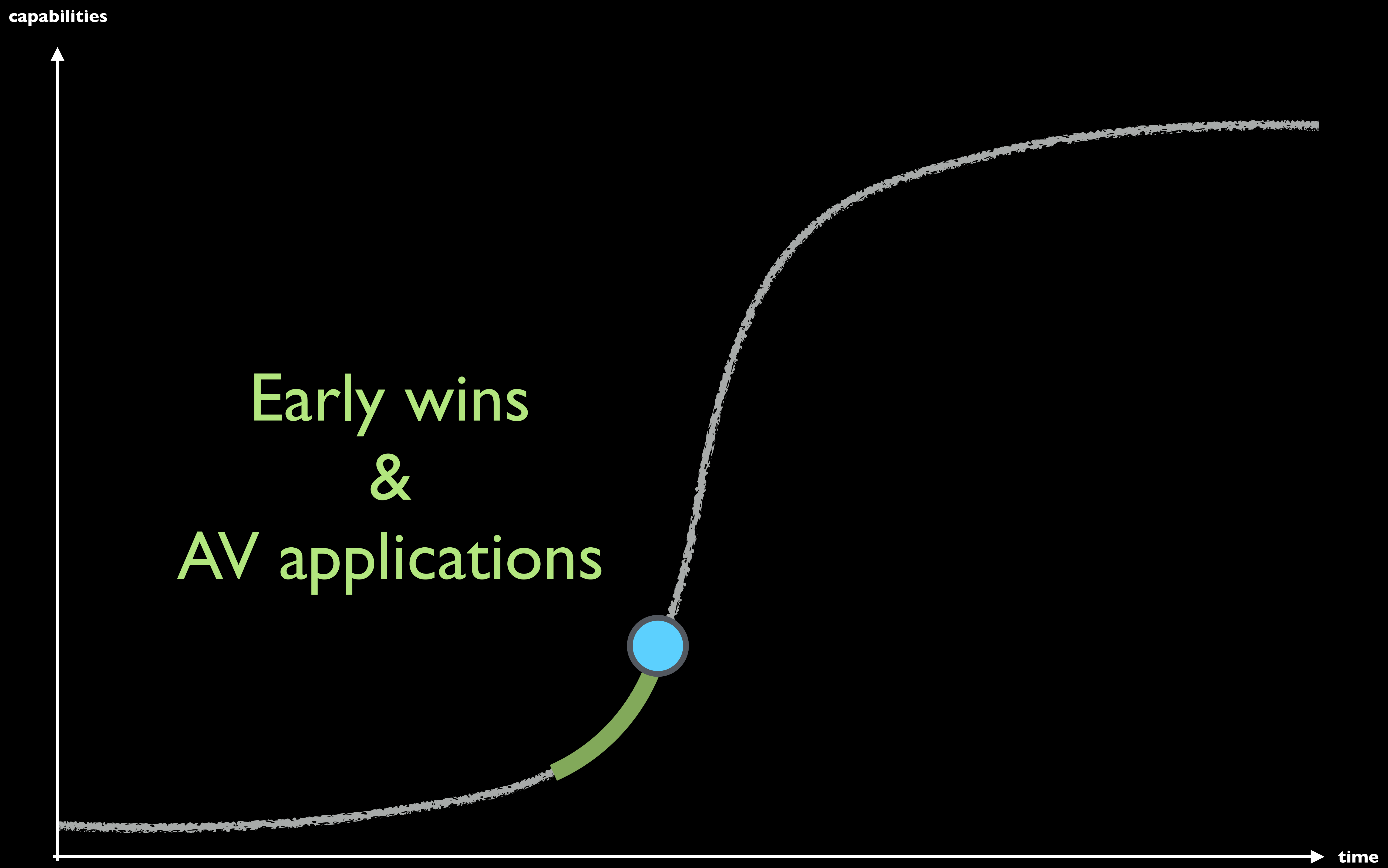
Robert Seidl
robert@motusventures.com




Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

# Deep Learning: Cambrian Explosion

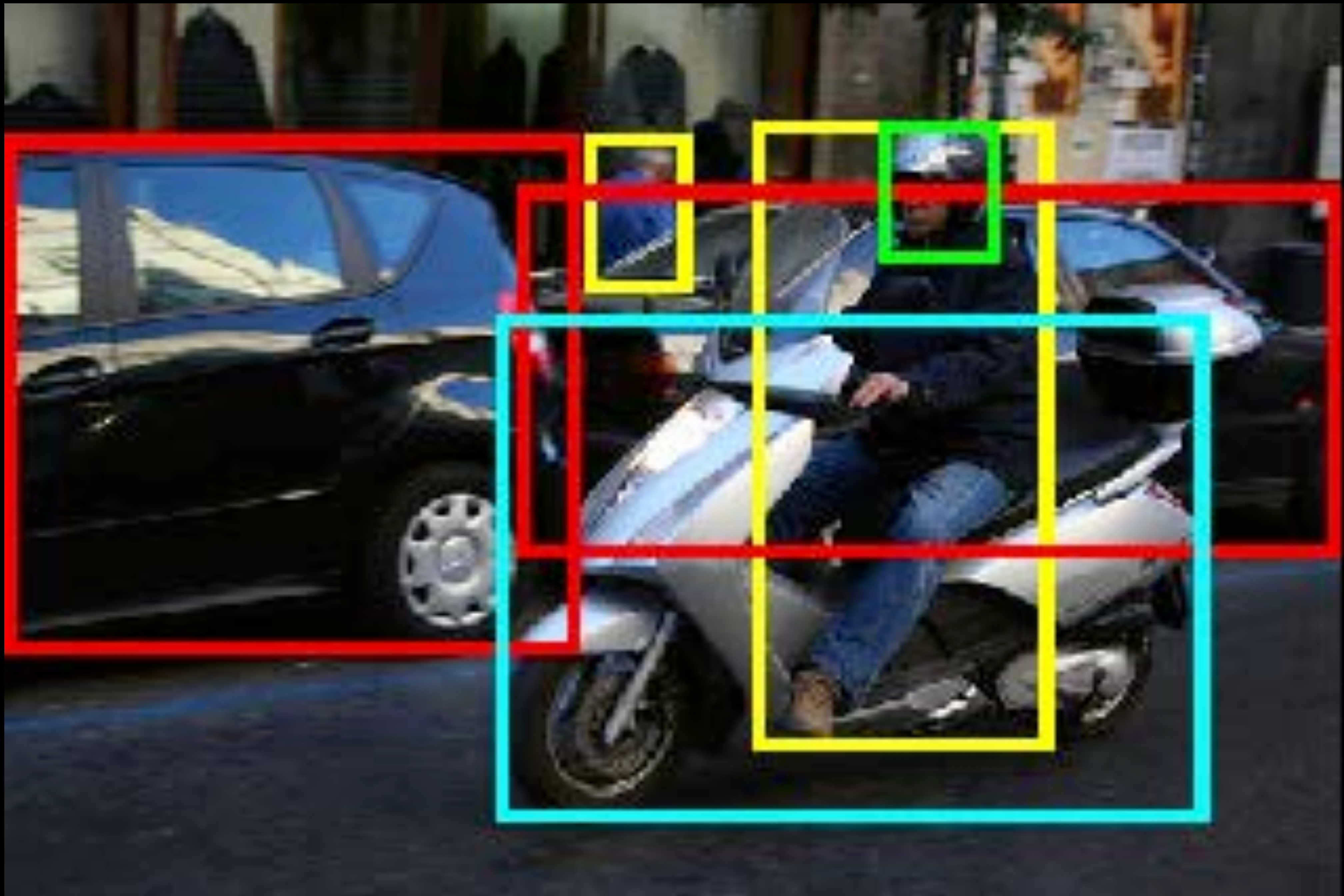



capabilities
Early wins
&
AV applications
time

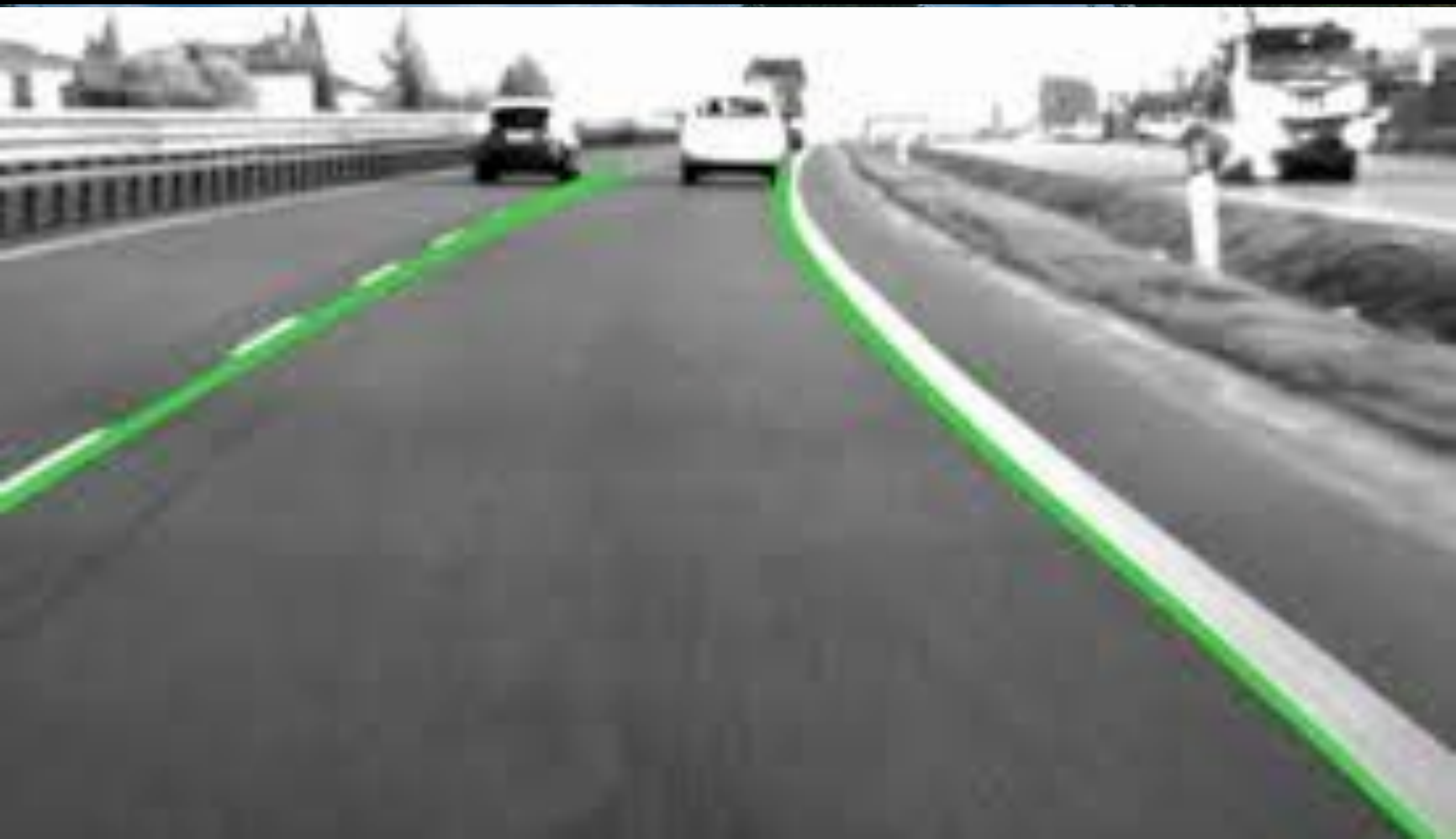
# Deep Learning Success Stories

- Object recognition (cats, faces, cars, voices etc.)
- Image matching, Machine translation, Games, Lip reading
- Healthcare: iphone dermatology diagnostics, Deep Patient
- Generative networks
  - Voice synthesis
  - Chat bots
  - Image Captioning
  - Image completion, inpainting

# Object detection & bounds




# Road & Lane markings



# Image Segmentation

# Free space detection

# HOW A DEEP NEURAL NETWORK SEES

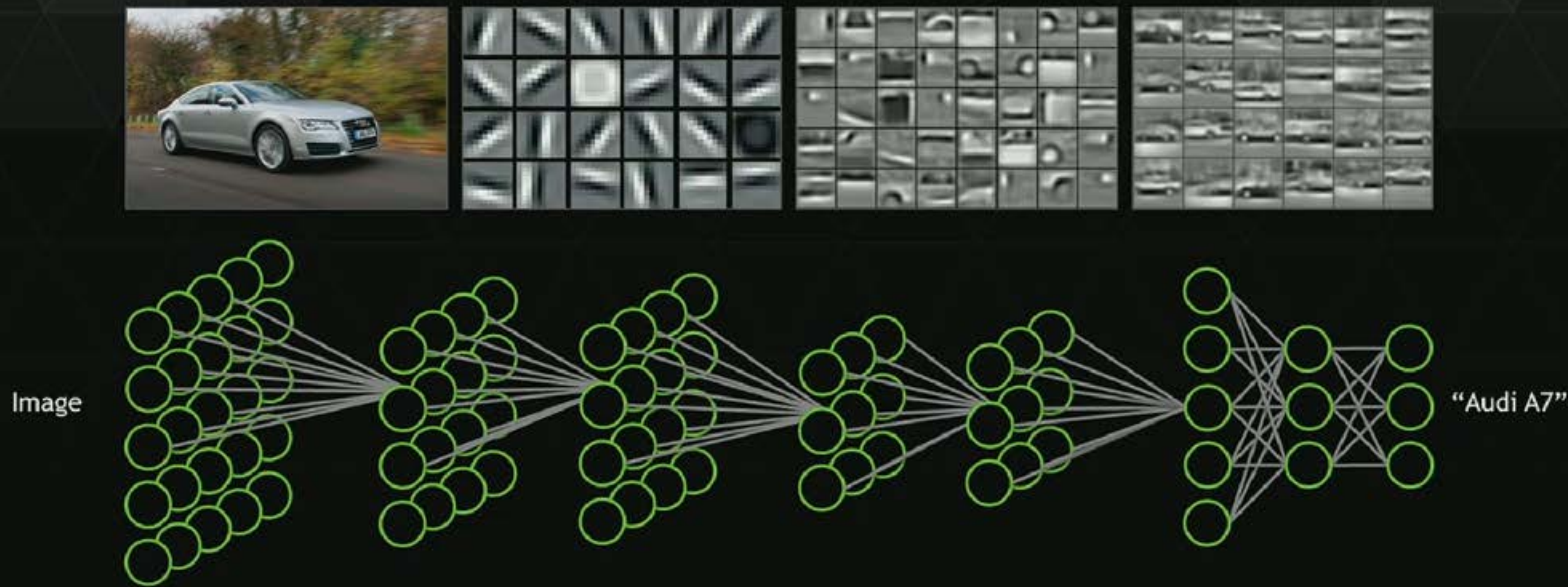
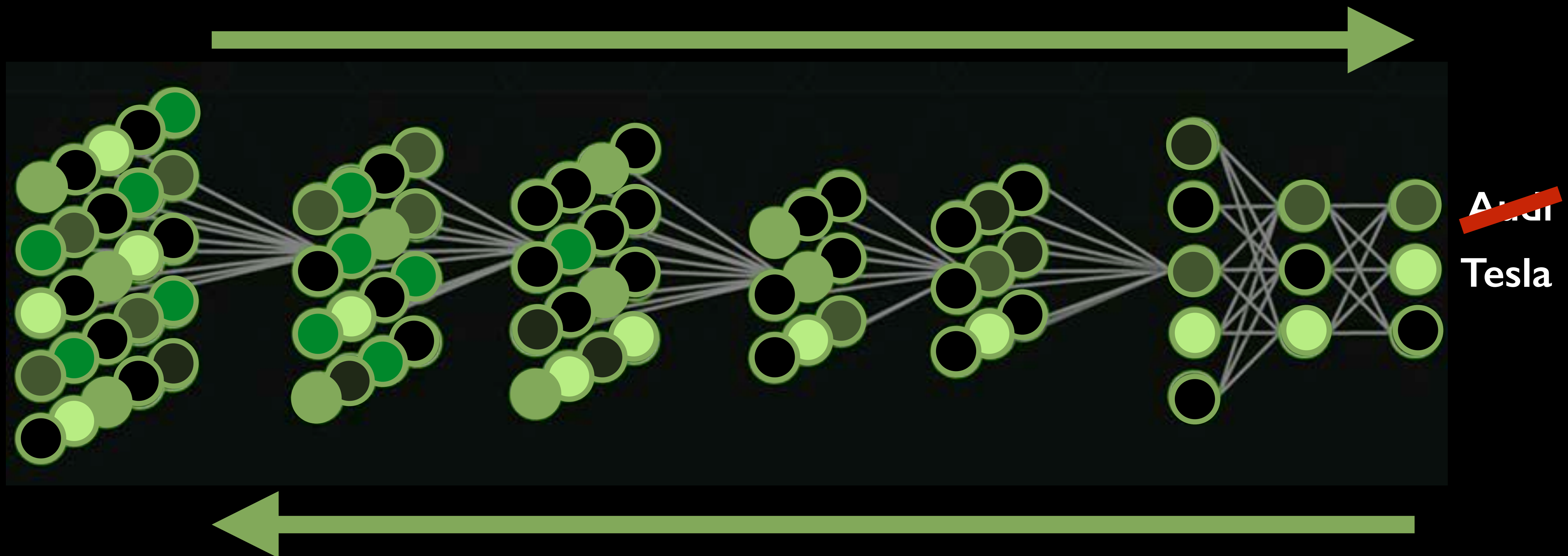



*Image source: "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks" ICML 2009 & Comm. ACM 2011. Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng.*

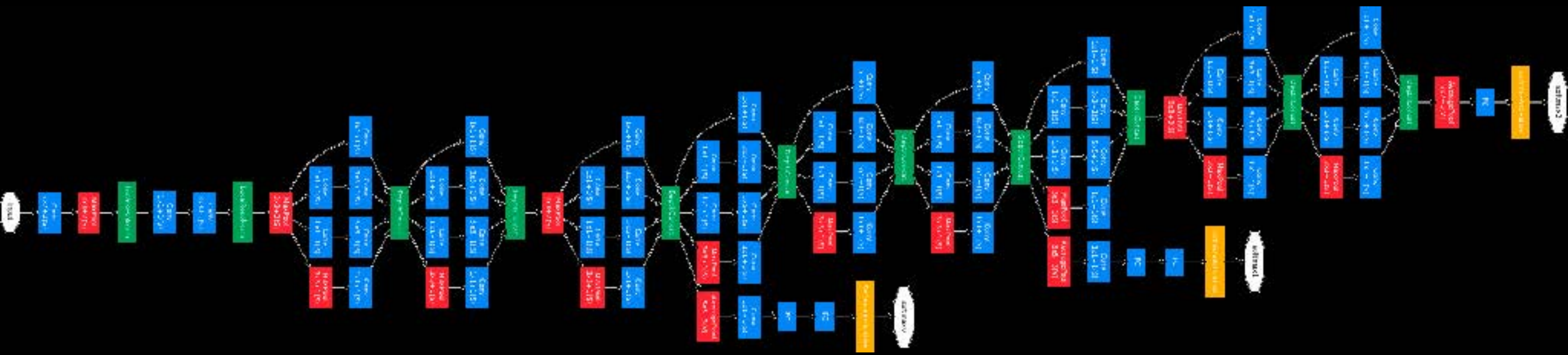
# Training

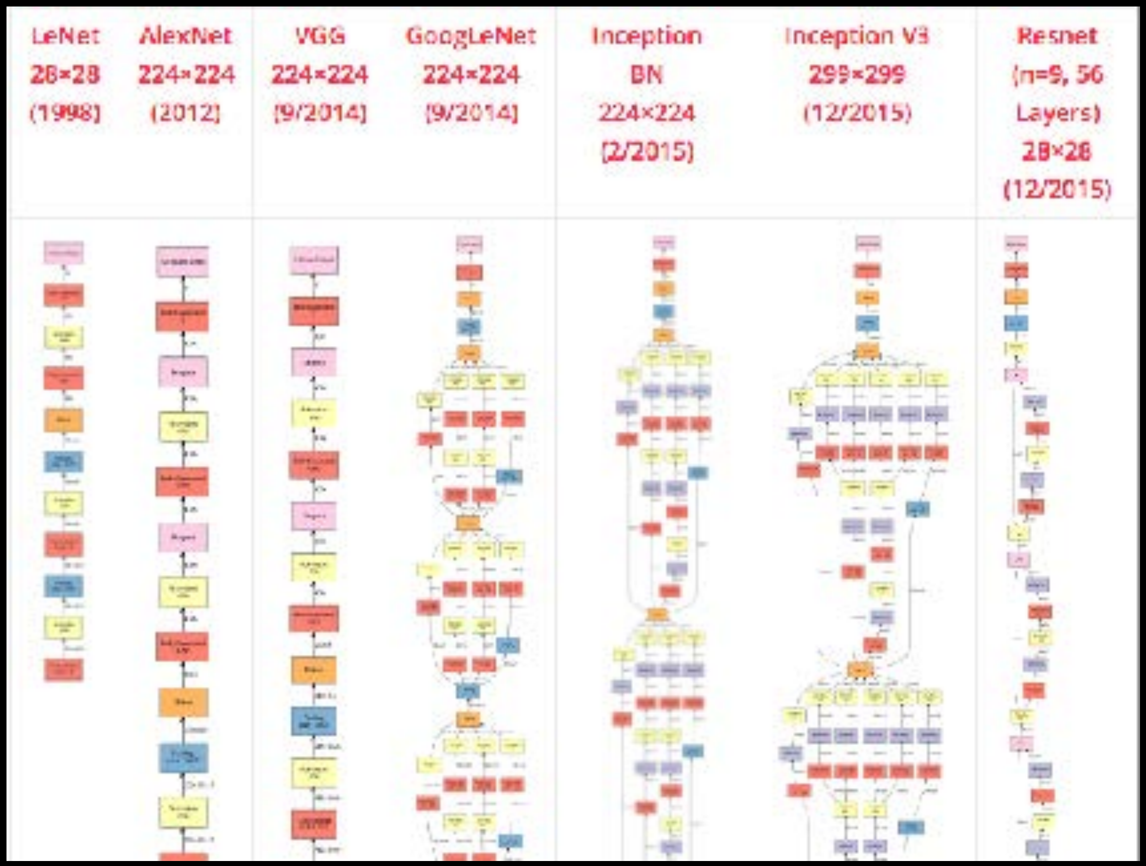
Forward / Inference pass




Back propagation / Error adjustment pass

# GoogleNet




LeNet 28×28 (1998)
AlexNet 224×224 (2012)
VGG 224×224 (9/2014)
GoogLeNet 224×224 (9/2014)
Inception BN 224×224 (2/2015)
Inception V3 299×299 (12/2015)
Resnet (n=9, 56 Layers) 28×28 (12/2015)


Convolution
Pooling
Softmax
Other

# AV Architecture


**Sense**

- Sensors & fusion
- Object Detection & tracking
- Behavior Detection
- Failure detection

- camera, lidar, radar, ultrasound
- V2x, 5G
- GPS, IMU, map/localization
- lead vehicle

**Plan**

- Trajectory control
- Maneuver planning
- e-horizon, maps
- Traffic signs, rules, lights
- Safe distances
- Comfortable dynamics

**Actuate**

- steering
- brakes, accelerator
- belt tighteners
- ECUs

# End-to-End learning


Raw sensor inputs

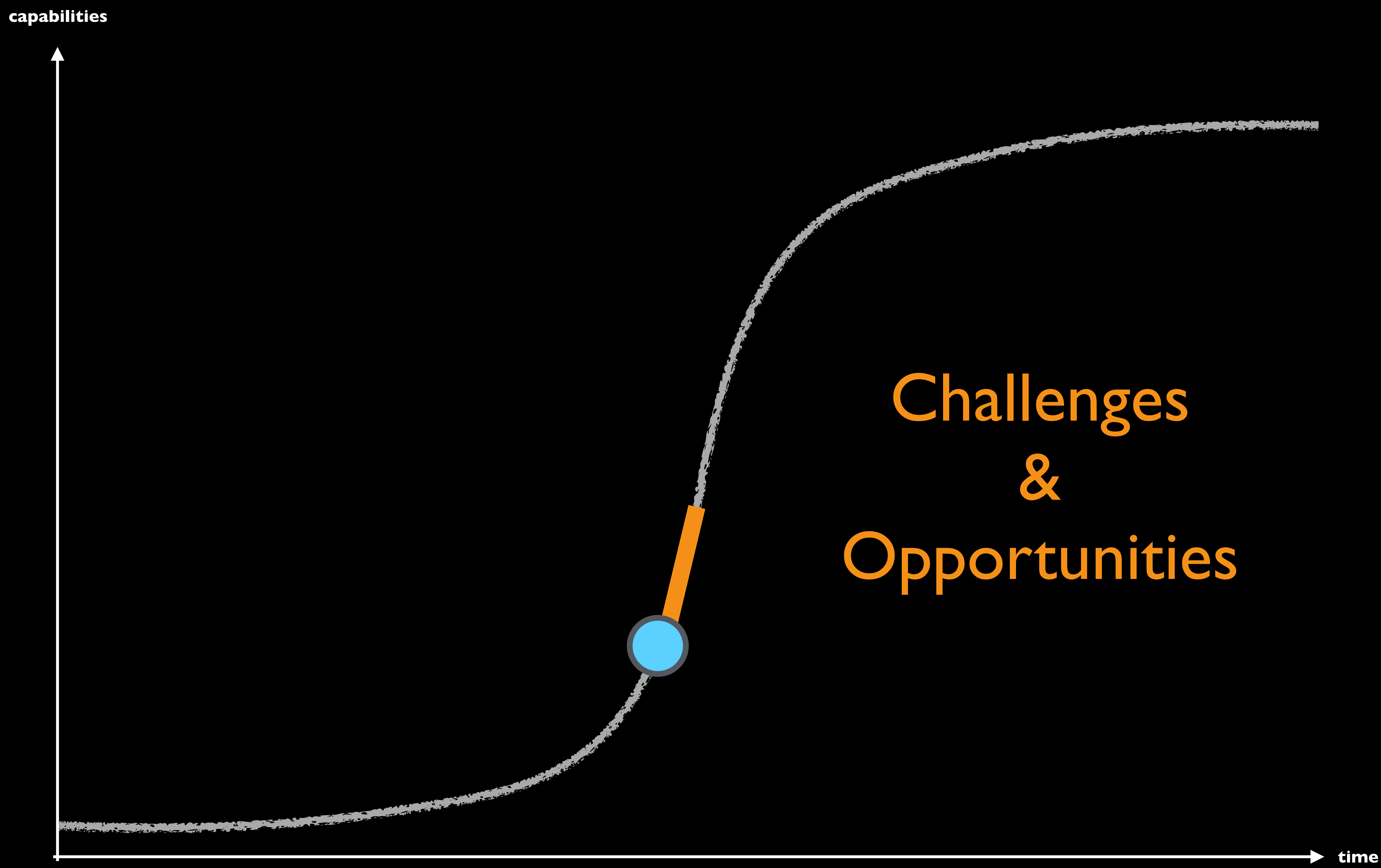
steering, brake outputs

http://www.cs.nyu.edu/~yann/research/dave/

14 years ago
Urs Muller, Yann LeCun
Training data: human driver,
100K images + stick input
72K nodes, 3M connections
no hand-coded algorithms

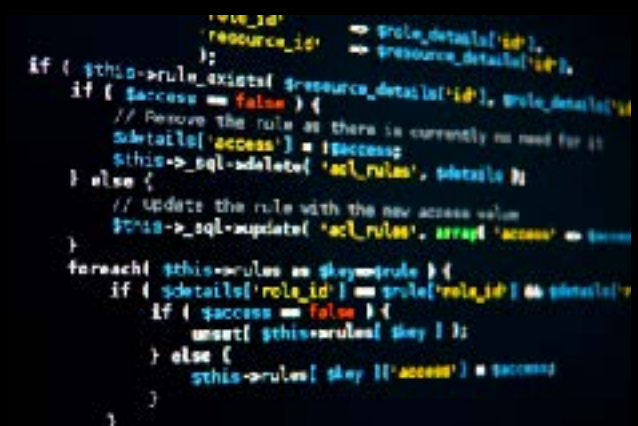
# End-to-End learning

NVidia BB8 car running PilotNet


capabilities
time
Challenges
&
Opportunities

# Challenges: Introspection & Development Tools

# Engineering "V"



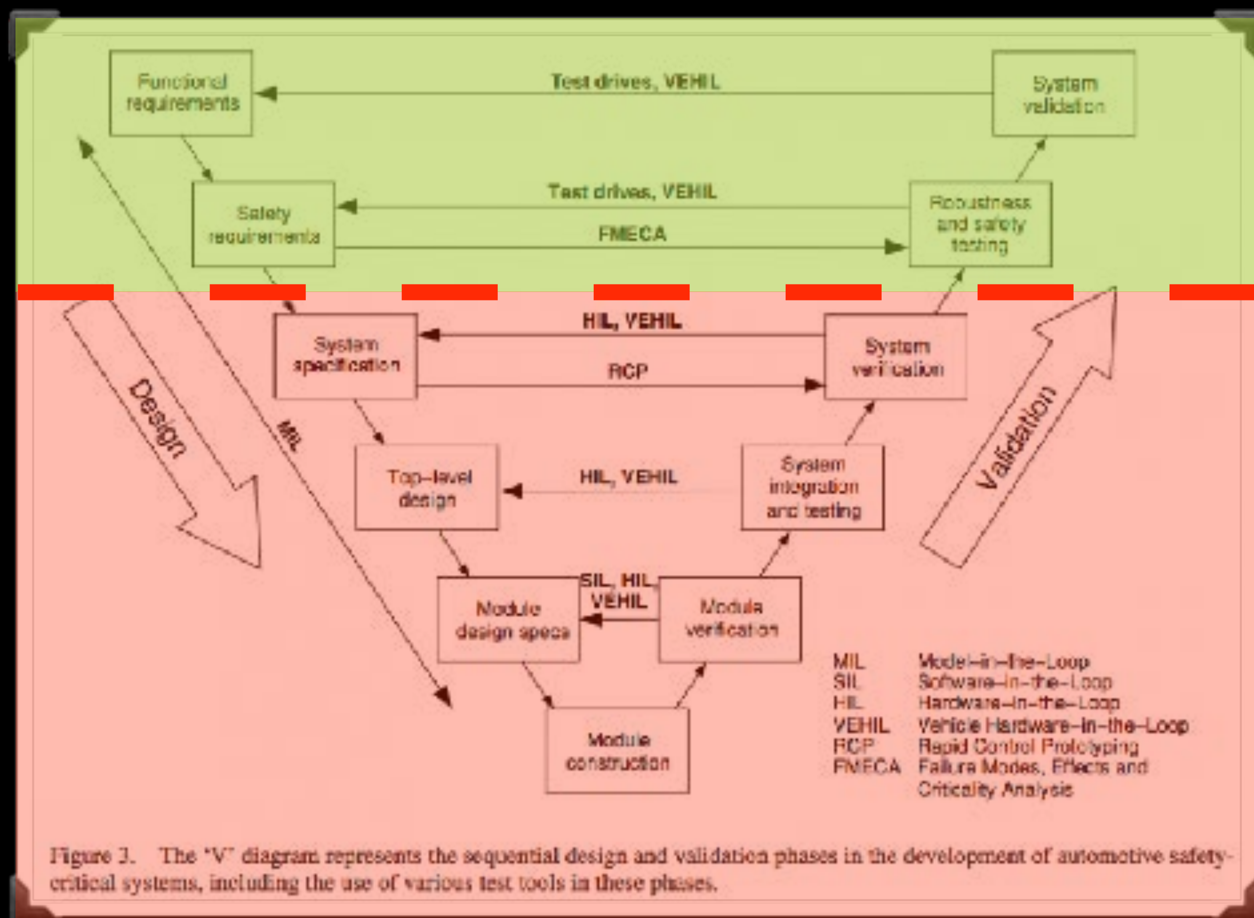
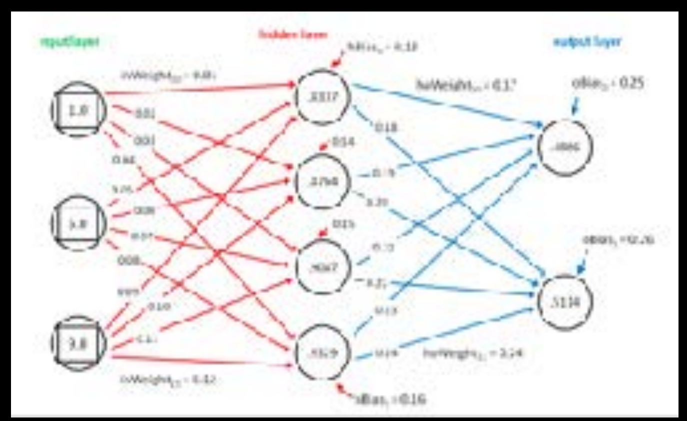
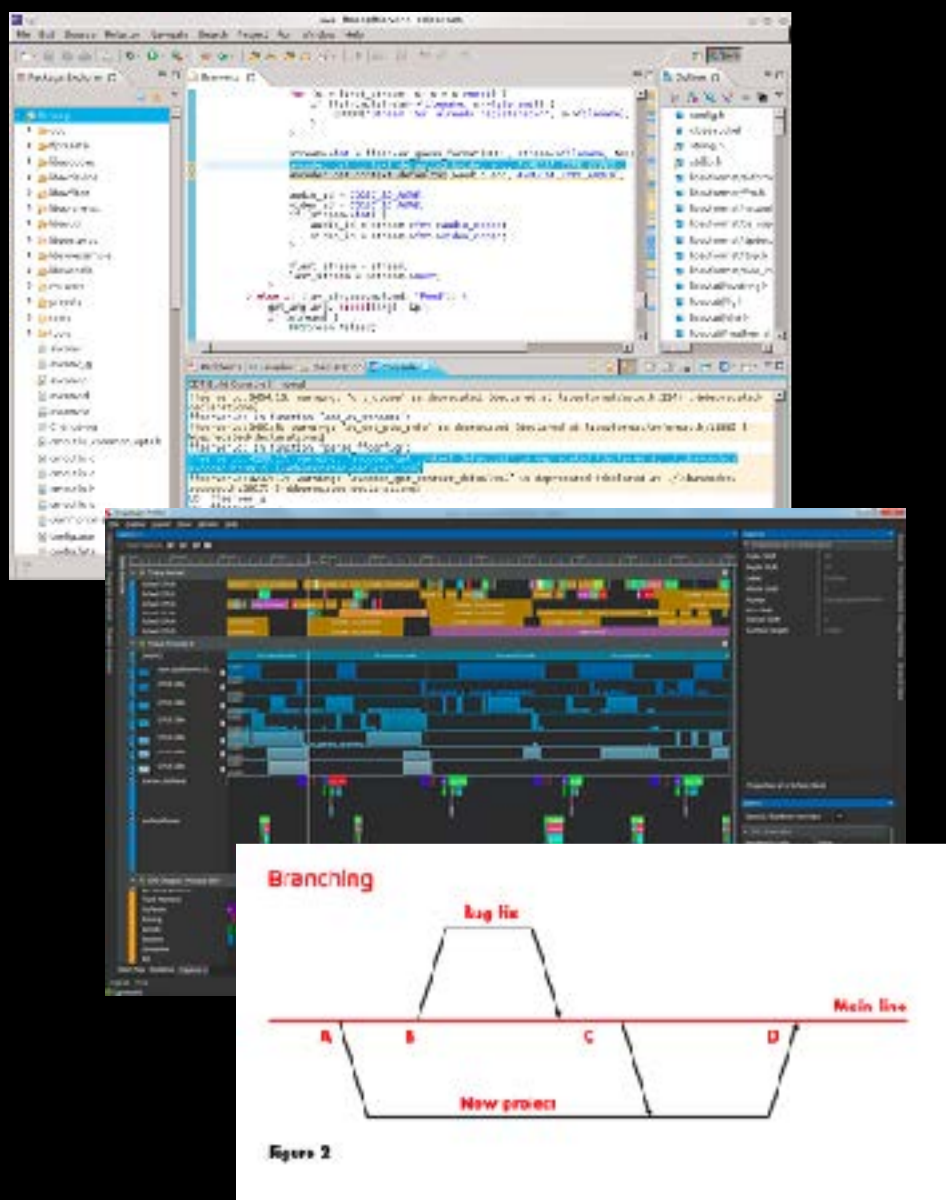


Figure 3. The "V" diagram represents the sequential design and validation phases in the development of automotive safety-critical systems, including the use of various test tools in these phases.




How does Deep Learning map to industry standard processes, ISO 26262 etc. ?

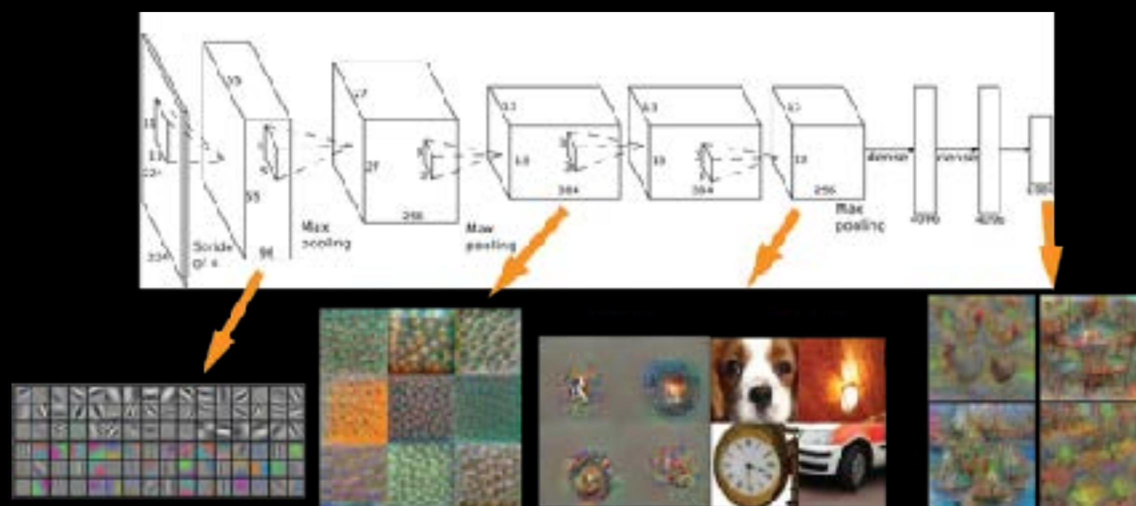
# Modern Coding Tools+Methods

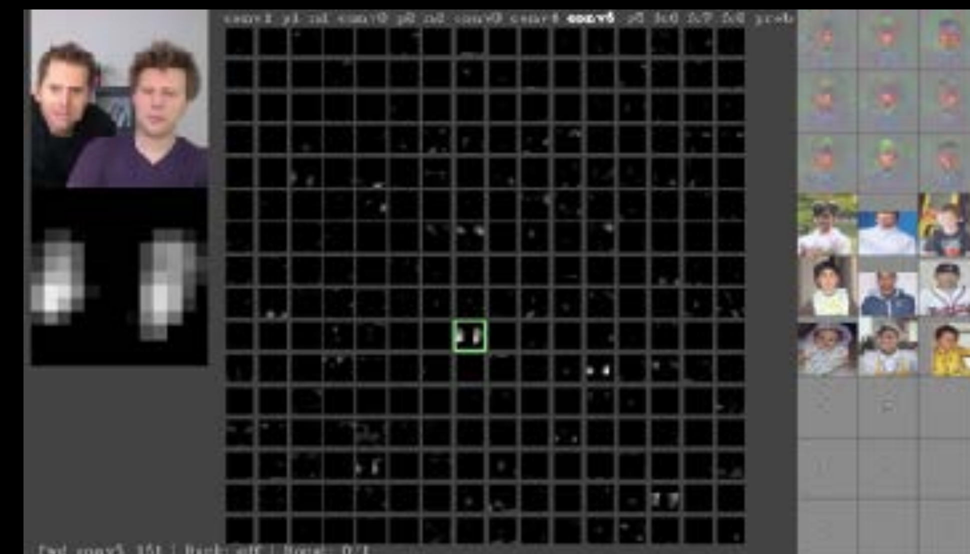



- Editor
- Debugger
- Profiler
- Static code analysis
- Versioning
- Testing
- Code coverage
- Requirements tracking

- Agile, SCRUM, CI, TDD

# NN: still more art than science





Deep Visualization Toolbox

- Network architecture
  - Height / Width / Reusable modules
- Training set size, backprop f’s
- Don’t run over grandma
- Adding outputs w/o retraining
- Activation semantics
- Undertrained parts

Spatial Saliency visualization (NVidia)

# Faster learning of rules

**Sense** → **Plan** → **Actuate**

Deep Learning
Data driven
Perception

Procedural
Rule driven
Planning




Your perception
planning
control code

ROS, Tensorflow
Drive works, AutoSAR
OpenGL, CUDA
Linux, QNX

DrivePX, Freescale etc.

$
Development costs

3-4x $
Maintenance costs

# Opportunities

# Functional & Safety validation

using

Sensor Simulation


Autoliv
Autoliv
30 TESTS / DAY

# Cityscapes dataset

20K images hand labeled data, 30 classes, 50 cities, Daimler/Max Planck/TU Darmstadt

https://www.cityscapes-dataset.com/

# Tesla collecting driver behavior


>100K vehicles, >3M miles per day

The world is way too big to just *randomly* sample it.

Phillip Isola (MIT/UC Berkeley)

# TESTING

day

night

different light conditions

rain

fog, snow

# THOUSANDS OF CASES

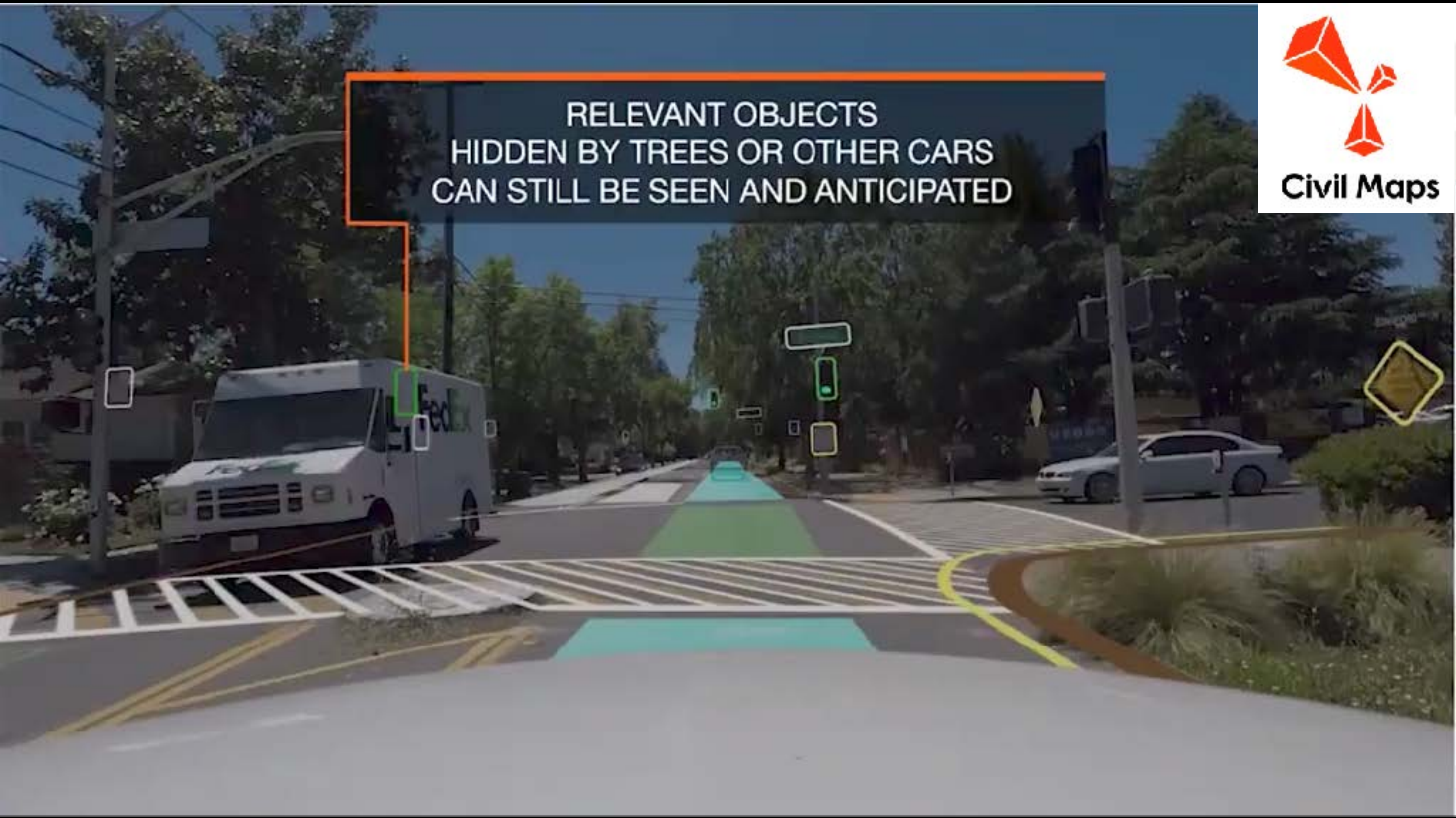

3 MILLION
TESTS / DAY

# Simulation: a key driver for AV success

- Controllable (weather, lighting, sensor configs etc.)
- Repeatable, fuzz-able
- Scalable
- Dangerous = ok
- Lots of training data, fully labeled, with ground truth
- **Sim Training is transferable**

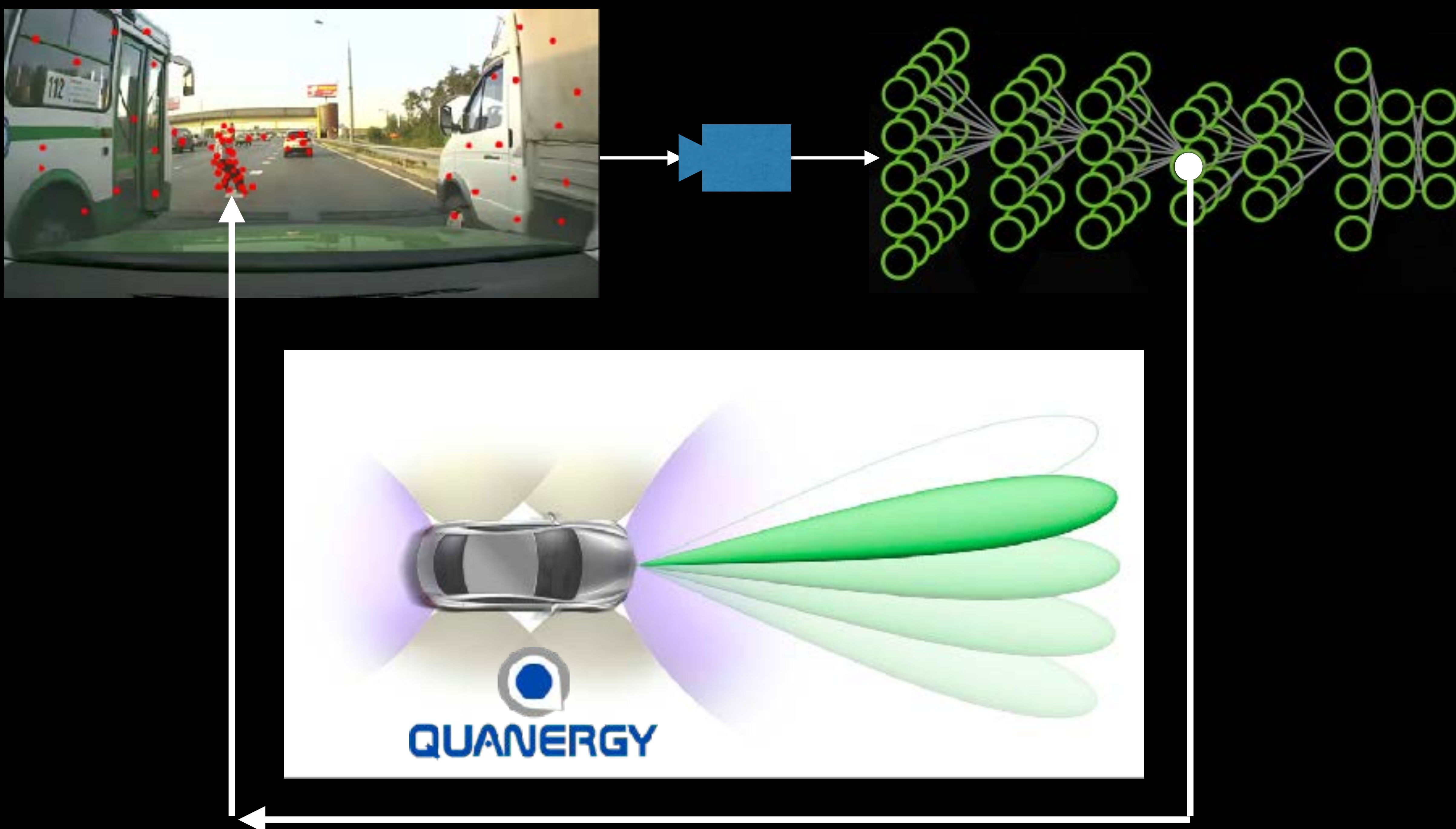
# Sensor Fusion


038402
698 TSW
Scirocco.


RELEVANT OBJECTS
HIDDEN BY TREES OR OTHER CARS
CAN STILL BE SEEN AND ANTICIPATED
Civil Maps

# Sensor Steering

foveation requires loops rather than just feedforward

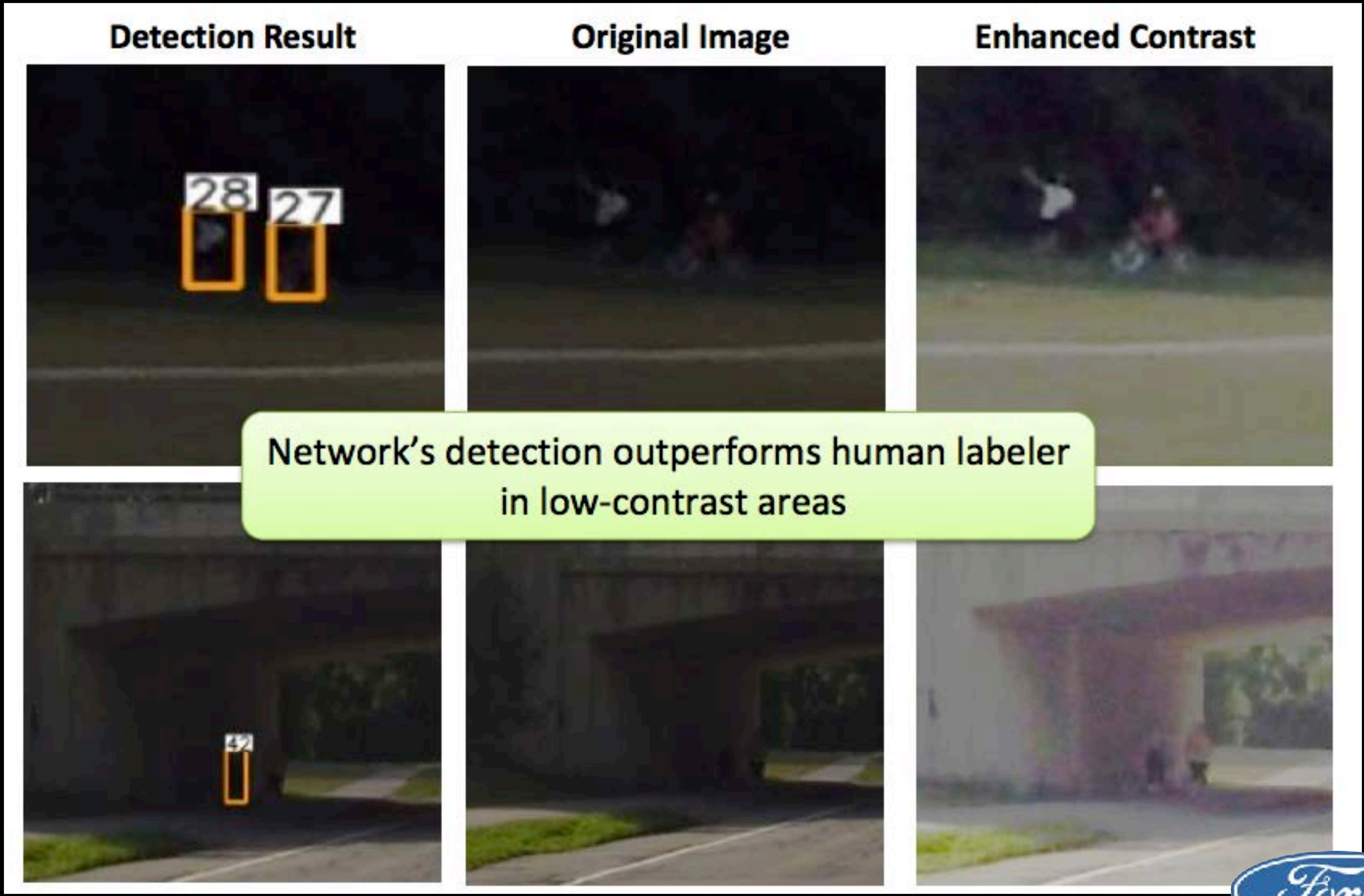


Behavior

not just

Object recognition

# RNNs to better detect "difficult" objects



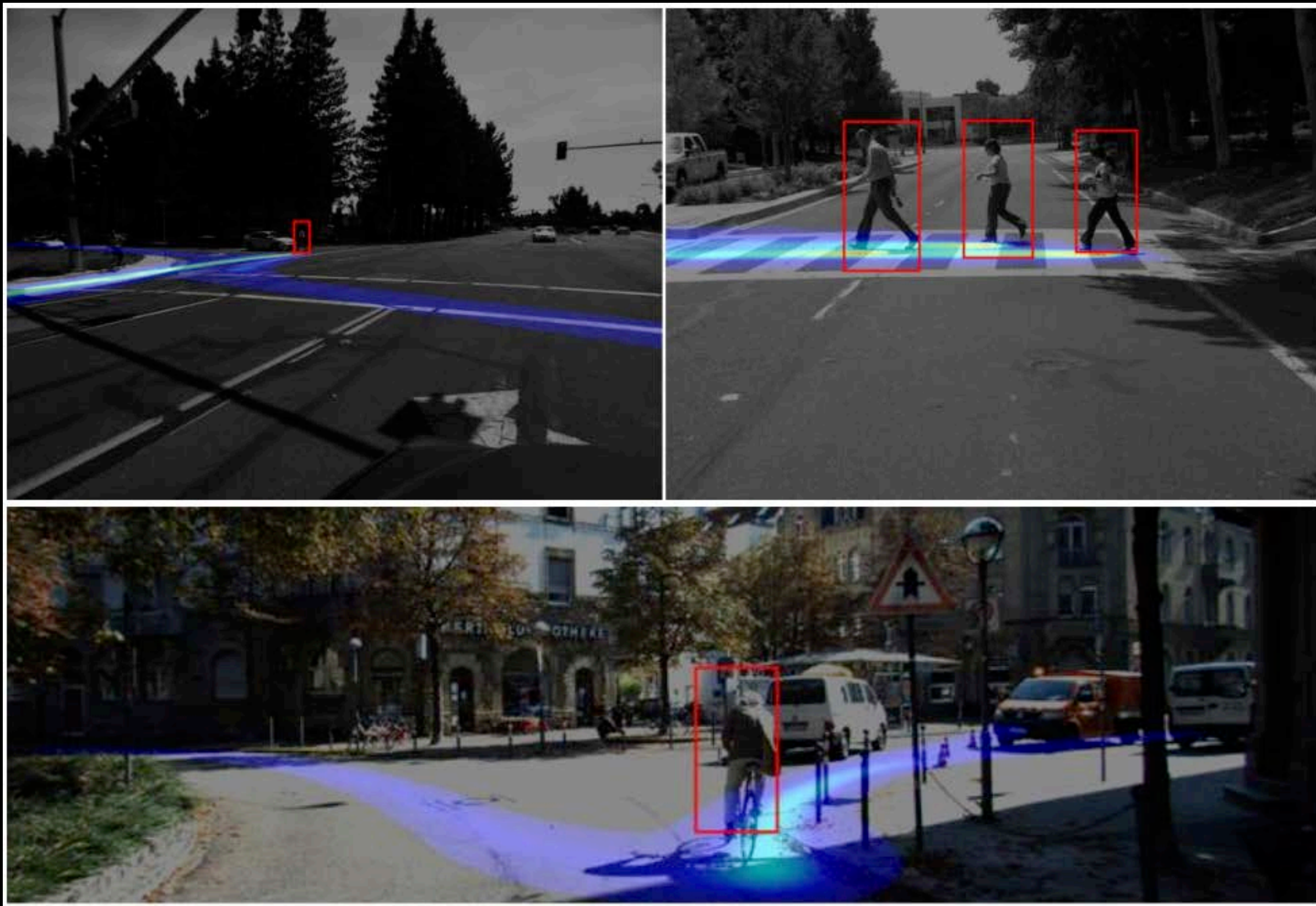

Network's detection outperforms human labeler in low-contrast areas

# Behavior classification


# Eye contact

Head pose, gaze estimation

# Intent-Aware Long-Term Prediction of Pedestrian Motion



http://vision.ucla.edu/papers/karasevAHS16.pdf

# Sharing Data & Best Practices

- Scenario catalog
    - scenarios (geometry, traffic, environment etc.)
    - eg freeway maneuvers
    - German project:
- Accident Data
- Realistic traffic & pedestrian data

PEGASUS

http://www.pegasus-projekt.info/

# This has only just begun