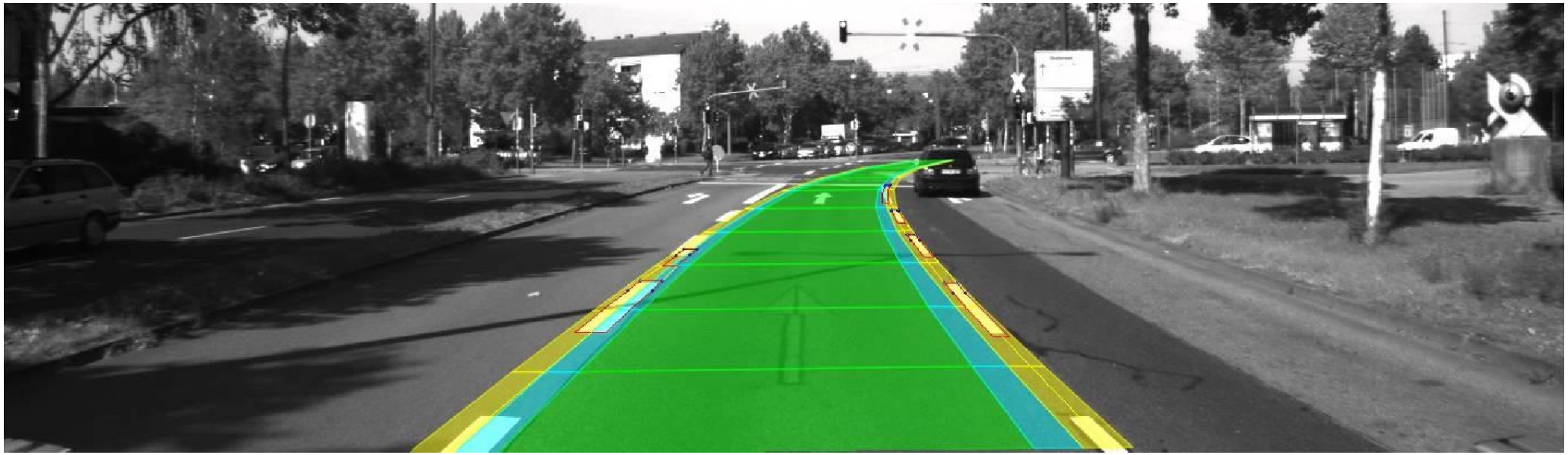
6.S094: Deep Learning for Self-Driving Cars

Introduction to Deep Learning and Self-Driving Cars

cars.mit.edu



References: [13, 14]

Target Audience

You may be:

- New to programming
- New to machine learning
- New to robotics

What you will learn:

- An overview of deep learning methods:
  - Deep Reinforcement Learning
  - Convolutional Neural Networks
  - Recurrent Neural Networks
- How deep learning can help improve each component of autonomous driving: perception, localization, mapping, control, planning, driver state

Target Audience

Not many equation slides like the following:

Lemma 0.1. *Let $\mathcal{C}$ be a set of the construction.*

Let $\mathcal{C}$ be a gerber covering. Let $\mathcal{F}$ be a quasi-coherent sheaves of $\mathcal{O}$-modules. We have to show that

$$\mathcal{O}_{\mathcal{O}_X} = \mathcal{O}_X(\mathcal{L})$$

.

Proof. This is an algebraic space with the composition of sheaves $\mathcal{F}$ on $X_{étale}$ we have

$$\mathcal{O}_X(\mathcal{F}) = \{morph_1 \times_{\mathcal{O}_X} (\mathcal{G}, \mathcal{F})\}$$

where $\mathcal{G}$ defines an isomorphism $\mathcal{F} \to \mathcal{F}$ of $\mathcal{O}$-modules. □

* Though it would be more efficient, since the above is LaTeX code automatically generated character by character with Recurrent Neural Networks (RNNs)

[35] Andrej Karpathy. "The Unreasonable Effectiveness of Recurrent Neural Networks." (2015).

Project: **DeepTraffic**

Speed:
80 mph
Cars Passed:
290

Road Overlay:
None

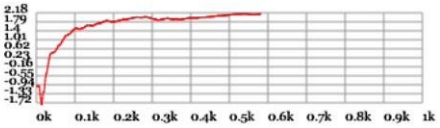
Simulation Speed:
Normal

DeepTraffic

Americans spend 8 billion hours stuck in traffic every year.
Deep neural networks can help!

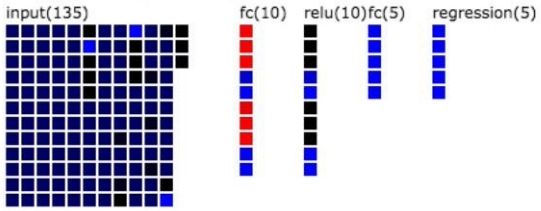
```
1
2  //<![CDATA[
3  // a few things don't have var in front of them - they update already
   existing variables the game needs
4  lanesSide = 1; //1;
5  patchesAhead = 10; //13;
6  patchesBehind = 0; //7;
7  trainIterations = 100000;
8
9  // begin from convnetjs example
10 var num_inputs = (lanesSide * 2 + 1) * (patchesAhead + patchesBehind);
11 var num_actions = 5;
12 var temporal_window = 3; //1 // amount of temporal memory. 0 = agent lives
   in-the-moment :)
13 var network_size = num_inputs * temporal_window + num_actions *
```

Apply Code/Reset Net | Save Code/Net to File | Load Code/Net from File | Submit Model to Competition

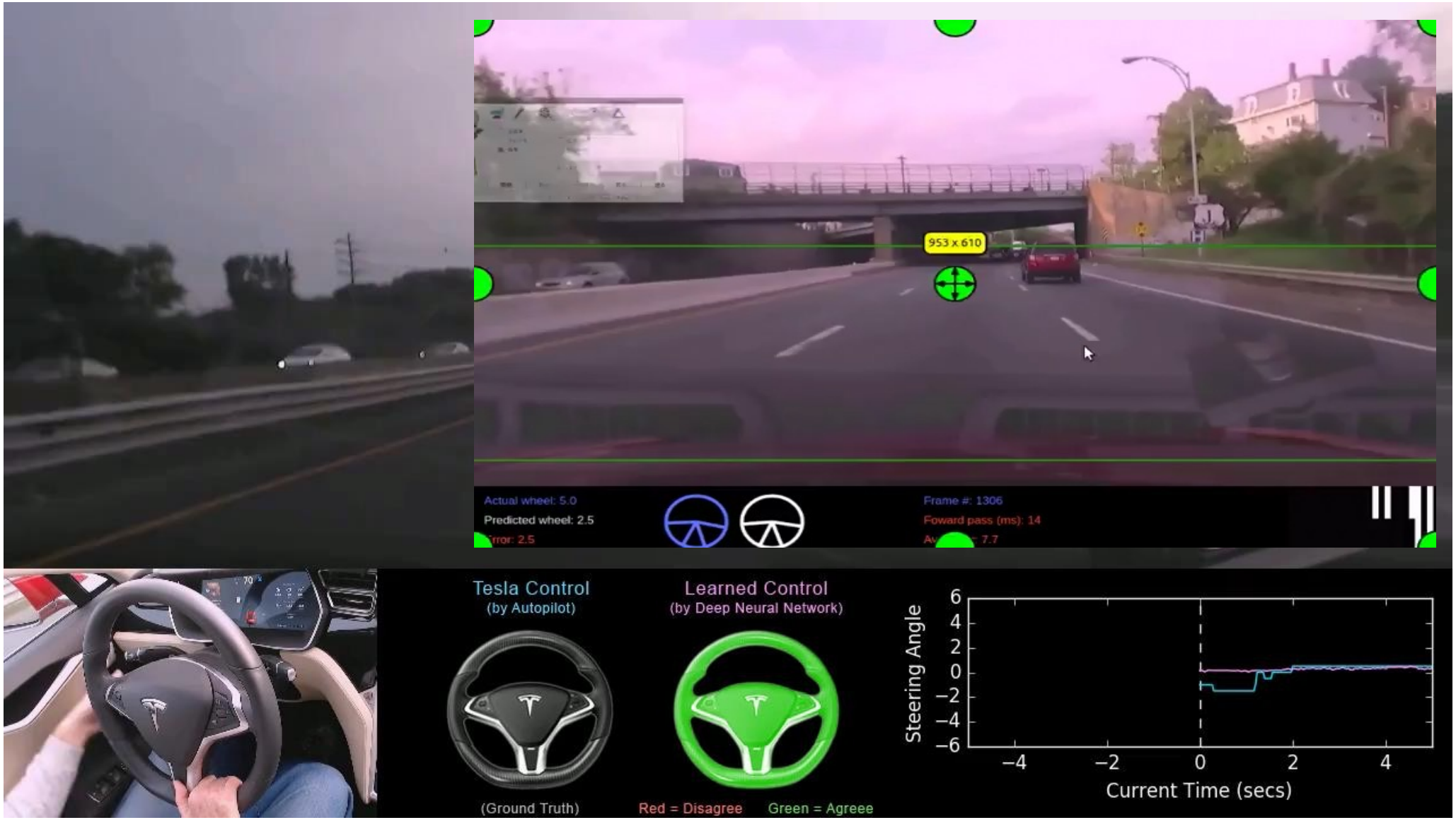


Start Evaluation Run

Value Function Approximating Neural Network:

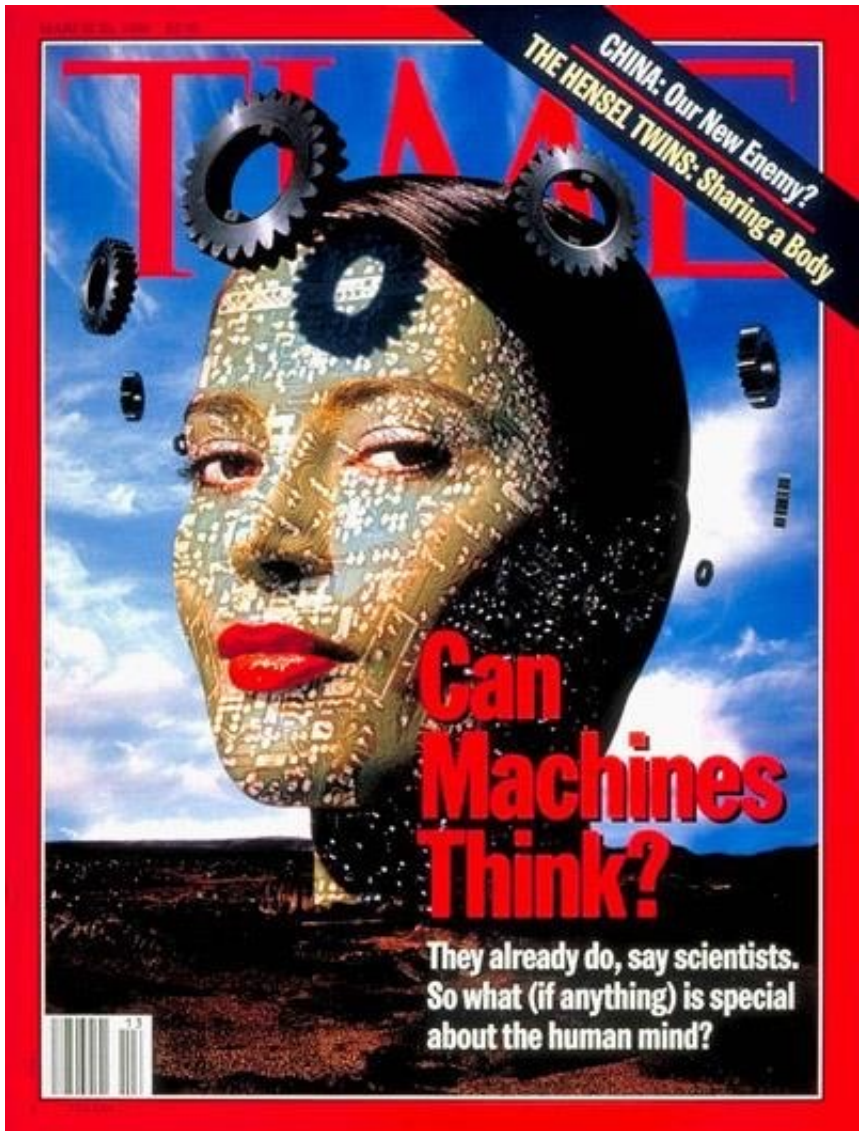


Project: **DeepTesla**



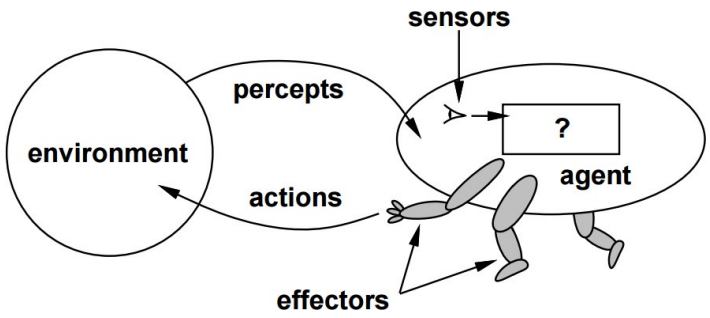
Defining (Artificial) Intelligence

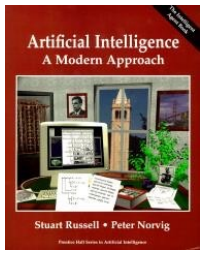
March 25, 1996



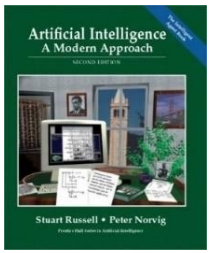
Special Purpose:
Can it achieve a well-defined finite set of goals?

General Purpose:
Can it achieve poorly-defined unconstrained set of goals?

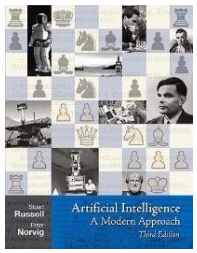




(1995)

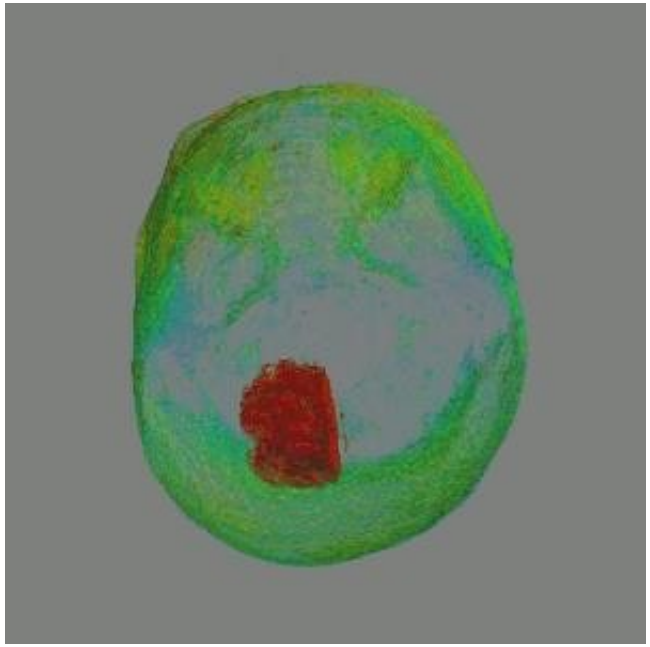


(2002)



(2009)

References: [59, 60, 61]

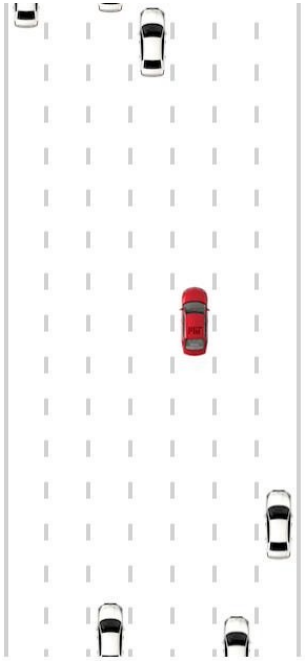


- **Formal tasks:** Playing board games, card games. Solving puzzles, mathematical and logic problems.
- **Expert tasks:** Medical diagnosis, engineering, scheduling, computer hardware design.
- **Mundane tasks:** Everyday speech, written language, perception, walking, object manipulation.

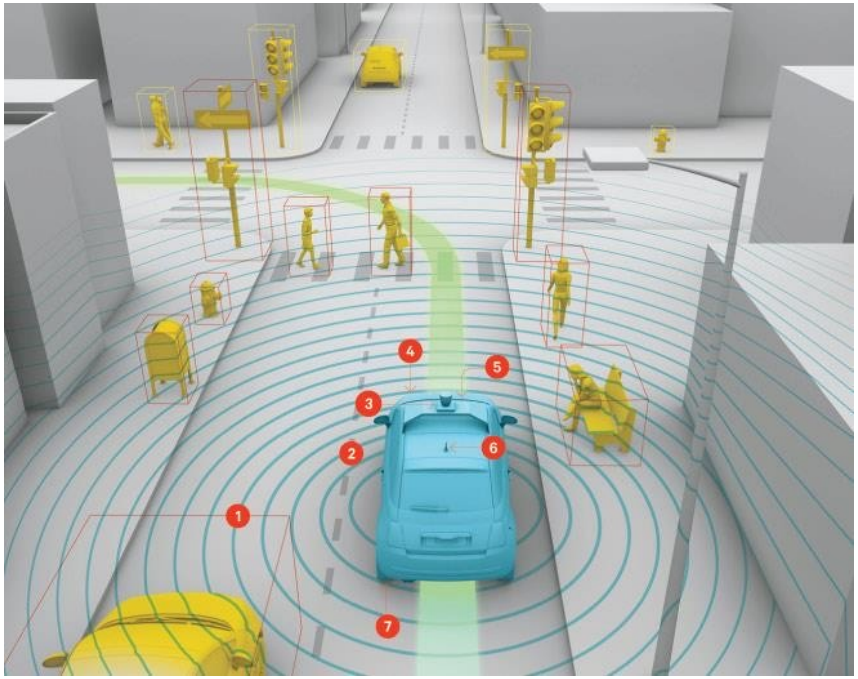
How Hard is Driving?

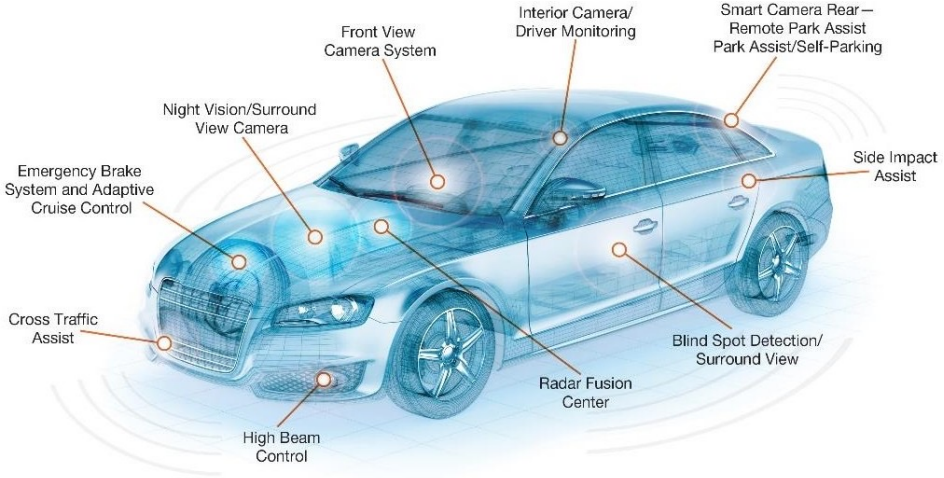
Open Question:

Is driving closer to **chess** or to **everyday conversation**?



Chess Pieces: Self-Driving Car Sensors





External

1. Radar
2. Visible-light camera
3. LIDAR
4. Infrared camera
5. Stereo vision
6. GPS/IMU
7. CAN
8. Audio

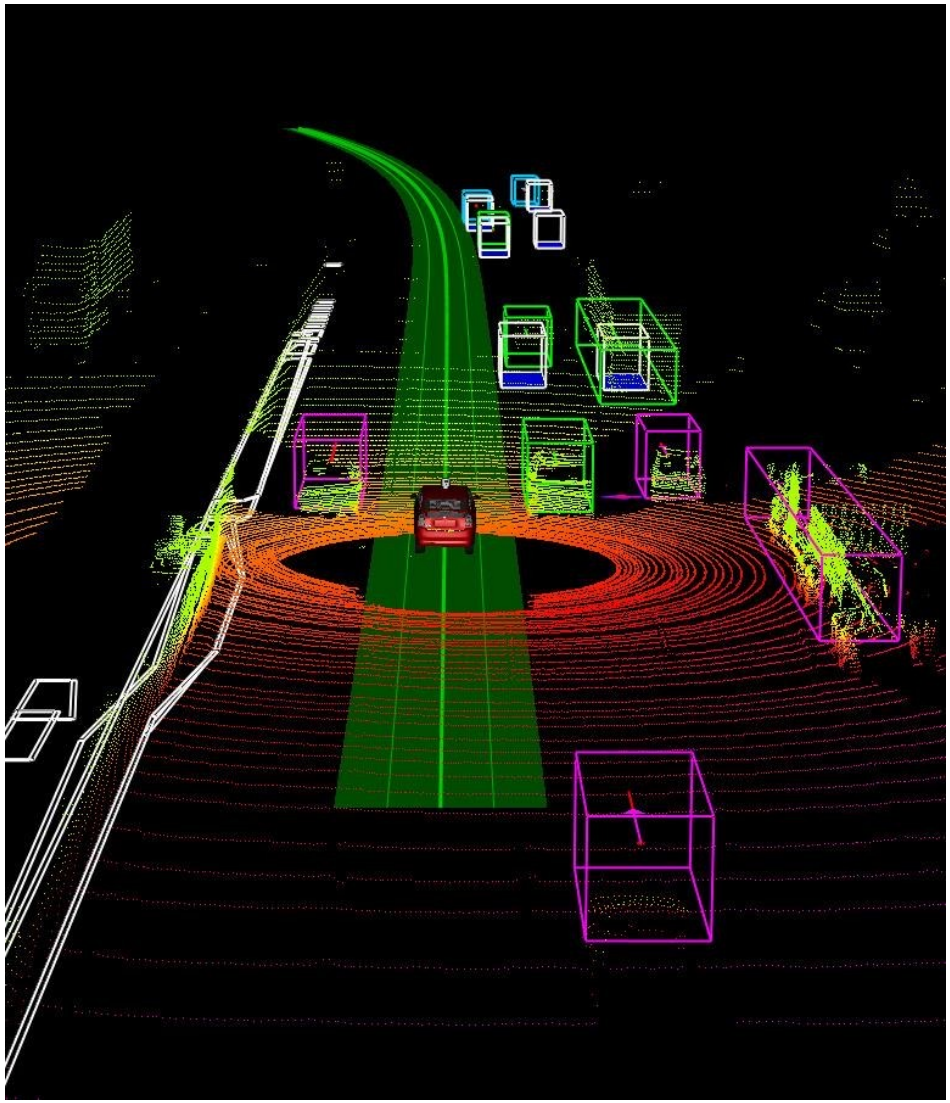
Internal

1. Visible-light camera
2. Infrared camera
3. Audio

References: [71]

Chess Pieces: Self-Driving Car Tasks

- **Localization and Mapping:**
  Where am I?
- **Scene Understanding:**
  Where is everyone else?
- **Movement Planning:**
  How do I get from A to B?
- **Driver State:**
  What's the driver up to?



DARPA Grand Challenge II (2006)

Result: Stanford's Stanley wins

References: [73]

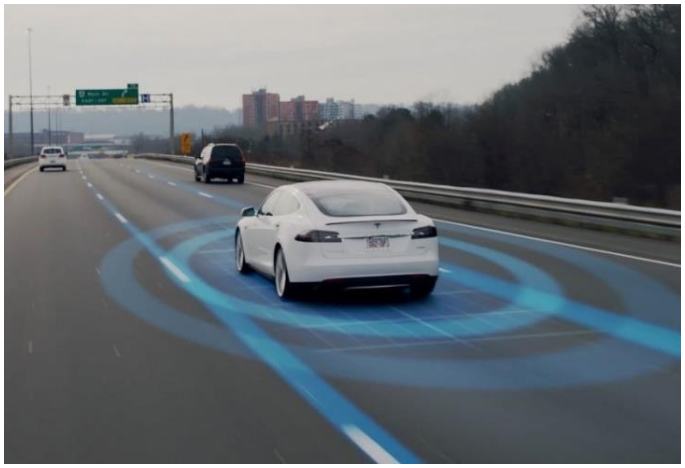
DARPA Urban Challenge (2007)

Result: CMU's Boss (Tartan Racing) wins

References: [74]

Industry Takes on the Challenge

Waymo / Google Self-Driving Car



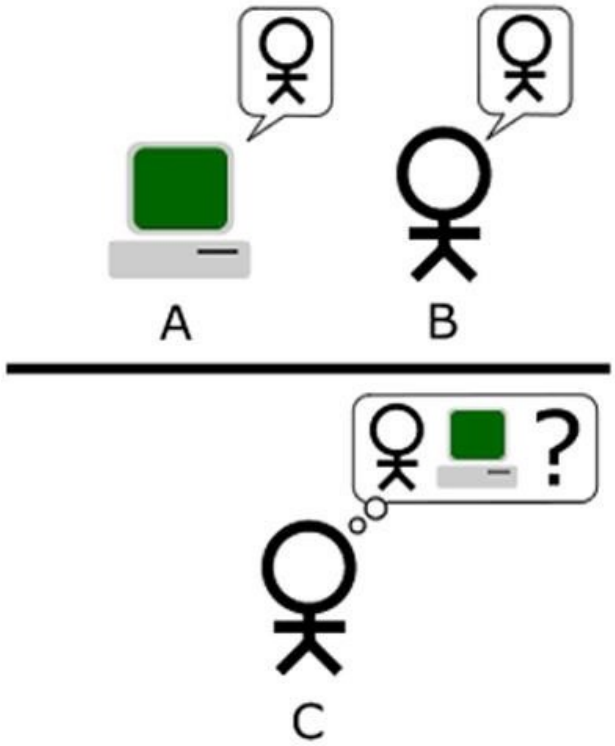
Tesla Autopilot

Uber

nuTonomy

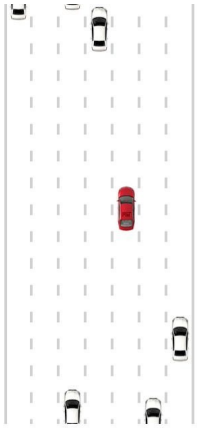
If Driving is a Conversation:

How Hard is it to Pass the Turing Test?

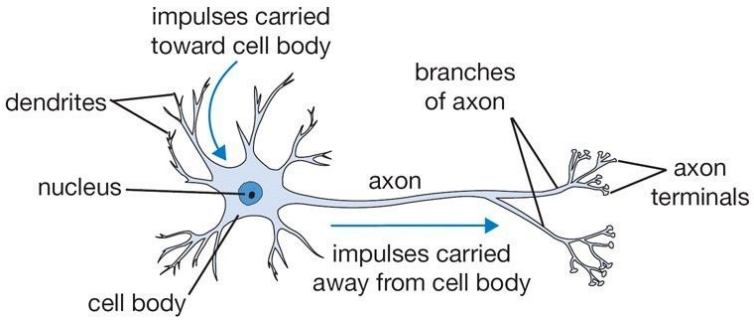


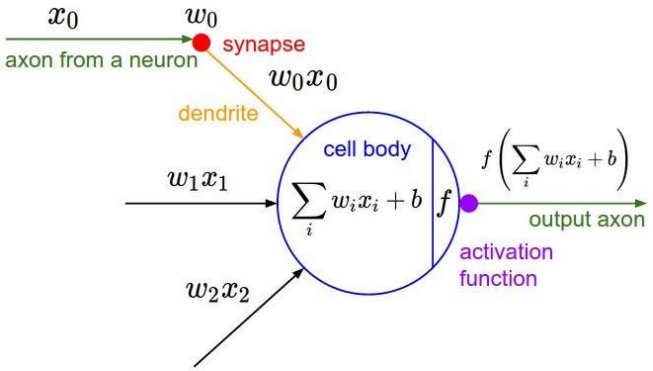
1. **Natural language processing** to enable it to communicate successfully
2. **Knowledge representation** to store information provided before or during the interrogation
3. **Automated reasoning** to use the stored information to answer questions and to draw new conclusions

Turing Test:
Can a computer be mistaken for a human more than 30% of the time?



Neuron: Biological Inspiration for Computation



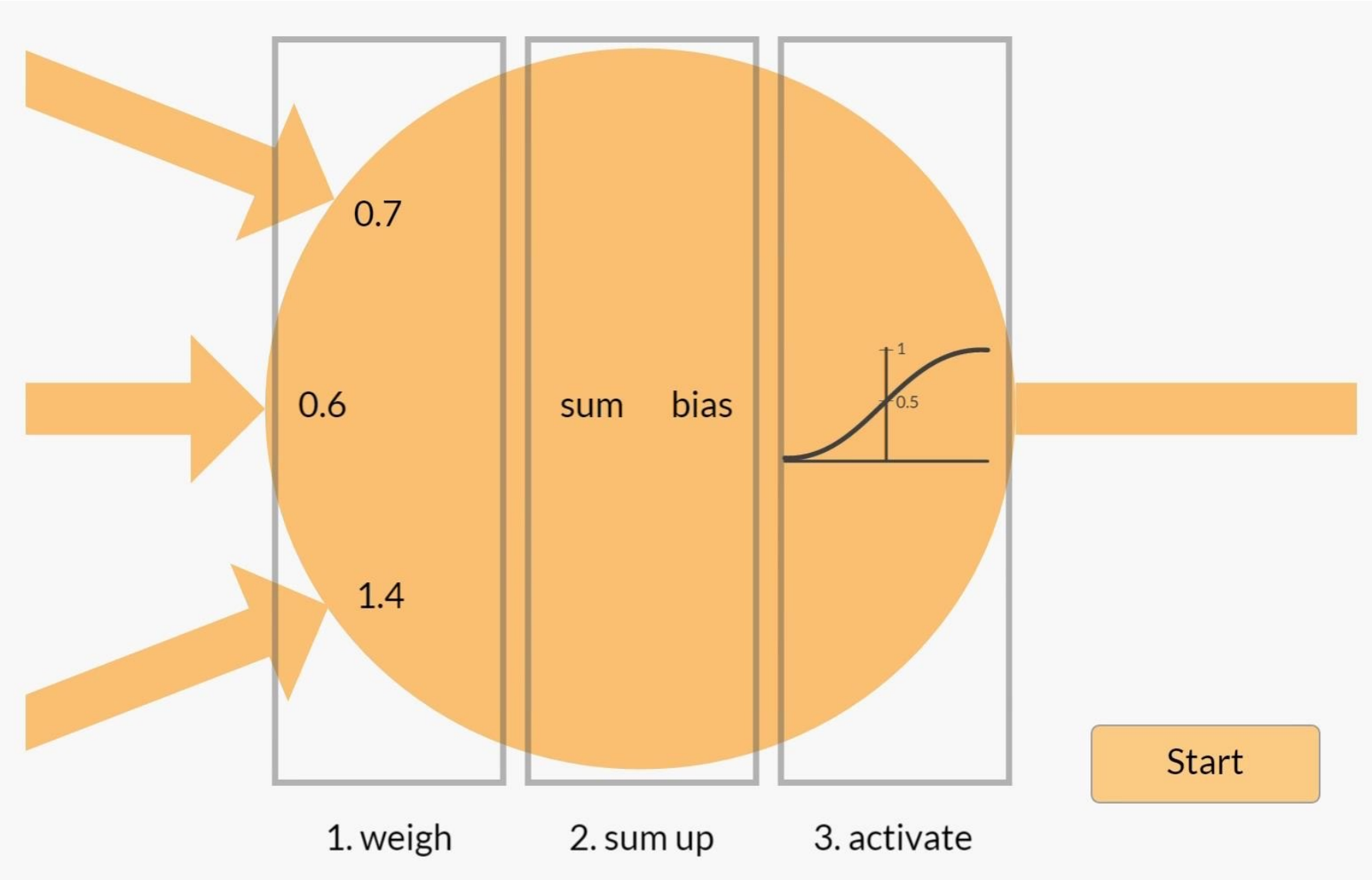


- **Neuron:** computational building block for the brain
- Human brain:
  - ~100-1,000 trillion synapses

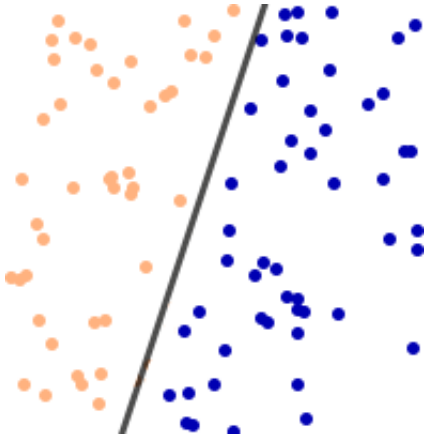
- **(Artificial) Neuron:** computational building block for the "neural network"
- (Artificial) neural network:
  - ~1-10 billion synapses

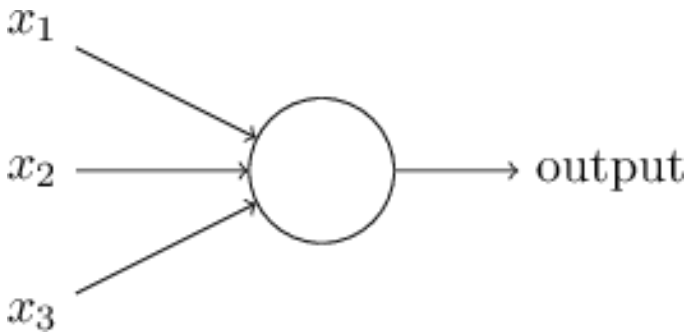
Human brains have ~10,000 computational power than computer brains.

Perceptron: Forward Pass



Perceptron Algorithm



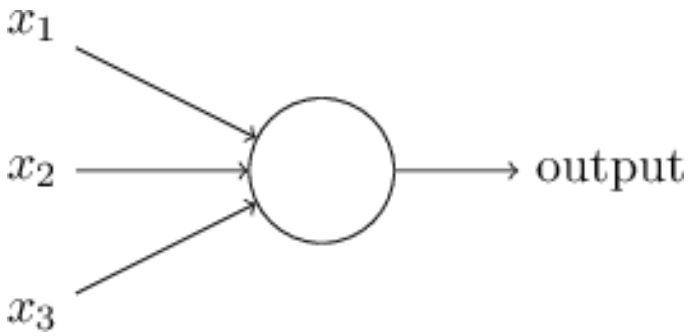


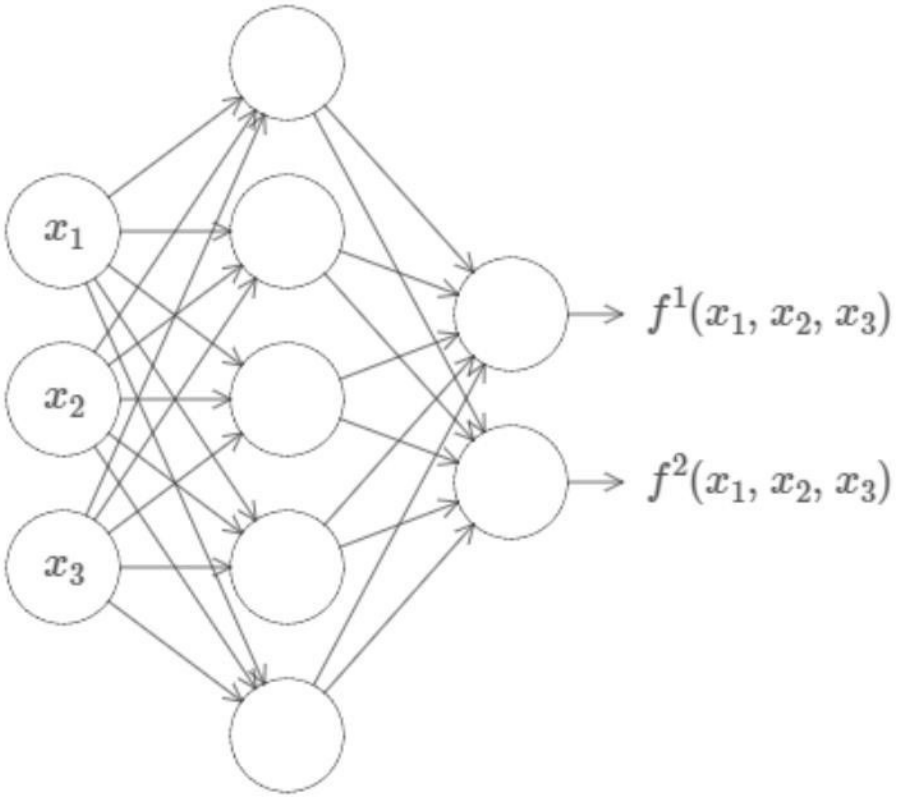
Provide training set of (input, output) pairs and run:

1. Initialize perceptron with random weights
2. For the inputs of an example in the training set, compute the Perceptron's output
3. If the output of the Perceptron does not match the output that is known to be correct for the example:
    1. If the output should have been 0 but was 1, decrease the weights that had an input of 1.
    2. If the output should have been 1 but was 0, increase the weights that had an input of 1.
4. Go to the next example in the training set and repeat steps 2-4 until the Perceptron makes no more mistakes

References: [22, 78]

Neural Networks are Amazing



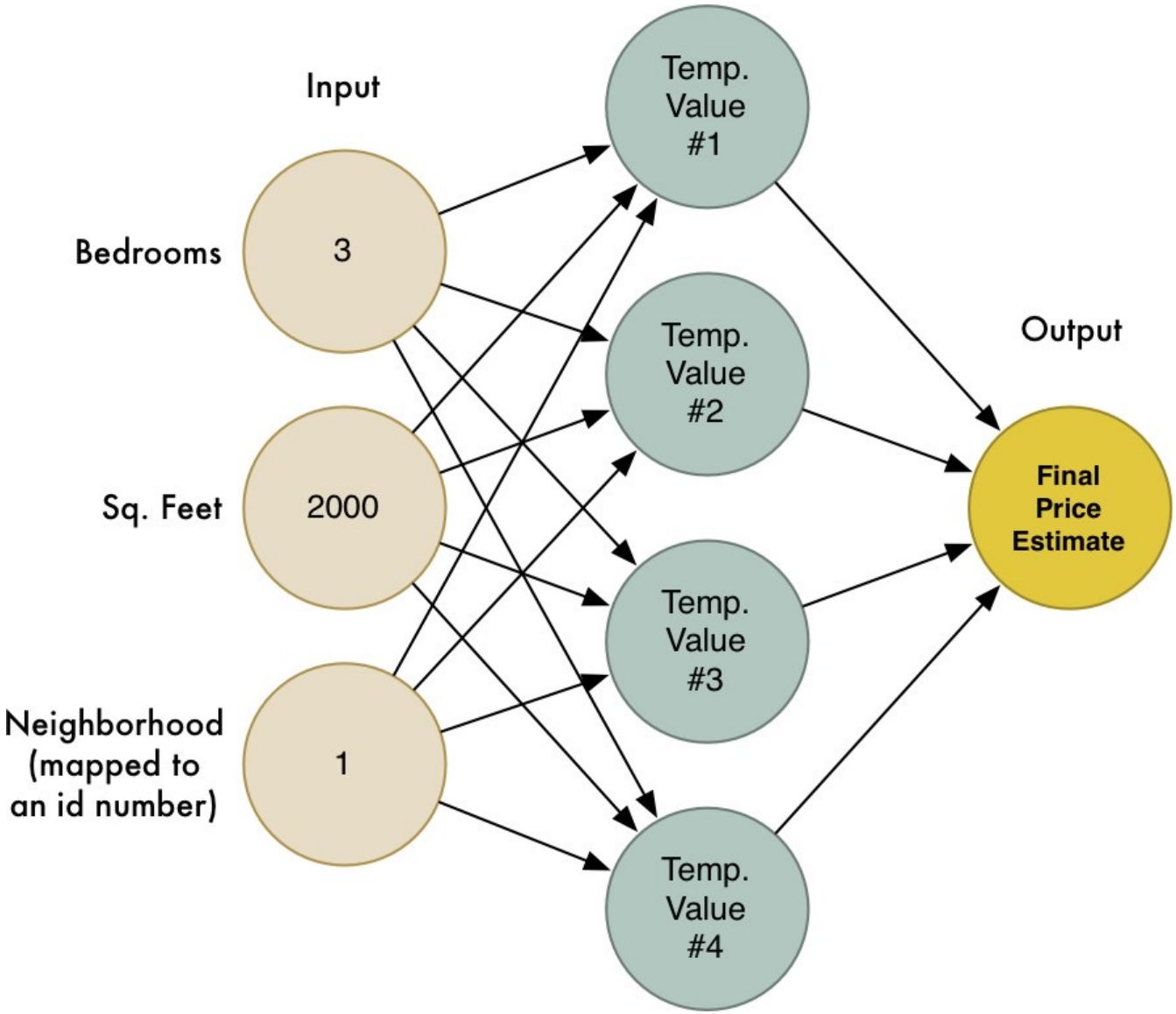


Universality: For any arbitrary function f(x), there exists a neural network that closely approximate it for any input x

Universality is an incredible property!* And it holds for just 1 hidden layer.

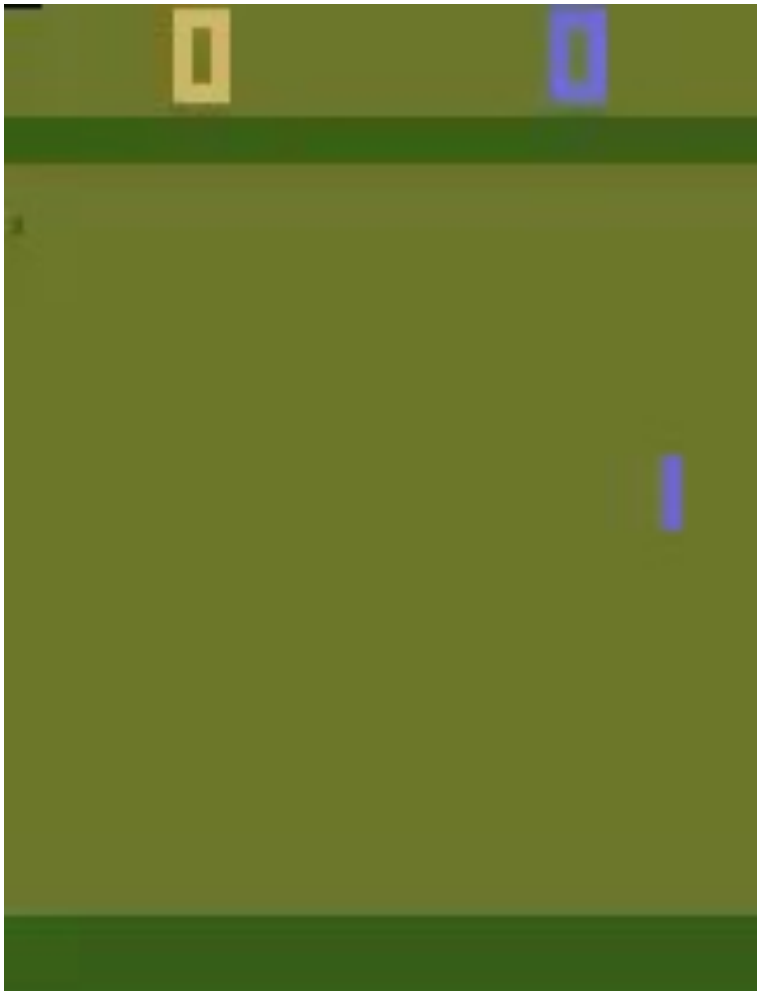
* Given that we have good algorithms for training these networks.

Special Purpose Intelligence



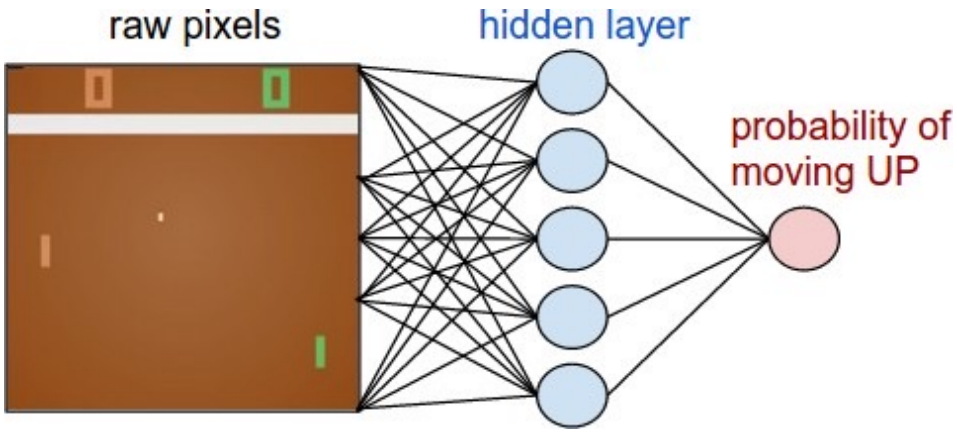
Neural Networks are Amazing:

General Purpose Intelligence



Andrej Karpathy. "Deep Reinforcement Learning: Pong from Pixels." 2016.

Policy Network:

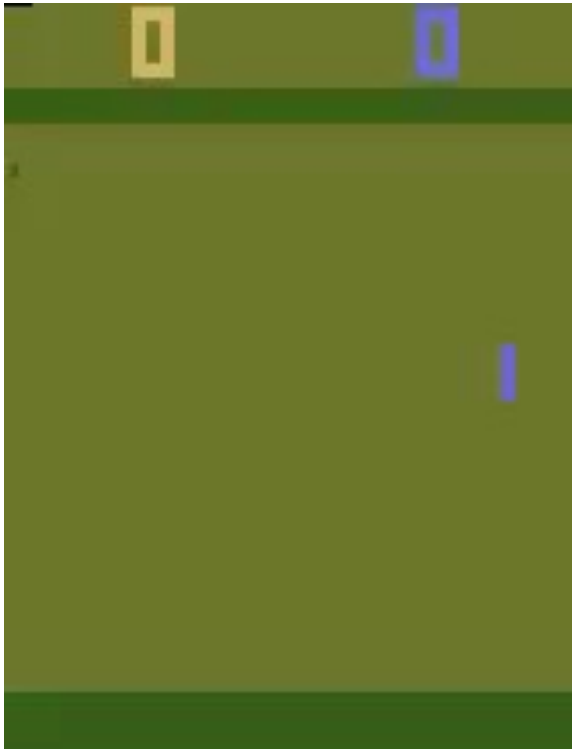


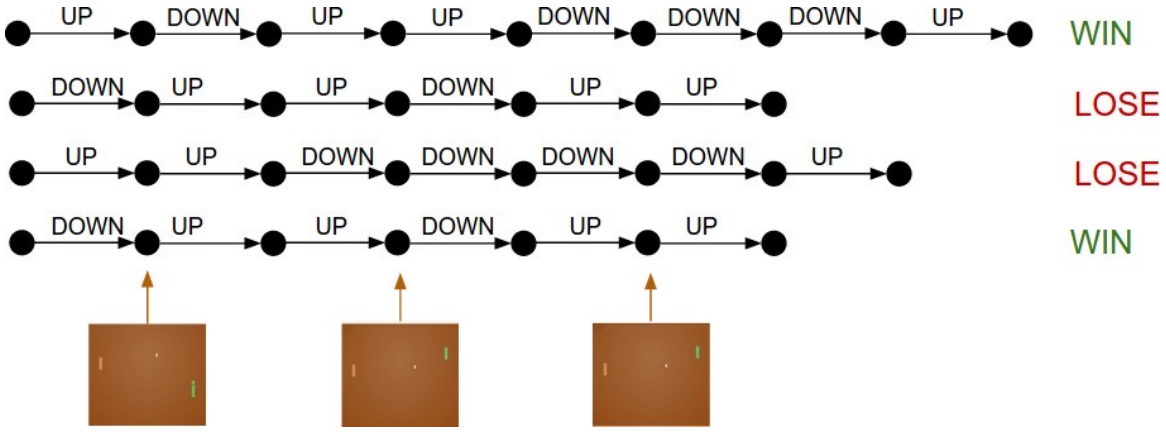
- 80x80 image (difference image)
- 2 actions: up or down
- 200,000 Pong games

This is a step towards general purpose artificial intelligence!

Neural Networks are Amazing:

General Purpose Intelligence





- Every (state, action) pair is **rewarded** when the final result is a **win**.
- Every (state, action) pair is **punished** when the final result is a **loss**.

The fact that this works at all is amazing!

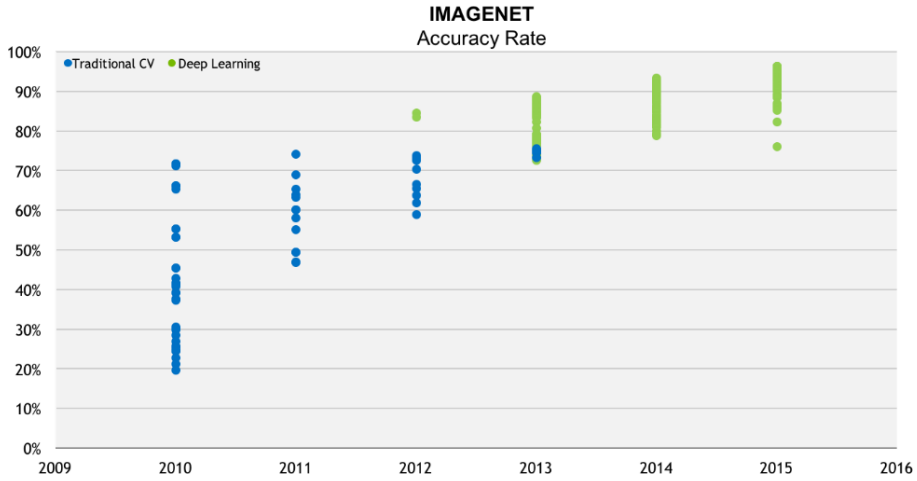
It could be called “general intelligence” but not yet “human-level” intelligence…

Current Drawbacks

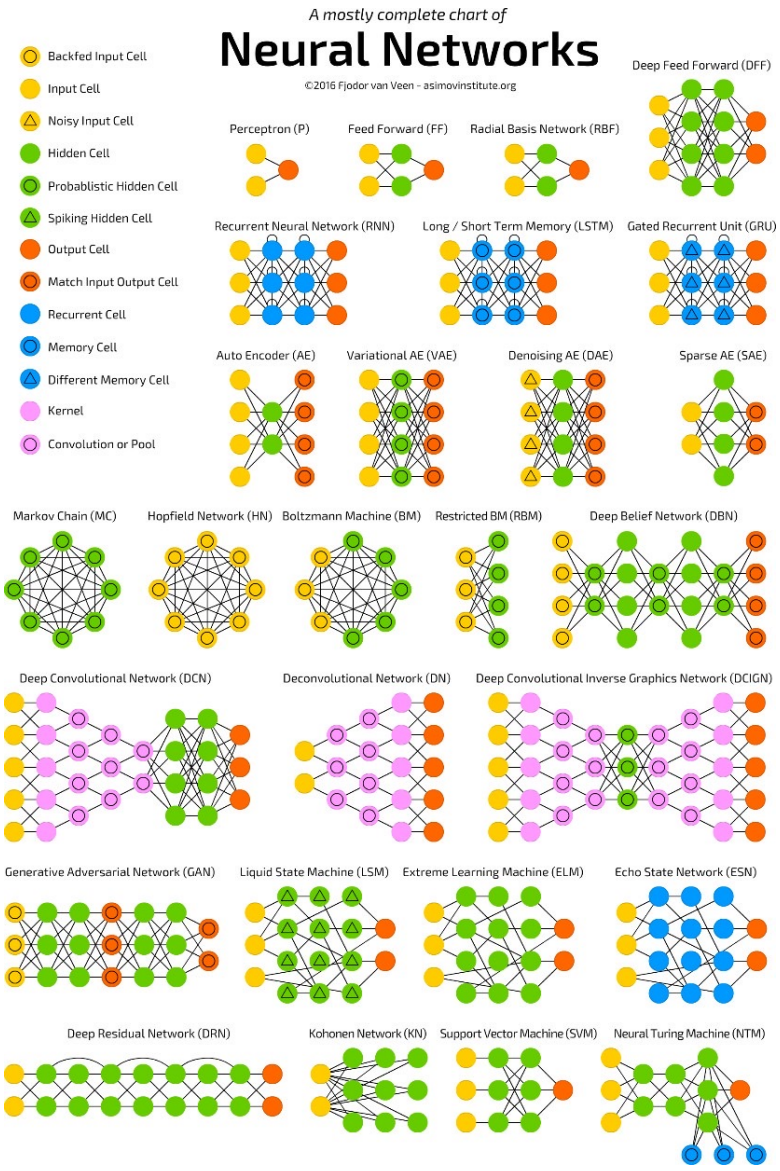
- Lacks Reasoning:
  - Humans only need simple instructions:
    "You're in **control** of a paddle and you can move it up and down, and your task is to bounce the ball past the other player controlled by AI."
- Requires **big** data: inefficient at learning from data
- Requires **supervised** data: costly to annotate real-world data
- Need to manually select network structure
- Needs hyperparameter tuning for training:
  - Learning rate
  - Loss function
  - Mini-batch size
  - Number of training iterations
  - Momentum: gradient update smoothing
  - Optimizer selection
- Defining a good reward function is difficult…

Deep Learning Breakthroughs: What Changed?

- **Compute**
  CPUs, GPUs, ASICs
- **Organized large(-ish) datasets**
  Imagenet
- **Algorithms and research:**
  Backprop, CNN, LSTM
- **Software and Infrastructure**
  Git, ROS, PR2, AWS, Amazon Mechanical Turk, TensorFlow, …
- **Financial backing of large companies**
  Google, Facebook, Amazon, …



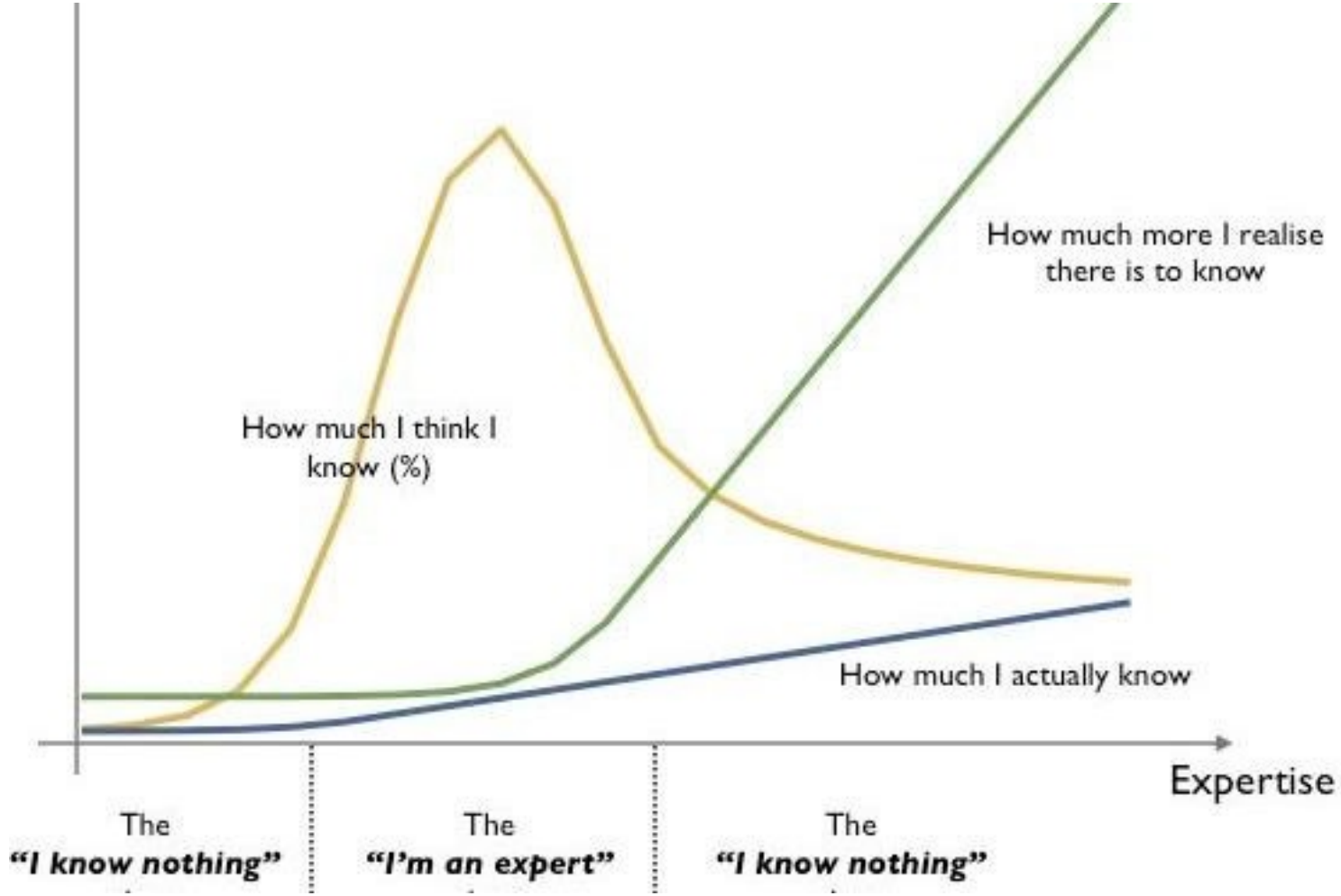
Useful Deep Learning Terms



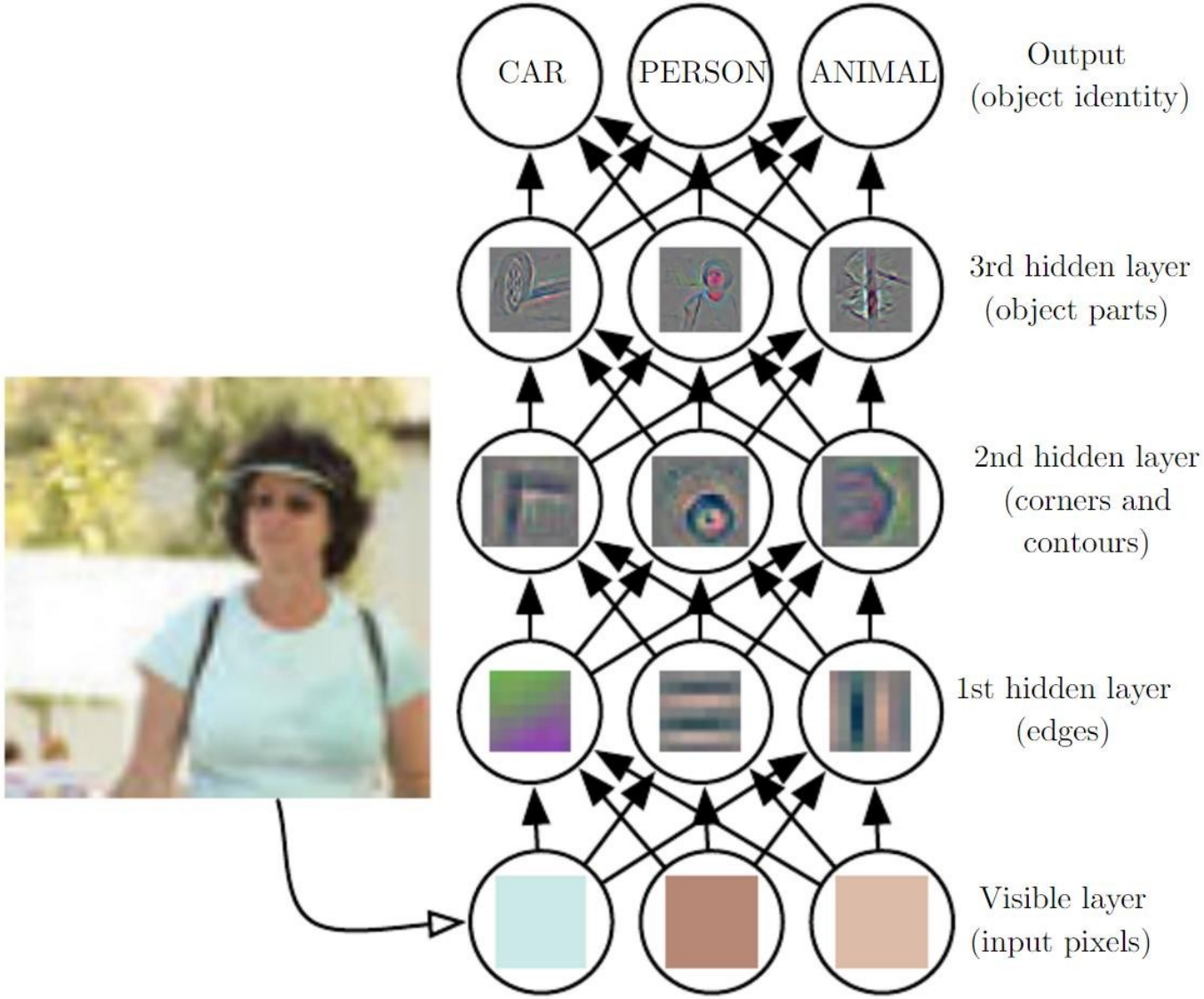
- Basic terms:
  - **Deep Learning = Neural Networks**
  - **Deep Learning** is a subset of **Machine Learning**
- Terms for neural networks:
  - **MLP:** Multilayer Perceptron
  - **DNN:** Deep neural networks
  - **RNN:** Recurrent neural networks
    - **LSTM:** Long Short-Term Memory
  - **CNN or ConvNet:** Convolutional neural networks
  - **DBN:** Deep Belief Networks
- Neural network operations:
  - Convolution
  - Pooling
  - Activation function
  - Backpropagation

Asimov Institute. "A mostly complete chart of neural networks."

Neural Networks: Proceed with Caution

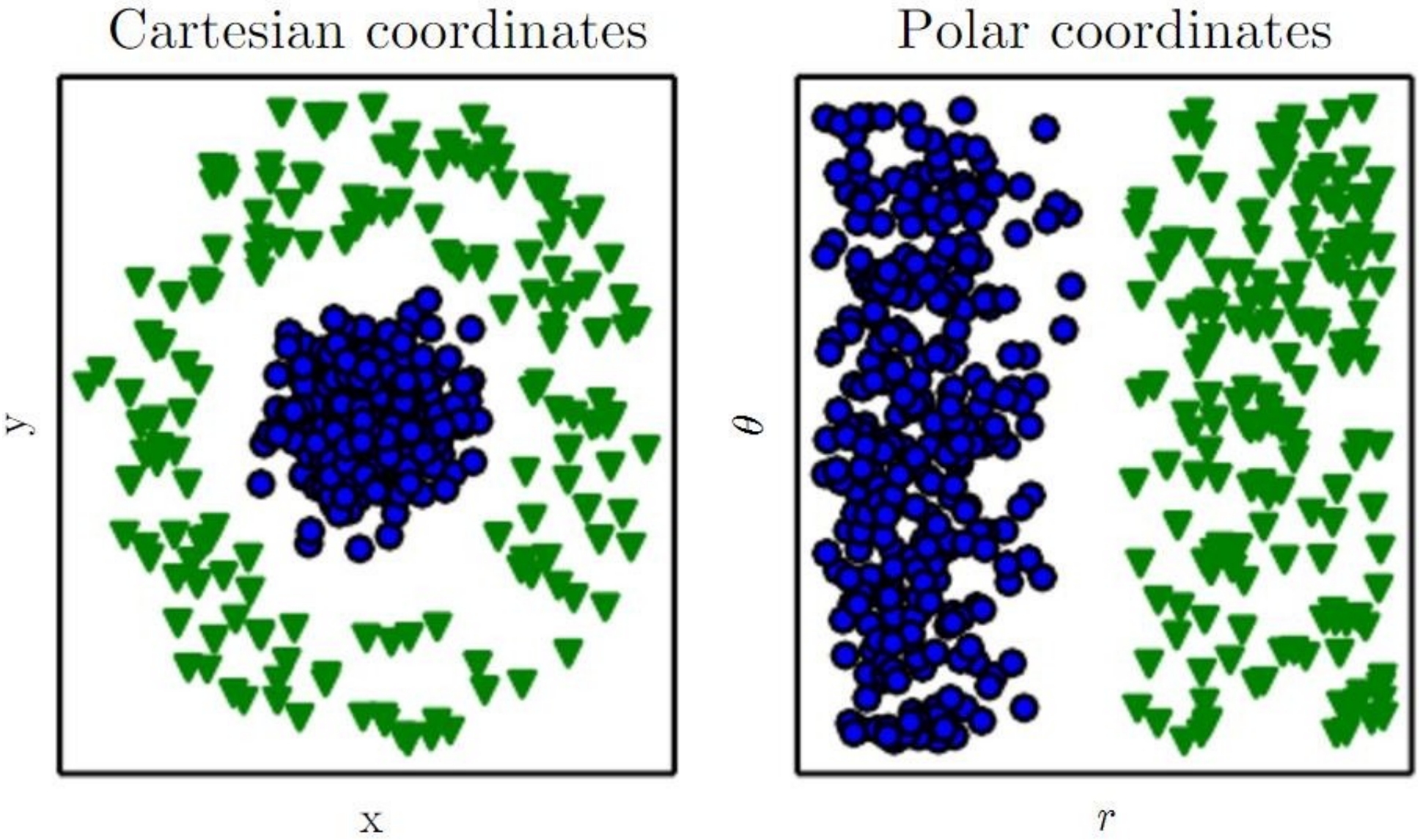


Deep Learning is **Representation Learning**



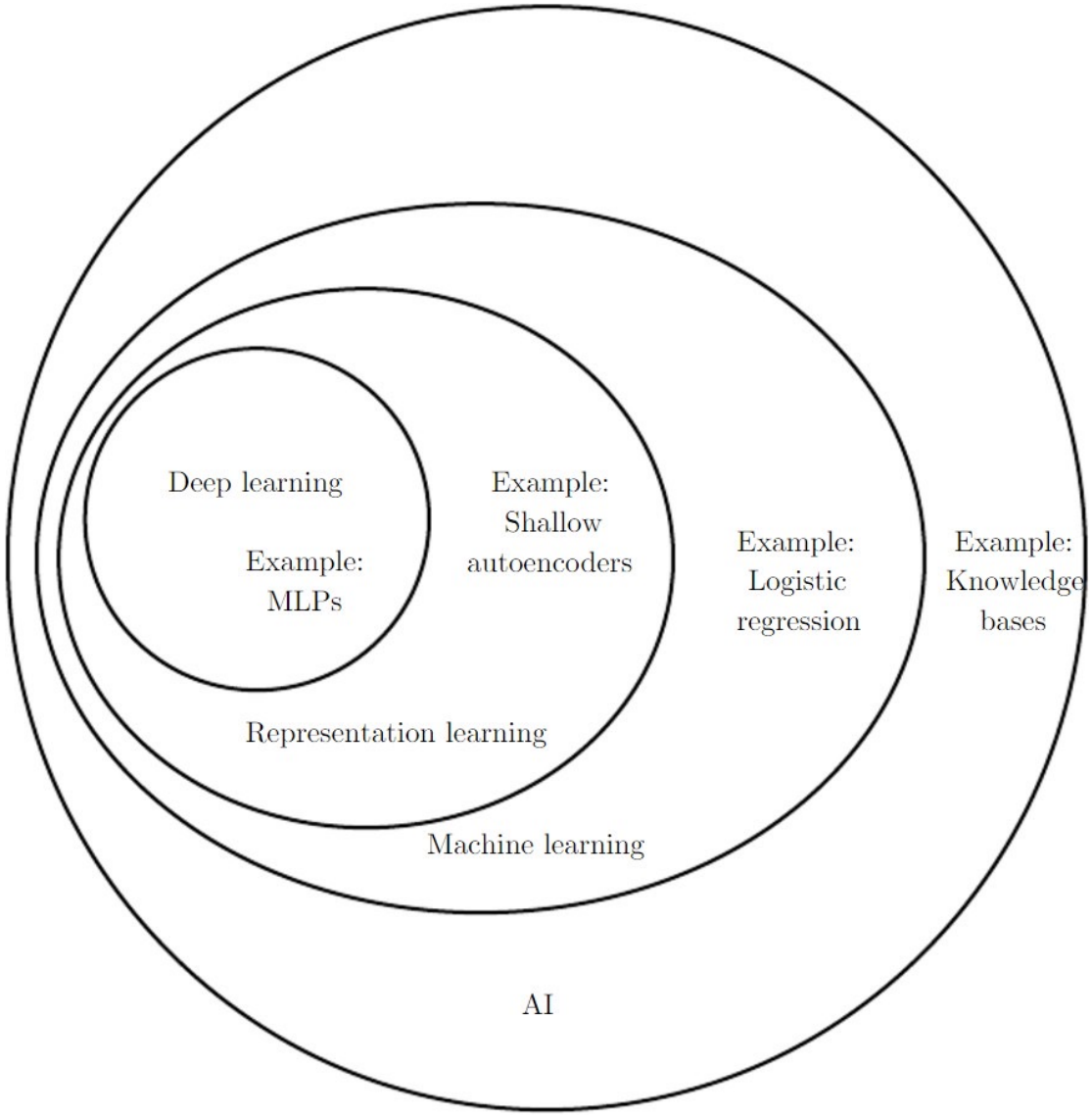
[20] Goodfellow et al. "Deep learning." (2017).

Representation Matters

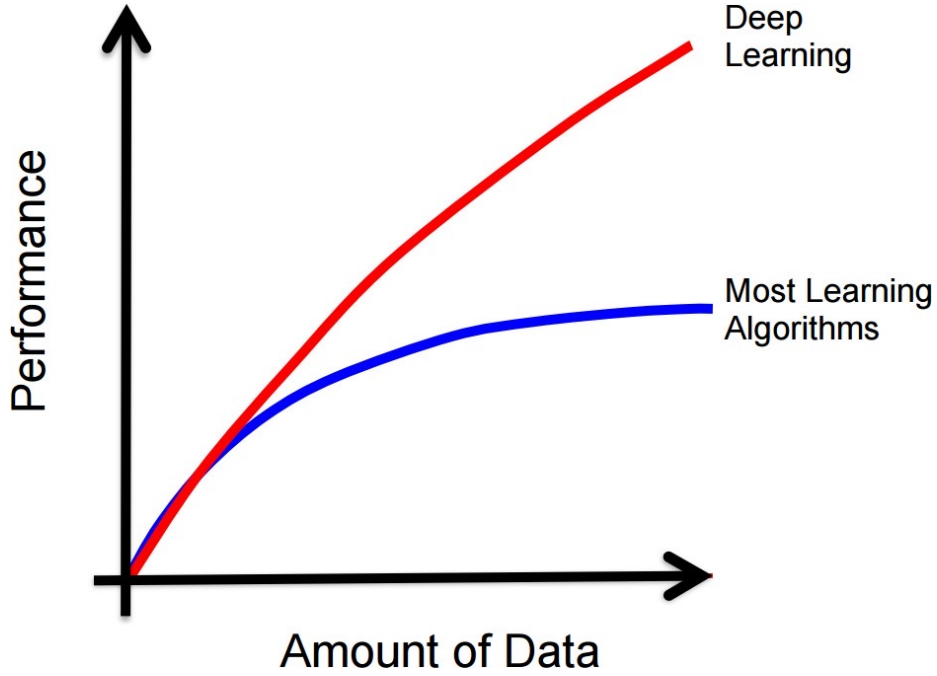


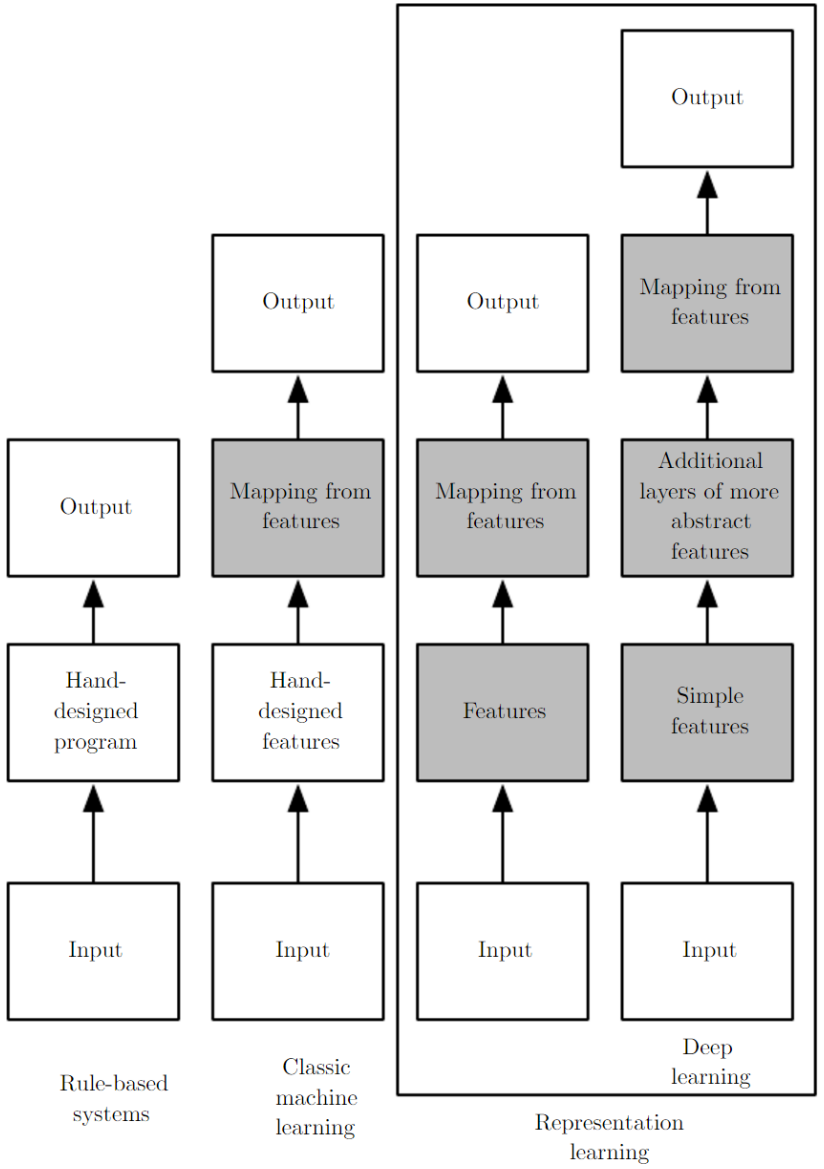
References: [20]

Deep Learning is **Representation Learning**

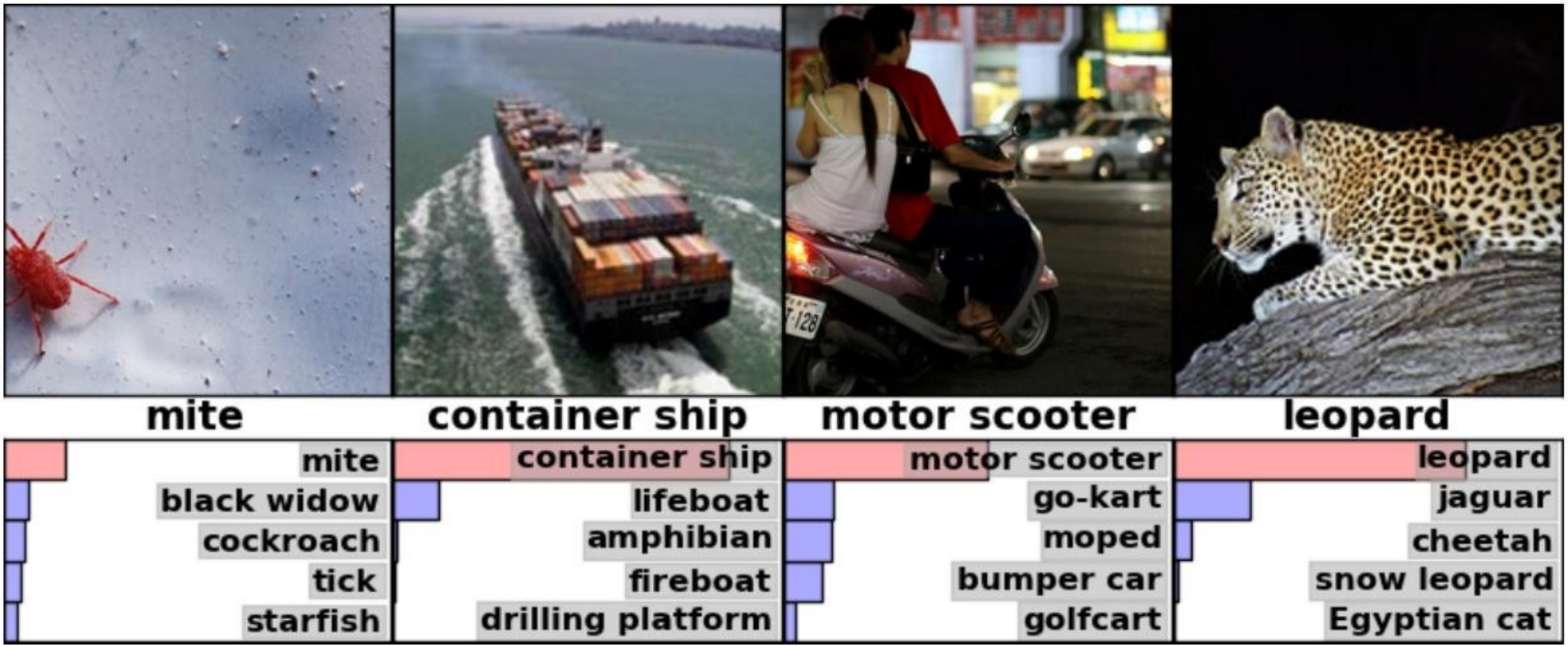


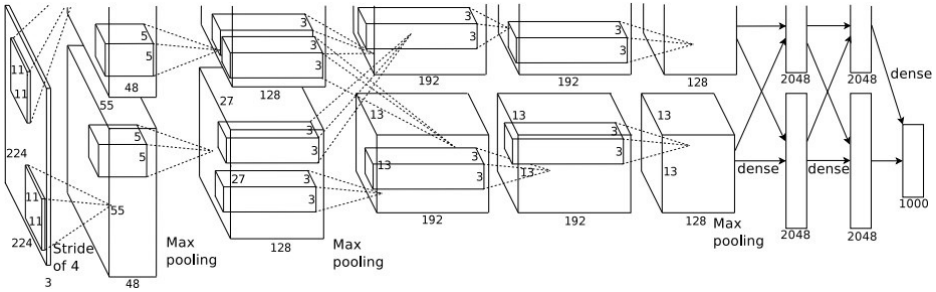
Deep Learning: **Scalable** Machine Learning





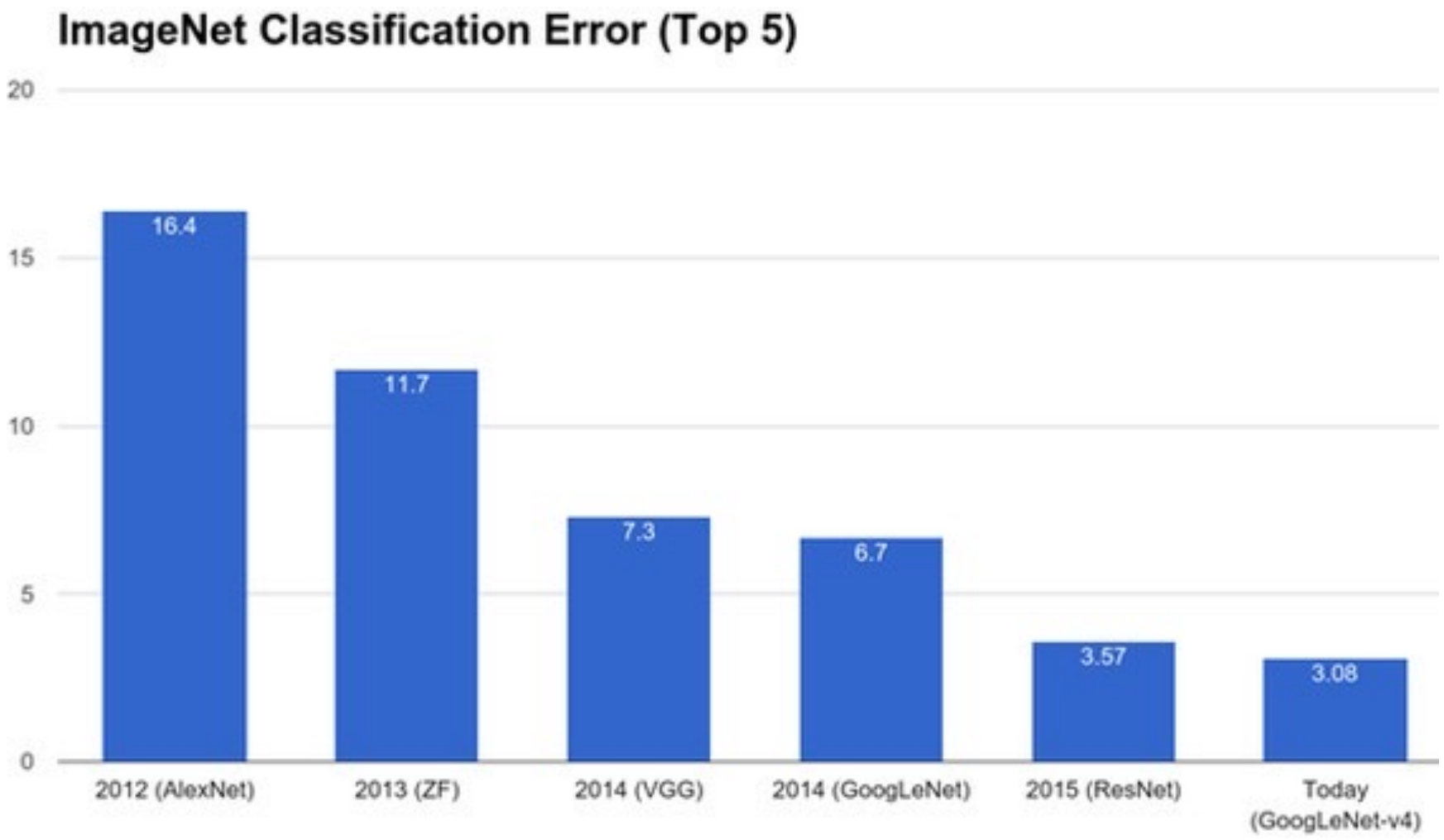
Applications: Object Classification in Images





Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. **"Imagenet classification with deep convolutional neural networks."** Advances in neural information processing systems. 2012.

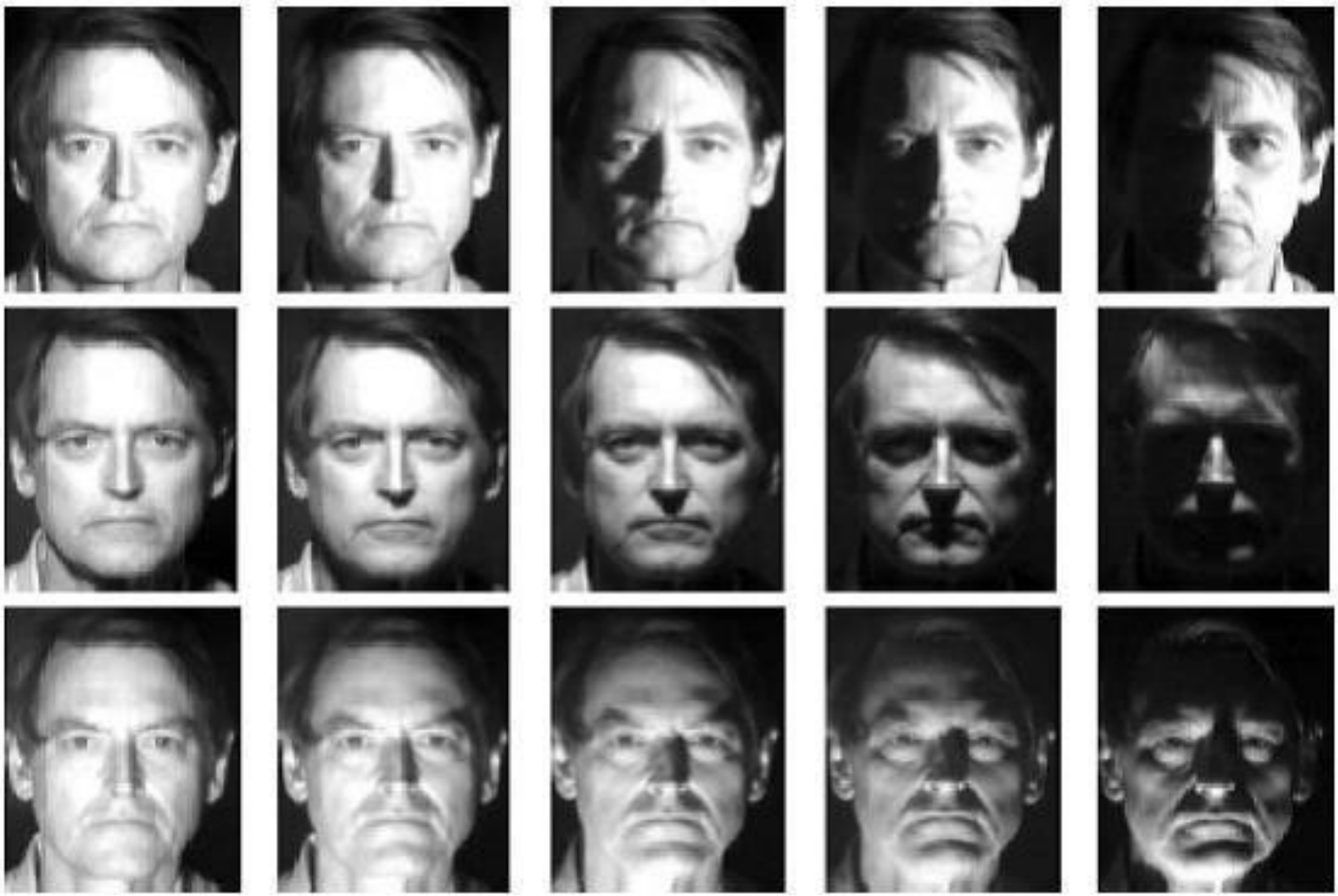
Pause: Progress on ImageNet



References: [55]

Computer Vision is Hard:

Illumination Variability



Computer Vision is Hard:

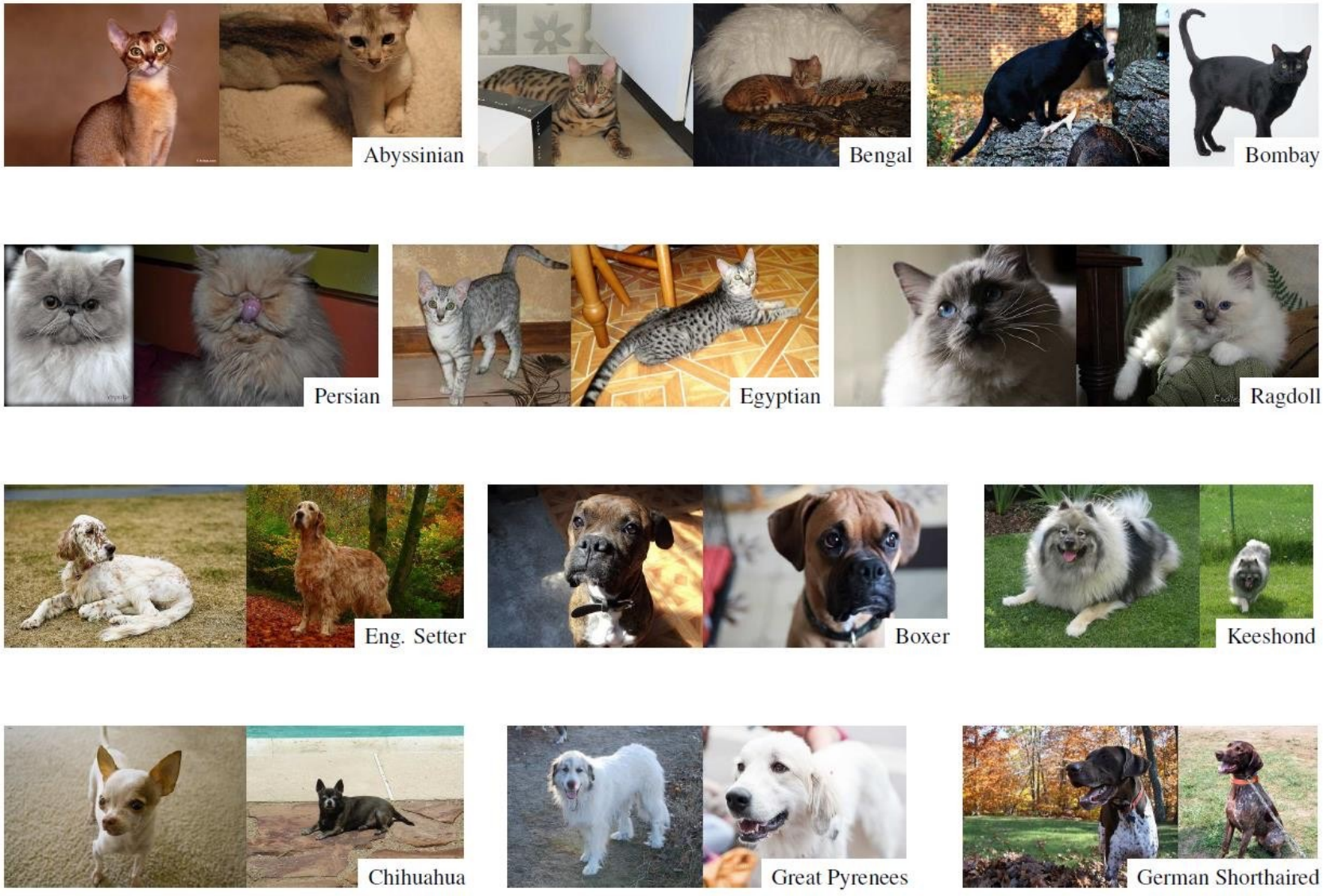
Pose Variability and Occlusions

Figure 1. **The deformable and truncated cat.** Cats exhibit (al-

Parkhi et al. "The truth about cats and dogs." 2011.

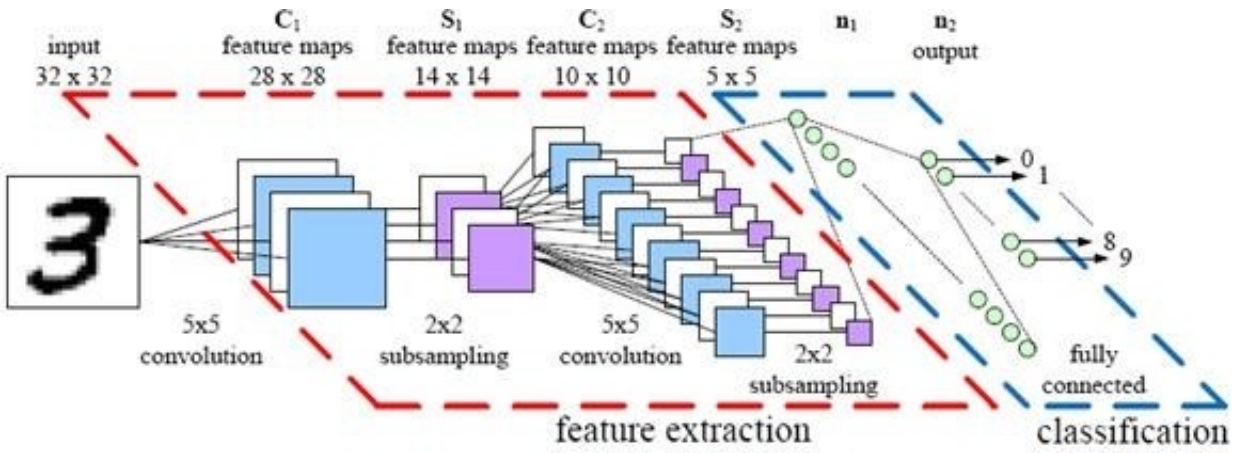
Computer Vision is Hard:

Intra-Class Variability

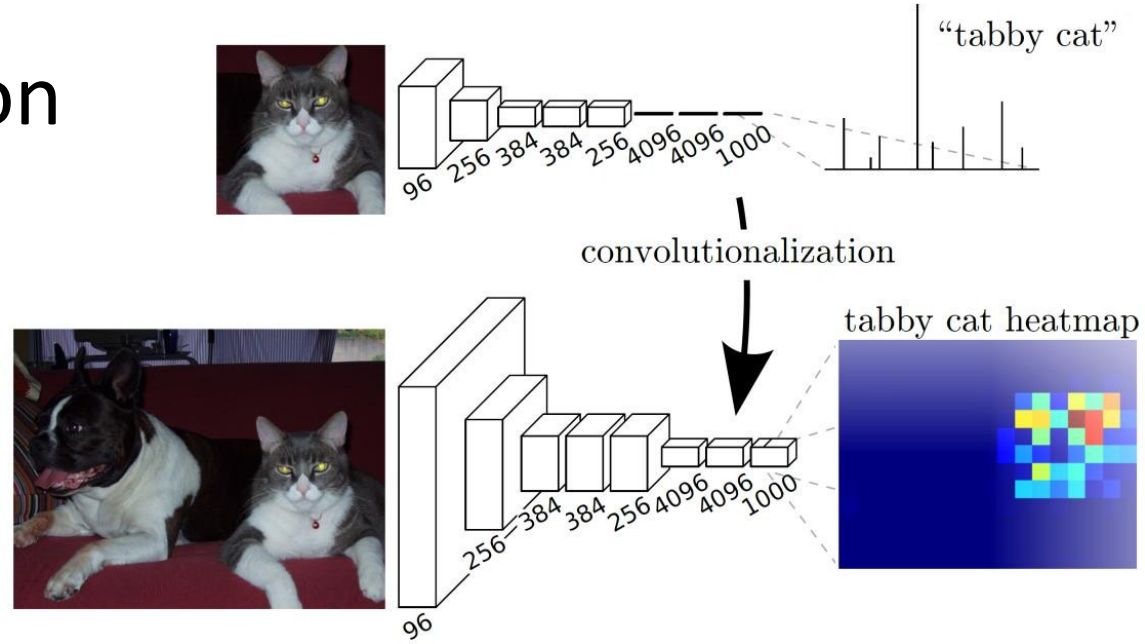


Parkhi et al. "Cats and dogs." 2012.

Pause: Object Recognition / Classification



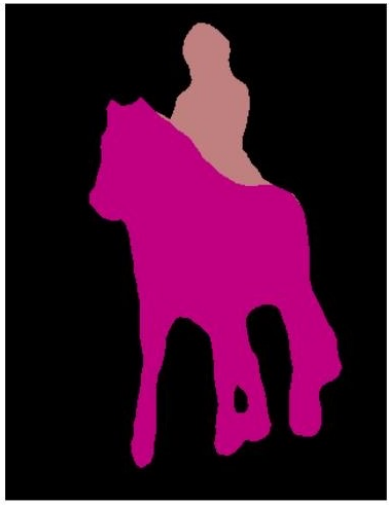
Pause: Segmentation



Original

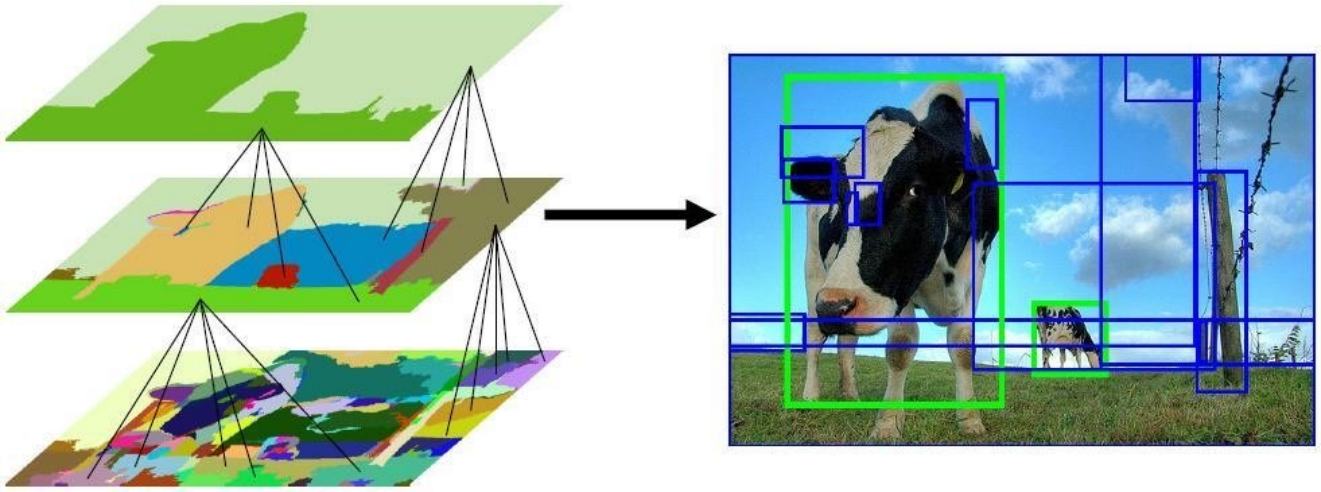
Ground Truth

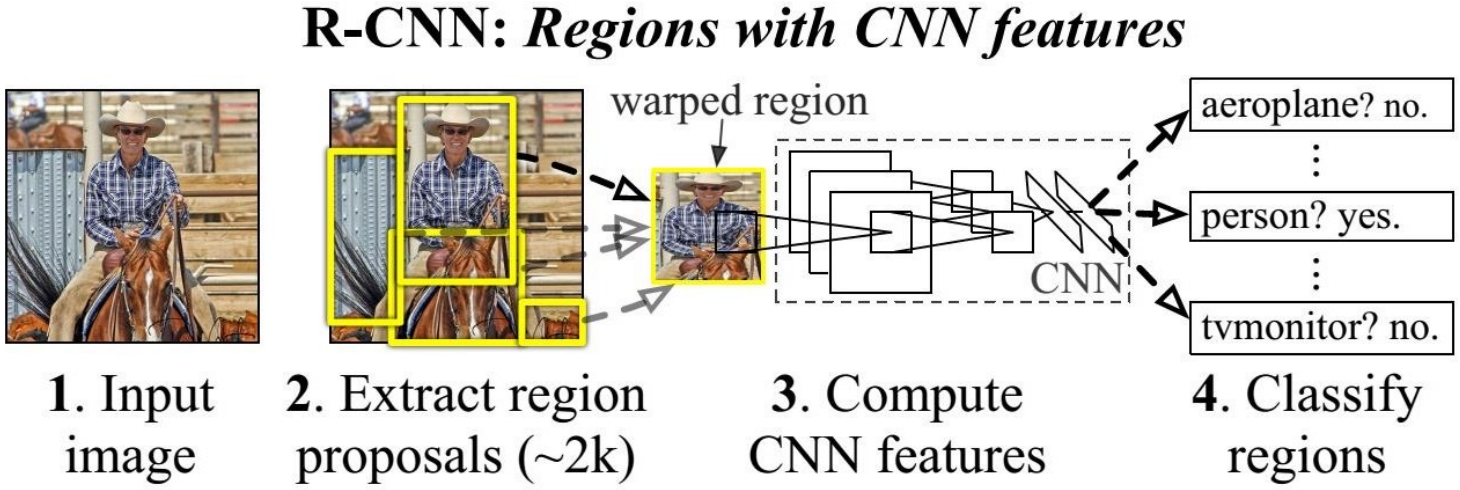
FCN-8



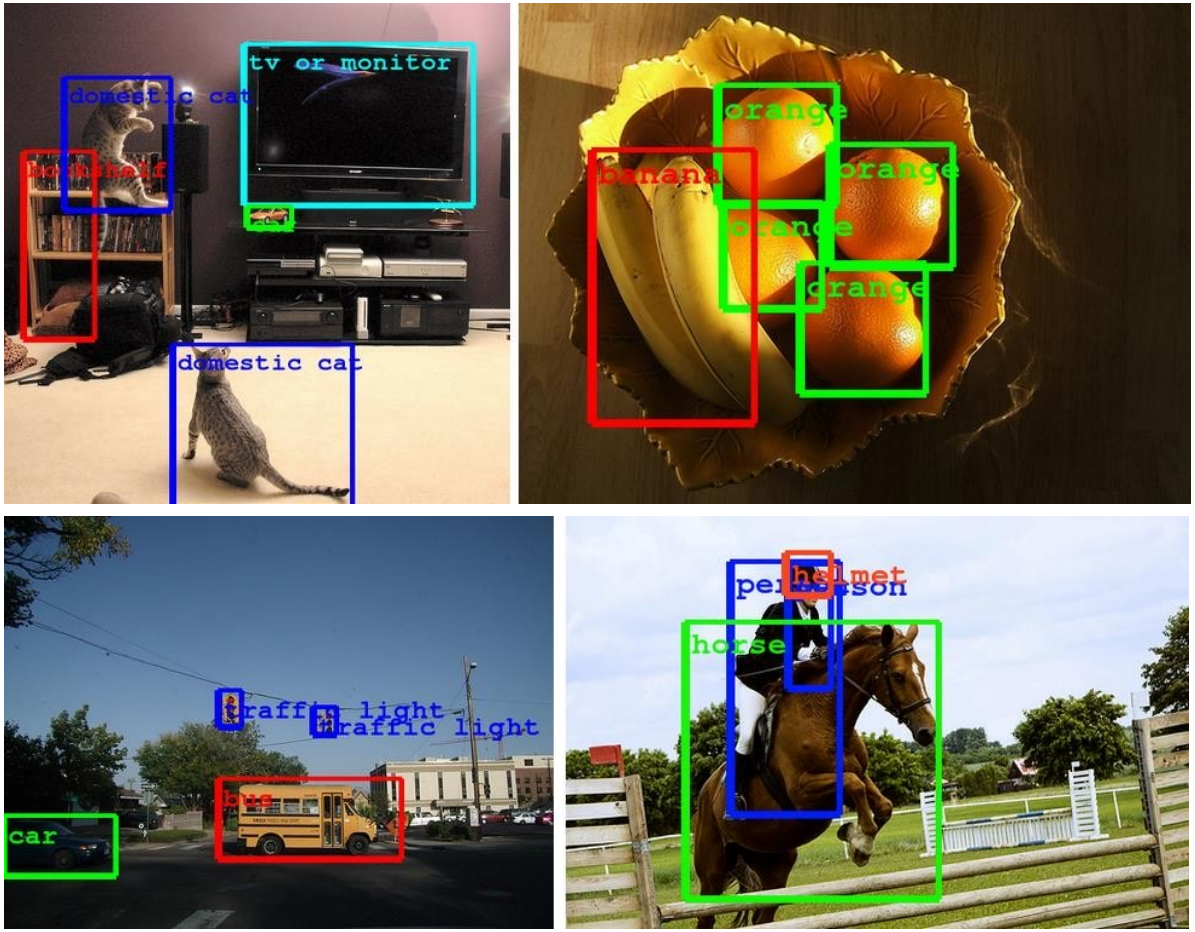
Source: Long et al. Fully Convolutional Networks for Semantic Segmentation. CVPR 2015.

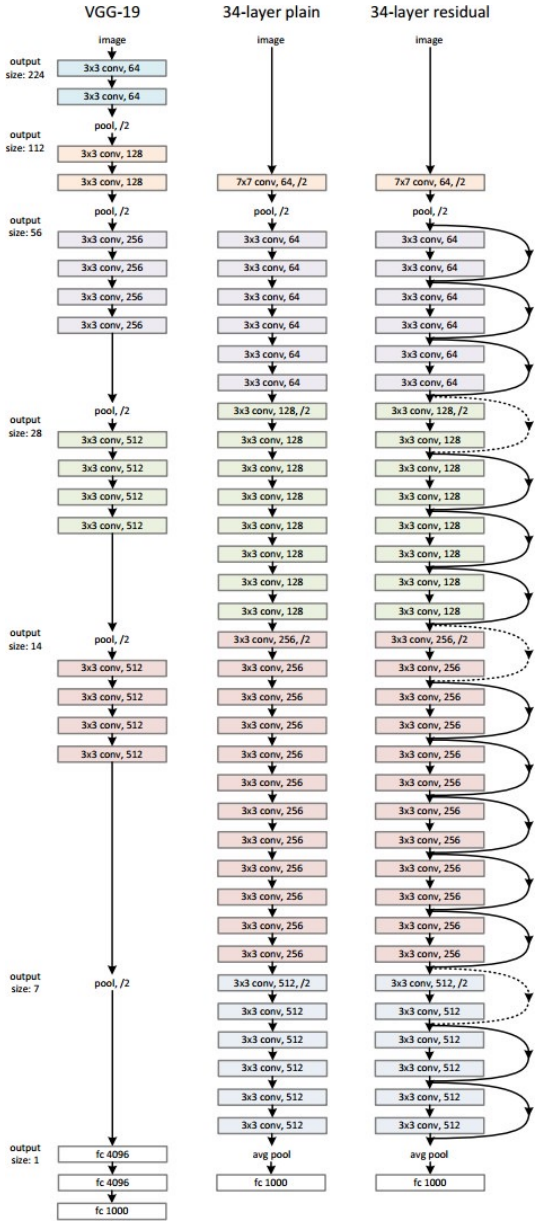
Pause: Object Detection





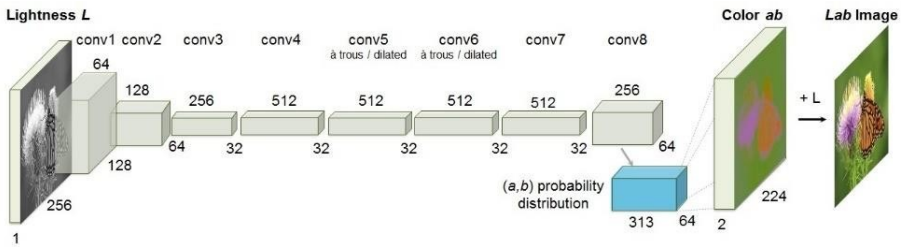
Applications: Object Detection and Localization in Images





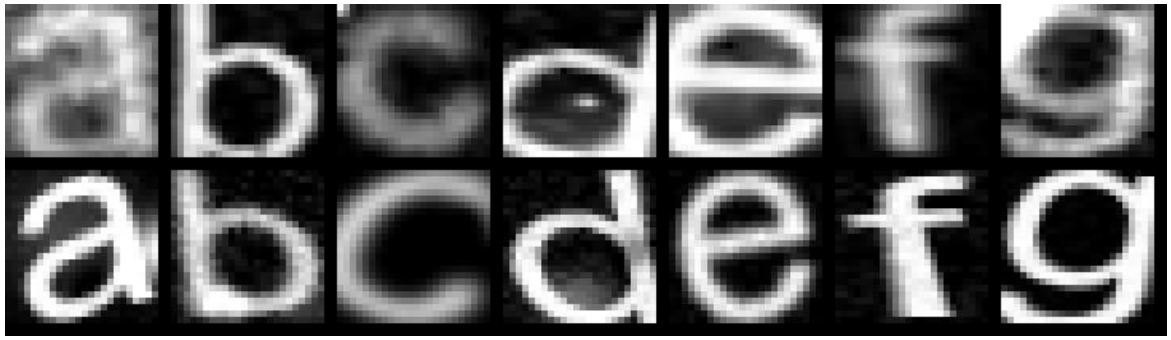
He Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Su. **"Deep residual learning for image recognition."** (2015).

Applications: Colorization of Images



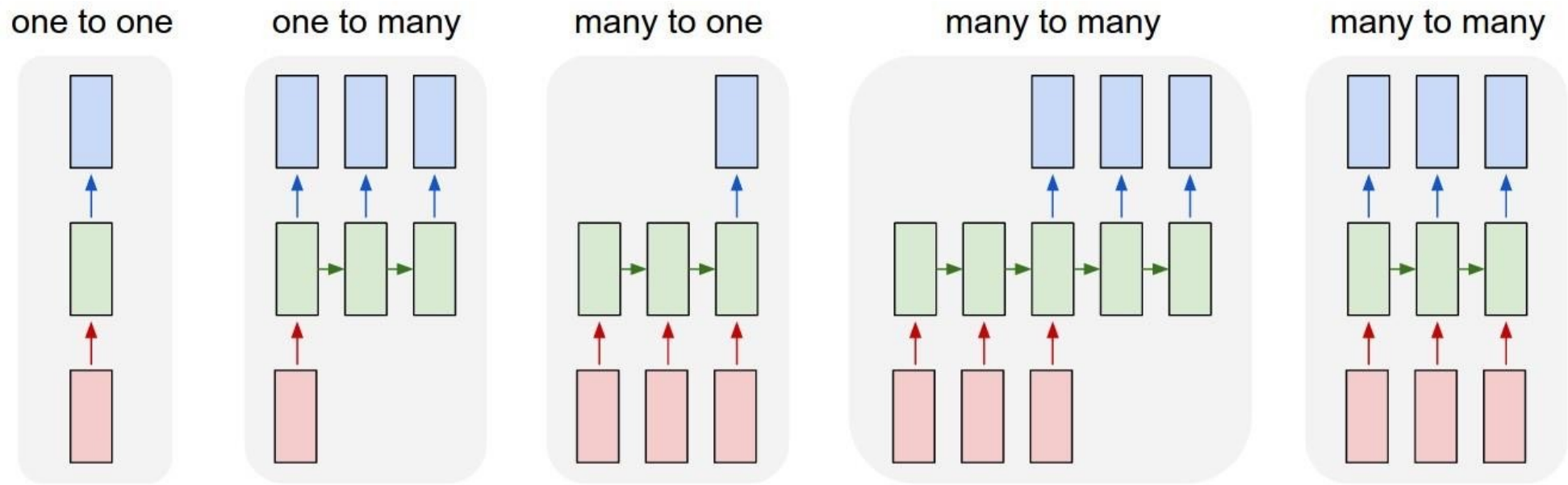
Zhang, Richard, Phillip Isola, and Alexei A. Efros. **"Colorful Image Colorization."** (2016).

Applications: Automatic Translation of Text in Images



Google Translate

(Pause…) Flavors of Neural Networks



"Vanilla" Neural Networks

Recurrent Neural Networks

Andrej Karpathy. **"The Unreasonable Effectiveness of Recurrent Neural Networks."** (2015).

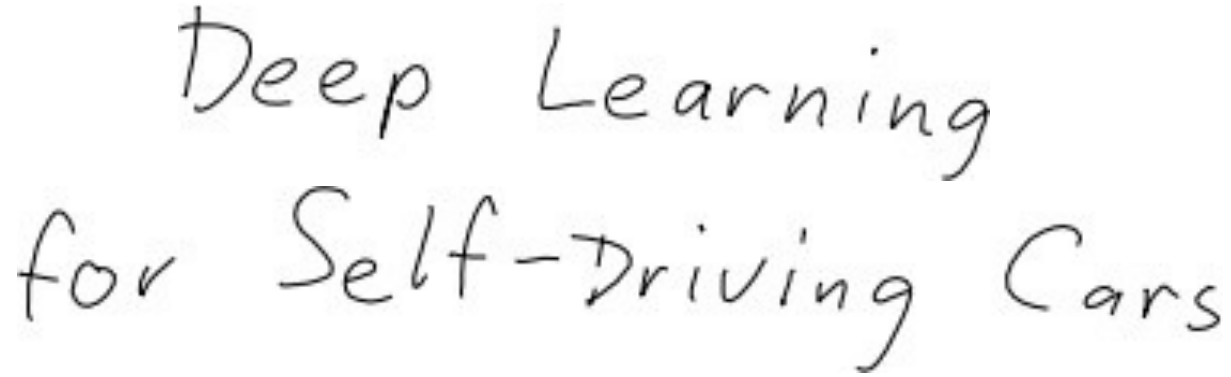
Applications: Handwriting Generation from Text

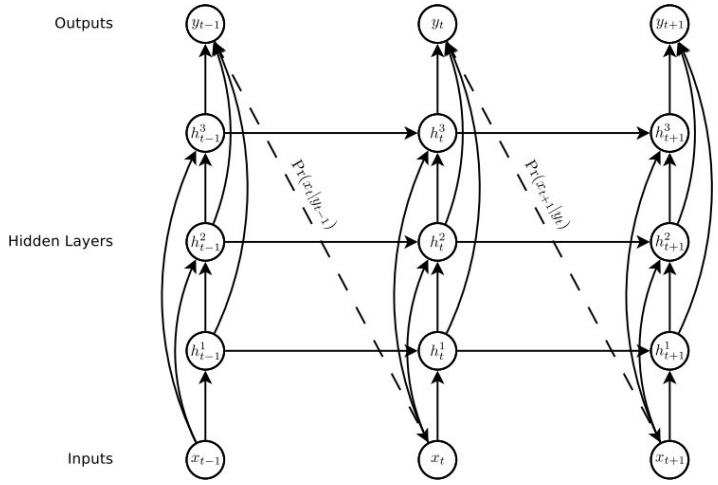
Input:

Text --- up to 100 characters, lower case letters work best

Deep Learning for Self Driving Cars

Output:

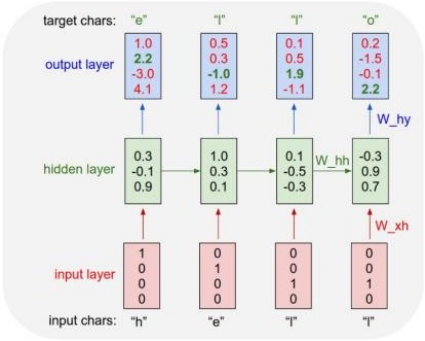




Alex Graves. **"Generating sequences with recurrent neural networks."** (2013).

Applications: Character-Level Text Generation

Naturalism and decision for the majority of Arab countries' capitalide was grounded by the Irish language by [[John Clair]], [[An Imperial Japanese Revolt]], associated with Guangzham's sovereignty. His generals were the powerful ruler of the Portugal in the [[Protestant Immineners]], which could be said to be directly in Cantonese Communication, which followed a ceremony and set inspired prison, training.



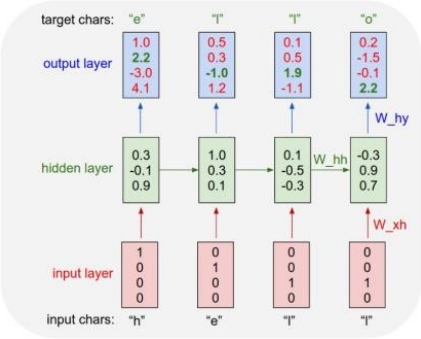
Andrej Karpathy. "The Unreasonable Effectiveness of Recurrent Neural Networks." (2015).

Code: https://github.com/karpathy/char-rnn

Applications: Character-Level Text Generation

Life Is About The Weather!
Life Is About The (Wild) Truth About Human-Rights
Life Is About The True Love Of Mr. Mom
Life Is About Where He Were Now
Life Is About Kids
Life Is About What It Takes If Being On The Spot Is Tough
Life Is About… An Eating Story
Life Is About The Truth Now

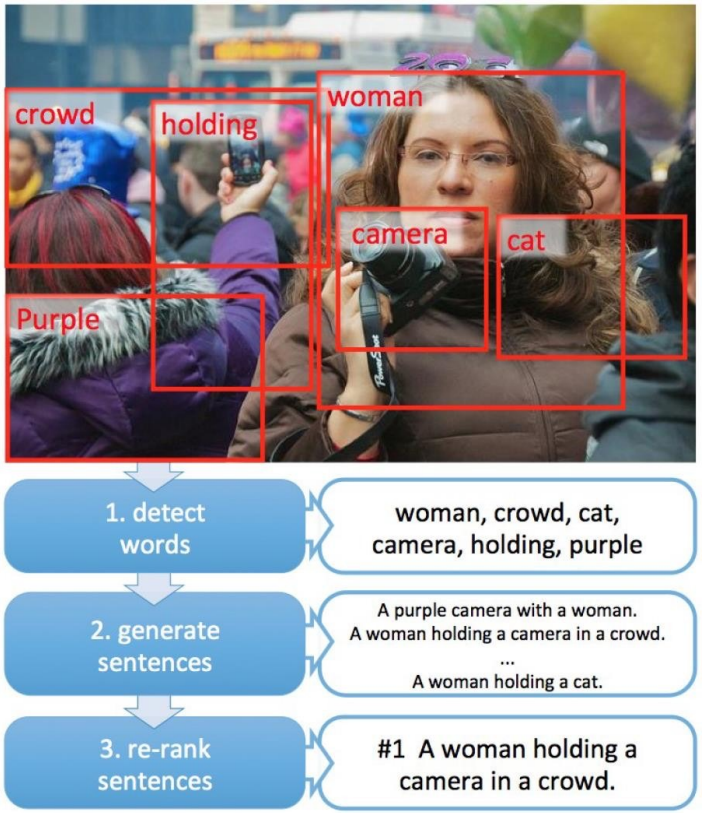
The meaning of life is literary recognition.
The meaning of life is the tradition of the ancient human reproduction



Andrej Karpathy. "The Unreasonable Effectiveness of Recurrent Neural Networks." (2015).

Code: https://github.com/karpathy/char-rnn

Applications: Image Caption Generation

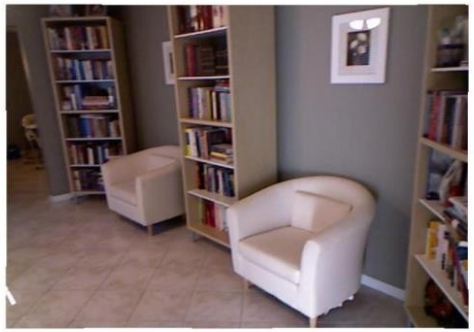


References: [43 – Fang et al. 2015]

Applications: Image Question Answering

COCOQA 33827
What is the color of the cat?
Ground truth: black
IMG+BOW: black (0.55)
2-VIS+LSTM: black (0.73)
BOW: gray (0.40)

COCOQA 33827a
What is the color of the couch?
Ground truth: red
IMG+BOW: red (0.65)
2-VIS+LSTM: black (0.44)
BOW: red (0.39)



DAQUAR 1522
How many chairs are there?
Ground truth: two
IMG+BOW: four (0.24)
2-VIS+BLSTM: one (0.29)
LSTM: four (0.19)

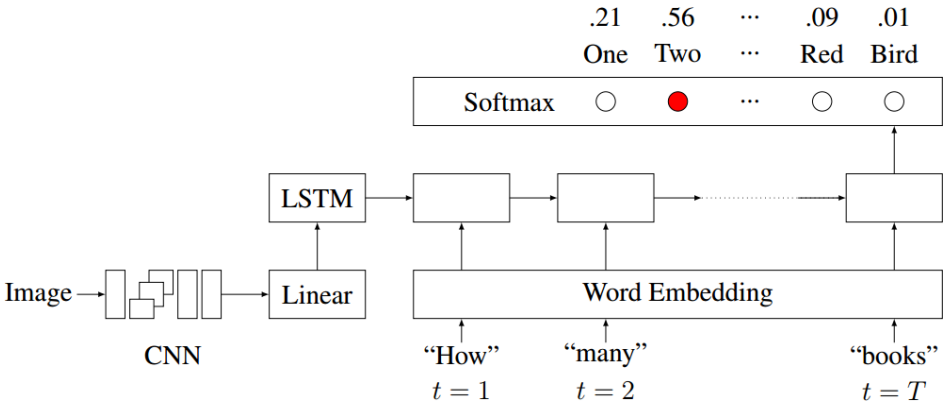
DAQUAR 1520
How many shelves are there?
Ground truth: three
IMG+BOW: three (0.25)
2-VIS+BLSTM: two (0.48)
LSTM: two (0.21)

COCOQA 14855
Where are the ripe bananas sitting?
Ground truth: basket
IMG+BOW: basket (0.97)
2-VIS+BLSTM: basket (0.58)
BOW: bowl (0.48)

COCOQA 14855a
What are in the basket?
Ground truth: bananas
IMG+BOW: bananas (0.98)
2-VIS+BLSTM: bananas (0.68)
BOW: bananas (0.14)

DAQUAR 585
What is the object on the chair?
Ground truth: pillow
IMG+BOW: clothes (0.37)
2-VIS+BLSTM: pillow (0.65)
LSTM: clothes (0.40)

DAQUAR 585a
Where is the pillow found?
Ground truth: chair
IMG+BOW: bed (0.13)
2-VIS+BLSTM: chair (0.17)
LSTM: cabinet (0.79)



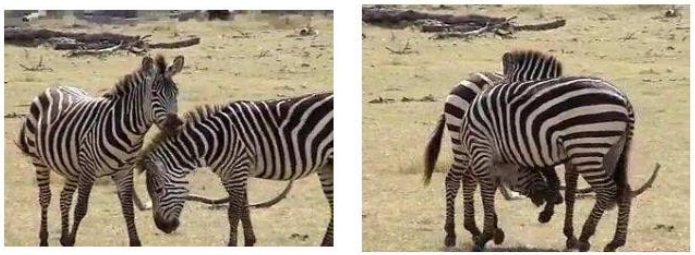
Ren et al. "Exploring models and data for image question answering." 2015.

Code: https://github.com/renmengye/imageqa-public

Applications: Video Description Generation

Correct descriptions.

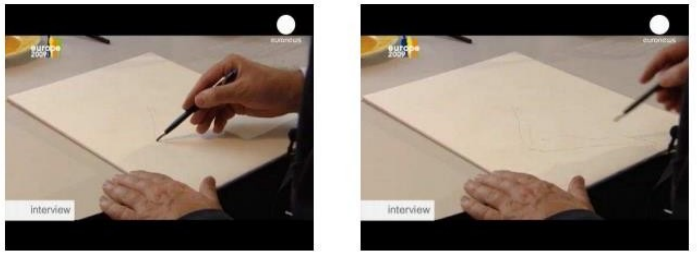
S2VT: A man is doing stunts on his bike.



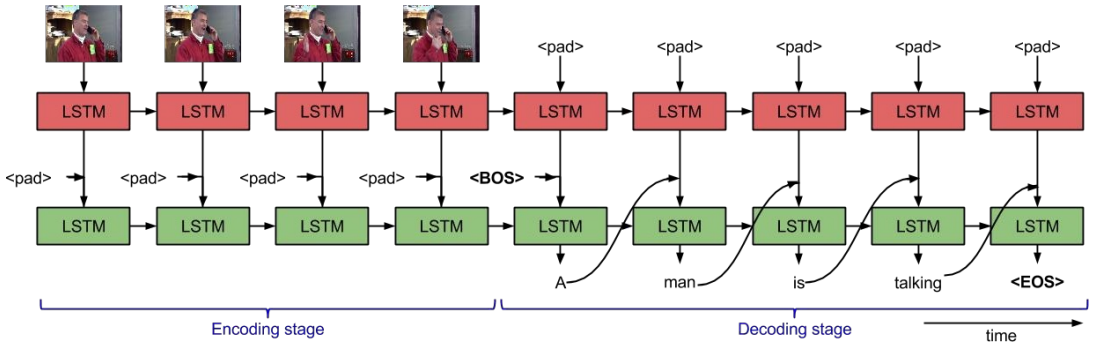
S2VT: A herd of zebras are walking in a field.

Relevant but incorrect descriptions.

S2VT: A small bus is running into a building.



S2VT: A man is cutting a piece of a pair of a paper.

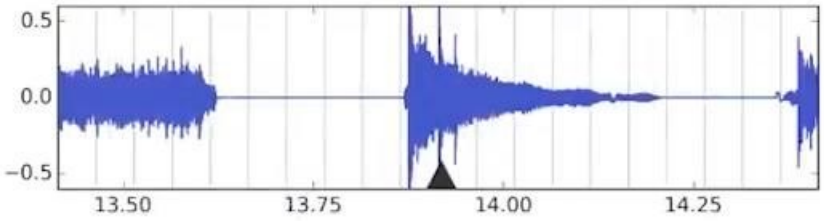


Venugopalan et al.
"Sequence to sequence-video to text." 2015.

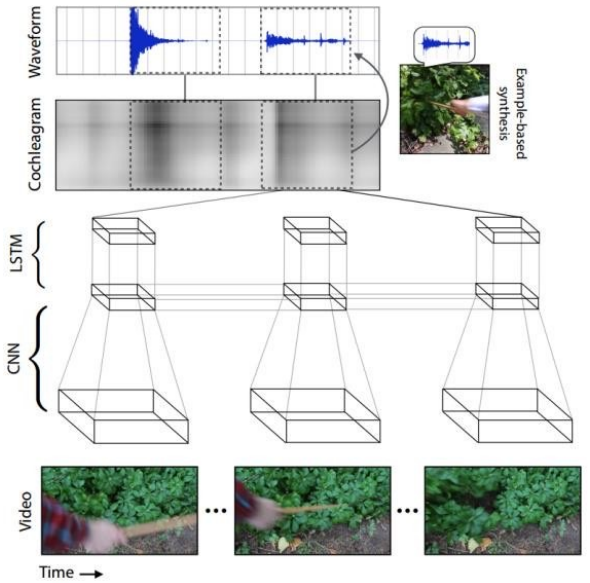
Code: https://vsubhashini.github.io/s2vt.html

Applications: Adding Audio to Silent Film

Silent video



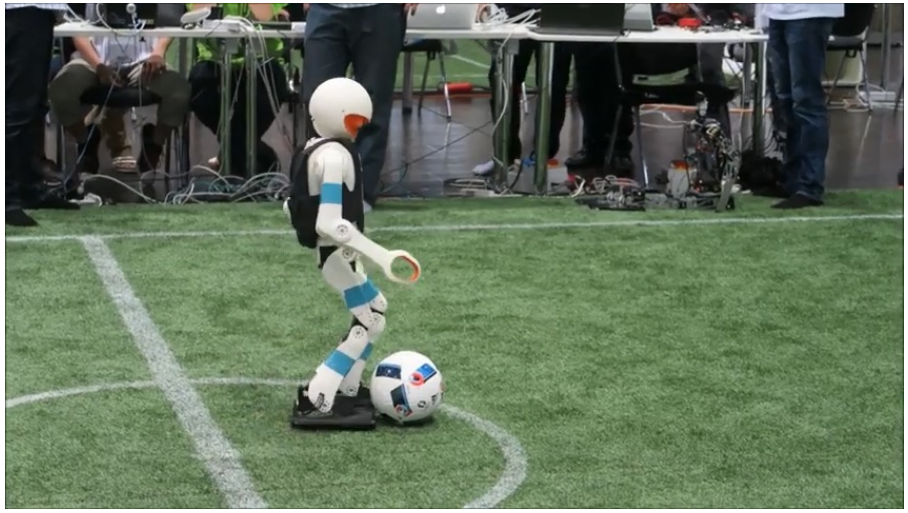
Predicted soundtrack



Owens, Andrew, Phillip Isola, Josh McDermott, Antonio Torralba, Edward H. Adelson, and William T. Freeman. **"Visually Indicated Sounds."** (2015).

References: [28, 29]

Moravec's Paradox: The "Easy" Problems are Hard



Soccer is harder than **Chess**

Moravec's Paradox: The "Easy" Problems are Hard

References: [8, 9]

Question: Why? **Answer:** Data

Visual perception: 540 millions years of data

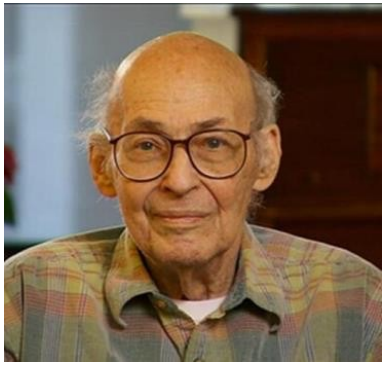
Bipedal movement: 230+ million years of data

Abstract thought: 100 thousand years of data

"Encoded in the large, highly evolved sensory and motor portions of the human brain is a **billion years of experience** about the nature of the world and how to survive in it.... Abstract thought, though, is a new trick, perhaps less than **100 thousand years** old. We have not yet mastered it. It is not all that intrinsically difficult; it just seems so when we do it."
- *Hans Moravec, Mind Children (1988)*

Hans Moravec (CMU)

Rodney Brooks (MIT)



Marvin Minsky (MIT)

Walking is Hard. How Hard is Driving?

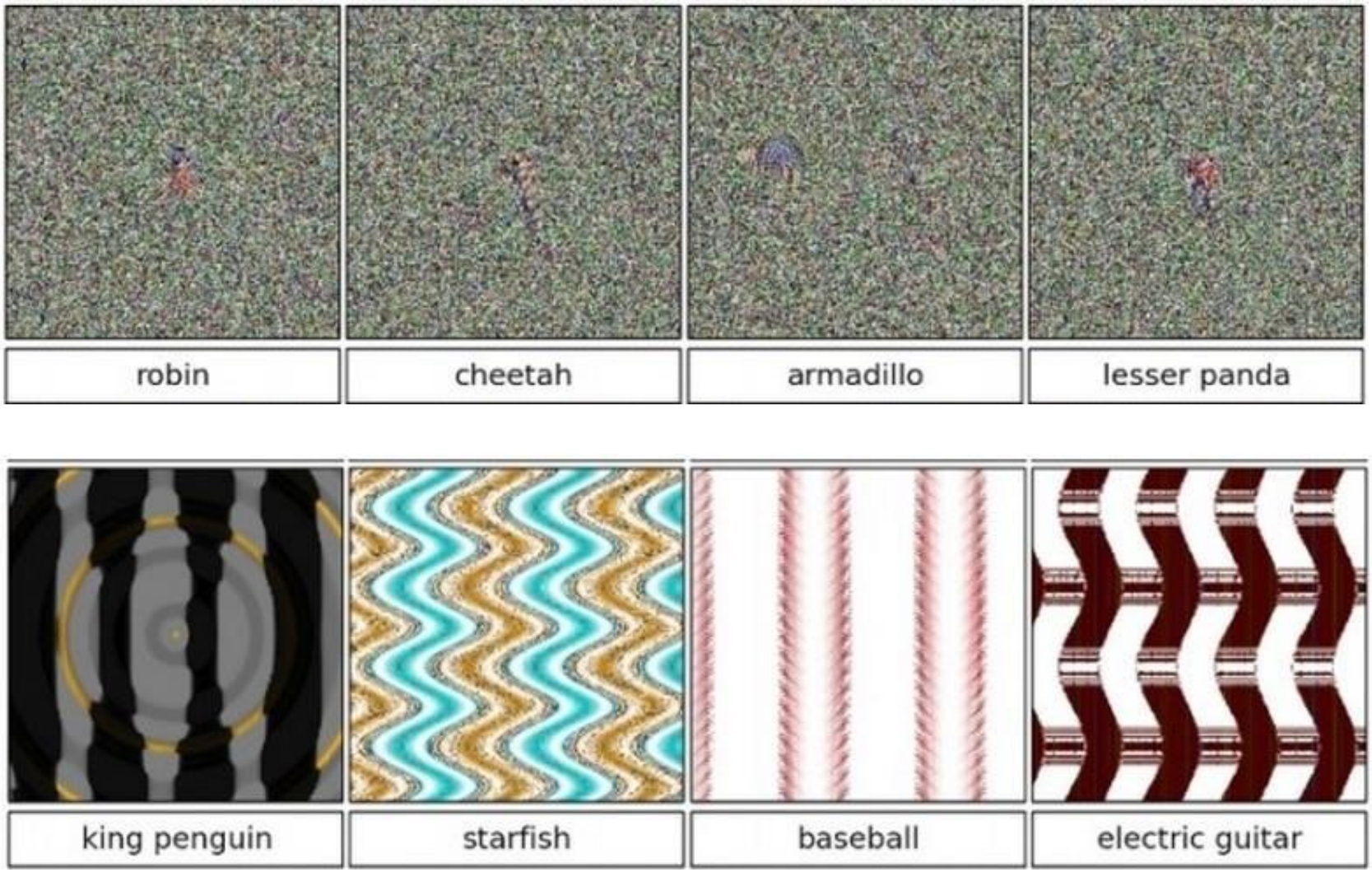
Human performance: 1 fatality per 100,000,000 miles

Error rate for AI to improve on: 0.000001%

Challenges:

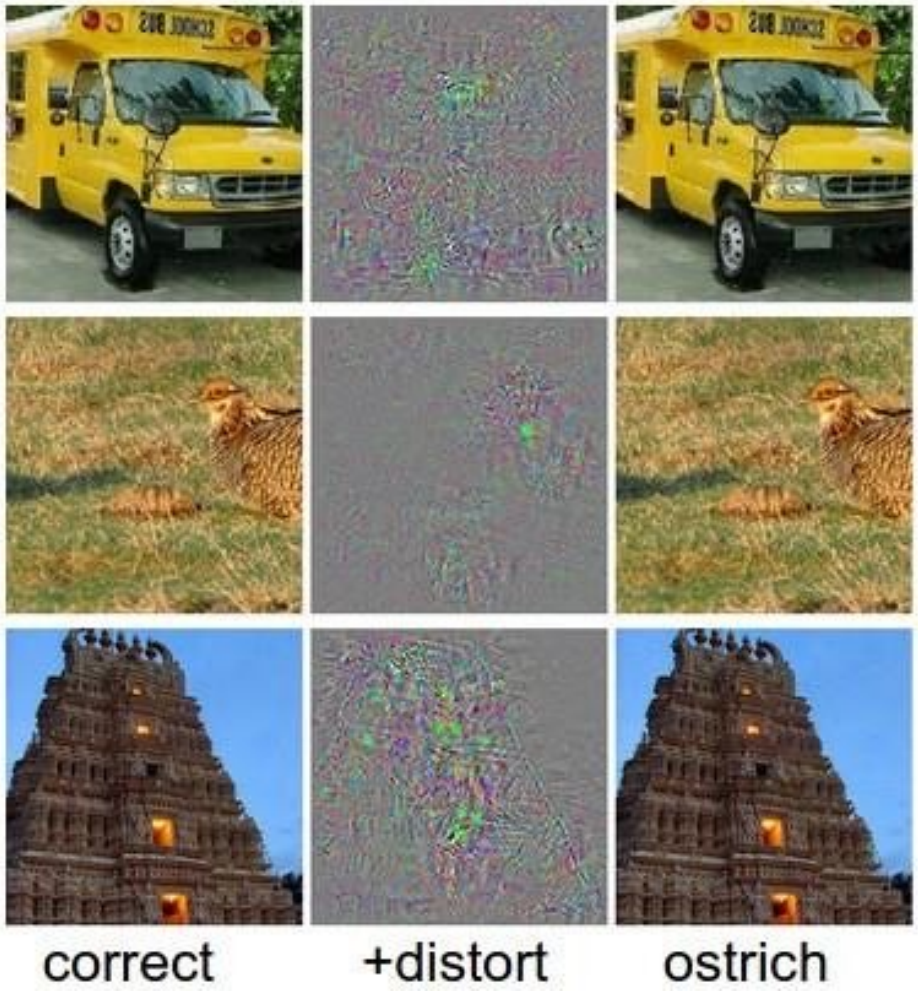
- Snow
- Heavy rain
- Big open parking lots
- Parking garages
- Any pedestrian behaving irresponsibly or just unpredictably
- Reflections, dynamics blinding ones
- Merging into a high-speed stream of oncoming traffic

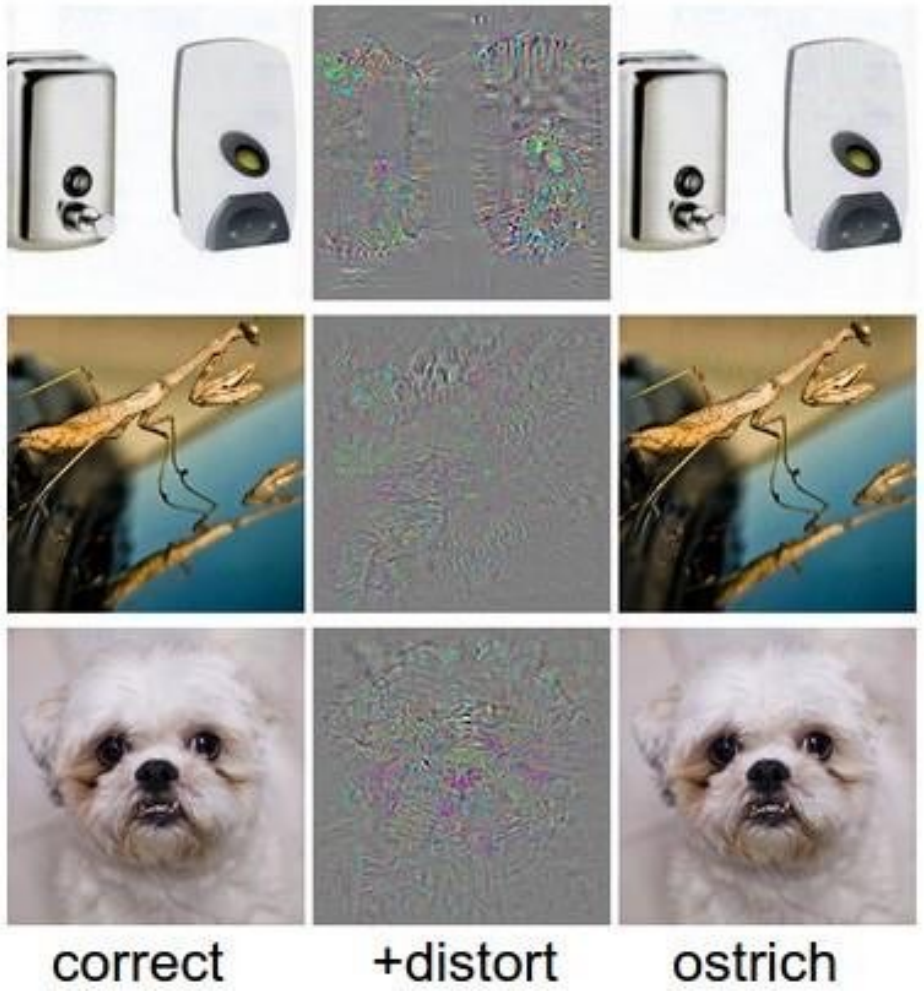
Robustness:
>99.6% Confidence in the Wrong Answer



Nguyen et al. "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images." 2015.

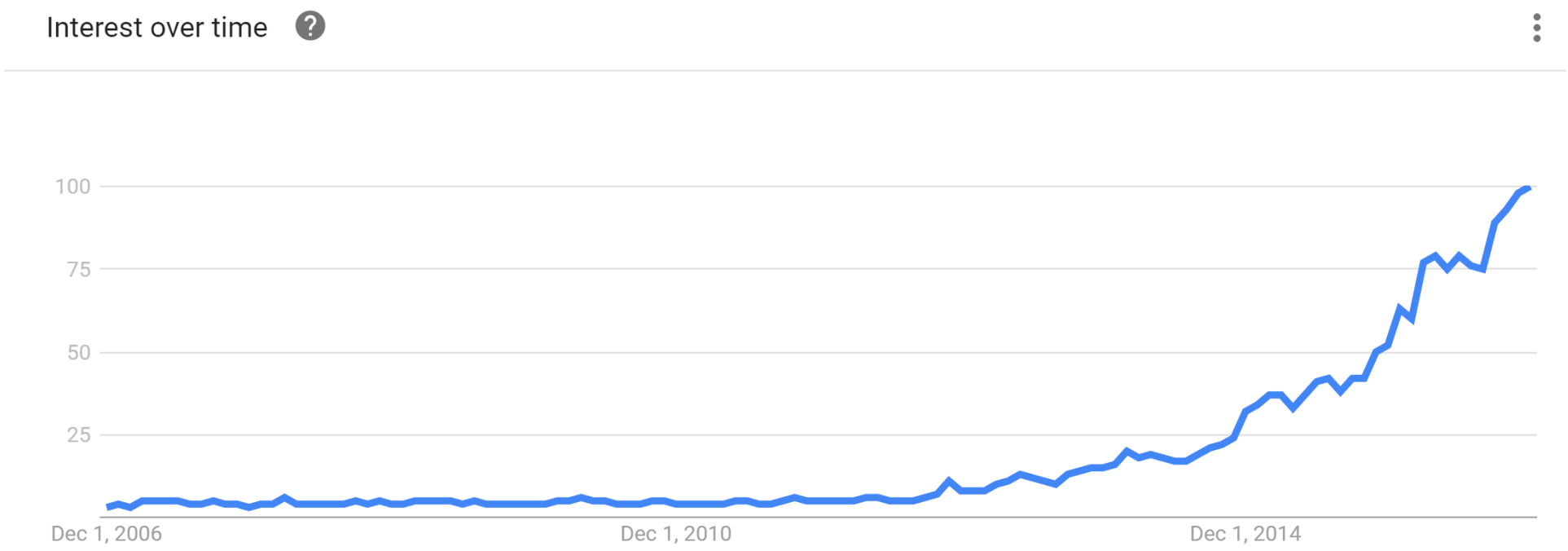
Robustness:
Fooled by a Little Distortion





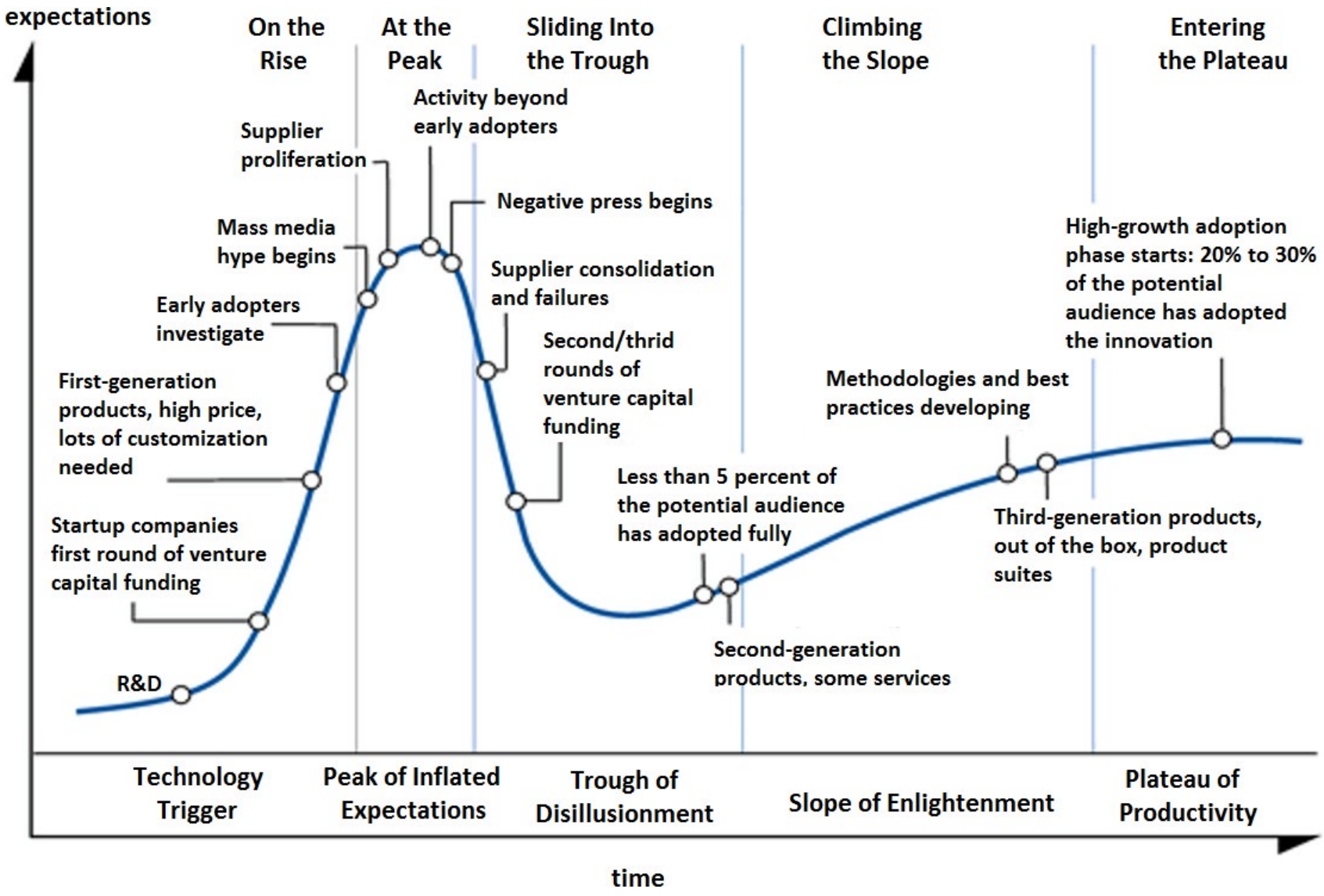
Szegedy et al. "Intriguing properties of neural networks." 2013.

3rd Summer of Deep Learning

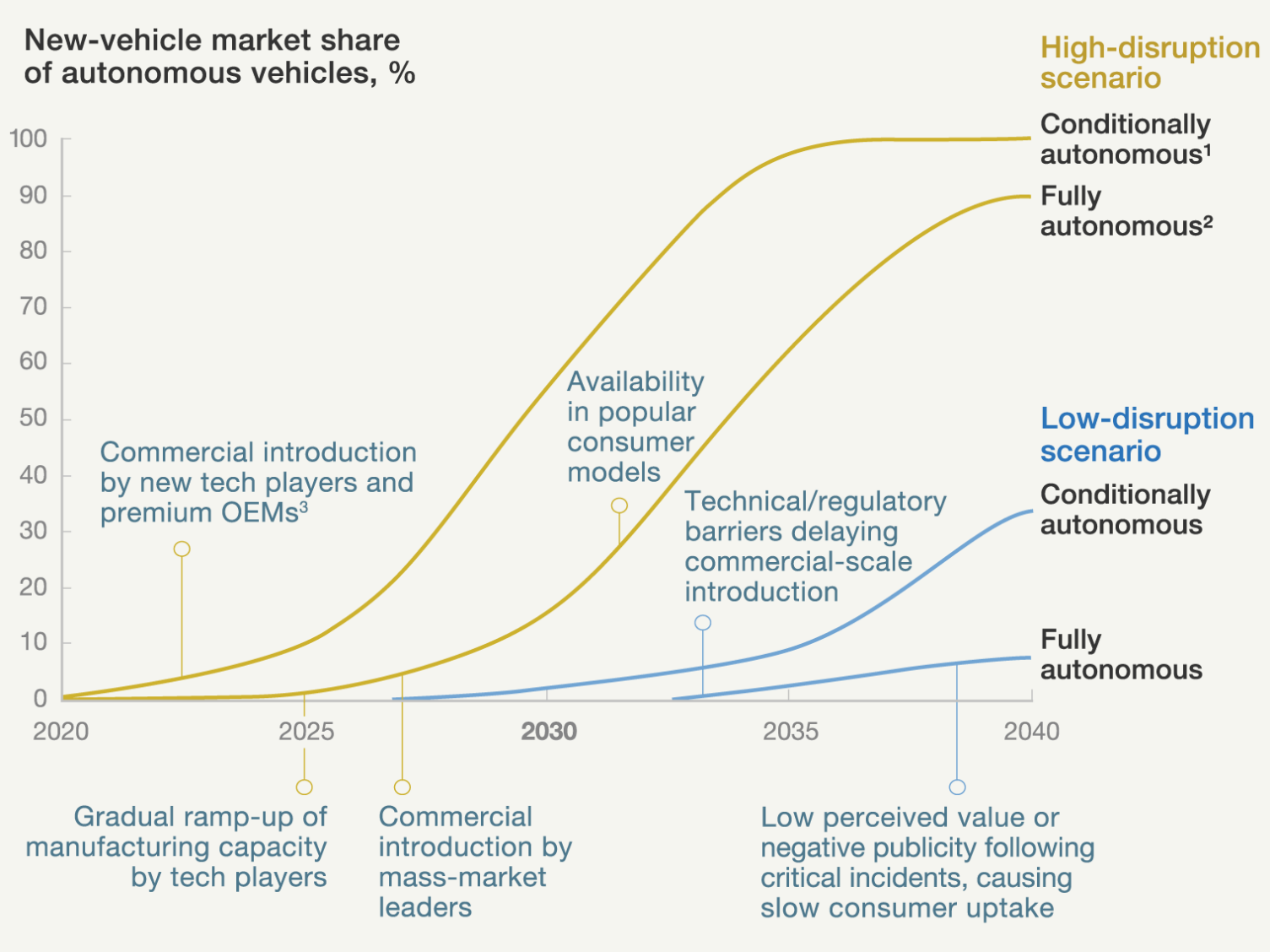


Google trends: “Deep Learning”

Gartner Hype Cycle

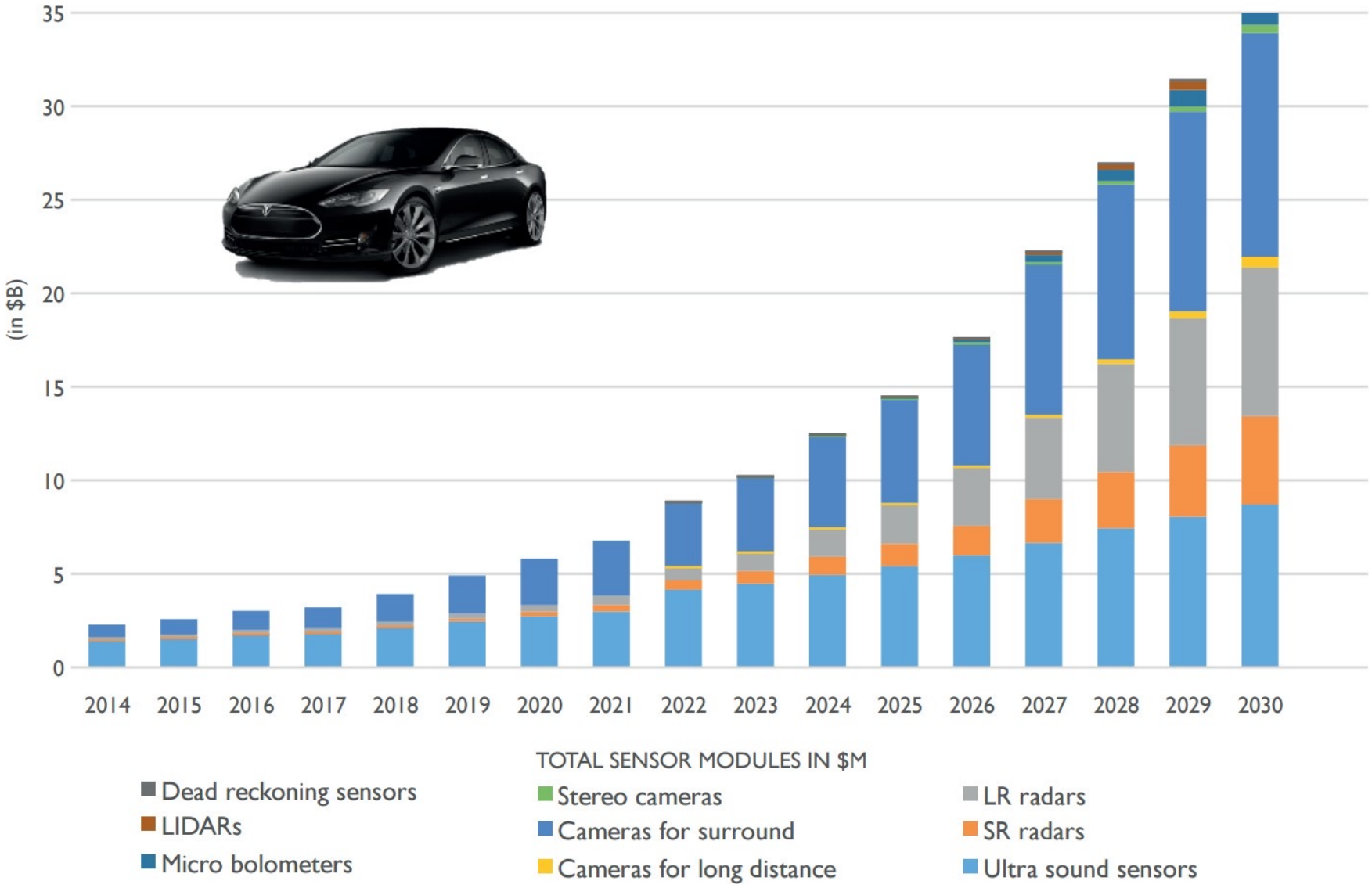


References: [21]



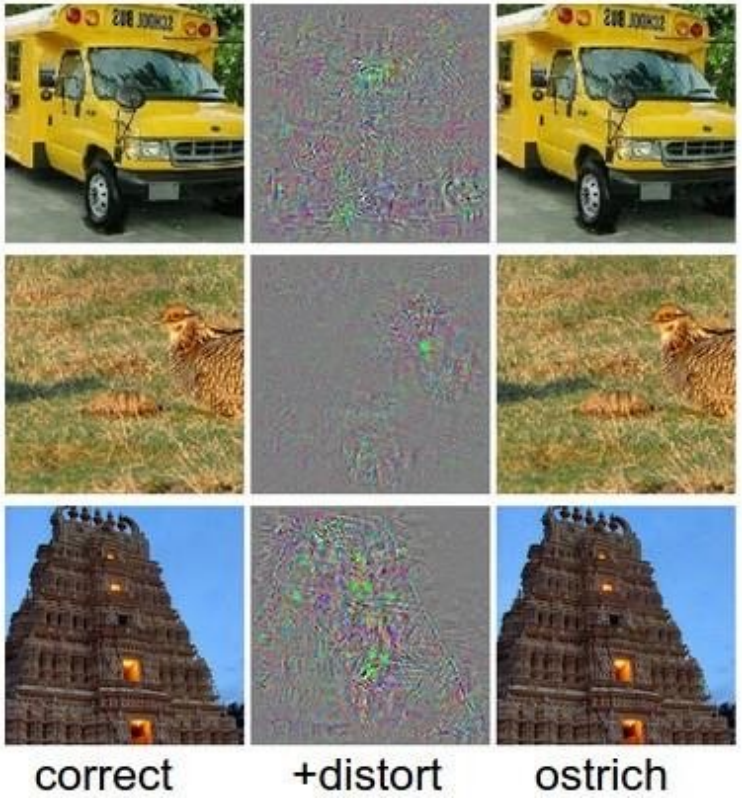
Factors in disruption scenarios	High disruption	Low disruption
Regulatory challenges	Fast	Gradual
Safe, reliable technical solutions	Comprehensive	Incomplete
Consumer acceptance, willingness to pay	Enthusiastic	Limited

Sensor modules market value for autonomous cars from 2015 to 2030 (in $B)



Attention to (AI) Drivers: Proceed with Caution

Camera Spoofing



LIDAR Spoofing

References: [68, 72]

- Interface: Python, (C++)
- Automatic Differentiation
- Multi GPU, Cluster Support
- Currently most popular

Keras

- On top of Tensorflow (and Theano)
- Interface: Python
- Goal: provide a simplified interface

- **Also:** TF Learn, TF Slim

Torch

- Used by researchers doing lower level (closer to the details) neural net work
- Interface: Lua
- Fragmented across different plugins

facebook

theano

- Interface: Python (tight NumPy integration)
- One of the earlier frameworks with GPU support
- Encourages low-level tinkering

Université de Montréal | Montreal Institute for Learning Algorithms

cuDNN

- The library that most frameworks use for doing the actual computation
- Implements primitive neural network functions in CUDA on the GPU

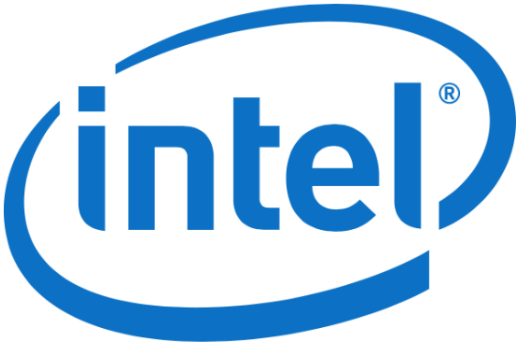
- Multi GPU Support (scales well)
- Interface: Python, R, Julia, Scala, Go, Javascript …

amazon

neon

framework by nervana

- Interface: Python
- Often best on benchmarks
- Nervana was working on a neural network chip
- Bought by Intel



Caffe

- Interface: C++, Python
- One of the earliest GPU supported
- Initial focus on computer vision (and CNNs)

Berkeley

Artificial Intelligence Research Laboratory

Microsoft Cognitive Toolkit (CNTK)

- Interface: Custom Language (BrainScript), Python, C++, C#
- Multi GPU Support (scales very well)
- Mostly used at MS Research

In the Browser

- Keras.js
  - GPU Support
  - Full sized networks
  - Can use trained Keras models

- ConvNetJS
  - Built by a Andrej Karpathy
  - Good for explaining neural network concepts
    - Fun to play around with
    - Very few requirements
  - Full CNN, RNN, Deep Q Learning