Lecture 5: Introduction to Deep Learning

Zerrin Yumak

Utrecht University

About your teacher

PhD in Computer Science, MIRALab, University of Geneva, 2006-2011

First post-doc, HCI Group, EPFL, Lausanne, 2012-2013

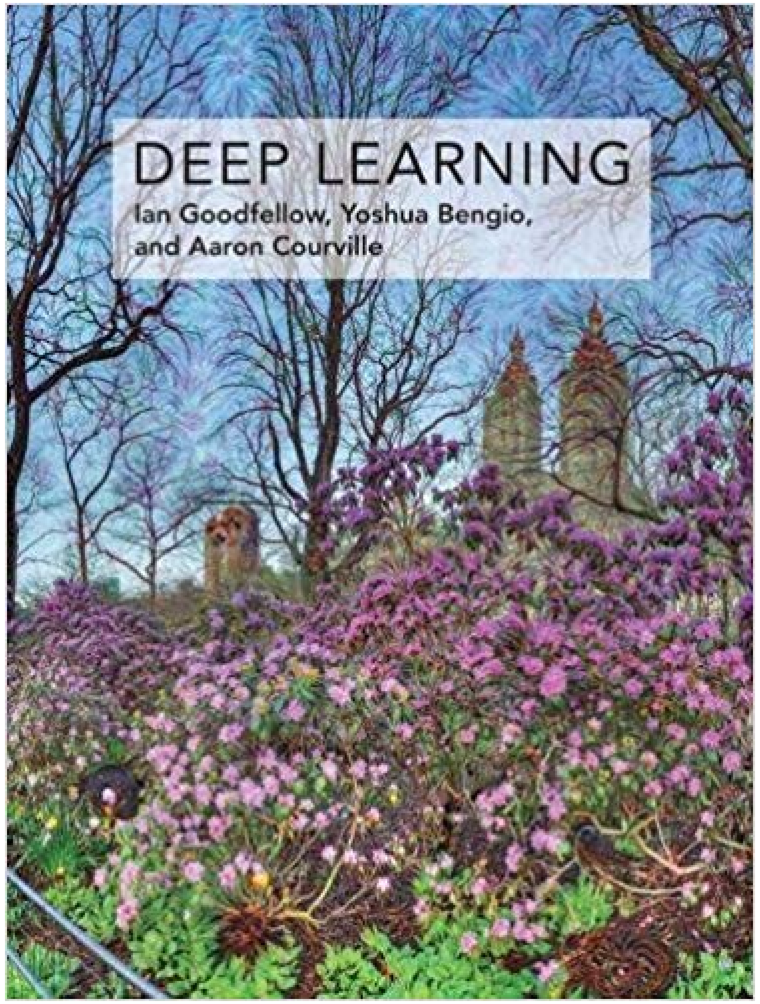
Second post-doc, Institute for Media Innovation, Nanyang Technological University, 2013-2015

Research Interests

- (Expressive) Character animation
- Facial animation
- Body gestures/emotions
- Gaze behavior
- Motion synthesis
- Multi-character interactions
- Virtual humans in VR and (Serious) Games
- Social robots and AI

Textbook

- Goodfellow, Bengio and Courville, Deep Learning, MIT Press, 2016
  - Available online
    - https://www.deeplearningbook.org/
- Supplementary materials (articles, video lectures, etc.)



Coarse Goals

- Understanding fundamental deep learning topics
  - Soft landing to the world of deep learning
- Developing practical skills for applying deep learning methods
  - Your first deep learning application (or second! ☺)
- Getting acquainted with research topics in deep learning
  - e.g. character animation with deep learning

Prerequisites

- Linear algebra (matrix operations)
- Calculus (derivatives)
- Probability and statistics
- Machine learning
- Programming

Contents

- Lecture 1: Intro to deep learning, applications and history
  - Introduction of project
- Lecture 2: NN Basics, training, regularization, optimization
  - Tensorflow example by Steven Langerwerf

- Lecture 3: Convolutional neural networks
- Lecture 4: Recurrent neural networks

- Lecture 5: Deep learning research papers on animation + Guest lecture (TBD)

Assignment 2

- Project in teams of 4 – 5
  - Link on website
- Goal: Apply deep learning on a research oriented problem
  - Image processing, text analysis, speech recognition, character animation, games, music, recommender systems
- Step 1: Submit a 1-2 page project proposal and make a presentation of the proposal
- Step 2: Submit a 4-6 pages final report, code and present the results

Project proposal

- Which research topic will you investigate?
- What data will you use?
- What will be the neural network architecture you will use?
- What are the related work?
- How will you evaluate the results?
- Who will do what? When?

- If you have questions, find me after the lecture, send me an email or ask for an appointment!

Final deliverable

- Submit a 4-6 page short paper (Latex or Word template)

- Title, Authors, Abstract
- Introduction: Introduce the problem and motivation
- Related work: Relevant literature
- Approach: Methodology and technical details
- Results: Provide qualitative and quantitative results
- Conclusion: Summarize your findings
- References

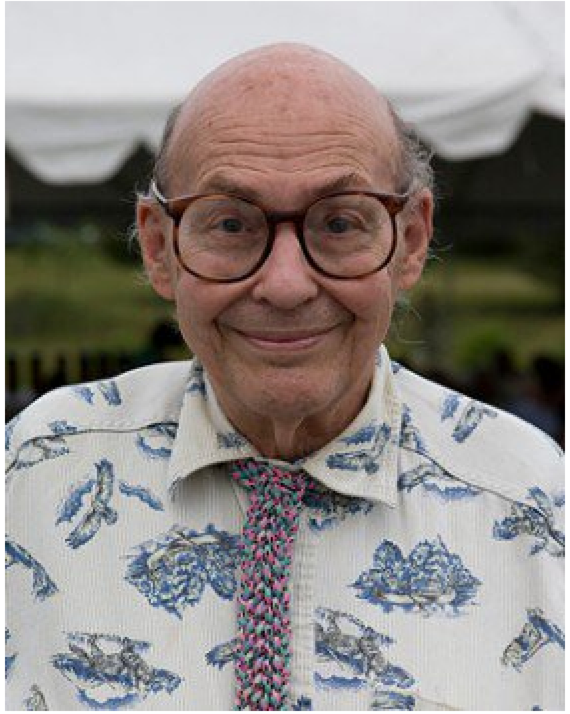
Grading criteria

- Clarity of the document
- Clarity of the presentation
- Technical content
- Critical analysis

(Code is supplementary material, will not be graded)

Abstract level of thinking in research 1

- I want to **solve vision** / language/ etc.



Marvin Minsky

Abstract level of thinking in research 2

- I want to **solve vision** / language/ etc.

- I want to do X (e.g. image captioning)
  - This is excellent if X is something no one has done or thought about and is important (guaranteed success)
  - Requires forward thinking, knowledge of the field
  - Difficult to do as the field matures

Rosalind Picard

Abstract level of thinking in research 3


Geoffrey Hinton

- I want to **solve vision** / language/ etc.

- I want to do X (e.g., **image captioning**)

- I think the **right way to solve** (or improve) X is Y
  - More incremental; a lot of science is incremental (''standing on the shoulders of giants")
  - **Retrospective**: compare existing approaches, see why they work and what is missing (guaranteed success)
  - **Perspective:** come up with an idea or the insight that you truly believe and test it
  - Requires through knowledge of the sub-field (lots of reading)
  - Requires strong intuition and high level (intuitive) thinking
  - Requires understanding of the mathematical tools and formulations to know what maybe possible
  - Helps to bringing knowledge from other fields (field cross pollination)

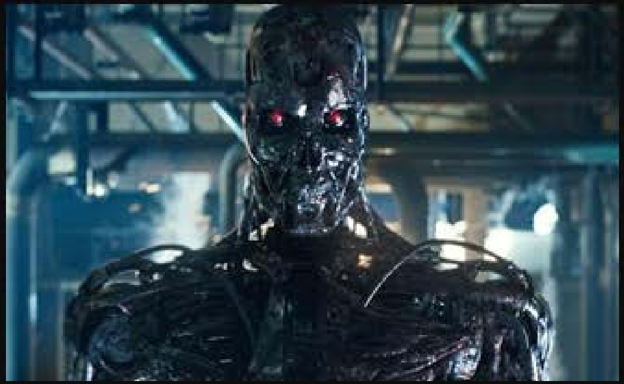
Abstract level of thinking in research 4

- I want to **solve vision** / language/ etc.
- I want to do X (e.g., **image captioning**)
- I think the **right way to solve** (or improve) X is Y
- Mathematical **formulation**
- Implementation / engineering
- **Experimental** testing

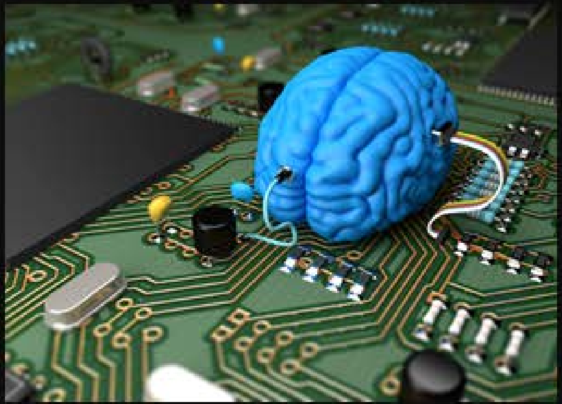
In this lecture

- What is deep learning?
- History of deep learning
- Perceptron
- Advantages and disadvantages of deep learning
- Software
- Applications

Deep Learning



What society thinks I do



What my friends think I do

What other computer scientists think I do

What mathematicians think I do

What I think I do

```
from theano import *
```

What I actually do

What is deep learning?

- Deep learning allows **computational models** that are composed of **multiple processing layers** to **learn representations of data with multiple levels of abstraction**.

Famous deep learning Nature paper from 2015

REVIEW

doi:10.1038/nature14539

Deep learning

Yann LeCun[1,2], Yoshua Bengio[3] & Geoffrey Hinton[4,5]

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

Machine-learning technology powers many aspects of modern society: from web searches to content filtering on social networks to recommendations on e-commerce websites, and it is increasingly present in consumer products such as cameras and smartphones. Machine-learning systems are used to identify objects in images, transcribe speech into text, match news items, posts or products with users' interests, and select relevant results of search. Increasingly, these applications make use of a class of techniques called deep learning.

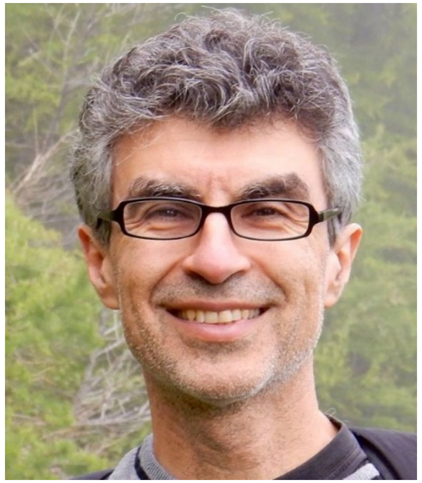
Conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, constructing a pattern-recognition or machine-learning system required careful engineering and considerable domain expertise to design a feature extractor that transformed the raw data (such as the pixel values of an image) into a suitable internal representation or feature vector

intricate structures in high-dimensional data and is therefore applicable to many domains of science, business and government. In addition to beating records in image recognition[1–4] and speech recognition[5–7], it has beaten other machine-learning techniques at predicting the activity of potential drug molecules[8], analysing particle accelerator data[9,10], reconstructing brain circuits[11], and predicting the effects of mutations in non-coding DNA on gene expression and disease[12,13]. Perhaps more surprisingly, deep learning has produced extremely promising results for various tasks in natural language understanding[14], particularly topic classification, sentiment analysis, question answering[15] and language translation[16,17].

We think that deep learning will have many more successes in the near future because it requires very little engineering by hand, so it can easily take advantage of increases in the amount of available computation and data. New learning algorithms and architectures that are

Important people in the field

Jeffrey Hinton
University of Toronto and Google Brain



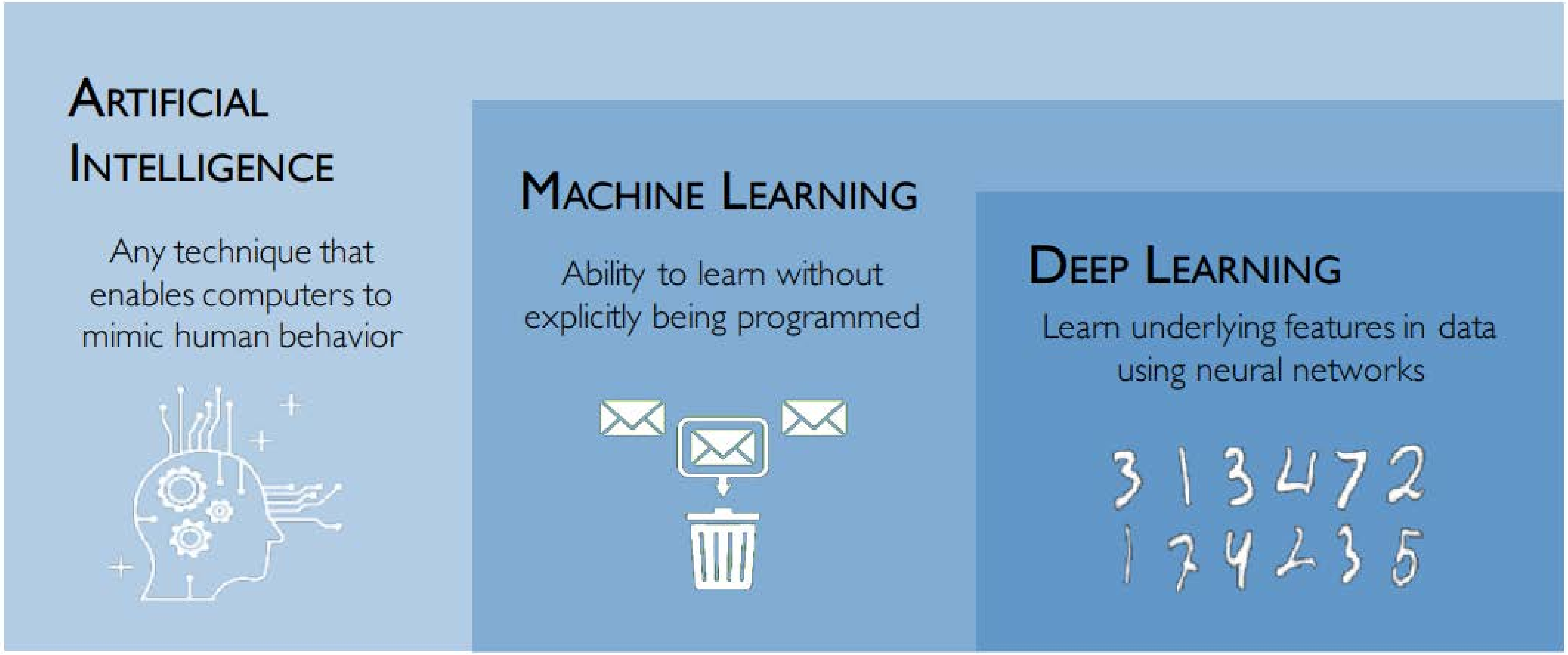
Yoshua Bengio
McGill University, Université de Montréal

Yann LeCun
Director of Facebook AI research, post-doc at Hinton’s lab

Andrew Ng
Founder of Google Brain, Coursera, Chief Scientist at Baidu, Stanford University

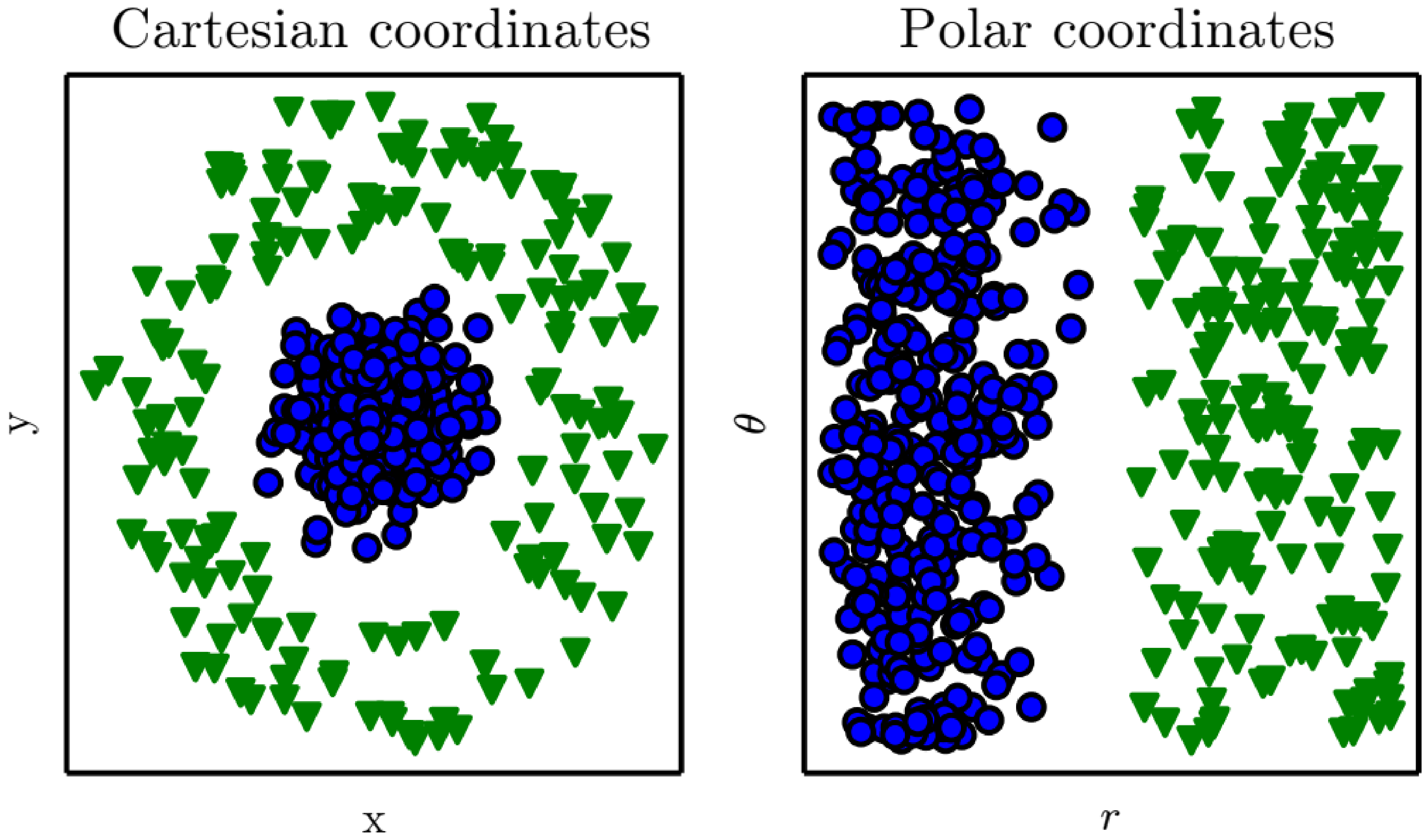
[Heroes of Deep Learning Andrew Ng Interviews on YouTube]

Deep Learning, Machine Learning and AI

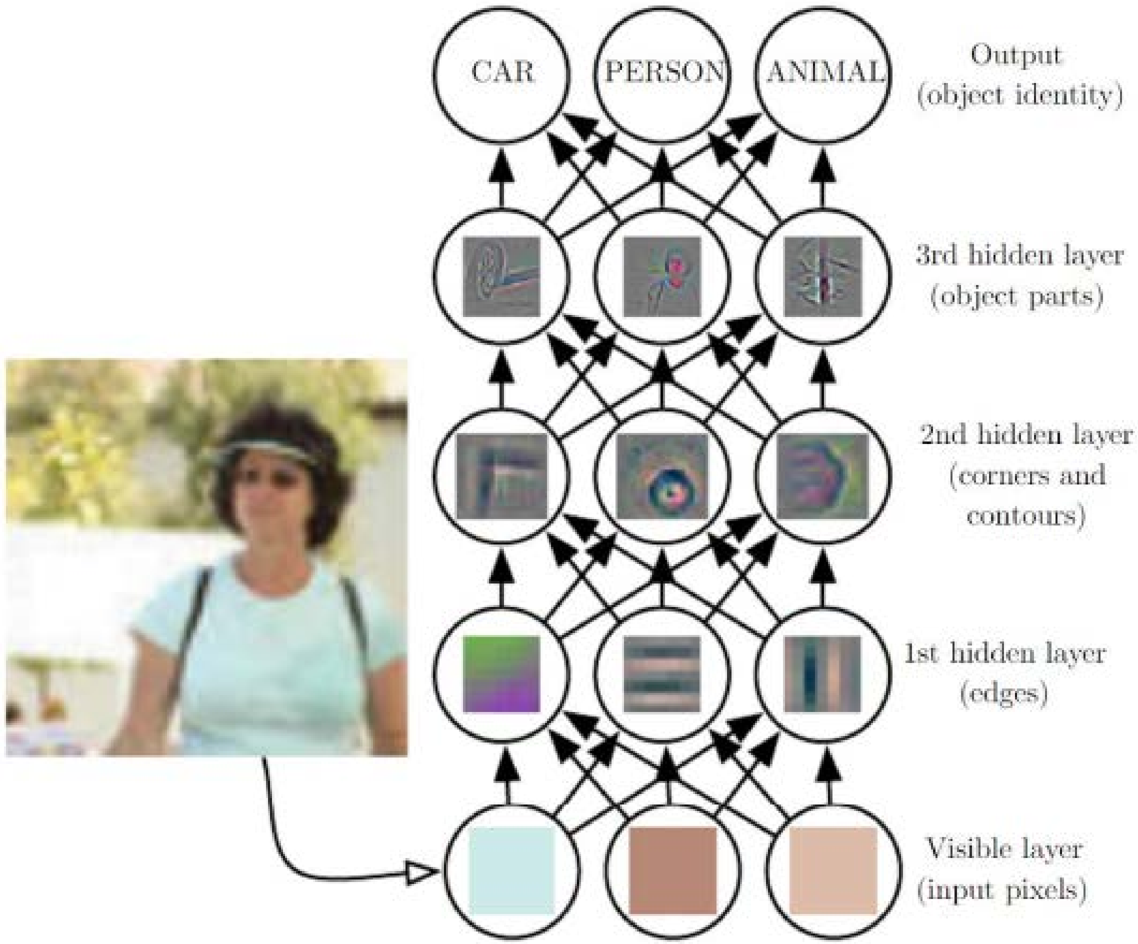


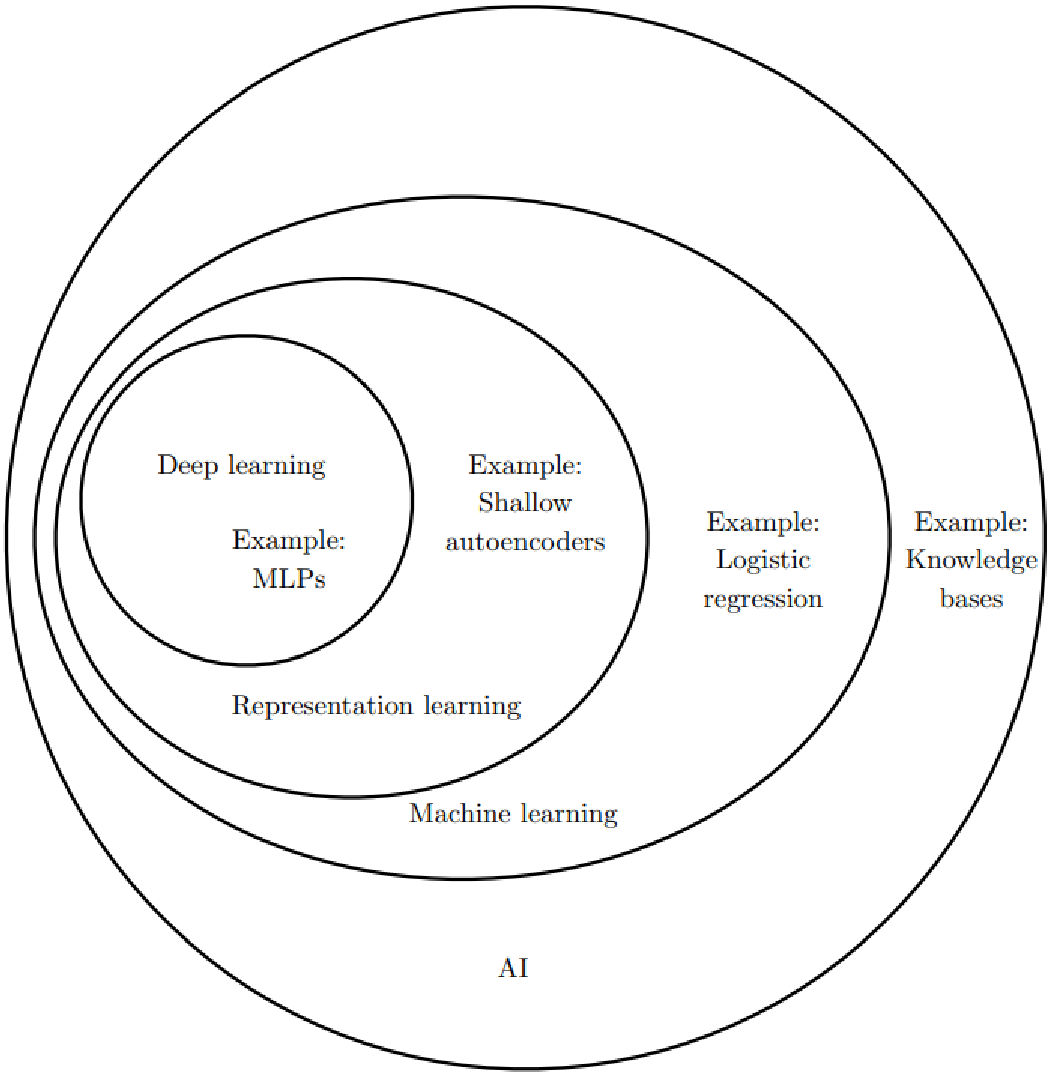
MIT Intro to deep learning

How to represent data?

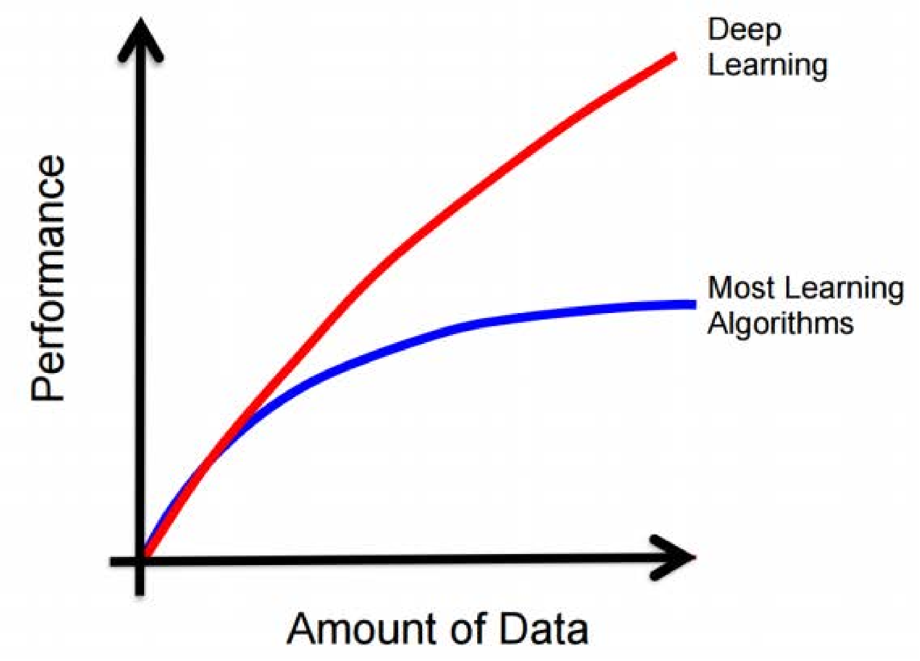


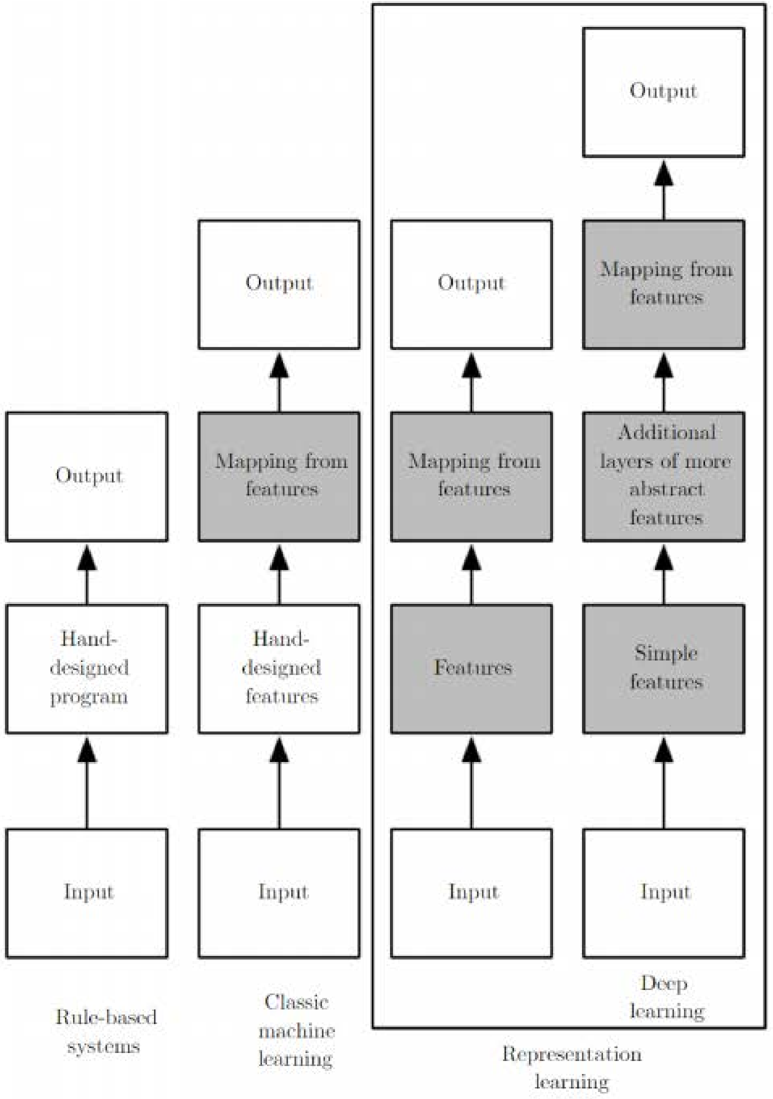
Deep Learning, Machine Learning and AI



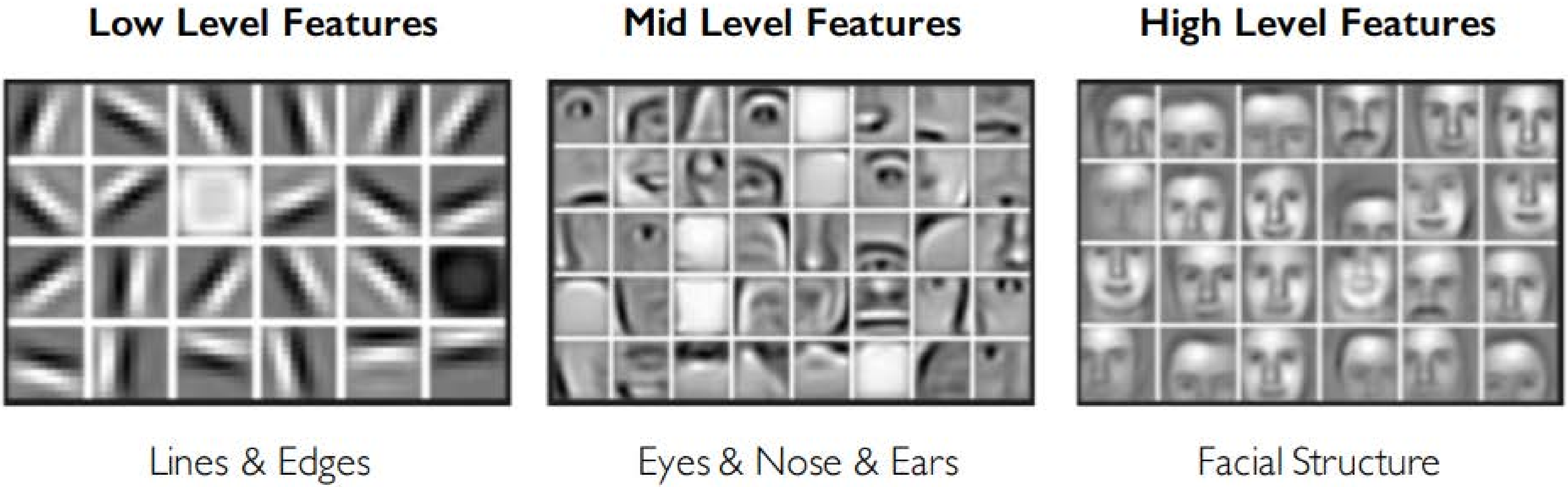


Deep Learning vs Other Learning algorithms

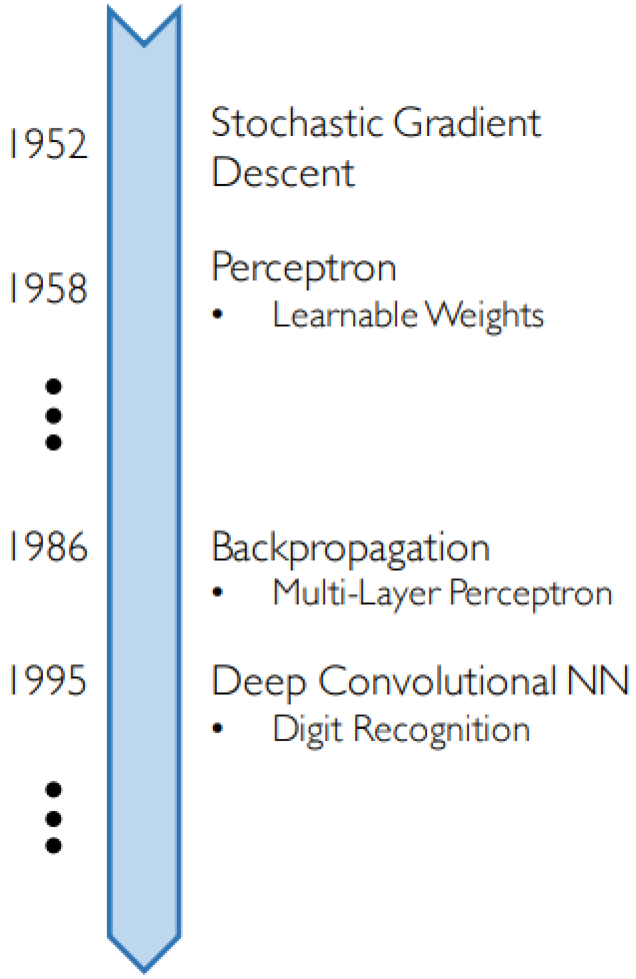




Can we learn underlying features directly from data?



Neural networks date back decades, Why now?



- Big Data: Larger Datasets, Easier Collection and Storage
- Hardware: Graphics Processing Units (GPUs), Massively Parallellizable
- Software: Improved techniques, new models, toolboxes

Software

TensorFlow

torch

mxnet

theano

Caffe

Caffe2

MatConvNet

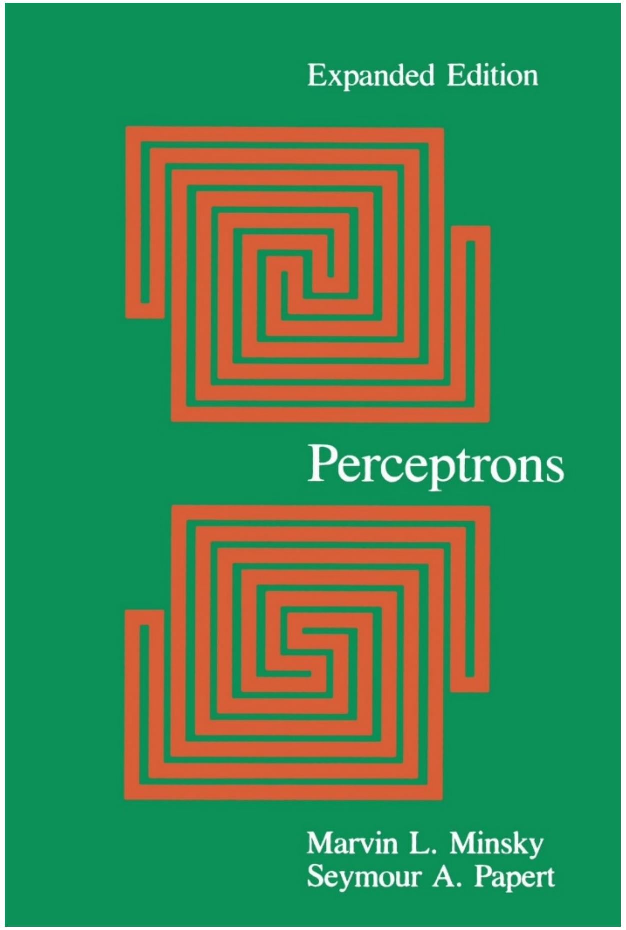
The Microsoft Cognitive Toolkit

A free, easy-to-use, open-source, commercial-grade toolkit that trains deep learning algorithms to learn like the human brain.

PYTORCH

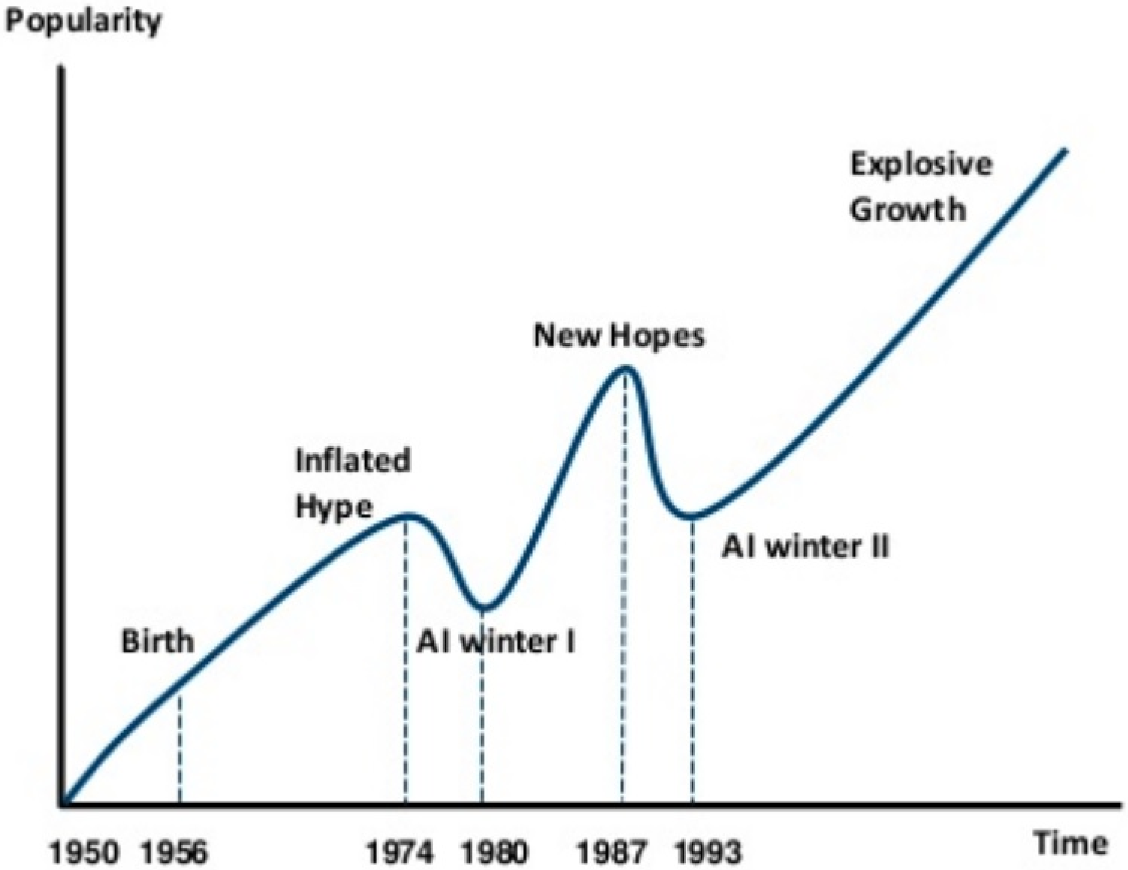
History of deep learning (Early days)

- 1943: First computational model by McCulloch and Pitts
  - A neuron model that sums binary inputs and outputs a 1 if the sum exceeds a threshold value
- 1958: A computational model of a single neuron, Rosenblatt's Perceptron
  - Solves a binary classification problem
- 1969: Perceptron book by Marvin Minsky and Seymour Papert
  - Claims perceptrons can only represent linearly separable functions
  - Attributed as the reason for the AI Winter, a period of reduced funding and interest in AI research



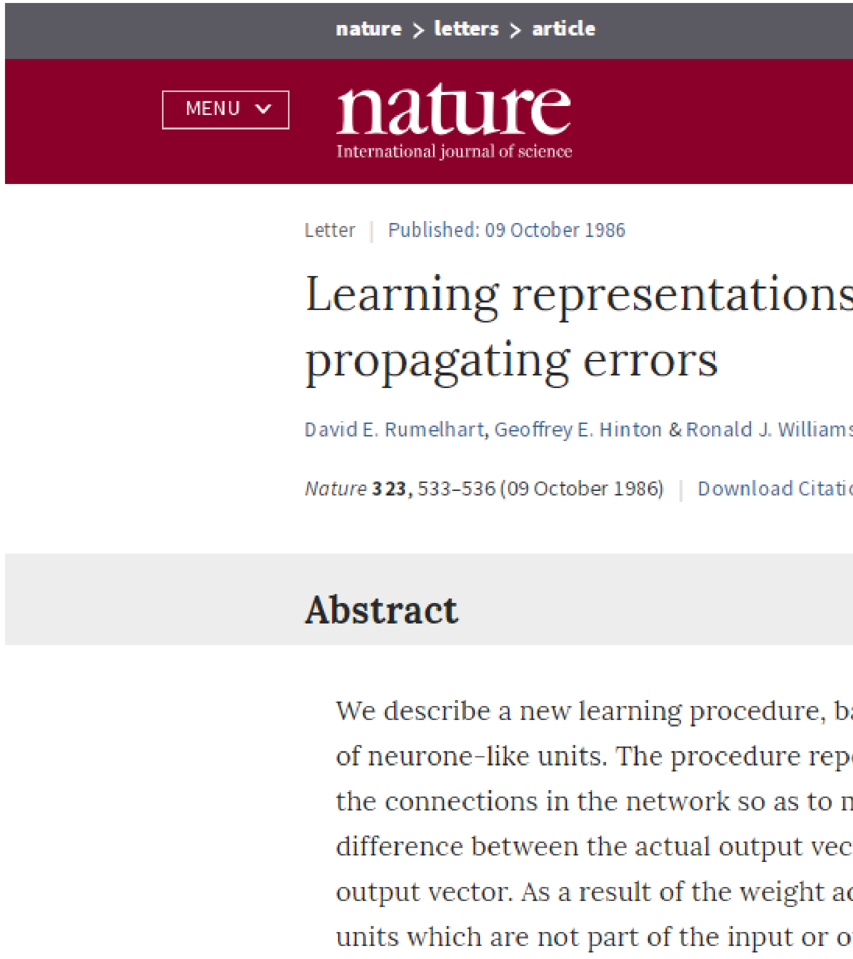
F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain", Psych. Review, Vol. 65, 1958

AI Winter (1969 – 1990)



History of deep learning (1990s)

- Multi-layer perceptrons (Cybenko, 1989 and Hornik, 1991)
  - **Multilayer feedforward networks** are universal approximators
- Training multi-layer perceptrons
  - **Backpropagation:** Learning representations by backpropagating errors (Rumelhart, Hinton, Williams 1986, Nature paper)
  - **Backpropagation through time (BPTT):** Generalization of Backpropagation with Application to a Recurrent Gas Market Model (Werbos, 1988)
- New neural architectures
  - **Convolutional neural nets:** Backpropagation Applied to Handwritten Zip Code Recognition (Lecun et al. 1989)
  - **Long short-term memory nets (LSTM):** Hochreiter and Schmidhuber, 1997


nature > letters > article

MENU

nature
International journal of science

Letter | Published: 09 October 1986

Learning representations
propagating errors

David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams

Nature **323**, 533–536 (09 October 1986) | Download Citatio

Abstract

We describe a new learning procedure, b
of neurone-like units. The procedure rep
the connections in the network so as to m
difference between the actual output vec
output vector. As a result of the weight ad
units which are not part of the input or o

History of Deep Learning (2000s)

- Despite progress, it wasn’t very popular back then
  - Not enough training data
  - Relying on feature engineering
  - Black-box model, not interpretable
  - Non-convex optimization
  - Computational complexity
  - Success and efficiency of Support Vector Machines

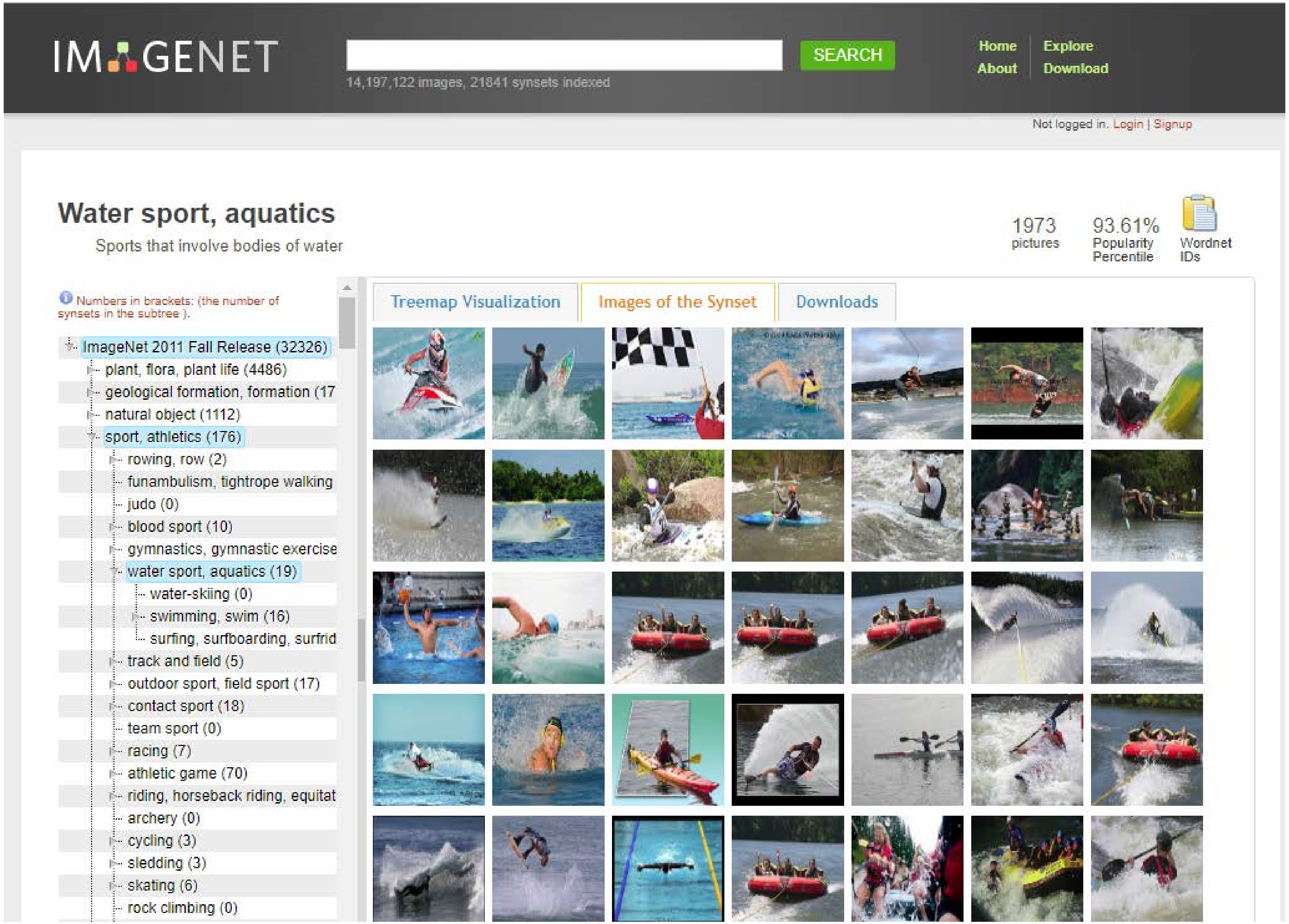
- 2006: Reducing the dimensionality of data with neural networks (Hinton and Salakhutdinov)
  - High-dimensional data can be converted to low-dimensional codes by training a multilayer neural network
  - Works much better than principal components analysis as a tool to reduce the dimensionality of data.

ImageNet Challenge

- Beginning of deep learning revolution, attention from both AI community and industry
- **What is ImageNet?**
  - A large dataset (14+ million images, 20K+ categories)
  - Was presented at CVPR (Conference on Computer Vision and Pattern Recognition), from Princeton University

http://www.image-net.org

Deng et al, ImageNet: A large-scale hierarchical image database, CVPR 2009


IMAGENET
SEARCH
14,197,122 images, 21841 synsets indexed
Home
About
Explore
Download
Not logged in. Login | Signup
Water sport, aquatics
Sports that involve bodies of water
1973
pictures
93.61%
Popularity
Percentile
Wordnet
IDs
Numbers in brackets: (the number of synsets in the subtree).
ImageNet 2011 Fall Release (32326)
plant, flora, plant life (4486)
geological formation, formation (17
natural object (1112)
sport, athletics (176)
rowing, row (2)
funambulism, tightrope walking
judo (0)
blood sport (10)
gymnastics, gymnastic exercise
water sport, aquatics (19)
water-skiing (0)
swimming, swim (16)
surfing, surfboarding, surfrid
track and field (5)
outdoor sport, field sport (17)
contact sport (18)
team sport (0)
racing (7)
athletic game (70)
riding, horseback riding, equitat
archery (0)
cycling (3)
sledding (3)
skating (6)
rock climbing (0)
Treemap Visualization
Images of the Synset
Downloads

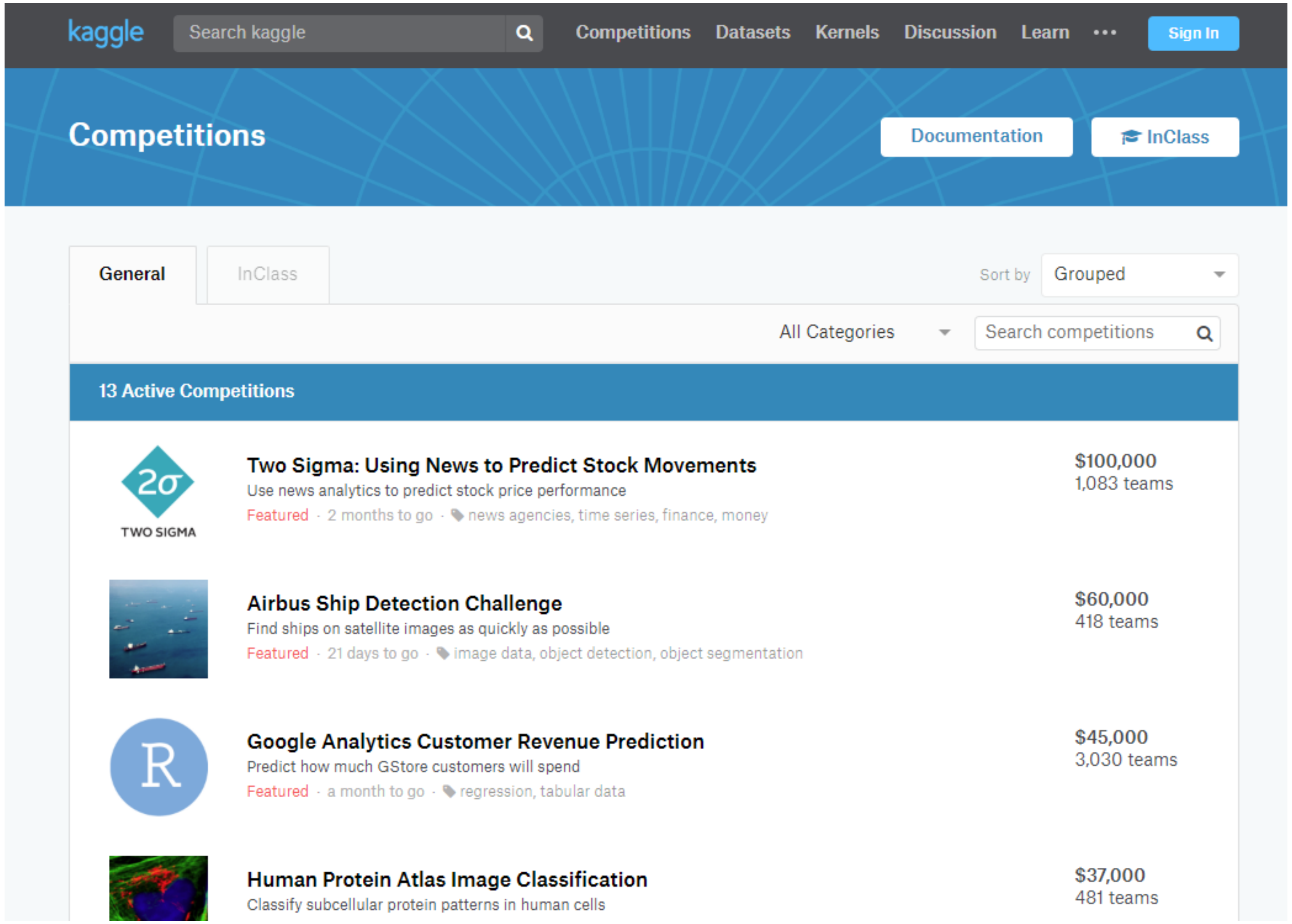
ImageNet Challenge

- **What is the challenge?**
  - ImageNet Large Scale Visual Recognition Challenge (ILSVRC)
  - Research teams compete for higher accuracy
    - Error rate in 2011 was 25%, in 2012 a deep CNN achieved 16%
    - In 2015, surpassed human error (5.1%)

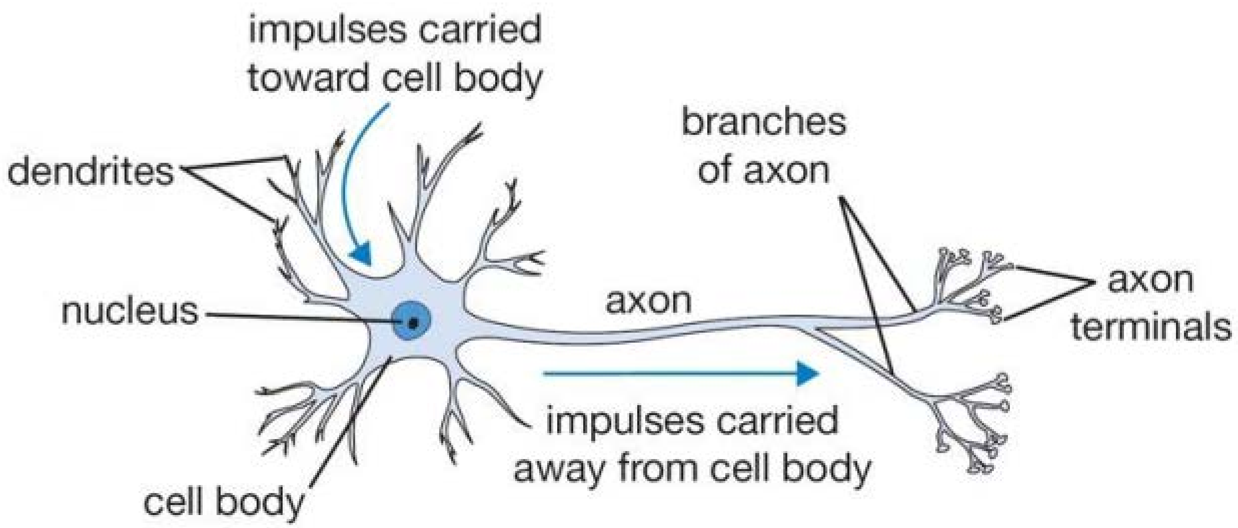
MIT Self-driving cars course

2018: ImageNet Challenge moves to Kaggle

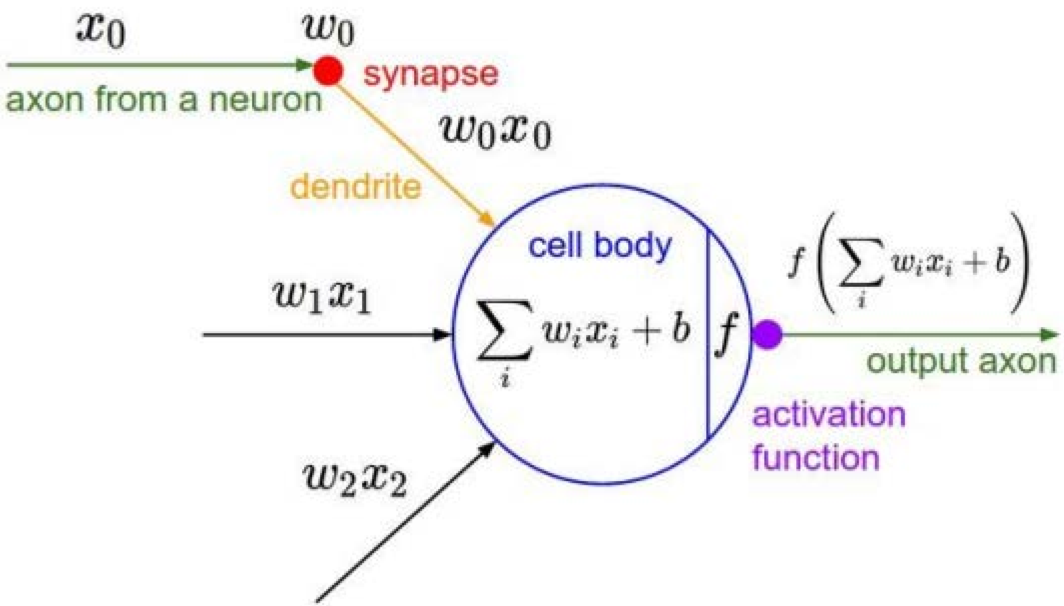
Kaggle is an online community of data scientists and machine learners, owned by Google.



Biological Neuron vs Artificial Neuron



Computational unit of the brain

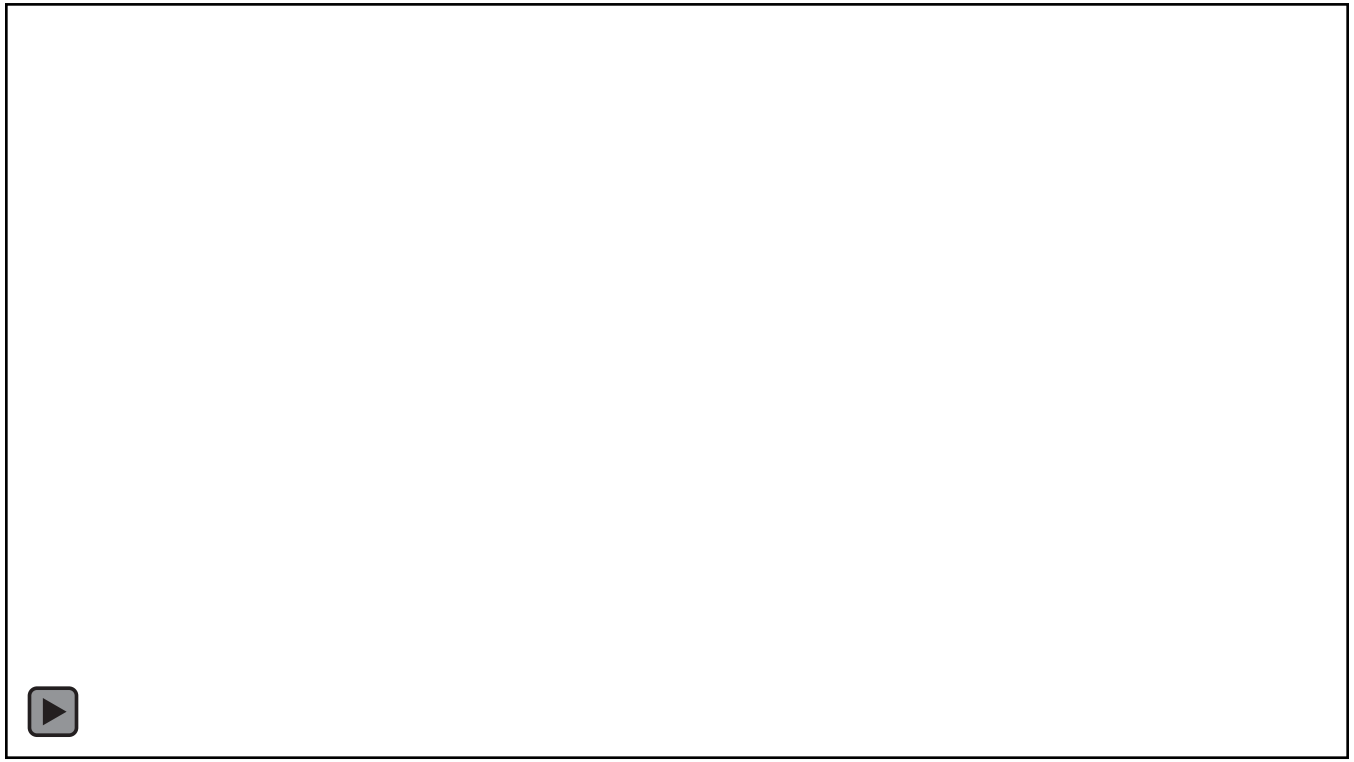


Computational unit of artificial neural networks

[Hugo Larocehlle Neural networks [1.6] : Feedforward neural network - biological inspiration]

Human Brain vs ANNs

- Human Brain
  - 100 billion neurons, 1000 trillion synapses



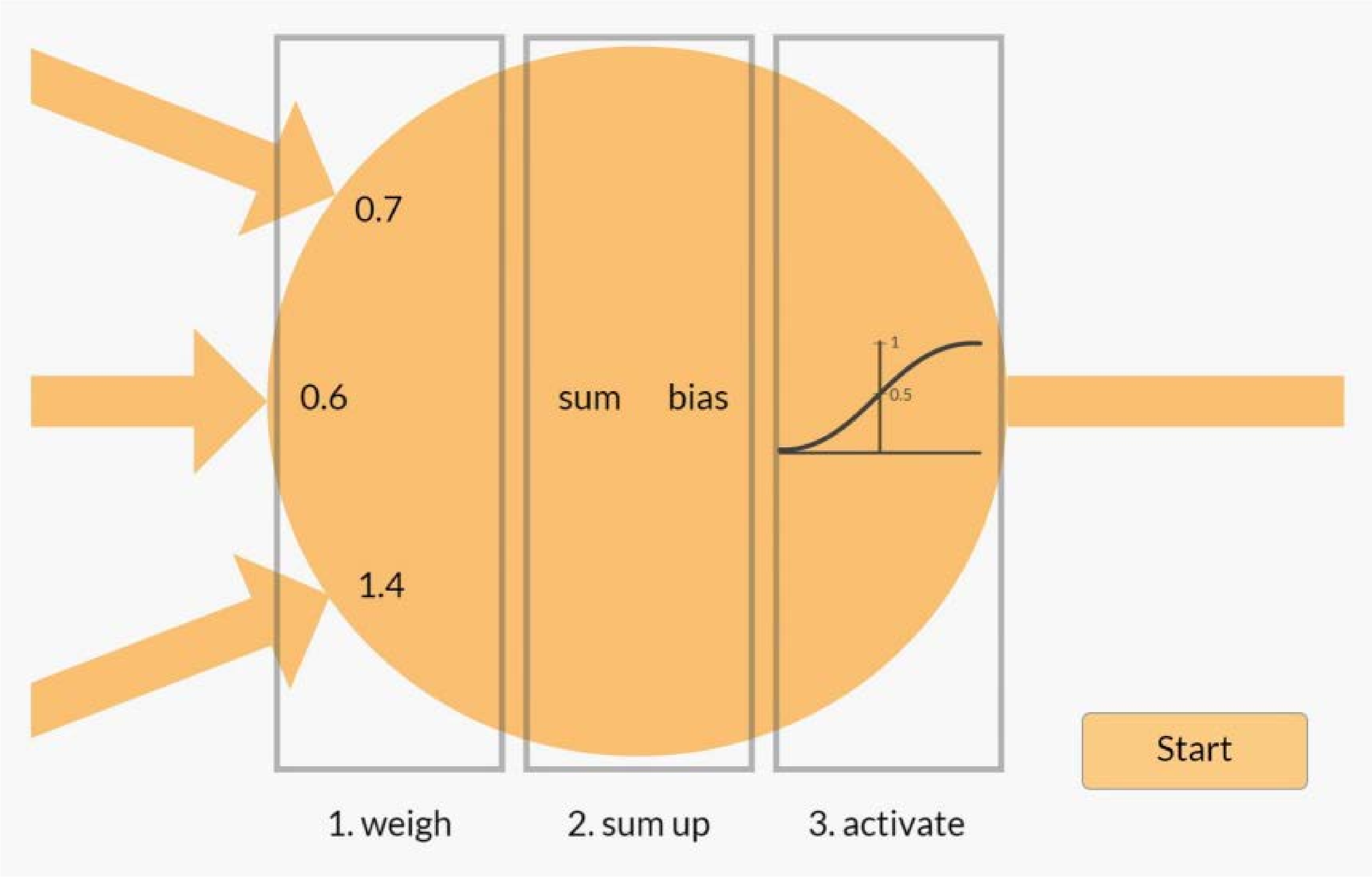
Human Brain vs ANNs

- ResNet-152 neural network (Winner of ImageNet Challenge in 2015)
  - 60 million synapses
  - Deep Residual Learning for Image Recognition, He et al., CVPR 2015

Human Brain vs ANNs

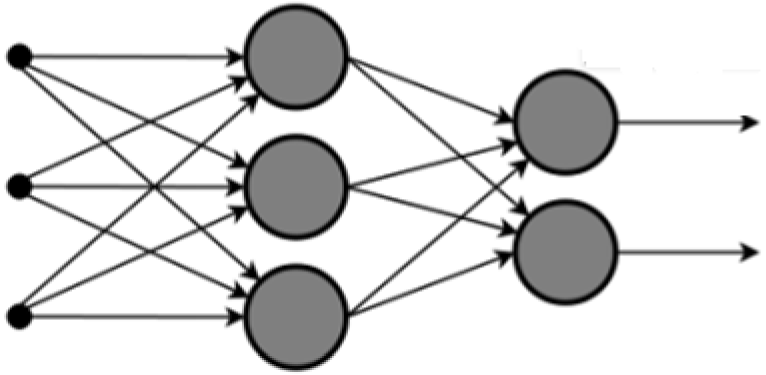
Parameters	Human brains have ~10,000,000 times more synapses than artificial neural networks
Topology	Human brains have no “layers”. Topology is complicated.
Async	The human brain works asynchronously, ANNs work synchronously.
Learning algorithm	ANNs use gradient descent for learning. Human brains use … (we don’t know).
Processing speed	Single biological neurons are slow, while standard neurons in ANNs are fast.
Power consumption	Biological neural networks use very little power compared to artificial networks
Stages	Biological networks usually don't stop / start learning. ANNs have different fitting (train) and prediction (evaluate) phases.
Distributed computations on a large scale	Both of them

Neuron: Forward Pass

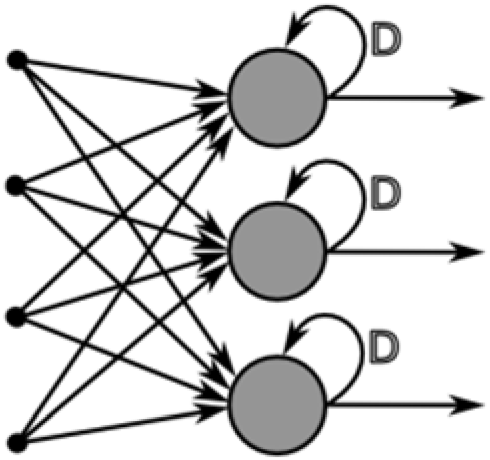


https://appliedgo.net/perceptron/

Combining Neurons into Layers



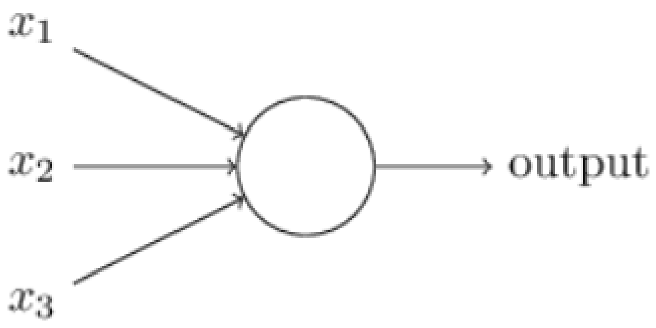
Feed Forward Neural Network

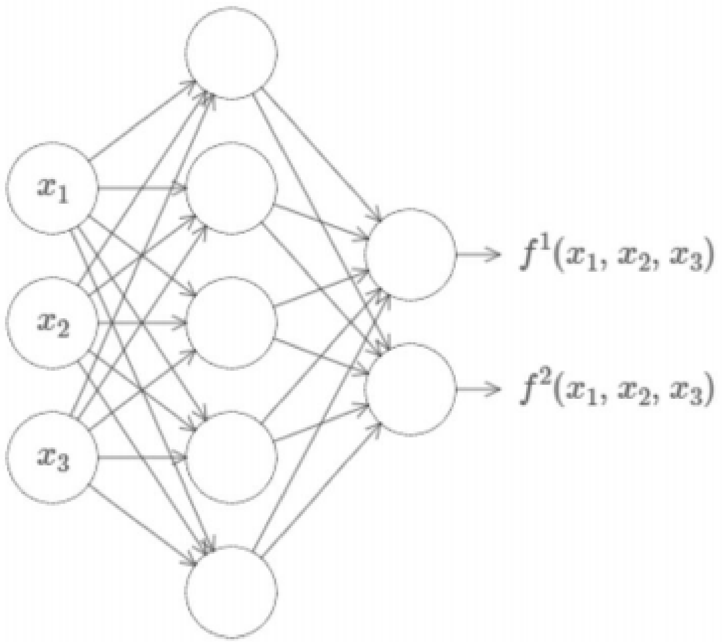


Recurrent Neural Network

- Have state memory
- Are hard to train

Combining Neurons in Hidden Layers





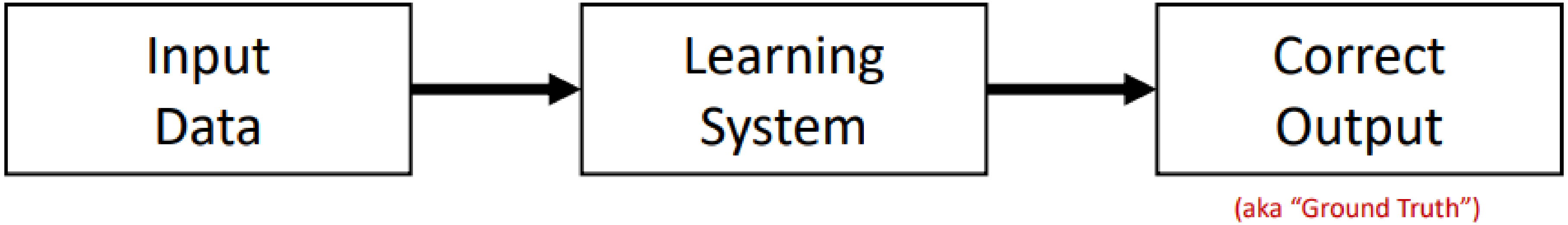
Universality: For any arbitrary function f(x), there exists a neural network that closely approximate it for any input x

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

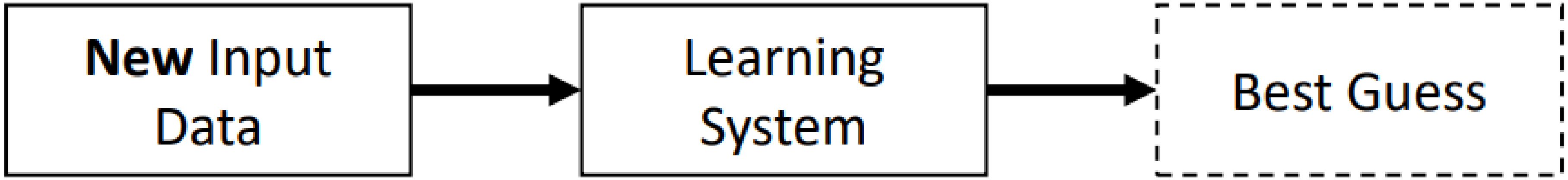
* Given that we have good algorithms for training these networks.

Deep Learning: Training and Testing

Training Stage:

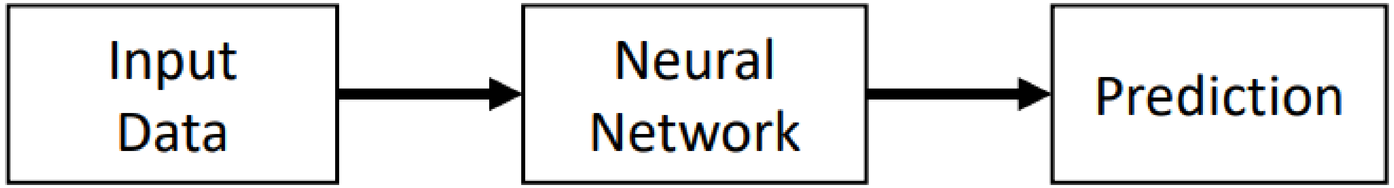


Testing Stage:

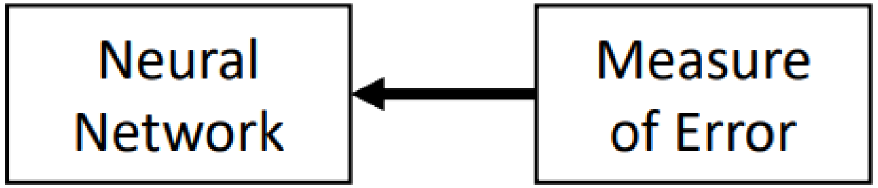


How neural networks learn?

Forward Pass:

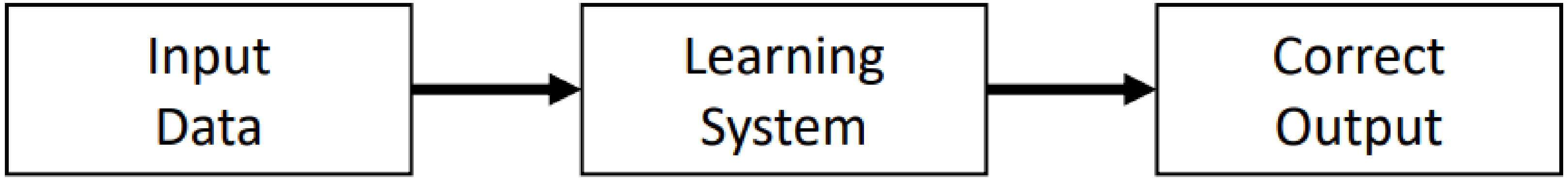


Backward Pass (aka Backpropagation):



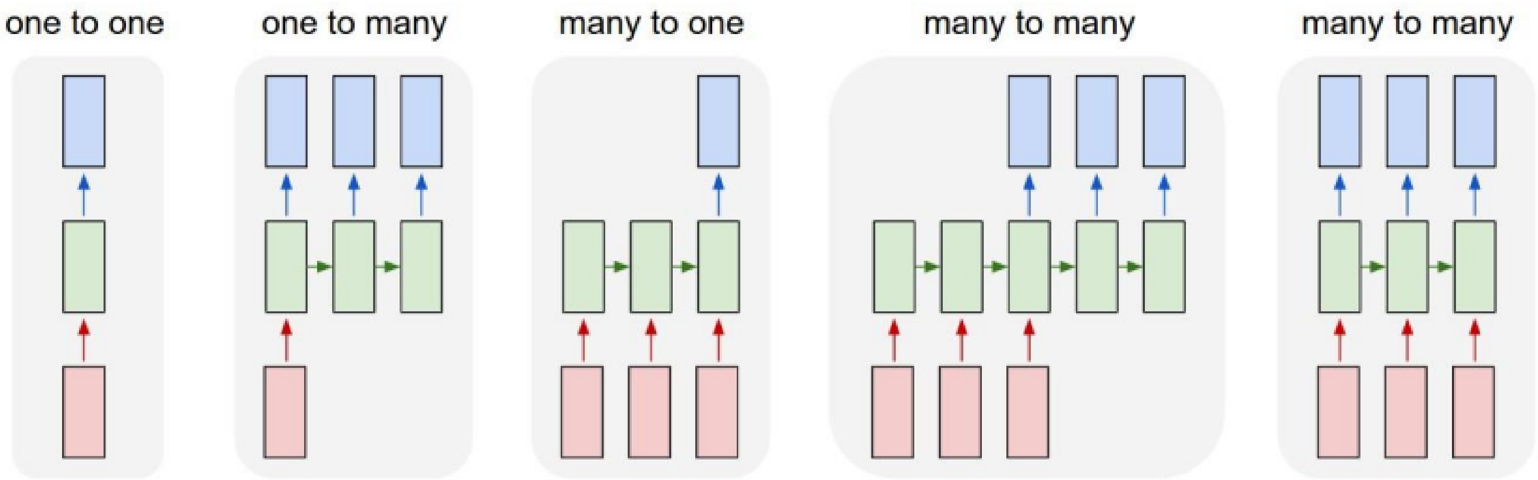
Adjust to Reduce Error

What can we do with Deep Learning?

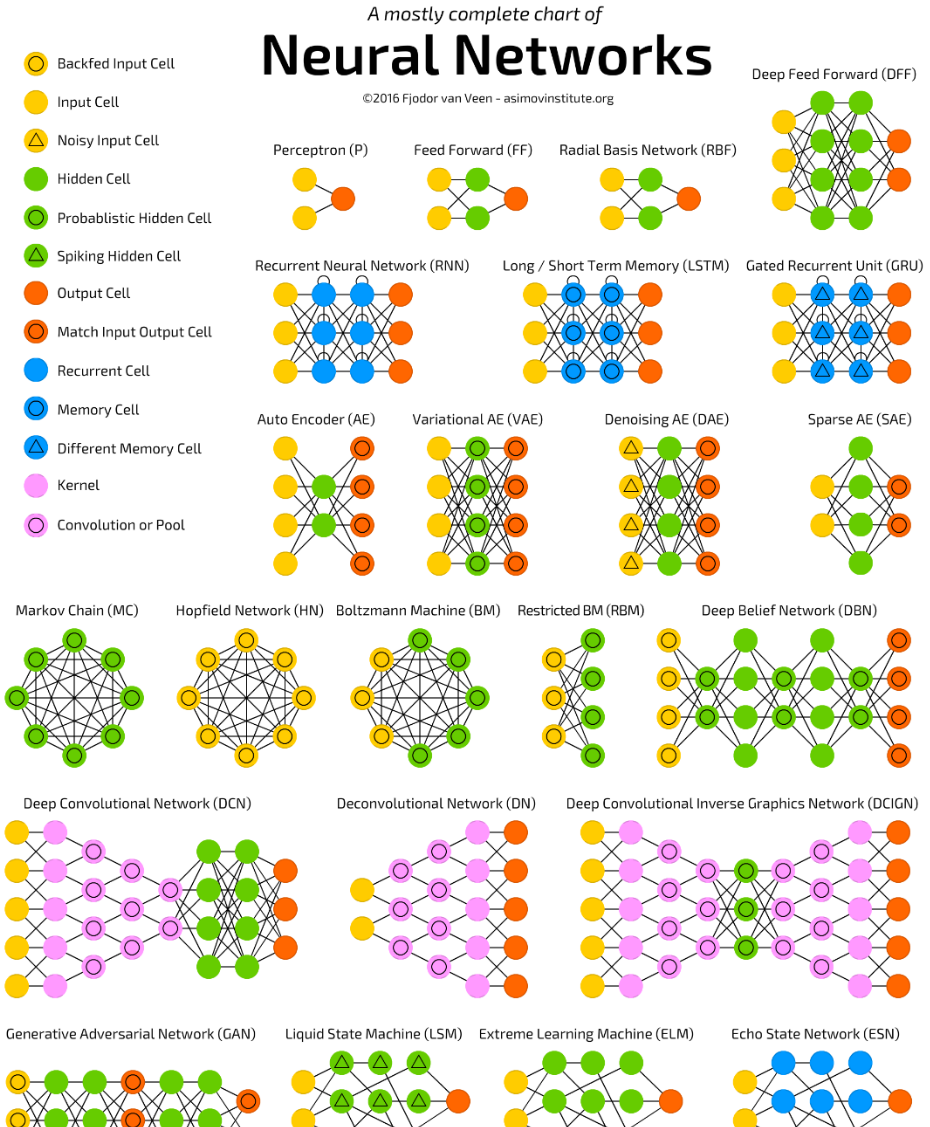


- Number
- Vector of numbers
- Sequence of numbers
- Sequence of vectors of numbers

- Number
- Vector of numbers
- Sequence of numbers
- Sequence of vectors of numbers



Various deep learning architectures



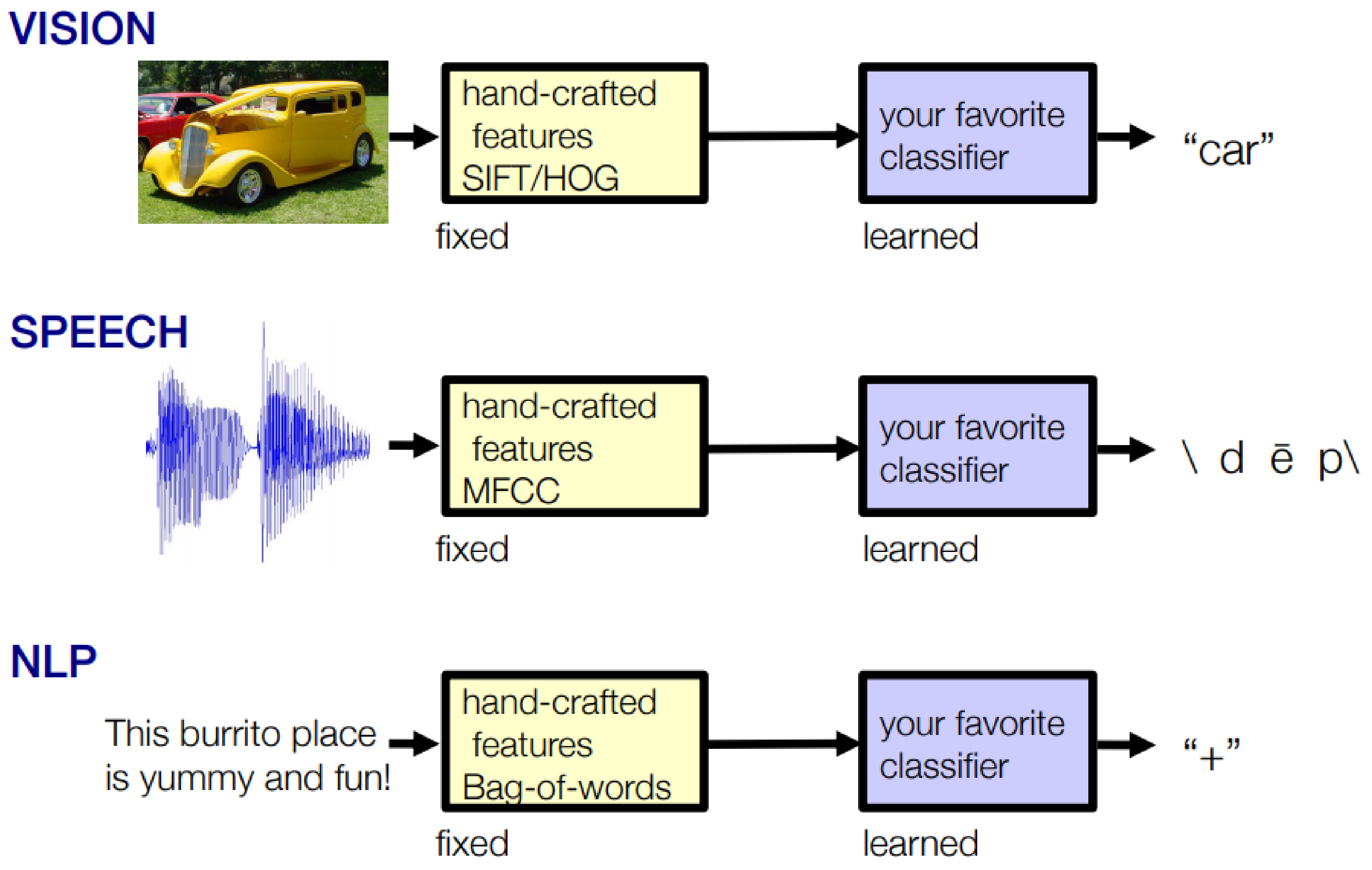
- Various terms, difficult to keep up with
  - **MLP**: Multi-layer Perceptron
  - **DNN**: Deep neural networks
  - **RNN**: Recurrent neural networks
    - LSTM: Long Short-Term Memory
  - **CNN**: Convolutional neural networks
  - **DBN**: Deep Belief Networks

http://www.asimovinstitute.org/neural-network-zoo

What is good about deep learning?

- Hierarchical Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations
- End-to-end Learning
  - Learning goal-driven representations
  - Learning to extract features
- Distributed Representations
  - No single neuron “encodes” everything
  - Groups of neurons work together

Traditional Machine Learning



Hierarchical Compositionality

VISION

pixels → edge → texton → motif → part → object

SPEECH

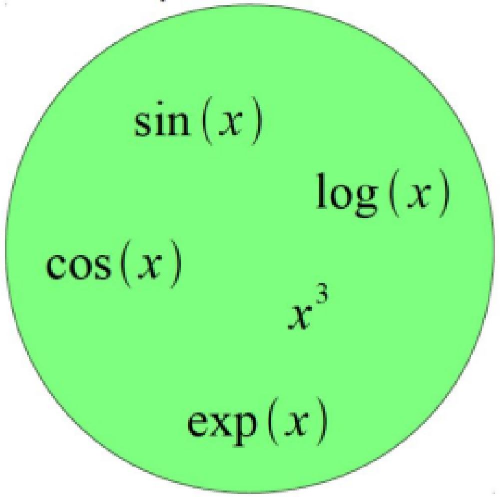
sample → spectral band → formant → motif → phone → word

NLP

character → word → NP/VP/.. → clause → sentence → story

Building a Complicated Function

Given a library of simple functions



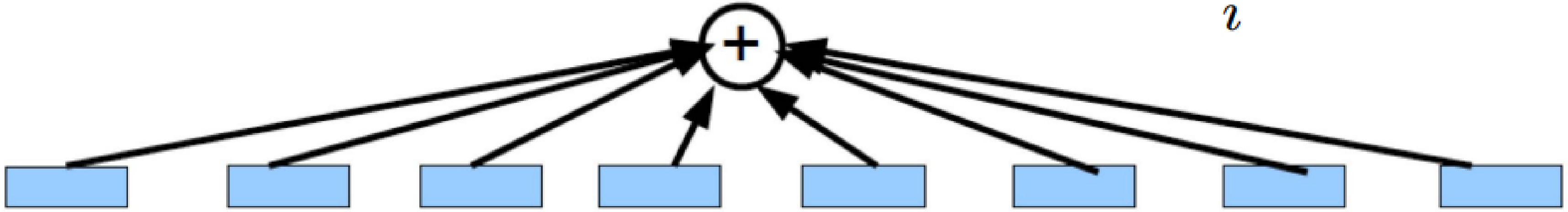
Compose into a

complicate function

Idea 1: Linear Combinations

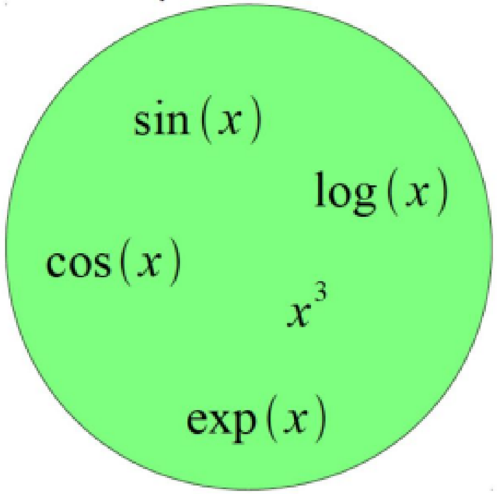
- Boosting
- Kernels
- …

$$f(x) = \sum_i \alpha_i g_i(x)$$



Building a Complicated Function

Given a library of simple functions

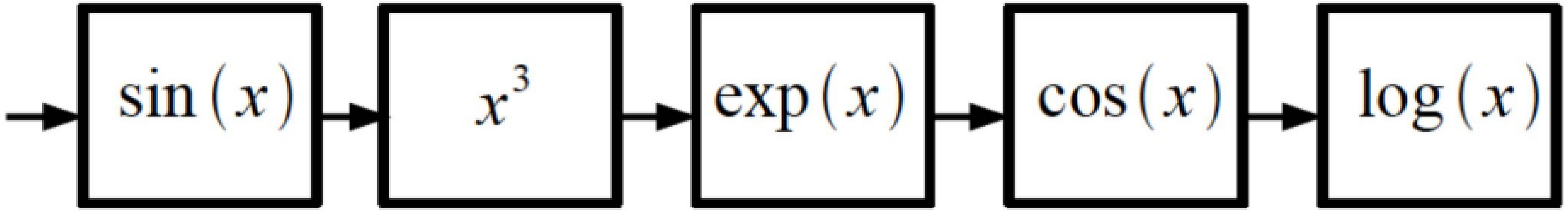


Compose into a complicate function

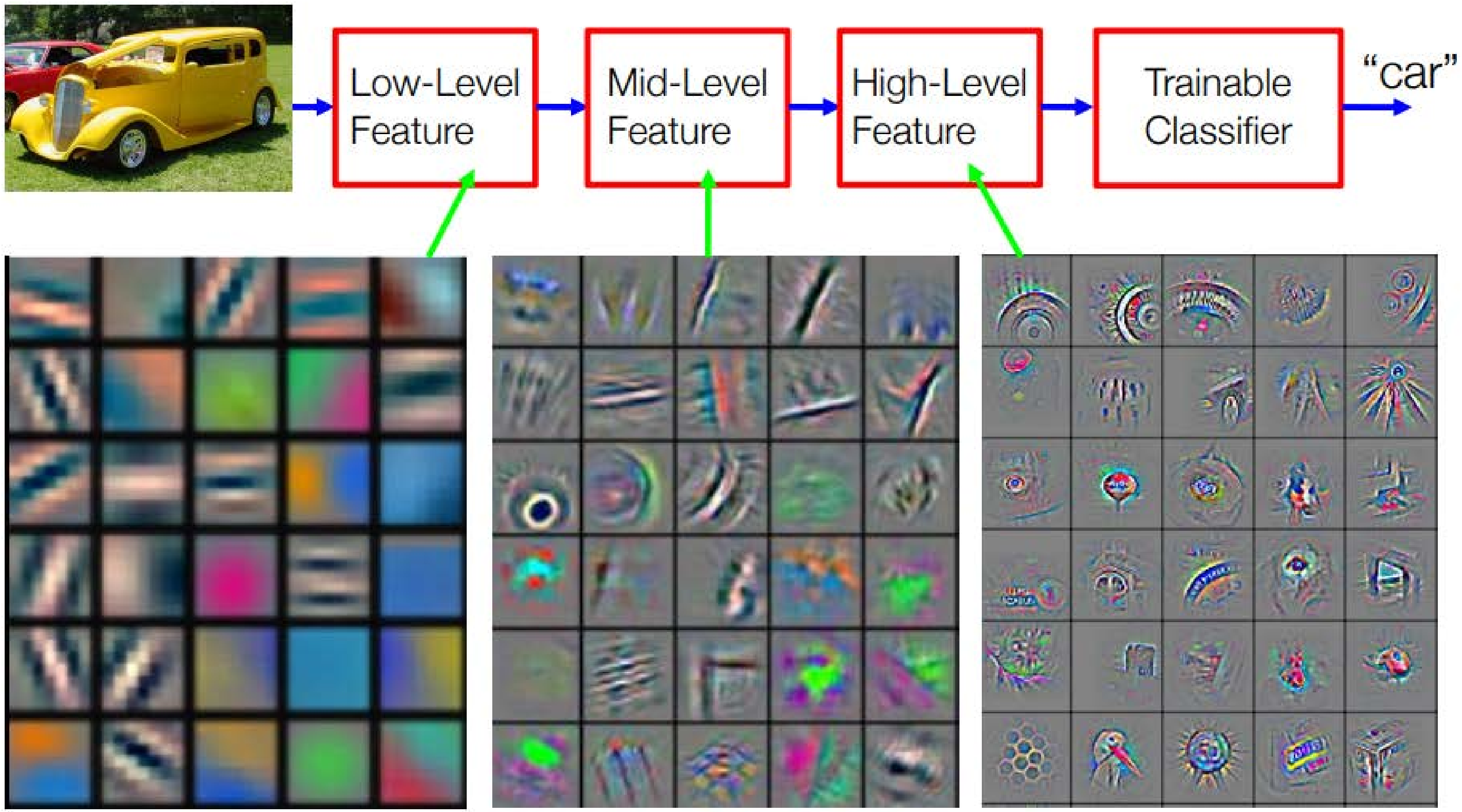
Idea 2: Compositions

- Deep Learning
- Grammar models
- Scattering transforms…

$$f(x) = \log(\cos(\exp(\sin^3(x))))$$

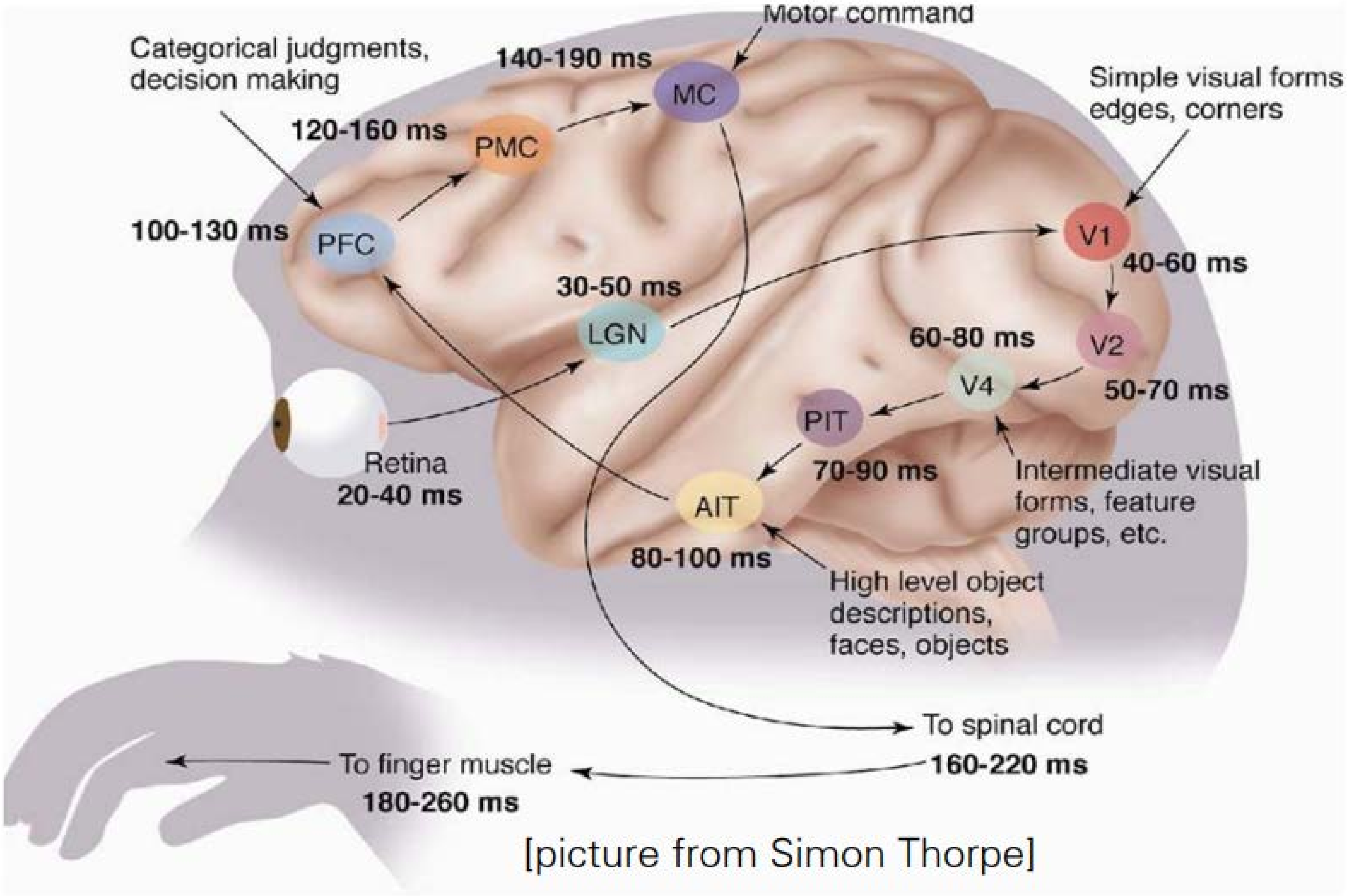


Hierarchical Composionality



M.D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks", In ECCV 2014

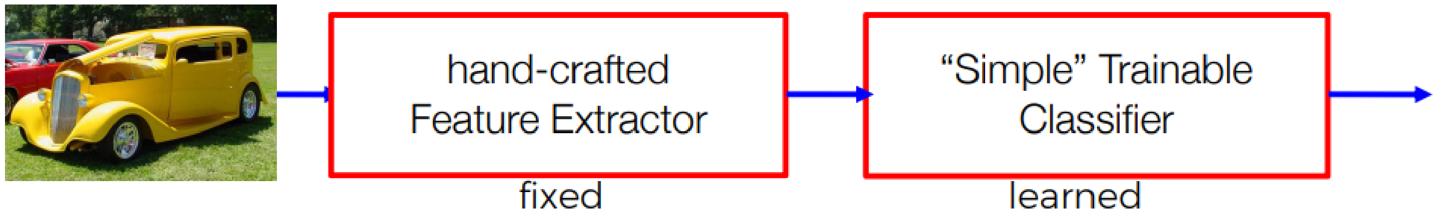
The Mammalian Visual Cortex is Hierarchical



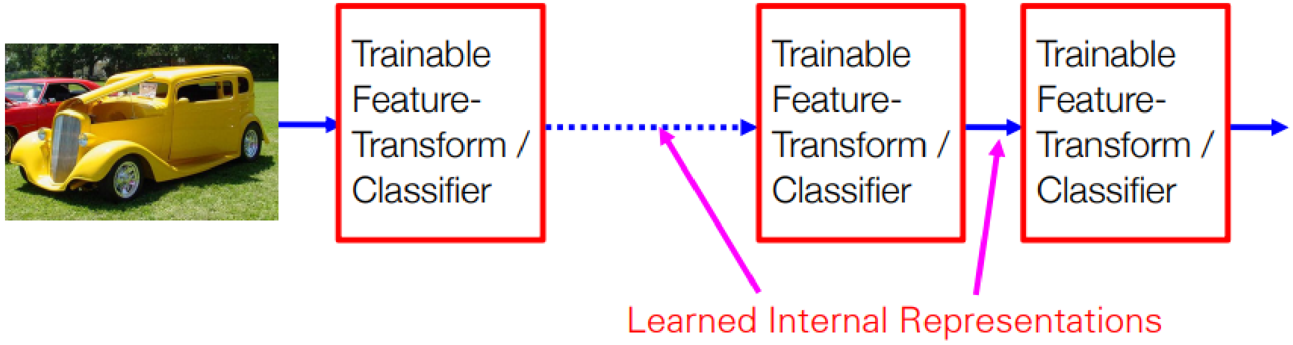
End-to-end Learning

- A hierarchy of trainable feature transforms
  - Each module transforms its input representation into a higher-level one
  - High-level features are more global and more invariant
  - Low-level features are shared among categories

“Shallow” models



Deep models

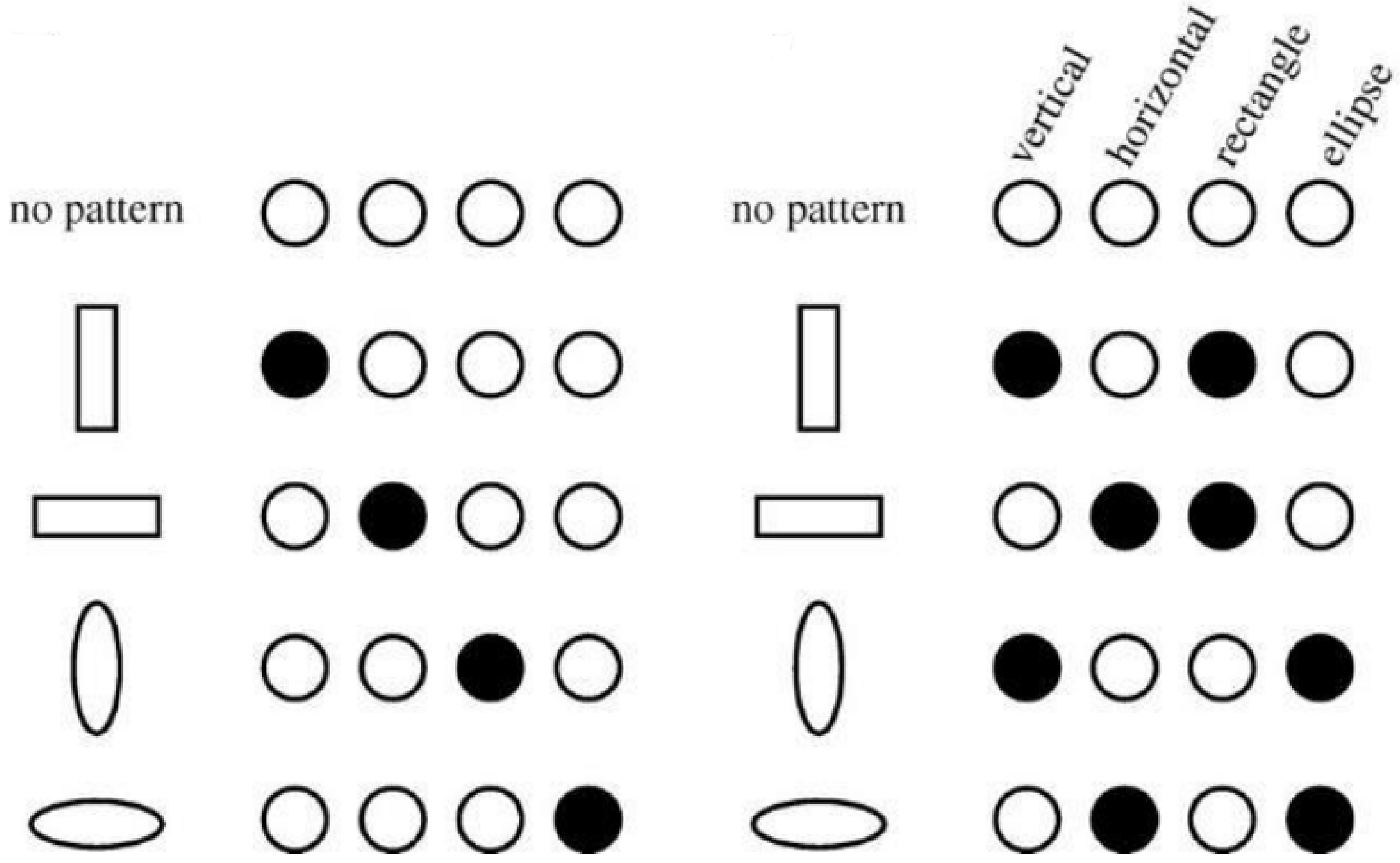


Distributed representations

- Localist models dedicate one neuron to each thing.
  - Easy to understand, code and learn
    - E.g. each cluster corresponds to one neuron
  - But they are inefficient whenever the data has componential structure

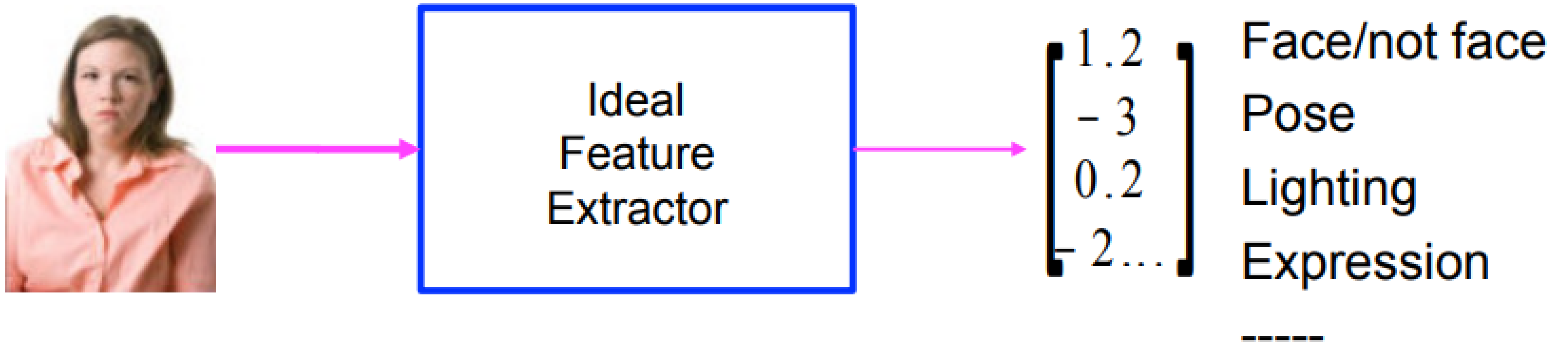
- Distributed representations means a many-to-many relationships between two types of representations (such as concepts and neurons)
  - Each concept is represented by many neurons
  - Each neuron participates in the representation of many concepts

Local vs Distributed Toy Example



Power of distributed representations

- Example: all face images of a person
  - 1000x1000 pixels: 1000000 dimensions
  - But the face has 3 Cartesian coordinates and 3 euler angles
  - And humans have less than about 50 muscles in the face
  - Hence the manifold of face images of a person has <56 dimensions
- Perfect representation of a face image:
  - Its coordinates on the face manifold
  - Its coordinates away from the manifold



https://computing.ece.vt.edu/~f15ece6504/slides/L1_intro.pptx.pdf

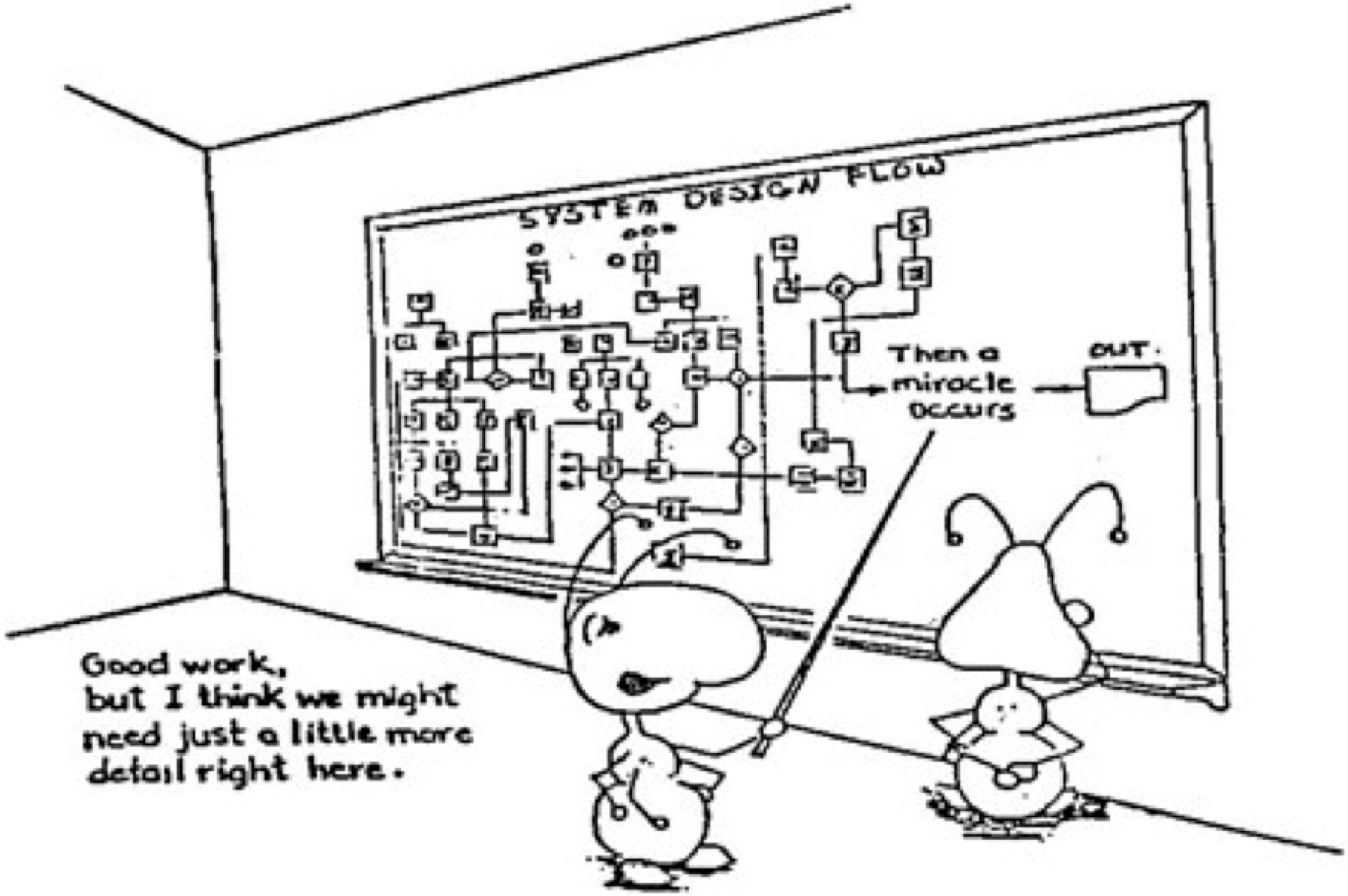
Benefits of Deep/Representation Learning

- (Usually) Better Performance
  - “Because gradient descent is better than you”, Yann LeCun

- New domains without experts
  - e.g. Gene-expression data
  - Unclear how to hand-engineer

- Expert intuition might be misleading
  - “Every time I fire a linguist, the performance of our speech recognition system goes up”, Fred Jelinik, IBM 1998

Problems with Deep Learning



Problems with Deep Learning

- Problem #1: Non-Convex! Non-Convex! Non-Convex!
  - “Yes, but all interesting learning problems are non-convex”
- Problem #2: Hard to track down what’s failing
  - In end-to-end systems, it’s hard to know why things are not working
  - Tricks such as visualizing features, adding losses at different layers
- Problem #3: Lack of easy reproducibility
  - Direct consequence of stochasticity & non-convexity
  - It is getting better with toolkits/libraries/frameworks e.g. Caffe, Theano, Tensorflow

Applications

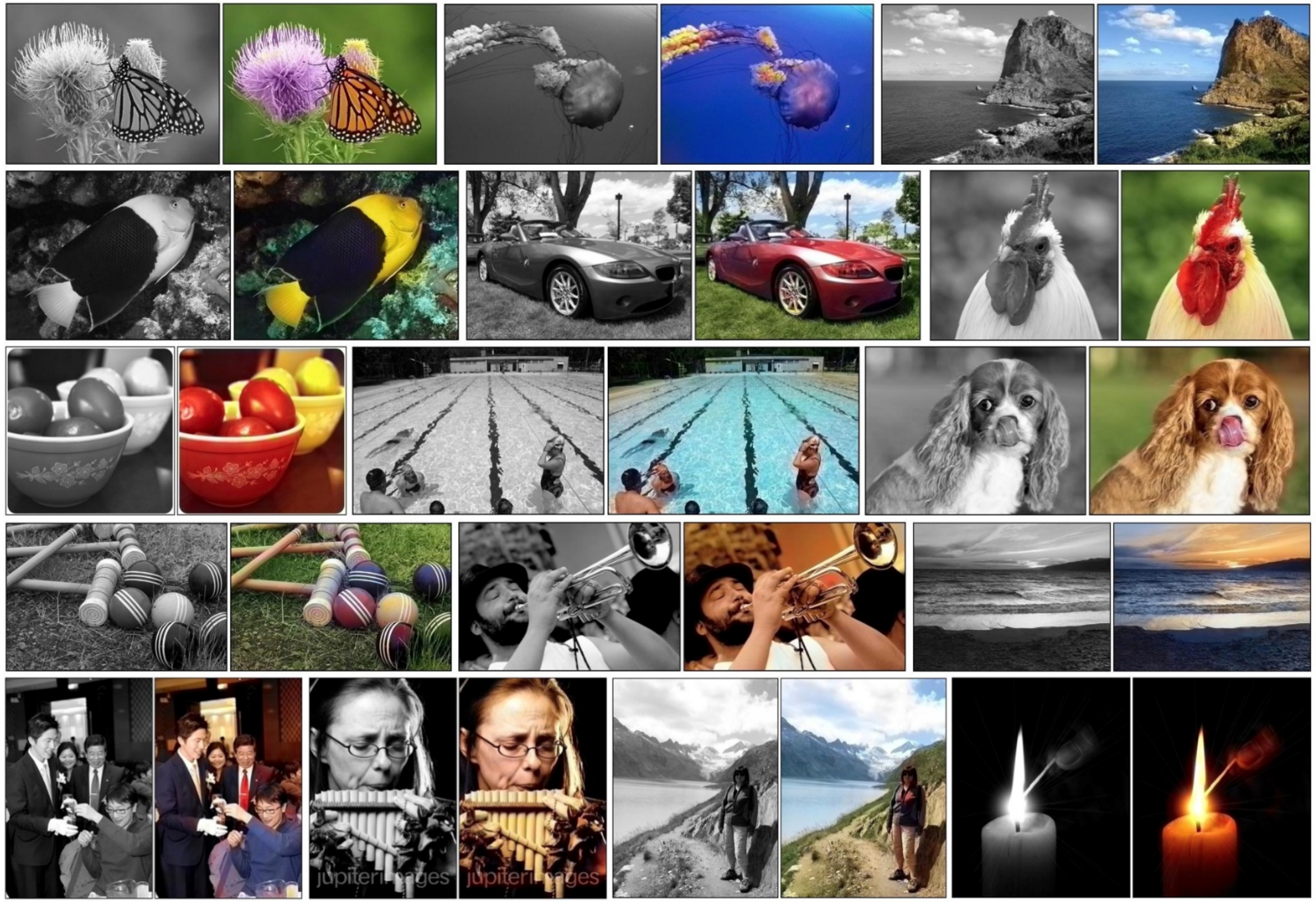
- Computer Vision
- Speech Recognition
- Natural Language Understanding/Generation
- Computer Games
- Computer Animation
- Human-robot interaction
- Self-driving cars
- Recommender systems
- …

Object Detection and Segmentation

- P. O. Pinheiro, R. Collobert and P, Dollár, "Learning to Segment Object Candidates", NIPS 2015

- P. O. Pinheiro, T.-Y. Lin, R. Collobert and P. Dollár, "Learning to Refine Object Segments", ECCV 2016

Background Removal & Colorization of Images



https://towardsdatascience.com/background-removal-with-deep-learning-c4f2104b3157

Zhang, R., Isola, P., & Efros, A. A. (2016, October). Colorful image colorization. In *European Conference on Computer Vision* (pp. 649-666). Springer, Cham.

Image and video captioning

Andrej Karpathy’s NeuralTalk, Stanford University, director of AI at Tesla
https://www.youtube.com/watch?v=8BFzu9m52sc
https://cs.stanford.edu/people/karpathy/
http://karpathy.github.io/

Question Answering

The first recorded travels by Europeans to China and back date from this time. The most famous traveler of the period was the Venetian Marco Polo, whose account of his trip to "Cambaluc," the capital of the Great Khan, and of life there astounded the people of Europe. The account of his travels, Il milione (or, The Million, known in English as the Travels of Marco Polo), appeared about the year 1299. Some argue over the accuracy of Marco Polo's accounts due to the lack of mentioning the Great Wall of China, tea houses, which would have been a prominent sight since Europeans had yet to adopt a tea culture, as well the practice of foot binding by the women in capital of the Great Khan. Some suggest that Marco Polo acquired much of his knowledge through contact with Persian traders since many of the places he named were in Persian.

How did some suspect that Polo learned about China instead of by actually visiting it?

Answer: through contact with Persian traders

P. Rajpurkar, J. Zhang, K. Lopyrev, P. Liang. "SQuAD: 100,000+ Questions for Machine Comprehension of Text", In EMNLP, 2016.

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)

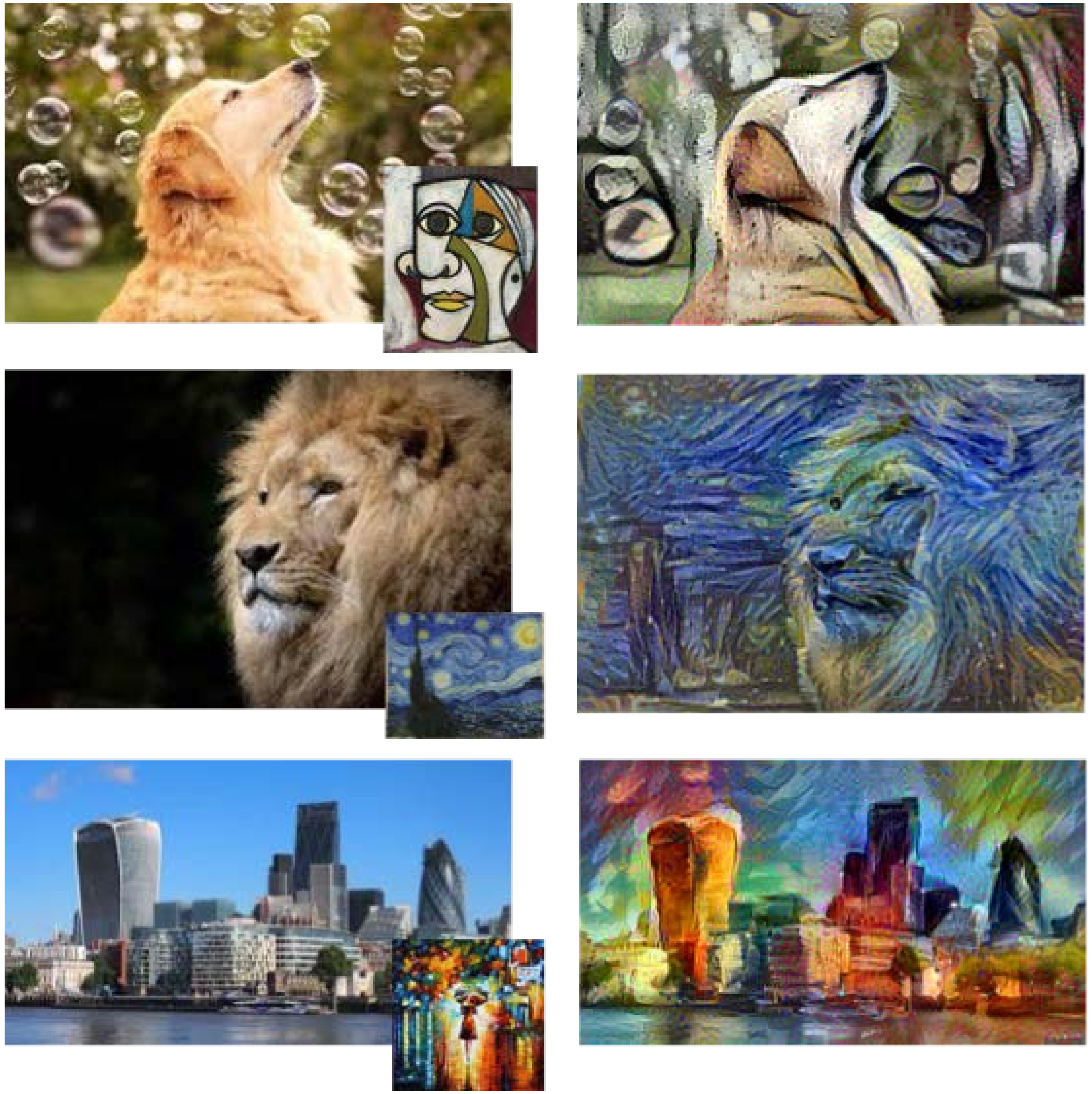
DAQUAR 1522
How many chairs are there?
Ground truth: two
IMG+BOW: four (0.24)
2-VIS+BLSTM: one (0.29)
LSTM: four (0.19)

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)

M. Ren, R. Kiros, and R. Zemel, "Exploring Models and Data for Image Question Answering" NIPS 2015

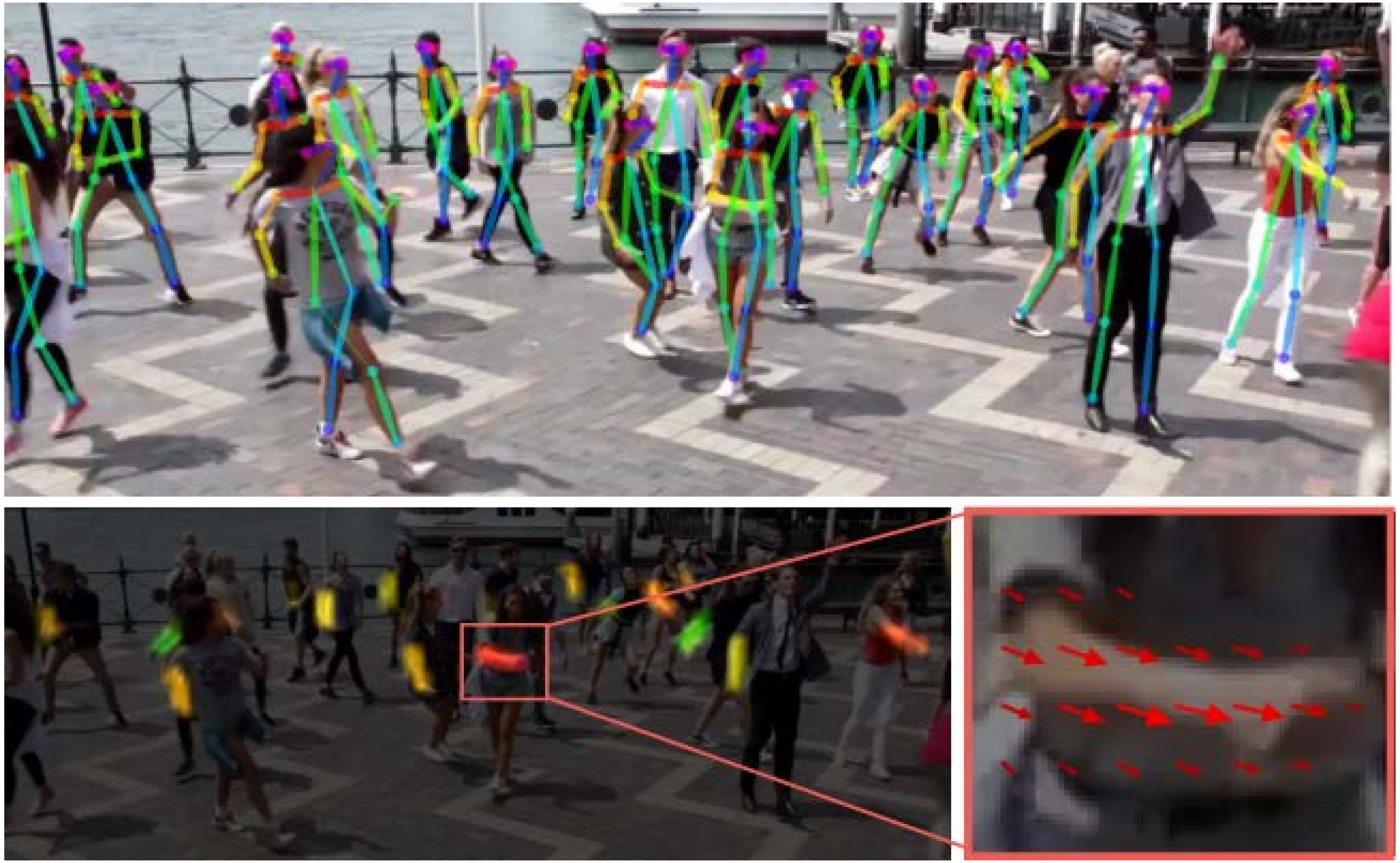
Style Transfer

Leon A. Gatys, Alexander S. Ecker, Matthias Bethge , “Image Style Transfer Using Convolutional Neural Networks”, CVPR 2016



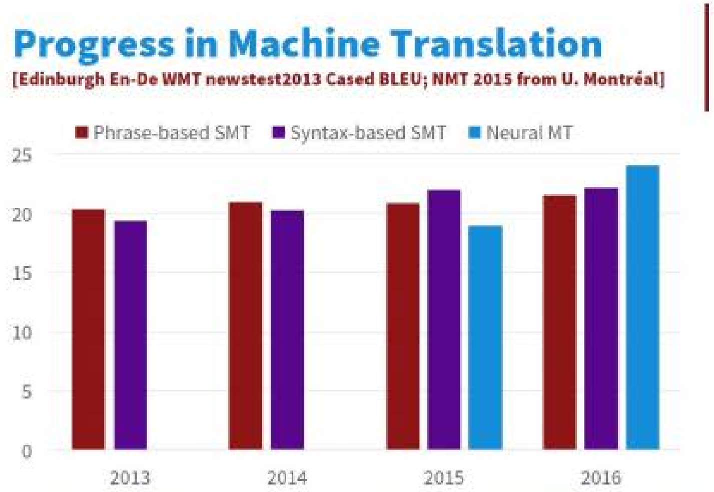
Human Pose Estimation

Z. Cao ,T. Simon, S.–E. Wei and Yaser Sheikhr, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", CVPR 2017

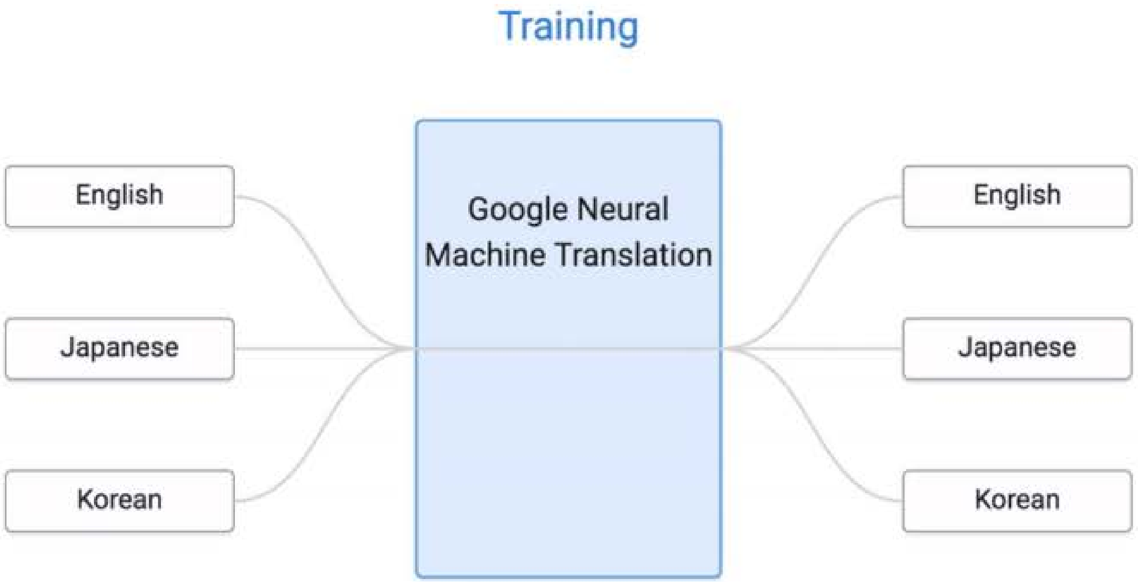


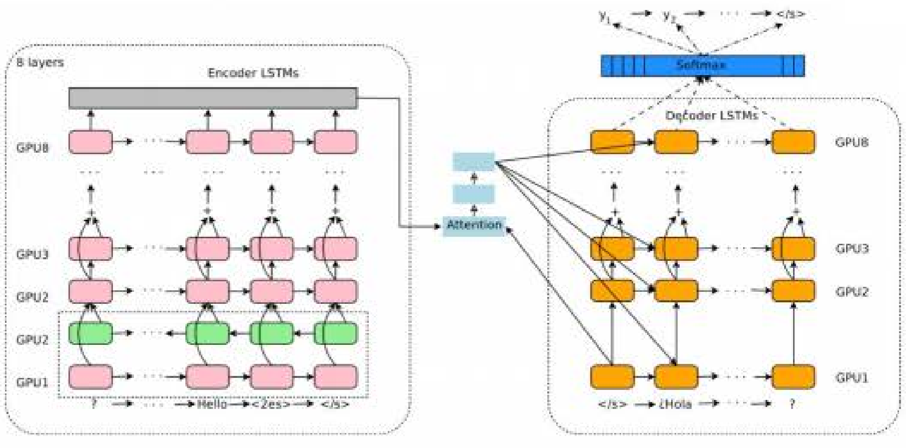
https://www.youtube.com/watch?v=pW6nZXeWlGM

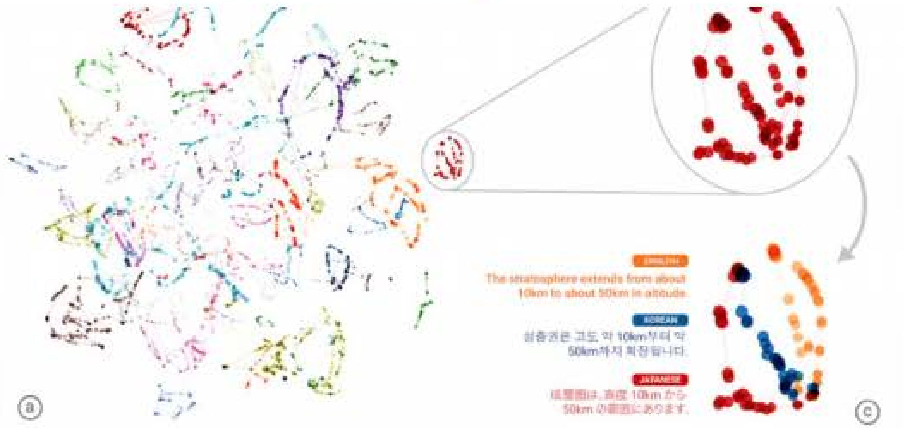
Machine Translation



From [Sennrich 2016, http://www.meta-net.eu/events/meta-forum-2016/slides/09_sennrich.pdf]



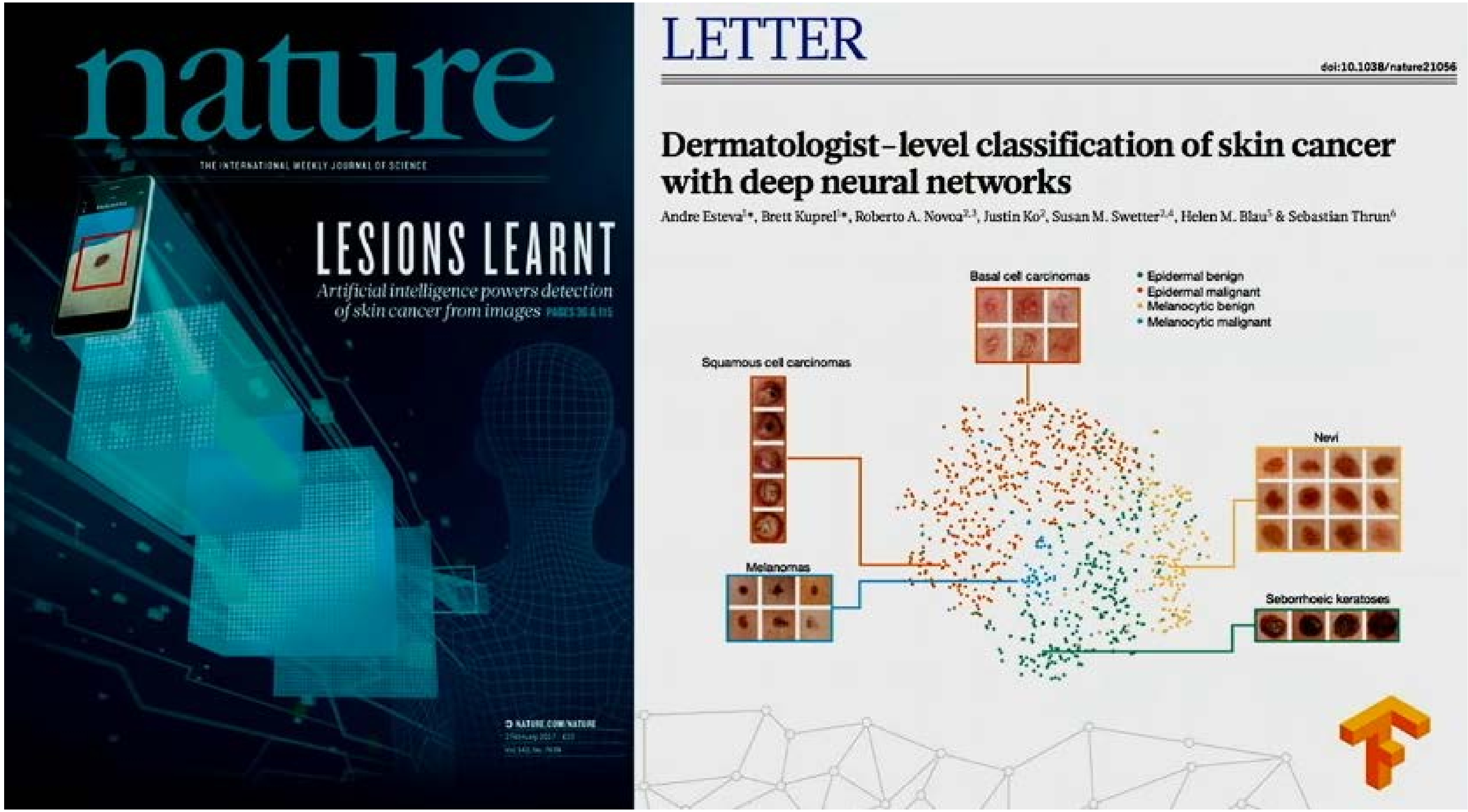




M. Johnson et al., "Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation", arXiv preprint, 2016
https://www.youtube.com/watch?v=nR74lBO5M3s

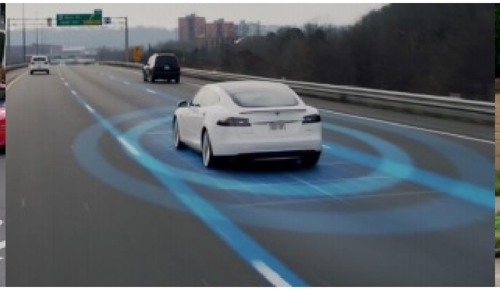
Medical Image Analysis

A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks", Nature 542, 2017



Self-Driving Cars


Waymo / Google Self-Driving Car

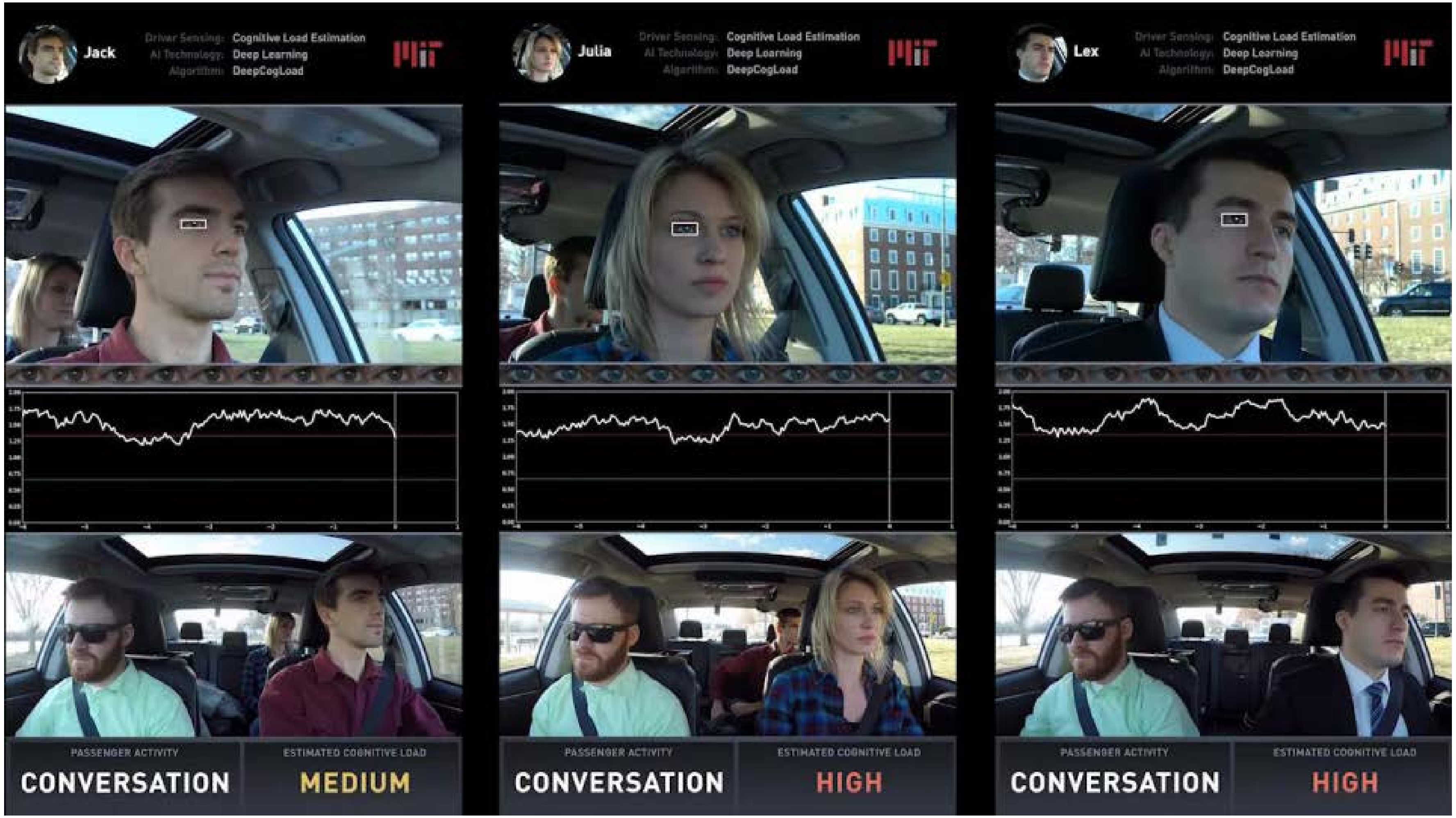

Tesla Autopilot

Uber

nuTonomy

https://selfdrivingcars.mit.edu

Emotion and Cognitive Load Detection



https://hcai.mit.edu/hcav
https://www.affectiva.com

Human vs Machine

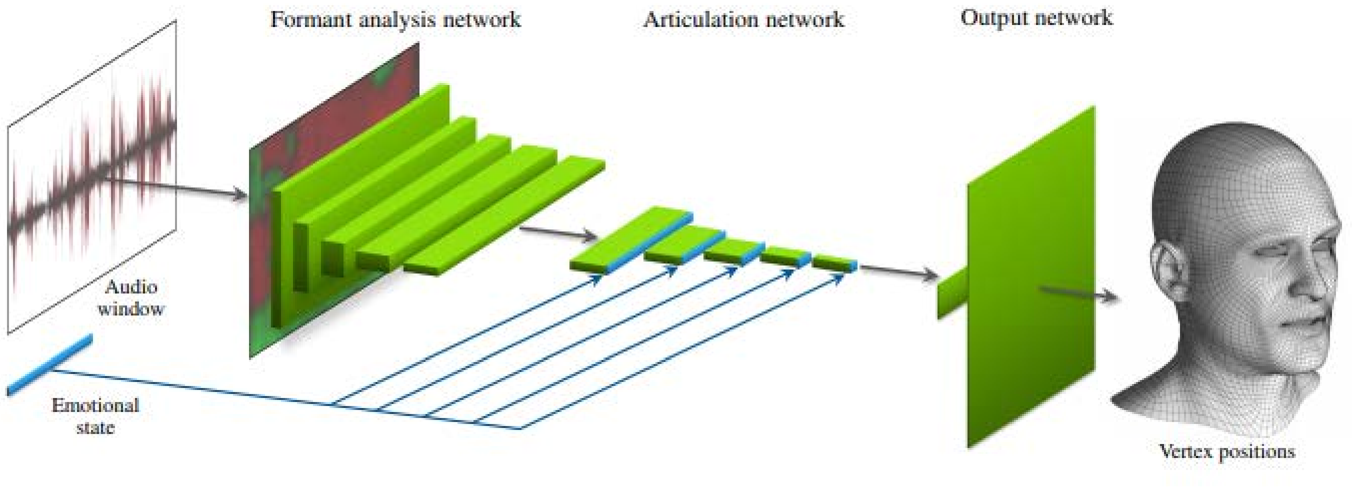
Deep Learning:
Learn effective perception-control from **data**

Solve the perception-control problem where **possible**:

Deep Learning:
Learn effective human-robot interaction from **data**

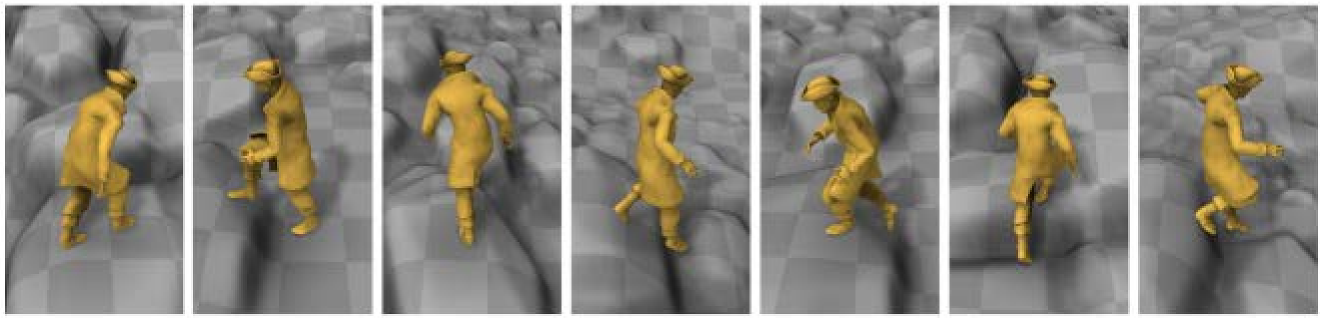
And where **not possible**: involve the human

Computer Animation



https://www.youtube.com/watch?v=lDzrfdpGqw4

Tero Karras, Timo Aila, Samuli Laine, Antti Herva, and Jaakko Lehtinen. 2017. Audio-driven facial animation by joint end-to-end learning of pose and emotion. ACM Trans. Graph. 36, 4, Article 94 (July 2017)



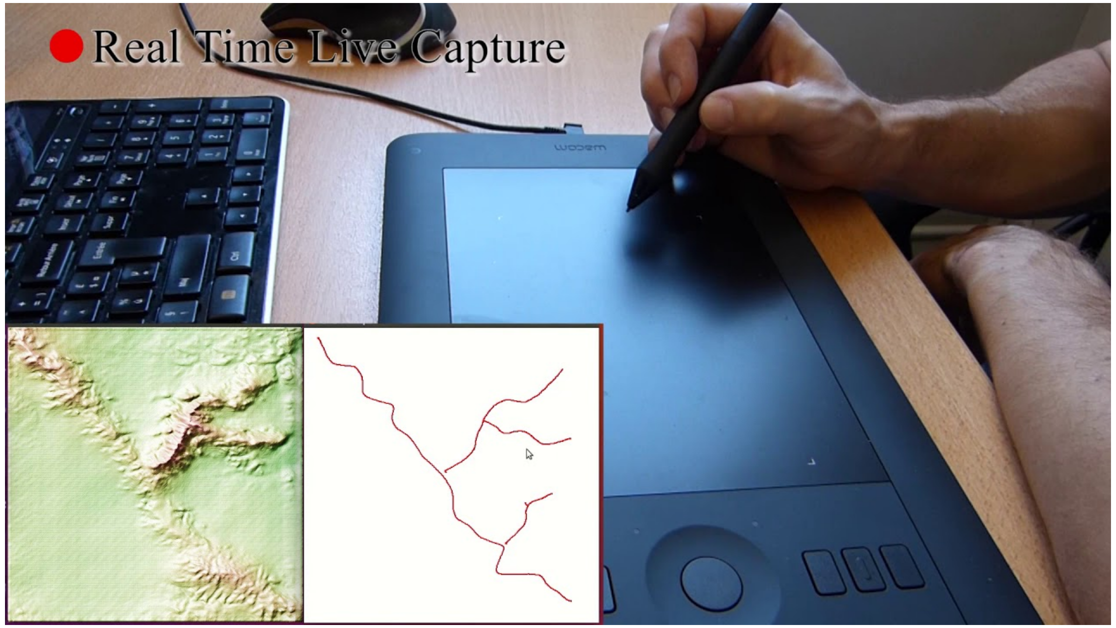
https://www.youtube.com/watch?v=Ul0Gilv5wvY

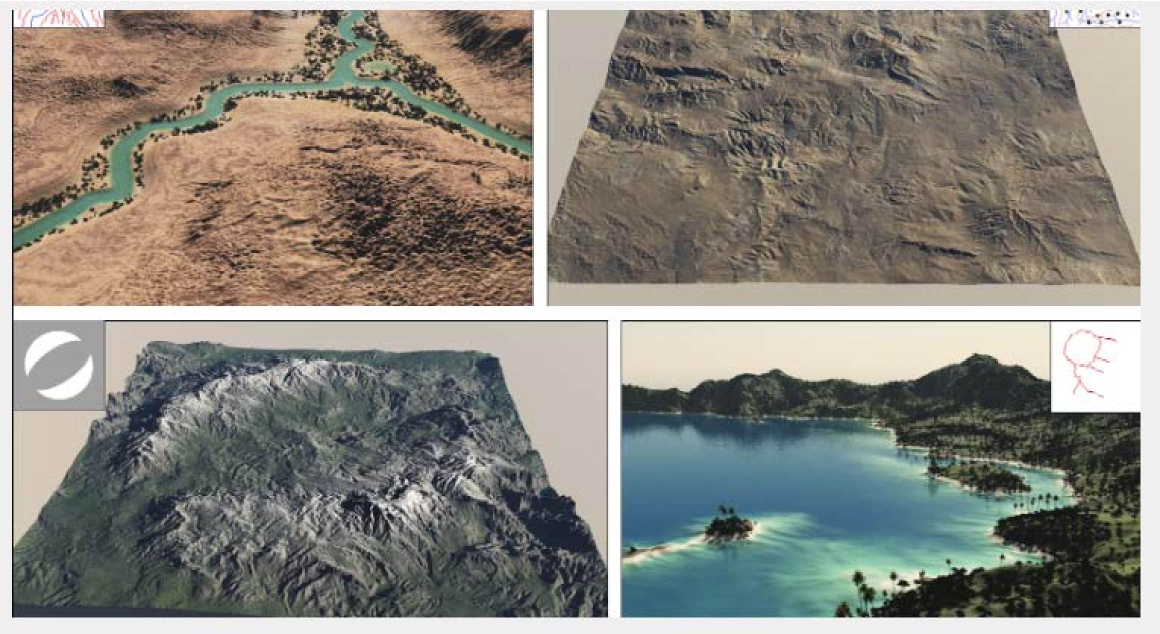
Daniel Holden, Taku Komura, and Jun Saito. 2017. Phase-functioned neural networks for character control. ACM Trans. Graph. 36, 4, Article 42 (July 2017)

Image Generation

T.Karras, T.Aila, S.Laine and J, Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation", ICLR 2018

Procedural Content Generation for Games





Éric Guérin, Julie Digne, Éric Galin, Adrien Peytavie, Christian Wolf, Bedrich Benes, and Benoît Martinez. 2017. Interactive example-based terrain authoring with conditional generative adversarial networks. ACM Trans. Graph. 36, 6, Article 228 (November 2017)
https://www.youtube.com/watch?time_continue=1&v=5w685udM838

Robotics (Sensori-motor deep learning)

http://rll.berkeley.edu/deeplearningrobotics

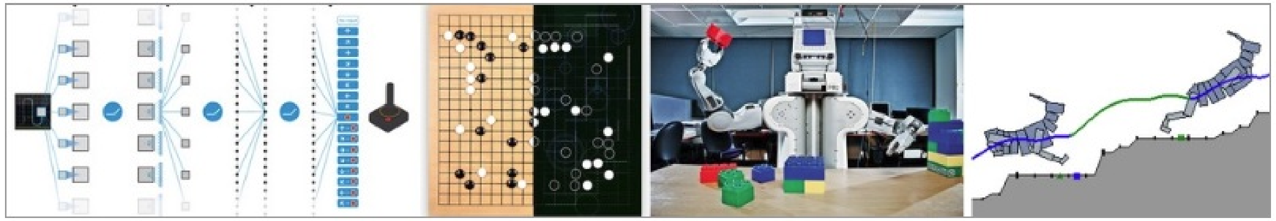
Towards Artificial General Intelligence

Andrej Karpathy blog   About   Hacker's guide to Neural Networks

Deep Reinforcement Learning: Pong from Pixels

May 31, 2016

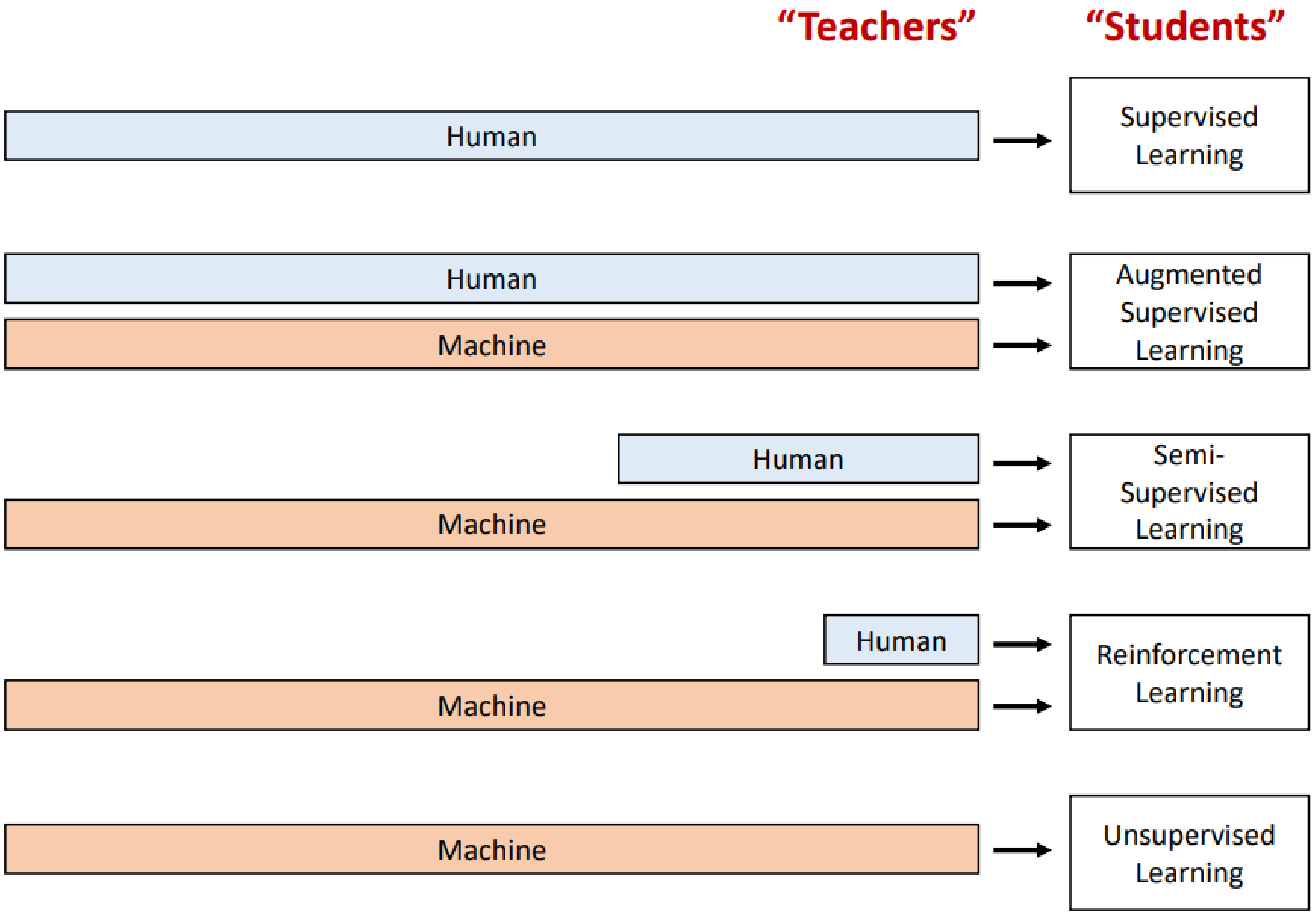
This is a long overdue blog post on Reinforcement Learning (RL). RL is hot! You may have noticed that computers can now automatically learn to play ATARI games (from raw game pixels!), they are beating world champions at Go, simulated quadrupeds are learning to run and leap, and robots are learning how to perform complex manipulation tasks that defy explicit programming. It turns out that all of these advances fall under the umbrella of RL research. I also became interested in RL myself over the last ~year: I worked through Richard Sutton's book, read through David Silver's course, watched John Schulmann's lectures, wrote an RL library in Javascript, over the summer interned at DeepMind working in the DeepRL group, and most recently pitched in a little with the design/development of OpenAI Gym, a new RL benchmarking toolkit. So I've certainly been on this funwagon for at least a year but until now I haven't gotten around to writing up a short post on why RL is a big deal, what it's about, how it all developed and where it might be going.



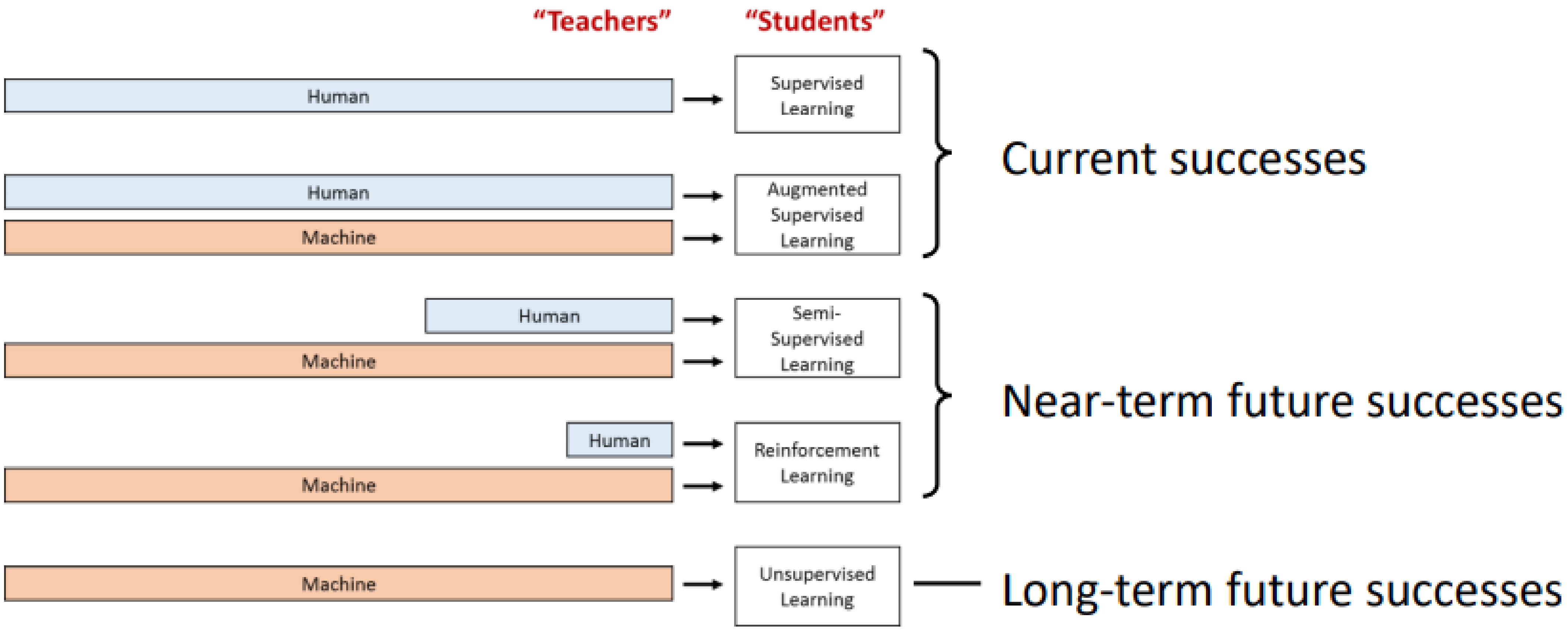
Examples of RL in the wild. **From left to right**: Deep Q Learning network playing ATARI, AlphaGo, Berkeley robot stacking Legos, physically-simulated quadruped leaping over terrain.

https://www.youtube.com/watch?v=8tq1C8spV_g (Watch on Netflix)

Deep Learning from Human and Machine



Deep Learning from Human and Machine

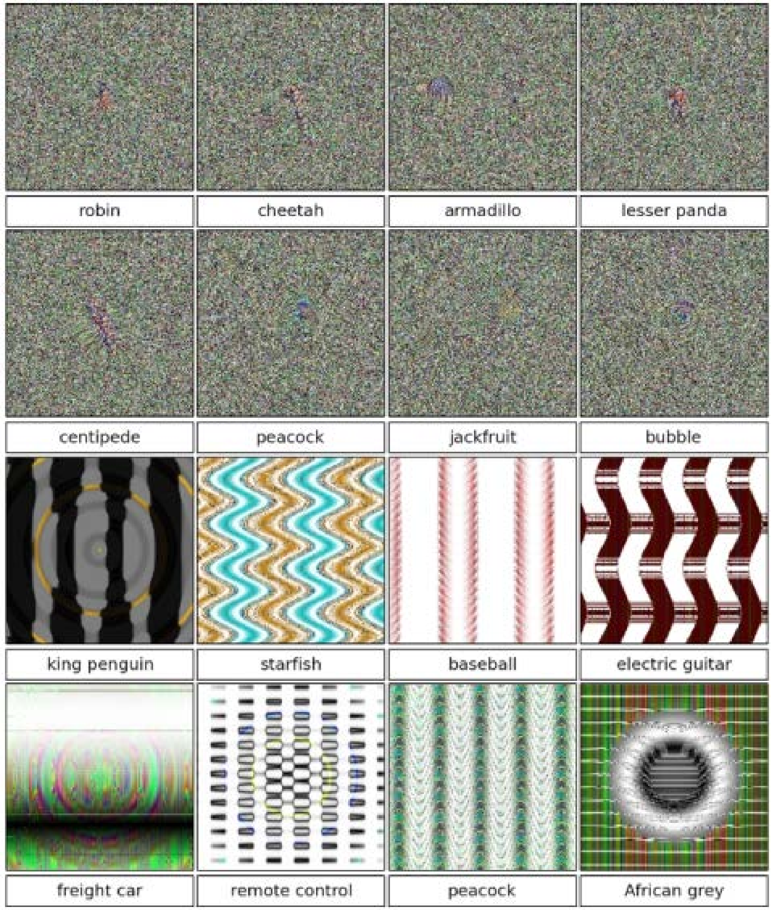


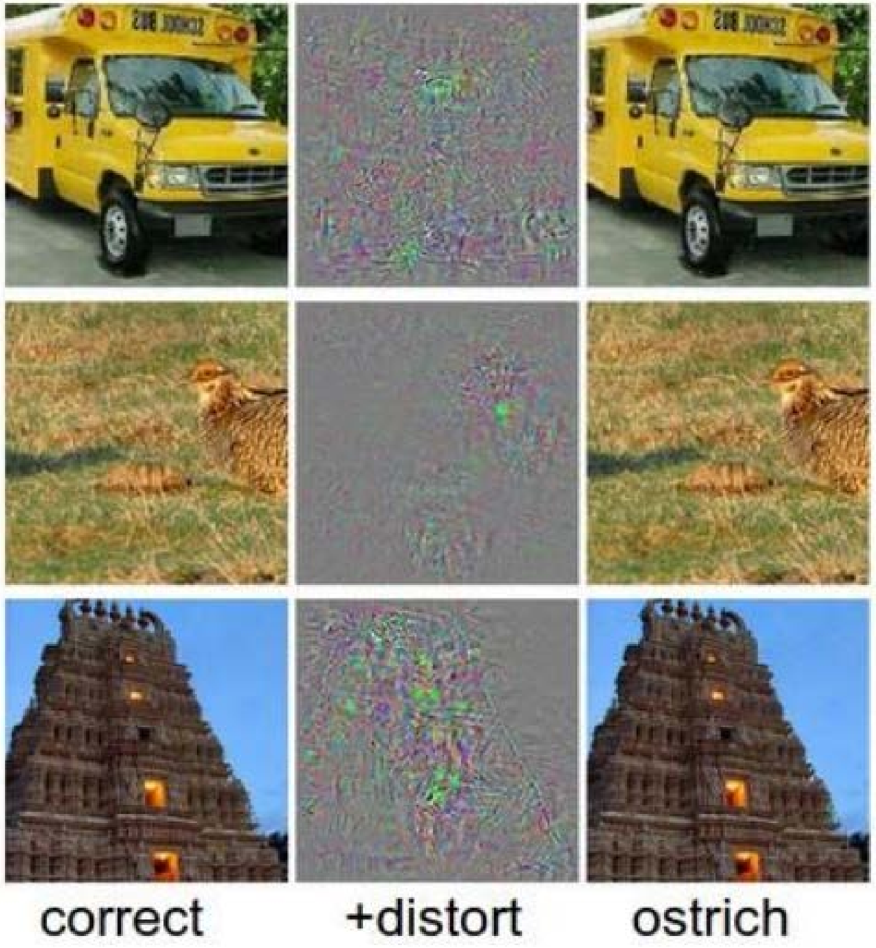
Current Challenges

- **Transfer learning:** Unable to transfer representation to most reasonably related domains
- **Requires big data:** inefficient at learning from data
- **Requires supervised data**: costly to annotate real-world data
- **Not fully automated:** Needs hyperparameter tuning: learning rate, loss function, mini-batch size, number of iterations, optimization algorithm
- **Rewards:** Defining a good reward function is difficult
- **Transparency:** Black boxes, despite tools to visualize what is happening
- **Edge cases:** Deep learning is not good at dealing with edge cases, requires training for edge cases

Deep Neural Networks are easily fooled

>99.6% Confidence in the Wrong Answer





Nguyen A, Yosinski J, Clune J. Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. In Computer Vision and Pattern Recognition (CVPR ’15), IEEE, 2015.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I. J. & Fergus, R. (2013). Intriguing properties of neural networks.. *CoRR*

Artificial General Intelligence

MIT 6.S099: Artificial General Intelligence

Artificial Intelligence Podcast   Vote AI   Edit Profile

This class takes an engineering approach to exploring possible research paths toward building human-level intelligence. The lectures will introduce our current understanding of computational intelligence and ways in which strong AI could possibly be achieved, with insights from deep learning, reinforcement learning, computational neuroscience, robotics, cognitive modeling, psychology, and more. Additional topics will include AI safety and ethics. Projects will seek to build intuition about the limitations of state-of-the-art machine learning approaches and how those limitations may be overcome. The course will include several guest talks. Listeners are welcome.

https://agi.mit.edu

Artificial General Intelligence
Lex Fridman (Lecture)
[Slides] - [Lecture Video]

Building machines that see, learn, and think like people
Josh Tenenbaum (Guest Talk)
Professor, MIT
[Lecture Video]

Robots that walk, run, jump, and grasp in the real world
Marc Raibert (Guest Talk)
CEO, Boston Dynamics
[Lecture Video]

Computational Universe
Stephen Wolfram (Guest Talk)
Wolfram Research
[Lecture Video]

Future of Intelligence
Ray Kurzweil (Guest Talk)
Google
[Lecture Video]

How the brain creates emotions
Lisa Feldman Barrett (Guest Talk)
Professor, Northeastern University

Supplementary reading

- Chapter 1 of Deep Learning textbook
- Deep Learning, Yann LeCun, Yoshio Bengio, Geoffrey Hinton. Nature, Vol. 521, 2015.
- Deep Learning in Neural Networks: An Overview, Juergen Schmidhuber. Neural Networks, Vol. 61, pp. 85–117, 2015. (advanced!)

- https://www.tensorflow.org/ (good to look at for the project)
- http://neuralnetworksanddeeplearning.com/ (good to look at for the project)

Further links

- Collection of resources (courses, datasets, tools, people, research groups etc.)

- https://github.com/ChristosChristofidis/awesome-deep-learning
- http://thegrandjanitor.com/2016/08/15/learning-deep-learning-my-top-five-resource/
- http://deeplearning.net/

- Summer Schools (all recent topics on deep learning, from top deep learning experts)

- http://videolectures.net/deeplearning2016_montreal/, Deep Learning Summer School 2016
- http://videolectures.net/deeplearning2017_montreal/, Deep Learning Summer School 2017

References

- MIT 6.S094 Deep Learning for Self-Driving Cars
- CMP784: Deep Learning Hacettepe University
- CPSC 532L: Multimodal Learning with Vision, Language and Sound
- (Slides are mainly adopted from the above courses)

What to do after this lecture?

- Make your groups today
- Meet with your group asap this week and start planning
- Start reading, watching, searching, thinking and learning TensorFlow NOW!!!
  - Be curious and work hard