COMP/EECE 7/8740 Neural Networks

Topics:

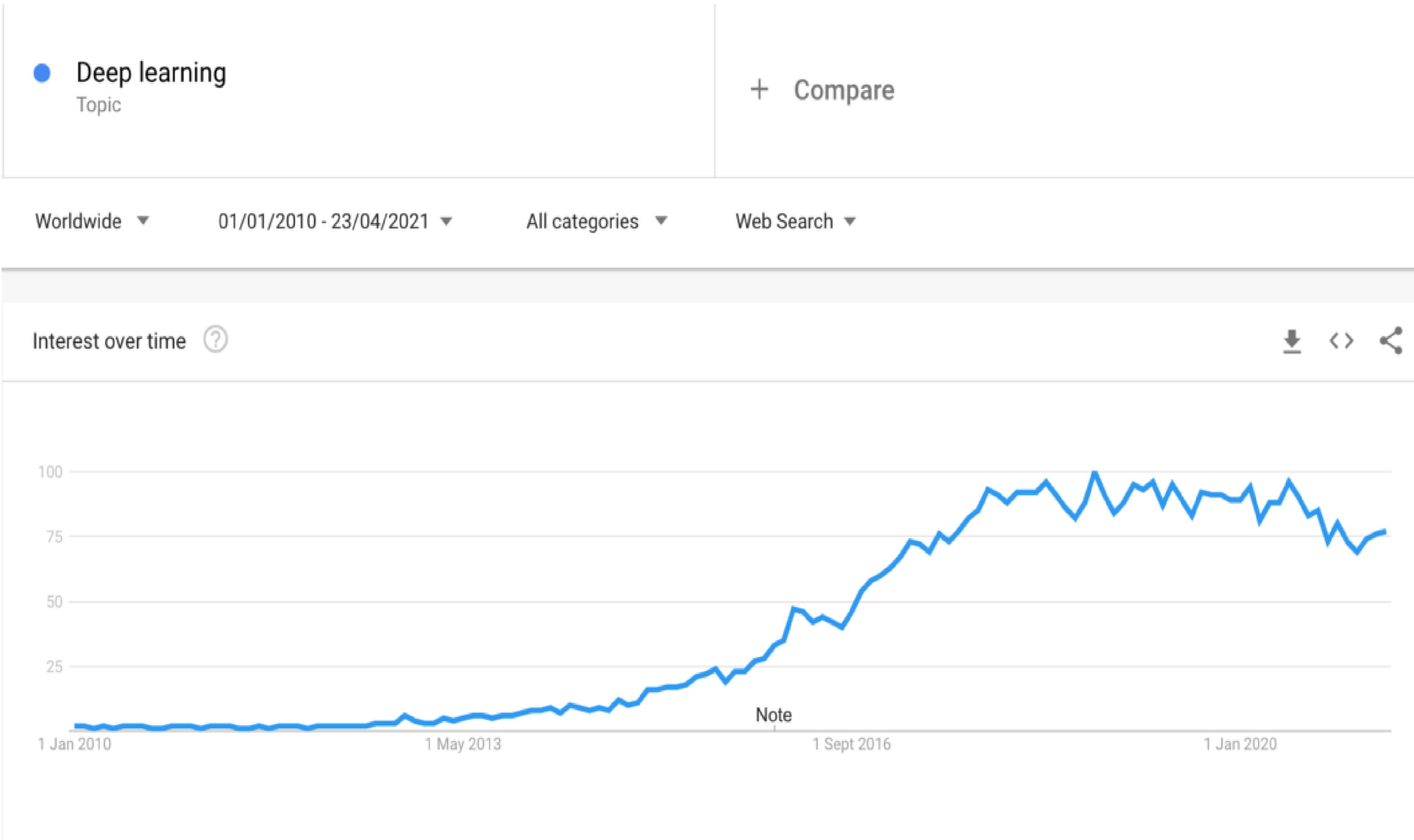
Evolution of neurocomputing

- What is Artificial Intelligence (AI), Machine Learning (ML), and Neural Networks (NN).
- What is Deep Neural Networks (DNN)?
- Why DNN ?

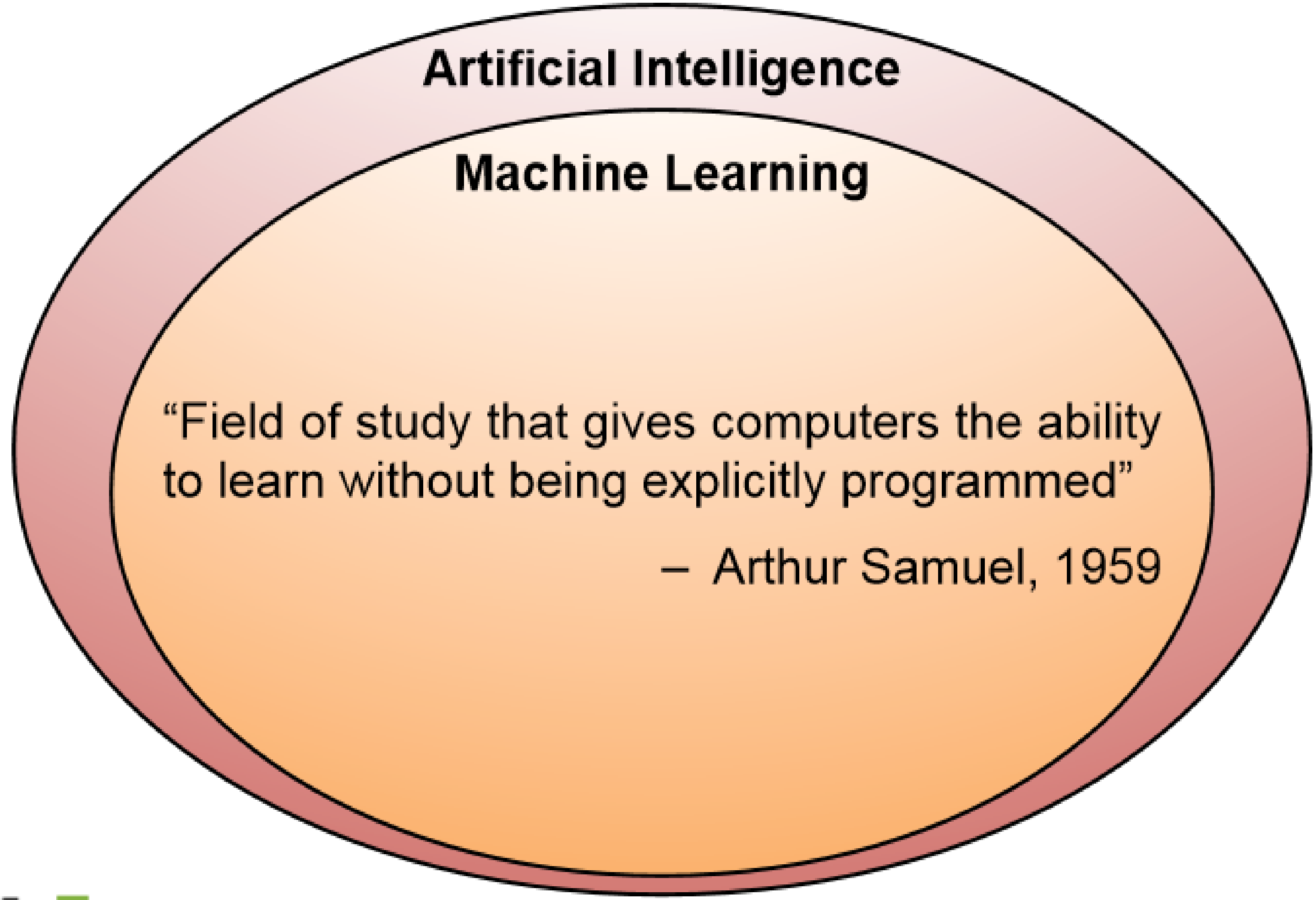
Md Zahangir Alom
Department of Computer Science
University of Memphis, TN

Deep learning: got lots of attention.

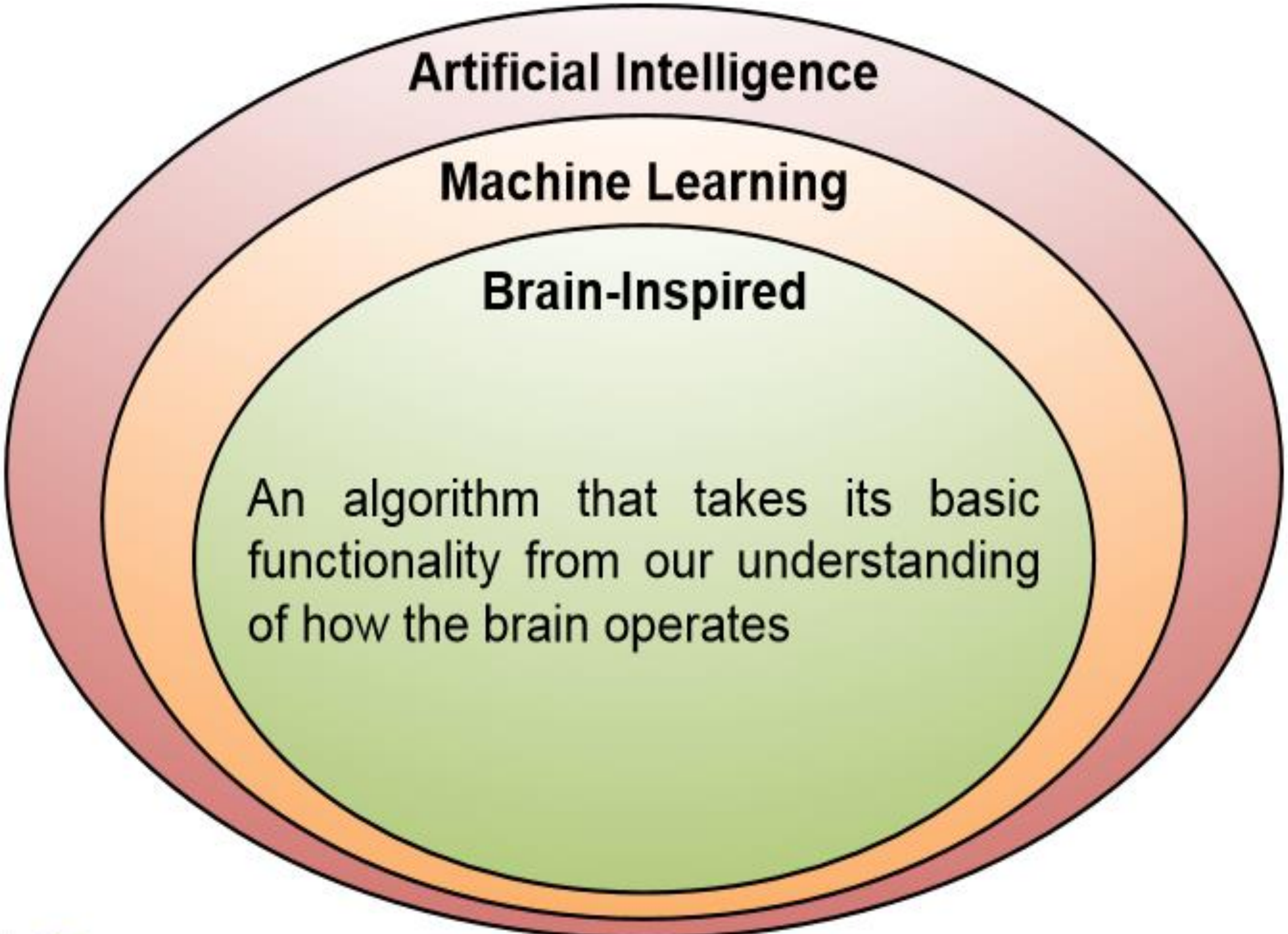
- Google Trends :



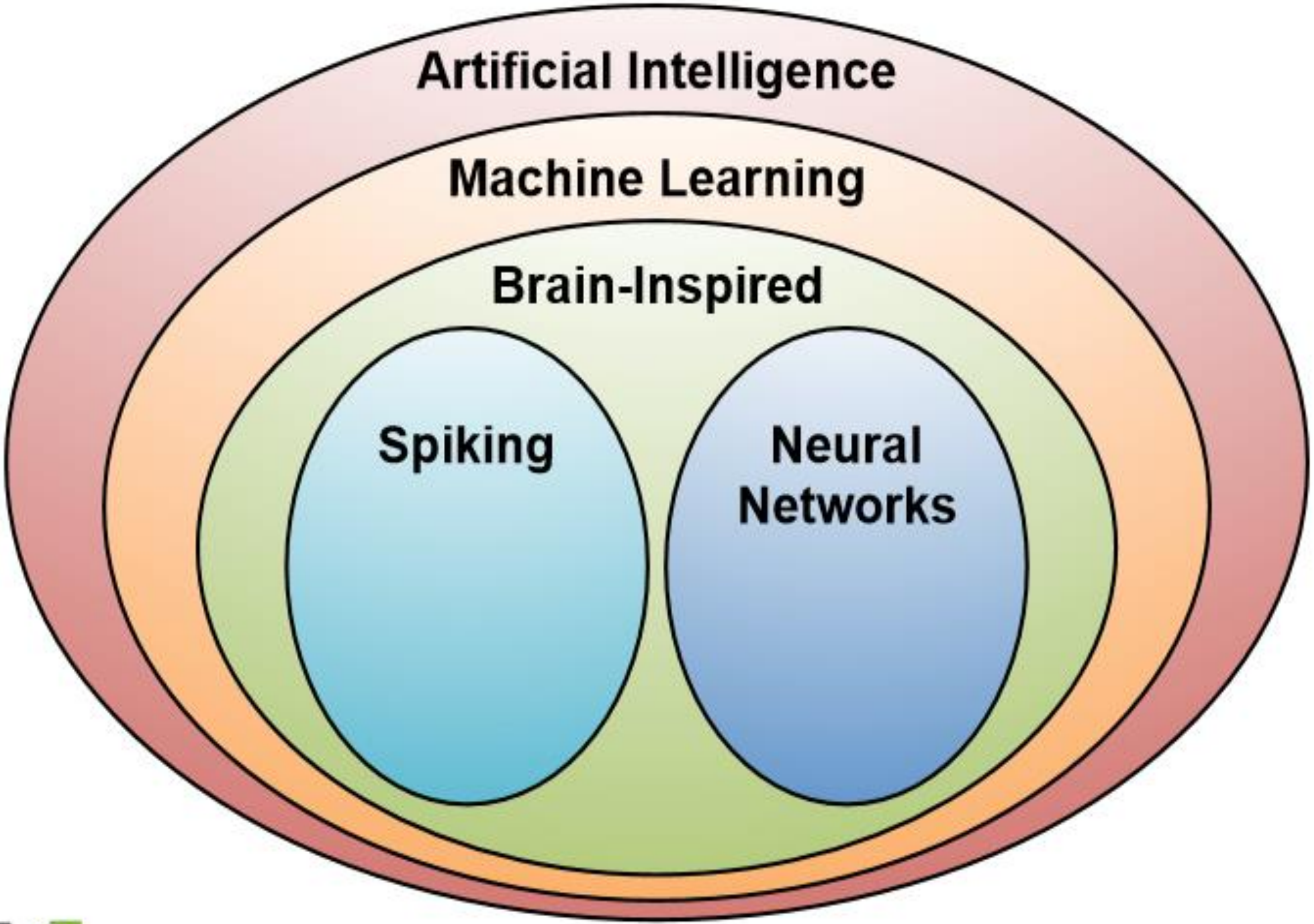
Artificial Intelligence (AI) and Machine Learning (ML)



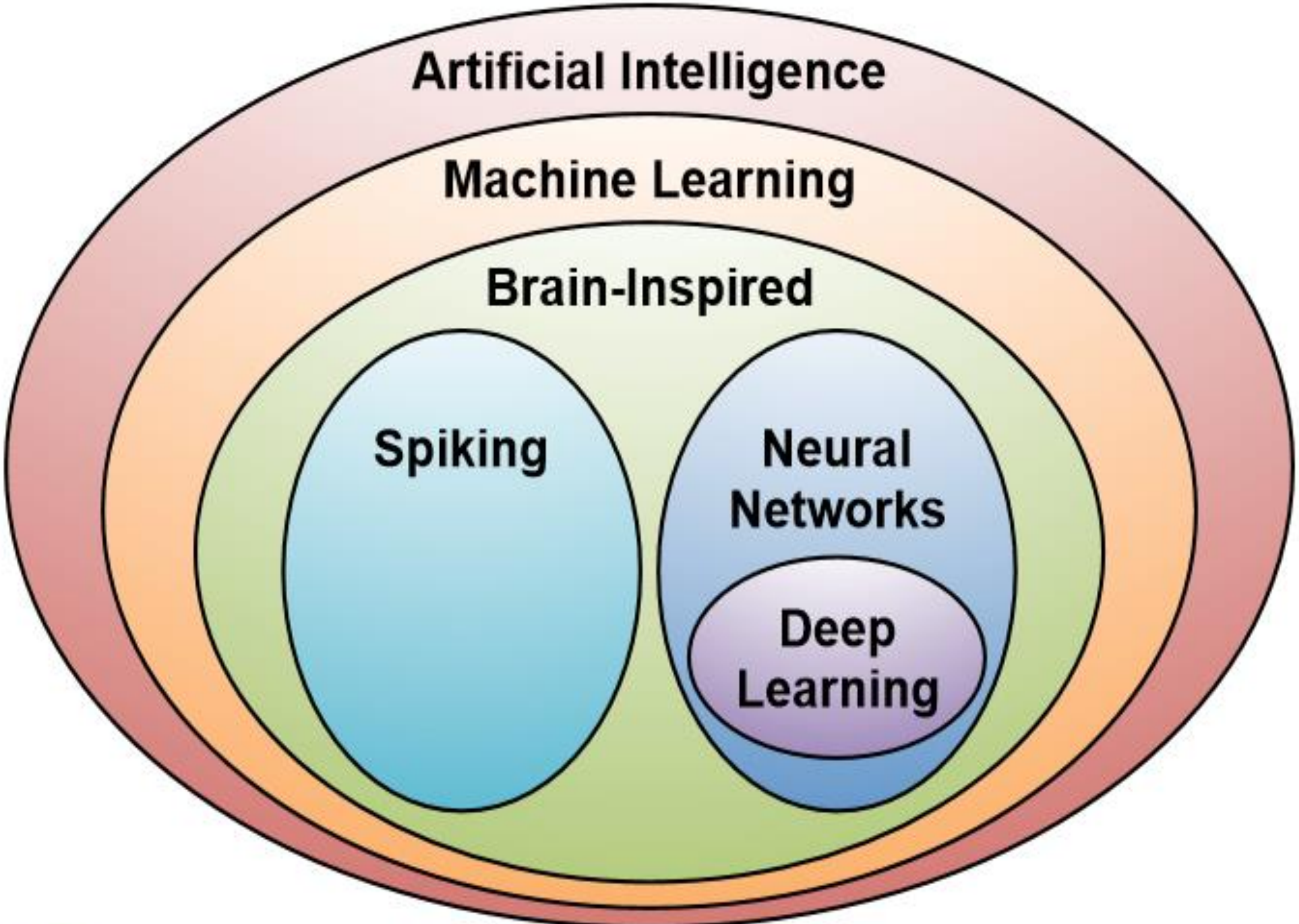
AI,ML, and Brain-inspired computing



AI, ML and Neural Networks

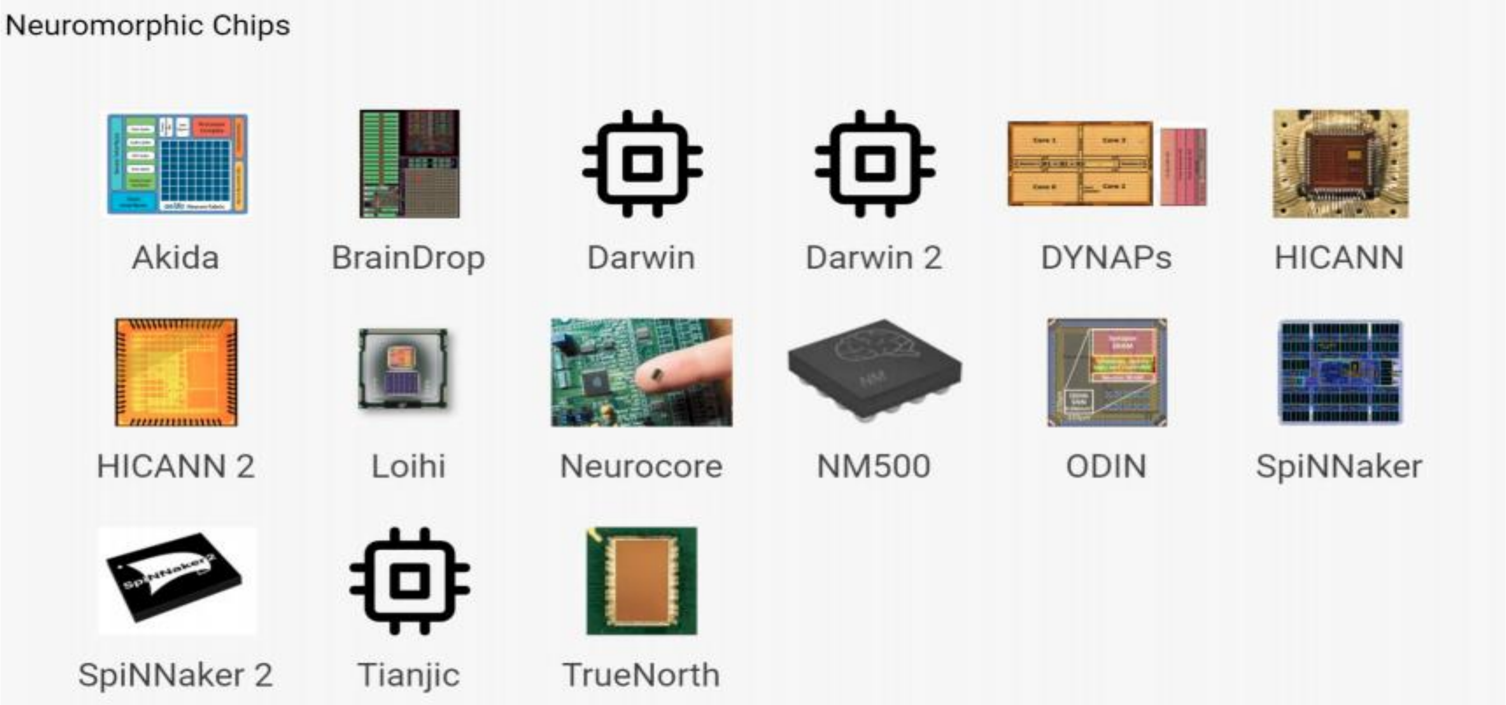


Deep Learning



Neuromorphic computing

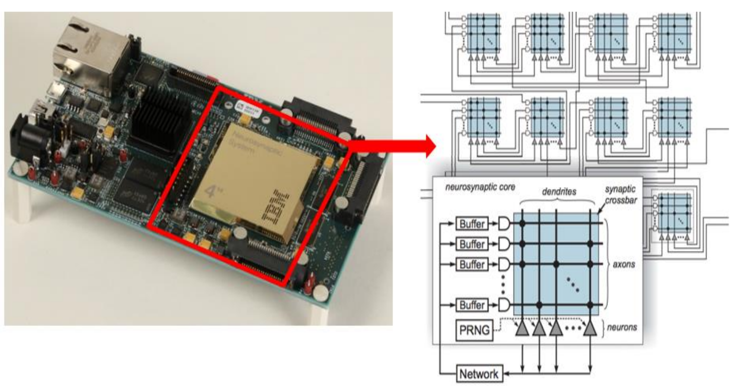
- Neuromorphic computing, which is concerned with **emulating the neural structure and operation of the human brain**, as well as probabilistic computing for **dealing with the uncertainty, ambiguity, and contradiction** in the natural world.



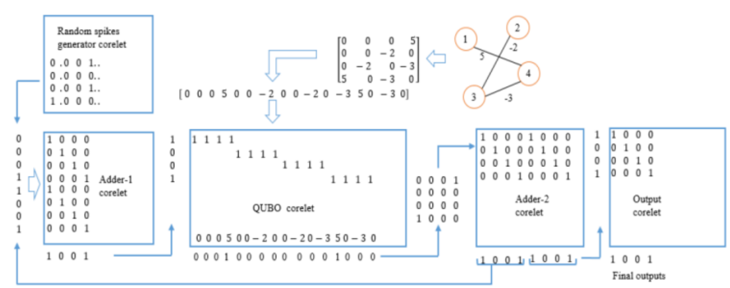
https://www.intel.com/content/www/us/en/research/neuromorphic-computing.html

Neuro-Synaptic Cognitive System

- **IBM TrueNorth** Neuro-synaptic cognitive system, non-von Neumann architecture.
- **4096 cores** per chip, each core consists of
  - 256 input axons and 256 output neurons connected with a 256x256 crossbar of configurable synapses
  - A chip contains
    - 1 million programmable neurons and
    - 256 million synapses.
- **Corelet programming language** to make software that operates like the human brain



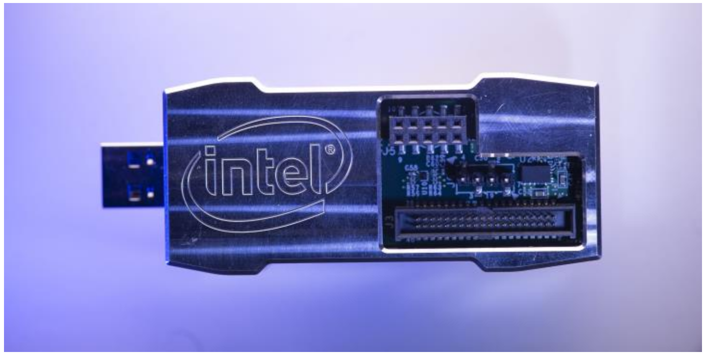
IBM TrueNorth Neuro-synaptic cognitive system



Alom, Md Zahangir, Brian Van Essen, Adam T. Moody, David Peter Widemann, and Tarek M. Taha. "Quadratic Unconstrained Binary Optimization (QUBO) on neuromorphic computing system." In *Neural Networks (IJCNN), 2017 International Joint Conference on*, pp. 3922-3929. IEEE, 2017.

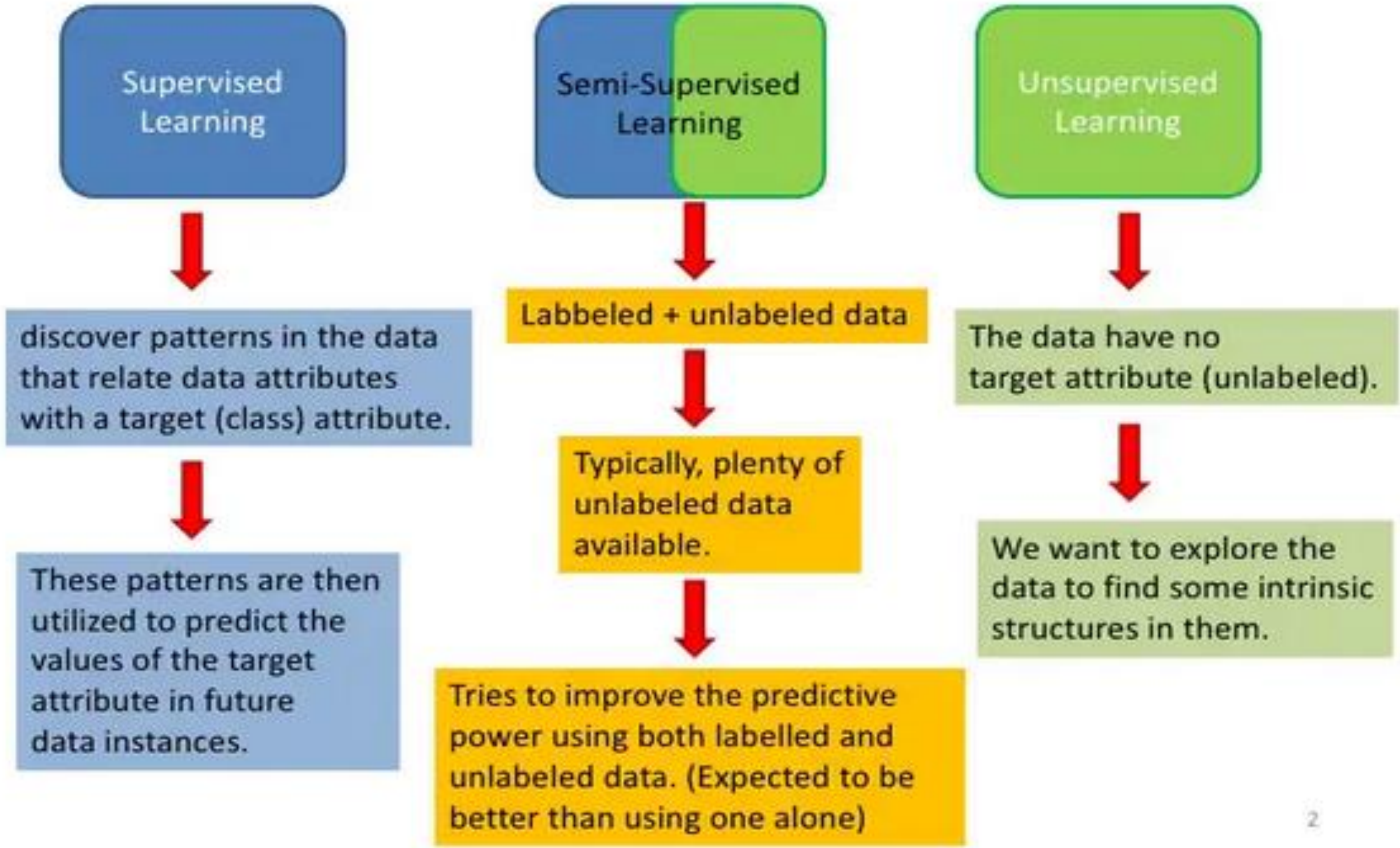
Intel : Loihi

- Loihi (pronounced low-ee-hee) is a neuromorphic **research test chip designed by Intel Labs** that uses an asynchronous spiking neural network (SNN)
- Implemented with **adaptive self-modifying event-driven fine-grained parallel system** used to implement for
  - learning and
  - inference with high efficiency.



Intel's Neuromorphic System Hits 8 Million Neurons, 100 Million released in 2020

Machine Learning Areas



Reinforcement Learning: An agent interacting with the world makes observations, takes actions, and is rewarded or punished; it should learn to choose actions in such a way as to obtain a lot of reward

Machine Learning Areas

Unlabeled data is easy to obtain

Labelled data can be difficult to obtain

- human annotation is boring
- may require experts
- may require special equipment
- very time-consuming

Examples:

- Web page classification (billions of pages)
- Email classification (SPAM or No-SPAM)
- Speech annotation (400h for each hour of conversation)
- ...

Supervised Learning : Important Concepts

- Data: labeled instances <x_i, y>, e.g. emails marked spam/not spam
  - Training Set
  - Held-out Set
  - Test Set
- Features: attribute-value pairs which characterize each x
- Experimentation cycle
  - Learn parameters (e.g. model probabilities) on training set
  - (Tune hyper-parameters on held-out set)
  - Compute accuracy of test set
  - Very important: never "peek" at the test set!
- Evaluation
  - Accuracy: fraction of instances predicted correctly
- Overfitting and generalization
  - Want a classifier which does well on test data
  - Overfitting: fitting the training data very closely, but not generalizing well

Example: Digit Recognition

Input: images / pixel grids

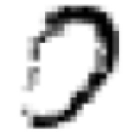
Output: a digit 0-9

Setup:

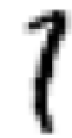
- Get a large collection of example images, each labeled with a digit
- Note: someone has to hand label all this data!
- Want to learn to predict labels of new, future digit images

Features: The attributes used to make the digit decision

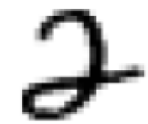
- Pixels: (6,8)=ON
- Shape Patterns: NumComponents, AspectRatio, NumLoops
- …



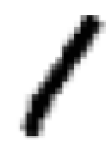
0



1



2



1

??

Slide from Mackassy

Example: Spam Filter

Input: email

Output: spam/ham

Setup:

- Get a large collection of example emails, each labeled "spam" or "ham"
- Note: someone has to hand label all this data!
- Want to learn to predict labels of new, future emails

Features: The attributes used to make the ham / spam decision

- Words: FREE!
- Text Patterns: $dd, CAPS
- Non-text: SenderInContacts
- …

> Dear Sir.
>
> First, I must solicit your confidence in this transaction, this is by virture of its nature as being utterly confidencial and top secret. …

> TO BE REMOVED FROM FUTURE MAILINGS, SIMPLY REPLY TO THIS MESSAGE AND PUT "REMOVE" IN THE SUBJECT.
>
> 99 MILLION EMAIL ADDRESSES FOR ONLY $99

> Ok, Iknow this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power nothing happened.

Slide from Mackassy

How and what does machine learn?

Input: X

Output: Y

Label"motorcycle"

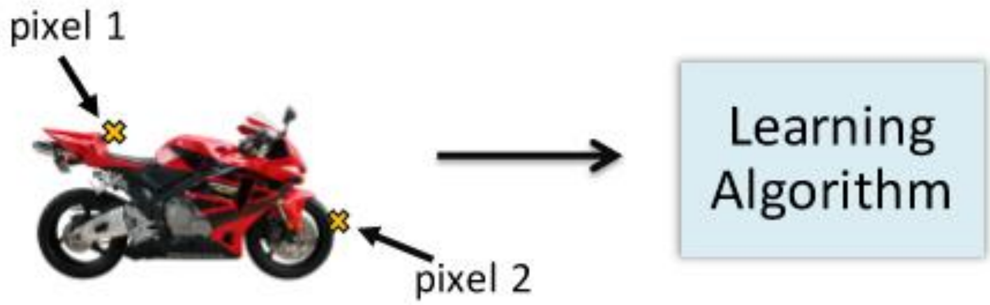
Why is it hard?

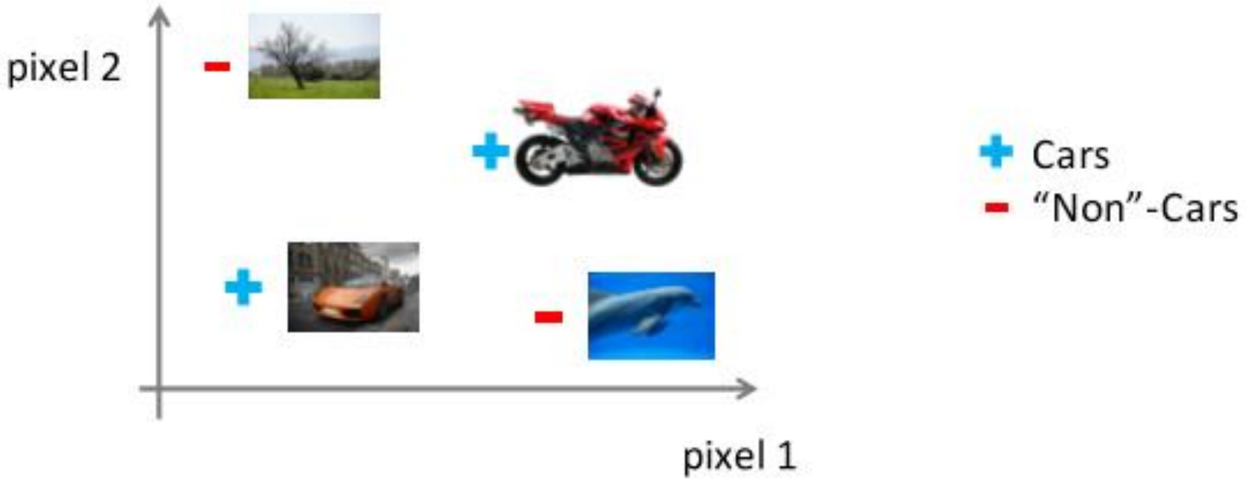
You see this

But the camera sees this:

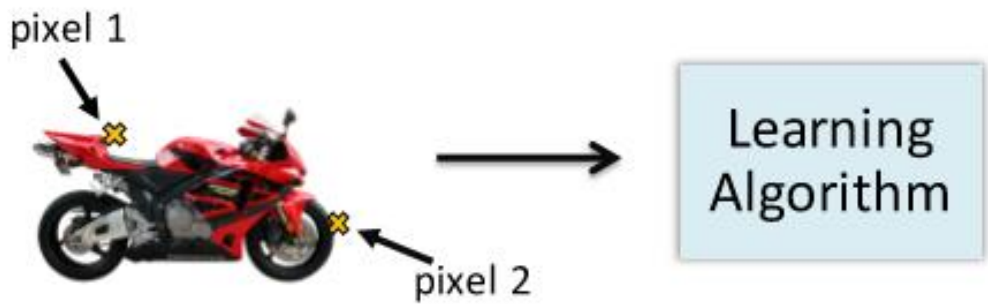
194	210	201	212	199	213	215	195	178	158	182	209
180	189	190	221	209	205	191	167	147	115	129	163
114	126	140	188	176	165	152	140	170	106	78	88
87	103	115	154	143	142	149	153	173	101	57	57
102	112	106	131	122	138	152	147	128	84	58	66
94	95	79	104	105	124	129	113	107	87	69	67
68	71	69	98	89	92	98	95	89	88	76	67
41	56	68	99	63	45	60	82	58	76	75	65
20	43	69	75	56	41	51	73	55	70	63	44
50	50	57	69	75	75	73	74	53	68	59	37
72	59	53	66	84	92	84	74	57	72	63	42
67	61	58	65	75	78	76	73	59	75	69	50

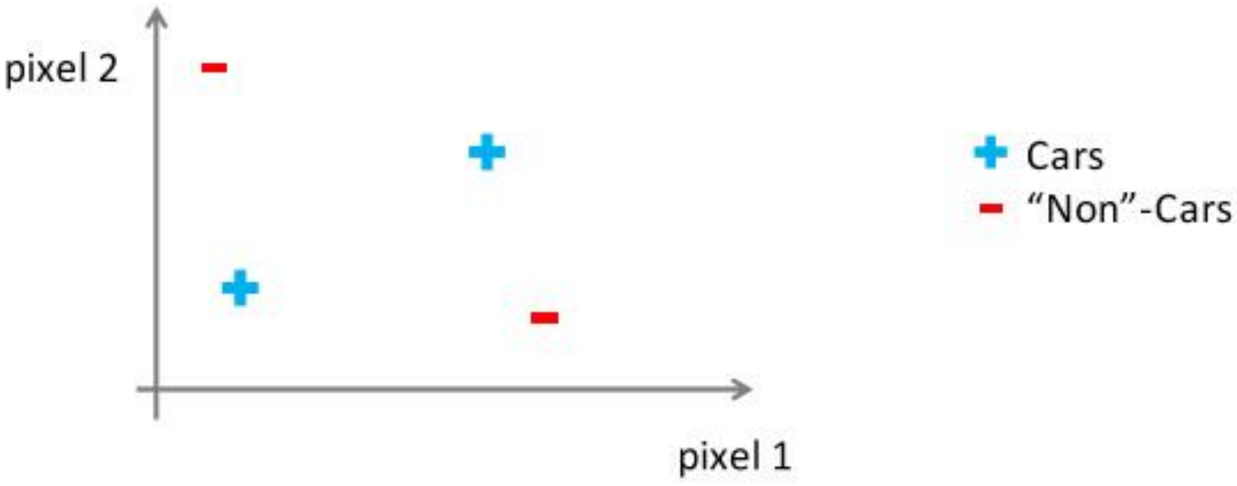
Raw Image Representation



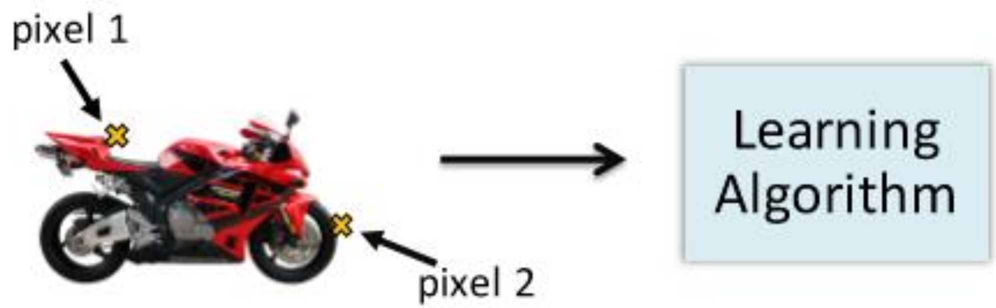


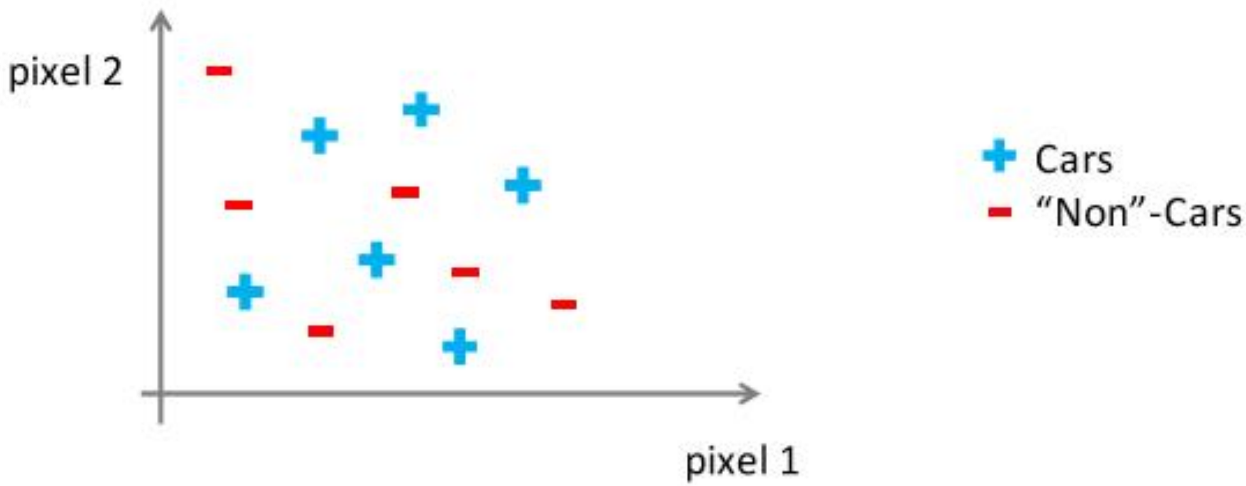
Raw Image Representation



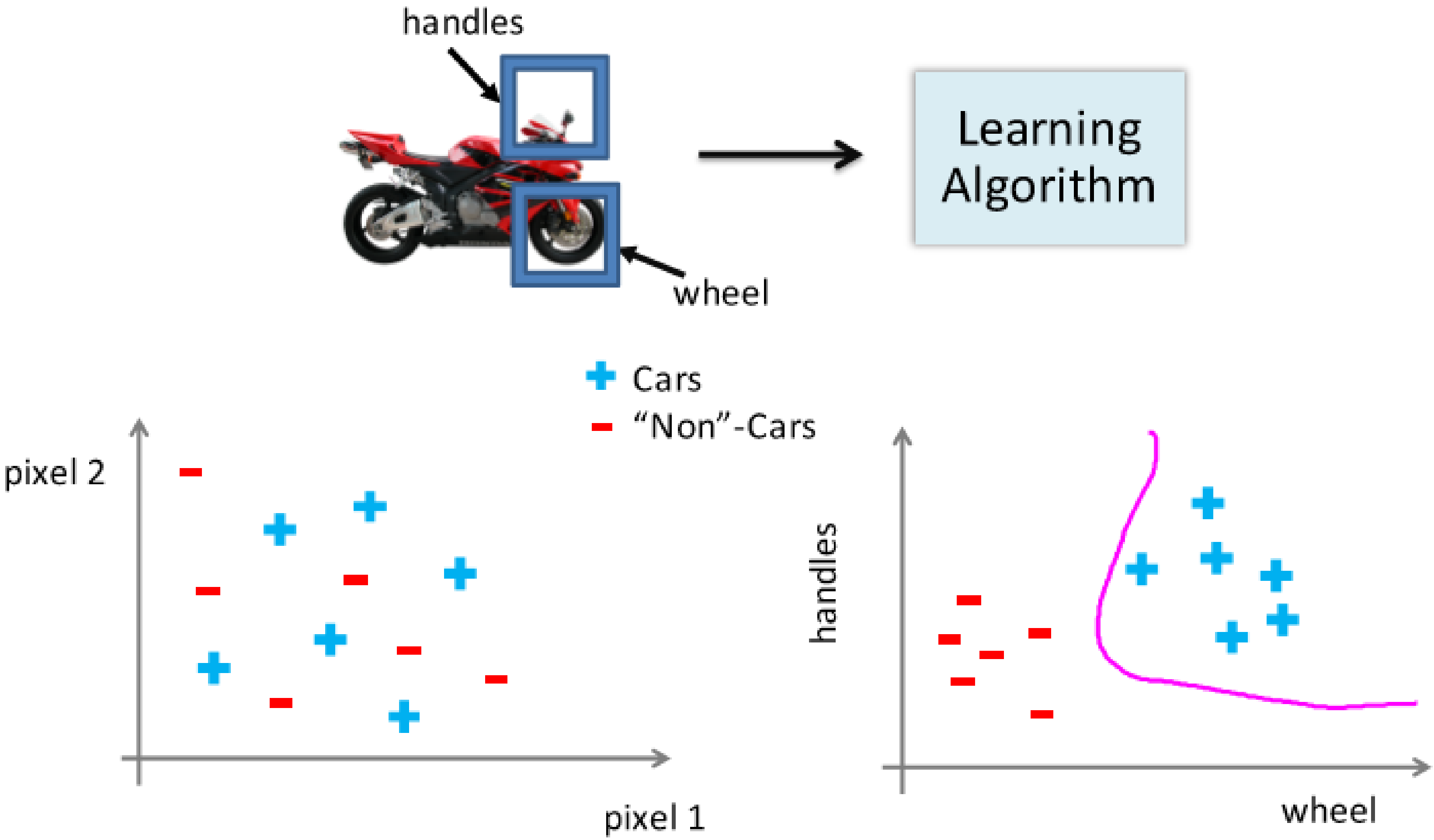


Raw image representation





Better feature representation



Traditional Machine Learning

VISION

hand-crafted features
SIFT/HOG

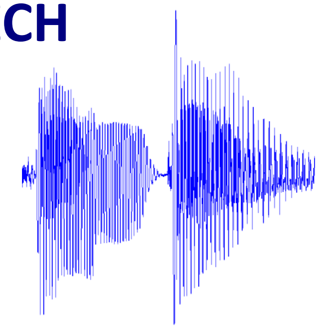
fixed

your favorite classifier

learned

"car"

SPEECH



hand-crafted features
MFCC

fixed

your favorite classifier

learned

\ˈd ē p\

NLP

This burrito place
is yummy and fun!

hand-crafted features
Bag-of-words

fixed

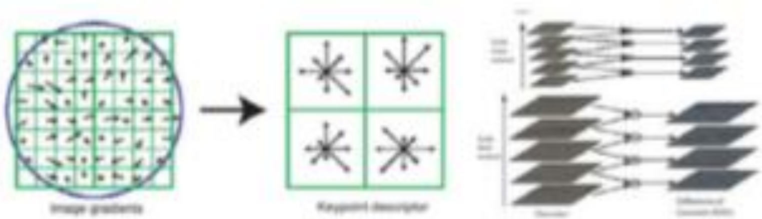
your favorite classifier

learned

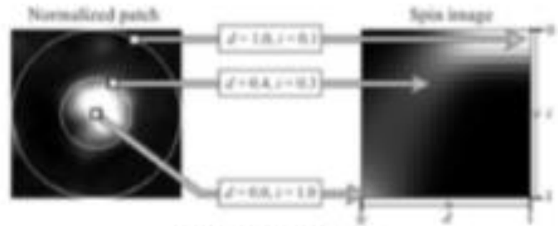
"+"

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

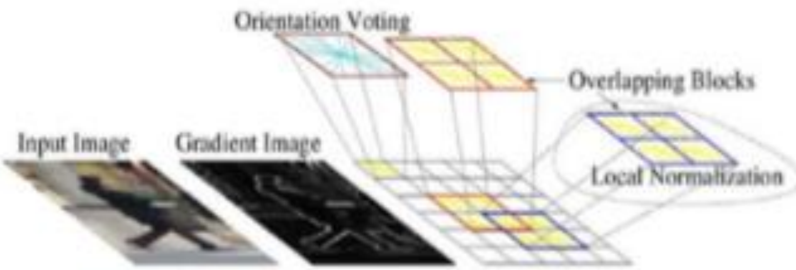
Feature representation methods



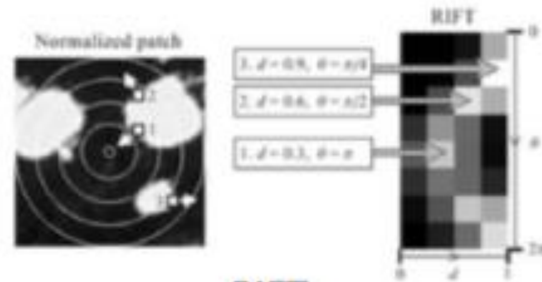
SIFT



Spin image

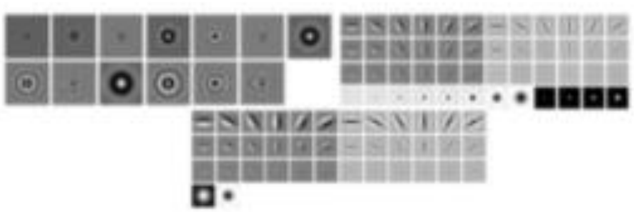


HoG

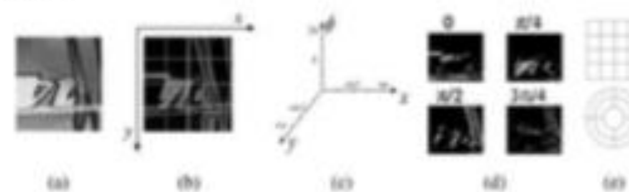


RIFT

Expert Knowledge!



Textons



GLOH

Source: feature representations in computer vision(Honglak lee)

Traditional Machine Learning (more accurately)

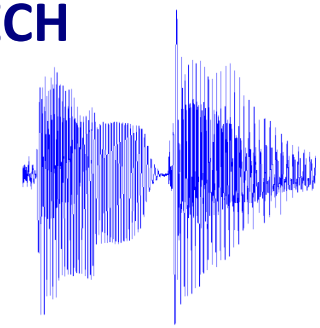
"Learned"

VISION

SIFT/HOG → K-Means/ pooling → classifier → "car"

fixed | unsupervised | supervised

SPEECH



MFCC → Mixture of Gaussians → classifier → \ˈd ē p\

fixed | unsupervised | supervised

NLP

This burrito place
is yummy and fun!

Parse Tree Syntactic → n-grams → classifier → "+"

fixed | unsupervised | supervised

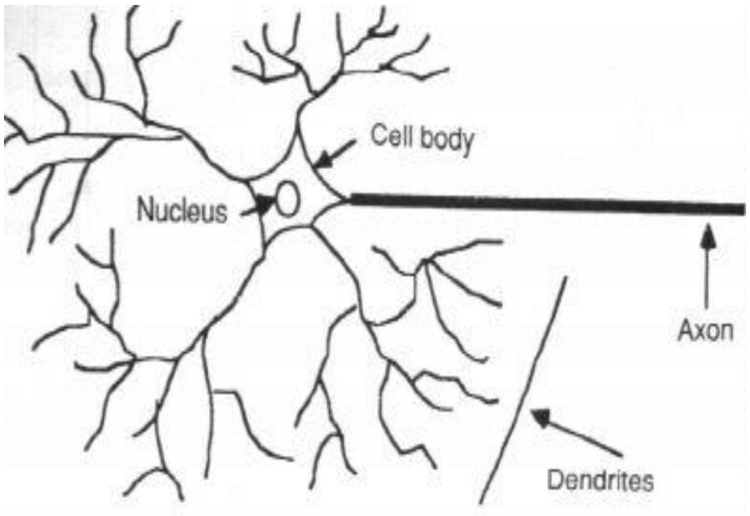
Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

[Mel Frequency Cepstral Coefficient (MFCC)]

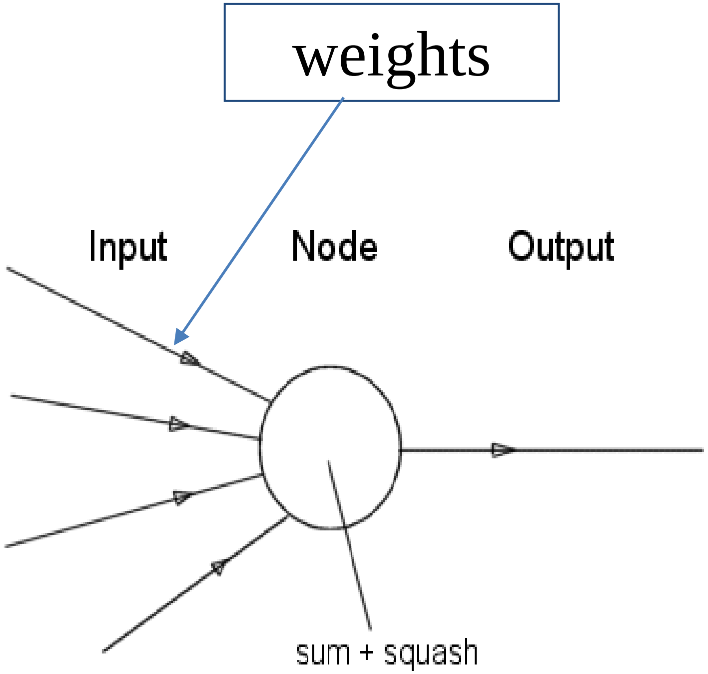
Artificial Neural Networks (ANNs) – The basics

ANNs incorporate the two fundamental components of biological neural nets:

1. Neurones (nodes)
2. Synapses (weights)



Biological neurons



Artificial neurons

What exactly is deep neural networks?

The short answers

A Neural Network with **Several Layers of nodes(neurons)** between inputs and outputs.

Why it is better than other methods?

The series of layers between input & output do **feature identification, representation and processing** in a series of stages, just as our brains seem to.

Properties of Deep (Machine) Learning approaches

- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction

- Distributed Representations , Scalability, and Genericity
  - No single neuron “encodes” everything
  - Groups of neurons work together

Hierarchical Compositionality

VISION

pixels → edge → texton → motif → part → object

SPEECH

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

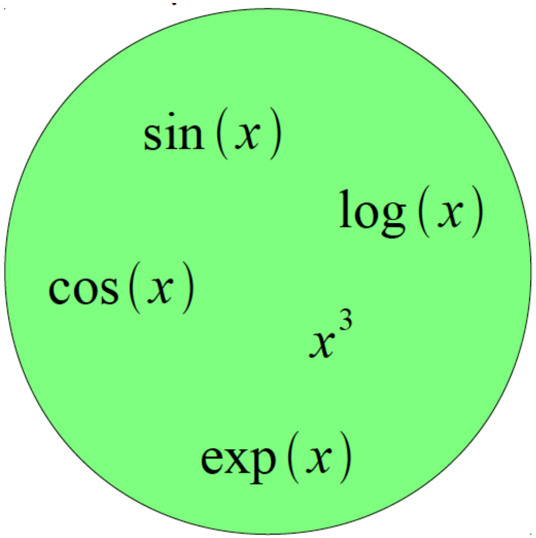
NLP

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

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Building A Complicated Function

Given a library of simple functions

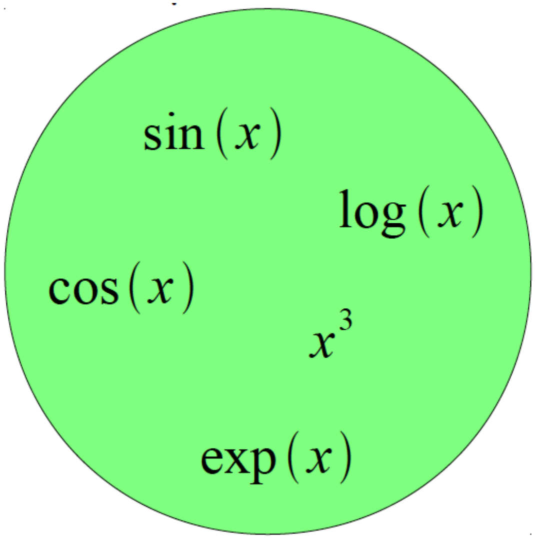


Compose into a

complicate function

Building A Complicated Function

Given a library of simple functions

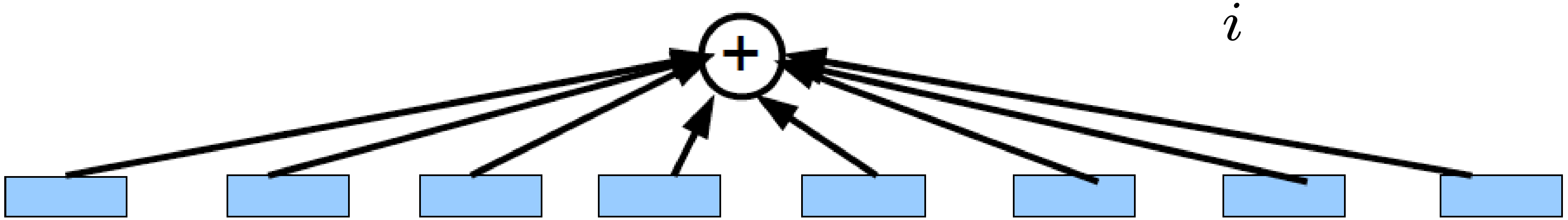


Compose into a complicate function

Idea 1: Linear Combinations

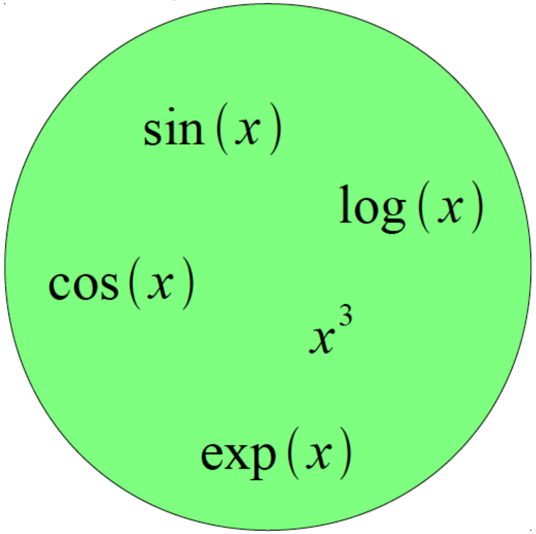
- Boosting
- Kernels
- …

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



Building A Complicated Function

Given a library of simple functions

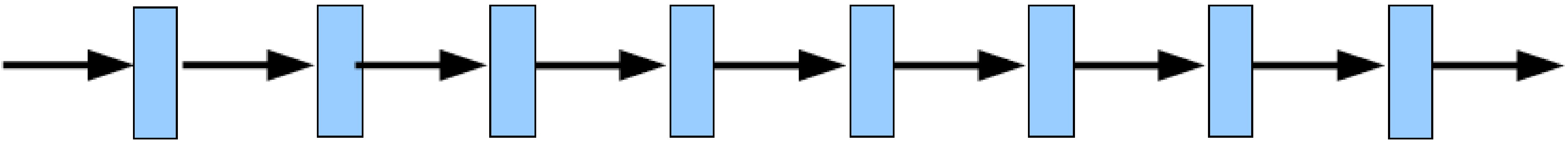


Compose into a complicate function

Idea 2: Compositions

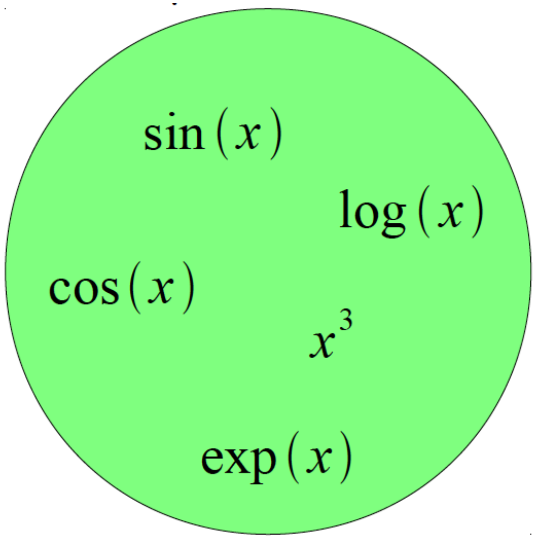
- **Deep Learning**
- Grammar models
- Scattering transforms…

$$f(x) = g_1(g_2(\ldots(g_n(x)\ldots))$$



Building A Complicated Function

Given a library of simple functions

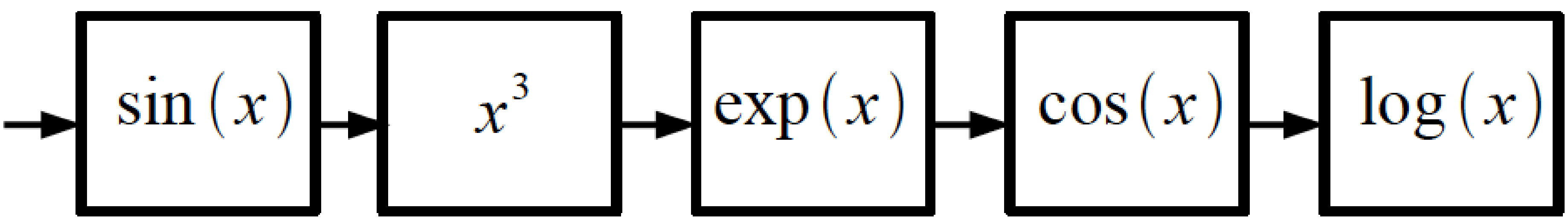


Compose into a complicate function

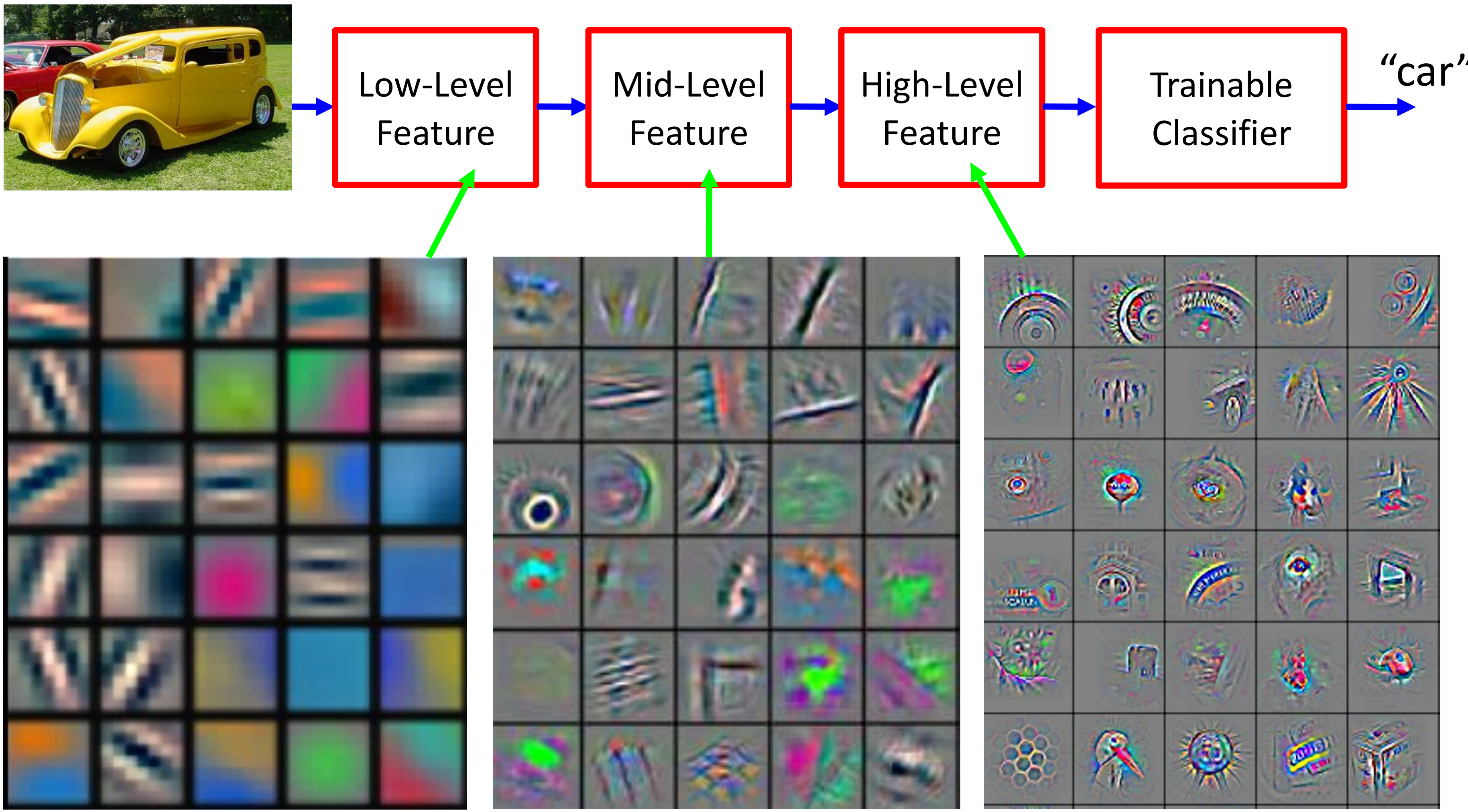
Idea 2: Compositions

- **Deep Learning**
- Grammar models
- Scattering transforms…

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



Deep Learning = Hierarchical Compositionality



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Slide Credit: Marc Aurelio Ranzato, Yann LeCun

Properties of Deep (Machine) Learning approaches

- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction

- Distributed Representations , Scalability, and Genericity
  - No single neuron “encodes” everything
  - Groups of neurons work together

Deep Learning = End-to-End Learning

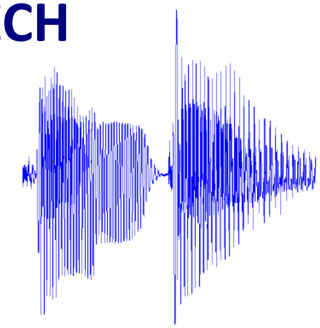
"Learned"

VISION

SIFT/HOG → K-Means/ pooling → classifier → "car"

fixed | unsupervised | supervised

SPEECH



MFCC → Mixture of Gaussians → classifier → \ˈd ē p\

fixed | unsupervised | supervised

NLP

This burrito place is yummy and fun!

Parse Tree Syntactic → n-grams → classifier → "+"

fixed | unsupervised | supervised

"Shallow" vs Deep Learning

- "Shallow" models

hand-crafted

Feature Extractor

fixed

"Simple" Trainable

Classifier

learned

- Deep models

Trainable Feature-Transform / Classifier

Trainable Feature-Transform / Classifier

Trainable Feature-Transform / Classifier

Learned Internal Representations

Properties of Deep (Machine) Learning approaches

- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations
- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction
- Distributed **Representations , Scalability, and Genericity**
  - No single neuron “encodes” everything
  - Groups of neurons work together

One Model To Learn Them All

Kaiser, Lukasz, Aidan N. Gomez, Noam Shazeer, Ashish Vaswani, Niki Parmar, Llion Jones, and Jakob Uszkoreit. "One model to learn them all." *arXiv preprint arXiv:1706.05137*(2017).

Part II:

Why Deep Neural Networks (DNN)?

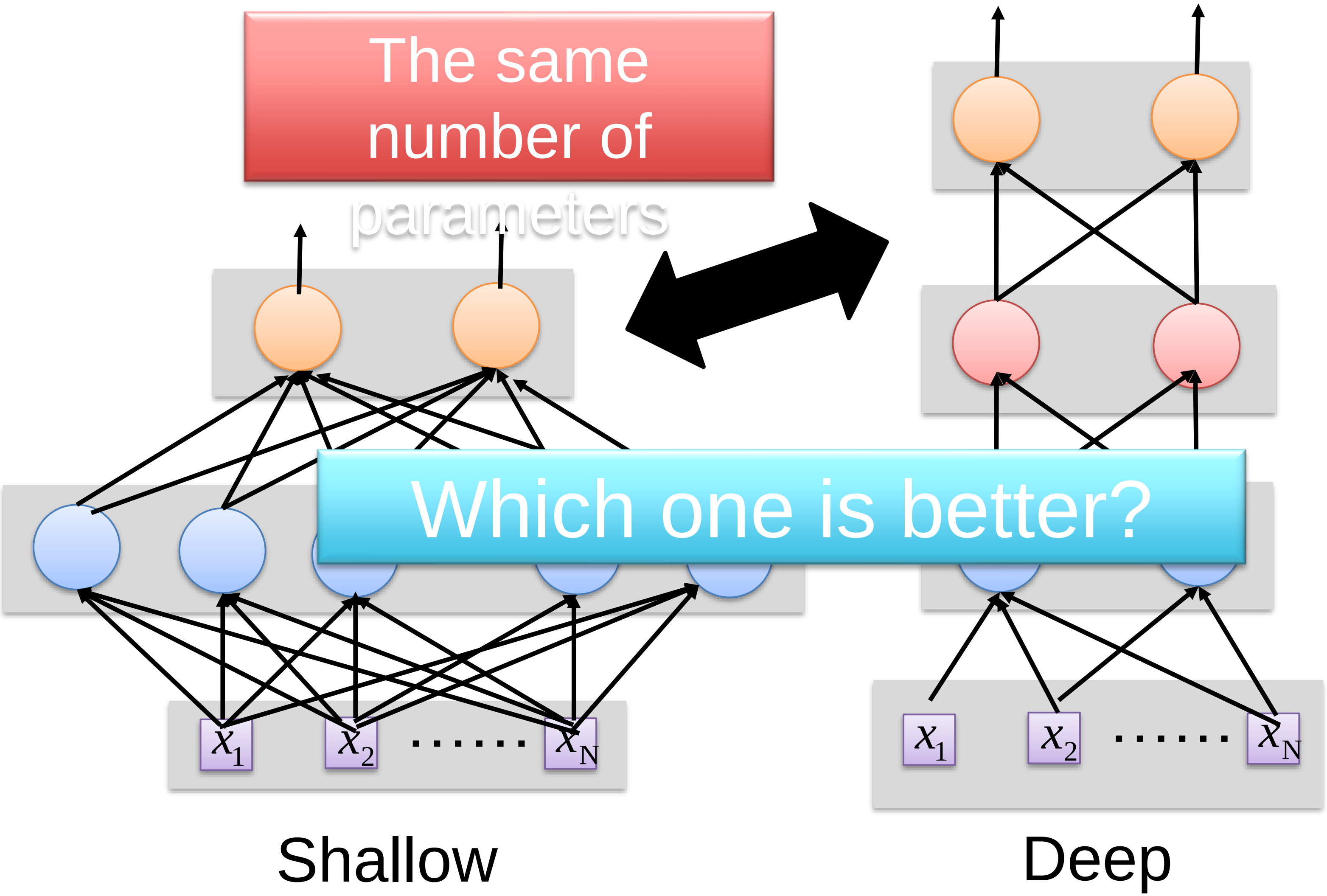
Deeper is Better?

Layer X Size	Word Error Rate (%)
1 X 2k	24.2
2 X 2k	20.4
3 X 2k	18.4
4 X 2k	17.8
5 X 2k	17.2
7 X 2k	17.1

Not surprised, more parameters, better performance

Seide, Frank, Gang Li, and Dong Yu. "Conversational Speech Transcription Using Context-Dependent Deep Neural Networks." *Interspeech*. 2011.

Fat + Short v.s. Thin + Tall

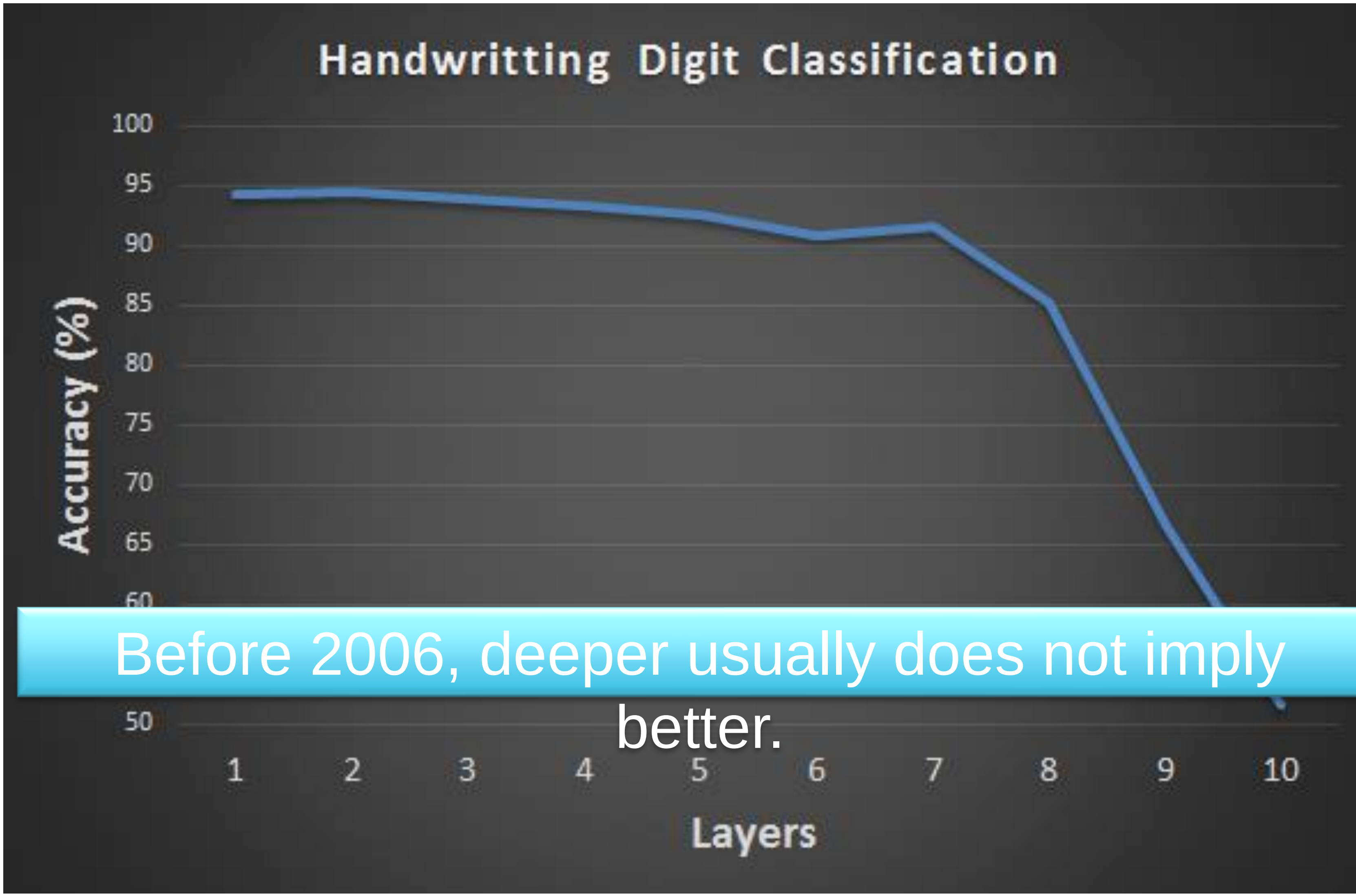


Fat + Short v.s. Thin + Tall

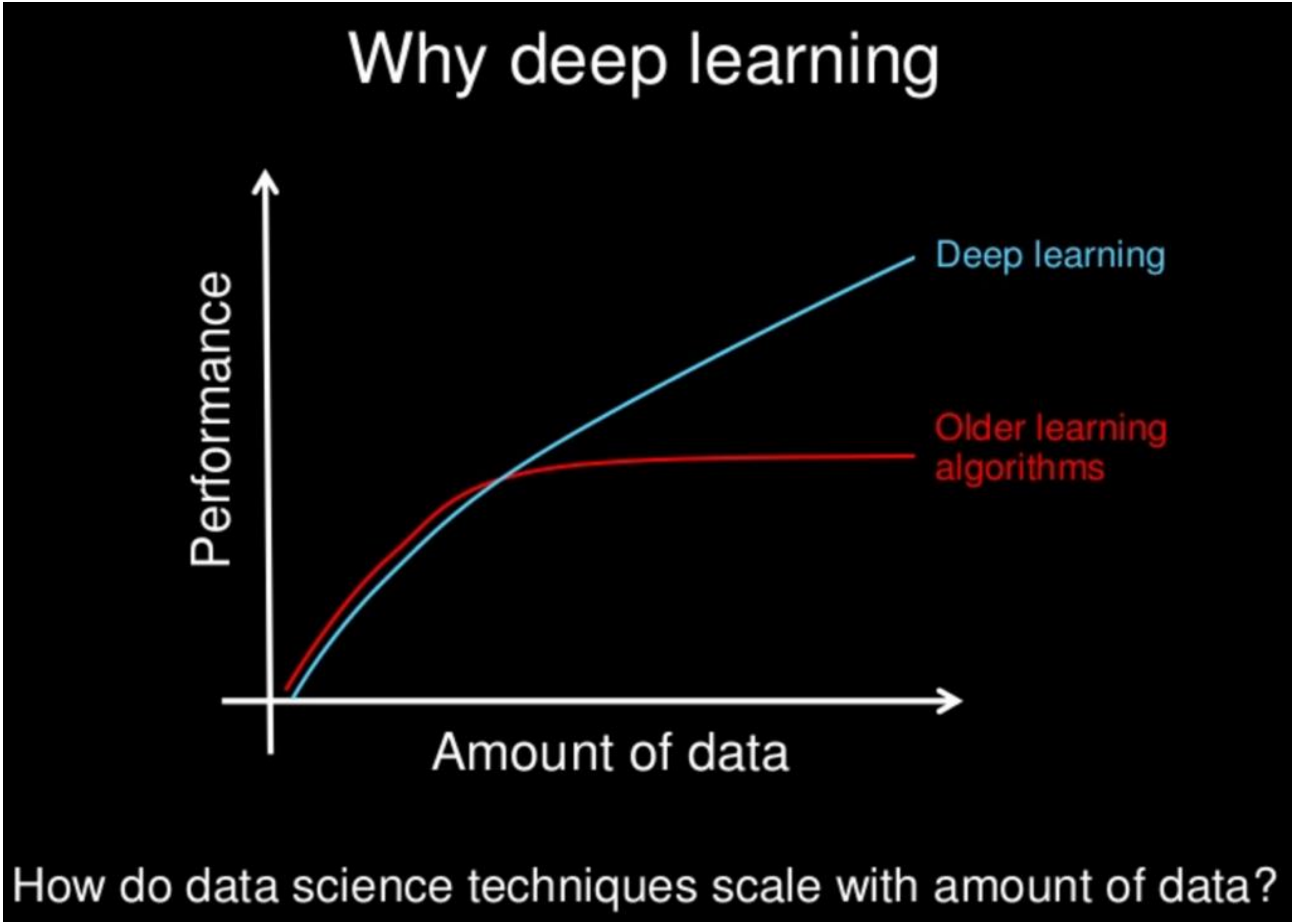
Layer X Size	Word Error Rate (%)	Layer X Size	Word Error Rate (%)
1 X 2k	24.2		
2 X 2k	20.4		
3 X 2k	18.4		
4 X 2k	17.8		
5 X 2k	17.2	1 X 3772	22.5
7 X 2k	17.1	1 X 4634	22.6
		1 X 16k	22.1

Seide, Frank, Gang Li, and Dong Yu. "Conversational Speech Transcription Using Context-Dependent Deep Neural Networks." *Interspeech*. 2011.

Hard to get the power of deep …



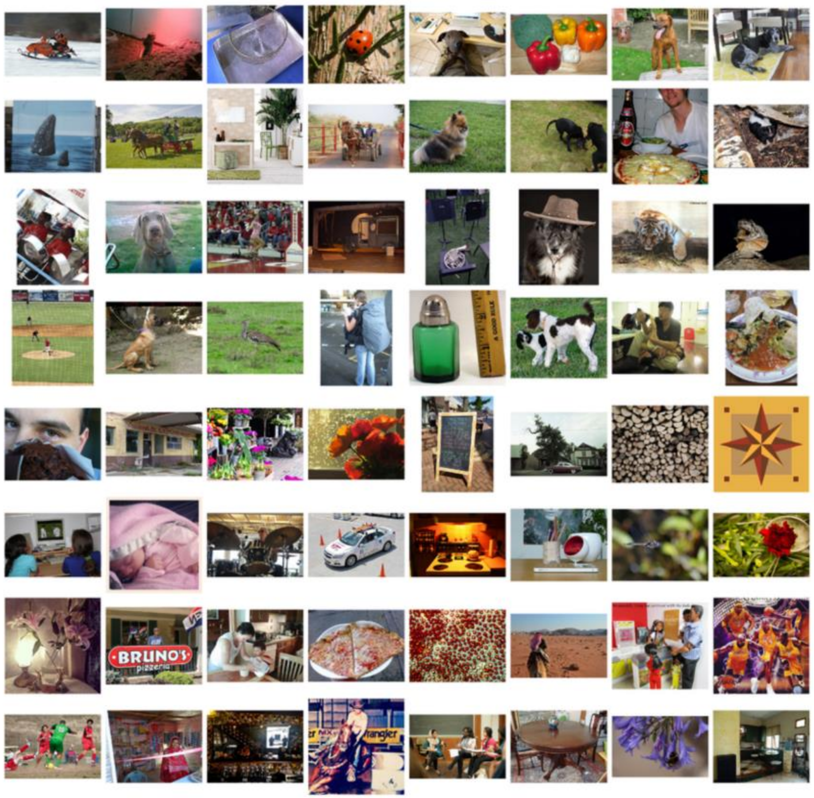
Before 2006, deeper usually does not imply better.


Why deep learning
Deep learning
Older learning
algorithms
Performance
Amount of data
How do data science techniques scale with amount of data?

Applications

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

21,841 object classes **14+ M images**



https://www.image-net.org/update-mar-11-2021.php

Image Classification

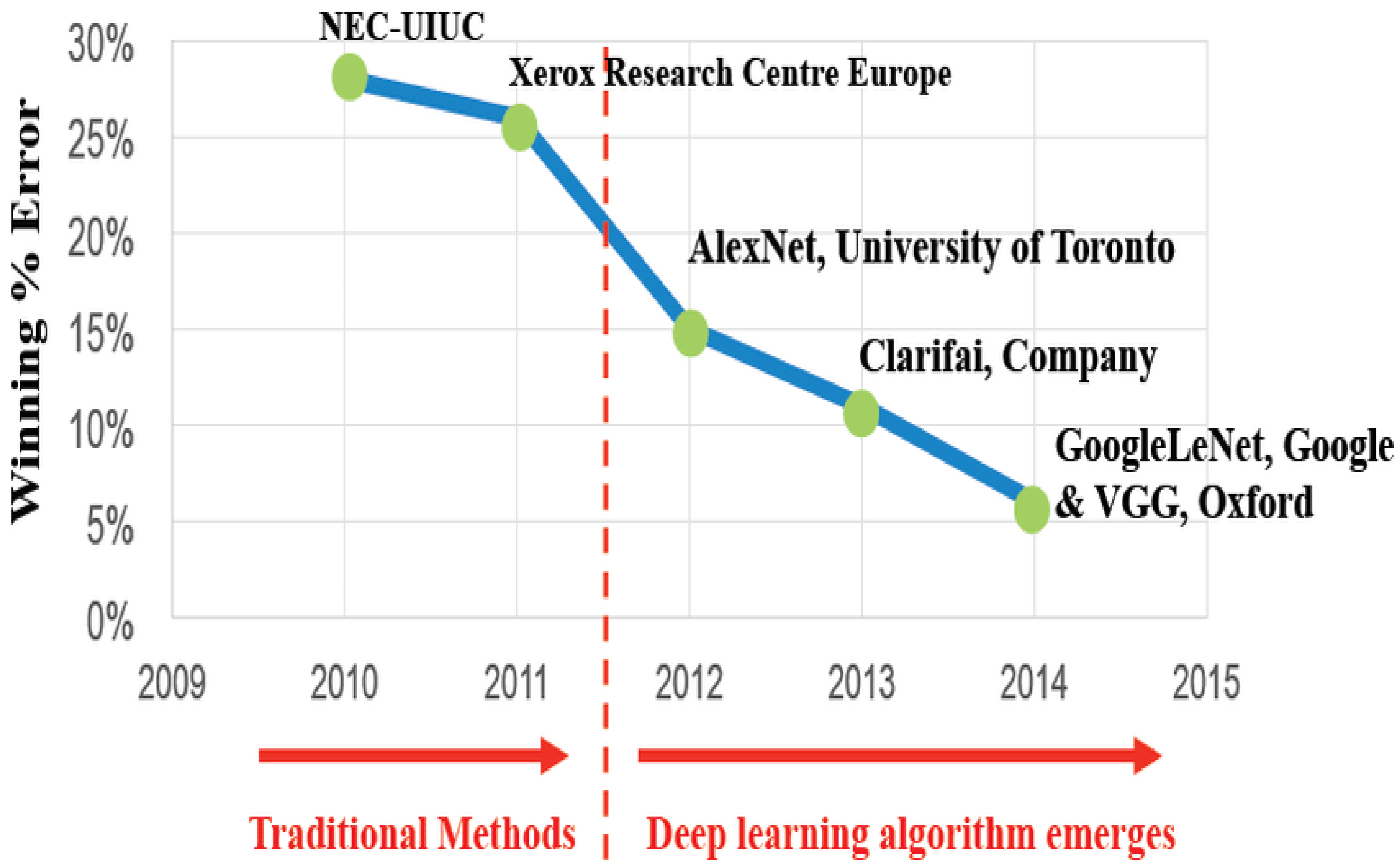
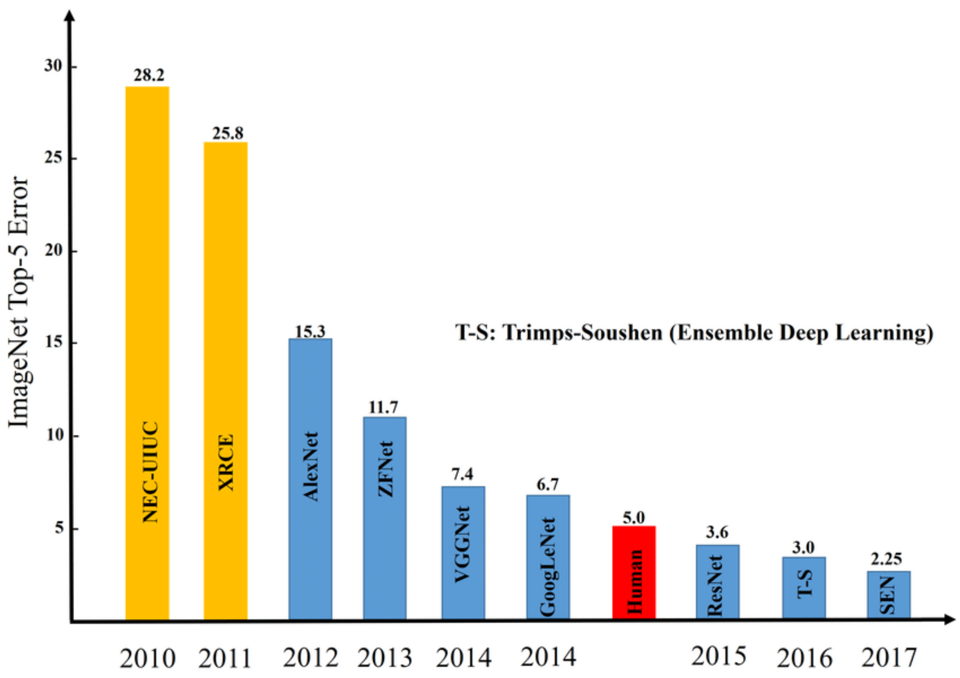
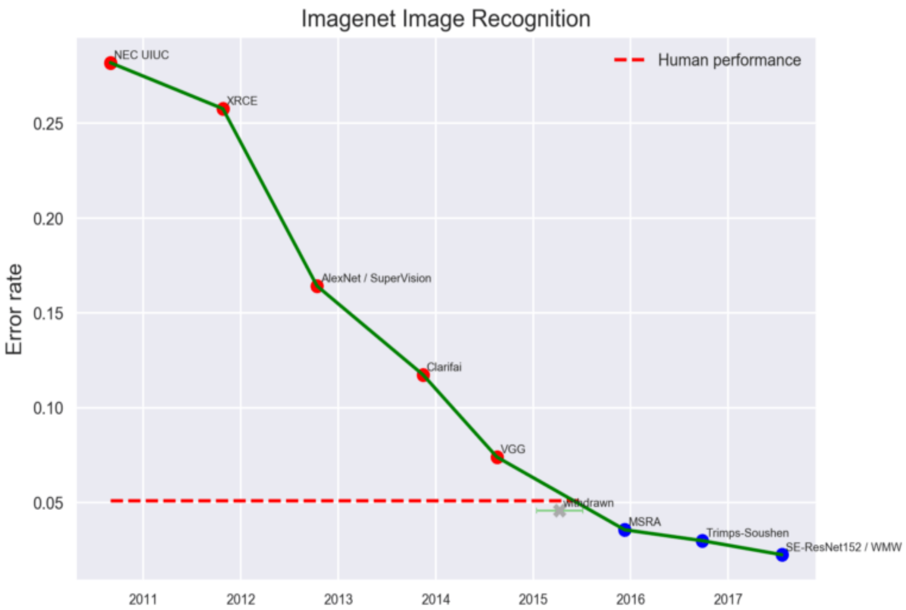


Image classification





https://paperswithcode.com/sota/image-classification-on-imagenet

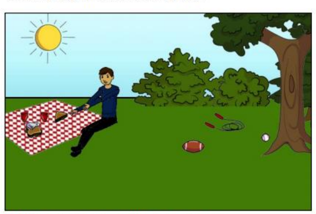
Tasks are getting bolder

Image captioning: A group of young people playing a game of Frisbee

Vinyals et al., 2015

What color are her eyes?
What is the mustache made of?

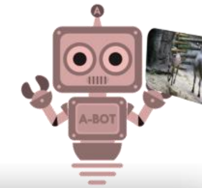
How many slices of pizza are there?
Is this a vegetarian pizza?

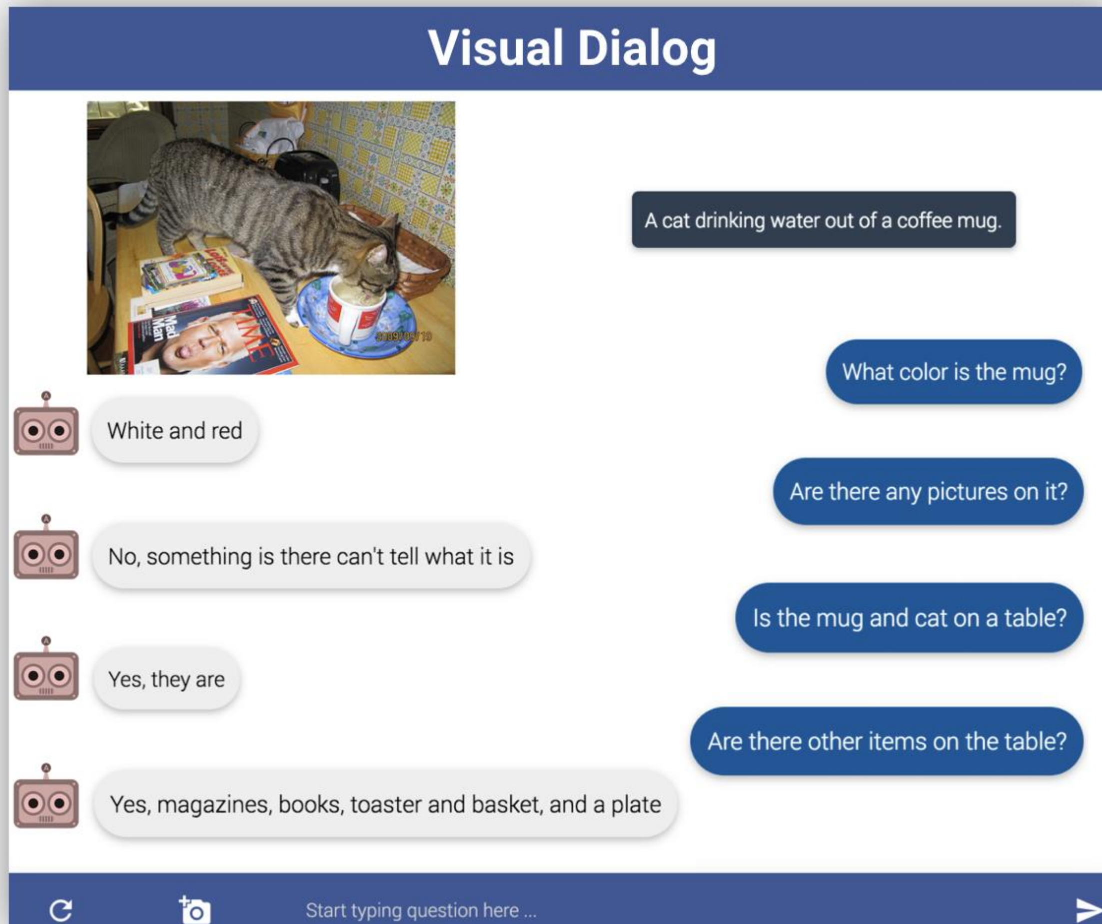


Is this person expecting company?
What is just under the tree?

Does it appear to be rainy?
Does this person have 20/20 vision?

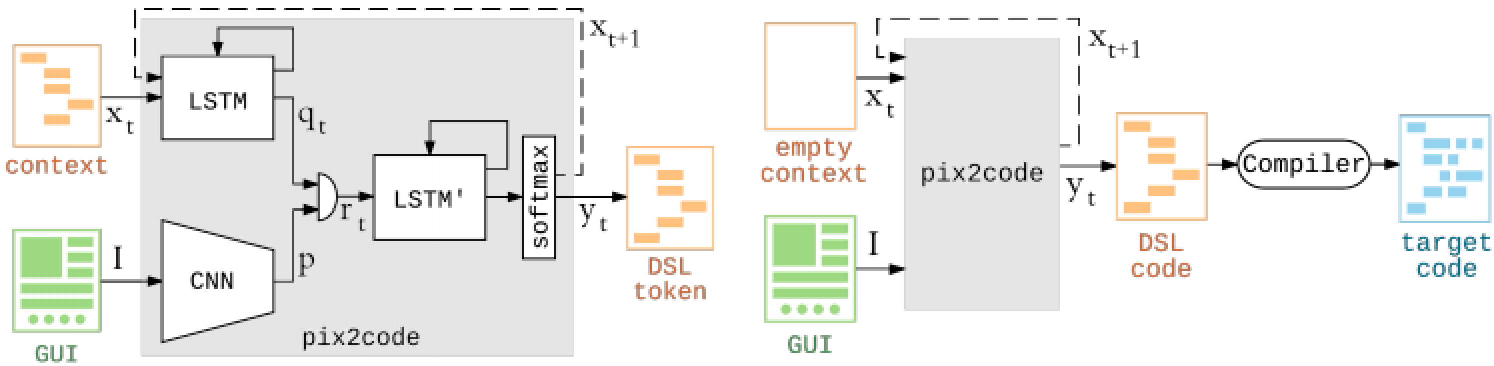
Visual Question Answering (VQA) : Antol et al., 2015





Das et al., 2017

Pix2code



(a) Training (b) Sampling

Overview of the pix2code model architecture

- Transforming a graphical user interface screenshot into computer code
- Over 77% of accuracy for three different platforms (i.e. iOS, Android and web-based technologies)

```
stack {
  row {
    label, switch
  }
  row {
    label, btn-add
  }
  row {
    label, slider, label
  }
  row {
    img, label
  }
}
footer {
  btn-more, btn-contact, btn-search, btn-download
}
```

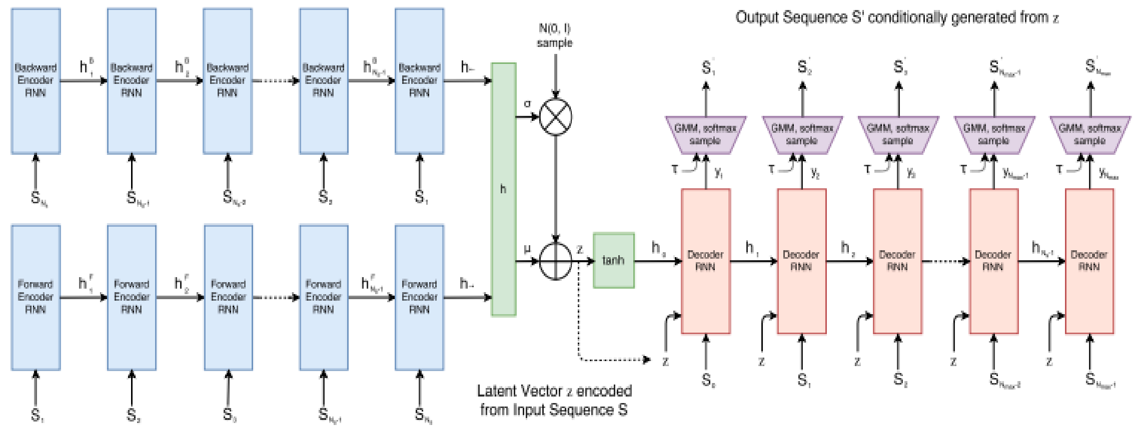
(a) iOS GUI screenshot (b) Code describing the GUI written in our DSL

An example of a native iOS GUI written in our markup-like DSL.

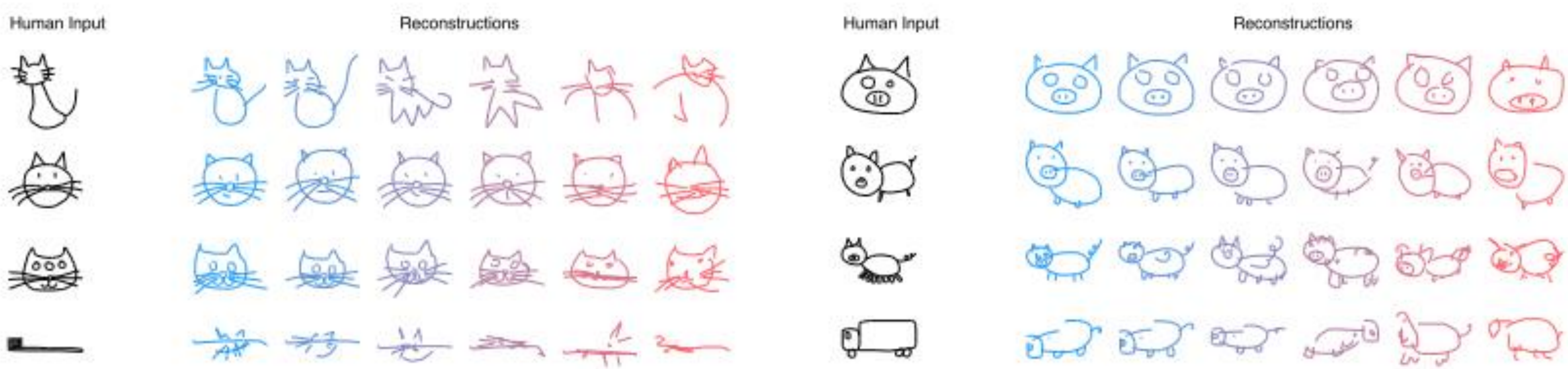
Beltramelli, Tony. "pix2code: Generating code from a graphical user interface screenshot." *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. 2018.

SketchRNN: teaching a machine to draw

- A recurrent neural network (RNN) able to construct stroke-based drawings of common objects



Sketch-RNN



Conditional generation of cats (left) and pigs (right).

Ha, David, and Douglas Eck. "A neural representation of sketch drawings." *arXiv preprint arXiv:1704.03477* (2017).

Language understanding

- Recognize phrases and sentences being spoken by a talking face, with or without the audio

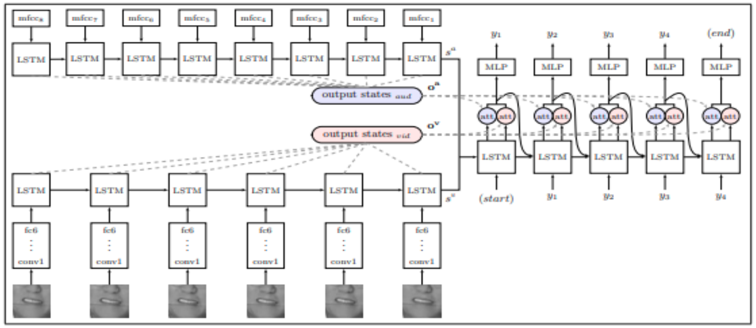


Figure 1. *Watch, Listen, Attend and Spell* architecture. At each time step, the decoder outputs a character y_i, as well as two attention vectors. The attention vectors are used to select the appropriate period of the input visual and audio sequences.

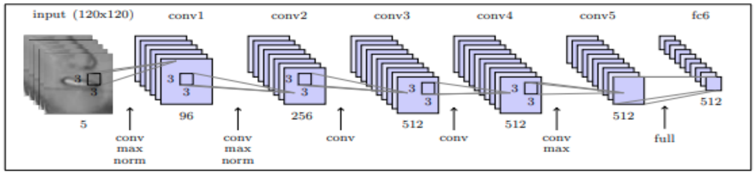
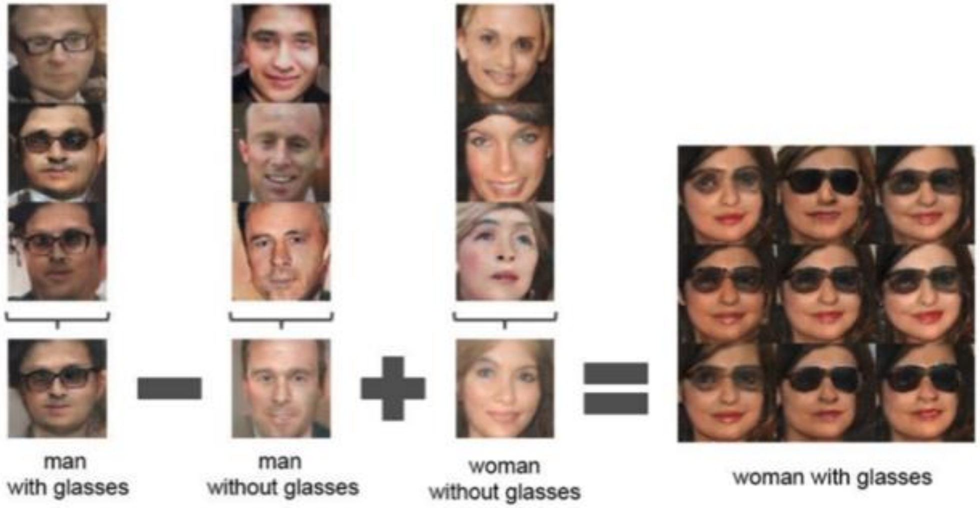


Figure 2. The ConvNet architecture. The input is five gray level frames centered on the mouth region. The 512-dimensional fc6 vector forms the input to the LSTM.

Chung JS, Senior A, Vinyals O, Zisserman A. Lip reading sentences in the wild. In2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017 Jul 21 (pp. 3444-3453). IEEE.

Application of generative models

- GANs Generated images



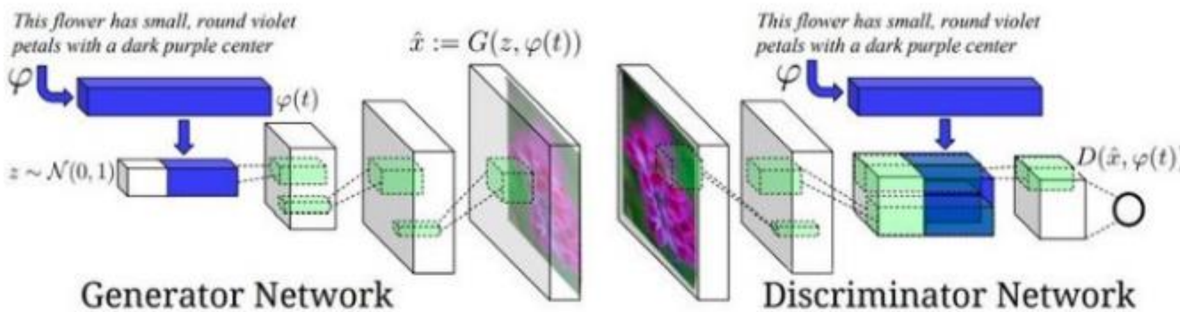
StyleGAN results

Picture: *These people are not real – they were produced by our generator that allows control over different aspects of the image.*

Karras, Tero, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4401-4410. 2019.

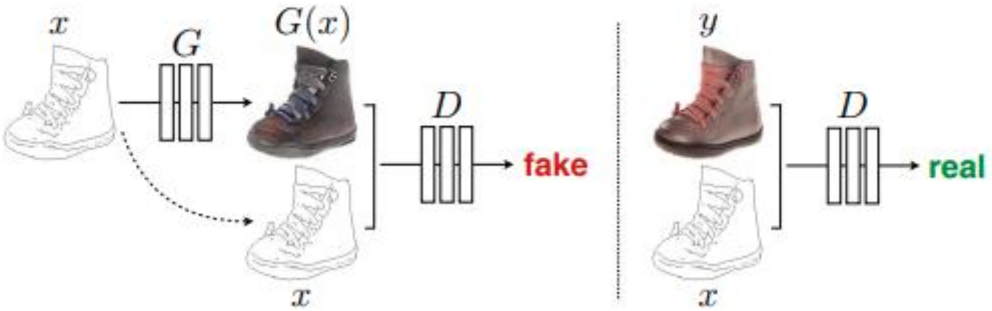
Synthesization of an image from a text description

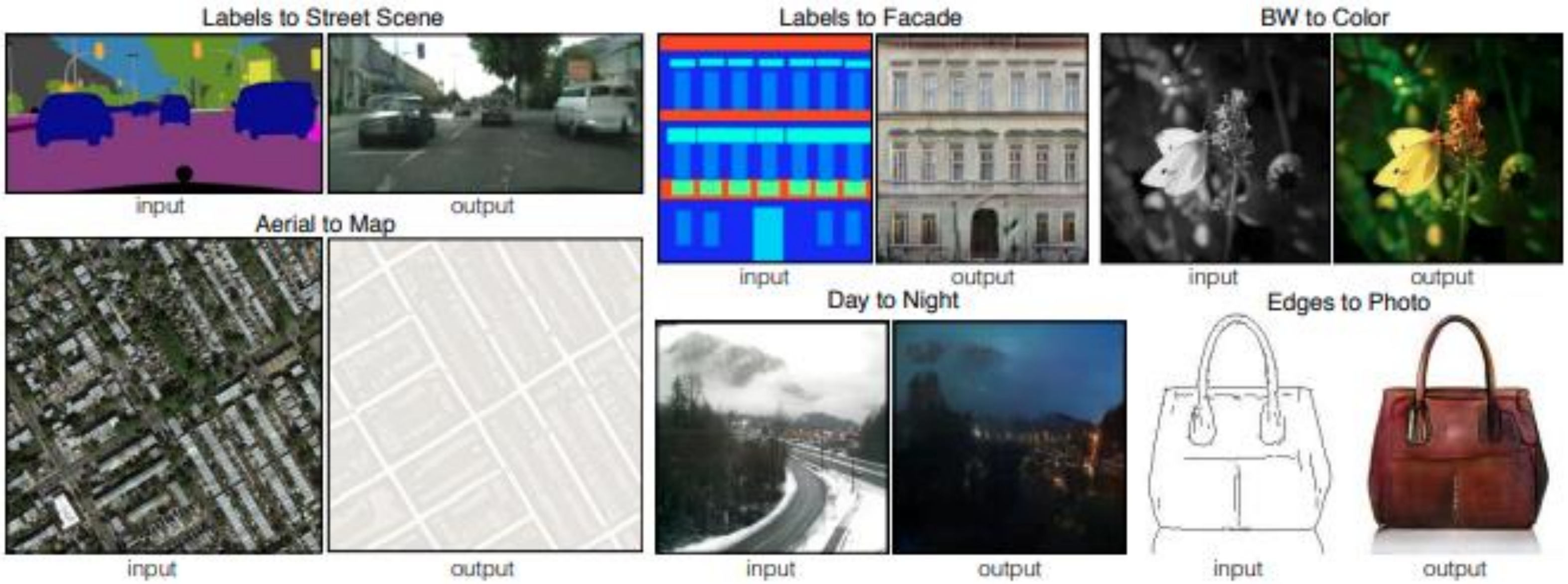
- Text2Image:



Pix2pix

- Image2Image…





Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. "Image-to-image translation with conditional adversarial networks." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125-1134. 2017.

Google DeepMind AlphaGo vs Lee Sedol

AlphaGo seals 4-1 victory over Go grandmaster Lee Sedol

DeepMind's artificial intelligence astonishes fans to defeat human opponent and offers evidence computer software has mastered a major challenge

The world's top Go player, Lee Sedol, lost the final game of the Google DeepMind challenge match. Photograph: Yonhap/Reuters

Google DeepMind's AlphaGo program triumphed in its final game against South Korean Go grandmaster Lee Sedol to win the series 4-1, providing further evidence of the landmark achievement for an artificial intelligence program.

https://www.youtube.com/watch?v=WXuK6gekU1Y&t=899s

DeepMind solves protein folding | AlphaFold 2

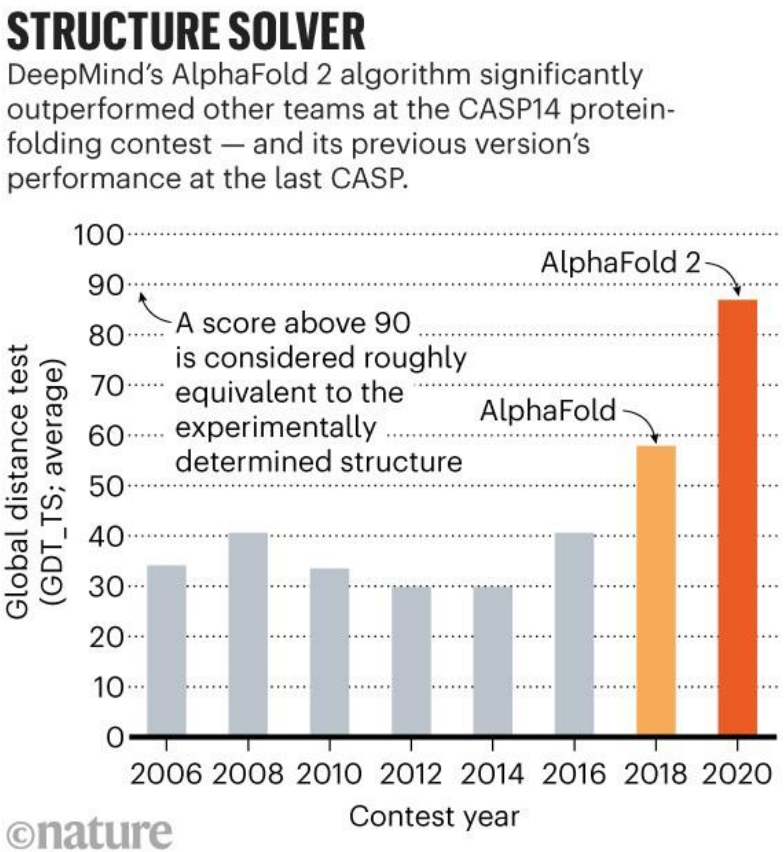
What happened?

DeepMind's AlphaFold2 solves protein folding **(50 years old grand challenge)**

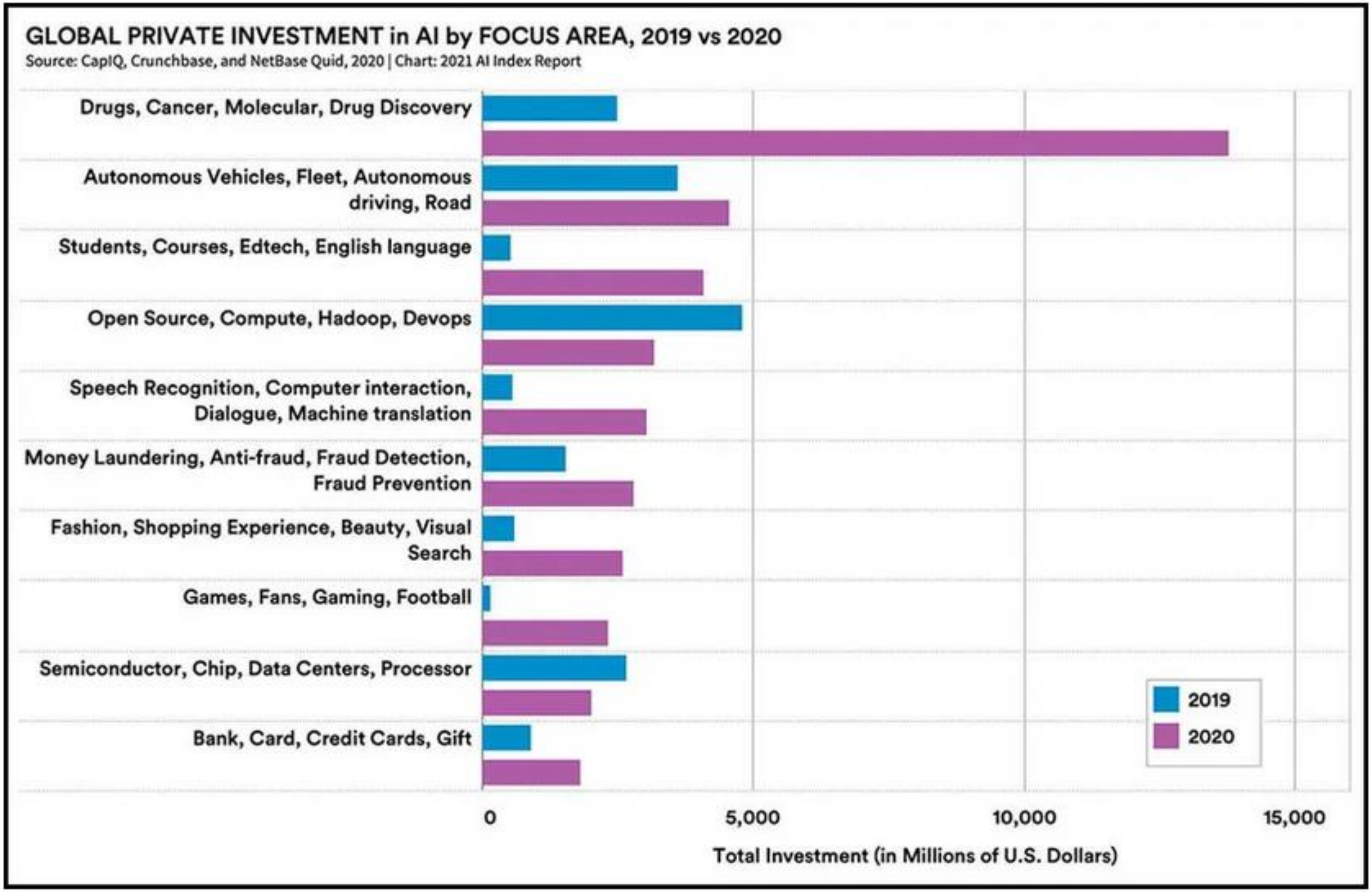
→ "Solves" = Achieves 87+ GDT of CASP

How big is this accomplishment?

- Biggest advancement is structural biology of the past 20+ years
- Biggest advancement in artificial intelligence of the past 20+ years
  - ImageNet moment (AlexNet)
- **Prediction: First Nobel Prize for machine learning model**



Deep learning applications



Source: https://spectrum.ieee.org/the-state-of-ai-in-15-graphs/9-the-startup-frenzy-is-over

Course organization & deliverables

- 5 ~ 6 Assignments (50%)
  - Mix of theory and implementation
  - Codes are available in : https://github.com/zahangircse/COMP_EECE_7or8740_NNs
  - **First one goes out end of next week**
    - Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early
- Examinations (10%)
- Progress Reports (10%)
- Term project (30%)
  - Project will be done in groups of 1- 2

Final project

- Goal
  - To explore Deep Learning models
  - Encouraged to apply on Computer vision, Speech, NLP, Medical imaging, Robotics, Bioinformatics and so on.
  - Must be done this semester.
- Main categories
  - Application/Survey
    - Compare a bunch of existing algorithms on a new application of your interest
  - Formulation/Development
    - Formulate a new model or algorithm for a new or old problem
  - Theory
    - Theoretically analyze an existing deep learning approach

Computing

- Major bottleneck
  - GPUs

- Options
  - Your own / group / advisor's resources

  - Google COLAB for free : https://towardsdatascience.com/getting-started-with-google-colab-f2fff97f594c

  - Google Cloud Credits
    - $50 credits to every registered student courtesy Google

  - UM / CS Department GPU cluster (if available)

Summary

- What is Artificial Intelligence (AI), Machine Learning (ML), and Neural Networks (NN).
- Neuromorphic computing
- Machine learning system and types
- How does machine learning system work?
- What is Deep Neural Network(DNN) and why DNN?
- Applications
- What’s next:
  - Neural Networks (NN)
  - Back-propagation for NN
  - Momentum and batch learning approaches
  - Ecosystem for Deep Learning (DL)