Intro to Deep Learning

John Ramey
uStudio
@ramhiser

Call me Ramey

- Lead Data Scientist @uStudio
- Background: Statistics / Computer Science
- Baylor alum

Deep Hype

VANITY FAIR

HIVE

BUSINESS | POLITICS | TECHNOLOGY | VF SUMMIT | VF.COM | SUBSCRIBE

ELON MUSK: A MACHINE TASKED WITH GETTING RID OF SPAM COULD END HUMANITY

BY LESSLEY ANDERSON | OCTOBER 2014

Source: http://www.vanityfair.com/news/tech/2014/10/elon-musk-artificial-intelligence-fear future tense ASU | NEW AMERICA | SLATE

Learn more about Future Tense »

future tense THE CITIZEN'S GUIDE TO THE FUTURE | FEB. 14 2014 2:07 PM

Why Watson Is Real Artificial Intelligence

By Miles Brundage and Joanna Bryson

30

Source: http://www.slate.com/blogs/future_tense/2014/02/14/watson_is_real_artificial_intelligence_despite_claims_to_the_contrary.html

OREN ETZIONI BUSINESS 06.15.16 7:00 AM

DEEP LEARNING ISN'T A DANGEROUS MAGIC GENIE. IT'S JUST MATH

Machine Learning in One Slide

Image

Label/Search

Text

NLP/Search

Audio

Speech Recognition

Video

Label/Search

Image

Label/Search

Text

NLP/Search

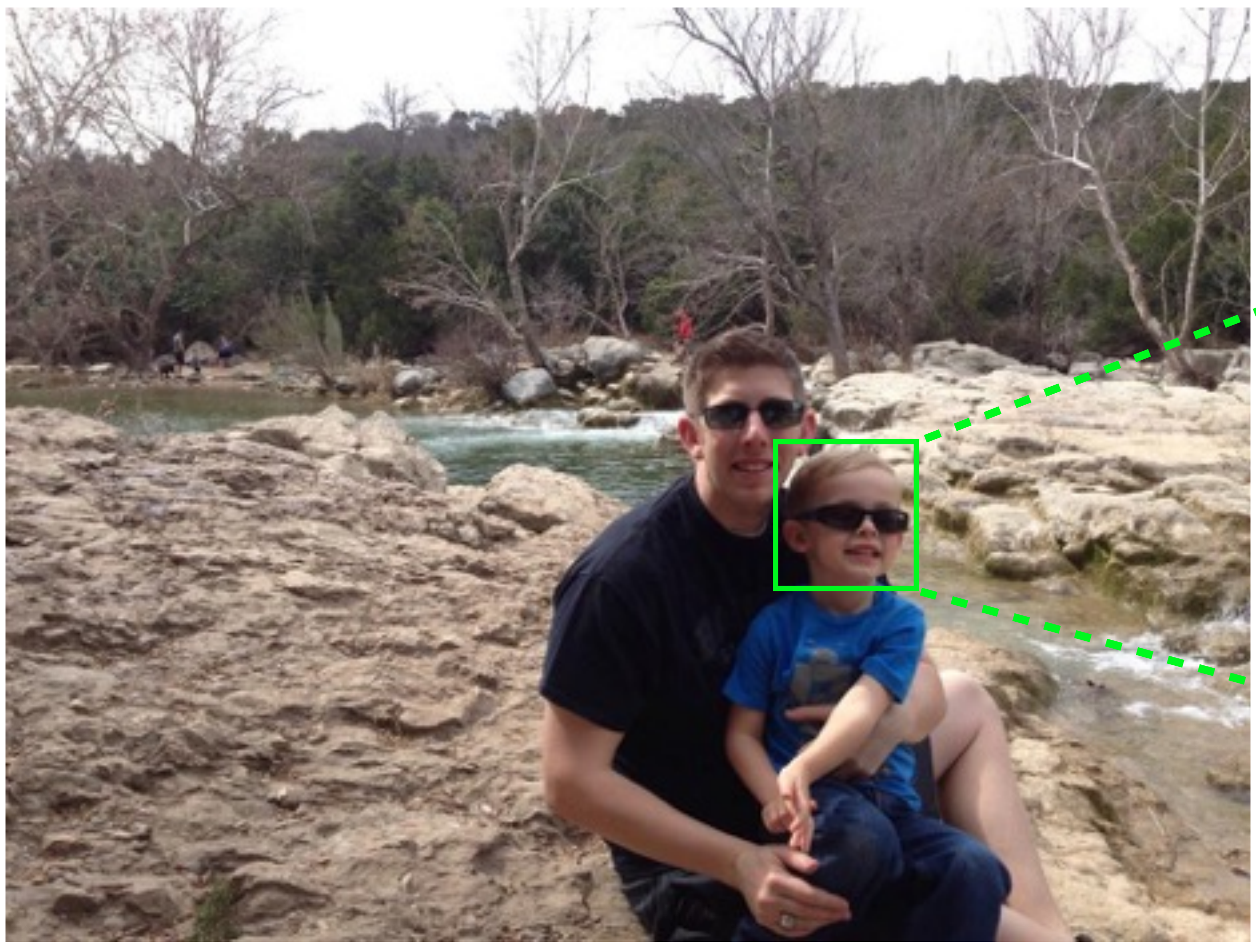
Audio

Speech Recognition

Video

Label/Search

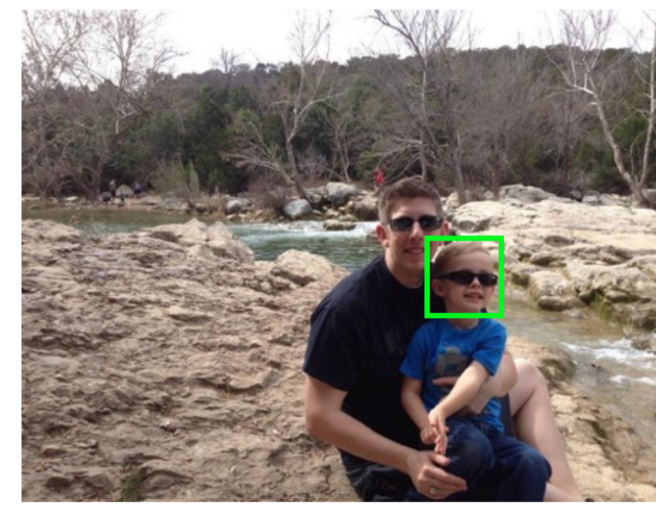
Why is computer vision **difficult**?



210	217	203	229	230	229
208	211	233	223	203	161
178	210	185	208	166	181
170	169	172	214	182	223
229	212	232	194	224	211
227	209	197	236	232	215

Simple Task

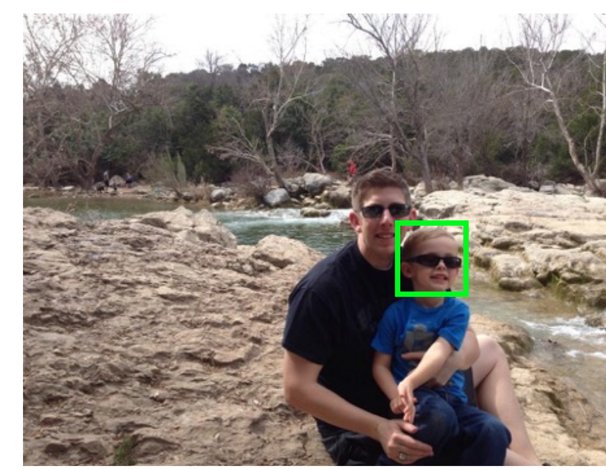
Learning Algorithm



Features Needed

Feature
Representation

Learning
Algorithm



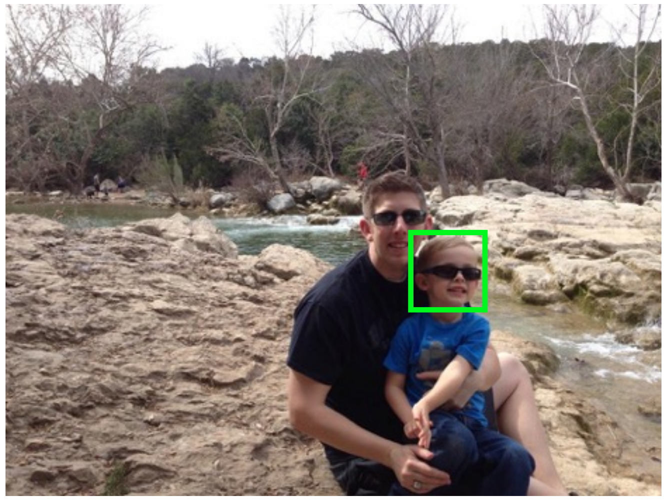
Features: Computer Vision

SIFT

SURF

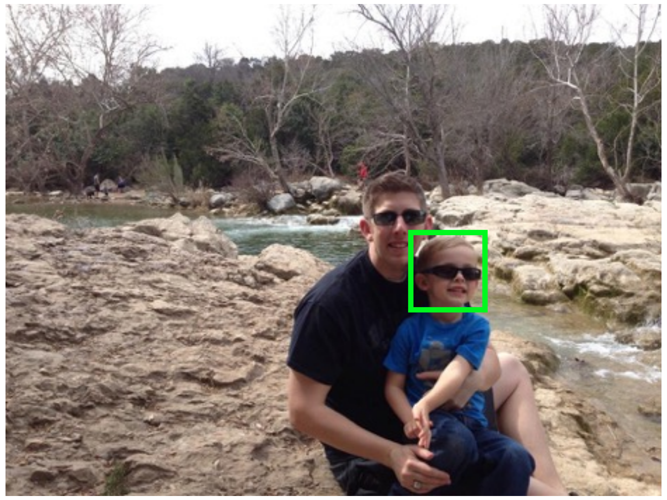
We Prefer…

Learning
Algorithm



We Prefer…

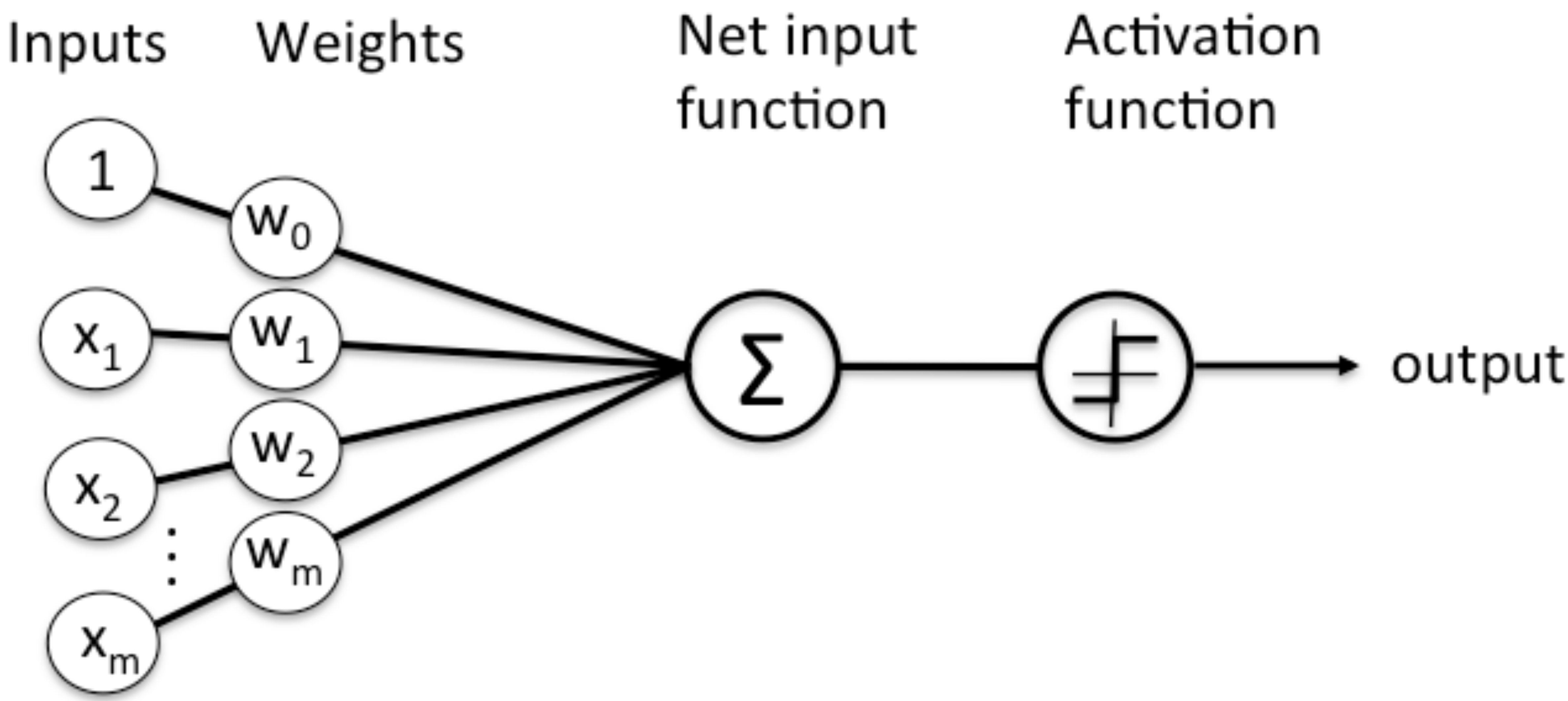
Deep Learning



What is **Deep Learning**?

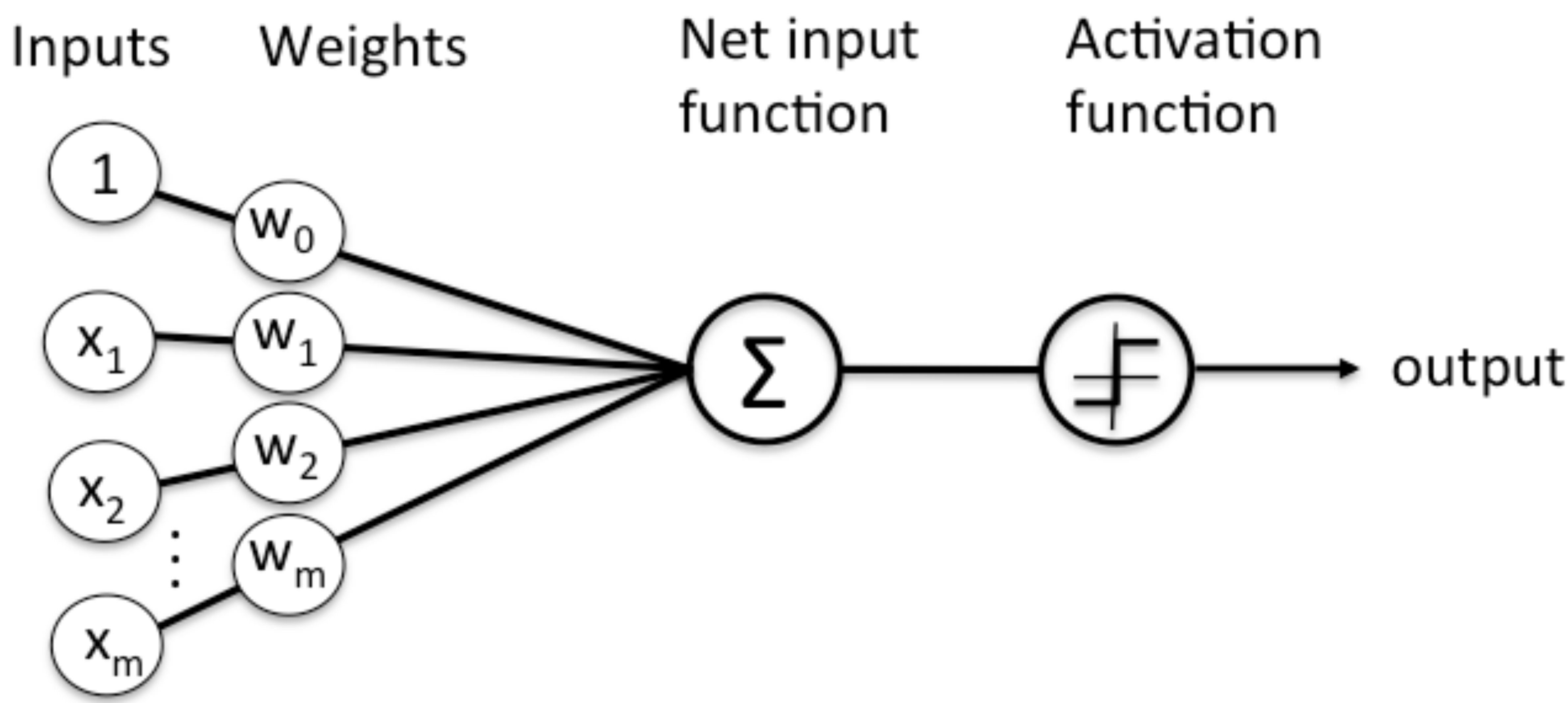
Deep Learning is a **neural network** with many **hidden layers**

No Hidden Layers



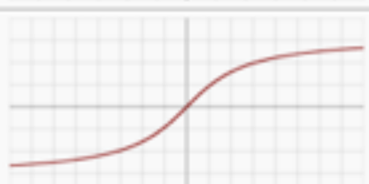
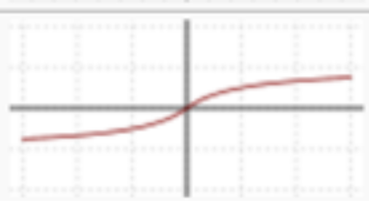
Source: http://sebastianraschka.com/Articles/2015_singlelayer_neurons.html

Logistic Regression!

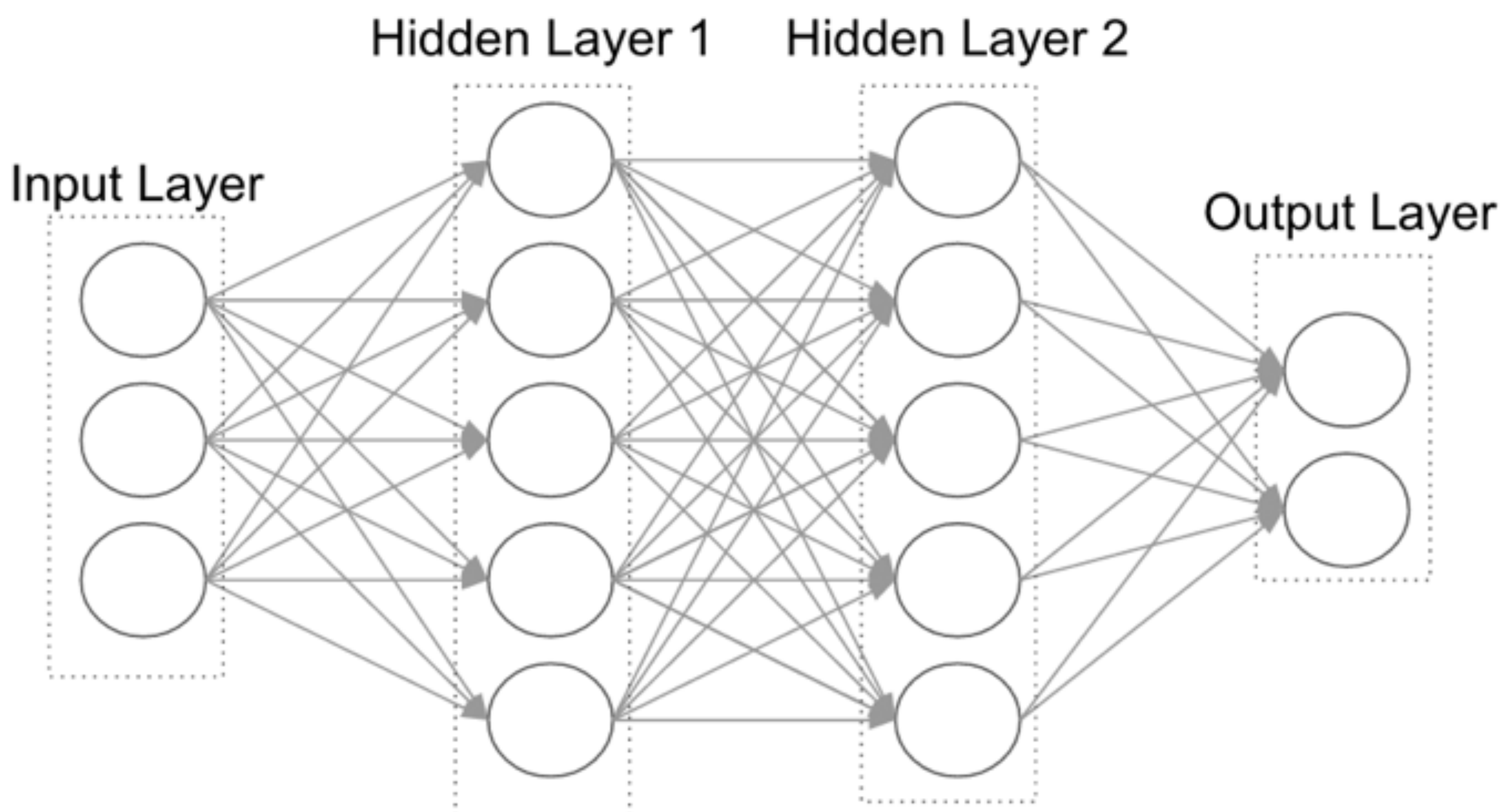


Source: http://sebastianraschka.com/Articles/2015_singlelayer_neurons.html

Activation Functions

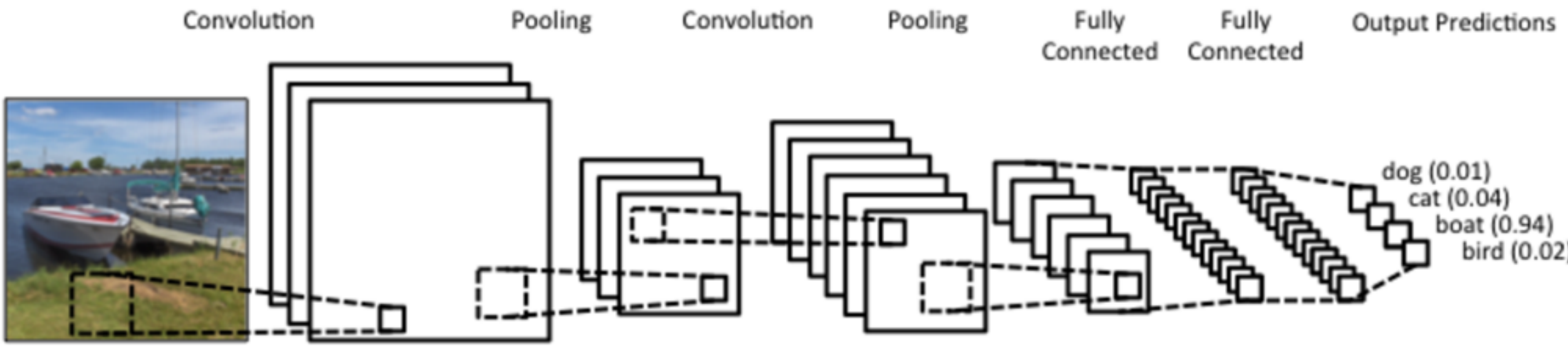
Name	Plot	Equation	Derivative
Identity		$f(x) = x$	$f'(x) = 1$
Binary step		$f(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 & \text{for} \quad x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for} \quad x \neq 0 \\ ? & \text{for} \quad x = 0 \end{cases}$
Logistic (a.k.a Soft step)		$f(x) = \frac{1}{1 + e^{-x}}$	$f'(x) = f(x)(1 - f(x))$
TanH		$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1$	$f'(x) = 1 - f(x)^2$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \frac{1}{x^2 + 1}$
Softsign [7][8]		$f(x) = \frac{x}{1 + \lvert x \rvert}$	$f'(x) = \frac{1}{(1 + \lvert x \rvert)^2}$
Rectifier[9]		$f(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 & \text{for} \quad x \geq 0 \end{cases}$

Neural Net with 2 Layers

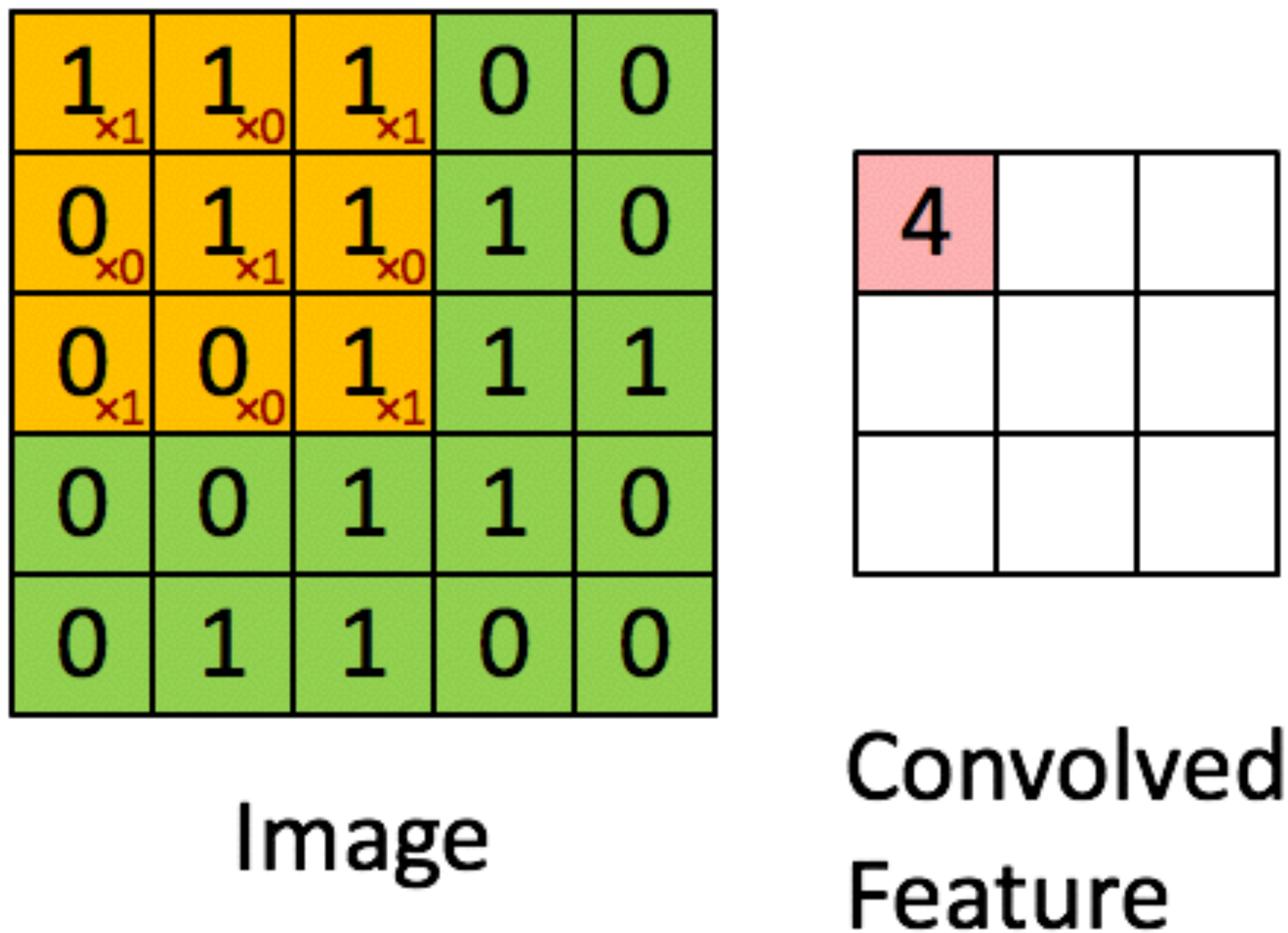


Feedforward neural network with 2 hidden layers

The Most Controversial Slide I've Ever Made

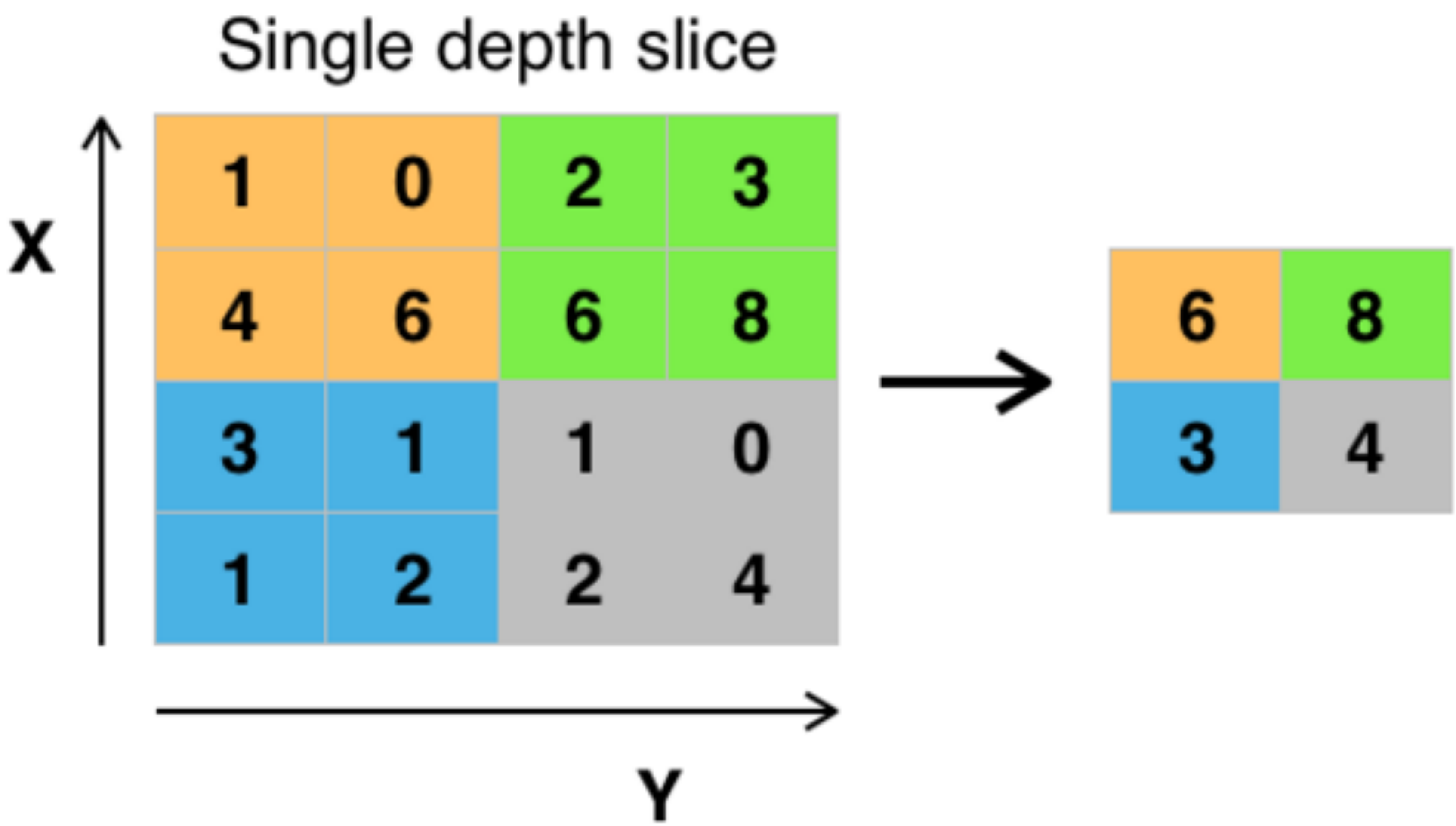


What is a Convolution?

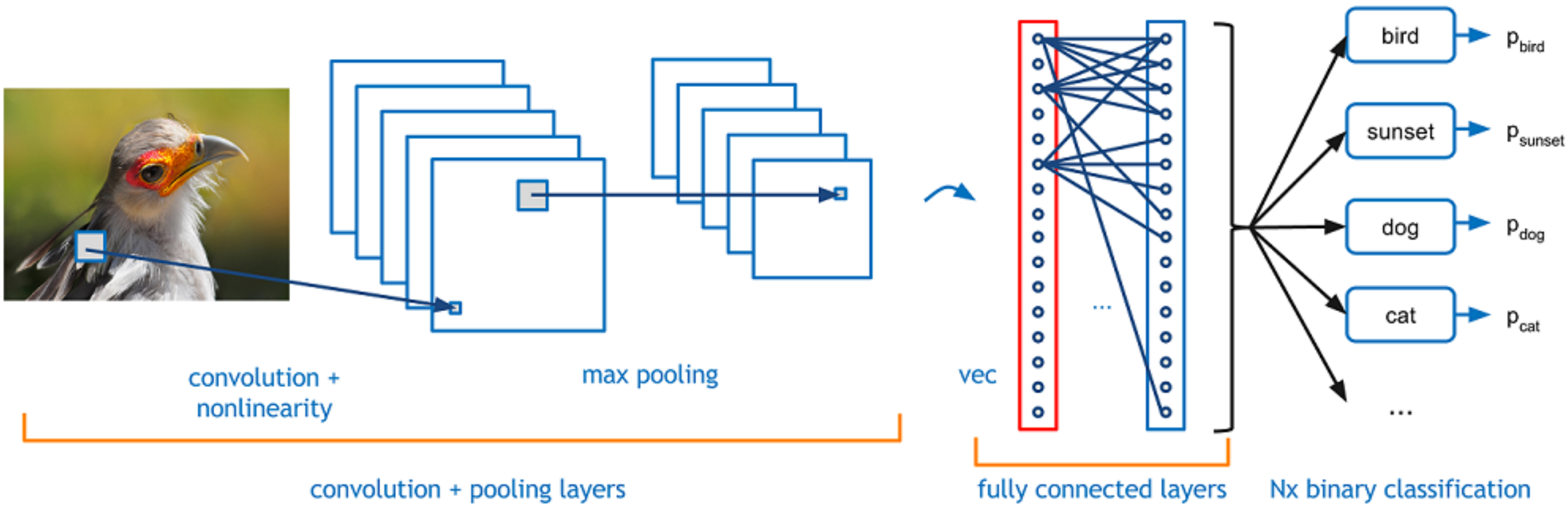


Source: http://deeplearning.stanford.edu/wiki/index.php/Feature_extraction_using_convolution

Max Pooling

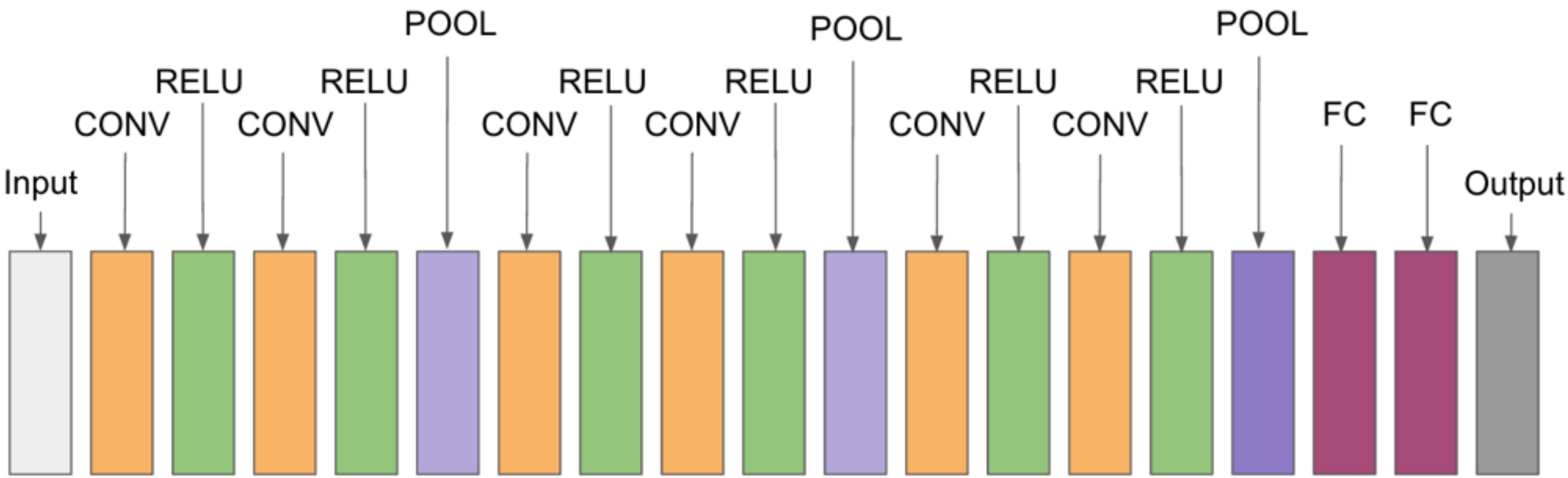


Max pooling with a 2x2 filter and stride = 2
(source: Wikipedia)



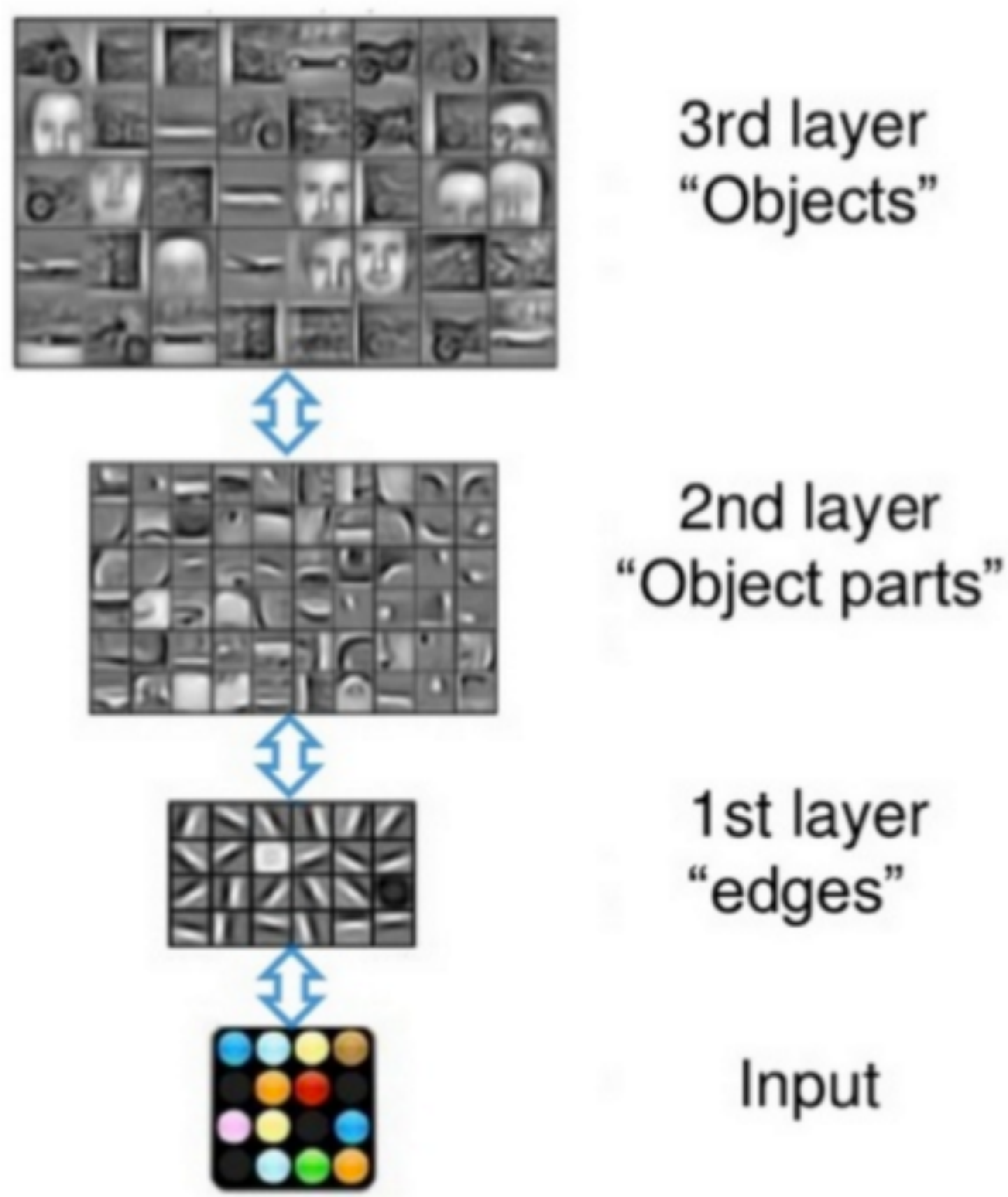
Source: https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner’s-Guide-To-Understanding-Convolutional-Neural-Networks/

Example Architecture



Example of a CNN Architecture

Visualizing Layers



Source: Lee et al. (2009), ICML

How to **Train**?

- Typically with **Stochastic Gradient Descent**
- Standard Objective Function: **Cross Entropy**
- Covered more in a later talk!

Other Deep NNs

- Autoencoders
- Recurrent Neural Networks (RNN)
- Lots more
  - Plenty of opportunities for talks!

Why **now**?

Not New

PROC. OF THE IEEE, NOVEMBER 1998 1

Gradient-Based Learning Applied to Document Recognition

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner

Abstract— **Multilayer Neural Networks trained with the backpropagation algorithm constitute the best example of a successful Gradient-Based Learning technique. Given an appropriate network architecture, Gradient-Based Learning algorithms can be used to synthesize a complex decision surface that can classify high-dimensional patterns such as handwritten characters, with minimal preprocessing. This paper reviews various methods applied to handwritten character recognition and compares them on a standard handwritten digit recognition task. Convolutional Neural Networks, that are specifically designed to deal with the variability of 2D shapes, are shown to outperform all other techniques.**

Real-life document recognition systems are composed of multiple modules including field extraction, segmentation, recognition, and language modeling. A new learning paradigm, called Graph Transformer Networks (GTN), allows such multi-module systems to be trained globally using Gradient-Based methods so as to minimize an overall performance measure.

I. Introduction

Over the last several years, machine learning techniques, particularly when applied to neural networks, have played an increasingly important role in the design of pattern recognition systems. In fact, it could be argued that the availability of learning techniques has been a crucial factor in the recent success of pattern recognition applications such as continuous speech recognition and handwriting recognition.

The main message of this paper is that better pattern recognition systems can be built by relying more on automatic learning, and less on hand-designed heuristics. This is made possible by recent progress in machine learning and computer technology. Using character recognition as a case study, we show that hand-crafted feature extrac-

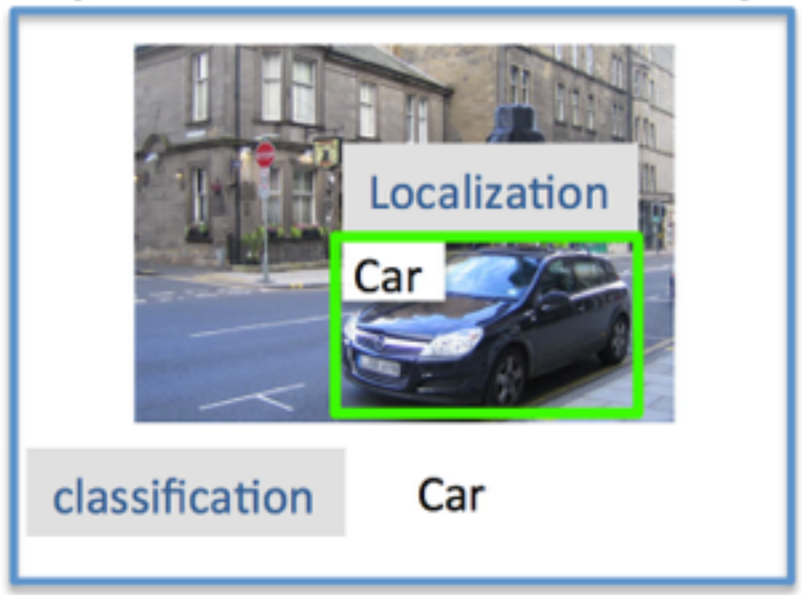
Source: http://yann.lecun.com/exdb/publis/pdf/lecun-98.pdf

ImageNet Challenge 2012

Task 1: Classification

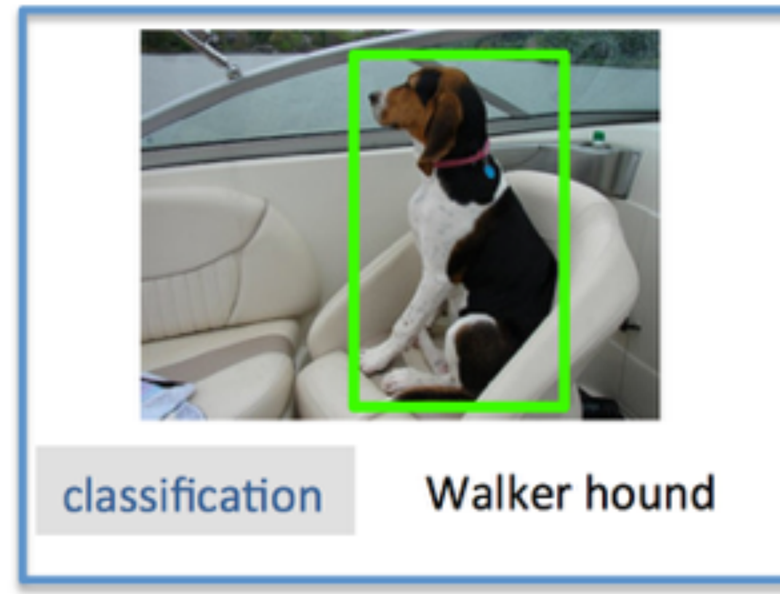
- Predict a class label
- 5 predictions / image
- 1000 classes
- 1,200 images per class for training
- Bounding boxes for 50% of training.

Task 2: Detection (Classification + Localization)



- Predict a class label and a bounding box
- 5 predictions / image
- 1000 classes
- 1,200 images per class for training
- Bounding boxes for 40% of training.

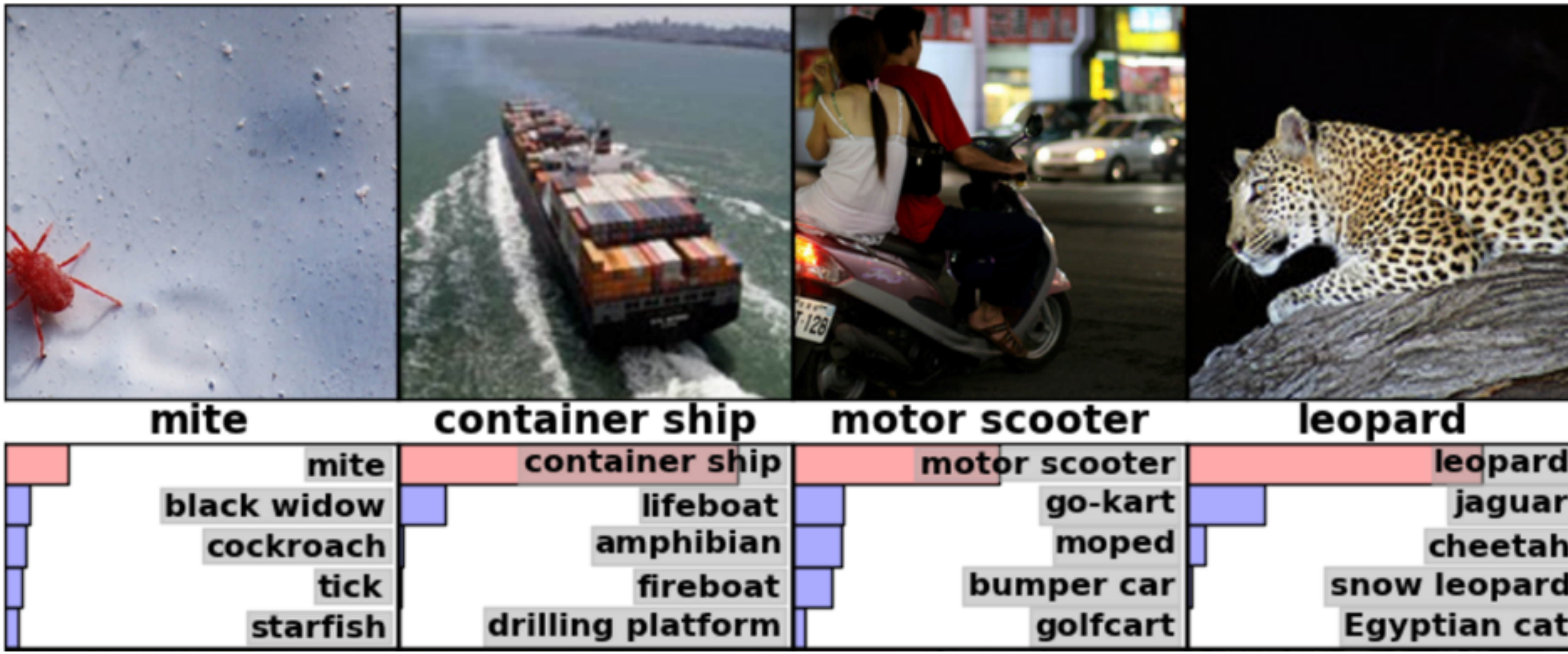
Task 3: Fine-grained classification



- Predict a class label given a bounding box in test
- 1 prediction / image
- 120 dog classes (subset)
- ~200 images per class for training (subset)
- Bounding boxes for 100% of training

Source: http://image-net.org/challenges/LSVRC/2012/ilsvrc2012.pdf

ImageNet Challenge 2012



Source: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks

ImageNet 2012 Results

Top 5	Error Rate
AlexNet	**16%**
ISI	26%
OXFORD_VGG	27%
XRCE/INRIA	27%
LEAR-XCE	34%

GPUs

Frameworks

TensorFlow

theano

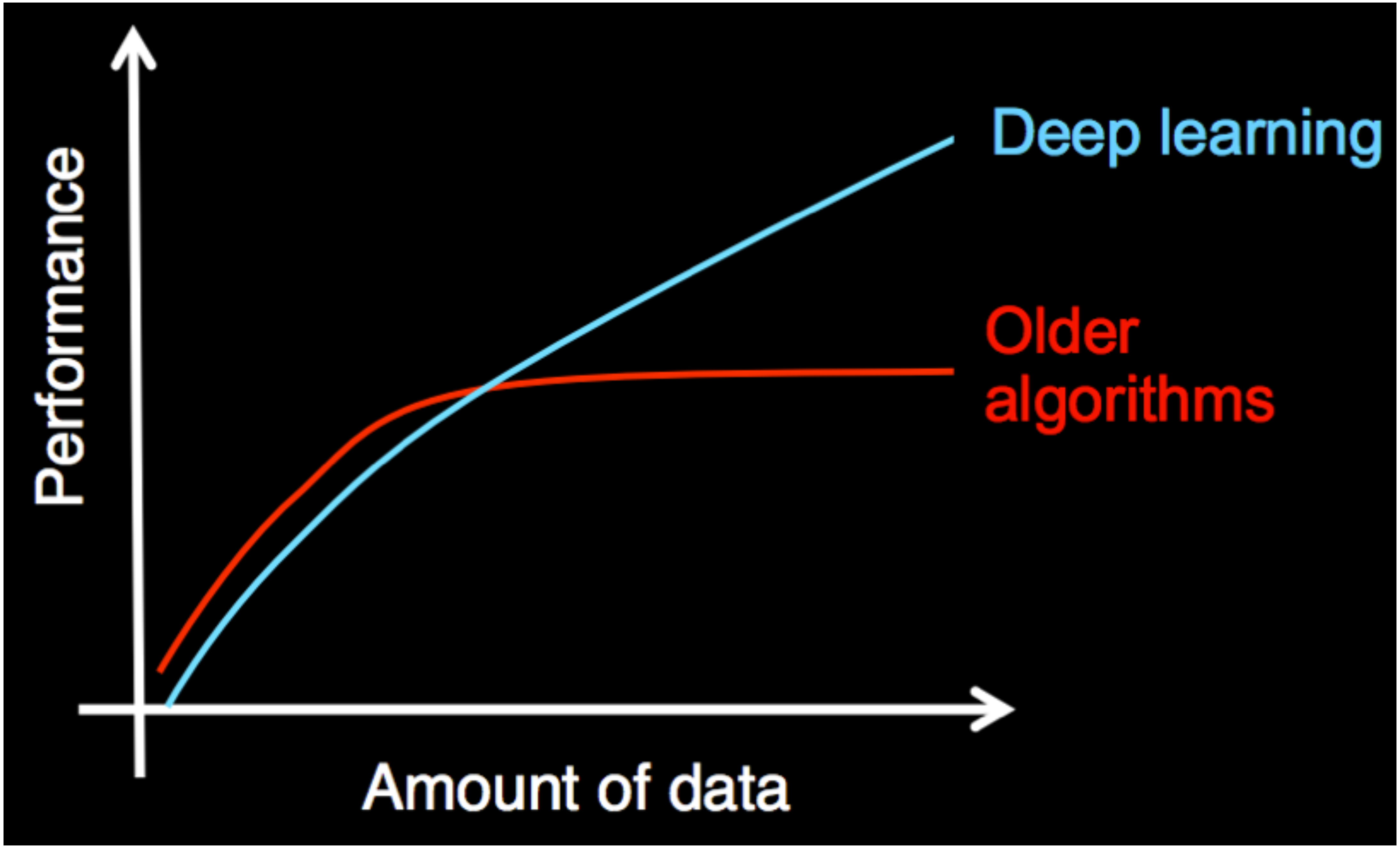
torch

Facebook AI Research

K

DEEPLEARNING4J

The Claim



Source: http://cs229.stanford.edu/materials/CS229-DeepLearning.pdf

Highly Overparameterized

- Deep Neural Nets — highly overparameterized
  - 1.2 million images in ImageNet
  - Billions of parameters to estimate
  - **Regularization is a must!**

Computers are slow

Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification

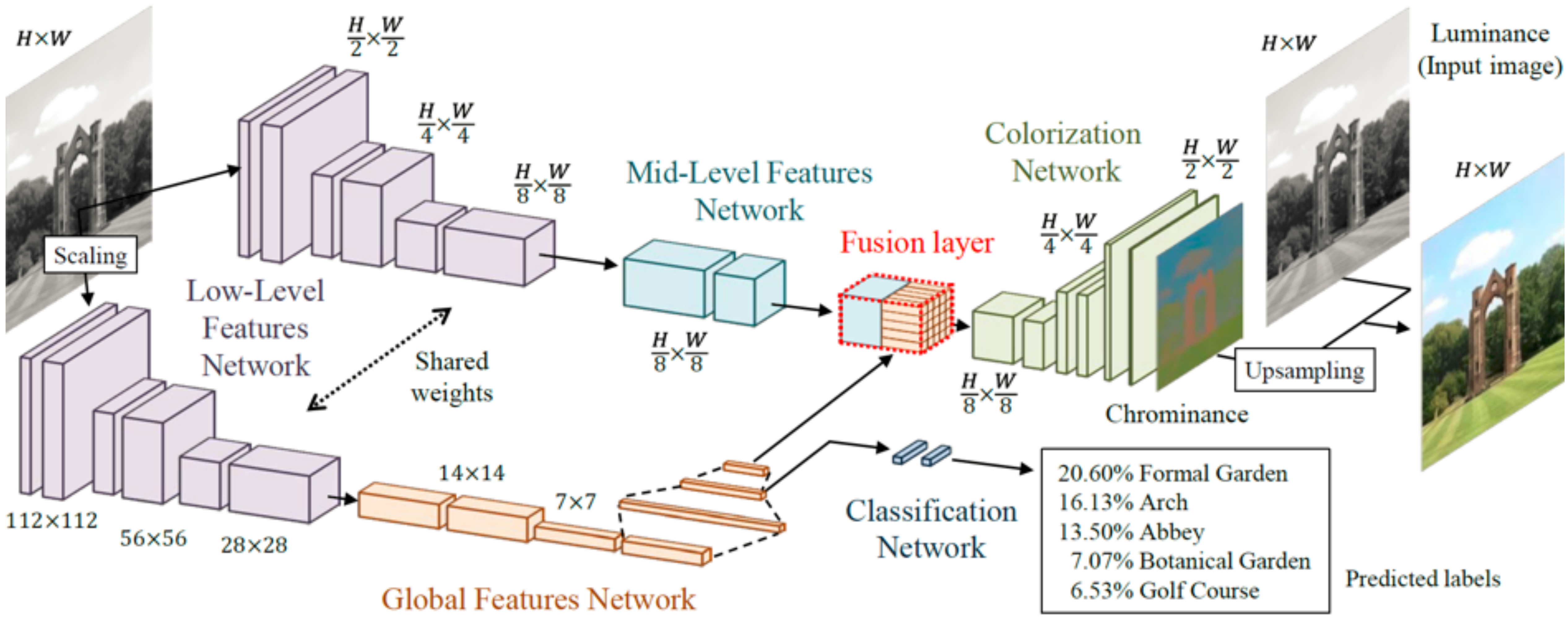
Satoshi Iizuka* Edgar Simo-Serra* Hiroshi Ishikawa (*equal contribution)

Waseda University

Colorado National Park, 1941 | Textile Mill, June 1937 | Berry Field, June 1909 | Hamilton, 1936

Source: http://hi.cs.waseda.ac.jp/~iizuka/projects/colorization/en/

Computers are slow

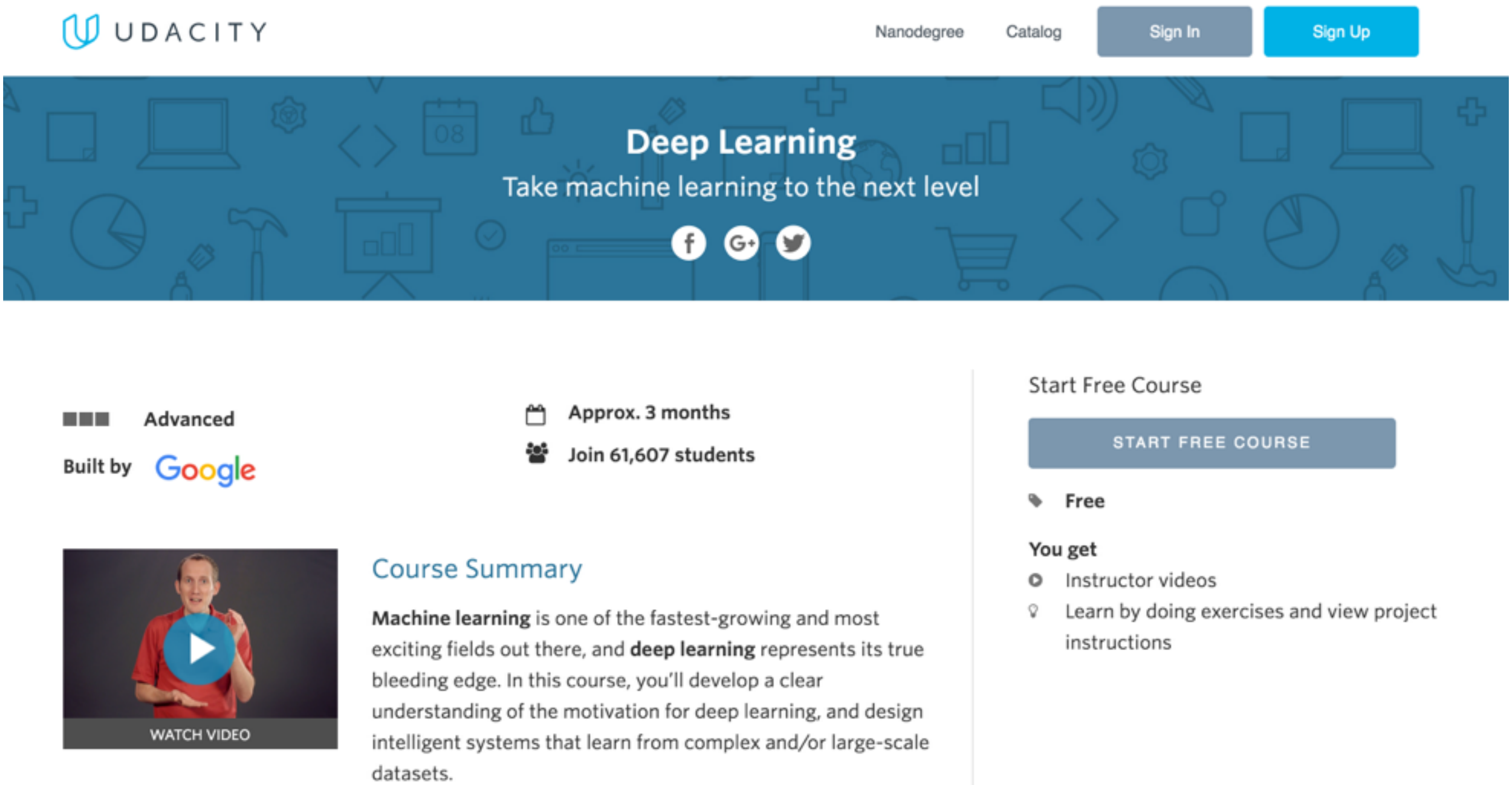


Source: http://hi.cs.waseda.ac.jp/~iizuka/projects/colorization/en/

Computers are slow

- Training Time: ~3 weeks on NVIDIA Tesla K80 GPU
- Colorization Time: ~1 second

How to **Begin**?



Source: https://www.udacity.com/course/deep-learning--ud730

Other Tutorials

- Official TensorFlow Tutorial
- Keras Documentation
- DeepLearning.net Tutorial
- Neural Networks and Deep Learning (Free Book)

Conclusion

- Deep Learning is **powerful**
- **Network architecture** instead of feature representation
- **Overfitting** can be a challenge
- **Fast** computer is warranted

Questions?