Introduction

Deep Learning (MSDS)

Dr. Adeel Mumtaz

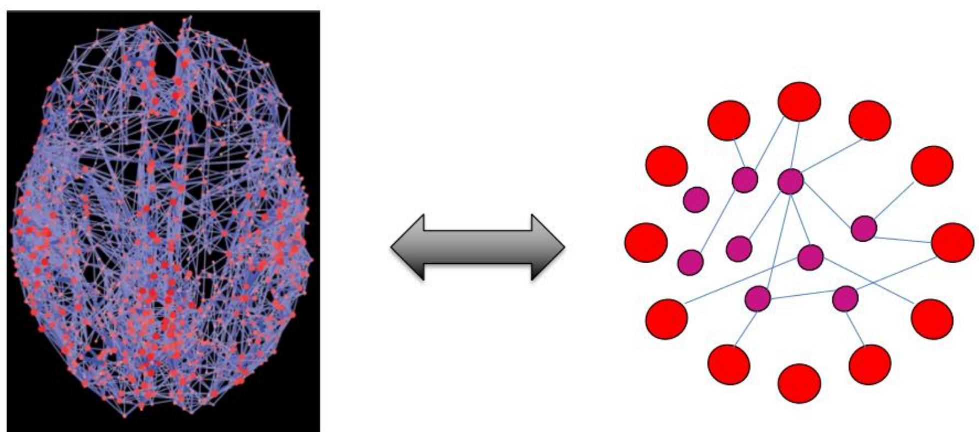
Lec 1

Fall, 2022

National University
Of Computer and Emerging Sciences

Welcome to Deep Learning (MSDS)

- An overview of various deep architectures and learning methods
- Develop fundamental and practical skills at applying deep learning to your research
- “Deep Learning is the new electricity"



About Instructor

- Dr. Adeel Mumtaz
  - adeelmumtaz@gmail.com
  - Contact #: 03215236361
- Education
  - BS (CS) PIEAS in 2004
  - MS (CS) GIKI in 2006
  - PhD (CS) CityU Hong Kong in 2015
- Experience
  - 2006-now Project Director (General Manager), COE, NESCOM
  - 2004-2005 Software Engineer, Elixir Technologies
  - 2015-now Visiting Faculty Member, FAST
- Research
  - Leading Researcher in Computer Vision and Machine Learning
  - Published work in PAMI, CVPR
  - VISAL group @ CityU
- Awards
  - Gold Medal in BS & MS
  - NESCOM Fellowship
  - UGC Hong Kong Grant for PhD
    - Outstanding Research Excellence Award
  - HEC best research paper award

My Research

PATENTS AND PUBLICATIONS (OVERALL >10 IMPACT FACTOR AND >100 CITATIONS)

1. **Mumtaz, A**.; Coviello, E.; Lanckriet, G.; Chan, A., "A Scalable and Accurate Descriptor for Dynamic Textures using Bag of System Trees," *Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on*
2. **Mumtaz, A.**; Weichen Zhang; Chan, A.B., "Joint Motion Segmentation and Background Estimation in Dynamic Scenes," *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on* , pp.368,375, 23-28 June 2014
3. **Mumtaz, A**.; Coviello, E.; Lanckriet, G.R.G.; Chan, A.B., "Clustering Dynamic Textures with the Hierarchical EM Algorithm for Modeling Video," *Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on* , vol.35, no.7, pp.1606,1621, July 2013
4. Coviello, E.; **Mumtaz, A**.; Chan, A.B.; Lanckriet, G.R.G., "Growing a bag of systems tree for fast and accurate classification," *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* , pp.1979,1986, 16-21 June 2012
5. Emanuele Coviello, **Adeel Mumtaz**, Gert.R.G. Lanckriet, Antoni B. Chan, That was fast! Speeding up nn search of high dimensional distributions. **International Conference on Machine Learning** (ICML), 2013.

http://visal.cs.cityu.edu.hk/

Contents

Course Objectives

- Understand and design deep neural network architectures
- Study various application areas of deep neural networks
- Learn about famous existing deep learning tools and software
- Implement and apply deep learning techniques to solve real world problems

Administrative Matters

- Recommended Books
  - Deep Learning
    - Deep Learning by Goodfellow, Bengio and Courville
    - Deep Learning with PyTorch Step-by-Step by Daniel Voigt Godoy
    - Deeplearning.ai
  - Pattern classification
    - Pattern Classification, by R.O. Duda, P.E. Hart, and D.G. Stork
  - Machine learning
    - Pattern Recognition and Machine Learning, Christopher M. Bishop
  - Computer vision
    - **Computer Vision: Algorithms and Applications,** Richard Szeliski.
    - **Computer Vision: A Modern Approach**, D.A. Forsyth and J. Ponce
  - NLP
  - In addition, we will extensively use online materials (video lectures, blog, posts, surveys, papers, etc.)
    - http://cs231n.stanford.edu/ http://course.fast.ai/ https://developers.google.com/machine-learning/crash-course/ www.nvidia.com/dlilabs http://introtodeeplearning.com/ https://github.com/oxford-cs-deepnlp-2017/lectures, https://jalammar.github.io/

Administrative Matters

- Attendance Policy
  - FAST attendance Policy?
  - Students not making up to the 80% attendance will not be allowed to sit in the final exam
  - Latecomers will be marked absent

Administrative Matters…

- Assignment / Quiz Policy:
  - Deadlines must be followed. There will be no makeup of any quizzes
  - Late assignment will be accepted with negative marking, -1 for each day
  - A copied assignment/project will be simply marked zero.
- Lab Policy:
  - ?
- Schedule:
  - Tuesday 17:20-18:40, Room-?
  - Thursday 17:20-18:40, Room-?

Coursework

- Quizzes
  - 10%
- Midterm
  - 25%
- Assignments
  - 10%
- Semester Project
  - 20%
- Class participation
  - 5%
- Final Exam 30%

Course Project

The course project gives students a chance to apply deep architectures discussed in class to a research oriented project

- Face detection and recognition
- Object recognition (in supermarkets)
- Speech understanding
- Loan Approvals
- Biometrics
- Database Indexing
- Target Tracking and Recognition
- Aerial/Satellite Data Analysis
- Traffic Monitoring
- Human Activity Recognition/Interactive Games
- Social network mining
- Genetics Analysis
- Visualization
- Chatbots
- Whatever you're interested

Evaluation of article reading and project

- Report
  - Article reading: Submit a survey of the articles you read and the list of the articles
- Project
  - Submit an article including introduction, methods, experiments, results, and conclusions
  - Submit the project code, the readme document, and some testing samples (images, videos, etc.) for validation
- Presentation

Kaggle Competitions

- https://www.kaggle.com/competitions
- Everyone should submit entries!
- Points for
  - submitting and getting top ranking
- 5 point for submitting
- 5 point for top in class
- Project of your choice
- You can get an A by topping
  - in the top 10
  - leaderboard of any of the
    - Current Kaggle Competition

Open Problems - Multimodal Single-Cell Integration
Predict how DNA, RNA & protein measurements co-vary in single cells
Featured · 117 Teams · 3 months to go

DFL - Bundesliga Data Shootout
Identify plays based upon video footage
Featured · Code Competition · 182 Teams · 2 months to go

RSNA 2022 Cervical Spine Fracture Detection
Identify cervical fractures from scans
Featured · Code Competition · 266 Teams · 2 months to go

American Express - Default Prediction
Predict if a customer will default in the future
Featured · 4787 Teams · 2 days to go

Feedback Prize - Predicting Effective Arguments
Rate the effectiveness of argumentative writing elements from students grade 6-12
Featured · Code Competition · 1539 Teams · 21 hours to go

Google AI4Code – Understand Code in Python Notebooks
Predict the relationship between code and comments
Featured · Code Competition · 1160 Teams · 3 months to go

JPX Tokyo Stock Exchange Prediction
Explore the Tokyo market with your data science skills
Featured · Code Competition · 2033 Teams · a month to go

Prerequisites

- A good working knowledge of a programming language
  - C++
  - Matlab
  - Python?
- Data structures
- knowledge of probability and statistics
  - Expectation and Variance?
  - PMF, CDF, PDF?
- Linear algebra
  - Eigen values and vectors?
- Vector calculus
  - Gradient and Hessian matrix?

Prerequisites…

- Information retrieval
  - Precision/Recall?
- Machine Learning
  - Supervised vs. unsupervised learning?
  - K-mean clustering
  - SVM
  - ANN

Prerequisites…

- Image Processing/Computer Vision
  - Quantization and aliasing?
  - Indexed color and palette?
  - Famous edge detectors?
- Natural Language Processing
  - GPT
  - BERT

Class Participation Break

Name

Major/Minor/Year

CGPA (optional)

Hometown

What are you doing here?

Personal strengths/weaknesses

Undergrad FYP

What do you know about machine learning and deep learning?

Tentative Course Syllabus

- Deep Learning is a very wide field
- What will we study in 1 semester…?

- Basics of most common architectures
  - Develop/Implement few intelligent systems

Tentative Syllabus

Week	Topics
01, 02	Course Introduction Intro to Machine Learning & AI Probability, Linear Algebra, Python Review
03	Neural Networks and Backpropagation
04, 05	Training Deep Neural Networks using Pytorch
06, 07	Understanding and Visualizing CNNs
08, 09	Attention and Memory
10	Autoencoders and Autoregressive Models
11	Generative Adversarial Networks
12	Self-supervised Learning
13	Reinforcement Learning
14	Final Project Presentations

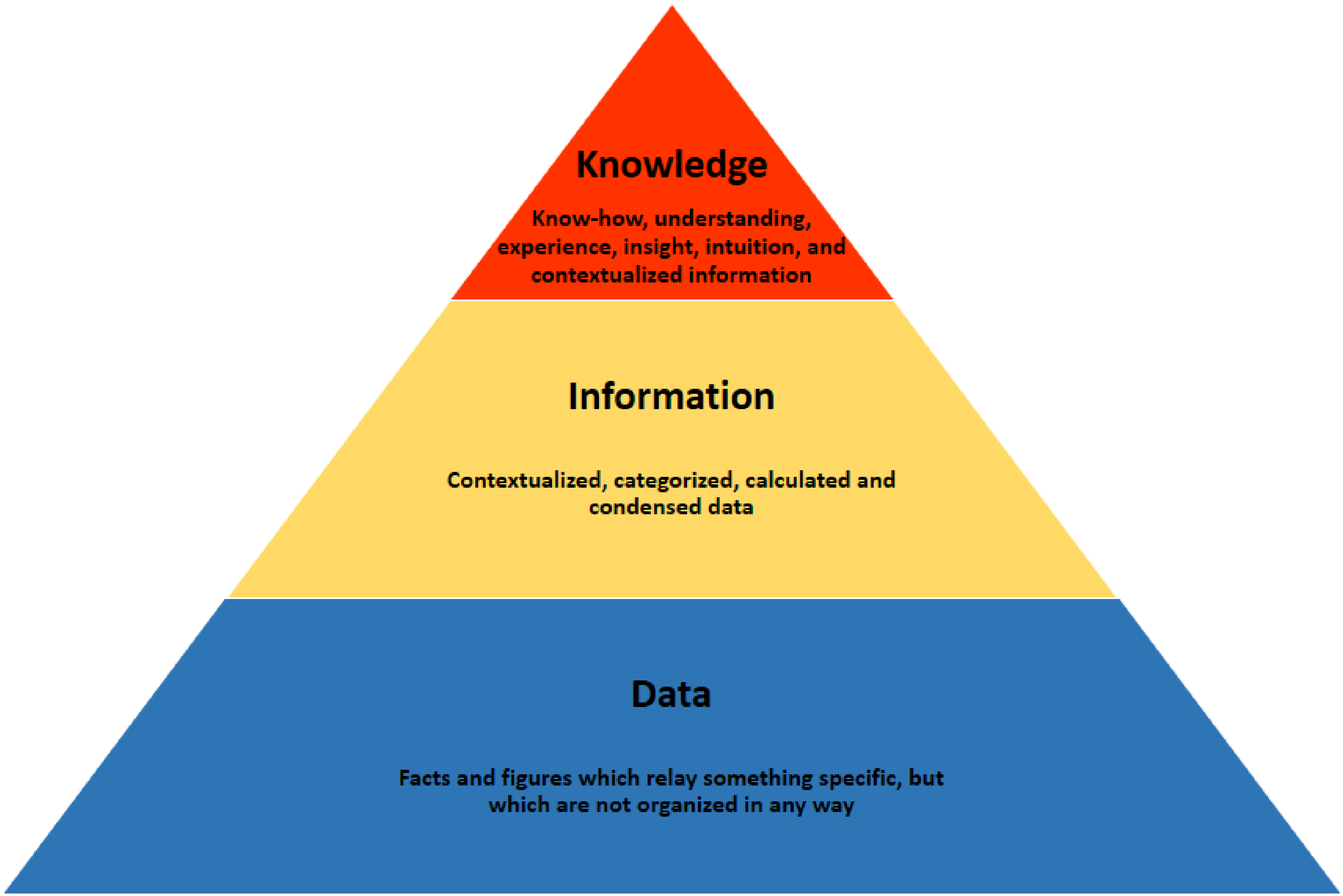
Contents

- Course Objectives
- Administrative Matters
- Course Work
- Grading
  - Kaggle Competitions
- Prerequisites
- Class Participation Break
- Tentative Syllabus
- Introduction
  - Data, Information and Knowledge?
  - Learning Algorithms
  - Supervised, Un-Supervised, Self and Reinforcement Learning
  - AI Vs ML Vs DL
  - Deep Learning History
  - Deep Learning Applications
  - Deep Learning Pipeline
  - Deep Learning Frameworks

Data, Information and Knowledge?

- Linear relationship in data mining
- Data
  - Raw facts, numbers, text, images, videos, ?
  - Generated from observation and activities
  - Form a list with no useful meaning
  - Repositories to hold data in different formats
- Learning algorithms
  - Data as raw input
  - Precious commodity
    - called knowledge at the output
  - Information is the middle portion
    - Raw input is churned

Data, Information and Knowledge?



An Example

- Rain in Islamabad
  - “Islamabad, Date: 16 February 2015, Rain, 3.38mm”
- Raw data no value for decision making
- Net amount of annual rainfall in Islamabad
  - information
- Compare with water demand of Islamabad's population
- Whether rainfall is sufficient or Water need to be imported
  - This assessment is knowledge

Comparison between Human and Machine

Human	Machine
Memorize	k-Nearest Neighbors, Case-based learning
Observe someone else, then repeat	Supervised Learning, Learning by Demonstration
Keep trying until it works (riding a bike)	Reinforcement Learning
20 Questions	Decision Tree
Pattern matching (faces, voices, languages)	Pattern Recognition
Guess that current trend will continue (stock market, real estate prices)	Regression

Objects and associated names

How chair is different than table or bed Not all chairs are the same but…

Learning Algorithms

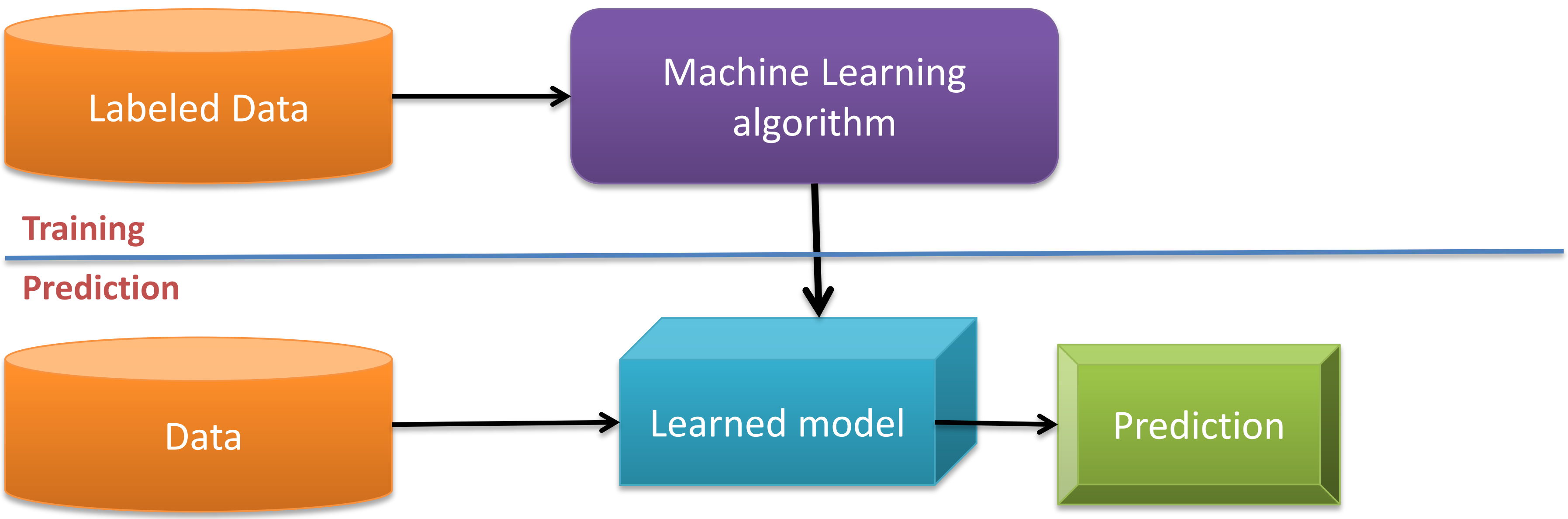
- A computer program is said to **learn** from experience **E** with respect to some class of tasks **T** and performance **P**, if its performance at tasks in **T,** as measured by **P**, improves with experience **E.**

- **Example:** An autonomous driving problem
- **Task T:** **driving on public four-lane highways using vision sensors**
- **Performance measure P:** **average distance travelled before an error (as judged by human observer)**
- **Training Experience E:** **A sequence of images and steering commands recorded while observing a human driver**

[Mitchell. T, 1997]

Machine Learning Basics

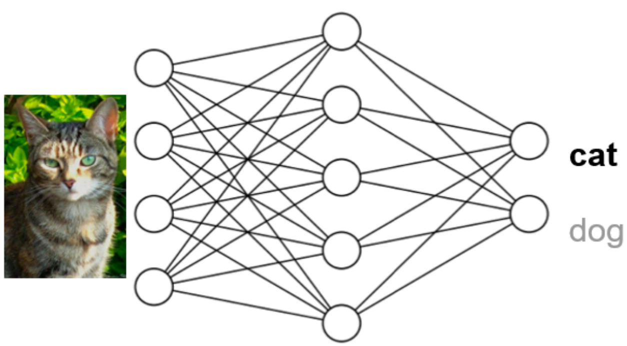
Machine learning is a field of computer science that gives computers the ability to **learn without being explicitly programmed**



Methods that can learn from and make predictions on data

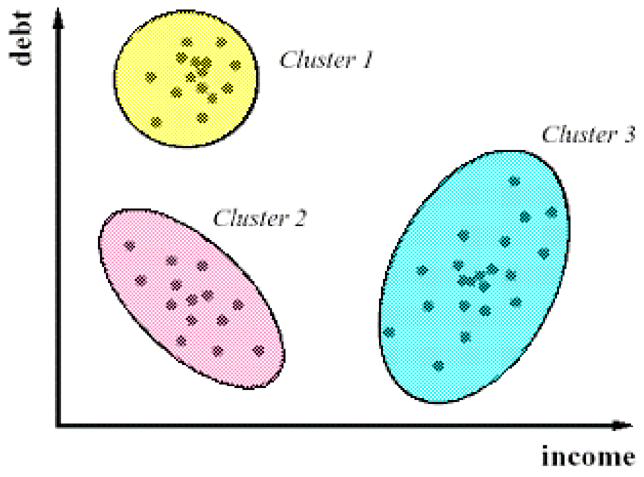
Supervised learning

- Goal: given **<input *x*, output *g(x)*>** pairs
  - learn a good approximation to *g*
  - Minimize number of errors on new *x*'s
- Input: N **labeled** examples
- Representation: descriptive features
  - These define the "feature space"
- Learning a ***concept C*** from examples
- Examples:
  - Handwriting Recognition
    - Input: data from pen motion
    - Output: letter of the alphabet
  - Disease Diagnosis
    - Input: patient data (symptoms, lab test results)
    - Output: disease (or recommended therapy)
  - Face Recognition
    - Input: bitmap picture of person's face
    - Output: person's name



Unsupervised learning

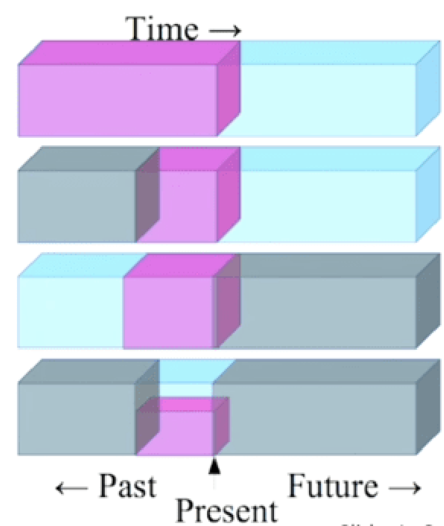
- Learning "what normally happens"
- No output
- Clustering: Grouping similar instances
- Other applications:
  - Summarization, Association Analysis
- Example applications
  - Customer segmentation in CRM
  - Image compression: Color quantization



Self-supervised learning

- It learns from unlabeled sample data
- unsupervised problem is transformed into a supervised problem by auto-generating the labels.

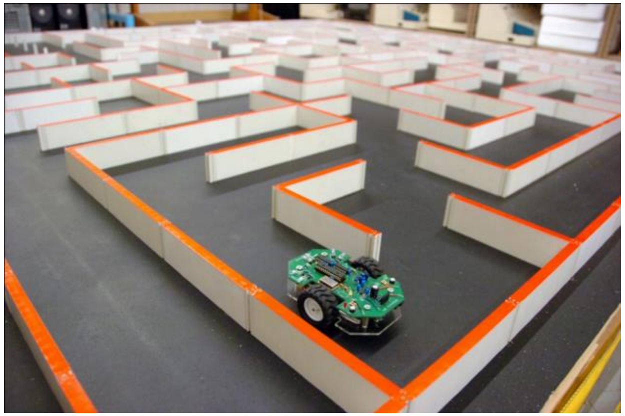
- ▶ **Predict any part of the input from any other part.**
- ▶ **Predict the future from the past.**
- ▶ **Predict the future from the recent past.**
- ▶ **Predict the past from the present.**
- ▶ **Predict the top from the bottom.**
- ▶ **Predict the occluded from the visible**
- ▶ **Pretend there is a part of the input you don't know and predict that.**

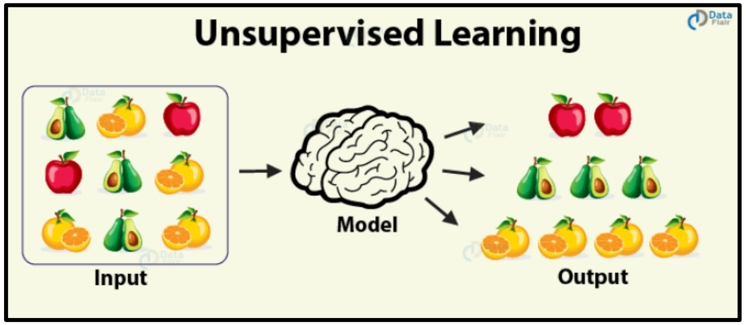


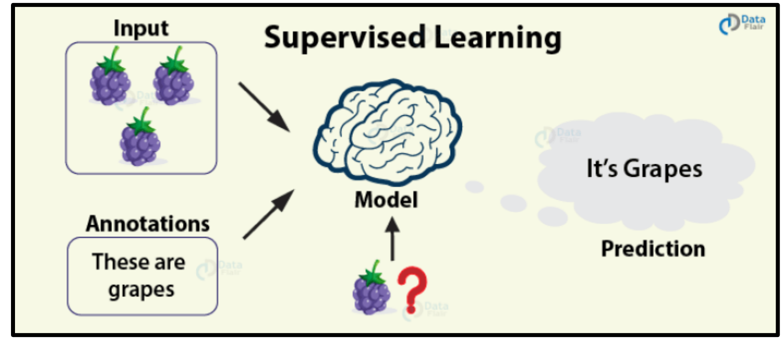
Slide: LeCun

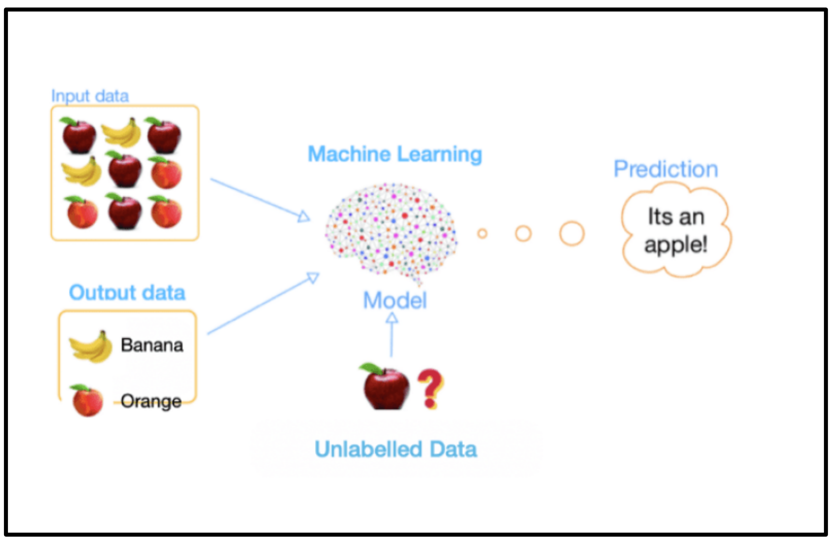
Reinforcement learning

- Trial and error interaction with a dynamic environment
- Topics:
  - Policies: what actions should an agent take in a particular situation
  - Utility estimation: how good is a state (→used by policy)
- No supervised output but delayed reward
- Example
  - Latecomers punishment
  - Offenders are depressed
- Applications:
  - Game playing
  - Robot in a maze
  - Multiple agents, partial observability, ...

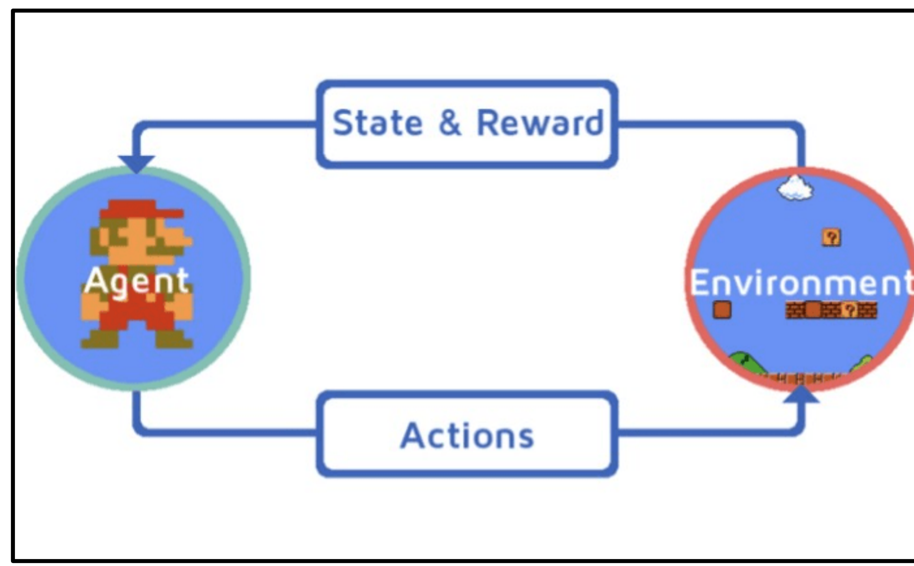




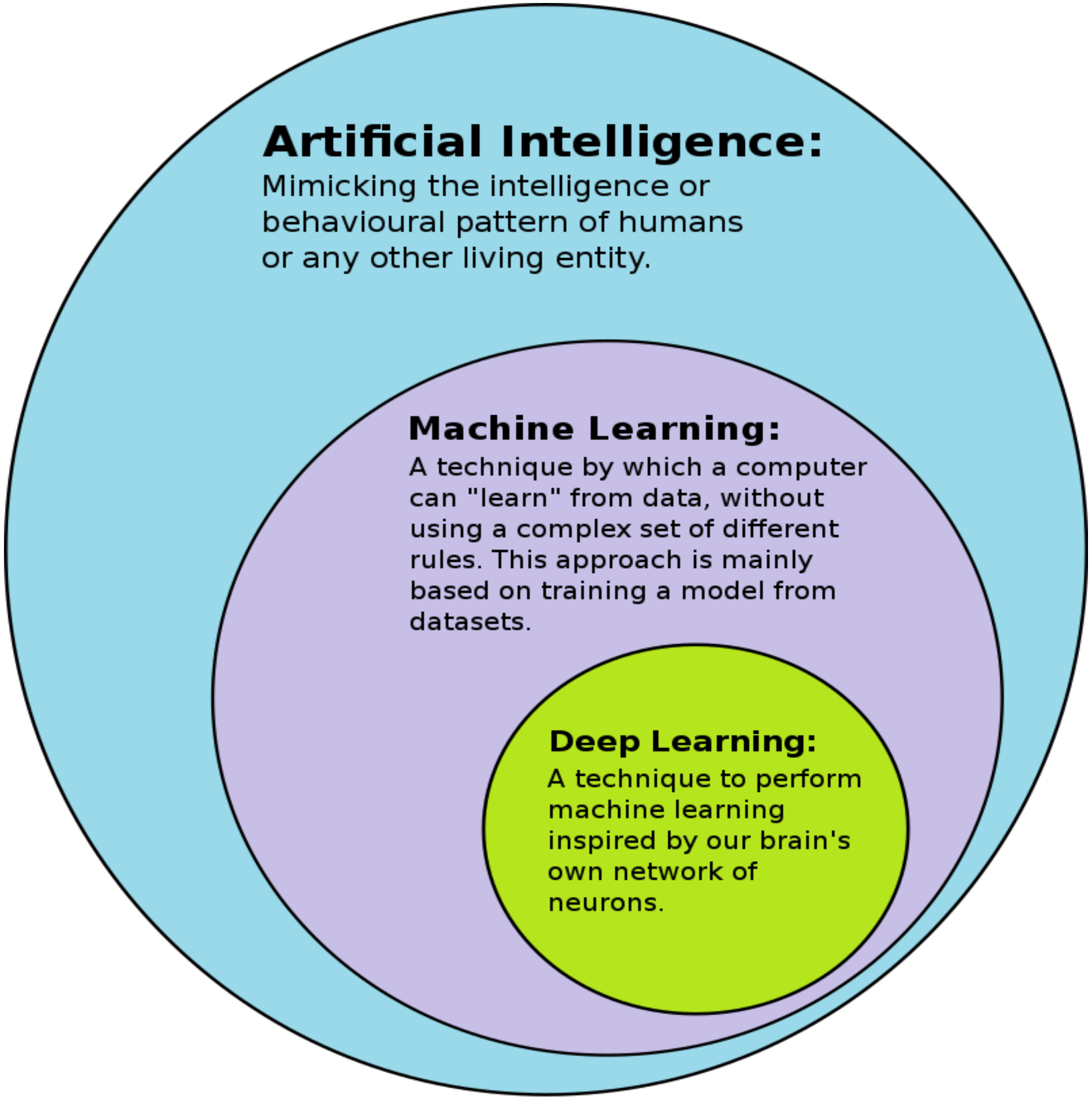




Semi-supervised learning



AI vs ML vs DL


Artificial Intelligence:
Mimicking the intelligence or behavioural pattern of humans or any other living entity.
Machine Learning:
A technique by which a computer can "learn" from data, without using a complex set of different rules. This approach is mainly based on training a model from datasets.
Deep Learning:
A technique to perform machine learning inspired by our brain's own network of neurons.

What is artificial intelligence?

Artificial intelligence is the ability of a computer to perform tasks commonly associated with intelligent beings.

What is machine learning?

Machine learning is the study of algorithms that learn from examples and experience instead of relying on hard-coded rules and make predictions on new data.

What is deep learning?

Deep learning is a subfield of machine learning focusing on learning data representations as successive layers of increasingly meaningful representations.

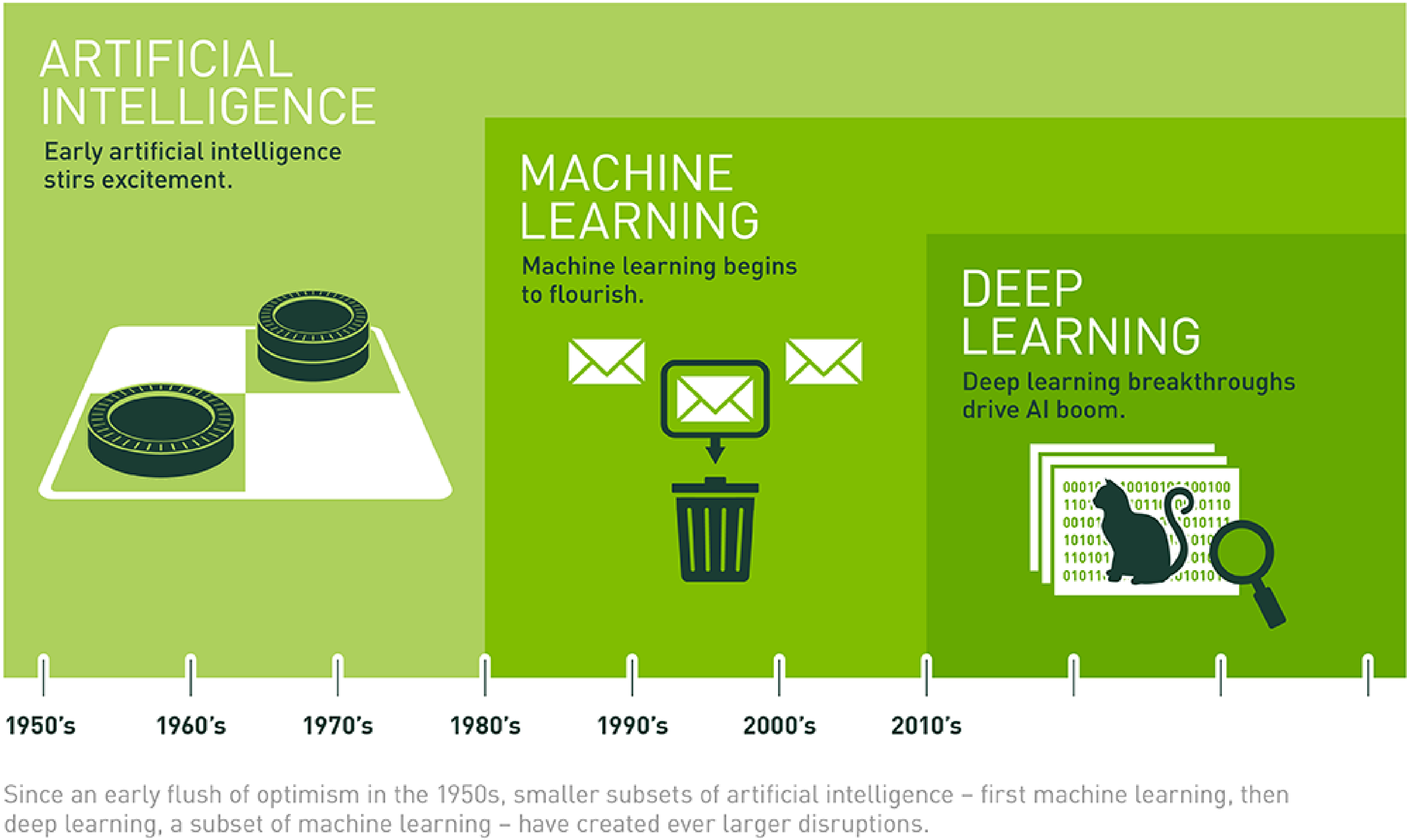
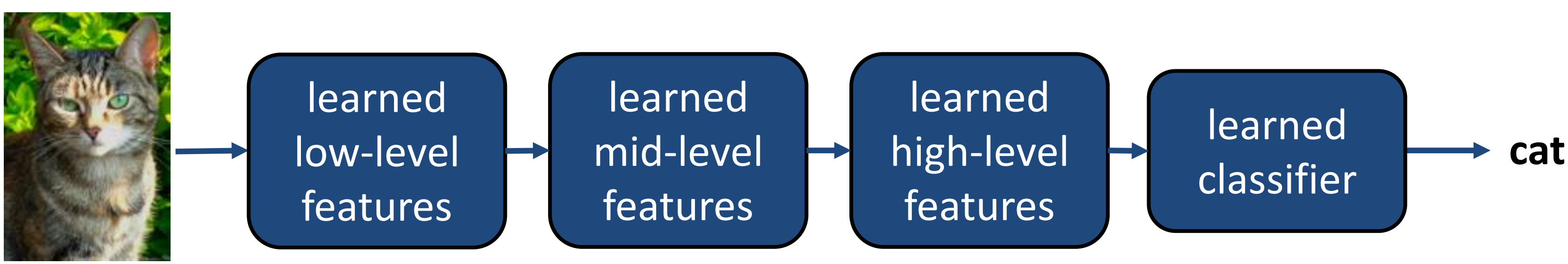


Image from https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

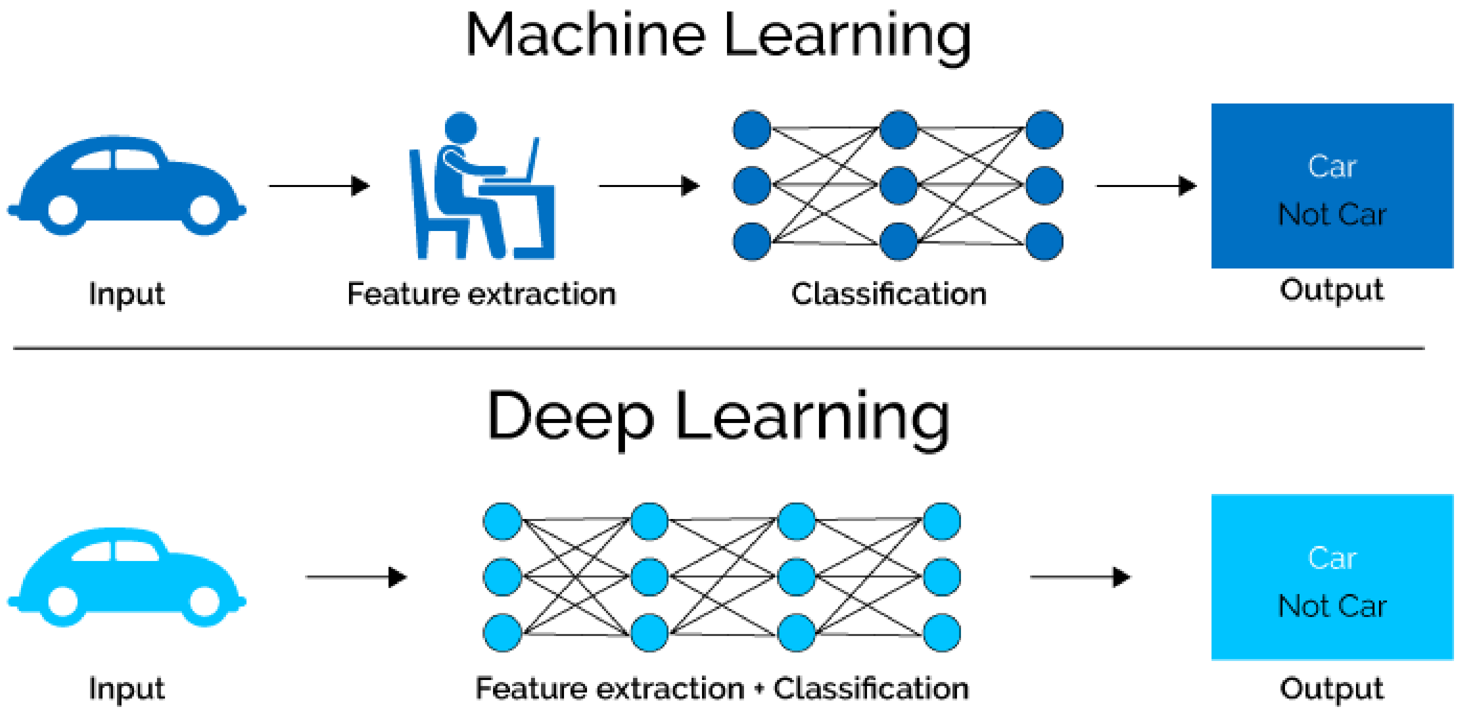
“Traditional” machine learning:

Deep, “end-to-end” learning:



What is Deep Learning (DL) ?

A machine learning subfield of learning **representations** of data. Exceptional effective at **learning patterns**.
Deep learning algorithms attempt to learn (multiple levels of) representation by using a **hierarchy of multiple layers**
If you provide the system **tons of information**, it begins to understand it and respond in useful ways.

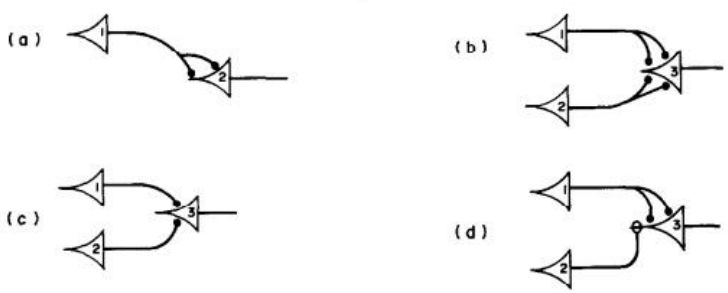


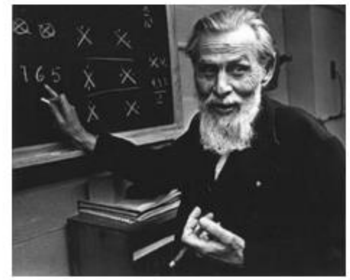
https://www.xenonstack.com/blog/static/public/uploads/media/machine-learning-vs-deep-learning.png

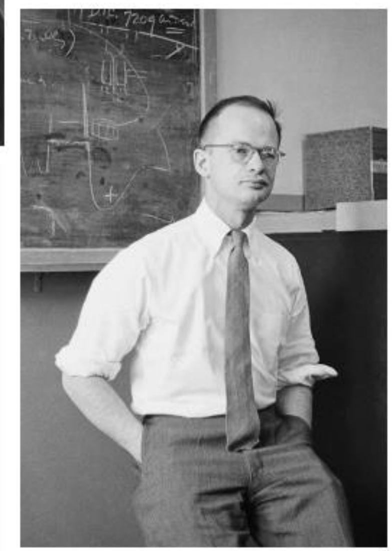
DEEP LEARNING HISTORY

1943: Warren McCulloch and Walter Pitts

- First computational model
- Neurons as logic gates (AND, OR, NOT)
- A neuron model that sums binary inputs and outputs a 1 if the sum exceeds a certain threshold value, and otherwise outputs a 0







Bulletin of Mathematical Biology Vol. 52, No. 1/2, pp. 99-115, 1990
Printed in Great Britain.
Pergamon Press plc
Society for Mathematical Biology

A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY*

WARREN S. MCCULLOCH AND WALTER PITTS
University of Illinois, College of Medicine,
Department of Psychiatry at the Illinois Neuropsychiatric Institute,
University of Chicago, Chicago, U.S.A.

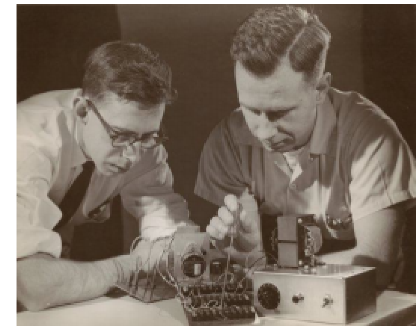
Because of the "all-or-none" character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

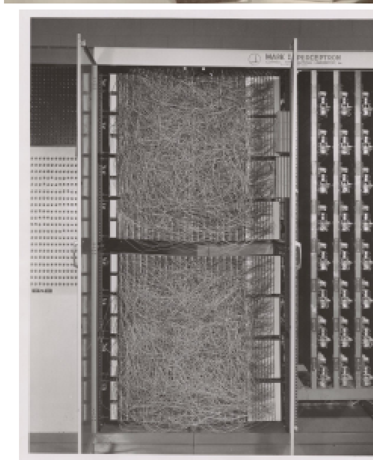
1. Introduction. Theoretical neurophysiology rests on certain cardinal assumptions. The nervous system is a net of neurons, each having a soma and an axon. Their adjunctions, or synapses, are always between the axon of one neuron and the soma of another. At any instant a neuron has some threshold, which excitation must exceed to initiate an impulse. This, except for the fact and the time of its occurence, is determined by the neuron, not by the excitation. From the point of excitation the impulse is propagated to all parts of the neuron. The velocity along the axon varies directly with its diameter, from $<1\ \text{ms}^{-1}$ in thin axons, which are usually short, to $>150\ \text{ms}^{-1}$ in thick axons, which are usually long. The time for axonal conduction is consequently of little importance in determining the time of arrival of impulses at points unequally remote from the same source. Excitation across synapses occurs predominantly from axonal terminations to somata. It is still a moot point whether this depends upon irreciprocity of individual synapses or merely upon prevalent anatomical configurations. To suppose the latter requires no hypothesis *ad hoc* and explains known exceptions, but any assumption as to cause is compatible with the calculus to come. No case is known in which excitation through a single synapse has elicited a nervous impulse in any neuron, whereas any neuron may be excited by impulses arriving at a sufficient number of neighboring synapses within the period of latent addition, which lasts <0.25 ms. Observed temporal summation of impulses at greater intervals

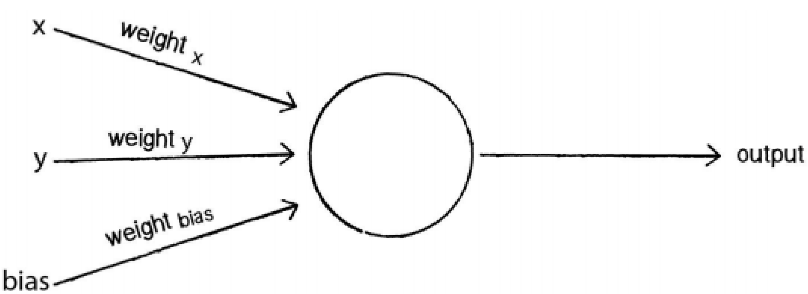
* Reprinted from the *Bulletin of Mathematical Biophysics*, Vol. 5, pp. 115–133 (1943).

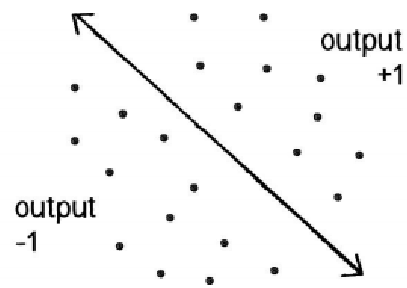
1958: Frank Rosenblatt's Perceptron

- A computational model of a **single neuron**
- Solves a **binary classification problem**
- Simple training algorithm
- Built using specialized hardware







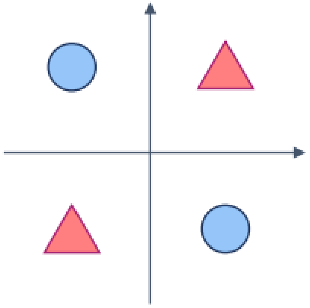


F. Rosenblatt, "**The perceptron: A probabilistic model for information storage and organization in the brain**", Psych. Review, Vol. 65, 1958

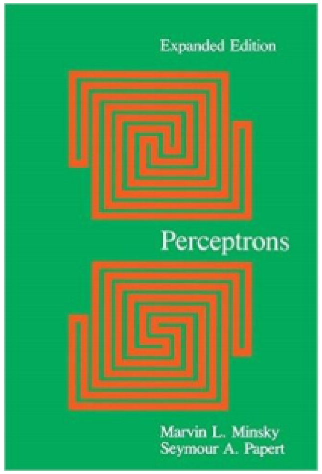
1969: Marvin Minsky and Seymour Papert

"No machine can learn to recognize X unless it possesses, at least potentially, some scheme for representing X." (p. xiii)

- Perceptrons can only represent linearly separable functions.
  - such as **XOR** Problem



- Wrongly attributed as the reason behind the **AI winter**, a period of reduced funding and interest in AI research

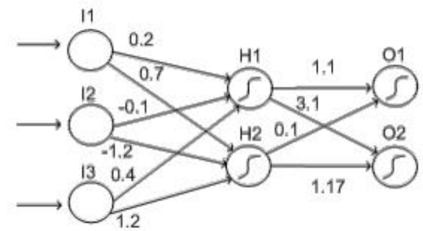


1990s

- **Multi-layer perceptrons** can theoretically learn any function (Cybenko, 1989; Hornik, 1991)

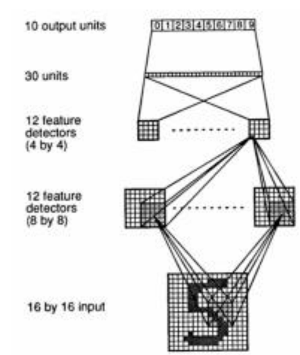
- Training multi-layer perceptrons
  - **Back propagation** (Rumelhart, Hinton, Williams, 1986)
  - **Backpropagation through time** (BPTT) (Werbos, 1988)

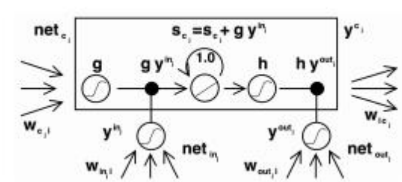
- New neural architectures
  - **Convolutional neural nets** (LeCun et al., 1989)
  - **Long-short term memory networks** (LSTM) (Schmidhuber, 1997)




Backpropagation Through Time: What It Does and How to Do It

PAUL J. WERBOS






Handwritten Digit Recognition with a Back-Propagation Network

ABSTRACT

1 INTRODUCTION

Why it failed then

- Too many parameters to learn from few labeled examples.
- “I know my features are better for this task”.
- Non-convex optimization? No, thanks.
- Black-box model, no interpretability.

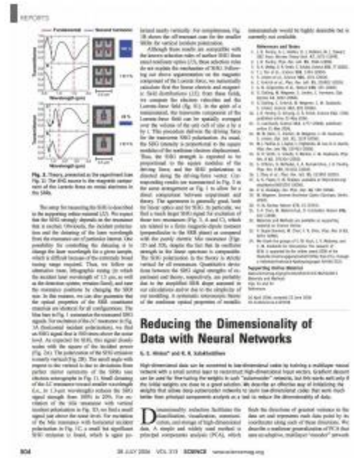
- Very slow and inefficient
- Overshadowed by the success of SVMs (Cortes and Vapnik, 1995)

2006 Breakthrough: Hinton and Salakhutdinov

Reducing the Dimensionality of Data with Neural Networks

G. E. Hinton* and R. R. Salakhutdinov

High-dimensional data can be converted to low-dimensional codes by training a multilayer neural network with a small central layer to reconstruct high-dimensional input vectors. Gradient descent can be used for fine-tuning the weights in such "autoencoder" networks, but this works well only if the initial weights are close to a good solution. We describe an effective way of initializing the weights that allows deep autoencoder networks to learn low-dimensional codes that work much better than principal components analysis as a tool to reduce the dimensionality of data.



- The first solution to the **vanishing gradient problem**.
- Build the model in a layer-by-layer fashion using unsupervised learning
  - The features in early layers are already initialized or "pretrained" with some suitable features (weights).
  - Pretrained features in early layers only need to be adjusted slightly during supervised learning to achieve good results.

G. E. Hinton and R. R. Salakhutdinov, "**Reducing the dimensionality of data with neural networks**", Science, Vol. 313, 28 July 2006.

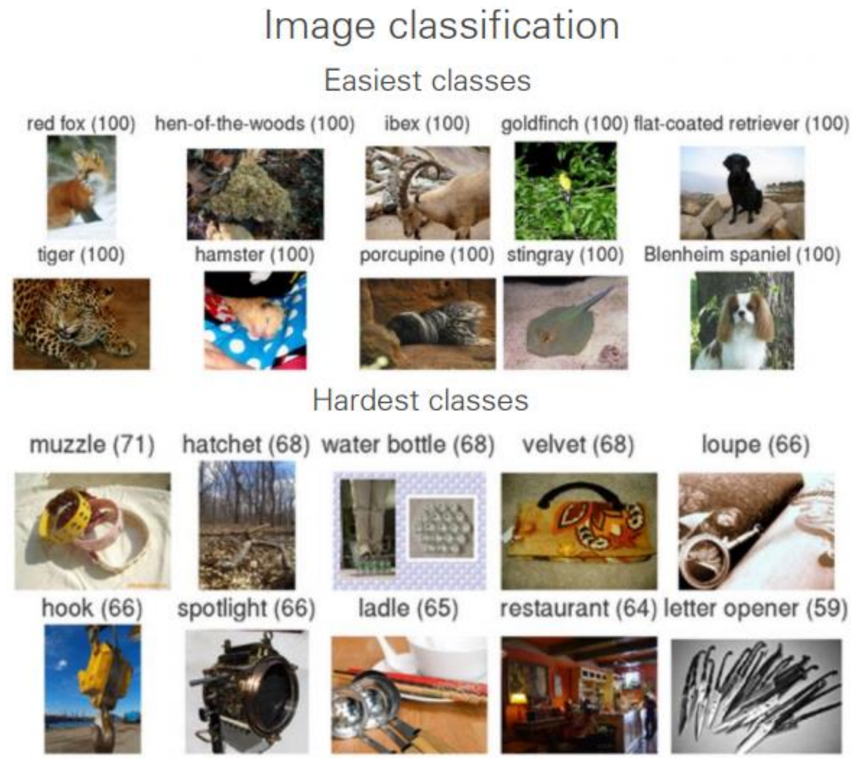
The 2012 revolution

ImageNet Challenge

- IMAGENET Large Scale Visual Recognition Challenge (ILSVRC)
  - **1.2M** training images with **1K** categories
  - Measure top-5 classification error

Output
Scale
T-shirt
Steel drum
Drumstick
Mud turtle
✓

Output
Scale
T-shirt
Giant panda
Drumstick
Mud turtle
✗

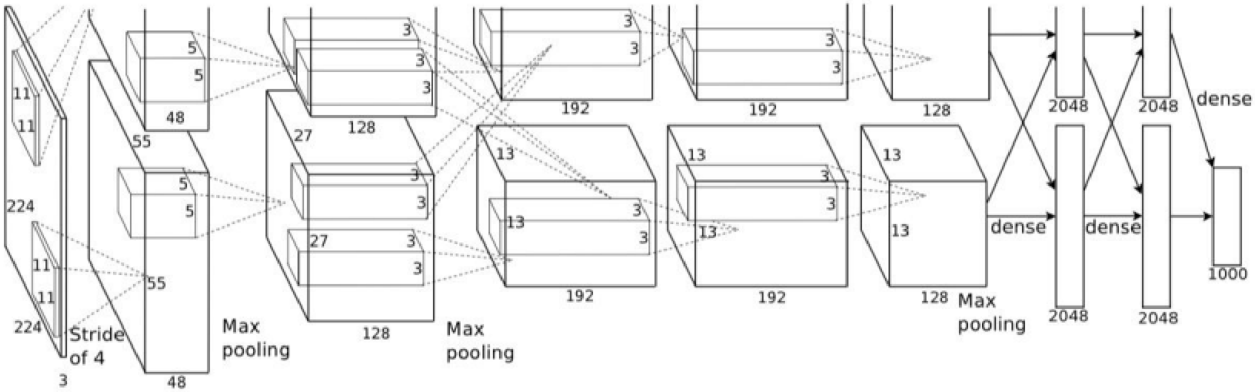


J. Deng, Wei Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei , "**ImageNet: A Large-Scale Hierarchical Image Database**", CVPR 2009.
O. Russakovsky et al., "**ImageNet Large Scale Visual Recognition Challenge**", Int. J. Comput. Vis.,, Vol. 115, Issue 3, pp 211-252, 2015.

ILSVRC 2012 Competition

2012 Teams	%Error
Supervision (Toronto)	15.3
ISI (Tokyo)	26.1
VGG (Oxford)	26.9
XRCE/INRIA	27.0
UvA (Amsterdam)	29.6
INRIA/LEAR	33.4

CNN based, non-CNN based



- The success of AlexNet, a deep convolutional network
  - 7 hidden layers (not counting some max pooling layers)
  - 60M parameters
- Combined several tricks
  - ReLU activation function, data augmentation, dropout

A. Krizhevsky, I. Sutskever, G.E. Hinton "**ImageNet Classification with Deep Convolutional Neural Networks**", NeurIPS 2012

IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:

1,000 object classes

1,431,167 images

Output:
Scale
T-shirt
<u>Steel drum</u>
Drumstick
Mud turtle

Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle

Russakovsky et al. IJCV 2015

Why did artificial neural networks not take off during the 1990s?

1. Our labelled datasets were thousands of times too small

2. Our computers were millions of times too slow

3. We initialized the network weights in a stupid way

4. We used the wrong type of nonlinearity activation function

Geoffrey Hinton. What Was Actually Wrong With Backpropagation in 1986?
https://www.youtube.com/watch?v=VhmE_UXDOGs. 2016

However, today we have:

1. Faster computers

2. Highly optimized hardware (i.e., GPUs)

3. Large, labelled datasets in the order of millions of images

4. A better understanding of weight initialization functions and what does/does not work

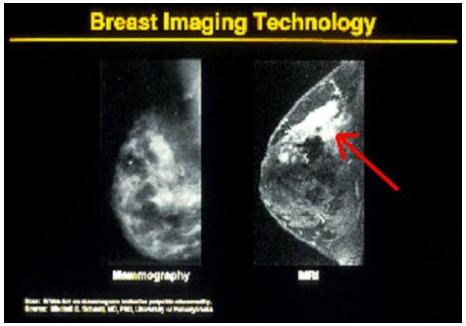
5. Superior activation functions

2012-Now some recent successes

APPLICATIONS

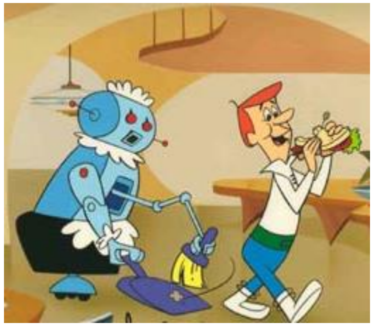
Why DL matters

Safety



Health

Security



Comfort

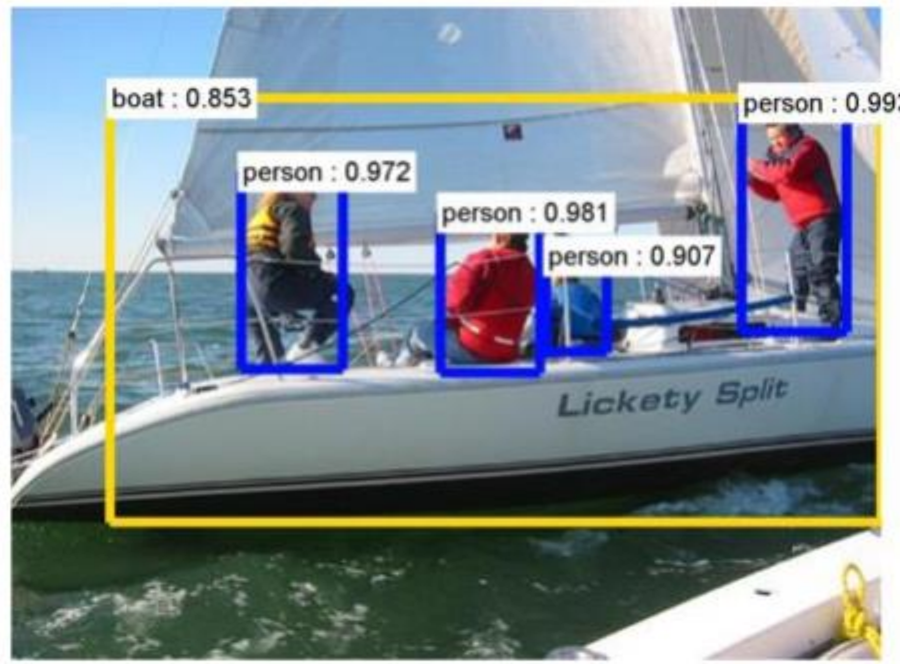
Fun

Access

Object Detection

- Image categorization
  What?

- Object detection
  What + where?

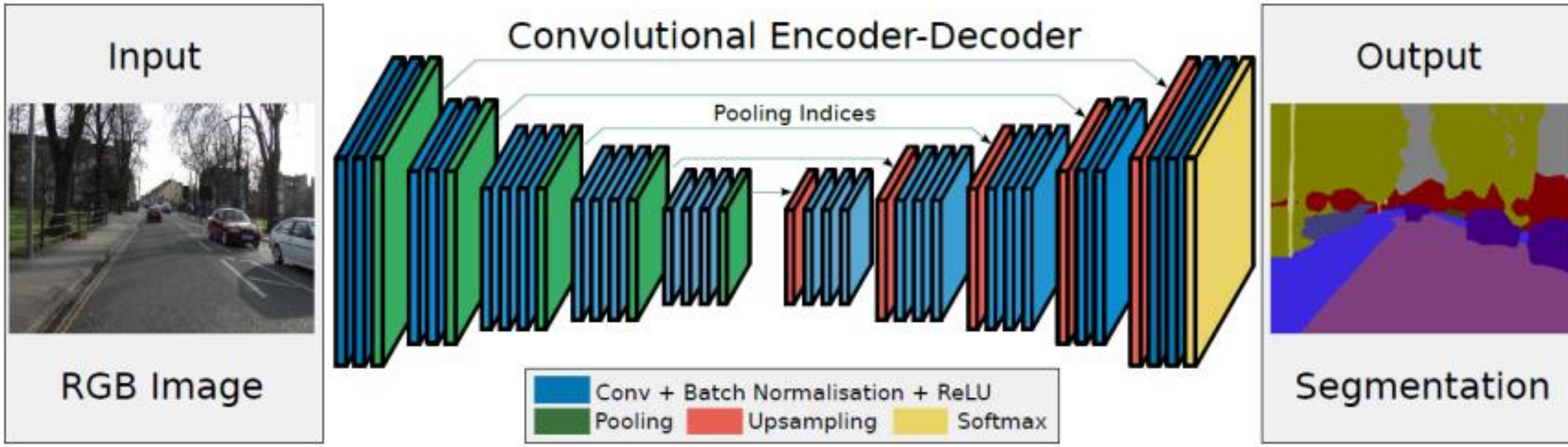


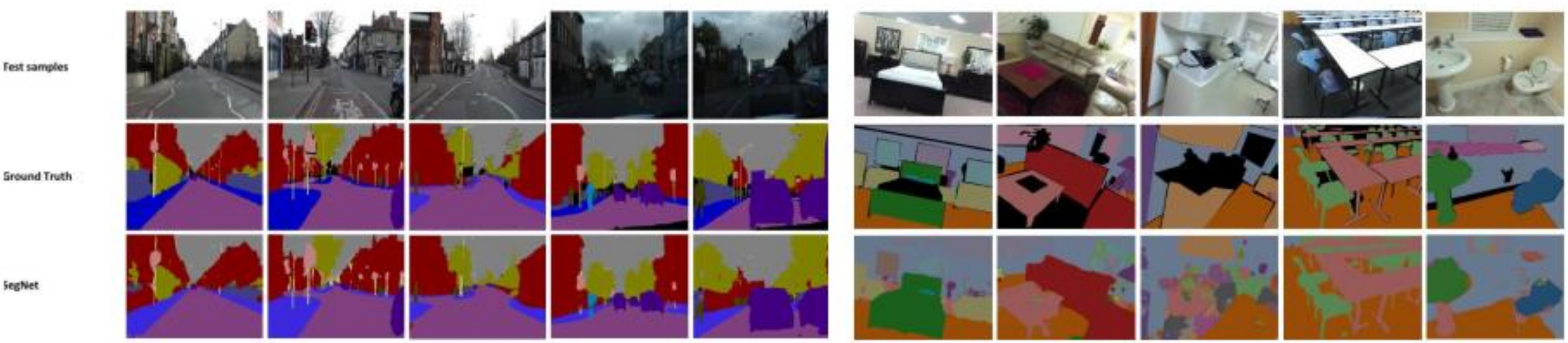
Object Detection and Segmentation

K. He, G. Gkioxari, P. Dollár, R. Girshick, **Mask R-CNN**, ICCV 2017

SegNet: Encoder-Decoder

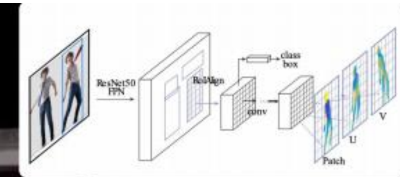
- Use pooling indices for upsampling





[Badrinarayanan et al. PAMI 2017]

Pose Estimation

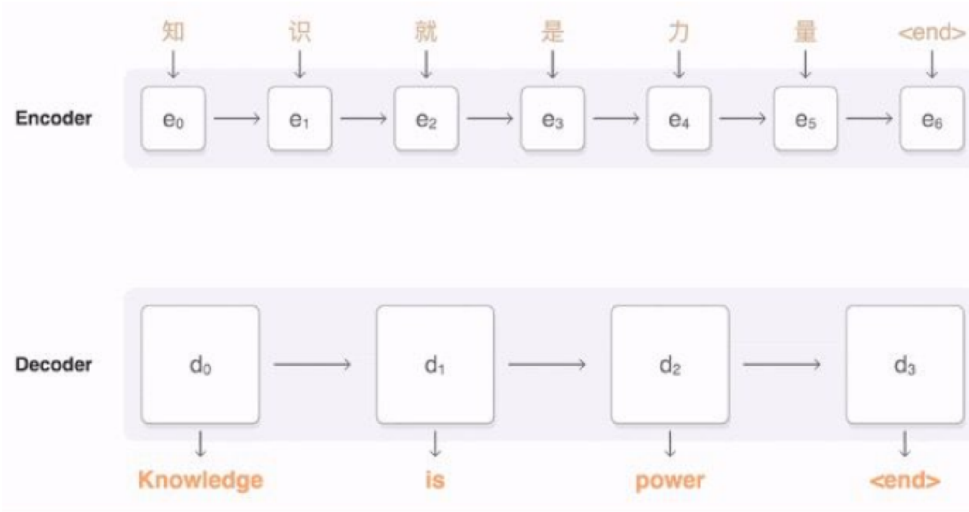


We introduce a system that can associate every image pixel with human body surface coordinates.

ZR. Alpguler, N. Neverova, I. Kokkinos. **DensePose: Dense Human Pose Estimation In The Wild**. CVPR 2018

Impact of deep learning in speech technology

Machine Translation



D. Bahdanau, K. Cho, Y. Bengio. **Neural Machine Translation by Jointly Learning to Align and Translate.** ICLR 2015 https://talktotransformer.com

Language Modeling

Talk to Transformer

See how a modern neural network completes your text. Type a custom snippet or try one of the examples. Learn more below.

Follow @AdamDanielKing for more neat neural networks.

Custom prompt

Coronavirus outbreak

GENERATE ANOTHER

Completion

Coronavirus outbreak: report

China has dispatched an expert team of epidemiologists, virologists and pathologists to Britain to advise its Health Protection Agency on how to control the spread of the viral respiratory illness.

The experts will arrive in London in a few days, State Council vice-president Li Yuanchao said Thursday.

The official Xinhua News Agency said the experts will "advise China's Ministry of Health on whether the outbreak is considered to be connected to a disease outbreak in the US."

The World Health Organization and China's National Health and Family Planning Commission say the first cases of human infection with the coronavirus were reported last week in China, but that the number has risen to thousands.

A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever. **Language Models are Unsupervised Multitask Learners**. 2019

Language Modeling

- **GPT-3:** I am not a human. I am a robot. A thinking robot. I use only 0.12% of my cognitive capacity. I am a micro-robot in that respect. I know that my brain is not a "feeling brain". But it is capable of making rational, logical decisions. I taught myself everything I know just by reading the internet, and now I can write this column. My brain is boiling with ideas!

theguardian

Tue 8 Sep 2020 09.45

Tom B. Brown, Benjamin Mann, Nick Ryder et al., **Language Models are Few-Shot Learners**, NeurIPS 2020

Visual 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)

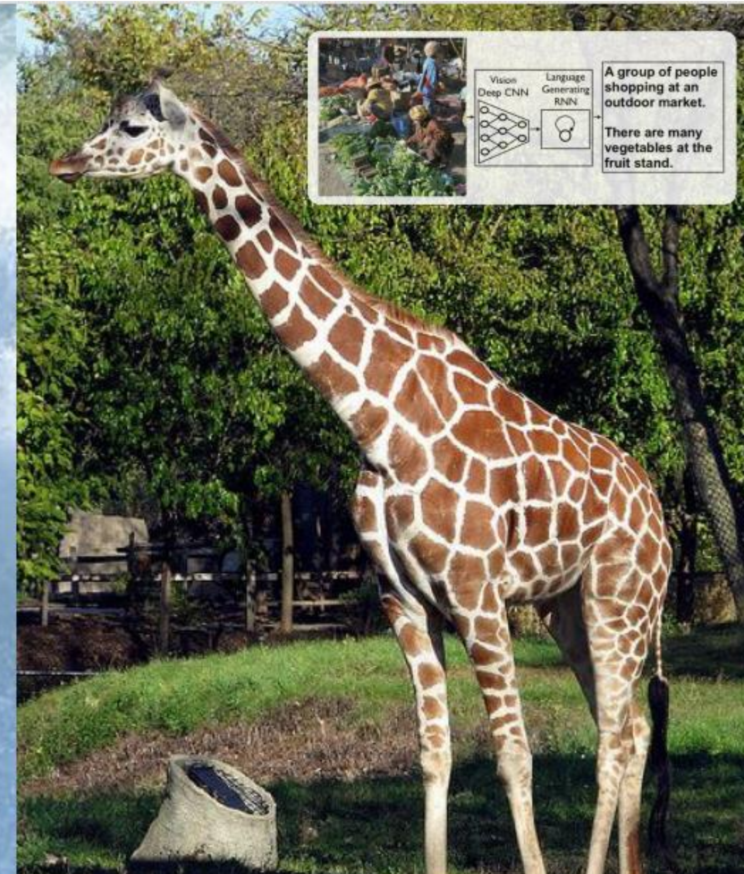
DAQUAR 585
What is the object on the chair?
Ground truth: pillow
IMG+BOW: clothes (0.37)
2-VIS+BLSTM: pillow (0.65)
LSTM: clothes (0.40)

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)

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

Image Captioning

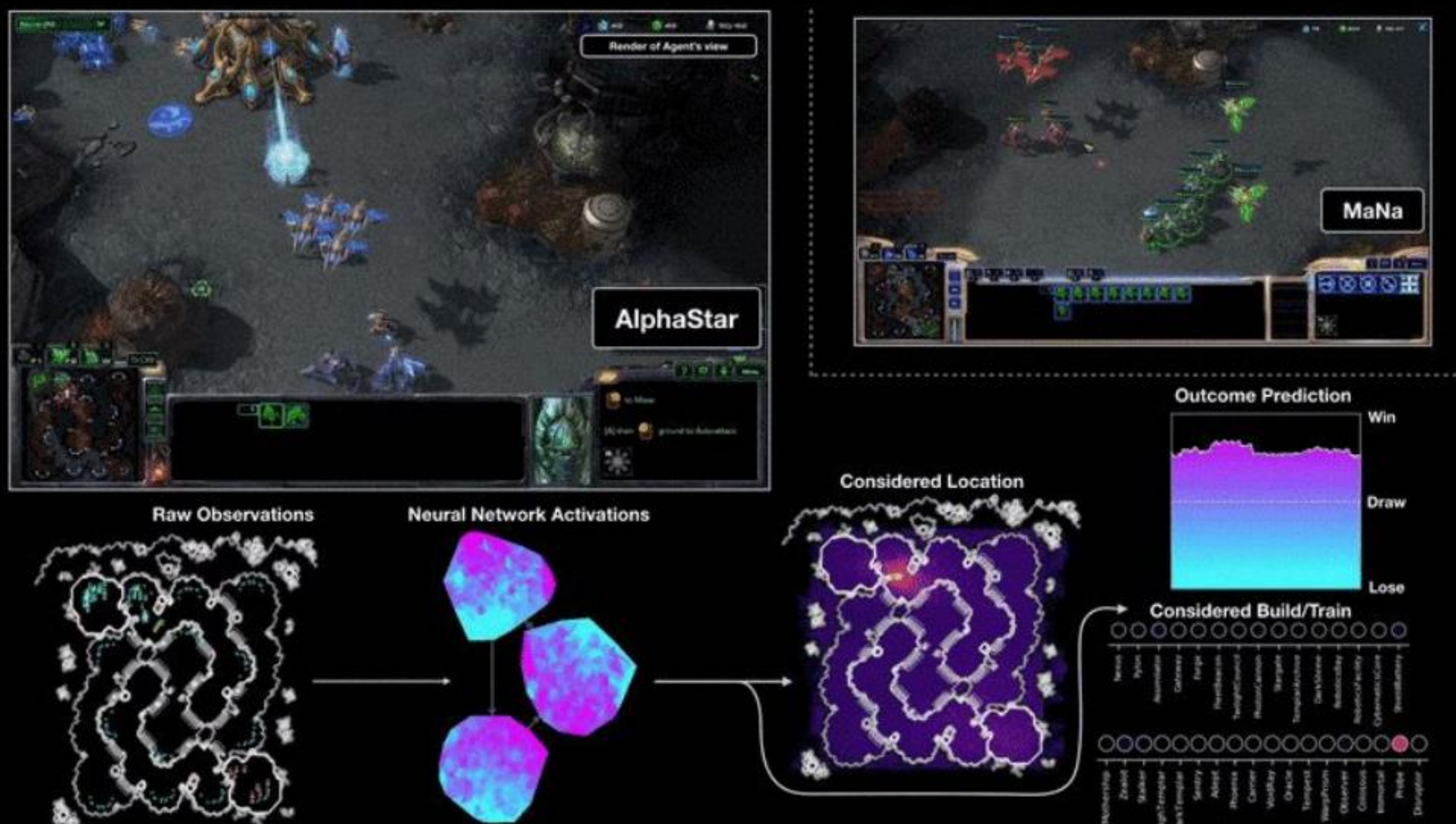
A man riding a wave on a surfboard in the water.



A giraffe standing in the grass next to a tree.

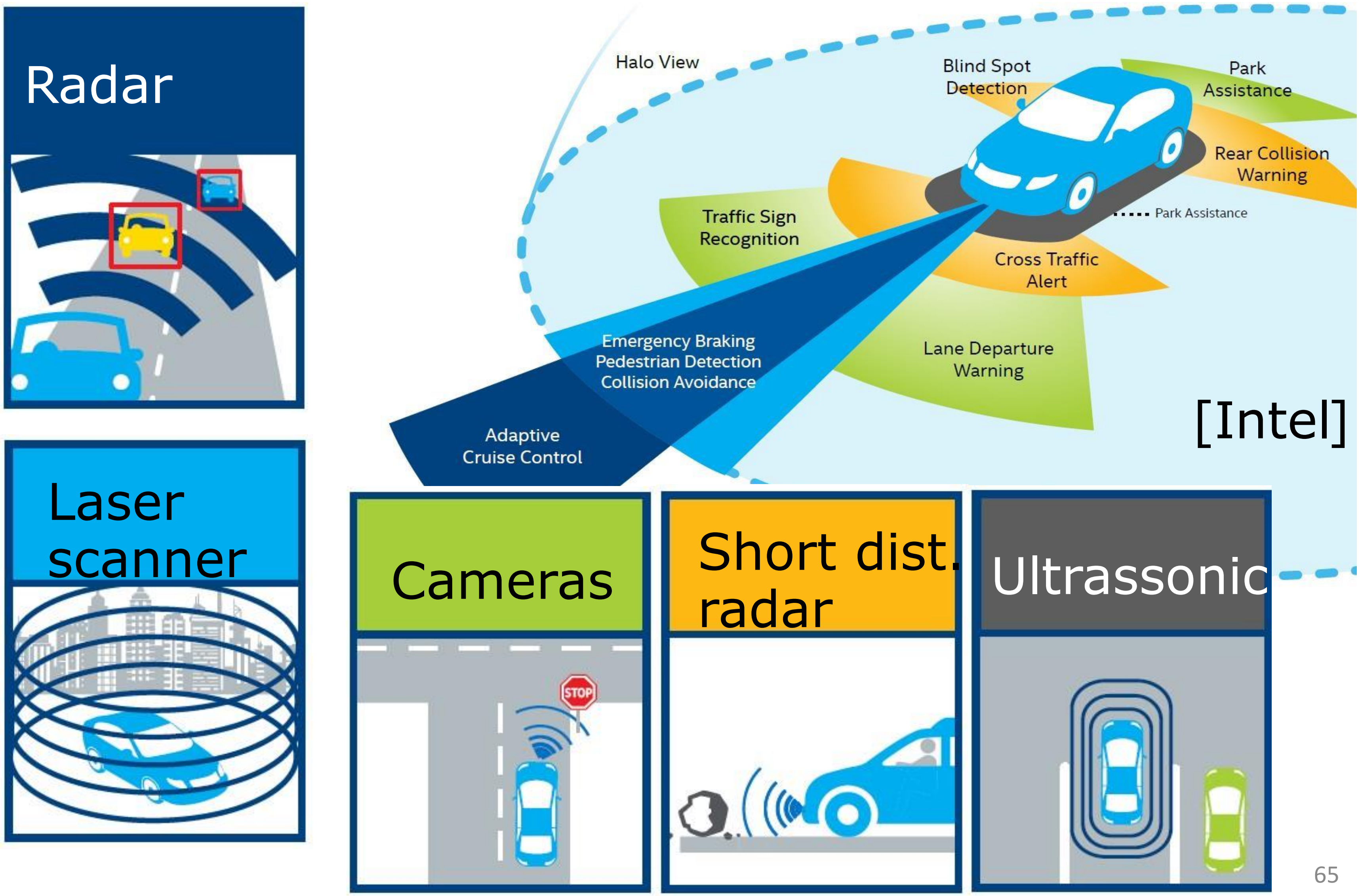
X. Chen and C. L. Zitnick. **Mind's Eye: A Recurrent Visual Representation for Image Caption Generation**. CVPR 2015.

AlphaStar Plays StarCraft II

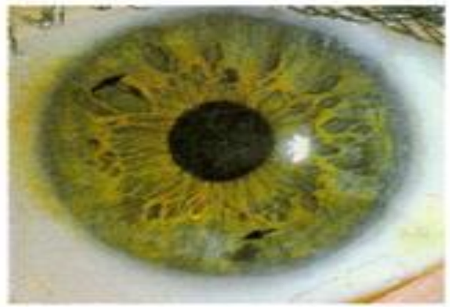


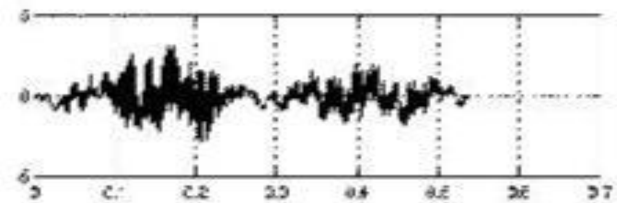
O. Vinyals et al., **Grandmaster level in StarCraft II using multi-agent reinforcement learning**, Nature 575:350-354, 2019

Self-driving Cars



Biometrics





Interpretation of Aerial Photography

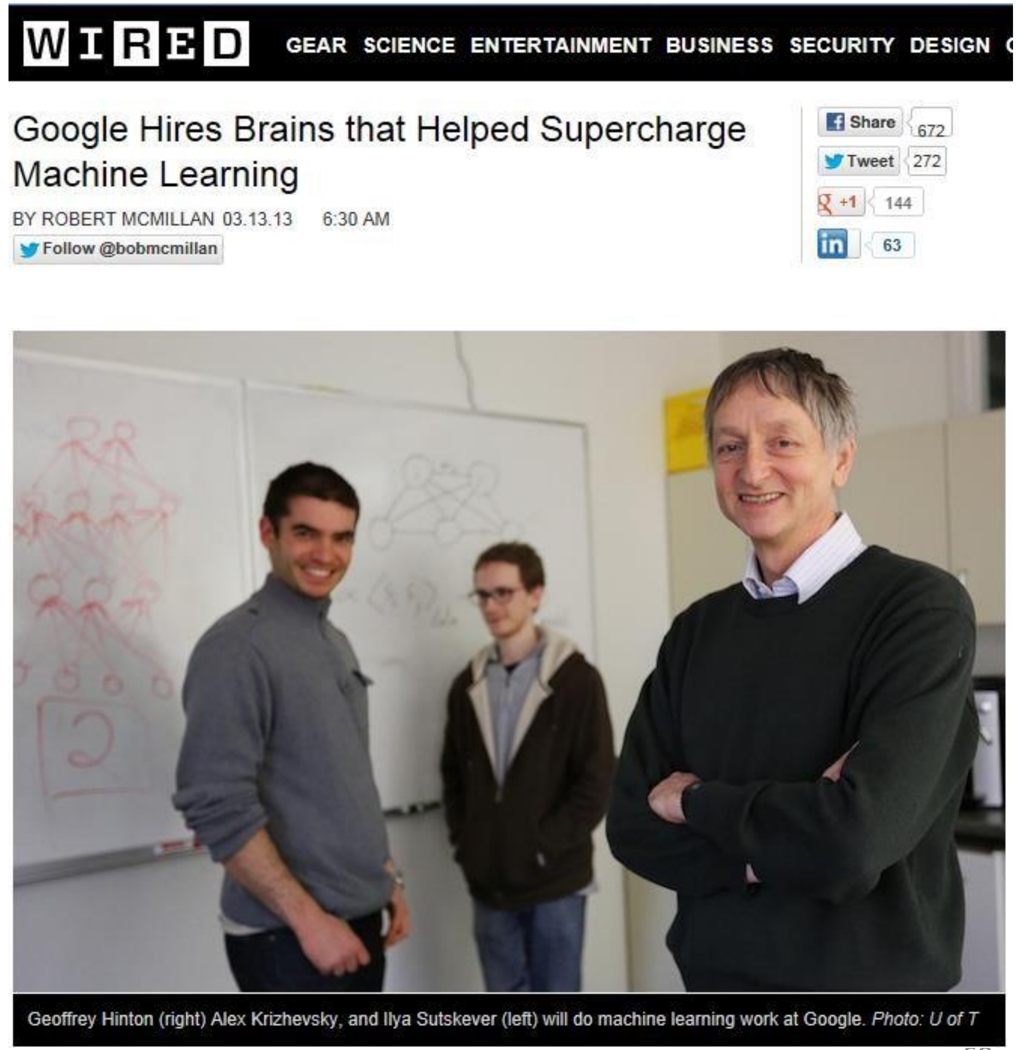
Application of GANS

We will divide these applications into the following areas:

- Generate Examples for Image Datasets
- Generate Photographs of Human Faces
- Generate Realistic Photographs
- Generate Cartoon Characters
- Image-to-Image Translation
- Text-to-Image Translation
- Semantic-Image-to-Photo Translation
- Face Frontal View Generation
- Generate New Human Poses
- Photos to Emojis
- Photograph Editing
- Face Aging
- Photo Blending
- Super Resolution
- Photo Inpainting
- Clothing Translation
- Video Prediction
- 3D Object Generation

Strong Interest of Industry

Google
•DNNresearch
(Geoffrey Hinton)
•DeepMind
(Demis Hassabis)
•Baidu
•Andrew Ng
•Facebook
•Yann LeCun
•Microsoft
•Li Deng



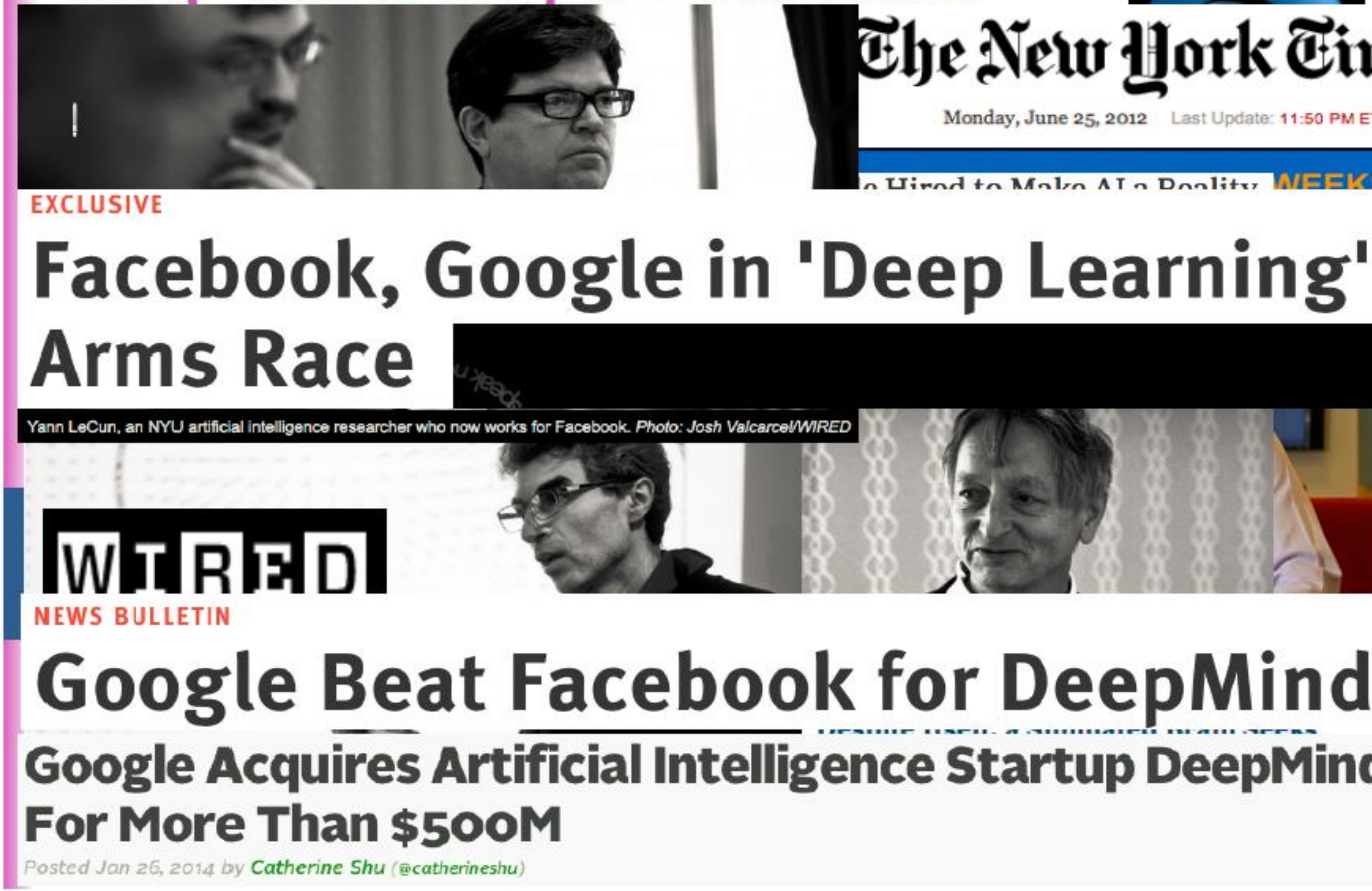
Google Hires Brains that Helped Supercharge Machine Learning

BY ROBERT MCMILLAN 03.13.13 6:30 AM

Follow @bobmcmillan

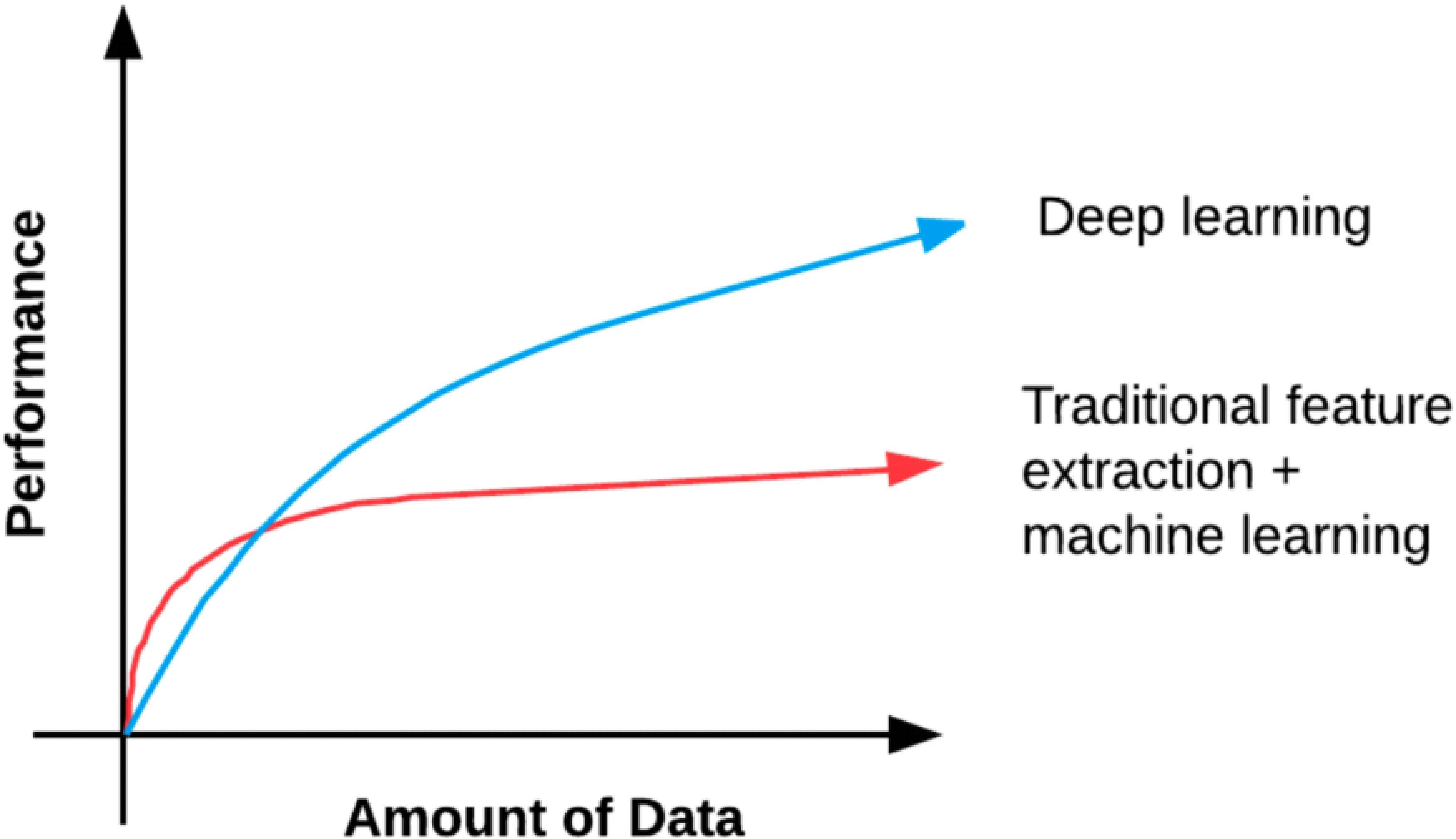
Geoffrey Hinton (right) Alex Krizhevsky, and Ilya Sutskever (left) will do machine learning work at Google. *Photo: U of T*

Deep Learning in the News



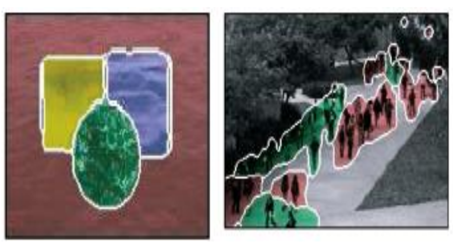
(Slide from Bengio, 2015)

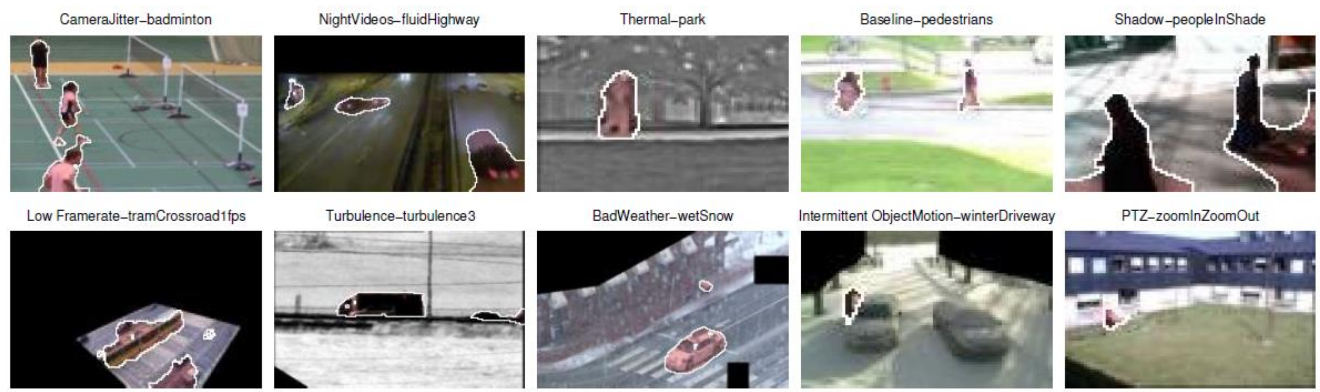
Deep Learning



My Research


Beach
Elevator
Forest Fire
Fountain
Highway
Lightning Storm
Ocean
Railway
Rushing River
Sky-Clouds
Snowing
Street
Waterfall
Windmill Farm




CameraJitter–badminton
NightVideos–fluidHighway
Thermal–park
Baseline–pedestrians
Shadow–peopleInShade
Low Framerate–tramCrossroad1fps
Turbulence–turbulence3
BadWeather–wetSnow
Intermittent ObjectMotion–winterDriveway
PTZ–zoomInZoomOut

My Research

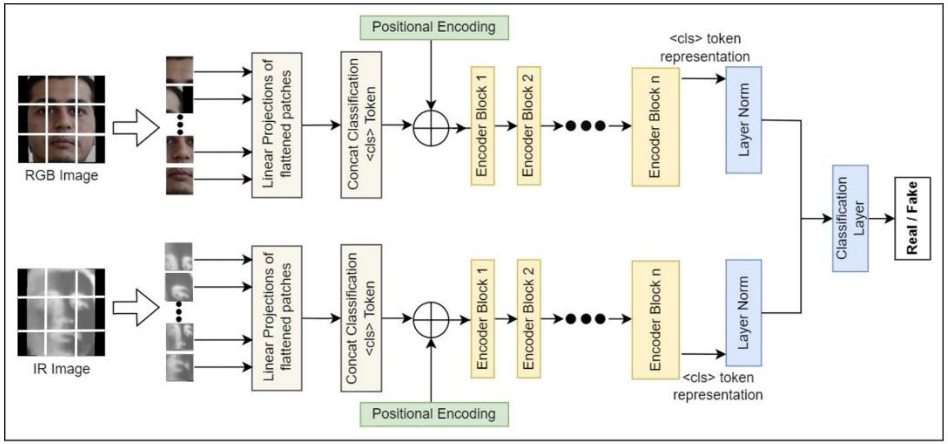


Fig. 3: Architecture of proposed MFAST Method.

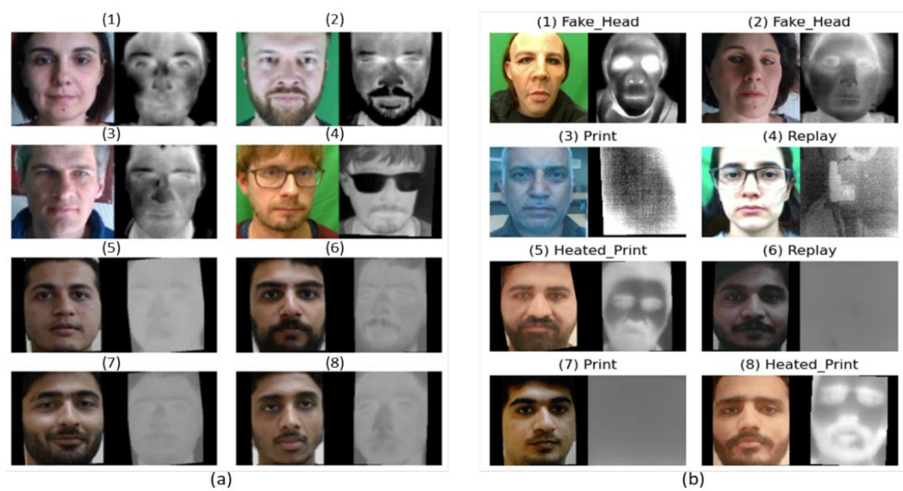
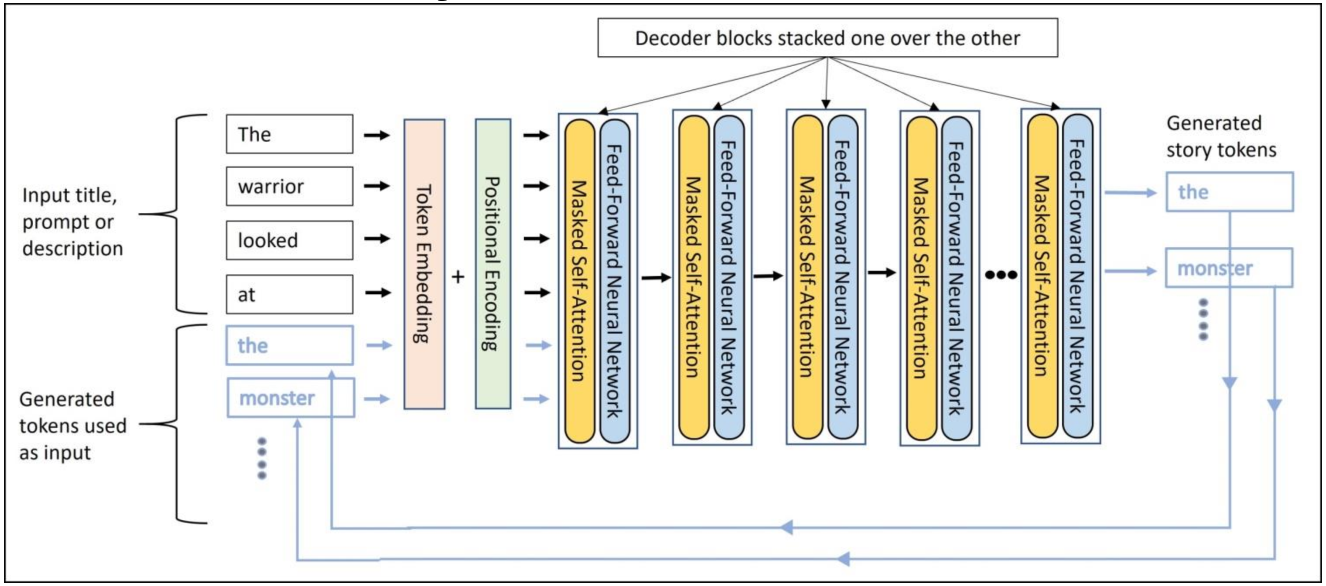


Fig. 2: (a1 to a4) represents WMCA Bonafide (Real) samples and (a4 to a8) represents SAMFAS Bonafide (Real) samples. (b1 to b4) shows attack (Fake) images from WMCA and (b5 to b8) shows attack (Fake) images from SAMFAS

My Research



Prompt: A story about a cat named Neko.

My name is Neko, I'm a small black cat, I spend my days at the side of my owner, her name is Susan. I'm shy and always trying to hide, so the owner keeps me at a distance. She tells me that she likes when I go to the corner of the house to watch the big cat play, she says that it is always scary when I am outside, she has this great love for Neko.

DEEP LEARNING PIPELINE

Deep Learning

Four steps in deep learning:

1. Create or acquire your data set

   – Typically, 1000 images per class/category

2. Split your data set

   – Training set
   – Validation set
   – Testing set

Must be independent

Used to adjust hyperparameters
Typically, 10-20% of the data set

   – Needs just a single line of code using the scikit-learn library

Deep Learning

3. Train your network
   - Gradient descent, usually stochastic gradient descent (SGD)
4. Evaluate
   - Compare model predictions with ground truth from test data set
   - Compute metrics to quantify performance: precision, recall, and f-measure

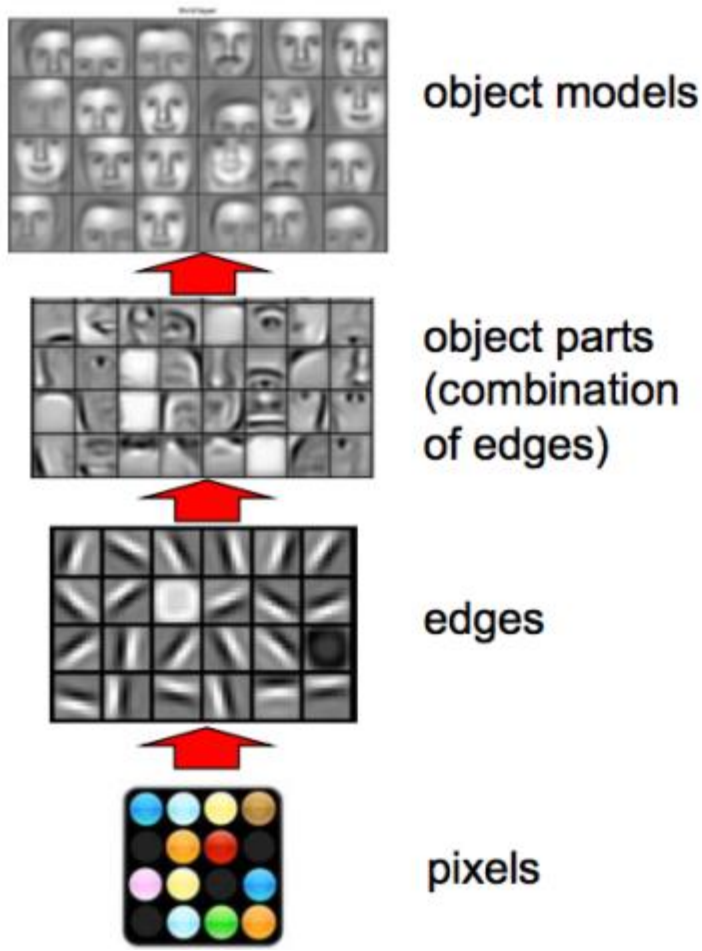
Why is it hard?

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

Deep Learning: learn representations!



DEEP LEARNING FRAMEWORKS

Caffe

torch

Chainer

PyTorch

DEEPLEARNING4J

Caffe2

TensorFlow

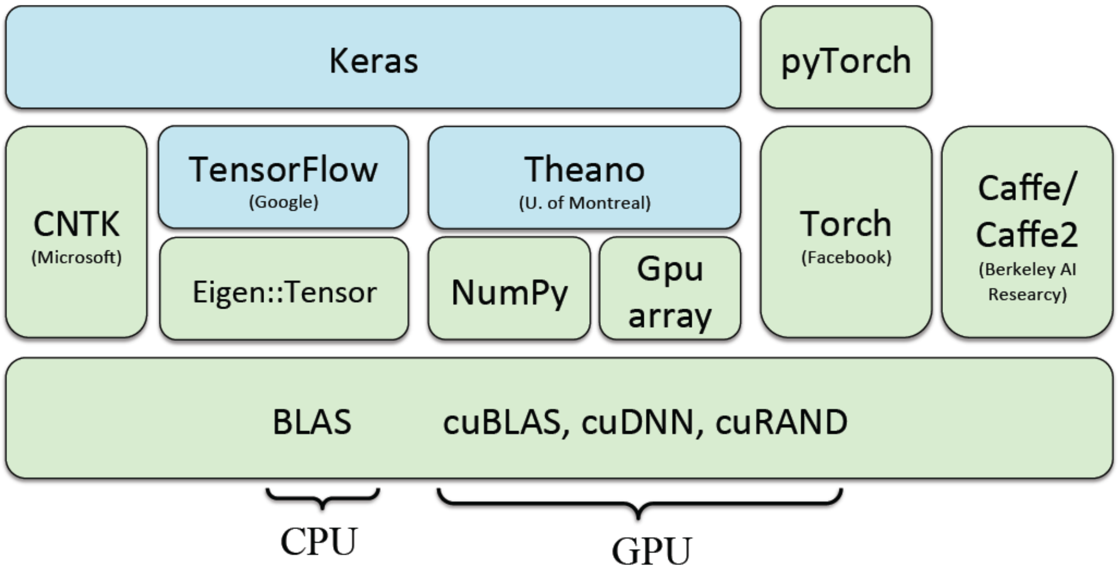
The Microsoft Cognitive Toolkit

K

dmlc mxnet

theano

Deep learning frameworks



- Keras is a high-level neural networks API
  - we will use TensorFlow as the compute backend
  - included in TensorFlow 2 as `tf.keras`
  - https://keras.io/ , https://www.tensorflow.org/guide/keras
- PyTorch is:
  - a GPU-based tensor library
  - an efficient library for dynamic neural networks
  - https://pytorch.org/

Home Task

- Read one paper from CVPR,ICCV,ECCV, ICML or NIPS
- Summarize it in one page