ARTIFICIAL INTELLIGENCE

Alexander Schindler
Scientist
Information Management
Center For Digital Safety & Security

AIT Austrian Institute Of Technology Gmbh
Alexander.schindler@ait.ac.at

DR. ALEXANDER SCHINDLER

- **Scientist @ AIT**
  - Multi-Modal Machine Learning
  - Center for Safety and Security
- **Project-Assistant - TU-Wien (part-time)**
  - Research
  - Teaching
    - Data Science
    - Information Retrieval
    - Intelligent Audio and Music Analysis
- **Research Interests**
  - Audio / Music Analysis
  - Audio-Visual Analysis
  - Machine Learning / Deep Learning
  - Artificial Intelligence
- **Event Organization**
  - Vienna Deep Learning Meetup
  - AI-Summit 2017
  - Ethics & Bias in AI 2018
  - WeAreDevelopers AI Congress 2018 (Partner)
  - Tutorials on Deep Learning (ML-Prague 2018, ISMIR 2018)
  - Int. Workshop on Music Speech and Mind (SMM2019, Vienna)

OUTLINE

- Introduction to Artificial Intelligence
- Introduction to Deep Learning
- Practical Session: How to code a Deep Learning experiment in Python?

QUESTION

- What is an Artificial Intelligence?
- What do you imagine under an artificial intelligence?
- Who knows an Artificial Intelligence?

BEST AI IN TV AND MOVIES

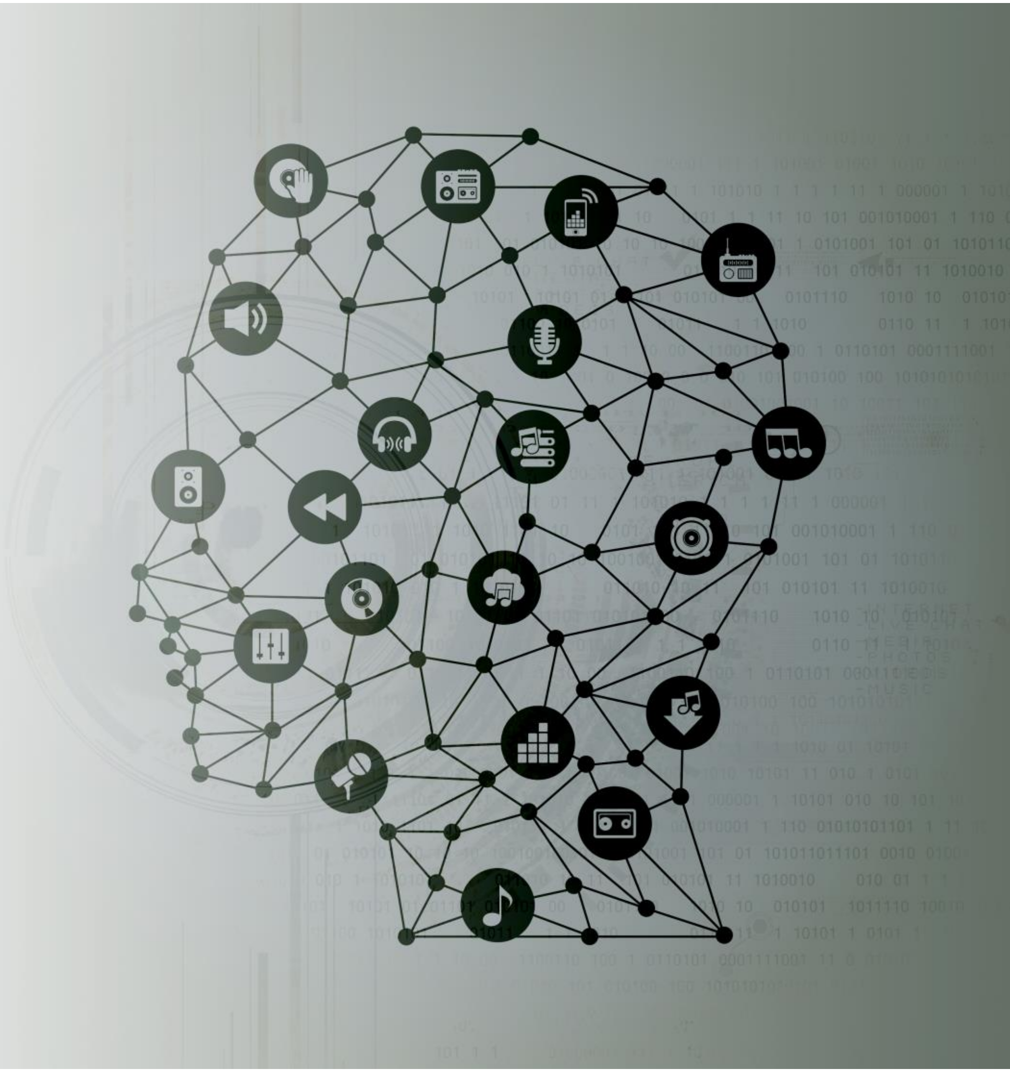
BEST AI IN TV AND MOVIES

- **Self-Driving**
- **Listens, understands & answers**
  - Including sarcasm / is sarcastic
  - Sense of humor / tells jokes
- **Sees, hears, „feels“**
- **Plans ahead**
  - Roads, actions, complex, interactions
- **Has human emotions**
  - Is offended, easily bruised, jealuous
  - Interprets human interaction
- **Feels social responsibility**
  - start a union for cars
- **Self-awareness**
- **Common knowledge**
- **Recommends**
  - Music, books, social events, …

AUSTRIAN INSTITUTE OF TECHNOLOGY

Introduction to

Artificial Intelligence



REAL INTELLIGENCE

Definition

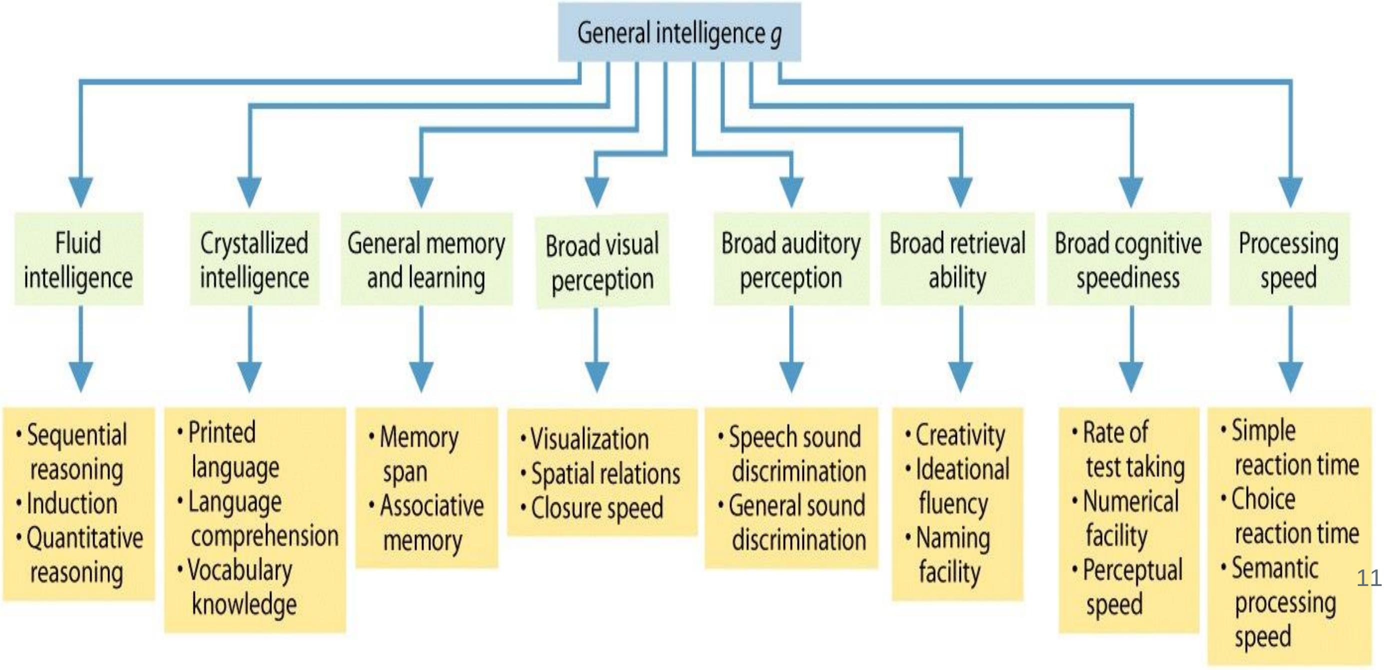
WHAT IS INTELLIGENCE?

Intelligence is an **inferred process** that humans use to explain the different degrees of adaptive success in people’s behaviour

- The mental **abilities** that enable one to **adapt** to, **shape**, or **select** one’s **environment**
- The ability to **judge**, **comprehend**, and **reason**
- The ability to **understand and deal** with **people**, **objects**, and **symbols**
- The ability to **act purposefully**, **think rationally**, and **deal effectively** with the environment

DEFINITIONS OF INTELLIGENCE

- Many confliction theories
  - **Spearman**'s Psychometric Approach
    - Intelligence as a Single Trait
  - **Catell -** Intelligence as a few basic abilities
    - Fluid Intelligence: think on the spot, solve novel problems
    - Crystalized Intelligence: factual knowledge about the world
  - **John Caroll** – Three Statum Theory



BROADER THEORY OF INTELLIGENCE

- **Howard Gardener**
  - Theory of *multiple intelligences*
  - 9 distinct types of intelligence
  - first three intelligences included in psychometric theories of intelligence
    - Linguistic intelligence
    - Logical-Mathematical Intelligence
    - Spatial Intelligence
  - Remaining 6 defined by Gardener
    - Musical
    - Bodily-kinesthetic
    - Interpersonal
    - Intrapersonal
    - Naturalistic
    - Existential intelligence

GARDNER‘S THEORY OF INTELLIGENCE

- **Musical:** Sensitivity to individual tones and phrases of music, an understanding of ways to combine tones and phrases into larger musical rhythms and structures, awareness of emotional aspects of music
- **Bodily-Kinesthetic:** Use of one’s body in highly skilled ways for expressive or goal-directed purposes, capacity to handle objects skilfully
- **Interpersonal:** Ability to notice and make distinctions among the moods, temperaments, motivations, and intentions of other people and potentially to act on this knowledge
- **Intrapersonal:** access to one’s own feelings, ability to draw on one’s emotions to guide and understand one’s behaviour, recognition of personal strengths and weaknesses
- **Naturalistic:** sensitivity and understanding of plants, animals, and other aspects of nature
- **Existential:** sensitivity to issues related to the meaning of life, death, and other aspects of the human condition

ARTIFICIAL INTELLIGENCE

Definition

WHAT IS AI?

- Academic Discipline / Computer Science

- *"Artificial Intelligence (AI) is the part of computer science concerned with* ***designing intelligent computer systems****, that is, systems that exhibit* ***characteristics we associate with intelligence in human behavior****"* (Barr & Feigenbaum, 1981)
  - Understanding language
  - Learning
  - Reasoning
  - solving problems

- ***Scientific Goal:*** To determine which ideas about knowledge representation, learning, rule systems, search, and so on, explain various sorts of real intelligence.

- ***Engineering Goal:*** To solve real world problems using AI techniques such as knowledge representation, learning, rule systems, search, and so on.

WHAT IS AI?

- **Roots / Foundations of AI**
  - Philosophy (e.g. Descartes, Leibnitz)
  - Logic / Mathematics (e.g. Gödel)
  - Computation (e.g. Turing, von Neumann)
  - Psychology / Cognitive Science (knowledge representation)
  - Biology / Neuroscience (Connectionism, Neural Networks)
  - Evolution (Genetic Programming)

- **Sub-fields of AI**
  - Neural Networks / Machine Learning
  - Evolutionary Computation
  - Computer Vision
  - Robotics
  - Expert Systems
  - Speech Processing
  - Natural Language Processing
  - Planning

WHAT IS AI?

- **AI-Effect**
  - AI successfully solves a problem
  - the problem is no longer a part of AI
  - Examples:
    - Digit Recognition
    - Optical Character Recognition (OCR)

  - "AI is whatever hasn't been done yet.“ (D. Hofstadter, 1980)

- **AI Hype Today**
  - Inverse AI-Effect
  - Companies claim to use AI
    - linear regression
    - Rule-based Systems

AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

HISTORY

of Artificial Intelligence

QUESTION

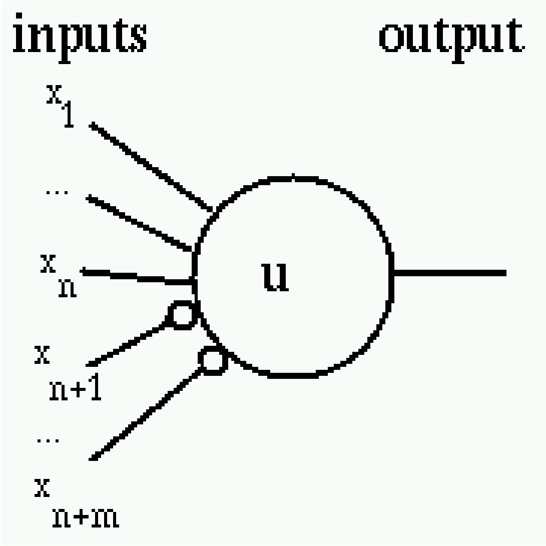
- How old is the research field of *Artificial Intelligence*?

- 5 years?
- 10 years?
- 20 years?
- 50 years?
- 80 years?
- 100 years?

A BRIEF HISTORY OF AI

- 1943: McCulloch and Pitts propose a model of artificial neurons
  - *"A Logical Calculus of the Ideas Immanent in Nervous Activity"*

- 1956: Minsky and Edmonds build first neural network computer, the *Stochastic neural analog reinforcement calculator (SNARC)*



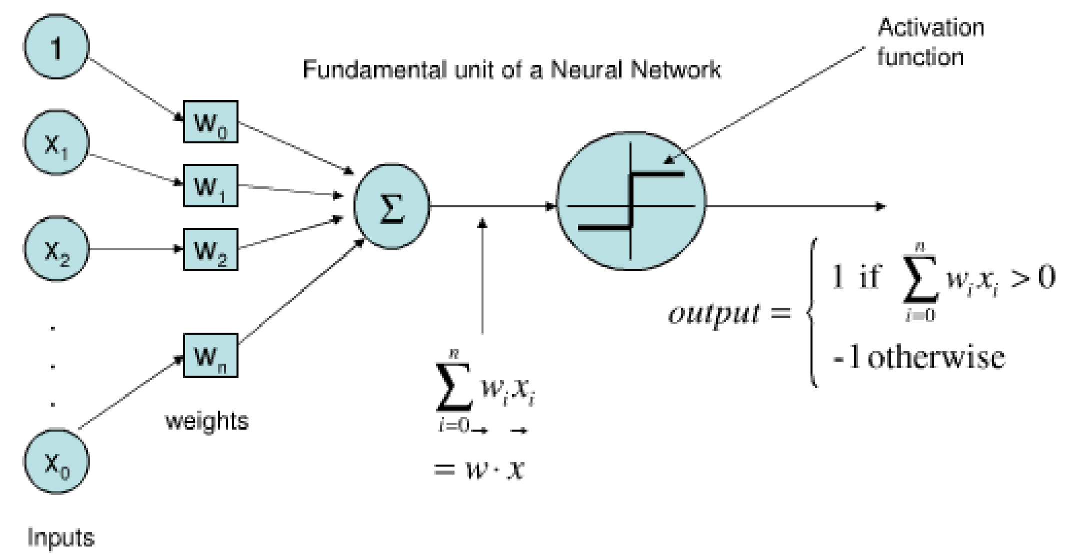
1956: **The Dartmouth Conference**

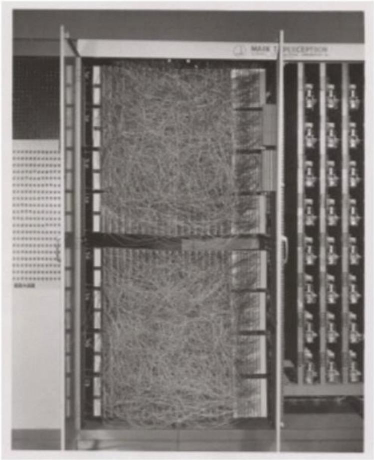
- **two-month workshop** for researchers interested in
  - neural networks
  - the study of intelligence

- Organizers
  - John McCarthy (Stanford)
  - Marvin Minsky (MIT)
  - Herbert Simon (CMU)
  - Allen Newell (CMU)
  - Arthur Samuel (IBM)

- **Agreement** to adopt a **new name** for this field of study:
  - **Artificial Intelligence**

1958: **Perceptron** by Frank Rosenblatt





Weights were encoded in potentiometers, and weight updates during learning were performed by electric motors.

Linear binary classifier using a step function

For the first time a NN could solve simple classification problems merely from training data

UPS & DOWNS OF NEURAL NETWORKS

1952 - 1969: Golden years of AI (funded by DARPA):

- Solve algebra
  - 1956 Logic Theorist
  - 1961 SAINT
- Reasoning
- Semantic nets

- 1952-1962: Checkers player (by Arthur Samuel)
- 1957: Newell and Simon: "within ten years a digital computer will be the world's chess champion"
- 1967: MacHack achieved class-C rating in tournament chess

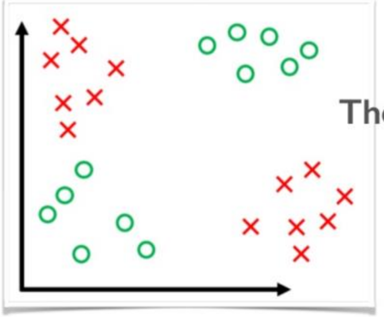
UPS & DOWNS OF NEURAL NETWORKS

1969: *Perceptrons: An Introduction to Computational Geometry*
by Marvin Minsky and *Seymour Papert*

- shown that XOR problem cannot be solved by Perceptron
  - But: they argued for locally connected neurons
    - Easier to implement in the 1960s
  - Fully connected three-layered Perceptrons can model XOR function

- Book was popular and is often cited as a show-stopper for AI

It is impossible to linearly separate the XOR function



The first winter: 1969 - 86

Is it?

1970s: **First AI-Winter**

- Book by Minsky and Papert was not the only problem
- AI problems appear to be too big and complex
- Computers are very slow, very expensive, and have very little memory (compared to today)
  - neural networks were tiny and could not achieve (the expected) high performance on real problems
  - Datasets were small

- Pessimism in the AI community
- followed by pessimism in the press
- followed by a severe cutback in funding
- followed by the “end” of serious research on Neural Networks
- First AI-Winter

1969 - 1979: **Knowledge-based Systems**

- Birth of expert systems
- Idea is to give AI systems lots of information to start with
    - Rule-based Systems
    - Fuzzy Logic

WHY NO DEEP LEARNING IN THE 1980S?

Neural Networks could not become “deep” yet - because:

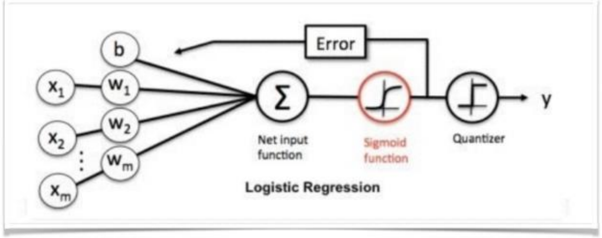
- Computers were slow. So the neural networks were tiny and could not achieve (the expected) high performance on real problems.
- Datasets were small. There were no large datasets that had enough information to constrain the numerous parameters of (hypothetical) large neural networks.
- Nobody knew how to train deep nets. Today, object recognition networks have > 25 successive layers of convolutions. In the past, everyone was very sure that such deep nets cannot be trained. Therefore, networks were shallow and did not achieve good results.

citing Ilya Sutskever
http://yyue.blogspot.co.at/2015/01/a-brief-overview-of-deep-learning.html

1980s: Governments (starting in Japan) and industry provide AI with billions of dollars. **Boom of “expert systems”.**

1986: **Backpropagation** had been invented in the 1970s, but only 1986 it became popular through a famous paper by David Rumelhart, Geoffrey Hinton, and Ronald Williams. It showed that also complex functions became solvable through NNs by using multiple layers.

Late 1980s: Investors - despite actual progress in research - became disillusioned and withdrew funding again.



SECOND AI-WINTER

1991: Hornik proved 1 hidden layer network can model any continuous function (universal approximation theorem)

1991/92 Vanishing Gradient: problem in multi-layer networks where training in front layers is slow due to backpropagation diminishing the gradient updates through the layers. Identified by Hochreiter & Schmidhuber who also proposed solutions.

1990s - mid 2000s:
Due to lack of computational power, interest in NNs decreased again and other Machine Learning models, such as Bayesian models, Decision Trees and Support Vector Machines became popular.

UPS & DOWNS OF NEURAL NETWORKS

1996: Deep Blue (IBM) beats world chess champion

2005: Stanford robot won the DARPA Grand Challenge by driving autonomously for 131 miles along an unrehearsed desert

2011: Watson (IBM), defeated the two greatest Jeopardy! champions

Wikipedia

Wikipedia

RESURRECTION OF DEEP LEARNING IN THE 2000S

2000s: Hinton, Bengio and LeCun (“The fathers of the age of deep learning”) join forces in a project

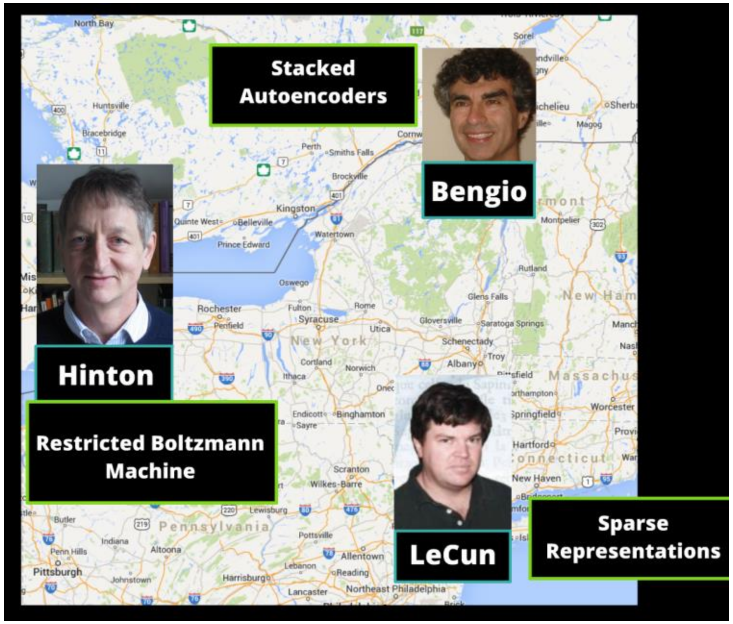
They overcome some problems that caused deep networks not to learn anything at all

2006: Breakthrough with Layer-wise pre-training by unsupervised learning (using RBMs)

2010s: Important new contributions:

- Simpler initialization (without pre-training)
- Dropout
- Simplier activations: Rectifier Units (ReLUs)
- Batch Normalization

→ not a re-invention of NNs but paved the way for very deep NNs



https://www.datarobot.com/blog/a-primer-on-deep-learning/

GPUs (2004)

Pattern Recognition

Volume 37, Issue 6, June 2004, Pages 1311–1314

Rapid and Brief Communication

GPU implementation of neural networks

Kyoung-Su Oh, Keechul Jung
Show more

doi:10.1016/j.patcog.2004.01.013

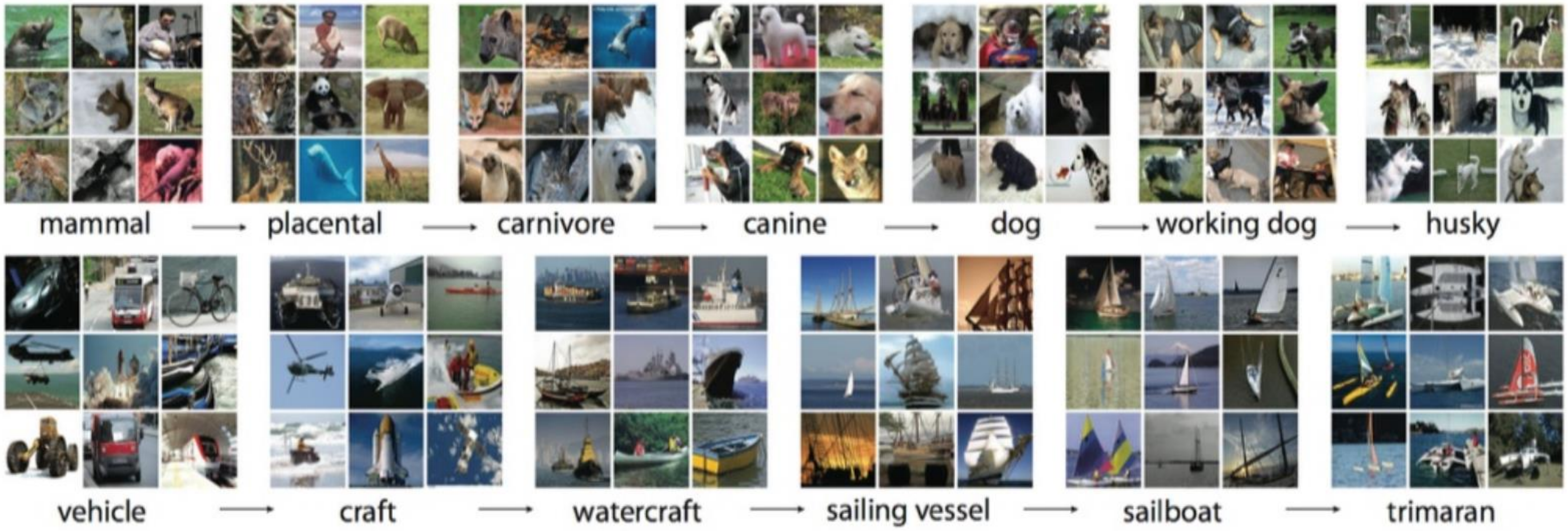
Get rights and content

Abstract

Graphics processing unit (GPU) is used for a faster artificial neural network. It is used to implement the matrix multiplication of a neural network to enhance the time performance of a text detection system. Preliminary results produced a 20-fold performance

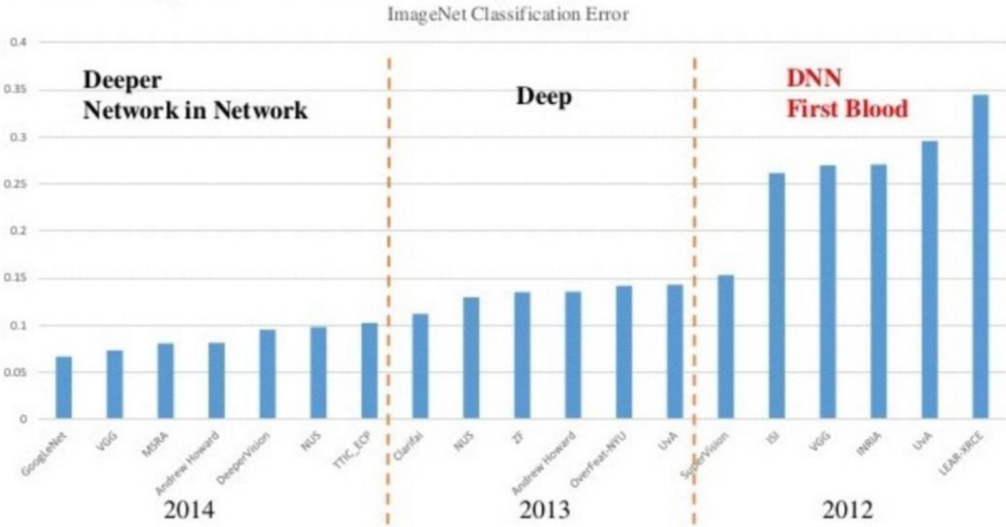
IMAGENET (2009)

- Large Database for Visual Object Recognition
- More than 20.000 categories
- Aligned to WordNet
- Large Scale Visual Recognition Challenge (ILSVRC)
- Significant impact on Deep Learning

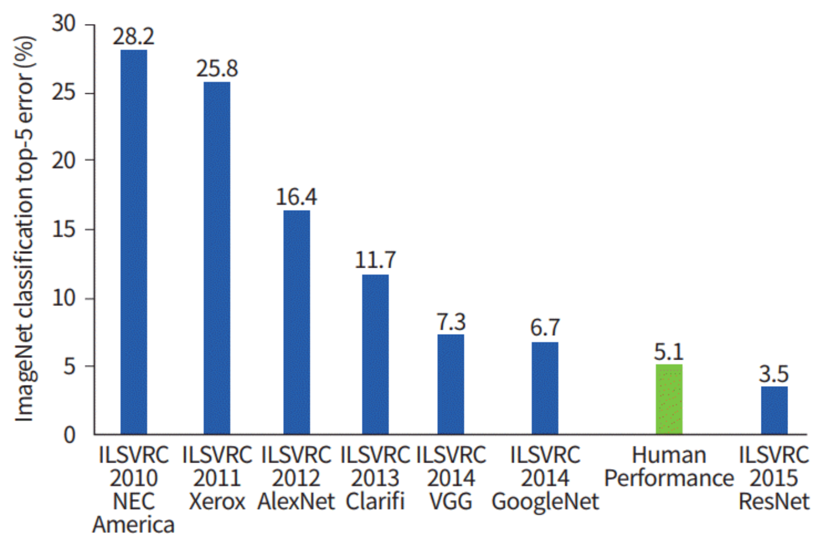


DROPPING ERROR RATES SINCE THEN

- **1000** categories and **1.2** million training images

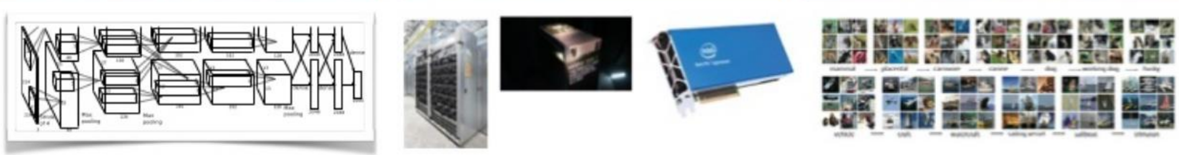


Li Fei-Fei: ImageNet Large Scale Visual Recognition Challenge, 2014 http://image-net.org/



SUMMARY:
THREE EPOCHS OF NEURAL NETWORKS

	techniques / tricks	hardware	data
1957-69 dawn	perceptron	early mainframes	toy linear, small images, XOR
1986-95 golden age	early NNs	workstations	MNIST
2006- deep learning	deep NNs	GPU, TPU, Intel Xeon Phi	Imagenet



Introduction to

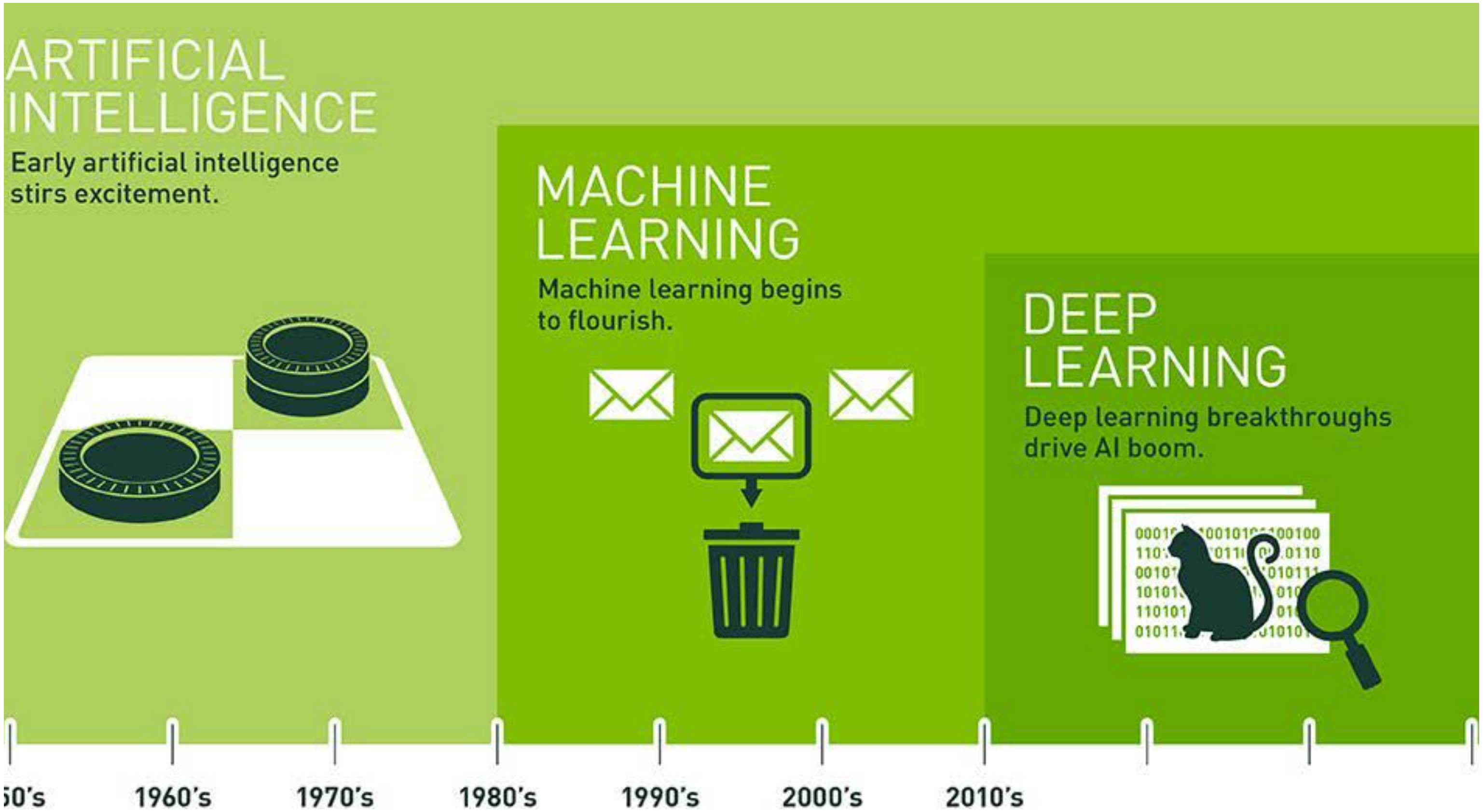
Deep Learning

Deep Learning

Definition

WHAT IS DEEP LEARNING?

- **Deep Learning** is
  - **Machine Learning**
  - using **Deep Neural Networks**

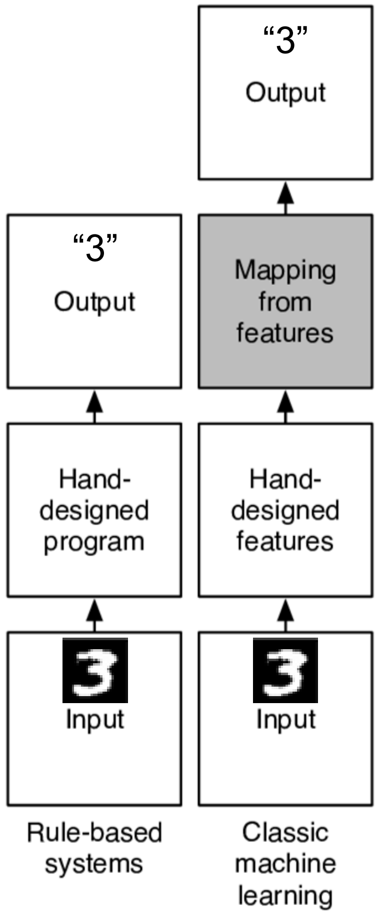


Rule-based systems:
Write algorithm by hand.

Rule-based systems

graphic: Y. Bengio, Deep Learning, MLSS 2015

MACHINE LEARNING PARADIGMS

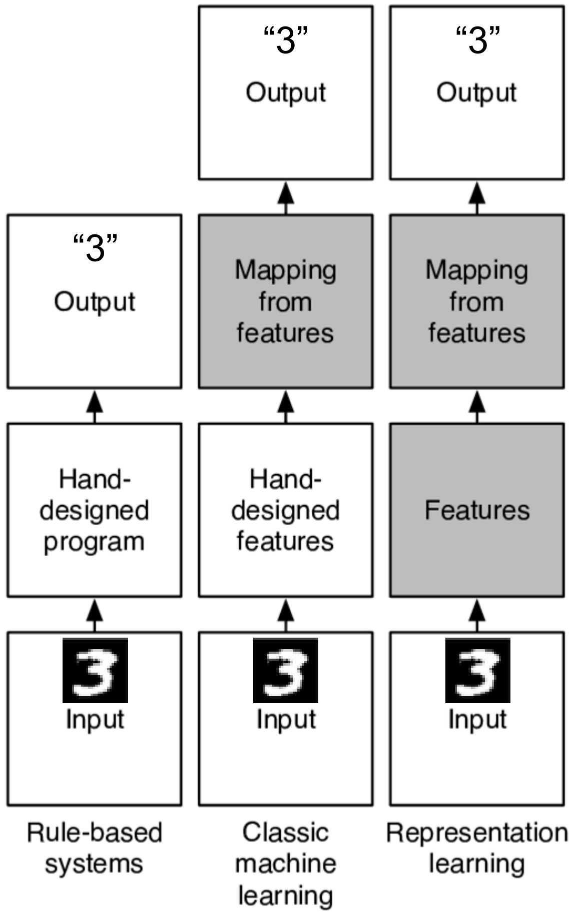


Rule-based systems:
Write algorithm by hand.

Classic machine learning:
Write feature extractor by hand, train classifier on top.

graphic: Y. Bengio, Deep Learning, MLSS 2015

MACHINE LEARNING PARADIGMS



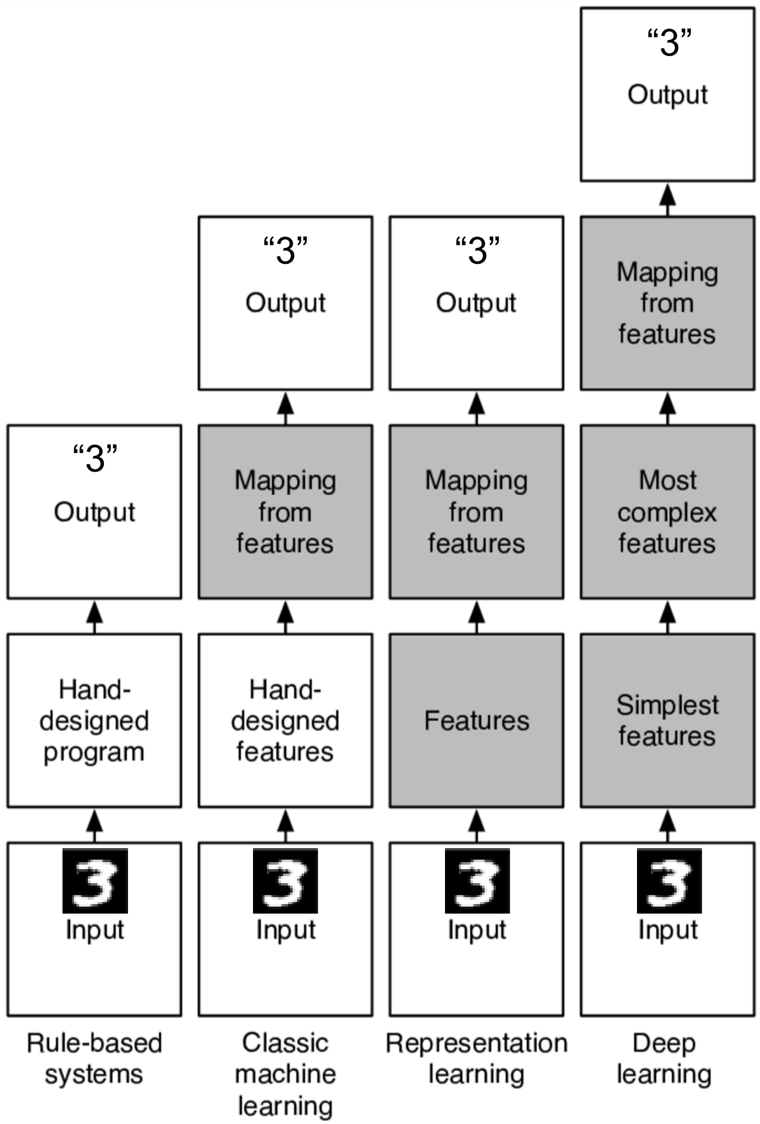
Rule-based systems:
Write algorithm by hand.

Classic machine learning:
Write feature extractor by hand, train classifier on top.

Representation learning:
Learn feature extractor (often unsupervised), train classifier on top.

graphic: Y. Bengio, Deep Learning, MLSS 2015

MACHINE LEARNING PARADIGMS



Deep learning:
Learn a stack of many simpler functions to map input to output.

- Often, that stack is a ***neural network***.
- Often, it is trained on raw input: optimize features & classification together, ***minimize hand-crafting***. (“end-to-end learning”)

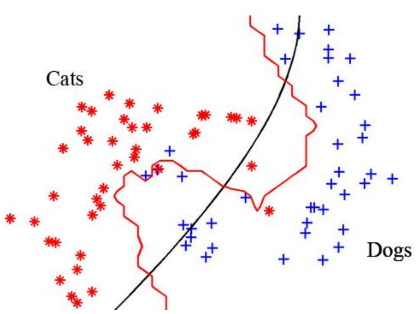
graphic: Y. Bengio, Deep Learning, MLSS 2015

MACHINE LEARNING PARADIGMS

Supervised

- Labelled data
- Learn from association
  - Input ⇔ Output

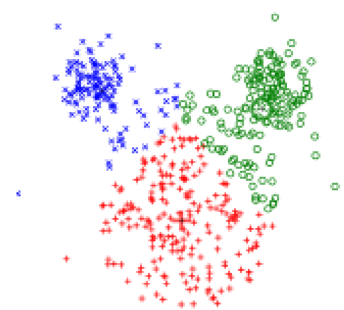
- Examples:
  - Classification:
    - Predict a category
  - Regression:
    - Predict a value (on a continuous scale)



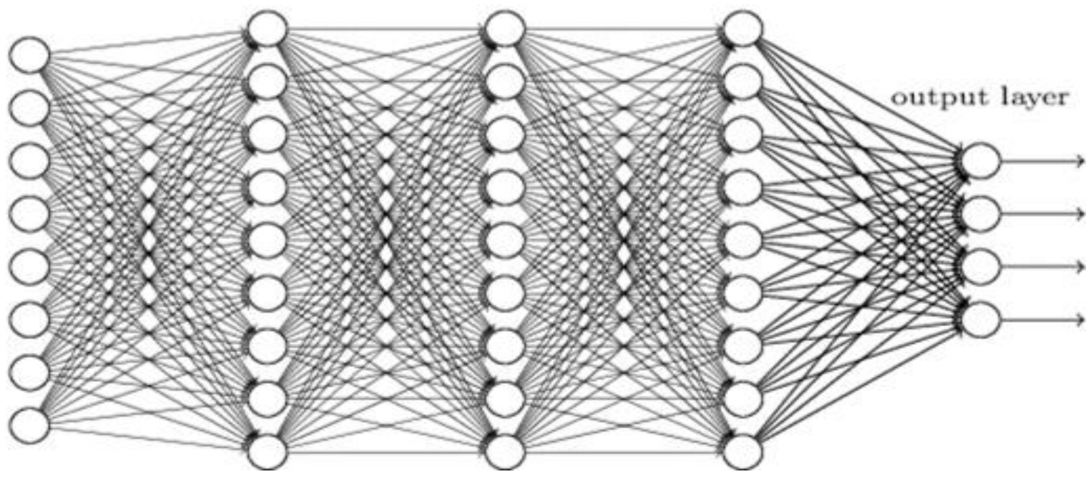
Unsupervised

- Unlabelled data
- Learn from inherent structures

- Examples:
  - Clustering:
    - Group the data into clusters of similar data, Self-organizing systems
  - Data Mining:
    - Discover patterns



So What is Deep Learning?



WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

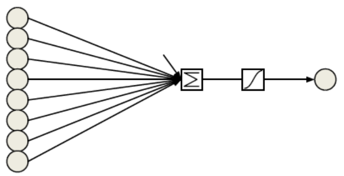
$$y = \sigma(b + \mathbf{w}^T\mathbf{x}) \qquad \text{(equivalent to logistic regression)}$$

WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

$y = \sigma(b + \mathbf{w}^T\mathbf{x})$ (equivalent to logistic regression)

expression can be visualized as a graph:

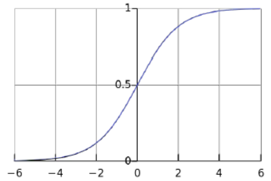


$\mathbf{x}$ $b + \mathbf{w}^T\mathbf{x}$ y

Output value is computed as a **weighted sum of its inputs,**

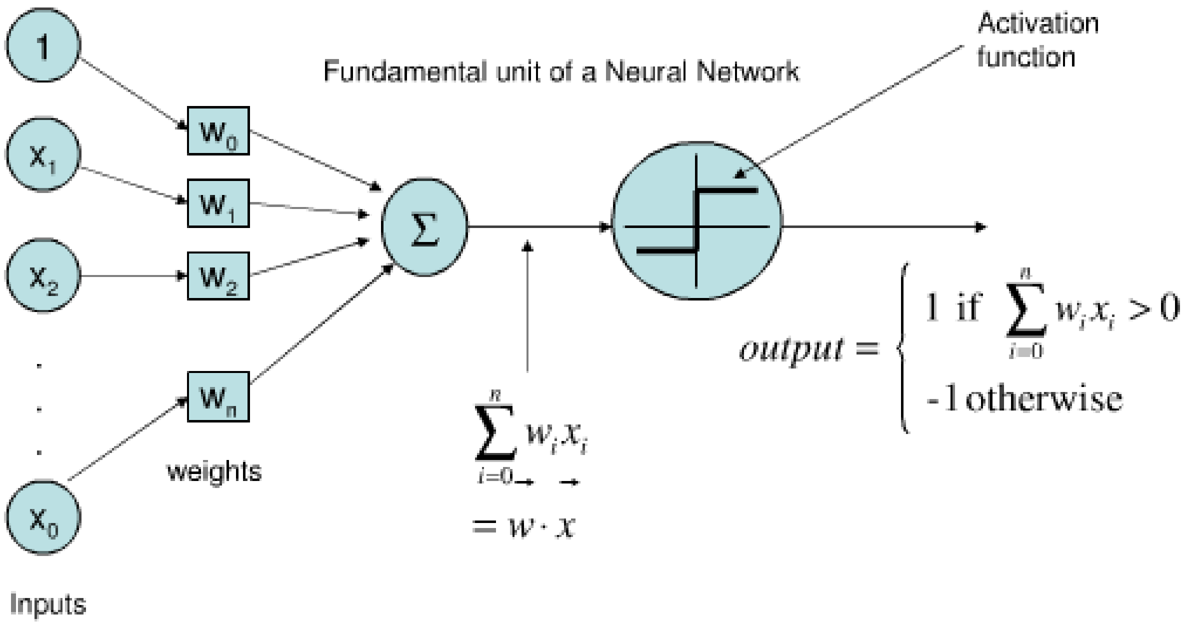
$$b + \mathbf{w}^T\mathbf{x} = b + \Sigma_i w_i x_i$$

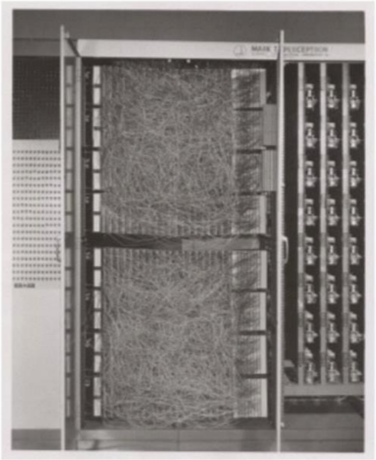
followed by a nonlinear function.



Origins of Neural Networks: The Perceptron

1958 by Frank Rosenblatt





Weights were encoded in potentiometers, and weight updates during learning were performed by electric motors.

Linear binary classifier using a step function

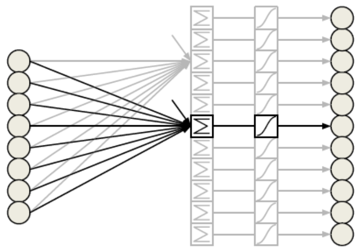
For the first time a NN could solve simple classification problems merely from training data

WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

$\mathbf{y} = \sigma(\mathbf{b} + \mathbf{W}^T\mathbf{x})$ (multiple logistic regressions)

expression can be visualized as a graph:

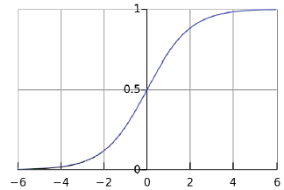


$\mathbf{x}$ $\mathbf{b} + \mathbf{W}^T\mathbf{x}$ $\mathbf{y}$

Output values are computed as **weighted sums of their inputs,**

$$\mathbf{b} + \mathbf{W}^T\mathbf{x} = b_j + \Sigma_i w_{ij} x_i$$

followed by a nonlinear function.

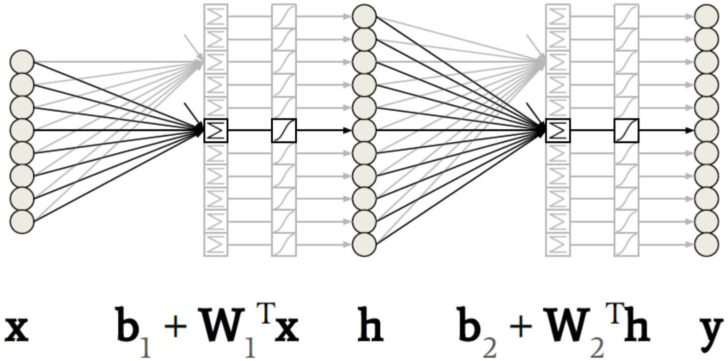


WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

$$\mathbf{y} = \sigma(\mathbf{b}_2 + \mathbf{W}_2{}^T\sigma(\mathbf{b}_1 + \mathbf{W}_1{}^T\mathbf{x})) \qquad \text{(stacked logistic regressions)}$$

expression can be visualized as a graph:

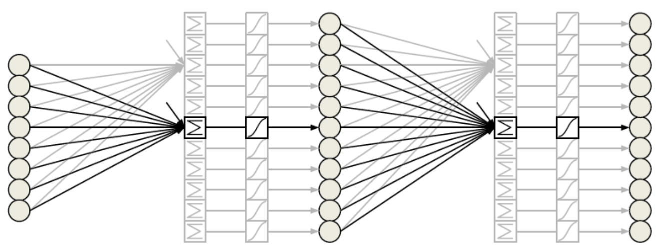


WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

$$\mathbf{y} = \sigma(\mathbf{b}_2 + \mathbf{W}_2^{\mathrm{T}}\sigma(\mathbf{b}_1 + \mathbf{W}_1^{\mathrm{T}}\mathbf{x})) \qquad \text{(stacked logistic regressions)}$$

expression can be visualized as a graph:



$\mathbf{x} \quad \mathbf{b}_1 + \mathbf{W}_1^{\mathrm{T}}\mathbf{x} \quad \mathbf{h} \quad \mathbf{b}_2 + \mathbf{W}_2^{\mathrm{T}}\mathbf{h} \quad \mathbf{y}$

Universal Approximation Theorem:
This can model any continuous function from $\mathbb{R}^n$ to $\mathbb{R}^m$ arbitrarily well (if $\mathbf{h}$ is made large enough).

MATHEMATICAL REASONS FOR GOING "DEEP"

A neural network with a single hidden layer of enough units can approximate any continuous function arbitrarily well. In other words, it can solve whatever problem you're interested in!
(Cybenko 1998, Hornik 1991)

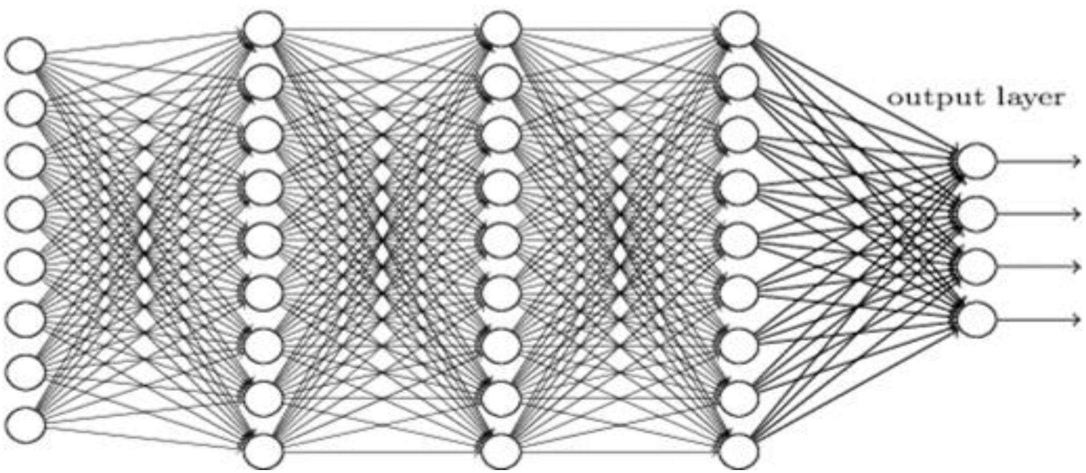
But:

- "Enough units" can be a very large number. There are functions representable with a small, but deep network that would require exponentially many units with a single layer.
  (e.g., Hastad et al. 1986, Bengio & Delalleau 2011)

- The proof only says that a shallow network *exists*, it does not say how to find it. Evidence indicates that it is easier to train a deep network to perform well than a shallow one.

WHAT MAKES THEM DEEP?

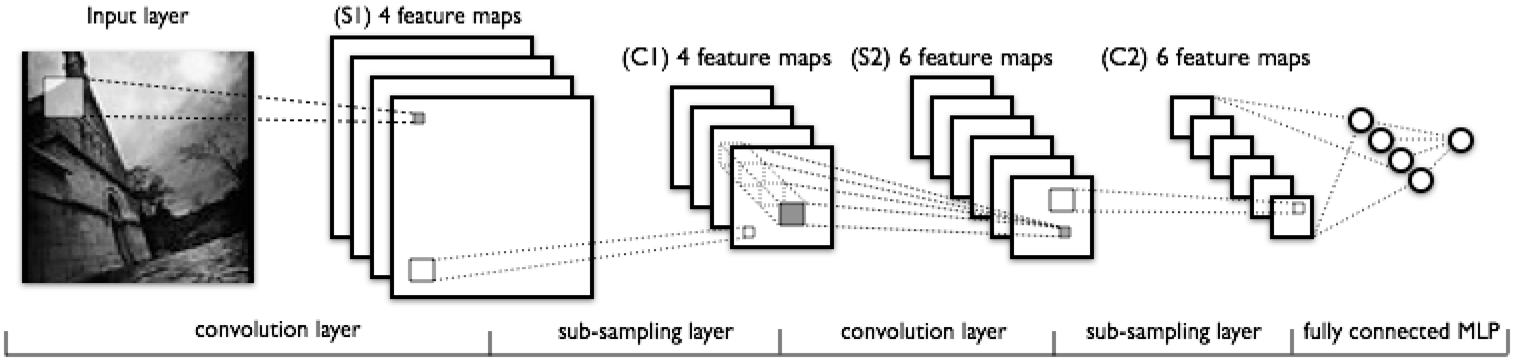
- Neural Networks can learn any arbitrary function
- That‘s what makes them so powerful
- The more layers they have, the more expressive they can be
- However, mathematically the problem gets more difficult to solve
- And computationally it becomes extreme (or unfeasible)
  (that‘s why GPUs are needed)



More hidden layers = deep

Deep Learning Concepts

CONVOLUTIONAL NEURAL NETWORK (CNN)



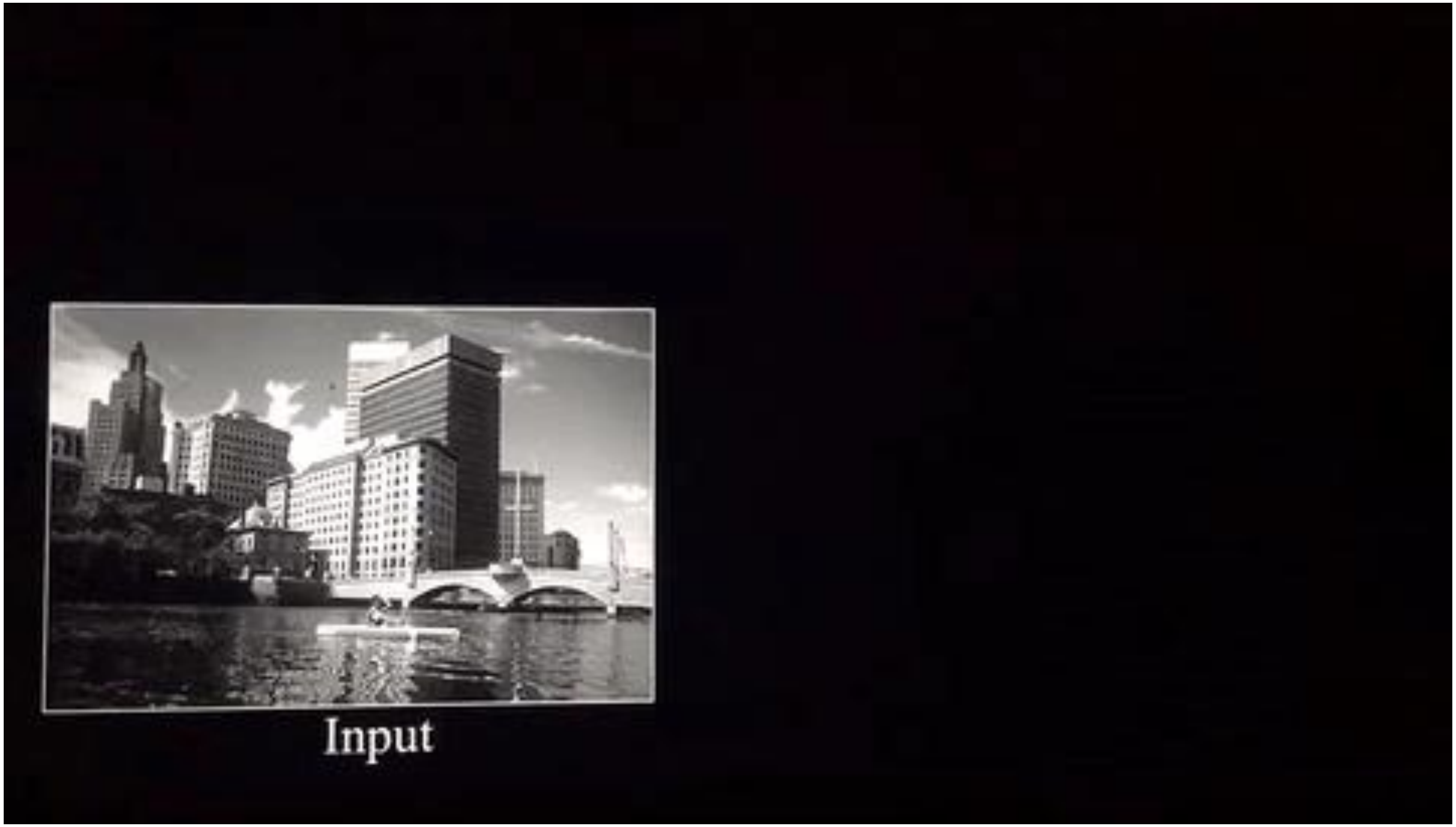
http://deeplearning.net/tutorial/lenet.html

Combines three types of layers:

- **Convolutional layer:** performs 2D convolution of 2D input with multiple learned 2D kernels
- **Subsampling layer:** replaces 2D patches by their maximum ("max-pooling") or average
- **Fully-connected layer:** computes weighted sums of its input with multiple sets of learned coefficients

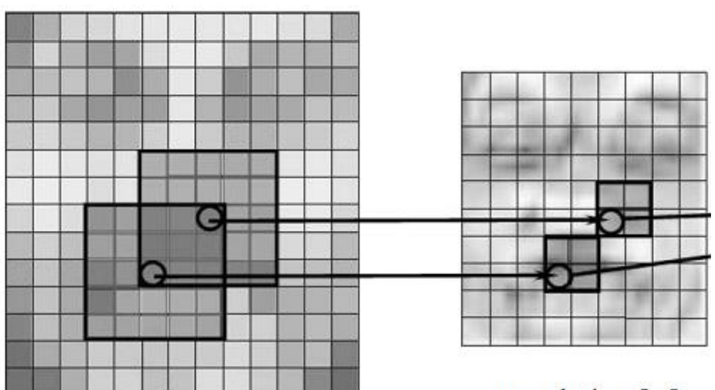
Applies a nonlinear function after each linear operation (without, a deep network would be linear despite its depth).

IMAGE PROCESSING
CONVOLVING FILTERS

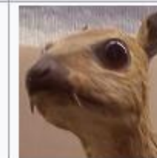


MOTIVATION FOR CONVOLUTIONS

- Apply local filter kernels
- These kernels are the neurons that are learned



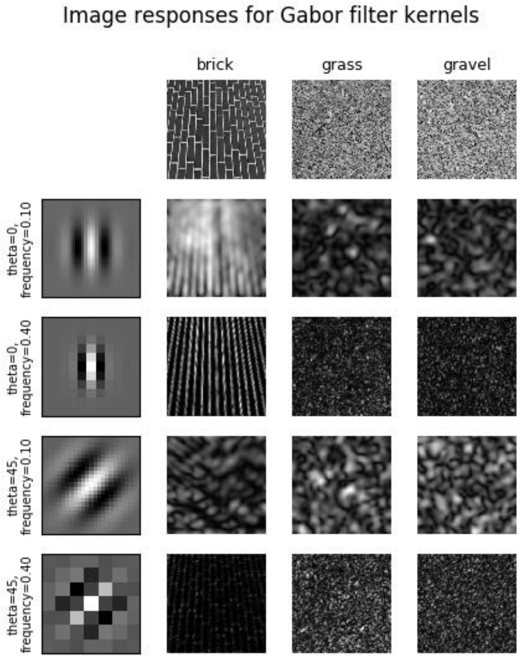
convolution 5x5

Operation	Kernel	Image result
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	

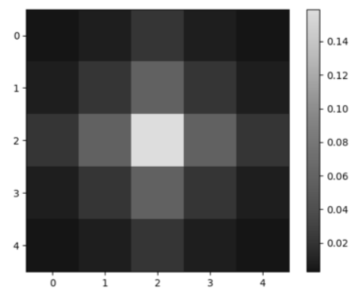
Images: http://sanghyukchun.github.io/75/
https://en.wikipedia.org/wiki/Kernel_(image_processing)

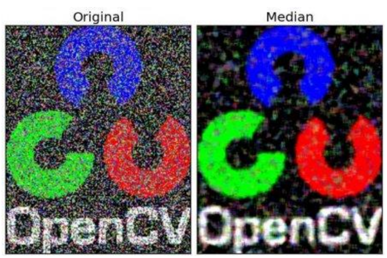
FILTER EXAMPLES

Gabor Filters (Texture)



Gauss Filters (Texture)

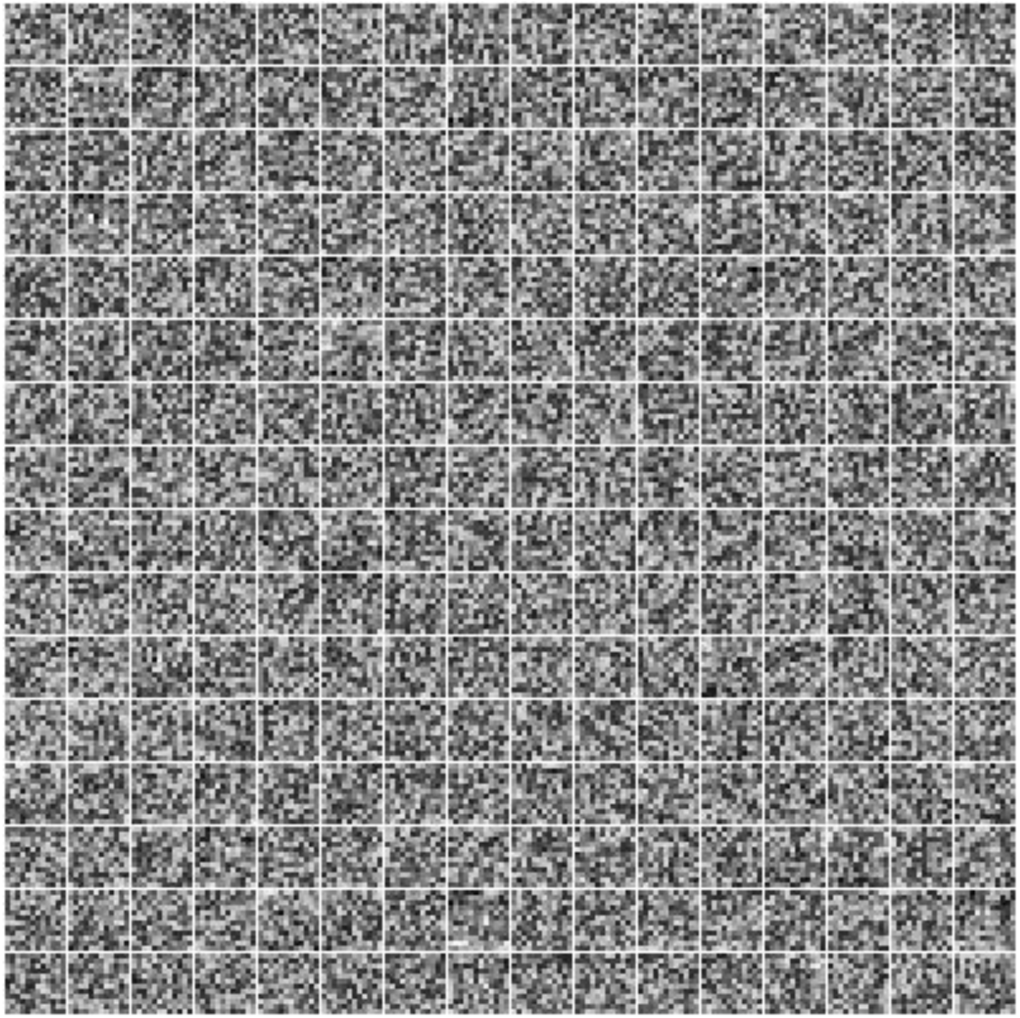




Gabor filter banks for texture classification https://scikit-image.org/docs/dev/auto_examples/features_detection/plot_gabor.html#sphx-glr-auto-examples-features-detection-plot-gabor-py

CV Basics https://yoyoinwanderland.github.io/CV-Basics/

IMAGE PROCESSING



iteration no 0

Second very important aspect of a CNN:
(also called subsampling or downsampling)

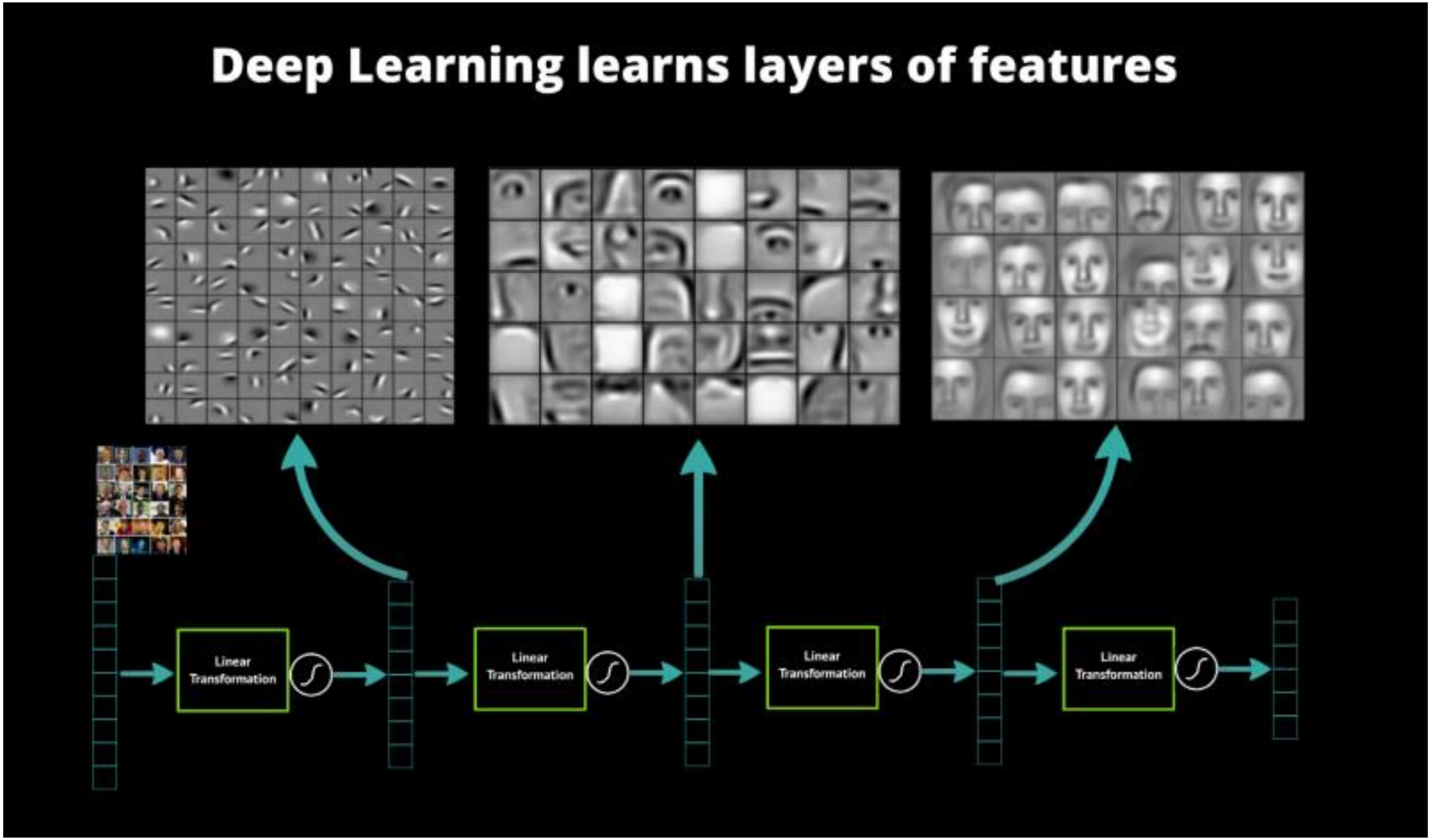
A **pooling layer** reduces the size of feature maps (i.e. output of a CNN layer and thus the input to the next layer)

1	**3**	**5**	-3
-2	**3**	2	-1
1	2	-2	**2**
-1	**3**	-2	1

3	5
3	2

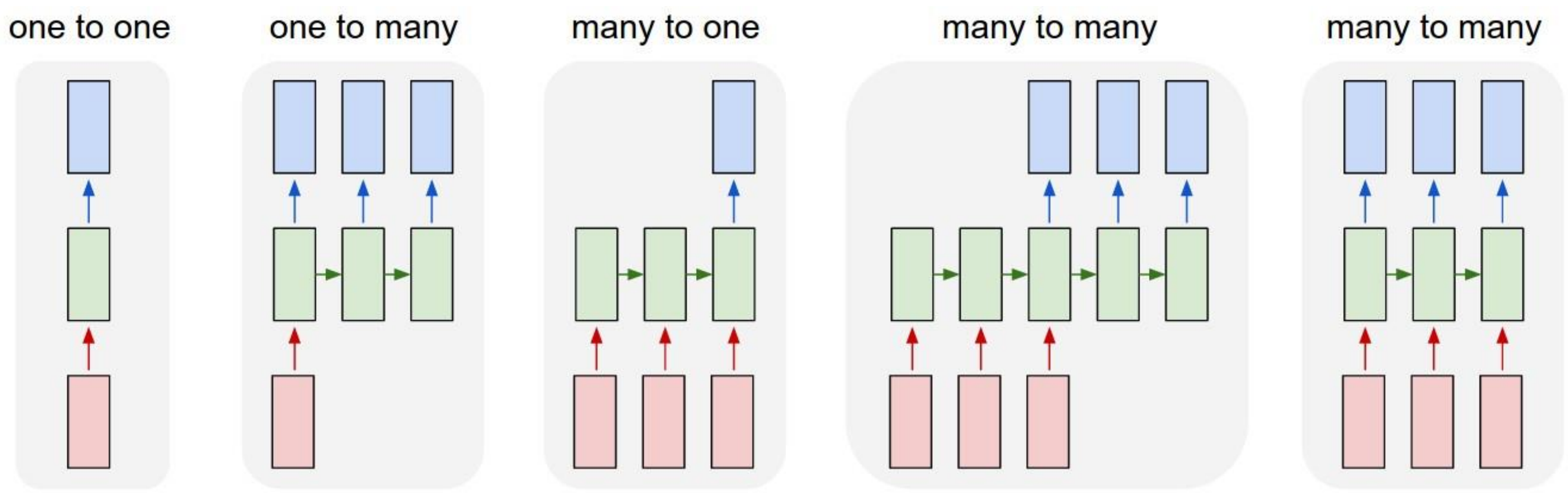
Max pooling: take the max. activation across small regions (e.g. 2x2, as in the example above)

FACE RECOGNITION WITH CNNS



RECURRENT NEURAL NETWORKS

Process **sequences** of data in various ways:



2) Image description (output: multiple words or sentences)
3) Sentiment analysis (input: text, output: mood category)
4) Translation (English -> French)
5) Synced input / output (e.g. frame-wise video categorization)

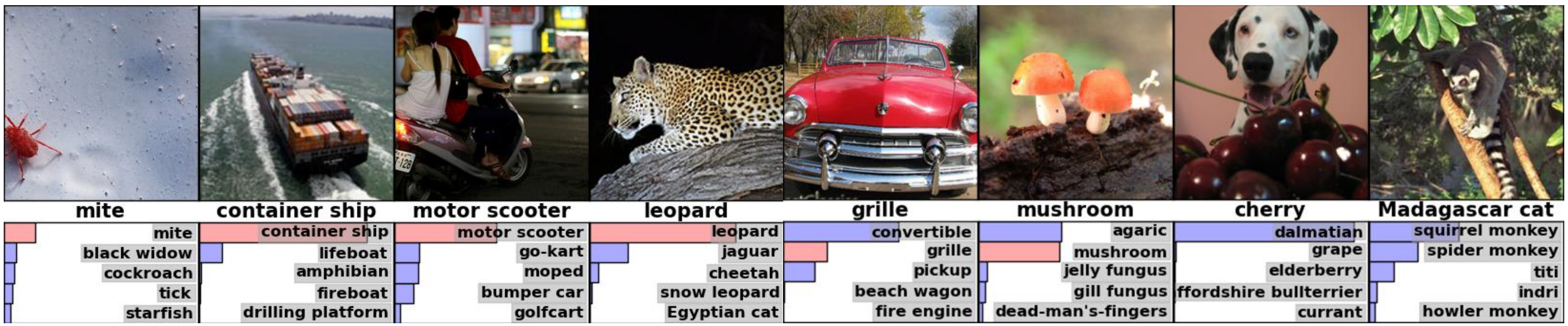
Image: http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Deep Learning

Application Examples

OBJECT CLASSIFICATION

IMAGENET arge Scale Visual Recognition Challenge:
1.2 million training images of 1000 classes



- 2013: Clarifai: 11.7% top-5 error, or 11.2% with extra data (most contestants used deep nets now).
- 2014: GoogLeNet: 6.7%. Andrej Karpathy (a human): 5.1%.
- 2015: ResNets: 3.6%.

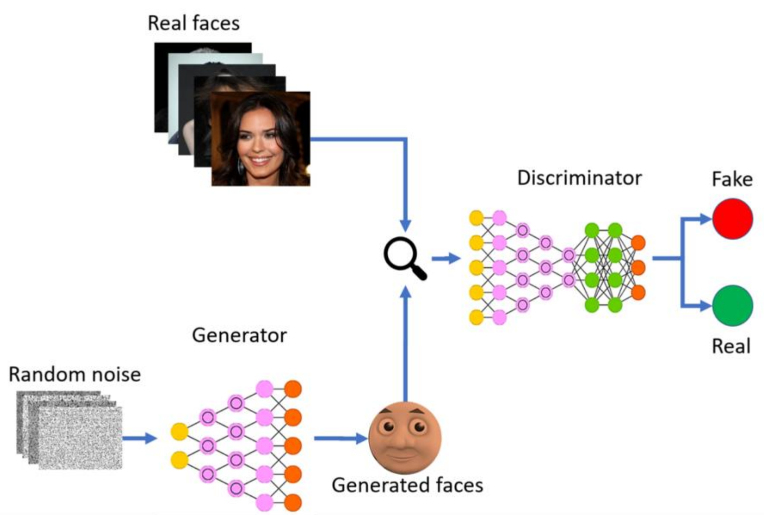
IMAGE CAPTIONING

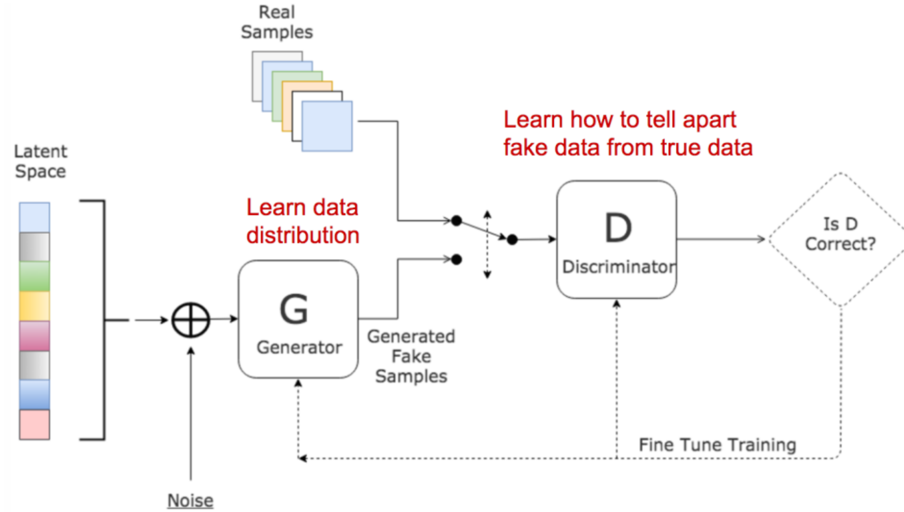
Input: Photograph. **Output:** Textual description.
Method: CNN to analyze image, RNN to output text

Nov 2014: Show and Tell: A Neural Image Caption Generator, http://arxiv.org/abs/1411.4555

GENERATIVE ADVERSARIAL NETWORKS (GAN)

Generative adversarial nets
I Goodfellow, J Pouget-Abadie, M Mirza... - Advances in neural ..., 2014 -
... Like **generative adversarial networks**, variational autoencoders pair **network** with a second neural **network**. Unlike **generative adversarial network** in a VAE is a recognition model that performs approximate infer
Zitiert von: 13618 Ähnliche Artikel Alle 29 Versionen





A Brief Introduction To GANs
https://medium.com/sigmoid/a-brief-introduction-to-gans-and-how-to-code-them-2620ee465c30

IMAGE GENERATION

Generative Adversarial Networks: One network transforms random noise into images, a second one tries to distinguish generated from real images. Both are trained at the same time.

Photographs of bed rooms that do not actually exist

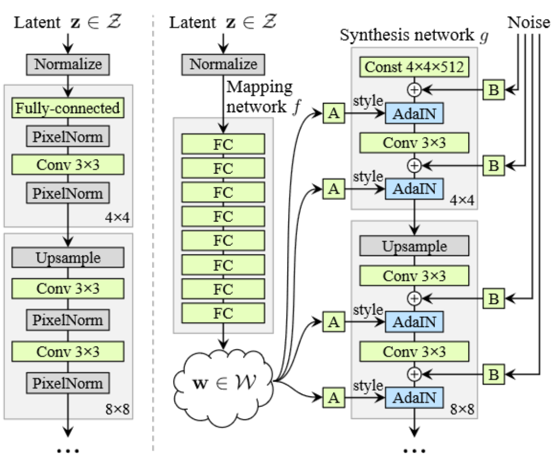
Nov 2015: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, http://arxiv.org/abs/1511.06434, https://github.com/Newmu/dcgan_code

A Style-Based Generator Architecture for Generative Adversarial Networks

Tero Karras
NVIDIA
tkarras@nvidia.com

Samuli Laine
NVIDIA
slaine@nvidia.com

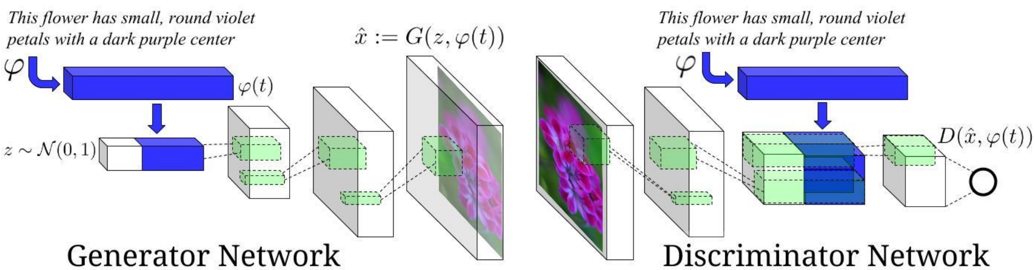
Timo Aila
NVIDIA
taila@nvidia.com



(a) Traditional (b) Style-based generator

Generative Adversarial Text to Image Synthesis

Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran REEDSCOT[1], AKATA[2], XCYAN[1], LLAJAN[1]
Bernt Schiele, Honglak Lee SCHIELE[2],HONGLAK[1]
[1] University of Michigan, Ann Arbor, MI, USA (UMICH.EDU)
[2] Max Planck Institute for Informatics, Saarbrücken, Germany (MPI-INF.MPG.DE)



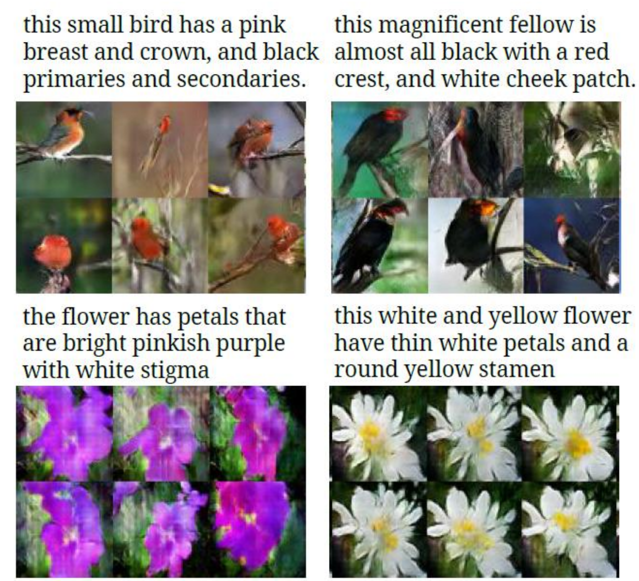
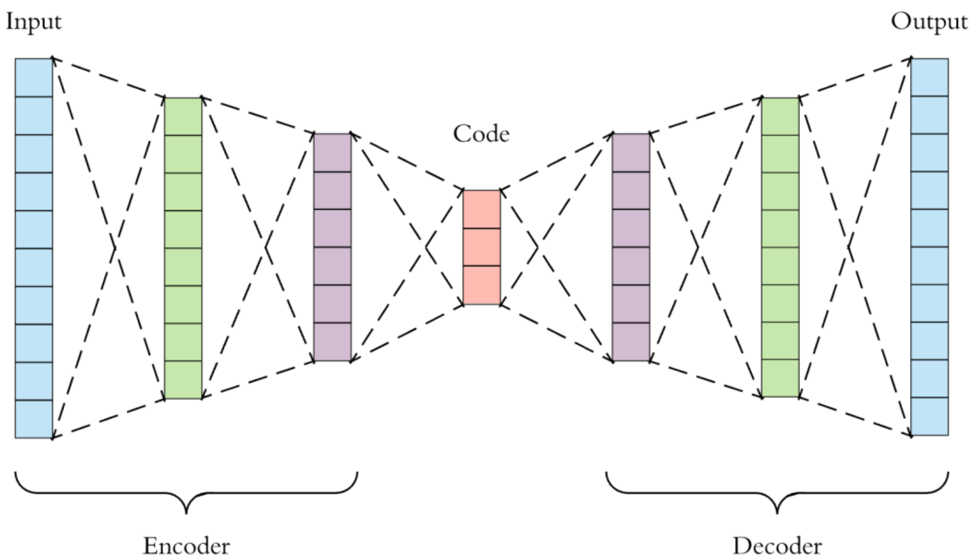
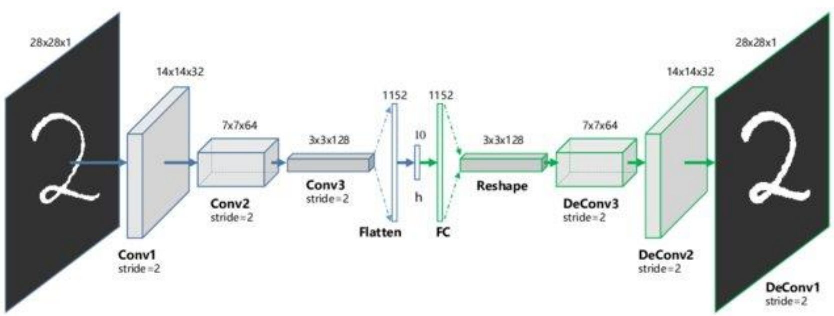


Figure 1. Examples of generated images from text descriptions. Left: captions are from zero-shot (held out) categories, unseen text. Right: captions are from the training set.

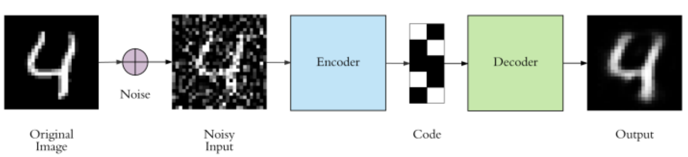
- Paper: http://arxiv.org/abs/1605.05396
- Github: https://github.com/paarthneekhara/text-to-image

AUTO ENCODERS



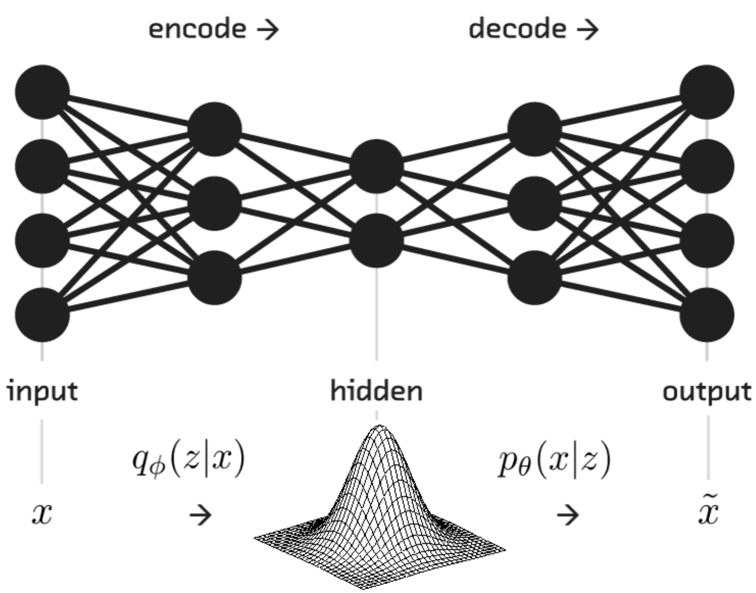


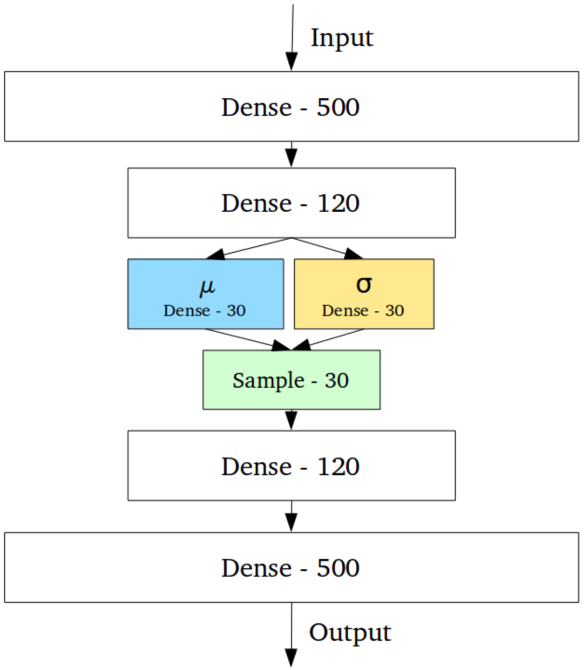
Guo, Xifeng, et al. "Deep clustering with convolutional autoencoders." International Conference on Neural Information Processing. Springer, Cham, 2017.



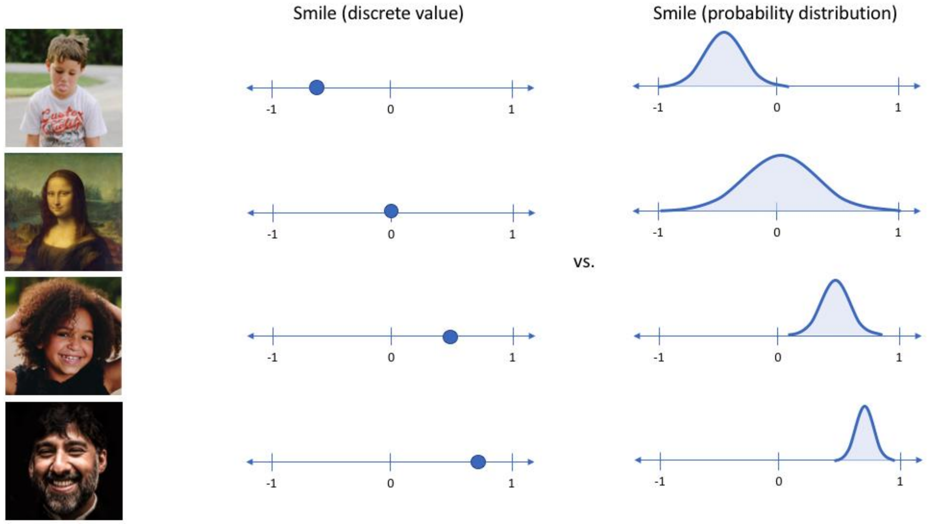
Applied Deep Learning - Part 3: Autoencoders
https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798

VARIATIONAL AUTO ENCODERS





PROBABILITY DISTRIBUTION OF LATENT FACTORS



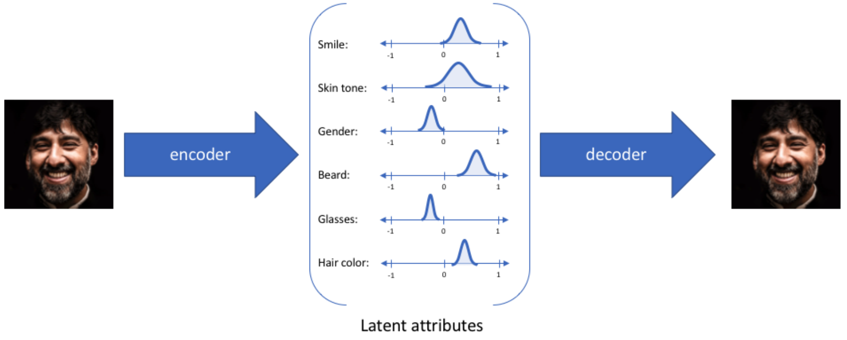
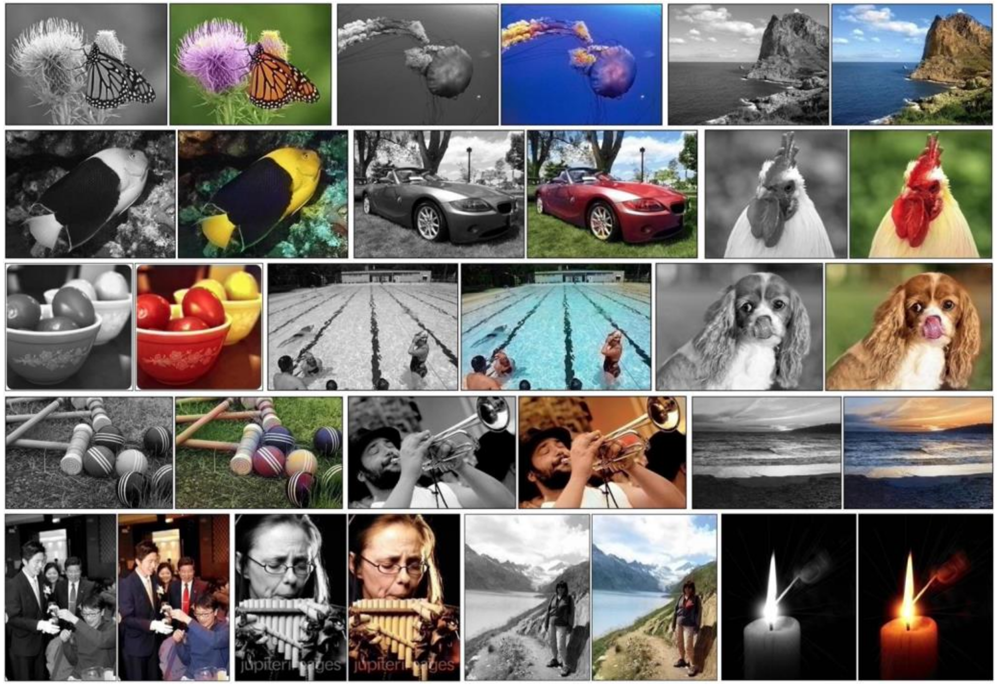


IMAGE COLORIZATION

Input: Gray-scale image. **Output:** Colored image.
Method: fully-convolutional network



Mar 2016: Colorful Image Colorization, http://arxiv.org/abs/1603.08511, http://richzhang.github.io/colorization/

NEURAL STYLE TRANSFER

Starting point: ConvNet trained for object classification
Use network in reverse: Find image that matches high-level representation of photograph and low-level repr. of painting

Aug 2015: A Neural Algorithm of Artistic Style, http://arxiv.org/abs/1508.06576

WAVENET

ConvNet that can **predict next step of time sequence**, using a clever architecture for processing a large temporal context (about 3000-6000 past time steps)

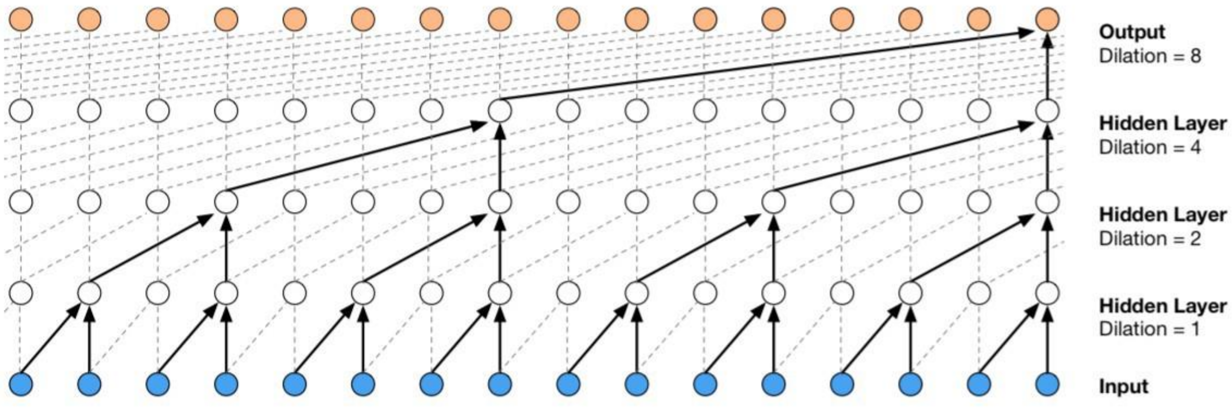


Figure 3: Visualization of a stack of *dilated* causal convolutional layers.

https://deepmind.com/blog/wavenet-generative-model-raw-audio/

TACOTRON-2 EXAMPLES

- Synthetic or Real?

"That girl did a video about Star Wars lipstick."

"She earned a doctorate in sociology at Columbia University."

"George Washington was the first President of the United States."

Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., ... & Saurous, R. A. (2018, April). Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4779-4783). IEEE.

- Stress and intonation
- Questions
- Prosody
  - Intonation, rhythm, tone

Style / Reference

Base Voice Model

Reference text: Alice was not much surprised at this, she was getting so used to queer things happening.

Result

Perturbed text: Eric was not much surprised at this, he was getting so used to TensorFlow breaking.

Singing

Skerry-Ryan, R. J., Battenberg, E., Xiao, Y., Wang, Y., Stanton, D., Shor, J., ... & Saurous, R. A. (2018). Towards end-to-end prosody transfer for expressive speech synthesis with tacotron. arXiv preprint arXiv:1803.09047.

DEEP FAKES

- **Visual Style Transfer**
  - Only face using Face recognition
- **No Deep Fake for Audio!!!**
  - Deep Fakes with voice require
    - Traditional audio manipulation
    - Impressionist
- **Approaches to style-transfer in audio**
  - Not promising

Deep Fake VFX - Pity the poor impressionist by Jim Meskimen https://www.youtube.com/watch?v=Wm3squcz7Aw

FURTHER RESOURCES

Code & Tutorials

AIT AUSTRIAN INSTITUTE OF TECHNOLOGY

Vienna Deep Learning Meetup

Austria's largest monthly event on Deep Learning & AI

The Organizers:

Thomas Lidy
Musimap

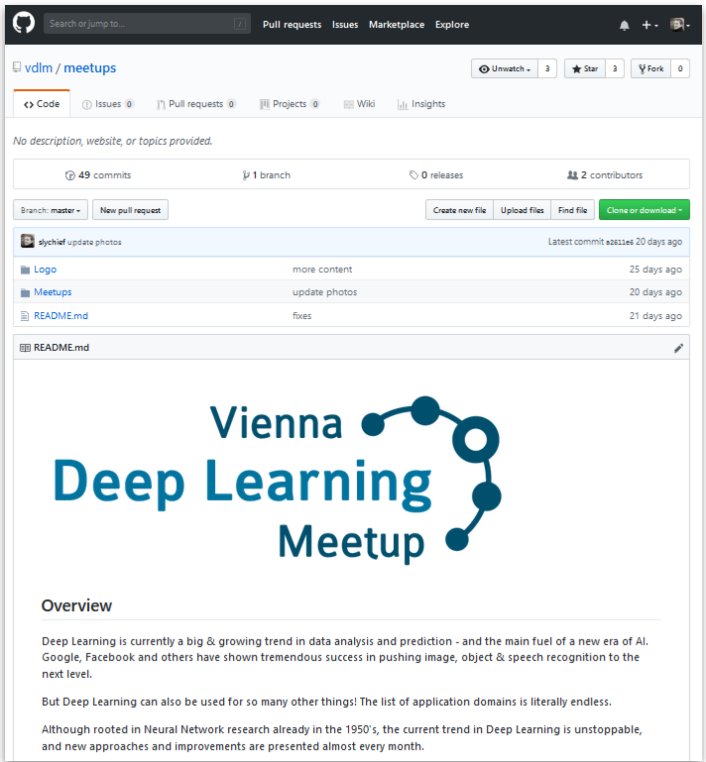
Alex Schindler
AIT & TU Wien

René Donner
contextflow

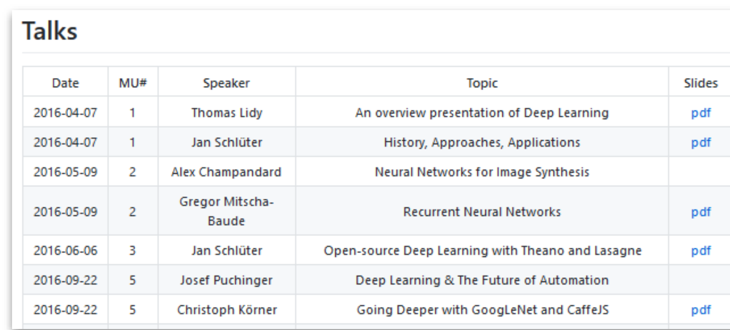
Jan Schlüter
OFAI & UTLN

www.meetup.com/Vienna-Deep-Learning-Meetup

VDLM ON GITHUB



- ➔ Talks
- ➔ Slides
- ➔ Videos
- ➔ Wiki with beginner's resources



Talks

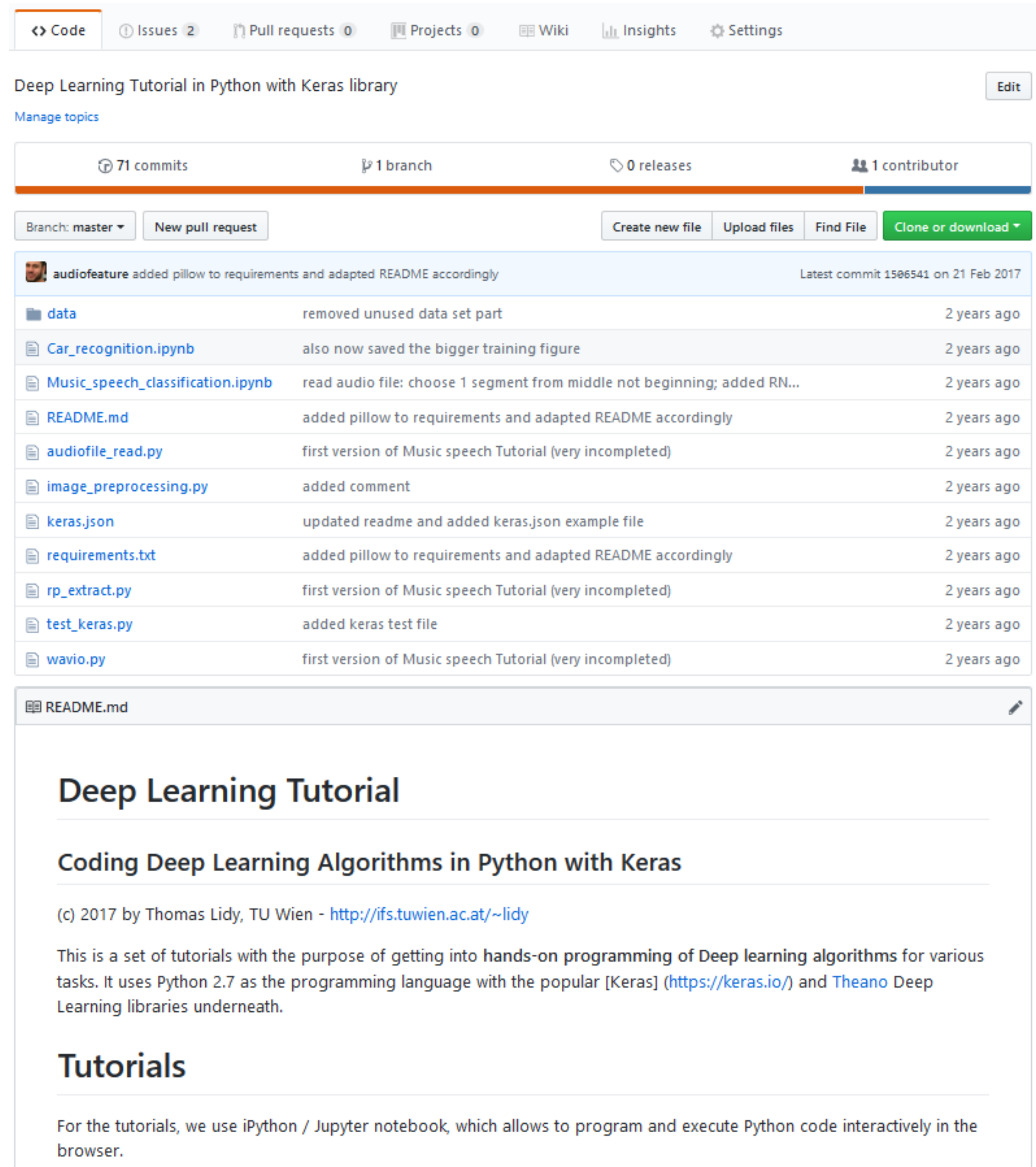
Date	MU#	Speaker	Topic	Slides
2016-04-07	1	Thomas Lidy	An overview presentation of Deep Learning	pdf
2016-04-07	1	Jan Schlüter	History, Approaches, Applications	pdf
2016-05-09	2	Alex Champandard	Neural Networks for Image Synthesis	
2016-05-09	2	Gregor Mitscha-Baude	Recurrent Neural Networks	pdf
2016-06-06	3	Jan Schlüter	Open-source Deep Learning with Theano and Lasagne	pdf
2016-09-22	5	Josef Puchinger	Deep Learning & The Future of Automation	
2016-09-22	5	Christoph Körner	Going Deeper with GoogLeNet and CaffeJS	pdf

github.com/vdlm/meetups

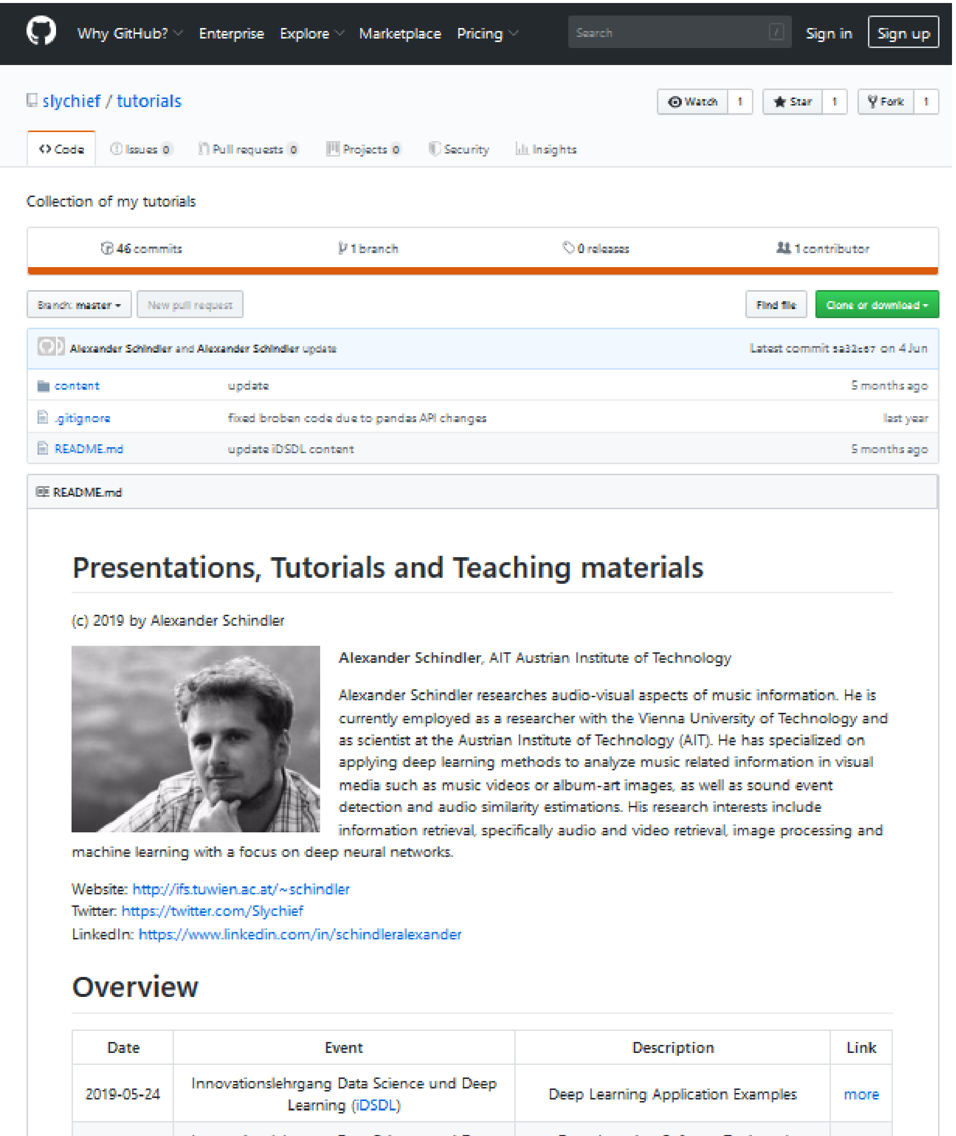
DEEP LEARNING TUTORIAL

- Beginners Tutorial
  - Simple Image Processing
  - Simple Audio Processing
  - Python
  - Keras
  - Tensorflow

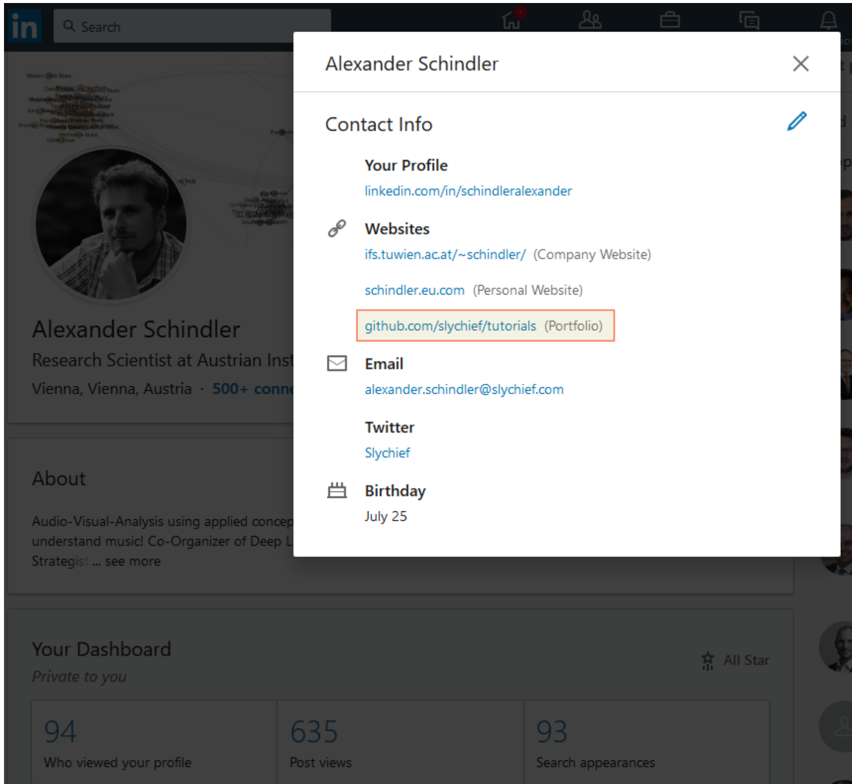
https://github.com/tuwien-musicir/DL_Tutorial



SLIDES OF THIS TALK



- Further Questions?
- Connect via LinkedIn



AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

THANK YOU!