Deep Learning

What society thinks I do

What my friends think I do

What other computer scientists think I do

What mathematicians think I do

What I think I do

```
from theano import *
```

What I actually do

Overview!

Introduction to deep learning!

Big data + Big computational power (GPUs)!

Feature extraction!

Examples!

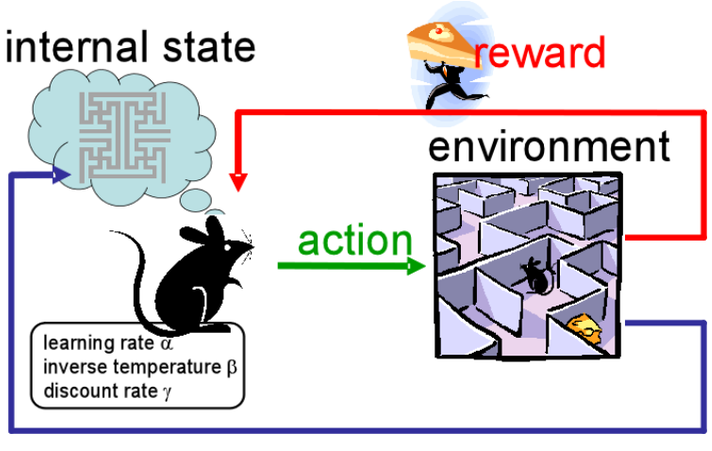
Latest updates of deep learning!

Machine Learning!

Supervised Learning

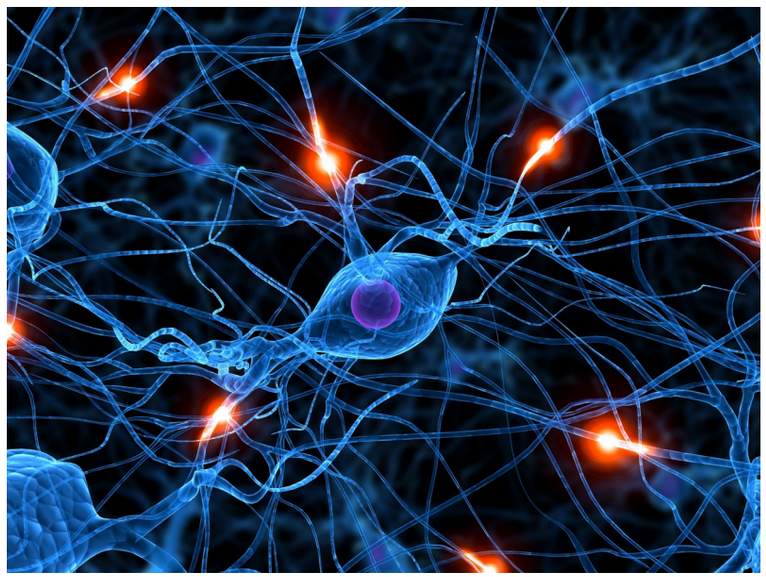
Unsupervised Learning

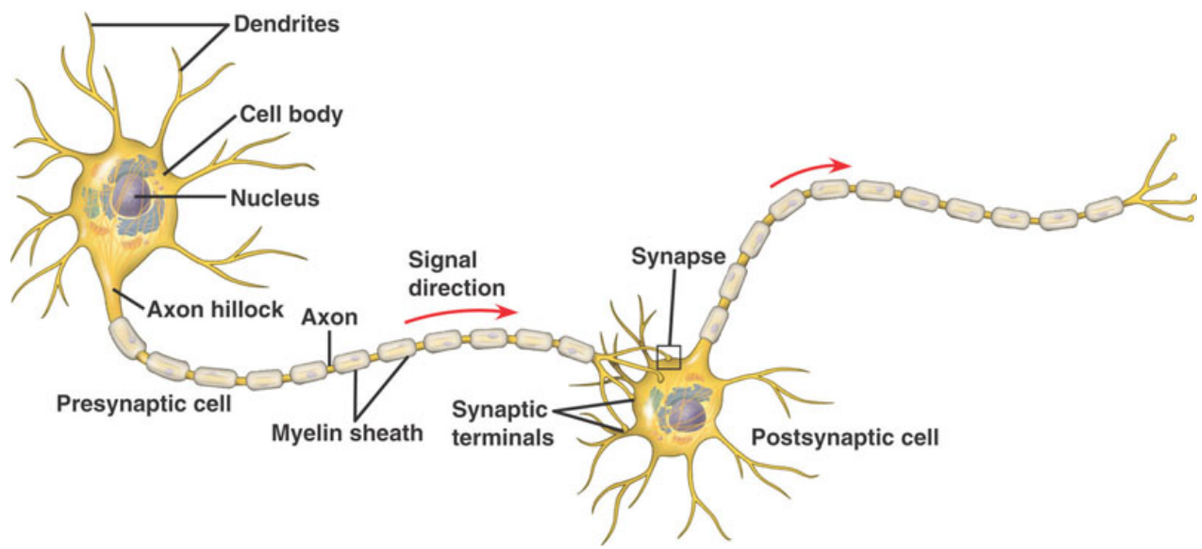
Reinforcement learning



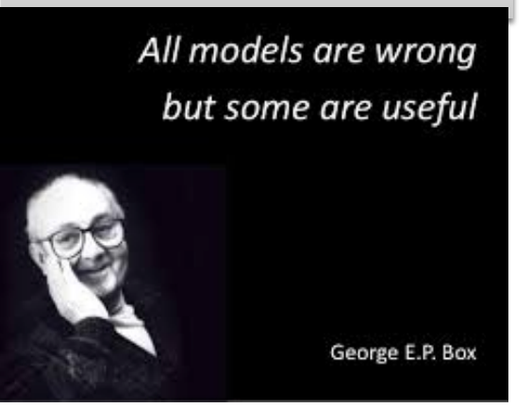
source : http://www.cs.utexas.edu/~eladlieb/RLRG.html!

Inspiration from Human brain!

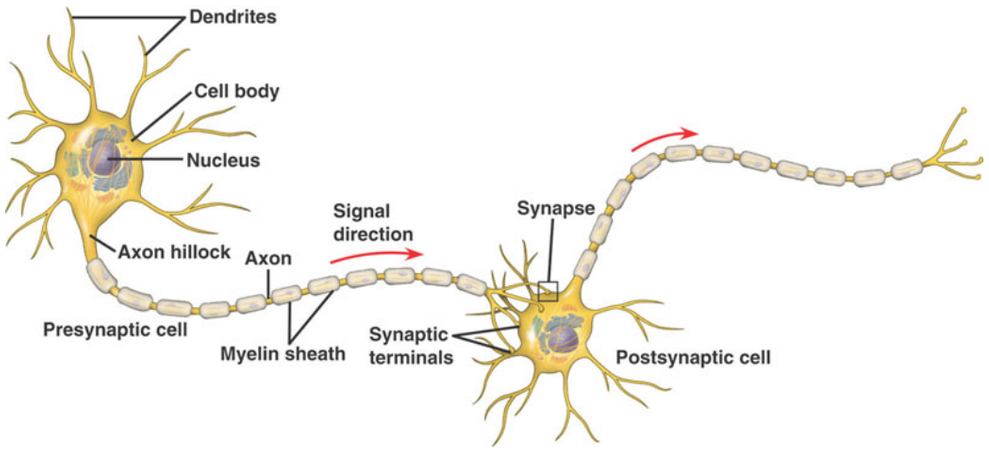


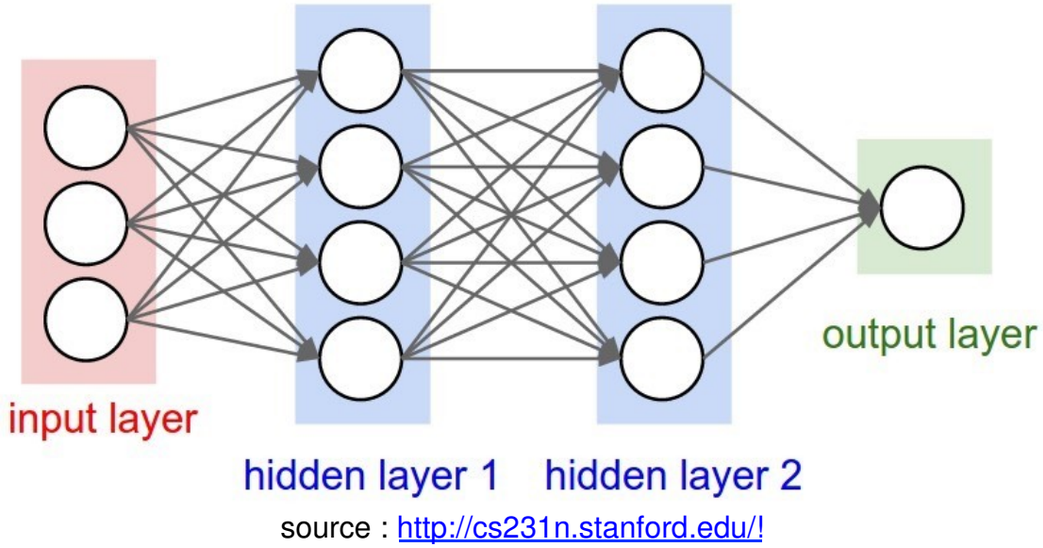


- Over 100 billion neurons !
- Around 1000 connections !
- Electrical signal transmitted through axon, can be inhibitory or excitatory!
- Activation of a neuron depends on inputs!



Artificial Neuron!

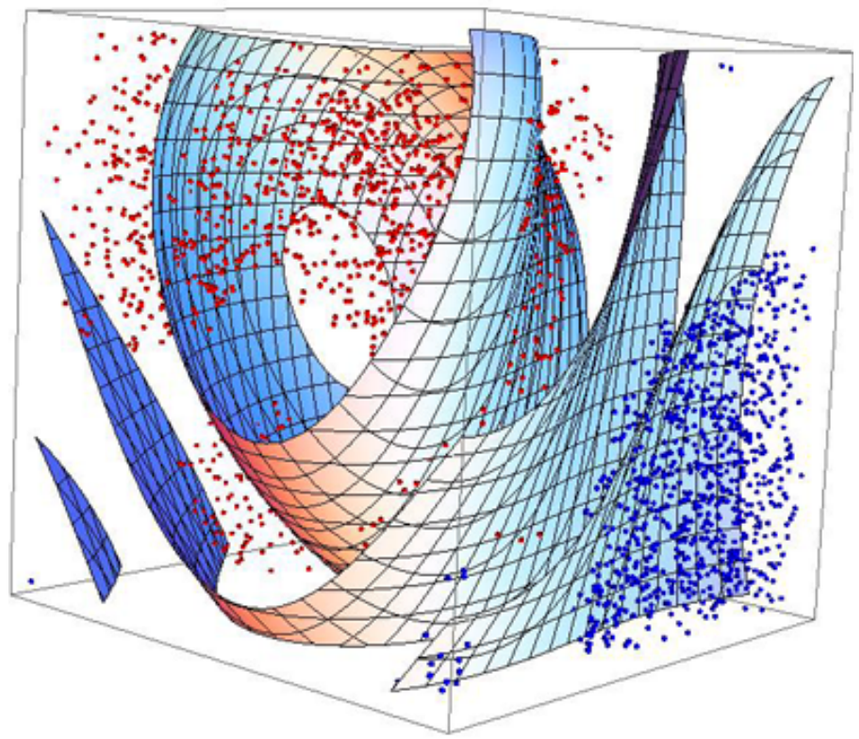




source : http://cs231n.stanford.edu/!

- ? Circles ——————————————- Neuron!
- ? Arrow ——————————————— Synapse!
- ? Weight ——————————————- Electrical signal!

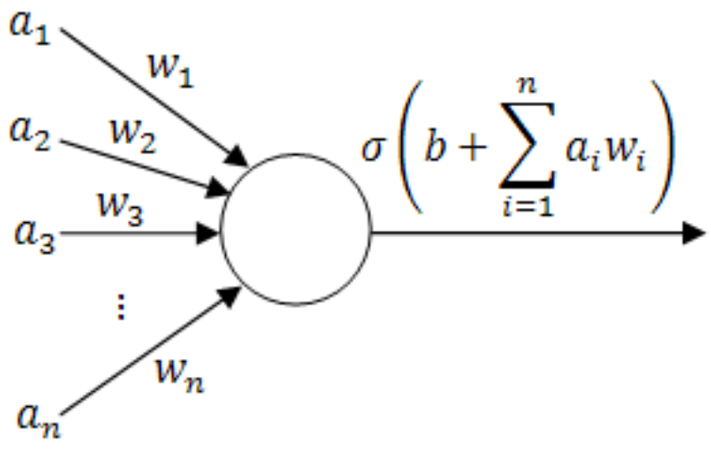
Nonlinearity!



- ? Neural network learns complicated structured data with non-linear activation functions!

Perceptron!

- Perceptron with linear activation!



- In1969, cannot learn XOR function and takes very long to train!

Winter of AI!

Back-Propogation!

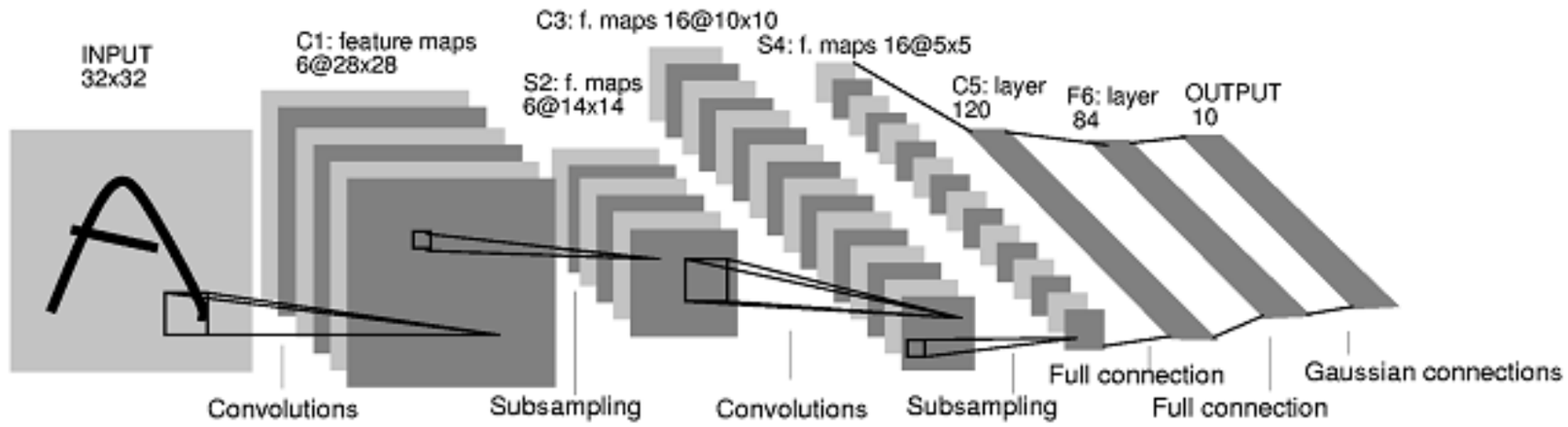
- 1970, got BS on psychology and keep studying NN at University of Edinburg !

- 1986, Geoffrey Hinton and David Rumelhart, Nature paper "Learning Representations by Back-propagating errors”!

  - Hidden layers added!
  - Computation linearly depends on # of neurons !
  - Of course, computers at that time is much faster!

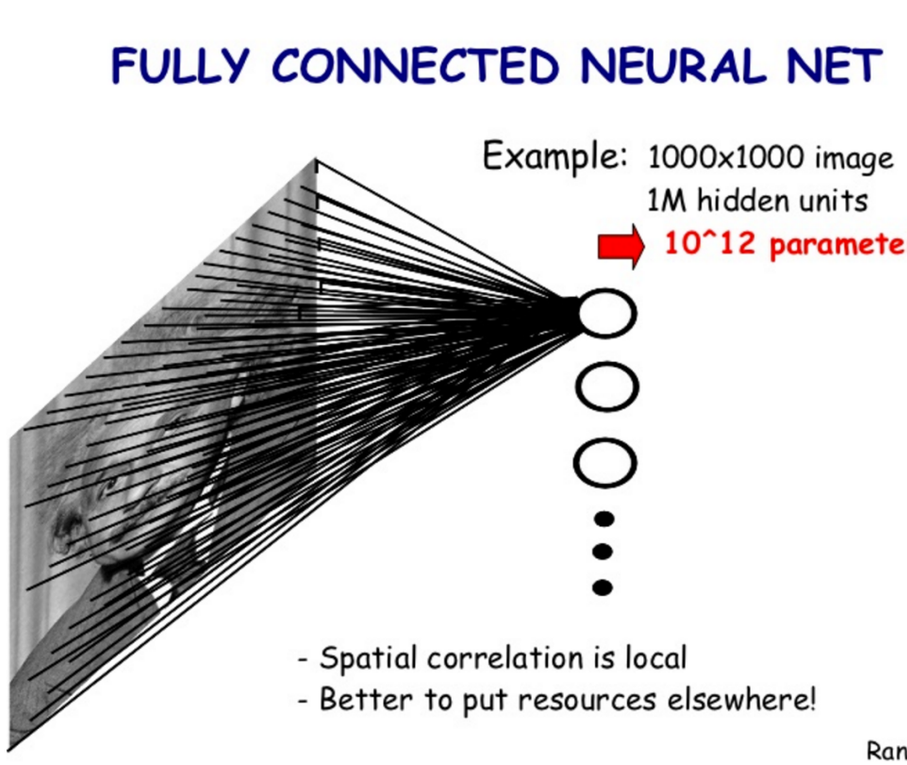
Convolutionary NN!

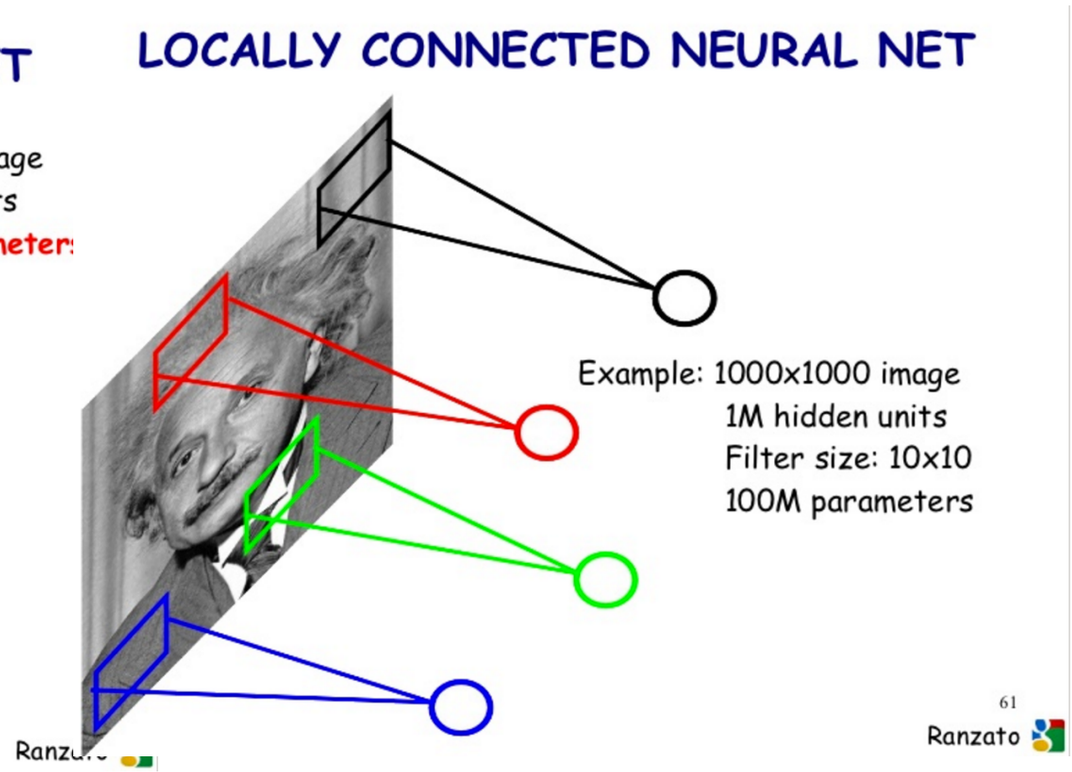
- Yann Lecun, born at Paris, post-doc of Hinton at U Toronto.!
- LeNet on hand-digital recognition at Bell Labs, 5% error rate!



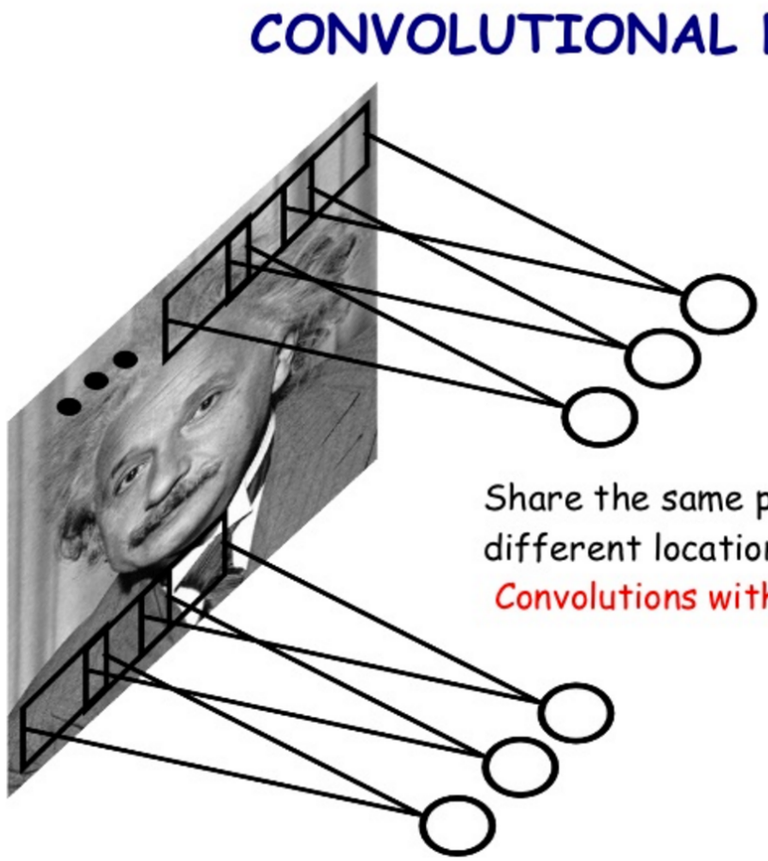
- Attack skill : convolution, pooling!

Local connection!

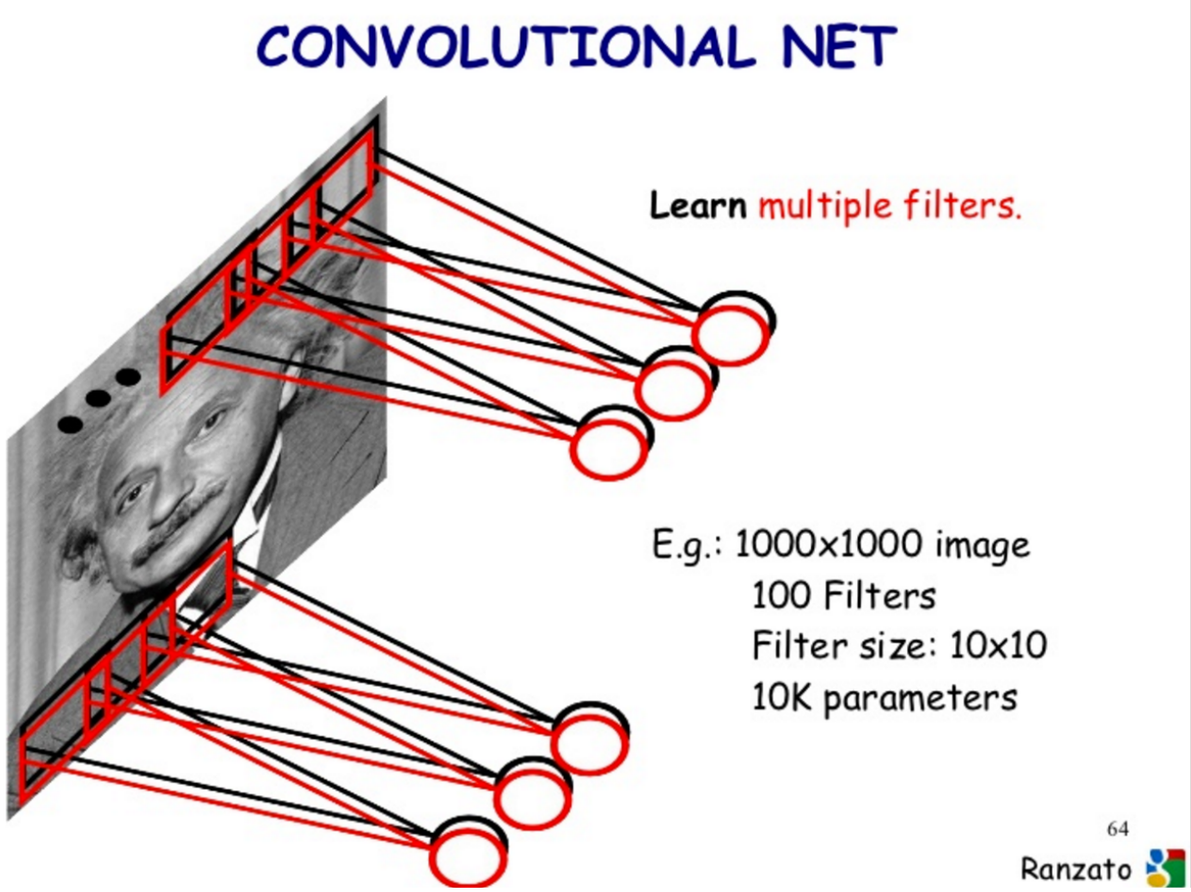




Weight sharing!

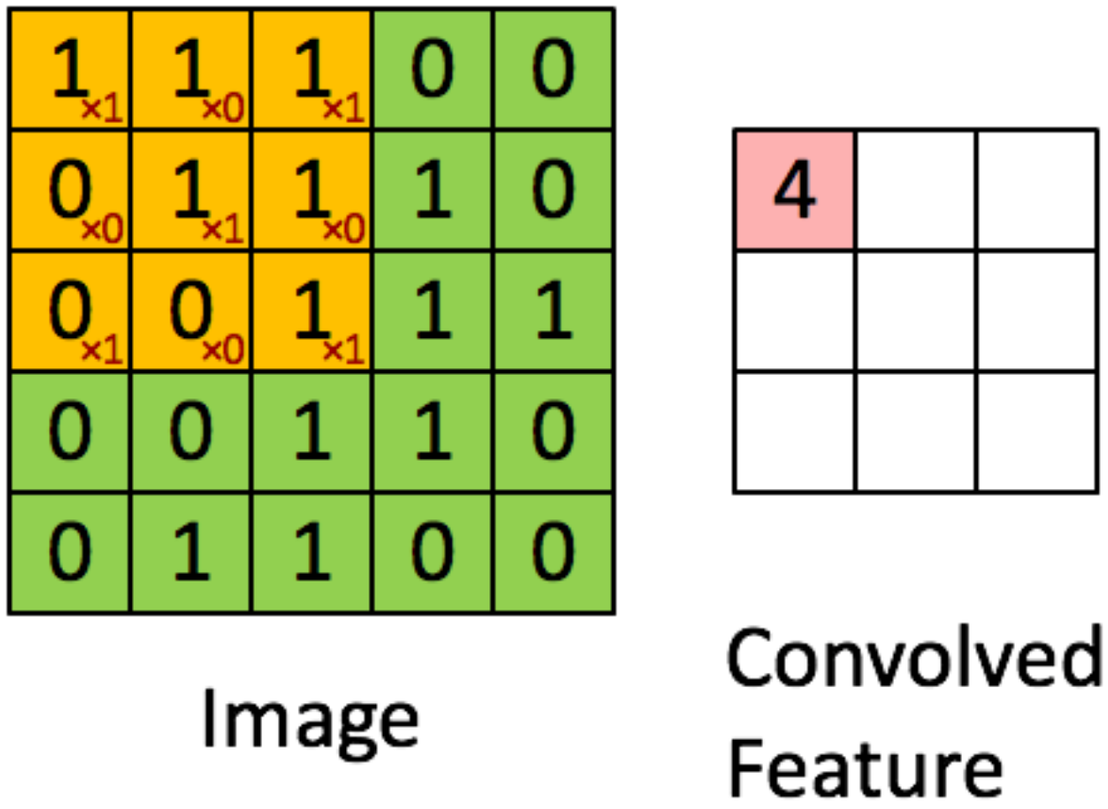


63
Ranzato



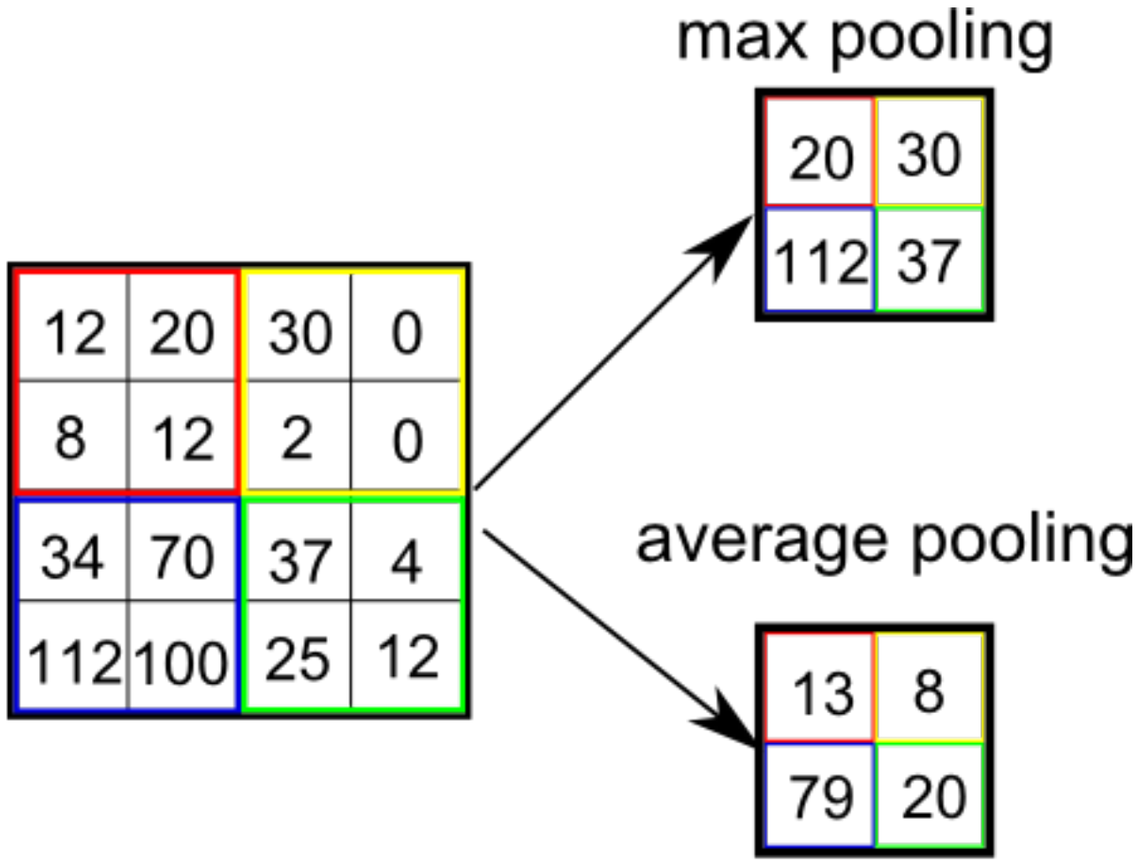
64
Ranzato

Convolution!

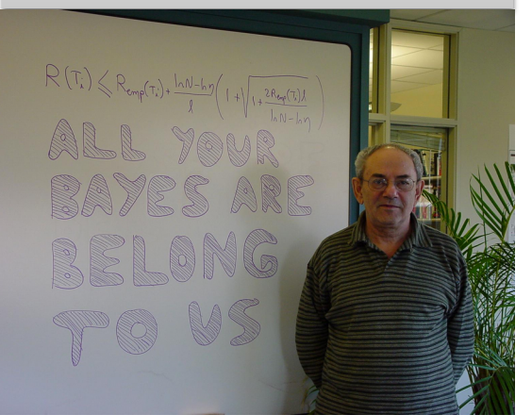


- ? Feature map too big!

Pooling!

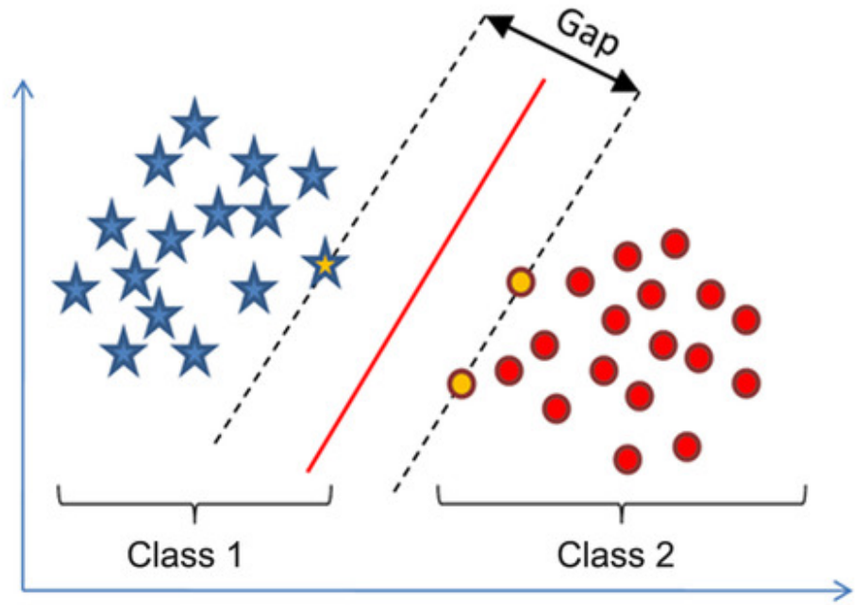


http://vaaaaaanquish.hatenablog.com/entry/2015/01/26/060622!



Support Vector Machine!

- Vladmir Vapnik, born at Soviet Union, colleague of Yann Lecun , invented SVM at 1963!



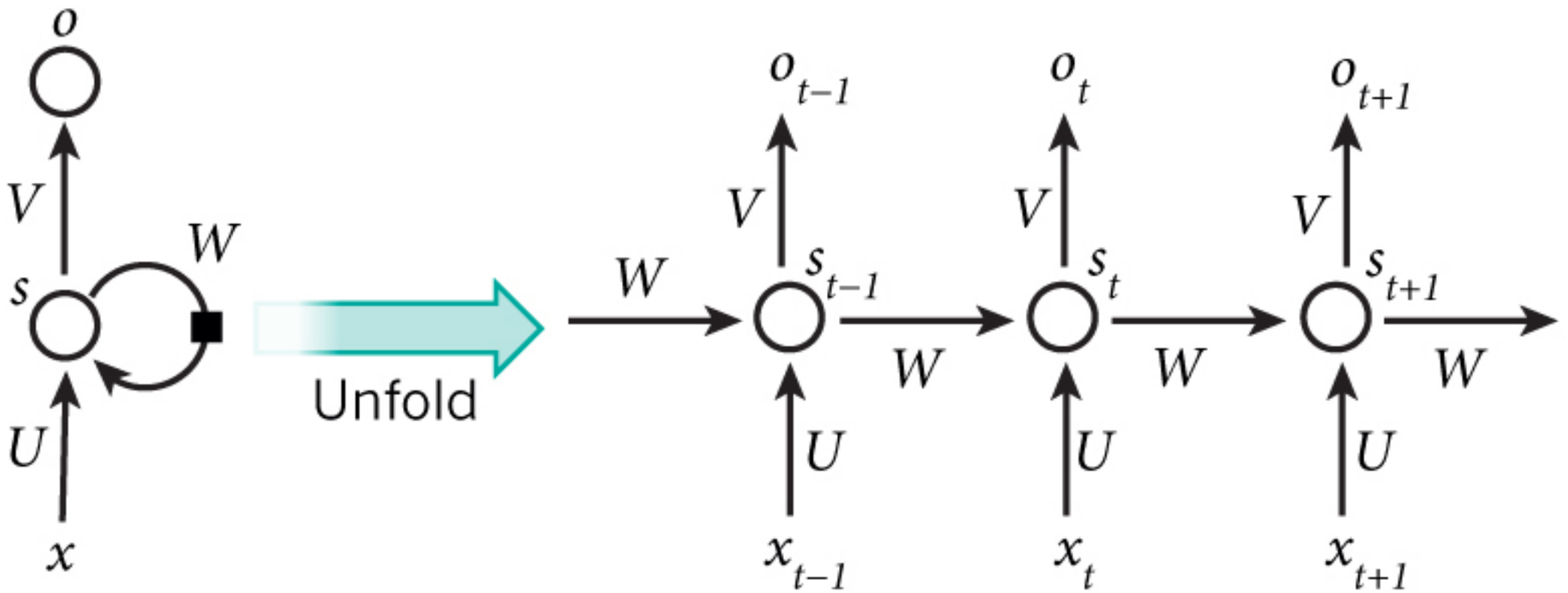
- Attack skill : Max-margin + kernel mapping!

Another winter looming!

- SVM : good at “capacity control”, easy to use, repeat!
- NN : good at modelling complicated structure!

- SVM achieved 0.8% error rate as 1998 and 0.56% in 2002 on the same hand-digit recognition task. !
- NN stops at local optima, hard and takes long to train, overfitting!

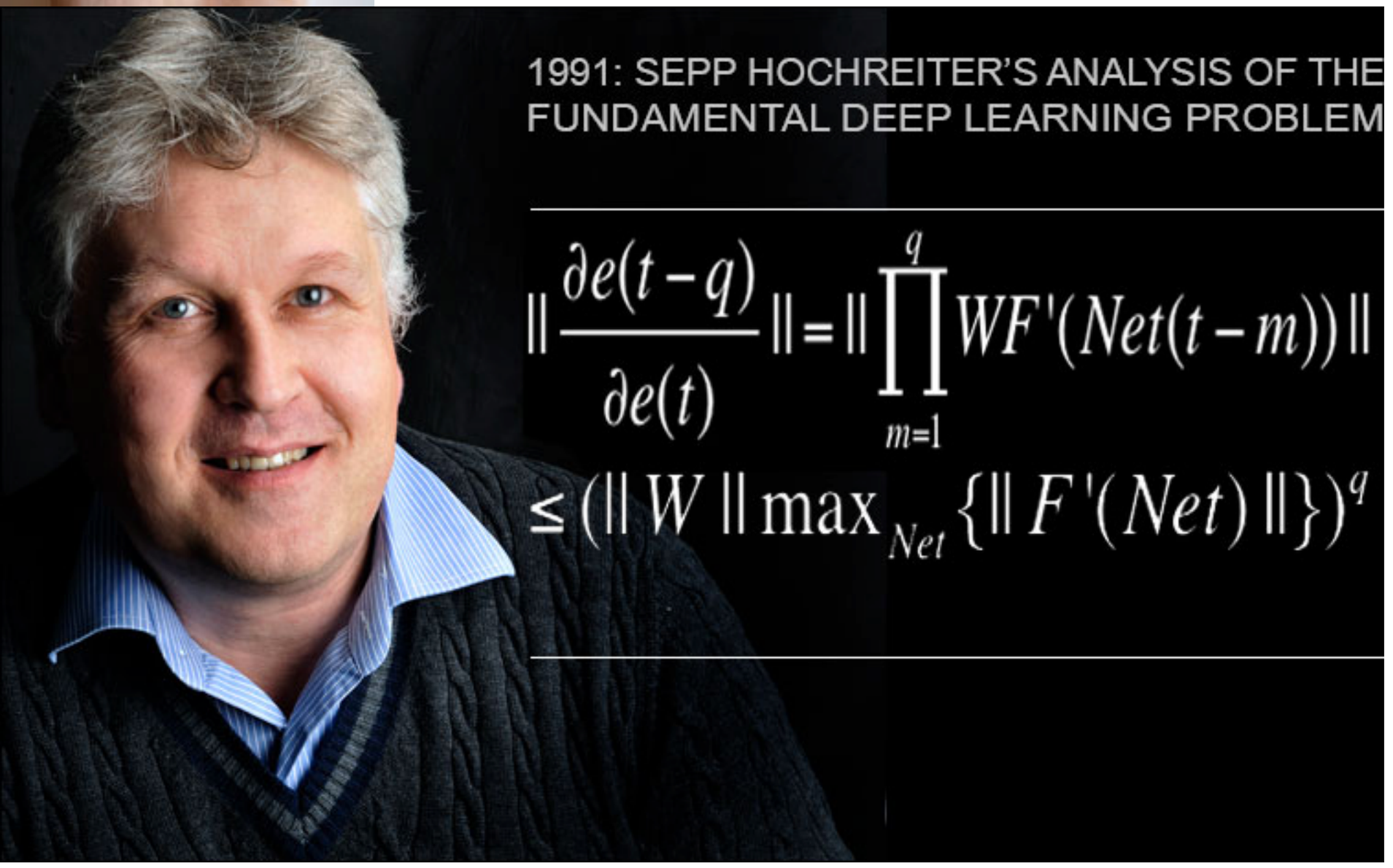
Recurrent Neural Nets!



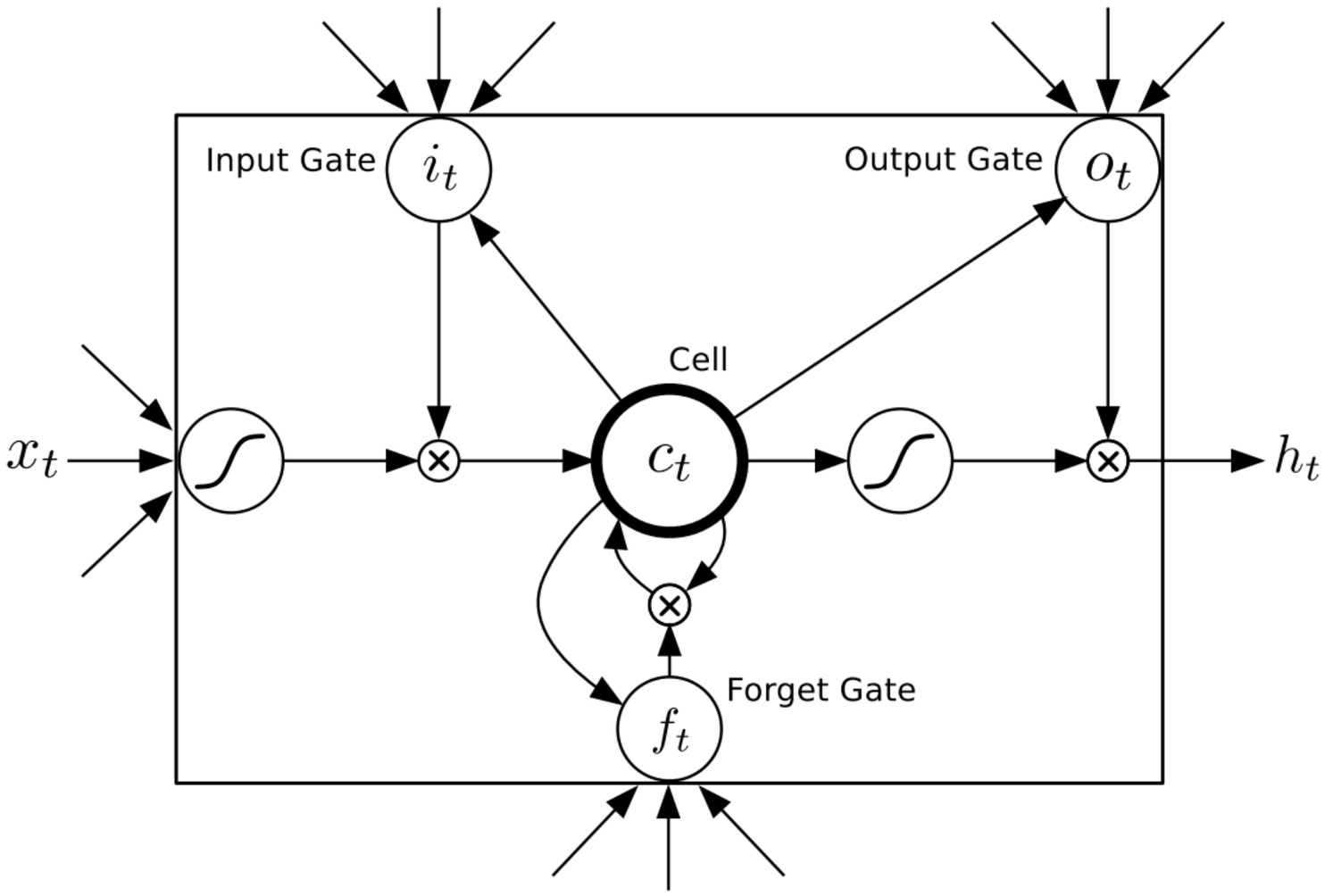
Model sequential data!

http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-2-implementing-a-language-model-rnn-with-python-numpy-and-theano/!

Vanish gradient!



LSTM!



- ? http://www.cs.toronto.edu/~graves/handwriting.html!

10 years funding!

”A fast learning algorithm for deep belief nets”
-- Hinton et al., 2006

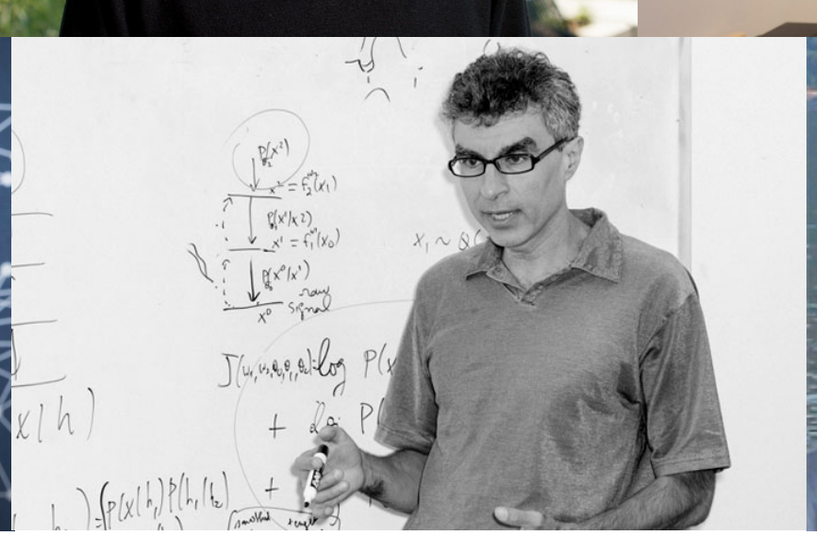
”Reducing the dimensionality of data with neural networks”
-- Hinton & Salakhutdinov

Geoffrey Hinton
University of Toronto

- ? Hinton was happy and rebranded its neural network to deep learning. The story goes on ……..!

Other Pioneers!

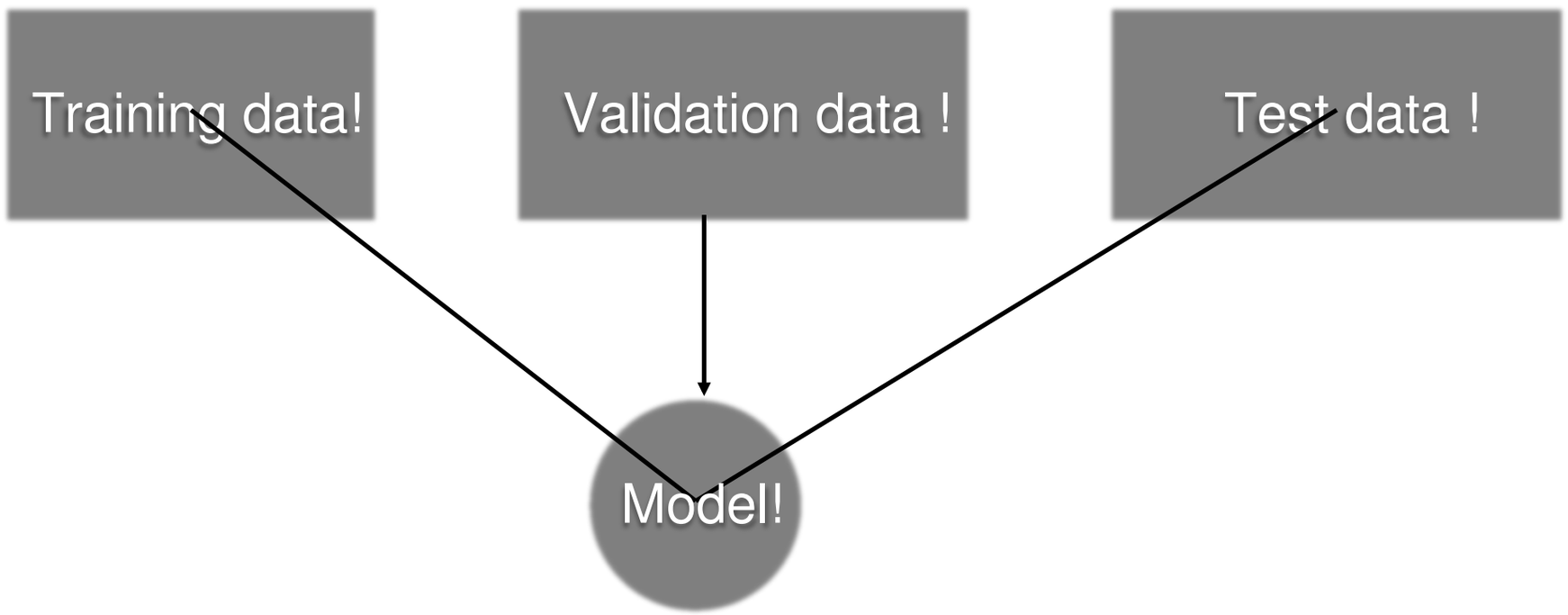





21!

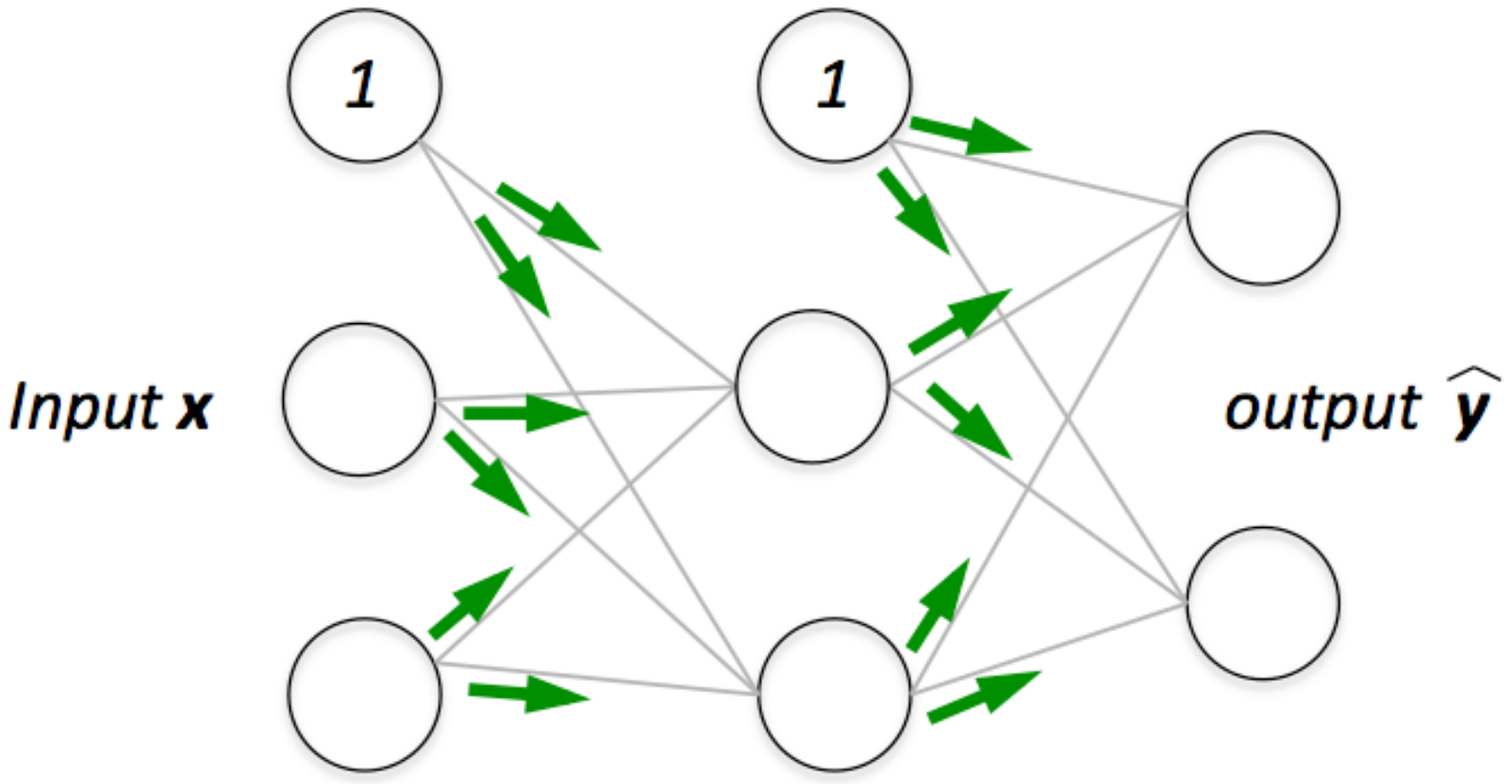
Data preparation!

- Input : Can be image, video, text, sounds….!
- Output : Can be a translation of another language, location of an object, a caption describing a image ….!



Training !

- ? Forward pass, propagating information from input to ouput!



Training!

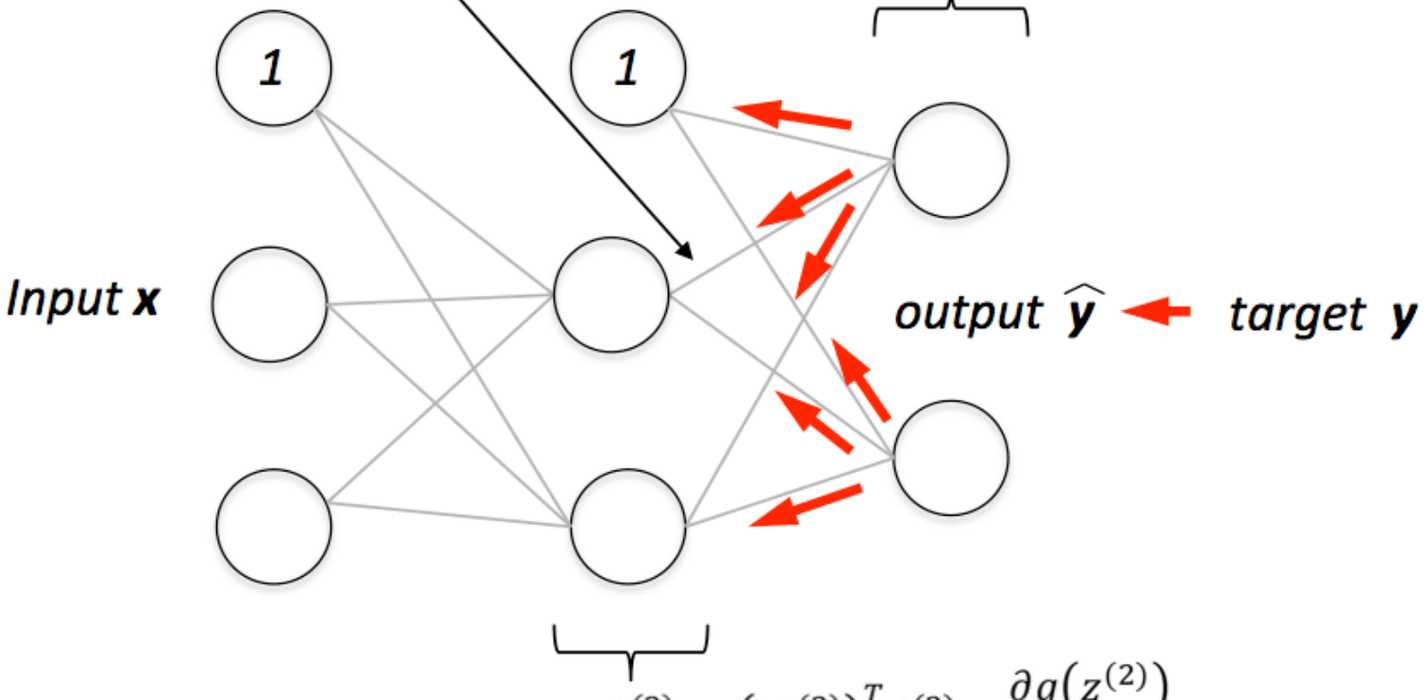
- ? Backward pass, propagating errors back and update weights accordingly!

$$\frac{\partial}{\partial w_{i,j}^{(l)}} J(\boldsymbol{W}) = a_j^{(l)} \delta_i^{(l+1)}$$

(compute gradient)

(error term of the output layer)

$$\boldsymbol{\delta}^{(3)} = \boldsymbol{a}^{(3)} - \boldsymbol{y}$$

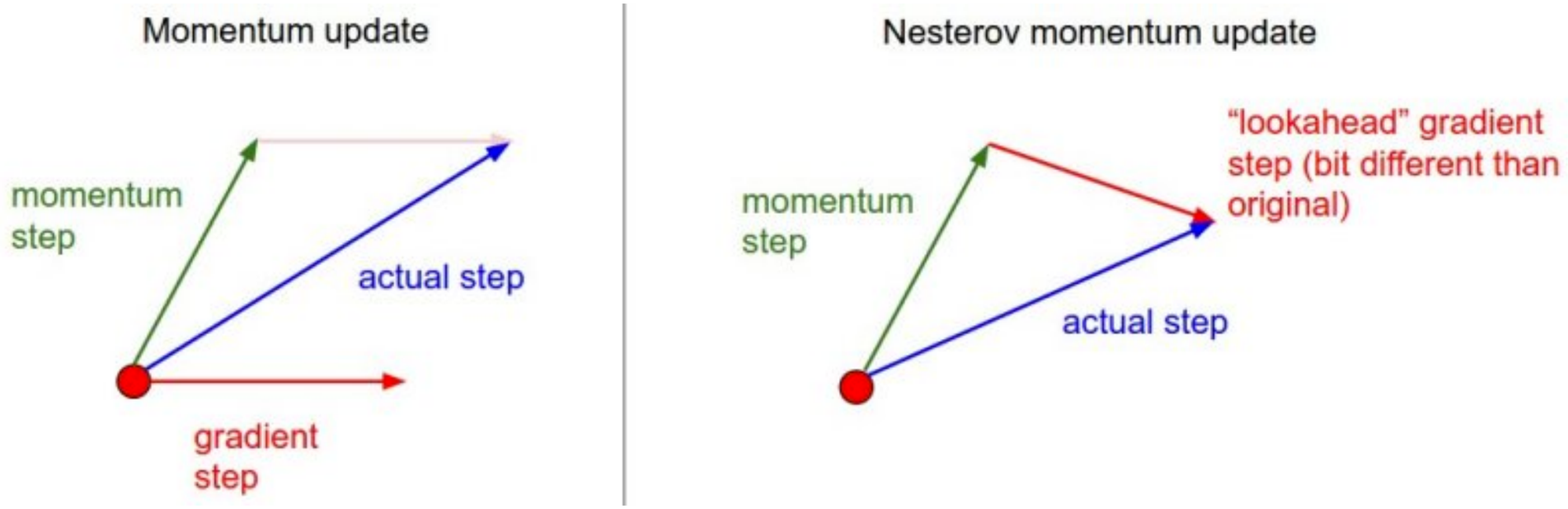


$$\boldsymbol{\delta}^{(2)} = \left(\boldsymbol{W}^{(2)}\right)^T \boldsymbol{\delta}^{(3)} * \frac{\partial g\left(z^{(2)}\right)}{\partial z^{(2)}}$$

(error term of the hidden layer)

Parameter update!

- Gradient descent (sgd, mini-batch)!
- x = x - lr * dx!

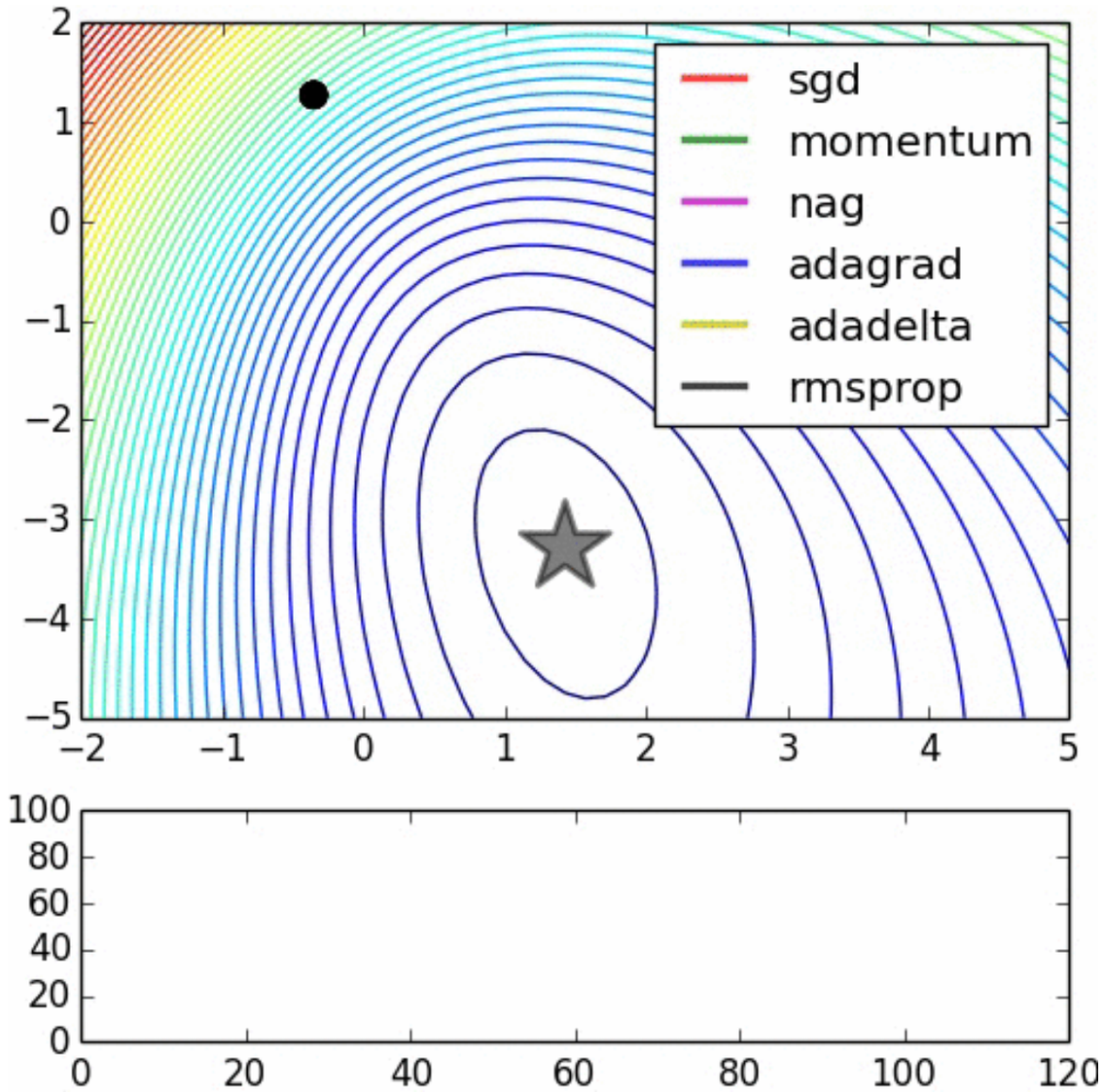


source : http://cs231n.stanford.edu/!

animation by Alec Radford!

Parameter update!

- ? Gradient descent!
- ? x = x - lr * dx!



animation by Alec Radford)!

- ? Second order method ? Quasi-Newton’s method!

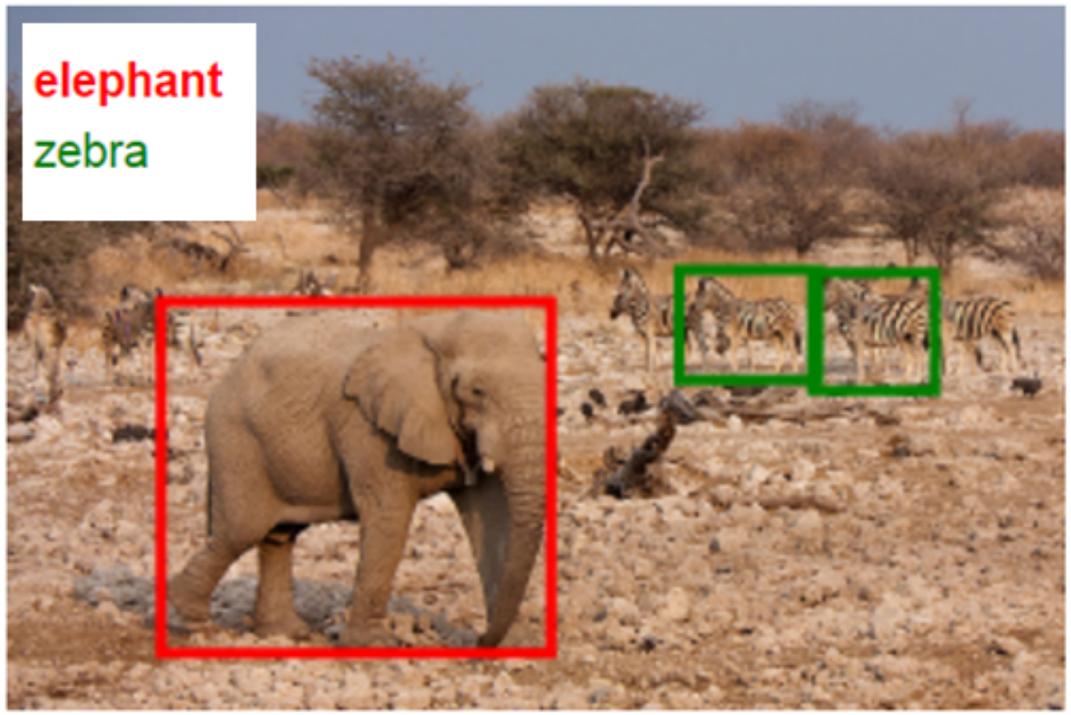
Problems!

- While neural networks can approximate any function of input X to output y, it is very hard to train with multiple layers!
- Initialization!
- Pre-training!
- More data/Dropout to avoid overfitting!
- Accelerate with GPU!
- Support training in parallel!

Big data + Big computational power!

ImageNet!

ILSVRC2014_train_00012929

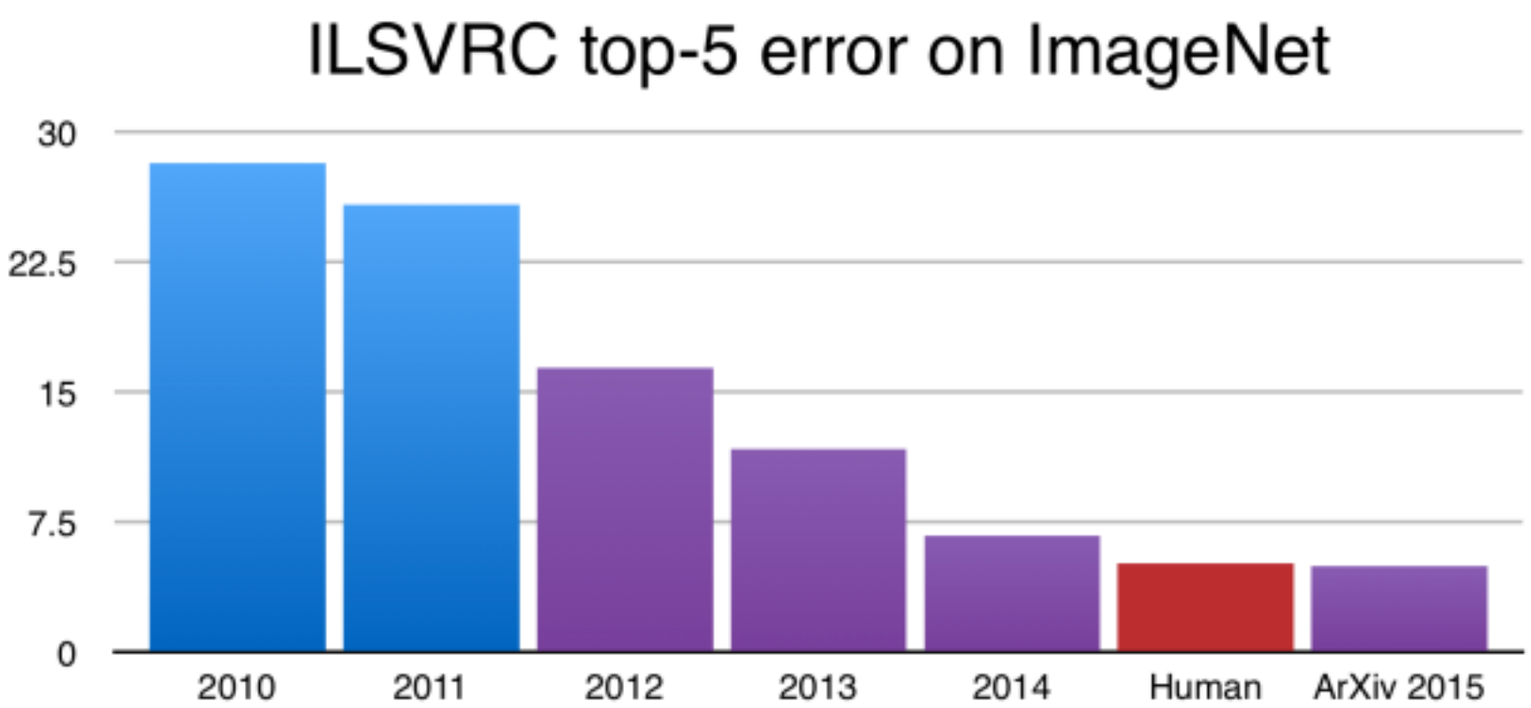


ILSVRC2014_train_00038948

- ? >1M well labeled images of 1000 classes!

- ? source : https://devblogs.nvidia.com/parallelforall/nvidia-ibm-cloud-support-imagenet-large-scale-visual-recognition-challenge/!

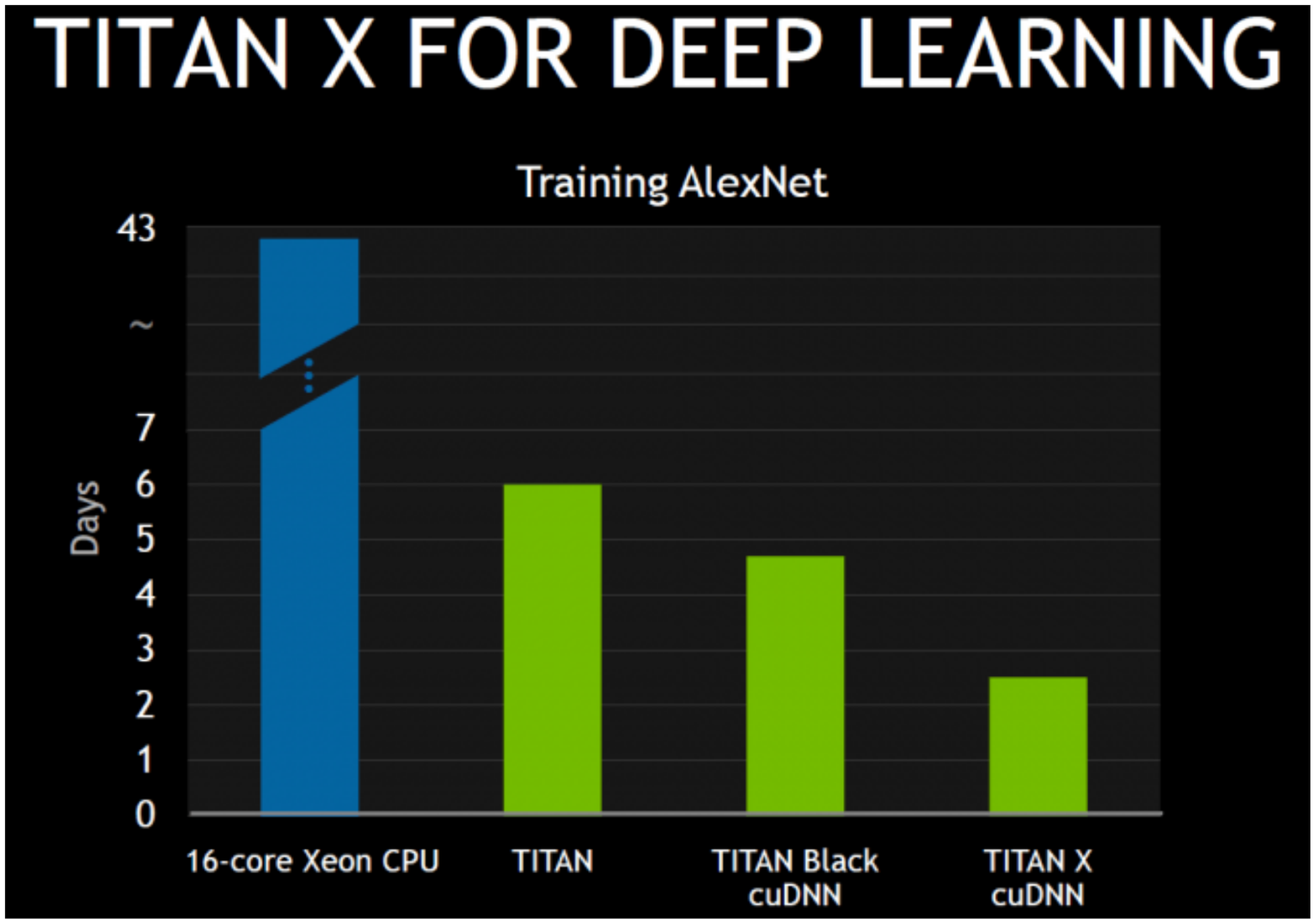
Big data + Big computational power!



29!

Big data + Big computational power!

GPUs!

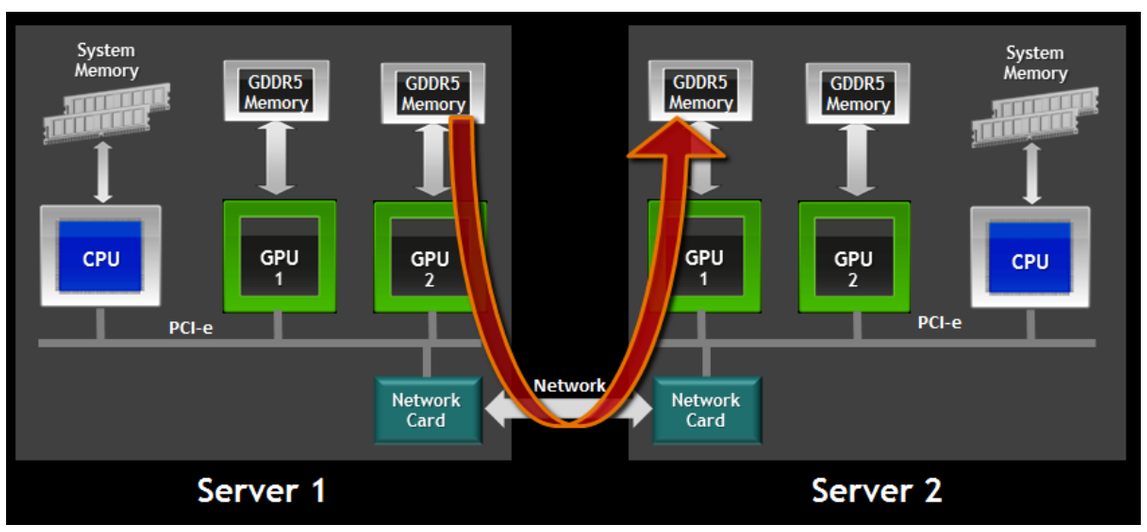


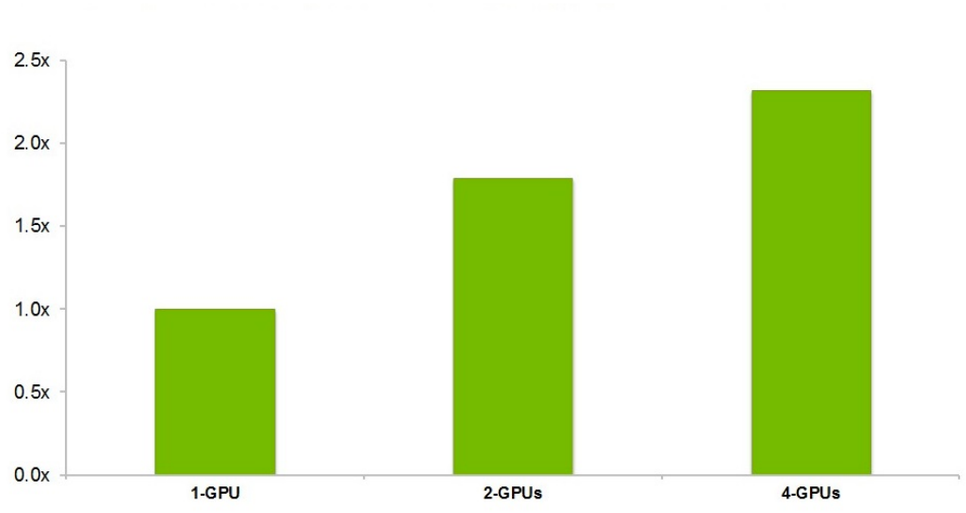
- ? source : http://techgage.com/article/gtc-2015-in-depth-recap-deep-learning-quadro-m6000-autonomous-driving-more/!

Big data + Big computational power!

Parallelization!

- Data Parallelization: same weight but different data, gradient needs to be synchronized!
  - Not scale well with size of cluster!
  - Works well on smaller clusters and easy to implement !





source : Nvidia!

Big data + Big computational power!

Parallelization!

- Model Parallelization: same data, but different parts of weight!

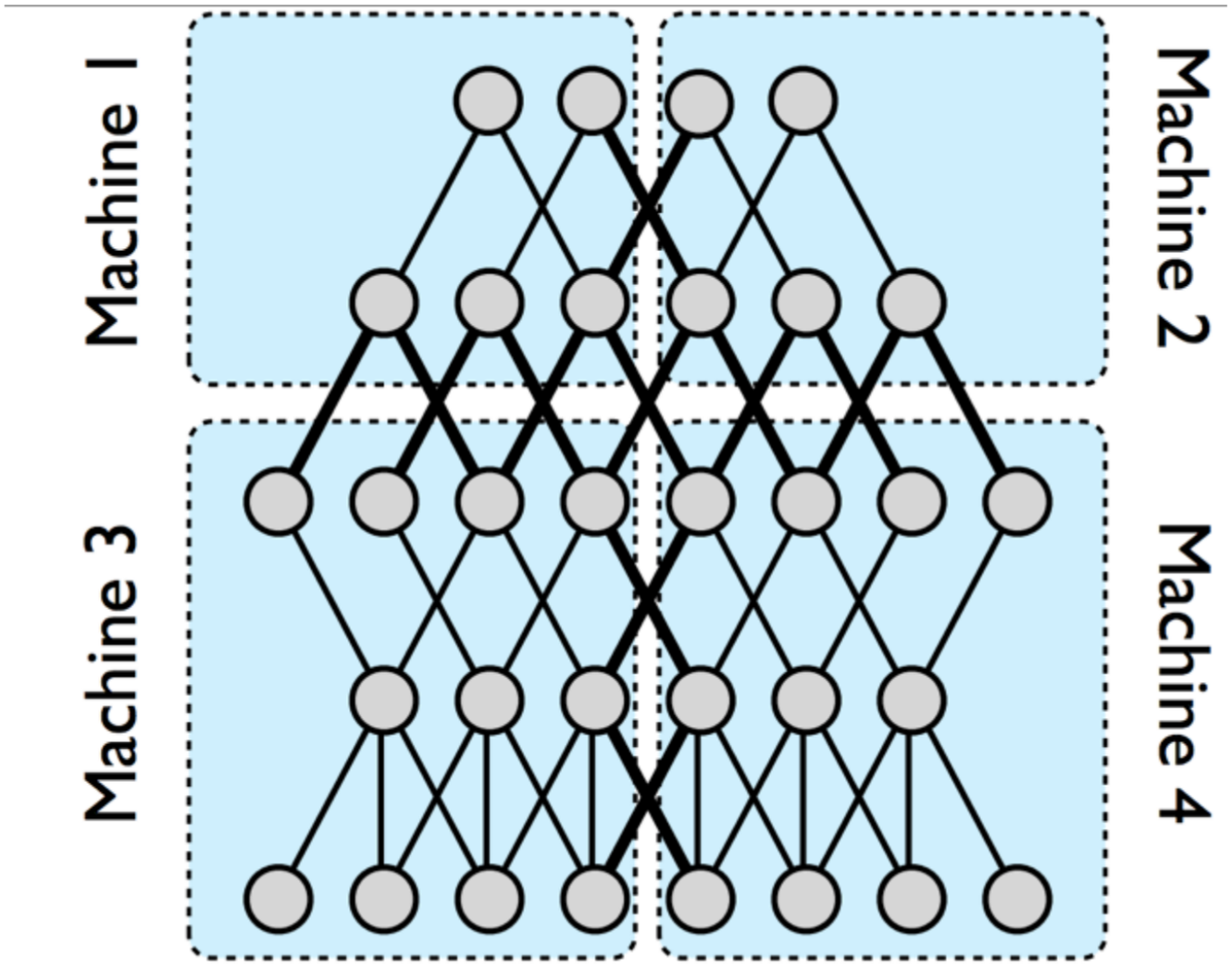
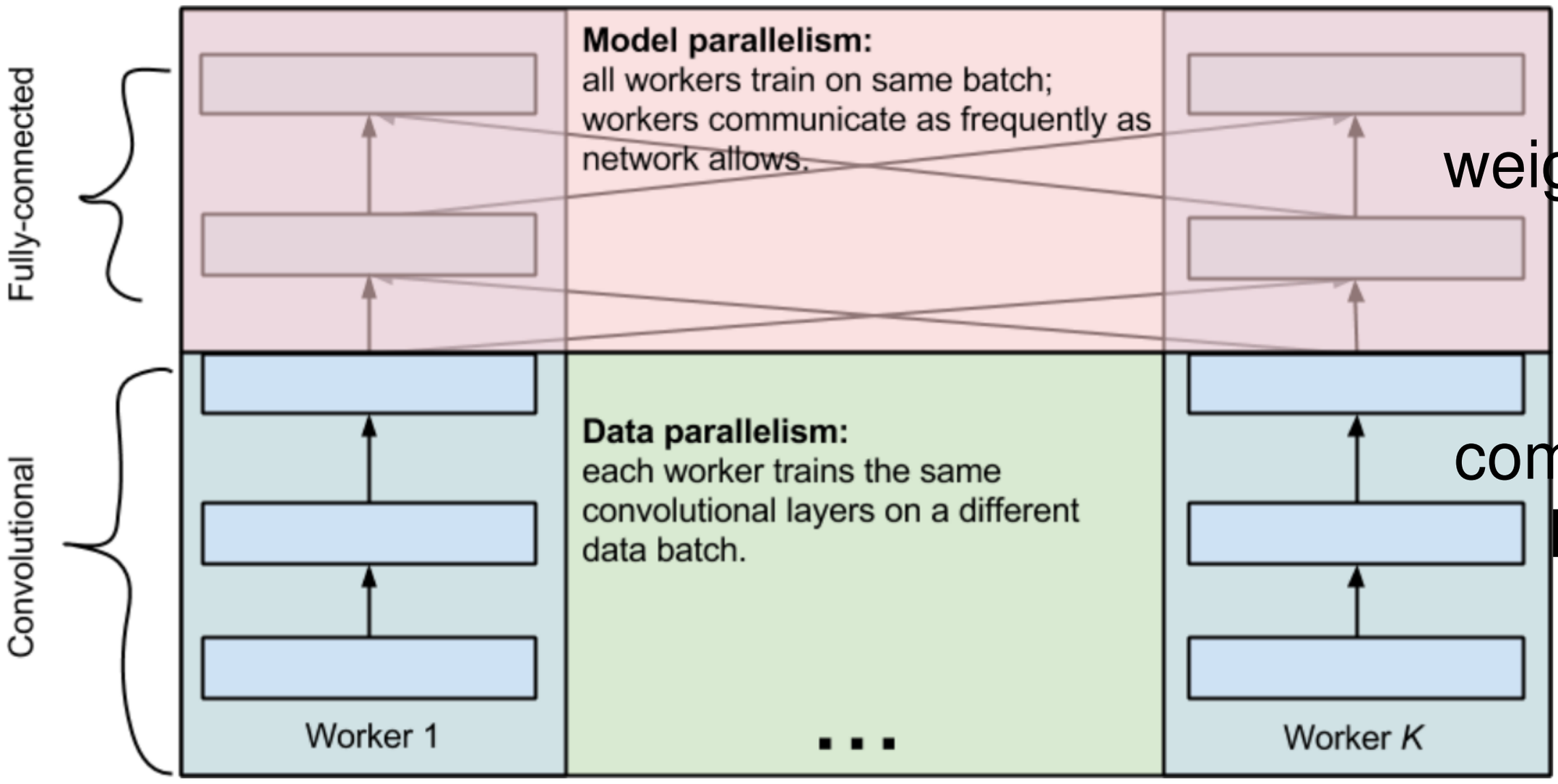


Figure from Yi Wang!

- Parameters in a big Neural net can not fit into memory of a single GPU!

Big data + Big computational power!

Model+Data Parallelization!



weight heavy!

computation!
heavy!

Krizhevsky et al. 2014!

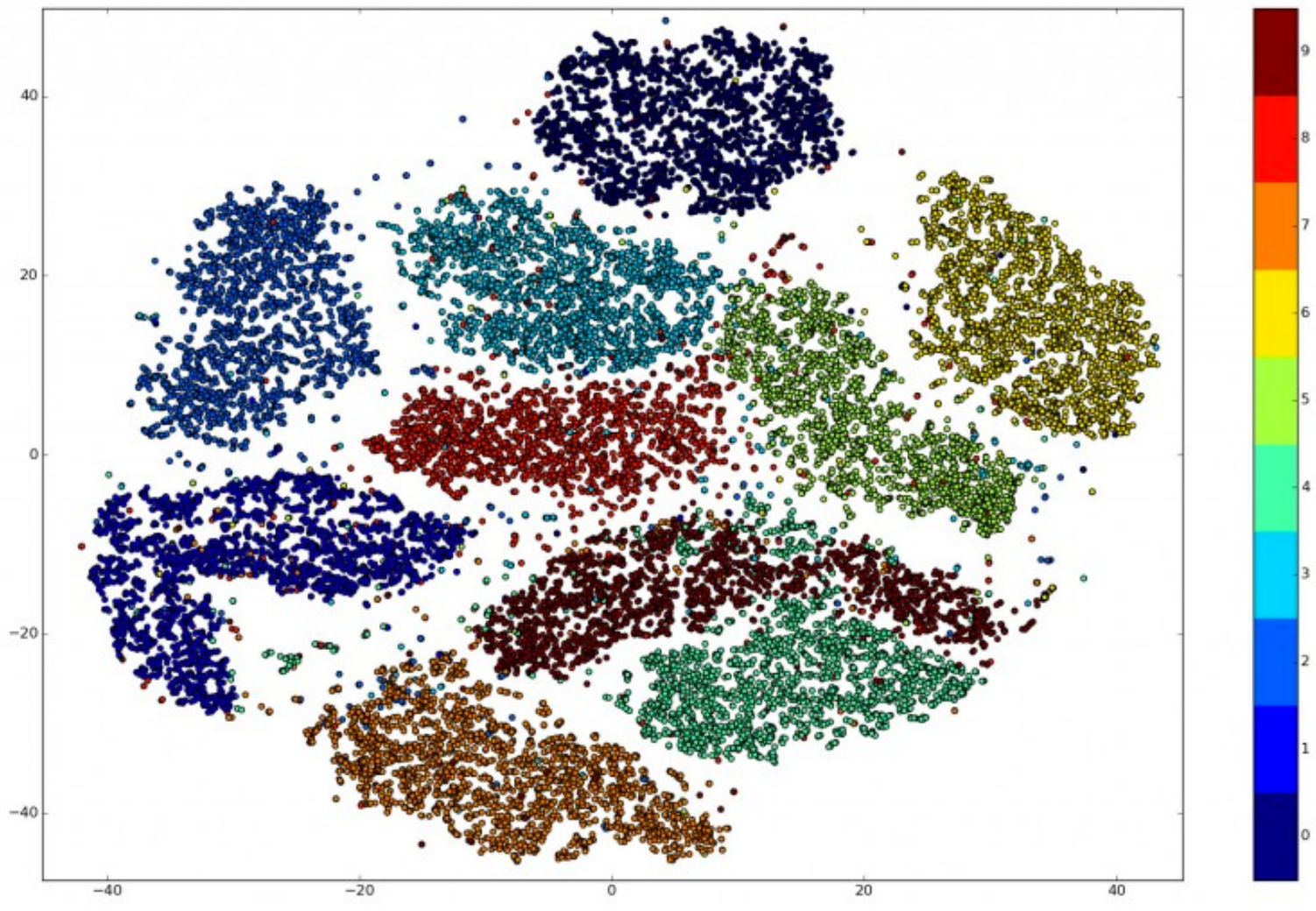
Big data + Big computational power!

AlphaGO	Lee Se-dol
1202 CPUs, 176 GPUs,	1 Human Brain,
100+ Scientists.	1 Coffee.

- ? Trained on 30M moves + Computation + Clear Reward scheme!

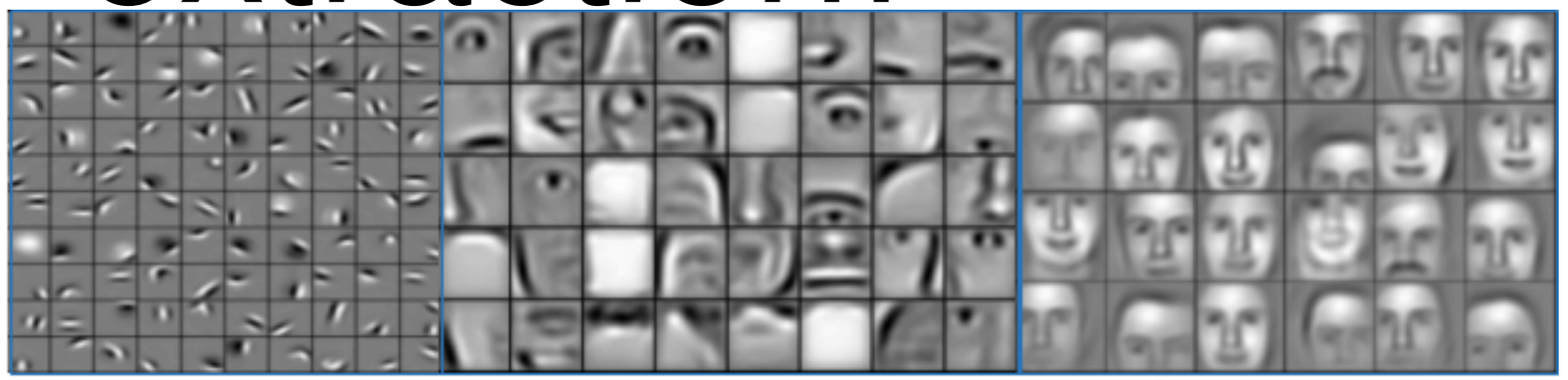
Feature extraction!

- ? Task : A classifier to recognize hand written digits!

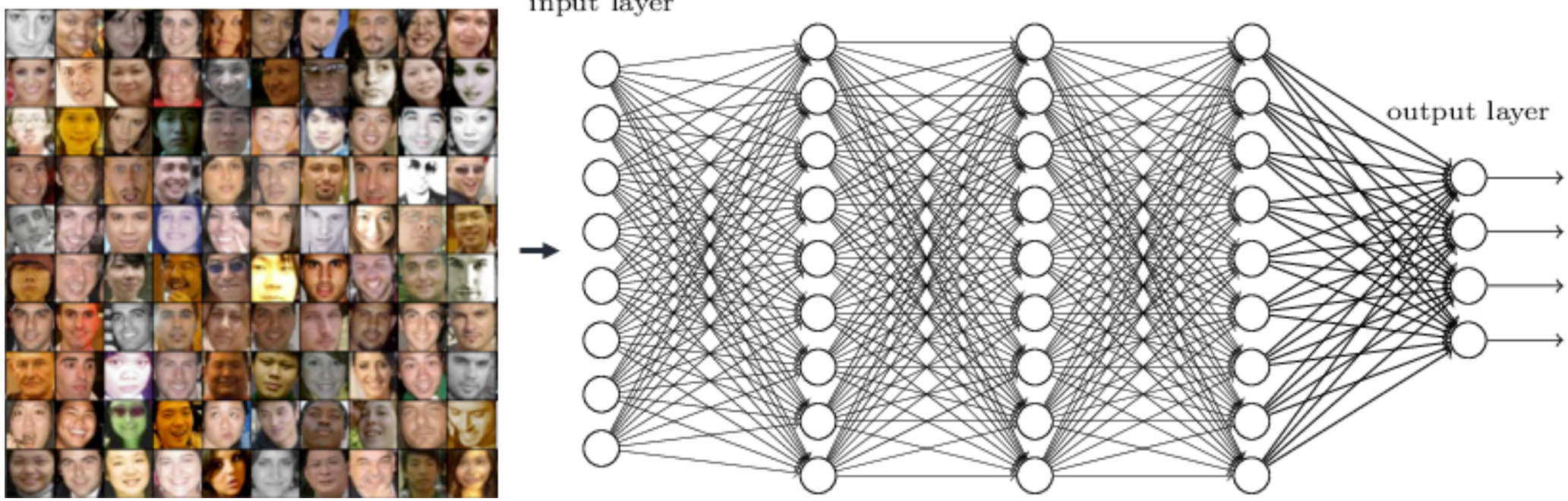


source :https://indico.io/blog/visualizing-with-t-sne/!

Hierarchal feature extraction!



Deep neural networks learn hierarchical feature representations

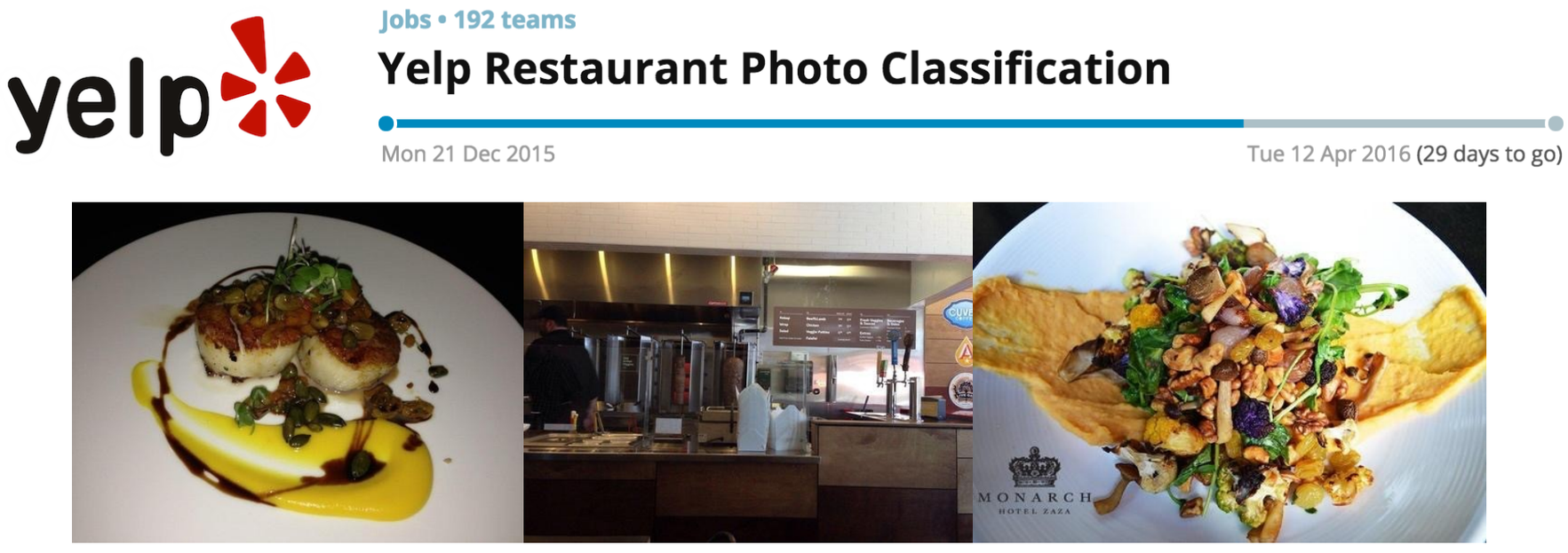


- ? source : http://www.rsipvision.com/exploring-deep-learning/!

Feature extraction!

Transfer learning!

- Use features learned on ImageNet to tune your only classifier on other tasks.!

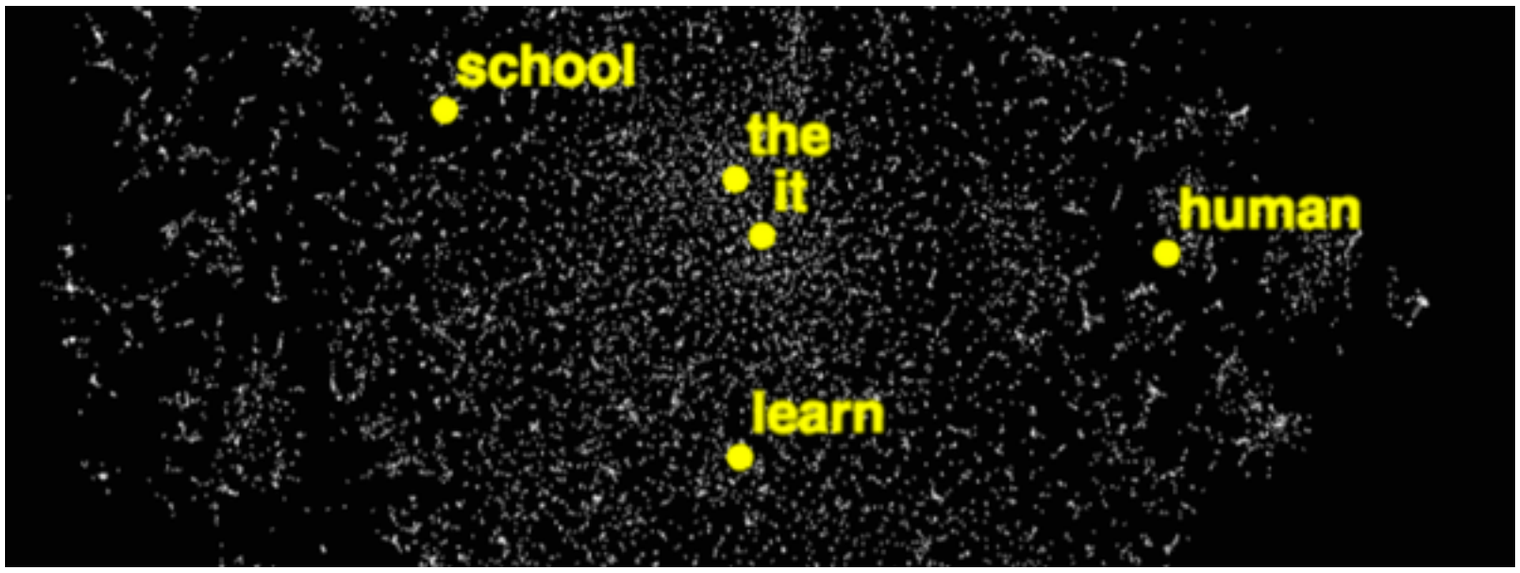


- 9 categories + 200, 000 images!

- extract features from last 2 layers + linear SVM!

Word embedding!

- Embed words into vector space such that similar words are more close to each other in vector space!



- Calculate cosine similarity !

Pairs - France+ Italy = Rome!

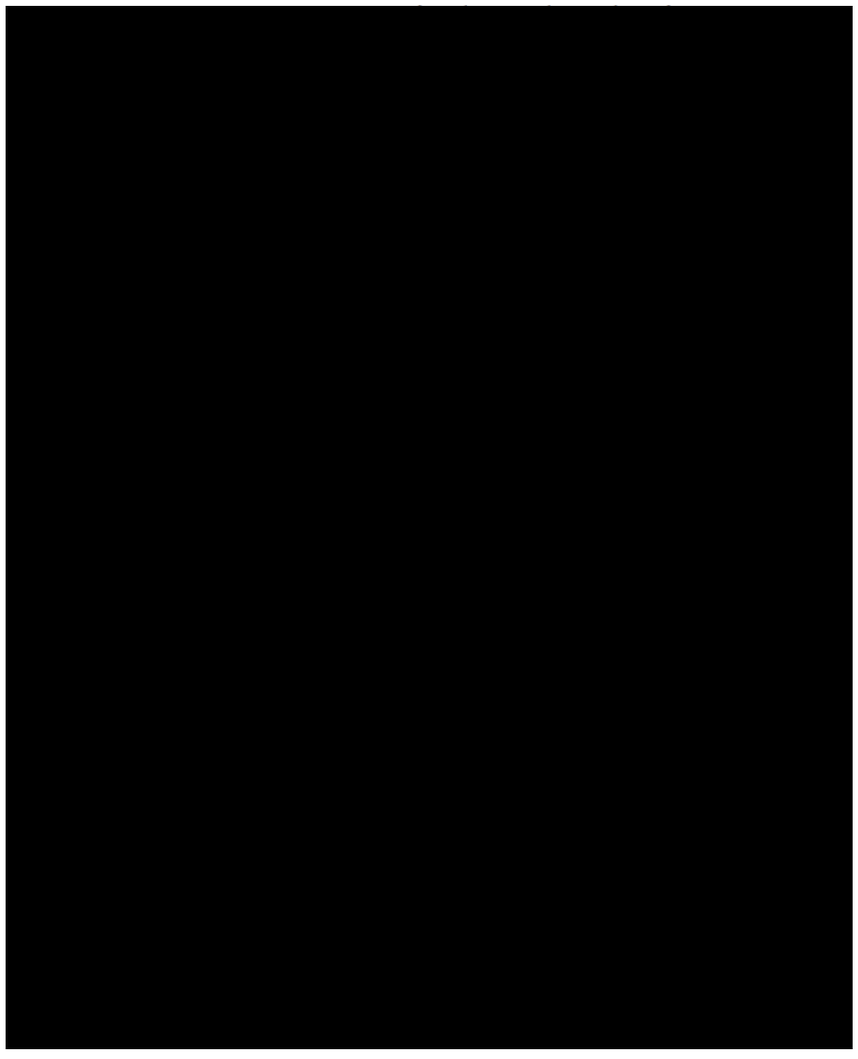
source : http://anthonygarvan.github.io/wordgalaxy/!

Examples!

Playing Atari with Deep Reinforcement Learning

Volodymyr Mnih **Koray Kavukcuoglu** **David Silver** **Alex Graves** **Ioannis Antonoglou**

Daan Wierstra **Martin Riedmiller**



Go!

- ? 3^{361} different legal positions, roughly 10^{170}!

  Brute Force does not work !!!!

- ? Monto Carlo Tree Search !

  Heuristic Searching method!

  Still slow since the search tree is so big!

AlphaGo!

- Offline learning : from human go players, learn a Deep learning based **policy network** and a **rollout policy**

- Use self-playing result to train a **value network**, which evaluate the current situation!

Online Playing!

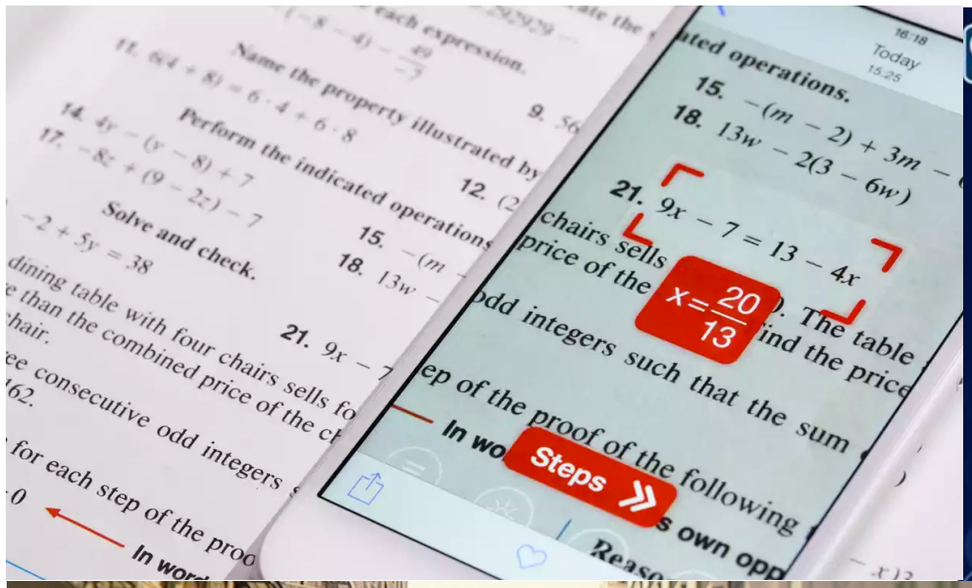
- ? When playing with Lee Sedol, policy network give possible moves (about 10 to 20), the value network calculates the weight of position.!

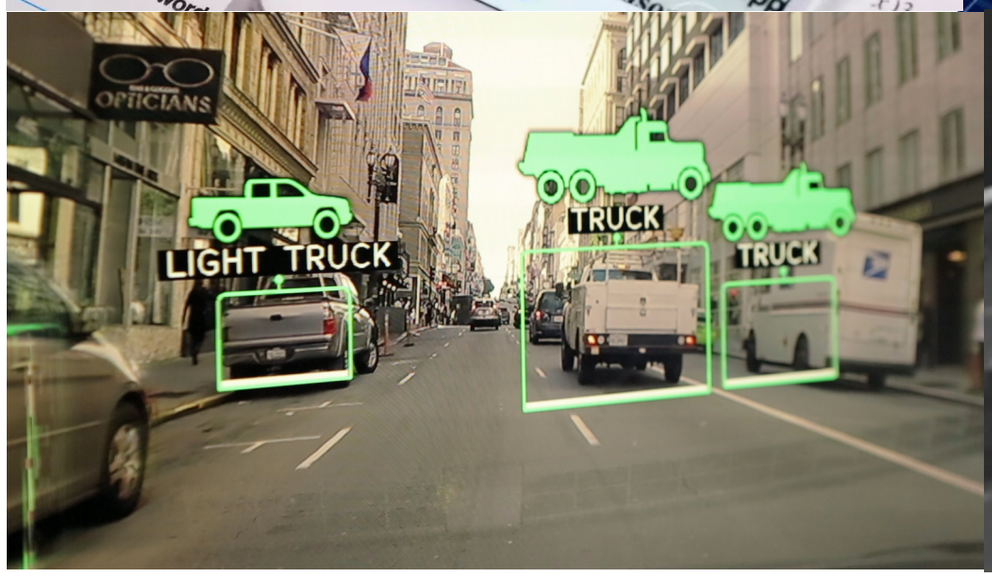
- ? MCTS update the weight of each possible moves and select one move.!

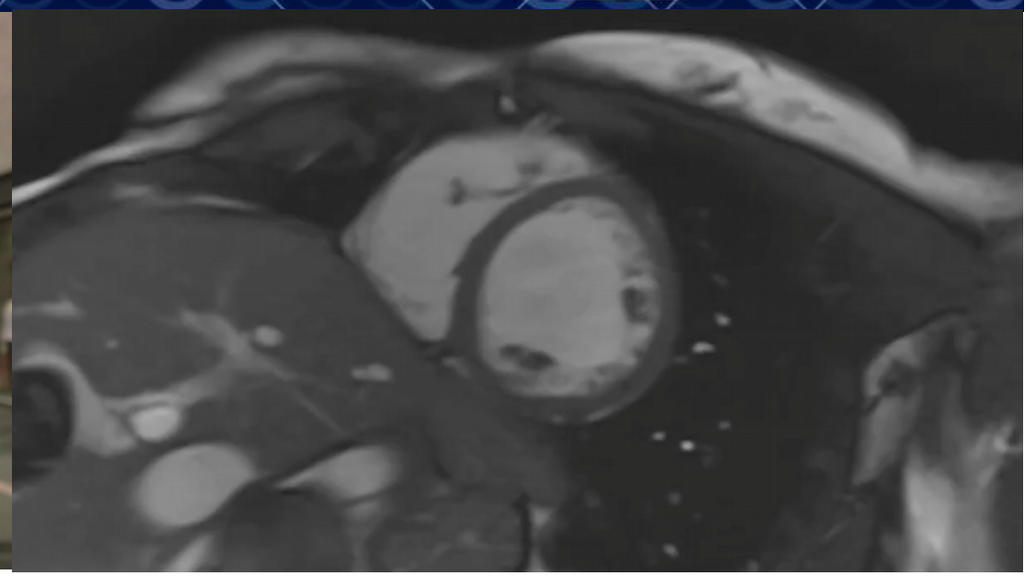
- ? **The engineering effort on systems with CPU and GPU with data , model parallelism matters**

Examples!

Image recognition!







Train in a supervised fashion !

Examples!

Caption generation!

A [[[woman|2]]] in a white [[[dress|0]]] and gold [[[boots|5]]] leaning on a [[[car|3]]] .

A [[[woman|2]]] poses along a [[[car|3]]] .

[[[Woman|2]]] dressed in white with gold [[[boots|5]]] poses next to a police [[[car|3]]]

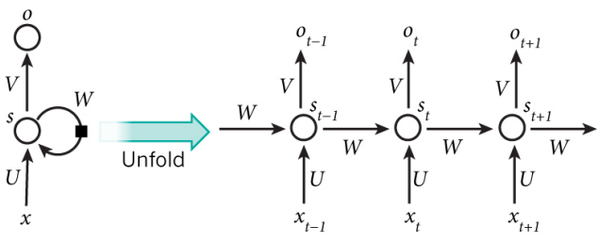
A [[[woman|2]]] dressed in white leans against a white [[[car|3]]] .

A [[[woman|2]]] is leaning against a [[[car|3]]] .

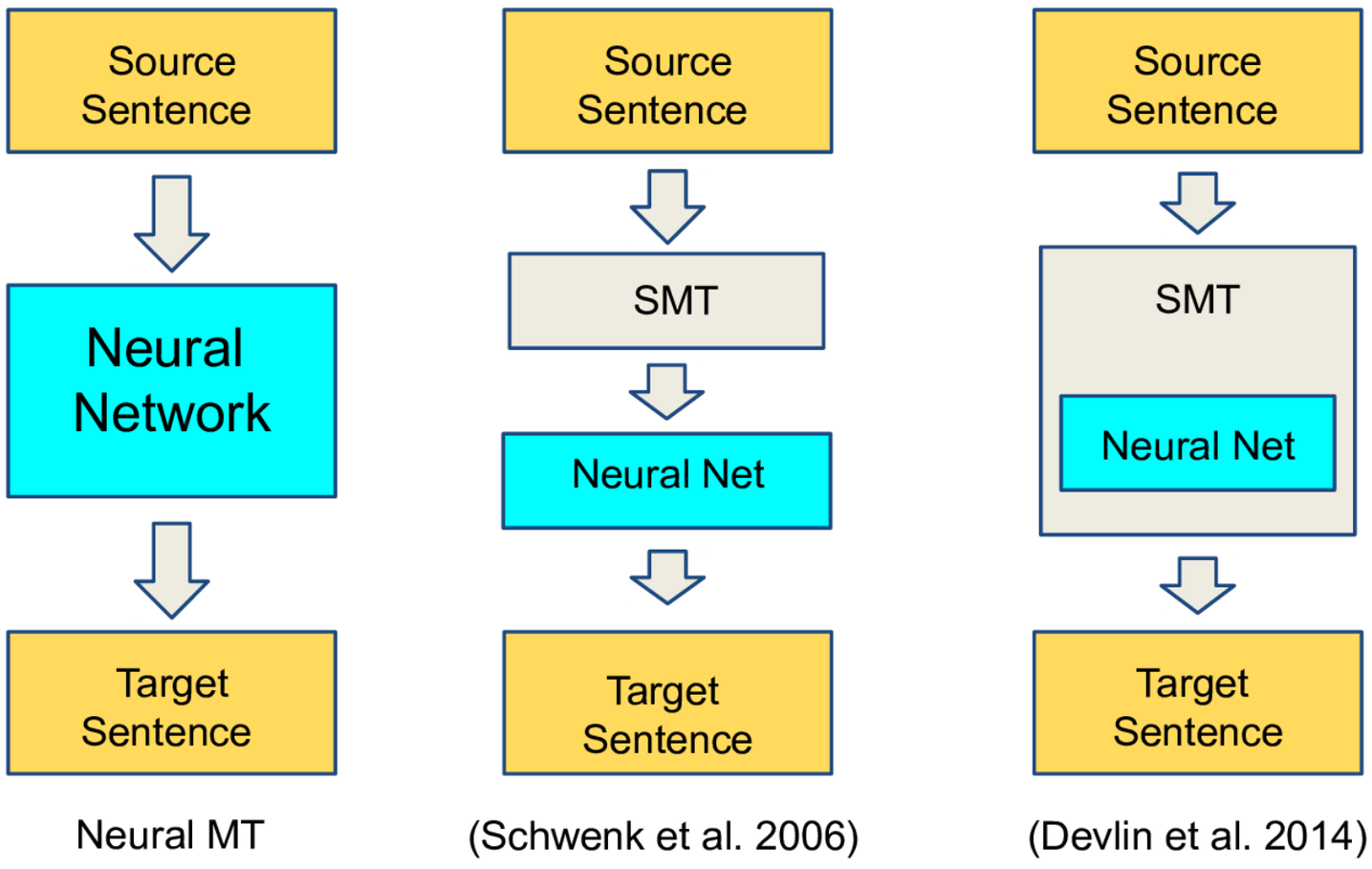
A [[[woman|2]]] with gold [[[boots|5]]] leans against an Indy pace [[[car|3]]] .

A blonde [[[woman|2]]] wearing gold shiny [[[boots|5]]] , a white [[[top|0]]] and short white skirt is leaning on a [[[car|3]]] .

Examples!



Neural MT!



- ? The main idea behind RNNs is to compress a sequence of input symbols into a fixed-dimensional vector by using recursion.!

source : https://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-with-gpus/!

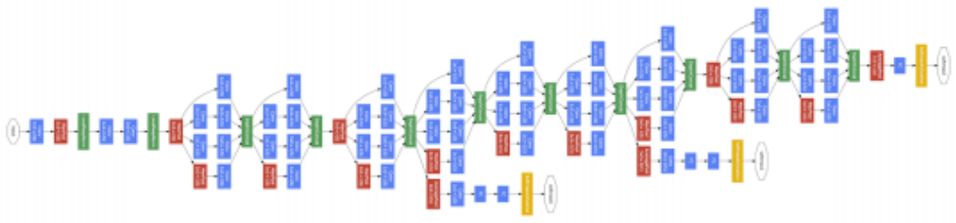
Examples!

Neural arts!

46!

Updates of deep learning!

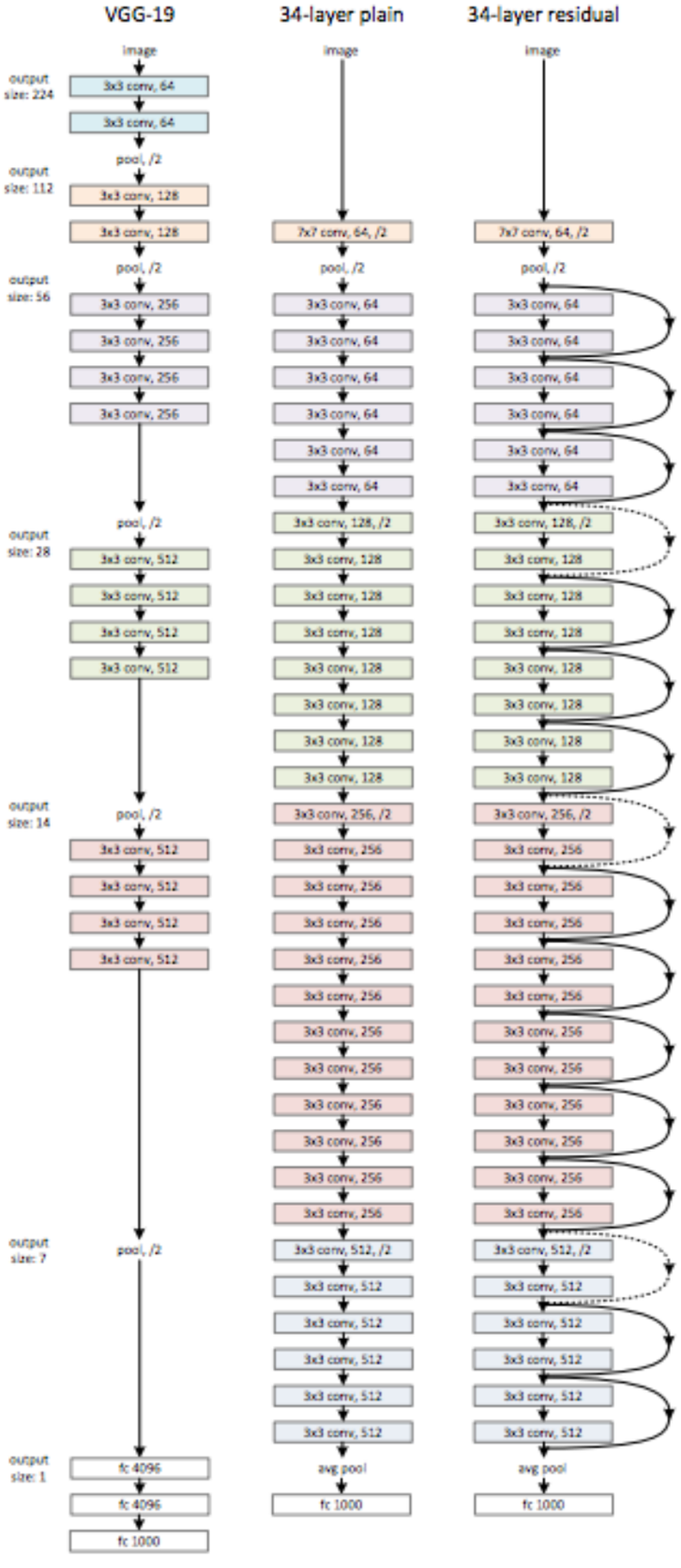
Go deeper



[1]Inception 5 (GoogLeNet)

Inception 7a

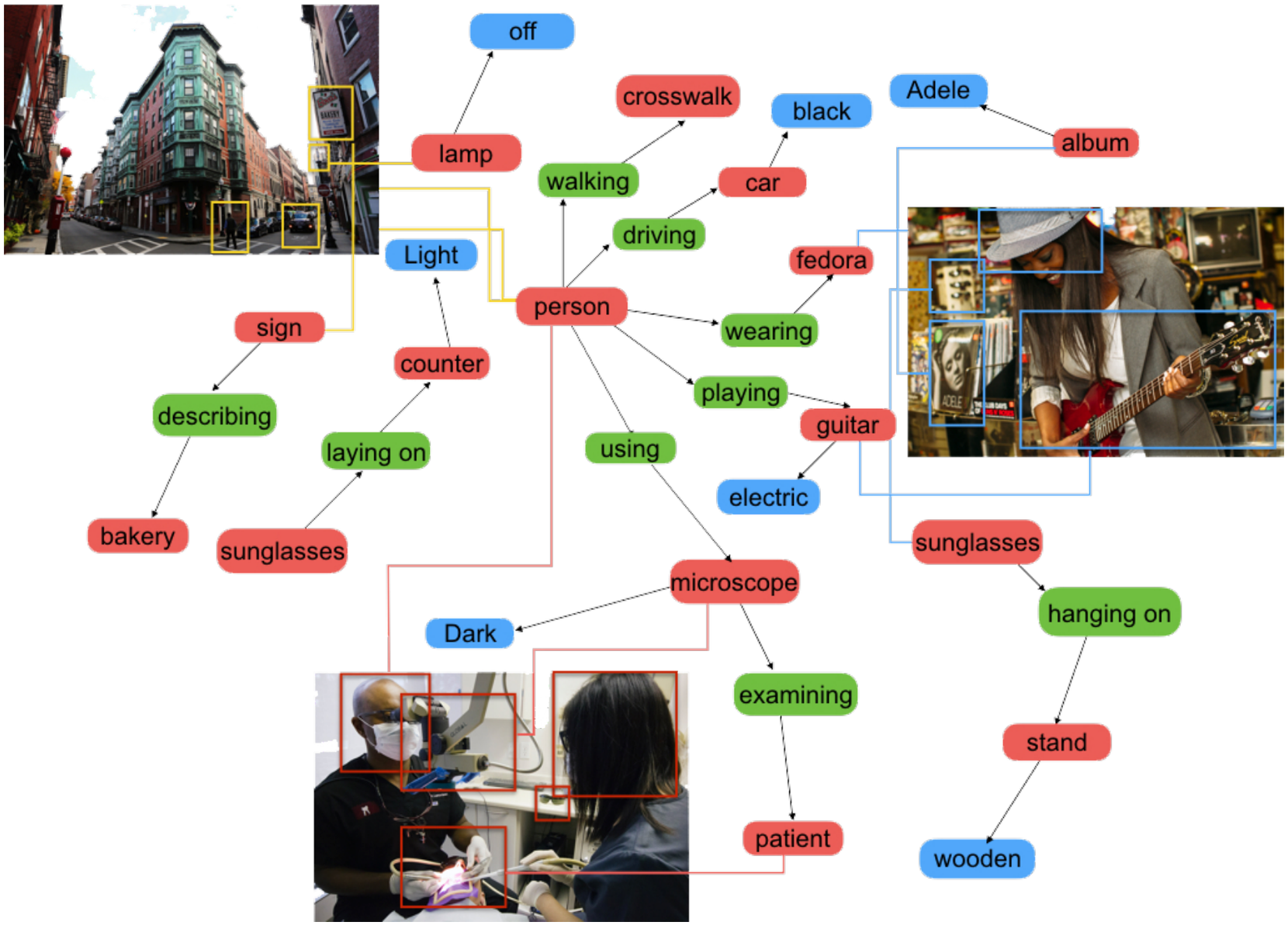
[1]Going Deeper with Convolutions, [C. Szegedy et al, CVPR 2015]

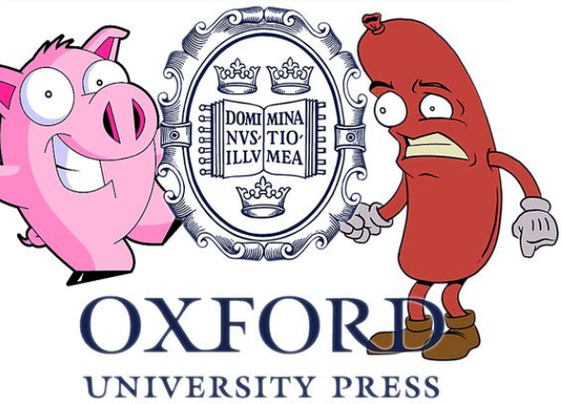


- Deep Residual Learning for Image Recognition, MSRA, 2015!
  - ?

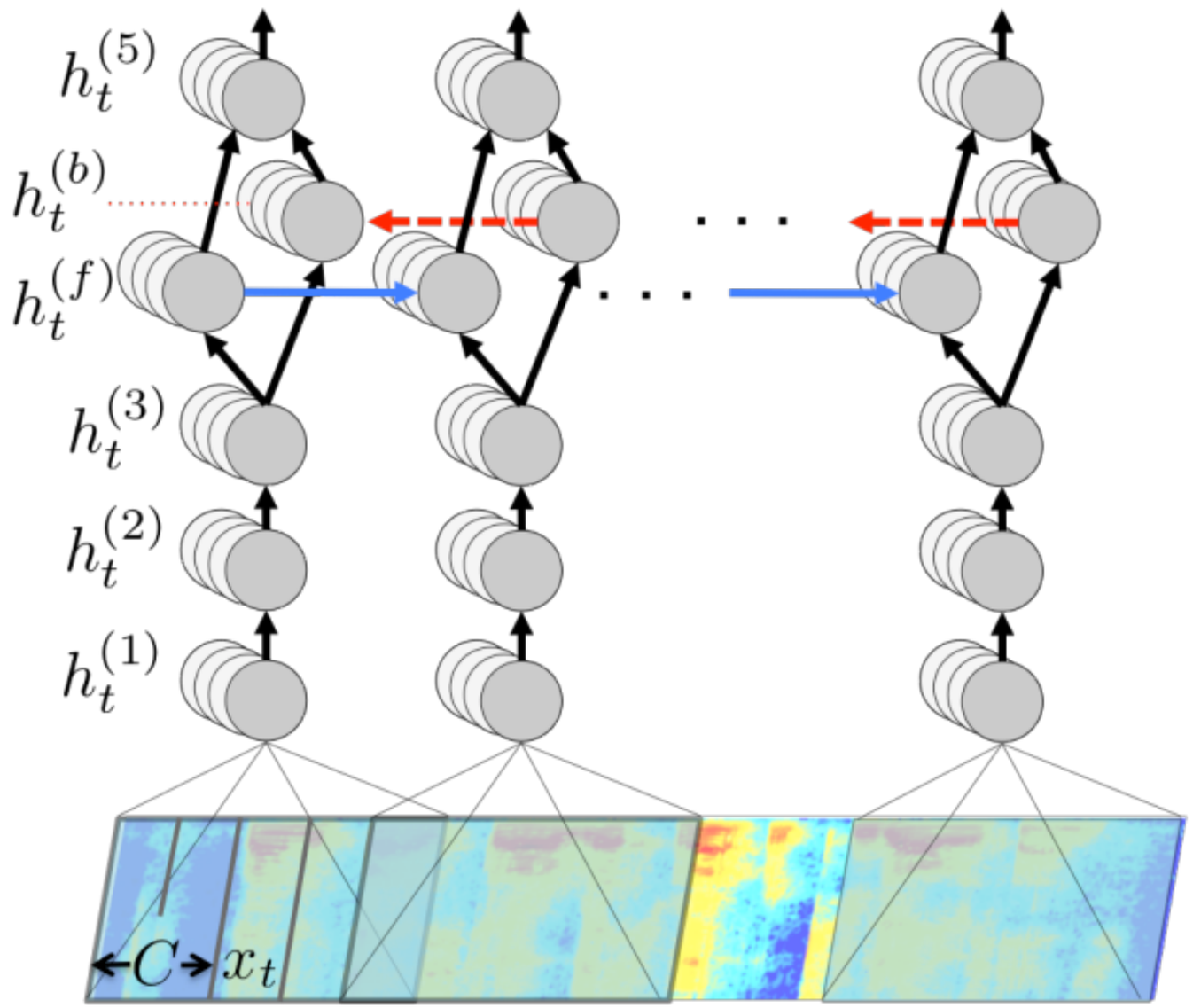
Updates of deep learning!

Go multi-modal!





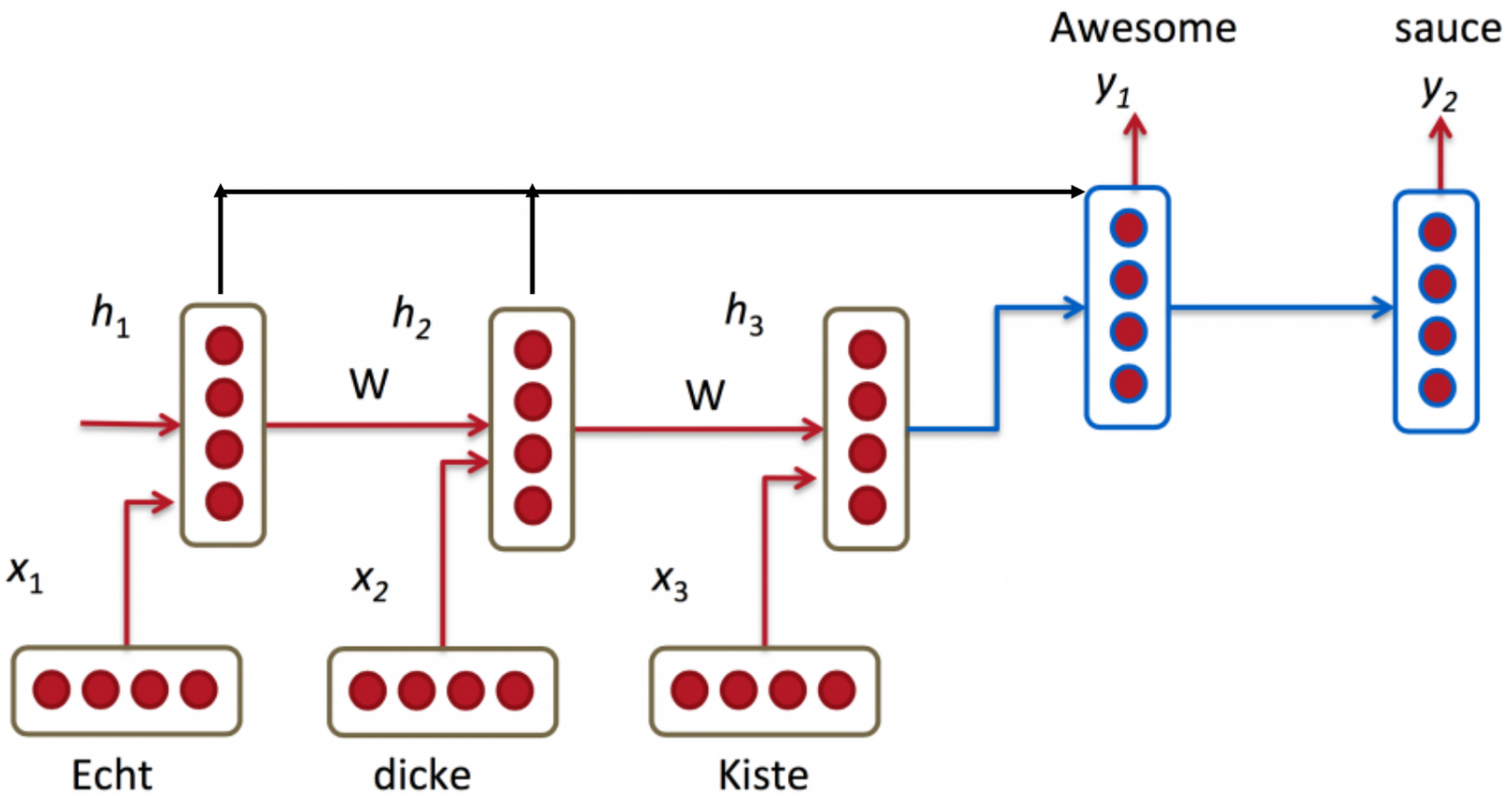
Go End-End!



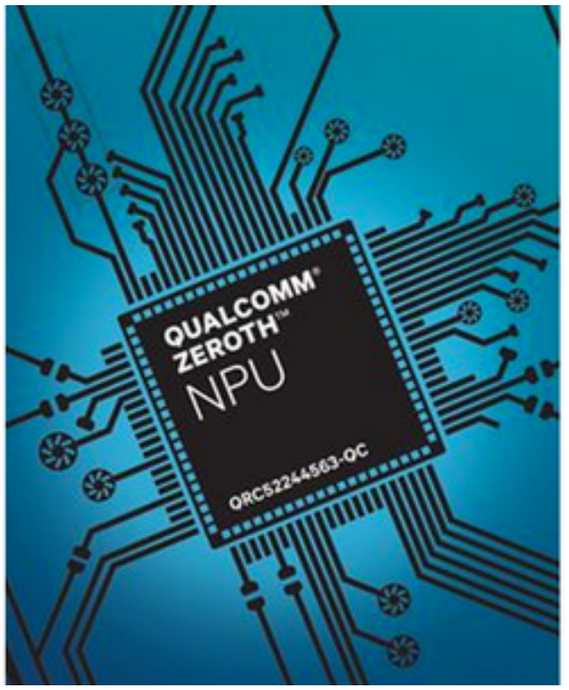
ps://devblogs.nvidia.com/parallelforall/deep-speech-accurate-speech-recognition-gpu-accelerated-deep-learning/#.VO56E8ZtEkM.google_pluson

Updates of deep learning!

Go with attention!



Go with Specalized Chips!



- ? Prototypes!

Go where!

- ? Video understanding!
- ? Deep reinforcement learning!
- ? Natural language understanding!
- ? Unsupervised learning !!!!

Action classification!

ACTIVITYNET

Home Explore People Download About

Annotations

Release 1.2(10-2015)

Inhouse Curated ActivityNet

In this release, we include inhouse curated annotations:

- 100 activity classes
- 4,819 training videos (7151 instances)
- 2,383 validation videos (3582 instances)
- 2,480 testing videos (labels witheld)
- Evaluation server

Release 1.1(06-2015)

Raw ActivityNet

Initial release including post-processed crowdsourced annotations:

- 203 activity classes
- 20,565 training videos
- 6,862 validation videos

53!

ALLEN INSTITUTE *for* ARTIFICIAL INTELLIGENCE

Completed • $80,000 • 170 teams

The Allen AI Science Challenge

Wed 7 Oct 2015 – Sat 13 Feb 2016 (35 days ago)

Dashboard

Private Leaderboard - The Allen AI Science Challenge

This competition has completed. This leaderboard reflects the final standings.

See someone using multiple accounts? Let us know.

#	Δrank	Team Name ‡ model uploaded * in the money	Score	Entries	Last Submission UTC (Best – Last Submission)
1	↑1	Cardal ‡ *	0.59308	2	Mon, 08 Feb 2016 06:54:27
2	↑1	poweredByTalkwalker ‡ *	0.58344	4	Fri, 12 Feb 2016 07:28:58 (-0h)
3	↓2	Alejandro Mosquera ‡ *	0.58257	2	Sat, 06 Feb 2016 08:20:27
4	↑5	Capuccino Monkeys	0.56242	8	Fri, 12 Feb 2016 14:55:43
5	↓1	A Pure Logical Approach ‡	0.56154	5	Thu, 11 Feb 2016 16:51:28

54!

Yann LeCun

March 14 at 3:15am ·

Follow

Statement from a Slashdot post about the AlphaGo victory: "We know now that we don't need any big new breakthroughs to get to true AI"

That is completely, utterly, ridiculously wrong.

As I've said in previous statements: most of human and animal learning is unsupervised learning. If intelligence was a cake, unsupervised learning would be the cake, supervised learning would be the icing on the cake, and reinforcement learning would be the cherry on the cake. We know how to make the icing and the cherry, but we don't know how to make the cake.

We need to solve the unsupervised learning problem before we can even think of getting to true AI. And that's just an obstacle we know about. What about all the ones we don't know about?

#deeplearning #AI #AlphaGo

Slashdot

M.SLASHDOT.ORG

55!

Summary!

- “The takeaway is that deep learning excels in tasks where the basic unit, a single **pixel**, a single **frequency**, or a single **word/character** has little meaning in and of itself, but a combination of such units has a useful meaning.” — Dallin Akagi!

- learning in a End-End fashion is nice trend!

- Structured data + well defined cost function/ rewards !

Thoughts!

- Deep leaning will change our lives in a positive way!
- Still long way to go towards real AI!
- What might happen in the future!
  - **We learn fast by either competing with friend like AlphaGo, or they directly teach us in a effective way**