Deep Learning 101

Cognitive Robotics

Marco Ciccone

Dipartimento di Informatica Elettronica e Bioingegneria
Politecnico di Milano

About me

Present

- **PhD student** in Deep Learning and Computer Vision
- supervised by Prof. Matteo Matteucci
- ML engineer @ Horus

Background

Machine Learning, Signal Processing

- MSc at Politecnico di Milano, Como Campus
- BSc at Università degli studi di Firenze

Contacts

marco.ciccone@polimi.it

Tentative Outline of the course

- Deep Learning introduction
- Optimization background
- Convolutional Neural Networks
- Recurrent Neural Networks
- Training Tricks
- Applications and very (very) brief TensorFlow tutorial [very optimistic]

Let’s start!

Deep Learning breakthrough

“With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.”

https://www.technologyreview.com/s/513696/deep-learning/

“One of the 10 breakthrough technologies 2013”

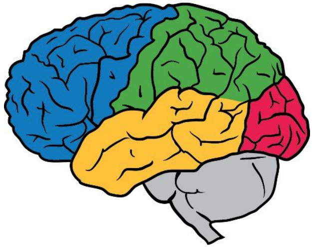
MIT Technology Review

Companies

nnaisense

DeepMind

facebook



NVIDIA

Baidu 百度

amazon echo

Microsoft

UBER AI Labs

intel + nervana

Big players in Academy

UNIVERSITY OF TORONTO

MILA

USI/SUPSI

Istituto Dalle Molle di studi sull'intelligenza artificiale

IDSIA

UNIVERSITY OF OXFORD

Carnegie Mellon University

Berkeley UNIVERSITY OF CALIFORNIA

NYU

UNIVERSITY OF CAMBRIDGE

Main conferences

ICLR

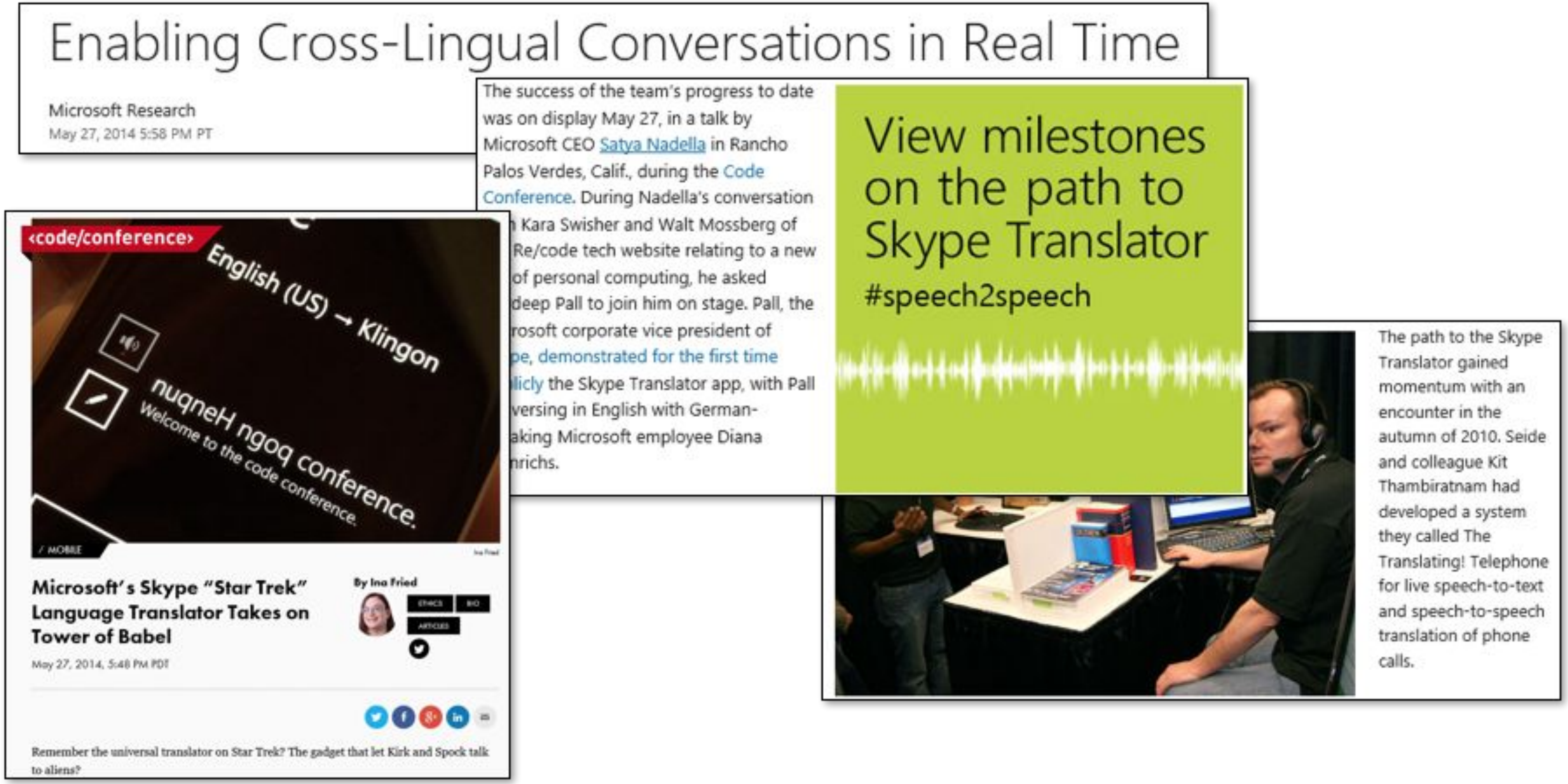
ICML

Deep Learning is growing faster also in other conferences…

CVPR **ICCV** **ICASSP**

ICRA **IROS**

Natural Language Understanding



Enabling Cross-Lingual Conversations in Real Time

Microsoft Research
May 27, 2014 5:58 PM PT

The success of the team's progress to date was on display May 27, in a talk by Microsoft CEO Satya Nadella in Rancho Palos Verdes, Calif., during the Code Conference. During Nadella's conversation h Kara Swisher and Walt Mossberg of Re/code tech website relating to a new of personal computing, he asked deep Pall to join him on stage. Pall, the rosoft corporate vice president of pe, demonstrated for the first time licly the Skype Translator app, with Pall versing in English with German- aking Microsoft employee Diana nrichs.

View milestones on the path to Skype Translator
#speech2speech

The path to the Skype Translator gained momentum with an encounter in the autumn of 2010. Seide and colleague Kit Thambiratnam had developed a system they called The Translating! Telephone for live speech-to-text and speech-to-speech translation of phone calls.

Microsoft's Skype "Star Trek" Language Translator Takes on Tower of Babel

By Ina Fried

May 27, 2014, 5:48 PM PDT

Remember the universal translator on Star Trek? The gadget that let Kirk and Spock talk to aliens?

Speech and Music Generation

WaveNet (DeepMind)

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

Magenta Team (Google Brain)

https://magenta.tensorflow.org/2016/12/16/nips-demo

https://magenta.tensorflow.org/

https://www.youtube.com/watch?v=vM5NaGoynjE

Neural Style Transfer

Code to have fun

https://github.com/jcjohnson/neural-style

https://github.com/jcjohnson/fast-neural-style

https://ml4a.github.io/ml4a/style_transfer/

Deep Photo Style Transfer

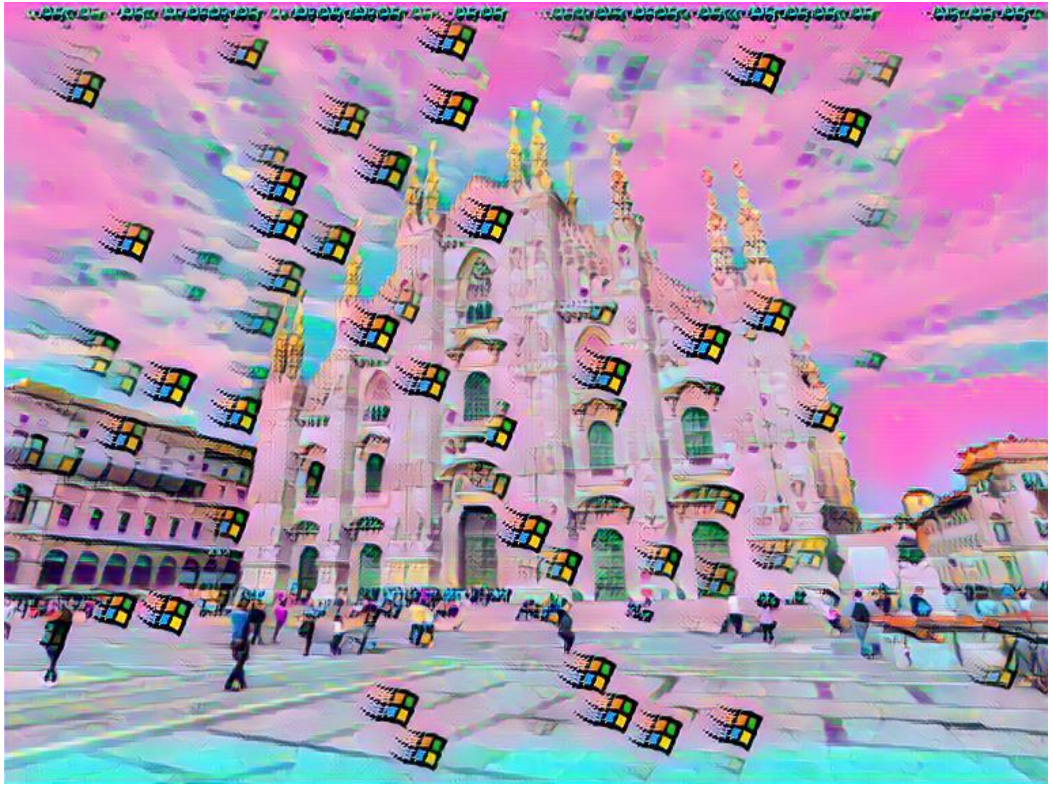
Input

Style

Output

https://github.com/luanfujun/deep-photo-styletransfer

Neural Vaporwave Style Transfer



Neural Vaporwave

provided by Spooky

Mask R-CNN on Image Instance Segmentation

Mask R-CNN, He et Al

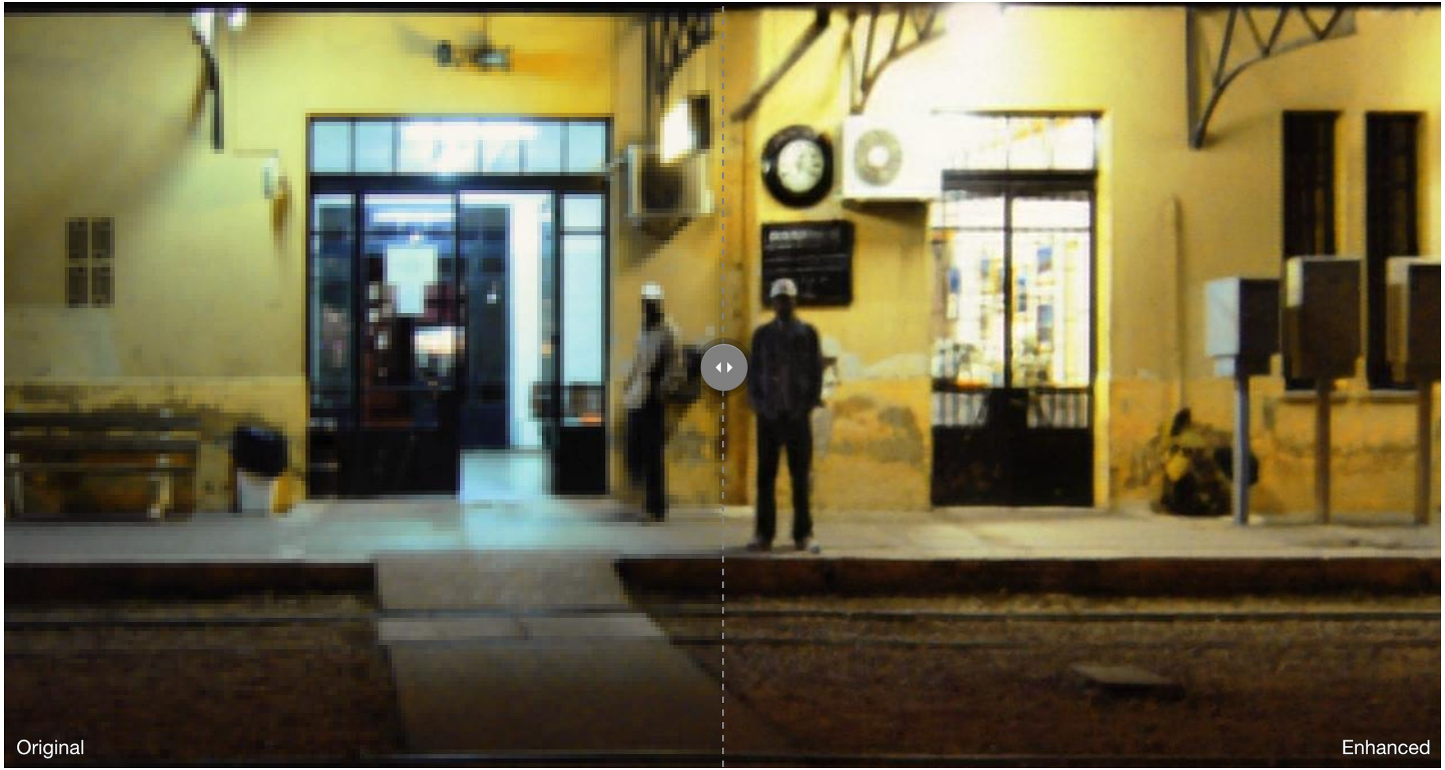
Facebook Research

Mask R-CNN on Detection and Segmentation

Mask R-CNN on Keypoint detection

Super Resolution

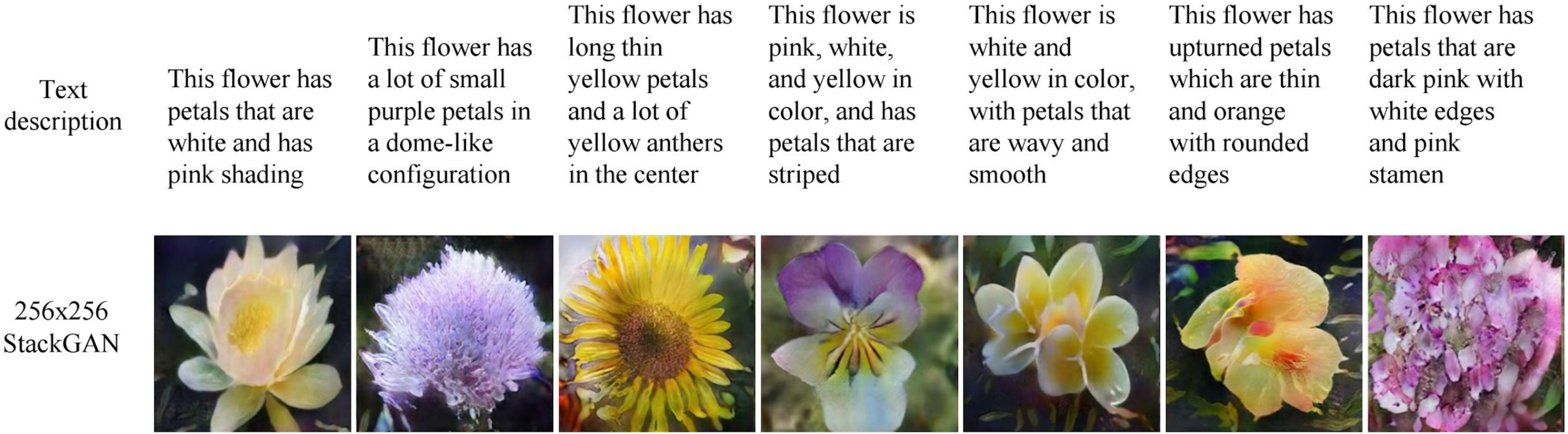
[Live Demo]



https://github.com/alexjc/neural-enhance

Generative Adversarial Network (GAN)

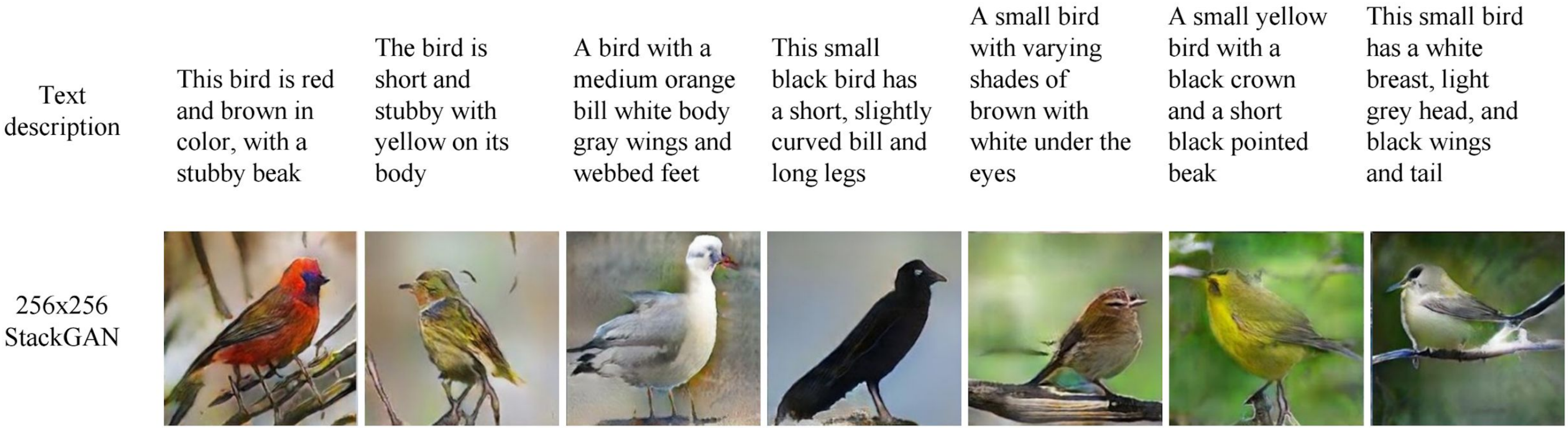
Text-to-Photo Realistic Image Synthesis



https://github.com/hanzhanggit/StackGAN

Generative Adversarial Network (GAN)

Text-to-Photo Realistic Image Synthesis

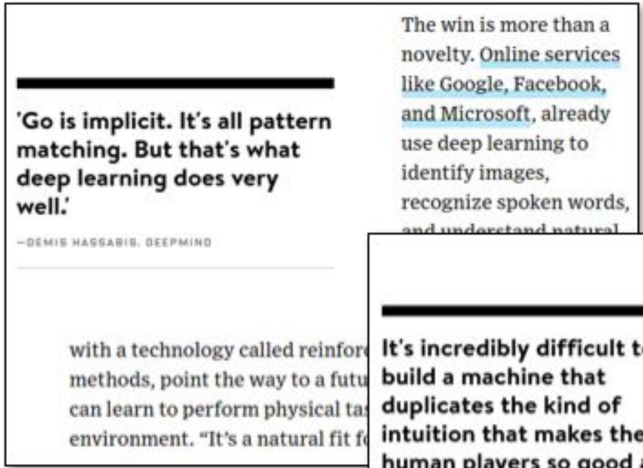


Recommender Systems

https://research.google.com/pubs/pub45530.html

http://benanne.github.io/2014/08/05/spotify-cnns.html

Deep Reinforcement Learning



'Go is implicit. It's all pattern matching. But that's what deep learning does very well.'

—DEMIS HASSABIS, DEEPMIND

The win is more than a novelty. Online services like Google, Facebook, and Microsoft, already use deep learning to identify images, recognize spoken words, and understand natural

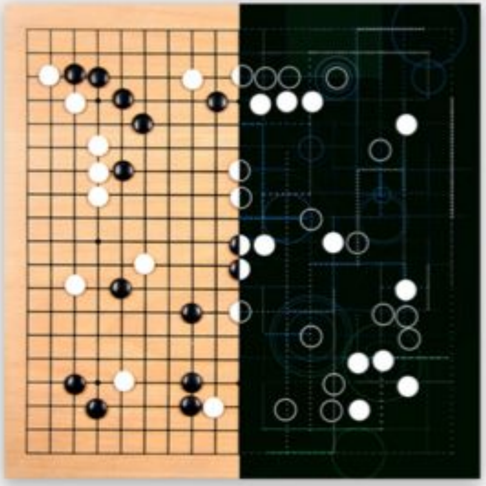
with a technology called reinfor
methods, point the way to a futu
can learn to perform physical ta
environment. "It's a natural fit f

It's incredibly difficult to build a machine that duplicates the kind of intuition that makes the top human players so good at

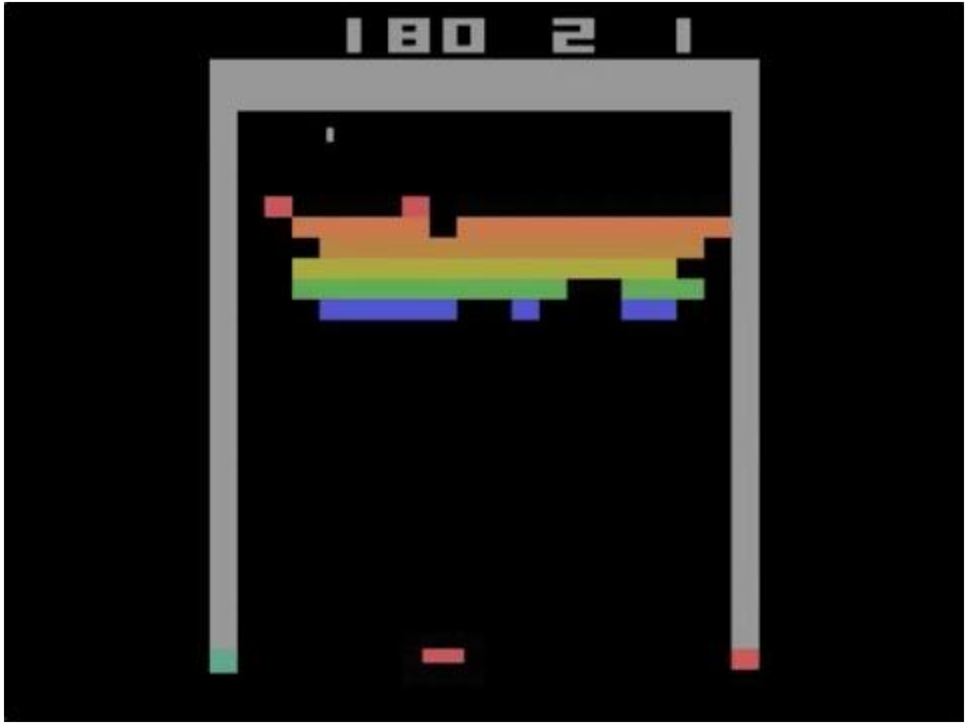
In the mid-'90s, a computer program called Chinook beat the world's top player at the game of checkers. A few years later, IBM's Deep Blue supercomputer shocked the chess world when it wiped the proverbial floor with world champion Gary Kasparov. And more
BM machine, Watson, topped the best
dy!, the venerable TV trivia game.
o mastered Othello, Scrabble,
poker. But in the wake of Crazy Stone's
Coulom predicted that another ten years
a machine could beat a grandmaster
rt.

IN A HUGE BREAKTHROUGH, GOOGLE'S AI BEATS A TOP PLAYER AT THE GAME OF GO

WIRED



Playing Atari with Deep Reinforcement Learning



AlphaGo

Nice article:

https://medium.com/@karpathy/alphago-in-context-c47718cb95a5

Autonomous Driving

Check other NVIDIA automotive partners

NVIDIA self-driving car demo

HORUS

Wearable device for blind people

Text Reading

Face Recognition

Object Recognition

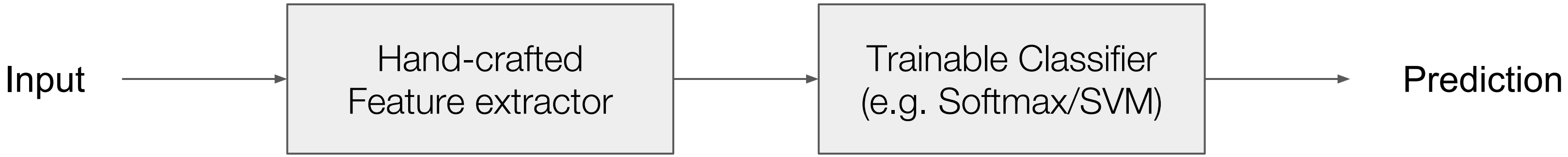
HORUS

eyra

Ok, many funny applications but how does this black magic work?

It’s all about extracting good features!

In the beginning were the hand engineered features



Descriptors are based on fixed heuristics

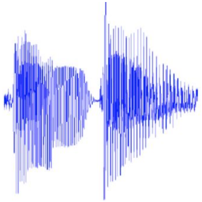
- SIFT, HoG, BRIEF (...) for visual tasks
- MCCF for speech recognition

Bad:

- They need to be carefully designed depending on the task
- Time consuming
- Fixed and cannot deal with data variability

Traditional Pattern Recognition pipelines

Speech Recognition (early 90s - 2011)



MFCC
Low Level Features

Mixture of Gaussian
Mid Level Features

Classifier

Fixed

Unsupervised

Supervised

Object Recognition (2006 - 2012)

SIFT/HoG
Low Level Features

K-means
Sparse Coding
Mid Level Features

Classifier

Fixed

Unsupervised

Supervised

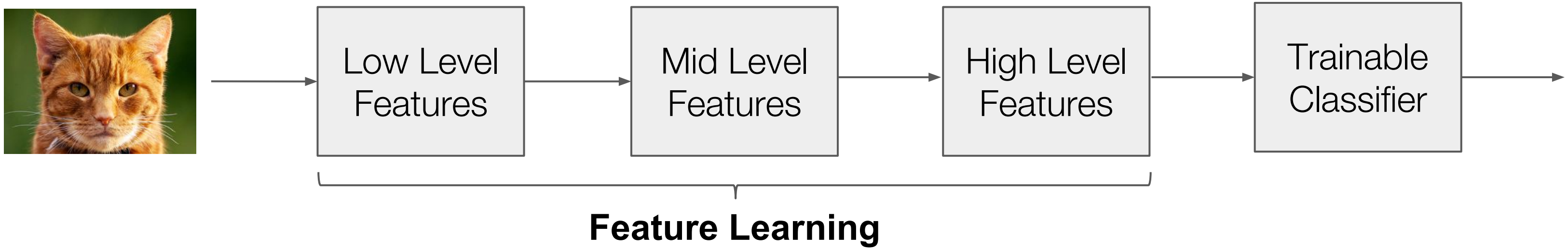
Hand-engineered features still work very well for task that involves feature matching!

But Deep Learning is coming…

LIFT: DL-based descriptor

https://arxiv.org/abs/1603.09114

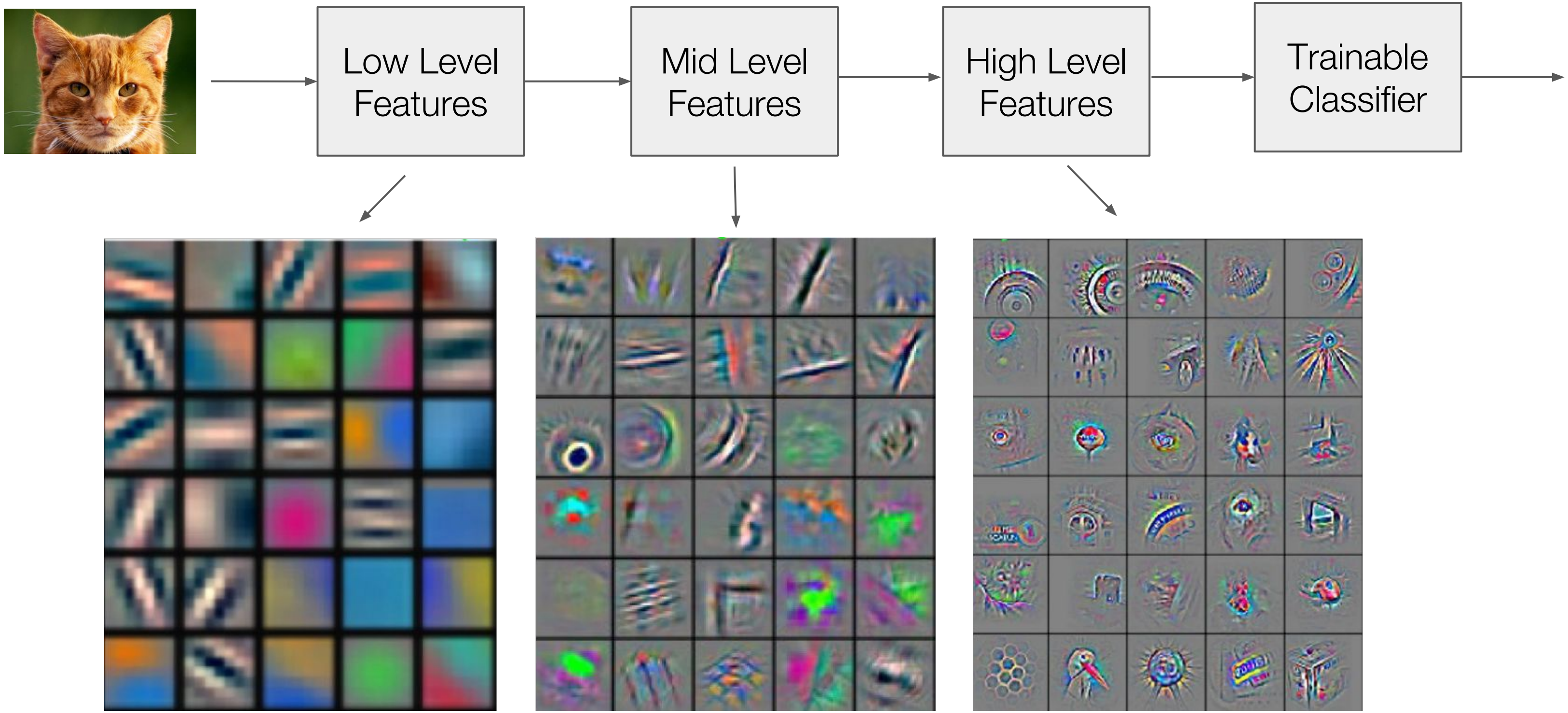
Deep Learning pipeline



Representation Learning address the problem of learning a general and hierarchical feature representation that can be exploited for different tasks.

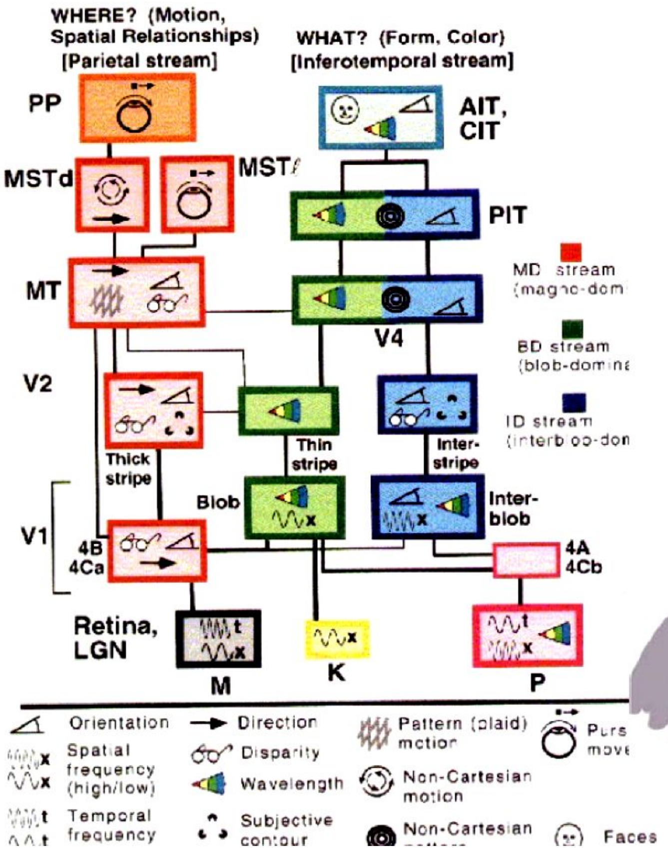
Deep Learning puts together **Representation Learning + Trainable Classifier** in a single *end-to-end training* procedure stacking multiple layers of nonlinear transformation.

Deep Learning pipeline

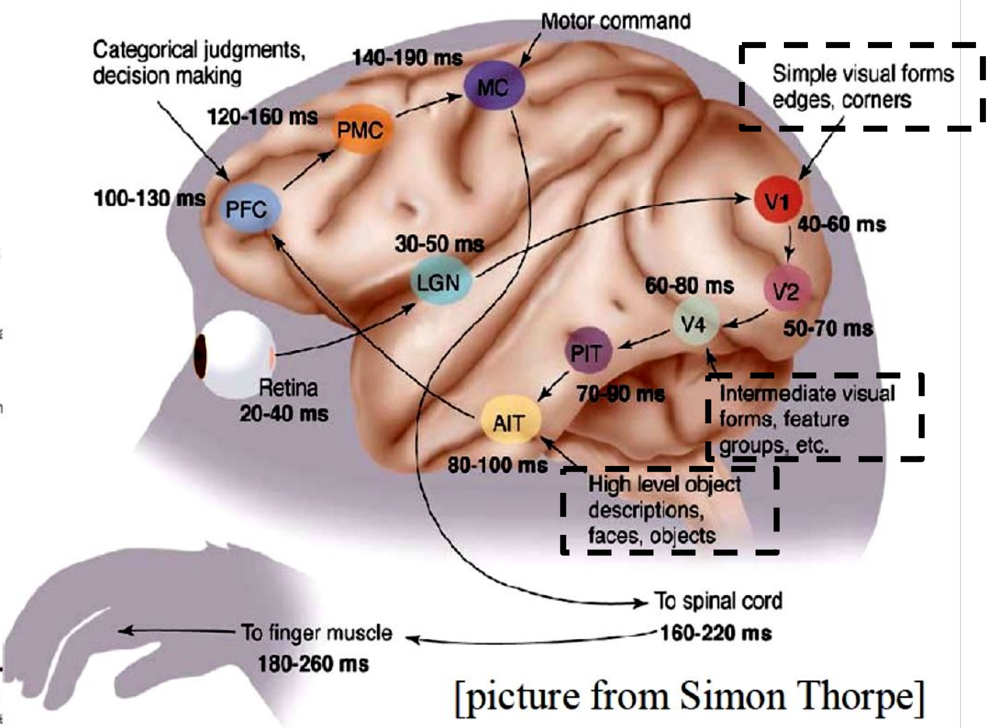


[Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]]

Mammalian visual cortex is hierarchical



[Gallant & Van Essen]



[picture from Simon Thorpe]

Hierarchical Feature Learning

In Deep learning layers are *nonlinear trainable* *feature transforms* that learn a hierarchy of descriptors with increasing abstraction, i.e.,

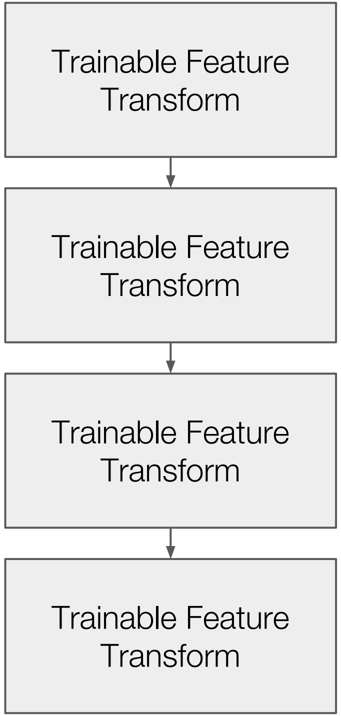


Image recognition

Pixel → edge → texton → motif → part → object

Text analysis

Character → word → word group → clause → sentence → story

Speech recognition

Sample → spectral band → sound → phone → phoneme → word

Representation Learning

Learning the representation is a challenging problem that can have several interpretations

Cognitive perspective

- How can a perceptual system build itself by looking at the external world?
- How much prior structure is necessary?

Neuroscience

- Does the cortex «run» a single, general learning algorithm? Or multiple simpler ones?

ML/AI Perspective

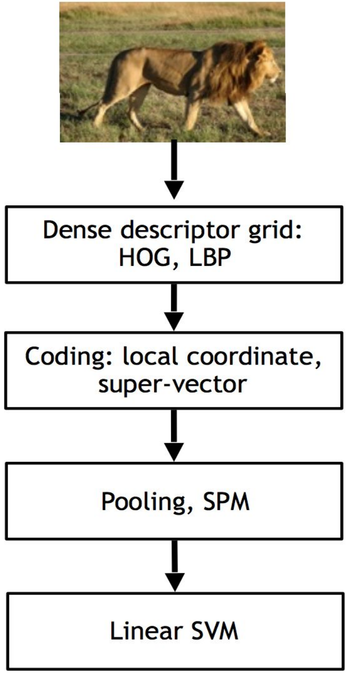
- What is the fundamental principle?
- What is the learning algorithm?
- What is the architecture?

DL addresses the problem of learning hierarchical representations with a single algorithm

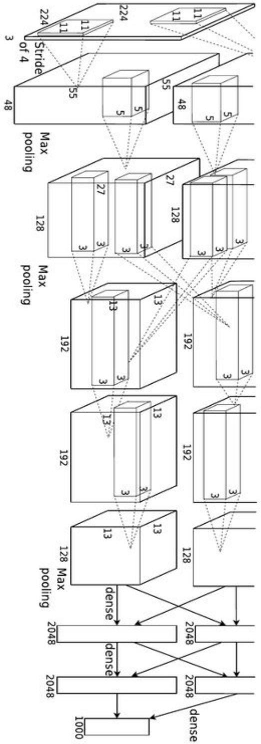
It all started from Image Classification

Large Scale Visual Recognition Challenge (ILSVRC)

[2010] NEC-UIUC

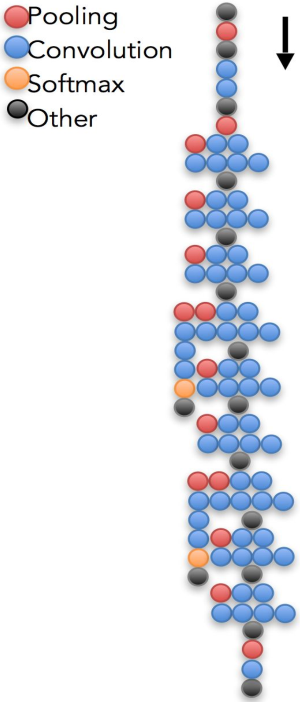


[2012] AlexNet



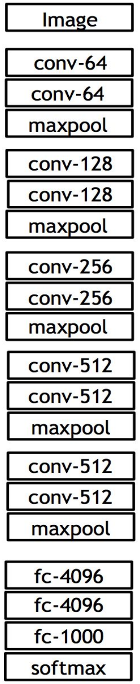
[Krizhevsky et Al.](#)

[2014] GoogleNet



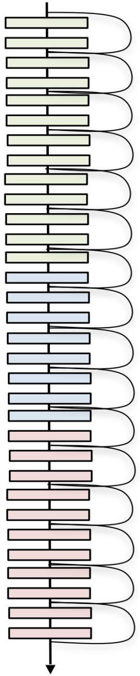
[Szegedy et Al.](#)

VGG



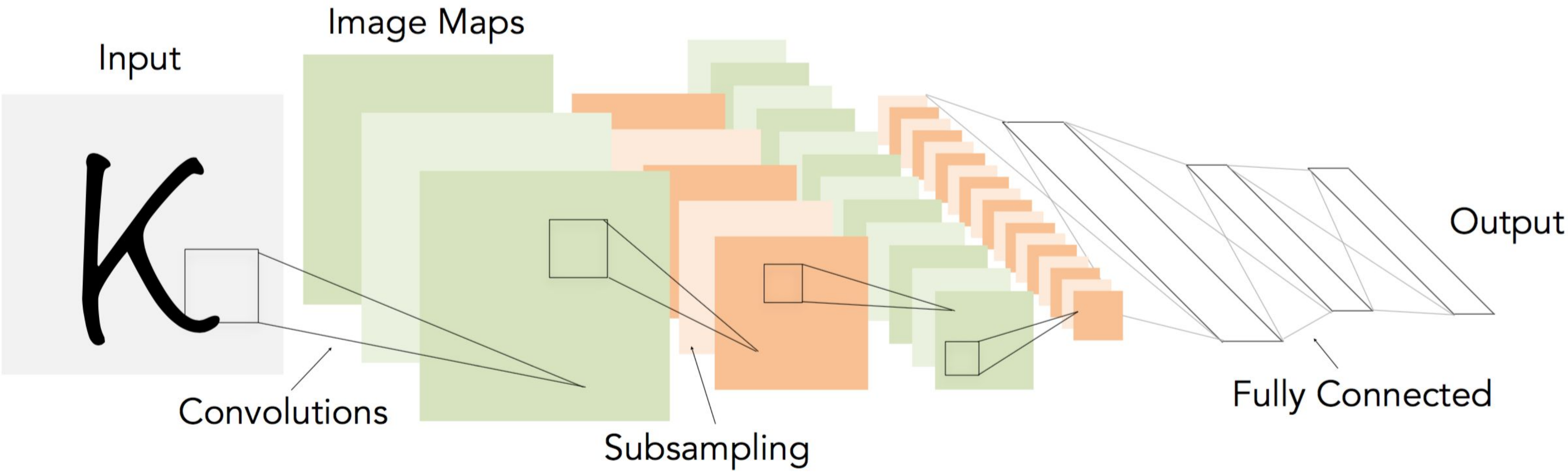
[Symonian & Zisserman](#)

[2015] ResNet



[He et Al.](#)

CNNs are not a new idea!



[1998] LeNet-5

[LeCun et al., 1998]
Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner.
Gradient-based learning applied to document recognition.

But, why Deep Learning didn’t work until now?

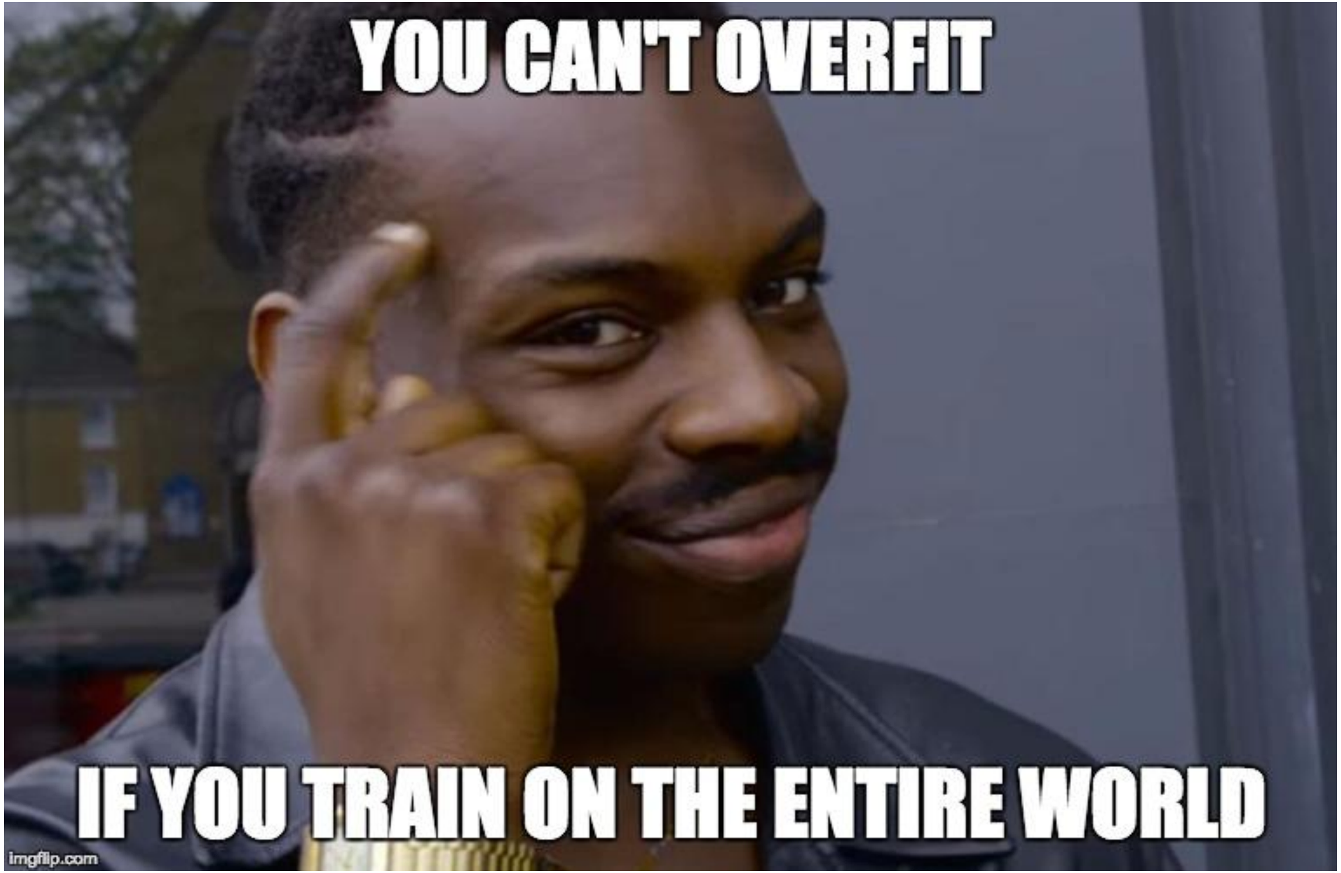
Why DL didn't work until now?

- **Our labeled datasets were thousands of times too small.**
- Our computers were millions of times too slow.
- We initialized the weights in a stupid way.
- We didn't know the Loss surface we were optimizing.
- We used the wrong type of nonlinearity.

IMAGENET

Huge datasets

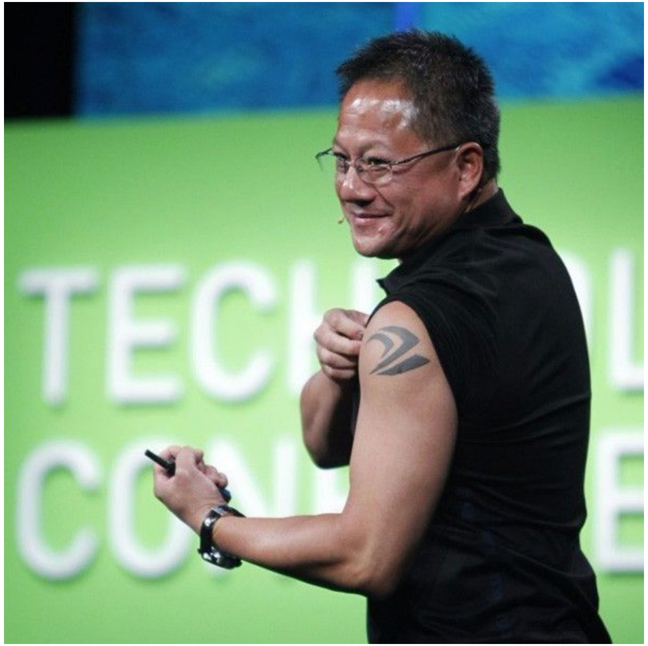
22K categories and 14M images


YOU CAN'T OVERFIT
IF YOU TRAIN ON THE ENTIRE WORLD
imgflip.com

Why DL didn’t work until now?

- Our labeled datasets were thousands of times too small.
- **Our computers were millions of times too slow.**
- We initialized the weights in a stupid way.
- We didn’t know the Loss surface we were optimizing.
- We used the wrong type of nonlinearity.

GPU acceleration and scalable algorithms



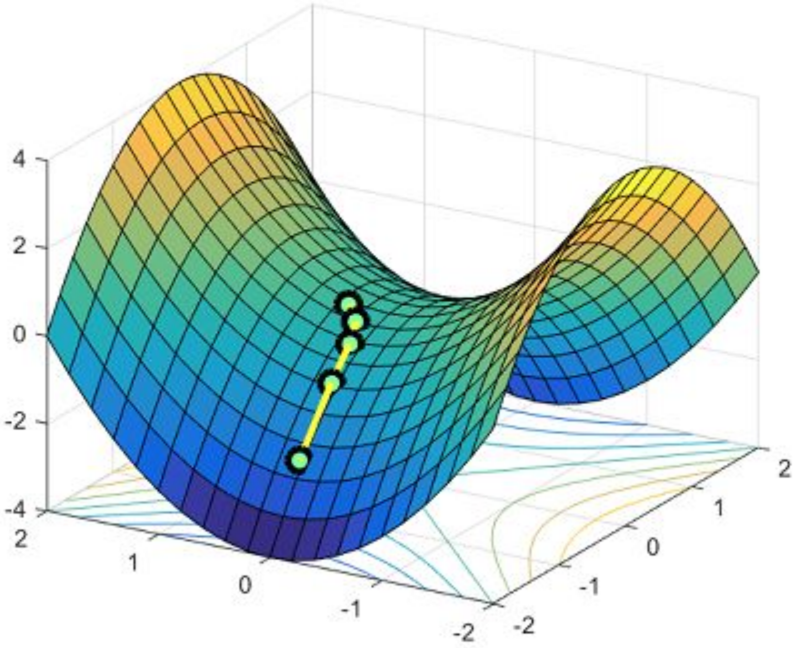
Approved by Jensen (Gianni)

Why DL didn’t work until now?

- Our labeled datasets were thousands of times too small.
- Our computers were millions of times too slow.
- **We initialized the weights in a stupid way.**
- We didn’t know the Loss surface we were optimizing.
- We used the wrong type of nonlinearity.

Why DL didn't work until now?

- Our labeled datasets were thousands of times too small.
- Our computers were millions of times too slow.
- We initialized the weights in a stupid way.
- **We didn't know the Loss surface we were optimizing.**
- We used the wrong type of nonlinearity.



References

- On the saddle point problem for non-convex optimization, Pascanu, Dauphin, Ganguli, Bengio, 2014
- Identifying and attacking the saddle point problem in high-dimensional non-convex optimization, YN Dauphin, R Pascanu, C Gulcehre, K Cho, S Ganguli, Y Bengio, NIPS 2014
- **Escaping from Saddle Points** (13-minutes read)

Why DL didn’t work until now?

- Our labeled datasets were thousands of times too small.
- Our computers were millions of times too slow.
- We initialized the weights in a stupid way.
- We didn’t know the Loss surface we were optimizing.
- **We used the wrong type of nonlinearity (we’ll see).**

What should I know before starting?

Pre-requisites

We will be formulating *cost functions*, taking derivatives and performing optimization with gradient descent so...

It's strictly recommended to have a strong background of:

- Calculus
- Linear Algebra, Matrix Calculus
- Probability Theory
- Machine Learning basic concepts (e.g. Supervised Training, Overfitting…)

I will give you a brief overview on *Optimization theory* and techniques.

Resources

- **Slides provided by me,** mainly based on other brilliant material such as:

  Hugo Larochelle, slides and videos

  Andrej Karpathy, Stanford CS231n Course Notes

  Laurent Dinh, Introduction to DL with Theano

- **The Deep Learning Book** by Ian Goodfellow, Yoshua Bengio, Aaron Courville

  Available online for free: http://www.deeplearningbook.org/