Deep Learning

Lecture 1: Introduction and the success of deep learning

Prof. Stéphane Gaïffas
https://stephanegaiffas.github.io

Introduction and logistics

Your favorite professor:

- Pr. Stéphane Gaïffas
- LPSM, Univ. de Paris and DMA, Ecole normale supérieure
- Data Science, Machine Learning and Statistics
- https://stephanegaiffas.github.io

Your favorite website (bookmark it):

- https://stephanegaiffas.github.io/deep_learning

When ?

EDT M2 MIDS

How ?

- 1/2 time spent in lectures
- 1/2 time on hands-on (`tensorflow`, `pytorch`, etc.)
- Assessment through homeworks by pairs of students
- Check regularly the course website
  https://stephanegaiffas.github.io/deep_learning
  to get up-to-date information

Deep learning. Why learning ?

What do you see?

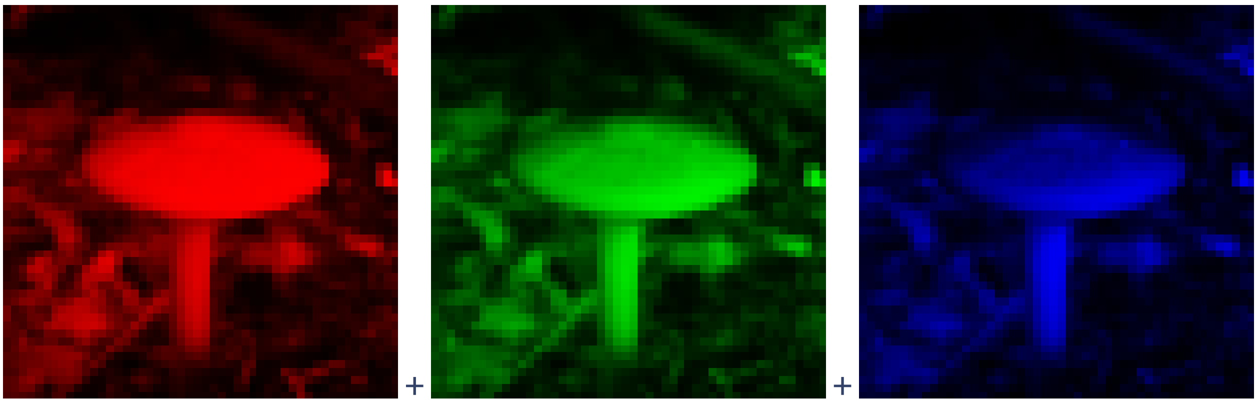
Sheepdog or mop?

———
Credits: Karen Zack, 2016.

Chihuahua or muffin?

This is a mushroom.

This is a mushroom.

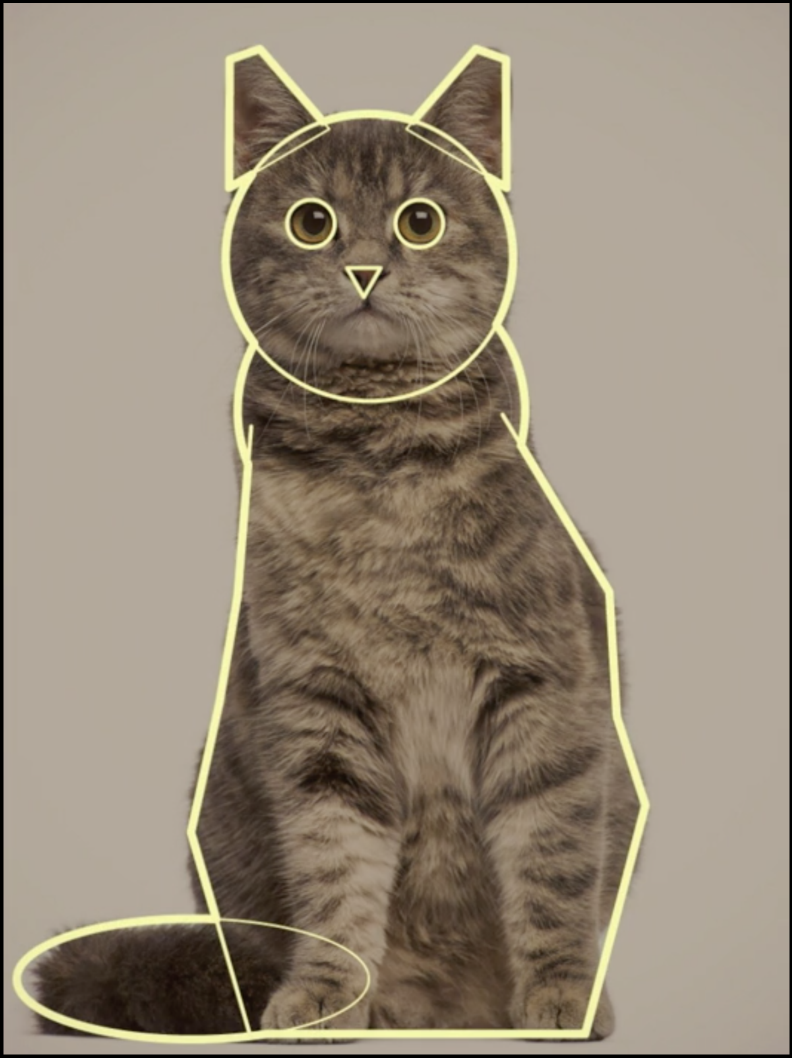


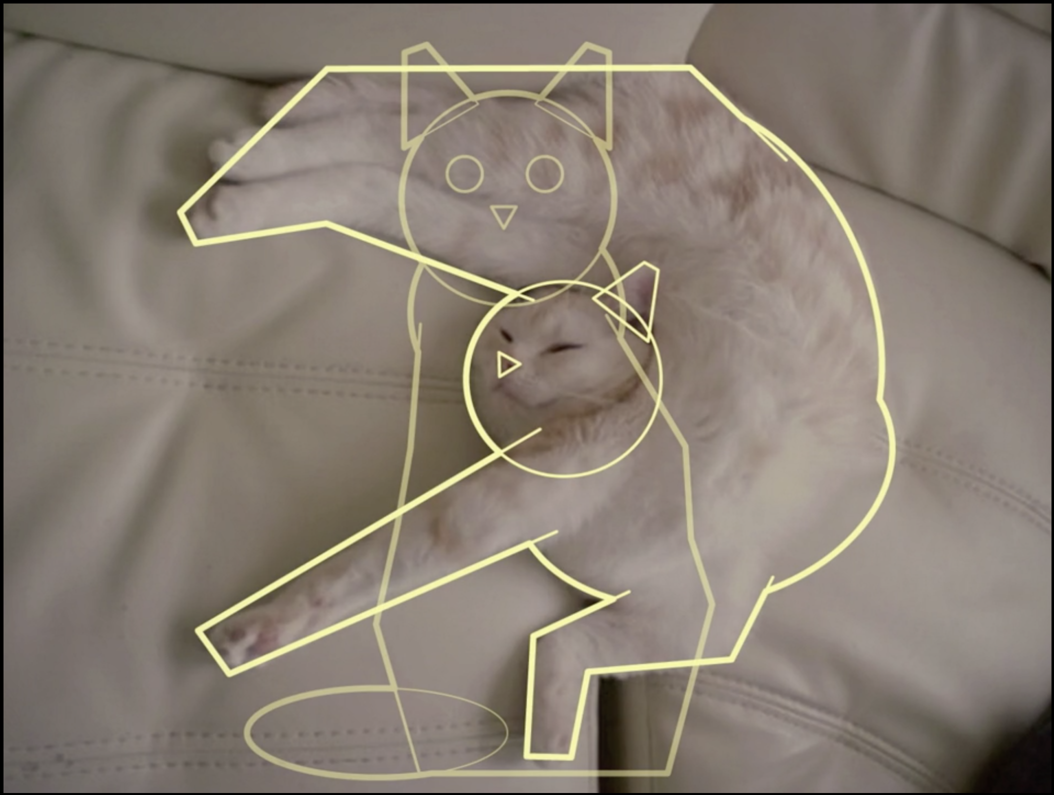
This is a mushroom.

```
array([[[0.03921569, 0.03529412, 0.02352941, 1.        ],
        [0.2509804 , 0.1882353 , 0.20392157, 1.        ],
        [0.4117647 , 0.34117648, 0.37254903, 1.        ],
        ...,
        [0.20392157, 0.23529412, 0.17254902, 1.        ],
        [0.16470589, 0.18039216, 0.12156863, 1.        ],
        [0.18039216, 0.18039216, 0.14117648, 1.        ]],

       [[0.1254902 , 0.11372549, 0.09411765, 1.        ],
        [0.2901961 , 0.2509804 , 0.24705882, 1.        ],
        [0.21176471, 0.2       , 0.20392157, 1.        ],
        ...,
        [0.1764706 , 0.24705882, 0.12156863, 1.        ],
        [0.10980392, 0.15686275, 0.07843138, 1.        ],
        [0.16470589, 0.20784314, 0.11764706, 1.        ]],

       [[0.14117648, 0.12941177, 0.10980392, 1.        ],
        [0.21176471, 0.1882353 , 0.16862746, 1.        ],
        [0.14117648, 0.13725491, 0.12941177, 1.        ],
        ...,
        [0.10980392, 0.15686275, 0.08627451, 1.        ],
        [0.0627451 , 0.08235294, 0.05098039, 1.        ],
        [0.14117648, 0.2       , 0.09803922, 1.        ]],

       ...,
```

This is a mushroom.

Writing a computer program that sees?





- Extracting semantic information requires models of high complexity, which cannot be designed by hand.
- However, one can write a program that learns the task of extracting semantic information.

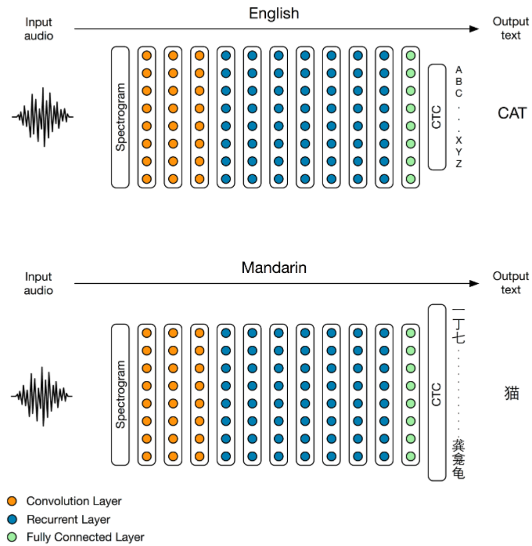
Techniques used in practice consist of:

- defining a parametric model with high capacity
- optimizing its parameters, by "making it work" on the training data

- Very roughly reminiscent of biological systems for which the model (e.g., brain structure) and parameters (e.g., synaptic weights) are tuned through life experiences
- Deep learning encompasses software and hardware technologies allowing to scale-up to billions of model parameters and as many training examples

Applications and successes

Speech-to-text

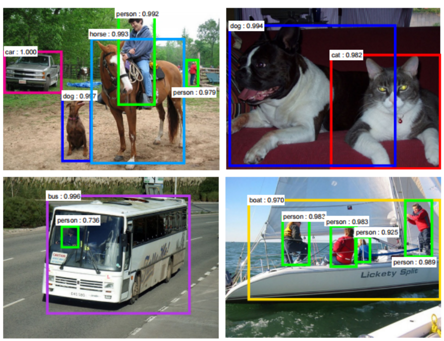


[Baidu 2014]

Computer vision

[Krizhevsky 2012]

[Ciresan et al. 2013]

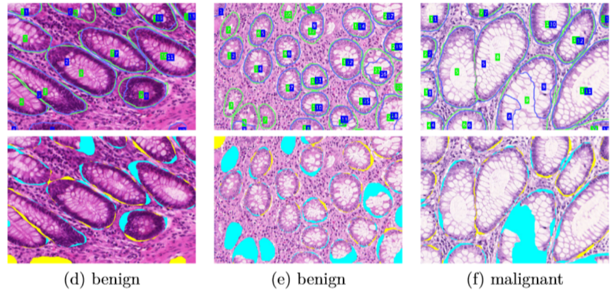


[Faster R-CNN - Ren 2015]

[NVIDIA dev blog]

Computer vision

[Stanford 2017]



[Nvidia Dev Blog 2017]

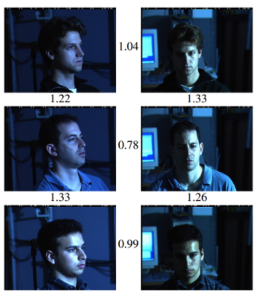
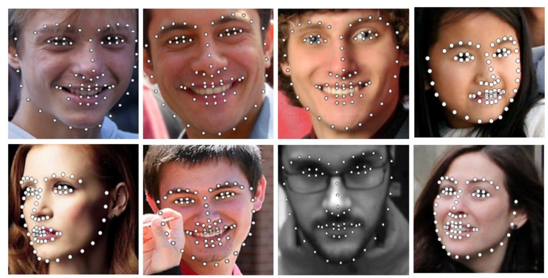


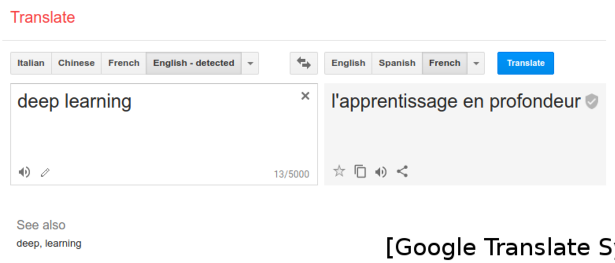
Figure 1. **Illumination and Pose invariance.**

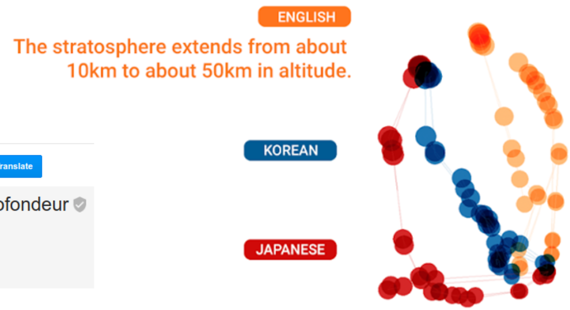
[FaceNet - Google 2015]



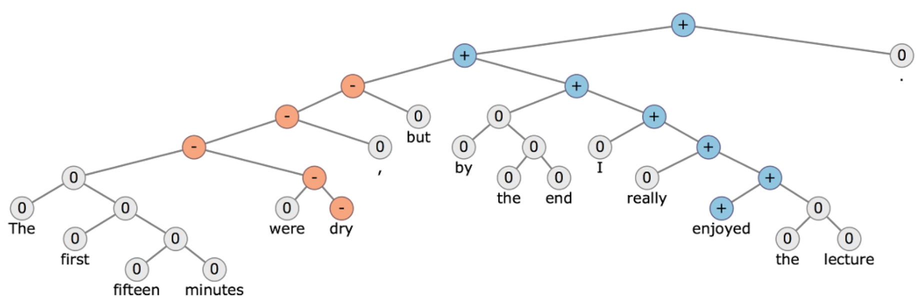
[Facial landmark detection CUHK 2014]

Natural Language Processing (NLP)





[Google Translate System - 2016]



[Socher 2015]

Natural Language Processing (NLP)

[Google Inbox Smart Reply]

[Amazon Echo / Alexa]

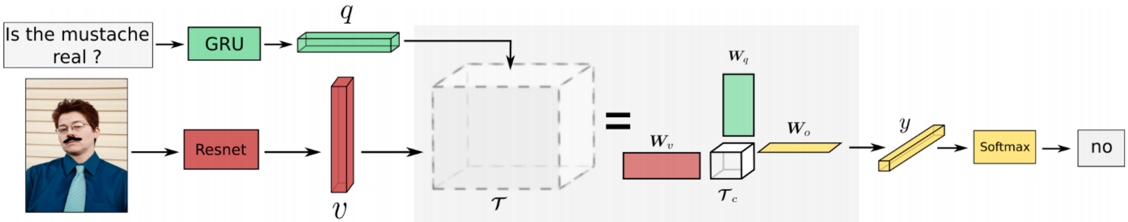
Natural Language Processing (NLP)

[Google Inbox Smart Reply]

[Amazon Echo / Alexa]

However, most chatbots claiming "AI" do not use deep learning (yet?)

Computer vision + NLP



[VQA - Mutan 2017]

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

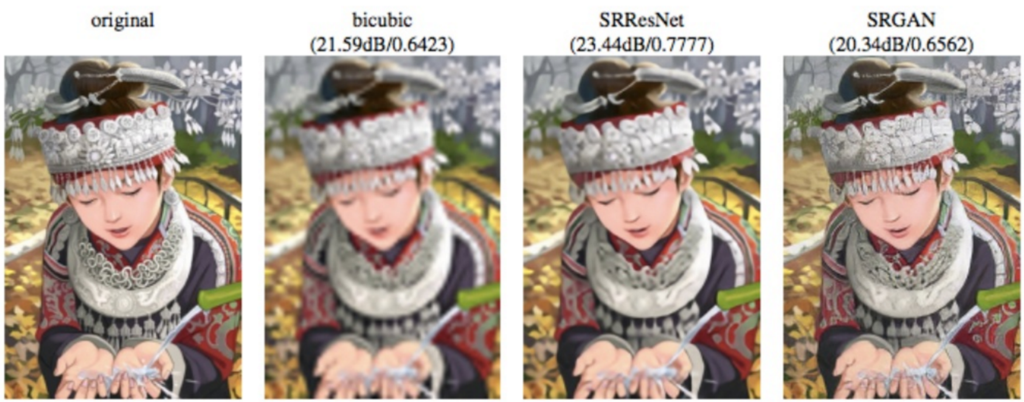
"boy is doing backflip on wakeboard."

[Karpathy 2015]

Image translation

[DeepDream 2015]

[Gatys 2015]



[Ledig 2016]

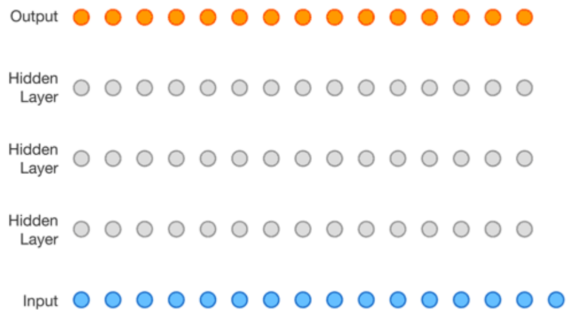
Generative models

Sampled celebrities [Nvidia 2017]

Generative models

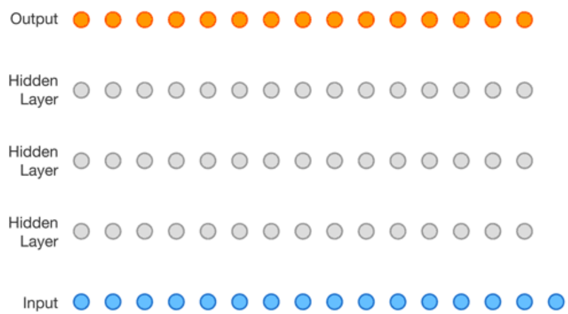
StackGAN v2 [Zhang 2017]

Generative models



Sound generation with WaveNet [DeepMind 2017]

Generative models



Sound generation with WaveNet [DeepMind 2017]

Which one is generated?

Tacotron 2 Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions, 2017

Language / Image models

Open-AI GPT-3 or DALL-E: https://openai.com/blog/dall-e/

TEXT PROMPT

an armchair in the shape of an avocado [...]

AI-GENERATED IMAGES

View more or edit prompt ↓

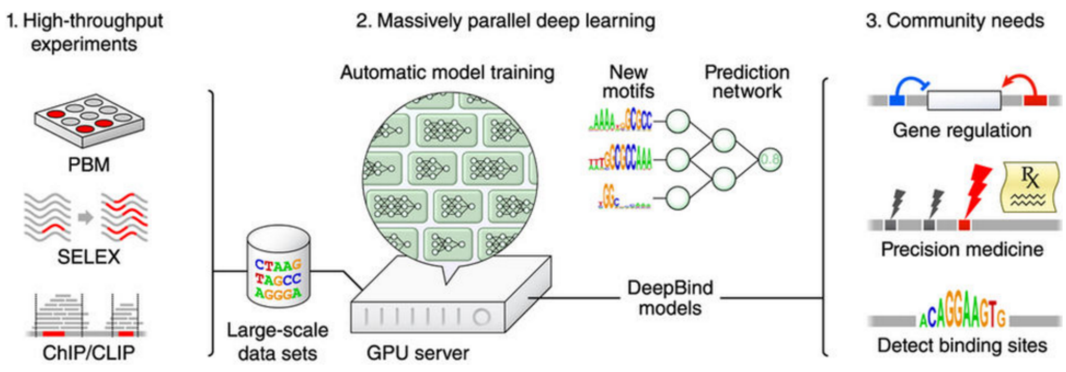
TEXT PROMPT

a store front that has the word 'openai' written on it [...]

AI-GENERATED IMAGES

View more or edit prompt ↓

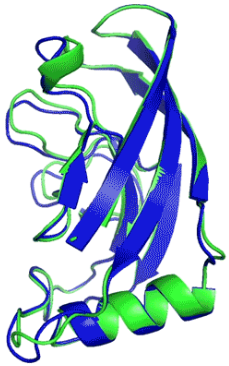
Genomics / biology



[Deep Genomics 2017]

Genomics / biology

T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)

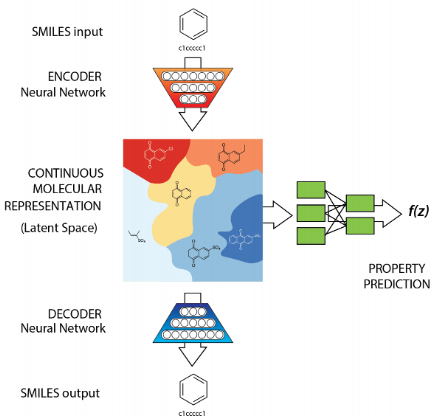


T1049 / 6y4f
93.3 GDT
(adhesin tip)

● Experimental result
● Computational prediction

AlphaFold by DeepMind

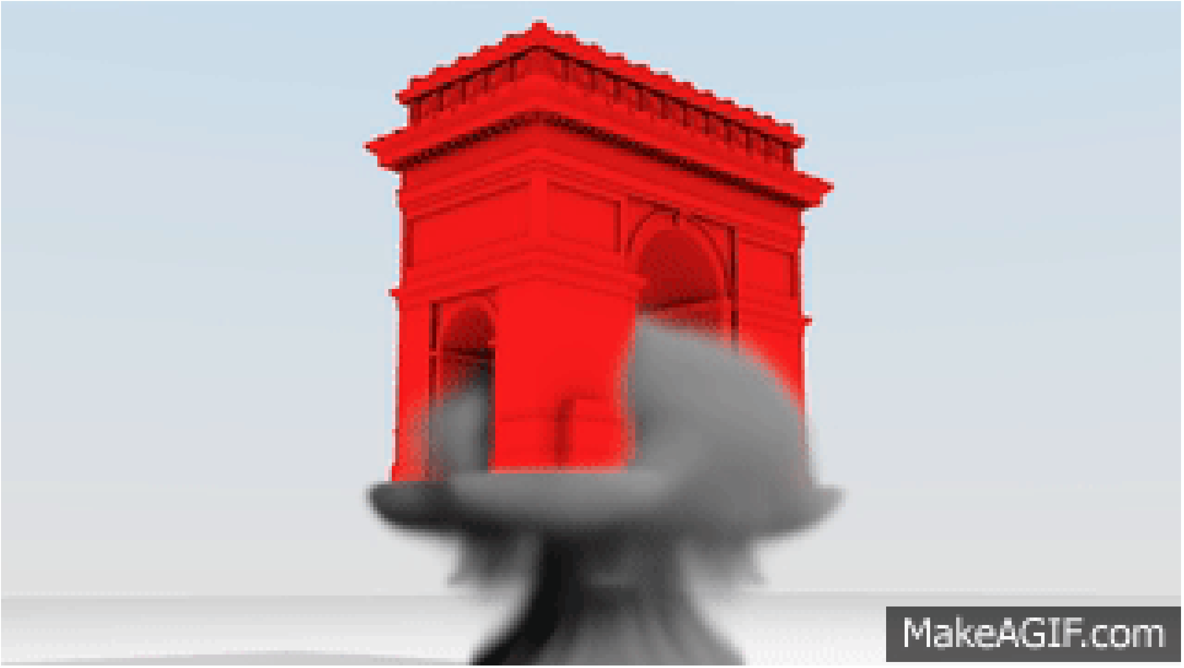
Chemistry, physics



[Gómez-Bombarelli 2016]

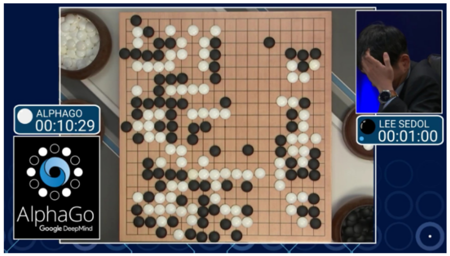
[Tompson 2016]

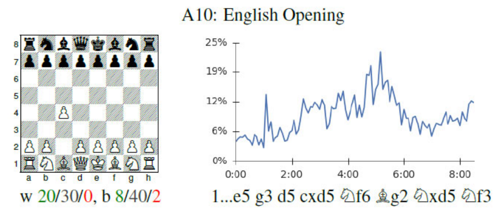
Chemistry, physics



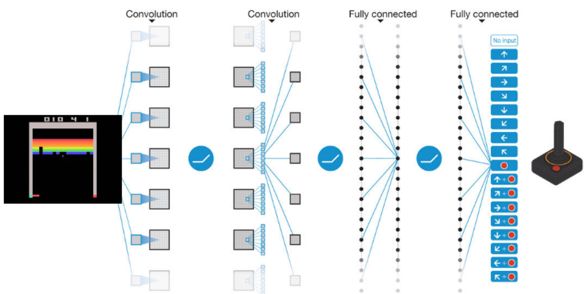
Finite element simulator accelerated ($\approx$ 100 fold) by a 3D convolutional network

AI in games

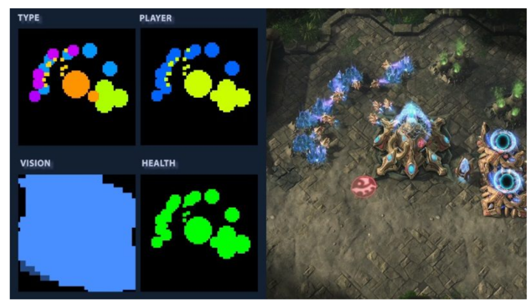




[Deepmind AlphaGo / Zero 2017]



[Atari Games - DeepMind 2016]

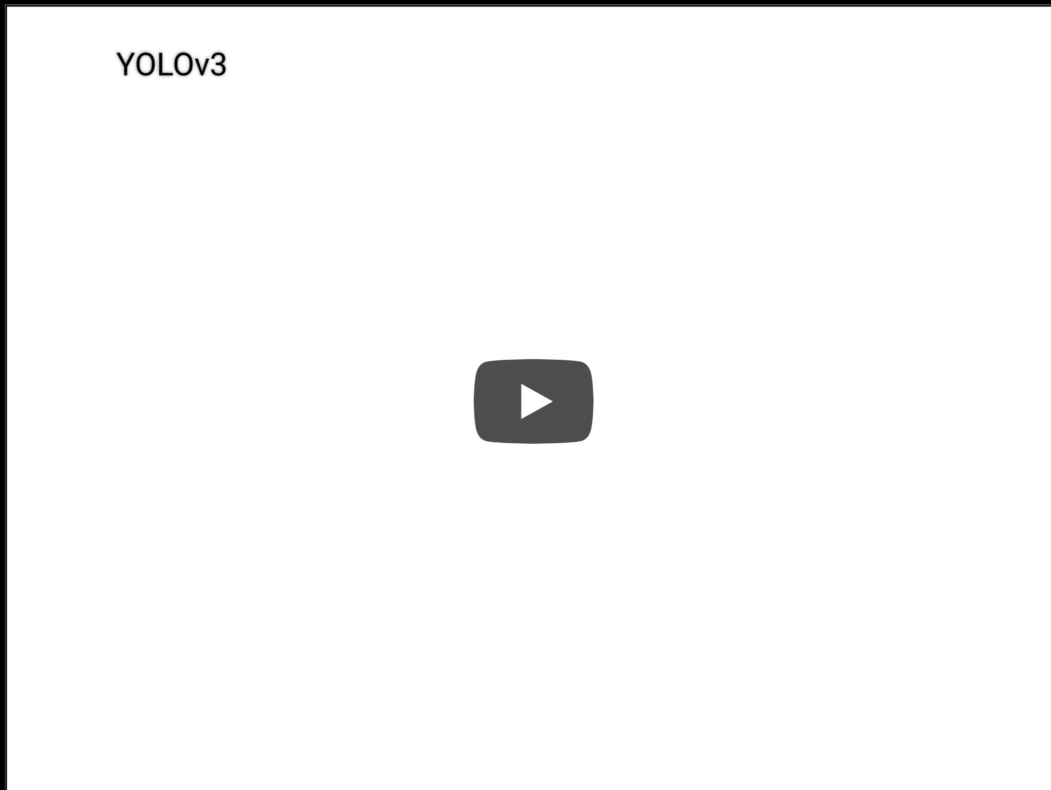


[Starcraft 2 for AI research]

AlphaGo/Zero: Monte Carlo Tree Search, Deep Reinforcement Learning, self-play

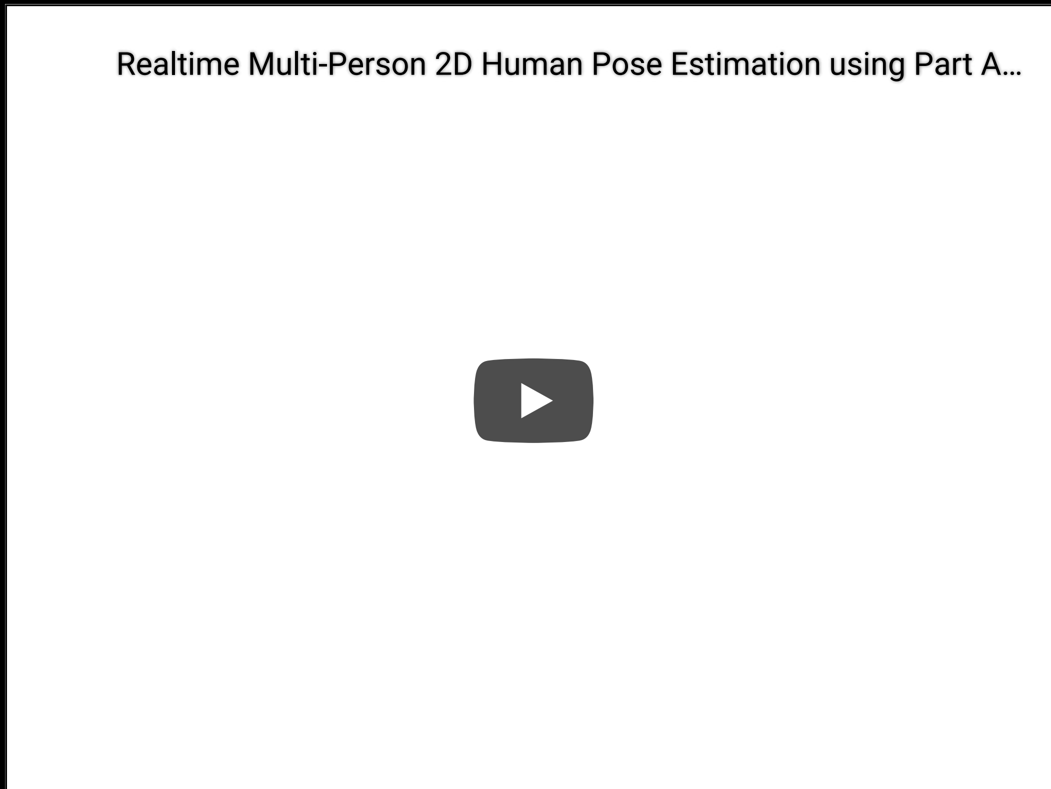
More examples in the next videos

(watch these later...)

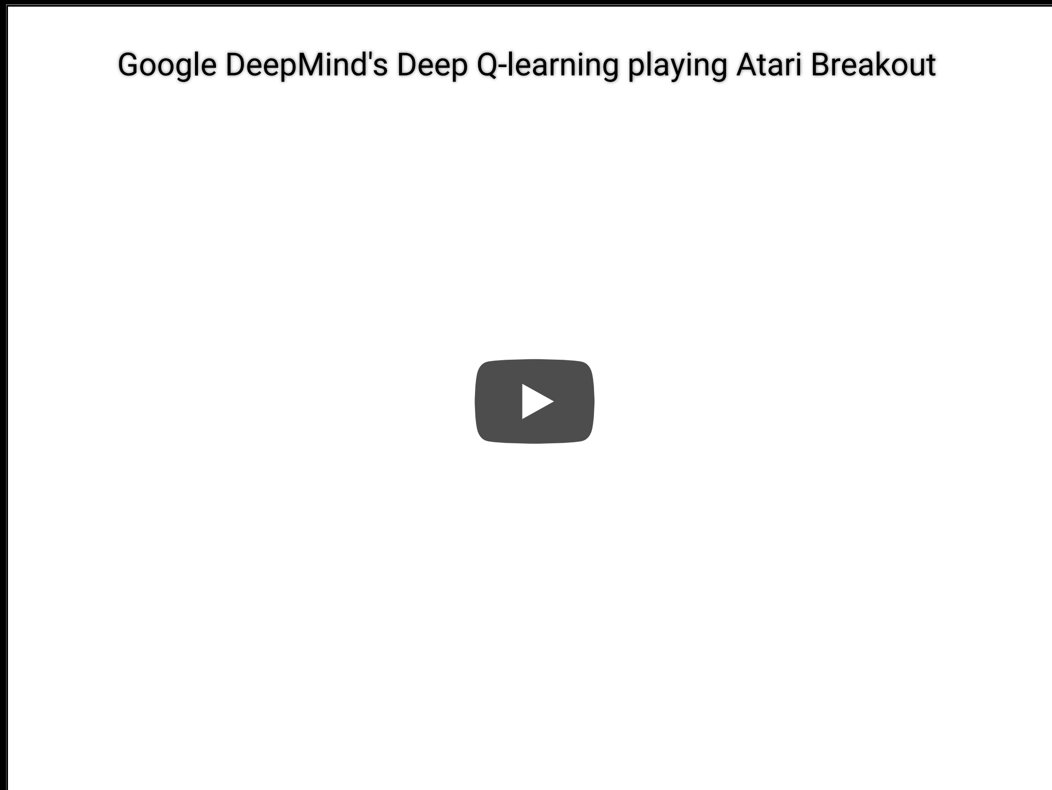


Real-time object detection (Redmon and Farhadi, 2018)

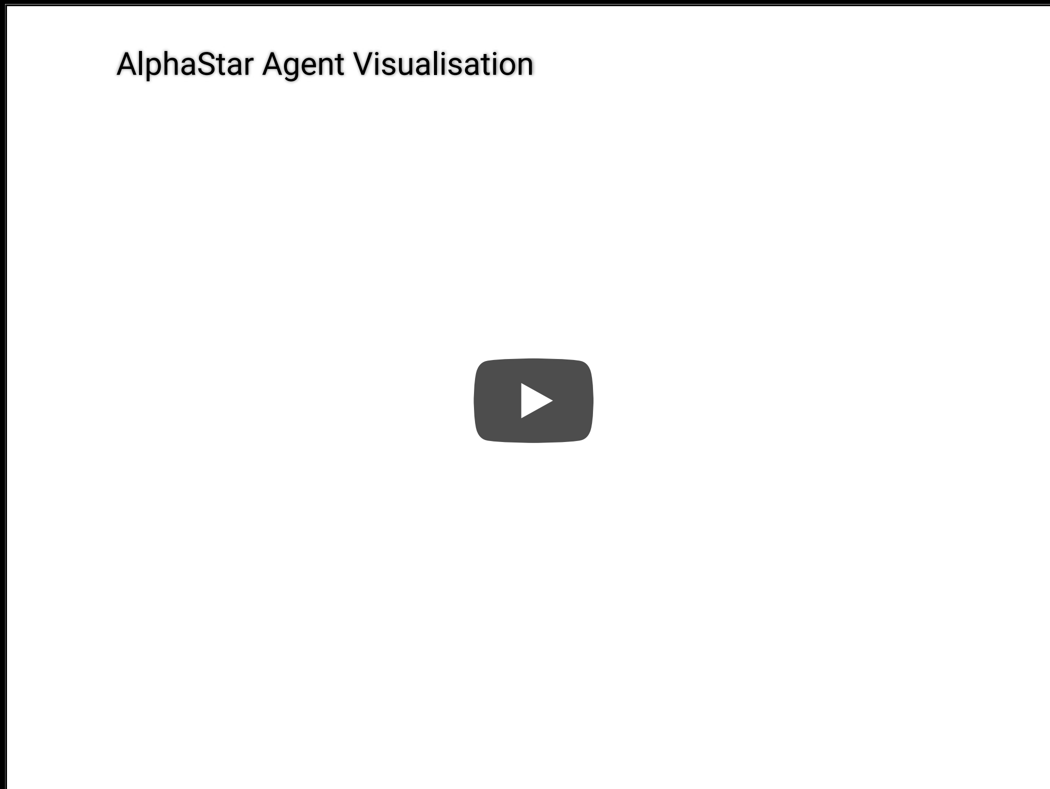
Segmentation (Hengshuang et al, 2017)



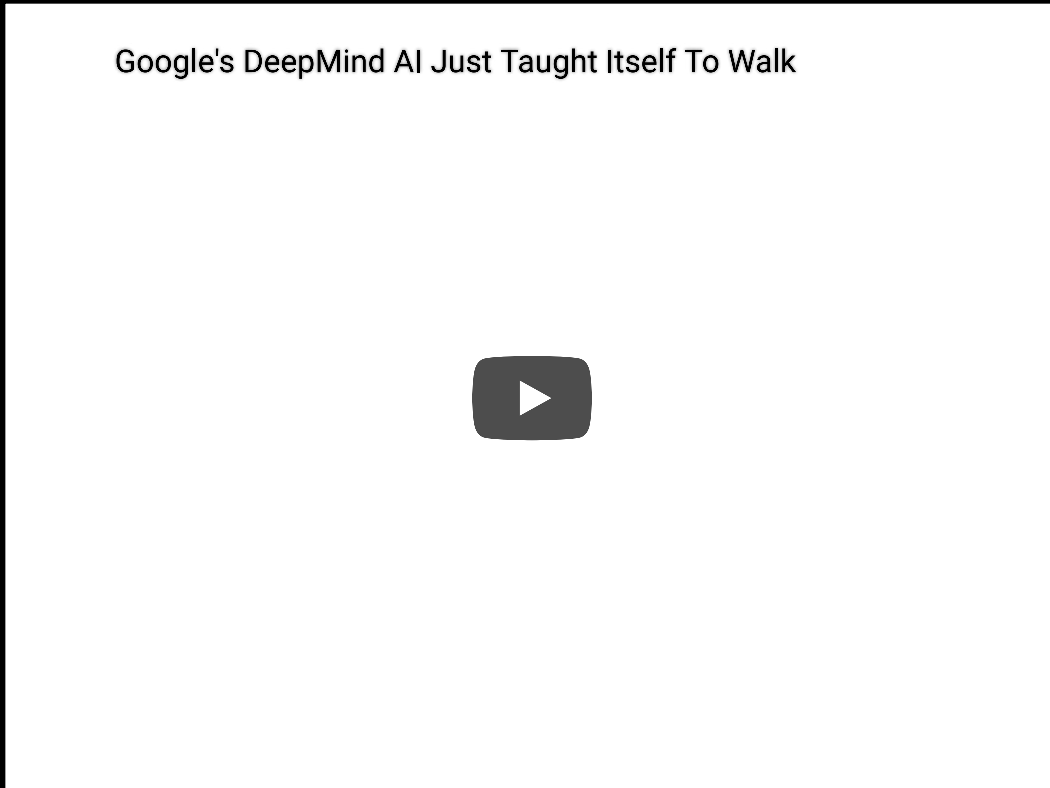
Pose estimation (Cao et al, 2017)



Reinforcement learning (Mnih et al, 2014)

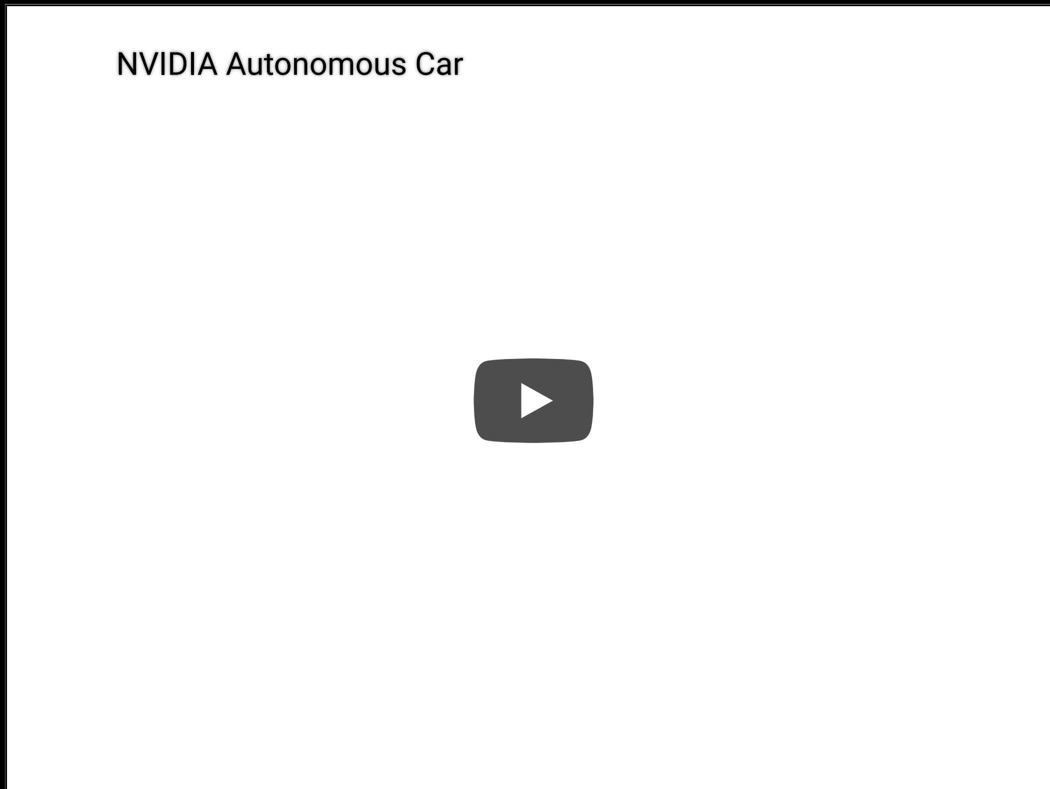


Strategy games (Deepmind, 2016-2018)

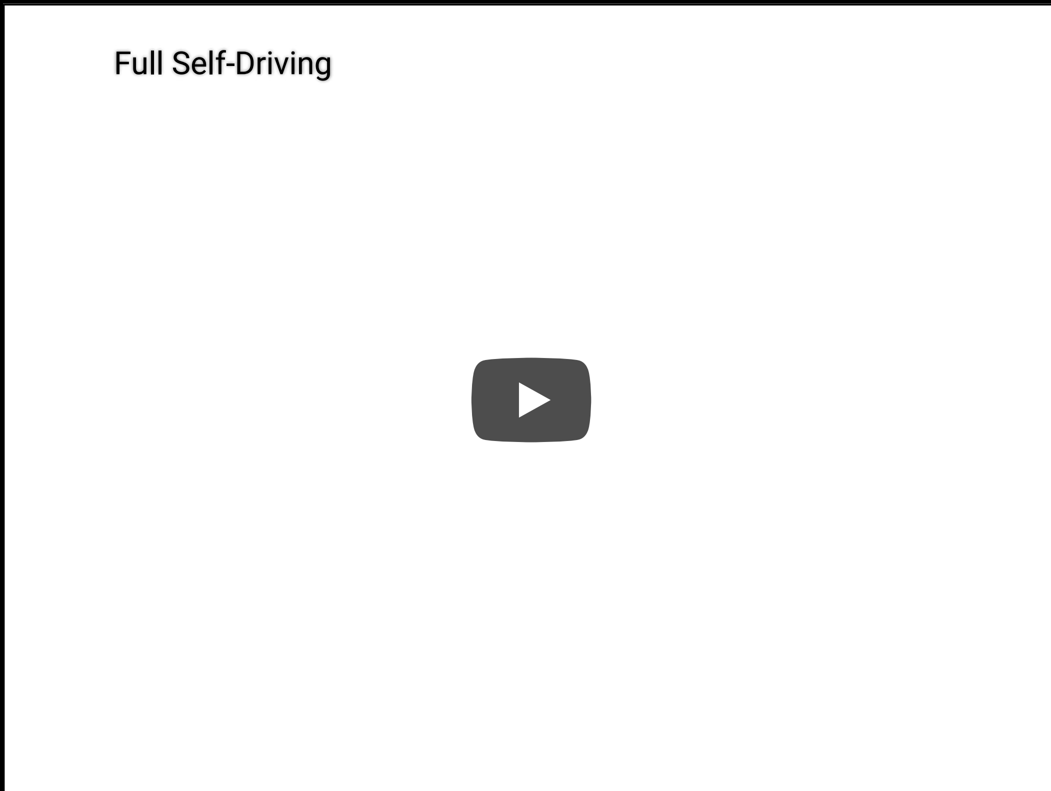


Learning to walk (2017)

- Single algorithm for learning! Nothing is hardcoded.
- Similar to a baby learning to walk.



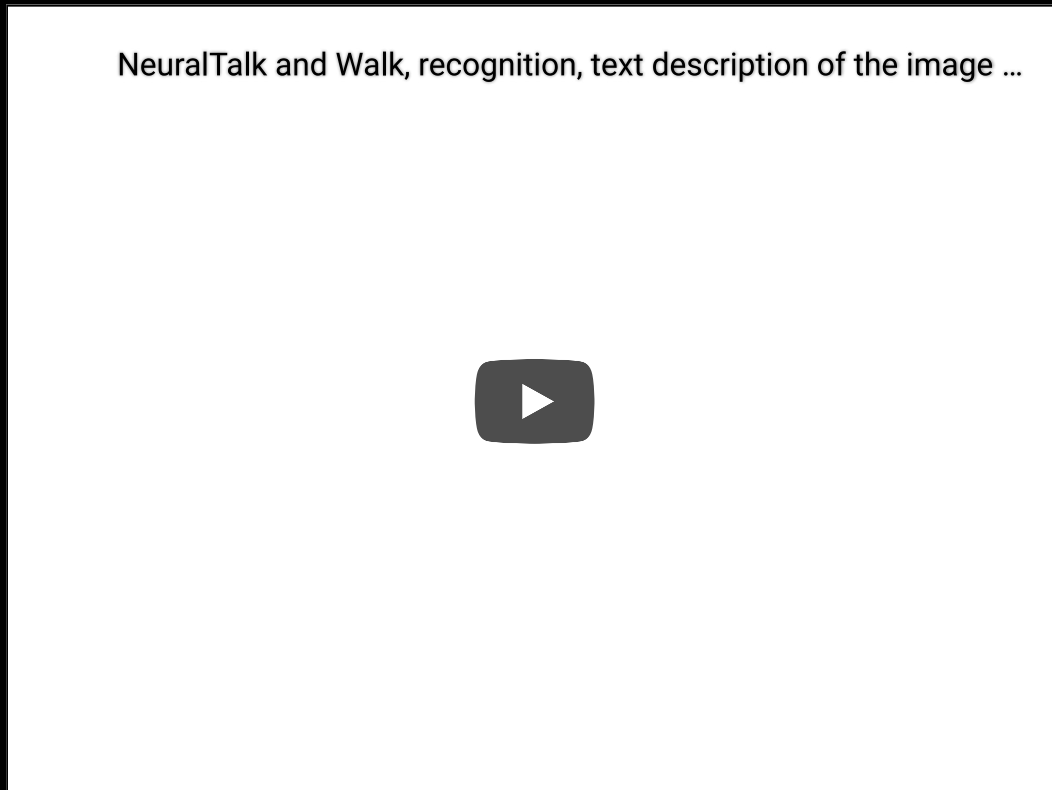
Autonomous cars (NVIDIA, 2016)



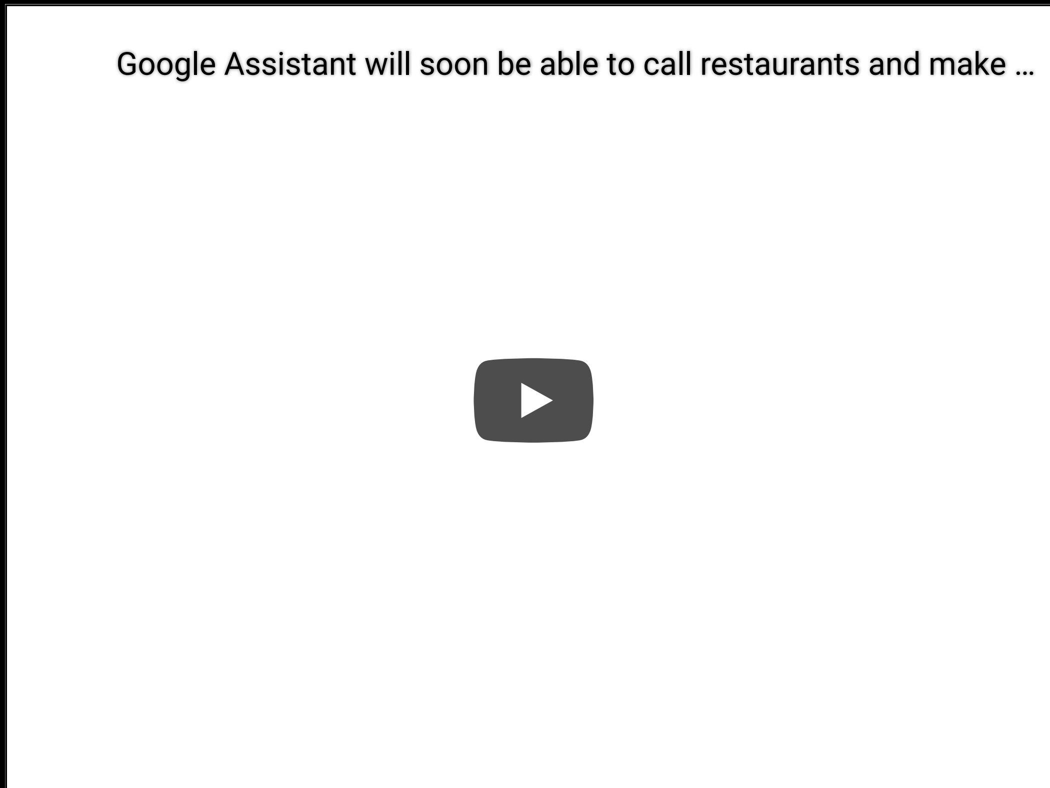
Autopilot (Tesla, 2019)

- A full build of Autopilot neural networks involves 48 networks that take 70,000 GPU hours to train 🔥
- Together, they output 1,000 distinct tensors (predictions) at each timestep.

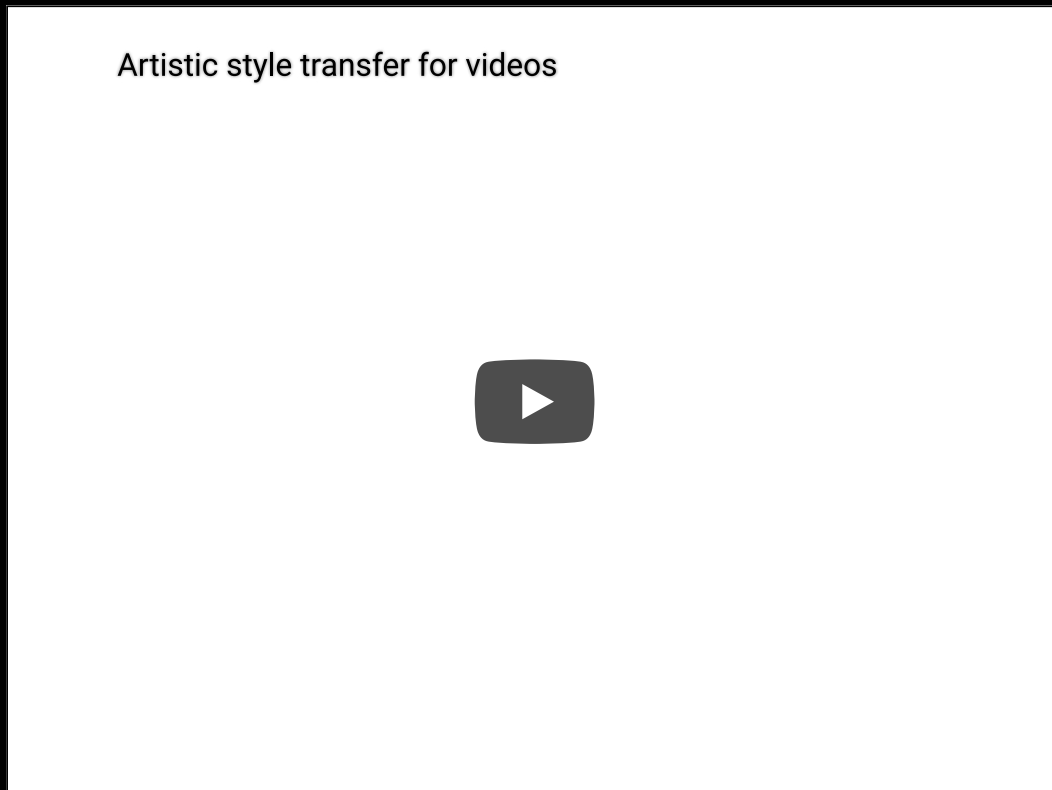
Speech recognition, translation and synthesis (Microsoft, 2012)



Auto-captioning (2015)



Speech synthesis and question answering (Google, 2018)



Artistic style transfer (Ruder et al, 2016)

Image generation (Karras et al, 2018)

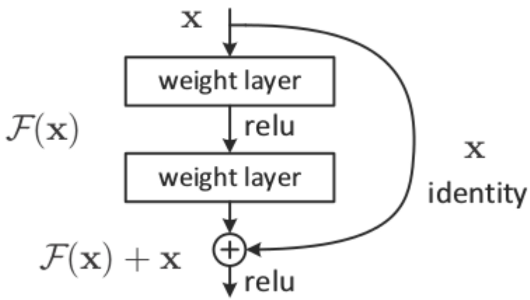
Music composition (NVIDIA, 2017)

Dali Lives (2019)

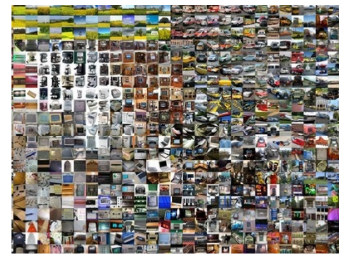
Why does it work now?

A multi-factorial success

Better algorithms and maths



More labeled data

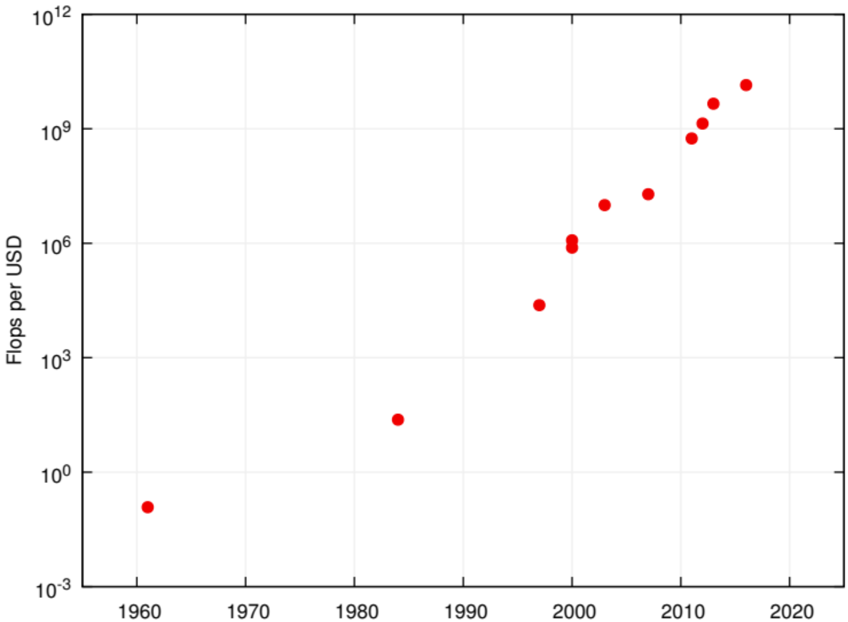


Open source software and models

Caffe2 · Microsoft CNTK · TensorFlow · theano · Chainer · PYTORCH · dmlc mxnet · K

Computing power (GPUs, TPUs, etc.)

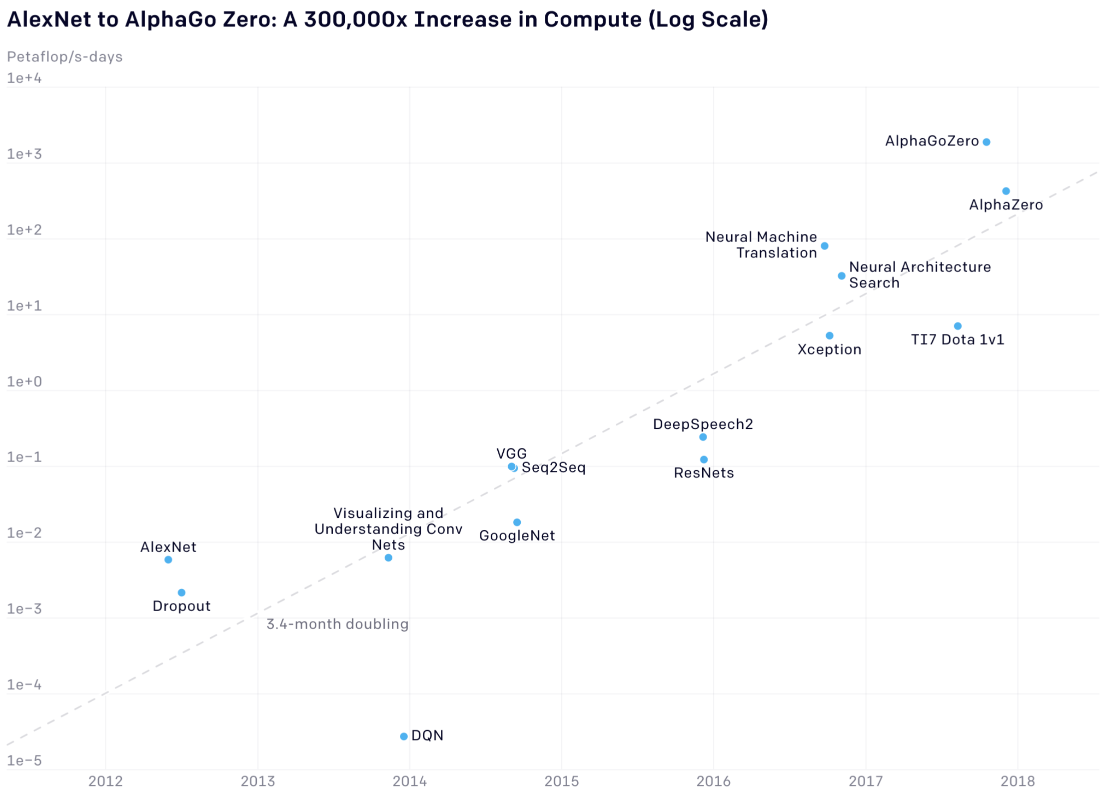
More computation power



	TFlops (10^{12})	Price	GFlops per $
Intel i7-6700K	0.2	$344	0.6
AMD Radeon R-7 240	0.5	$55	9.1
NVIDIA GTX 750 Ti	1.3	$105	12.3
AMD RX 480	5.2	$239	21.6
NVIDIA GTX 1080	8.9	$699	12.7

Credits: Francois Fleuret, EE559 Deep Learning, EPFL.

More computation power



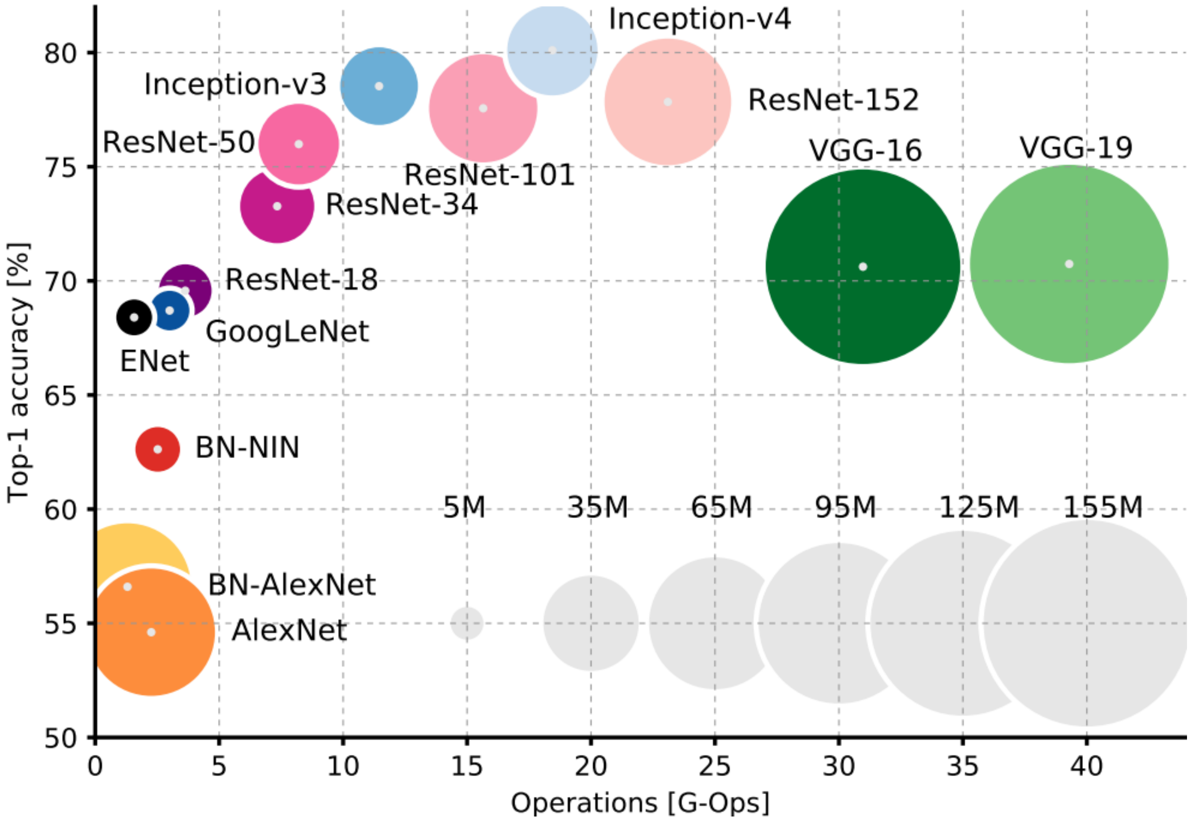
———

Credits: AI and Compute, Open AI, 2018.

More computation power

- 1 petaflop/s-day
- $= 10^{15}$ neural net operations per second for one day
- $= 10^{20}$ operations
- ≈ 100 GPUs for one day
- $\approx 500\text{kWh}$
- https://twitter.com/eturner303/status/1223976313544773634

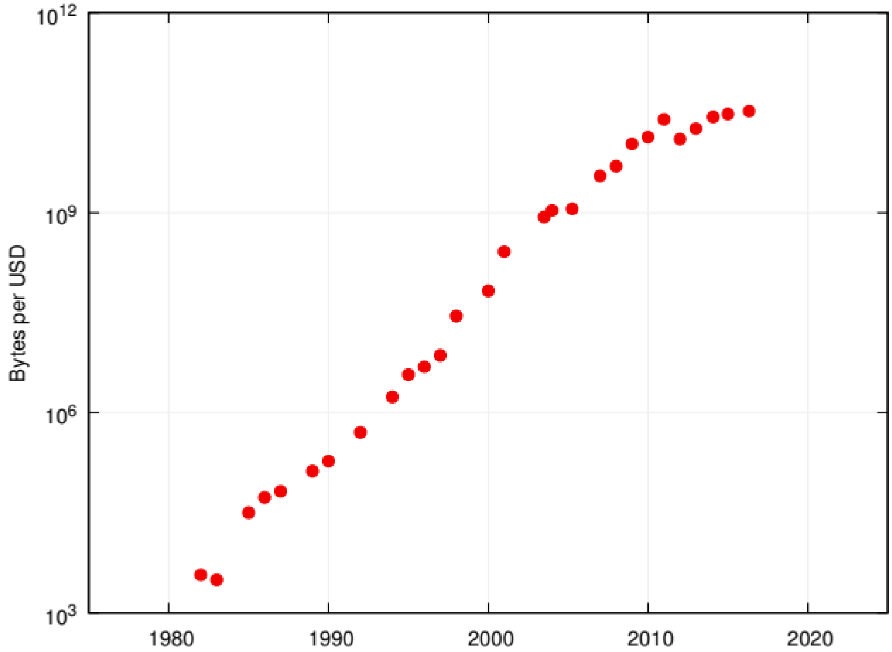
Bigger models



———

Image credits: Canziani et al, 2016, arXiv:1605.07678.

More data



The typical cost of a 4Tb hard disk is less than 100 USD (February 2020).

Open source libraries and frameworks

Automatic differentiation

- `TensorFlow`, `MXnet`, `CNTK`, `Flux.jl`

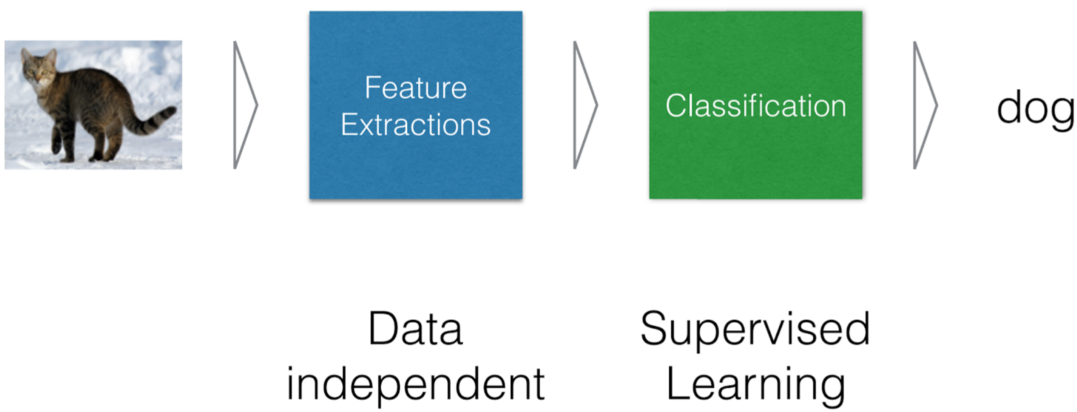
Dynamic and high level

- `TensorFlow2`, `PyTorch`, `Chainer`, `MinPy`, `DyNet`, etc.

A multi-factorial success

- Good old Neural Networks, with more layers or modules
- Non-linear, hierarchical, abstract representations of data
- Very flexible models with any input/output types and sizes
- Differentiable functional programming

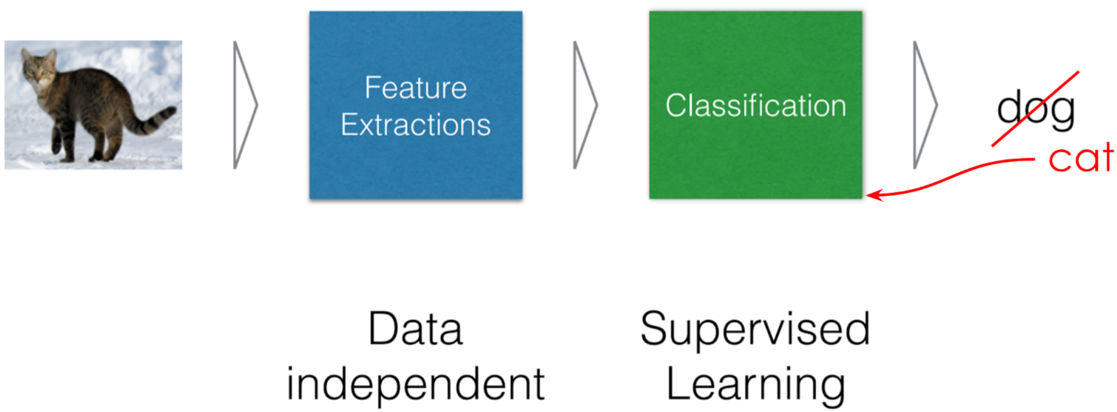
Typical machine learning system



A multi-factorial success

- Good old Neural Networks, with more layers or modules
- Non-linear, hierarchical, abstract representations of data
- Very flexible models with any input/output types and sizes
- Differentiable functional programming

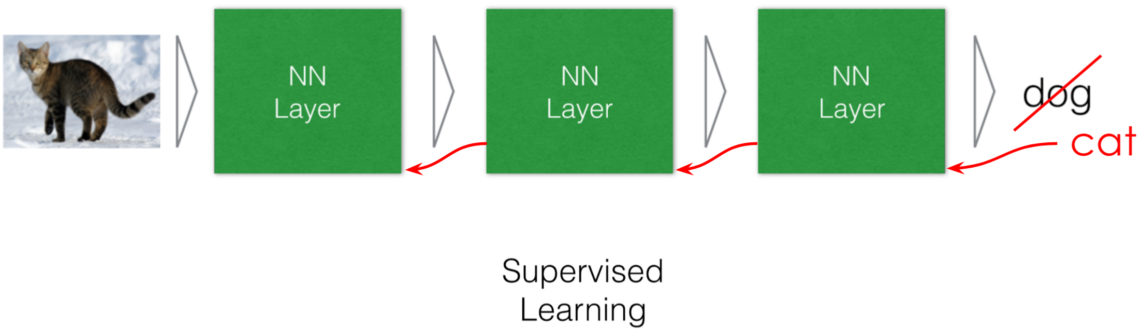
Typical machine learning system



A multi-factorial success

- Good old Neural Networks, with more layers or modules
- Non-linear, hierarchical, abstract representations of data
- Very flexible models with any input/output types and sizes
- Differentiable functional programming

Deep learning (representation learning) system



Feature engineering versus representation learning

Most problems require feature engineering

- Create features by hand using domain knowledge / expertise

Deep learning is representation learning

- Data-driven learning of representations
- Mitigates (or even removes) feature engineering

Example

- Let us illustrate this on a toy example: polynomial logistic regression versus multi-layer perceptron
- We'll show this using the tensorflow playground and in today's lab

Playground time !

Let's go to https://playground.tensorflow.org

Thank you !