

Deep Learning and it's applications in Computer Vision and NLP

Course Introduction

Mikhail Yurushkin
m.yurushkin@broutonlab.com

September 3, 2020



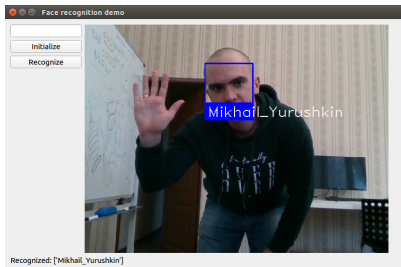
<https://broutonlab.com>

Our projects: Accurate face recognition

- ▶ There are many open-source solutions, but..
- ▶ The aim is to develop highly accurate system (fraud detection, resolving of lighting issues, relatives distinction)

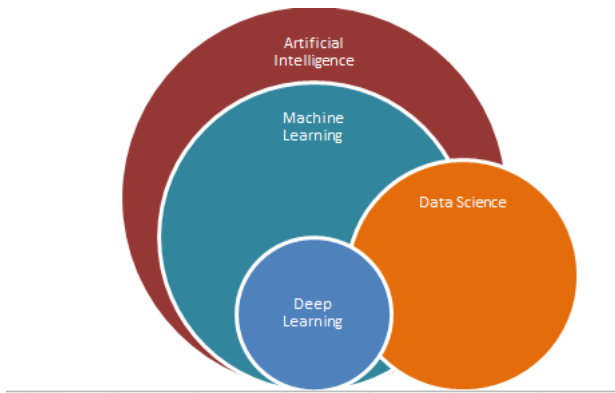


Intel's Depth Camera D415



Face recognition demo

AI vs ML vs DL



Source: <http://becominghuman.ai/ultimate-guide-and-resources-for-data-science-2019-f663f9384fc7>

The structure of course

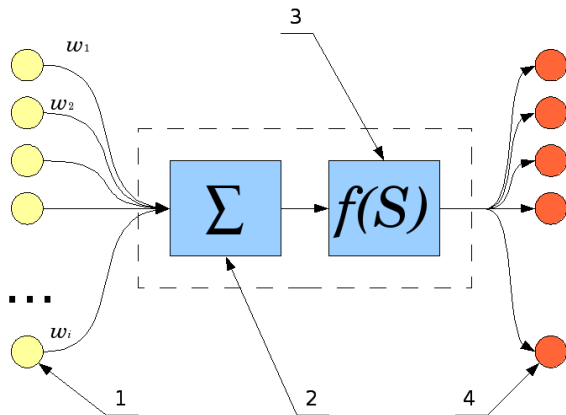
- ▶ Basics of Machine Learning
- ▶ Neural Networks fundamentals: learning via gradient descent, back-propagation, regularization methods, tensor mathematics, batchnorm/dropout
- ▶ Deep Learning in Natural Language Processing (NLP)
 1. Recurrent layers internals (RNN/GRU/LSTM)
 2. Word Embeddings (word2vec'13), Named entity Recognition, Text classification
 3. seq2seq, attention mechanisms'14
 4. neural text translation, image captioning
 5. Google Transformer'17, BERT'18
- ▶ Deep Learning in Computer Vision
 1. Convolutions, poolings
 2. Image classification, Image segmentation, object detection
 3. Siamese networks, Autoencoders, Variational encoders, etc
- ▶ Basics of Deep Learning in Reinforcement Learning. Beating Atari games (frozen-lake, atari/kung fu master, Doom)

Course specifics

- ▶ Practical mostly
- ▶ python 3.6
- ▶ python notebooks
- ▶ Discussion in telegram channel: DL-mmcs20-fall
- ▶ used packages: sklearn, pandas, numpy, matplotlib
- ▶ deep learning packages: keras, tensorflow
- ▶ technical requirements: $RAM \geq 8GB$, $GPU \geq GTX1060$

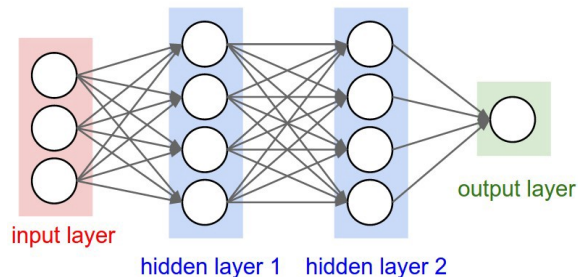
McCullock-Pitts Linear Neuron Model

- ▶ Mathematical model of human neuron
- ▶ 1 - weights of input signals
- ▶ 3 - Activation function
- ▶ red circles - outputs



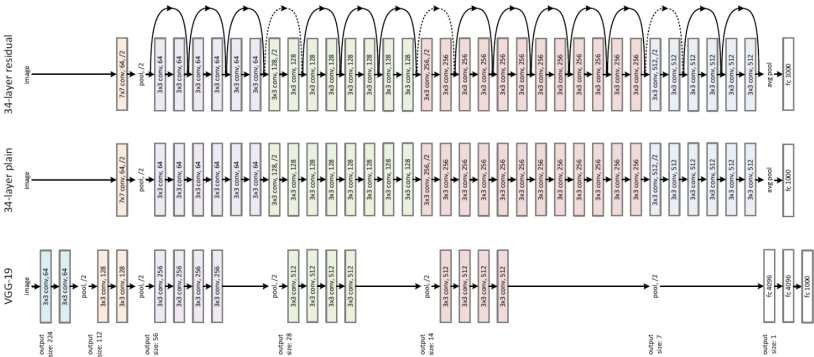
Source: wikipedia

Neural network with small number of layers



- ▶ Were used 30 years ago
- ▶ Don't have big generative power... :(

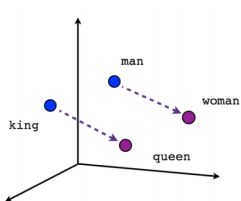
Neural network with **big** number of layers



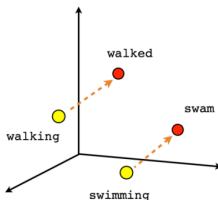
ResNet Architecture

Word Embeddings

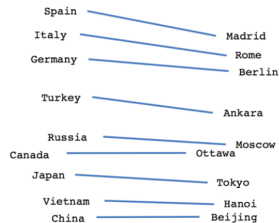
- ▶ Word2Vec'13
- ▶ King - Man + Woman = Queen
- ▶ Paris - France + Italy = Roma
- ▶ Actively used in NLP



Male-Female



Verb tense



Country-Capital

Source: www.tensorflow.org/images/linear-relationships.png

Named Entity Recognition (NER)

- ▶ NER systems are widely used for raw text understanding.
- ▶ The aim is to extract words (or phrases) of particular type.
- ▶ Example listed below: there are 3 types of entities (style, date and person).

The Russian **avant-garde** **ART-STYLE** was a large, influential wave of modernist art that flourished in **Russia** **COUNTRY** from approximately **1890** **DATE** to **1930** **DATE**. The term covers many separate, but inextricably related art movements that occurred at the time, namely **neo-primitivism** **ART-STYLE**, **suprematism** **ART-STYLE**, **constructivism** **ART-STYLE**, **rayonism** **ART-STYLE**, and **Russian Futurism** **ART-STYLE**. Notable artists from this era include **El Lissitzky** **PERSON**, **Kazimir Malevich** **PERSON**, **Wassily Kandinsky** **PERSON**, and **Marc Chagall** **PERSON**.

Neural translation

Source sentence:

Source sentence: Ich denke, dass wir in der nächsten Zeit vor allem an der Entwicklung neuer Sensor typen arbeiten werden müssen, die es ermöglichen, Reaktionen von Personen genauer zu erfassen

Right example of translation: Translation produced by NN:

I think in the near future, we will mainly work on developing new sensors that will enable more precise determination of a person's reaction.

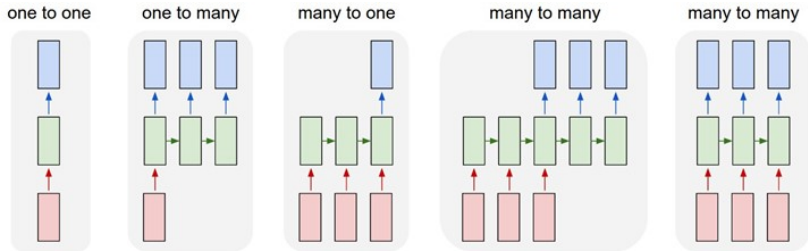


I believe that, in the next time, we need to be able to work in the next few years, so that it can be possible to remove the number of people.



Classification of popular recurrent architectures

- ▶ most of the problems can be classified in this way
- ▶ red blocks - input tokens
- ▶ blue blocks - output tokens

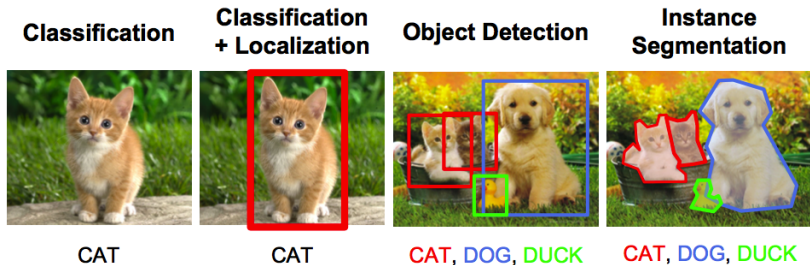


Issues with deep models training

- ▶ Resolving of issues with training of Seq2Seq
- ▶ Invention of attention mechanism (Dzmitry Bahdanau'14)
- ▶ Comparison of LSTM and Convolutional Neural Networks
- ▶ Invention of Google Transformer

Computer Vision: Image Classification, Detection and Segmentation ...

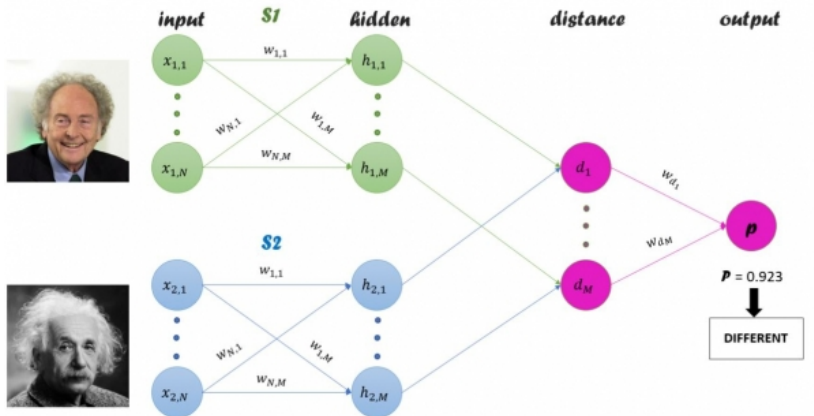
- ▶ The most popular tasks in Computer Vision
- ▶ Still subjects of active research



Source: <http://cs224d.stanford.edu/index.html>

Siamese networks

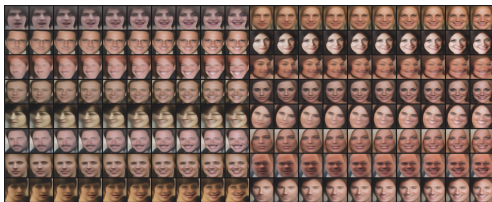
Widely used in ranking, processing of big number of queries



Source: <https://computervision.tecnalia.com/en/2018/07/siamese-neural-networks/>

Variational Autoencoders

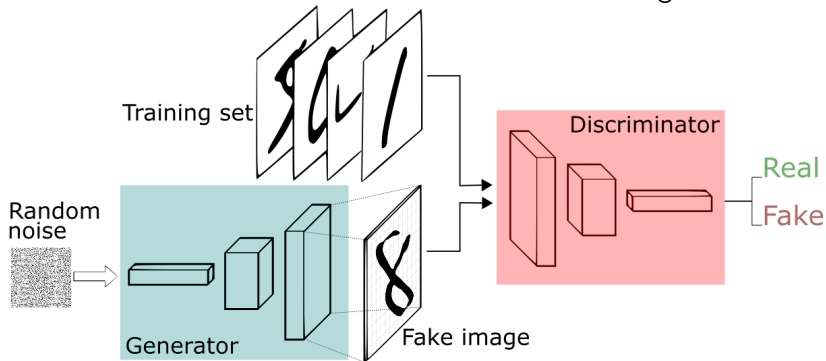
- ▶ Variational autoencoders embed images with extra properties.
- ▶ These properties define high level (semantic) features.
- ▶ You can change feature vector and generate new faces.



Source: github.com/davidsandberg/facenet/wiki/Variational-autoencoder

Generative Adversarial Networks - GAN's (1/2)

- ▶ Game of two participants: Generator and Discriminator
- ▶ Generator learns to produce fake images
- ▶ Discriminator learns to distinct real and fake images



Source:

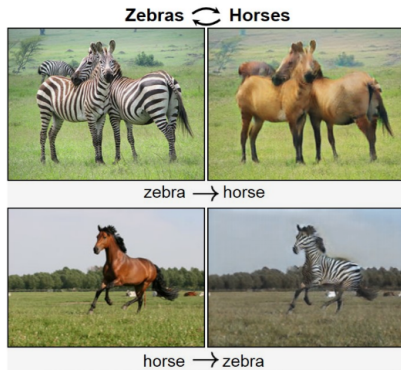
<https://skymind.ai/wiki/generative-adversarial-network-gan>

Generative Adversarial Networks - GAN's (2/2)

- Requires a lot of computation power
- Problems with convergence



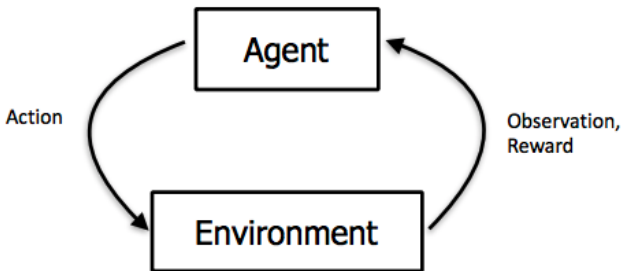
Source: <https://arxiv.org/abs/1710.10196>



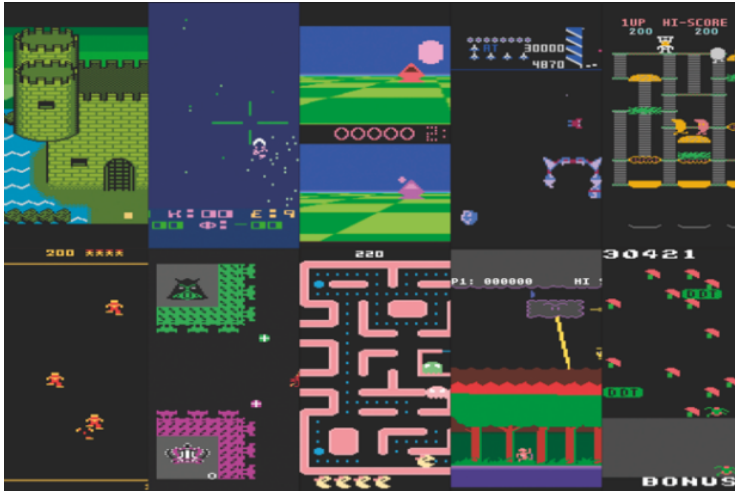
Source: <https://github.com/junyanz/CycleGAN>

Reinforcement Learning (1/3)

- ▶ Agent communicates with environment (via actions)
- ▶ Environment returns agent's new state and reward
- ▶ The aim is to maximize reward - it can be extremely difficult



Reinforcement Learning - is very practical (2/3)



Source: www.retrogamer.net/top_10/top-ten-atari-2600-games/

Reinforcement Learning (3/3)

- ▶ AlphaGo'15, AlphaGo Zeo'17
- ▶ Dota (much more complex environment then 2D games have)
- ▶ Moving robots from Boston dynamics
- ▶ New kinds of weapon are coming...



Extra materials

- ▶ Vorontsov's lectures on Machine Learning:
<http://www.machinelearning.ru/>
- ▶ DL course co-developed by HSE, YSDA and Skoltech
https://github.com/yandexdataschool/Practical_DL