



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Artificial Intelligence, Machine Learning and Deep Learning

part 2

Luca Casini

PhD in Data Science and Computation

Introduction

- The first lesson was about history and theory to get on the same page
- This second lesson is more practice-oriented
- We will briefly talk about some tools and resources you can use to manage data
- We will focus on models that solve common problems that could be useful in your projects
 - Computer Vision
 - Natural Language
 - Audio and Music
 - Health
 - Reinforcement Learning
 - Physics Simulations



USEFUL TOOLS



Data Management: Pandas

- Most common library for data management
- Uses a flexible data structure called DataFrame, similar to an excel table
- Easy and fast manipulation using numpy
 - numpy is the math and algebra library for python
- Offers common statistical analysis function and basic plotting
- Data can be saved in various formats such as .csv and .xlsx



Visualization: Matplotlib & Seaborn

- Most common libraries for plotting and visualization
- useful for plotting
 - data distribution and timeseries
 - training and testing metrics
 - confusion matrices
 - performance bar graphs

The image shows two overlapping web pages. The top page is the Seaborn website, titled 'seaborn: statistical data visualization'. It features a navigation bar with links to 'Gallery', 'Tutorial', 'API', 'Site', and 'Page'. Below the title is a row of six small plots showing various statistical visualizations like histograms, density plots, scatter plots, box plots, and regression lines. A description states: 'Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.' There are buttons for 'Contents' and 'Features'. The bottom page is the Matplotlib website, titled 'matplotlib'. It has a navigation bar with links to 'Installation', 'Documentation', 'Examples', 'Tutorials', and 'Contributing'. Below the title is a code block showing a simple plot:

```
import numpy as np
import matplotlib.pyplot as plt

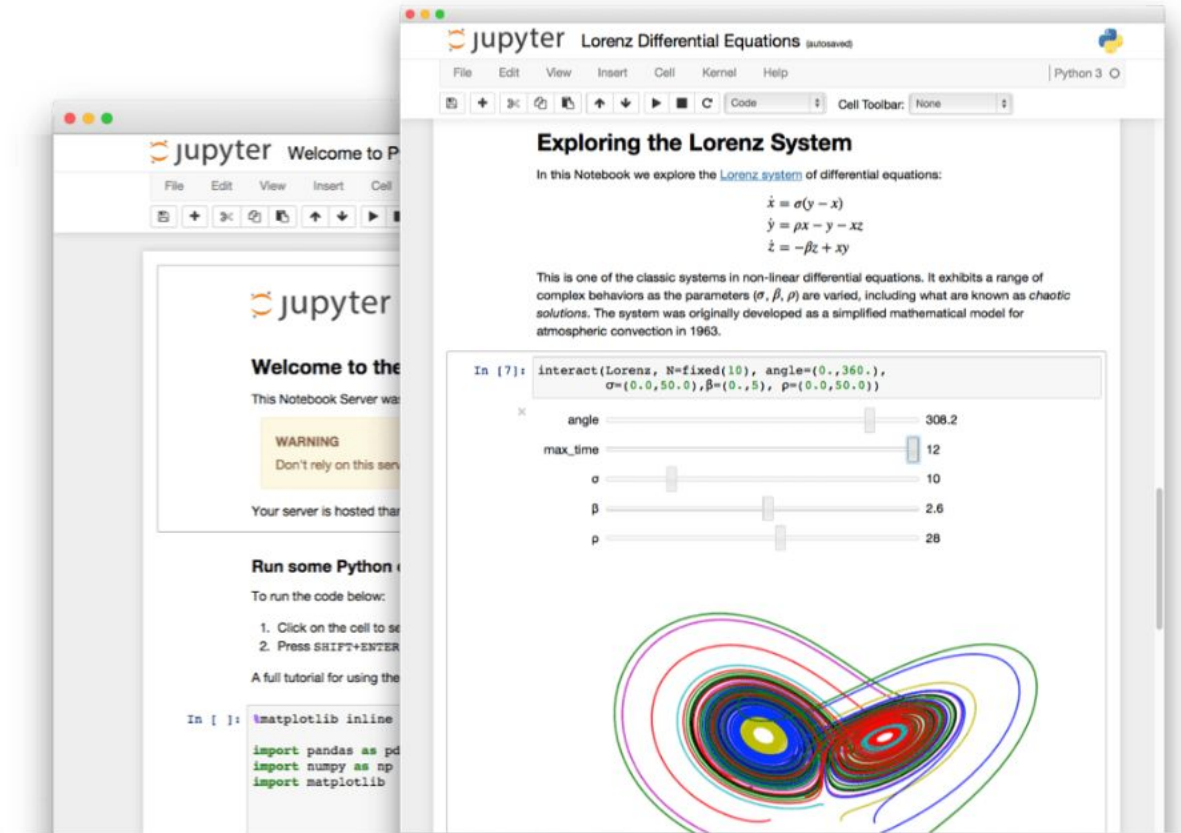
x = np.arange(0, 5, 0.1)
y = np.sin(x)
plt.plot(x, y)
```

 The page also includes a 'Quick search' bar, a 'Table of Contents' section, and a 'Related Topics' section.



Jupyter Notebooks

- Notebook system similar to Wolfram Mathematica
- Code is organized in cells that can be executed independently
- Support Markdown notation
- Output is visualized under the cell and it can also show images, plots and tables automatically formatted if you are using pandas, seaborn, etc.



Anaconda

- Integrated environment for all platforms
- Works with Python and R
- Includes all popular software libraries for Data Science
- Comes with its own packet manager: conda
- Can create virtual python environment so you can use different versions of the same libraries without conflicts.
- It is the easiest way to install and manage deep learning frameworks in windows



COMPUTER VISION



Image Classification

- Maybe the most common vision task of all
- Pytorch and Keras have many precompiled models available
 - They also come with weights if you just want to use them without training
- Famous models like Inception and ResNet are quite big and maybe not needed
- There are models specific for mobile applications like MobileNet

<https://pytorch.org/docs/stable/torchvision/models.htm>

TORCHVISION.MODELS

The models subpackage contains definitions of models for addressing different tasks, including: image classification, pixelwise semantic segmentation, object detection, instance segmentation, person keypoint detection and video classification.

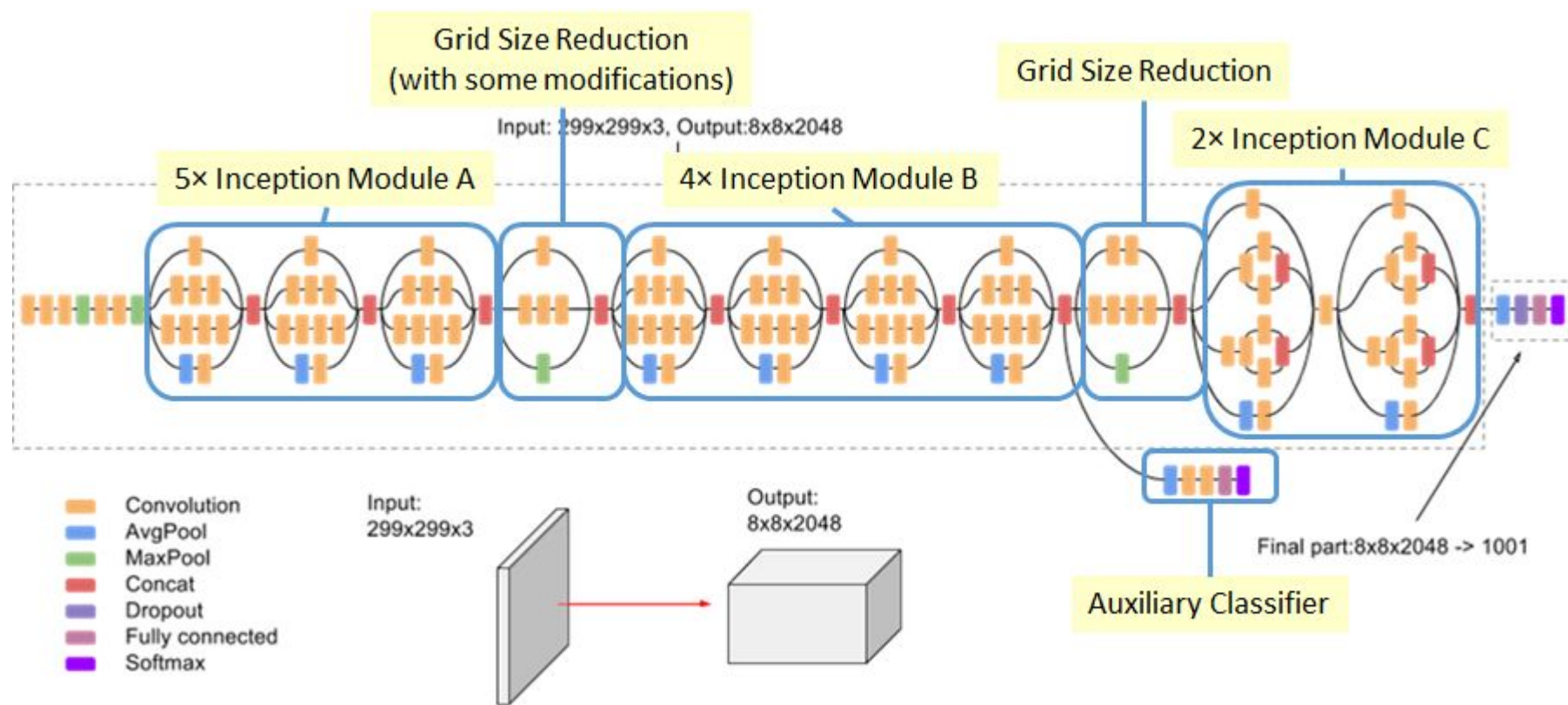
Classification

The models subpackage contains definitions for the following model architectures for image classification:

- AlexNet
- VGG
- ResNet
- SqueezeNet
- DenseNet
- Inception v3
- GoogLeNet
- ShuffleNet v2
- MobileNet v2
- ResNeXt
- Wide ResNet
- MNASNet



Inception v3

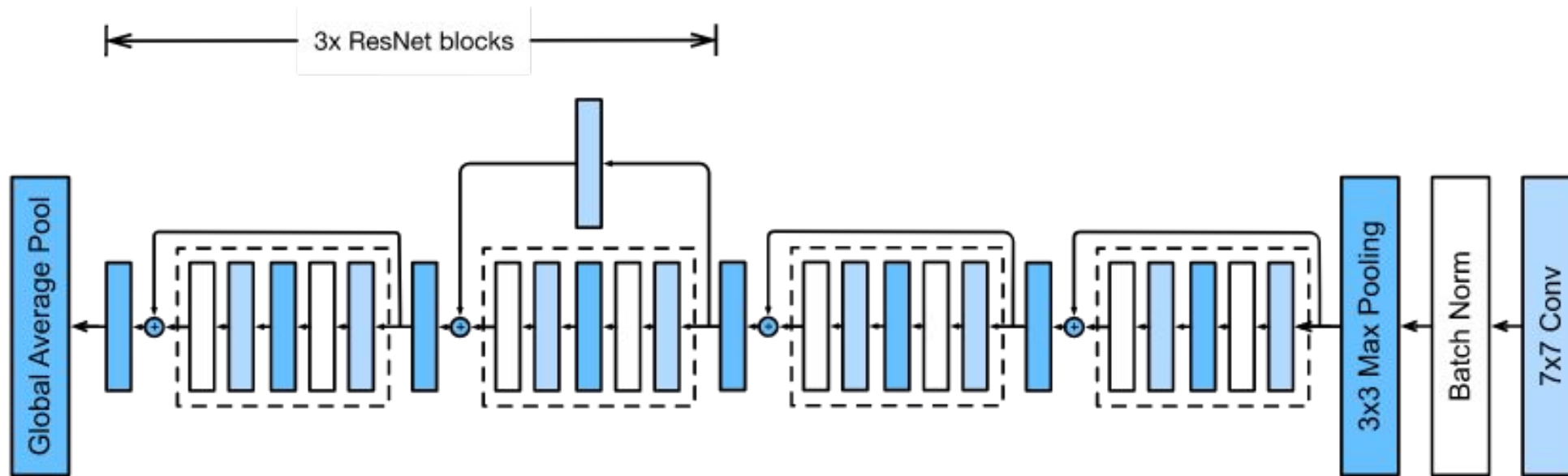


[Review: Inception-v3 — 1st Runner Up \(Image Classification\) in ILSVRC 2015](#)



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ResNet



[7.6. Residual Networks \(ResNet\) — Dive into Deep Learning 0.7 documentation](#)



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Object Detection: YOLO

- Object detection is the task of finding a bounding box around a specific object. basically image classification + regression for the coordinates
- There are two very famous models R-CNN and YOLO
- R-CNN is precise but slow
 - find proposal regions iteratively, classify each of them
- YOLO is less accurate but very fast (real time)
 - divide the image into tiles, for each of the predict a bounding box, then classify
 - remove low confidence bounding box and predictions

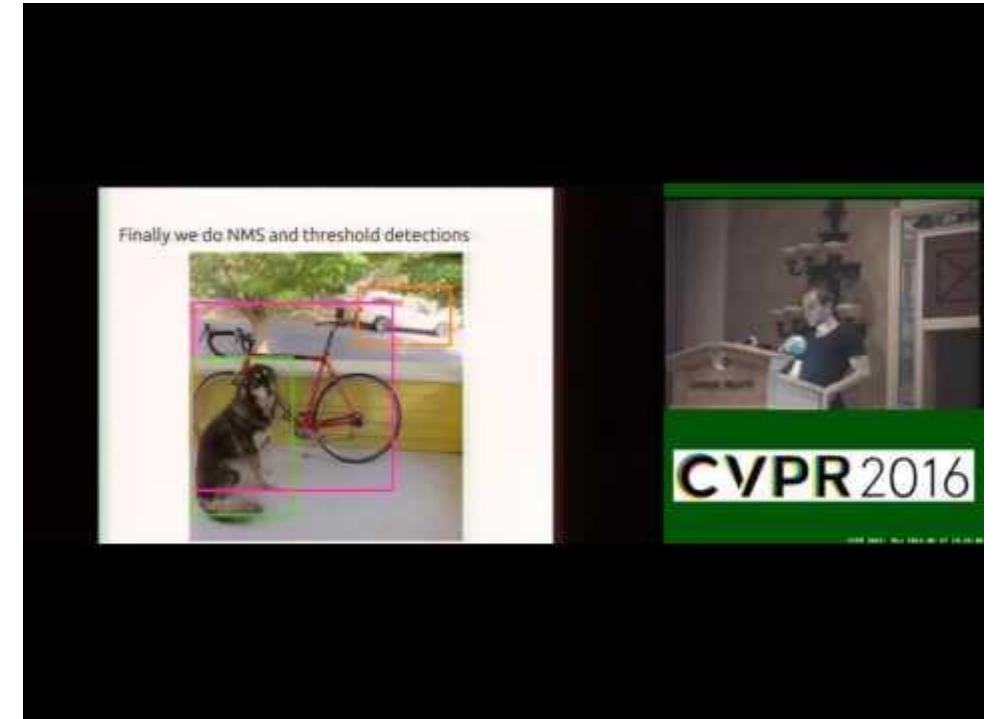

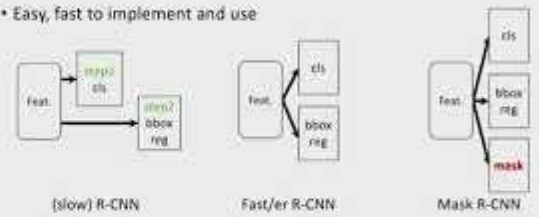


Image Segmentation




What is Mask R-CNN: Parallel Heads

- Easy, fast to implement and use

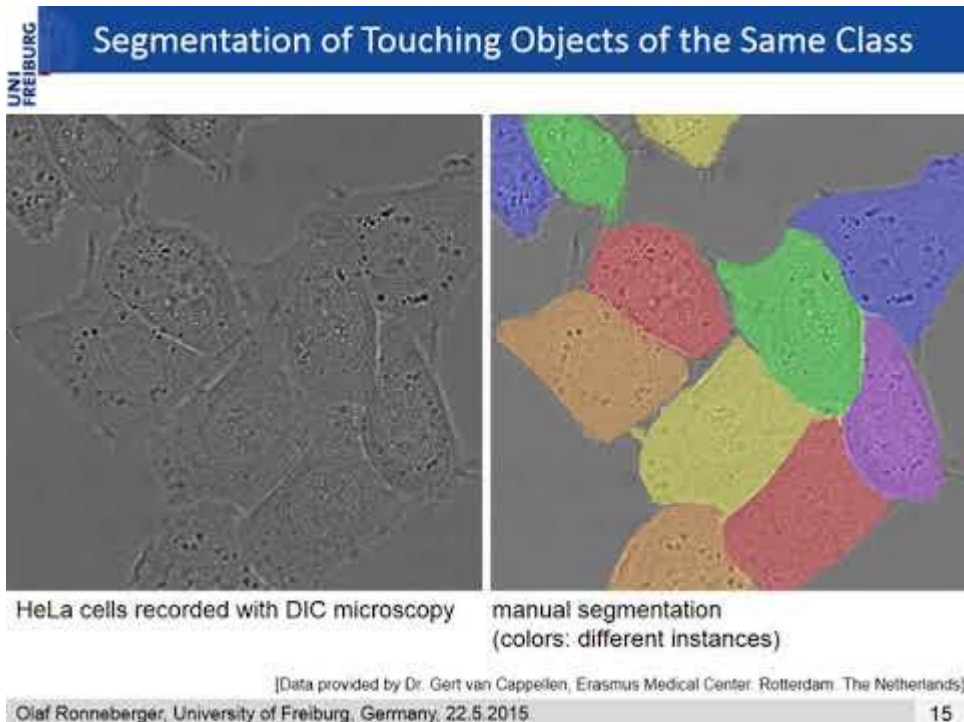


(slow) R-CNN Fast(er) R-CNN Mask R-CNN



- Image segmentation is the task of finding the shapes of the objects present in an image, not just bounding boxes
- Mask-R-CNN is an evolution of R-CNN for object detection that also outputs a mask for each object
- U-Net is a very popular network for segmentation that gets its name from its shape. It uses inverse convolutions and residual connections

Image Segmentation

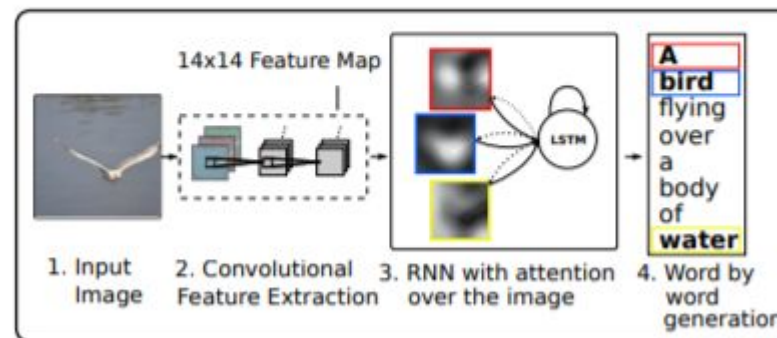


- Image segmentation is the task of finding the shapes of the objects present in an image, not just bounding boxes
- Mask-RCNN is an evolution of R-CNN for object detection that also outputs a mask for each object
- U-Net is a very popular network for segmentation that gets its name from its shape. It uses inverse convolutions and residual connections

Image Captioning

- Give a textual description of an image
- Google Show, Attend and Tell
 - CNN based on VGG19 analyses the image
 - A RNN with attention looks at the feature
 - Text is generated by the RNN
 - it's a seq2seq model

Figure 1. Our model learns a words/image alignment. The visualized attentional maps (3) are explained in Sections 3.1 & 5.4



A stop sign is on a road with a mountain in the background.

A man wearing a hat and a hat on a skateboard.

[Show, Attend and Tell: Neural Image Caption Generation with Visual Attention](#)



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Super Resolution

- Starting from a low resolution image obtain a high resolution one
- Upscaling and Interpolation give blurry results
- Using CNN partially solved the problem
 - Upscale the image
 - refine the results with a CNN
- GANs are now the state of the art for super-resolution

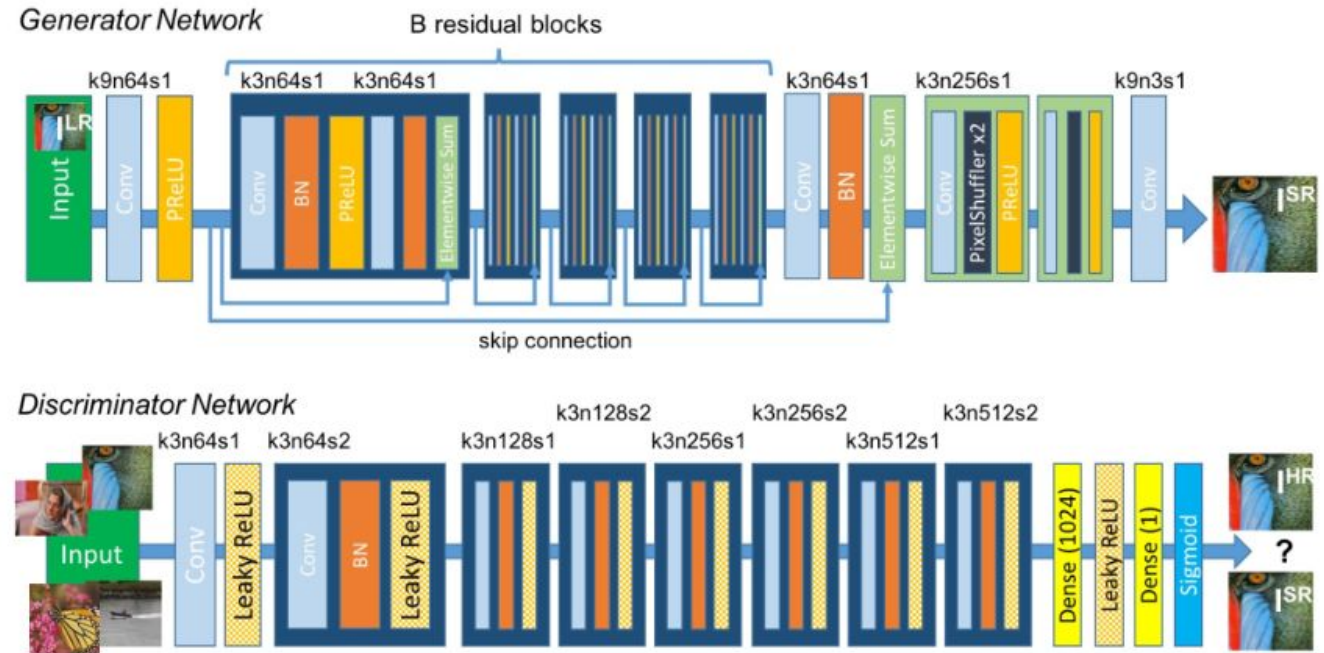


Figure: Architecture of Generator and Discriminator Network in SRGAN. (source)

[An Evolution in Single Image Super Resolution using Deep Learning](#)

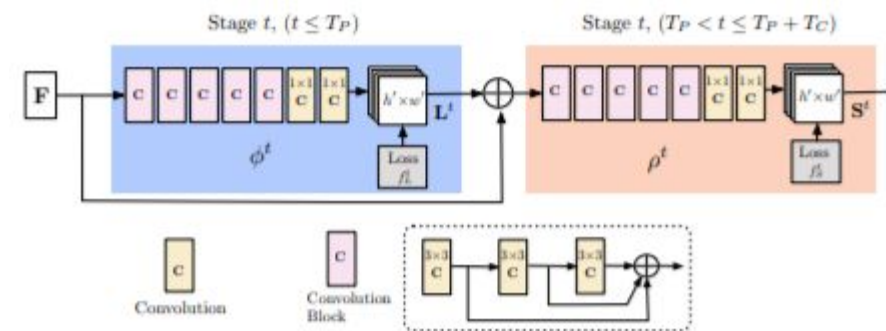


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Pose Estimation

- Given an image, output a skeleton representing the pose of the subject
- Ideally it should be fast enough to be used on video
 - e.g. gesture recognition
- Openpose uses multistage CNNs (based on VGG19)
 - works for 2d and 3d estimation
 - can be applied to webcam footage in near real-time
 - also has a unity plug-in for motion capture

[CMU-Perceptual-Computing-Lab/openpose: OpenPose: Real-time multi-person keypoint detection library for body, face, hands, and foot estimation](https://cmu-perceptual-computing-lab.github.io/openpose/)



NATURAL LANGUAGE PROCESSING



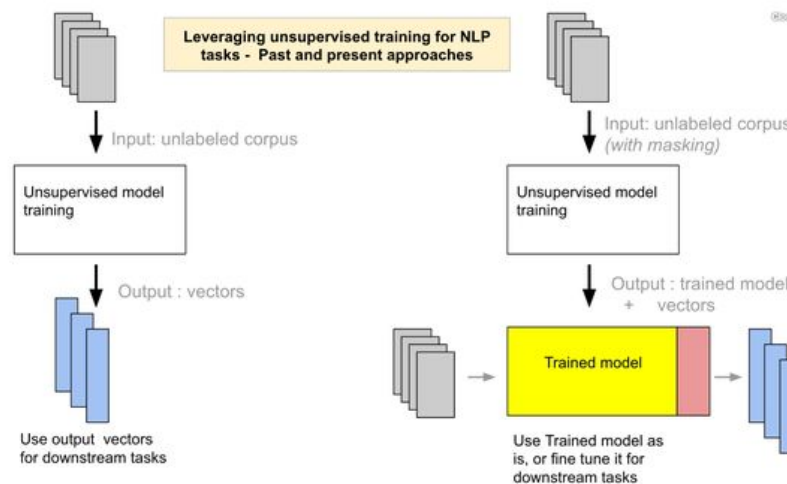
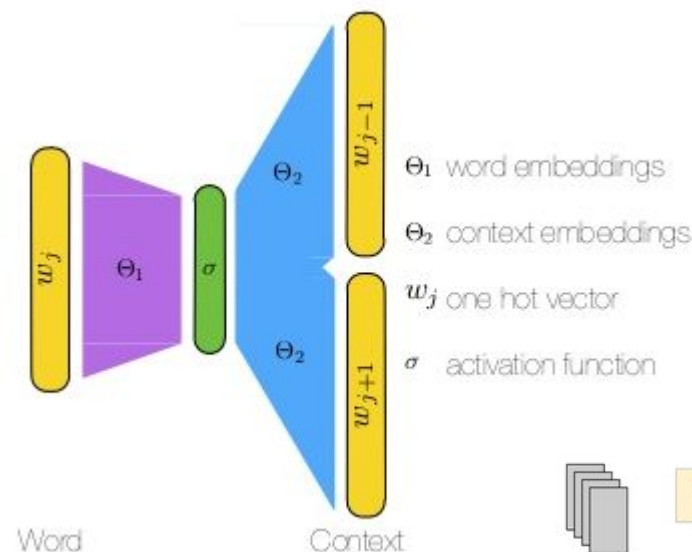
Natural Language Processing

- Beside vision is perhaps the most researched field of application of AI (and deep learning)
- Until a few years ago Recurrent Neural Networks ruled the NLP world
- Transformers are now the SotA architectures
 - BERT and GPT-2 are the best performing
- NLP task include
 - Text understanding
 - Sentiment Analysis
 - Topic Detection
 - ecc.
 - Machine translation
 - Text-to-Speech and Speech-to-Text



Word Embeddings

- Word Embeddings are vector representations of words
- A large vocabulary cannot be represented as one-hot vector easily
- Neural nets can learn of to project the dictionary into a vector space that conserves the semantic relationship of words
- Word2Vec and GLOVE are two of the most popular embeddings, there are pretrained models available



Past approach. Trained models (often shallow) are discarded. Only vector outputs of training used for downstream tasks (model examples: Word2vec, Glove, Fasttext etc.)

Present approach. Output of training - model and vectors. The model is used as is for generating word/sentence vectors, or can be fine tuned (supervised) with additional layers for specific tasks like NER. (model example: e.g. BERT)

[The Illustrated Word2vec – Jay Alammar – Visualizing machine learning one concept at a time](#)

GPT-2 and BERT

- Google BERT and OpenAI GPT-2 are the two best available models at the moment for NLP tasks
- Based on transformers
- Unsupervised pre-training followed by fine-tuning for specific tasks
- They can answer questions and create text summations, as well as solve classic nlp problems like POS-tagging, NER, etc.

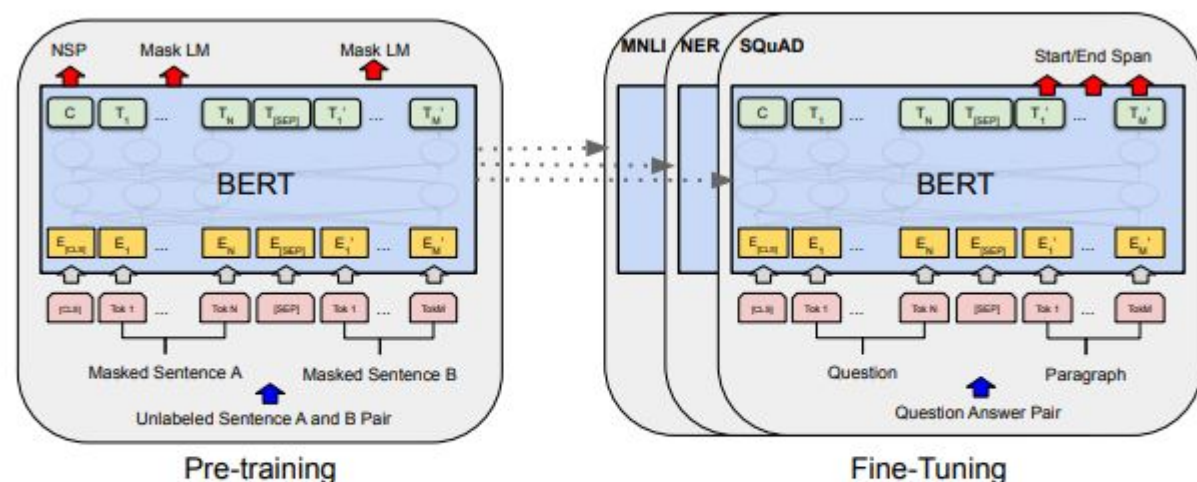


Figure 1: Overall pre-training and fine-tuning procedures for BERT. Apart from output layers, the same architectures are used in both pre-training and fine-tuning. The same pre-trained model parameters are used to initialize models for different down-stream tasks. During fine-tuning, all parameters are fine-tuned. [CLS] is a special symbol added in front of every input example, and [SEP] is a special separator token (e.g. separating questions/answers).

<https://arxiv.org/pdf/1810.04805.pdf>

Sentiment Analysis

- It's the problem of recognizing if a text's sentiment is positive or negative (or something in-between)
- Useful for analysing reviews and user comments when there is not an assigned score (e.g. Likert scale)
- The dataset is comprised of pair of sentences and label corresponding to the sentiment assigned to it. Some models are able to output sentiments of part of the sentence

<https://github.com/bentrevelt/pytorch-sentiment-analysis>



Automatic Translation and Chatbots

Neural Machine Translation (NMT)

- sequence-to-sequence models (seq2seq)
 - RNN-based or Transformer-based
- you need a paired dataset of sentences in the two languages
- The network learns how to associate a word embedding in a language to the word embedding in another
- BLEU (bilingual evaluation understudy) algorithm for scoring

Chatbots

- also a seq2seq model
- dataset is comprised of paired sentences: question and answer
- the neural network learns how to answer questions
- Scoring can be some sort of distance with expected output but human validation is important
- BERT + GPT-2 reddit comment bot



Text-to-Speech and Speech-to-text

Text-to-Speech

- transform text into phonemes
- the phonemes can be sound samples or the output can be the input of synthesizer
- Google Wavenet model is a SotA architecture that uses CNN to create a directly a soundwave
 - it's able to replicate incidental sounds like breathing
 - can also generate other sounds like music

[WaveNet: A generative model for raw audio](#)

[Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone](#)

Speech-to-text

- Transform Audio into text
- The model must recognize the right phoneme but also the right word
 - slight pronunciation differences should be allowed
- Spectrogram of audio is analyzed and the letter is predicted
 - slice the spectrogram and classify
 - use a seq2seq model (output the word)

[The 3 Deep Learning Frameworks For End-to-End Speech Recognition That Power Your Devices](#)



AUDIO AND MUSIC



Music Information Retrieval

- Working with audio signals with deep learning means using Spectrograms
- Spectrogram are two dimensional and can be seen as images. CNNs are a good solution
- Audio is also a sequence. RNNs are useful for finding relationships through time
 - Transformers are now being applied to those task too
- A model of this kind can be used as a base for recommender systems, genre classification, editing tasks, etc.

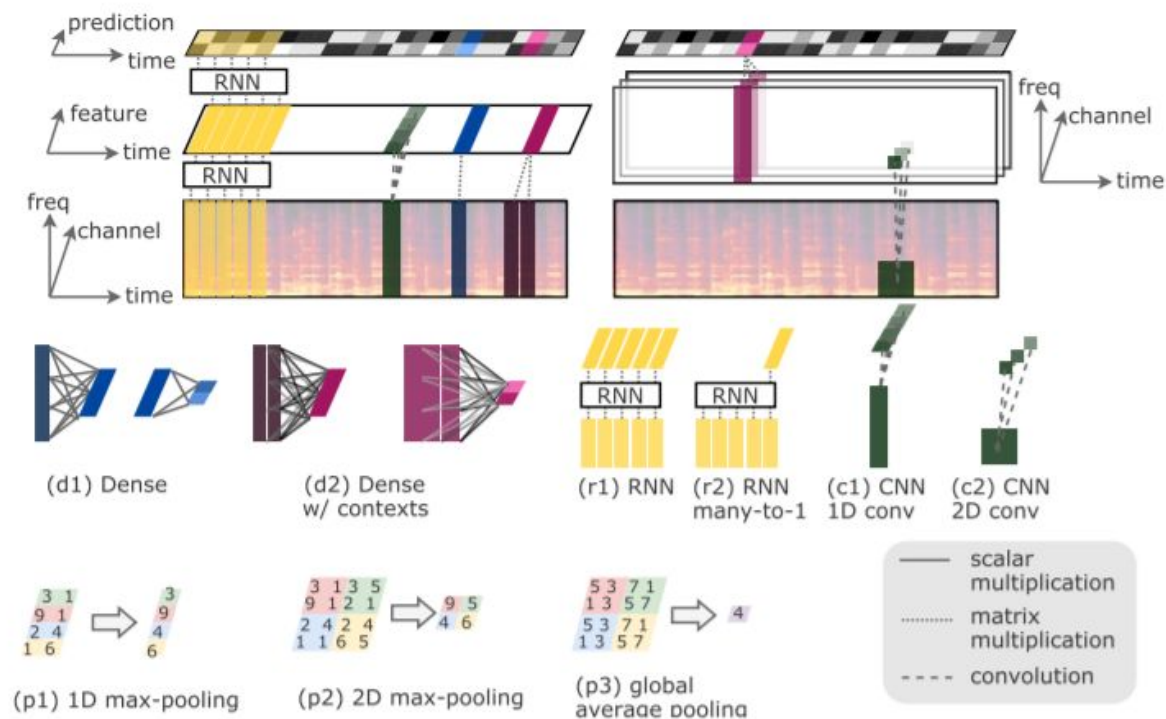


Figure 8: Neural network layers in the context of MIR

[\[1709.04396\] A Tutorial on Deep Learning for Music Information Retrieval](#)

Music Generation: Google Magenta

- Music can also be treated as symbolic language using techniques from NLP
 - Similarly to NLP, RNN are very popular for modelling music and now Transformer are taking over
- Google Magenta is a very successful project that aims at creating DL tools for musicians and creatives in general
 - Generative models for symbolic music
 - Sound interpolation in the latent space
 - Automatic Transcription of piano scores
 - Automatic interpretation of piano music



[Magenta Tensorflow](#)



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

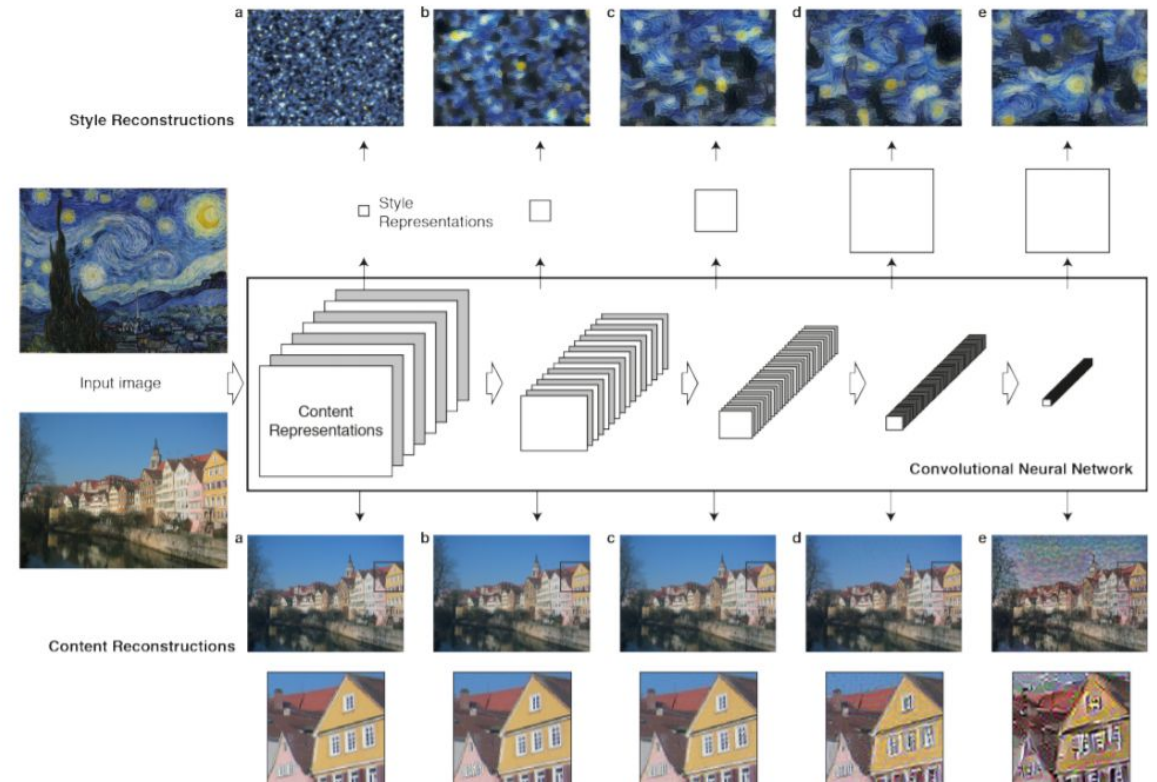
ARTS



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Style Transfer


- Became very popular in 2015
- Use CNN to transfer the style of a painting into a photo
- Take a CNN classifier and use it as a feature extractor
 - A special loss for difference between original image and result (first layers)
 - A special loss for difference between style image and result (last layers)
 - Minimizing the sum of losses gives a good style transfer
- GANs can also be used for the same effect now



[\[1508.06576\] A Neural Algorithm of Artistic Style](#)


Visual Art with GANs

- GANs are becoming very popular for art generation
- Beside style transfer techniques they can be used to generate basically everything in a random or constrained way
- [GAN Paint](#)
- [MuseGAN](#)
- [AI Can Manipulate Video to Make Everybody Dance Now](#)
- [David, 2018 | The first art sculpture made with a generative adversarial network \(GAN\)](#)

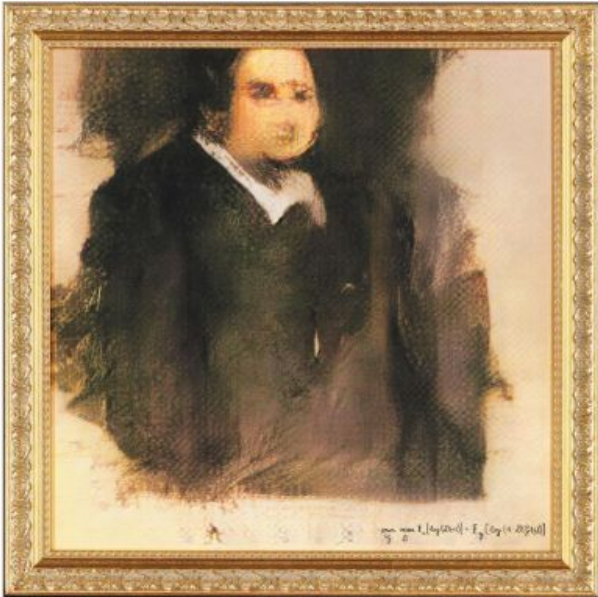


Oct 25, 2018, 08:29pm EDT

What Happens Now That An AI-Generated Painting Sold For \$432,500?




William Falcon Former Contributor
Science
I write about AI, its business use and ethical use.



Portrait of Edmond Belamy, 2018, created by GAN (Generative Adversarial Network).
COURTESY CHRISTIE'S IMAGES LTD.

Today at Christie's New York, the first AI-generated painting sold for \$432,500 - it was expected to sell for \$10,000. A type of AI algorithm known as *Generative Adversarial Networks (GAN)* and trained by a group called *Obvious art* generated the work.

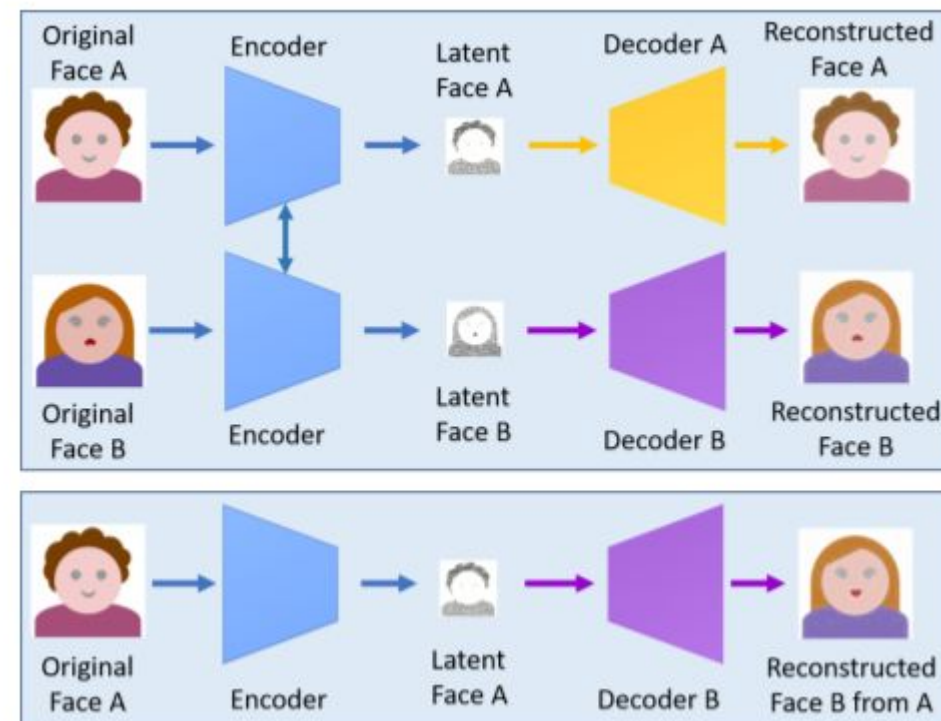


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Deepfakes

- Maybe the most popular application of GANs and DL in the 2 years
- Use encoder-decoder models to learn how to represent faces and then swap them
- GANs can be used to create a generator of very high quality that can fool humans
- The concept works on everything beyond faces
- Very powerful but very dangerous

[\[1909.11573\] Deep Learning for Deepfakes Creation and Detection](#)

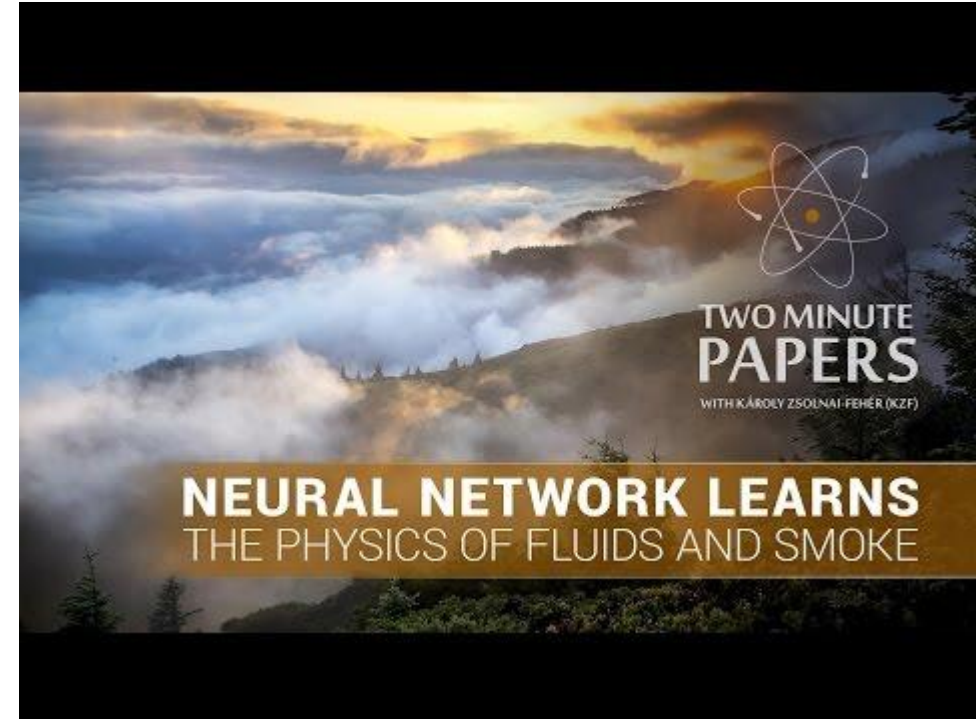


PHYSICS SIMULATIONS



PHYSICS SIMULATIONS

- Physical models can be very complicated and difficult to simulate
- If equations exist and only difficult to compute a NN can be trained to approximate them
- Not very useful for scientific calculations but it can be a game changer in the entertainment world
 - e.g. water simulation



[Accelerating Eulerian Fluid Simulation With Convolutional Networks](#)



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

HEALTH



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Computational Drug Discovery

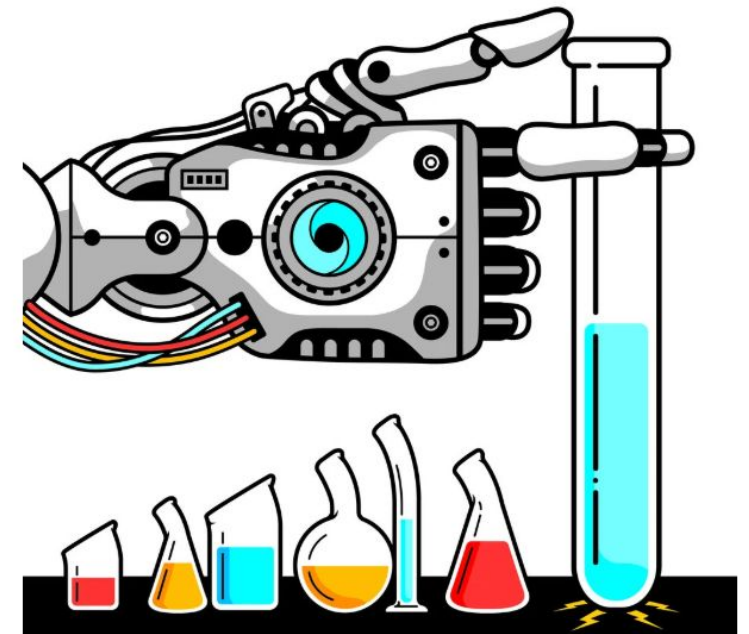
- NN can help in finding new drugs and molecules
- Protein folding is a very important task in the field
- in 2019 Deepmind took part in the CASP13 competition on protein folding against field experts
- Their model, AlphaFold, dominated the competition by using neural networks to model the physical properties of the protein folding process

[AlphaFold: Using AI for scientific discovery](#)

[Making New Drugs With a Dose of Artificial Intelligence](#)

The New York Times

Making New Drugs With a Dose of Artificial Intelligence



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Computer-aided Diagnosis

- AI model can be used to help doctors in their diagnoses
- Very delicate task as each error could mean someone is dying
- Data is difficult to gather
- Big scale projects show that neural network could surpass medical experts and that AI and doctors together can become extremely efficient
 - Google Breast Cancer prediction
01/01/2020 *"there was a 3.5 percent reduction in false positives and an 8.1 percent reduction in false negatives,"*

[International evaluation of an AI system for breast cancer screening](#)

[Using AI to improve breast cancer screening](#)

MENU ▾

nature

Article | Published: 01 January 2020

International evaluation of an AI system for breast cancer screening

Scott Mayer McKinney , Marcin Sieniek, [...] Shravya Shetty 

Nature **577**, 89–94(2020) | [Cite this article](#)

40k Accesses | **12** Citations | **3460** Altmetric | [Metrics](#)



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Thank You!

Luca Casini
PhD in Data Science and Computation

luca.casini7@unibo.it

www.unibo.it