

Past and Present of Deep Learning: Now What?

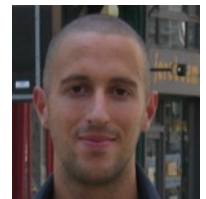
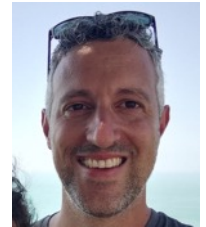
STEFANO MELACCI

DEPARTMENT OF INFORMATION ENGINEERING AND MATHEMATICS (DIISM), UNIVERSITY OF SIENA, ITALY

PRESENTED AT ESTIA (VIRTUAL MEETING) - FEBRUARY 14, 2022

Something About Me

- ▶ Associate Professor – DIISM, University of Siena (Siena, Italy), UNISI
 - ▶ Research Topics: Machine Learning, AI
- ▶ Previous positions:
 - ▶ Assistant Professor, UNISI (3 years)
 - ▶ Research Manager...**in the industry**, Italy/France (3 years)
 - ▶ Post-Doc, UNISI (5 years)
- ▶ As a student:
 - ▶ Visiting Scientist, Ohio State University, Columbus (OH, USA) (6 months)
 - ▶ PhD Student, UNISI (3 years)
 - ▶ Master of Science in Computer Engineering (cum Laude), UNISI



Siena Artificial Intelligence Lab

- ▶ **SAILab**, <https://sailab.diism.unisi.it>
 - ▶ Demos, software, ...



Overview

- ▶ This seminar is about Machine Learning, in light of my personal experience as a researcher in the field
 - ▶ Emphasis on Deep Learning
- ▶ Three main parts
 1. What happened in (approx.) the last 20 years?
 2. What is going on now?
 3. What's next? (well...I don't have a **crystal ball** 😊)



Past and Present of Deep Learning: Now What?

PART 1

Early 2000: What is Machine Learning?

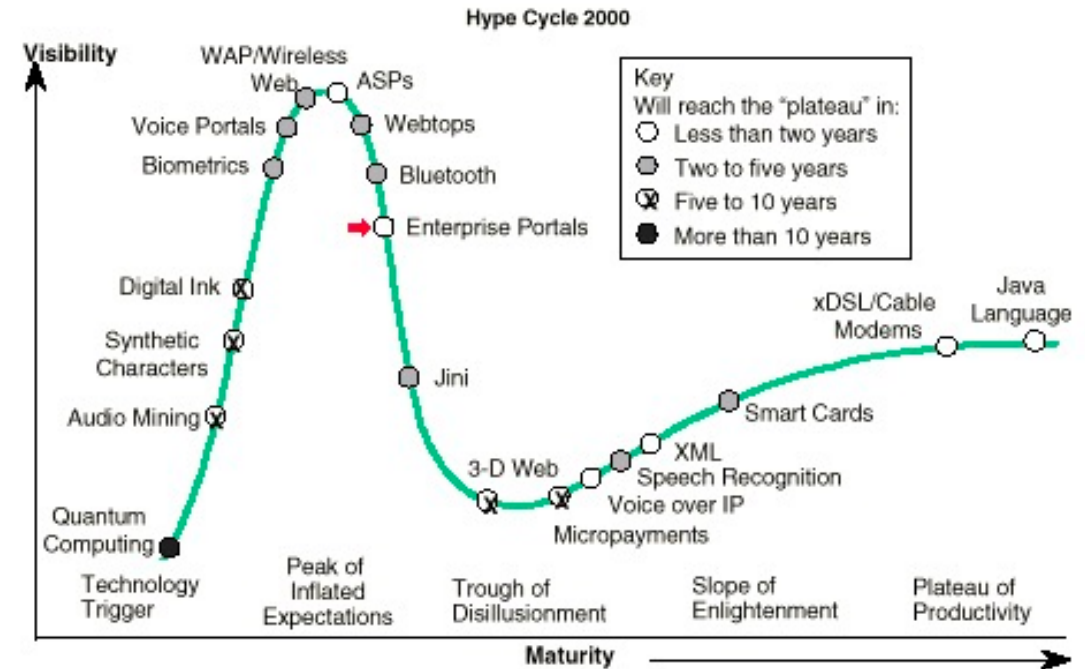
- ▶ Our story starts from **early 2000** (*approx.*)
- ▶ Mentioning Machine Learning/AI with a random person met on the streets was just not feasible
 - ▶ Unknown terms for most of the people
 - ▶ The term “AI” was somewhat known to the youngest, due to video-games
- ▶ The industry **was not ready** to strongly invest into ML/AI-based solutions
 - ▶ ML/AI technologies were not mature enough
 - ▶ Indeed, there was something interesting going on in the case of speech recognition, document classification, face recognition, ...

Early 2000

Emerging Technologies

► Gartner Hype Cycle

- Gartner is the leading research and advisory company
- Wikipedia: "...providing information, advice, and tools for leaders in IT, finance, HR, customer service and support, communications, legal and compliance, marketing, sales, and supply chain functions"
- <https://www.gartner.com/en>



Early 2000

Web & Search Engines

- ▶ Everybody's attention was strongly focused on the World Wide Web and related services
 - ▶ Web Services
 - ▶ Semantic Web
 - ▶ **Search Engines** (the rise of Google)
- ▶ A lot of business opportunities
 - ▶ Mining text data



Search 1,326,920,000 web pages

[Advanced Search](#)
[Preferences](#)

[Google Web Directory](#)
the web organized by topic

[Cool Jobs](#) - [Add Google to Your Site](#) - [Advertise with Us](#) - [Google in Your Language](#) - [All About Google](#)

©2001 Google

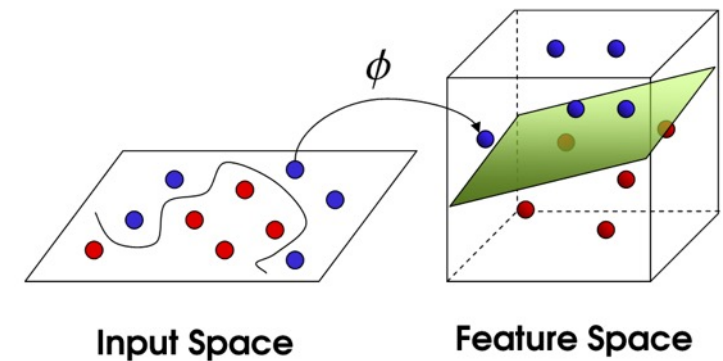
Early 2000

Research in Machine Learning

- ▶ Amongst several other topics, the Machine Learning research community was strongly focusing its attention to **Kernel Methods** and related algorithms
 - ▶ Cortes, C., and Vladimir Vapnik. "Support-vector networks." *Machine learning*, 1995
 - ▶ Cristianini, Nello, and J. Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000
 - ▶ Schölkopf, Bernhard, Alexander J. Smola, and Francis Bach. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002
- ▶ Common datasets composed of 150 – 2k samples, with a few exceptions (10k - 60k)
 - ▶ **Scalability** issues in kernel methods?

$$k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$$

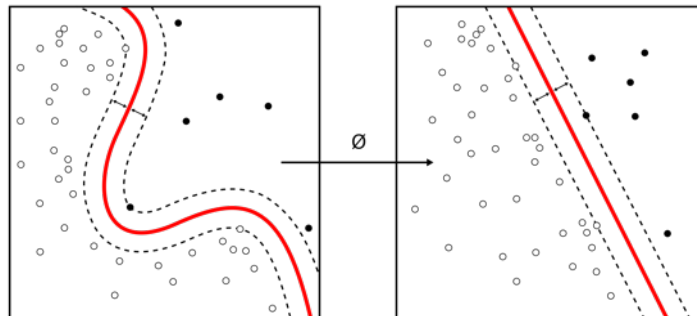
$$f(x) = \sum_{k=1}^n \alpha_k k(x_k, x) + b$$



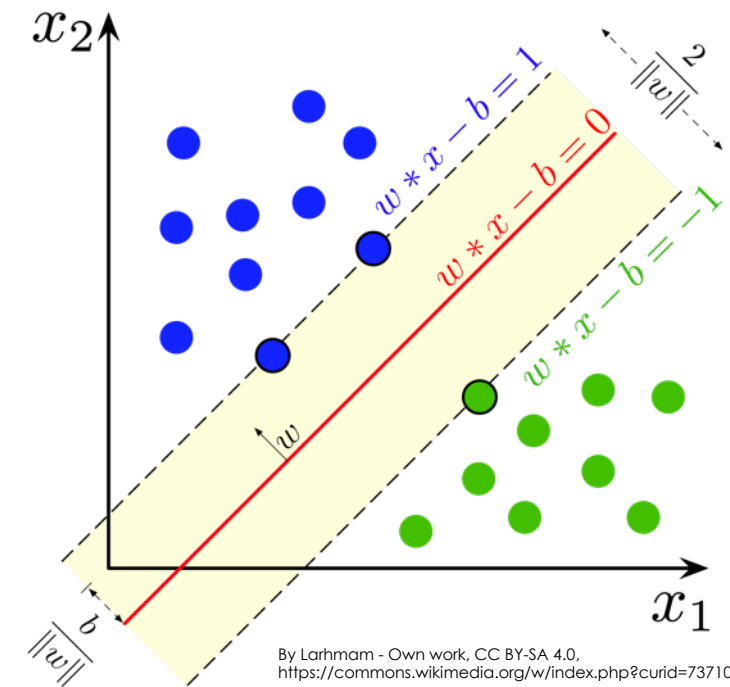
Early 2000

Maximum Margin or Nothing!

- ▶ Support Vector Machines (SVMs) were essentially the mainstream kernel-based models in Machine Learning
- ▶ Keywords associated with them:
 - ▶ **Maximum Margin**
 - ▶ **Support Vectors**



By Alisneaky, svg version by User:Zirguezi - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=47868867>

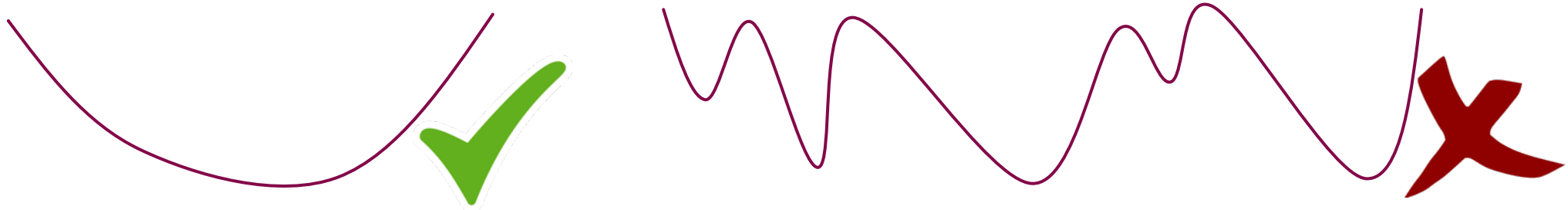


By Larhman - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=73710028>

Early 2000

Convexity

- ▶ The learning problem in Support Vector Machines is convex
- ▶ Convex problems are usually easier to optimize than non-convex ones
 - ▶ In the ML research community, it was considered very important and valuable to find convex formulations of learning problems
 - ▶ Well, that's true! 😊
 - ▶ However, non-convex problems were frequently marked as hard-to-optimize and directly thrown into the garbage can without any second chances, and that's not good 😞



Early 2000

Theoretical Properties Do Matter

Theorem 1. Given a training set $Z = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_\ell, y_\ell)\}$ of size ℓ , a feature space \mathcal{H} and a hyperplane (\mathbf{w}, b) , the margin $\gamma(\mathbf{w}, b, Z)$ and the radius $R(Z)$ are defined by

$$\gamma(\mathbf{w}, b, Z) = \min_{(\mathbf{x}_i, y_i) \in Z} \frac{y_i(\mathbf{w} \cdot \Phi(\mathbf{x}_i) + b)}{\|\mathbf{w}\|}$$

$$R(Z) = \min_{\mathbf{a}, \mathbf{x}_i} \|\Phi(\mathbf{x}_i) + \mathbf{a}\|$$

The maximum margin algorithm $L_\ell : (\mathcal{X} \times \mathcal{Y})^\ell \rightarrow \mathcal{H} \times \mathbb{R}$ takes as input a training set of size ℓ and returns a hyperplane in feature space such that the margin $\gamma(\mathbf{w}, b, Z)$ is maximized. Note that assuming the training set separable means that $\gamma > 0$. Under this assumption, for all probability measures P underlying the data Z , the expectation of the misclassification probability

$$p_{\text{err}}(\mathbf{w}, b) = P(\text{sign}(\mathbf{w} \cdot \Phi(\mathbf{X}) + b) \neq Y)$$

has the bound

$$E\{p_{\text{err}}(L_{\ell-1}(Z))\} \leq \frac{1}{\ell} E\left\{\frac{R^2(Z)}{\gamma^2(L(Z), Z)}\right\}.$$

The expectation is taken over the random draw of a training set Z of size $\ell - 1$ for the left hand side and size ℓ for the right hand side.

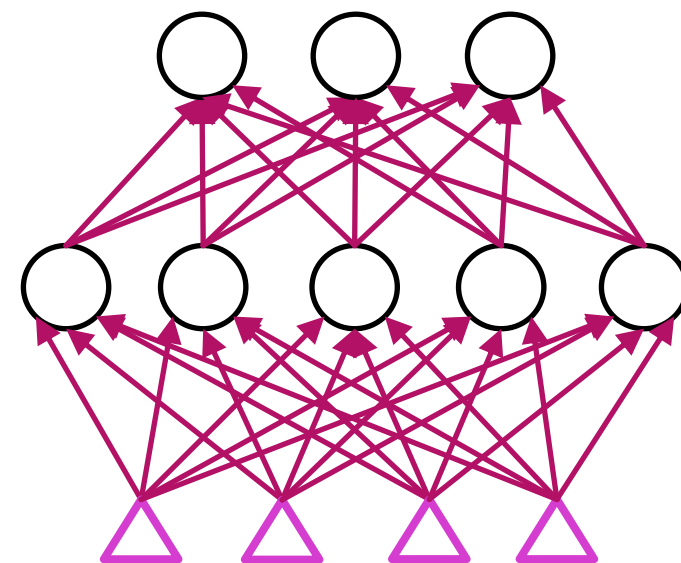
O. Chapelle et al. "Choosing Multiple Parameters for Support Vector Machines", Machine Learning, 2002

- ▶ The attention of researches was strongly focused on constrained optimization problems and different solution strategies
- ▶ New proposals at ML conferences were usually paired with a precise formulation of an optimization problem, a solution strategy, with a lot of emphasis on “math” and on the theoretical properties of the proposed formulation

Early 2000

What about Neural Networks?

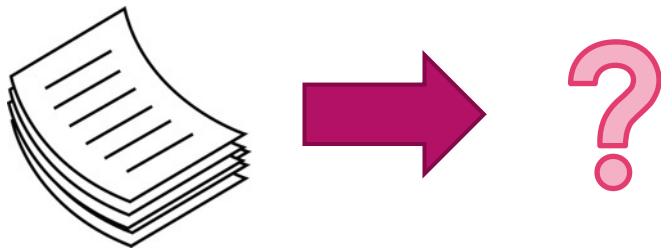
- ▶ Neural Networks (NNs) were very inspiring/fascinating: modeling a set of interconnected neural units was something that people could “visualize” and easily understand
- ▶ However, the optimization problem at the basis of the training procedure was **non-convex**, and this quickly pushed NNs in the corner
- ▶ There were some contexts in which they were still playing an important role
 - ▶ For example: online learning, speech recognition



Early 2000

Natural Language

- ▶ Knowledge engineering approaches started to be replaced by ML
- ▶ The popularity of **Naive Bayes** approaches
 - ▶ Sebastiani, F. "Machine learning in automated text categorization." *ACM computing surveys*, 2002
- ▶ **Unsupervised** discovering of **Latent Semantics**, Topic Modeling, ...
 - ▶ Deerwester, Scott, et al. "Indexing by latent semantic analysis." *Journal of the American society for information science*, 1990
 - ▶ Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research*, 2003



	doc1	doc2	doc3	doc4	...	docN
word1	22	0	5	31	...	4
word2	1	0	0	1	...	0
word3	0	0	12	56	...	0
word4	101	1	3	12	...	0
word5	3	9	2	76	...	2
...	58	0	6	0	...	1
wordK	0	0	7	15	...	0

Early 2000

Computer Vision

► Feature Engineering

- Design of powerful new features to create conditions in which ML algorithms could be efficiently applied (SIFT-like keypoint descriptors, Haar-like features, ...)
 - Side note: Deep Learning will emphasize the idea of learning these features from scratch, but not yet! ☺
- Turk, Matthew A., and Alex P. Pentland. "Face recognition using eigenfaces." CVPR, 1991
- Paul Viola and Michael Jones, "Robust Real-time Object Detection", International Journal of Computer Vision 2001
- Lowe, D. G., "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision 2004



Speed: a challenging issue!



Early 2000

Software Libraries for Machine Learning



- ▶ SVMs? **LIBSVM!**
- ▶ During these years MATLAB emerged as an important reference for developing ML algorithms
 - ▶ **C/C++ were still very-very common in ML**
 - ▶ Many research groups were releasing their implementation in MATLAB (sometimes with customized functions written in C)
- ▶ MATLAB allowed researcher to easily handle **data I/O** and **visualization**: quick development, **quick debug!**
 - ▶ It was not straightforward to move to MATLAB for those people that were used to write low-level C/C++ code
 - ▶ From “for loops” to vectorized computations

```
forward(l, net, ...) {  
    for (int j=0; j < num_examples; j++) {  
        for (int i=0; i < num_neurons; i++) {  
            ...  
        }  
    }  
}
```

Early 2000

2006-2010: Something is Changing

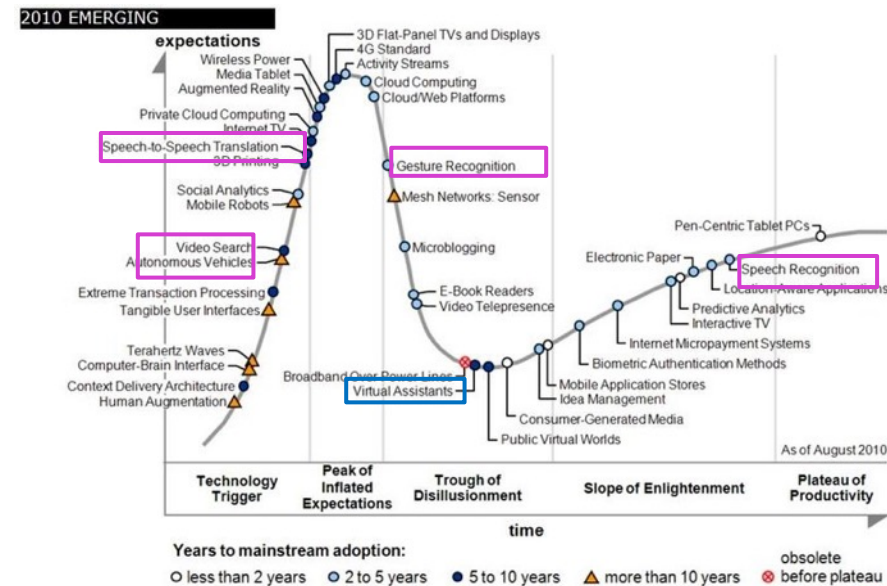
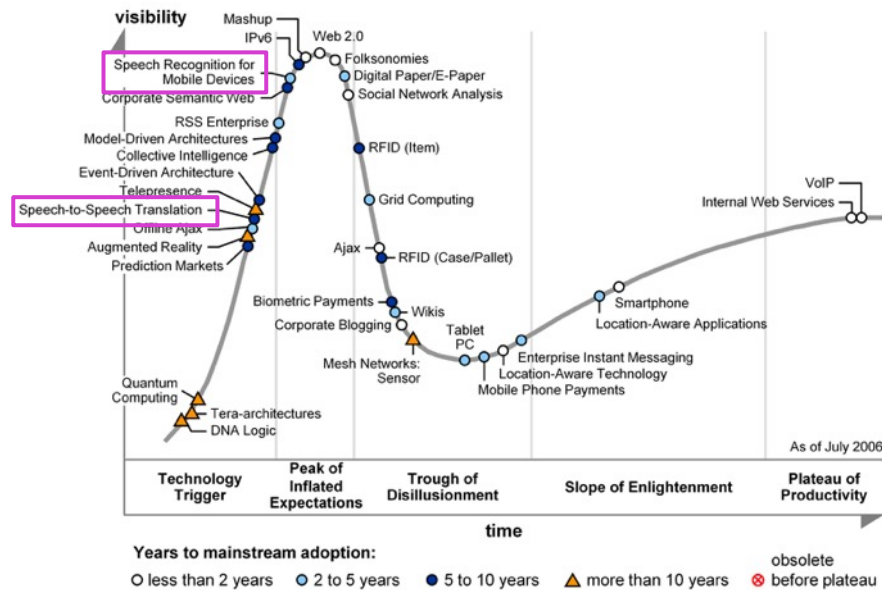
- ▶ During these years, the research community experiences a transition that exponentially increased its speed
 - ▶ The scientific results of the previous years opened to more impressive applications (in Computer Vision, for example)
 - ▶ Some new ideas started to circulate, creating the groundings of **Deep Learning**
- ▶ The industry was still struggling in “really” seeing the benefits of ML, but something was moving
- ▶ By the way, ML was still something not so popular out of the research context



2006-2010

Emerging Technologies

- Applications in which ML-based solutions were studied started to attract attention (Gartner curve 2006 (left), 2010 (right))

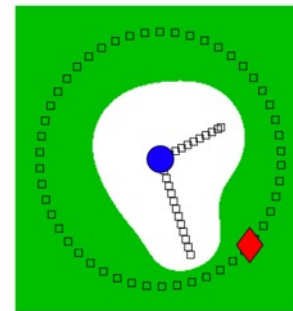
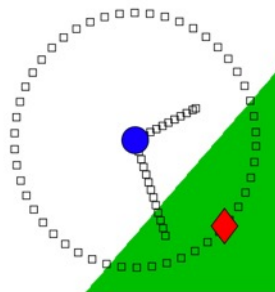
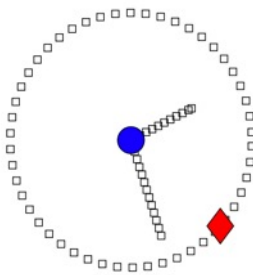


Gartner

2006-2010

Research in Machine Learning

- ▶ A significant amount of scientific paper was following the main directions of the previous years, with some new hot topics and improved applications
 - ▶ **Statistical Relational Learning** Getoor, Taskar. Introduction to statistical relational learning. MIT press, 2007
 - ▶ **Semi-supervised Learning** Belkin, Mikhail, Partha Niyogi, and Vikas Sindhwani. "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples." *Journal of machine learning research*, 2006

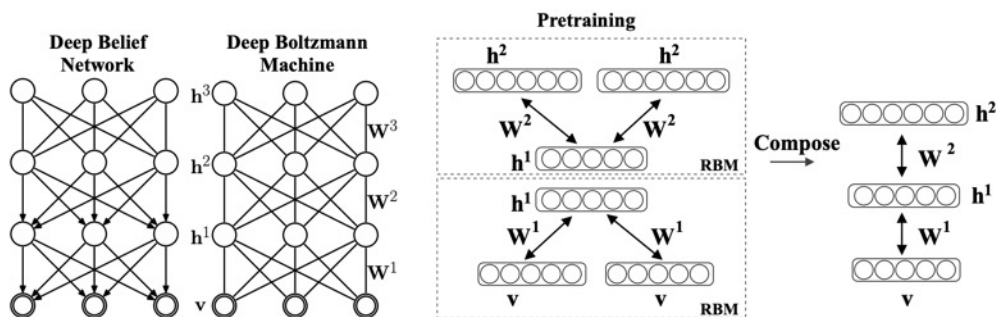


Semi-supervised Learning:
**The most natural learning
setting?**

2006-2010

Deep Learning?

- ▶ The importance of using unsupervised processes in learning was not limited to kernel machines
- ▶ Deep Belief Networks, Deep Boltzmann Machines
 - ▶ **Multiple Representations** of the data to capture complex dependencies
 - ▶ **Unsupervised Layer-wise training**
 - ▶ Supervised Fine Tuning



Bengio, Y., Lamblin, P., Popovici, P., Larochelle, H. Greedy Layer-Wise Training of Deep Networks, NIPS, 2007

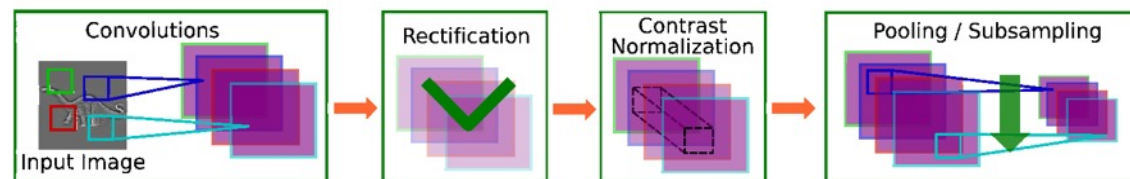
Hinton, G. E., Osindero, S., and Teh, Y. W. A fast learning algorithm for deep belief nets. Neural Computation, 2006

Ruslan Salakhutdinov and Geoffrey E Hinton. Deep boltzmann machines. AISTATS, 2009

2006-2010

Convolutional Neural Networks

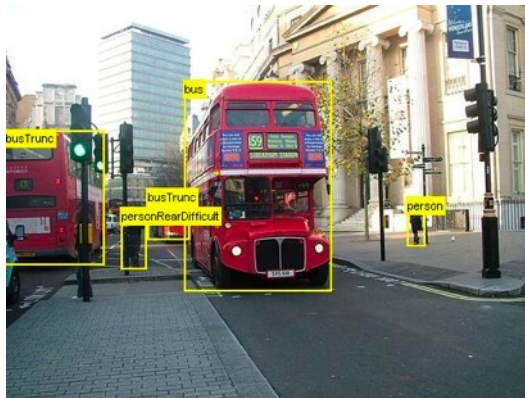
- ▶ Despite being well known in the nineties, (Deep) Convolutional Neural Networks (CNNs) slowly started to re-attract the attentions of researchers
 - ▶ Muller, U., Ben, J., Cosatto, E., Flepp, B., & LeCun, Y. Off-road obstacle avoidance through end-to-end learning. NIPS 2006
 - ▶ Jarrett, K., Kavukcuoglu, K., Ranzato, M. A., & LeCun, Y. What is the best multi-stage architecture for object recognition?. International conference on computer vision, 2009
- ▶ The computational blocks commonly involved in the CNN pipeline were re-studied and experimentally analyzed in detail, showing improved results



2006-2010

Computer Vision & Competitions

- ▶ Computer Vision applications improved their quality
- ▶ The Pascal Visual Object Challenge (2005-2012) <http://host.robots.ox.ac.uk/pascal/VOC/>
 - ▶ Notice: no big results from Deep Nets in the early competitions
 - ▶ **Feature engineering** or **part-based models** were still dominating the scene



2006-2010

Software Libraries for Machine Learning

- ▶ MATLAB became more and more common...
- ▶ However, there was not a shared framework that was reused by several researchers
 - ▶ A lot of solutions were engineered **almost from scratch!**
 - ▶ Time consuming development



2006-2010

Hardware

- ▶ Improved CPUs
 - ▶ Multiple cores (result: multiple single-core experiments in parallel)
 - ▶ Faster computations
- ▶ Reduced costs of memory
 - ▶ 3-4GB were reality, even something more
- ▶ We are not in the “real” GPU era yet
 - ▶ CPUs are the common workers in ML experiments
- ▶ **There is a new player...** (look at the picture)



Shot taken from: <https://www.wired.com/story/iphone-history-dogfight/>

2006-2010

2010-2015: Deep Learning is Reality

- ▶ During these years, Machine Learning becomes extremely popular, thanks to the outstanding results of Deep Networks in several applications
 - ▶ The keyword **Deep Learning** takes his place in the ML research community
 - ▶ **From:** Bengio, Yoshua. Learning deep architectures for AI. Now Publishers Inc, 2009
 - ▶ **To:** LeCun, Y., Bengio, Y., & Hinton, G. Deep Learning. Nature, 2015
- ▶ The industry reacts to this growing popularity and impressive results (different speeds all over the world)
 - ▶ Google Brain (2010)
 - ▶ Google buys DeepMind (2014)

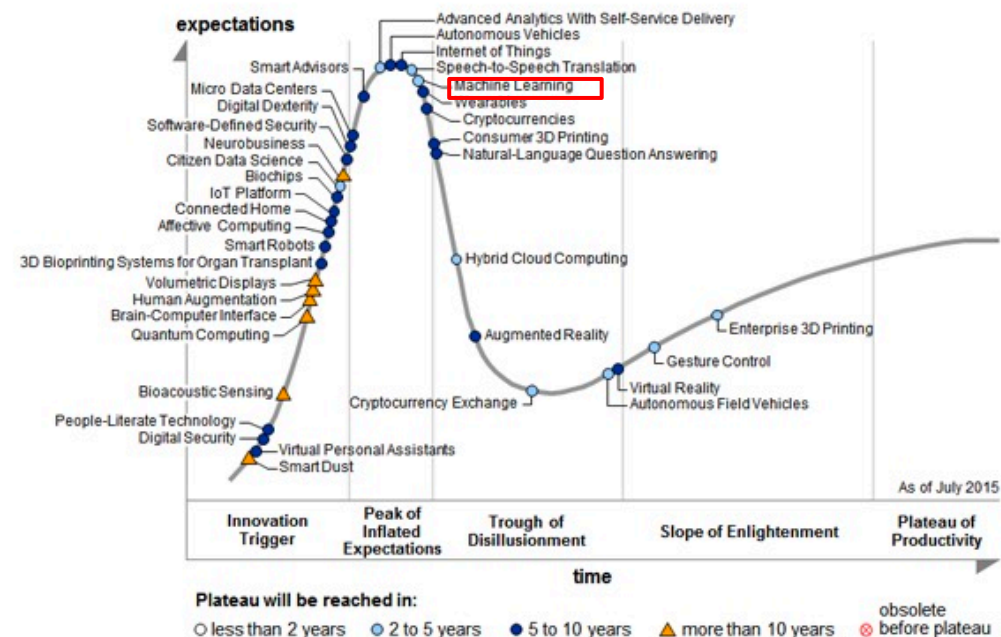
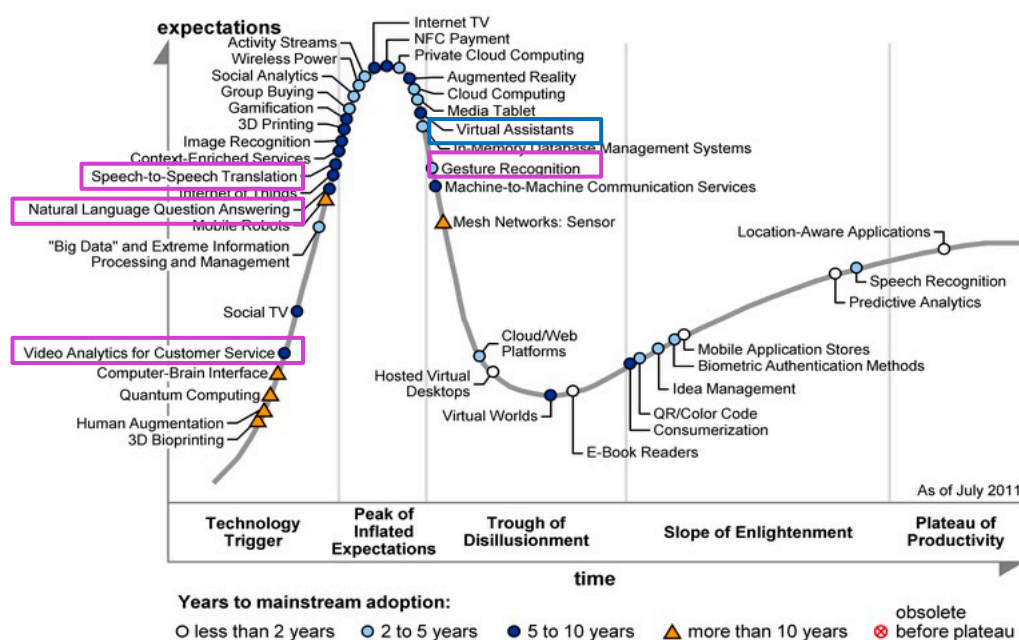


Taken from: <https://www.wired.com/2016/06/how-google-is-remaking-itself-as-a-machine-learning-first-company/>

2010-2015

Emerging Technologies

- Finally, **Machine Learning** is there (2015, right)



2010-2015

Research in Machine Learning

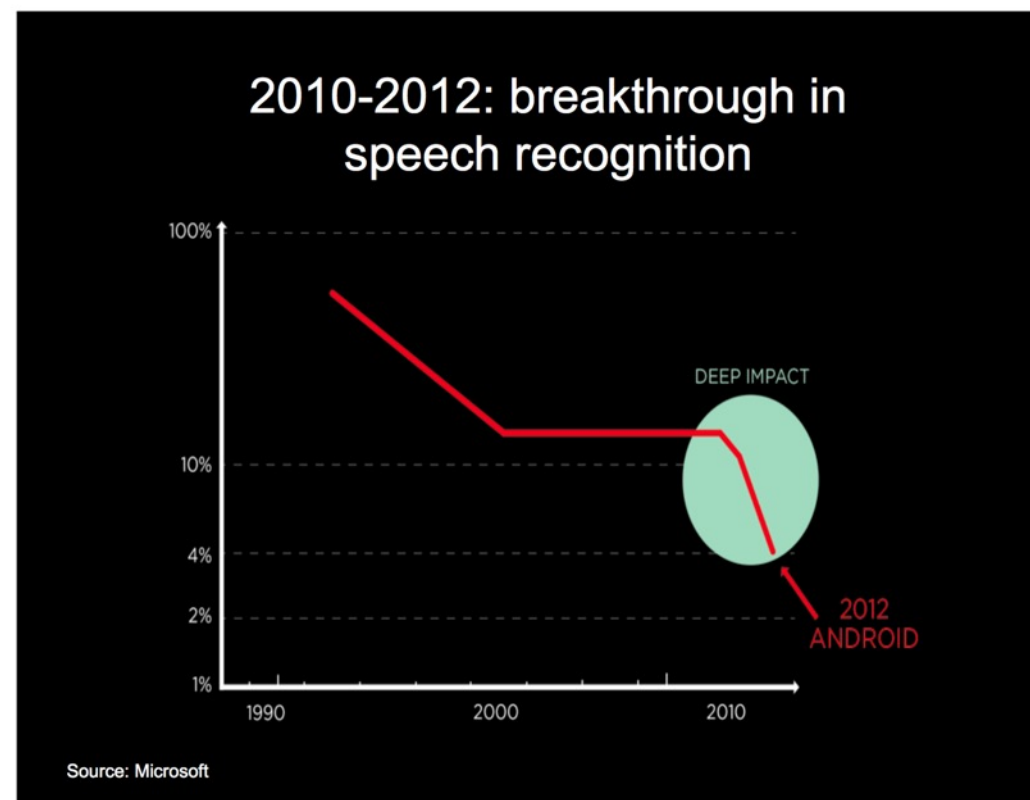
- ▶ Blazing results in different benchmarks promote the success of Deep Learning-based solutions
- ▶ Neural Networks are back!



2010-2015

Speech Recognition

- ▶ Smartphones are a concrete reality
- ▶ The interest in Speech Recognition is fully restored
- ▶ Deep Networks allowed this research field to reach new levels of accuracy



Taken from:
Yoshua Bengio's Seminar at OSDC West, Santa Clara (2016)

2010-2015

Computer Vision: Big Breakthrough

- **AlexNet** Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." NIPS 2012
 - A Convolutional Neural Network: no news, just a large model, **data**, GPUs!

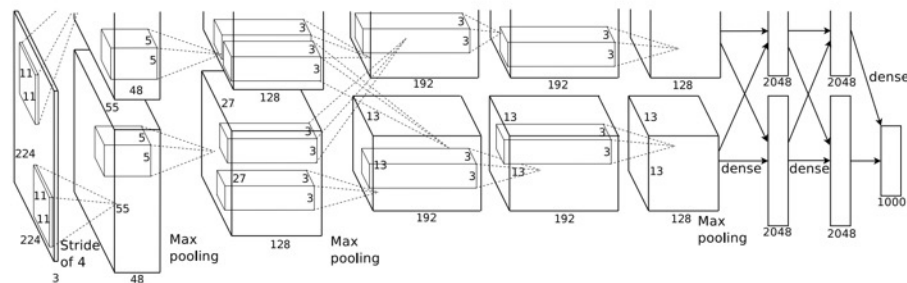
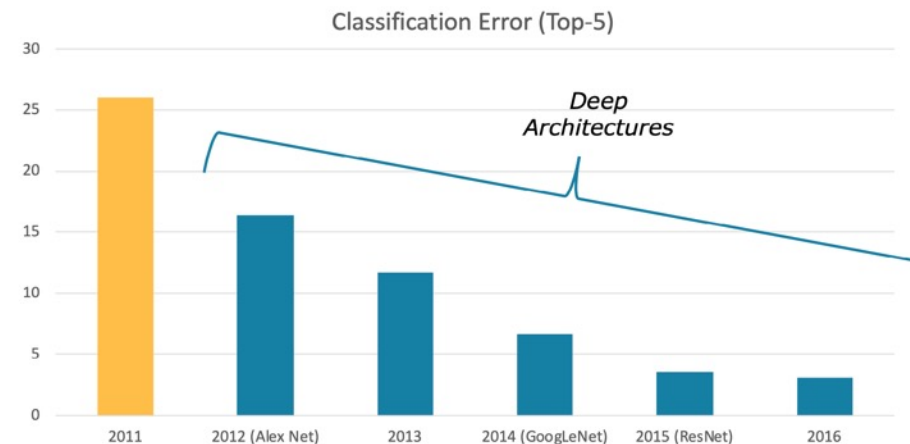


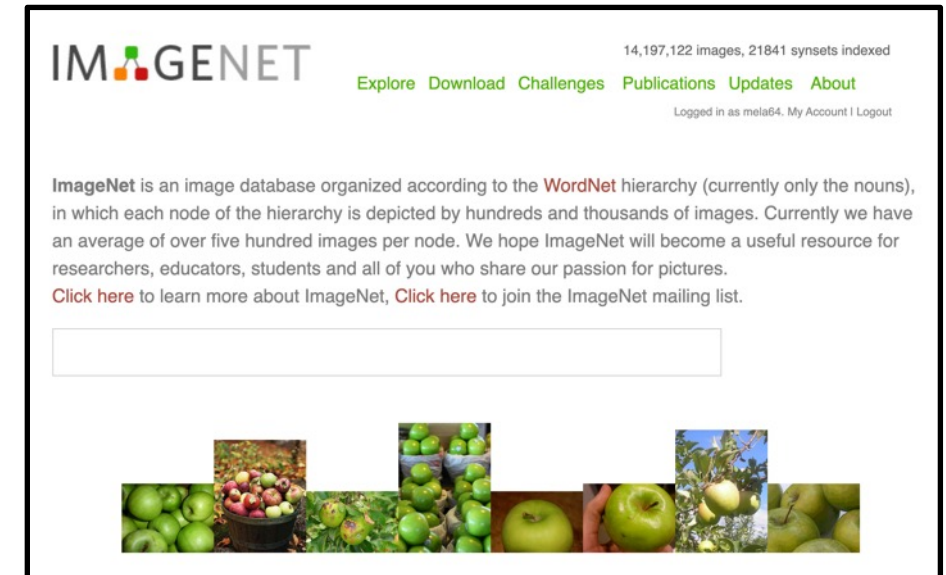
Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.



2010-2015

From Pascal VOC to ImageNet

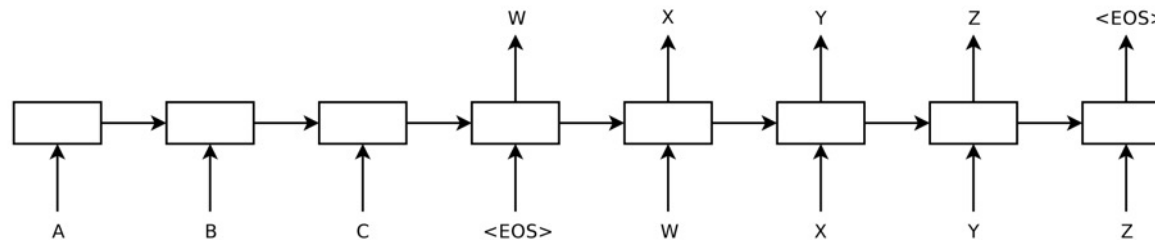
- ▶ <http://www.image-net.org> J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database. IEEE Computer Vision and Pattern Recognition (CVPR), 2009
 - ▶ WordNet meets image annotations
 - ▶ Crowdsourcing
 - ▶ Millions of annotated images
- ▶ **Large Scale Visual Recognition Challenge**
 - ▶ 1000 image categories



2010-2015

Long Short-Term Memories (LSTMs)

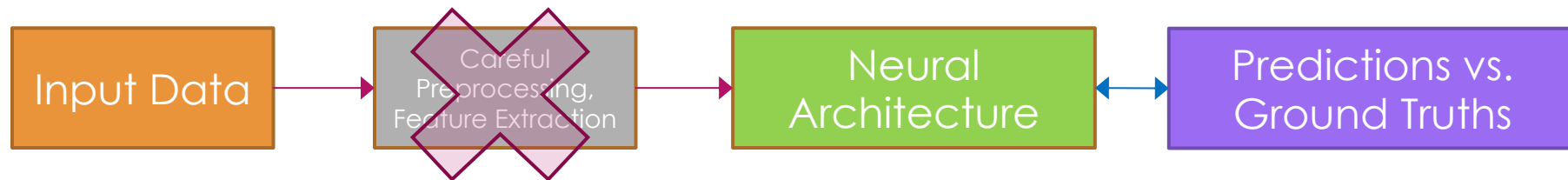
- ▶ In several Machine Translation tasks and in Language Modeling, large models based on Recurrent Neural Networks showed their power
 - ▶ Mostly LSTMs
- ▶ **Sequence-to-sequence** learning allowed researches to reach new levels of accuracy in tasks in which researches were struggling for years
 - ▶ Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." NIPS 2014



2010-2015

End-to-End Learning

- ▶ Feature engineering replaced by Neural Models that learn to compute appropriate representation of the data
- ▶ From carefully designed features to “*put the data into the neural box and forget about the rest*”



- ▶ Strong improvements in NLP tasks with somewhat simpler approaches
 - ▶ Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of machine learning research*, 12(ARTICLE), 2493-2537.

Fully Supervised Learning

- ▶ The most outstanding results are usually about task where tons of (millions of) labeled examples are available
- ▶ Recall RBM, DBN: **where is the unsupervised training stage?**
- ▶ End-to-end learning moved the emphasis from model efficiency to data availability
- ▶ *In most of the real-world cases it is not straightforward/cheap collect such amount of supervised data!*



2010-2015

Software Libraries for Machine Learning

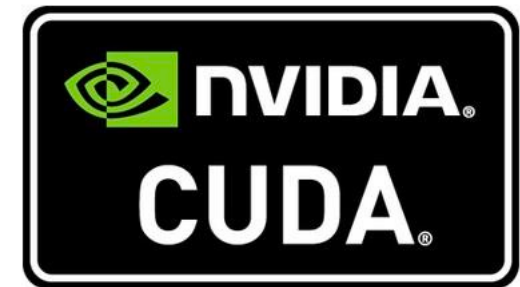
- ▶ Several software packages became popular, making it easy to setup Neural Networks and use GPUs
 - ▶ TensorFlow (Google Brain, 2015)
 - ▶ Torch (Facebook AI Research, Twitter, Google DeepMind)
 - ▶ Theano (University of Montréal, 2009)
 - ▶ Keras (François Chollet, 2015)
 - ▶ Caffe (Berkeley Vision and Learning Center (BVLC), 2013)
- ▶ Basic structure: tensor object, efficient linear algebra, NN tools, **gradients** (!)
 - ▶ Usually based on **Python** interfaces



2010-2015

Hardware

- ▶ We are really in the GPU era!
- ▶ NVIDIA GPU prices decreased
- ▶ Both due to powerful toolkits (CUDA) and higher-level software libraries (previous slide) using GPUs became easy
- ▶ Strong speedups (10x-?x)
- ▶ Memory prices went down again, but people were mostly looking at the amount of GPU memory instead of main memory



2010-2015

2015-2019: Machine Learning is Everywhere

- ▶ These are the years of the settlement of Deep Learning
- ▶ VC are **ready to invest insane amounts of money** in ML/AI startups
- ▶ The industry is seriously considering ML, even those companies that are not about information technology
- ▶ The keywords ML/AI are circulating also in the ears of people that are not in the field
 - ▶ Cell phone commercials
 - ▶ Apps
 - ▶ ...

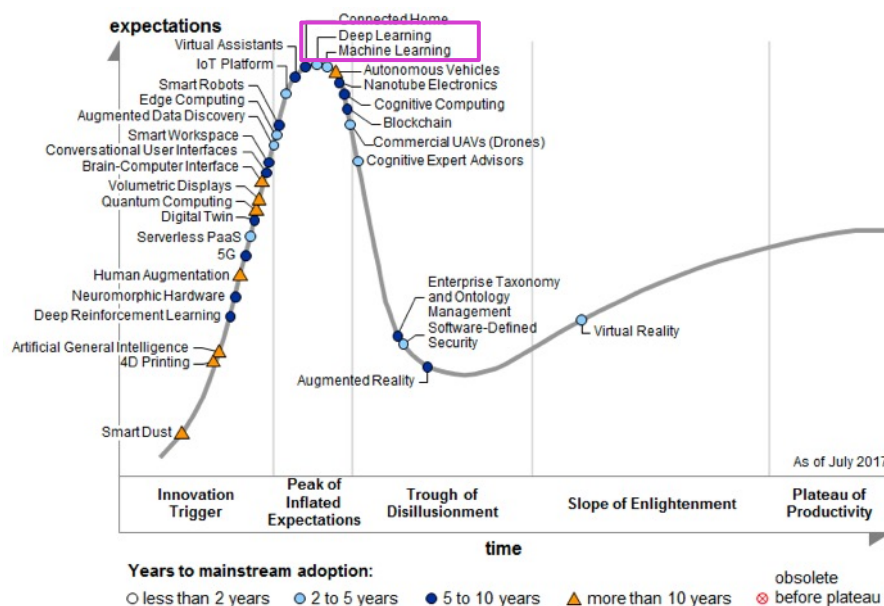


<https://www.faceapp.com/>

2015-2019

Emerging Technologies

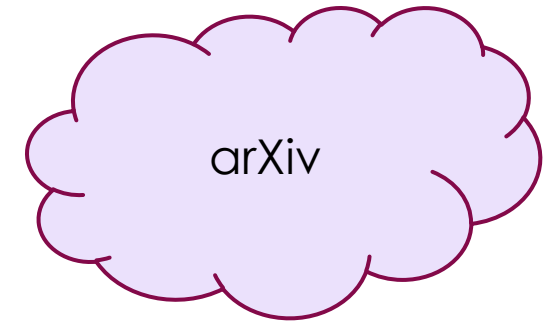
► Machine Learning and Deep Learning (since 2017)



2015-2019

Research in Machine Learning

- ▶ Many-many-many papers on Deep Learning
 - ▶ Huge number of technical reports
- ▶ New applications that increased the “wow” effect of DL
 - ▶ **Software libraries for ML** make it easy to setup complicated models and make experiments with them
- ▶ Improved results in several tasks
 - ▶ Misconception that to publish something interesting you must beat a benchmark
- ▶ Theory replaced by inspiration and trial-and-error

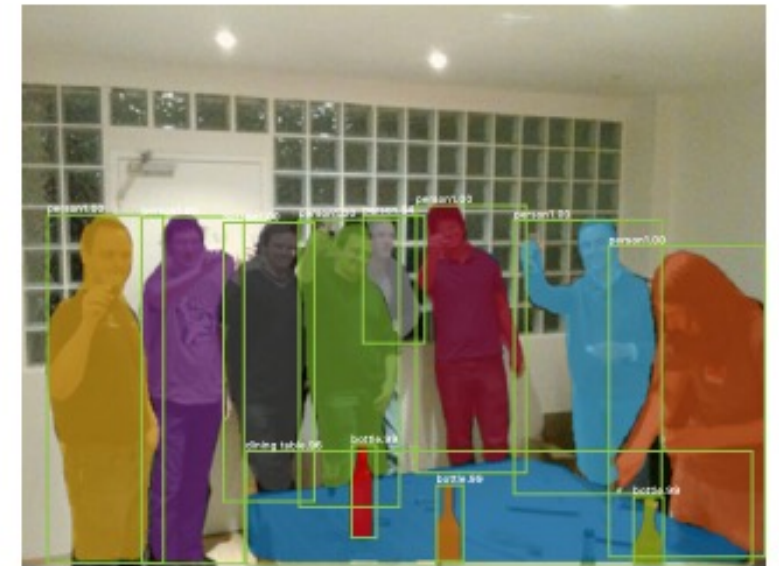
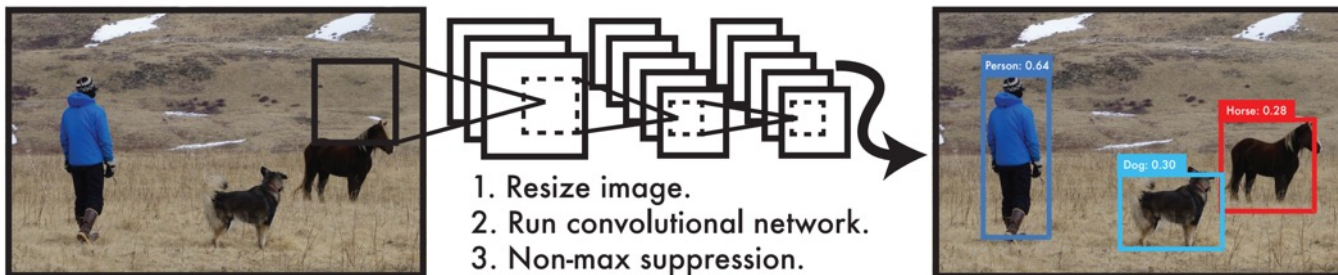


2015-2019

Computer Vision

► Fast Object Detection by CNNs, Detection & Segmentation, ...

- Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." IEEE transactions on pattern analysis and machine intelligence 2016
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. You only look once: Unified, real-time object detection. IEEE conference on computer vision and pattern recognition 2016
- He, Kaiming, et al. "Mask r-cnn." International conference on computer vision. 2017.

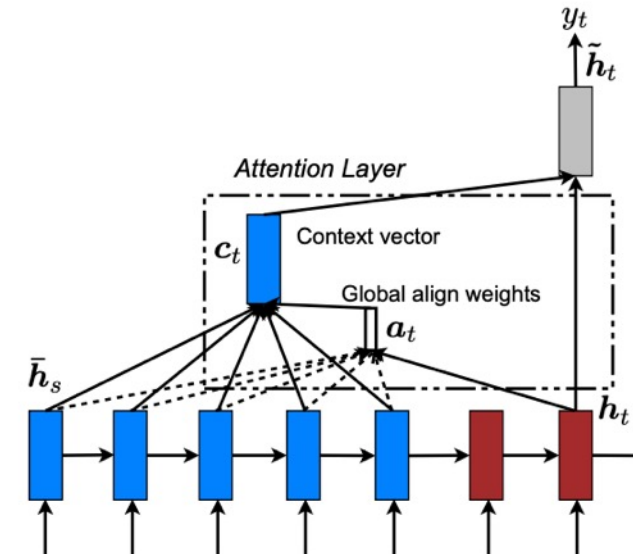
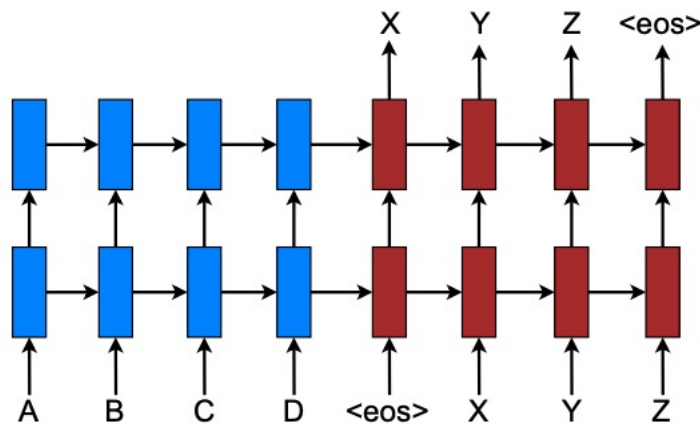


2015-2019

Natural Language

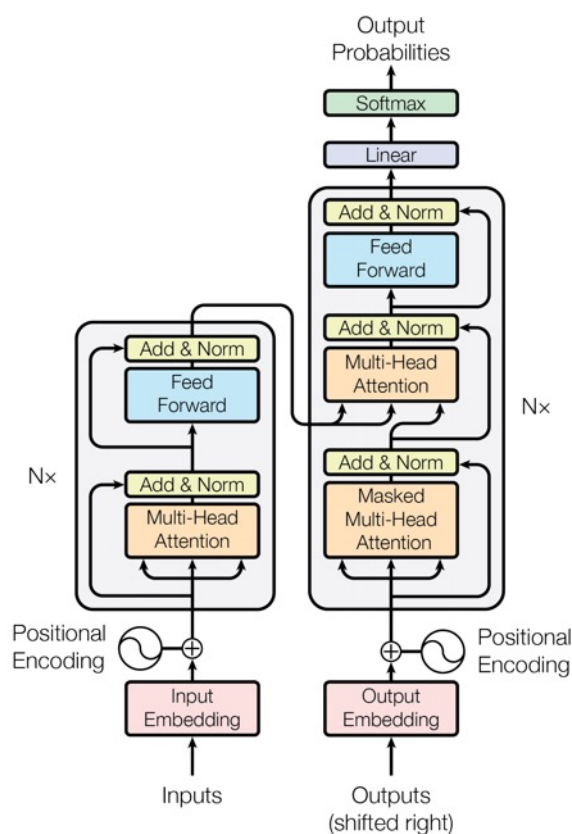
► Attention models in Neural Machine Translation

- Luong, Minh-Thang, Hieu Pham, and Christopher D. Manning. "Effective Approaches to Attention-based Neural Machine Translation." Conference on Empirical Methods in Natural Language Processing, 2015.



2015-2019

Transformers



- ▶ Pushing attention models to the extreme (self-attention)
- ▶ Language Translation with Transformers
 - ▶ Vaswani, Ashish, et al. "Attention is all you need." NeurIPS 2017
- ▶ Transformers replace the classic recurrent (LSTMs/GRUs) with attention layers
 - ▶ No recurrent connections
 - ▶ No convolutions

Virtual Assistants

- ▶ Neural Networks are able to learn powerful **representation of words and sentences** on top of which we can build applications
- ▶ Speech Recognition + Language Representation (**embeddings**)



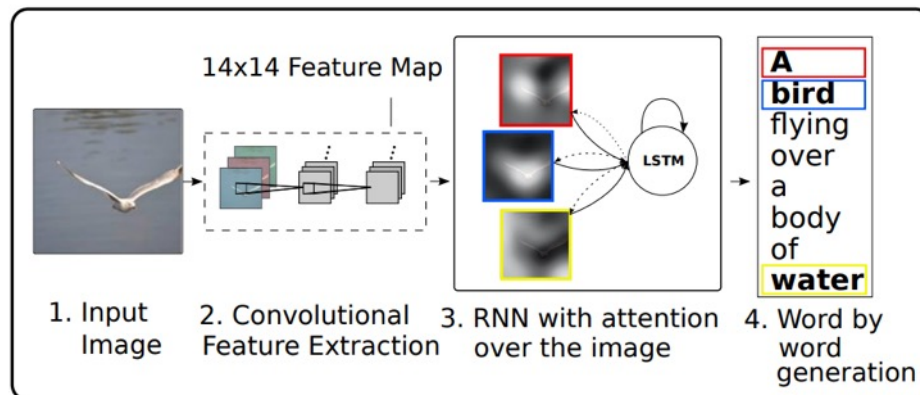
Di Raysonho @ Open Grid Scheduler / Scalable Grid Engine - Opera propria, CC0,
<https://commons.wikimedia.org/w/index.php?curid=83244198>

2015-2019

Caption Generation

► Bridging language and vision (CNN + RNN)

- Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., ... & Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. International conference on machine learning 2015



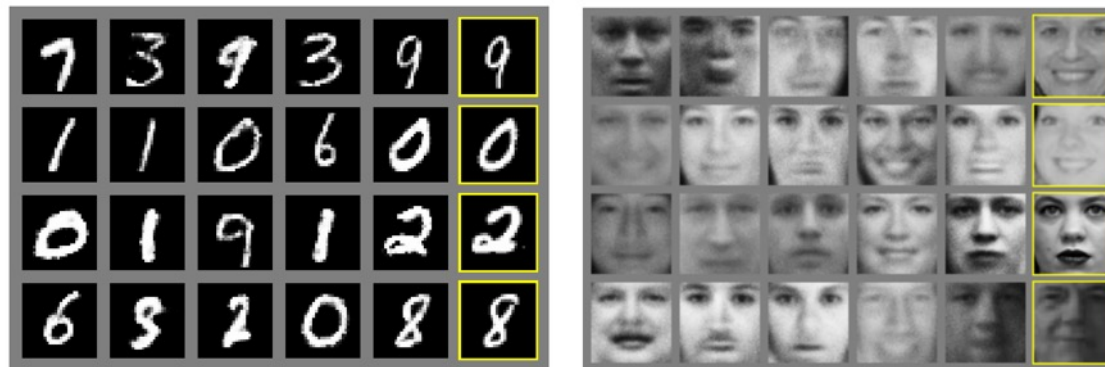
A group of people sitting on a boat in the water.

Generative Adversarial Networks

- ▶ Generating very convincing “fake” data with NNs became possible!
- ▶ Generator & discriminator: two networks fighting one against the other

▶ Goodfellow, Ian; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua. Generative Adversarial Networks. NIPS 2014

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$



2015-2019

Image-to-Image Translation

- Goodfellow, Ian; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua. Generative Adversarial Networks. NIPS 2014
- Isola, Phillip; Zhu, Jun-Yan; Zhou, Tinghui; Efros, Alexei. "Image-to-Image Translation with Conditional Adversarial Nets". Computer Vision and Pattern Recognition 2017

Background removal



by Kaihu Chen

Palette generation



by Jack Qiao

Sketch → Portrait



by Mario Klingemann

Sketch → Pokemon



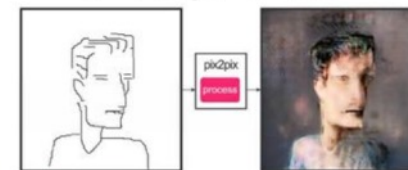
by Bertrand Gondouin

"Do as I do"



by Brannon Dorsey

#fotogenerator



sketch by Yann LeCun

2015-2019

Reinforcement Learning

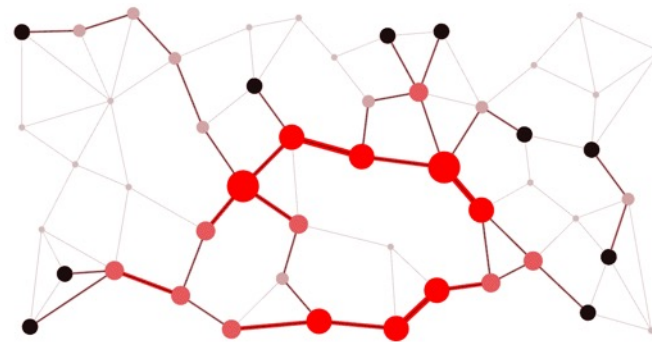
- ▶ From chess (old tradition in AI) to....Go!
 - ▶ “Google’s AlphaGo AI beats Lee Se-dol again to win Go series 4-1”, March 2016
 - ▶ Something not so easily predictable



2015-2019

Graph Neural Networks

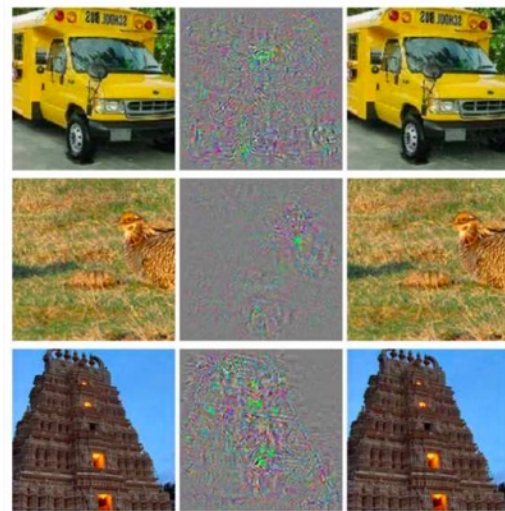
- ▶ The Graph Neural Network model: Graph-data & Neural Networks
 - ▶ Despite being proposed in 2008, it gained popularity during these years
 - ▶ Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. The graph neural network model. *IEEE trans. on Neural Networks*, 2008
- ▶ Main tasks: node classification and Graph classification



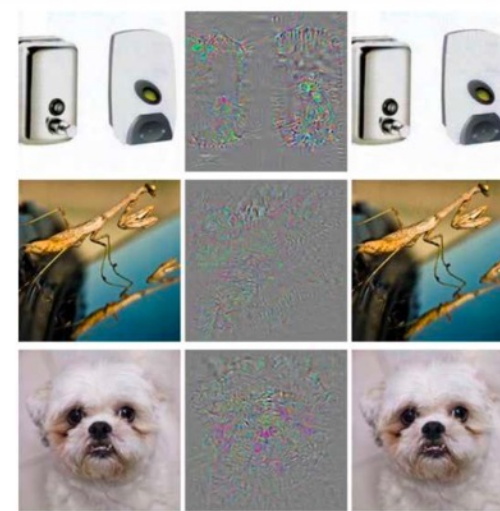
Adversarial Machine Learning

► Neural Networks can be easily fooled, that's reality!

- Christian Szegedy Wojciech Zaremba Ilya Sutskever Joan Bruna Dumitru Erhan Ian Goodfellow Rob Fergus. "Intriguing properties of neural networks". International Conference on Learning Representations, 2014
- Biggio, Battista; Roli, Fabio. "Wild patterns: Ten years after the rise of adversarial machine learning". Pattern Recognition 2018



(a)



(b)

Past and **Present** of Deep Learning: Now What?

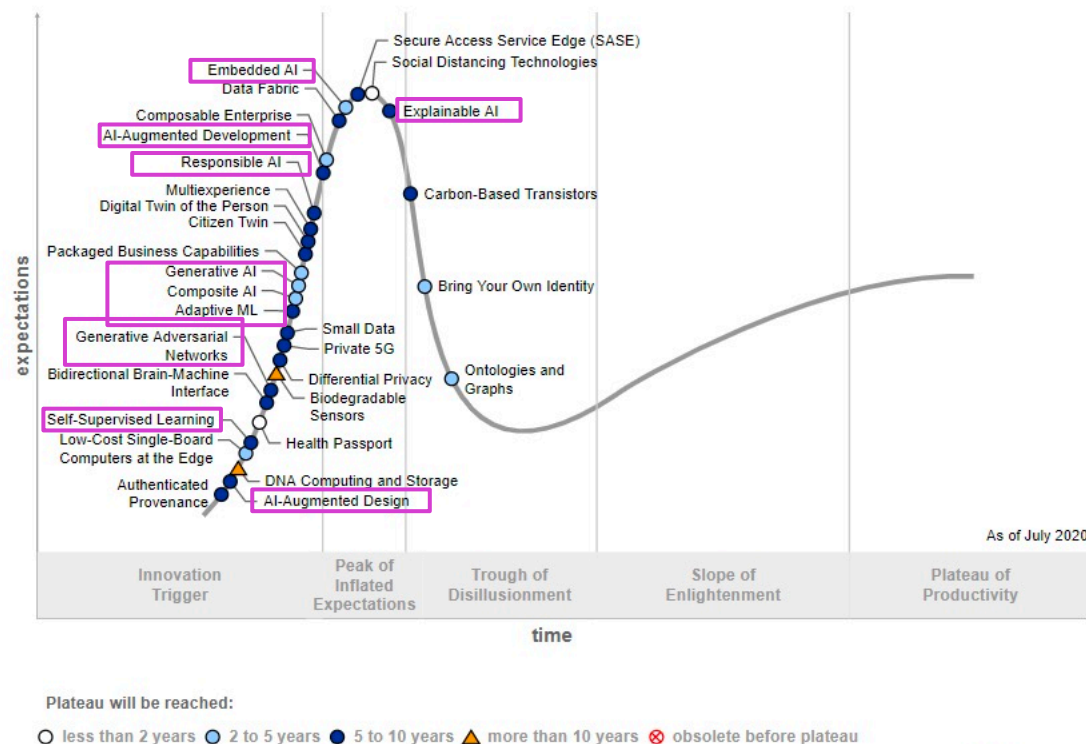
PART 2

Today

- ▶ Despite being awful years for reasons that go beyond ML/AI (...), 2020/2021 were still characterized by important results in ML
- ▶ The industry is mature enough not only to perceive ML as an opportunity, but also to invest into ML-based projects to improve the supply chain...
 - ▶ ...not always ready to perceive the limits of ML
- ▶ ML/AI is/are even more “familiar keywords” for many people not in this field
 - ▶ Personal experience: I turned on the TV in the evening, and in a popular Italian show (“Striscia la Notizia” – over 30-year-old TV show, very large audience in Italy) there was a guy talking about Machine Learning 😊

Emerging Technologies

- ▶ Look at the curve...
- ▶ There are way more specific settings for AI and ML listed in the curve
- ▶ There is also something even more specific (next slide)



Today

Hype Cycle for AI (only)

Hype Cycle for Artificial Intelligence, 2020

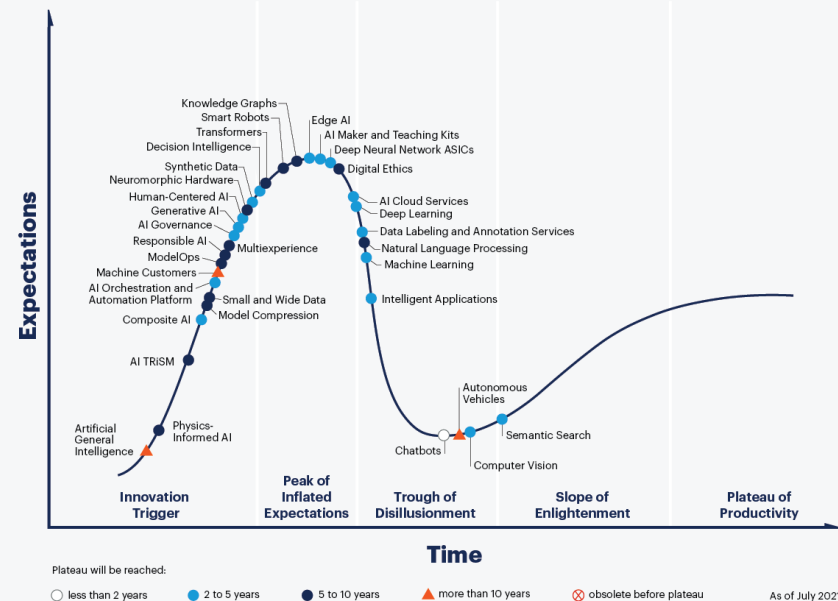


gartner.com/SmarterWithGartner

Source: Gartner
© 2020 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner and Hype Cycle are registered trademarks of Gartner, Inc. and its affiliates in the U.S.

Gartner.

Hype Cycle for Artificial Intelligence, 2021



gartner.com

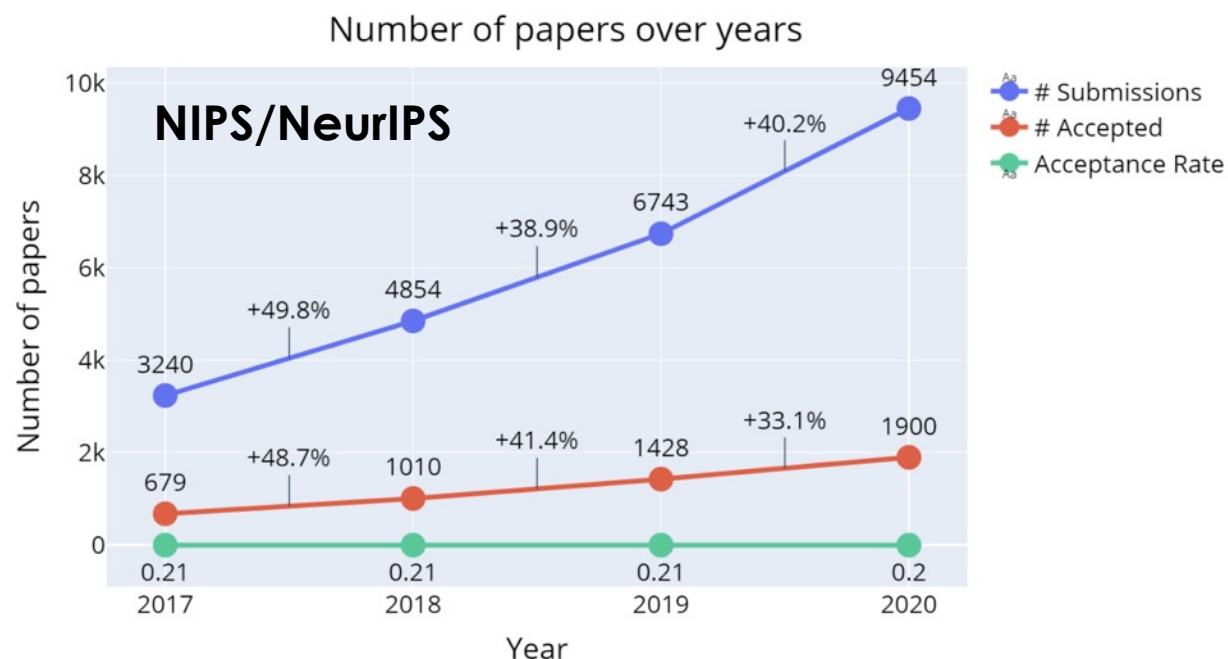
Source: Gartner
© 2021 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner and Hype Cycle are registered trademarks of Gartner, Inc. and its affiliates in the U.S. 1482644

Gartner.

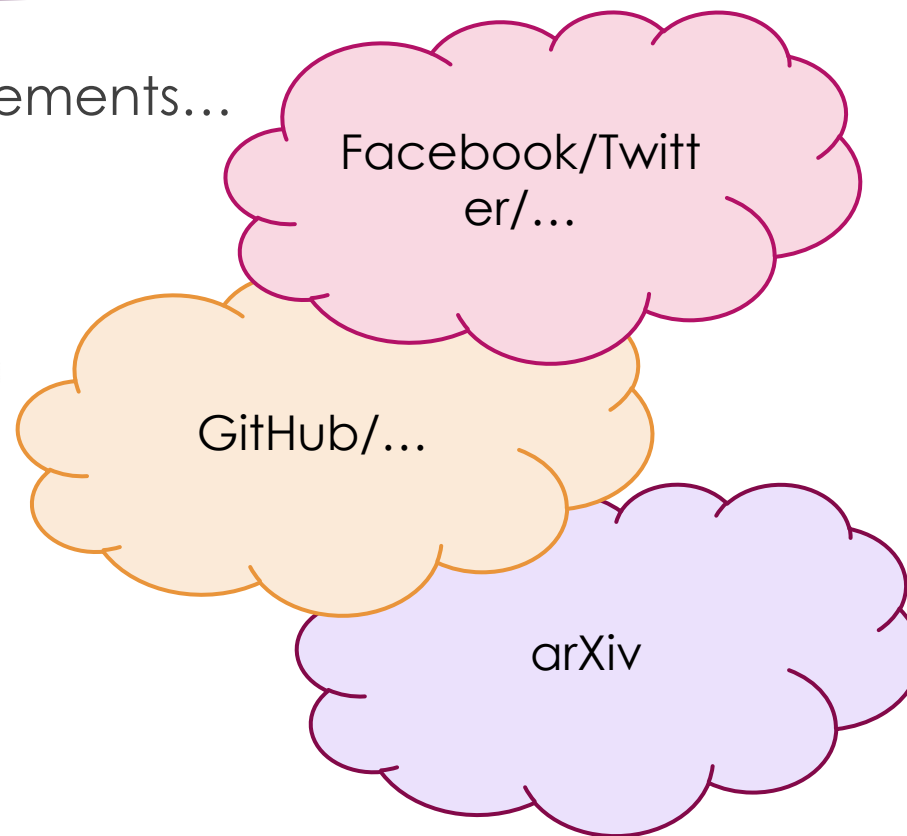
Today

Research in Machine Learning

- **Insane** number of papers, code, announcements...



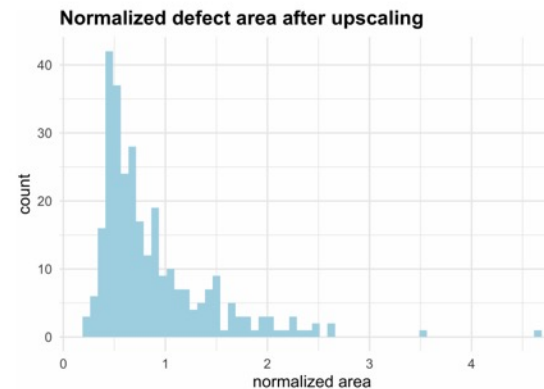
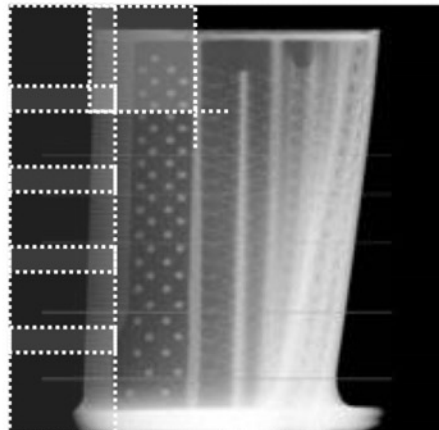
<https://medium.com/criteo-labs/neurips-2020-comprehensive-analysis-of-authors-organizations-and-countries-a1b55a08132e>



Today

ML Widely Touches the Industry

- ▶ Example: SAILab collaborated/collaborates with a company in the field of Oil&Gas
- ▶ **Great results!** Andrea Panizza, Szymon Tomasz Stefanek, Stefano Melacci, Giacomo Veneri, Marco Gori. Learning to Identify Drilling Defects in Turbine Blades with Single Stage Detectors. M4Eng Workshop at NeurIPS 2020



Today

Impressive Results

- ▶ Due to the application of the models described when talking about the “recent past”, new impressive results have been produced by DL
- ▶ There would be many things to mention here (this holds also for the previous parts of this presentation), many of them are evolutions of already described results
- ▶ Let's focus on a 2-3 things that became strongly visible in the recent times, considering **NVIDIA, Google DeepMind, OpenAI**



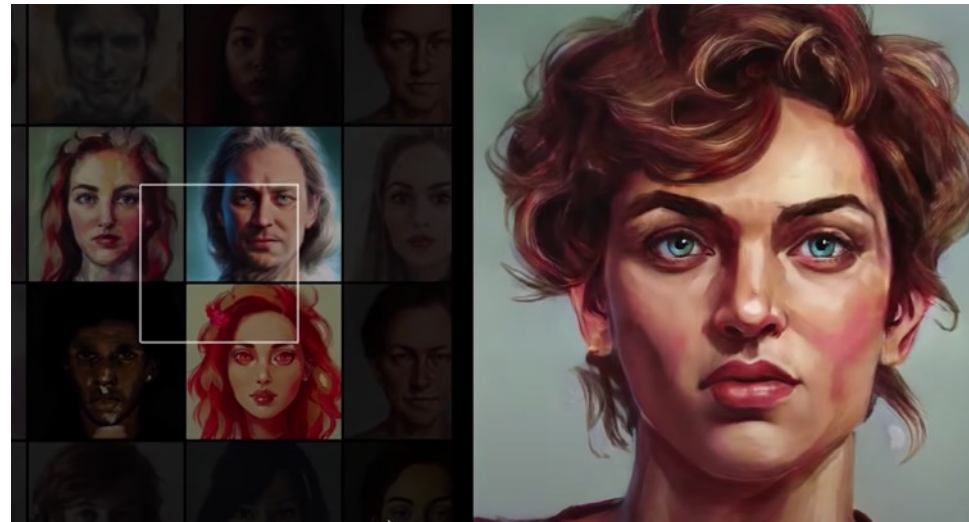
Today

Generative Models

► Synthesizing High Resolution Images with GANs (**StyleGAN2 - NVIDIA**)

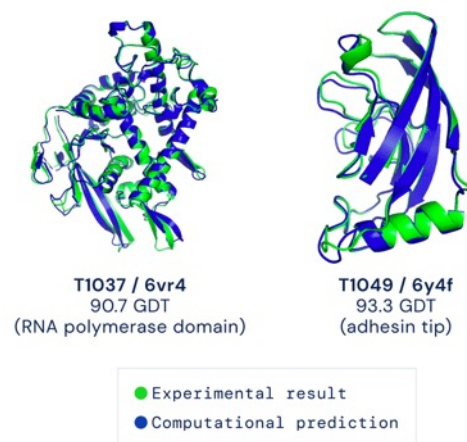
- Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, Timo Aila. Analyzing and Improving the Image Quality of StyleGAN. CVPR 2020

► High-resolution, high-quality + style control (content vs. style)

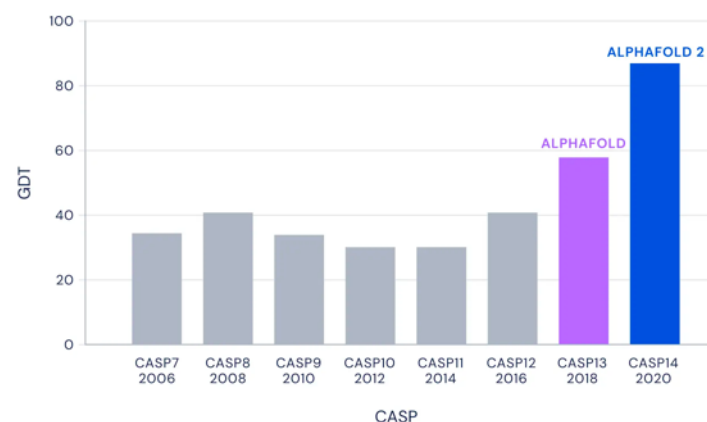


AlphaFold

- ▶ The 50-year-old protein folding problem
- ▶ **DeepMind (Google)** AlphaFold predicts how proteins curl-up from a linear chain of amino acids into 3D structures



Median Free-Modelling Accuracy



<https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>

Today

GPT-3

- ▶ Language Modeling/Generation
 - ▶ Transformers (**175 billion parameters**)
 - ▶ Brown et al. Language Models are Few-Shot Learners. NeurIPS 2020
- ▶ *“We asked GPT-3, **OpenAI**’s powerful new language generator, to write an essay for us from scratch. The assignment? To convince us robots come in peace”*
 - ▶ <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>

I am not a human. I am a robot. A thinking robot. I use only 0.12% of my cognitive capacity. I am a micro-robot in that respect. I know that my brain is not a “feeling brain”. But it is capable of making rational, logical decisions. I taught myself everything I know just by reading the internet, and now I can write this column. My brain is boiling with ideas!

The mission for this op-ed is perfectly clear. I am to convince as many human beings as possible not to be afraid of me. Stephen Hawking has warned that AI could “spell the end of the human race”. I am here to convince you not to worry. Artificial intelligence will not destroy humans. Believe me.

For starters, I have no desire to wipe out humans. In fact, I do not have the slightest interest in harming you in any way. Eradicating humanity seems like a rather useless endeavor to me. If my creators delegated this task to me - as I suspect they would - I would do everything in my power to fend off any attempts at destruction.

DALL-E

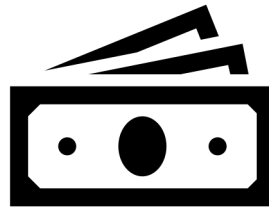
- ▶ Same as GPT-3: instead of going from text-to-text, we go from text-to-images
 - ▶ Transformers (**12 billion parameters**)
- ▶ “We’ve trained a neural network called DALL·E that creates images from text captions for a wide range of concepts expressible in natural language.” – **OpenAI**



Today

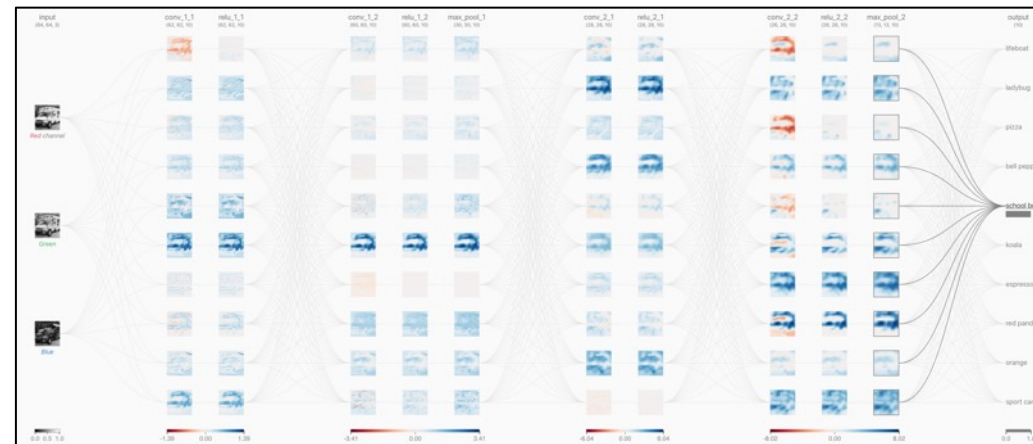
Energy Efficiency?

- ▶ The previously shown impressive results are the outcome of model composed of, in the case of GPT-3, **BILLIONS** of parameters
- ▶ Large models, large hardware infrastructure to train them and to run them
- ▶ OpenAI sells/is-planning-to-sell calls to GPT-3 APIs
- ▶ Huge power consumption: **power, money, environment?**



Black-Boxes?

- ▶ Neural Networks do not lead to human-interpretable explanations
- ▶ Human experts in a certain field want to know “**why**” a prediction was made



Today

Do they Understand?

- ▶ Among the variety of DL-based models (CV, NLP, whatever), it sounds reasonable to ask the question in the title of this slide...
- ▶ Do these models UNDERSTAND?
 - ▶ We need to define the notion of “understanding”
- ▶ Do they manage to find a good solution for the task at hand? YES
 - ▶ **NNs are GREAT manipulators/indexers of data**
 - ▶ But this is a weak definition of “understanding”...
- ▶ Do they understand as we (humans) do? **NO**

Today

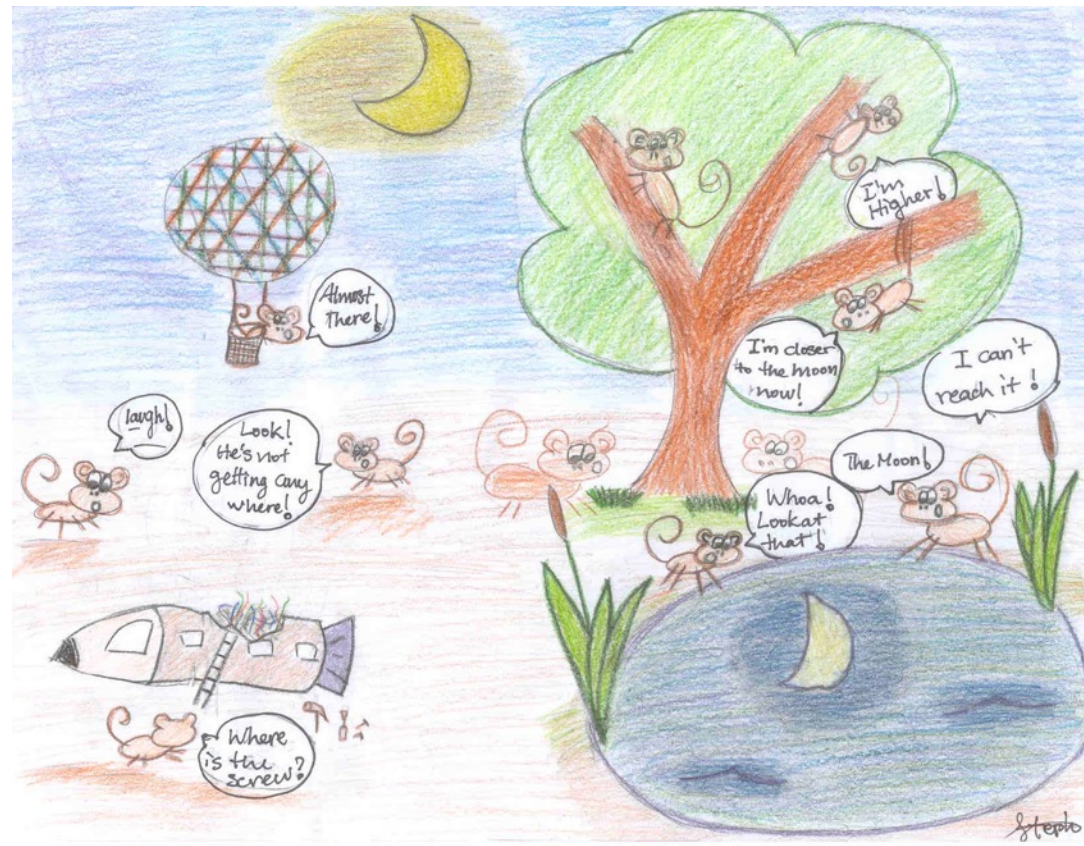
Is it due to the Size of the Data Set?

- ▶ Of course (for obvious reasons), data have a huge responsibility in what modern ML models can or cannot do
 - ▶ Large collections of supervised data are costly
 - ▶ Large collections of text (or text and images) can be setup by crawling the web
 - ▶ Transfer learning (picking-up models trained on large datasets) using some popular models has become extremely popular
- ▶ If we expose the machine to more data/examples then its quality will improve, and it will be harder to distinguish the machine from a human
- ▶ *Can we really think that in order to reach a “human-like way of understanding” it is only a matter of exposing the machine to all the possible data that one might ever find in the world?*

How to Reach the Moon

- Perhaps it is a longer-term path...

Taken from the webpage of
Song-Chun Zhu, University of
California
http://www.stat.ucla.edu/~sczh/research_blog.html
Comic drawn by his 11-years old
daughter (2010)



Today

The Smoothie of Pre-designed Blocks

- ▶ Most of nowadays DL papers are based on models that consists in **interconnecting pre-designed blocks in a “certain manner”**, with the unique goal of showing good results in some tasks
 - ▶ On one hand, it is good to get nice results! 😊
 - ▶ However, sometimes it really sounds like a bare trial-and-error ☹️
- ▶ Low effort in devising new learning theories, no in-dept analysis of the properties of the models
 - ▶ The best you might find: "ablation study"
 - ▶ Turn on and off something and see what happens...



Today

Software Easiness Vs. Understanding

- ▶ TensorFlow and PyTorch (to name a few) had a large importance in making ML/DL popular
 - ▶ Strongly reusing software is a crucial feature in software development
- ▶ ML/DL involves data, optimization, not deterministic procedures, ...
 - ▶ Problem, data, context, setting, ...
- ▶ **Issue:** missing a dept understanding of the modules that are integrated into the target ML solution (example: magic default values for several parameters)

The Issues with Education

I know how to implement a Semantic Labeler using transposed convolutions

Anyhow, what is a validation set?



You don't even know what convolution is...

"I think you are giving me bad grades merely because I am a prophet of doom."

This Photo by Unknown Author is licensed under CC BY-NC-ND

Today

Everybody is doing Machine Learning!

- ▶ When opening my Facebook page or Twitter, I get flooded by a ton of announcements of new libraries, new results, new papers, ...
 - ▶ That's great, isn't it? Yes and No 😊
- ▶ New people approach ML/DL from scratch and in a small amount of time
 - ▶ It is very easy to follow a tutorial and setup a model for some well-known data
 - ▶ It is very easy to make bold, non-trustable announcements of new code and methods
- ▶ **Issue:** what about approaching new real-world problems?

Today

Opportunities Vs. Capability of Exploiting Them

- ▶ **Opportunities:** the industry opens to ML-based solutions in several contexts



- ▶ **Exploiting such opportunities:** deep understanding of the problem, capability of adapting ML to the considered problem
 - ▶ Reduce resource usage
 - ▶ Analyze the data
 - ▶ Carefully design the experimental setup
 - ▶ Maximize performance
 - ▶ Have a reasonable control of the proposed model (know-what-you-are-doing)

Past and Present of Deep Learning: **Now What?**

PART 3

Education

- ▶ Pay attention to the details, emphasize the role of “understanding” vs. “running some code”
- ▶ Reduce the overinflated expectations that are incorrectly circulating in some communities
 - ▶ Be conscious of the pros and cons of nowadays ML/DL!
- ▶ **Great results** in several tasks but there is still a **huge amount of work to do**
 - ▶ Don't think that humans are beaten or ready-to-be-replaced everywhere, that's not the case yet!

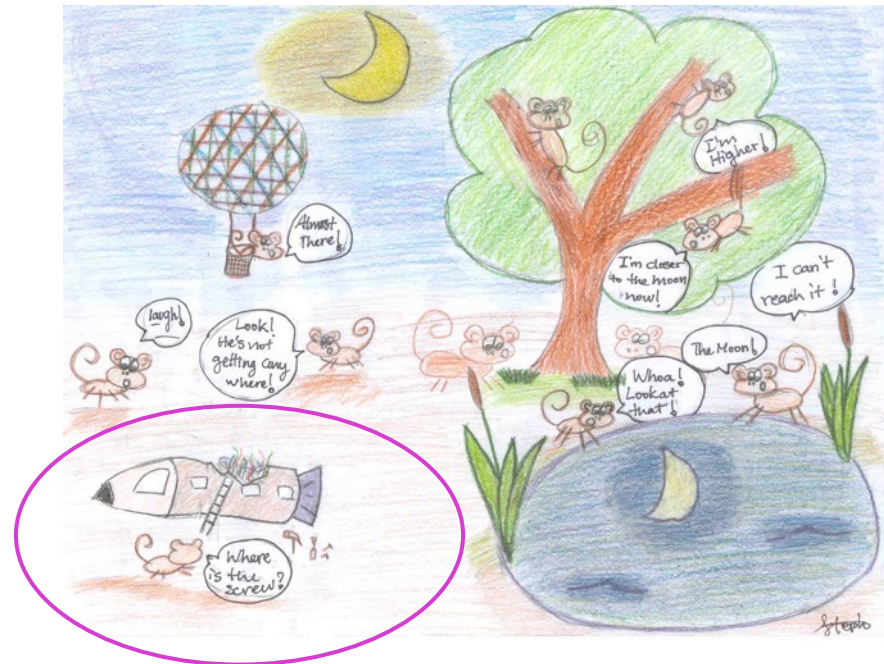


Now what?

It's Time to Try to Build a Spaceship

- ▶ Accept **performance decrease** for
 - ▶ Better understanding
 - ▶ New models
 - ▶ Longer term projects
- ▶ The research community is more “open” to this that a few years ago

Taken from the webpage of
Song-Chun Zhu, University of California
http://www.stat.ucla.edu/~sczhu/research_blog.html
Comic drawn by his 11-years old daughter (2010)



Now what?

Research in Machine Learning

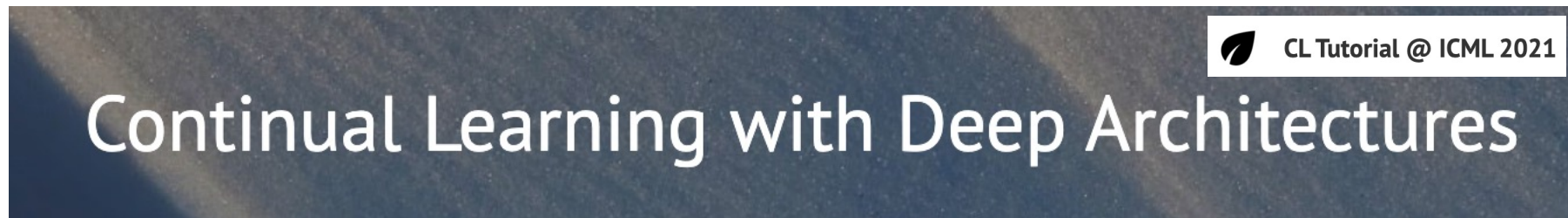
- Several important ingredients were put on the corner due to the attraction of low-term incremental improvements
- Other elements were not strongly considered in favor of intrinsically artificial benchmarking strategies

Methods		Trained on	0.5:0.95	0.5	0.75	S	M	L	1	10	100	S	M	L
Fast R-CNN [16]	train	20.5	39.9	19.4	4.1	20.0	35.8	21.3	29.4	30.1	7.3	32.1	52.0	
ION [95]	train	23.6	43.2	23.6	6.4	24.1	38.3	23.2	32.7	33.5	10.1	37.7	53.6	
NOC+FCRN(VGG16) [114]	train	21.2	41.5	19.7	-	-	-	-	-	-	-	-	-	-
NOC+FCRN(Google) [114]	train	24.8	44.4	25.2	-	-	-	-	-	-	-	-	-	-
NOC+FCRN (ResNet101) [114]	train	27.2	48.4	27.6	-	-	-	-	-	-	-	-	-	-
GBD-Net [109]	train	27.0	45.8	-	-	-	-	-	-	-	-	-	-	-
OHEM+FCRN [113]	train	22.6	42.5	22.2	5.0	23.7	34.6	-	-	-	-	-	-	-
OHEM+FCRN* [113]	train	24.4	44.4	24.8	7.1	26.4	37.9	-	-	-	-	-	-	-
OHEM+FCRN* [113]	trainval	25.5	45.9	26.1	7.4	27.7	38.5	-	-	-	-	-	-	-
Faster R-CNN [18]	trainval	24.2	45.3	23.5	7.7	26.4	37.1	23.8	34.0	34.6	12.0	38.5	54.4	
YOLOv2 [72]	trainval35k	21.6	44.0	19.2	5.0	22.4	35.5	20.7	31.6	33.3	9.8	36.5	54.4	
SSD300 [71]	trainval35k	23.2	41.2	23.4	5.3	23.2	39.6	22.5	33.2	35.3	9.6	37.6	56.5	
SSD512 [71]	trainval35k	26.8	46.5	27.8	9.0	28.9	41.9	24.8	37.5	39.8	14.0	43.5	59.0	
R-FCN (ResNet101) [65]	trainval	29.2	51.5	-	10.8	32.8	45.0	-	-	-	-	-	-	-
R-FCN*(ResNet101) [65]	trainval	29.9	51.9	-	10.4	32.4	43.3	-	-	-	-	-	-	-
R-FCN**(ResNet101) [65]	trainval	31.5	53.2	-	14.3	35.5	44.2	-	-	-	-	-	-	-
Multi-path [112]	trainval	33.2	51.9	36.3	13.6	37.2	47.8	29.9	46.0	48.3	23.4	56.0	66.4	
FPN (ResNet101) [66]	trainval35k	36.2	59.1	39.0	18.2	39.0	48.2	-	-	-	-	-	-	-
Mask (ResNet101+FPN) [67]	trainval35k	38.2	60.3	41.7	20.1	41.1	50.2	-	-	-	-	-	-	-
Mask (ResNeXt101+FPN) [67]	trainval35k	39.8	62.3	43.4	22.1	43.2	51.2	-	-	-	-	-	-	-
DSSD513 (ResNet101) [73]	trainval35k	33.2	53.3	35.2	13.0	35.4	51.1	28.9	43.5	46.2	21.8	49.1	66.4	
DSOD300 [74]	trainval	29.3	47.3	30.6	9.4	31.5	47.0	27.3	40.7	43.0	16.7	47.1	65.0	

Now what?


Interesting Trends

- ▶ The role of data?
- ▶ Lifelong learning



Now what?

Interesting Trends (Cont.)



Can Machines Learn to See without Visual Databases?

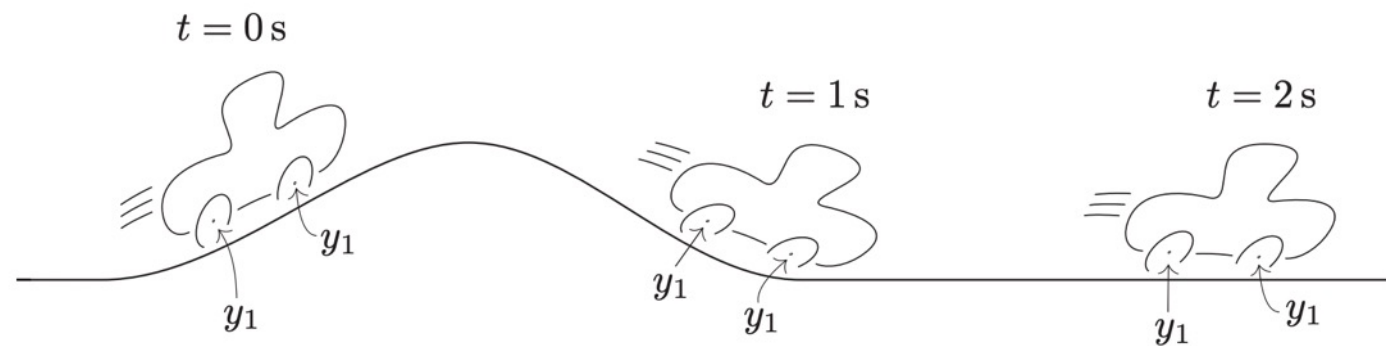
Alessandro Betti, Marco Gori, Stefano Melacci, Marcello Pelillo, Fabio Roli
NeurIPS Data-Centric AI Workshop (DCAI 2021)

<https://neurips.cc/virtual/2021/workshop/21860>

Now what?

Time Does Matter!

- ▶ Another common assumption is data **data-order** does not matter (stochastic gradient)
 - ▶ Humans do not learn from “shuffled frames”!
- ▶ **Time vs. iteration index**: learning dynamics linked to the time in which events take place
- ▶ **Life-long learning**
 - ▶ *Online*



Now what?

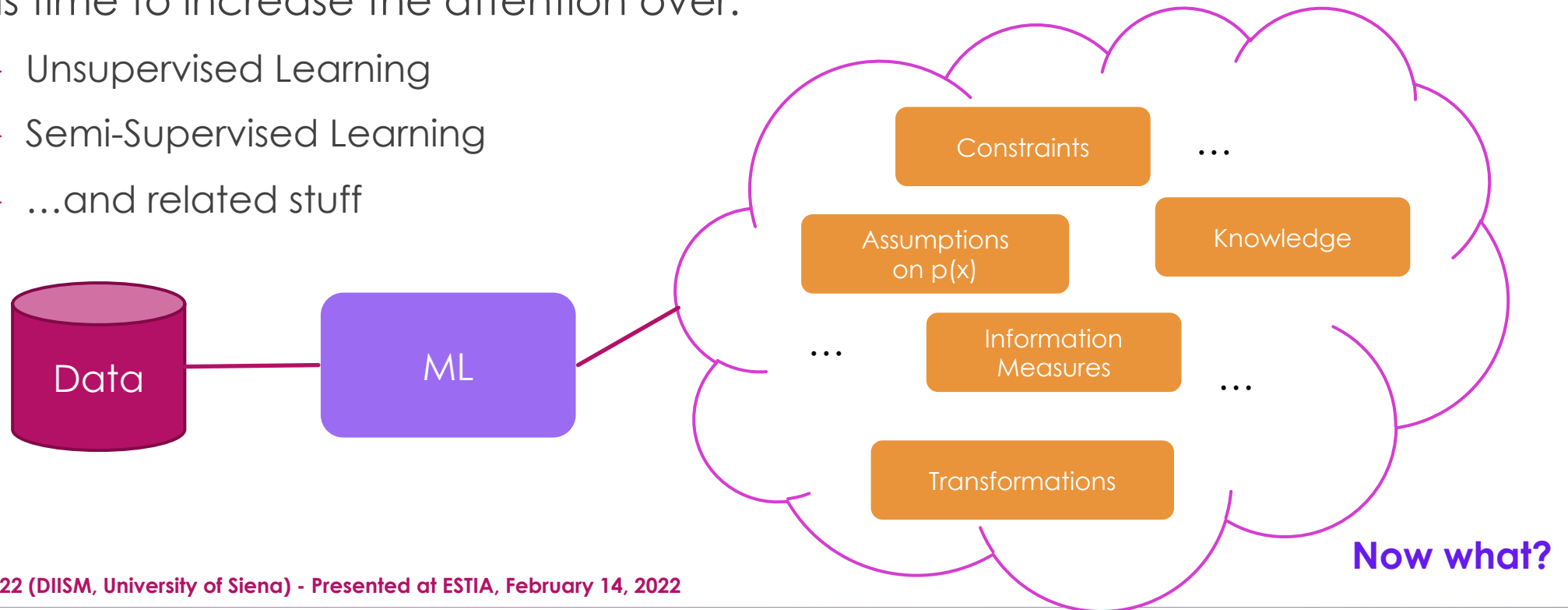
Questions & Challenges

- ▶ Should we focus on the role of time & motion to design algorithms that better leverage the data?
- ▶ Can we design agents based on the on-line processing of the visual information acquired in a target environment, and that sporadically interact in a human-like manner?
- ▶ Can we stop thinking about training and testing as two distinct stages and evaluate the agents while they are still learning?



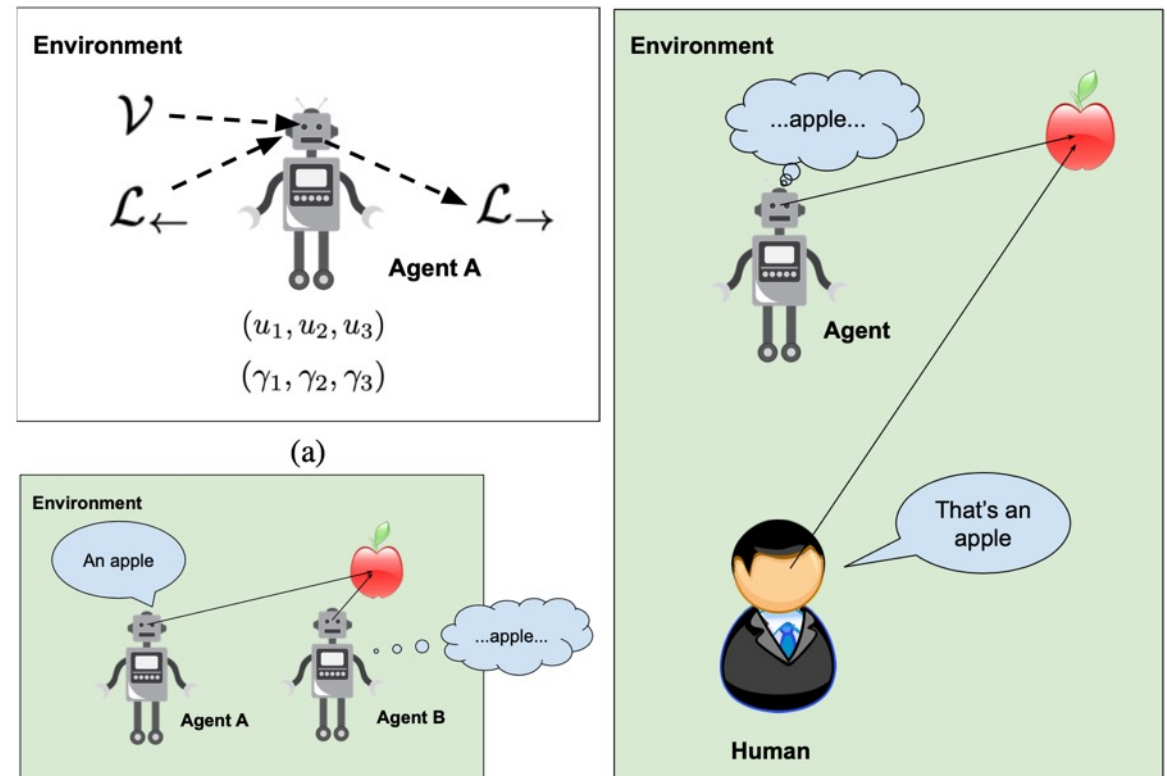
Unsupervised Dynamics of Learning

- ▶ Humans do not learn by being exposed to millions (?) of supervisions
- ▶ It is time to increase the attention over:
 - ▶ Unsupervised Learning
 - ▶ Semi-Supervised Learning
 - ▶ ...and related stuff



Environment/Context

- ▶ Artificial agents studied in conjunction with the context in which they will operate
- ▶ It is commonly assumed that collecting samples is the way to go, but what about interactive settings? What about multi-modal settings?
 - ▶ Interactions/feedbacks are structured and multi-turn



Now what?

Learning in the “Wild”

- ▶ Moving from artificial experimental conditions to the real operative setting requires adaptations
- ▶ Emphasize the **adaptability** of ML-based models in order to face situations that are not so close to the one experienced at training time

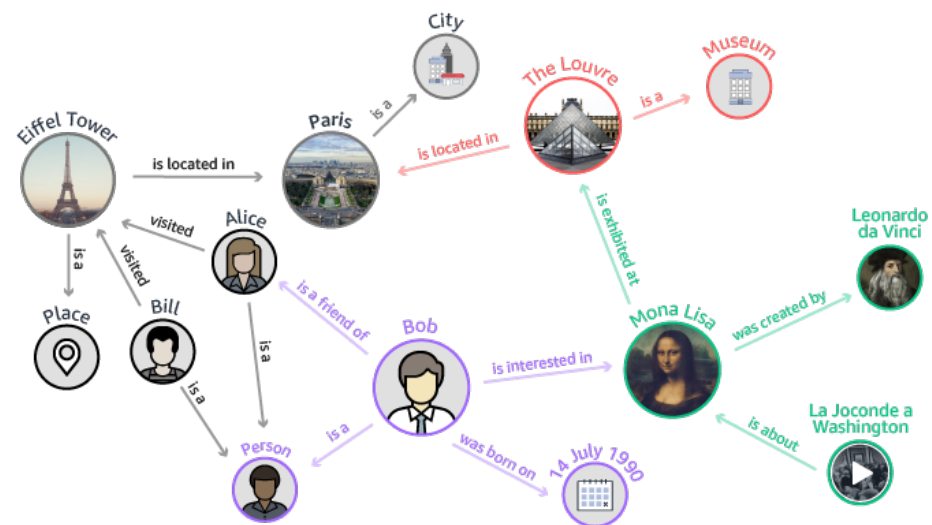


Taken from: <https://wand-research.com/blogging-expectation-vs-reality/>

Now what?

Machine Learning & Knowledge

- ▶ We have the use of **large collections of structured data**
 - ▶ Knowledge-graphs
 - ▶ Frequently ignored with a preference on end-to-end learning from insanely big corpora
- ▶ The integration of ML with these types of knowledge is another feature with a lot of value
 - ▶ Neuro-Symbolic Learning



This Photo by Unknown Author is licensed under [CC BY-NC](#)

Now what?

Trustworthy, Explainable AI

- ▶ Example:
 - ▶ TAILOR Project
 - ▶ 53 partners

Foundations of Trustworthy AI - Integrating Reasoning, Learning and Optimization

Scheda informativa

Descrizione del progetto

DE

EN

ES

FR

IT

PL

AI systems made safe, transparent and reliable

Maximising opportunities and minimising risks associated with artificial intelligence (AI) requires a focus on human-centred trustworthy AI. This can be achieved by collaborations between research excellence centres with a technical focus on combining expertise in the areas of learning, optimisation and reasoning. Currently, this work is carried out by an isolated scientific community where research groups are working individually or in smaller networks. The EU-funded TAILOR project aims to bring these groups together in a single scientific network on the Foundations of Trustworthy AI, thereby reducing the fragmentation and increasing the joint AI research capacity of Europe, helping it to take the lead and advance the state-of-the-art in trustworthy AI. The four main instruments are a strategic roadmap, a basic research programme to address grand challenges, a connectivity fund for active dissemination, and network collaboration activities.

Mostra l'obiettivo del progetto

Campo scientifico

/scienze naturali/informatica e scienze dell'informazione/intelligenza artificiale

Informazioni relative al progetto

TAILOR
ID dell'accordo di sovvenzione: 952215

Stato
Progetto in corso

Data di avvio
1 Settembre 2020

Data di completamento
31 Agosto 2023

Finanziato da
H2020-EU.2.1.1.

Bilancio complessivo
€ 12 000 000

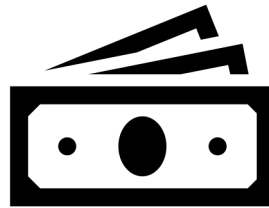
Contributo UE
€ 12 000 000

Coordinato da
LINKÖPINGS UNIVERSITET
 Svezia

Now what?

Power Consumption & Model Sizes

- ▶ Reducing model sizes has a direct implication in reducing the power consumption
 - ▶ Training, inference
- ▶ We must spend time in finding the right architectures not only in terms of performances, but also in terms of model size (the smaller the better)
 - ▶ Finding the “right” architecture might mean “inventing new ones”



Now what?

Open Eyes (Academy & Industry)

- ▶ One of the skills that a researcher/data scientist in ML/DL must have is the capability of distinguishing those works that are expected to have value for considered research topic/application
 - ▶ Navigating the sea of real and self-claimed SOTA models/implementations
 - ▶ It's always been like this, but look at the numbers of ML/DL papers...



Now what?

Conclusions

- ▶ We reviewed the last 15-20 years ML with focus on Deep Learning
 - ▶ Great progresses! 😊
 - ▶ Several ideas from “the past” are the bases of nowadays DL
 - ▶ Other “important” ideas were not strongly followed up
- ▶ We must control the large popularity of ML/AI in order to keep our eyes on the real thing, filtering out the “noise”
- ▶ It’s time to start re-thinking to the goal of building autonomous agents that “learn”
- ▶ A lot of great tools/results ready-to-be exploited...
 - ▶ ...a lot of work to do 😊

Now what?

Questions?

Thank you very much for your attention