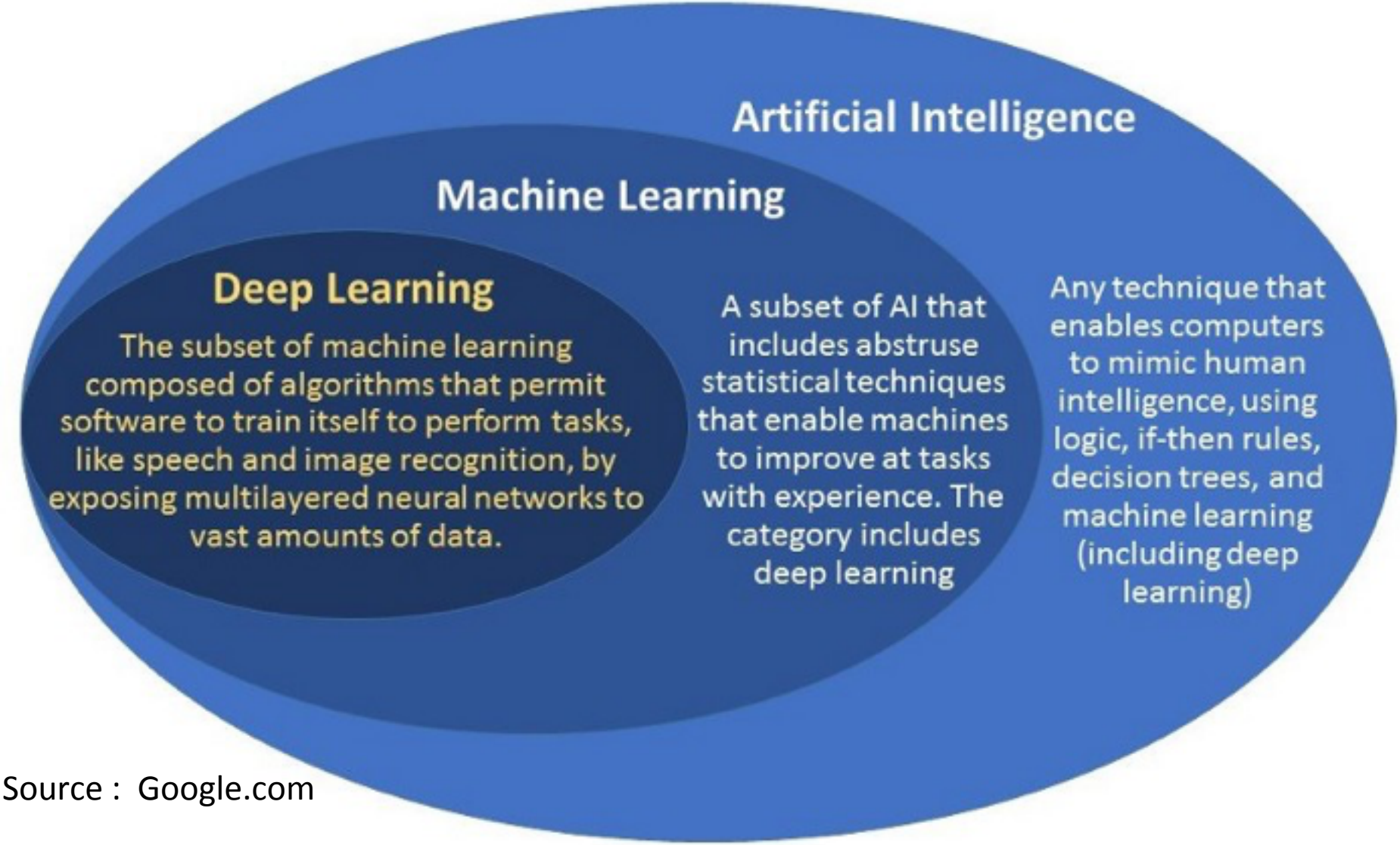
Elements
Machine Learning/Deep Learning

Olivier J.J. MICHEL,
Pr. Grenoble-INP
Olivier.Michel@grenoble-inp.fr

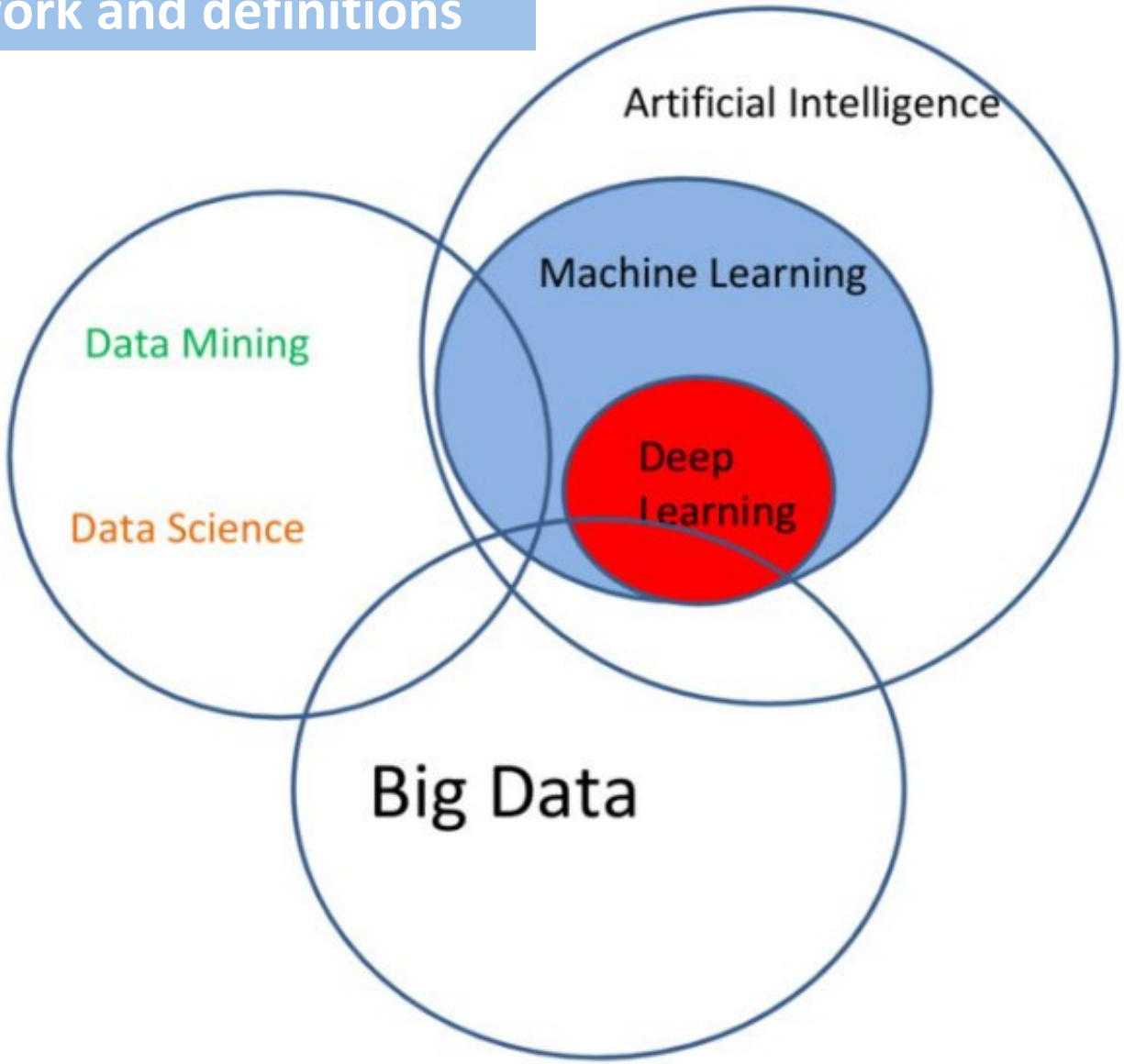
GIPSA-Lab, UMR 5216 CNRS

Framework and definitions



Source : Google.com

Framework and definitions



Machine Learning /Deep Learning goal and features :

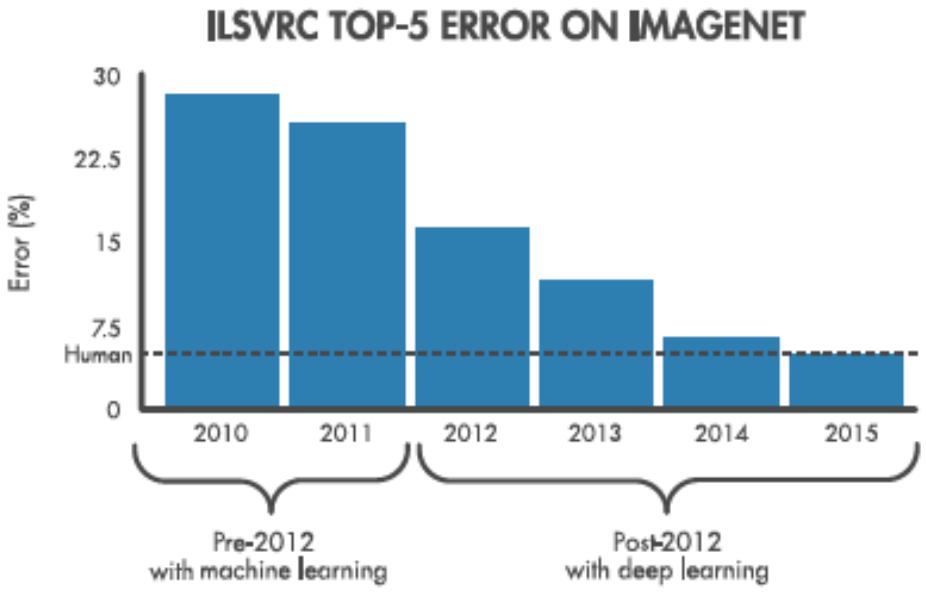
To construct computer systems that automatically improves through experience

Key feature = far easier to train a system by showing input-output examples, that to program output by anticipating all possible inputs.

= scalability to « Big Data » sets, data mining ability

Approaches allowing to analyse high throughput data in a novel way

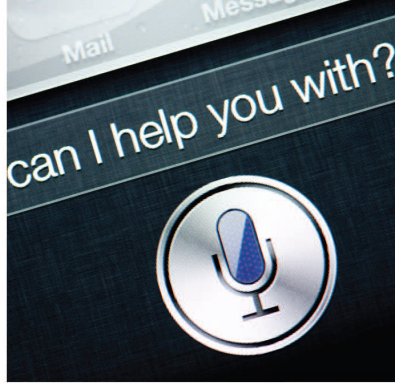
Recent Evolution of Performances in Image classification

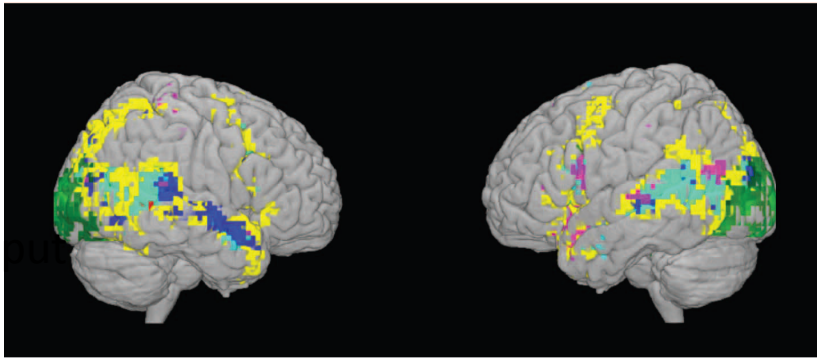


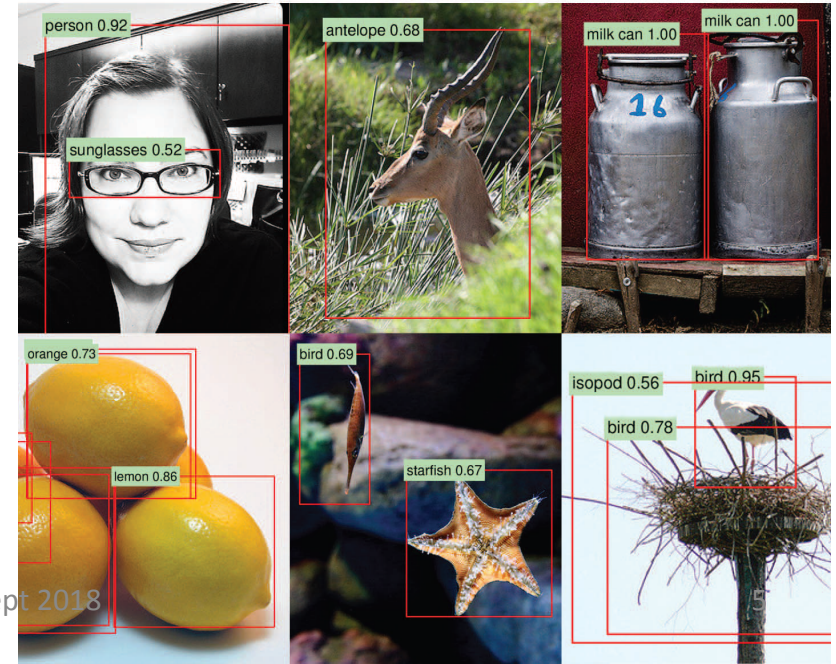
Huge incidence in

- in computer science
- in industry concerned with data intensive issues
- in « empirical sciences »
  - social sciences
  - biology
  - cosmology…

Again …. Because ML allows to analyse high throughput data in a novel way







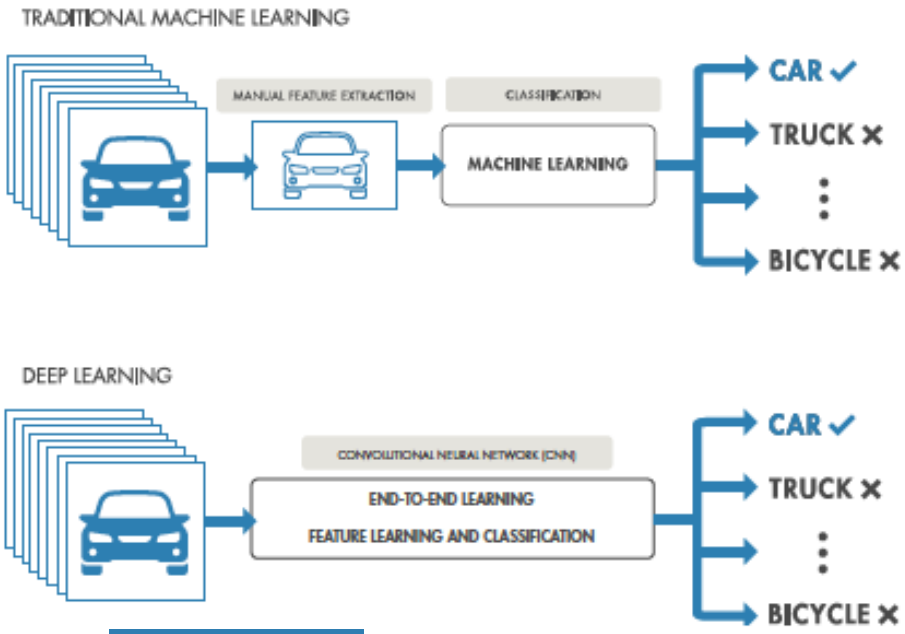
Machine learning is having a substantial effect on many areas of technology and science; examples of recent applied success stories include robotics and autonomous vehicle control (top left), speech processing and natural language processing (top right), neuroscience research (middle), and applications in computer vision (bottom).

Source :
Machine Learning: Trends, perspectives, and prospects
M.I.Jordan, T.M.Mitchell, Science 349, 255 (2015)

What is the Difference Between Deep Learning and Machine Learning?

Deep learning is a subtype of machine learning. With machine learning, you manually extract the relevant features of an image. With deep learning, you feed the raw images directly into a deep neural network that learns the features automatically.

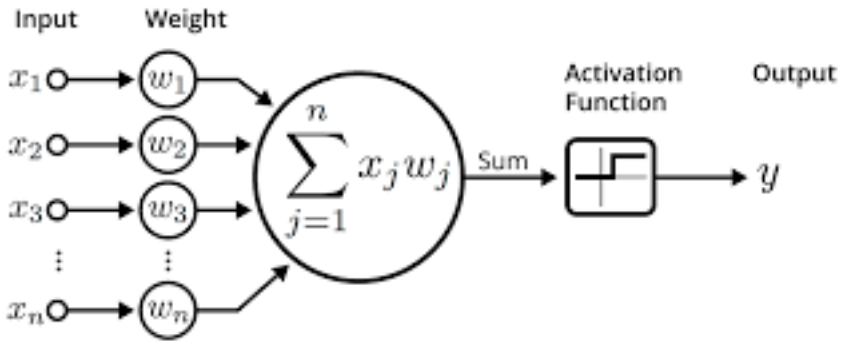
Deep learning often requires hundreds of thousands or millions of images for the best results. It's also computationally intensive and requires a high-performance GPU.



Machine Learning	Deep Learning
+ Good results with small data sets	— Requires very large data sets
+ Quick to train a model	— Computationally intensive
— Need to try different features and classifiers to achieve best results	+ Learns features and classifiers automatically
— Accuracy plateaus	+ Accuracy is unlimited

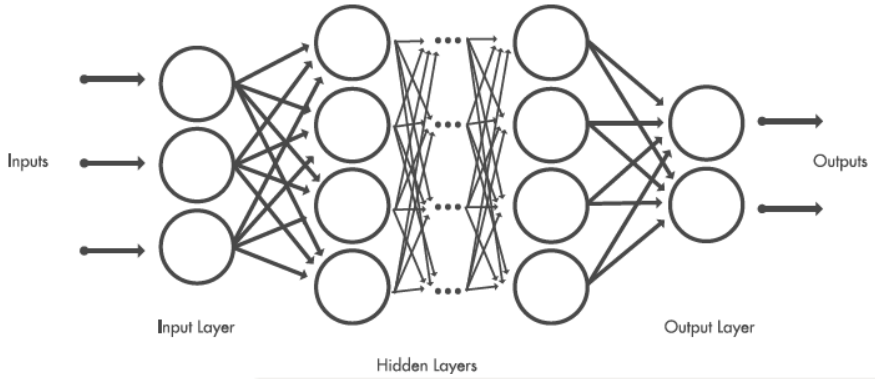
source MathWorks

Basic components for NN, DNN, CNN

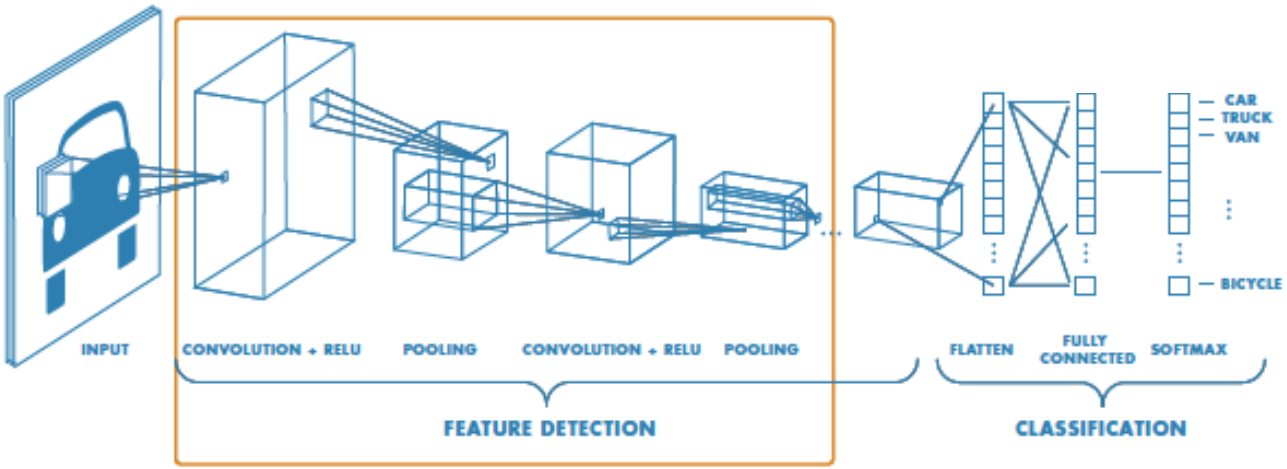


An illustration of an artificial neuron. Source: Becoming Human.

Deep NN



There is no exact formula for selecting layers. The best approach is to try a few and see how well they work—or to use a pretrained network.



Convolutional NN

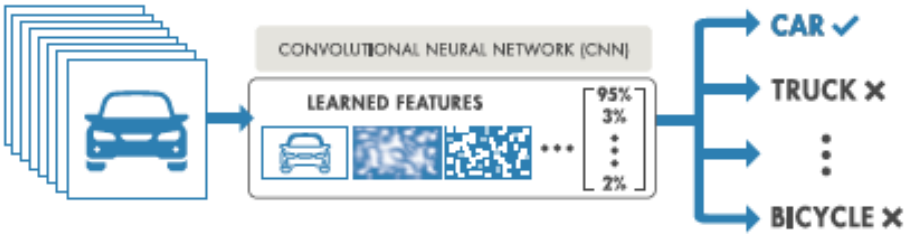
=> On n'attend pas des perf « optimales »…multiplicité de minima locaux…

What Makes Deep Learning State-of-the-Art?

What are the challenges/problems raised ?

« Easy » access to massive sets a (labeled) data

(-> Privacy/security/Network throughput)

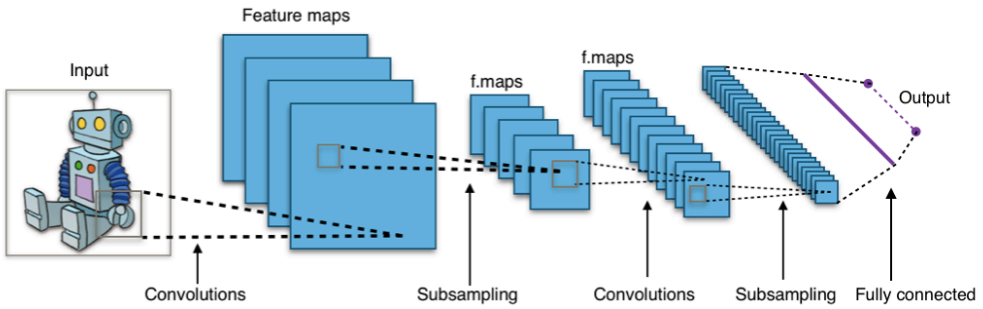


High computing power, high perf GPU

->Cost of Training from massive amount of data

Algorithmic developments, learning strategies (Transfer Learning, Reinforcement learning, Optimization methods…)

-> Perf analysis, bounds and guarantees?



A data science perspective to define key scientific and practical goals :

-> How to theoretically characterize the capabilities of a (deep) learning algo?

*how accurately does the algo learn from a given type and volume of training data?
*how robust is the ML algo to errors in the model or sensitivity to missing data?
*Given a learning task + training data,
is the design of an algorithm possible (how/ tractable?)
(*how much data, which computational cost for a given task?*)

Rk : DL not (yet) used in critical contexts (Defense, medicine, …)

→ requires blends of (<u>in addition to knowledge in the engineering field of expertise</u>)

Statistical decision theory,
Computational complexity theory
Information theory, Sampling theory,
Optimization methods, to provide bounds on convergence rates
Computer programming, architecture, data manipulation

=> Grenoble has many strength in this perspective

I.E. MUCH MORE THAN « BLACK BOX BASED » solution design skills often found
- in the increasing demand from
  - many students
  - many companies offering internships
- or even in some published (rank A) literature :

Guides for deep learning submissions in IEEE T-IFS

Aiming to strengthen scientific reproducibility among our peers, this document lays down some guidelines for submissions to IEEE T-IFS.
Reproducibility is one aspect that should be assessed by reviewers and considered by the AE in the final acceptance decision.
The guidelines explicitly refer to papers dealing with steganalysis and multimedia forensics, however they are applicable to all papers submitted to T-IFS in which the proposed system relies on deep learning.

Thanks to recent advances in machine learning, forensic and steganalytic attacks have grown in sophistication and complexity. **With more parameters, it becomes easier to omit key information.**
This is particularly so for Deep Learning (DL) methods, which have many variants and details. When some of these are omitted from the paper it weakens reproducibility, and often extends the review process.
Authors are encouraged to make both code and data sets available1 (preferably in time for the review process) but this is not a formal requirement at the moment.
In any case, supplementary code is not a substitute

for a proper written description of the method. To help authors, the T-IFS EB has drawn up this checklist of details, that **DL submissions on IFS topics should take particular care** to include.

- **The type of network (CNN, RNN, BDRNN, …) and loss function.**
- **The topology:**
  - type of each layer (pooling, convolution, non-linearity, …),
  - activation function of each layer (tanh, sigmoid, max- or average-pooling, …)
  - any parameters (strides, weights size, etc).
- **Any pre-processing of network inputs, and interpretation of outputs**. Note that there are different re-scalings commonly called "normalization", so this needs to be specified.

Training phase

- **How the network is initialized.**
  - If randomly, give the exact parameters of the initializing distributions.
  - If there is pre-training, the procedure used (SAE, RBM, etc) and any parameters.
- **The learning algorithm** (SGD, Adagrad, AdaDelta, HF, etc).
  - Hyperparameters of the learning algorithm: (momentum, etc).
  - Mini-batches and (if appropriate) number of representatives of each class in learning batches. Where appropriate (e.g. steganalysis), specify if the data are stratified (i.e. paired cover and stego kept together).
  - If used, exactly what form of batch normalization and pooling.
  - Initial learning rate and its evolution during the learning, number of iterations of epochs, dropout or other regularization factors, etc.
  - The order that training data is visited during the learning process, and whether shuffled between epochs.

From the Transactions website: "The Transactions encourages authors to make their publications reproducible by making all information needed to reproduce the presented results available online.
This typically requires publishing the code and data used to produce the publication`s figures and tables on a website; see the supplemental materials section of the information for authors.
It gives other researchers easier access to the work, and facilitates fair comparisons."

- Stopping criterion (is there a fixed number of iterations or not?)
- Which data was used to determine the hyperparameters, learning rates, regularization parameters, and stopping criterion; the policy for hyperparameter search.

Test phase

- How data is broken down into training, validation, and testing sets.
- If cross-validation is used, how it is arranged and whether it is repeated.

Implementation

- If a GPU is used, which model.
- If you have used a standard deep learning tool, the exact version and any tool-specific parameters.
- If you propose techniques for scalability (kernel approximations, etc) then specify.

We encourage the inclusion of a "methodology" section in the paper to contain these details, which could be in an appendix, or as supplementary material providing that this is available for review.
Algorithms and methods need not be described from scratch: a good citation is adequate, as long as the exact variety is clear and all parameters have been specified.
While we encourage the publication of negative results contradicting previously published research, in these cases a higher standard of reproducibility applies. Barring exceptional circumstances, data sets should be made available for the review process (if not, this should be explained to the AE). **Finally, authors should demonstrate the statistical significance of results claimed.**

Sources : Google.com, Mathworks.com, Sciences; IEEE SP magazine, IEEE T-IFS, Cours College de France S.Mallat, …

For fun / or intuition ….

http://playground.tensorflow.org