

Machine Learning

Examples you know ?

Machine Learning

Examples you know ? the 'woaw effect' of daily life

- ChatGPT and other conversational agents
- Image generators (image to text, text to image, etc)
- Google beats the world-champion of **Go** (and starcraft 2 ?)
- Picture **classification** (10, 100, 1000 categories)
- **DeepL** (natural language translation)
- **Health**: lots of talking, not lots of action (yet?)
- Less glamorous:
 - targeted **marketing** (Amazon...)
 - **Financial** Markets
- + Many others ...

Examples so old that we forget:

- spam filters
- Face detection in cameras, from 2000' (this is just SVM !)
- Smile detection → similar
- ...

MNIST (1998)

from 4% to 0.21% error



CIFAR10:

21% error in 2010, 1% in 2018

airplane



automobile



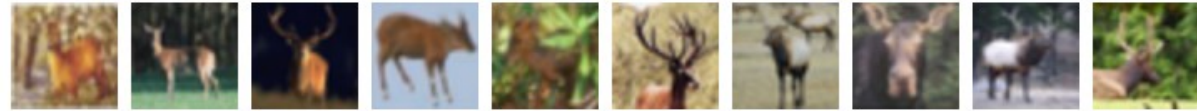
bird



cat



deer



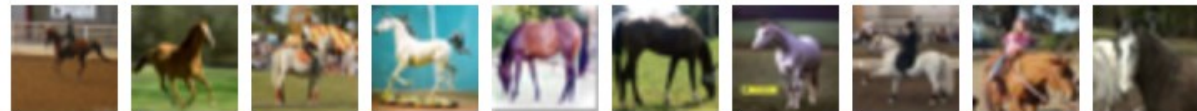
dog



frog



horse



ship



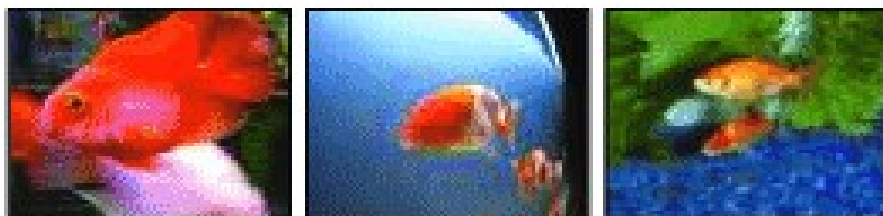
truck



ImageNet:

today: Beyond-human accuracy

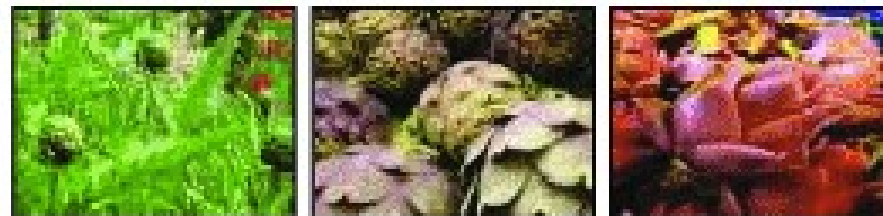
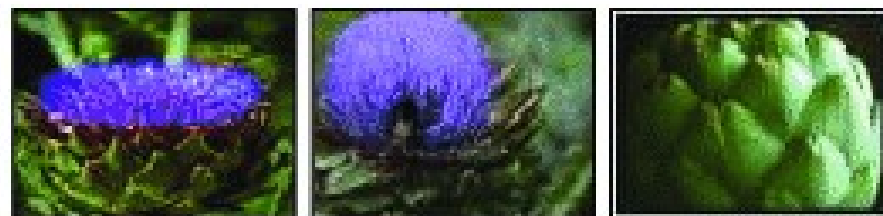
easy



hard



Goldfish - easy (23 blocks) vs. hard (29 blocks)

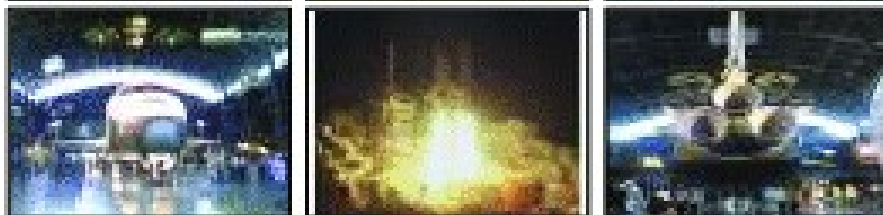


Artichoke - easy (18 blocks) vs. hard (28 blocks)

easy



hard

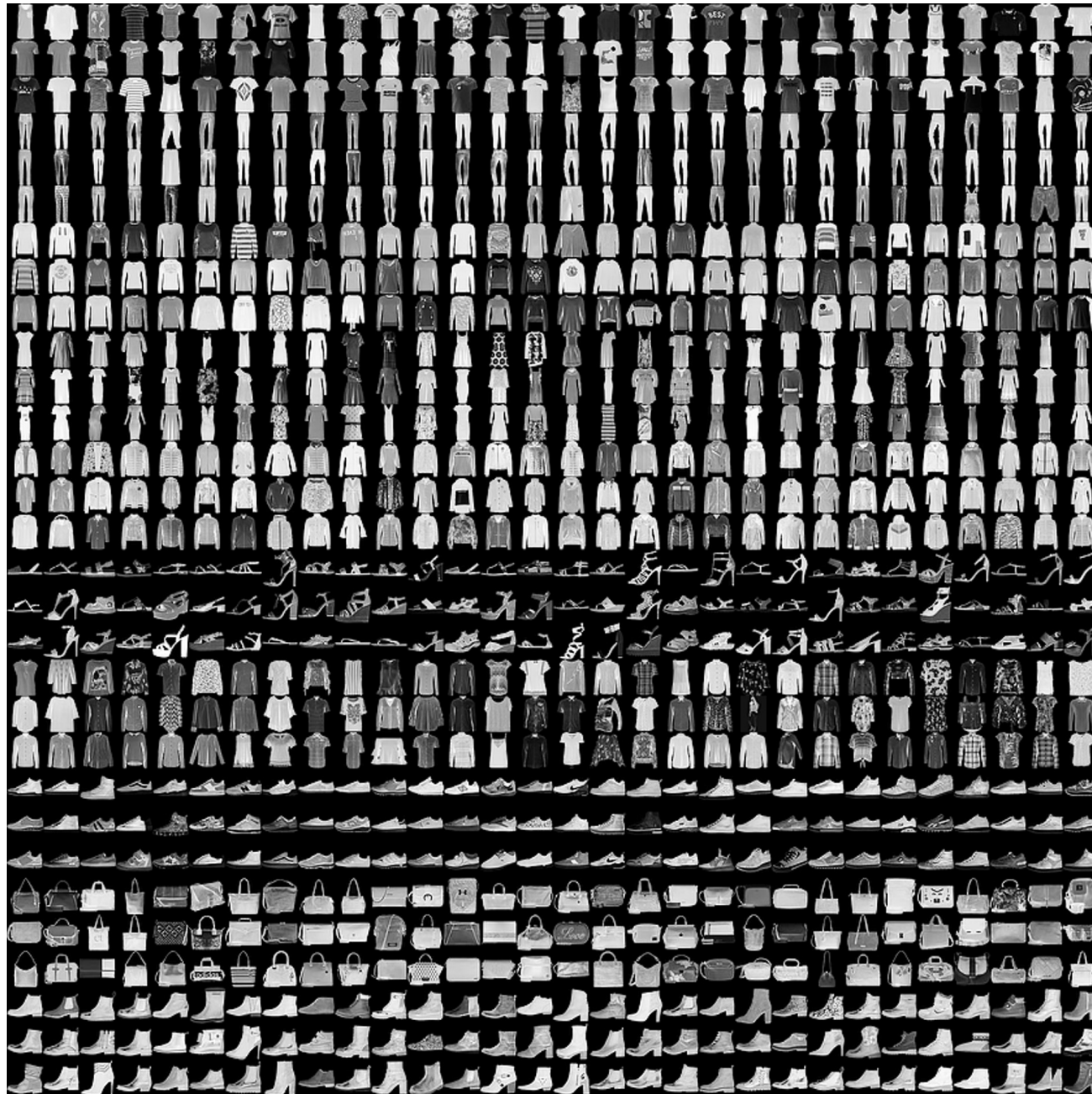


Spacecraft - easy (23 blocks) vs. hard (29 blocks)



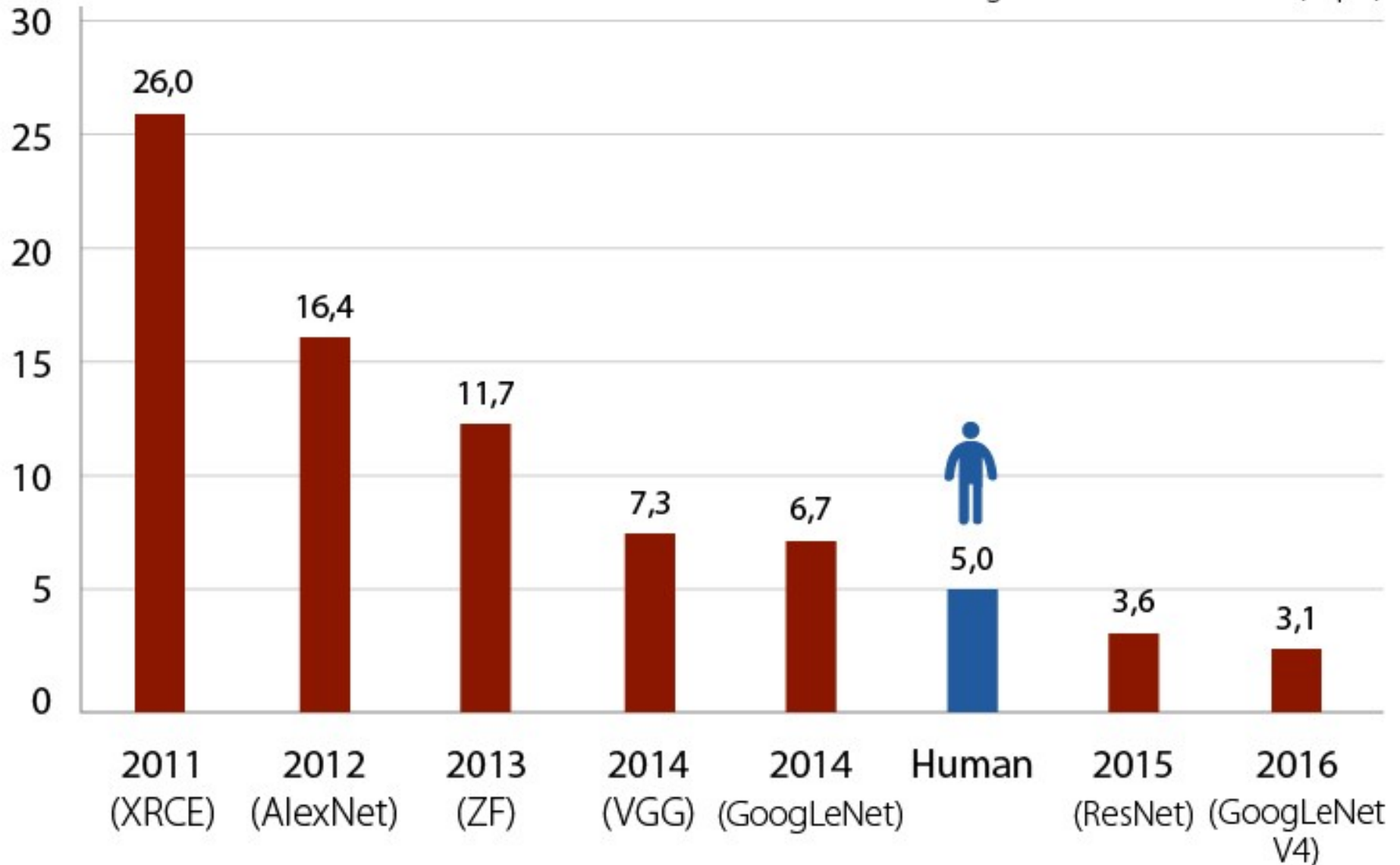
Bridge - easy (24 blocks) vs. hard (29 blocks)

Fashion-Mnist



Error rate over time

ImageNet Classification Error (Top 5)



Autres performances, rapportées ici:

https://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html

Data Sets / Benchmarks

- MNIST: 1998 (based on NIST) → *(this class)* between 4% and 0.21% error
- CIFAR10: 21% error in 2010, 1% in 2018
- CIFAR100: 55% accuracy 2012, 91.3% in 2018
- ImageNet: from ~30%?(2011) to 15% (2012) ...
... to 5% (2015) !

Why? **2012 = development of Deep Learning**
→ renewed interest in ML

- Other performances, see :

https://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html

Why this surge of success ?

Several factors explain the recent **success** (since ~2012) of ML (and thus, the renewal of *interest* for ML):

- Large sets of data available (increasingly true)
- GPUs = cheap, fast intensive computations
- Automatic differentiation / user-friendly software
- **New algorithms** (CNNs taken seriously after 2012, and many others since)

In this class we talk about the algorithms (basic ones).
But the first 3 points are very important !! Each ones reinforces the other.