Deep Learning

Deep Learning
23 September 2024
Profs. Luigi Cinque, Fabio Galasso and Marco Raoul Marini

SAPIENZA
UNIVERSITÀ DI ROMA

Lecturers

- Lecturers
  - Prof. Luigi Cinque
  - Prof. Fabio Galasso
  - Prof. Marco Raoul Marini

- Teaching Assistants
  - Federico Fontana
  - Romeo Lanzino
  - Leonardo Plini
  - Diego Bellani

Lecture and Exercise

- Lectures:
  - Mondays 14:00-17:00 @Aula Magna (Via Regina Elena, 295)
  - Thursdays 16:00-18:00 @Aula Magna (Via Regina Elena, 295)
- Office hours
  - Fabio: Thursdays 13:30-15:30 @Room 43, 2° floor, build. G, via Regina Elena 295
  - Marco: Monday 9:00-11:00 @Room 309, 3° floor, via Salaria 113
- Website: https://sites.google.com/di.uniroma1.it/DL-2024-2025
- Mailing: https://groups.google.com/u/1/a/di.uniroma1.it/g/DL-24-25
  - For slides, course material, assignments and news
- **Subscribe to it now**

Exam

- Exam
  1. Written
  2. Sample questions will be provided and discussed at lecture

- Kaggle Competition:
  - 1st: early in the course (worth +1 of your course grade)
  - 2nd: in the second part of the course (worth +2 of your course grade)

Physical and Learning Disabilities

- Sapienza provides counseling and support

- You may reach out to:
  sportellodisabili@uniroma1.it and counselingdsa@uniroma1.it

- Or directly to:
  Prof. Tiziana Calamoneri
  Coordinator for Disabilities and DSA for the I3S Faculty
  http://wwwusers.di.uniroma1.it/~calamo/

Material

- Slides and coding scripts are distributed after lectures
- There is much material online (papers, blogs, twits, etc.)
- Books (more at https://sites.google.com/di.uniroma1.it/dl-2023-2024/resources)
  - Aston Zhang, Zachary Lipton, Alexander J. Smola, Mu Li, 2023. Dive Into Deep Learning (available at: https://d2l.ai/)
  - Francois Fleuret, 2024. The Little Book of Deep Learning. (available at: https://fleuret.org/dlc/)
- Additional reference books
  - Deisenroth, Faisal, Ong (2020). Mathematics for Machine Learning (https://mml-book.github.io/book/mml-book.pdf)
  - Antonio Torralba, Phillip Isola and William T. Freeman, 2024. Foundations of Computer Vision (https://mitpress.mit.edu/9780262048972/foundations-of-computer-vision/)
  - Frank Dellaert, 2024, Robotics. (available at: https://www.roboticsbook.org/)

Coding References

- Coding classes will be in Python and Pytorch
- Books for Python/Pytorch
  - Aston Zhang, Zachary Lipton, Alexander J. Smola, Mu Li, 2023. Dive Into Deep Learning (available at https://d2l.ai/ , select Notebooks/Pytorch)
  - Jake VanderPlas, 2016. Python Data Science Handbook: Tools and Techniques for Developers: Essential Tools for working with Data (Book and notebooks available at: https://github.com/jakevdp/PythonDataScienceHandbook)
- Online tutorials for Python: https://docs.python.org/3/tutorial/
- Online tutorials for Pytorch: https://pytorch.org/tutorials/
  - https://pytorch.org/tutorials/beginner/deep_learning_60min_blitz.html

Setup and computing

- A Linux OS is recommended
  - but Python and Pytorch also run on Windows and MacOS
- Recommended Python distribution: Anaconda which includes the IDE Spyder and Jupyter Lab
  - https://www.anaconda.com/distribution/
- For the coding classes you will need a GPU
  - For “local” users, you will need CUDA and CUDnn
    - Check for your GPU model: https://pytorch.org/get-started/locally/
  - Consider Google colab: https://colab.research.google.com

Pre-requisites

- General notions of calculus and linear algebra
  - taking derivatives, understanding matrix vector operations and notation
- Basics of probability and statistics
  - Gaussian distributions, mean, standard deviation, etc.
- General notions of artificial intelligence and machine learning
  - cost functions, derivatives, backpropagation and optimization with gradient descent and the computational graph (we will briefly review some concepts)
  - for your review, consider chapters 5, 6, 7 of:
    - Deisenroth Faisal Ong Book 2020 Mathematics for Machine Learning
- Proficiency in Python, some experience with Pytorch

Syllabus

- Review of linear filtering with convolutions, neural networks and backpropagation
- Convolutional neural networks, basics and modern architectures
- Training deep neural networks
- Recurrent neural networks
- Transformer networks and attention
- Generative neural networks, including GANs, VAEs and Diffusion models

Provided time

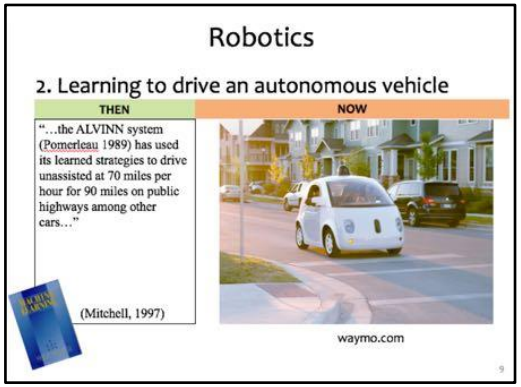
- Self-supervised Learning, meta-learning, continual learning, Q-learning, reinforcement learning, binary neural networks, visualization, ethics of modern deep learning

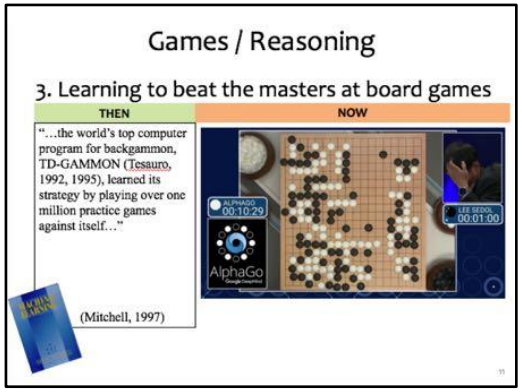
What is Machine Learning

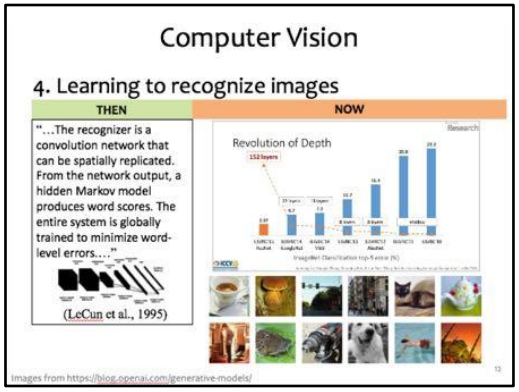
SAPIENZA
UNIVERSITÀ DI ROMA

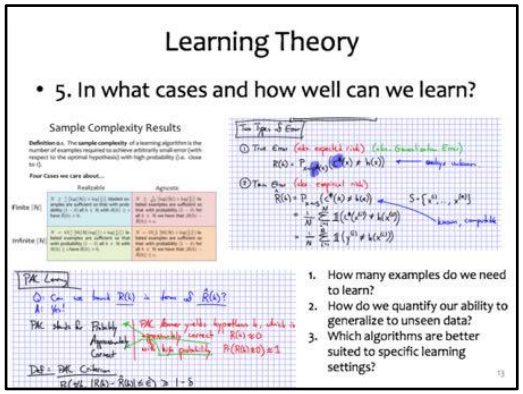
What is ML?


Speech Recognition
1. Learning to recognize spoken words
THEN
NOW
"…the SPHINX system (e.g. Lee 1989) learns speaker-specific strategies for recognizing the primitive sounds (phonemes) and words from the observed speech signal…neural network methods…hidden Markov models…"
(Mitchell, 1997)
Source: https://www.stonetemple.com/great-knowledge-box-showdown/#VoiceStudyResults


Robotics
2. Learning to drive an autonomous vehicle
THEN
NOW
"…the ALVINN system (Pomerleau 1989) has used its learned strategies to drive unassisted at 70 miles per hour for 90 miles on public highways among other cars…"
(Mitchell, 1997)
waymo.com

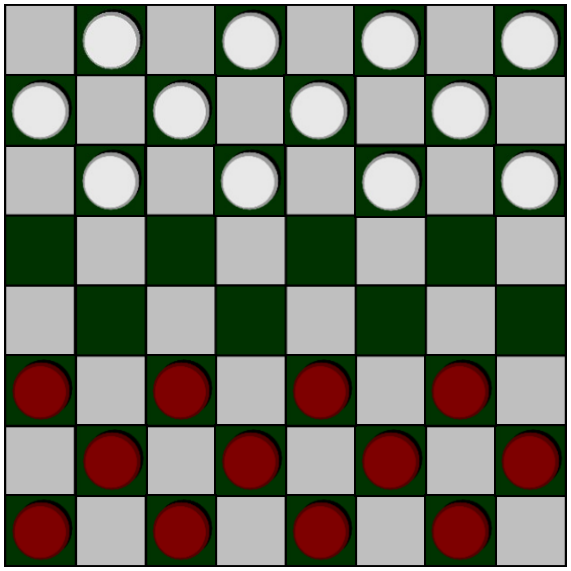

Games / Reasoning
3. Learning to beat the masters at board games
THEN
NOW
"…the world's top computer program for backgammon, TD-GAMMON (Tesauro, 1992, 1995), learned its strategy by playing over one million practice games against itself…"
(Mitchell, 1997)
AlphaGo


Computer Vision
4. Learning to recognize images
THEN
NOW
"…The recognizer is a convolution network that can be spatially replicated. From the network output, a hidden Markov model produces word scores. The entire system is globally trained to minimize word-level errors…"
(LeCun et al., 1995)
Revolution of Depth
Images from https://blog.openai.com/generative-models/


Learning Theory
• 5. In what cases and how well can we learn?
Sample Complexity Results
1. How many examples do we need to learn?
2. How do we quantify our ability to generalize to unseen data?
3. Which algorithms are better suited to specific learning settings?

What is Machine Learning

- Arthur Samuel (1959). Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed



What is Machine Learning

- Arthur Samuel (1959). Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed

- Tom Mitchell (1998) Well-posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E

What is Machine Learning

- "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

- Suppose your email program watches which emails you do or do not mark as spam and, based on that, learns how to better filter spam. What is the task T in this setting?
  - Classifying emails as spam or not spam.
  - Watching you label emails as spam or not spam.
  - The number (or fraction) of emails correctly classified as spam/not spam.
  - None of the above—this is not a machine learning problem.

Capturing the Knowledge of Experts

Solution #1: Expert Systems

- Over 20 years ago, we had rule based systems
- Ask the expert to
  1. Obtain a PhD in Linguistics
  2. Introspect about the structure of their native language
  3. Write down the rules they devise

Give me directions to Starbucks

```
If: "give me directions to X"
Then: directions(here, nearest(X))
```

How do I get to Starbucks?

```
If: "how do i get to X"
Then: directions(here, nearest(X))
```

Where is the nearest Starbucks?

```
If: "where is the nearest X"
Then: directions(here, nearest(X))
```

Capturing the Knowledge of Experts

Solution #2: Annotate Data and Learn

- Experts:
  - **Very good at** answering questions about specific cases
  - **Not very good at** telling **HOW** they do it
- 1990s: So why not just have them tell you what they do on **SPECIFIC CASES** and then let **MACHINE LEARNING** tell you how to come to the same decisions that they did

Capturing the Knowledge of Experts

Solution #2: Annotate Data and Learn

1. Collect raw sentences $\{x_1, \ldots, x_n\}$
2. Experts annotate their meaning $\{y_1, \ldots, y_n\}$

x_1: How do I get to Starbucks?

y_1: `directions(here, nearest(Starbucks))`

x_2: Show me the closest Starbucks

y_2: `map(nearest(Starbucks))`

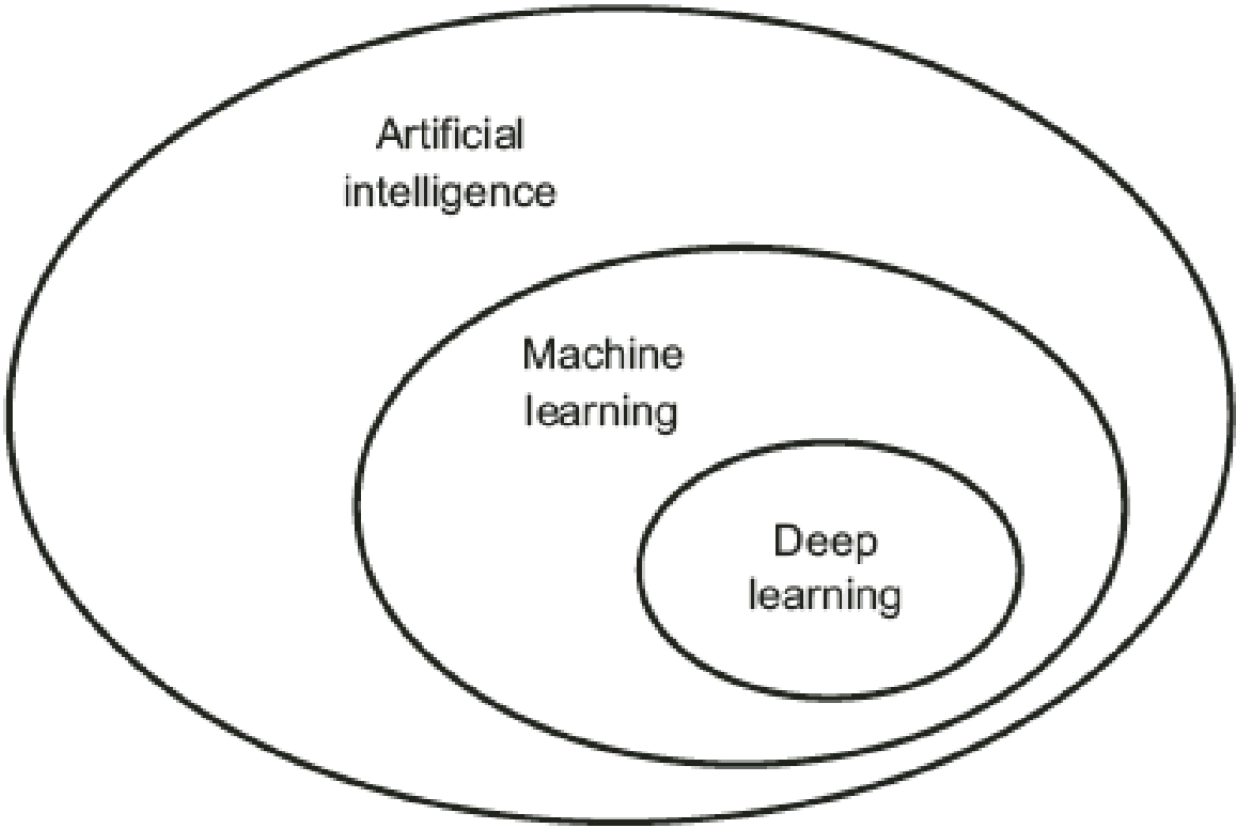
x_3: Send a text to John that I'll be late

y_3: `txtmsg(John, I'll be late)`

x_4: Set an alarm for seven in the morning

y_4: `setalarm(7:00AM)`

What is Deep Learning?

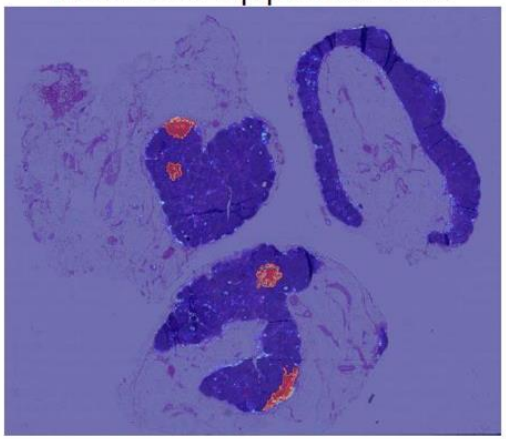


What is Perception (Computer Vision)

SAPIENZA
UNIVERSITÀ DI ROMA

Machine Learning and Perception (Computer Vision)

Medical applications



Robotics

Security

Gaming

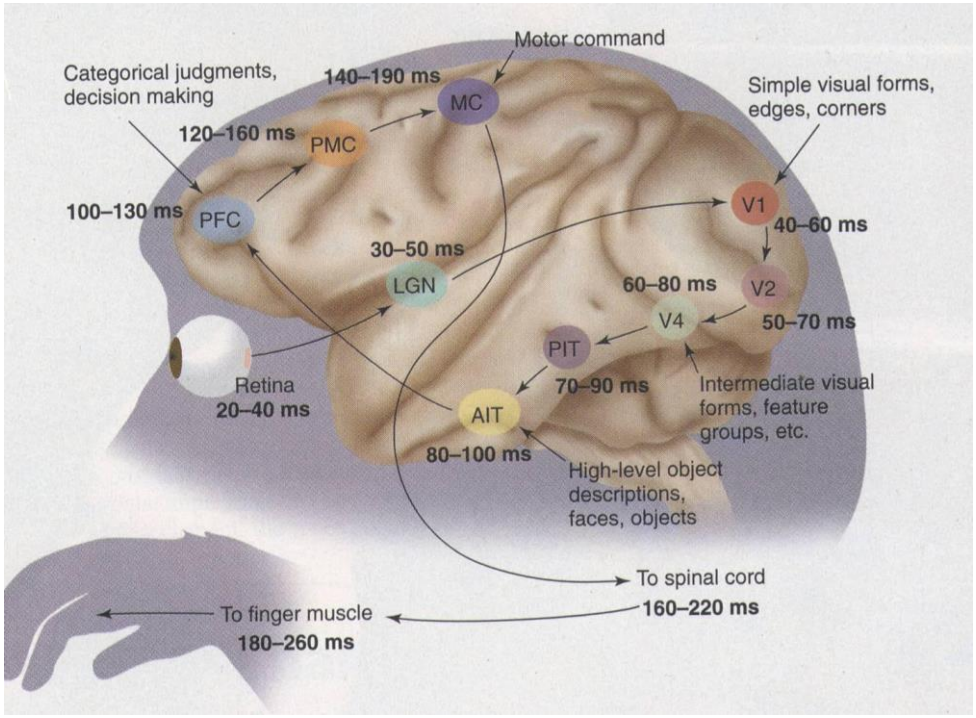
Mobile devices

Driving

Courtesy of A. Torralba, @ICVSS'18

Computer Vision

- Science
  - Foundations of perception. How do *WE* see?
  - computer vision to explore “computational model of human vision”



Computer Vision

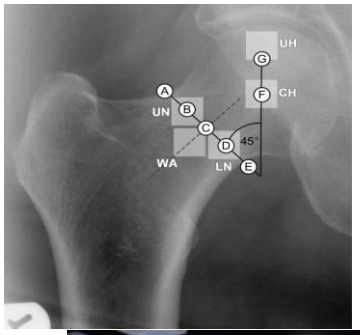
- Science
  - Foundations of perception. How do *WE* see?
  - computer vision to explore “computational model of human vision”
- Engineering
  - How do we build systems that perceive the world
  - computer vision to solve real-world problems: cars to detect pedestrians
- Applications
  - medical imaging (computer vision to support medical diagnosis, visualization)
  - surveillance (to follow/track people at the airport, train-station, ...)
  - entertainment (vision-based interfaces for games)
  - graphics (image-based rendering, vision to support realistic graphics)
  - car-industry (lane-keeping, pre-crash intervention, …)
  - …

Some Applications

- US Post office
  - At the mail processing plant, ***machines*** separate mail by shape and size, and orient them so their addresses are right-side up and facing the same direction
  - ***An optical scanner*** scans the address, and then a bar code representing the specific address is sprayed on the front of the envelope
  - If the scanner can't read the address, the letter is manually sorted

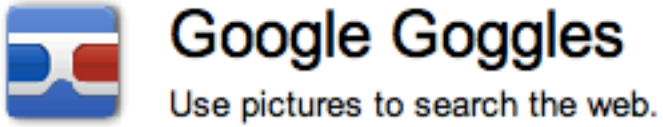
Some Applications

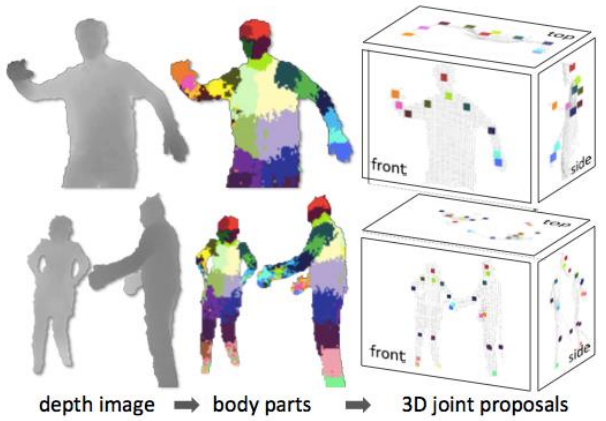
- License Plate Recognition
  - London Congestion Charge
    https://tfl.gov.uk/modes/driving/congestion-charge
- Security/Surveillance
  - Face Recognition
    - Apple's Face ID: chance of 1-in-1-million that a random person could unlock your phone
  - Biometric passport (*aka* e-passport) has an embedded electronic chip which contains biometric information
    - Currently standardized biometrics are facial recognition, fingerprint recognition, and iris recognition
  - Airport Security
    (People Tracking)
- Medical Imaging
  - (Semi-)automatic segmentation and measurements
- Robotics
- Autonomous driving



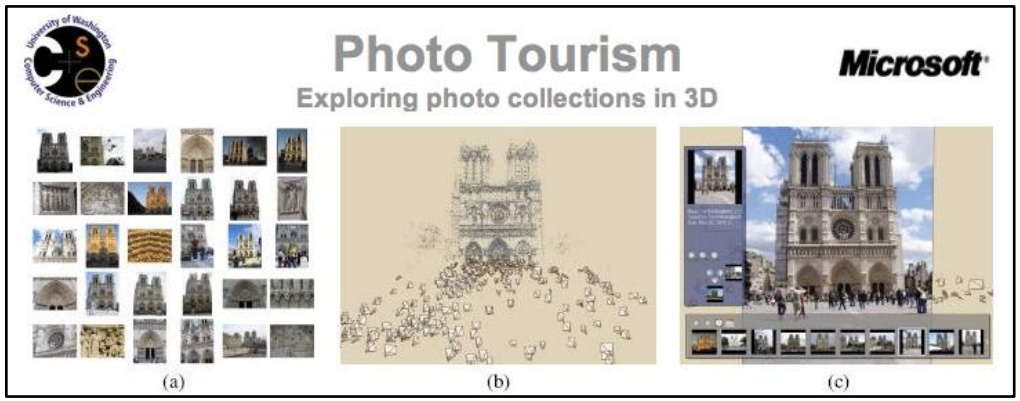
More Applications

- Vision on Cellphones:
  - e.g. Google Goggles
- Vision for Interfaces:
  - e.g. Microsoft Kinect
- Reconstruction





Microsoft



The rise of Deep Learning

Artificial Intelligence in the news

Artificial Intelligence in the news

Scuola Università

In evidenza Criptovalute Spread BTP-Bund FTSE-MIB Petrolio

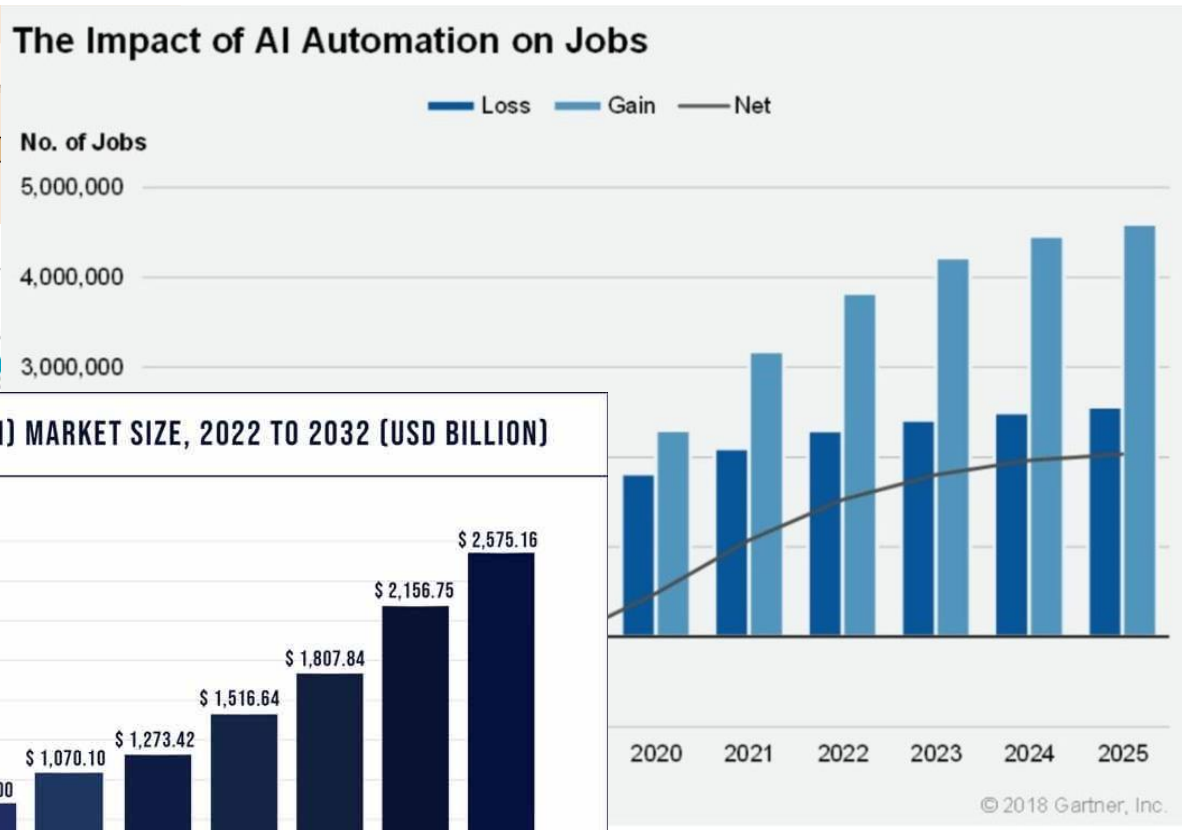
Servizio | I progetti europei

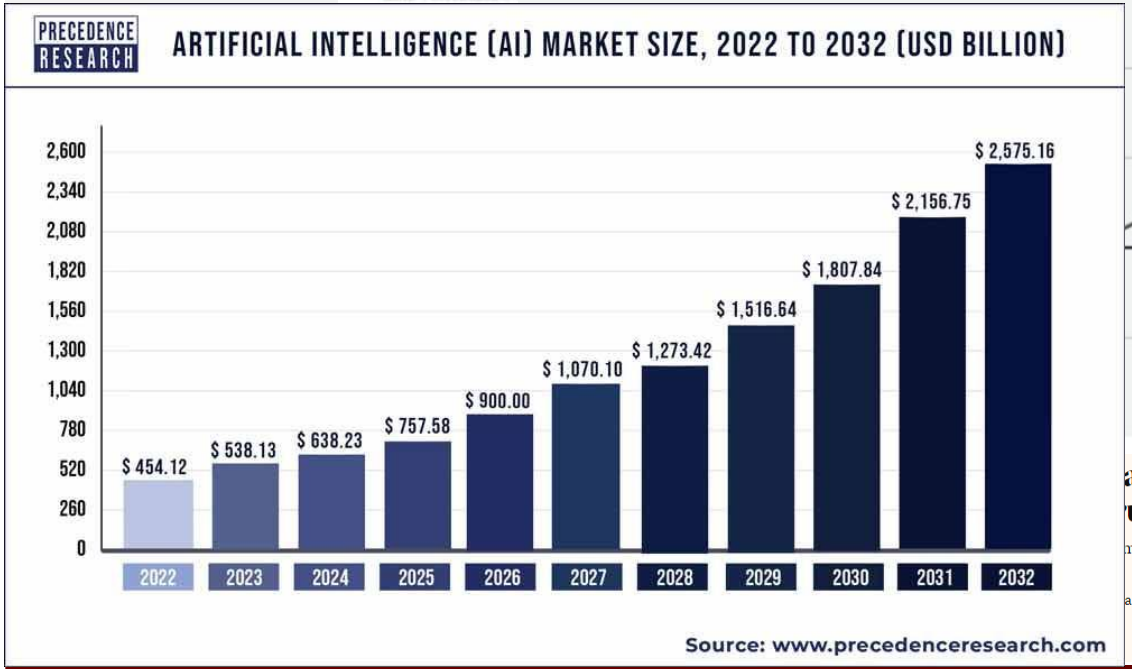
Intelligenza artificiale, oltre mi
progetti di ricerca con fondi Ue

sky tg24 L'APP DI SKY TG24

MONDO

Intelligenza artificiale,



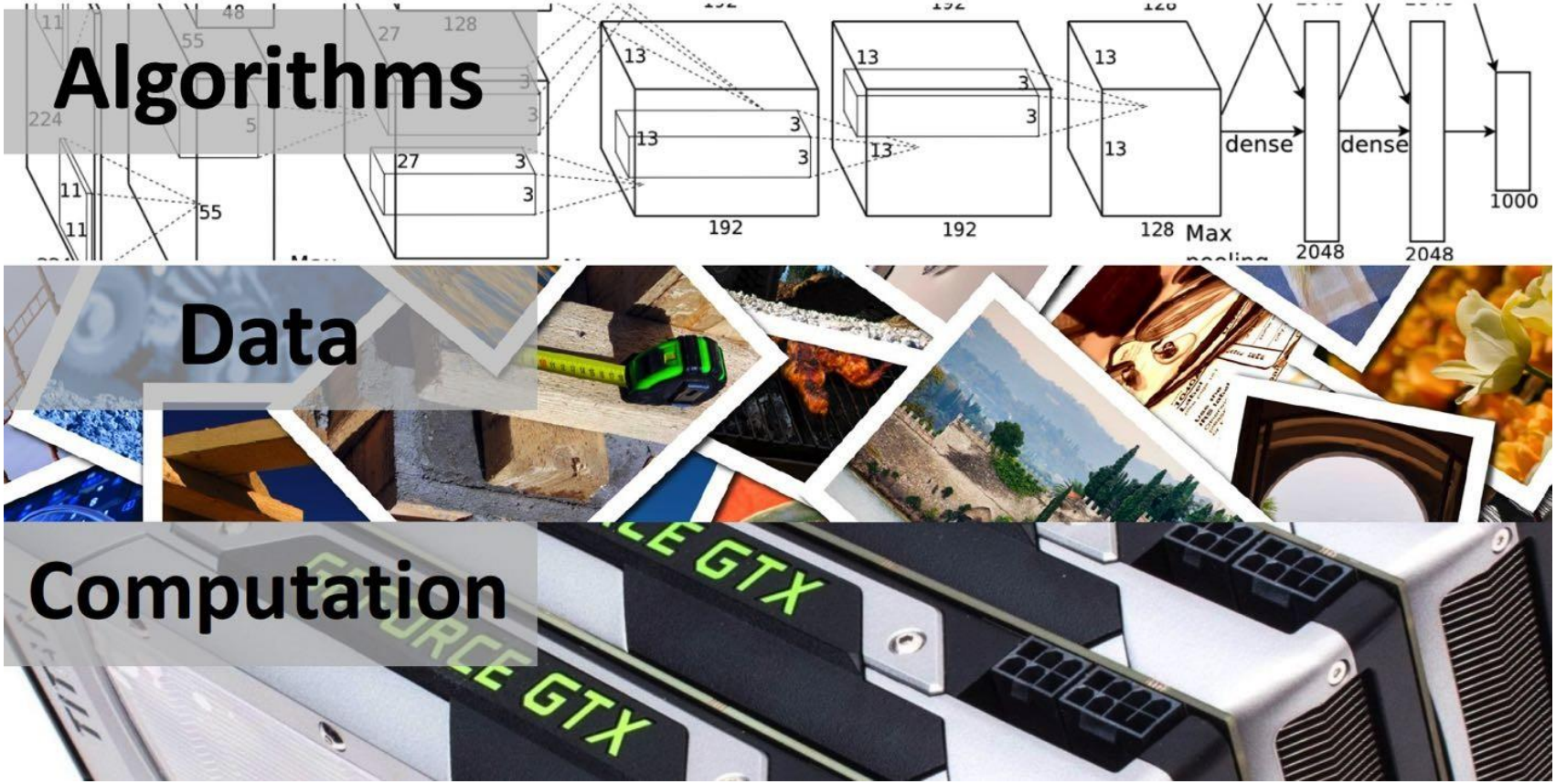


a artificiale. In Borsa vince chi
uttarla bene

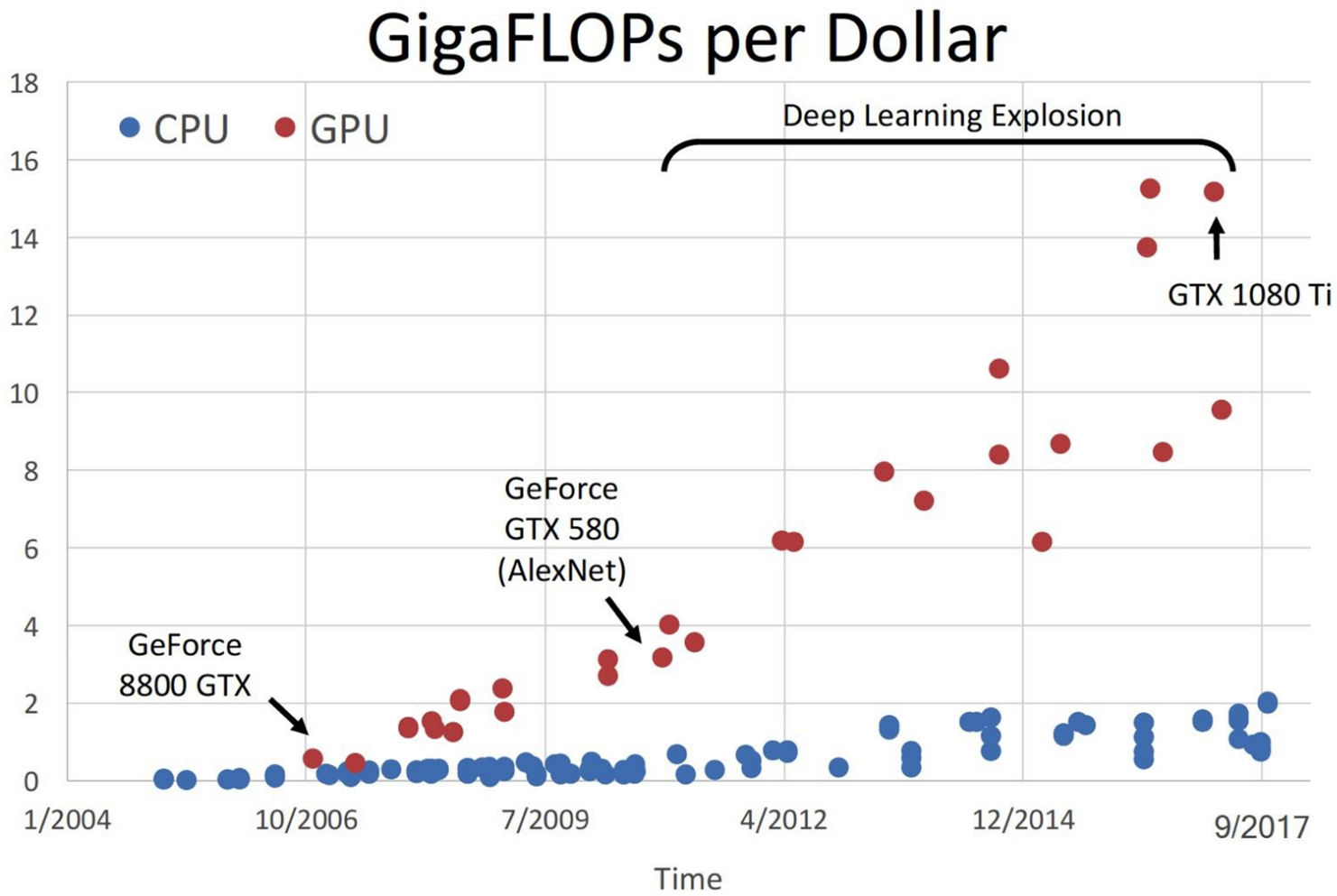
niconduttori opportunità tra gli utenti più tecnologici

aprile 2024 3 min

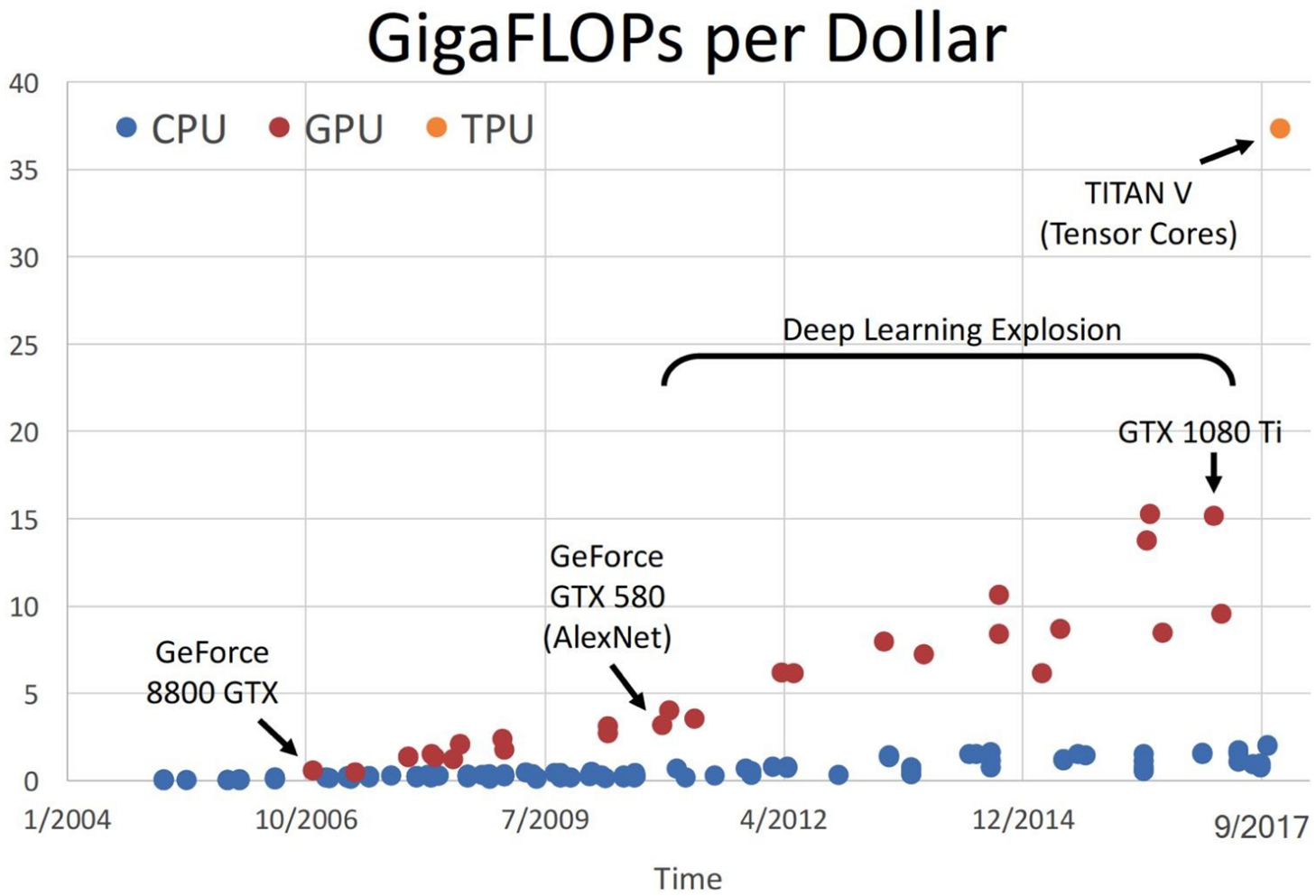
Keys to successes



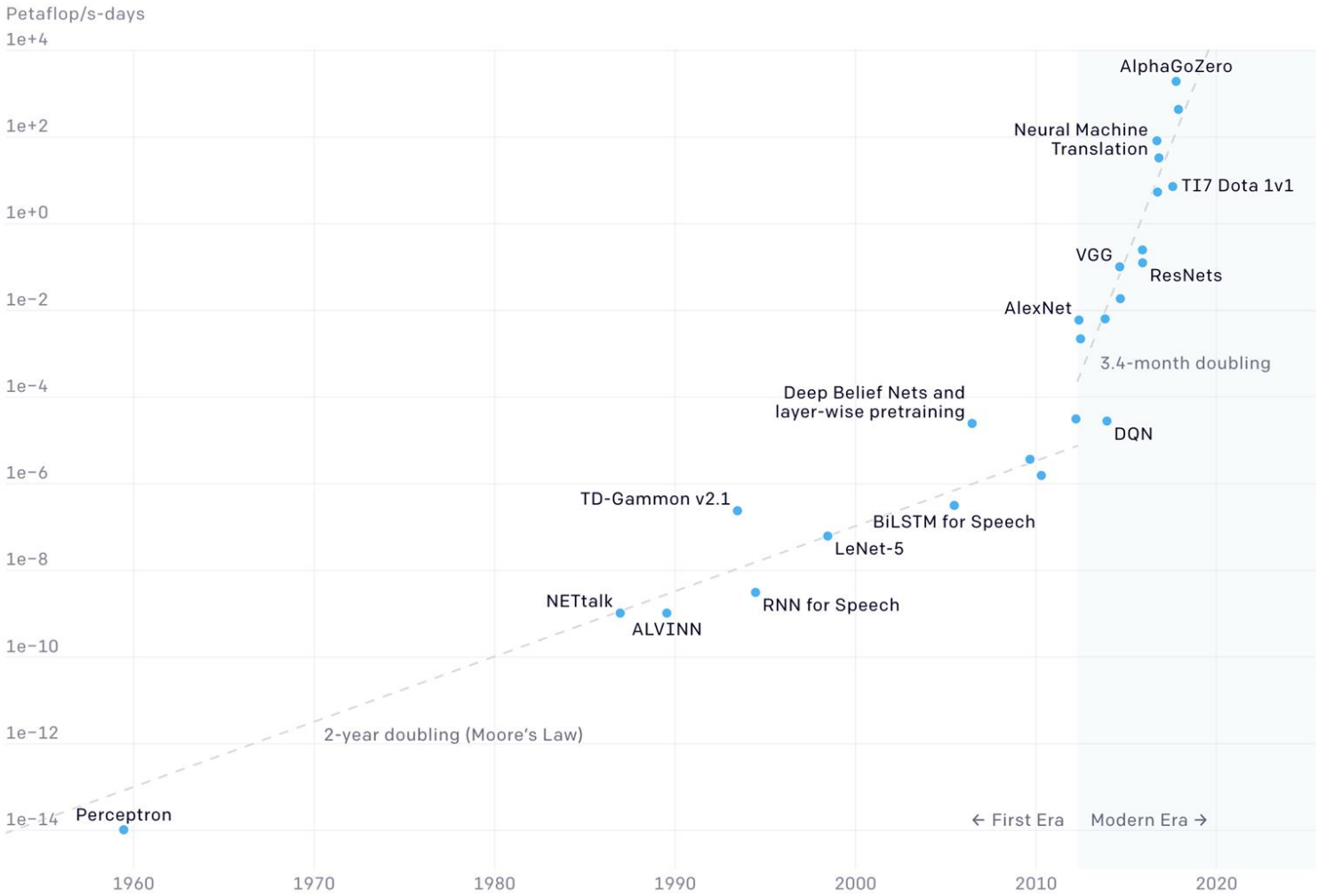
Keys to successes
Computation



Keys to successes
Computation

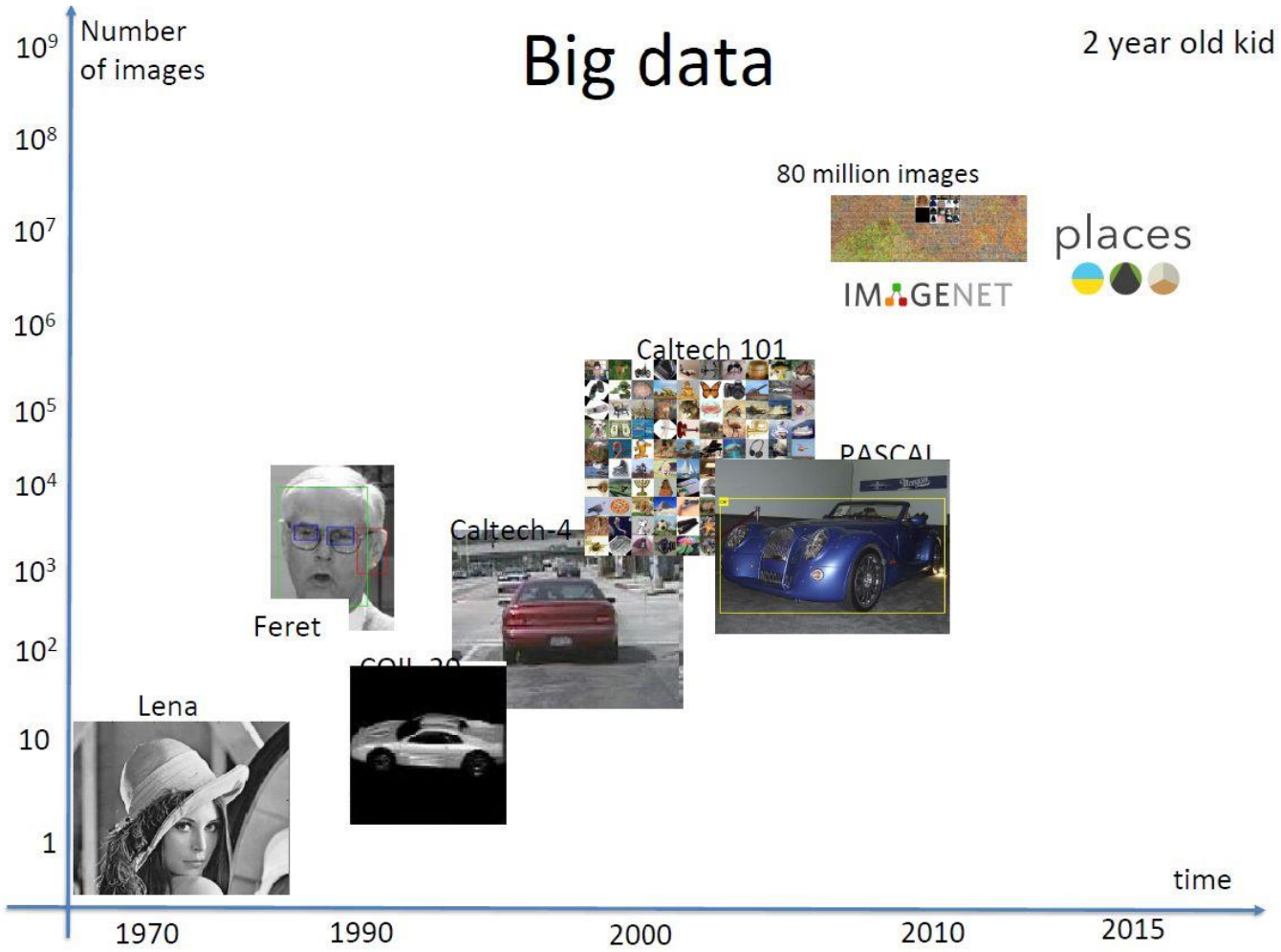


Keys to successes
Computation



Courtesy of OpenAI

Keys to successes
Data



Courtesy of A. Torralba @ICVSS'18

IMAGENET

www.image-net.org

22K categories and **14M** images

- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
- Food
- Materials
- Structures
- Artifact
  - Tools
  - Appliances
  - Structures
- Person
- Scenes
  - Indoor
  - Geological Formations
- Sport Activities

Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009

IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:
1,000 object classes
1,431,167 images

Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

✔

Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle

✗

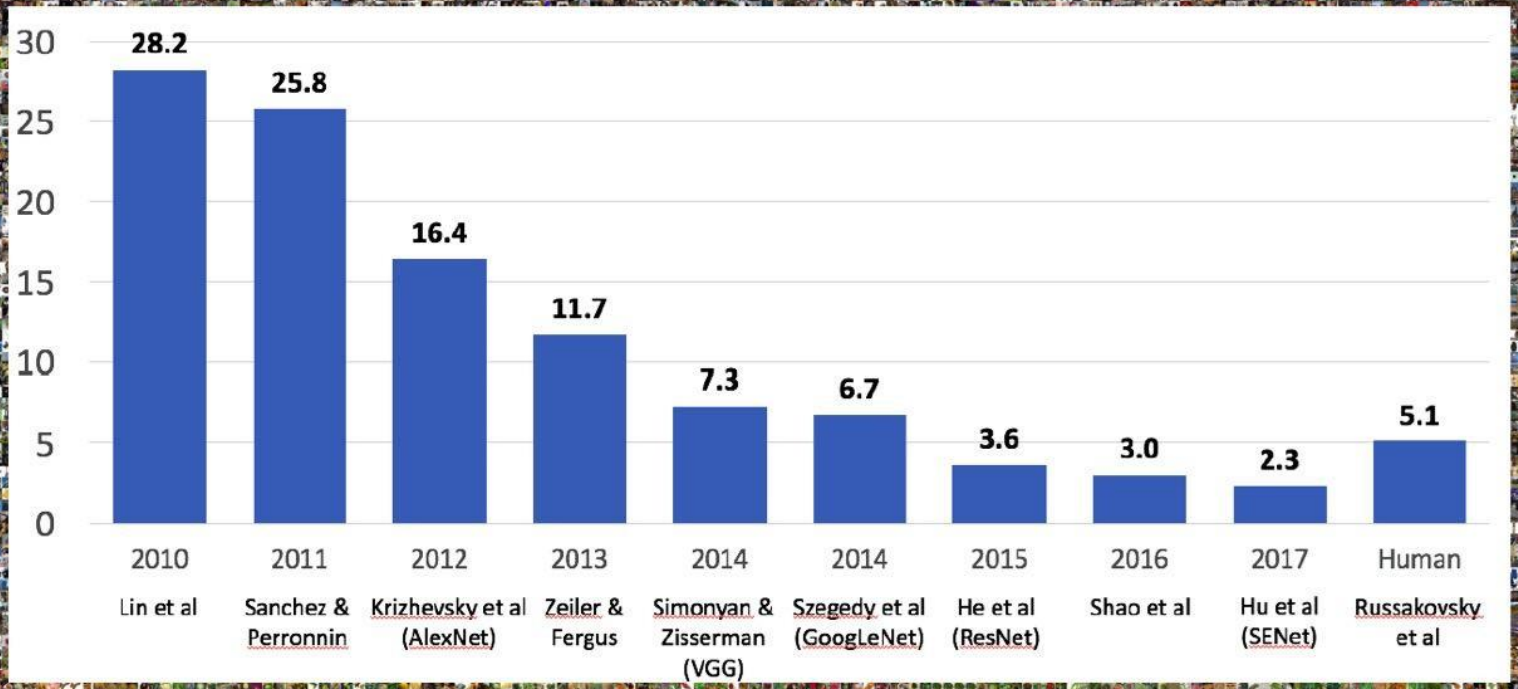
Russakovsky et al. IJCV 2015

IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:

1,000 object classes

1,431,167 images



Russakovsky et al. IJCV 2015

Keys to successes
Algorithms

IMAGENET Large Scale Visual Recognition Challenge

Year 2010

NEC-UIUC

Dense descriptor grid: HOG, LBP

Coding: local coordinate, super-vector

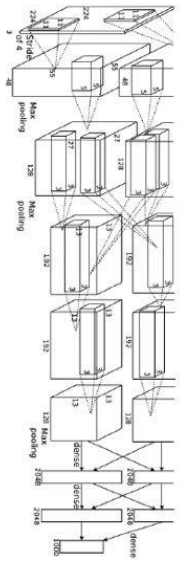
Pooling, SPM

Linear SVM

[Lin CVPR 2011]

Year 2012

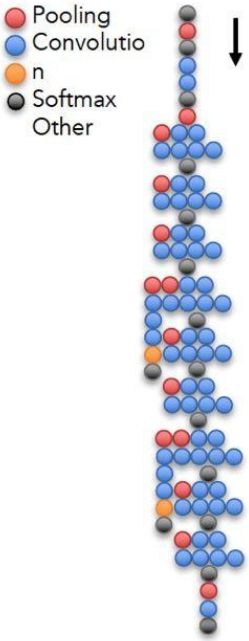
SuperVision



[Krizhevsky NIPS 2012]

Year 2014

GoogLeNet



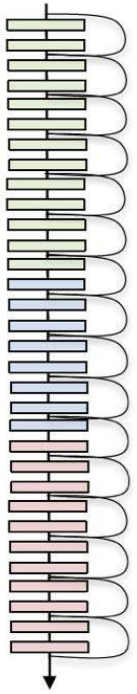
[Szegedy arxiv 2014]

VGG

[Simonyan arxiv 2014]

Year 2015

MSRA



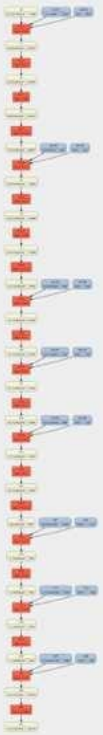
[He ICCV 2015]

How deep is deep enough?

AlexNet (2012)

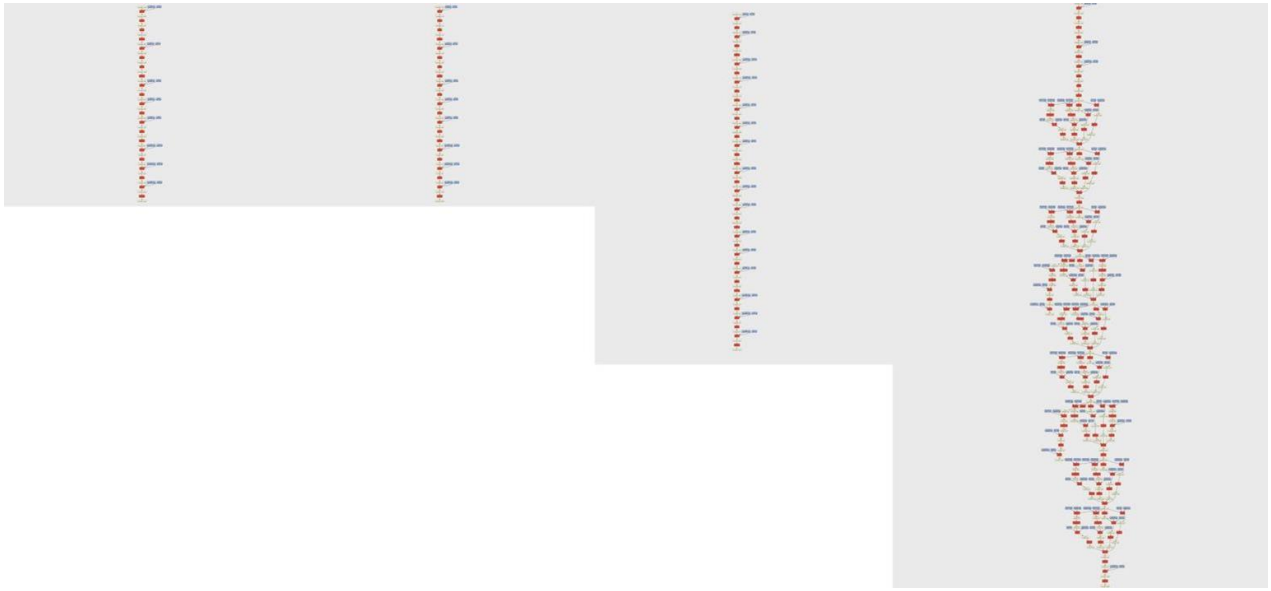
5 convolutional layers

3 fully-connected layers

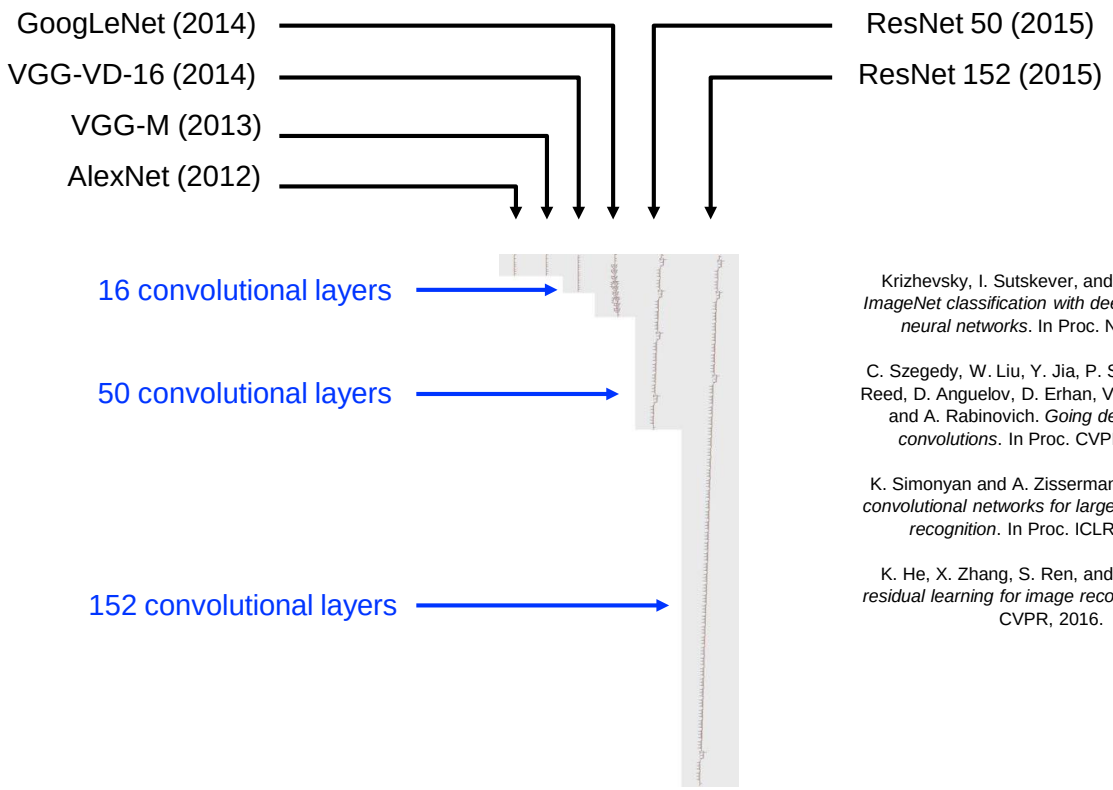


How deep is deep enough?

AlexNet (2012) VGG-M (2013) VGG-VD-16 (2014) GoogLeNet (2014)



How deep is deep enough?



Krizhevsky, I. Sutskever, and G. E. Hinton. *ImageNet classification with deep convolutional neural networks*. In Proc. NIPS, 2012.

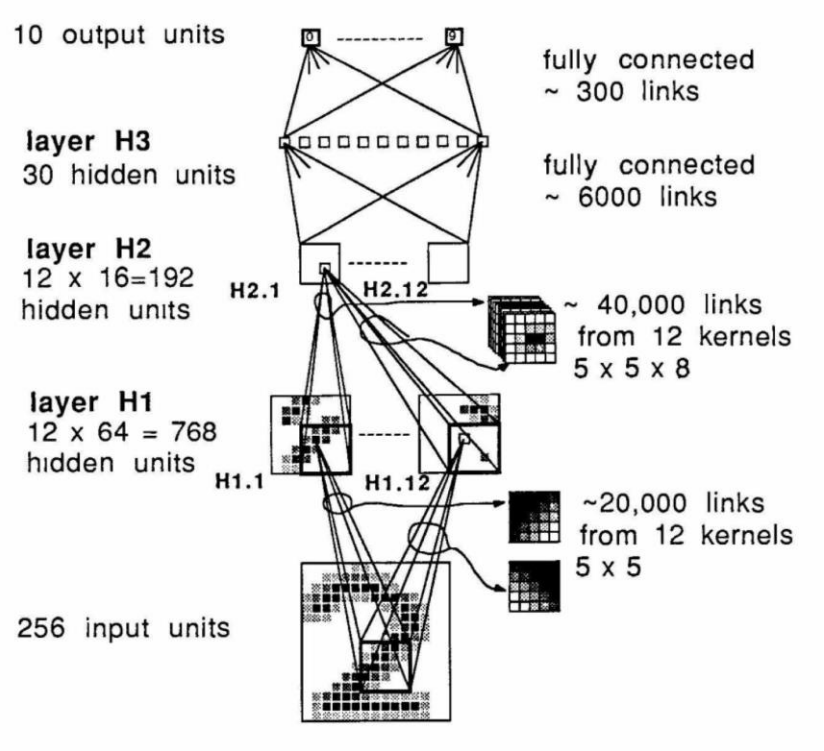
C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. *Going deeper with convolutions*. In Proc. CVPR, 2015.

K. Simonyan and A. Zisserman. *Very deep convolutional networks for large-scale image recognition*. In Proc. ICLR, 2015.

K. He, X. Zhang, S. Ren, and J. Sun. *Deep residual learning for image recognition*. In Proc. CVPR, 2016.

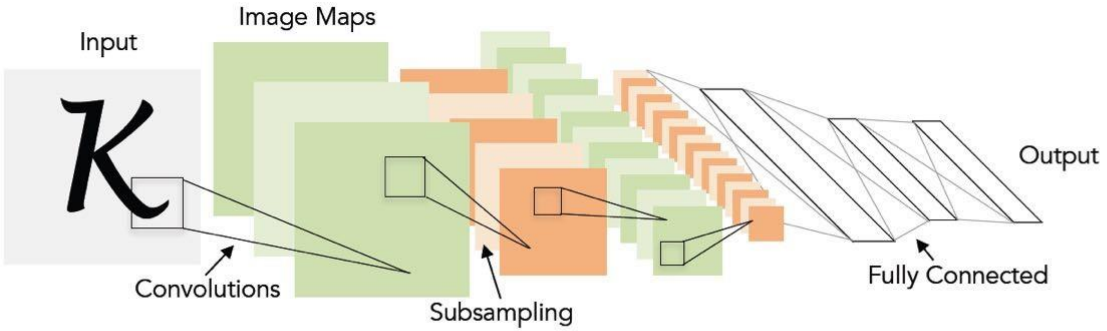
Keys to successes
Algorithms

- Progress in modelling
  - Cognitron/Neocognitron [Fukushima 1971-1982]
  - Pooling [Riesenhuber and Poggio 1999]
  - Convnet's [LeCun et al. 1989]
  - Non-linearities [Nair, Hinton 2010]
  - DropOut [Krizhevsky et al. 2012]
  - Batch Normalization [Ioffe Szegedy 2015]
  - Identity mapping [He et al. 2015]
  - Attention [Bengio et al. 2015]
  - …



Keys to successes

1998
LeCun et al.



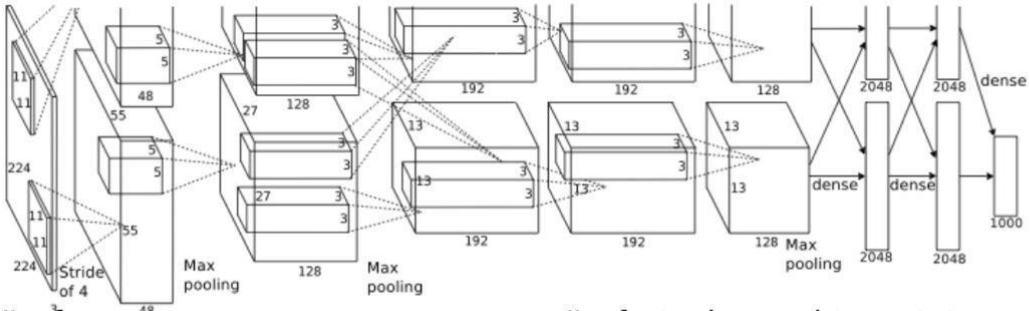
# of transistors

10^6

# of pixels used in training

10^7 NIST

2012
Krizhevsky et al.



# of transistors

10^9

GPUs

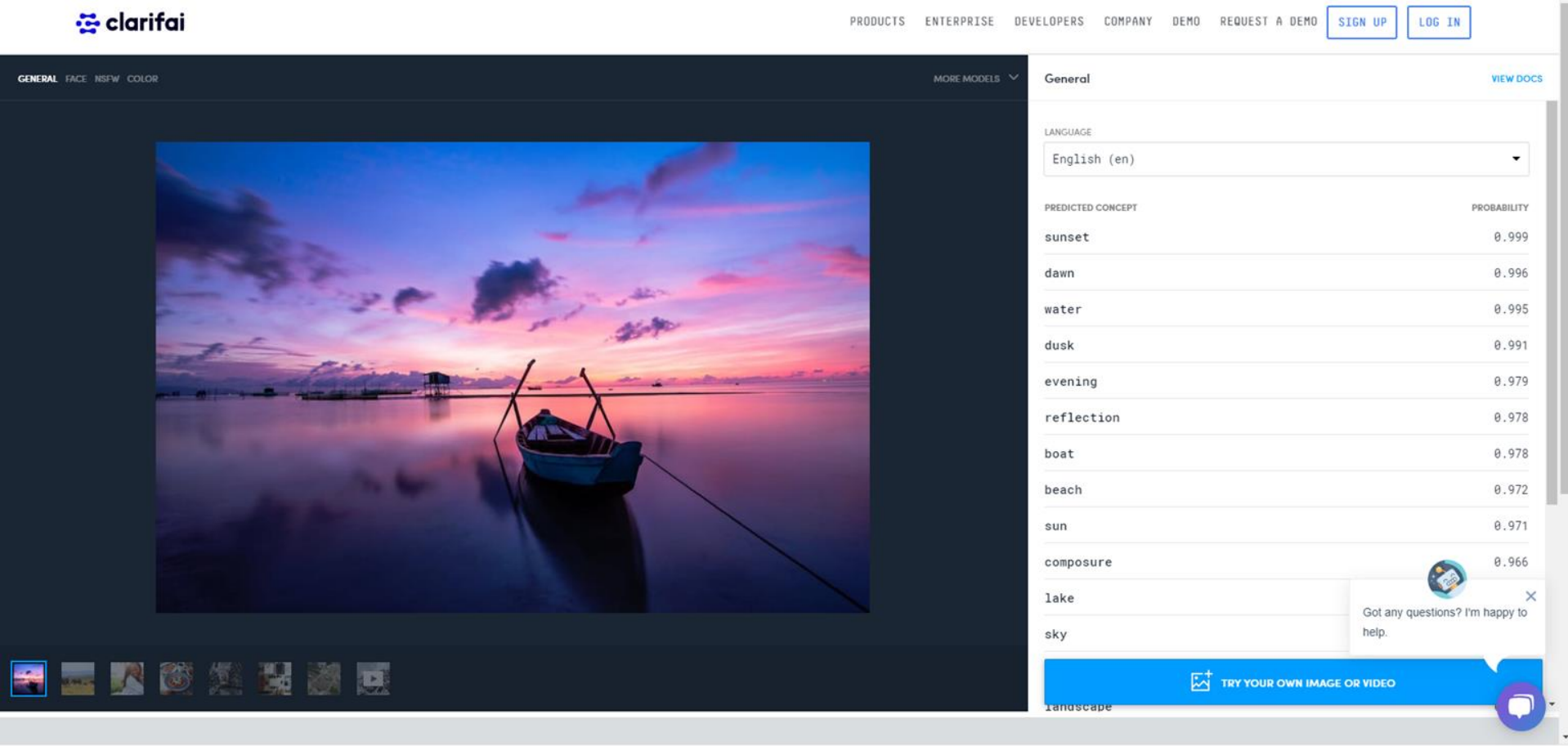
# of pixels used in training

10^{14} IMAGENET

Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

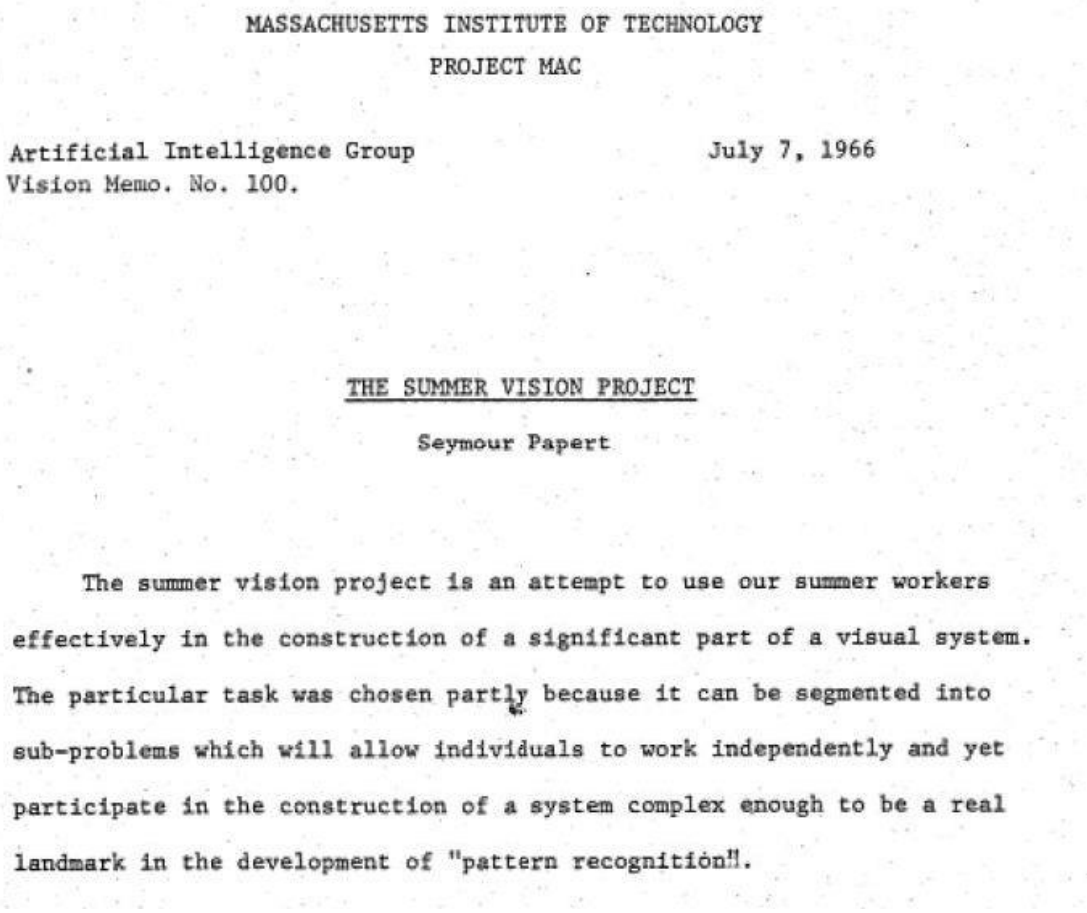
Test it out

- https://www.clarifai.com/models/general-image-recognition



Challenges

- Deceivingly simple, difficult to foresee progress



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group — July 7, 1966
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

AI Hype

- "AI innovations continue to deliver big benefits to business and adoption rates will accelerate in coming years" Gartner (July 2022)

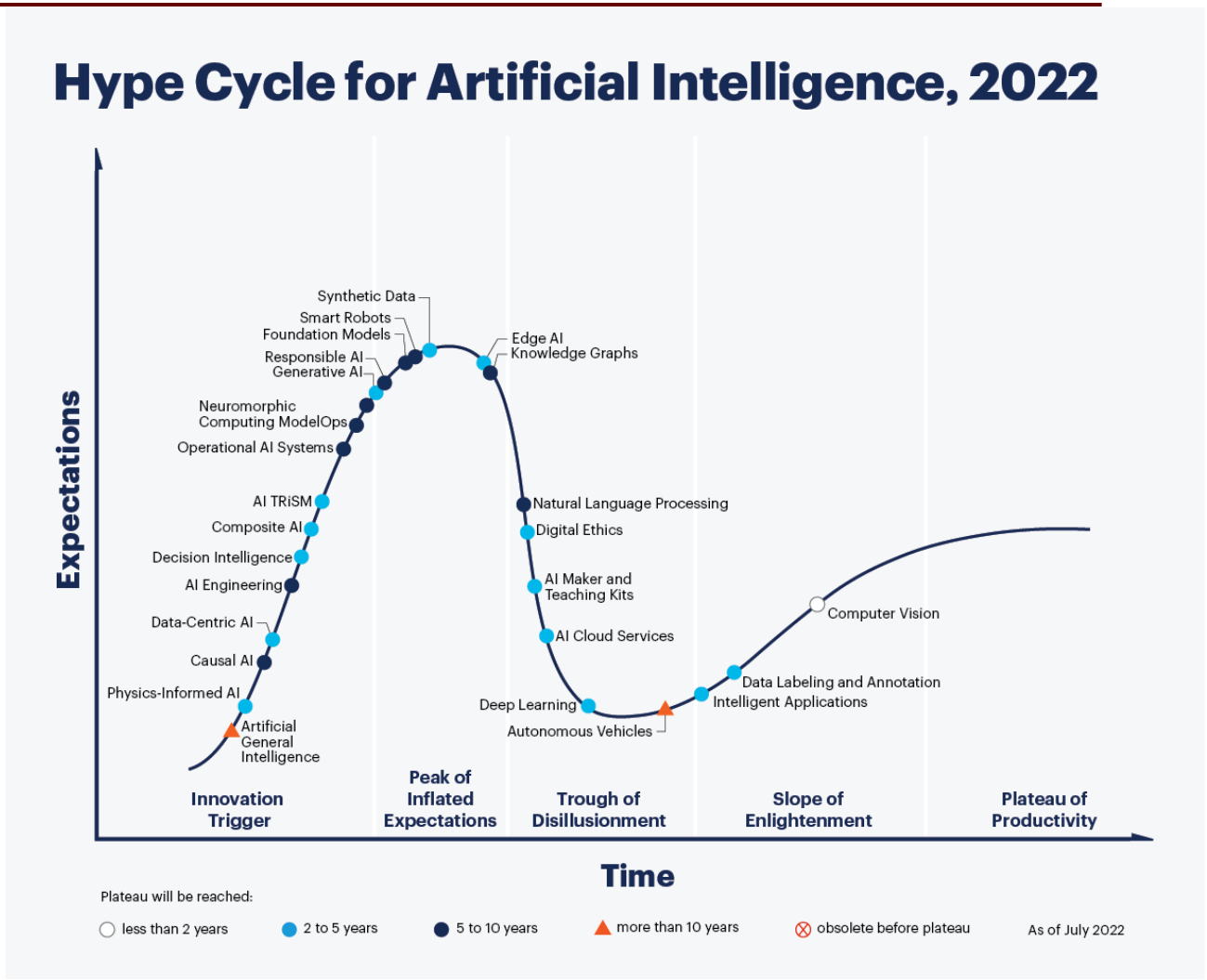


Image credit: https://www.gartner.co.uk/en/articles/what-is-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle

Thank you

Acknowledges: some slides and material from Bernt Schiele, Mario Fritz, Michael Black, Bill Freeman, Fei-Fei, Justin Johnson, Serena Yeung, Matthew R. Gormley

SAPIENZA
UNIVERSITÀ DI ROMA