Introduction to Deep Learning

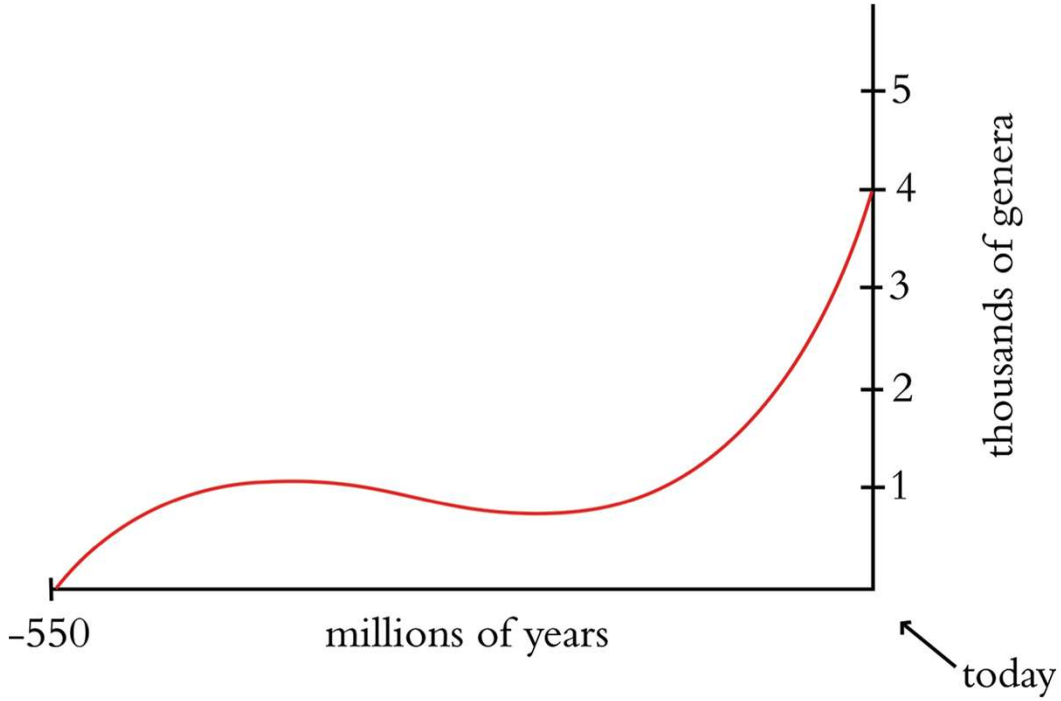
TIN0255/09P9M80/D82 - Aprendizagem Profunda

2026.2 - Prof. Pedro Moura

Biological Neuroanatomy

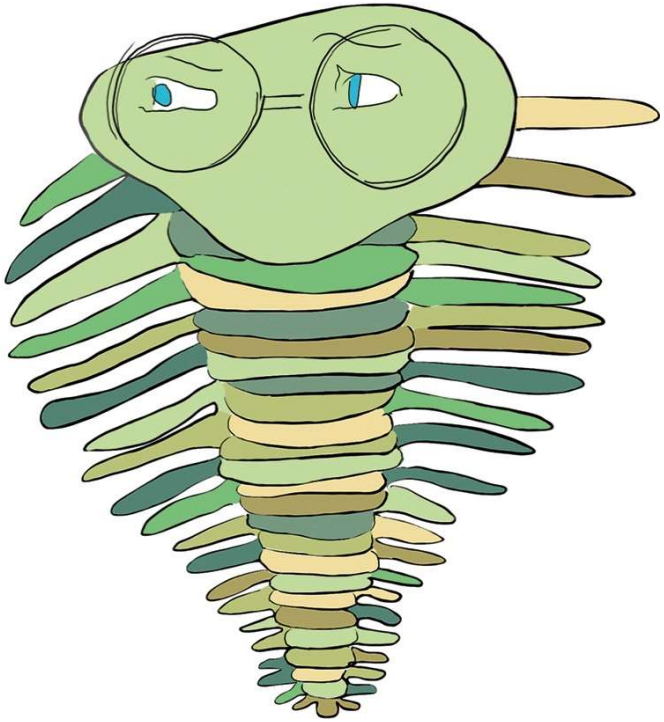
- 550 million years ago, in the prehistoric Cambrian period, the number of species on the planet began to surge.
- Fossil record: explosion was driven by the development of light detectors in the trilobite.
  - Small marine animal ≈ modern crabs

Biological Background



Increase on the number of species during the prehistoric Cambrian period.

Biological Background



A bespectacled trilobite.

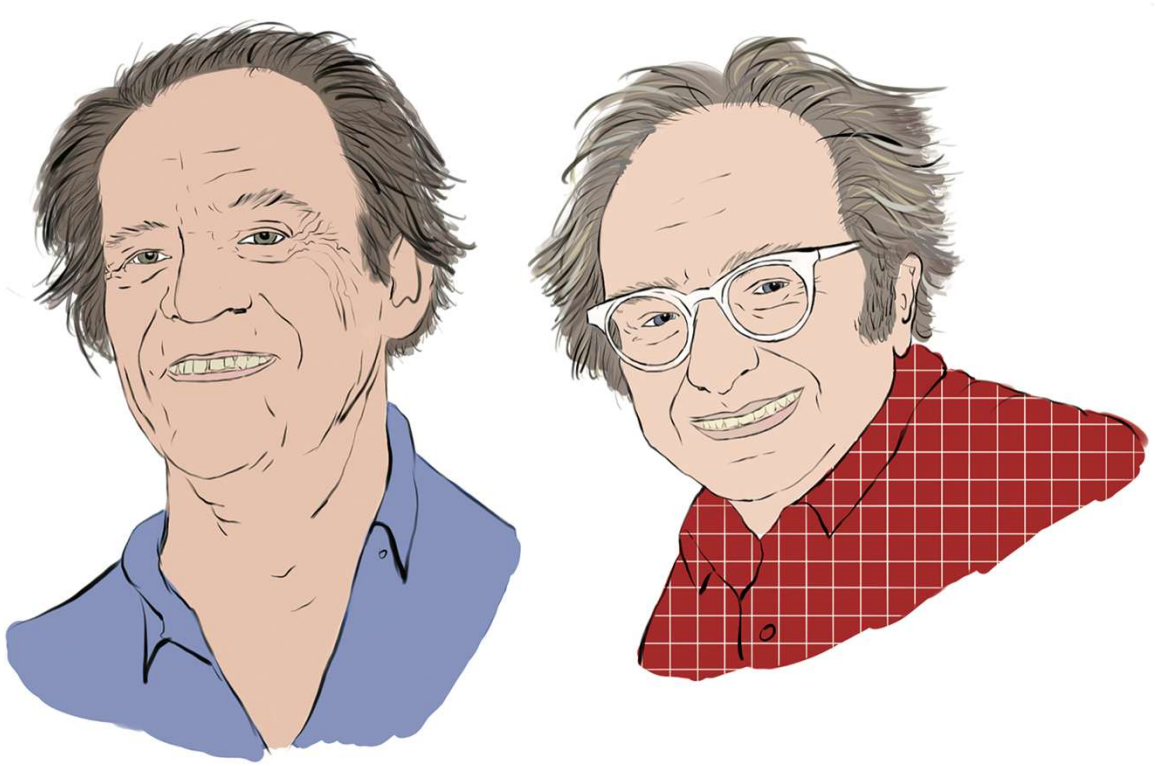
Biological Background

- A visual system, even a primitive one, bestows a delightful set of fresh capabilities.
  - For example, one can spot food, foes and friend-looking mates at some distance.
- Other senses, such as smell, also enable animals to detect as well.
  - But not with the accuracy and light-speed pace of vision.
- Once the trilobite could see: arms race that produced the Cambrian explosion.
  - They had to evolve in order to survive.

Biological Background

- Since then, the complexity of vision has increased considerably.
- In modern mammals, a large portion of the *cerebral cortex* is involved in visual perception.
- In the late 1950s, at John Hopkins University, the physiologists David Hubel and Torsten Wiesel started researching on **how visual information is processed** in the mammalian cerebral cortex.
  - They won a Nobel Prizer many years later.

Biological Background

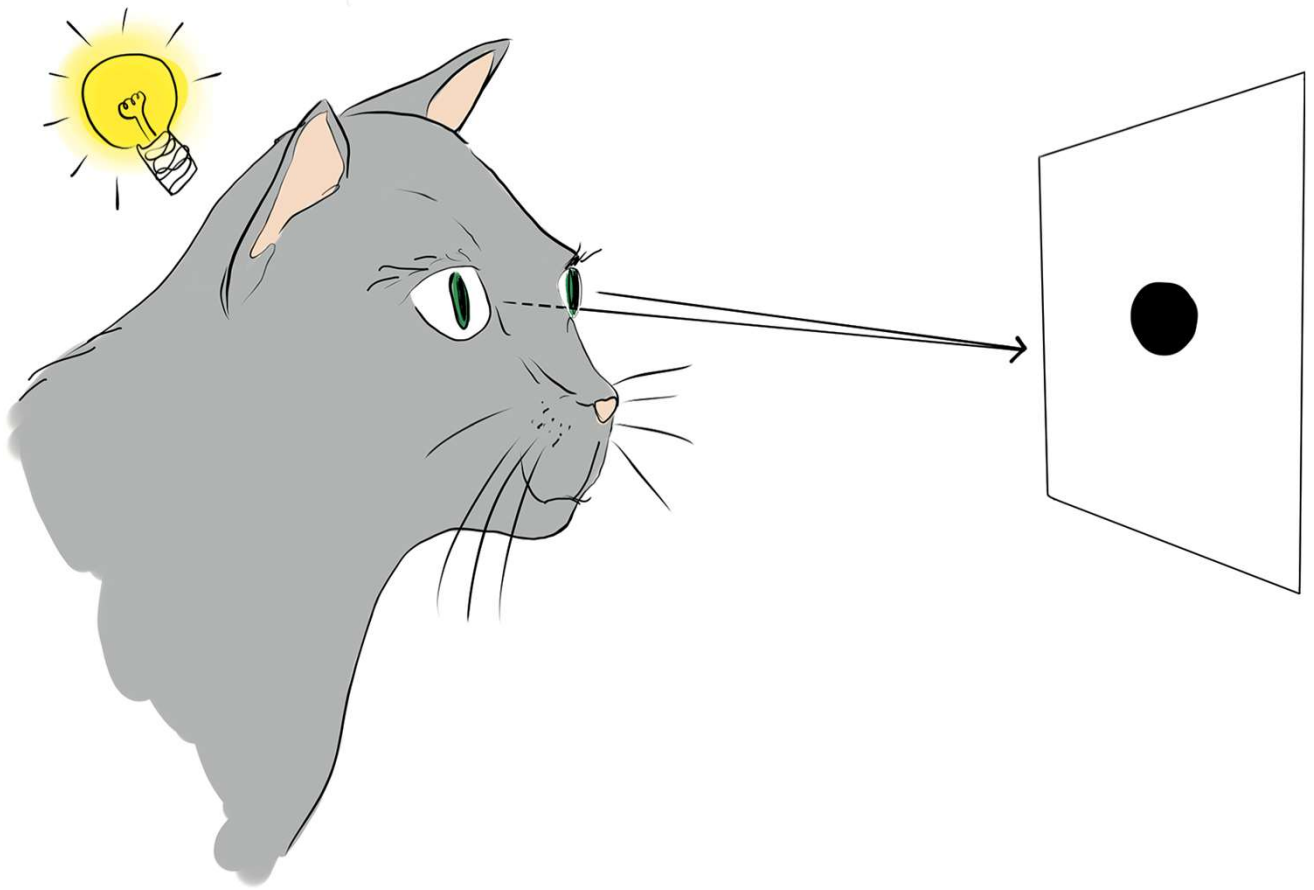


The Nobel Prize-winning neurophysiologists Torsten Wiesel (left) and David Hubel (right).

Biological Background

- They showed images to anesthetized cats while simultaneously recording the activity of **individual neurons** from the primary visual cortex.
  - The first part of the cerebral cortex to receive visual **input** the eyes.
- Projecting slides onto a screen.
  - Started by simple shapes like a dot.
- The cats had electrical recording equipment implanted within their skull.

Biological Background

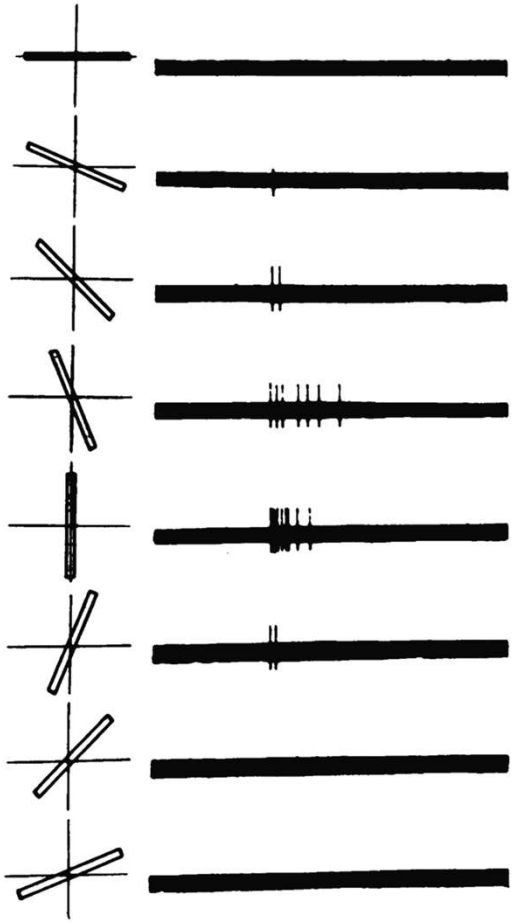


Hubel and Wiesel used a light projector to present slides to anesthetized cats with electrical recording equipment implanted within its skull.

Biological Background

- They tried in vain to stimulate the neurons by jumping and waving their arms in front of the cat.
  - Nothing happened.
- While they were removing slides from the projector, a straight edge elicited their record equipment.
  - A primary visual cortex **neuron** was **firing**.
- A large group of such **simple neurons** together is able to represent all 360 degrees of orientation.
  - They are then called **complex neurons**.
- **Hierarchically** organized layers.

Biological Background

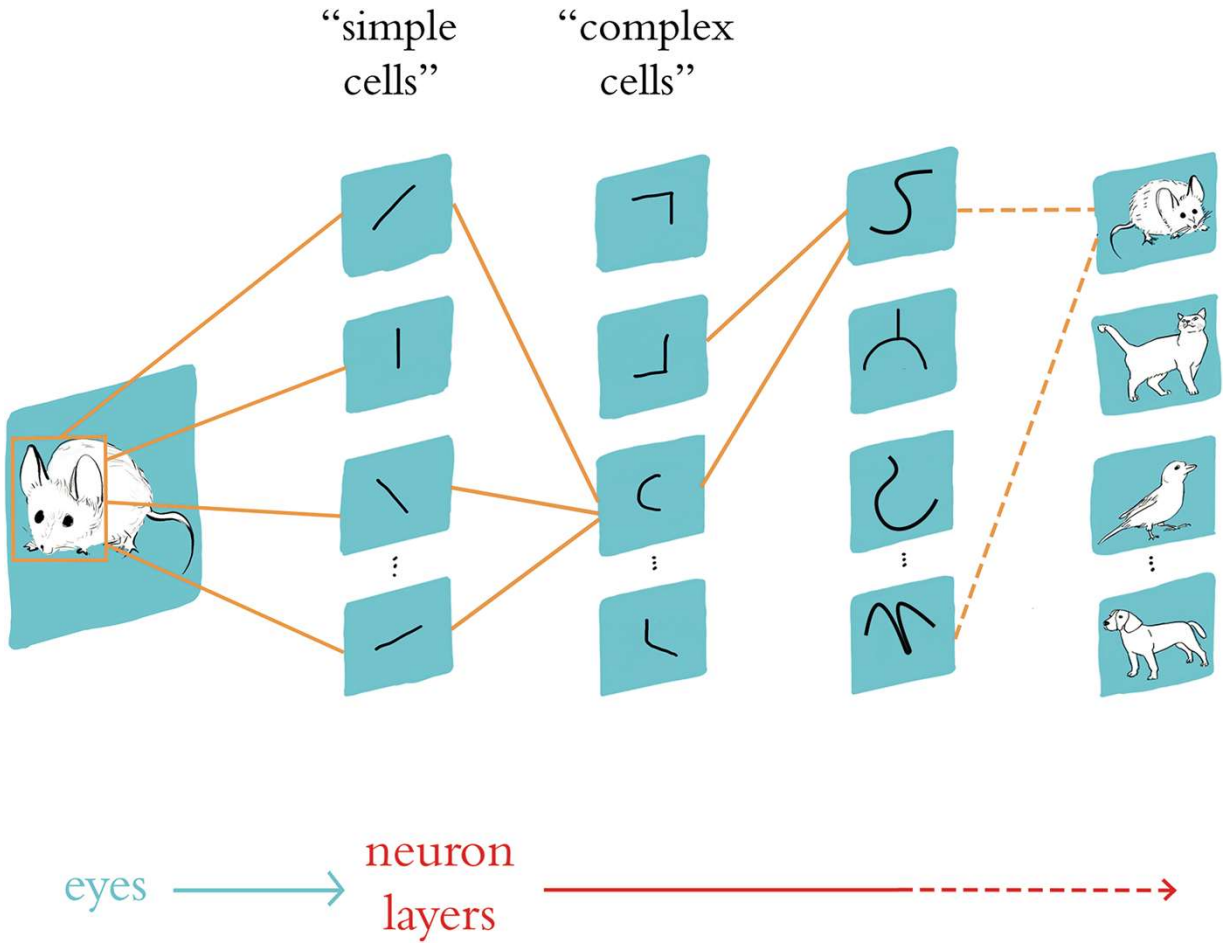


A simple cell in the primary visual cortex of a cat fires at different rates, depending on the orientation of a line shown to the cat. **A vertical line causes the most electrical activity for this particular simple cell.**

Biological Background

- Hierarchically organized **layers of neurons** feeding information into increasingly higher-order neurons.
  - More complex representation.
- Photons of light stimulate neurons located in the retina of each eye.
  - Transmitted from the eyes to the primary visual cortex of the brain.
- In subsequente layers: representation of visual stimuli becomes more complex and abstract.

Biological Background

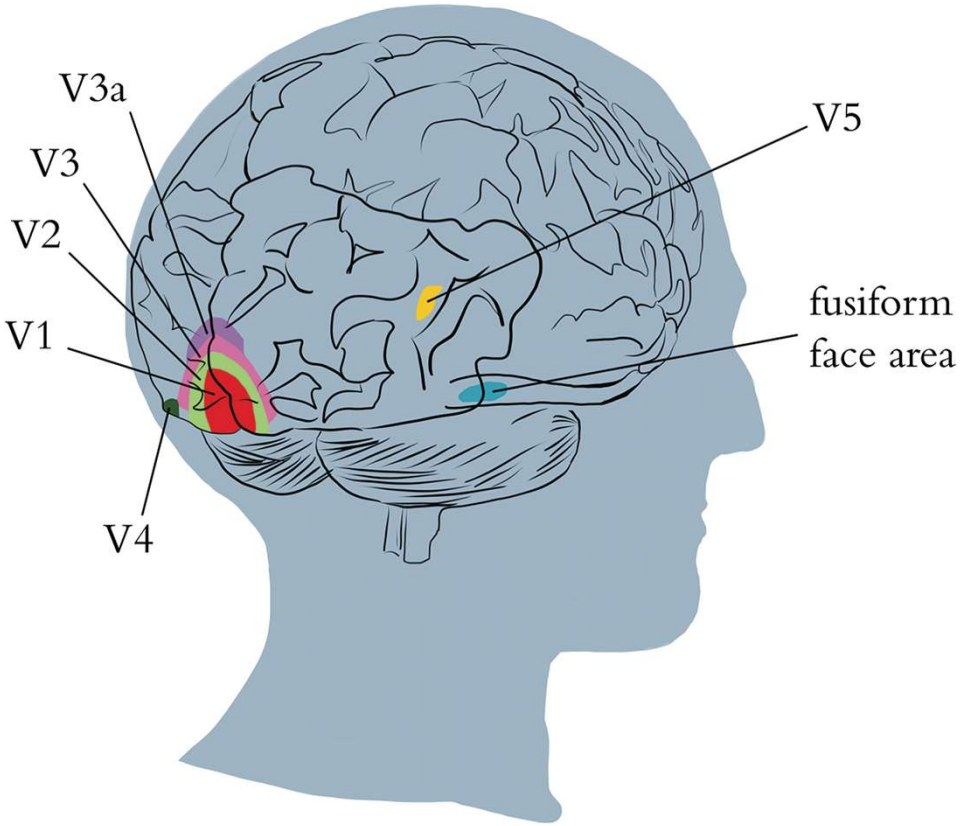


Layers of biological neurons representing visual information in the brain.

Biological Background

- Recordings from the cortical neurons of brainsurgery patients as well as noninvasive techniques like Magnetic Resonance Imaging.
  - High-resolution map of regions to process particular visual stimuli.

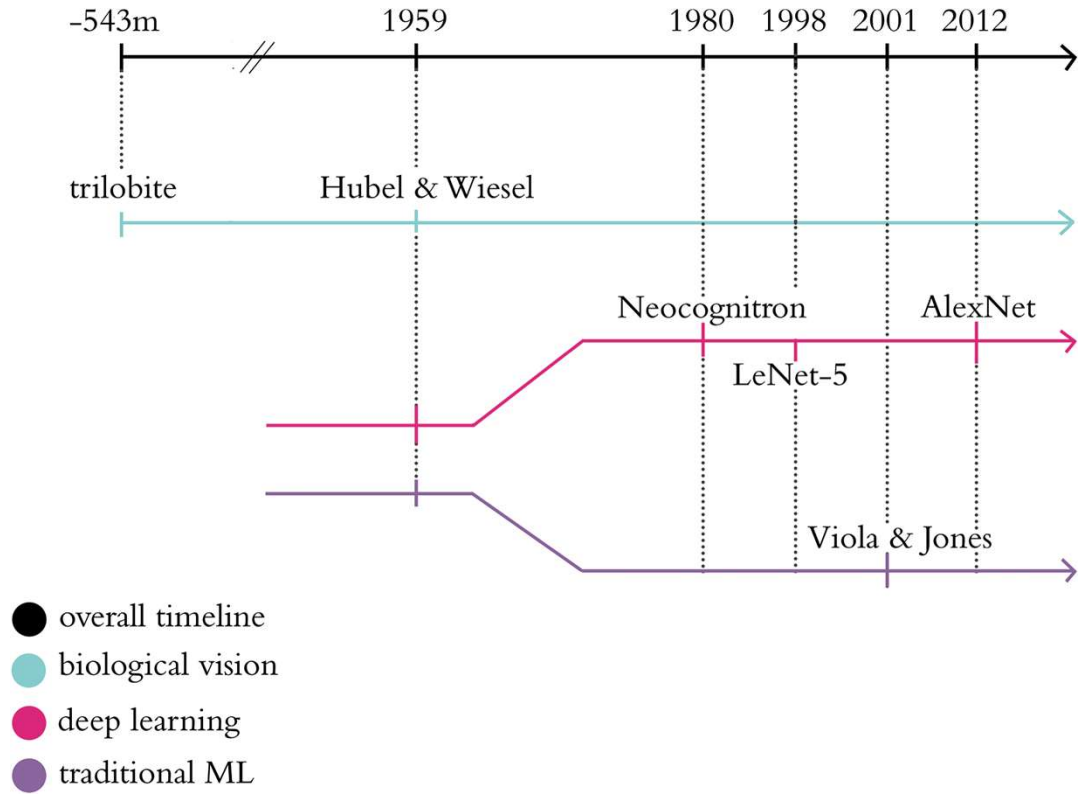
Biological Background



The V1 region receives input from the eyes and contains the simple cells that detect edge orientations. Through the recombination of information via myriad subsequent layers of neurons (including within the V2, V3, and V3a regions), increasingly abstract visual stimuli are represented. Specializations: detection of color (V4), motion (V5), and people's faces (fusiform face area).

Machine Vision

- Biological visual system serves as inspiration for the **modern Deep Learning** approaches to **Machine Vision**.



Abridged timeline of biological and machine vision.

The Neocognitron

- Inspired by Hubel and Wiesel's discovery, in the late 1970s the Japanese electrical engineer Kunihiko Fukushima proposed an architecture for machine vision: **Neocognitron**.
  - He explicitly referred to Hubel and Wiesel's work, citing three papers of them.

- Arranged artificial neurons in hierarchical manner, in which neurons represent line orientations in the cells of they layers closest to the raw image.

- Successively deeper layers represent successively complex and abstract objects.

- **It was able to identify handwritten digits.**

LeNet-5

- Later, Yann LeCun and Yoshua Bengio proposed the **LeNet-5** (1998) model, which achieved better accuracy and efficiency.
  - Built on both Fukushima's lead and the biological inspiration uncovered by Hubel and Wiesel.

- They benefited from **superior data** for training their model, **faster processing power** and, critically, the **backpropagation algorithm**.

- **Backpropagation** facilitated efficient learning throughout the layers of artificial neurons within a deep learning model.
  - Together with **researcher's data** and **processing power**, backprop rendered LeNet-5 sufficiently reliable to become an early commercial application of deep learning.

LeNet-5

Paris-born Yann LeCun is one of the preeminent figures in artificial neural network and deep learning research.

LeNet-5

- In his book, Yann LeCun says that his interesting to Artificial Neural Networks started after attending a Cerisy colloquium on the innate and the acquired, in 1980.

- The linguist Noam Chomsky asserted **it exists in the brain pre-established structures** that make it possible to learn to speak.

- The developmental psychologist Jean Piaget defended **that everything is learnable**, including part of the structures and that the **acquisition of language is done in stages**, as intelligence builds.
  - Intelligence would then be the result of learning based on exchanges with the outside world.

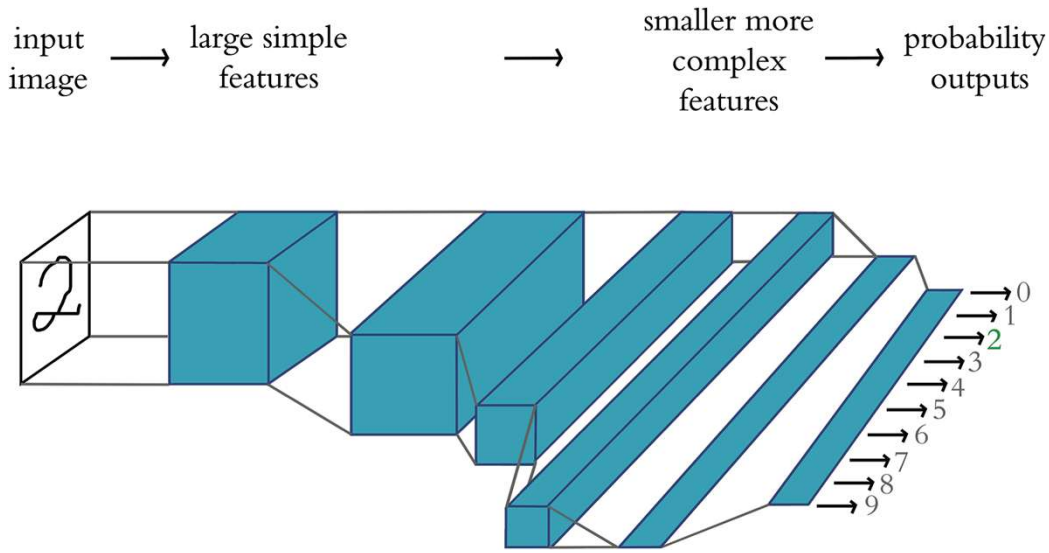
LeNet-5

- He says that this idea seduced him and he then wondered how this idea could be applied to machines!
- Some relevant researchers took part in the discussion, such as Seymour Papert (one of the **creators of Artificial Intelligence**).
- Seymour Papert describes the **Perceptron** as a simple machine capable of learning complex tasks.
- **LeCun then discovered about the existence of this learning machine and became fascinated about it.**
  - He was later surprised that the book that put an end to researches on perceptron was written by Seymour Papert.

LeNet-5

Yoshua Bengio is another of the leading characters in artificial neural networks and deep learning. Born in France, he is a computer science professor at the University of Montreal and codirects the renowned Machines and Brains program at the Canadian Institute for Advanced Research.

LeNet-5

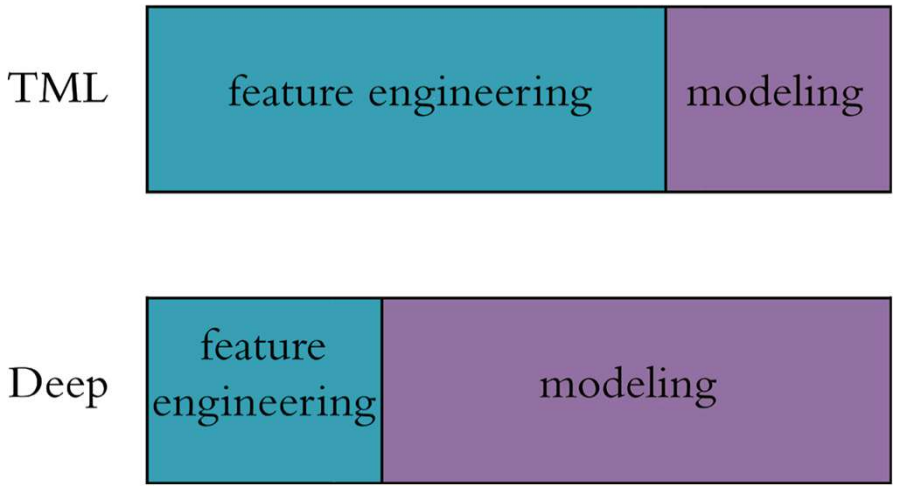


LeNet-5 retains the hierarchical architecture uncovered in the primary visual cortex by Hubel and Wiesel and leveraged by Fukushima in his neocognitron. The leftmost layer represents simple edges, while successive layers represent increasingly complex features.

LeNet-5

- It became an early commercial application of Deep Learning: it was used by the United States Postal Service to automate the **reading of ZIP codes**.
- It did not need to include any expertise about handwritten digits in their code.
- **Feature engineering:** practitioners investing the bulk of their efforts into engineering features.
- **Application of clever, and often elaborate, algorithms to raw data in order to preprocess the data into input variables that can be readily modeled by traditional statistical techniques.**

LeNet-5



Traditional Machine Learning vs Deep Learning.

The Traditional Machine Learning Approach

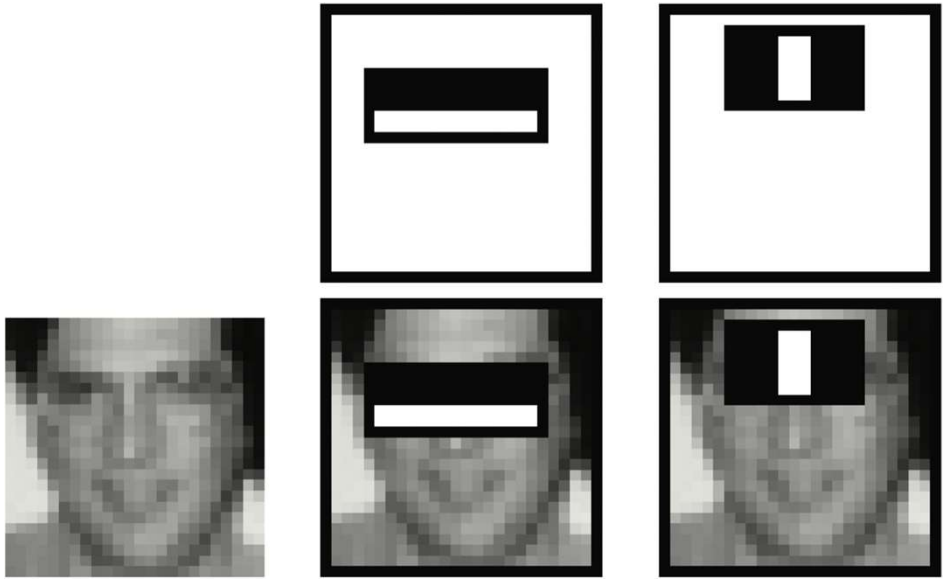
- Following LeNet-5 → research into artificial neural networks, including deep learning, **fell out of favor**.
- Consensus: approach's automated feature generation was not pragmatic.
  - Even though it worked well for handwritten character recognition, the **feature-free ideology** was perceived to have limited breadth of applicability.
- Traditional machine learning, including its feature engineering, appeared to hold more promise, and **funding shifted away from deep learning research**.

The Traditional Machine Learning Approach

- To make clear what feature engineering, let's take a look at Viola-Jones algorithm, proposed by Paul Viola and Michael Jones in the early 2000s.

- Viola and Jones employed rectangular filters such as the vertical or horizontal black-and-white bars shown in the Figure in the next slide.

Machine Vision



Engineered features leveraged by Viola and Jones (2001) to detect faces reliably.

The Traditional Machine Learning Approach

- Features generated by passing these filters over an image can be fed into machine learning algorithms to reliably detect the presence of a face.

- **The algorithm was efficient enough to be the first real-time face detector outside the realm of biology.**

The Traditional Machine Learning Approach

- Devising clever face-detecting filters to process raw pixels into features for input into a machine learning model was accomplished **via years of research and collaboration** on the **characteristics of faces**.

- However, it is limited to detecting faces in general, as opposed to being able to recognize a particular face as, say, Angela Merkel's or Oprah Winfrey's.

- What if we wanted to develop features for detecting Oprah in particular, or for detecting some non-face class of objects like houses, cars, or Yorkshire Terriers?
  - It would neeed development of expertise in that category, which could take years of academic community collaboration to execute both **efficiently** and **accurately**.

- **How to cope with all that time and effort?**

Before the Deep Learning era ...

- There was a lot of prejudice with respect to neural networks (until the results of 2012).
- Researchers were not able to published papers dealing with neural networks, neither obtain grants for their reseaches.
- **General scientific consensus: neural networks didn't provide good results and researchers/professionals should not use or research them.**

Before the Deep Learning era ...

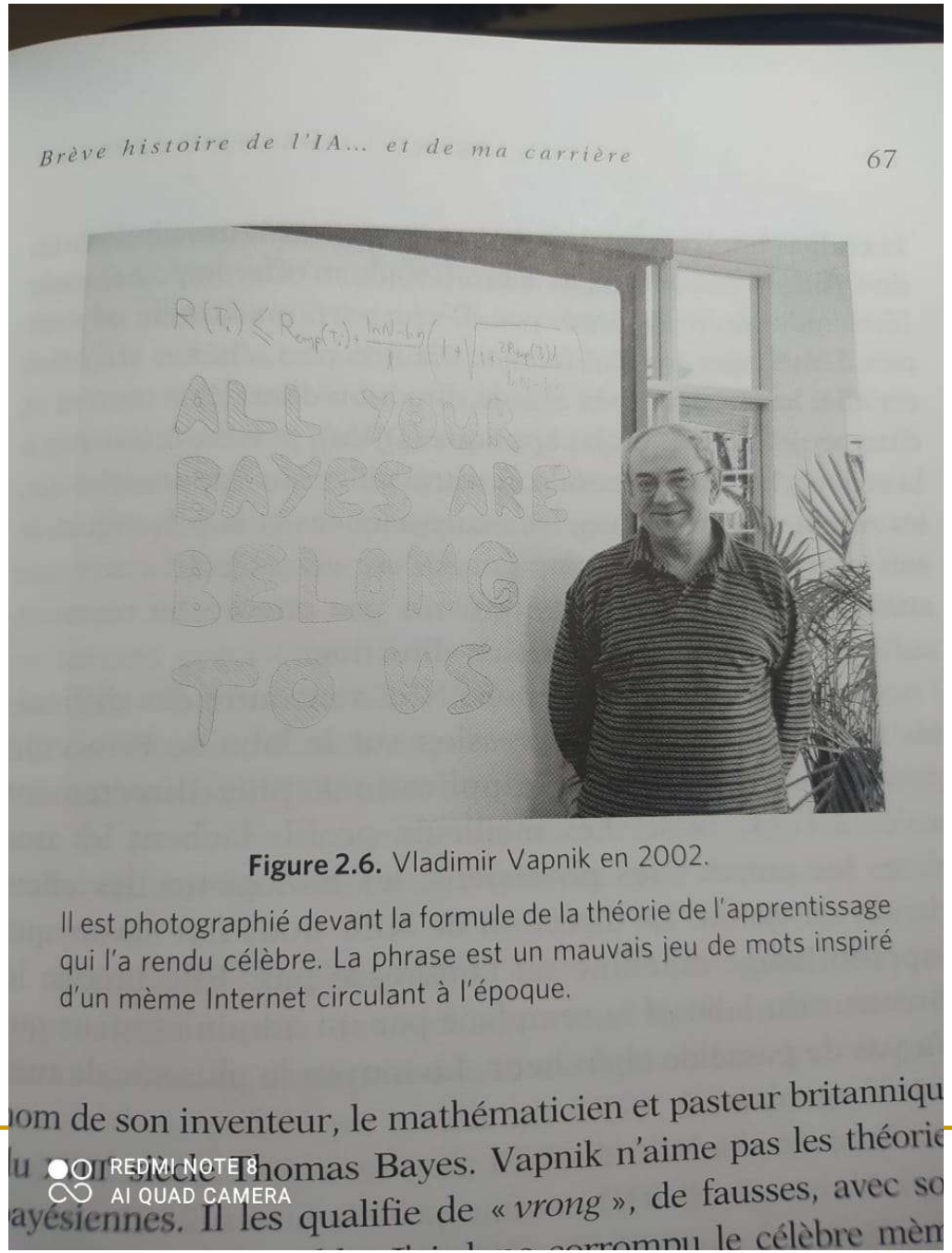
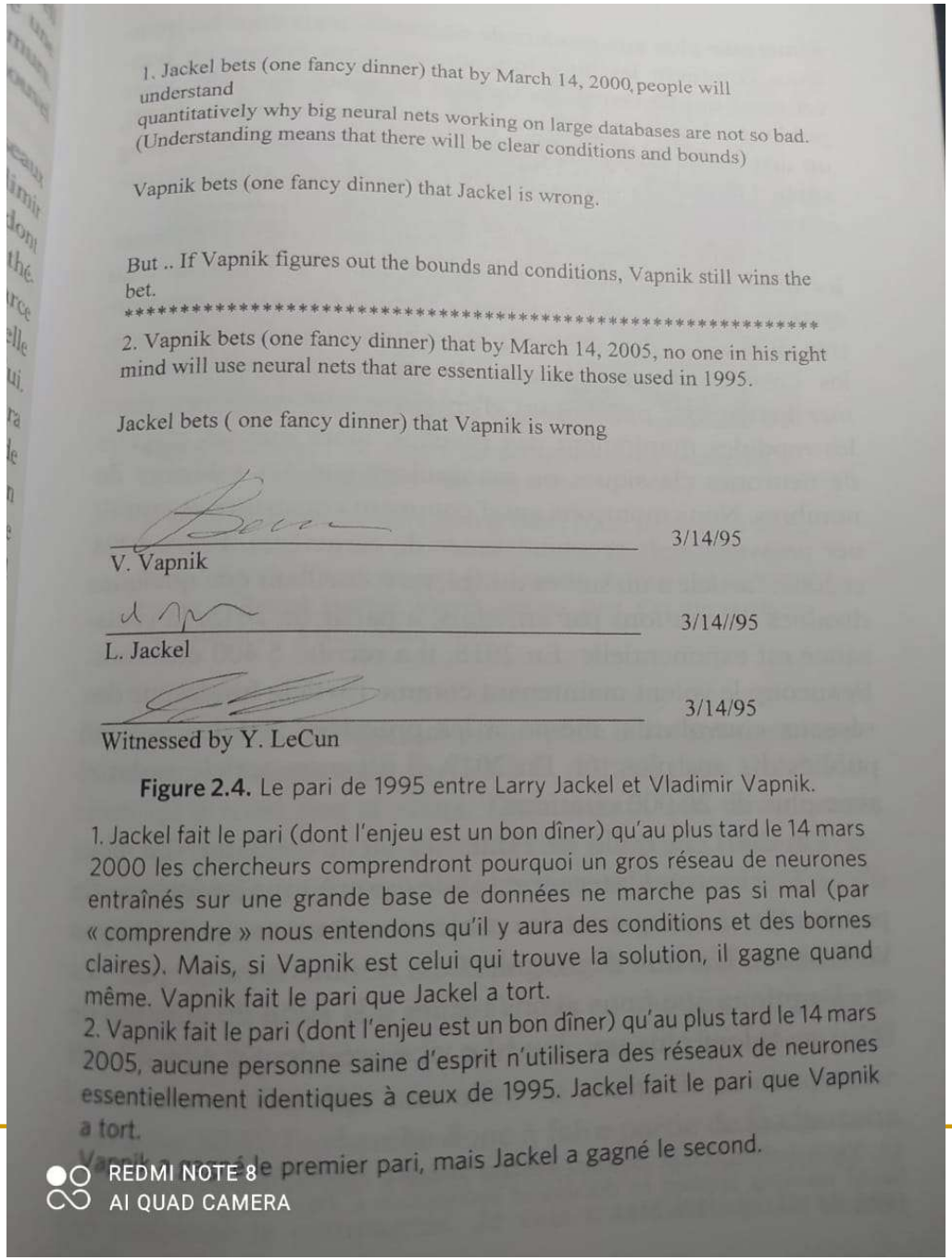


Figure 2.6. Vladimir Vapnik en 2002.

Il est photographié devant la formule de la théorie de l'apprentissage qui l'a rendu célèbre. La phrase est un mauvais jeu de mots inspiré d'un mème Internet circulant à l'époque.

Before the Deep Learning era ...



1. Jackel bets (one fancy dinner) that by March 14, 2000, people will understand
quantitatively why big neural nets working on large databases are not so bad.
(Understanding means that there will be clear conditions and bounds)

Vapnik bets (one fancy dinner) that Jackel is wrong.

But .. If Vapnik figures out the bounds and conditions, Vapnik still wins the bet.

****

2. Vapnik bets (one fancy dinner) that by March 14, 2005, no one in his right mind will use neural nets that are essentially like those used in 1995.

Jackel bets (one fancy dinner) that Vapnik is wrong

3/14/95
V. Vapnik

3/14//95
L. Jackel

3/14/95
Witnessed by Y. LeCun

Figure 2.4. Le pari de 1995 entre Larry Jackel et Vladimir Vapnik.

1. Jackel fait le pari (dont l'enjeu est un bon dîner) qu'au plus tard le 14 mars 2000 les chercheurs comprendront pourquoi un gros réseau de neurones entraînés sur une grande base de données ne marche pas si mal (par « comprendre » nous entendons qu'il y aura des conditions et des bornes claires). Mais, si Vapnik est celui qui trouve la solution, il gagne quand même. Vapnik fait le pari que Jackel a tort.
2. Vapnik fait le pari (dont l'enjeu est un bon dîner) qu'au plus tard le 14 mars 2005, aucune personne saine d'esprit n'utilisera des réseaux de neurones essentiellement identiques à ceux de 1995. Jackel fait le pari que Vapnik a tort.

Vapnik a gagné le premier pari, mais Jackel a gagné le second.

ImageNet and the ILSVRC

- One of the advantages LeNet-5 had over the neocognitron: **larger, high-quality set of training data**.

- Next breakthrough in neural networks was also facilitated by a high-quality public dataset: **ImageNet**.
  - Labeled index of photographs devised by Fei-Fei Li, armed machine vision researchers with an immense catalog of training data.

  - Handwritten data used to train LeNet-5 contained **tens of thousands of images**.

  - **ImageNet contains tens of millions of data.**

ImageNet and the ILSVRC

The hulking ImageNet dataset was the brainchild of Chinese-American computer science professor Fei-Fei Li and her colleagues at Princeton in 2009. Now a faculty member at Stanford University, Li is also the chief scientist of A.I./ML for Google's cloud platform.

ImageNet and the ILSVRC

- The 14 million images in the ImageNet are spread across 22,000 categories.
- Categories are as diverse as: ships, leopards, starfish, and elderberries.
- Since 2010, Li has run an open challenge called **ILSVRC** (the ImageNet Large Scale Visual Recognition Challenge).
  - It uses a subset of the ImageNet data, containing 1.4 million images across 1,000 categories.
  - It has become the premier ground for assessing the world's state-of-the-art machine vision algorithms.
  - Broad range of categories + many categories are breeds of dogs → **evaluates the algorithms' ability not only to distinguish widely varying images but also to specialize in distinguishing subtly varying ones.**

ImageNet and the ILSVRC

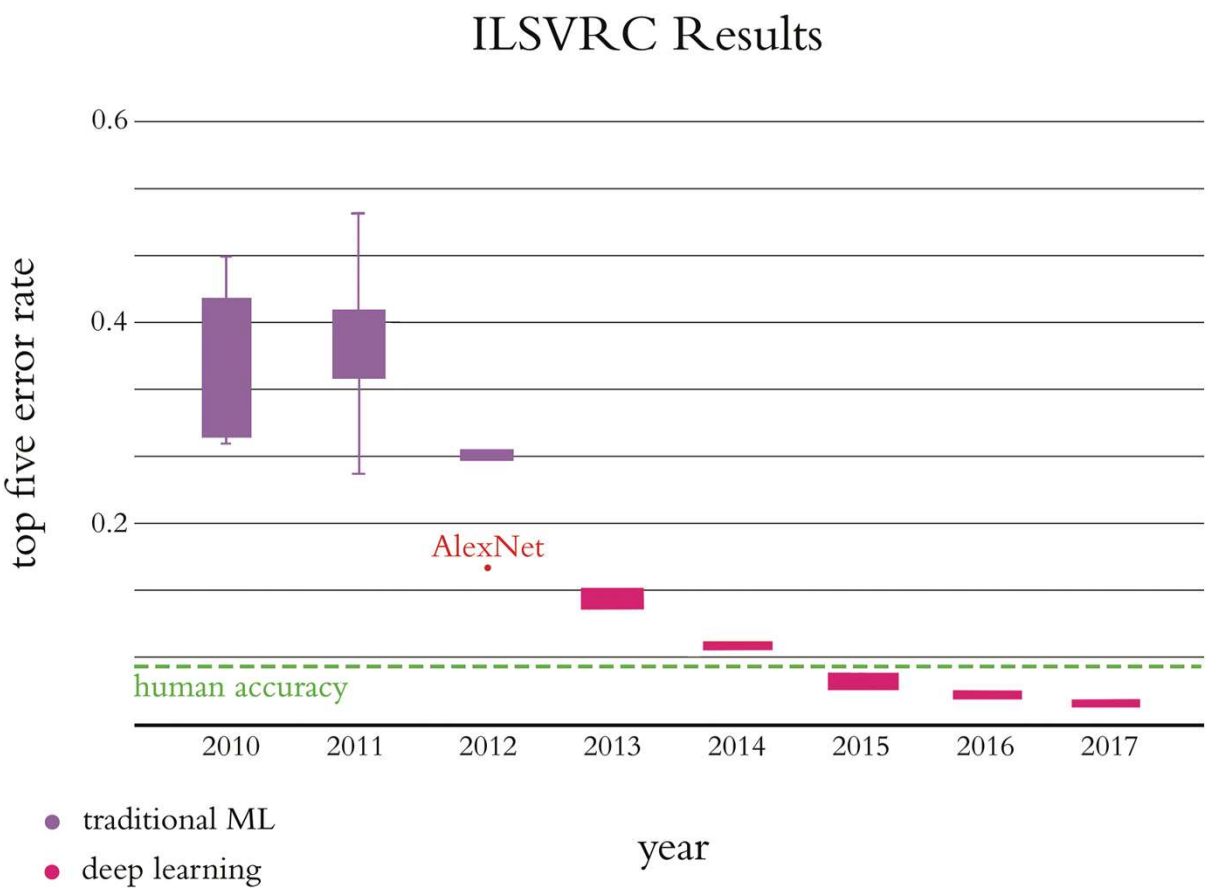
- Let's check the following site: deepdreamgenerator.com

- *"Initially it was invented to help scientists and engineers to see what a deep neural network is seeing when it is looking in a given image. Later the algorithm has become a new form of psychedelic and abstract art."*

AlexNet

- In the first two years of the ILSVRC → all algorithms entered into the competition using the **feature-engineering-driven traditional machine learning ideology**.
- In the third year, all entrants **except one** were traditional ML algorithms.
- If that one deep learning model in 2012 had not been developed → then the year-over-year image classification accuracy would have been negligible.
- Instead, Alex Krizhevsky and Ilya Sutskever – from the University of Toronto lab led by Geoffrey Hinton - **crushed the existing benchmarks with their submission, today referred to as AlexNet**.

AlexNet

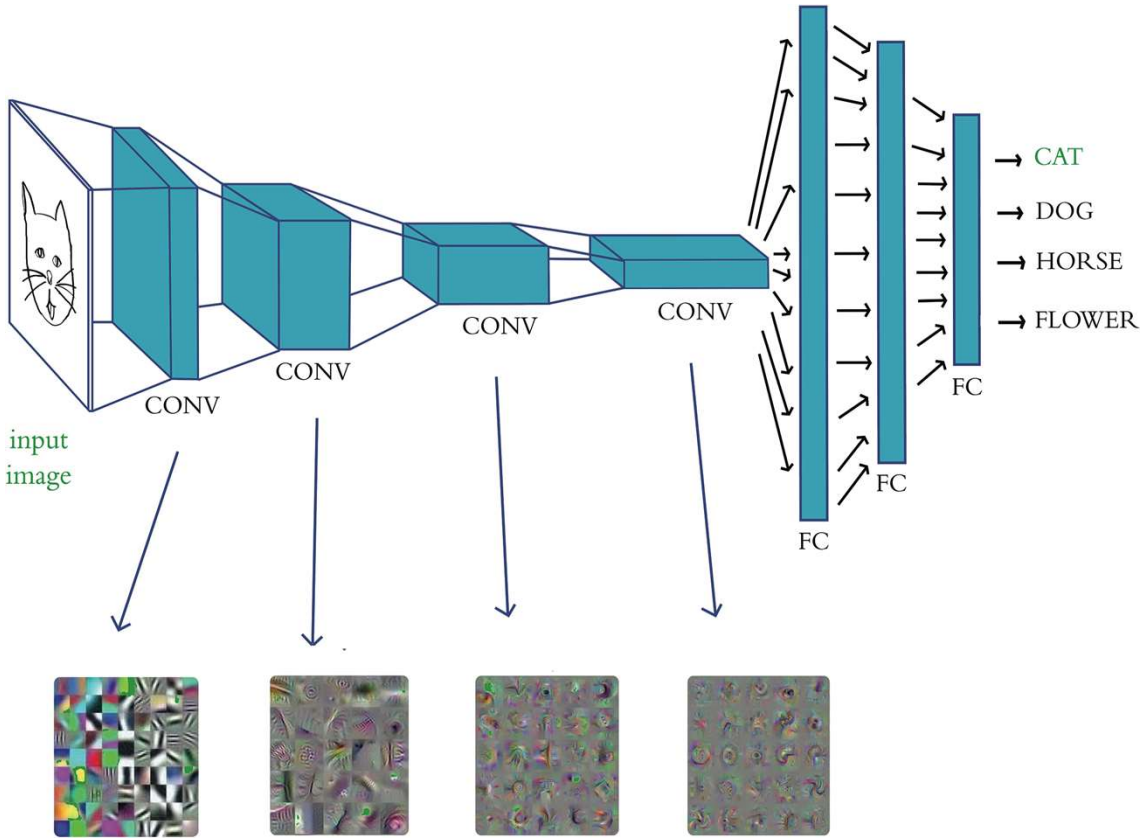


Performance of the top entrants to the ILSVRC by year.

AlexNet

The eminent British-Canadian artificial neural network pioneer Geoffrey Hinton, habitually referred to as “the godfather of deep learning” in the popular press.

Machine Vision



AlexNet's hierarchical architecture is reminiscent of LeNet-5.

AlexNet

- In an instant, deep learning architectures emerged from the fringes of machine learning to its fore.
- Academics and commercial practitioners scrambled to grasp the fundamentals of artificial neural networks.
- They also invested in the creation of software libraries - **many of them open-source** - to experiment with deep learning on their own data and use cases.
  - Be they machine vision or otherwise.

AlexNet

- **In the years since 2012 all of the top-performing models in the ILSVRC have been based on deep learning.**

- Hierarchical architecture of AlexNet is reminiscent of LeNet-5.

- However, there are **three** principal factors that enabled AlexNet to be the state-of-the-art machine vision algorithm in 2012:
  - Training data: massive ImageNet index + they also artificially expanded the data available to them **by applying transformations to the training images**.
  - Processing power: computing power per unit of cost increased dramatically from 1998 to 2012 + but **they also programmed two GPUs to train their large datasets with previously unseen efficiency**.
  - Architectural advances: **AlexNet is deeper (has more layers)** than LeNet-5, and it takes advantage of both a **new type of artificial neuron (ReLU)** and a **nifty trick (dropout)** that helps generalize deep learning models beyond the data they're trained on.

AlexNet

- We will build both LeNet-5 and AlexNet later and use them to classify images.

- This ILSVRC case study underlines why deep learning models are so widely useful and disruptive across industries and computational applications → **they dramatically reduce the subject-matter expertise required for building highly accurate predictive models.**

- Trend away from **expertise-driven feature engineering** and toward powerful **automatic-feature-generating deep learning models**:
  - Vision applications;
  - Complex games;
  - Natural language processing.

AlexNet

- **Vision:** One no longer needs to be a specialist in the visual attributes of faces to create a face-recognition algorithm.
- **Games:** One no longer requires a thorough understanding of a game's strategies to write a program that can master it.
- **Language:** One no longer needs to be an authority on the structure and semantics of each of several languages to develop a language translation tool.
- **For many use cases (and growing!), one's ability to apply deep learning techniques outweighs the value of domain-specific proficiency.**

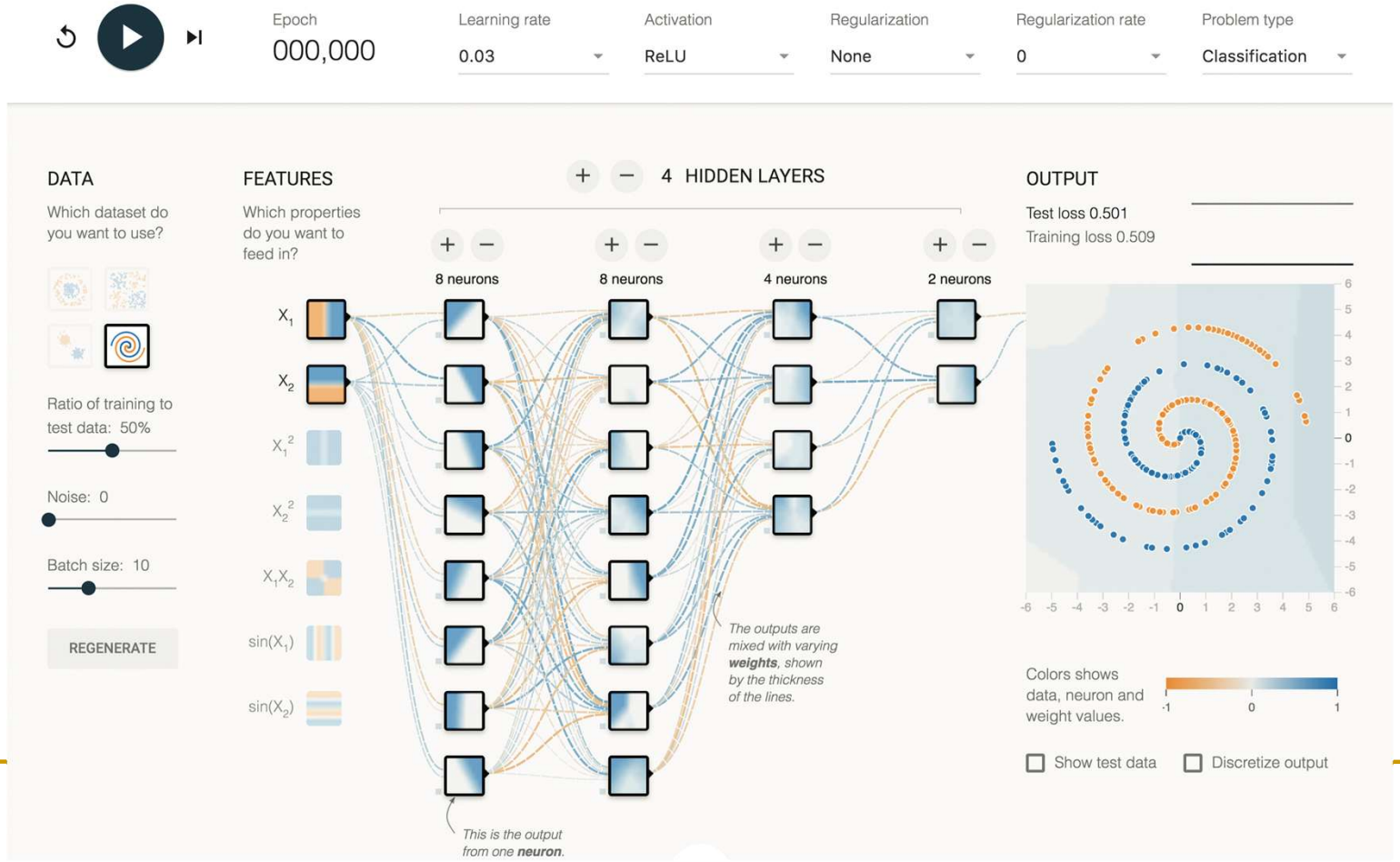
AlexNet

- Such proficiency formerly may have necessitated a **doctoral degree** or perhaps **years of postdoctoral research** within a given domain.

- **While a functional level of deep learning capability can be developed with relative ease**

TensorFlow Playground

Let's crystallize the hierarchical, feature-learning nature of deep learning in the TensorFlow Playground.



TensorFlow Playground

- **This is a deep learning model.**
- Model architecture contains **six layers** of artificial neurons:
  - Input layer on the left (called “features”) .
  - Four “Hidden Layers”, which bear the responsibility of learning.
  - Output Layer (the grid ranging from -6 to +6).
- Goal is to learn how to distinguish **orange dots** (negative cases) from **blue dots** (positive cases) **based solely on their location on the grid**.
- We are only feeding in two pieces of information about each dot: its horizontal position (X) and its vertical position (X).

TensorFlow Playground

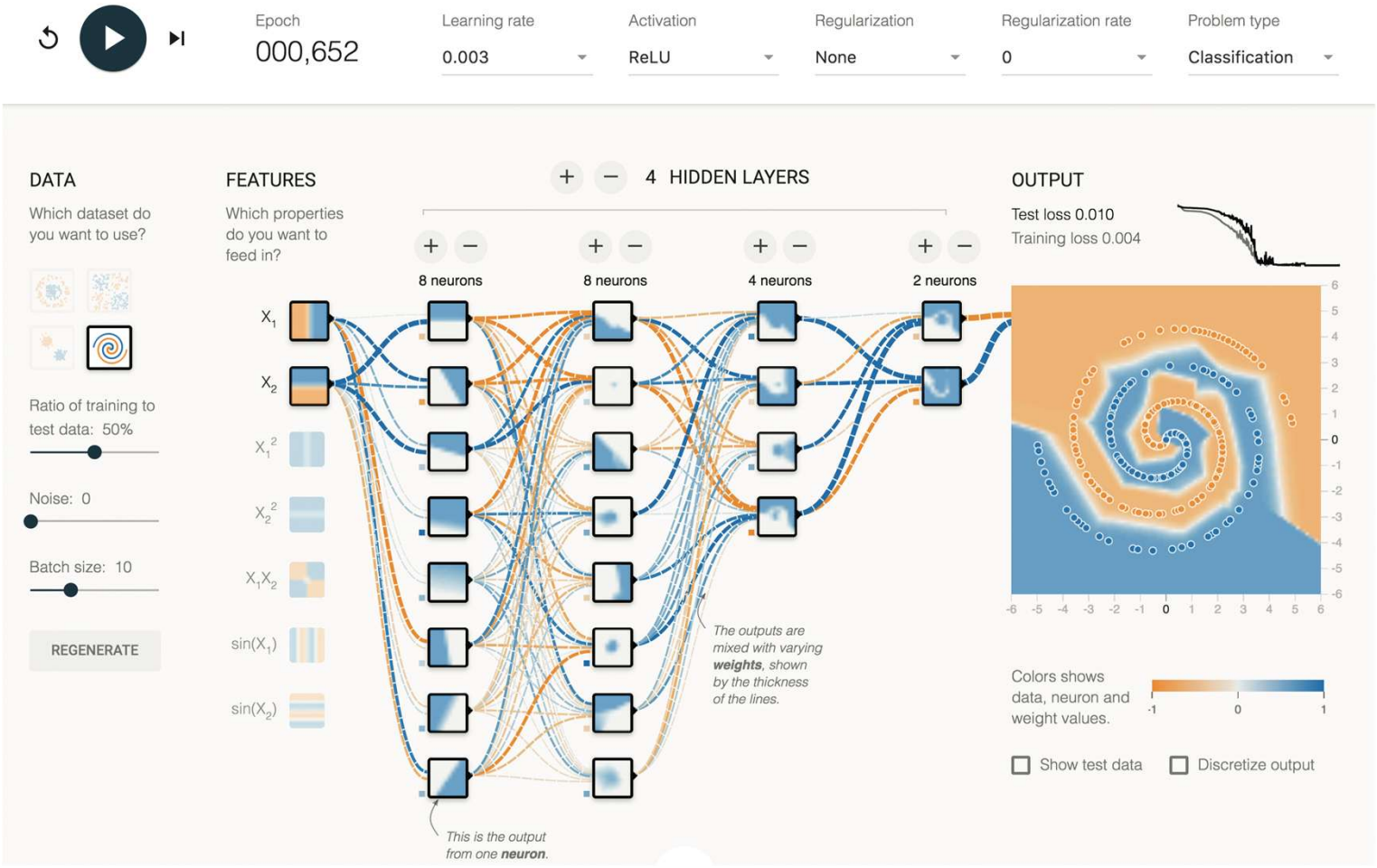
- The dots that will be used as **training data** are shown by default on the grid.
- *Show test* data toggle → we can also see the location of dots that will be used to **assess the performance of the network** as it learns.
- These **test data are not available** to the network while it's learning, so they help us ensure that the network **generalizes well to new, unseen data**.

TensorFlow Playground

- Let's enable the network to train until the "*Training loss*" and "*Test loss*" have both approached zero (less than 0.05).

- The network's artificial neurons represent the input data, with increasing complexity and abstraction the deeper (further
- to the right) they are positioned.
  - as in the neocognitron, LeNet-5, and AlexNet.

TensorFlow Playground



Referências

1. Capítulo 1 do livro *“Deep Learning Illustrated: A Visual, Interactive Guide to Artificial Intelligence”* de Jon Krohn, Grant Beyleveld e Aglaé Bassens.

2. Capítulo 2 do livro *“Quand la machine apprend: La révolution des neurones artificiels et de l’apprentissage profond”* de Yann Le Cun.