

Deep Learning for Computer Vision

Lecture 1: Introduction

Dr. José Ramón Iglesias

DSP-ASIC BUILDER GROUP

Director Semillero TRIAC

Ingenieria Electronica

Universidad Popular del Cesar

Introducción

1



Universidad
Popular del Cesar
4-Apr-23

Artificial Intelligence

Introducción

2



Universidad
Popular del Cesar
4-Apr-23

Slide inspiration: Justin Johnson

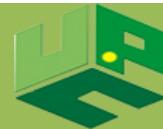
Artificial Intelligence

Machine Learning

Computer
Vision

Introducción

3



Universidad
Popular del Cesar
4-Apr-23

Slide inspiration: Justin Johnson

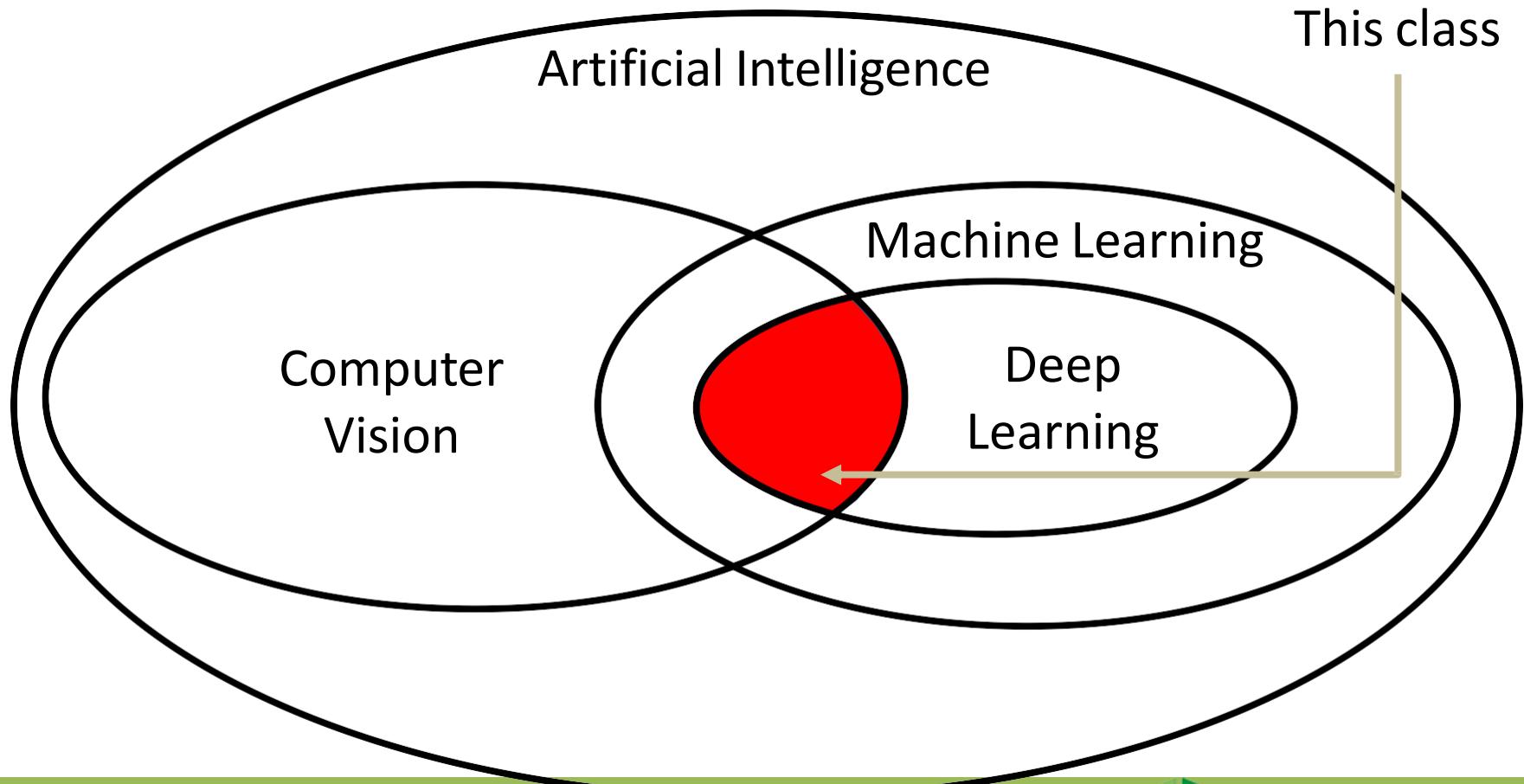
Artificial Intelligence

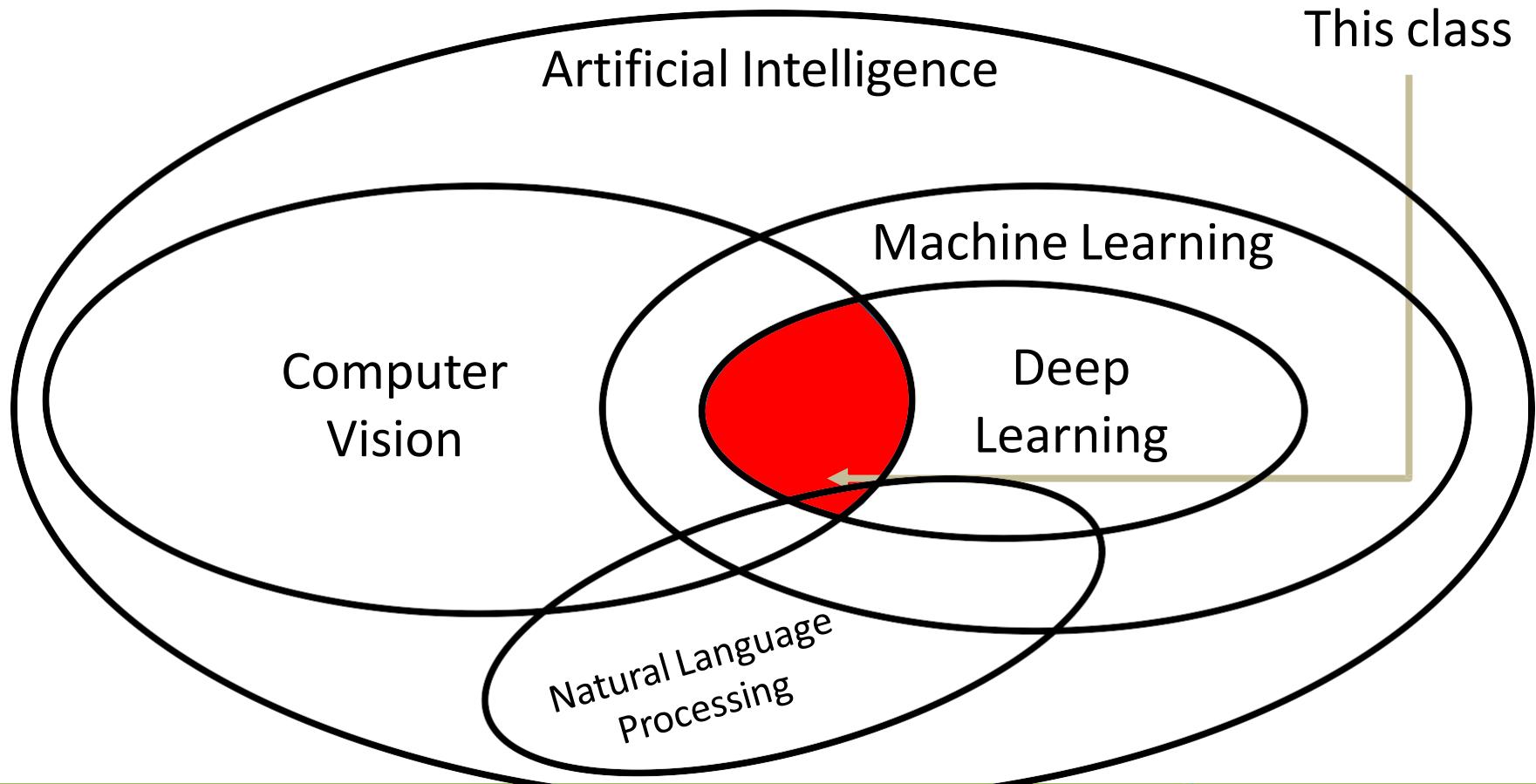
Computer
Vision

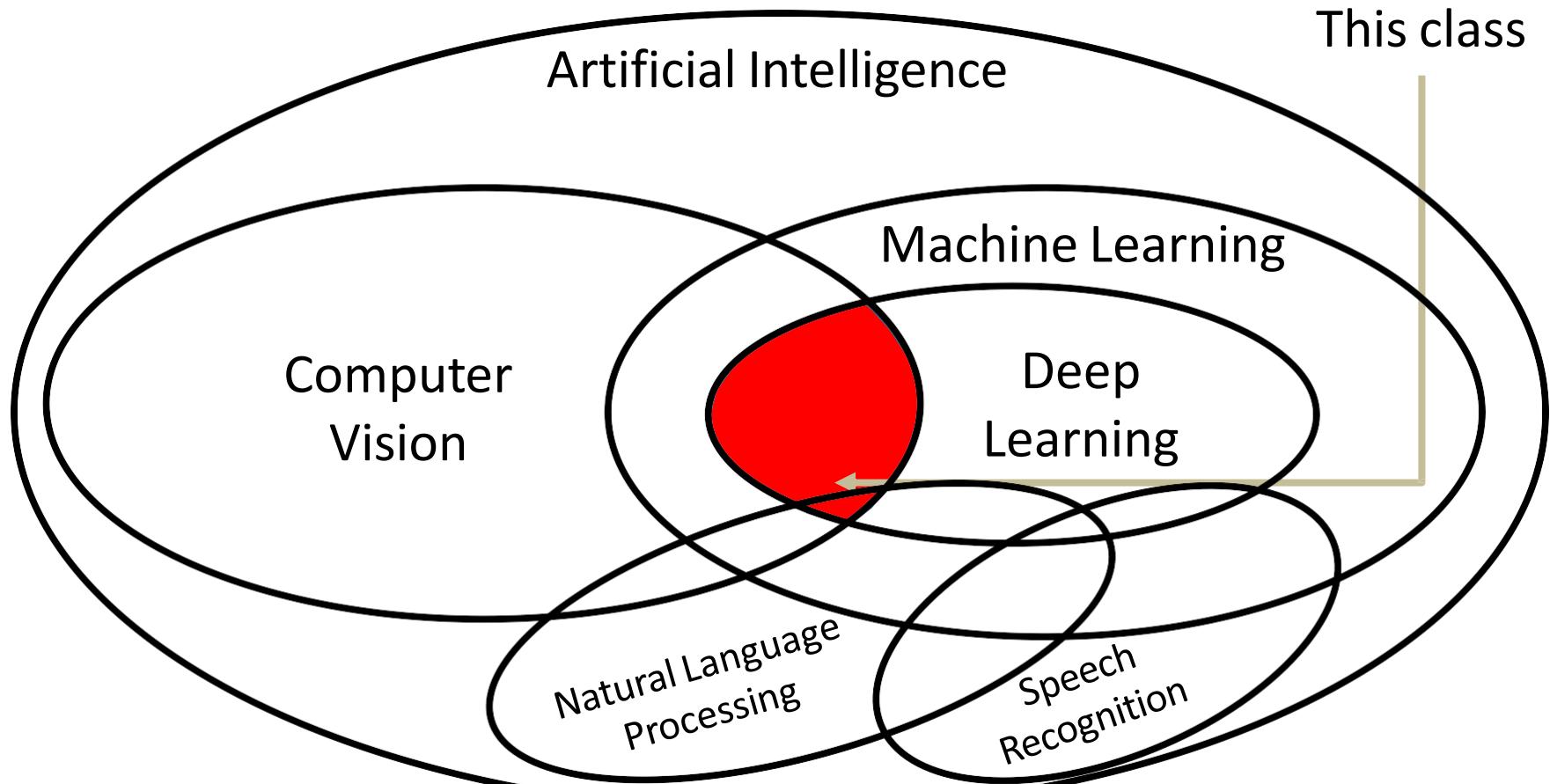
Machine Learning

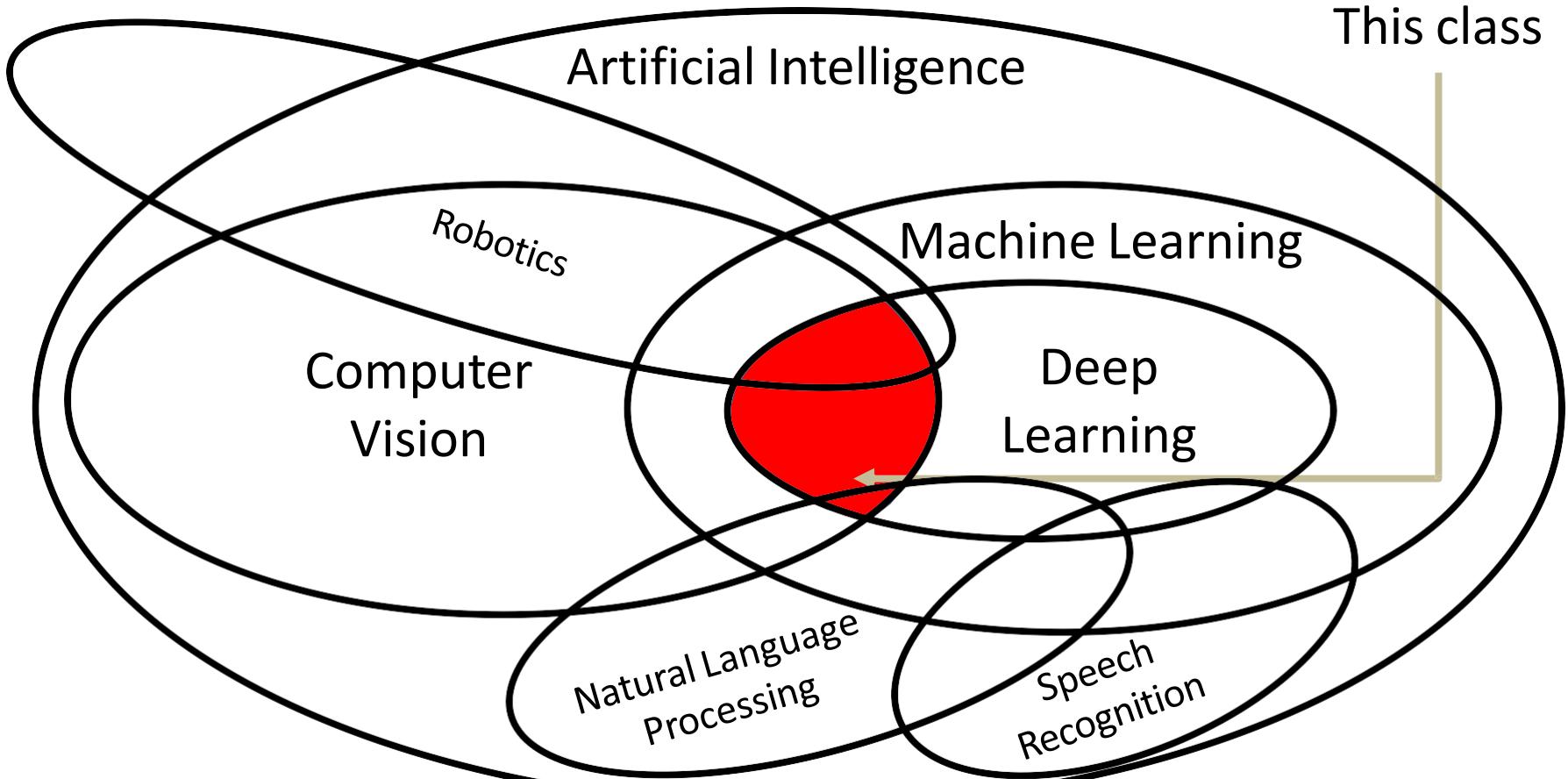
Deep
Learning

Introducción

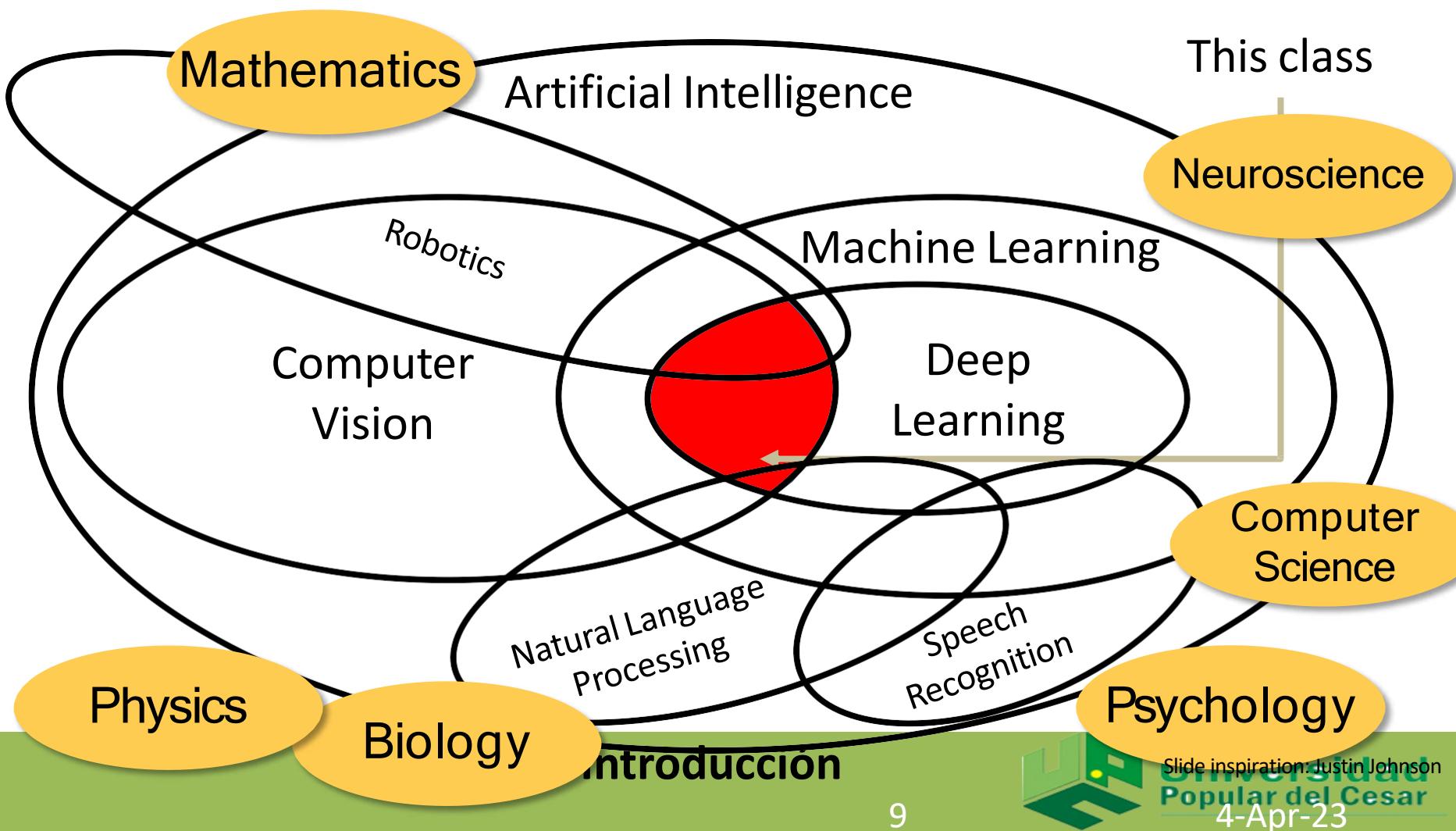








Introducción



Today's agenda

- A brief history of computer vision and deep learning
- overview

Evolution's Big Bang: Cambrian Explosion, 530-540million years, B.C.



This image is licensed under CC-BY 2.5



This image is licensed under CC-BY 2.5



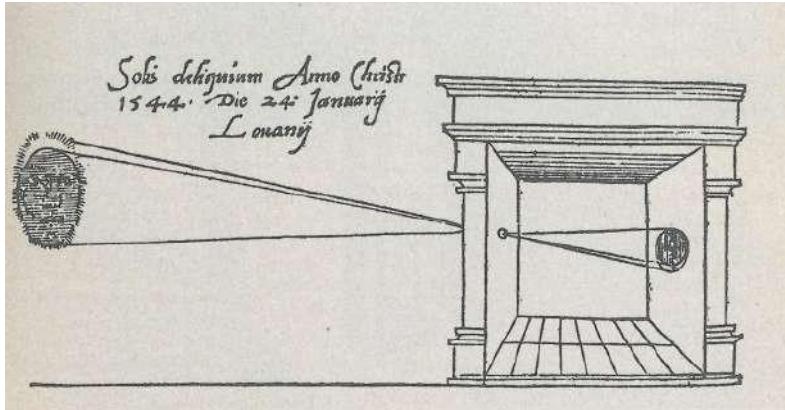
This image is licensed under CC-BY 3.0

Introducción

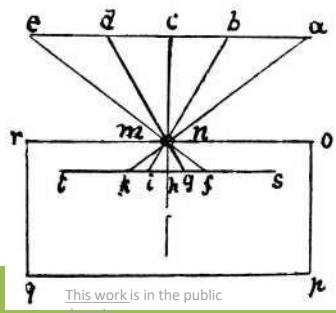


Camera Obscura

Gemma Frisius, 1545

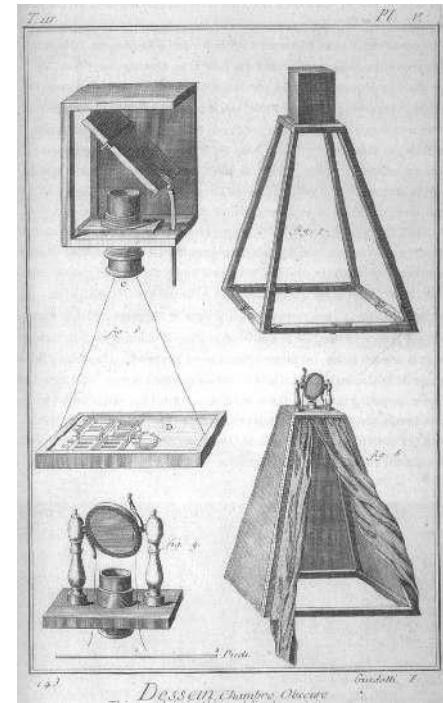


This work is in the public domain



Leonardo da Vinci,
16th Century AD

Encyclopedia, 18th Century

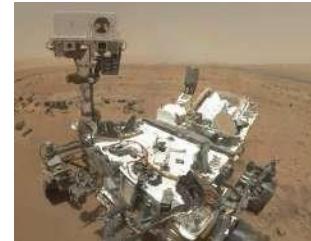


Introducción

Computer Vision is everywhere!



Left to right:
Image by Roger H Goun is licensed under CC BY 2.0
Image is CCO 1.0 public domain
Image is CCO 1.0 public domain
Image is CCO 1.0 public domain



Left to right:
Image is free to use
Image is CCO 1.0 public domain
Image by NASA is licensed under CC BY 2.0
Image is CCO 1.0 public domain



Bottom row, left to right
Image is CCO 1.0 public domain
Image by Derek Keats is licensed under CC BY 2.0, changes made
Image is public domain
Image is licensed under CC-BY 2.0, changes made

introduction

Where did we come from?

Introducción

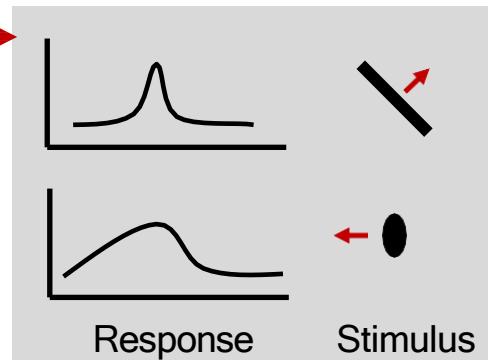
15



Universidad
Popular del Cesar
4-Apr-23

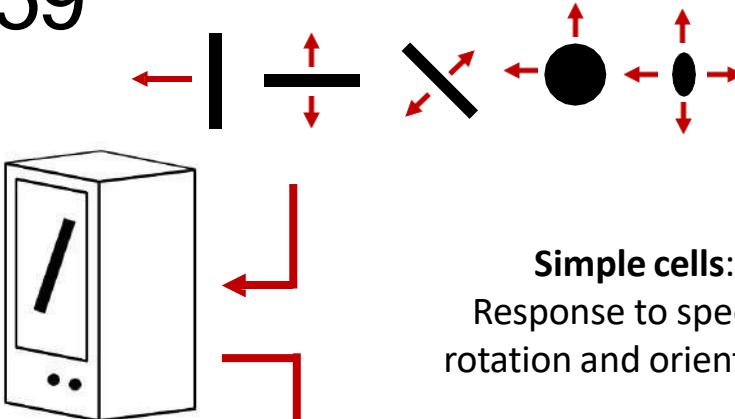
Hubel and Wiesel, 1959

Measure
brain activity



Cat image by CNX OpenStax is licensed under CC BY 4.0; changes made

1959
Hubel & Wiesel

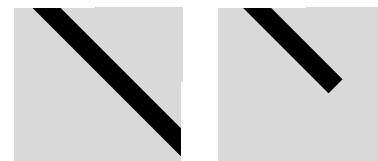


Simple cells:

Response to specific rotation and orientation

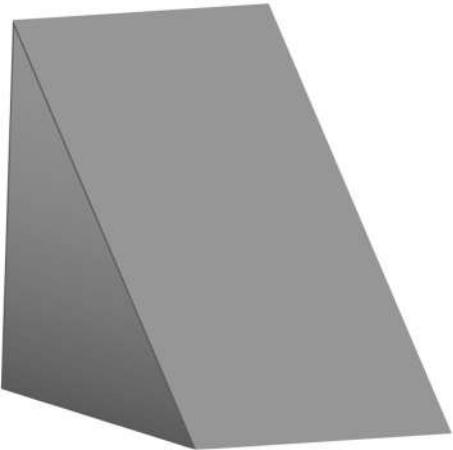
Complex cells:

Response to light orientation and movement, some translation invariance

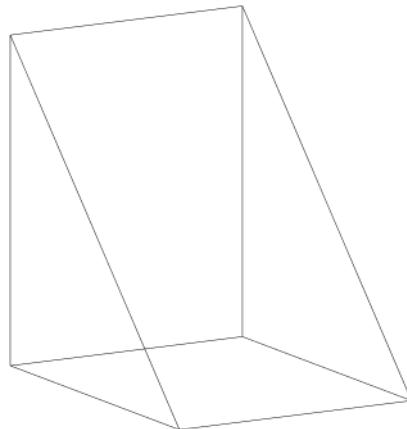


Introducción

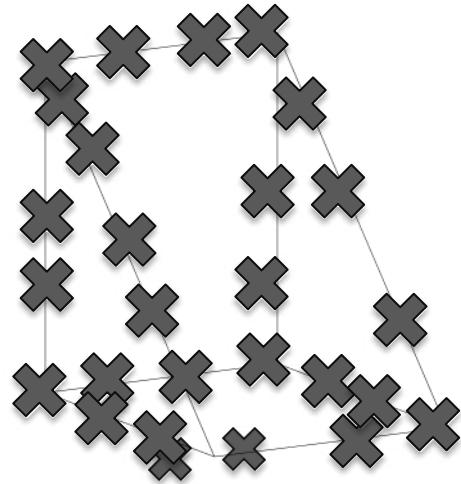
Larry Roberts, 1963



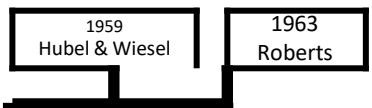
(a) Original picture



(b) Differentiated picture



(c) Feature points selected



Lawrence Gilman Roberts, "Machine Perception of Three-Dimensional Pictures," 1963

Introducción

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

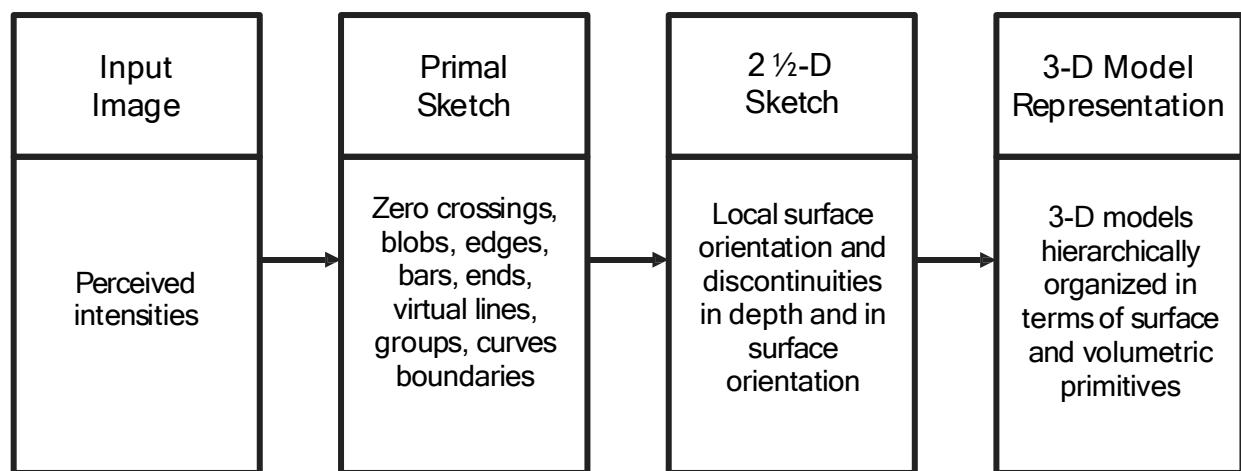
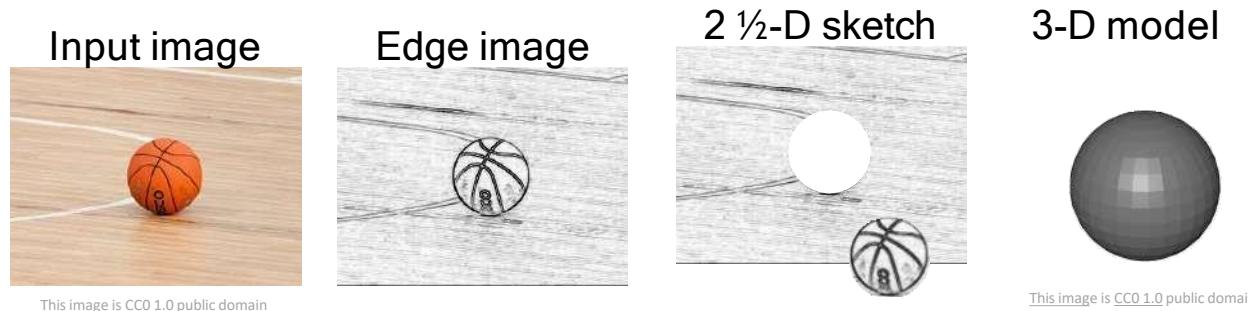
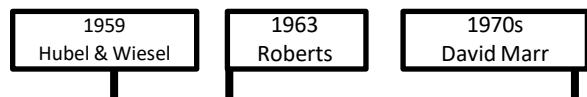
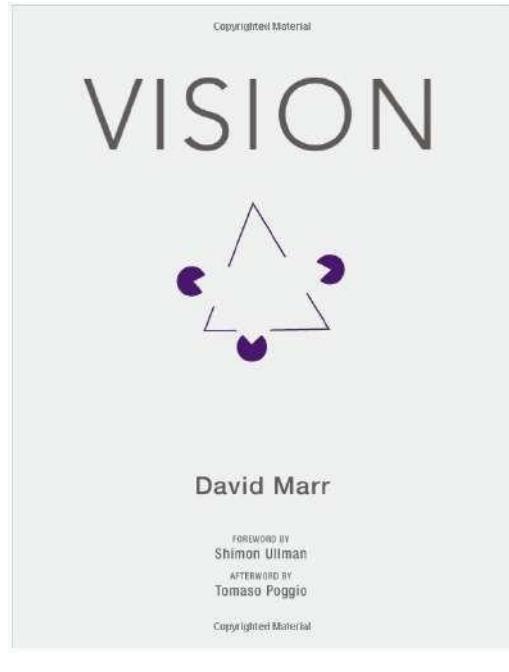
THE SUMMER VISION PROJECT

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".

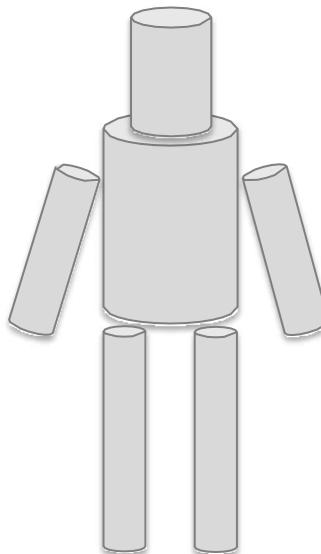
1959
Hubel & Wiesel

1963
Roberts

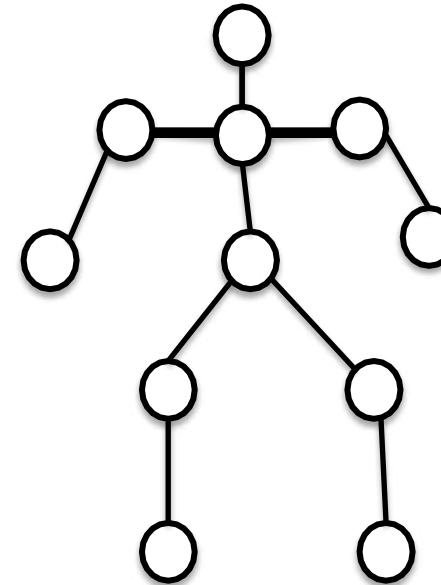


Stages of Visual Representation, David Marr, 1970s
Introducción

Recognition via Parts (1970s)



Generalized Cylinders,
Brooks and Binford,
1979



Pictorial Structures,
Fischler and Elshlager, 1973



Introducción

Recognition via Edge Detection (1980s)



1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

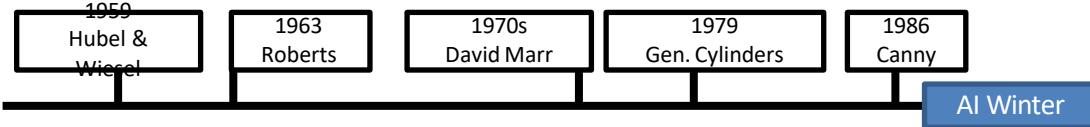
1986
Canny

John Canny, 1986
David Lowe, 1987

Introducción

Arriving at an “AI winter”

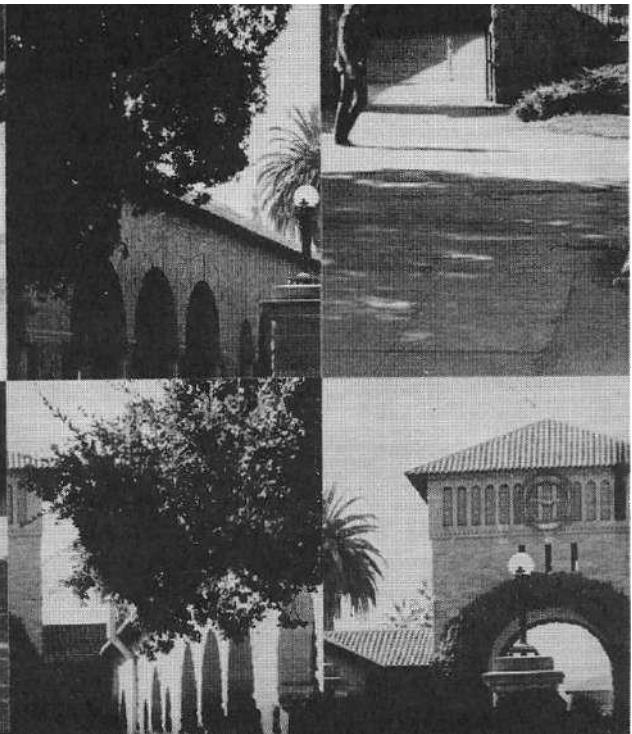
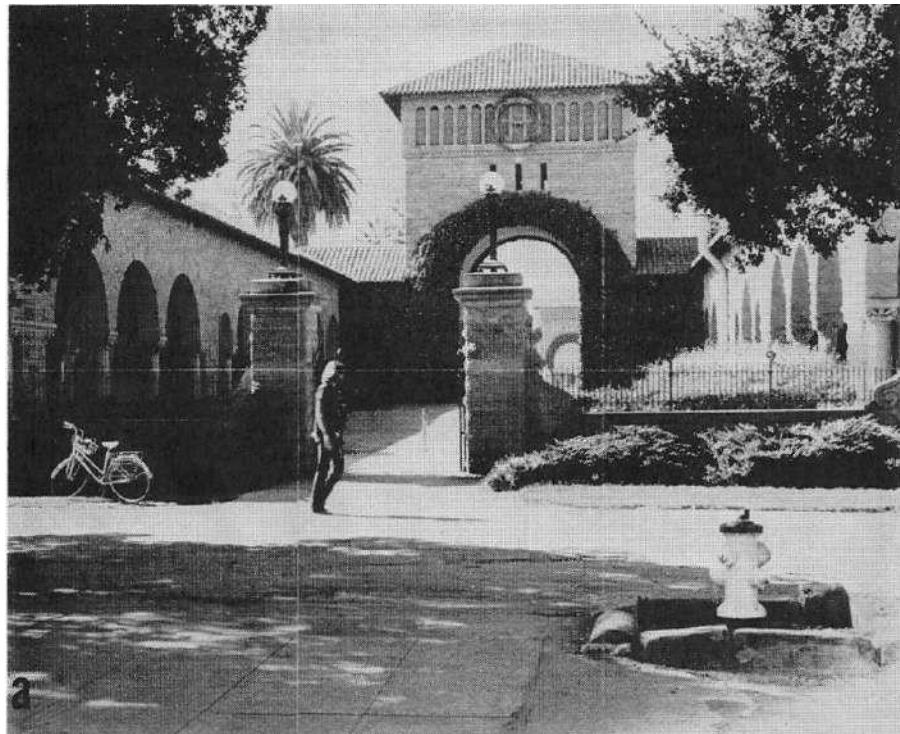
- Enthusiasm (and funding!) for AI research dwindled
- “Expert Systems” failed to deliver on their promises
- But subfields of AI continues to grow
 - Computer vision, NLP, robotics, compbio, etc.



In the meantime...seminal work in
cognitive and neuroscience

Perceiving Real-World Scenes

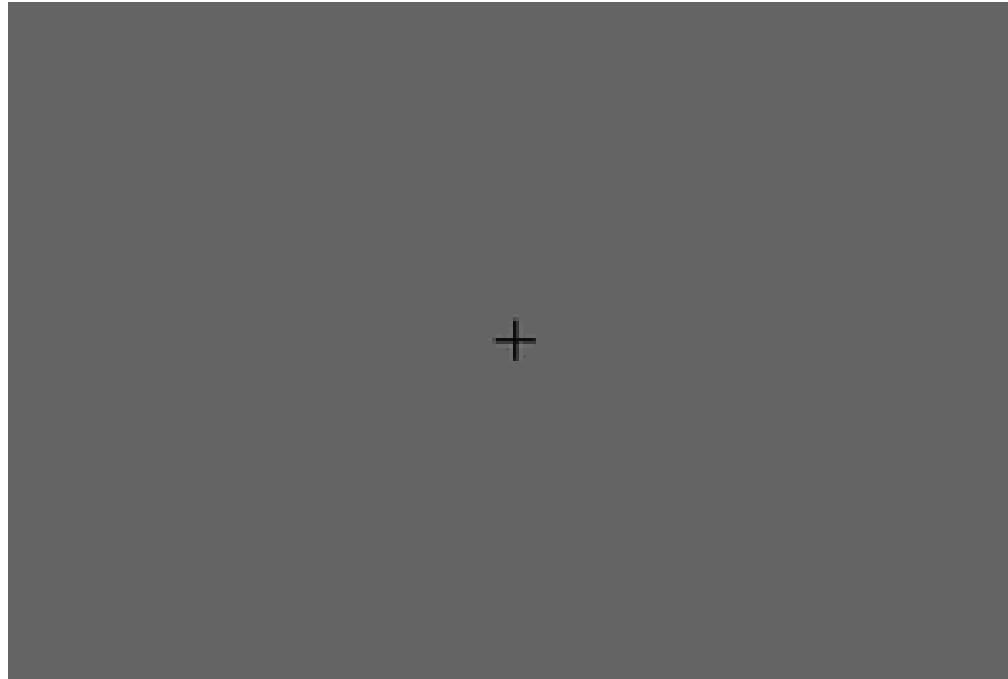
Irving Biederman



Introducción

24

Rapid Serial Visual Perception (RSVP)



Potter, etc. 1970s

Introducción

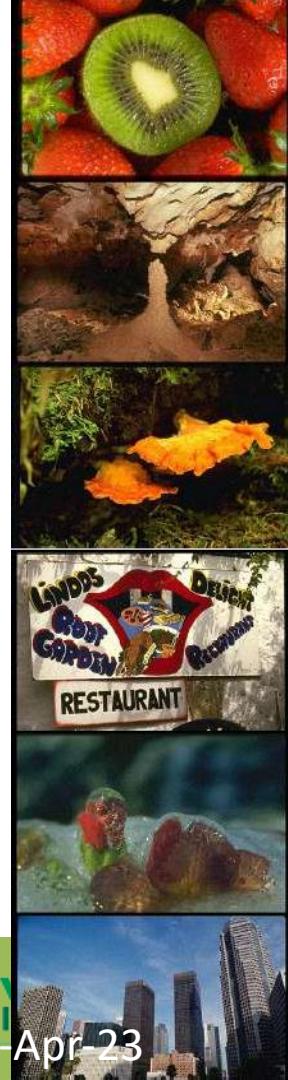
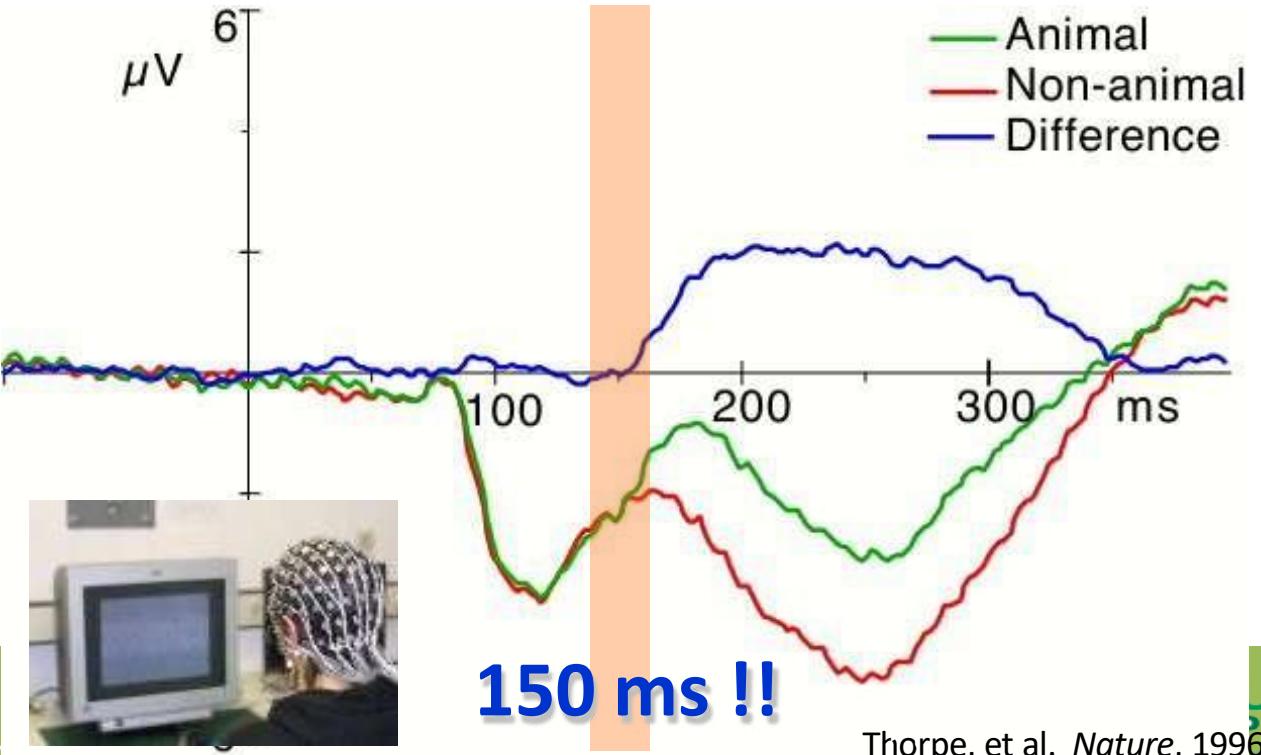
25



Universidad
Popular del Cesar
4-Apr-23

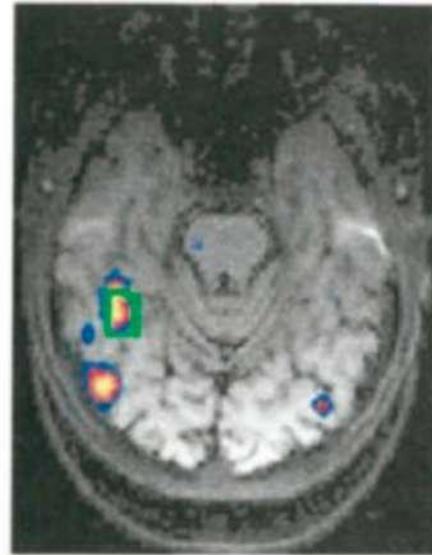
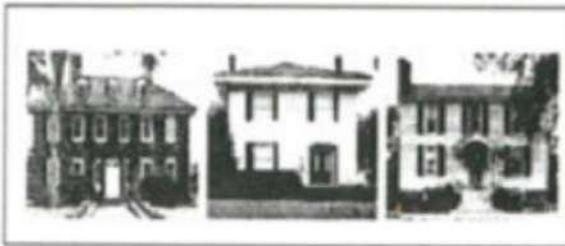
Speed of processing in the human visual system

Simon Thorpe, Denis Fize & Catherine Marlot

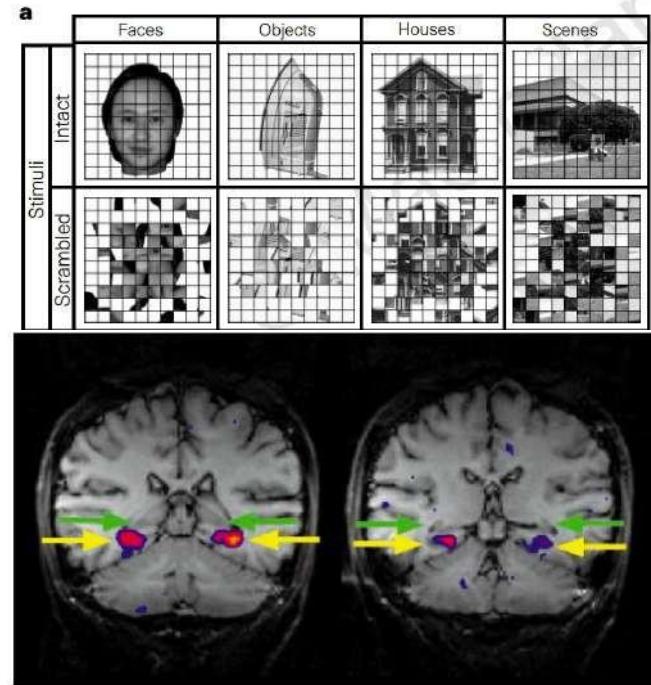


Neural correlates of object & scene recognition

Faces > Houses



% signal change

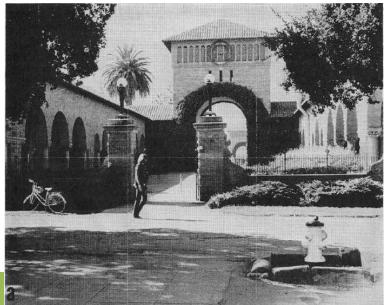


Kanwisher et al. J. Neuro. 1997

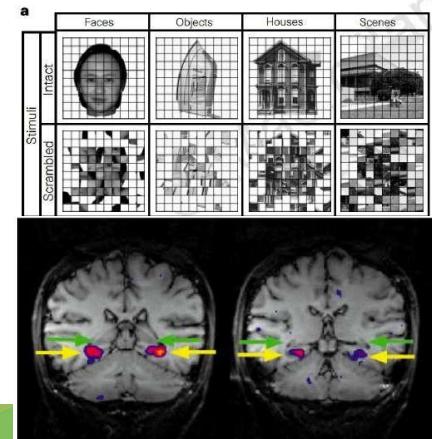
Introducción

Epstein & Kanwisher, Nature, 1998

Visual recognition is a fundamental task for visual intelligence



introducción



Recognition via Grouping (1990s)



1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

1986
Canny

1997
Norm. Cuts

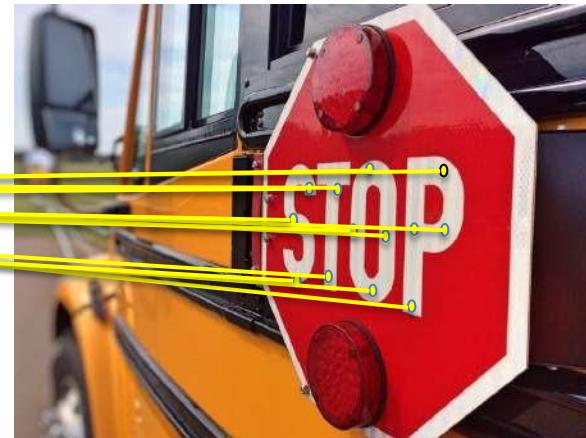
AI Winter

Normalized Cuts, Shi and Malik, 1997

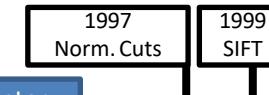
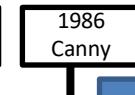
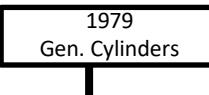
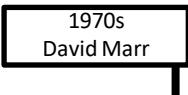
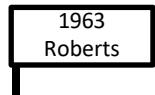
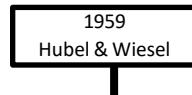
Recognition via Matching (2000s)



[Image](#) is public domain



[Image](#) is public domain



AI Winter

SIFT, David Lowe, 1999

Introducción

Face Detection

Viola and Jones, 2001

One of the first successful applications of machine learning to vision



1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

1986
Canny

1997
Norm. Cuts

1999
SIFT

2001
V&J

AI Winter

Introducción

Caltech 101 images



1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

1986
Canny

1997
Norm. Cuts

1999
SIFT

2001
V&J

2004, 2007
Caltech101;
PASCAL

AI Winter

PASCAL Visual Object Challenge

Image is CC0 1.0 public domain



Image is CC0 1.0 public domain

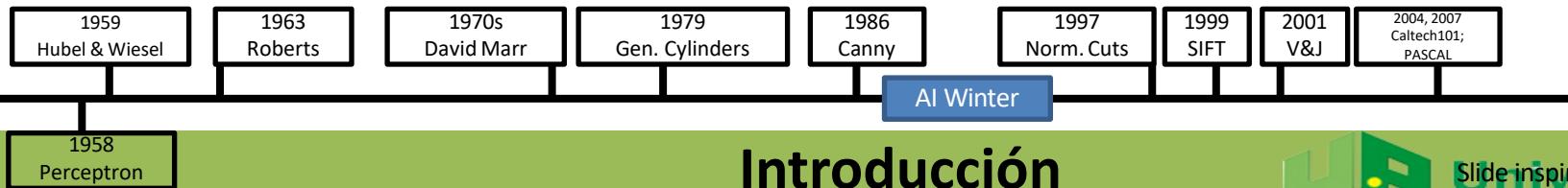
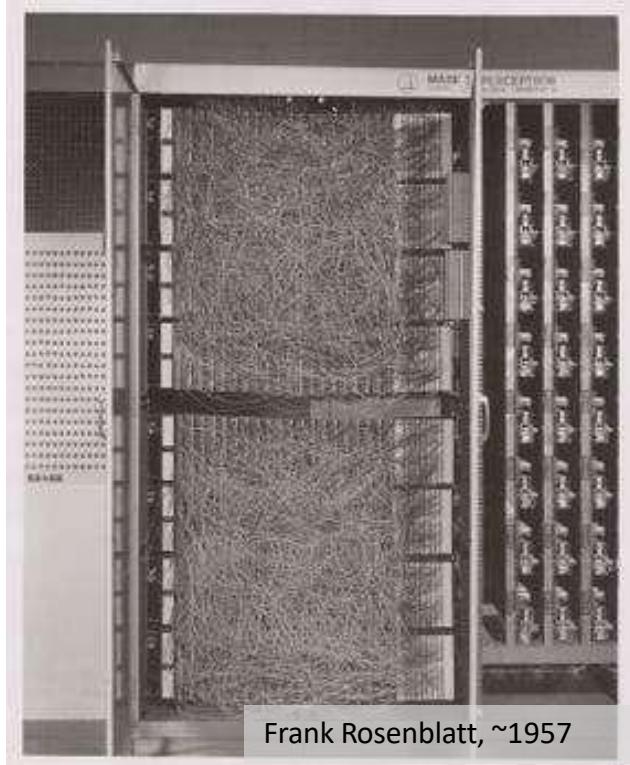
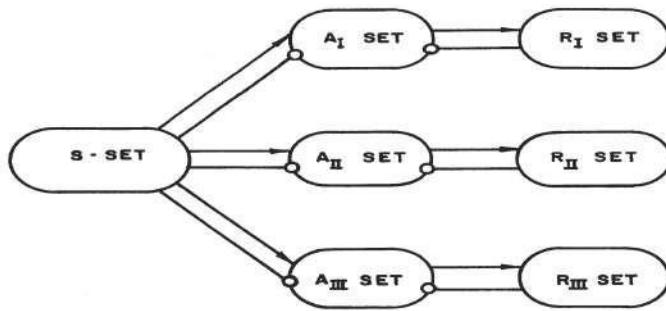
Introducción

Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton†
& Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California,
San Diego, La Jolla, California 92093, USA

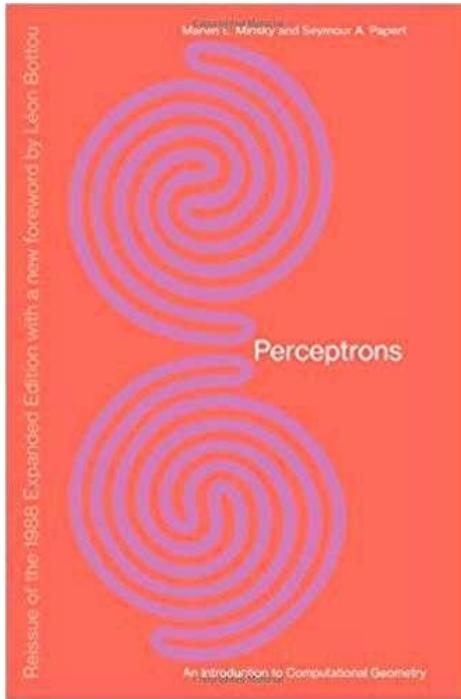
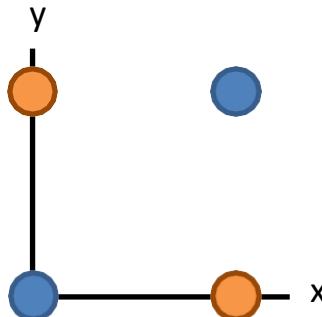
† Department of Computer Science, Carnegie-Mellon University,
Pittsburgh, Philadelphia 15213, USA



Introducción

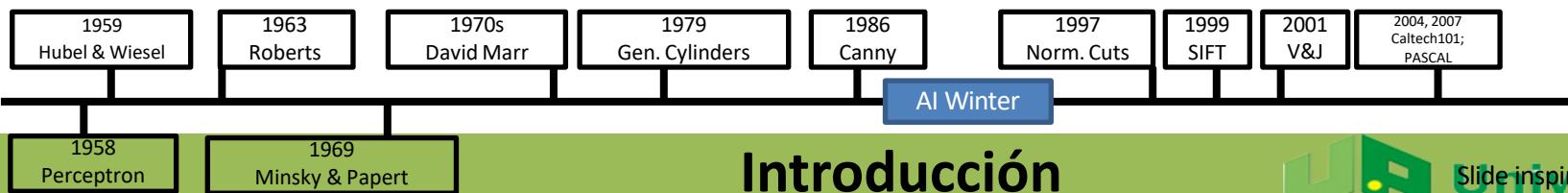
Minsky and Papert, 1969

X	Y	F(x,y)
0	0	0
0	1	1
1	0	1
1	1	0



Showed that Perceptrons could not learn the XOR function

Caused a lot of disillusionment in the field



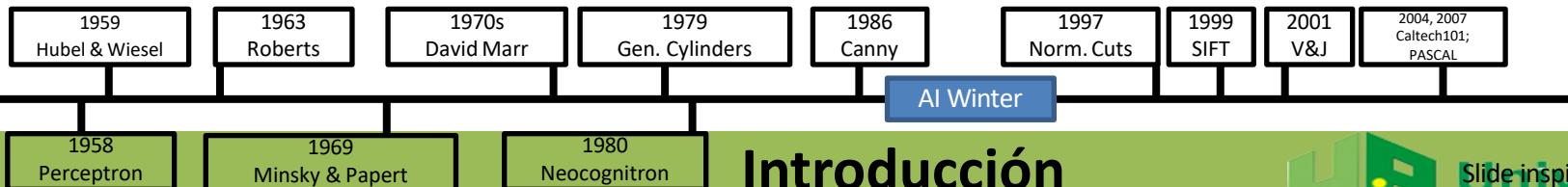
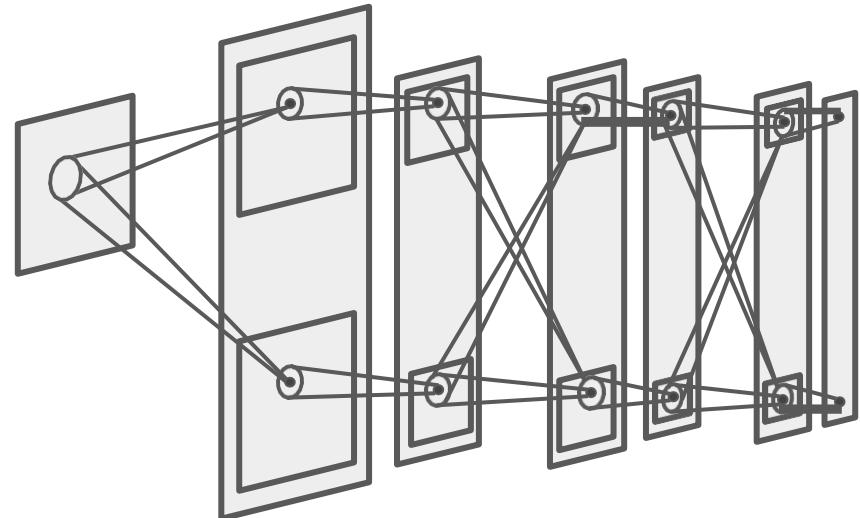
Introducción

Neocognitron: Fukushima, 1980

Computational model the visual system,
directly inspired by Hubel and Wiesel's
hierarchy of complex and simple cells

Interleaved simple cells (convolution)
and complex cells (pooling)

No practical training algorithm

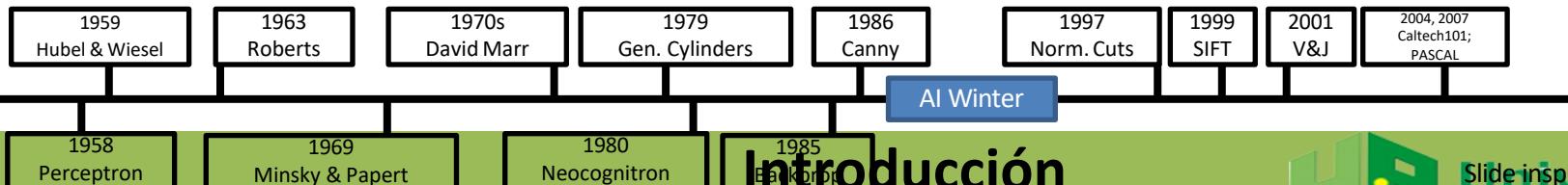
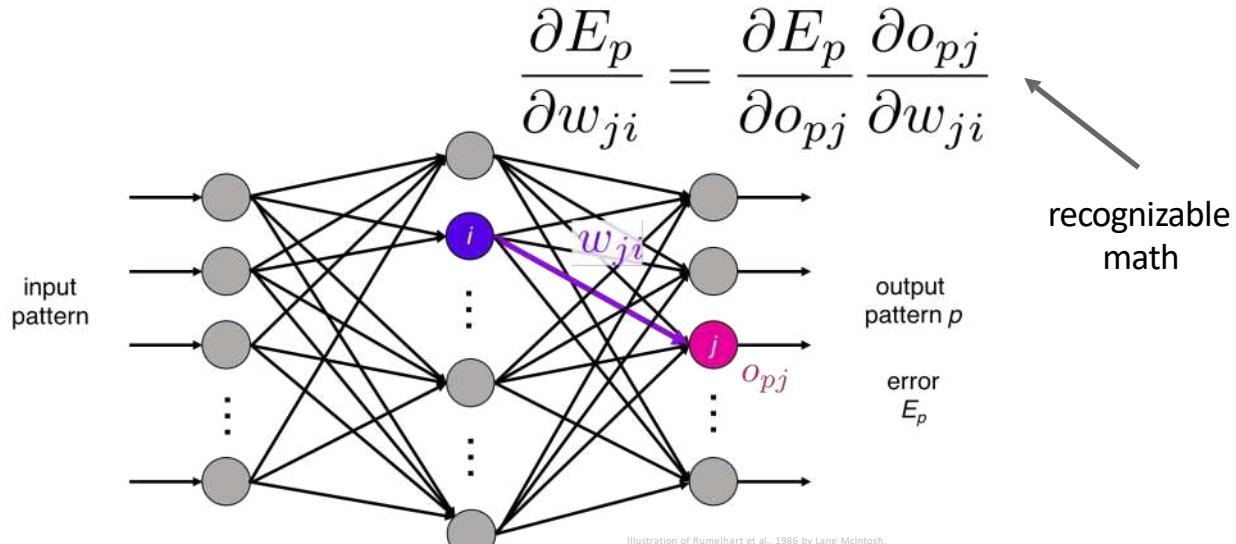


Introducción

Backprop: Rumelhart, Hinton, and Williams, 1986

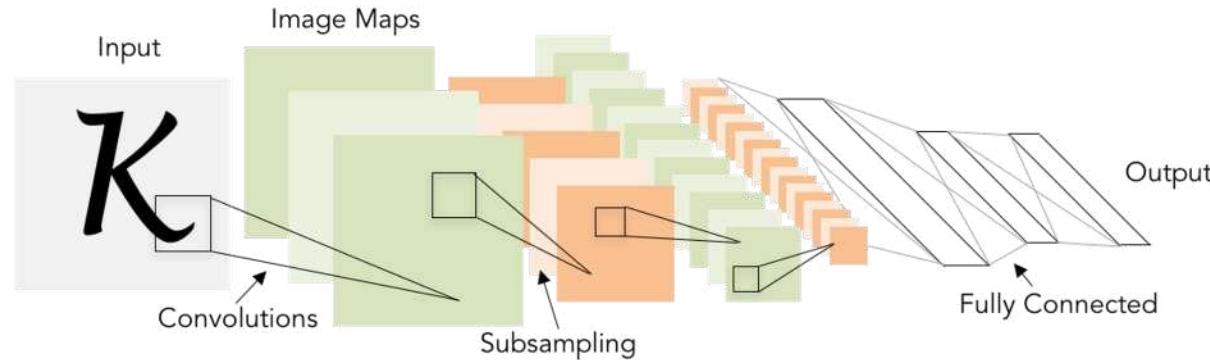
Introduced backpropagation for computing gradients in neural networks

Successfully trained perceptrons with multiple layers



Introducción

Convolutional Networks: LeCun et al, 1998

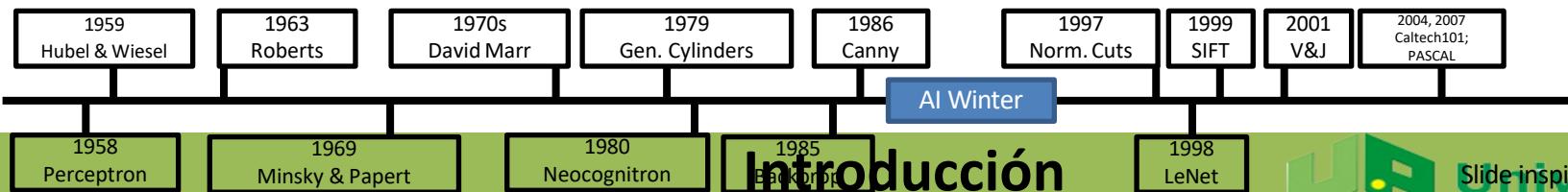


Applied backprop algorithm to a Neocognitron-like architecture

Learned to recognize handwritten digits

Was deployed in a commercial system by NEC, processed handwritten checks

Very similar to our modern convolutional networks!



Introducción

2000s: “Deep Learning”

People tried to train neural networks that were deeper and deeper

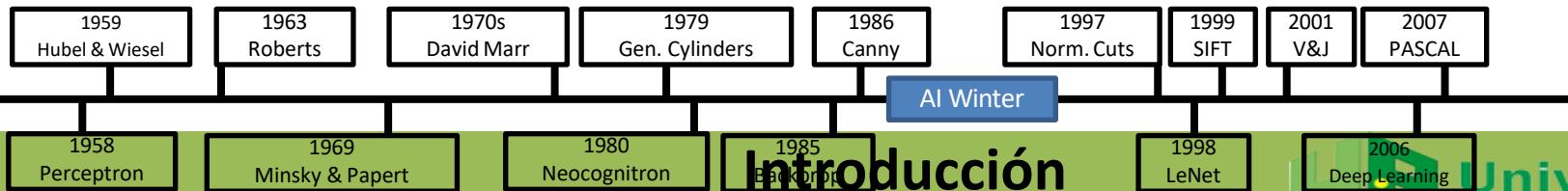
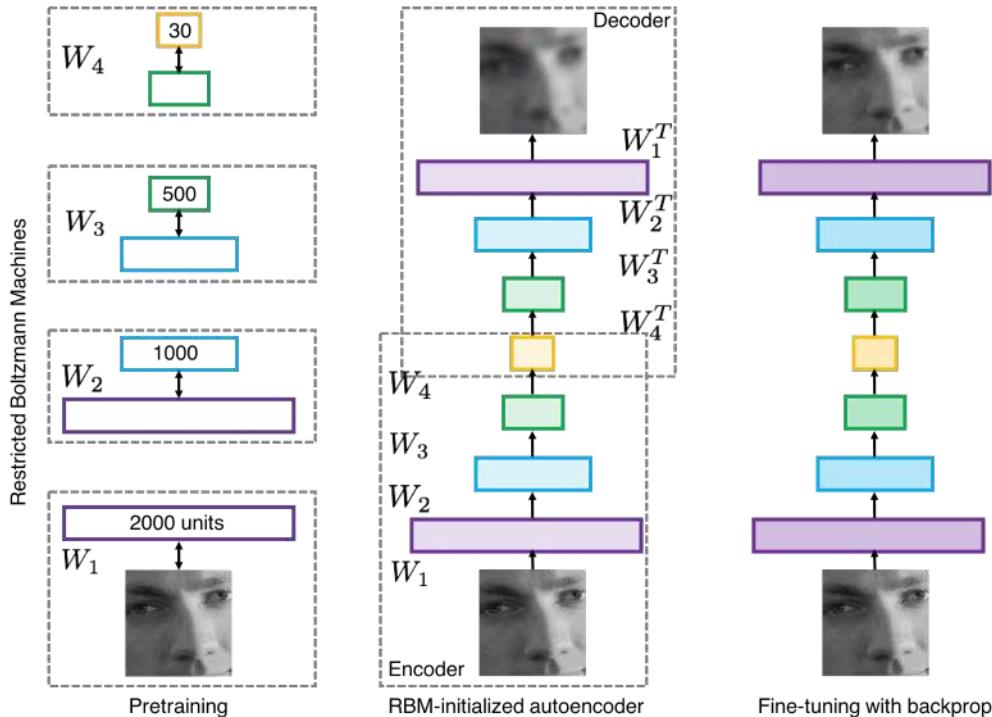
Not a mainstream research topic at this time

Hinton and Salakhutdinov, 2006

Bengio et al, 2007

Lee et al, 2009

Glorot and Bengio, 2010



Introducción

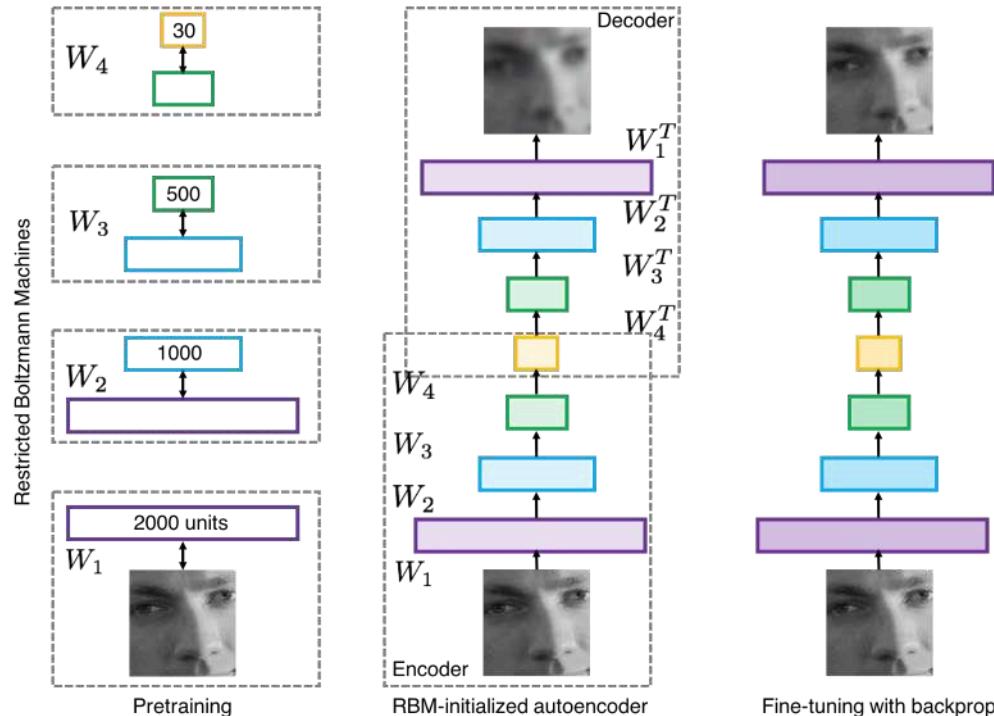
2000s: “Deep Learning”

People tried to train neural networks that were deeper and deeper

Not a mainstream research topic at this time

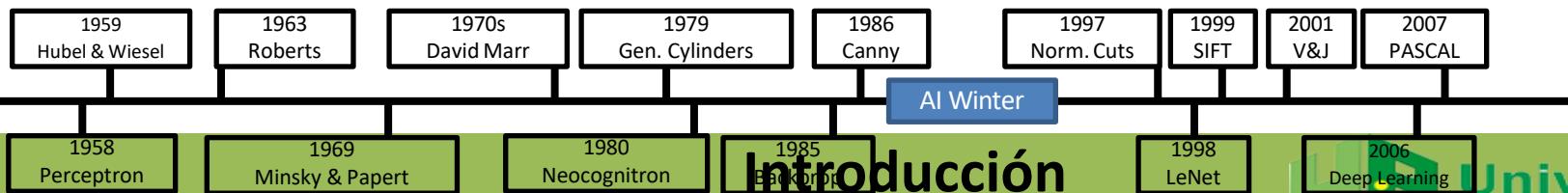
No good dataset to work on

Hinton and Salakhutdinov, 2006
Bengio et al, 2007
Lee et al, 2009
Glorot and Bengio, 2010



Fine-tuning with backprop

Slide inspiration: Justin Johnson



Introducción

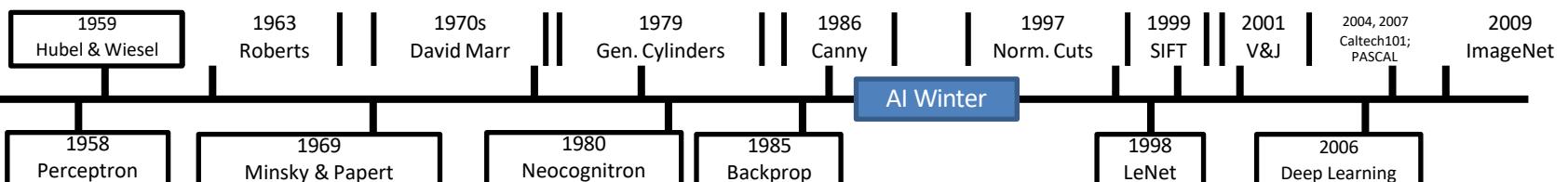
IMAGENET Large Scale Visual Recognition Challenge

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

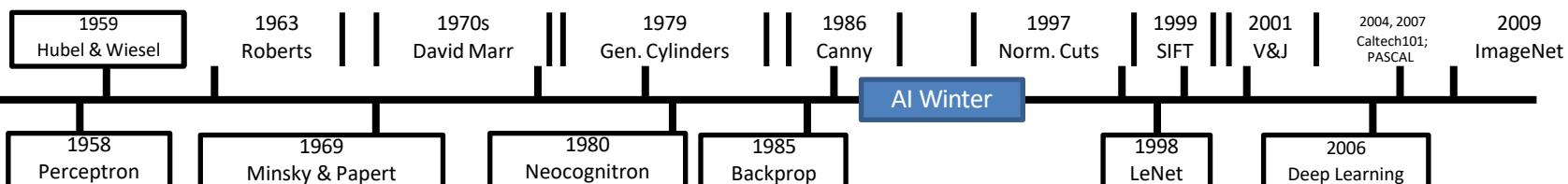
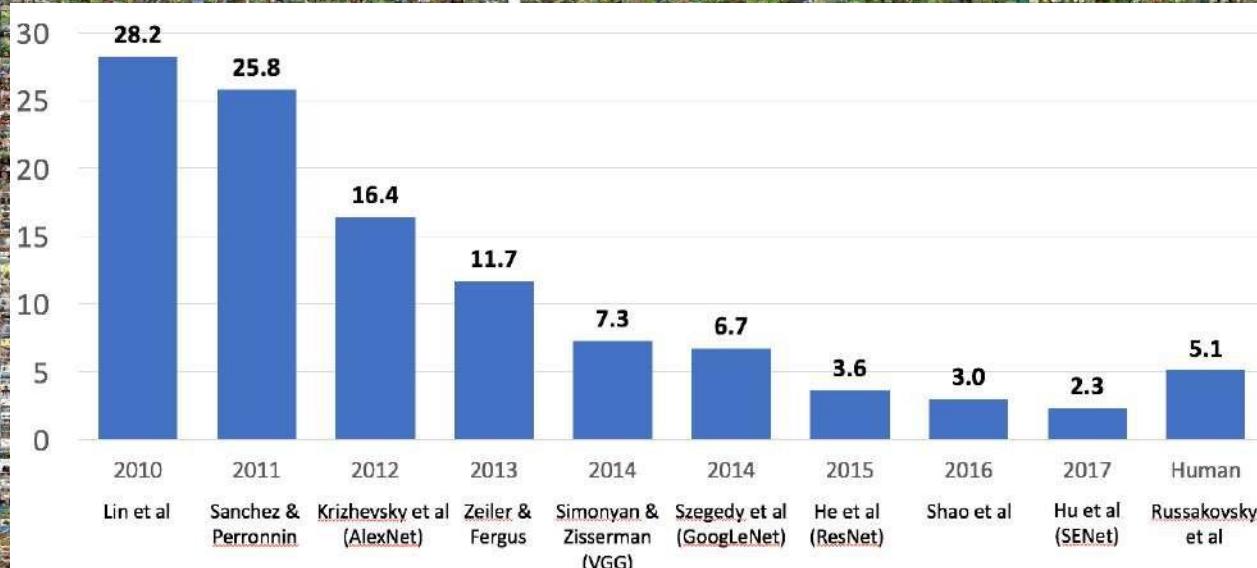


Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

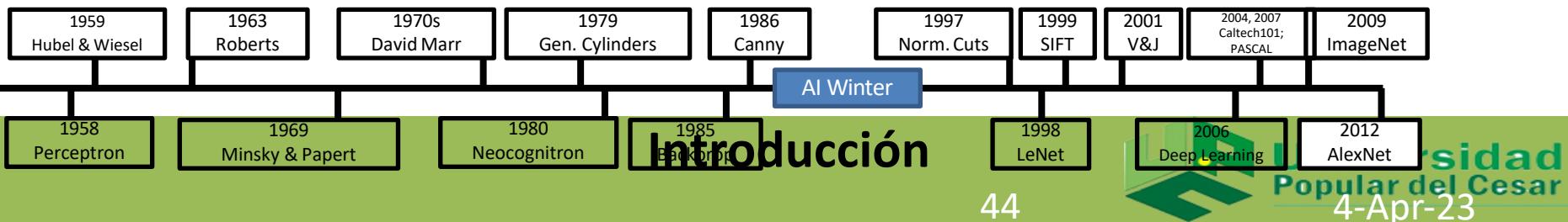
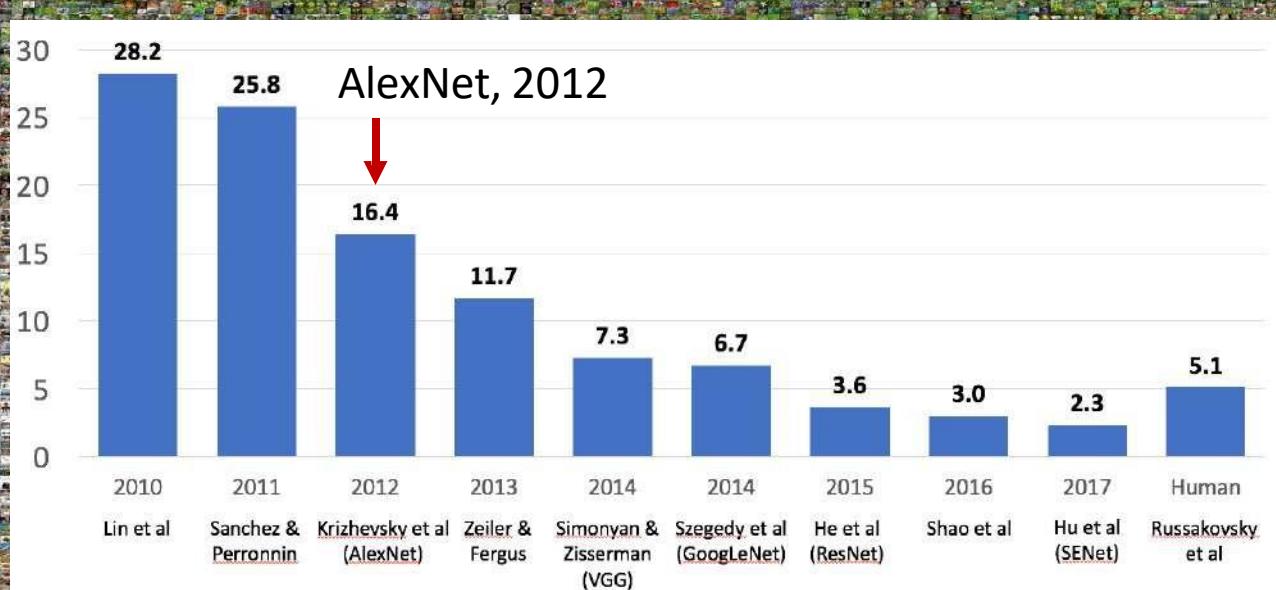
Deng et al, 2009
Russakovsky et al. IJCV 2015



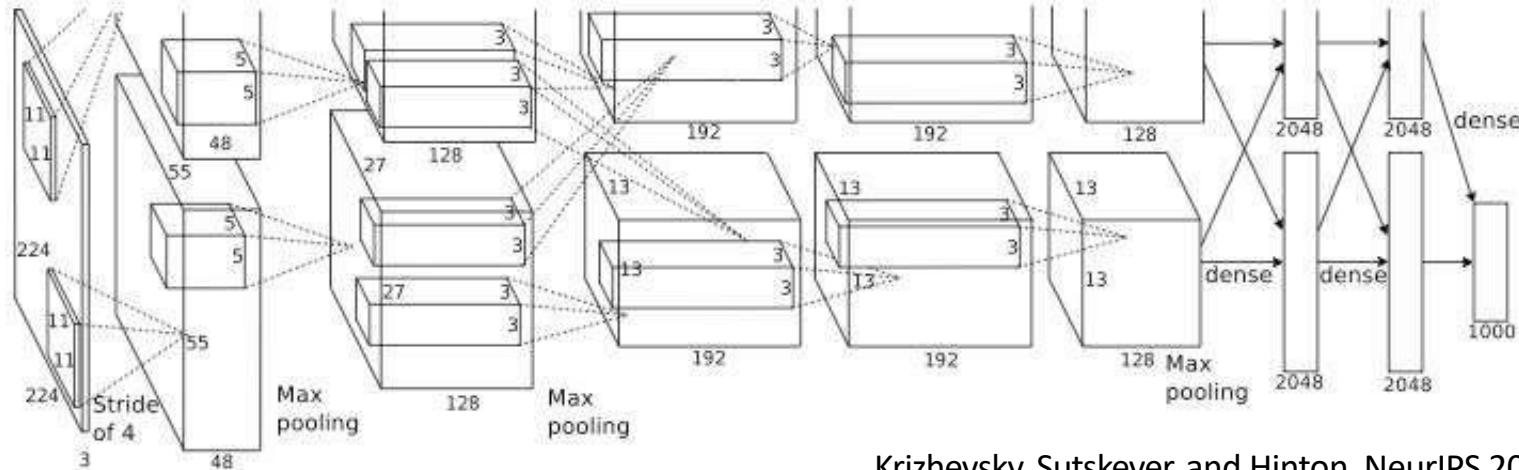
IMAGENET Large Scale Visual Recognition Challenge



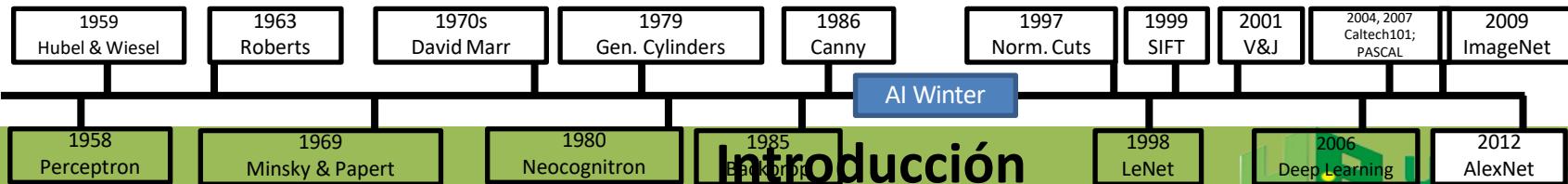
IMAGENET Large Scale Visual Recognition Challenge



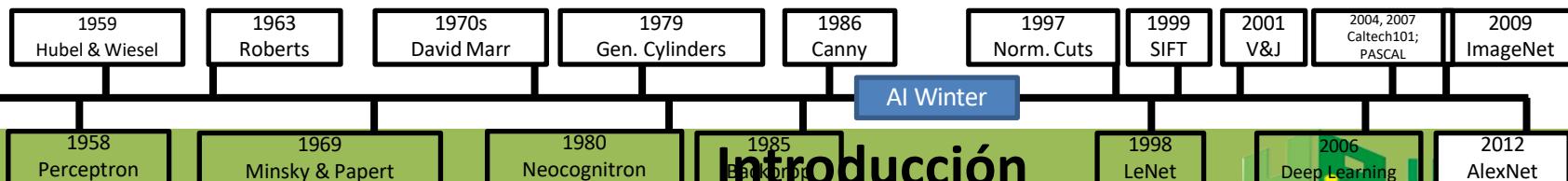
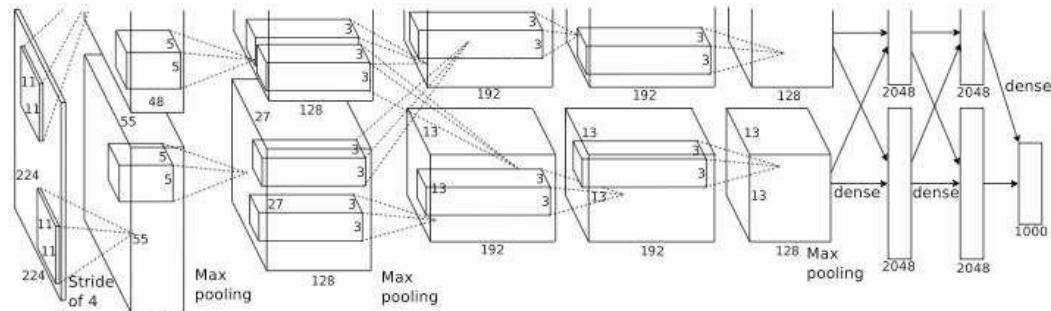
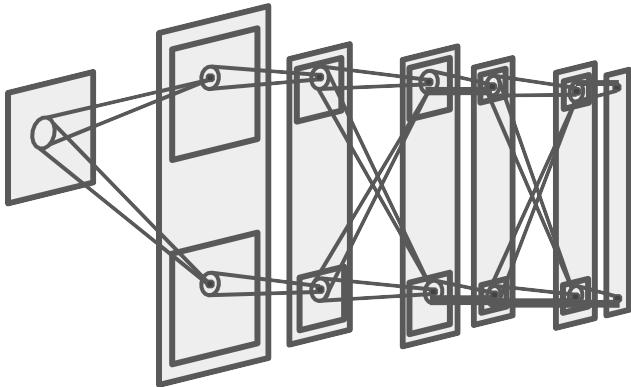
AlexNet: Deep Learning Goes Mainstream



Krizhevsky, Sutskever, and Hinton, NeurIPS 2012



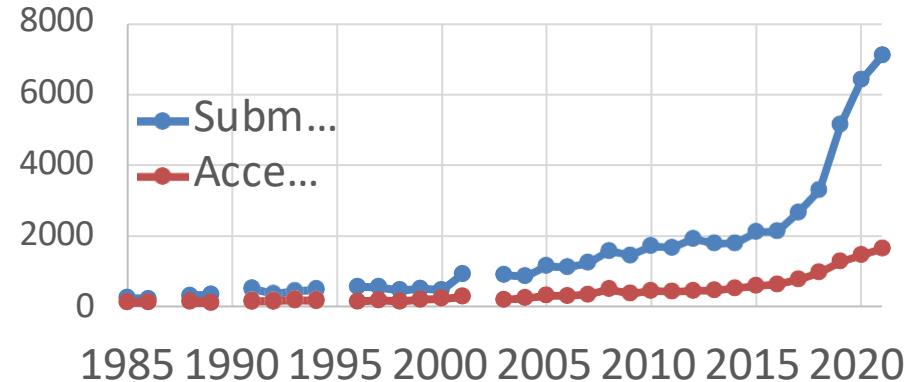
AlexNet vs. Neocognitron: 32 years apart



Introducción

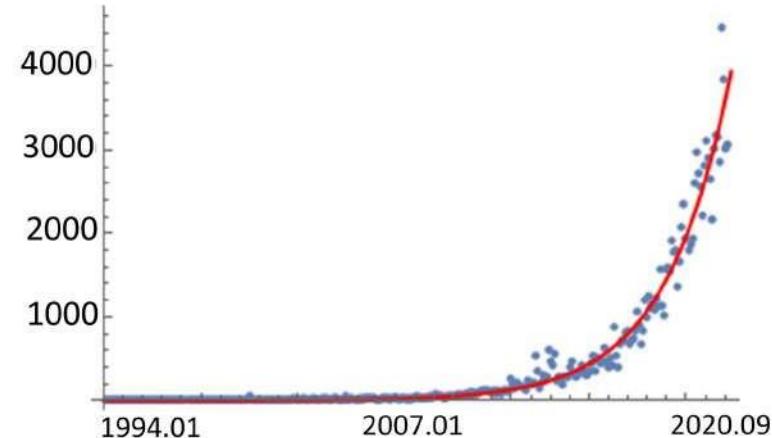
2012 to Present: Deep Learning Explosion

CVPR Papers

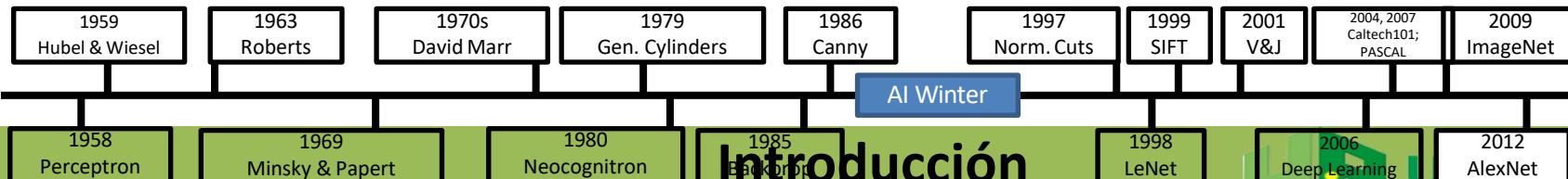


Publications at top Computer Vision conference

ML+AI arXiv papers per month



arXiv papers per month ([source](#))

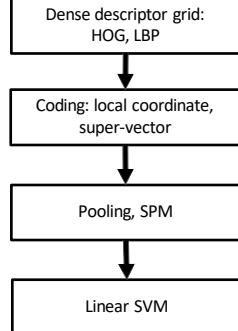


Introducción

2012 to Present: Deep Learning is Everywhere

Year 2010

NEC-UIUC

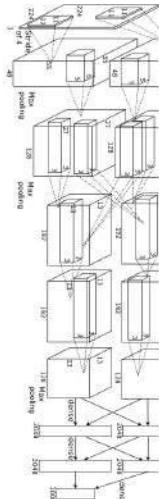


[Lin CVPR 2011]

Lion image by Swissfrog
is
licensed under CC BY 3.0

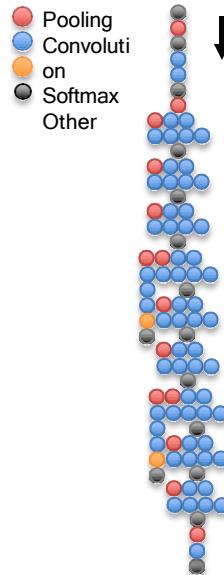
Year 2012

SuperVision



[Krizhevsky NIPS 2012]

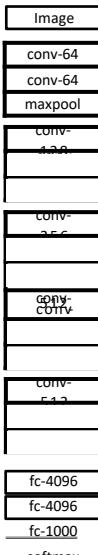
Year 2014
GoogLeNet



[Szegedy arxiv 2014]

Introducción

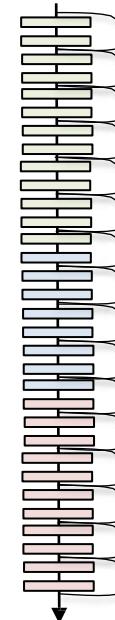
VGG



[Simonyan arxiv 2014]

Year 2015

MSRA



[He ICCV 2015]

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

2012 to Present: Deep Learning is Everywhere

Image Classification



Image Retrieval

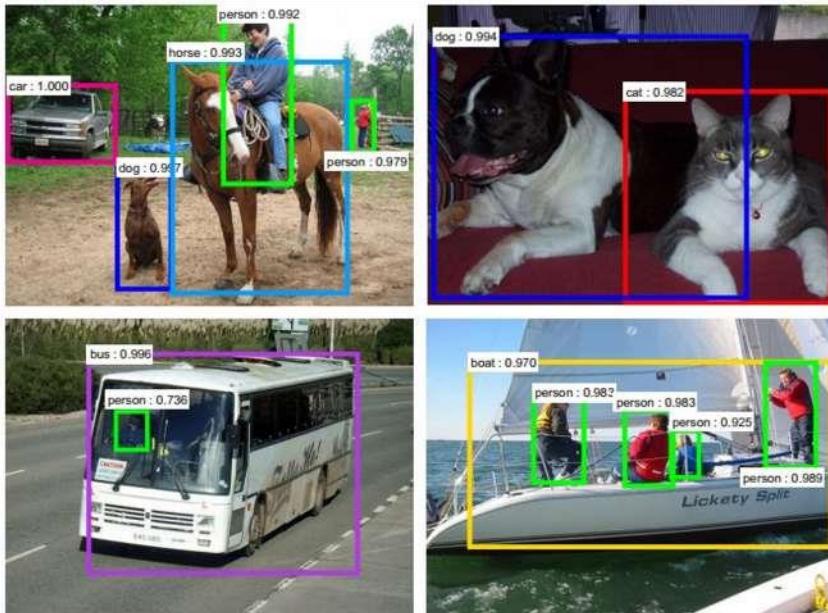


Figures copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

Introducción

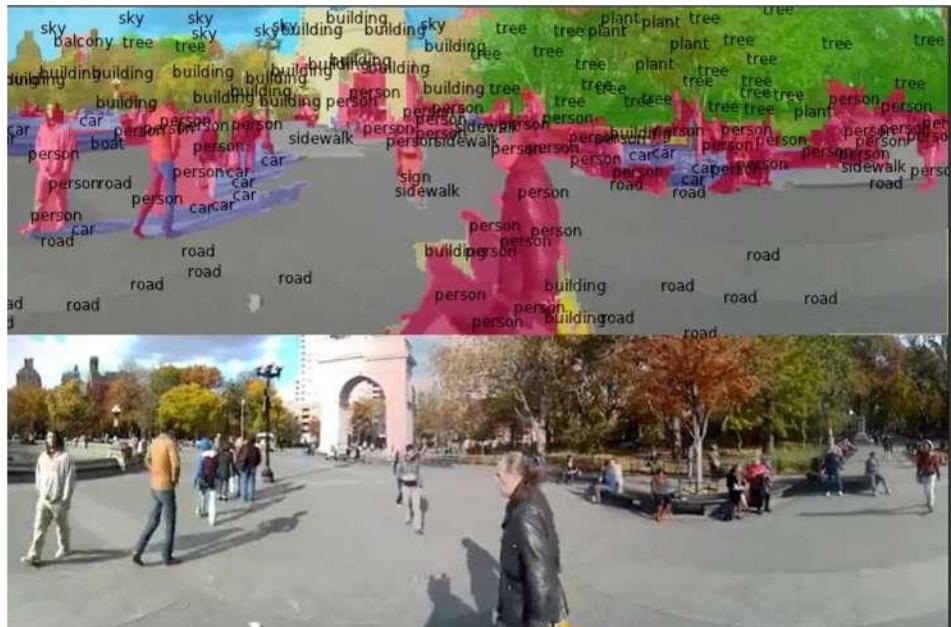
2012 to Present: Deep Learning is Everywhere

Object Detection



Ren, He, Girshick, and Sun, 2015

Image Segmentation

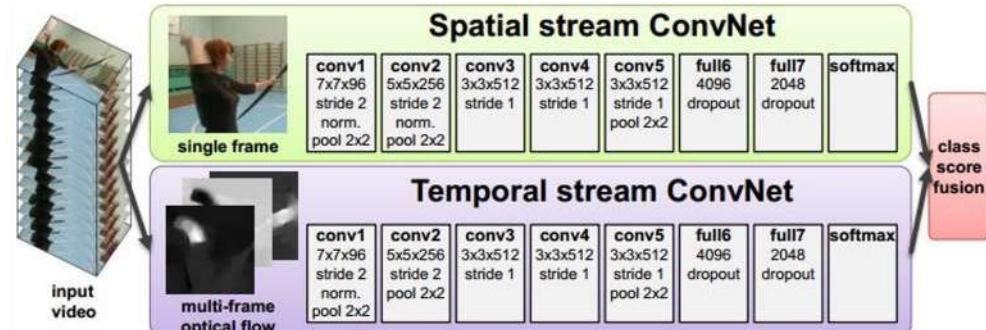


Fabaret et al, 2012

Introducción

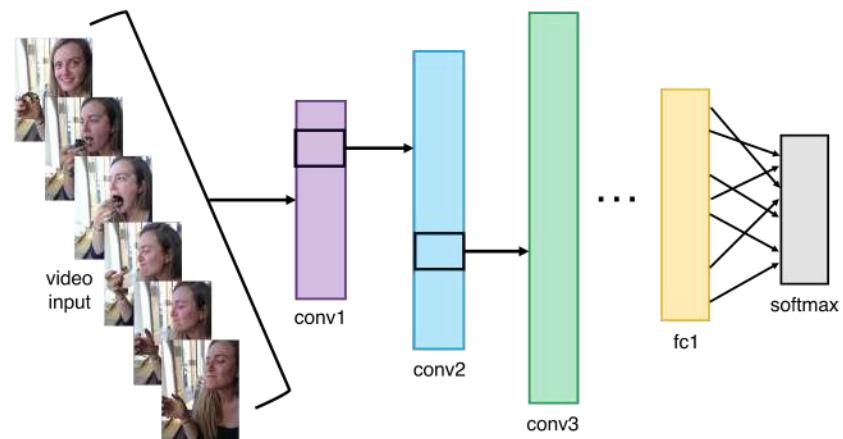
2012 to Present: Deep Learning is Everywhere

Video Classification



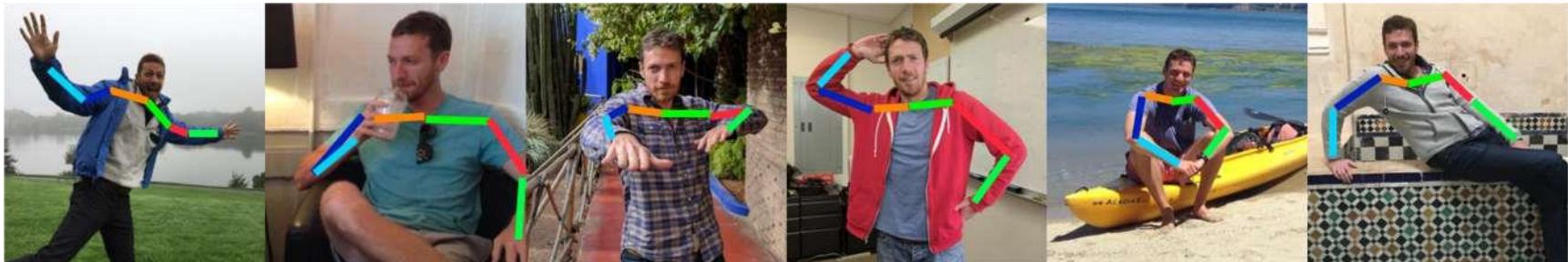
Simonyan et al, 2014

Activity Recognition

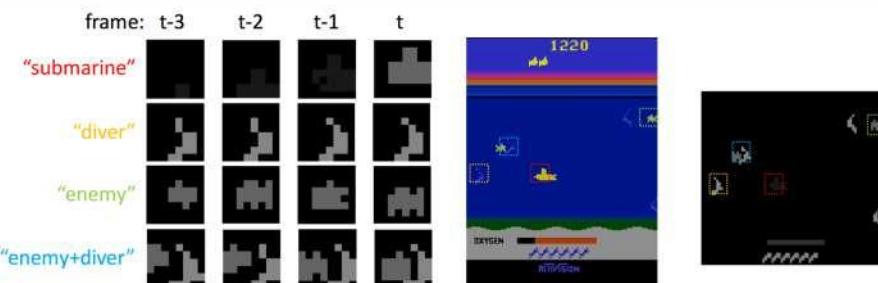


2012 to Present: Deep Learning is Everywhere

Pose Recognition (Toshev and Szegedy, 2014)



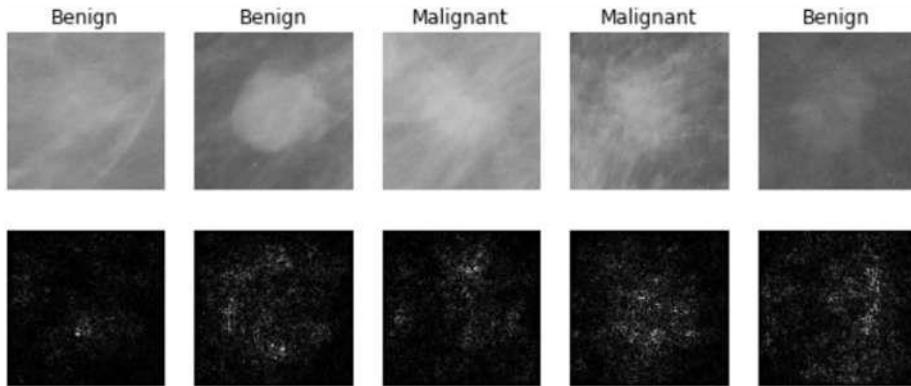
Playing Atari games (Guo et al, 2014)



Introducción

2012 to Present: Deep Learning is Everywhere

Medical Imaging



Levy et al, 2016

Figure reproduced with permission

Whale recognition



Galaxy Classification



Dieleman et al, 2014

From left to right: public domain by NASA; image submitted by
ESA/Hubble, public domain by NASA; image submitted by
Public domain

Introducción

Kaggle Challenge

51



This image by Christin Khan is in the public domain and originally came from the U.S. NOAA

Universidad
Popular del Cesar

4-Apr-23

2012 to Present: Deep Learning is Everywhere



*A white teddy bear
sitting in the grass*



*A man in a baseball
uniform throwing a ball*



*A woman is holding
a cat in her hand*



*A man riding a wave
on top of a surfboard*



*A cat sitting on a
suitcase on the floor*



*A woman standing on a
beach holding a surfboard*

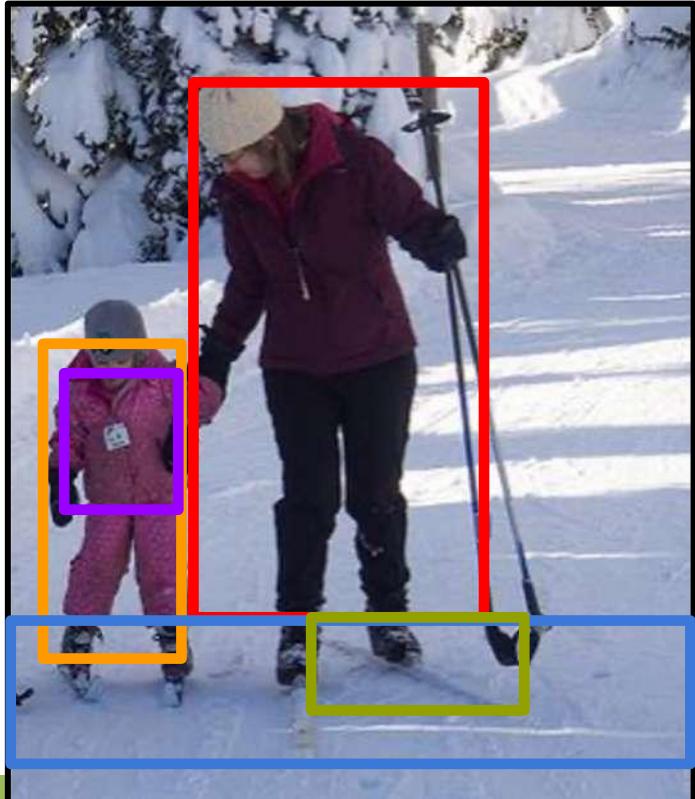
Image Captioning
Vinyals et al, 2015
Karpathy and Fei-Fei,
2015

All images are CC0 Public domain:
<https://pixabay.com/en/luggage-antique-cat-1643010/>
<https://pixabay.com/en/teddy-plush-bears-cute-teddy-bear-1673436/>
<https://pixabay.com/en/sun-wave-summer-sport-litral-1668716/>
<https://pixabay.com/en/woman-female-model-portrait-adult-983967/>
<https://pixabay.com/en/baseball-player-shortstop-infield-1045263/>
<https://pixabay.com/en/surfing-surfboard-ocean-waves-1045263/>

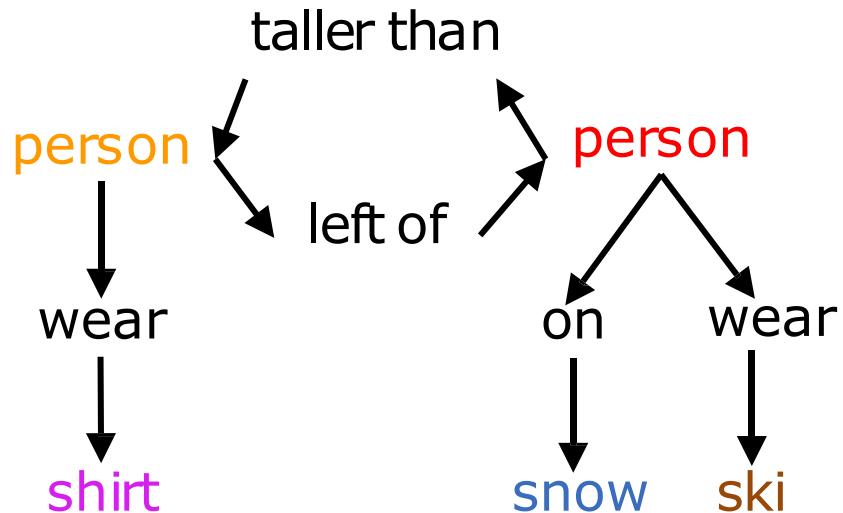
Captions generated by Justin Johnson using NeuralTalk2

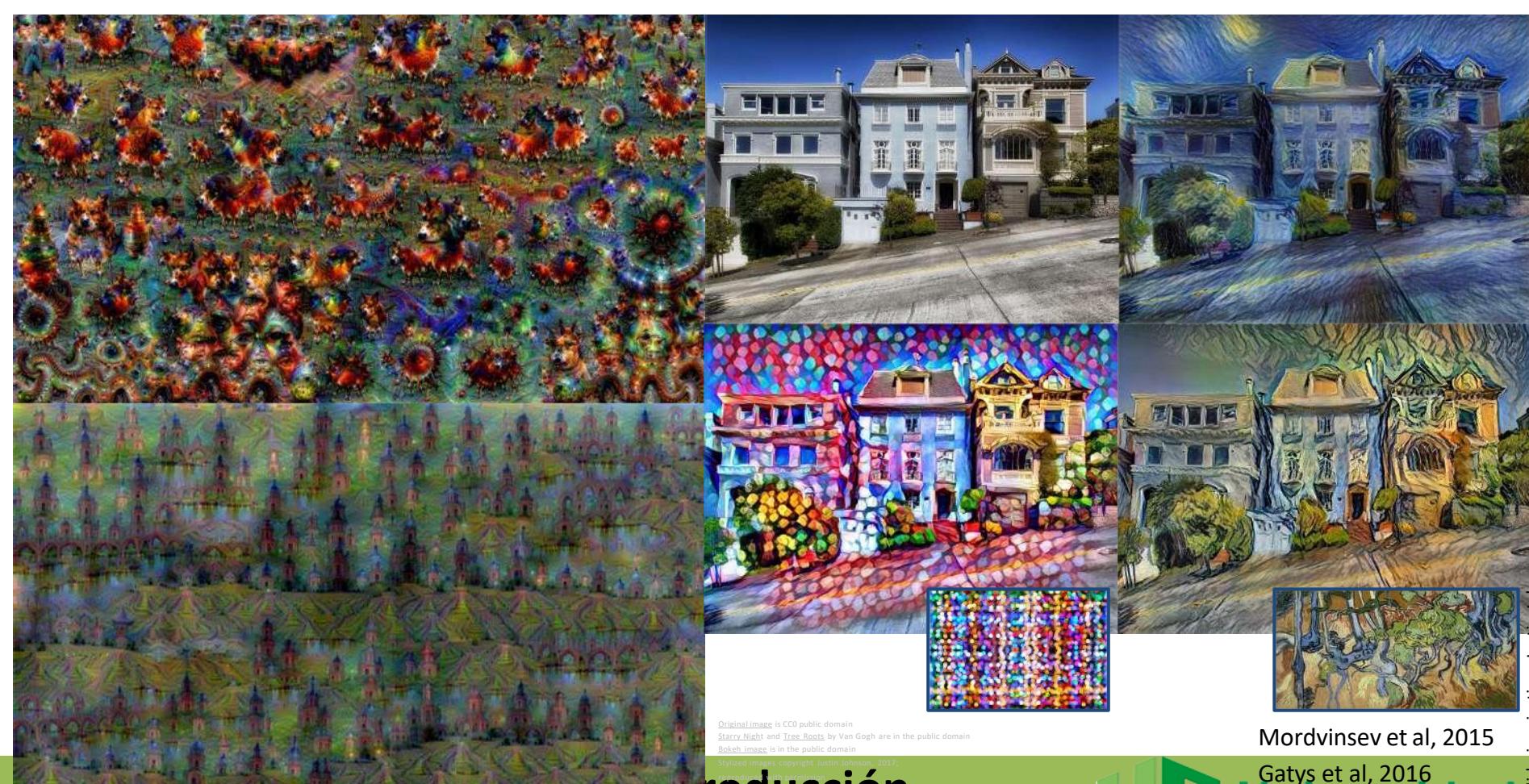
Introducción

2012 to Present: Deep Learning is Everywhere



Results:
spatial, comparative, asymmetrical, verb,
prepositional





Introducción

Figures copyright Justin Johnson, 2015. Reproduced with permission. Generated using the Inceptionism approach from a blog post by Google Research.

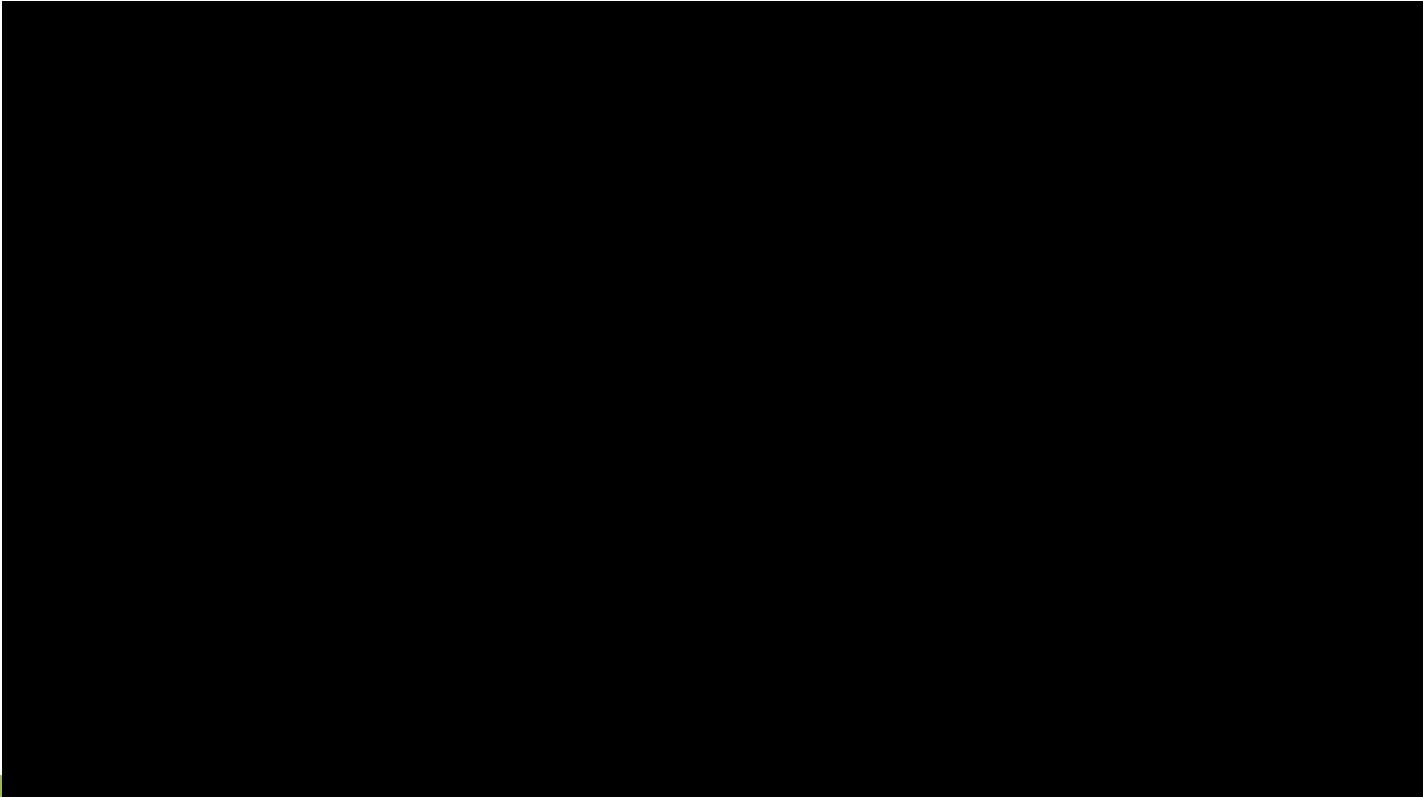
Mordvinsev et al, 2015

Gatys et al, 2016

Universidad
Popular del Cesar

4-Apr-23

2012 to Present: Deep Learning is Everywhere



2012 to Present: Deep Learning is Everywhere

TEXT PROMPT

an armchair in the shape of an avocado. an armchair imitating an avocado.

AI-GENERATED IMAGES



Slide inspiration: Justin Johnson

2012 to Present: Deep Learning is Everywhere

TEXT PROMPT

an armchair in the shape of a peach. an armchair imitating a peach.

AI-GENERATED IMAGES

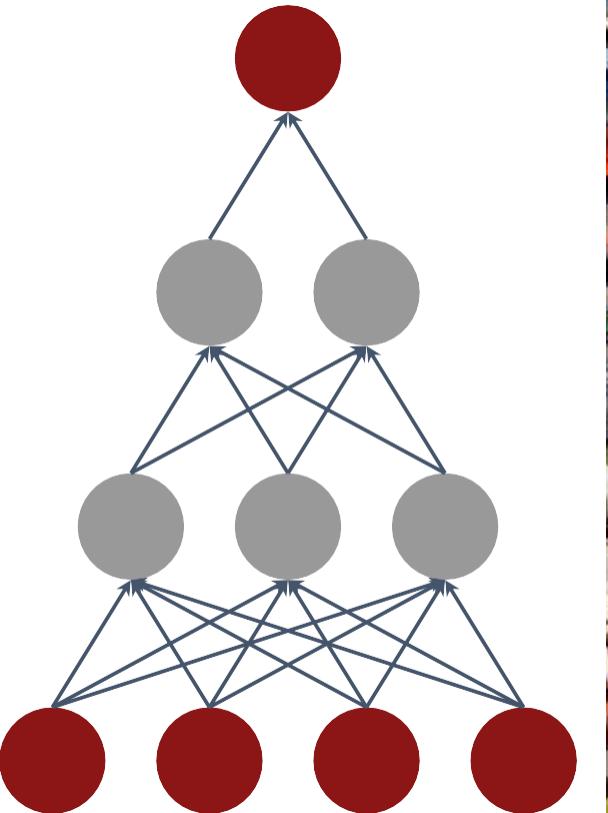


Slide inspiration: Justin Johnson



Computation

4 Apr 23

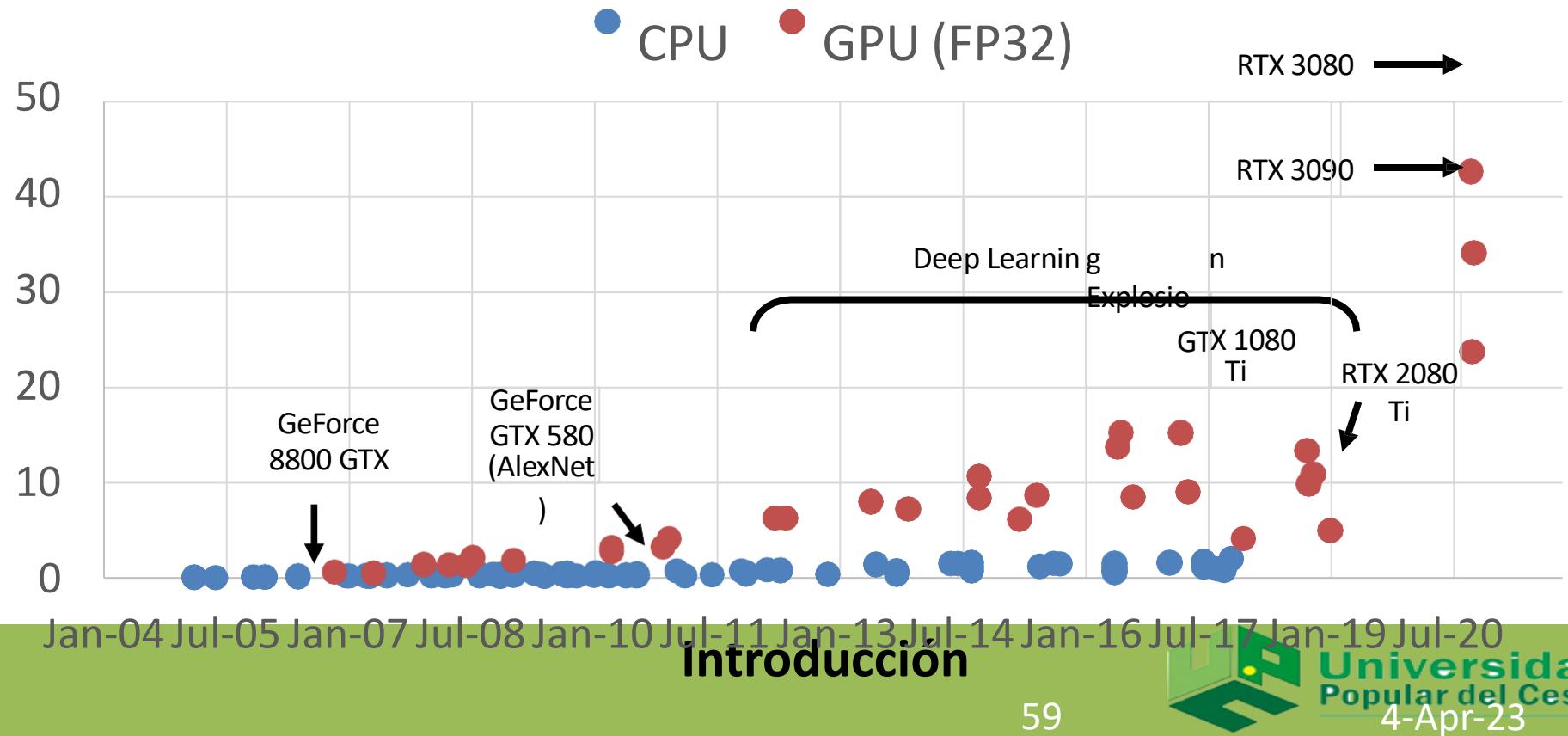


Algorithms



Data

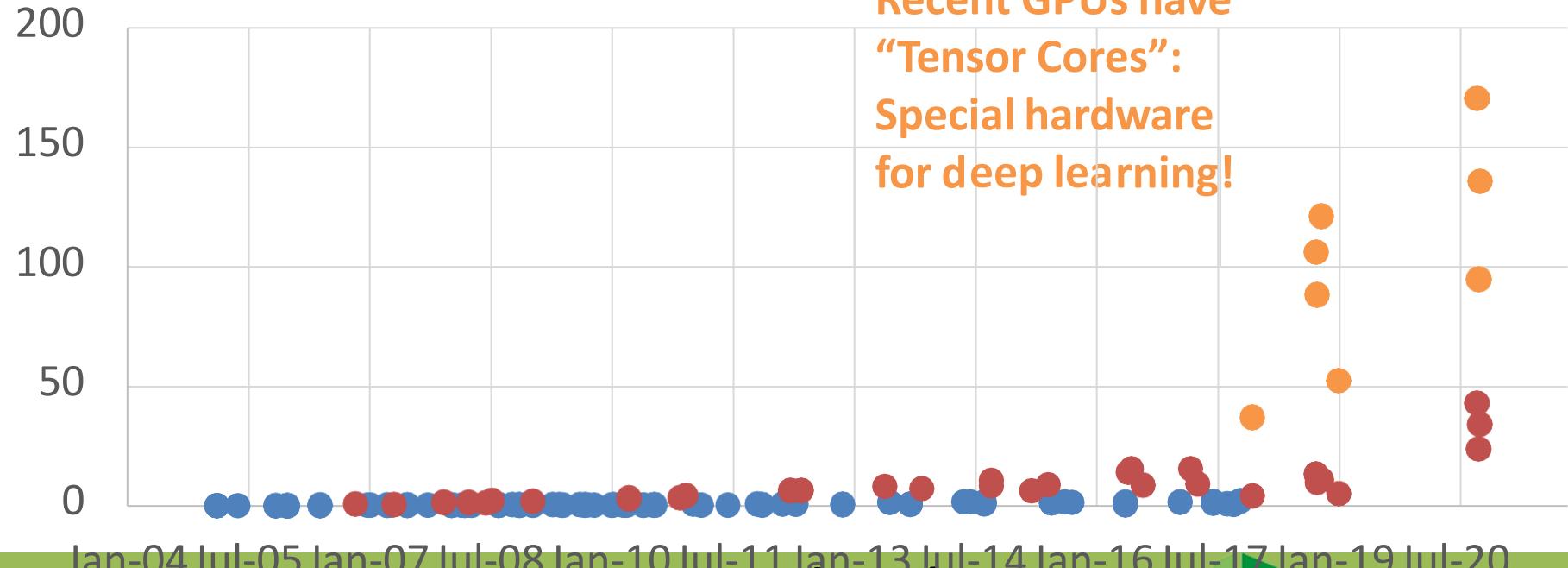
GFLOP per Dollar



GFLOP per Dollar

● CPU ● GPU (FP32) ● GPU (Tensor Core)

Recent GPUs have
“Tensor Cores”:
Special hardware
for deep learning!



Introducción

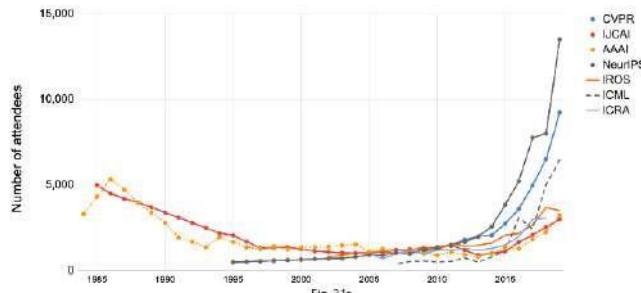


Universidad
Popular del Cesar

4-Apr-23

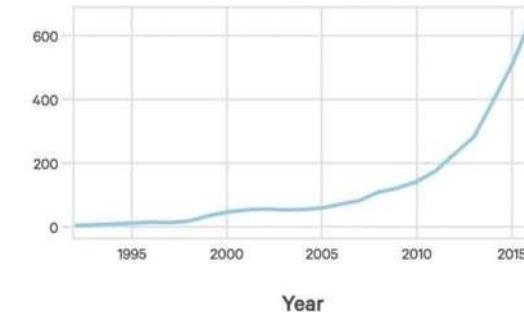
AI's Explosive Growth & Impact

Attendance at large conferences (1984-2019)
Source: Conference provided data.



Number of attendance
At AI conferences

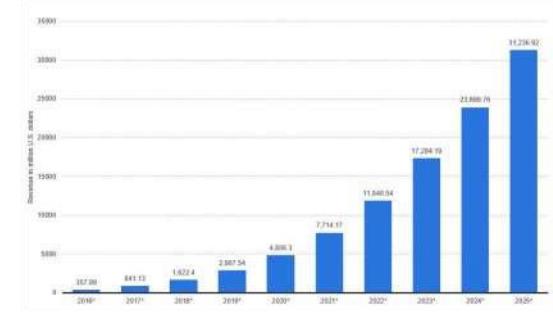
Source: The Gradient



Startups Developing AI
Systems

Source: Crunchbase, VentureSource, Sand
Hill Econometrics

Introducción



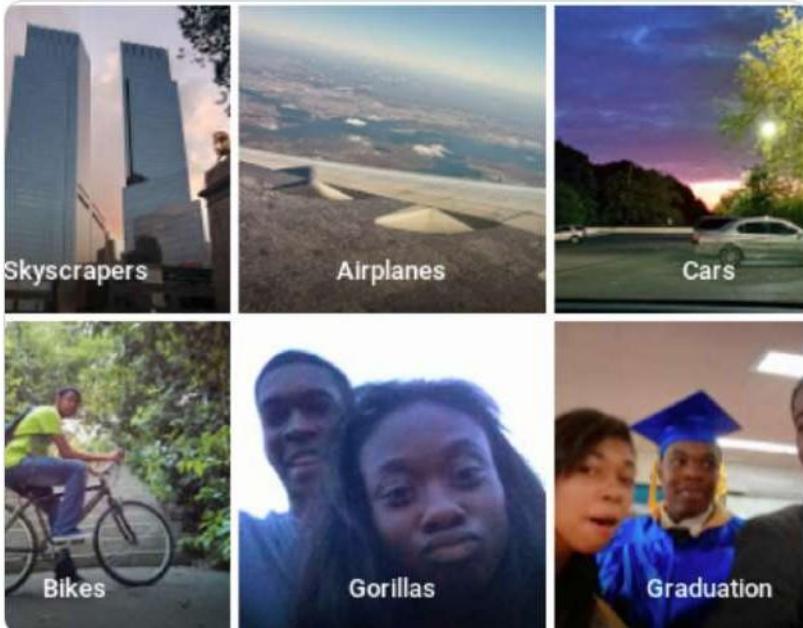
Enterprise Application AI
Revenue

Source: Statista

Despite the successes, computer vision still has a long way to go

Computer Vision Can Cause Harm

Harmful Stereotypes

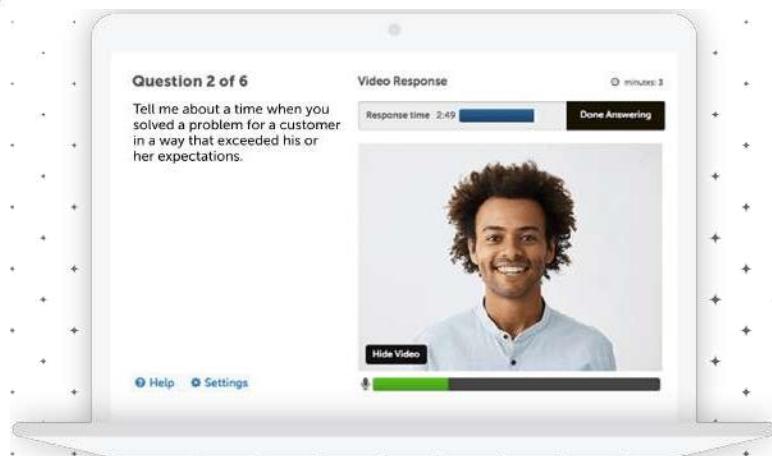


Affect people's lives

Technology

A face-scanning algorithm increasingly decides whether you deserve the job

HireVue claims it uses artificial intelligence to decide who's best for a job. Outside experts call it 'profoundly disturbing.'



Barocas et al., "The Problem With Bias: Allocative Versus Representational Harms in Machine Learning", SIGCIS 2017

Kate Crawford, "The Trouble with Bias", NeurIPS 2017 Keynote

Source: <https://twitter.com/jackyallene/status/615329515909156865> (2015)

Introducción

source: <https://www.washingtonpost.com/technology/2019/10/22/ai-hires-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/>

<https://www.hirevue.com/> is a video interviewing software

Example Credit: Timnit Gebru

Computer Vision Can Save Lives

**How to take care of seniors
while keeping them safe?**



Early Symptom Detection
of COVID-19



Monitor Patients with
Mild Symptoms



Manage Chronic Conditions

Versatile



Mobility



Infection



Sleep



Diet



Scalable



Low-cost



Burden-free



And there is a lot we don't know how to do



https://fedandfit.com/wp-content/uploads/2020/06/summer-activities-for-kids_optimized-scaled.jpeg



This image is
copyright-free [United States government work](#)

Introducción