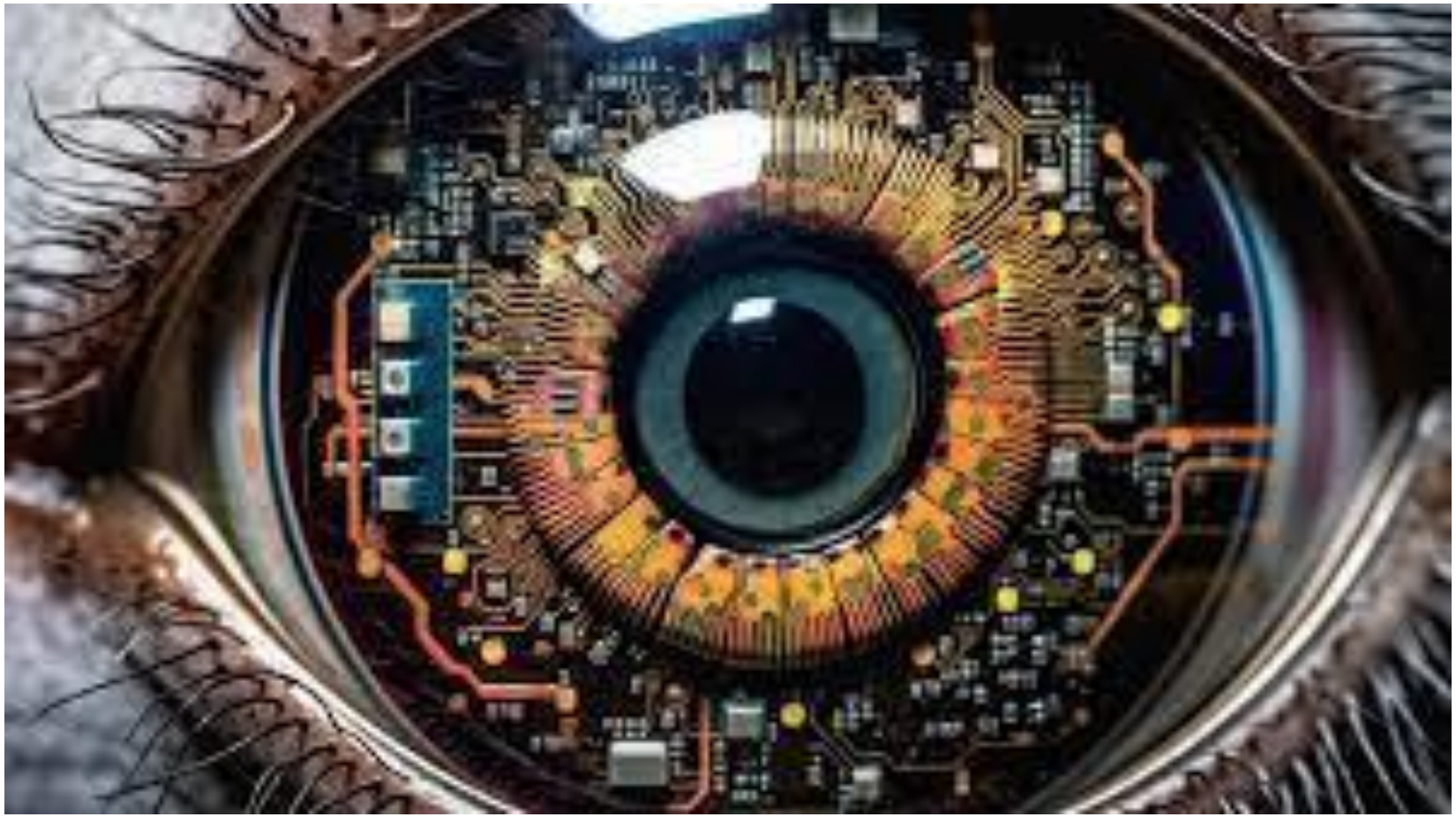
INTRODUCTION

- ThS. Đoàn Chánh Thống
- ThS. Nguyễn Cường Phát
- ThS. Nguyễn Hữu Lợi
- ThS. Trương Quốc Dũng
- ThS. Nguyễn Thành Hiệp
- ThS. Võ Duy Nguyên
- ThS. Nguyễn Văn Toàn
- ThS. Lê Ngô Thục Vi
- TS. Nguyễn Duy Khánh
- TS. Nguyễn Tấn Trần Minh Khang

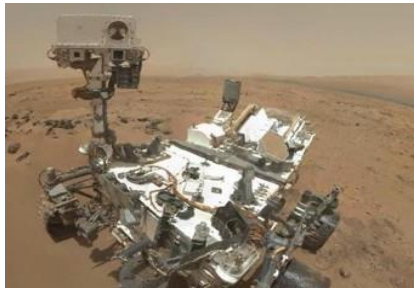
Welcome to CS231n



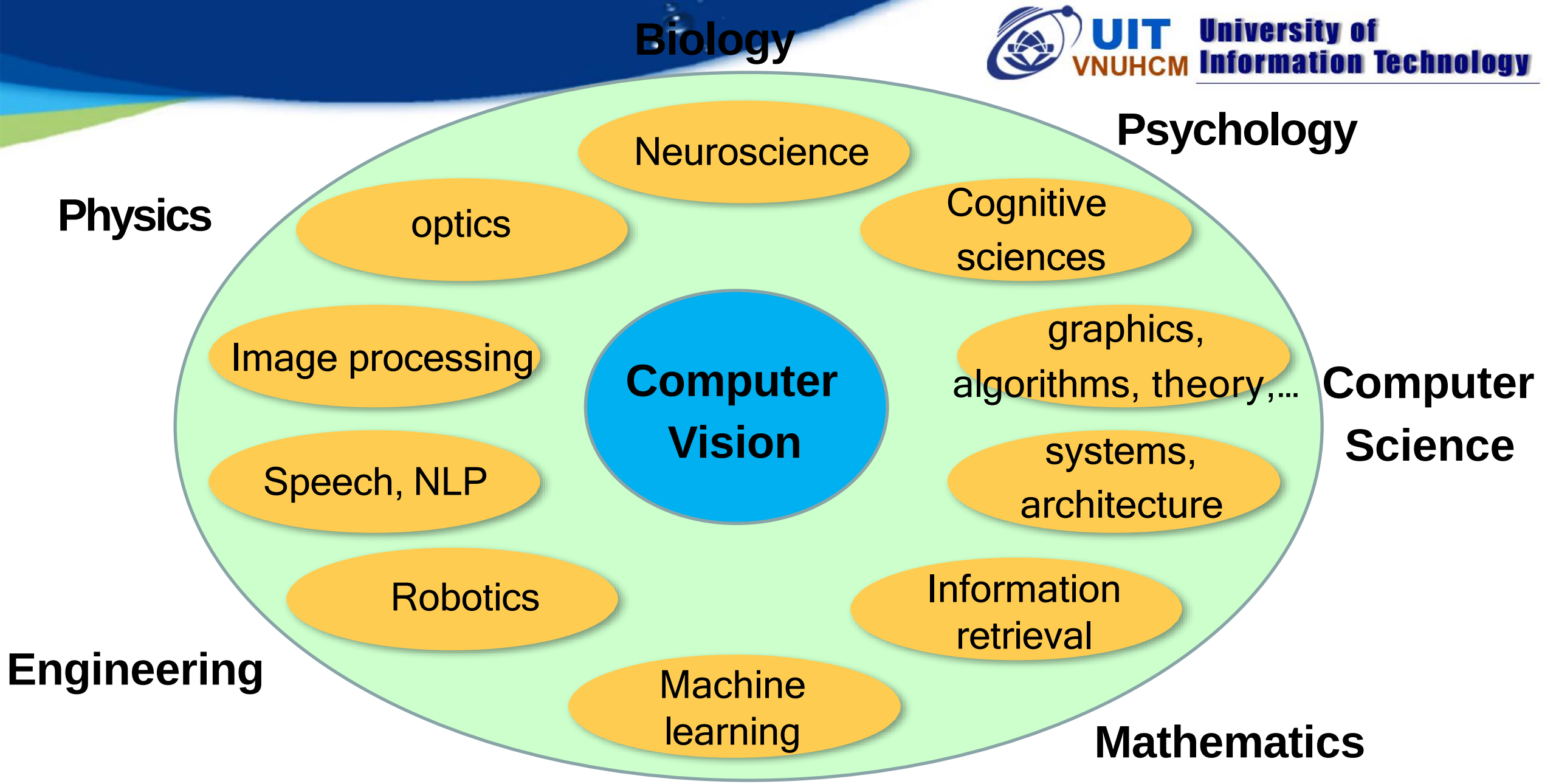
Welcome to CS231n

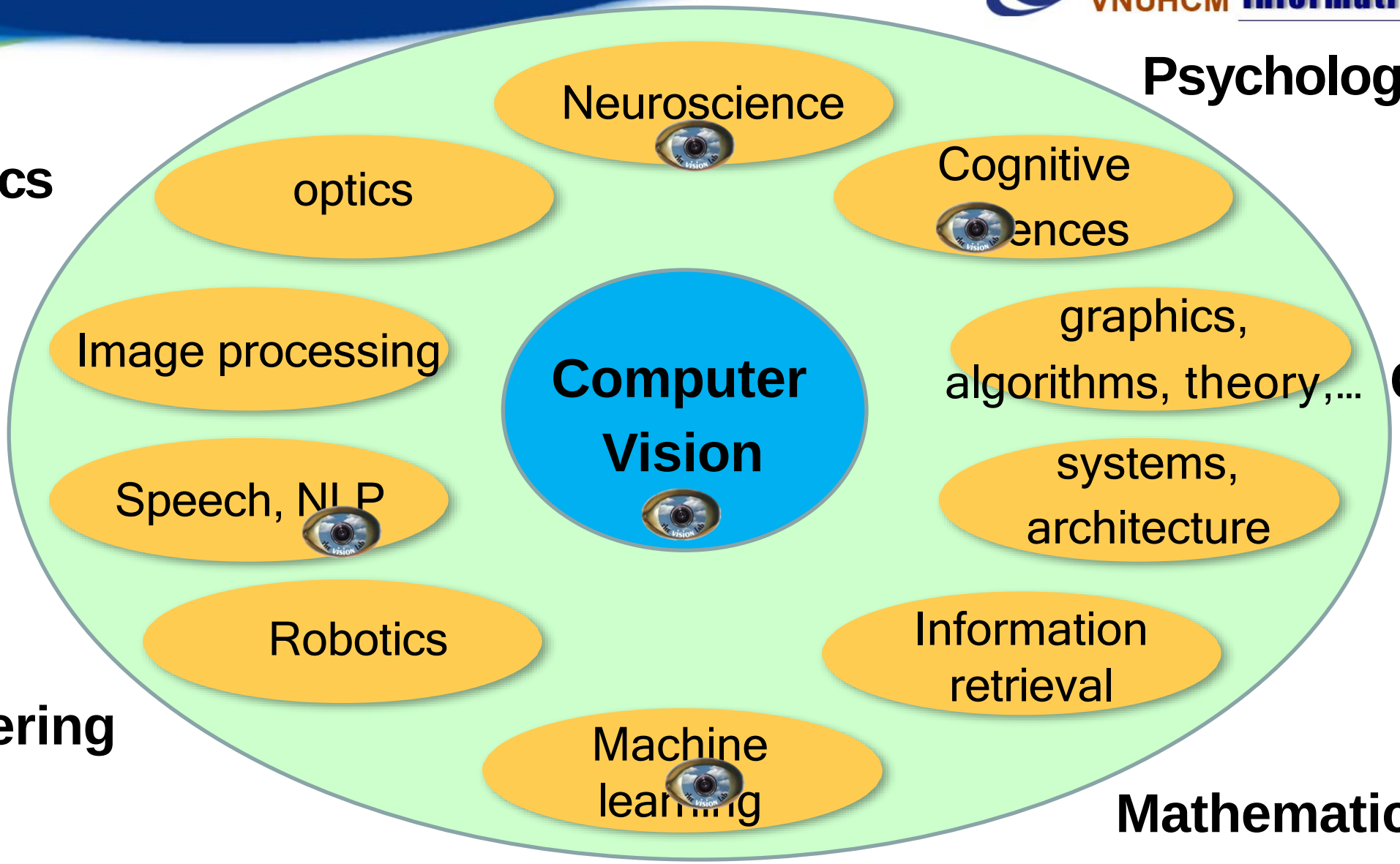






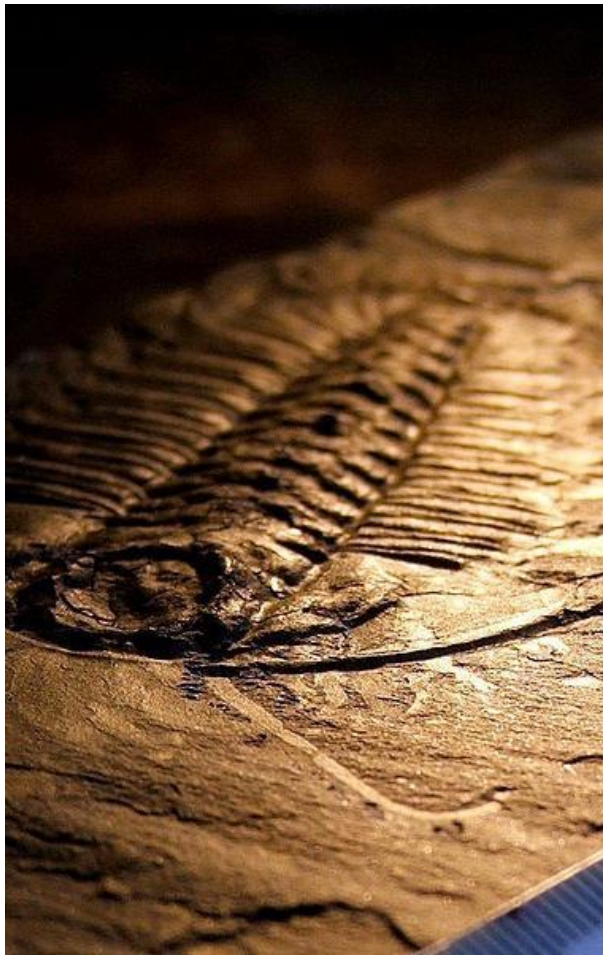


Biology
Psychology
Physics
Neuroscience
optics
Cognitive sciences
Image processing
Computer Vision
graphics, algorithms, theory,...
Computer Science
Speech, NLP
systems, architecture
Robotics
Information retrieval
Engineering
Machine learning
Mathematics


Biology
Psychology
Physics
Neuroscience
optics
Cognitive
ences
Image processing
Computer Vision
graphics, algorithms, theory,...
Computer Science
Speech, NLP
systems, architecture
Robotics
Information retrieval
Machine learning
Engineering
Mathematics

Today's agenda

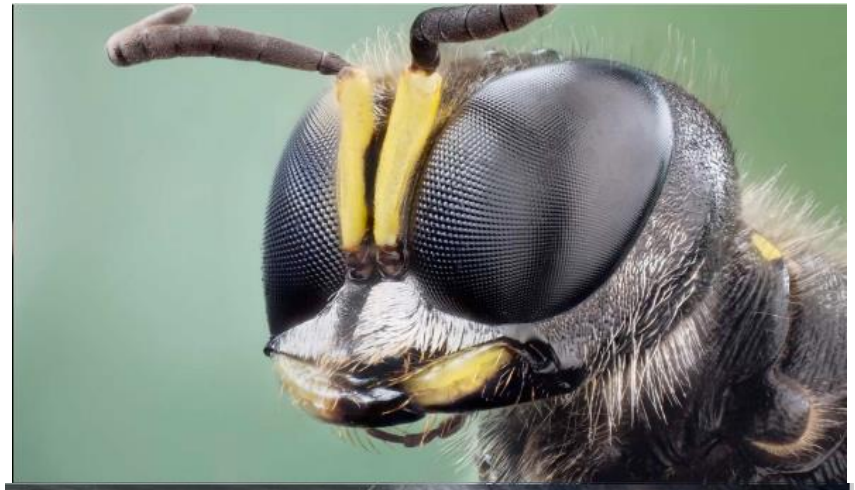
- A brief history of computer vision.
- Một lịch sử ngắn gọn về thị giác máy tính.
- CS231n overview.
- Tổng quan về CS231n.

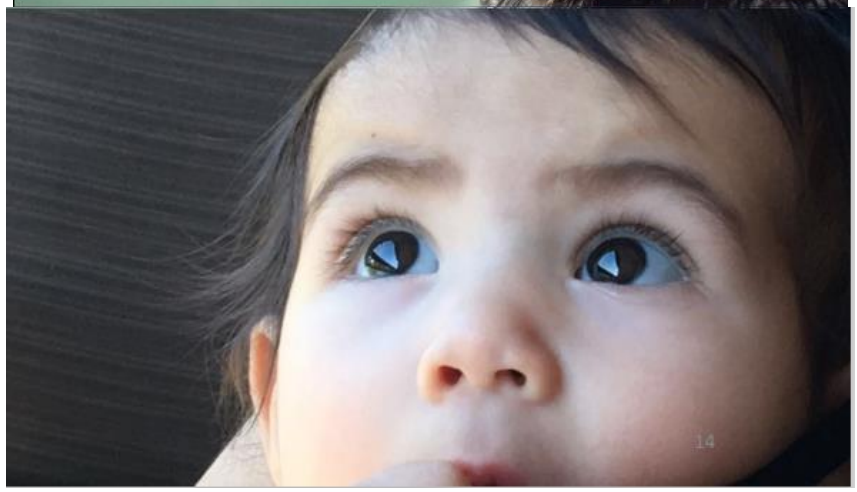
Evolution's Big Bang



- 543 million years, B.C.
- After 10 million years (533 million years, B.C.)

Evolution's Big Bang





Computer Vision is everywhere!





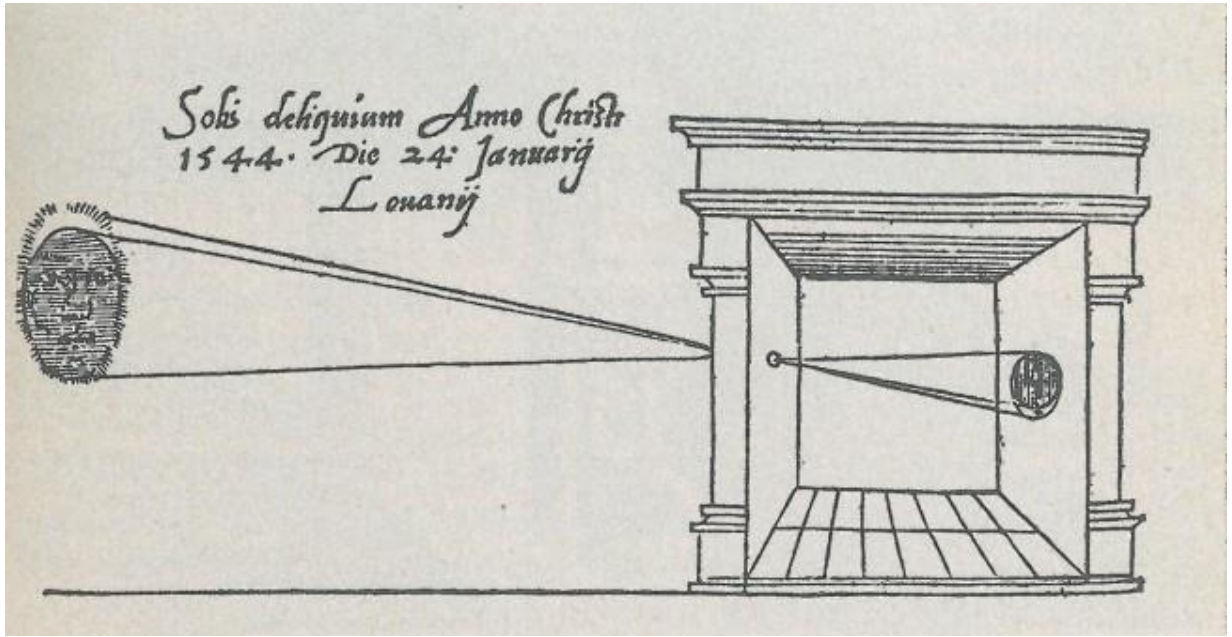






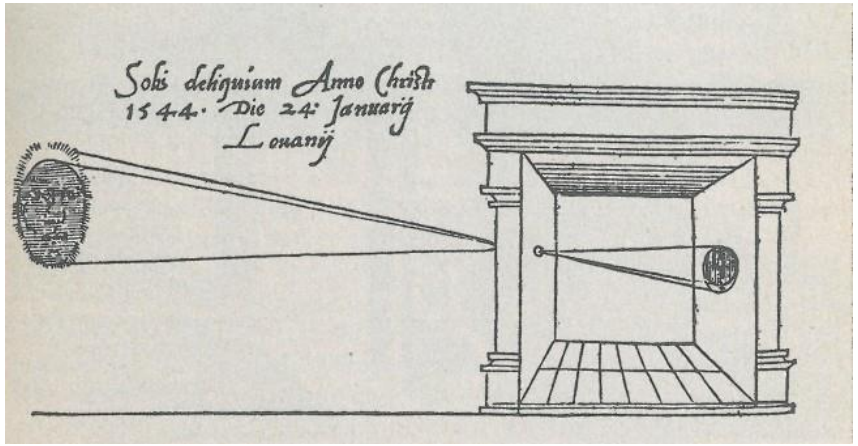
Camera Obscura, 1545

– Camera Obscura by Reinerus Gemma – Frisius in 1545.



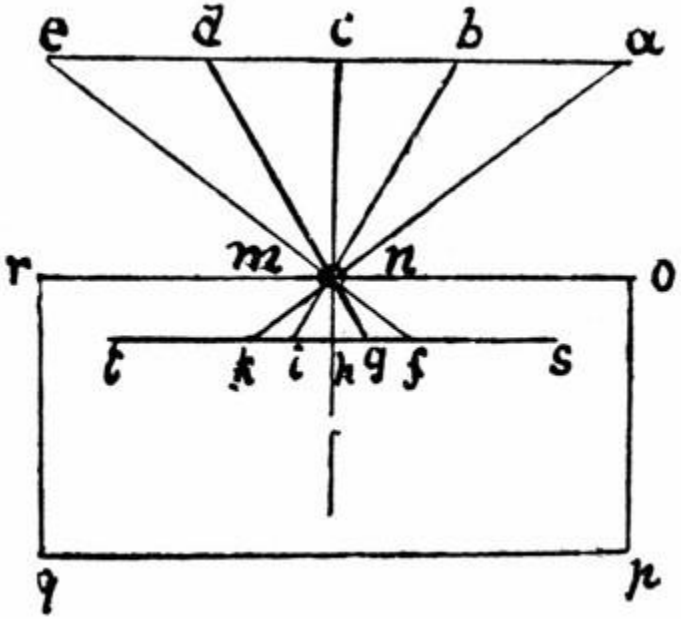
Camera Obscura, 1545

- Camera Obscura, một thuật ngữ tiếng Latin có nghĩa là "phòng tối," là một thiết bị quang học mà từ đó nhiếp ảnh hiện đại phát triển. Đây là một trong những phát minh quan trọng nhất trong lịch sử khoa học và nghệ thuật, cung cấp cơ sở lý thuyết cho các phát triển sau này trong lĩnh vực quang học và nhiếp ảnh.



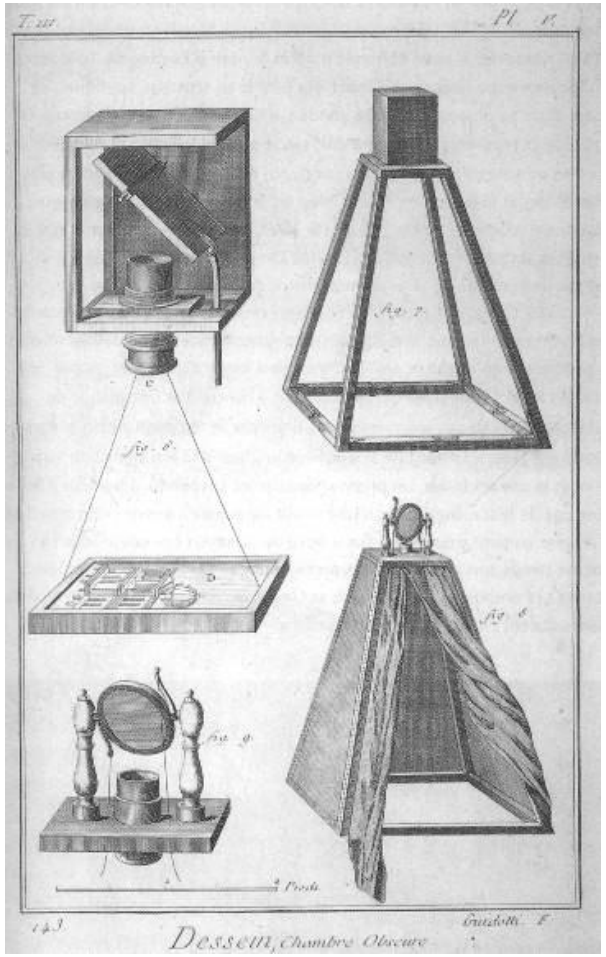
Camera Obscura, 15xx

- Leonardo da Vinci, 16th Century AD.
- Nghiên cứu về nguyên lý của Camera Obscura.
- Leonardo đã miêu tả cách một căn phòng tối có một lỗ nhỏ có thể chiếu hình ảnh của cảnh bên ngoài lên tường đối diện. Ông đã hiểu rằng ánh sáng di chuyển theo đường thẳng và tạo ra hình ảnh lộn ngược của cảnh bên ngoài.

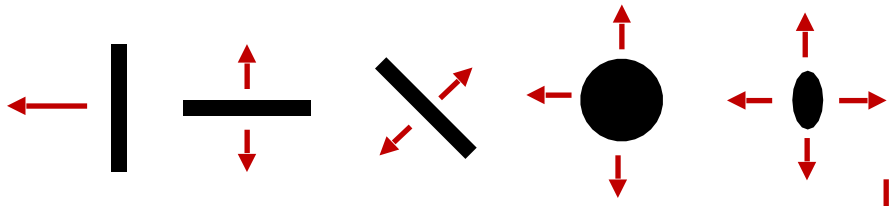


Camera Obscura, 18xx

- Encyclopedie, 18th Century
- Camera Obscura, một thiết bị quang học có từ thời cổ đại, đã được ghi chép và phổ biến rộng rãi trong suốt nhiều thế kỷ. Đến thế kỷ 18, thiết bị này đã được mô tả chi tiết trong "Encyclopédie," một bộ bách khoa toàn thư nổi tiếng được biên soạn bởi Denis Diderot và Jean le Rond d'Alembert.



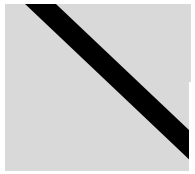
Hubel & Wiesel, 1959



Simple cells: Response to light orientation

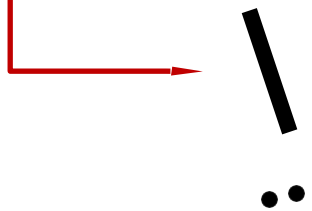
Complex cells: Response to light orientation and movement

Hypercomplex cells: Response to movement with an end point

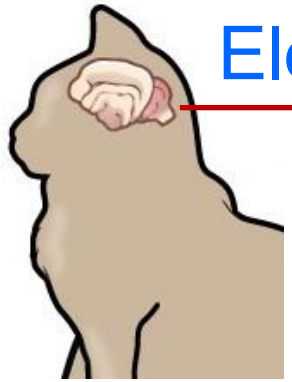


No response

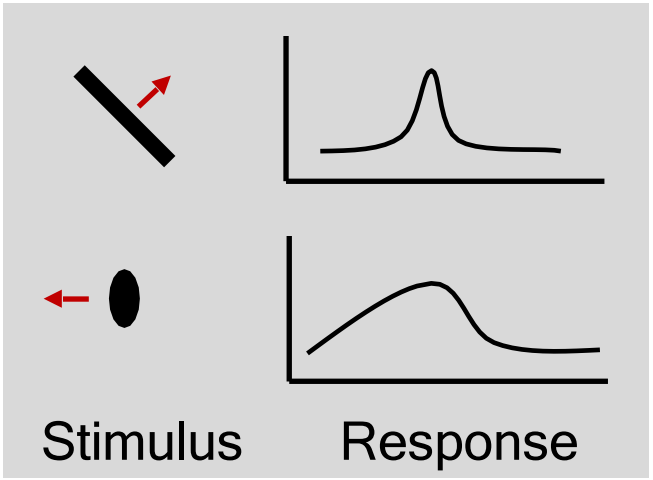
Response (end point)



Stimulus



Electrical signal from brain

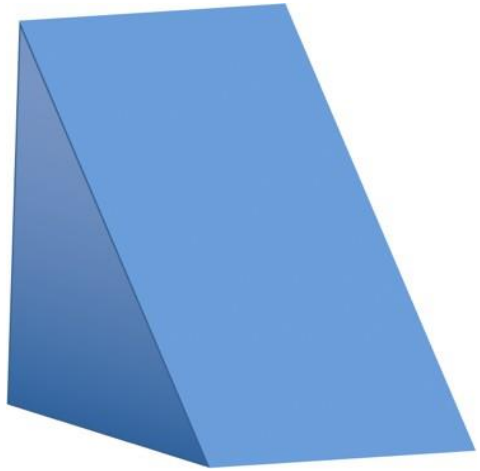


First Thesis, 1963

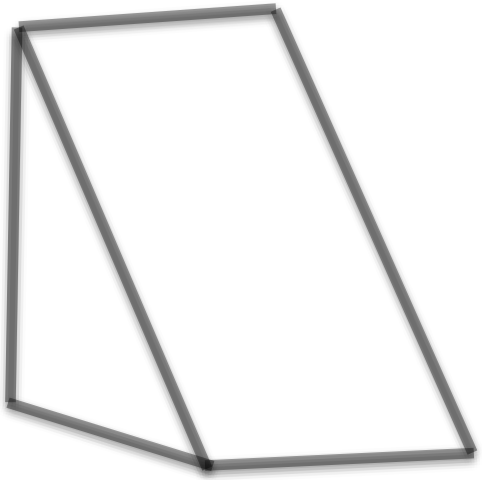
- In May 1963, MIT graduate student Larry Roberts submitted a PhD thesis outlining how machines can perceive solid three-dimensional objects by breaking them down into simple two-dimensional figures.

First Thesis, 1963

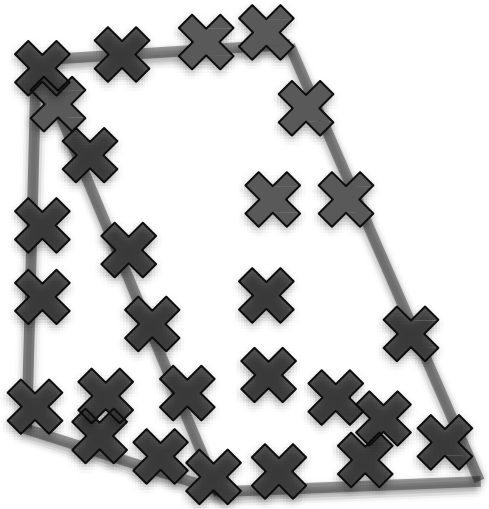
- Larry Roberts, 1963
- Block world



(a) Original picture

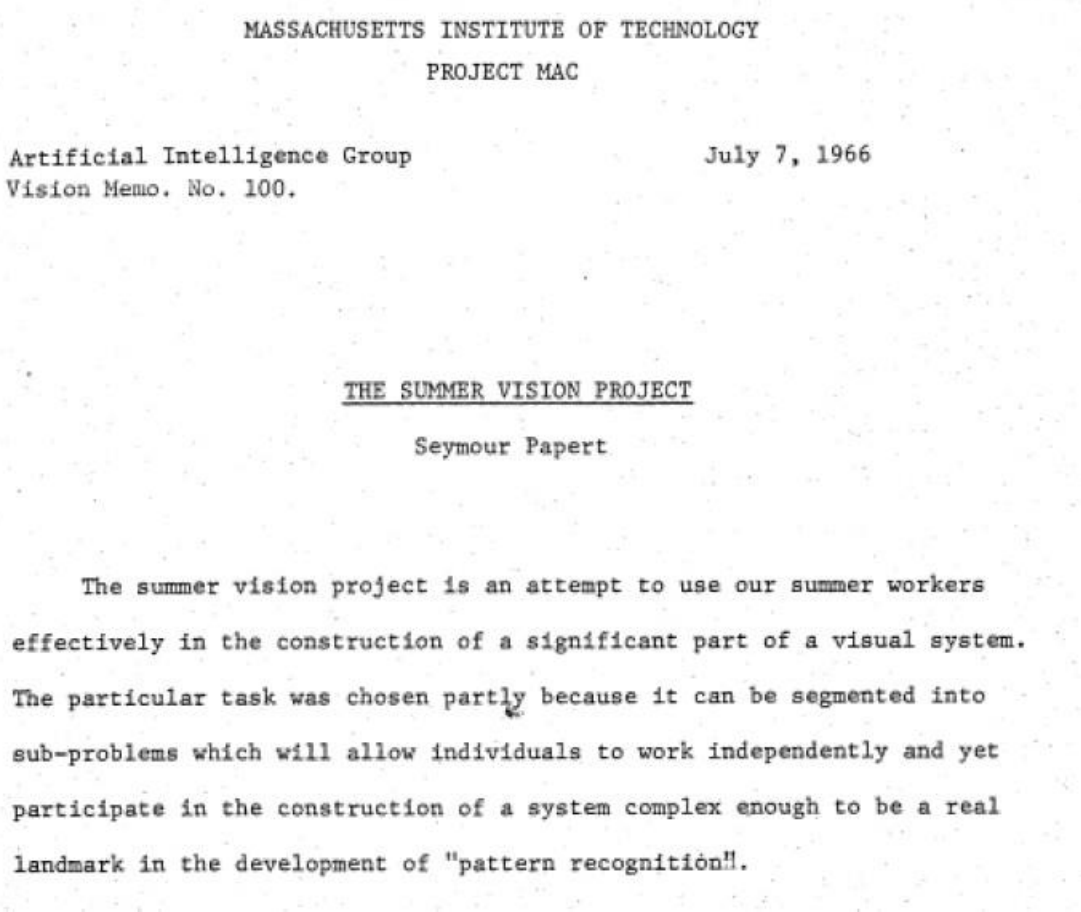


(b) Differentiated picture



(c) Feature points selected

First Project, 1966



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

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

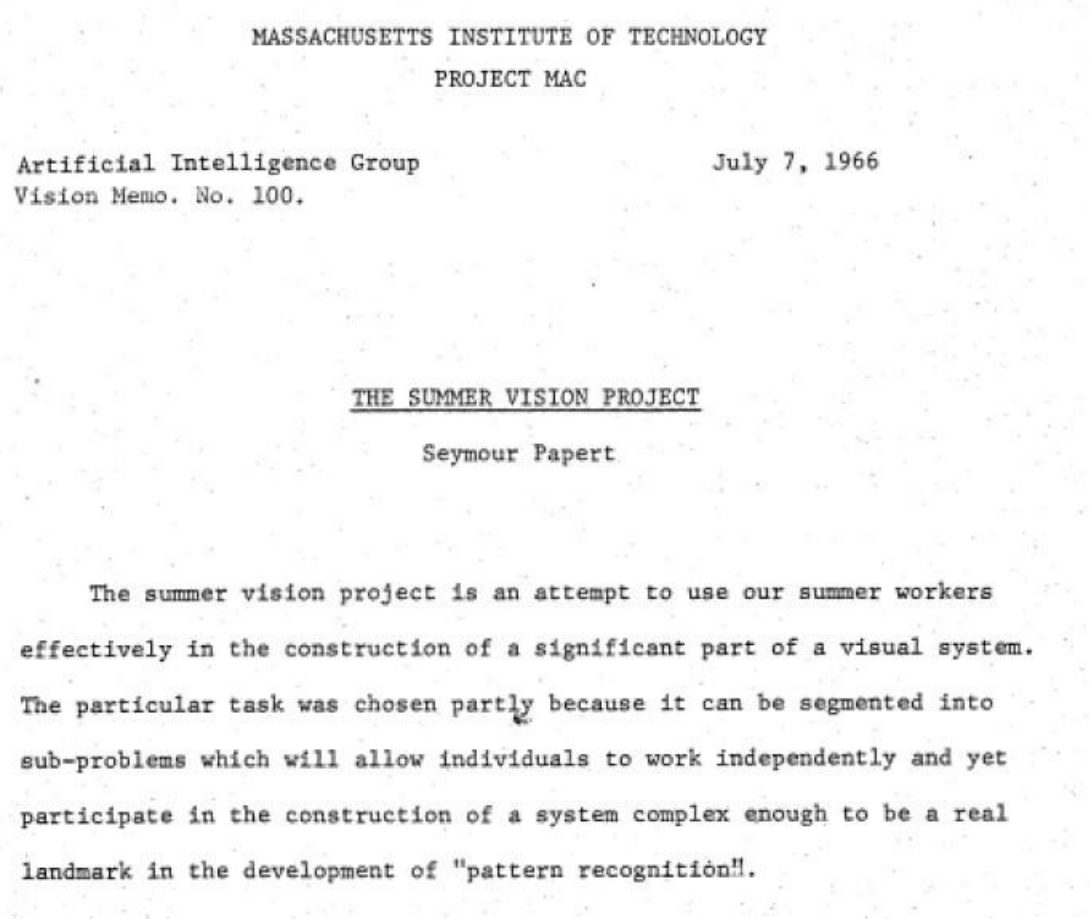
THE SUMMER VISION PROJECT

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".

- Năm 1966, Seymour Papert và Marvin Minsky, hai nhà tiên phong về trí tuệ nhân tạo, đã khởi động một dự án mang tên “Summer Vision Project“
- Thời gian thực hiện dự án: hai tháng.
- Số người tham gia: 10 người.

First Project, 1966



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

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

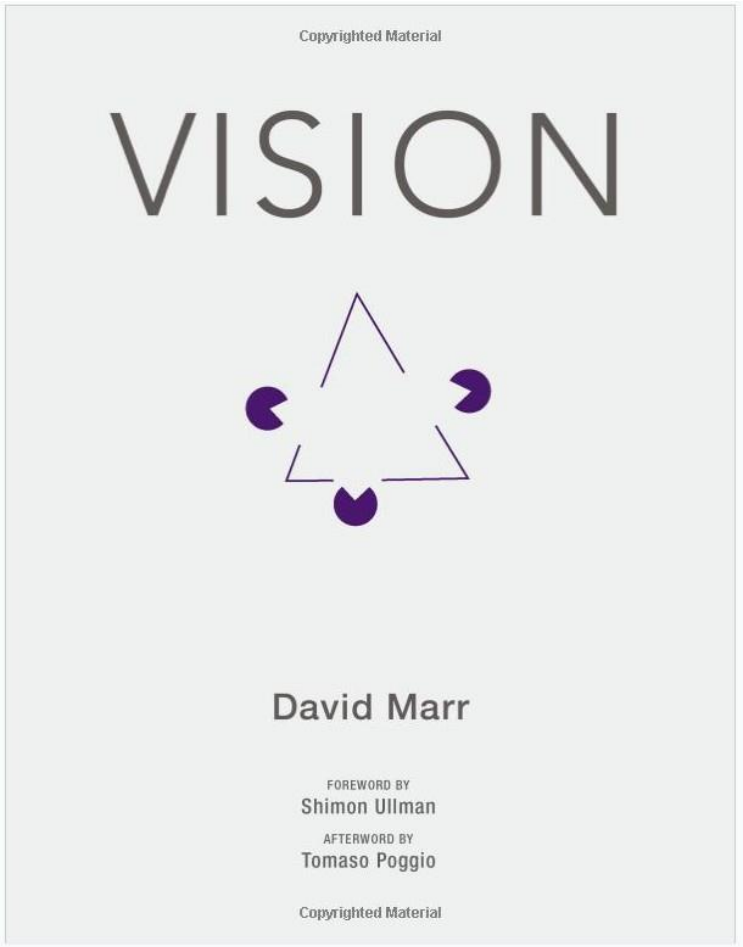
THE SUMMER VISION PROJECT

Seymour Papert

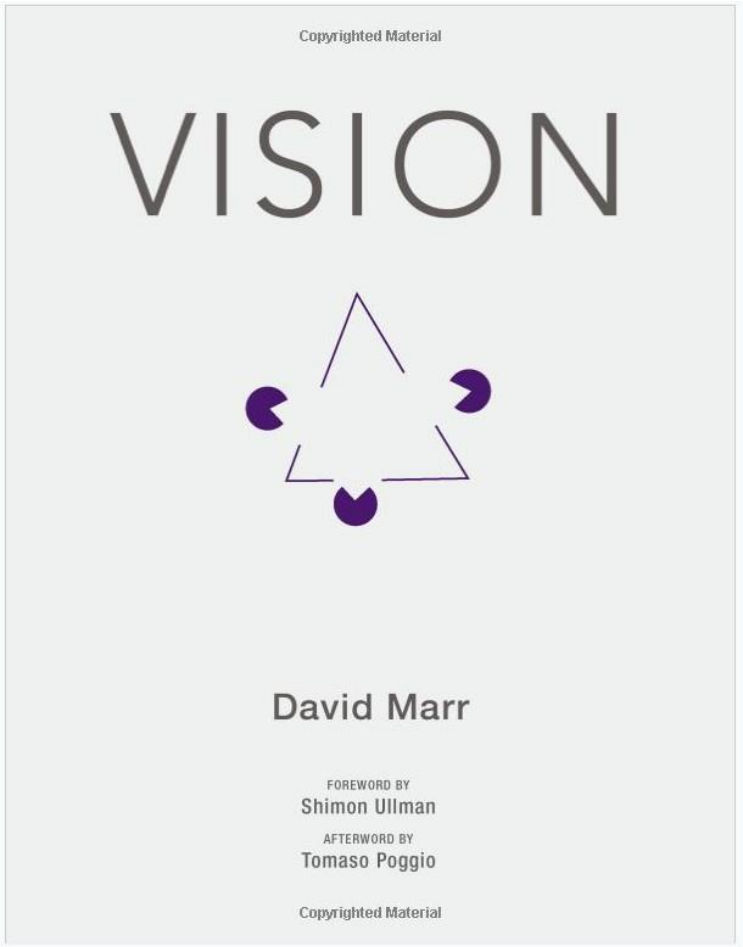
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the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

- Mục tiêu dự án: Xây dựng một hệ thống máy tính có thể nhận dạng các vật thể trong ảnh.
- Để hoàn thành nhiệm vụ, dự án cần viết một chương trình máy tính có khả năng xác định pixel nào thuộc về đối tượng nào.

First Book, 1970



First Book, 1970

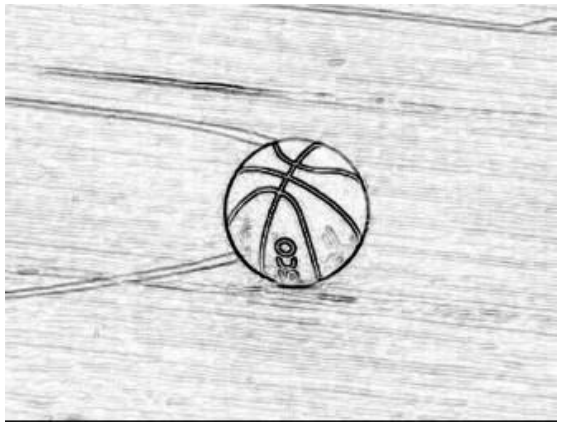


- David Marr là một nhà khoa học nổi tiếng trong lĩnh vực thị giác máy tính và thần kinh học. Vào những năm 1970, ông đã phát triển các lý thuyết quan trọng về cách thức thị giác hoạt động, đặc biệt là cách mà não bộ xử lý thông tin thị giác để tạo ra sự hiểu biết về thế giới xung quanh.

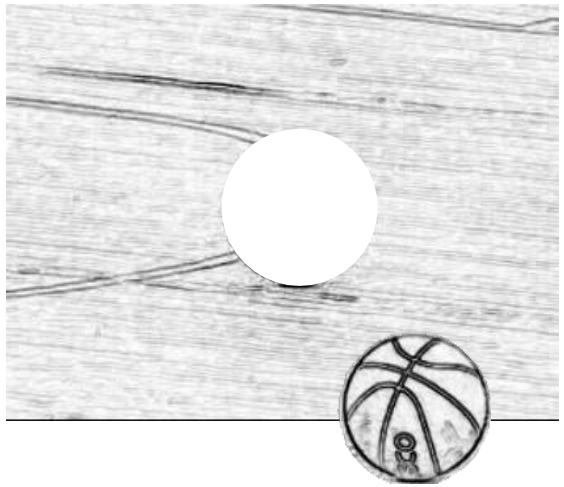
Stages of Visual Representation, 1970

Input image

Edge image



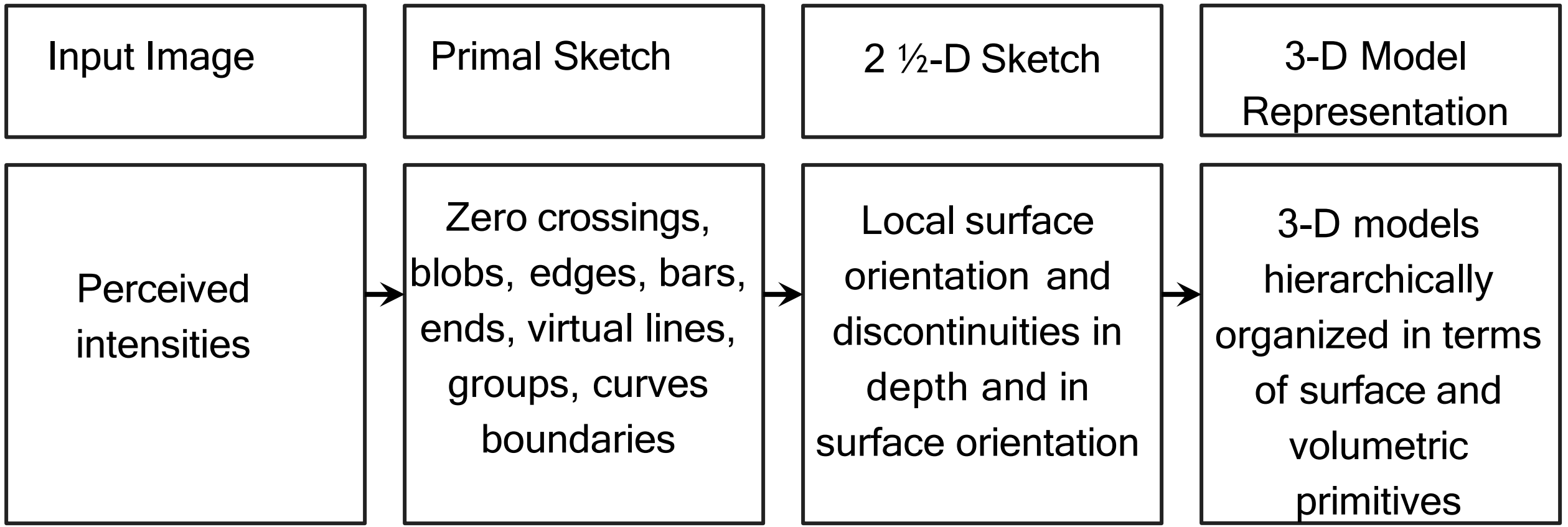
2 ½-D sketch



3-D model

Stages of Visual Representation, David Marr, 1970s

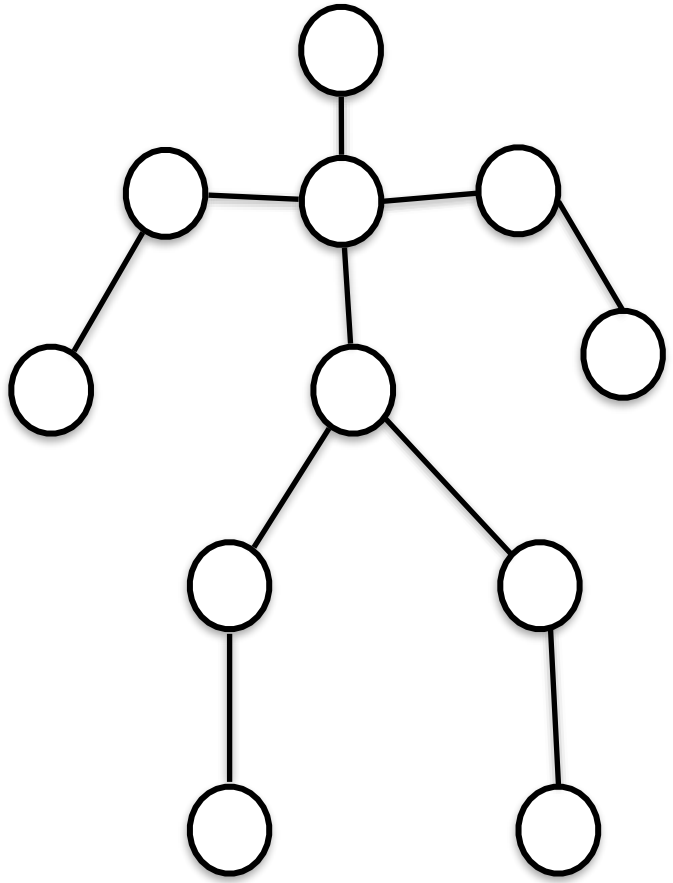
Stages of Visual Representation, 1970



Stages of Visual Representation, David Marr, 1970s

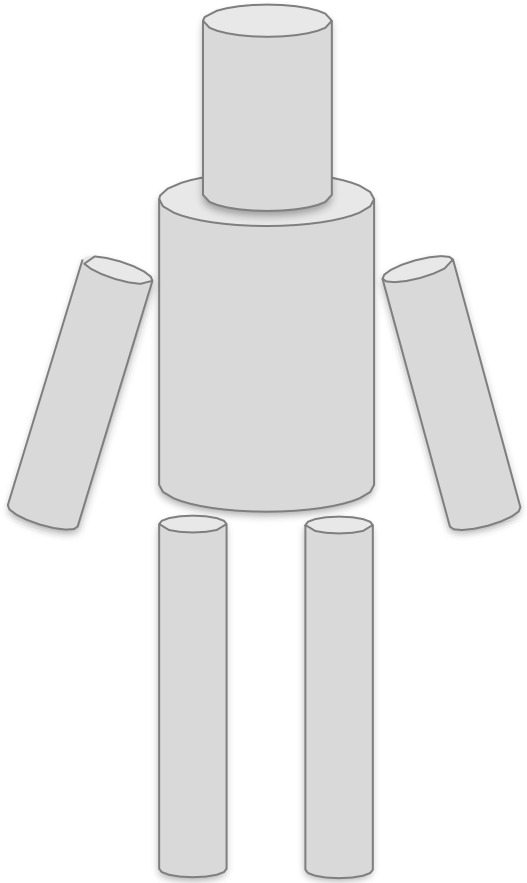
Pictorial Structure, 1973

- Fischler and Elschlager, 1973
  + Pictorial Structure
  + Structural Matching
- Công trình là một trong những nghiên cứu đầu tiên áp dụng cách tiếp cận cấu trúc để giải quyết vấn đề nhận dạng hình ảnh, một bước tiến lớn so với các phương pháp cục bộ trước đó.



Generalized Cylinder, 1979

- Brooks & Binford, 1979
- Generalized Cylinder
- Khái niệm generalized cylinder cung cấp một phương pháp mạnh mẽ và linh hoạt để biểu diễn các đối tượng ba chiều, mở ra nhiều hướng nghiên cứu và ứng dụng mới trong lĩnh vực này.

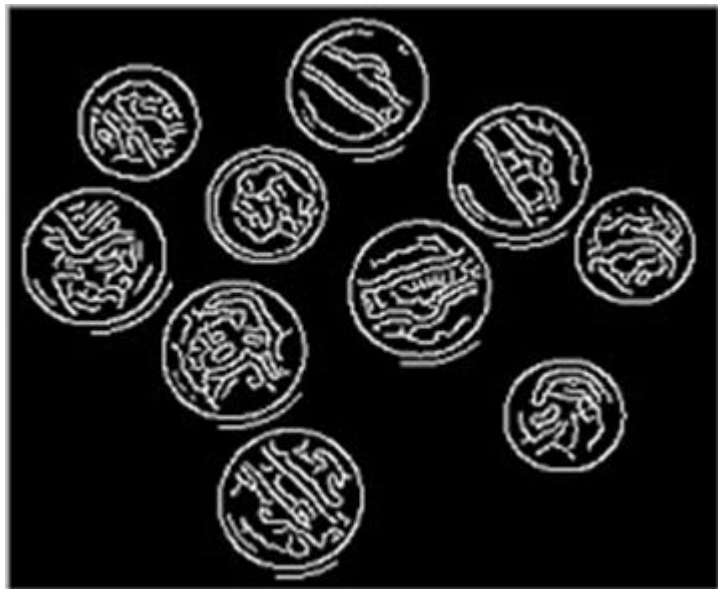


Recognition via Edge Detection (1980s)

David Lowe, 1987

- David Lowe phát triển các kỹ thuật để nhận dạng và phân đoạn đối tượng ba chiều từ dữ liệu hình ảnh.
- Lowe đã giới thiệu khái niệm về các đặc trưng bất biến, tức là các đặc trưng không thay đổi khi đối tượng trải qua các biến đổi như quay, thay đổi tỷ lệ và chiếu sáng.

David Lowe, 1987

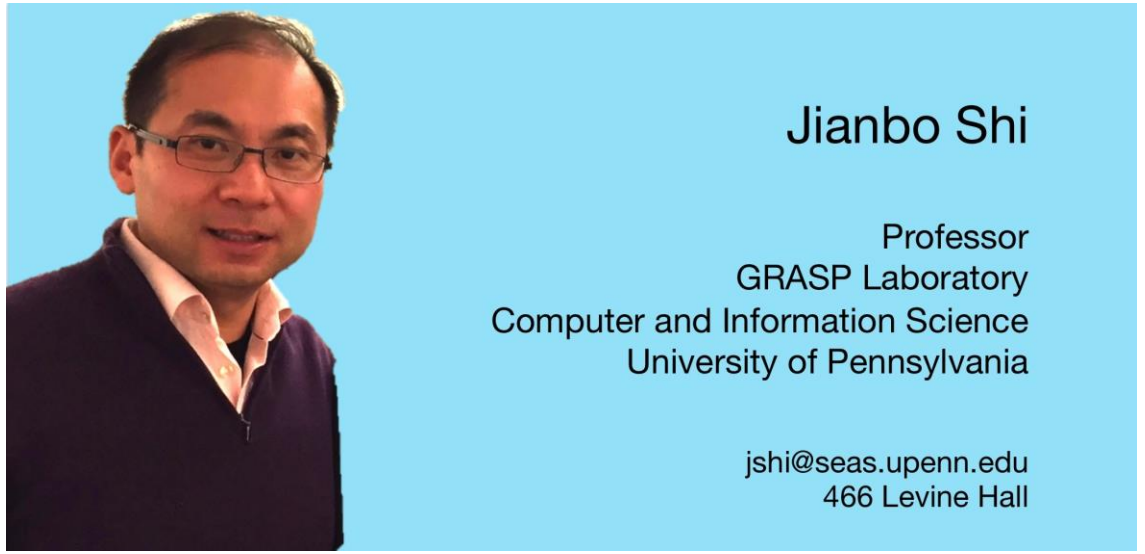


Normalized Cut (Shi & Malik, 1997)

- Công trình của Jianbo Shi và Jitendra Malik vào năm 1997 về phương pháp cắt chuẩn hóa để phân đoạn hình ảnh là một bước tiến lớn trong lĩnh vực thị giác máy tính.

Normalized Cut (Shi & Malik, 1997)

- Công trình của Jianbo Shi và Jitendra Malik vào năm 1997 về phương pháp cắt chuẩn hóa để phân đoạn hình ảnh là một bước tiến lớn trong lĩnh vực thị giác máy tính.



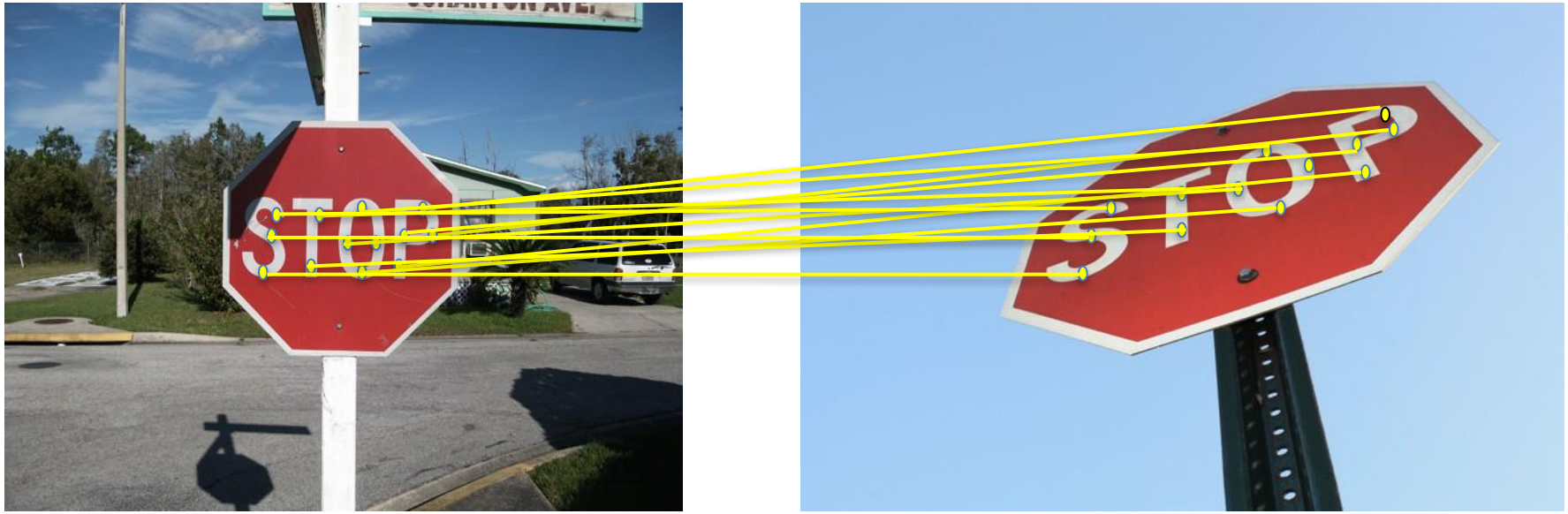
Normalized Cut (Shi & Malik, 1997)

Normalized Cut (Shi & Malik, 1997)

Normalized Cut (Shi & Malik, 1997)

SIFT & Object Recognition, 1999

– SIFT & Object Recognition, David Lowe, 1999

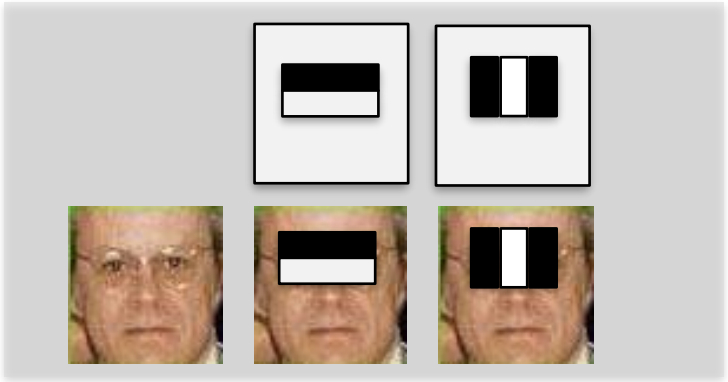


SIFT, David Lowe, 1999

Face Detection, Viola & Jones, 2001

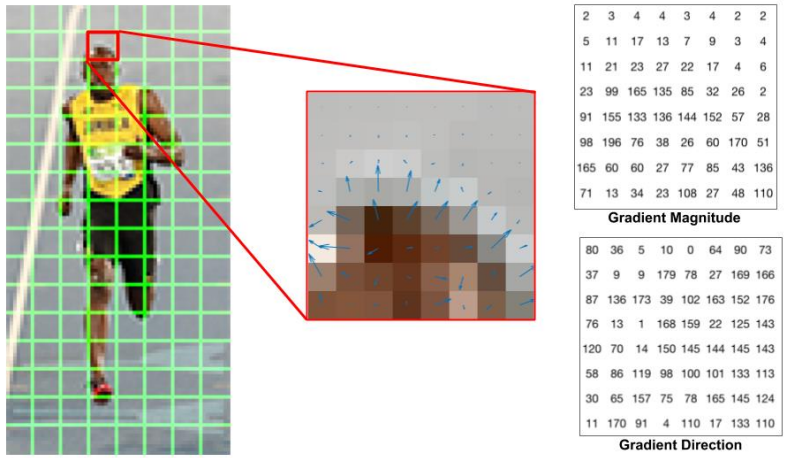
- Viola-Jones là một thuật toán quan trọng trong việc nhận diện khuôn mặt, được phát triển bởi Paul Viola và Michael Jones vào năm 2001.

Face Detection, Viola & Jones, 2001



Histogram of Gradients (HoG), 2005

- Histogram of Gradients (HoG) Dalal & Triggs, 2005
- Histogram of Gradients (HOG) là một kỹ thuật quan trọng trong xử lý ảnh và thị giác máy tính, được sử dụng để mô tả và nhận dạng các đối tượng trong hình ảnh. Đây là một phương pháp mạnh mẽ để rút trích đặc trưng từ hình ảnh để nhận diện vật thể.



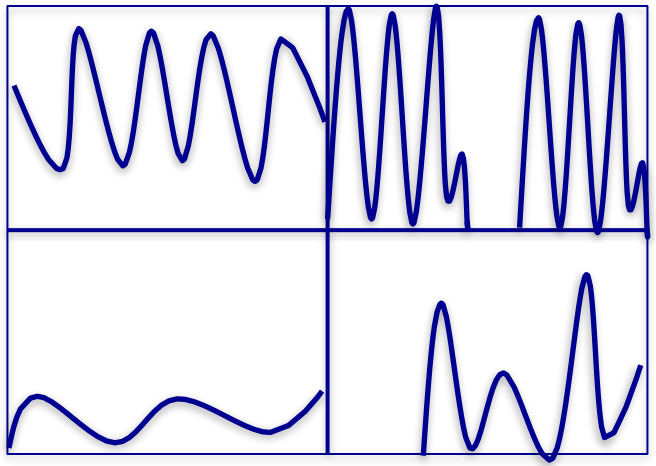
Spatial Pyramid Matching, 2006

– Spatial Pyramid Matching là một phương pháp cải tiến của Histogram of Oriented Gradients (HOG) và các kỹ thuật rút trích đặc trưng dựa trên vùng (region-based features).

Level 0

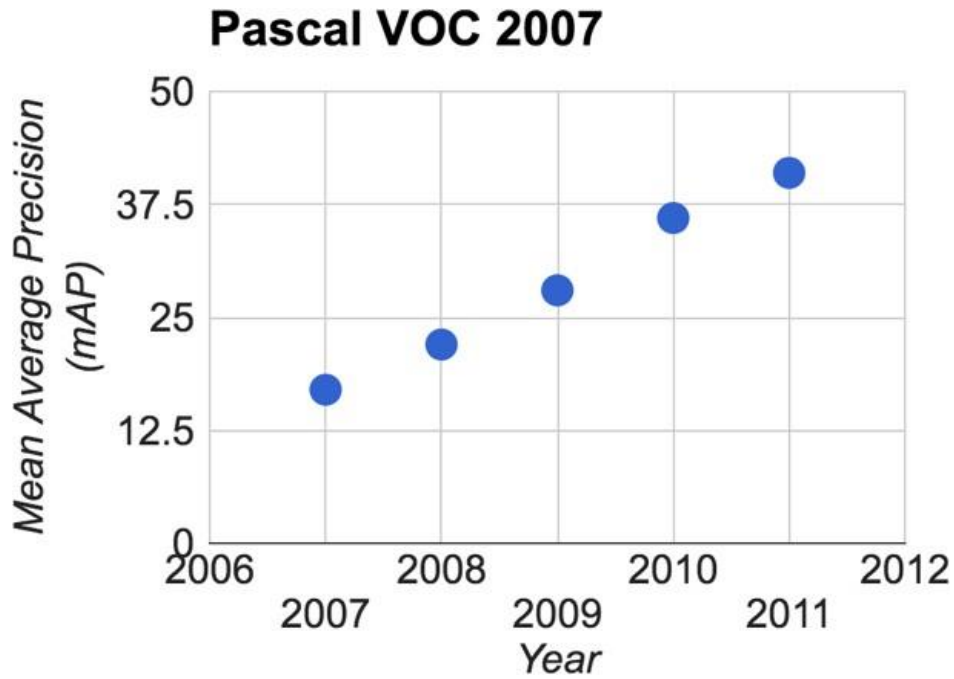
Spatial Pyramid Matching, 2006

- Spatial Pyramid Matching là một phương pháp cải tiến của Histogram of Oriented Gradients (HOG) và các kỹ thuật rút trích đặc trưng dựa trên vùng (region-based features).



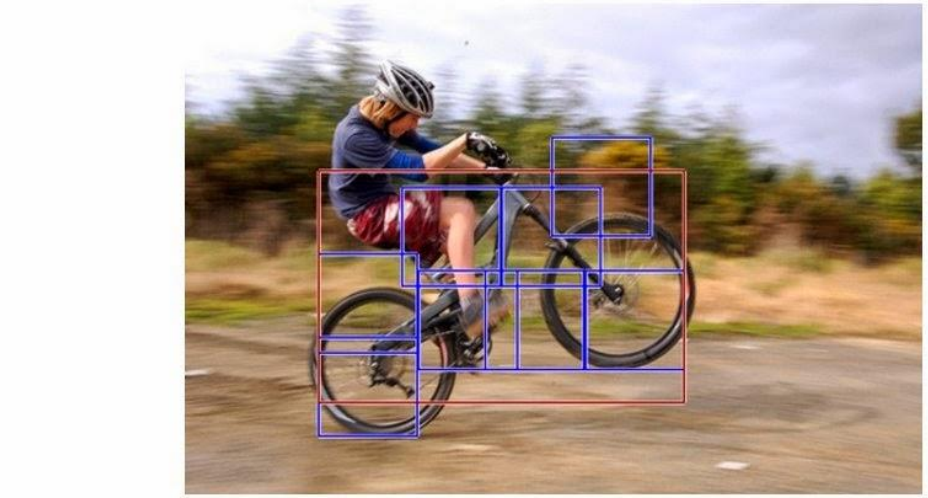
PASCAL VOC, 2006 – 2012

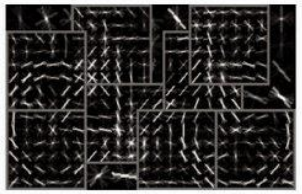
- PASCAL Visual Object Challenge (20 object categories)
- Everingham et al. 2006-2012

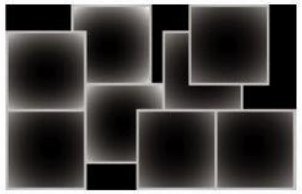


Deformable Part Model, 2009

- Deformable Part Model (DPM) Felzenswalb, McAllester, Ramanan, 2009.
- Mô hình DPM là sự phát triển từ các mô hình trước đó, nhằm cải thiện khả năng nhận dạng đối tượng trong các hình ảnh có sự biến đổi về hình dạng và vị trí của đối tượng.







ImageNet, 2009

– 22K categories and 14M images

IMAGENET

ImageNet, 2009

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

The Image Classification 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

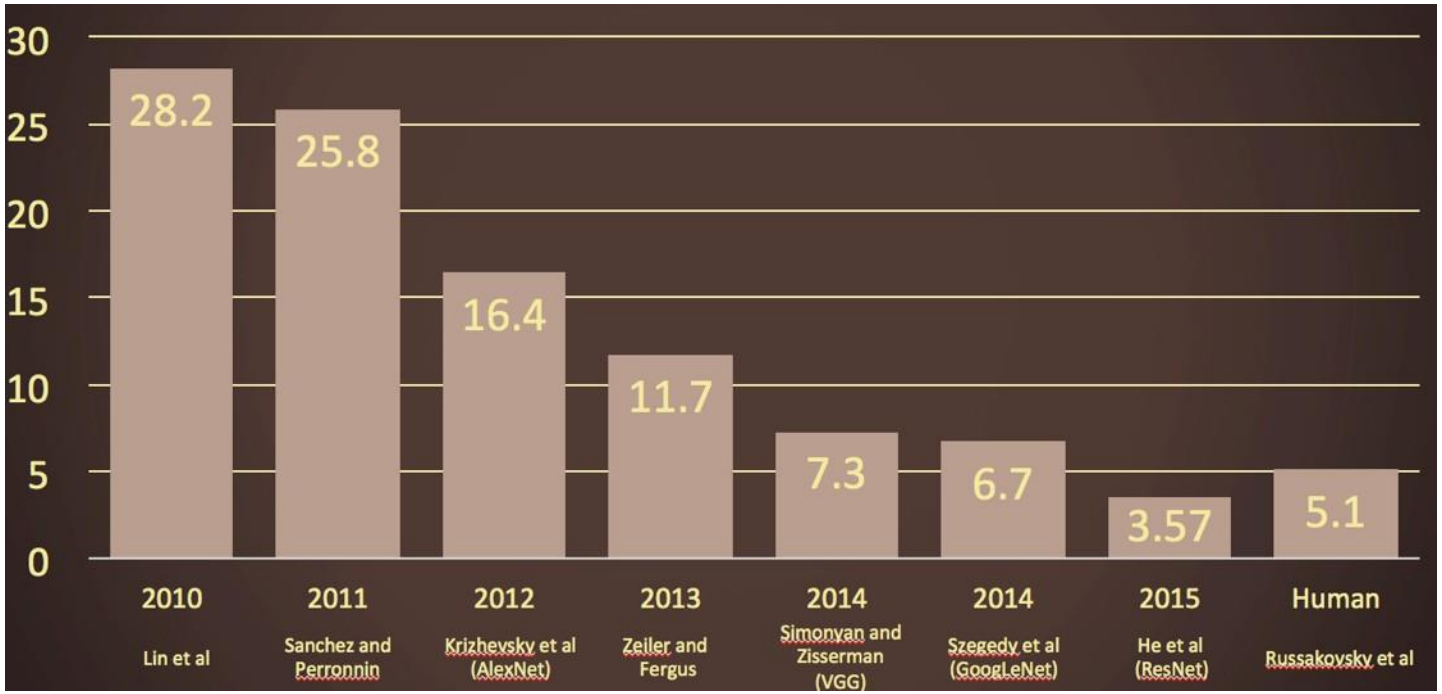
✗

IMAGENET **Large Scale Visual Recognition Challenge**

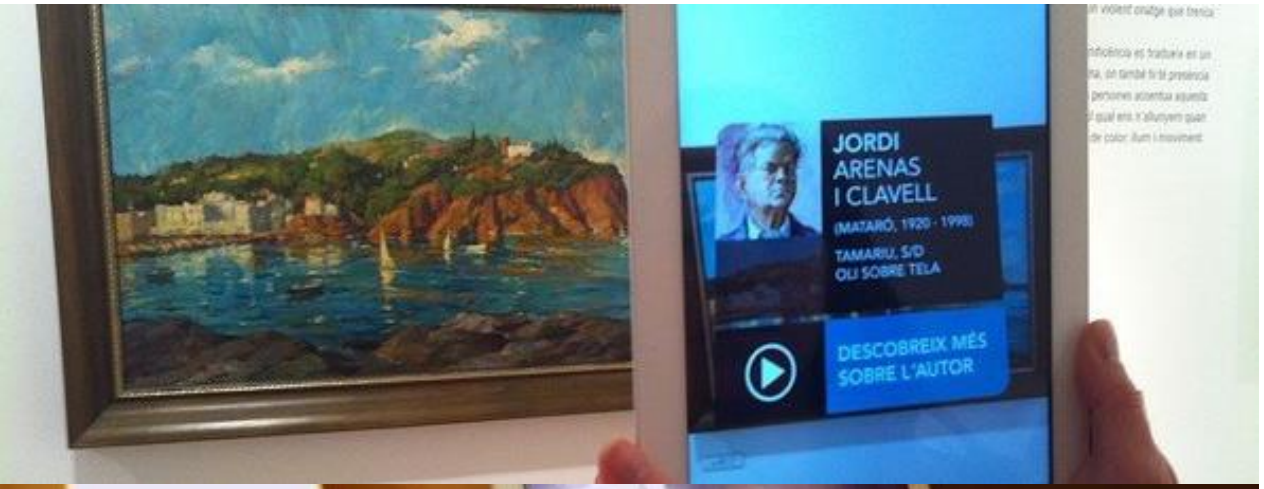
The Image Classification Challenge

– The Image Classification Challenge:

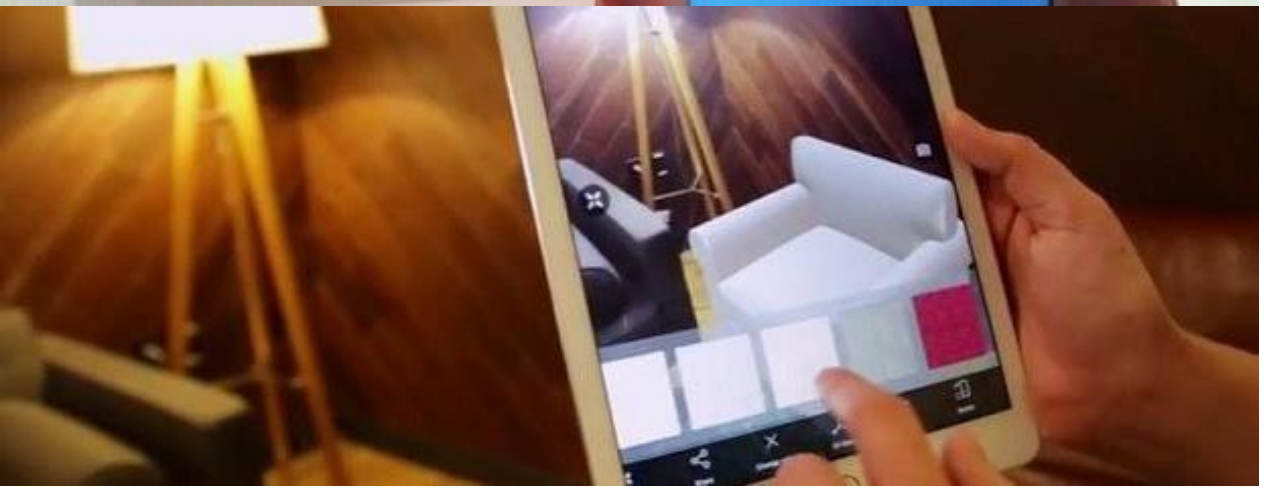
+ 1,000 object classes – 1,431,167 images



CS231n focuses on one of the most important problems of visual recognition – image classification


JORDI
ARENAS
I CLAVELL
(MATARÓ, 1920 - 1998)
TAMARIU, S/D
OLI SOBRE TELA
DESCOBREIX MÉS
SOBRE L'AUTOR


DC
PWR
Capture Left Iris
UNCLASSIFIE
Press trigger to acquire




0.892 Burger
0.529 Cup
0.322 Plate

There is a number of visual recognition problems that are related to image classification, such as object detection, image captioning

Object detection

Action classification

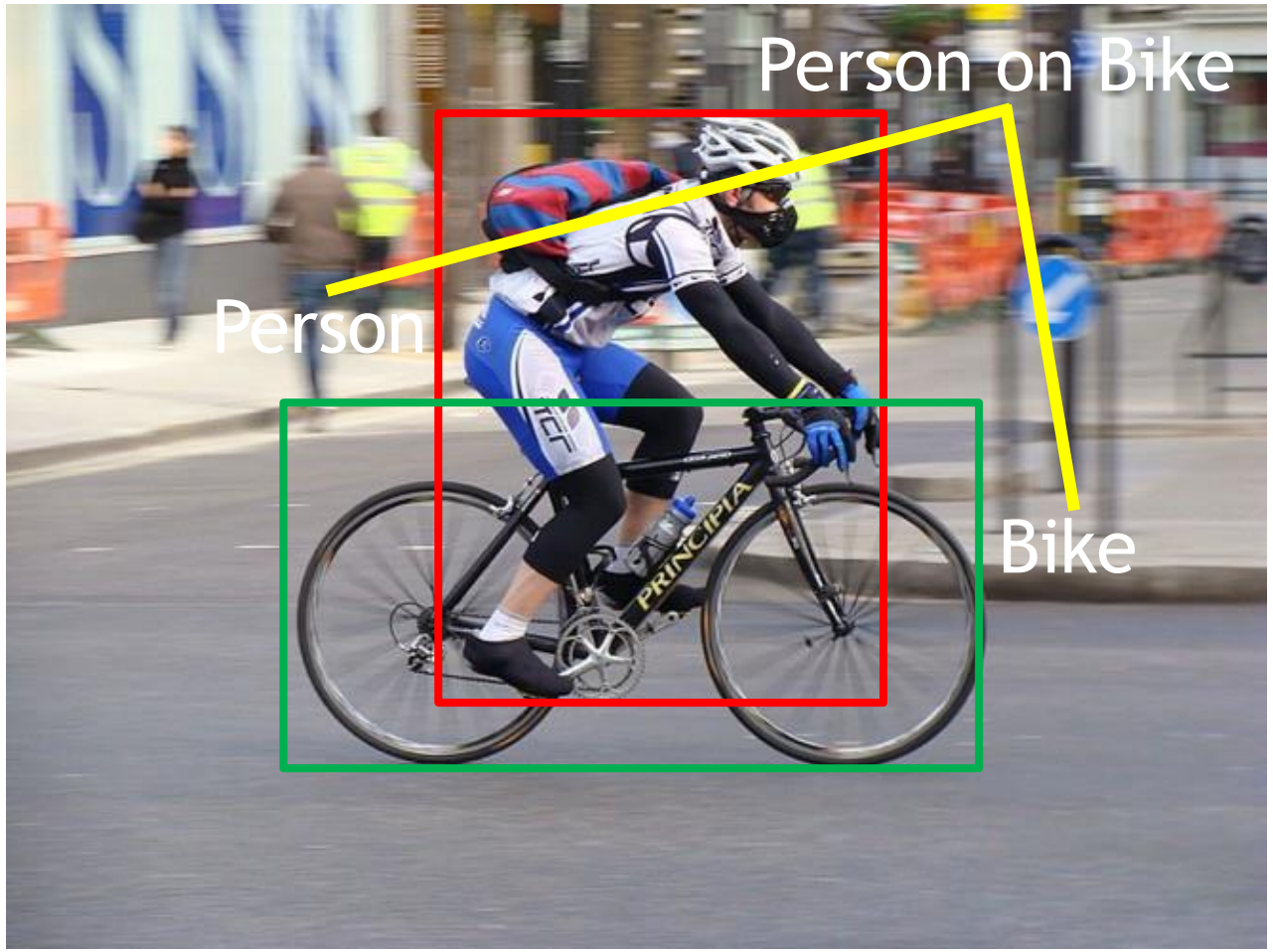


Image captioning

Person

Hammer

Convolutional Neural Networks (CNN) have become an important tool for object recognition

Year 2010

NEC-UIUC

Dense descriptor grid: HOG, LBP

Coding: local coordinate, super-vector

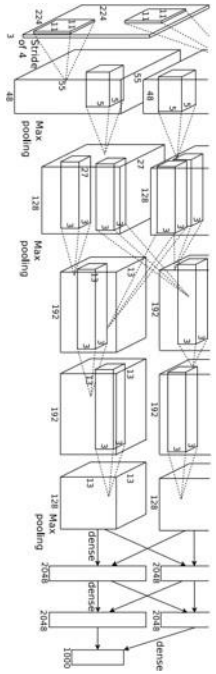
Pooling, SPM

Linear SVM

[Lin CVPR 2011]

Year 2012

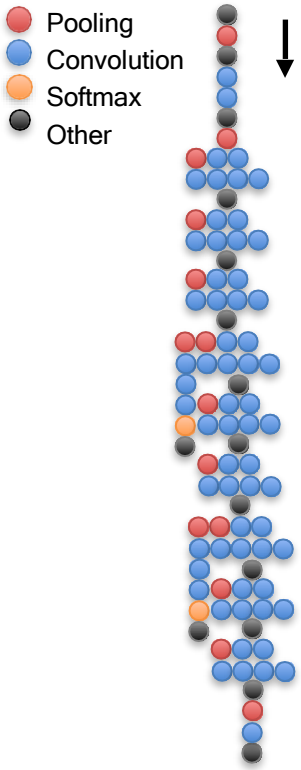
SuperVision



[Krizhevsky NIPS 2012]

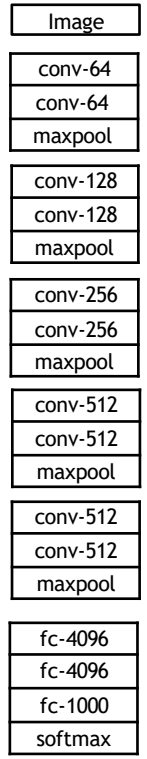
Year 2014

GoogLeNet



[Szegedy arxiv 2014]

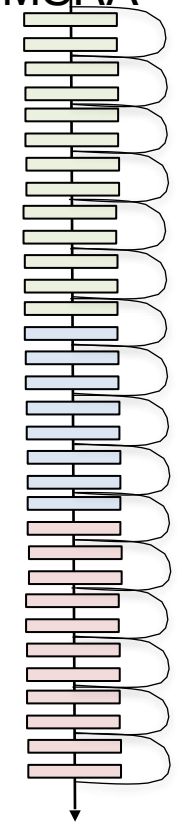
VGG



[Simonyan arxiv 2014]

Year 2015

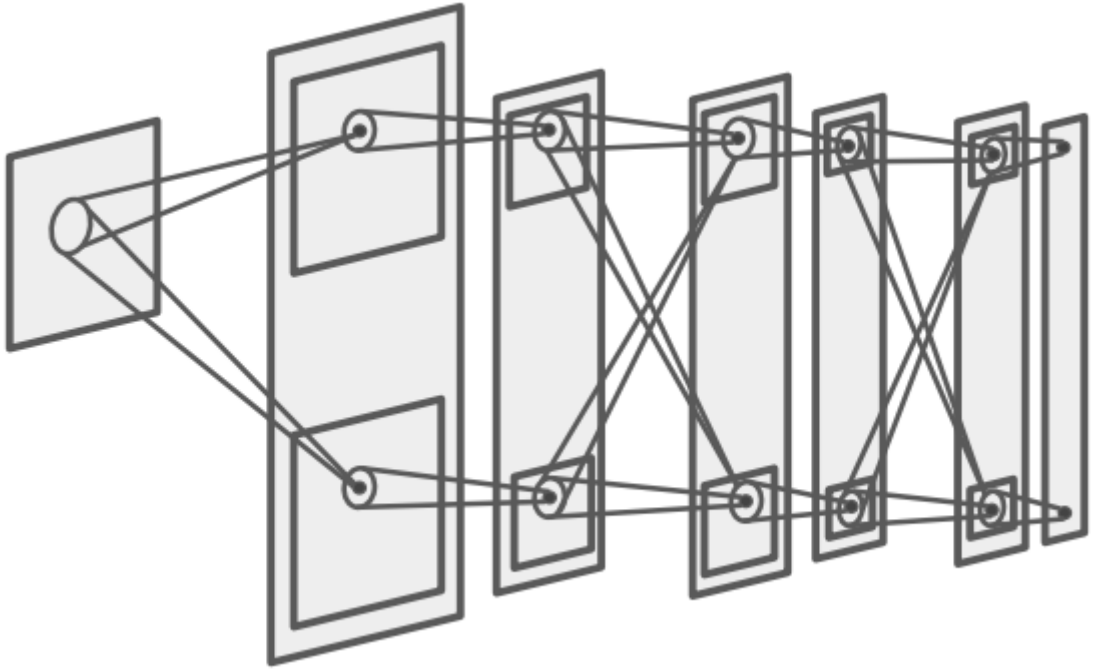
MSRA



[He ICCV 2015]

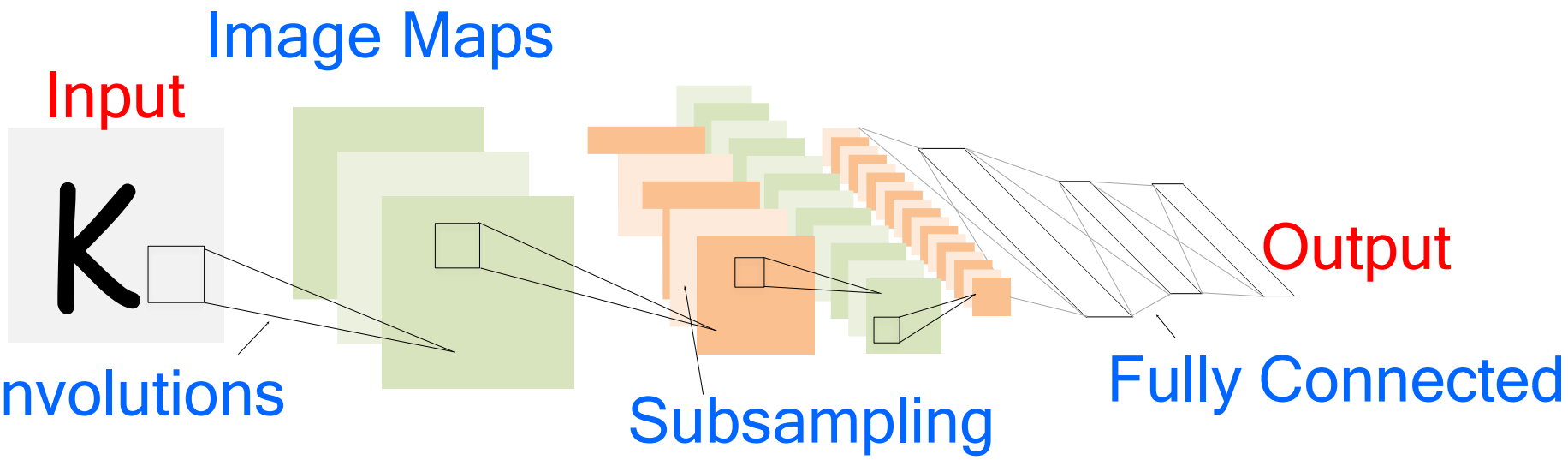
Convolutional Neural Networks (CNN) were not invented overnight

Neocognitron, 1980



LeCun, 1998

1998
LeCun et al.



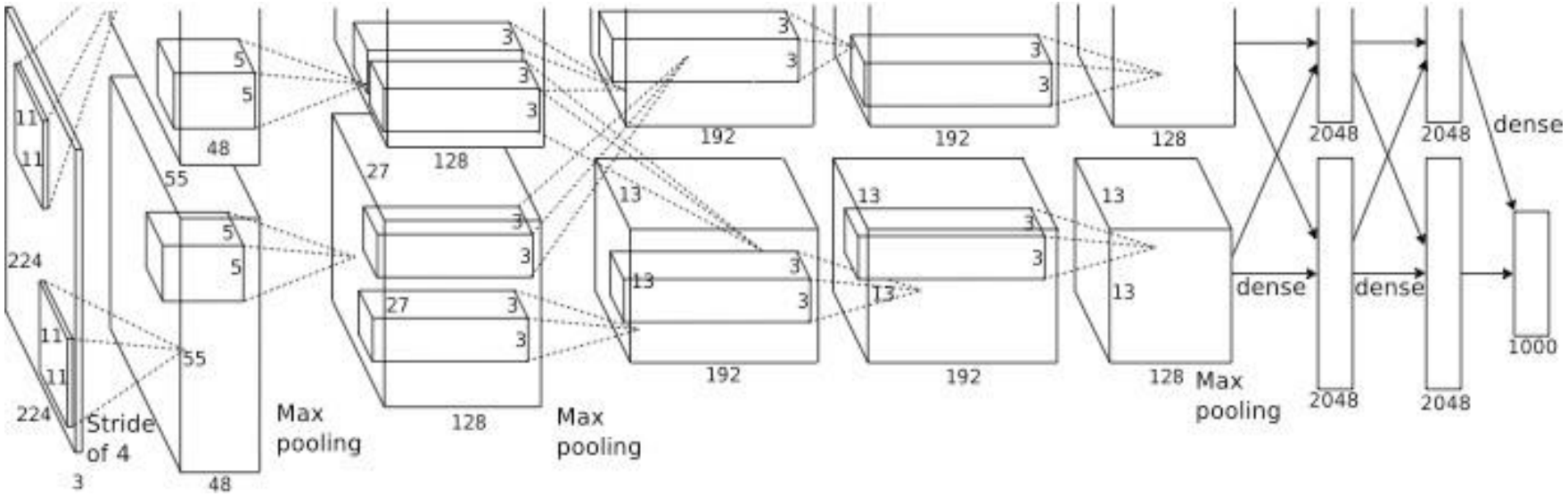
# of transistors

10^6

# of pixels used in training

10^7 NIST

Krizhevsky, 2012



# of transistors

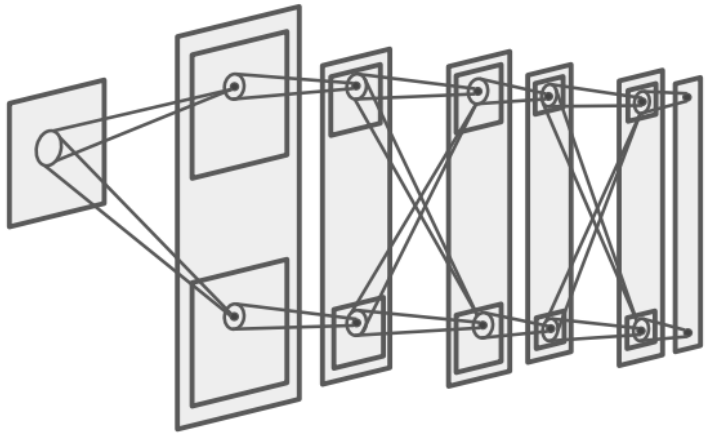
10^{14}

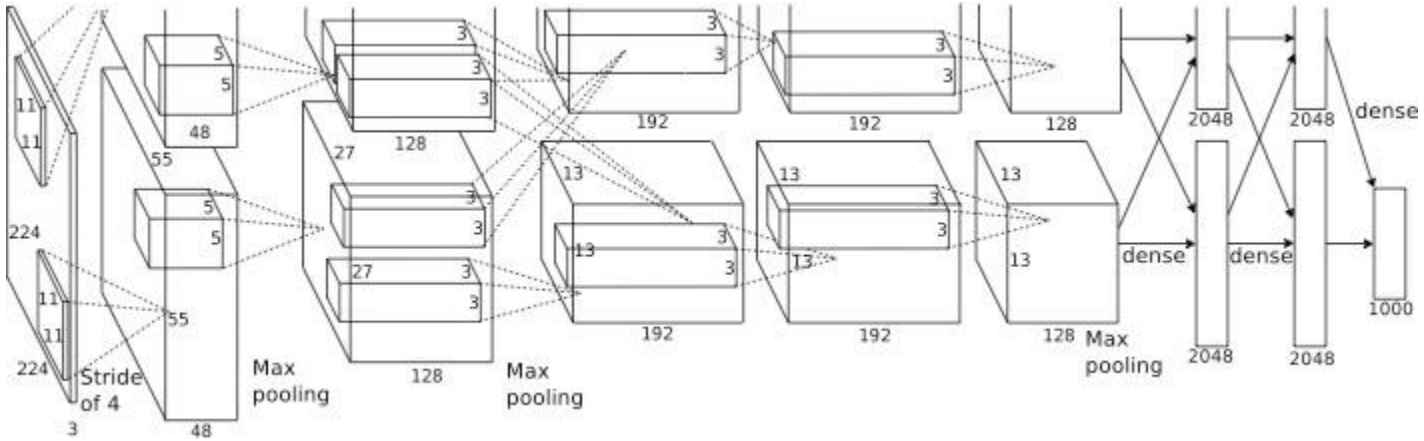
GPUs

# of pixels used in training

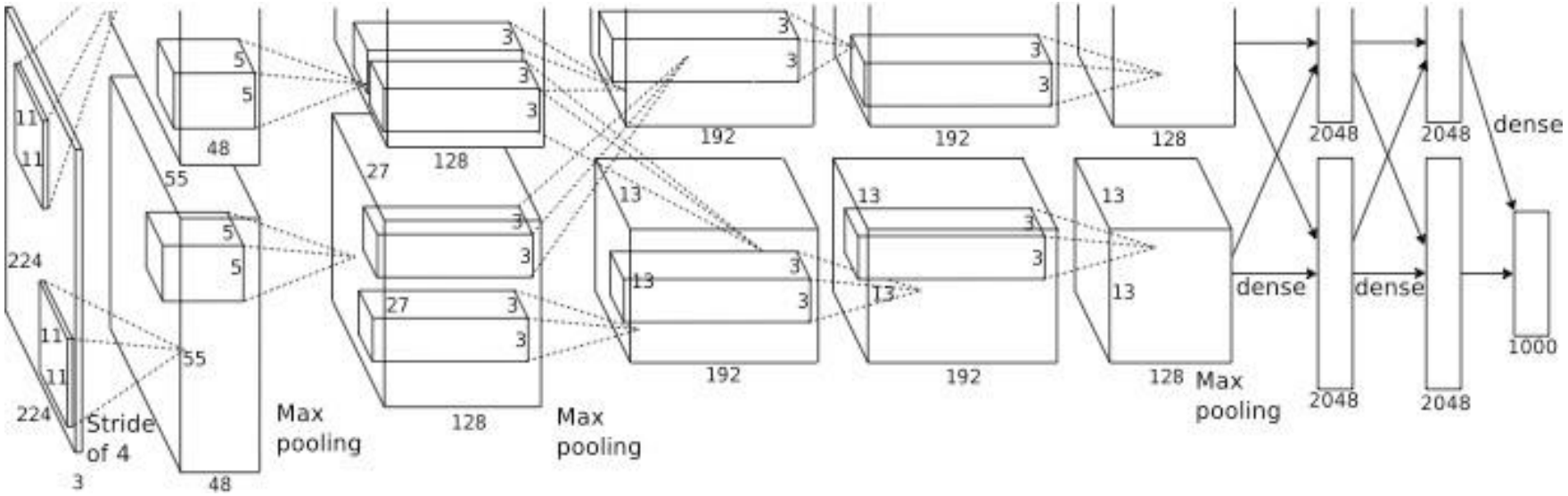
10^{14} IMAGENET

AlexNet – Neocognitron: 32 years apart





Deep Learning Goes Mainstream



# of transistors

10^{14}

GPUs

# of pixels used in training

10^{14} IMAGENET

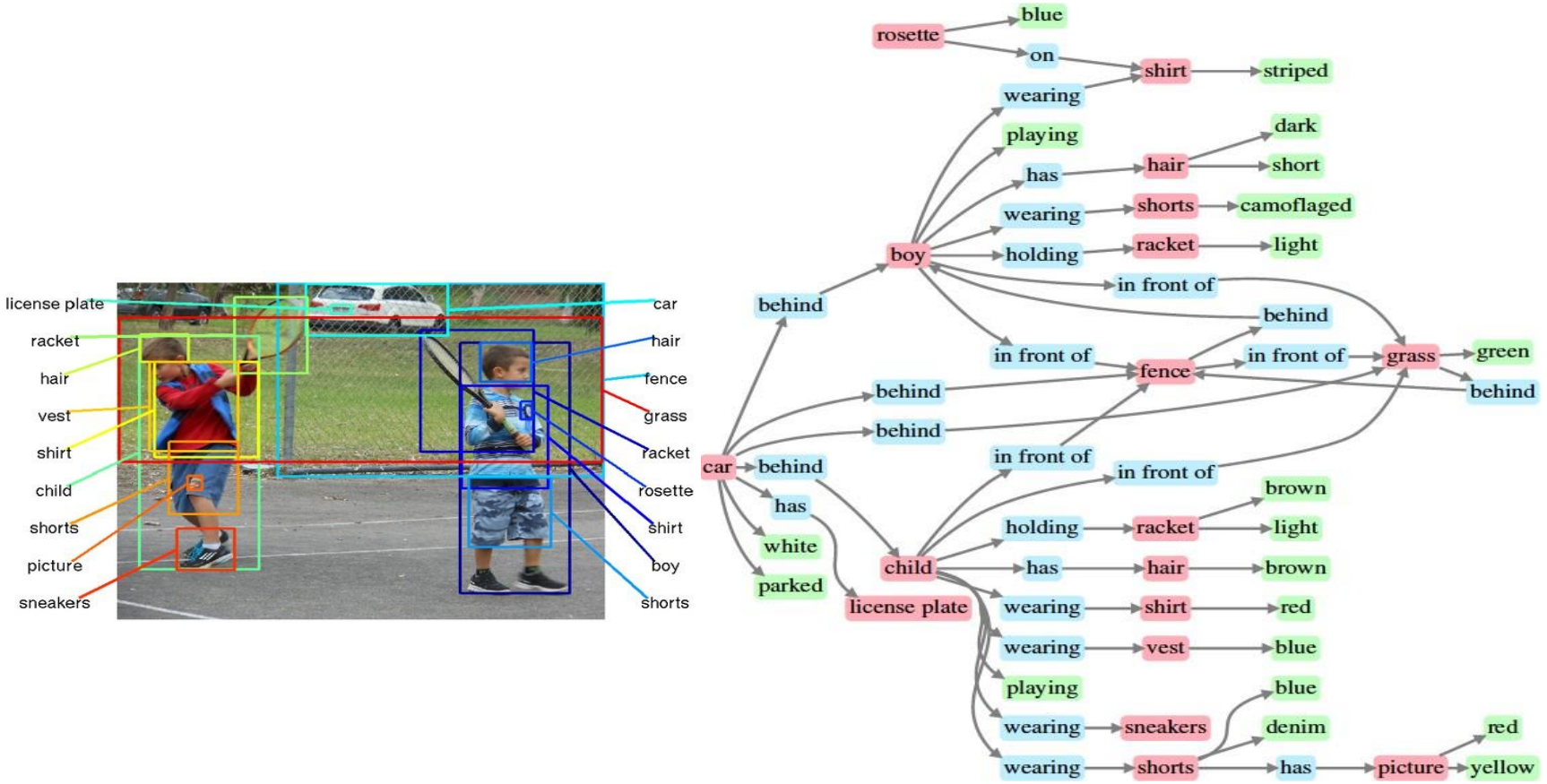
The quest for visual intelligence goes far beyond object recognition…

Deep Learning is Everywhere

Image Classification

Image Retrieval

Image Retrieval using Scene Graphs



Johnson et al., "Image Retrieval using Scene Graphs", CVPR 2015

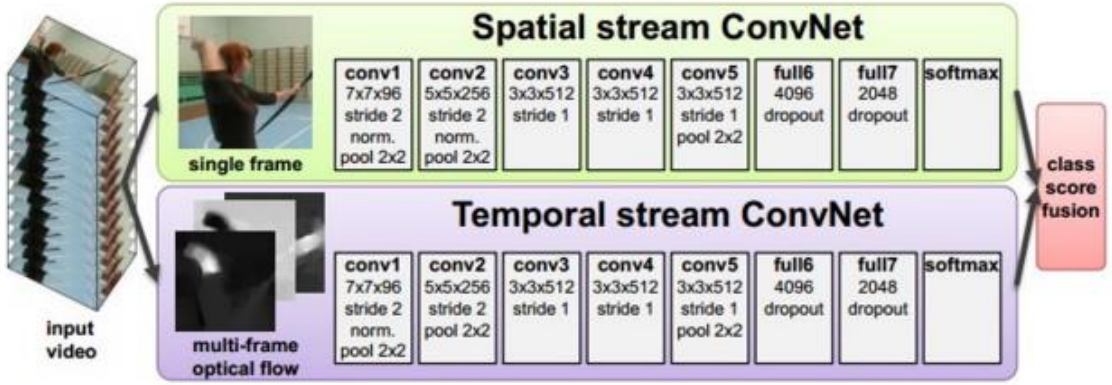
Deep Learning is Everywhere

Object Detection

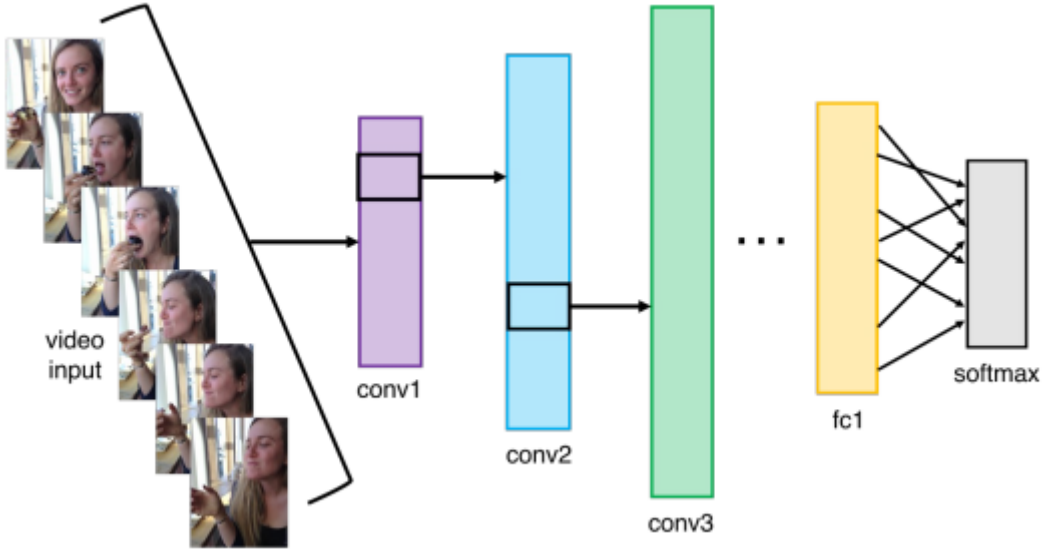
Image Segmentation

Deep Learning is Everywhere

Video Classification



Activity Recognition

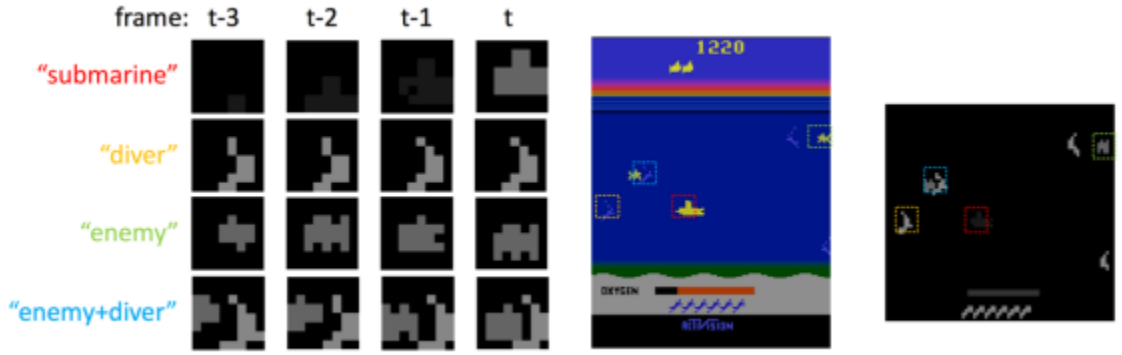


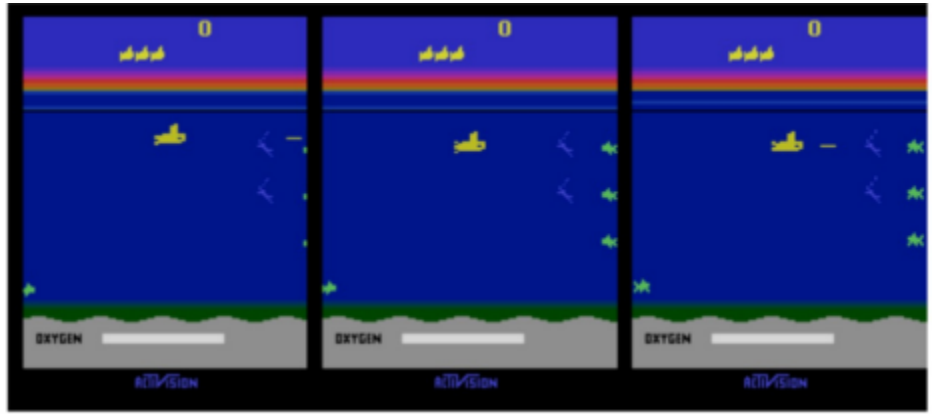
Deep Learning is Everywhere

– Pose Recognition (Toshev and Szegedy, 2014)

Deep Learning is Everywhere

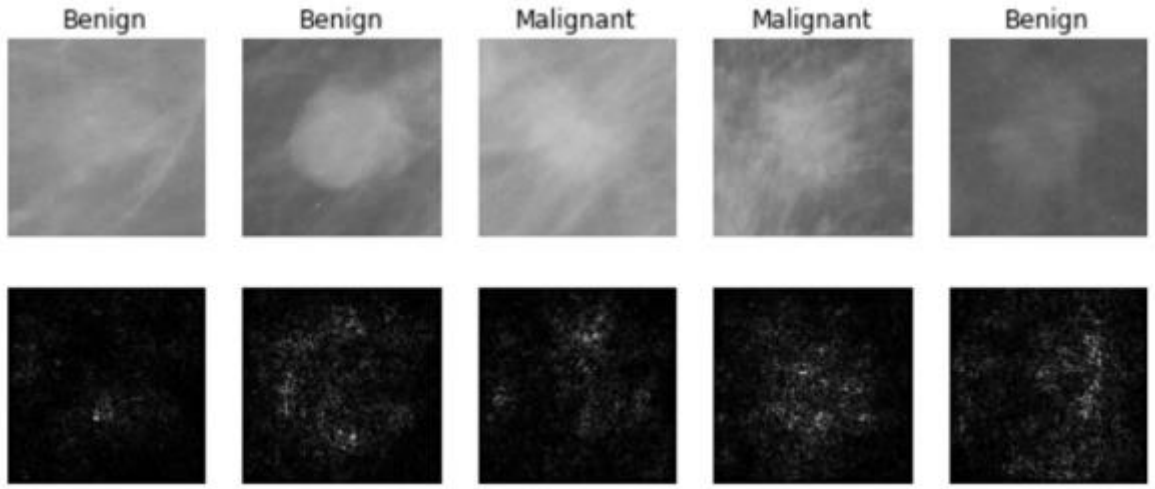
– Playing Atari games (Guo et al, 2014)





Deep Learning is Everywhere

- Medical Imaging
- Levy et al, 2016



Deep Learning is Everywhere

- Galaxy Classification
- Dieleman et al, 2014

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

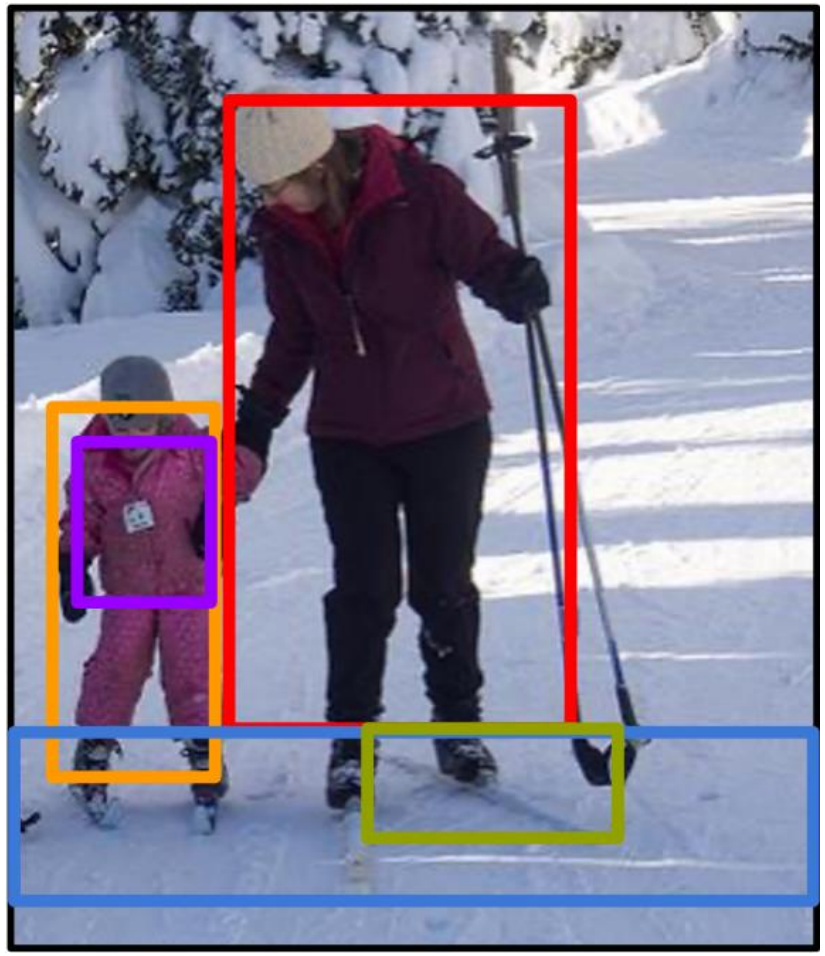
A man riding a wave on top of a surfboard

A cat sitting on a suitcase on the floor

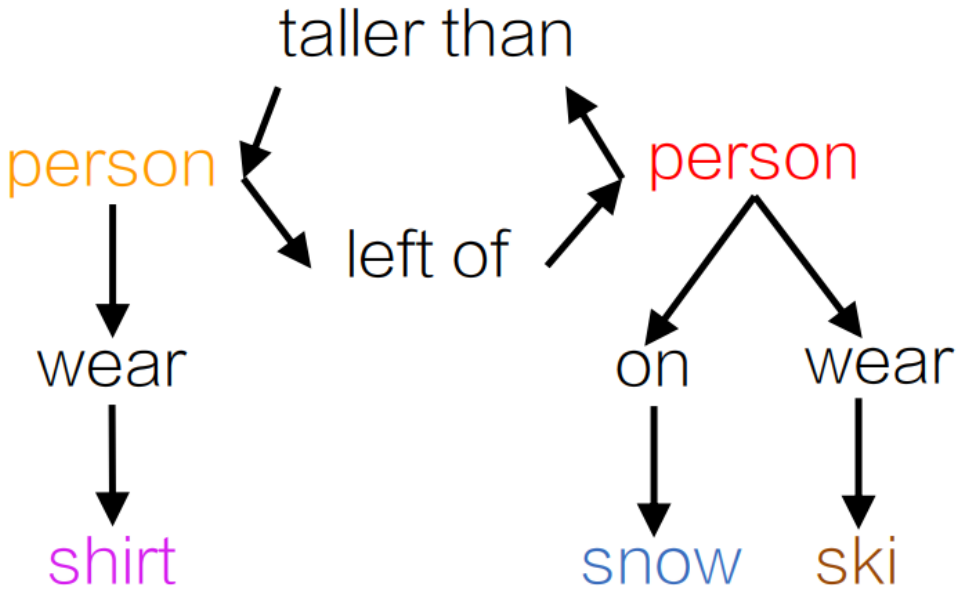
A woman standing on a beach holding a surfboard

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

Deep Learning is Everywhere

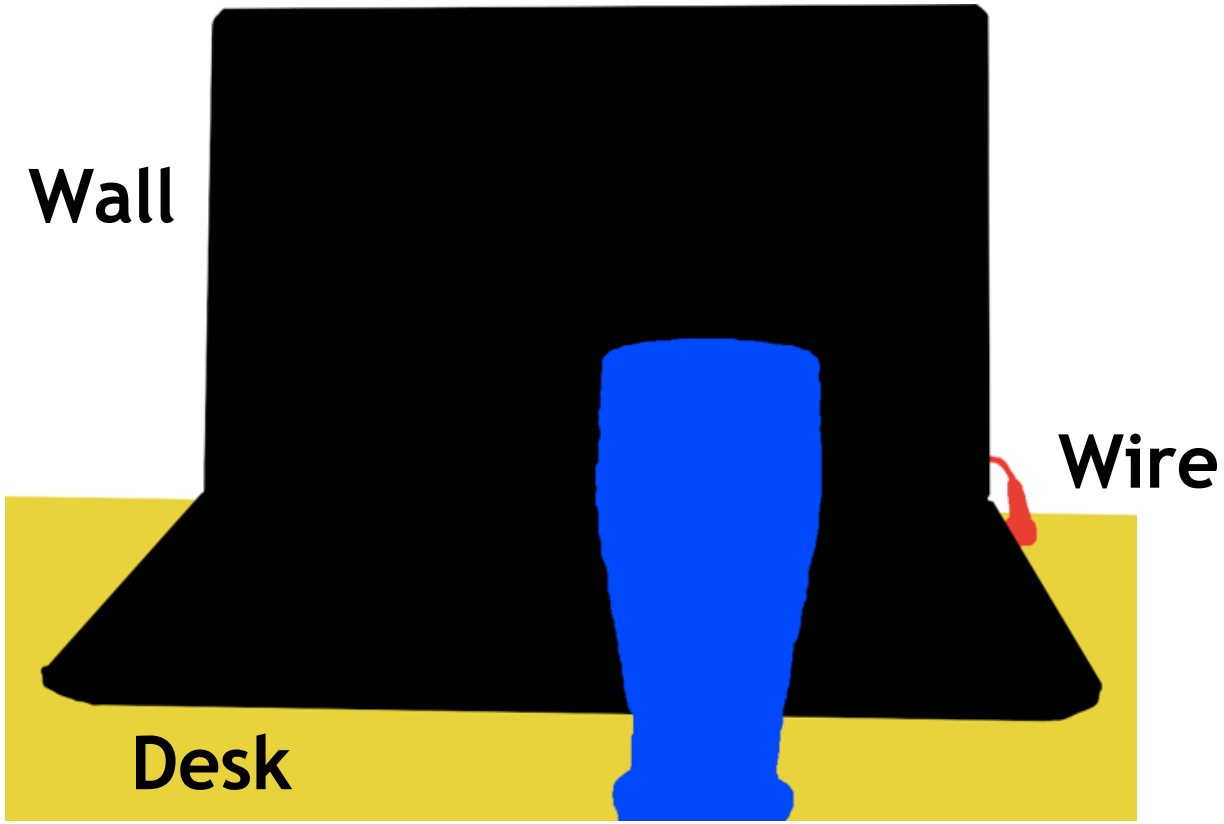


– Results: spatial, comparative, asymmetrical, verb, prepositional.



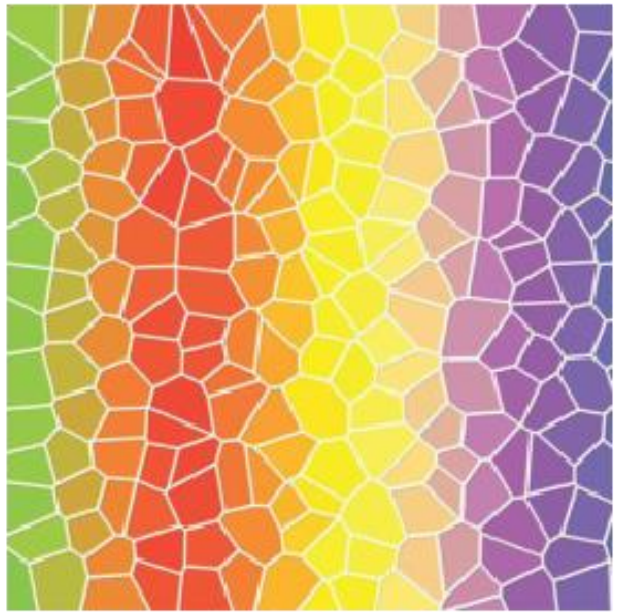
Deep Learning is Everywhere

Whale recognition



Deep Learning is Everywhere

Style Image



Content Image

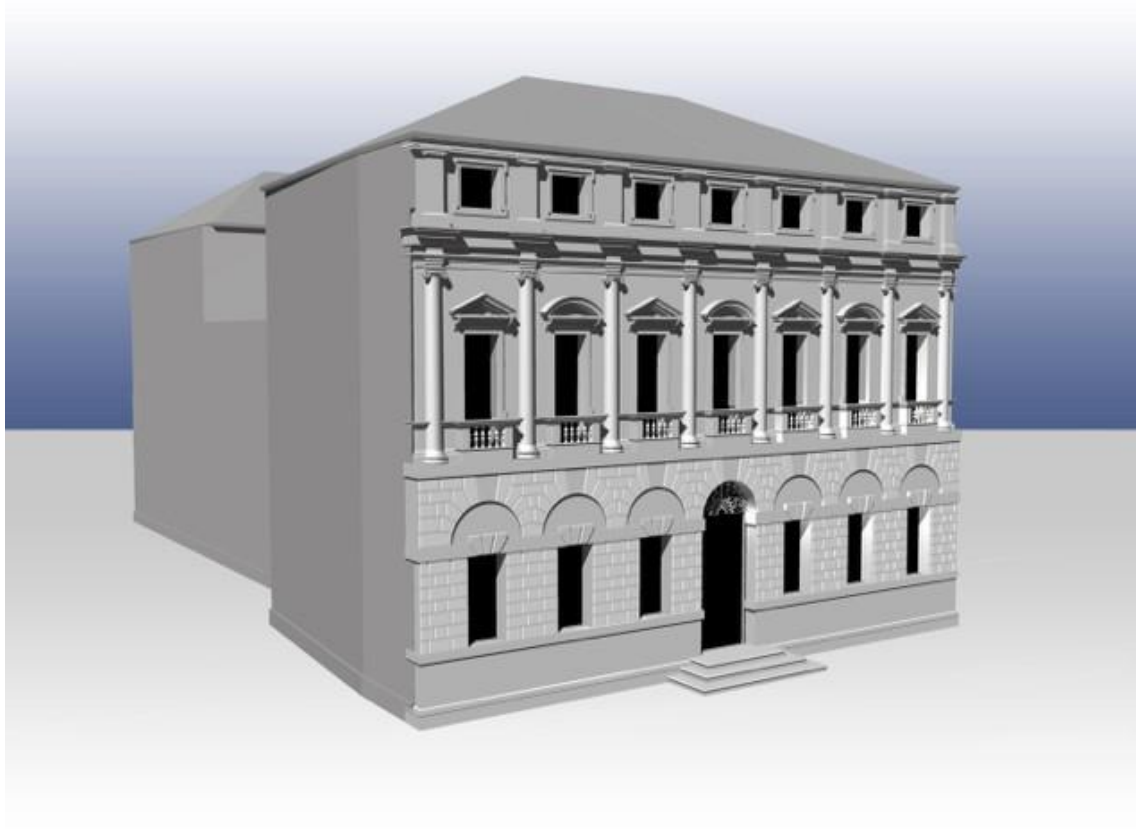
Output

Neural Style Transfers

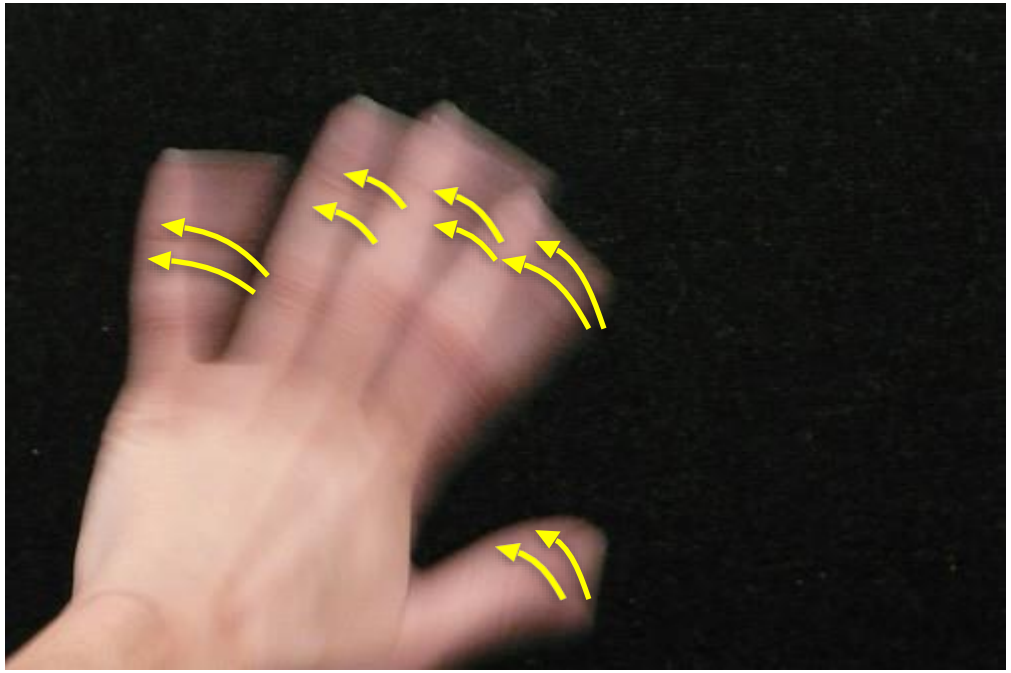
Deep Learning is Everywhere

Neural Style Transfers

Deep Learning is Everywhere



Deep Learning is Everywhere



Deep Learning is Everywhere

TEXT PROMPT

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

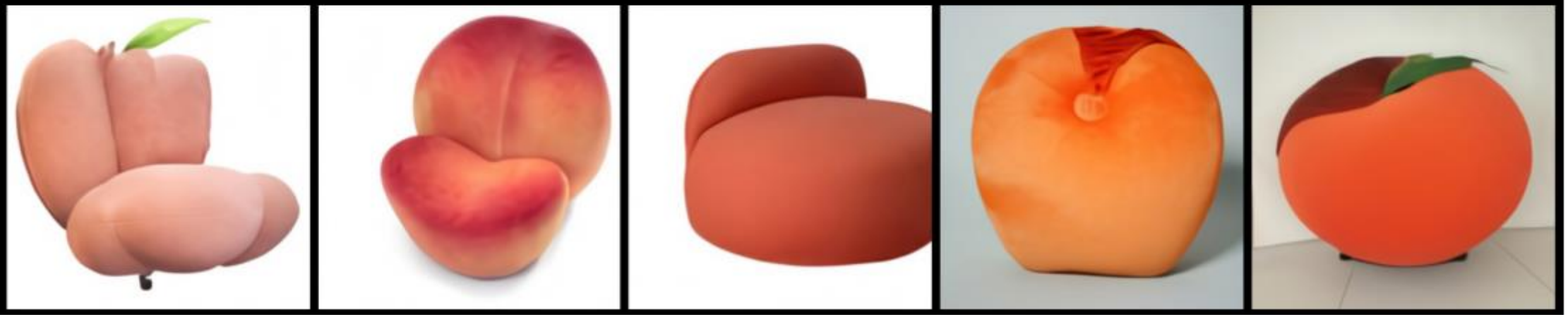
AI-GENERATED IMAGES

Deep Learning is Everywhere

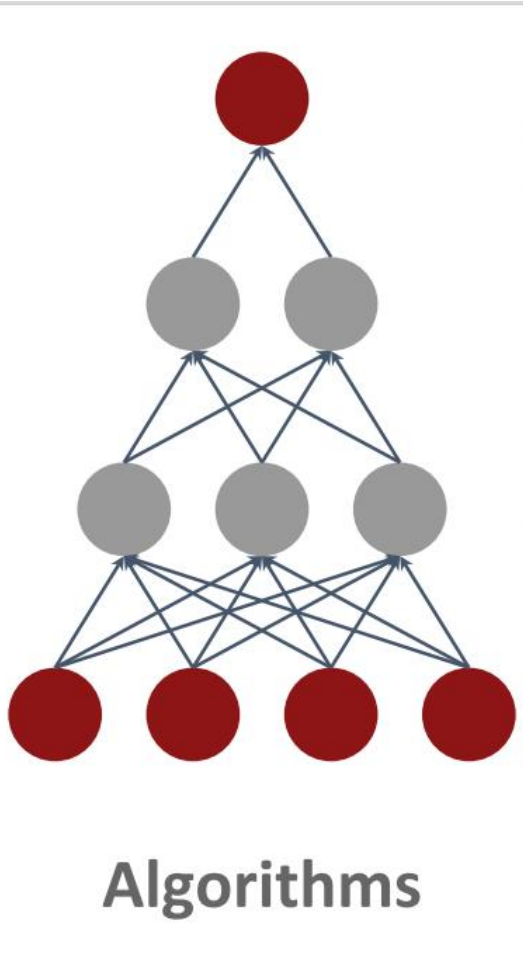
TEXT PROMPT

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

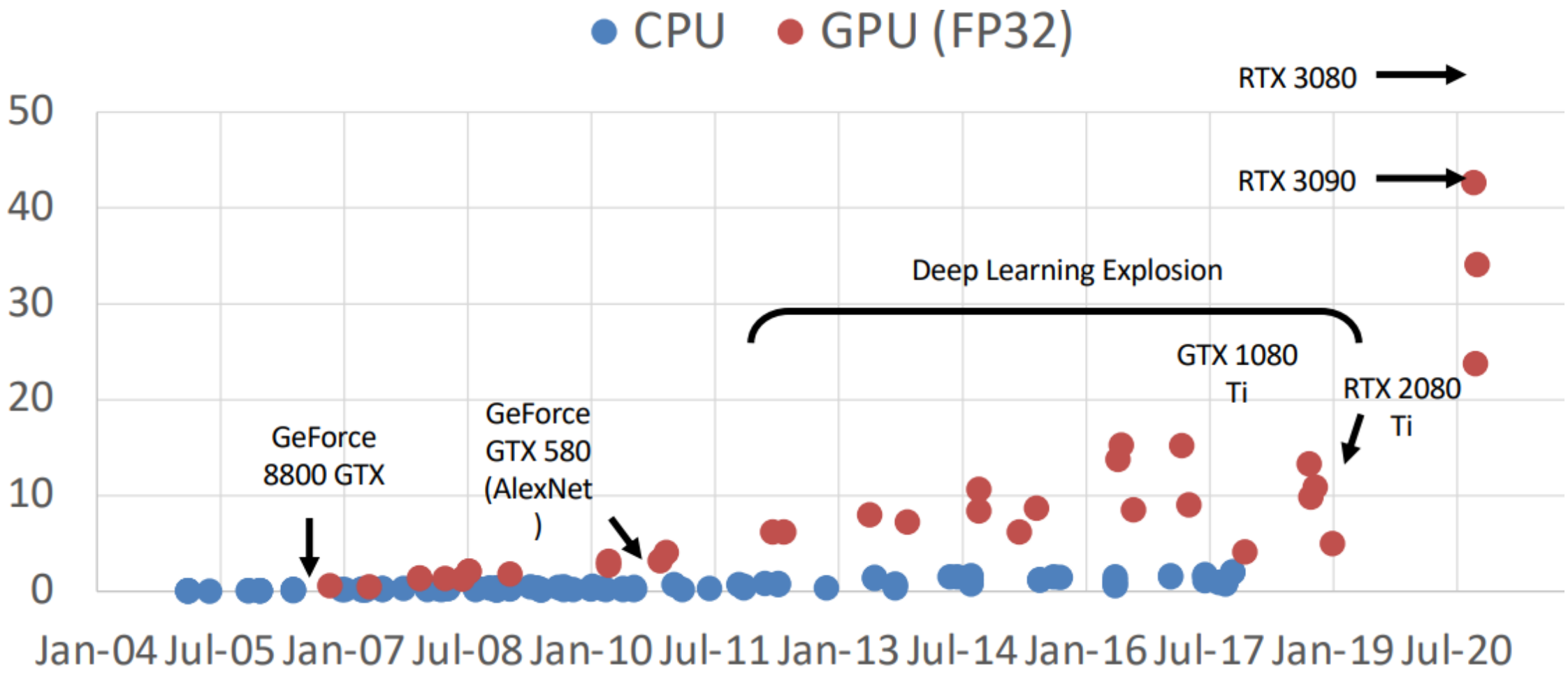
AI-GENERATED IMAGES



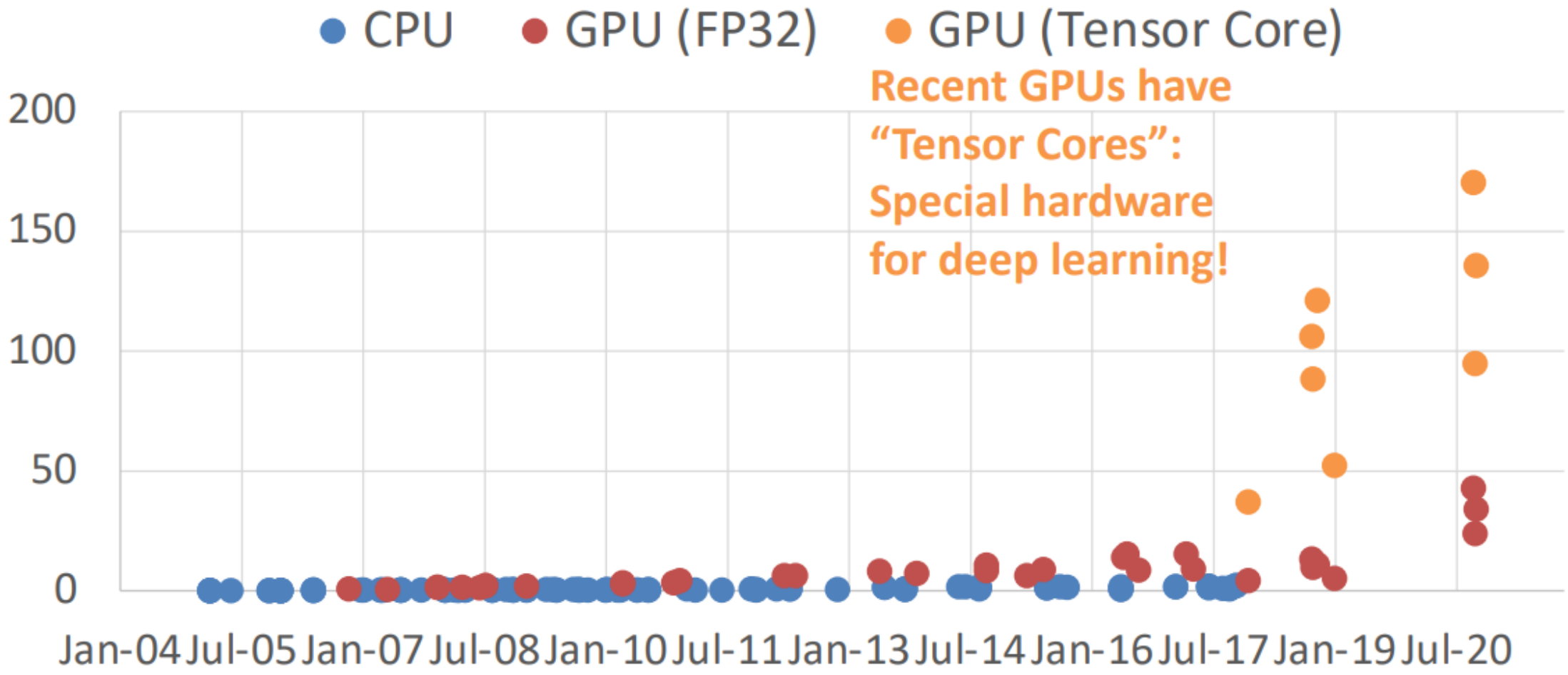
Deep Learning is Everywhere



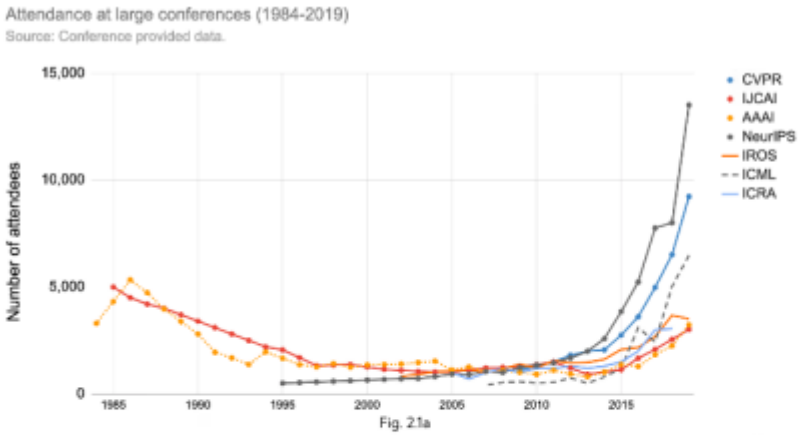
GFLOP per Dollar



GFLOP per Dollar

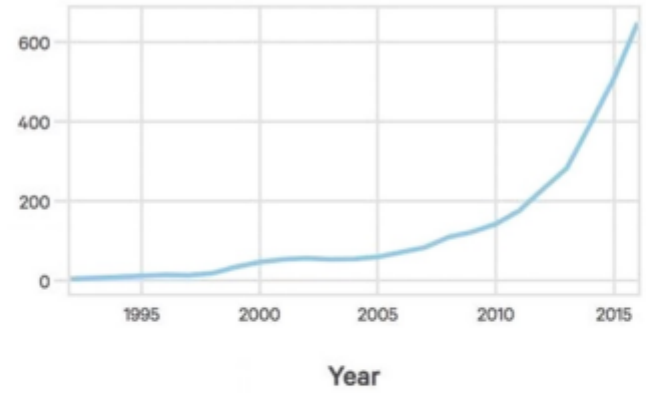


AI's Explosive Growth & Impact



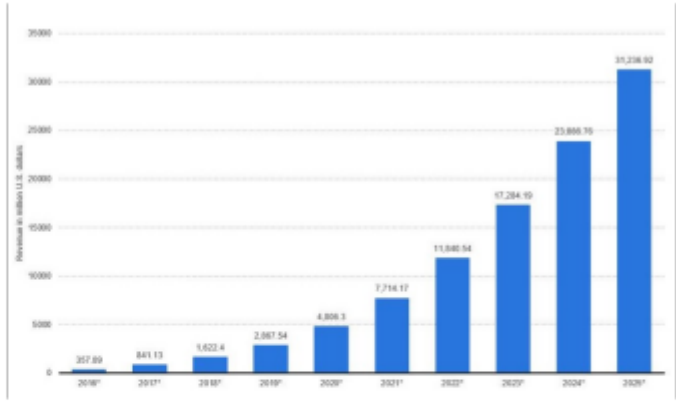
Number of attendance At AI conferences

Source: The Gradient



Startups Developing AI Systems

Source: Crunchbase, VentureSource, Sand Hill Econometrics



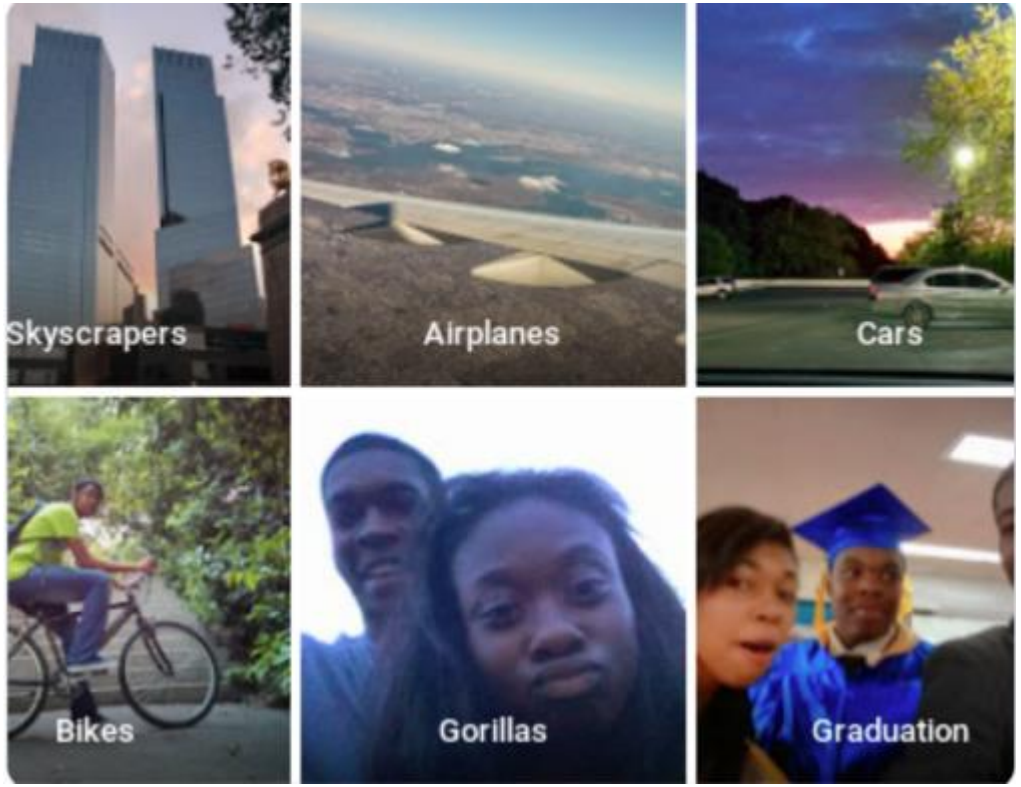
Enterprise Application AI Revenue

Source: Statista

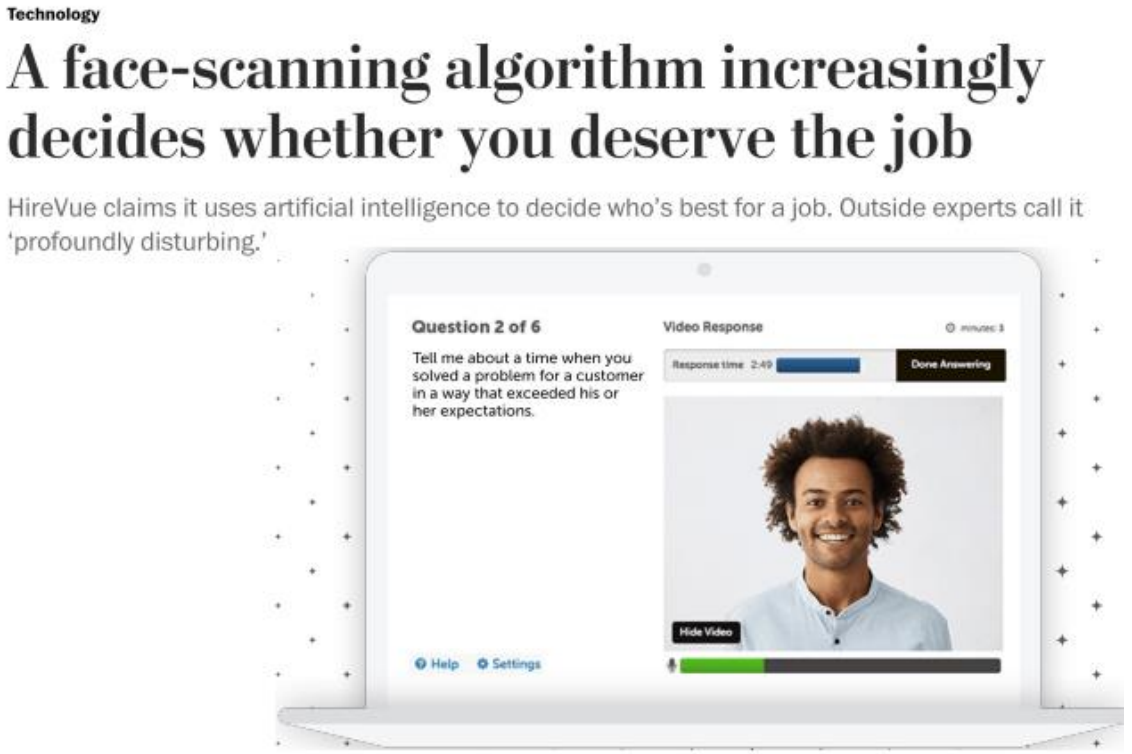
Despite the successes,
computer vision still has a long way to go

Computer Vision Can Cause Harm

Harmful Stereotypes



Affect people's lives



Computer Vision Can Save Lives

There is a lot we don’t know how to do

PT = 500ms

- Some kind of game or fight. Two groups of two men? The man on the left is throwing something. Outdoors seemed like because i have an impression of grass and maybe lines on the grass? That would be why I think perhaps a game, rough game though, more like rugby than football because they pairs weren't in pads and helmets, though I did get the impression of similar clothing. maybe some trees? in the background. (Subject: SM).

There is a lot we don’t know how to do

Chúc các bạn học tốt

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN TP.HCM

Nhóm UIT-Together

TS. Nguyễn Tấn Trần Minh Khang