Applications of Computer Vision and Deep Learning

Overview of the course

Jimut Bahan Pal
Research Scholar - CMInDS
PMRF, CMInDS Fellow

Conducted Online
RKMVERI
IIT Bombay
Saturday, January 13, 2024

Who Am I & Introductions!

- M.Sc. in Computer Science, RKMVERI, 2019-2021 batch!

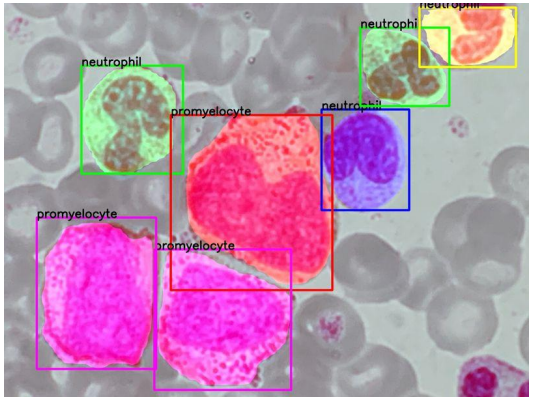
Who Am I & Introductions!

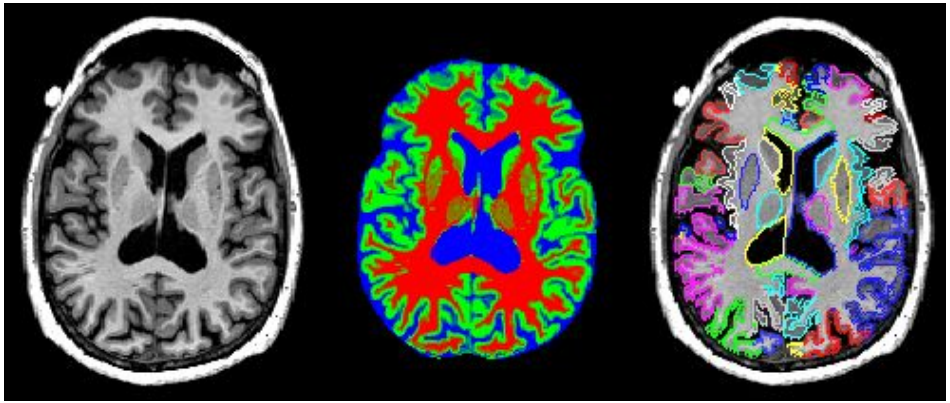
- IIT Bombay CMInDS, 2022 - Current (No idea when this will end) Ph.D. batch

Who Am I & Introductions!

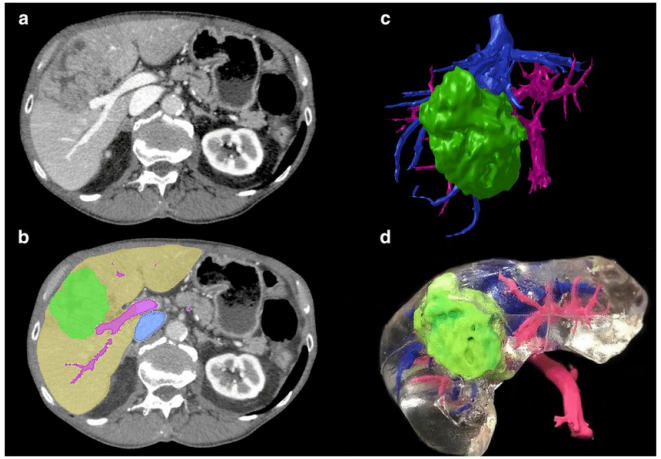
- M.Sc. in Computer Science, RKMVERI, 2019-2021 batch!
- Tell me about yourself – Planning for job/research/interest/background etc.
- This course is for the completion of my PMRF requirements, where each semester, I shall complete 25 hours of teaching! (With exception, this sem about 40 hours)
- First time offering a course!
- Mostly related to programming/hands on experience.
- Flexible syllabus
- Philosophy of learning together! (I don’t know everything)
- Concept of recitation and other stuffs.
- Mostly conducted @ Saturday between 4:00 - 6:00 P.M.

Application of Computer Vision - on Medical Images





Pathology, Blood cell counting

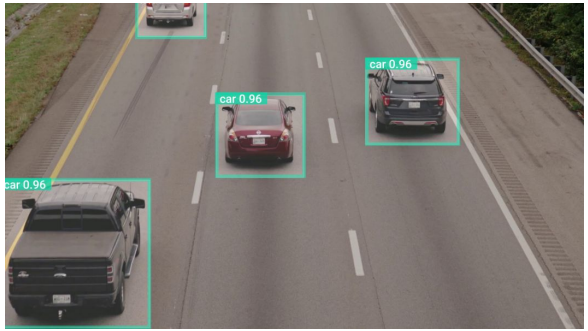


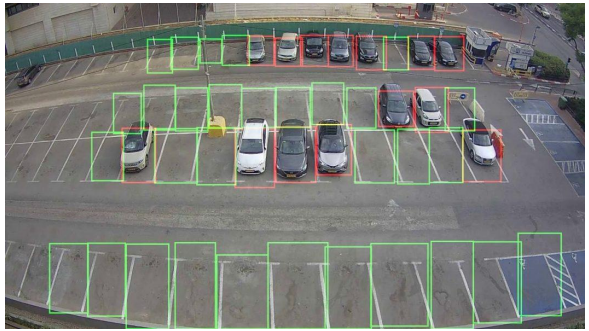
3D study of organs

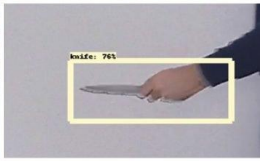
http://dx.doi.org/10.1007/s00330-019-06511-2
https://www.nih.gov/news-events/nih-research-matters/annual-chest-x-rays-dont-cut-lung-cancer-deaths https://soundray.org/maper/
https://www.mountsinai.org/health-library/tests/abdominal-ct-scan
https://jewelwc.org/wp-content/uploads/2020/11/ultrasound-compressed.jpg

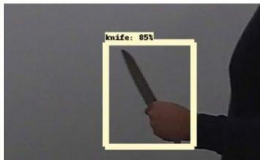
Disease Classification, Detection and Segmentation from X-Rays, MRI, CT (Abdominal), Ultrasound

More Applications - on Surveillance







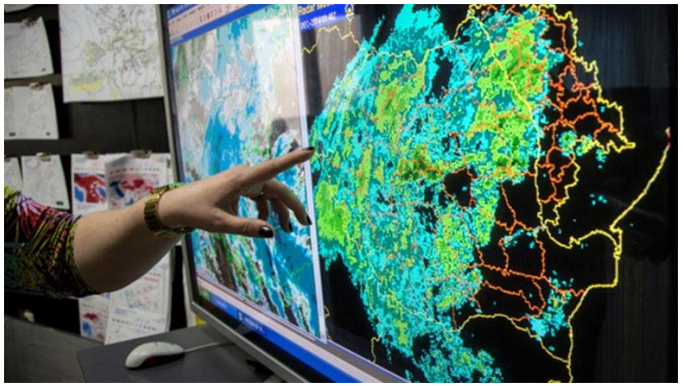


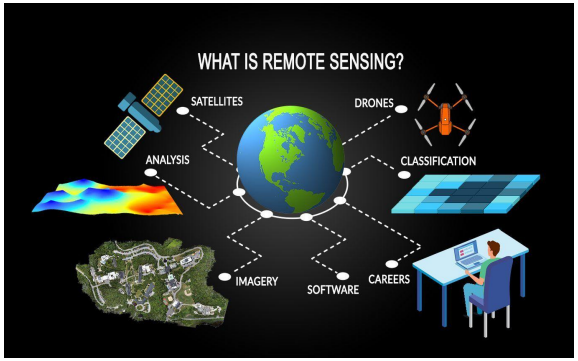
https://viso.ai/applications/computer-vision-applications-in-surveillance-and-security/

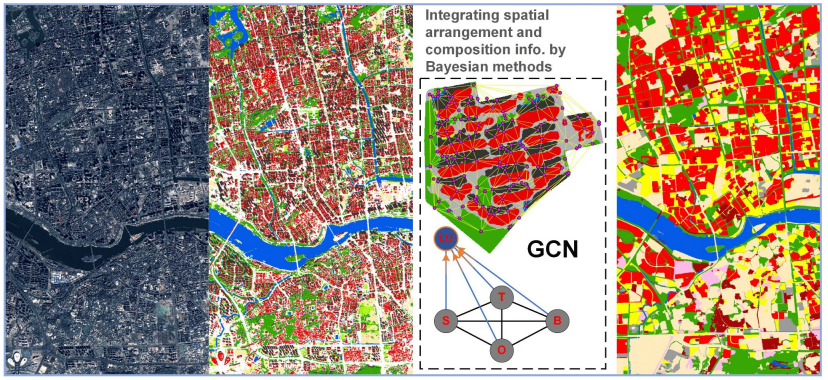
https://github.com/agentmorris/agentmorrispublic/blob/main/drone-datasets.md

Traffic speed analysis (congestion etc), automated parking ticket, farm animal drone surveillance, trajectory analysis of pedestrians, weapon detection in airports, number plate recognition for automated fine etc.

More Applications - Geospatial Imaging and Analysis



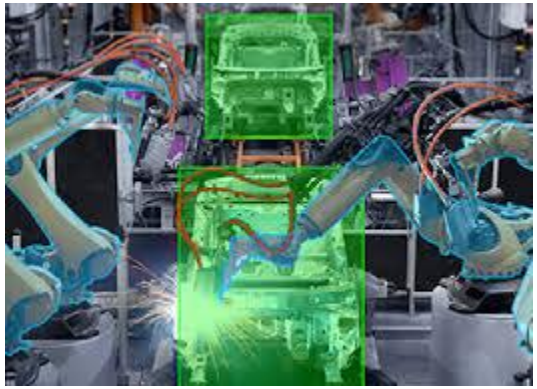




Satellite images -> Street Maps
Satellite images -> 3D models
Automated weather forecasting
Remote sensing, land cover analysis
Segmentation and analysis of objects in maps via Graph Convolutional Networks

https://www.worldbank.org/content/dam/photos/780x439/2017/sep/afr-improving-weather-forecasts-can-reduce-losses-to-development-in-africa-780x439.jpg
https://www.researchgate.net/profile/Aileen-Chang-3/publication/26690935/figure/fig2/AS:203158279200778@1425448116641/Example-of-Printed-Satellite-Map-for-Data-Collection-in-Neighborhood-Beholden-Public.png
https://mms.businesswire.com/media/20220420005388/en/1425965/5/BLSK-MAXAR_PR_image-01.jpg
https://pub.mdpi-res.com/remotesensing/remotesensing-12-04158/article_deploy/html/images/remotesensing-12-04158-ag.png?1608542088
https://scene.iitmandi.ac.in/uploads/specialization/1687553803146823.jpg

More Applications - Robotics, Cars & Surgery

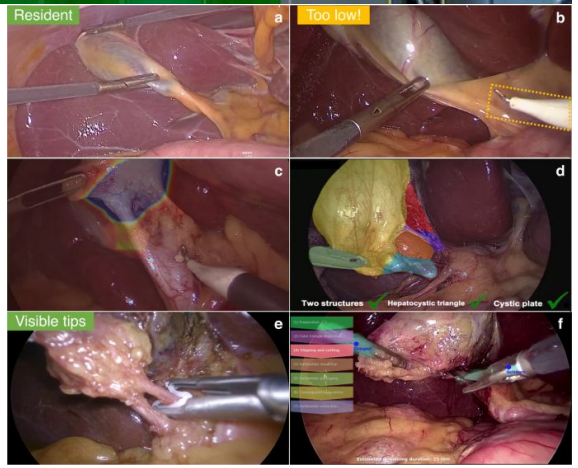


Robots in pipeline- factory processing

Autonomous vehicles with lidar + image segmentation methods.

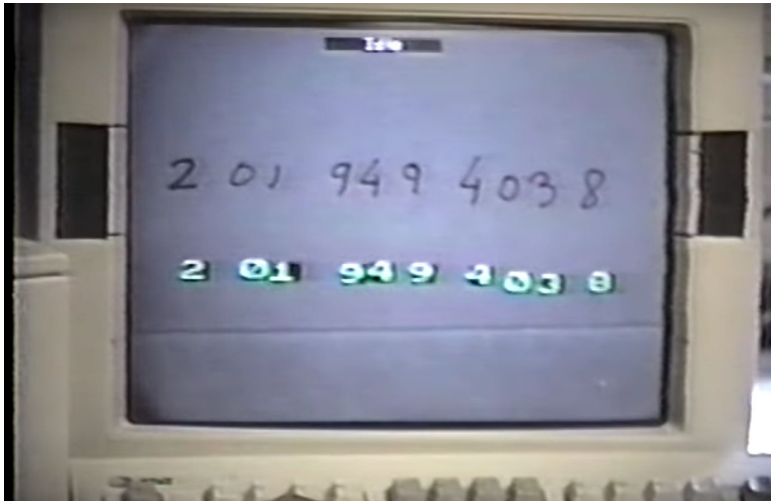
3D Image Segmentation for cars Detection of objects during driving

Robotic surgery of human organs for assistance

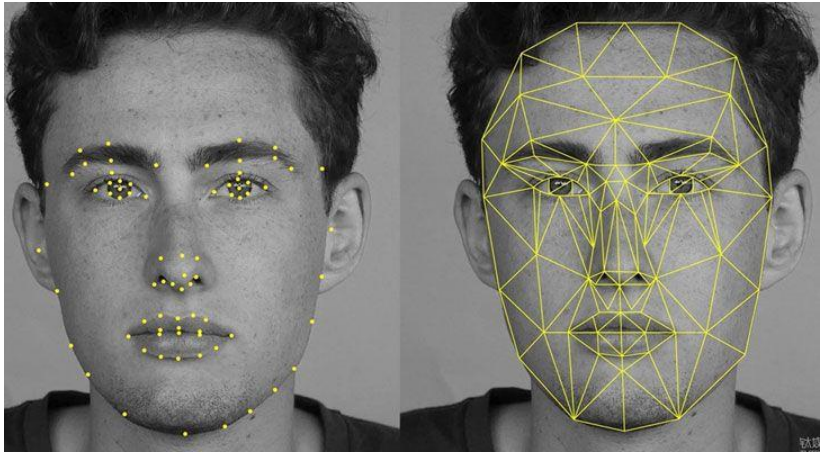


https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcTIeXU5vPVKSxDxli_BSh-7AIZZzWLNYl9j4Q&usqp=CA
https://images.inc.com/uploaded_files/image/1920x1080/Example_65th-(1)-1_379364.jpg
https://www.sustained-quality.com/wp-content/uploads/sites/24/2018/06/computer-visions-impact-automotive-industry.jpg
https://media.springernature.com/lw685/springer-static/image/art%3A10.1038%2Fs41746-022-00707-5/MediaObjects/41746_2022_707_Fig2_HTML.png

More Applications - OCR, Biometrics & VRs

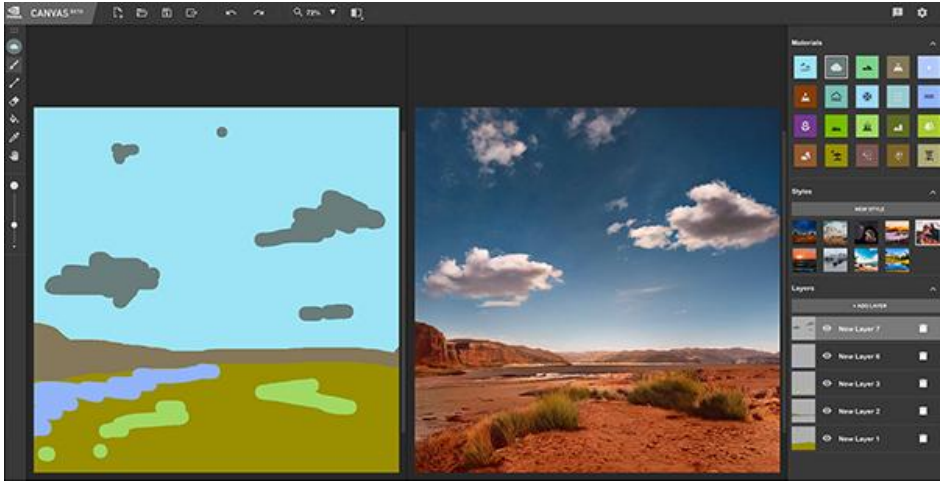


Yann LeCunn OCR 1989 application of LeNet, Fingerprint recognition for biometrics, Apple Face Unlock feature and Entertainment via VR requires computer vision.



https://www.youtube.com/watch?v=FwFduRA_L6Q
https://www.globalsign.com/application/files/5315/6260/6129/biometrics-fingerprint-.png
https://im.indiatimes.in/media/content/2017/Sep/facial-recog_1505726689.jpg
https://av-eks-lekhak.s3.amazonaws.com/media/__sized__/article_images/Screenshot_2023-06-07_164658-thumbnail_webp-600x300.webp
http://yann.lecun.com/exdb/lenet/rotation.html

Applications - Art & More…



NVIDIA's Canvas App for turning paint-strokes into realistic images with different textures and scene styles.

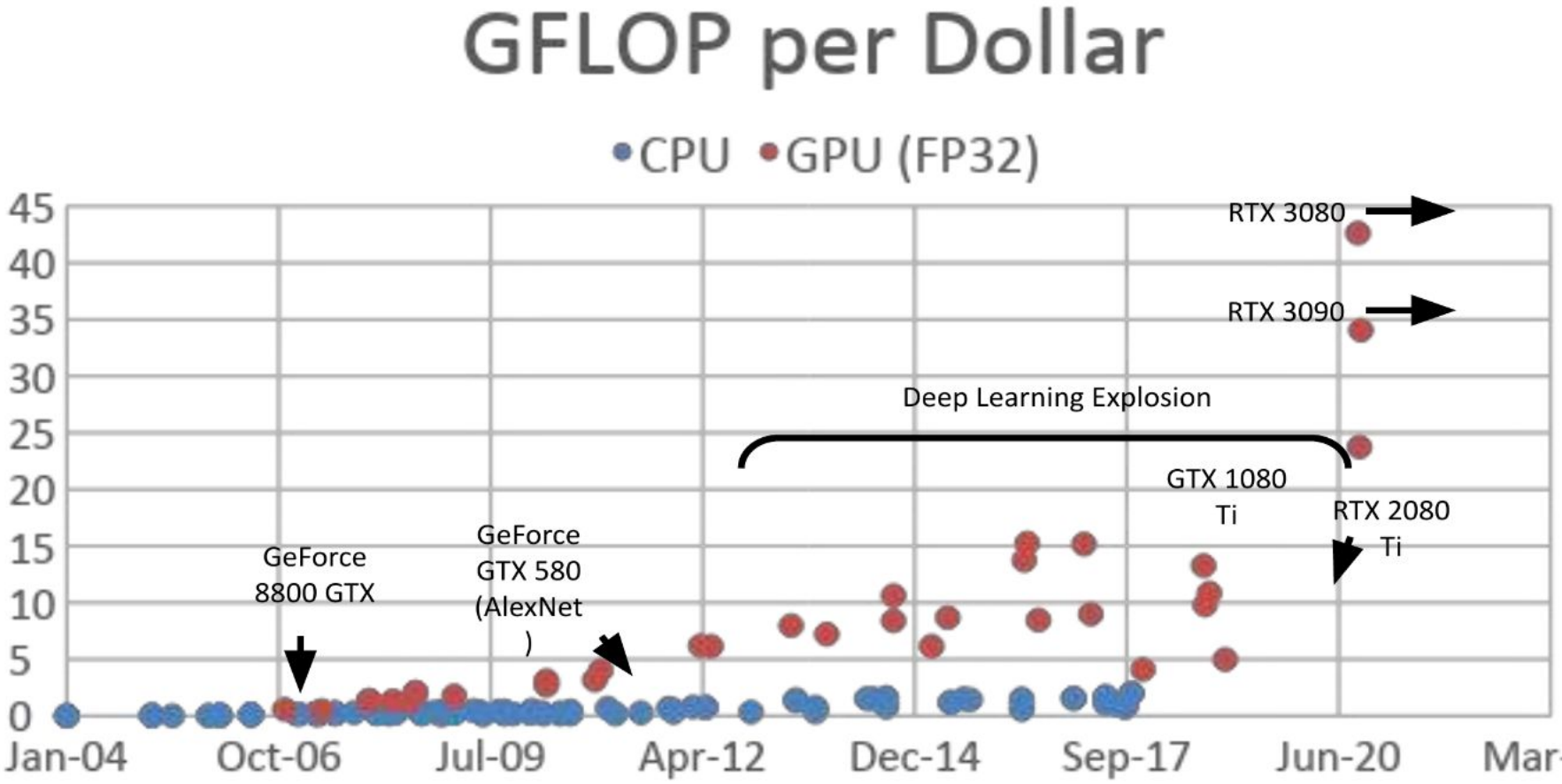
Neural Style transfer to capture different artist's styles.

Images generated by DALL-E, Midjourney, stable diffusion etc, uses texts as inputs.

https://www.nvidia.com/en-in/studio/canvas/
https://images.squarespace-cdn.com/content/v1/5bfd6f322487fdb33ce53474/a8c2b51c-94db-4260-9034-3abec09430b2/ImageGeneration
https://assets-global.website-files.com/5d7b77b063a9066d83e1209c/627d124493a74d43a0294052_613ec308797490827ea06868_mona%2520lisa.png

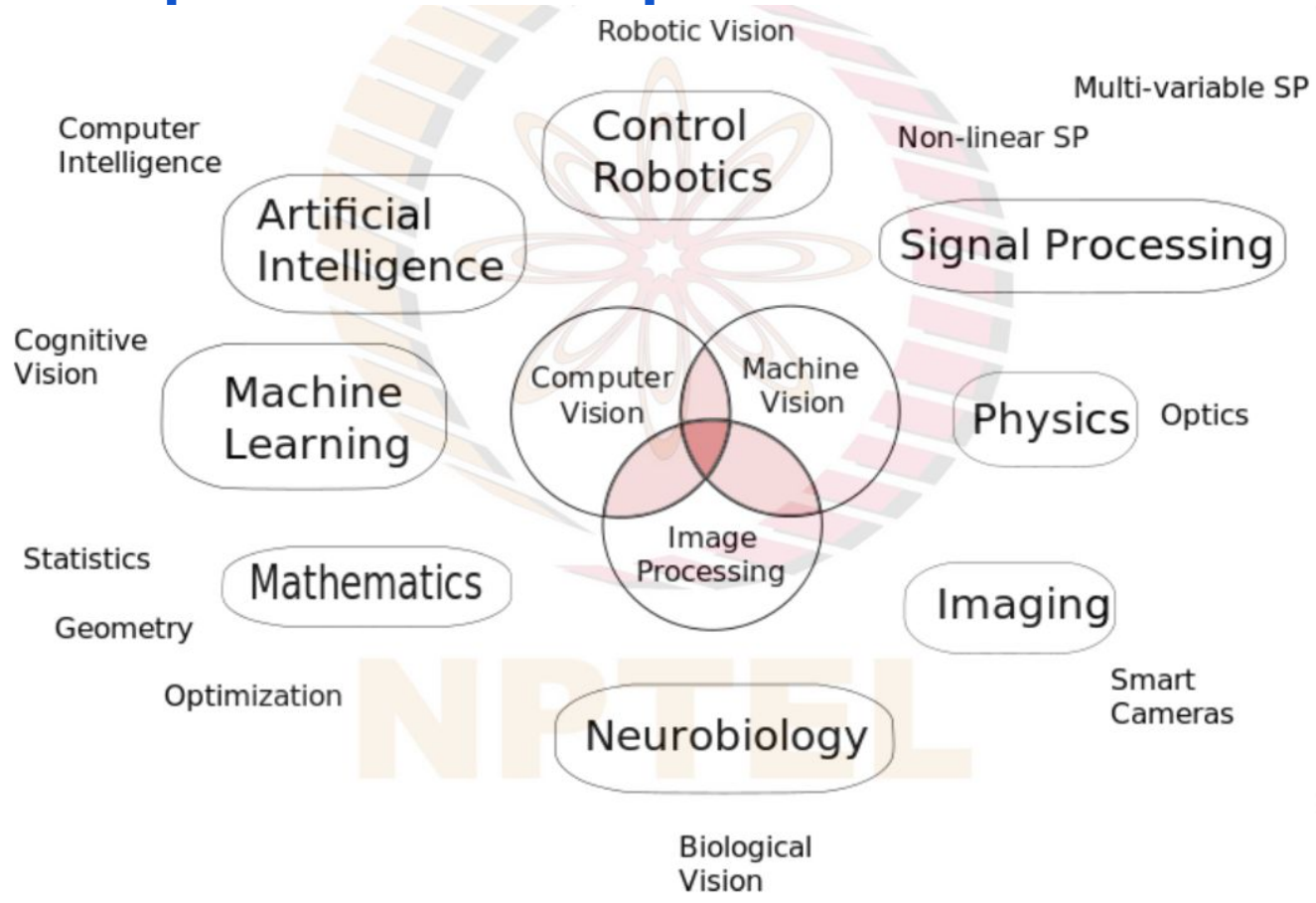
"A cat and a female character in a spaceship exploring a hidden galaxy. With detailed backgrounds, expressive characters, including magical elements, illustration made by hand."

Cost GPUs



Slide inspiration: Justin Johnson

Perspective of Computer Vision



https://dl4cv-nptel.github.io/DL4CVBK/intro.html
https://medium.com/@anishchopra/computer-vision-why-is-this-so-difficult-2b4f22e94efe

A brief History of Computer Vision and Deep Learning

- **A wide variety of animals burst onto the evolutionary scene in an event known as the Cambrian explosion 530 - 540 million years…**

T-rex's insanely brilliant vision system

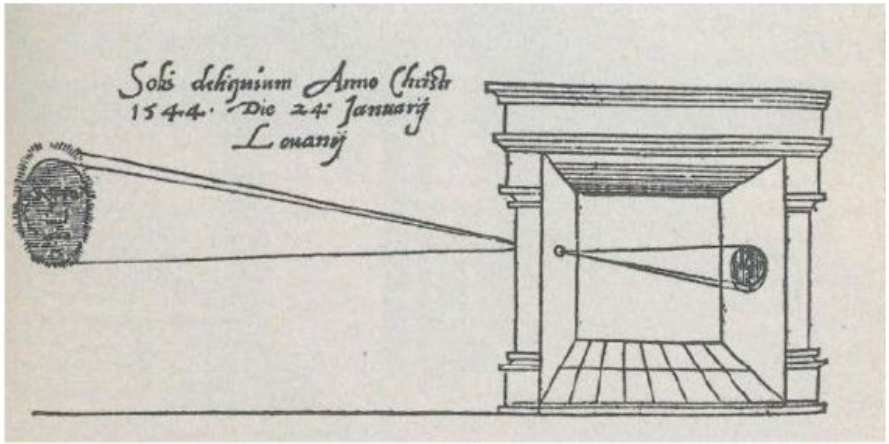
- T rex's eye was around 10-13 centimeters, and had binocular vision!
- Was **13 times better than human,** and probably 100 times better than my eyesight!

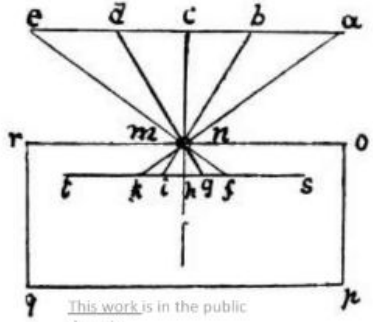
https://www.tandfonline.com/doi/pdf/10.1671/0272-4634(2006)26%5B321%3ABVITD%5D2.0.CO%3B2
https://daily.jstor.org/five-things-you-probably-have-wrong-about-the-t-rex/

Stanford CS-231n
course materials

Pin hole camera/Camera Obscura

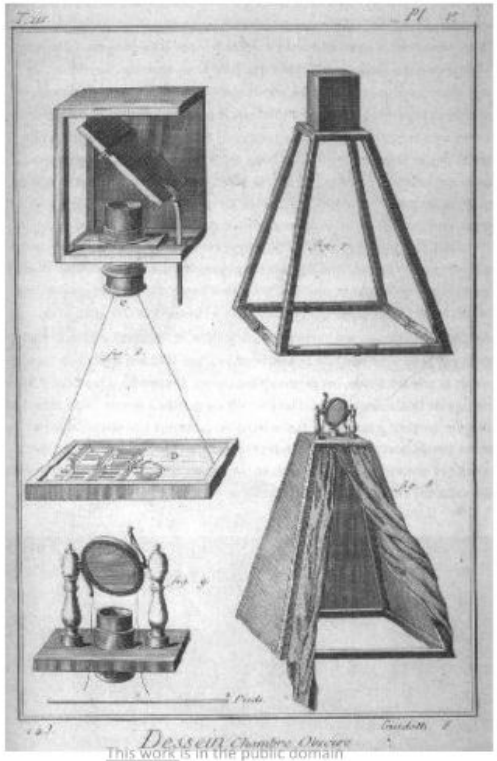
Gemma Frisius, 1545





Leonardo da Vinci, 16th Century AD

Encyclopedia, 18th Century

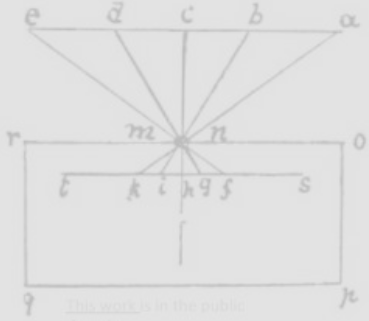


Pin hole camera/Camera Obscura

Gemma Frisius, 1545

This work is in the public domain

Leonardo ‘the’ Vinci

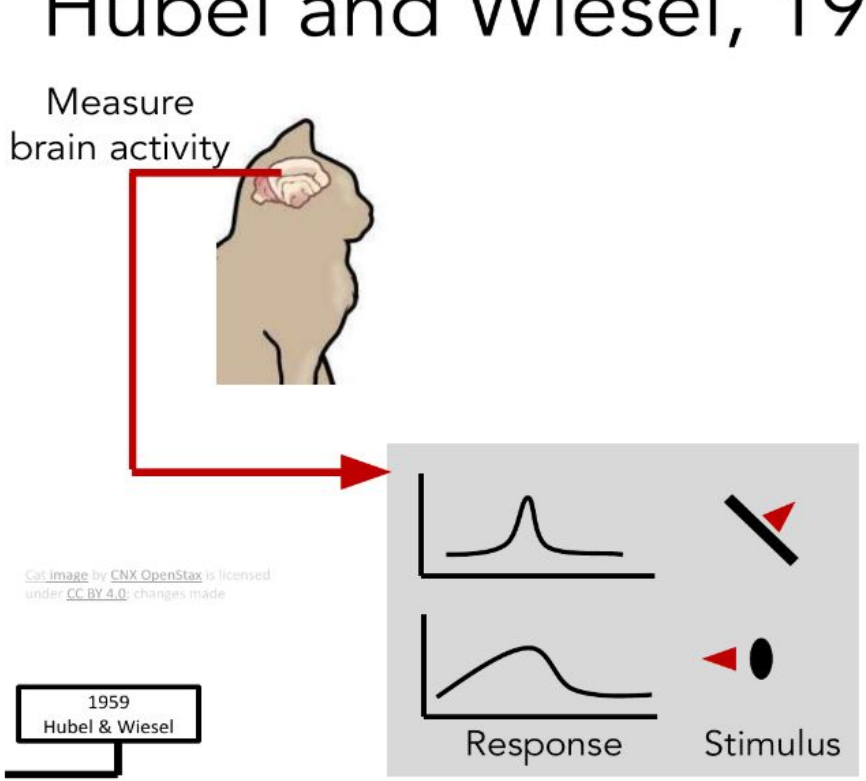


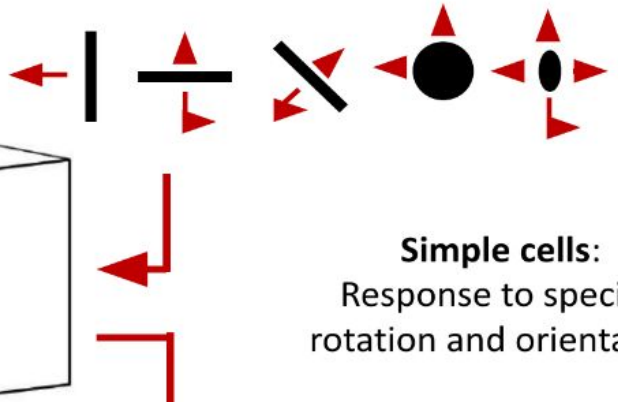
This work is in the public domain

Leonardo da V
16th Century A

Cat's response to stimulus

Hubel and Wiesel, 1959





Simple cells:
Response to specific rotation and orientation

Complex cells:
Response to light orientation and movement, some translation invariance

Slide inspiration: Justin Johnson

Era of Computer Vision

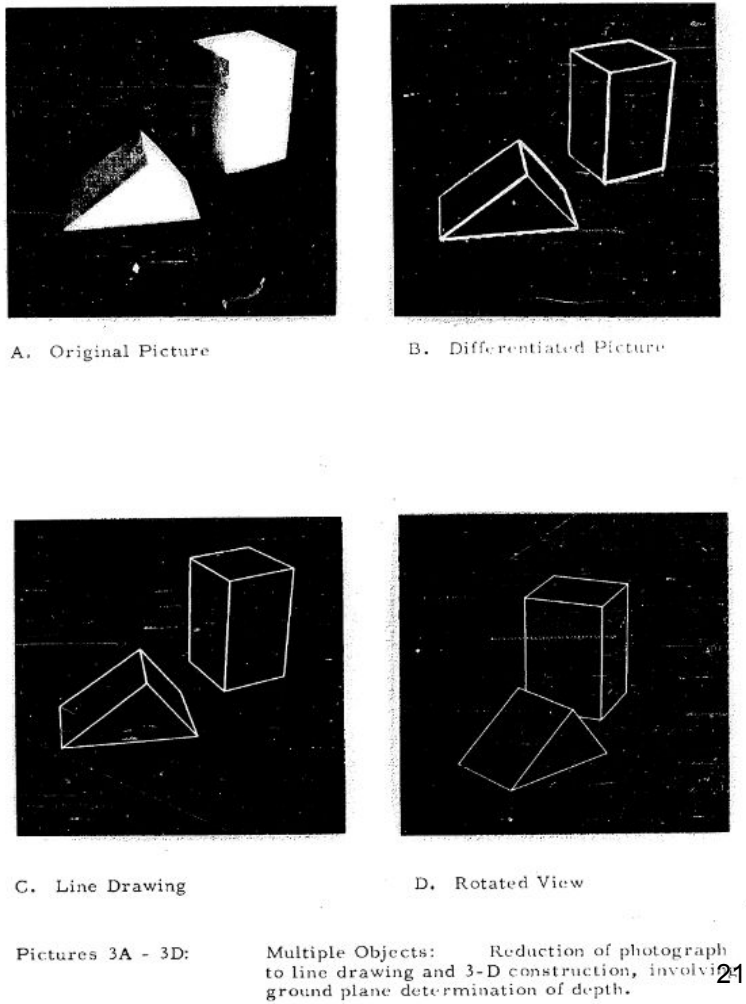
- 1960 - Early 1990's — The Geometric Era
- 1990's — Appearance based models
- Mid 90's — Sliding Window Based Approaches
- Late 1990's — Local features
- Early 2000's — parts and shape models
- Mid 2000's — bags of features
- Present trends — data-based methods, knowledge graphs, combination of local and global methods, contexts etc.

Era of Computer Vision

- **1960 - Early 1990's — The Geometric Era**
- 1990's — Appearance based models
- Mid 90's — Sliding Window Based Approaches
- Late 1990's — Local features
- Early 2000's — parts and shape models
- Mid 2000's — bags of features
- Present trends — data-based methods, knowledge graphs, combination of local and global methods, contexts etc.

Era of Computer Vision - 1960's

- **Recognition as an alignment problem: Block world - 1963 - Roberts!**
- Find edge, select feature point, connect via line, line fitting, 3D shape.
- Rotate the 3D shape in any axis for creating world view of the model
- Different innovative techniques were created back then.
- Please check the thesis: https://dspace.mit.edu/handle/1721.1/11589



A. Original Picture

B. Differentiated Picture

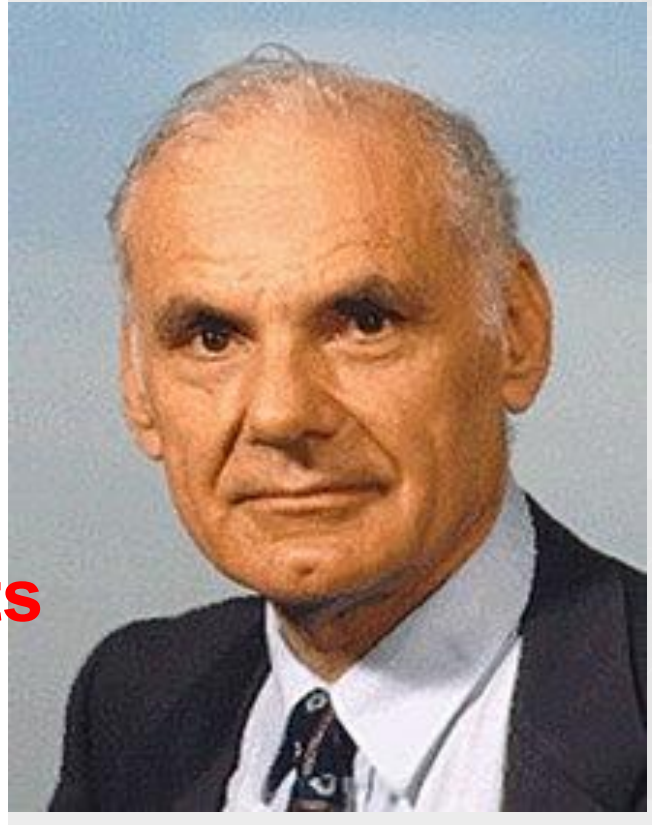
C. Line Drawing

D. Rotated View

Pictures 3A - 3D: Multiple Objects: Reduction of photograph to line drawing and 3-D construction, involving ground plane determination of depth.

Era of Computer Vision - 1960's

- **Recognition as an alignment problem: Block world - 1963 - Roberts!**
- Find edge, select feature point, connect via line, line fitting, 3D shape.
- Rotate the 3D shape in any axis for creating world view of the model
- Different innovative techniques were created back then.
- Please check the thesis: https://dspace.mit.edu/handle/1721.1/11589

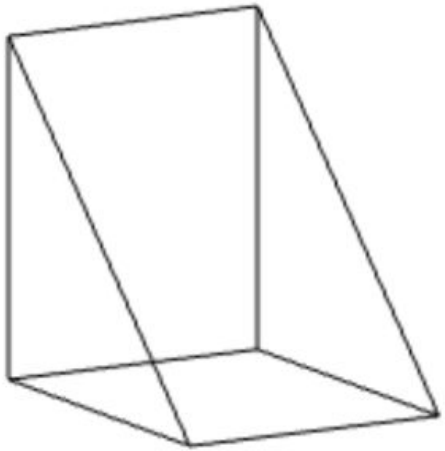


Larry Roberts

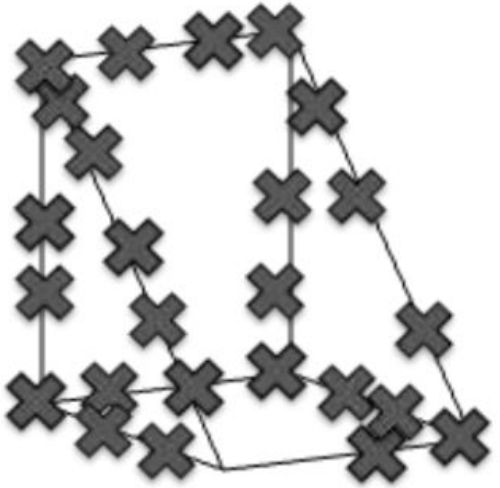
Era of Computer Vision - 1960's

Larry Roberts, 1963

(a) Original picture



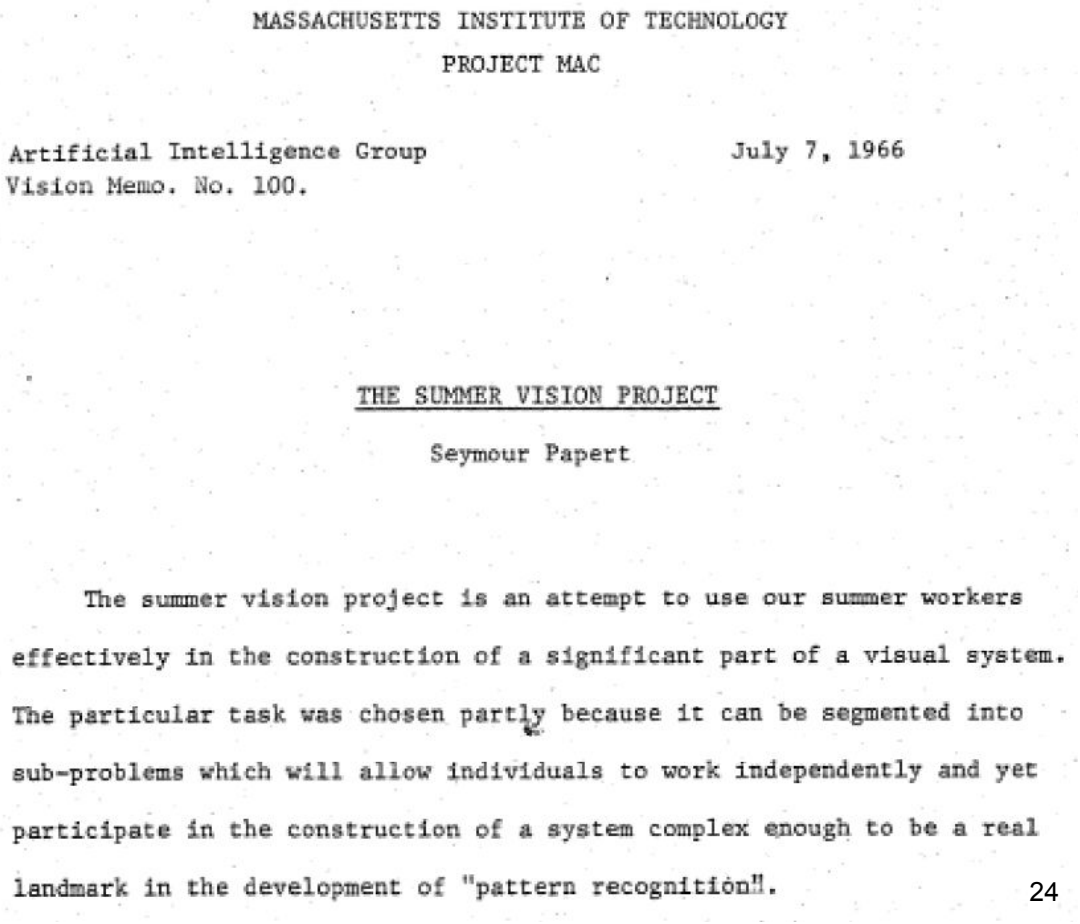
(b) Differentiated picture



(c) Feature points selected

Era of Computer Vision - 1966's - Papert

- Building a visual Recognition system by using summer workers, to be completed in a semesters.
- **Tasks divided into subtasks - More to be added as project progresses.**
- Getting more funding for more ideas.
- **Little did the know…**
- https://dspace.mit.edu/handle/1721.1/6125



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

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

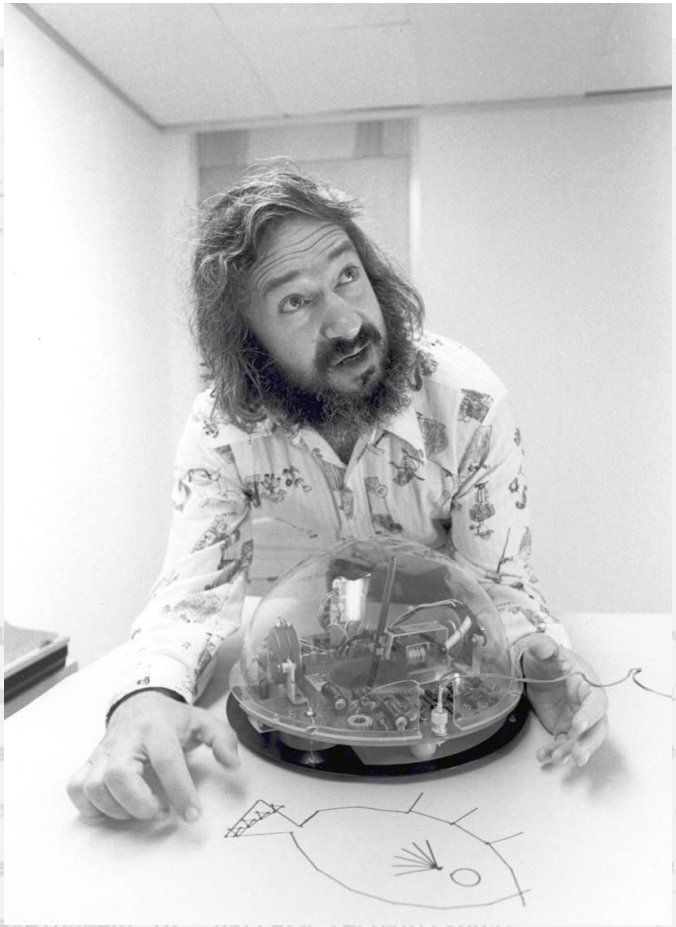
THE SUMMER VISION PROJECT

Seymour Papert

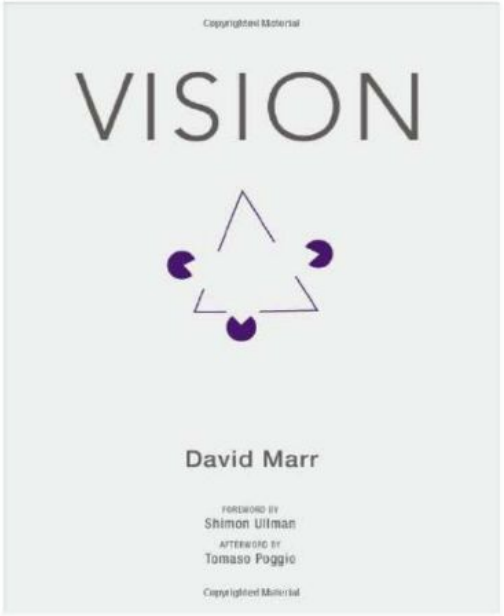
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the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

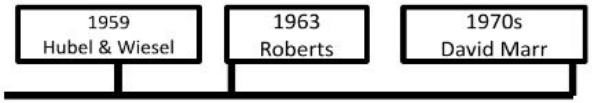
Era of Computer Vision - 1966's - Papert

- Building a visual Recognition system by using summer workers, to be completed in a semesters.
- **Tasks divided into subtasks - More to be added as project progresses.**
- Getting more funding for more ideas.
- **Little did the know…**
- https://dspace.mit.edu/handle/1721.1/6125

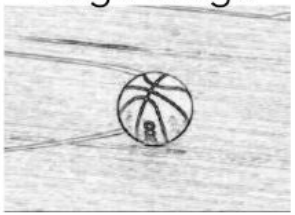


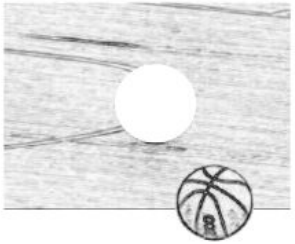
Era of Computer Vision - 1970's - David Marr

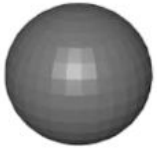




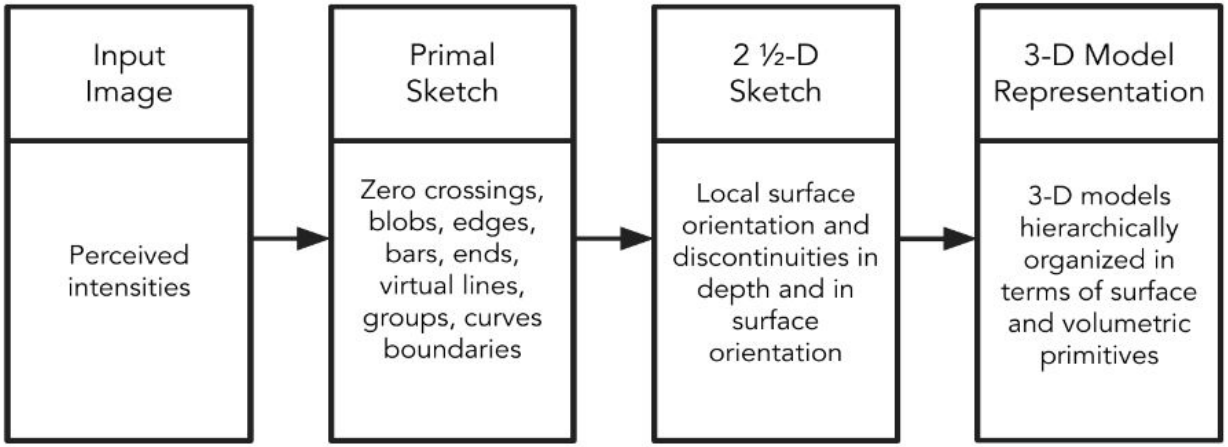
This image is CC0 1.0 public domain





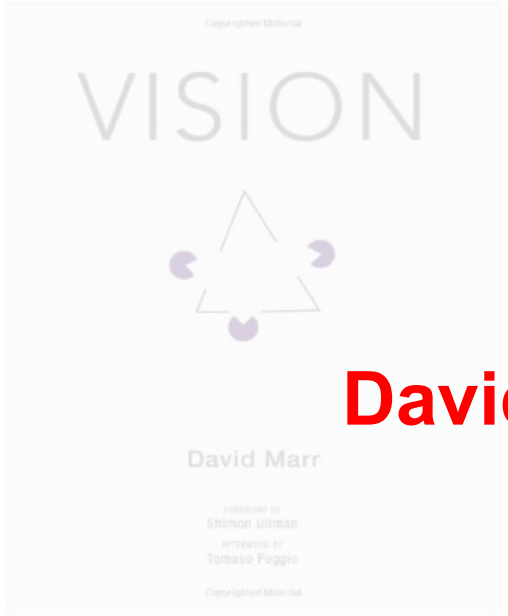


This image is CC0 1.0 public domain

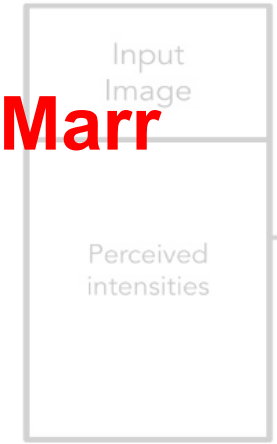


Stages of Visual Representation, David Marr, 1970s

Era of Computer Vision - 1970's - Da



This image is CC0 1.0 public domain

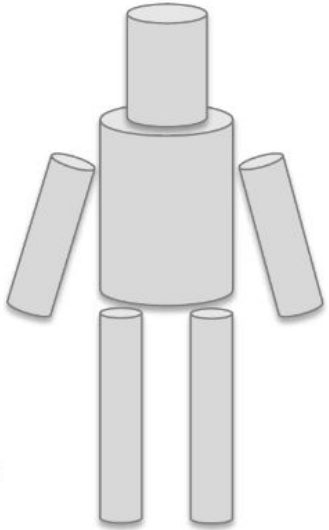


Stages of Visual Representation, David Marr, 1970s

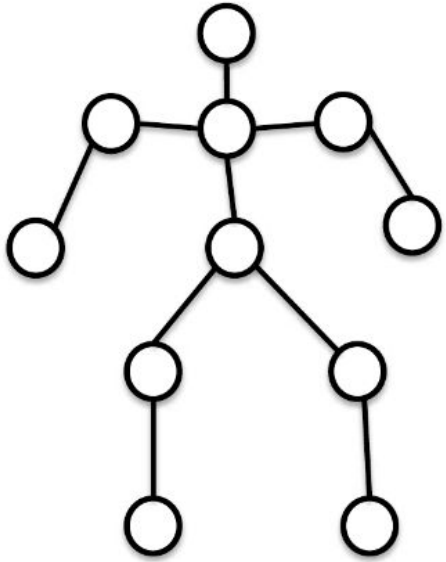
David Marr

Era of Computer Vision - 1979's - General Cylinders

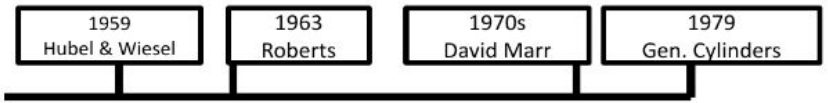
Recognition via Parts (1970s)



Generalized Cylinders,
Brooks and Binford,
1979



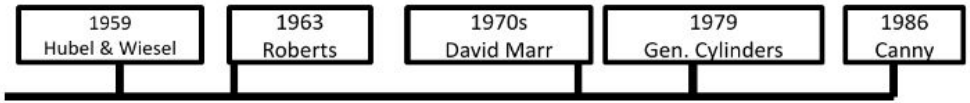
Pictorial Structures,
Fischler and Elshlager, 1973



Slide inspiration: Justin Johnson

Era of Computer Vision - 1986's - Canny

Recognition via Edge Detection (1980s)



John Canny, 1986
David Lowe, 1987

Stanford CS-231n course materials

Slide inspiration: Justin Johnson

Era of Computer Vision - 1986's - Canny

Recognition via Edg

John F. Canny

1959 Hubel & Wiesel

1963 Roberts

1970s David Marr

1979 Gen. Cylinders

1986 Canny

Stanford CS-231n course materials

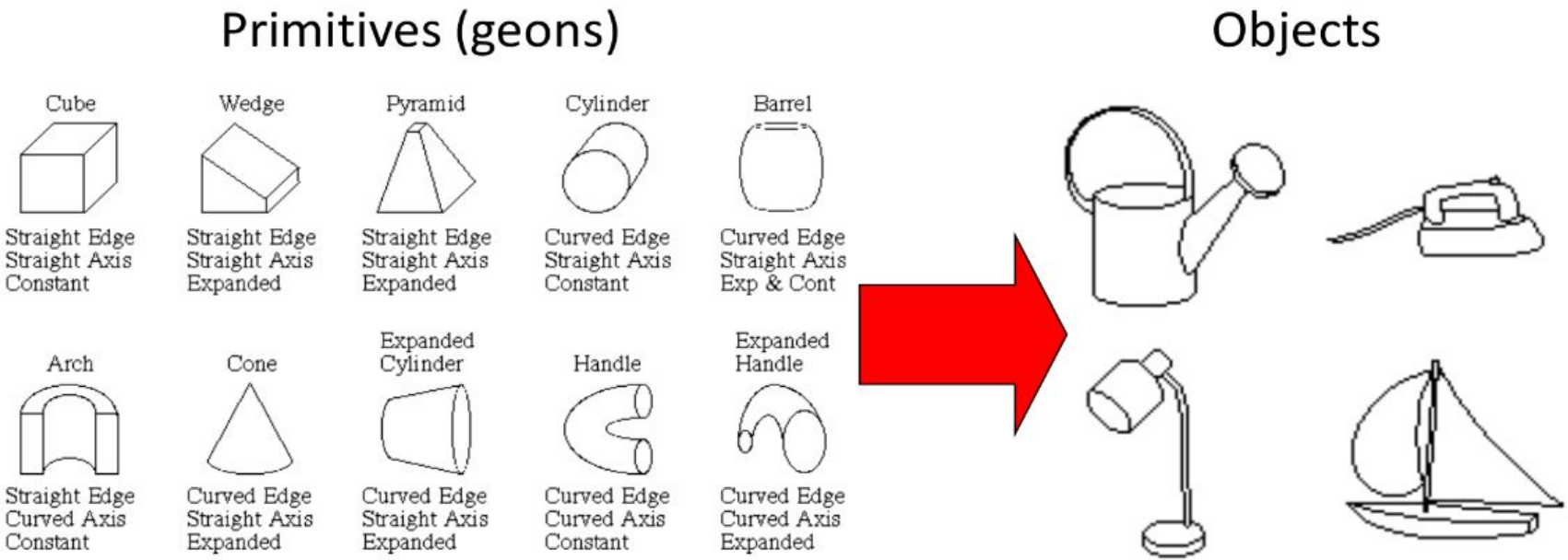
Slide inspiration: Justin Johnson

Era of Computer Vision - 1987's - Biederman

Recognition by components

GNR 638 course materials

Biederman (1987)



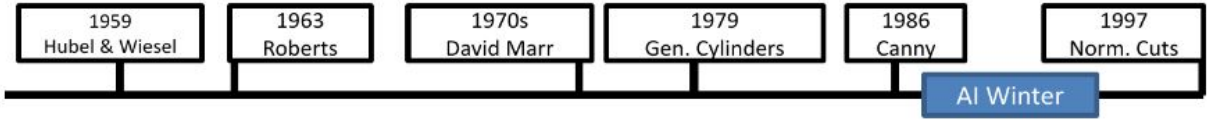
Arrival of AI Winter

- Enthusiasm for AI dwindled.
- “Expert Systems” - failed to deliver their promises
- Subfileds of AI continued to grow
  - NLP
  - Computer Vision
  - Robotics
  - Reinforcement Learning
  - Bayesian
  - Neural etc.

Era of Computer Vision

- 1960 - Early 1990's — The Geometric Era
- **1990's — Appearance based models**
- Mid 90's — Sliding Window Based Approaches
- Late 1990's — Local features
- Early 2000's — parts and shape models
- Mid 2000's — bags of features
- Present trends — data-based methods, knowledge graphs, combination of local and global methods, contexts etc.

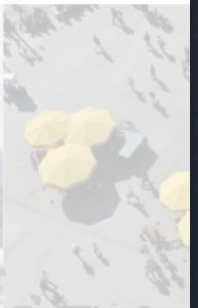
Era of Computer Vision - 1990's - Group based Recogn.

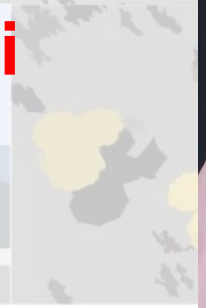


Normalized Cuts, Shi and Malik, 1997

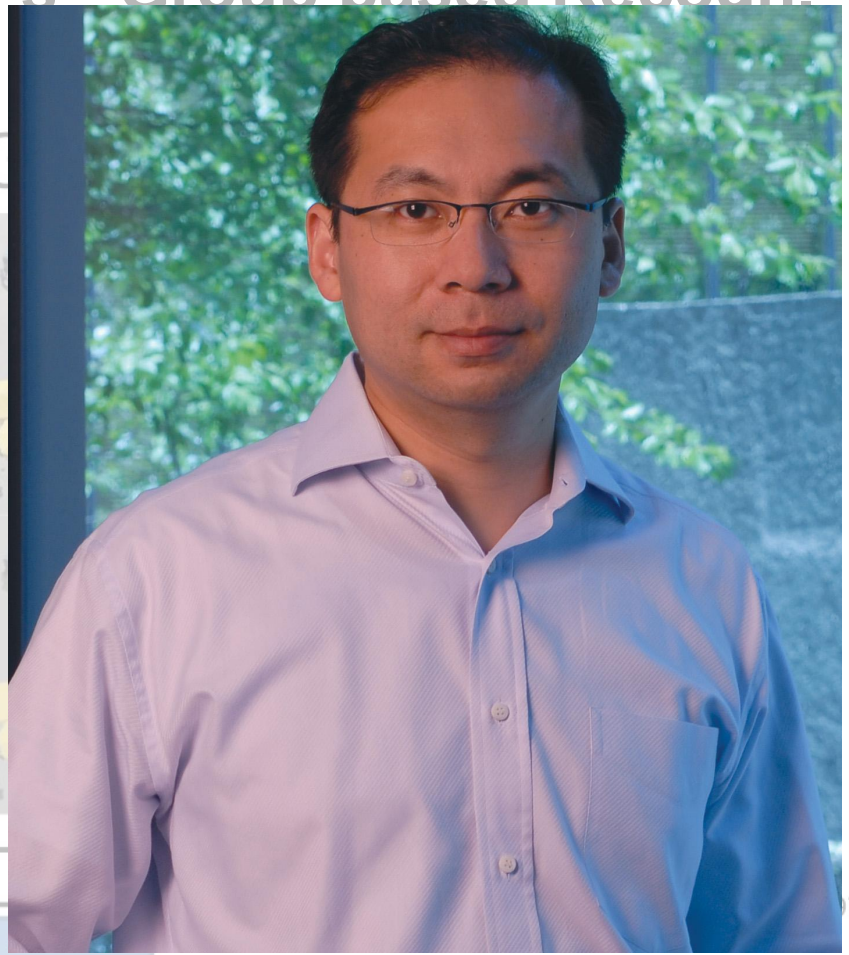
Era of Computer Vision - 1990's - Group based Recogn.

Recognition via G





Jianbo Shi



1959 Hubel & Wiesel

1963 Roberts

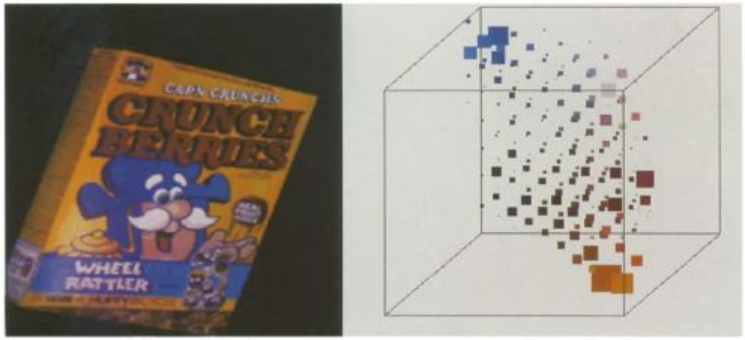
1970s David Marr

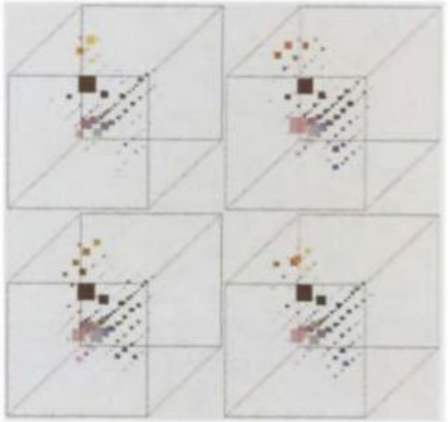
1979 Gen. Cylinders

1986 Canny

Era of Computer Vision - 1991's - Color histogram

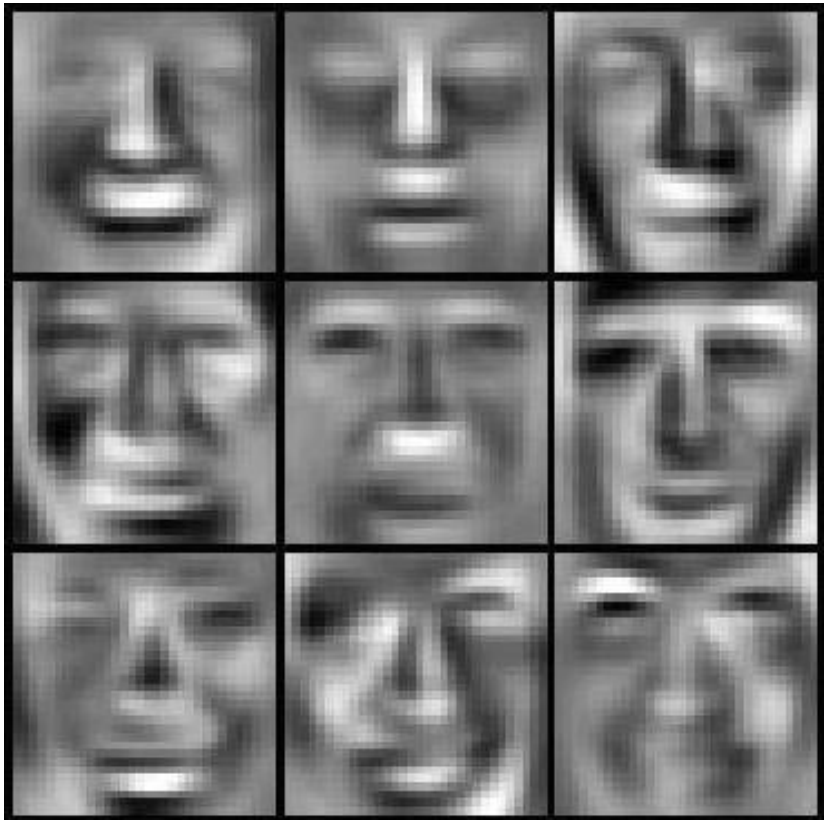
- Histogram based indexing of objects
- Real time performance.
- New innovative technique for image registration and recognition
- https://www.inf.ed.ac.uk/teaching/courses/av/LECTURE_NOTES/swainballard91.pdf





Era of Computer Vision - 1991's - Eigen Faces

- **Turk and Pentland - 1991**
- Face recognition - by computing eigenvector of faces.
- Comparing the eigenvector to get the closest distance and using them for recognition.
- So - Encoding faces just for compact and accurate representation.
- Can use SVD - Singular Value Decomposition for calculating the eigen vectors etc.
- Check the paper



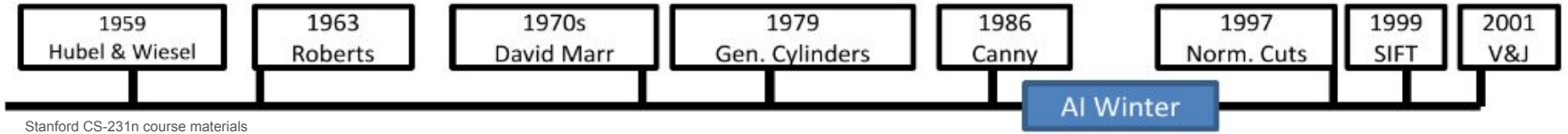
https://towardsdatascience.com/eigenfaces-recovering-humans-from-ghosts-17606c328184

Era of Computer Vision

- 1960 - Early 1990's — The Geometric Era
- 1990's — Appearance based models
- **Mid 90's — Sliding Window Based Approaches**
- Late 1990's — Local features
- Early 2000's — parts and shape models
- Mid 2000's — bags of features
- Present trends — data-based methods, knowledge graphs, combination of local and global methods, contexts etc.

Era of Computer Vision - 2001's - Sliding window based

- Viola and Jones - First application of Machine Learning to Computer Vision.
- Sliding window based
- Haar Cascade - uses features - hand engineered for face/ eye etc detection.
- Efficient and fast, but sometimes - just wrong… Inbuilt in OpenCV



Era of Computer Vision - 2001's - Sliding window based

- Viola and Jones - First application of Machine Learning to Computer Vision.
- Sliding window based
- Haar Cascade - uses features - Hand engineered for face/ eye etc detection.
- Efficient and fast, but sometimes - just wrong
- Inbuilt in OpenCV

Paul Viola and Michael Jones

1959 Hubel & Wiesel | 1963 Roberts | 1970s David Marr | 1979 Gen. Cylinders | 1986 Canny | AI Winter | 1997 Norm. Cuts | 1999 SIFT | 2001 V&J

Era of Computer Vision

- 1960 - Early 1990's — The Geometric Era
- 1990's — Appearance based models
- Mid 90's — Sliding Window Based Approaches
- **Late 1990's — Local features**
- Early 2000's — parts and shape models
- Mid 2000's — bags of features
- Present trends — data-based methods, knowledge graphs, combination of local and global methods, contexts etc.

Era of Computer Vision - 2000's - Matching based Recogn.

Recognition via Matching (2000s)

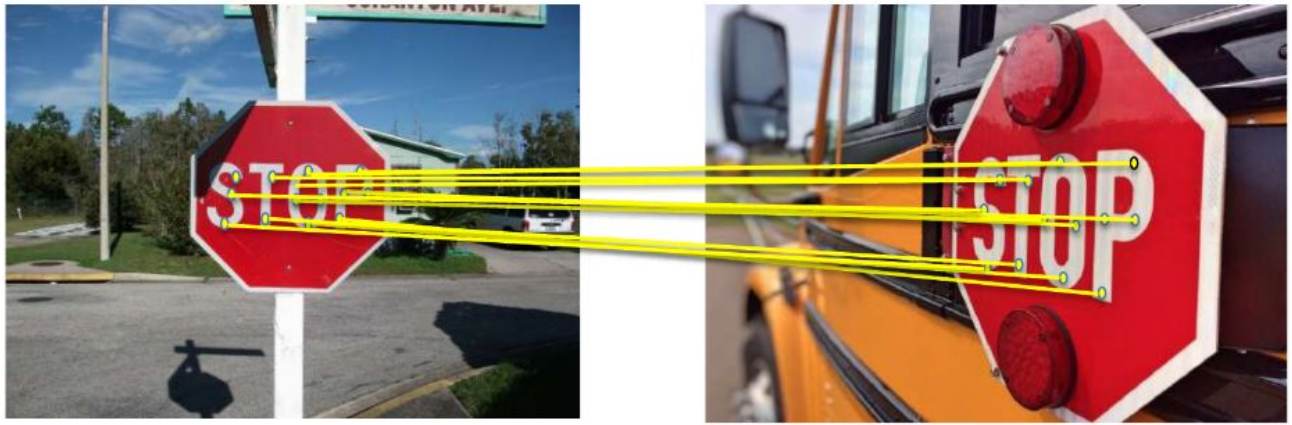
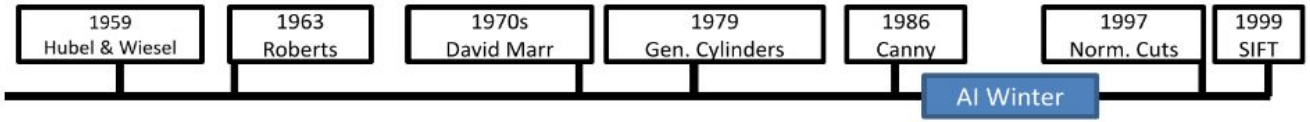


Image is public domain

Image is public domain



SIFT, David Lowe, 1999

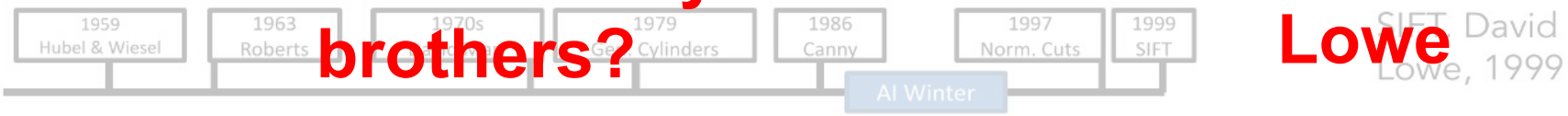
Era of Computer Vision - 2000's - Matching based Recogn.

Image is public domain

Image is public domain

But … are they brothers?

David G. Lowe



Era of Computer Vision - 2000's - Local features (D. Lowe)

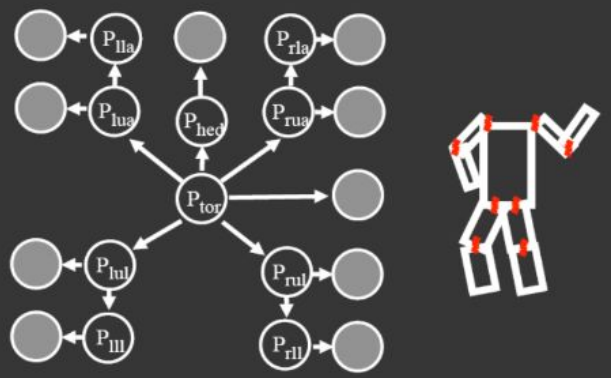
Era of Computer Vision

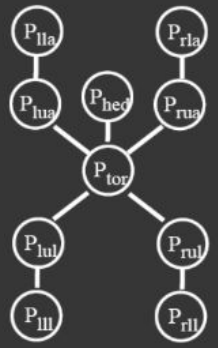
- 1960 - Early 1990's — The Geometric Era
- 1990's — Appearance based models
- Mid 90's — Sliding Window Based Approaches
- Late 1990's — Local features
- **Early 2000's — parts and shape models**
- Mid 2000's — bags of features
- Present trends — data-based methods, knowledge graphs, combination of local and global methods, contexts etc.

Parts and Shape Model

Pictorial structure model

Fischler and Elschlager(73), Felzenszwalb and Huttenlocher(00)

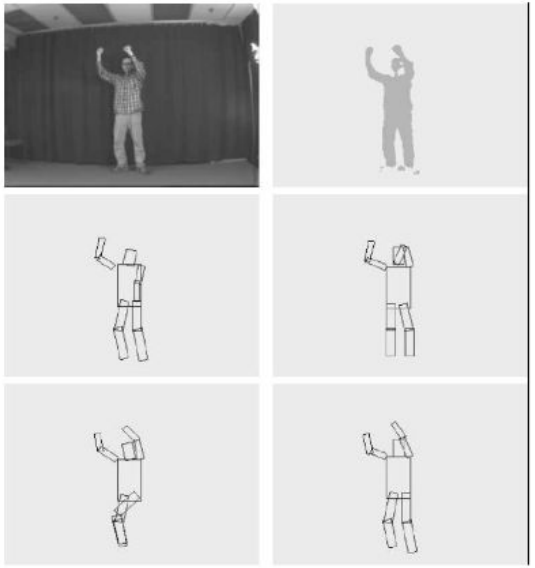




$$\Pr(P_{tor}, P_{arm}, \ldots | \mathrm{Im}) \propto \prod_{i,j} \Pr(P_i | P_j) \prod_i \Pr(\mathrm{Im}(P_i))$$

part geometry

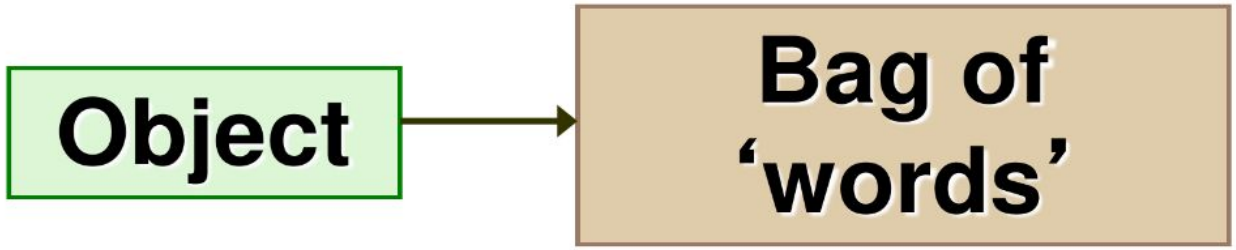
part appearance



Era of Computer Vision

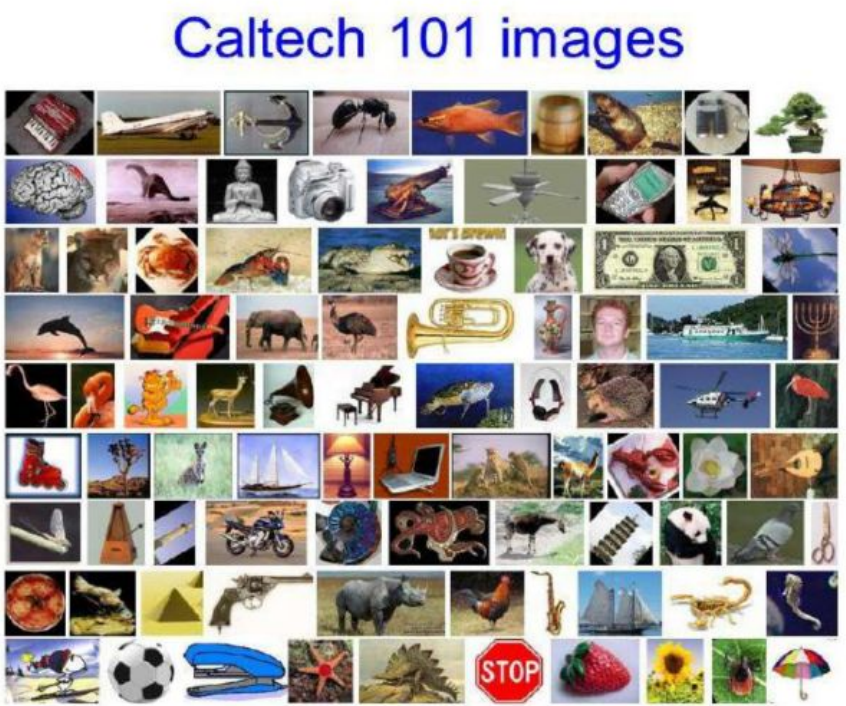
- 1960 - Early 1990's — The Geometric Era
- 1990's — Appearance based models
- Mid 90's — Sliding Window Based Approaches
- Late 1990's — Local features
- Early 2000's — parts and shape models
- **Mid 2000's — bags of features**
- Present trends — data-based methods, knowledge graphs, combination of local and global methods, contexts etc.

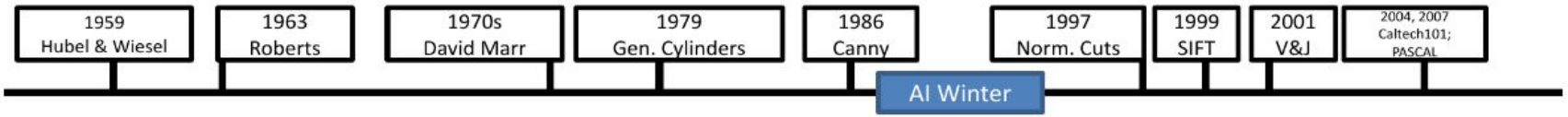
Bag of words



Svetlana Lazebnik

Era of Computer Vision - 2004 - 2007's - Challenges!





But…
there was a parallel timeline …

But…
there was a parallel timeline …

Some people were secretly working …
unnoticed until…

But…
there was a parallel timeline …

Some people were secretly working … unnoticed until…

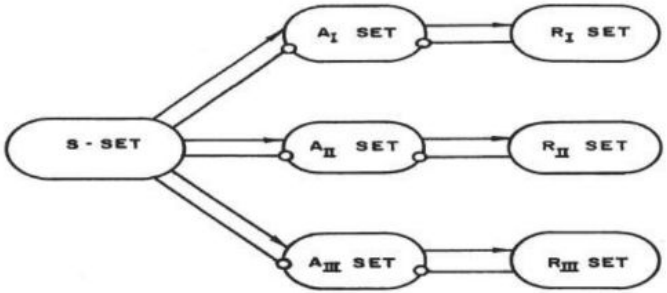
Let’s see what happened by going back in time…

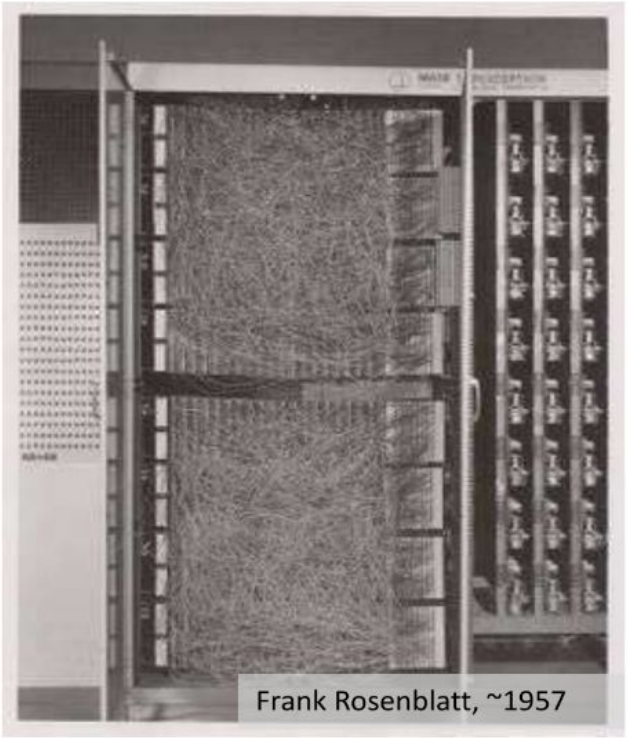
Era of Deep Learning - 1958 - Perceptron

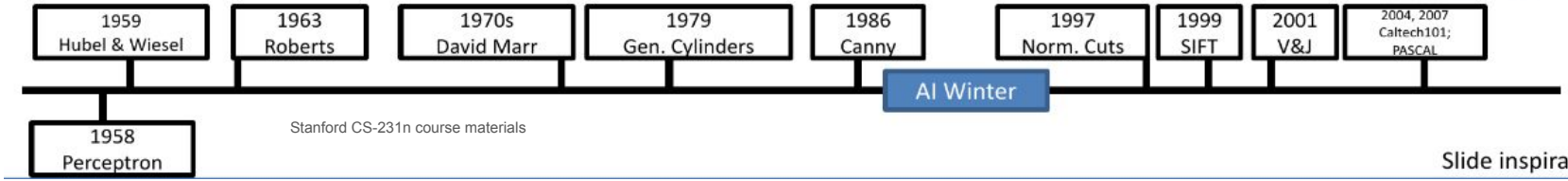
Learning representations by back-propagating errors

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




Frank Rosenblatt, ~1957



Stanford CS-231n course materials

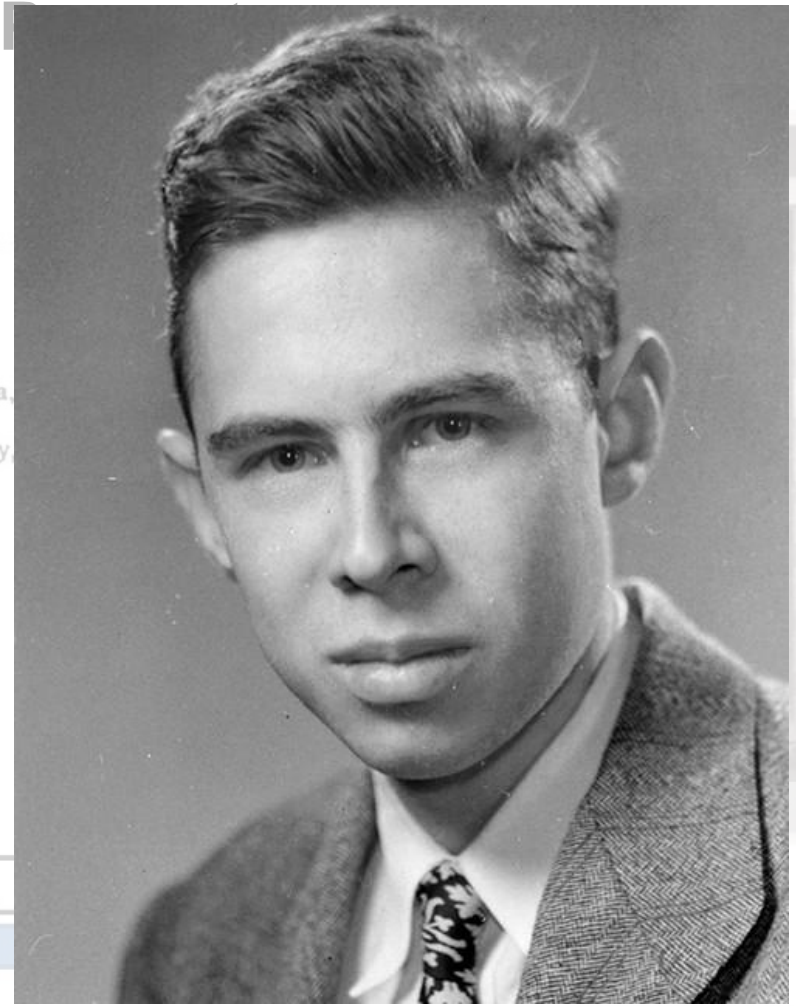
Slide inspiration: Justin Johnson

Era of Deep Learning - 1958 - P

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



Frank Rosenblatt
- Perceptron

Slide inspiration: Justin Johnson

Era of Deep Learning - 1986 - Backprop

Backprop: Rumelhart, Hinton, and Williams, 1986

Introduced backpropagation for computing gradients in neural networks

Successfully trained perceptrons with multiple layers

$$\frac{\partial E_p}{\partial w_{ji}} = \frac{\partial E_p}{\partial o_{pj}} \frac{\partial o_{pj}}{\partial w_{ji}}$$

recognizable math

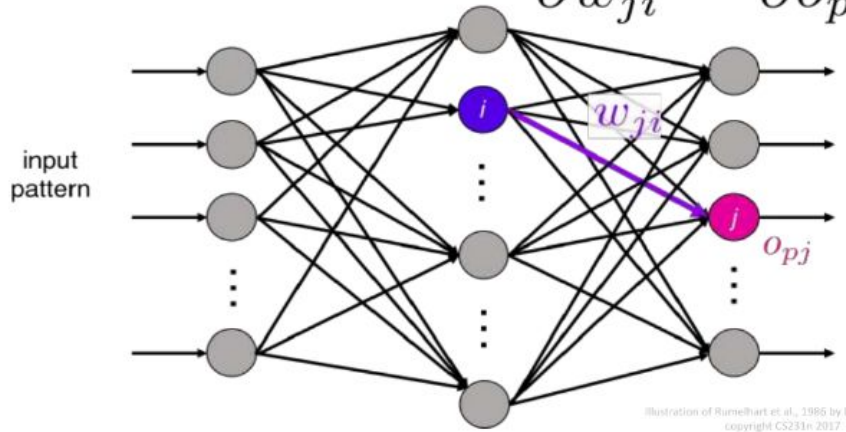
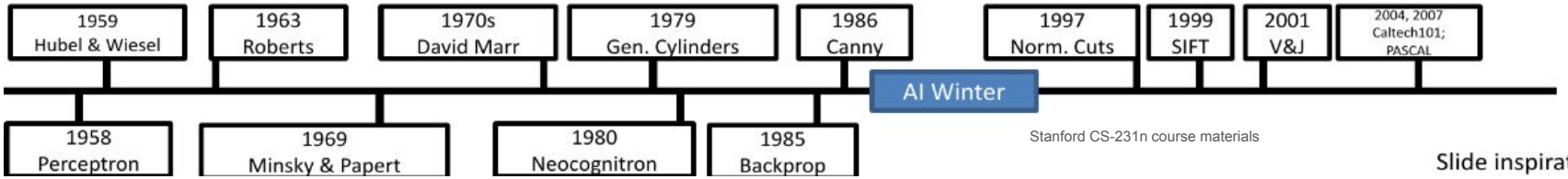


Illustration of Rumelhart et al., 1986 by Lane McIntosh, copyright CS231n 2017



Stanford CS-231n course materials

Slide inspiration: Justin John

Era of Deep Learning - 1958 - Perceptron

David Rumelhart

Hinton!

1959 Hubel & Wiesel

1963 Roberts

1997 Norm. Cuts

AI Winter

1958 Perceptron

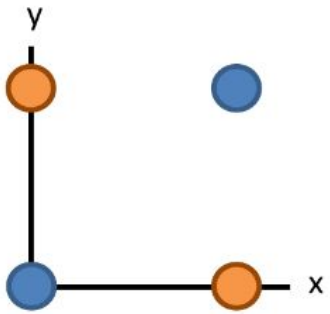
Stanford CS-231n course materials

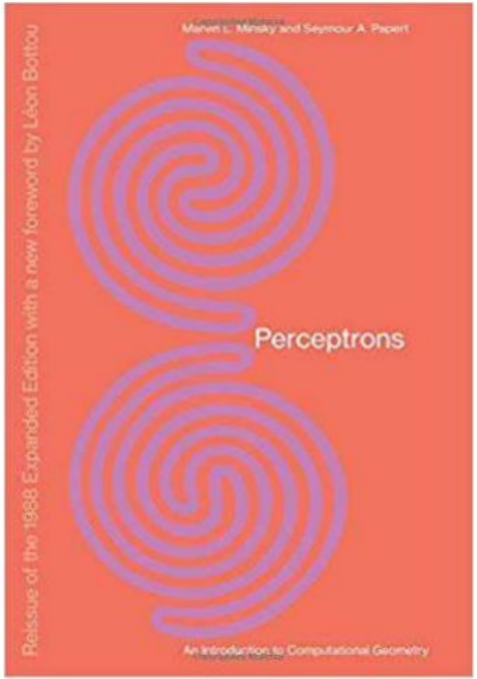
Slide inspiration: Justin Johnson

Era of Deep Learning - 1969 - Papert

Minsky and Papert, 1969

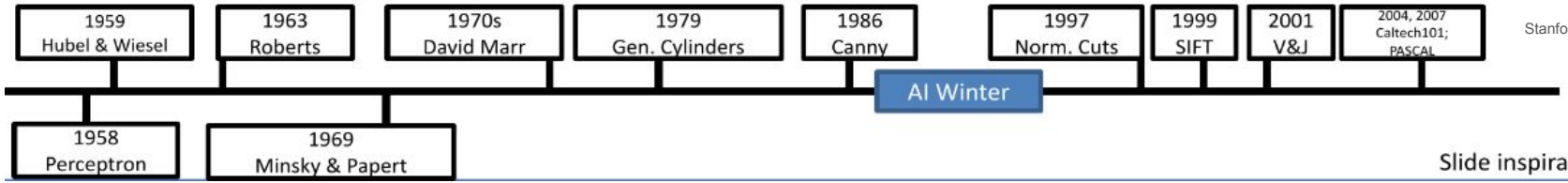
X	Y	F(x,y)
0	0	0
0	1	1
1	0	1
1	1	0





Showed that Perceptrons could not learn the XOR function

Caused a lot of disillusionment in the field

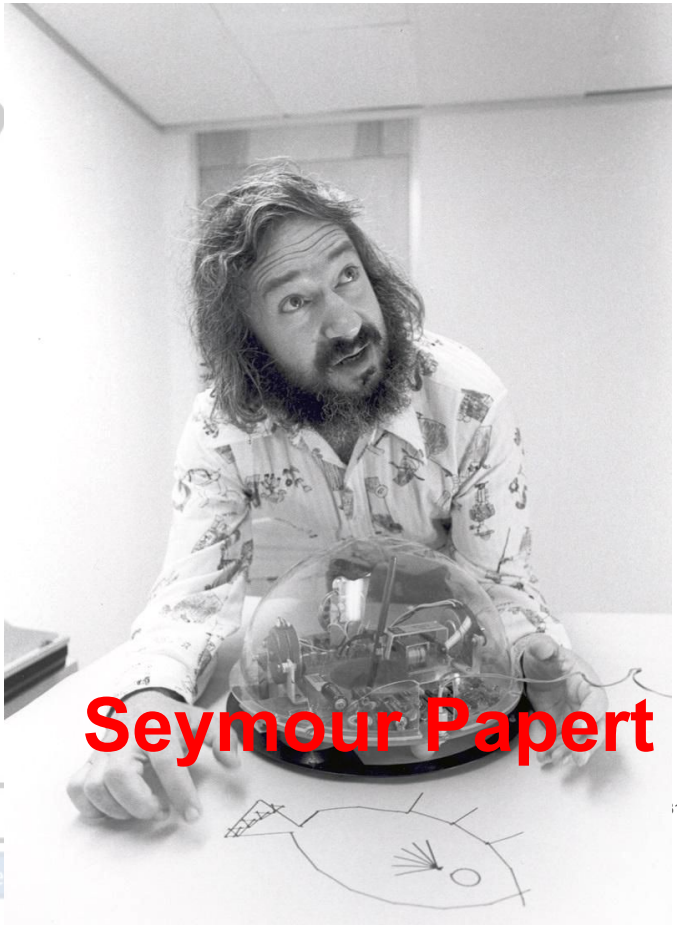


Stanford CS-231n course materials

Slide inspiration: Justin Johnson

Era of Deep Learning - 1969 - Papert

Marvin Minsky



Seymour Papert

1n course materials

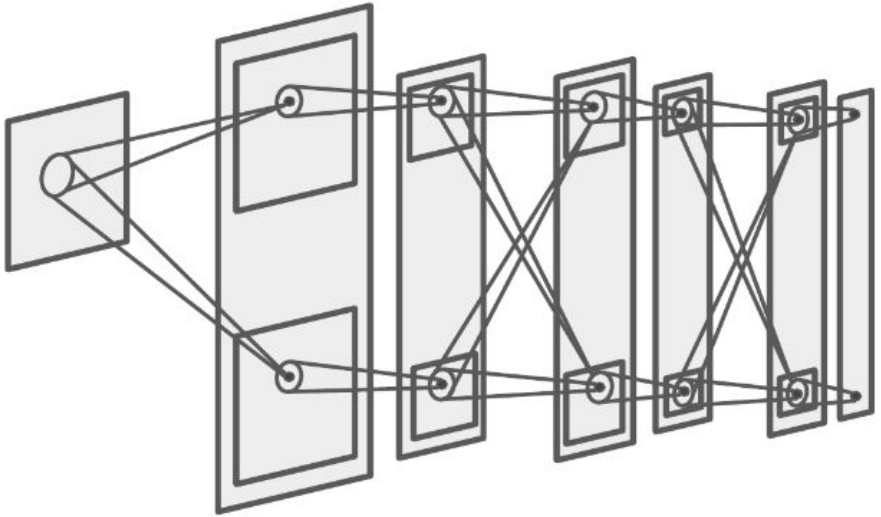
Slide inspiration: Justin Johnson

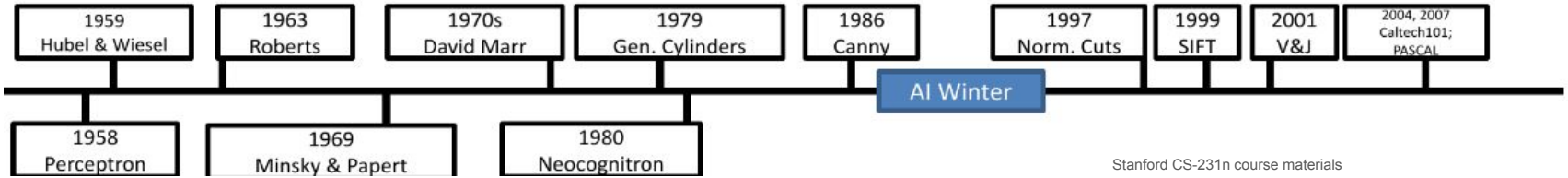
Era of Deep Learning - 1980 - Neocognitron - Fukushima

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





Stanford CS-231n course materials

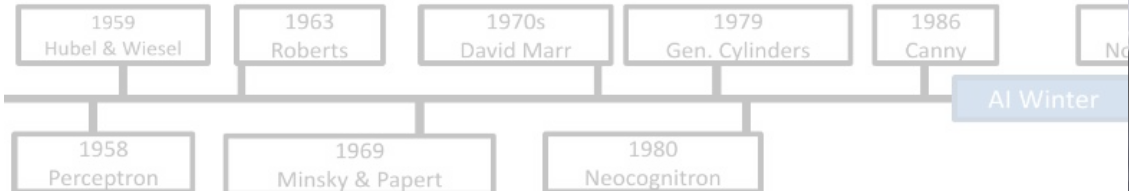
Era of Deep Learning - 1980 - Neocognitron - Fukushima

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)

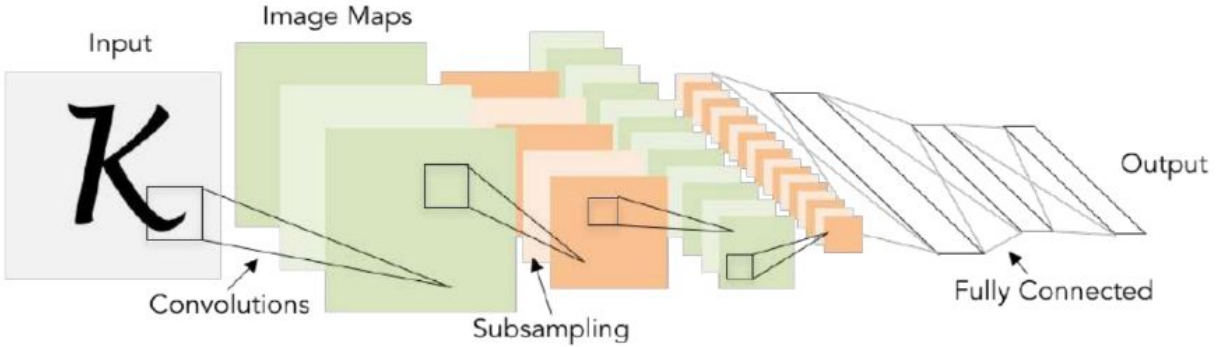
No practical training algorithm

Kunihiko Fukushima



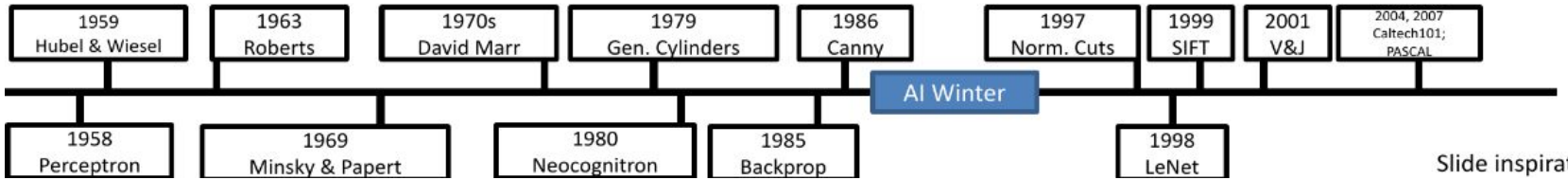
Era of Deep Learning - 1998 - CNN - LeNet

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!

Stanford CS-231n course materials



Slide inspiration: Justin Johnson

Era of Deep Learning - 1998 - CNN - LeNet

Convolutional Netw 998

Image Maps

Input

Convolutions

Sub

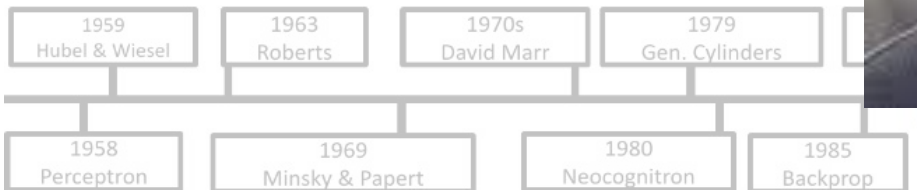
Applied backprop algorithm to a Ne

Learned to recognize handwritten

Very similar to our modern convolu

Yann LeCun

d CS-231n course materials



Slide inspiration: Justin Johnson

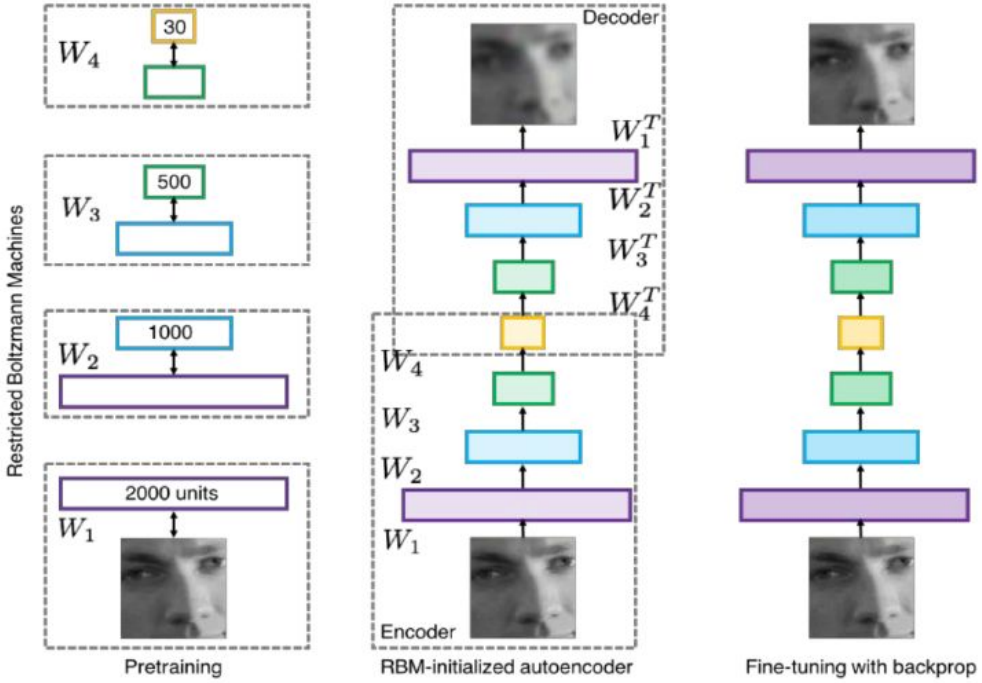
Era of Deep Learning - 2000s - Restricted Boltzmann Mac.

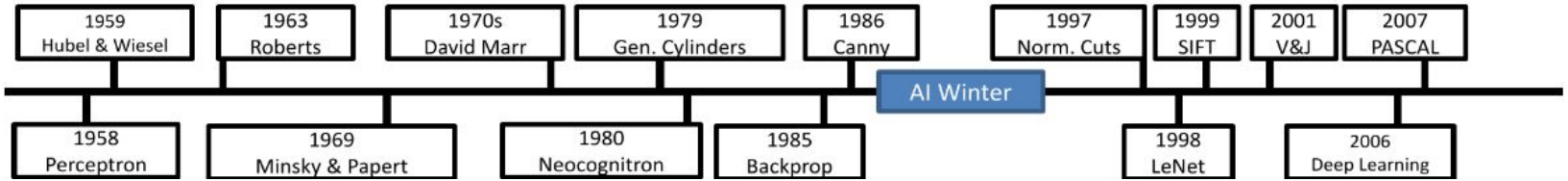
2000s: “Deep Learning”

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





Stanford CS-231n course materials

Slide inspiration: Justin Johnson

Era of Deep Learning - 2000s - Restricted Boltzmann Mac.

2000s: “Deep Learning”

People tried to train neural networks that were deeper and deeper

Not a mainstream research topic at this time

Hinton and Salakh
Bengio et al, 2007
Lee et al, 2009
Glorot and Bengio, 2010

Yoshua Bengio

Restricted Boltzmann Machines

ine-tuning with backprop

1959 Hubel & Wiesel
1963 Roberts
1970s David Marr
1979 Gen. Cylinders
1958 Perceptron
1969 Minsky & Papert
1980 Neocognitron
1985 Backprop
LeNet
Deep Learning

Stanford CS-231n course materials

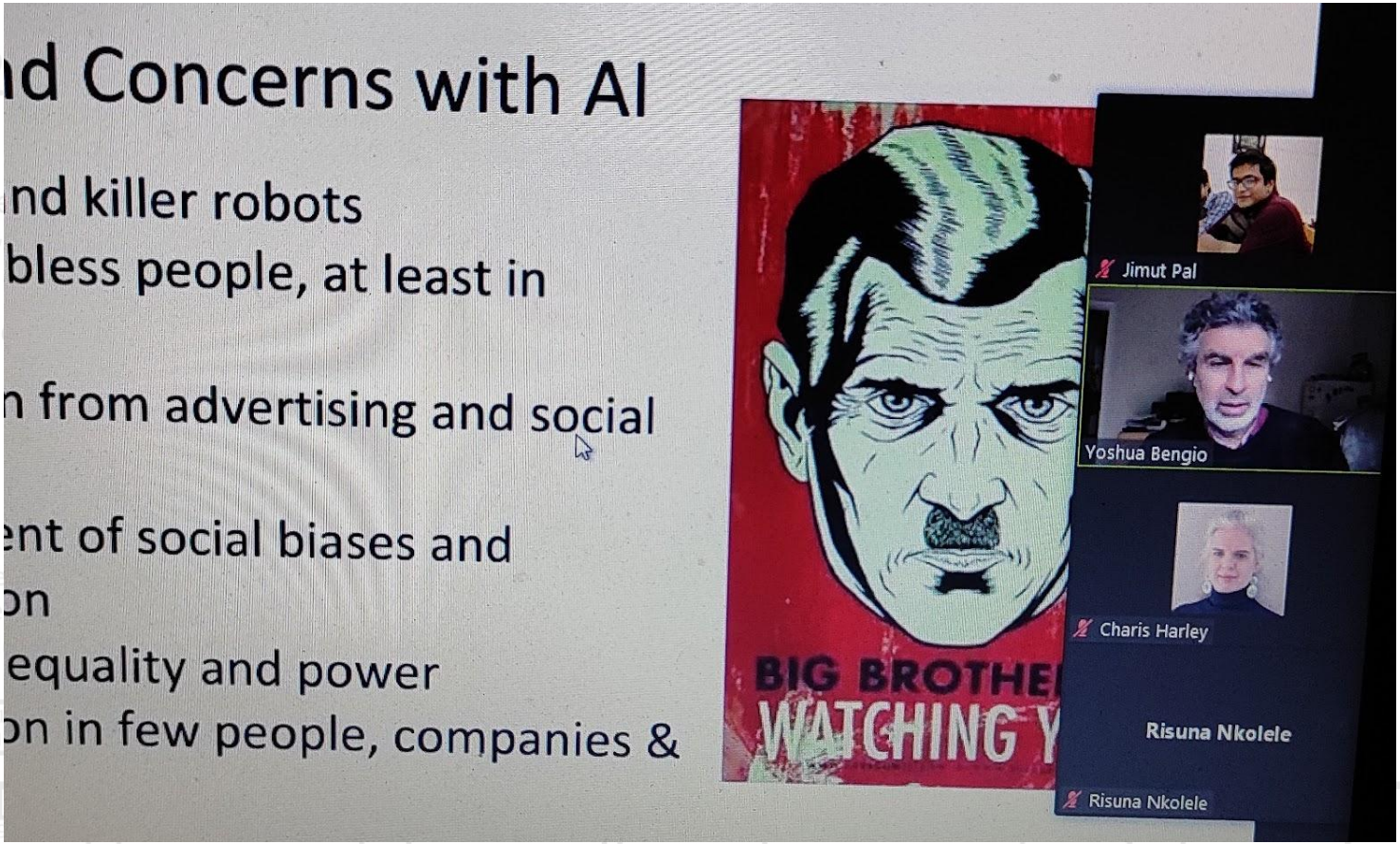
Slide inspiration: Justin Johnson

Deep Learning - 2018 - Turing Award

For conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing.

https://awards.acm.org/about/2018-turing

Era of Deep Learning - 2000s - Restricted Boltzmann Mac.



Slide inspiration: Justin Johnson

Era of Deep Learning - 2009s - ImageNet

IMAGENET Large Scale Visual Recognition Challenge

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

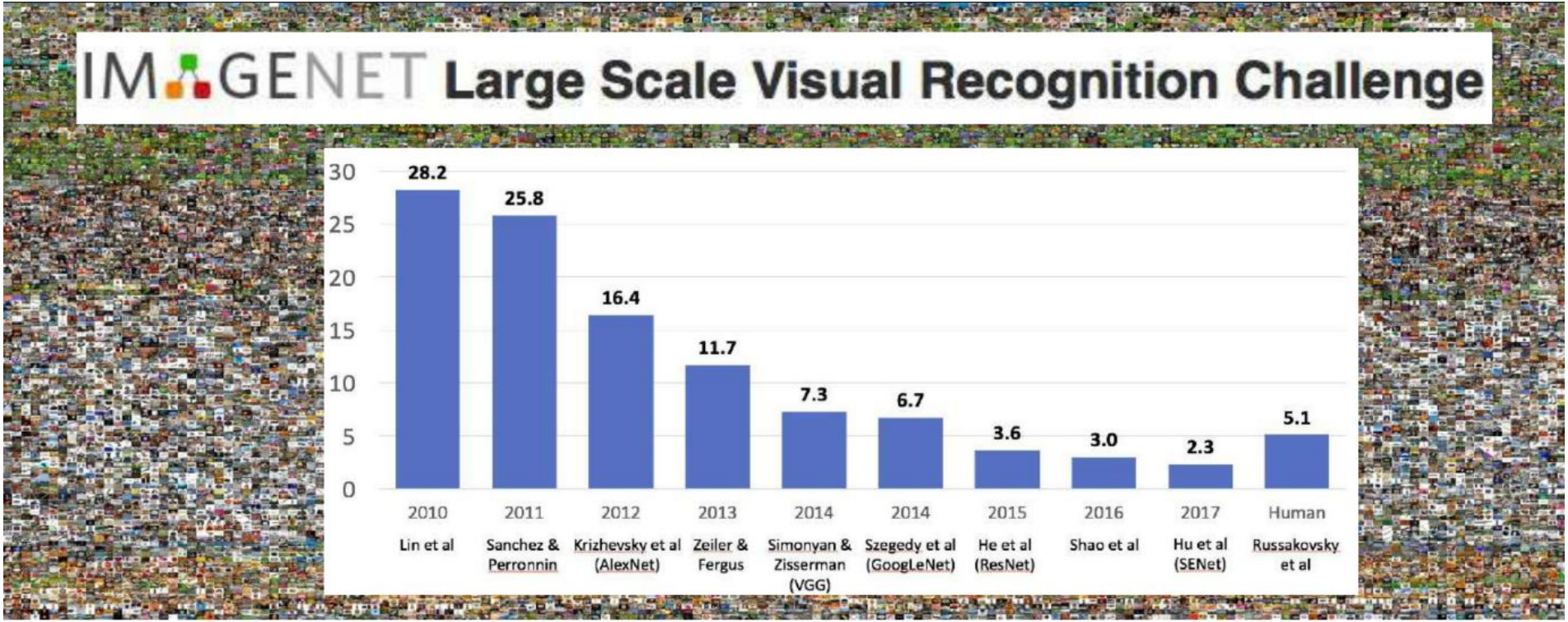
Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

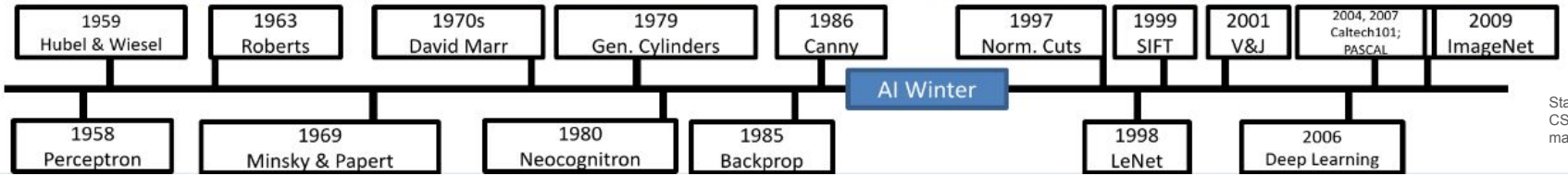
Deng et al, 2009
Russakovsky et al. IJCV 2015

1959 Hubel & Wiesel
1958 Perceptron
1963 Roberts
1969 Minsky & Papert
1970s David Marr
1979 Gen. Cylinders
1980 Neocognitron
1986 Canny
1985 Backprop
AI Winter
1997 Norm. Cuts
1998 LeNet
1999 SIFT
2001 V&J
2004, 2007 Caltech101; PASCAL
2006 Deep Learning
2009 ImageNet

Stanford CS-231n course materials

Era of Deep Learning - 2017s - ImageNet Challenge Cont…

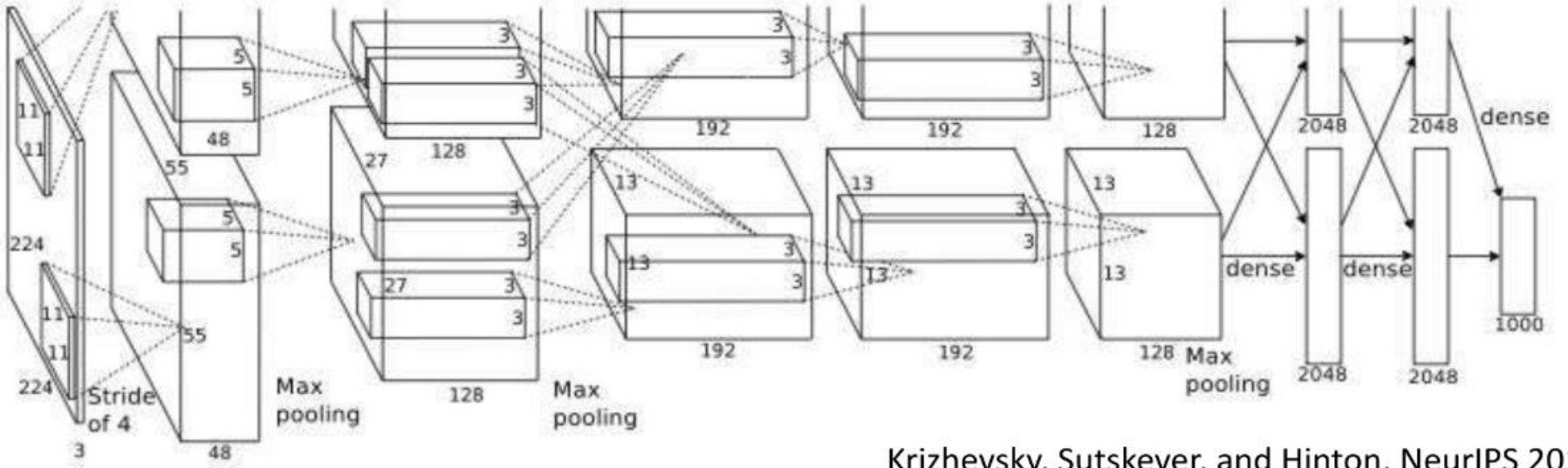




Stanford CS-231n course materials

Era of Deep Learning - 2012s - AlexNet

AlexNet: Deep Learning Goes Mainstream



Krizhevsky, Sutskever, and Hinton, NeurIPS 2012

AlexNet: Deep Learning Goes Mainstream

Alex Krizvesky

Notable work: **Imagenet classification with deep convolutional neural networks**

Total Citations: **212087**

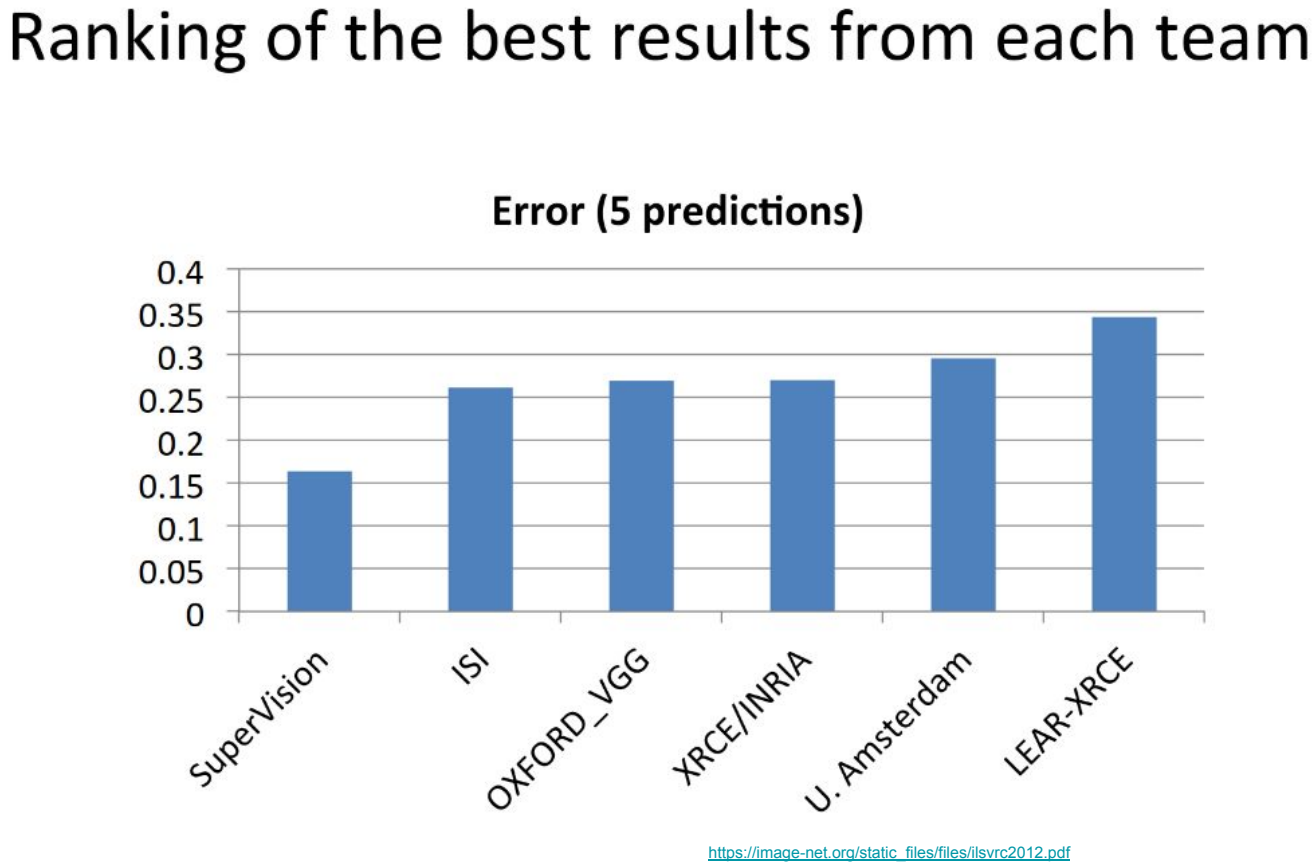
Alma Mater: **University of Toronto**

Current: Didn't complete his PhD (UltraChad) Retired

Krizhevsky, Sutskever, and Hinton, NeurIPS 2012

Era of Deep Learning - 2012s - AlexNet

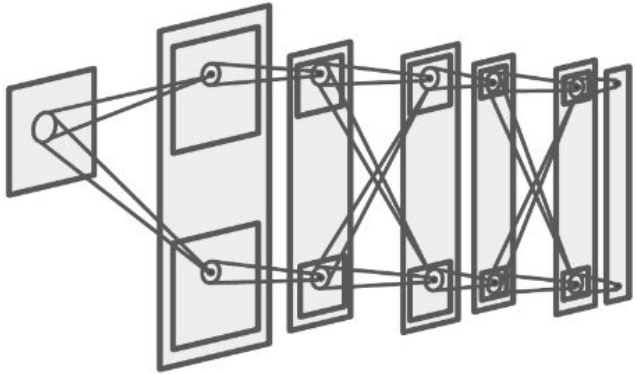
- SuperVision - Alex Kriz's team
- Note: Not our ISI, Kolkata
- Other methods used SIFT based classical techniques.
- First Deep Learning based technique trained on GPUs.

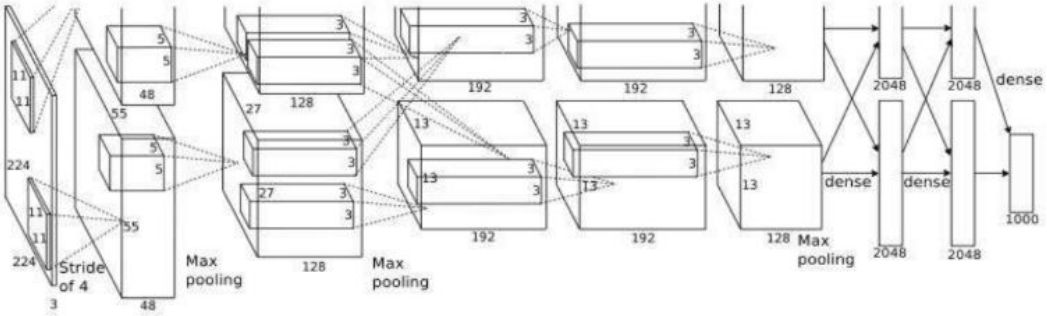


https://image-net.org/static_files/files/ilsvrc2012.pdf

Era of Deep Learning - 2012s - AlexNet

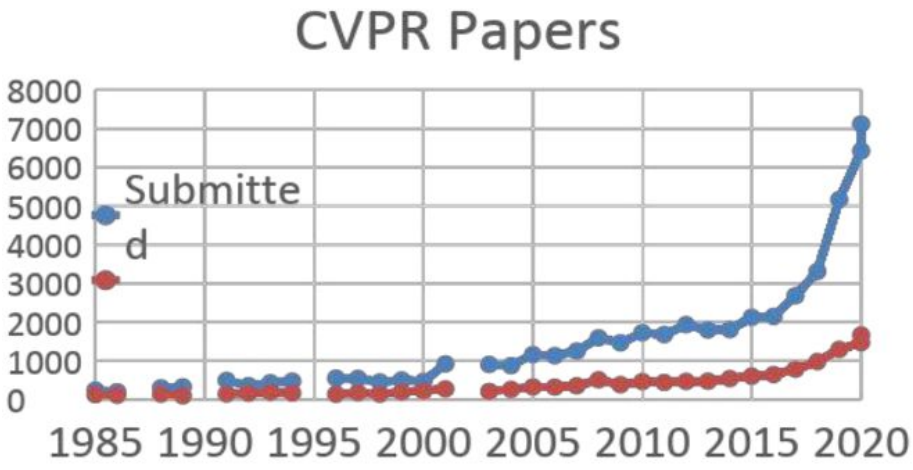
AlexNet vs. Neocognitron: 32 years apart



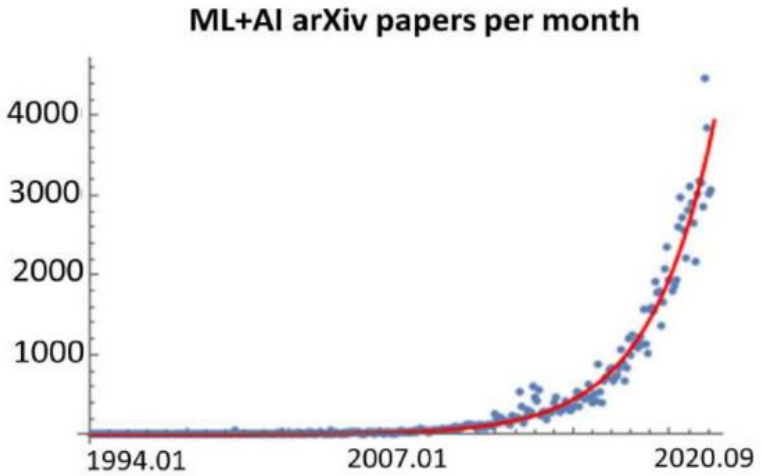


Era of Deep Learning - 2012 and beyond!

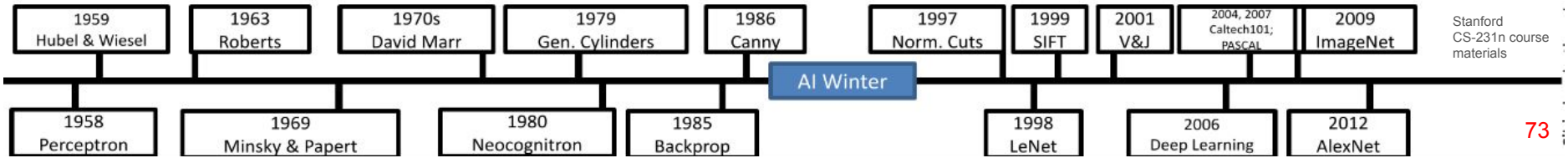
2012 to Present: Deep Learning Explosion



Publications at top Computer Vision conference



arXiv papers per month (source)



Stanford CS-231n course materials

Era of Deep Learning - 2012 and beyond!

2012 to Present: Deep Learning is Everywhere

Year 2010

NEC-UIUC

Dense descriptor grid: HOG, LBP

Coding: local coordinate, super-vector

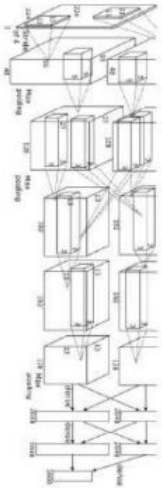
Pooling, SPM

Linear SVM

[Lin CVPR 2011]

Year 2012

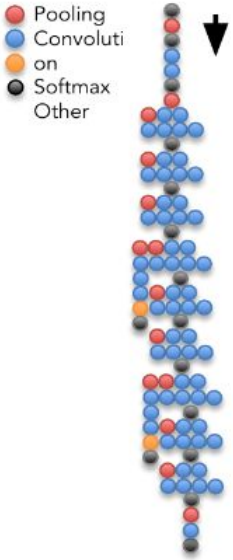
SuperVision



[Krizhevsky NIPS 2012]

Year 2014

GoogLeNet



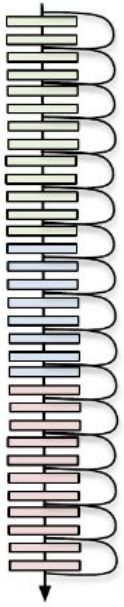
[Szegedy arxiv 2014]

VGG

[Simonyan arxiv 2014]

Year 2015

MSRA



[He ICCV 2015]

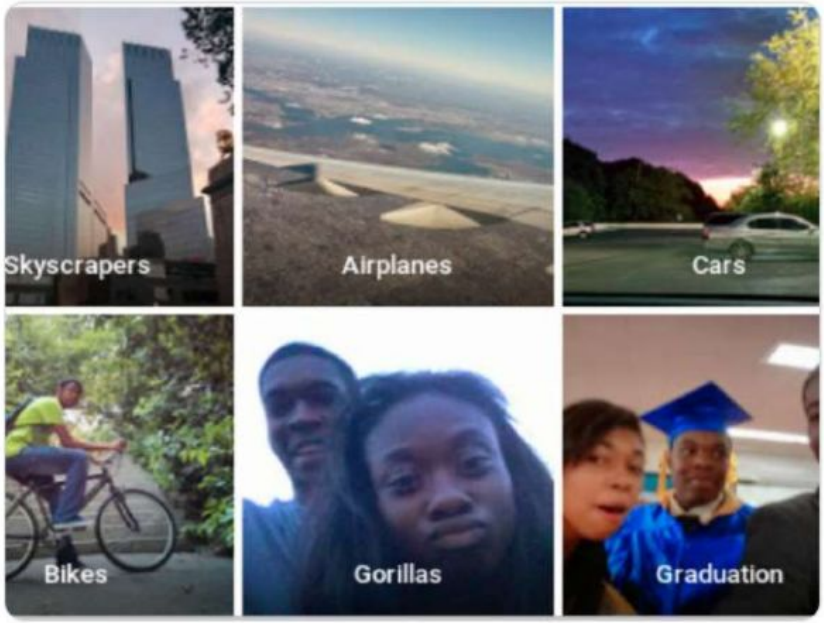
Stanford CS-231n course materials

Vision Has A LONG Way to go…

Computer Vision Can Cause Harm

Stanford CS-231n course materials

Harmful Stereotypes



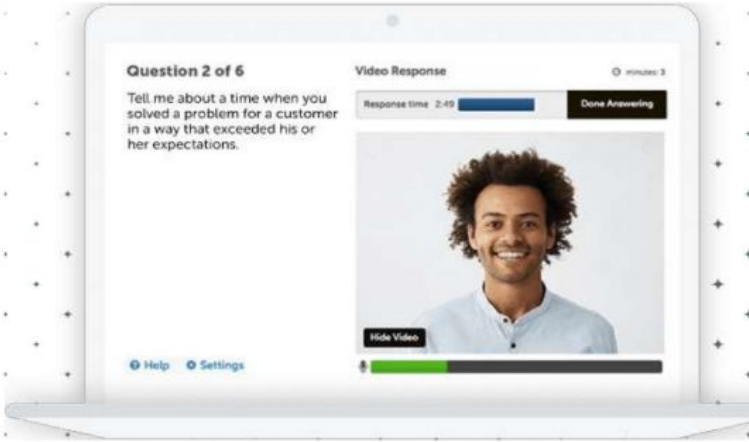
Barocas et al, "The Problem With Bias: Allocative Versus Representational Harms in Machine Learning", SIGCIS 2017
Kate Crawford, "The Trouble with Bias", NeurIPS 2017 Keynote
Source: https://twitter.com/jackyalcine/status/615329515909156865 (2015)

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.'



Source: https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-jo
https://www.hirevue.com/platform/online-video-interviewing-softw
Example Credit: Timnit Ge

Vision Has A LONG Way to go…

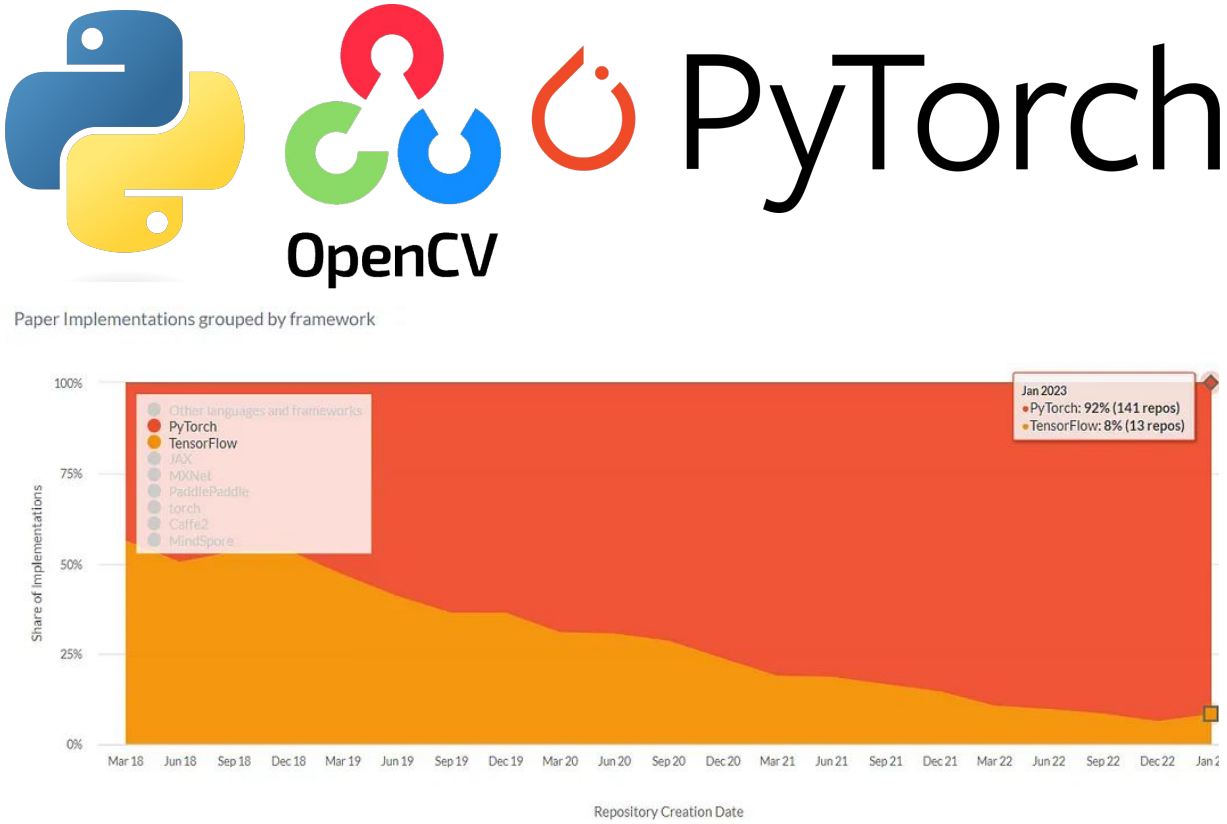
And there is a lot we don't know how to do

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

Stanford CS-231n course materials

Grading and Course Structure

- Grading (Tentative)
  - **Assignments** (about 1-2): 30%
  - **Recitation**: 20%
  - Mid-semester examination: 20%
  - End-semester examination: 30%
- Course Webpage: https://jimut123.github.io/courses/vision/offering1_sem2_2024.html



https://medium.com/@valkont/battle-of-the-giants-tensorflow-vs-pytorch-2023-fd8274210a38
https://upload.wikimedia.org/wikipedia/commons/thumb/c/c3/Python-logo-notext.svg/1869px-Python-logo-notext.svg.png
https://upload.wikimedia.org/wikipedia/commons/thumb/c/c6/PyTorch_logo_black.svg/2560px-PyTorch_logo_black.svg.png
https://opencv1.b-cdn.net/wp-content/uploads/2020/07/OpenCV_logo_black-2.png

More on course structure

- **Tutorials will be conducted in Pytorch.**
- **Show recitation guidelines.**
- Exams - **A necessary evil**
  - Midsems - Mostly 20-30% subjective/MCQs, and rest coding (Tentative).
  - Endsems - follows a similar pattern.
- (Give suggestion on this!)
- **Relative grading will be used (A - 10, B - 8, C - 6)**
- We will use Google Chat and emails for discussions - related to recitation, assignment submission, sharing of materials etc (Pointers & Resources will be provided)
- Each student should adhere to all principles of academic honesty during homeworks, exams and projects!

TAKE ATTENDANCE HERE

WITH ROLL NUMBERS

More on course structure

- **Regarding plagiarism - (only instructors are allowed to plagiarise course content/slides) STUDENTS must AVOID fabrication of results/ copying at all cost.**
- **Report to HOD: if found cheating!**
- **Assignments -**
  - **Evaluation criteria: Discuss the assignment one to one with the instructor once done.**
  - **Coding assignments will be given.**
  - **Might give independent Kaggle Challenge assignment related to Computer Vision - 1 week of time investment necessary.**
  - **Present/ Describe the solution one to one via a VIVA voce.**
- **Discuss about time - Mid Sem, End Sem, Dates and plans etc. (5-6 months course).**

Necessary tools…

- Install Python3
- Install OpenCV
- Use Chat-GPT
- Use Github Co-Pilot – Student Developer (Use Insti ID)
- Use GPU servers with permission, but mostly, things will be done using Google Colab.
- Use Google Colaboratory (but using PC is recommended!)
- USE LINUX if possible… installation of all the things are easy

Why Pytorch - OpenAI

OpenAI Research API ChatGPT Safety Company Search Log in Try ChatGPT

As technical context: at the heart of our infrastructure is a large-scale deployment of GPU nodes running in dozens of Kubernetes clusters across regions. Some core technologies we build with include Python, PyTorch, CUDA, Triton, Redis, Infiniband, NCCL, NVLink

This role is exclusively based in our San Francisco HQ. We offer relocation assistance to new employees.

Compensation, Benefits and Perks

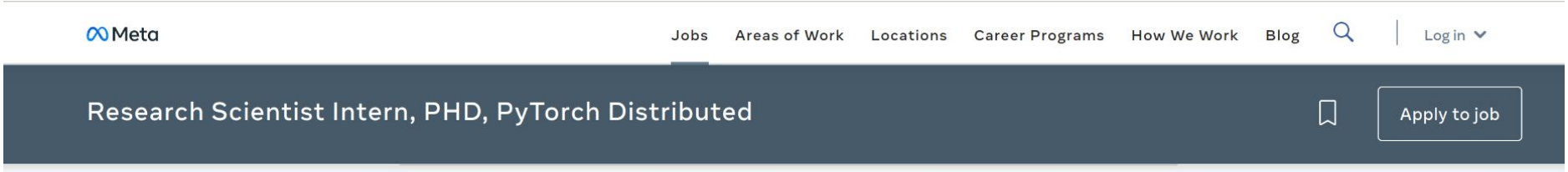
Total compensation also includes generous equity and benefits.

- Medical, dental, and vision insurance for you and your family
- Mental health and wellness support
- 401(k) plan with 4% matching
- Unlimited time off and 18+ company holidays per year
- Paid parental leave (20 weeks) and family-planning support
- Annual learning & development stipend ($1,500 per year)

Annual Salary Range

$350,000—$500,000 USD

Why Pytorch - Meta aka Facebook



Research Scientist Intern, PHD, PyTorch Distributed Responsibilities

- Apply relevant AI and machine learning techniques to advance the state-of-the-art in machine learning frameworks.
- Collaborate with users of PyTorch to enable new use cases for the framework both inside and outside Meta.
- Develop novel, accurate AI algorithms and advanced systems for large scale distributed training and inference.

Minimum Qualifications

Prior experience with Meta can be considered to supplement an applicant's prior years of experience or types of prior experience to meet the minimum qualifications of the position.

- Currently has, or is in the process of obtaining, PhD degree in the field of Computer Science or a related STEM field

Why Pytorch - Microsoft

Job you selected

Research Scientist

19 days ago Redmond, Washington, United States

Up to 100% work from home

We are seeking a highly skilled and passionate Research Scientist to join our Responsible & OpenAI Research (ROAR) team in Azure Cognitive Services. As a Research Scientist, you will play a key role in advancing the field of Responsible Artificial Intelligence...

See details

Similar jobs

Jobs aligned to the same profession, discipline, and location as the one you selected.

Senior Research Scientist

1 month ago Redmond, Washington, United States

We are seeking a highly skilled and passionate Research Scientist to join our Responsible & OpenAI Research (ROAR) team in Azure Cognitive Services.

As a Research Scientist, you will play a key role in advancing the field of Responsible Artificial Intelligence (AI) to ensure safe releases of the rapidly advancing AI technologies, such as GPT-4, GPT-4V, DALL-E 3 and beyond, as well as to expand and enhance our standalone Azure AI Content Safety Service.

Microsoft's mission is to empower every person and every organization on the planet to achieve more. As employees we come together with a growth mindset, innovate to empower others, and collaborate to realize our shared goals. Each day we build on our values of respect, integrity, and accountability to create a culture of inclusion where everyone can thrive at work and beyond.

In alignment with our Microsoft values, we are committed to cultivating an inclusive work environment for all employees to positively impact our culture every day.

Responsibilities

- Conduct cutting-edge research to develop Responsible AI definitions, methodologies, algorithms, and models for both measurement and mitigation of AI risks.
- Stay abreast of the latest advancements in the field and contribute to the scientific community through publications at top venues.
- Contribute to the development of Responsible AI policies, guidelines, and best practices and ensure the practical implementation of these guidelines within various AI technology stacks across Microsoft, promoting a consistent approach to Responsible AI.
- Enable the safe release of new Azure OpenAI Service features, expand and enhance the Azure AI Content Safety Service with new detection technologies.
- Develop innovative approaches to address AI safety challenges for diverse customer scenarios.
- Other: Embody our Culture and Values

Required/Minimum Qualifications:

- Master's Degree in relevant field AND 1+ year(s) related research experience
  - OR equivalent experience.
- Research/Applied Science experience in Natural Language Processing, Computer Vision, Multimodal analysis, Deep Learning, Machine Learning, or a related field.
- Experience with algorithmic problem solving and software development skills (Python/Java/C++).
- Experience with open-source tools such as PyTorch, as well as experience with statistical analysis and data visualization tools.

Other Requirements:
Ability to meet Microsoft, customer and/or government security screening requirements are required for this role. These requirements include but are not limited to the following specialized security screenings:

Why Pytorch - Apple

has made contributions to the field by building new models that can infer physiological, health or comfort related signals from videos. The ideal candidate has a passion to create transformative bio-feedback experiences that are inspired by cognitive and systems neuroscience. We are looking for a teammate with excellent interpersonal skills, and a dedication to conducting impactful research in a fast-paced environment.

Key Qualifications

- Hands-on experience with applying both traditional CV models (for ex: optical flow, structure from motion, state estimation, feature descriptors, Eulerian magnification) and deep learning models (for ex: body/face detection, segmentation, stereo depth, saliency, recurrent and convolutional nets) on physiological datasets.
- Fluency in a high-level programming language (Python and/or Matlab) and experience with ML tools (scikit-learn, pandas and deep learning toolboxes such as TensorFlow and PyTorch).
- Experience with OpenCV or similar.
- Experience in signal processing, linear algebra and time-frequency analysis.
- Excellent programming, problem solving and debugging skills.
- Interest in applying machine learning techniques to help improve user's wellbeing, health and comfort.
- (Preferred) Experience with human visual system and perception.
- (Preferred) Experience with translating exploratory data analyses into production models.

Why Pytorch - Qualcomm

- Deep generative models
- Representation learning
- Adversarial attacks, examples

Minimum Qualifications
- Python Programming
- Deep learning with hands-on experience of deep learning tools, e.g. PyTorch
- Basic knowledge and development experience on machine learning and signal processing

Applicants: If you need an accommodation, during the application/hiring process, you may request an accommodation by sending email to accommodationsupport

Qualcomm expects its employees to abide by all applicable policies and procedures, including but not limited to security and other requirements regarding protection of Company confidential information and other confidential and/or proprietary information, to the extent those requirements are permissible under applicable law.

To all Staffing and Recruiting Agencies: Our Careers Site is only for individuals seeking a job at Qualcomm. Staffing and recruiting agencies and individuals being represented by an agency are not authorized to use this site or to submit profiles, applications or resumes, and any such submissions will be considered unsolicited. Qualcomm does not accept unsolicited resumes or applications from agencies. Please do not forward resumes to our jobs alias, Qualcomm employees or any other company location. Qualcomm is not responsible for any fees related to unsolicited resumes/applications.

If you would like more information about this role, please contact Qualcomm Careers.

Job Application Privacy Notice

Job Application Privacy Notice

Use of AI in the Application Process

Use of AI in the Application Process

Syllabus! - Would be enough if we can cover 50-60%

- Overview of the course
- Lecture 1: Python3 Programming Refresher.
- Lecture 2: Installation, GUI features, Basic Operations, Colorspaces, Geometric Transformations, Image Thresholding, Image Smoothing.
- Lecture 3: Morphological Operations, Image Gradients, Canny Edge Detection, Image Pyramids.
- Lecture 4: Conoutours, Histogram, Fourier Transform.
- Lecture 5: Template matching, Hough transform - Line, Hough transform - Circle, Watershed, Grab-Cut.
- Lecture 6: Feature detection, Harris Corner, Shi-Tomasi Detector, SIFT, SURF.
- Lecture 7: FAST, BRIEF, Feature Matching, Homography.

Syllabus!

- Lecture 8: Video (Optical Flow, Background subtraction methods, Meanshift and Camshift).
- Lecture 9: HaarCascade, Camera Calibration, Pose Estimation, Epipolar Geometry, Depth Map from stereo image.
- Lecture 10: K-Means Clustering, K-Nearest Neighbor, Image Denoising, Image Inpainting, High Dynamic Range (HDR).
- Lecture 11: Introduction to Pytorch and Tensors, Feed Forward Neural Networks, Back-propagation, Gradient descent, Classifier example, Activation functions.
- Lecture 12: Optimizers, Convolutional Neural Networks, Convolution operation, Maxpool.
- Lecture 13: Basic Methodology - Train - Validation - Test Pipeline, Design a Classifier using Pytorch, Autoencoders - theory and example.

Syllabus!

- Lecture 14: Segmentation - Single class and Multi-Class, Theory: U-Net, Attention U-Net, ResU-Net, DeepLabV3+.
- Lecture 15: Application - Design a Segmentation Network using Pytorch.
- Lecture 16: VAEs - Theory.
- Lecture 17: Application - Design a VAE using Pytorch.
- Lecture 18: GANs - Theory.
- Lecture 19: Application - Design a GAN using Pytorch.
- Lecture 20: (Tentative) Guest Lecture: Transformer & Vision Transformer Theory and Tutorial, Seshadri Mazumder, CVIT, IIIT Hyderabad.
- Lecture 21: (Tentative) Guest Lecture: Stable Diffusion Theory and Tutorial, Seshadri Mazumder, CVIT, IIIT Hyderabad.

Flexibility in adding/ removing stuffs from syllabus …

- How much python do you guys know (?)
- Perceptron training (?) - Coding?
- Neural Network from scratch (?) - Coding?
- Simulated Annealing (Theory)
- Fuzzy-C-Means Clustering (Theory)
- EM Optimization (Theory)
- Transformers - coding (?)
- Recurrent Neural Network - coding (?)
- Did you guys do NLP based course ?

Flexibility in adding/ removing stuffs from syllabus …

- A big assignment (with deadline of 2 weeks?), some serious extension to something that I will teach, but can be doable.
- Kind of a research based assignment.
- Example
  - Say I teach you how to segment stuffs in 2D, you will segment stuffs in 3D. Will help you in joining the current 3D segmentation competitions.
  - Only dataset will be provided - need to do EDA etc.
  - No skeleton code for 3D will be provided, however you may extend 2D code to 3D
- Need to submit code, demo, presentation (15 minutes) - Pytorch.
- Or two small assignments, whose worth of work is same as the big one (?)

On a lighter note …

https://i.kym-cdn.com/photos/images/newsfeed/001/244/486/559.png