Deep Learning for Computer Vision

Dr. Konda Reddy Mopuri
Mehta Family School of Data Science and Artificial Intelligence
IIT Guwahati
Aug-Dec 2022

Outline

- Computer Vision: What and Why?

Outline

- Computer Vision: What and Why?
- Brief history of Computer Vision

Outline

- Computer Vision: What and Why?
- Brief history of Computer Vision
- This Course: structure, organization

Outline

- Computer Vision: What and Why?
- Brief history of Computer Vision
- This Course: structure, organization
- Logistics and Resources

What is Computer Vision?

What is Computer Vision?

- Field of AI that enables machines to
  - Extract meaningful information from the visual world via digital images and videos

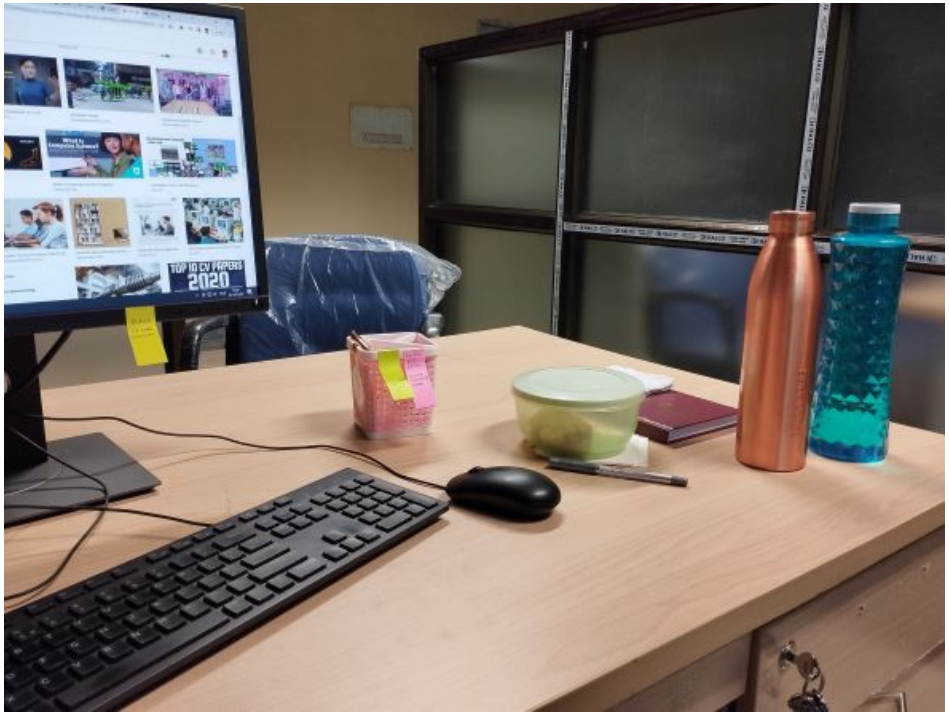
What is Computer Vision?

- Field of AI that enables machines to
  - Extract meaningful information from the visual world via digital images and videos
  - And, recommend appropriate actions based on that

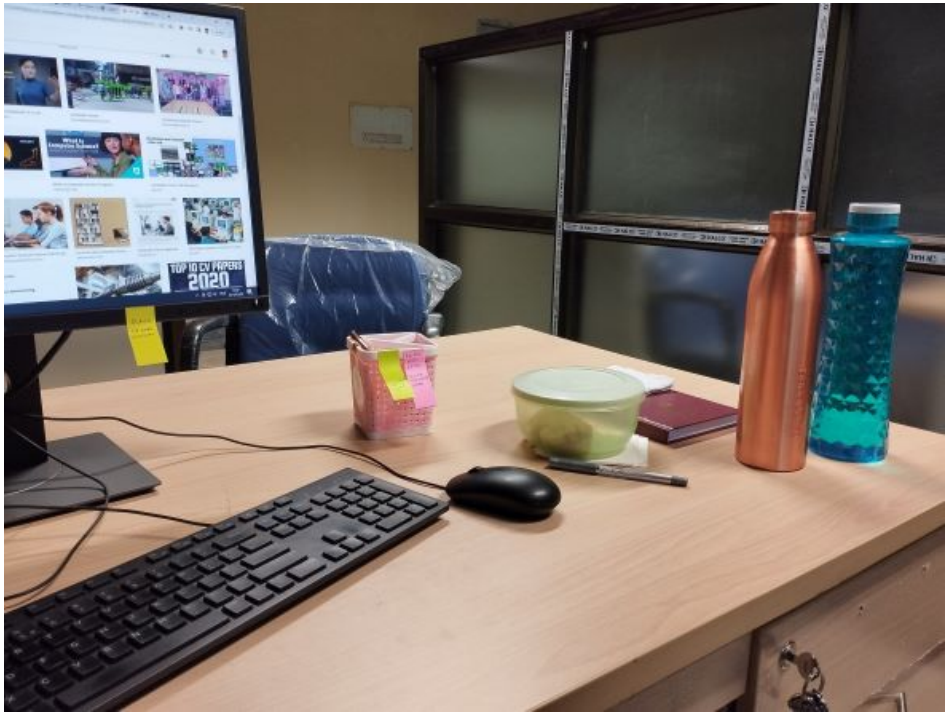
What is Computer Vision?

- Field of AI that enables machines to
  - Extract meaningful information from the visual world via digital images and videos
  - And, recommend appropriate actions based on that
- Simply, enabling machines to see as humans do!

What is Computer Vision?

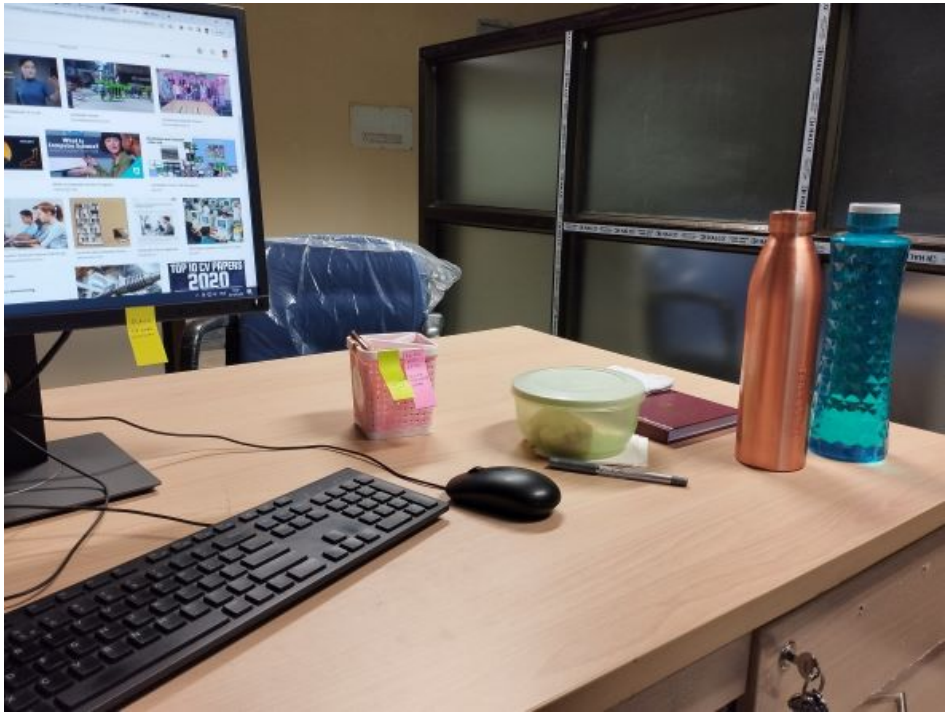


What is Computer Vision?



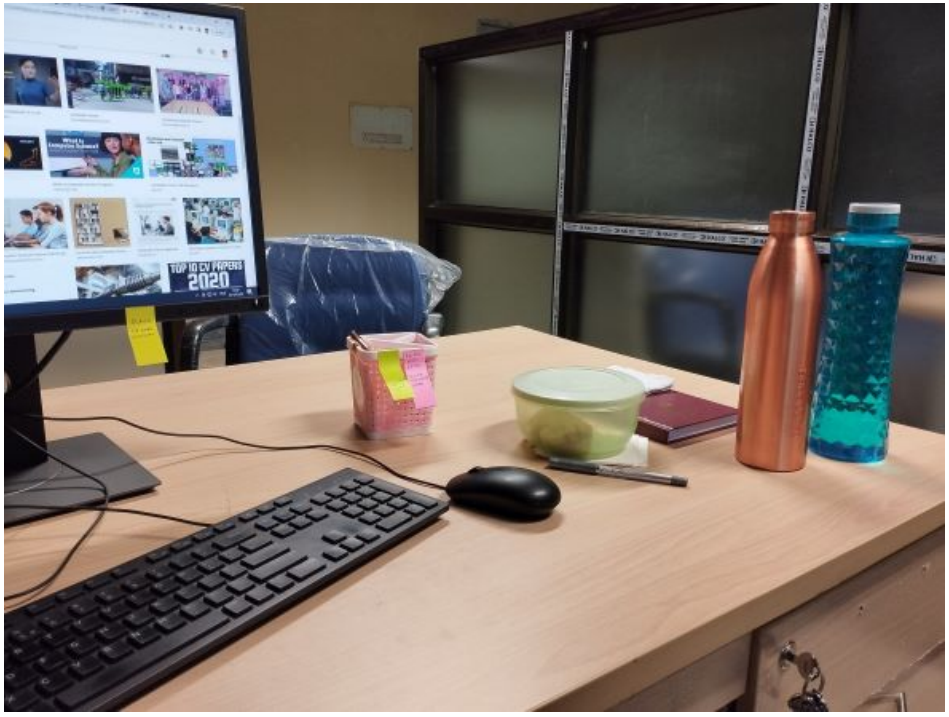
- How many sticky notes are there?

What is Computer Vision?



- How many sticky notes are there?
- What is the object that is new in the scene?

What is Computer Vision?



- How many sticky notes are there?
- What is the object that is new in the scene?
- Is there something to eat/drink here?

What is Computer Vision?

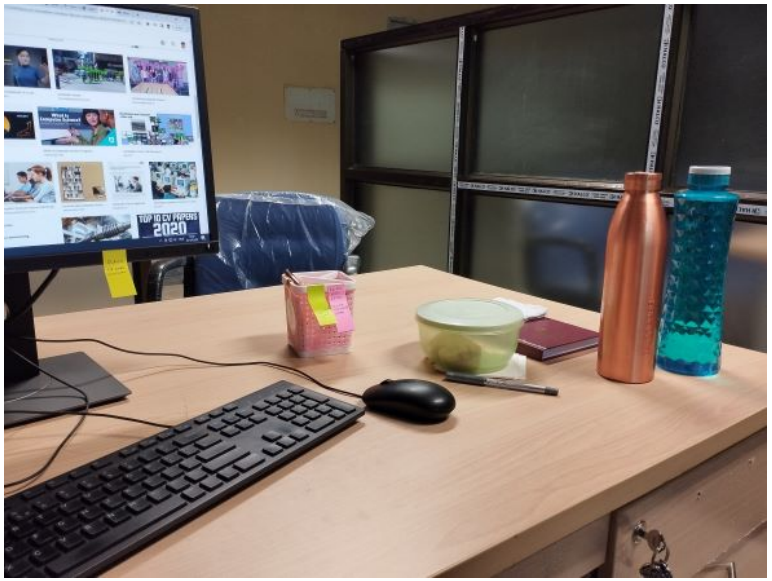
Images from the 'Objects out of the context' dataset

What is Computer Vision?

Images from the 'Objects out of the context' dataset

- What is wrong with each of these images?

What is Computer Vision?



- How many sticky notes are there?
- What is the object that is new in the scene?
- What is wrong with each of these images?

Computer Vision

Can we make machines answer these questions?

What is Computer Vision?

More formally

Building artificial systems that can process, perceive, and reason about the visual world (Taken from Justin Johnson, U.Mich.)

What is Computer Vision?

More formally

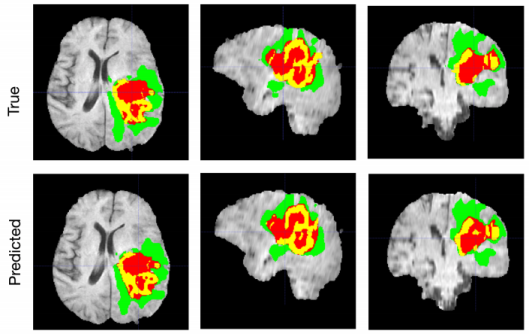
Building artificial systems that can process, perceive, and reason about the visual world (Taken from Justin Johnson, U.Mich.)

Other definitions

- “construction of explicit, meaningful descriptions of physical objects form images" (Ballard & Brown, 1982)
- “computing properties of the 3D world from one or more images" (Trucco & Verri, 1998)
- “to make useful decisions about real physical objects and scenes based on sensed images" (Sockman & Shapiro, 2001)

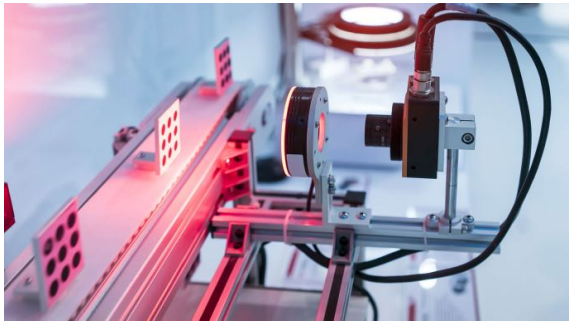
Why CV? Application Areas

Autonomy
(Credits: Getty Images)



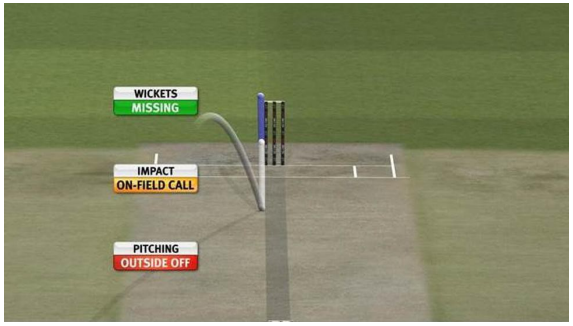
Healthcare
(Credits:Nvidia.Developer)

Surveillance
(Credits:Flickr)



Manufacturing
(Credits: Moonvision)

HCI
(Credits:X-tech.am)



Sports
(Credits:Medium and Sasank Gurajapu)

Why is it hard?

Why is it hard?

- Partly because it is an inverse problem

Why is it hard?

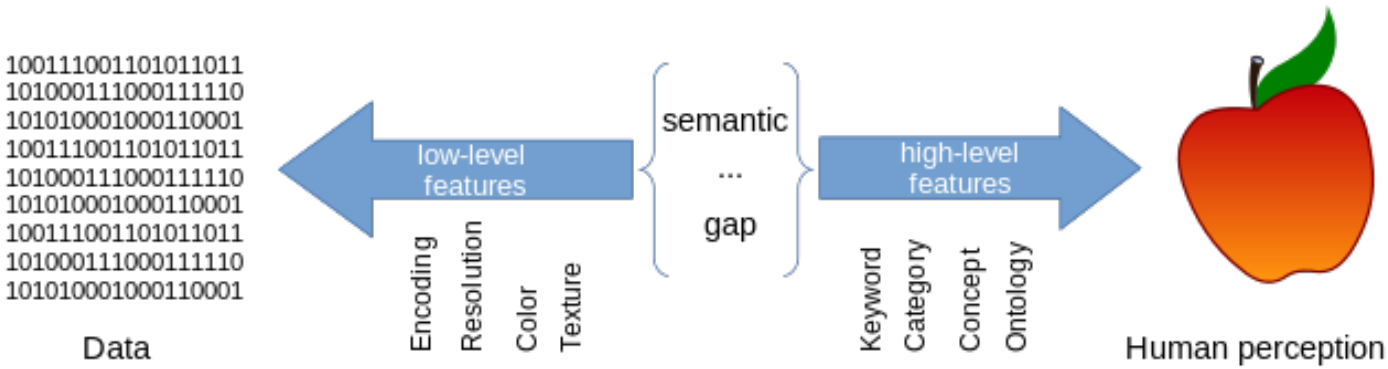
- Partly because it is an inverse problem
- i.e., seek to find something from insufficient information about the solution

Why is it hard?

- Partly because it is an inverse problem
- i.e., seek to find something from insufficient information about the solution
- Forward models are generally developed in physics and computer graphics

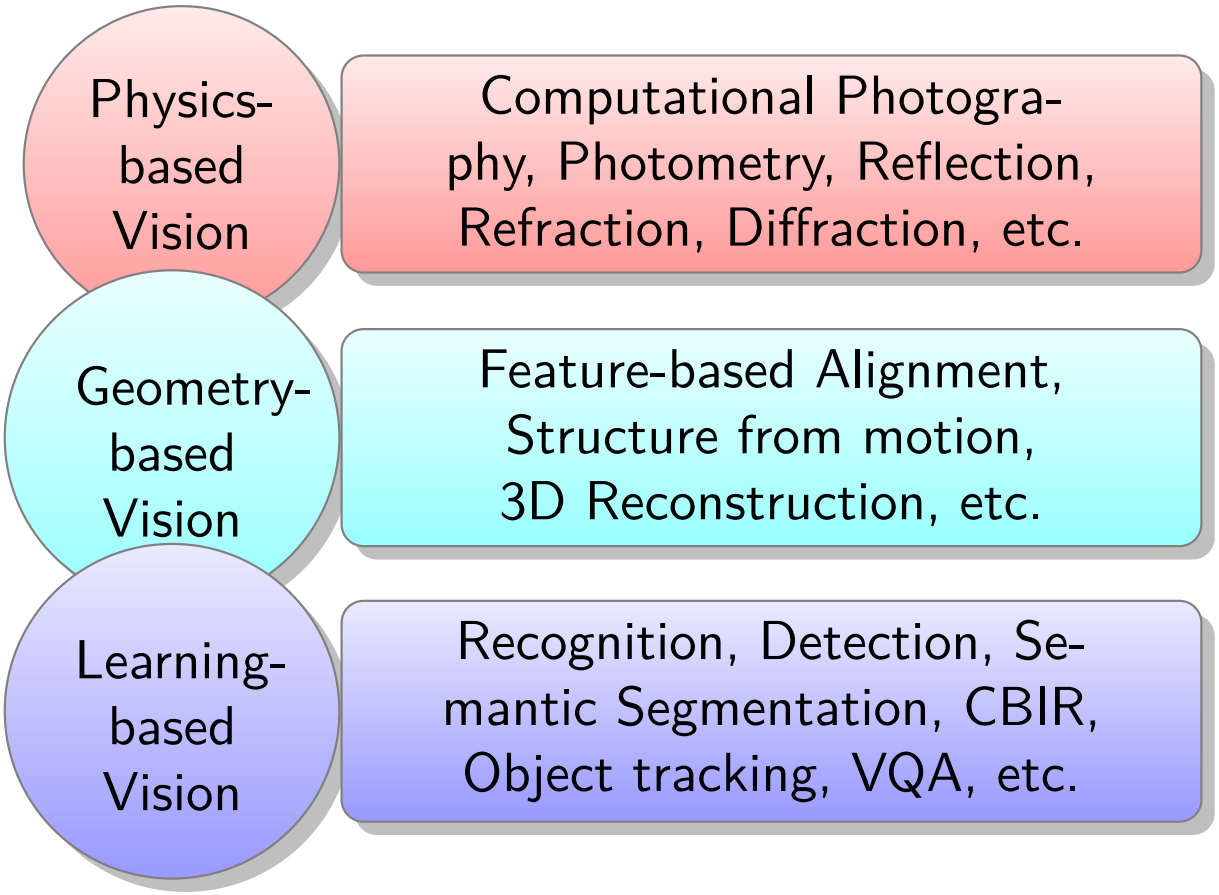
Why is it hard?

- Semantic gap



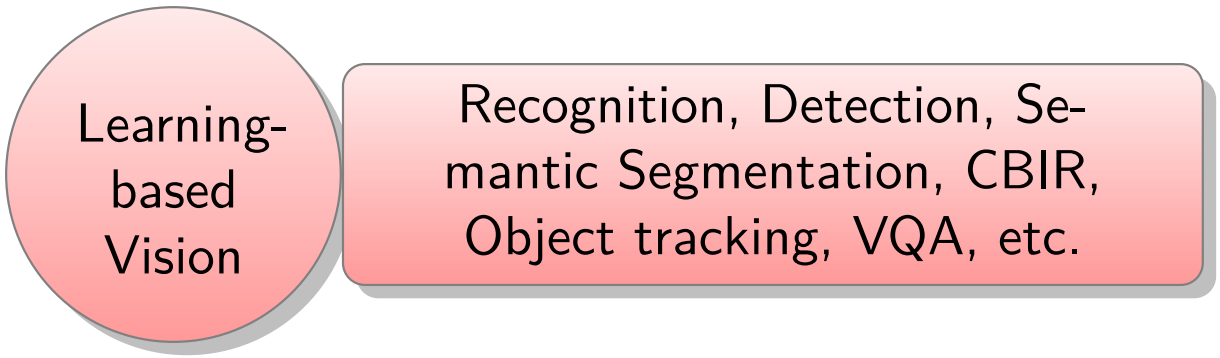
Source: Wikipedia

Computer Vision: Themes



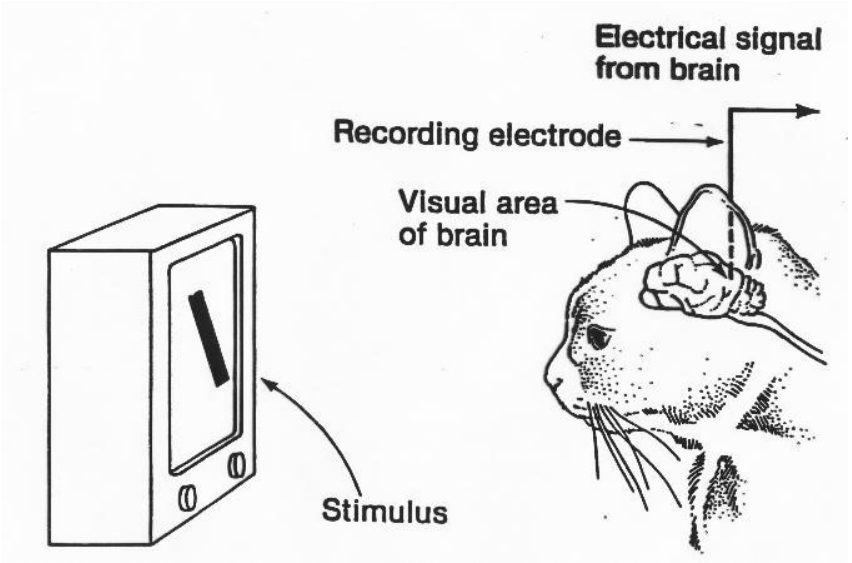
Taken from Prof. Vineet N Subramanian, IITH

Computer Vision: this course



Brief History: David Hubel and Torsten Wiesel (1959)

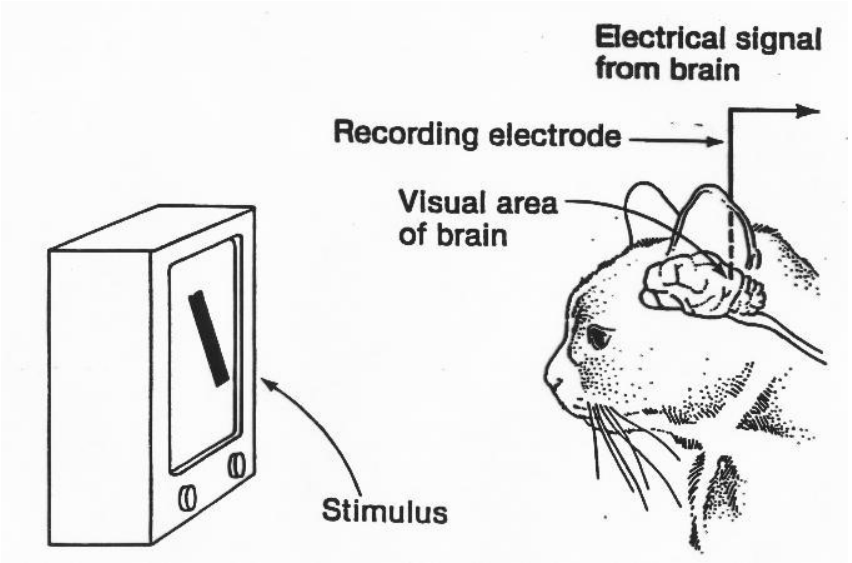
- Receptive fields of single neurons in the cat's striate cortex [Link to the experiment]



Source

Brief History: David Hubel and Torsten Wiesel (1959)

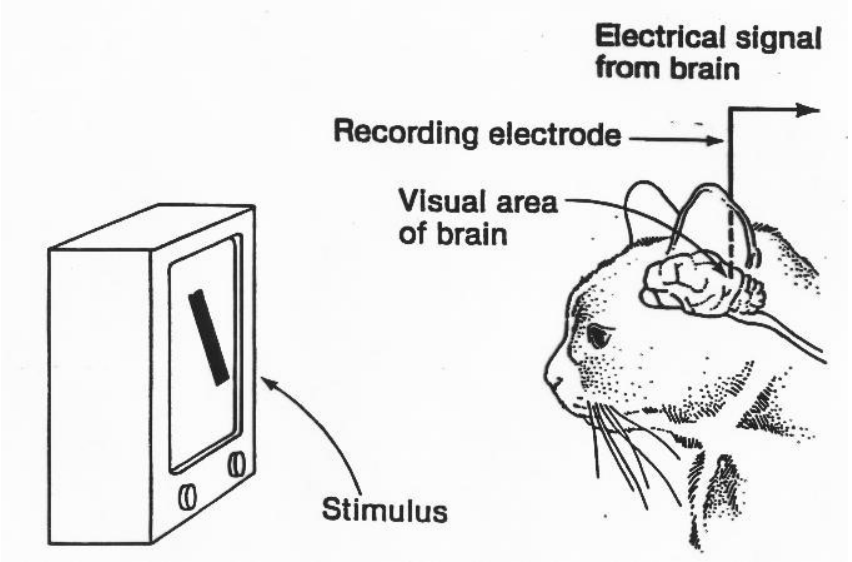
- Receptive fields of single neurons in the cat's striate cortex [Link to the experiment]
- Established that simple and complex neurons exist in visual cortex



Source

Brief History: David Hubel and Torsten Wiesel (1959)

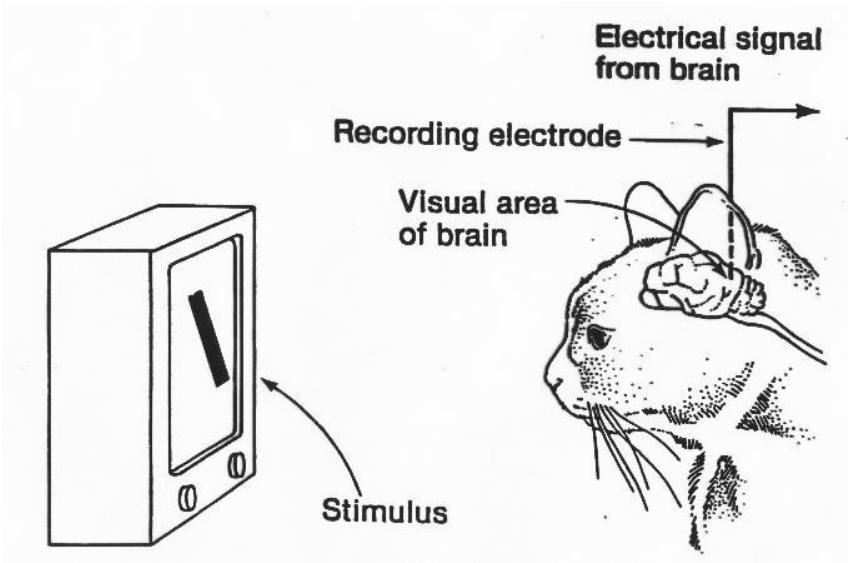
- Receptive fields of single neurons in the cat's striate cortex [Link to the experiment]
- Established that simple and complex neurons exist in visual cortex
- Visual processing starts with simple structures such as oriented edges (Remember this!)



Source

Brief History: David Hubel and Torsten Wiesel (1959)

- Receptive fields of single neurons in the cat's striate cortex [Link to the experiment]
- Established that simple and complex neurons exist in visual cortex
- Visual processing starts with simple structures such as oriented edges (Remember this!)
- Went on to win a Nobel in 1981!



Source

Brief History: Russel kirsch (1959)

- First digital image

Source

Brief History: Russel kirsch (1959)

- First digital image
- 176×176, 5cm in size

Source

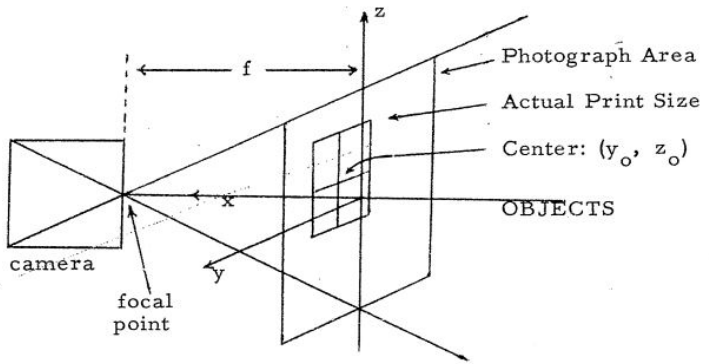
Brief History: Russel kirsch (1959)

- First digital image
- 176×176, 5cm in size
- Preserved in the Portland Art Museum

Source

Brief History: Lawrence Roberts (1963)

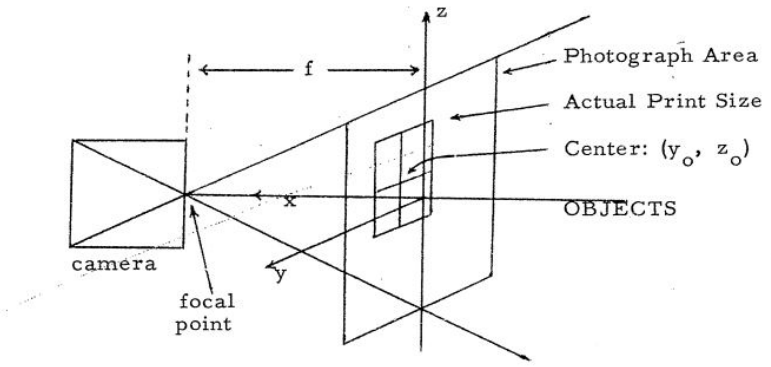
- Machine perception of three-dimensional solids



Camera Transformation

Brief History: Lawrence Roberts (1963)

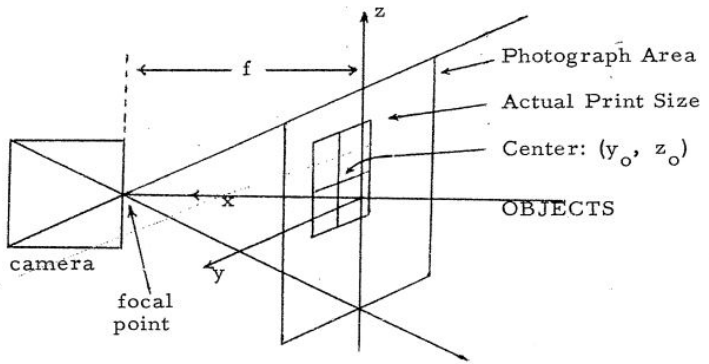
- Machine perception of three-dimensional solids
- Thesis described the process of deriving 3D info about solid objects from their 2D images of line drawings



Camera Transformation

Brief History: Lawrence Roberts (1963)

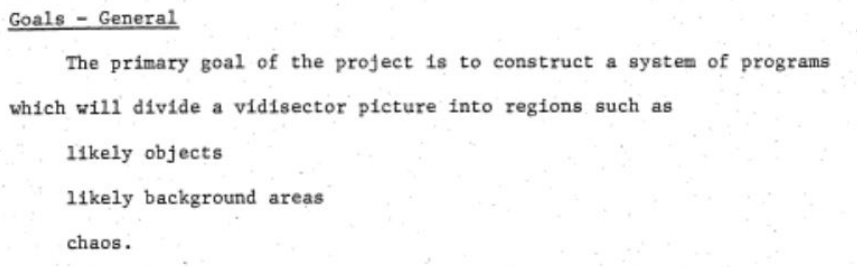
- Machine perception of three-dimensional solids
- Thesis described the process of deriving 3D info about solid objects from their 2D images of line drawings
- Camera transformations, perspective effects, depth perception, etc.



Camera Transformation

Brief History: Summer vision project (1966)

- Seymour Papert and Gerald Sussman (Aim document)



Goals - General

The primary goal of the project is to construct a system of programs which will divide a vidisector picture into regions such as

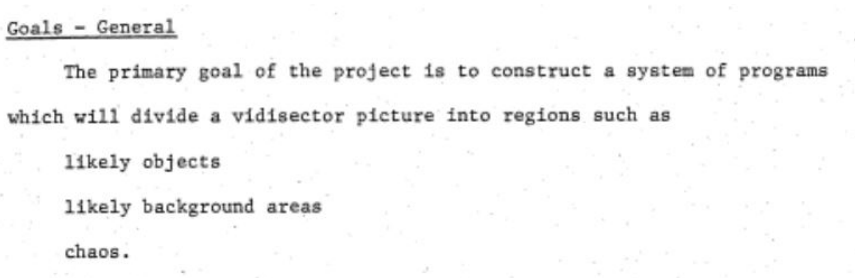
likely objects

likely background areas

chaos.

Brief History: Summer vision project (1966)

- Seymour Papert and Gerald Sussman (Aim document)
- Intended to develop a system for FG/BG segmentation, extracting non-overlapping objects from the real-world images



Goals - General

The primary goal of the project is to construct a system of programs which will divide a vidisector picture into regions such as

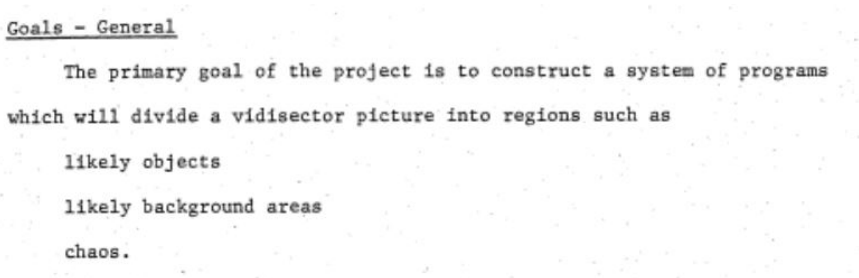
likely objects

likely background areas

chaos.

Brief History: Summer vision project (1966)

- Seymour Papert and Gerald Sussman (Aim document)
- Intended to develop a system for FG/BG segmentation, extracting non-overlapping objects from the real-world images
- 60 years later, the world is still working on it!



Goals - General

The primary goal of the project is to construct a system of programs which will divide a vidisector picture into regions such as

likely objects

likely background areas

chaos.

Brief History: 1970s (AI winter)

- MIT's AI lab offered first Machine Vision course

Brief History: 1970s (AI winter)

- MIT's AI lab offered first Machine Vision course
- First CV product for OCR (by Raymond Kurzweil)

Brief History: 1970s (AI winter)

- MIT's AI lab offered first Machine Vision course
- First CV product for OCR (by Raymond Kurzweil)
- Object recognition through shape analysis (Generalized Cylinders, Skeletons, etc.)

Brief History: David Marr (1982)

- “Vision: A computational investigation into the human representation and processing of visual information"

Brief History: David Marr (1982)

- "Vision: A computational investigation into the human representation and processing of visual information"
- Established the "Hierarchy" of the vision: high-level understanding of visual data is built on top of the low-level tools for detecting edges, curves, corners, etc.

David Marr's Representational framework (1982)

- Primal sketch of the image (edges, boundaries, etc.) are represented

Rostyslav Demush, medium.com

David Marr's Representational framework (1982)

- Primal sketch of the image (edges, boundaries, etc.) are represented
- $2.5D$ representation: depth and discontinuities are represented

Rostyslav Demush, medium.com

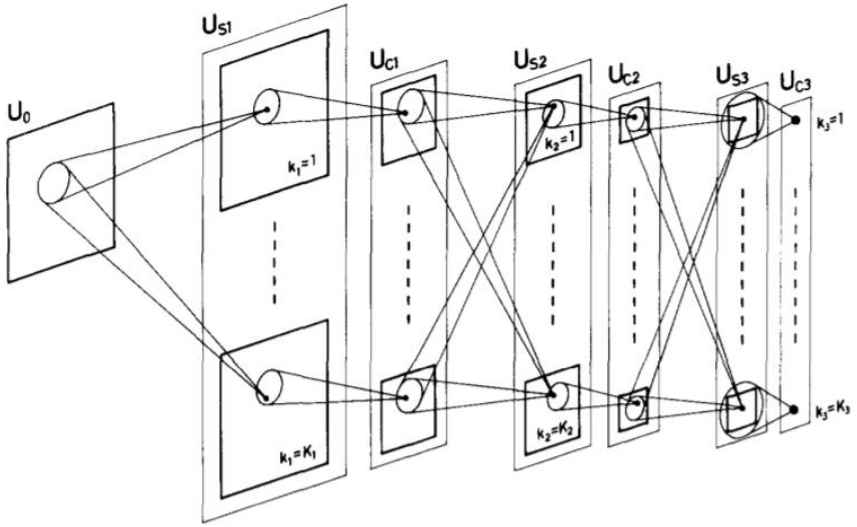
David Marr's Representational framework (1982)

- Primal sketch of the image (edges, boundaries, etc.) are represented
- $2.5D$ representation: depth and discontinuities are represented
- $3D$ model hierarchically organized in terms of surface and volumetric primitives

Rostyslav Demush, medium.com

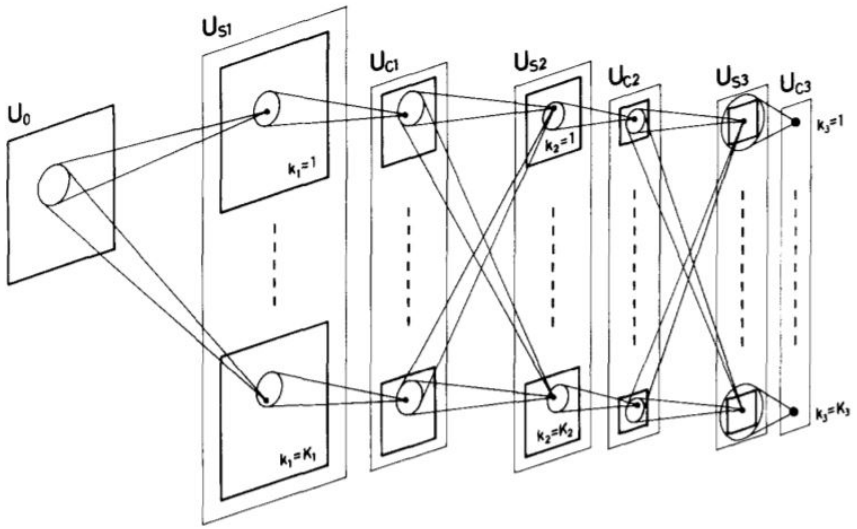
Brief History: Neocognitron (1979-82)

- Fukushima implements the Hubel and Wiesel's principles



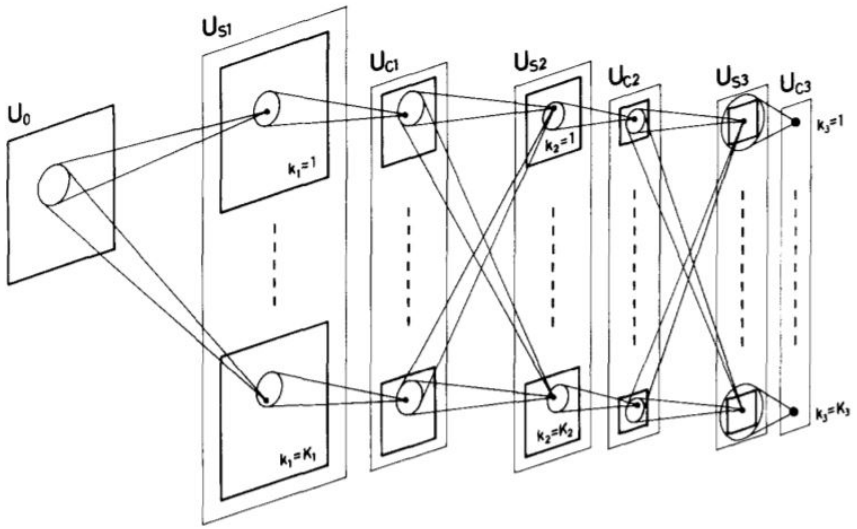
Brief History: Neocognitron (1979-82)

- Fukushima implements the Hubel and Wiesel's principles
- Used for hand-written digit recognition



Brief History: Neocognitron (1979-82)

- Fukushima implements the Hubel and Wiesel's principles
- Used for hand-written digit recognition
- Viewed as precursor for the modern CNNs (had conv filters and layers, spatial invariance)



Brief History: Optical Flow (1981)

- Determining Optical Flow by Horn and Schunck

Source

Brief History: Optical Flow (1981)

- Determining Optical Flow by Horn and Schunck
- Estimates the direction and speed of moving objects across pair of images

Source

Brief History: Canny Edge detection (1986)

- Multi-stage approach for detecting the edge content in an image

Source:OpenCV

Brief History: Canny Edge detection (1986)

- Multi-stage approach for detecting the edge content in an image
- Signal variations are dealt with calculus (simple but popular method)

Source:OpenCV

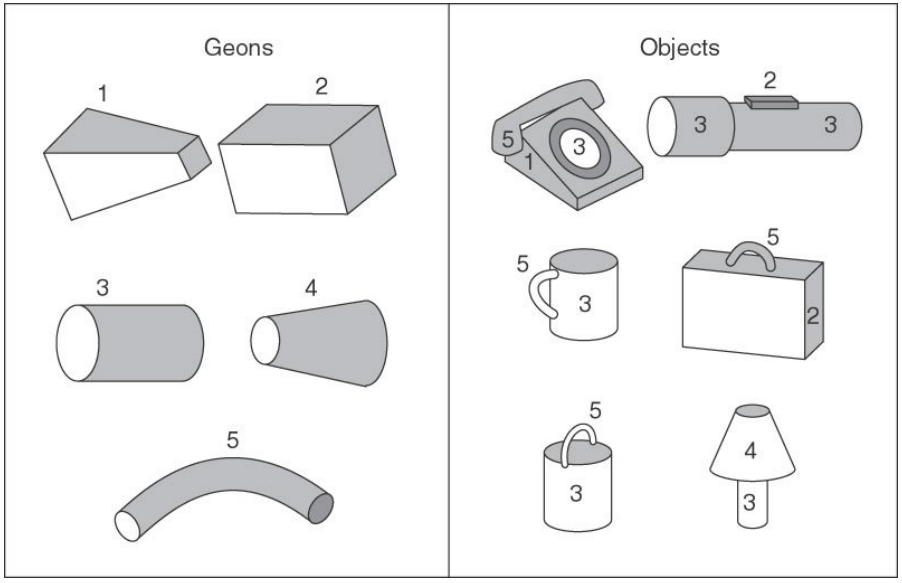
Brief History: Canny Edge detection (1986)

- Multi-stage approach for detecting the edge content in an image
- Signal variations are dealt with calculus (simple but popular method)
- Developed as a masters student, published in Trans. on PAMI, 1986 (Link)

Source:OpenCV

Brief History: Recognition by components (1987)

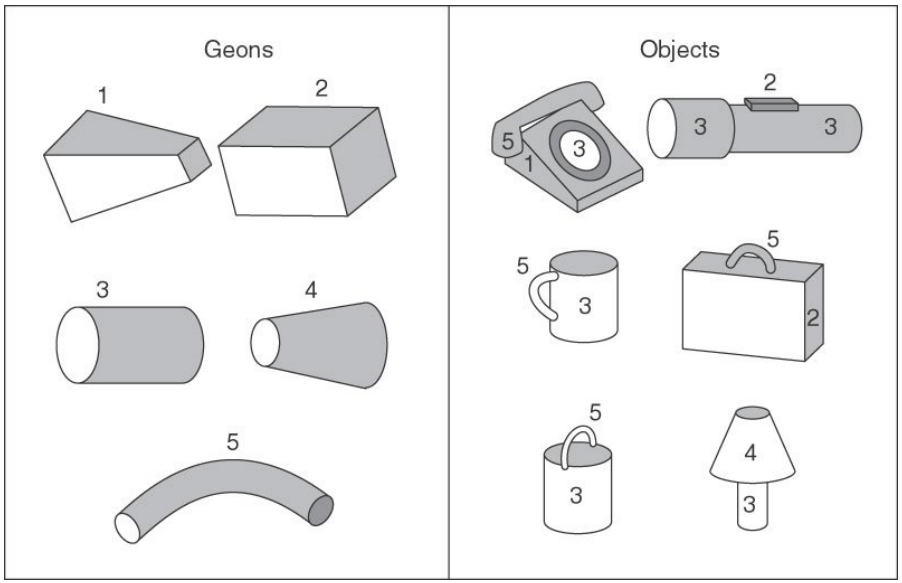
- Bottom-up process for object recognition proposed by Irving Biederman



Source: Prof. Kenneth M. Steele

Brief History: Recognition by components (1987)

- Bottom-up process for object recognition proposed by Irving Biederman
- Simple $3D$ shapes (geons) such as cones and cylinders compose objects



Source: Prof. Kenneth M. Steele

Brief History: snakes and Contours (1988)

- Active contour models (Snakes) aim to outline the objects of interest from the images

Source

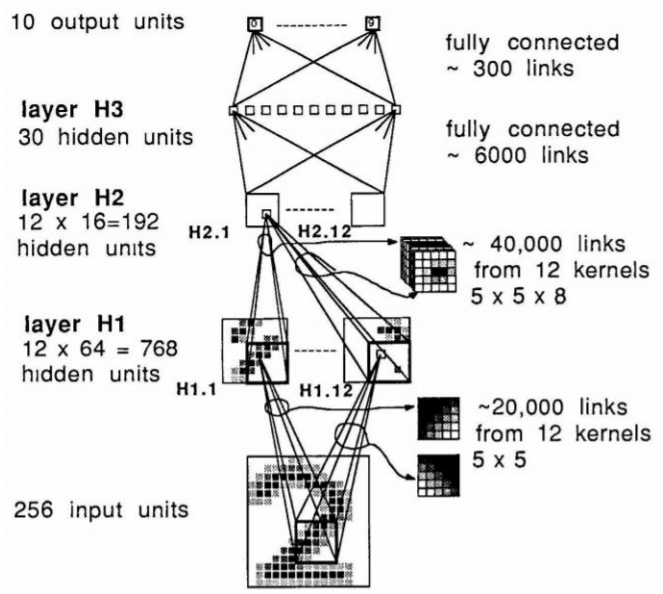
Brief History: snakes and Contours (1988)

- Active contour models (Snakes) aim to outline the objects of interest from the images
- Widely applied in edge detection, segmentation, shape recognition, object tracking, etc.

Source

Brief History: Backpropagation (1989)

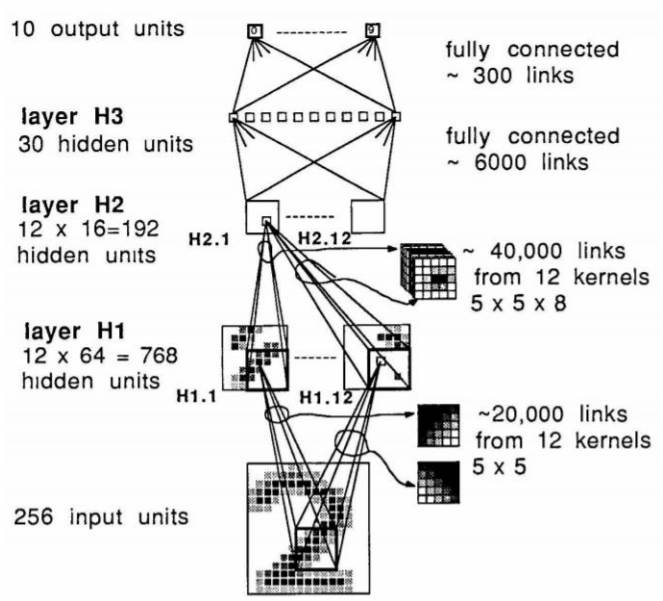
- Prof. Yan Lecun applied a backprop style learning algorithm to Fukushima's convolutional neural network



Source

Brief History: Backpropagation (1989)

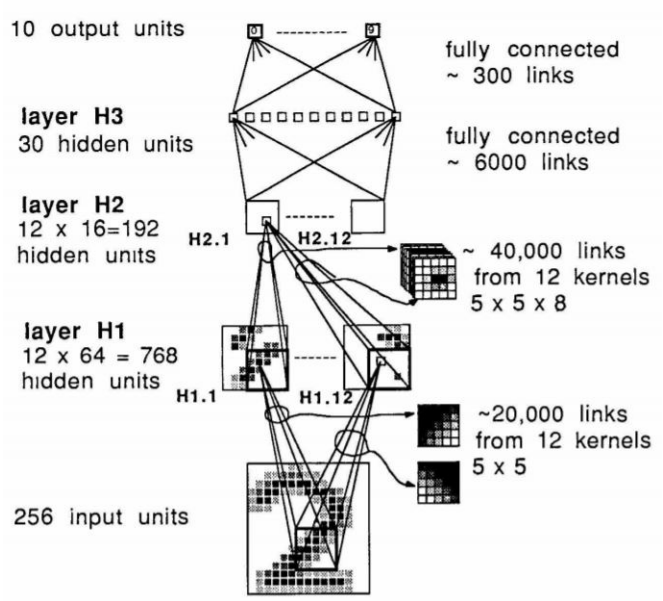
- Prof. Yan Lecun applied a backprop style learning algorithm to Fukushima's convolutional neural network
- Developed a commercial product for digit recognition and released MNIST dataset



Source

Brief History: Backpropagation (1989)

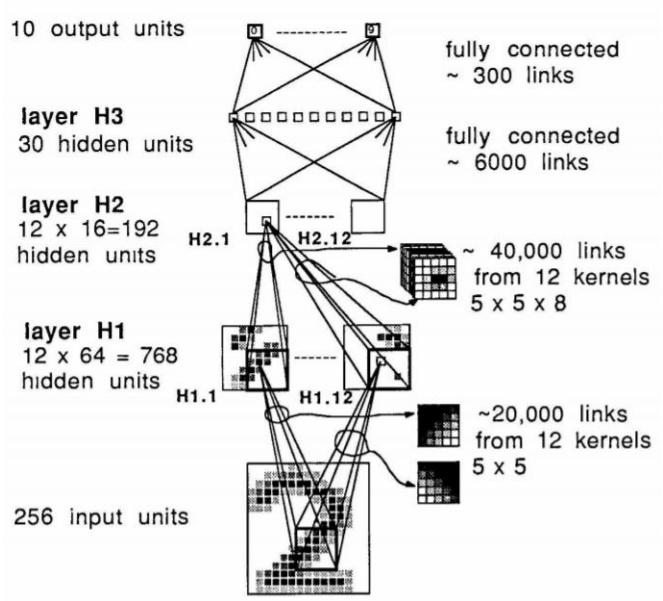
- Prof. Yan Lecun applied a backprop style learning algorithm to Fukushima's convolutional neural network
- Developed a commercial product for digit recognition and released MNIST dataset
- Very similar to modern CNN architectures



Source

Brief History: Backpropagation (1989)

- Prof. Yan Lecun applied a backprop style learning algorithm to Fukushima's convolutional neural network
- Developed a commercial product for digit recognition and released MNIST dataset
- Very similar to modern CNN architectures
- Backpropagation is attributed 'Majorly' to Paul Werbos 1974 (although it was independently discovered by multiple from 1960s)



Source

Brief History: Other works in 1980s

- Image Pyramids and scale-space processing

Brief History: Other works in 1980s

- Image Pyramids and scale-space processing
- Wavelets

Brief History: Other works in 1980s

- Image Pyramids and scale-space processing
- Wavelets
- Markov Random Fields

Brief History: Other works in 1980s

- Image Pyramids and scale-space processing
- Wavelets
- Markov Random Fields
- Variational Optimization Methods

Brief History: Moving from low-level processing in 1990s

- Face recognition by Eigen faces (Turk and Pentland, 1991)

Brief History: Moving from low-level processing in 1990s

- Face recognition by Eigen faces (Turk and Pentland, 1991)
- Normalized cuts (J Shi and J Malik, 1997)

Brief History: Moving from low-level processing in 1990s

- Face recognition by Eigen faces (Turk and Pentland, 1991)
- Normalized cuts (J Shi and J Malik, 1997)
- Particle filters and Mean-shift for tracking (Liu and Chen, 1998) (Cheng, 1998)

Brief History: Moving from low-level processing in 1990s

- Face recognition by Eigen faces (Turk and Pentland, 1991)
- Normalized cuts (J Shi and J Malik, 1997)
- Particle filters and Mean-shift for tracking (Liu and Chen, 1998) (Cheng, 1998)
- Scale-Invariant Feature Transform (D Lowe, 2004)

Brief History: Moving from low-level processing in 1990s

- Face recognition by Eigen faces (Turk and Pentland, 1991)
- Normalized cuts (J Shi and J Malik, 1997)
- Particle filters and Mean-shift for tracking (Liu and Chen, 1998) (Cheng, 1998)
- Scale-Invariant Feature Transform (D Lowe, 2004)
- ....

Brief History: Moving from low-level processing in 2000s

- Face detection by Viola Jones (2001)

Brief History: Moving from low-level processing in 2000s

- Face detection by Viola Jones (2001)
- Conditional Random Fields (Lafferty et al, 2001)

Brief History: Moving from low-level processing in 2000s

- Face detection by Viola Jones (2001)
- Conditional Random Fields (Lafferty et al, 2001)
- PASCAL VOC Dataset $\rightarrow$ boost Recognition applications

Brief History: Moving from low-level processing in 2000s

- Face detection by Viola Jones (2001)
- Conditional Random Fields (Lafferty et al, 2001)
- PASCAL VOC Dataset $\rightarrow$ boost Recognition applications
- Constellation methods (R Fergus, Perona and A Zisserman, 2007)

Brief History: Moving from low-level processing in 2000s

- Face detection by Viola Jones (2001)
- Conditional Random Fields (Lafferty et al, 2001)
- PASCAL VOC Dataset $\rightarrow$ boost Recognition applications
- Constellation methods (R Fergus, Perona and A Zisserman, 2007)
- Deformable parts model (Felzenszwalb et al, 2009)

Brief History: Deep learning era in 2010s

- Fully connected networks (FCN) $\rightarrow$ semantic segmentation (2015)

Brief History: Deep learning era in 2010s

- Fully connected networks (FCN) $\rightarrow$ semantic segmentation (2015)
- COCO and VQA datasets came up (2015)

Brief History: Deep learning era in 2010s

- Fully connected networks (FCN) $\rightarrow$ semantic segmentation (2015)
- COCO and VQA datasets came up (2015)
- SSD and YOYO for object detection; Visual Genome dataset (2016)

Brief History: Deep learning era in 2010s

- Fully connected networks (FCN) $\rightarrow$ semantic segmentation (2015)
- COCO and VQA datasets came up (2015)
- SSD and YOYO for object detection; Visual Genome dataset (2016)
- Scene-graphs; (2017)

Brief History: Deep learning era in 2010s

- Fully connected networks (FCN) $\rightarrow$ semantic segmentation (2015)
- COCO and VQA datasets came up (2015)
- SSD and YOYO for object detection; Visual Genome dataset (2016)
- Scene-graphs; (2017)
- Higher-levels of abstraction: VCR dataset, panoptic segmentation, etc. (2018-19)

Brief History: Deep learning era in 2010s

- Fully connected networks (FCN) $\rightarrow$ semantic segmentation (2015)

Brief History: Deep learning era in 2010s

- Fully connected networks (FCN) $\rightarrow$ semantic segmentation (2015)
- COCO and VQA datasets came up (2015)

Brief History: Deep learning era in 2010s

- Fully connected networks (FCN) $\rightarrow$ semantic segmentation (2015)
- COCO and VQA datasets came up (2015)
- SSD and YOYO for object detection; Visual Genome dataset (2016)

Brief History: Deep learning era in 2010s

- Fully connected networks (FCN) $\rightarrow$ semantic segmentation (2015)
- COCO and VQA datasets came up (2015)
- SSD and YOYO for object detection; Visual Genome dataset (2016)
- Scene-graphs; (2017)

Brief History: Deep learning era in 2010s

- Fully connected networks (FCN) $\rightarrow$ semantic segmentation (2015)
- COCO and VQA datasets came up (2015)
- SSD and YOYO for object detection; Visual Genome dataset (2016)
- Scene-graphs; (2017)
- Higher-levels of abstraction: VCR dataset, panoptic segmentation, etc. (2018-19)

Brief History: Deep learning era in 2010s

Turing Award winners for 2018

Course Contents

- Part-1: Foundations of Deep learning (Implementing and training different types of neural networks)
  - (MP, perceptron), MLP, CNNs, and RNNs (LSTM and GRU)
  - Gradient Descent technique using the Backpropagation
  - Implement them in PyTorch framework (this is not a lab course, so it is majorly your responsibility!)

Course Contents

- Part-1: Foundations of Deep learning (Implementing and training different types of neural networks)
  - (MP, perceptron), MLP, CNNs, and RNNs (LSTM and GRU)
  - Gradient Descent technique using the Backpropagation
  - Implement them in PyTorch framework (this is not a lab course, so it is majorly your responsibility!)
- Part-2: Applications in Computer Vision (with a slight research flavour)
  - Object recognition, detection, semantic segmentation Vision and Language
  - Generative models: GANs and VAEs
  - Recent trends

Prerequisites

Theory

- Knowledge on basics of probability, linear algebra, and calculus
- Basic course on ML
- Exposure to Deep learning (a course greatly helps)

Practicals

- Programming in Python
- Knowledge of a deep learning framework (we work with PyTorch)

Time slot

- D1 slot
  - Monday 4 - 4:55 PM
  - Tuesday 4 - 4:54 PM
  - Friday 3 - 3:55 PM

Time slot

- Open elective (Final year B.Tech, M.Tech, and Ph.D.)

Time slot

- Open elective (Final year B.Tech, M.Tech, and Ph.D.)
- Class Room - 4104 (CORE-4, First Floor)

Logistics

- Course website: https://krmopuri.github.io/dl4cv/
  - Course updates
  - Lecture slides and other material
  - Assignments
  - etc.

Evaluation (Tentative)

- Assignments - 30%
- Mid-semester - 20%
- End-semester - 30% Mini-project - 20%

Textbooks and References

- Computer Vision
  - Computer Vision: A Modern Approach, Forsyth and Ponce
  - Computer Vision: Algorithms and Applications, Richard Szeliski
- Deep Learning
  - Deep Learning textbook by Ian Goodfellow *et al.*
  - NPTEL course by Prof. Mitesh Khapra, IITM.
  - Michael Nielsen's text book on NN & DL
  - DL course by François Fleuret, EPFL and Uni. of Geneva
  - PyTorch - `https://pytorch.org/`
  - Many more that I could not list and am not aware of...
- DL for CV
  - NPTEL Course by Prof. Vineet Balasubramanian, IITH.
  - Course by Dr. Justin Johnson, University of Michigan