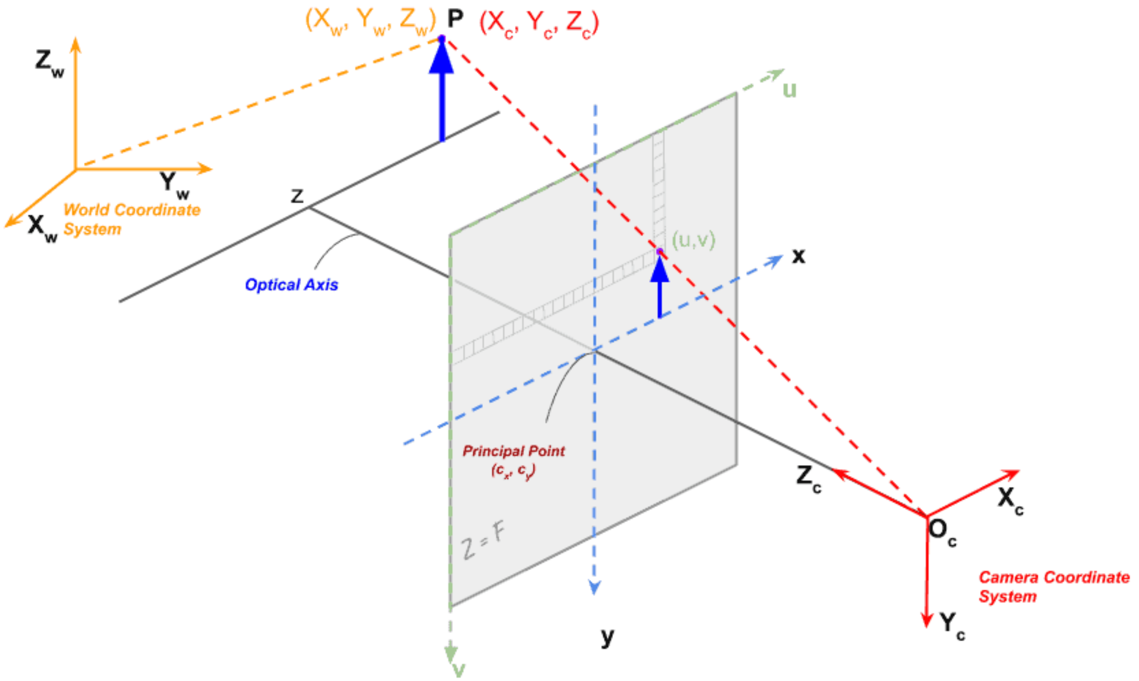
BLG 453E

Week 1
Introduction to Computer Vision and Image Data

Dr. Yusuf H. Sahin

Computer Vision & Image Formation

- In a broad sense, vision is the inverse of image formation.
  - **Image formation:** How objects create images
    Computer Vision: How to use the image to understand objects in space.



https://learnopencv.com/geometry-of-image-formation/

In order to establish a precise correspondence between points in 3-D space (with respect to a fixed global reference frame) and their projected images in a 2-D image plane (with respect to a local coordinate frame), a mathematical model for this process must account for three types of transformations:

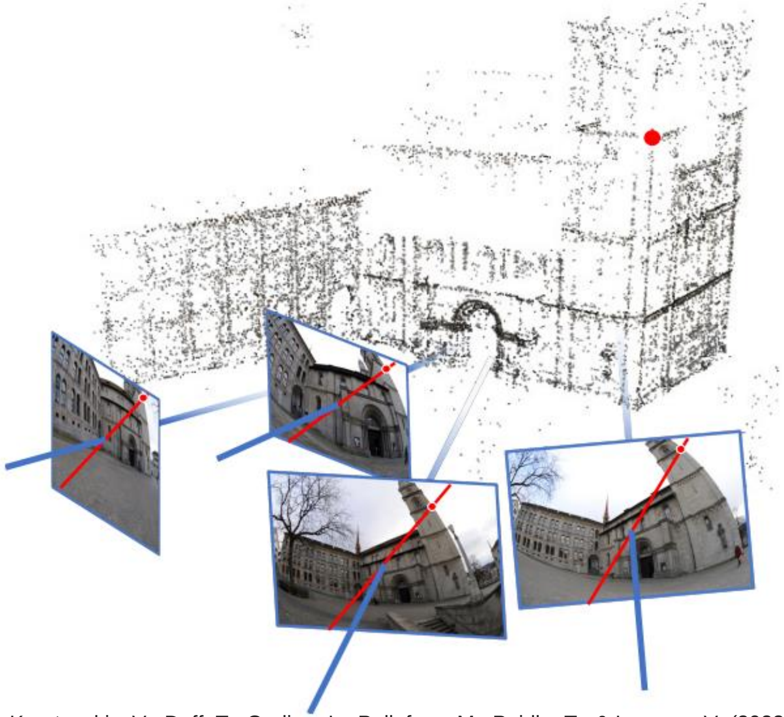
I. coordinate transformations between the camera frame and the world frame (coordinate transform)

II. projection of 3-D coordinates onto 2-D image coordinates (perspective transform)

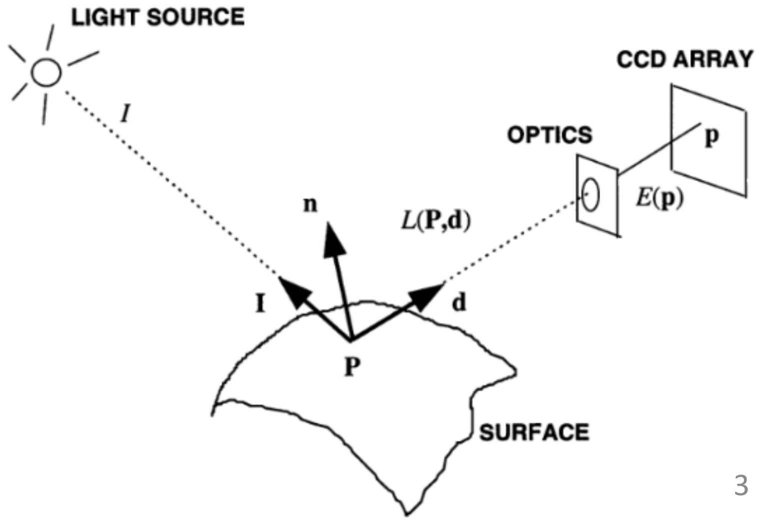
III. coordinate transformation between possible choices of image coordinate frame (pixel coordinates)

Steps of Image Formation

- Geometric image formation deals with points, lines, and planes, and how these are mapped onto images using projective geometry and other models (including radial lens distortion).
- Photometric image formation covers radiometry, which describes how light interacts with surfaces in the world.
- Optics deals with the light rays which enter the camera through an angular aperture and hit a screen or image plane, the camera's photosensitive device which registers light intensities.

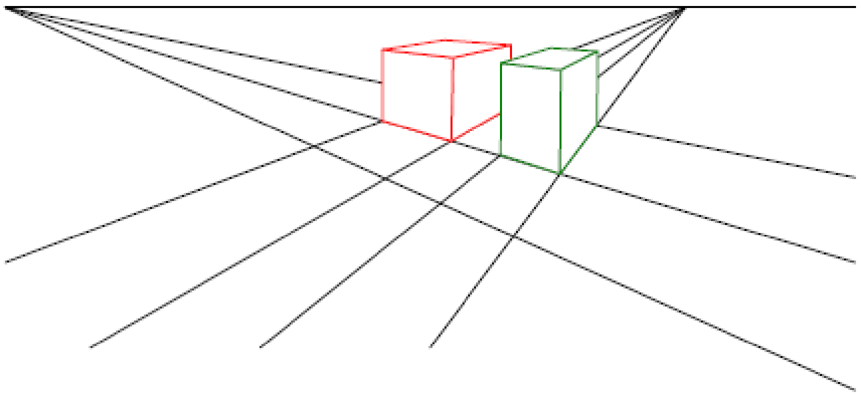


Hruby, P., Korotynskiy, V., Duff, T., Oeding, L., Pollefeys, M., Pajdla, T., & Larsson, V. (2023). Four-view geometry with unknown radial distortion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8990-9000).

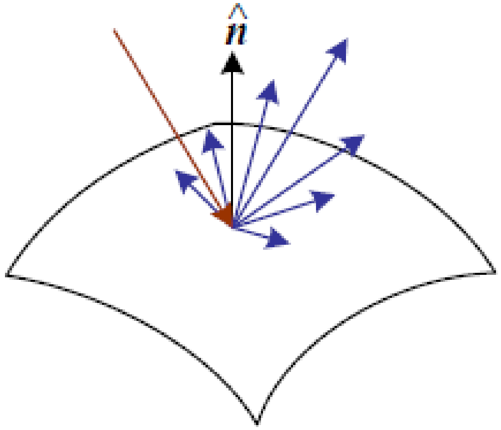


The Full Process

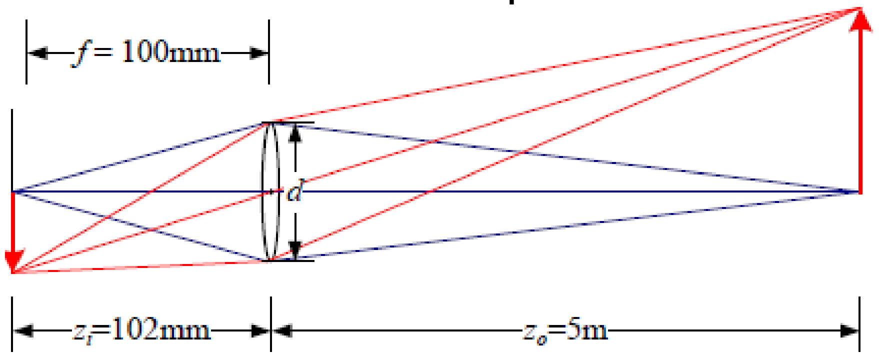
perspective projection



light scattering



lens optics

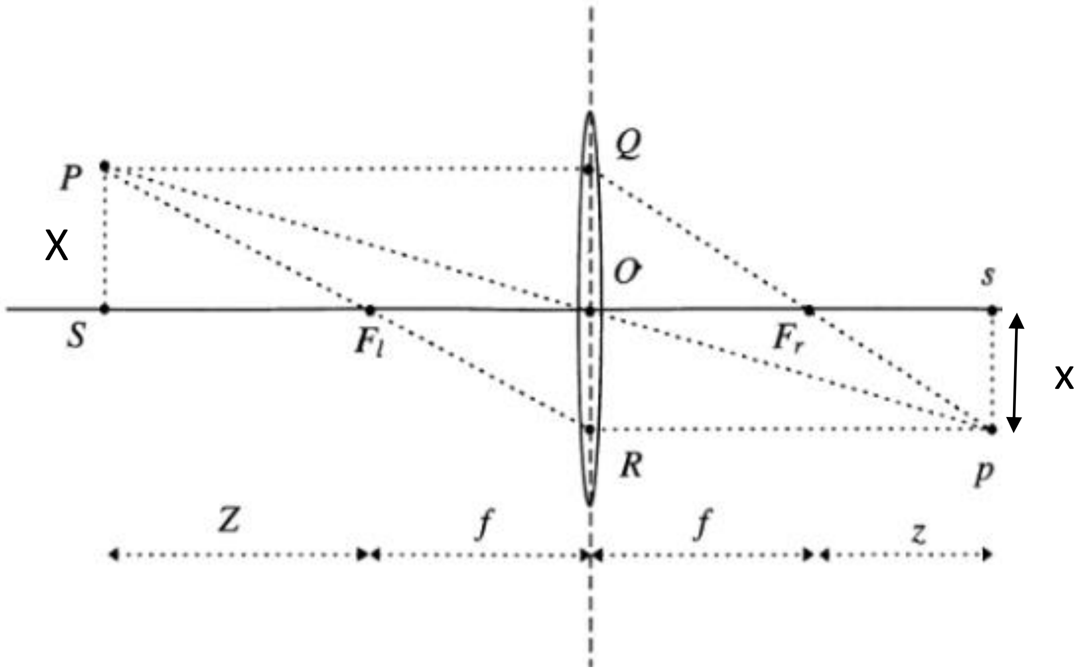


Charge-Coupled Devices with Bayer pattern

G	R	G	R
B	G	B	G
G	R	G	R
B	G	B	G

Basic Optics: Thin Lens Model

- A thin lens deflects all rays parallel to the optical axis and coming from one side onto the focus on the other side, as described by two basic properties:
  - (1) Any ray entering the lens parallel to the axis on one side goes through the focus on the other side.
  - (2) Any ray entering the lens from the focus on one side emerges parallel to the axis on the other side.



For a point P, not too far from the optical axis,

- Z + f be the distance of P from the lens along the optical axis.
  - Using the two pairs of similar triangles

$$< P, F_l, S > and < R, O, F_l >$$
$$< p, s, F_r > and < Q, O, F_r >$$

$$\frac{x}{z} = \frac{X}{f} \quad \frac{X}{Z} = \frac{x}{f} \longrightarrow Zz = f^2$$

- Setting $\hat{Z} = Z + f$ and $\hat{z} = z + f$

$$f^2 = (\hat{Z} - f)(\hat{z} - f)$$
$$f(\hat{Z} + \hat{z}) = \hat{Z}\hat{z}$$
$$\frac{1}{f} = \frac{\hat{Z} + \hat{z}}{\hat{Z}\hat{z}}$$

$$\boxed{\frac{1}{f} = \frac{1}{\hat{Z}} + \frac{1}{\hat{z}}}$$

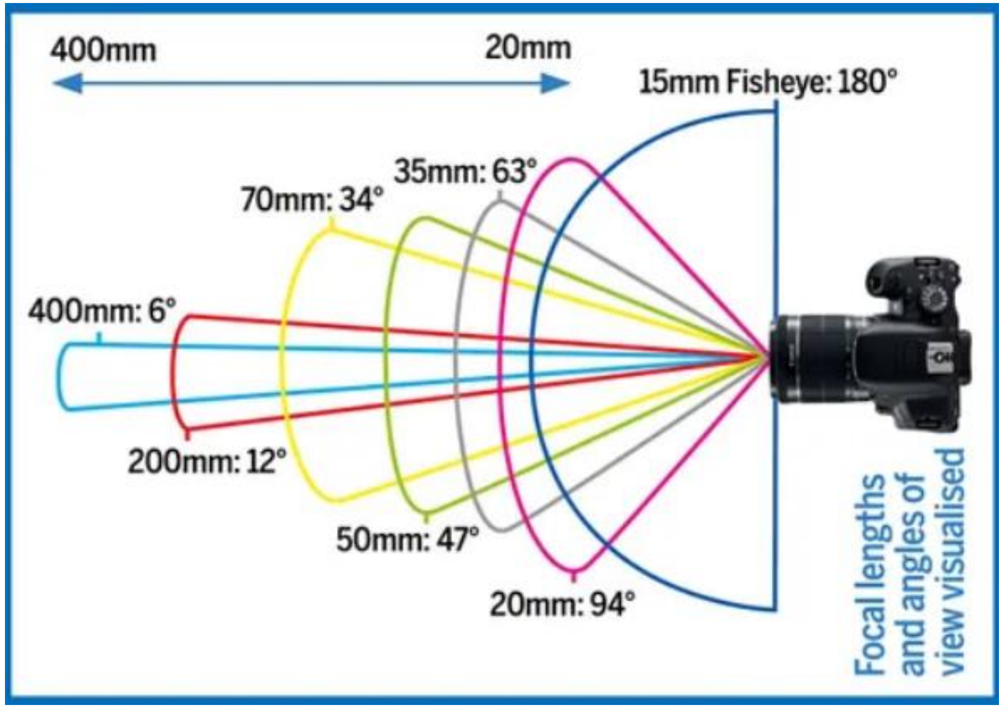
Field of View

- The thin lens has two parameters: its focal length f and its diameter 2r.
- We define the field of view (FOV) to be the angle subtended by the spatial extent of the sensor as seen from the optical center. If 2r is the largest spatial extension of the sensor (e.g., the side of the CCD), then the field of view is

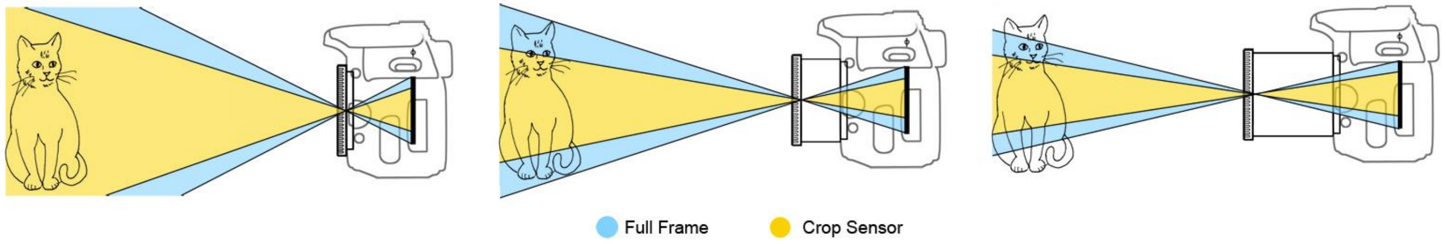
$$\theta = 2 * \arctan\left(\frac{r}{f}\right)$$

- For a full-frame sensor (36 mm in modern cameras), and focal length of 35 mm:

$$2 * \arctan\left(\frac{18}{35}\right) \cong 63^o$$

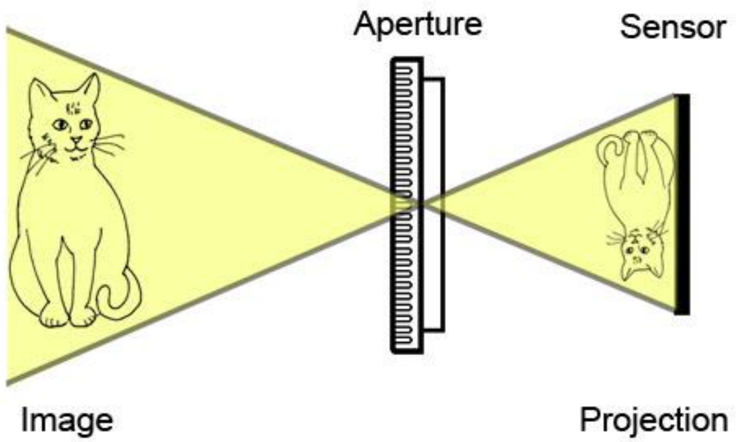


https://www.digitalcameraworld.com/tutorials/photography-cheat-sheet-what-is-field-of-view-fov

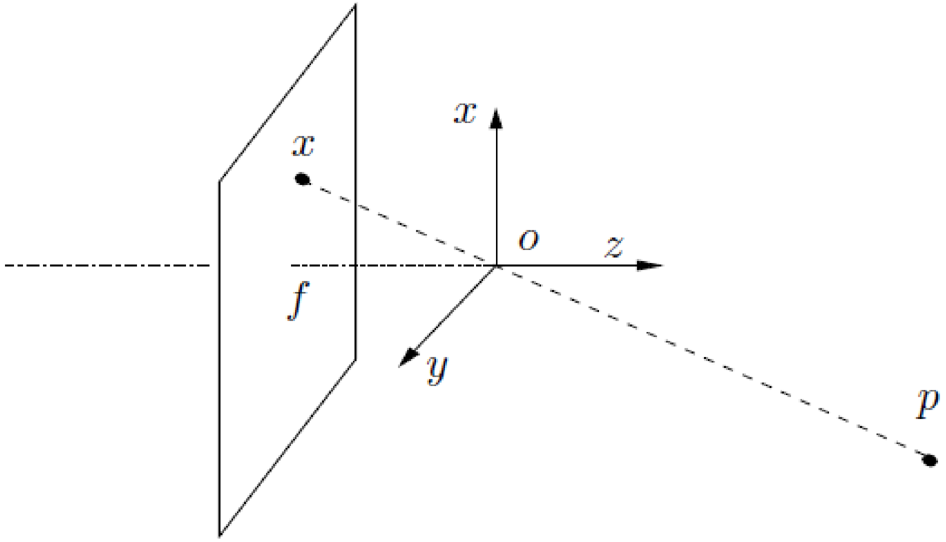


https://daystarlaser.com/education/pinhole-photography-primer/

Pinhole Camera Model



- Reducing the camera's aperture to a point, called a pinhole.
- Only one ray from any given point can enter the camera, and creates a one-to-one correspondence between visible points, rays, and image points.
- This results in very sharp, undistorted images of objects at different distances from the camera.



- If a point p has coordinates $X = [X,\ Y,\ Z]^T$ relative to a reference frame centered at the optical center o, with its z-axis being the optical axis (of the lens), then it is immediate to see from similar triangles
  - The ideal perspective projection.

$$x = -f\frac{X}{Z} \qquad y = -f\frac{Y}{Z}$$

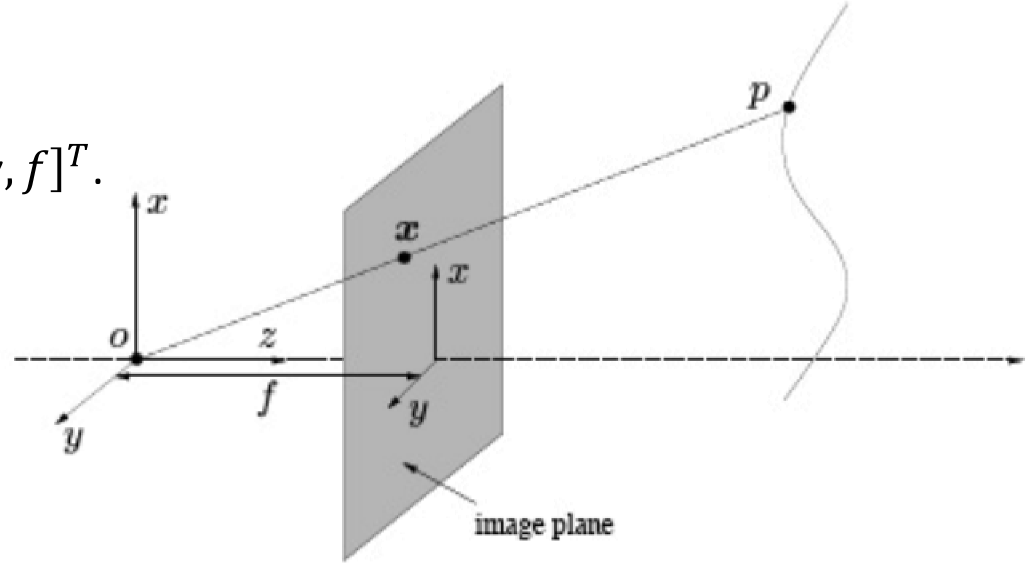
  - Negative sign: Upside down

Frontal Pinhole Camera Model

- To eliminate negative sign, we can simply flip the image: $(x, y) \rightarrow (-x, -y)$.
- This corresponds to placing the image plane $\{z = -f\}$ in front of the optical center instead $\{z = +f\}$.

$$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z}$$

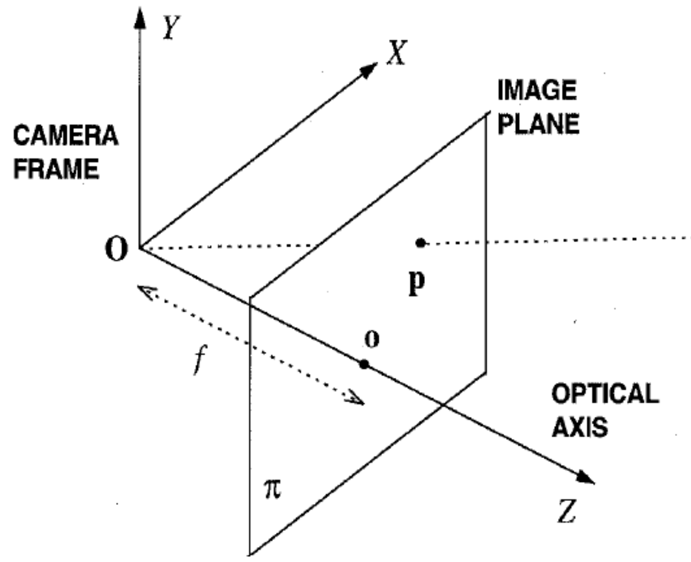
- In the camera frame, the third component of an image point is always equal to the focal length (as the equation of the plane n is $z = f$).
- For this reason, we will often write $p = [x, y]$ instead of $p = [x, y, f]^T$.



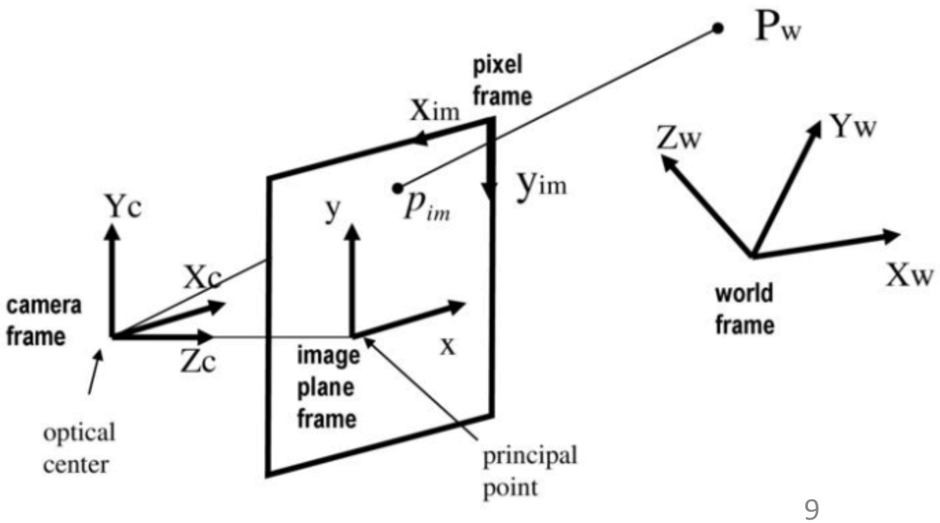
Ideal Perspective Camera Model

- Sometimes, we simply write the projection as a map π:
  - $\pi: \mathbb{R}^3 \rightarrow \mathbb{R}^2; X \mapsto x$
  - $x = \pi(X)$

$$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z}$$



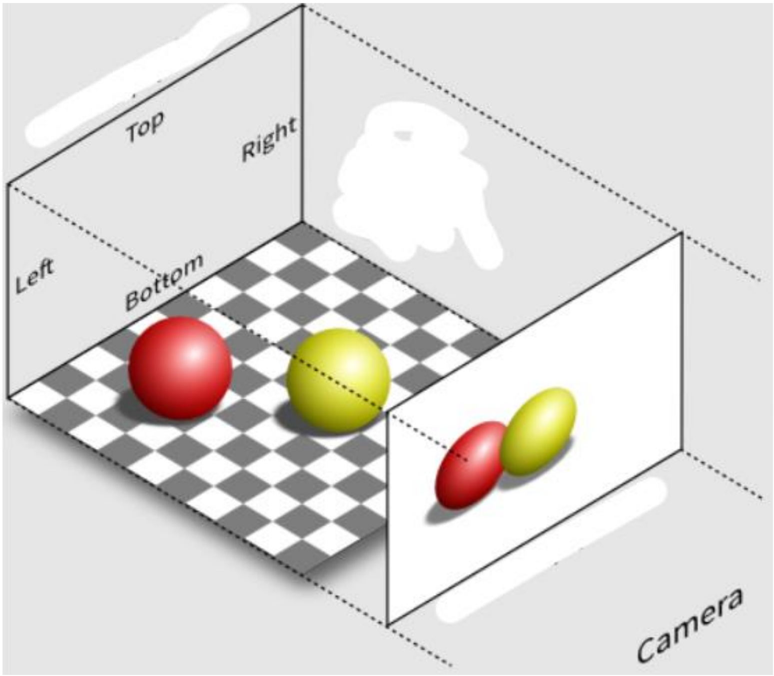
- The intersection between image plane and the optical axis, is named principal point or image center.
- p, the image of P, is the point at which the straight line through P and O intersects the image plane.
- Camera Frame: The 3D reference frame in which O is the origin and the plane n is orthogonal to the Z axis.



Approximate Camera Models

- Ortographic Projection
  - The light rays in the orthographic model travel along the lines parallel to the optical axis.

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

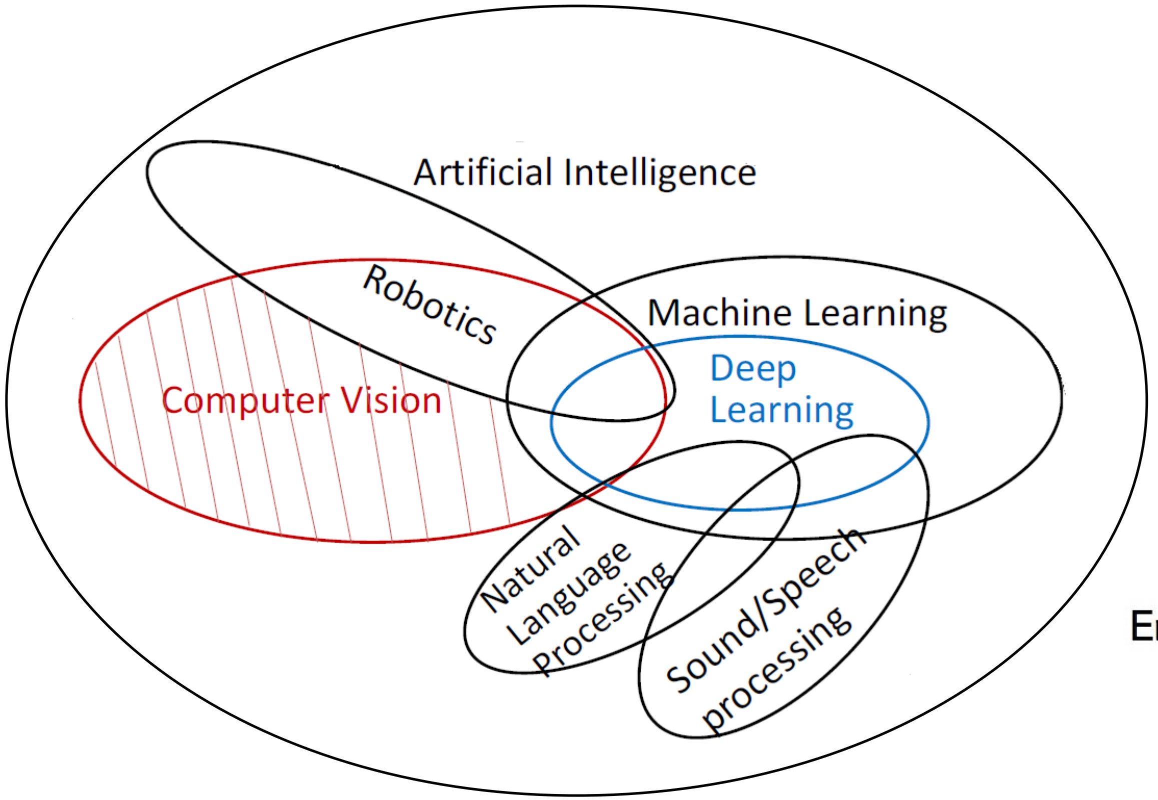


- Weak Perspective Camera Model (Scaled Ortographic)
  - If the average depth of the scene, $\bar{Z}$, is much larger than the relative distance

$$x = f\frac{X}{\bar{Z}} \qquad y = f\frac{Y}{\bar{Z}}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = s \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, s = \frac{f}{\bar{Z}}$$

Computer Vision



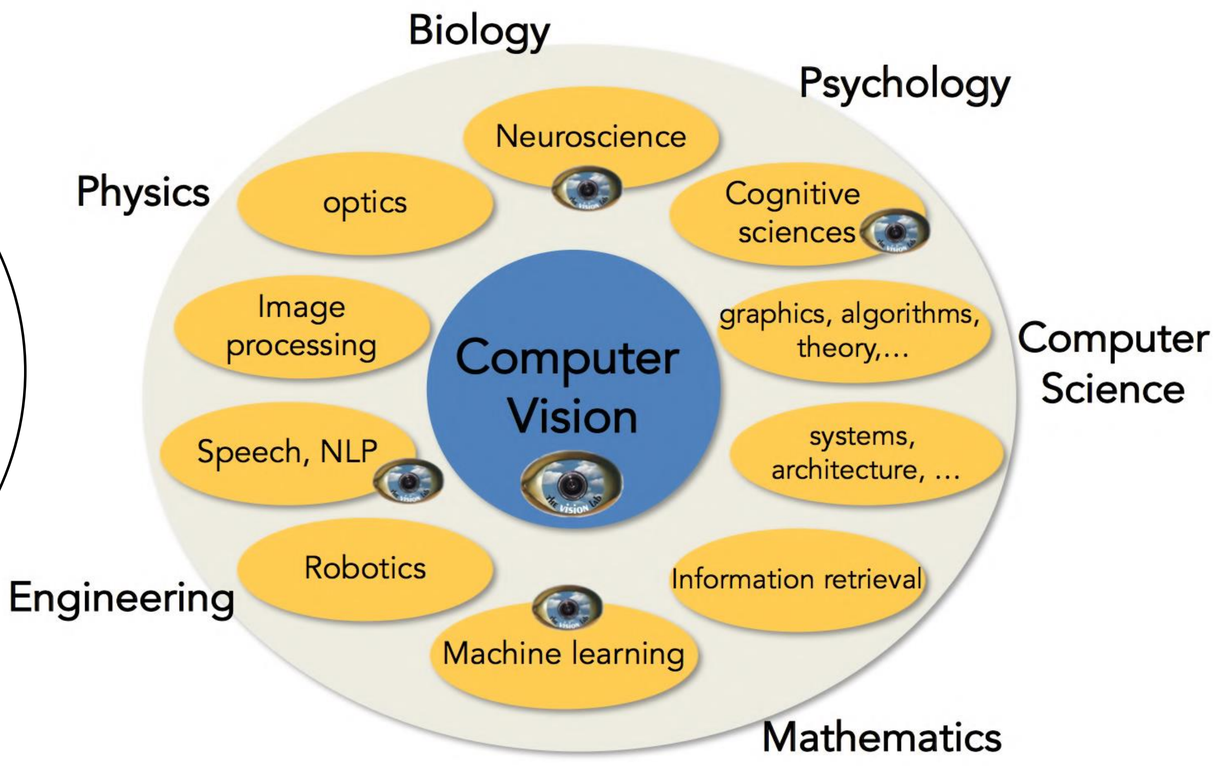
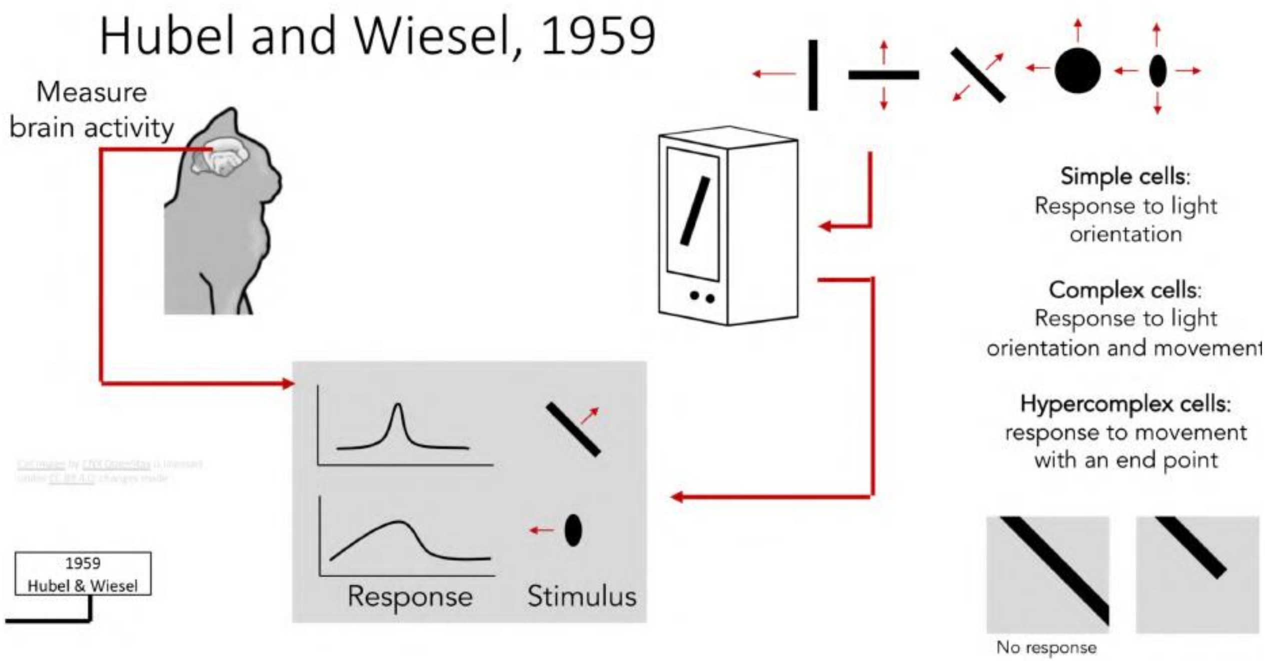
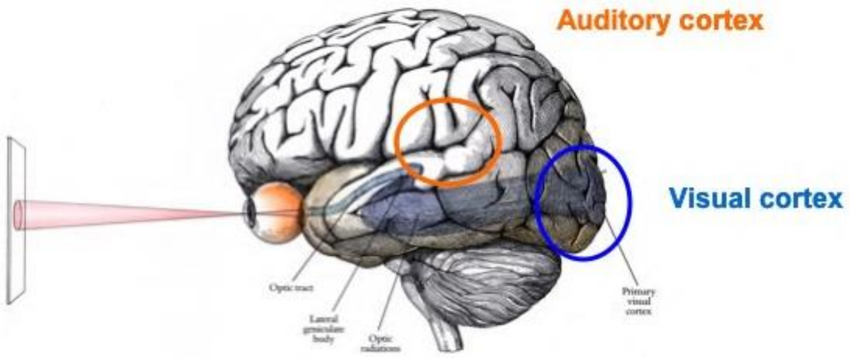


Image Formation In Brain

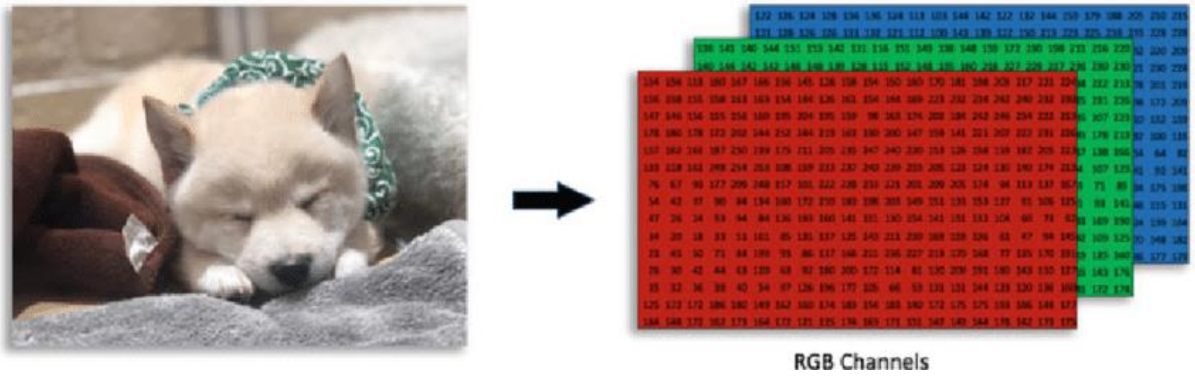
- Hubel and Wiesel received the Nobel Prize in 1981 for their discoveries concerning the visual system.
- Their findings helped identify how early visual processing occurs through hierarchical cell types that respond to progressively complex visual stimuli.





Preliminaries for the Lecture

- Python programming
  - Numpy, matplotlib, opencv
  - Pytorch (For some homeworks)
- Linear Algebra
  - Matrices, SVD, Eigenvalue decompositon etc.
- Probability Theory
  - Probability for Machine Learning

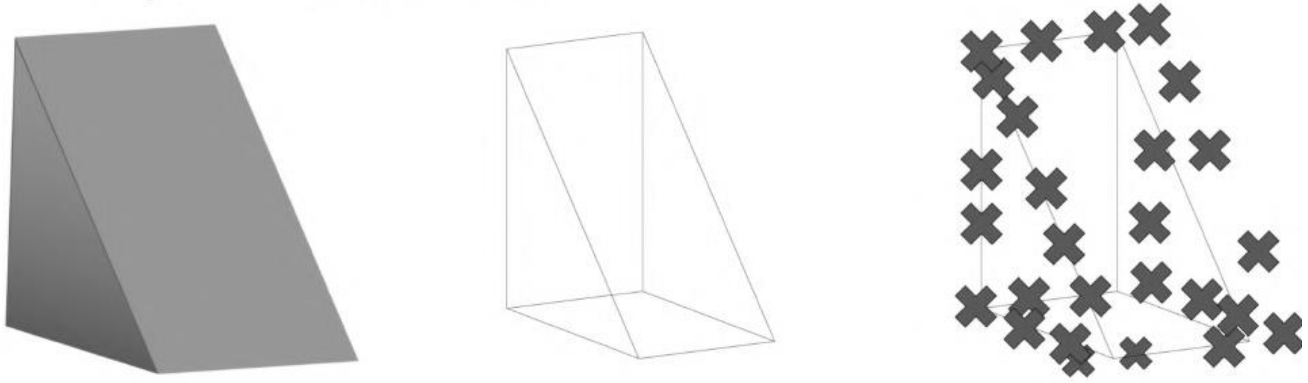


Ng, D., & Feng, M. (2020). Medical Image Recognition: An Explanation and Hands-On Example of Convolutional Networks. *Leveraging Data Science for Global Health*, 263-284.

A bit of history…*

- Block World
  - The first PhD thesis on CV

Larry Roberts, 1963



(a) Original picture (b) Differentiated picture (c) Feature points selected

- The Summer Vision Project in MIT (1966)

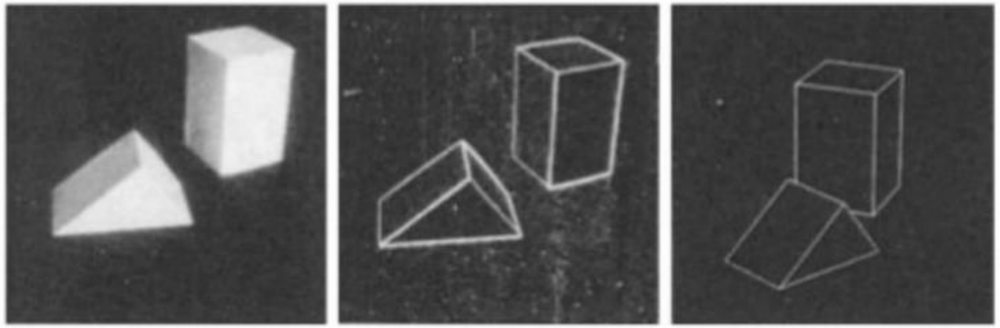
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

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".

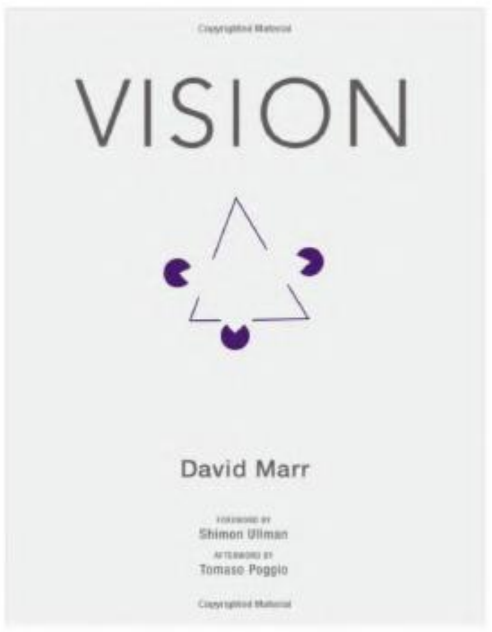


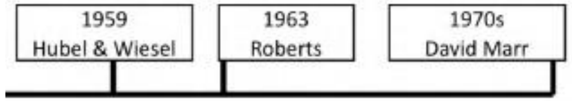
Input image 2x2 gradient operator computed 3D model rendered from new viewpoint

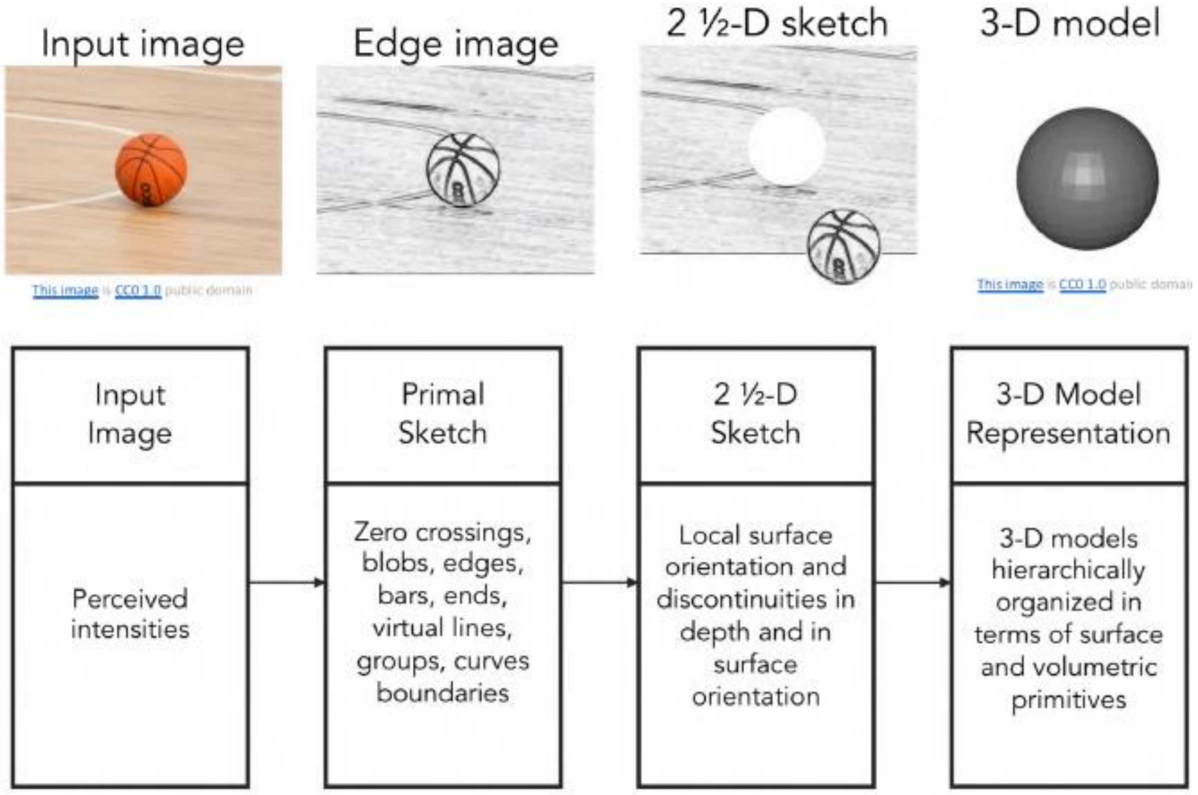
*Borrowed from Prof. Fei Fei Li's slides

A bit of history...

- To obtain a complete 3D representation from a 2D image, multiple steps must be followed
  - inspired from evidence in neuroscience



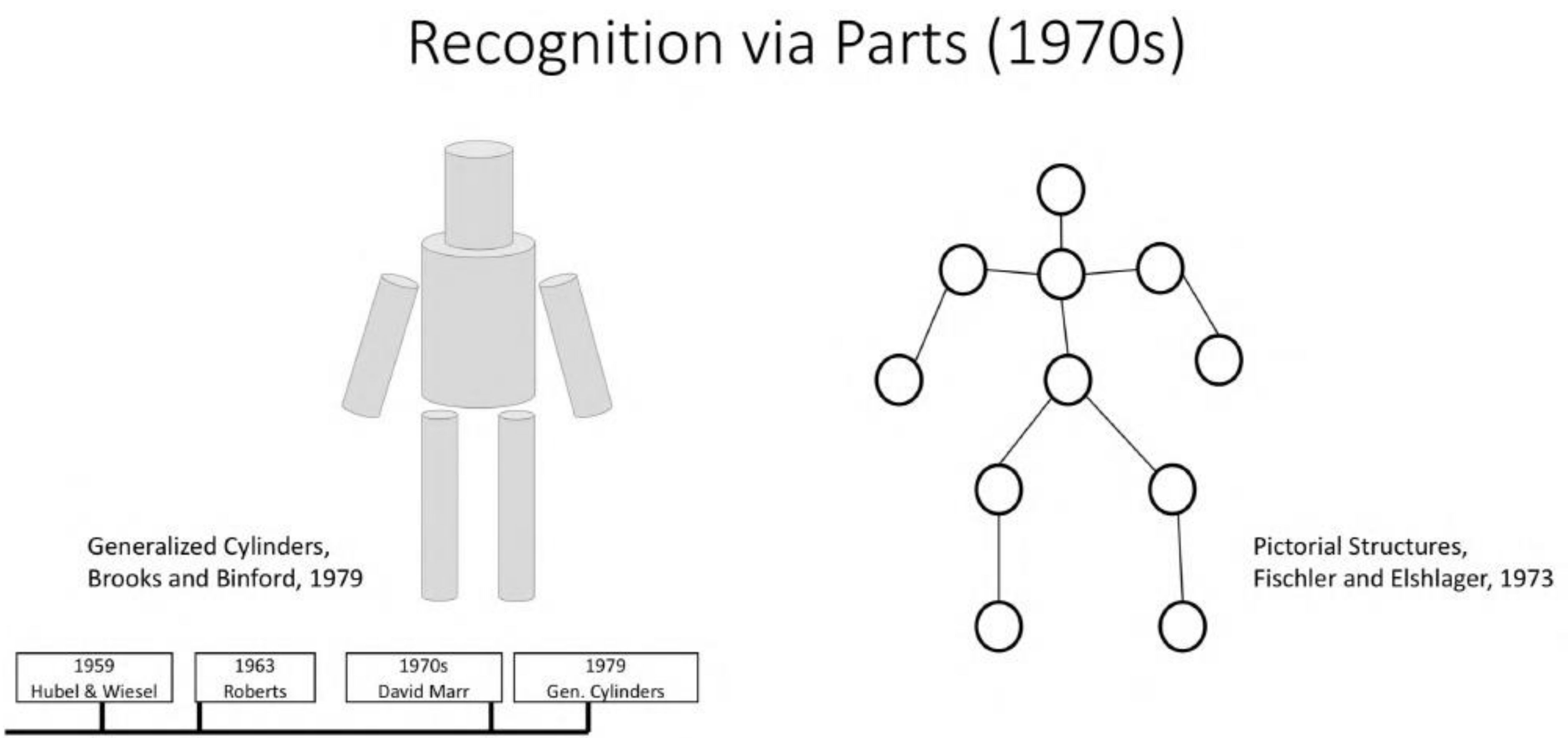




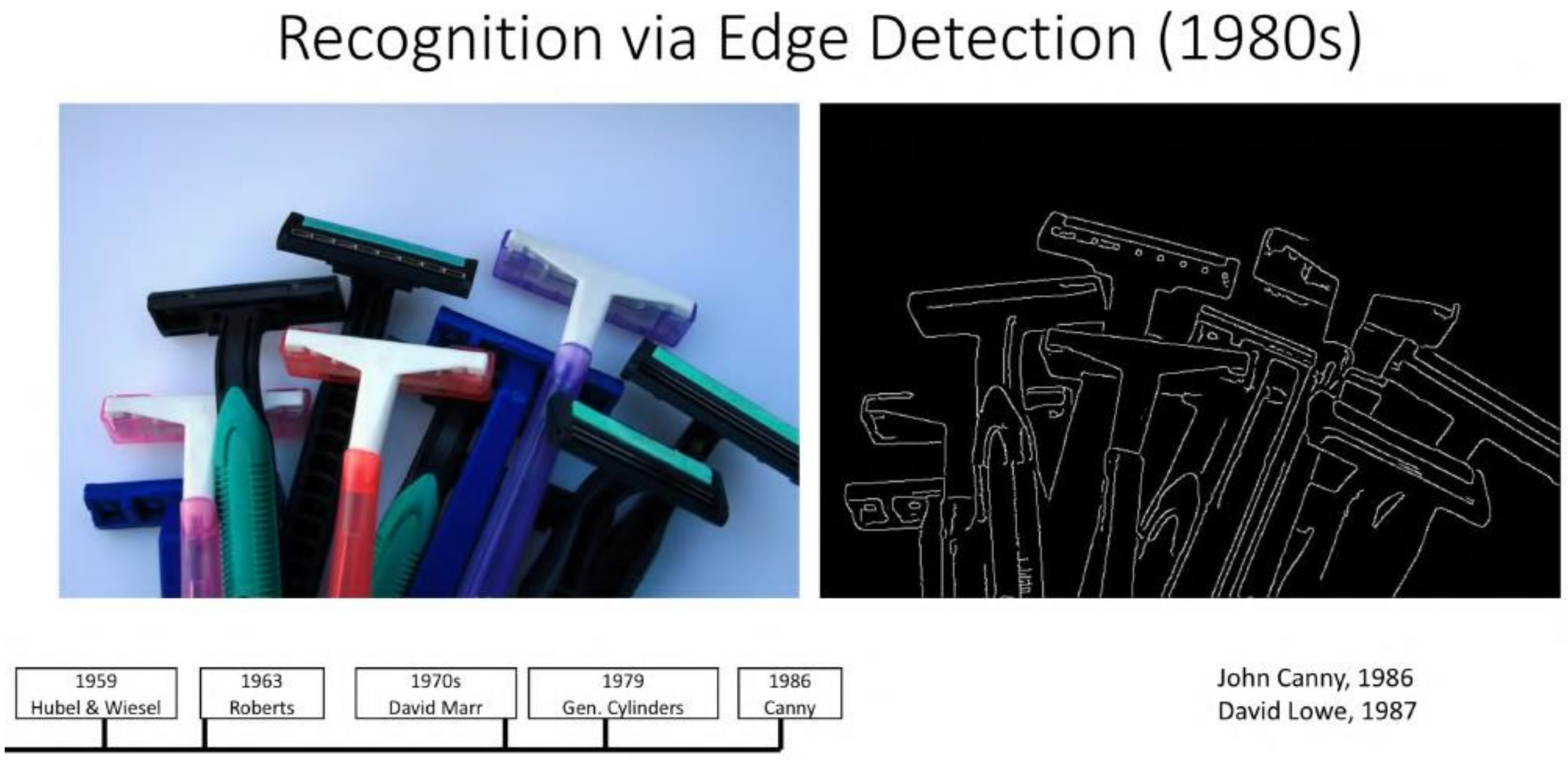
Stages of Visual Representation, David Marr, 1970s

A bit of history…

- Every geometric object consist of simple geometric primitives.

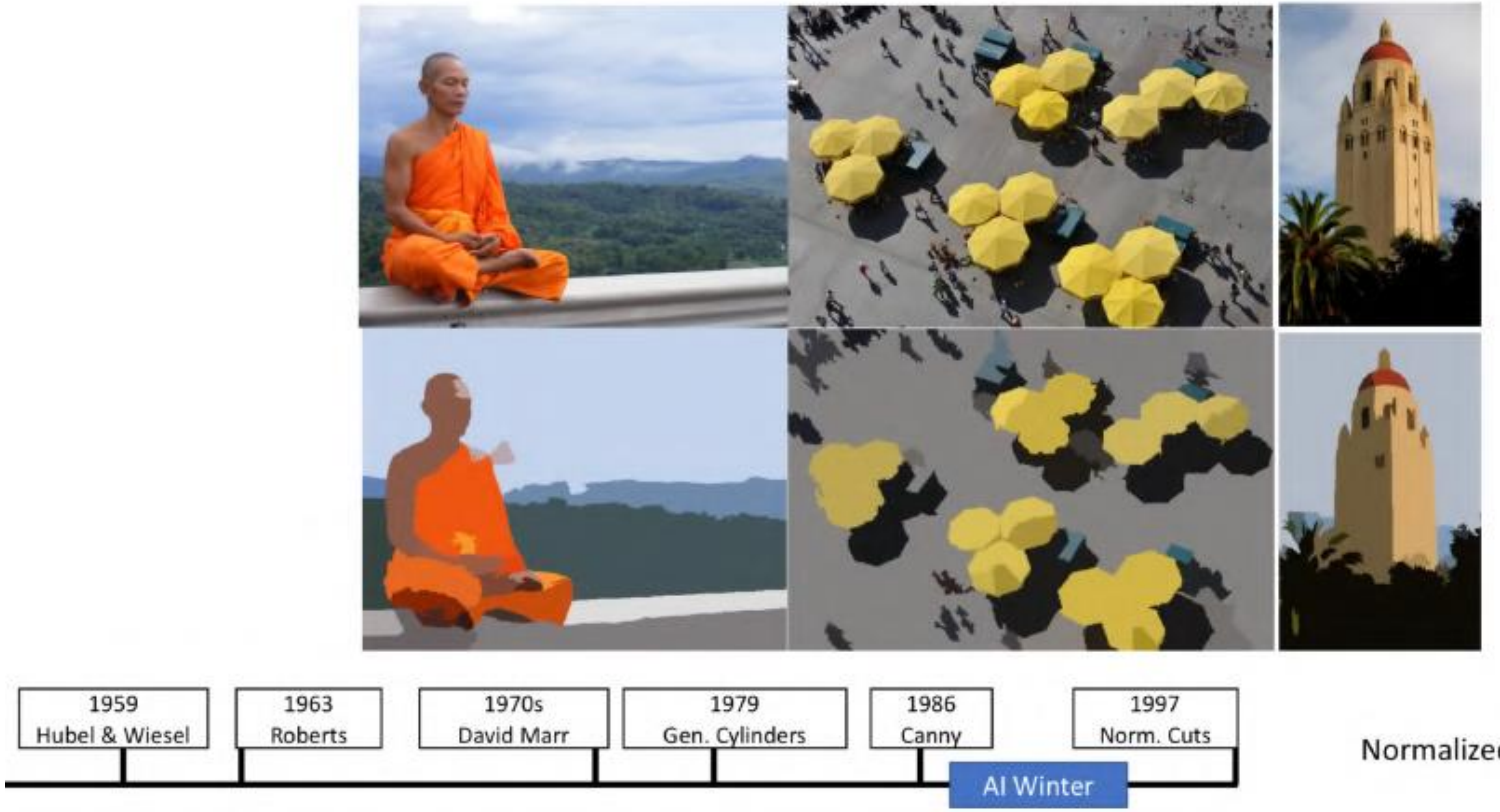


A bit of history...



A bit of history...

Recognition via Grouping (1990s)



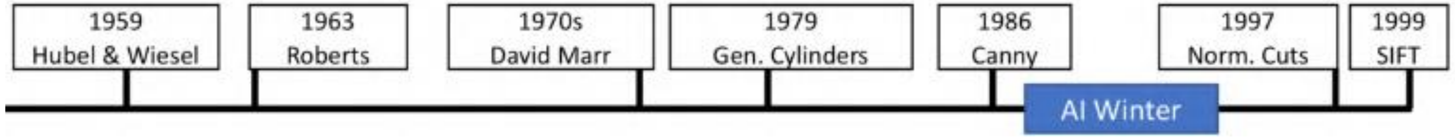
Normalized Cuts, Shi and Malik, 1997

A bit of history…

Recognition via Matching (2000s)

Image is public domain

Image is public domain



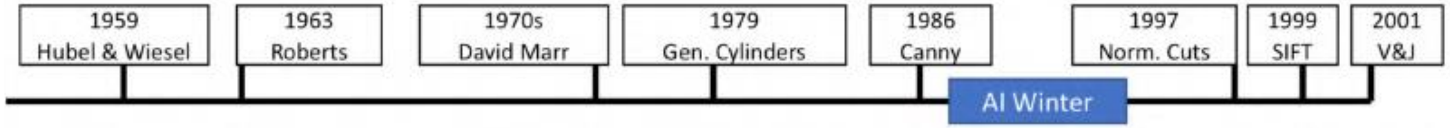
SIFT, David Lowe, 1999

A bit of history…

Face Detection

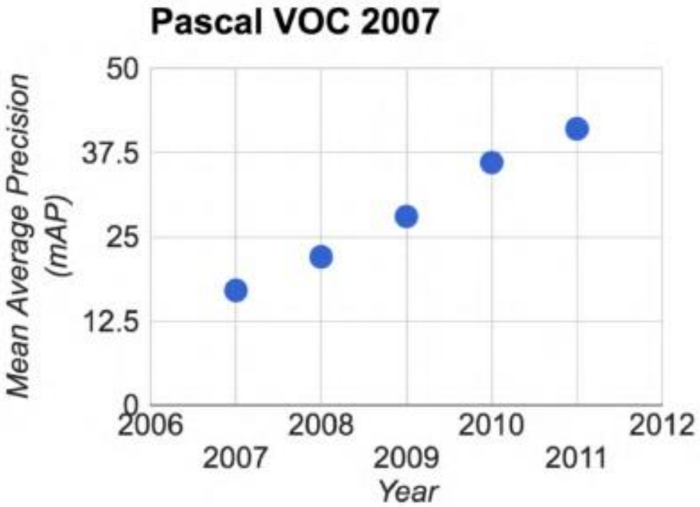
Viola and Jones, 2001

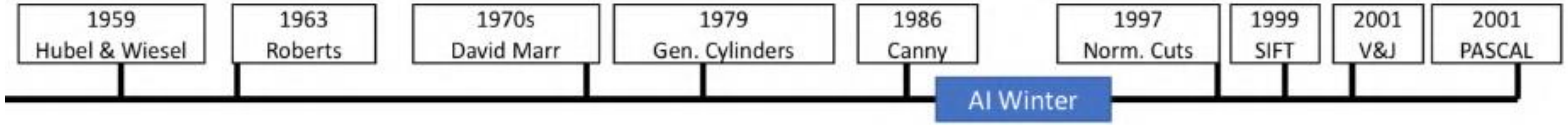
One of the first successful applications of machine learning to vision



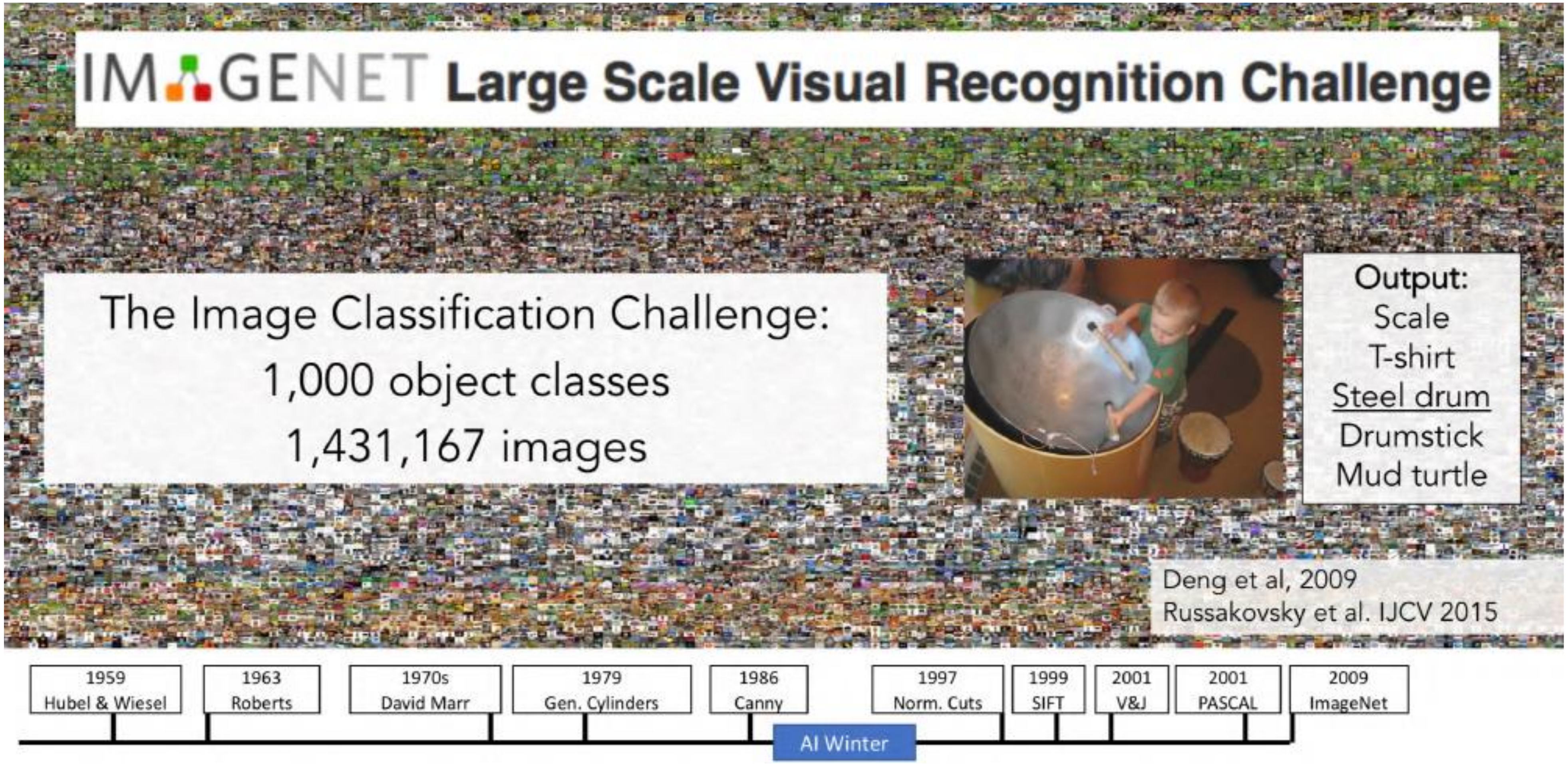
A bit of history…

PASCAL Visual Object Challenge





A bit of history…

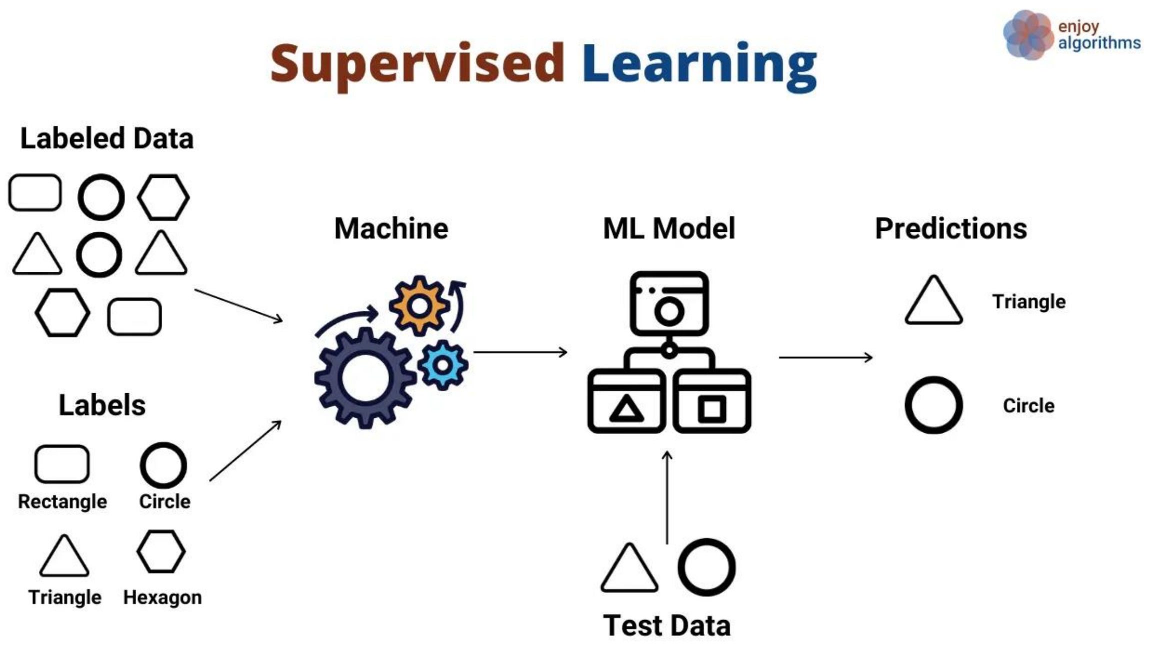


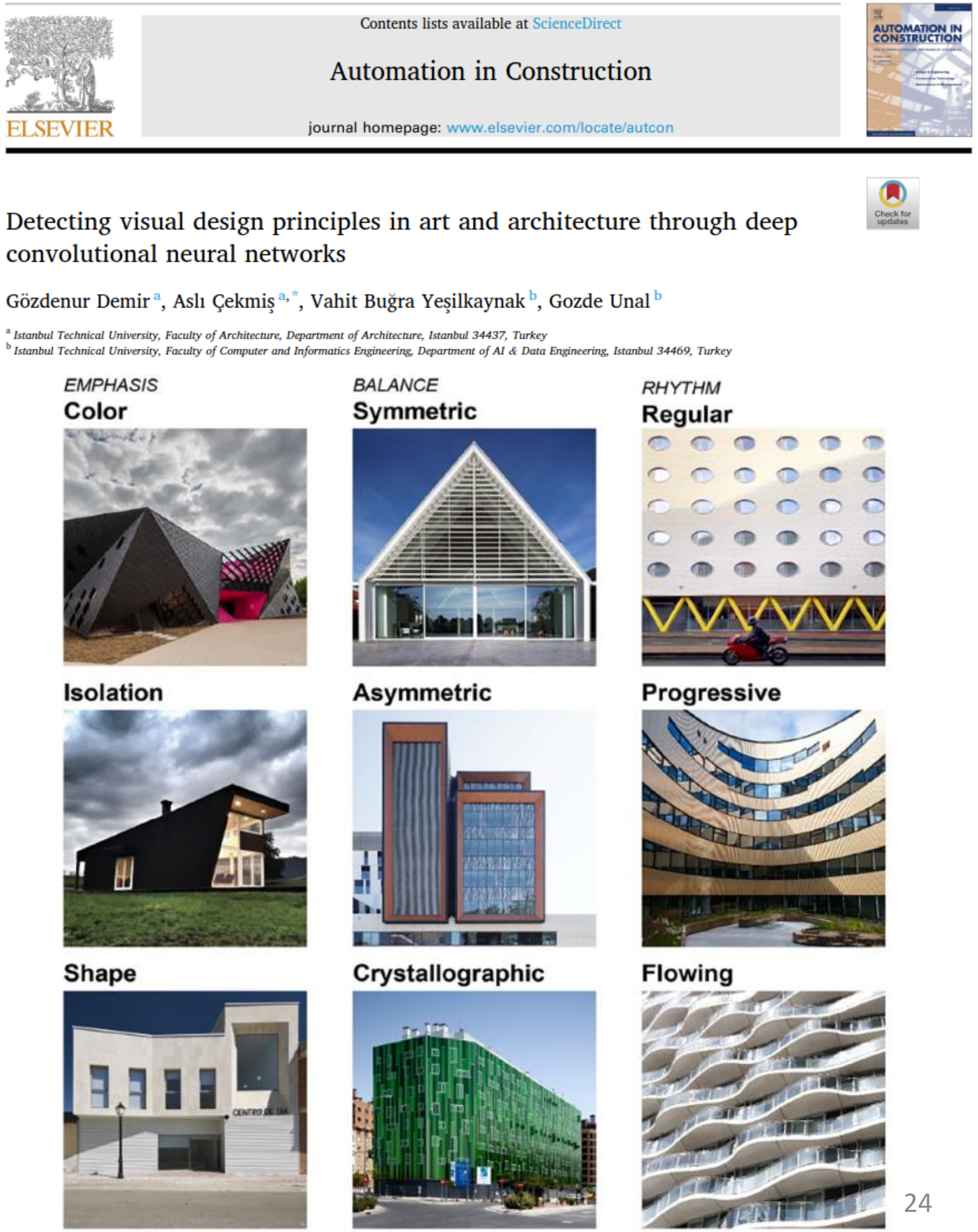
Machine Learning Objectives and CV

- Supervised Learning
  - Input: Training data + Example training outputs
- Unsupervised Learning
  - Input: Training data (independent of outputs)
- Semi-Supervised Learning
  - Input: Training data + Some example training outputs
- Reinforcement Learning
  - Input: Action processes with reward/punishment outcomes

Supervised Learning

Classification:
Determining the class for given examples





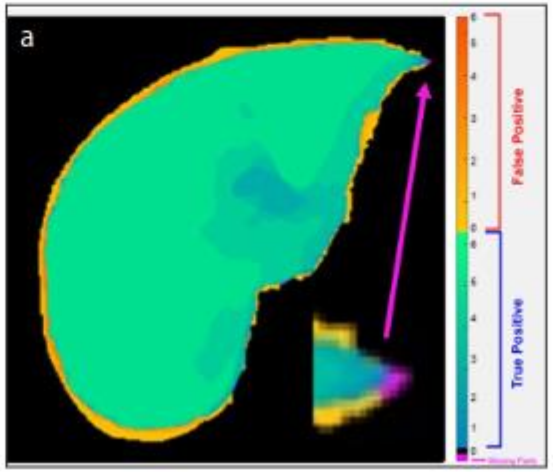
Supervised Learning

Segmentation:
Pixelwise classification

DIR
Diagn Interv Radiol 2020; 26:11–21

ABDOMINAL IMAGING
ORIGINAL ARTICLE

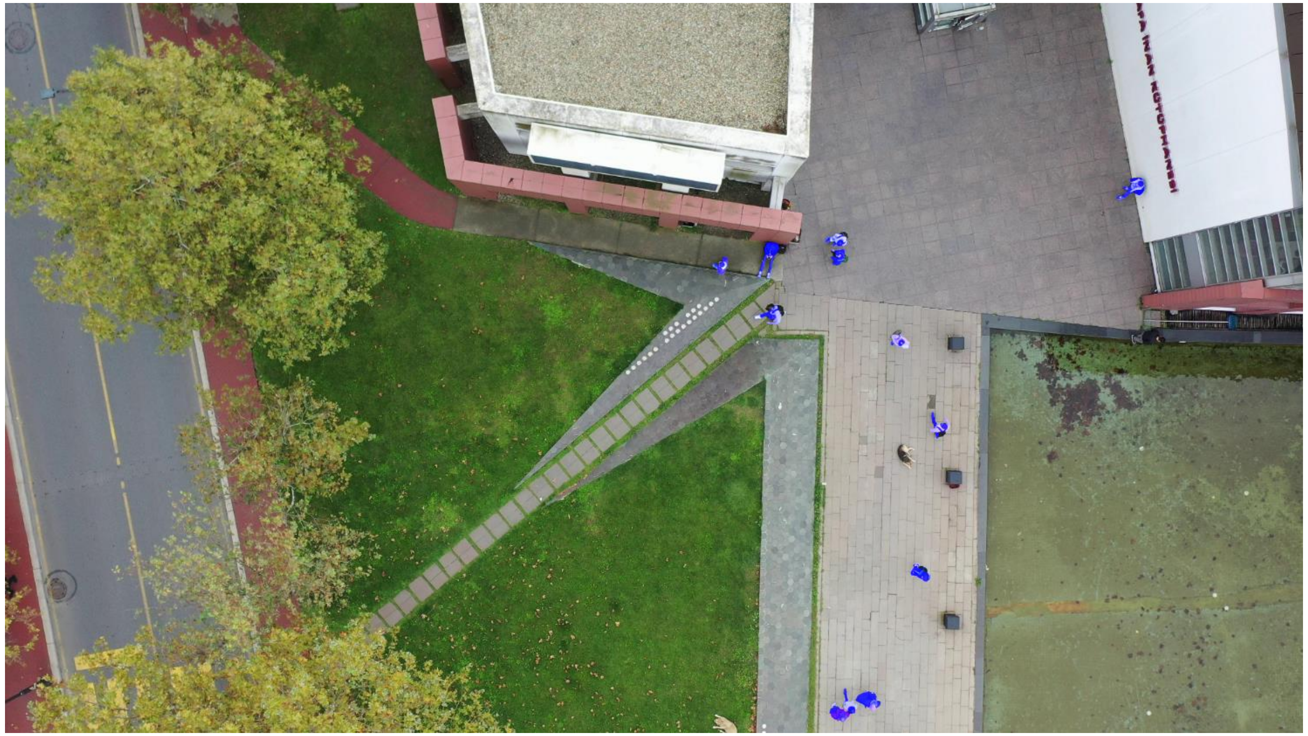
Comparison of semi-automatic and deep learning-based automatic methods for liver segmentation in living liver transplant donors



P2-23
18th International Conference on Machine Vision Applications (MVA)
Hamamatsu, Japan, July 23-25, 2023.

TinyPedSeg: A Tiny Pedestrian Segmentation Benchmark for Top-Down Drone Images

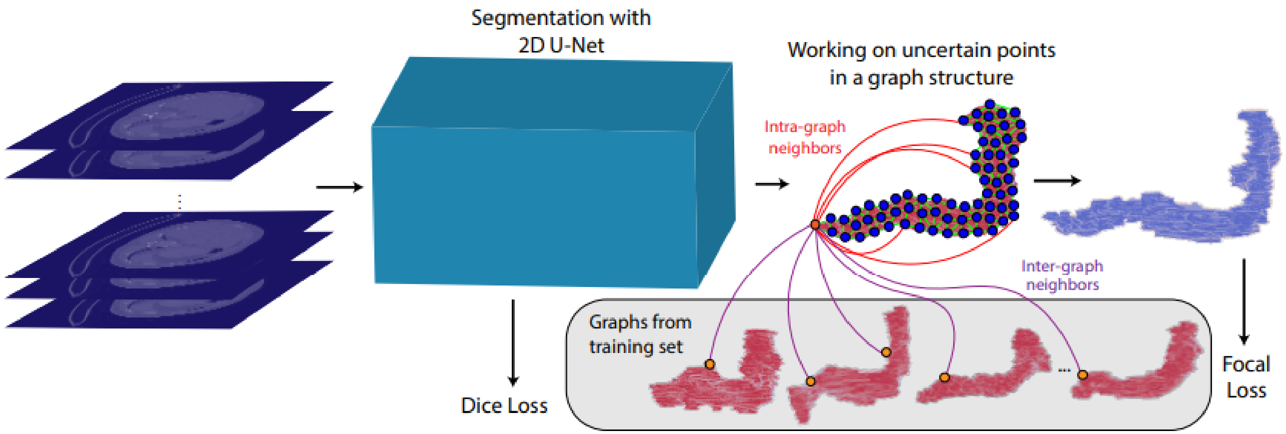
Yusuf H. Sahin, Elvin Abdinli, M. Arda Aydın, Gozde Unal
Istanbul Technical University, Faculty of Computer and Informatics
{sahinyu, abdinli18, aydinmu19, gozde.unal}@itu.edu.tr

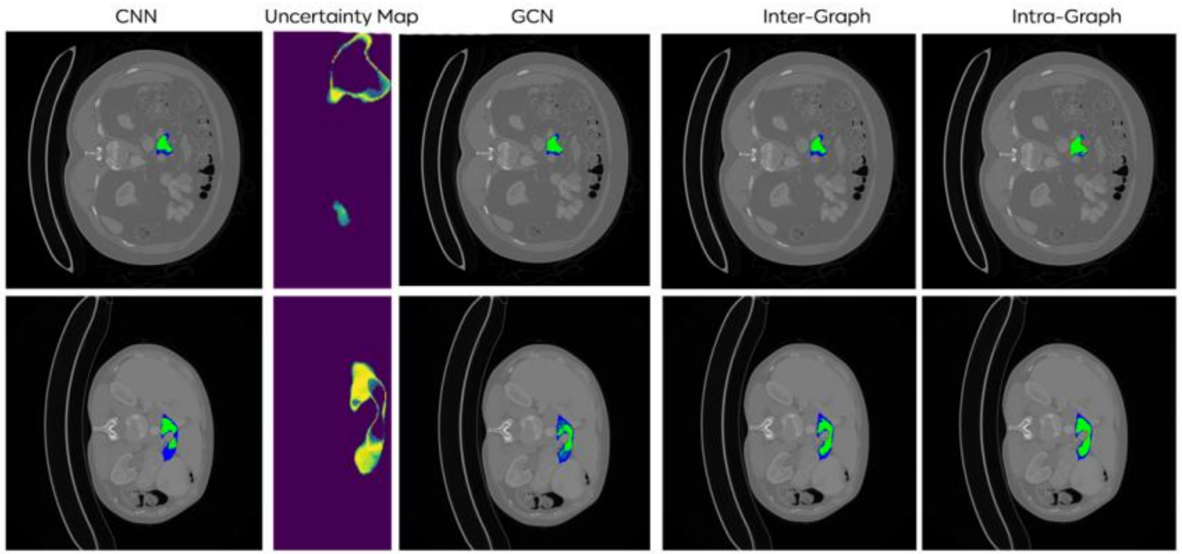


Semi-Supervised Learning

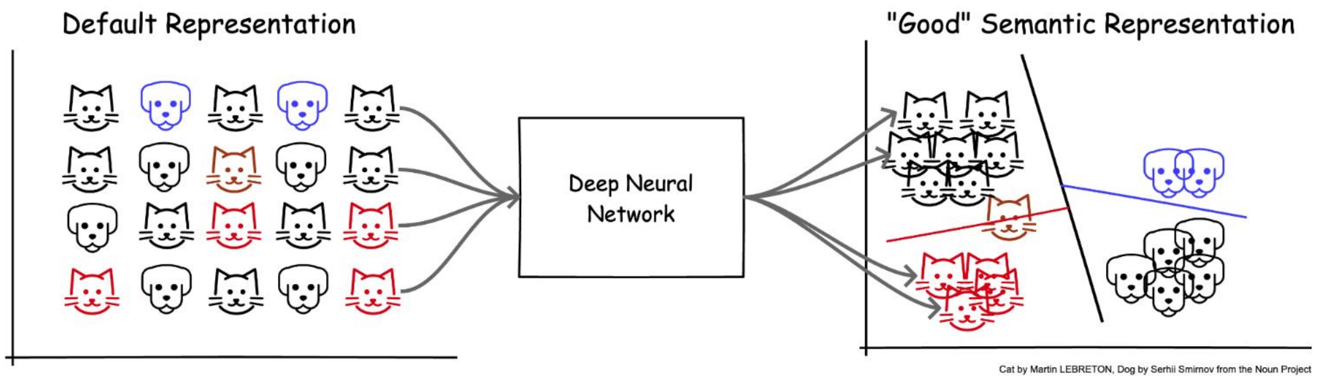
Uncertainty-Based Dynamic Graph Neighborhoods For Medical Segmentation

Ufuk Demir*[1], Atahan Ozer*[1], Yusuf H. Sahin[1], and Gozde Unal[2]





Unsupervised Learning



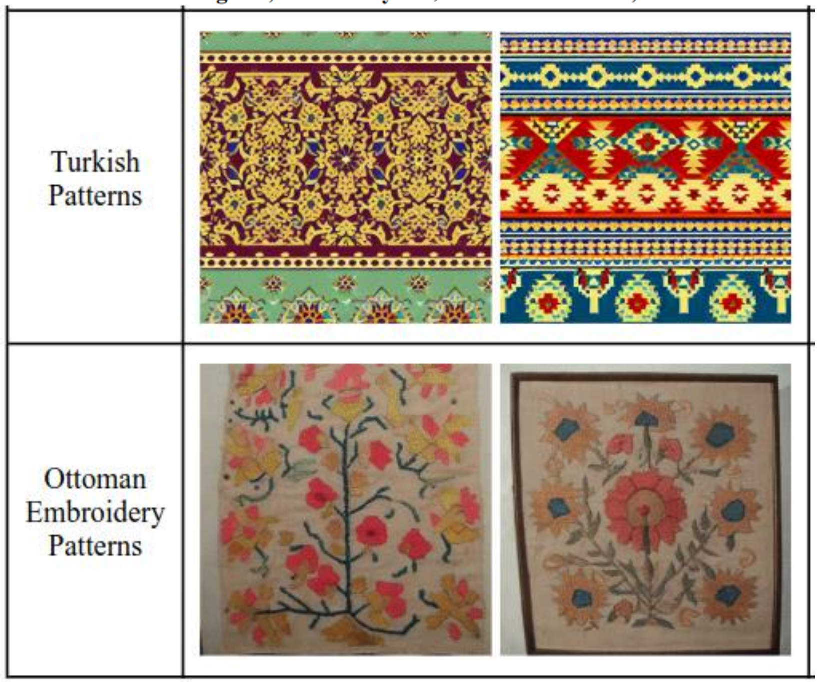
Cat by Martin LEBRETON, Dog by Serhii Smirnov from the Noun Project

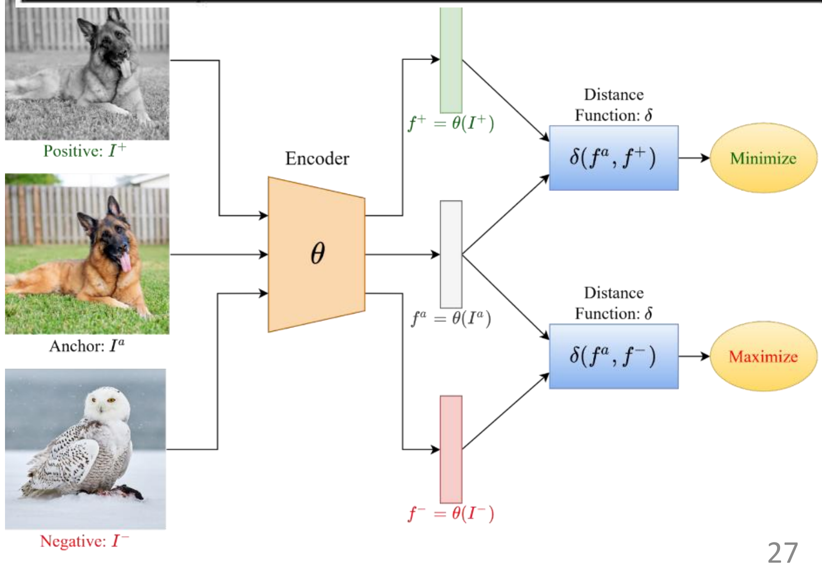
Rethinking CNN-Based Pansharpening: Guided Colorization of Panchromatic Images via GANs

Furkan Ozcelik, Ugur Alganci, Elif Sertel, and Gozde Unal

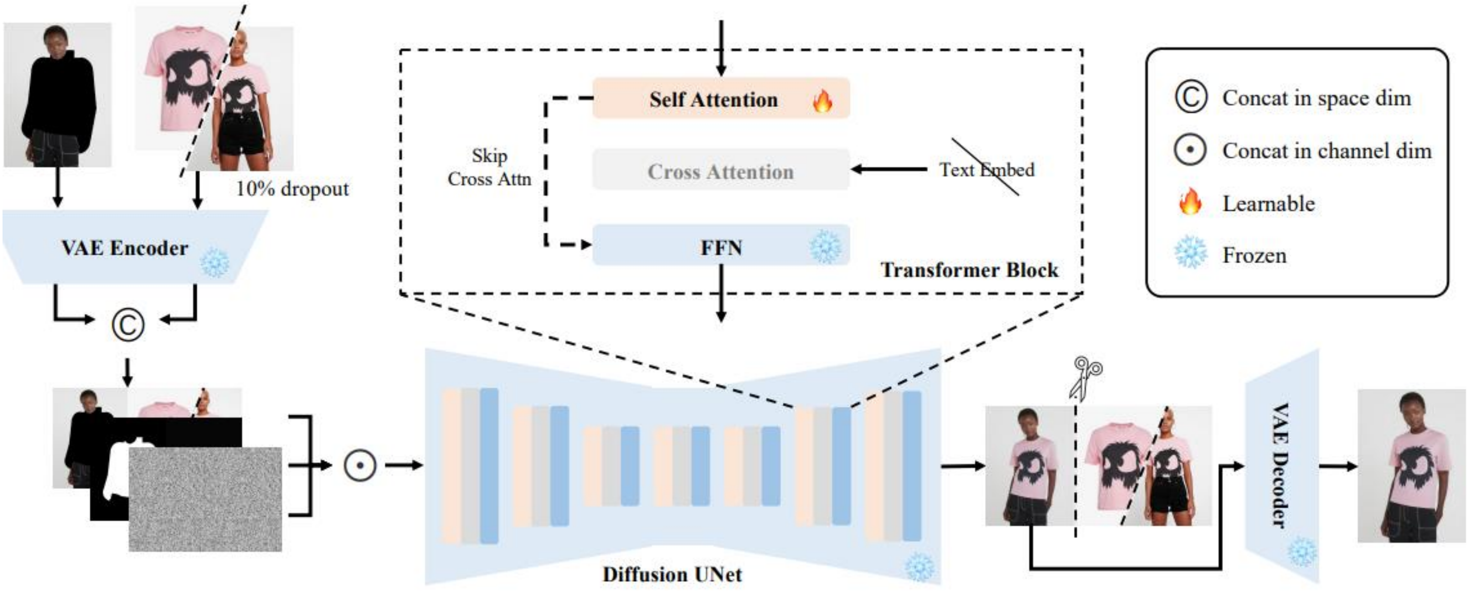
Textile Pattern Generation Using Diffusion Models

Halil Faruk Karagoz[1,*], Gulcin Baykal[1], Irem Arikan Eksi[2], Gozde Unal[3]

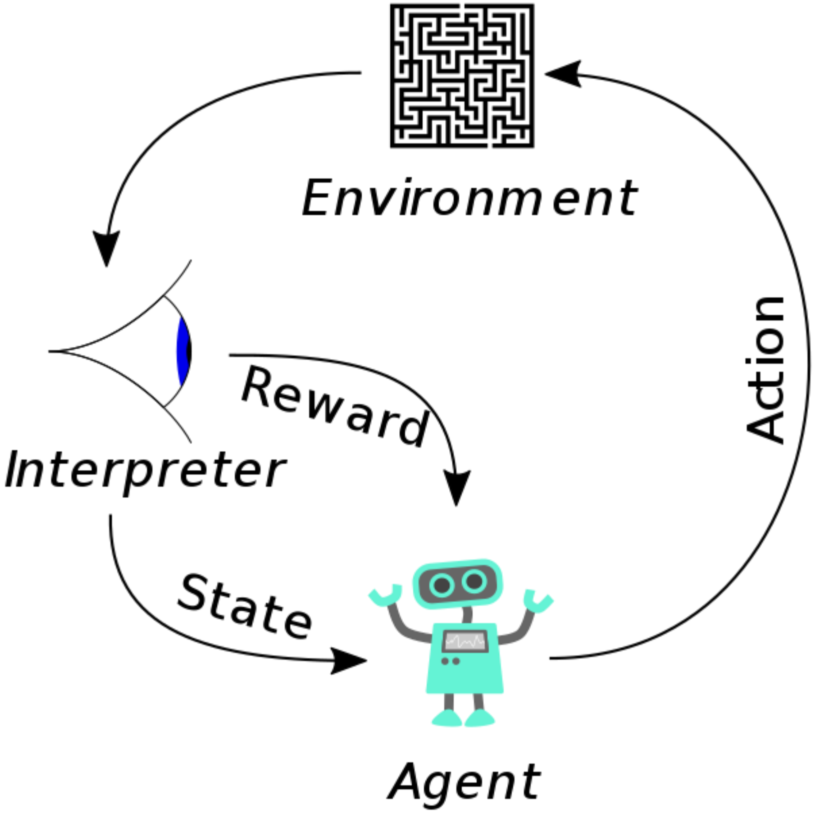




Unsupervised Learning



Reinforcement Learning



https://www.youtube.com/watch?v=0xo1Ldx3L5Q