AMRITA VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY

19CSE437
DEEP LEARNING FOR COMPUTER VISION
L-T-P-C: 2-0-3-3

Click to add text

Introduction to

- Artificial Intelligence
- Machine Learning
- Computer Vision

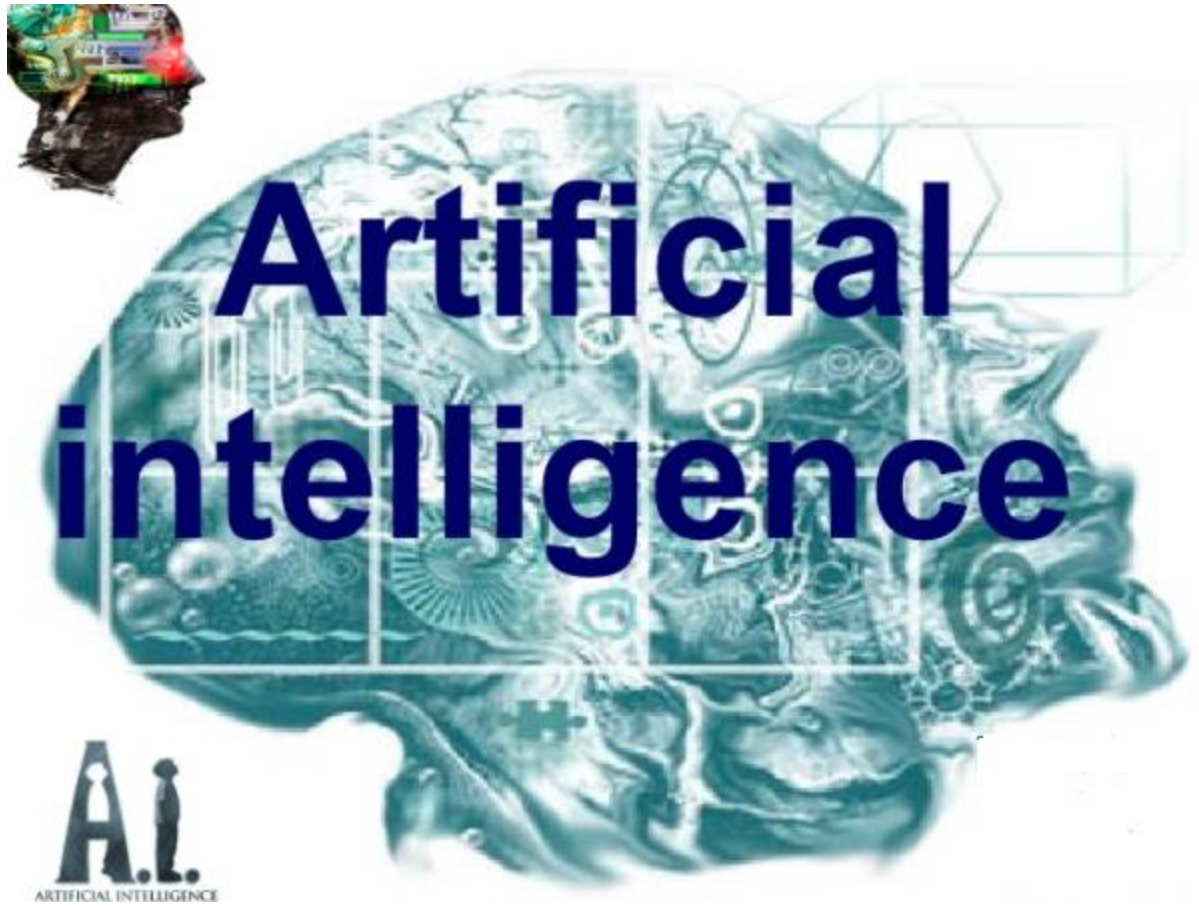
Artificial Intelligence

AI refers to 'Artificial Intelligence'

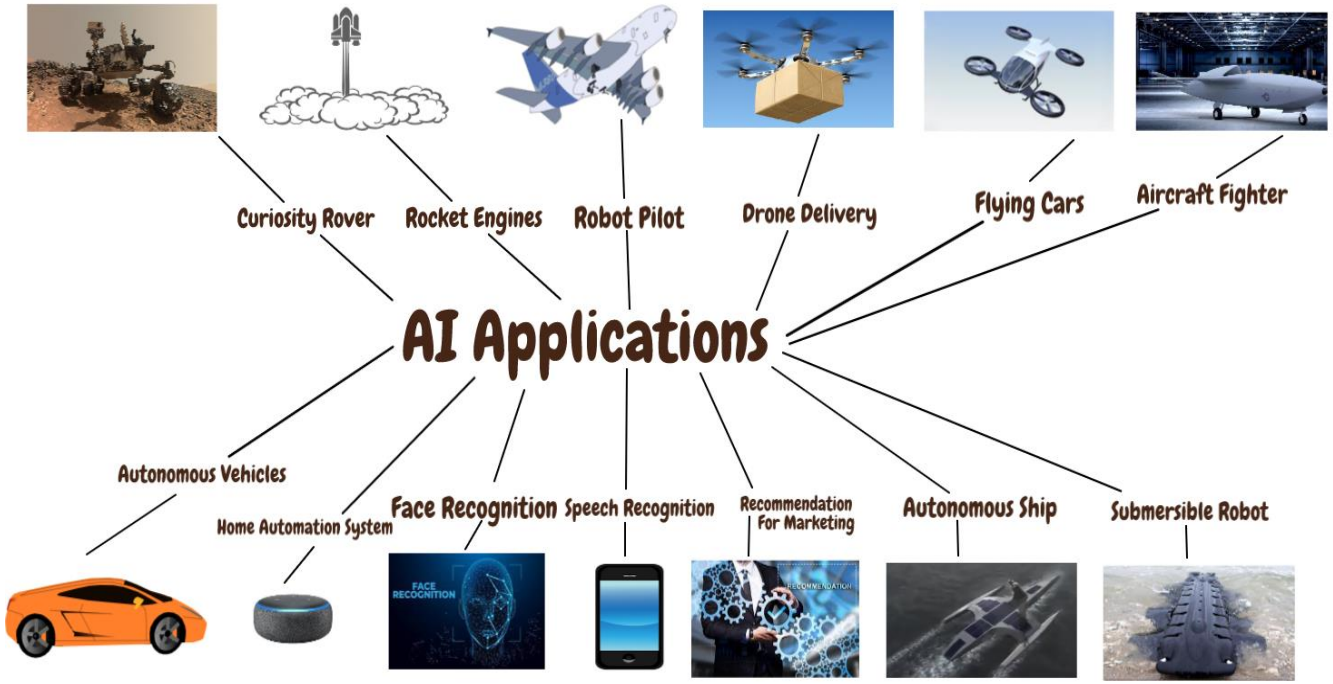
which means making machines capable of performing quick tasks like human beings.

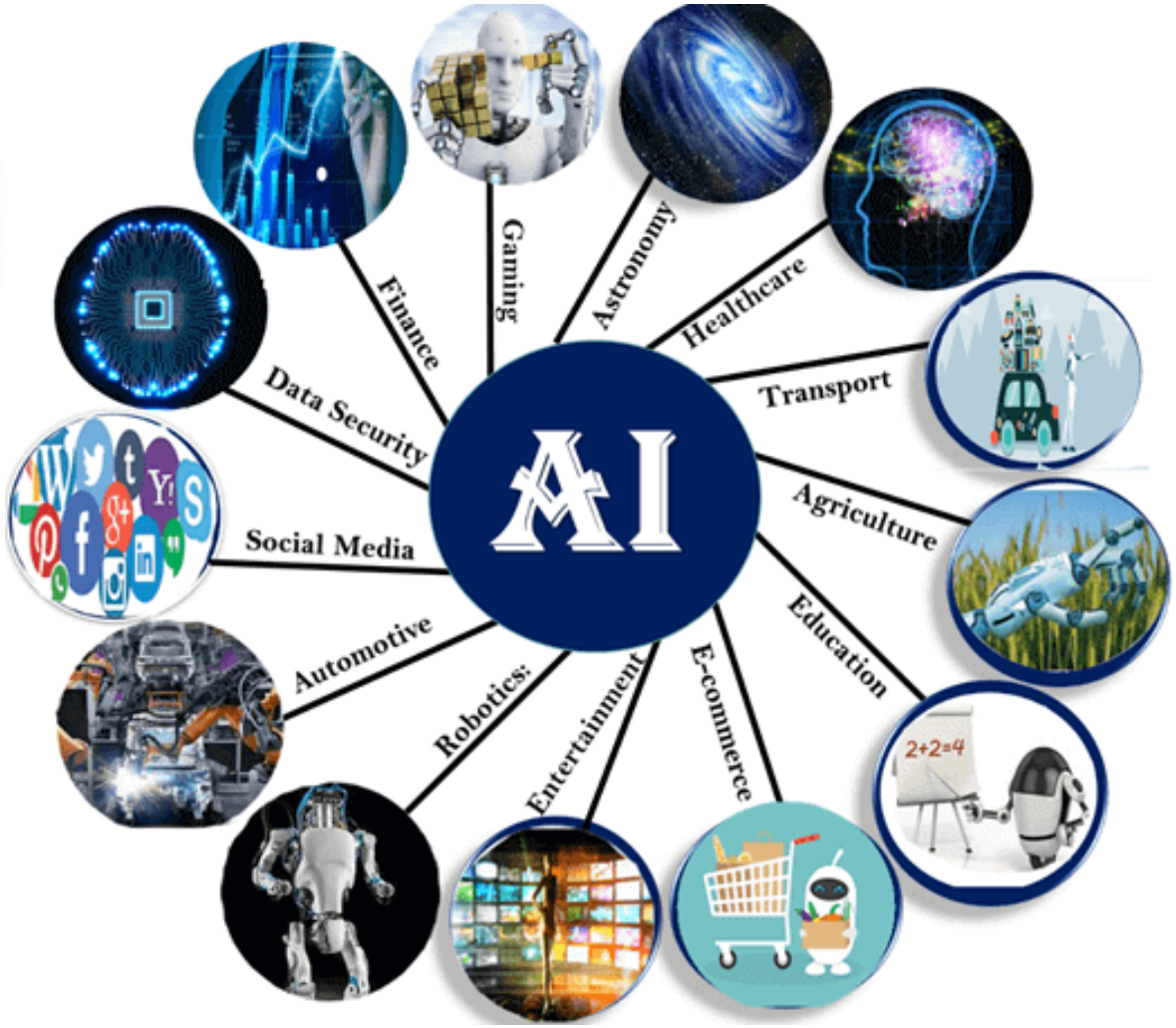
AI performs automated tasks using intelligence.

It has two key components –
- **Automation**
- **Intelligence**

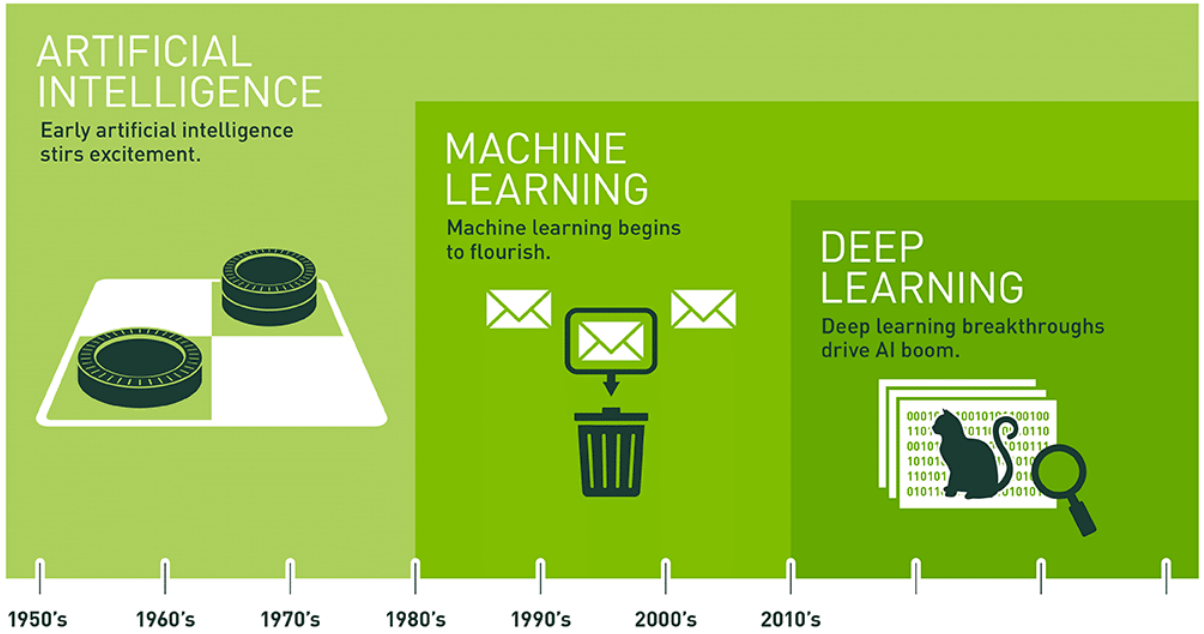


AI Application Domains

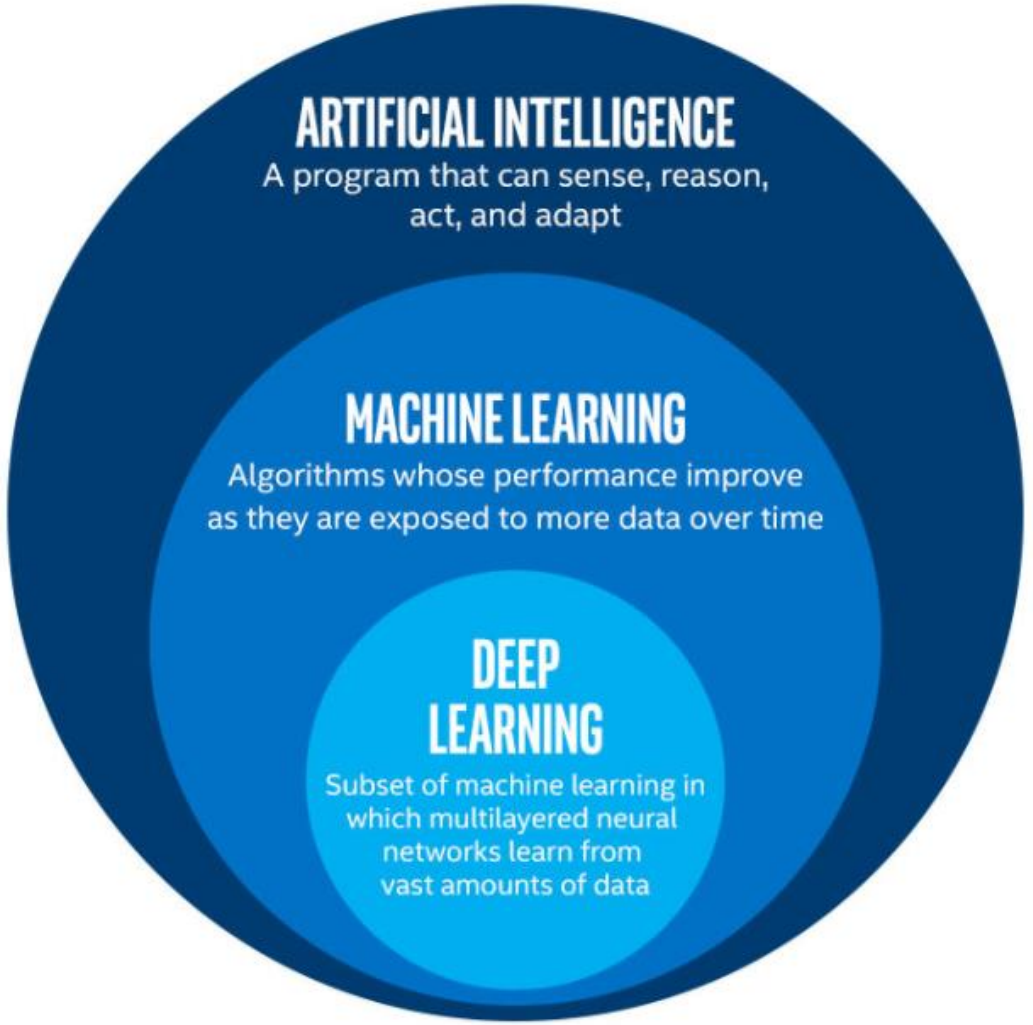




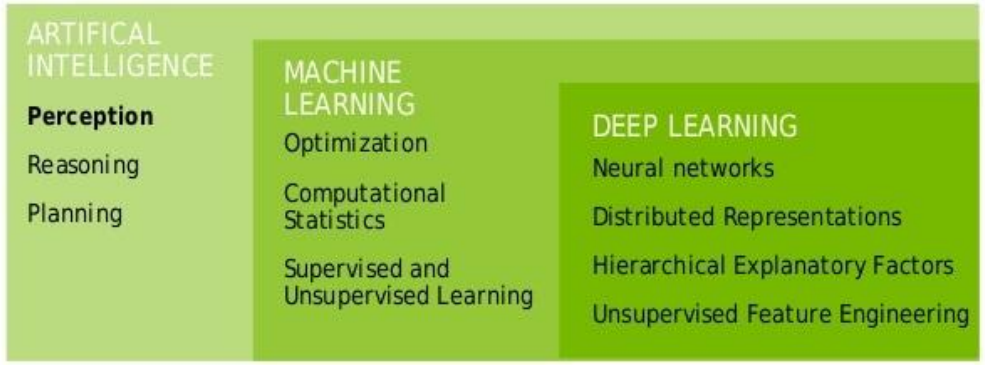
AI applications few eg: advanced web search engines, recommendation systems (used by YouTube, Amazon and Netflix), understanding human speech (such as Siri or Alexa), self-driving cars (e.g. Tesla), and competing at the highest level in strategic game systems (such as chess and Go) As machines become increasingly capable, tasks considered to require "intelligence" are often removed from the definition of AI, a phenomenon known as the AI effect.



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep l

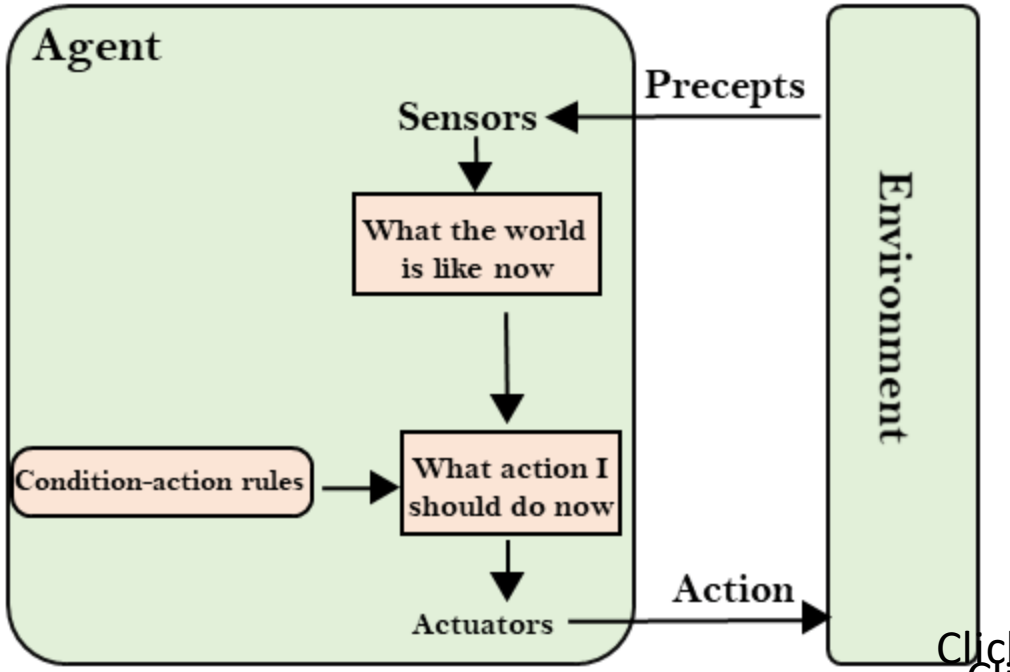


WHAT IS DEEP LEARNING?



Intelligent agents must be able to set goals and achieve them

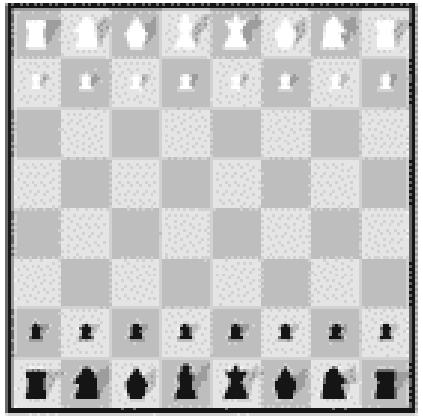
Intelligent Agent



Intelligent Agent is a system (ie: software program) which observes the world through sensors and acts upon an environment using actuators. It directs it's activity towards achieving goals. Intelligent agents may also learn or use knowledge to achieve their goals.

Click to add text

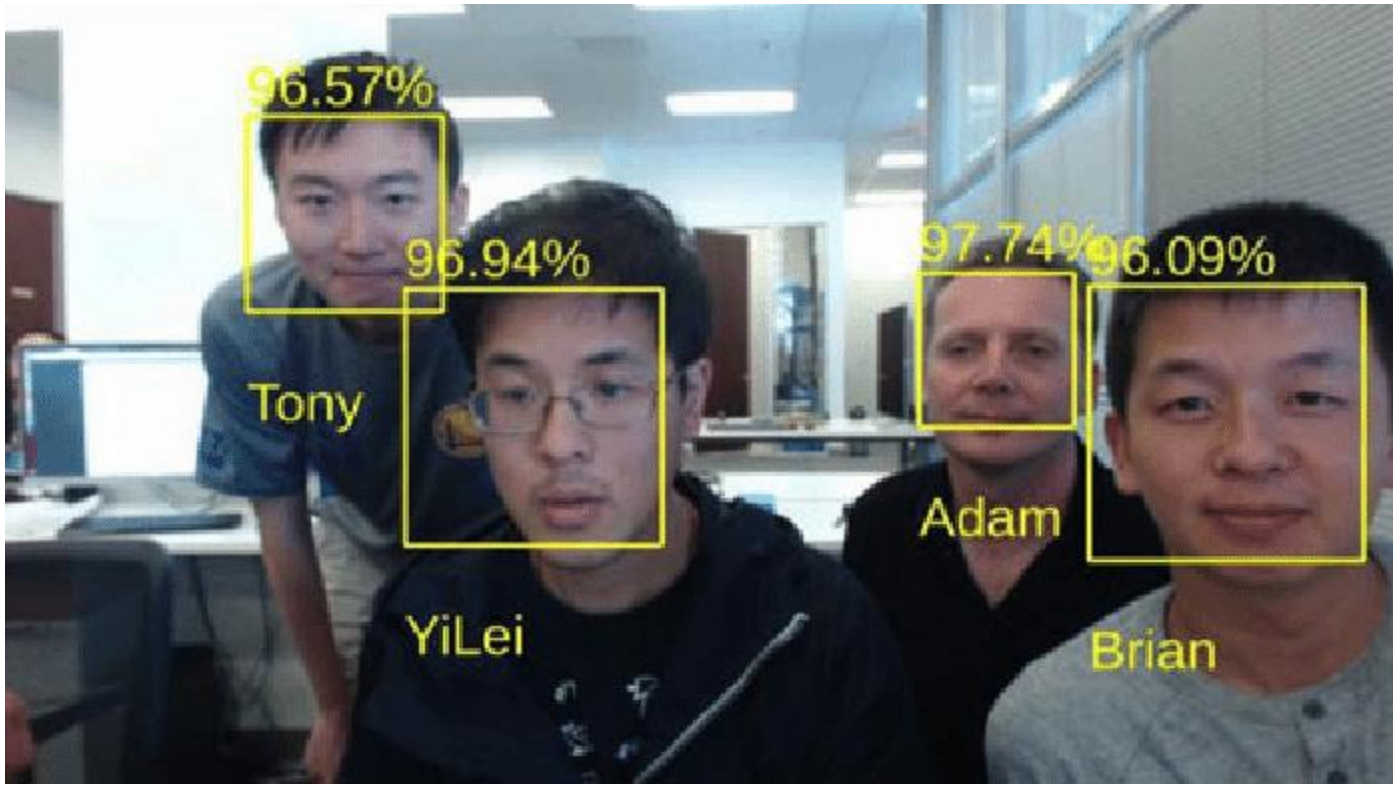
Click to add text



Central AI problems:

1.Reasoning
2.Knowledge
3.Planning: Make prediction about their actions,
4.Learning
5.Perception
6.Natural Language Processing (NLP)

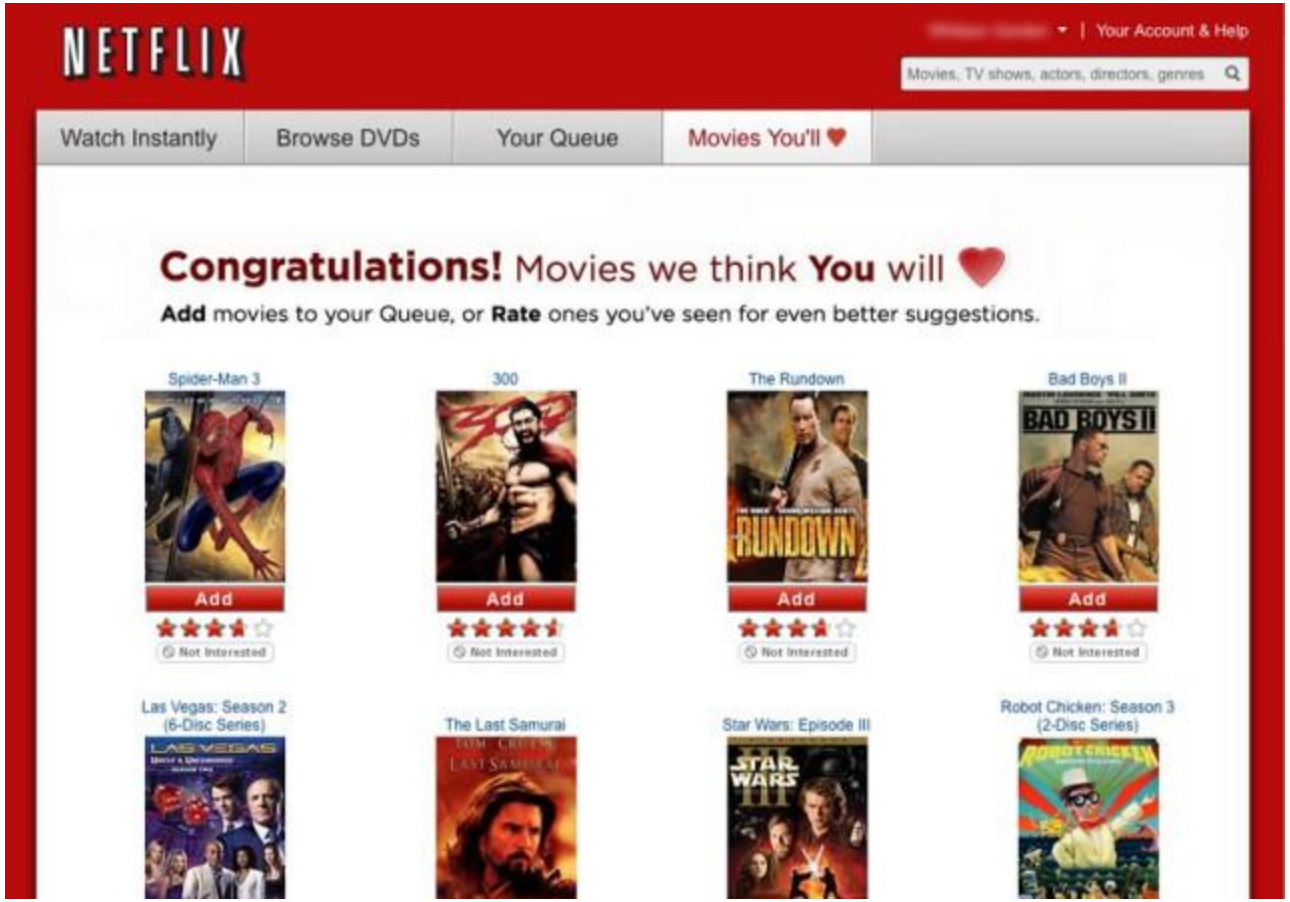
An example- Computer Vision Problem



Can a machine recognise people just like human do ?

Machine Learning Applications -Recommendation Systems

Netflix, YouTube, Tinder, and Amazon are all examples of recommender systems in use. The systems entice users with relevant suggestions based on the choices they make.



YouTube

amazon.com

Recommended for You

Amazon.com has new recommendations for you based on items you purchased or told us you own.

The Little Big Things: 163 Ways to Pursue EXCELLENCE | Fascinate: Your 7 Triggers to Persuasion and Captivation | Sherlock Holmes [Blu-ray] | Alice in Wonderland [Blu-ray]

Machine Learning Applications -Virtual Assistant

An **intelligent virtual assistant** (**IVA**) or **intelligent personal assistant** (**IPA**) is a software agent that can perform tasks or services for an individual based on commands or questions

Amazon Alexa
Google assistant
Siri (Apple)
Cortana (Microsoft)

Intelligent Personal Assistants, answer the queries and perform actions via voice commands using a natural language user interface.

Self Driving car

A **self-driving car** (sometimes called an **autonomous car** or driverless **car**) is a **vehicle** that uses a combination of sensors, cameras, radar and **artificial intelligence** (**AI**) to travel between destinations without a human operator.

IoT Sensors

IoT Connectivity

Image Processing

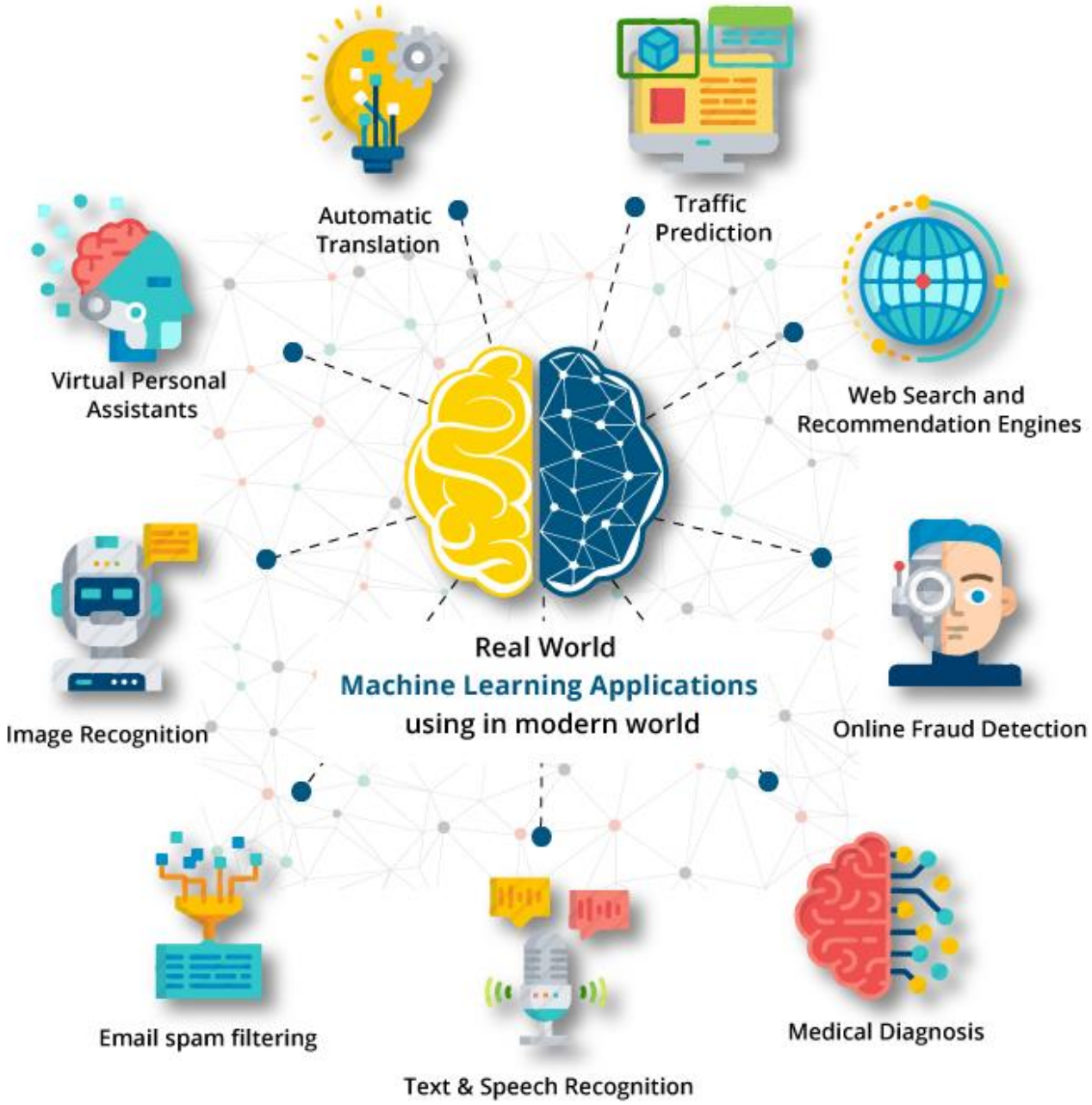
Video processing

Graphics -High-resolution maps of the world,

Learning machine learning data models

Robotics - avoiding obstacles

Mechanics- Motor Gear


Automatic Translation
Traffic Prediction
Virtual Personal Assistants
Web Search and Recommendation Engines
Real World
Machine Learning Applications
using in modern world
Image Recognition
Online Fraud Detection
Email spam filtering
Text & Speech Recognition
Medical Diagnosis

Introduction to

Machine Learning

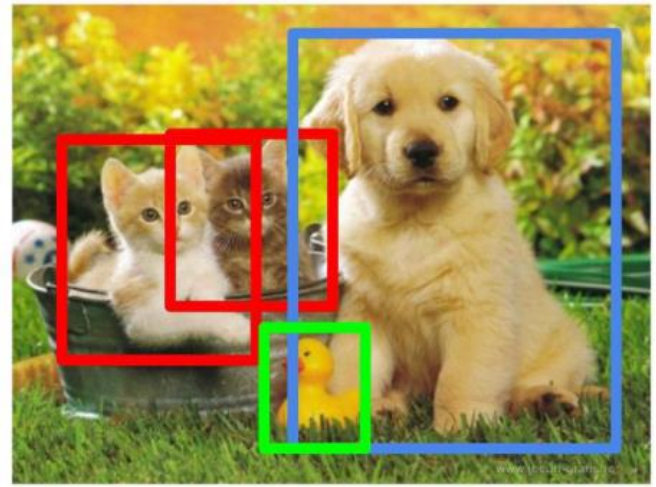
Classification

CAT

Classification + Localization

CAT

Object Detection



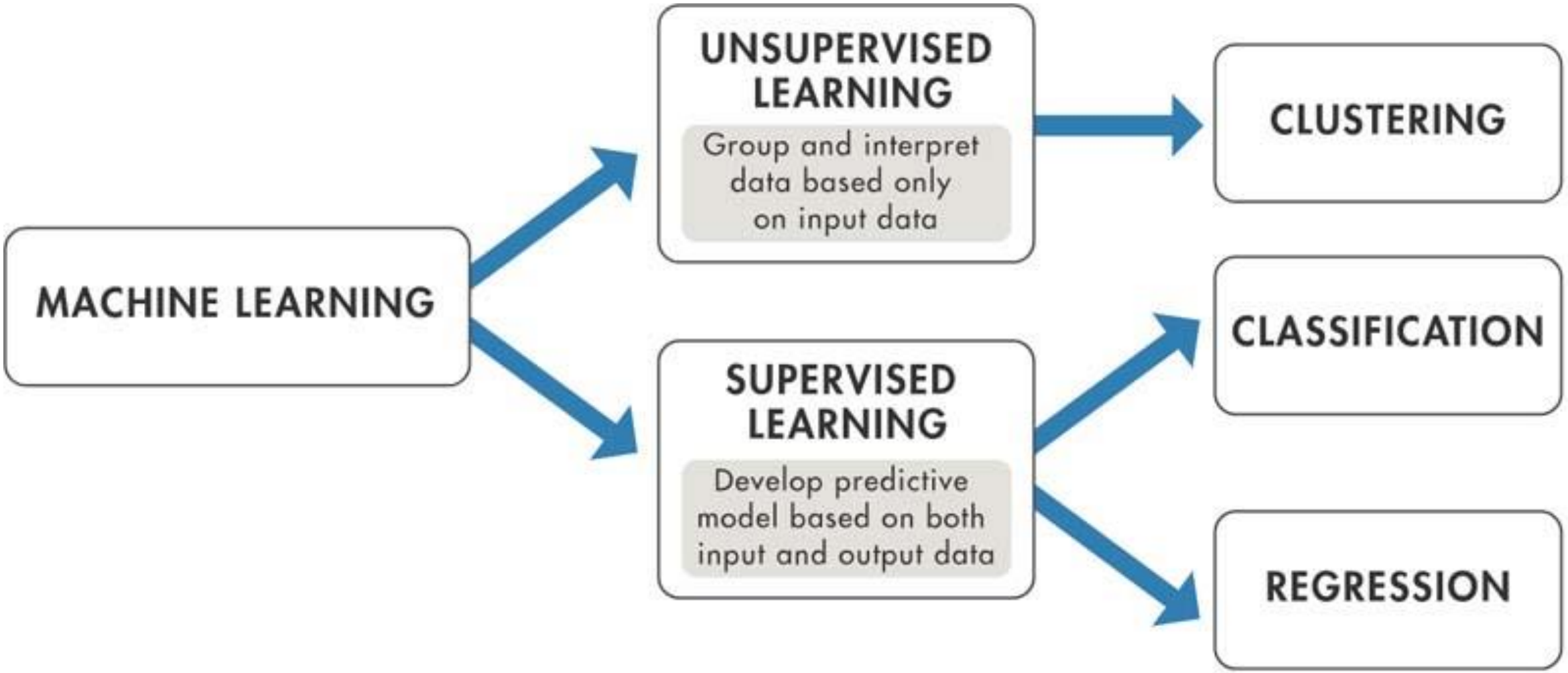
CAT, DOG, DUCK

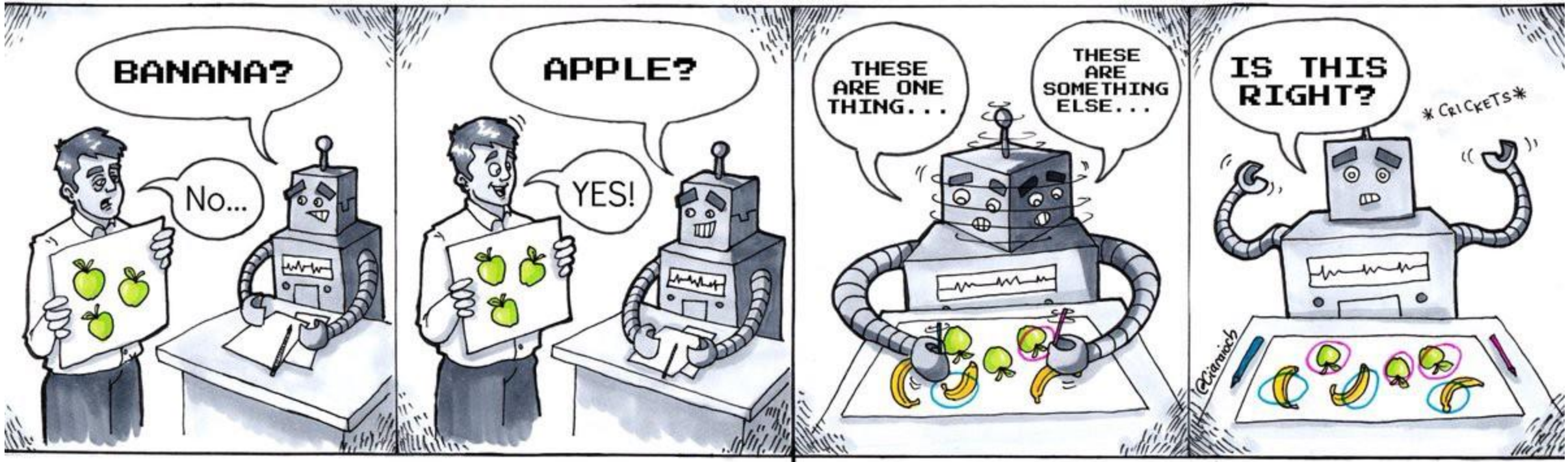
Instance Segmentation

CAT, DOG, DUCK

Single object

Multiple objects

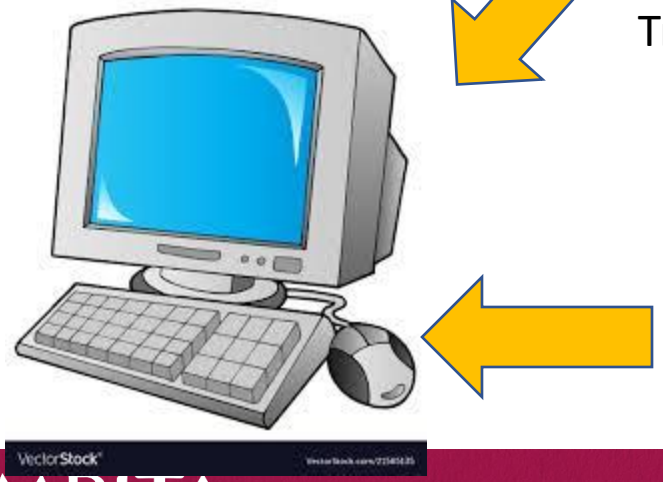

MACHINE LEARNING
UNSUPERVISED LEARNING
Group and interpret data based only on input data
SUPERVISED LEARNING
Develop predictive model based on both input and output data
CLUSTERING
CLASSIFICATION
REGRESSION



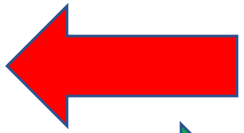
Supervised Learning

Unsupervised Learning

Id	Cl.thickness	Cell.size	Cell.shape	Marg.adhesion	Epith.c.size	Bare.nuclei	Bl.cromatin	Normal.nucleoli	Mitoses	Class
1000025	5	1	1	1	2	1	3	1	1	benign
1002945	5	4	4	5	7	10	3	2	1	benign
1015425	3	1	1	1	2	2	3	1	1	benign
1016277	6	8	8	1	3	4	3	7	1	benign
1017023	4	1	1	3	2	1	3	1	1	benign
1017122	8	10	10	8	7	10	9	7	1	malignant
1018099	1	1	1	1	2	10	3	1	1	benign
1018561	2	1	2	1	2	1	3	1	1	benign
1033078	2	1	1	1	2	1	1	1	5	benign
1033078	4	2	1	1	2	1	2	1	1	benign
1035283	1	1	1	1	1	1	3	1	1	benign
1036172	2	1	1	1	2	1	2	1	1	benign
1041801	5	3	3	3	2	3	4	4	1	malignant



Train the model with the above data set



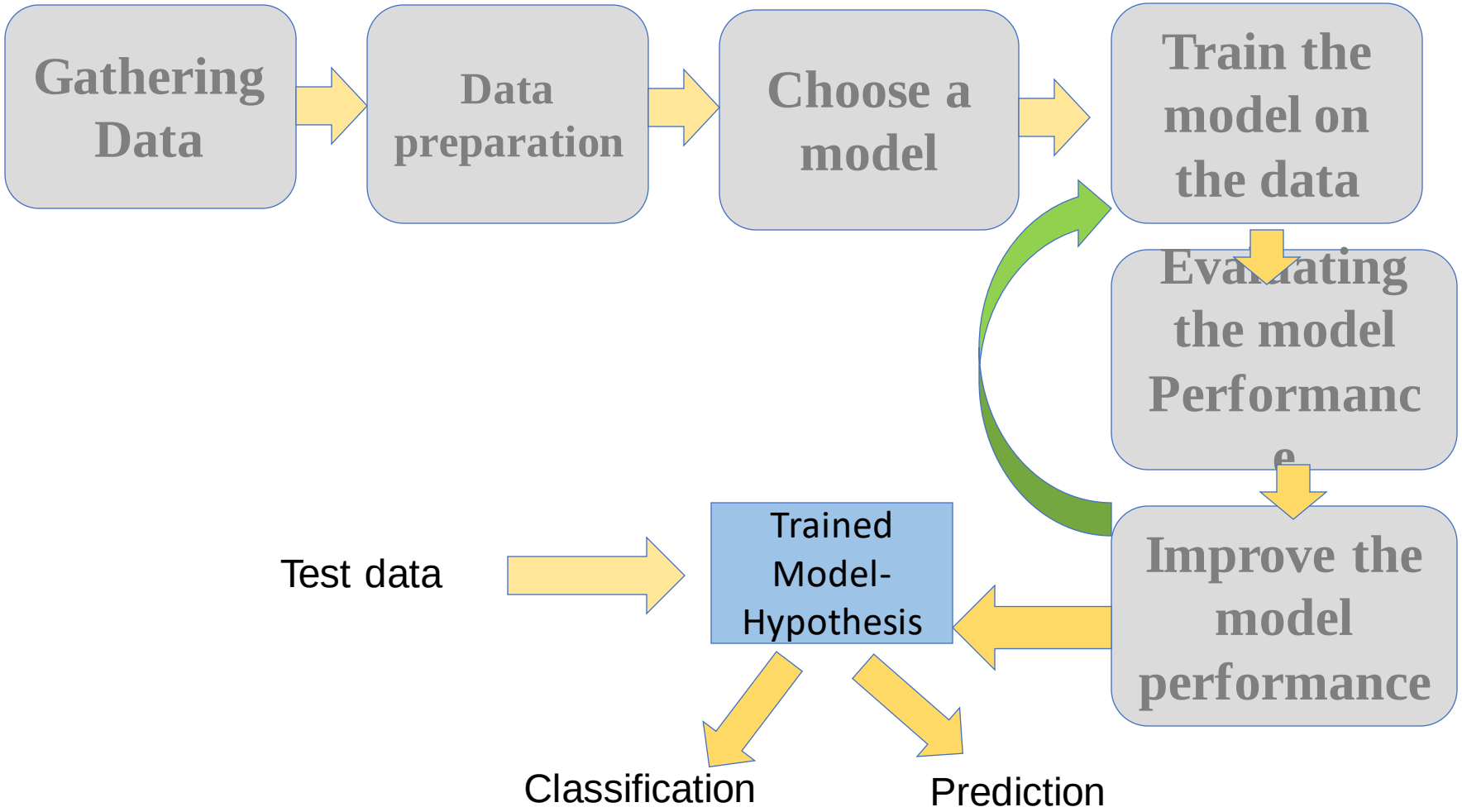
Is my cancer Malignat or benign

NO it is benign

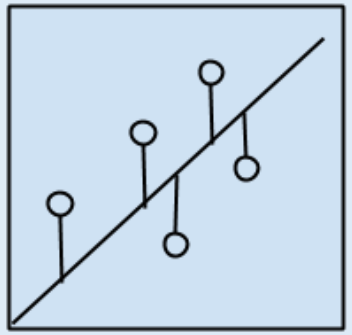
Test the model with a new real time input.

1041801	5	3	3	3	2	3	4	4	1

Steps of Supervised Learning

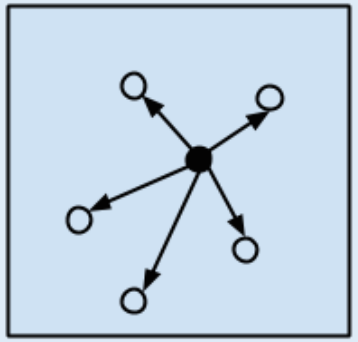


Regression Algorithms



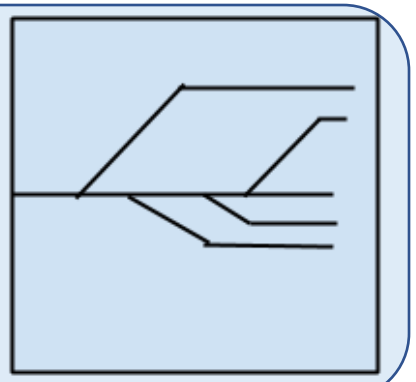
Regression Algorithms

Instance-based Algorithms



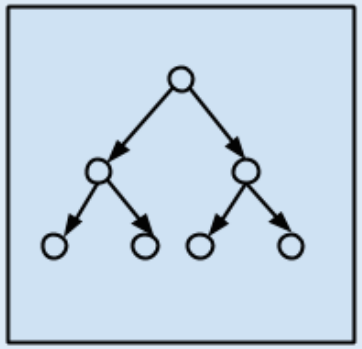
Instance-based Algorithms

Regularization Algorithms



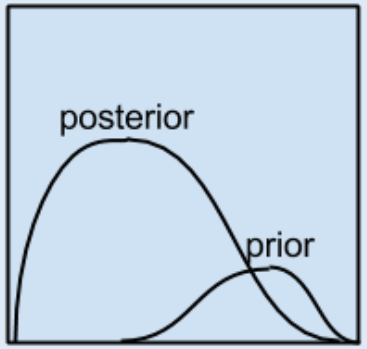
Regularization Algorithms

Decision Tree Algorithms



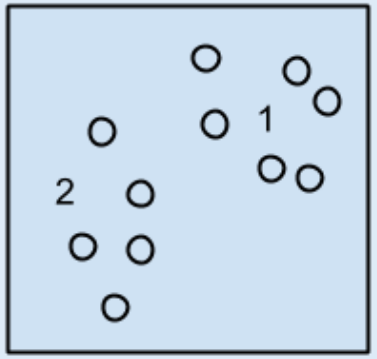
Decision Tree Algorithms

Bayesian Algorithm



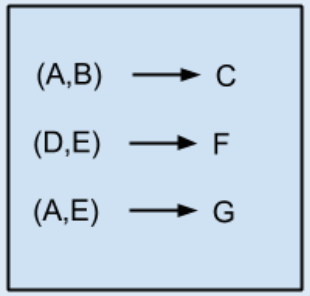
Bayesian Algorithms

Clustering Algorithms



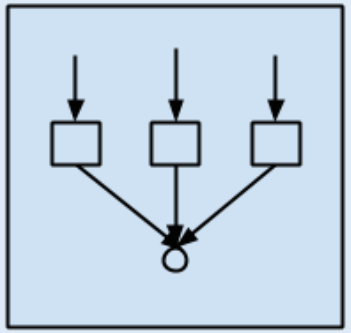
Clustering Algorithms

Association Rule Learning Algorithms



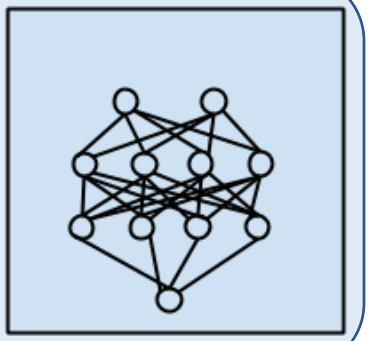
Association Rule Learning Algorithms

Ensemble Algorithms



Ensemble Algorithms

Artificial Neural Network / Deep Learning algorithms

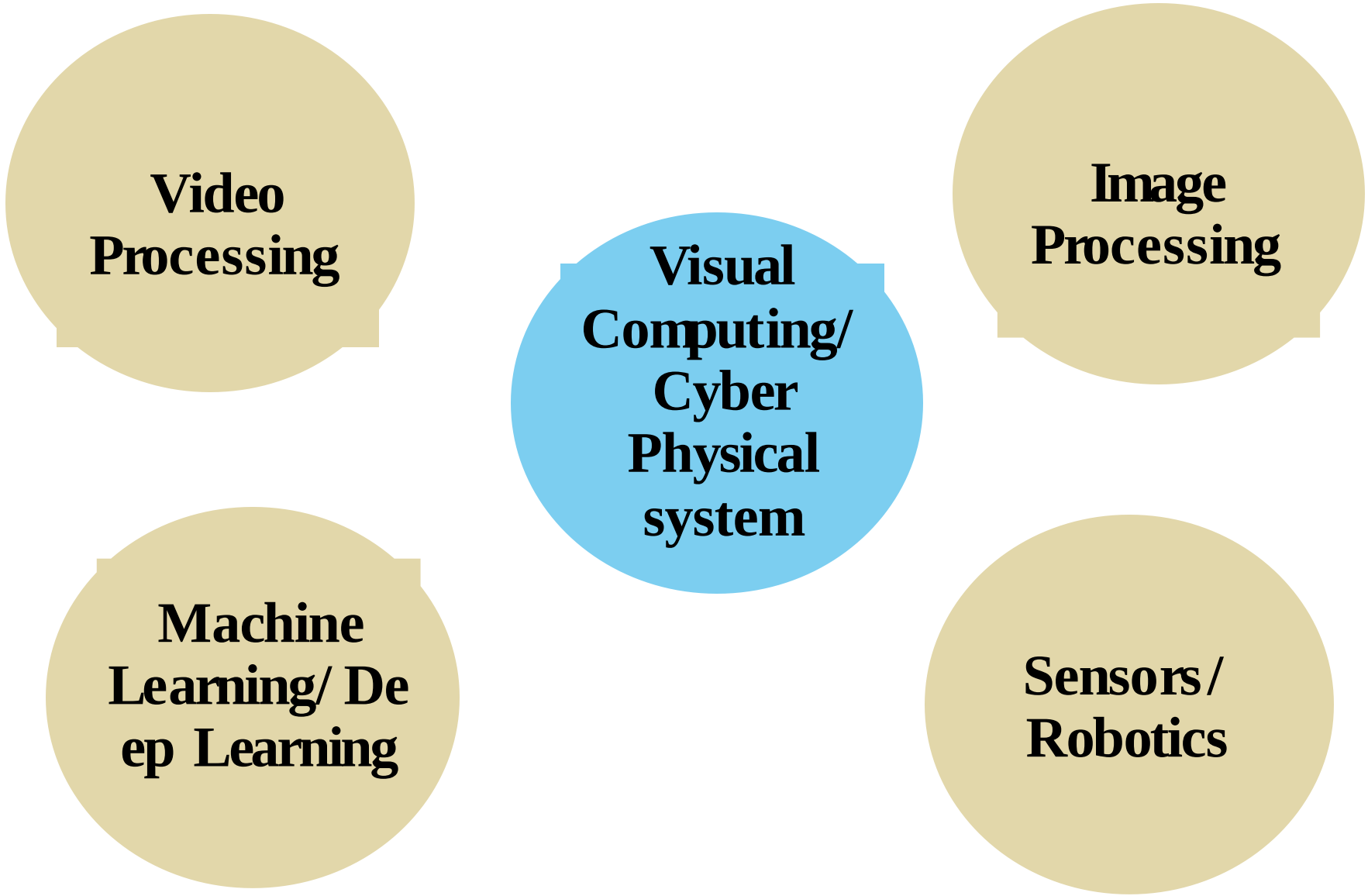


Deep Learning Algorithms

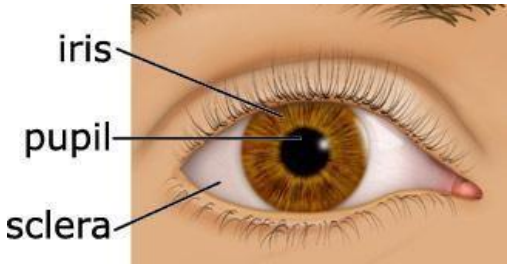
Introduction to

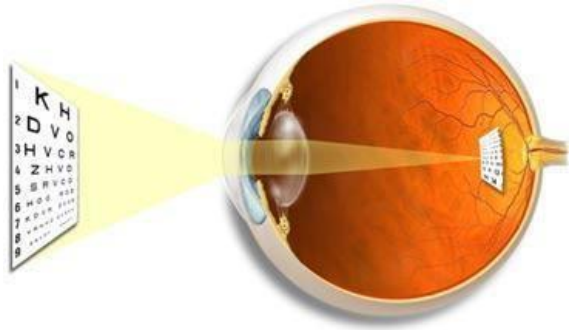
Computer Vision

Visual computing



Computer Vision- Emulate Human vision





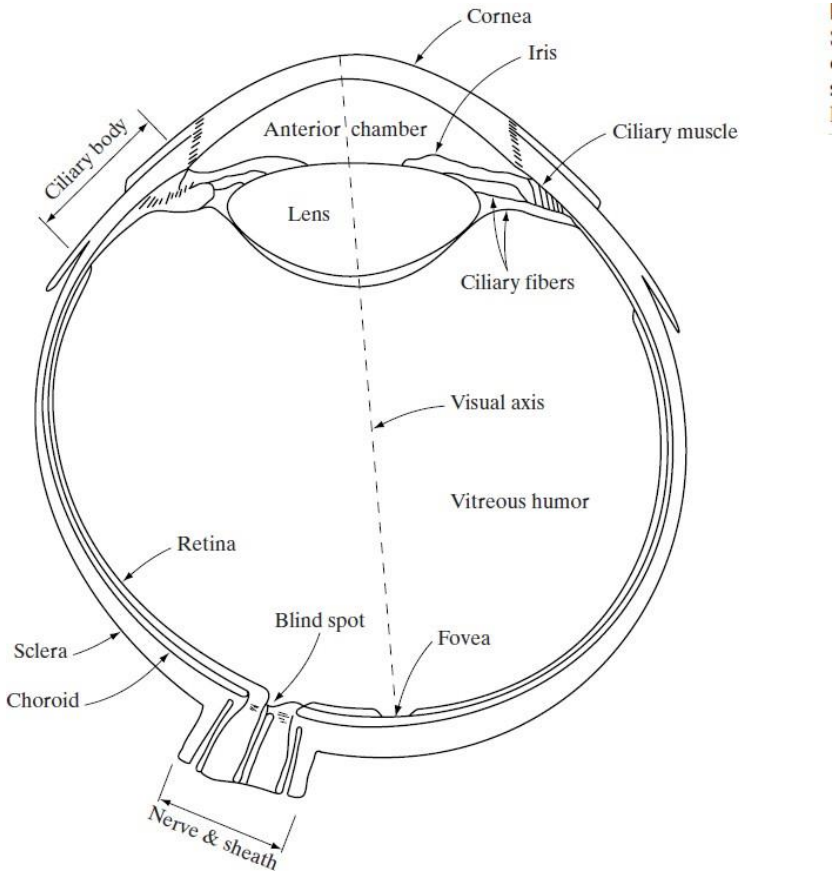
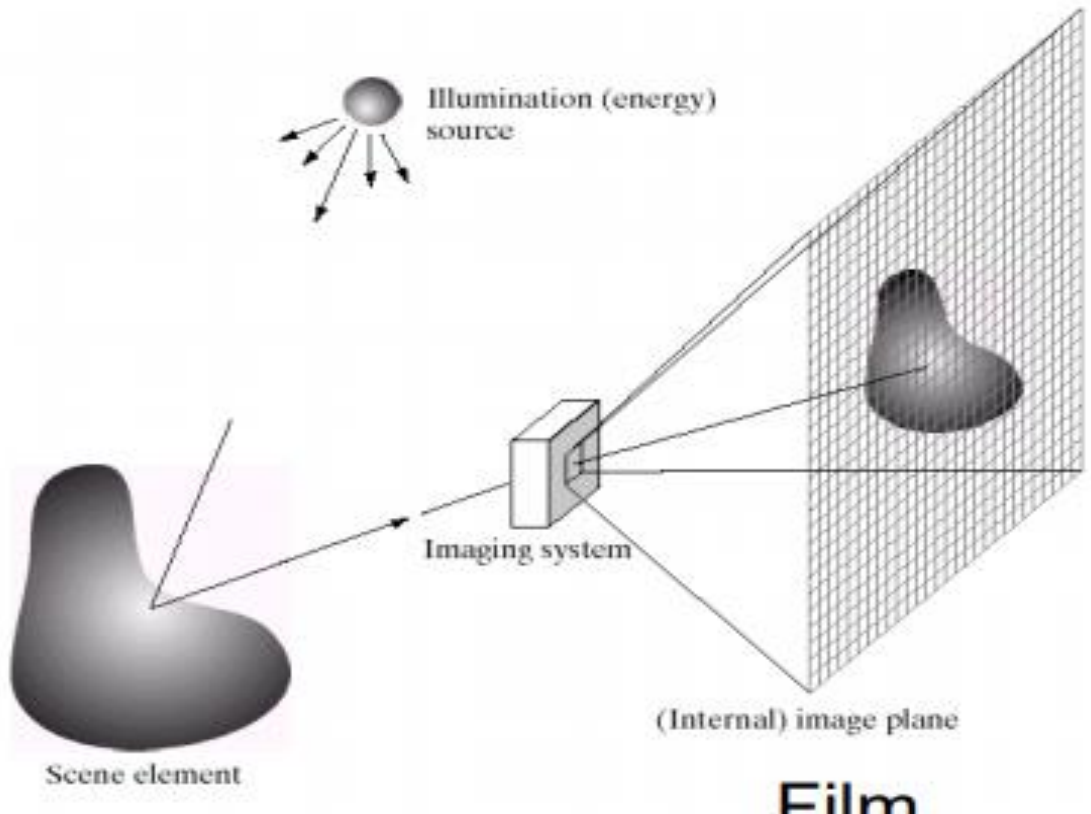
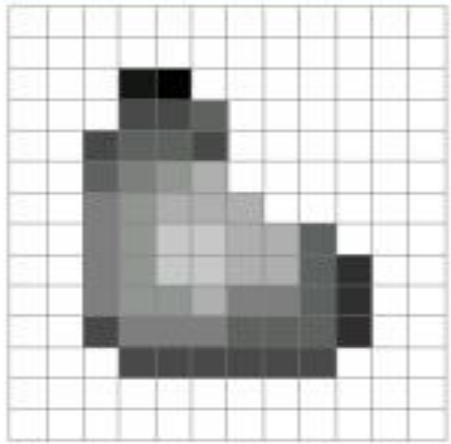


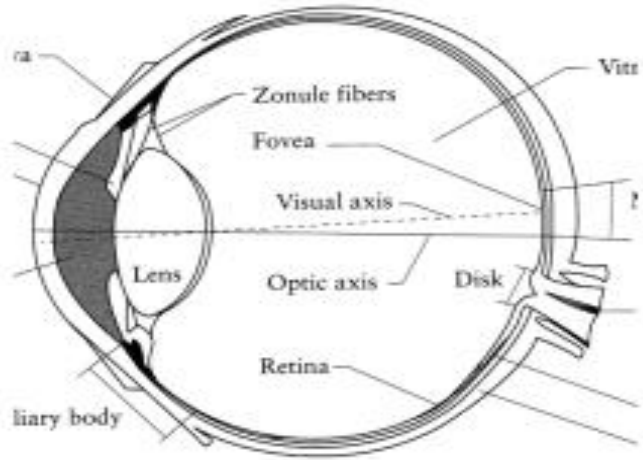
Image Formation: Simple Model



Film



Digital Camera



The Eye

Image Matrix- pixel,resolution(spatial, graylevel)

$F(x,y)$

148	123	52	107	123	162	172	123	64	89
147	130	92	95	98	130	171	155	169	163
141	118	121	148	117	107	144	137	136	134
82	106	93	172	149	131	138	114	113	129
57	101	72	54	109	111	104	135	106	125
138	135	114	82	121	110	34	76	101	111
138	102	128	159	168	147	116	129	124	117
113	89	89	109	106	126	114	150	164	145
120	121	123	87	85	70	119	64	79	127
145	141	143	134	111	124	117	113	64	112

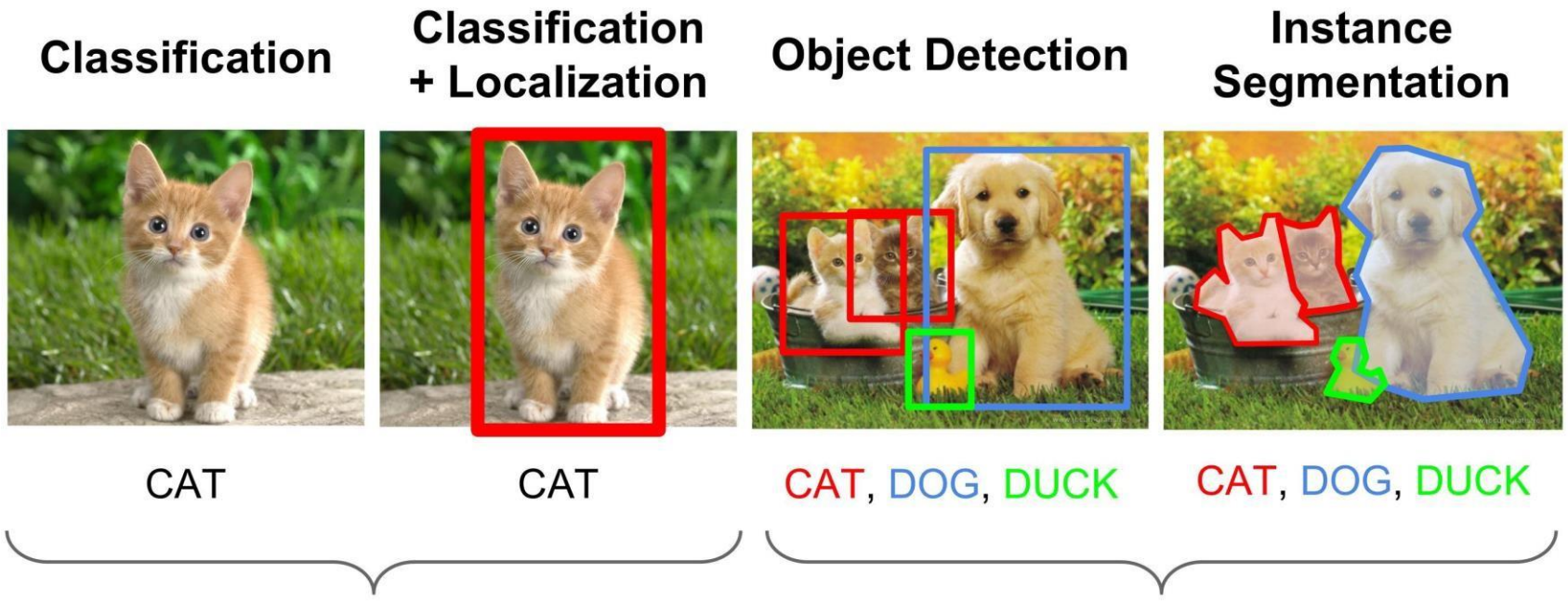
$I(u,v)$

	165	187	209	58	7	
	14	125	233	201	98	159
253	144	120	251	41	147	204
67	100	32	241	23	165	30
209	118	124	27	59	201	79
210	236	105	169	19	218	156
35	178	199	197	4	14	218
115	104	34	111	19	196	
32	69	231	203	74		

Color image

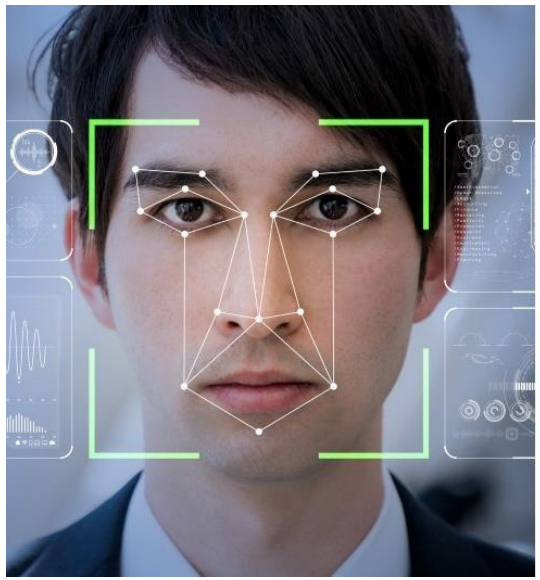
Gray Scale image 8 bit gray scale image 0-255 0-black, 255- white

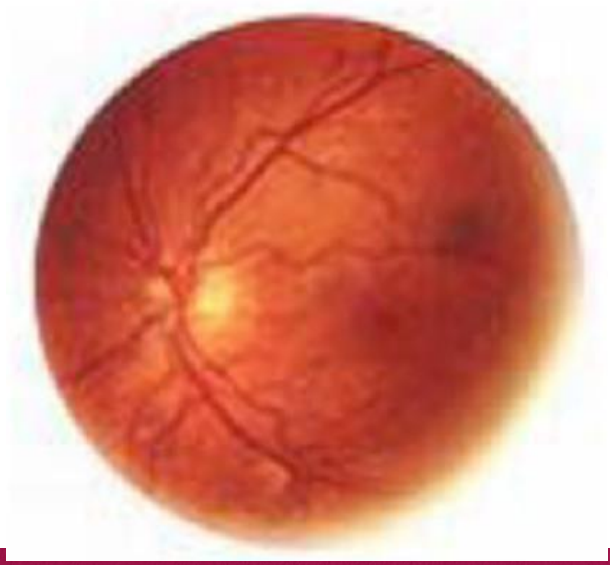
Computer Vision Applications


Classification
Classification + Localization
Object Detection
Instance Segmentation
CAT
CAT
CAT, DOG, DUCK
CAT, DOG, DUCK

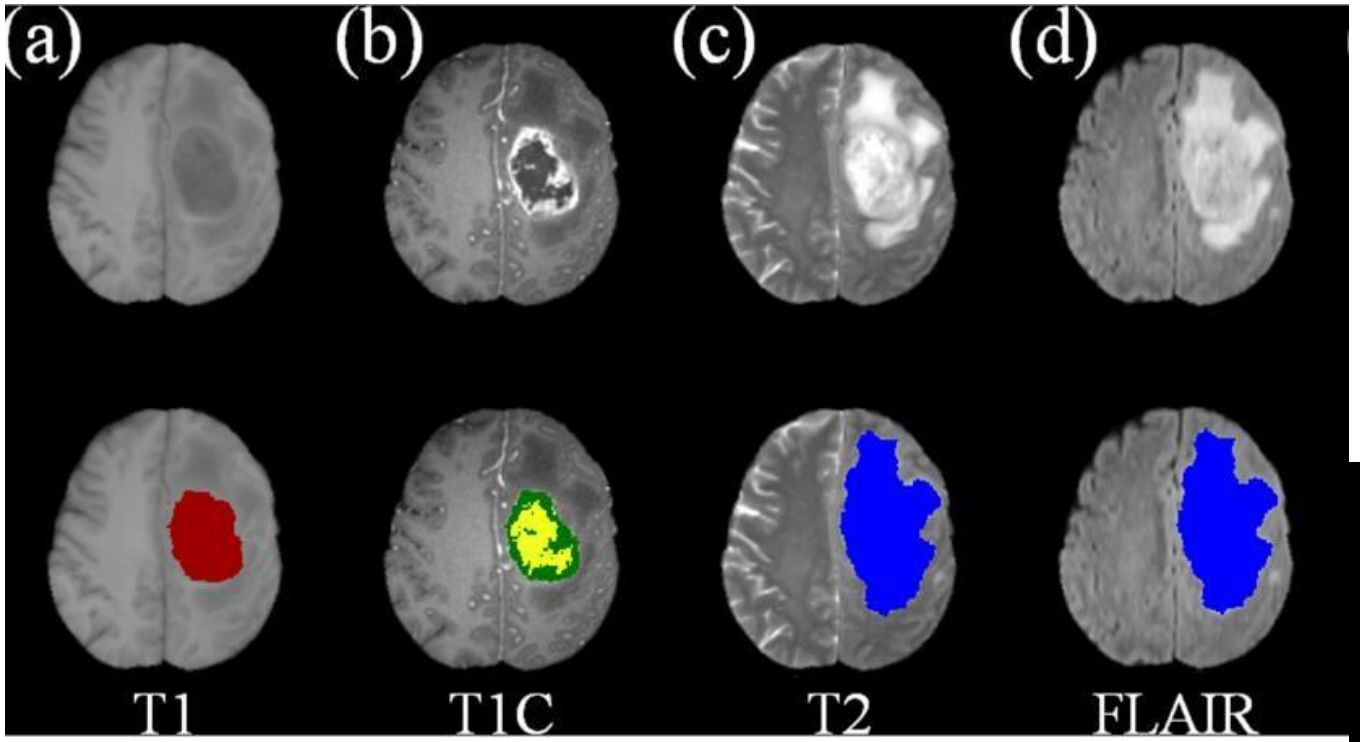

PERSON, CAT, DOG
(A) Classification
(B) Detection
(C) Segmention

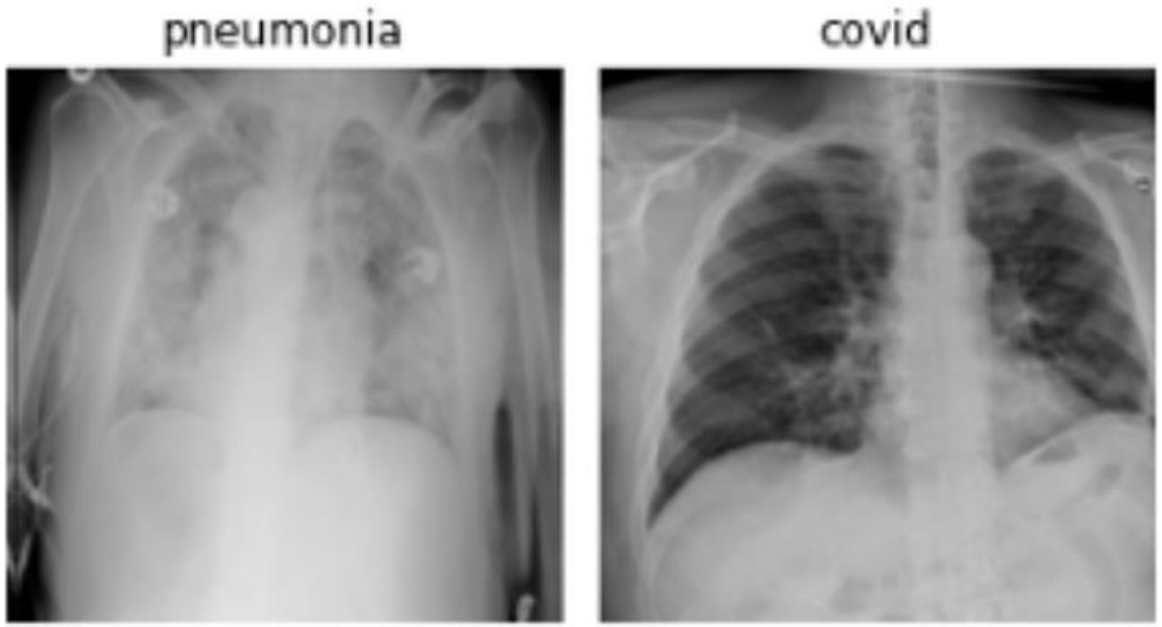
Biometric Applications

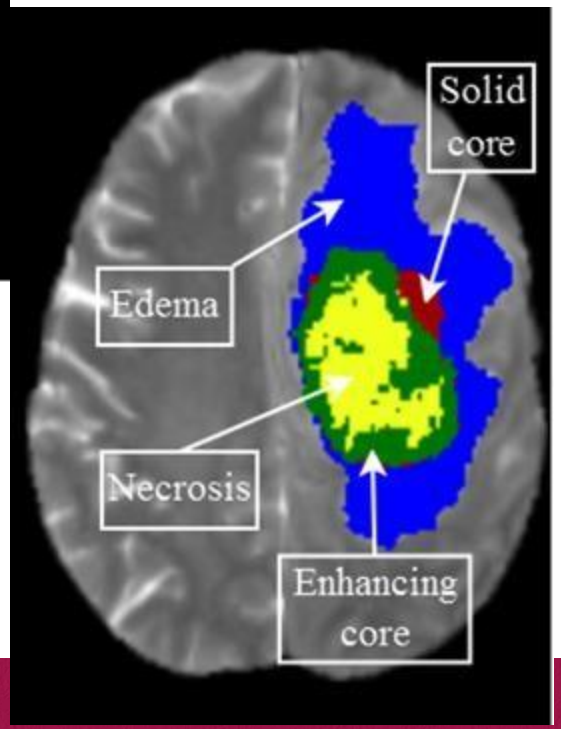




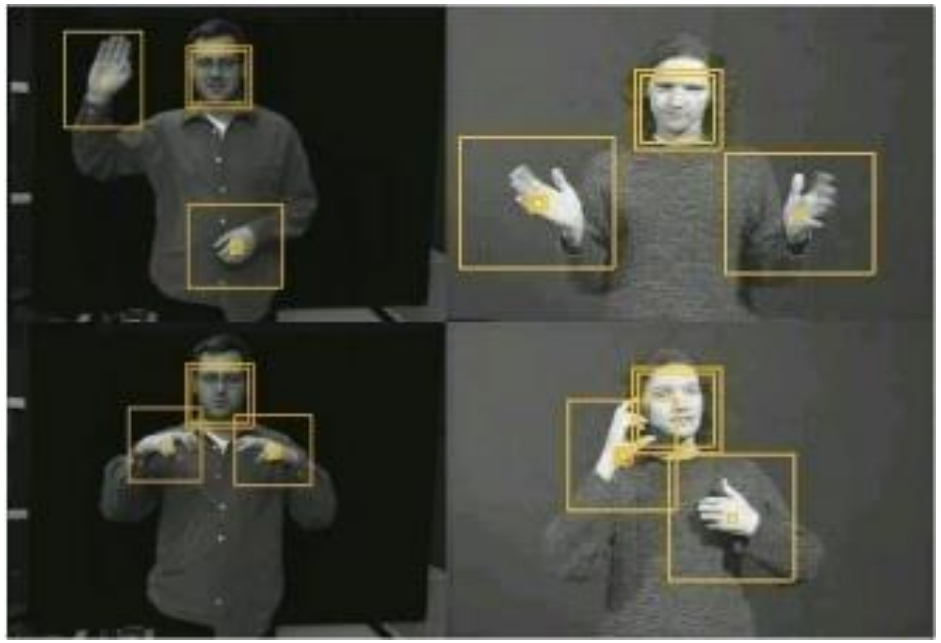
Medical Image Analysis

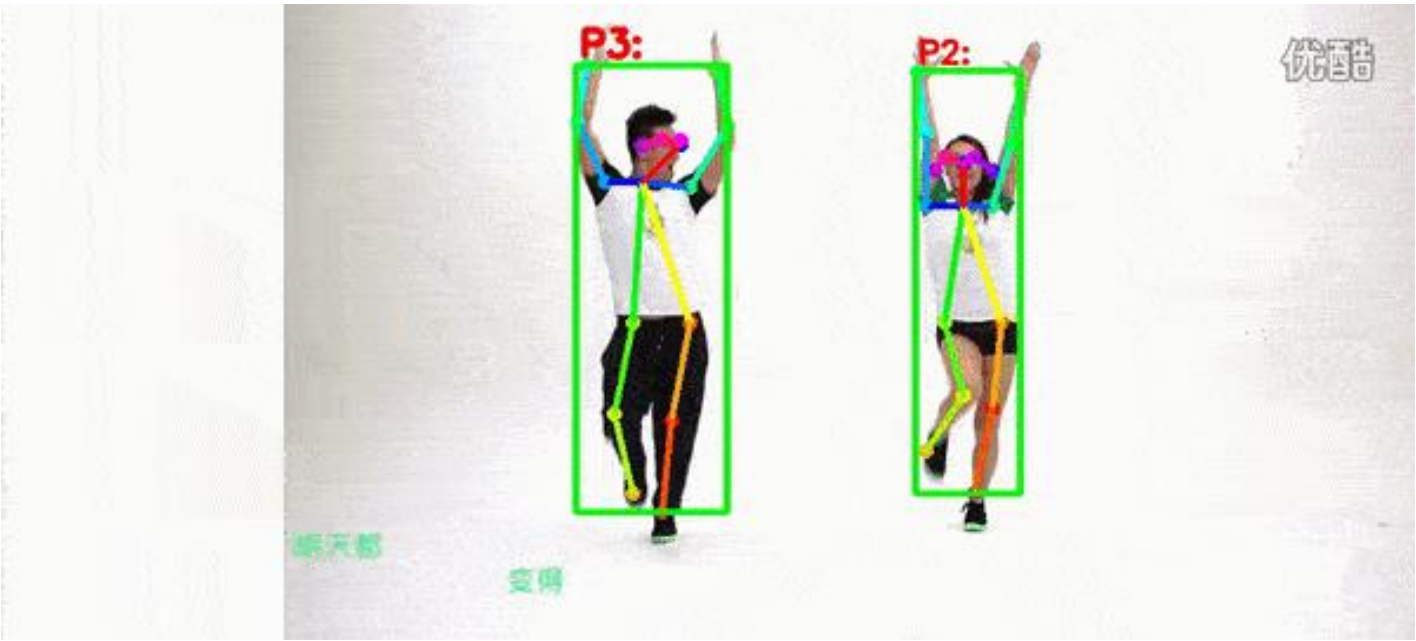




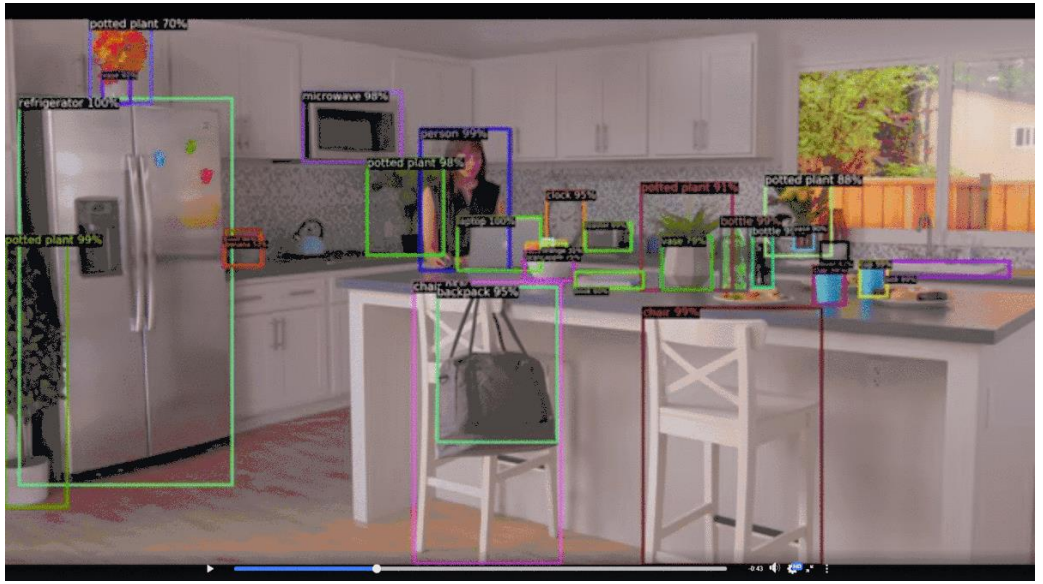


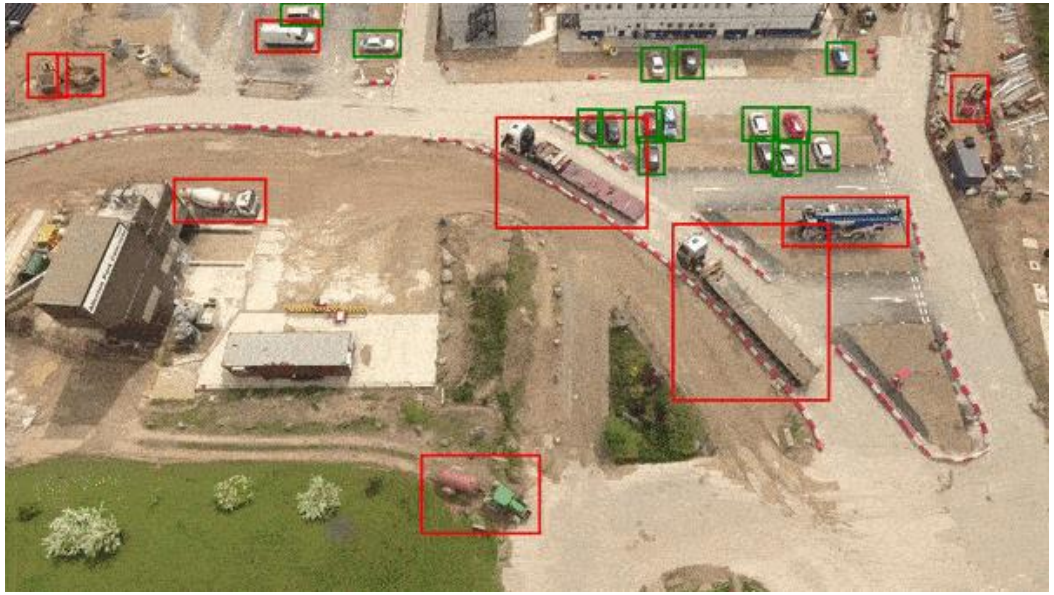
Video Processing - Gesture Recognition/Action Recognition

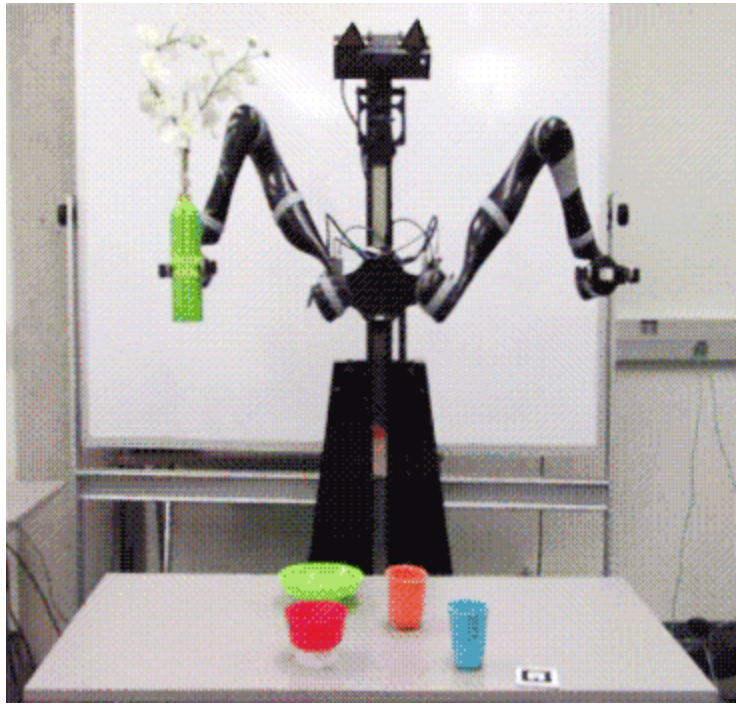


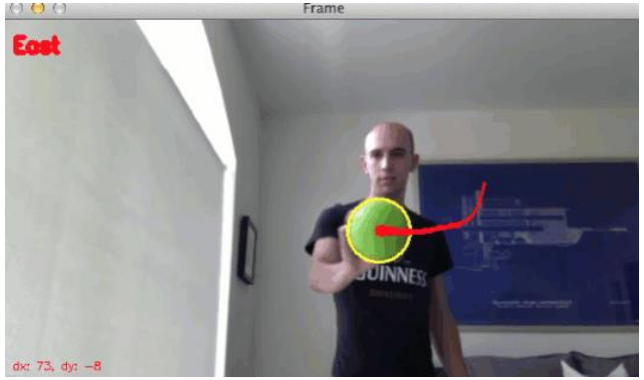



car : 0.967
person
car : 0.787
0.904
car : 0.997

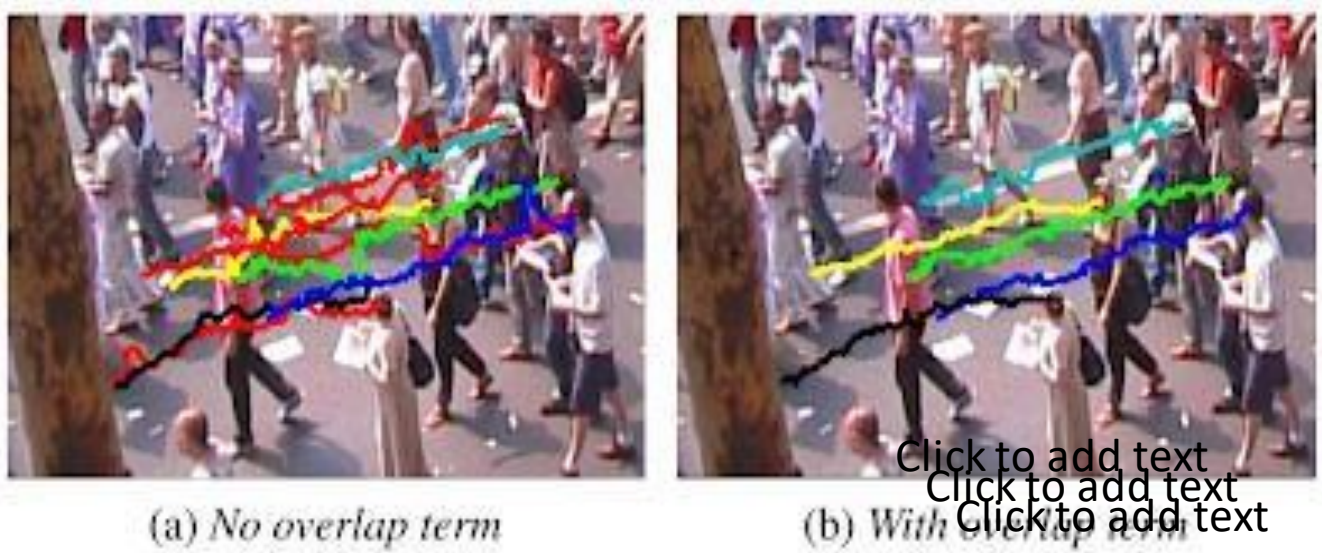

potted plant 70%
refrigerator 100%
microwave 98%
person 99%
potted plant 98%
potted plant 91%
potted plant 88%
clock 95%
laptop 100%
potted plant 99%
chair 96%
backpack 95%
chair 99%






Frame
East
dx: 73, dy: −8

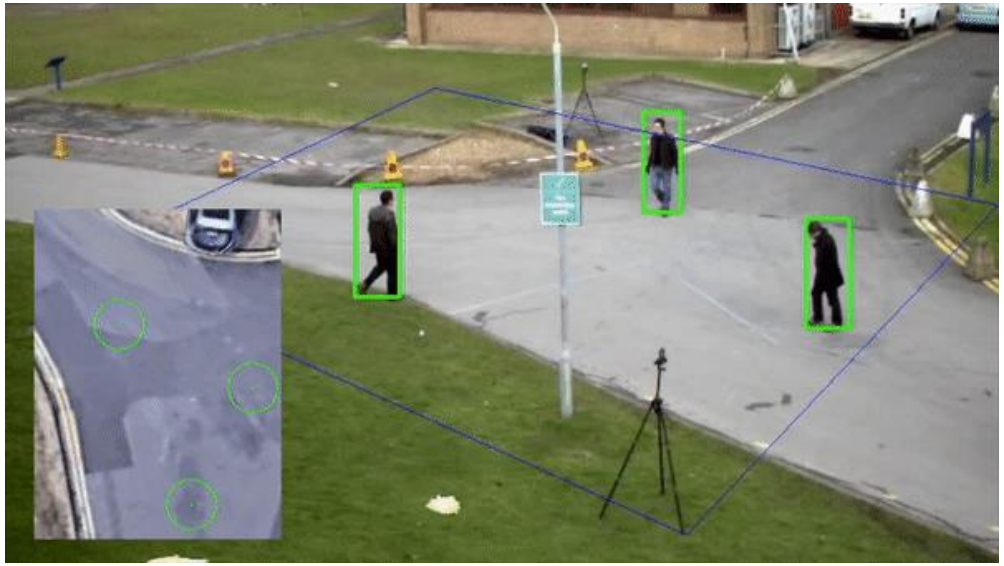
Object Tracking



(a) *No overlap term* (b) *With overlap term*

(c) *No co-occurrence term* (d) *With co-occurrence term*

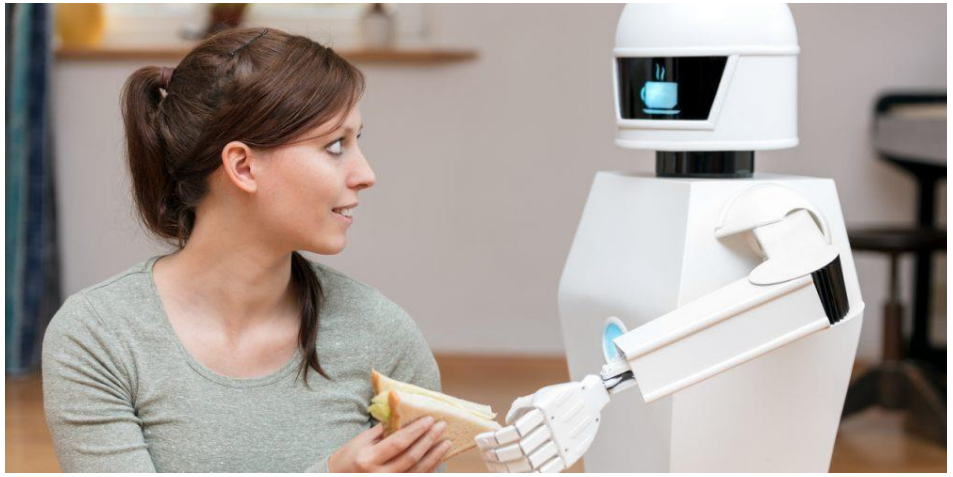
Click to add text
Click to add text
Click to add text



Tracking Social Distancing??

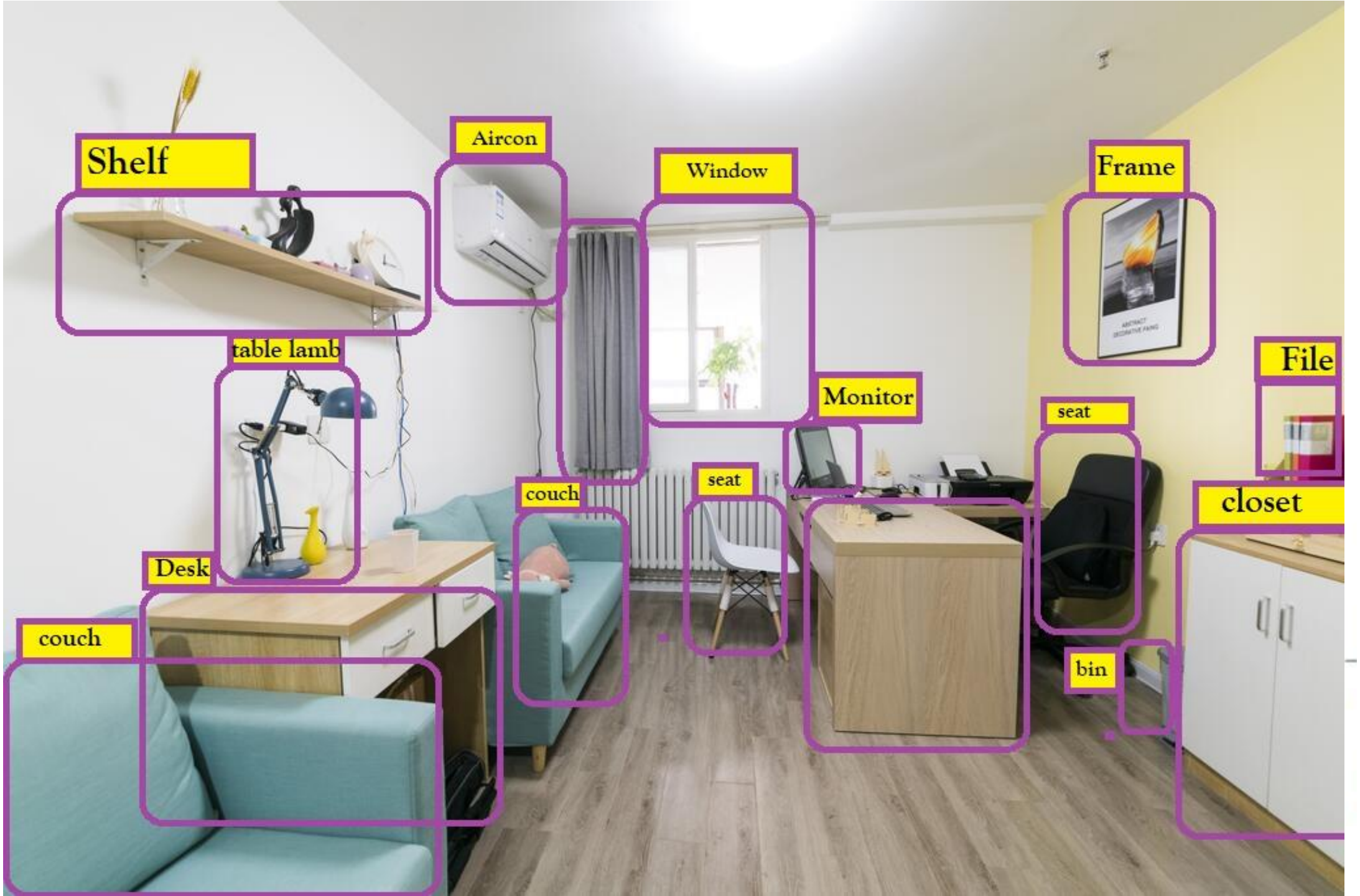
Vision based Robotic Applications-

Manufacturing companies



Automatic Navigation robots/Social robots

Scene Understanding



Foil

Yes

No

What is the sandwich laying on?

What is the food on?

Is this a sandwich?

Has the sandwich been cut?

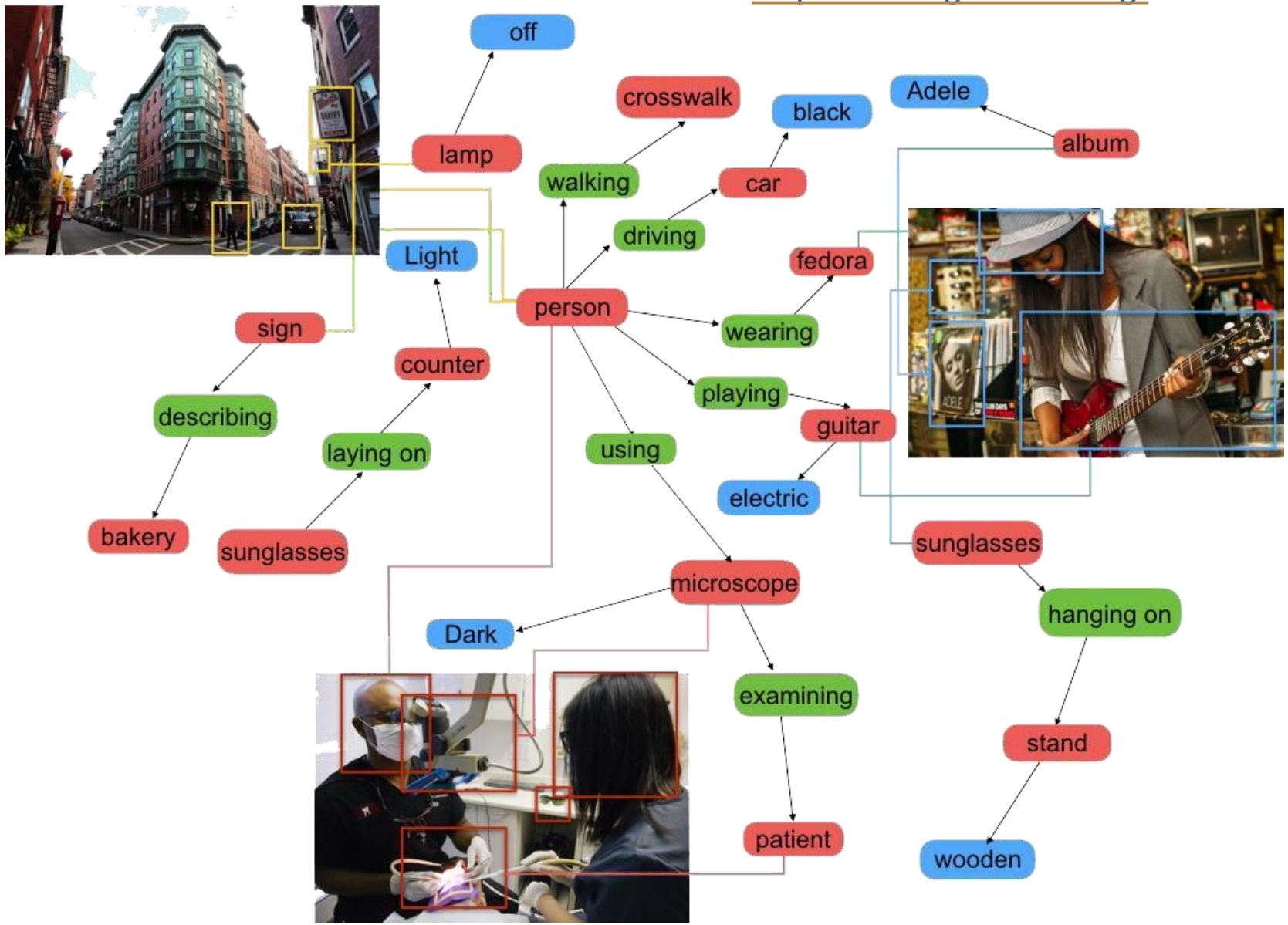
Is the boy wearing a hat?

Is the boy playing baseball?

Is this a professional baseball player?

Did the boy hit the ball?

http://visualgenome.org/



Detect Badminton Ball in User uploaded videos

- Add more of training data and detect new objects

Object Detection in Sports Videos

M. Burić, M. Pobar, M. Ivašić-Kos
University of Rijeka/Department of Informatics, Rijeka, Croatia
matija.buric@hep.hr, marinai@inf.uniri.hr, mpobar@inf.uniri.hr

***Abstract* - Object detection is commonly used in many computer vision applications. In our case, we need to apply the object detector as a prerequisite for action recognition in handball scenes. Object detection, to be successful for this task, should be as accurate as possible and should be able to deal with a different number of objects of various sizes, partially occluded, with bad illumination and deal with cluttered scenes. The aim of this paper is to provide an overview of the current state-of-the-art detection methods that rely on convolutional neural networks (CNNs) and test their performance on custom video sports materials acquired during handball training and matches. The comparison of the detector performance in different conditions will be given and discussed.**

Keywords – object detectors; sports scenes; Mixture of Gaussians; YOLO; Mask R-CNN

I. INTRODUCTION

Object detection is one of the fundamental tasks in

There are many other factors which can degrade the detection of players, of the ball and of the lines on the playground that humans don't even notice since it comes naturally to us.

To tackle the object detection problem, many approaches have been proposed including the Viola-Jones detector with Haar Cascades [2], HOG gradient-based approaches [3], segmentation and template matching approaches, and recent state-of-the-art methods that rely on deep convolutional neural networks (CNNs). In the last few years, CNNs have achieved a tremendous increase in the accuracy of object detection and are widely considered as the de facto standard approach for the most image recognition tasks.

Object detection in videos presents additional challenges, as it is usually desirable to track the identity of various objects between frames. It can be performed applying an object detector frame by frame, similarly as in case of images, or by using some kind of multi-frame

ɔs://bib.irb.hr/datoteka/941522.5075-ect-Detection-in-Sports-Videos-v2.pdf

Tracking Badminton Ball Once Marked

TrackNet: A Deep Learning Network for Tracking High-speed and Tiny Objects in Sports Applications

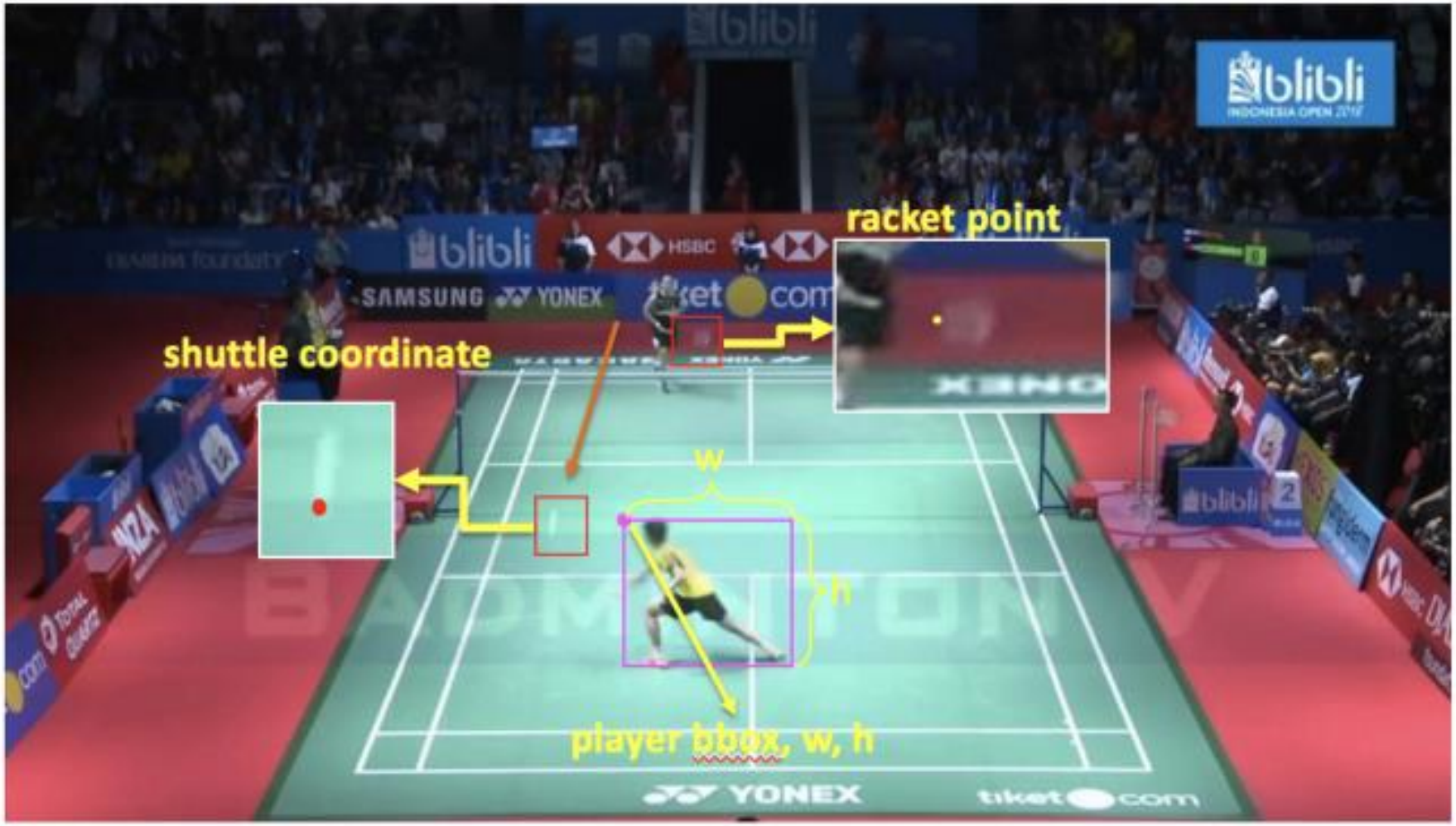
Yu-Chuan Huang I-No Liao Ching-Hsuan Chen Tsì-Uí İk* Wen-Chih Peng
Department of Computer Science, College of Computer Science
National Chiao Tung University
1001 University Road, Hsinchu City 30010, Taiwan
*Email: cwyi@nctu.edu.tw

***Abstract*—Ball trajectory data are one of the most fundamental and useful information in the evaluation of players' performance and analysis of game strategies. Although vision-based object tracking techniques have been developed to analyze sport competition videos, it is still challenging to recognize and position a high-speed and tiny ball accurately. In this paper, we develop a deep learning network, called TrackNet, to track the tennis ball from broadcast videos in which the ball images are small, blurry, and sometimes with afterimage tracks or even invisible. The proposed heatmap-based deep learning network is trained to not only recognize the ball image**

topic in the areas of image processing and deep learning. In the applications of sports analyzing and athletes training, videos are helpful in the post-game review and tactical analysis. In professional sports, high-end cameras have been used to record high resolution and high frame rate videos and combined with image processing for referee assistance or data collection. However, this solution requires enormous resources and is not affordable for individuals or amateurs. Developing a low-cost solution for data acquisition from broadcast videos will be significant

https://arxiv.org/pdf/1907.03698.pdf

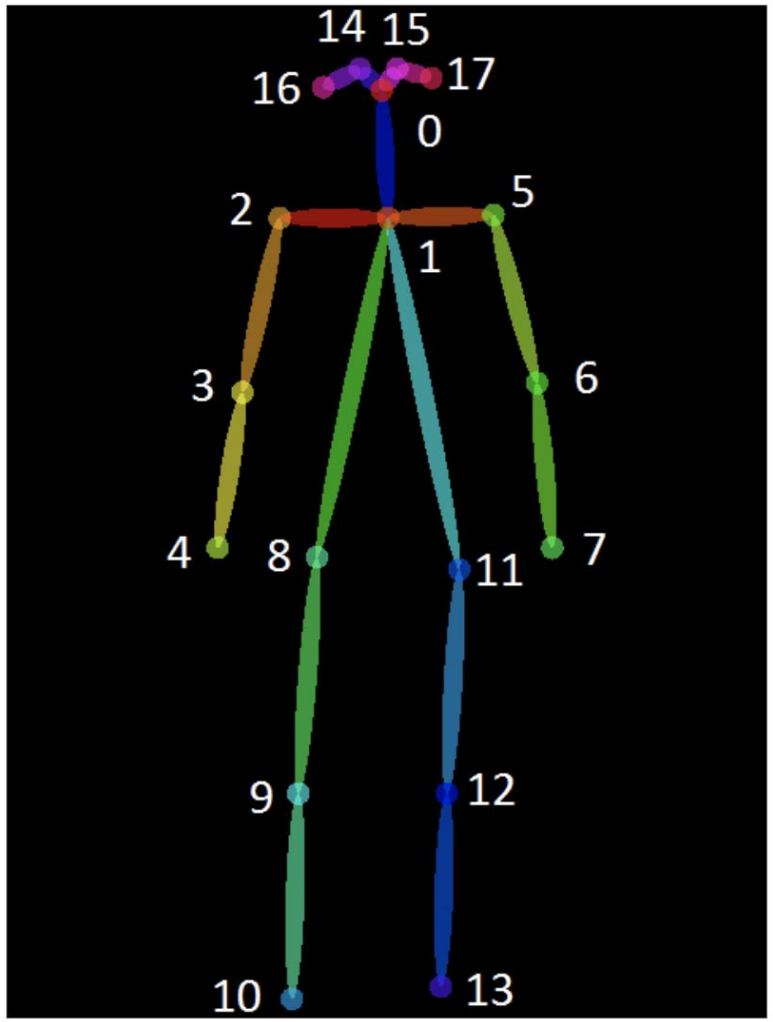
Tracking Players in Badminton Game

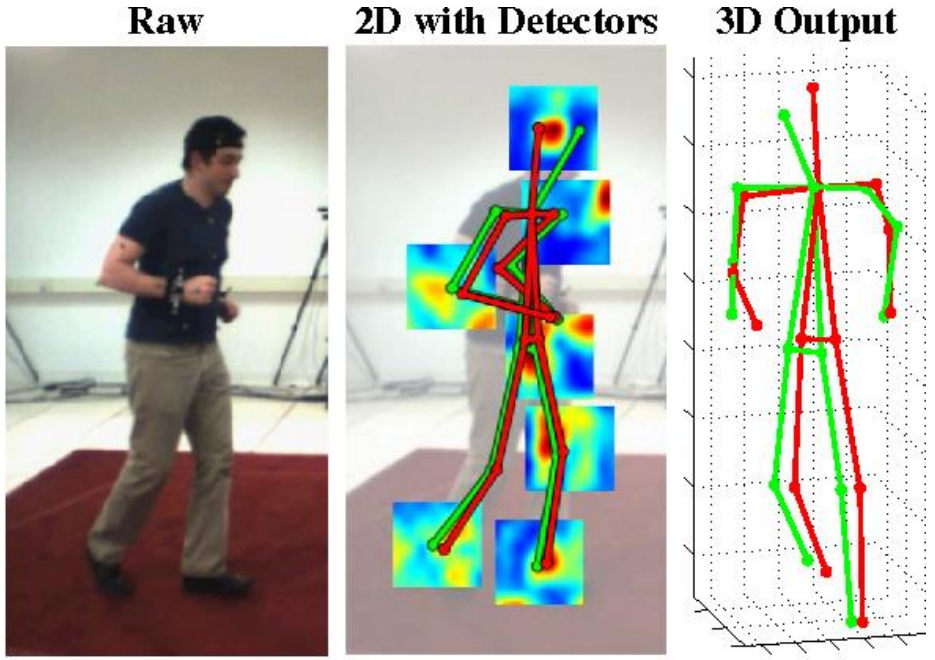


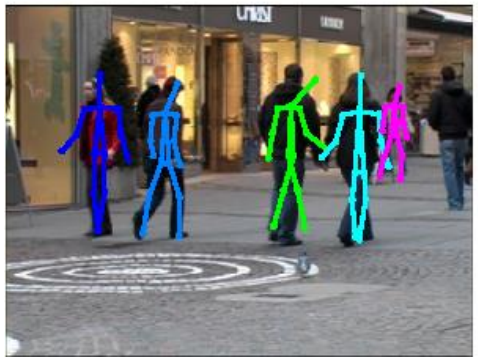
https://arxiv.org/pdf/1907.12888.pdf

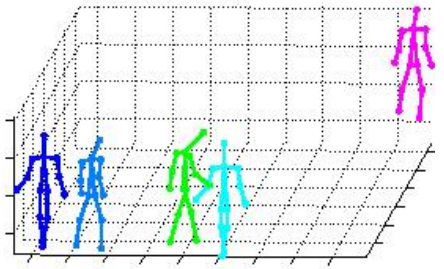
3. Human Pose Estimation

Human Pose Estimation is defined as the problem of localization of human joints (also known as keypoints - elbows, wrists, etc) in images or videos







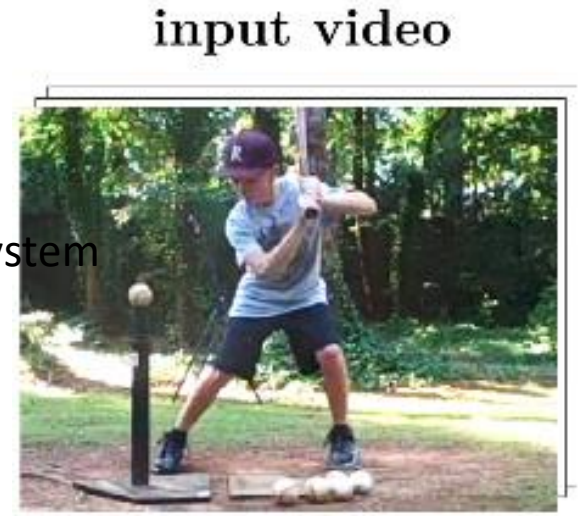


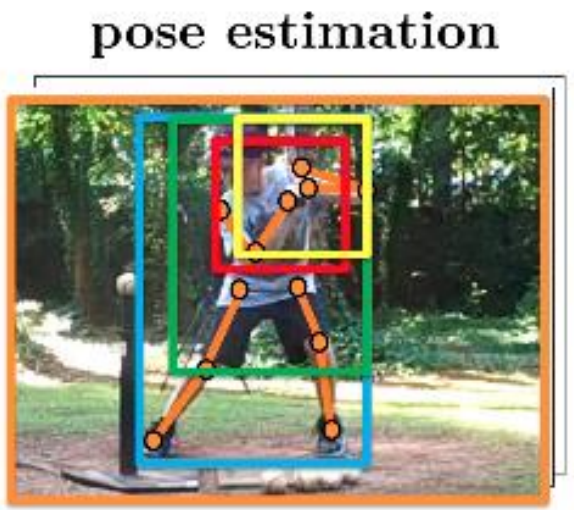
https://nanonets.com/blog/human-pose-estimation-2d-guide/

Pose Estimation

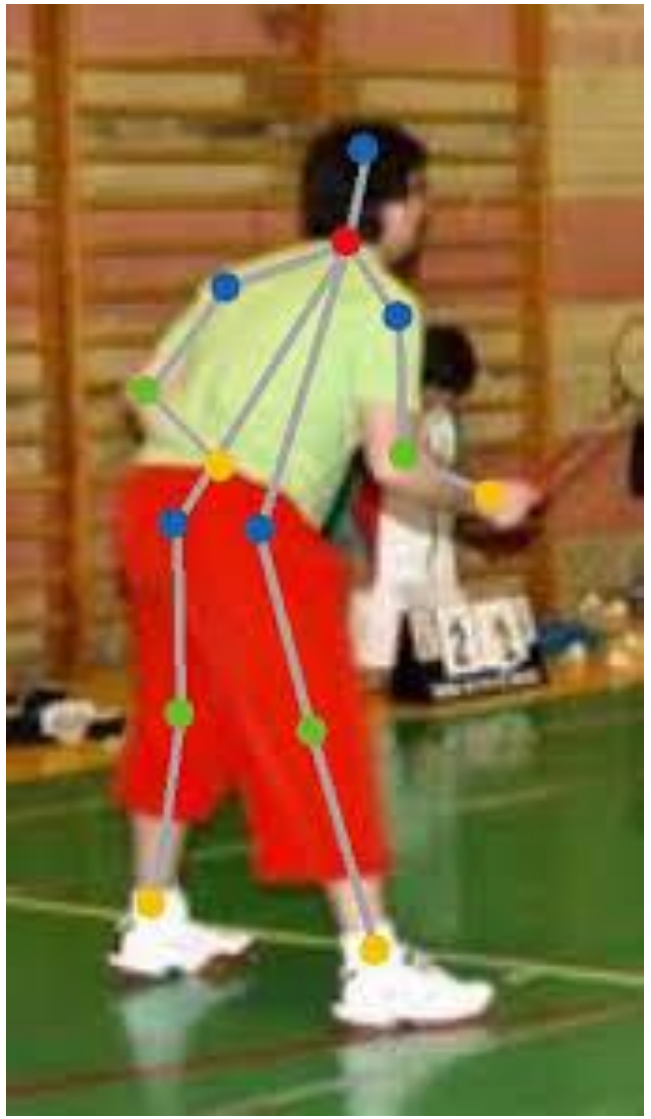
Applications

- Assisted living
- Character animation
- Intelligent driver assisting system
- Video games
- Medical Applications
- Other applications



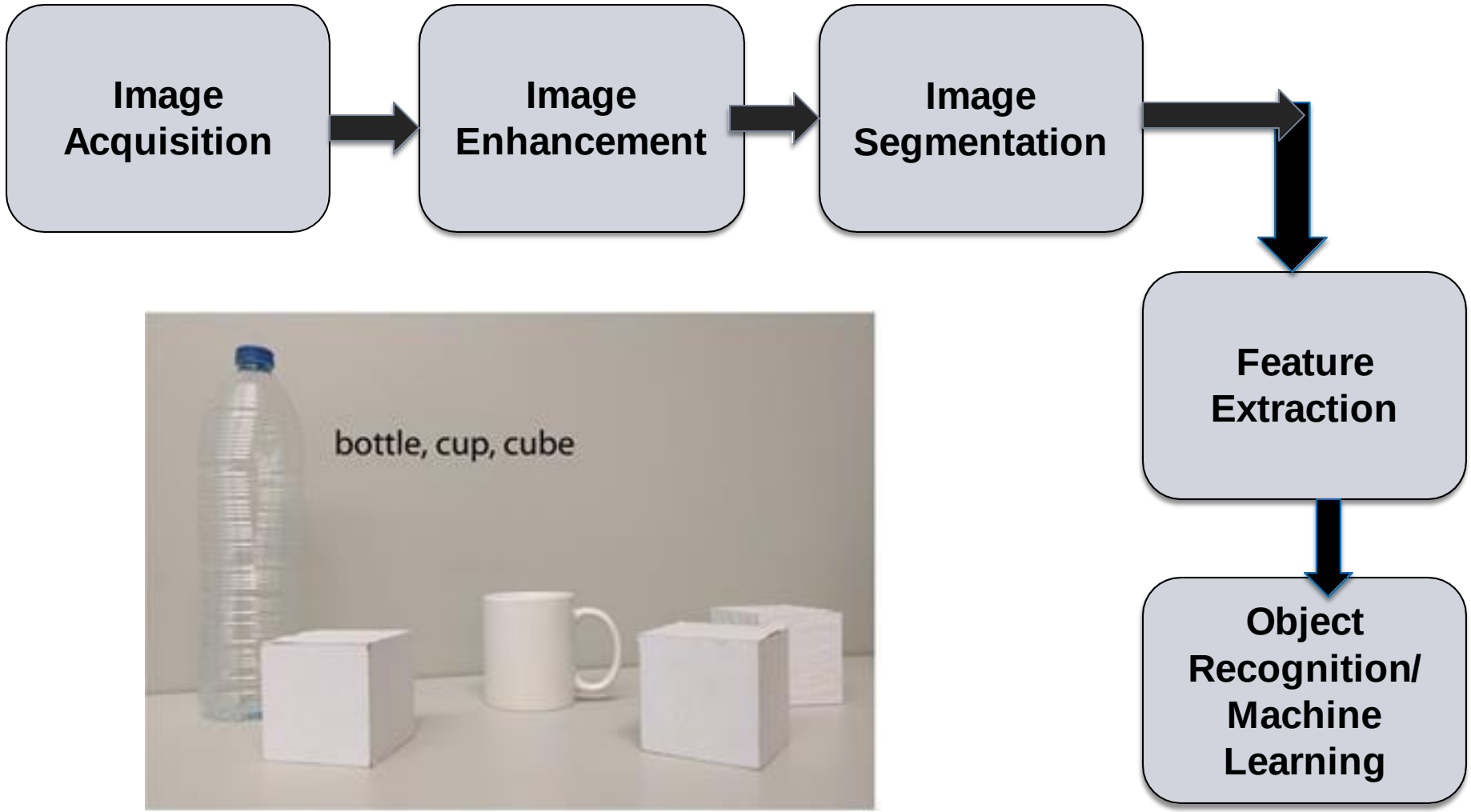


3. Pose Estimation of Selected Player in Badminton Game

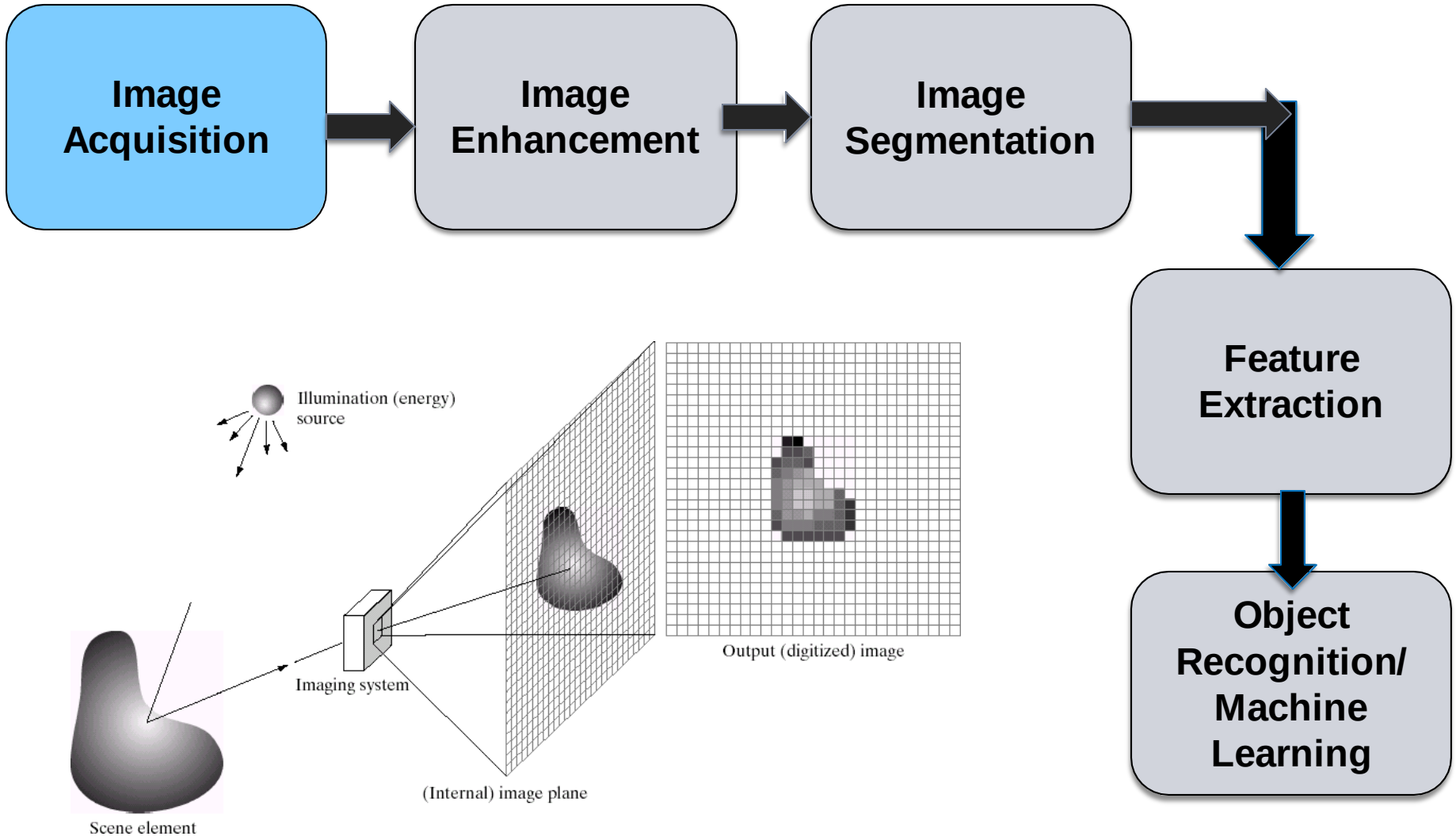


Steps of Image Analysis

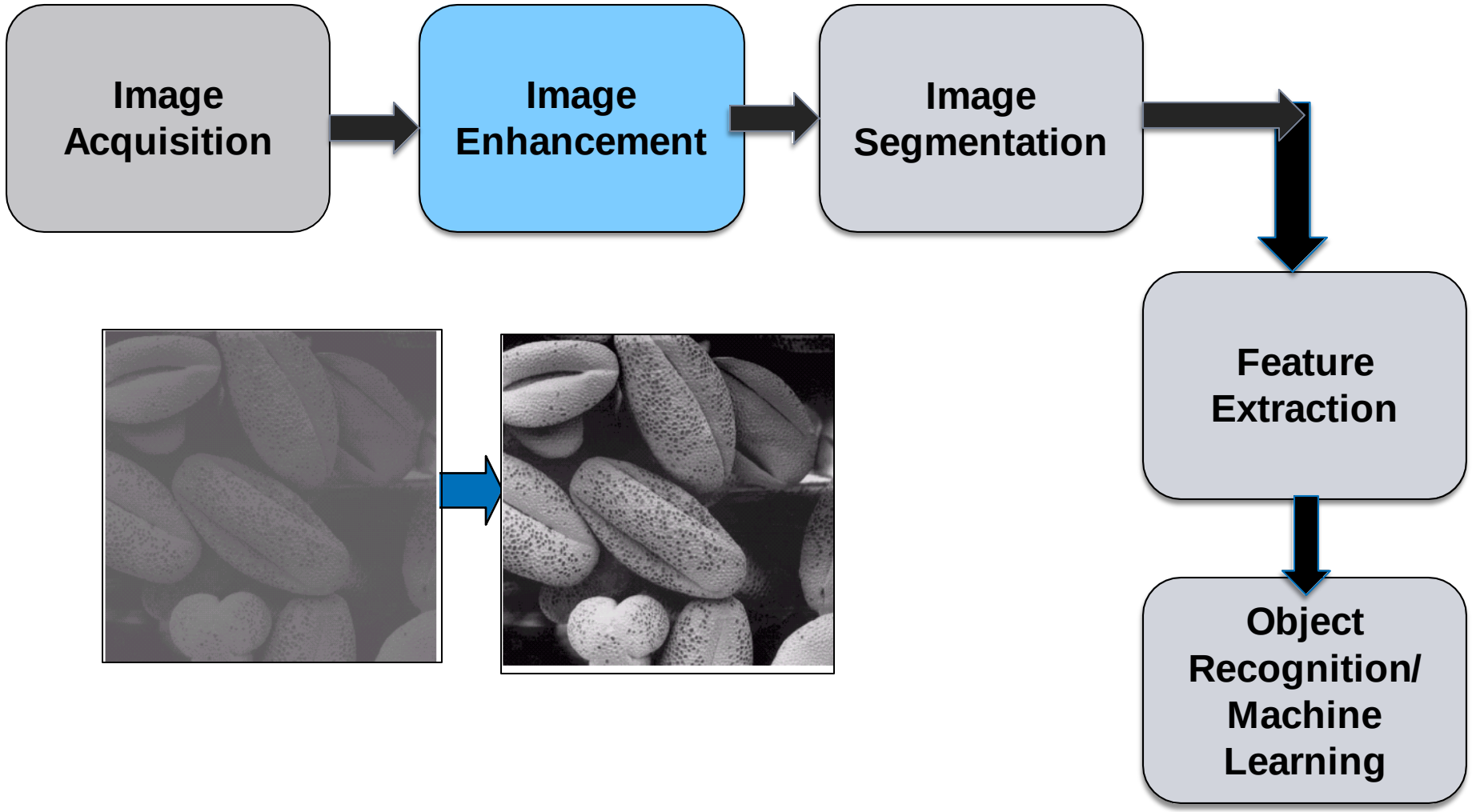
KEY STEPS-IMAGE ANALYSIS



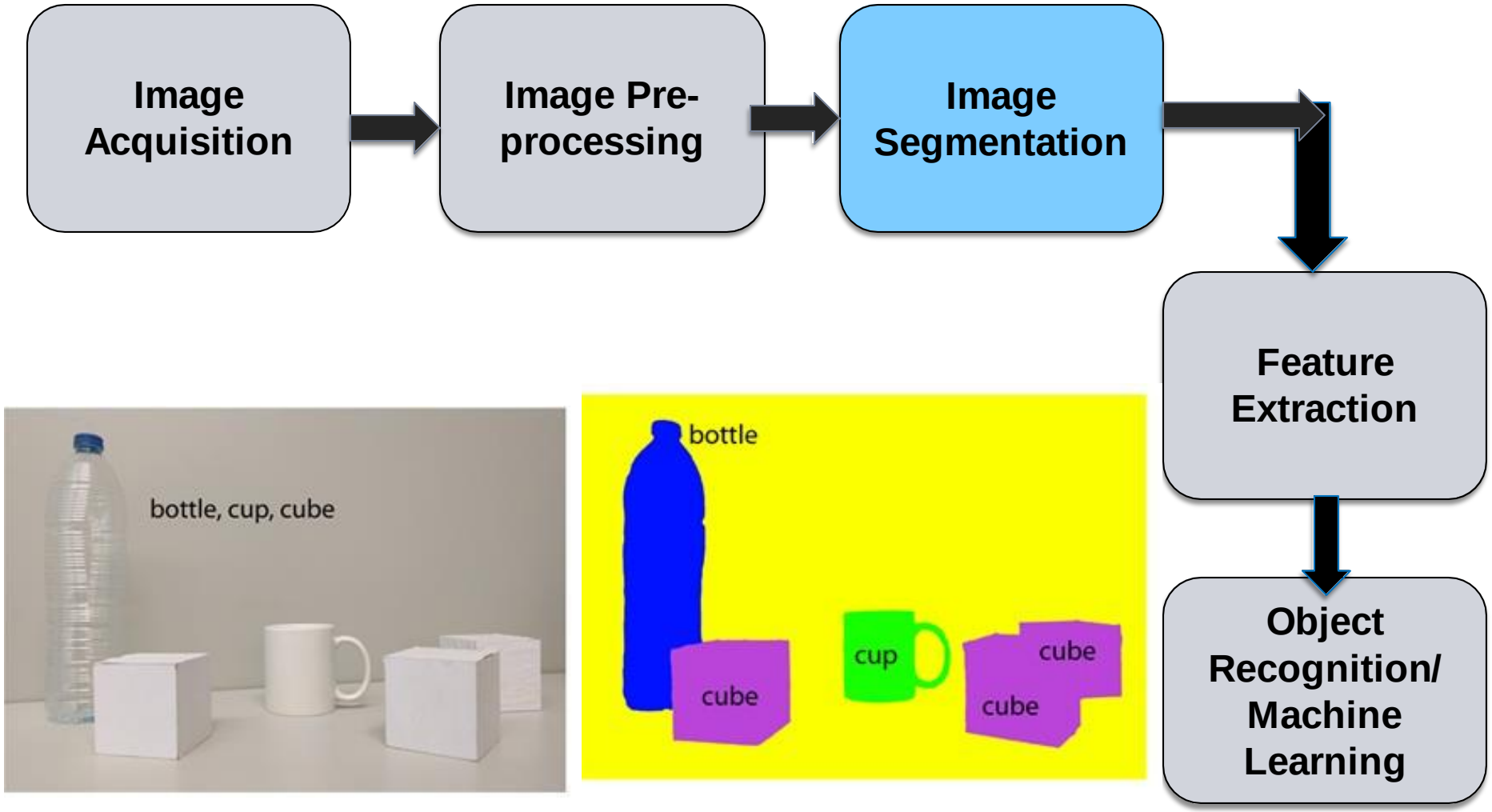
KEY STEPS-IMAGE ANALYSIS



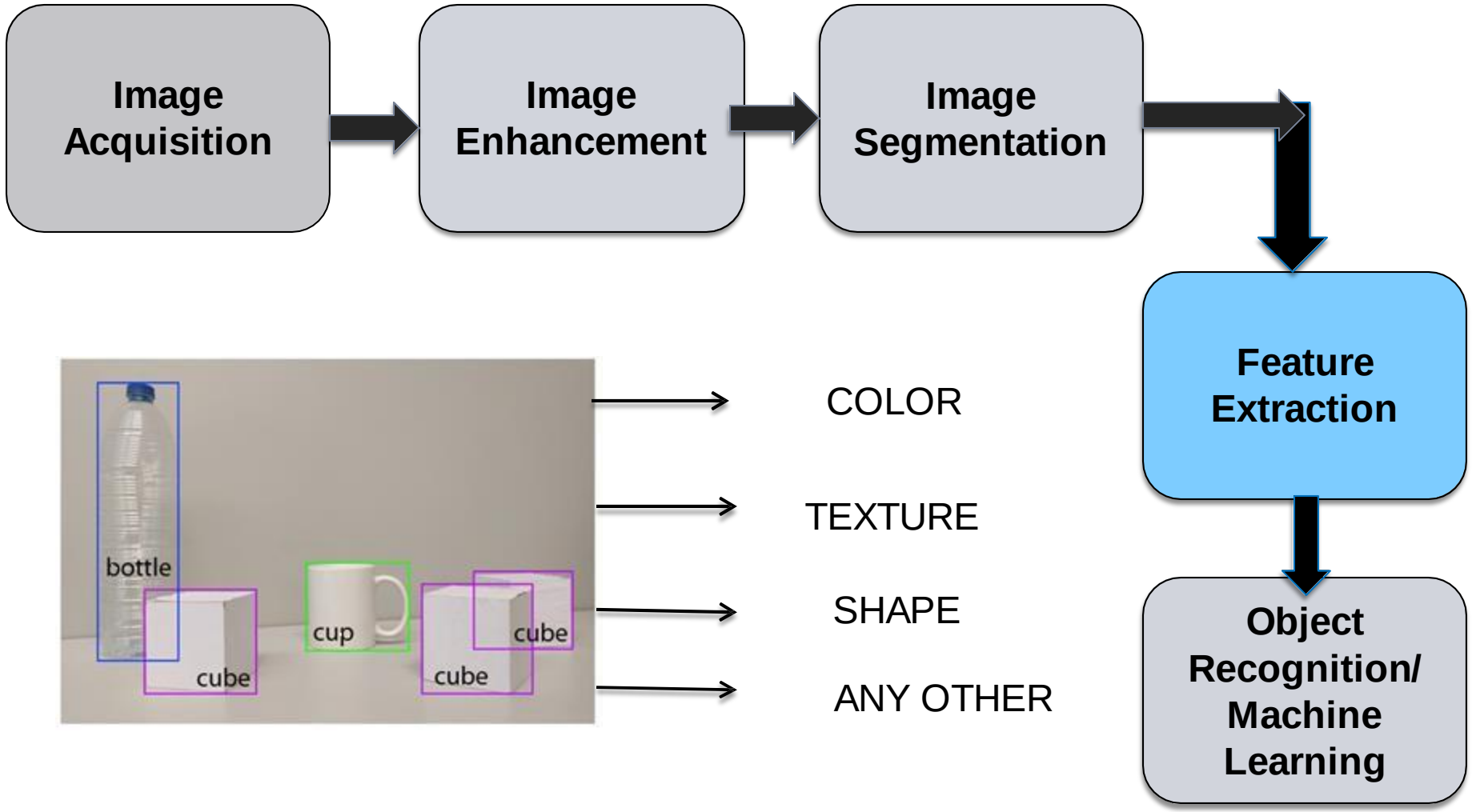
KEY STEPS-IMAGE ANALYSIS



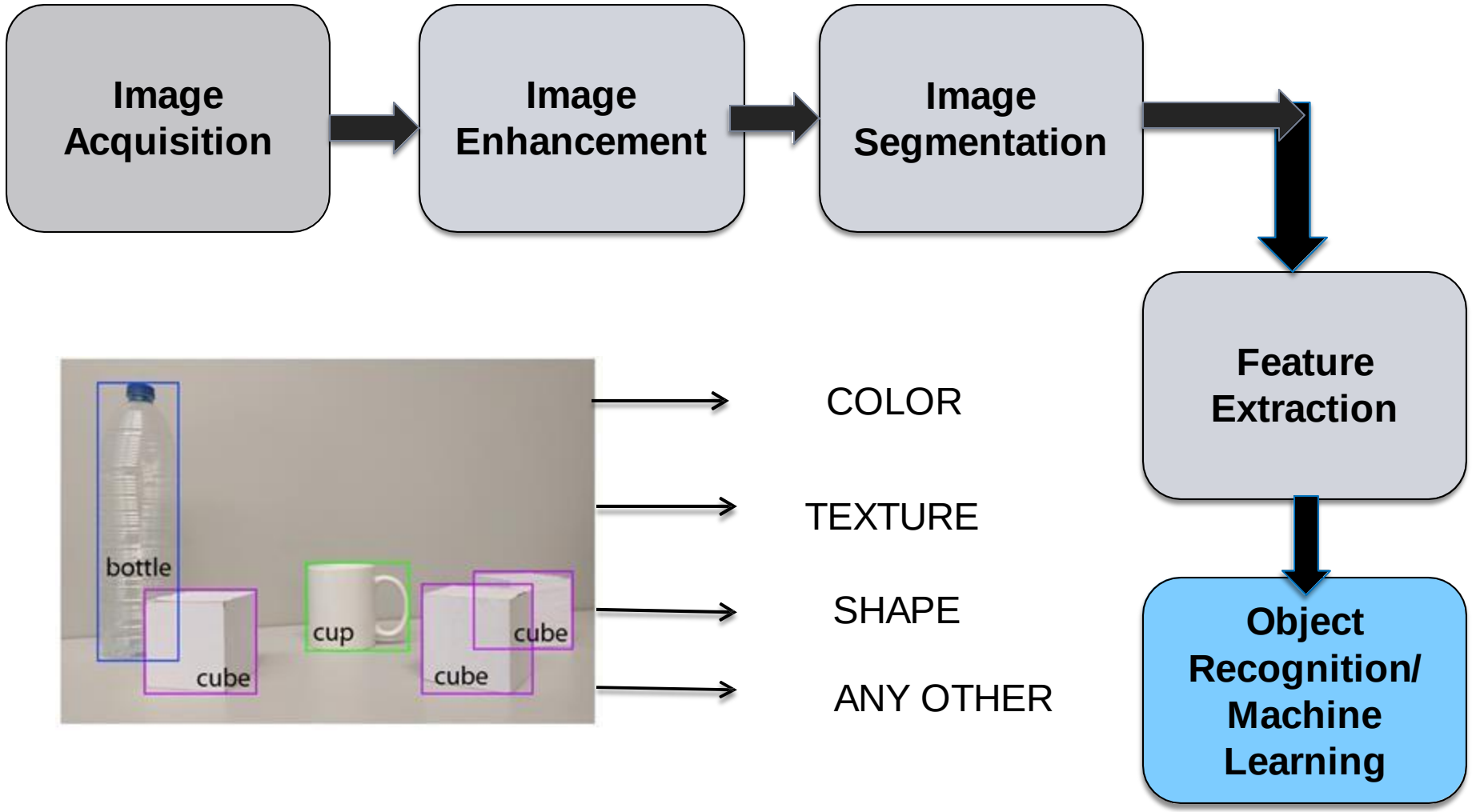
KEY STEPS-IMAGE ANALYSIS



KEY STEPS-IMAGE ANALYSIS



KEY STEPS-IMAGE ANALYSIS



KEY STEPS-IMAGE ANALYSIS

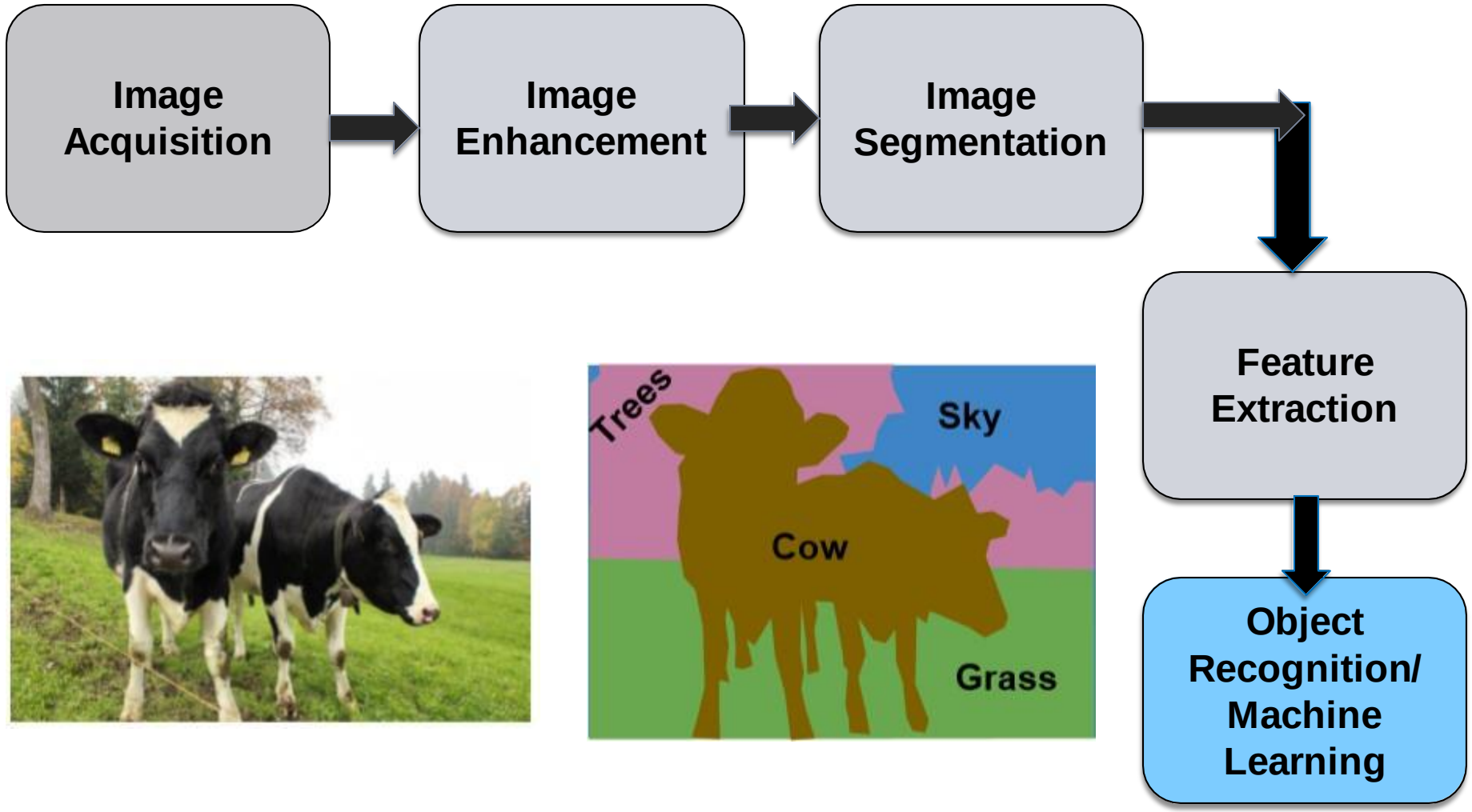


Image Matrix- pixel, resolution(spatial, graylevel)

$F(x,y)$

148	123	52	107	123	162	172	123	64	89
147	130	92	95	98	130	171	155	169	163
141	118	121	148	117	107	144	137	136	134
82	106	93	172	149	131	138	114	113	129
57	101	72	54	109	111	104	135	106	125
138	135	114	82	121	110	34	76	101	111
138	102	128	159	168	147	116	129	124	117
113	89	89	109	106	126	114	150	164	145
120	121	123	87	85	70	119	64	79	127
145	141	143	134	111	124	117	113	64	112

$I(u,v)$

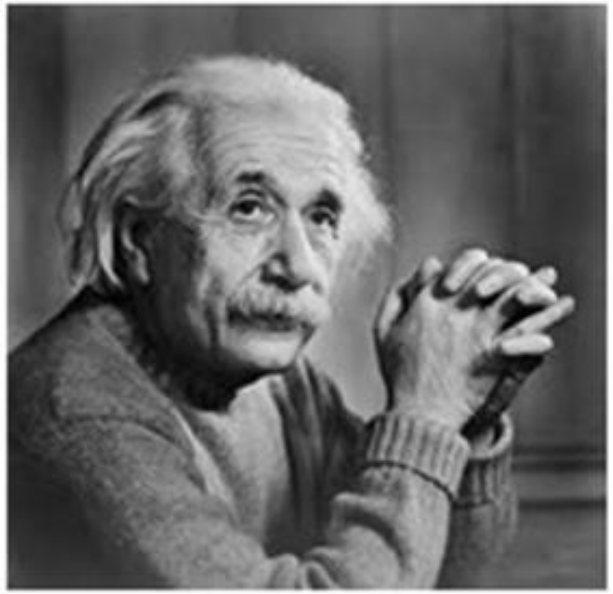
	165	187	209	58	7	
	14	125	233	201	98	159
253	144	120	251	41	147	204
67	100	32	241	23	165	30
209	118	124	27	59	201	79
210	236	105	169	19	218	156
35	178	199	197	4	14	218
115	104	34	111	19	196	
32	69	231	203	74		

Color image

Gray Scale image 8 bit gray scale image 0-255 0-black, 255- white

Image Preprocessing

Image Negative



- Spatial domain enhancement methods can be generalized as
  - $g(x,y) = T[f(x,y)]$

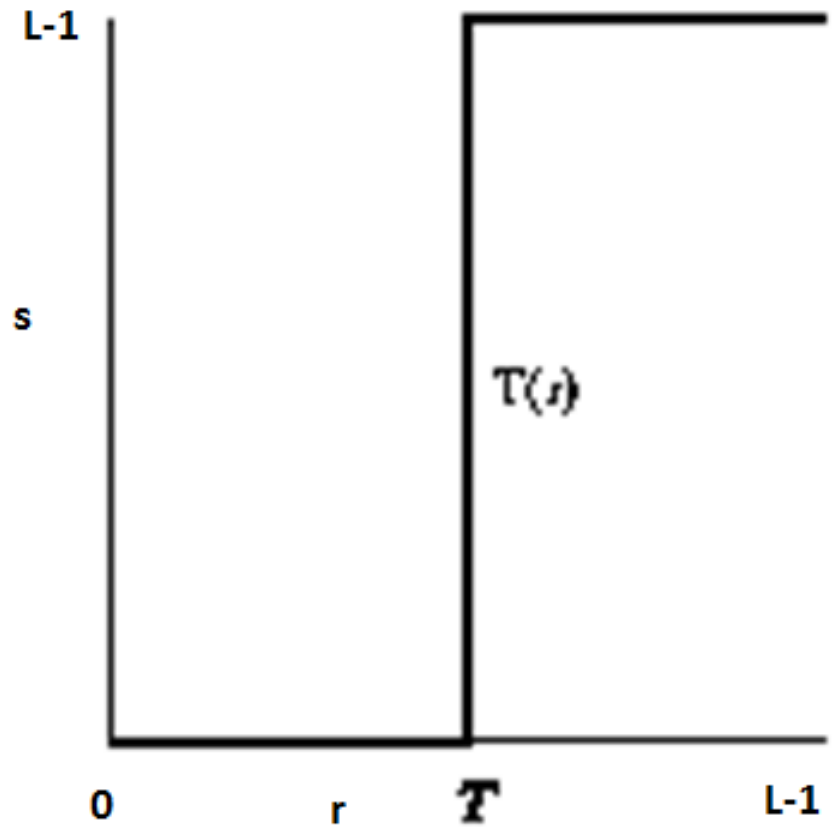
  $f(x,y)$: input image

  $g(x,y)$: processed (output) image

  $T[*]$: an operator on f (or a set of input images), defined over neighborhood of (x,y)

Negative: $s = L - 1 - r$

Thresholding Function

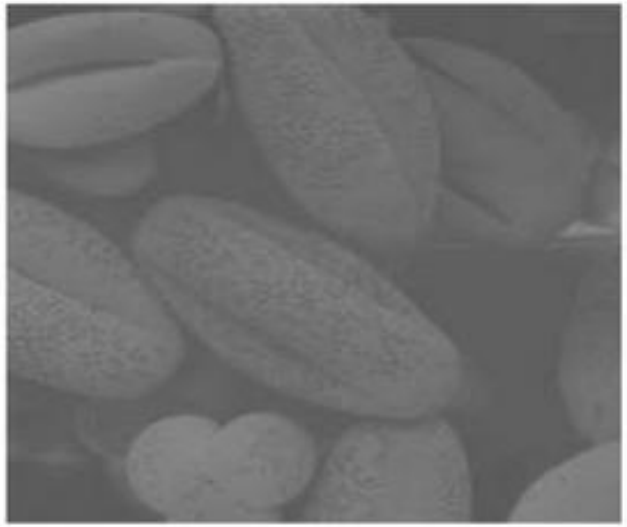


Thresholding Example

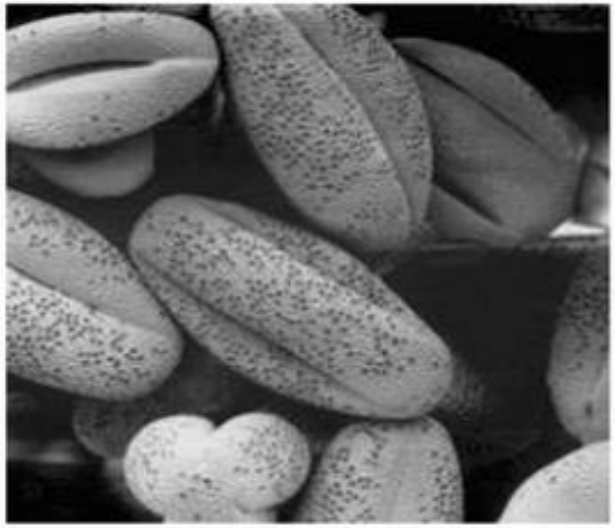
Original Image

Thresholded Image

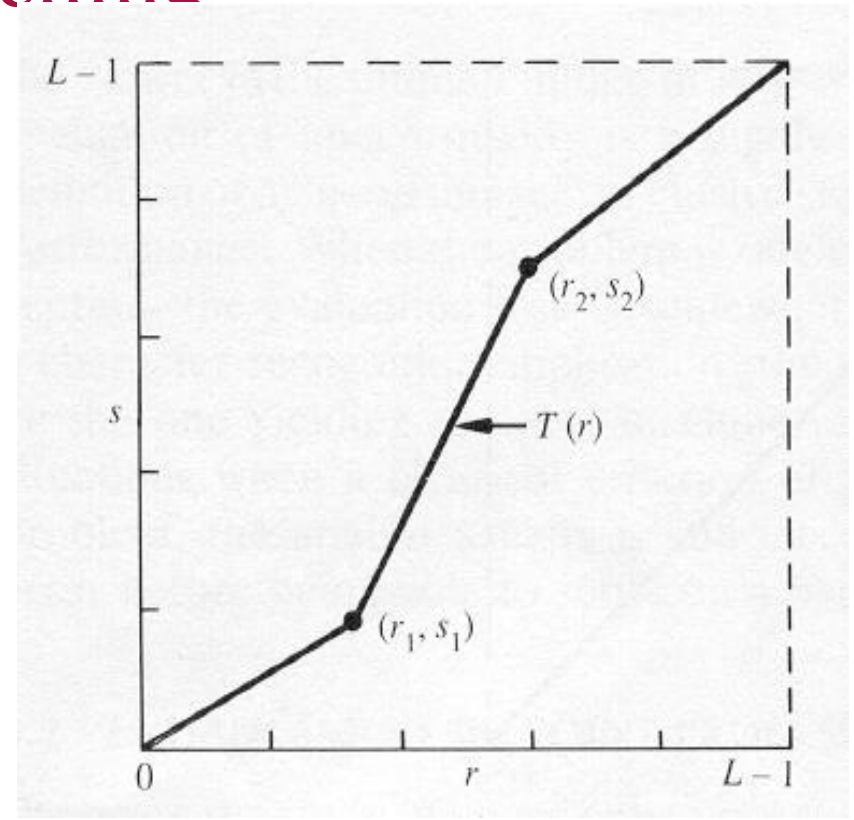
Image Preprocessing- Contrast Stretching



Original Image



Contrast Enhanced Image

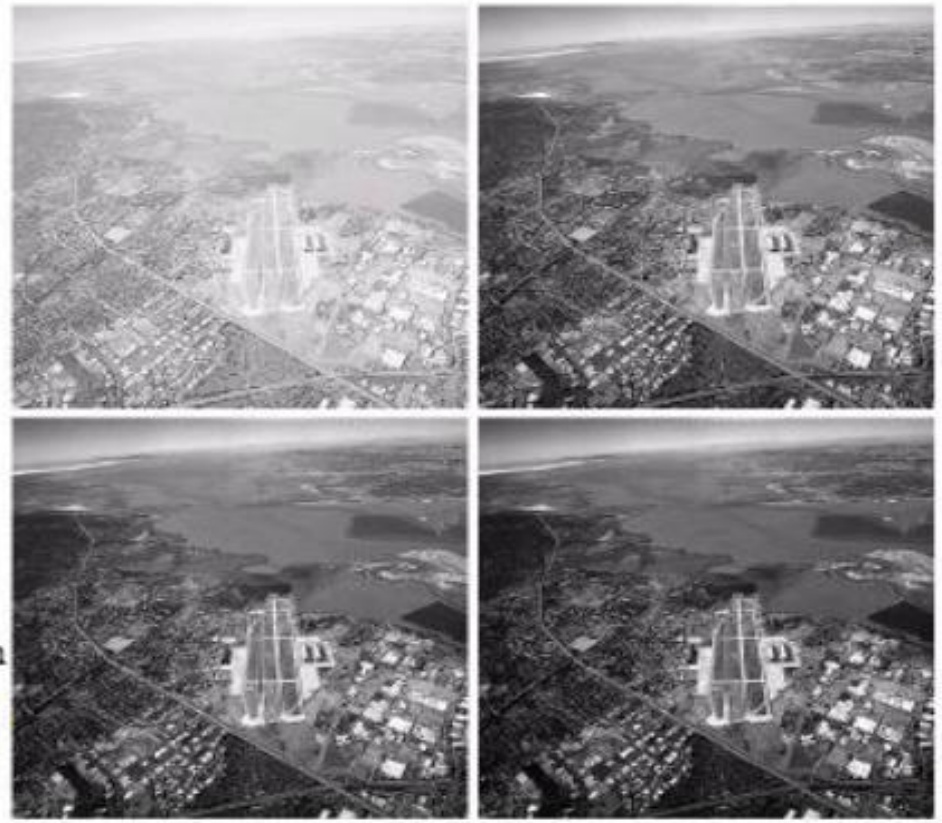


If $r_1 = s_1$ and $r_2 = s_2$ the transformation is a **linear function** and produces no changes.

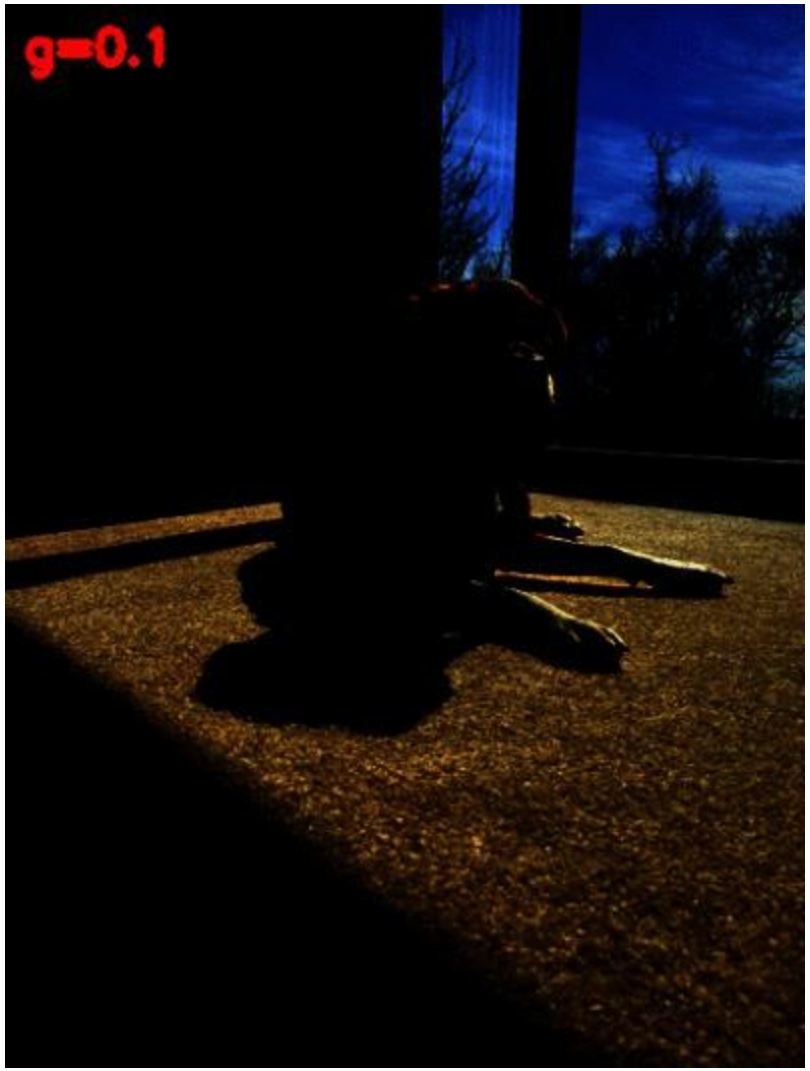
Power Law Transformation: Example

- An aerial photo of a runway is shown
- This time Power Law Transform is used to darken the image
- Different curves highlight different details

$$s = cr^{1/\gamma}$$

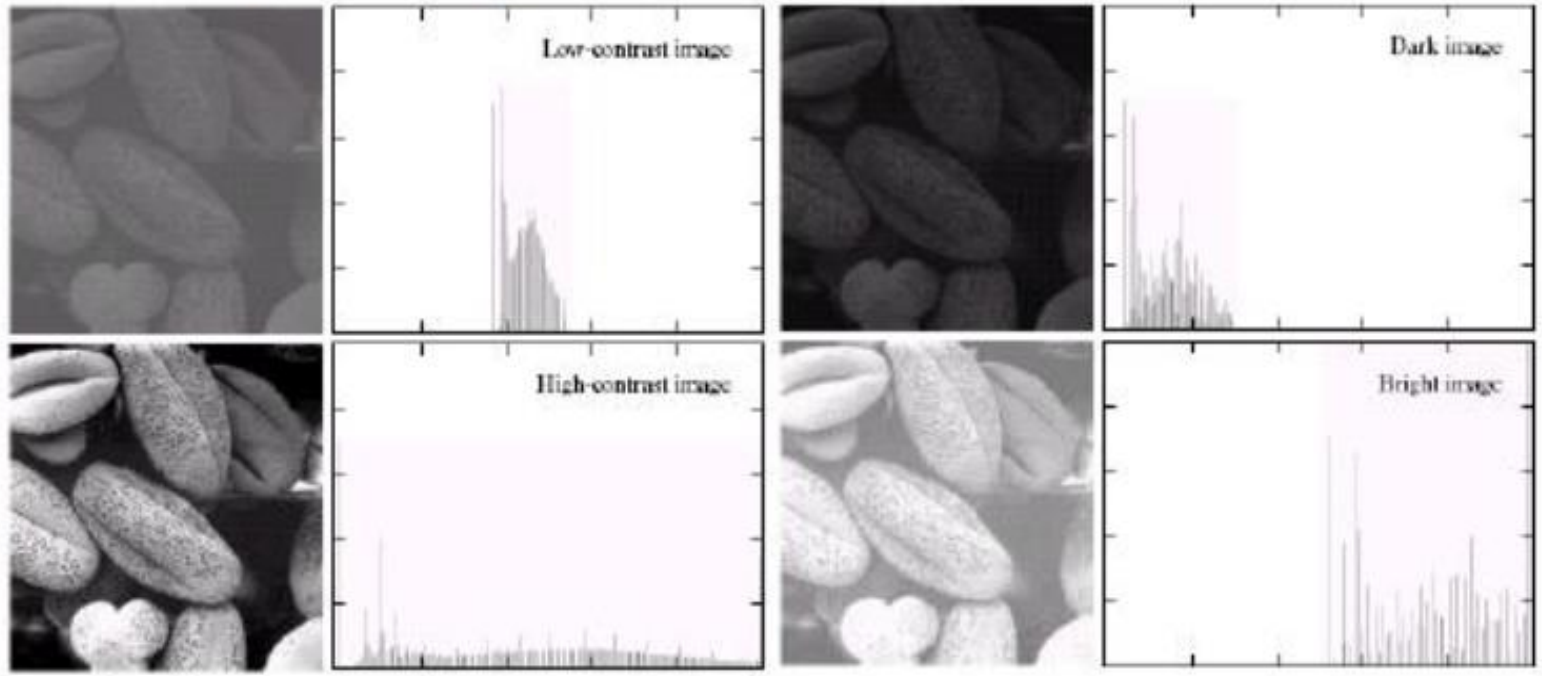


Gama Correction



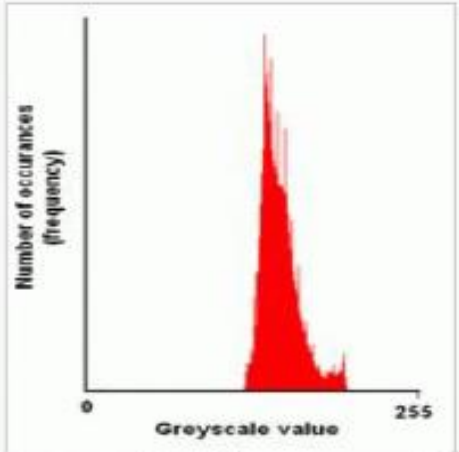
Histogram : Example

- A selection of images and their Histograms
- Note that the high contrast image has the most evenly spaced histogram
- Histograms of low contrast images are located in certain portions and not in the entire gray scale range



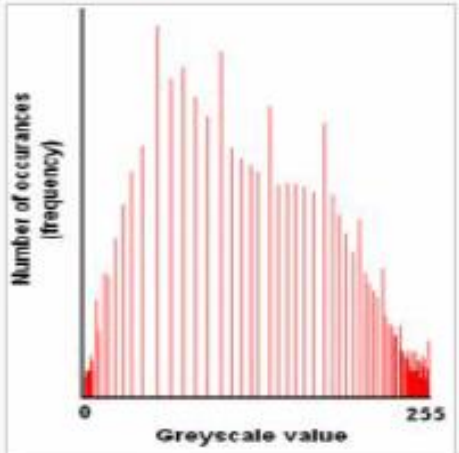
Histogram Equalization: Example

An unequalized image



Corresponding histogram

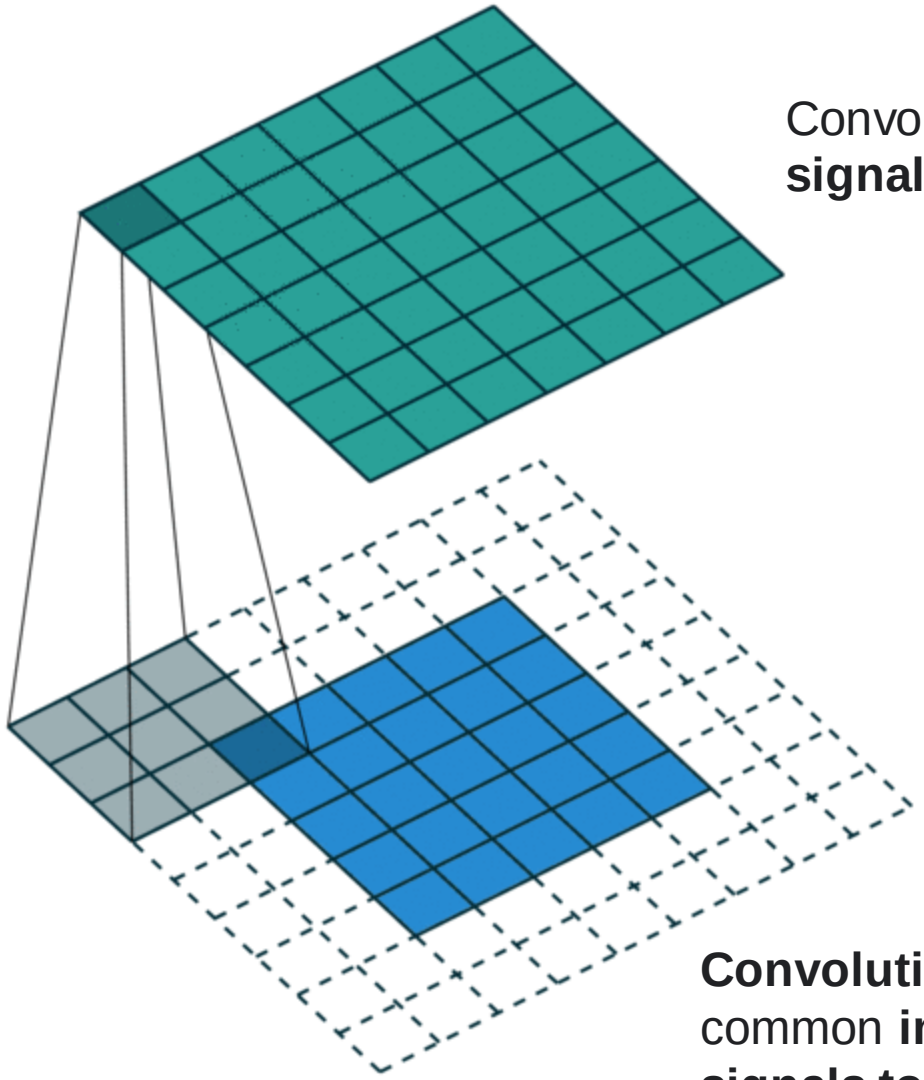
The same image after histogram equalization

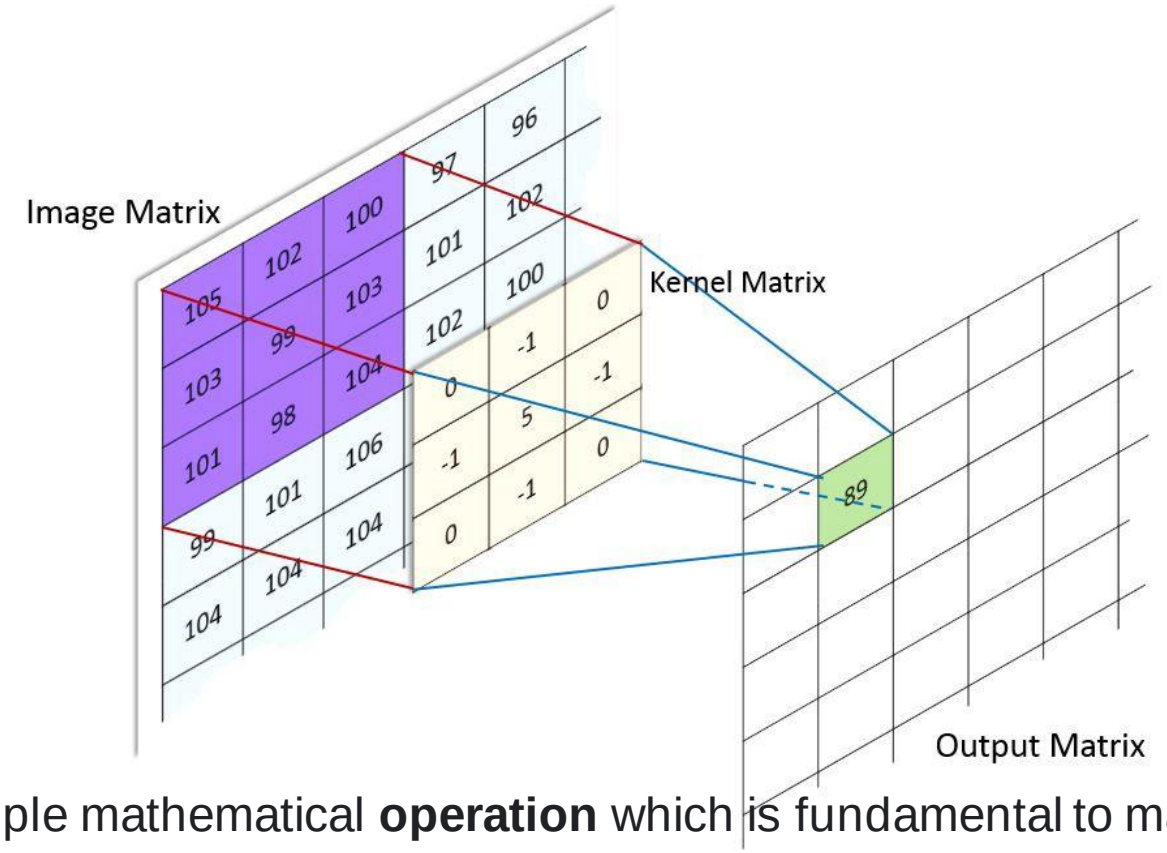


Corresponding histogram

Convolution operation in images

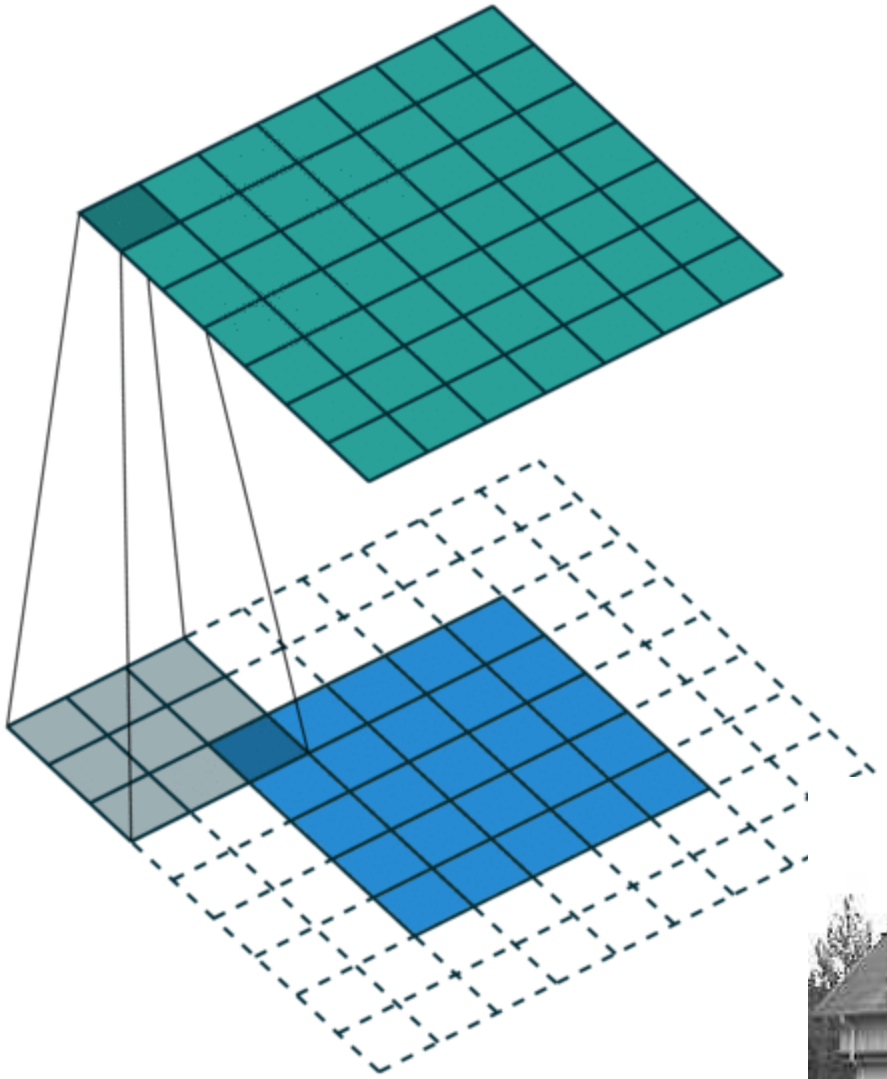
Convolution is **a mathematical way of combining two signals to form a third signal**.





Convolution is a simple mathematical **operation** which is fundamental to many common **image proces**Convolution is **a mathematical way of combining two signals to form a third signal**.**sing** operators. **Convolution** provides a way of `multiplying together' two arrays of numbers, generally of different sizes, but of the same dimensionality, to produce a third array of numbers of the same dimensionality

Convolution operation in images



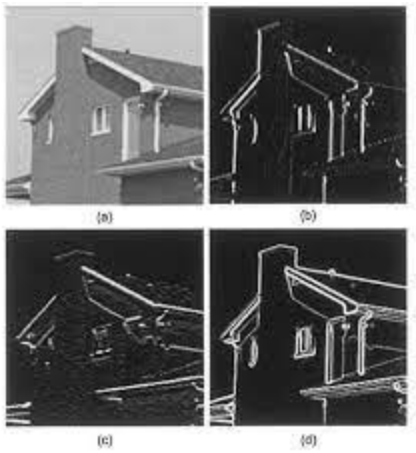
Image

100	100	200	200
100	100	200	200
100	100	200	200
100	100	200	200

Kernel/Filter

-1	0	1
-2	0	2
-1	0	1

-100
-200
-100
200
400
+200
=400



Edge detection using sobel operator

We have seen that convolving an input of 6 X 6 dimension with a 3 X 3 filter results in 4 X 4 output.

- **Input:** n X n
- **Filter size:** f X f
- **Output:** (n-f+1) X (n-f+1)

So, convolving a 6 X 6 input with a 3 X 3 filter gave us an output of 4 X 4. Consider one more example:

10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0

6 X 6 image

*

1	0	-1
1	0	-1
1	0	-1

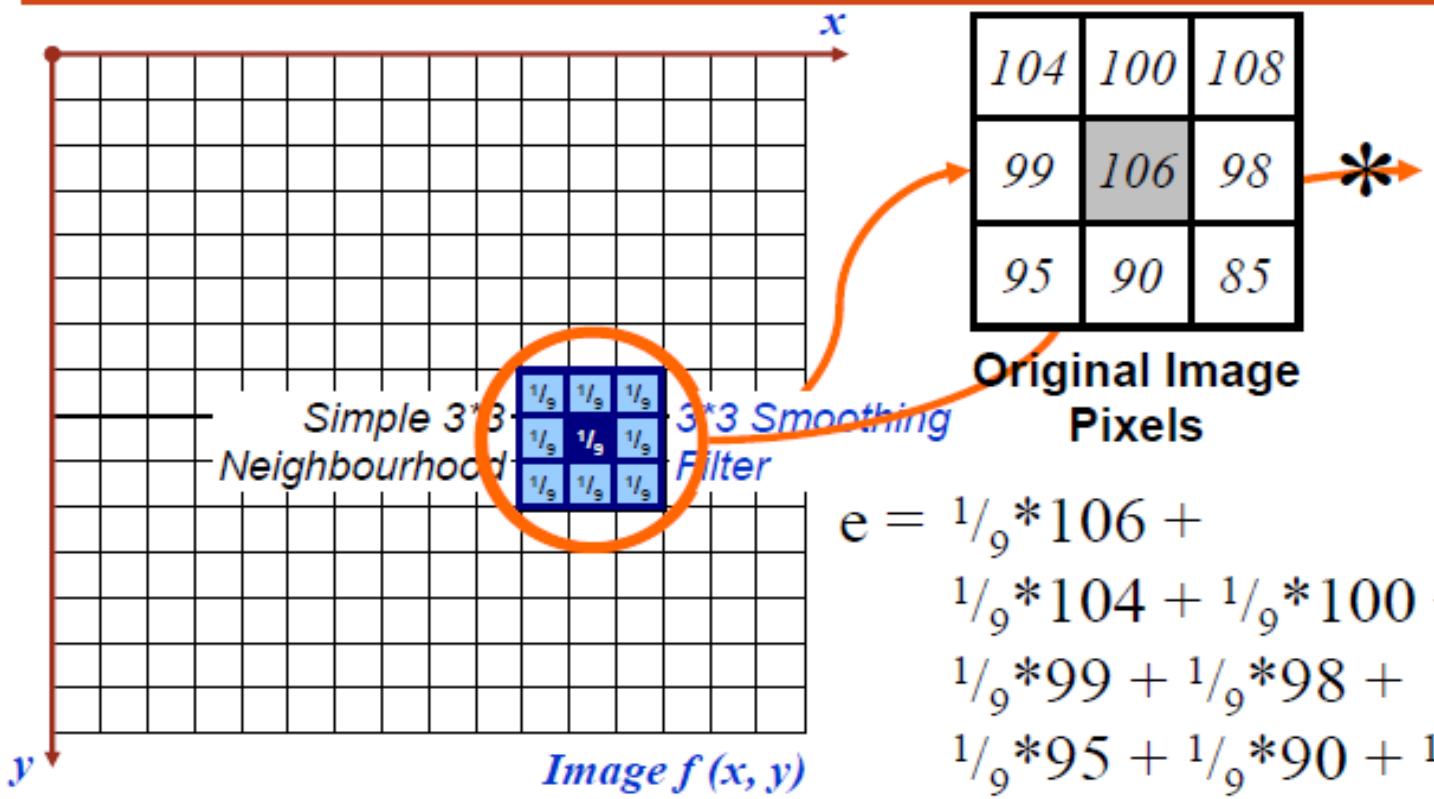
3 X 3 filter

=

0	30	30	0
0	30	30	0
0	30	30	0
0	30	30	0

4 X 4 matrix

Smoothing Spatial Filters



$$e = {}^1/_9*106 + {}^1/_9*104 + {}^1/_9*100 + {}^1/_9*108 + {}^1/_9*99 + {}^1/_9*98 + {}^1/_9*95 + {}^1/_9*90 + {}^1/_9*85 = 98.3333$$

- The above is repeated for every pixel in the original image to generate the smoothed image

Smoothing Spatial Filters

- One of the simplest spatial filtering operations we can perform is a smoothing operation
  - Simply average all of the pixels in a neighbourhood around a central value
  - Especially useful in removing noise from images

$^1/_9$	$^1/_9$	$^1/_9$
$^1/_9$	$^1/_9$	$^1/_9$
$^1/_9$	$^1/_9$	$^1/_9$

Simple Averaging Filter

or

$$\frac{1}{9}\times\begin{array}{|c|c|c|}\hline 1 & 1 & 1\\\hline 1 & 1 & 1\\\hline 1 & 1 & 1\\\hline\end{array}$$

Common Edge Detectors

•Given a 3*3 region of an image the following edge detection filters can be used

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9

−1	−1	−1		−1	0	1		
0	0	0		−1	0	1		
1	1	1		−1	0	1		

Prewitt

−1	0		0	−1		−1	−2	−1
0	1		1	0		0	0	0
						1	2	1

Roberts

−1	−2	−1		−1	0	1		
0	0	0		−2	0	2		
1	2	1		−1	0	1		

Sobel

Edge Detection Example

Original Image

Horizontal Gradient Component

Vertical Gradient Component

Combined Edge Image

Images taken from Gonzalez & Woods, Digital Image Processing (2002)

Edge Detection Example

Images taken from Gonzalez & Woods, Digital Image Processing (2002)

Edge Detection Example

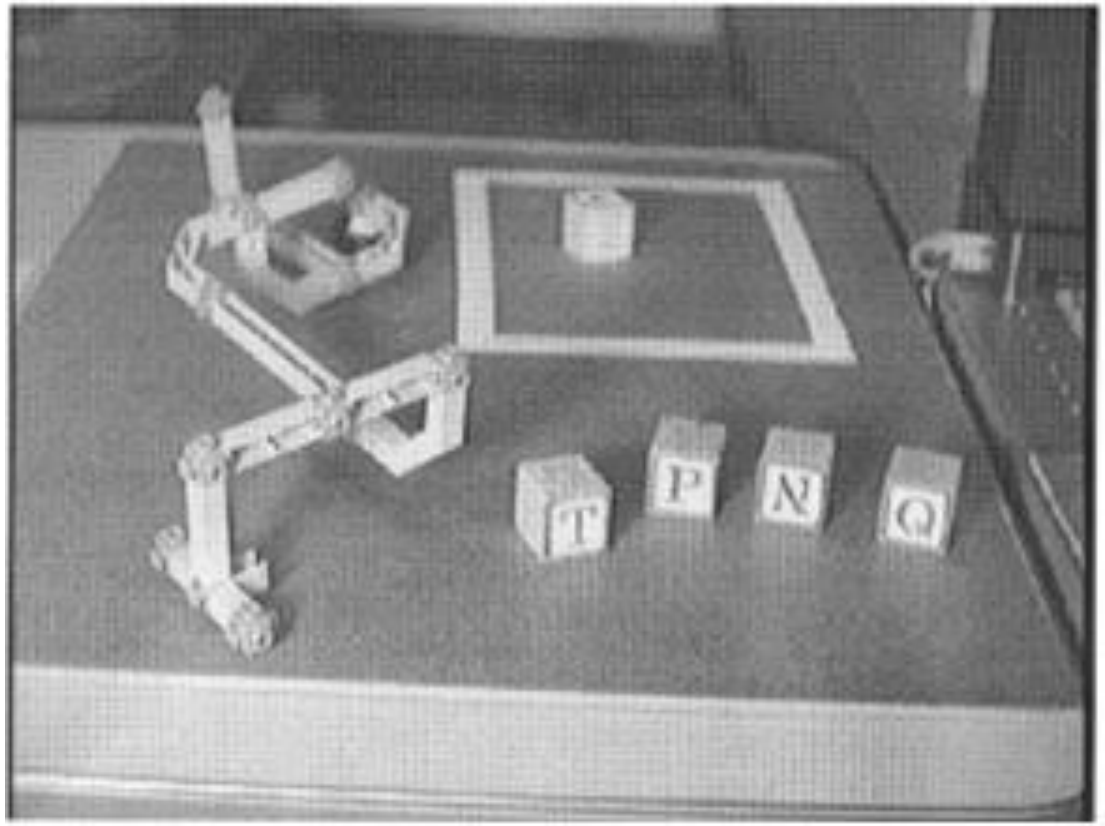
Images taken from Gonzalez & Woods, Digital Image Processing (2002)

Edge Detection Example

Images taken from Gonzalez & Woods, Digital Image Processing (2002)

Edge Detection Example

Images taken from Gonzalez & Woods, Digital Image Processing (2002)


T
P
N
Q

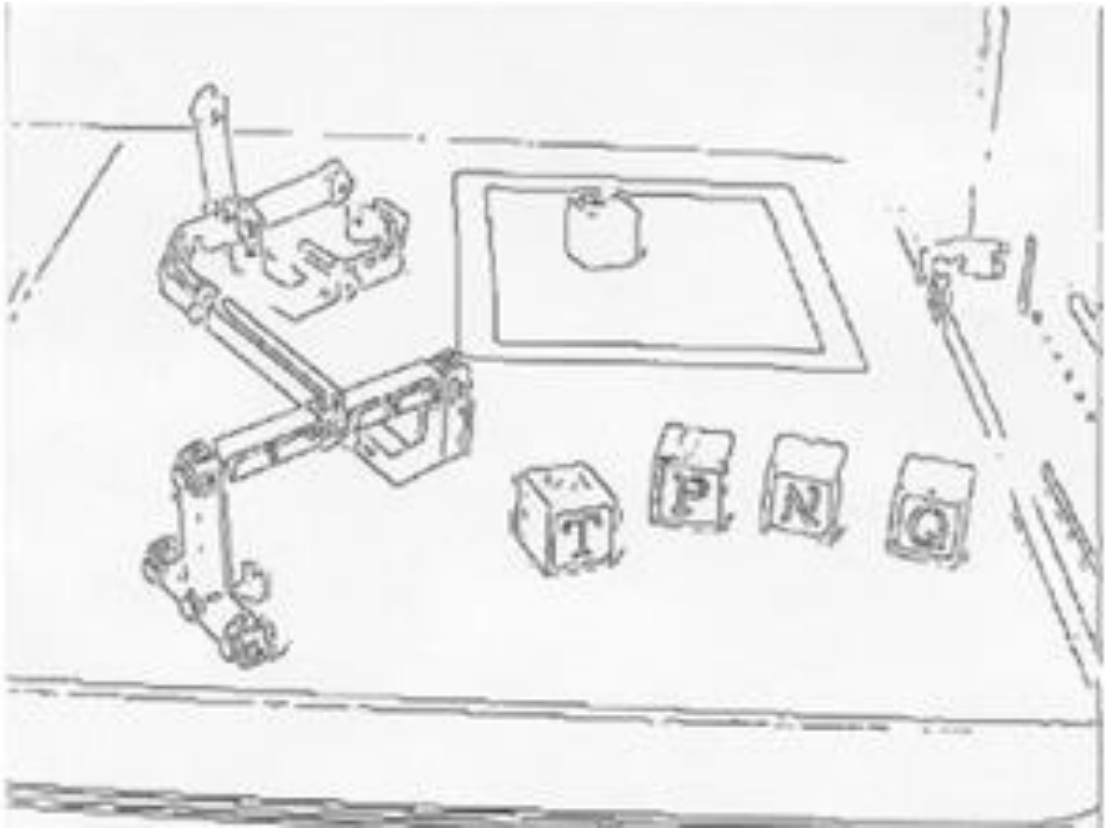
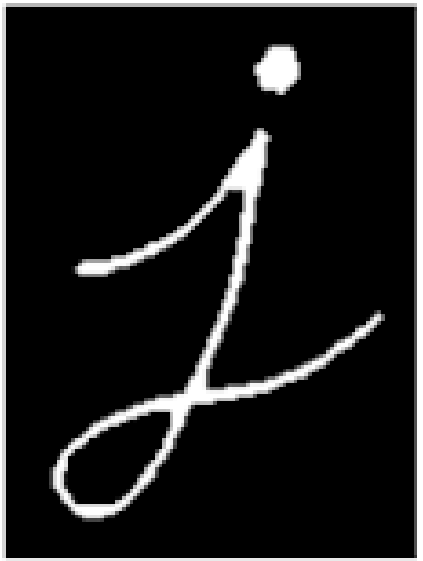
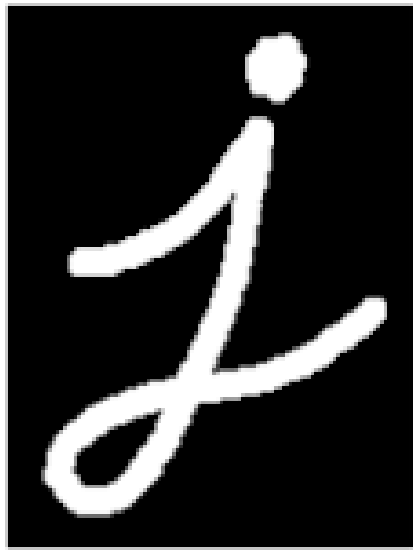

T
P
N
Q

Image Preprocessing- Erosion and dilation



erosion

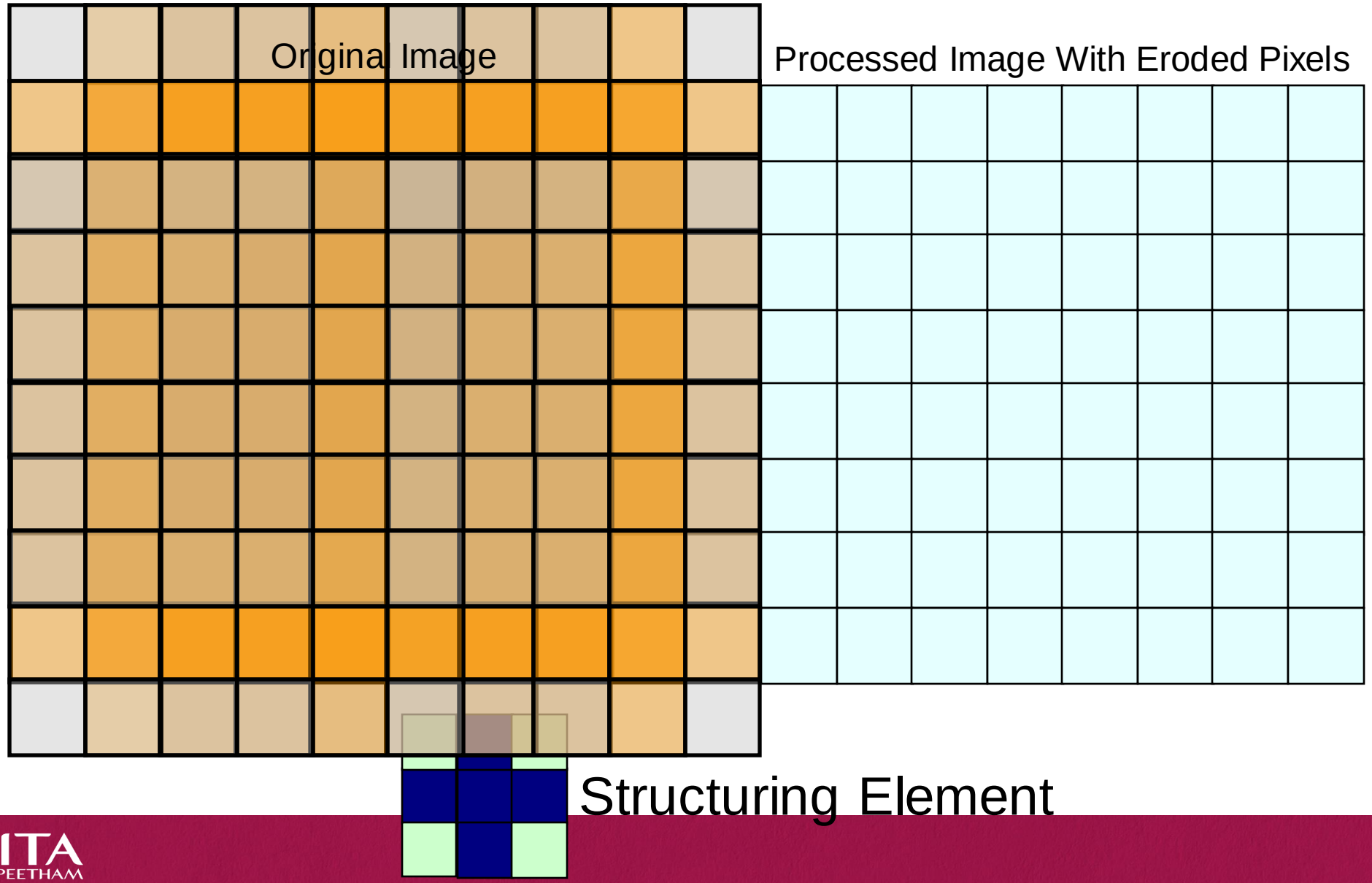


original

dilation

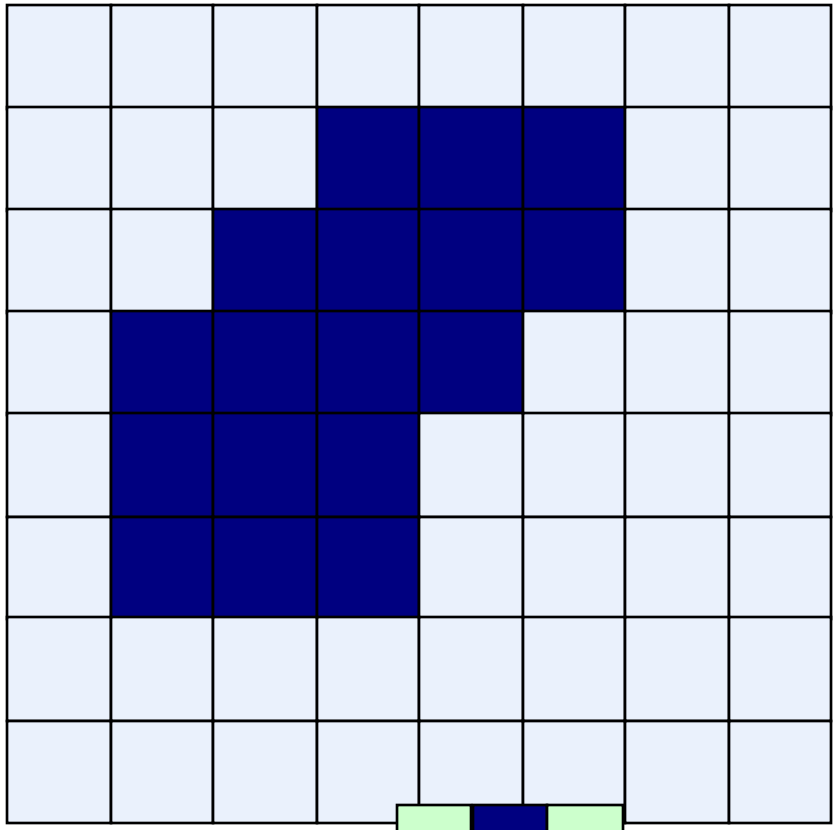
字母"j":(左）侵蚀，（中）原始图像，（右）扩张

Erosion Example

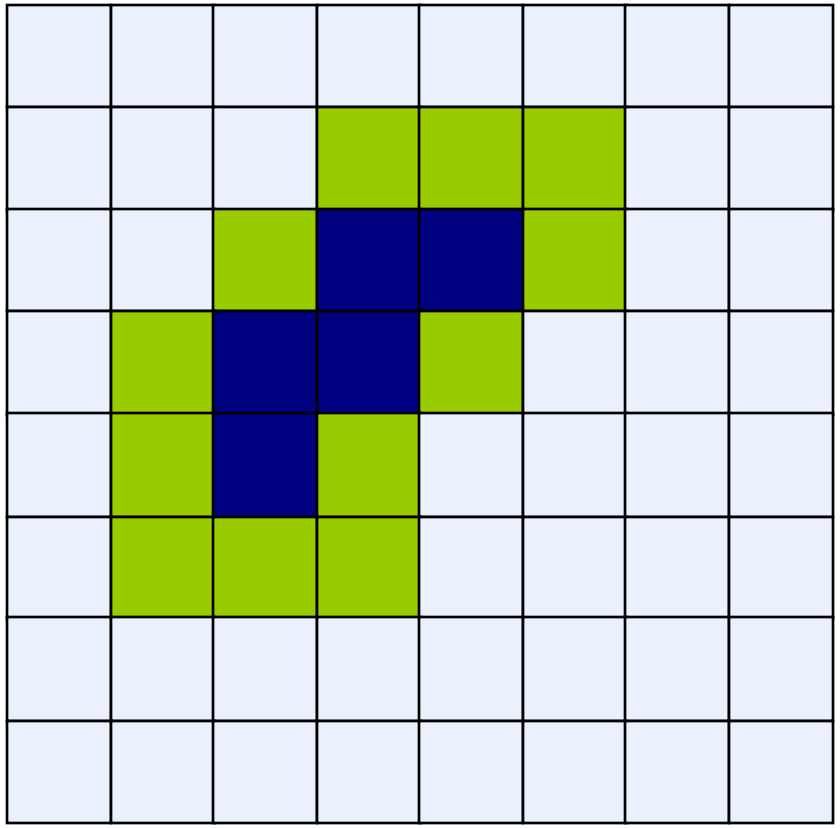


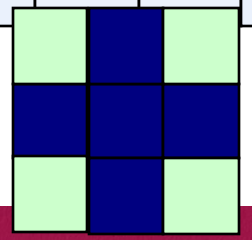
Erosion Example

Original Image



Processed Image



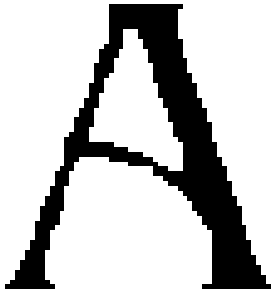


Structuring Element

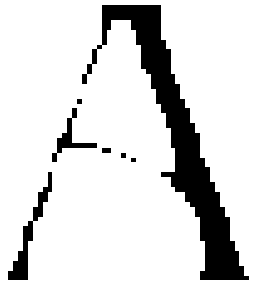
Erosion Example 1

Watch out: In these examples a 1 refers to a black pixel!

Original image



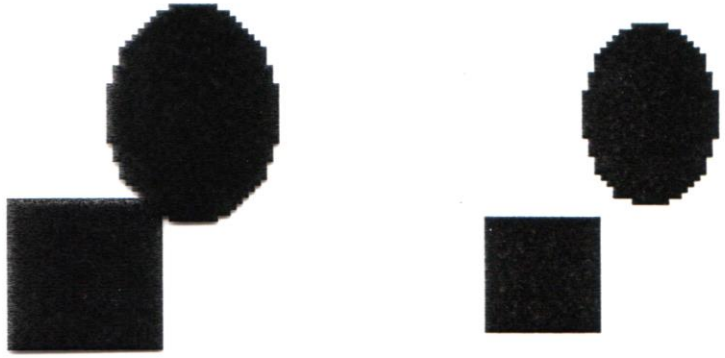
Erosion by 3*3 square structuring element



Erosion by 5*5 square structuring element

What Is Erosion For?

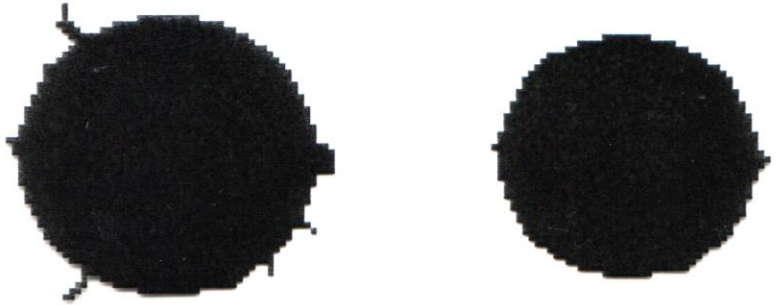
Erosion can split apart joined objects



Erosion can split apart

Erosion can strip away extrusions

Watch out:

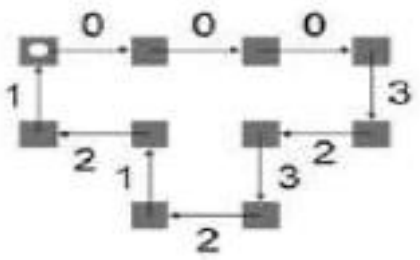


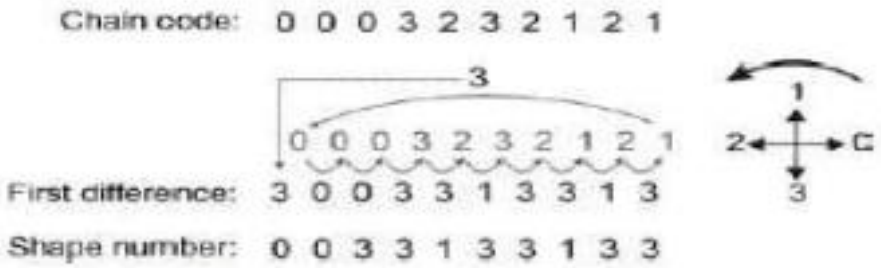
Feature Extraction and Classification

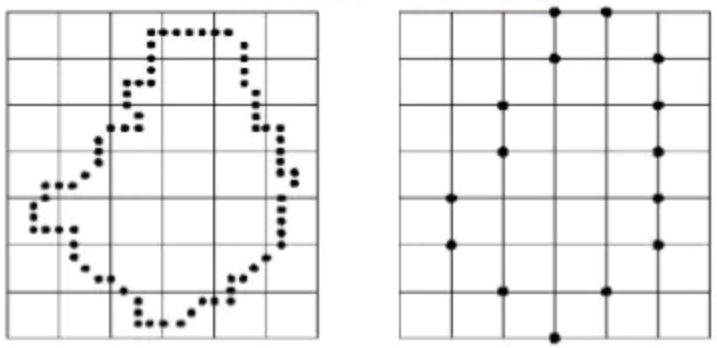
Feature Extraction- Hand Crafted

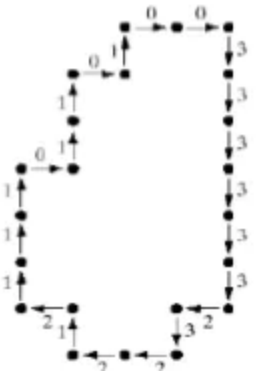
- Chaincode (shape)
- Fourier Descriptors (shape)
- Harris Corner Detection
- Gray Level Co-Occurrence Matrix (GLCM)- texture
- Histogram of Oriented Gradients (HOG)- shape
- Moment based feature (shape)
- Haar Cascades
- Color features- color spaces
- Scale-Invariant Feature Transform (SIFT)- keypoint based
- Speeded Up Robust Feature (SURF)- keypoint
- Lot more…..

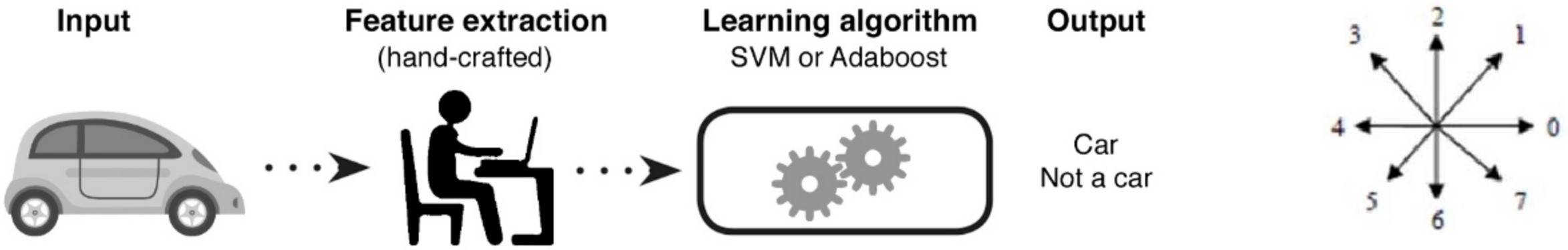
A simple example –chain code







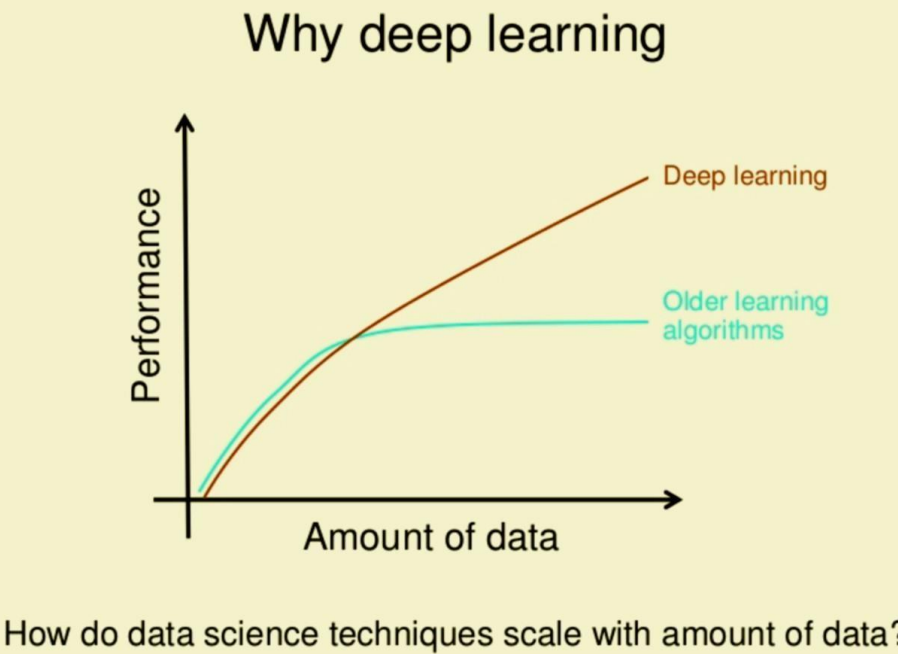


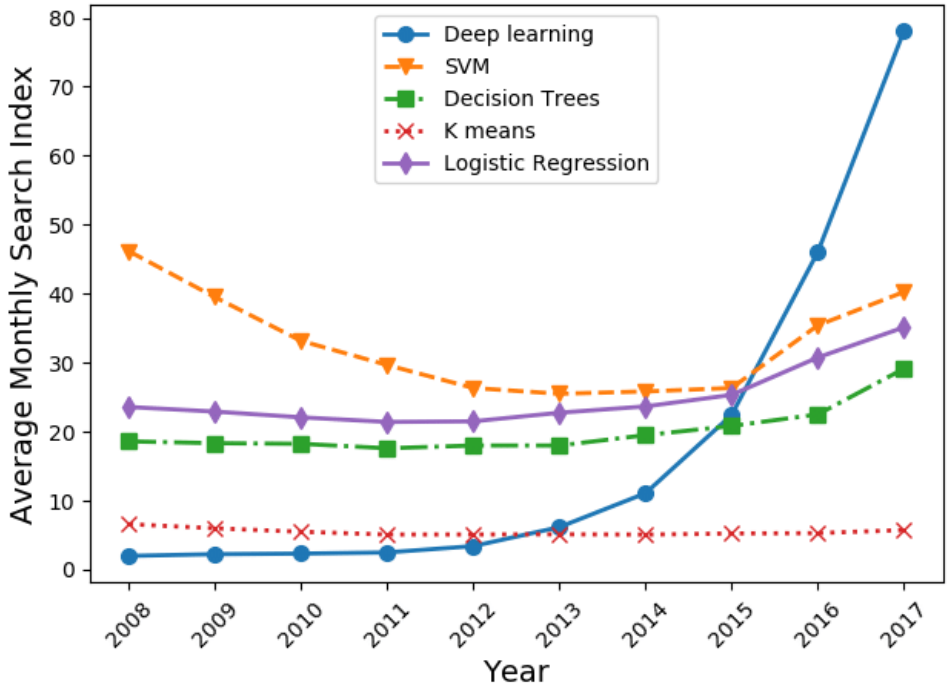


Deep Learning- A breakthrough

Given the Availability of Data, DeepLearning performance has surpassed all traditional algorithms

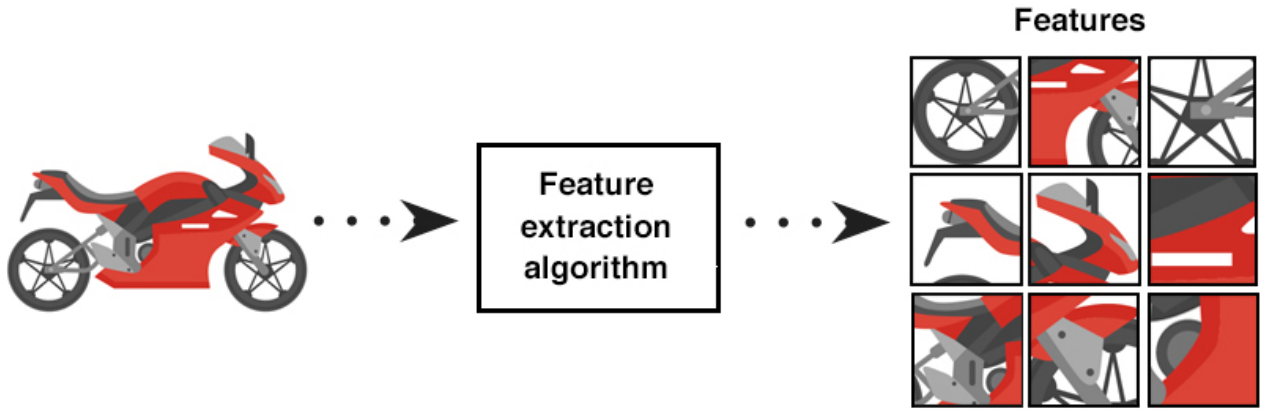
Artificial Intelligence Paradigmshift!-
Machine Learning and **Deep Learning?**



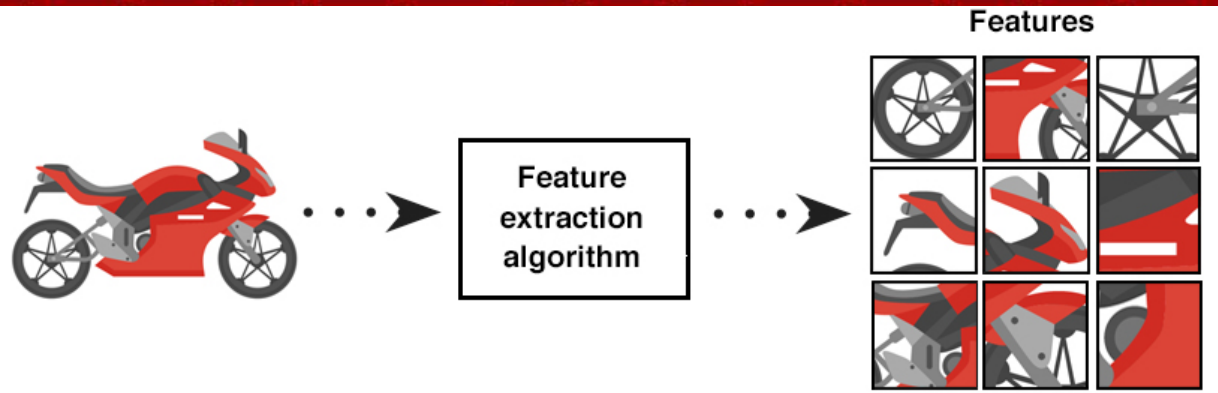


Promising results if trained with lot of data!

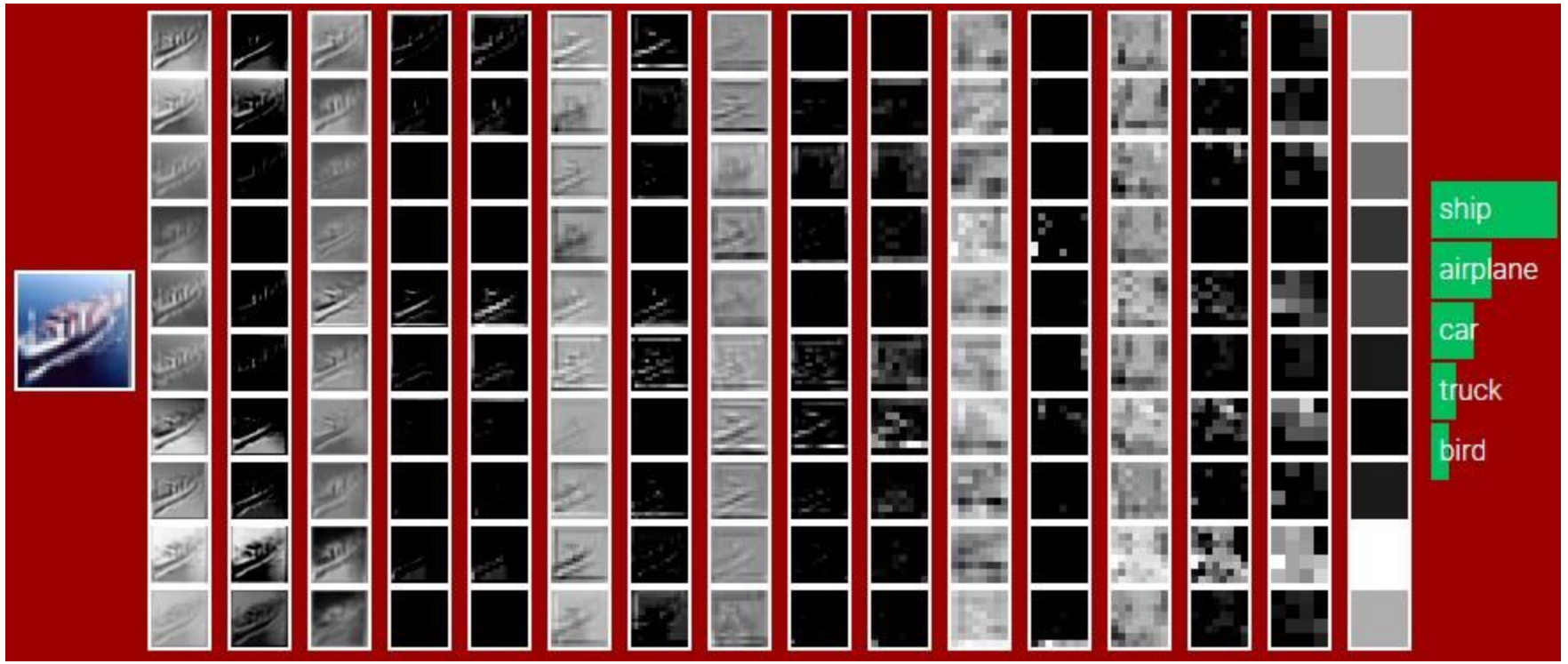
Deep Learning

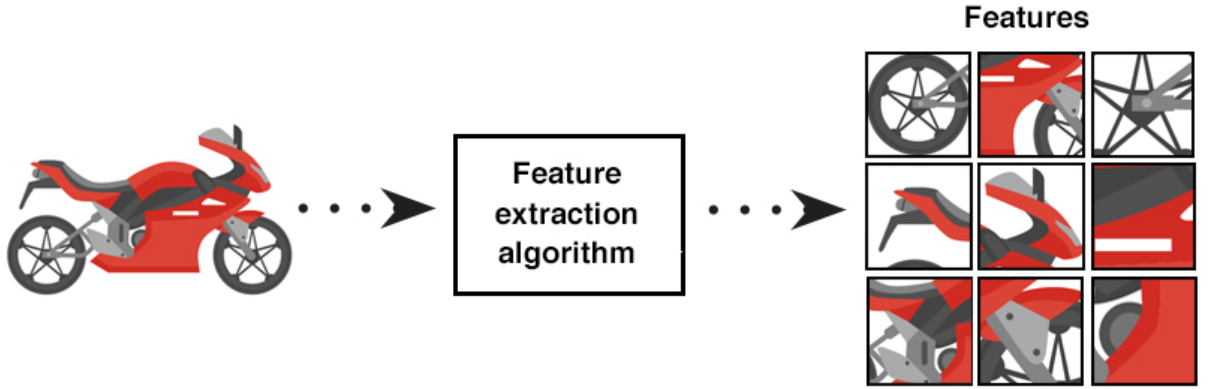


Deep Learning-



Deep Learning-





Thank you