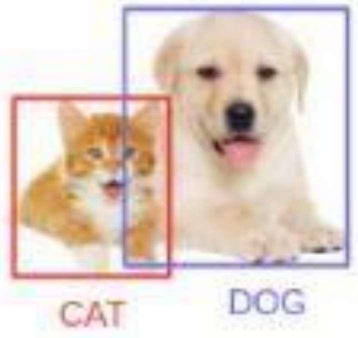
Computer Vision (CV) and Natural Language Processing (NLP)

Refer to class notes for Introduction section for the following topics

Computer vision, Object detection and Image segmentation

Image Segmentation

Image Localization



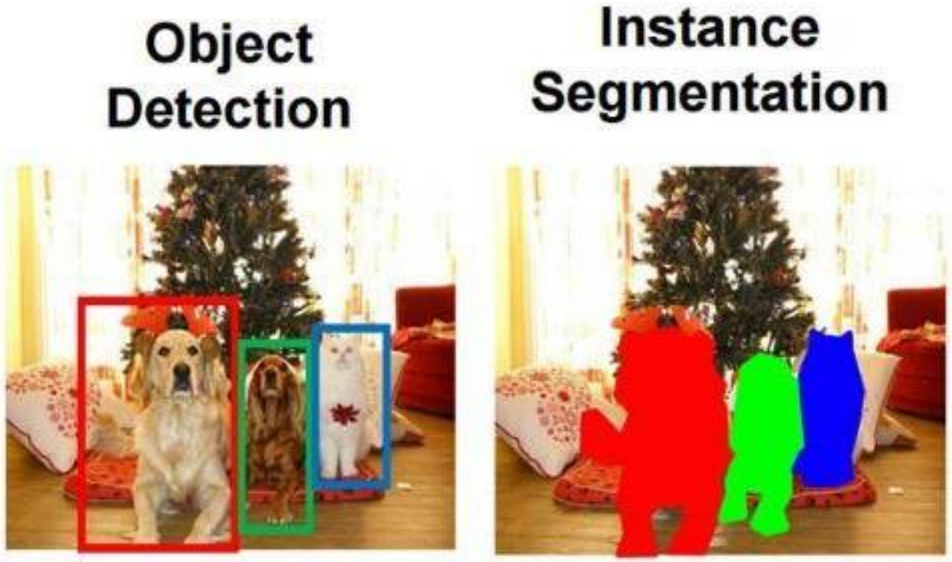
Object Detection

- Image localization helps us to identify the location of a single object in the given image.
- In case we have multiple objects present, we then rely on the concept of object detection (OD). We can predict the location along with the class for each object using OD.
- Before detecting the objects and even before classifying the image, we need to understand what the image consists of. Enter – Image Segmentation.

How does image segmentation work?

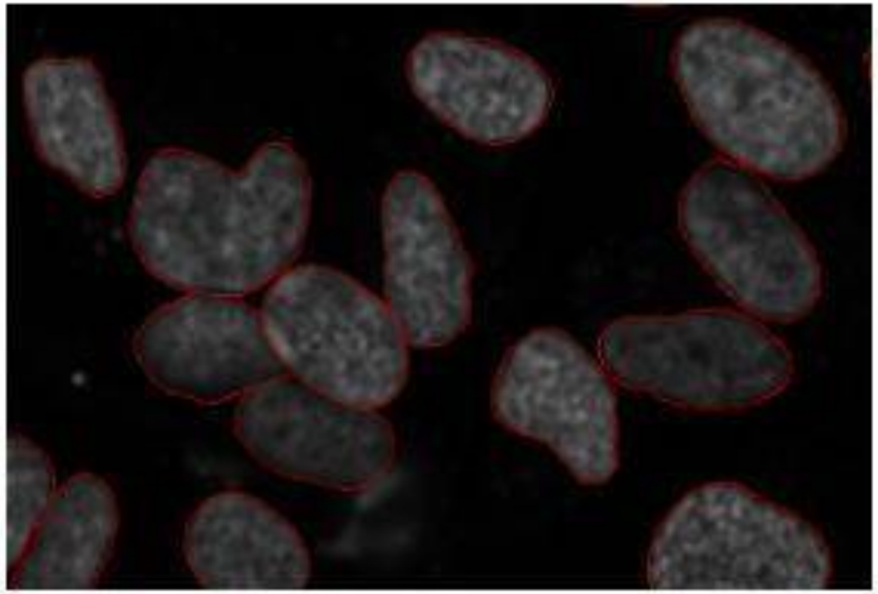
- We can divide or partition the image into various parts called segments.
- It’s not a great idea to process the entire image at the same time as there will be regions in the image which do not contain any information. By dividing the image into segments, we can make use of the important segments for processing the image.
- An image is a collection or set of different pixels. We group together the pixels that have similar attributes using image segmentation.

Object Detection vs Image Segmentation



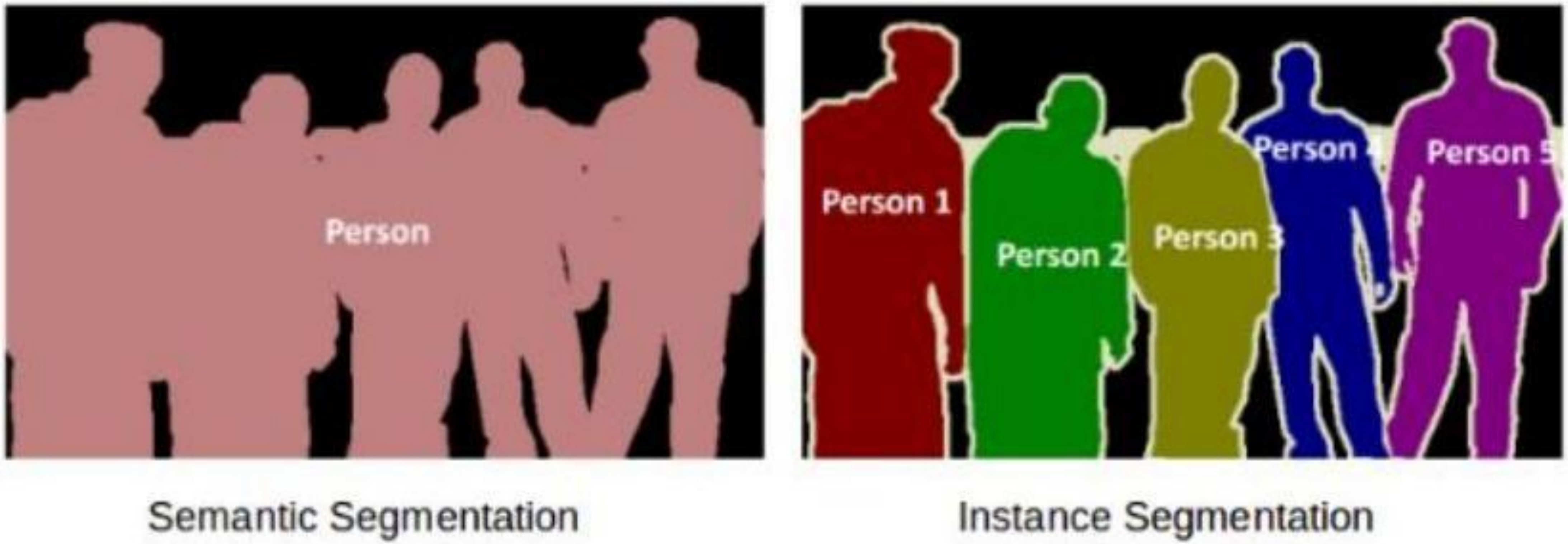
- Object detection builds a bounding box corresponding to each class in the image. But it tells us nothing about the shape of the object.
- Image segmentation creates a pixel-wise mask for each object in the image. This technique gives us a far more granular understanding of the object(s) in the image.

Why do we need Image Segmentation?



- The shape of the cancerous cells plays a vital role in determining the severity of the cancer.
- object detection will not be very useful here. We will only generate bounding boxes which will not help us in identifying the shape of the cells.
- Image Segmentation techniques approach this problem in a more granular manner and get more meaningful results

Different Types of Image Segmentation



Different Types of Image Segmentation

- **Semantic segmentation:** every pixel belongs to a particular class (either background or person). Also, all the pixels belonging to a particular class are represented by the same color (background as black and person as pink). semantic segmentation will focus on classifying all the people as a single instance.
- **Instance segmentation:** has also assigned a particular class to each pixel of the image. However, different objects of the same class have different colors (Person 1 as red, Person 2 as green, background as black, etc.). Instance segmentation, on the other hand. will identify each of these people individually.

Region-based Segmentation

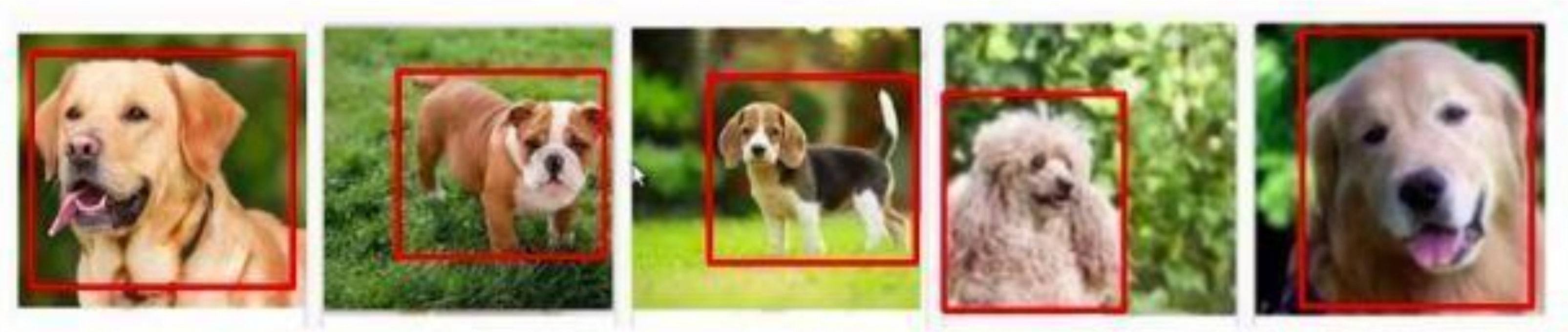
- The pixel values will be different for the objects and the image's background if there's a sharp contrast between them.
- The pixel values falling below or above that threshold can be classified accordingly (as an object or the background). This technique is known as Threshold Segmentation.
- If we want to divide the image into two regions (object and background), we define a single threshold value. This is known as the global threshold.
- If we have multiple objects along with the background, we must define multiple thresholds. These thresholds are collectively known as the local threshold.

Region-based Segmentation : Advantages

- Calculations are simpler
- Fast operation speed
- When the object and background have high contrast, this method performs really well
- limitations: When we don’t have significant grayscale difference, or there is an overlap of the grayscale pixel values, it becomes very difficult to get accurate segments.

Object Detection

- Object detection is a computer vision technique for locating instances of objects in images or videos.



- The box around the object in the image is formally known as a bounding box.
- This becomes an image localization problem where we are given a set of images and we have to identify where is the object present in the image.

Object Detection

- In the case of object detection problems, we have to classify the objects in the image and also locate where these objects are present in the image.

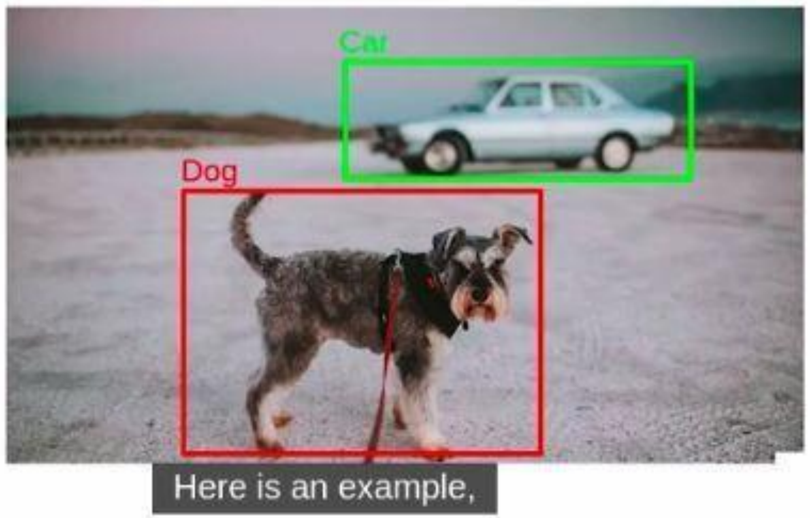
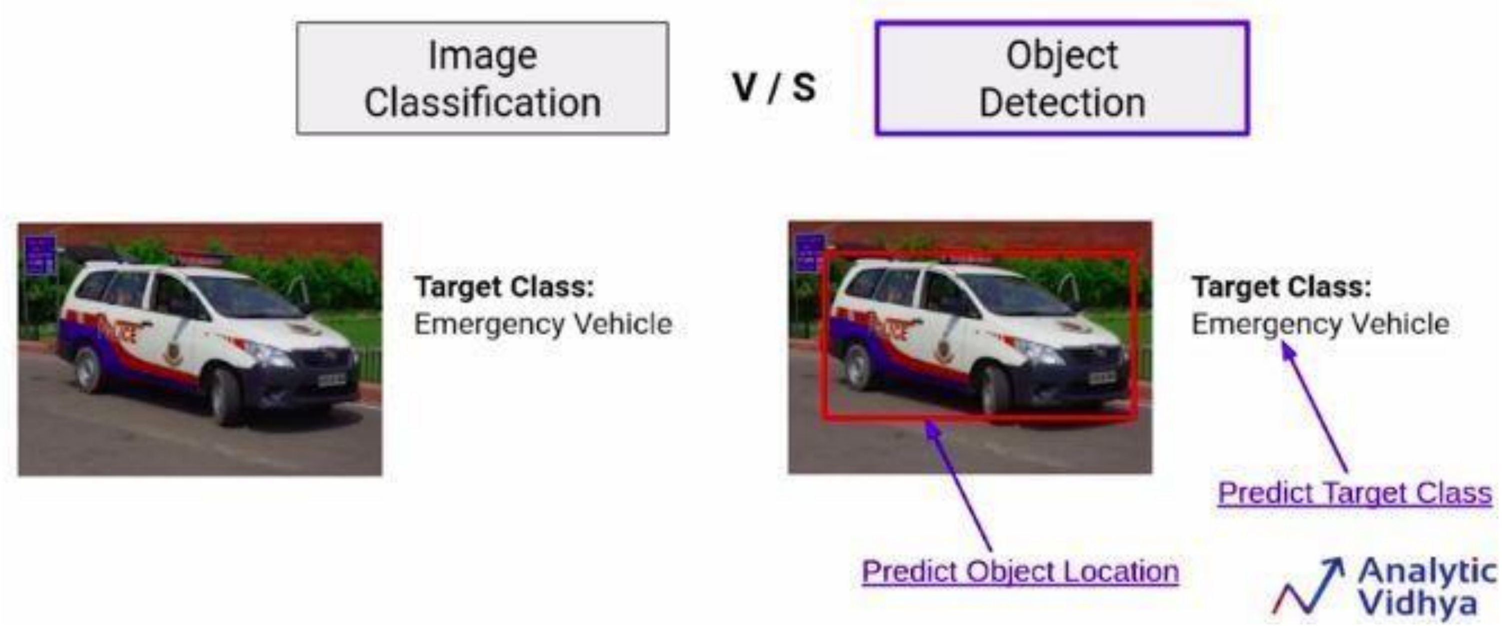
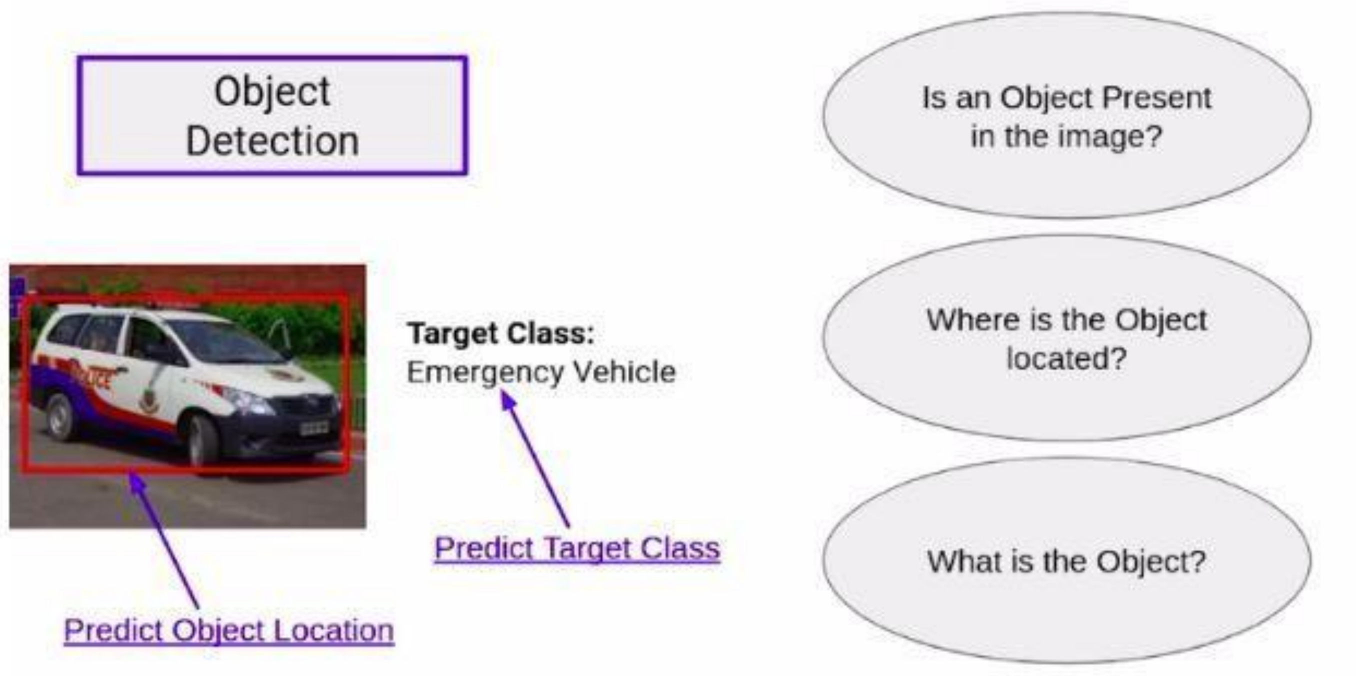


Image Classification vs Object Detection

Image Classification	V / S	Object Detection
• Object Classification		• Object Classification • Object Localization

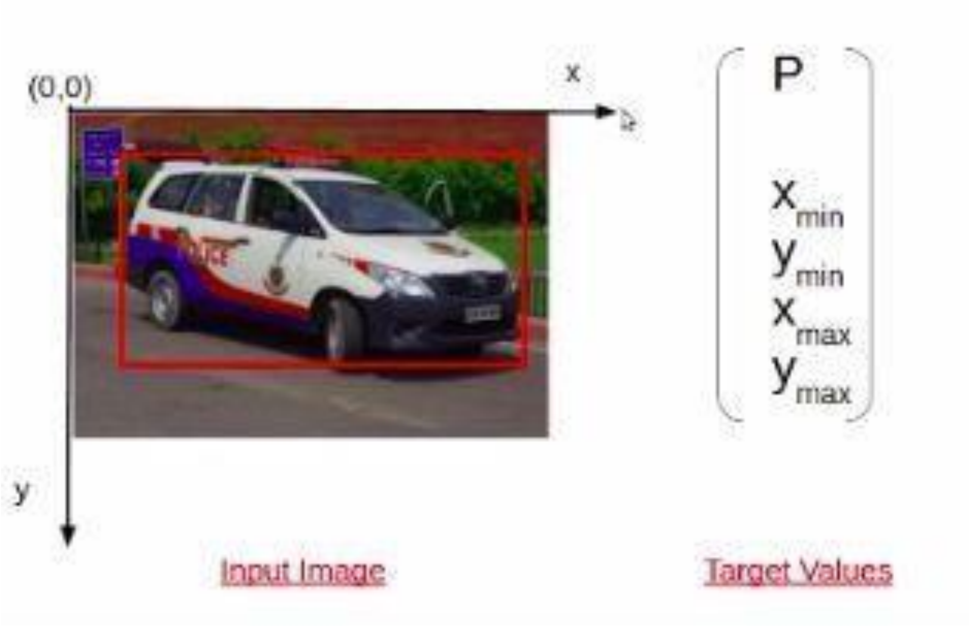


How Object Detection Works?



- Three tasks for object detection problems:
- 1. To identify if there is an object present in the image,
- 2. where is this object located,
- 3. what is this object?

Training Data for Object Detection

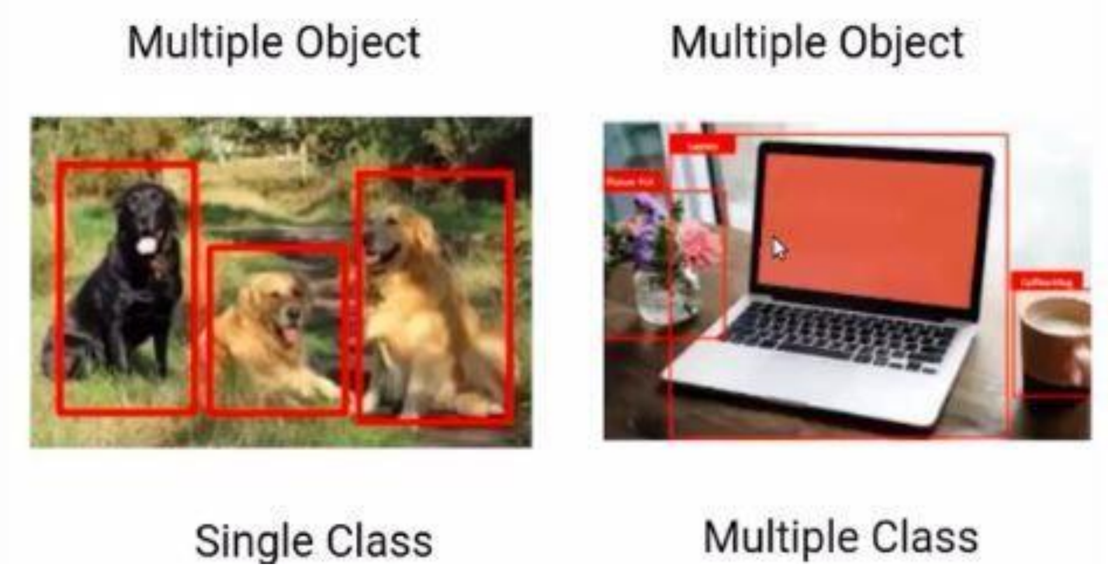


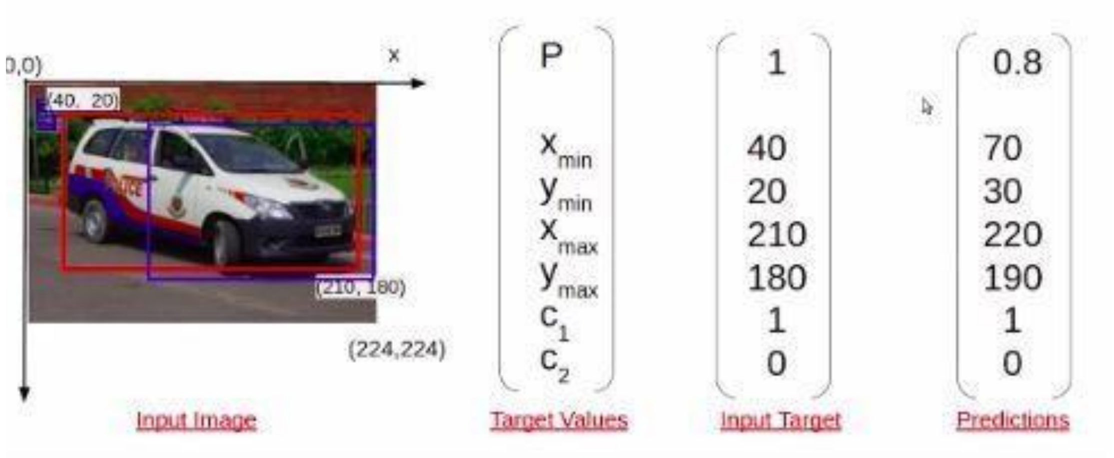
When you have images that have only one object. That is you can have 1000 images in the data set, and all of these images will have only one object. And if all these objects belong to a single class, that is all the objects are cars, then this will be an image localization problem. That is you already know what class these objects belong to, you only have to locate where these objects are present in the image.

Training Data for Object Detection

- Target variable has five values the value p denotes the probability of an object being in the above image whereas the four values Xmin, Ymin, Xmax, and Ymax denote the coordinates of the bounding box.
- If an object is not present then p will be zero and when there is an object present in the image p will be one.

Training Data for Object Detection





if you have two classes which are an emergency vehicle and a non-emergency vehicle, you'll have two additional values c1 and c2 denoting which class does the object present in the above image belong.

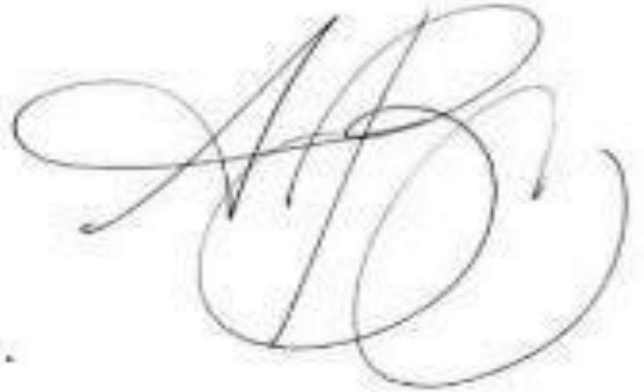
So if we consider this example, we have the probability of an object present in the image as one. We have the given Xmin, Ymin, Xmax, and Ymax as the coordinates of the bounding box. And then we have c1 is equal to 1 since this is an emergency vehicle and c2 would be 0 because of a non-emergency vehicle.

Pattern Recognition

- Pattern Recognition is the method of identifying and distinguishing the patterns, from the images that are fed as input, and the output is obtained in the form of patterns.
- The practice of distinguishing the patterns using Artificial Intelligence and Machine Learning tools with algorithms.
- Applications such as facial expression recognition, speech recognition, MDR, medical image recognition, etc., are a part of PR systems.

Pattern Recognition

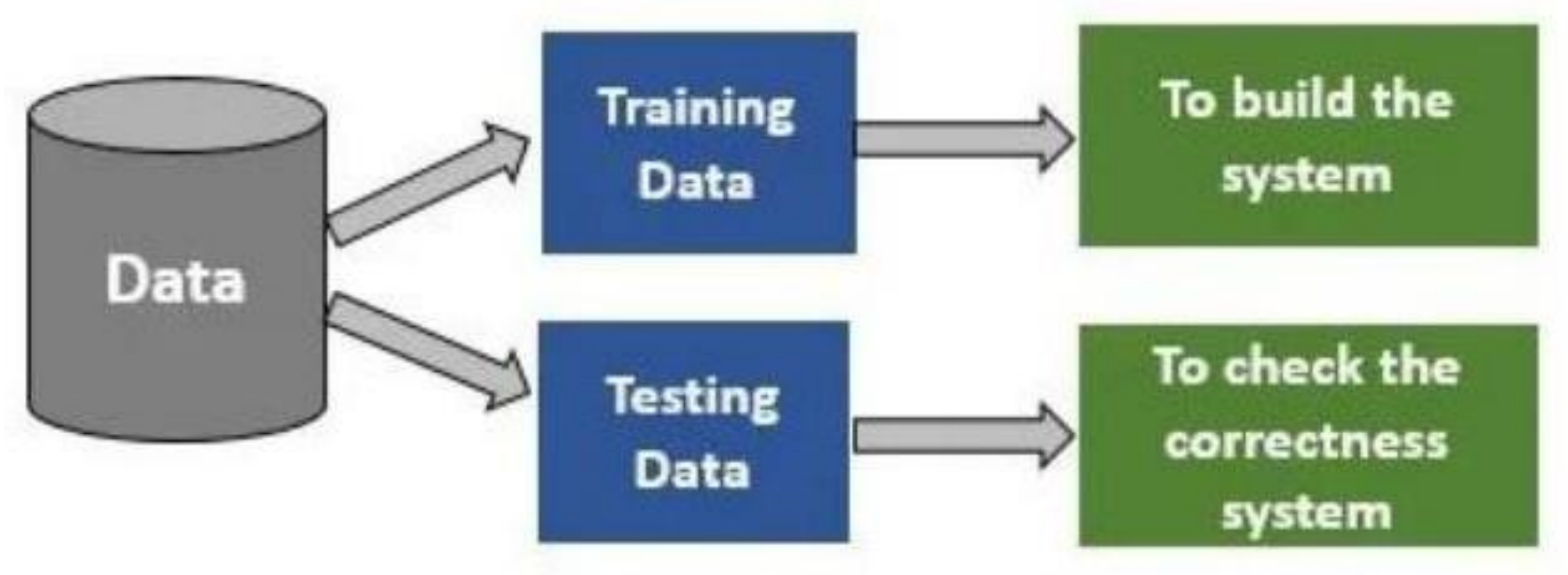
- A pattern is a set of objects or phenomena or concepts where the elements of the set are similar to one another in certain ways or aspects.



- A pattern is an entity , that could be given a name .

  Example : Fingerprint Image, handwritten word , human face , speech signal , DNA sequence etc.

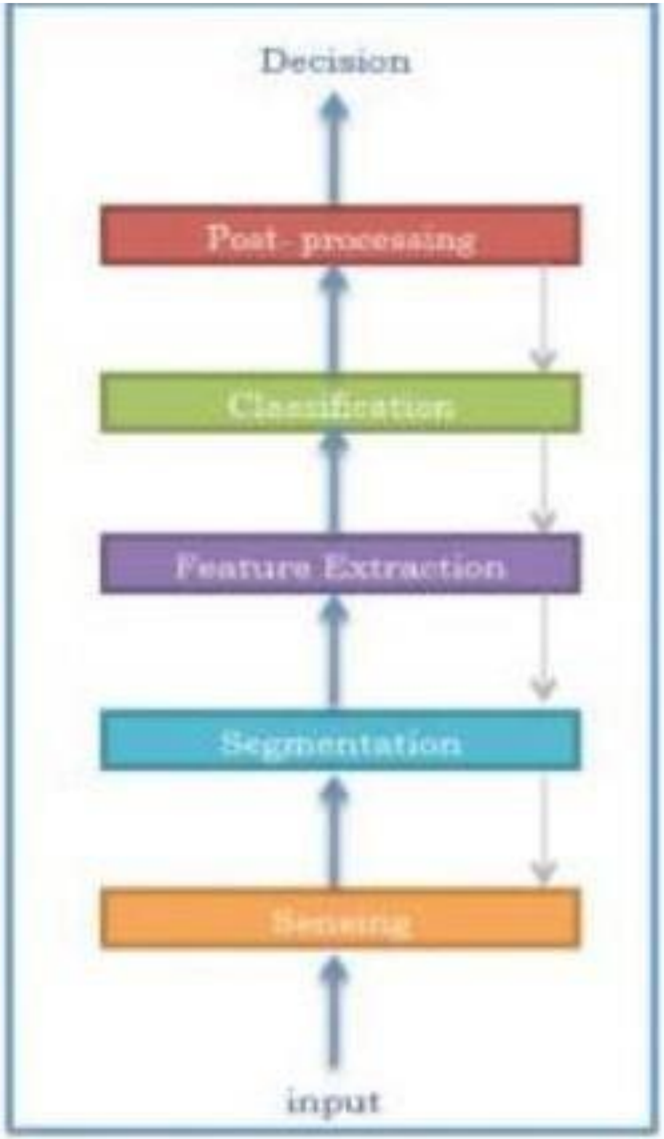
How Pattern Recognition Works?



- When we refer to the classification technique, **we assign appropriate class labels to any pattern that we want to recognize.**
- we have one face and its features include eyes, facial structure, nose, ears, etc.
- Thus the feature vector includes the above-mentioned features.
- The features, i.e. eyes, nose, structure, ears, etc. are taken together in aligned sequence as [eyes, nose, structure, ears]
- The feature vector used here is the series of certain features that are epitomized in the form of a k-dimensional column vector.

How Pattern Recognition Works?

- ***Data acquisition and sensing:***
  - Measurements of physical variables.
  - Important issues: bandwidth, resolution , etc.
- ***Pre-processing:***
  - Removal of noise in data.
  - Isolation of patterns of interest from the background.
- ***Feature extraction:***
  - Finding a new representation in terms of features.
- ***Classification***
  - Using features and learned models to assign a pattern to a category.
- ***Post-processing***
  - Evaluation of confidence in decisions.



How Pattern Recognition Works?

- **1. Data acquisition and sensing:**
  - Measurements of physical variables (capturing image, recording speech or video etc.)
  - Important factors (parameters) in this step that affects the patterns due to data acquisition: bandwidth, resolution, sensitivity, distortion, SNR, latency, etc.
  - Data collected is divided to 3 sections (train, test, validation). The training data is that used for learning models. From time to time you need to make a test to adjust the system parameters (thresholds) or rules. In such case, validation data is used. When the system performs quite well (without over fitting), test section is used for final evaluation.

How Pattern Recognition Works?

- **2. Pre-processing:**
  - an essential task that precedes the tasks of image representation and recognition. Its importance is derived from the fact that the discrimination power is directly proportional to the digital image quality, in the sense that, the higher the image quality the less confusions we have and thus more powerful classification we can make.

  - Some common operations performed as preprocessing are: *binarization*, the task of converting gray-scale image into a binary black-white image; *noise removal*, the extraction of the foreground textual matter by removing textured background, salt and pepper noise or interfering strokes; *image enhancement and restoration*, the task of converting the image to be more suitable than the original image for a specific application ; *morphological image processing*, the task of extracting image components that are useful in the representation and description of region shape.

How Pattern Recognition Works?

- **3. Feature extraction:**
- Finding a new representation in terms of features

- **4. Model learning and estimation:**
  - Learning a mapping between features and pattern groups and categories.

- **5. Classification:**
  - Using features and learned models to assign a pattern to a class.

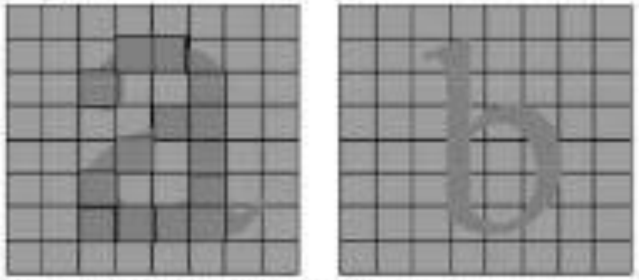
- **6. Post-processing (extra optional stage to improve performance):**
  - Evaluation of confidence in decisions.
  - Exploitation of context to improve performance.
  - Combination of multiple classifiers decisions.

Pattern Recognition Model

- ***Statistical model:*** Pattern recognition systems are based on statistics and probabilities.
- ***Syntactic model:*** Structural models for pattern recognition and are based on the relation between features. Here the patterns are represented by structures .
- ***Template matching model:*** In this model, a template or a prototype of the pattern to be recognized is available.
- ***Neural network model:*** An artificial neural network (ANN) is a self-adaptive trainable process that is able to learn and resolve complex problems based on available knowledge.

Pattern Recognition Model

- ***1. Template Matching,*** i.e. using a template (as a mask) to compute certain features. For example:



- using a grid to compute the ratio of foreground black pixels to the back ground white pixels, or the count of black pixels per column, or the number of connected pixels. The template can be circular mask with sectors and the foreground black pixels area could be the feature, and so on.
  - Template Matching is very sensitive to pattern/sample size, rotation and translation variations as well as to noise. I.e., two typical patterns may have different feature vectors if one of them is bigger in size, rotated, or translated a little. Thus, it has limited application.

Advantage of Pattern Recognition

- Pattern recognition cracks the problems in classification.
- There are various problems in day to day life that are handled by the intelligent PR systems such as Facial expression recognition systems.
- Visually impaired people are also benefitting the PR systems in many domains.
- The speech recognition systems are doing wonders and helping in research fields.
- The object detection is a miraculous achievement of PR systems that is helpful in many industries such as aviation, health, etc.

Disadvantage of Pattern Recognition

- The process is quite complex and lengthy, which consumes time.
- The dataset needs to be large for accuracy.
- The logic is not certain of object recognition.

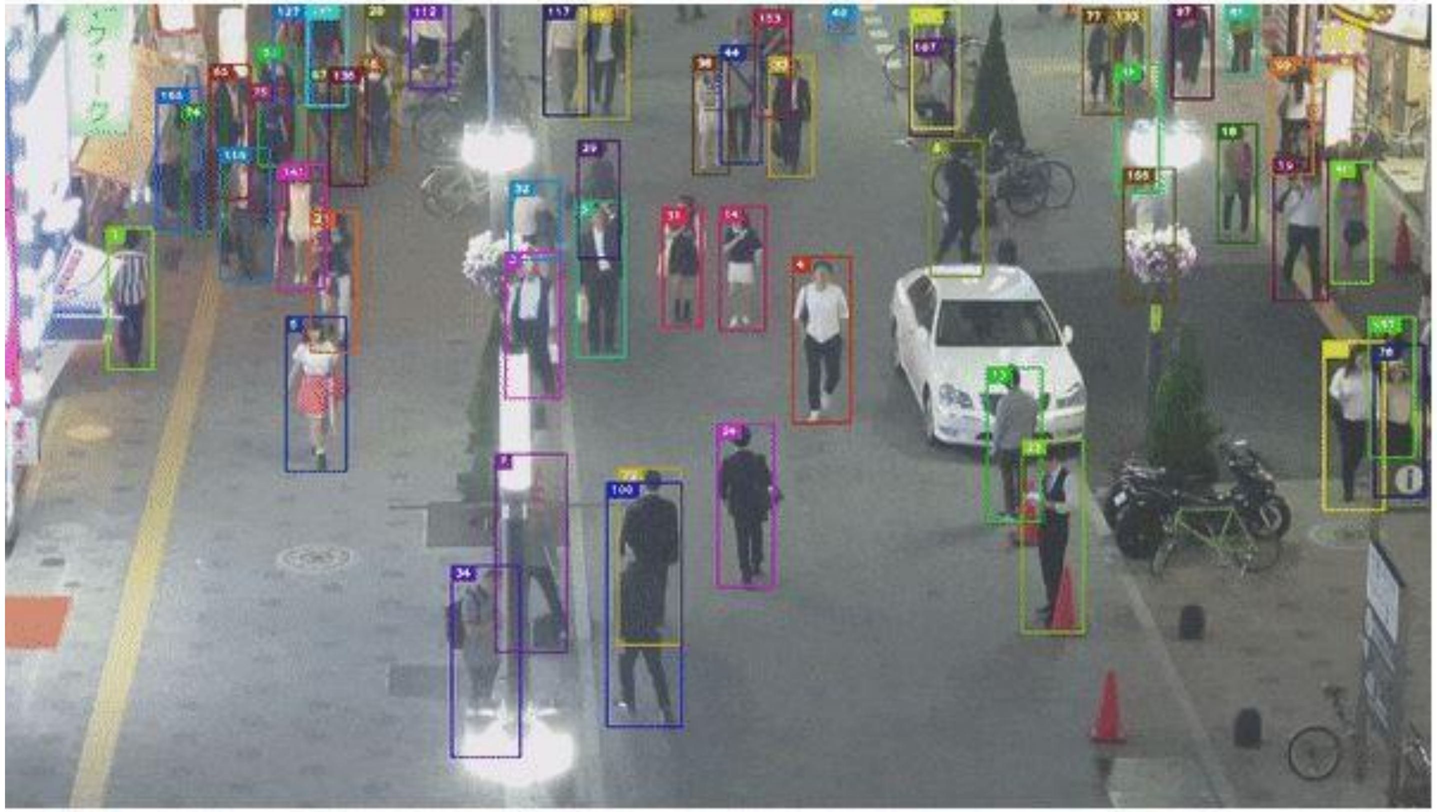
Application of Pattern Recognition

- Facial expression recognition
- Iris recognition
- Handwriting recognition
- Speaker recognition
- Object Tracking
- Fingerprint identification
- Computer vision
- Seismic analysis
- Radar signal classification/analysis

Object Tracking

- Object tracking is a field within computer vision that involves tracking objects as they move across several video frames.
- Object tracking has many practical applications including surveillance, medical imaging, traffic flow analysis, self-driving cars, people counting and audience flow analysis, and human-computer interaction.
- Object tracking starts with object detection —identifying objects in an image and assigning them bounding boxes.
- The object tracking algorithm assigns an ID to each object identified in the image, and in subsequent frames tries to carry across this ID and identify the new position of the same object.

Object Tracking



Types of Object Tracking

- **Offline object tracking**—object tracking on a recorded video where all the frames, including future activity, are known in advance.

- **Online object tracking**—object tracking done on a live video stream, for example, a surveillance camera. This is more challenging because the algorithm must work fast, and it is not possible to take future frames and combine them into the analysis.

People Counting and Tracking

- The objective is to build a system that has the following features.
- Read the frames from the Video.
- Draw a desired reference line on the input frame.
- Detect the people using the object detection model.
- Mark the centroid on the detected person.
- Track the movement of that marked centroid.
- Calculate the direction of centroid movement (whether it is moving upwards or downwards).
- Count the number of people coming in or going out of a reference line.
- Based on the counting, increment the up or down counter.

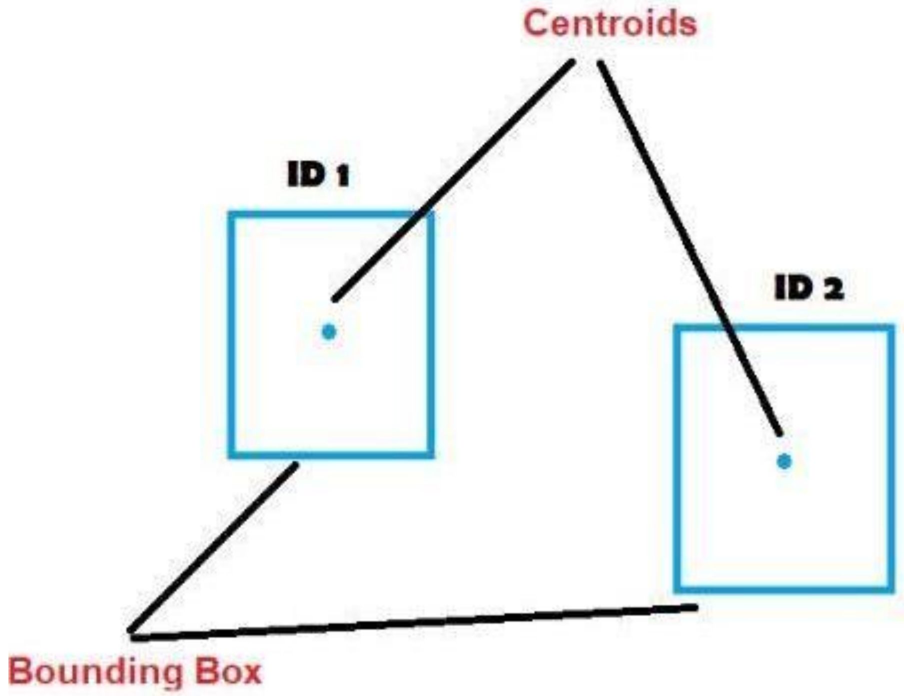
People Counting and Tracking

- Object tracking techniques use methods like **deep sort, centroid tracker, csrt, kcf,** and camshift which track the detected object by comparing the similarity of detected objects with each other in each processed frame.
- If the object has the same similarity metric throughout the frame then it will track the same object throughout the sequence of frames and retain the same object ID for that object.
- This constant object ID for a particular object makes it easier for us to do the counting operations.

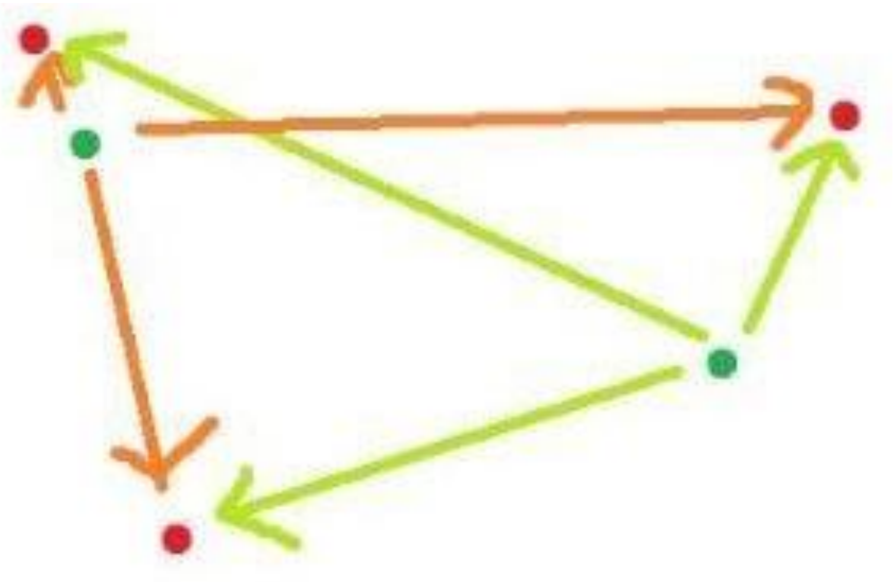
Centroid Tracker

- The centroid tracker has the following steps:

- Accepts the bouding box coordinates and computes the centroid.
- The algorithm accepts the bounding box coordinates that are xmin, ymin, xmax, and ymax and the gives (x_center, y_center) coordinates for each of the detected objects in each frame.
- The Centroid is calculated is given below:
- X_cen = ((xmin + xmax) // 2
- Y_cen = ((ymin + ymax) // 2)

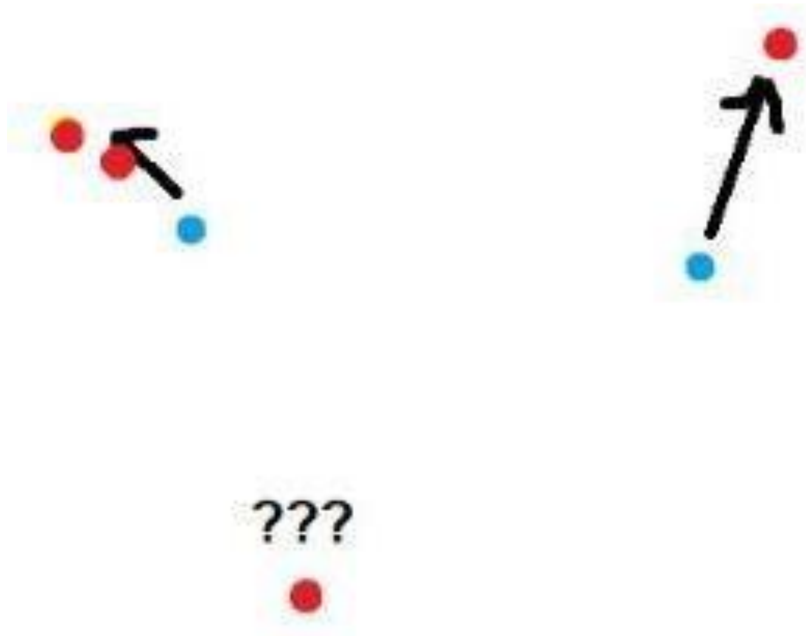


Centroid Tracker



Then, it calculates the euclidean distance between the new detected bounding box and the existing object (mostly in consecutive video frames).

Centroid Tracker



Update the centroid for the existing object. After calculating the euclidian distance between the detected bounding box and the existing bounding box, it will update the position of the centroid in the frame, there by tracking the object.

Centroid Tracker

Registering new objects. When a new object enters or the same object is been detected, the centroid tracker will register the new object with a unique ID, so that it becomes helpful for different applications.

De-registering the previous objects. Once the object is not in the frame, the algorithm will de-register the object ID, stating that the object is not available or left the frame.

Face Detection

- Face detection -- also called facial detection -- is an artificial intelligence (AI) based computer technology used to find and identify human faces in digital images.
- Face detection technology can be applied to various fields -- including security, biometrics, law enforcement, entertainment and personal safety -- to provide surveillance and tracking of people in real time.

How Face Detection Works?

- Face detection applications use algorithms and ML to find human faces within larger images, which often incorporate other non-face objects such as landscapes, buildings and other human body parts like feet or hands.
- Face detection algorithms typically start by searching for human eyes -
  - one of the easiest features to detect.
- The algorithm might then attempt to detect eyebrows, the mouth, nose, nostrils and the iris.
- Once the algorithm concludes that it has found a facial region, it applies additional tests to confirm that it has, in fact, detected a face.

How Face Detection Works?

- To help ensure accuracy, the algorithms need to be trained on large data sets incorporating hundreds of thousands of positive and negative images.
- The training improves the algorithms' ability to determine whether there are faces in an image and where they are.

How Face Detection Works?

- The methods used in face detection can **be knowledge-based, feature-based, template matching or appearance-based.** Each has advantages and disadvantages:

1. **Knowledge-based, or rule-based methods,** describe a face based on rules. The challenge of this approach is the **difficulty of coming up with well-defined rules**.
2. **Feature-based methods** -- which use features such as a person's eyes or nose to detect a face -- can be negatively affected by noise and light.
3. **Template-matching methods** are based on comparing images with standard face patterns or features that have been stored previously and correlating the two to detect a face. Unfortunately these methods do not address variations in pose, scale and shape.
4. **Appearance-based methods** employ statistical analysis and machine learning to find the relevant characteristics of face images. This method, also used in feature extraction for face recognition.

How Face Detection Works?

- Some of the more specific techniques used in face detection include:
- Removing the background. For example, if an image has a plain, mono-color background or a pre-defined, static background, then removing the background can help reveal the face boundaries.
- In color images, sometimes skin color can be used to find faces; however, this may not work with all complexions.
- Using motion to find faces is another option. In real-time video, a face is almost always moving, so users of this method must calculate the moving area. One drawback of this method is the risk of confusion with other objects moving in the background.

Advantage Face Detection

- Improved security. Face detection improves surveillance efforts and helps track down criminals and terrorists. Personal security is also enhanced since there is nothing for hackers to steal or change, such as passwords.
- Easy to integrate. Face detection and facial recognition technology is easy to integrate, and most solutions are compatible with the majority of security software.
- Automated identification. In the past, identification was manually performed by a person; this was inefficient and frequently inaccurate. Face detection allows the identification process to be automated, thus saving time and increasing accuracy.

Face Detection Challenges

- Massive data storage burden. The ML technology used in face detection requires powerful data storage that may not be available to all users.
- Detection is vulnerable. While face detection provides more accurate results than manual identification processes, it can also be more easily thrown off by changes in appearance or camera angles.