COMPUTER
VISION &
INTELLIGENCE
CLUB

OBJECT DETECTION WORKSHOP

FROM HOG TO YOLO

FROM EARLY CONVENTIONAL MODELS TO STATE-OF-THE-ART, LEARN EVERYTHING FROM SCRATCH RIGHT HERE

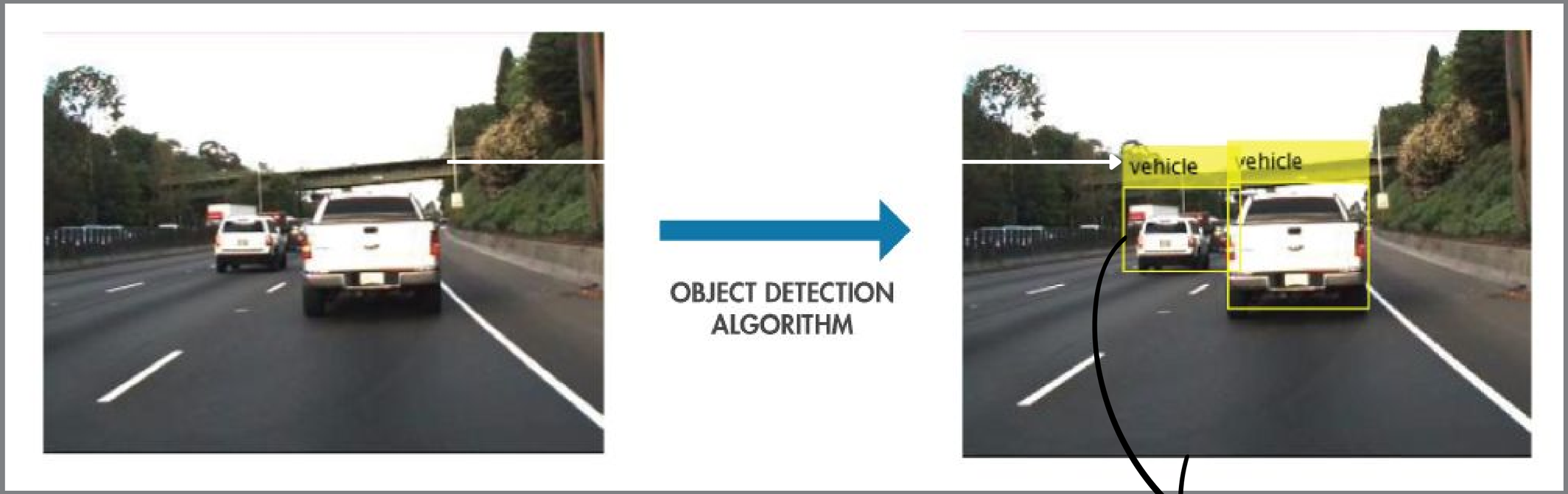
17TH-19TH OCT 2022
CRC 204

Welcome!

Session 1
Introduction to Histogram of Oriented Gradients and Support vector machines

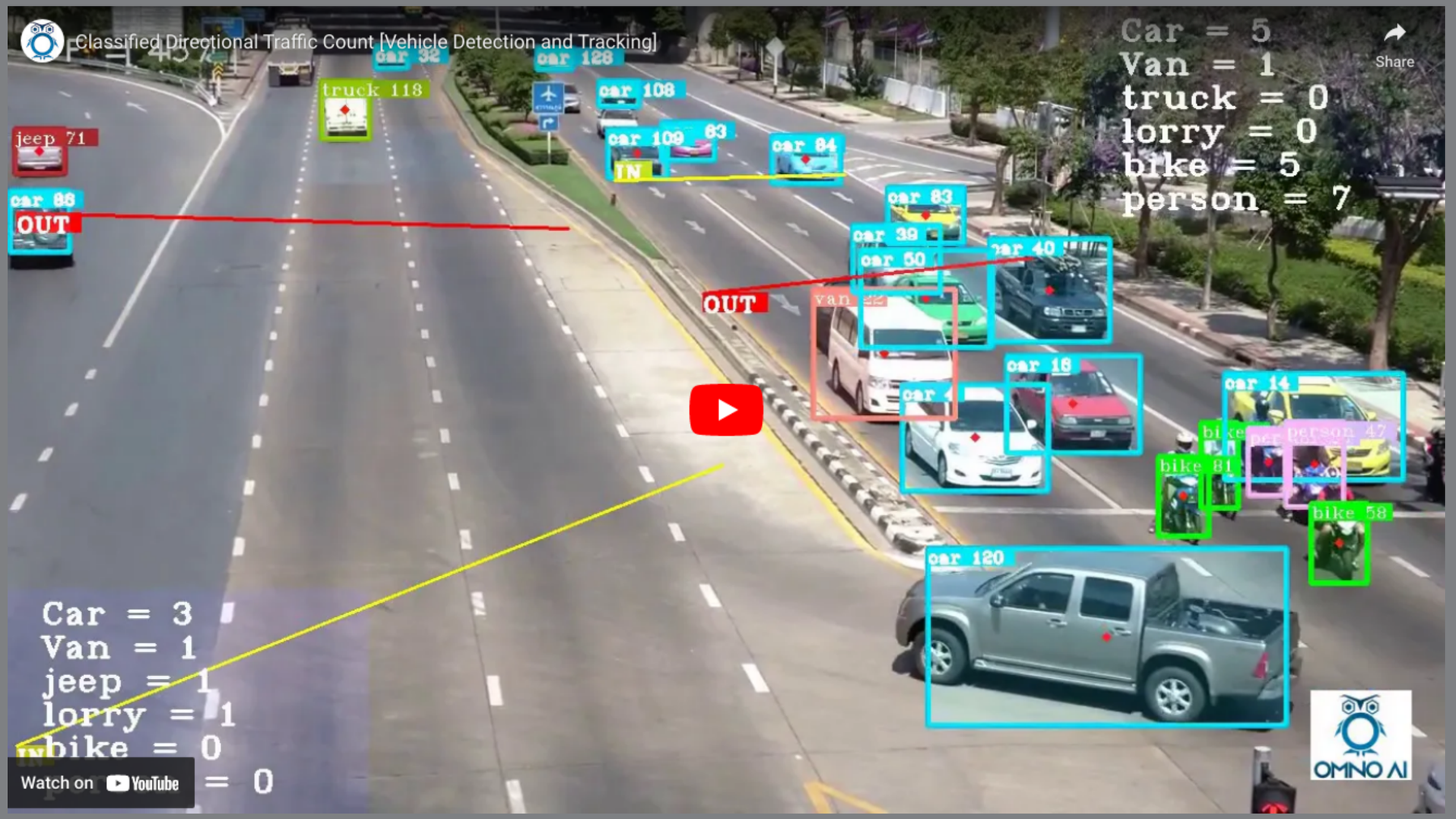
Object Detection

Object Detection is a computer Vision Technique to locate instances of an object in image or videos
It involves checking if an object is present, locating and identifying it.



Object Detection Model applied on a video

- **Object Detection model applied on traffic data for counting vehicle flow bidirectonally in junctions.**

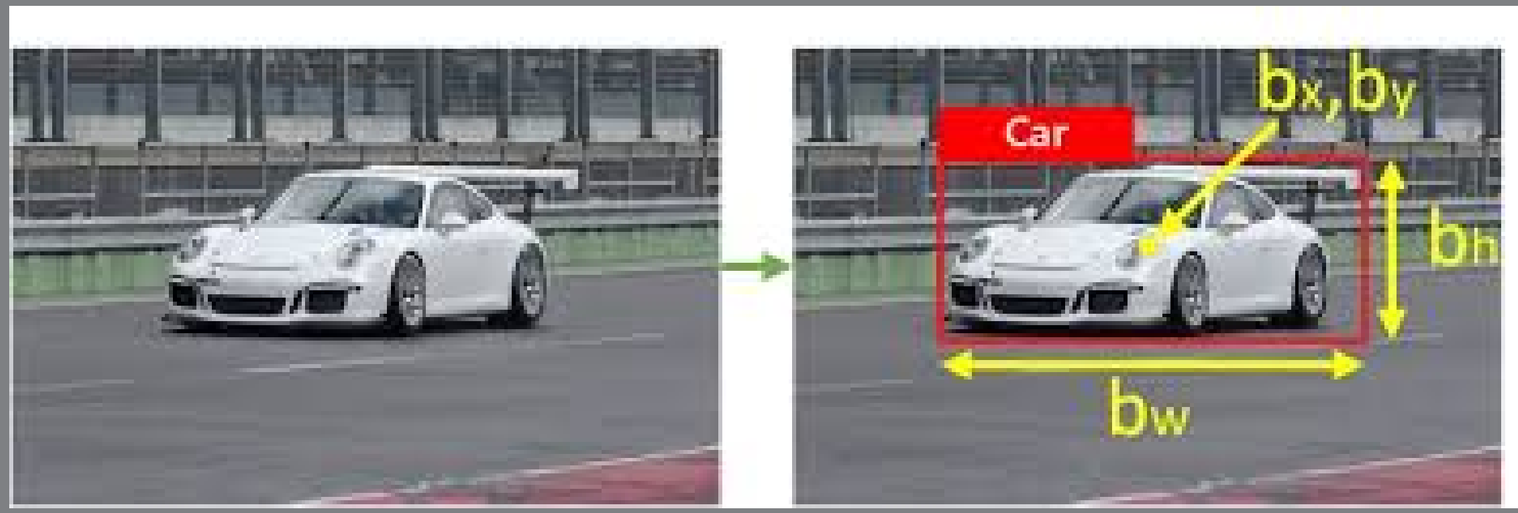


Task Definition

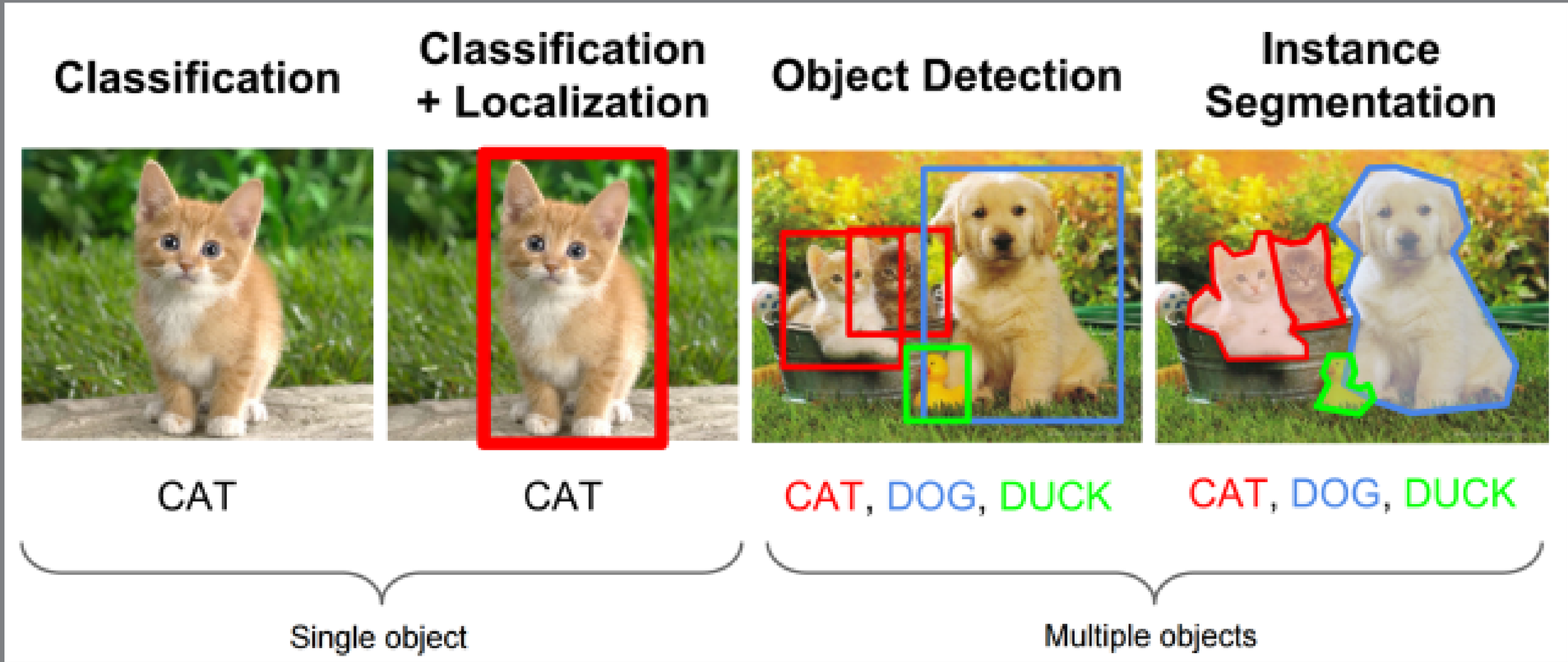
Input: A single RGB Image

Output:
A set of detected object,
For each pair of object predict:

- Category Label(from known, fixed set of categories)
- Bounding box: 4 numbers(x, y, width, height)

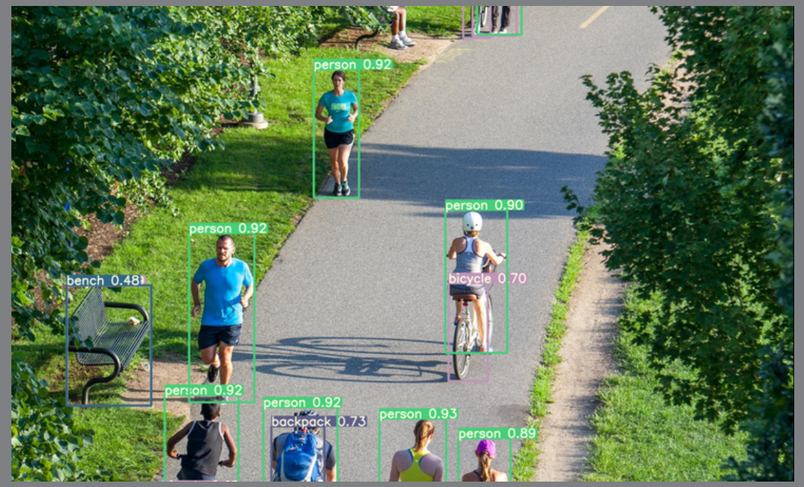


Different Computer Vision Tasks



Applications in Real World

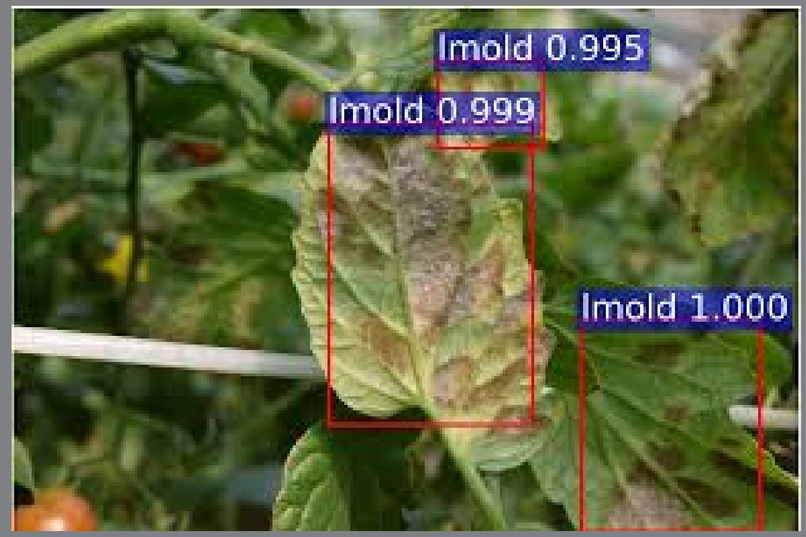
Object Tracking



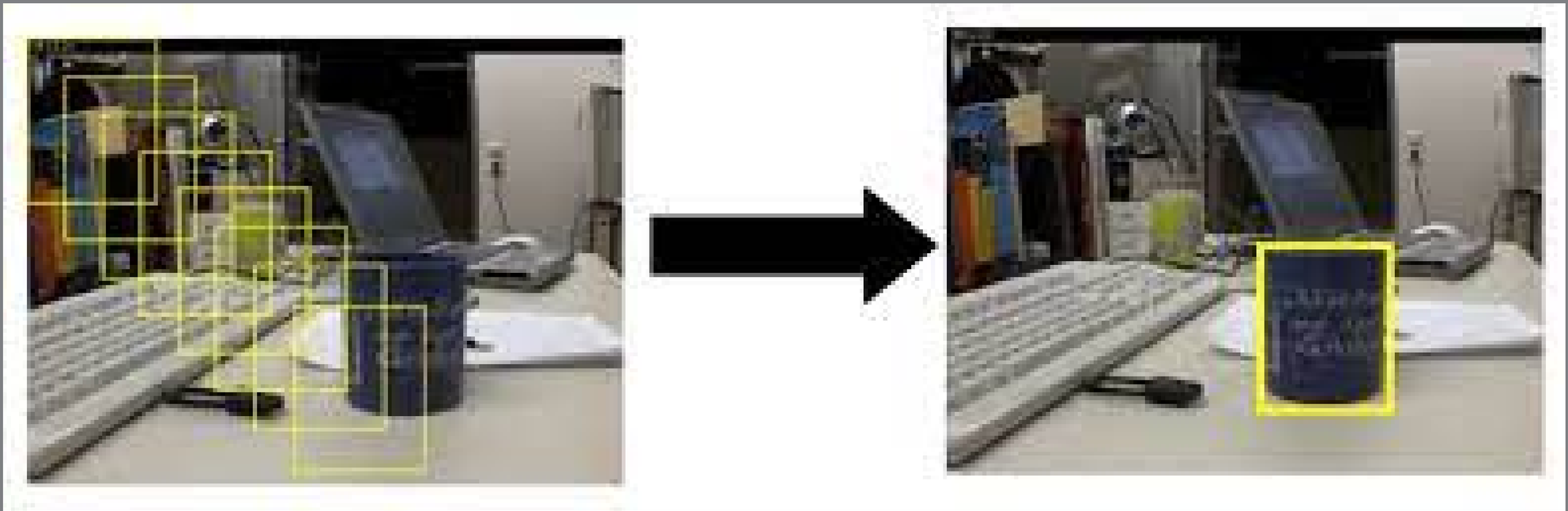
Crowd counting

Security systems

Crop Anomaly Detection



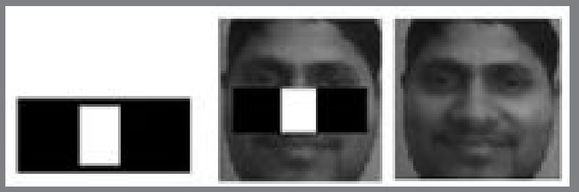
Sliding Window Intuition for Object Detection

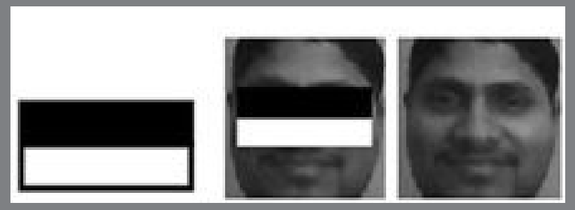


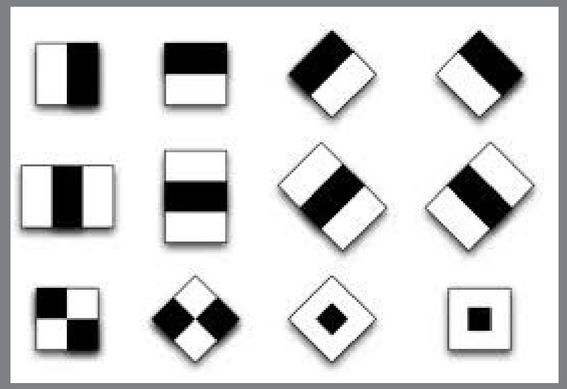
Evolution of Object Detection algorithms

Viola Jones Face Detector(2001)

- Detects Whether an image contains a face in it
- Slides a window covering all locations and regions in an image
- The sliding window searches for Haar like features



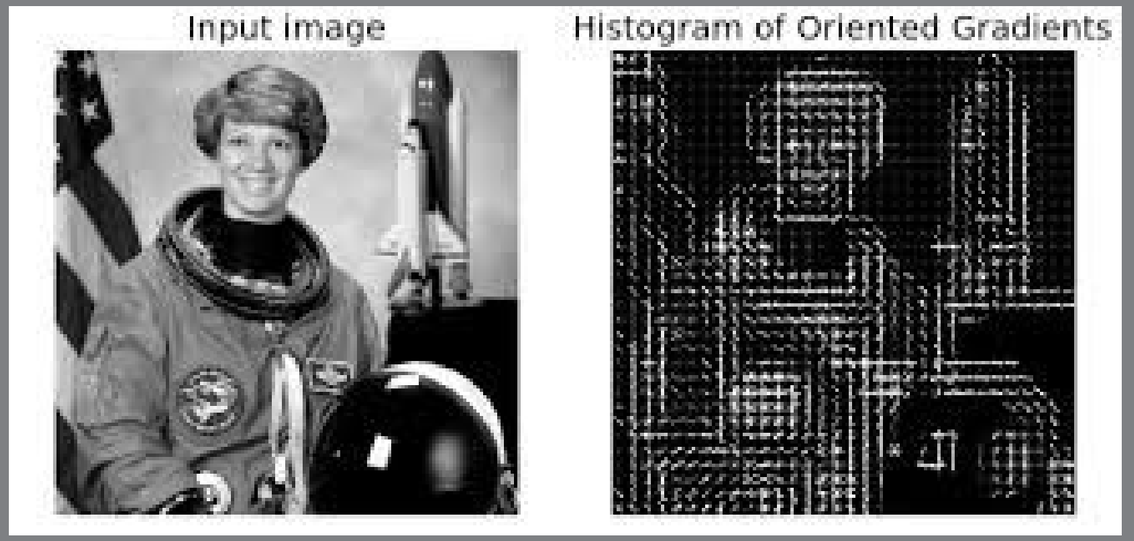




Haar features

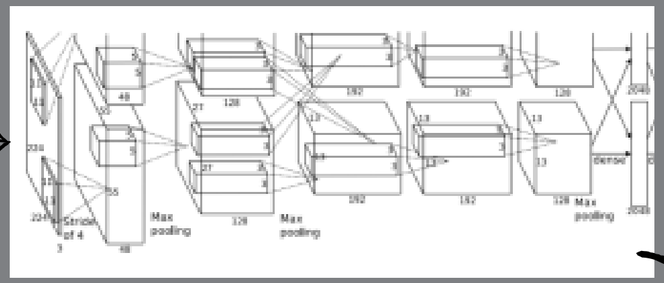
HOG Based Detector(2005)i

- Makes a histogram on the basis of gradients strengths and directions for each pixel.
- creates a vector of gradients averaged over adjacent cells for the image
- An svm or similar model can be used for determining whether an object is present or not



Application of CNNs in object detection

Input Image



CNN layers

Class scores
Cat:0.15
Dog 0.80
..

Softmax loss

+

Weighted Net Loss

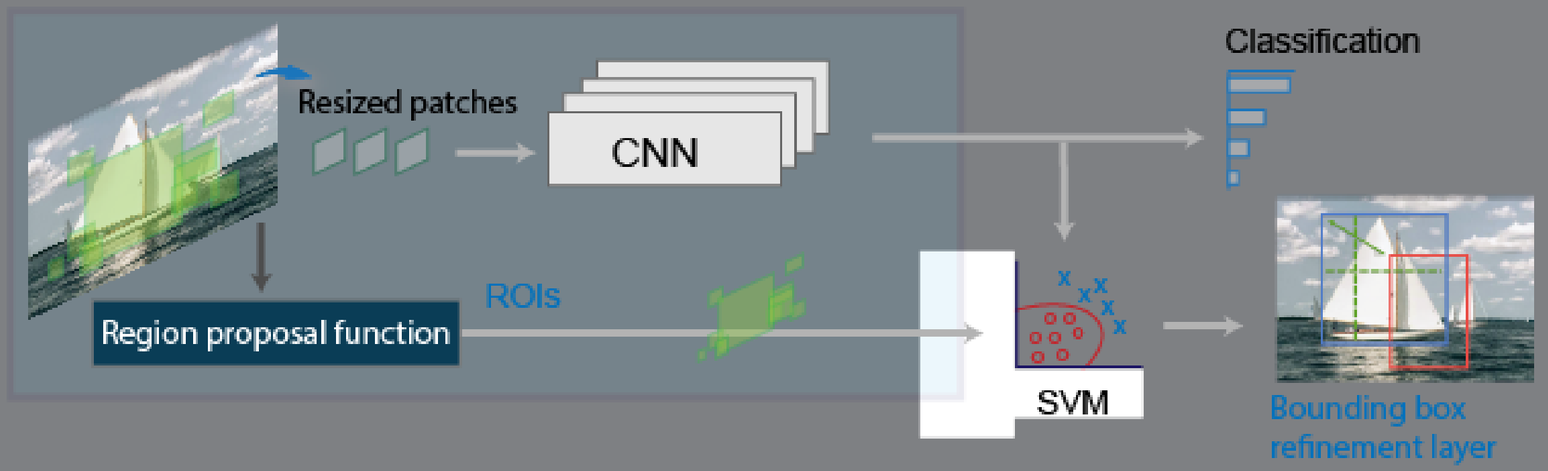
Bounding Box coordinates:
(x, y, w, h)
.

L2 Loss: mean squared error

Deep Learning based models

RCNNs (2014)

- Applying sliding window technique on a image and then feeding it to a CCN Model comes with high computational complexity
- An image of m*n dimensions has approximately M2*N2 bounding boxes which needs to be fed into CNN
- Here's where RCCN(Region based CNN) comes in, Selective Search algorithm of RCNN proposes regions which has maximum possibility of object presence in it and only these images are fed into CNNs
- It reduces the computational power required to identify objects and captures less noise objects



YOLO Detection Algorithm(2015)

- All of the previous object detection algorithms use regions to localize the object within the image.
- But YOLO algorithm looks at the image only and generates bounding boxes for all objects in it.
- It splits an object into multiple s∗s grids and finds a probability if an object is present in the particular grid
- Bounding boxes below a particular threshold are removed resulting in bounding boxes for only