Object Detection (Yolo Algorithm)

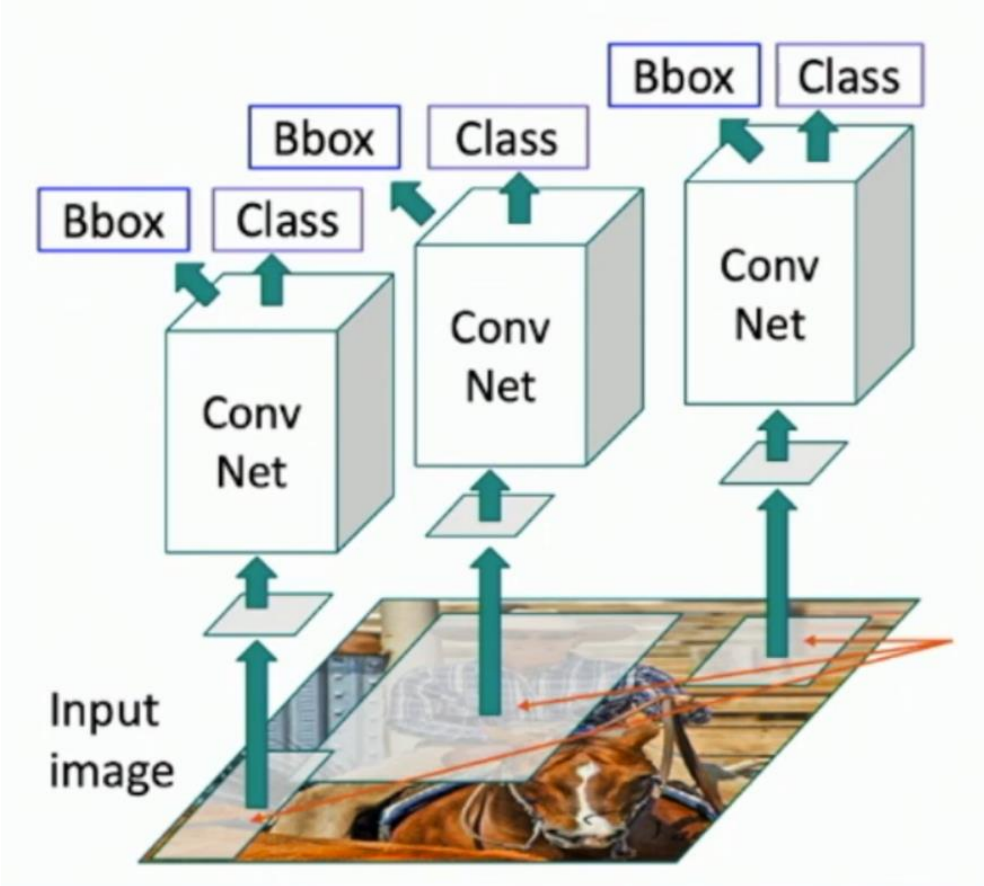
Presented by

Zyad Hassan Hosney 1806090

Dina Ibrahim Elkady 18q3800

Convolutional Neural Networks (CNN) Family

- (CNN) have high performance in the fields of object recognition and classification. The strength of CNNs comes from the fact that they are able to extract information from raw-pixel content and learn features automatically without explicit programming.



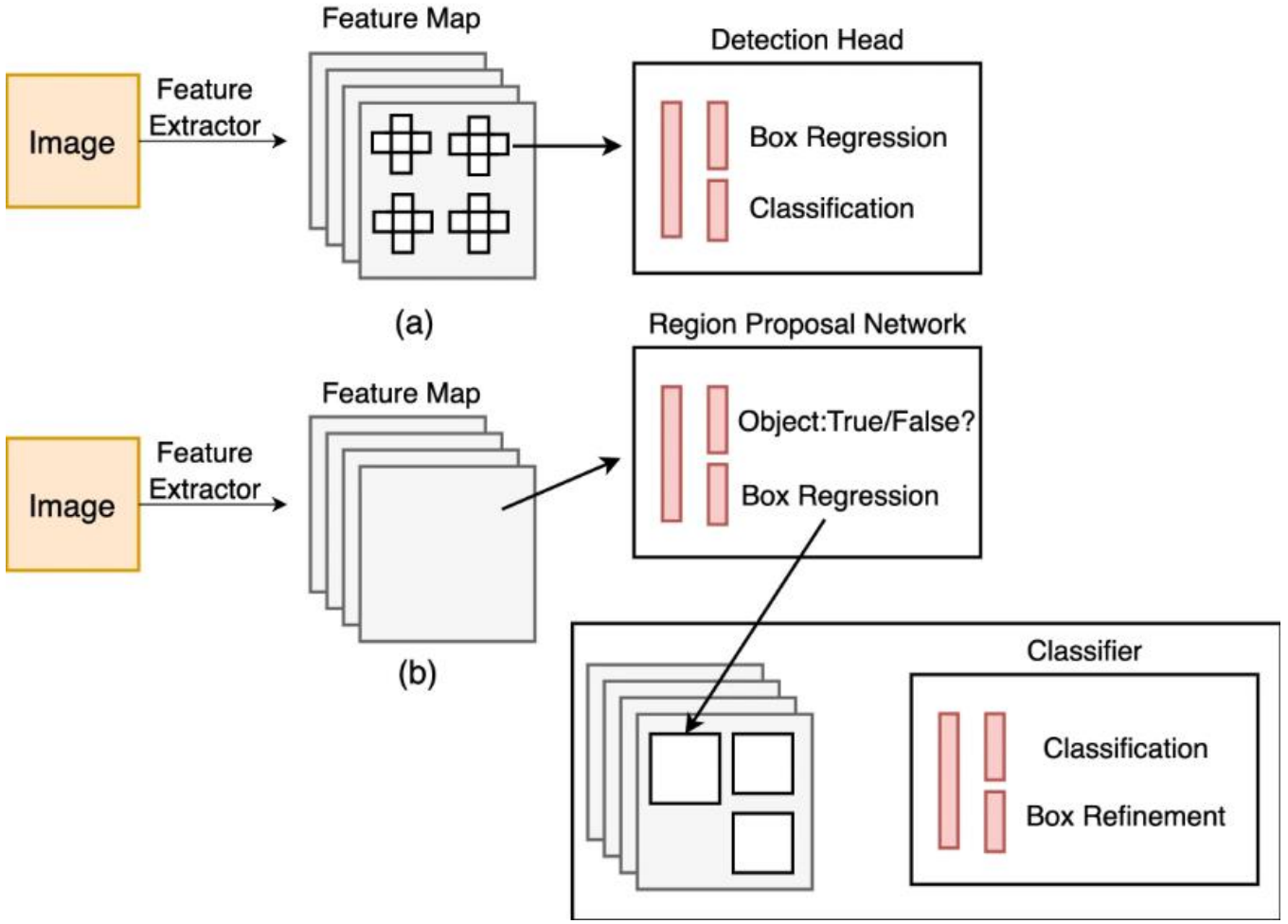
Two-stage object detectors VS. Single-stage object detectors

- A single-stage detector removes the ROI extraction process and directly classifies and regresses the candidate anchor boxes.

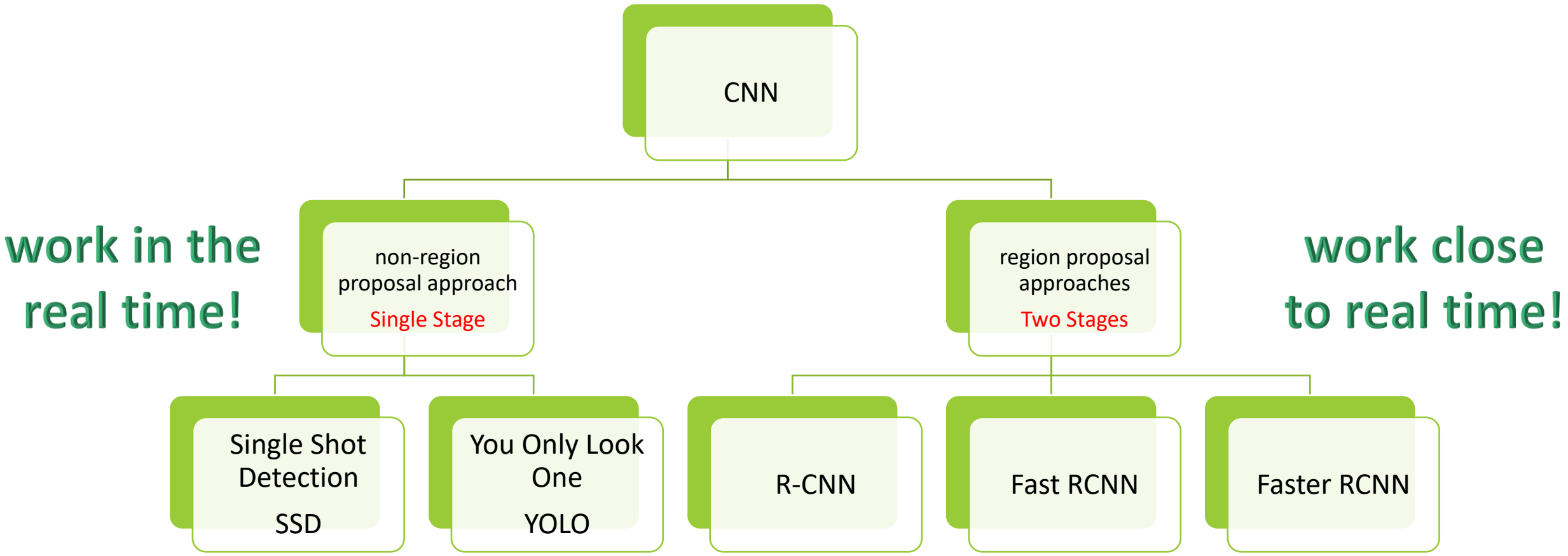
- Two-stage detectors divide the object detection task into two stages:

1- Extract ROIs (Region of interest),

2- Classify and regress the ROIs.



Convolutional Neural Networks (CNN) Family

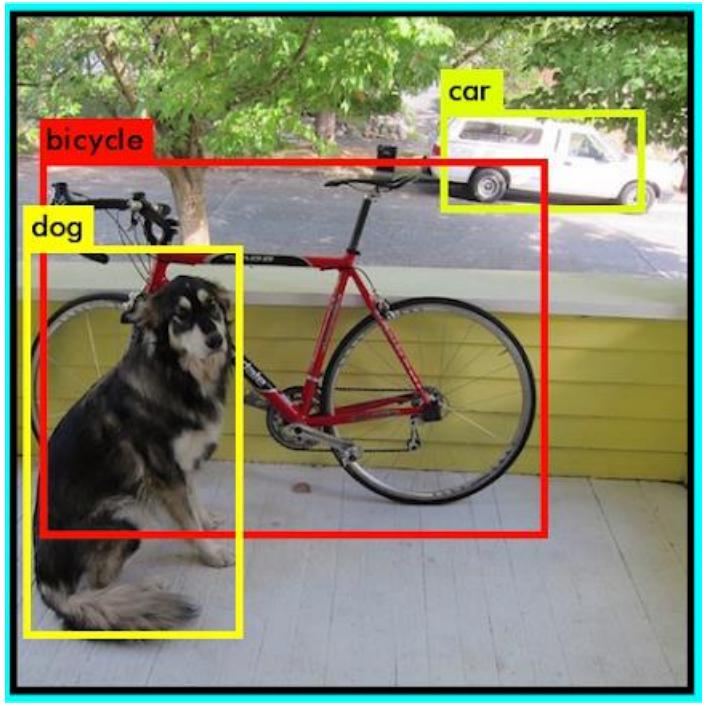


work in the real time!

work close to real time!

You Only Look Once

- It’s object detection algorithm.
- single convolutional network predicts the bounding boxes and the class probabilities for these boxes all at once .
- The used network reasons globally about all objects in the image. Its design enables end-to-end training, real time speed, and high average precision.
- running at as high as 45 FPS.



How it works ?

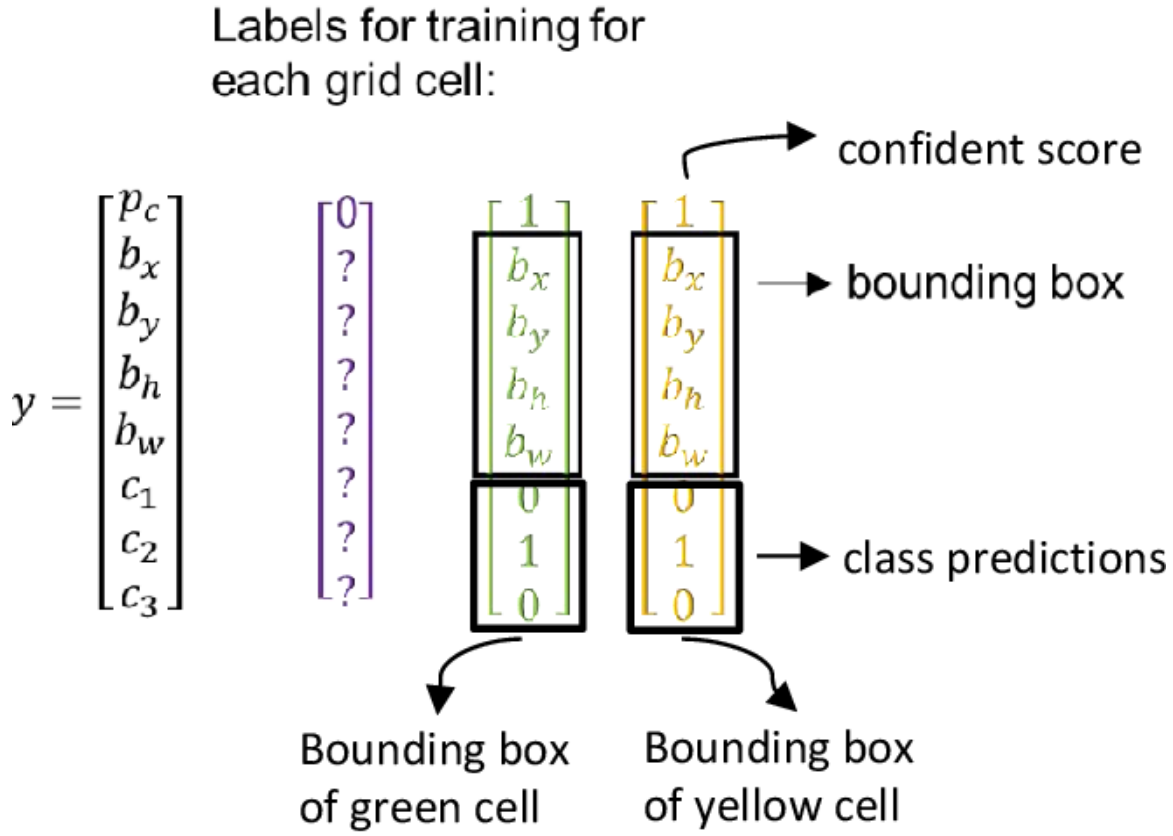
Step One :

Divide the image into a SxS grid.

Each cell will predict :

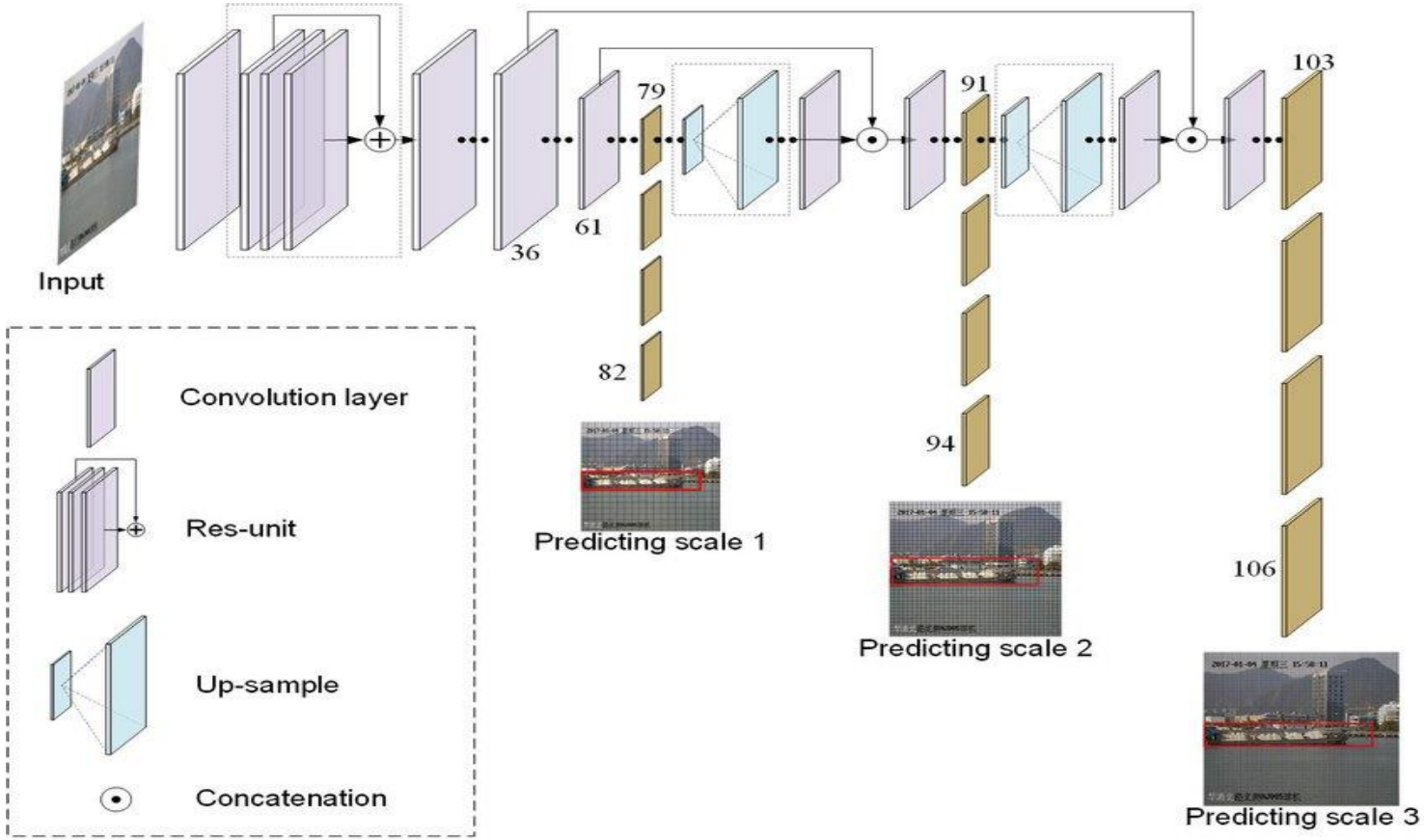
- Predict if an object exists
- Coordinates of the center of the object
- Size of bounding box
- The class of the object

- Each cell should predict the vector which have the class probability and the bonding box
- YOLO needs to be trained with labeled dataset first to can predict



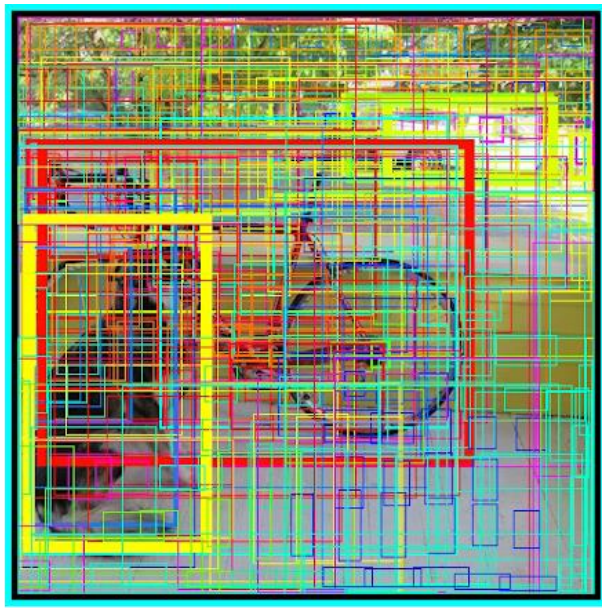
Architecture

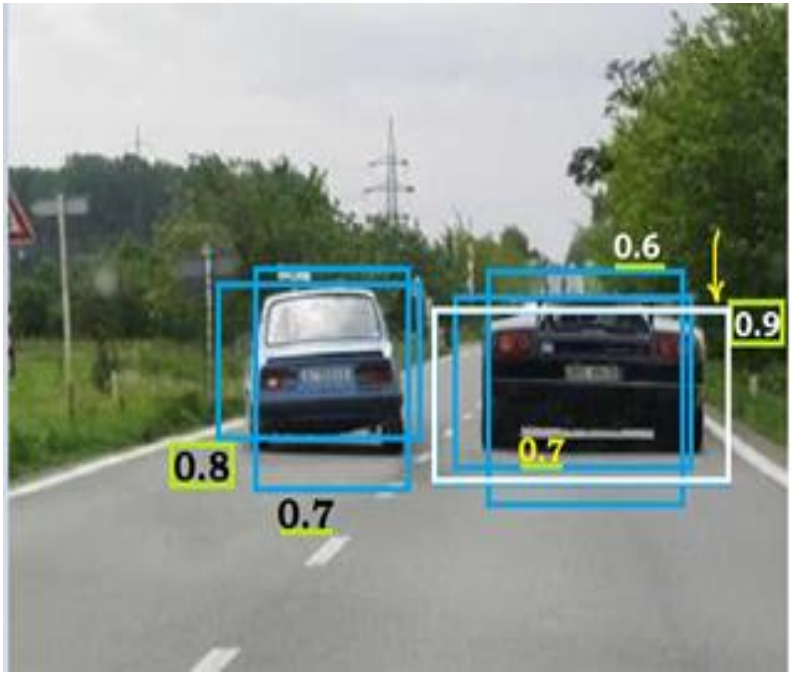
- In the Architecture of YOLO v3 it takes 3 different scales of cell to improve the accuracy of detecting the object



Step two:

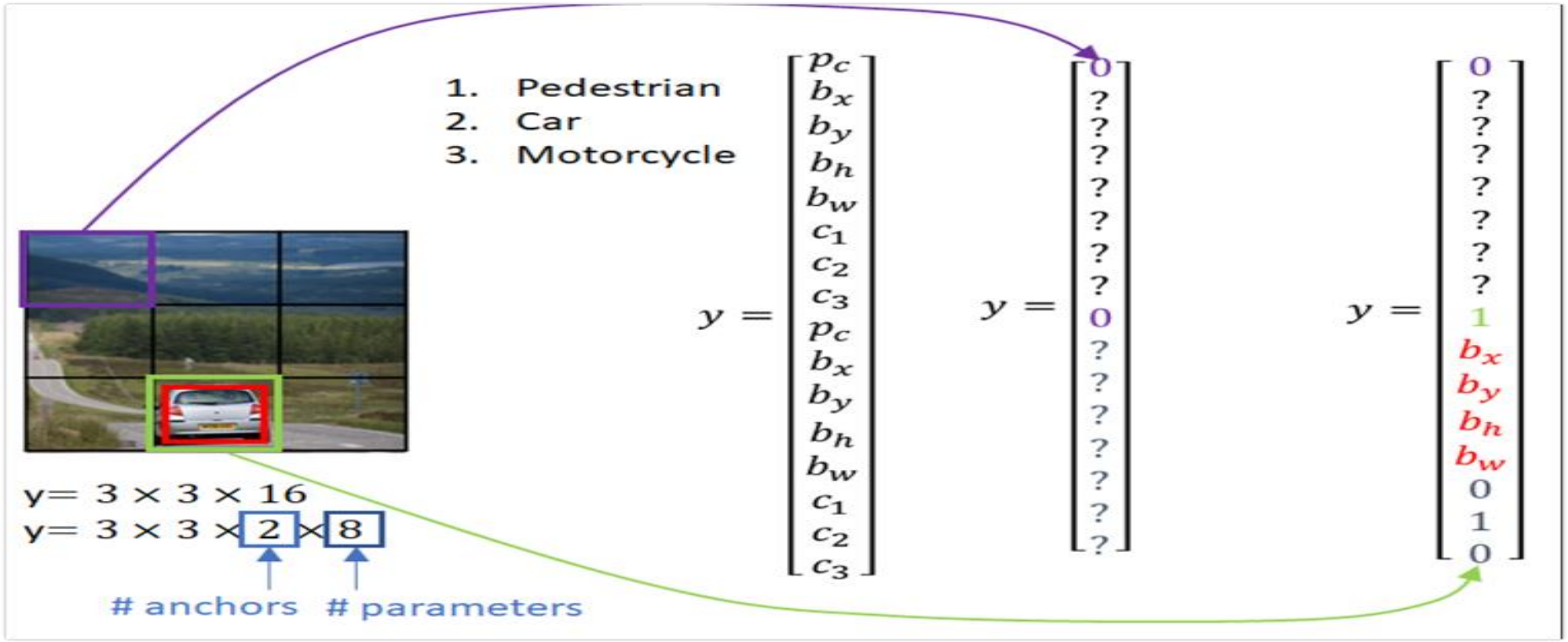
- every cell predict its vector (bounding box)
- Every box has probability of the object in it and the boxes intersects
- We can’t take the maximum to avid the case that existing two objects close to each other and consider it as one object
- Applying IOU (intersection over union) : compute the intersection area of the boxes and apply thresholding if its higher apply non maximum suppression and take the higher probability



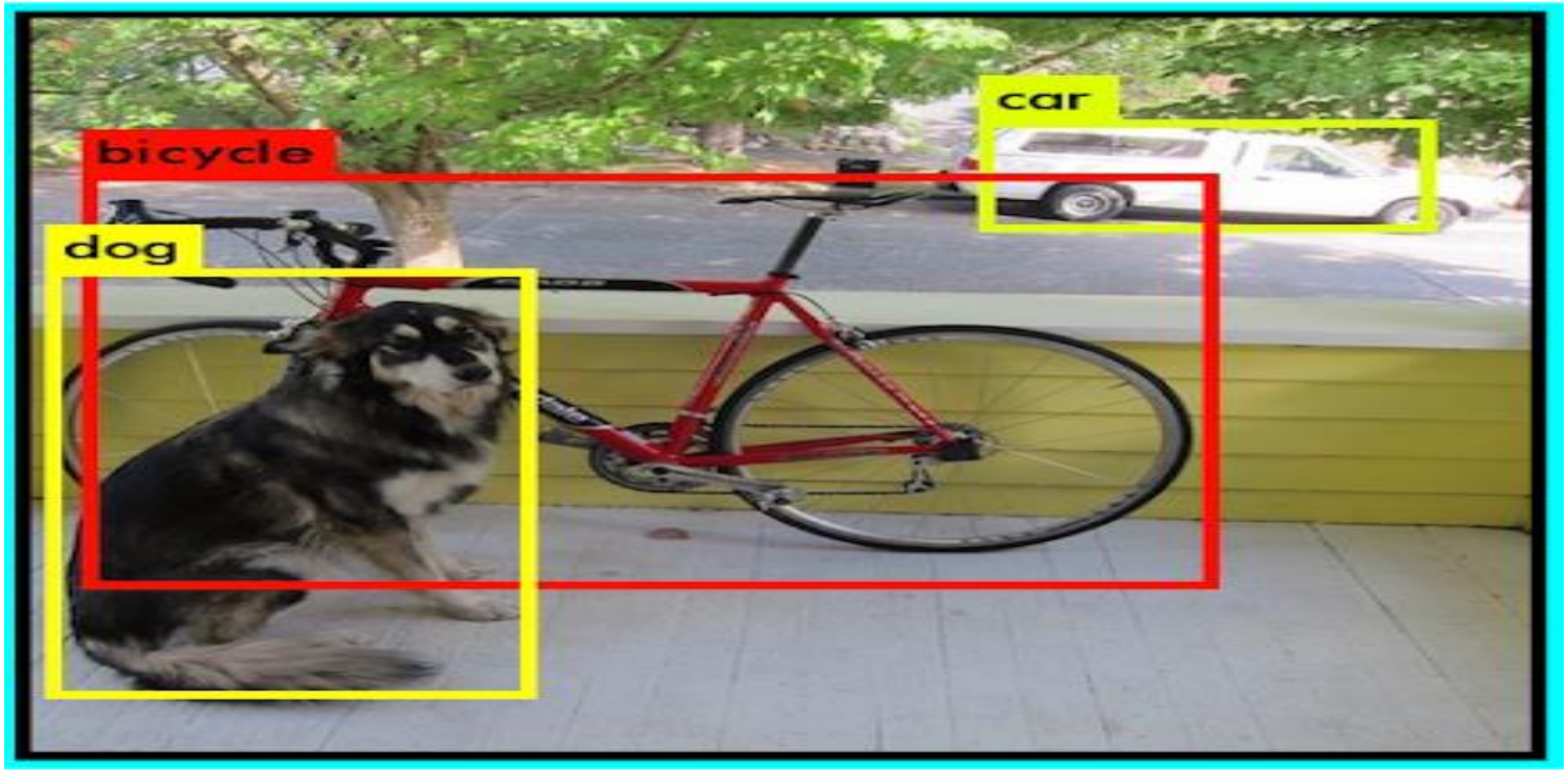


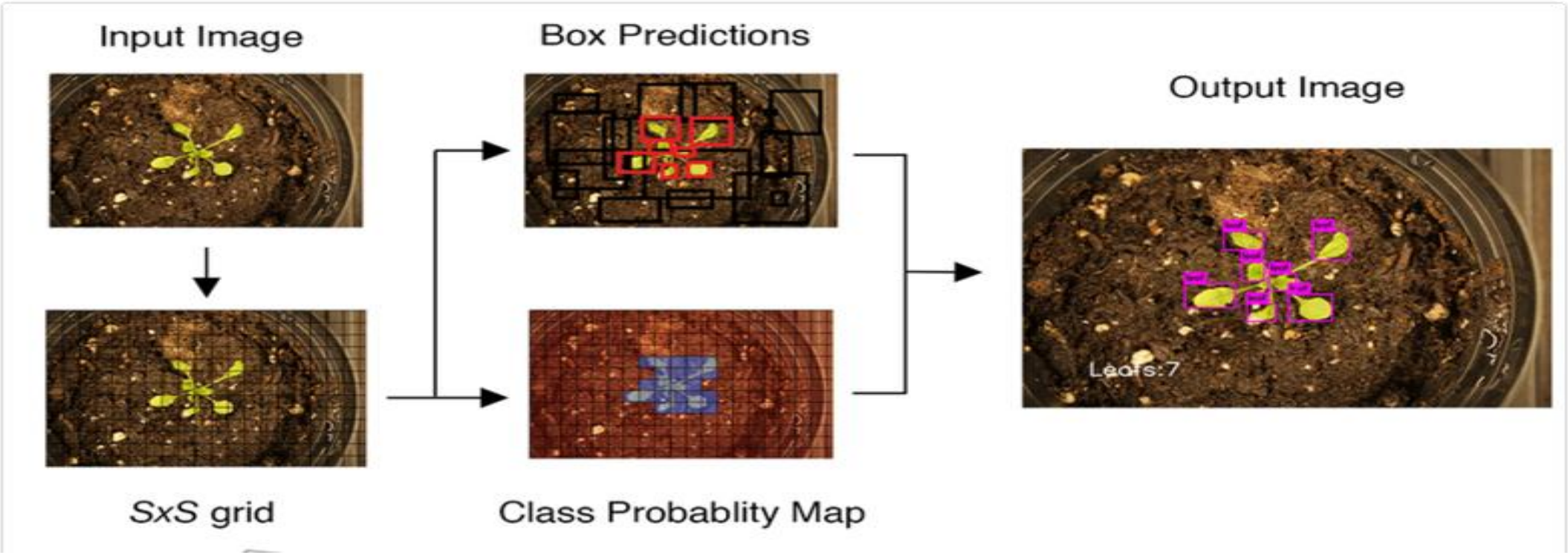
In case of have multiple objects in the same cell :

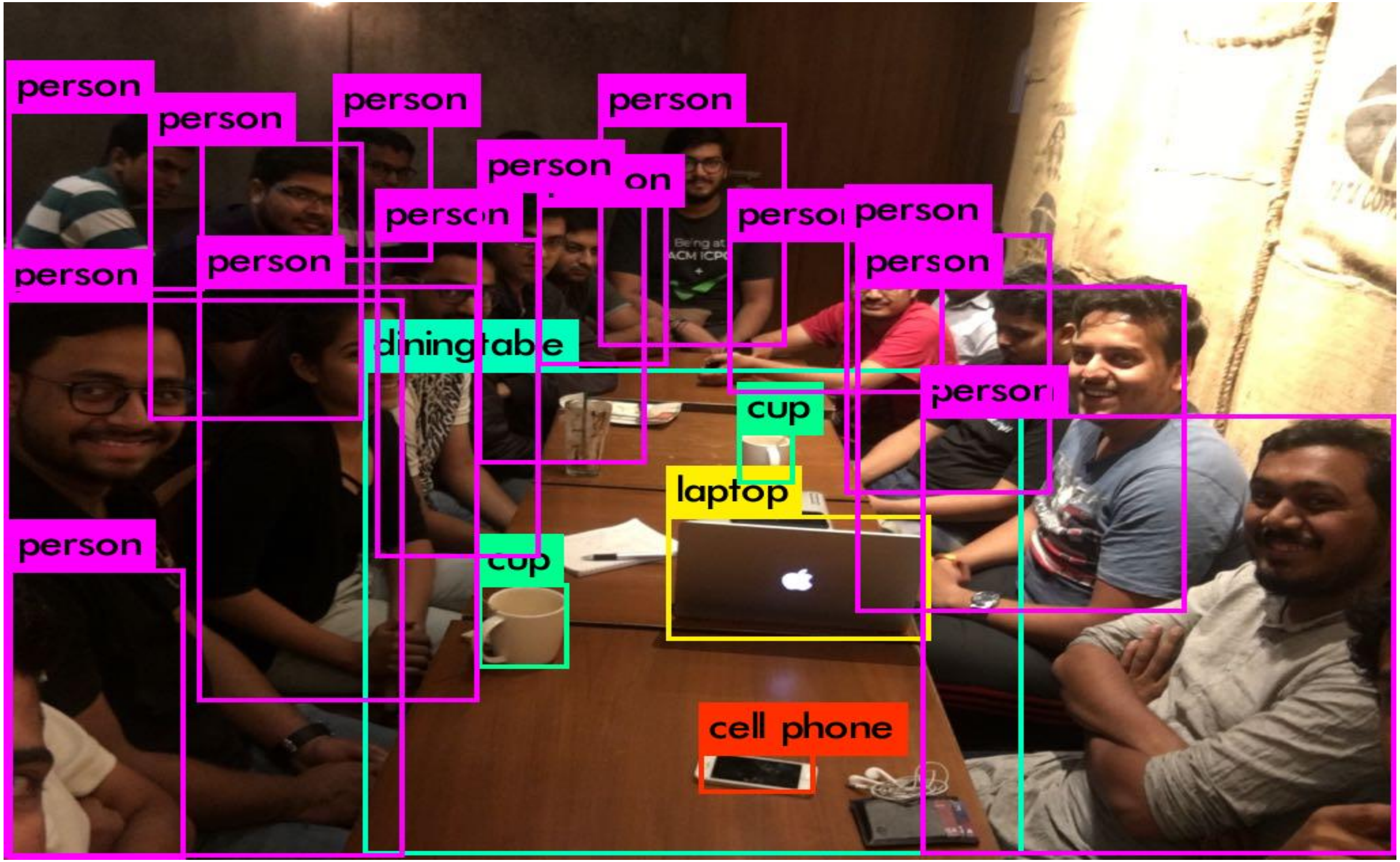
The vector can be improved to contain the predictions of the two objects



Finally it detects the object in the image




Input Image
Box Predictions
Output Image
Leafs:7
SxS grid
Class Probablity Map


person
person
person
person
person
person
person
person
person
person
person
diningtable
cup
laptop
cup
cell phone

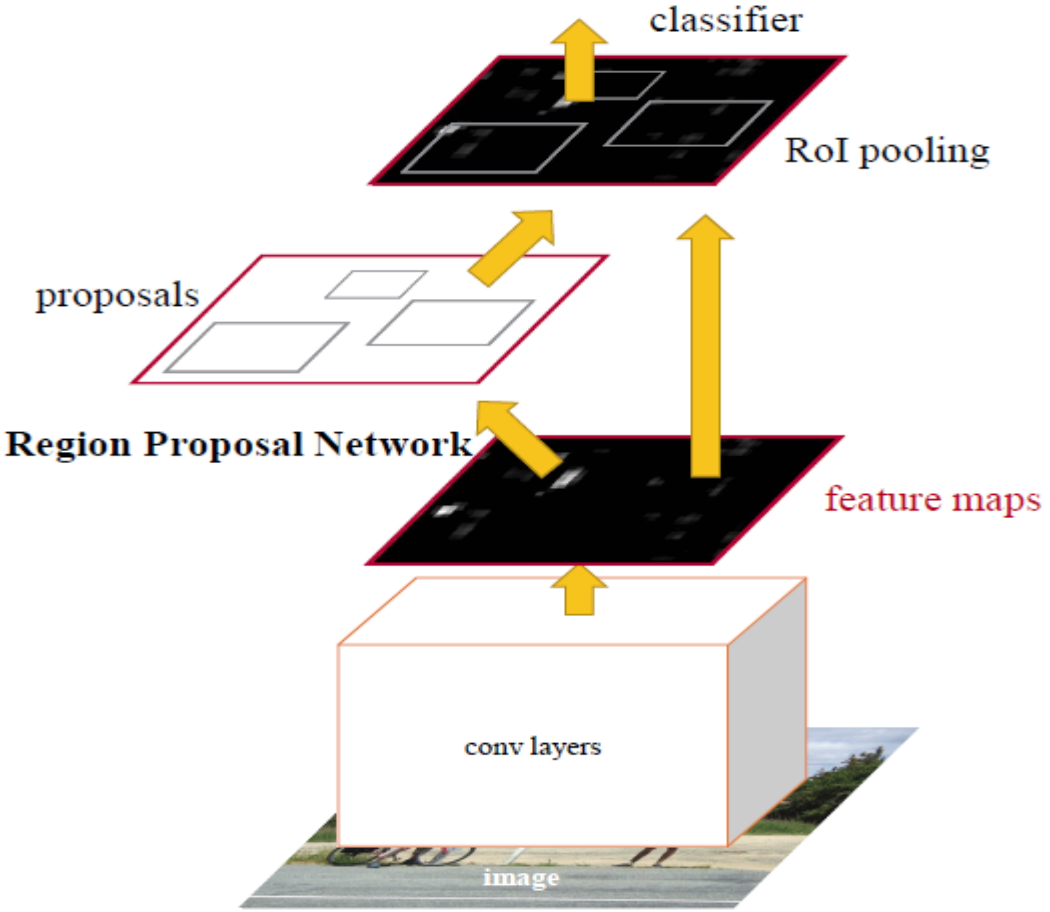
YOLO or Faster R-CNN?

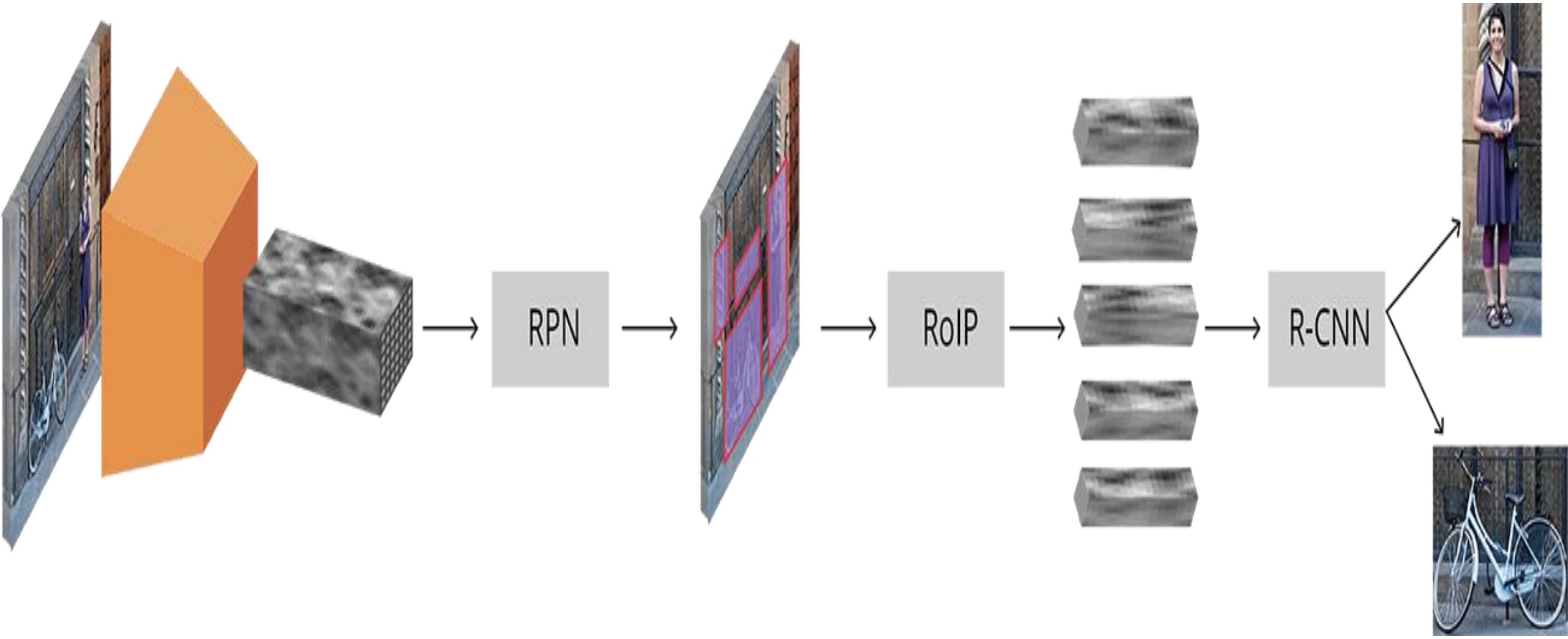
Faster R-CNN:

It is the closer algorithm to YOLO as it offers end-to-end training but more complex architecture

RPN is trained to produce region proposals directly without the need for any external mechanism like Selective Search (RPN is a CNN functioning by predicting object bounds and scores for those bounds simultaneously).

After this we use ROI pooling and an upstream classifier and bounding box regressor similar to Fast R-CNN.




RPN
RoIP
R-CNN

YOLO or Faster R-CNN?

YOLO

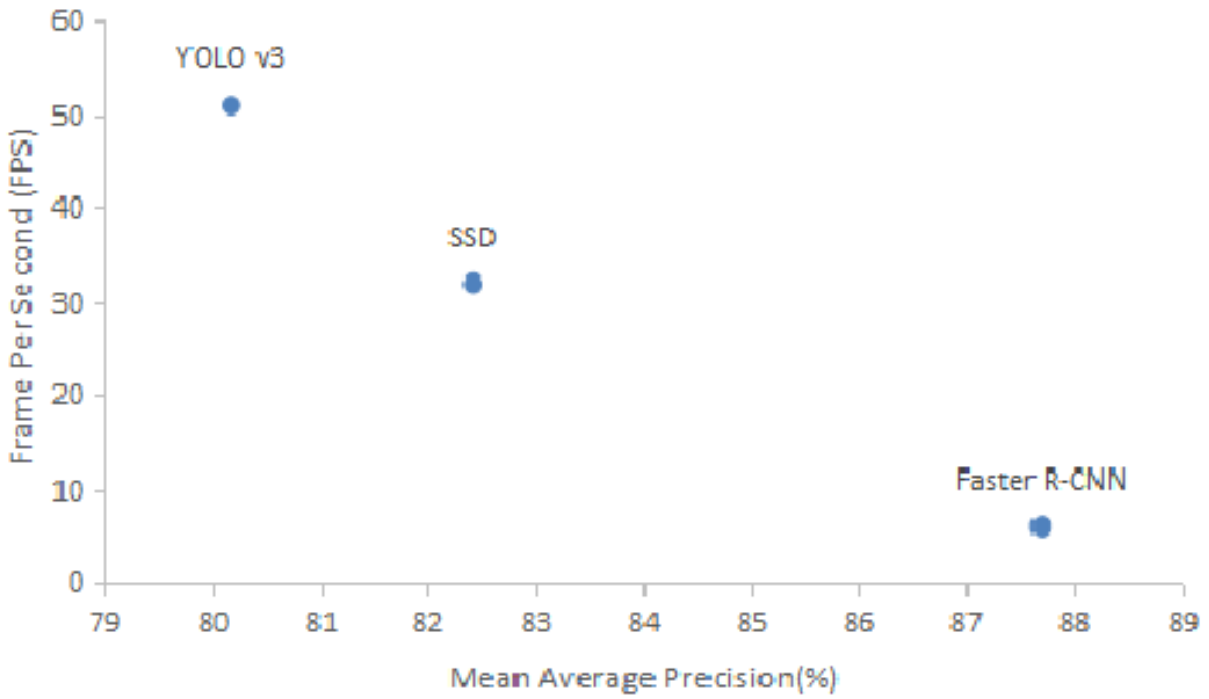
- Single Stage
- Running at as high as 45 FPS.
- Work at real time well
- Divides the whole image at the very beginning and later uses CNN to process.
- YOLOv2 uses 5 anchor boxes.
- YOLOv2 creates many new techniques to improve precision, such as dimension clusters: uses three feature maps with different scales to predict the bounding box
- YOLO v3 can operate on low performance platforms and in environments with requirements for high speed of detection.

Faster R-CNN

- Two Stages
- Running at as high as 21 FPS
- Detect small objects well
- Has 9 anchor boxes
- Work near to the real time
- keeps the traditional general frame of R-CNN: using CNN dealing with the whole input image at first and dividing proposals later

YOLO or Faster R-CNN for Real-Time Pill Identication?

- The Faster R-CNN model has a high MAP (87.69%), but the detection speed (FPS : 7) is not fast enough for real-time application.
- YOLO v3 does not have the highest MAP (80.17%), it can greatly improve the detection speed and achieve real-time performance (FPS : 51).



In busy hospital pharmacies, pill identification requires not only a high enough MAP, but also detection speed.

YOLO or Faster R-CNN for drone detection?

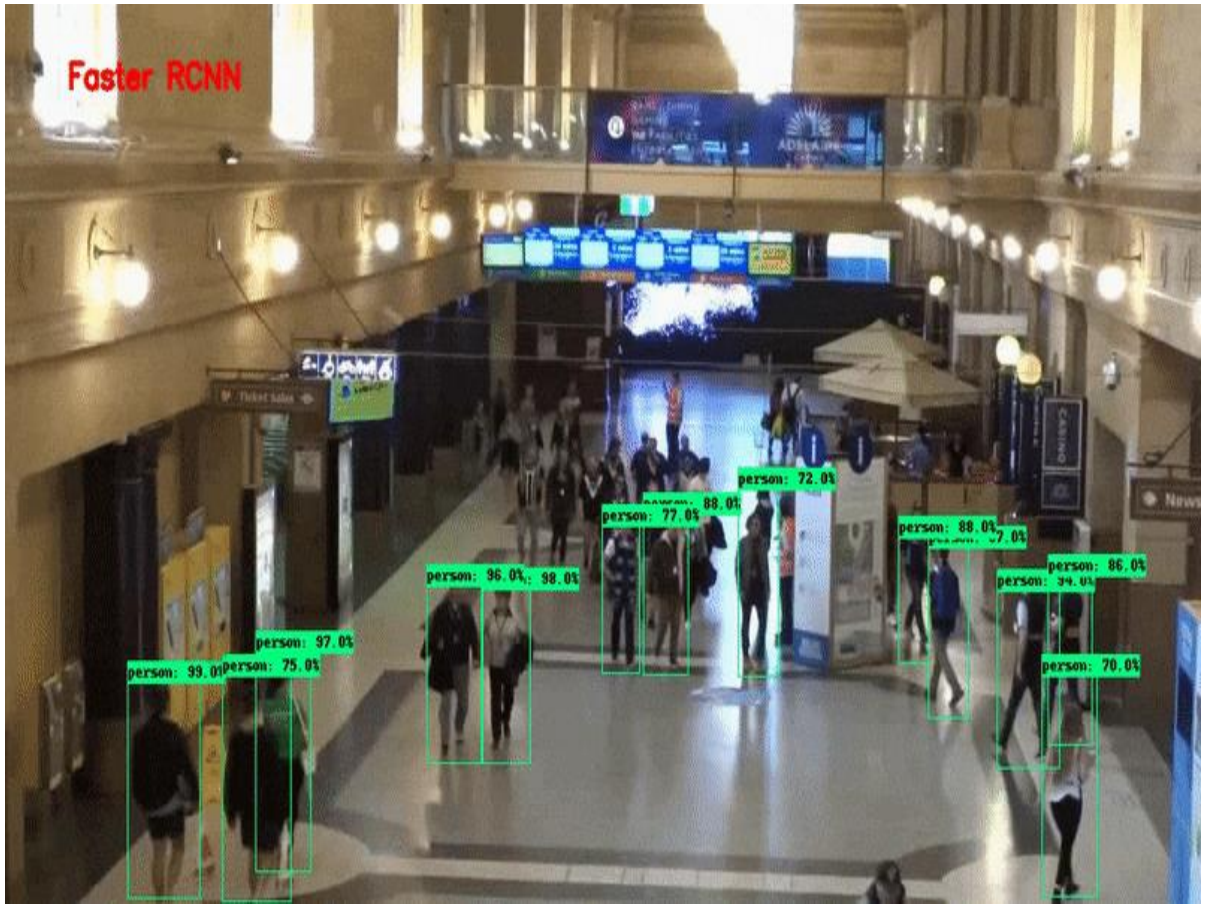
- this study showed that Faster RCNN still has an edge on accuracy and detection ability over other algorithms but suffers slow speed.

Method	FPS	mAP%
Faster 1000	7.69	**6.05**
Faster 544	16.10	**1.78**
Faster 300	27.18	**0.78**
SSD 300	27.93	**0.84**
YOLO 832	45.10	**1.74**
YOLO 312	208.48	**0.71**
YOLO 632	72.67	**1.13**
YOLO 544	87.22	**1.19**

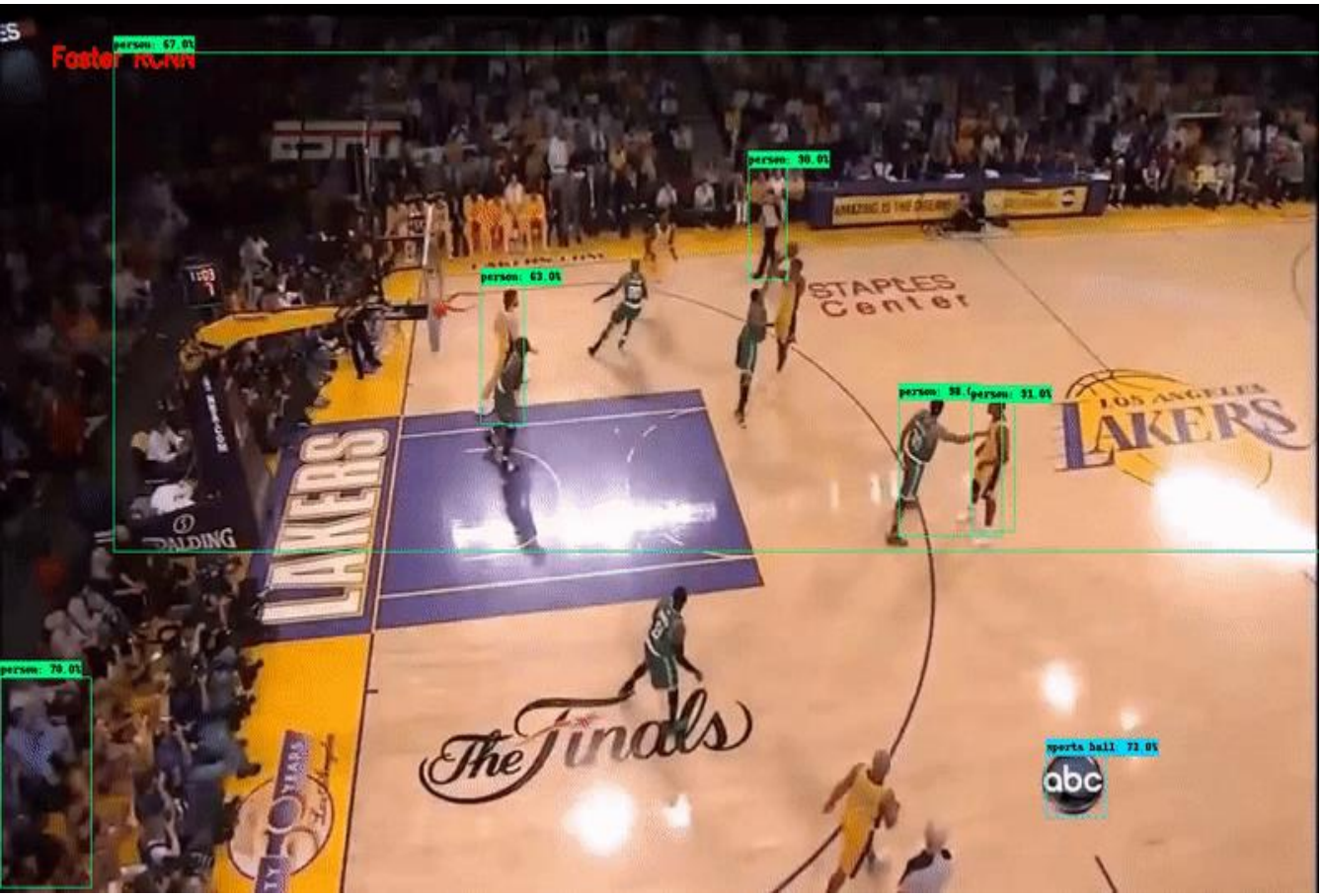
#	Method	TP	FP	FN	F1
1	Faster 1000	1779	0	2973	**0.54**
2	Faster 544	502	0	4250	**0.19**
3	Faster 300	88	0	4664	**0.03**
4	SSD 300	662	677	4090	**0.217**
5	YOLO 832	562	0	4190	**0.211**
6	YOLO 312	137	0	4615	**0.05**
7	YOLO 632	289	0	4463	**0.07**
8	YOLO 544	203	0	4549	**0.08**

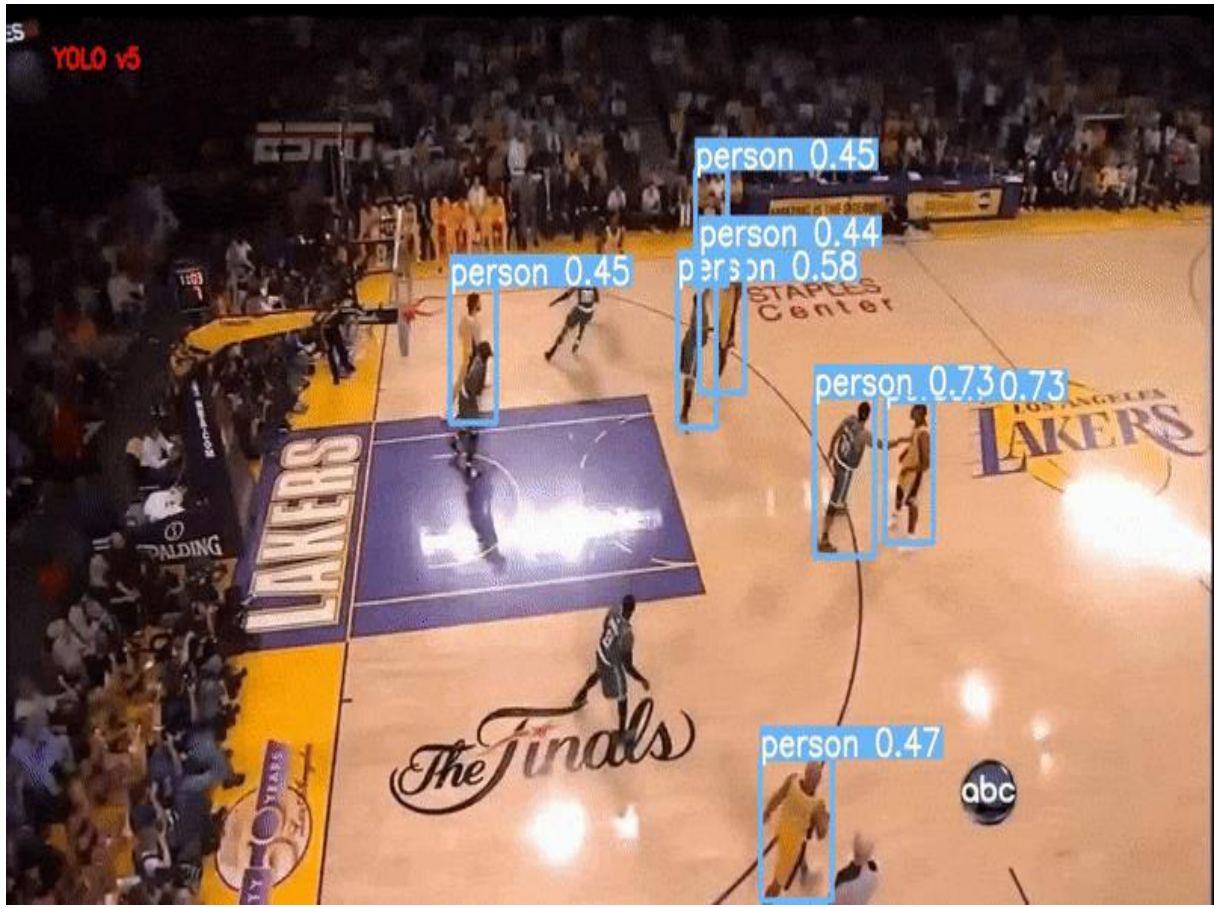
Detecting drones requires high accuracy, long range, and low false positives!

	YOLO v5	Faster RCNN
Inference Speed	✓	
Detection of small or far away objects	✓	
Little to no overlapping boxes	✓	
Missed Objects	✗	✗
Detection of Crowded objects	✓	✓


Faster RCNN
person: 99.0%
person: 75.0%
person: 97.0%
person: 96.0%
person: 98.0%
person: 77.0%
person: 72.0%
person: 88.0%
person: 86.0%
person: 70.0%


YOLO v5
person 0.46
person 0.55
person 0.43
person 0.58


Faster RCNN
person: 67.0%
person: 30.0%
person: 63.0%
person: 98.1%
person: 91.0%
person: 79.0%
sports ball: 72.0%
STAPLES Center
LAKERS
The Finals
abc


YOLO v5
person 0.45
person 0.44
person 0.45
person 0.58
person 0.73
0.73
person 0.47
STAPLES Center
LAKERS
The Finals
abc