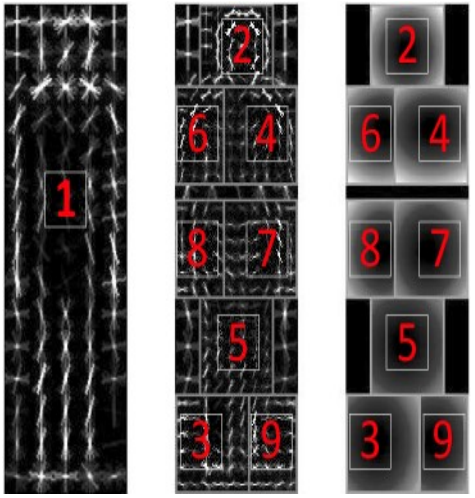
YOLO
(You Only Look Once)

R170023
R170010
R170360

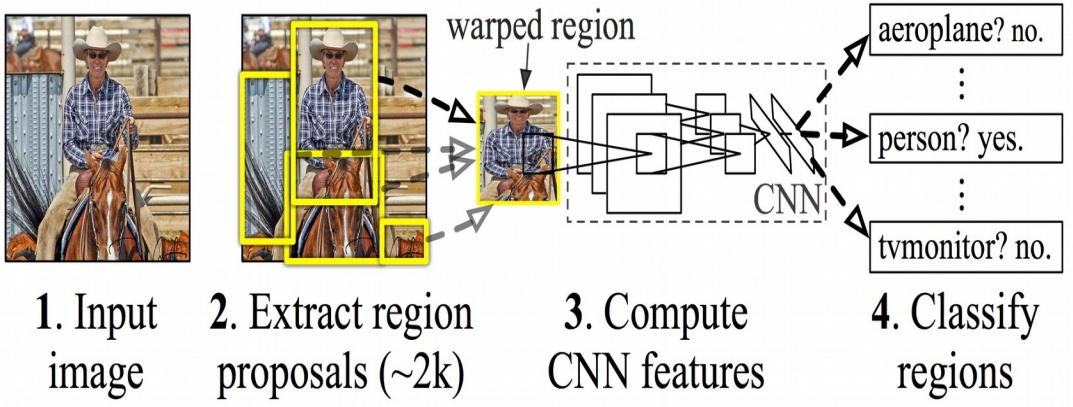
INTRODUCTION

- Object detection is a computer technology related to computer vision and Image processing
- Fast,accurate algorithms for object detection would allow computers to drive cars without specialized sensors,enable assistive devices to convey real time scene information to human users.
- Current detection systems repurpose classifiers to perform detection means the systems take a classifier to detect object and evaluate it at various locations and scales in a test image.
- You Only Look Once (YOLO),a new approach to object detection.

Previously : Object Detection by Classifiers

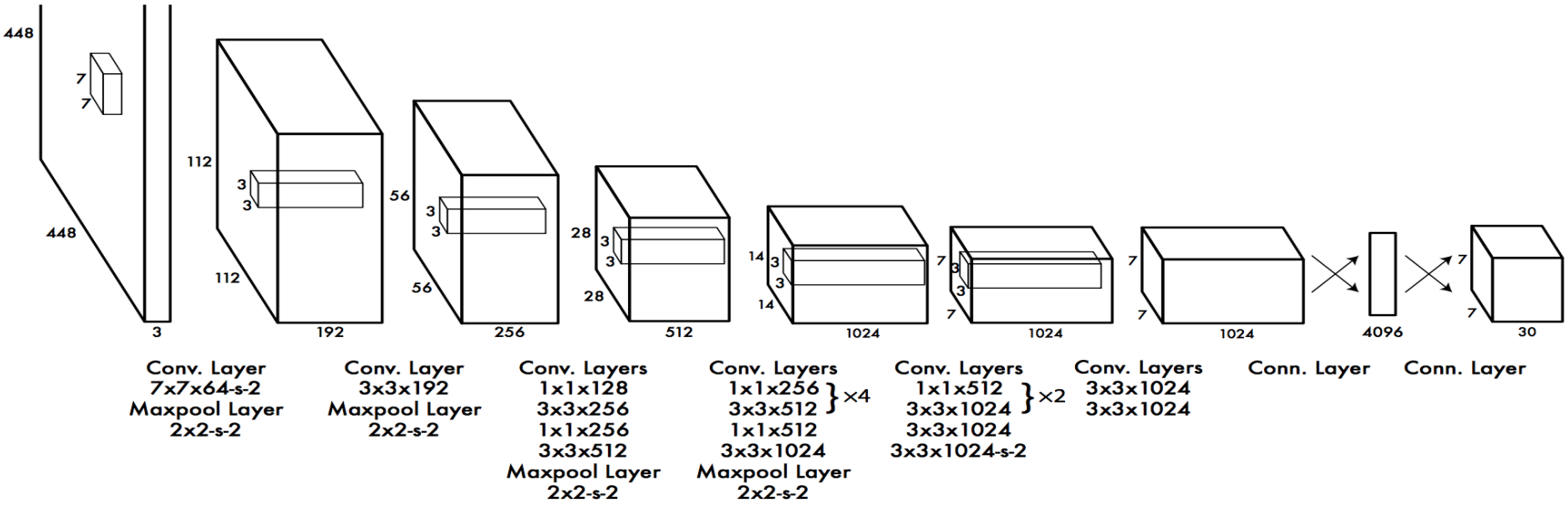


- **DPM (Deformable Parts Model)**
  - use a sliding window approach where the classifier is run at evenly spaced locations over the entire image.
- **R-CNN**
  - use regional proposal methods to first generate potential bounding boxes
  - run a classifier on the proposed boxes
  - Post processing
- **YOLO**
  - Resize image,run convolutional network non-max suppression.



YOLO

- A single regression Problem (single neural network),straight from image pixels to bounding box coordinates and class probabilities
- Predicts directly from full images in one evaluation,about all classes
- The Architecture : Inspired by the GoogleNet
- 24 convolutional layers followed by the 2 fully connected layers.



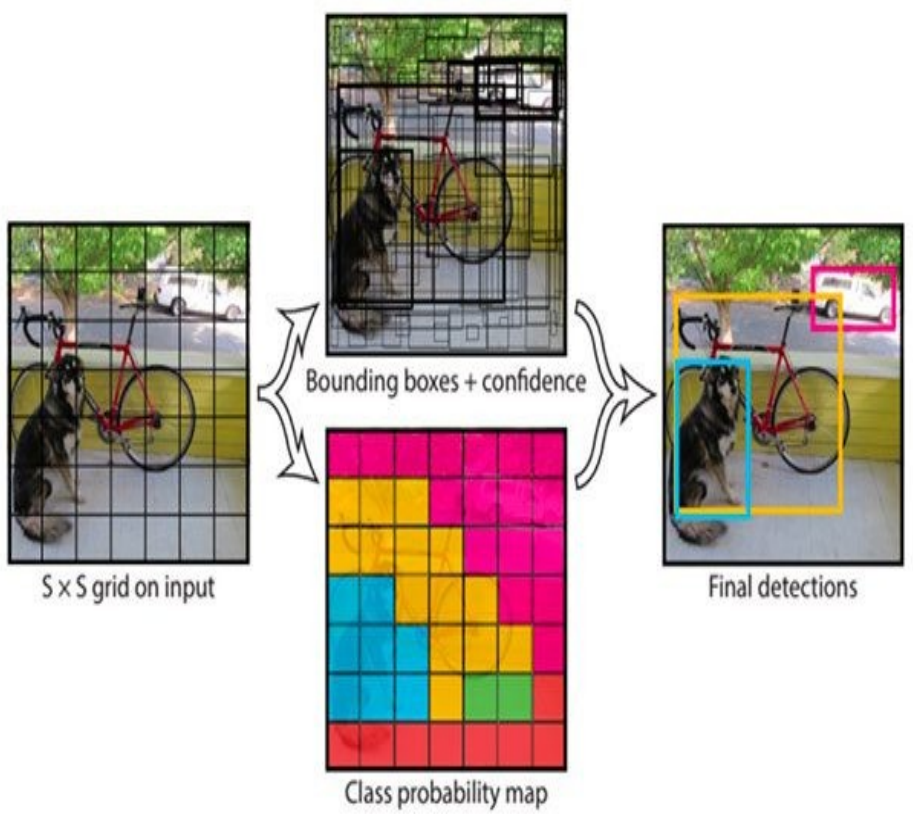
The Model

- This models detection as a regression problem.
- It divides the image into an S*S grid
- Each grid cell predicts B bounding boxes,

  Confidence for those boxes and

  C class probabilities.
- These predictions are encoded as an

  S*S*(B*5+C) tensor



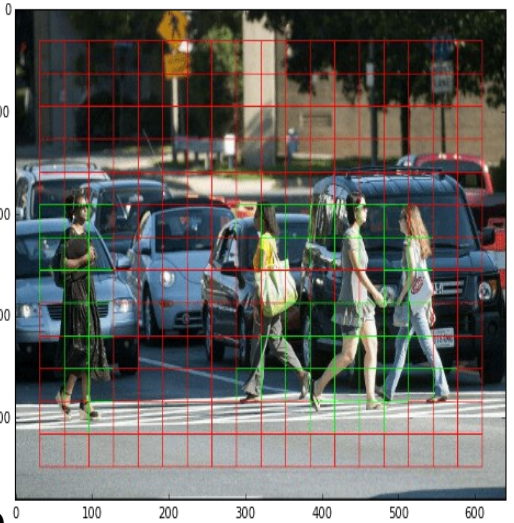
How the YOLO algorithm works

- **Residual blocks**

  The input image is divided into grids,each grid has dimension s*s

  if an object center appears within a certain grid cell, then this cell

  will be responsible for detecting it

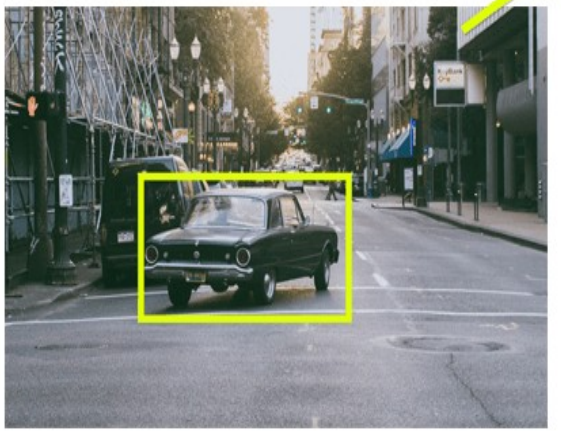


- **Bounding box regression**

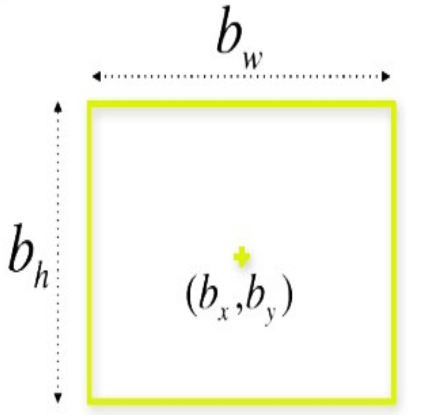
  A bounding box is an outline that highlights an object in an image.

  Used to predict width (bw),height(bh)

  class(c),bounding box centre(bx,by).



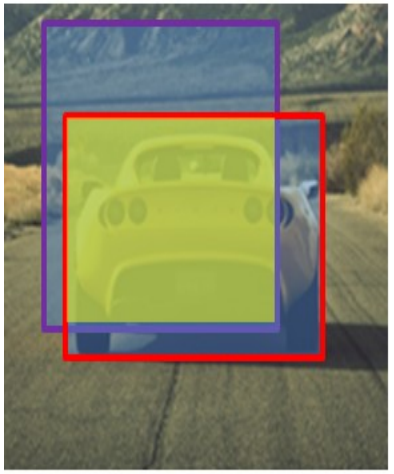
$$y = (p_c, b_x, b_y, b_h, b_w, c)$$



How the YOLO algorithm works

- **Intersection over union (IOU)**

  if IOU = 1,then the predicted bounding box is same as real box.so that it eliminates all remaining boxes which are not equal to real boxes.



Intersection over union (IoU)

$$= \frac{\text{size of}}{\text{size of}}$$

- **non-max suppression(NMS)**

  discard all boxes with Pc<=0.6

  for remaining boxes

  pick largest Pc box and discard boxes with IOU>=0.5 with the box o/p in the previous step.

"Correct" if IoU $\geq$ 0,5

Limitations & Advantages of YOLO

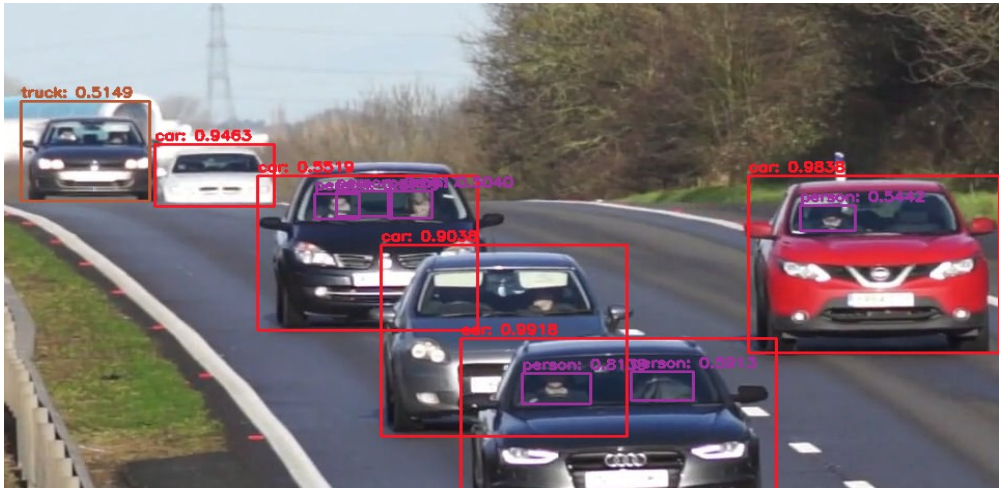
Limitations

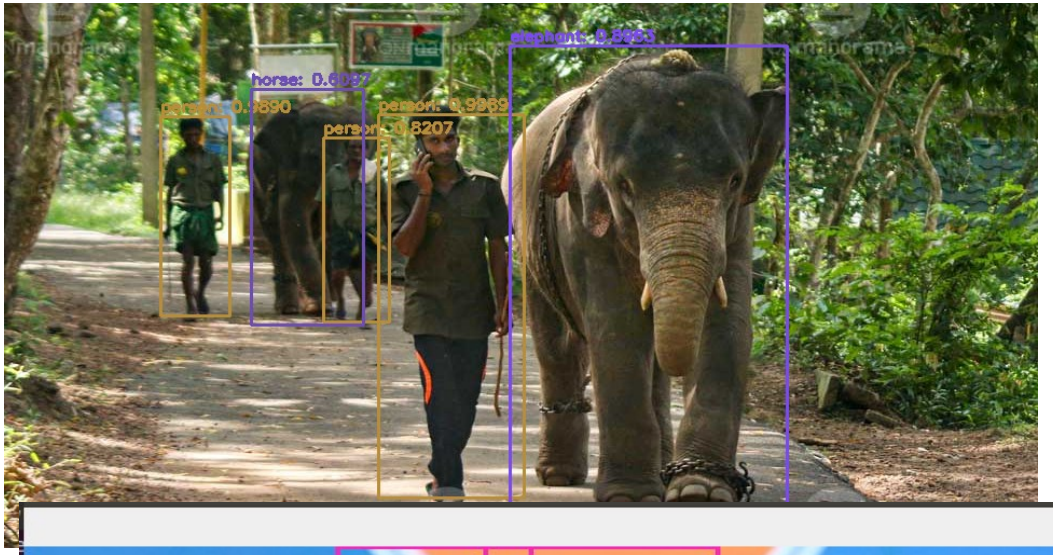
- Struggles with small objects that appear in groups,such as flock of birds.
- Struggles to generalize with different aspects and ratio of objects
- Loss function treats errors as same for both small and large bounding boxes.

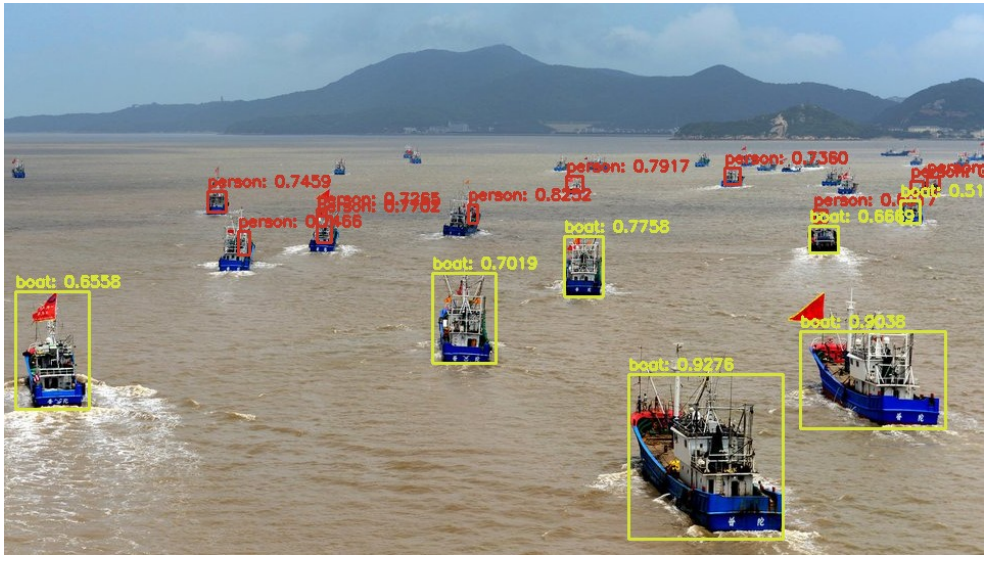
Advantages

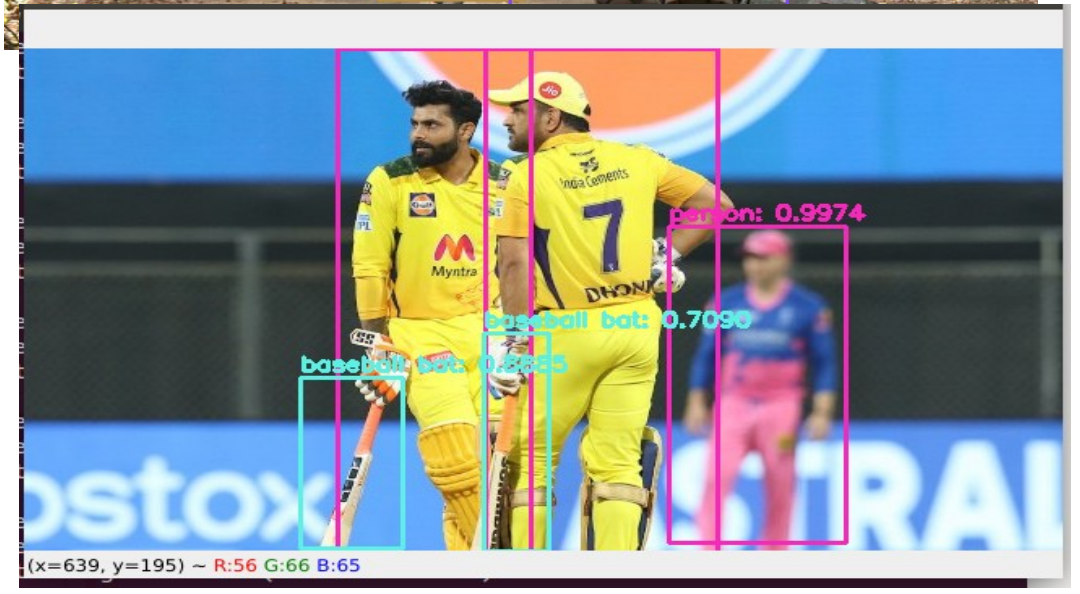
- Speed
- High Accuracy
- Learning capabilities
- Robust object detection

Results


truck: 0.5149
car: 0.9463
car: 0.9838
person: 0.5442


horse: 0.6097
person: 0.9889
person: 0.8207


person: 0.7459
person: 0.8252
person: 0.7917
person: 0.7360
boat: 0.6558
boat: 0.7758
boat: 0.7019
boat: 0.9276


person: 0.9974
baseball bat: 0.7090
(x=639, y=195) ~ R:56 G:66 B:65