Social Distancing Using YOLO V3

MACE

Outline

- Aim of My Application
- Review Object Detection
- Person Detection/Mask detection

Aim of My Application:

- Detect Person/person with mask from images and live video frames
- Detect Crowd
- Calculate safe distance (Social Distance tracking)

Crowd Management:

Due to the growing population,Crowd management become an important issue and challenge for security agencies across the world where effective crowd management can prevent serious accidents and in some cases mortalities.

DataSet Exploration

Datasets

- COCO (Common Objects in Context) data set for object detection.
- Class of the dataset :Person

  Train images : 83K
  Test images : 41K
  Validation images : 41k

- COCO is a set of challenging, high quality datasets for computer vision.
- COCO consists of 80 classes

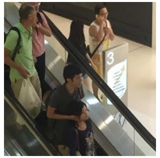
Datasets

COCOPerson


CrowdHuman


WiderPerson


CUHK Person


WiderPerson19

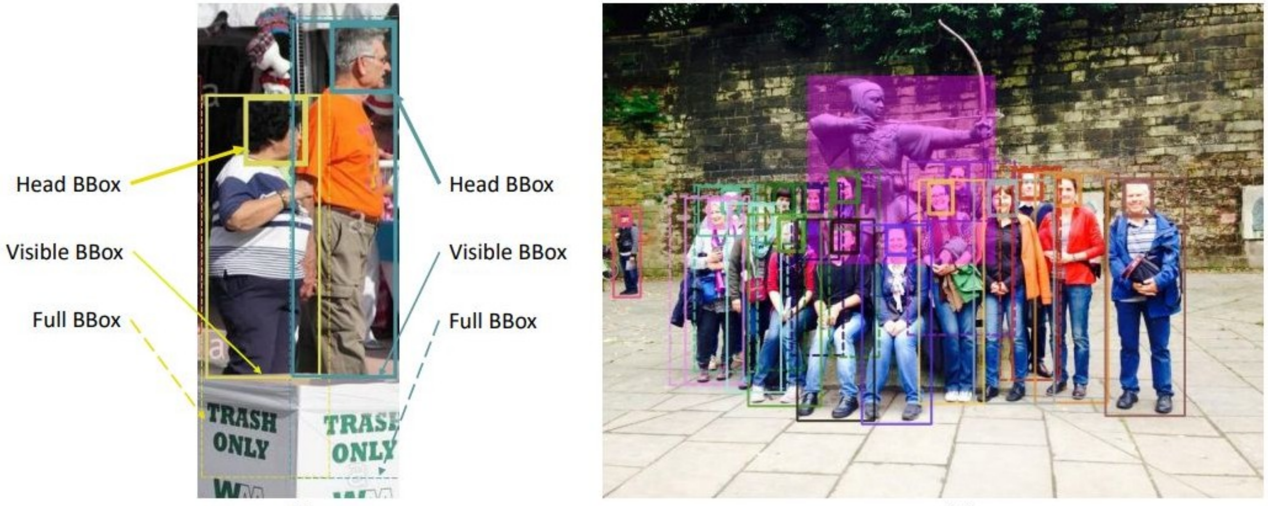

Caltech pedestrian


dataset	#of img	#of person	density
COCO Person	64,115	257,252	4.01
CrowdHuman	15,000	339,565	22.64
WiderPerson	9,000	399,786	39.87
CUHK Person	18,184	99,809	5.48
WiderPerson19 sur/ad	8,240/ 88,260	58,190/ 248,993	7.05/ 2.82
Caltech pedestrian	72,782	13,674	0.32
CityPerson	2,975	19,654	6.61

train, **test**, **benchmark**

Dataset: COCOPerson

Dataset: CrowdHuman



(a) (b)

	Caltech	KITTI	CityPersons	COCOPersons	CrowdHuman
# images	42,782	3,712	2,975	**64,115**	15,000
# persons	13,674	2,322	19,238	257,252	**339,565**
# ignore regions	50,363	45	6,768	5,206	**99,227**
# person/image	0.32	0.63	6.47	4.01	**22.64**
# unique persons	1,273	< 2,322	19,238	257,252	**339,565**

Annotations

- Full box
- Visible box
- Head box

Features

- Aim Crowdness issue

Dataset: CrowdHuman

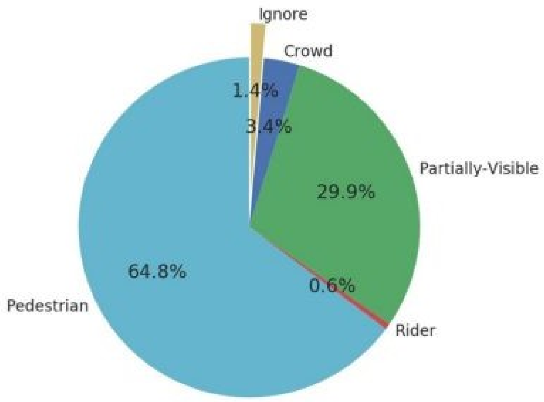
Dataset: WiderPerson

Features

- Questionable annotation quality
- Limited scence distribution (by observation)

Annotations

- Full box
- class, tag

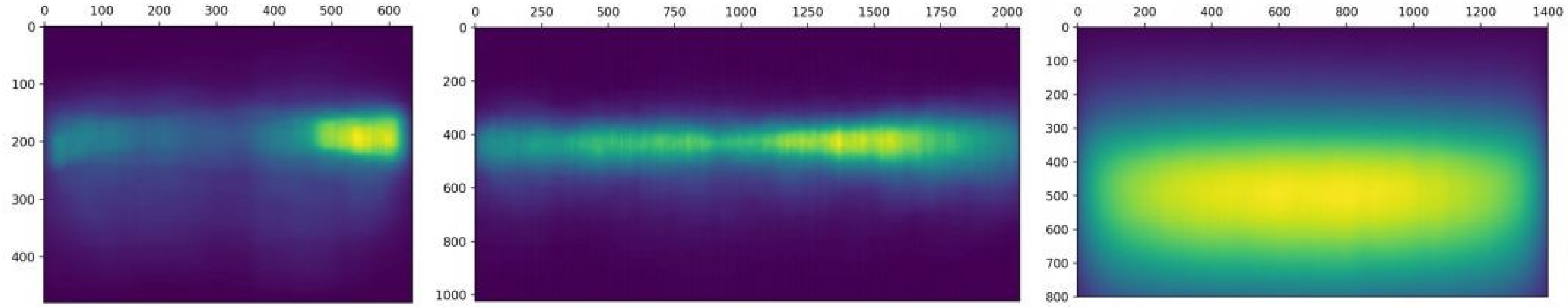


TMM2019 http://www.cbsr.ia.ac.cn/users/jwan/papers/TMM2019-WiderPerson.pdf

Dataset: WiderPerson

Features

- More balanced location distribution



(a) Caltech-USA (b) CityPersons (c) WiderPerson

Fig. 6. The location distribution of pedestrians on the image. Pedestrians on the Caltech-USA and CityPersons dataset are distributed in a narrow band across the center of the image, while WiderPerson has an uniform location distribution.

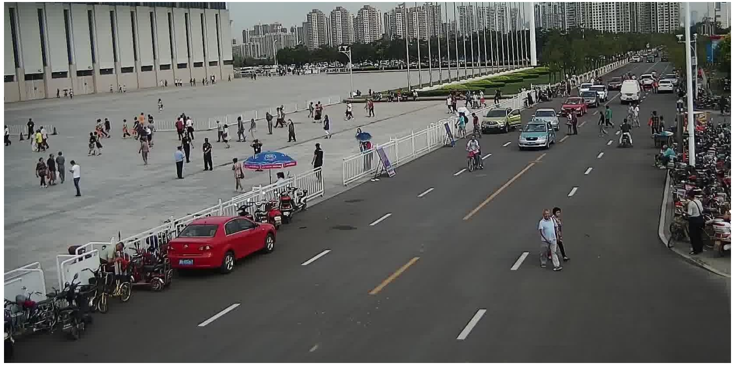
TMM2019 http://www.cbsr.ia.ac.cn/users/jwan/papers/TMM2019-WiderPerson.pdf

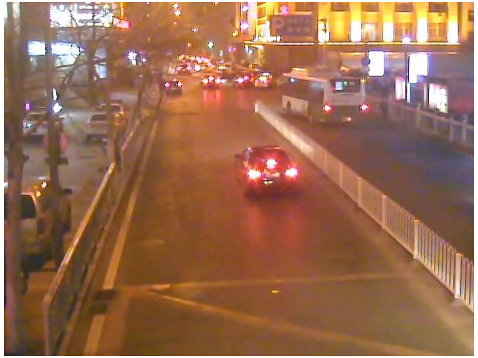
Dataset: WiderPerson





Dataset: WiderPerson2019





Features

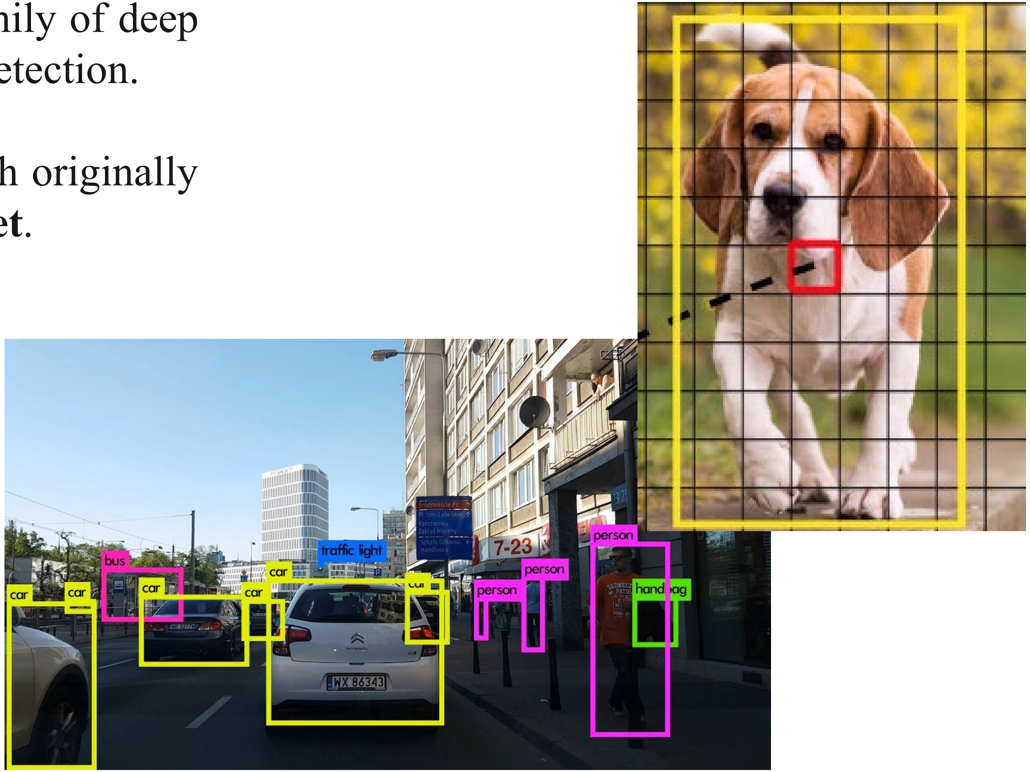
- vehicle & surveillance
- low quality but high resolution images

https://wider-challenge.org/2019.html

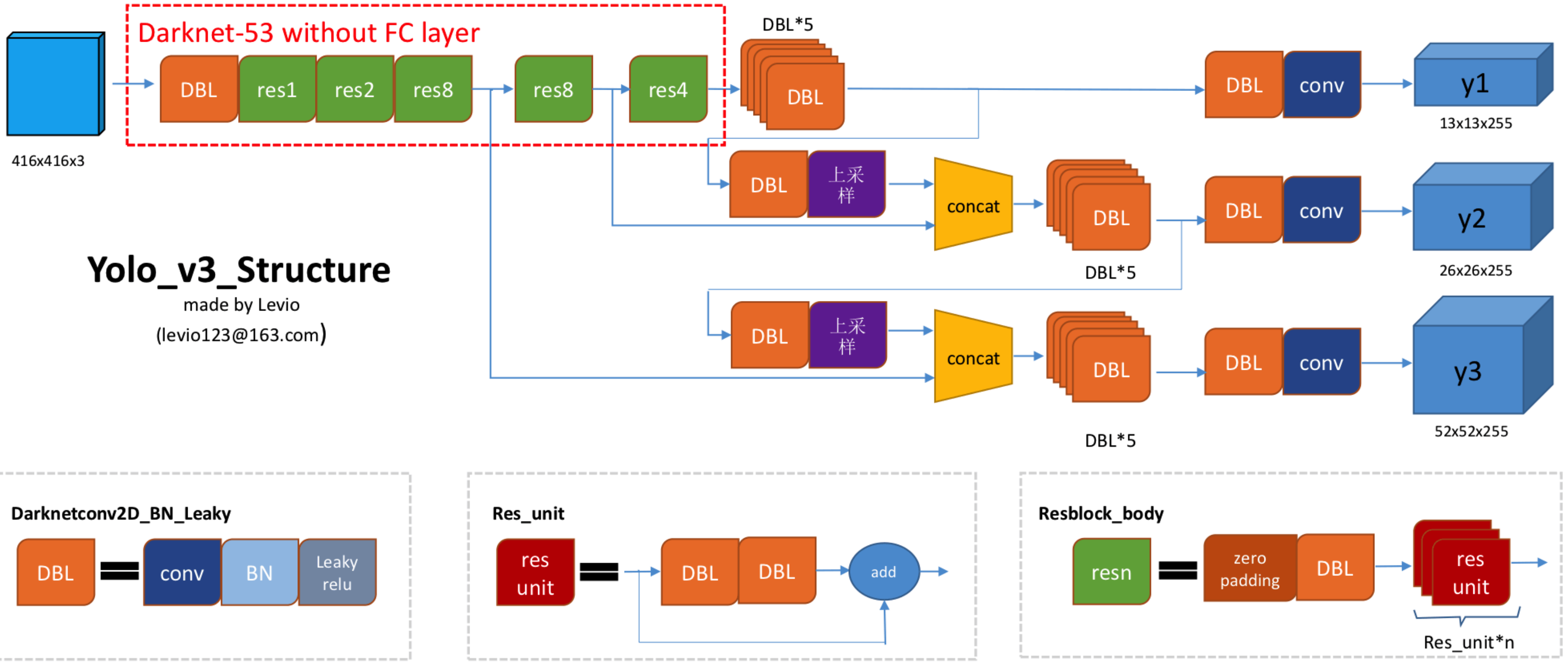
YOLO V3

YOLO V3

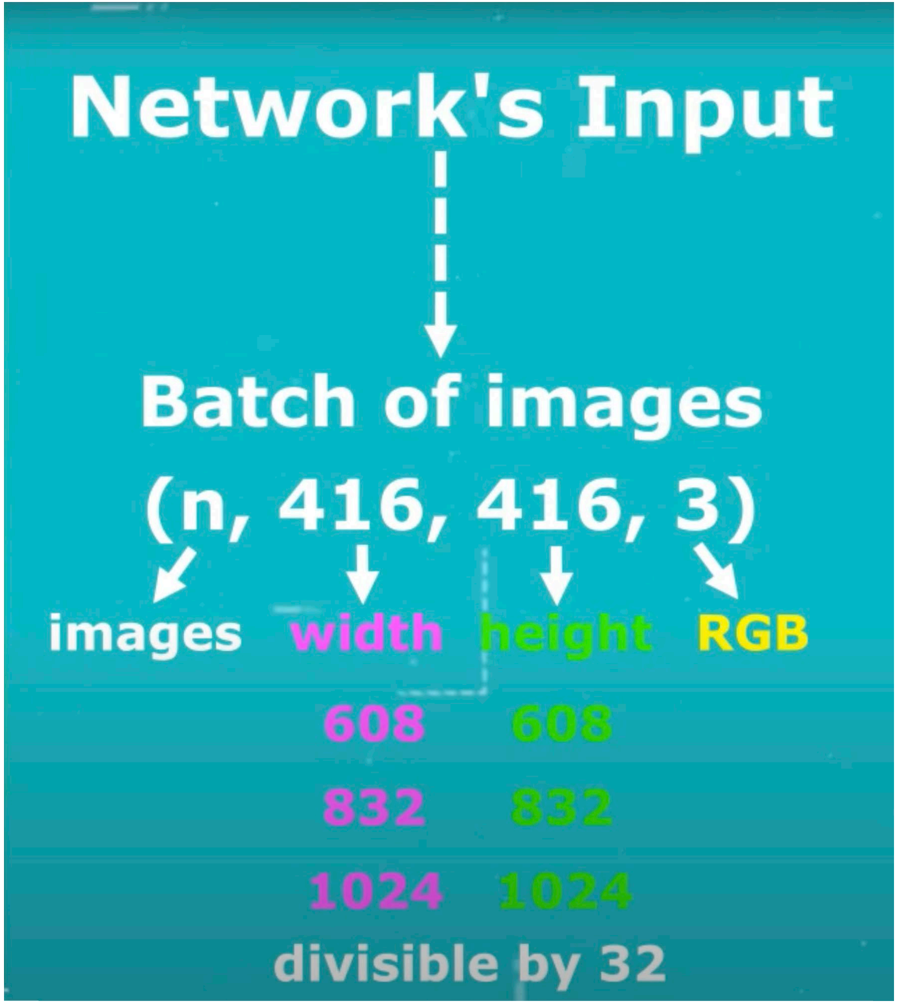
- "You Only Look Once" or YOLO is a family of deep learning models designed for fast object Detection.
- YOLO v3 uses a variant of Darknet, which originally has **53 layer network trained on Imagenet**.
- Classification & Localization



YOLO V3



Darknet-53 without FC layer
416x416x3
DBL
res1
res2
res8
res8
res4
DBL*5
DBL
DBL
conv
y1
13x13x255
DBL
上采样
concat
DBL
DBL*5
DBL
conv
y2
26x26x255
DBL
上采样
concat
DBL
DBL*5
DBL
conv
y3
52x52x255
Yolo_v3_Structure
made by Levio
(levio123@163.com)
Darknetconv2D_BN_Leaky
DBL
conv
BN
Leaky relu
Res_unit
res unit
DBL
DBL
add
Resblock_body
resn
zero padding
DBL
res unit
Res_unit*n


Network's Input
Batch of images
(n, 416, 416, 3)
images
width
height
RGB
608
608
832
832
1024
1024
divisible by 32

Network strides:

32, 16, 8

at layers

82, 94, 106

Size of outputs:

13x13, 26x26, 52x52

large medium small

Network's input

416 x 416

Equation:

✓ (b * (5 + c)),

where:

✓ b = 3, number of BB

✓ (5 + c), BB attributes

✓ c = 80, COCO classes

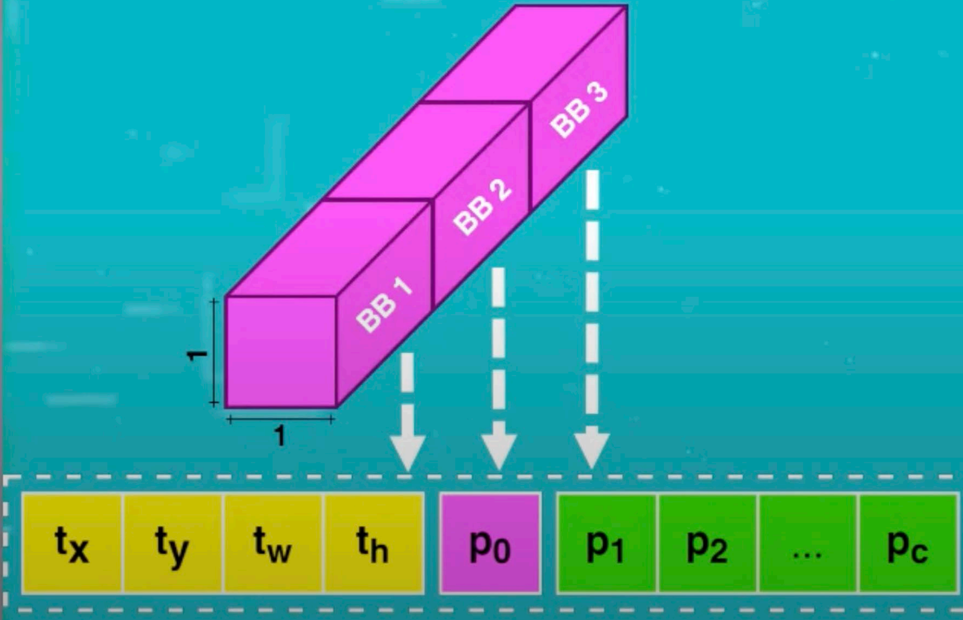
Result:

✓ (3 * (5 + 80)) = 255

Feature maps:

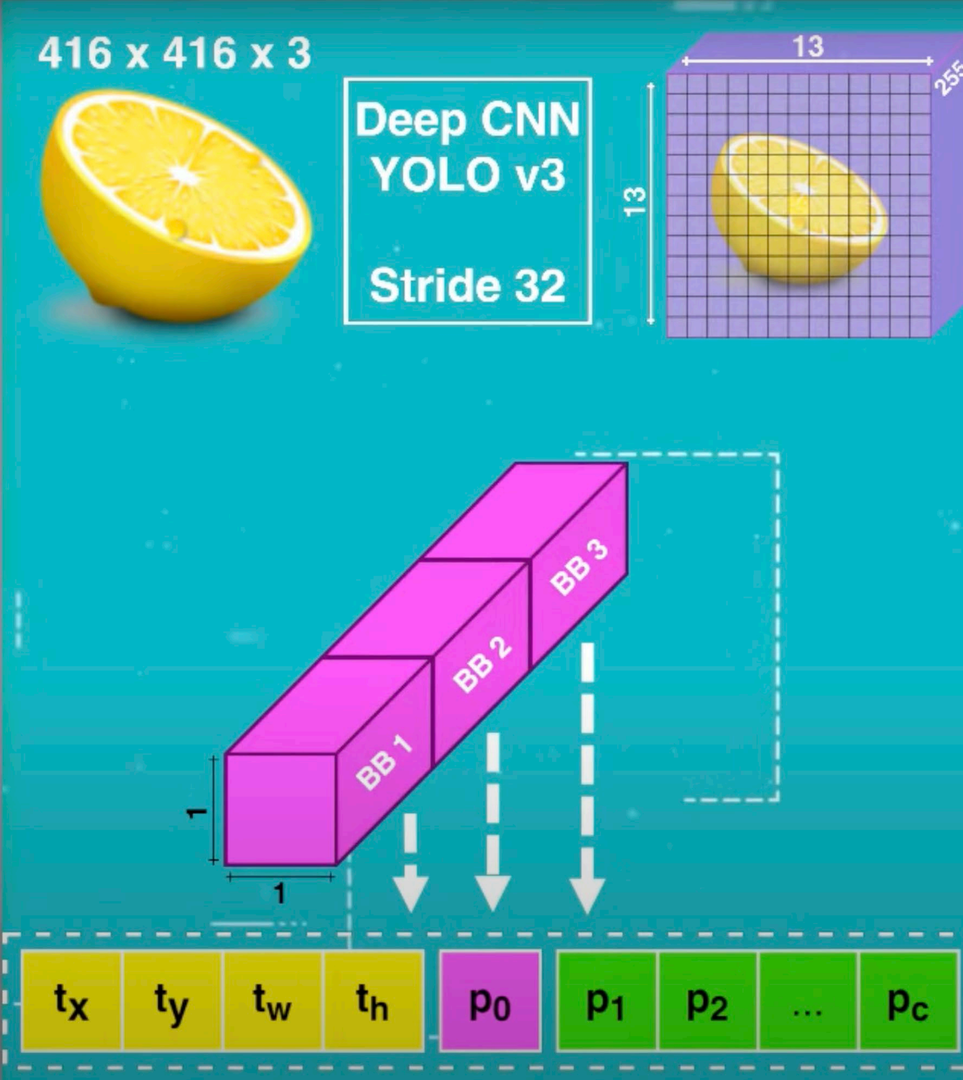
(13x13), (26x26), (52x52)

1x1 kernels' shape:



BB attributes obtained by:

- ✓ Input (416, 416, 3)
- ✓ Deep CNN of YOLO v3
- ✓ Downsampled input image by stride 32
- ✓ (13, 13, 255) feature map at Scale-1



BB probabilities:

- ✓ Extract probabilities of BB
- ✓ Compute elementwise product
- ✓ Find maximum probability

$$BB_1 \text{ score} = P_0 * \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ pc \end{bmatrix} = \begin{bmatrix} 0.08 \\ 0.03 \\ \vdots \\ 0.05 \end{bmatrix} \xrightarrow{\text{MAX}} \boxed{0.55}$$

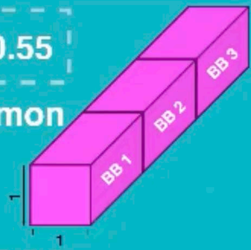
class: **lemon**

$$BB_2 \text{ score} = P_0 * \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ pc \end{bmatrix} = \begin{bmatrix} 0.02 \\ 0.07 \\ \vdots \\ 0.09 \end{bmatrix} \xrightarrow{\text{MAX}} \boxed{0.33}$$

class: **lime**

$$BB_3 \text{ score} = P_0 * \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ pc \end{bmatrix} = \begin{bmatrix} 0.01 \\ 0.07 \\ \vdots \\ 0.05 \end{bmatrix} \xrightarrow{\text{MAX}} \boxed{0.48}$$

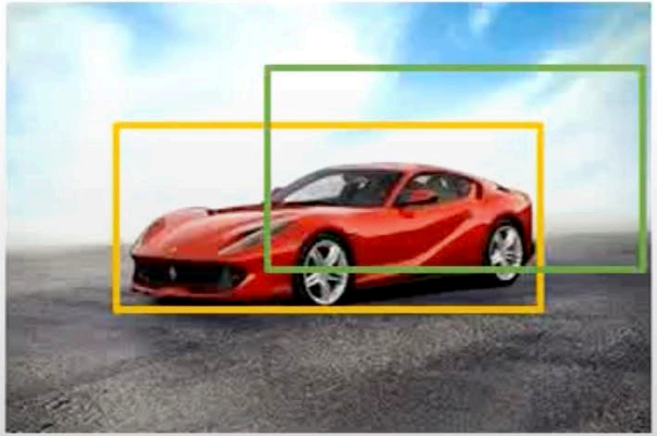
class: **orange**



IOU and NMS

Intersection over union (IoU) computes the intersection over union on these 2 bounding blocks

$$intersection\ over\ union = \frac{Size\ of\ intersection}{Size\ of\ union}$$

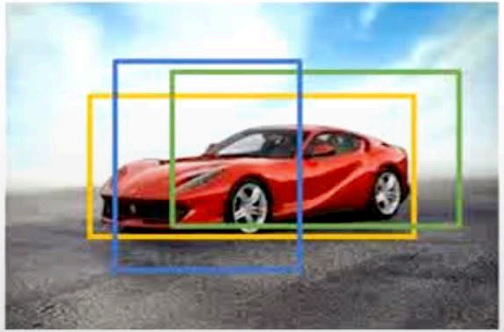


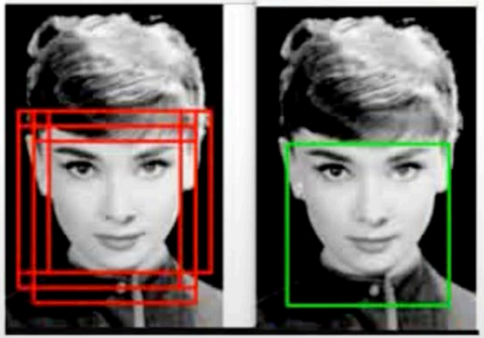
"Correct" if IoU ≥ 0.5

NMS

Non-max suppression is a common algorithm used for cleaning up when multiple boxes are predicted for the same object.

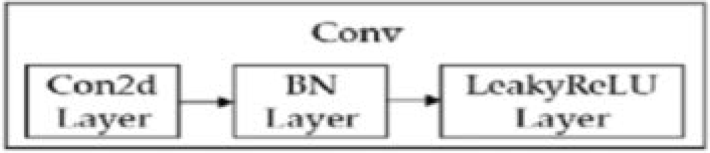
We have used 3 x 3 bounding boxes.

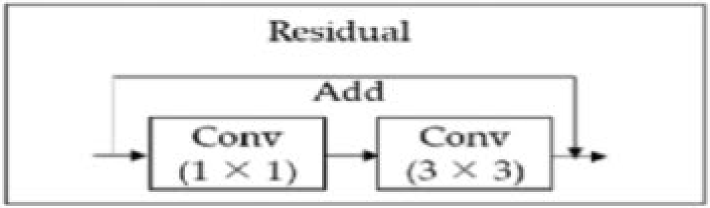




Dimension Table

Layer	Filters size	Repeat	Output size
Image			416 × 416
Conv	32 3 × 3/1	1	416 × 416
Conv	64 3 × 3/2	1	208 × 208
Conv Conv Residual	32 1 × 1/1 64 3 × 3/1	Conv Conv Residual × 1	208 × 208 208 × 208 208 × 208
Conv	128 3 × 3/2	1	104 × 104
Conv Conv Residual	64 1 × 1/1 128 3 × 3/1	Conv Conv Residual × 2	104 × 104 104 × 104 104 × 104
Conv	256 3 × 3/2	1	52 × 52
Conv Conv Residual	128 1 × 1/1 256 3 × 3/1	Conv Conv Residual × 8	52 × 52 52 × 52 52 × 52
Conv	512 3 × 3/2	1	26 × 26
Conv Conv Residual	256 1 × 1/1 512 3 × 3/1	Conv Conv Residual × 8	26 × 26 26 × 26 26 × 26
Conv	1024 3 × 3/2	1	13 × 13
Conv Conv Residual	512 1 × 1/1 1024 3 × 3/1	Conv Conv Residual × 4	13 × 13 13 × 13 13 × 13





Thank You