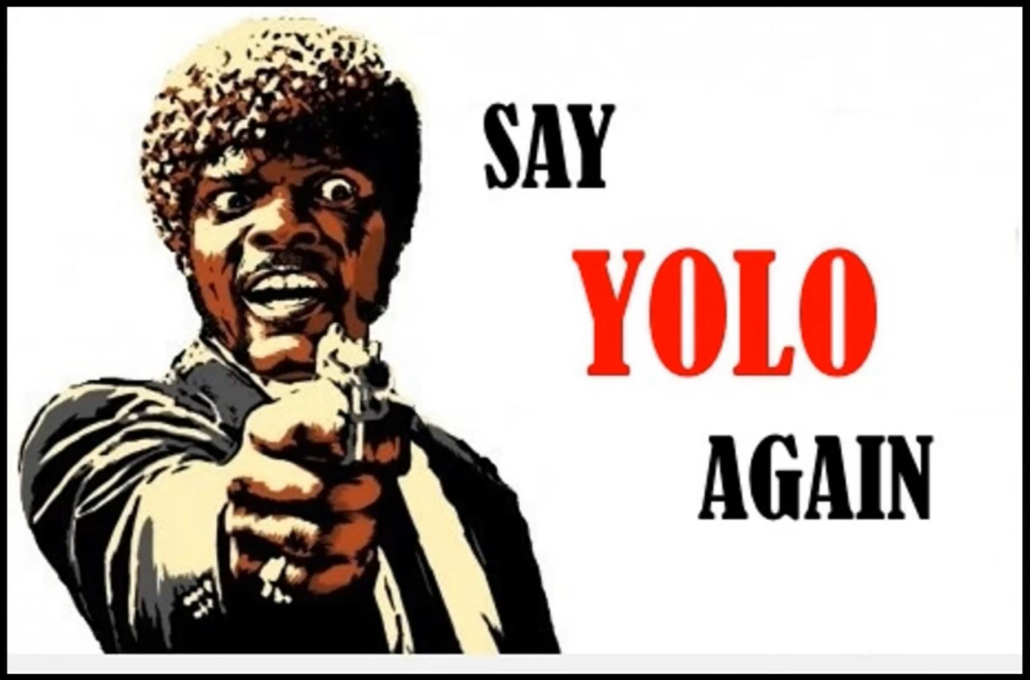


YOLO

You Only Look Once

Image Classification

Is this a dog or a person?

Neural Network Output

Dog = 1
Person = 0

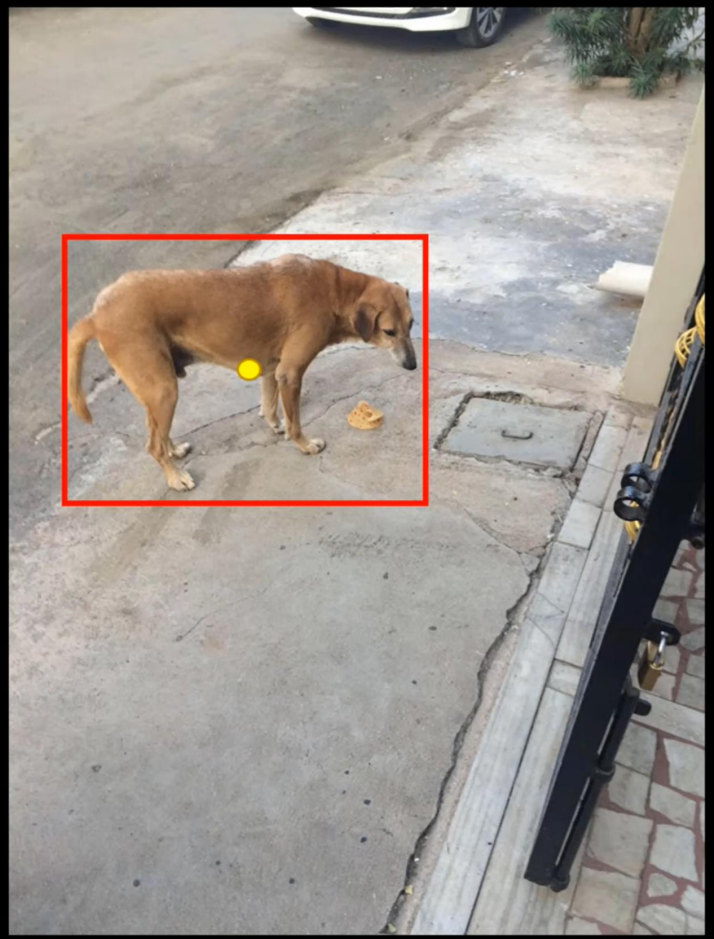
Object Localization

Where exactly is the dog in this image?

Neural Network Output

Dog = 1
Person = 0
+
Bounding Box

Object Localization

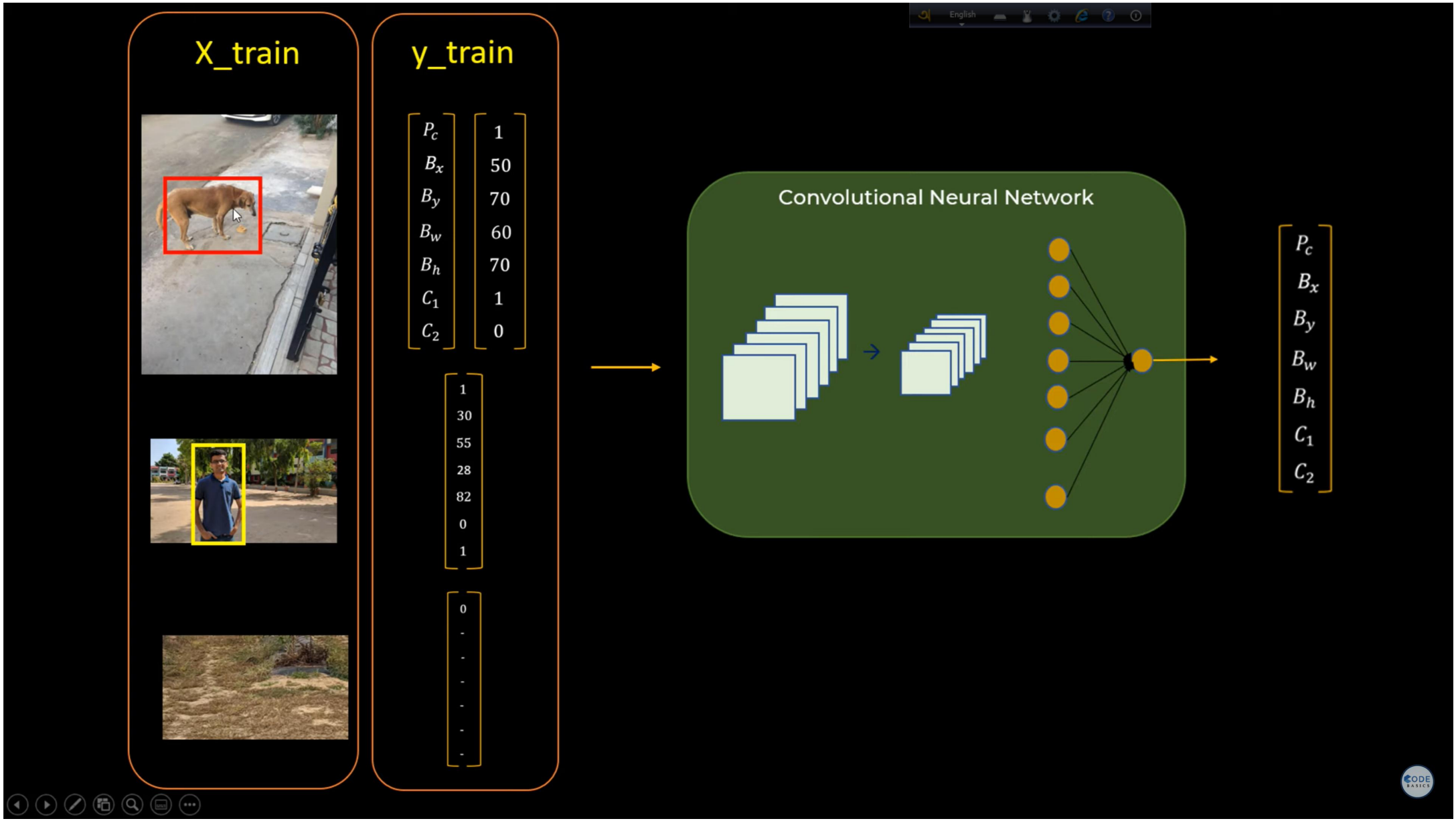


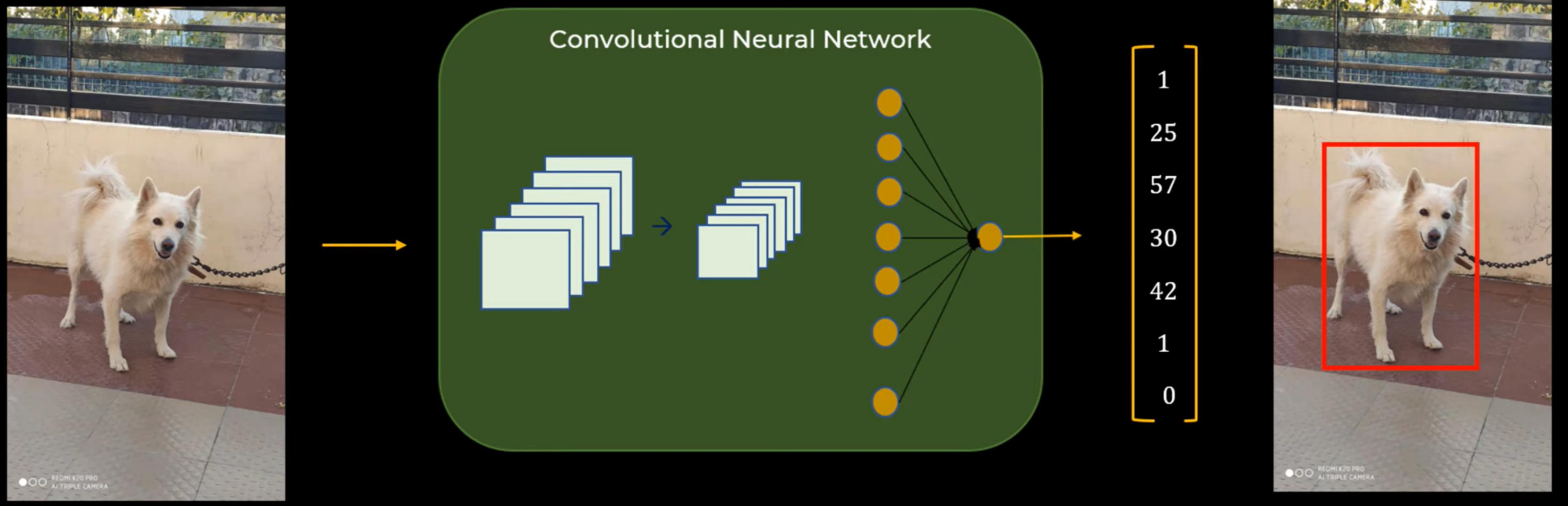
$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 50 \\ 70 \\ 60 \\ 70 \\ 1 \\ 0 \end{bmatrix}$$

C_1 = *Dog class*
C2 = Person Class

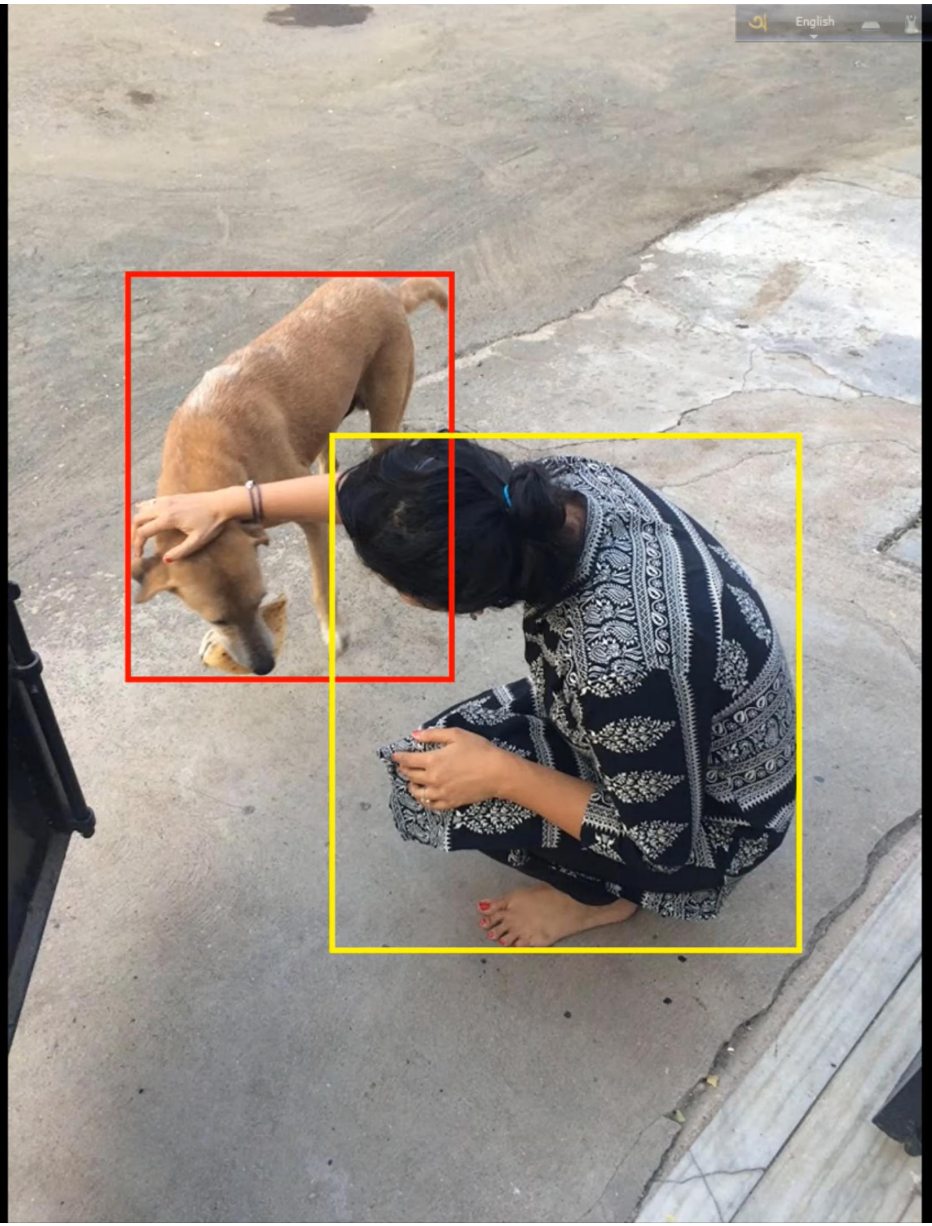
$$\begin{bmatrix} 1 \\ 30 \\ 28 \\ 28 \\ 82 \\ 0 \\ 1 \end{bmatrix}$$

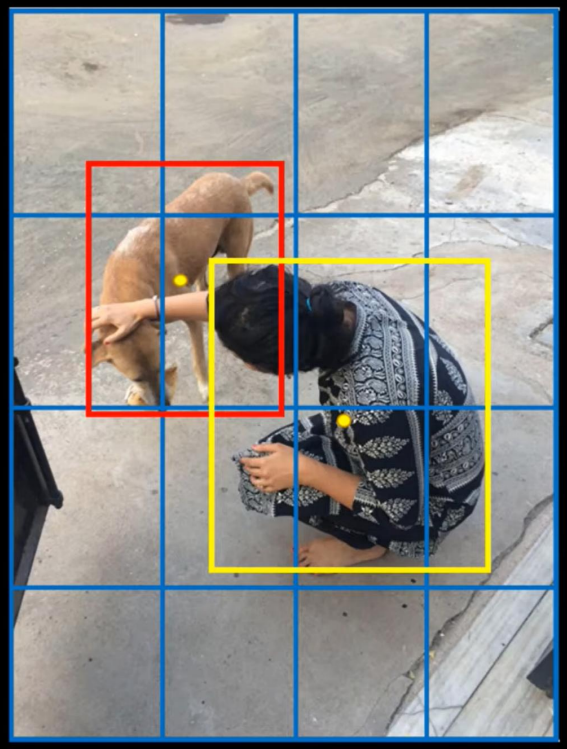
$$\begin{bmatrix} 0 \\ - \\ - \\ - \\ - \\ - \\ - \end{bmatrix}$$


X_train
y_train
P_c
B_x
B_y
B_w
B_h
C_1
C_2
1
50
70
60
70
1
0
1
30
55
28
82
0
1
0
-
-
-
-
-
-
Convolutional Neural Network
P_c
B_x
B_y
B_w
B_h
C_1
C_2

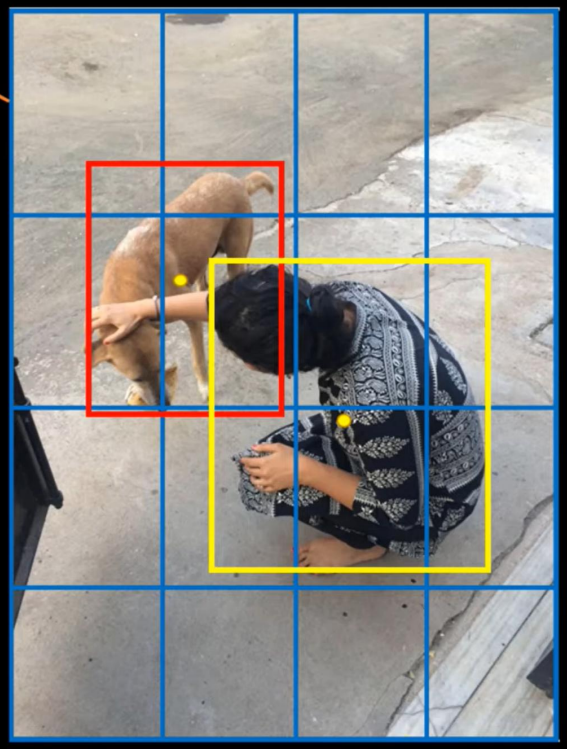

Convolutional Neural Network
1
25
57
30
42
1
0

This works ok only for single object. What about multiple objects in an image?

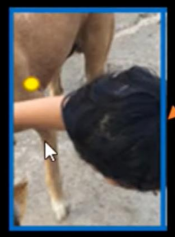

English

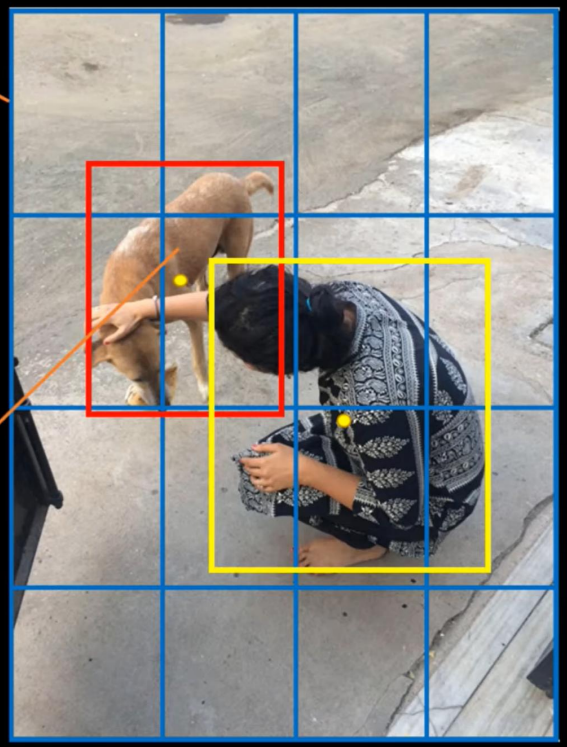


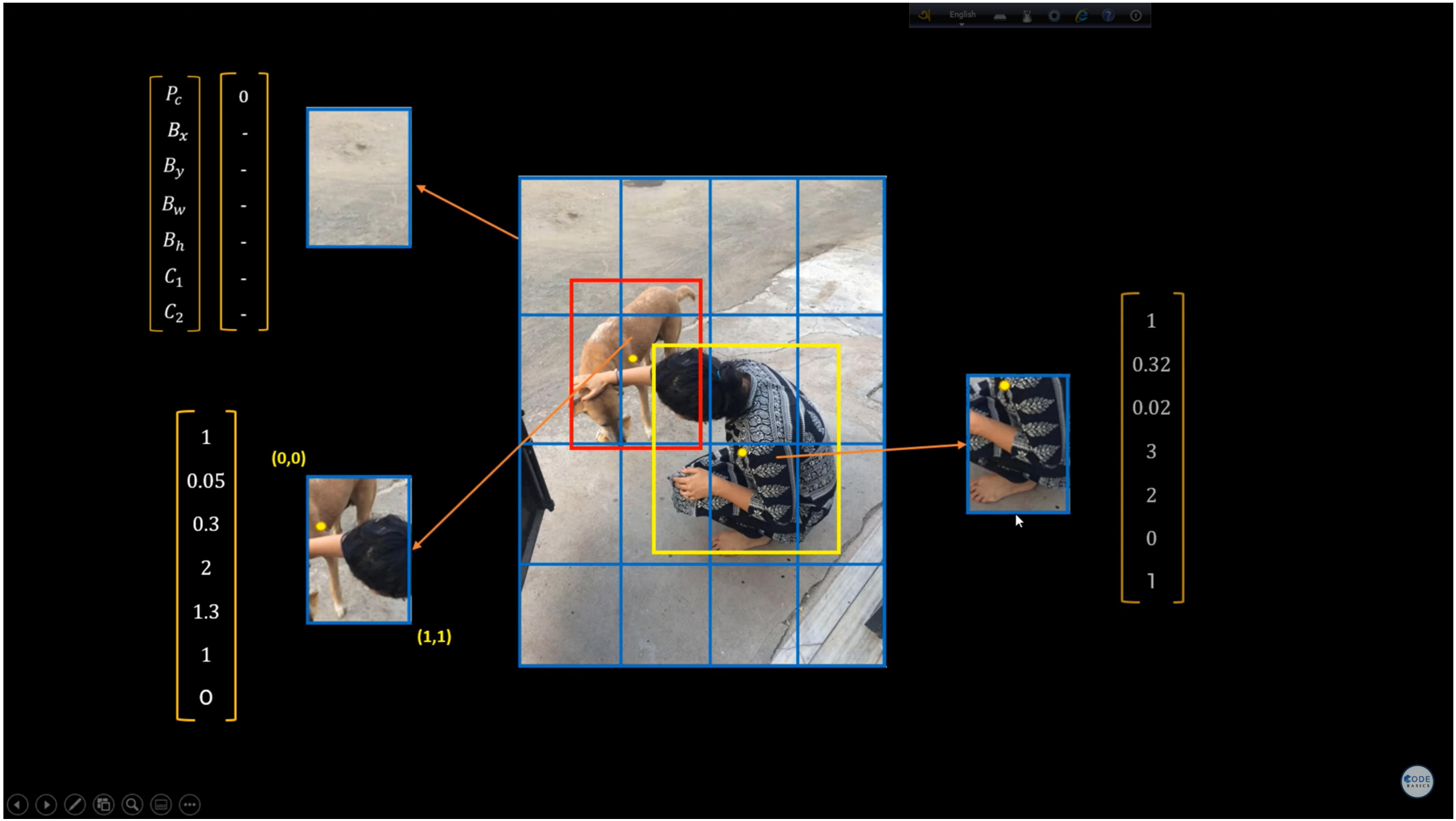
$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix} \quad \begin{bmatrix} 0 \\ - \\ - \\ - \\ - \\ - \\ - \end{bmatrix}$$

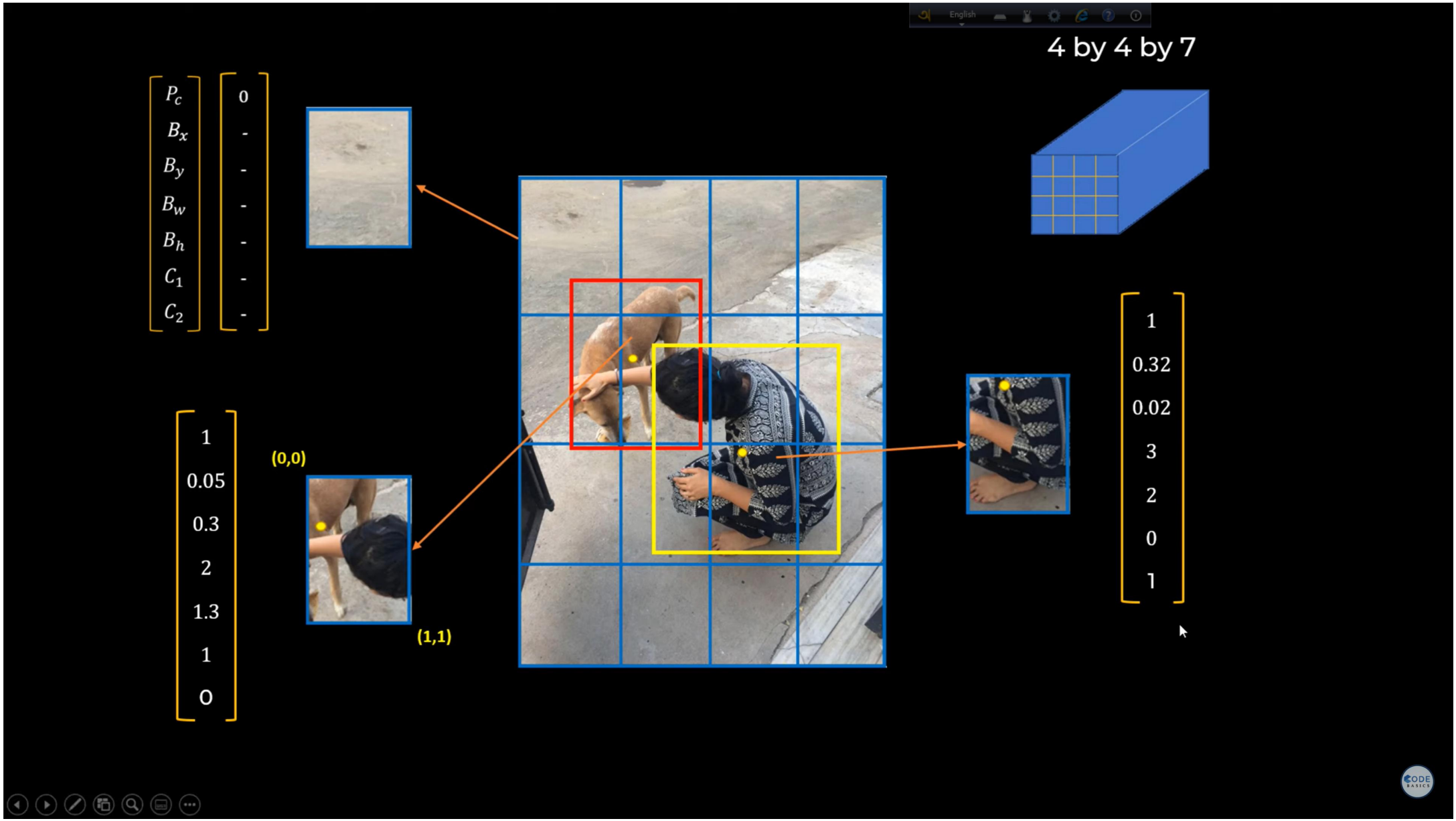


$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix} \quad \begin{bmatrix} 0 \\ - \\ - \\ - \\ - \\ - \\ - \end{bmatrix}$$



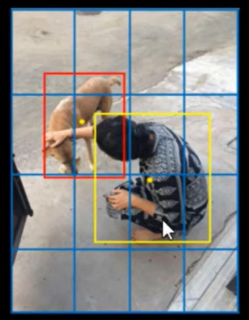



P_c
B_x
B_y
B_w
B_h
C_1
C_2
0
-
-
-
-
-
-
1
0.05
0.3
2
1.3
1
0
(0,0)
(1,1)
1
0.32
0.02
3
2
0
1


4 by 4 by 7
P_c
B_x
B_y
B_w
B_h
C_1
C_2
0
-
-
-
-
-
-
1
0.05
0.3
2
1.3
1
0
(0,0)
(1,1)
1
0.32
0.02
3
2
0
1

Training

X_train



y_train

16 such vectors

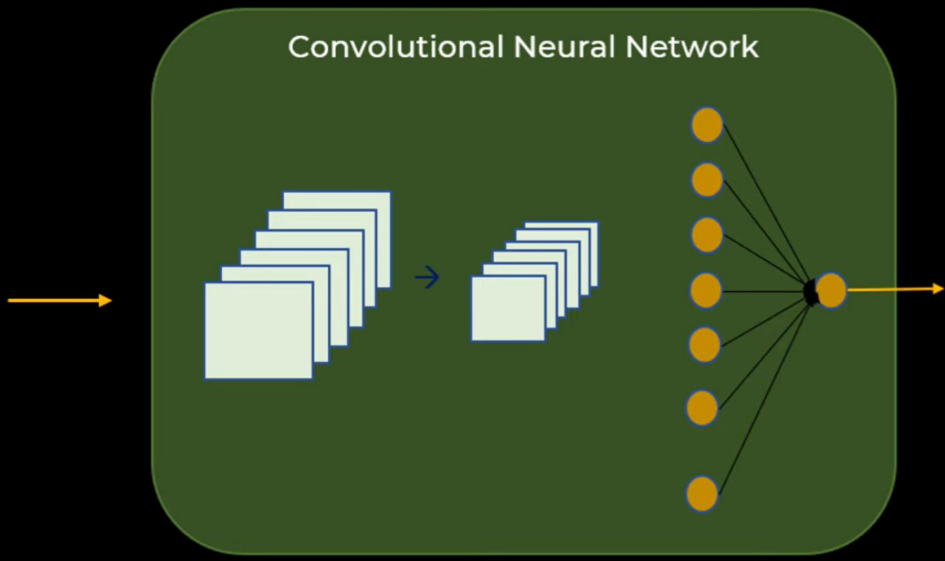
$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix}$$

16 such vectors

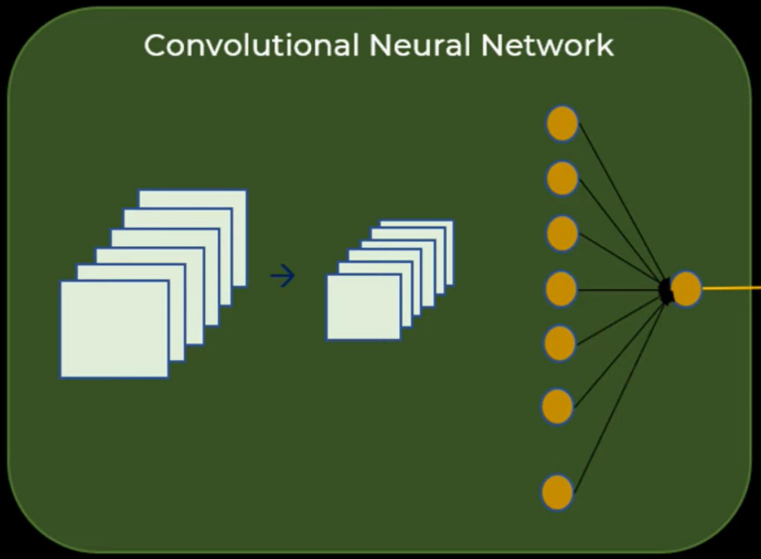
$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix}$$

16 such vectors

$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix}$$



Prediction



16 such vectors

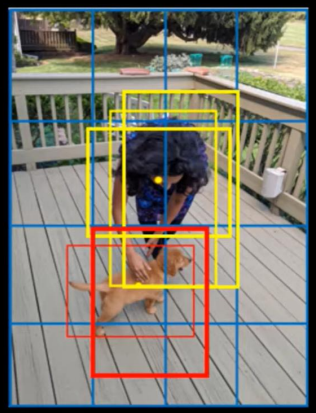
$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix}$$

Prediction

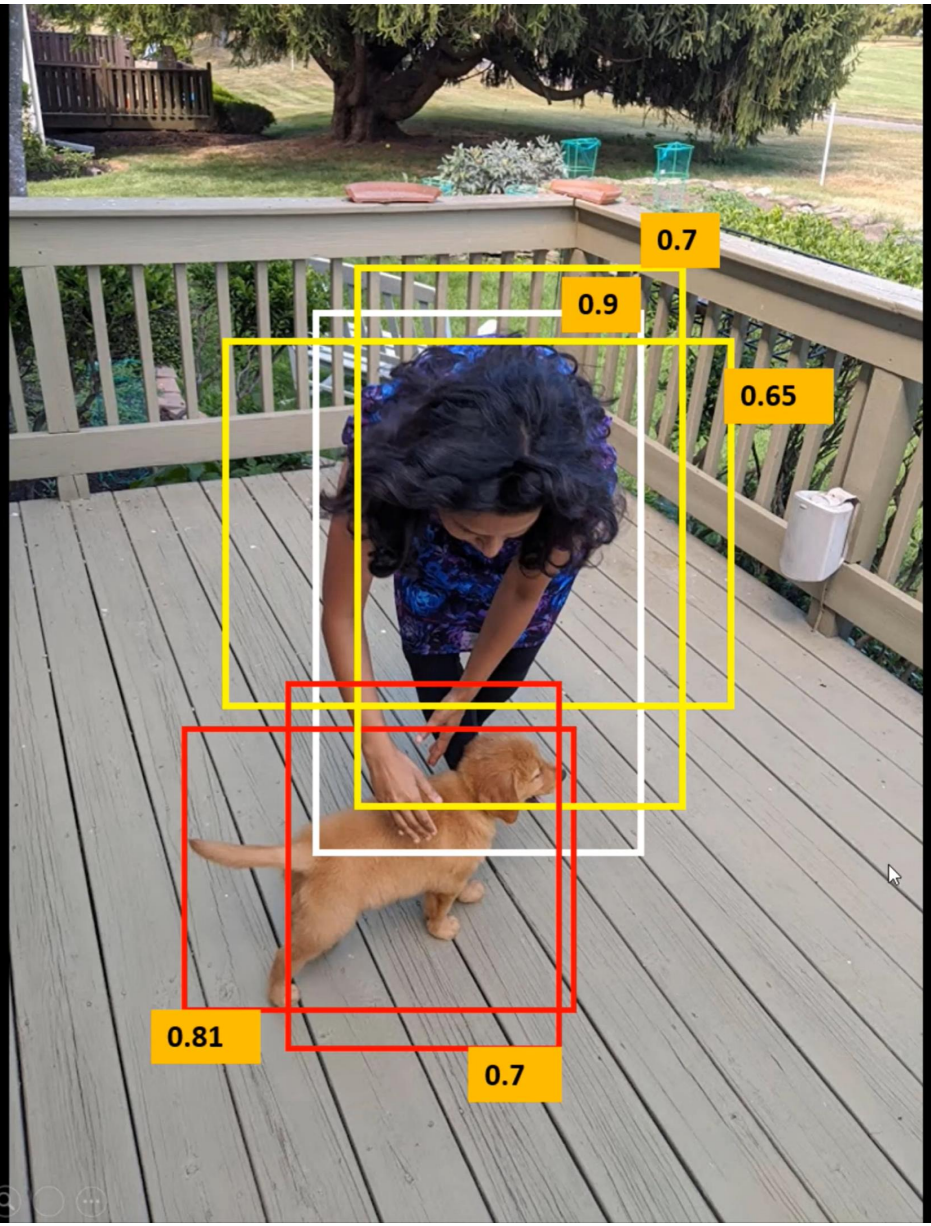
Convolutional Neural Network

16 such vectors

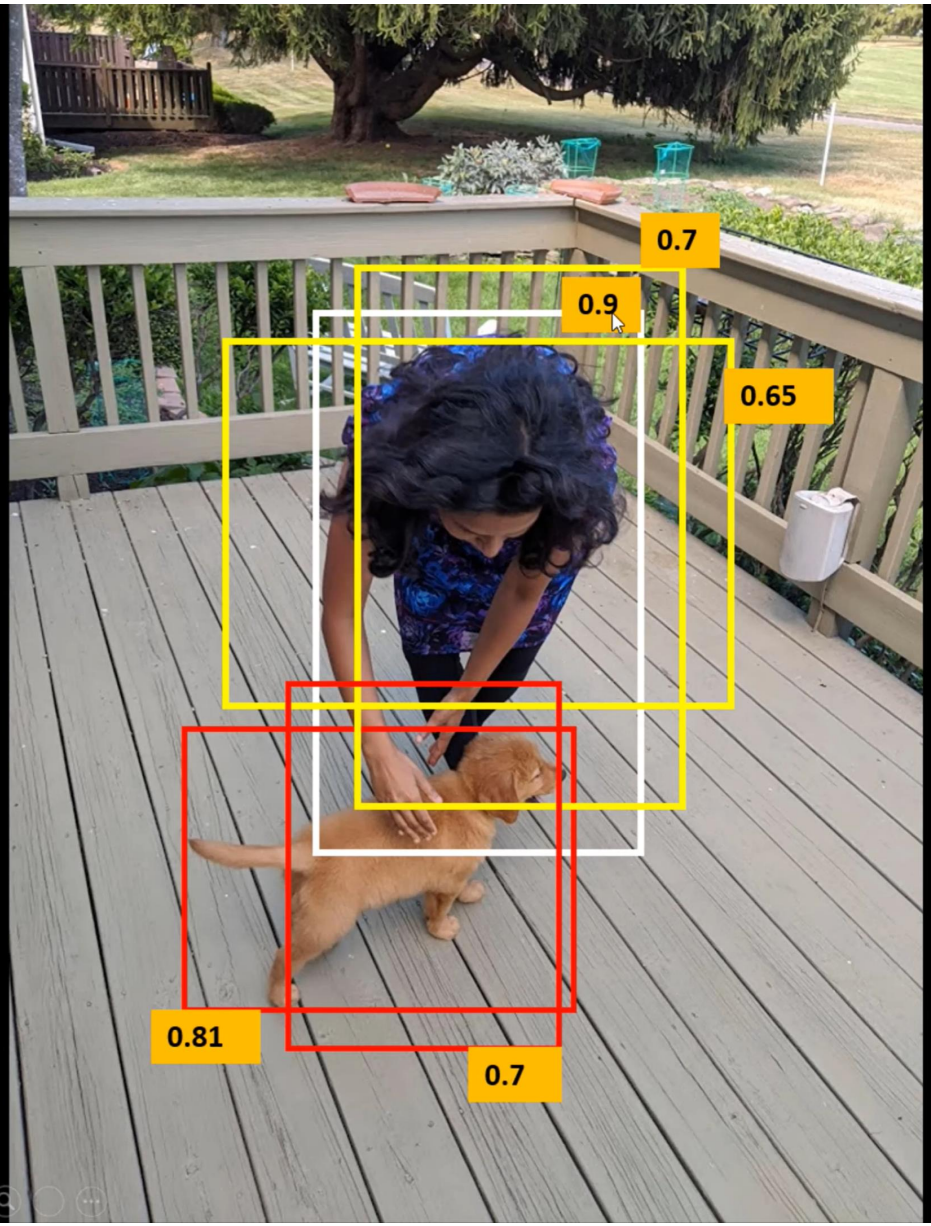
$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix}$$

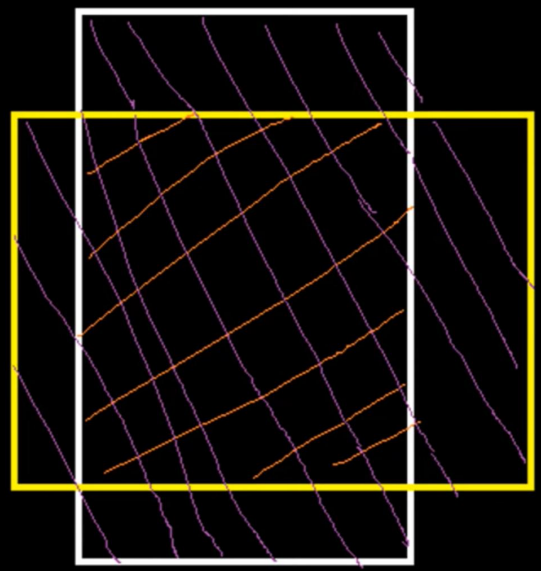


Multiple bounding boxes



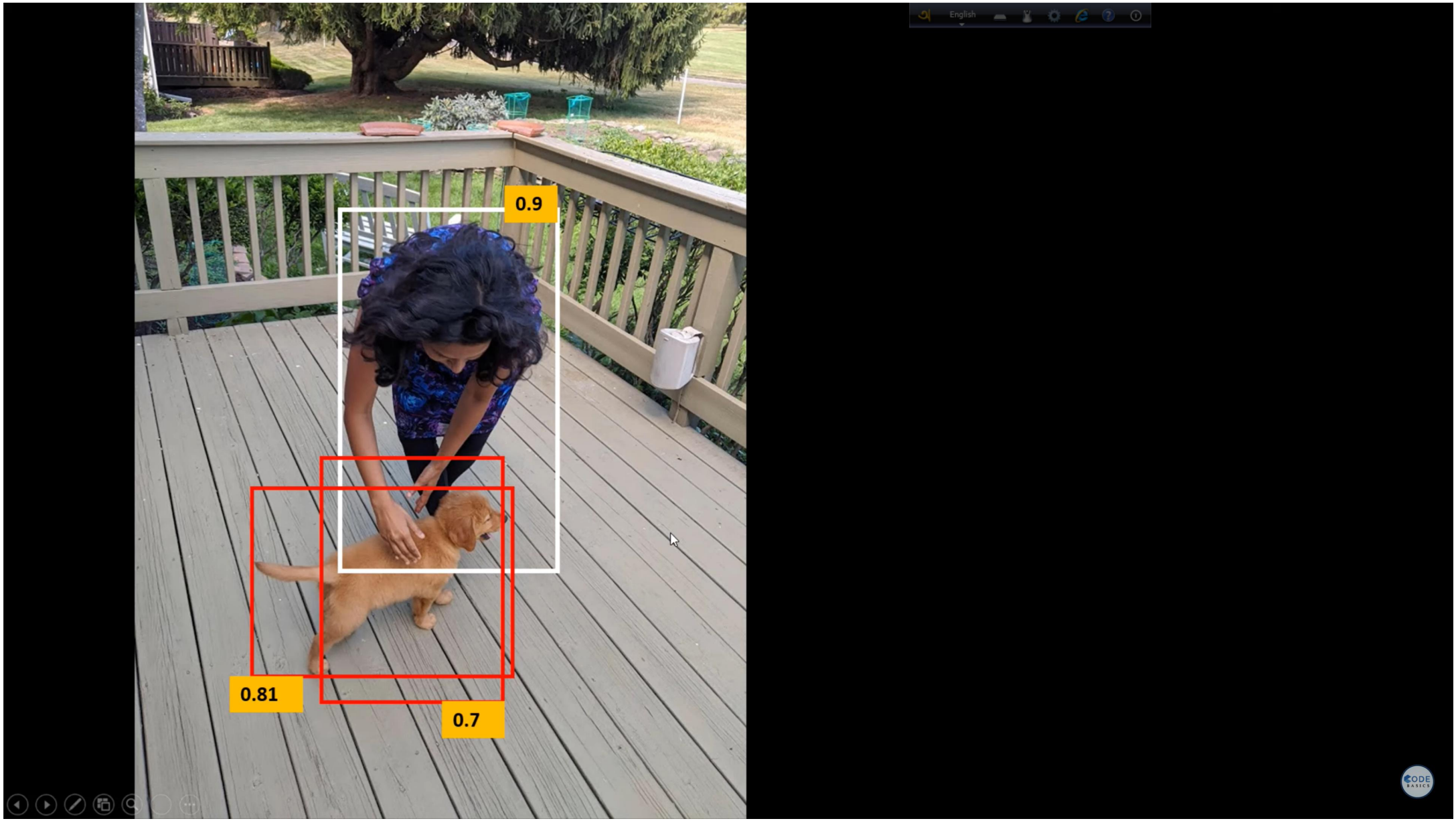
Can we just take max for each class?

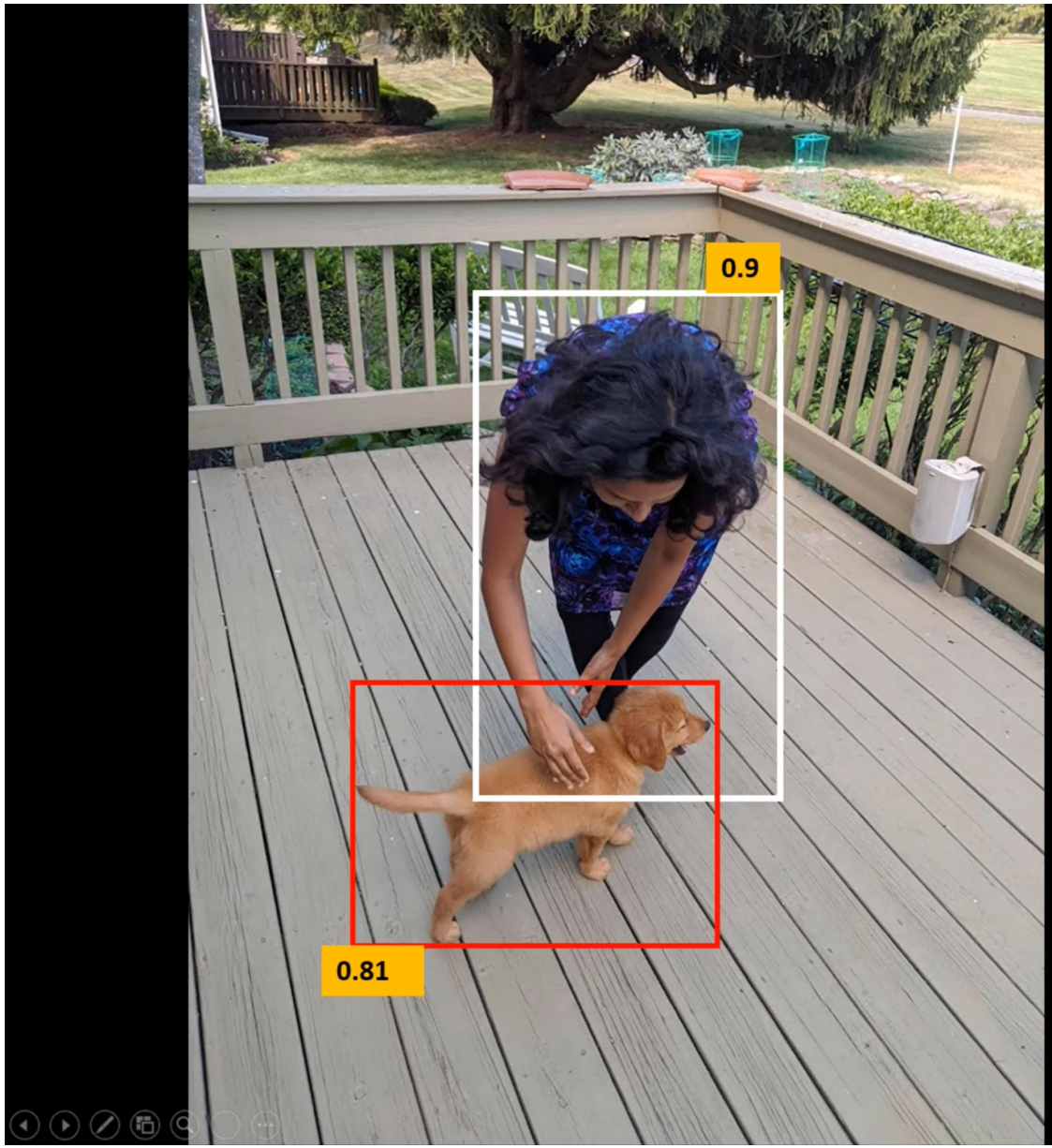




Intersection over union = intersect area / union area

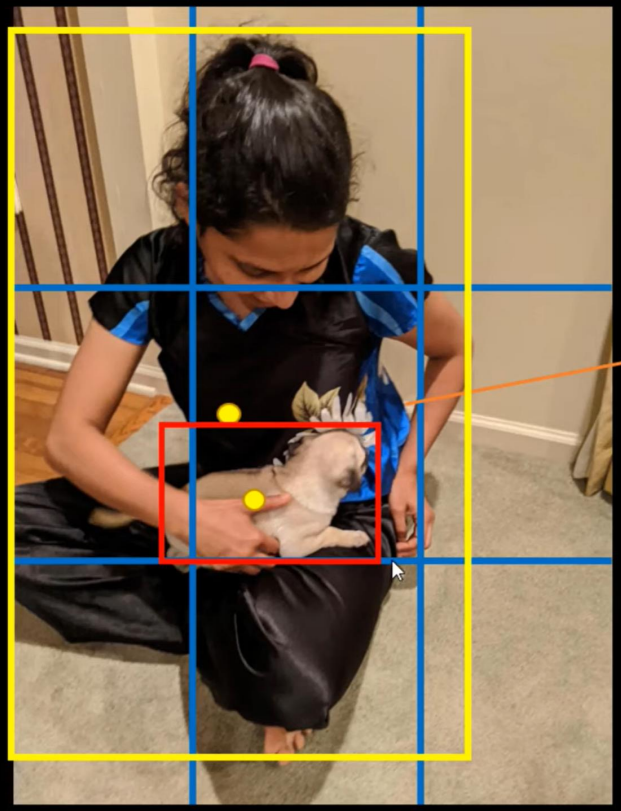
Intersection over union : IOU

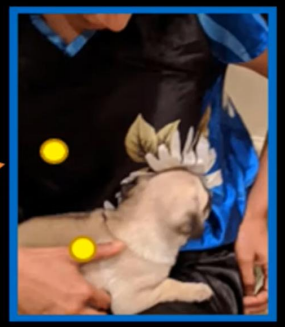

0.9
0.81
0.7



Non max suppression

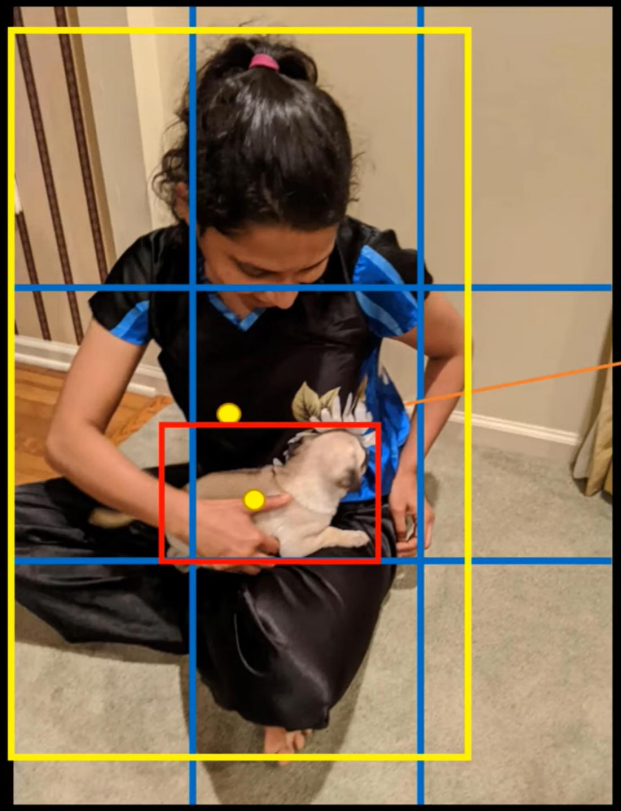
What if one grid cell has center of two objects?

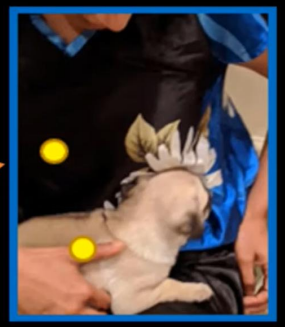




$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix}$$

What if one grid cell has center of two objects?



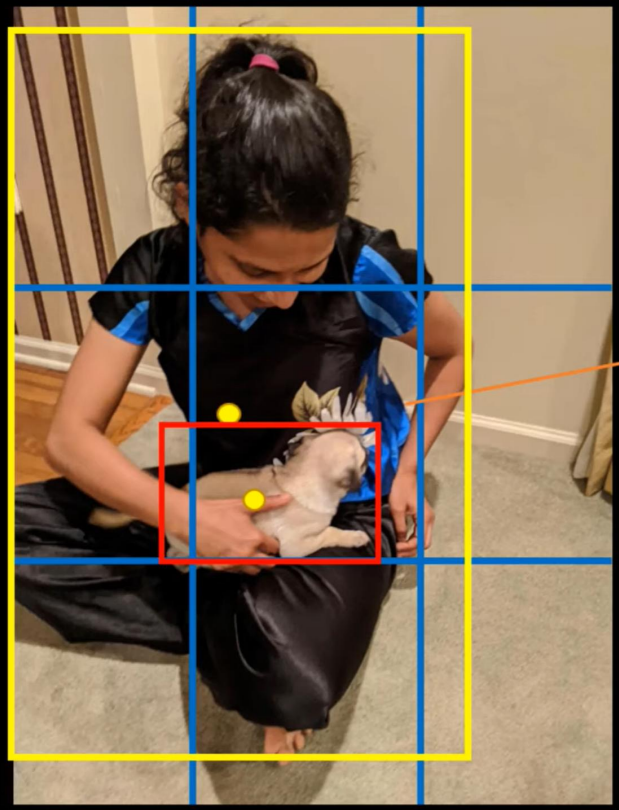


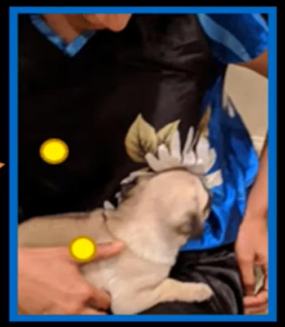
$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 0.22 \\ 0.45 \\ 1 \\ 0.7 \\ 1 \\ 0 \end{bmatrix}$$ Dog

$$\begin{bmatrix} 1 \\ 0.13 \\ 0.38 \\ 2.1 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$ Person

What if one grid cell has center of two objects?





$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 0.22 \\ 0.45 \\ 1 \\ 0.7 \\ 1 \\ 0 \end{bmatrix}$$

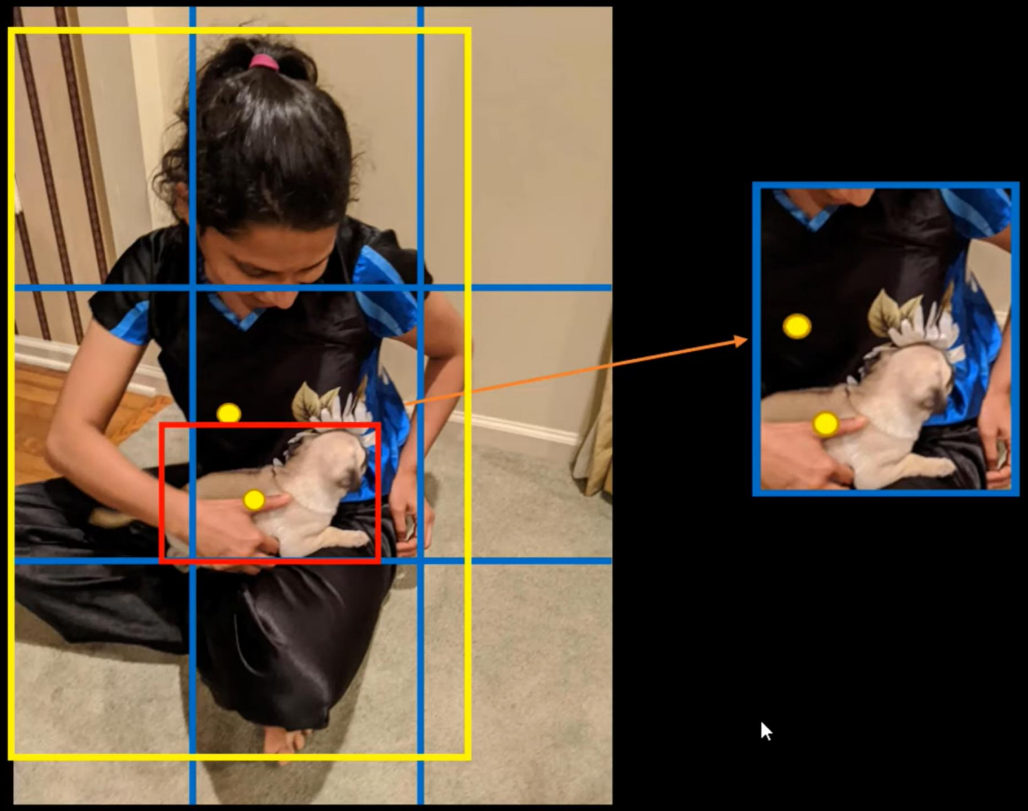
Dog

$$\begin{bmatrix} 1 \\ 0.13 \\ 0.38 \\ 2.1 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

Person

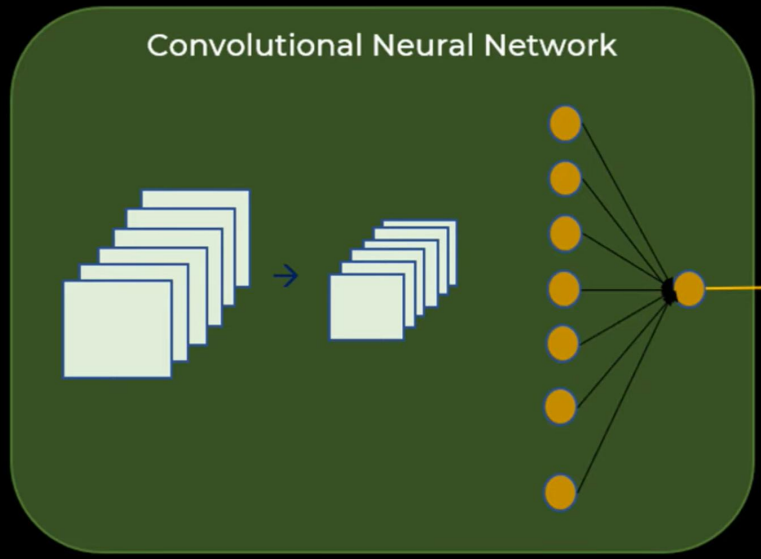
$$\begin{bmatrix} 1 \\ 0.22 \\ 0.45 \\ 1 \\ 0.7 \\ 1 \\ 0 \\ 1 \\ 0.13 \\ 0.38 \\ 2.1 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

This concept is called anchor boxes



$$\begin{bmatrix} 1 \\ 0.22 \\ 0.45 \\ 1 \\ 0.7 \\ 1 \\ 0 \\ 1 \\ 0.13 \\ 0.38 \\ 2.1 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

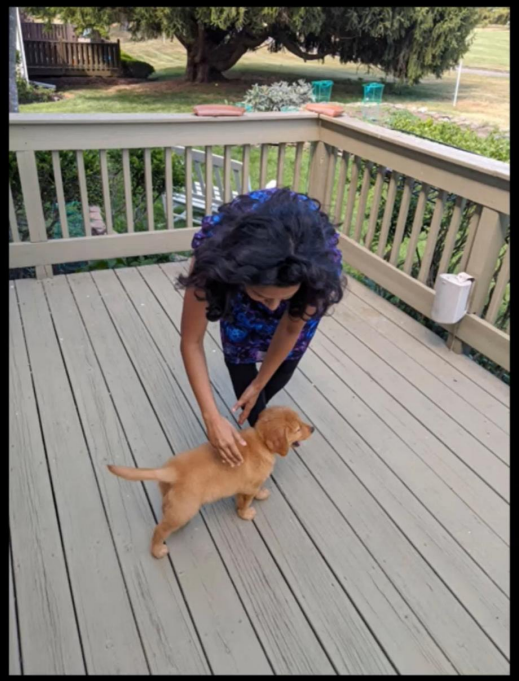
CNN with Two anchor boxes

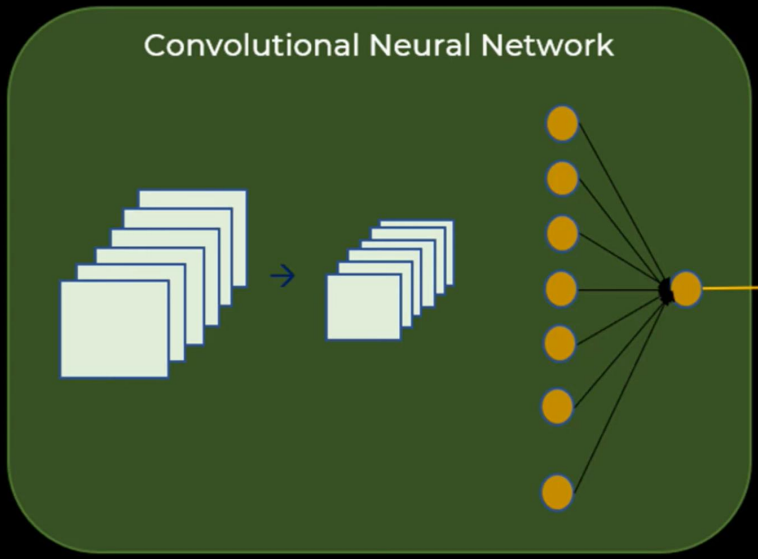


16 such vectors

P_{c1}
B_{x1}
B_{y1}
B_{w1}
B_{h1}
C_{11}
C_{21}
P_{c2}
B_{x2}
B_{y2}
B_{w2}
B_{h2}
C_{12}
C_{22}

CNN with Two anchor boxes





16 such vectors

$$\begin{bmatrix} P_{c1} \\ B_{x1} \\ B_{y1} \\ B_{w1} \\ B_{h1} \\ C_{11} \\ C_{21} \\ P_{c2} \\ B_{x2} \\ B_{y2} \\ B_{w2} \\ B_{h2} \\ C_{12} \\ C_{22} \end{bmatrix}$$