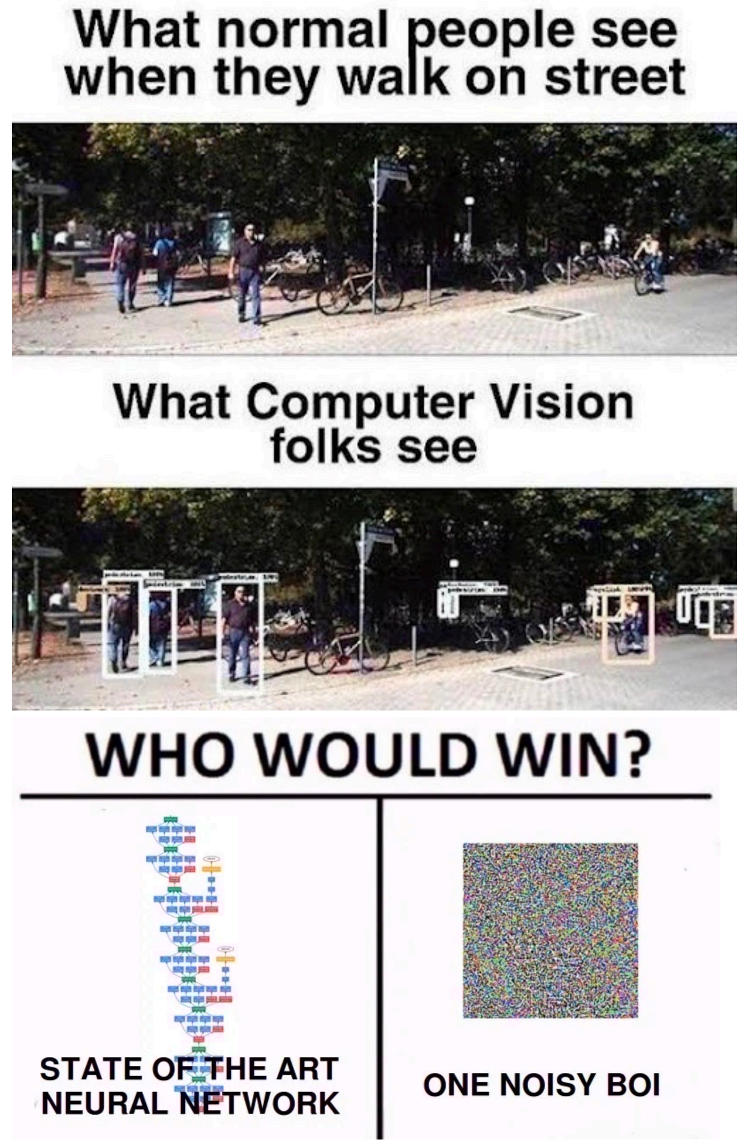
April 10, 2023

DSML: Computer Vision.

Image Segmentation.

Class starts @ 9:05 pm.



Recap: Object Detection.

* R-CNN family:
  * R-CNN.
  * Fast R-CNN
  * Faster R-CNN. – comes close to real time.
    ≈ 5 frames per second.

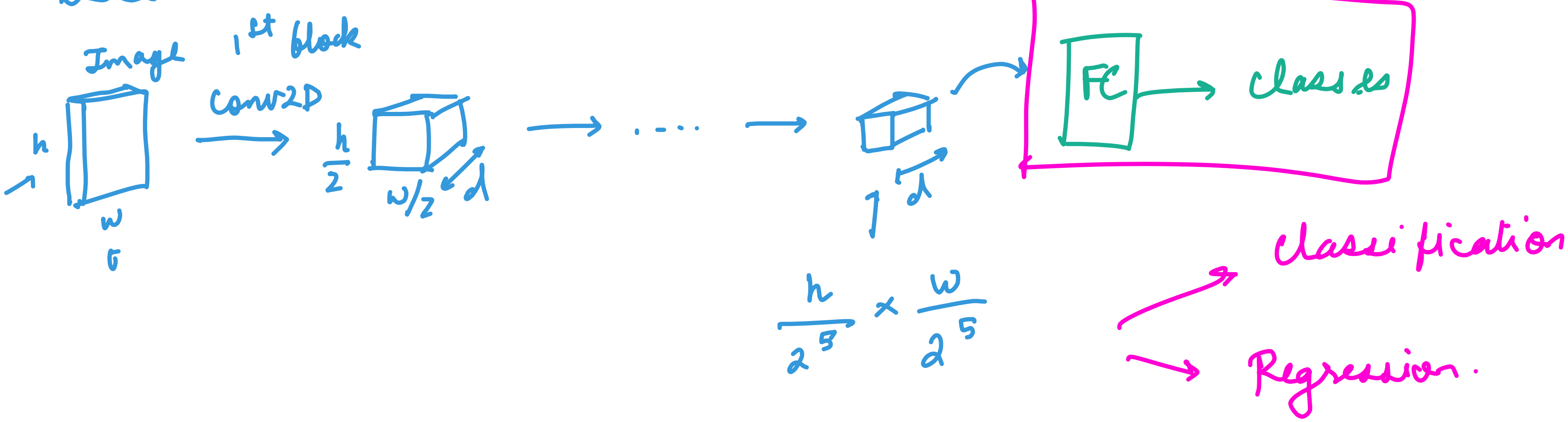
* YOLO family:
  → Faster object detection, good enough for real time.

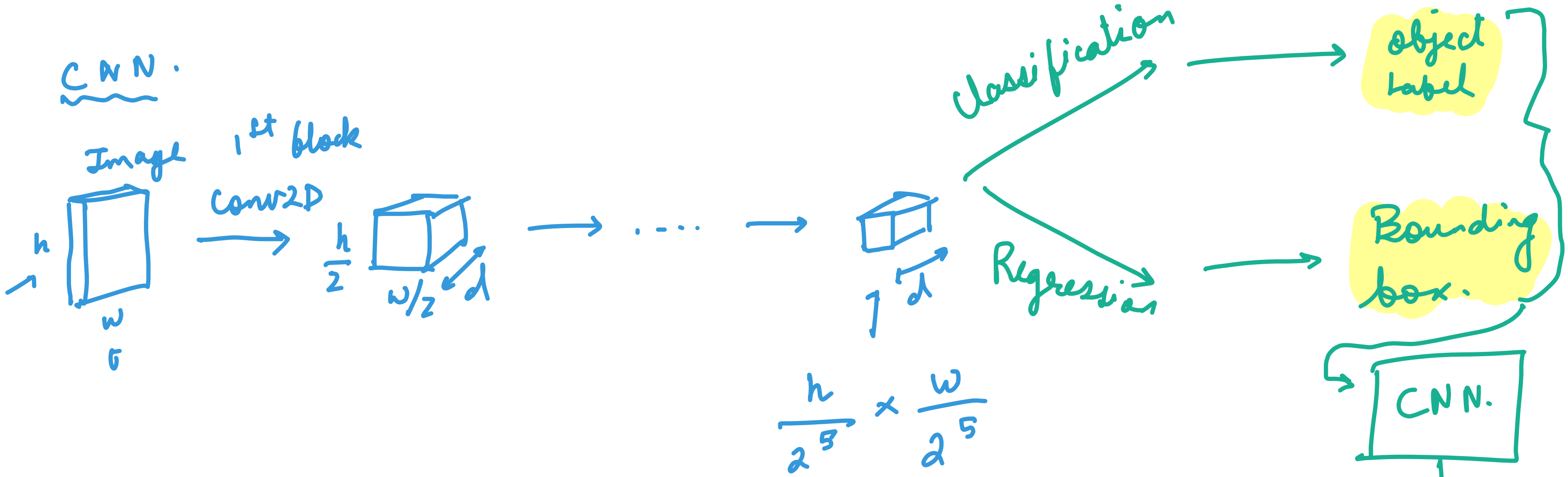
Image Segmentation:

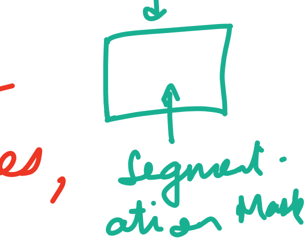
* Most popular – Based on the Fast R-CNN architecture, called Mask-RCNN.

Image Segmentation :

CNN.







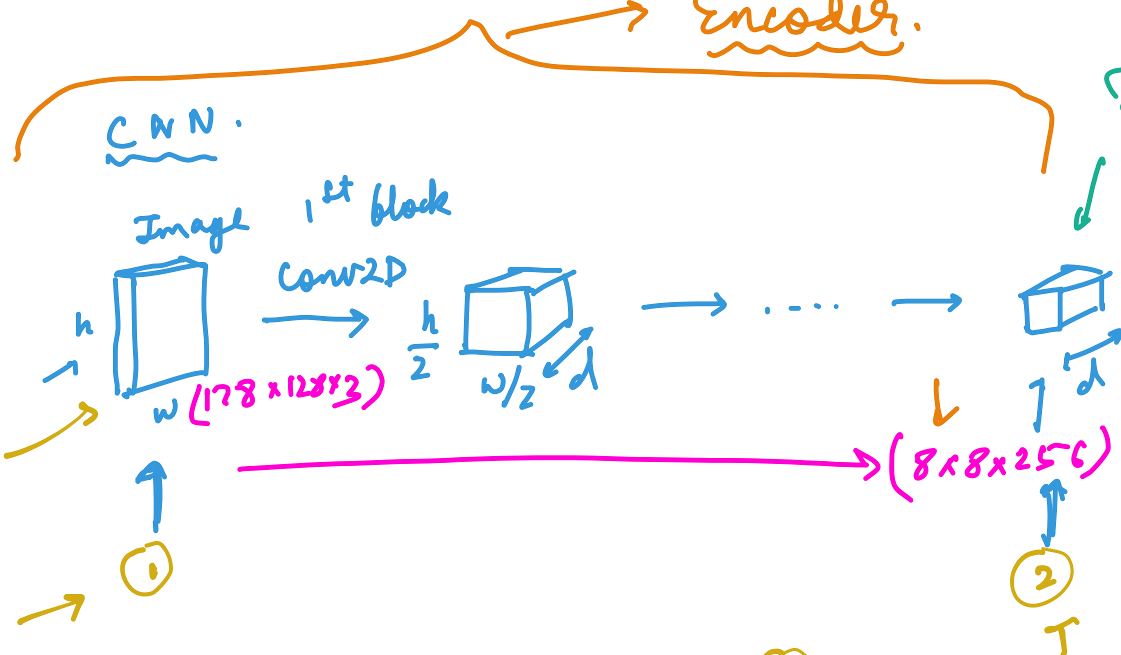
Possible strategies

① Instead of patches, classify pixels

↳ We have to change the way we get the output from the network,

② Boundary detection inside a bounding box.

Image Segmentation from Scratch :



This is a better representation for classification & for detection.

Similar Images : 2.

Image Classification : 2

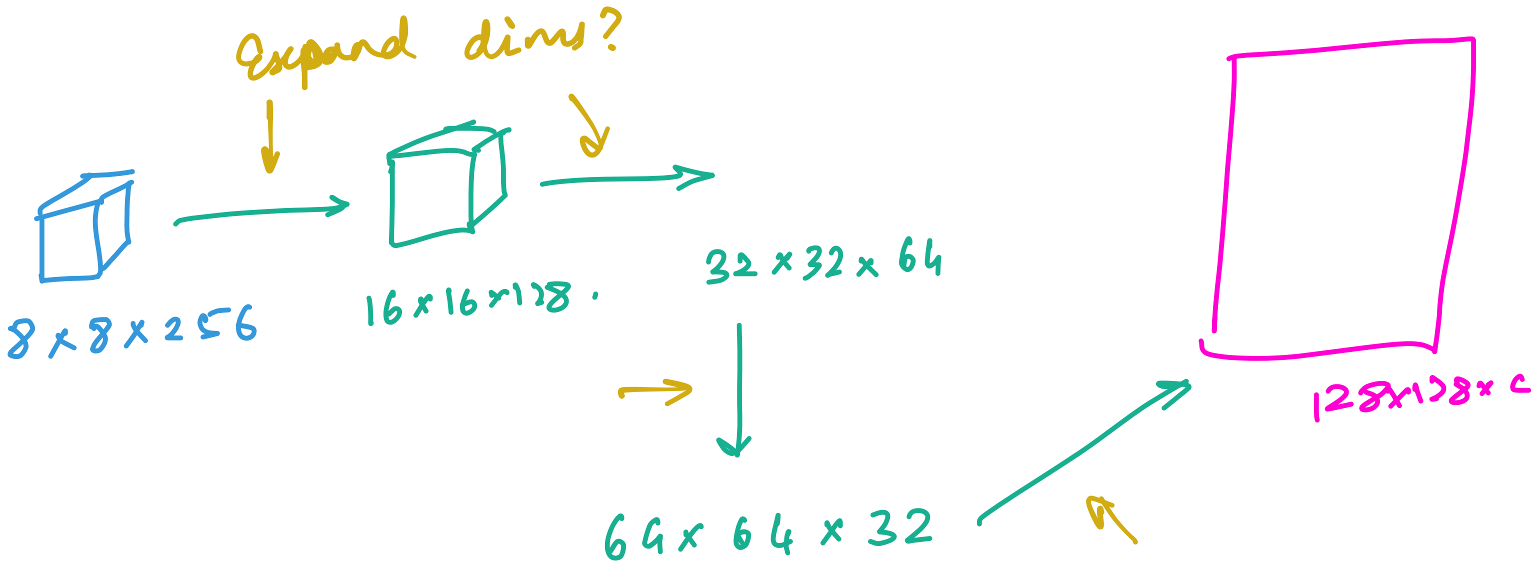
Image Segmentation?

NO. we cannot use this directly:

Dimension
(a) Loss of clarity.
(b) Loss of spatial info.

Output.

128x128xc
↳ no. of classes.

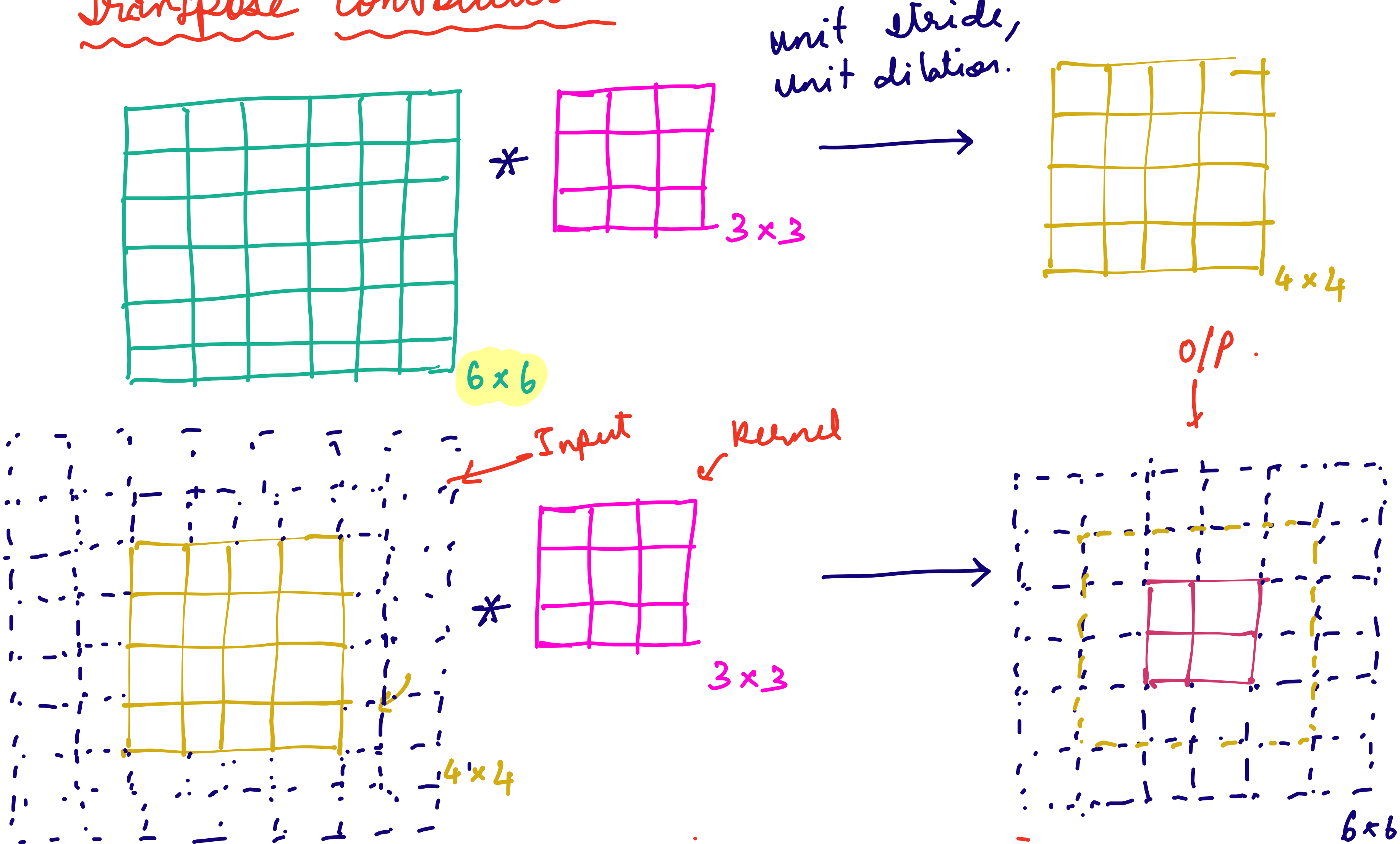


Suggestions:

1. Opposite of max pool.
2. Transpose convolution.
3. Upsampling.

1. FCN - Fully convolutional network.
2. UNet → Very popular in the Medical Imaging community.

Transpose convolution.



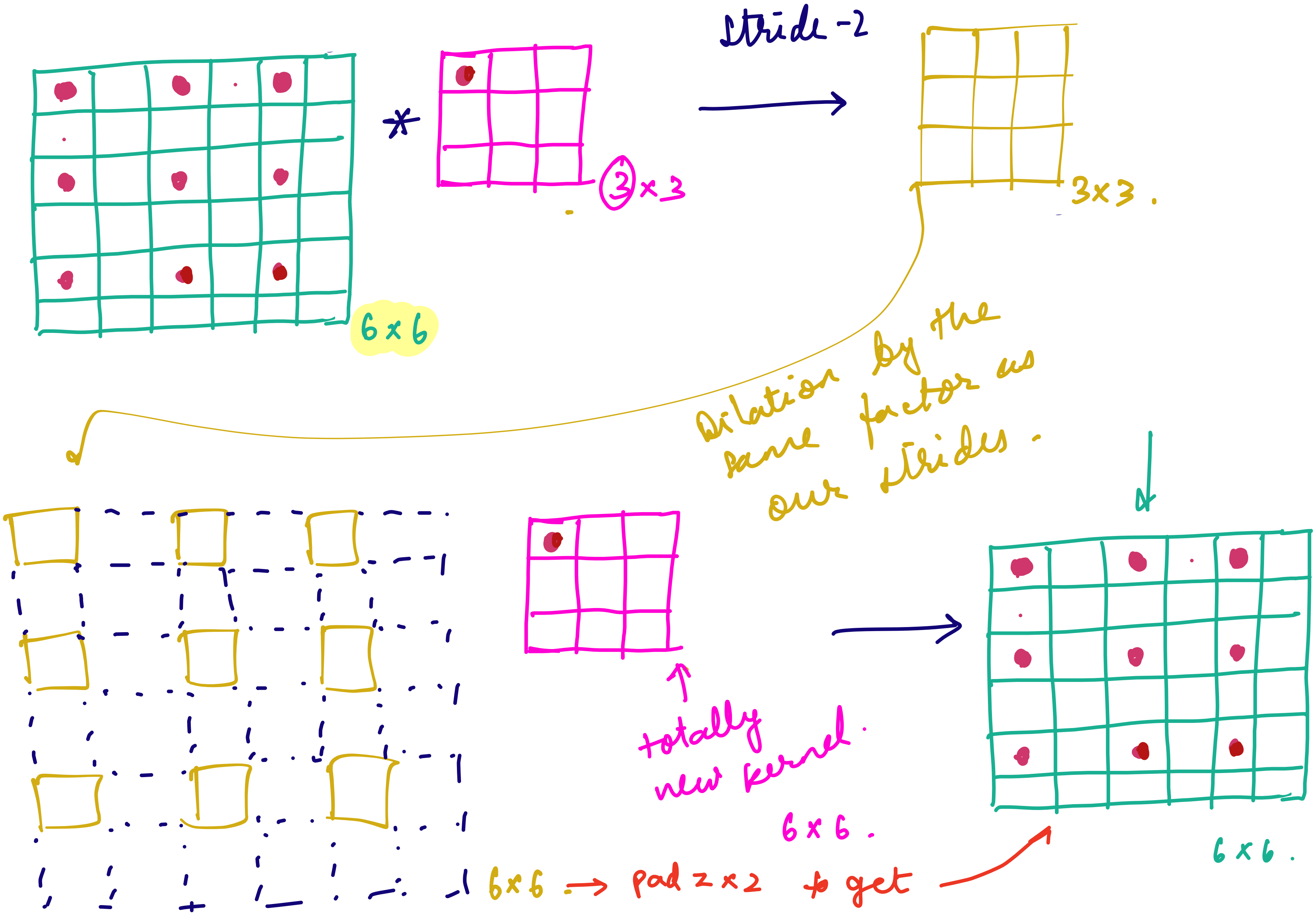
Using the normal convolution operation, how to increase the output size? → Pad & convolve!!

Transpose convolution:

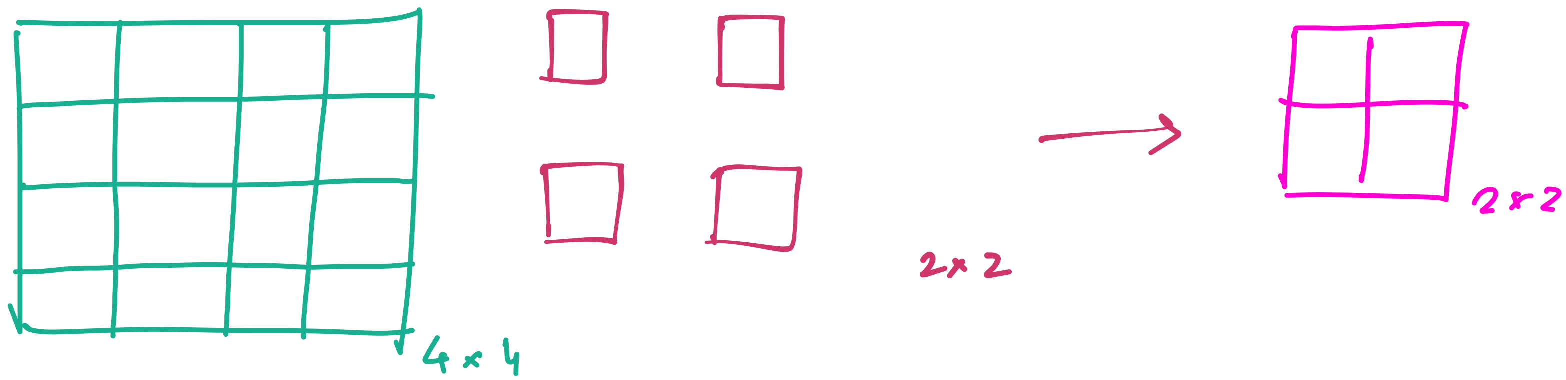
→ Normal convolution.

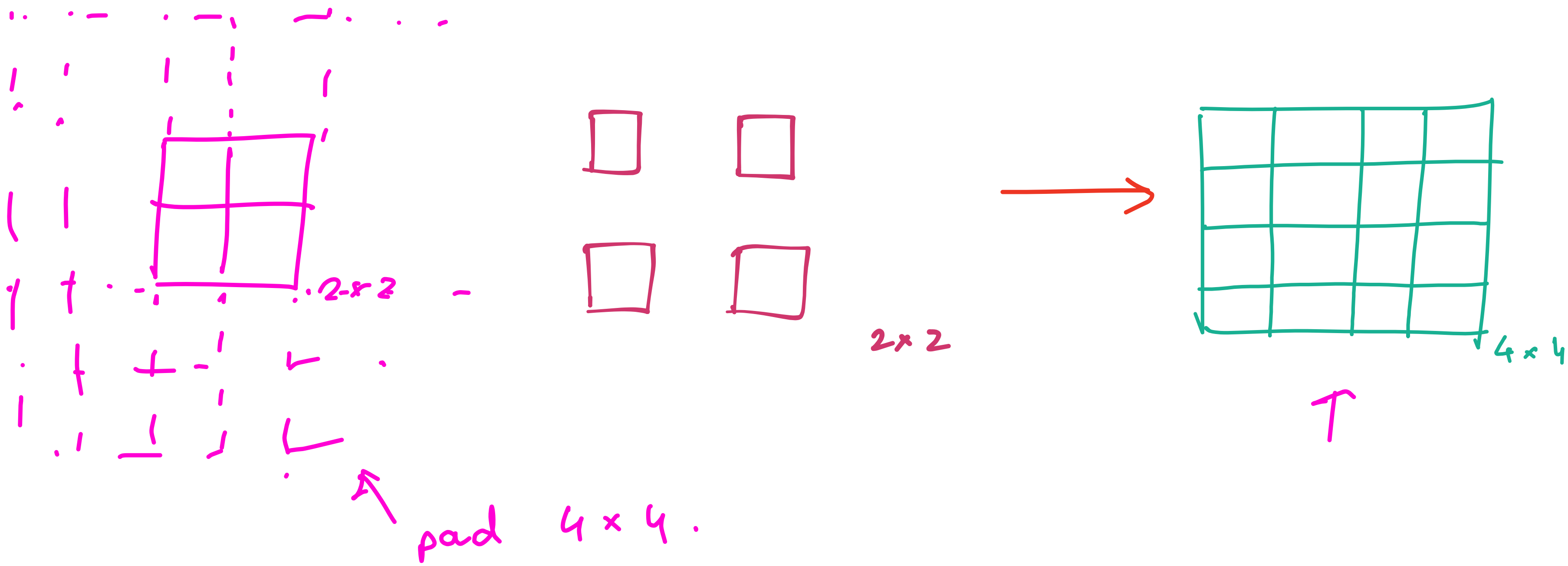
→ It will take the kernel size and the output shape, and automatically pad the input so that after convolution, the output size is obtained.

Transpose of Strided Convolutions:

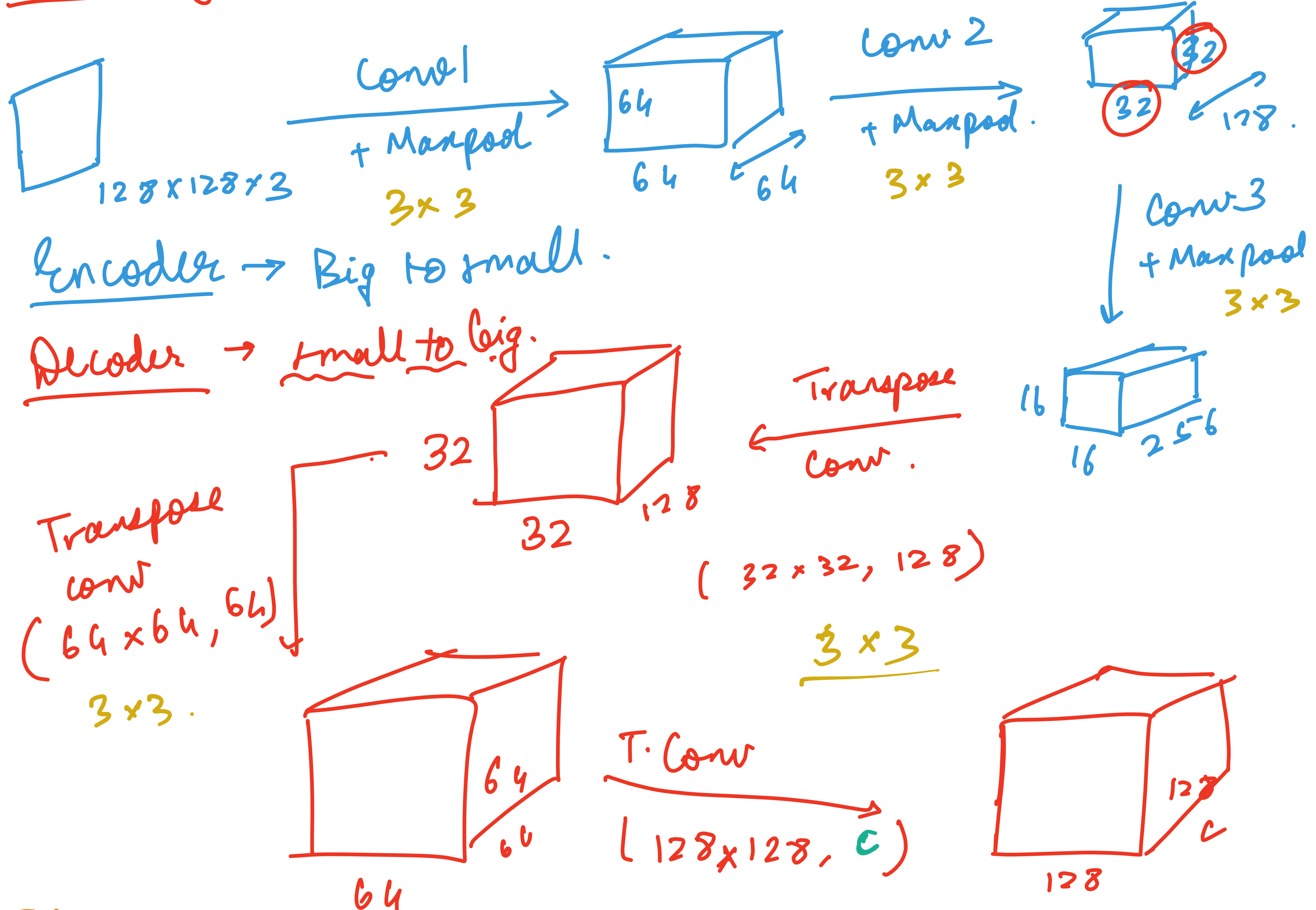


Transpose of dilated .





Applying Tranpose conv to mirror a VGG-like

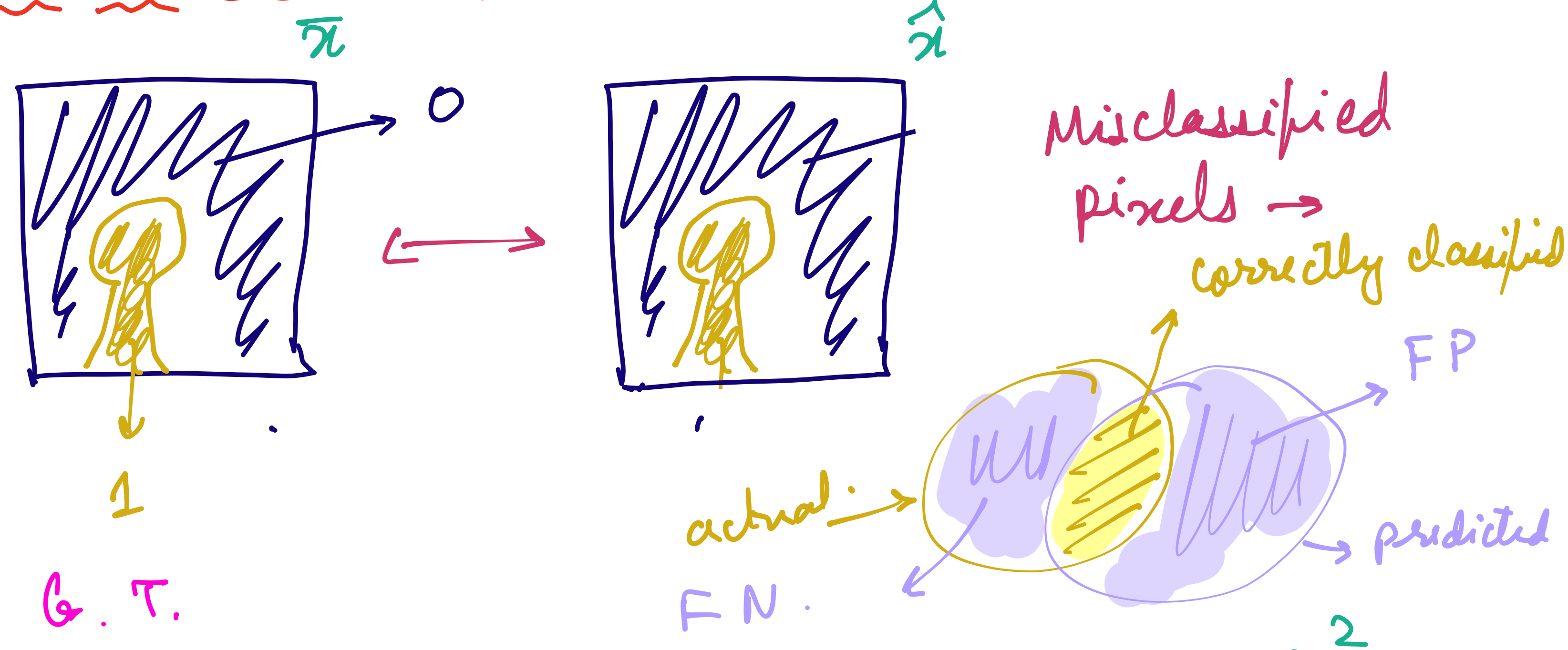


Encoder → Big to small.

Decoder → small to big.

FCN → Fully convolutional Neural network.

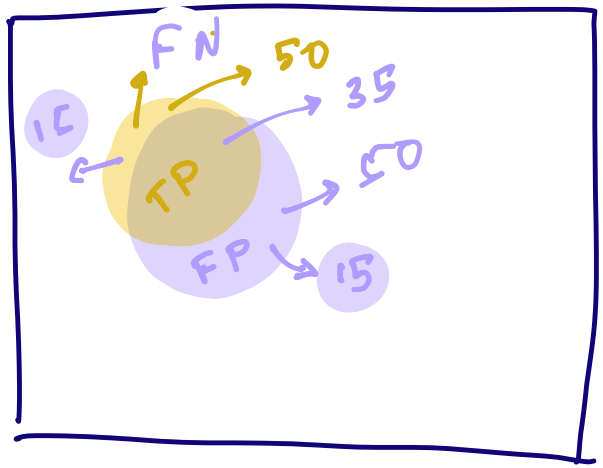
How to evaluate our model's performance?



1] Squared difference. $\sum_{i,j} (x_{ij} - \hat{x}_{ij})^2$

2] IoU → Intersection over Union.

3] DICE coefficient. ← $\left[\dfrac{2TP}{2TP + FP + FN} \right]$ We can use this to check how well our model is doing.



$$35/50 \simeq 70\%$$

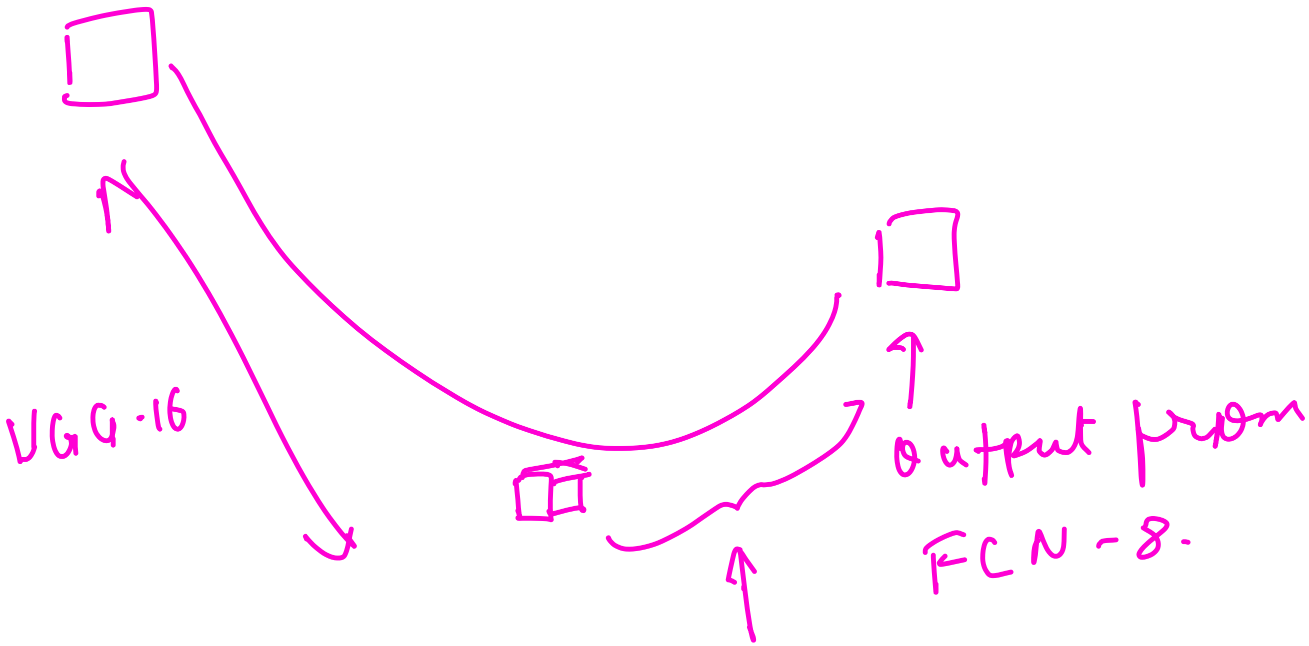
$$\text{I.O.U.} = \frac{TP}{TP + FP + FN}$$

$$= \frac{35}{35 + 30}$$

$$= 0.33$$

$$\text{DICE} = \frac{2 \cdot TP}{2TP + FP + FN} = \frac{70}{70 + 30}$$

$$= \frac{70}{100} \approx 70\%$$



VGG-16
output from
FCN-8.

1

1 only if $i = 24$, $j = 1$.

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

person

0

error in prediction?

error in prediction!

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\underbrace{\sqrt{w_i} - \sqrt{\hat{w}_i}}_{\text{width}} \right)^2 + \left(\underbrace{\sqrt{h_i} - \sqrt{\hat{h}_i}}_{\text{height}} \right)^2 \right]$$

checks if BBox center is in the correct cell.

if the dimensions are correct.

$$+ \underbrace{\sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2}_{\text{confidence term.}}$$

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \quad (3)$$

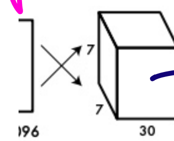
20

no. of classes

$+ \ (5 \times 2)$

num outputs

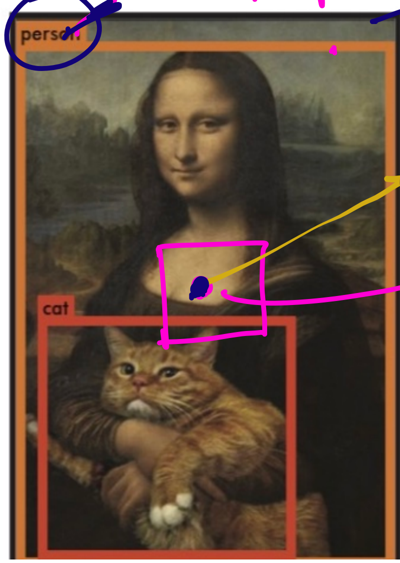
num boxes.



Conn. Layer

$\underbrace{7 \times 7}_{\text{grid dimensions.}} \times \ 30$

BBox for person.



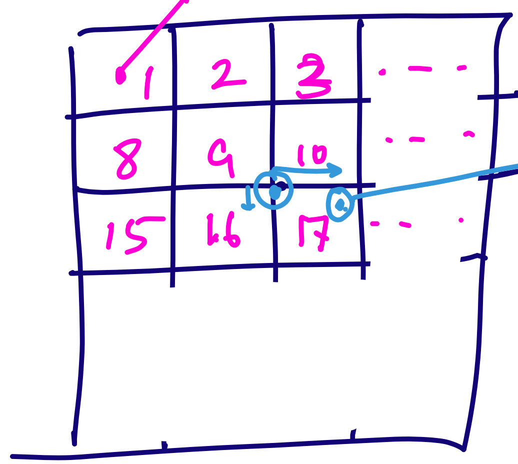
0

1

2

24

cell.



$i \to 1$ to 49.

$j = 1, 2$ (since we have 2 boxes per cell.)

box center x

box center y.

① $$\lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$ } check if center is correct

② $$+ \lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$ } check dimensions.

confidence of prediction.

③ $$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$ } → check confidence.

negative class weight

④ $$+ \boxed{\lambda_{\textbf{noobj}}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$ }

where objects not present, reduce confidence

0.5

$$\underbrace{+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c) \right)^2}_{}\qquad (3)$$

⑤ controls prediction accuracy of the object.

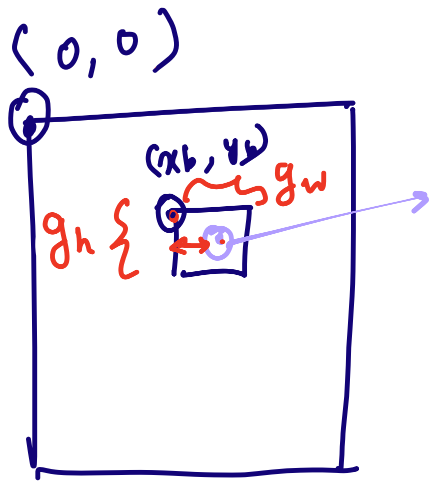
$$\left\{ \underbrace{\quad p_i(c),\ \hat{p}_i(c) \quad}_{20 \text{ classes.}} \quad \underbrace{}_{c_1} \quad \underbrace{}_{x_1, y_1, h_1, w_1} \quad \underbrace{}_{c_2} \quad \underbrace{}_{x_2, y_2, h_2, w_2.} \right\} \times 7 \times 7.$$

O.P.

c_1, c_2: Binary prediction

1 if obj,

0 if no object.



$(x_p, y_p) \longrightarrow (x_p', y_p')$

$$x_p' = \frac{(x_p - x_b)}{g_w}$$

$$y_p' = \frac{(y_p - y_b)}{g_h}$$