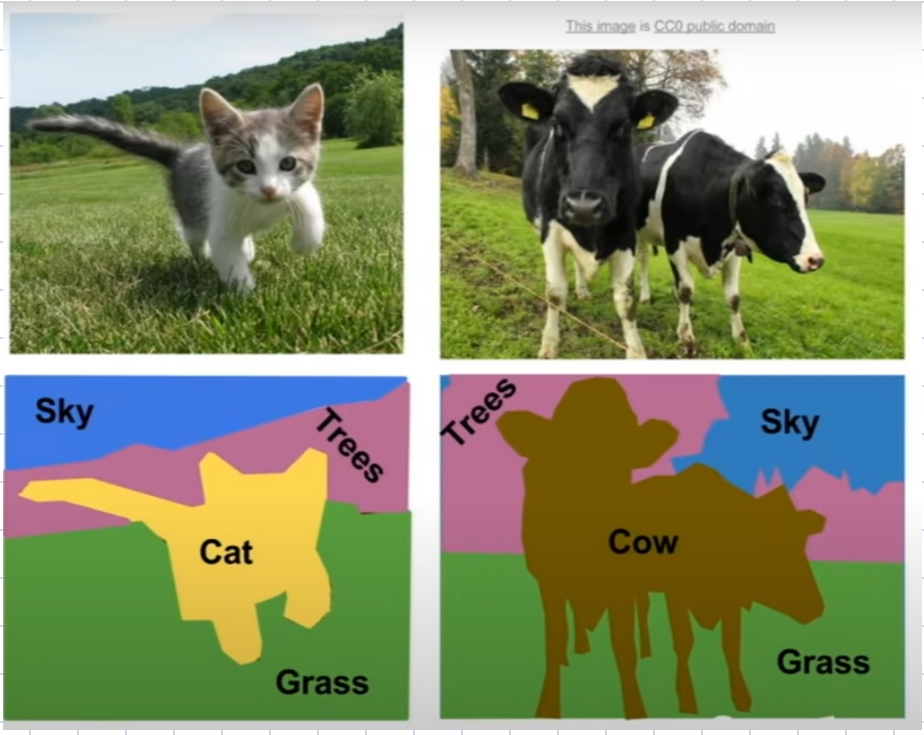
Semantic segmentation:

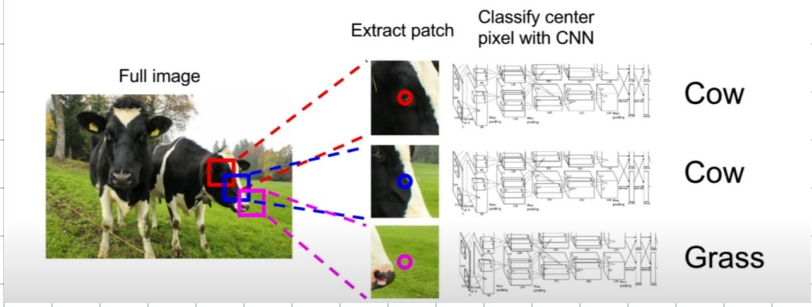
↳ Outputs a decision of a category for every pixel in that image.

↳ Doesnot differentiate between instances



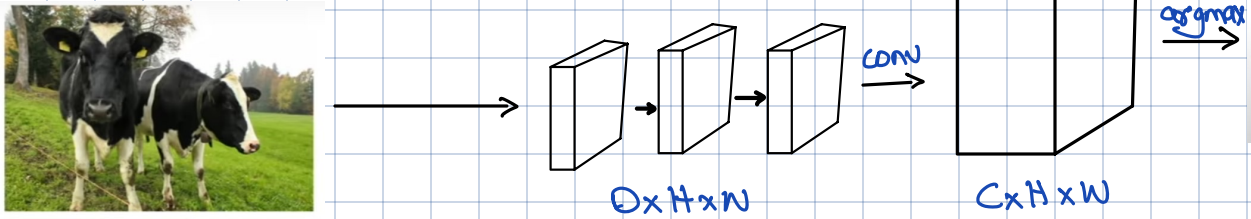
Sliding window:

↳ Extract patches from the image
↳ Apply CNN on the patch
↳ Classify it

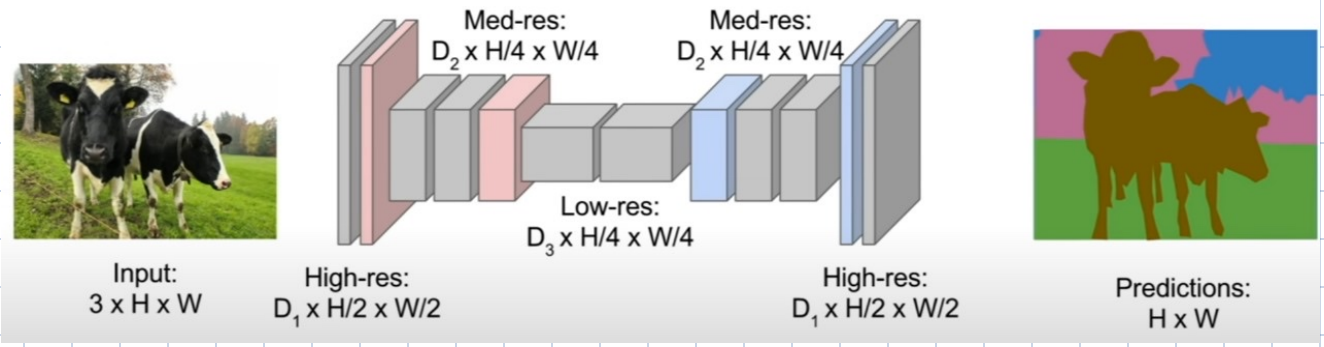


↳ Computationally expensive

Fully convolutional:



↳ A bunch of conv layers
↳ Final conv layer gives classification scores for every pixel
↳ Too computationally expensive
↳ You use upsampling & downsampling



↳ Conv → Down sampling → Upsample → Predictions

Unpooling:

↳ Nearest neighbour unpooling

1	2
3	4

input: 2x2

→

1	1	2	2
1	1	2	2
3	3	4	4
3	3	4	4

output: 4x4

↳ Bed of nails:

1	2
3	4

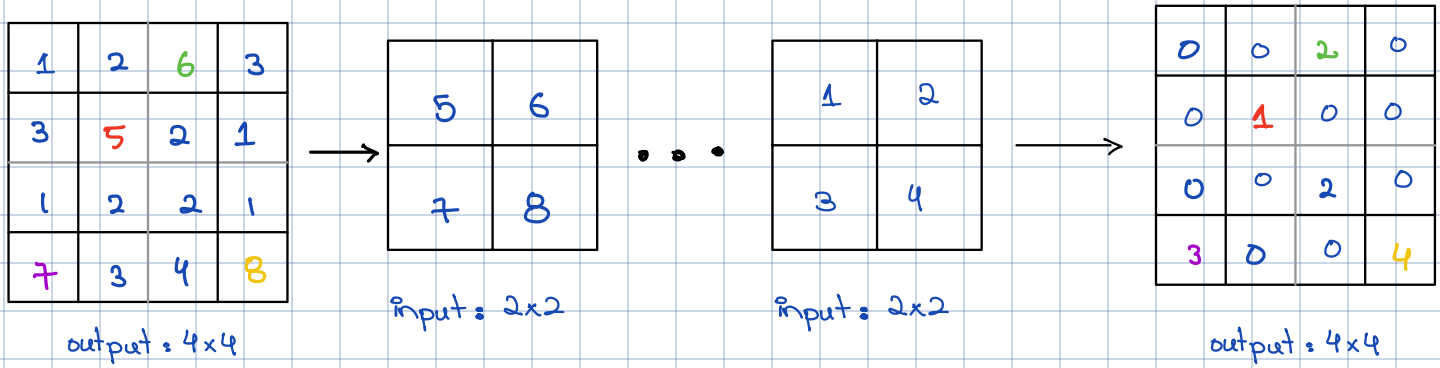
input: 2x2

→

1	0	2	0
0	0	0	0
3	0	4	0
0	0	0	0

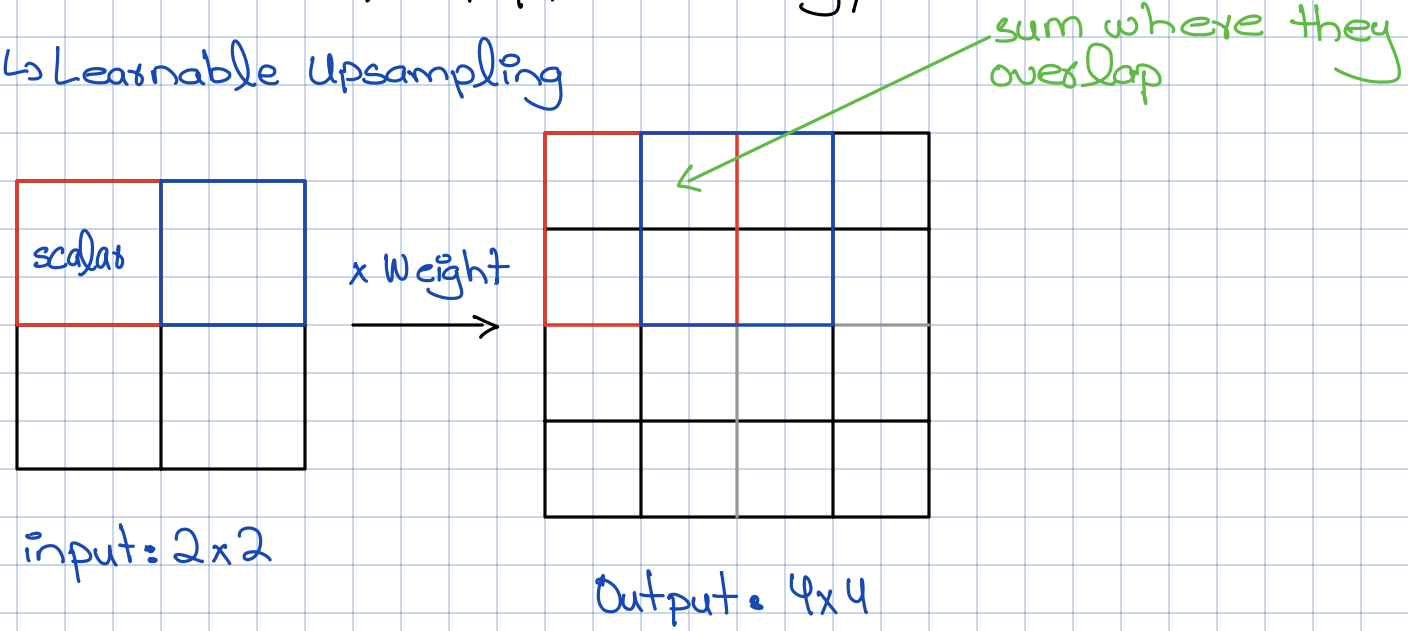
output: 4x4

Max unpooling:

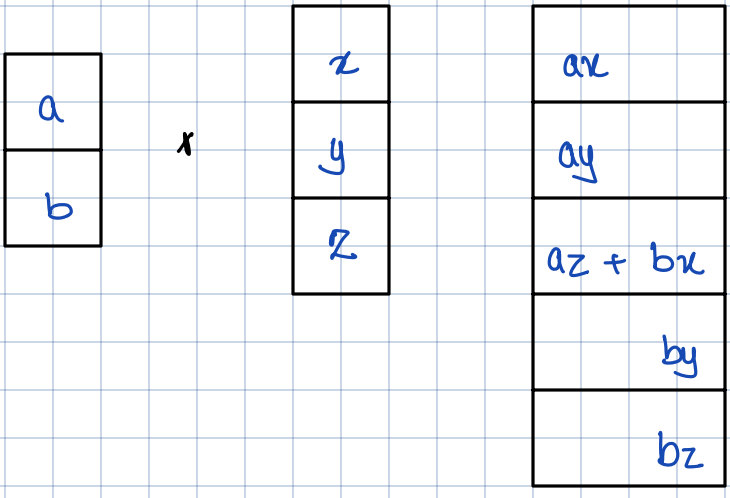


Transpool Convolution/De/up/fractionally/Backward strided conv

↳ Learnable Upsampling

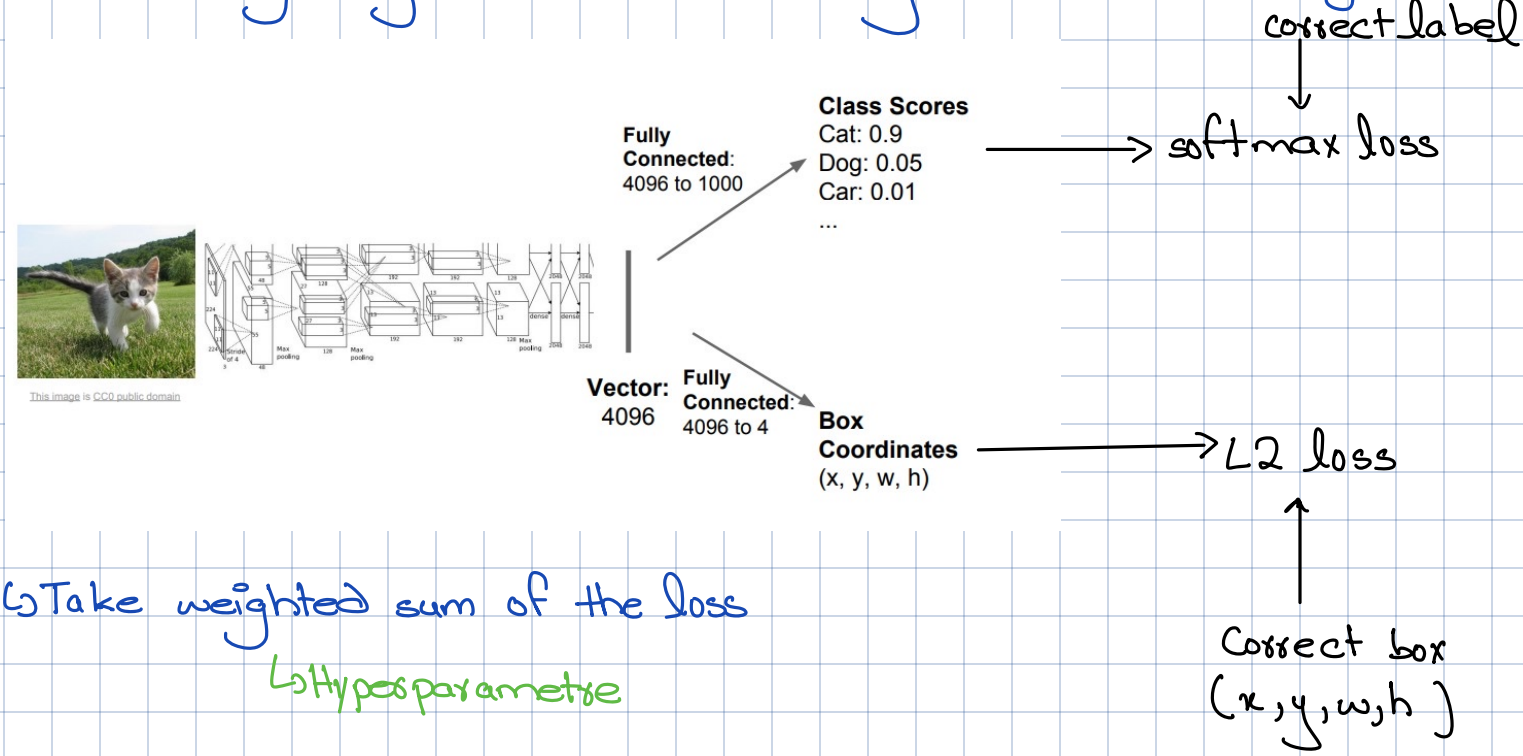


Eg: 1D:



Classification + localization:

↳ Classify image & the draw boundary box around that object

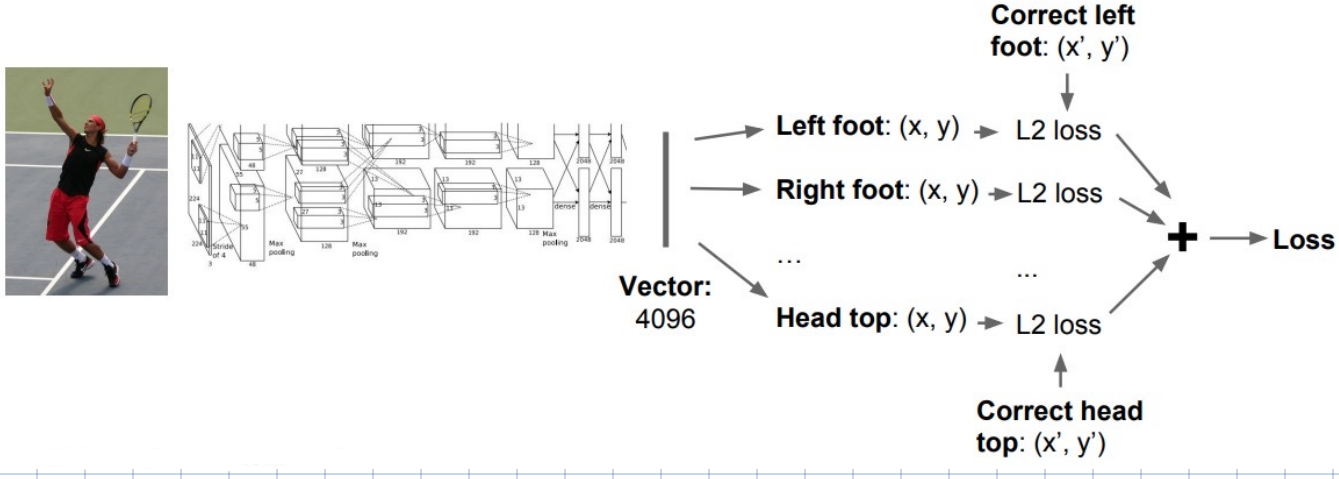


↳ Take weighted sum of the loss

↳ Hyperparameter

Human pose estimation:

↳ Represent by 14 joint locations



Object detection:

↳ Can't use above due many objects thus many coordinates

↳ Can't use sliding window because "How to choose the crop?", Objects can anywhere, at any size & too computationally expensive

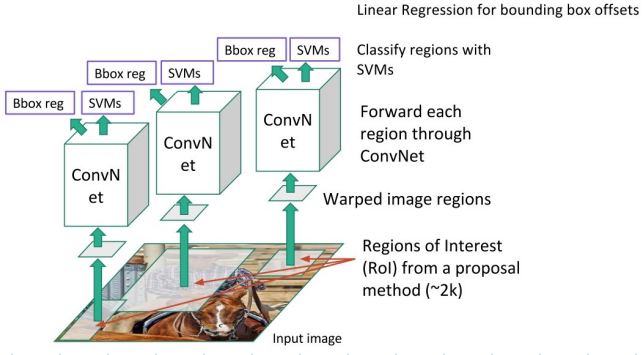
Region proposals:

↳ Gives regions where object might be present

↳ Selective search gives 2000 regions where objects might be found

R-CNN:

↳ Take proposal regions
↳ Change them into same size
↳ Pass them through CNN
↳ Use SVM
↳ Predicts bounding box (x,y,w,h)

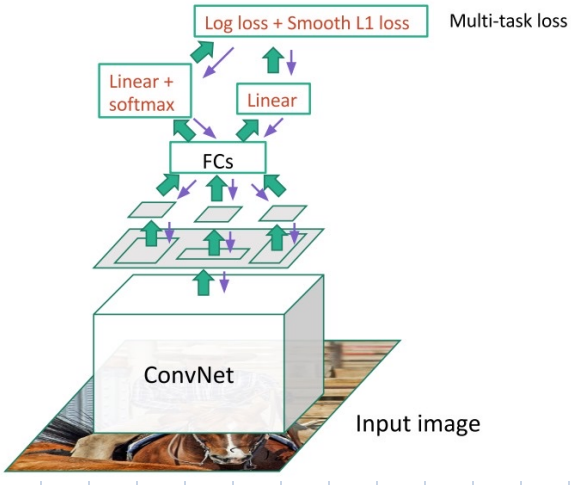


Problems:

↳ Computationally expensive
↳ Region proposals NOT learned
↳ Training is slow

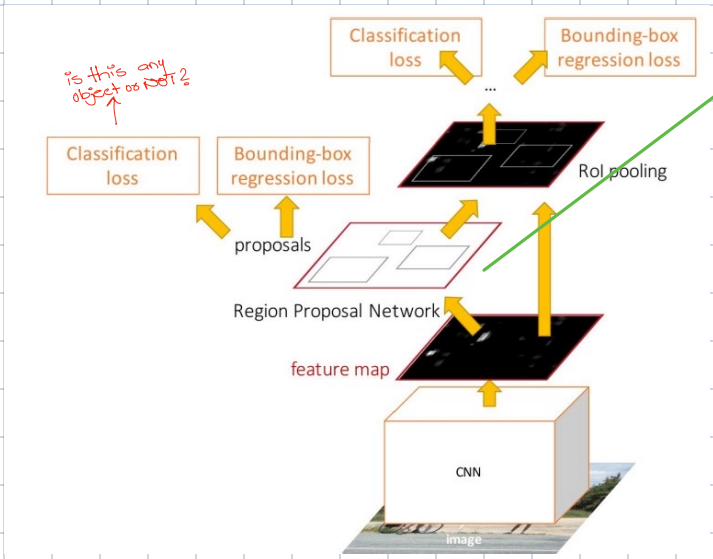
Fast R-CNN

↳ Run image through CNN
↳ Crop patches corresponding to the feature map on the input



↳ This ends up getting slowed by the process of computing region proposals

Faster R-CNN:



Any time region proposal has some overlap with any groundtruth object then that's positive region proposal

You only look once

Detections without proposals: YOLO / SSD → single shot detection

↳ Divide image into grids 7x7

↳ Image a set of base bounding boxes, B = 3

↳ Within each grid cell:

↳ Predict offset from base bounding box

↳ Predict scores

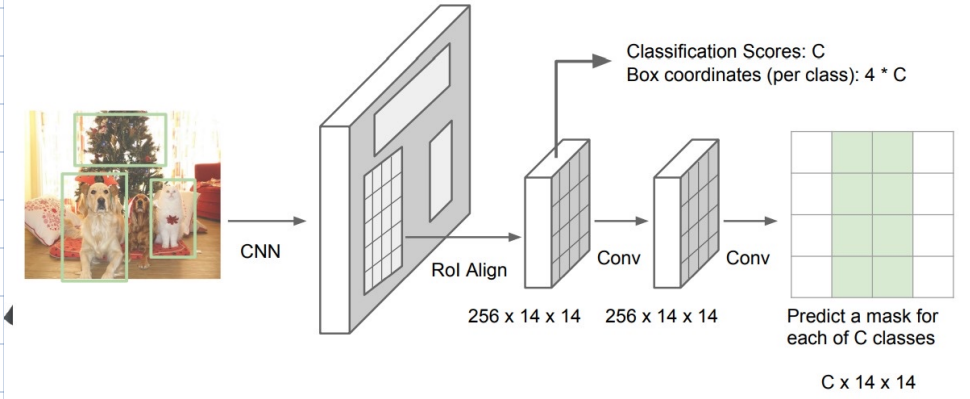
↳ Output: 7x7 x (5xC + C)

Instance Segmentation:

↳ Which pixels corresponds to object
↳ Differentiates between multiply same instances

Mask R-CNN:

↳ Same as faster Faster R-CNN
↳ Now, it labels the exact pixels in the region proposals



↳ Can also do pose detection