(11)

Detection & Segmentation & Localization

Semantic Segmentation

Classification + Localization

Object Detection

Instance Segmentation

★ Semantic Segmentation

- Image (Input) ⟶ desision of catagory for every Pixel in the image (Output)

⇒ Don't differntiate instances, only care about pixels

Idea1: Sliding Window

↳ Problem: Very inefficient! Not reusing shared features between overlapping patches.

Idea 2: Fully Convolutiond network

⇒ Design a network as a bunch of convolutiond layers to make predictions for pixels all at once.

⇒ Design network as a bunch of convolutiond layers, with downsampling & upsampling inside the network.

In-Network upsampling: Unpooling

1	2
3	4

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

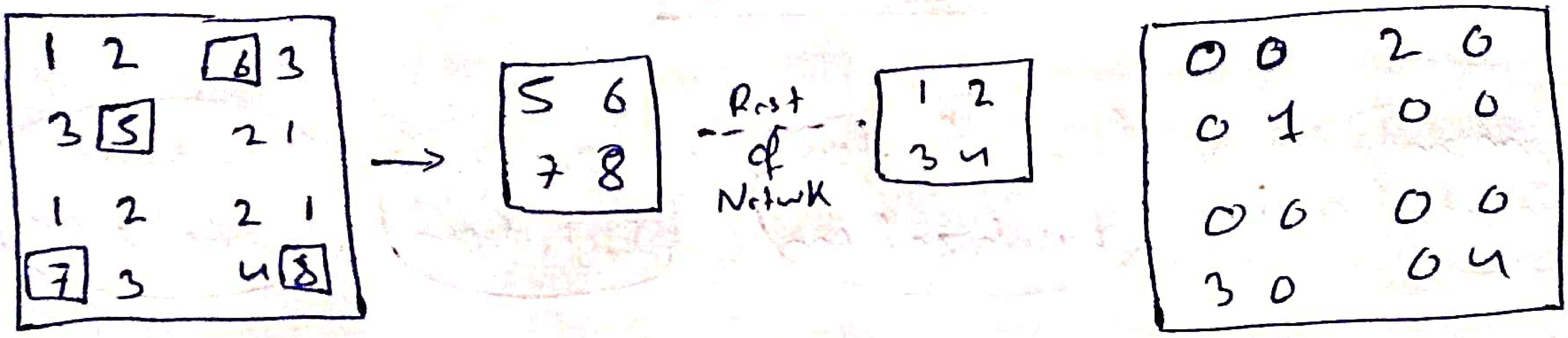
Nearest Neighbor

1	2
3	4

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

Bed of Nails

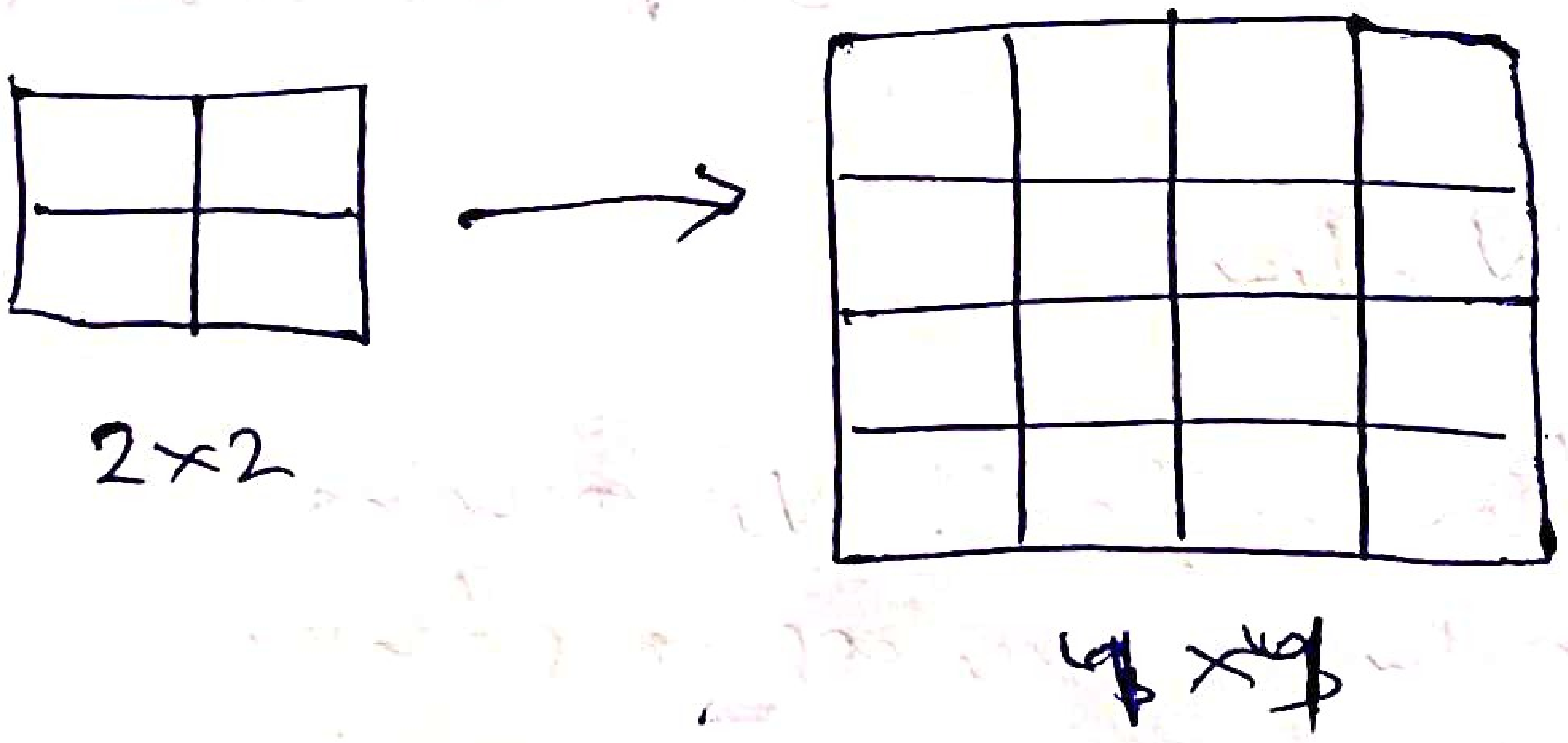
Max Pooling Max Unpooling



Remember which element was max!

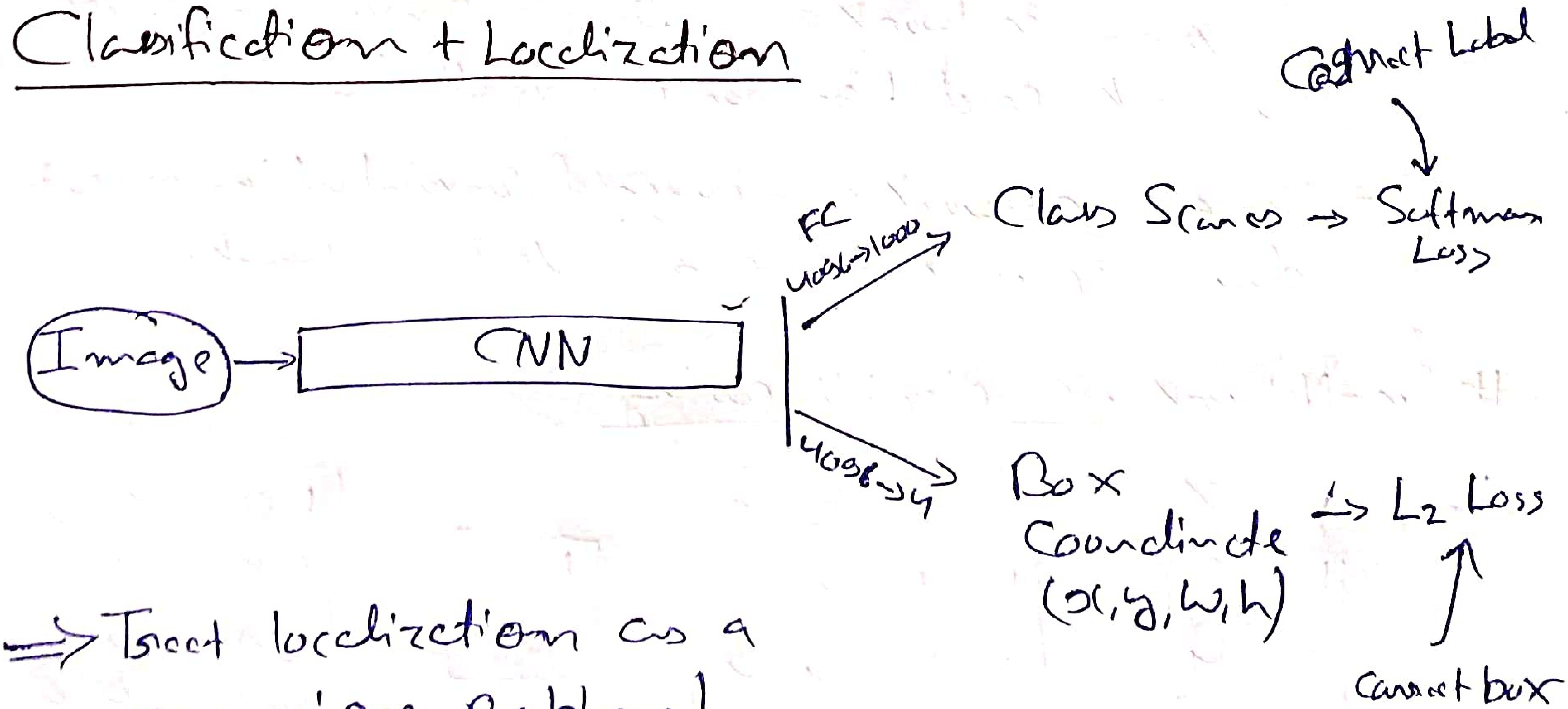
use position from Pooling layer.

Learnable Upsampling: Transpose Convolution



⇒ It uses filter for upsampling.

Classification + Localization



⇒ Treat localization as a regression problem!

★ Object Detection

⇒ Given an image draw bounding box around all the class in the image and give the class label.

⇒ Each image needs a different number of output.

⇒ One approach: Sliding Window

↳ Problem: Need to apply CNN to huge number of locations & scales, very computationally expensive!

Region Proposal

⇒ Find "blobby" image regions that are likely to contain object.

R-CNN

{Region Based Convolutional Neural Networks}

Fast R-CNN

Faster R-CNN

↳ Jointly train with 4 losses:

① → RPN classify object / not object

② → RPN regress box coordinate

③ → Find classification score (object class)

④ ↳ Find box coordinate

⊕ **Detection without Proposals: YOLO / SSD**

{You Only Look Once} ←

{Single-Shot Detection} ←

⇒ Within each grid cell:

- → Regress from each of the B base boxes to a final box with 5 numbers.

  (dx dy dh dw, confidence)

- → Predict score for each of C classes.

  (Including background as a Class)

★ **Instance Segmentation**

Mask R-CNN