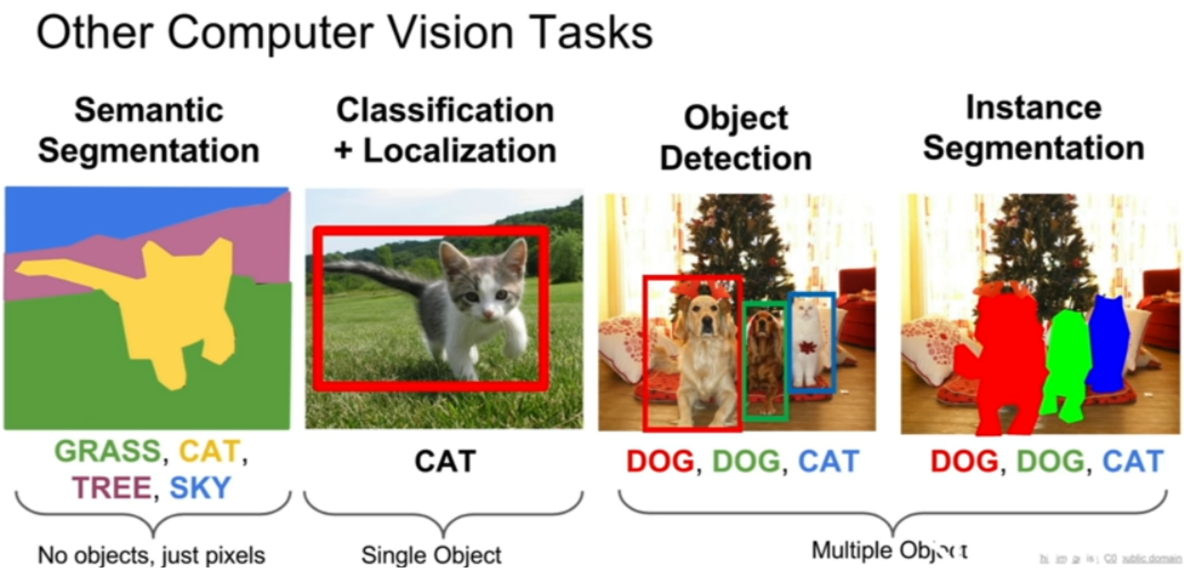
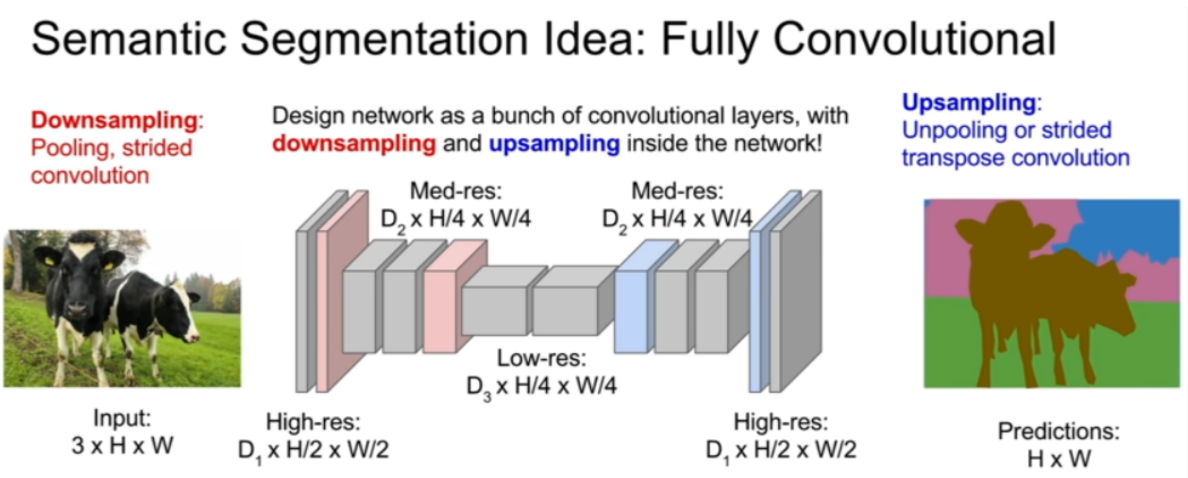
Detection & Segmentation



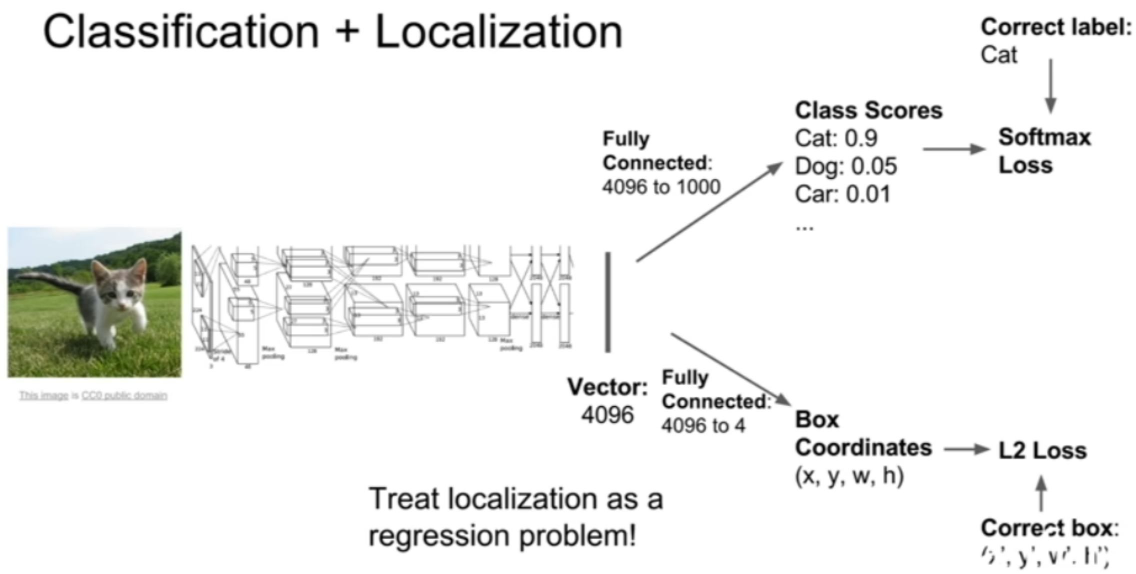
Semantic Segmentation

Label each pixel in the image with a category label

Ideas: sliding window (too expensive) → fully convolutional (downsampling & upsampling using average unpooling / max unpooling / transpose convolution)



Classification + Localization



Object Detection

Object detection as classification: sliding window (too expensive)

Region proposals: R-CNN, ~2k regions of interest → CNN, separated

Fast R-CNN, whole image → convolutional feature map → CNN

Faster R-CNN, insert Region Proposal Network (RPN) - faster region proposal

Detection without Proposals: YOLO (you only look once), SSD (single shot detector)

$$\text{Intersection over Union (IoU)} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

Mean Average Precision (mAP) metric:

- Calculate the precision and recall metrics.
- Calculate the area under the precision-recall curve.
- Measure the average precision.

With YOLO, you only look once at an image to perform detection

- We split the image into a grid
- Each cell predicts boxes and confidences: P(Object)
- Each cell also predicts a class probability
- Conditioned on object: P(Car | Object)

- Then we combine the box and class predictions
- Finally we do Non-maximum Suppression (NMS) and threshold detections

R-CNN: Region based ConvNets for Object Detection, Regions of Interest (RoI) from a proposal method (~2k) on input images

Fast R-CNN: Forward whole image through ConvNet, RoI from a proposal method on “conv5” feature maps (after ConvNet)

Faster R-CNN: Solely based on CNN, no external region proposals, Region Proposal Net after feature map instead

Mask R-CNN: Faster R-CNN with FCN on RoIs, mask other areas