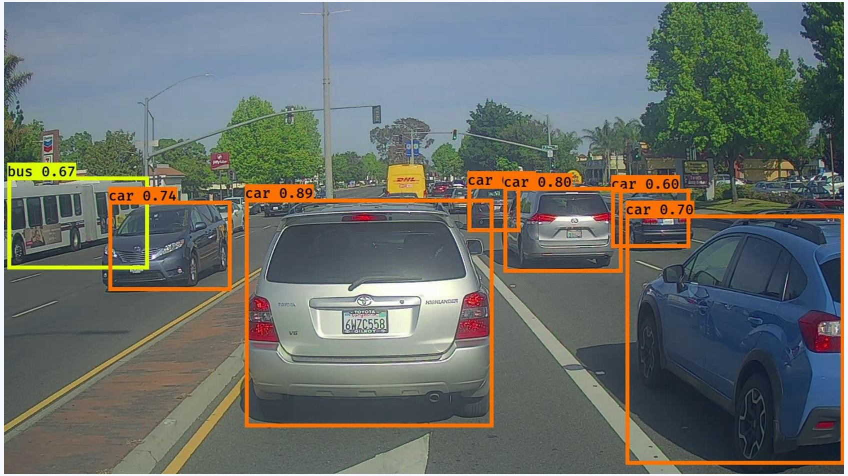
Object Detection and Image Segmentation

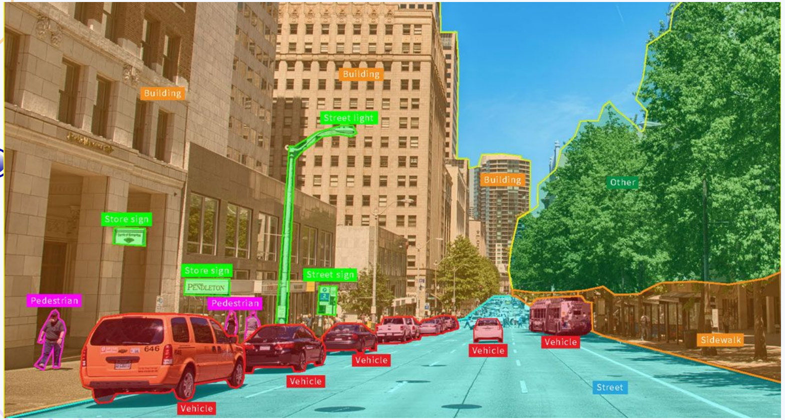
Object Detection and Image Segmentation

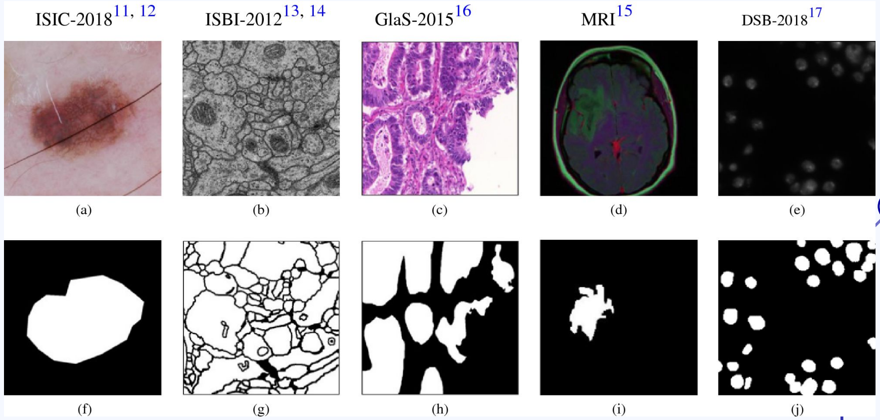
- Often, what we need is not just to classify an image into one of some distinct categories
- We need to locate where an object is in an image!



Object Detection and Image Segmentation

- Useful for scenarios such as self-driving cars, biomedical image analysis, etc.



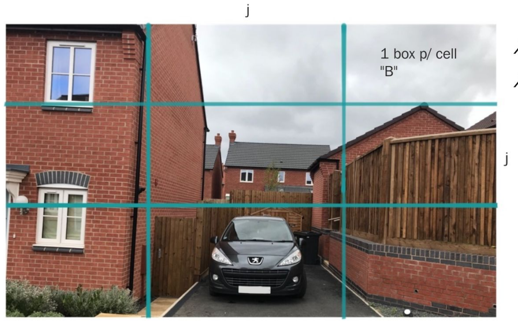


Object Detection

- Detects objects in an image (for each pixel whether there’s an object) and draws bounding boxes around them
- Classify each bounding box into a certain class of image

$1_{ij}^{obj} \longrightarrow$ 1 if box j and cell i match, 0 otherwise

$1_{i}^{obj} \longrightarrow$ 1 if cell has object, 0 otherwise



$$\lambda_{coord} = 5{,}0$$
$$\lambda_{noobj} = 0{,}5$$

$$Loss_{yolo} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2 \right] + \longrightarrow$$ Bounding Box coord

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

$$\sum_{i=0}^{S^2} 1_{ij}^{noobj} \sum_{C \in classes} \left[\left(p_i(C) - \hat{p}_i(C)\right)^2 \right] \longrightarrow$$ Classification

$1_{ij}^{noobj} \longrightarrow$ 1 if box j and cell i no match, 0 otherwise

Object Detection

- Advantages: easy to postprocess results; relatively easier to annotate quality bounding boxes
- Disadvantages: does not tell where the object actually is in the bounding box

Image Segmentation

- Semantic segmentation: classifies each pixel in an image into a class out of many
- Instance segmentation (sometimes known as panoptic segmentation in its more complete form): classifies each pixel in an image into a class AND separate different instances of the same class

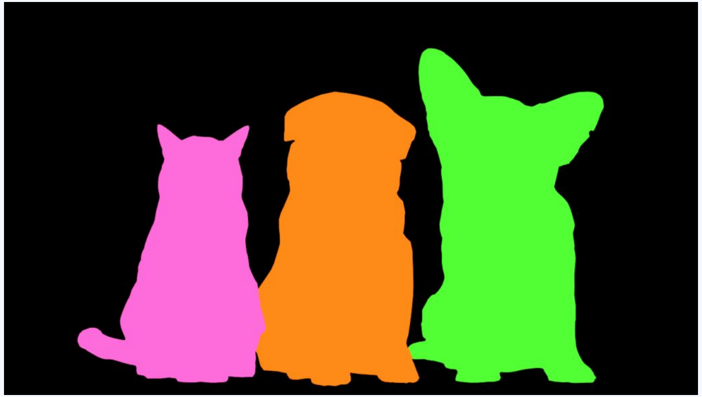


Image Segmentation

- Advantages: provides complete information about the location of the subjects in an image
- Disadvantages: requires more effort in carefully labeling data (needs pixel-perfect annotations); requires more sophisticated postprocessing techniques

Image Segmentation

- How many channels should the output segmentation map contain?

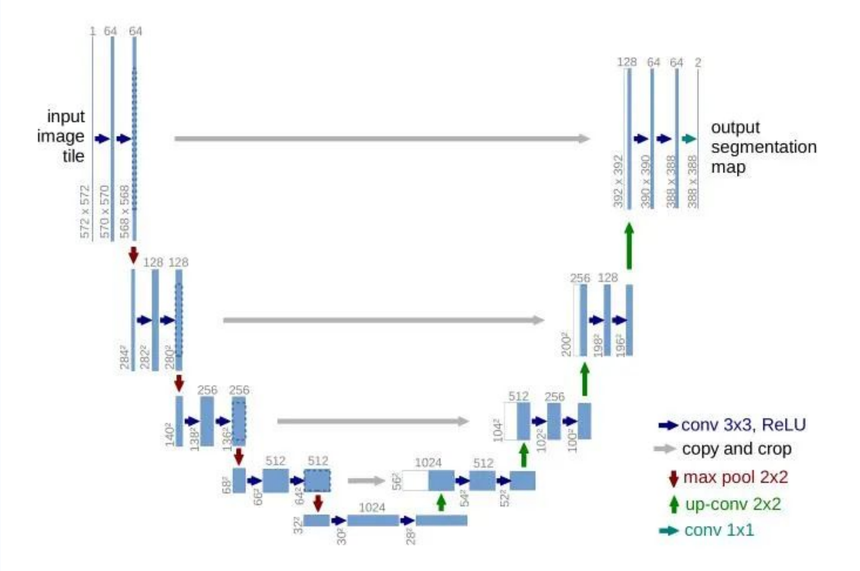
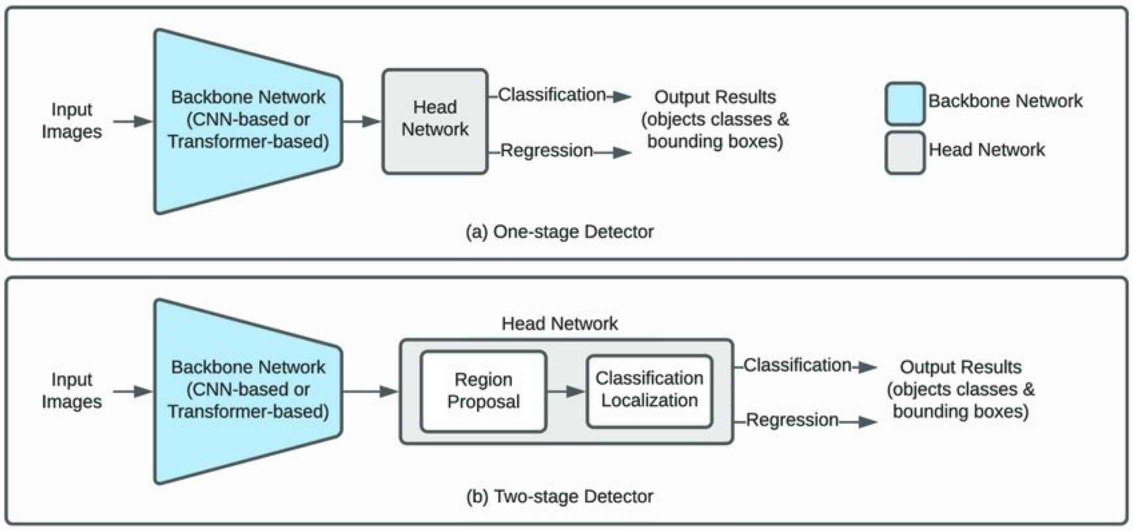


Image Segmentation

- The number of channels should be the same as the number of classes
- One channel is reserved for each class
- Therefore you can run a standard classification for each pixel

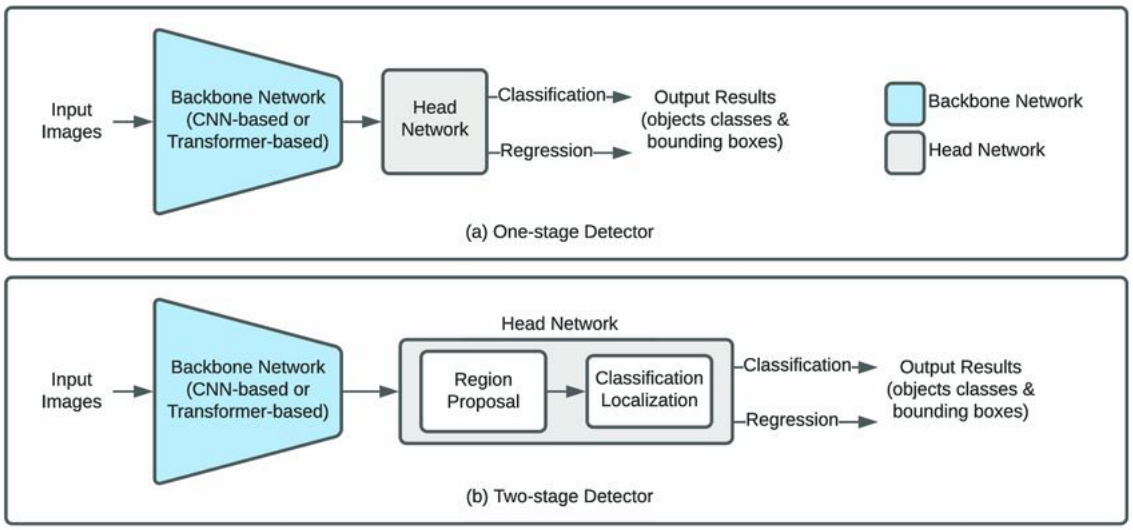
Object Detection

- Two-stage detectors: the first stage is region proposal, in which a region proposal network extracts regions of interest of an image
- This is then fed into a head which carries out image classification (whether there's any object, and which object is it) and bounding box localization



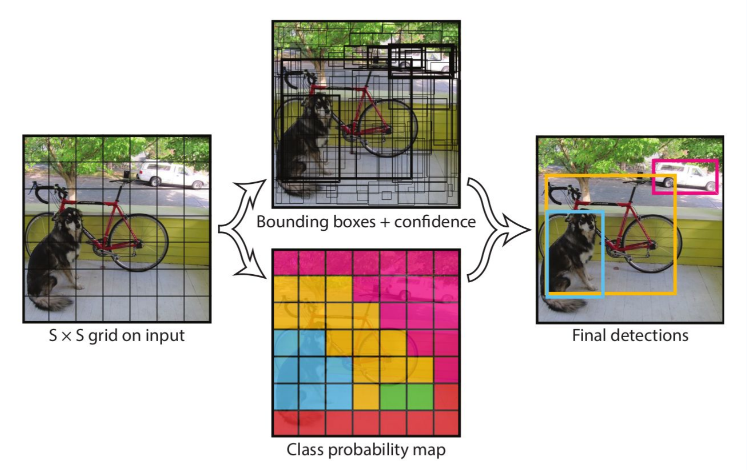
Object Detection

- One-stage detector: after feature extraction, a specialized head directly processes the features extracted and runs classification and localization at the same time



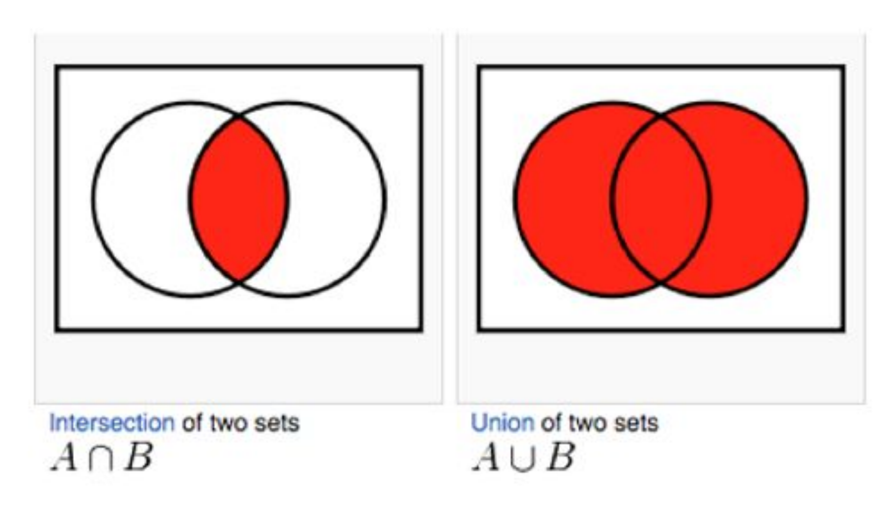
Object Detection

- For each predefined "anchor point" in the image, predict a bounding box and a confidence score
- Using non-maximum suppression (NMS), less confident bounding boxes which intersect much with more confident bounding boxes are removed, leading to clean object detections



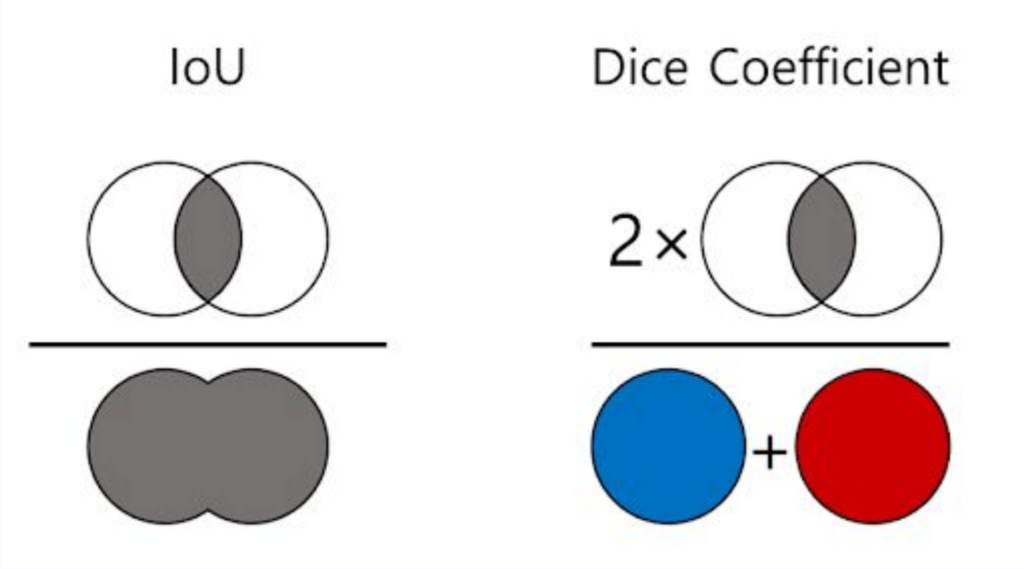
Metrics

- There are some metrics which assess the goodness of the predictions of a detection or segmentation network
- Some of them involve combination of intersection and union between predictions and ground truth



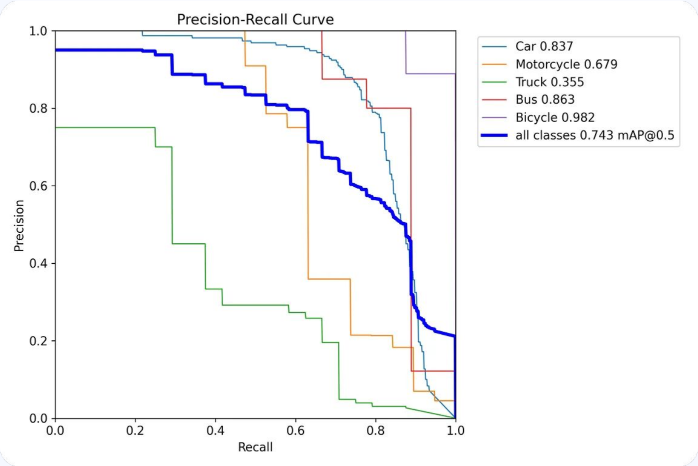
Metrics

- IoU: intersection / union
- Dice score: 2 * intersection / (size of predicted mask + size of ground truth mask)



Metrics

- Average precision (AP): area under precision-recall curve when trying out different thresholds across confidence scores
- To count as a correct prediction, the IoU must be over a threshold (hence the 0.5 in mAP@0.5 for example)



Try out yourself

https://colab.research.google.com/drive/1xlNuI2aq3r9szqTymSP8T4fauWXXnYfI#scrollTo=YsGiVYcaGaMr