

# Instance Segmentation

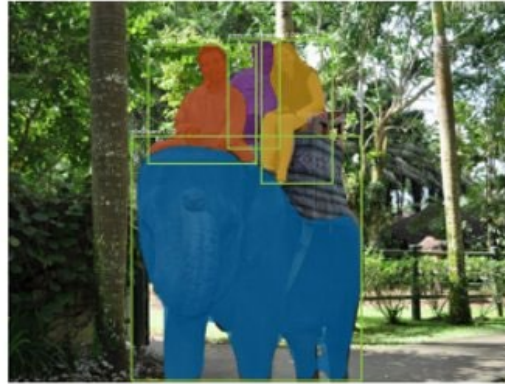
## Computer Vision

# Why still use a two-stage object detector?

- Better recall of RPN as compared to SSD/YOLO
  - Trained with all object instances
  - Generic first stage, usable for multitask
- Finer control over training classifier
  - Custom minibatch (sampling 3:1 negative samples)
- Instance-level multitask (Mask-RCNN)

# Why still use a two-stage object detector?

- Better recall of RPN as compared to SSD/YOLO
  - Trained with all object instances
  - Generic first stage, usable for multitask
- Finer control over training classifier
  - Custom minibatch (sampling 3:1 negative samples)
- Instance-level multitask (Mask-RCNN)



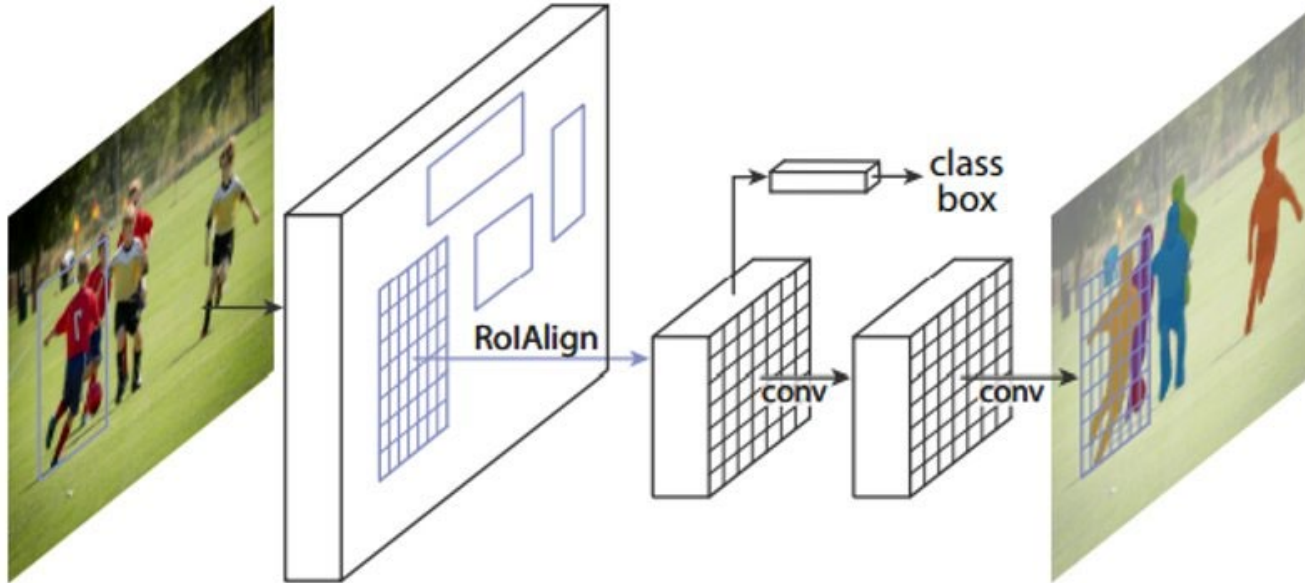
# Mask R-CNN – Towards Instance-Level Understanding



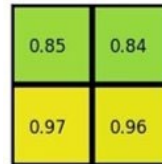
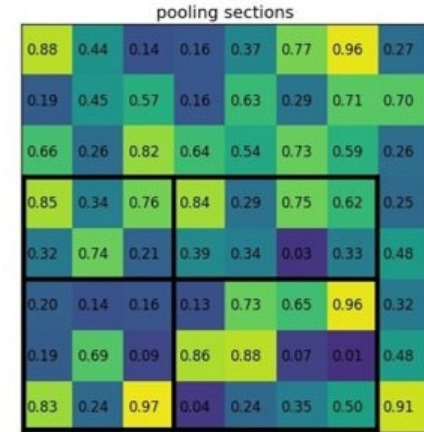
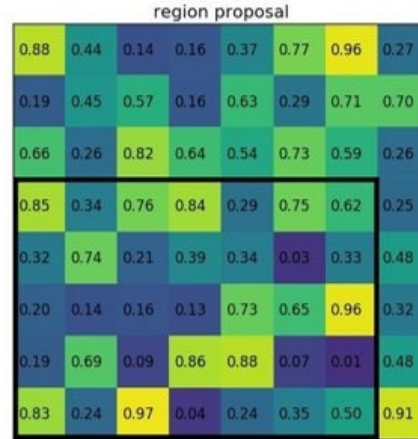
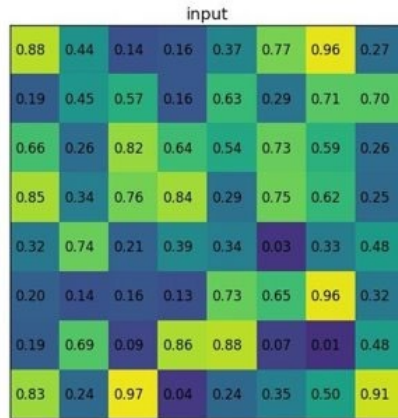
Zoom in on instances

# Mask R-CNN

Preserves pixel-to-pixel alignment

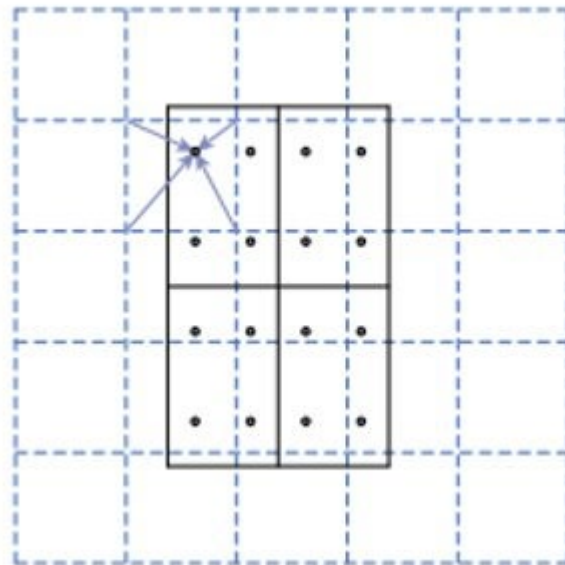


# Quantization – loss of pixel-to-pixel alignment



# RoI Align - Improvement on RoI Pooling

- Input: Feature map (5x5 here) and region proposal (normalized float coordinates)
- Output: 2x2 'pooled' bins
- Sample 4 points in every bin uniformly
- Compute value at each bin using bilinear interpolation
- Max or average the 4 bins





# Class Imbalance in Training a Classifier

- While training detectors, maximum samples are background (negatives)
- Faster R-CNN: Ratio of 3 negatives to 1 positive is maintained while training classifier head Custom minibatch
- Not easy in single stage detectors

# Class Imbalance in Training a Classifier

- Cross entropy loss

$$\text{CE}(p_t) = -\log(p_t)$$

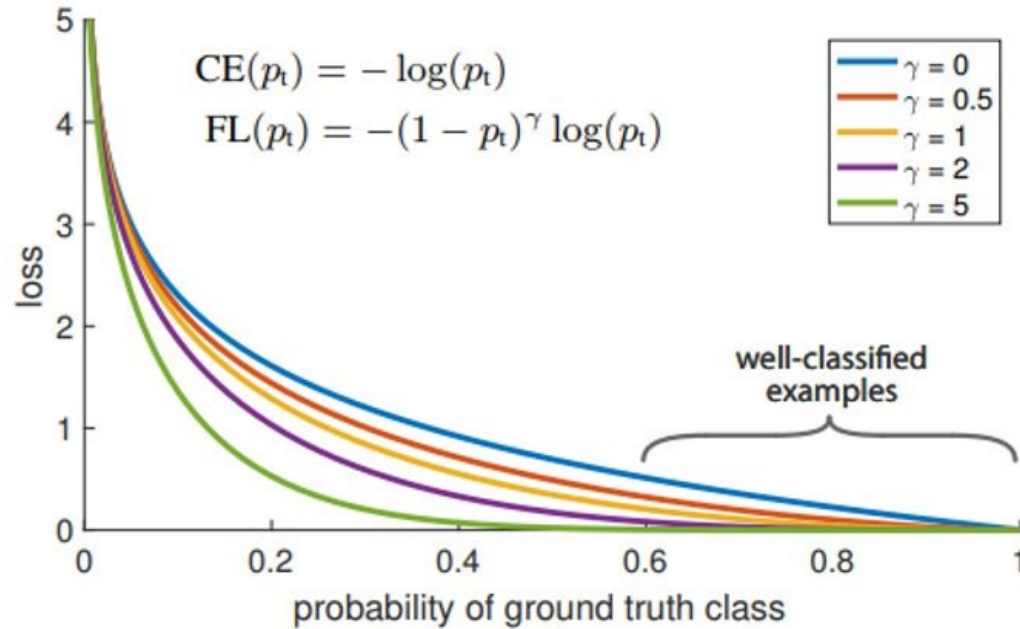
- Balanced cross entropy loss

$$\text{CE}(p_t) = -\alpha_t \log(p_t)$$

- Focal Loss

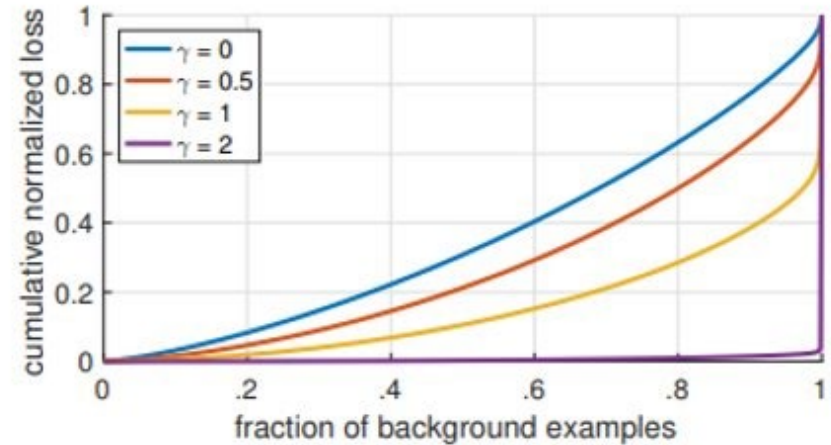
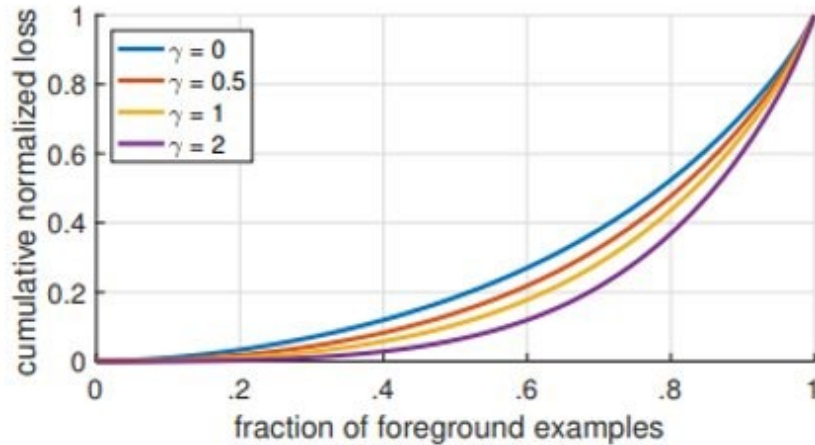
$$\text{FL}(p_t) = -(1 - p_t)^\gamma \log(p_t)$$

# Focal Loss



# Focal Loss

$$\text{FL}(p_t) = -(1 - p_t)^\gamma \log(p_t)$$



# Thank you!

Happy Learning :)