Practice of Computer Vision

Lecture 3

TA: Wonjong Jang

Object detection

RCNN

Fast RCNN

Faster RCNN

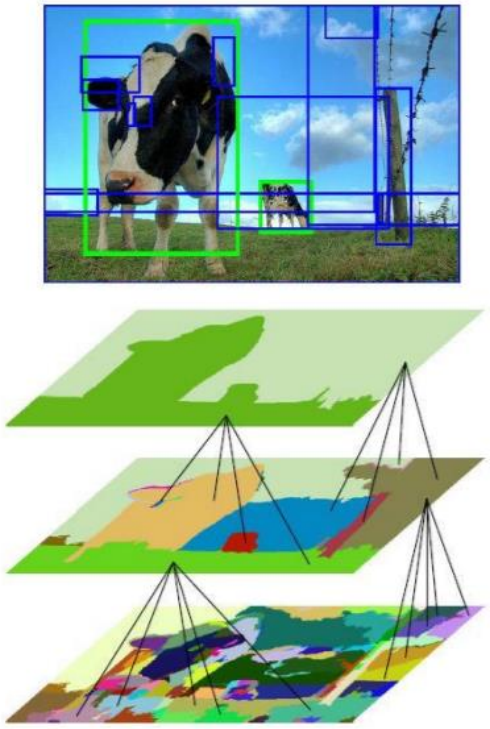
Object Detection: A Naïve Approach

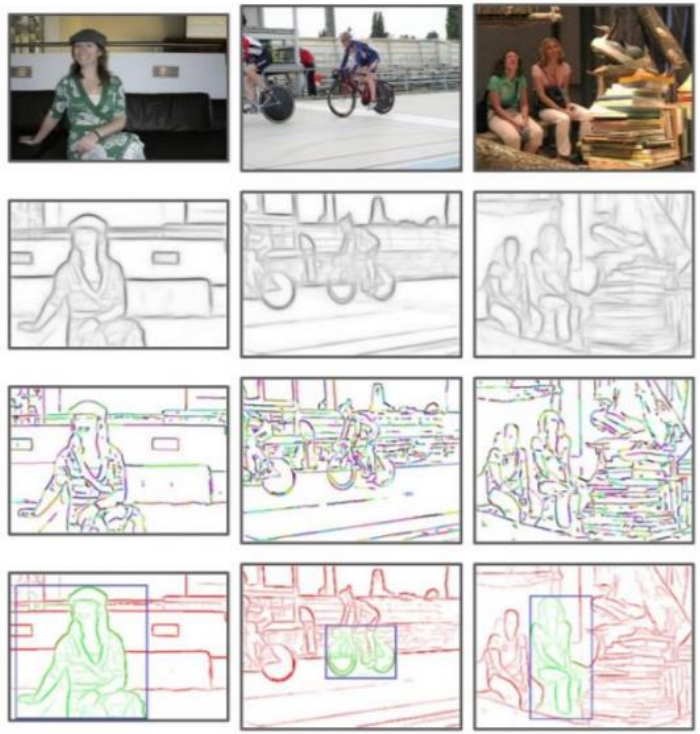
- Motivated by the great success of deep learning in image classification

Object Detection = Box localization + Box classification

Box localization: Object proposals

Box classification: CNN

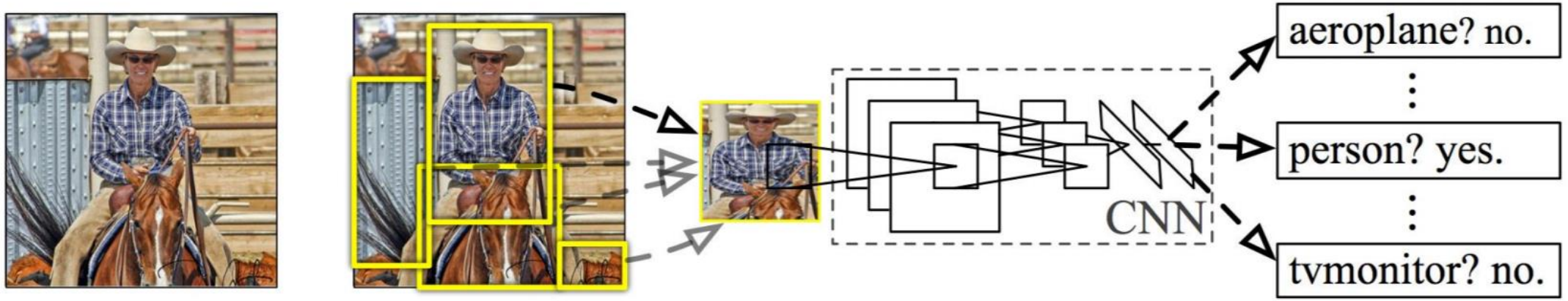




(*left*) Selective Search for Object Detection, IJCV 2013

(*right*) Edge Boxes: Locating Object Proposals from Edges, ECCV 2014

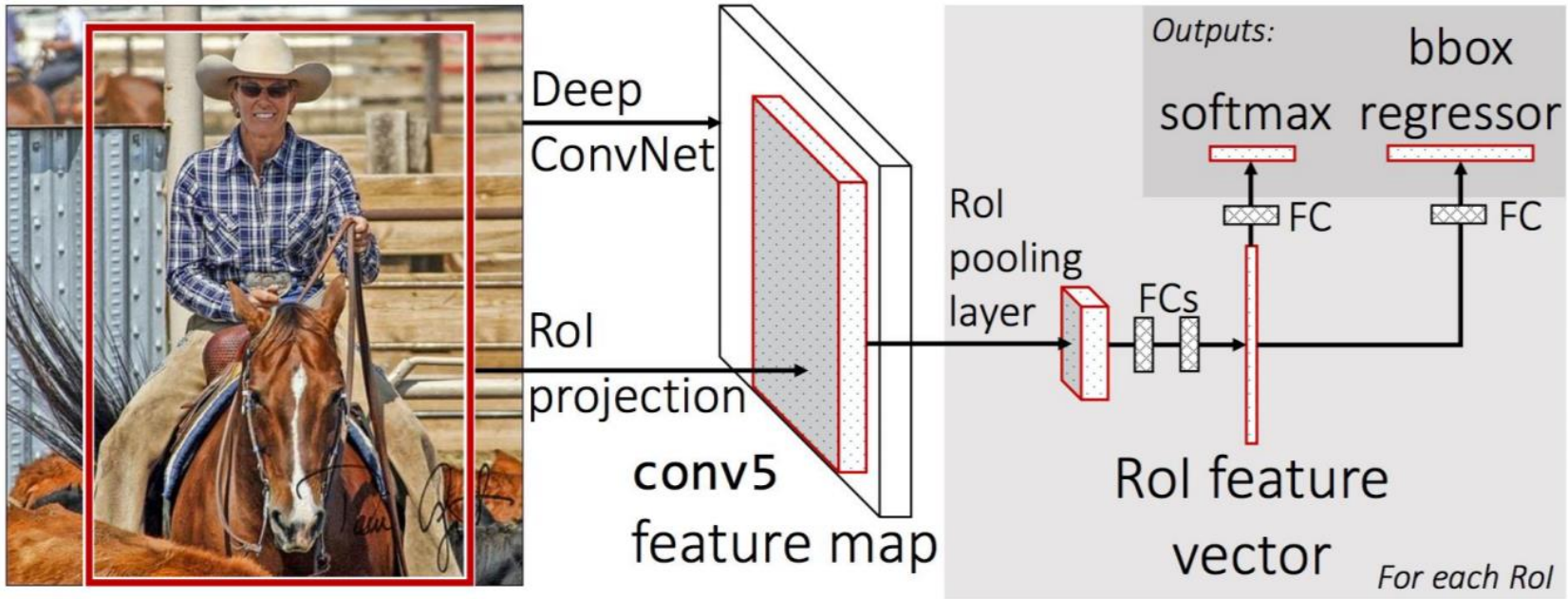
Region-based CNN (R-CNN)



- Summary
  - Independent evaluation of each proposal
  - Bounding box regression improves detection accuracy.
  - Mean average precision (mAP): 53.7% with bounding box regression in VOC 2010 test set

*Girshick et al., *Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation*, CVPR 2014

Fast R-CNN

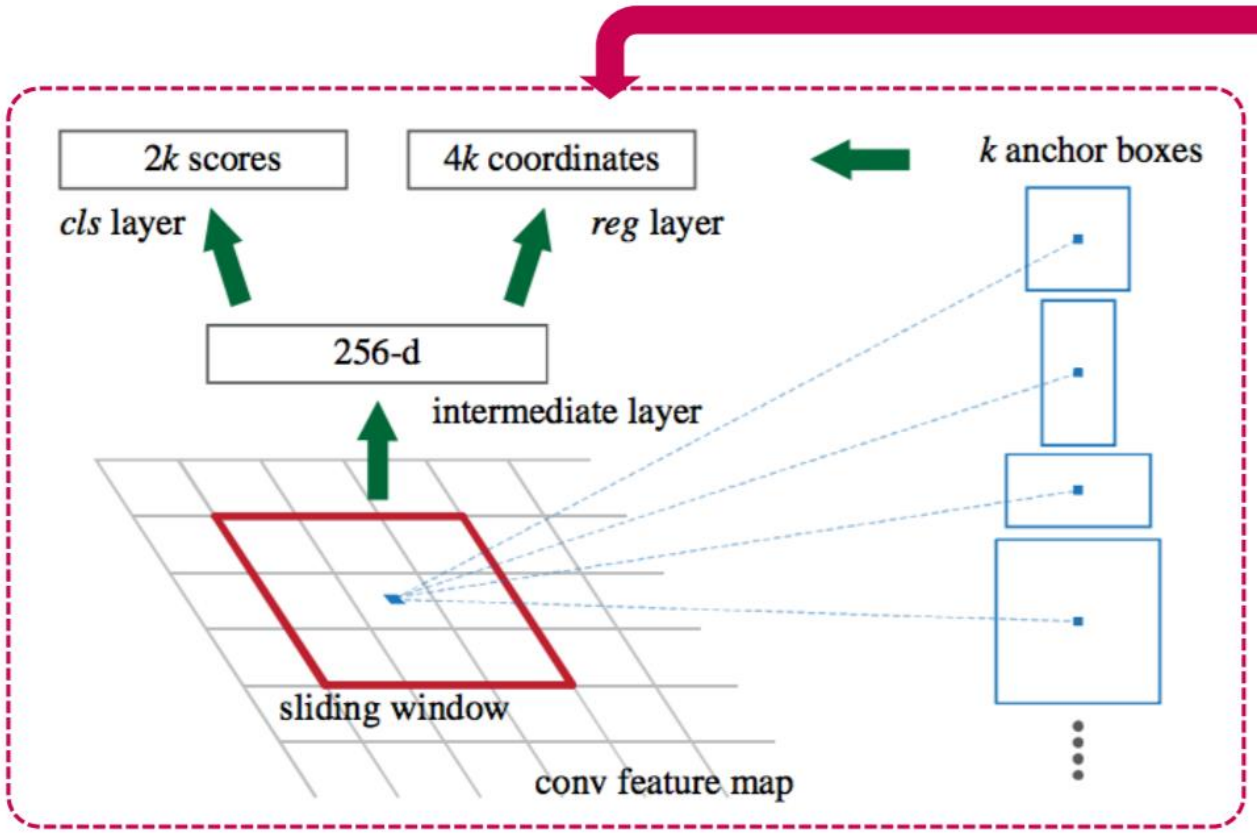


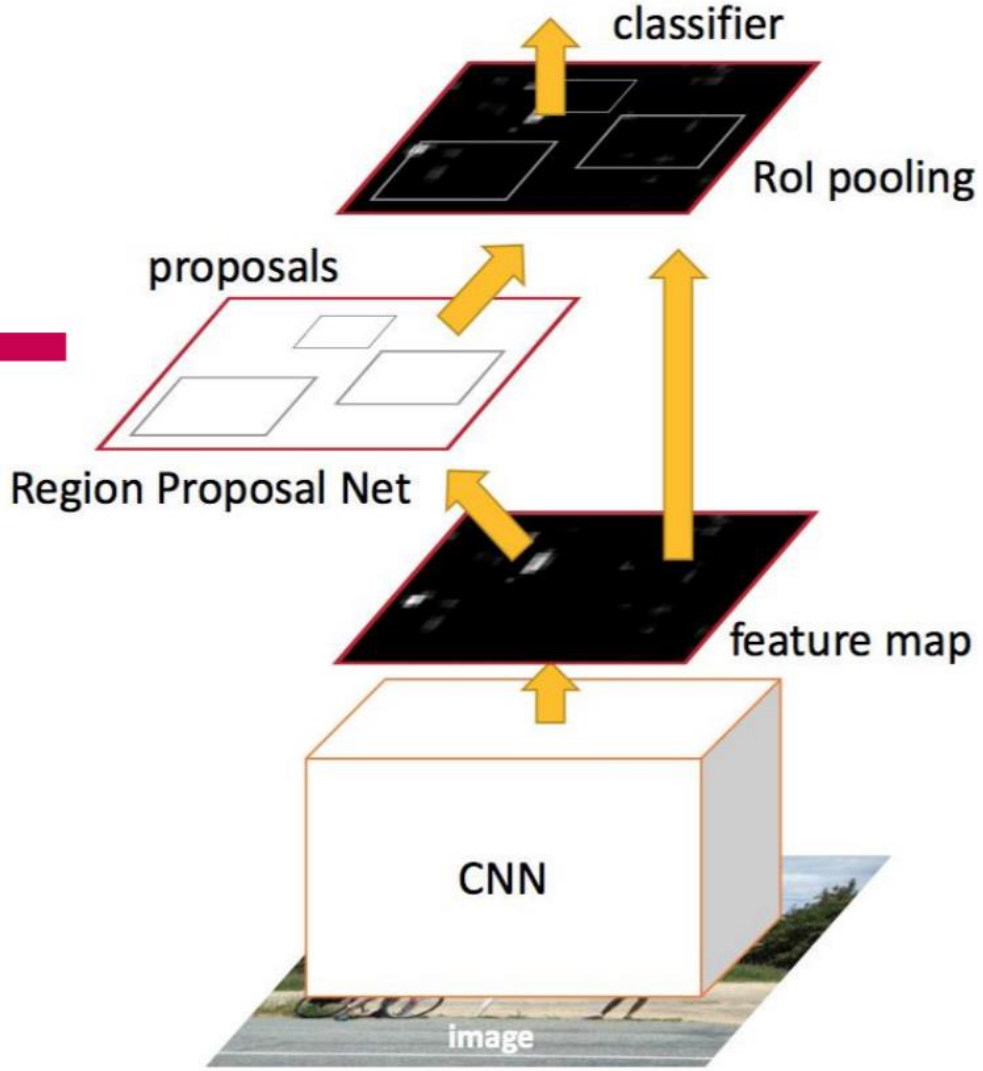
- A fast version of R-CNN
  - 9x faster in training and 213x faster in testing than R-CNN
  - A single feature computation and ROI pooling using object proposals
  - Bounding box regression into network
  - Single stage training using multi-task loss

*Girshick, *Fast R-CNN*, ICCV 2015

Faster R-CNN

- Fast R-CNN + Region Proposal Network
  - Proposal computation into network
  - Marginal cost of proposals: 10ms





*Ren et al., *Faster R-CNN*, NIPS 2015

Faster R-CNN

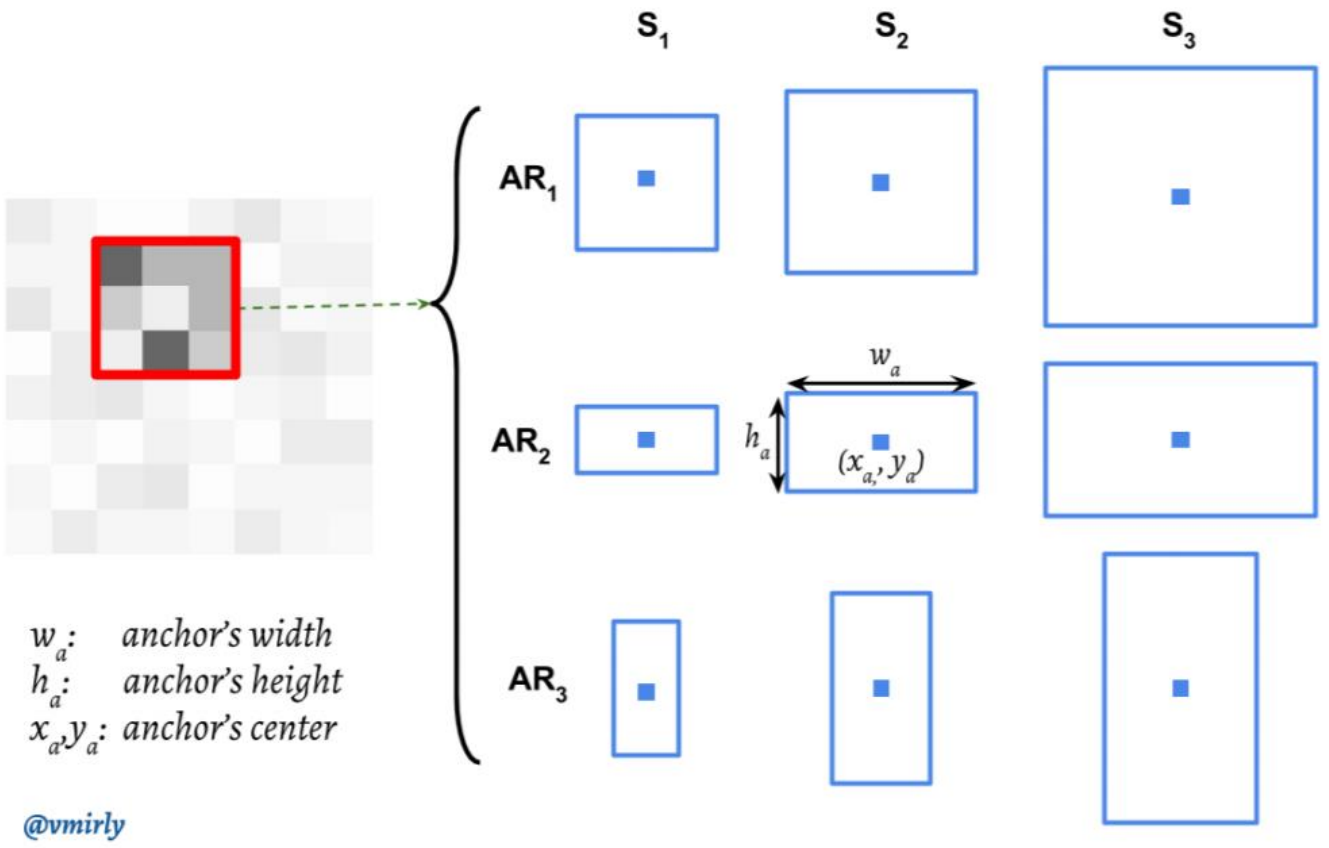
- Details of the region proposal network
  - 9 anchors per location (3 aspect ratios x 3 scales)
  - Groundtruth label per anchor

$$p^* = f(x) = \begin{cases} -1, & \text{if IoU} < 0.3, \\ 1, & \text{if IoU} > 0.7, \\ 0, & \text{otherwise.} \end{cases}$$

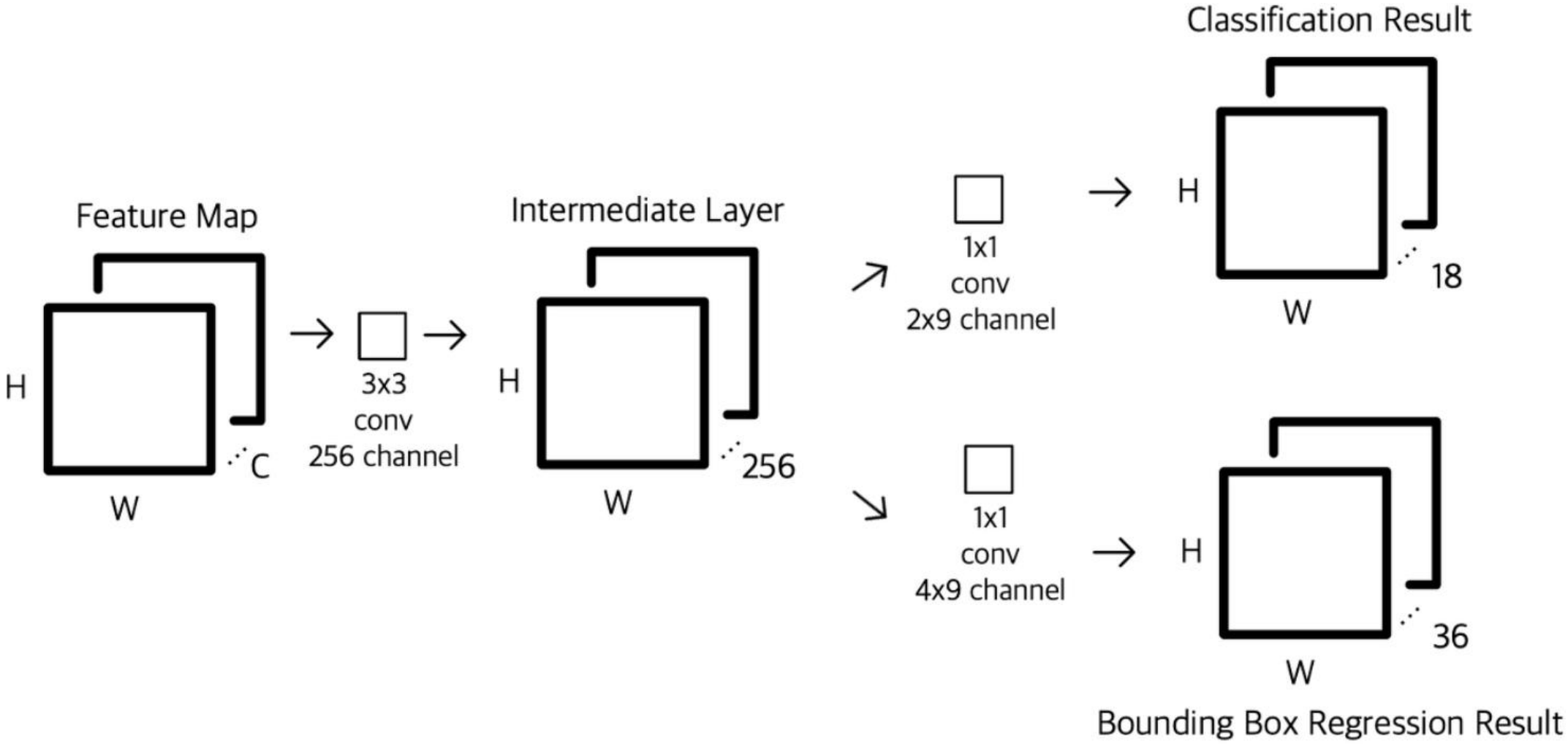
where IoU is intersection over union:

$$\text{IoU} = \frac{\text{Anchor} \cap \text{GTBox}}{\text{Anchor} \cup \text{GTBox}}$$

  - Trained with *a binary classification loss* for anchor selection and *a regression loss* for box refinement



Faster R-CNN



RPN Structure

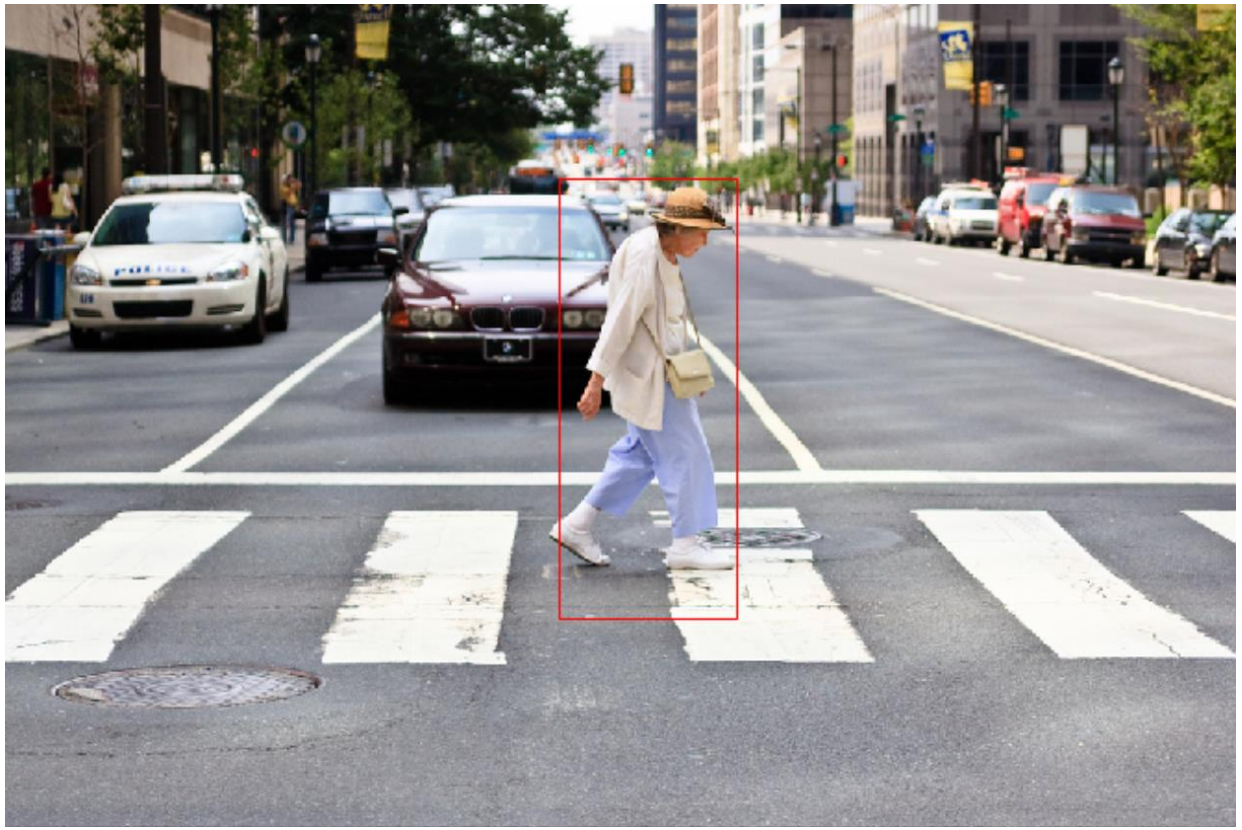
Faster R-CNN

Summary of the process in Faster R-CNN

Image → (Feature extractor network) → Feature map → (Region Proposal Network) → Region Proposal

Feature map → (RoI pooling using Region Proposal) → RoI pooled feature map → (Classifier) → label
→ (Regreesor) → bbox

Today’s practice - pedestrian detection



Thanks