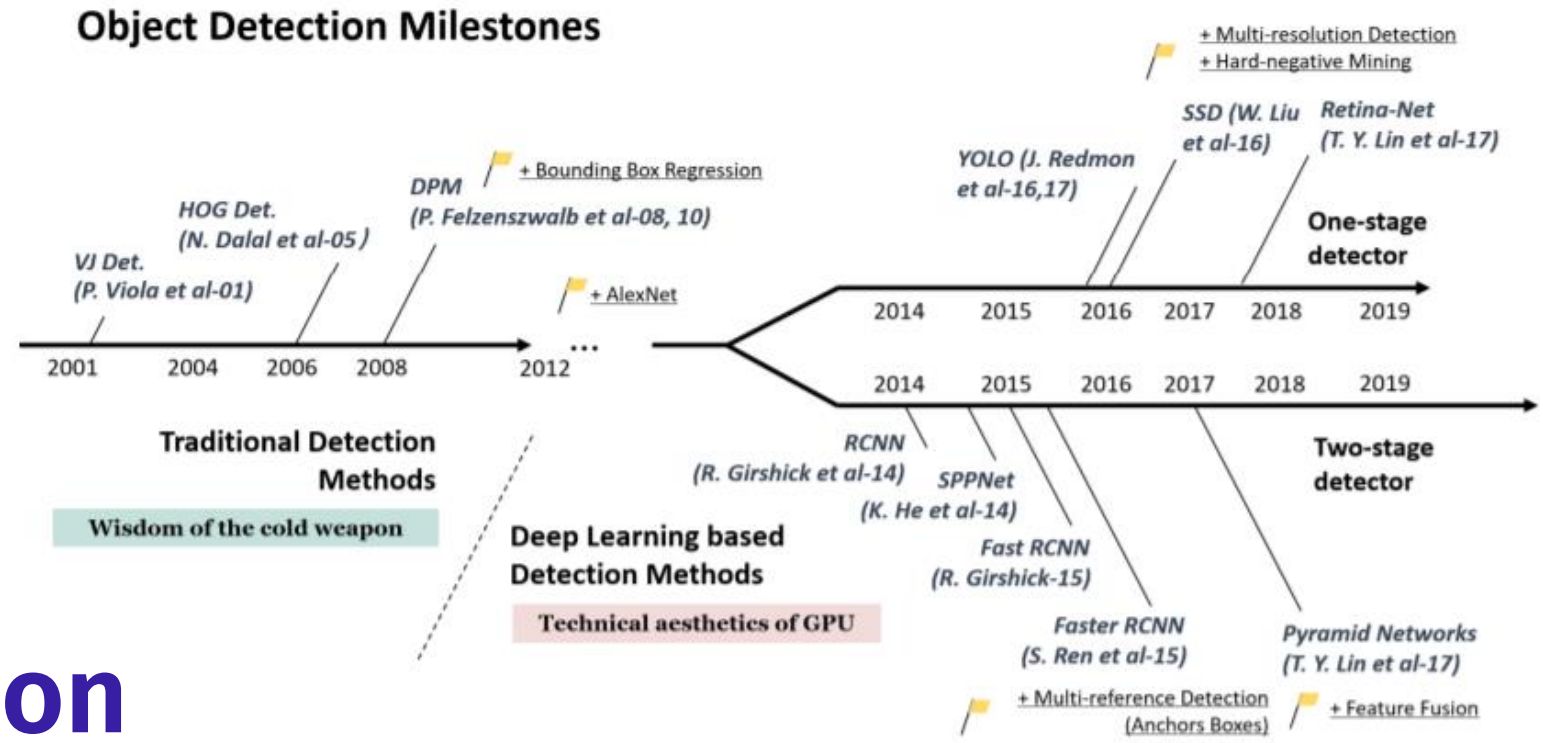
A brief introduction to object detection

Jakub Wiśniewski

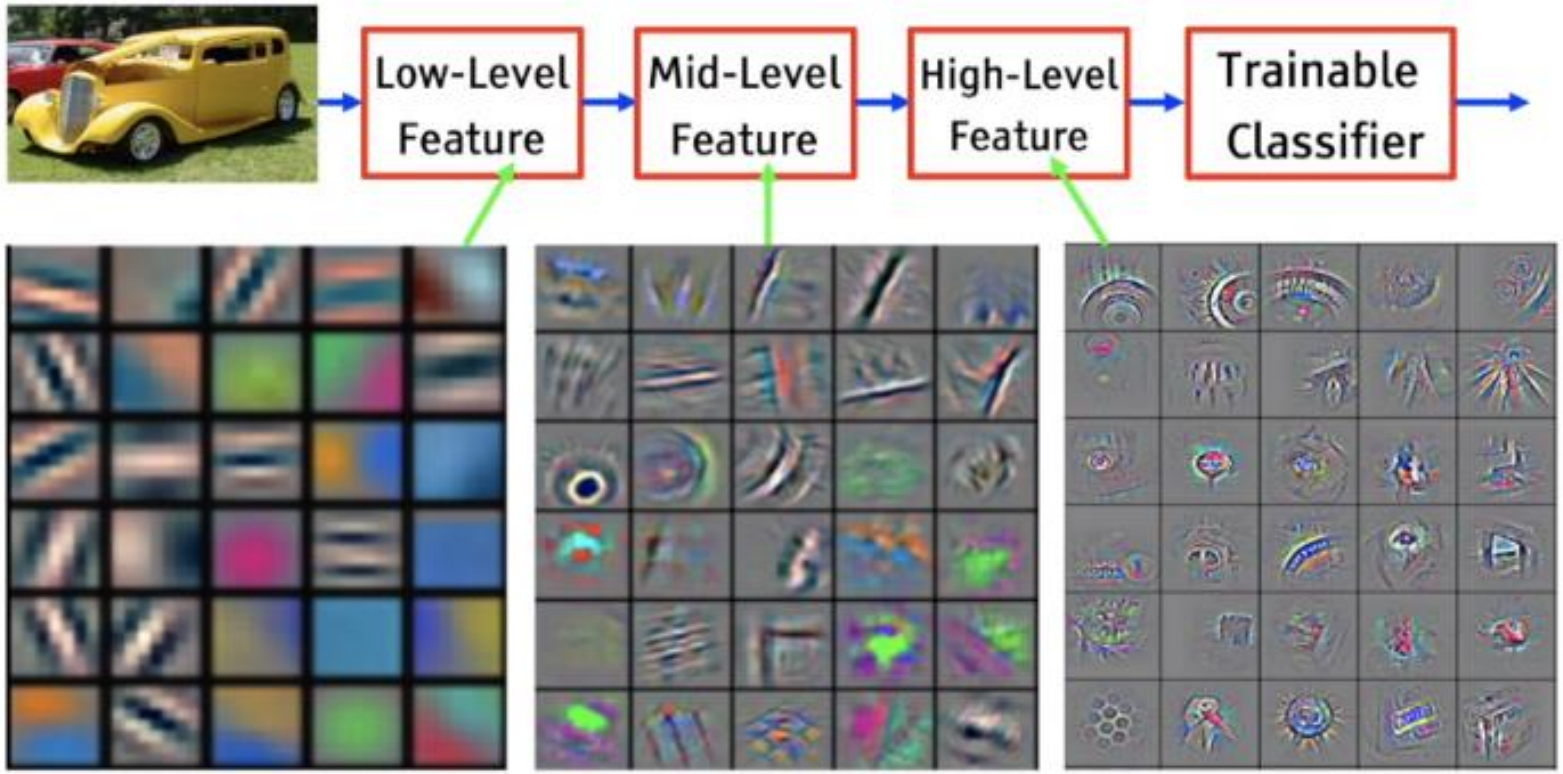
Agenda

- Influence of convolutions
- One-stage & Two-stage detectors
- R-CNN
- Fast R-CNN
- Faster R-CNN
- Faster R-CNN + FPN
- Sparse R-CNN

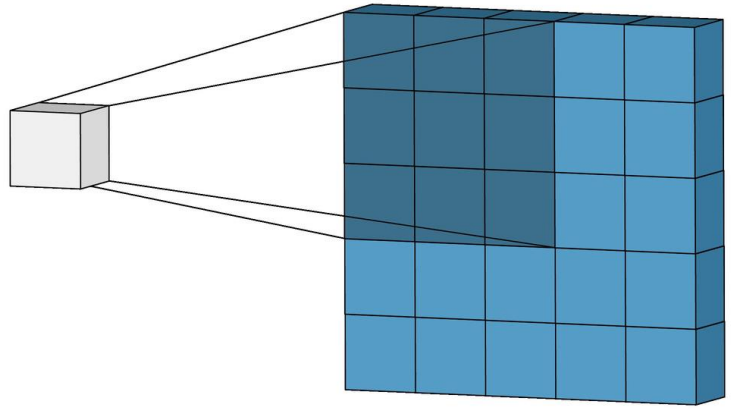


Aim : Gaining intuition

Convolutions

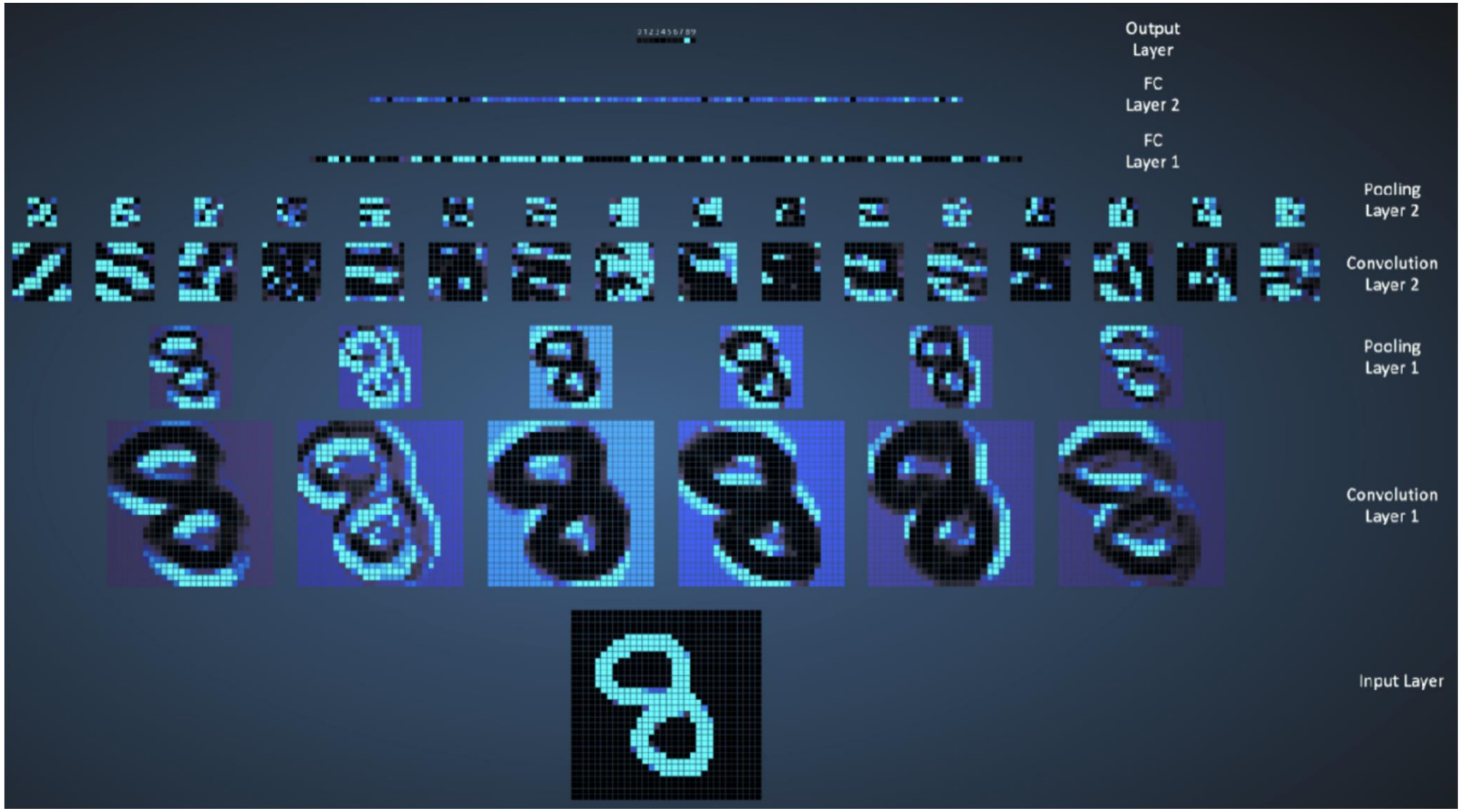


Shiv Vignesh - The world through the eyes of CNN



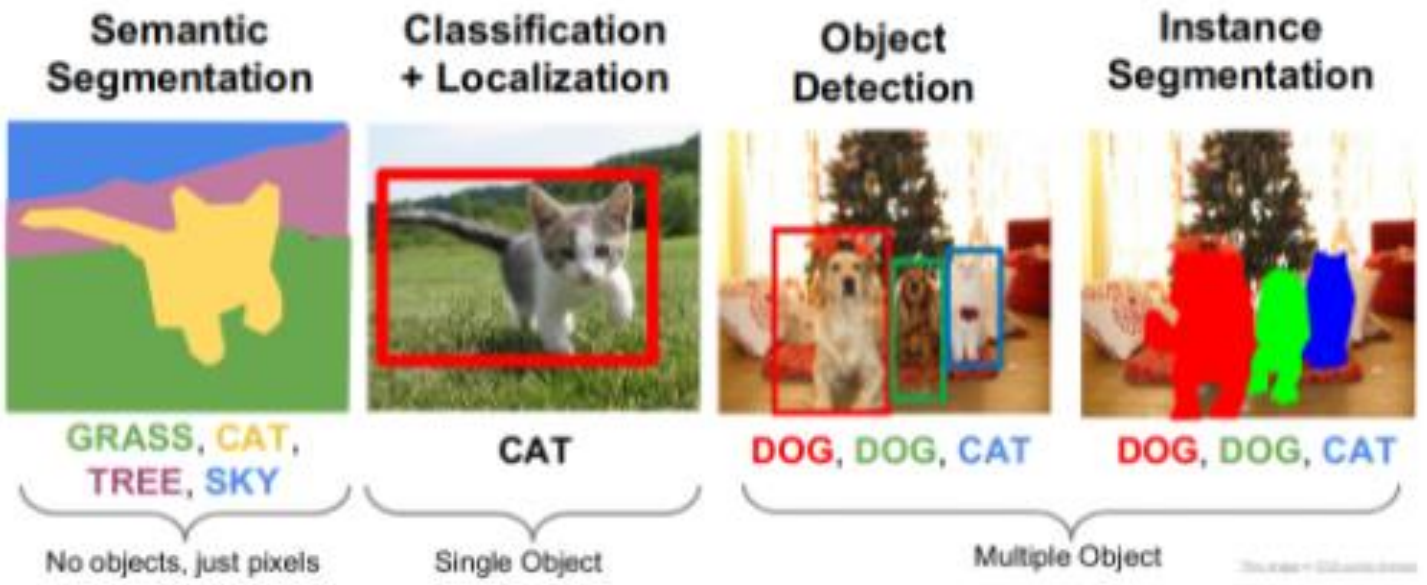
Irhum Shafkat - Intuitively Understanding Convolutions for Deep Learning

Convolutions



https://wandb.ai/site/articles/intro-to-cnns-with-wandb

Defining the term



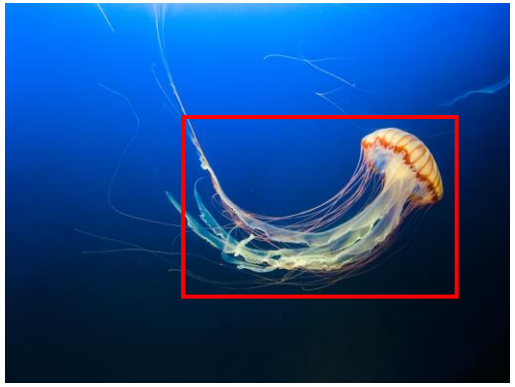
CS231 Stanford course

$$\text{Loss} = L_{CLF} + \alpha * L_{REG}$$

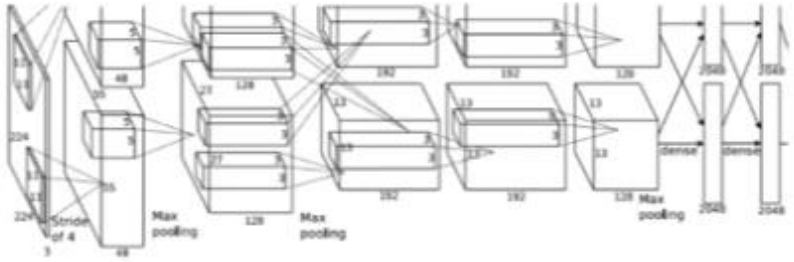
Bounding box proposal
+
CNN

Classification

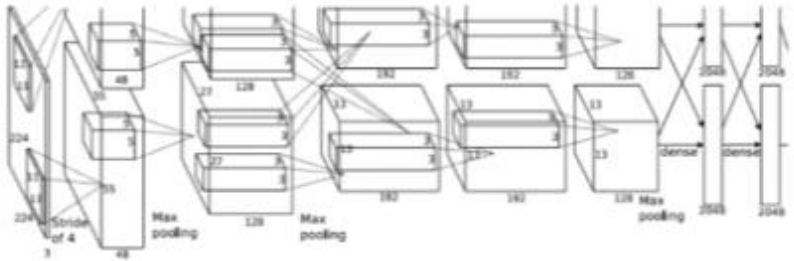
Bounding box regression



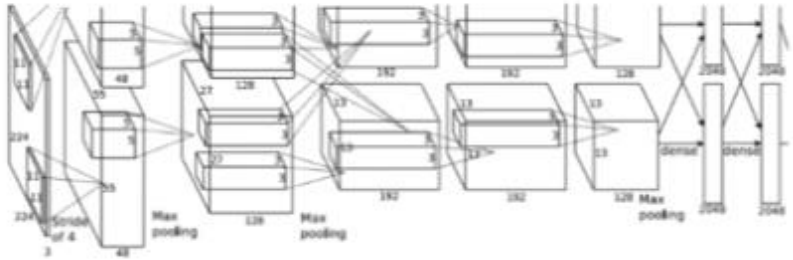
Switching to object detection



CAT: (x, y, w, h)



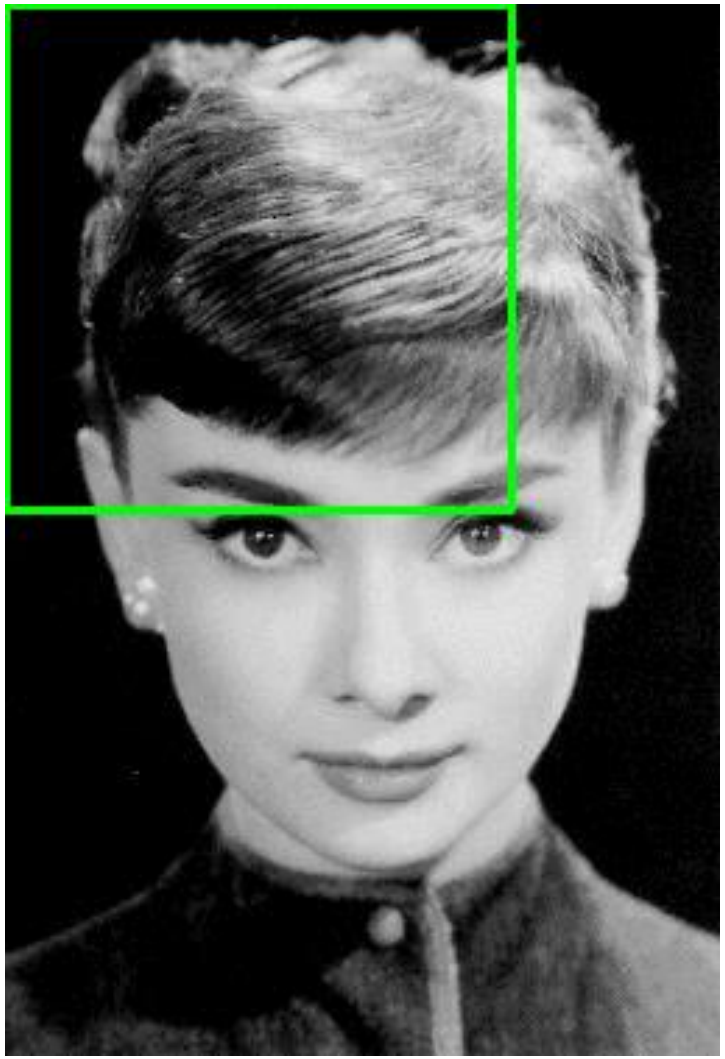
DOG: (x, y, w, h)
DOG: (x, y, w, h)
CAT: (x, y, w, h)

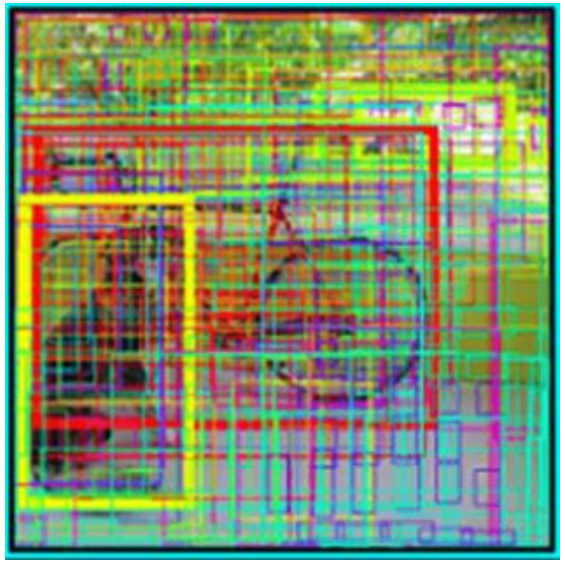


DUCK: (x, y, w, h)
DUCK: (x, y, w, h)
....

CS231 Stanford course

Naive approach



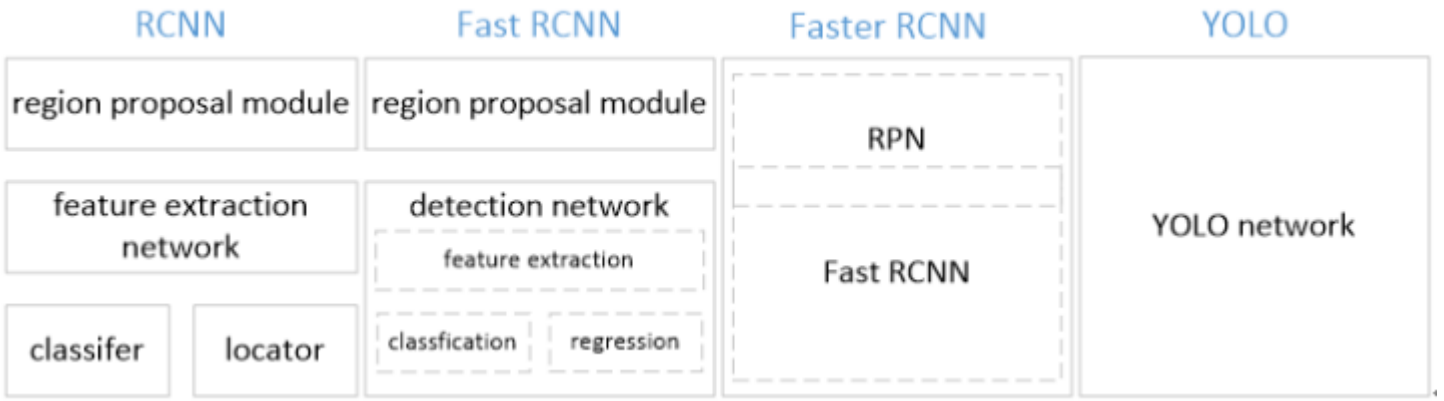


A picture containing text, road, outdoor, street

Description automatically generated

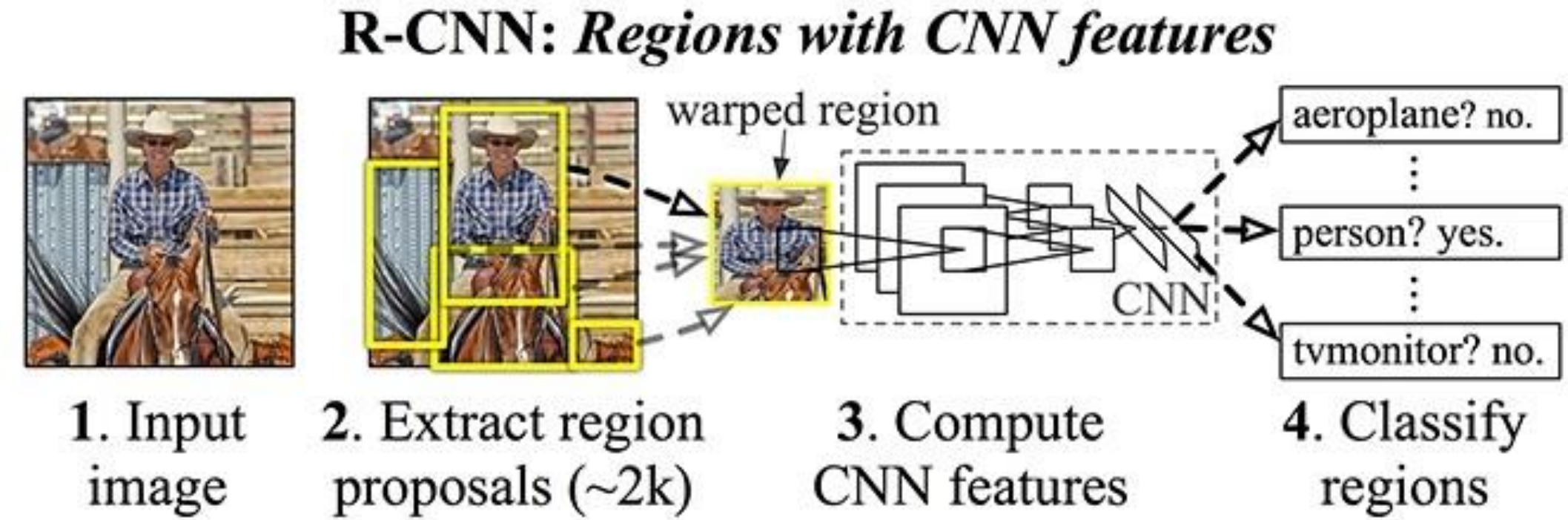
Two-stage vs One-stage detectors

- Two-stage detection
  - Finds regions of interest and sends it down the pipeline
  - Uses region proposal network or selective search
  - Coarse to fine
  - Slower
- One-stage detection
  - Only using single deep neural network
  - Faster (real time object detection)
  - Not that accurate



https://github.com/yehengchen/Object-Detection-and-Tracking

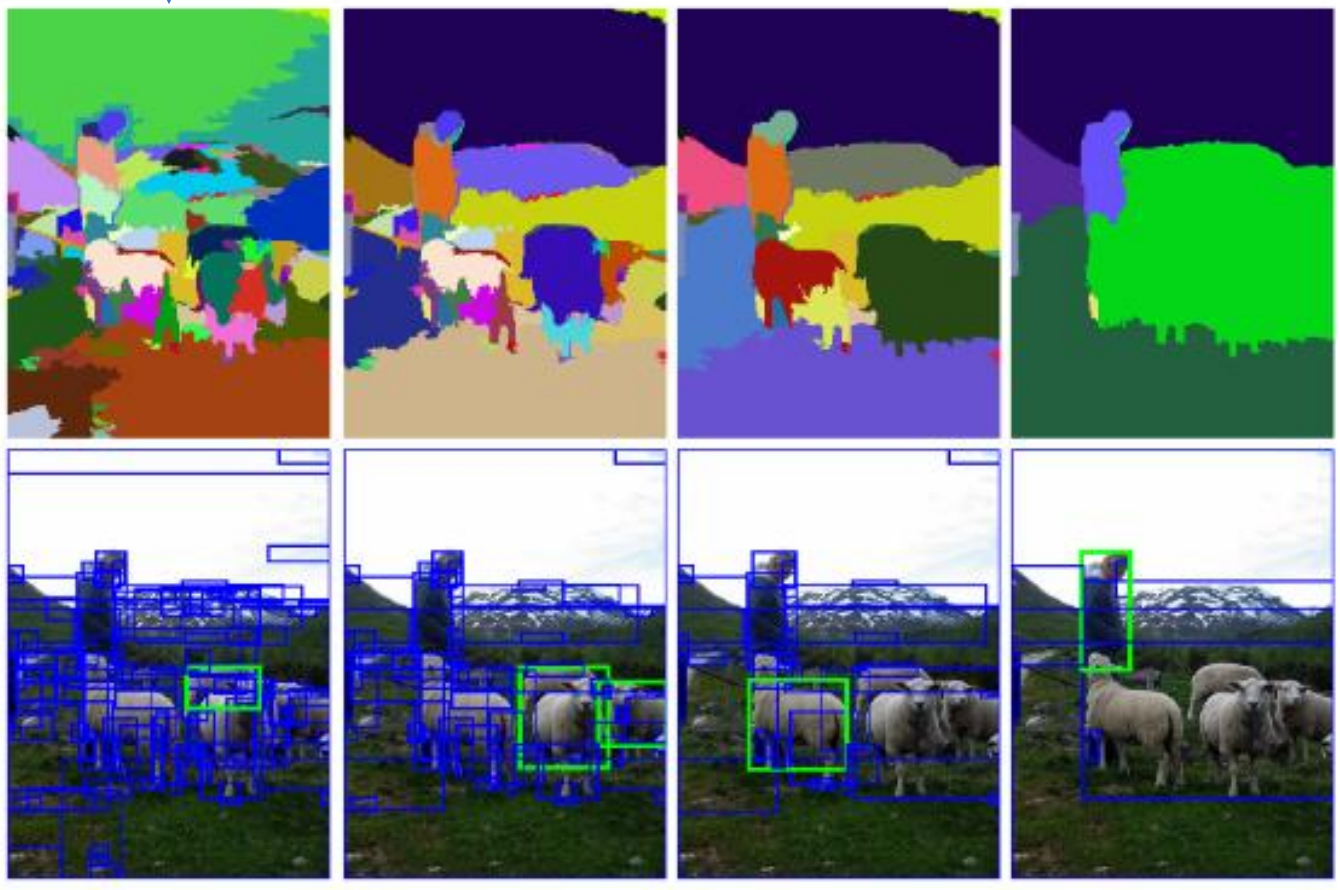
R-CNN (Regions with CNN)



R. Girshick et al., 2014

R-CNN – how to pick good regions

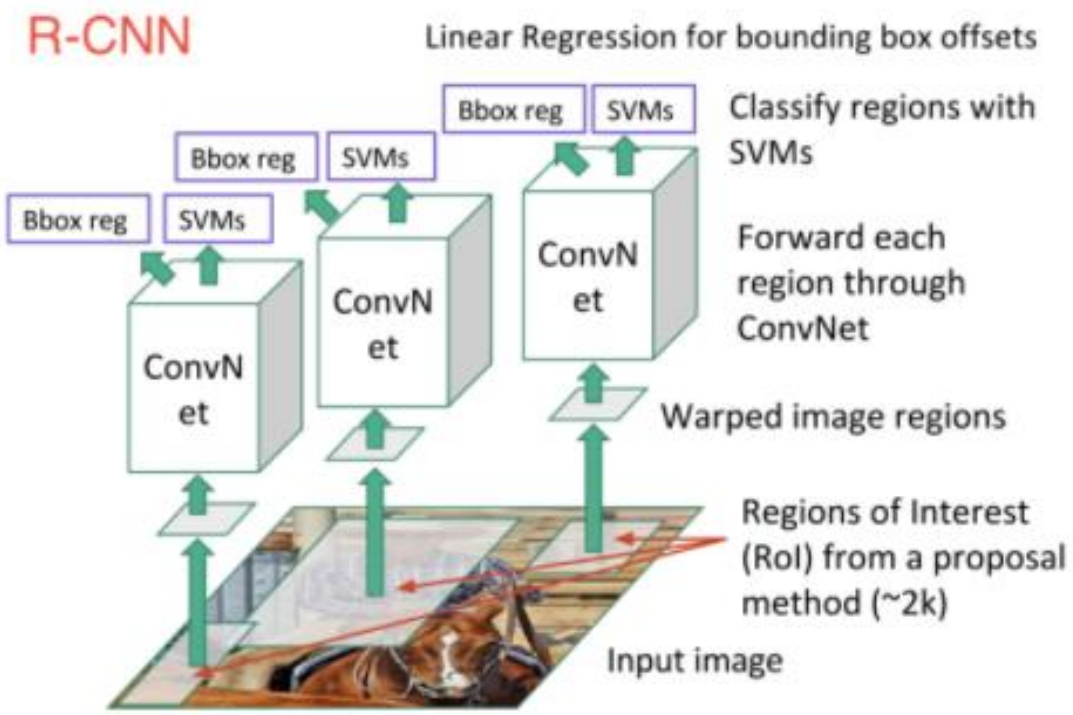
Efficient Graph-Based Image Segmentation, Felzenszwalb et al., 2004



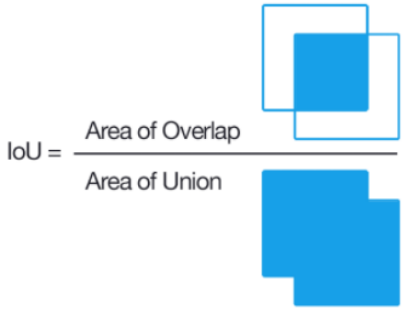
Selective Search - Van de Sande et al., 2011

$$s(r_i, r_j) = a_1 s_{colour}(r_i, r_j) + a_2 s_{texture}(r_i, r_j) + a_3 s_{size}(r_i, r_j) + a_4 s_{fill}(r_i, r_j),$$

R-CNN architecture



- 2k proposals from picture
- Propagating to CNN and obtaining features
- We use SVM to obtain scores for each class
- If some threshold of IoU is exceeded for bounding boxes keep one with higher score (Non max suppression)
- Regressing the bounding boxes

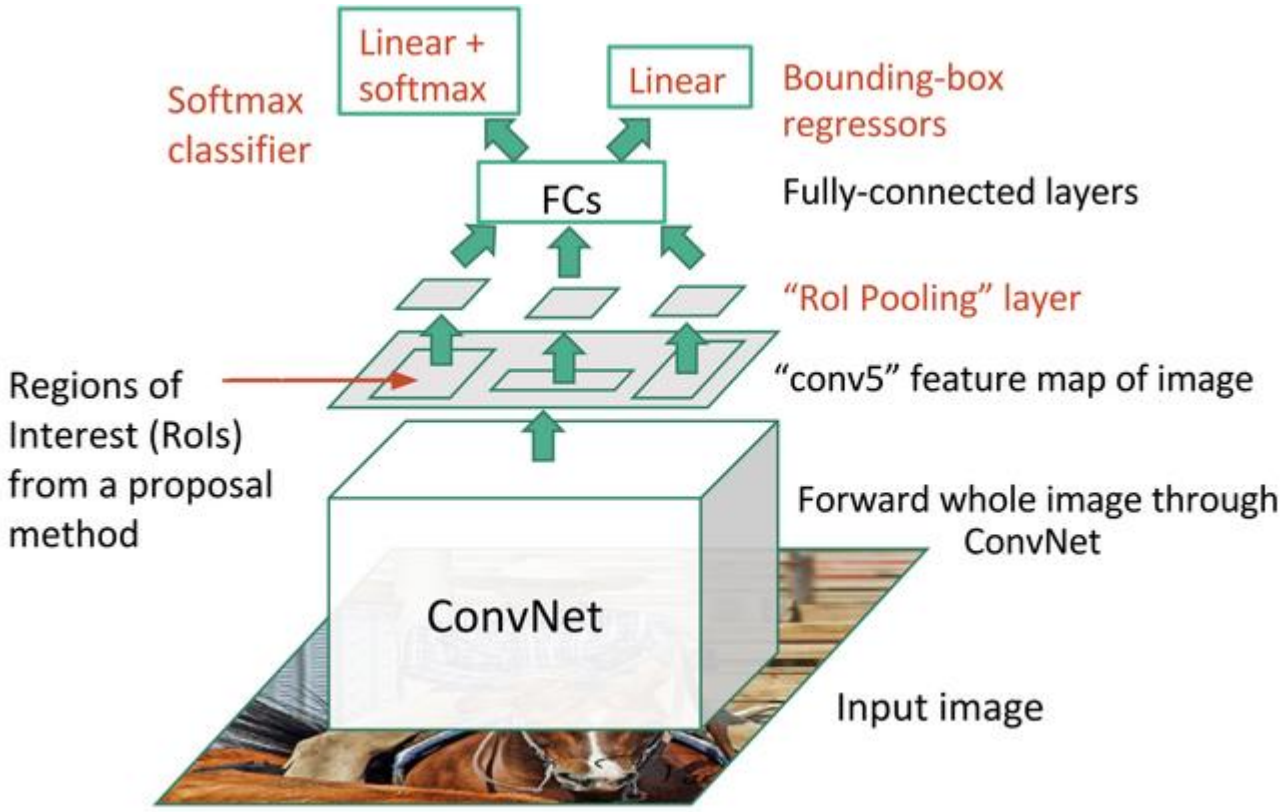


R. Girshick et al., 2014

Bottlenecks

- 40-50 sec. per photo
- Selective search is given “as is”
- Long training hours

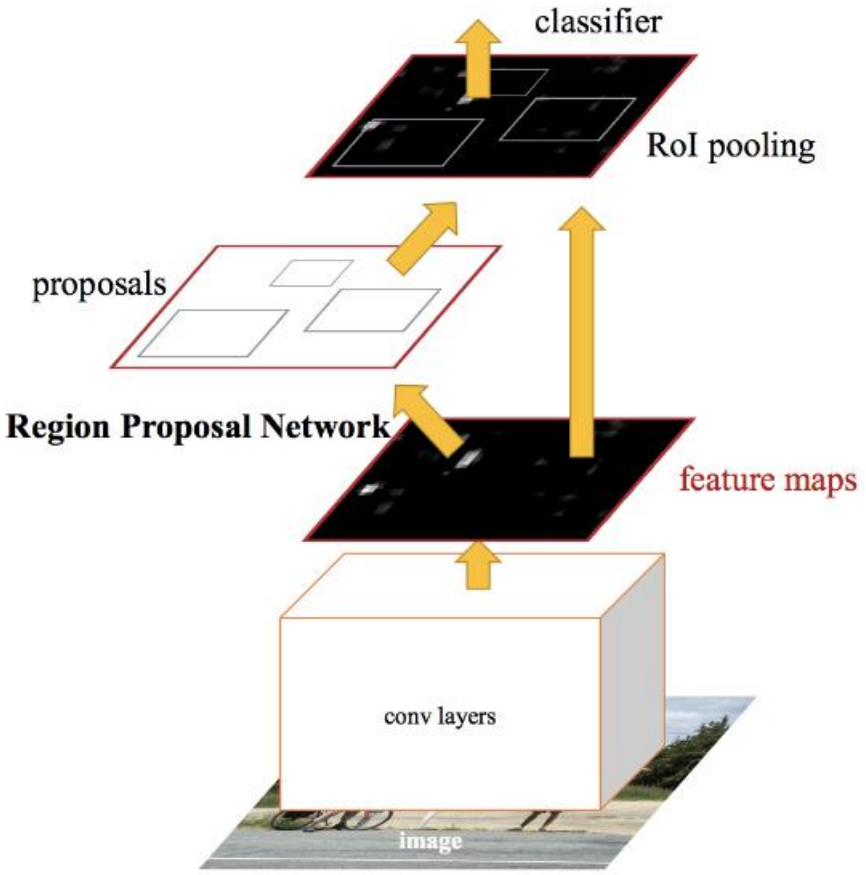
Fast R-CNN architecture



- Selective search is proposing boxes
- Features are extracted through ConvNet
- Then feature maps are transformed into lower, fixed size dimensions (RoI Pooling)
- Such feature vectors are input to FC layers
- Classification is now used with softmax

R. Girshick et al., 2015

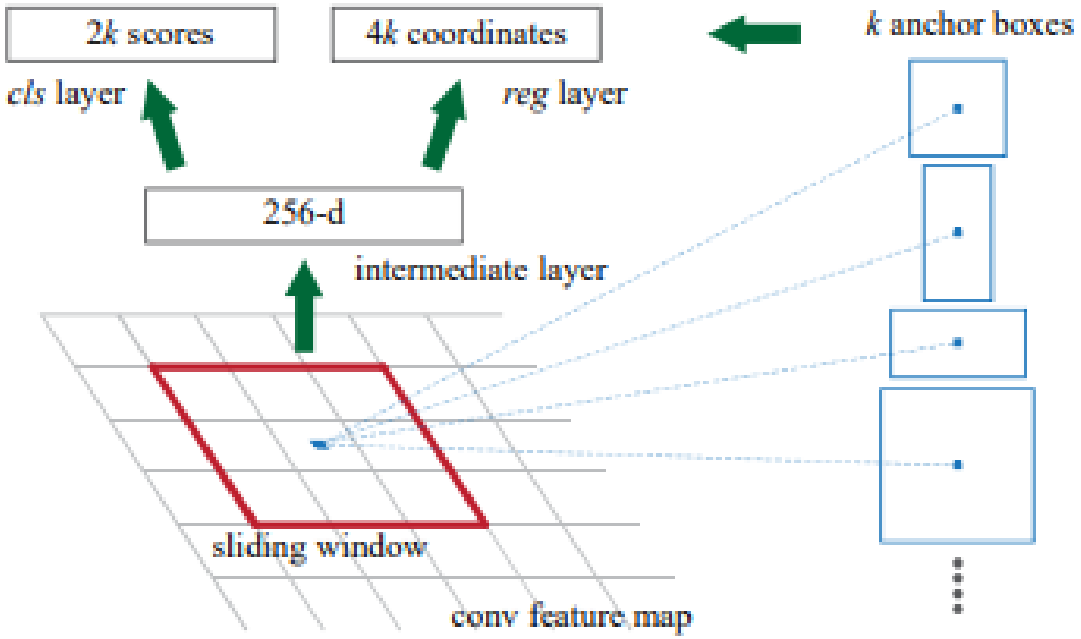
Faster R-CNN



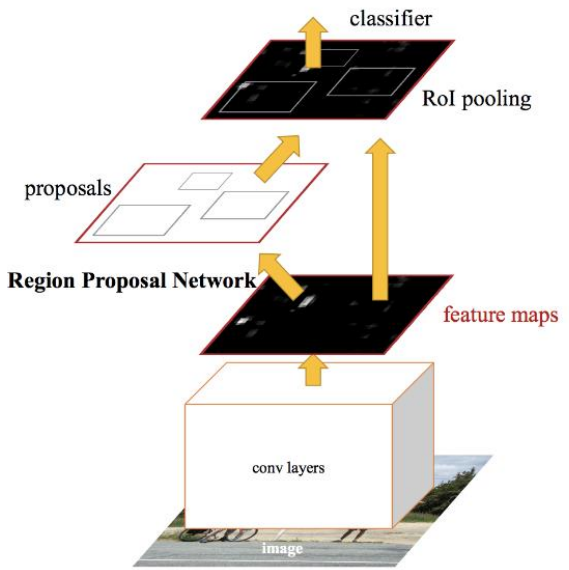
- 2 modules: RPN + Fast R-CNN
- RPN tells Fast R-CNN where to look
- RPN outputs proposals with ‘objectness’ score

S. Ren et al., 2015 (but with R. Girshick as co-author)

How RPN works



- Sliding window on the feature map
- Different anchor boxes (9)
- Mapping to lower dimensional vector
- RPN is trainable



Faster R-CNN adventages & drawbacks

- As this is one trainable system and feature maps are shared there is an improvement in performance and time
- 0.2 s per image
- Still not enough for Real time object detection
- 2 trainable nets instead of one

Image pyramid

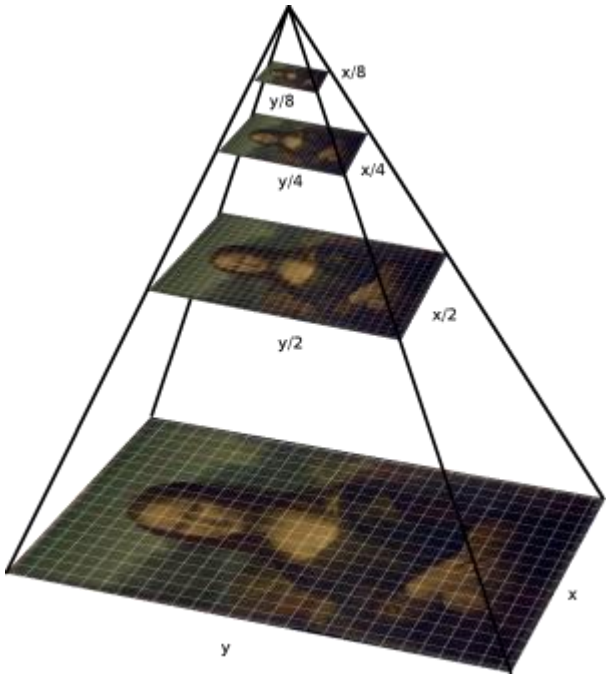
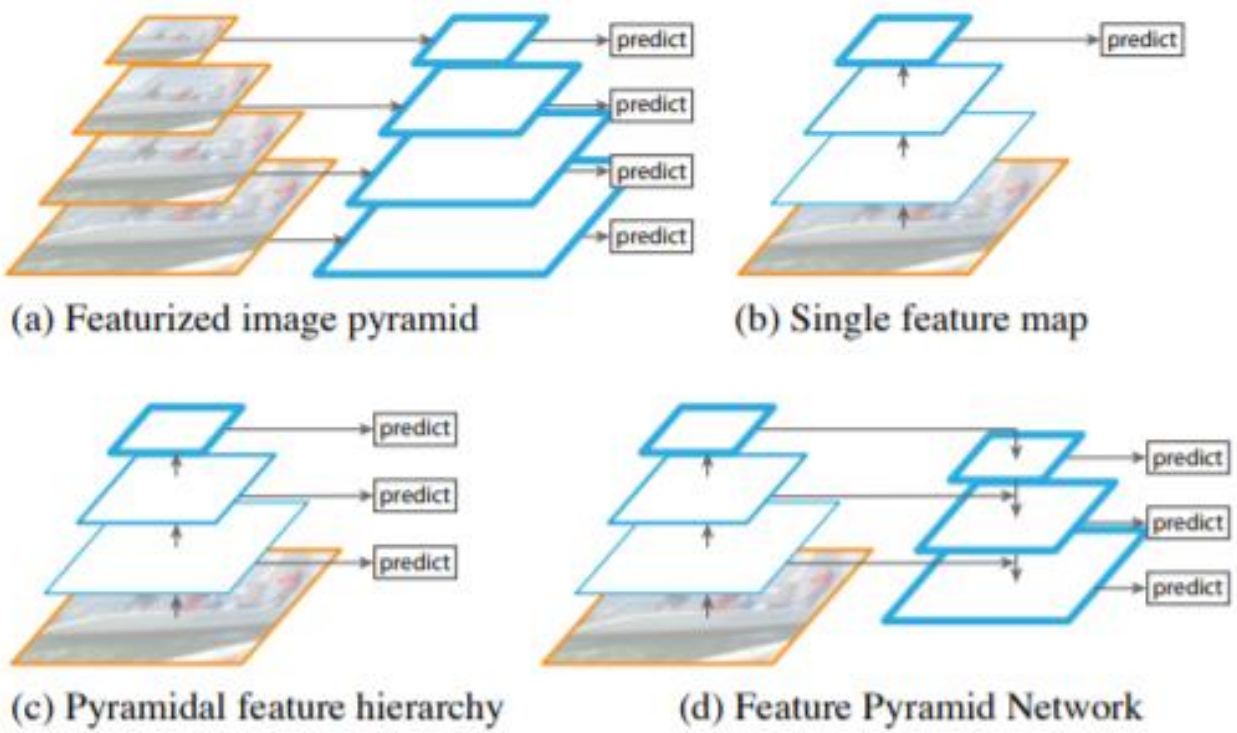


Image source

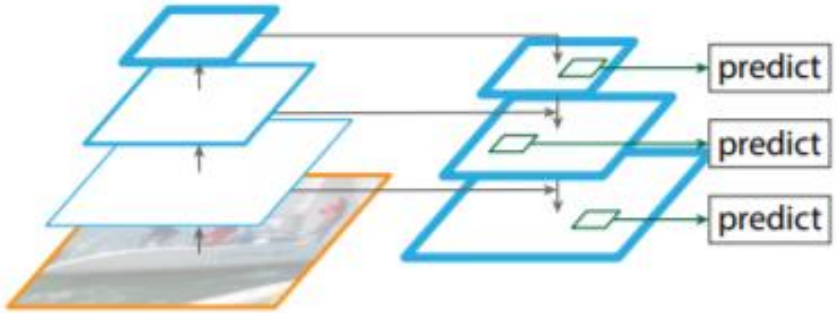
- Image pyramids used by networks improve accuracy
- They tend to decrease speed and memory efficiency

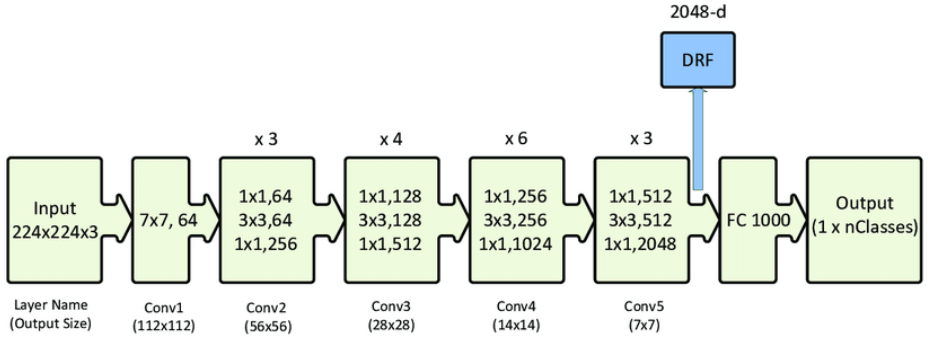
FPN



- A – Running CNN on images of different scale
- B – classical CNN
- C – A + B
- D – what authors suggest

Faster R-CNN + FPN

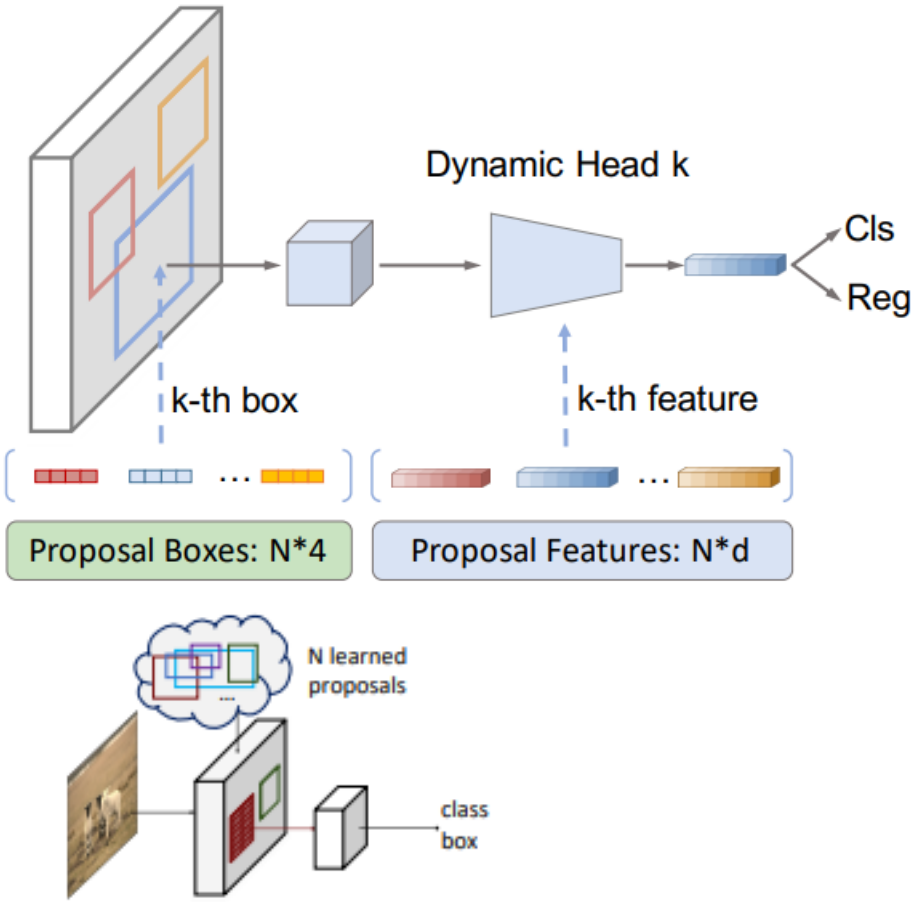




- Each feature map after Conv block
- Upsampling and merging feature maps (element-wise)
- Predictions at each level
- After obtaining feature map(s) steps like in Faster R-CNN

T. Lin et al., 2017 (with R. Girshick as co-author)

Sparse R-CNN



- ~ 100 learnable boxes per image
- ~ 100 learnable proposal features
- One feature per box
- Features may encode for example pose, shape etc.
- No non-maximum suppression

P. Sun et al., 2021

Additional refereces and further reads

- http://www.robots.ox.ac.uk/~tvg/publications/talks/fast-rcnn-slides.pdf
- https://arxiv.org/pdf/1905.05055.pdf
- https://dudeperf3ct.github.io/object/detection/2019/01/07/Mystery-of-Object-Detection/
- https://www.youtube.com/playlist?list=PL3FW7Lu3i5JvHM8ljYj-zLfQRF3EO8sYv
- https://arxiv.org/pdf/1905.05055.pdf