

Deep Object Detection

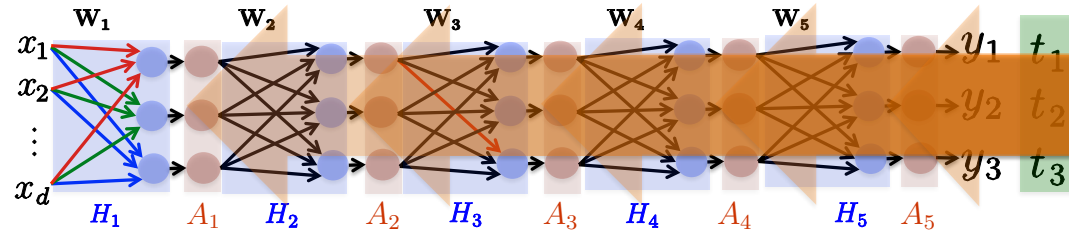
Ali Farhadi

Mohammad Rastegari

CSE 576

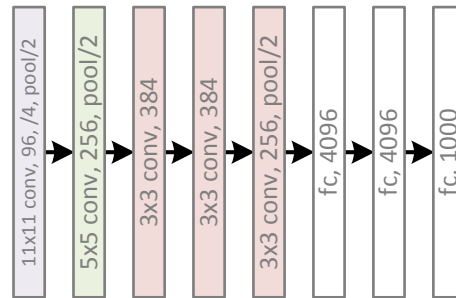
So Far

- Backpropagation

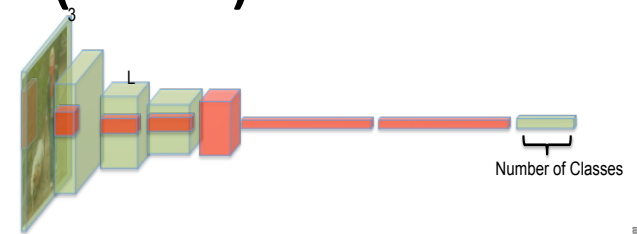


- Convolutional Neural Networks(CNN)

- AlexNet



- Deeper Architectures

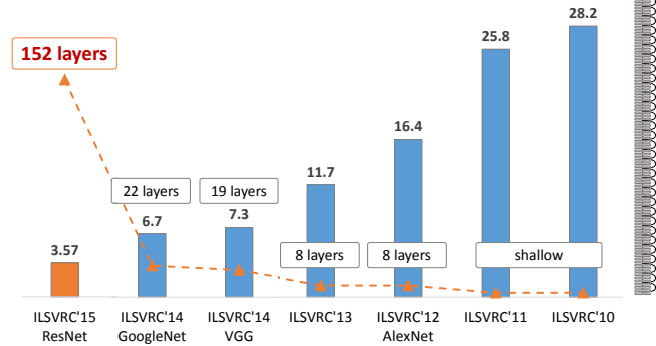


Revolution of Depth

AlexNet, 8 layers
(ILSVRC 2012)

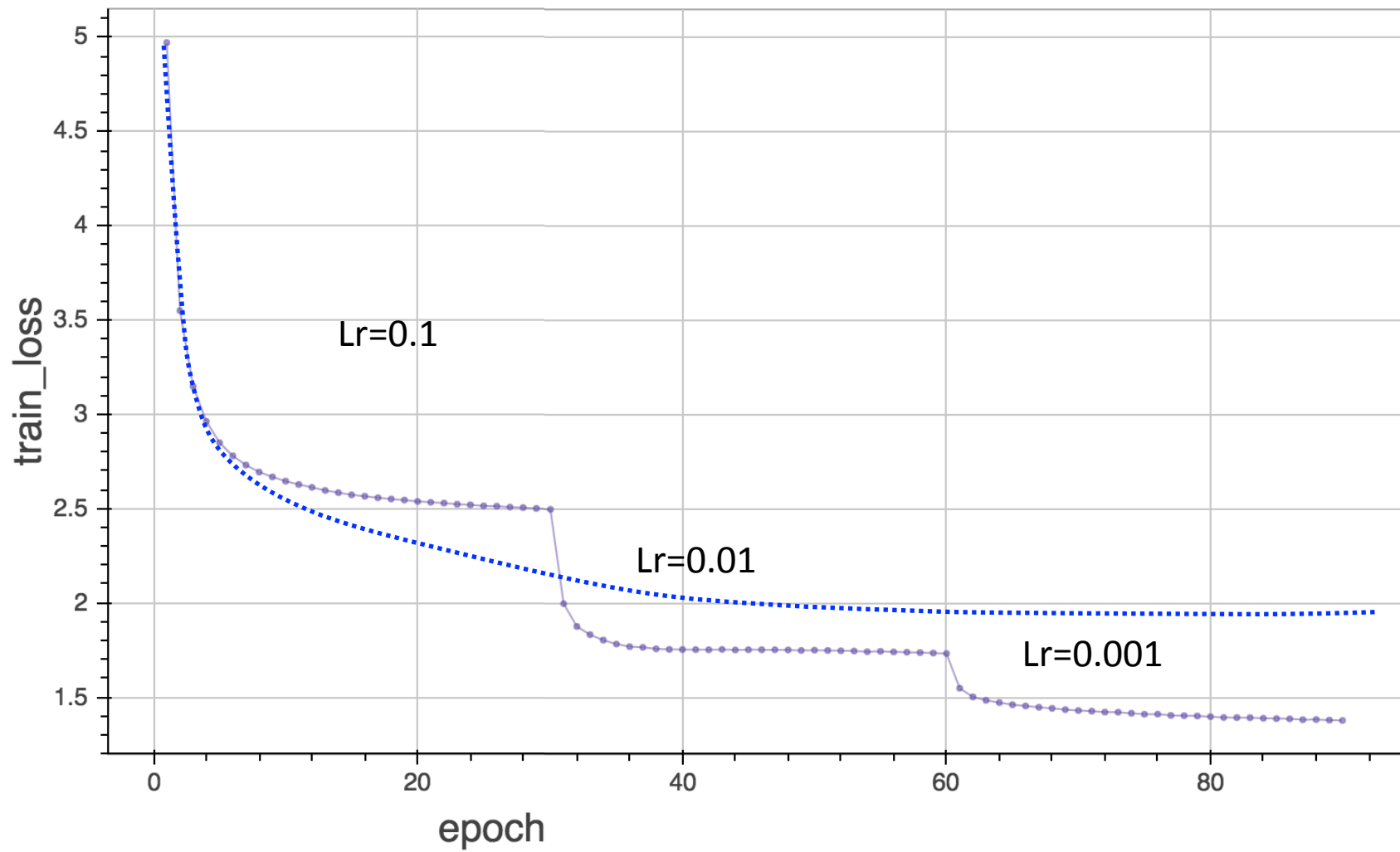
VGG, 19 layers
(ILSVRC 2014)

ResNet, 152 layers
(ILSVRC 2015)

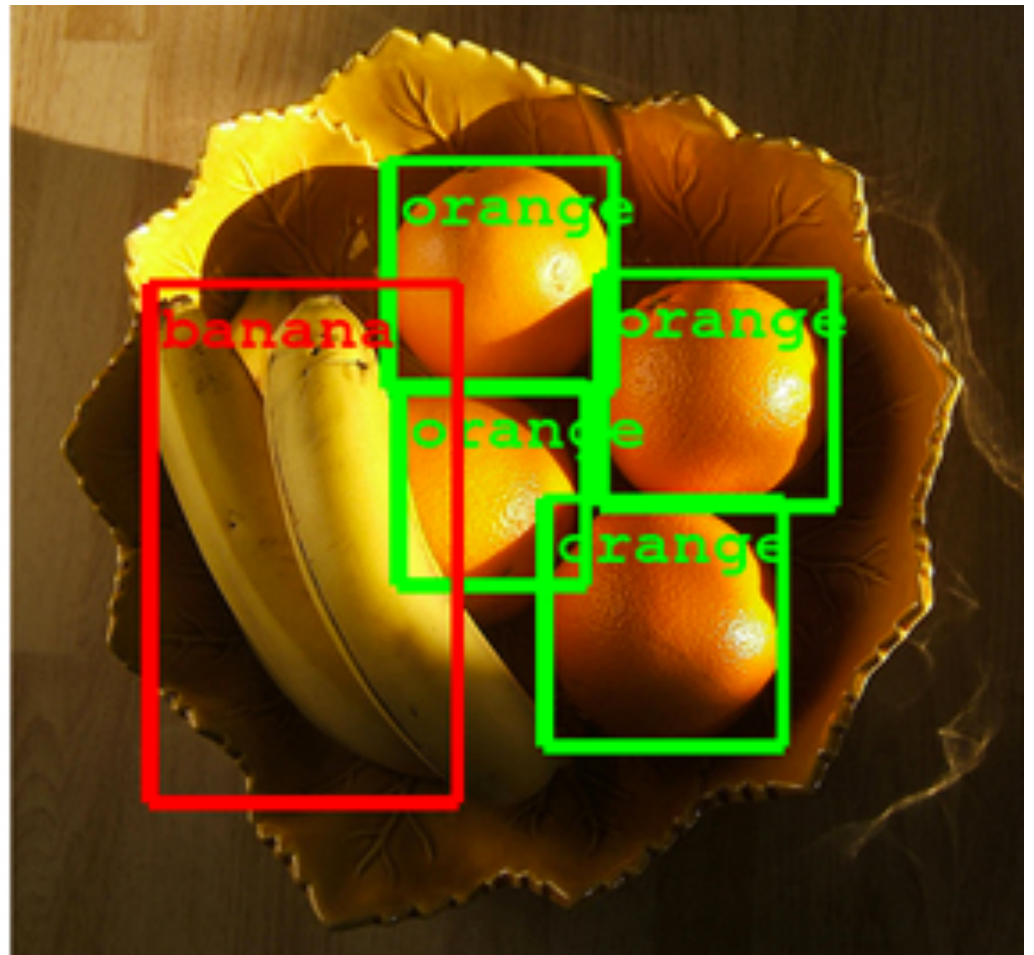
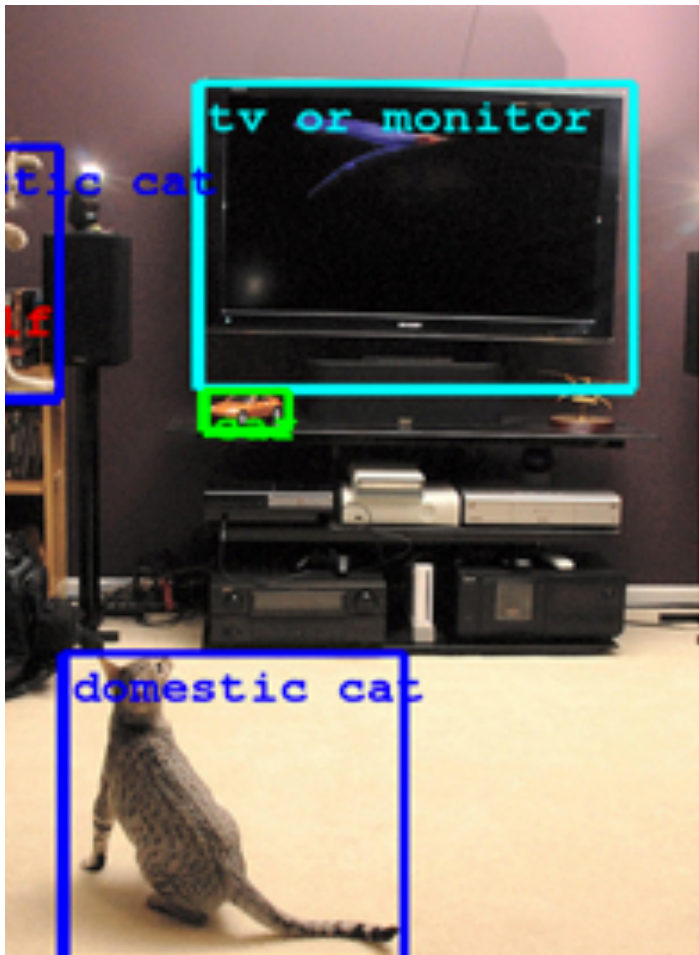


Deep Learning Practical Tips

- Use off-the-shelf architectures
- Verify the correctness of your network by training over a single batch.
 - Overfit : Good to go!
 - Did not converge : Something is wrong with forward/backward functions or data!
- Use a proper learning rate regime.



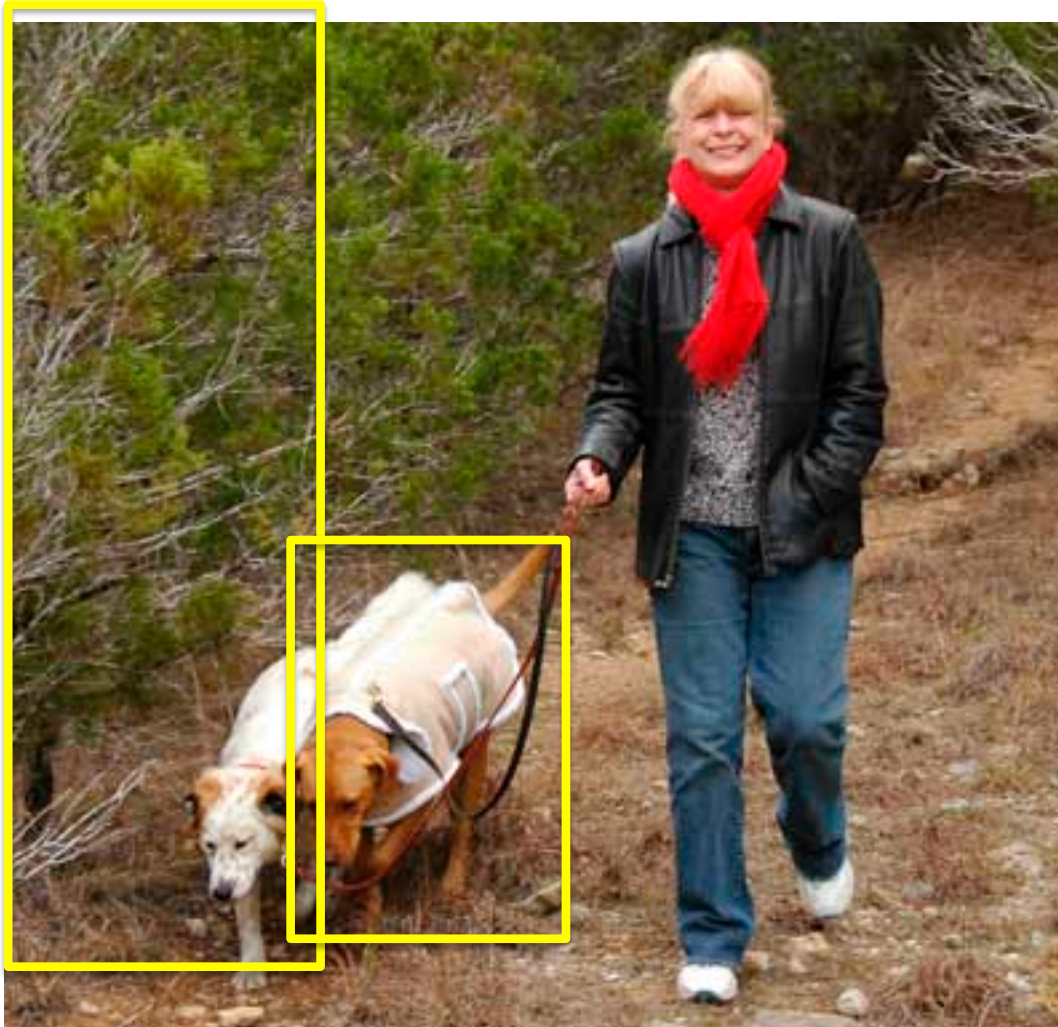
Object Detection



Sliding Window



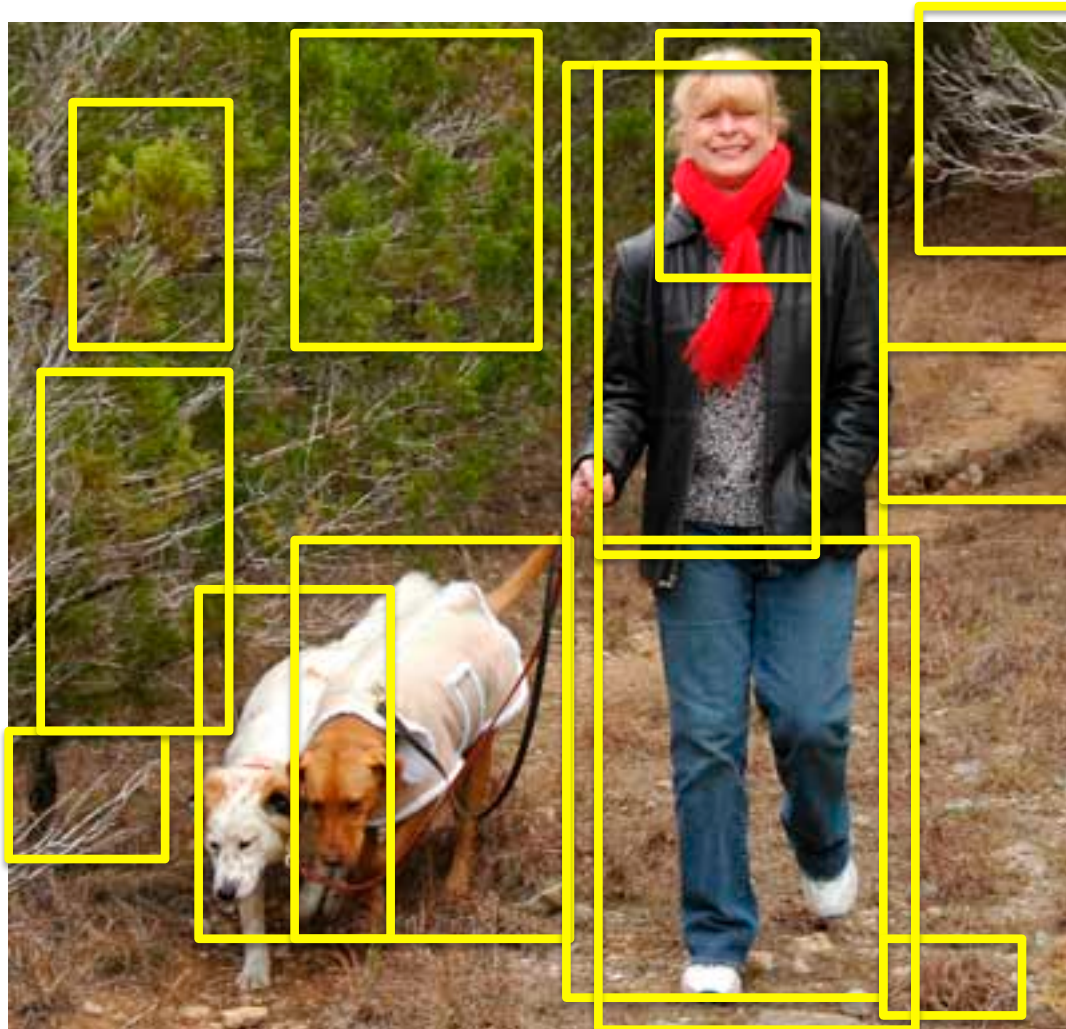
Sliding Window



Sliding Window



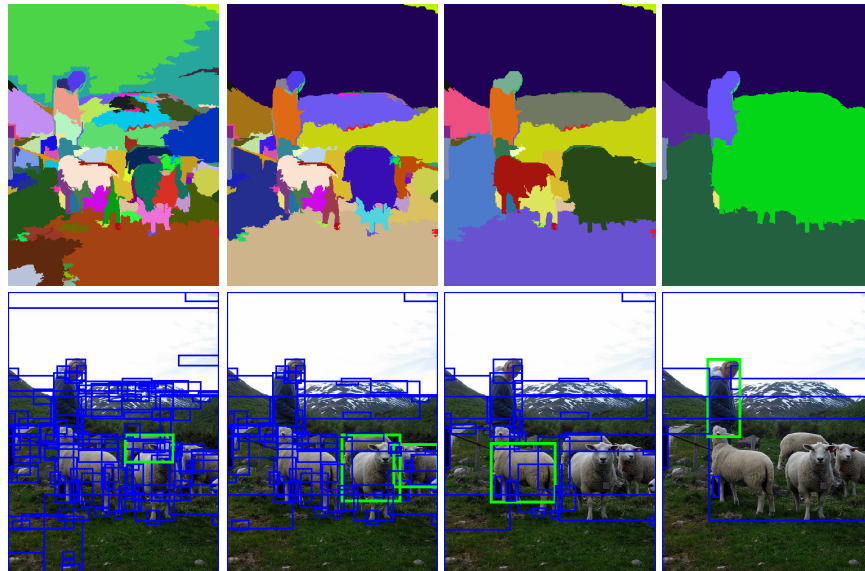
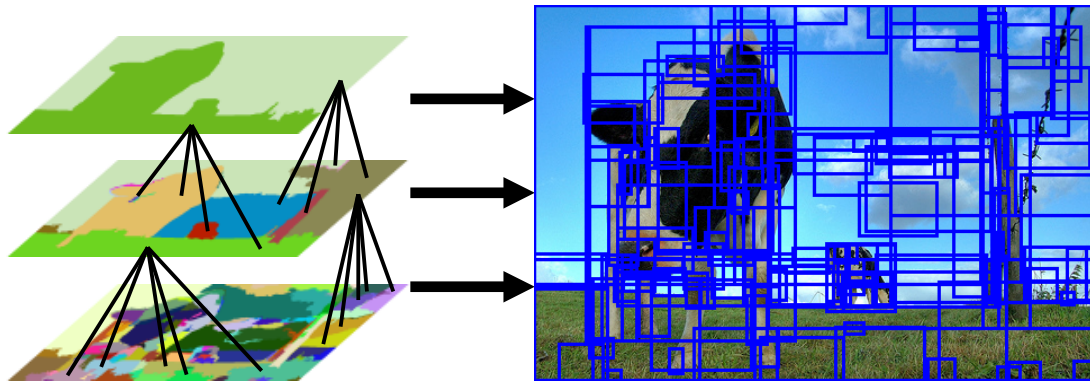
Object Proposal



Selective Search

Uijlings, Jasper RR, et al. "Selective search for object recognition."
International journal of computer vision (2013).

Object hypotheses



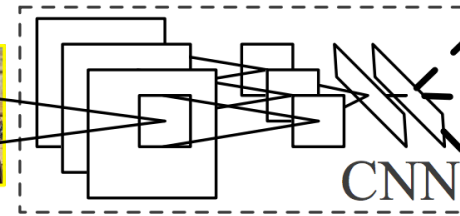
Region-Based CNN (R-CNN)



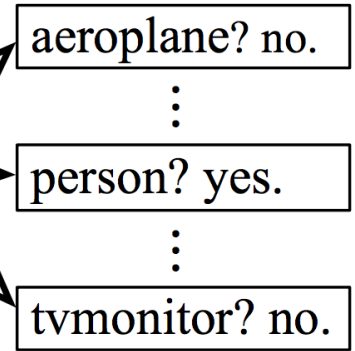
Input
image



Extract region
proposals (~2k / image)

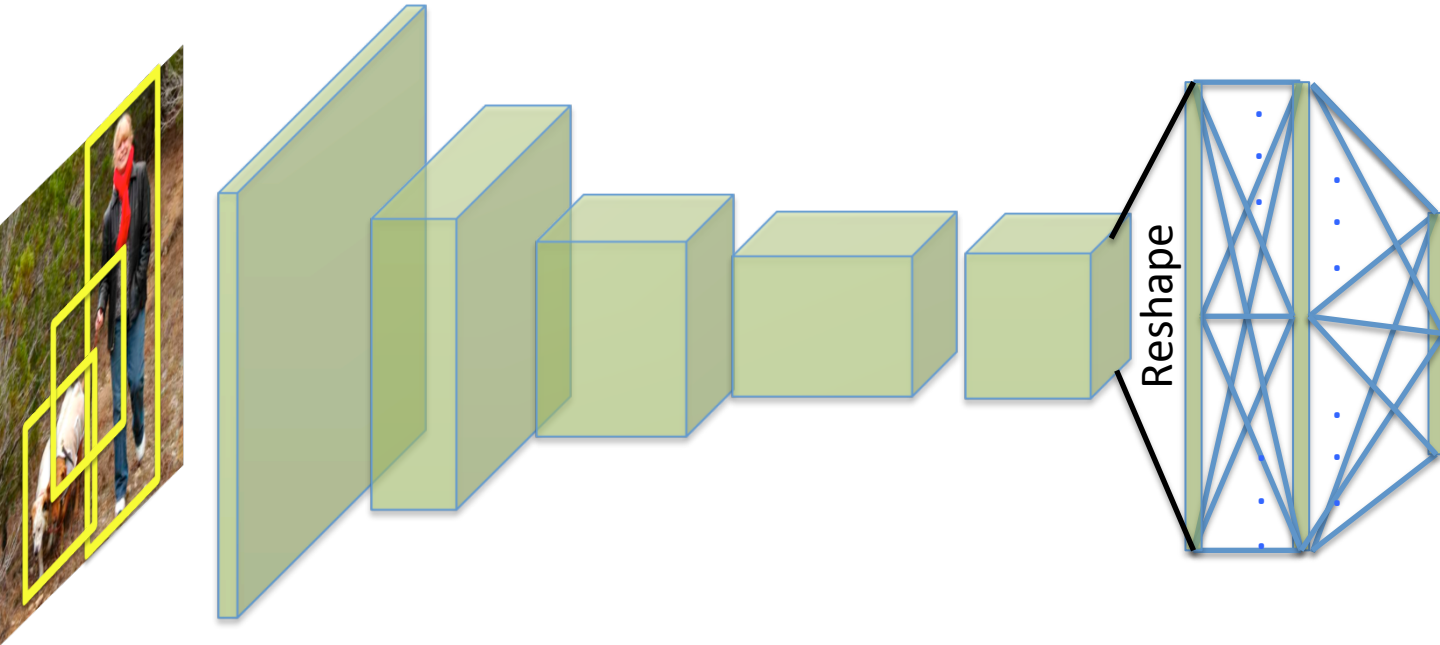


Compute CNN
features

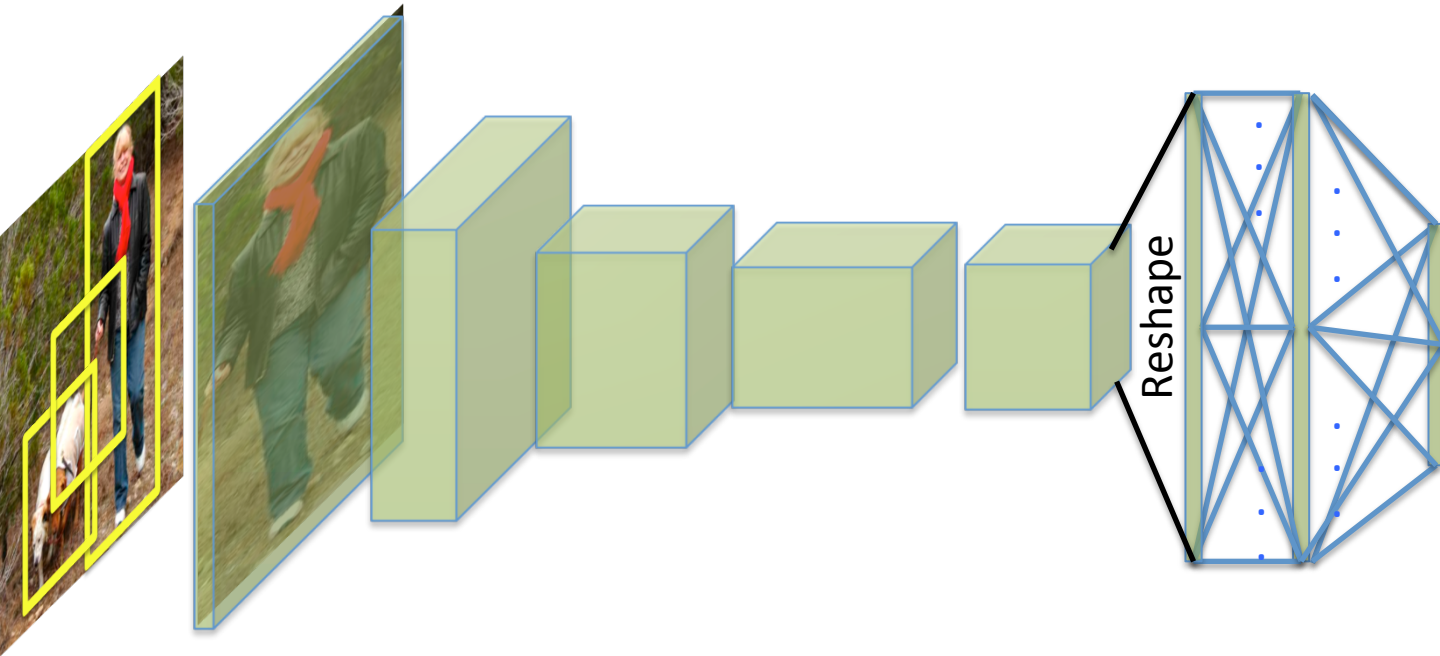


Classify regions
(linear SVM)

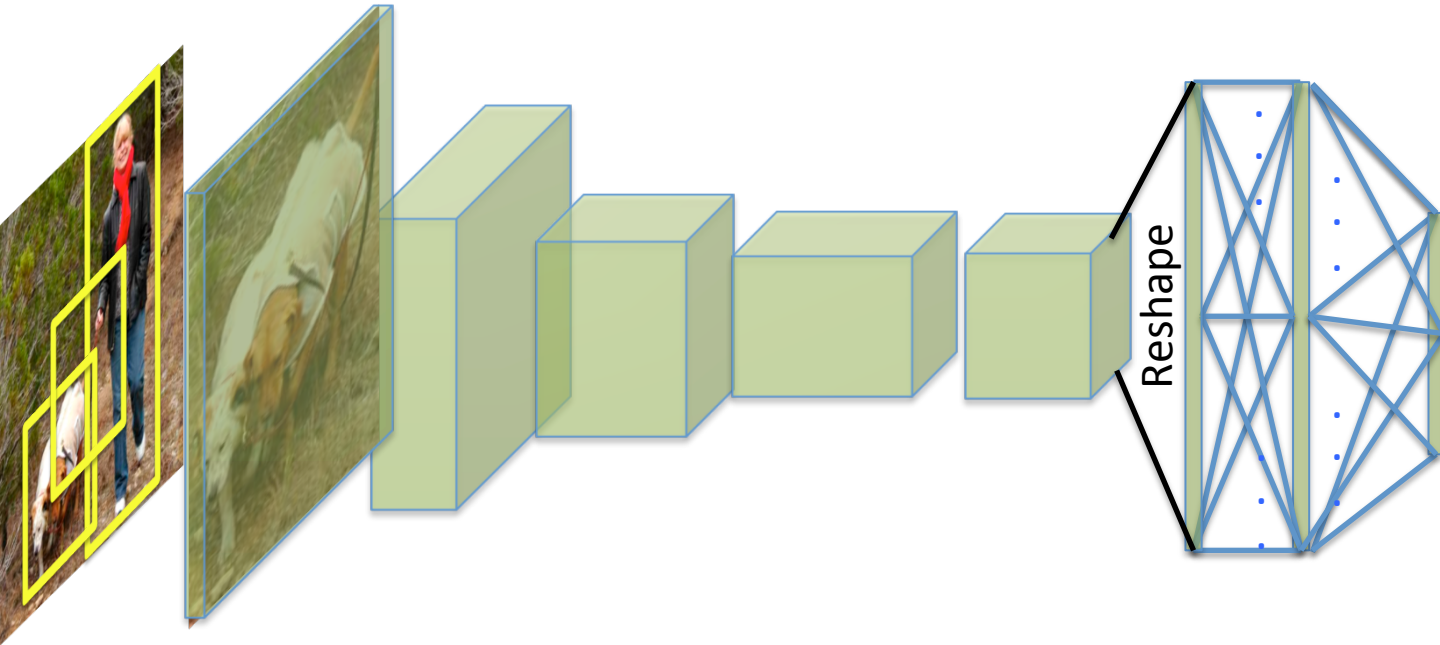
Object Detection by R-CNN



Object Detection by R-CNN

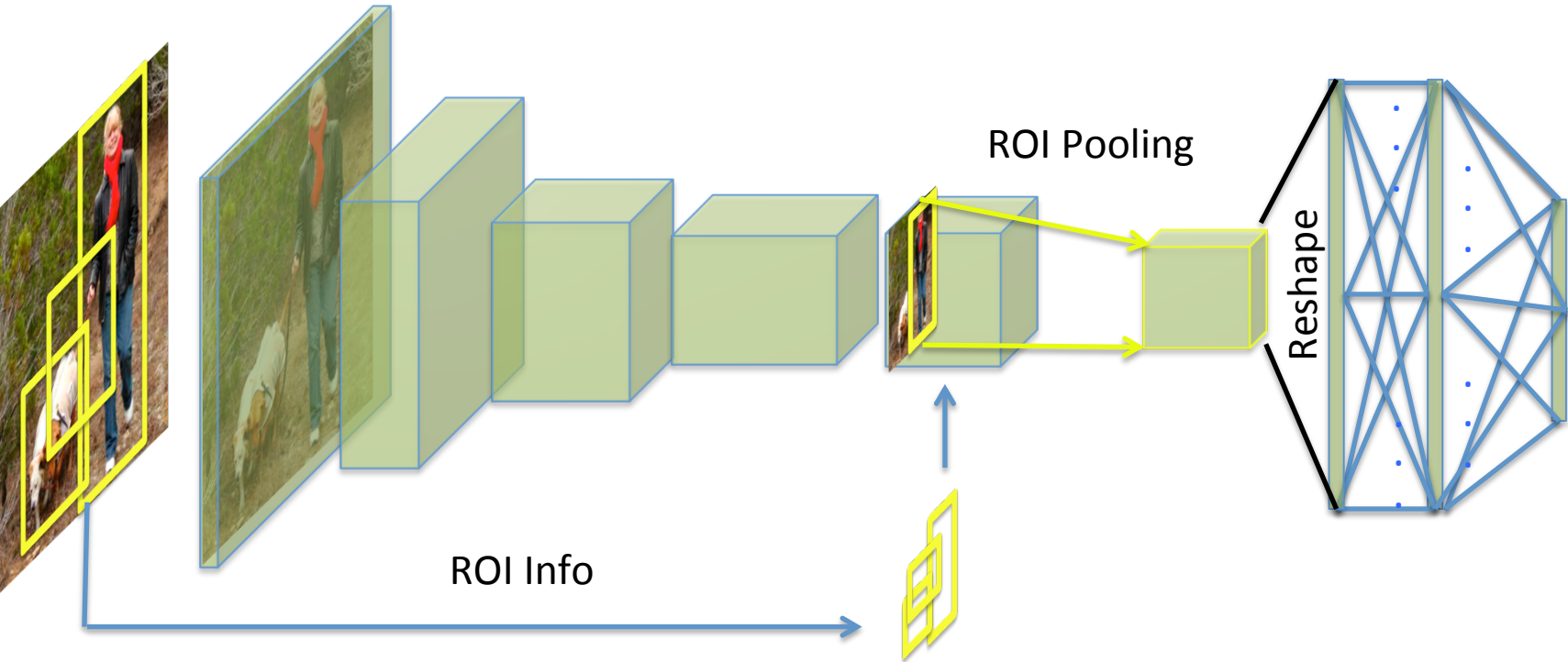


Object Detection by R-CNN

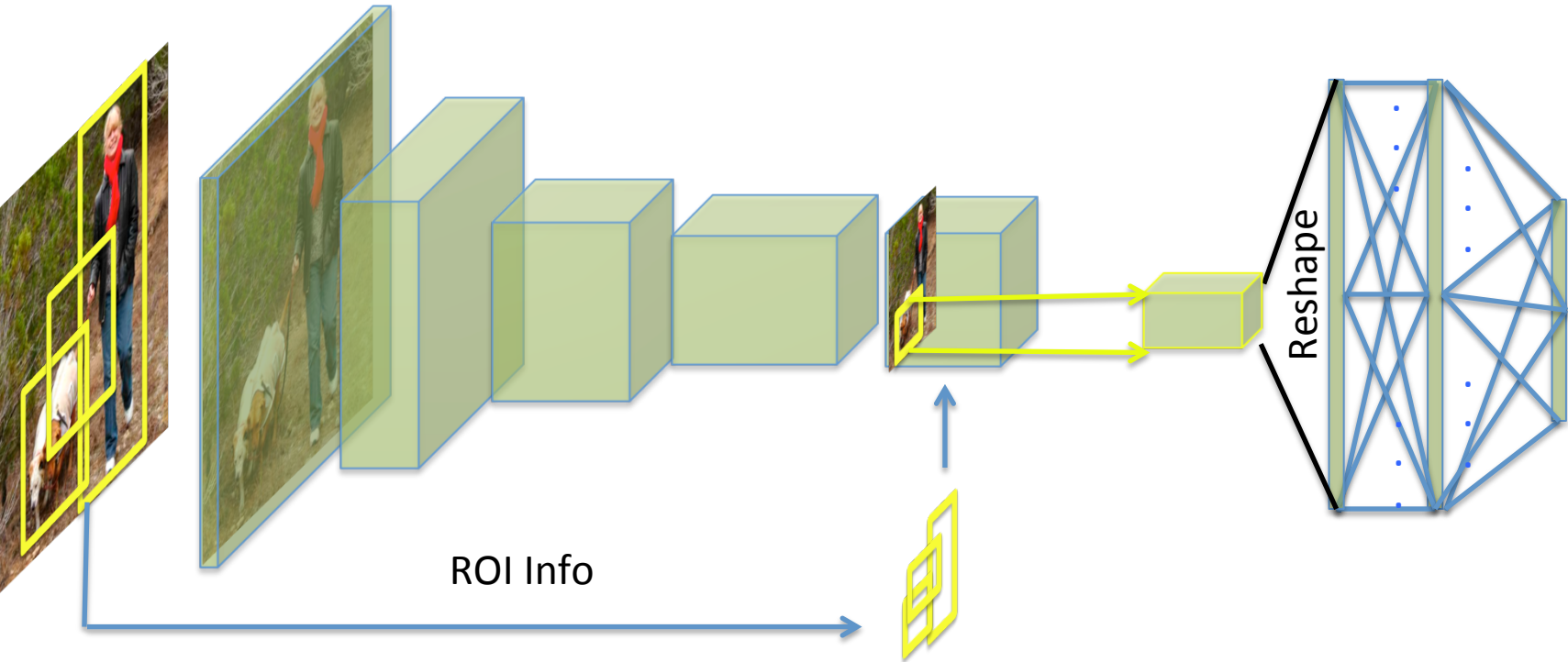


- Depending on region proposal
- Need to apply CNN $\sim 2K$ times per image

Fast R-CNN



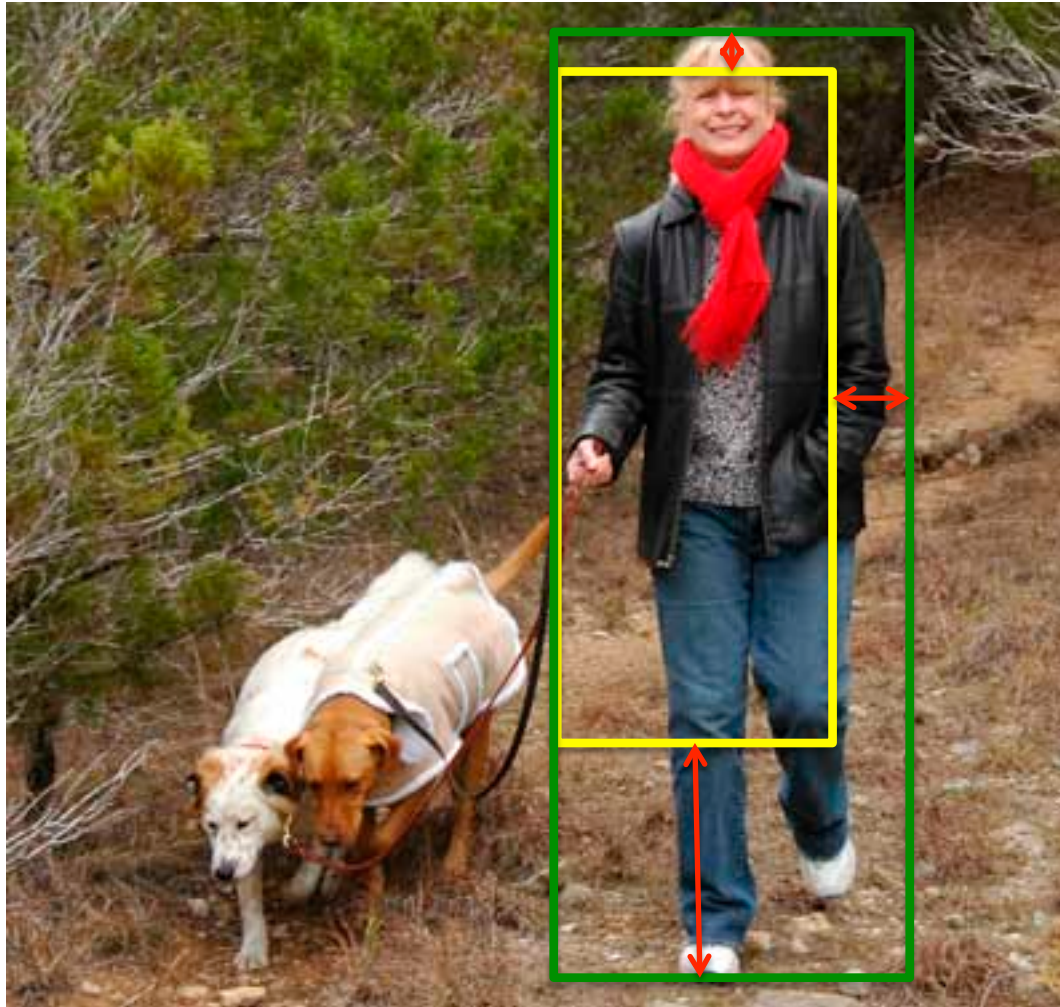
Fast R-CNN



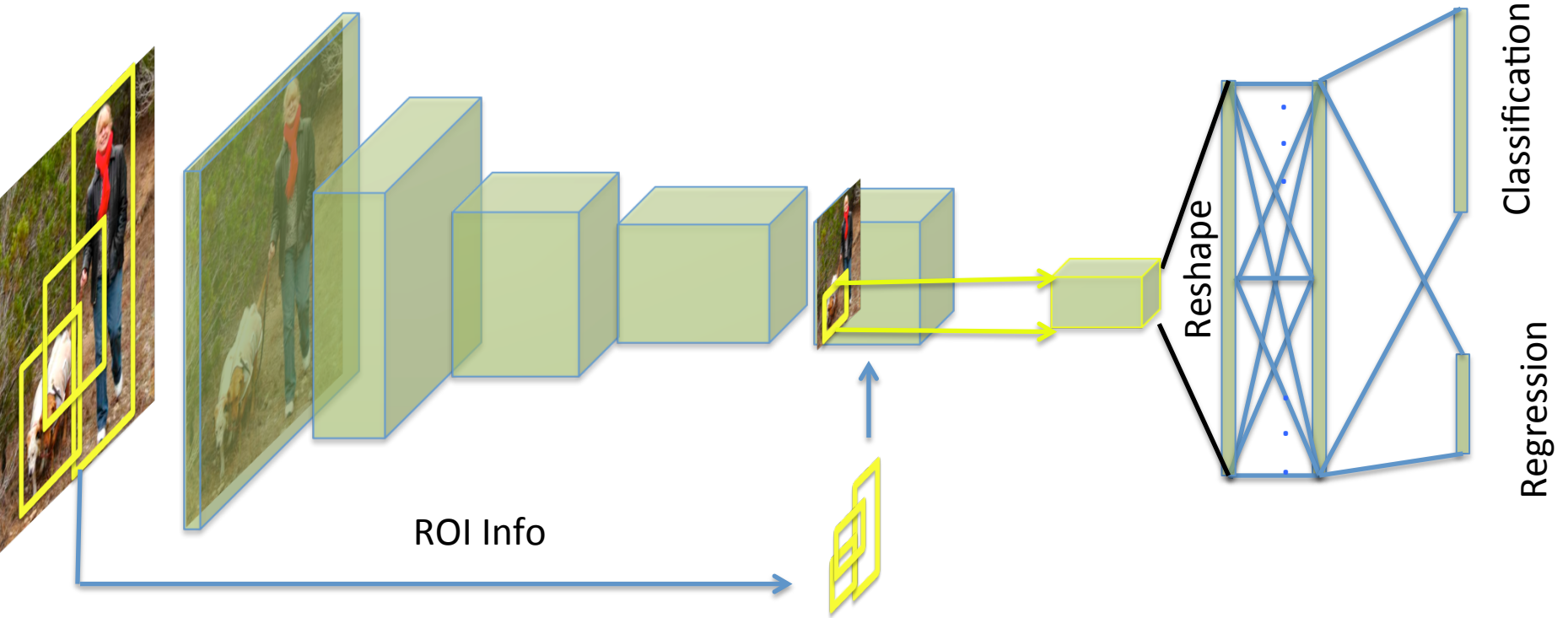
Bounding Box Regression



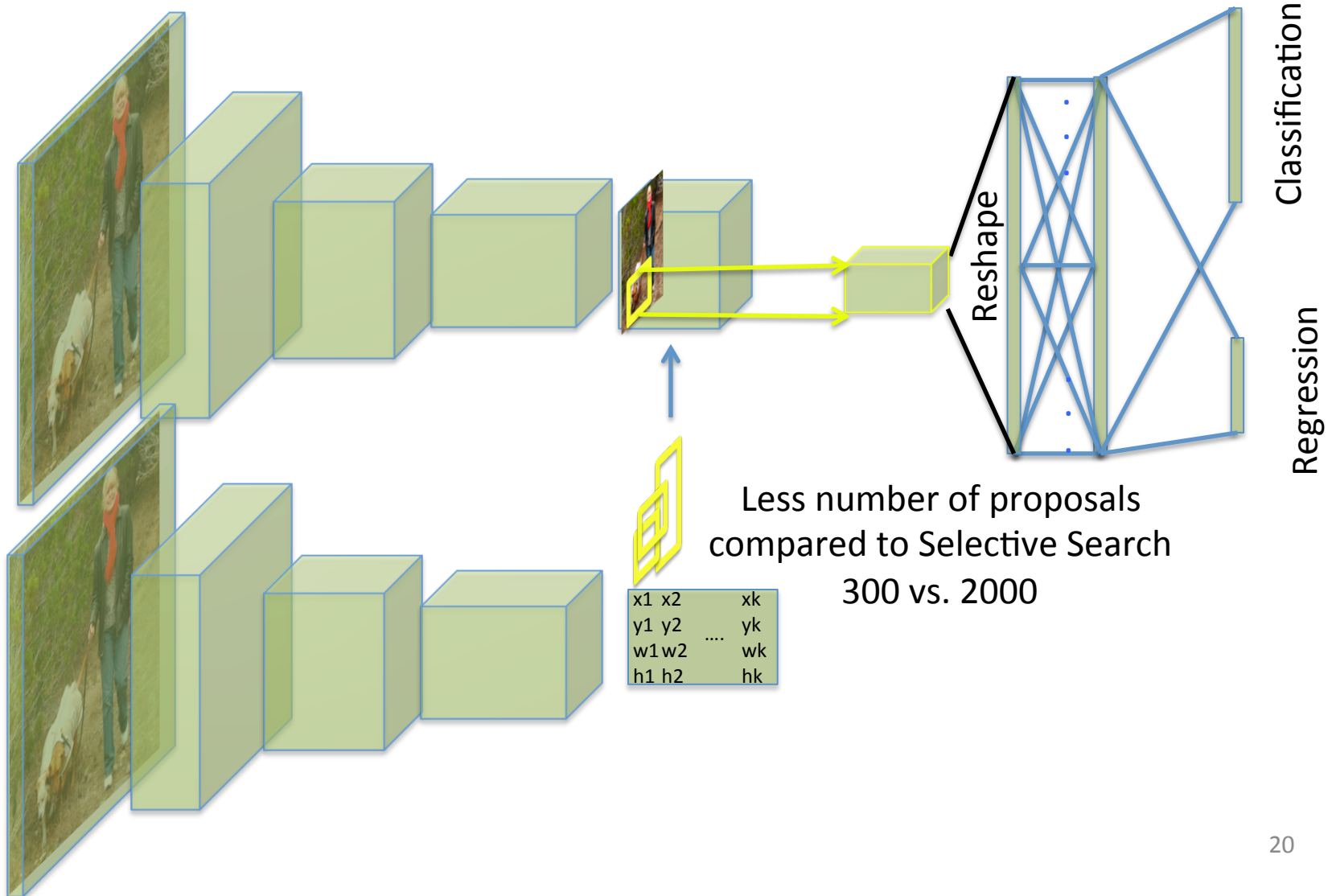
Bounding Box Regression



Bbox Regression



Faster R-CNN

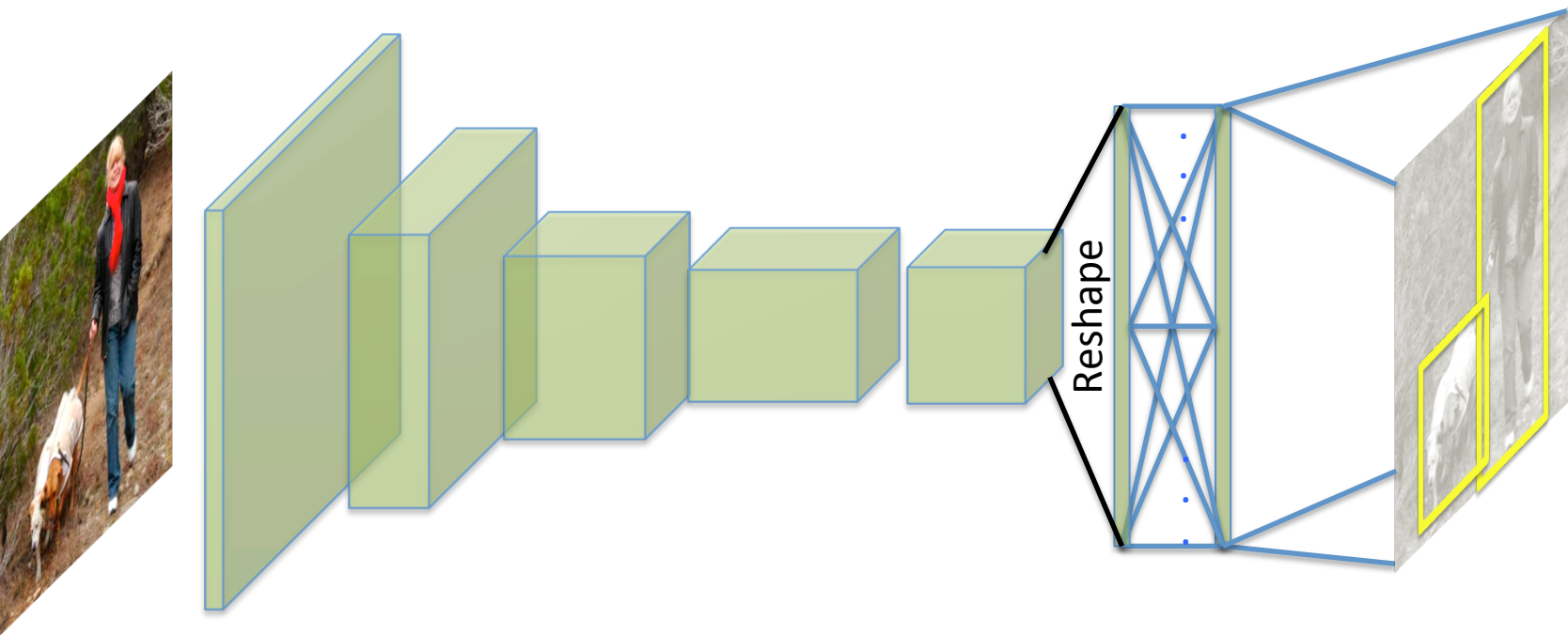


Pascal 2012

Method	mAP	Sec/im
R-CNN	59.2	20
Fast R-CNN	68.4	2
Faster R-CNN	72.1	0.5

Direct Regression

No Proposal



We do not know the number of objects in an image

JOSEPH
REDMON

ROSS
GIRSHICK

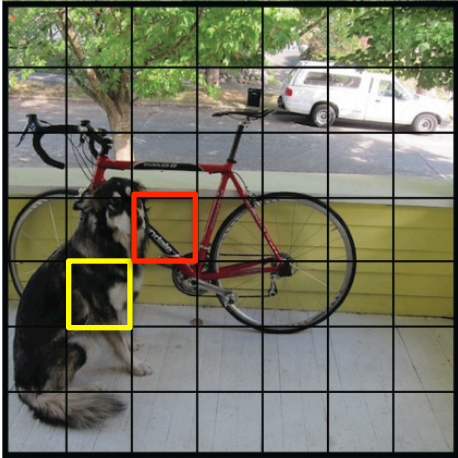
SANTOSH
DIVVALA

ALI
FARHADI

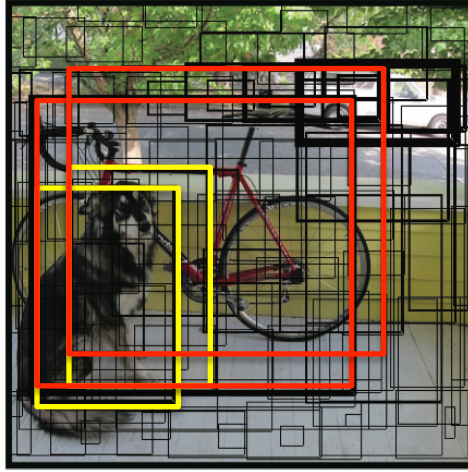
Dog



“YOU ONLY LOOK ONCE”
REAL-TIME
DETECTION



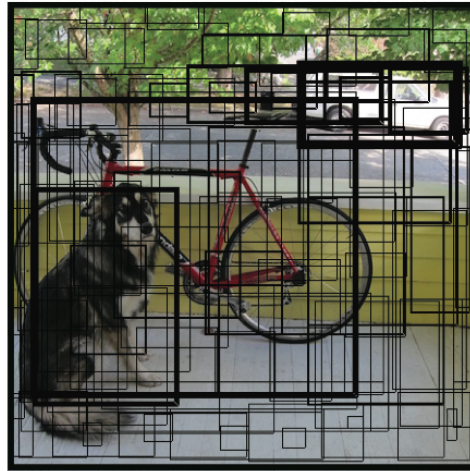
$S \times S$ grid on input



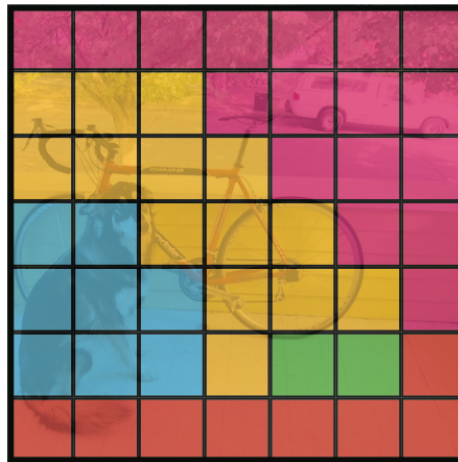
Bounding boxes + confidence



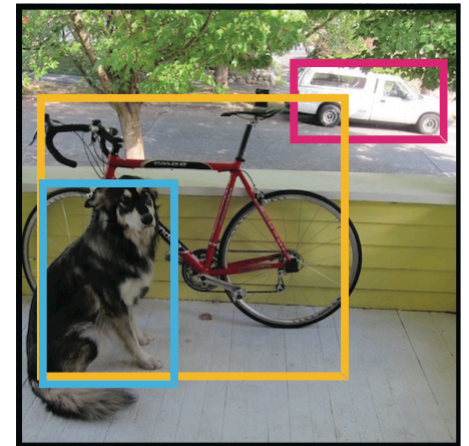
$S \times S$ grid on input



Bounding boxes + confidence

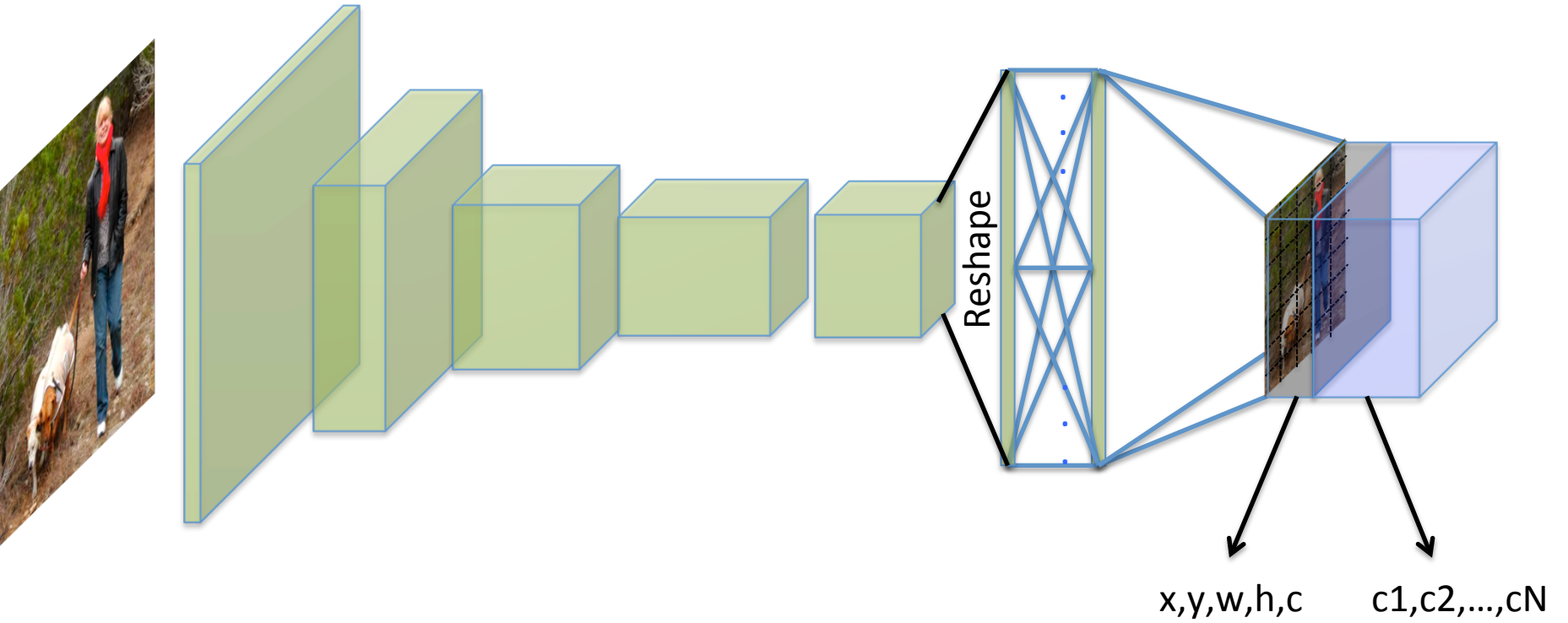


Class probability map



Final detections

YOLO



Pascal 2012

Method	mAP	Sec/im
R-CNN	59.2	20
Fast R-CNN	68.4	2
Faster R-CNN	72.1	0.5
YOLO	57.9	0.02

Source Code

- Fast R-CNN
 - <https://github.com/mahyarnajibi/fast-rcnn-torch>
 - <https://github.com/rbgirshick/fast-rcnn>
- YOLO
 - <https://github.com/pjreddie/darknet/blob/master/src/yolo.c>