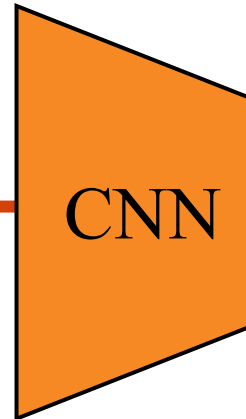# Localization and Detection

CS 281b
02/26/2018

Da Zhang

Image Classification
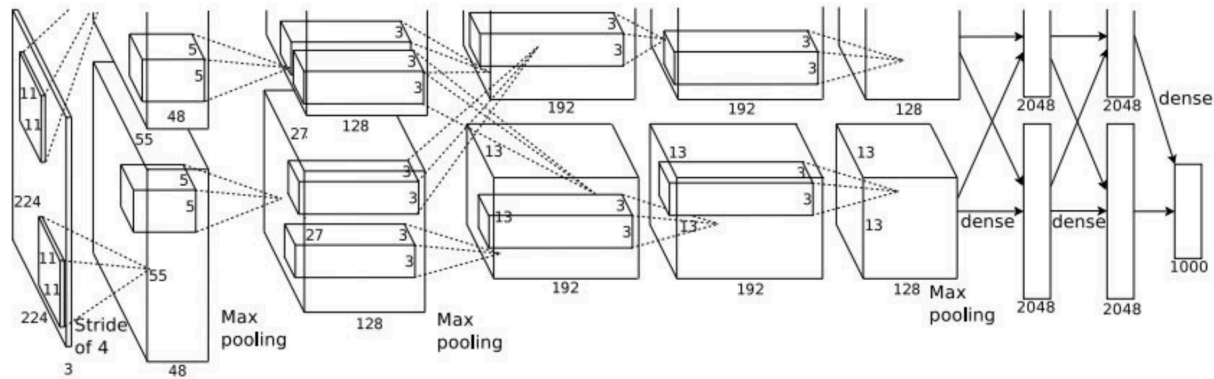

Class Scores
Cat: 0.9
Dog: 0.03
Car: 0.002
…

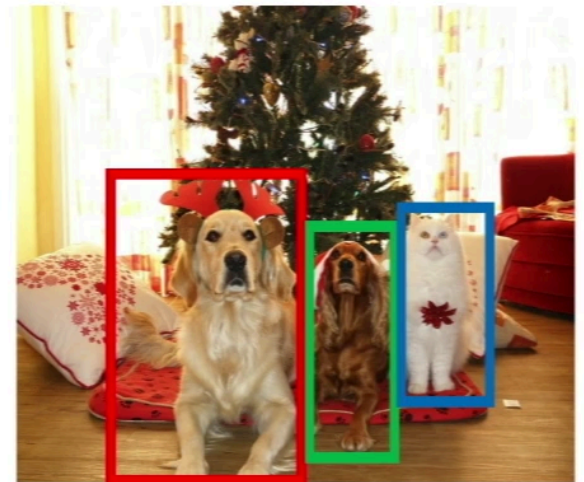# Localization and Detection
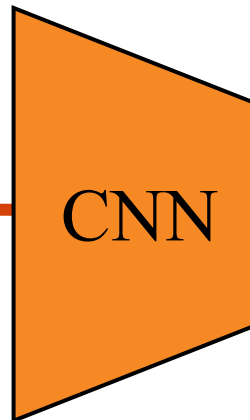
Classification

Localization

Detection

CAT

CAT

DOG, DOG, CAT

Single Object

Multiple Object

Localization

Class Scores
Cat: 0.9
Dog: 0.03
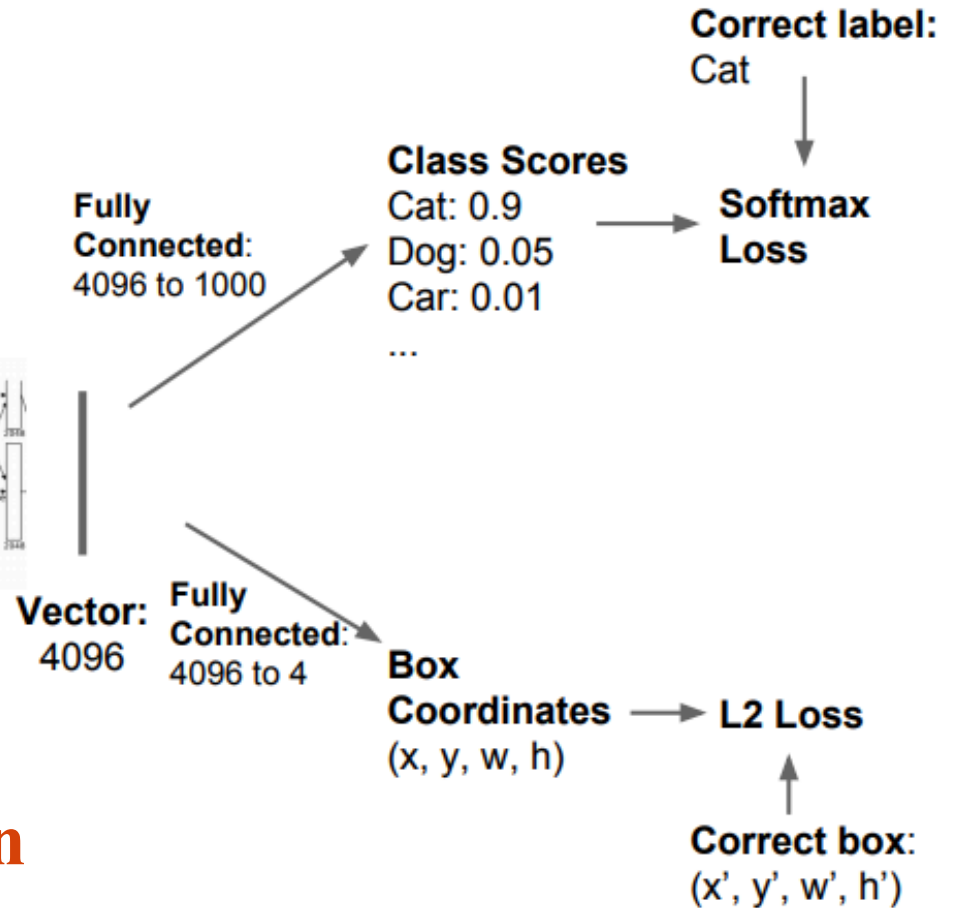Car: 0.002
…

Box Coordinates
(x, y, w, h)

Localization

Correct label:
Cat

Class Scores
Cat: 0.9
Dog: 0.05
Car: 0.01
...

Softmax Loss

Fully Connected:
4096 to 1000

Vector:
4096

Fully Connected:
4096 to 4

Box Coordinates
(x, y, w, h)

L2 Loss

Correct box:
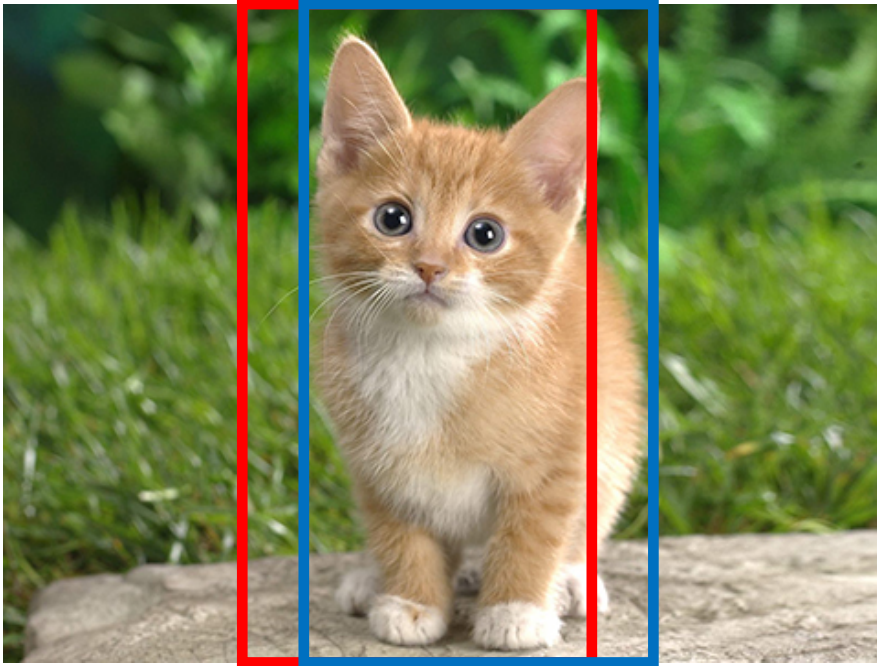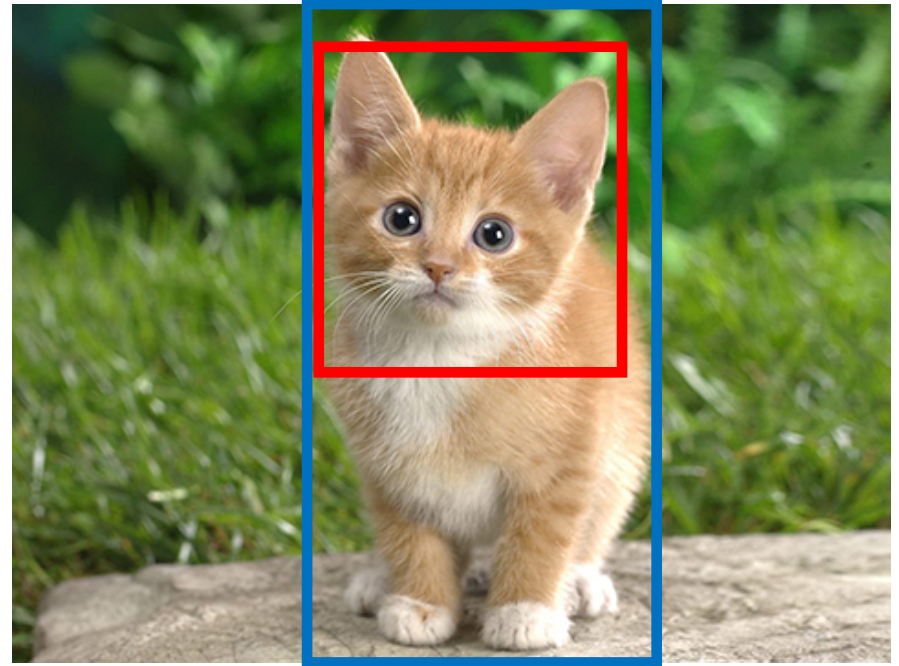(x', y', w', h')
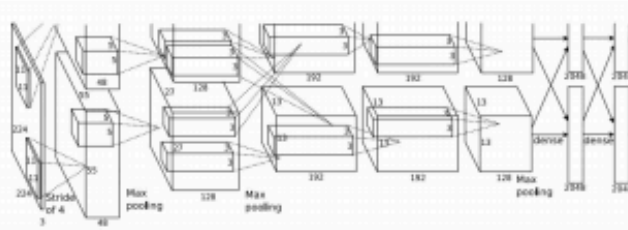
**Localization as regression
Multi-task loss**

5

Evaluation: IoU
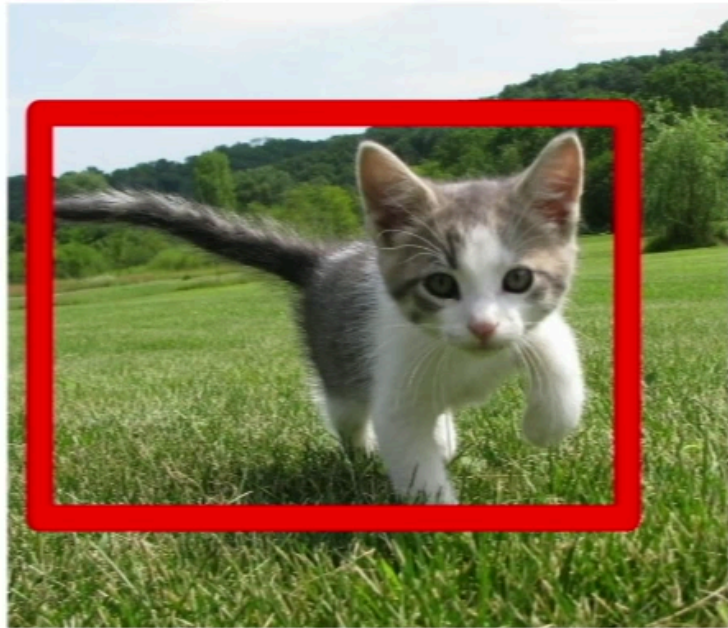

IoU: 66.7%

IoU: 45.0%

Object Detection

CAT
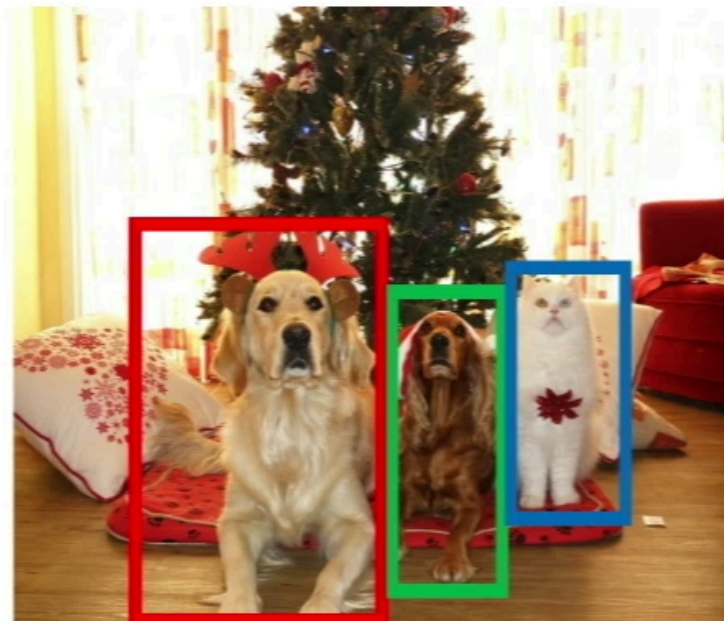
DOG, DOG, CAT

Object detection as regression?
Each image needs a different number of outputs!

Object Detection: Sliding Window

Apply a CNN to many different crops of the image, CNN classifies each crop as some object or background.
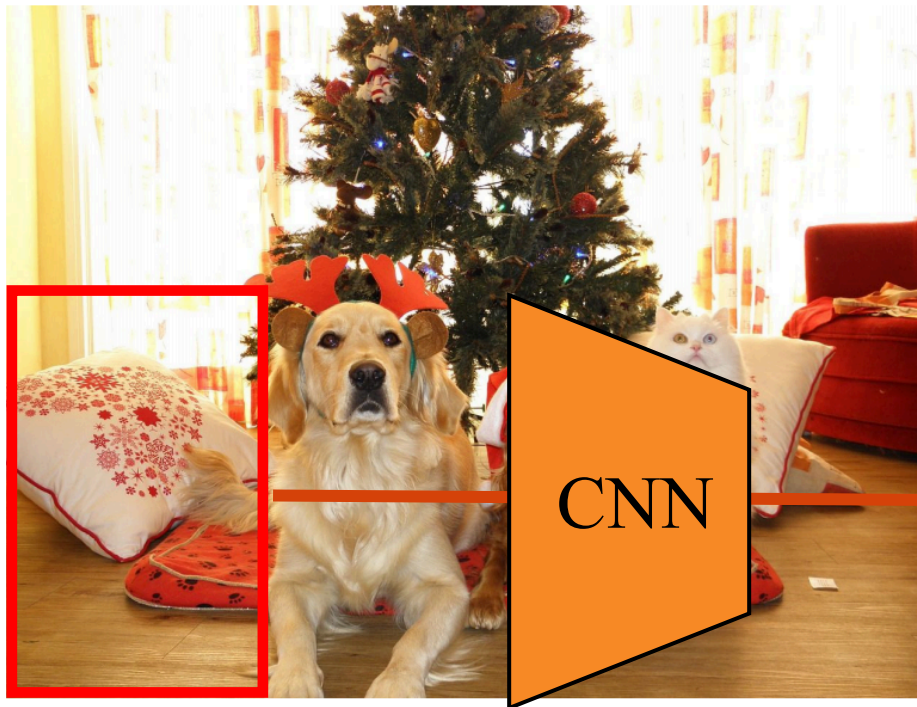
Class Scores
BG: 0.87
Dog: 0.07
Car: 0.05
…

Object Detection: Sliding Window

Apply a CNN to many different crops of the image, CNN classifies each crop as some object or background.

Class Scores
Dog: 0.95
Cat: 0.02
BG: 0.01
…

Object Detection: Sliding Window

Apply a CNN to many different crops of the image, CNN classifies each crop as some object or background.
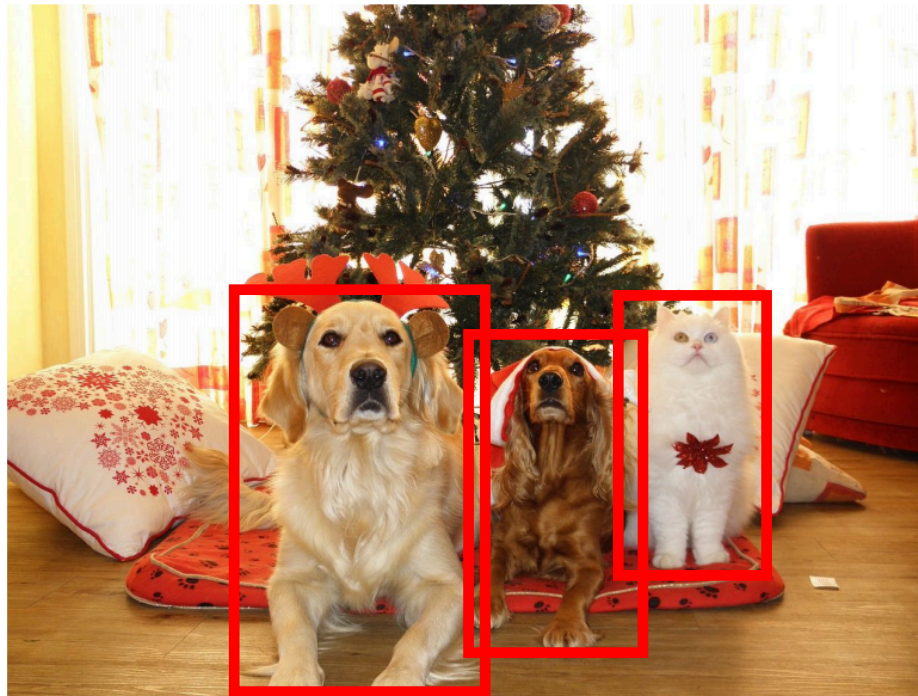
Problem: Need to apply CNN to huge number of locations and scales, very computationally expensive!
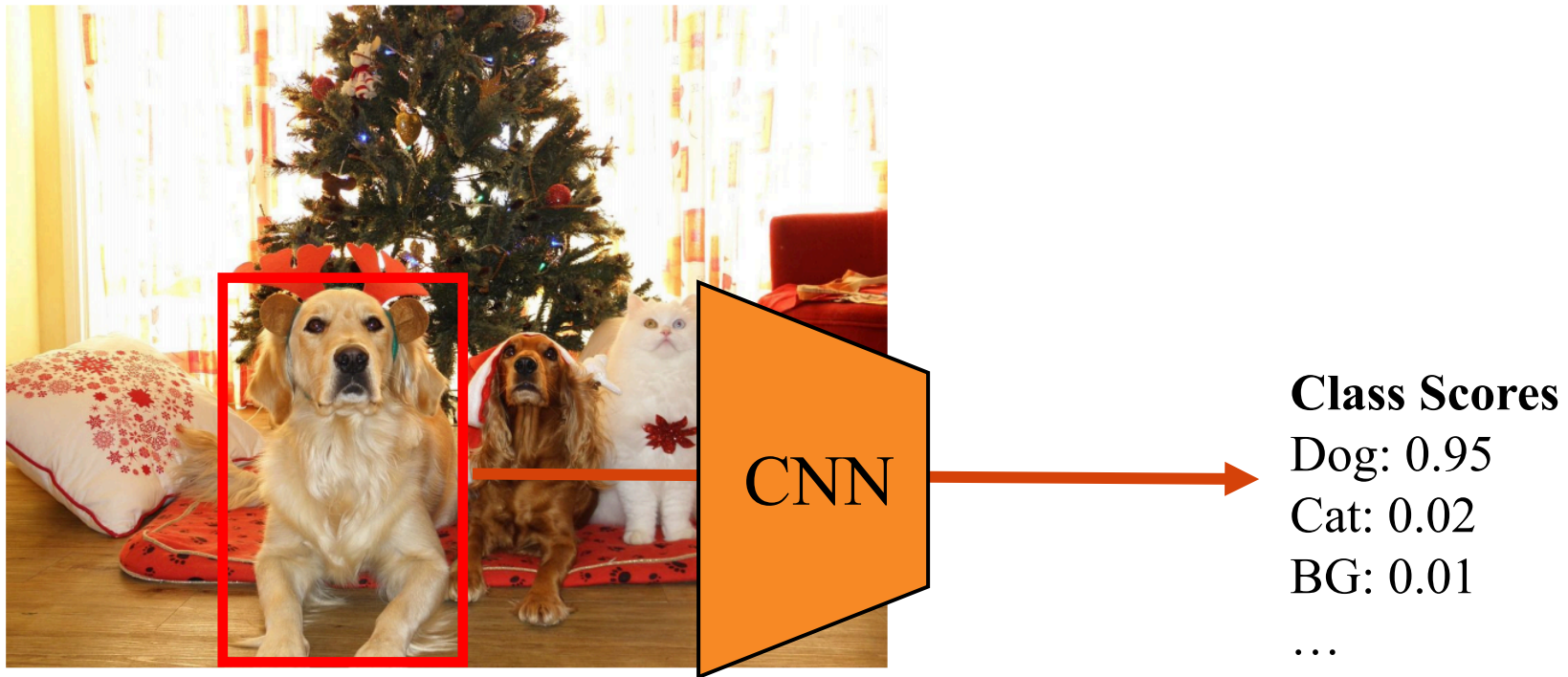
Object Detection: Region Proposal

❑Find potential image regions that are likely to contain objects.

❑Methods such as Selective Search, Edge Boxes etc.

❑Relatively fast to run; generates reasonable amount of region proposals in a few seconds on CPU

❑High recall rate
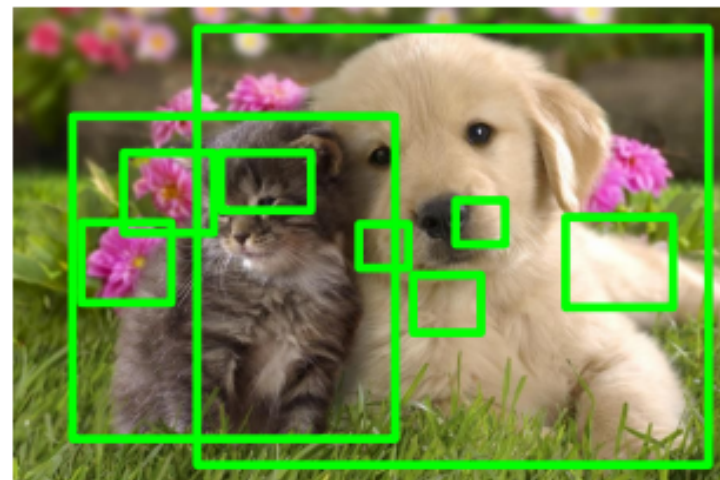
Region CNN
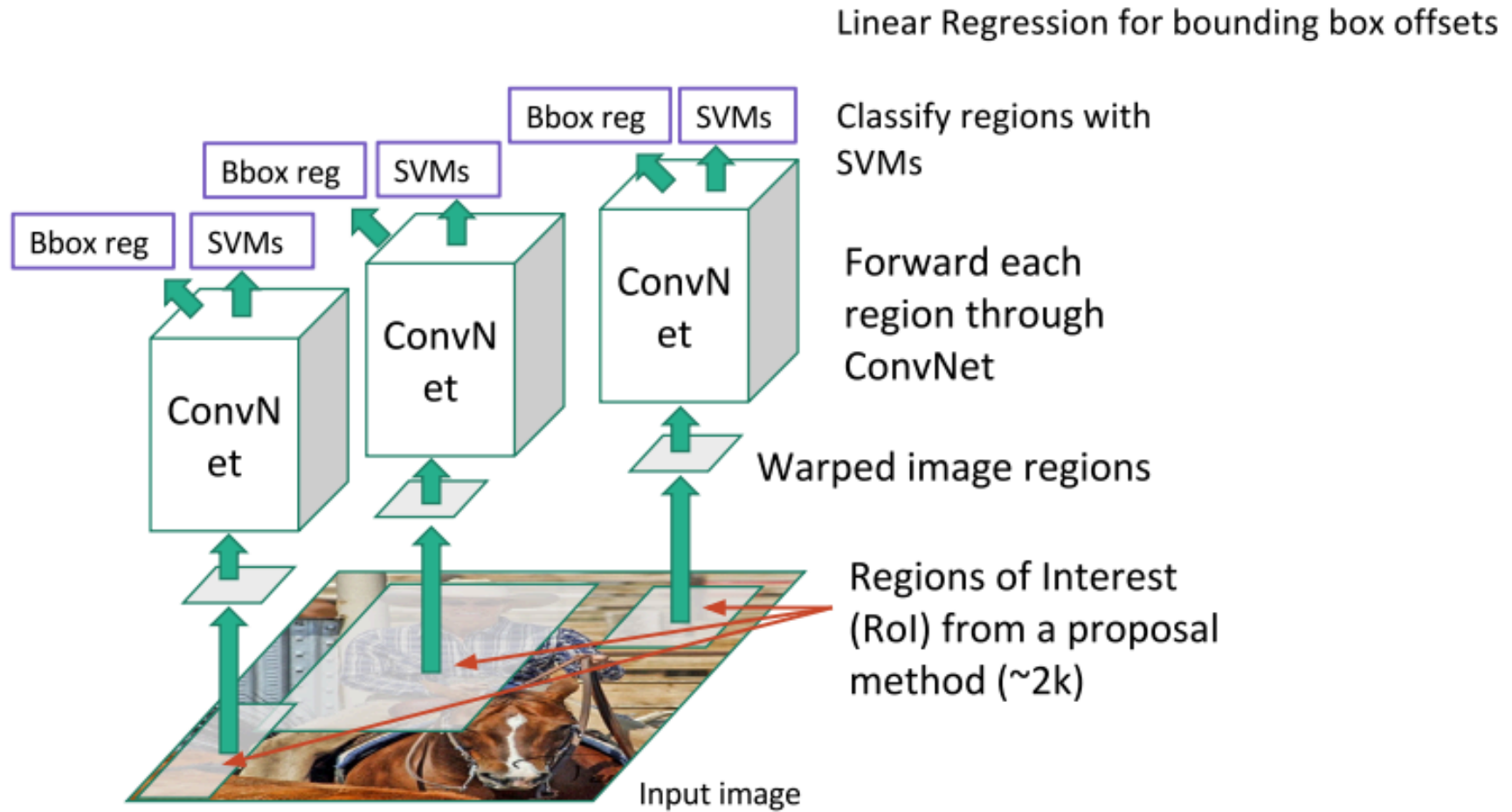
Linear Regression for bounding box offsets

Classify regions with SVMs

Forward each region through ConvNet

Warped image regions

Regions of Interest (RoI) from a proposal method (~2k)

Input image

[1]   Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.

Problem with R-CNN

❑Ad hoc training objectives
❑Inference is slow
  ❑47s / image with VGG16 network

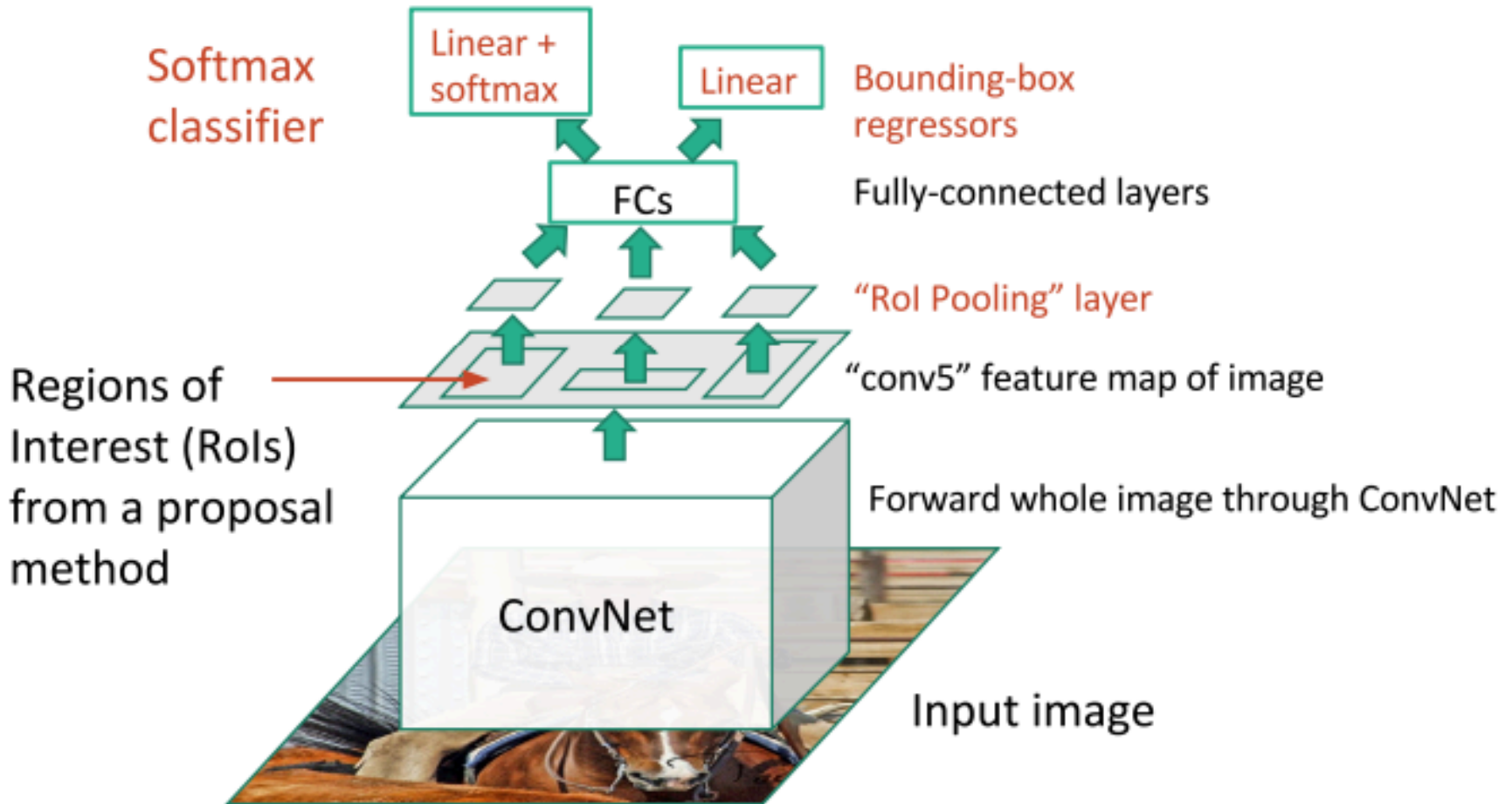[1] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.

Fast R-CNN

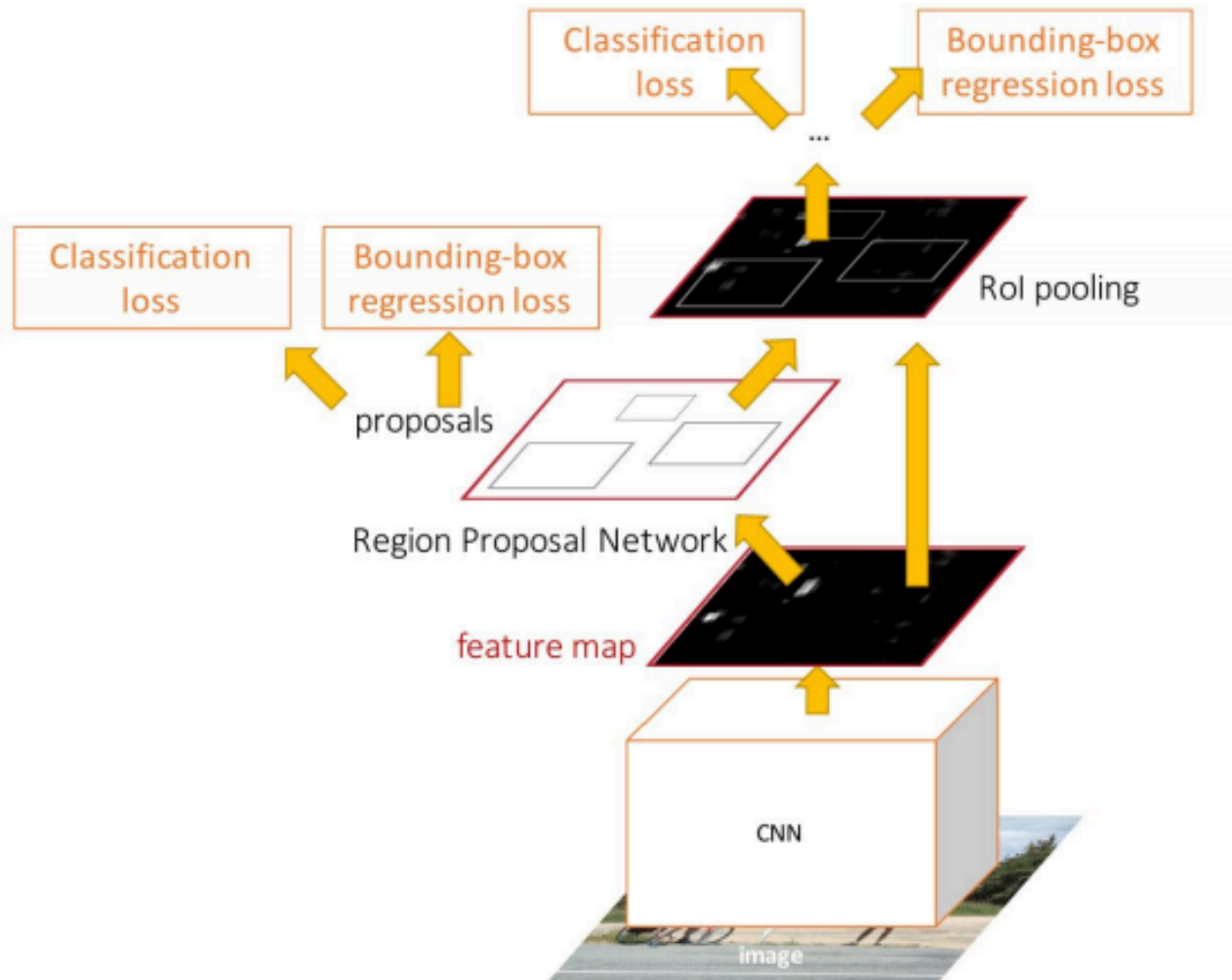[1]   Girshick, Ross. "Fast r-cnn." *Proceedings of the IEEE International Conference on Computer Vision*. 2015.

Faster R-CNN

Insert Region Proposal Network (RPN) to predict proposals from features.

An end-to-end trainable neural network architecture using the same CNN feature map for both region proposal and proposal classification.

[1] Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems*. 2015.

15

Summary

- ❑ R-CNN
  - ❑ Traditional region proposal + CNN classifier for each proposal
- ❑ Fast R-CNN
  - ❑ Traditional region proposal + CNN classifier for entire image
- ❑ Faster R-CNN
  - ❑ An unified CNN architecture for region proposal & proposal classification

	R-CNN	**Fast R-CNN**	**Faster R-CNN**
Test time	50 s	2 s	0.2 s
mAP (%)	66.0	66.9	66.9

	AlexNet	**VGG-16**	**ResNet-101**
mAP (%)	62.1	73.2	76.4

Using different CNNs in faster R-CNN

Object Detection: Lots of variations

❑Base Network
- ❑ Inception V2
- ❑ Inception V3
- ❑ DenseNet
- ❑ MobileNet
- ❑ ...

❑Other architectures
- ❑ YOLO
- ❑ R-FCN
- ❑ SSD
- ❑ …

Thank You !