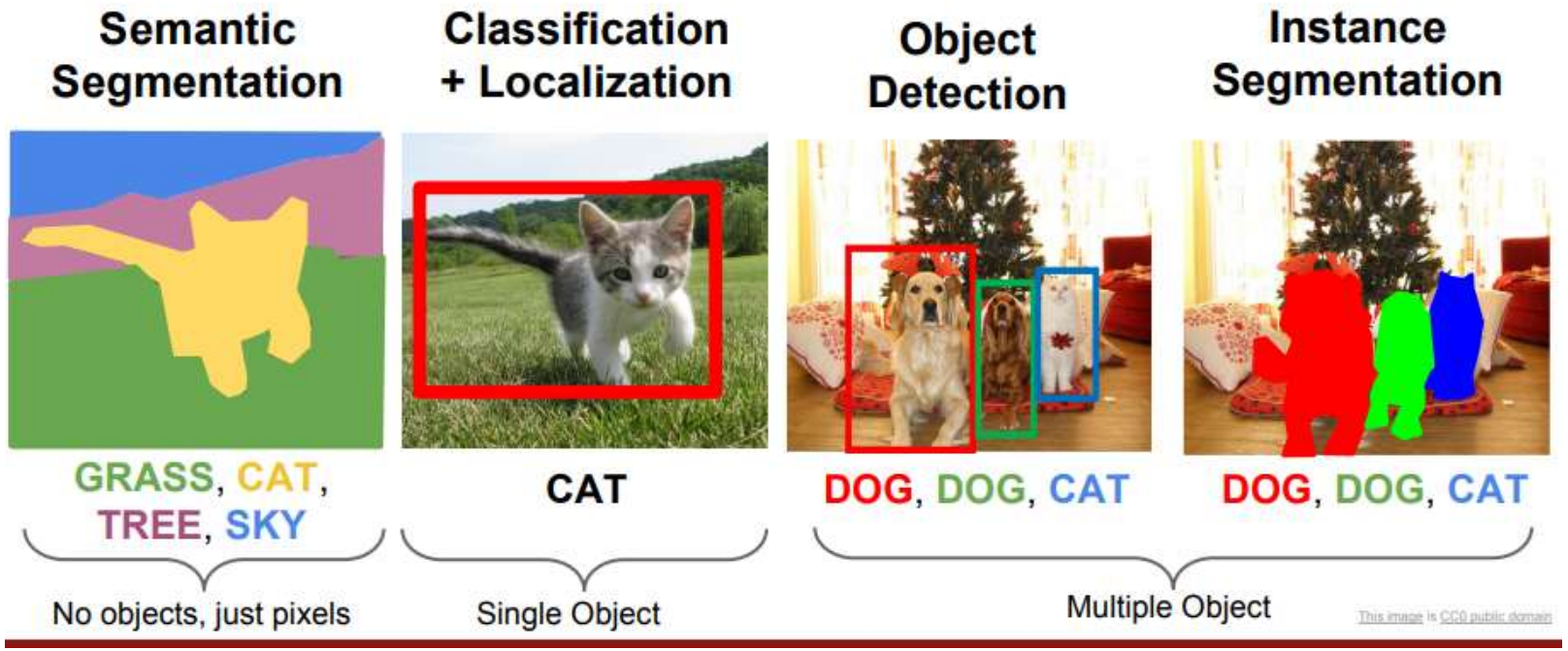
Object Detection

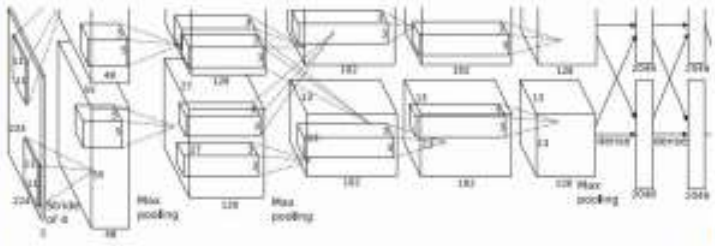
Computer vision tasks



http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf

Classification + Localization

This image is CC0 public domain



Vector: 4096

Fully Connected: 4096 to 1000

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

Softmax Loss

Correct label: Cat

Fully Connected: 4096 to 4

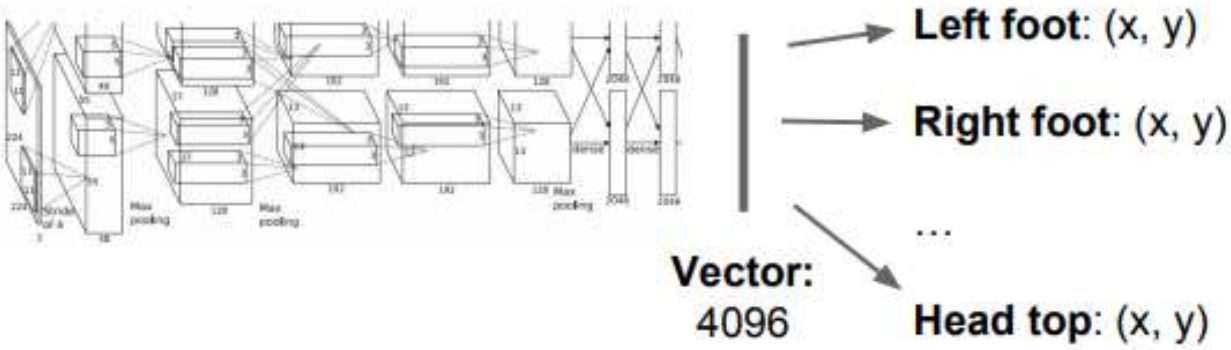
Box Coordinates (x, y, w, h)

L2 Loss

Correct box: (x', y', w', h')

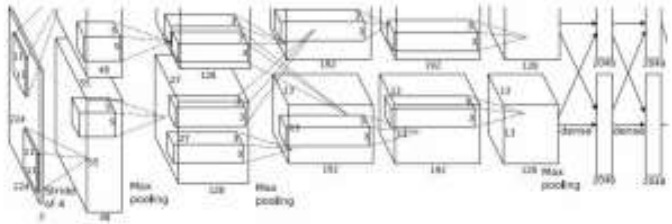
Treat localization as a regression problem!

Aside: Human Pose Estimation



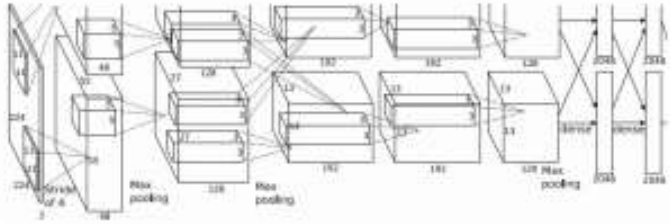
Toshev and Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks", CVPR 2014

Object detection



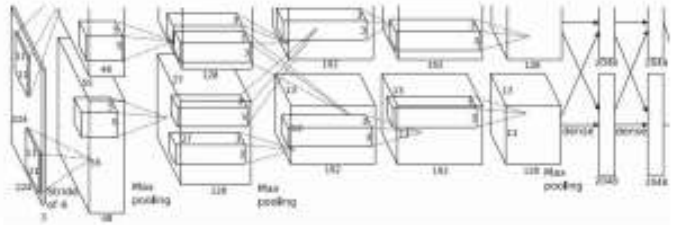
CAT: (x, y, w, h)

4 numbers



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

16 numbers

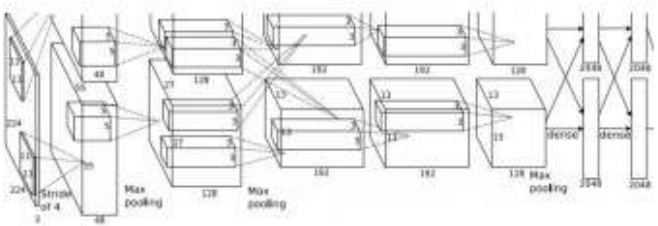


DUCK: (x, y, w, h)
DUCK: (x, y, w, h)
....

Many numbers!

Object Detection as Classification: Sliding Window

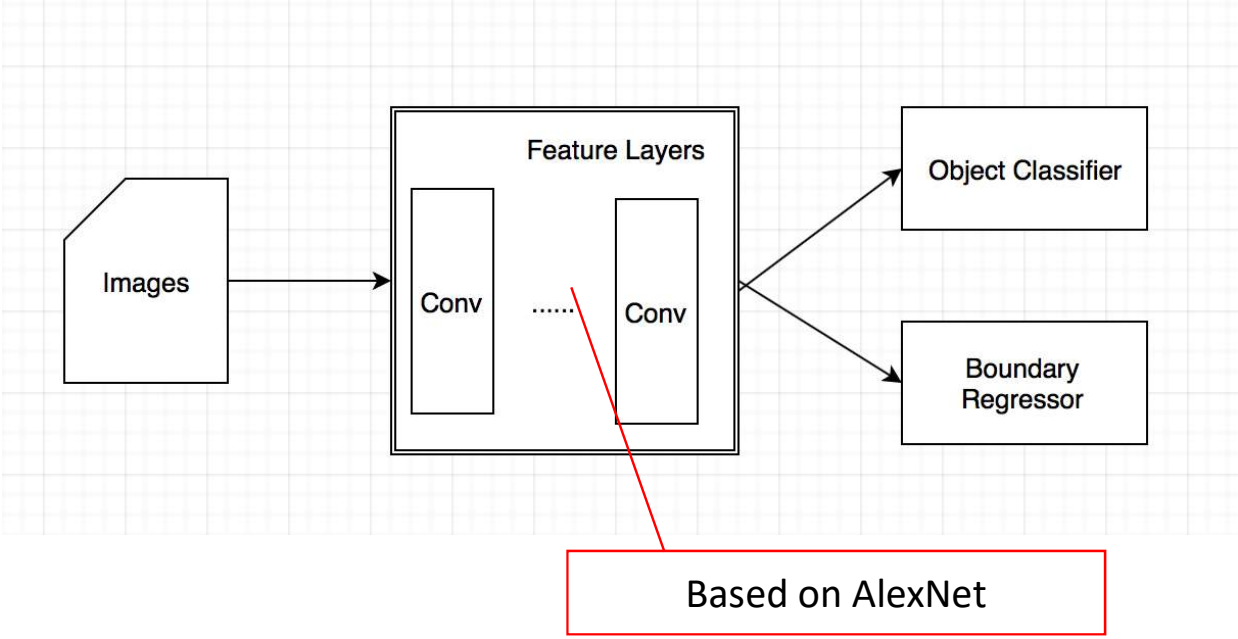
Apply a CNN to many different crops of the image, CNN classifies each crop as object or background



Dog? NO
Cat? NO
Background? YES

Overfeat (2013)

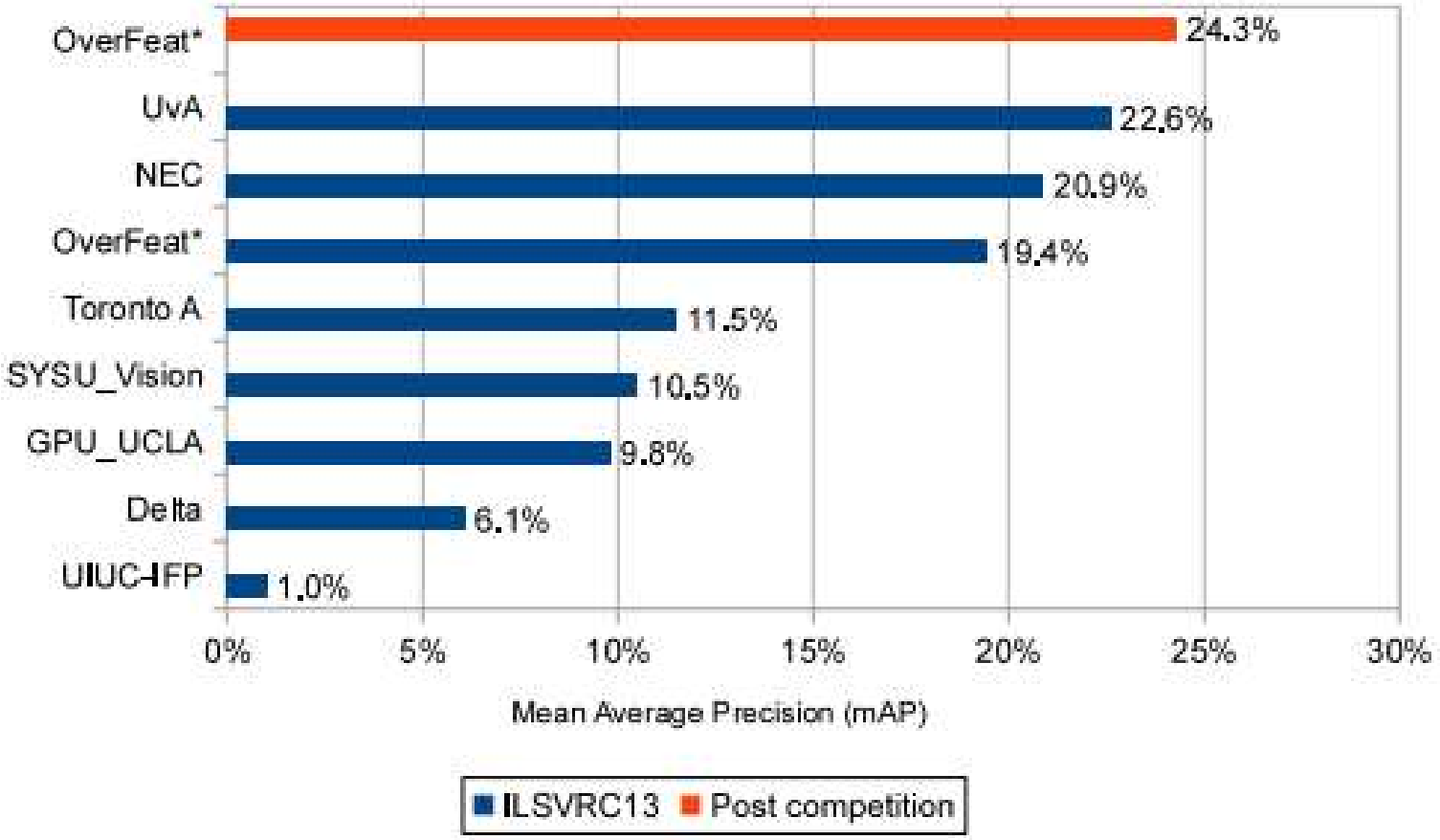
- the first published neural net based object localization architecutre



- Sliding window approach
- Applies classification at multiple evenly spaced square windows over multiple scale of the image.
- Due to extensive pooling, windows do not align too well aligned with the objects.

KMJ1

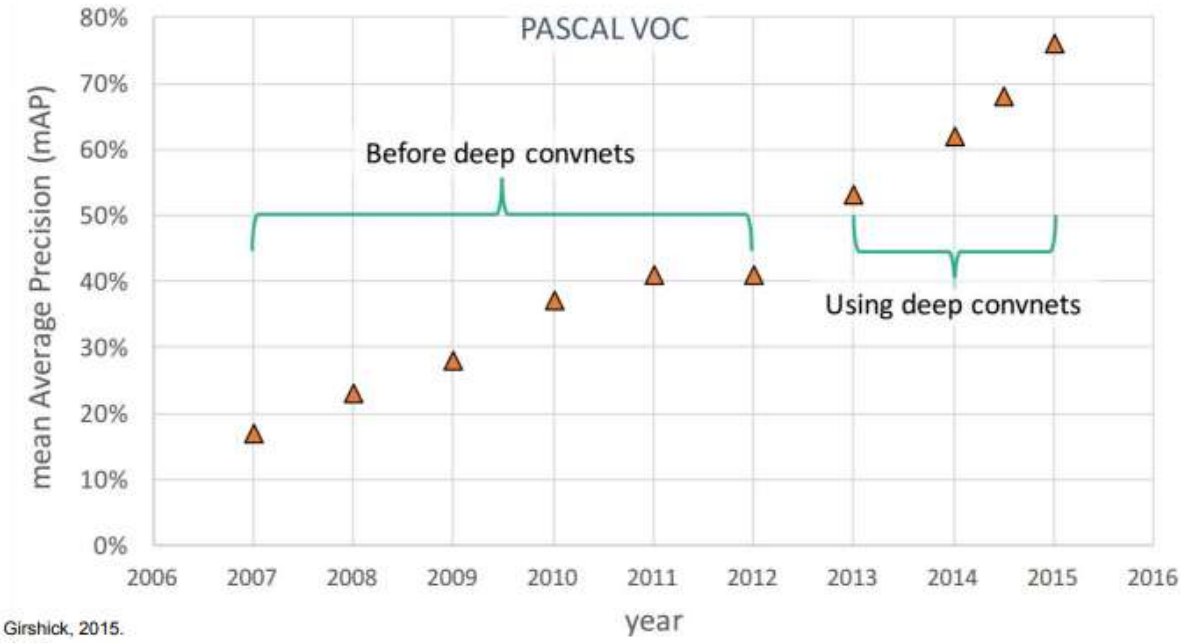
ImageNet 2013 Detection results



Slide 8

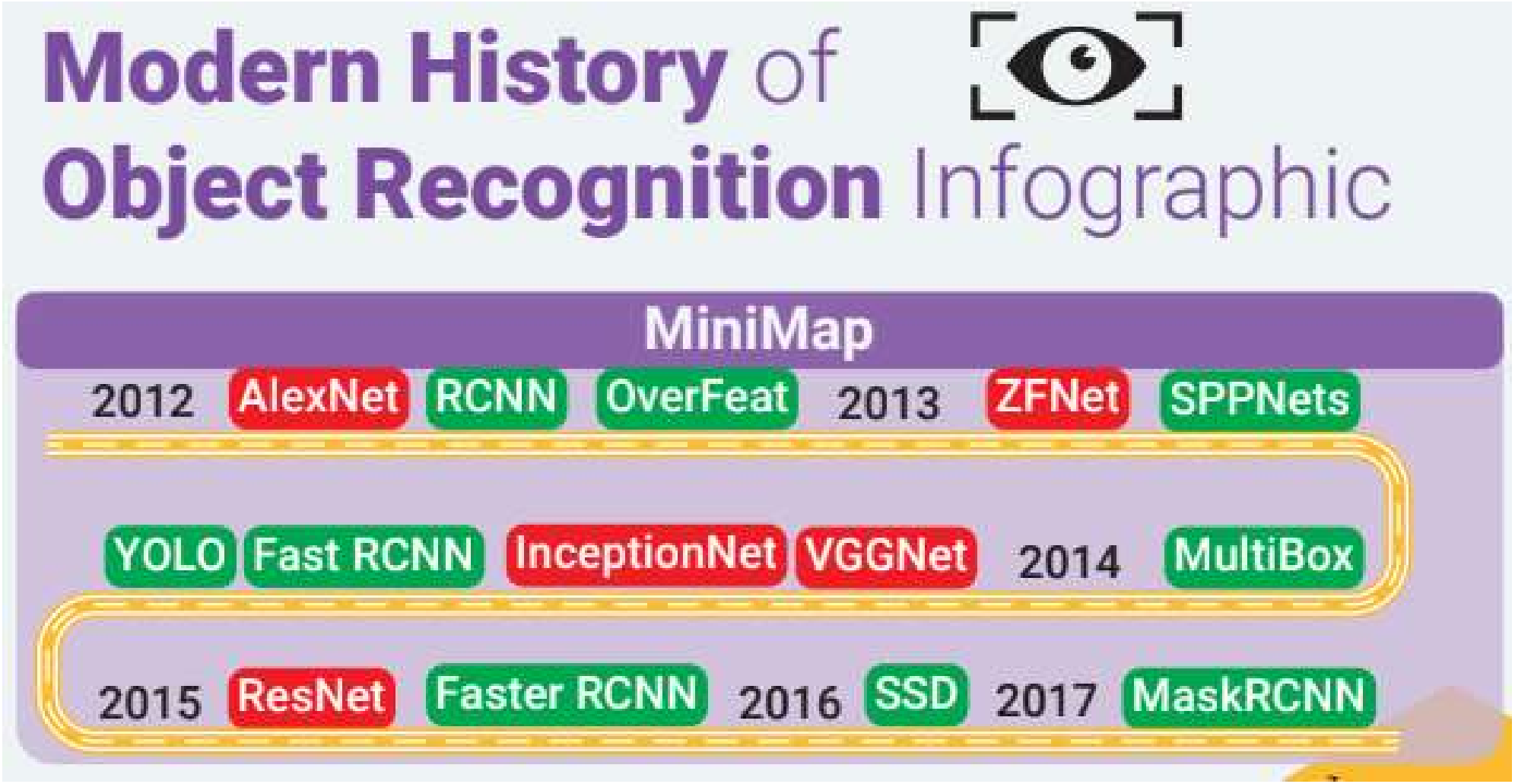
KMJ1 Kari Meling Johannessen, 08/10/2019

Object Detection: Impact of Deep Learning



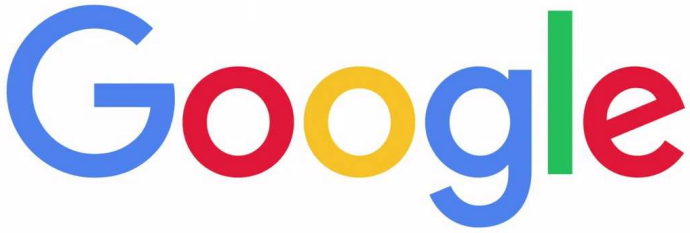
Problems with sliding window

- How to choose crop: size, position aspect ratio.
- Brute force … very computationally expensive



https://medium.com/@nikasa1889/the-modern-history-of-object-recognition-infographic-aea18517c318

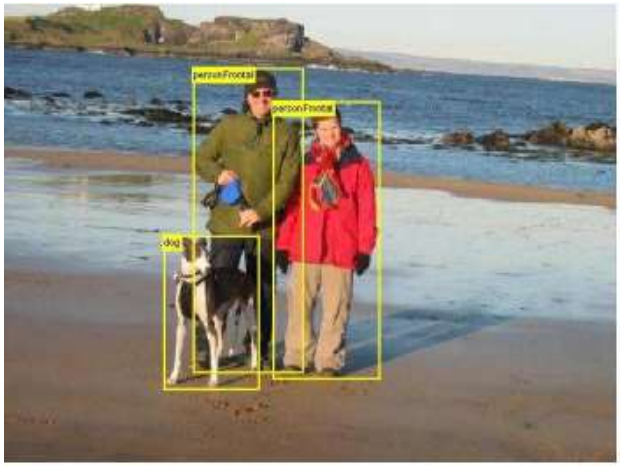
Object Detection
datasets and challenges



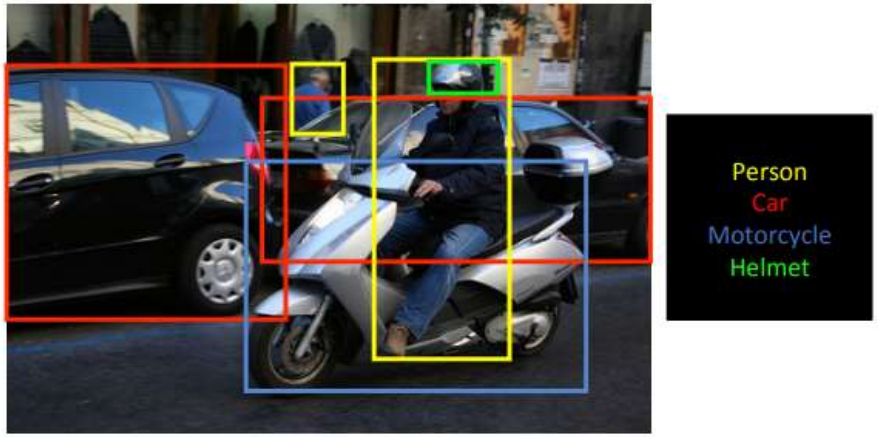
PASCAL VOC

PASCAL VOC

- The PASCAL VOC dataset (2012) is well-known an commonly used for object detection and segmentation.
- More than 11k images compose the train and validation datasets while 10k images are dedicated to the test dataset.
- Metric: Intersection over Union (IoU)



IMAGENET



200 object classes **527,982 images** **DET**

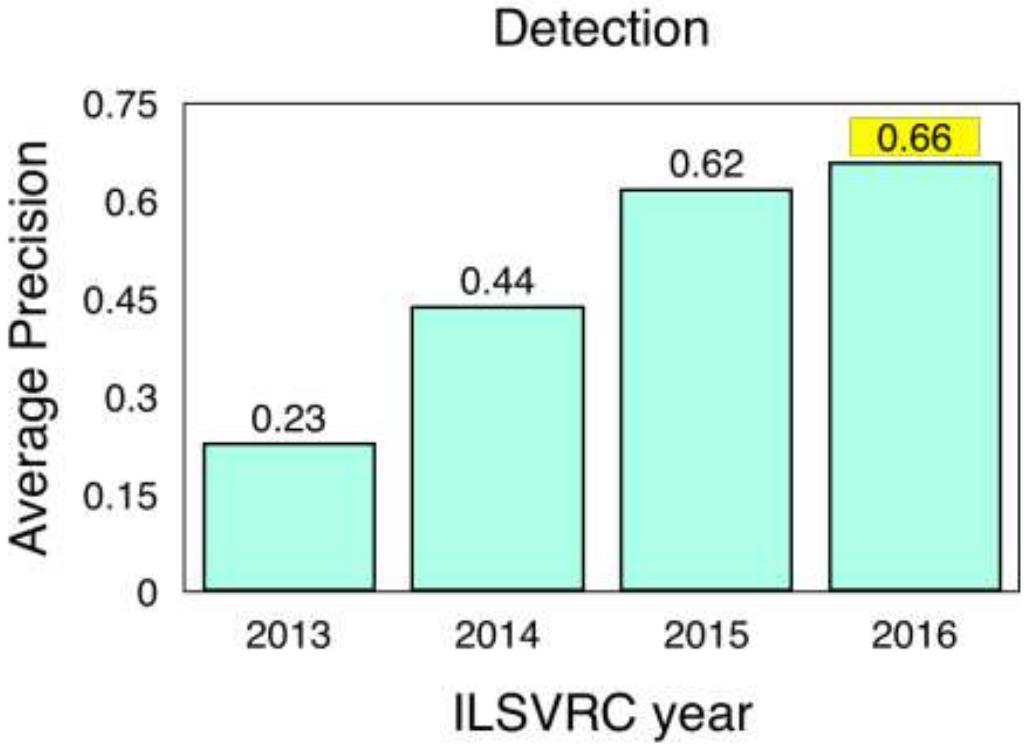
- **The ILSVRC competition has its own image dataset that is a subset of the ImageNet dataset.**
- 1,000 object categories spread over ~1.2 million images.
- Half of these images also have bounding boxes around the class category object

Evaluation based after PASCAL VOC:

- Algorithm outputs a list of bounding box detections with confidences
- A detection is considered correct if intersection over union (IoU) overlap with ground truth > threshold (0.5)
- Evaluated by average precision per object class
- Winners of challenge is the team that wins the most object categories

http://image-net.org/challenges/talks/2016/ECCV2016_ilsvrc_coco_detection_segmentation.pdf

IMAGENET


Detection
Average Precision
0.75
0.6
0.45
0.3
0.15
0
0.23
0.44
0.62
0.66
2013
2014
2015
2016
ILSVRC year

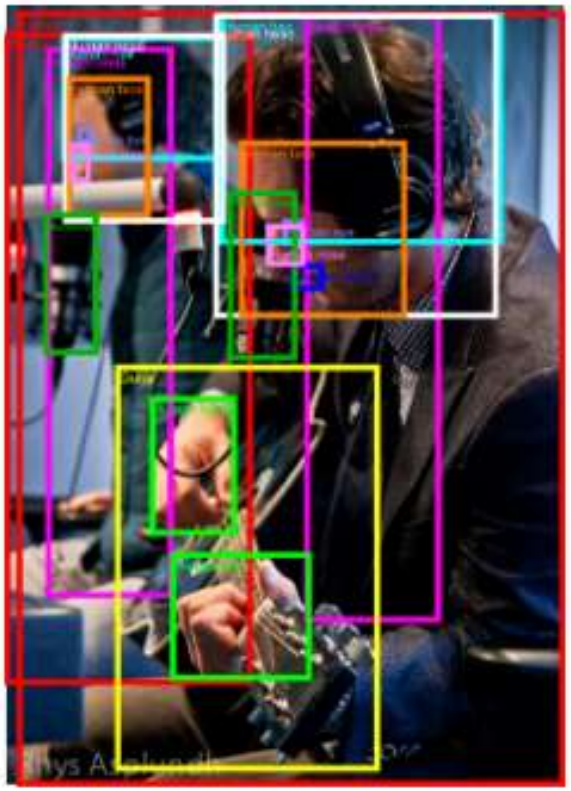
COCO
Common Objects in Context

- 80 object categories
- 200k images
- 1.2M instances (350k people)
- Every instance segmented
- **106k people with keypoints** NEW

Available for download at
mscoco.org

Detection Segmentation Keypoints

Open Images Dataset



Open Images example image with bounding box annotation

- first released in 2016
- complex scenes that span thousands of classes of objects
- ~15.4 million bounding boxes of 600 classes of objects