# Object Detection

Honghui Shi

IBM Research

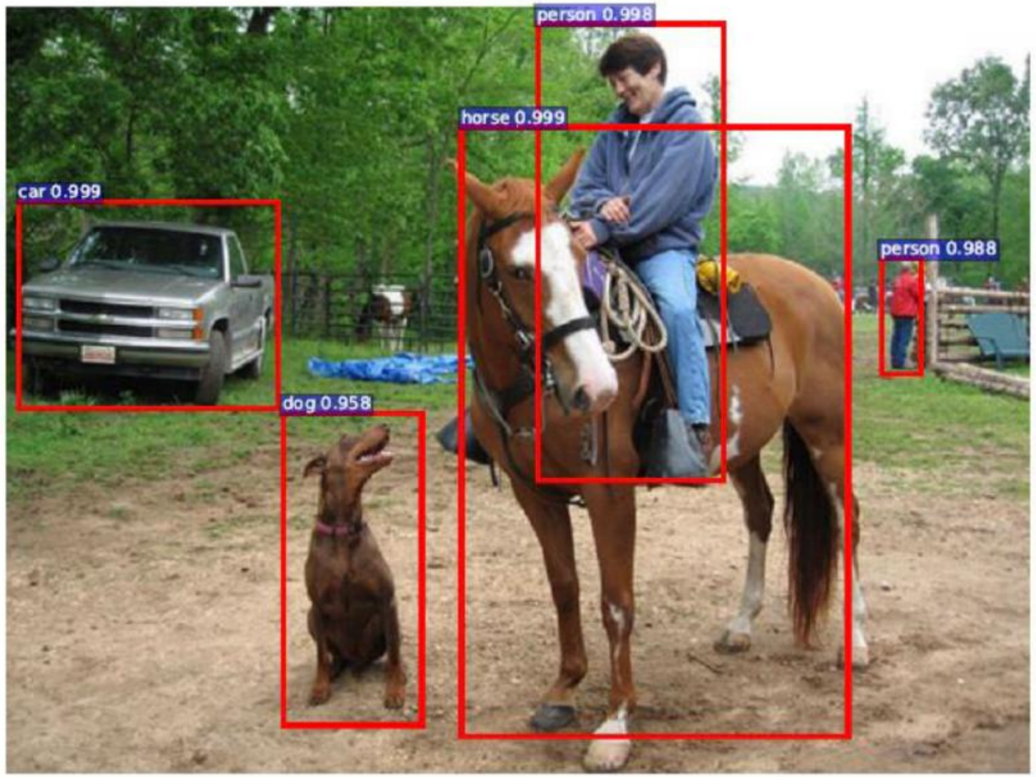
2018.11.27 @ Columbia

# Outline

- Problem
- Evaluation
- Methods
- Directions

# Problem

Image classification: Horse (People, Dog, Truck…)




Object detection: categories & **locations** of objects

# Challenges

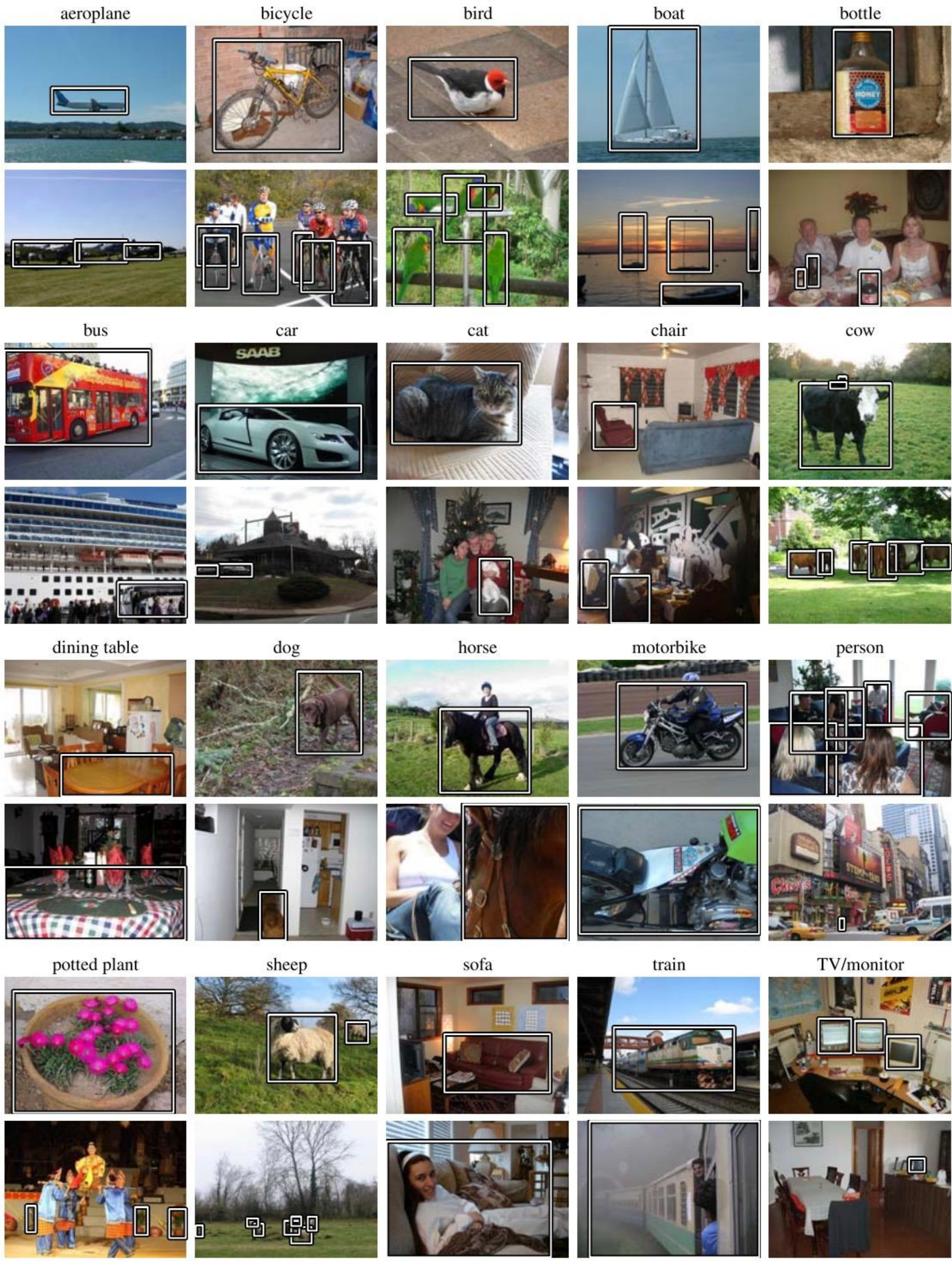
- From single image-level label to **multiple object instances**
- Object localization
- Object classification

# Datasets

- PASCAL VOC
- ImageNet
- COCO
- Google Open Images
- KITTI
- Nvidia AI City

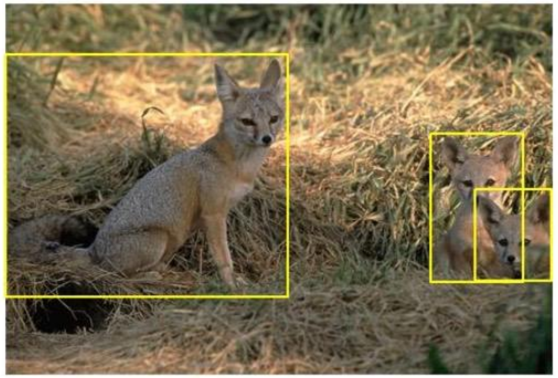
# PASCAL VOC

- Dataset (voc2012)
  - 20 classes
    - Person
    - Animal (bird, cat, cow, dog, horse, sheep)
    - Vehicles (aeroplane, bicycle, boat, bus, car, motorbike, train)
    - Indoor (bottle, chair, dinning table, potted plant, sofa, tv monitor)
  - ~ 11k train/val, 27k boxes, 7k segmentations
- Challenge
  - 2005 - 2012

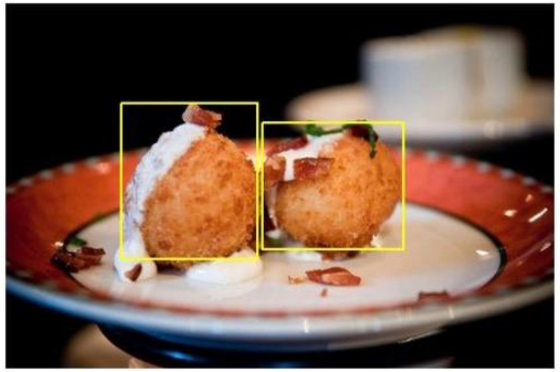


# ImageNet

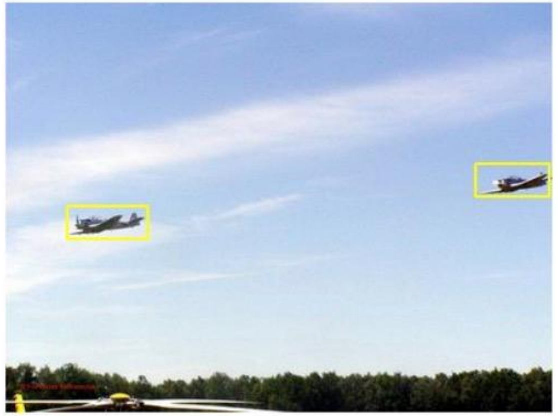
- Image Dataset
  - 200 categories
  - ~ 450k images
- Video Dataset
  - 30 categories
  - ~ 4000 videos



kit fox



croquette

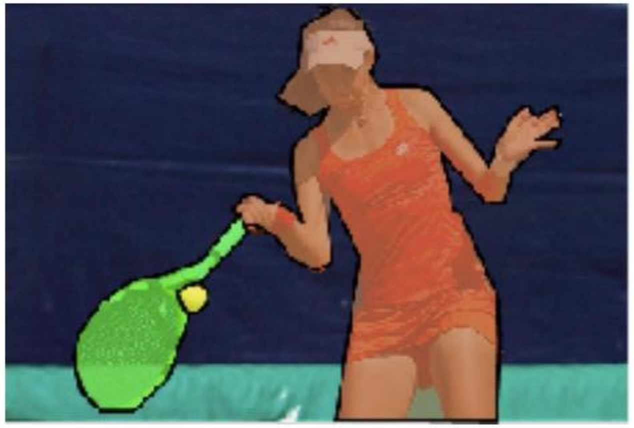


airplane

frog

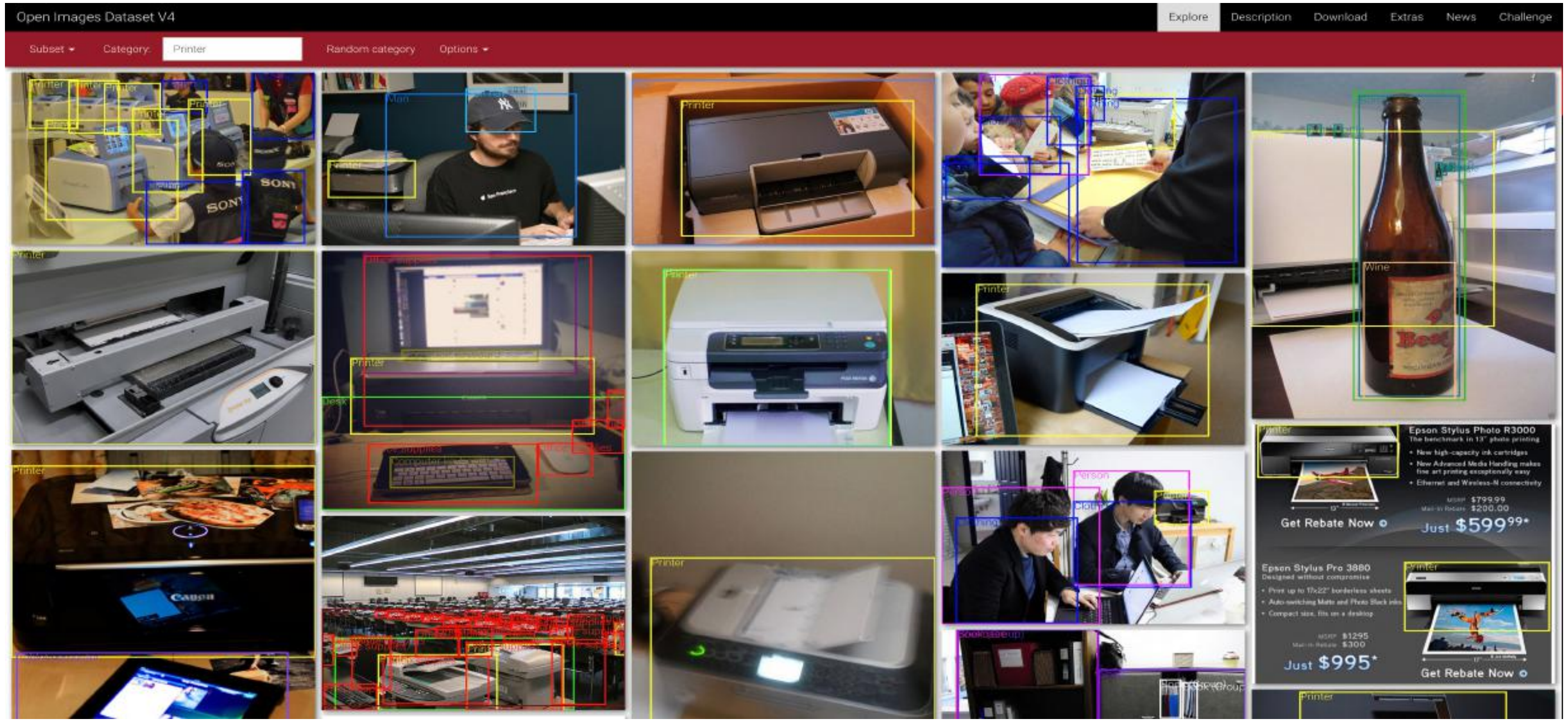
# COCO

- Dataset
  - 80 categories
  - ~ 200k images
- Challenge
  - 2015 ~ now



# Google OpenImages

- ~ 9M images overall
- 14.6M bounding boxes for 600 object classes on 1.74M images
- Complex scenes with several objects (8.4 per image on average).

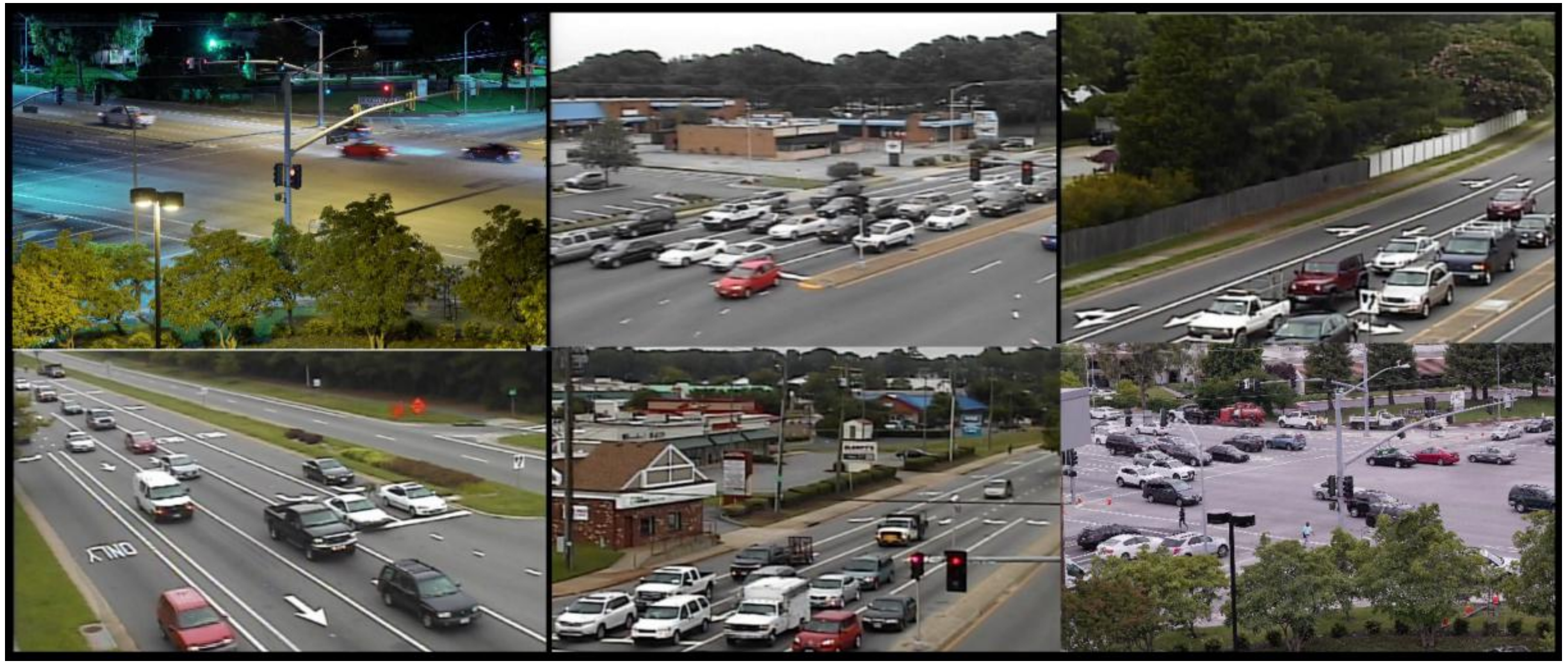


# KITTI

- Dataset
  - 7481/7518 train/val, 80k objects
- Leaderboard
  - 100+ entries

# Nvidia AI City Dataset

- Traffic cameras, challenge to be hosted in 2019

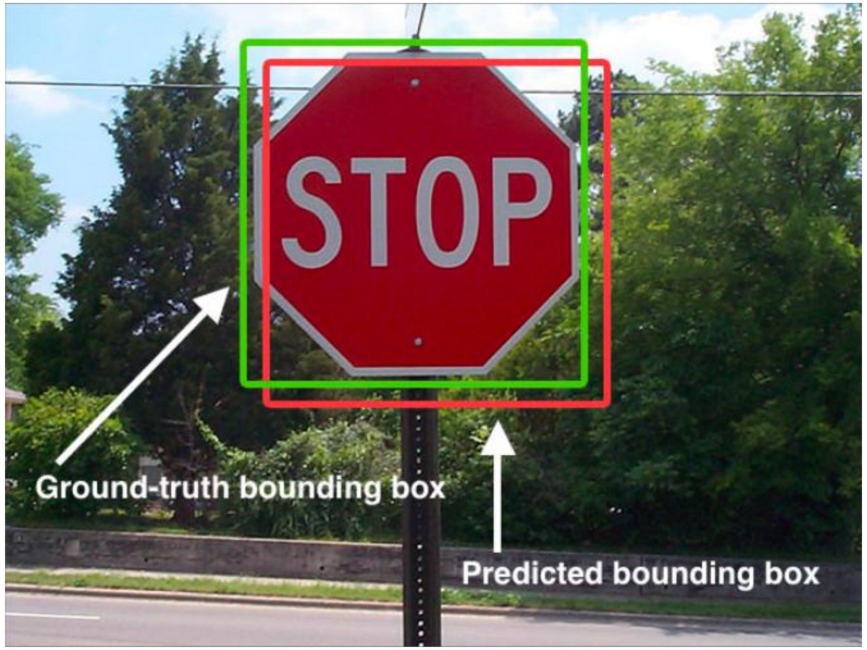


# Evaluation Metrics

- Evolving Metrics
  - VOC
  - COCO
  - OpenImages

- Two core concepts
  - IoU
  - AP

# IoU: Intersection over Union



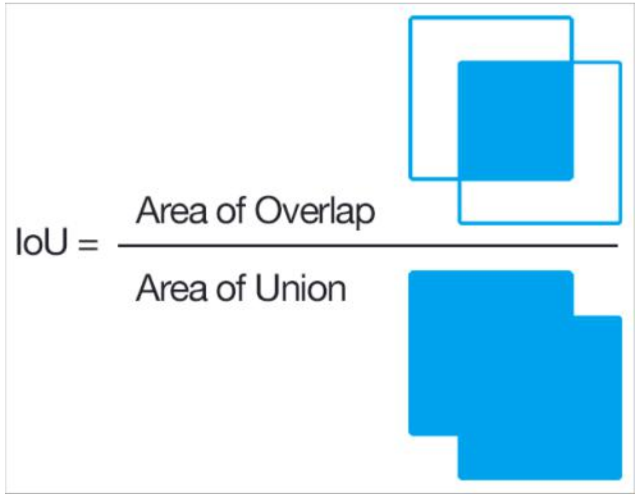
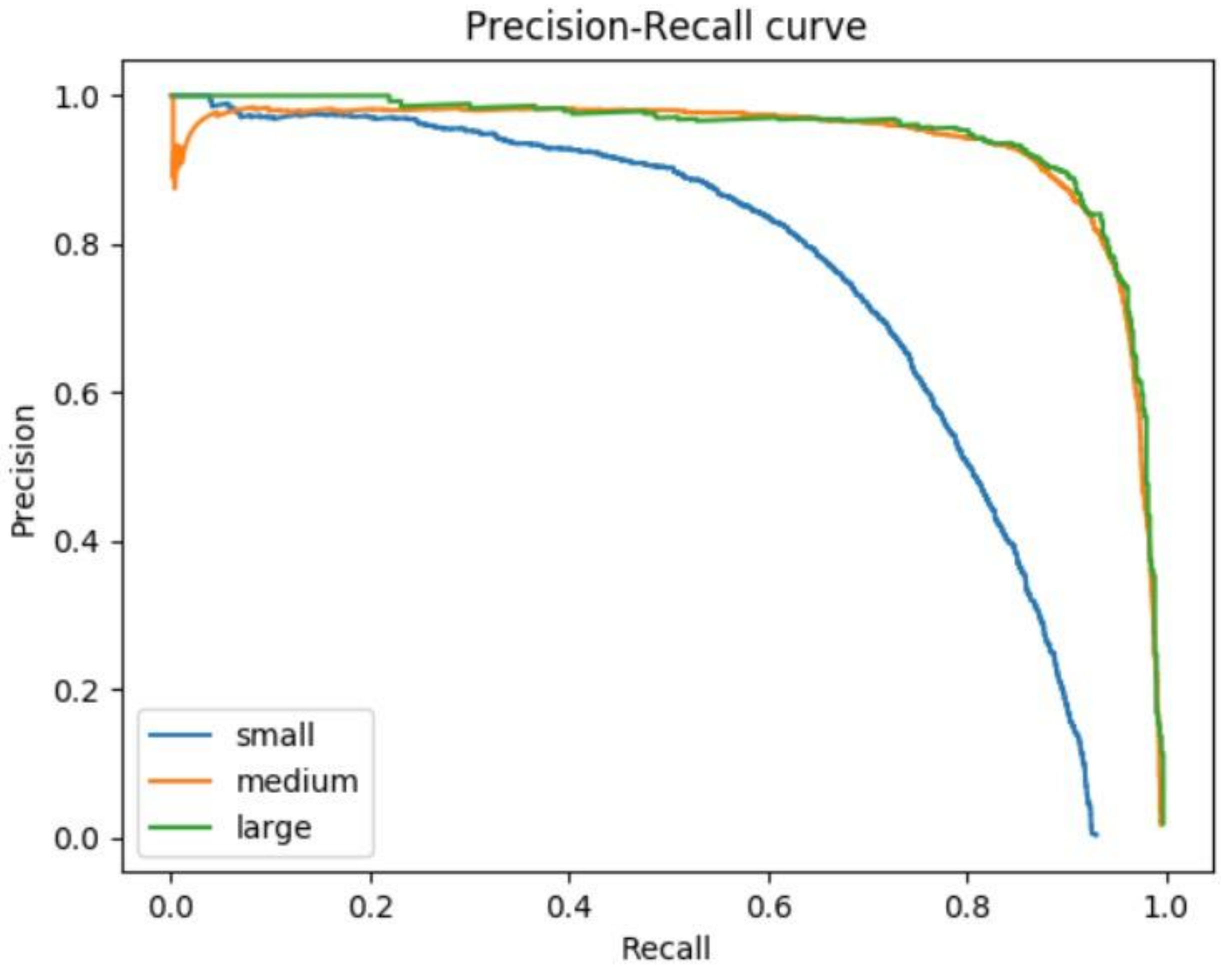




Image source

# AP: Average Precision



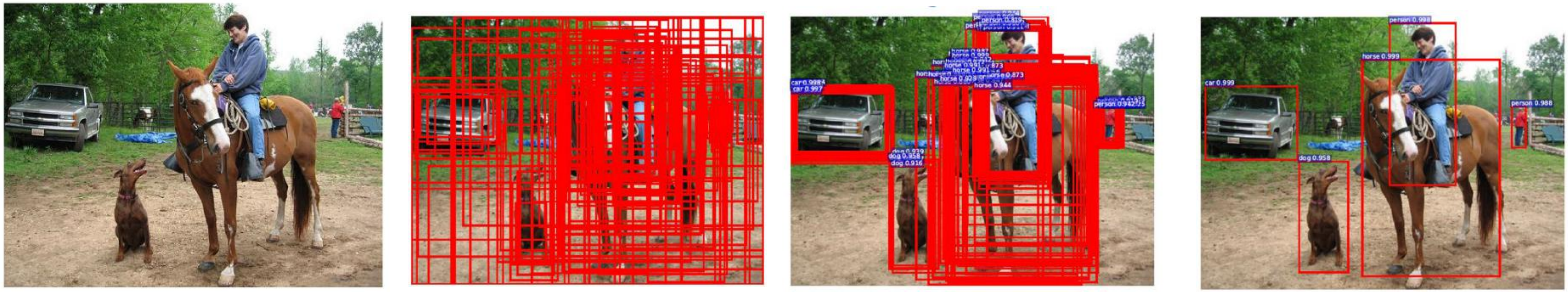
# COCO Metric

```
Average Precision (AP):
   AP                % AP at IoU=.50:.05:.95 (primary challenge metric)
   AP^IoU=.50        % AP at IoU=.50 (PASCAL VOC metric)
   AP^IoU=.75        % AP at IoU=.75 (strict metric)
AP Across Scales:
   AP^small          % AP for small objects: area < 32²
   AP^medium         % AP for medium objects: 32² < area < 96²
   AP^large          % AP for large objects: area > 96²
Average Recall (AR):
   AR^max=1          % AR given 1 detection per image
   AR^max=10         % AR given 10 detections per image
   AR^max=100        % AR given 100 detections per image
AR Across Scales:
   AR^small          % AR for small objects: area < 32²
   AR^medium         % AR for medium objects: 32² < area < 96²
   AR^large          % AR for large objects: area > 96²
```

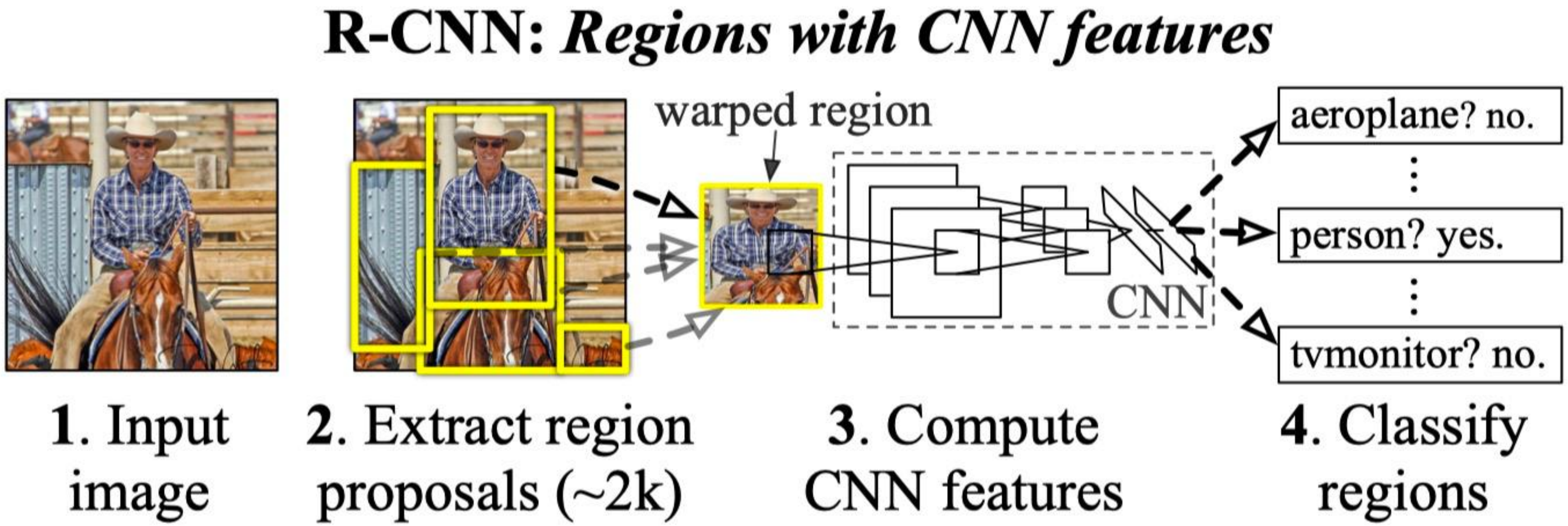
[Data source]()

# Methods

- Methods before ImageNet
  - DPM
- CNN based detectors
  - Proposal-based
  - Proposal-free

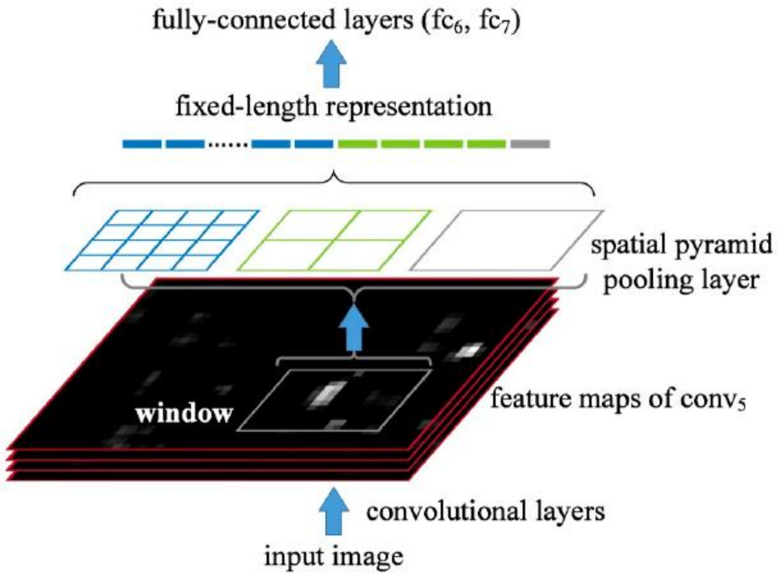


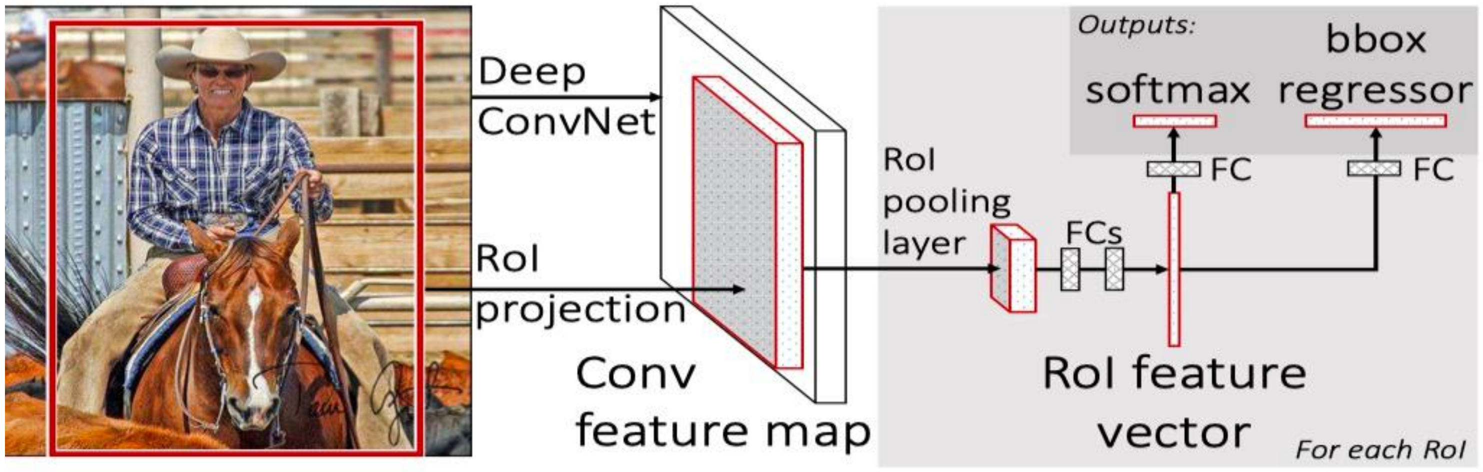
# RCNN




Ross Girshick et al., 2014

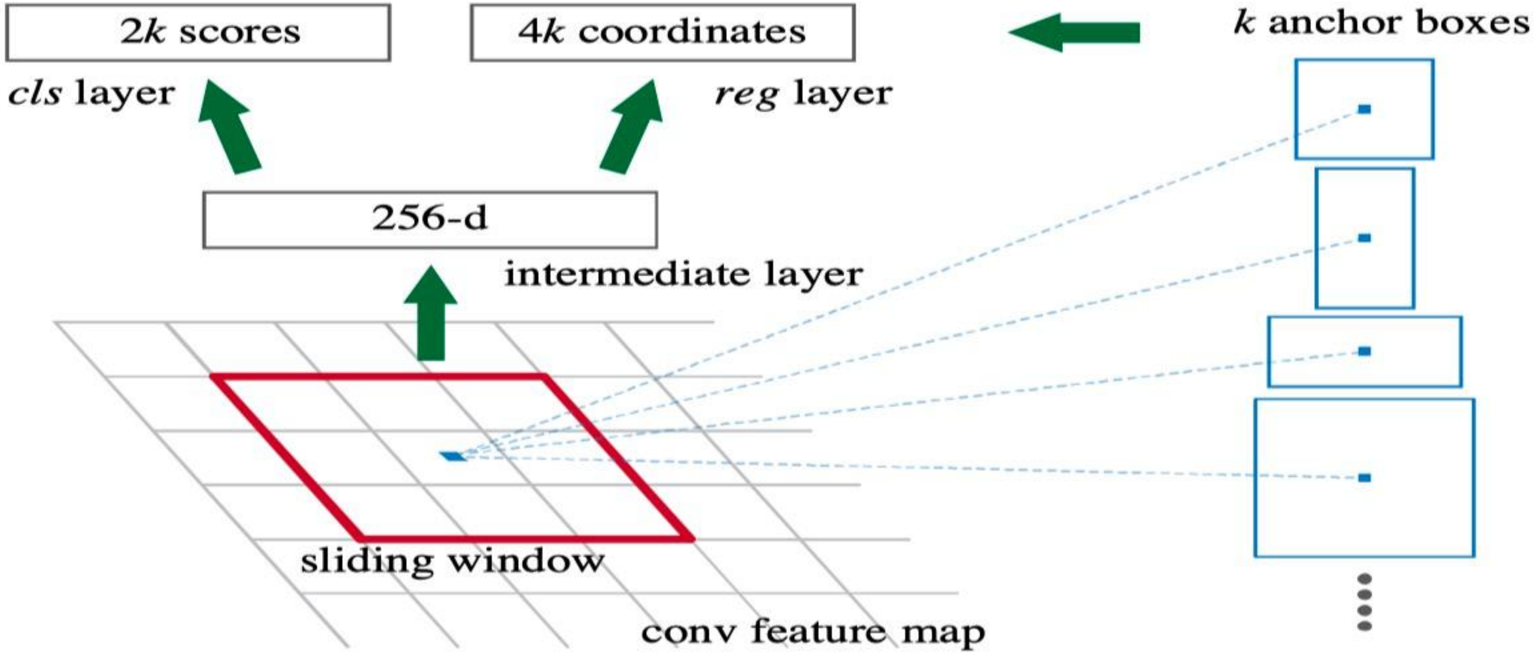
# Fast RCNN



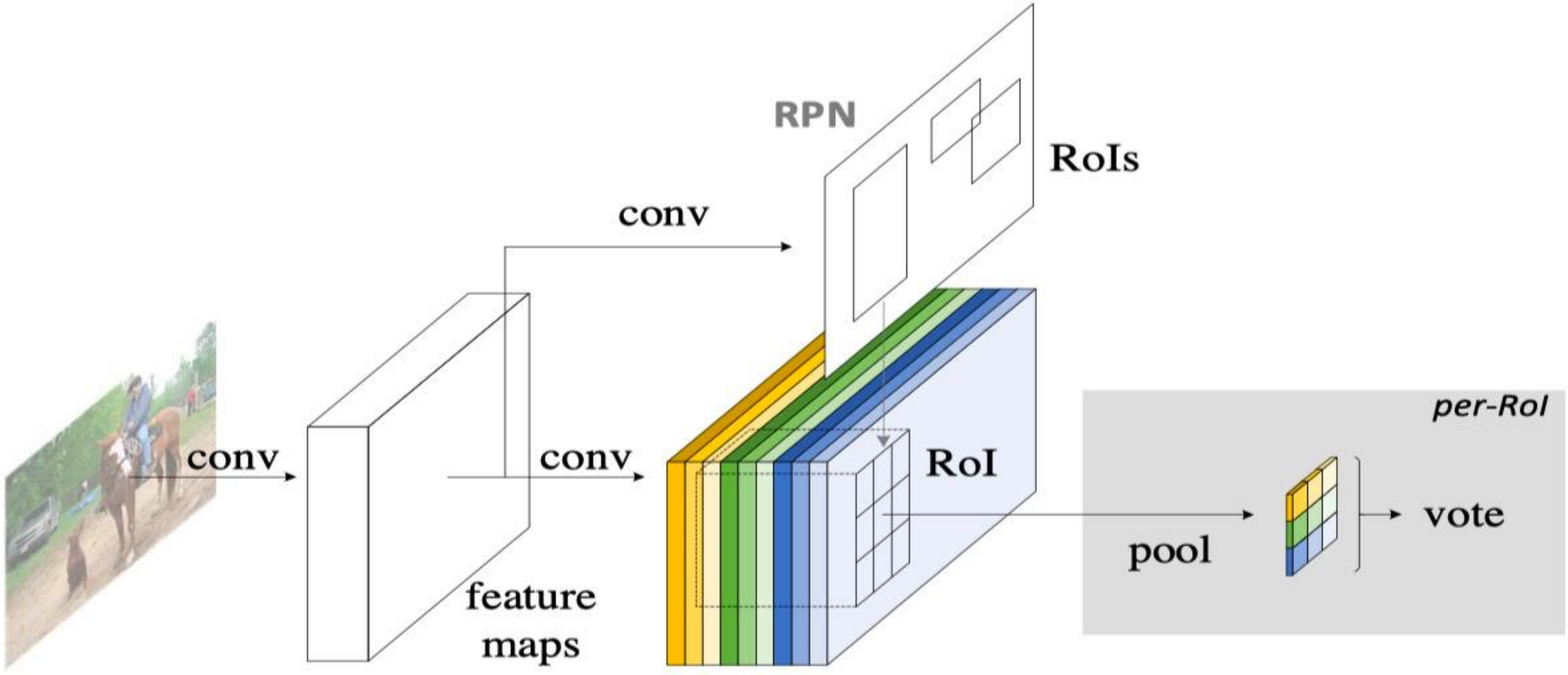

Ross Girshick et al., 2015

# Faster RCNN



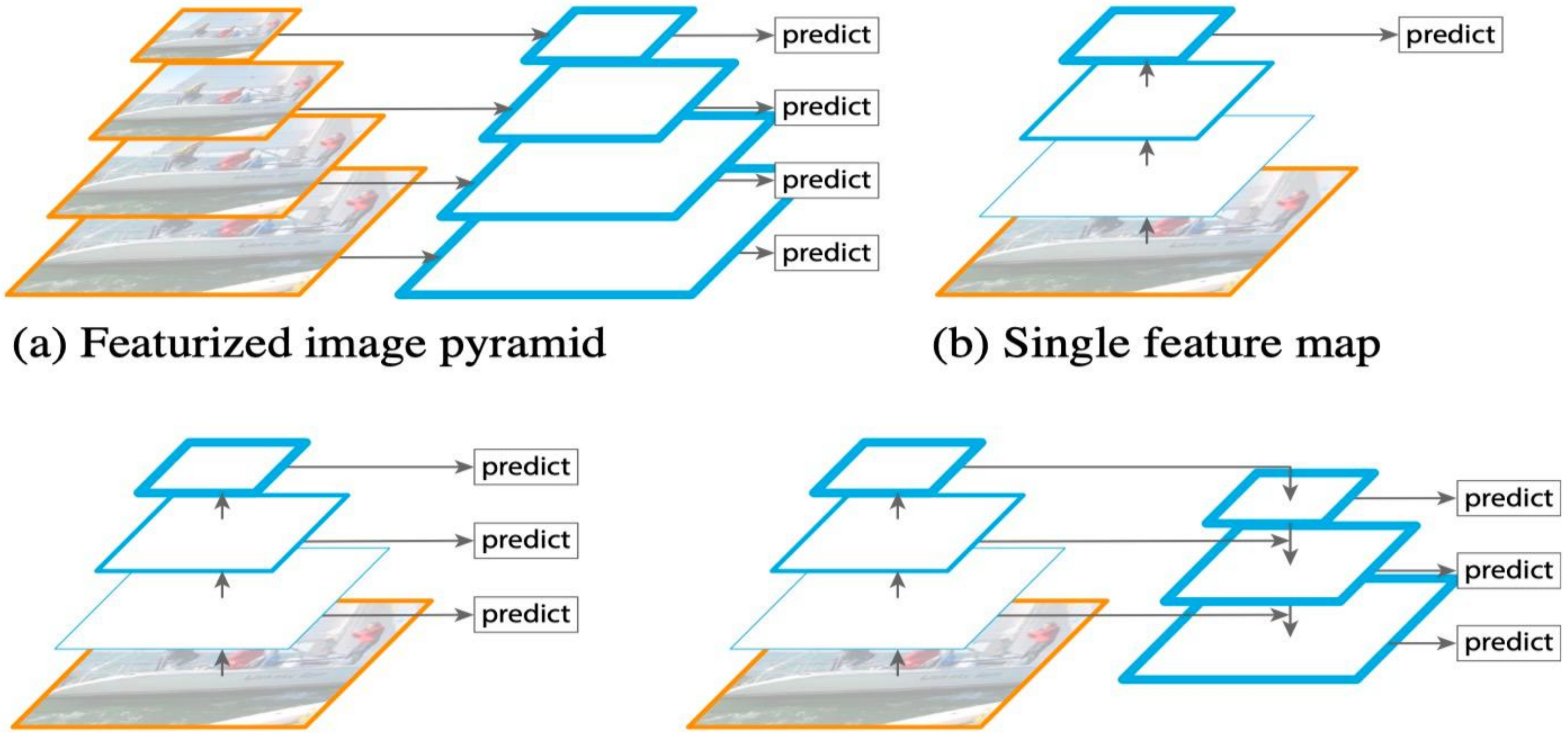

Shaoqing Ren et al., 2015

# R-FCN




Jifeng Dai et al., 2016

# FPN

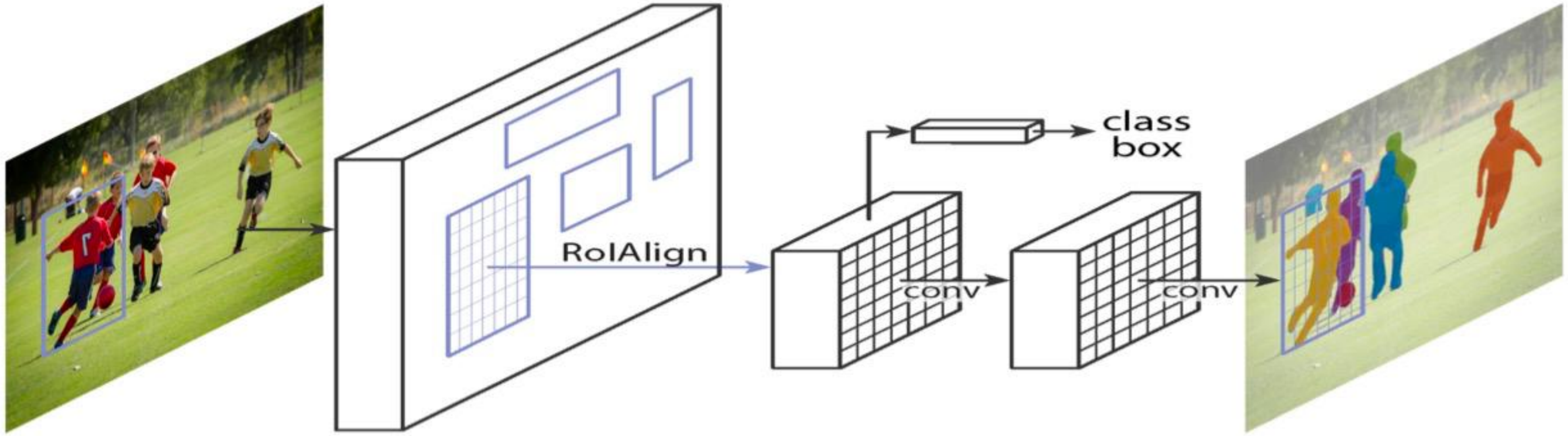



(a) Featurized image pyramid (b) Single feature map

(c) Pyramidal feature hierarchy (d) Feature Pyramid Network

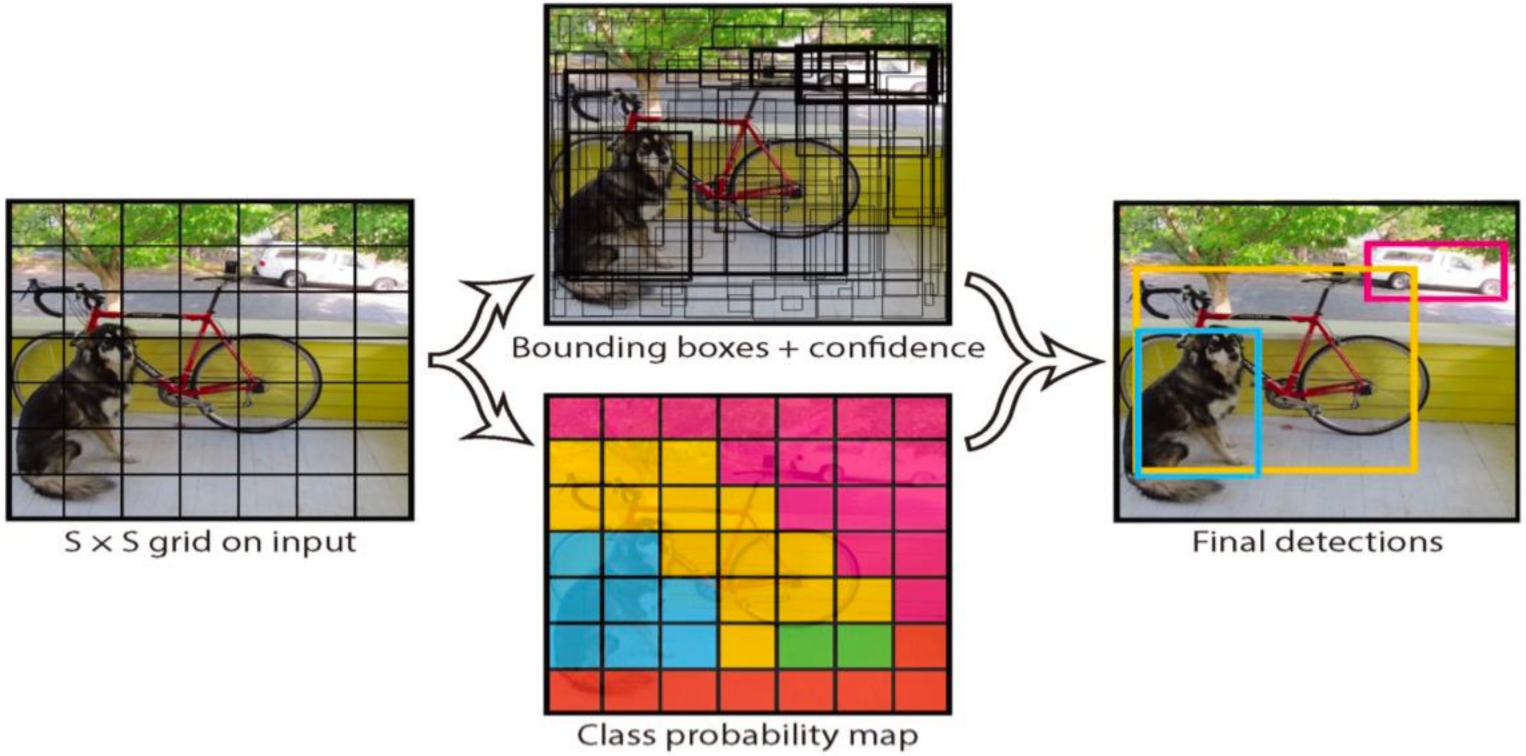
T.Y. Lin et al., 2017

# Mask RCNN



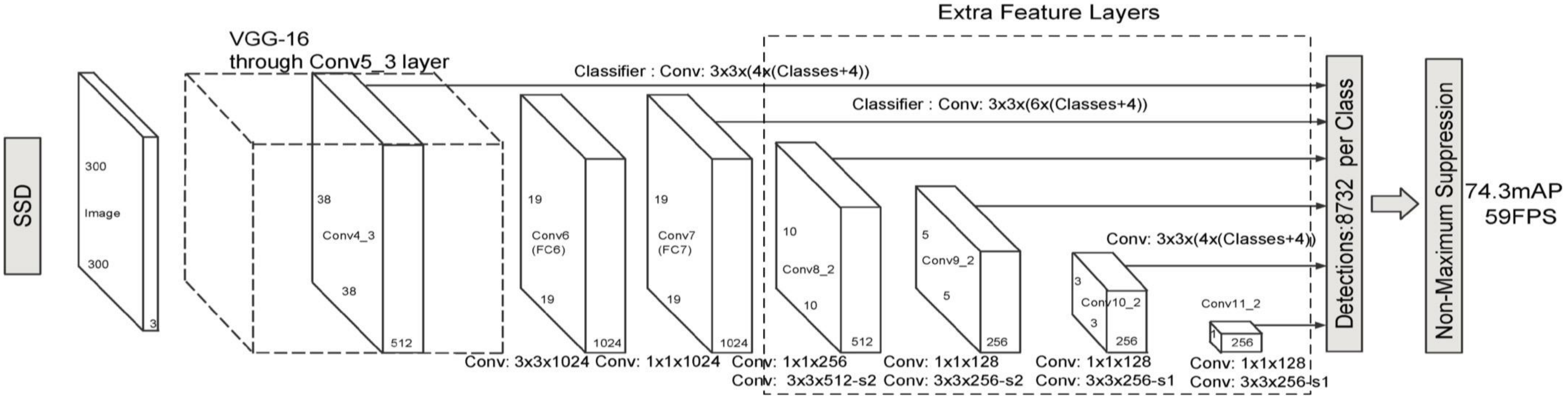

Wei Liu et al., 2016

# YOLO




Joseph Redmon et al., 2016

# SSD




Wei Liu et al., 2016

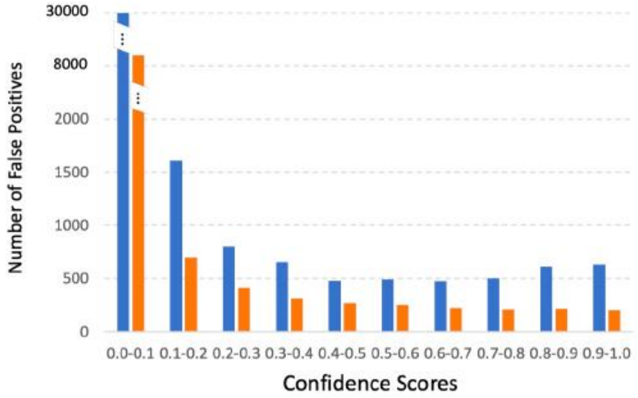
# Detection as a Multitask Learning Problem

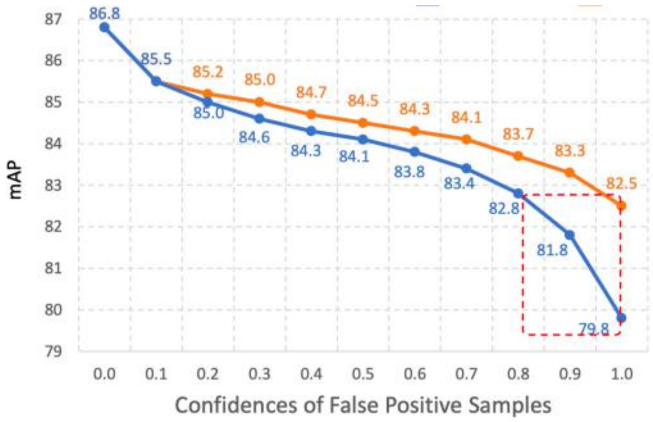
- How to achieve the best result for both localization and classification tasks in object detection?
- DCR as an example

Revisiting RCNN: Awakening the Power of Classification in Faster RCNN, Bowen Cheng, et al., 2018

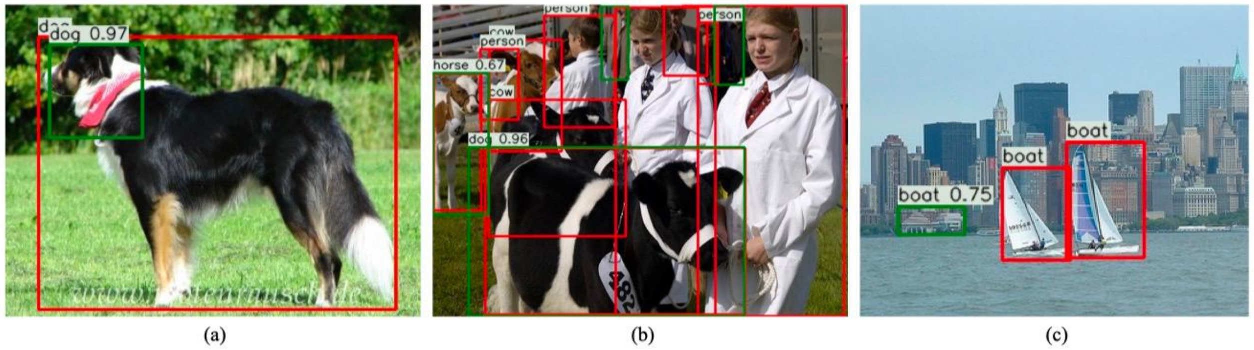
# DCR Motivation

- Question: How can we improve state-of-the-art object detectors?
- Observations with Faster RCNN:




Reducing *hard false positives* (those with high confidence scores) can improve the detection mAP significantly

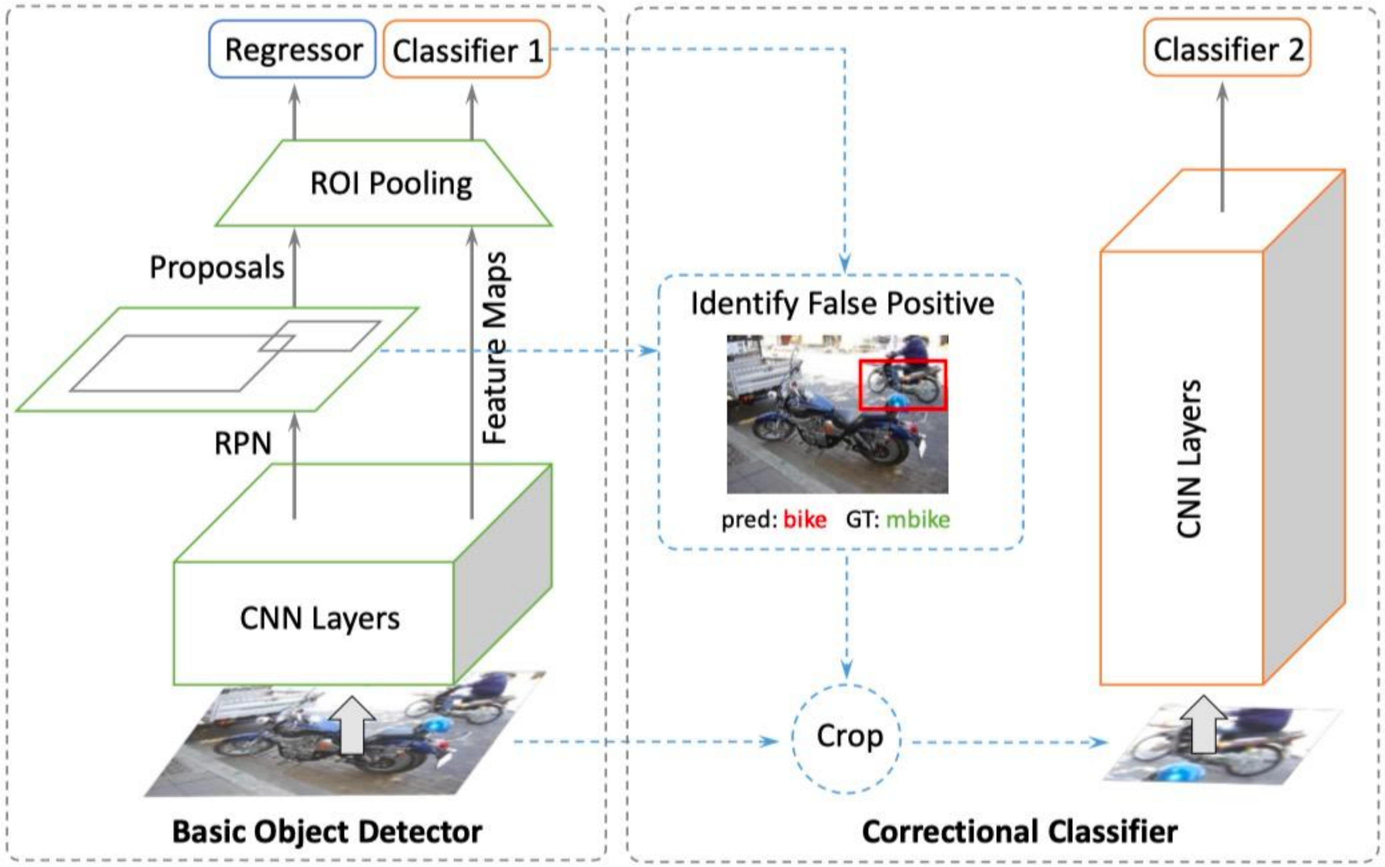



Can we use iterative proposal classification to improve object detection?

Examples of hard false positives

Revisiting RCNN: Awakening the Power of Classification in Faster RCNN, Bowen Cheng, et al., 2018

# DCR Our Approach




Proposed decoupled classification refinement (DCR) module

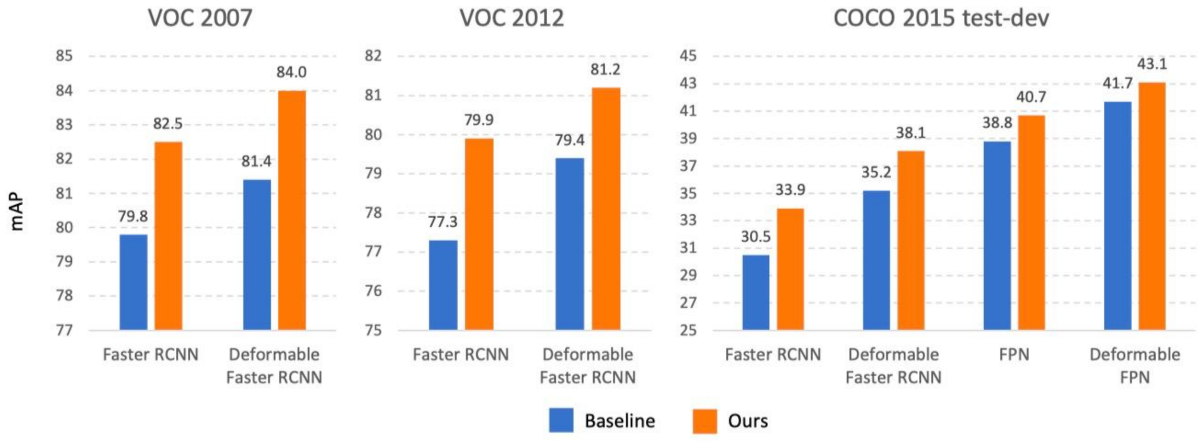
Networks design:

- Decoupled features
- Decoupled optimization
- Adaptive receptive field

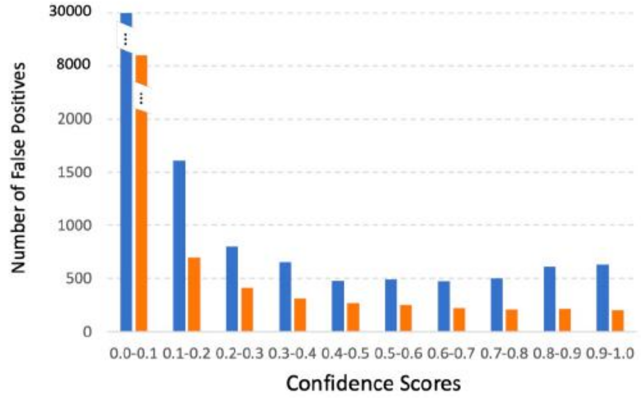
Revisiting RCNN: Awakening the Power of Classification in Faster RCNN, Bowen Cheng, et al., 2018

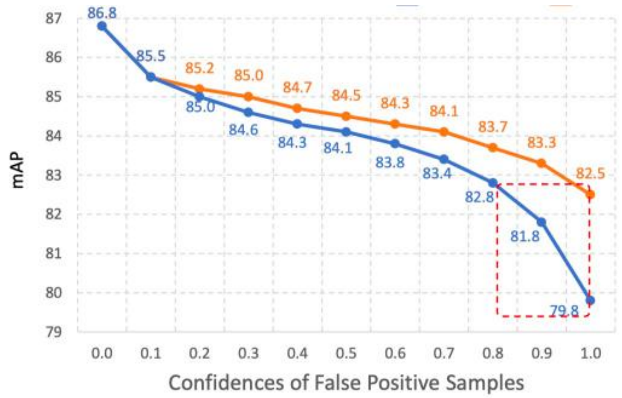
# DCR Results

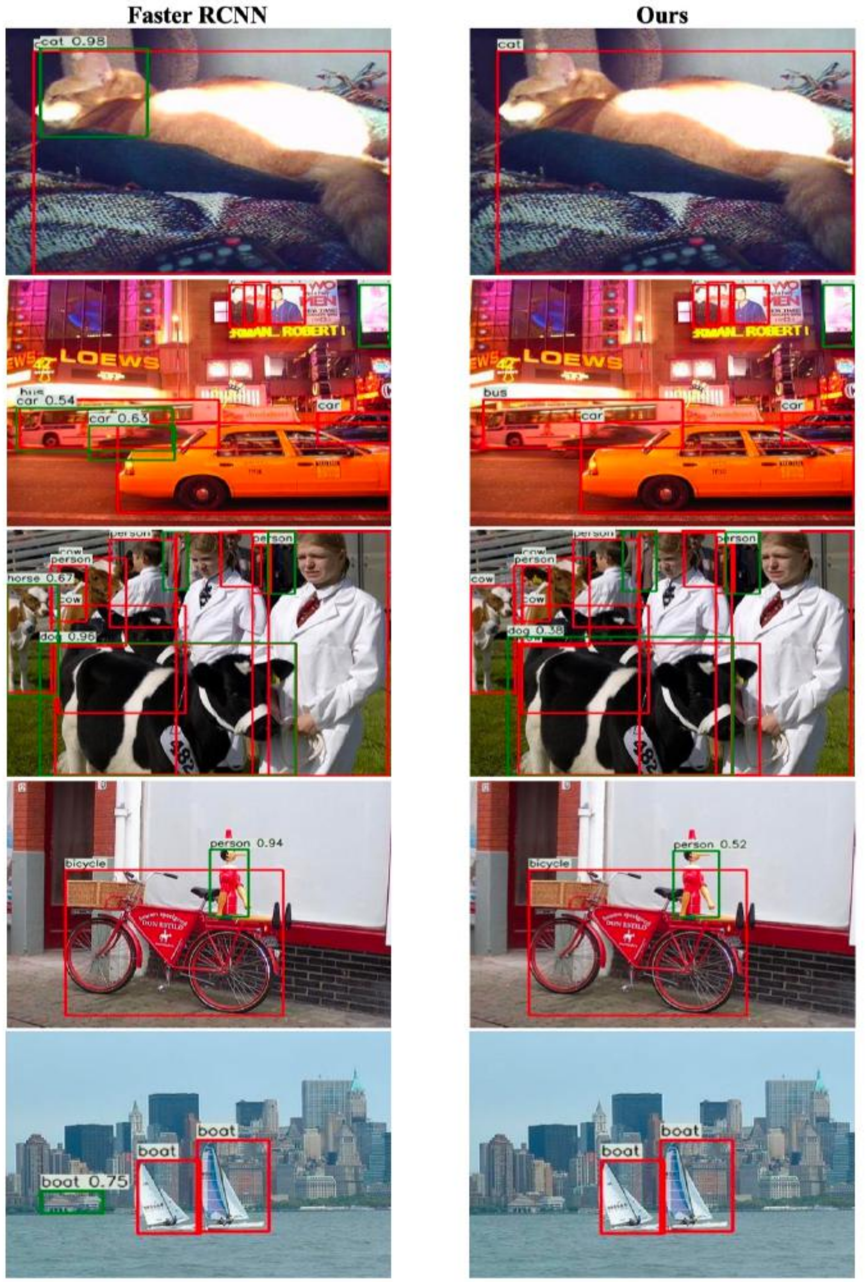
- Results on Pascal VOC & COCO




- How are we doing on false posotiveis?



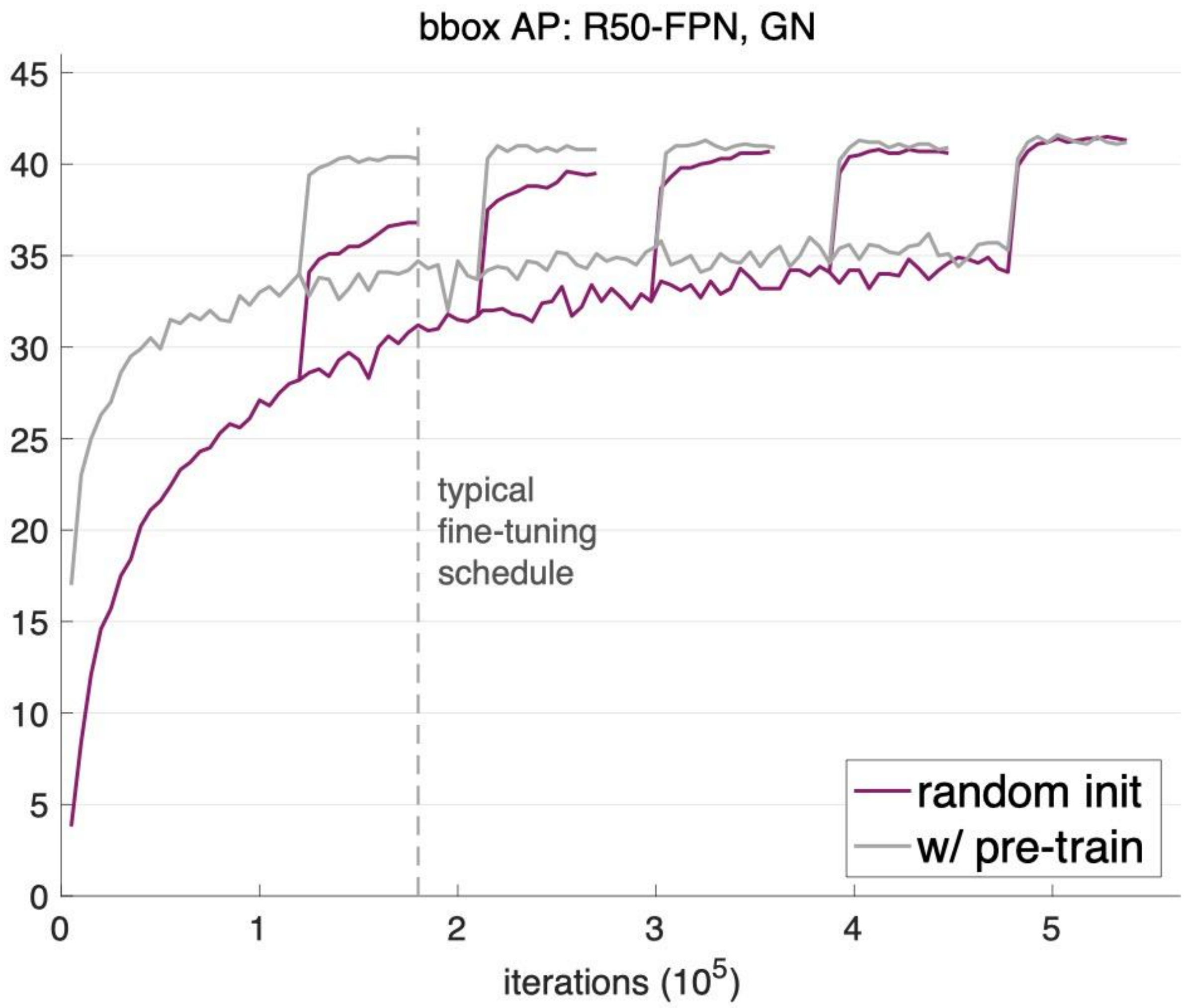




Revisiting RCNN: Awakening the Power of Classification in Faster RCNN, Bowen Cheng, et al., 2018

# On Pre-training: Use ImageNet or from Scratch

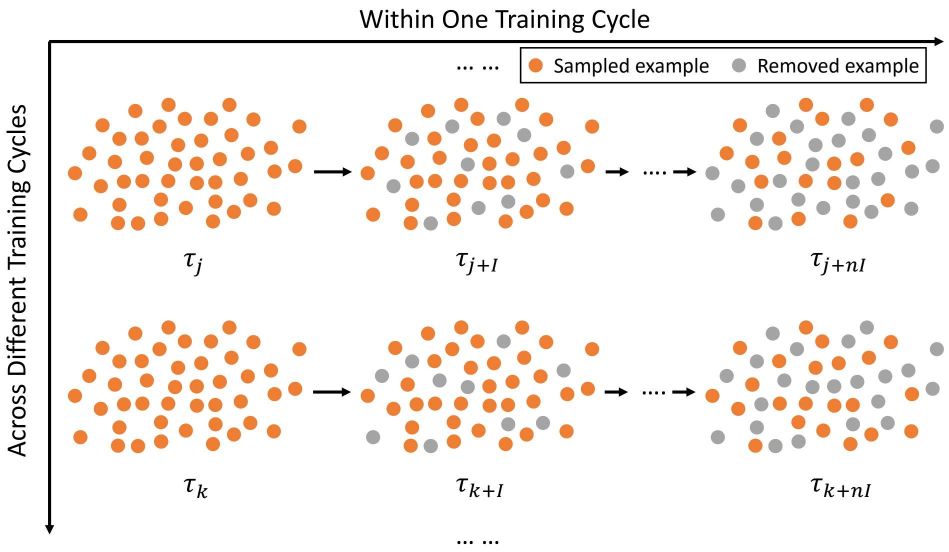
- Detectors without ImageNet pre-training can be trained as good as those with
  - When dataset is large
  - Use more iterations
  - Use initialization/normalization techniques
- What does this imply?




Rethinking ImageNet Pre-training, Kaiming He et al., 2018

# On Pre-training: Knowledge Transferability

- Training can be more efficient
- Specialized knowledge learned on ImageNet pre-training can be effectively transferred to downstream tasks.




| Method | Backbone | Pre-train Computation | $AP^{bb}$ | $AP^{bb}_{50}$ | $AP^{bb}_{75}$ | $AP^{bb}_{S}$ | $AP^{bb}_{M}$ | $AP^{bb}_{L}$ | $AP^{m}$ | $AP^{m}_{50}$ | $AP^{m}_{75}$ | $AP^{m}_{S}$ | $AP^{m}_{M}$ | $AP^{m}_{L}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FPN [15] | ResNet-50 | 100% | 35.8 | 57.7 | 38.2 | 20.2 | 39.5 | 45.9 | - | - | - | - | - | - |
| FPN [15] | ResNet-50 | 75% | 35.9 | 57.8 | 38.4 | 21.1 | 39.7 | 46.6 | - | - | - | - | - | - |
| FPN [15] | ResNet-50 | 86% | 36.1 | 57.9 | 38.9 | 20.6 | 39.9 | 46.9 | - | - | - | - | - | - |
| Mask R-CNN [7] | ResNet-50 | 100% | 36.6 | 58.0 | 39.5 | 20.8 | 40.1 | 47.6 | 33.5 | 54.8 | 35.4 | 17.0 | 36.8 | 46.0 |
| Mask R-CNN [7] | ResNet-50 | 75% | 36.8 | 58.4 | 39.7 | 21.4 | 40.5 | 47.7 | 33.8 | 55.3 | 35.8 | 17.7 | 37.2 | 46.3 |
| Mask R-CNN [7] | ResNet-50 | 86% | 36.9 | 58.5 | 39.9 | 21.0 | 40.5 | 47.8 | 33.8 | 55.1 | 35.9 | 17.5 | 37.1 | 46.2 |

Revisiting Pre-training: An Efficient Training Method for Image Classification, Bowen Cheng, et al., 2018

# Questions and Contact

shihonghui3@gmail.com