Deep Object Detection summary

V. Nousi, E. Patsiouras, A. Tefas, I. Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 3.8

Artificial Intelligence &
Information Analysis Lab

Object Detection for UAV sports cinematography

Artificial Intelligence & Information Analysis Lab

Object Detection for UAV sports cinematography

Target/object examples: athletes, boats, bicycles.

Object Detection

- Object detection = classification + localization:
- Find ***what*** is in a picture as well as ***where*** it is.

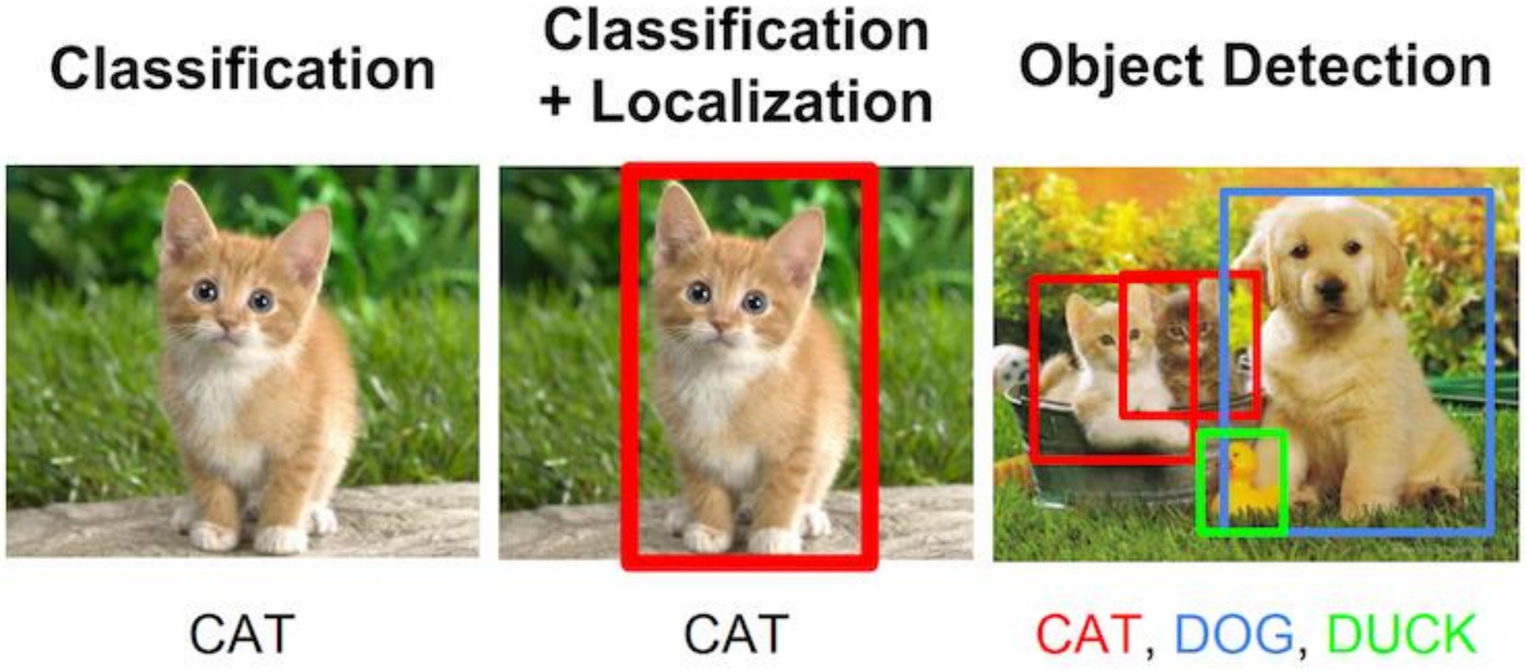
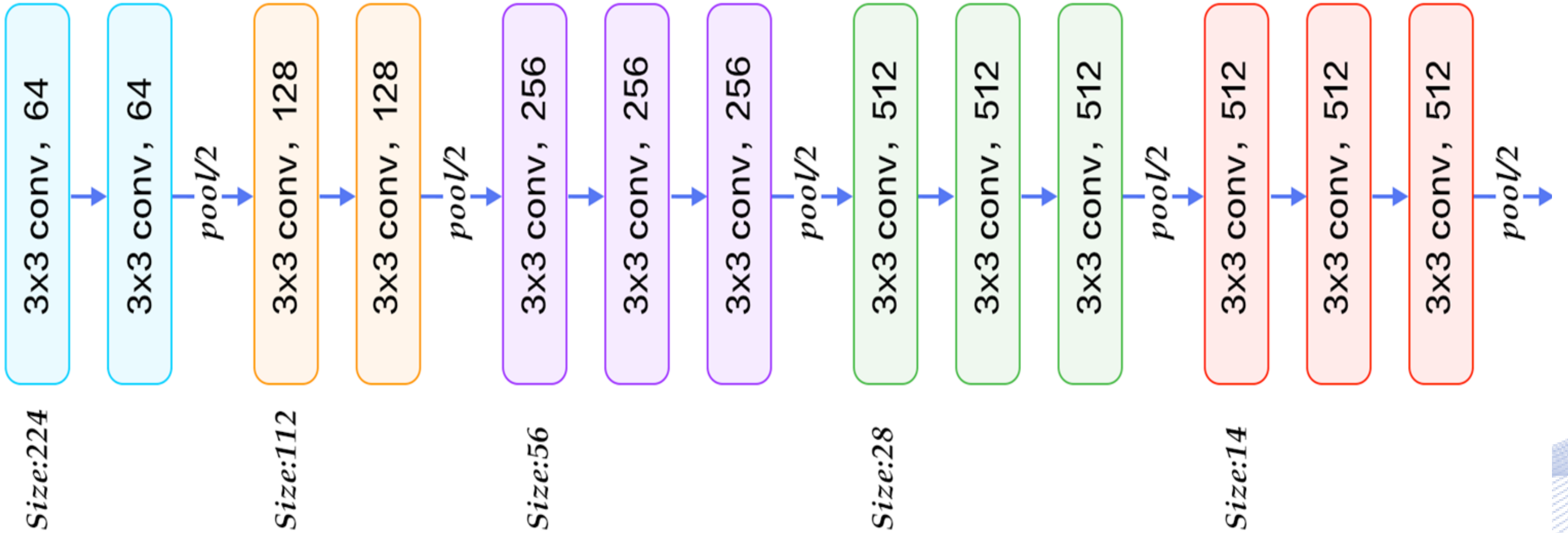


Figure: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

Object Detection with CNNs



Object detection: CNN pipeline for bounding box regression.

CNN Object Detection

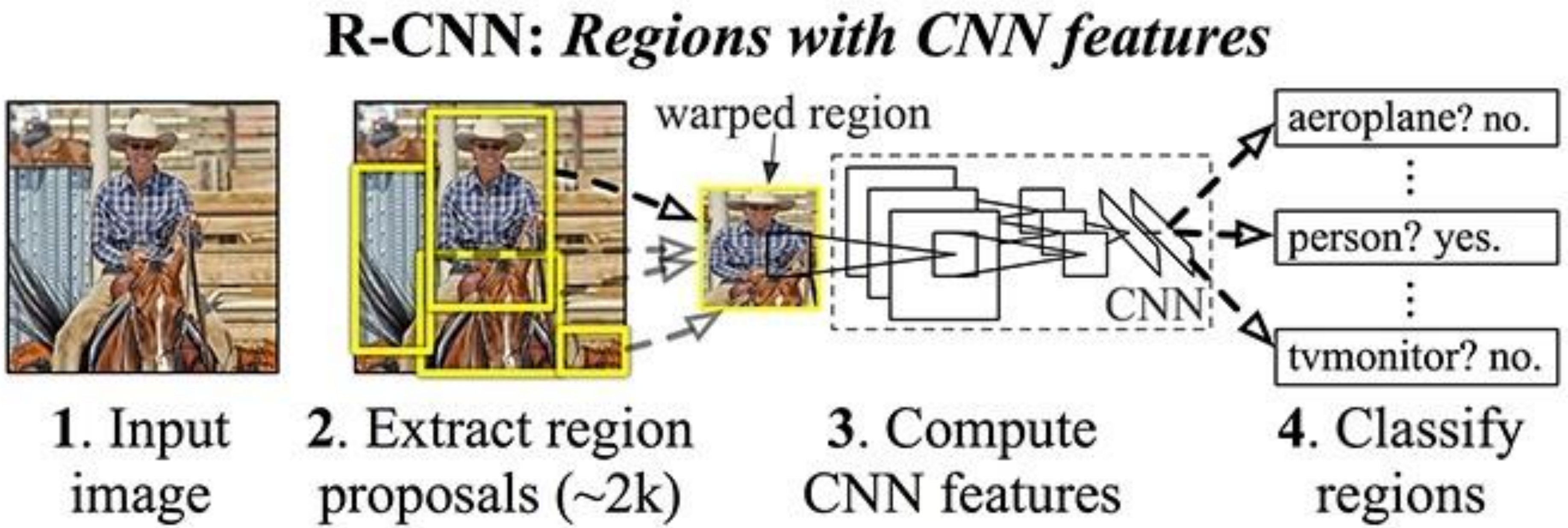
Region proposal-based detectors

- R-CNN, Fast R-CNN, Faster R-CNN
- R-FCN

Single Stage Detectors

- YOLO
- SSD
- YOLO v2, v3, v4
- RetinaNet, RBFnet
- CornerNet, CenterNet
- DETR.

R-CNN

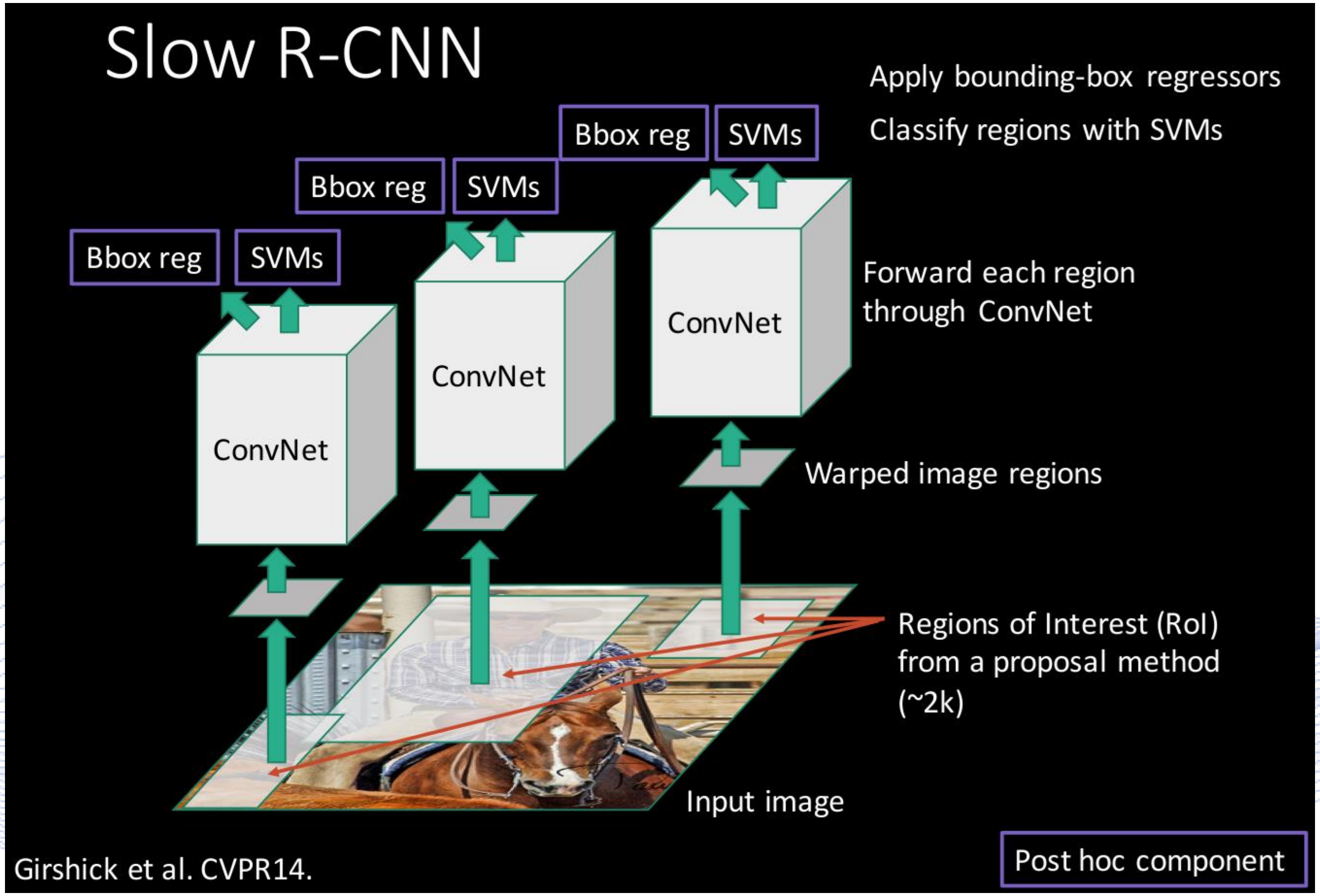


[GIR2014]

R-CNN

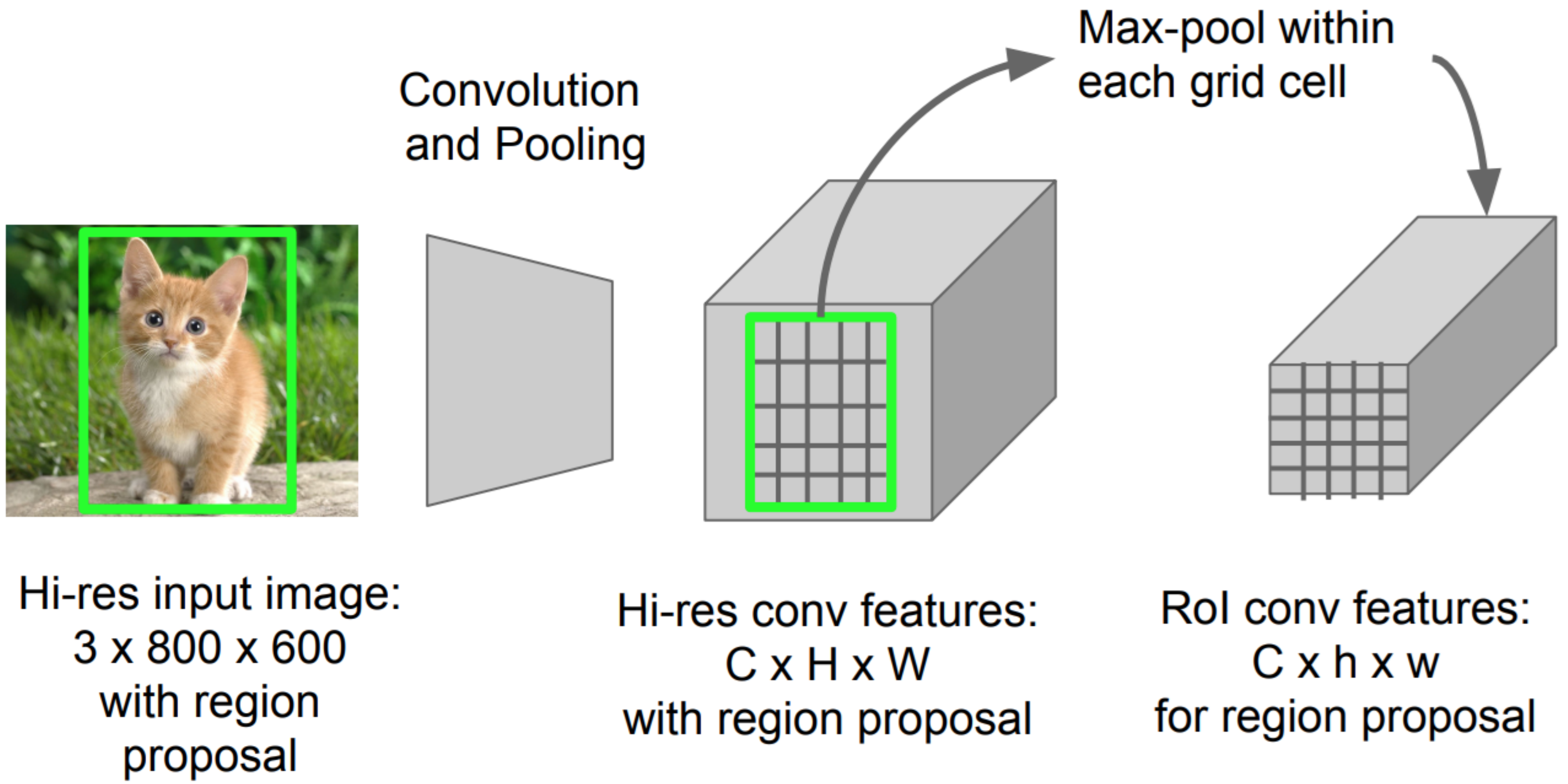
Selective Search: start from oversegmentation, merge similar regions.

R-CNN



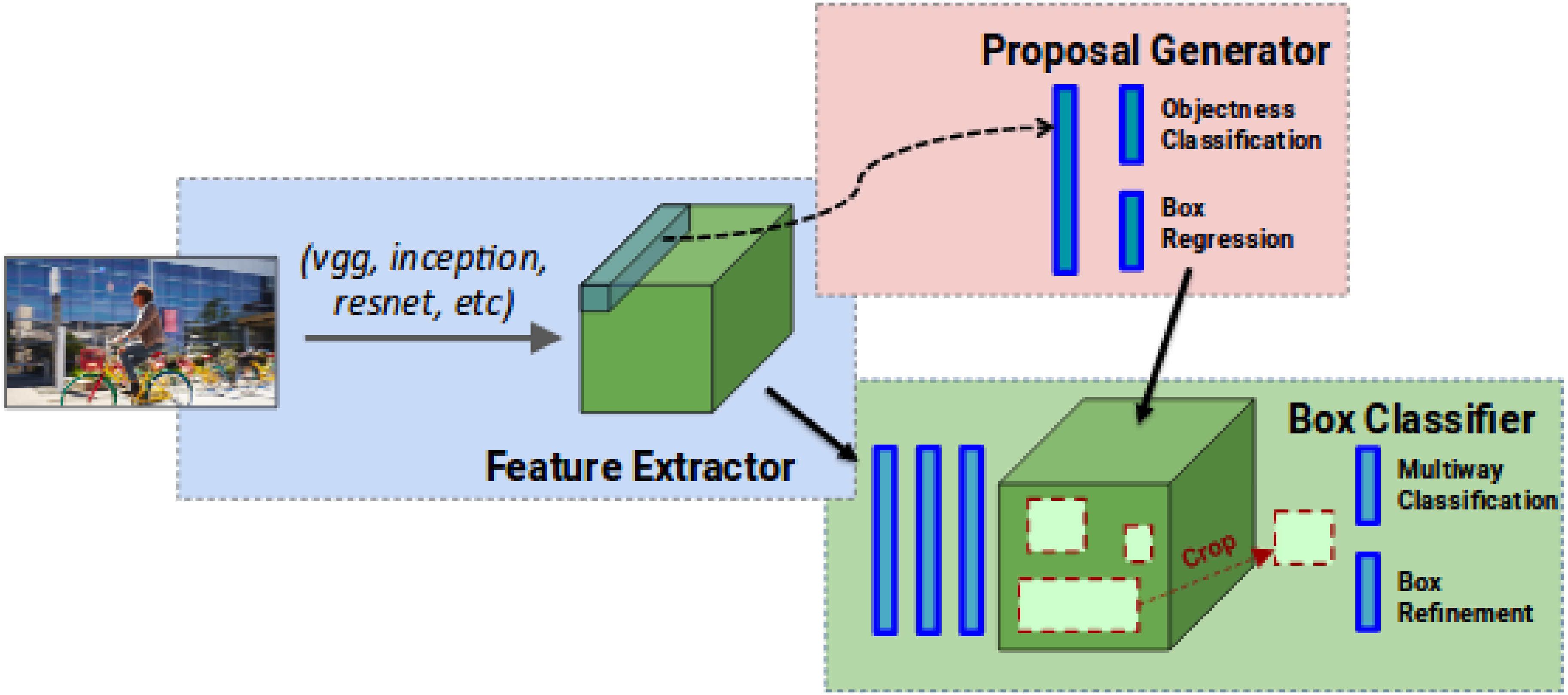
Slow R-CNN
Apply bounding-box regressors
Classify regions with SVMs
Bbox reg
SVMs
Bbox reg
SVMs
Bbox reg
SVMs
ConvNet
ConvNet
ConvNet
Forward each region
through ConvNet
Warped image regions
Regions of Interest (RoI)
from a proposal method
(~2k)
Input image
Girshick et al. CVPR14.
Post hoc component

Fast R-CNN



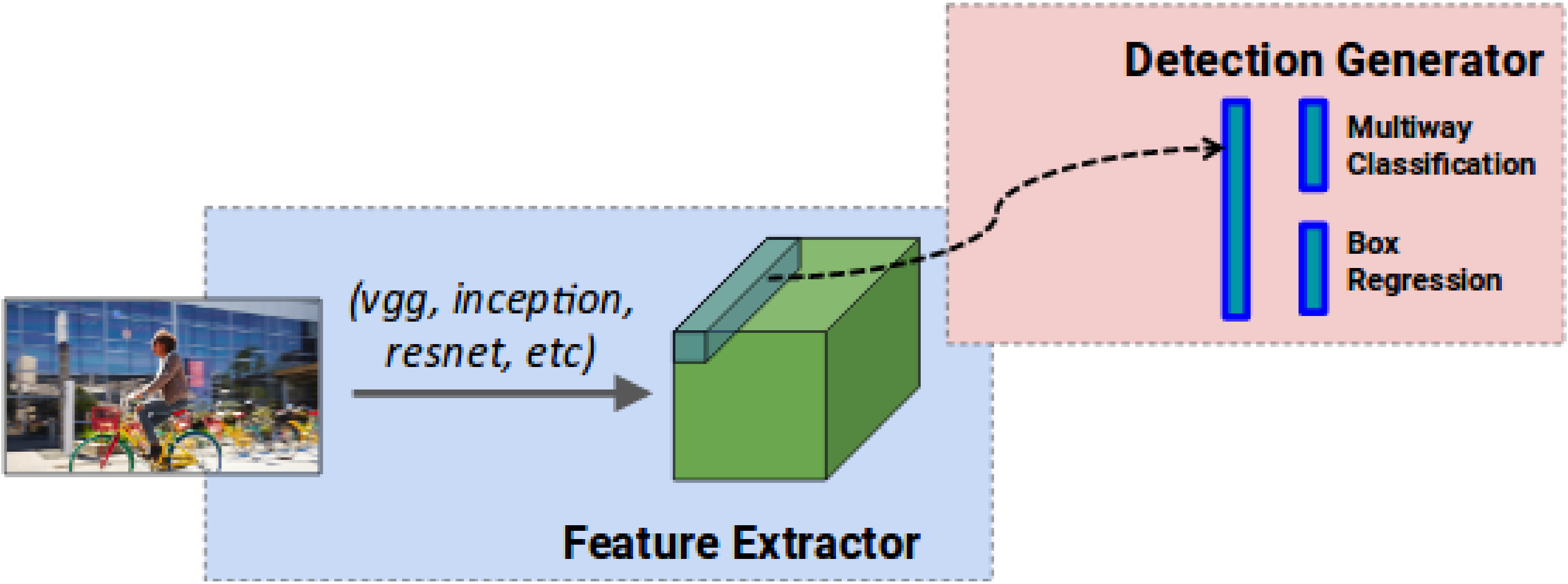
ROI pooling.

R-FCN



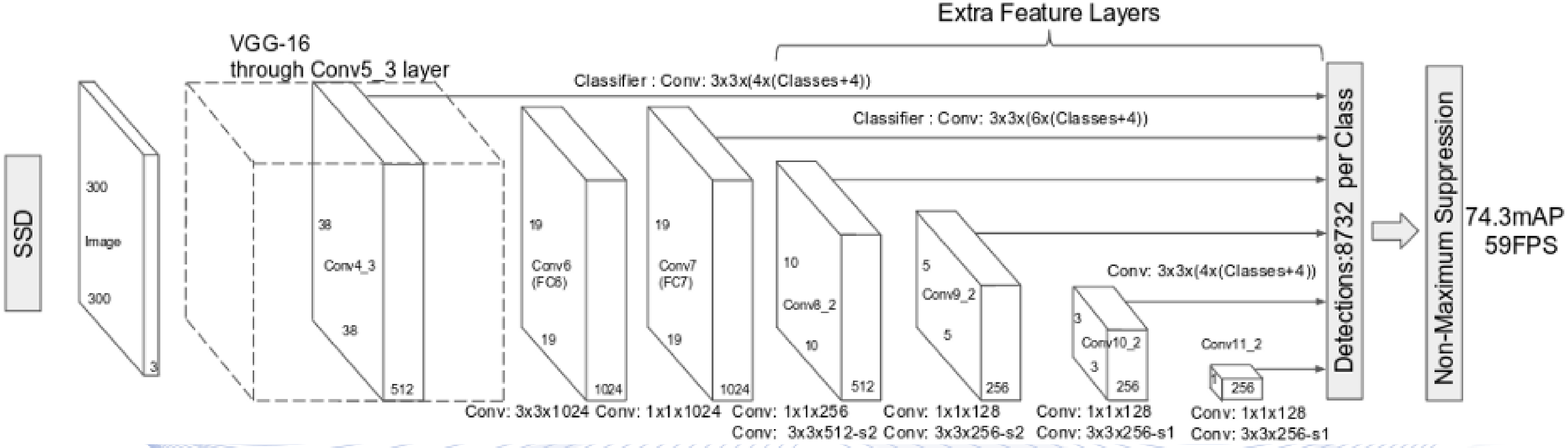
[HUA2017]

SSD

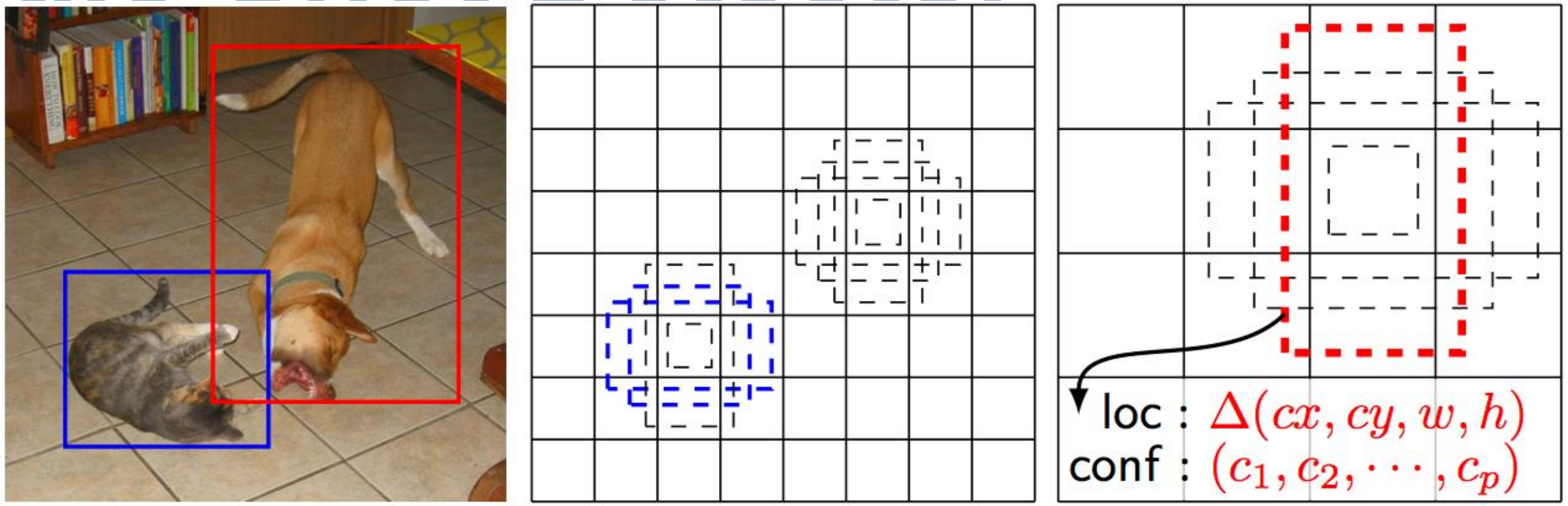


SSD architecture [HUA2017].

Single Shot Detector



Single Shot Detector



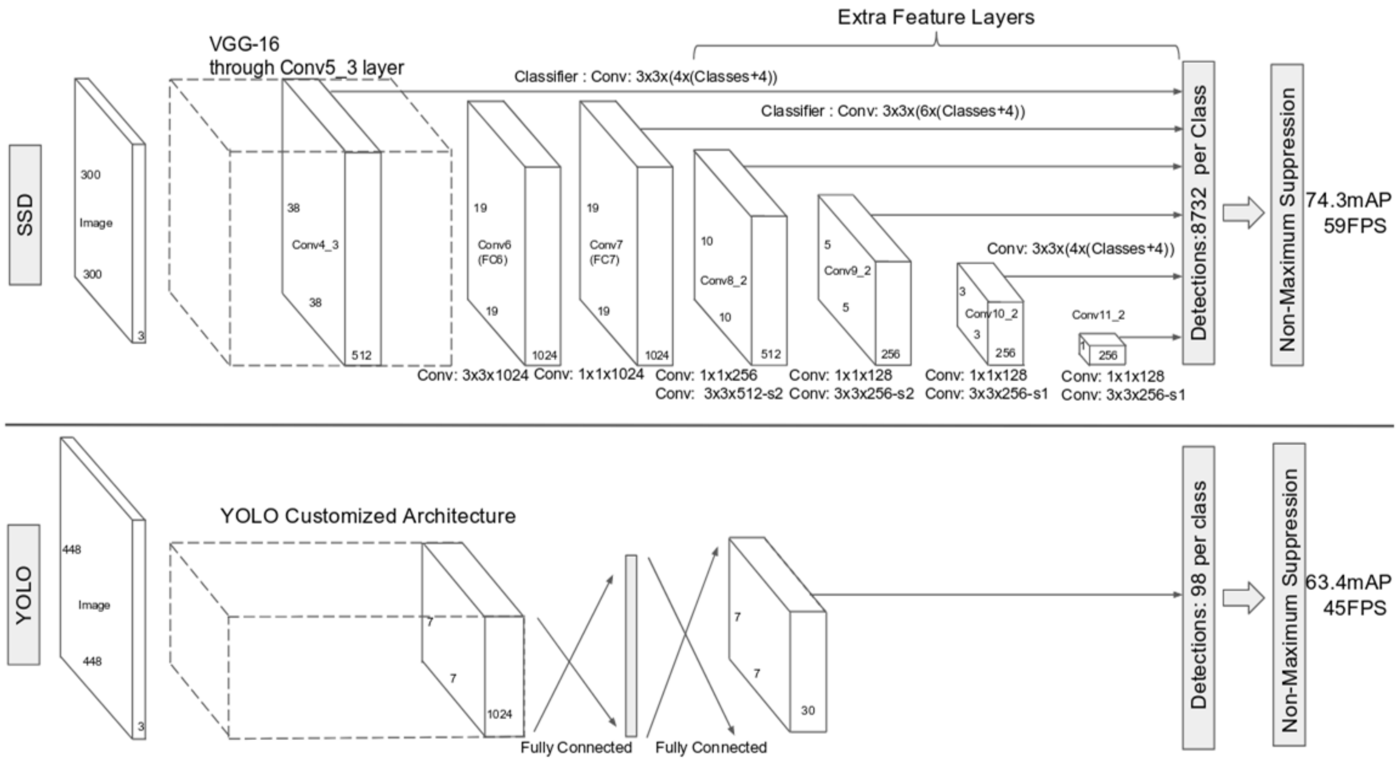
(a) Image with GT boxes (b) 8×8 feature map (c) 4×4 feature map

- Example: The cat has 2 anchors (ROIs) that match on the 8×8 feature map, but none match the dog.
- On the 4×4 feature map there is one anchor that matches the dog and it is refined.

YOLO

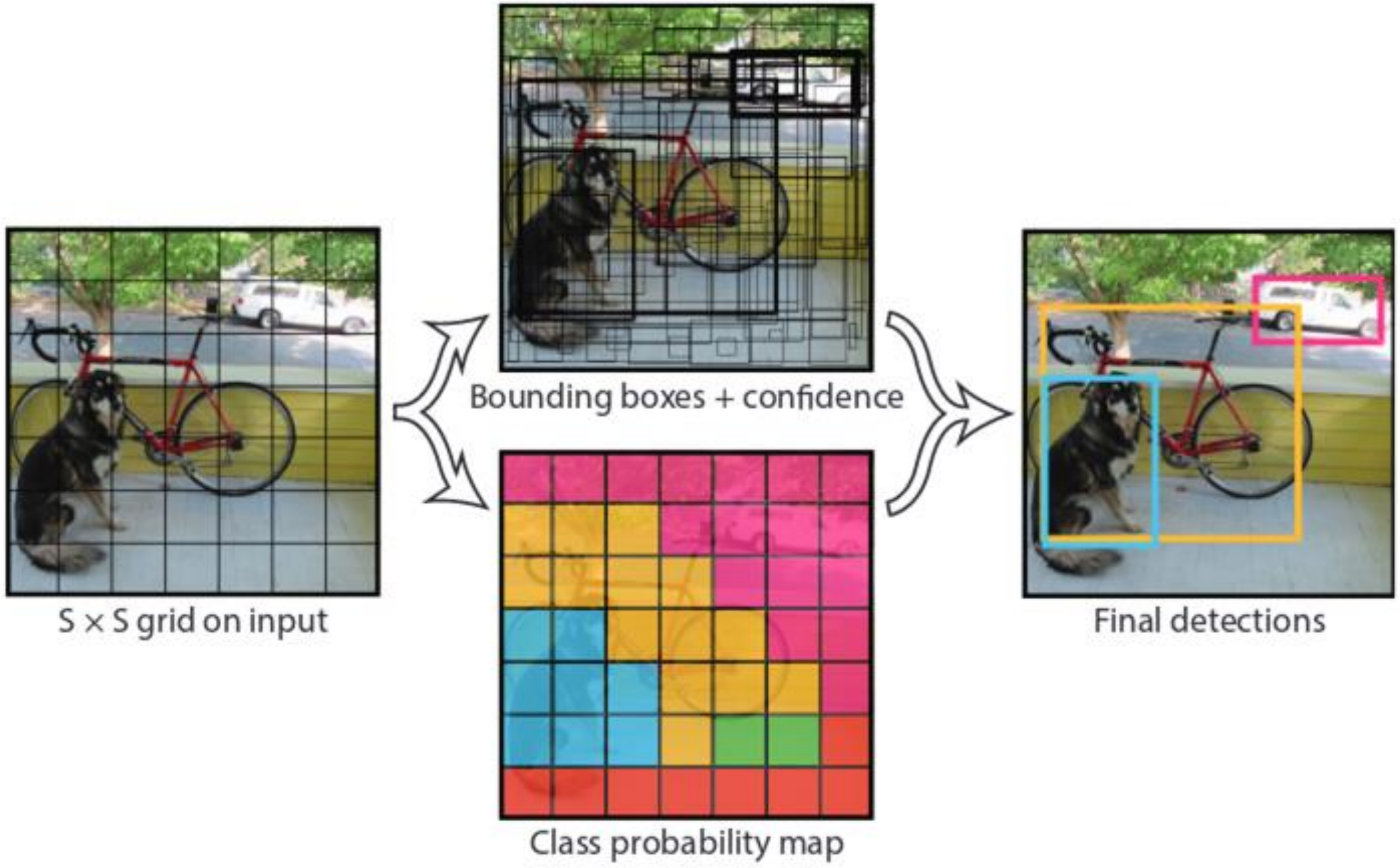
YOLO (***You Only Look Once***) architecture:

- Darkenet19 convolutional network plus FC layer.
- Prediction only at the final convolutional feature map.



[LIU2016]

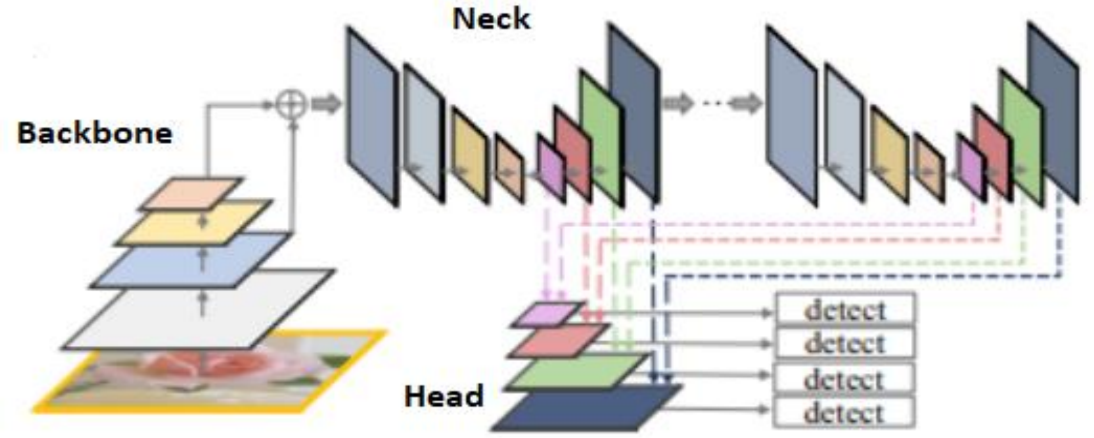
YOLO



[RED2016]

YOLO v4

YOLO v4 design:

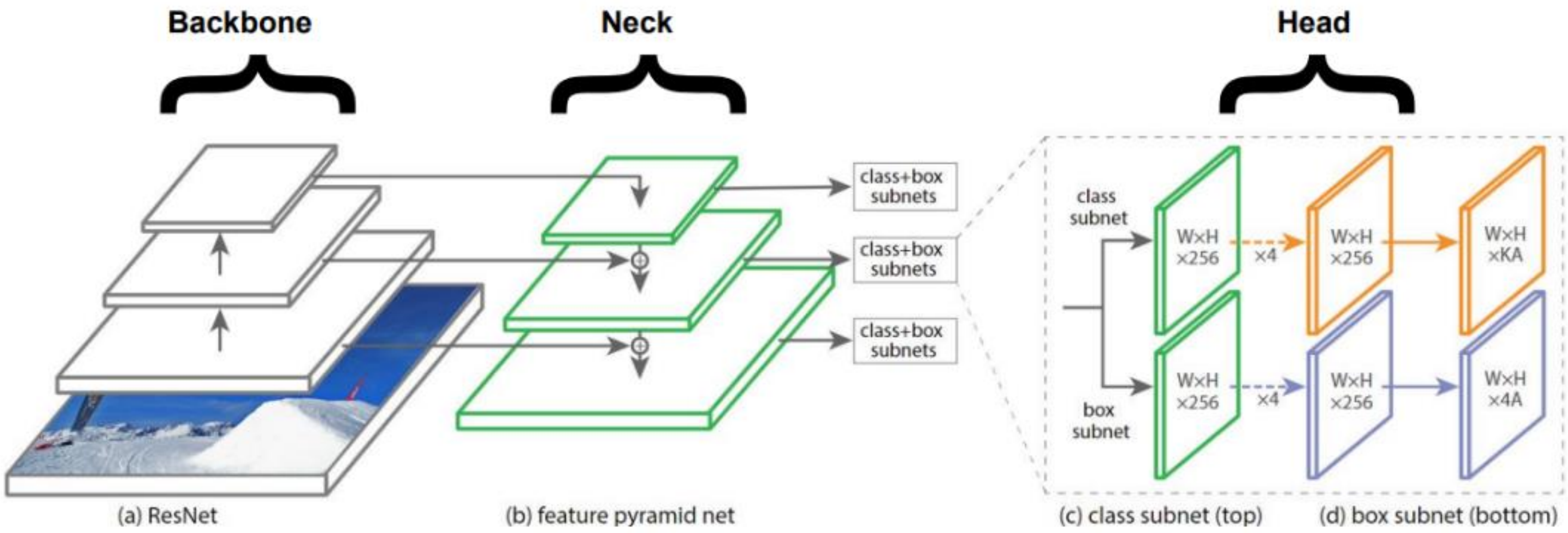


[BOC2020]

- **Backbone:** CSPDarknet53.
- **Neck:** Spatial pyramid pooling (SPP) and Path Aggregation Network (PAN).
- **Head:** Same as YOLO v3.

RetinaNet

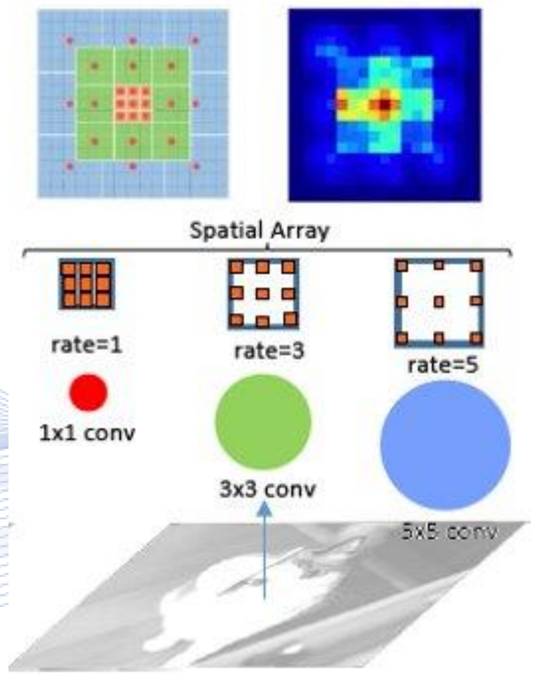
- ResNet is used as a backbone for feature extraction.
- ***Feature Pyramid Network (FPN)*** is used as a neck on top of ResNet for constructing a rich multi-scale feature pyramid from one single resolution image.



Architecture of RetinaNet [LIN2017].

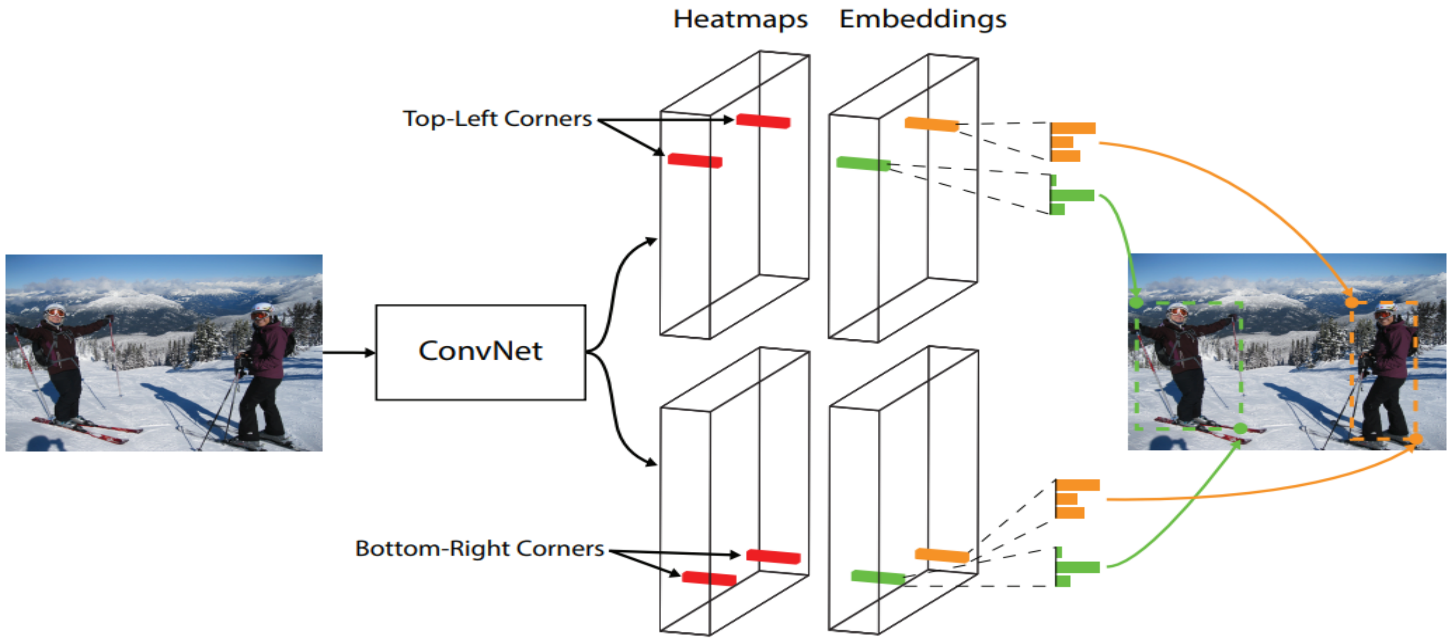
RFBNet

- It inspired by the structure of receptive fields in human visual system [LIU2018].
- Use of multiple dilated convolutions with different kernel sizes in each convolutional layer.
- State-of-the-art results and fast inference time.



CornerNet

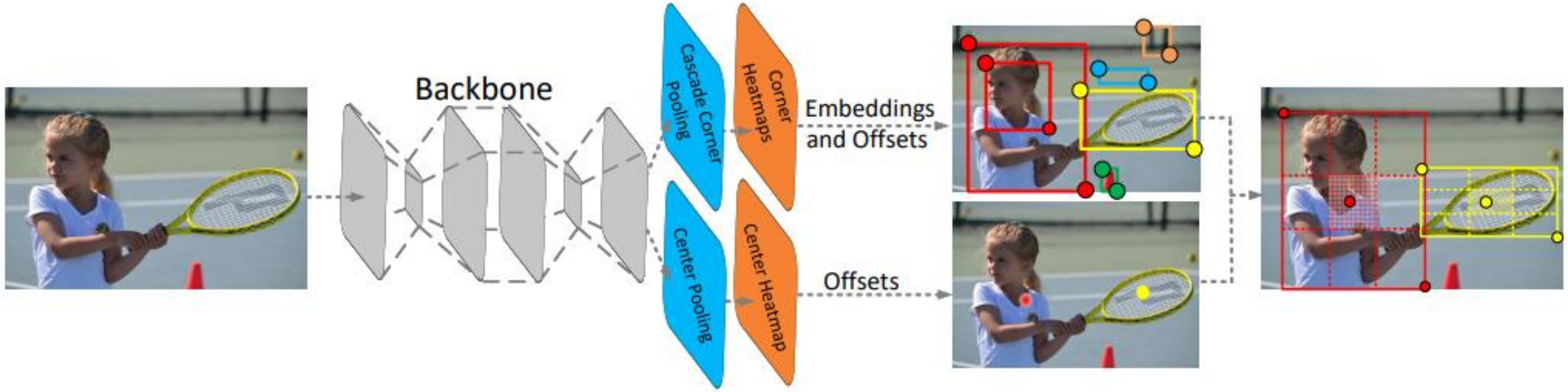
- Each set of heatmaps has C channels and is of size $h \times w$ pixels:
  - C: number of categories to detect.



Pipeline of CornerNet object detection [LAW2018].

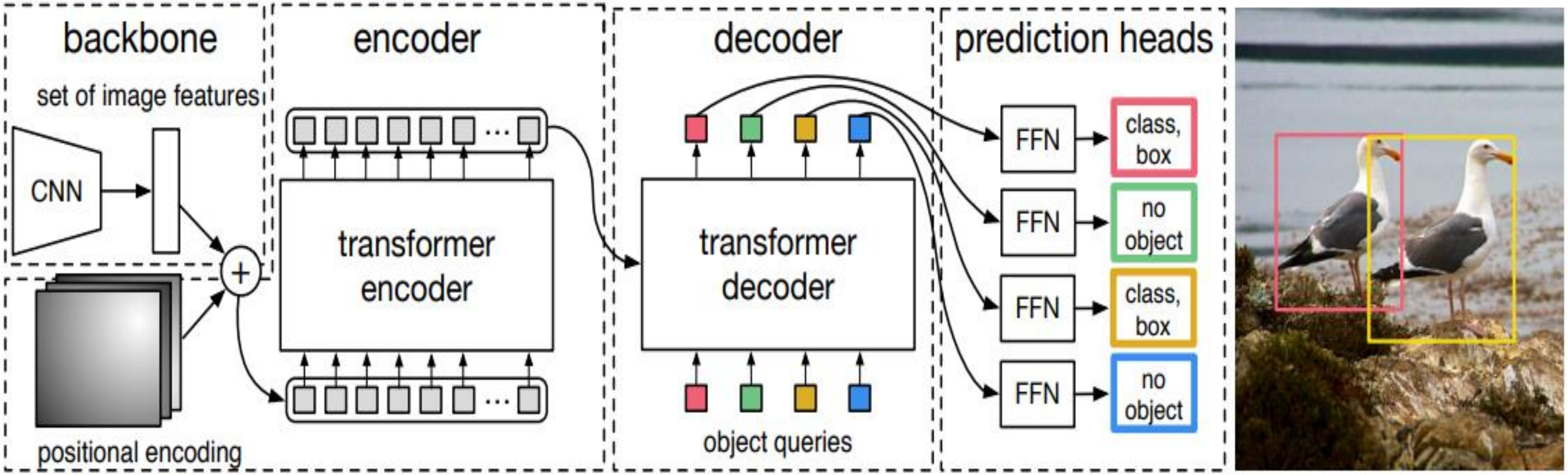
CenterNet

- A CNN backbone applies cascade cornel pooling and center pooling in order to output two corner heatmaps and a center keypoint heatmap, respectively.



Architecture of CenterNet. [DUA2019].

DETR



DETR architecture [CAR2020].

Using object detectors for drone-based shooting

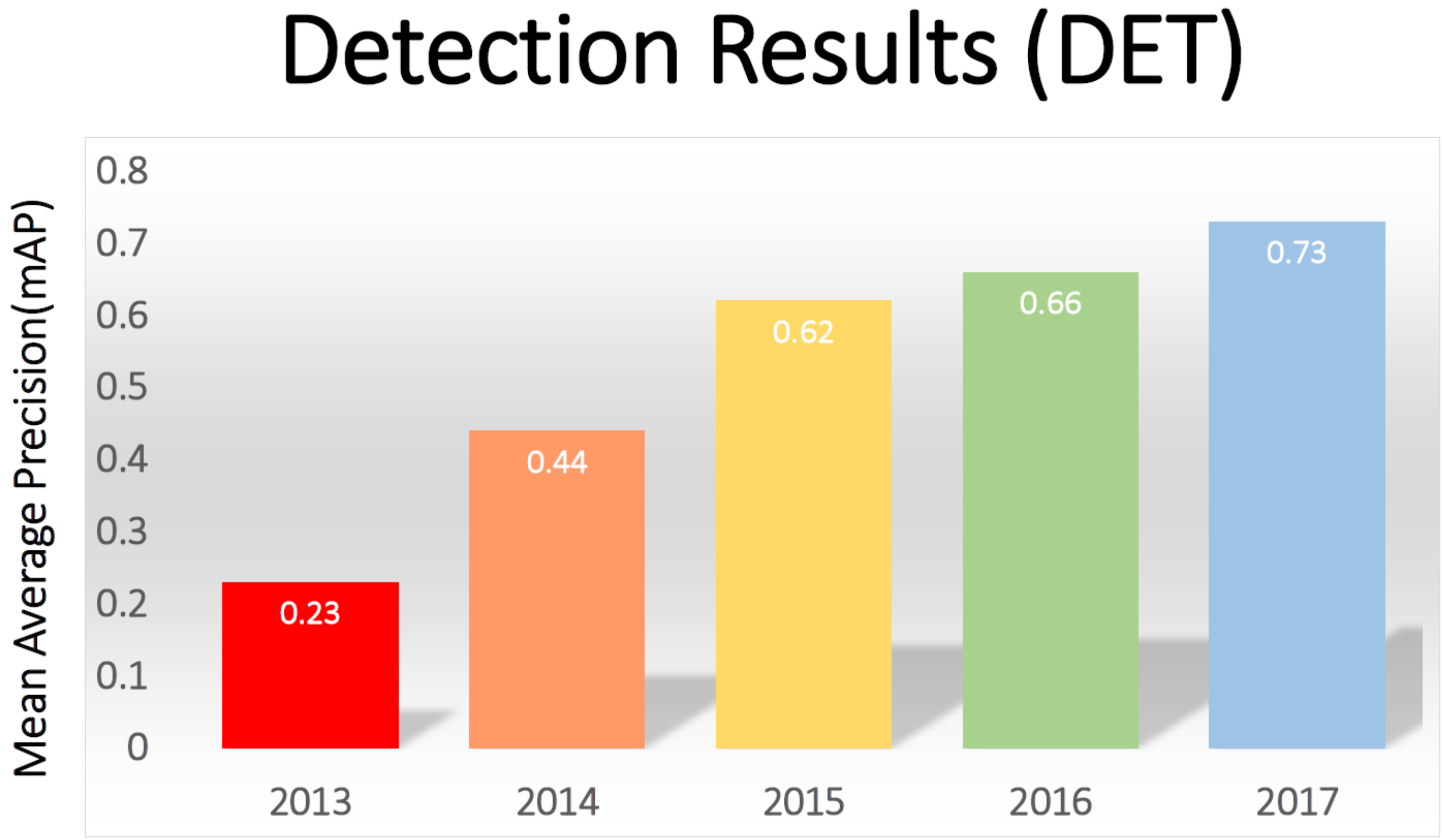
- **Reducing the input image size can also increase the detection speed**
  - However, this can **significantly impact the accuracy** when detecting very small objects (which is the case for drone shooting)

Model	Input Size	Pascal 2007 test mAP*
YOLO v.2	544x544	77.44
YOLO v.2	416x416	74.60
YOLO v.2	288x288	67.12
YOLO v.2	160x160	48.72
YOLO v.2	128x128	40.68

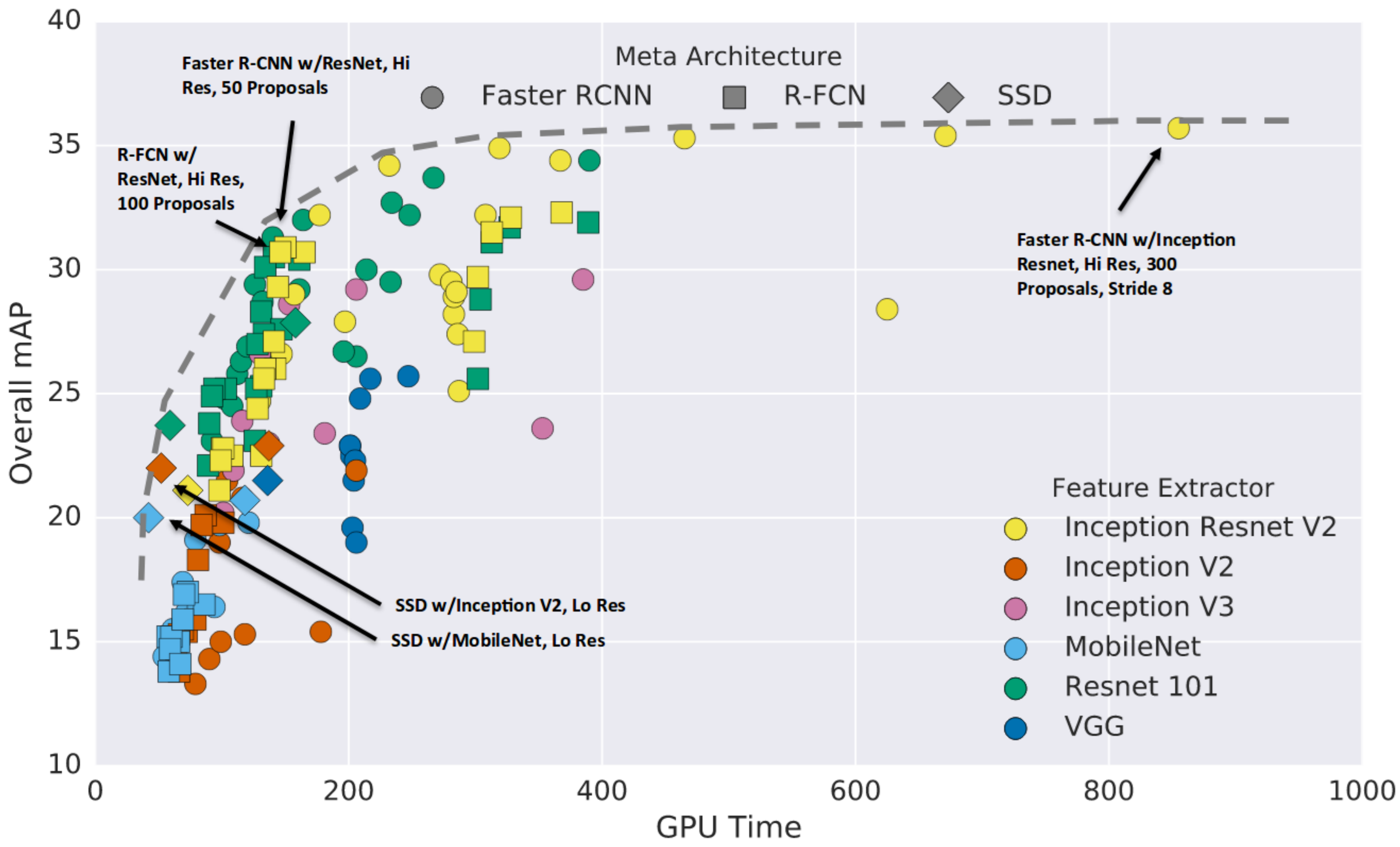
Object Localization Performance Metrics

Object detection: a) $J(\mathcal{A},\mathcal{B}) = 0.67$; b) $J(\mathcal{A},\mathcal{B}) = 0.27$.

Object Detection Performance Metrics



CNN comparison



[HUA2017]

Input Size NxN

[HUA2017]

Artificial Intelligence &
Information Analysis Lab

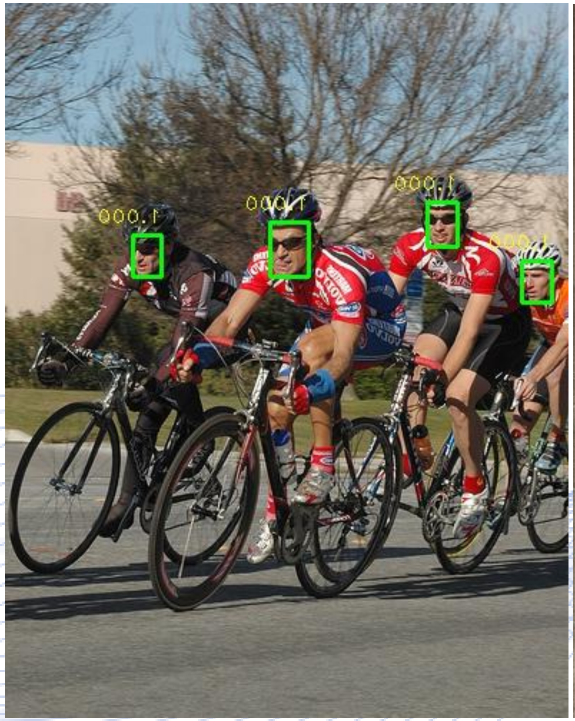
CNN comparison

- **Faster R-CNN is more accurate but slower.**
- **YOLO, SSD are much faster** but not as accurate.
- YOLO, SSD make **more mistakes when objects are small** and have trouble correctly predicting the exact location of such objects.

Object detection acceleration

- Examples of acceleration techniques:
  - Input size reduction.
  - Specific object detection instead of multi-object detection.
  - Parameter reduction.
  - Post-training optimizations with TensorRT (NVIDIA), including FP16 (floating point 16 bit) computations.

Face detection examples


Object Detection for UAV powerline inspection

Insulator, dumper & tower detection.

Q & A

Thank you very much for your attention!

Contact: Prof. I. Pitas
pitas@csd.auth.gr