Object Detection Introduction

Presentation by Jamyoung Koo

2019.03.20

주식회사 에스아이에이
SI ANALYTICS

Contents

- Object Detection using CNN

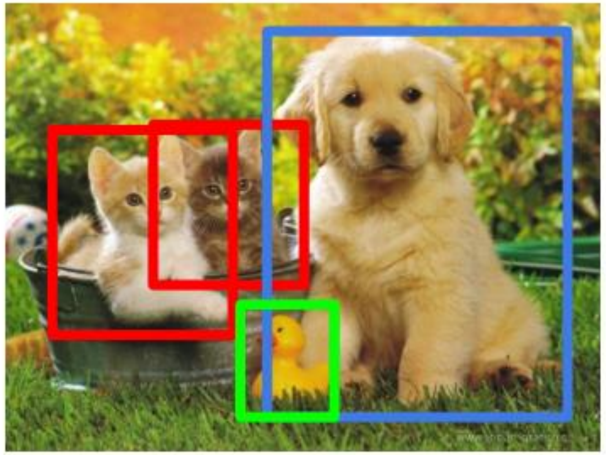
  1. Basics

  2. Recent Approach

  3. In Remote Sensing

01 Basics

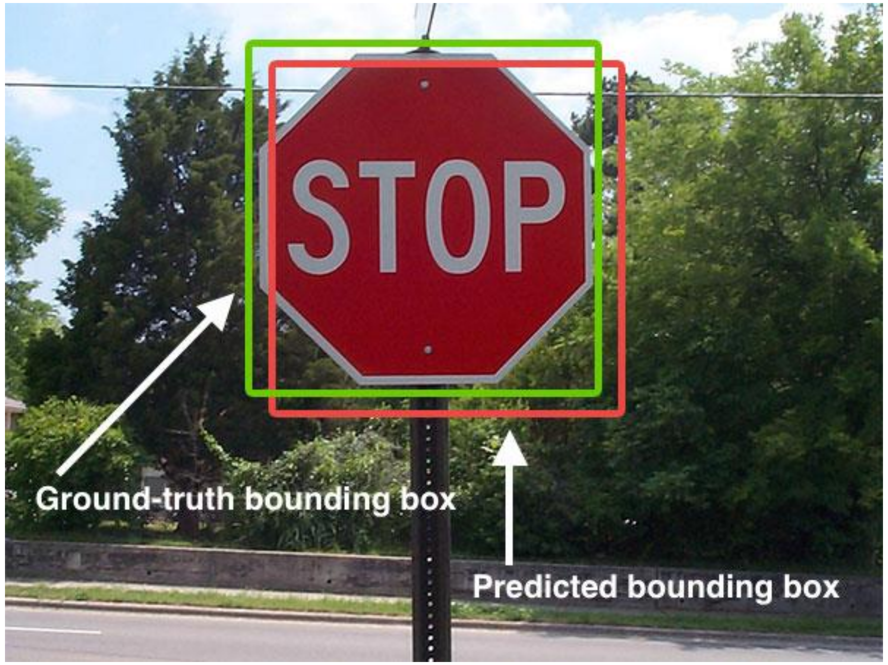
Terms: Object Detection



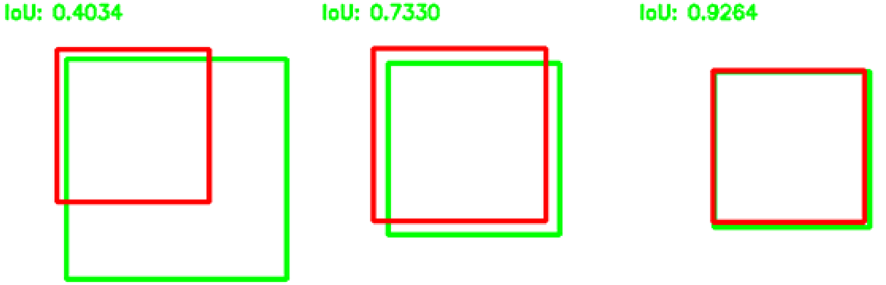
CAT, DOG, DUCK

The task of assigning a **label** and a **bounding box** to all objects in the image

Terms: IoU(Intersection over Union)



$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$



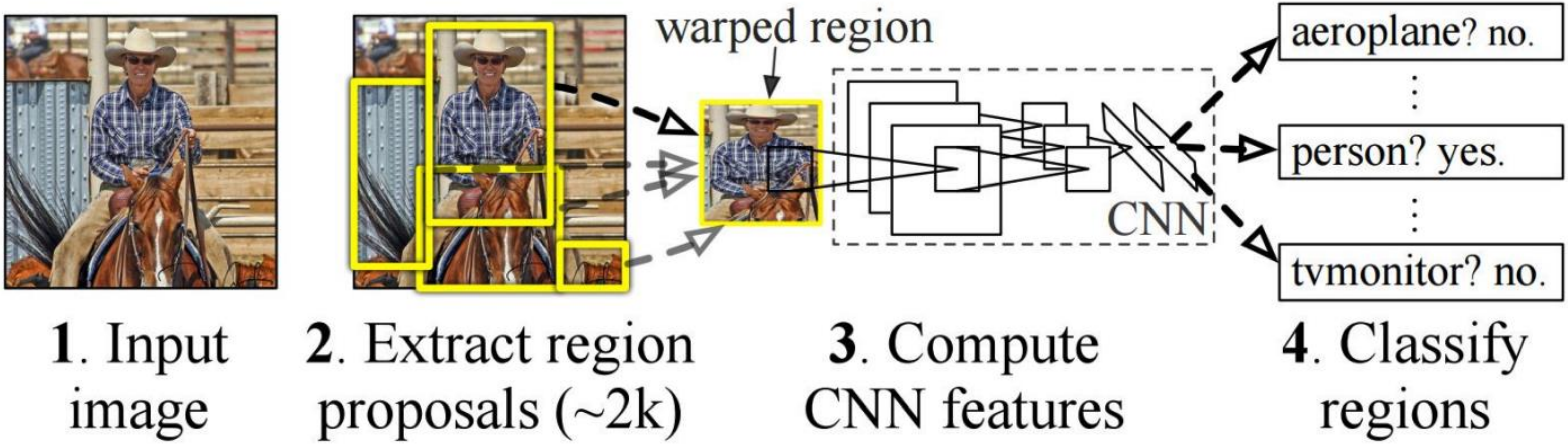
Poor Good Excellent

Terms: NMS(Non-Maximum Suppression)

Modified NMS. Boxes are ranked by their foreground ROI score

R-CNN: Architecture

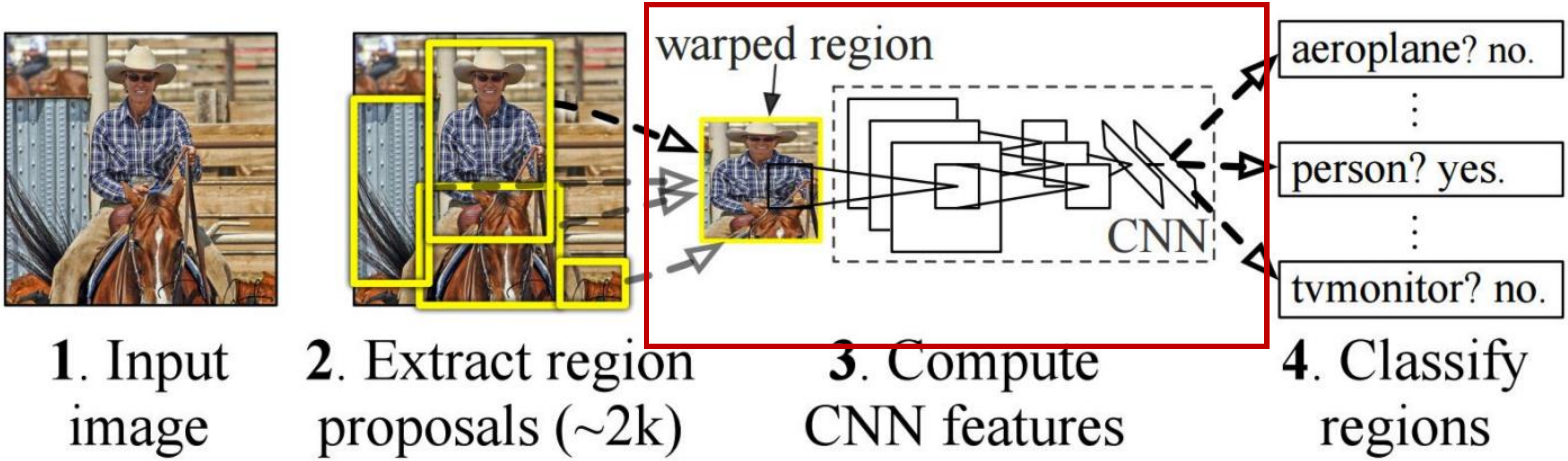
Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic seg mentation." *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2014.



R-CNN: Problem

Slow at test-time

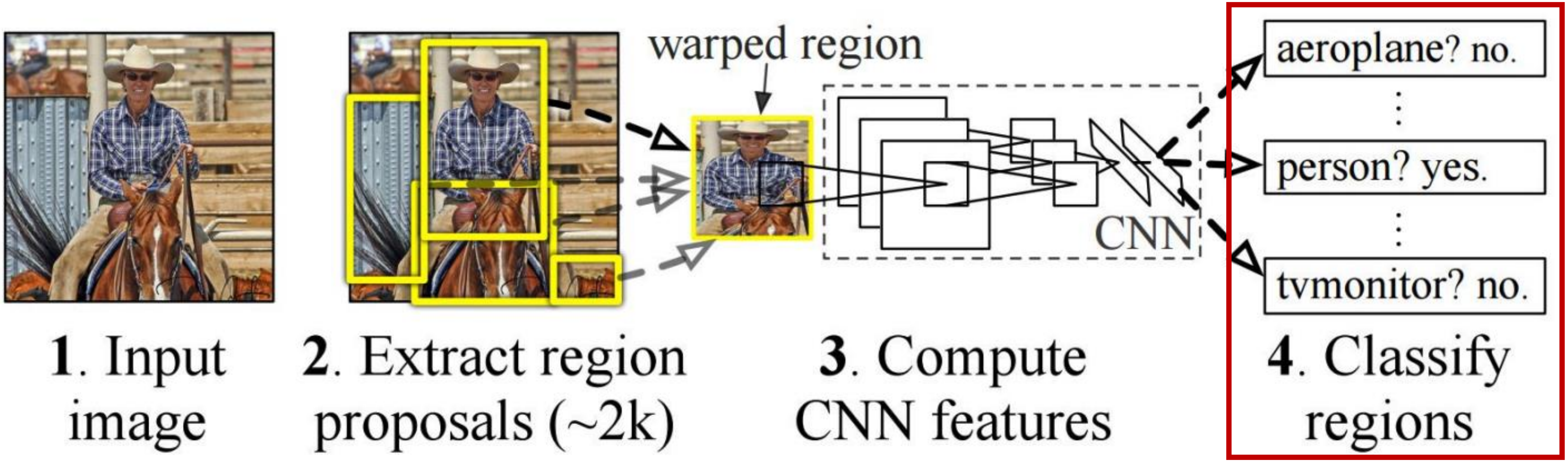
- Need to run full forward pass of CNN for each region proposal



R-CNN: Problem

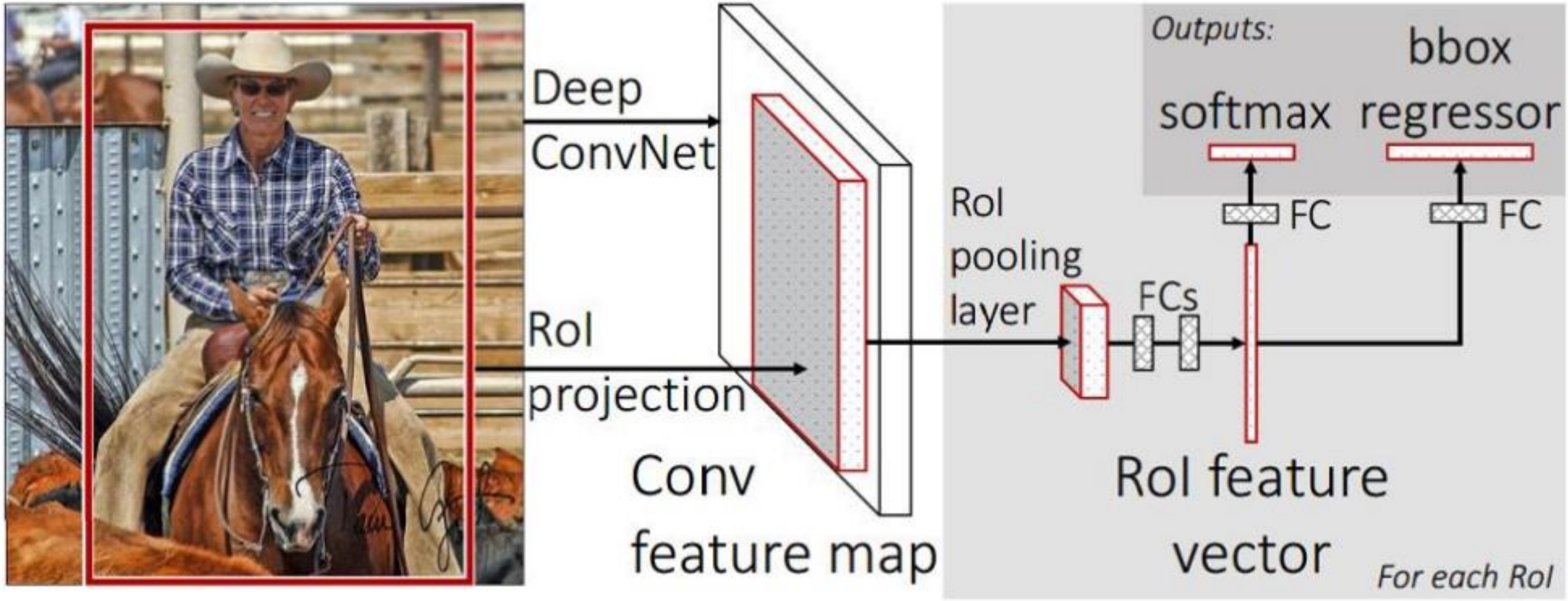
SVMs and regressor

- CNN features not updated in response to SVMs and regressors



Fast R-CNN: Architecture

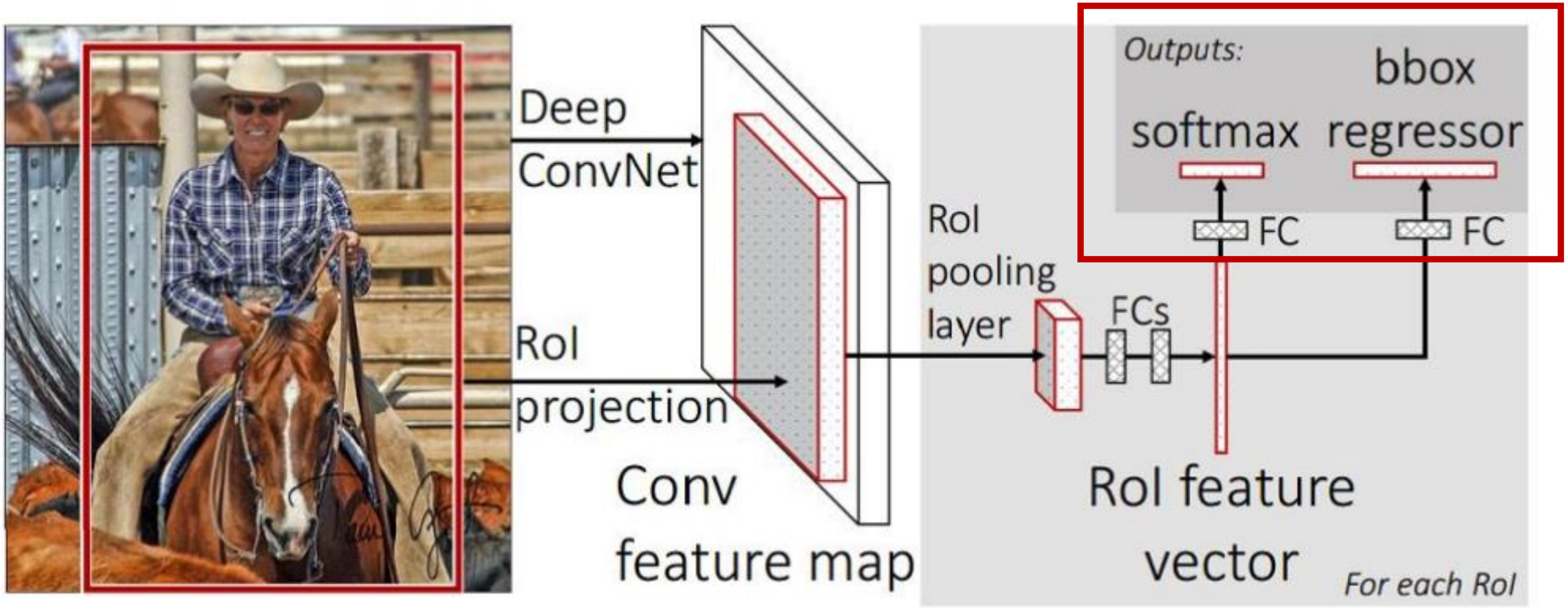
Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.



Fast R-CNN: Architecture

Solution: SVMs and regressor

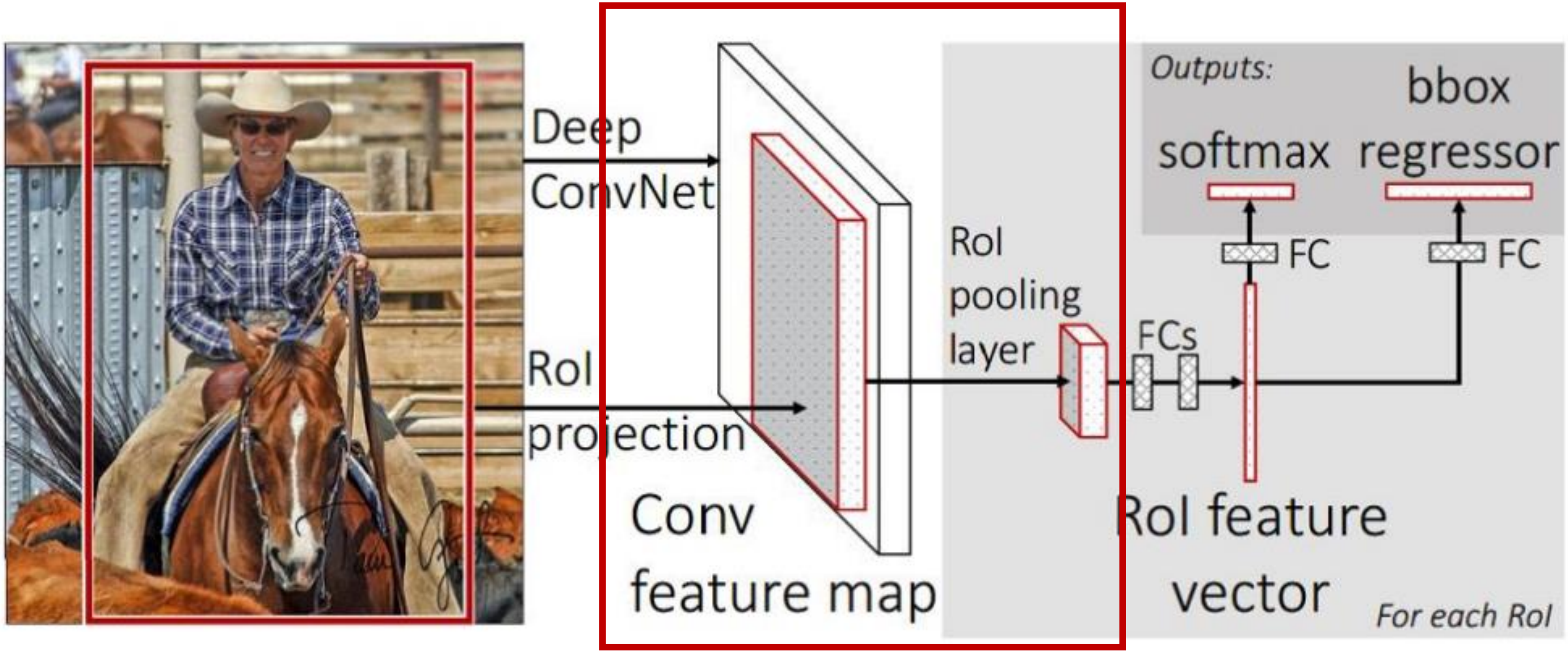
- Train it all at together End-to-End by change SVMs to FC layer



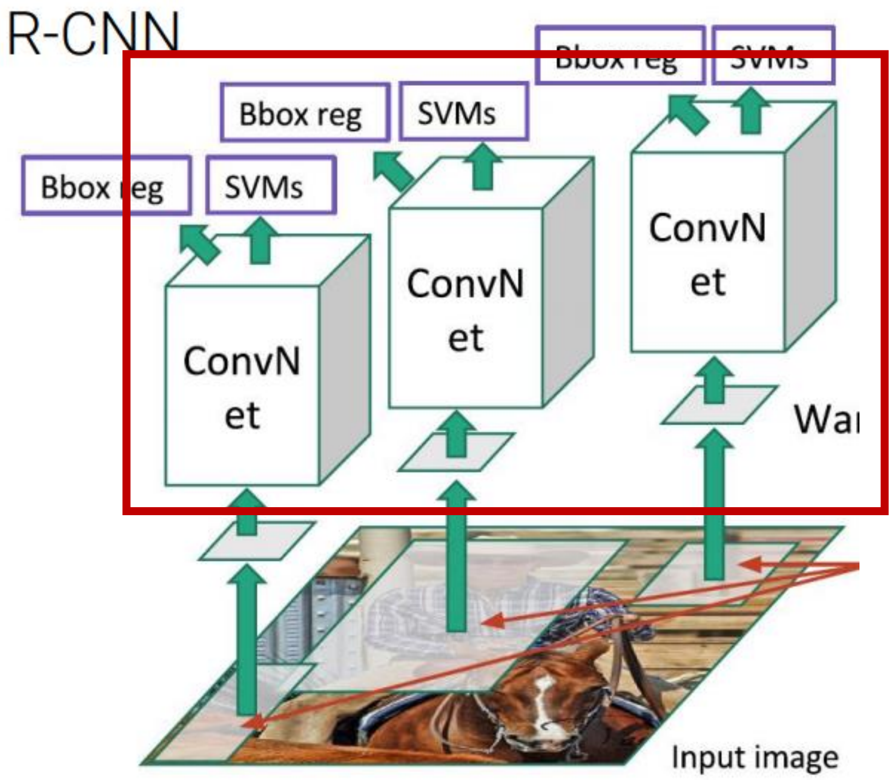
Fast R-CNN: Architecture

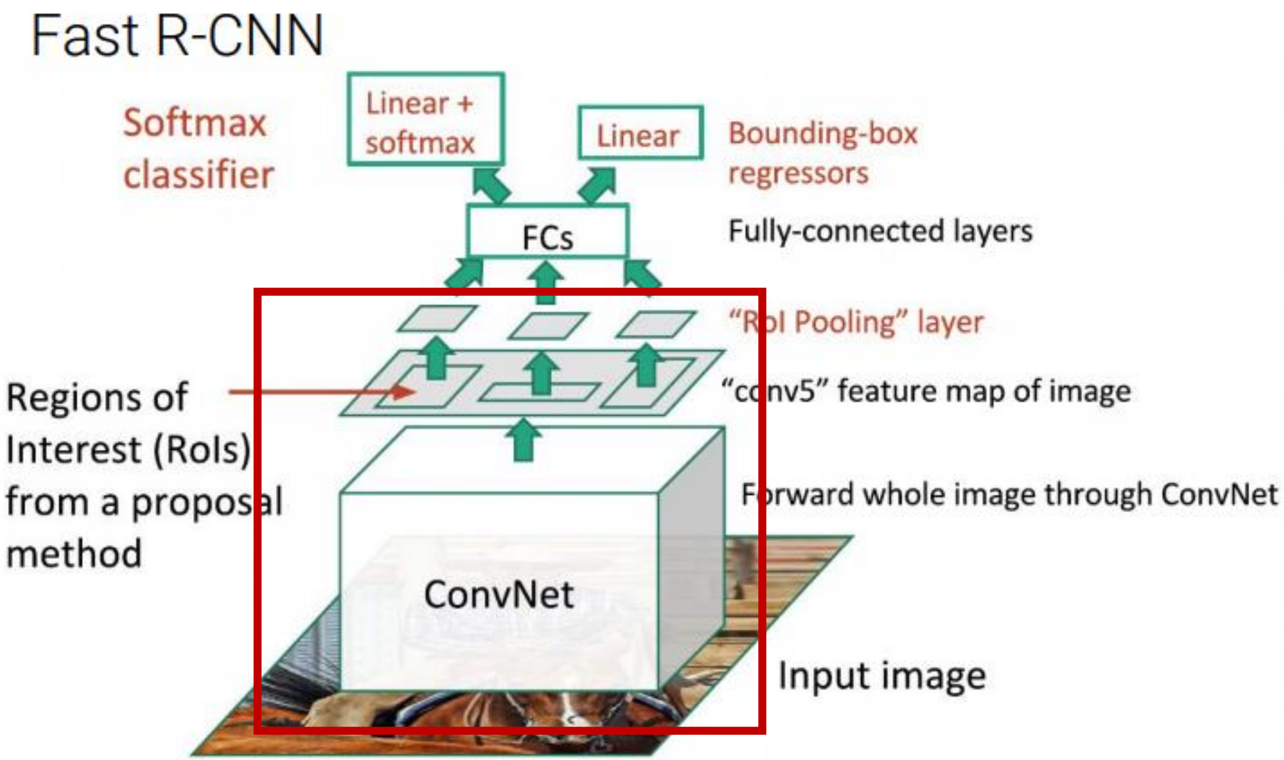
Solution: Slow at test-time

- Share computation of convolutional layers by RoI pooling layer

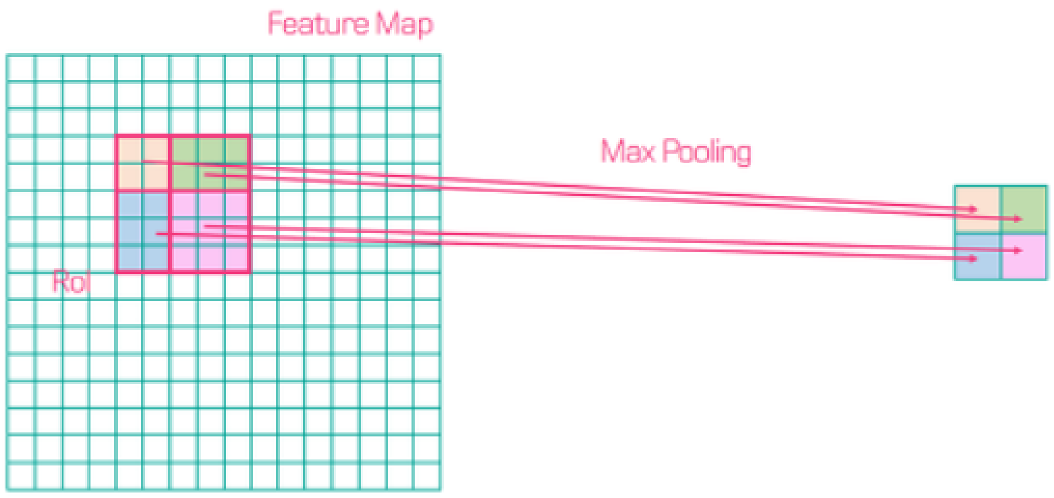


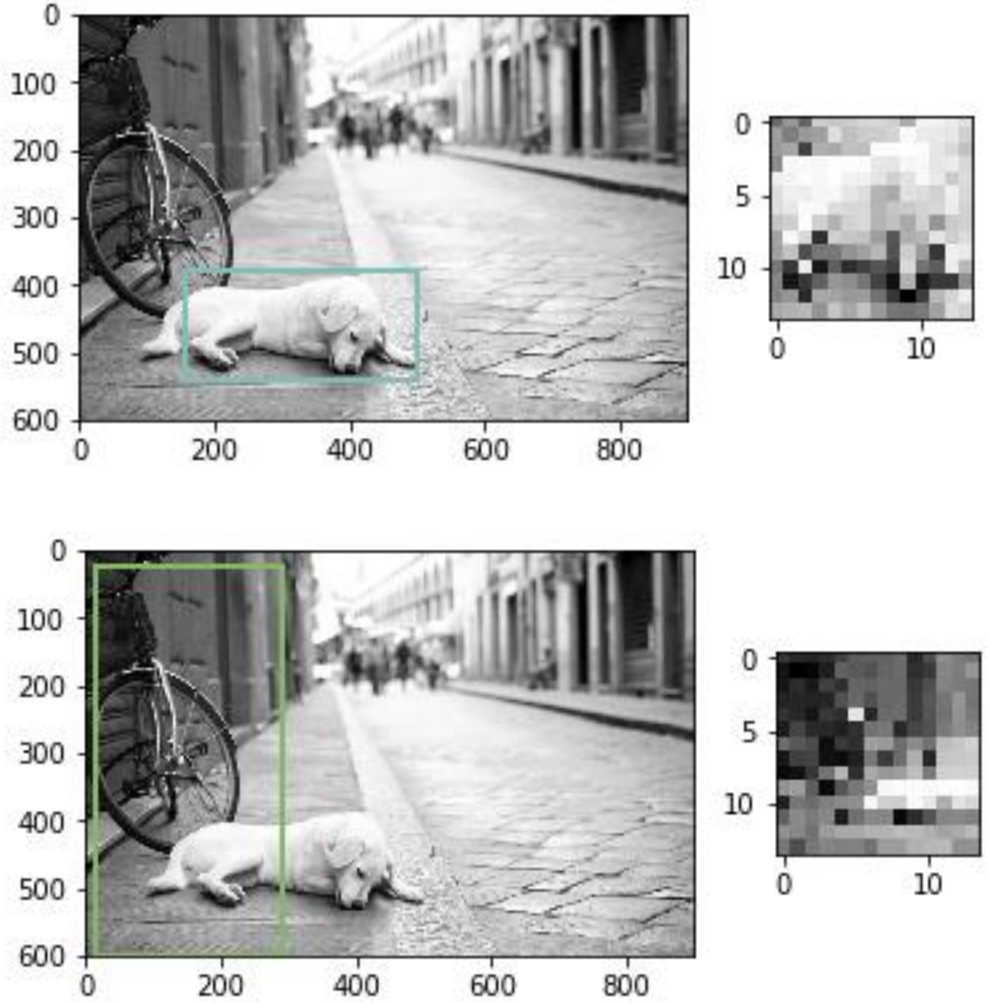
Fast R-CNN: RoI pooling layer





Fast R-CNN: RoI pooling layer





<Example in Image for visuality>

Fast R-CNN: Problem

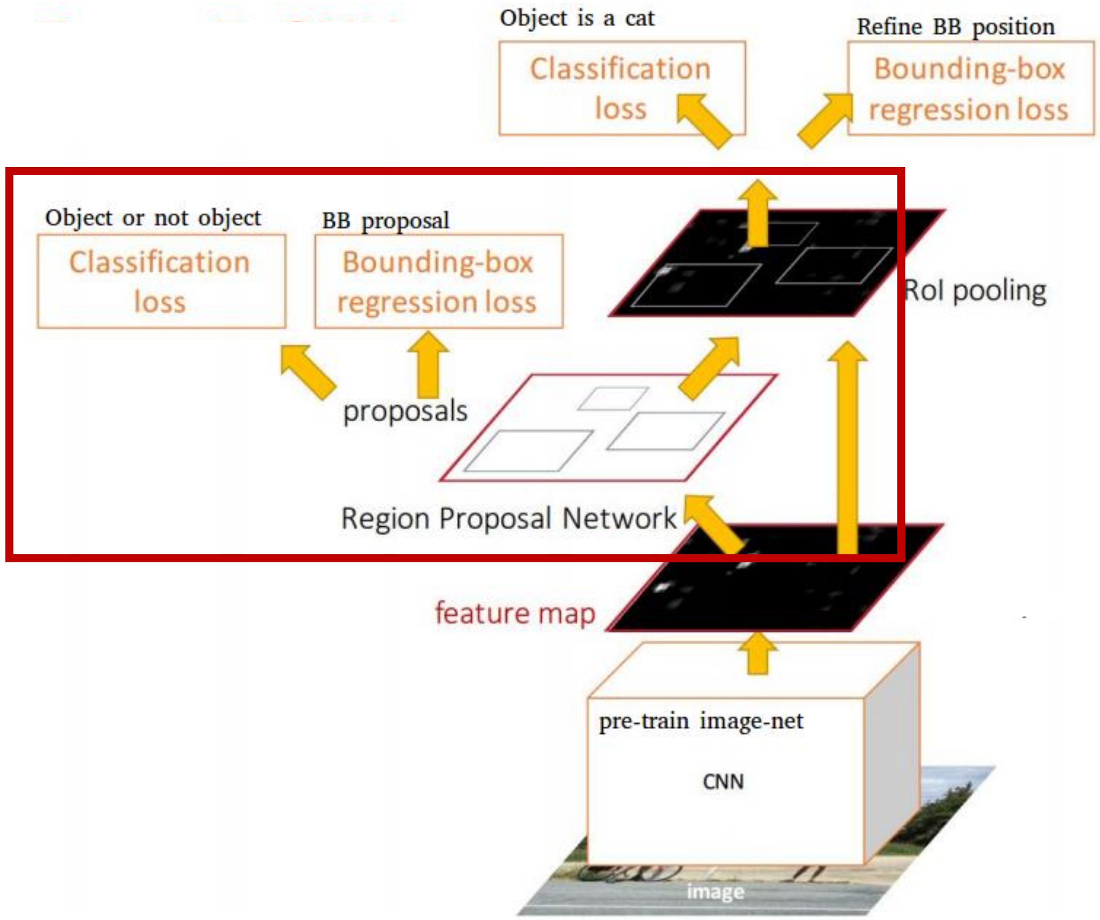
Selective Search

- Selective Search is very slow and is not trainable

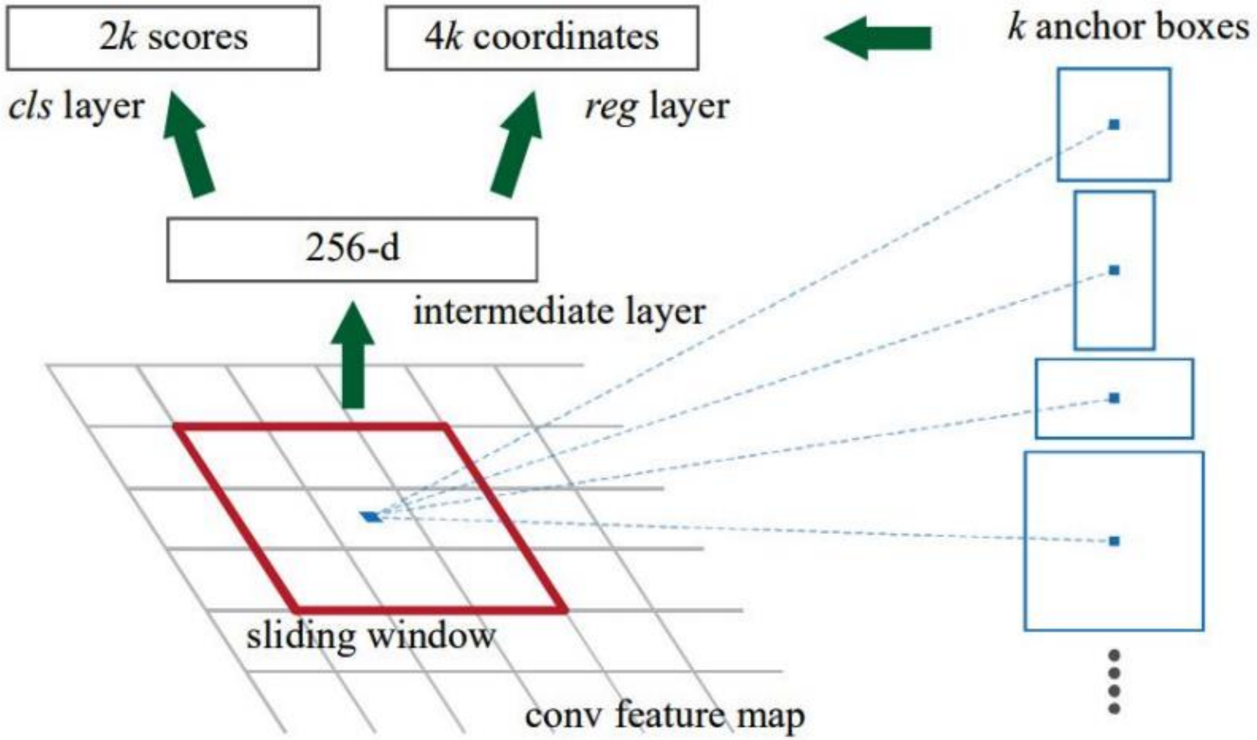
	R-CNN	Fast R-CNN
Test time per image	47 seconds	**0.32 seconds**
(Speedup)	1x	**146x**
Test time per image with Selective Search	50 seconds	**2 seconds**
(Speedup)	1x	**25x**

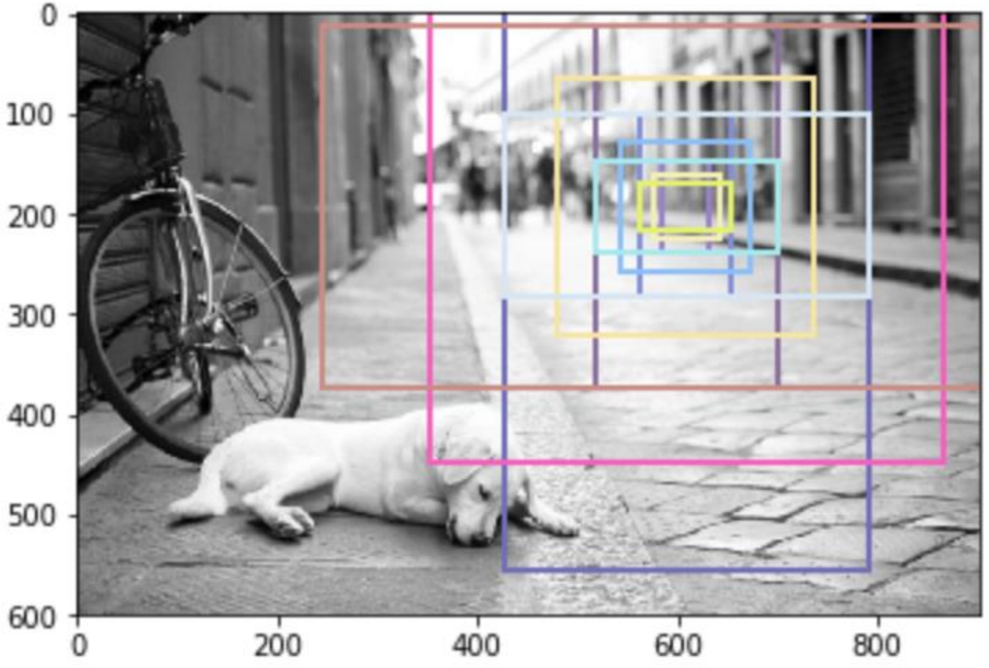
Faster R-CNN: Architecture

Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems*. 2015.



Faster R-CNN: Architecture

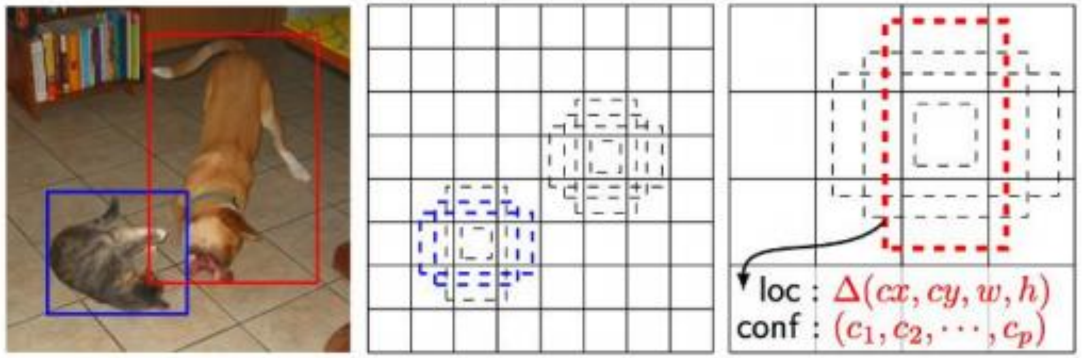




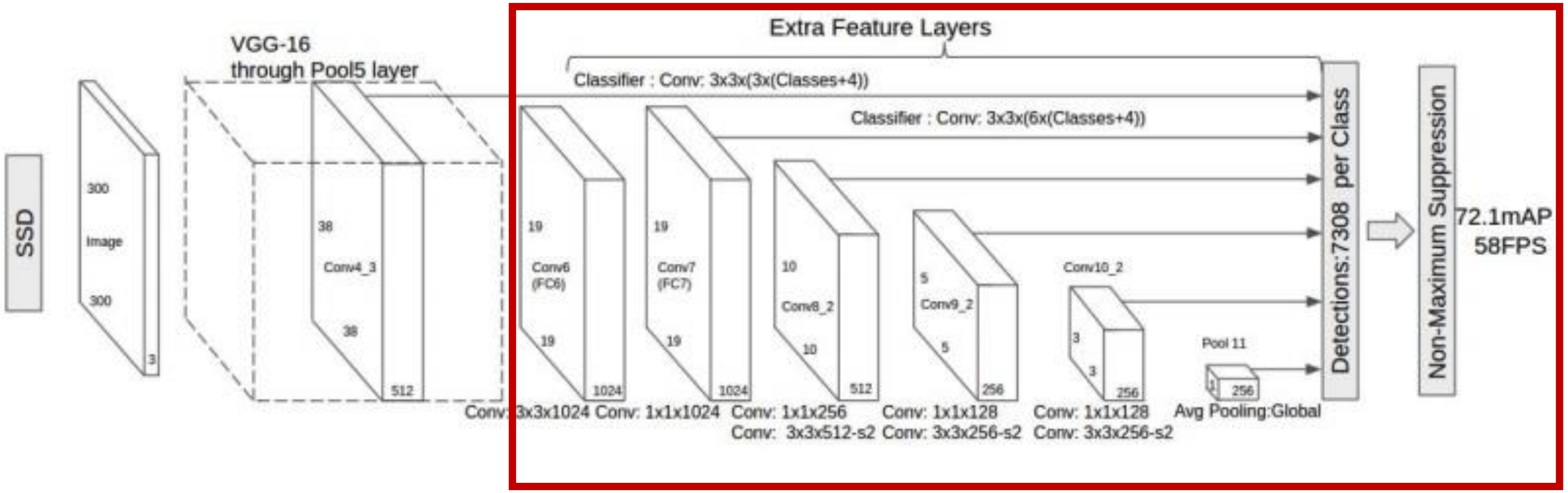
<Example in Image for visuality>

SSD: Single Shot multibox Detector

Liu, Wei, et al. "Ssd: Single shot multibox detector." *European conference on computer vision*. Springer, Cham, 2016.

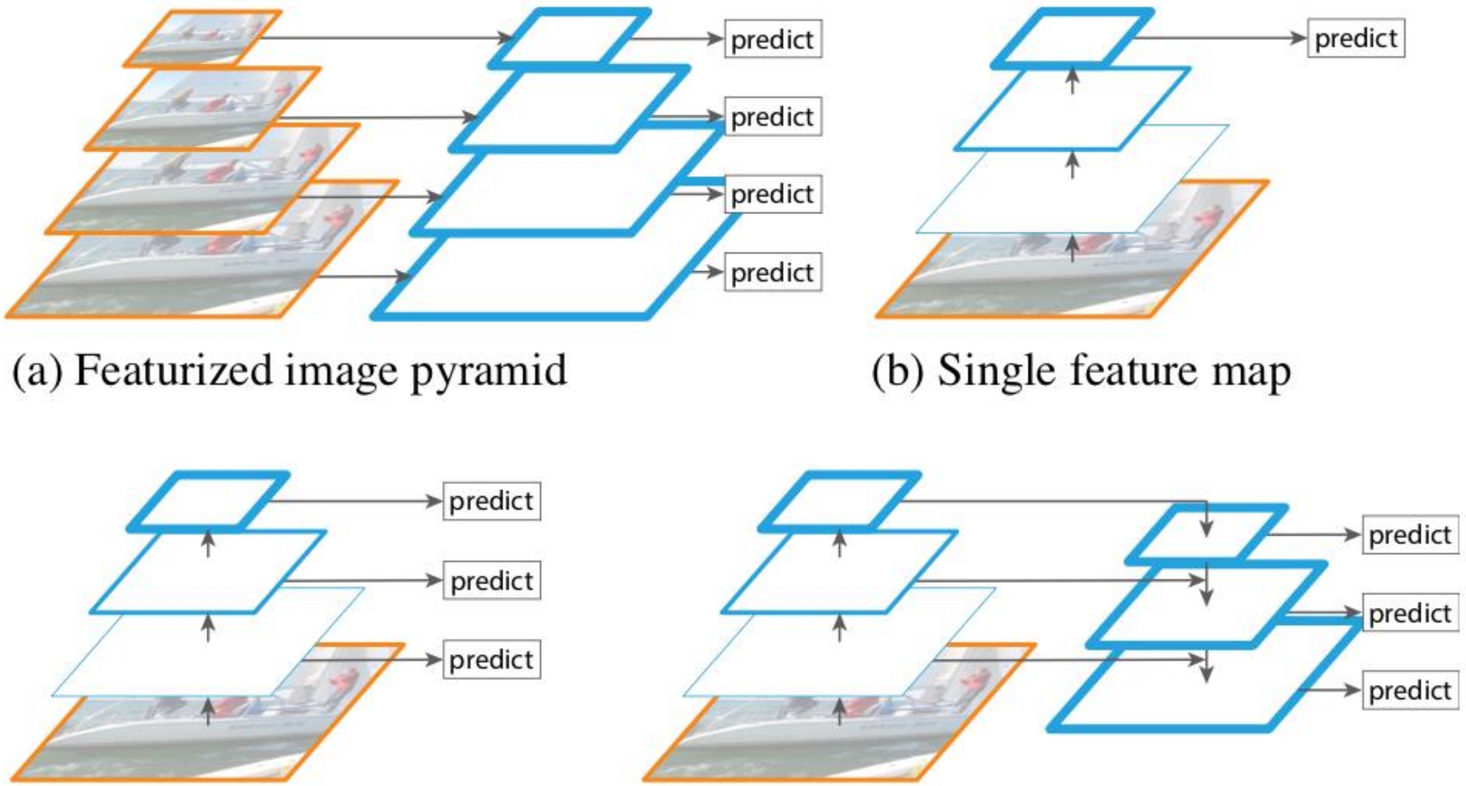


(a) Image with GT boxes (b) 8×8 feature map (c) 4×4 feature map



FPN: Feature Pyramid Network

Lin, Tsung-Yi, et al. "Feature pyramid networks for object detection." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.



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

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

02 Recent Approach

CVPR'18: 28 papers

- Relation Networks for Object Detection
- Finding Tiny Faces in the Wild With Generative Adversarial Network
- **An Analysis of Scale Invariance in Object Detection - SNIP**
- MegDet: A Large Mini-Batch Object Detector
- Multi-Oriented Scene Text Detection via Corner Localization and Region Segmentation
- Single-Shot Object Detection with Enriched Semantics
- R-FCN-3000 at 30fps: Decoupling Detection and Classification
- Multi-Scale Location-Aware Kernel Representation for Object Detection
- Geometry-Aware Scene Text Detection With Instance Transformation Network
- Real-Time Rotation-Invariant Face Detection with Progressive Calibration Networks
- Pseudo Mask Augmented Object Detection
- Feature Selective Networks for Object Detection
- Single-Shot Refinement Neural Network for Object Detection
- Learning Globally Optimized Object Detector via Policy Gradient
- Structure Inference Net: Object Detection Using Scene-Level Context and Instance-Level Relationships
- ClusterNet: Detecting Small Objects in Large Scenes by Exploiting Spatio-Temporal Information
- Objects as context for detecting their semantic parts
- Dynamic Zoom-in Network for Fast Object Detection in Large Images
- **Cascade R-CNN: Delving into High Quality Object Detection**
- DecideNet: Counting Varying Density Crowds Through Attention Guided Detection and Density Estimation
- Repulsion Loss: Detecting Pedestrians in a Crowd
- …

ECCV'18: 19 papers

- **Acquisition of Localization Confidence for Accurate Object Detection**
- Receptive Field Block Net for Accurate and Fast Object Detection
- Revisiting RCNN: On Awakening the Classification Power of Faster RCNN
- Deep Feature Pyramid Reconfiguration for Object Detection
- SOD-MTGAN: Small Object Detection via Multi-Task Generative Adversarial Network CornerNet: Detecting Objects as Paired Keypoints
- Zero-Shot Object Detection
- Learning Region Features for Object Detection
- Graininess-Aware Deep Feature Learning for Pedestrian Detection
- DetNet: Design Backbone for Object Detection
- PyramidBox: A Context-assisted Single Shot Face Detector
- Quantization Mimic: Towards Very Tiny CNN for Object Detection
- Object Detection with an Aligned Spatial-Temporal Memory
- Localization Recall Precision (LRP): A New Performance Metric for Object Detection Context Refinement for Object Detection
- Context Refinement for Object Detection
- Bi-box Regression for Pedestrian Detection and Occlusion Estimation
- Deep Feature Pyramid Reconfiguration for Object Detection
- Parallel Feature Pyramid Network for Object Detection
- Occlusion-aware R-CNN: Detecting Pedestrians in a Crowd
- Learning Region Features for Object Detection
- Where are the Blobs: Counting by Localization with Point Supervision(Counting)Geometry-Aware Scene Text Detection With Instance Transformation Network

SNIP

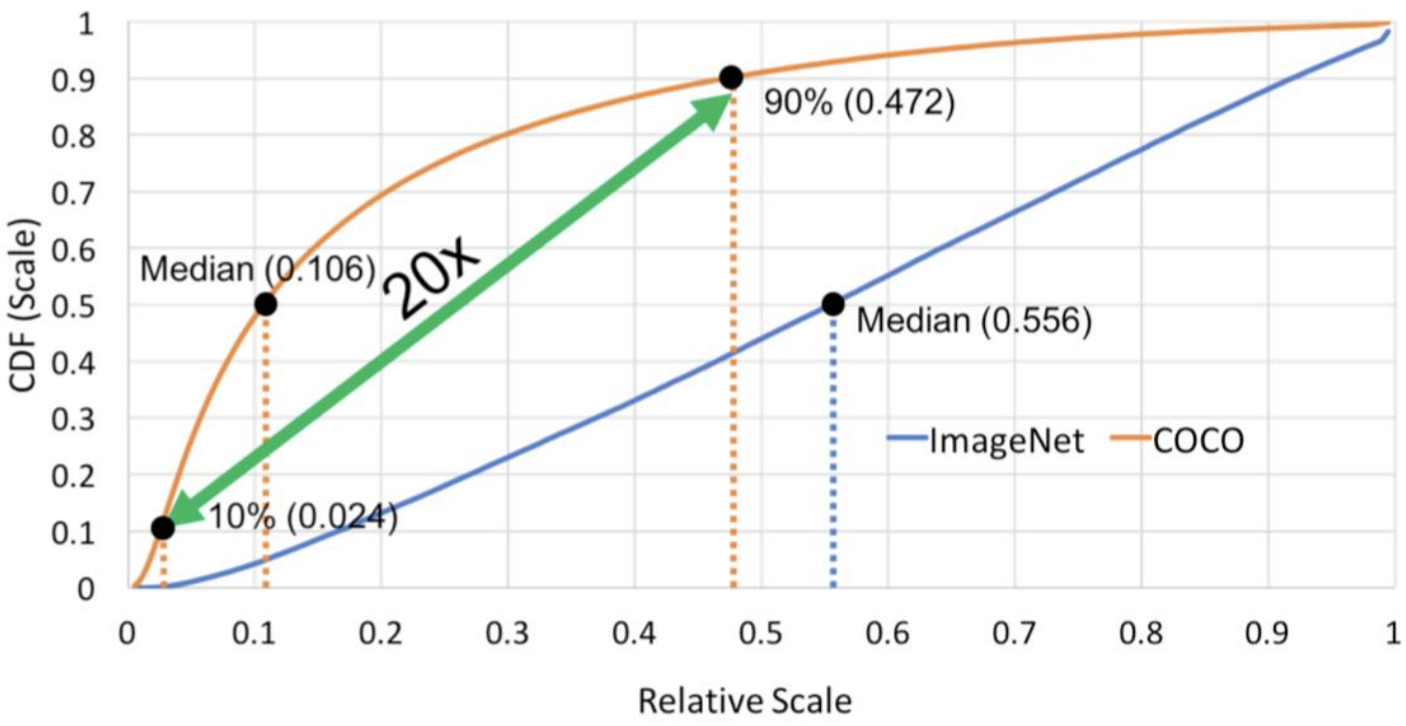
An Analysis of Scale Invariance in Object Detection – SNIP

Bharat Singh Larry S. Davis
University of Maryland, College Park
{bharat,lsd}@cs.umd.edu

SNIP

Problem Definition: Large scale variation

- Why is object detection so much harder than image classification?
  → Large scale variation across object instances



SNIP

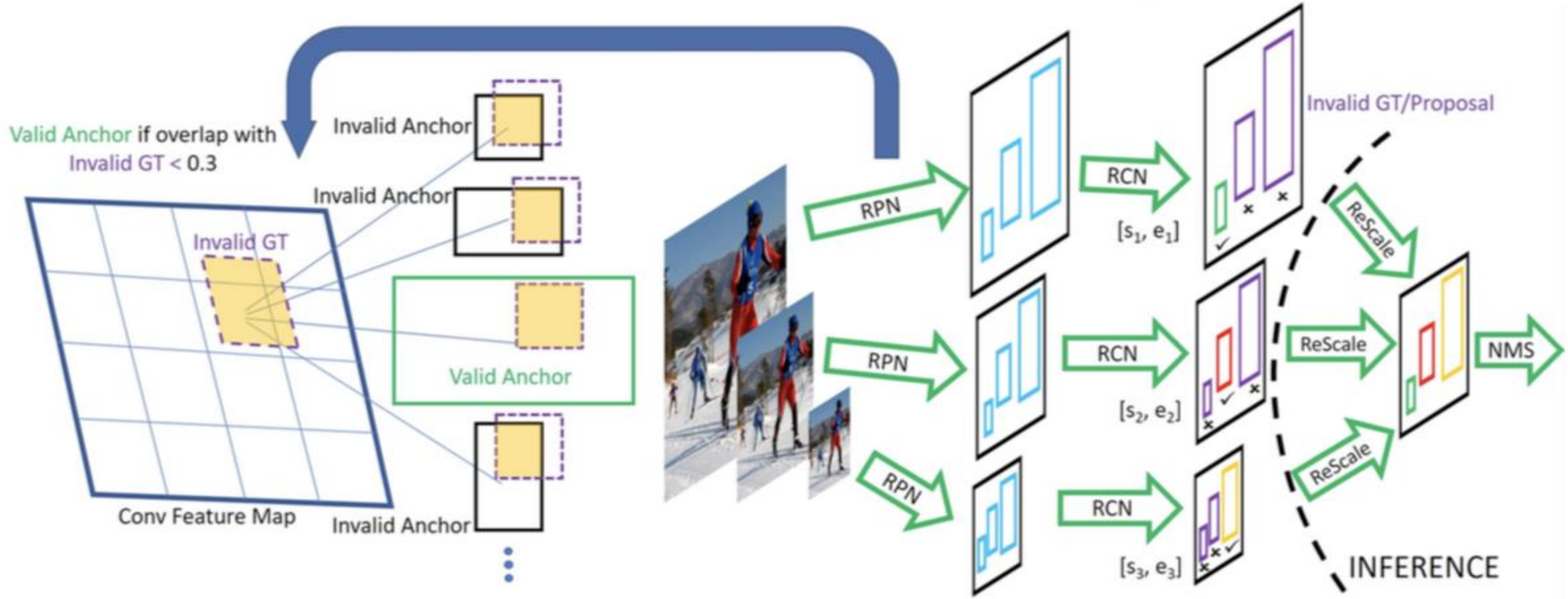


Figure 6. SNIP training and inference for IPN is shown. Invalid RoIs which fall outside the specified range at each scale are shown in purple. These are discarded during training and inference. Each batch during training consists of images sampled from a particular scale. Invalid GT boxes are used to invalidate anchors in RPN. Detections from each scale are rescaled and combined using NMS.

Cascade R-CNN

Cascade R-CNN: Delving into High Quality Object Detection

Zhaowei Cai
UC San Diego
zwcai@ucsd.edu

Nuno Vasconcelos
UC San Diego
nuno@ucsd.edu

Cascade R-CNN

~~Accuracy~~

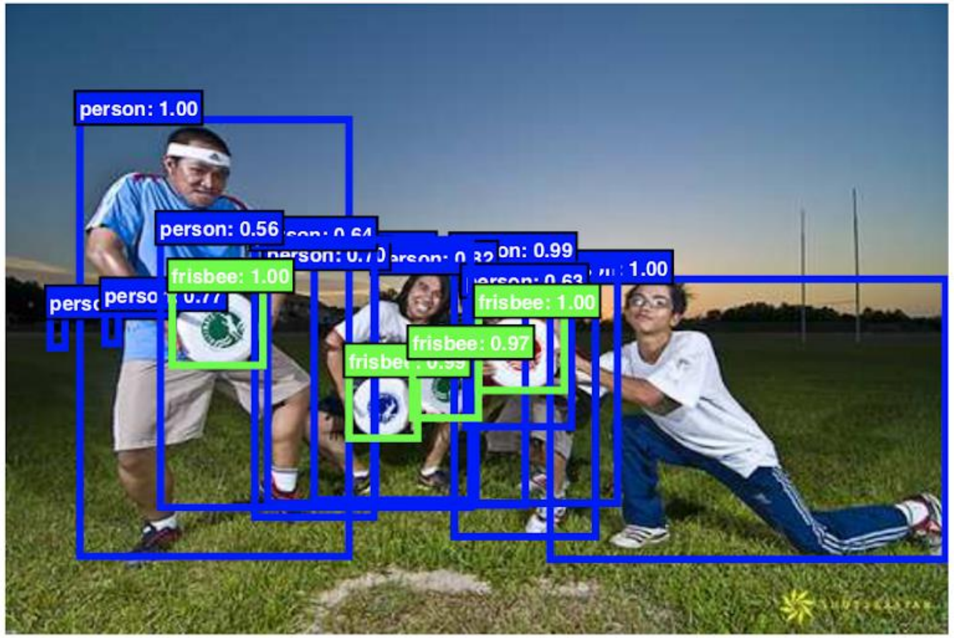
Cascade R-CNN: Delving into High Quality Object Detection

Zhaowei Cai
UC San Diego
zwcai@ucsd.edu

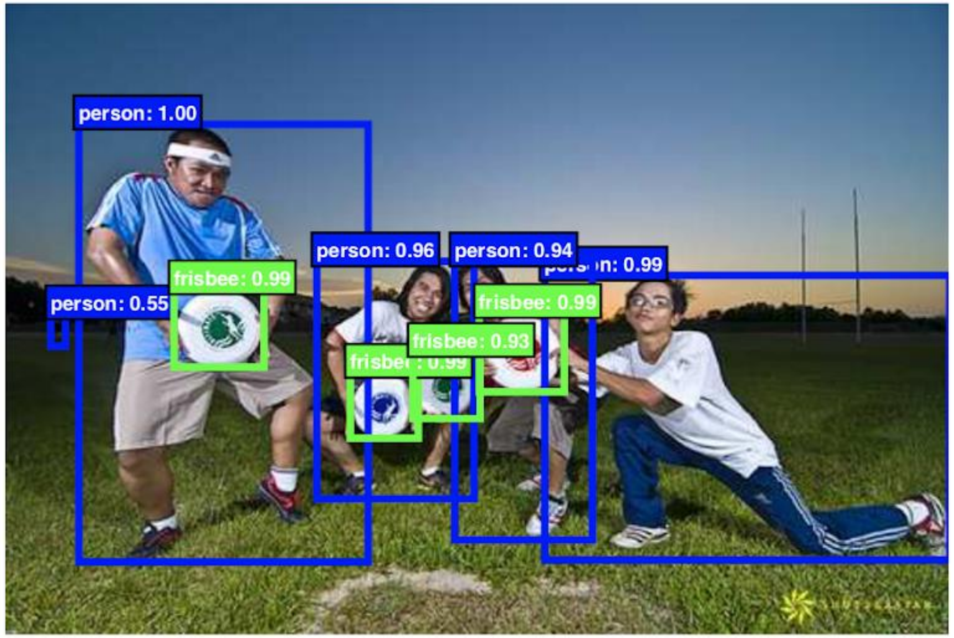
Nuno Vasconcelos
UC San Diego
nuno@ucsd.edu

Cascade R-CNN: Quality

In this work, we define the *quality* of an hypothesis as its IoU with the ground truth, and the *quality of the detector* as the IoU threshold u used to train it. The goal is to investi-



(a) Detection of $u = 0.5$

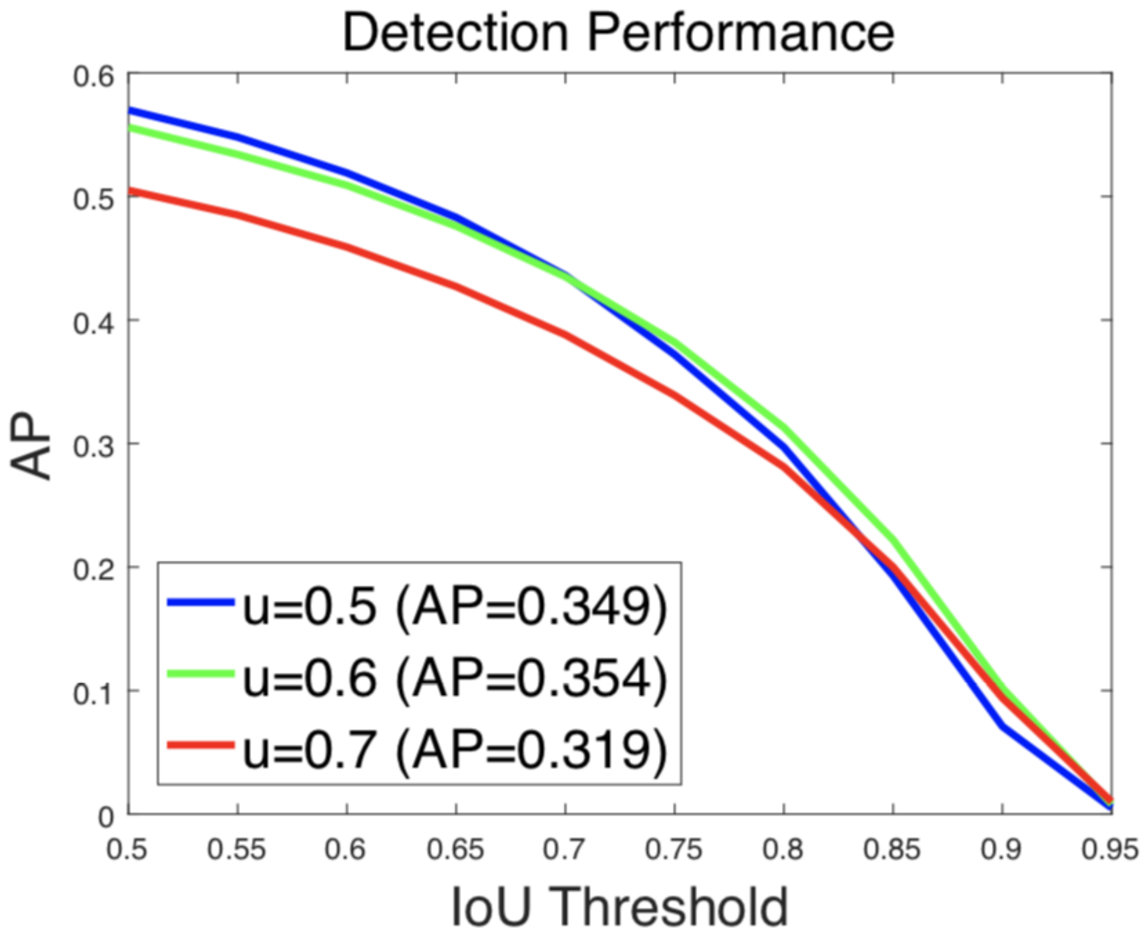


(b) Detection of $u = 0.7$

High quality detector

Cascade R-CNN: Problem

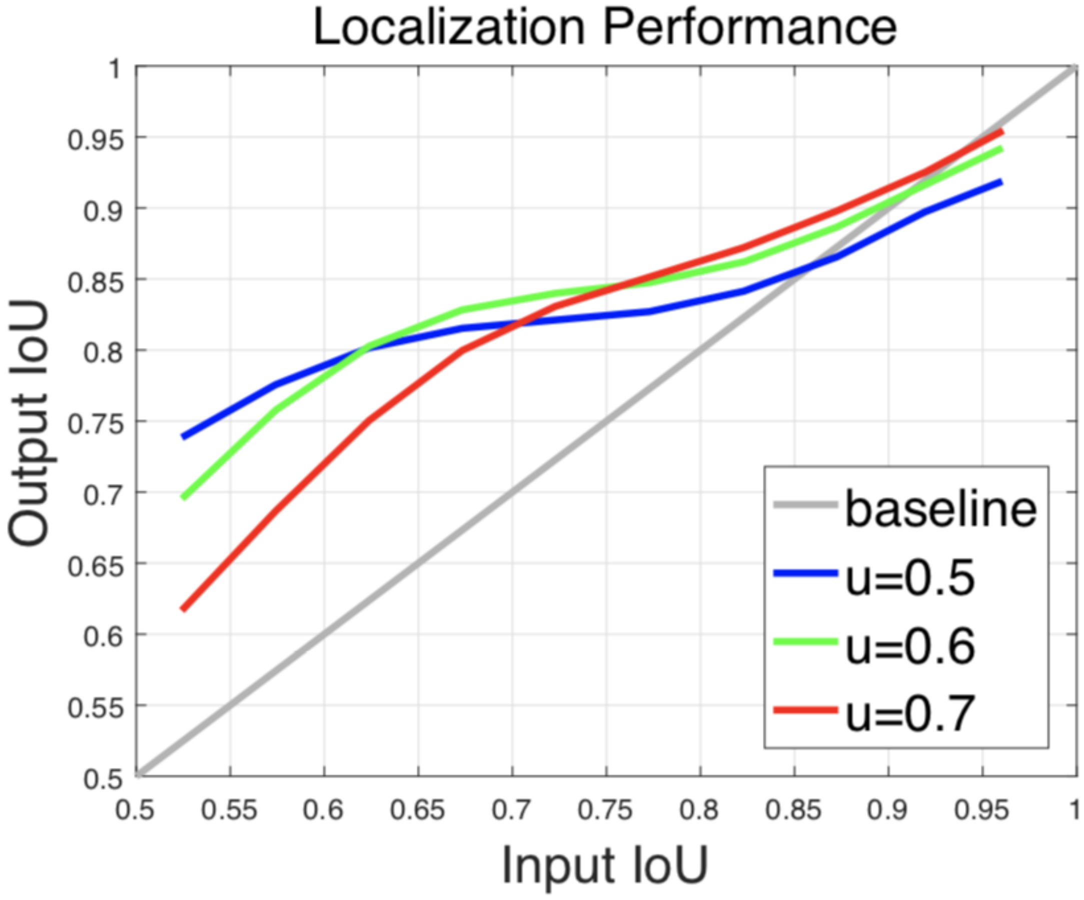
Overfitting during training, due to exponentially vanishing positive samples



SI ANALYTICS

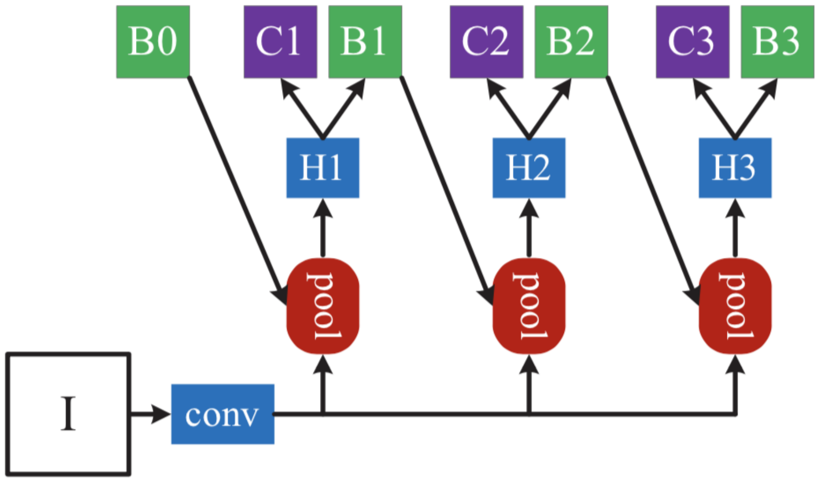
Cascade R-CNN: Main Idea

are trained sequentially, using the output of one stage to train the next. This is motivated by the observation that the output IoU of a regressor is almost invariably better than the input IoU.



SI ANALYTICS

Cascade R-CNN: Architecture



(d) Cascade R-CNN

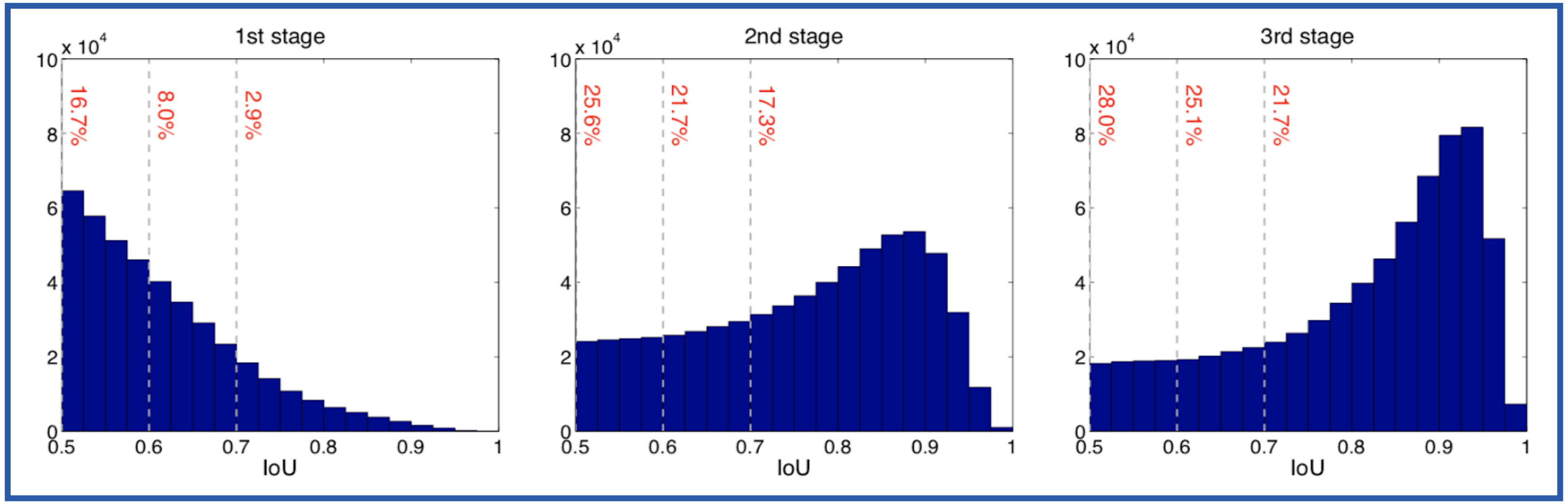


Figure 4. The IoU histogram of training samples. The distribution at 1st stage is the output of RPN. The red numbers are the positive percentage higher than the corresponding IoU threshold.

IoU-Net

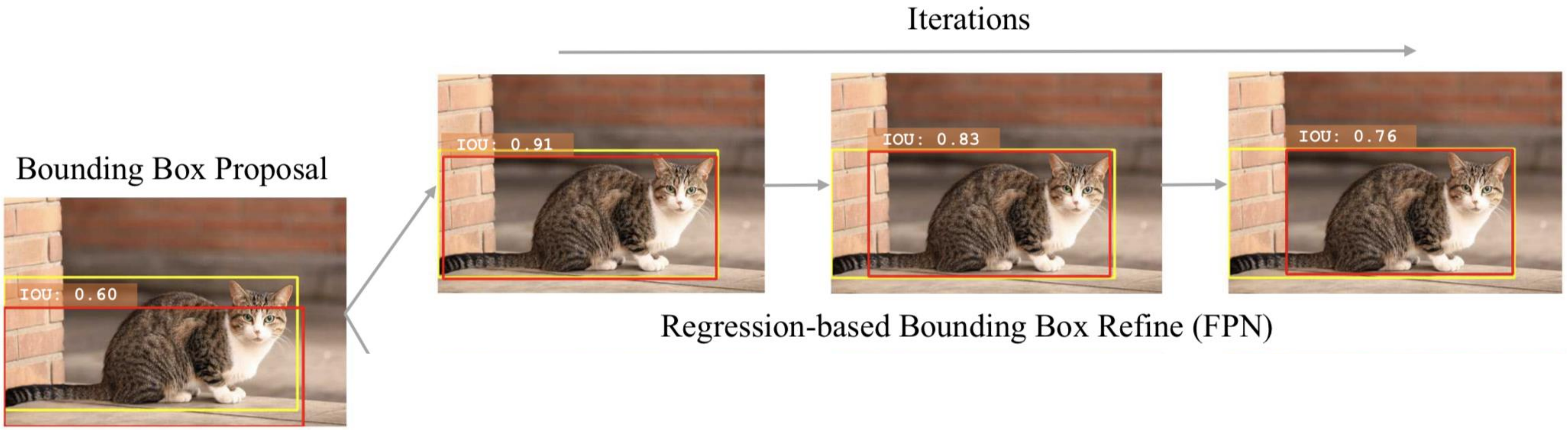
Acquisition of Localization Confidence for Accurate Object Detection

Borui Jiang*[1,3], Ruixuan Luo*[1,3], Jiayuan Mao*[2,4], Tete Xiao[1,3], and Yuning Jiang[4]

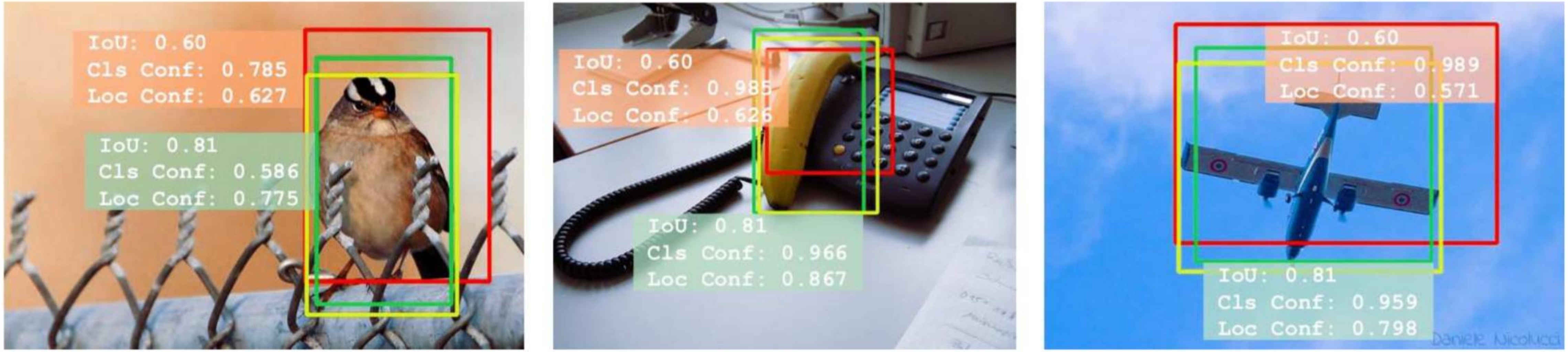
[1] School of Electronics Engineering and Computer Science, Peking University
[2] ITCS, Institute for Interdisciplinary Information Sciences, Tsinghua University
[3] Megvii Inc. (Face++) [4] Toutiao AI Lab
{jbr, luoruixuan97, jasonhsiao97}@pku.edu.cn,
mjy14@mails.tsinghua.edu.cn, jiangyuning@bytedance.com

IoU-Net: Problem

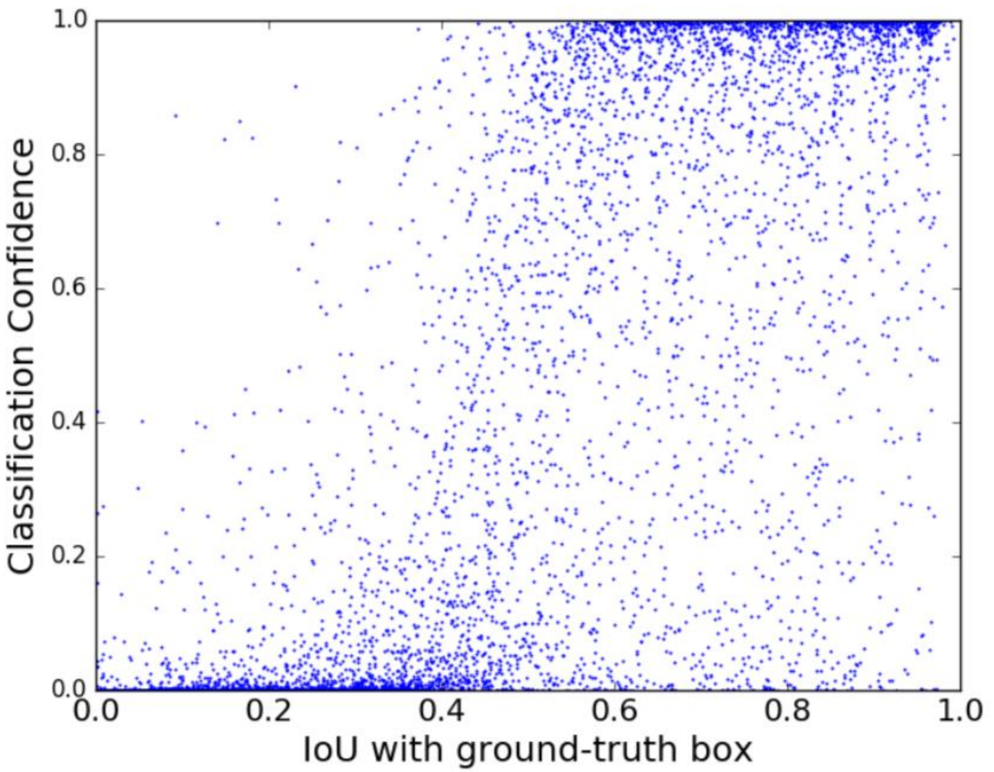
Abstract. Modern CNN-based object detectors rely on bounding box regression and non-maximum suppression to localize objects. While the probabilities for class labels naturally reflect classification confidence, localization confidence is absent. This makes properly localized bounding boxes degenerate during iterative regression or even suppressed during NMS. In the paper we propose IoU-Net learning to predict the IoU



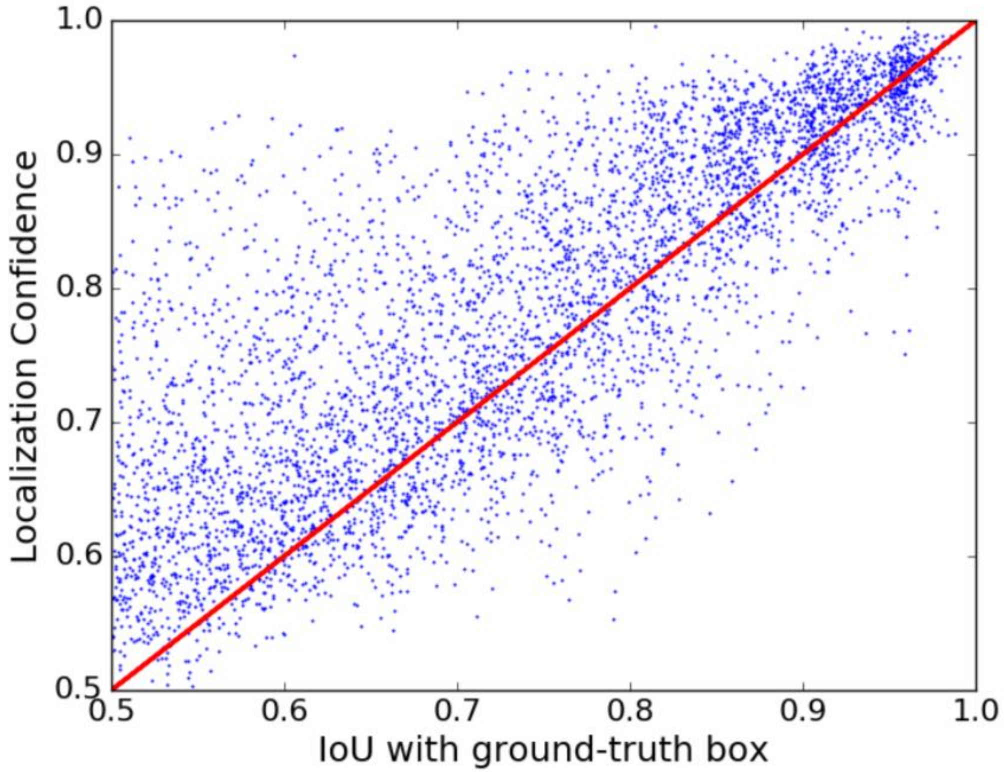
IoU-Net: Main Idea



IoU-Net

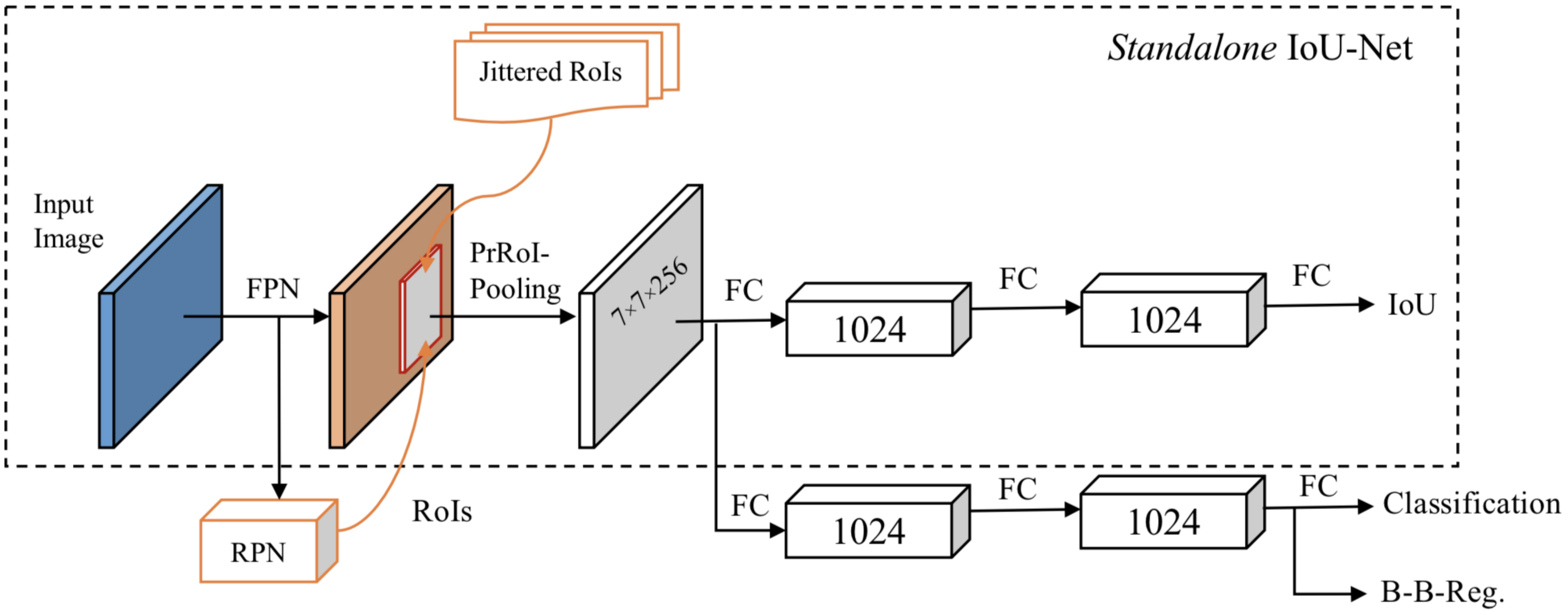


(a) IoU vs. Classification Confidence



(b) IoU vs. Localization Confidence

IoU-Net



Standalone IoU-Net
Jittered RoIs
Input
Image
FPN
PrRoI-
Pooling
7×7×256
FC
1024
FC
1024
FC
IoU
RPN
RoIs
FC
1024
FC
1024
FC
Classification
B-B-Reg.

Meta-Anchor

MetaAnchor: Learning to Detect Objects with Customized Anchors

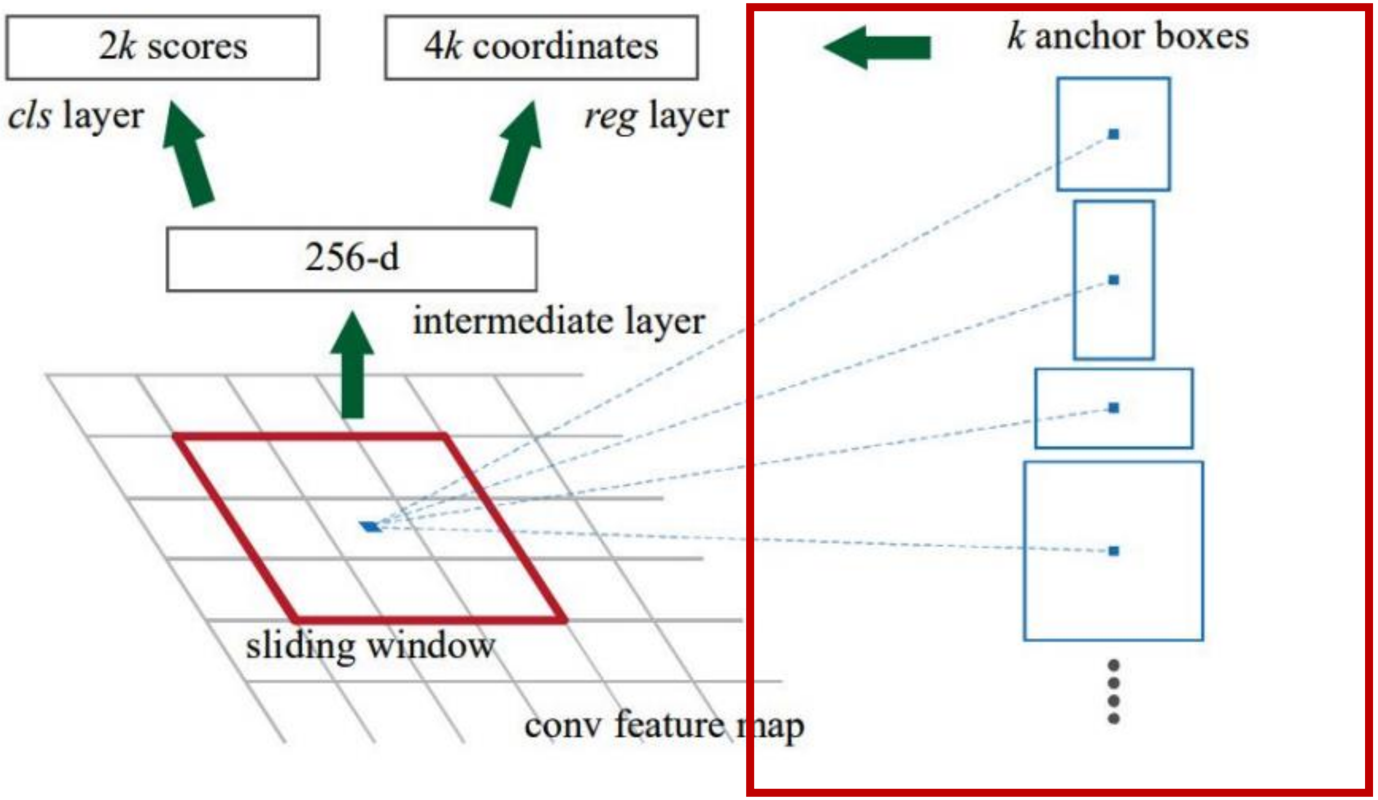
Tong Yang[*†] Xiangyu Zhang[*] Zeming Li[*] Wenqiang Zhang[†] Jian Sun[*]

[*] Megvii Inc (Face++)
{yangtong,zhangxiangyu,lizeming,sunjian}@megvii.com

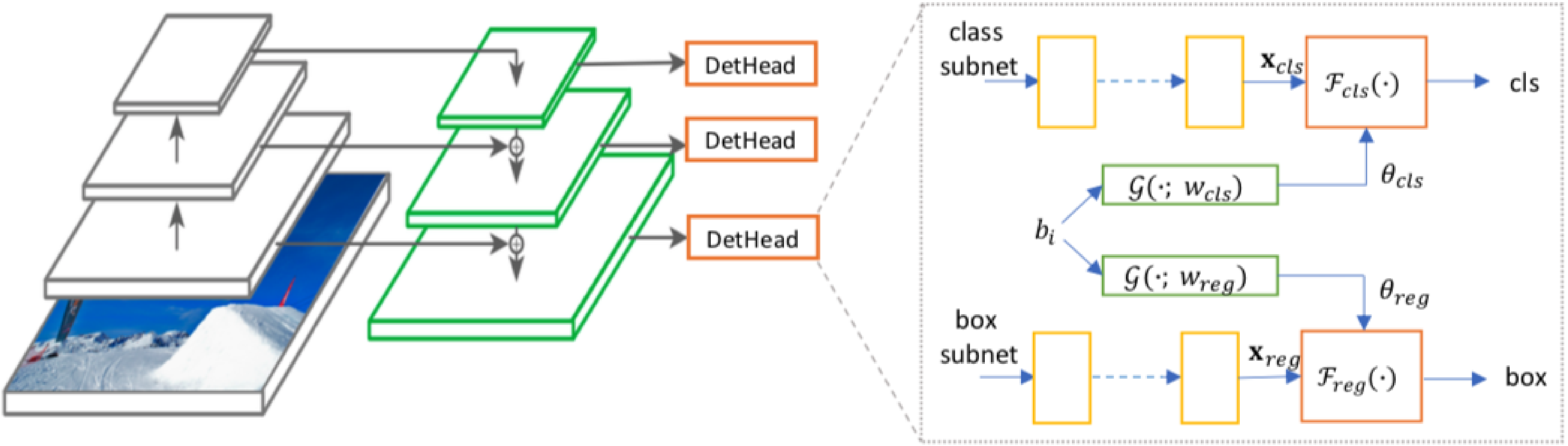
[†] Fudan University
wqzhang@fudan.edu.cn

Meta-Anchor: Problem

Anchors are fixed

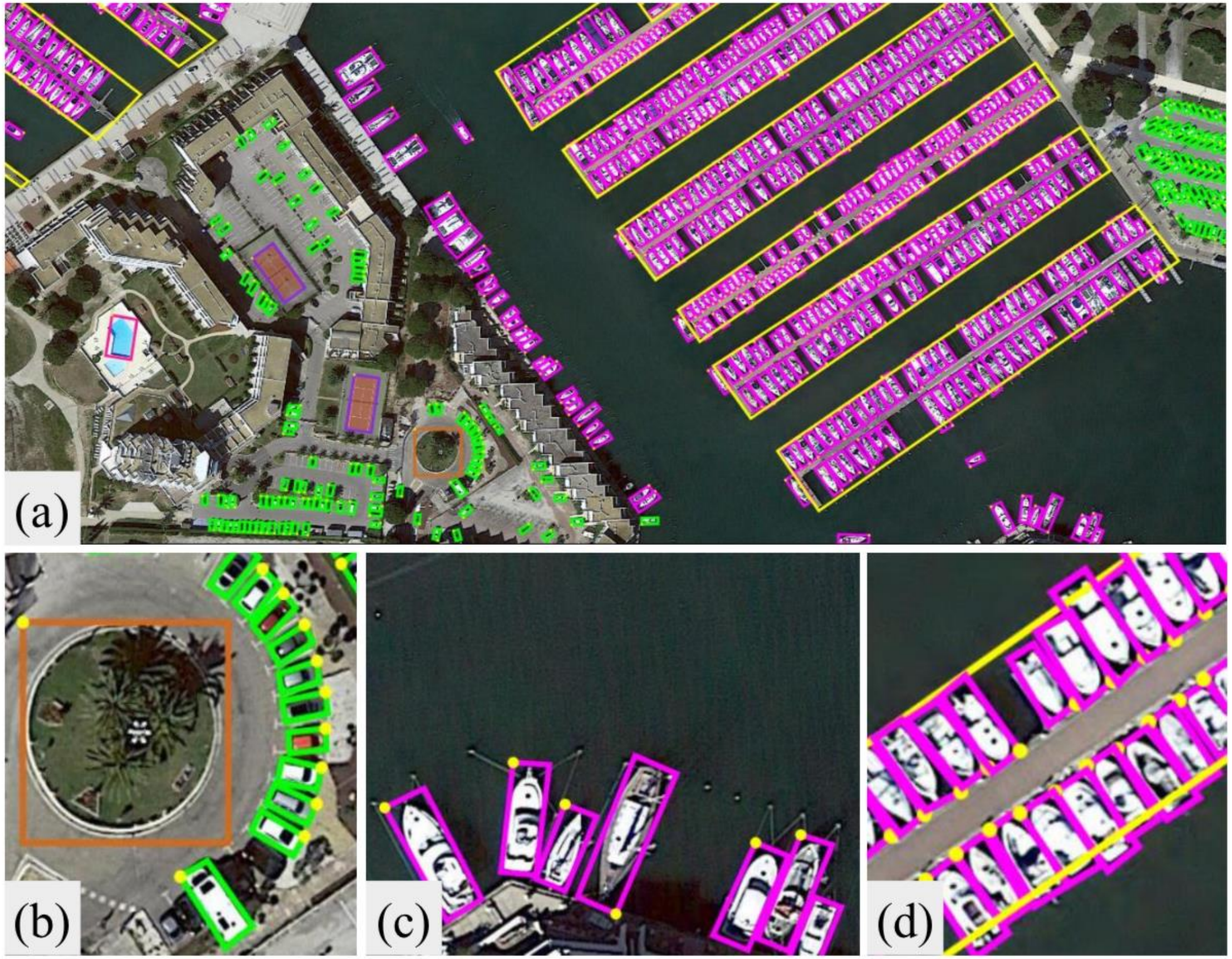


Meta-Anchor: Architecutre



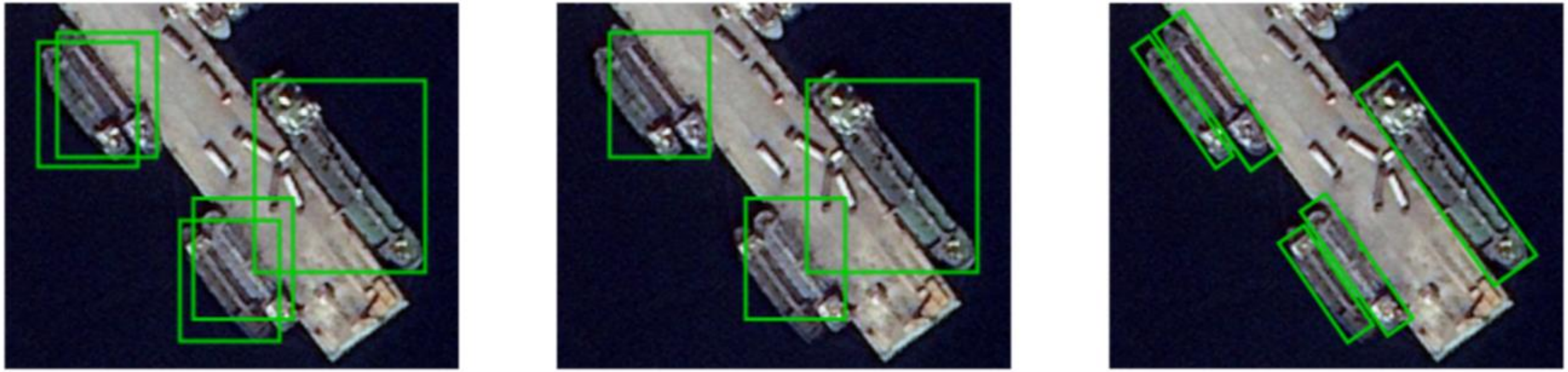
03 In Remote Sensing

Oriented Objects



Oriented Objects

Bounding Box V.S. Rotated Bounding Box



RBox-CNN

RBox-CNN: Rotated Bounding Box based CNN for Ship Detection in Remote Sensing Image

Jamyoung Koo
Satrec Initiative, Korea
jmkoo@satreci.com

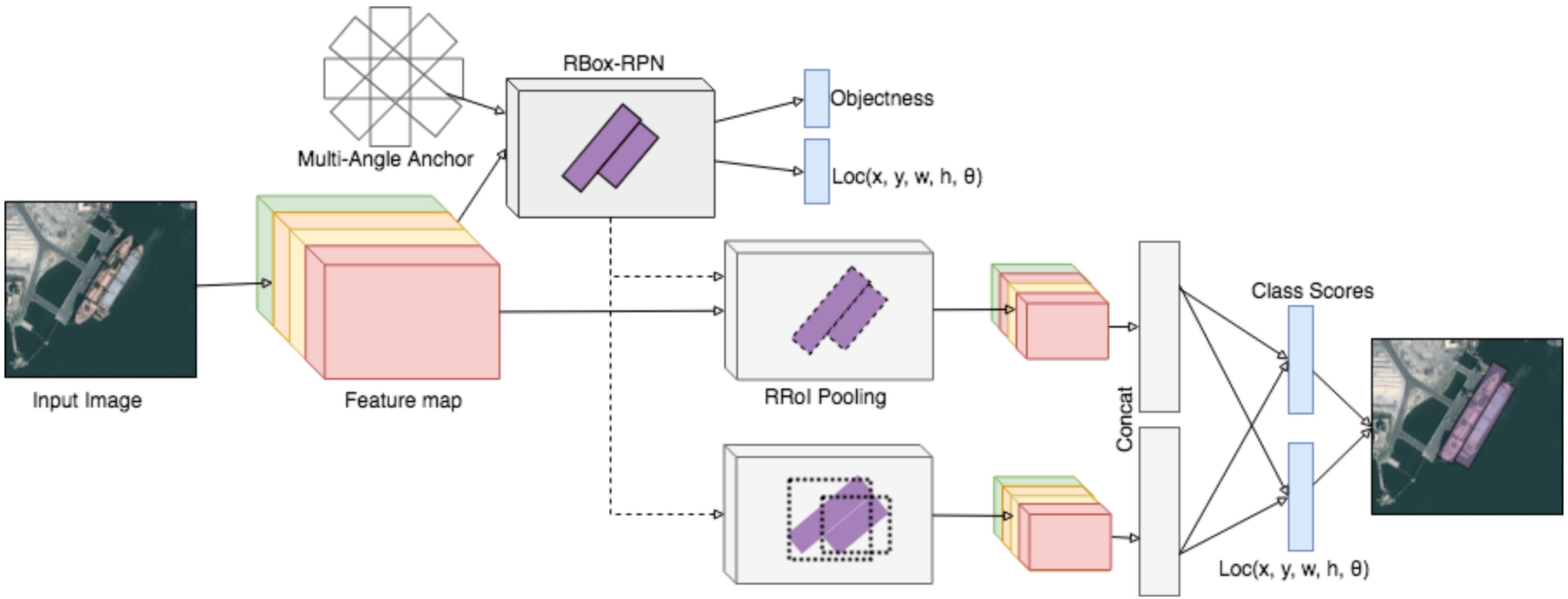
Junghoon Seo
Satrec Initiative, Korea
sjh@satreci.com

Seunghyun Jeon
Satrec Initiative, Korea
jsh@satreci.com

Jeongyeol Choe
Satrec Initiative, Korea
cjy@satreci.com

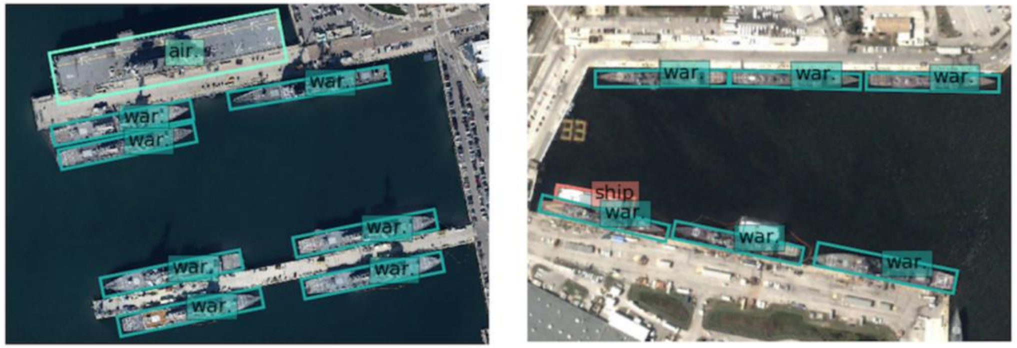
Taegyun Jeon
Satrec Initiative, Korea
tgjeon@satreci.com

RBox-CNN



RBox-RPN
Objectness
Loc(x, y, w, h, θ)
Multi-Angle Anchor
Input Image
Feature map
RRoI Pooling
Concat
Class Scores
Loc(x, y, w, h, θ)

RBox-CNN



(a) HRSC2016

(b) ODAI

Team Name	mAP
USTC-NELSLIP	0.705
jmkoo	0.622
HUST_MCLAB	0.598
NWPU_SAIP	0.578
changzhonghan	0.531
madebyrag	0.506
mfhan	0.42

<ICPR’18 ODAI Contest>

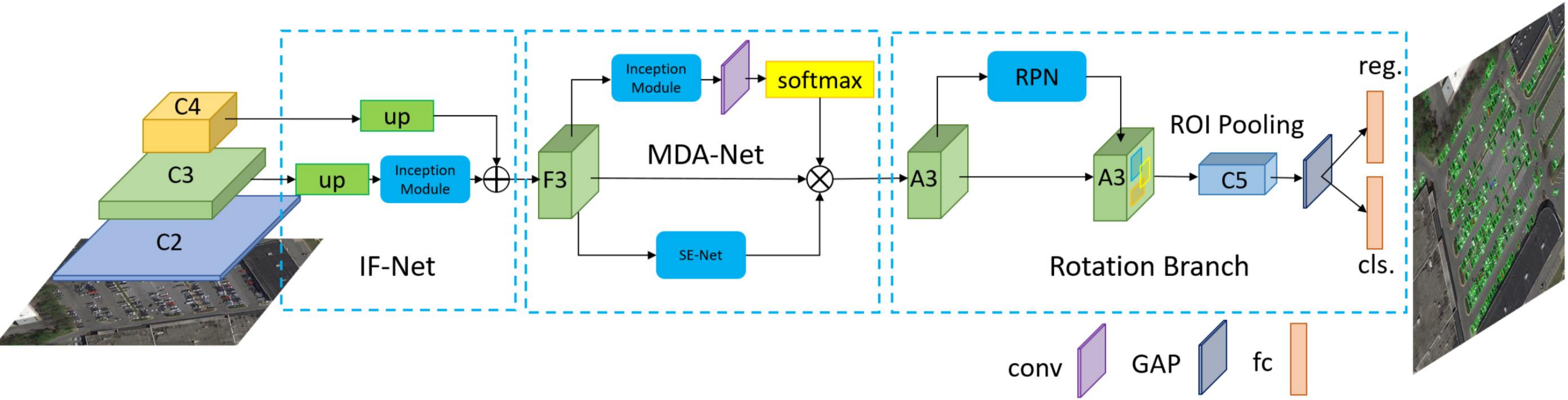
R2CNN++

R²CNN++: Multi-Dimensional Attention Based Rotation Invariant Detector with Robust Anchor Strategy

Xue Yang[1,3], Kun Fu[1,2,3], Hao Sun[1], Jirui Yang[1,3], Zhi Guo[1]
Menglong Yan[1], Tengfei Zhang[1,3], Sun Xian[1]
[1,2]Institute of Electronics, Chinese Academy of Sciences, Beijing (Suzhou), China.
[3]University of Chinese Academy of Sciences, Beijing, China.
yangxue16@mails.ucas.ac.cn

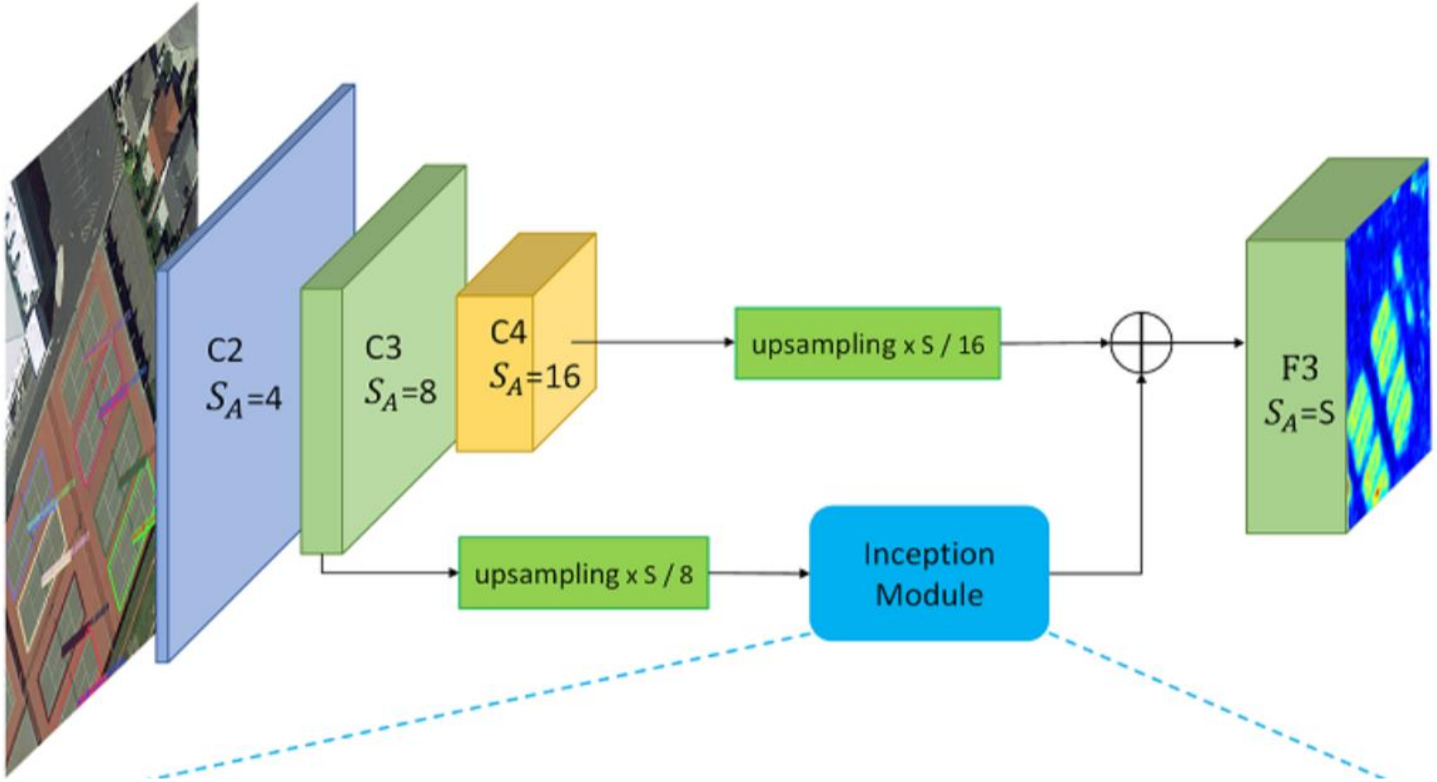
R2CNN++

Yang, Xue, et al. "R2CNN++: Multi-Dimensional Attention Based Rotation Invariant Detector with Robust Anchor Strategy." *arXiv preprint arXiv:1811.07126* (2018).



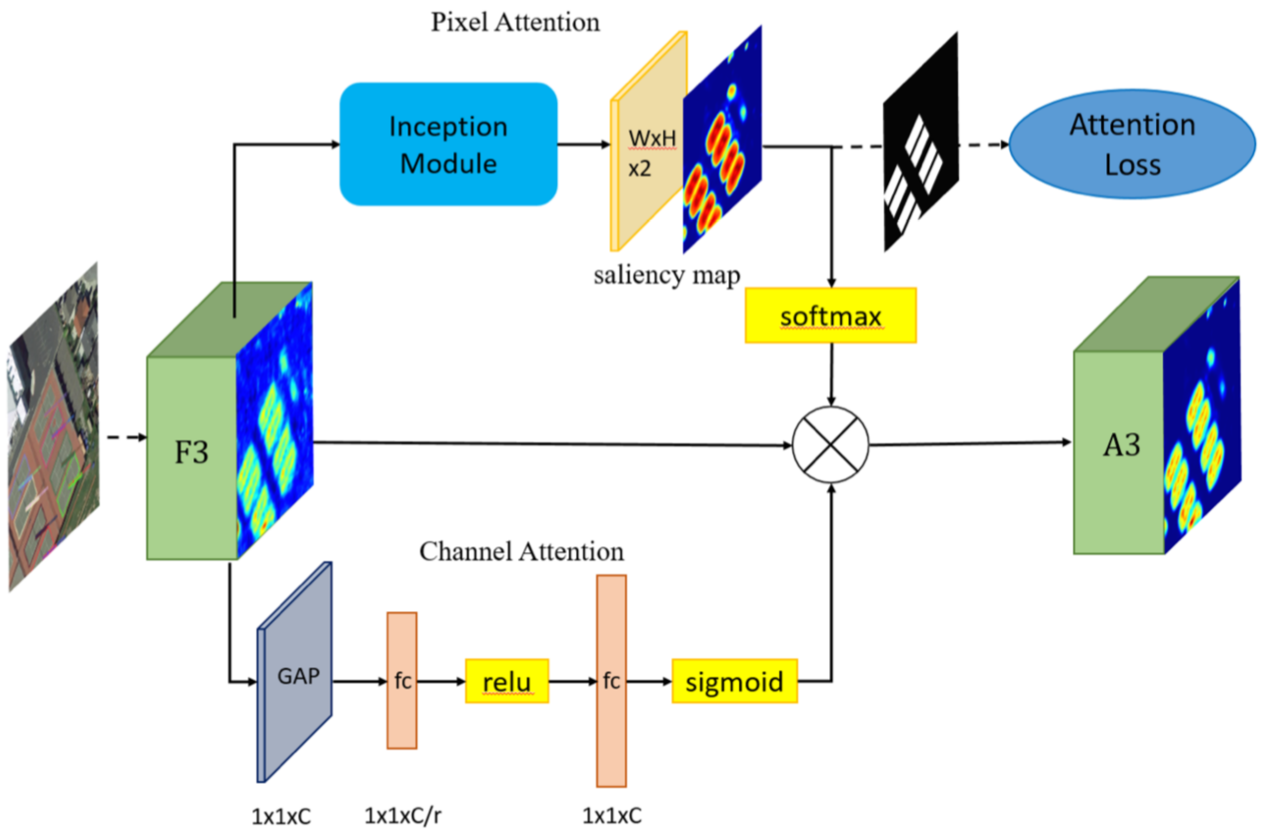
R2CNN++

IF-Net: Inception Fusion Network



R2CNN++

MDANet: Multi-Dimensional Attention Network



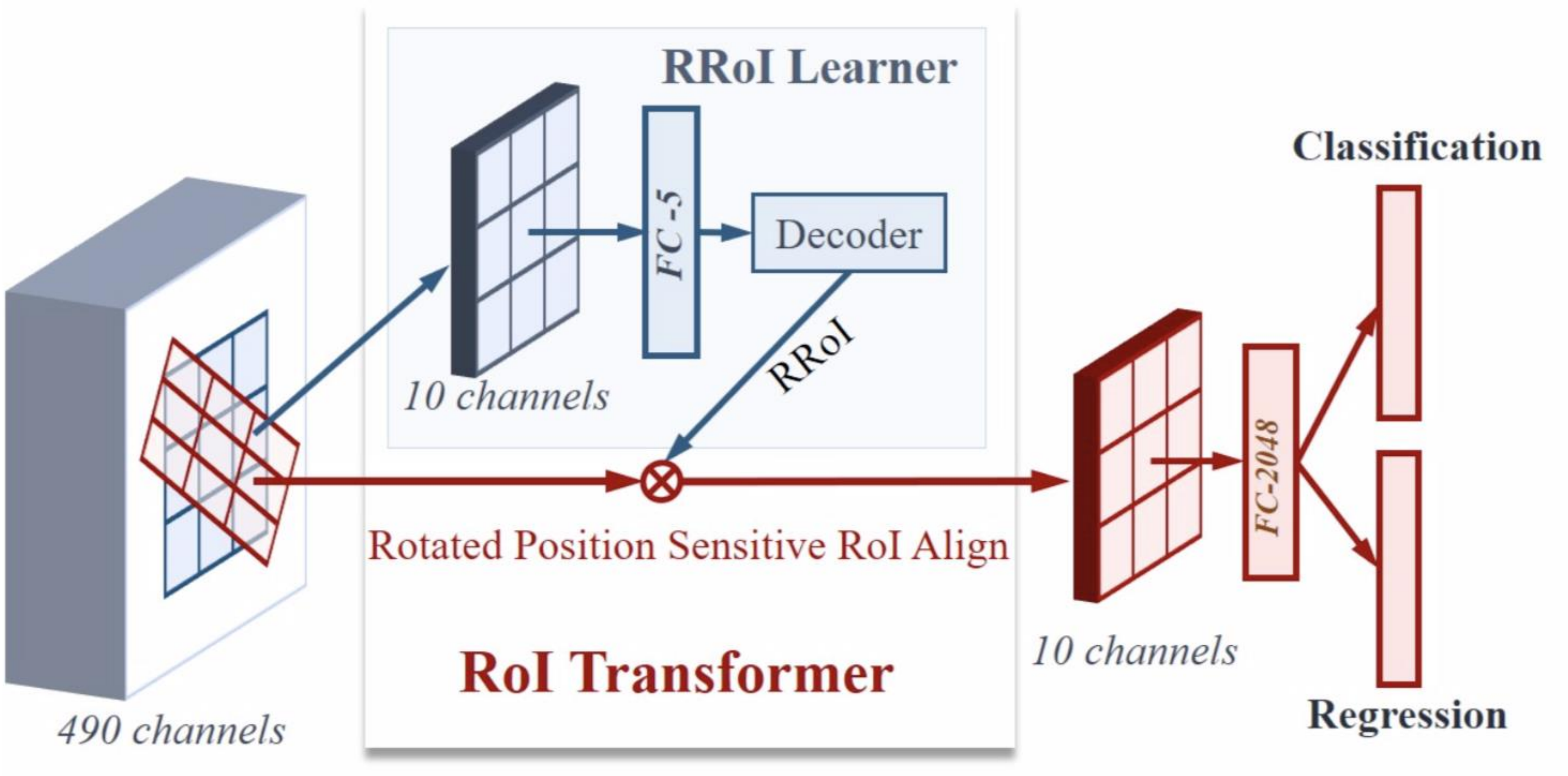
Learning RoI Transformer

Learning RoI Transformer for Detecting Oriented Objects in Aerial Images

Jian Ding, Nan Xue, Yang Long, Gui-Song Xia*, Qikai Lu
LIESMARS-CAPTAIN, Wuhan University, Wuhan, 430079, China
{jian.ding, xuenan, longyang, guisong.xia, qikai_lu}@whu.edu.cn

December 4, 2018

Learning RoI Transformer

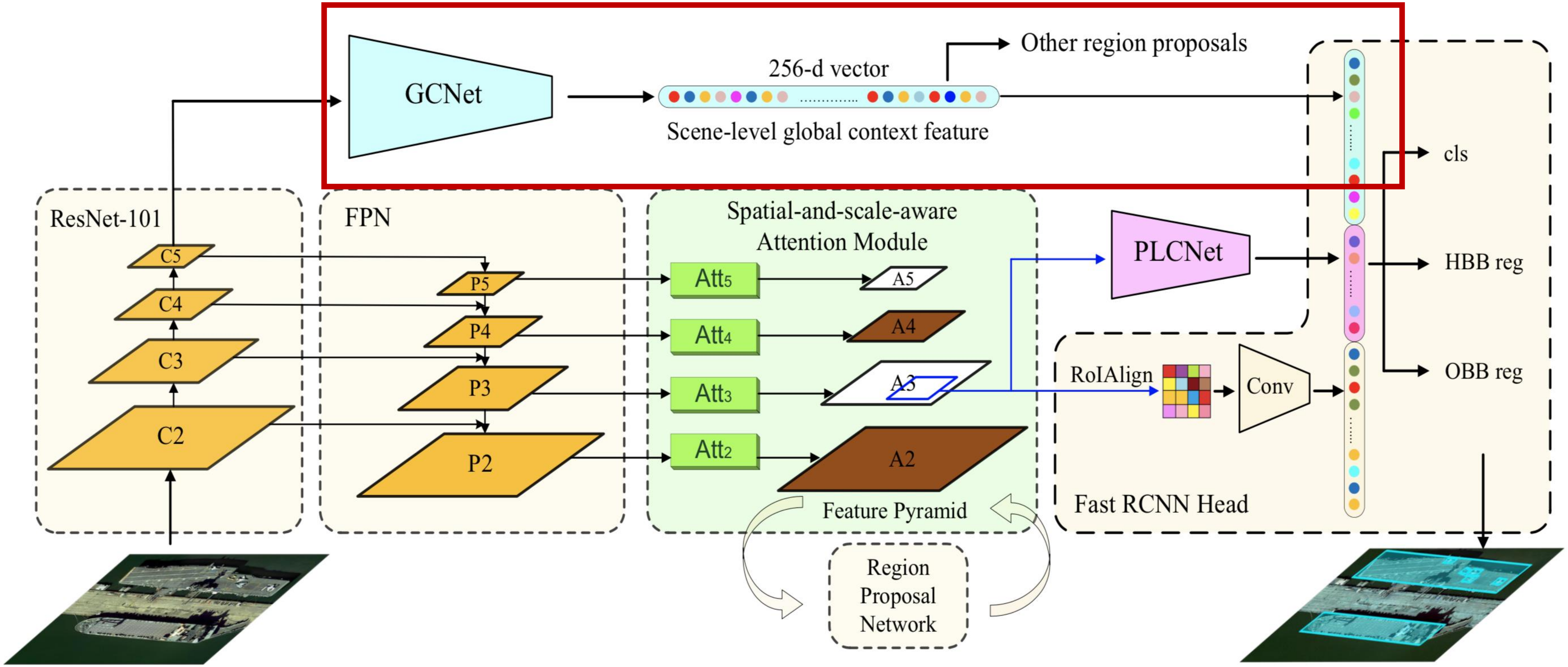


CAD-Net

CAD-Net: A Context-Aware Detection Network for Objects in Remote Sensing Imagery

Gongjie Zhang, Shijian Lu and Wei Zhang

CAD-Net



GCNet
256-d vector
Other region proposals
Scene-level global context feature
ResNet-101
C5
C4
C3
C2
FPN
P5
P4
P3
P2
Spatial-and-scale-aware
Attention Module
Att5
Att4
Att3
Att2
A5
A4
A3
A2
Feature Pyramid
Region
Proposal
Network
PLCNet
RoIAlign
Conv
Fast RCNN Head
cls
HBB reg
OBB reg

Appendix: CVPR Workshop

News

- **2019-03-19**: Release of evaluation code for DOTA-v1.5. **New!**
- **2019-03-19**: Little revision of annotation. **New!**
- **2019-03-05**: A small problem of annotations is fixed and a new-version annotation has been released.
- **2019-03-04**: The challenge is open.
- **2019-03-04**: The new dataset DOTA-v1.5 is available online.

Important Dates

- Registration open: March 5, 2019
- Training/Validation dataset available: March 5, 2019
- The evaluation server of task 1-2 will be available: March 20, 2019
- Challenge submission deadline: April 15, 2019

Thank you