Object Detection
Introduction

Presentation by Jamyoung Koo
2019.03.20
Object Detection using CNN

1. Basics
2. Recent Approach
3. In Remote Sensing
Terms: Object Detection

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

CAT, DOG, DUCK
Terms: IoU (Intersection over Union)

\[
\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}
\]

- Poor: IoU: 0.4034
- Good: IoU: 0.7330
- Excellent: IoU: 0.9264
Terms: NMS (Non-Maximum Suppression)

Modified NMS. Boxes are ranked by their foreground ROI score.
R-CNN: Architecture

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

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

<table>
<thead>
<tr>
<th></th>
<th>R-CNN</th>
<th>Fast R-CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test time per image</td>
<td>47 seconds</td>
<td>0.32 seconds</td>
</tr>
<tr>
<td>(Speedup)</td>
<td>1x</td>
<td>146x</td>
</tr>
<tr>
<td>Test time per image with Selective Search</td>
<td>50 seconds</td>
<td>2 seconds</td>
</tr>
<tr>
<td>(Speedup)</td>
<td>1x</td>
<td>25x</td>
</tr>
</tbody>
</table>
Faster R-CNN: Architecture

Faster R-CNN: Architecture
SSD: Single Shot multibox Detector

FPN: Feature Pyramid Network

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
An Analysis of Scale Invariance in Object Detection – SNIP

Bharat Singh               Larry S. Davis
University of Maryland, College Park
{bharat, lsd}@cs.umd.edu
Problem Definition: Large scale variation

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

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
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.
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 Xiao1,3, and Yuning Jiang4

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

Input Image → FPN → PrRoI-Pooling → RoIs → Jittered RoIs

7x7x256 → FC → 1024 → FC → 1024 → FC → IoU

Standalone IoU-Net

FC → 1024 → FC → 1024 → FC → Classification

B-B-Reg.
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: Architecture
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-CNN

(a) HRSC2016

(b) ODAI

<table>
<thead>
<tr>
<th>Team Name</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>USTC-NELSLIP</td>
<td>0.705</td>
</tr>
<tr>
<td>jmkoo</td>
<td>0.622</td>
</tr>
<tr>
<td>HUST_MCLAB</td>
<td>0.598</td>
</tr>
<tr>
<td>NWPU_SAIP</td>
<td>0.578</td>
</tr>
<tr>
<td>changzhonghan</td>
<td>0.531</td>
</tr>
<tr>
<td>madebyrag</td>
<td>0.506</td>
</tr>
<tr>
<td>mfhan</td>
<td>0.42</td>
</tr>
</tbody>
</table>

<ICPR’18 ODAI Contest>
R2CNN++: Multi-Dimensional Attention Based Rotation Invariant Detector with Robust Anchor Strategy

Xue Yang¹,³, Kun Fu¹,²,³, Hao Sun¹, Jirui Yang¹,³, Zhi Guo¹
Menglong Yan¹, Tengfei Zhang¹,³, Sun Xian¹
¹,²Institute of Electronics, Chinese Academy of Sciences, Beijing (Suzhou), China.
³University of Chinese Academy of Sciences, Beijing, China.

yangxuel6@mails.ucas.ac.cn
R2CNN++

**IF-Net:** Inception Fusion Network
**R2CNN++**

**MDANet**: Multi-Dimensional Attention Network
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

Gongjie Zhang, Shijian Lu and Wei Zhang
CAD-Net

GCNet

256-d vector
Scene-level global context feature

Other region proposals

ResNet-101
C5
C4
C3
C2

FPN
P5
P4
P3
P2

Spatial-and-scale-aware Attention Module
Atte
Att4
Att3
Att2
A5
A4
A3
A2

Feature Pyramid

PLCNet

RoIAlign
Conv

Fast RCNN Head

cls
HBB reg
OBB reg

Region Proposal Network
Appendix: CVPR Workshop

Challenge-2019 on
Object Detection in Aerial Images

June 16, 2019, Long Beach, California.

News

- 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