

Object Detection and Localization

Halil Eralp Koas

January 18, 2020

1 Object Detection

1.1 Motivation

Various models are developed through years in the field of object detection. All these models are developed using different methods. Although all of these models are based on convolutional neural networks, their way of identifying objects in given images or frames are varies. Also, their backbone networks, scanning methods, multi-scale object detection, error functions are varies. These varieties affect the performance of models in different aspects such as small and large object detections, speed of the model, etc.

The aim in this article is to investigate, analyze, and compare the performances of various state-of-the-art object detectors through the time trained on video data. Then, promising detectors will be selected and these selected detectors will be analyzed on Video Object Detection dataset on both mean average precision and frame rate per second basis.

1.2 Literature in Static Object Detection

1.3 Literature in Video Object Detection

2 Detector Features

2.1 Backbone Networks

2.2 Scanning Methods

2.3 Multi-scale Handling

2.4 Loss Functions

2.5 Bells-And-Whistles

3 Static Detector Types

3.1 Two-Stage Detectors

3.1.1 Faster R-CNN

3.1.2 Mask R-CNN

3.2 One-Stage Detectors

3.2.1 You Only Look Once: Unified, Real-Time Object Detection

3.2.2 YOLO9000: Better, Faster, Stronger

3.2.3 YOLOv3: An Incremental Improvement

3.2.4 SSD: Single Shot MultiBox Detector

3.2.5 DSSD: Deconvolutional Single Shot Detector

3.2.6 CornerNet: Detecting Objects as Paired Keypoints

3.2.7 RetinaNet

3.2.8 FPN

3.2.9 EfficientDet: Scalable and Efficient Object Detection

3.2.10 FreeAnchor: Learning to Match Anchors for Visual Object Detection

3

4 Datasets and Metrics

5 Performance Comparison

5.1 Accuracy and Real-Time Applicable

5.2 On COCO

5.2 In Video