Object Detection - Comparison between YOLOv3 and Faster RCNN

An Pham Viet[1], Minh Nguyen Vu Quang[1], and Duc Le Duong Cong[1]

Advanced Program in Computer Science

Abstract. The project is an lucrative opportunity for us to learn and to engage to the field of Machine Learning by first reading and understanding their paper then secondly applying the idea to solve essential problems. This work is the very first approach to penetrate into the field by studying, examining and investigating the object detection problem. The goal of our work is properly using and comparing the execution time of Yolo vs. Faster RCNN algorithm. *abstract* environment.

1 Introduction

Due to the rapid development of computer vision with video analysis and image understanding, object detection has attracted much research attention in recent years. Moreover, it also plays a major role in object tracking for various purposes, such as self driving cars, face detection, medical imaging, robotics, and many more. Those important demands motivate the extensive research into object detection.

2 Solution

In this work, our solution is to apply Yolo algorithm for the object detection problem using COCO, bounding box dataset. To be more precise, we apply YOLOv3 bounding box algorithm for this problem. Furthermore, with an immense curiosity, we planned on applying manifold state-of-the-art algorithms like fast R-CNN with diverse datasets to evaluate and to compare their performance.

2.1 Dataset

Our solution will be examined with COCO dataset. There are 330K images with more than 200K labeled images. One advantage of COCO is it contains 1.5 million object instances belong to 80 object categories and 91 stuff categories. Each image appears with 5 captions per one. Nevertheless, in the *prototype* report we use a dataset of 50 raw images collected randomly from the internet.

2.2 Implementation

We have written a script for automatically fetching *darknet* and a script for executing the Yolo algorithm to draw a bounding box around detected objects (within 80 different object classes).

In the next report, we continue to investigate Faster RCNN algorithm with the small dataset as well as COCO dataset. We will also perform measurements in running time and conduct the comparison between two algorithms.

Fig. 1. Original image

Fig. 2. Result image

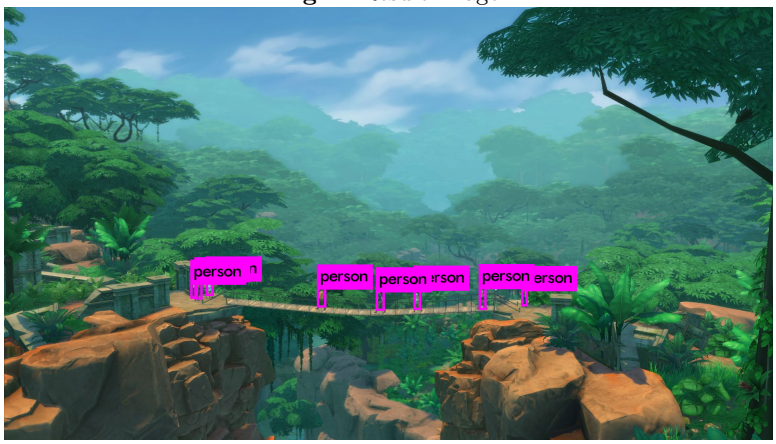
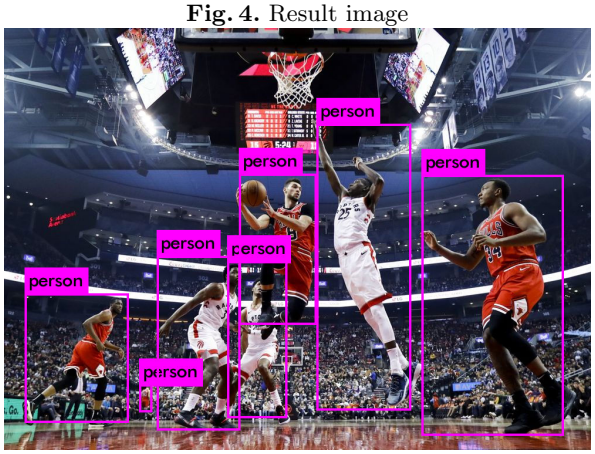


Fig. 3. Original image

Fig. 4. Result image



3 Related Work

There are several methods were proposed in object detection problem:

R-CNN Model[1]: There are two main approaches in this model: applying high-capacity convolutional neural networks to bottom-up region proposals so as to localize and segment objects and supervised pre-training for auxiliary tasks. This model use selective search to extract just 2000 regions from the image called region proposals.Therefore, instead of trying to classify a huge number of regions, it can just work with 2000 regions. However, this approach has some drawbacks: training is a multi-stage pipeline, training is expensive in space and time because of deep networks such as VGG16, which take up huge amounts of space, object detection is slow because it performs a ConvNet forward pass for each object proposal.

Fast R-CNN[2]: The same author of the previous paper(R-CNN) solved some of the drawbacks of R-CNN to build a faster object detection algorithm and it was called Fast R-CNN. In comparison to the R-CNN, Fast R-CNN is faster because it doesn't have to feed 2000 region proposals to the convolutional neural network every time. Instead, the convolution operation is done only once per image and a feature map is generated from it.

Objects as Points[3]: This paper proposes modeling an object as a single point. It uses key point estimation to find center points and regresses to all other object properties. These properties include 3D location, pose orientation, and size. It uses CenterNet, a center point based approach that's faster and more accurate compared to other bounding box detectors.

References

1. Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation"
2. Ross Girshick, "Fast R-CNN"
3. Xingyi Zhou, Dequan Wang, Philipp Krahenbuhl, "Objects as Points"