Real-Time Object Detection

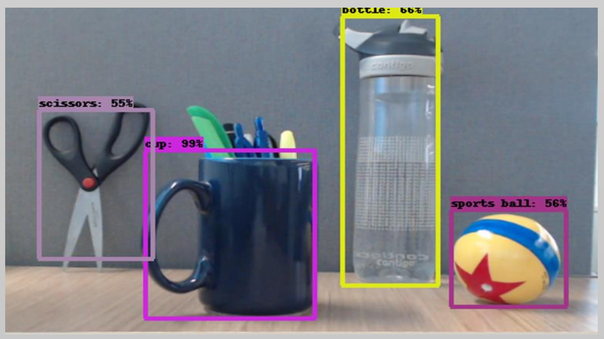
Prerna Kaul
Data Science, Georgetown University

Abstract

Computer Vision has seen a lot of advancement since the introduction of neural networks. Object detection has been a quintessential part of this field and a lot of state-of-the-art techniques exist to tackle tasks in the same area. This paper talks about real time object detection using transfer learning from Mask R-CNN and OpenCV programming functions. The main objective of this technique is to segment the different objects in an image clicked in real-time from a webcam, mask those objects distinguishably and classify them.

Introduction

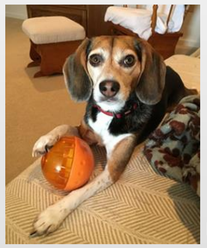
Humans glance at an image and instantly know what objects are in the image. Computers have always been good at dealing with numerical data but analyzing huge amount of data in images was not easy until the rise of deep learning techniques. Making the machines to “see” objects is a significantly harder task in Computer Vision than the traditional Image Classification. However, the most successful approaches to object detection are currently extensions of image classification models.



Data

Common Objects in Context (COCO) is a large-scale object detection dataset. It has over 200k labeled images with 80 object classes. All object instances are annotated with a detailed segmentation mask.

Few of the object class examples are:


DOG


PERSON

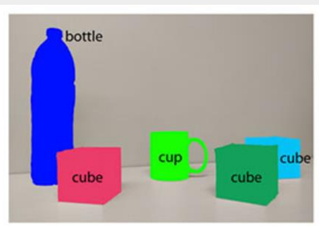

LAPTOP


CHAIR

Methodology

- *Image classification* involves predicting the class of one object in an image.
- *Object localization* refers to identifying the location of one or more objects in an image and drawing a bounding box around their extent.
- *Instance segmentation* refers to computing a pixel-wise mask for every object in the image.
- *Object detection* combines all these tasks and localizes, masks and classifies one or more objects in an image.

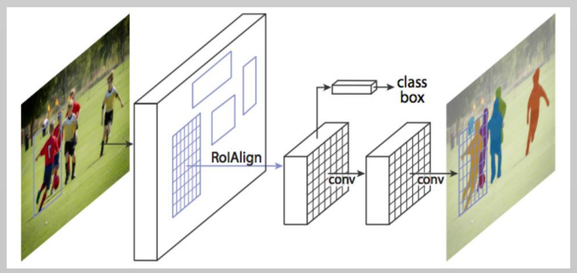
(a) Image classification (b) Object localization



(c) Instance segmentation

This paper uses transfer learning from Mask R-CNN (Region Based CNN) to detect and segment objects in an image. It has 2 modules:

1. Generating proposals about regions of interest (ROI) using deep fully convolutional network
2. Predicting class of the object, refining the bounding box and generating a mask in pixel level of the object based on proposed ROI through Faster RCNN

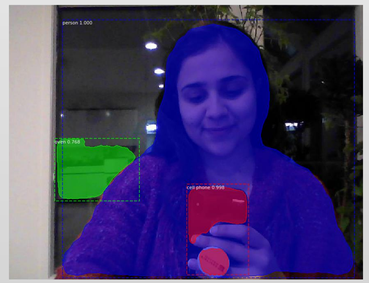


Results

The images fed into the model were clicked in real-time through webcam or android device. The experimented images included people, bottle, bag, remote, cell phone, cars and many more objects.

- The general overall accuracy of classifying and masking objects was pretty good
- Objects that were overlapping were masked appropriately
- The overall time to respond with results after capturing was on an average 19 seconds
- Sometimes the model used to get confused between visually similar objects

Result instances:



References

[1] *Mask RCNN* - Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick
[2] Matterport, Inc - https://github.com/matterport/Mask_RCNN