Object Detection

CS725: Project Proposal

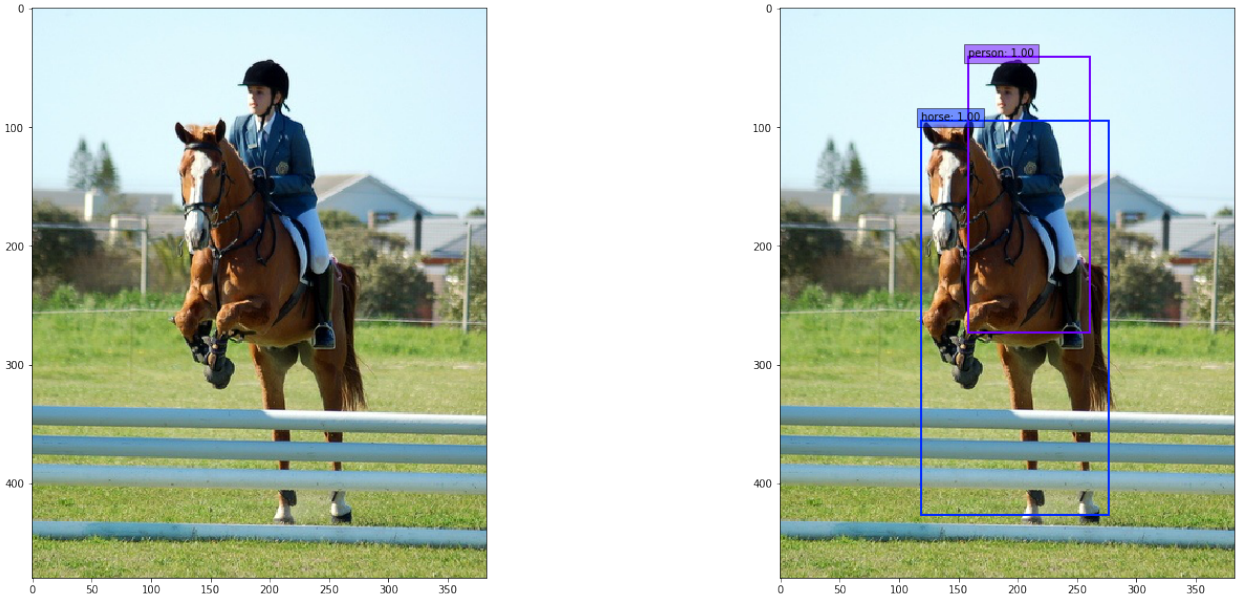
Pratik Kalshetti (163050048) Ashish Jaiswal (163050055)
Naman Rastogi (163050056) Prafull Gangawane (163050080)

Project Description

The aim of this project is to implement a system which is capable of detecting objects in a scene. The input to this system will be an image and the output will be a bounding box around the object along with the class label of the corresponding object. This box will be defined by 4 real numbers; x and y coordinate of the centre of the box, height and width of the box. The output class label will have a score associated with it, which denotes the confidence of the class prediction.

This problem can be formulated as a regression problem for estimating the bounding box and a classification problem for predicting the class. Thus this project serves as a platform to develop a system based on the concepts of classification and regression which are the main contents of the course syllabus.

The goal is to achieve accurate object detection at a high speed.



(a) Input (b) Output

Figure 1: System Specification

Approach

Background The most successful methods in object detection consist of two major approaches - two step detection [1, 2, 3], and unified approach [4, 5]. The two step detection strategy produces accurate results, however it lacks the performance speed which is tackled by a unified approach. Also this approach achieves the goals (accuracy and speed) of this project.

Details The plan is to implement an end-to-end deep neural network based on the ideas of [5]. The task of feature extraction from the images will be achieved by using a pre-trained network on ImageNet data [6] and then fine tune the network for the detection task. The deep network will be a fully convolutional network (FCN). The advantage of using a FCN is that any input image size is allowed. Also the spatial information is preserved by these kinds of networks for images and they tend to be computationally efficient when compared to fully connected layers. The initial layers will help detect smaller objects, while the later layers will be useful for bigger objects. The output from some of these layers will be passed onto specialized networks which will work as classifier and localizer. Finally a non-maximum suppression algorithm will be applied to filter the multiple boxes per object based on their confidence in prediction. During training the error function to minimize consists of a combined classification and regression loss.

A more ambitious plan is to generate a new data set of different objects and test the network for evaluating cross-dataset performance. Another intention is to detect objects in videos for real-time object detection.

Research Papers

The ideas in [5] will help in implementing the network. This paper introduces Single Shot Multibox Detector (SSD) which is the algorithm used in the standard API for object detection in tensorflow, thus proving its impact.

Dataset

For the purpose of this project, the publicly available PASCAL VOC dataset will be used. It consists of 10k annotated images with 20 object classes with 25k object annotations. These images are downloaded from flickr. This dataset is used in the PASCAL VOC Challenge which runs every year since 2006.

Figure 2: Dataset

References

[1] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[2] Ross Girshick. Fast R-CNN. In *International Conference on Computer Vision (ICCV)*, 2015.

[3] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems (NIPS)*, 2015.

[4] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[5] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: Single shot multibox detector. In *ECCV*, 2016.

[6] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.