YOU ONLY LOOK ONCE V1

Group 1

Chinmay Polya Ramesh - 5135125
Godwin Rayan Chandran - 5347998
Nan Lin - 5227623

Object Identification involves identifying which objects are in an image and identify the class. Before YOLO, object detectors were quite slow as it involves extraction of regions in an image and further processed for classification. To overcome such difficulties, YOLO was built by which can be trained, optimized end to end and achieve higher frames/second in real-time.

Introduction

This project relates to the reproduction of YOLO V1 [1] originally developed by Joseph Redmon et al. The convolutional neural network was initially built with C programming language. We aim to implement YOLO V1 using PyTorch to train, test to achieve results as in Table 1 in [1]. Additionally, the results are visualized with webcam images as dataset .

Objective

To understand the classic architecture of YOLO V1 to dive deeper into Object detection. Get hands-on experience with PyTorch and OpenCV framework to execute Object-Detection Projects.

Methodology

We begin the project by collecting the Pascal VOC dataset from 2007 - 2012 for training and testing . Various back-bone architectures such as VGG-16, Resnet-50 etc. are trained with YOLO to identify the best performing based on mAP. The network is loaded with pre-trained ImageNet Dataset weights. The fine-tuning and data augmentation are performed similarly to the paper.

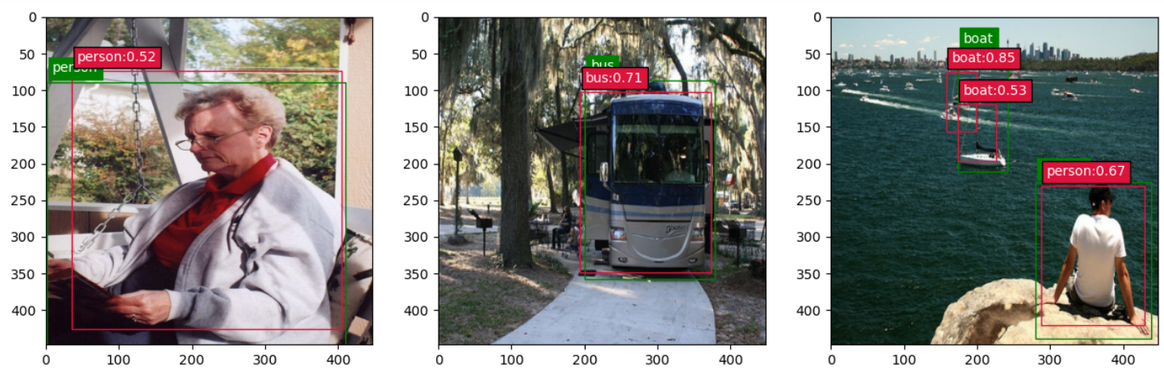
Results

Our reproduced model achieved an mAP of 46.5% on the Pascal Visual Object Classes dataset (2007-2012). And it runs at 120 fps on Nvidia V100, 10 fps on Nvidia GeForce 1660s. It is also implemented for identifying objects in picture/video captured by web cameras.

Model	mAP	FPS	GPU
YOLO	63.4	45	Titan X
ResNet+YOLO	46.5	120	v100

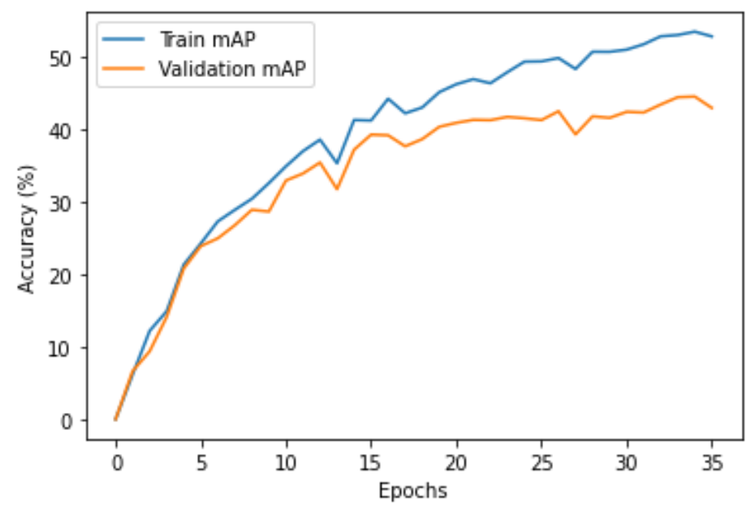
Reproduction

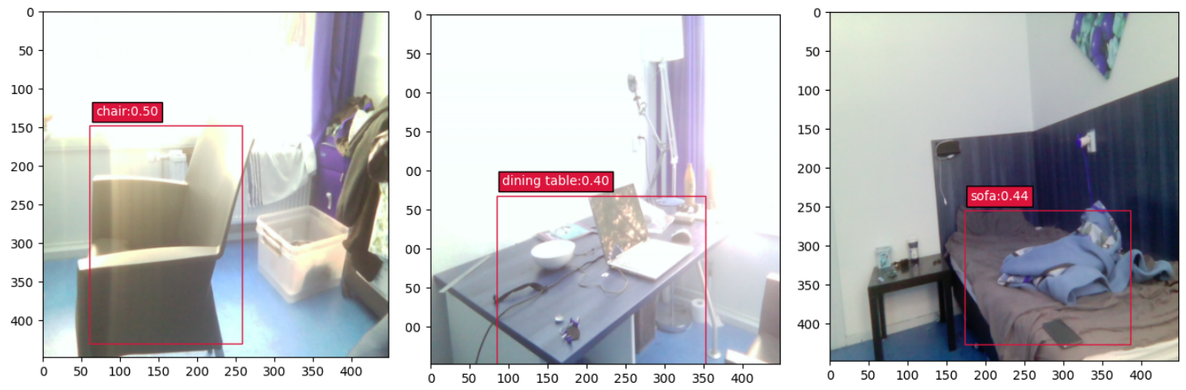
We adopted most of the original settings of the paper while doing some adaptations. Adaptations like Resnet50 instead of darknet 19, more data augmentation, and BatchNorm were used in our model. Additionally, the Pascal VOC dataset was extended by using all images from 2007 to 2012 instead of only 2007 and 2012.



Test images with predicted bounding boxes and ground truth

Train and Validation mAP





WebCam images and predicted bounding boxes

Conclusion

YOLO (v1) is an inspiring architecture for object detection. It is monumental at the time when it was released. Despite that there are better algorithms now, it served as a foundation for those successors and as a touching stone for beginners.

Our reproduction result has proved Yolo to be a solid method in terms of object detection. However, we do wish to further improve it by, for example, adapting its loss function and architecture in future work.

Related Literature

[1] Redmon, J. (2015, June 8). You Only Look Once: Unified, Real-Time Object Detection. ArXiv.Org. https://arxiv.org/abs/1506.02640