MyVision : Vision for Visually Impaired People

Project-2 (CSD405)

Presented By :
Puneet Saluja (2015KUCP1019)
Tanmay Sonkusle (2015KUCP1023)
Krishna Sharma (2015KUCP1040)

Supervisor:
Dr. Isha Pathak Tripathi

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, KOTA

May 17, 2019

Overview

INTRODUCTION

- Object detection is one of the areas of computer vision that is maturing very rapidly.

Figure: Detection and Classification[1]

[1] From paper by Joseph Redmon *et. al* [1]

OBJECTIVES

- To create an Android Application that would assist people with visual impairment in analyzing their surroundings.
- Learn the objects in the picture and will provide voice assistance, about the types of objects nearby.

OVERVIEW

- User can open the application with help of any virtual assistance.
- The application capture the image of the surrounding.
- The image is then fed into the pre-trained CNN and detected objects in the picture will get labeled.
- Labels will then be passed to the text to speech engine to convert the text to speech and assist the visually impaired user.

LITERATURE REVIEW [1/2]

Paper	Methodology
Joseph Redmon *et. al.* [5] (2016)	Single neural network predicts bounding boxes and class probabilities.
Ross B. Girshick *et. al.* [3] (2013)	Used region-based convolutional network (R-CNN) to scan possible object regions using selective search for object detection
R. Girshick *et. al.* [2] (CVPR, 2015)	Performed feature extraction over image before proposing regions and then run CNN for object detection.

LITERATURE REVIEW [2/2]

Paper	Objective
Kaiming and Girshick *et. al.* [6] (2015)	Introduced region proposal network (RPN) to generate region proposal and detect objects.
Jifeng Dai *et. al.* [1] (2016)	Used region-based fully convolutional network (R-FCN) to detect objects.
Liu *et. al.* [4] (2016)	Used single shot detector(SSD) to simultaneously predict the boundary box and class as it process the image.

WHY YOLO ?

- Detection as a single Regression Problem.
- Developed as a single Convolution Network.
- Reason Globally on the entire Image.
- Learns Generalizable Representations.

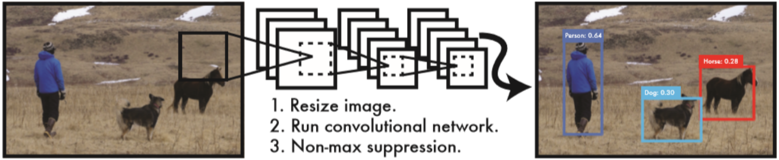
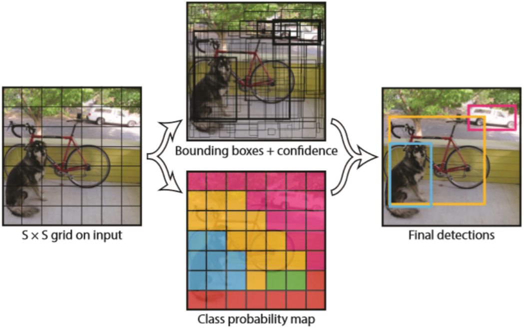


Figure: The YOLO Detection System[2]

[2] From paper by Joseph Redmon *et. al* [1]

METHODOLOGY

- Divide the image into a SxS grid.
- Each grid cell predict:
  - B bounding boxes
  - B confidence scores as C=Pr(Obj)*IOU
  - C conditional Class probability as P=Pr(Class/Object)
- Confidence Prediction is obtained as IOU of predicted box and any ground truth box.



ARCHITECTURE

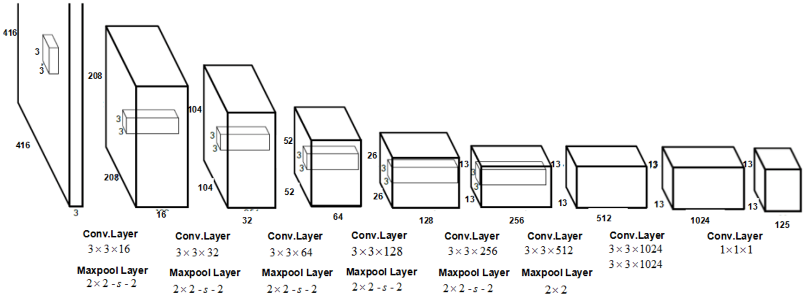


Figure: Architecture[3]

[3] From paper by Joseph Redmon *et. al* [1]

Loss Function

$$\lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$
$$+ \lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$
$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$
$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$
$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

Figure: Loss Function

FlowChart

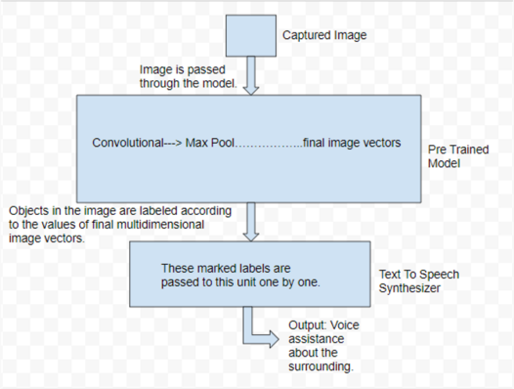


Figure: FlowChart

Android Application

- Trained the tiny YOLOv2 model on Pascal VOC dataset.
- After the training of the data we got the weights file for the model.
- The weights file is the converted into to the tensorflow model using the darkflow.
- Then the tensorflow model is used in the android application to detect the labels in the images and the labels are then passed to the android text to speech API for voice assistance.

- Device Used : OnePlus 5

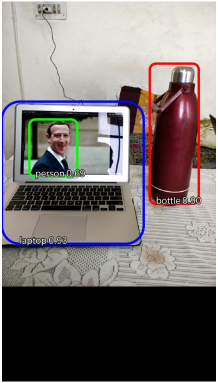


Figure: Working of Android Application

Results and Conclusion

- We have tested YOLOv1 and YOLOv2 on android
- YOLOv2 is successfully able to detect the objects with upto 70% to 80% confidence value with minimum confidence threshold of 60%.
- YOLOv1 is a little slow and for classification the confidence threshold has to be reduced to 25%.

	mAP	FPS
YoloV1	43.3	181
YoloV2	57.1	207

FUTURE WORK

- This work can be extended to do object detection for some custom objects.
- To further improve the accuracy of the given model and make it more robust for processing in mobile devices.

References I

J. Dai, Y. Li, K. He, and J. Sun.
R-FCN: object detection via region-based fully convolutional networks.
CoRR, abs/1605.06409, 2016.

R. Girshick.
Fast r-cnn.
In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

R. B. Girshick, J. Donahue, T. Darrell, and J. Malik.
Rich feature hierarchies for accurate object detection and semantic segmentation.
CoRR, abs/1311.2524, 2013.

References II

W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg.
Ssd: Single shot multibox detector.
In B. Leibe, J. Matas, N. Sebe, and M. Welling, editors, *Computer Vision – ECCV 2016*, pages 21–37, Cham, 2016. Springer International Publishing.

J. Redmon, S. Divvala, R. Girshick, and A. Farhadi.
You only look once: Unified, real-time object detection.
In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

References III

S. Ren, K. He, R. Girshick, and J. Sun.
Faster r-cnn: Towards real-time object detection with region proposal networks.
In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 91–99. Curran Associates, Inc., 2015.