



OBJECT DETECTION SYSTEM FOR THE BLIND WITH VOICEGUIDANCE

Miss Rajeshvaree Ravindra Karmarkar

Department Of Computer Science And Engineering Walchand College Of Engineering
Sangli, India

Prof.V.N. Honmane

Department Of Computer Science And Engineering Walchand College Of Engineering
Sangli, India

Abstract—As object recognition technology has developed recently, various technologies have been applied to autonomous vehicles, robots, and industrial facilities. However, the benefits of these technologies are not reaching the visually impaired, who need it the most. This paper proposed an object detection system for the blind using deep learning technologies. Furthermore, a voice guidance technique is used to inform sight impaired persons as to the location of objects. The object recognition deep learning model utilizes the You Only Look Once (YOLO) algorithm and a voice announcement is synthesized using text-to-speech (TTS) to make it easier for the blind to get information about objects. As a result, it implements an efficient object-detection system that helps the blind find objects in a specific space without help from others, and the system is analyzed through experiments to verify performance.

Index Terms—object detection, TTS technology, YOLO algorithm, position estimation

I. INTRODUCTION

With the recent rapid development of information technology (IT), a lot of research has been carried out to solve inconveniences in everyday life, and as a result, various conveniences for people have been provided. Nevertheless, there are still many inconveniences for the visually impaired. The greatest inconveniences that a blind person feels in everyday life include finding information about objects and indoor mobility problems. Previous studies included object analysis using ultrasonic sensors. However, with these methods, it is difficult to know exactly where an object is located, especially in the presence of obstacles. In this paper, we analyze accurate object information and obtain a location using a deep learning object recognition technique. Object recognition algorithms are designed based on the You

Only Look Once architecture, an object recognition deep learning model, to detect objects using a camera. In this dissertation, system analyzes accurate object information and obtain a location using a deep learning object recognition technique. In addition, voice guidance technologies are synthesized so the visually impaired can know the location of the objects with voice guidance. In the voice guidance technology, the technique of synthesizing the position of the object so it can be output, and synthesizing the name of the object, is done by using text-to-speech (TTS).

II. LITERATURE SURVEY

[1] This paper explained the current methods for detection models and also gave the standard datasets. This paper covered different detectors like one-stage, two-stage which helped in analyzing different object detection methods and also gathered some traditional as well as new applications. It also listed some branches related to object detection. Also pointed out some development trends to better follow the set of art algorithm and further process

[2] This paper presented region based fully convolution network. R-FCNN is used for precise and efficient object detection. Thus this paper can easily adopt ResNets like fully convolution image classifier backbones for detecting the object. This paper provided a simple but efficient R-FCNN framework for object detection. This method achieves same accuracy as compared to faster R-FCNN. Thus this helped to adopt the state of art image classification backbones.

[3] This Challenge is considered as the guideline for object classification and object detection. This paper classified and detected more 100+ object categories and 1M+ images. This paper provided the large scale data collection process. And also described the most efficient algorithm for this data and also discussed the success and failure of other algorithms.

[4] The result of this research showed that grids of



oriented gradient performs better than current feature set for human recognition.

III. METHODOLOGY

A. Preprocessing Image:

- Read image: -Very first step of image processing is read the training as well as testing images.

- Converting color image to grayscale image: -color isn't required to recognize and detect an image -Color images are represented in three channels (i.e. Red, Blue and Green) perhaps gray image contains only one channel. - converting to gray scale image reduces time consumption for processing images as well as extra space required to store in memory

- Standardize Images: -Images must be resized to basic heights and widths before they are fed to algorithm. -for that standardizing is done in case of preprocessing and scaling of image.

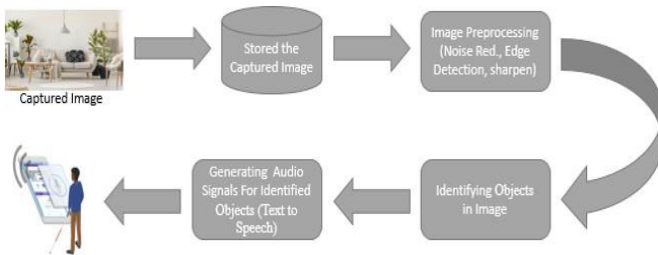


Fig. 1. System Design

B. Identifying objects in image

- YOLO(You Only Look Once): YOLO algorithm predicts from which class object is from and also helps to give the location of object Following components of bounding boxes are used to get the location of object center, width, height, value c defines which class object is from and In addition pc value is used to get the possibility of object present in the bounding box.

In this algorithm only region containing object is not taken into account but the whole image is divided into grids. This is an algorithm used for object recognition Other algorithms such as variants of cnn does not predict the location of object. But in case of YOLO it useful to predict the location using bounding boxes. Also YOLO can detect multiple objects in a single image unlike other algorithms which classifies objects into categories only. YOLO uses single neural network on the image and not uses multiple multiple neural network.

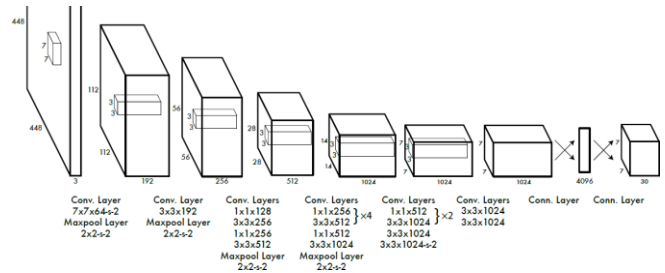


Fig. 2. YOLO Architecture

fig 2 represents the architecture of YOLO Algorithm. There are 2 FC layers with 24 convolutional layers in detection n/w. That makes alternating convolutional layer. These convolutional layers decrease the feature space of previous layers.

In this algorithm, network divides the images into mxm grids which is also called as regions.

Further this algorithm decides the bounding boxes and also gives the class probability map which predicts the possibility of object present in certain grid.

These bounding boxes contains weight depending on how much object is present in the region.

YOLO estimates the type and location of objects using regression inference on the problem of area selection and classification.

YOLO processes 45 frames per second so it is incredibly fast and also efficient as it predicts more than one objects from one image.

- Bounding Box Regression: Bounding Box is used to high- light the object

y =	pc
	bx
	by
	bh
	bw
	c1
	c2
c3	

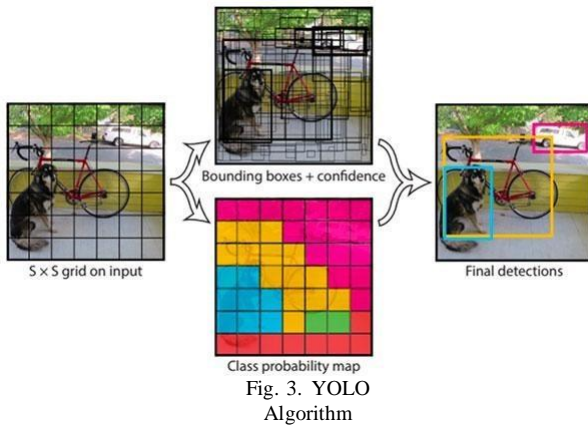
Here,

Pc:Probability of object present in region bx,by,bh,bw: Dimension of bounding box c1,c2,c3:Classes of predicted object

YOLO uses above dimensions to predict the label of object.

- Intersection Over Union(IOU): It determines how predicted box overlaps on determined box. YOLO ensures that these two boxes are overlapped equally.

- Combination of three techniques: Given image shows how three techniques are combined by YOLO.



As Shown in fig 3, it will divide the input image in $S \times S$ grid, then on each grid edge detection filters (Canny Filter) will be applied. With Edges and their connectedness bounding boxes (The rectangular box in which object is enclosed) are generated. Features of each grid in particular bounding box is concatenated and object is predicted using trained model Dataset: In YOLO we will train our object detector model on COCO dataset.

Dataset: The COCO dataset consists of 80 labels and 330k images. Dataset consists labels like:

- Animals
- Flowers
- People
- Traffic signals
- buildings
- devices like mobile phone, computer, laptop, etc.
- Food items
- And many like this...

C. Estimating the position of image:

To estimate the position of image, we need to create the bounding box for every detected object. Then with the height width of particular bounding box, with respect to image frame.

5 values are used to estimate the position of object in bounding box.

The first 4 values b_x, b_y, b_w, b_h shows the position of object.

The 5th value is BC: defines how much box contains an object

BC is calculated as multiplication of object exist in box and IOU (i.e Intersection Over Union)

As labelled data should be passed forward for training 'y' is considered as label and it can be defined as,

- Triangle Similarity [11] As object gets closer to the camera its width increases and angle at the camera. triangle similarity has formula as:

$$F = (P \times D) / W$$

As continue to move the camera both closer and farther away from the object/marker, applied the triangle similarity to determine the distance of the object to the camera:

$$D' = (W \times F) / P$$

D. Generating audio signal for identified object:

Voice guidance is a function that conveniently provides information to specific users, such as people who are visually impaired, when the system finds the location of the object they want to find

The identified class label of detected object and the estimated portion of object is concatenated as text and for voice feedback Google Text-to-Speech API can be used

Google Cloud TTS API:- It allows users to merge hundreds of natural voices

- It is also available in multiple languages.

- It is an easy to use api as developer can interact with user with multiple languages, applications and devices.

IV. RESULT AND DISCUSSION

In this study we have developed an android application for object detection

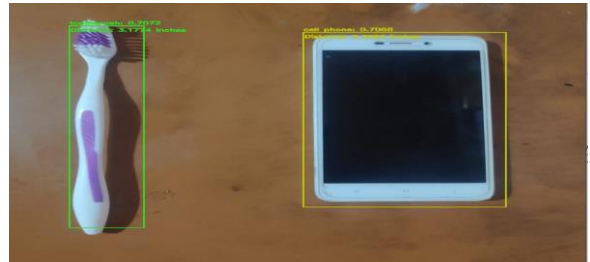


Fig. 4. Object detection with Position

The system displays output as the name of the object and the probability of the object in percentage. So the system will only detect those objects having probability greater than given threshold probability.

As well as system uses google API to convert text into speech and gives the position of object in inches using device's speakers.

tiny YOLO is used in order to implement the system on android platform as a result it reduces the accuracy of object detection.



Fig. 5. Android Application screenshot

Fig represents the snapshot of android application. It displays the name of object and its probability.

whilst the application also tells the class label and how far the object from camera through device's speakers.

V. SIGNIFICANCE

Current system provides help for blind people.

-This system can be embed in mobile phones or any similar handy devices.

-This system will help visually impaired people to know the information (like name and location) of object.

VI. CONCLUSION AND FUTURE WORK

Develop the object detection system with the help of deep learning algorithm such as YOLO and estimate the position of object with the help of same.

-This system provides the voice guidance for visually impaired people. This system is solely designed to help the blind people effectively.

However the accuracy can be increased. Also the current system works on android Operating system which can be further developed in order to be compatible on every convenience devices.

VII. ACKNOWLEDGEMENT

This paper is inspired from "Object Detection System for the Blind with Voice Command and Guidance" presented from Wonju National University / Gangneung,

Gangwon 25457, Korea.

REFERENCES

- [1] Choi D., and Kim M. , "Trends on Object Detection Techniques Based on Deep Learning," *Electronics and Telecommunications Trends*, Vol. 33, No. 4, pp. 23-32, Aug. 2018.
- [2] J. Dai et al. , "R-FCN: Object Detection via Region-based Fully Convolutional Networks." *Conf. Neural Inform. Process. Syst., Barcelona, Spain*, Dec. 4-6, 2016, pp. 379-387
- [3] Dalal N. and Triggs B., , "Histograms of Oriented Gradients for Human Detection," *IEEE Comput. Soc. Conf. Comput. Vision Pattern Recogn., San Diego, CA, USA*, June 20-25, 2015, pp. 886-893
- [4] O. Russakovsky et al., , "ImageNet Large Scale Visual Recognition Challenge," *Int.J. Comput. Vision*, vol. 115, no. 3, Dec. 2015, pp. 211-252.
- [5] Ahonen T., and Pietikainen M., , "Face Description with Local Binary Patterns: Application to Face Recognition." *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, , VOL. 28, NO.12, DECEMBER 2006
- [6] Raj A., Kannaujiya M., Bharti A., Prasad R., Singh N., Bhardwaj I., , "Model for Object Detection using Computer Vision and Machine Learning for Decision Making .," *International Journal of Computer Applications (0975 - 8887)*, Volume 181 - No. 43, March 2019
- [7] W. Liu et al., , "SSD: Single Shot MultiBox Detector," *European Conference on Computer Vision*, Vol. 9905, pp. 21-37, Sept. 2016
- [8] J. Redmon et al., , "You Only Look Once: Unified, Real-Time Object Detection." *in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, Jun. 2016
- [9] Moonsik K., , "Object Detection System for the Blind with Voice Command and Guidance." *IEIE Transactions on Smart Processing and Computing*, vol. 8, no. 5, October 2019
- [10] YOLO: Real-Time Object Detection, <https://pjreddie.com/darknet/yolo>
- [11] Triangle Similarity <https://www.pyimagesearch.com/>