REAL TIME OBJECT DETECTION USING YOLOV4 FOR VISUALLY CHALLENGED PEOPLE

Dr. Ramkumar. M[1], Shivani. K[2], Snegha. C[3], Sowmya. S[4]

[1]Assistant Professor
Department of Electronics and Communication Engineering
Sri Krishna College of Engineering & Technology, Coimbatore
ramkumarm@skcet.ac.in

[2]U.G Scholar
Department of Electronics and Communication Engineering
Sri Krishna College of Engineering and Technology, Coimbatore
19euec137@skcet.ac.in

[3]U.G Scholar
Department of Electronics and Communication Engineering
Sri Krishna College of Engineering and Technology, Coimbatore
19euec145@skcet.ac.in

[4]U.G Scholar
Department of Electronics and Communication Engineering
Sri Krishna College of Engineering and Technology, Coimbatore
19euec150@skcet.ac.in

Abstract

Vision is the most important sense for humans. Visually challenged people face many difficulties to travel independently and also, they face issues in accessing information and communication. Our project will help visually challenged people to travel alone without any external help by using the real-time object detection system. To determine the real time objects the application uses image processing and machine language which assists visually challenged people about the object present before them through the voice output. The object detection model is done using YOLOv4 algorithm, where classification is done using darknet53 algorithm, and the output of this model is fed into gtts API which gives the voice output of the detected objects.

Keywords*: YOLO, object detection, visuallychallenged, image processing.*

1. INTRODUCTION

Any decreased ability to visualize the outside world is termed as low vison or blindness. This decreased ability makes the daily activity of the people very difficult. This proposed system provides solution for visually challenged people by identifying the objects in real time. Web application or android phone cameras are used to capture the image of the object. When the application is launched it immediately starts to capture the object present in real time and the identified object is conveyed via voice output. Detection of any kind of objects is possible through this object detection. Object can be captured and imported or it can be able to identify by video streaming. Examples are pen, dress, bat, ball, pencil, utensils, animals, birds, trees, food and other such necessary objects can also be identified which are very useful to the visually impaired people to carry out their daily activities. This system is aimed at providing easy to use application which makes daily activities of visually challenged people very convenient.

2.LITERATURE SURVEY

YOLOv4 detects the object with high accuracy when compared to other algorithms like CNN, RCNN etc. YOLOv4 is more efficient because it is a mixture of Convolutional Neural Networks (CNNs) and sliding window approach which can achieve 65.7% Average Precision accuracy in accordance with the Microsoft COCO dataset. The first model that produced a 30% improvement in object detection is RCNN. The very similar approach to RCNN is Fast RCNN, Selective search was used in Fast RCNN to detect an object. After Fast RCNN, Faster RCNN came in for object detection.

Though Selective Search was serviceable it took lot of time to detect the object. In 2015, SSD [Single Shot MultiBox Detector] came to detect multiple objects at single shot. It also increased the detection rate. Anchors were used in the SSD to count the default regions. From the name, we can clearly say that it takes a single shot to detect multiple images. The YOLO group of architectures were constructed in the same vein as the SSD architectures but YOLO is more advantageous than any other method for object detection as it has high accuracy in object detection.

3.EXISITING SYSTEM

There is various real time object detection with voice output models using different algorithms like CNN, RCNN, Faster RCNN, YOLOv3 etc. The problem with these algorithms is accuracy is less and real time speed of object detection is low.

4.PROPOSED SYSTEM

The proposed system uses YOLOv4 algorithm. YOLOv4 is termed as single stage detection model in which mean Average precision is increased by 10% and No. of frames per second processed is increased by 12% when compared to other algorithms. Other than this we are using python 3. The camera in the 2penCV library starts to capture the images. It captures at a rate

of 30 images per second. These images are fed into the YOLOv4 algorithm for object detection. Objects placed before camera is identified by Darkne-53 algorithm.Google Text to Speech convertor converts the identified object into audio. This audio is the output of the system which tells all the objects present to the person.
This helps the visually challenged people to visualize the objects around them. This system will also protect them from colliding with the objects and hence securing from injuries.

5.IMPLEMENTAION

A. Difference between YOLOV4 and other algorithms

Other algorithms (CNN, RCNN etc..) uses a selective search to find whether an object is present, detects and predicts. But YOLO separates the image into grids and check whether the object is present in each grid or not, detects and predicts it. Boundary box is done using IOU (Intersection Over Union), non-max suppression techniques. Number of objects detected in single frame is fixed in other algorithms but in YOLO any number of objects can be detected.

YOLO is one stage detection model while CNN, RCNN are two stage object detection model. Mean average precision increased by 10% when compared to object detection using other algorithms and also number of frames processed per second is increased by 12%.

B. Data set:

The COCO [Common Objects in Context] dataset is used in this project. COCO dataset consists of 91 object dataset categories, almost 82 objects have 5000 labelled instances. It has 3,28,000 images and 25,00,000 instances.

C. YOLOv4 Architecture:

In real time object detection using YOLOv4 the image goes through several convolutional layers to form a feature map. The images are divided into grid cells. Each grid cells in YOLO algorithm generate two anchors each. In every object detection model the following steps take place, data augmentation is done on the input images where the different orientation of same images is trained to improve the accuracy. Then Normalization is done to improve the quality of images.
Regularization is done to adjust the output within the range. Loss functions are used to calculate the losses and tries to reduce the loss using backtracking algorithms. Classification and bounding boxes are done on the images captured by camera and detection is done. The work is done by dividing the detection task into two categories, one is detection and other one is classification. We use darknet framework for implementation of YOLOv4.
YOLOv4 has three important parts:

- Backbone
- Neck

- Head

CutMix,Mosaic data augumentaion,class label smoothing and dropblock regularization are used to increase the classifier training accuracy.Mish activation function is also used in addition for classification and training.

Instead of using a single image Mosaic data augmentation uses 4-image at the same time for better processing.

CutMix data augmentation: we use different orientation and random patches between the training images. Localization ability is increased on less discriminative partsof the object to be classified.

Class label smoothing: *Label smoothing* is a regularization technique that addresses overfitting and overconfidence while classification.Mish Activation function: It is a self-regularized function which is defined in mathematical terms below:

f(x)=x tanh(soft plus(x))

The detector performance is increased by using SPP,PAN, and SAM.

SPP: SPP (Spatial pyramid pooling) On simultaneous pooling on multiple kernel sizes Spatial pyramid pooling acquires both the coarse and fine information which are required for further process.

PAN: Path Aggregation Network is a technique that uses maximum information in layers close to the input by inspecting features from different convolution layers.

SAM: Modified Spatial Attention Module is used to highlight the most important and miniature features.

For further optimization CSPNet divides the feature map of the base layer into two segments and then merging them together using a cross-stage hierarchy. In yolov4 CIoU-loss is used to reduce the errors in boundingboxes by using midpoints of actual bounding boxes and predicted bounding boxes. CmBN(cross mini batch normalization) technique is used, and the results is that it decreases the cost of training. Bag Of Freebies for backbone are Class label smoothing, DropBlock regularization, CutMix . Bag of Specials for backbone are Cross-stage partial connections and Mish activation. Bag of Freebies for detector are CmBN, DropBlock regularization, Self- Adversarial Training, eliminate grid sensitivity, Cosine annealing scheduler, Optimal hyperparameters, Random training shapes. Bag of Specials (BoS) for detector are Mish activation, SPP block, SAM-block, PAN path-aggregation block, DIoUNMS. After the introduction of Bag of features and Bag of species classification and image detection became very easy and anyone can use this for training the model.

6. SYSTEM DESIGN

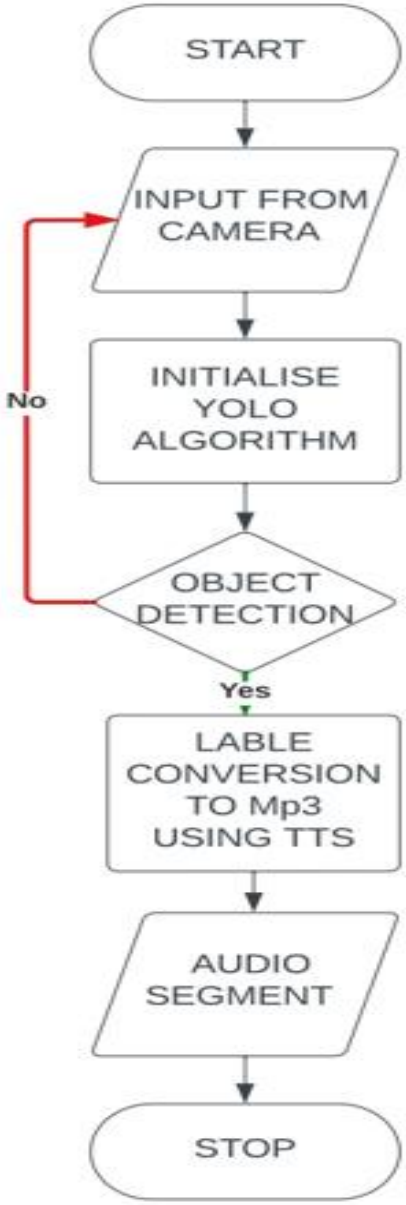

Fig1.System Design

The camera is launched when the user starts the application. This application captures the images of real time objects process it, and the results are fed into text-to-speech converter and the final output is conveyed to the user through voice output. There is no lag between the process since it is asynchronous and tends to be fast.

7.BLOCK DIAGRAM

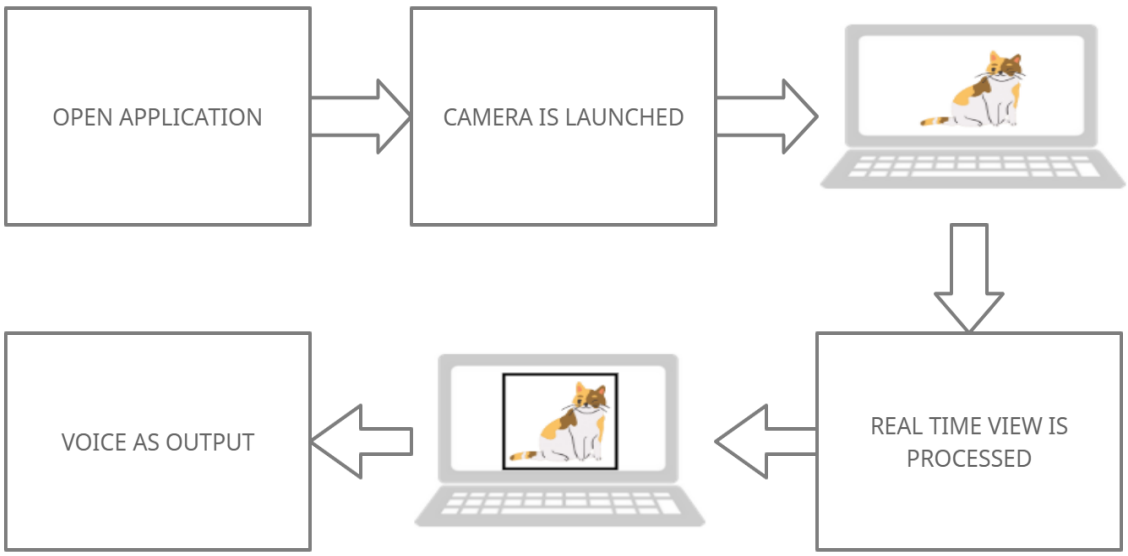

Fig 2.Block Diagram

ALGORITHM	SPEED	IMAGES
R-CNN	.05FPS	20S/IMG
FAST R-CNN	.5FPS	2S/IMG
FASTER R-CNN	.7FPS	140MS/IMG
YOLO	45FPS	22MS/IMG

Table 1: Different Object Detection Methods

8.RESULT ANALYSIS

Results are analyzed by comparing YOLO algorithm with other algorithms like CNN, fast RCNN.Instead of taking the whole image RCNN process by checking the probability of the image and it takes more time. In case of fast R-CNN as it uses the selective search, the accuracy is very low compared to other algorithms.

By comparing the results from different algorithms, we conclude that YOLO is the best suited algorithm for object detection and identification. YOLO (You Only Look Once) takes the whole image and process the whole image in a single instance with high speed and accuracy. The main feature of YOLO is that it can process 45 frames per second.

9.FUTURE SCOPE

In future the project focus on designing a system with better accuracy and less detection time. Enhancing the system by adding the additional features which will give complete support to visually challenged people. In addition to face recognition feature, the application can be trained to store the information about the people closely related to the person, which would help themto differentiate between peers and strangers.

REFERENCES

[1] Real-Time Object Detection for Visually Challenged People Sunit Vaidya Information Technology Department Sardar Patel Institute of Technology Mumbai, India sunit.vaidya1@gmail.com

[2] Jamal S. Zraqou, Wissam M. AlKhadour andMohammad Z. Siam Real-Time Objects Recognition Approach for Assisting Blind People 2017 International Journal of Current Engineering and Technology

[3] Sik-Ho Tsang YOLOv1 — You Only Look Once (Object Detection) https:// towardsdatascience.com/yolov1-you-only-look- once-objectdetection-e1f3ffec8a89

[4] Rui Li , Jun Yang Improved YOLOv2 Object Detection Model 2018 6th International Conference onMultimedia Computing and Systems (ICMCS)

[5] YOLOv3 explained https://medium.com/analytics- vidhya/yolo-v3- theory-explained-33100f6d193

[6] Anitha.J, Subalaxmi.A, Vijayalakshmi.G RealTime Object Detection for Visually Challenged Persons International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-8 June, 2019

[7] A. Vidyavani , K. Dheeraj, M. Rama Mohan Reddy, KH. Naveen Kumar Object Detection MethodBased on YOLOv3 using Deep Learning Networks ISSN: 2278-3075, Volume-9 Issue-1, November 2019.

[8] Zhong-Qiu Zhao, Member, IEEE, Peng Zheng, Shou-tao Xu, and Xindong Wu, Fellow, IEEE-Object Detection with Deep Learning: A Review 2019.

[9] Mingxing Tan Ruoming Pang Quoc V. Le Google Research, Brain Team- EfficientDet: Scalable and Efficient Object Detection 2017.

[10] Xingyi Zhou UT Austin zhouxy@cs.utexas.edu DequanWang UC Berkeley dqwang@cs.berkeley.edu
Philipp Krahen uhl UT Austin philkr@cs.utexas.edu-Objects as Points,2019.

[11] Object detection using convolutional neural network,IEEE explore,2018.

[12] Jason Brownlee-Introduction to Object recognition with deep learning,2017

[13] Manjula.S,Lakshmi Krishnamurthy-A Study on Object Detection,2016.

[14] Sunil,Gangadeep-Study of Object Detection Methods and Application on digital image 2019.

[15] Joseph Redmon,Santhosh Divvala,Ross Girshick,Ali Farhadi-You Only Look Once-unified,Real Time Object Detection,2015.

[16] Procedia Computer Science-A Review of YOLO algorithm and Developments,2022.
Upulie Handalage,Lakshini Kuganandamurthy-Real Time Object Detection Review,May 2021.