International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST

Object Detection and Distance Estimation Using Deep Learning

***[1]K Usha Rani,[2]Dosala Srinishma, [3]Ancha Vidisha**

[1]HOD &Associate Professor, Department of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, India

[2,3]Students, Department of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, India

ARTICLEINFO

Article History:

Accepted: 20 April 2023
Published: 08 May 2023

Publication Issue
Volume 10, Issue 3
May-June-2023

Page Number
69-73

ABSTRACT

Object detection is a computer vision technique for locating instances of objects in videos. When we as humans look at images or videos, we can recognize and locate objects within a matter of moments. The main goal of this project is to clone the intelligence of humans in doing that using Deep Neural Networks and IOT, Raspberry Pi and a camera. This model could be used for visually disabled people for improved navigation and crash free motion. When we consider real time scenarios, numerous objects come into a single frame. To identify different items simultaneously as they are captured, a strong model needs to be developed. YOLO (You Only Look Once) is a clever convolutional neural network (CNN) that helps in reaching that objective. The algorithm applies a single neural network to the full image, and then divides the image into regions and predicts bounding boxes and probabilities for each region. The bounding boxes are nothing but weighted by the predicted probabilities. The second objective of this model is to calculate distance of humans from the camera, to achieve that haar classifier is created and used. This classifier also helps in enhancing human detection along with distance calculation. Haar is just like a kernel in CNN where the kernel values are determined by training while in Haar they are determined manually. Whenever a person is detected by both YOLO and Haar classifier, a formula which considers height and width of human contours is applied to calculate the distance of it from the camera. As the objects are identified they will be read out using a text-to-speech engine known as gTTS(google text-to-speech) and ,which stores the text in an mp3 file. The package known as Pygame will load and play the mp3 file dynamically as the objects are detected. This developed Deep Learning model is integrated with Raspberry Pi using OpenCV. Though this project is primarily developed to aid visually disabled people, it can have various other applications such as, self-driving cars, video surveillance, pedestrian detection, face detection.

Keywords : Deep learning, IOT, YOLO

I. INTRODUCTION

With the world moving towards automation, there's a rising demand for an efficient Obstacle (Object) detector [1], [2], [7]. There are no low cost object detectors available which are efficient and most of the available obstacle detectors use ultrasonic sensor which doesn't take a broader frame into its line of sight. This project focuses on integrating Deep learning algorithms with Raspberry Pi and Camera to detect the type of object and its distance efficiently. The human visual system is fast and accurate and can perform complex tasks like identifying multiple objects and detect obstacles [9]. With the availability of huge amounts of data, faster GPUs, and better algorithms, computers can now be trained to detect and classify multiple objects within an image with high accuracy [10].In this project we will explore Object detection and localization using convolutional networks and integrate it with Raspberry Piand camera. A raspberry pi board with camera and speaker will be used to detect the obstacle and alert the user by reading out the distance and type of obstacle/object in the path of motion. Machine Learning algorithm will be used to train the system. The underlying algorithm is CNN(Convolutional Neural Network) . The model will be trained to identify the type of object and calculate the approximate, if not exact, distance of the humans from the camera. Once the object is detected, the type of object and the distance only for humans will be read out using a speaker.

II. RELATED WORK

A. Real-Time Objects Recognition Approach for Assisting Blind People. In computer vision, such as navigation and path finding, blind assistance is posing a significant challenge. To give the essential information about the environment, this research employs two cameras installed on a blind person's glasses, a GPS-free service, and an ultrasonic sensor. A dataset of objects acquired from everyday scenarios is built in order to apply the requisite recognition. Faces, bicycles, seats, doors, and tables are ubiquitous in blind situations, and object identification is a technique for distinguishing objects in the real world from a digital image. The disparity map is created using two cameras, the GPS service is used to group objects depending on their locations, and the sensor is utilised to identify any barrier at a medium to long range. The descriptor of the Speeded-Up Robust Features approach is optimised to conduct the recognition. The proposed technique for the blind intends to provide greater possibilities for those with vision loss to attain their full potential. The proposed work operates well in a real-time environment, according to the experimental results. for the blind is to provide greater possibilities for persons who have lost their vision to attain their full potential. The outcomes of the experiments show how well the suggested work functions in real-world scenarios. [1]
B. Object Detection Combining Recognition andSegmentation. Liming Wang1, Jianbo Shi2, Gang Song2, and I-fan Shen; Liming Wang1, Jianbo Shi2, Gang Song2, et al created a method for detection of objects that combines topdown recognition and also bottom-up image segmentation. The two essential steps in this strategy are generating hypotheses and verifying them. In the top-down hypothesis generation process, they increase the Shape Context feature, making it more resistant to object deformation and backdrop clutter. With the improved Shape Context, a collection of item placement hypotheses and figure ground masks with a high recall and low precision rate are generated. In the

verification step, they also compute a set of viable segmentations that are consistent with top-down object assumptions, and suggest a False Positive. [2]

C. Human objects detection, behavior recognition SeyedYahyaNikouei et al created Human object detection, behavior recognition, and prediction in smart surveillance fall into this category, where a substantial volume of video streaming data can take a long time to transition and impose a lot of strain on communication networks Video processing and object detection are widely accepted as computationally difficult and too costly for resource-constrained edge devices to handle. A lightweight Convolutional Neural Network (L-CNN) inspired by depth wise separable convolution and the Single Shot Multi-Box Detector is introduced in this paper (SSD). By restricting the classifier's searching space to focus on human objects in surveillance video frames, the proposed L-CNN technique detects pedestrians with a low compute workload on an edge device. [3]

III. PROPOSED SYSTEM

As shown in Fig 1 the phases of the proposed system design are described as follows :

Phase 1: The raspberry pi camera captures the input video live to detect objects in real time.

Phase 2: The video captured is broken down into frames and the model trained extracts specific features from the frames.

Phase 3: The boundary boxes are drawn and object proposals are generated [1].

Phase 4: Object proposals from phase 3 and the surrounding proposals from the vicinity are grouped together. Each object proposal grouped together detects a particular part of the object. These object proposals are refined to get the actual object that needs to be detected.

Phase 5: The refined objects from phase 4 are classified according to the trained categories. The number of categories that the model can identify us 80

Phase 6: Based on the contours of the humans detected, the distance between the detector and human is calculated [4]. A Haar classifier is used to identify the contours.

Phase 7: The type of object along with the distance for humans is read out using the speaker attached to the Raspberry Pi

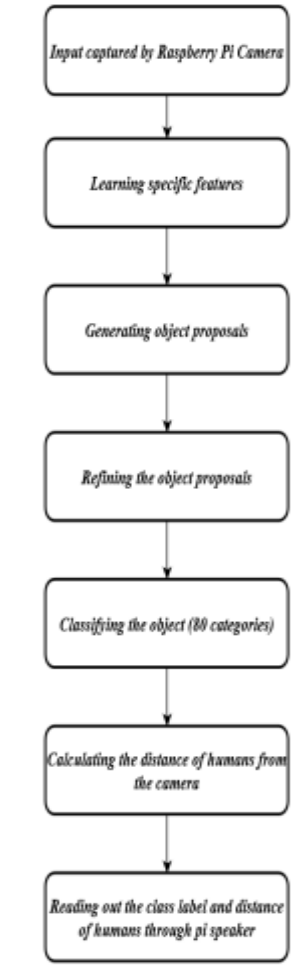

Fig 1 : Flow Diagram

The underlying neural network for YOLO is a Convolutional Neural Network which predicts multiple bounding boxes and class probabilities for those boxes simultaneously. YOLO trains on full images and directly optimizes detection performance [5], [6]. It has many benefits over traditional method for object detection, three of which are mentioned below : YOLO is extremely fast. To predict detections at test time, the neural network is run on a new image.

The network runs at 45 frames per second with no batch processing on a Titan X GPU. This means the streaming video in real-time can be processed with less than 25 milliseconds of latency.

ii. In contrast to the region based proposals and sliding window method, YOLO considers the entire image during training and test time where the information about classes as well as their appearance is encoded. In Fast R-CNN the larger context is not considered and the smaller objects in the background are identified as images. Compared to Fast R-CNN, YOLO makes very few background errors.

iii. Generalizable representations of objects are easily learnt by YOLO which makes it less likely to breakdown when applied on unexpected and new domain inputs. The network uses features from the entire image to predict each bounding box. For all classes for an image, bounding boxes are simultaneously predicted. The YOLO design enables end-to-end training and real-time speeds while maintaining high average precision [8].

The following points explain the process of object detection and how bounding boxes are drawn . As shown in Fig 3, An image is split into SxS grid, within each of the grid m bounding boxes are taken.

2. For each of the bounding box generated, a class probability and confidence scores values for the bounding box are given as output.

3. These confidence scores show how confident the model is that the box contains an object .

4. Each of the bounding box generated consists of 5 predictions: x, y, w, h, and confidence. The (x, y) coordinates represent the center of the box relative to the boundaries of the grid cell. Relative to the whole image, the width and height are predicted.

5. Only those bounding boxes that have the class probability greater than or equal to the set threshold value are selected. These bounding boxes are used to locate the object in the image/frame.

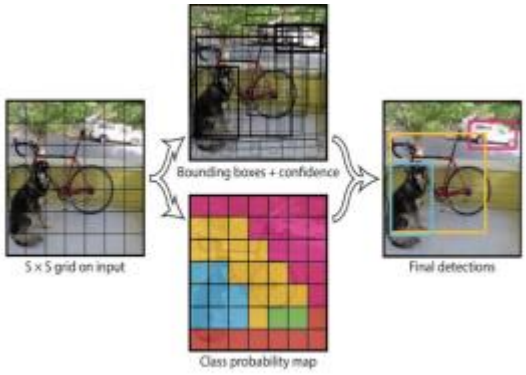

Fig 2. YOLO Working

IV. CONCLUSION

The project conducted more rigorous detection under the use of YOLO to provide suitable outcomes and help the project culminate into being assistance to impacted users, out of all the detection algorithms tested. The tensorFlowLite model ran seamlessly and successfully on the mobile device, offering a cost-effective and reliable medium for harnessing and instilling the benefits of machine learning.

V. REFERENCES

[1]. J. Redmon and A. Angelova, "Real-time grasp detection using convolutional neural networks," 2015 IEEE International Conference on Robotics and Automation (ICRA), 2015.

[2]. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[3]. R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), 2015.

[4]. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, 2017.

[5]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[6]. H. Caesar, J. Uijlings, and V. Ferrari, "COCO-Stuff: Thing and Stuff Classes in Context," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.

[7]. S. Gupta, R. Girshick, P. Arbeláez, and J. Malik, "Learning Rich Features from RGB-D Images for Object Detection and Segmentation," Computer Vision – ECCV 2014 Lecture Notes in Computer Science, pp. 345–360, 2014.

[8]. A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[9]. J. Xiao, K. Ramdath, M. Iosilevish, D. Sigh, and A. Takacs, "A low cost outdoor assistive navigation system for blind people," 2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA), 2013.

[10]."TensorFlow Lite | TensorFlow," TensorFlow. [Online]. Available: https://www.tensorflow.org/lite. [Accessed: 24-Mar-2019].

[11]."An introduction to Text-To-Speech in Android," Android Developers Blog, 23-Sep-2009. [Online]. Available: https://android developers.googleblog.com/2009/09/introduction -totext to-speech-in.html. [Accessed: 24-Mar-2019]

Cite this article as :

K Usha Rani, Dosala Srinishma, Ancha Vidisha, "Object Detection and Distance Estimation Using Deep Learning", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 3, pp. 69-73, May-June 2023.

Journal URL : https://ijsrst.com/IJSRST523102113