DESIGN AND SIMULATION LAB: PROJECT REPORT

OBJECT DETECTION USING YOLO AND NAVIGATION GENERATOR

ARKAJIT PAL – ROLL NO. 22SP06008

M.TECH (1st YEAR)

SIGNAL PROCESSING AND COMMUNICATION

SCHOOL OF ELECTRICAL SCIENCES

INDIAN INSTITUTE OF TECHNOLOGY, BHUBANESWAR

ABSTRACT

The Objective is to detection of objects using You Only Look Once (YOLO) approach. This method has several advantages as compared to other object detection algorithms. In other algorithms like Convolutional Neural Network, Fast Convolutional Neural Network the algorithm will not look at the image completely but in YOLO the algorithm looks the image completely by predicting the bounding boxes using convolutional network and the class probabilities for these boxes and detects the image faster as compared to other algorithms.

Along with the detection of objects, I have implemented the navigation controller.

THEORY

Working of YOLO –

Prior detection systems repurpose classifiers or localizers to perform detection. They apply the model to an image at multiple locations and scales. High scoring regions of the image are considered detections.

YOLO uses a totally different approach. It applies a single neural network to the full image. This network divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities.

YOLO model has several advantages over classifier-based systems. It looks at the whole image at test time so its predictions are informed by global context in the image. It also makes predictions with a single network evaluation unlike systems like R-CNN which require thousands for a single image. This makes it extremely fast, more than 1000x faster than R-CNN and 100x faster than Fast R-CNN.

METHODOLOGY

For this project a pre-trained YOLOv3 model which can detect 80 classes of objects. The first step is to load the pre-trained model for object detection on individual frames of our video stream.

Next, we establish a live video stream from our webcam feed using python OpenCV library. This stream is processed frame by frame by the YOLO detection system. As mentioned earlier, for each frame of the live video stream YOLO yields some detections in the form of bounding boxes. In order to find the most dominant detection we sort the detections based on the area of each of their bounding boxes. The bounding box with the largest area is chosen to be the most dominant detection and the decision of moving to the left or right is given based on the position of this bounding box with respect to the centre of the frame.

Whenever we make a decision to turn left or right a flag external to the main loop is set which dictates which voice command will be played. The voice command playback is controlled by a separate thread so that it does not interfere with or slow down the detection and decision-making process.

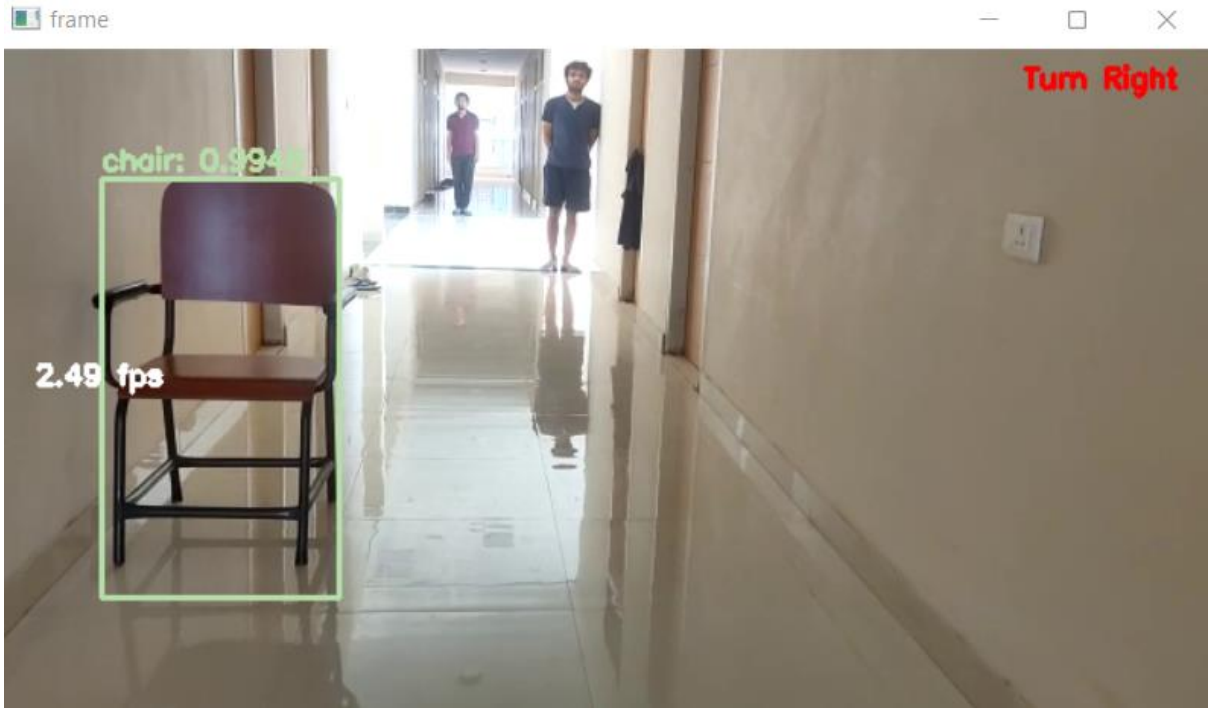
Apart from this we measure the time it requires to process a single frame and provide a rough estimate of the frames per second count on the output panel.

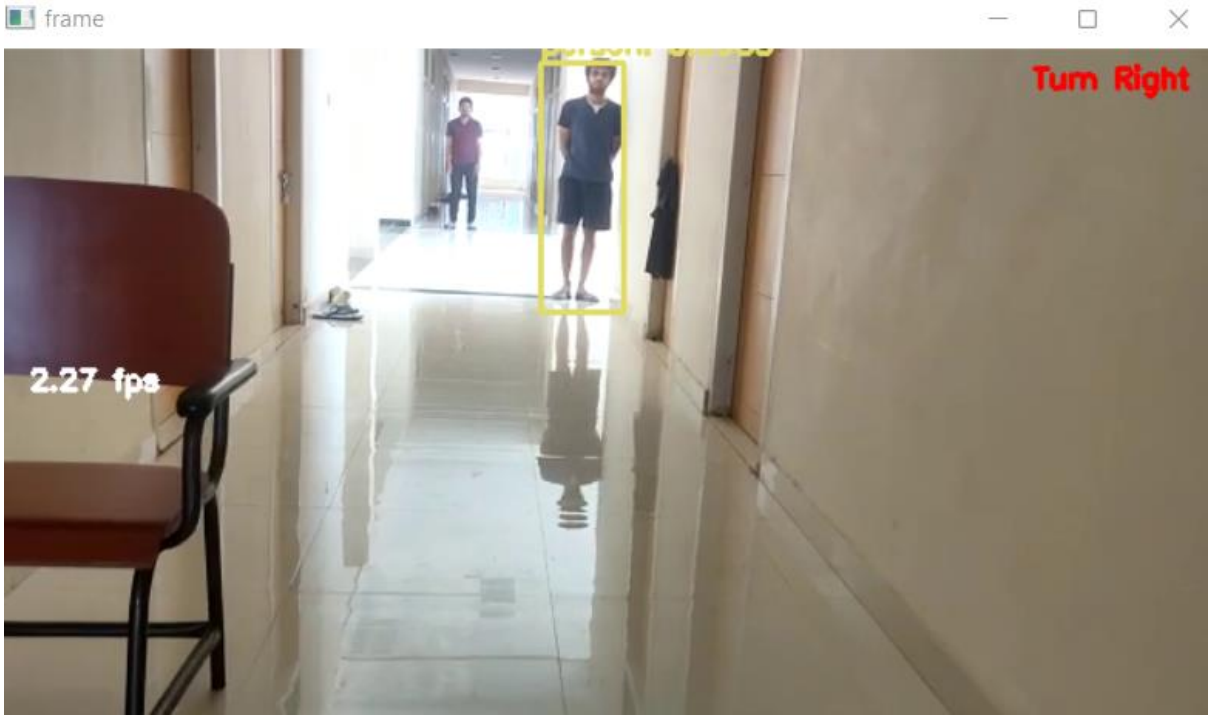
To exit the detection and decision loop we assign the ‘Esc’ key to stop the live video stream.

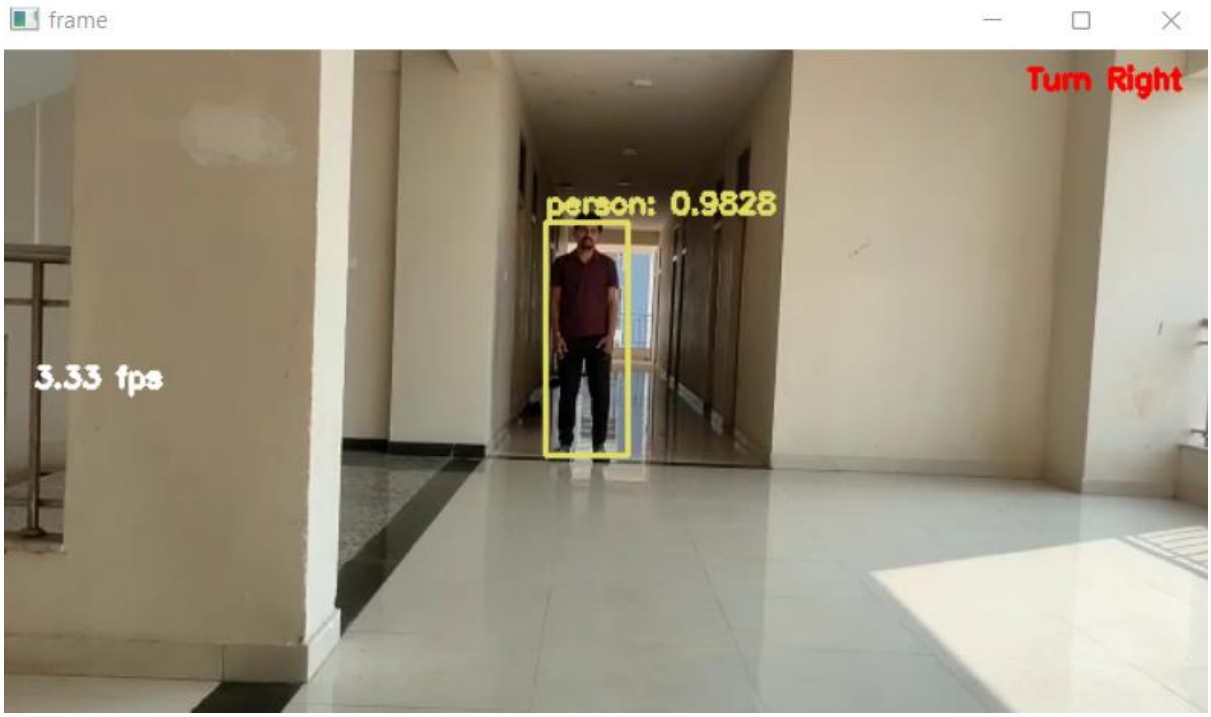
All of this is also tested on pre-recorded videos having different number and positions of obstacles.

RESULTS

Attached below are some images of the output panel of the detection.







The notebook attached can be run to clearly see the output on live video stream and also on the attached test videos.

From the above images we can see that the model is easily able to identify the biggest obstacle in the path and is suggesting a viable direction to move.

DISCUSSIONS

This approach to obstacle navigation, though accurate and simple, is not very effective in obstacle navigation as it is able to make only very crude identifications and suggest very general movement patterns. Although this may be refined further for better pathing and performance if it is ever used in real world robotics.

CONCLUSION

This detection and navigation system will prove useful for solving simple pathing problems. The voice prompts will act as an added accessibility feature. Although the system is accurate in terms of obstacle detection and providing decisions but its performance can still be further improved to improve the frames processed per second.

REFERENCES

[1] YOLO: Real-Time Object Detection. https://pjreddie.com/darknet/yolo/