Evolution in Electrical and Electronic Engineering Vol. 4 No. 1 (2023) 329-336

EEEE

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/eeee
e-ISSN : 2756-8458

Object Detection Using YOLO for Quadruped Robot Manipulation

Nor Aiman Alimie Norizan[1], **Mohd Razali Md Tomari**[1], **Wan Nurshazwani Wan Zakaria**[2]*

[1]Faculty of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia, Batu Pahat, 86400, MALAYSIA

[2]School of Mechanical Engineering, College of Engineering
Universiti Teknologi MARA, Shah Alam, 40450, MALAYSIA

*Corresponding Author Designation

DOI: https://doi.org/10.30880/eeee.2023.04.01.039
Received 15 January 2023; Accepted 07 April 2023; Available online 30 April 2023

Abstract: This paper focuses on the robot arm's vision system for integration with the quadruped robot. The quadruped robot has high indoor and outdoor maneuverability but lacks the ability to manipulate tasks. The additional robot arm with a good vision system needs to be integrated into the quadruped robot that allows manipulation capabilities. Hence, the objective of this study is to develop and apply custom-trained datasets on different YOLO algorithm architectures for the robot arm's vision. Prior to the vision system development, a suitable robot arm for the application is identified in which the robot motion and control are developed. The custom datasets preprocessing and training are done is trained using Roboflow and Google Colab respectively. Two versions of YOLO have been developed for the object detection algorithm which are Yolov3-tiny and Yolov5s. The comparison study in terms of speed, confidence level, and probability of detection is conducted to evaluate both YOLO version. It was found that the YOLOv5s provides an overall better performance compared to the YOLOv3-Tiny. All the aspects (speed, confidence level, probability) are essential to ensure the robot arm can identify, move and grasp the object efficiently.

Keywords: You Only Look Once(YOLO), Detection Algorithm, Custom Dataset, Google Colab, Robot Operating System(ROS)

1. Introduction

The quadruped robot can move efficiently indoor and outdoor environments but cannot manipulate objects. The quadruped robot that allows manipulation skills must have an extended robot arm with a powerful vision system. A robotic arm is quick, dependable, and precise and can be programmed to perform an endless number of jobs in various conditions. The articulated robotic arm comprises links

*Corresponding author: wshazwani@uitm.edu.my

connecting a series of rotating joints to an end effector [1]. The COVID-19 pandemic is forcing industries to use remote working tech [2].

For example, Shell Corporation applied Boston Dynamics' Spot Robot [2], and Petroliam Nasional Berhad deployed ANYbotic [3] for inspection purposes. The aspects that must be highlighted for task manipulation are the object detection and identification processes. The arm must be able to identify the thing it is intended to manipulate and be aware of its surroundings. Hence, the vision system is crucial to ensuring the smoothness of the developed motion planning in the future.

The YOLO detection algorithm is chosen as the detection algorithm in this study as the algorithm provides fast detection speed and is suitable for target detection in a real-time environment [4]. The fast detection is due to one stage-structure that allows the prediction of bounding boxes and class probabilities simultaneously [5]. The custom datasets trained for further implementation in the YOLO algorithm which is necessary for the development of YOLO algorithm. Yolov3-Tiny and Yolov5s versions of the YOLO algorithms were chose to test their capability in object identification and recognition in terms of processing speed, processing distance, and light intensity.

2. Methodology

This section describes the proposed method throughout the project, starting from the overview of the proposed robotics system, YOLO, for object recognition and the experimental program to test the efficiency of the developed detection algorithm.

2.1 Overview of Proposed Robotic System

The Adroit HDT 6 DOF has been chosen to integrate with the robot system based on the Design Requirements as tabulated in Table 1. It has 6 degrees of freedom, a 600 mm configurable length, and it is possible to lift a 5.4 kg payload at maximum reach. The end effector is replaceable with 2 types of actuators (A48 and A24) provided by HDT Robotics. In addition, the arm's joint produces maximum 48Nm torque and has sensing feedback that allows stable torque, angle, and angular velocity.

Table 1: Robot Arm Design Requirement

Robot Arm Design Requirements	
Configurable Length	1m
Weight*	Less than 5kg
Payload*	Up to 5kg
DOF	6 DOF
End Effector	Gripper with torque sensing
Camera	Will be mounted on one of the arm links
Environmental Requirements	IP66/IP67

The illustration of the proposed system can be seen in Figure 1. The Adroit HDT 6 DOF robotic arm was required to be mounted on the quadruped robot, and the Logitech C930 HD PRO was attached to the arm's end effector. It is medium size camera with 43.4 mm height, 94 mm width, and 71 mm depth. It has a 1080p/30 frame rate and a 720p/60 frame rate. The camera has a 78-degree diagonal field of view and 3 megapixels (dFoV). The fact that it is compatible with the ROS platform is essential because the YOLO algorithms were running on it.

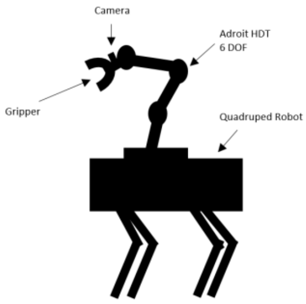

Figure 1: The Illustration of Robot Arm on Quadruped

2.2 YOLO for Object Recognition

There are many thousands of image datasets available to choose. There are several famous pre-trained datasets, such as ImageNet, COCO, PASCAL VOC, BDD100K, and DOTA v2.0. Nevertheless, only some objects were trained in that dataset. Hence, custom datasets are required to detect specific objects. The process of training new custom dataset is presented in Figure 2. The gathered dataset images is about 100 images of each class. The classes are divided into spanner, screw, air regulator filter, door handle, and switch. Each of the images has been annotated using the Roboflow Annotation platform. Image annotation is done to attach label to an image frame to recognize, count, or track object boundaries in the images.

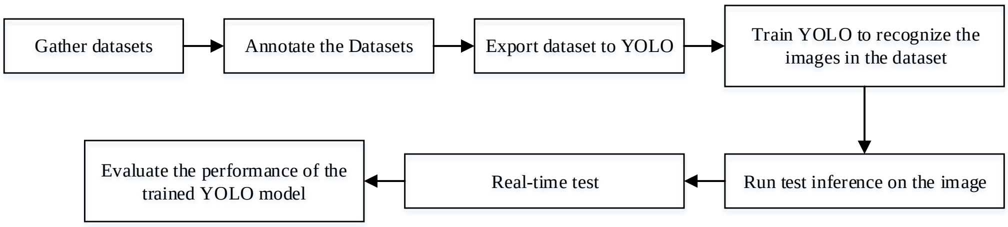

Figure 2: Custom Train Dataset Process

The images have been split into 3 sets: the training set, the validation set, and the testing set, as shown in Figure 3. The training set was used to train and make the model learn the hidden features in the images. The validation set was used to give information to tune the model's hyper-parameters and configurations, while the test set used was to test the model after completing the training. The dataset then goes to the pre-processing and augmentation stage before starting to train and finally conversion to the YOLO format. Yolov5 Pytorch and YOLO v3 Keras are used for Yolov5 and Yolov3-Tiny algorithm respectively.

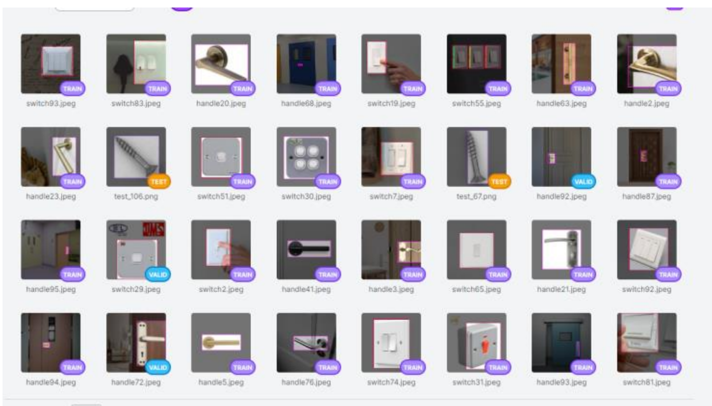

Figure 3: Split of the Dataset

The training process was done using Google Colab. It offers free access to GPUs, which are necessary for swiftly and effectively training a model. Google Colab is one solution to that issue because not all PCs and notebooks have GPUs to carry out the training process. The training process was done using pre-trained convolutional weights: darknetnet53.conv.74 for Yolov3-Tiny [6] and Yolov5s.pt for Yolov5s [1]. After the training, the developed detection algorithm was tested on the images to the accuracy. Then, it was tested on the camera for real-time detection. The performance of the trained YOLO model was analyzed at the end of the test.

2.3 Object Detection Test Design

The developed object detection algorithm for the Adroit HDT 6 DOF’s camera is tested for further analysis. The test focus on the confidence level of the detection, processing time, and optimum condition. The detections are as follows:

A. Detection in Optimum Condition

The test will concentrate on the impact of light intensity variations on the detection confidence level. The item will be set up under two conditions: daytime and night as illustrated in Figure 4. The test is done outdoors with the plane background. There is a 0.5-meter gap between the camera and the objects. In this test, only one test subject has been measured. The object is a plastic bottle.

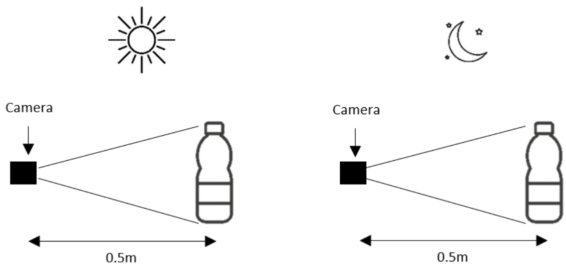

Figure 4: Test in Different Conditions

B. Detect Indoor and Outdoor

In addition, the object detection test has also been done in two separate environments: the indoor environment and the outdoor environment, as shown in Figure 5. The test will analyze the effectiveness of the detection algorithm in different environments. The bottle will be the test subject with a fixed 0.5-meter gap between the camera and the bottle.

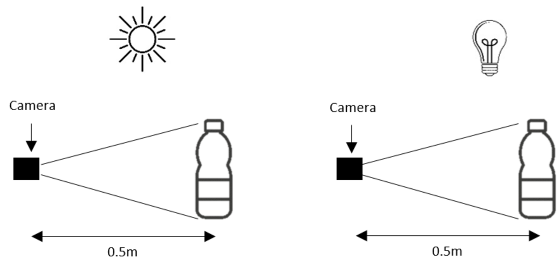

Figure 5: Test in Different Environments

C. Probability Test

The third designed test is to analyze the detection algorithm's efficiency in sensing the test subject's presence. The probability test is done on all listed test subjects (door handle, air filter regulator, spanner, screw, and switch). The test subject will be in the camera's field of sight while it moves in a random direction (left to right and back to the normal position of the camera), as in [7]. The probability of the camera identifying the test subject is recorded.

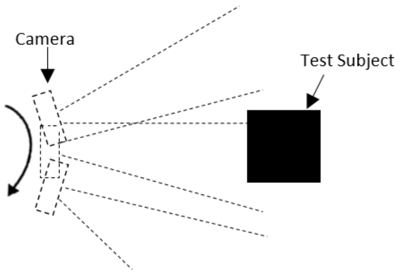

Figure 6: The Top View of the Camera Movement to Detect the Object

D. Processing Time Test

In task manipulation, the execution time for the camera to recognize the objects is essential. To make a smooth arm movement, image processing time is required. The velocity of the arm needs to meet the image processing. Each test participant will have their processing time data recorded in five sets.

E. Difference Distance Test

The last test is manipulating the camera's distance from the test subjects. The range distance will be from 0.2 to 2 meters, as shown in Figure 7. The confidence level of the test subjects is recorded. This test is done for all listed test subjects.

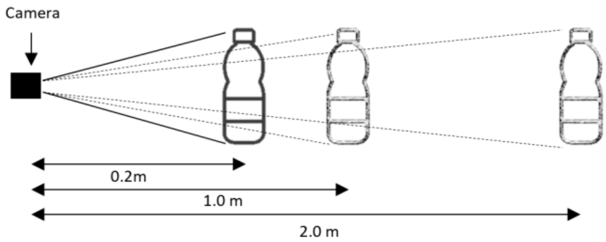

Figure 7: The Manipulated Distance of the Camera with the Test Subject

3. Results and Discussion

This section discussed the simulation results of for the developed system that running on ROS. Different YOLO algorithm architectures performed differently when the same custom-trained dataset was implemented. The comparison of the confidence level and processing time with the lighting, and distance as a manipulated variable was measured. Figure 8 shows the attachment of the camera on the Adroit HDT 6 DOF with the A24 pincer end effector.

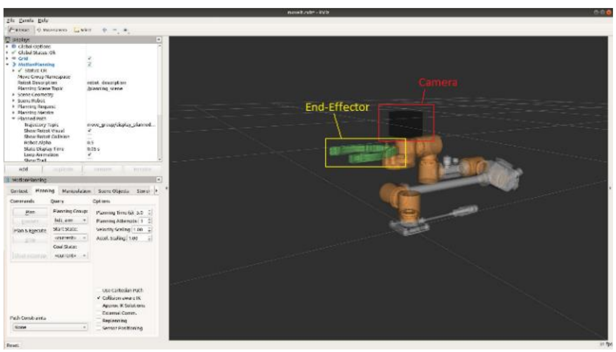

Figure 8: The Simulation of Adroit in Rviz

3.1 Object Detection Test Results

The inference test was done on Google Colab to test the custom-trained weights in the static images. The result of the test can be seen in Figure 9 for Yolov3-Tiny and Figure 10 for Yolov5s. Both algorithms generated high confidence levels for all test subjects.

Figure 9: The Result of Run Inference with Yolov3-Tiny Trained Weights

Figure 10: The Result of Run Inference with Yolov5s Trained Weights

Next, for the real-time detection, the result can be seen in Table 2 while the images along the test can be seen in Figure 11 and Figure 12.

Table 2: Real-time Detection Result

Algorithm	Average Confidence Level,%	Average Probability, %	Total Average Processing Time, s
Yolov3-Tiny	49.0	40.0	1.778
Yolov5s	72.5	72.0	0.504

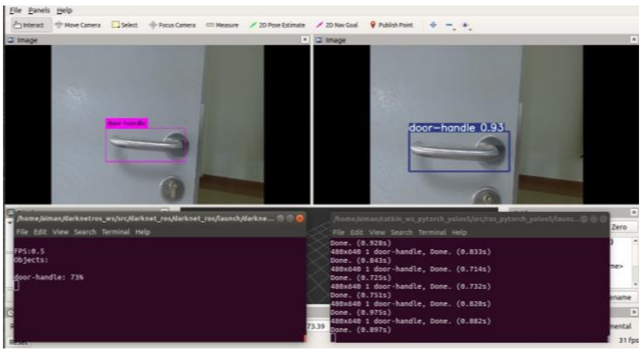

Figure 11: The detection of the Door Handle in Both Algorithms

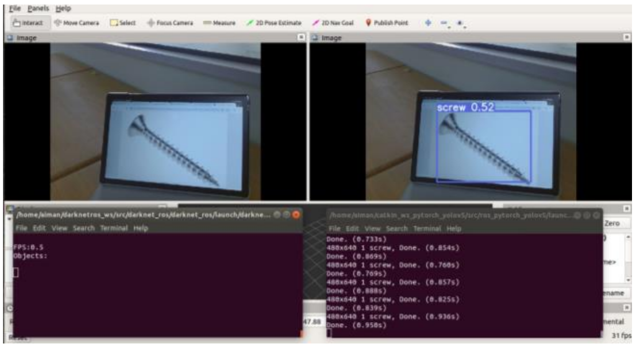

Figure 12: The Detection Test Done on Air Regulator Filter

3.2 Discussion

Based on the test results the Yolov5s performed better compared to the Yolov3-Tiny with average confidence level of 72.7% and 49.0% respectively. In addition, the Yolov5s received a 32% higher average probability score (72%) than the Yolov3-Tiny at 40%. The Yolov5s was able to execute the data three times faster than the Yolov3-Tiny. The Yolov5s processes data on average in 0.504 seconds, whereas the Yolov3-Tiny processes information on average in 1.778 seconds.

For the optimum condition test, the Yolov5s was able to identify the test subjects better in both conditions (daytime and night) compared to the Yolov3-Tiny lack of confidence level when in the night condition. However, for all trials, neither algorithm was able to identify the existence of a screw. It may be due to the small size of the screw, and the camera resolution was not high enough to capture the frame. The reason for the statement can be proved when the detection is done on the tablet with the display of the screw image on it. The Yolov5s was able to detect the presence of the screw, as shown in Figure 12 (right side of the figure).

4. Conclusion

In conclusion, the Adroit HDT 6 DOF has been chosen to manipulate object for the quadruped robot. Motion Planning algorithm was successfully established but further work needs to be done. Due to time constraints, the system integration will be conducted at a later stage. The object identification algorithm was successfully developed for YOLO. The custom-trained dataset using Google Colab offers unlimited opportunities to all researchers and students to train their dataset, even if their laptop does not have the GPU. Based on the real-time test result, the Yolov5s has been chosen as the detection algorithm for the Adroit HDT 6DOF robot arm. The Yolov5s has better performance in every aspect of the test is done. It has a faster processing time, better confidence level, more fitting bounding box, and sharper image captured. A better detection can be achieved if greater laptop performance has been used in this project. For the continuity of the project, a few aspects need to improve. The additional custom dataset can be trained to suit the oil and gas platform environment. The better laptop performance can be used to get better image processing, and the latest Yolo version can be used.

Acknowledgement

The authors would also like to thank the Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, for its support and for providing related facilities.

References

[1] M. Farman, M. Al-Shaibah, Z. Aoraiath, and F. Jarrar, “Design of a three degrees of Freedom Robotic Arm,” *International Journal of Computer Applications*, vol. 179, no. 37, pp. 12–17, 2018.

[2] Berry Mulder, Ross Doak, and Adam Serblowski, THE ROAD TRAVELLED SO FAR TO ground surveillance robots , 2021.

[3] A. S. S. Ahmad Jamil, “Meet anymal!,” PETRONAS Global. [Online]. Available: https://www.petronas.com/flow/technology/meet-anymal. [Accessed: 08-Apr-2022].

[4] W. Yang and Z. Jiachun, "Real-time face detection based on YOLO," 2018 1st IEEE International Conference on Knowledge Innovation and Invention (ICKII), Jeju, Korea (South), 2018, pp. 221-224, doi: 10.1109/ICKII.2018.8569109.

[5] J. Fan, J. Lee, I. Jung and Y. Lee, "Improvement of Object Detection Based on Faster R-CNN and YOLO," 2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), Jeju, Korea (South), 2021, pp. 1-4, doi: 10.1109/ITC-CSCC52171.2021.9501480.

[6] J. Redmon, *Yolo: Real-time object detection*. [Online]. Available: https://pjreddie.com/darknet/yolo/. [Accessed: 2022].

[7] Jocher, G. (2020), Yolov5 by Ultralytics version(7.0)[Source Code] https://github.com/ultra lytics/yolov5