King's College London

Deep Learning Based Object Detection for Robotic Arm

Submitted by:
Haoyi Zhang

Project Proposal Submitted in Partial Fulfillment of the Requirements of the MSc Project Module

Engineering Department

Supervised by:
Dr. Shan Luo

London
February/2024

Abstract

High accuracy of object detection is always crucial for robotic arms, as they directly affect the subsequent gripping accuracy and robustness of robots, which could bring more economic income for users. However, it is usually hard to detect an object for a robotic arm when the target is small or being occluded. In this paper, we plan to realize the recognition of targets for a robotic arm by using the deep learning method YOLO v8 model and further improve the recognition accuracy of targets under occlusion by refining the network. In the meantime, by comparing the detection results with those under other deep learning models, the recognition accuracy of the model proposed by this paper will be quantitatively verified. The feasibility and effectiveness of the model will be tested on both a simulation environment and an actual robotic arm. And through the experimental results will confirm that the method proposed in this paper can truly improve the accuracy of recognition in the robotic arm environment and enhance the detection ability in the case of occlusion.

Table of Contents

List of Tables

List of Figures

Page

List of Symbols

Parameters

Q	Matrices of query
K	Matrices of key
V	Matrices of value vectors

Notation

TP	True Positives
FP	False Positives
TN	True Negatives
FN	False Negatives

Chapter 1

Introduction

1.1 Background

With the advancement of modernization, more and more robots are being used in our society. They are widely used in factories, restaurants, hotels, and there will be more of them show up in normal people’s houses in the future. And make robots can interact with human better is one the most important part in robotic research. Robotic arm is a kind of common robot can be seen in the market, which is often being used in factories to do assembly or just as a part of humanoid robot. This kind of robot can get information in the environment through their vison sensor, and realize the function of object detection by using computer vision or deep learning and then grasp the target.

For robotic arms, improving accuracy, robustness, and detect occluded targets with high success rate have always been the difficulties to achieve. However, the development of deep leaning makes solving these problems possible. This paper aims to achieve the function of object detection for robot arms using deep learning method and realize a relatively high accuracy during object being occluded

1.2 Problem Statement

During the daily life, it is simple for us to recognize and pick up a spoon which is occluded by a cup, however it is hard for a detection algorithm. Most detection algorithms these days are based on deep learning, which takes thousands of pictures containing the target to be detected as an input data set for a neural network to train and extract features, then outputs a detector for detection. But for occluded objects, some important information or features cannot be extracted due to occlusion thus leading to difficulties in recognition. Many scholars have mentioned these problems in their papers and also proposed methods to try to solve them. For example, Tsung-Yi Lin, et al, Wei Liu, et al. proposed Feature Pyramid Networks [1], SSD [2] respectively to solve the problem of small targets, and K He proposed Mask R-CNN[3] not only for image segmentation but also for occluded targets recognition.

Robot arms used in factories or mailing companies for sorting items are often equipped with the function of object detection. Occluded objects can often appear in these areas due to the complexity of the environment, which brings challenge for the sorting task. The mis-recognition will cause the problem of wrong grip or failure to grip in a timely manner, which will bring economic losses for the user or even lead to serious factory shutdowns, courier mailing chaos and other issues resulting in huge losses. On the contrary, when the recognition accuracy increase, it will not only improve the accuracy of the grip but also speed up the work process then improve profits for users and enhance customer satisfaction.

Although many improvements have been proposed in the academics to enhance the approach in facing these difficulties, they cannot be applied to every situation due to the complexity of real-world environments and the different situations faced in different usage environments. At the same time, these algorithms require high computational resources in order to improve accuracy, which limits the usage environment. Therefore, it is necessary to optimize the recognition accuracy for occluded targets for the specific robotic arm usage environment.

1.3 Proposed Model

In this paper, we will firstly train a YOLO v8 model for object detection for a robotic arm usage environment. Then the detection results will be compared to other deep learning methods. Furthermore, we will optimize accuracy of recognition of occluded targets by improving the network structure. The refined model will be tested in both gazebo simulation environment and on a real robotic arm. The assumptions made in this paper are as follows:

1. YOLO v8 shows a better detection performance comparing to most major deep learning algorithms under the usage environment of robotic arms.
2. Through the refined network proposed by this paper, the recognition accuracy in the case of occlusion can be achieved better than before.
3. In the simulation and actual robotic arm test process, the method proposed in this paper can quickly and effectively identify the target.

1.4 Solution Approach, Expected Outcomes, and Contribution

Firstly, we will use the latest version of the YOLO series, YOLO v8 to see how well it fits on the robotic arm by comparing its accuracy with other deep learning algorithms. Then, we will try to refine the deep learning network to improve the problem that low accuracy of detection when target being occluded.

To assess the success of the proposed method, the refined detection algorithm will be tested in a Gazebo simulation environment and a real robot arm environment. Different cases of experiment will be conducted. For example, we can place the target behind another object and keep moving the target until it is fully displayed to check our detection algorithm under different occlusion situations, or cover different areas of the target to see at what percentage of the area being displayed the target can be successfully detected. The detection results will be quantitatively recorded for further optimization of the model.

Through the model proposed in this paper, robotic arm can grasp target better due to the high accuracy of target recognition, especially in the case of occlusion. This will bring a good improvement for the actual customer experience of the robotic arm. At the same time this paper has some potential disadvantages. The experimental environment of this paper is on a stationary tabletop where objects will be placed. Due to the limited experimental equipment, it is not possible to simulate a rolling conveyor belt to simulate a factory production environment. In the future, we will improve this aspect to simulate a more general usage environment.

This paper proposes a recognition algorithm suitable for robotic arms, which has high recognition accuracy, especially when the target is occluded. Applying the improved algorithm can improve the robotic arm's grabbing accuracy, improve user satisfaction, and further promote the development of industrialization and robots.

1.4 Outline

This project proposal is organized as follows. Chapter 2 reviews the literature on the robotic arms and recent advances in object detection algorithms. Chapter 3 defines the problem. Chapter 4 presents the solution and methodology. Chapter 5 presents the expected results and the required computational setup. Finally, Chapter 6 summarizes the project proposal.

Chapter 2
Literature Review

2.1 Introduction

Robotic arms are often used in factories, logistics companies, some restaurants and other environments in modern society, which are mostly used to detect and classify objects. As for robotic arms, it is very important to achieve high target recognition accuracy for it will directly affect its subsequent grasp accuracy and speed. The method combining computer vision and deep learning is currently widely used in the target recognition process of robotic arms these days, and many scholars have made certain improvements in different deep learning methods to improve its recognition capabilities. This section will present the research related to this topic.

2.2 Reviewed Literature

2.2.1 Robotic Arms

Current development of robotic arms is increasingly focused on intelligence, usually combined with computer vision and AI to achieve better interaction with the environment and users. ABB YuMi [4] is a collaborative robot developed by the Swiss company ABB. It has two independent robotic arms, each with 7 joints, vision and multiple sensors. It can realize the function of identifying obstacles and automatically stopping when encountering obstacles. It can also use a graphical interface for programming and most of them are used in food packaging and industrial assembly. Franka Emika Panda[5] Robotic Arm is a kind of robotic arm that is currently widely used. This robotic arm has seven joints and was developed by Franka Emika. It contains force, touch, temperature, inertia, and vision sensors and can be used in a variety of environments. In addition to being programmed by ROS, it is equipped with a graphical programming interface as well, allowing users to create programs more easily.

Although robotic arms are currently developing rapidly, they are also facing many problems and challenges in some aspects. For example, the accuracy of the robotic arm still needs to be improved, and whether high accuracy and repeatability can be guaranteed in different usage environments. Although currently more robotic arms are equipped with visual sensors, the working environment of robotic arms is complex and changeable. Ensuring recognition accuracy under dark light, complex background, obstruction by other objects, etc. is a problem that needs to be solved. Things like stronger human-computer interaction, lower computational cost, and safety are all problems that today's robotic arms developers have to deal with.

2.2.2 Development of Objection Detection

a) Traditional Object Detection

Traditional object detection methods use computer vision merely, which through image processing retains the target only to achieve the detection function. Among these methods, the key and commonly used technologies such as Canny detection, Harris corner detection, template matching, SIFT transformation, HOG, etc. Dong Hyeon Kim [6] and others realized the target recognition of a 6-degree-of-freedom classification manipulator by using the Canny algorithm combined with Gaussian filtering. They first extract the edges of the object and calculate the centroid point. The intrinsic parameters of the camera are obtained by calibration, and the corresponding position of the object center in the world frame is obtained through the transformation matrix. But what they did was recognize when the background was simple and the shape of the detection target was continuous (symmetrical structure). When the background or the shape of the detected object becomes complex, object segmentation will be difficult to implement with this method. Kunwei Song et al.［7］improved the TLD algorithm by decreasing the amounts of candidate samples and optimizing the real-time detection capability. Then they used the SIFT algorithm to achieve the function of detecting moving targets with 90% tracking accuracy. However, they gave a specific usage environment, and it would be difficult for this method to achieve a high accuracy in other environments. Generally speaking, the advantages of using traditional computer vision methods are low computing resource requirements, good real-time performance, and can be deployed in many places. However, once the environment changes or becomes complex, inaccurate detection may occur.

b) Deep Learning Based Object Detection Method:
With the continuous development of neural network structures and GPU, people have gradually begun to use deep learning for object detection. Compared with traditional computer vision methods, deep learning can identify deeper image features and has better generalization ability. It can better solve problems with changes in lightness and target occlusion and can handle more complex problems as well.

Guohao Yu et al. [8] explored the issues of low accuracy when detecting small objects with Faster R-CNN under real-time target detection. They modified the convolutional layers and introduced SE attention to increase the ability to extract features of small objects. At the same time, the pooling layers had been improved to reduce the impact of resizing the input image for input convolution. They test this method by detecting the front view of vehicles. The result shows that compared with the traditional Faster R-CNN method, the average accuracy was increased by 1.29%. However, this paper only compared the front view of vehicles, it did not use its improved model for other object detection to verify its generalization ability. Meanwhile, its detection effect on occluded objects was poor.

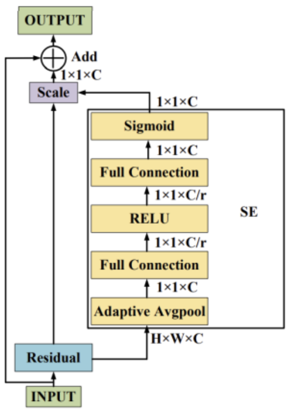


Fig.1 Structure of SE Attention[8]

Mask R-CNN is an improvement method also based on Faster R-CNN made by He Kaiming[3] and others. Based on Faster R-CNN, an image segmentation mask is added to the detected object, which can achieve clear segmentation of the detected object. What’s more, RoI Align is introduced to ensure the accuracy of segmentation. Based on

Mask R-CNN, the target detection accuracy and speed are also improved. A. N. Yumang et al. [9] achieved the recognition of 4 different beans through Mask R-CNN. The size of these beans is relatively small, and the differences between beans are also small, making it difficult to extract features. They use a Raspberry Pi with a camera as the hardware and they make the recognition results display on the Raspberry Pi’s screen. The accuracy of identifying different beans can reach 87.5%. However, the backgrounds of the detecting images they used are very simple and the beans may be more numerous and be obscured in some actual situations. The next step should be to try to conduct model training and testing under complex backgrounds and occlusions. At the same time, the article didn’t mention the relevant content about the consuming time, and further research can be conducted on this part.

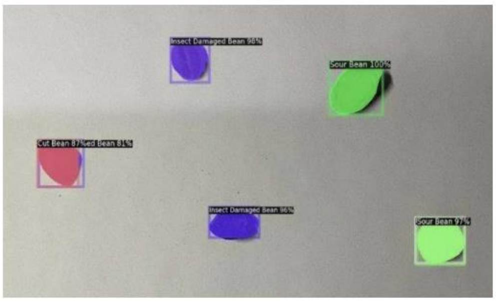


Fig.2 Output with Mask R-CNN[9]

In terms of algorithms based on R-CNN, region proposals are first generated, and then classification and regression are performed on the proposed regions. These kinds of methods are also called two-stage detectors. This type of algorithm has relatively high accuracy and can handle small targets and complex environments well, but the detection speeds are relatively slow. Other one-stage detectors treat the entire process as a regression or classification. The representative algorithm of this type is YOLO. Regarding the usage of YOLO in various fields. Xiaodong Yu[10] and others used different versions (v5l, v5x, v5s, etc.) of YOLO v5 to identify targets of an automatic sweeping robot in real time. The result shows that YOLO v5s shows better accuracy when the frame rate is low and the detection target is small. However, the confidence rate when the object with its shadow is not very high, the impact of shadows on recognition can be taken into consideration when training the model in future work. K. Patel et al. proposed a detection system for workers' safety helmets through YOLO v8[11], which can detect whether workers wear safety helmets at the construction site and achieve an accuracy of 99%, which is the highest score compared to other YOLO methods.

2.3 Research Gap

Tab.1 Research Gap

Reference	Operating Environment	Method/ Model	Backbone	accurancy	Comment
Kim et al.（2021）	Robotic Arm	Canny	——	Not Mentioned	Low computational resource requirements, good real-time performance, but lacking advanced feature recognition capabilities
Song et al.（2022）	Moving Object	SIFT	——	90%	Capable of recognizing the same features at different scales but lacks flexibility in handling complex images computationally.
Yu et al.（2022）	Car Front Detection	Faster R-CNN with SE Attention	Resnet50-FPN	AP:64.87	The modified model is not universally applicable
Yumang et al.（2022）	Fresh Beans Detection	Mask R-CNN	Resnet 50 integrated with Feature Pyramid Network	87.50%	Capable of segmenting smaller objects but has not been further studied under complex backgrounds and occlusion scenarios.
Yu et al.（2022）	Self-Driving Sweeping Bot	YOLO v5s	Not Mentioned	Not Mentioned	In cases where objects have shadows, confidence is not high. Further research can be conducted in this aspect.
Patel et al.（2023）	Helmet Detection	YOLO v8	Hybrid Backbone	Average precision: 0.99	Detailed comparisons were made between YOLO v8 and other YOLO series detectors. The results can serve as a reference for future applications of YOLO v8 in different domains.

With the rapid development of deep learning, target detection is becoming more and more accurate and faster. But generally speaking, the main difficulty of current recognition algorithms lies in the recognition of small objects and occluded objects in complex environments. YOLO v8, as the new generation of the YOLO algorithm in 2023, has not yet been used in the field of robotic arm recognition and it has better accuracy than other detection methods proved by some papers. Therefore, in this paper, YOLO v8 will be used and tested on a robotic arm usage environment to check its applicability and effectiveness. At the same time, the problem of low recognition accuracy under targets being occluded will be explored by improving its neural network.

Chapter 3
Problem Statement

3.1 Introduction

With the advancement of industrialization, robotic arms are widely used in various places. The recognition accuracy of the robotic arm affects the grasping accuracy of the robotic arm. Inaccurate grabbing will reduce the user's economic benefits and experience. This paper aims to improve the recognition accuracy of the robotic arm by using refined deep learning methods, thereby improving the performance of the robotic arm.

3.2 Problem Definition

Currently, deep learning is widely used in target detection and recognition. Due to the characteristics of the deep learning neural network, the detection accuracy on occluded objects is not high, and detection errors often occur. What is more, occluded objects can often appear in the usage of robotic arm environment, where mis-recognition could bring significate economic loss. In recent years, many scholars have conducted relevant research on these issues, but due to the complexity of the real environment, it is impossible to use one method to apply to all situations. Thus, it is important to solve the problem of low accuracy when detecting occluded targets for the usage of robotic arm environment. Therefore, there will be two problems this paper aiming to solve:

a) Test the object detection accuracy and suitability with the latest deep learning methods, YOLO v8, in the usage of the robot arm environment.
b) Refine the deep learning network to improve the accuracy during targets being occluded for robot arms.

3.3 Proposed Model

Firstly, at least 1k-5k images containing target object (especially those obscured object) will be collected and augmented as dataset to be input to deep learning models. The detection accuracy of different models will be compared and verified. Moreover, in order to solve the problem of targets being occluded, which often occurs in the robotic usage environment, the model will be refined by different methods like refining the attention mechanism or using prior knowledge. Finally, an improved model which can achieve high accuracy to obscured object will be obtained. Then the model proposed in this paper will be tested in both simulation and physics environments.

It is expected that the improved model will show better recognition accuracy in the robotic usage environment, which will lead to high performance in the subsequent grasp process for a robotic arm.

Chapter 4
Methodology

4.1 Introduction

Deep learning is a branch of machine learning. It learns the characteristics of the input information based on the neural network, which can extract the input picture information and realizes the target recognition through training. This article will use deep learning methods to realize target recognition for robotic arms. In order to verify the results, the trained model will be tested in the simulation environment Gazebo and a real robotic arm environment respectively. The details are as follows.

4.2 Objective Identification

4.2.1 Data Acquisition and Processing

Data acquisition will be achieved in two ways. The first way is to use existing widely used data sets such as COCO or Kaggle. The second way is to directly use manual annotation to mark the target items in the image. Through manual labeling, training images containing obscured targets, poor angles, and complex backgrounds will be added to increase the strength of the data set. At the same time, data enhancement algorithms will be used to further enhance the data by rotating, cropping, and adding noise to the images of the data set. The amount of data is expected to be at least 1k-5k, and will be divided into training sets and test sets respectively.

4.2.2 Deep learning model selection and training

In this paper, we will use the latest YOLO v8 of the YOLO series for object detection model training. The detection results will be compared with those of other deep learning models, such as Faster R-CNN, YOLO v3. YOLO v5, SSD and so on. Then the recognition accuracy, recall rate and other parameters of different models will be compared.

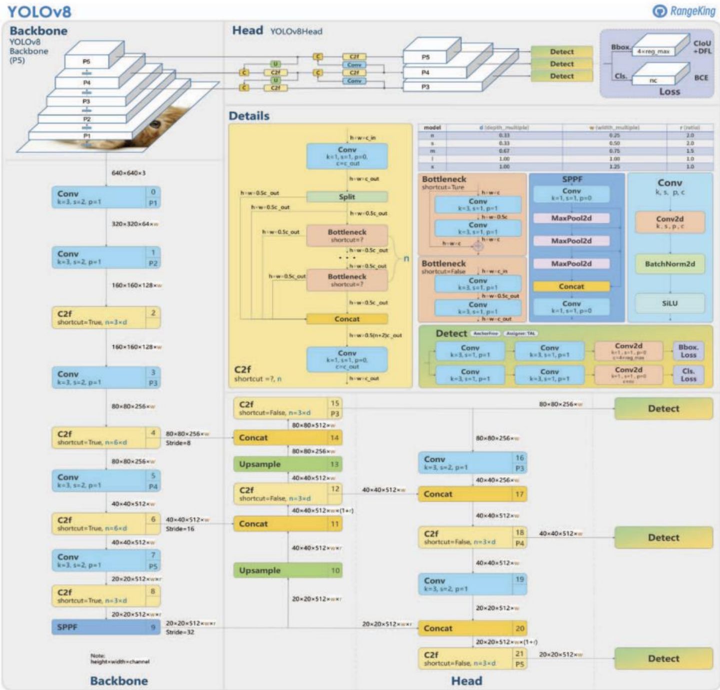


Fig.3 Architecture of YOLO v8[11]

4.3 Model Optimization for Occluded Targets

Since the robot arm is used in an environment where the grasping target could be blocked, resulting in inaccurate or unrecognizable target recognition, this article will modify the network to achieve the goal of detect the target under occlusion. And these methods could be used to refined the model:

a) Refining the attention mechanism: By refining the attention mechanism to make the detector focused on some key features of the target. Self-Attention is often used in deep learning enhance the interactions between elements in time sequence. The attention scores can be computed by:

$$Q = W^Q * x_i$$

$$K = W^K * x_i$$

$$V = W^V * x_i$$

$$Attention = soft\max(Q * K^T / \sqrt{d_K})V$$

Where, Q, K, V represent the matrices of query, key, and value vectors respectively, and d_K is the dimension of the key vectors.

And the output is obtained by:

$$out_i = sum(Attention_i * V_j),\ j = 1, 2, \ldots, n$$

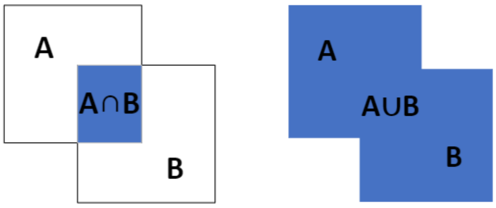
Where, out_i represent the i output in time sequence, and V_j represent the j element, $Attention_i$ is the attention of i.

b) Using prior knowledge: Inferring object by using 3D Model libraries established previously.

c) Refining Intersection over union: Refining the IoU threshold based on detecting objects accordingly.

$$IoU = \frac{A \cap B}{A \cup B}$$

Where:



d) Dataset Augmentation: Augmenting dataset by adding more images that contain occluded target.

4.4 Verification

The detection results will be verified on both Gazebo simulation and a real robot arm. Through the simulation, it is capable to find latent problems of code and to refined them before a real physic test. The algorithms that perform well in the simulation will be then used for testing in a real robotic arm, aiming to check what problems will occur on a real environment, and whether the algorithms with high performance. And the performance will be evaluated by accuracy, recall and precision.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\mathrm{recall} = \frac{TP}{TP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

Where TP, FP, TN, FN represent True Positives, False Positives, True Negatives, False Negatives respectively.

Chapter 5

Expected Results

5.1 Introduction

This paper will use deep learning methods to achieve relatively accurate object detection for the robotic arm and improve the issue of low detection accuracy of occluded targets, thereby providing precise information for subsequent robotic arm grabbing. The specific aim and objective are as follows:

Aim: Through the detection algorithm proposed by this paper, the robot arm can achieve high-precision object detection capabilities and also perform well on occluded objects.

Objectives:
1. Obtain a strong training data set.
2. Use YOLO v8 to achieve accurate object detection, and compare the results with other deep learning models.
3. Improve the object recognition accuracy under occlusion by modifying its network.
4. Achieve accurate object recognition in the testing environment

5.2 Obtaining Results

The data set in this article will be constructed through a combination of manual annotation and widely used data set like COCO to train the model. The training set will be input into different deep learning models for training and test the trained model to obtain the accuracy. Afterwards, gazebo simulation and actual robotic arm testing will be carried out. The specific test process is as follows:

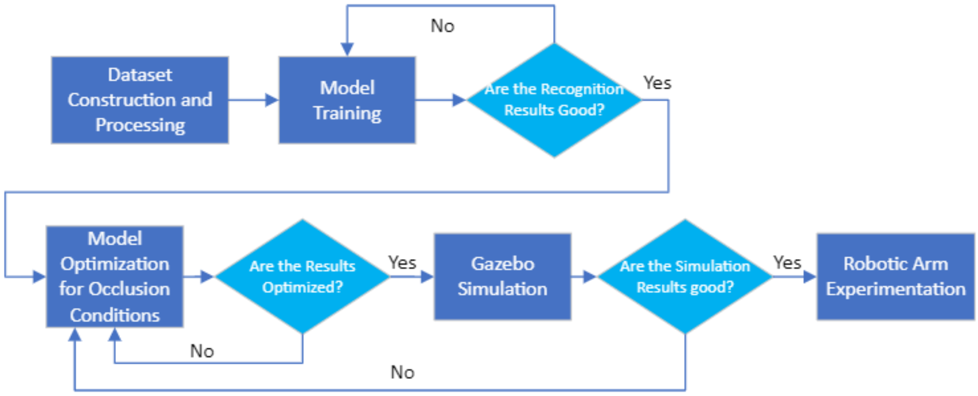


Fig.4 Test Process

5.3 Analysis of Results

YOLOv8 has further improved the feature extraction layer on the basis of v7, and the extraction effect has become better. In some papers, it has showed that the detection accuracy of YOLO v8 is better than most recognition algorithms. We predict that YOLO v8 will have a better performance in object recognition for robotic arms as well. At the same time, by refining the network, the accuracy of recognition of occluded object will raise, and better performance will be conducted in both simulation environments and an actual robotic arm experiment.

5.4 Inferences

By using the same training data set and under the same usage environment, it can be proved that YOLO v8 performs better or worse than other algorithms in target recognition for robotic arms. This will provide a reference for subsequent research in target recognition for robotic arms for other researchers. At the same time, by comparing the algorithms before and after the improvement, it can be proved whether the refined network proposed in this paper is effectiveness. Success of detection can be determined by accuracy and recall rate.

Chapter 6

Conclusion

In terms of the problem that the target recognition accuracy of the robotic arm is not high and the target is difficult to identify when it is obscured, this paper will use YOLO v8 algorithm of deep learning and modify the network to improve its performance. Meanwhile through horizontal comparison with other commonly used deep learning algorithms, the accuracy and advantages of the model proposed in this paper will be confirmed. The proposed algorithm will be tested in the simulation environment Gazebo and the actual robotic arm environment to check its effect and the improvement in occluded target recognition. This paper constructs a specific robot arm usage environment, which cannot represent all usage environments. In the future, more realistic usage environments and more complex backgrounds may be constructed for further confirmation. What’s more, more improvements will be conducted in the neural network part to achieve faster and more accurate recognition.

References

[1]Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S., 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).

[2] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C., 2016. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14* (pp. 21-37). Springer International Publishing.

[3] He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

[4]Lagerberg, T., & Jonson, J. (2016). Une bio du robot. Petite histoire de la robotique industrielle. ABB Review 3/16.

[5]Gaz, C., Cognetti, M., Oliva, A., Giordano, P.R. and De Luca, A., 2019. Dynamic identification of the franka emika panda robot with retrieval of feasible parameters using penalty-based optimization. *IEEE Robotics and Automation Letters*, *4*(4), pp.4147-4154.

[6] D. H. Kim and M. C. Lee, "6-DOF Robot Arm Visual Servoing with Canny Edge-Based Object Detection," *2021 21st International Conference on Control, Automation and Systems (ICCAS)*, Jeju, Korea, Republic of, 2021, pp. 147-150, doi: 10.23919/ICCAS52745.2021.9649807.

[7]K. Song, F. Zhu and L. Song, "Moving Target Detection Algorithm Based on SIFT Feature Matching," *2022 International Conference on Frontiers of Artificial Intelligence and Machine Learning (FAIML)*, Hangzhou, China, 2022, pp. 196-199, doi: 10.1109/FAIML57028.2022.00045.

[8] G. Yu, P. Yu, H. Li and H. Li, "An Improved Faster R-CNN Method for Car Front Detection," 2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Beijing, China, 2022, pp. 7-12, doi: 10.1109/IAEAC54830.2022.9929605.

[9]Yumang, A.N., Juana, M.C.M.S. and Diloy, R.L.C., 2022, March. Detection and classification of defective fresh excelsa beans using mask r-CNN algorithm. In *2022 14th International Conference on Computer and Automation Engineering (ICCAE)* (pp. 97-102). IEEE.

[10] X. Yu, T. W. Kuan, Y. Zhang and T. Yan, "YOLO v5 for SDSB Distant Tiny Object Detection," *2022 10th International Conference on Orange Technology (ICOT)*, Shanghai, China, 2022, pp. 1-4, doi: 10.1109/ICOT56925.2022.10008164.

[11] Patel, K., Patel, V., Prajapati, V., Chauhan, D., Haji, A. and De

gadwala, S., 2023, June. Safety Helmet Detection Using YOLO V8. In *2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)* (pp. 22-26). IEEE.

Appendix A
Workplan

A.1: Gantt Diagram of Workplan

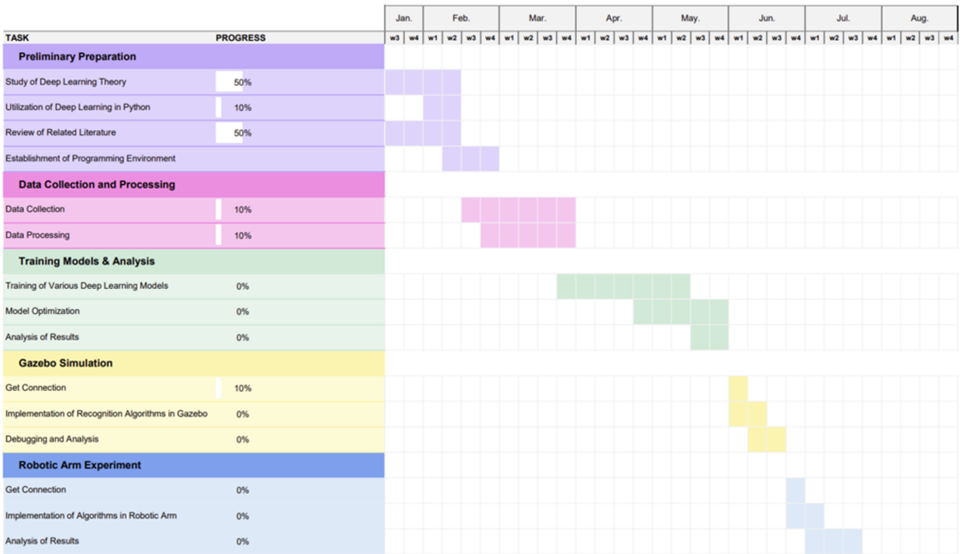


Fig.5 Gantt Diagram

A.2: Risks and Mitigation Plan

Table 2 describes risks associated with completing this project and a corresponding mitigation plan.

Tab.2 Risks and Plan

Risk	Likelihood	Impact	Contingency plan
Insufficient data volume or poor quality dataset	Medium	Medium	Utilize existing datasets such as COCO or Kaggle, and employ methods like data augmentation to increase dataset size and robustness.
Some deep learning models are unable to run	Medium	Medium	Switch to alternative deep learning approaches for comparison with YOLO v8.
Modified networks fail to achieve accuracy improvements.	Medium	Medium	Consider adopting alternative strategies to enhance accuracy, such as incorporating methods from image studies, or adjusting aspects like the size of anchor boxes and the values of Intersection Over Union (IOU).
Gazebo connection issues	Medium	Medium	Do not use the existing Gazebo models; instead, download other available Gazebo models. Alternatively, skip Gazebo and proceed directly with robotic arm experiments.
Robotic arm connection failures or can not move	Medium	Medium	Conduct recognition tests only without performing grasping tests. If Gazebo's simulation results are satisfactory, consider using Gazebo for additional simulation experiments.

Appendix B
Professional and Ethical Issues

B.1: Introduction

As an engineer, we need to consider whether the products we design and manufacture are ethical and friendly to society and the environment. This is also our responsibility as an engineer. This section mainly introduces the impact and avoidance of environmental, social, ethical and other issues that would be related to this paper.

B.2: PESTEL Analysis

PESTEL is a good analysis tool for assessing the environmental and social impacts of projects. In this part, PESTEL will be used to analyze the project proposed in this paper.

Political: Currently, most countries are actively promoting innovation and technological progress, and have published some relevant policies to regulate artificial intelligence, robots and other related fields as well. We will make sure to comply with the policy requirements of different countries in the whole process of researching, manufacturing, and selling our robotic arms and detection algorithm.

Economic: Robotic arms with high accuracy can greatly improve the production speed and the quality of products for factories. It will bring benefits to both factories and robot arm sellers. However, an excessively high utilization rate of robotic arms may bring about the risk of factory dependence. If the robotic arms are damaged in large quantities, it will cause greater losses to the factory. Therefore, it is necessary to have a complete warranty system and regular robotic arm testing for users.

Social: Extensive use of robotic arms may cause a large number of workers to lose their jobs. Therefore, factories must determine the number of robotic arms to use based on their own conditions and consider the balance between cost, revenue and expenditure. At the same time, robotic arms can also bring more jobs, such as robotic arm maintenance and code writing.

Technological: A robotic arm with high recognition accuracy can speed up its grasping process. At the same time, this recognition system can also be applied to other fields to promote the further development of robots and artificial intelligence.

Environmental: The production and disposal of robotic arms will cause certain pollution to the environment. Therefore, when manufacturing robotic arms, environmentally friendly materials should be considered to make them less polluting during production and easy to decompose.

Legal: Different countries have different laws and regulations on robots and robotic arms. Therefore, in the early stages of production, it is necessary to ensure that global development requirements are met to reduce legal disputes in the later stage. At the same time, if the requirements of some countries cannot be met, it is prohibited to export to this country or for sale in this country.

B.3: Ethical Considerations

Regarding the environment, the production and disposal processes will have a certain impact on the environment. Therefore, the material of the robotic arm must be environmentally friendly. At the same time, the convenience of transportation and the impact on the environment during transportation should be considered during the design process. In terms of law, it must meet the global design requirements and regulations for robotic arms and respect the laws of various countries. In terms of ethics, measures should be taken to protect robots from harm to human users, such as using tactile sensors to stop immediately when hit humans or adding protective covers to prevent people from approaching them while working. What is more, courses about programming or the introduction of such robotic arms can be opened to the public free of charge, while also educating them and creating more jobs. In terms of data safety, we will add data protection functions to ensure customer data security.

B.4: Inclusive Engineering Outcomes

- This plan takes into account the relevant interests and perspectives from developers, to producers, to users.
- This solution complies with the global common robotic arm development requirements and avoids prejudice and discrimination.
- This solution provides an identification solution that is more suitable for robotic arms, improves detection accuracy in industrial environments, and is in line with future development trends.
- This solution provides everyone with equal access to results as much as possible.
- This program is in line with the United Nations Sustainable Development Goals.

B.5: Conclusion

With the rapid development of the fields of artificial intelligence and robotics, more and more artificial intelligence products have entered people's lives. However, they have also brought about ethical issues that did not exist before. What if a robot causes harm to humans? What should we do if an auto-pilot car hits a pedestrian? The current field of artificial intelligence is still in the development stage. The relevant laws and regulations are not yet perfect, which makes it difficult to solve these problems when encountering them. Therefore, it is necessary to consider these issues when designing and strive to avoid the occurrence of such problems. In addition to meeting the issues under PESTEL analysis, some issues need to be considered. For example, the health issues of the product to producers and users, product quality issues, etc. Materials that are harmless to the human body must be used during production while ensuring the quality and reliability of the product and protecting the interests of customers.

For the SDG requirements, the development of high-precision robotic arms can bring about the development of SDG9 (industry, innovation and infrastructure), while supporting SDG3 (good health and well-being) and SDG12 (responsible consumption and production). It will create more jobs by creating more robotic arm development and maintenance personnel. The popularity of robotic arms accelerates the production speed and accuracy of factories and creates economic growth, so it achieves SDG8 (decent work and economic growth), which will bring more vitality to the city and ensure SDG11 (sustainable cities and communities) as well.