ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 12 N. 1 (2023), e31586
eISSN: 2255-2863
DOI: https://doi.org/10.14201/adcaij.31586

Computer Vision-Assisted Object Detection and Handling Framework for Robotic Arm Design Using YOLOV5

Ajmisha Maideen[a] Dr. A Mohanarathinam[b]

[a] Research Scholar, Faculty of Engineering, Department of Electronics and Communication Engineering. Karpagam Academy of Higher Education, Coimbatore, India-641021
[b] Assistant Professor, Faculty of Engineering, Department of Biomedical Engineering. Karpagam Academy of Higher Education, Coimbatore, India-641021
ajimsha06@gmail.com, mohanarathinam@gmail.com

KEYWORDS

Convolutional Neural Networks; Object Detection; Transfer Learning; Object Categorization; YOLOV5

ABSTRACT

In recent years, there has been a surge in scientific research using computer vision and robots for precision agriculture. Productivity has increased significantly, and the need for human labor in agriculture has been dramatically reduced owing to technological and mechanical advancements. However, most current apple identification algorithms cannot distinguish between green and red apples on a diverse agricultural field, obscured by tree branches and other apples. A novel and practical target detection approach for robots, using the YOLOV5 framework is presented, in line with the need to recognize apples automatically. Robotic end effectors have been integrated into a Raspberry Pi 4B computer, where the YOLOV5 model has been trained, tested, and deployed. The image was taken with an 8-megapixel camera that uses the camera serial interface (CSI) protocol. To speed up the model creation process, researchers use a graphical processing computer to label and preprocess test images before utilizing them. Using YOLOV5, a computer vision system-assisted framework aids in the design of robotic arms capable of detecting and manipulating objects. The deployed model has performed very well on both red and green apples, with ROC values of 0.98 and 0.9488, respectively. The developed model has achieved a high F1 score with 91.43 for green apples and 89.95 for red apples. The experimental findings showed that robotics are at the forefront of technological advancement because of the rising need for productivity, eliminating monotonous work, and protecting the operator and the environment. The same discerning can be applied to agricultural *robots, which have the potential to improve productivity, safety, and profit margins for farmers while reducing their impact on the environment. The system's potential could be seen in an assortment of fields, including sophisticated object detection, nuanced manipulation, multi-robot collaboration, and field deployment.*

1. Introduction

Recent years have seen substantial development in computer vision to aid with object recognition and manipulation in robotics. Convolutional neural networks (CNNs) have recently become a prominent method for object identification. The cutting-edge CNN model You Only Look Once (YOLOV5) has been shown to perform very well in real-time object detection and localization. In this research, we provide a unique framework design for an autonomous robotic arm that uses the YOLOV5 model for object detection and manipulation. The primary goal of this framework is to enhance the overall performance of the robotic arm by increasing the speed and precision with which object recognition and handling tasks are completed. Traditional object identification techniques need help with issues including sluggish detection speed, restricted item recognition, and an inability to handle various objects. We hope to address these issues by combining computer vision capabilities with the robotic arm. The YOLOV5 model overcomes these difficulties by tracking several items simultaneously, pinpointing their precise locations, and classifying them instantly into categories.

Multiple parts comprise the proposed framework: a camera module to record live video, a deep learning model based on YOLOV5 to identify and categorize objects, and a control module to coordinate robotic arm movements in response to object detection. Using YOLOV5's robust object recognition and manipulation capabilities, our system paves the way for the robotic arm to carry out complicated tasks in a dynamic and unstructured setting. The effectiveness of our framework in terms of item identification accuracy, speed, and the capacity to handle varied objects is measured by rigorous simulations and real-world trials. The results prove the efficacy and sturdiness of our method, emphasizing its potential uses in industries including manufacturing, supply chain management, and medicine. Our framework may significantly contribute to robotics by boosting the robotic arm's capabilities via cutting-edge object identification and handling methods, hence opening new pathways for automation and efficiency in a wide range of fields.

Nowadays, computer vision (CV) and deep learning (DL) have become more popular topics in robotics, healthcare sectors, and artificial intelligence. They deserve credit for their ability to convert, what were formerly thought to be insurmountable hurdles of data, into simpler information. This ability also enables the development of object detection applications. While computers can digest data far more quickly than humans, they still have trouble distinguishing between similar-looking objects in videos and images. Improvements in computer vision and image processing have allowed for more flexibility in the development and management of prosthetics in recent years (Intisar et al., 2021). Muscle-controlled (myoelectric) prostheses are much more widespread than their cable-operated predecessors. To help people with prosthetic hands grab and manipulate everyday products, biomedical engineers have set a challenge for themselves, to create a computer vision system for prosthetic hands (Starke et al., 2022). Myoelectric impulses from the muscles in the stump operate today's gripping limb prostheses, although mastering this method requires time and practice. To overcome these issues, it is vital to create a vision-based system that is both smart and effective. By integrating computer vision into the design of a bionic hand, it will be feasible for the user to reach out and grab any item with only a glimpse in the appropriate direction.

Object detection is a computer vision method that enables us to find specific things in images or videos. Labeling and counting items in a scene, pinpointing their positions, and following their movement are all possible because of the ability of object detection to precisely identify and localize them (Savio et al., 2022). Object identification methods can generally be categorized into machine learning (ML) and deep learning methods (Ji et al., 2023). Classical machine learning methods examine an image's color histogram and edge to determine whether a particular cluster of pixels represents an object (Hasan et al., 2022). A regression model is given these characteristics to predict where an item is and what label it has. To accomplish end-to-end, unsupervised object recognition where features do not need to be created and extracted individually, deep learning-based systems use convolutional neural networks. The family of region-based CNN models include the most widely used models for object detection (Junos et al., 2022). Models such as recurrent convolutional neural networks (RCNN) (Z. Wang et al., 2022), You Only Look Once (Reis et al., 2019), etc., are only a few of the many options for object identification. This research demonstrated the significant influence of computer vision on robotics industries in object recognition and handling.

This research used the YOLO framework for object identification and localization to guide the robotic arm's movements. Algorithms for object localization can identify the existence of an item in a photograph and use a bounding box to describe its precise position. These algorithms take as input an image containing one or more items and provide the coordinates of one or more bounding boxes based on the objects' positions, dimensions, and sizes. Methods for detecting objects can be roughly divided into two: those that rely on neural networks and those that do not. To classify images without resorting to neural networks, one must first define features by using feature engineering methods and then use a technique such as support vector machine (SVM) (Zhao et al., 2022), residual network (ResNet) (X. Zhang et al., 2019), visual geometric group (VGGNet) etc. Single shot detector (SSD), and region-based convolutional neural networks are some of the most well-known neural network algorithms (Abubeker & Baskar, 2023). Even though neural network approaches are more precise; they need a lot of labeled data to train well.

Researchers proposed the YOLO technique to address these difficulties, allowing pre-trained and fine-tuned models to test data. Since it delivers excellent accuracy on most real-time processing tasks while maintaining a respectable speed and frames per second, even on devices accessible to practically everyone, the YOLO algorithm has become one of the most common ways to identify an object in real-time. YOLO has become a popular real-time object identification technique since it just requires one run through the neural network to identify objects. Because of the computational requirements of the YOLO method, such as parallel processing, customized architecture, availability of libraries and frameworks, and scalability, graphics processing units are well suited for implementing YOLO. Due to the parallel nature of convolutions in neural networks, GPUs are much quicker than CPUs when used for YOLO object detection because of the large number of convolutions and computations performed in each picture frame. Everyday neural network operations, such as matrix multiplications and convolutions are handled efficiently by specific hardware components, such as tensor cores in many contemporary GPUs. The YOLO training and inference procedures are sped up considerably because of these specific components. YOLO models are complex because of the moving objects and variables. GPUs have the memory and processing capability to deal with such complex models effectively. Considering a series of criteria, including deployment scale, price, and application needs; we chose the Raspberry Pi 4B GPU hardware for model development and deployment.

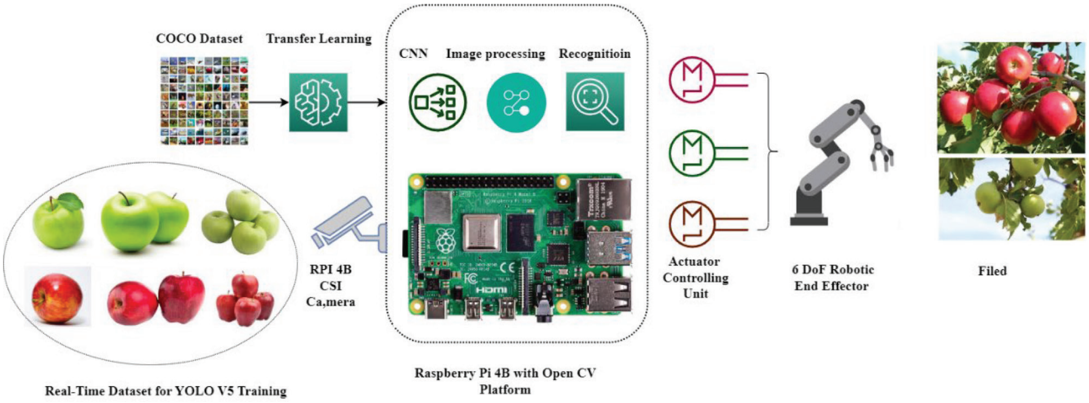


Figure 1. Graphical abstract of the computer vision-assisted object handling framework using YOLO V5

Figure 1 shows a Raspberry Pi 4B device operating with a YOLOV5 framework for object recognition and handling. Open CV is a platform that helps in image processing. In the past ten years, the area of computer vision has flourished in large part thanks to the advancements made possible by deep learning. This led to unprecedented precision in solving widespread computer vision issues, including image categorization, object recognition, and segmentation, with practical applications in the industry. Several state-of-the-art object detection architectures can be deployed immediately on real-world datasets to get acceptable detection results. The sole requirement is that the test dataset and the pre-trained detector have the same classes. However, the regularly created applications could use objects not represented in the pre-trained object detector classes. For instance, the distribution of the dataset is entirely dissimilar to the conditions in which the detector was trained. In this case, researchers can employ transfer learning, taking a detector that has already been trained and tweaking it for the new dataset. Using the pre-trained YOLOV5 object detector and a personalized dataset, YOLOV5 has terrific support and is much simpler to use in production. It provides a series of object identification architectures pre-trained on the MS COCO dataset (Lin et al., 2014). The most admirable aspect is that the constraints of the Darknet framework are removed since YOLOV5 is a native implementation in PyTorch.

Real-time fruit images are acquired using a NoIR (no infrared) camera equipped with a camera serial interface (CSI) protocol. A NoIR camera is a specific kind of camera that is devoid of an infrared filter, enabling it to record both visible and infrared light. NoIR cameras provide distinct advantages in applications that need infrared photography, such as those in the agriculture industry. Infrared cameras can detect and record infrared light, which is reflected but not perceptible to the human visual system. Farmers and agricultural researchers can use these cameras to identify the first indications of plant stress, illnesses, and nutrient deficits, allowing for prompt treatments that enhance crop health. In addition to their many applications, NoIR cameras are used to monitor water stress in crops. Plants under stress exhibit decreased infrared light reflection due to decreased water content. Alterations in the parameters of leaf reflectance characterize numerous plant diseases. NoIR cameras can detect these alterations in advance of the manifestation of apparent symptoms. The timely diagnosis of illnesses in agriculture enables the quick application of remedies, therefore mitigating the spread of diseases and reducing the extent of crop losses. In general, NoIR cameras provide a non-intrusive, economically viable, and effective means of monitoring and controlling crops, hence enhancing productivity, minimizing resource depletion, and promoting sustainable agricultural methodologies such as fruit categorization and sorting mechanisms. The main contributions of the research are listed below.

1. To identify which apples are green and red, we trained a YOLOV5 model using data from our own collection and the Kaggle dataset (Apple Historical Dataset, 2022).
2. The designed model has been implemented in the Raspberry Pi 4B model, tested in various environments, and has shown satisfactory results.
3. Contributes to a significant increase in productivity, reducing the need for human labor in agriculture.

The sections of this article are organized as follows: Section 2 discusses current advancements in fruit categorization and plucking using different algorithms and computer vision technology. Section 3 outlines the approach used in this study, including the dataset, image labeling, training, and testing procedures used with the YOLOV5 model. Finally, the constructed model's performance is tested in terms of accuracy, precision, recall, and F1 score across several testing scenarios in Section 4, followed by the conclusion.

2. Related Work

Robotics with built-in computer vision have significantly altered the automation landscape across several sectors. Robotic arm technology has significantly aided the automation of manufacturing procedures. Traditional robotic arms, on the other hand, can only move along predetermined lines and cannot adjust on the fly to accommodate different items and settings. Our goal is to improve robotic arms' ability to recognize and manipulate items by using computer vision methods. You Only Look Once is a fast and accurate real-time object identification technique. The newest version of the YOLO series, YOLOV5, is based on deep learning architectures and can perform object recognition in real time across a variety of hardware platforms. In this research, we introduce the YOLOV5 robotic arm framework, a state-of-the-art computer vision-assisted object recognition and handling system. By incorporating YOLOV5, robotic arms are given the ability to see their surroundings in real time, allowing for improved item detection, manipulation, and environment adaptation. When applied to automation processes in a wide range of sectors, the suggested framework has the potential to dramatically improve both efficiency and security. New possibilities for robotic automation could arise as this area of study as it progresses.

The research presented by Cognolato et al. (2021) looked at whether a multimodal method that uses expected human behavior (namely, eye-hand coordination) might help with the difficulty of categorizing different kinds of grasps for use with hand prostheses. Positive findings indicate that offline grip type classification ability is improved for transradial amputees when data from electromyography, gaze, and first-person video are fused. Therefore, the findings support the strategy based on eye-hand synchronization and demonstrate the value of a multimodal categorization of grip types. In addition, the dataset's accessibility paves the way for more research and enhancements, both of which are necessary to arrive at a strategy that can be evaluated in web-based software. A brain-computer interface (BCI) using invasive electroencephalography (EEG) has been demonstrated to control robotic arms with many DOFs. However, a multi-degree-of-freedom (DOF) robotic arm is challenging to control noninvasively due to the limitations of EEG decoding capabilities, making it difficult to reach and grip the intended object correctly in a complex 3D space. Using an EEG-based BCI, as proposed by Xu et al. (2022), the users of a robotic arm control it to perform multi-target reach and grasp tasks, as well as avoid obstacles with hybrid control.

An adaptive user profile and recognition system presented by Maroto-Gómez et al. (2023) for social robots to facilitate one-on-one encounters. Their objective is to use convolutional neural networks for flag identification. They built up a training and testing collection of items, trained a neural network to execute the detection task, ran several trials to see how well it did, and evaluated the results. These findings apply to general computer vision tasks and cognitive and robotics technology. Yan et al. (2021) proposed a lightweight object detection YOLOV5-CS model to count wild green citrus. Image rotation algorithms improved the model's reusability and generalization. Secondly, a detection layer was placed as the backbone to improve the citrus detection accuracy. The cosine annealing process and CIoU loss function enhance model training. They employ "virtual zone" scene segmentation to count green citrus in an embedded edge computing system.

In response to this real-world challenge, Yan et al. (2021) proposed a lightweight apple targets detection method for picking robots using enhanced versions of the YOLOV5 algorithm. This allowed the robots to automatically identify the graspable and ungraspable apples in the image of an apple tree. The original YOLOV5s network's BottleneckCSP module was upgraded and renamed the BottleneckCSP-2 module before being inserted into the network's backbone architecture. The enhanced backbone network included the SE module and enhancements were made to the inputs for the original YOLOV5s network's medium-sized target detection layer. Finally, the original network's anchor box size was enhanced. Orchard blooming level estimate on both a global and a block scale was the focus of a new approach proposed by Chen et al. (2022). Apple blossom data from different sources were used to assess the detector's resilience in varying lighting conditions and its ability to generalize between years. On average, the trained apple flower detector was 77.5% accurate. Orchard blooming was calculated using the blooming level estimator, which considered the relative abundance of flowers at various phases of development.

In the present proposal, one of the robot's central technologies is a sophisticated algorithm for recognizing apples as targets. While there are apple detection algorithms out there, the vast majority of them cannot tell the difference between apples obscured by tree branches and those obscured by other apples. In response to this real-world issue, we propose the use of YOLOV5 to implement a lightweight apple target detection method for picking robots, allowing them to automatically identify graspable and ungraspable apples within an image of an apple tree or target position.

3. Methodology

The Raspberry Pi 4B is a significant upgrade over its predecessor in processing speed, multimedia capabilities, memory, and connection, making it an ideal platform for use in embedded and robotic computing. High-performance 64-bit quad-core CPU, 4K micro-HDMI connectors, hardware video decoder, 4GB of RAM, dual-band 2.4/5.0 GHz wireless LAN, Bluetooth 5.0, Gigabit Ethernet, USB 3.0, and Power over Ethernet (PoE) functionality are all included in the Raspberry Pi 4B system on module computer.

This research first uses the MS COCO dataset, a massive resource for object recognition, segmentation, key-point detection, and captioning developed by Microsoft. The dataset contains 328 thousand photos, of which 118 thousand is for training and five thousand for validation. The compact nature of YOLOV5 makes its use more practical on portable devices. The COCO dataset serves as the basis for YOLO, a collection of object identification architectures and models that have already been pre-trained

(Dewi et al., 2020). YOLO, a representative one-stage detector, employs regression to forecast all categories and their confidence and bounding box information, which speeds up detection but reduces accuracy. Compared to previous versions of YOLO, YOLOV5's smaller model size means it might be used in novel ways, perhaps for things such as edge AI or machine learning on a microcontroller. This paper covers the specifics of the technologies used and the process of constructing the YOLOV5 machine learning framework. Moreover, apple color recognition (both green and red) is tested to ensure the framework's viability. Figure 2 depicts the steps taken to train and label a custom dataset using YOLOV5.

The initial phase of any item identification process is gathering training data; in this case, researchers are using publicly available datasets and custom apple dataset to develop a detection framework. Image annotation is the act of labeling or categorizing an image using annotation tools so that a model can learn to identify the data characteristics independently. Object detection differs from image classification, which assigns a label to the whole image by classifying individual items. Figure 3 below shows annotated apple data from the dataset.

Creating a training dataset for computer vision models can be accomplished with the help of the roboflow image annotation tool. It involves a human operator looking through a large number of pictures, picking out the important details, then marking them with labels and shapes to tell the computer what it is supposed to learn. Roboflow is a web-based annotation tool that allows researchers to easily input the video and image they captured before and output it as a sequence of photographs. This is the mechanism by which the dataset is divided into the training, validation, and testing sets. In addition,

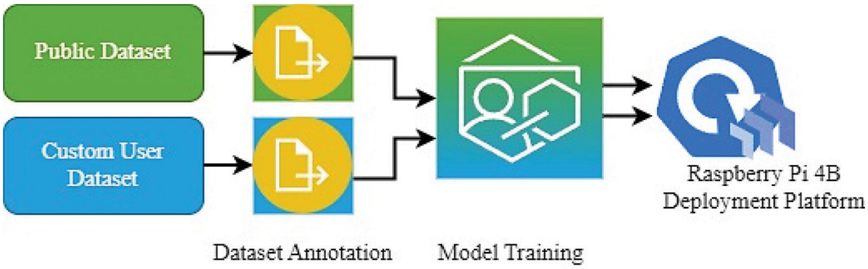


Figure 2. Workflow of the apple detection and categorization model developed in this research

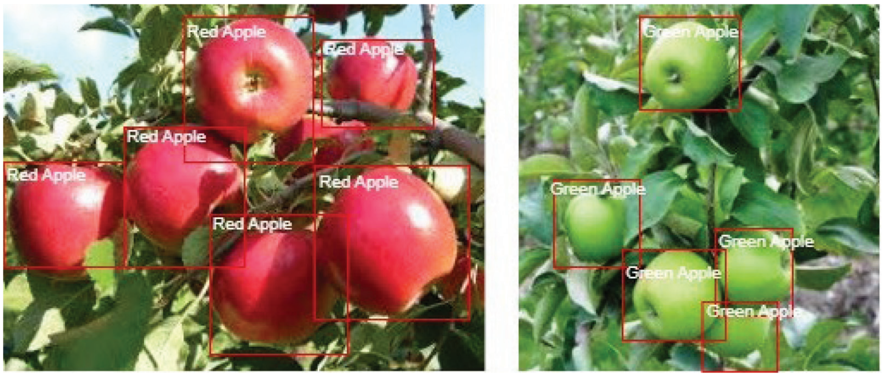


Figure 3. Labelled images used in the dataset for the training and testing process

the YOLOV5 PyTorch format has exported from these three files containing labeled datasets. Apples can be classified in part by their color and shape, considered throughout the training process.

To save money on computing and make a more effective model, developers are increasingly turning to transfer learning, which can apply to both bespoke and public datasets (Q. Wang et al., 2022; D. Wang et al., 2021; Xia et al., 2023). One machine learning method, transfer learning, uses a previously trained model to address unrelated issues. Simply defined, transfer learning is a machine learning technique where an already-trained model is used as the basis for another unrelated task. Transfer learning can outperform methods trained with fewer data when applied to a new problem. Instead, academics and data scientists like to start with a pre-trained model that has been taught how to identify things and has picked up on familiar visual cues, such as edges and shapes. In this paper, researchers have applied transfer learning to achieve a series of goals. Firstly, to eliminate the need to retrain several machine learning models to do the same tasks, which would have wasted both time and resources. Secondly, to save time and energy in the resource-intensive facets of machine learning, such as red and green apple image classification. And finally, compensate for the scarcity of labeled training data in-house by relying on previously trained models.

However, throughout the deep learning-based research of apple target identification, while the recognition accuracy of most current apple detection models was good, the real-time performance of many of them was poor owing to their high complexity, large number of parameters, and large size. Consequently, to fulfill the needs of the picking robot for real-time identification, it was crucial to create a lightweight apple target detection algorithm that would assure the accuracy of fruit recognition. To automatically recognize the apples that are graspable by a picking robot and the ones that are not, a lightweight apple target real-time recognition method based on enhanced YOLOV5 has been presented for picking robots. The proposed technique has the potential to provide technical assistance for the apple-picking robot's real-time, precise identification of numerous fruit targets.

In Figure 4, a three DoF robotic manipulator and the deployment model of the developed YOLOV5 framework on Raspberry Pi 4B single board computer for real time apple detection and plucking is

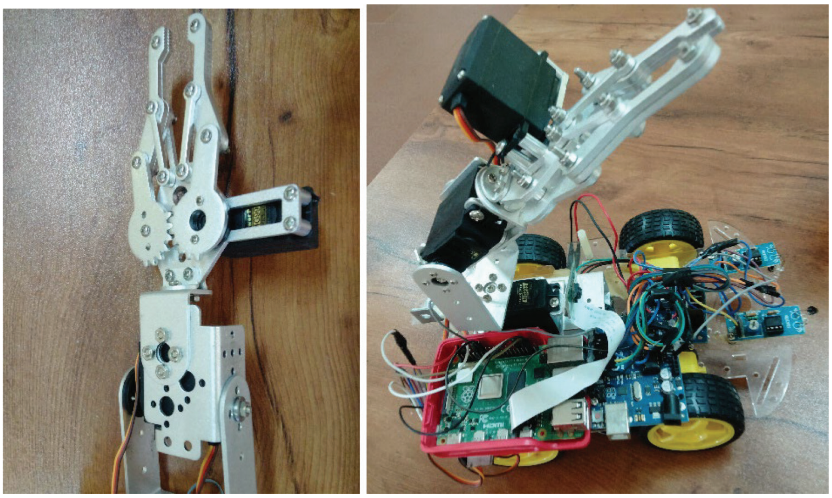


Figure 4. A 3 DoF robotic manipulator and the deployment model of the developed YOLOV5 framework on Raspberry Pi 4B single board computer for real time apple detection and plucking

presented. From the industrial sector to the space exploration industry, robotic manipulators play a key role. Using the joint angles or joint variables, forward kinematics could be used to calculate the location and orientation of the robot's end-effector. Here, we can go through the planning and derivation of forward kinematics equations for a basic robotic manipulator with three degrees of freedom. For example, a robotic manipulator with three degrees of freedom that consists of three revolute joints (rotational joints) linked in series. Each joint's rotation axis is aligned with the z-axis of its own coordinate frame (i = 1, 2, 3) where E is the coordinate frame representing the manipulator's end-effector (x, y, z). A careful consideration of the robot's kinematics, end-effector design, vision system, and motion planning, is required when designing a robotic manipulator to pick green and red apples. The apple-picking end-effector may safely and gently grab and remove apples from the tree. When handling soft fruits, pneumatic grippers are a common tool.

The classification challenge has been accomplished by training a convolutional neural network on a dataset of labelled apple photos. The forward kinematics equation was created to find the location and orientation of the end-effector in the workspace from the joint angles. Calculating the joint angles needed to reach a certain apple, depending on its detected location in the camera frame, would require the use of inverse kinematics. Create a motion planning algorithm that plots out safe routes for the robot to follow so that it can pick up each apple. Before implementing the robotic system in an actual apple orchard, we have completed thorough testing and validation in a controlled setting. Perception, gripping, and system performance are only few of the areas that need to be examined. The visual coordinate data is used to guide the robotic manipulator while it does the plucking operations, with the green and red apples going to the right of the robot and the red chilies going to the left. Inverse kinematics is a mathematical method used to compute the angle at which the manipulator must travel to achieve a desired goal position.

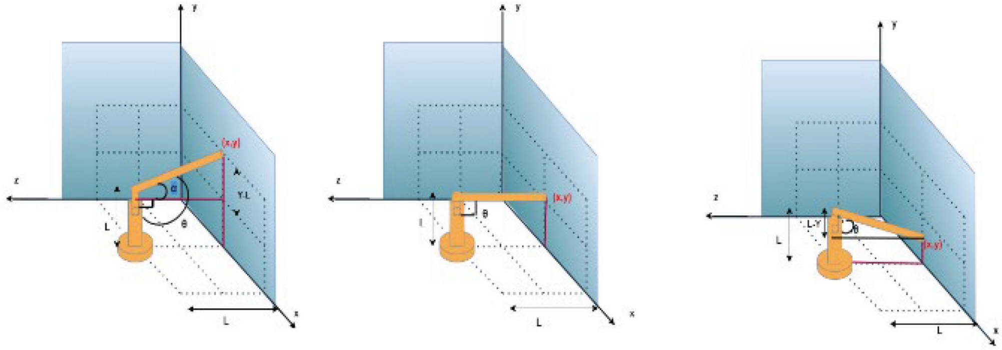


Figure 5. Design of inverse kinematics of elbow part with the coordinate system for 3 DoF robotic manipulator

Figure 5 depicts three distinct scenarios in which an angle with either the elbow or the central section can be determined using coordinates. (x,y)

$$\text{If } y > l, tan\alpha = \frac{y-l}{l} \tag{1}$$

$$\theta = 90^\circ + \alpha \tag{2}$$

$$\text{If } y = l,\ \theta = 90^\circ \tag{3}$$

$$\text{If } y < l, tan\theta = \frac{l}{l - y} \tag{4}$$

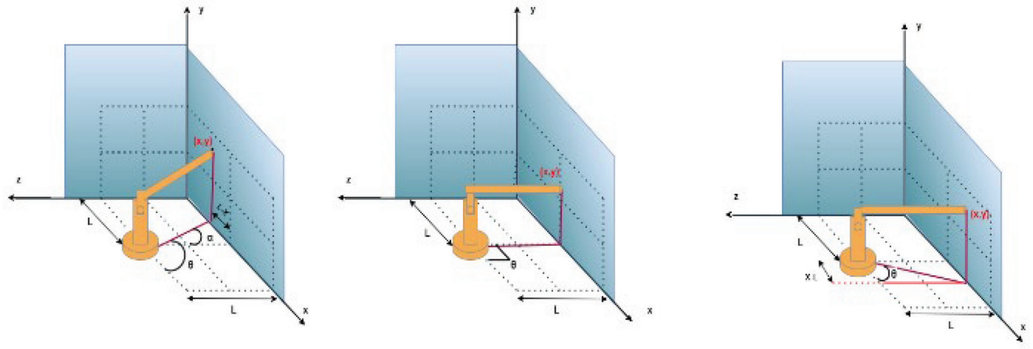


Figure 6. Design of inverse kinematics of base part with the coordinate system for 3 DoF robotic manipulator

Figure 6 presents three distinct scenarios that may be used to determine the angle of the base given its coordinates on the (x, y) plane.

$$\text{If } x < l, tan\alpha = \frac{l - x}{l} \tag{5}$$

$$\theta = 90^\circ + \alpha \tag{6}$$

$$\text{If } x = l,\ \theta = 90^\circ \tag{7}$$

$$\text{If } x > l, tan\theta = \frac{l}{x - l} \tag{8}$$

Using Equations 5 to 8, we can determine the exact angle at which the robotic manipulator will move. To determine the angles, they are designed to use the picture coordinates in real time.

4. Experimental Results

Training on a vast amount of image data gathered from various open-source platforms, and real-time images allows for constructing a deep learning model for target recognition. First, from the photographs in the Kaggle dataset and online repositories, 266 red apple images and 443 green apple images were randomly picked as the test set: the other 995 and 876 images as the training set for red and green apples, respectively. Table 1 displays the dataset image distribution details.

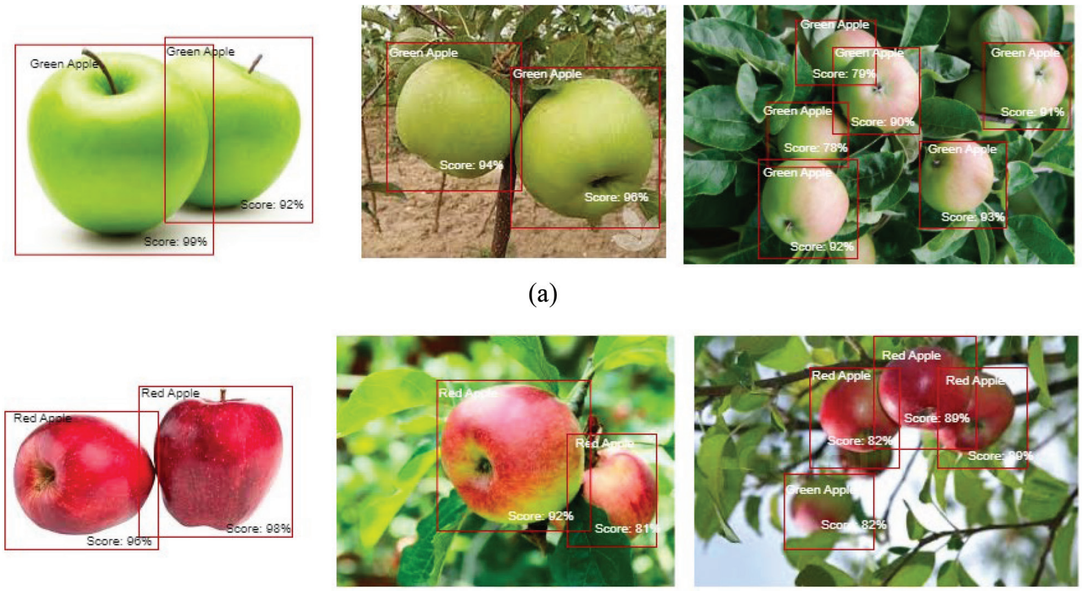
Table 1. Details of image dataset used in the green and red apple detection framework

Type	Training	Testing	Validation	Total
Green Apple	266	995	120	1381
Red Apple	443	876	134	1453
Total	709	1871	254	2834

Once decided on a splitting ratio, researchers can implement strategies for dividing up the data. The 80:20 split is often employed in this study, with the former indicating the usage of more training data than testing data. This rule is grounded in the well-known Pareto principle, which is, once again, just a rule of thumb utilized by professionals in the field. Second, the images are resized, brightened, and contrast-adjusted to the original 2834 images in the training, testing and validating dataset to boost the model's training efficiency with apple targets. Finally, processed images with apple targets were manually annotated by drawing rectangle boxes around them using image data annotation.

The red and green apple image classification model was trained, tested, and validated on a graphics processor enabled Raspberry Pi 4B system. The constructed model was introduced in a Raspberry Pi with a robotic manipulator to pluck or handle the two-class classification model. A total of 254 photos have been evaluated with the generated model, with acceptable classification and object detection accuracy. Figure 4 depicts some of the red and green apple images utilized in the validation process.

Figure 7 shows how we have verified the model's accuracy using a variety of image types, including images of green and red apples. The boundary boxes categorize the apples' varieties, and the scores are also displayed. All green was recognized in Figure 7.a with varying degrees of certainty. The score values



(a)

(b)

Figure 7. Results of object detection using the deployed YOLOV5 model on a Raspberry Pi 4B computer. (a). Green Apples detection and (b) Red Apples detection

represent the percentage of correct detections. Some red apples were mistakenly scored as green in Figure 7.b lowering their overall value. Due to the small number of datasets used during training, the model's accuracy suffered slightly. The model can be retrained with a new dataset using data augmentation and assembly approaches to reduce error and improve accuracy. Researchers have achieved a high level of model abstraction in graphics processing platforms using transfer learning and the YOLOV5 architecture.

The research team builds a real-time recognition model for the apple picking robot using YOLOV5 and tests it on a Raspberry Pi 4B computer to ensure the accurate detection of green and red apples. There are four metrics used to evaluate a model's efficacy: accuracy (Acc), precision (Pre), recall (Rec), and F1 score (F1). The ratio of correctly predicted observations to the total number of observations makes up accuracy. Precision is the ratio of successfully predicted positive observations to all expected positive observations. A recall is the percentage of correctly predicted positive observations in a class. The following are the definitions for the four performance indicators shown in Equations 1 to 4.

$$\text{Acc}= \text{TP+TN/TP+FP+FN+TN} \tag{9}$$

$$\text{Pre}= \text{TP/TP+FP} \tag{10}$$

$$\text{Rec}= \text{TP/TP+FN} \tag{11}$$

$$\text{F1 Score} = 2*(\text{Rec} * \text{Pre}) / (\text{Rec} + \text{Pre}) \tag{12}$$

In Equations 9 to 12, a true positive (TP) indicates that the observed value matches the predicted value (positive), a false positive (FP) indicates that the observed value does not match the predicted value (negative), a false negative (FN) indicates that the observed value does not match the predicted value (negative), and a true negative (TN) indicates that the observed value matches the predicted value (negative).

Table 2. Performance evaluation of the deployed model on a Raspberry Pi 4B computer

Type	Accuracy	Precision	Recall	F1 Score
Green Apple	87.50	91.95	90.91	91.43
Red Apple	85.82	91.40	88.54	89.95

In Table 2, F1 scores of 91.43 and 89.95 were achieved for red and green apple categorization, respectively. Using the YOLOV5 framework, have been able to create an apple identification system with high accuracy (87.50 and 85.82), precision (91.95 and 91.40), recall (90.91) and 88.54), and F1 score (91.43 and 89.95) for both red and green apples, respectively. We have opted to use the ROC curve when evaluating the efficacy of the YOLOV5 classification model we had created (receiver operating characteristic curve). The ROC curve compares the true positive rate (TPR) to the false positive rate (FPR) at various cutoff points for making a classification. Raising the threshold for classification causes more items to be falsely labelled as positive while simultaneously increasing the number of true positives.

Here,

$$\text{TPR}=(\text{TP}/(\text{TP+FN}) \tag{13}$$

$$\text{FPR}=(\text{FP}/(\text{FP+TN}) \tag{14}$$

In Figure 8, we have plotted the classification performnce of the proposd model for clasifying the green and red apples uisng RCO curve.

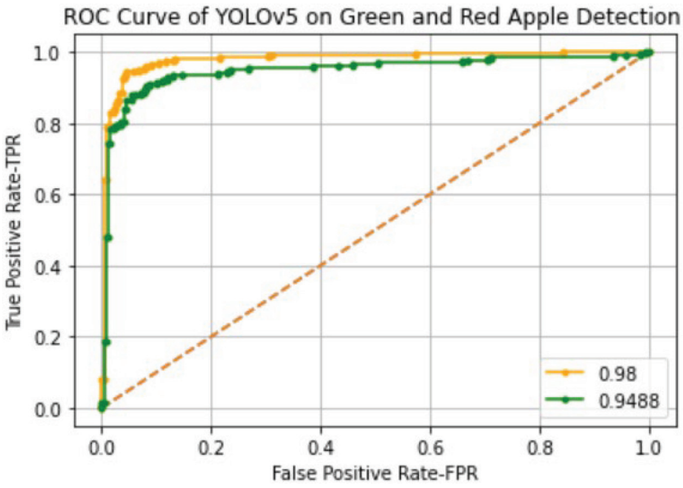


Figure 8. ROC curve of the developed YOLOV5 model for green and red apple detection process

TPR vs. FPR at varying cutoffs for classification are shown on a ROC curve. An increase in false positives and true positives results in a reduction in the categorization threshold. Figure 7 shows that the model is very accurate for both red and green apples (r = 0.98 and r = 0.9488, respectively). If a model's ROC is 1, then it successfully predicted the test data.

5. Conclusion

Robotic harvesters perform better in challenging agricultural settings when computer vision and related algorithms are used to enhance their performance in these areas and their intelligence and capacity for remote interaction. The future of advanced agricultural applications holds great promise for computer vision and emerging technology. Most harvesting robots have yet to reach accurate commercial applications because of the numerous technical challenges associated with machine vision and its precise positioning. An enhanced version of the YOLOV5 algorithm is presented as a real-time detection approach for the apple-picking robot to automatically recognize red and green fruits in the images of an apple tree. The original network's anchor box size was increased to prevent the network from incorrectly identifying little apples in the image's background. The existing capabilities and future enhancements of the framework are addressed, showcasing its wide range of industry application and revolutionary potential in the realm of automation. The current architecture makes use of a high-resolution camera attached to the robotic arm to take apple pictures in real time. YOLOV5 analyzes these photos in real time, allowing the robotic arm to locate and label items. The identified item's size, shape, and orientation are used to plot the robotic arm's most efficient route for handling and manipulating the object. Multiple domains, such as advanced object recognition, fine-grained manipulation, multi-robot collaboration, real-world deployment, etc., provide hope for the future of this system. This technology has the potential to change automated processes, boost productivity, and make workplaces safer if it continues to be developed and used. Test set detection results demonstrated that the YOLOV5 network model could reliably distinguish between apples within the picking robot's reach and those out of reach in the displayed apple tree image. The developed apple detection model with the YOLOV5 framework has produced good results for the red and green apples, respectively, in terms of accuracy (87.50, 85.82), precision (91.95, 91.40), recall (90.91, 88.54), and F1 score (91.43, 89.95). Results from prior studies were compared to those from the trained YOLOV5 algorithm proposed in this work for detecting two different types of subjects.

Conflict of Interest Statement

The authors Ajimsha Maideen and Dr. A Mohanarathinam certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Equal Contributions

All the authors have contributed equally to this manuscript and research work.

6. References

"Apple Historical Dataset" Apple Historical Dataset | Kaggle. /datasets/prasoonkottarathil/apple-lifetime-stocks-dataset (accessed Dec. 22, 2022).

Abubeker, K. M., & Baskar, S. (2023). B2-Net: an artificial intelligence powered machine learning framework for the classification of pneumonia in chest x-ray images. *Machine Learning: Science And Technology*, *4*(1), 1-23. https://doi.org/10.1088/2632-2153/acc30f

Chen, Z., Su, R., Wang, Y., Chen, G., Wang, Z., Yin, P., & Wang, J. (2022). Automatic Estimation of Apple Orchard Blooming Levels Using the Improved YOLOv5. *Agronomy*, *12*(10), 2483. https://doi.org/10.3390/agronomy12102483

Cognolato, M., Atzori, M., Gassert, R., & Müller, H. (2021). Improving Robotic Hand Prosthesis Control With Eye Tracking and Computer Vision: A Multimodal Approach Based on the Visuomotor Behavior of Grasping. *Frontiers In Artificial Intelligence*, *4*, 744476. https://doi.org/10.3389/frai.2021.744476

Dewi, T., Risma, P., & Oktarina, Y. (2020). Fruit sorting robot based on color and size for an agricultural product packaging system. *Bulletin Of Electrical Engineering And Informatics*, *9*(4), 1438-1445. https://doi.org/10.11591/eei.v9i4.2353

Hasan, S., Jahan, M. S., & Islam, M. I. (2022). Disease detection of apple leaf with combination of color segmentation and modified DWT. *Journal Of King Saud University - Computer And Information Sciences*, *34*(9), 7212-7224. https://doi.org/10.1016/j.jksuci.2022.07.004

Intisar, M., Khan, M. M., Islam, M. R., & Masud, M. (2021). Computer Vision based robotic arm controlled using interactive GUI. *Intelligent Automation And Soft Computing*, *27*(2), 533-550. https://doi.org/10.32604/iasc.2021.015482

Ji, S-J., Ling, Q., & Han, F. (2023). An improved algorithm for small object detection based on YOLO v4 and multi-scale contextual information. *Computers & Electrical Engineering*, *105*, 108490. https://doi.org/10.1016/j.compeleceng.2022.108490

Junos, M. H., Khairuddin, A. S. M., & Dahari, M. (2022). Automated object detection on aerial images for limited capacity embedded device using a lightweight CNN model. *Alexandria Engineering Journal*, *61*(8), 6023-6041. https://doi.org/10.1016/j.aej.2021.11.027

Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. *arXiv*. https://doi.org/10.48550/arXiv.1405.0312.

Maroto-Gómez, M., Marqués-Villarroya, S., Castillo, J. C., Castro-González, Á., & Malfáz, M. (2023). Active learning based on computer vision and human–robot interaction for the user profiling and behavior personalization of an autonomous social robot. *Engineering Applications Of Artificial Intelligence*, *117*, 105631. https://doi.org/10.1016/j.engappai.2022.105631

Reis, D. H., Welfer, D., De Souza Leite Cuadros, M. A., & Gamarra, D. F. T. (2019). Mobile Robot Navigation Using an Object Recognition Software with RGBD Images and the YOLO Algorithm. *Applied Artificial Intelligence, 33*(14), 1290–1305. https:// doi.org/10.1080/08839514.2019.1684778

Savio, A., Dos Reis, M. C., Da Mota, F. A. X., Marciano Martinez, M. A., & Auzuir Alexandria, A. R. (2022). New trends on computer vision applied to mobile robot localization. *Internet Of Things And Cyber-Physical Systems*, *2*, 63-69. https://doi.org/10.1016/j.iotcps.2022.05.002

Starke, J., Weiner, P., Crell, M., & Asfour, T. (2022). Semi-autonomous control of prosthetic hands based on multimodal sensing, human grasp demonstration and user intention. *Robotics And Autonomous Systems*, *154*, 104123. https://doi.org/10.1016/j.robot.2022.104123

Wang, D., & He, D. (2021). Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. *Biosystems Engineering*, *210*, 271-281. https://doi.org/10.1016/j.biosystemseng.2021.08.015

Wang, Q., Cheng, M., Huang, S., Cai, Z., Zhang, J., & Yuan, H. (2022). A deep learning approach incorporating YOLO v5 and attention mechanisms for field real-time detection of the invasive weed Solanum rostratum Dunal seedlings. *Computers And Electronics In Agriculture*, *199*, 107194. https://doi.org/10.1016/j.compag.2022.107194

Xia, R., Li, G., Huang, Z., Meng, H., & Pang, Y. (2023). Bi-path Combination YOLO for Real-time Few-shot Object Detection. *Pattern Recognition Letters*, *165*, 91-97. https://doi.org/10.1016/j.patrec.2022.11.025

Xu, B., Li, W., Liu, D., Zhang, K., Miao, M., Xu, G., & Song, A. (2022). Continuous Hybrid BCI Control for Robotic Arm Using Noninvasive Electroencephalogram, Computer Vision, and Eye Tracking. *Mathematics*, *10*(4), 618. https://doi.org/10.3390/math10040618

Yan, B., Fan, P., Lei, X., Liu, Z., & Yang, F. (2021). A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. *Remote Sensing*, *13*(9), 1619. https://doi.org/10.3390/rs13091619

Zhang, X., Fu, L., Karkee, M., Whiting, M., & Zhang, Q. (2019). Canopy Segmentation Using ResNet for Mechanical Harvesting of Apples. *IFAC-PapersOnLine*, *52*(30), 300-305. https://doi.org/10.1016/j.ifacol.2019.12.550

Zhao, K., Li, H., Zha, Z., Zhai, M., & Wu, J. (2022). Detection of sub-healthy apples with moldy core using deep-shallow learning for vibro-acoustic multi-domain features. *Measurement: Food*, *8*, 100068. https://doi.org/10.1016/j.meafoo.2022.100068