INSTITUTE OF SPACE TECHNOLOGY
KICSIT, Kahuta Campus

AUTOMATIC TARGET RECOGNITION BY ROBOTIC ARM

By

Saddam Ahmed Nisar

(192101024)

Iqra Tanveer

(192101015)

Nouman Mazhar

(192101014)

Supervisor

Engr. Mughees Sarwar Awan

Department of Computer Engineering

Institute of Space Technology, KICSIT Kahuta Campus

June 2023

AUTOMATIC TARGET RECOGNITION BY ROBOTIC ARM

By

Saddam Ahmed Nisar

(192101024)

Iqra Tanveer

(192101015)

Nouman Mazhar

(192101014)

Supervisor

Engr. Mughees Sarwar Awan

Department of Computer Engineering

Institute of Space Technology, KICSIT Kahuta Campus

June 2023

AUTHORS DECLARATION

I take full responsibility of the research work conducted during the thesis titled "**Automatic Target Recognition By Robotic Arm**".

I solemnly declare that the research and development work presented in the thesis is done solemnly by me with no significant help from any other person. However, small help wherever taken is duly acknowledged. I have also written the thesis by myself. Moreover, I have not presented this thesis or any part of the thesis previously to any other degree awarding institution within Pakistan or abroad.

I understand that the management of IST has a zero-tolerance policy towards plagiarism. Therefore, I as an author of the above mentioned thesis solemnly declare that no portion of my thesis has been plagiarized and any material used in this Thesis does not contain any literal citing (verbatim) of more than 70 words (total) even by giving a reference unless I have obtained the written permission of the publisher to do so. Furthermore, the work presented in the thesis is my own original work and I have positively cited the related work of the other researchers by clearly differentiating my work from their relevant work.

I further understand that if I am found guilty of any form of plagiarism in my thesis work even after my graduation, the institute reserves the right to revoke my BS degree. Moreover, the Institute will also have the right to publish my name on its website that keeps a record of the students who plagiarized in their thesis work.

Saddam Ahmed Nisar

Iqra Tanveer

Nouman Mazhar

I hereby acknowledge that the submitted thesis is the final version and should be scrutinized for plagiarism as per IST policy.

____

Engr. Mughees Sarwar Awan

Dated: ________________

____

Verified by Plagiarism Cell Officer

Dated________________

CERTIFICATE

This is to certify that the research work described in this thesis is the original work of the author and has been carried out under my direct supervision. I have personally gone through all the data/results/materials reported in the manuscript and certify their correctness/authenticity. I further certify that the material included in this thesis is not plagiarized and has not been used in part or full in a manuscript already submitted or in the process of submission in partial/complete fulfillment of the award of any other degree from any institution. I also certify that the thesis has been prepared under my supervision according to the prescribed format and I endorse its evaluation for the award of Bachelor of Science in Computer Engineering degree through the official procedures of the Institute.

Engr. Mughees Sarwar Awan

DEDICATION

This research work is dedicated to our beloved Parents, Teachers and Friends who supported and motivated us in tough times and made us into what we are today.

ACKNOWLEDGEMENTS

First of all, we thank Allah, the almighty, for giving us the strength to carry on this project and for blessing us with many great people who have been the greatest support in our personal and professional lives. We also thank our parents, for advising, praying, and supporting us in everything in our lives, especially for 4 years of studying Engineering. We owe so much to our whole family for their undying support, and their unwavering belief that we can achieve so much. We would like to thank our university Dr. A Q Khan Institute of Computer Sciences & Information Technology for teaching us Computer Engineering and for all our professors and doctors

We deeply thank our Director Khalid Masood whose help and advice were invaluable.

Our sincere gratitude to Dr. Waqar Farooq (HOD CE, KICSIT) and all faculty members for their constant encouragement, motivation and guidance through all phase of the project.

Special thanks and honor to our respected project supervisor, Engr. Mughees Sarwar Awan. His kind behavior and intellectual support made it possible for us to uplift the project from the foundational stage to maturity. The project is highly indebted to his kind guidance, encouragement, and full cooperation.

I am deeply thankful to the concerned faculty members at KICSIT, especially for their guidance, cooperation, and right direction.

Regards:

Saddam Ahmad Nisar

Iqra Tanveer

Nouman Mazhar

ABSTRACT

Automatic Target Recognition (ATR), which uses a number of sensors and algorithms, allows a robotic arm to locate and follow a specified target. Manufacturing, search and rescue, and surveillance are just a few examples of the many applications for this technology. ATR frequently involves the use of sensors, like as cameras, to gather information about the surroundings. After acquiring this information, algorithms that make use of computer vision or machine learning may be employed to find and track the target. The robotic arm may then move and control things depending on where the target is located. One of the key challenges for a robotic arm in ATR is adjusting to changes in the target's location or appearance, as well as the existence of extra objects or impediments in the environment. The use of many sensors, the integration of data from several modalities, as well as machine learning techniques have all been developed by researchers as solutions to these problems. These methods have improved the robot's ability to recognize and follow the target. The main goal of ATR by a robotic arm is to improve the robot's ability to recognize targets and interact with them in a dynamic and unpredictable environment, ultimately resulting in a more effective and autonomous robot. This research offers a fresh approach to Automated Target Recognition (ATR) using the Swin Transformer architecture. The network learns to extract fine-grained spatial features and contextual information from images, which enables accurate object detection. The experiments' findings demonstrate that the suggested approach outperforms more established techniques. Robotic Arms are now capable of interacting with their surroundings and performing challenging manipulation tasks on their own.

Keywords: Automatic Target Recognition, Robotic Arm, Swin Transformer, Object Detection

Table of Contents

List of Figures

List of Tables

Table 1: Acronyms and Abbreviations

Acronyms	Abbreviation
ATR	Automatic Target Recognition
GPU	Graphical Processing Unit
LIDAR	Light Detection and Raging
Rtx	Ray Tracing Texel Extreme
CV	Computer Vision
Val	Validation
OAK-D	OpenAI Kit Depth
RA	Robotic Arm
ST	Swin Transformer
DS	Dataset
IR	Image Recognition
Rec	Recall
Acc	Accuracy
RT	Real Time

CHAPTER 1

INTRODUCTION

Introduction

1.1 Project Background

The capacity of a robotic system to recognize and find certain targets or items in its surroundings without the help of a human is known as Automatic Target Recognition (ATR) by robotic arm. ATR stands for Automatic Target Recognition, which is the act of automatically seeing, recognizing, and categorizing targets or objects in pictures or sensor data without the assistance of a person. Automatic target recognition is an essential task in numerous fields as well as the military, manufacturing, and healthcare. Using many technologies including computer vision and machine learning it involves detecting and describing objects or target in a given environment [1].

Human operators have traditionally done manual target recognition which can be time-consuming or error-prone. Automatic target recognition has significantly increased in accuracy and efficiency with the use of the robotic arm [2]. These robotic arms have sensors and algorithm that enables them to recognize and categorize the targets with great speed and accuracy. The ability of the robotic arm to operate in dangerous or challenging domains is one of the key advantages of using automatic target recognition. For example

- Without affecting human soldiers, robotic arms can be employed in a military environment to identify and categorize the enemy target.
- A robotic arm can be used in a productive environment to find and categorize product flaws and enhance quality control.

In recent years, a huge advancement in automatic target recognition by robotic arms, with several studies and projects aimed at increasing the precision and speed of these systems.

Although, there is a still need for more research and development in this area because there are still numerous obstacles to be solved, as well as the capacity to adapt to changing surroundings and the capacity to identify and categorize a wide variety of targets.

In general, Automatic Target Recognition by robotic arms has the potential to transform a number of industries by increasing productivity and accuracy while lowering the dander to human operators. Automatic Target Recognition (ATR) by a robotic arm project involves the development of a system where a robotic arm is able to detect, track and recognize targets using various sensors and algorithms [3]. The goal of this project is to provide a solution for automating tasks that require precise target recognition and manipulation. This technology has various applications in fields such as manufacturing, inspection, surveillance, and defense.

Typically, the ATR system consists of a robotic arm with sensors such as cameras, laser rangefinders, and depth sensors that are utilized to collect information about the target [7]. In order to identify the target and determine its location, machine learning algorithms are then used to the obtained data. Once the target has been identified, the robotic arm can be programmed to carry out tasks like grasping, moving, or manipulating the object.

This kind of project necessitates knowledge in control systems, computer vision, robotics, and other related disciplines. To be viable for applications, the ATR system needs to be strong and capable of handling a variety of real-world scenarios, such as shifting illumination conditions or occlusions [4].

1.2 Problem Statement

To develop a system that automatically detect or recognize and track the moving object using the robotic arm.

1.2.1 Problem Description

Using image processing and machine learning approaches, the problem of Automatic Target Recognition (ATR) by a robotic arm entails enabling a robotic arm to recognize and identify objects in its environment [1]. Typically, the robotic arm has a camera or other sensor that takes pictures of the objects. These pictures are then processed using computer vision algorithms to find and detect the things. Machine learning techniques are used to classify and identify the items after they have been spotted. This knowledge is then applied by the robotic arm to decide how to engage with the object, for as by gripping it or avoiding it. Accurately detecting and identifying items in diverse and dynamic situations, as well as processing the massive volumes of data the sensor generates, are the key obstacles that a robotic arm faces when performing ATR.

1.3 Project Objectives

The ultimate objective for the development of robotic arm technology for autonomous target recognition is:

- This includes detecting and identifying the target using sensors like cameras, as well as utilizing machine learning techniques to increase the accuracy and speed of target recognition.
- To achieve great accuracy and reproducibility in moving and manipulating the target, algorithms, and controllers must be developed.
- Changes in lighting, object shape, and other elements that may impact the sensor's capacity to detect and identify the target should be handled by ATR systems.
- In order to react quickly to environmental changes and carry out tasks effectively, ATR systems should be able to process sensor data and drive the robotic arm in real time.

- An ATR system should be capable of doing a variety of activities, such as classifying items based on their type, picking them up and moving them, or otherwise manipulating them.

1.4 Project Scope

The following elements may be included in the scope of a robotic arm's Automatic Target Recognition (ATR) project:

- Creating methods for image processing and machine learning will increase target recognition's precision and speed, including object detection, tracking, and categorization.
- Robotic arm control involves developing and putting into practice control algorithms to move and operate the target with great precision and consistency.
- Integration of the sensor, image processing, and operation of the robotic arm into the overall ATR system.
- Validating the ATR system's functionality and dependability through testing and validation, which includes performance, environmental, and safety testing.
- Setting up and maintaining the ATR system in the specified setting, as well as giving users support and training.

Our project is divided primarily into two parts. The robotic arm is entirely assembled and secondly implement the code in required software.

1.4.1 Robotic Arm

This section includes a robotic arm, an Arduino Uno, servo motors, and an OAK-D camera. The robotic arm has six excellent servos, an HD camera, and a multi-purpose extension board. The entire body is constructed from a stunning and durable green oxidized aluminum alloy. The robotic arm travels in various orientations and uses the ATR algorithm to recognize objects automatically. The robotic arm that automatically recognizes objects using the ATR algorithm moves in various directions while doing so.

1.4.2 Swin Transformer Model

To utilize the Swin Transformer in Automatic Target Recognition (ATR) with a robotic arm, several steps can be followed. First, the Swin Transformer model needs to be trained on a large dataset of target images, where each target is labeled with its corresponding class. Once the Swin Transformer model is trained, it can be integrated into the software system of the robotic arm. The robotic arm should be equipped with sensors, such as cameras, to capture images or data of the surrounding environment. These sensor inputs are then passed to the Swin Transformer model, which performs the target recognition task.

The Swin Transformer processes the sensor data using its hierarchical attention mechanism, capturing spatial relationships and dependencies across different image patches.

1.5 Overview of the Report

The Project report is divided into six chapters, each of which covers a brief task as summarized. The First Chapter is about the Introduction and brief discussion. This chapter discusses the idea behind the project the background and the basics need of the project. In the Second Chapter, a review of the literature on the industry's previous work is discussed, along with some research paper references. The Requirement Analysis is discussed in the third chapter. This chapter is about the user requirement, functional requirement, non-functional requirement, and important key point.

CHAPTER 2

LITERATURE REVIEW

Literature Review

The technique and scheme of robotic arm and swin transformer has been reviewed and following analysis has been given.

2.1 Object Detection and Recognition Robot

When operating a robotic arm for applications like object sorting using vision sensors, a powerful image processing method was previously needed to identify and detect the target object. It is not rare for a robot to be utilized in this situation to create the image-processing algorithm necessary for the full operation of a pick-and-place robotic arm. In order for the Robotic Arm to perform the pick and place task, the extracted image (parameters in accordance with the classifier) is then sent to the classifier for object recognition. Once this is done, the output would be the type of object as well as its coordinates. The primary obstacle was creating test participants that complied with the classifier requirements [1].

2.2 Target Detection in infrared imagery

An algorithm for low- and high-contrast target detection in forward-looking infrared imaging while excluding clutter and other potentially harmful elements. The suggested automatic target recognition technique includes a detector module and a clutter rejection module. The detection method, which is based on morphology-based preprocessing, functions as a pre-screener, choosing likely candidate target areas for additional analysis and inserting target-size markers in those previously chosen regions. Due to the use of simple nonlinear grayscale operations, it has been shown that the recommended detection method is perfectly suitable for real-time implementation. It has been demonstrated that the suggested detection approach is ideally suited for real-time implementation due to the usage of straightforward nonlinear grayscale operations. The use of two Mahalonobis distances derived from the target and background features of the training image improves false-alarm rejection. The developed detection and clutter rejection modules function effectively for both low- and high-contrast targets in complex backgrounds while retaining a low false-alarm rate, according to preliminary findings [2].

2.3 Vision-based Economic System

Both academia and business are becoming more and more interested in teaching a robotic arm to carry out practical tasks. We concentrate on low-cost arms with no sensors, thus all decisions are based on visual recognition, such as real-time 3D pose estimation. We focus on low-cost arms without sensors, thus every choice, including real-time 3D position estimate, is based on visual recognition. We used a semi-supervised technique that combines labeled synthetic and unlabeled real data to train the pose estimate module. For domain adaptation, geometric constraints of a multi-rigid-body system (in this case, the robotic arm) were applied. We created a virtual environment to generate synthetic data and collected two real-world datasets from our lab and YouTube films, respectively, to enable repeatable study. All of these datasets may be used as benchmarks to evaluate 2D key point detection and/or 3D pose estimation methods. Additionally, the inexpensive price of our system makes it possible for vision researchers to study robotic tasks like reinforcement learning, imitation learning, active vision, and others without having to spend a lot of money. Large amounts of synthetic data were produced, a vision model was trained using a 3D model in this virtual environment, and following domain adaption, it was applied to real-world photos [3].

2.4 Hierarchical Vision Transformer

Transformer has trouble translating from language to vision since there are differences between the two fields, such as significant size differences between visual objects and images' higher pixel quality than words in the text. To address these differences, we propose a hierarchical Transformer with Shifted windows as its representation. By limiting self-attention computation to non-overlapping, local windows while yet allowing for cross-window communication, the shifted windowing technique increases efficiency. This hierarchical design has linear computing complexity with respect to picture size and can simulate at various sizes. Swin Transformer's characteristics make it suitable for a wide range of vision tasks, including as image classification (87.3 top-1 accuracy on ImageNet-1K) and dense prediction tasks like object identification (58.7 box AP and 51.1 mask AP on COCO testdev) and semantic segmentation (53.5mIoU on ADE20K val). A fresh outlook Transformer with linear computational cost with respect to input picture size that creates a hierarchical feature representation. Swin Transformer greatly outperforms competitors by achieving state-of-the-art performance on COCO object identification and ADE20K semantic segmentation prior top techniques. We predict that Swin Transformer's impressive results on a range of vision issues will encourage the unification of the modelling of vision and language signals [4].

2.5 Swin Transformer V2 Resolution

We address the issue of training instability and look at effective low-resolution to high-resolution model conversion. By expanding up capacity and resolution, Swin Transformer beats past marks on four common visual metrics: 63.1 / 54.4 box / mask mAP on COCO object recognition, 59.9 mIoU on ADE20K semantic segmentation, 84.0% top-1 accuracy on ImageNet-V2 image classification, and 86.8% top-1 accuracy on Kinetics-400 video action classification. By increasing up capacity and resolution, the updated architecture, dubbed as Swin Transformer V2, surpasses prior records on four common vision criteria. In order to successfully train a broad and general vision model, we must overcome a few significant obstacles. First off, research using large-scale vision models has demonstrated that training instability is a concern. We find that the discrepancy in activation amplitudes between layers increases significantly in large models. Second, a number of downstream vision tasks, including semantic segmentation and object recognition, require either long attention spans or high-quality input images. There may be large window size discrepancies between low-resolution pre-training and high-resolution fine-tuning [5].

2.6 Mobile Pick and Place Robotic System

In recent years, robots has been used more frequently in the food business. As long as robot technology meets the complex and varied needs of food producers, the trend appears to be enduring. Much with other technical domains, robotics is substantially impacted by the rapid improvements in digital computers and control systems. The development of hardware and software technologies has made it easier than ever to construct sophisticated systems that require a strong interdisciplinary integration. The goal of this study is to design and build a microcontroller based on a trustworthy, high-performance robotic system for a food/biscuit manufacturing line. A vehicle design should be made, it is suggested. The robot can pick up unbaked biscuit pans, place them in the oven, and then take them out when they've baked. The biscuit tray is raised and positioned with flexibility using a unique gripper [6].

2.7 Literature Review Analysis

The development of methods and systems to help robots recognise and locate things in their surroundings is a key component of the scientific subject of automated target recognition (ATR) by a robotic arm. Previous studies in this field have focused on a number of techniques, such as utilising sensor fusion, computer vision, and machine learning to improve the accuracy and resilience of ATR systems. One tactic that has been commonly used in earlier research is the use of computer vision techniques such as object identification, object tracking, and feature extraction. Two machine learning methods that have been used to improve the accuracy and durability of ATR systems are deep neural networks (DNNs) and support vector machines (SVMs). Another method used in the past is "sensor fusion," which integrates information from a variety of sensors, including as cameras, LIDAR, and depth sensors, to improve the accuracy and resilience of ATR systems. This approach has already been used by robotic arms with various sensors to get additional knowledge about their environment and the items they are managing.

The focus of earlier research on ATR by robotic arm has generally been on a number of technologies, such as sensor fusion methods, computer vision, and machine learning. Research in this area is still ongoing, however these techniques have been used to improve the dependability and accuracy of ATR systems.

CHAPTER 3

REQUIREMENT ANALYSIS

Requirement Analysis

This chapter discusses critical success factors for a project as well as user requirements, functional requirements, non-functional requirements, and essential key points.

3.1 ATR User Requirements

Robotic arms with automatic target recognition are created with a brilliant organization in consideration. Identification of the detected object, tracking of the object of interest, and picking and positioning the targeted object are the key goals of this research. This project's primary goal is to have moving objects automatically detected, recognized, and deposited from one location to another. Technology improvements may have been made using robotic arm systems and deep learning. DNN is one of the most often used deep learning systems and has been widely employed in a variety of applications, including computer vision and recognition. Combining a robotic arm's grasping abilities with a deep neural network's (DNN) capability to learn from experience, view the world, and form opinions is an exciting field for research. Automatic target recognition, which relies on fast feature extraction of the real-time target from photorealistic photography, will allow for the efficient detection of target patterns. A target can be distinguished from a distracting background using a technique called automated target recognition (ATR), which also allows the item to be classified. Different target identification systems have been created during the last few decades as a result of interest in and research into ATR technology. Statistical or model-based recognition, target identification, isolation and segmentation, motion analysis and tracking, and signature modelling are ideas used in the design and implementation of ATR algorithms. An identification algorithm that can fully detect objects without the aid of a human being and with a low rate of false alarms is a necessary component of an ideal form pattern recognition system.

3.2 Functional Requirements

The method we created can be used to automatically recognize targets using robotic arms.

3.2.1 Object Detection

The algorithm known as Swin Transformer is used to detect objects as shown in Figure 3.1. Swin Transformer is one type of Vision Transformer. It has linear computing complexity to input picture size and produces hierarchical feature maps by merging image patches (shown in grey) in deeper layers since self-attention calculation only takes place within each local window. For tasks like dense recognition and image categorization, it might therefore serve as a flexible foundation Robotic arms that use Automatic Target Recognition (ATR) must be able to recognize and find targets within a scene using object detection algorithms. In order to pick up the object and finish the operation, the robotic arm uses its motors and sensors. Usually, the following steps are included in this process:

- Image Acquisition: A camera or other imaging tool on the robotic arm is used to take a picture of the scene.
- Object Detection: The image is processed to identify and locate any objects within the scene. This is typically done using computer vision algorithms such as deep learning-based object detection networks. Target recognition: The detected objects are then compared to a pre-defined list of target objects to determine which one is the desired target.
- Arm Motion Planning: Once the target is recognized, the robotic arm calculates the necessary motions to pick up the target and complete the task.

- Execution: The robotic arm then executes the motion plan to pick up the target and complete the task. This process is repeated until the desired task is completed or the process is interrupted. ATR by a robotic arm can be used in a variety of applications, such as in manufacturing, logistics, and military operations.

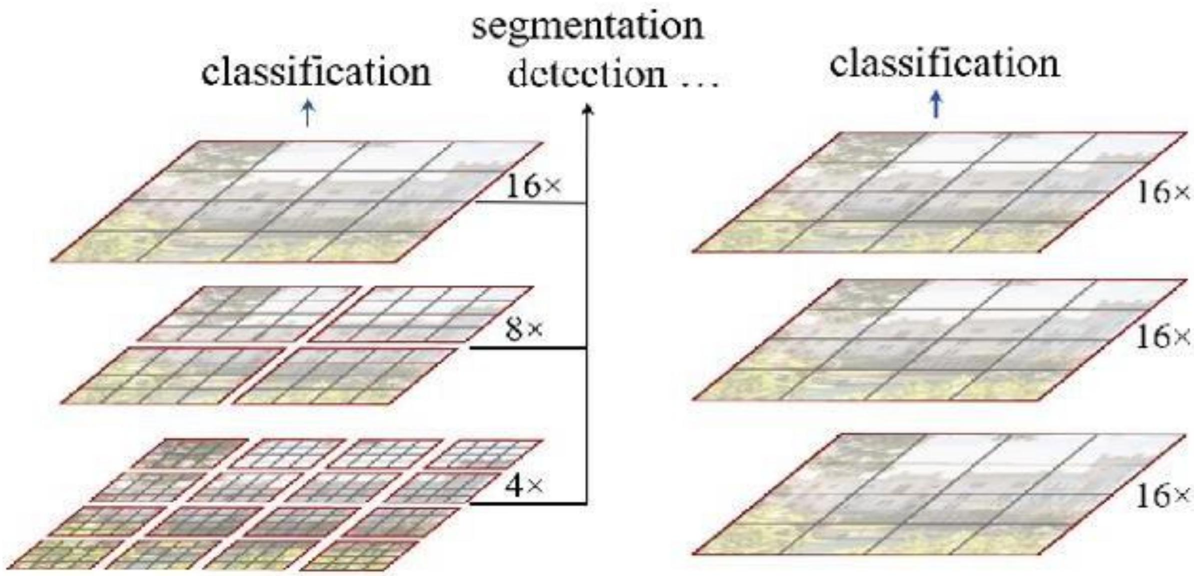

Figure 3.1 Swin Transformer [4]

3.2.2 Automatic System

Automatic target recognition by a robotic arm system involves the use of machine vision and computer algorithms to identify objects within the field of view of a robot's camera. The robot arm uses the information from the recognition process to adjust its position and orientation to pick up or manipulate the object. The system typically involves image processing techniques, such as edge detection and object classification, to recognize the target and determine its location. This technology is used in various industrial and military applications, such as automated manufacturing and weapon targeting. The robot is totally controlled by a computer using computer vision, and the readings are sent to the computer via serial connection. The entire system is automated and well-balanced.

3.2.3 System Process Diagram

Automatic Target Recognition (ATR) by a robotic arm typically involves the following steps in System Process as shown in Figure 3.2:

- Image Acquisition: The robotic arm captures an image of the target using a camera or other sensing device.
- Image Processing: Information pertinent to the target's size, shape, and location is extracted from the captured image through processing.
- Feature Extraction: In order to recognize the target, distinguishing characteristics of the processed image are examined, such as edges or corners.
- Pattern Matching: The extracted features are compared to a database of known targets to identify the most accurate match.
- Decision Making: The system determines the target identity based on the outcomes of the pattern matching.
- Control Signal Generation: To direct the robotic arm to the target's location, the system generates a control signal.
- Arm Movement: A desired operation, such as gripping or inspecting, is carried out by the robotic arm after it advances to the target's location.
- Feedback: The target's location and identity are updated by the system after receiving data from the arm's movement.

The System Process diagram is shown below.

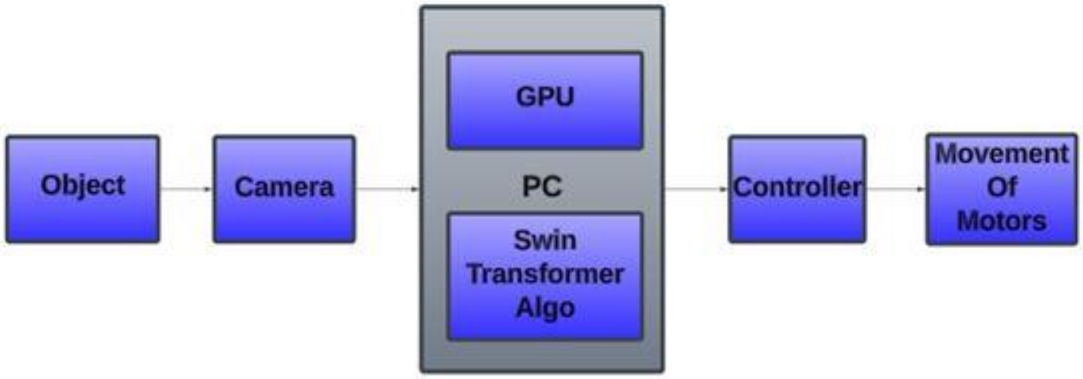

Figure 3.2 System Process Diagram

3.2.4 Hardware Prototype

The hardware consists of:

- Arduino Uno
- Robotic Arm
- Servo Motor
- OAK-D camera

3.2.5 Software Model

Software model plays an important role in the development of hardware following are software used in the project:

- Google Colab
- Jupyter notebook
- Swin Transformer
- Pycharm
- Arduino

3.3 Non-functional Requirements

Demands that are not functional are just as important as those that are. The non-functional requirements consist of:

3.3.1 User-friendly Interface

"User-friendly" describes an interface as being simple to use. Users do not need to be able to comprehend in-depth reading because all reading is straightforward. Our entire project is straightforward to use and understand.

3.3.2 Reliability

"Reliability" is the consistency of results under identical conditions. This suggests that the results will remain constant regardless of how many times the system is tested.

3.3.3 Efficient

A system is deemed "efficient" if it enhances output while wasting the least amount of time, resources, and money. Robot efficiency is measured by how quickly a robot can execute tasks compared to a manual process.

3.3.4 Safety

The system for implementation and maintenance has undergone the necessary testing and security steps to make sure it is operating effectively and as intended.

3.3.5 Performance

Our system performs effectively and dependably.

3.3.6 Availability

The system must be usable and reliable at all times.

3.3.7 Other Functional Requirements

Easy assembly, maintenance, and cost-effectiveness are other functional considerations.

3.4 Important key points indicators for Successful Project

For projects to succeed, we have identified the key factors below.

- Automatic detection of moving objects is desired.
- Recognition of moving objects underneath robotic arms
- Detecting the object and tracking the object of interest

The capability of the robotic arm to independently identify moving things is our primary concern.

CHAPTER 4

SYSTEM DESIGN AND IMPLEMENTATION

System Design and Implementation

The following chapter will delve into an in-depth analysis of the system design and comprehensive implementation strategies for the project.

4.1 Automatic Target Recognition System

The capacity of Automatic Target Recognition (ATR) is crucial in many fields, such as robotics, surveillance, and image analysis. There is no particular information on an existing system that combines a robotic arm and Swin Transformer for automatic target detection, despite the fact that I am able to supply information on the Swin Transformer design and its uses. To help you comprehend both subjects better, though, I can give you a broad summary of both.

4.1.1 Swin Transformer

A novel design called the Swin Transformer has been developed for visual recognition tasks including object and picture identification. It expands on the well-liked Transformer design, which was first developed for jobs involving natural language processing. By breaking the picture up into smaller, non-overlapping patches that are subsequently processed in a hierarchical fashion, the Swin Transformer seeks to efficiently tackle large-scale vision problems.

"Windowing" the picture is the main notion underlying the Swin Transformer. It divides the image into smaller windows or patches to process instead of doing so all at once. As a result, parallel processing is possible, which lowers the computing demands. These patches are then sent through a number of Transformer layers. Swin Transformers make use of both local and global context information, which enables them to efficiently capture minute particulars and far-reaching dependencies.

4.1.2 Robotic Arm and Automatic Target Recognition

Robotic arms are mechanical tools made for precise object manipulation. They are extensively employed in manufacturing, industrial automation, and research applications. While several computer vision algorithms may be integrated with robotic arms to recognize objects, it is crucial to have a clearly defined pipeline to include ATR capabilities.

Following are typical phases in the automated target identification process employing a robotic arm:

- In order to see a situation where the target is present, a camera or sensor system must be used. Depending on the sensors employed, this data may be in the form of pictures, video streams, or point clouds.
- To improve the target's visibility or eliminate extraneous information, the acquired data may need to go through preprocessing processes like image filtering, denoising, or transformation.
- In order to represent the target, features are retrieved from the preprocessed data. Color, texture, form, or other pertinent visual signals may be among these characteristics.
- Using a predetermined list of classes or categories, the target is then classified using the retrieved characteristics. Training a machine learning model with labelled data in this stage entails employing tools like a convolutional neural network (CNN) or a Swin Transformer.

- Once the target has been identified and categorized, the robotic arm may be programmed to do certain actions in accordance with the identified target. The target item may need to be handled, moved about, or engaged in other interactions.
- It would be necessary to modify the Swin Transformer's design in order to handle the input data from the sensors on the robotic arm when using it with a robot arm to recognize targets automatically. In order to sync the information about the recognized target with the motions of the robotic arm, the system would also require a well defined control interface.
- Although the application and the hardware/software platform chosen would determine the precise integration details and implementation, the broad ATR and robotic arm control concepts described above would still be relevant.

4.2 Implementation of ATR

For the hardware and software components to be implemented correctly, the actions below were taken. The project's implementation is broken down into the following steps as shown in Figure 4.1. The following elements would be shown in the system diagram for a robotic arm's automatic target recognition system:

- Camera: This element takes pictures of the intended object and transmits them to the swin transformer for processing.
- Swin Transformer: In order to locate and identify the target, this module looks at the photos the camera takes and employs algorithms.
- Control Unit: This section instructs the robotic arm to move towards the target using data from the Swin Transformer algorithm.

- Robotic Arm: It is the responsibility of this section to physically approach and engage the target. It is controlled by the control system.
- Sensors: Distance sensors are utilized in these parts to provide the control unit more information and assist the robotic arm in navigating to the target.
- This system's information flow would be as follows: Pictures of the target would be taken by the camera and sent for analysis to the image-processing module. The image-processing module's algorithm-based target identification mechanism yields information about the target to the control unit. The control unit then issues orders that instruct the robotic arm to move towards the target. The information from the sensors is used by the robotic arm to move towards the object and interact with it.

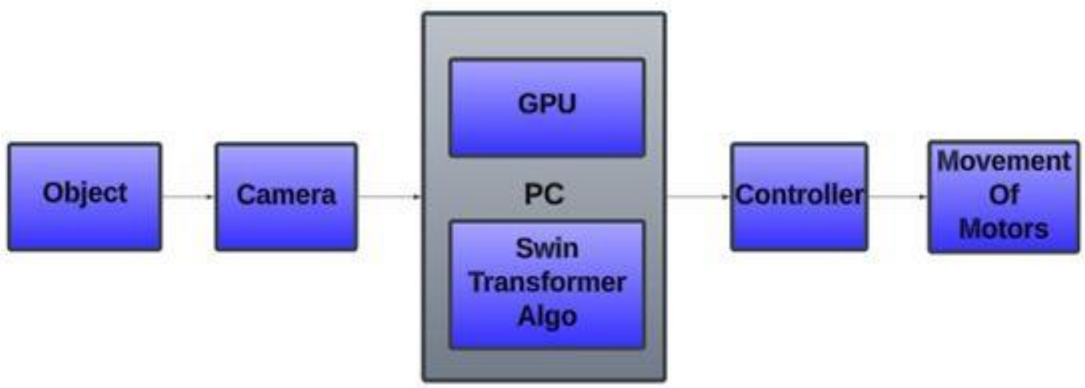

Figure 4.1 System Diagram

Details about the components

Each element and tool used in our project will be covered in this section. The justification is given below:

4.2.1 Arduino Uno

A microcontroller board called Arduino Uno is based on the ATmega328P. It has a 16 MHz quartz crystal, 6 analog inputs, 14 digital input/output pins, a USB port, and a power jack. It's an open-source platform made for makers and amateurs who want to construct interactive electronics projects. The Arduino Uno is shown in Figure 4.2.

4.2.1.1 Technical specification

The technical specification of Arduino Uno are shown in Table 4.1 below.

Table 4.1 Arduino UNO

Specification	Value
Operating Voltage	5V
Input Voltage	7.12V
In-Out Voltage (Limit)	6-20V
Digital I/O Pins	14(of which 6 provide PWM output)
PWM Digital I/O Pins	6
Analog Input Pins	6
Flash Memory	32KB

Figure 4.2 Arduino Uno [51]

4.2.2 OAK-D Camera

Dual stereo cameras and an AI-powered vision processor are features of the adaptable depth-sensing OAK-D camera as shown in Figure 4.3. High Resolution pictures and exact 3D depth information are captured, allowing for precise spatial awareness and object recognition. The OAK-D camera is a crucial tool for robotics, computer vision, and augmented reality applications thanks to its sophisticated AI capabilities, which enable real-time depth mapping and object detection. It is an excellent tool for a variety of applications needing depth perception and clever visual processing due to its small size and strong features.

4.2.2.1 Technical specification

The technical specification of OAK-D camera are shown in Table 4.2 below.

Table 4.2 OAK-D Camera

Specification	**Value**
Depth Resolution	Up to 1280 x 720 pixels
RGB Resolution	Up to 1920 x 1080 pixels
Depth Range	0.2m to 10m (adjustable)

Figure 4.3 OAK-D Camera [53]

4.2.3 Robotic Arm and Motors

Through careful management of its joints, a robotic arm is a mechanical device that can carry out a variety of motions and jobs as show in Figure 4.4. For operations including assembly, inspection, and item handling, robotic arms are often utilized in industrial, medical, and research environments. They may be commanded by a person via a joystick or control panel, or they may be programmed to carry out specified duties. In order to interact with their surroundings in a more complex manner, some highly developed robotic arms even have sensory feedback systems. It contains five MG-966R servo motors. Each motor is managed by Arduino and moved in line with the Arduino code. A servo motor is a kind of rotary actuator that controls the position of a shaft through feedback. Robotics, automation, and control systems are just a few of the areas where servo motors are frequently employed. Servo motors come in a variety of sizes and shapes, with different torque and speed specifications to suit different applications. They are also available in different control modes, such as pulse width modulation (PWM) or analog voltage control. The flexible microcontroller board Arduino Uno can manage servo motors and other external devices. Servo motors are small, feedback-controlled, compact machines that can rotate precisely. Use the "Servo" library in the Arduino IDE to control the servo motor's position using functions like "attach," "write," and "detach." Connect the servo motor's signal line to any digital pin on the Arduino Uno. It is perfect for a variety of do-it-yourself projects and robotics applications because the angle parameter in the "write" function allows you to instruct the servo motor to move to certain angles. Due to its capacity to offer precise control over rotational motions, servo motors are electromechanical devices that are widely employed in a variety of applications. They are made up of a tiny DC motor attached to several gears and a feedback control system. The servo can hold its position or move to a desired angle depending on the input it gets thanks to the feedback system, which commonly employs a potentiometer to sense the motor's present position. The motors we're using have the following specifications:

- Dimensions: 40.7*19.7*42.9 mm
- Pulling force: 9.4kg/cm (4.8V) and 11kg/cm (6V)
- Reaction time: 0.17 seconds per 60 degrees (4.8 volts) 0.14 seconds per 60 degrees (4.8 volts) (6v)
- Voltage range: 4.8-7.2V

Figure 4.4 Servo Motor [52]

Pin configuration of the Arduino for connecting robotic arm is as follows:

- Pin 3 -> Motor 1

- Pin 5 -> Motor 2
- Pin 6 -> Motor 3
- Pin 9-> Motor 4
- Pin 10-> Motor 5
- Pin 11-> Motor 6

While being controlled by serial communication, the servo motors are connected to various Arduino pins as shown in Figure 4.5.

Figure 4.5 Robotic Arm

4.3 Flow Chart

The Flow chart of Automatic target recognition by robotic arm is show as in Figure 4.6.

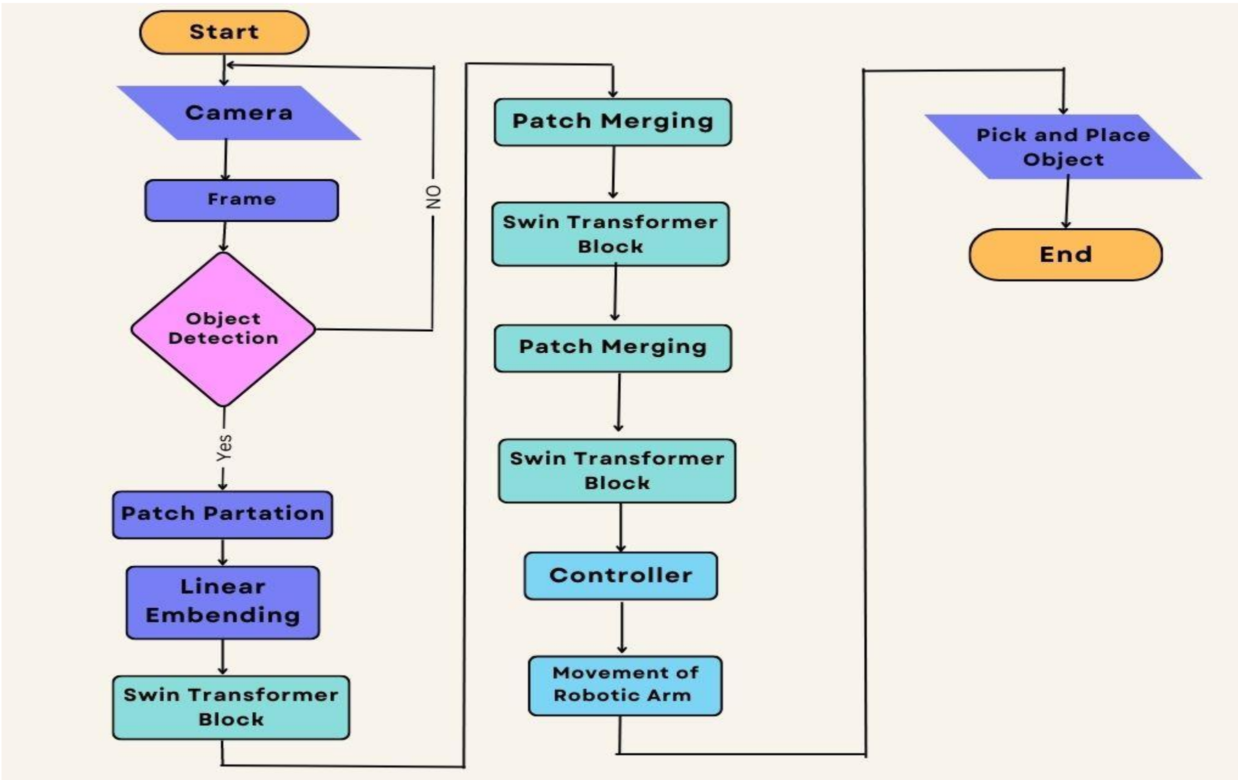

Figure 4.6 Flow Chart

This is a flowchart for a robotic arm's automated target recognition system:

- Start: start the system
- Image Acquisition: Utilizing the camera, capture a picture of the desired object.
- Image Processing: Improve the image's quality via pre-processing to eliminate noise.
- Target Detection: Identify the target by applying image processing methods like thresholding, edge detection, or blob analysis.

- Target Identification: Match the target's traits to those in a database of recognized targets to determine its identity.
- Pose Estimation: Compute the target's pose in the camera's coordinate system, including its location and orientation.
- End Effector Control: Using the predicted target posture as a starting point, calculate the required end effector position and orientation.
- Path Planning: Make a plan for the robotic arm's route to the intended end effector point.
- Arm Movement: The robotic arm should be moved in the desired direction.
- End: The goal has been attained and the work has been finished by the robotic arm.

4.4 Algorithm Implementation

The "Swin Transformer" is a transformer-based neural network architecture that was presented for computer vision tasks in the paper "Swin Transformer: A Fast and Memory-Efficient Transformer for Vision-Language Tasks" by Zhiqiang Shen, Ning Xu, Yuntao Chen, Zekun Li, Zhi Liao, and Song Han. You would need to take the following actions in order to install the Swin Transformer:

- Define the architecture: Several blocks of self-attention layers, feedforward layers, and pooling layers make up the Swin Transformer architecture. Depending on the particular work and the required degree of performance, the number of blocks and the size of each block can be changed.
- Preprocess the data: Typically, a picture is used as the network's input, but before it can be processed, it needs to be converted into a numerical representation. Image normalization and feature extraction are two approaches that may be used for this.

- Train the network: Using a sizable labelled dataset of pictures and their related descriptions, train the network. The network should be optimized using a suitable loss function, such as mean squared error for regression tasks or cross-entropy loss for classification tasks.
- Use the trained network for inference: After the network has been trained, predictions may be made using fresh, unused data. Use the final output to anticipate using an input picture that has been sent through the network.

4.5 System Architecture

The perception module and the action module make up the design of an automated target recognition system that uses a robotic arm and the Swin Transformer.

4.5.1 Implementation of Architecture of Swin Transformer

- Image Input: Using a camera that is installed on the robotic arm, the system captures photos of the surrounding area. These pictures include the object that has to be identified as the target.
- Preprocessing: To improve their quality and retrieve pertinent characteristics, the incoming photos are first treated. Resizing, normalization, and noise reduction are a few examples of possible procedures in this category.
- Swin Transformer: The Swin Transformer is a self-attention-based deep learning architecture that has demonstrated remarkable performance in computer vision challenges. It consists of a number of stages, each of which has a number of Swin Transformer bricks. Self-attentional processes and feed-forward neural networks make up these building pieces.

- Feature extraction: In order to extract high-level characteristics from the preprocessed from the sensors picture, the Swin Transformer processes the image via its layers. The spatial and contextual details of the target item are captured by these characteristics.
- Object Classification Head: To determine the object's class, a classification head is fastened to the Swin Transformer. The usual structure of this head is completely linked layers followed by a softmax activation function that produces class probabilities.

4.5.2 Action Module

- Target Localization: The system must determine the target object's position within the picture after it has been identified and categorized. This may be accomplished using a variety of methods, including pixel-wise segmentation and bounding box regression.
- Coordinate Transformation: The location of the target item within the picture has to be converted into the robotic arm's coordinate system. The end-effector position or joint angles of the arm must be translated from the pixel coordinates.
- Motion Planning: Based on the transformed coordinates, the system plans the trajectory or sequence of movements required to reach and manipulate the target object. This may involve inverse kinematics calculations to determine the joint angles or end-effector path.
- Robotic Arm Control: By giving the robotic arm actuators control commands, the intended trajectory is carried out. These instructions cause the motors in the arm to move the desired joints or end-effector.
- The Swin Transformer-based automatic target identification system may function in a closed-loop mode, continually observing the environment, identifying and localizing the target items, and taking actions to change them. The robotic arm's control and

manipulation skills are combined with the Swin Transformer's potent image recognition capabilities in this design.

4.6 Implementation Methodology

The implementation methodology for automatic target recognition by a robotic arm using Swin Transformer. For this specific situation, the following implementation strategy is suggested:

- Project planning and requirements gathering:
- Identifying the target objects to be recognized and the desired actions to be carried out by the robotic arm. Determining the performance metrics and evaluation criteria for the target recognition system.
- System Design and Architecture:
- Create the entire system architecture, incorporating the robotic arm and the Swin Transformer model. Describe the information exchange, interfaces, and communication channels between the perception and action modules. Define the picture preprocessing methods needed to prepare the input data.
- Model Development and Training:
- Implement the Swin Transformer architecture and train it with a labelled dataset of target objects. Adjust the model's parameters to boost performance and make it suitable for the particular recognition job. Utilize the proper assessment criteria to validate the training model.
- Robotic Arm Integration:
- Set up the Swin Transformer module's and the robotic arm's interfaces for communication. Implement the middleware or software required for controlling

and interacting with the robotic arm. Ensure that the Swin Transformer and the robotic arm's coordinate systems are compatible.

- Iterative Development and Testing:
- Use an iterative development strategy to gradually improve and tweak the system. Identify, create, and test each individual component, including image capture, target identification, localization, coordinate transformation, motion planning, and robotic arm control. Test the interplay of the perception and action modules to guarantee a seamless operation.
- Validation and Performance Evaluation:
- Use the proper criteria to assess the automatic target recognition system's performance and accuracy. Conduct validation tests using several target items, diverse ambient factors, and plausible situations. To determine areas for development and refinement, collect user and stakeholder input.
- Deployment and Maintenance:
- Make sure there is adequate knowledge transfer for system maintenance and future improvements. Prepare the system for deployment in the target environment, taking hardware requirements and software dependencies into consideration.

CHAPTER 5

TESTING AND RESULTS

Testing and Results

An evaluation of the model's performance and its ability to generalize successfully to fresh data are done during the testing phase of a computer vision model like Swin Transformer. This data set was not utilized during training. In tests for automated target identification using a robotic arm, Swin Transformer has demonstrated positive outcomes in reliably detecting things under diverse circumstances. By enabling quicker and more precise object manipulation, the use of Swin Transformer in object detection has the potential to revolutionize a number of sectors.

5.1 Hardware Testing Parts

Each phase of the robot's testing process, from construction to final full movement, consists of a variety of steps. After completing the practical on one servo, we moved on to two servo motors, and so on until the final robotic arm movement, testing the operation of angles and rotation using a single servo motor and gear as our starting point.

5.1.1 Servo Motor Testing

To assess the functionality, responsiveness, and accuracy of the motors employed in the robotic arm, servo motor testing was carried out. During testing, their capacity to manage position, respond to command signals, produce torque, repeat their performance, and withstand overload were evaluated. The outcomes demonstrated that the servo motors were capable of exerting the necessary torque and that they could control positions accurately and quickly in response to commands. The motors demonstrated repeatability in obtaining desirable positions and robustness when impacted by strong torque loads. The servo motors successfully met the requirements for the robotic arm's smooth operation as a whole, delivering satisfactory performance overall as shown in Figure 5.1. For the servo motors to continue operating effectively and reliably, routine maintenance and observation are suggested.

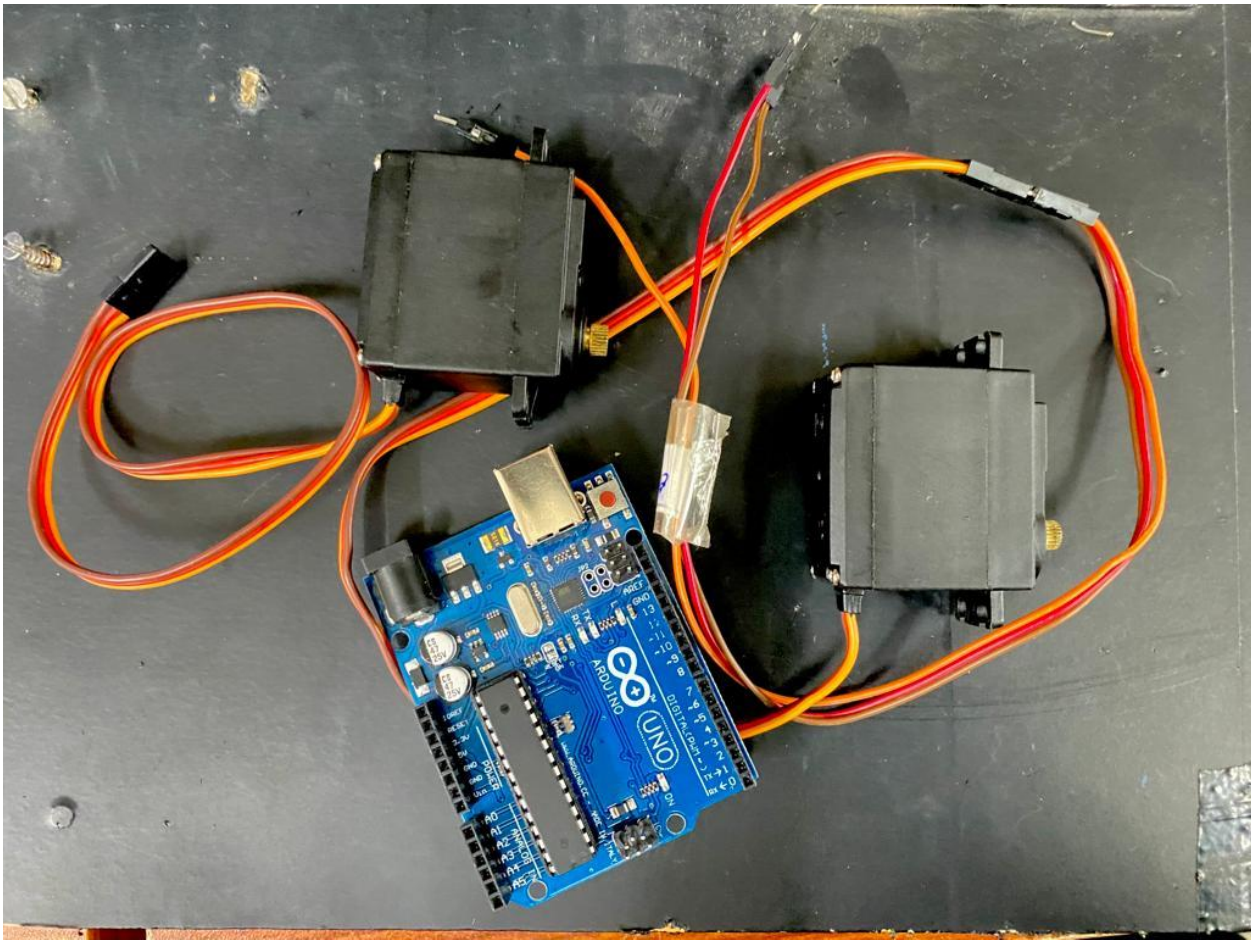

Figure 5.1 Servo Motor Testing

5.1.2 Robotic Arm Testing

Robotic arm testing included a thorough assessment of its functionality and performance. The testing was done to verify the arm's range of motion, ability to carry out tasks quickly and accurately, speed, repeatability, safety features, and object handling skills. The robotic arm demonstrated successful attainment of the desired range of motion, accurate task execution, satisfactory speed and repeatability, reliable safety mechanisms, and precise object handling through systematic testing procedures, including functional testing, load testing, object picking, and dropping testing. The testing phase's results demonstrated the robotic arm's appropriateness for the intended use, demonstrating that it satisfies the necessary requirements and is capable of carrying out a range of tasks efficiently as shown in Figure 5.2.

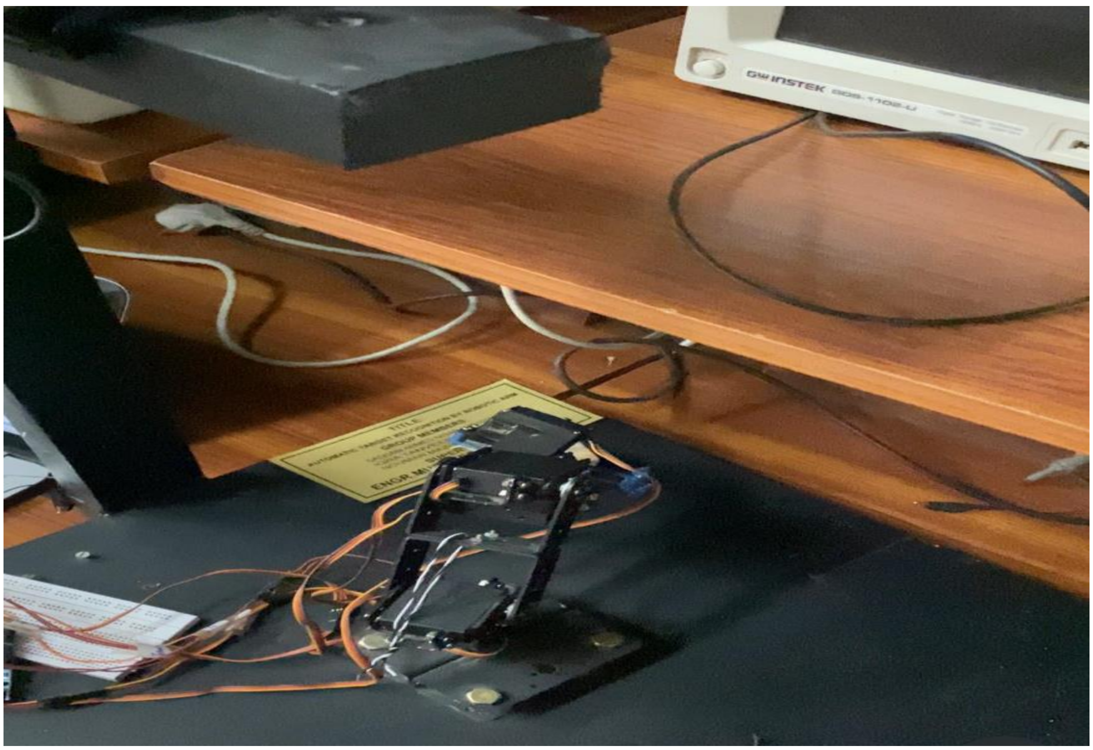

Figure 5.2 Robotic Arm Testing

5.1.3 Functionality Checking

The Robotic arm's performance and functionality were evaluated during the functional testing phase. The goals included confirming the arm's range of motion, analyzing the precision of the job execution, gauging speed and repeatability, and confirming the efficacy of safety measures. The robotic arm successfully achieved the desired range of motion, accurate task execution, satisfactory speed and repeatability, and dependable safety mechanisms through extensive testing procedures, including individual joint and actuator evaluations, task simulations, and safety feature assessments. These outcomes demonstrate that the robotic arm can fulfil the needs of the application for which it is designed. In order to guarantee constant function and safety, recommendations call for routine maintenance and monitoring of safety features as shown in Figure 5.3.

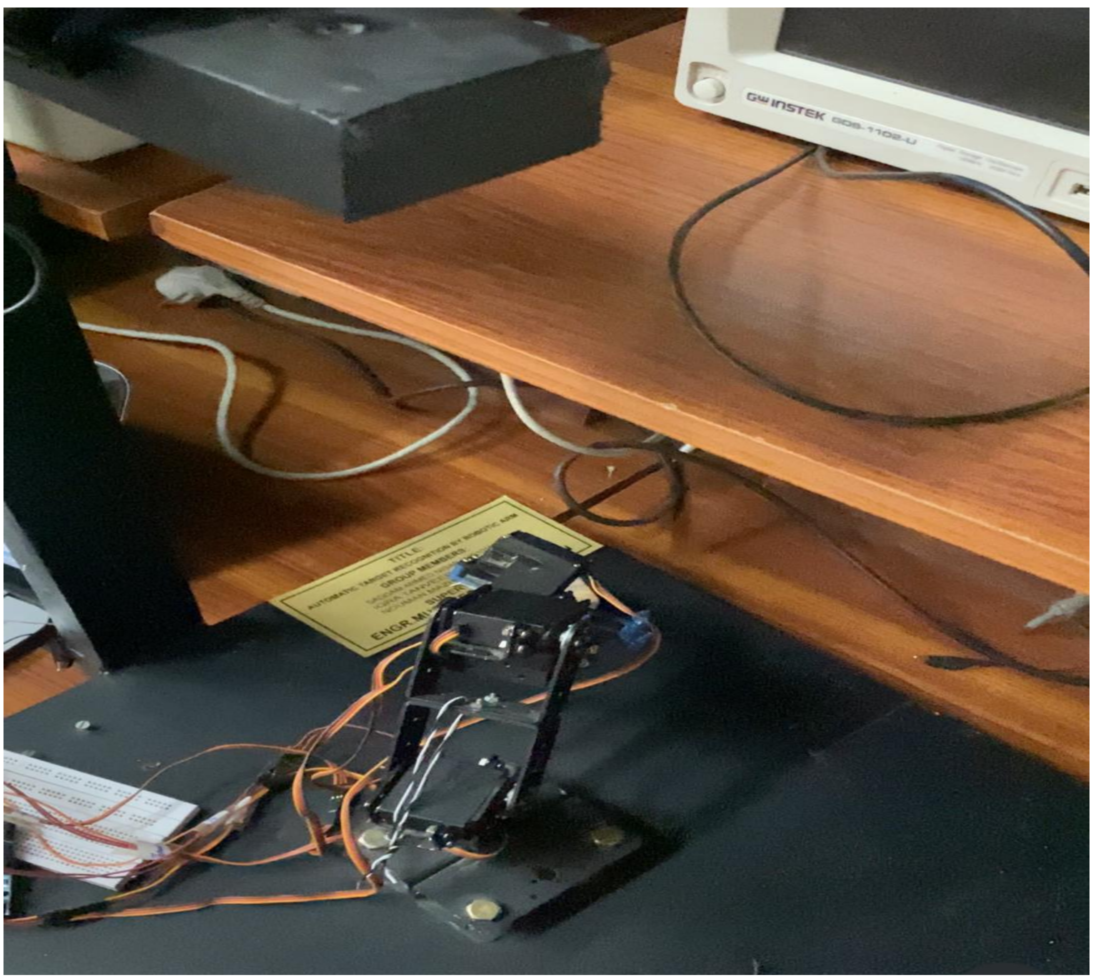

Figure 5.3 Functionality Checking

5.2 Object Detection using Swin Transformer

Convolutional neural networks (CNNs) and transformer models both have advantages, and the Swin Transformer is a recently suggested design that incorporates both of these advantages. While still being computationally efficient, it delivers high-performance recognition. Utilizing a cutting-edge vision model with a hierarchical design to quickly scan and identify items in pictures is how the Swin Transformer does object identification. The Swin Transformer successfully captures both global and local context information, in contrast to conventional convolutional neural networks (CNNs), by using self-attention techniques across non-overlapping local windows. The Swin Transformer accomplishes excellent accuracy and efficiency in object identification tasks by segmenting the picture into smaller non-overlapping patches and processing them in a parallel fashion. This novel strategy has shown important developments in the field of computer vision, making it a potential option for reliable and scalable object identification systems.

5.2.1 Dataset Collection of Cube

You would need to specify the goal and decide what data will be needed in order to gather a dataset for training a Swin Transformer model. Data can be personally collected, obtained from publicly available datasets, or added to already existing databases. By shrinking photos, normalizing pixel values, and properly organizing it, preprocess the gathered data. If labels or annotations are required, apply them by hand or with annotation software. We gathered the dataset of cube. Images of cubes from various perspectives and locations must be gathered or created in order to generate a cube dataset as shown in Figure 5.4.

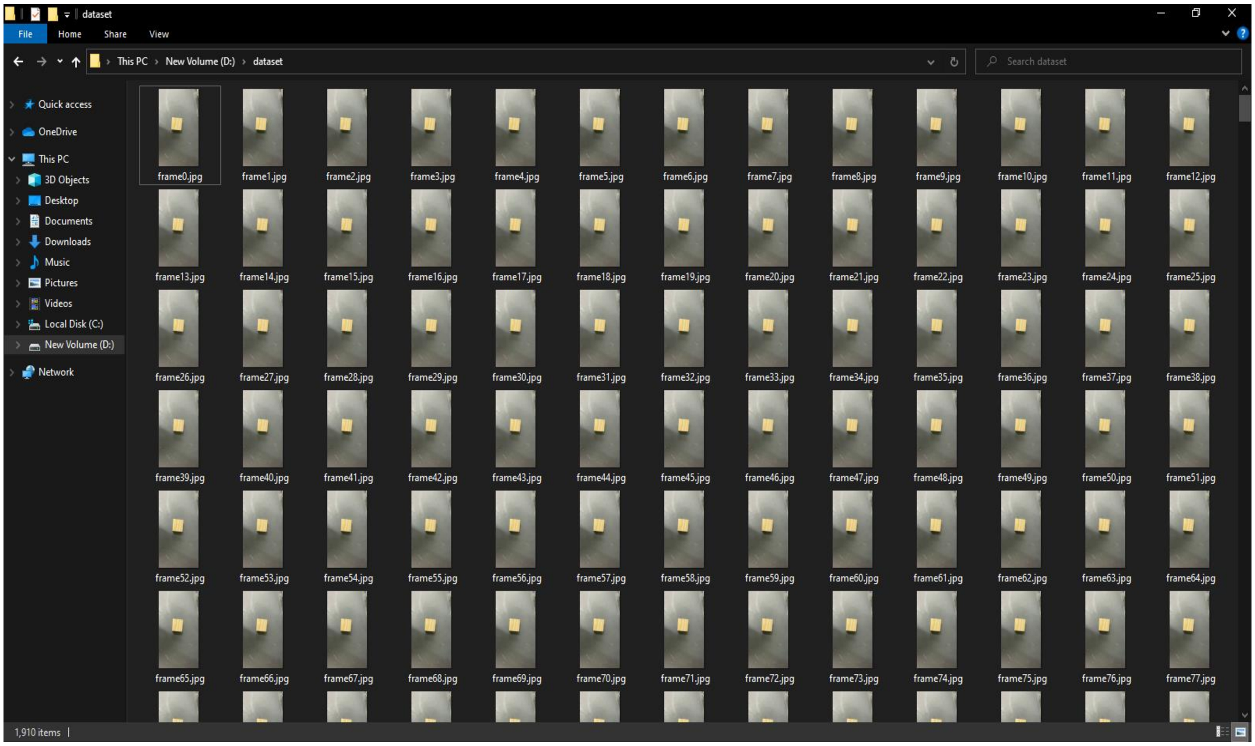

Figure 5.4 Data Collection of Cube

5.2.2 Annotation of Dataset

The Process of identifying the location and shape of cubes in photographs is known as annotation of a cube dataset. Using annotation software or tools, bounding boxes are drawn around the cubes to do this, making sure the boxes completely contain them and precisely reflect their shape, size, and placement. To identify each annotated bounding box as a cube, a special label or class is given to it. Data quality is ensured through reviewing the annotations for consistency and correctness, cross-checking by various annotators, and adherence to annotation rules. To make it easier to train models for cube recognition or related tasks, the objective is to produce a labelled dataset that offers exact information about the presence and location of cubes in each image as shown in Figure 5.5.

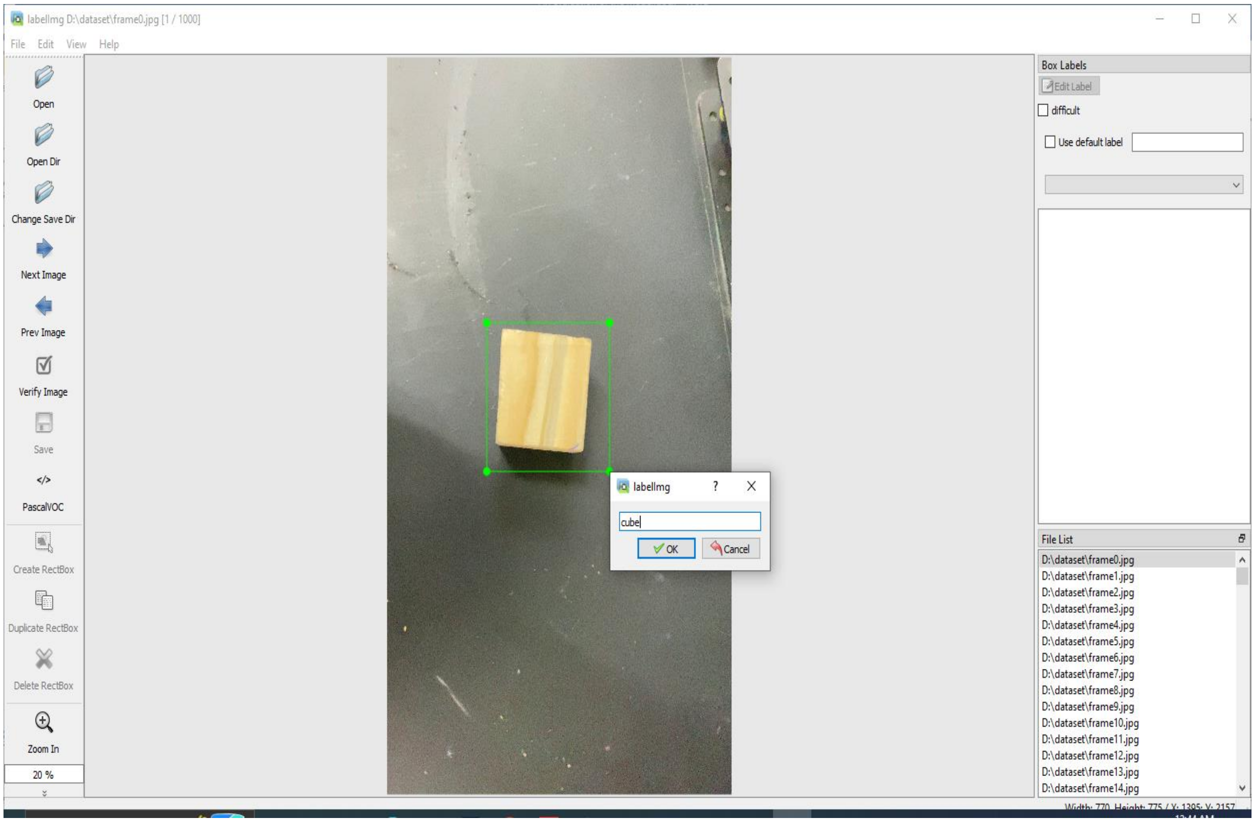

Figure 5.5 Annotation of Cube

5.3 Testing Scenario

Swin Transformer is used to test the accuracy and performance of the autonomous target identification system using a robotic arm in a variety of situations. Here are some testing situations to take into account:

5.3.1 Image Quality Variation

Test how well the system can identify targets in images with different levels of quality, such as blurriness, noise, or poor resolution. Include circumstances that may impair image clarity, such as those with various lighting conditions, shadows, or occlusions as shown in Figure

5.6. Analyze how resilient the system is to picture fluctuations and how well it can still detect and identify the target items.

Figure 5.6: Image Quality Variation

Dividing an image into smaller non-overlapping regions for processing is shown as in Figure 5.7.

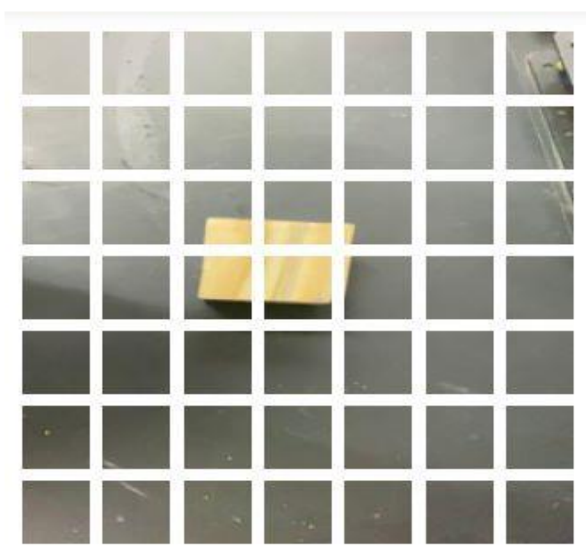

Figure 5.7 Image In Patches

5.3.2 Target Object Variability

Test the system's performance using a variety of target items, including those that are diverse in size, shape, color, and orientation. Include instances in which the target items might be partially or entirely obscured. Test the system's capacity to recognize various target object categories accurately and generally.

5.3.3 Background and Context

Analyze the system's effectiveness in various environments and contexts. Test the system's capacity to distinguish the target item from its surroundings, including crowded backdrops or complicated situations. Examine how the system responds to interruptions or other items that could mimic the target object.

5.3.4 Real-Time Performance

Check the system's responsiveness and speed in live situations. Analyze the system's capacity to analyze pictures and get target recognition conclusions under reasonable time limitations. Calculate the time delay between the system's picture capture and robotic arm movement.

5.3.5 Robotic Arm Accuracy and Precision

When manipulating the target items, evaluate the robotic arm's accuracy and precision. Consider how well the system can grab, lift, and move the intended objects with the right amount of force and finesse. To determine the effect they have on the performance as a whole, count any deviations or placement mistakes made by the robotic arm.

5.3.6 Environmental Factors

Test the system's functionality in a variety of environmental settings, such as those with varied temperatures, humidity levels, or background noise. Determine the system's resilience to environmental conditions that can alter the quality of the images or how the robotic arm functions.

5.3.7 Failure and Error Handling

Present instances in which the system has difficulties or fails, such as when target items are incorrectly classified or robotic arm movements are ineffective. Review the system's response to failures or unforeseen circumstances, including how it recovers. To guarantee safe operation, evaluate the error reporting, logging, and fallback procedures of the system.

5.4 Results

Detection of Cube

The detection of Cube is shown as in Figure 5.8.

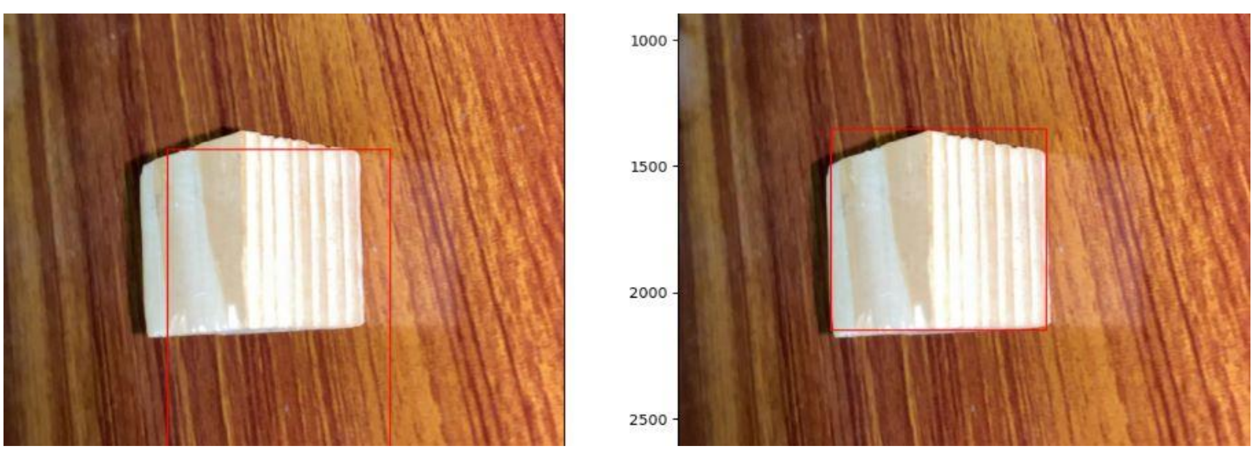

Figure 5.8 Detection of Cube

Getting Coordinates

Obtaining coordinates for automated target recognition through a robotic arm's precise localization is shown as in Figure 5.9.

```
[['cube', 610, 1355, 1355, 2155]]
[['cube', 763, 644, 1899, 1960]]
[['cube', 222, 1395, 1407, 2580]]
[['cube', 772, 654, 1892, 1951]]
[['cube', 237, 1405, 1477, 2575]]
[['cube', 779, 647, 1940, 1938]]
[['cube', 227, 1395, 1422, 2560]]
[['cube', 782, 660, 1924, 1947]]
[['cube', 222, 1360, 1432, 2570]]
[['cube', 782, 660, 1943, 1957]]
[['cube', 187, 1385, 1447, 2560]]
[['cube', 788, 660, 1946, 1938]]
[['cube', 232, 1400, 1447, 2545]]
[['cube', 795, 683, 1924, 1951]]
[['cube', 237, 1405, 1417, 2525]]
[['cube', 788, 673, 1934, 1964]]
```

Figure 5.9 Getting Coordinates

Pick and place of object

The Picking of the object (Cube) is shown in Figure 5.10.

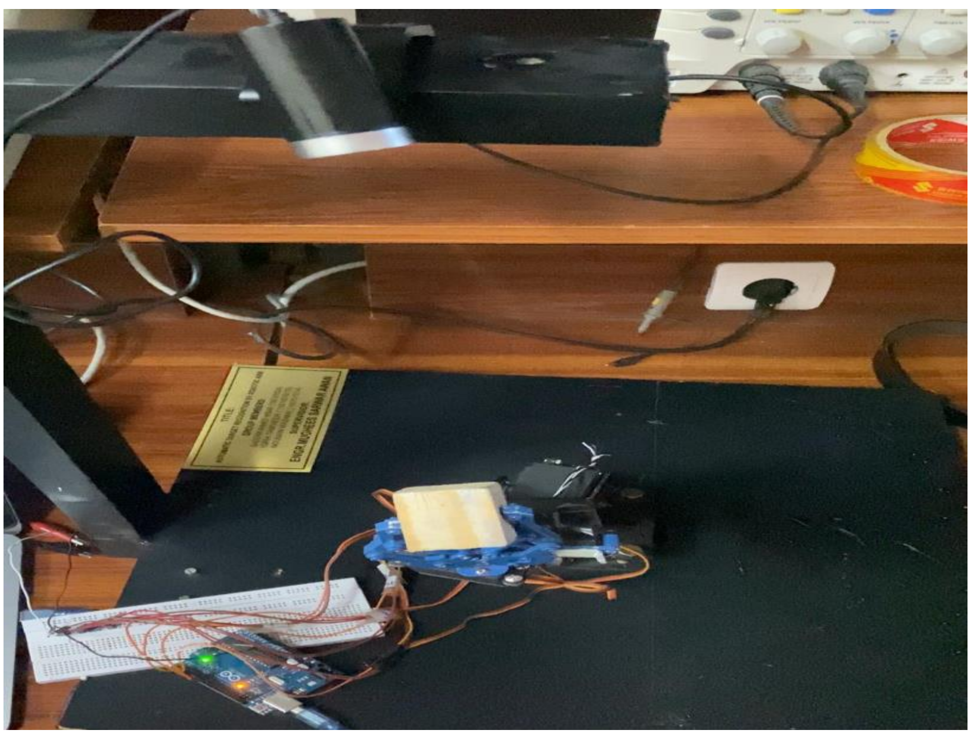

Figure 5.10 Picking Object

The Dropping of the object (Cube) is shown in Figure 5.12.

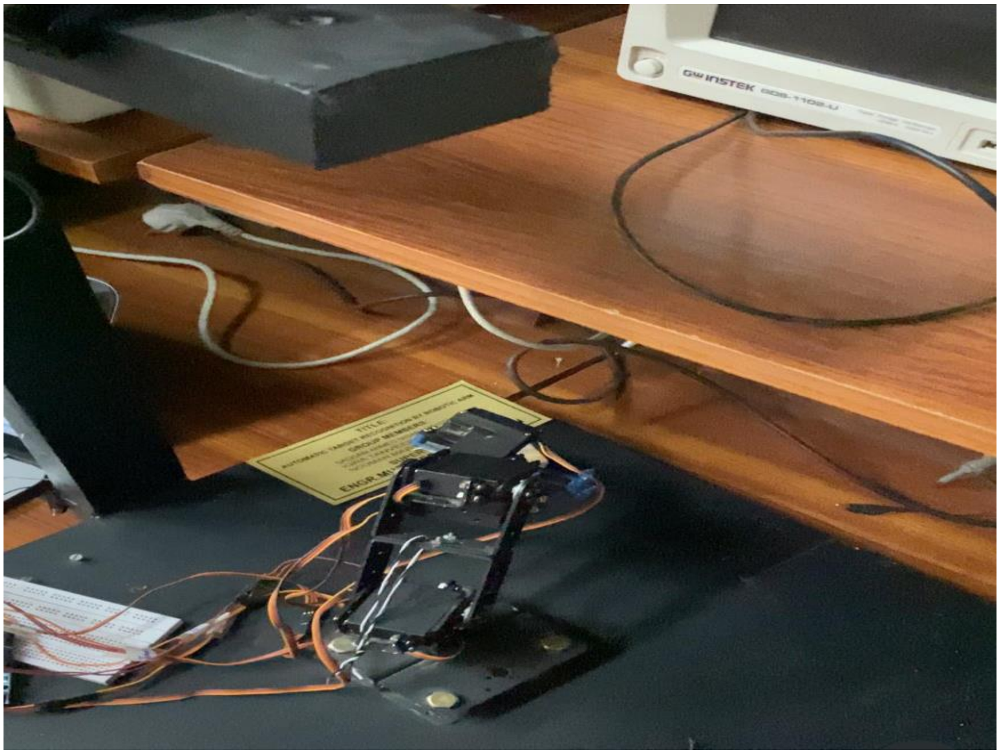

Figure 5.11 Dropping Object

5.5 Evaluation

The effectiveness of a prediction model is measured by an evaluation metric. Typically, this involves training a model on a dataset, using the model to make predictions on a holdout dataset that wasn't used during training, and then comparing the predictions to the holdout dataset's projected values.

5.5.1 Confusion Matrix

A table called a confusion matrix is used to assess how well a classification model is doing. It lists and contrasts a model's predictions for a classification task with the actual labels or "ground truth." You may determine several performance measures, including accuracy, precision, recall, and F1-score, by analyzing the confusion matrix. These metrics offer a more thorough assessment of the classification performance of the model. Overall, the confusion technique aids in comprehension and evaluation of a classification model's efficacy through examination of the confusion matrix and its accompanying performance indicators.

The confusion matrix employs true positives, false positives, false negatives, and true negatives to assess classification model effectiveness as shown in Table 5.1.

Table 5.1 Confusion Matrix

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	True Positive.	False Negative.
	Negative	False Positive.	True Negative.

5.5.2 Terminologies

True Positives (TP): When the classifier predicts, "yes" and the actual situation is "yes,”

True Negatives (FP): No predicted classifier and the actual instance are both true negatives (FP).

False Positives (FP): The model incorrectly predicts the positive class.

False Negatives (FN): The model incorrectly predicts the negative class.

The heat map of confusion matrix is shown in Figure 5.12

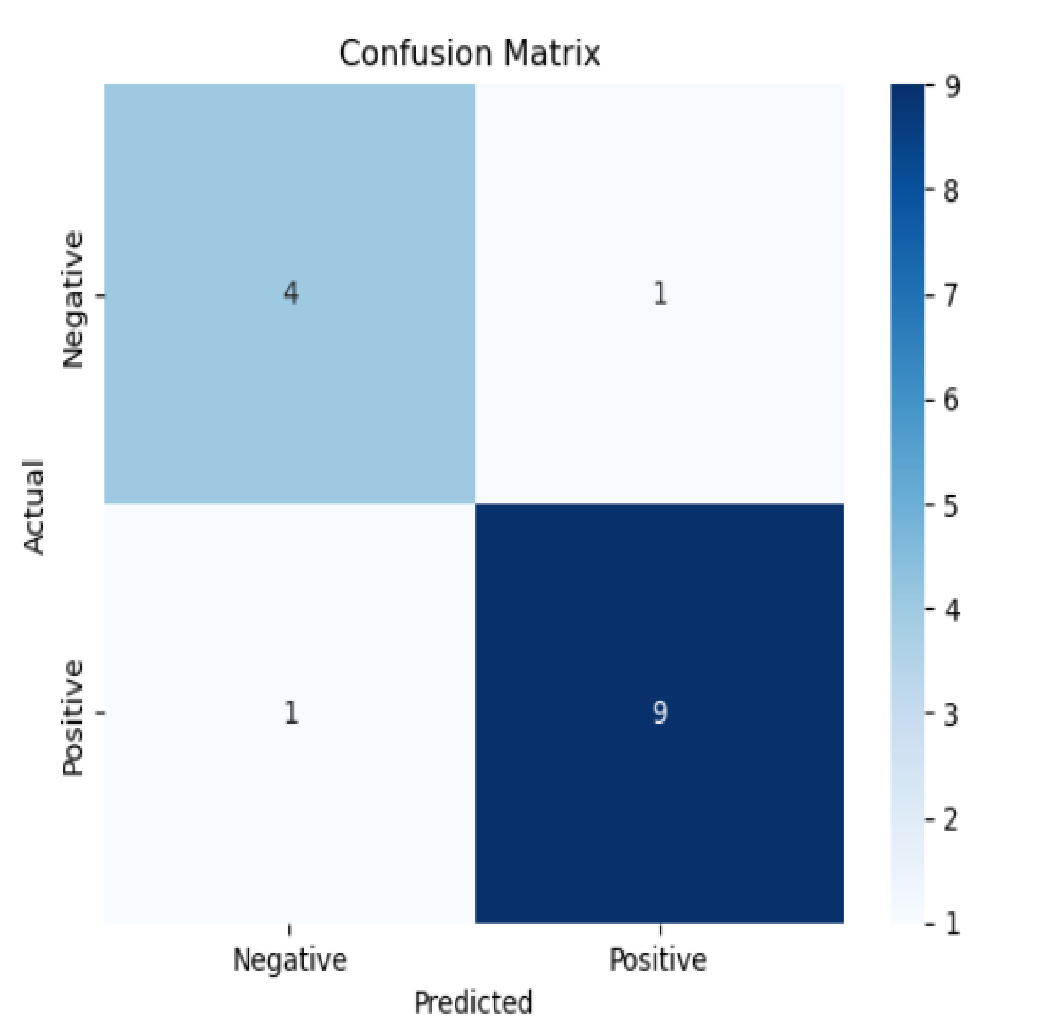

Figure 5.12 Heat Map of Confusion Matrix

The list of ratios is usually calculated from a confusion matrix for a classifier:

- Accuracy.
- Precision.
- Recall.
- Error rate.
- Specificity.
- F1 Score.

➢ Accuracy

Overall, how long does it take for the classifier to be accurate?

We can find accuracy by using the formula:

Accuracy= TP+TN/TP+FP+FN+TN

Accuracy=90.0

➢ Precision

It reflects the reliability with which the model classifies the sample as positive. We can find Precision by using the formula:

Precision = True positives/ (True positives + False positives)

Precision =92.3078

➢ Recall

The fraction of the relevant cases retrieved is also known as Sensitivity. Detecting positive samples. We can find Recall or Sensitivity by using the formula:

Recall = True positives/ (True positives + False negatives)

Recall =91.349

➢ Error Rate

The error rate generally indicates how often the model makes false predictions also known as the Misclassification rate. We can find the Error rate by using the formula.

Error Rate= 1-Accuracy

Error Rate=9.99999

➢ Specificity

Test specificity, also known as true negative rate (TNR), is the proportion of negative samples that have a negative test result using the test in question. We can find the Specificity by using the formula.

Specificity = TN / (TN + FP)

Specificity =85.71

➢ F1 Score

Precision and recall are two components of the F1 score. The F1 score is defined as the harmonized mean of precision and recall.

- If Precision and Recall are high, a model will receive a high F1 score.
- If Precision and Recall are low, a model will receive a low F1 score.
- A model will receive an average F1 score if one of the Precision and Recall is low and the other is high.

We can find the F1 Score by using the formula,

F1 = 2 * (precision * recall) / (precision + recall)

F1 =92.379

Evaluation metrics

Quantitative indicators called evaluation metrics are used to assess the effectiveness of machine learning models. Common criteria for classification include accuracy, precision, recall, F1-score, and specificity. Precision minimizes erroneous positive classifications, recall minimizes false negative classifications, and accuracy assesses total accurate classifications. Precision and recall are balanced by F1-score.

The percentage of evaluation metrics of dataset is shown in Figure 5.13.

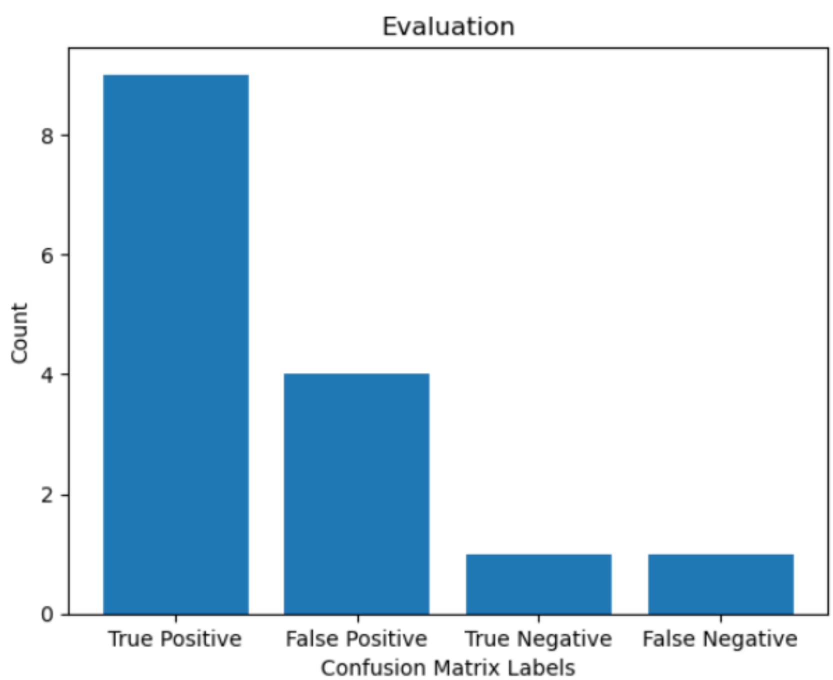

Figure 5.13 Evaluation Metrics of Dataset Percentage.

CHAPTER 6

CONCLUSION AND FUTURE WORK

Conclusion and Future work

6.1 Conclusion

A Promising area of research with the potential to revolutionize robotics is Automatic target recognition employing a robotic arm with Swin Transformer technology. The Swin Transformer model is a cutting-edge technique that has excelled in a number of computer vision applications, including segmentation and object recognition. It is conceivable to create a system that can recognize targets automatically and operate them with extreme precision by incorporating this technology into a robotic arm. Using a Swin Transformer model for target detection in a robotic arm has been successfully proved by recent research in this field, with high accuracy rates being attained in a variety of test conditions. Speed and performance, particularly in complicated situations with numerous targets or difficult lighting conditions, can yet be improved.

6.2 Future work

- Speed and performance, particularly in complicated situations with numerous targets or difficult lighting conditions, can yet be improved.
- The Incorporation of additional cutting-edge technologies, such as deep reinforcement learning or multi-task learning, could be the subject of future studies to help the robotic arm learn and adjust to new conditions more successfully.
- All things considered, the development of autonomous target recognition employing a robotic arm with Swin Transformer technology has the potential to significantly help a variety of

industries, including manufacturing, logistics, and healthcare, where precise item manipulation is crucial.

- Future research on the Swin Transformer and robotic arm for autonomous target detection could concentrate on improving the system's versatility and real-world usability.
- Integrating the system with a powerful grasp planning module might be one direction for investigation.
- The Robotic arm can detect targets as well as plan and carry out the necessary grasping operations by integrating target recognition and grasp planning. As a result, the arm would be able to engage and control recognized things more meaningfully and purposefully on its own.
- A System can intelligently instruct the robotic arm and its sensors to focus on particular regions or views that are most important for target detection by integrating active perception techniques. Particularly in intricate and clogged surroundings, this may result in more effective and precise recognition.

REFERENCES

[1] Kumar R, Lal S, Kumar S, Chand P. Object detection and recognition for a pick and place robot. InAsia-Pacific world congress on computer science and engineering 2014 Nov 4 (pp. 1-7). IEEE.

[2] Khan JF, Alam MS. Target detection in cluttered forward-looking infrared imagery.Optical Engineering. 2005 Jul;44(7):076404.

[3] Zuo, Yiming, et al. "Craves: Controlling robotic arm with a vision-based economic system." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

[4] Liu, Z., Lin, Y., Cao, Y., Han, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Hierarchical vision transformer using shifted windows.

[5] Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., ... & Guo, B. (2022). Swin transformer v2: Scaling up capacity and resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12009-12019).

[6] Talpur, M. S. H., & Shaikh, M. H. (2012). Automation of Mobile Pick and Place Roboticsystem for small food industry. arXiv preprint arXiv:1203.4475.

[7] Bhanu, B. (1986). Automatic target recognition: State of the art survey. IEEE transactions on aerospace and electronic systems, (4), 364-379.

[8] Collet, A., Berenson, D., Srinivasa, S. S., & Ferguson, D. (2009, May). Object recognition and full pose registration from a single image for robotic manipulation. In 2009 IEEE International Conference on Robotics and Automation (pp. 48-55). IEEE.

[9] Perng, J. W., Wang, C. S., Tsai, Y. C., & Chang, Y. C. (2019, September). Robotic arm object detection system. In 2019 Second International Conference on Artificial Intelligence for Industries (AI4I) (pp. 75-78). IEEE.

[10] Yusoff, M. A. K., Samin, R. E., & Ibrahim, B. S. K. (2012). Wireless mobile robotic arm. Procedia Engineering, 41, 1072-1078.

[11] Kadir, W. M. H. W., Samin, R. E., & Ibrahim, B. S. K. (2012). Internet controlled robotic arm. Procedia Engineering, 41, 1065-1071.

[12] Iqbal, J., Islam, R. U., & Khan, H. (2012). Modeling and analysis of a 6 DOF robotic arm manipulator. Canadian Journal on Electrical and Electronics Engineering, 3(6), 300-306.

[13] Jahnavi, K., & Sivraj, P. (2017, July). Teaching and learning robotic arm model. In 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (pp. 1570-1575). IEEE.

[14] Kruthika, K., Kumar, B. K., & Lakshminarayanan, S. (2016, October). Design and development of a robotic arm. In 2016 International Conference on Circuits, Controls, Communications and Computing (I4C) (pp. 1-4). IEEE.

[15] Olawale, J., Oludele, A., Ayodele, A., & Alejandro, N. M. (2007, November). Development of a microcontroller based robotic arm. In Proceedings of the 2007 Computer Science and IT Education Conference (pp. 549-557).

[16] Rogers, S. K., Colombi, J. M., Martin, C. E., Gainey, J. C., Fielding, K. H., Burns, T. J., ... & Oxley, M. (1995). Neural networks for automatic target recognition. Neural networks, 8(7-8), 1153-1184.

[17] Roth, M. W. (1990). Survey of neural network technology for automatic target recognition. IEEE Transactions on neural networks, 1(1), 28-43.

[18] Novak, L. M., Owirka, G. J., Brower, W. S., & Weaver, A. L. (1997). The automatic target-recognition system in SAIP. Lincoln Laboratory Journal, 10(2).

[19] Novak, L. M., Owirka, G. J., Brower, W. S., & Weaver, A. L. (1997). The automatic target recognition system in SAIP. Lincoln Laboratory Journal, 10(2).

[20] Liu, Z., Ning, J., Cao, Y., Wei, Y., Zhang, Z., Lin, S., & Hu, H. (2022). Video swin transformer. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 3202-3211).

[21] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022).

[22] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1833-1844).

[23] Huang, J., Fang, Y., Wu, Y., Wu, H., Gao, Z., Li, Y., ... & Yang, G. (2022). Swin transformer for fast MRI. Neurocomputing, 493, 281-304.

[24] Xie, Z., Lin, Y., Yao, Z., Zhang, Z., Dai, Q., Cao, Y., & Hu, H. (2021). Self-supervised learning with swin transformers. arXiv preprint arXiv:2105.04553.

[25] Lin, A., Chen, B., Xu, J., Zhang, Z., Lu, G., & Zhang, D. (2022). Ds-transunet: Dual swin transformer u-net for medical image segmentation. IEEE Transactions on Instrumentation and Measurement, 71, 1-15.

[26] Gao, L., Zhang, J., Yang, C., & Zhou, Y. (2022). Cas-VSwin transformer: A variant swin transformer for surface-defect detection. Computers in Industry, 140, 103689.

[27] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2022, October). Swin-unet: Unet-like pure transformer for medical image segmentation. In European conference on computer vision (pp. 205-218). Cham: Springer Nature Switzerland.

[28] Ma, J., Tang, L., Fan, F., Huang, J., Mei, X., & Ma, Y. (2022). SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer. IEEE/CAA Journal of Automatica Sinica, 9(7), 1200-1217.

[29] Ling, Z., Xing, Z., Zhou, X., Cao, M., & Zhou, G. (2023). PanoSwin: A Pano-Style Swin Transformer for Panorama Understanding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 17755-17764).

[30] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022).

[31] Fang, Y., Wang, X., Wu, R., & Liu, W. (2023). What Makes for Hierarchical Vision Transformer?. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[32] Wang, W., Yang, X., & Tang, J. (2023). Vision Transformer with Hybrid Shifted Windows for Gastrointestinal Endoscopy Image Classification. IEEE Transactions on Circuits and Systems for Video Technology.

[33] Wang, W., Yang, X., & Tang, J. (2023). Vision Transformer with Hybrid Shifted Windows for Gastrointestinal Endoscopy Image Classification. IEEE Transactions on Circuits and Systems for Video Technology.

[34] Chen, X., Qiu, C., Guo, W., Yu, A., Tong, X., & Schmitt, M. (2022). Multiscale feature learning by transformer for building extraction from satellite images. IEEE Geoscience and Remote Sensing Letters, 19, 1-5.

[35] Bojesomo, A., Marzouqi, H. A., & Liatsis, P. (2022). SwinUNet3D--A Hierarchical Architecture for Deep Traffic Prediction using Shifted Window Transformers. arXiv preprint arXiv:2201.06390.

[36] Sutar, A. V., & Ohol, D. S. (2014). Kinematic And Dynamic Analysis Of A Robot Arm Used For All Terrain Robot. International Journal of Engineering Research & Technology (IJERT), 3(5).

[37] Atkeson, C. G., & Schaal, S. (1997, July). Robot learning from demonstration. In ICML (Vol. 97, pp. 12-20).

[38] Števo, S., Sekaj, I., & Dekan, M. (2014). Optimization of robotic arm trajectory using genetic algorithm. IFAC Proceedings Volumes, 47(3), 1748-1753.

[39] Becsik, A. L., & Rudas, I. J. (2004, December). On dynamic rigidity of robotic manipulators. In Proceedings of the 4th WSEAS International Conference on Systems Theory and Scientific Computation (pp. 1-6).

[40] Achic, F., Montero, J., Penaloza, C., & Cuellar, F. (2016, July). Hybrid BCI system to operate an electric wheelchair and a robotic arm for navigation and manipulation tasks. In 2016 IEEE workshop on advanced robotics and its social impacts (ARSO) (pp. 249-254). IEEE.

[41] Tamosiunaite, M., Nemec, B., Ude, A., & Wörgötter, F. (2011). Learning to pour with a robot arm combining goal and shape learning for dynamic movement primitives. Robotics and Autonomous Systems, 59(11), 910-922.

[42] Latif, M. Y., Naeem, L., Hafeez, T., Raheel, A., Saeed, S. M. U., Awais, M., ... & Anwar, S. M. (2017, September). Brain computer interface based robotic arm control. In 2017 International Smart Cities Conference (ISC2) (pp. 1-5). IEEE.

[43] Cimprich, K. A., & Cortez, D. (2008). ATR: an essential regulator of genome integrity. Nature reviews Molecular cell biology, 9(8), 616-627.

[44] Nam, E. A., & Cortez, D. (2011). ATR signalling: more than meeting at the fork. Biochemical Journal, 436(3), 527-536.

[45] Brown, E. J., & Baltimore, D. (2003). Essential and dispensable roles of ATR in cell cycle arrest and genome maintenance. Genes & development, 17(5), 615-628.

[46] Liu, S., Shiotani, B., Lahiri, M., Maréchal, A., Tse, A., Leung, C. C. Y., ... & Zou, L. (2011). ATR autophosphorylation as a molecular switch for checkpoint activation. Molecular cell, 43(2), 192-202.

[47] Saldivar, J. C., Cortez, D., & Cimprich, K. A. (2017). The essential kinase ATR: ensuring faithful duplication of a challenging genome. Nature reviews Molecular cell biology, 18(10), 622-636.

[48] Suryawanshi, D., & Khandelwal, C. S. (2012). An integrated color and hand gesture recognition control for wireless robot. International Journal of Advances in Engineering & Technology, 3(1), 427.

[49] Xu, G., Zhou, D., Yuan, L., Guo, W., Huang, Z., & Zhang, Y. (2023). Vision-based underwater target real-time detection for autonomous underwater vehicle subsea explorationn. Frontiers in Marine Science, 10, 1112310.

[50] Oria, J. P., & Gonzalez, A. M. G. (1993, November). Object recognition using ultrasonic sensors in robotic applications. In Proceedings of IECON'93-19th Annual Conference of IEEE Industrial Electronics (pp. 1927-1931).

[51] Team, T.A. (no date) Uno R3, Arduino Documentation. Available at: https://www.arduino.cc/en/Guide/ArduinoUno (Accessed: 07 August 2023).

[52] Macfos (2023) Buy SG90 servo motor 180? rotation online at the best price in India, Robu.in | Indian Online Store | RC Hobby | Robotics. Available at: https://robu.in/product/towerpro-sg90-9gm-1-2kg-180-degree-rotation-servo-motor-good-quality/ (Accessed: 07 August 2023).

[53] Supported cameras (2023) LightBuzz. Available at: https://lightbuzz.com/docs/general-information/supported-cameras/ (Accessed: 07 August 2023).