Jurnal Teknologi dan Rekayasa Manufaktur

JTRM | Vol. 7 | No. 2 | Tahun 2025
ISSN (P): 2715-3908 | ISSN (E): 2715-016X
DOI: HTTPS://DOI.ORG/10.48182/JTRM.V7I2.149

Deep learning Implementation in Multi-Fingered Manipulator Robot for Pick and Place Food Serving Equipment

Ruminto Subekti[1], Ismail Rokhim[2], Muhammad Sulaeman Gheofani[3]

[1] Teknik Otomasi dan Mekatronika, Politeknik Manufaktur Bandung
[2] Teknik Otomasi dan Mekatronika, Politeknik Manufaktur Bandung
[3] Teknik Otomasi dan Mekatronika, Politeknik Manufaktur Bandung

Email: ruminto_s@polman-bandung.ac.id

Informasi Artikel:

Received:
02 September 2023

Accepted:
15 Oktober 2025

Available:
15 Desember 2025

ABSTRAK

Perusahaan perjalanan, pariwisata, dan perhotelan, mulai mengadopsi sistem RAISA dalam bentuk *chatbot*, robot pengiriman, mesin pencuci piring otonom, restoran konveyor, kios informasi swalayan, dan banyak lainnya. Penelitian ini berfokus pada implementasi *deep learning* jaringan saraf buatan untuk pengenalan objek dalam menentukan estimasi pose robot manipulator dan merencanakan cengkeraman pada efektor akhir. Manipulator robot dengan 4° kebebasan digunakan untuk mendukung estimasi sudut pose dan efektor akhir dalam bentuk *gripper* 5 jari digunakan untuk mendapatkan berbagai pegangan pada objek dengan bentuk acak. Kamera RGB digunakan untuk pengenalan objek dengan konfigurasi *eye-on-hand*, yang ditautkan ke efektor akhir untuk mendapatkan informasi visual pada objek menggunakan algoritma pembelajaran mendalam YOLOv3. Efektor akhir bekerja secara optimal pada benda dengan bentuk dasar tabung, prisma persegi panjang, prisma segi enam dan prisma sepuluh sisi dengan beban maksimum yang dapat diangkat 303 gram dengan tingkat keberhasilan 71,23%.

Kata Kunci:

Serving Robot,
RAISA,
Manipulator,
Multi-fingered Gripper,
Deep Learning

ABSTRACT

Travel, tourism and hospitality companies have started to adopt RAISA systems in the form of chatbots, delivery robots, autonomous dishwashers, conveyor restaurants, self-service information kiosks and many others. This research focuses on the implementation of deep learning artificial neural networks for object recognition in determining the pose estimation of the manipulator robot and planning the grip on the end effector. A robotic manipulator with 4 degrees of freedom is used to support the estimation of pose angles and an end effector in the form of a 5-finger gripper is used to obtain various grips on objects with random shapes. An RGB camera is used for object recognition with an eye-on-hand configuration, which is linked to the end effector to obtain visual information on objects using the YOLOv3 deep learning algorithm. The end effector works optimally on objects with the basic shape of a tube, rectangular prism, hexagon prism and ten-sided prism with a maximum load that can be lifted of 303 grams with a success rate of 71,23%.

1. INTRODUCTION

Following advances in the development of robots, artificial intelligence and service automation (RAISA) and the industrial revolution 4.0, companies from various economic sectors are using RAISA to improve operating processes, optimize costs, create customer experiences and expand service capacity [1]. Travel, tourism and hospitality companies have started to adopt RAISA systems in the form of chatbots, delivery robots, autonomous dishwashers, conveyor restaurants, self-service information kiosks and many others [1], [2]. The Covid-19 pandemic has also stimulated the acceleration of robotics adoption in travel, tourism and hospitality (TTH) in restaurant chains, hotel chains and airline companies.

Some entrepreneurs try to innovate through approaching customers by developing product delivery systems to customers to produce an interesting experience in the form of using robots [3]. In contrast to industrial robots that handle objects of a uniform type, serving robots are assigned to handle a wide variety of various food dishware objects. Tasks that seem simple to humans, such as picking up and placing objects of various shapes, sizes, materials, and surface properties, can be challenging for robots to grasp [4]. Therefore, it is necessary to develop a rendering robot system to solve the problem of holding objects with various class variants in practical applications [6]. The need to adjust product posture requires that the gripper shape be adjusted specifically for each product variant, where the more product variants, the more gripper variants that need to be replaced for each different product [6], [7].

Referring to the development of grippers in the industry, the multi-finger gripper is a structure that allows the robot to grip a wide variety of products because its structure is closest to that of the human hand [8], [9]. Then, an algorithm is needed that is capable of classifying certain objects so that each object variant can be given a variety of object gripping treatments according to its class. Basically there are several methods that can be implemented so that the system can find out the class of objects, one of which is by embedding a deep learning neural network system. Deep learning can be implemented to solve problems of image recognition, speech recognition, natural language processing and anomaly detection. In object recognition of food cooking equipment, the information that can be processed is visual information of the object.

2. METHOD

2.1. System Overview

This research is focused on object detection using basic devices on personal computers in the form of CPUs and webcams in computing image processing algorithms. Then the results of image processing will be processed to direct the robot manipulator to move closer to the detected object, then the effect of using the control method on the end effector of the prosthetic hand robot in holding the object is measured. An overview of the system can be seen in Figure 1 below.

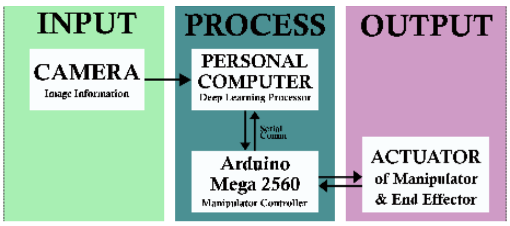

Figure 1 General System

The research conducted was experimental research which is a quantitative research method used to determine the effect of the independent variable (treatment) on the dependent variable (outcome) with all data used being numerical data. In this study, image data of dishware in the form of food serving utensils were used as input data to be processed in robot control in the pick and place task.

In the aspect of system design, this research uses the VDI 2206 system design guide which is a methodology in designing mechatronic systems, namely a combination of mechanics, electricity and informatics. The first stage is making a concept to clarify the needs of the system and as an illustration of what things must be considered in the system to be made. Some of the work requirements that must be met in the design process can be seen in Table 2. This list of specifications becomes a reference in selecting the components used.

Table 1 List of work requirements

No.	Objective	Spesification
1	Mechanical System	Simple and easy in assembly.
		The structure is modular.
		Has an actuator with a certain torque capacity to be able to lift a 500 gram object load.
		Easy to procure components and spare parts.
2	Electrical System	Simple and easy to assemble and troubleshoot.
		Easy to procure components and spare parts.
		The complexity of electrical circuits is made as simple as possible.
		Effective and efficient as well as modular to make modifications if needed.
3	Information System	C++ programming algorithm for programming Arduino.
		A programming algorithm in python to program deep learning on a computer.
		Uses the serial communication protocol for computer to Arduino communication.
4	Function	The system is able to detect and identify objects of food serving equipment.
		The system is able to grasp food dishware objects in random shapes.
		The system is able to move objects from point A to point B.
5	Process	Easy to operate (user friendly).
		The system can be operated manually and automatically.
		Class and object point detection by RGB camera, sending command result data. deep learning computing towards Arduino for manipulation using a stepper motor drive and a prosthetic robot hand end effector using a servo motor drive.

2.2. Schematics System Design

This stage is a detailed description of the general concepts that have been built in the previous section. At this stage, it is divided into 3 parts, namely: mechanical engineering, electronic engineering and information technology.

2.2.1. Mechanical Engineering

The manipulator structure consists of two main parts, namely links and joints. Joints are joints or hinges that make the manipulator move, while links are links between joints whose length can be adjusted as needed.

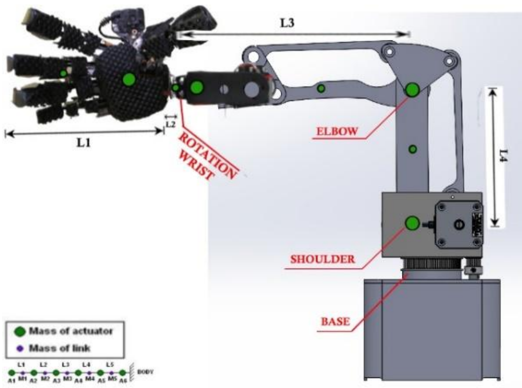

Figure 2 Mechanical Design

The five fingers on the hand consist of three different types of joints: carpometacarpal (CMC), metacarpophalangeal (MP) and interphalangeal (IP). At the same time, the hand is a complex structure and a tight space. It is very difficult to place the actuators individually in each of the 3 joints in the 5 fingers.

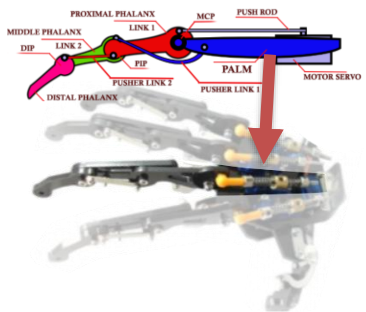

Figure 3 Five fingered gripper structure

To overcome this, the principle of under-actuation linkage mechanism is used as one of the strategies commonly used for finger design. A mechanism is said to be under-actuated when it has fewer actuators than the available degrees of freedom.

2.2.2. Electronic Engineering

When controlling a digital servo motor, it is only necessary to connect one cable to transmit a pulse signal, whereas on a stepper motor, up to 5 wires are needed for one motor unit.

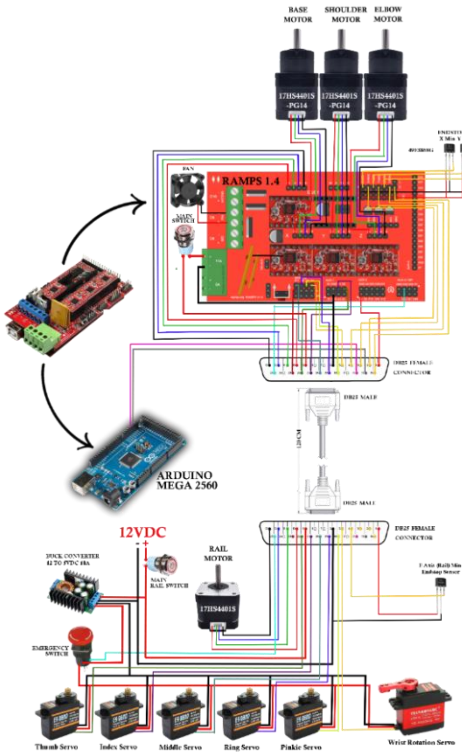

Figure 4 Electrical Design

Therefore, a circuit board is needed to support the stepper motor control circuit to make it simpler. The option that can be chosen is to use a variant of the circuit board for the 3D printer kit, including the RAMPS 1.4 circuit board (Reprap Arduino Mega Pololu Shield). RAMPS 1.4 is a shield for Arduino Mega which simplifies the circuit of stepper motor driver modules.

RAMPS 1.4 has a regulator circuit for power supply to be distributed to all stepper motors. In addition, I/O pins that are not integrated into the stepper motor control can still be accessed to add inputs and outputs. Digital servo motors to control end effectors utilize available digital pins and can be accessed via RAMPS 1.4.

2.2.3. Information Technology

The control system is divided into two sub-systems, namely the manipulator control sub-system using the Arduino Mega 2560 in C++ and the image processing sub-system using a personal computer in Python. The two sub-systems are then integrated via a serial communication protocol via a USB connection.

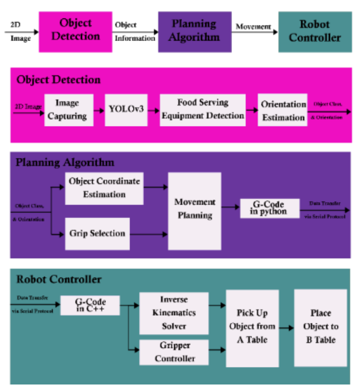

Figure 5 Program Data Routing

The manipulator control consists of two modes, namely manual mode and automatic mode. Manual mode defines the manipulator control conditions using a software called YAT (Yet Another Terminal). YAT is a popular terminal emulation program for windows which is used to send serial data strings from windows to devices with serial communication protocol via USB.

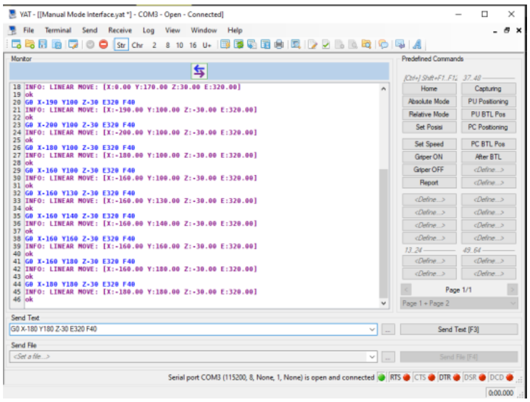

Figure 6 Manual mode interface on YAT

Whereas in automatic mode, the object identity recognized by the object recognition system is sent to the Arduino Mega 2560 along with the object's geometry information. The geometry of the objects obtained is then used as a reference for the system to be stored in a grasping configuration set containing information on the angle values of the thumb, forefinger, middle, ring and little finger.

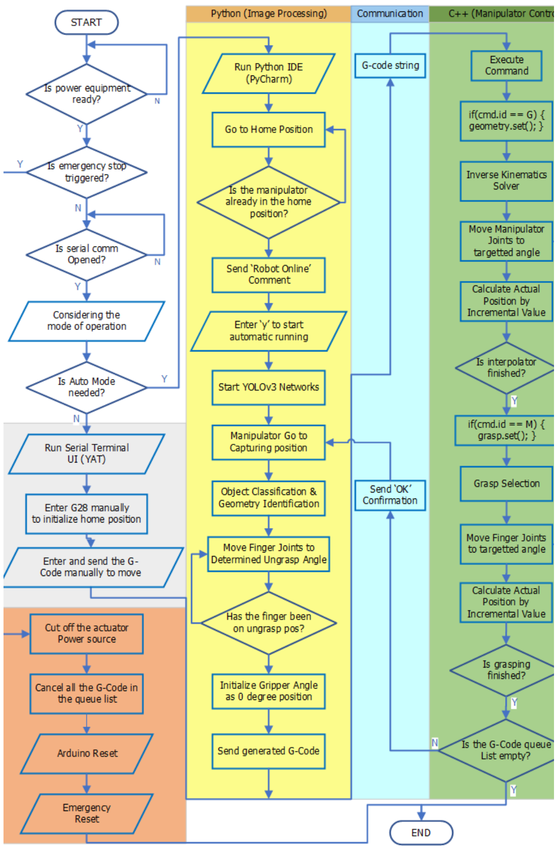

Figure 7 System flowchart

If an object class and its geometry value is detected, the system will call the previously prepared configuration set as a reference for the gripper command to do its job. Five fingers that move individually will move in sync with the movement of the wrist rotation joint to adjust the position of the object. The joint angle of the wrist rotation for each object class will be different according to its geometry.

3. RESULT

The resulting set of devices that are built simulates pick-up and serving from the restaurant kitchen table as point A to the serving table as point B. The transfer process from area A to area B is assisted by a conveyor rail.

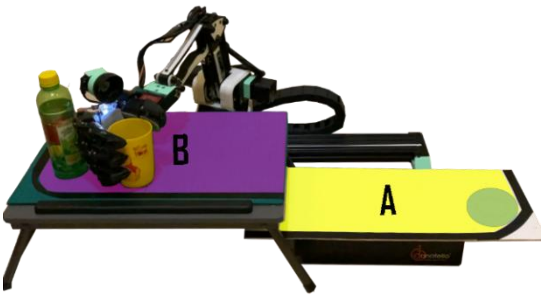

Figure 8 System on pick and place operation

The system that has been built has gone through several adjustments including adjusting the manipulator construction, adjusting the gripper surface structure to adjusting the camera optics used on the camera. Combines the complexity of the under-actuated mechanism on the gripper finger used with the functionality of the gripper to grip objects. The detrimental consequence of using an under-actuated mechanism is that the number of shafts in each frame causes a reduction in the static surface area which functions as the main component that makes direct contact with the surface of the object. The reduced static surface area requires adjustments to be made by adding a surface layer covering the gripper area. The added layer is a layer of dash mat material which has elastic and rough patterned properties.

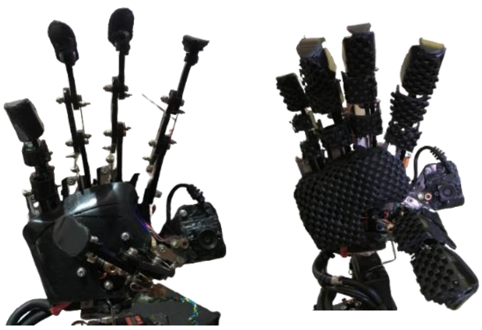

Figure 9 Gripper surface modification

This adjustment is the most important adjustment because the gripper is the only part that makes direct surface contact with the surface of the object. This system uses the camera as a receiver of image information where the performance of the camera in collecting visual data will greatly affect the overall performance of the system. Therefore, many adjustment trials are carried out in order to get the best quality captured images that contain the best visual information.

Figure 10 0.45x multiplication macro lens

The adjustments made are adjustments to the angle of the image that can be captured and the addition of an antidote, namely light interference. A macro lens is used with a multiplier of 0.45x which can widen the angle of the image capture range.

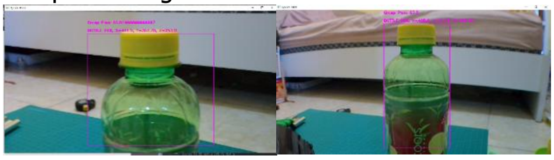

Figure 11 Visual comparison of the results using a macro lens

This image widening effect helps the system capture visual data from a more complete part of the object at the same distance. This is very influential because the camera placement configuration used is eye-on-hand where the camera position is very close to the object so that only a few parts of the object can be captured.

Besides being able to enlarge the angle captured by the camera, the lens structure which is shaped like a bowl and consists of a concave lens produces the effect of blocking light coming from directions other than the direction perpendicular to the lens.

3.1. System Test Result

3.1.1. Image Processing System

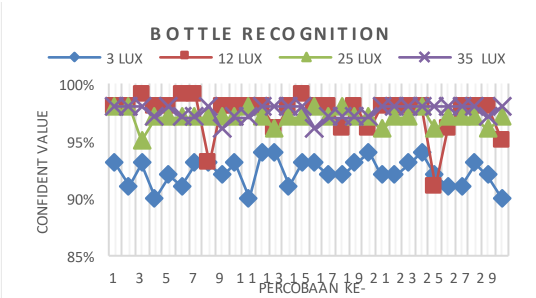

Figure 12 Response of the system's confident value to the light intensity of the bottle

Adjustment of light intensity was carried out with the help of a lux meter. Then it is measured as much as the confidence value by the object detection system. In the detection of bottle objects, the highest average confidence value is obtained at a light intensity of 35 lux, which is 98%, but for the peak value, the highest confidence is obtained when the intensity value is at 12 lux, which is 99%.

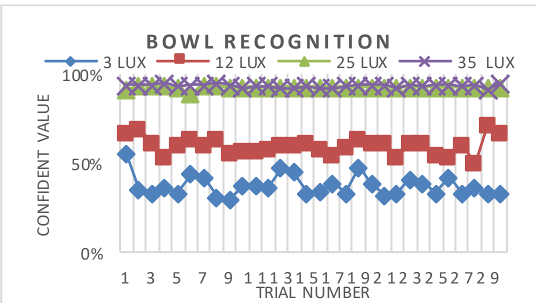

Figure 13 Response of the system's confident value to the light intensity of the bowl

In detecting bowl objects, the highest average confidence value is obtained at a light intensity of 35 lux, which is 93%, and the highest confidence peak value is obtained when the intensity value is at 35 lux, which is 95%.

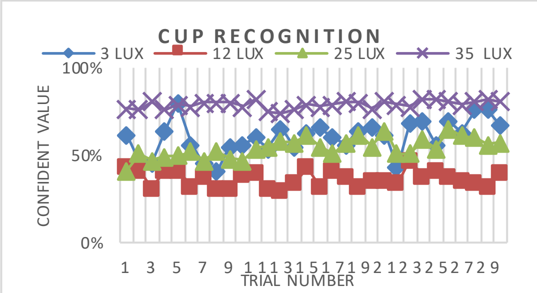

Figure 14 Response of the system's confident value to the light intensity of the cup

Adjustment of light intensity was carried out with the help of a lux meter. Then it is measured as much as the confidence value by the object detection system. In the detection of cup objects, the highest average confidence value is obtained at a light intensity of 35 lux, which is 79%.

3.1.2. Computation Duration

The identification process on the bottle object with the fastest average computational duration of 727.4 milliseconds was obtained when the light intensity was 12 lux.

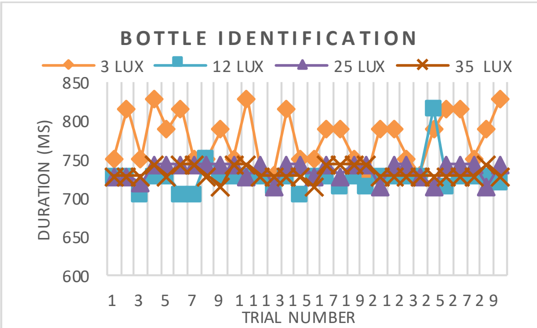

Figure 15 Computational duration of bottle object identification at different light intensities

While the bottle object identification process with the longest milliseconds average computational duration of 777.43 was obtained when the light intensity value was 3 lux.

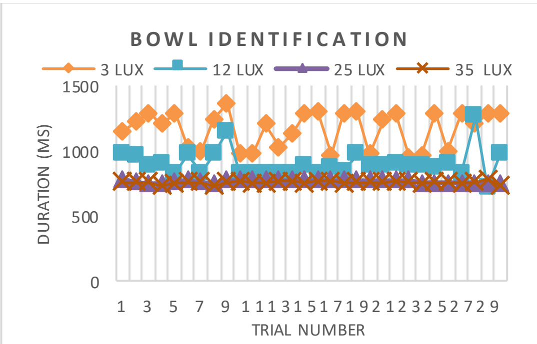

Figure 16 Computational duration of bowl object identification at different light intensities

The identification process on bowl objects with the fastest average computational duration of 754.2 milliseconds was obtained when the light intensity was 35 lux. While the process of identifying bowl objects with the longest average computational duration of 1166.33 milliseconds is obtained when the light intensity value is 3 lux. The identification process on the cup object with the fastest average computational duration of 773.7 milliseconds was obtained when the light intensity was 35 lux.

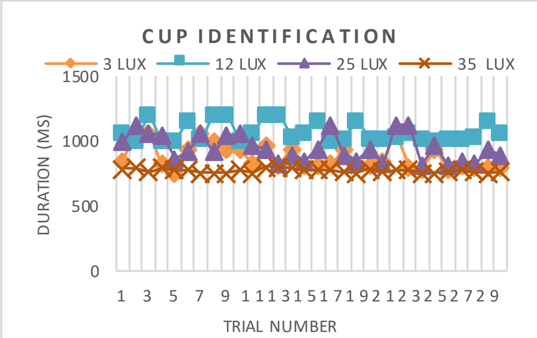

Figure 17 Cup object identification computational duration

While the process of identifying the cup object with the longest average computation duration of 1072.13 milliseconds is obtained when the light intensity value is 12 lux.

3.1.3. Grasping Result

In the bottle object, the results of slip begin to occur when the object's mass value is 150 grams. Bottle objects with an average slip of at least 9.4 times are obtained by bottle objects with a narrow or curvy cylindrical shape, namely slip occurs when the object's mass value is 300 grams.

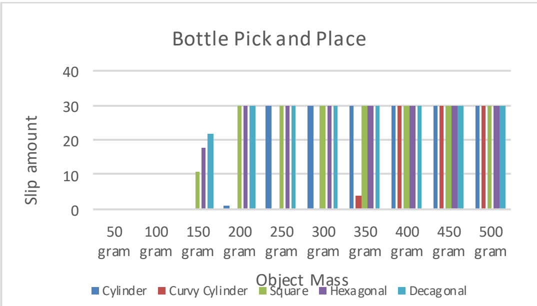

Figure 18 Chart of the amount of slip on the bottle object grasping

Whereas the bottle object with the highest average slip of 23.2 times was obtained by the bottle object with a hexagon shape, namely slip begins to occur when the object's mass is 150 grams.

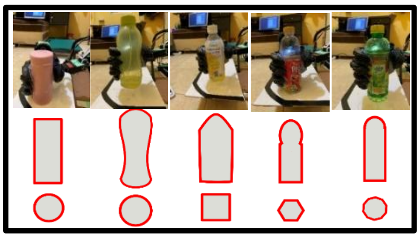

Figure 19 Bottle shape variant illustration

A bowl object with an average slip of at least 24.2 times is obtained by a bowl object with a circular shape with a diameter of 100mm, that is, slip occurs when the object's mass value is 100 grams.

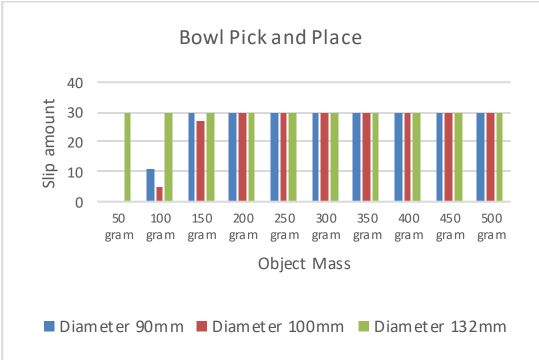

Figure 20 Chart of the amount of slip on the bowl object grasping

While the bowl object with the highest average slip of 30 times is obtained by the bowl object with a circular shape with a diameter of 90mm, namely slip begins to occur when the object's mass is 50 grams.

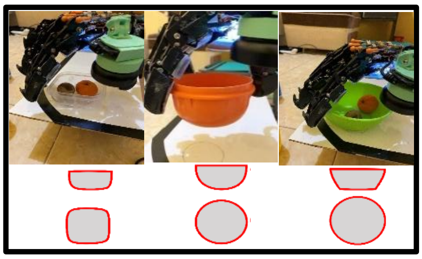

Figure 21 Bowl shape variant illustration

A cup object with an average slip of at least 19.7 times is obtained by a cup object with a cylindrical shape with a diameter of 92mm, namely the slip occurs when the object's mass value is 150 grams 5 times.

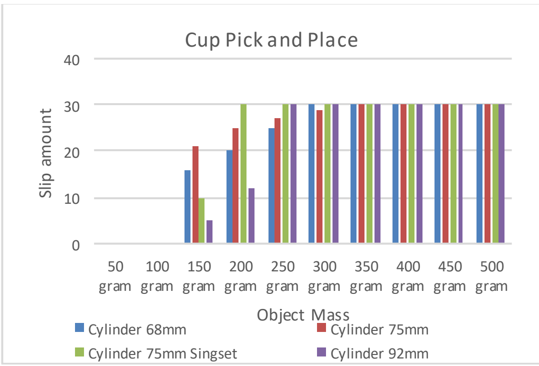

Figure 22 Chart of the amount of slip on the cup object grasping

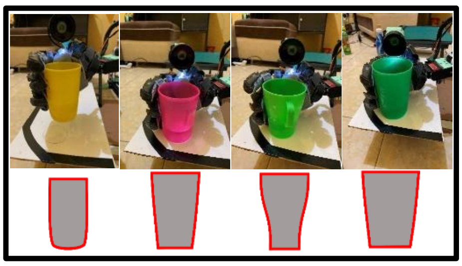

Figure 23 Cup shape variant illustration

While the cup object with the highest average slip of 22.2 times was obtained by the cup object with a cylindrical shape with a diameter of 75mm, namely slip began to occur when the object's mass was 150 grams 21 times.

3.1.4. Placing Coordinate Result

Cartesian coordinate references are used to specify the coordinates of the target object and the coordinates obtained by the object when placing. Binding is carried out on the robot device and serving table so that the distance between the robot and the serving table is consistent to produce accurate measurement data. When the coordinate test is carried out, the distance between the 0 point of the robot reference and the 0 point of the serving table reference on the X axis is 131mm and on the Y axis is 130mm.

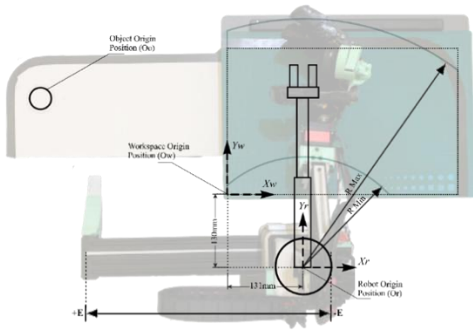

Figure 24 Illustration of cartesian coordinates reference of robot and work area

Therefore, calculations are needed to determine the actual target coordinate point that must be achieved at the serving table coordinate reference.

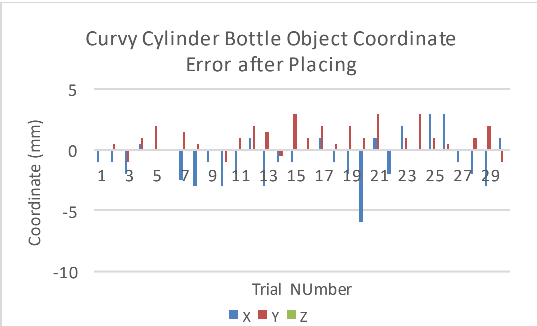

Figure 25 Chart of error values on placing curvy cylindrical bottle objects

The error value in the graph above is obtained from the deviation of the coordinates between the target coordinates and the results of measuring the actual coordinates of the object after the placing process. Placing a narrow cylindrical bottle object gets an average error of -0.83mm on the X axis and -0.917mm on the Y axis and 0mm on the Z axis.

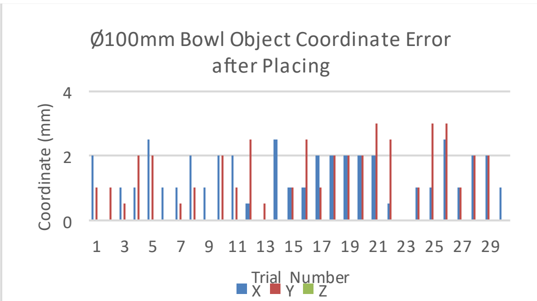

Figure 26 Chart of error values on placing 100mm bowl objects

Placing a bowl object with a diameter of 100mm gets an average error of 1.38mm on the X axis and 1.4mm on the Y axis and 0mm on the Z axis.

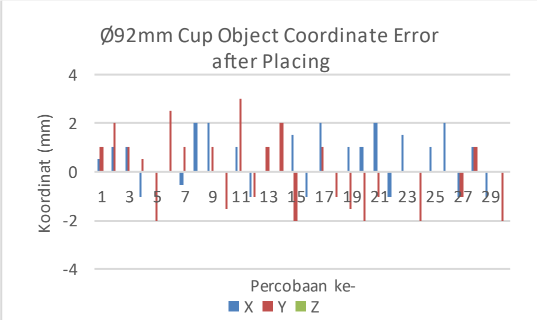

Figure 27 Chart of error values on placing 92mm Cup objects

Placing a 92 mm diameter cup object gets an average error of 0.467mm on the X axis and 0.0mm on the Y axis and 0mm on the Z axis.

4. ANALYSIS

4.1. Confident Value

The confidence value of the object detection system which includes 3 classes shows the light intensity at 35 lux is the best intensity to produce the highest confidence value obtained when detecting bowls and cups objects. Whereas in the detection of bottle objects, the intensity of 12 lux produces the highest confidence value. But even at the lowest light intensity condition, namely 3 lux which is a dim condition in an open room without lights, the system is still able to detect and classify correctly according to its class without any classification errors. This shows that the system can work according to its function in a range of light intensities that vary according to the intensity that humans generally carry out activities, namely in the range of 9 lux to 35 lux.

4.2. Computation Duration

The duration of deep learning computation in translating objects into g-code depends on the quality of the visual information obtained. For the bottle object, the fastest average computation is obtained at a light intensity of 12 lux while the slowest average computation is obtained at a light intensity of 3 lux. Meanwhile, for bowls and cups, the fastest average duration was obtained at a light intensity of 35 lux. While the slowest average duration for the bowl object is at a light intensity of 3 lux and for the bowl object at a light intensity of 12 lux. The data shows that low light intensity does not necessarily have a negative impact on system performance. In the identification of cup objects, the light intensity of 3 lux actually produces a faster computational speed compared to the intensity of 12 lux. At a light intensity of 3 lux, the longest average computation duration is 1166.3 milliseconds. However, with this duration it is still within reasonable limits and it makes sense to start the pick and place process.

4.3. Grasping Result

In the pick and place objects for bottles, bowls and cups, it was found that the grippers were able to grip and lift objects optimally at a maximum object mass value of 150 grams. Therefore, 150 grams is assumed to be the maximum object limit in this pick and place work. Furthermore, the value of 150 grams will be used as a reference to calculate the level of success of the end effector in gripping and moving objects.

Success rate = (Successful completion number of the task)/(Total trial number)
Success rate of bottle = 399/450 = 0.886
Success rate of bowl = 107/270 = 0.396
Success rate of cup = 308/360 = 0.855

Total Success Rate = (Bottle SR +Bowl SR+ Cup SR)/3
= (0.886 +0.396 +0.855)/3 x 100 = 71.23%

4.4. Placing Coordinate Result

Object position is the final result in pick and place work. In this work, the coordinates measured are the coordinates on the X and Y axes because the coordinates of the object on the Z axis are definitely attached to the base height value at point B, which is 60mm from the floor and 55mm on the Z axis of the robot. On the bottle object, the error value is obtained on the X axis of -0.83 mm and -0.917mm on the Y axis. In the bowl object, the error value obtained on the X axis is 1.38mm and the Y axis is 1.4mm. On the cup object, the error value obtained on the X axis is 0.467 mm and the Y axis is 0.0 mm. The average error obtained is not more than 1mm so that the deviation is still within reasonable limits. The farthest coordinate deviation is obtained by the bottle object, which is 6mm on the X axis and on the Y axis, the farthest error is 3mm on average. But in practice, 6mm does not have a significant effect on the work of presenting the robot so that the results of the coordinates of the objects reached are still considered relevant.

5. CONCLUSION

1) A deep learning algorithm tasked with recognizing object classes and geometric information embedded into the robot manipulator's control is capable of executing a 5-finger prosthetic robot end effector to grip 3 object classes and perform pick and place tasks on objects with the basic shape of a tube, rectangular prism, hexagonal prisms and ten-sided prisms with a maximum load of 303 grams that can be lifted. Meanwhile, for objects with a hemispherical basic shape such as a bowl with a diameter of 100mm and a load value of 66 grams and a maximum lifting capacity of 150 grams.

2) The optimal light intensity used is in the range of 12 lux to 35 lux which is the light intensity that can be obtained in a room with normal lighting. The best object detection performance was obtained at a light intensity of 35 lux with an average value of 90.1% confidence and an average duration of 753.67 milliseconds.

3) In a previous study entitled "Industrial Robot Grasping with Deep Learning using a Programmable Logic Controller (PLC)", with the use of a depth camera and the Fully Convolutional Grasp Quality Convolutional Neural Network (FC-GQ-CNN) object detection method the system was able to achieve a success value rate 90-95%. Whereas the multi finger manipulator robot system with YOLOv3 in pick and place operation with a maximum load capacity and object diameter when the light intensity is at 35 lux, the system is able to obtain a success rate of 71.23%. The system has not been able to produce better performance than previous studies because the computing speed of the devices used is still low and economical. But system performance can be improved again by improvising in several aspects which will be described in the suggestions section. In addition, the use of hardware and software used in this research is open source and can be accessed and developed by anyone.

4) The multi-finger robot manipulator system for picking and placing food serving equipment can be applied in restaurants by developing a larger end effector actuator capacity for greater grip and adding the length of the conveyor rail to adjust the distance between the kitchen and the serving table.

6. REFERENCE

[1] S. Ivanov and C. Webster, "ADOPTION OF ROBOTS, ARTIFICIAL INTELLIGENCE AND SERVICE AUTOMATION BY TRAVEL, TOURISM AND HOSPITALITY COMPANIES-A COST-BENEFIT ANALYSIS," International Scientific COnference "Contemporary Tourism - Traditions and Innovations," 2017, [Online]. Available: https://ssrn.com/abstract=3007577

[2] I. Voysey, T. George Thuruthel, and F. Iida, "Autonomous dishwasher loading from cluttered trays using pre-trained deep neural networks," Engineering Reports, vol. 3, no. 5, May 2021, doi: 10.1002/eng2.12321.

[3] A. O. C. C. M.Omar Parvez, "Does Coronavirus (COVID-19) Transform Travel and Tourism to Automation (Robots)?," Advances in global services and retail management: Volume 2, Sep. 2021, doi: 10.5038/9781955833035.

[4] L. Yang, S. Wu, Z. Lv, and F. Lu, "Research on manipulator grasping method based on vision", doi: 10.1051/matecconf/202030.

[5] Z. Samadikhoshkho, K. Zareinia, and F. Janabi-Sharifi, "A Brief Review on Robotic Grippers Classifications," 2019.

[6] W. Widhiada, T. G. T. Nindhia, and N. Budiarsa, "Robust Control for the Motion Five Fingered Robot Gripper," International Journal of Mechanical Engineering and Robotics Research, 2015, doi: 10.18178/ijmerr.4.3.226-232.

[7] W. Miao, G. Li, G. Jiang, Y. Fang, Z. Ju, and H. Liu, "OPTIMAL GRASP PLANNING OF MULTI-FINGERED ROBOTIC HANDS: A REVIEW," Appl. Comput. Math, pp. xx–xx, 2015.

[8] H.-S. Fang, C. Wang, and M. G. C. Lu, "GraspNet-1Billion: A Large-Scale Benchmark for General Object Grasping," 2020, [Online]. Available: www.graspnet.net.

[9] A. C. Hernandez, C. Gomez, J. Crespo, and R. Barber, "Adding uncertainty to an object detection system for mobile robots," in Proceedings - 6th IEEE International Conference on Space Mission Challenges for Information Technology, SMC-IT 2017, Dec. 2017, vol. 2017-December, pp. 7–12. doi: 10.1109/SMC-IT.2017.9.

[10] K. A. Tanwani, N. Mor, J. Kubiatowicz, J. E. Gonzales, K. Goldberg, and Institute of Electrical and Electronics Engineers, "A Fog Robotics Approach to Deep Robot Learning: Application toObject Recognition and Grasp Planning in Surface Decluttering," 2019.

[11] J. Xiong, W. Cui, W. Zhang, and X. Zhang, "YOLOv3-Darknet with Adaptive Clustering Anchor Box for Intelligent Dry and Wet Garbage Identification and Classification," in Proceedings - 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics, IHMSC 2019, Aug. 2019, vol. 2, pp. 80–84. doi: 10.1109/IHMSC.2019.10114.

[12] S. Soltan, A. Oleinikov, M. F. Demirci, and A. Shintemirov, "Deep learning-based object classification and position estimation pipeline for potential use in robotized pick-and-place operations," Robotics, vol. 9, no. 3, Sep. 2020, doi: 10.3390/ROBOTICS9030063.

[13] N. Saito, T. Ogata, S. Funabashi, H. Mori, and S. Sugano, "How to Select and Use Tools?: Active Perception of Target Objects Using Multimodal Deep Learning," IEEE Robot Autom Lett, vol. 6, no. 2, pp. 2517–2524, Apr. 2021, doi: 10.1109/LRA.2021.3062004.

[14] Q. X. Do, Y. S. Chan, and D. Roth, "Minimally supervised event causality identification," in EMNLP '11 Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2011, pp. 294–303.

[15] J. Atkinson and A. Rivas, "Discovering novel causal patterns from biomedical natural-language texts using Bayesian nets," IEEE Transactions on Information Technology in Biomedicine, vol. 12, no. 6, pp. 714–722, 2008, doi: 10.1109/TITB.2008.920793.

[16] G. Milette and A. Stroud, Professional Android Sensor Programming. Wrox; 1 edition, 2012.