Konya Mühendislik Bilimleri Dergisi, c. 11, s. 2, 423-438, 2023
Konya Journal of Engineering Sciences, v. 11, n. 2, 423-438, 2023
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1228275

DEEP LEARNING BASED HUMAN ROBOT INTERACTION WITH 5G COMMUNICATION

[1*] **Mücahid BARSTUĞAN**, [2] **Zeynep OSMANPAŞAOĞLU**

[1]*Konya Technical University, Engineering and Natural Sciences Faculty, Electrical and Electronics Engineering Department, Konya, TÜRKİYE*
[2]*Marmara University, Technology Faculty, Mechatronics Engineering Department, İstanbul, TÜRKİYE*
[1]mbarstugan@ktun.edu.tr, [2]zeynep.osmanpasaoglu@tr.mee.com

Highlights

- The remote control process of industrial collaborative robotic arm over 5G communication is presented.
- The control process is implemented over deep learning methods such as YOLOv4 and YOLOv5.
- The robotic hand is remotely controlled using a glove equipped with flex sensors.

Graphical Abstract

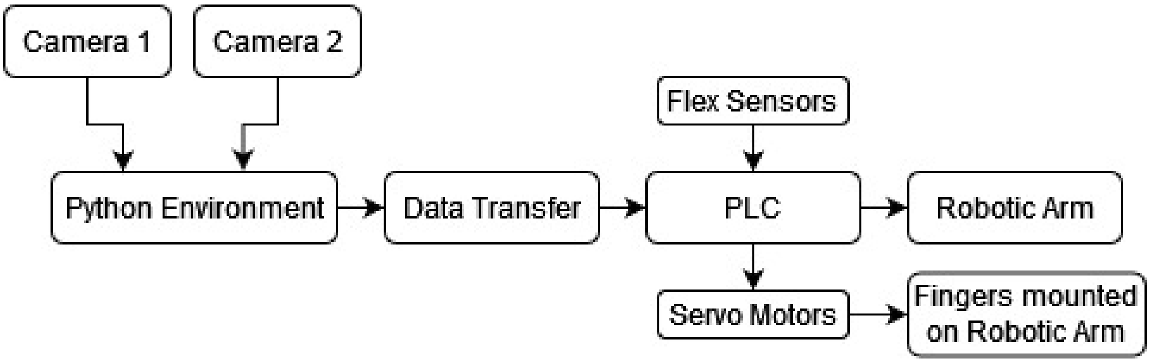

Figure Flowchart of the proposed method

***Corresponding Author:** Mücahid BARSTUĞAN, mbarstugan@ktun.edu.tr

Konya Mühendislik Bilimleri Dergisi, c. 11, s. 2, 423-438, 2023
Konya Journal of Engineering Sciences, v. 11, n. 2, 423-438, 2023
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1228275

DEEP LEARNING BASED HUMAN ROBOT INTERACTION WITH 5G COMMUNICATION

[1*] Mücahid BARSTUĞAN , [2] Zeynep OSMANPAŞAOĞLU

[1] *Konya Technical University, Engineering and Natural Sciences Faculty, Electrical and Electronics Engineering Department, Konya, TÜRKİYE*
[2] *Marmara University, Technology Faculty, Mechatronics Engineering Department, İstanbul, TÜRKİYE*
[1] mbarstugan@ktun.edu.tr, [2] zeynep.osmanpasaoglu@tr.mee.com

(Received: 02.01.2023; Accepted in Revised Form: 20.02.2023)

ABSTRACT: Factories focusing on digital transformation accelerate their production and surpass their competitors by increasing their controllability and efficiency. In this study, the data obtained by image processing with the aim of digital transformation was transferred to the collaborative robot arm with 5G communication and the robot arm was remotely controlled. A 3D-printed humanoid hand is mounted on the end of the robot arm for bin picking. Each finger is controlled by five servo motors. For finger control, the user wore a glove, and the finger positions of the user were transferred to the servo motors thanks to each flex sensor attached to the glove. In this way, the desired pick and place process is provided. The position control of the robot arm was realized with image processing. The gloves worn by the user were determined by two different YOLO (You only look once) methods. YOLOv4 and YOLOv5 algorithms were compared by using Python software language in object detection. While the highest detection accuracy obtained with the YOLOv4 algorithm during the test phase was 99.75% in the front camera, it was 99.83% in the YOLOv5 algorithm; YOLOv4 detection accuracy was the highest in the side camera of 97.59%, and YOLOv5 detection accuracy was 97.9%.

Keywords: *Collaborative Robot Arm, Digital Transformation, Object Detection, Remote Control, YOLO*

1. INTRODUCTION

Digital transformation has entered a new era with artificial intelligence, augmented reality, big data and 3D printer technologies. All data is transferred to digital environment by transforming the production processes of companies from analog (manual) controls to digital (automatic) control. Because digital transformation requires adapting to new conditions, it takes time for a company to complete the digital transformation process. Digital transformation mostly proceeds through automation systems. Automation systems are being developed and transforming into robotic systems. During this period, the need for operators is being eliminated. Robotic systems are evolving to do all the work a human can do. However, there may be a need for the robot control to be controlled by a human. Especially in studies for the disabled, robot control is carried out with brain signals. In a study using camera and EEG sensors, brain signals were controlled by an industrial robot arm [1, 2]. In studies using only cameras, predetermined hand movements were classified by deep learning methods and robot arm control was provided [3-5]. With the advancement of technology, studies involving human-robot interaction are increasing. Thanks to the digital twin of the factories, human-robot interaction can be transferred into the digital environment [6]. Digital twins enable all real-time data in the process to be stored and analyzed in a virtual environment. In this way, the appropriate and safe working environment required for human-robot interaction can be established. Human-robot interaction applications are also used in the fields of agriculture [7] and health [8] apart from production factories. In a study in the field of agriculture, the image taken by the robot is sent to the remote operator. The operator marks the product to be purchased and the robot takes the product [7]. In the study conducted in the field of health, the robot arm supports the doctors in surgeries with eye tracking made from the camera [8]. In the future, these studies will allow patients to have surgery in their own country in hospitals where appropriate

***Corresponding Author:** Mücahid BARSTUĞAN, mbarstugan@ktun.edu.tr

devices are available, without going to the country where the doctor is located. With the spread of 5G technology, communication will accelerate, and remote-control systems will become widespread. In this way, robots with 5G communication will be able to enter areas where humans cannot enter, and robots will be able to do things that humans cannot do with remote control systems. Remote control systems can be a controller, an artificial intelligence algorithm or a system with sensors.

The YOLO method (You Only Look Once), which is one of the popular artificial intelligence algorithms, is used in many different areas. Food [9-12], hand gesture detection [13], construction [14], agriculture [15-17], occupational health and safety [18-20], unmanned aerial vehicle [21-25], object detection [26, 27], medical [28], maritime [29], fire detection [30], traffic sign [31] and mask detection [32, 33]. With this study, real-time applications of the YOLO method are carried out. A summary of the literature studies is presented in Table 1.

Table 1. The literature summary

Paper	Year	Field of the Paper	Dataset	Detected object	Method
[9]	2021	Food	GWHD	Wheat Heads	YOLOv5, Faster-RCNN
[11]	2022	Food	Leaf	Leaf Disease	YOLOv5
[13]	2021	Posture Detection	Custom Dataset	Hand	YOLOv3, YOLOv4, YOLOv5
[14]	2022	Construction	Custom Dataset	Aggregate	YOLOv4, YOLOv5
[16]	2021	Agriculture	Custom Dataset	Insect on Soybean Crop	YOLOv4, YOLOv5
[15]	2022	Agriculture	Plant Village	Leaf Disease	YOLOv4, YOLOv5
[20]	2022	Occupational Safety	Safety Helmet Wearing	Helmet Detection	YOLOv5
[19]	2022	Occupational Safety	SHEL5K	Helmet Detection	YOLOv3, YOLOv4, YOLOv5
[22]	2022	Unmanned Aerial Vehicle	A2G	Drone Images	YOLOv3, YOLOv4, YOLOv5
[26]	2022	Object Detection	Custom Dataset	Pantry Objects	YOLOv5
[27]	2021	Object Detection	KAIST, FLIR	Infrared Image Objects	YOLO-FIRI
[28]	2021	Medical	BRATS	Brain Tumors	YOLOv3, YOLOv4, YOLOv5
[29]	2021	Marine	Maritime	Ship Types	YOLOv4, YOLOv5
[31]	2022	Object Detection	Custom Dataset	Traffic Signs	SSD, YOLOv5
[33]	2022	Mask Detection	ViDMASK	Face Mask	Mask R-CNN, YOLOv4, YOLOv5

The literature studies summarized in Table 1 show that the YOLO algorithm is used in real-time applications and has high success in most applications. In non-real-time artificial intelligence studies, hardware cost and performance, communication speed between products are ignored, and they are trained and tested on ready data sets. In this study, a data set was created from scratch and tested on real-time camera images. The generated algorithm has been optimized by testing the hardware performance and communication speeds of the system. Thanks to the 5G infrastructure, remote control was provided in the installed system. In the Table 1, remote control process was performed in [7] only.

This study established a real-time experimental environment and controlled a collaborative robot arm remotely with high performance. In the application, the basis of remote robot arm control was laid with 5G communication. In the system, the right arm of the operator was monitored with a camera from two different angles and the right arm movement detected by artificial intelligence was transferred to the robot arm. The flex sensor glove worn on the right hand sensed the finger movements of the operator and controlled the fingers mounted on the robot arm. Artificial intelligence data has been transferred to the robot and digital environment with 5G communication. Thanks to the data transferred to the digital environment, the operator can control the robot from different environments. Thanks to this technology, nuclear, chemical, poisonous gas, etc. areas, which people cannot enter, robots will be sent, and tasks will be completed by remote control. Another advantage of this technology will be the transfer of human abilities to the robot. The delicate workmanship of the craftsmen will be copied to the digital environment and more than one robot worker will be able to do the same delicate work.

The study consists of four parts. In Section 2, the introduction of the materials used in the study and the creation of a data set are explained. In Section 3, the results of the system are presented, and the operation of the system is explained. In Section 4, the findings are evaluated and information about future studies is given.

2. MATERIAL AND METHODS

2.1. Material

In this study, a computer with powerful hardware, a human hand model printed with a 3D printer, a collaborative robot arm, PLC, two sim cards attached to two 5G modems, servo motor drivers and two USB cameras (Logitech C270 webcam) were used. High speed metal gear dual ball bearing servo motors were used to control the fingers. The operation voltage of the servo motors is 5V. Each servo has an operation speed as 0.2sn/60o for 4.8V operation voltage. Identical webcams for used in the experiment. The resolution of webcam is 720P at 30 FPS (frame per second).

2.1.1 Human hand model and gloves

The human hand model designed and produced to pick products is mounted on the 6th axis of the robot arm. After the robot arm was moved to a certain position, the bin picking process was carried out by moving the fingers. The flex sensor glove that reads the finger movements and the servo motor block that moves the 3D printed fingers are shown in Figure 1.

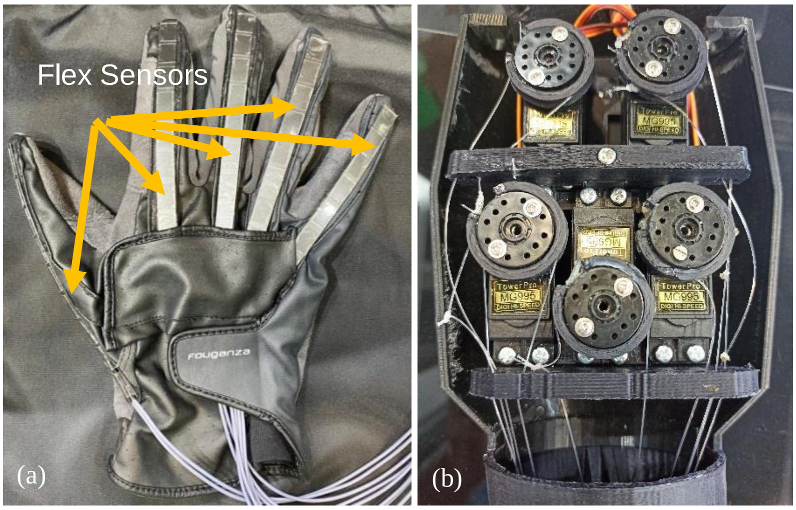

Figure 1. (a) Glove, (b) Servo motors

2.1.2. Collaborative robot arm

As an alternative to standard industrial robots working at high speeds with safety equipment, collaborative robots that operate at lower speeds (250 mm/sec) but can work in the same environment with humans without the need for any safety equipment are preferred in studies involving human interaction. In this study, 6 axis RV-5AS-D model Mitsubishi Electric collaborative robot arm model was used. It has a reach of 910 mm, a carrying capacity of 5 kg and a position repeatability of 0.03 mm. It has advanced torque sensing feature with J5 series Mitsubishi electric servo motors used in each axis, so it stops suddenly in case of any contact during operation. Apart from the collaborative working mode of the robot, there is also a high-speed operation mode (1000 mm/sec), this feature can be used by selecting the robot working area from the scanning sensors or the program. Thanks to the CC-Link IE Field Basic communication built into it, the data exchange in the project is provided with the communication feature over the directly assigned fields with the PLC. The axis characteristics and reach of the robot arm are presented in Table 2.

Table 2. Robot arm reach distance

Axis	Axis Motion Range (Degree)
J1	±240°
J2	±148°
J3	±150°
J4	±240°
J5	±120°
J6	±200°

2.1.3. Experimental setup

The experimental setup in which the remote robot arm control application is carried out is presented in Figure 2. It was established in Experimental Istanbul Expo Center and has fixed lighting in the indoor area.

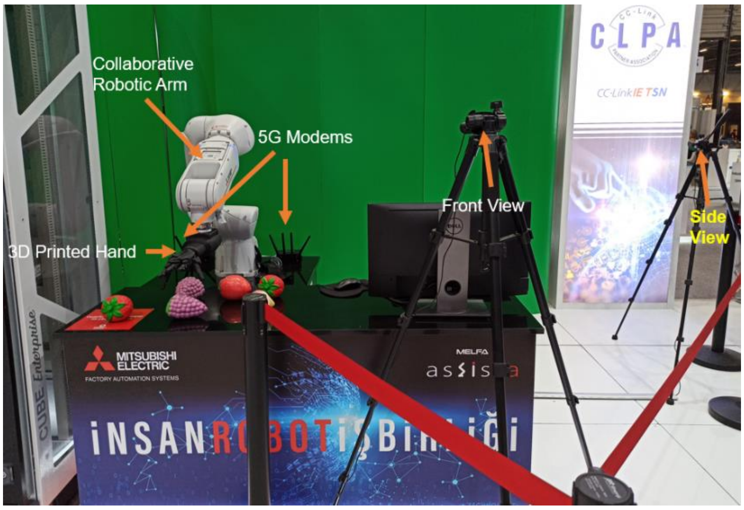

Figure 2. Experimental setup in which the control process is performed

2.1.4. Dataset preparation

The data set used in the study was created from images taken from two separate cameras. Both cameras are identical and 640*480 images were taken. The front-view model was created with the images taken from the first camera, and the side-view model was created with the images taken from the second camera. The images taken from the camera also include the face area, but those parts were cropped while putting them into the paper. Images taken from two cameras at the same time were recorded separately. Target regions were obtained by labeling the obtained images. For front and side views, 808 images were collected for both views from 10 different people and two different models were trained. The tagging process was done on the makesense.ai website and the tags containing the region to be detected were saved as *.txt. Example images and labels (the region to be detected) for front and side views are presented in Figure 3.

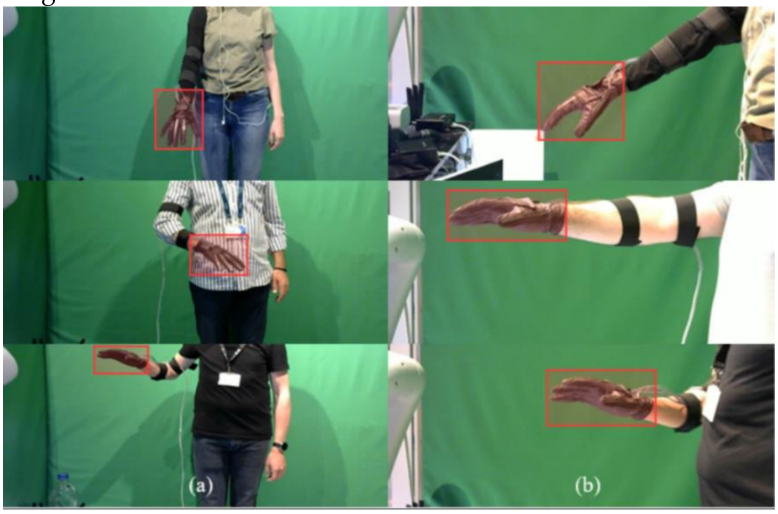

Figure 3. (a) Front view, (b) Side view

2.2. Method

Two different deep learning structures and communication methods were used in the study. YOLOv4 and YOLOv5 structures, which are used as deep learning structures, were compared in terms of training and test performances. The general scheme of the system used in the study is shown in Figure 4.

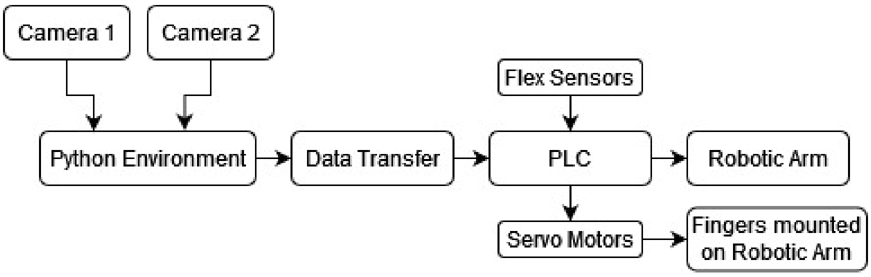

Figure 4. The general scheme of the system

The proposed method uses two cameras for hand position. The detected positions are transferred to PLC over Python. The PLC sends hand position data to the robotic arm. The PLC also reads analog inputs to detect glove finger positions to control the servo motors.

2.2.1. YOLO (You only look once)

Deep learning algorithms can perform object detection, classification, and object recognition with high accuracy. However, these algorithms require high processing power. The YOLO method, on the other hand, is a technique that can detect real-time objects with high accuracy with low processing power and is based on CNN (Convolutional Neural Network). R-CNN model, which is one of the other popular deep learning algorithms, analyzes each region with a network structure after dividing the image into regions. YOLO, on the other hand, presents the image to the structure once and then divides the image into grids. Searches for the object within each grid. In this way, the processing time is shortened, and the accuracy rate increases [34]. Grid and object search stages are shown in Figure 5.

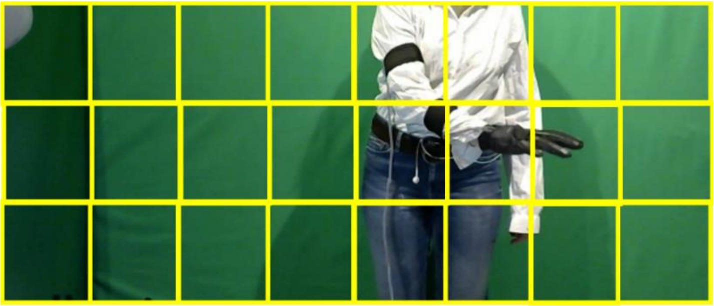

Figure 5. The grids and object searching process

The YOLO algorithm generates a confidence score for each grid. The confidence score includes the probability of detecting the object in the grid and the box surrounding the object. Intersection of Union (IoU), Precision (P), Recall (R), F1-Score, Mean Average Precision (mAP) and detection speed metrics are used for performance measurement of the YOLO algorithm. The IoU metric indicates how closely the tagged region (A) in the training data and the region (B) predicted by the training result intersect. An IoU value above 0.5 indicates a successful detection process. In cases where the IoU value is above 0.5, the frame number is calculated as TP (True Positive), while values below 0.5 are calculated as FP (False Positive). In cases where no frames are found, the frame count is calculated as FN (False Negative). Metric equations appear in Equation 1-5.

$$IoU = \left|\frac{A \cap B}{A \cup B}\right| \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1-score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

$$mAP = \frac{1}{k}\sum_{i=1}^{k} AP_i \text{, where } AP_k = \int_0^1 P_k(R_k)dR_k \tag{5}$$

The YOLOv4 algorithm runs on the Darknet framework and was developed with C/CUDA. Therefore, the Darknet framework has been downloaded and installed in the working environment for training and testing processes. CUDA-GPU connection was made by defining GPU (Graphical Processing Unit). Thanks to this process, the images taken are working on the GPU, and FPS (frame per second) values are obtained by performing high-speed image processing.

The YOLOv5 algorithm emerged shortly after YOLOv4 was published. While YOLOv4 works on Darknet, which was developed in C language, YOLOv5 was developed with PyTorch library. The advantage of YOLOv5 over YOLOv4 is that the trained network file is much smaller and thus training can be performed in a shorter time. The output vector of a YOLO algorithm includes information “p” (object present-not present information in grid), “x center”, “y center”, “width”, “height”, and “class information” [18]. The structure of the YOLOv5 algorithm is presented in Figure 6.

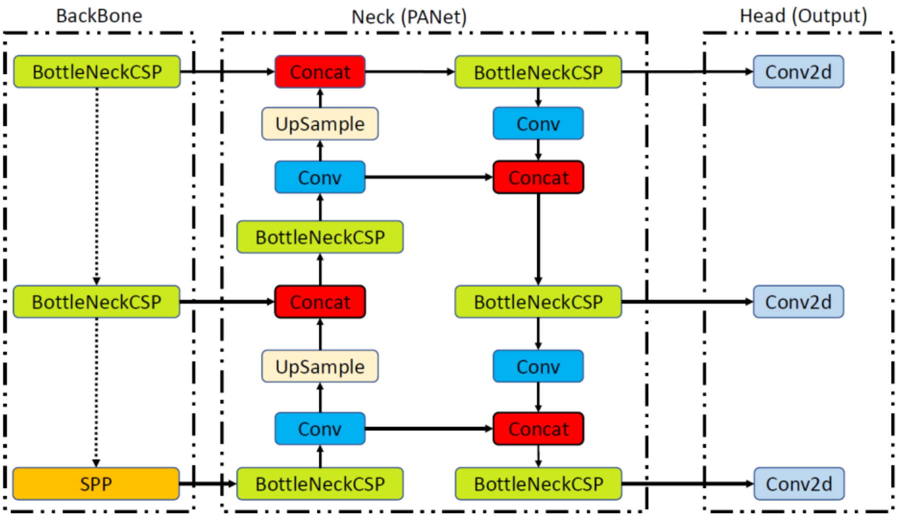

Figure 6. YOLOv5 structure

In the BackBone region seen in Figure 6, the properties of the input images (color, edge, contrast, etc.) are extracted. The model detects objects using these features. The features represent image data with high accuracy and low loss. The Neck area is placed between the BackBone and the Head. Features from the BackBone region are used for the task of detecting possible objects located in the Neck region. There may be many possible objects in the Neck region. The possible objects detected needs a classification process. This classification process is implemented in Head part. Head performs the detection process by classifying the objects [18].

2.2.2. The calibration of camera-robotic arm

Two cameras were used to take images in the experimental environment. The first camera looks at the operator and extracts the hand position information. As a result of the YOLO vector, "x center pixel" information is used to control the "Y axis" of the robot, and "y center pixel" information is used to control the "Z axis" of the robot. The second camera, on the other hand, extracts hand position information by looking from the side. The "x center pixel" information in the YOLO vector result is used to control the "X axis" of the robot. By using the position information from the camera, the robot arm-camera calibration was performed to move the robot arm at the desired level. The "x pixel" change in the first camera is between 250~370 pixels in the image. In this range, the robot arm moves between the position of -180~+300 mm in the "Y-axis". In any information coming out of the range of 250~370, the robot arm does not go beyond the limit of -180~+300 mm. In the first camera, the "y-pixel" change is between 50~450 pixels in the image. In this range, the robot arm moves between +150~+425 mm position in the "Z axis". In any information coming out of the range of 50~450, the robot arm does not go beyond the limit of -150~+425 mm. In the second camera, the "x pixel" change is between 130~400 pixels in the image. In this range, the robot arm moves between the +280~+480 mm position on the "X-axis". In any information coming out of the range of 130~400, the robot arm does not go beyond the limit of +280~+480 mm.

The conversion of pixel information to robot position information in the real world for each axis was done by linear regression and camera pixel value was converted to robot arm position information. The transformation equation and R^2 value obtained for each axis because of the linear regression process are shown in Equation 6-8.

Front View x axis → Robot y axis, $Robot\_y = 4*camera\_x - 1180$, $R^2=1$ (6)

Front View y axis→ Robot z axis, $Robot\_z = 0.6875*camera\_y + 115.63$, $R^2=1$ (7)

Side View x axis → Robot x axis, $Robot\_x = 0.7407*camera\_x + 183.7$, $R^2=1$ (8)

2.2.3. The calibration of glove-servo motors

PLC is one of the most durable hardware products that adapt to the industrial environment and has many features. The purpose of using PLC in this study is to read analog data and move motors with fast outputs. Five flex sensor data on the glove was read with Mitsubishi Electric brand FX5U series PLC and ADP485 analog module. The read data was sent to the servo motors as position information using the fast outputs of the PLC and the servo motors were moved to the desired position (fully open - fully closed - intermediate points). Linear Regression method was used in the conversion of analog information to servo motor position information. In the study, a separate estimation equation was created for each finger by finding the relationship between the analog information read from the flex sensors and the position information. When the operator changes, the analog information changes in the open and closed state of the hand. When the operator wearing the glove was changed, the calibration process was performed using the obtained regression equations and the fingers were fully opened and closed each time. The equations and R2 values obtained for each finger are as follows:

Thumb finer equation → Motor 1 position = Analog input 1* 0.17 + 1600, R^2=1 (9)
Forefinger equation → Motor 2 position = Analog input 2 * 0.19 + 1610, R^2=1 (10)
Middle finger equation → Motor 3 position = Analog input 3* 0.18 + 1662, R^2=1 (11)
Ring finger equation → Motor 4 position = Analog input * 0.17 + 1666, R^2=1 (12)
Little finger equation → Motor 5 position = Analog input * 0.25 + 1548, R^2=1 (13)

2.2.4. The system operation

This section explains the Python-Robot arm communication flow, robot arm movement and hand control with PLC. The detailed scheme of the work carried out is shown in Figure 7.

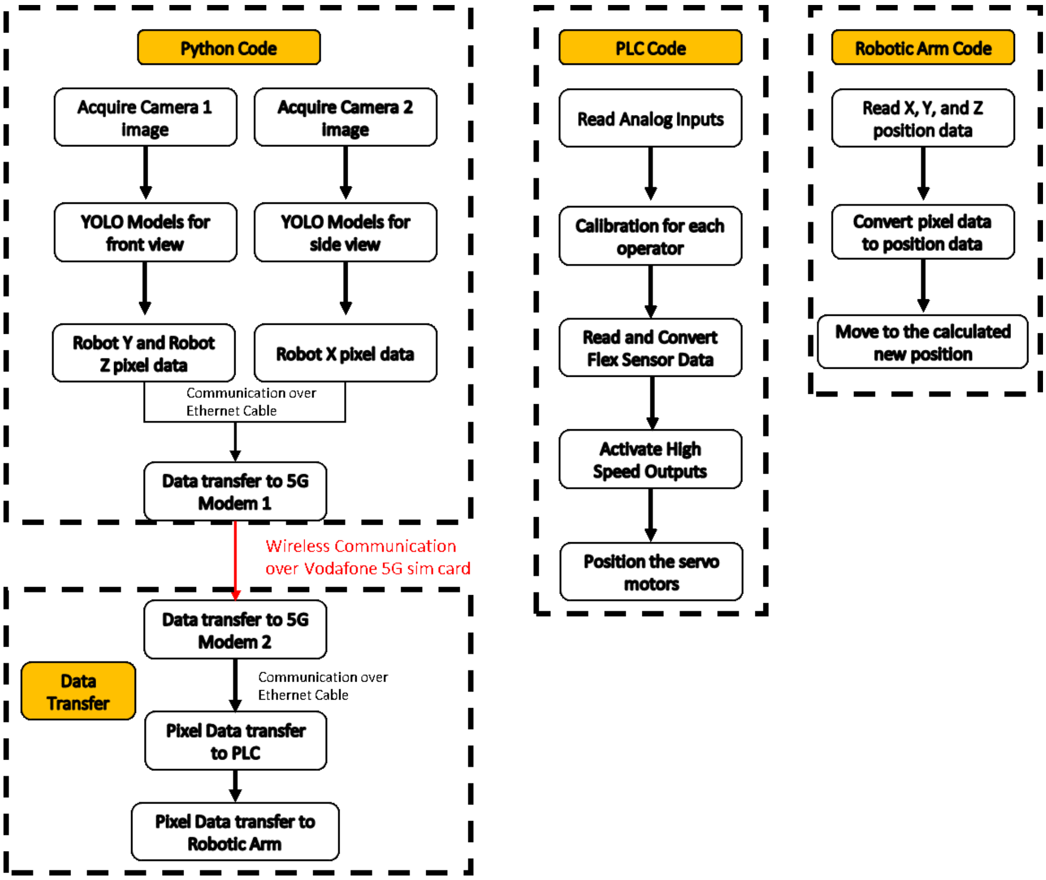

Figure 7. The block diagram of the system setup

After the operator puts on the gloves, first the calibration process is done manually, and the system starts to operate. Two cameras are started to find hand position. The hand is detected by YOLO methods, and the center of the hand area is sent to PLC. Y and Z data is obtained from the Camera 1, and the X data is obtained from Camera 2. X, Y and Z position information from image processing is transferred to Robot Arm via PLC via 5G communication. Two 5G modems were used for data transfer. The first modem is connected to the computer, and the second modem is connected to the PLC. The data is written to the first modem from Python and sent to the second modem. The PLC reads data from the second modem and send to the robot arm. The robot arm reads the incoming position information and moves towards the position. While moving towards the position, when a movement in the opposite direction is received, it stops and moves towards the new position. When there is not any movement, if the operator is standing still in front of the camera, it has been observed that the incoming information varies between 1-4 pixels. Therefore, the robot arm does not execute the motion command unless there is a pixel difference of more than 5 pixels. In this way, the robot arm is prevented from moving while the operator is stationary. When the operator moves his hand over the object to be taken, he closes his fingers and grabs the object and lifts it.

3. RESULTS AND DISCUSSION

In the YOLOv4 training phase, a computer with an Intel i7-8700 processor, Nvidia GTX 1080 graphics card and 32 GB RAM was used, and the training process was carried out on the GPU. The configuration file was configured in accordance with the data set and the training process was carried out. The YOLOv5 model was trained on the Google COLAB system. The Precision, Recall, F1-Score, mAP and IoU values obtained during the training process are presented in Table 3.

Table 3. Training results of YOLOv4 and YOLOv5

Evaluation Metric	Side View		Front View	
	YOLOv4 Result	YOLOv5 Result	YOLOv4 Result	YOLOv5 Result
Precision	1.00	0.975	1.00	0.99964
Recall	1.00	1.00	1.00	1.00
mAP@0.5	1.00	0.99	1.00	0.995
IoU	0.842	0.857	0.8604	0.9238
Training Time	12.52 h (GPU)	4.7 h (COLAB)	12.5 h (GPU)	4.7 h (COLAB)

Table 3 shows that YOLOv5 showed higher performance than YOLOv4 in front and side camera image training. After the training process was completed, the models obtained for the two cameras were recorded. Each model has been tested in real time for both cameras. Visual test results of images from two cameras for YOLOv4 and YOLOv5 are presented in Figure 8-11.

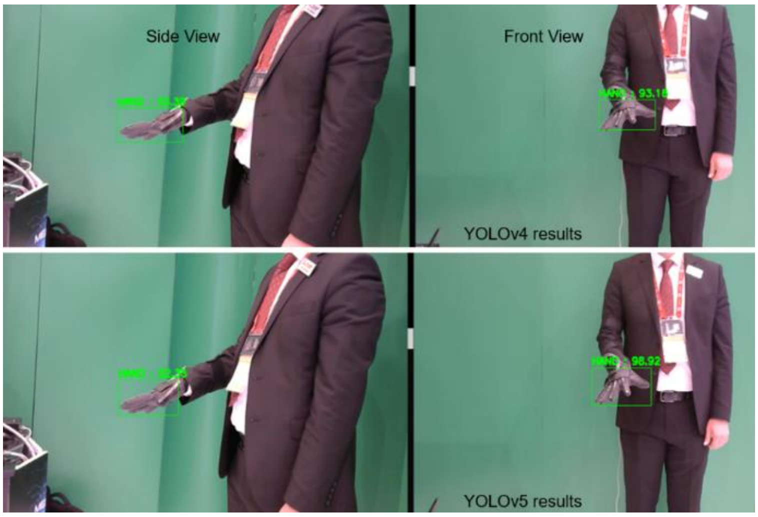

Figure 8. Comparative results for Test 1

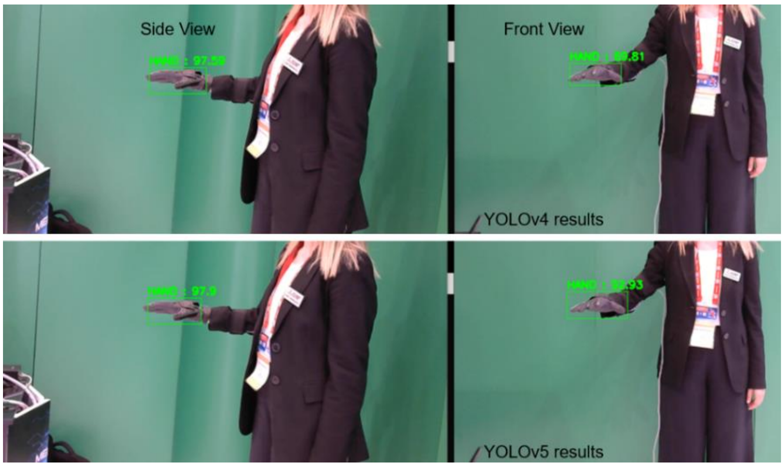

Figure 9. Comparative results for Test 2

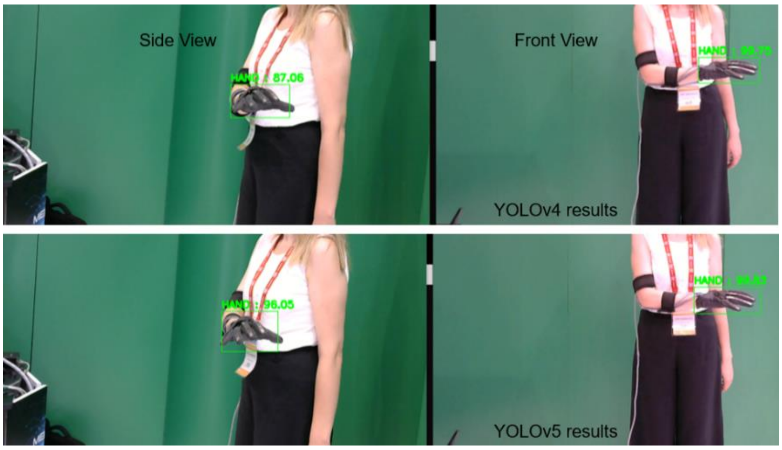

Figure 10. Comparative results for Test 3

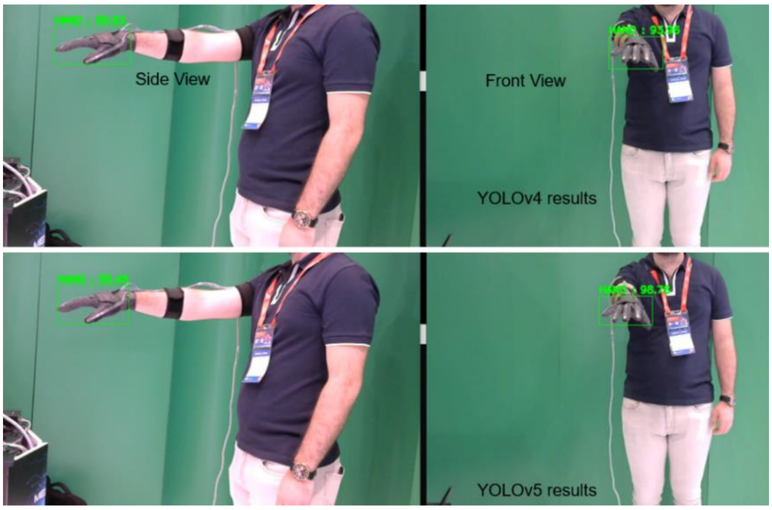

Figure 11. Comparative results for Test 4

Although training environments of YOLOv4 and YOLOv5 are different, the two models were tested on the same environment (Nvidia GTX 1080 graphics card). Four of the test results obtained in real time are presented in Table 4.

Table 4. Test results of YOLOv4 and YOLOv5

Evaluation Metric (Confidence Score)	Side View		Front View	
	YOLOv4 Result	YOLOv5 Result	YOLOv4 Result	YOLOv5 Result
Test 1	55.39	82.35	93.16	98.92
Test 2	97.59	97.9	69.81	92.93
Test 3	87.06	96.06	99.75	99.83
Test 4	89.93	95.45	93.96	98.75
FPS (CPU)	5.55	7.14	5.55	7.14
FPS (GPU)	44.54	81.3	44.54	81.3

Table 4 shows that the YOLOv5 algorithm gave higher performance than YOLOv4 in terms of training and test success on the dataset used in this study. While detection rates of 55.39% and 69.81% could be seen in YOLOv4, the detection rate did not fall below 80% in YOLOv5. Looking at the FPS numbers, it has been reported that the graphics card performs approximately 10 times the operation on the hardware where this work is performed, according to the processor. During the testing process, robot arm control and bin picking process are presented step by step in Figure 12.

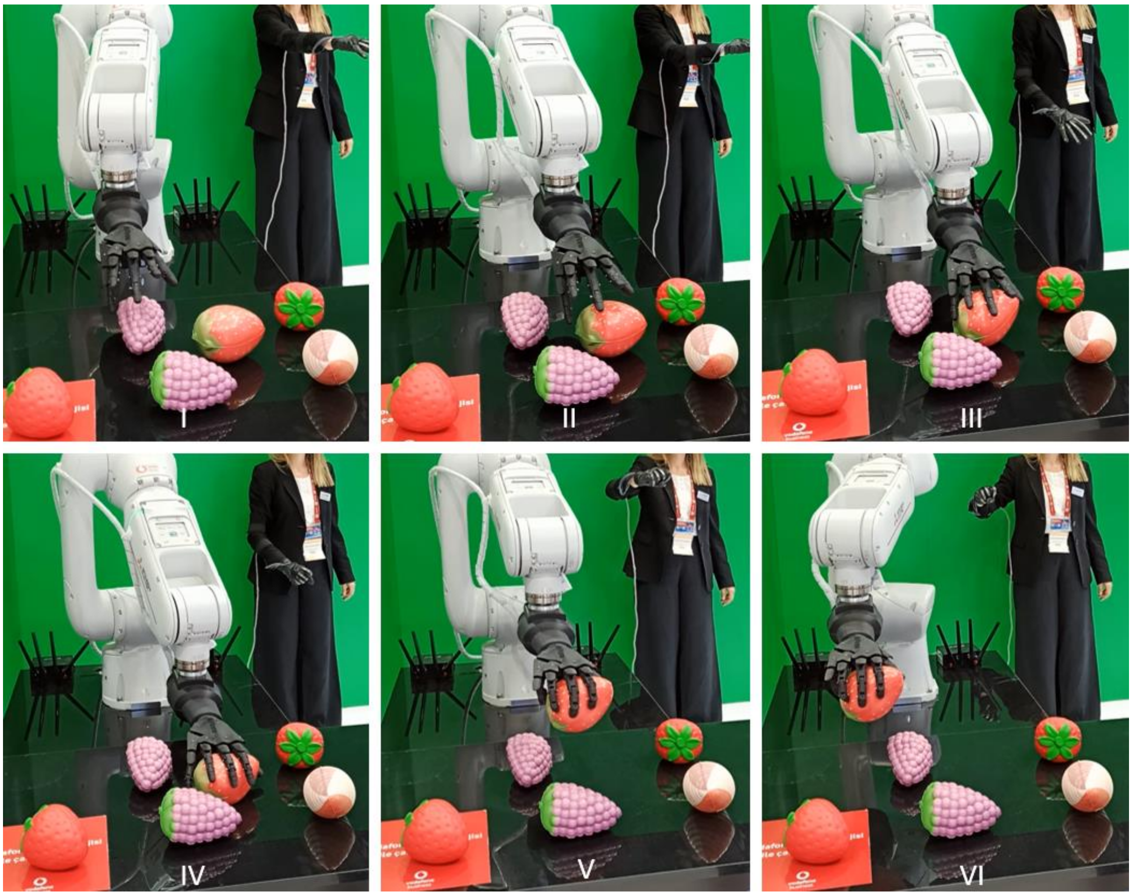

Figure 12. Robot motion according to the arm position

I. The operator directs the robot arm.
II. The robot arm moves in the desired direction and is located top of the bin.
III. When the robot arm is on top of the bin, the operator lowers the robot arm.
IV. When the robot arm comes over the product, the operator closes his hand and picks the bin.
V. Operator raises the closed hand and robot arm lifts bin.
VI. The operator directs the robot and place the bin at the drop zone

4. CONCLUSIONS

With 5G communication, remote quick control operations are made possible. By sending robots to dangerous areas where human hands cannot enter or where humans cannot go, it will be possible to solve problems more quickly and easily. When the problem-solving ability of humans and the knowledge and experience of the workforce are taught to robots, it will be possible for robots to work in the same environment with humans. For this purpose, the factories of the future should focus on the flexible development of production technologies for humans to work with robots. With the new phase of factory automation, production theories in factories will change. Trained and controllable collaborative robots will take part in the development of production theories. With this study, the infrastructure works of the factories of the future have started. Since the factories of the future will be equipped with artificial intelligence systems, the integration of artificial intelligence-factory automation has become an important issue. The YOLO algorithm, which is used in various fields and shows real-time high performance, is an algorithm that can be integrated into factory automation. In Table 5, YOLO methods studied on different subjects and successful results are presented.

In the next stage of this R&D work, the servo motors in the hand model will be replaced with industrial servo motors. The hand model extracted from the 3D printer will be updated. In this study, 3-axis control has been provided, and in future studies, necessary software and hardware additions will be made for 6-axis control. In addition, a sensor will be placed on the fingertip in each hand model and a haptic sensor hand structure will be developed, which will provide feedback to the user. For remote control, virtual reality glasses, in which camera images are transferred to the user, will be added to the system and the user and robot environment will be made independent of each other.

Table 5. The results of literature studies

Paper	Detected object	Method	mAP (%)
[9]	Wheat Heads	YOLOv5	71.92
		Faster-RCNN	70.85
[11]	Leaf Disease	YOLOv5	94.48
[13]	Hand	YOLOv3	71.2
		YOLOv4	72.3
		YOLOv5	75.3
[14]	Aggregate	YOLOv4	99.62
		YOLOv5	99.6
[16]	Insect on Soybean Crop	YOLOv4	94
		YOLOv5	99.5
[15]	Leaf Disease	YOLOv4	90.7
		YOLOv5	90.7
[20]	Helmet Detection	YOLOv5	93.8
[19]	Helmet Detection	YOLOv3	41.7
		YOLOv4	76.93
		YOLOv5	86.87
[22]	Drone Images	YOLOv3	82.3
		YOLOv4	81.5
		YOLOv5	91.3
[26]	Pantry Objects	YOLOv5	94
[27]	Infrared Image Objects	YOLO-FIRI	98.5
[28]	Brain Tumors	YOLOv3	84.3
		YOLOv4	88.71
		YOLOv5	95.07
[29]	Ship Types	YOLOv4	49.5
		YOLOv5	56.8
[31]	Traffic Signs	SSD	90.14
		YOLOv5	97.7
[33]	Face Mask	Mask R-CNN	74.71
		YOLOv4	68.22
		YOLOv5	65.5
This study	Hand Dataset	YOLOv4	86.04
		YOLOv5	92.38

Declaration of Ethical Standards

The authors declare that they have no declaration of ethical standards.

Credit Authorship Contribution Statement

The first author has contributed the writing of the paper, hand detection software, and PLC code.

The second author has contributed the writing of the robotic arm code.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

The first author reports equipment, drugs, or supplies and travel were provided by Mitsubishi Electric Turkey. The second author reports administrative support and equipment, drugs, or supplies were provided by Mitsubishi Electric Turkey

Funding / Acknowledgements

We would like to thank Can Tolga Bizel, who supported us in the realization of this work, and Mitsubishi Electric Türkiye, who funded us to use the products.

Data Availability

The research data is not available.

10. REFERENCES

[1] X. Chen, X. Huang, Y. Wang, and X. Gao, "Combination of augmented reality based brain-computer interface and computer vision for high-level control of a robotic arm," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 28, no. 12, pp. 3140-3147, 2020.

[2] Z. Zhang, Y. Huang, S. Chen, J. Qu, X. Pan, T. Yu, and Y. Li, "An intention-driven semi-autonomous intelligent robotic system for drinking," *Frontiers in Neurorobotics*, vol. 11, p. 1-14, 2017.

[3] S. M. Achari, S. G. Mirji, C. P. Desai, M. S. Hulasogi, and S. P. Awari, "Gesture based wireless control of robotic hand using image processing," *International Research Journal of Engineering and Technology*, vol. 5, no. 5, pp. 3340-3345, 2018.

[4] J. O. P. Arenas, R. J. Moreno, and R. D. H. Beleño, "Convolutional neural network with a dag architecture for control of a robotic arm by means of hand gestures," *Contemporary Engineering Sciences*, vol. 11, no. 12, pp. 547-557, 2018.

[5] P. Atre, S. Bhagat, N. Pooniwala, and P. Shah, "Efficient and feasible gesture controlled robotic arm," in *2018 Second International Conference on Intelligent Computing and Control Systems*, 2018, pp. 1-6: IEEE.

[6] A. A. Malik and A. J. R. Brem, "Digital twins for collaborative robots: A case study in human-robot interaction," *Robotics and Computer-Integrated Manufacturing*, vol. 68, pp. 1-16, 2021.

[7] J. P. Vasconez, G. A. Kantor, and F. A. A. Cheein, "Human–robot interaction in agriculture: A survey and current challenges," *Biosystems engineering*, vol. 179, pp. 35-48, 2019.

[8] K. Fujii, G. Gras, A. Salerno, and G.-Z. Yang, "Gaze gesture based human robot interaction for laparoscopic surgery," *Medical image analysis*, vol. 44, pp. 196-214, 2018.

[9] M. N. Datta, Y. Rathi, and M. Eliazer, "Wheat heads detection using deep learning algorithms," *Annals of the Romanian Society for Cell Biology*, vol. 25, no.5, pp. 5641-5654, 2021.

[10] F. Jubayer, J. A. Soeb, A. N. Mojumder, M. K. Paul, P. Barua, S. Kayshar, S. S. Akter, M. Rahman, and A. Islam, "Detection of mold on the food surface using YOLOv5," *Detection of mold on the food surface using YOLOv5*, vol. 4, pp. 724-728, 2021.

[11] M. Karakaya, M. F. Celebi, A. E. Gök, and S. Ersoy, "Discovery Of Agricultural Diseases By Deep Learning And Object Detection," *Environmental Engineering and Management Journal*, vol. 21, no. 1, pp. 163-173, 2022.

[12] R. Li and Y. J. E. Wu, "Improved YOLO v5 Wheat Ear Detection Algorithm Based on Attention Mechanism," *Electronics*, vol. 11, no. 11, pp. 1673-1694, 2022.

[13] C. A. Owusu-Agyei and J. Hou, "Hands Activities Detection in Egocentric Interactions Using YOLOv5," in *2021 International Conference on UK-China Emerging Technologies*, 2021, pp. 199-203: IEEE.

[14] S. Yuan, Y. Du, M. Liu, S. Yue, B. Li, and H. Zhang, "YOLOv5-Ytiny: A Miniature Aggregate Detection and Classification Model," *Electronics*, vol. 11, no. 11, pp. 1743-1758, 2022.

[15] M. P. Mathew and T. Y. Mahesh, "Leaf-based disease detection in bell pepper plant using YOLO v5," *Signal, Image and Video Processing*, vol. 16, no. 3, pp. 841-847, 2022.

[16] S. Verma, S. Tripathi, A. Singh, M. Ojha, and R. R. Saxena, "Insect Detection and Identification using YOLO Algorithms on Soybean Crop," in *TENCON 2021-2021 IEEE Region 10 Conference*, 2021, pp. 272-277: IEEE.

[17] B. Yan, P. Fan, X. Lei, Z. Liu, and F. Yang, "A real-time apple targets detection method for picking robot based on improved YOLOv5," *Remote Sensing*, vol. 13, no. 9, pp. 1619-1642, 2021.

[18] Y. Hathat, D. Samai, A. Benlamoudi, K. Bensid, and A. Taleb-Ahmed, "SNCF workers detection in the railway environment based on improved YOLO v5," in *7th International Conference on Image and Signal Processing and their Applications*, 2022, pp. 1-7: IEEE.

[19] M.E. Otgonbold, M. Gochoo, F. Alnajjar, L. Ali, T. H. Tan, J. W. Hsieh, and P. Y. Chen, "SHEL5K: an extended dataset and benchmarking for safety helmet detection," *Sensors*, vol. 22, no. 6, pp. 2315-2338, 2022.

[20] M. Sadiq, S. Masood, and O. Pal, "FD-YOLOv5: A Fuzzy Image Enhancement Based Robust Object Detection Model for Safety Helmet Detection," *International Journal of Fuzzy Systems*, vol. 24, pp. 2600-2616, 2022.

[21] K. Ding, X. Li, W. Guo, and L. Wu, "Improved object detection algorithm for drone-captured dataset based on yolov5," in *2nd International Conference on Consumer Electronics and Computer Engineering*, 2022, pp. 895-899: IEEE.

[22] L. Li, Z. Yao, Z. Miao, X. Qiu, and X. Yang, "YOLO-A2G: An air-to-ground high-precision object detection algorithm based on YOLOv5," in *6th International Conference on Machine Vision and Information Technology*, 2022, pp. 1-8:IOP Publishing.

[23] U. Nepal and H. Eslamiat, "Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs," *Sensors*, vol. 22, no. 2, pp. 464-479, 2022.

[24] V. Slyusar, M. Protsenko, A. Chernukha, V. Melkin, O. Biloborodov, M. Samoilenko, O. Kravchenko, H. Kalynychenko, A. Rohovyi, and M. Soloshchuk, "Improving The Model Of Object Detection On Aerial Photographs And Video In Unmanned Aerial Systems," *Eastern-European Journal of Enterprise Technologies*, vol. 1, no. 9, pp. 115-126, 2022.

[25] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2778-2788.

[26] L. A. J. Abel, T. C. N. Oconer, and J. C. D. Cruz, "Realtime Object Detection of Pantry Objects Using YOLOv5 Transfer Learning in Varying Lighting and Orientation," in *2nd International Conference on Innovative Research in Applied Science, Engineering and Technology*, 2022, pp. 1-7: IEEE.

[27] S. Li, Y. Li, Y. Li, M. Li, and X. Xu, "YOLO-FIRI: Improved YOLOv5 for Infrared Image Object Detection," *IEEE Access*, vol. 9, pp. 141861-141875, 2021.

[28] N. M. Dipu, S. A. Shohan, and K. A. Salam, "Brain Tumor Detection Using Various Deep Learning Algorithms," in *International Conference on Science & Contemporary Technologies*, 2021, pp. 1-6: IEEE.

[29] J.-H. Kim, N. Kim, Y. W. Park, and C. S. Won, "Object Detection and Classification Based on YOLO-V5 with Improved Maritime Dataset," *Journal of Marine Sciences and Engineering*, vol. 10, no. 3, pp. 377-391, 2022.

[30] W. Wahyono, A. Harjoko, A. Dharmawan, G. Kosala, and P. Y. Pranata, "A Comparison of Deep Learning Methods for Vision-based Fire Detection in Surveillance System," in *The 5th International Conference on Future Networks & Distributed Systems*, 2021, pp. 1-7.

[31] Y. Zhu and W. Q. Yan, "Traffic sign recognition based on deep learning," *Multimedia Tools and Applications*, vol. 81, no. 13, pp. 17779-17791, 2022.

[32] A. S. Hemanth, "Face Mask Detection Using YOLOv5," *International Journal of Novel Research Development*, vol. 7, no. 5, pp. 390-395, 2022.

[33] N. Ottakath, O. Elharrouss, N. Almaadeed, S. A. Maadeed, A. Mohamed, T. Khattab, K. Abualsaud," ViDMASK dataset for face mask detection with social distance measurement," *Displays*, vol. 73, p. 102235, 2022.

[34] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv:2004.10934*, 2020.