Motion Recognition System for Mobile Robot in Logistics Center

Trinh Tran Vinh An ∗ · *Bui Minh Hau* ∗∗ · † *Hwan-Seong Kim*

∗*Master Course, Department of Ocean Convergence Logistics Innovation, Korea Maritime and Ocean University, Busan, 49112, KOREA*
∗∗*PhD. Course, Department of Ocean Convergence Logistics Innovation, Korea Maritime and Ocean University, Busan, 49112, KOREA*
†*Professor, Dept. of Logistics, Korea Maritime and Ocean University, Busan, 49112, KOREA*

Abstract : *Motion recognition systems are crucial for mobile robots, particularly in logistics centers, where they need to interact with other robots or workers. Mobile robots require the capability to recognize movement patterns of surrounding objects to avoid collisions and maintain safety. Employing computer vision, machine learning, and deep learning techniques, cameras have become powerful sensors for object recognition and tracking. In this research, the YOLOv8 (You Only Look Once) model is utilized to detect and track object movements. The training data were gathered from videos of people and boxes in a lab setting. The system operates in real-time to meet the collision avoidance needs of mobile robots. The data captured by the camera is processed to analyze detected objects' movement patterns. The results demonstrate the success in real-time motion recognition and the capability of providing safety alerts when a tracked object enters the robot's safety perimeter.*

Key words : *computer vision, motion recognition, object tracking, mobile robot, logistics center*

1. Introduction

Along with the development of global supply chains and e-commerce, the flow of goods in Logistics centers is also increasing. The robot system has been applied in the Logistics Centers to operate continuously and ensure timely import and export of goods. However, safety and performance are always important when operating an automatic robot system. Managers need to pay attention to safety issues between workers and robots and between robots.

To make the robot environment safer, each robot needs to be equipped with modern technology to help identify and analyze the surrounding environment. With the development of technology, IoT sensors can help robots communicate and exchange information. AI algorithms have also been strongly developed to help robots better understand their surrounding environment. (Muftygendhis et al., 2022) designed a communication system for robots through the internet and robot operating system (ROS) so that robots can recognize each other's location. Information sharing helps the robot choose its path and avoid collisions. Lee has developed a cyber-physical system that uses the cloud to receive and store information about the robot's speed, state, and position. This information is used as input to the A∗ path-finding algorithm, which helps the robot's routes become more

optimal and continuously updated. IoT sensors and AI algorithms have helped mobile robots more widely perceive the surrounding work area and increase the ability to exchange information between robots in the system. (Vinh et al., 2023) proposed the methodology that combined 2 AI algorithms, object detection, and distance estimate from the camera, with the 3D map to visualize the vehicle's surrounding working place. This method allows managers to monitor vehicle workflows closely, enhancing vehicle safety and performance. (Syntakas et al., 2022) combined two laser sensors and a camera to develop an object detection and navigation algorithm for mobile robots.

In designing safety solutions for mobile robots, besides object detection, object motion recognition also provides a large amount of information for AI algorithms to predict and avoid collisions. When a mobile robot moves in the logistics center, there are many obstacles on its movement path. To avoid these obstacles, the mobile robot must recognize the obstacle's presence and motion to avoid collisions. The motion recognition systems in this paper consist of two parts: safety zone detection and motion direction recognition. Safety zone detection is used to detect or predict object violation with the mobile's robot safety zone based on the object's motion. The motion directions show the direction of the mobile robot's movement. These two modules provide mobile robots with more understanding of the objects around them to avoid

---

† Corresponding author, kimhs@kmou.ac.kr 051)410-4334

collisions and maintain safe operation. The motion recognition system is used to identify the movement and movement trends of objects obtained from the camera and the information if the object moves too near the robot Besides, it also provides the robot with a data package to make optimal decisions in collision situations. The motion recognition system works by identifying objects in front of the mobile robot and predicting the movement direction (left, right, up, down, or no movement) of that object.

This research aims to deploy a motion recognition system for mobile robots based on data received from the monocular camera. A mini PC that runs image processing models controls the robot and camera. The data obtained from the camera will be processed by the mini PC and return motion analysis and the object's movement direction in front of the camera.

However, to recognize the movement of objects, the system needs to identify objects that need motion analysis. In this research, people and boxes must be analyzed for motion because these objects commonly appear in the Logistic Center. Image data of these two subjects was collected from model samples in the laboratory.

This article uses the YOLOv8 object detection and tracking model to build the motion recognition system for detecting the motion and object's movement direction and also the notice if the object moves to the safety zone of the robot. In the experimental section of this paper, training results and practical test results are presented to clarify the implementation method and the achievable prediction results. Figure 1 shows the location of the camera installed on the mobile robot. The camera is installed in an unobstructed location to provide the best visibility and light.
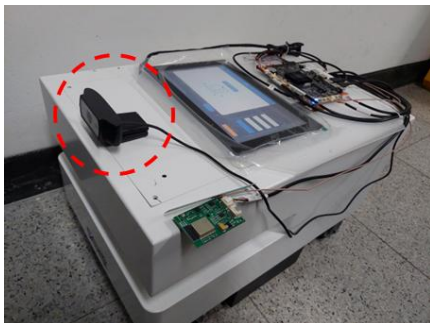

Fig, 1 Photo Camera on mobile robot

The structure of this article includes five sections. Section 2 discusses related research works and improvements of the motion recognition system in this paper. Section 3 presents how to process data and apply the YOLOv8 model to build the motion recognition system in this research. Section 4 gives the results of experimental cases tested in a laboratory to ensure the system's accuracy. Finally, Section 5 concludes.

2. Related Work

Object detection plays an important role in building modern automated systems. It helps the system identify the surrounding environment to make smarter and more optimal decisions. Although object detection models have been developed for a long time, the number of research articles on detection models is still very large today. (Lou et al., 2023) developed a new model from YOLOv8, called DC-YOLOv8, to detect the small size of objects in a special problem. (Chen et al., 2023) proposed a new methodology, DiffusionDet, to detect high-noise images. (Afdhal et al., 2023) applied the YOLOv8 model for real-time object detection in self-driving cars. Their application worked in a mixed-traffic environment with an accuracy of 0.8 in daylight scenarios. (Reinard et al., 2023) researched evaluating the accuracy of the YOLOv8 model in detecting objects at different distances.

Besides identifying objects, automatic systems also need to track the movement of objects. When the system determines the object's direction of movement, it can provide optimal solutions in cases where it is necessary to avoid obstacles, follow the sample, or track the object. (Luo et al., 2019) developed a mobile robot system that uses a stereo camera, IMU sensor, and machine learning algorithms to detect and track targets in real time. (Zhou et al., 2014) proposed a method that uses a ground plane projection (GPP) from an RGB-D camera to detect and track objects in indoor environments, helping minimize distortion and noise. This method allows the detection and tracking of targets in real-time with high accuracy. (Xu et al., 2023) proposed a lightweight 3D dynamic obstacle detection and tracking (DODT) method based on RGB-D cameras to meet the low computational requirements of small robots such as UAVs with limited capacity integrated computer technology. (Wang et al., 2023) developed a method called StrongFusionMOT to track multi-objects at the same time. This method needs to combine the data from LiDAR and the camera to implement. (Zheng et al., 2021) introduced AlphaTrack, a simple, effective algorithm for 3D object detection and

tracking that combines location information and changes in object appearance. (Xiao et al., 2022) proposed a target-tracking algorithm for mobile robots, combining a monocular camera and a 2D lidar sensor. (Zhu et al., 2023) used the YOLOv8CSM-tiny model, developed from YOLOv8, and correlation filtering of body key points BKPs-CF to make the human following function for mobile robots.

Most research articles on object detection and object tracking for autonomous vehicles use combined data from different types of sensors such as LiDAR, monocular camera, stereo camera, depth camera, or IMU to enhance the capabilities of tracking the movement of objects. However, this research aims to apply to mobile robots in Logistics Centers, where the operating environment, such as lighting, layout, and work area, is consistently maintained stably. Besides, the central processor for mobile robots also has limited memory capacity. Therefore, this research aims to use only one sensor, the monocular camera, for data collection and use the model's object detection and object tracking features to build a motion recognition system for the mobile robot. In addition, instead of just displaying a bounding box following the object that needs to be tracked like other research articles, the motion recognition system in this study also shows the object's direction of movement. The system also alerts if an object is too close to the robot. The main contributions of this research are as follows.

- Develop a motion recognition system based on a camera
- Determine the direction of movement of the tracked object
- Set a warning system if an object moves into the mobile robot's safety zone

3. Proposed Methodology

Motion recognition helps the robot recognize the object's motion by using the trajectory information obtained from the object detection and tracking method. From this trajectory, the change in object coordinate through the sequence of frames is used to recognize object motion. In this study, the YOLOv8 model is used, which is a popular model in object detection and tracking. From this combination, the trajectory of the objects is obtained and manipulated for the motion recognition method. This paper proposes motion recognition analysis based on the

trajectory from the object detection and tracking method – a computer vision method. In other papers, object tracking is used for obtaining the trajectory of the object but not using it for object motion recognition. This combination helps the autonomous mobile robot in the warehouse, which needs more understanding of the objects around them to avoid collision and maintain safe operations. The pipeline of the method includes two main parts: the first part is the object detection and tracking part, and the second part is the motion recognition part. Figure 2 shows the pipeline method. The input to the network is a frame sequence, which is taken from the camera. The object detection and tracking module processes this input to obtain the object trajectory. Then, the motion recognition module analyzes the object's trajectory to obtain the object's motion as output.
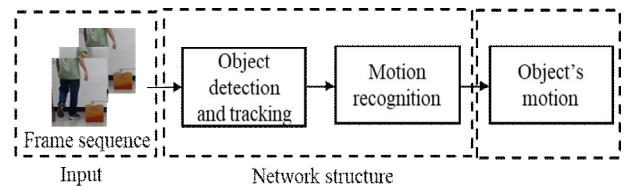

Fig. 2 The motion recognition pipeline

3.1 Object Detection and Tracking

The object detection module locates the object by bounding a box around it, classifying its type in the object, and tracking part. The tracking part links or matches the detected objects across multiple frames to perform the tracking capability process. The YOLOv8 model has embedded the object detection model with the tracking model. Figure 3 shows the blue trajectory points produced by the object detection and tracking module. The object detection and tracking module includes the object detection and tracking parts. These two combine to provide the object's trajectory, as shown in Figure 3
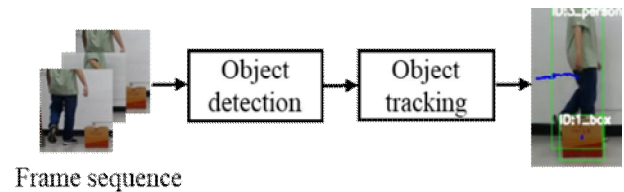

Fig. 3 Object detection and tracking part

The YOLOv8 model's training process as shown in Figure 4 includes four steps: the first is collecting data, and the second is processing the data by labeling it. In

this labeling process, the object is bounded by a box and classified as its type. As shown in Figure 3, the person is bound by the green box and classified as a person. The third step is training the model, and the last step is evaluating the model.

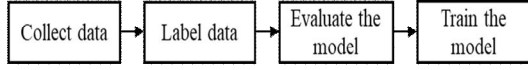Collect data → Label data → Evaluate the model → Train the model

Fig. 4 Training process

YOLO environment provides automated creation evaluations for Precision and Recall. Precision is the correct predilection overall predictions. It is the ratio of correct positive predictions and overall positive predictions by the model:

$$P = \frac{TP}{TP+FP} \tag{1}$$

Where P is the precision, TP is the number of objects correctly detected, FP is the number of objects incorrectly detected. The higher the precision, the more accurate the detection by the model. The recall is the ratio of correct object predictions to all the objects presented in the image. The recall is defined as:

$$R = \frac{TP}{TP+FN} \tag{2}$$

Where R is the recall, FN is the number of objects presented but the model failed to detect. High Recall means the model can find almost all the objects in the image. The object is correctly detected based on the Intersection over Union (IoU) between the predicted and ground truth bounding boxes. The IoU is defined as:
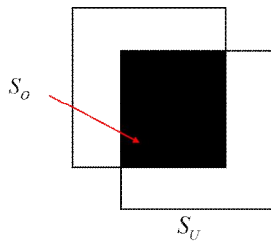
$$IoU = \frac{S_O}{S_U} \tag{3}$$

Fig. 5 Intersection over Union (IoU)

In the object detection model, two popular standards used to evaluate the effectiveness of the trained model are mAP50 and mAP50-95. The standard mAP50 is the mean average precession at the IoU threshold of 0.5, and the standard mAP50-95 is the mean average precession at the IoU threshold from 0.5 to 0.95.

3.2 Motion recognition

After obtaining the object's detection and tracking part, the motion recognition system receives the object trajectory as a sequence of points. Equation (4) shows the format of the object trajectory.

$$T = \{(x_0,y_0),(x_1,y_1),\cdots(x_n,y_n)\} \tag{4}$$

Where the trajectory T stores the last n center position of the bounding boxes around the object across the frame. (x_i,y_i) is the center coordinate $(0 \leq i \leq n)$, x_i is the horizontal coordinate of the object, y_i is the vertical coordinate of the object.

From this trajectory, the motion recognition system can obtain the motion recognition of objects. The research provides two functions in the motion recognition system: safe zone detection and object motion direction. In the safe zone detection function, the system used object bounding box location to determine if the object falls into the zone and used the trajectory to predict if the next motion of the object will fall into the zone. On the other hand, the object motion directions function shows the object movement status, such as stop, moving right, left, up, or down in the image plane.

For the first function, the system will define a safety zone inside the image frame collected from the camera. The safety zone is defined as the region in front of the mobile robot, as shown in Figure 6, the green area. If the object bounding box falls into this zone, the robot can recognize it, and the linear regression model is used to predict the next bounding box location in the next frames and then detect if it will fall into the safety zone.
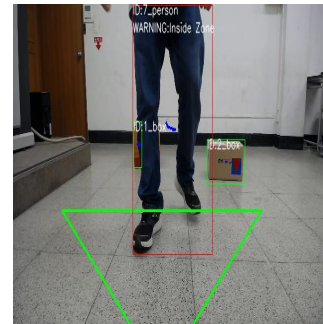
Fig. 6 Safety zone region is presented by green color

The center of the bounding box for the next frame can be predicted as:

$$\hat{x}_f = \beta_{x0} + \beta_{x1} \cdot f$$
$$\hat{y}_f = \beta_{y0} + \beta_{y1} \cdot f \tag{5}$$

where:

- $\hat{x}_f$ and $\hat{y}_f$ are predicted horizontal and vertical center coordinates of the object at point $f(f>n)$
- β_{x0} and β_{y0} are the intercepts for the x and y coordinates
- β_{x1} and β_{y1} are the slopes for the x and y coordinates

The parameter of the regression model is obtained by training the regression model with the center points from the trajectory. For each center point i, a tuple of center coordinates (x_i,y_i) is fitted to a linear regression model with the form:

$$\hat{x}(i)=\beta_{x0}+\beta_{x1}.i$$
$$\hat{y}(i)=\beta_{y0}+\beta_{y1}.i \tag{6}$$

The least square method is used to calculate the parameters of the regression model. The slope parameters β_{x1} and β_{y1} are calculated as follows:

$$\begin{cases} \beta_{x1}=\dfrac{\sum\limits_{i=1}^{n}(i-\bar{i})(x_i-\bar{x})}{\sum\limits_{i=1}^{n}(i-\bar{i})^2} \\ \beta_{y1}=\dfrac{\sum\limits_{i=1}^{n}(i-\bar{i})(y_i-\bar{y})}{\sum\limits_{i=1}^{n}(i-\bar{i})^2} \end{cases} \tag{7}$$

where $\bar{i}$ is calculated as the equation (8):

$$\bar{i}=\frac{\sum\limits_{i=1}^{n}i}{n} \tag{8}$$

$\bar{x}$ is mean of observed x coordinate:

$$\bar{x}=\frac{\sum\limits_{i=1}^{n}x_i}{n} \tag{9}$$

$\bar{y}$ is mean of observed y coordinate:

$$\bar{y}=\frac{\sum\limits_{i=1}^{n}y_i}{n} \tag{10}$$

The intercept β_{x0} and β_{y0} are calculated as follows:

$$\begin{cases} \beta_{x0}=\bar{x}-\beta_{x1}.\bar{i} \\ \beta_{y0}=\bar{y}-\beta_{y1}.\bar{i} \end{cases} \tag{11}$$

Then, using the predicted future center $(\hat{x}_f,\hat{y}_f)$ and the width and height of the bounding boxes at points n ,check if this predicted bounding box intersects with the safety zone region.

The second function is recognizing object motion, direction, and status. For the sequences of length trajectory, the changing coordinates of moving objects change significantly due to the movement of objects in the image plane. For recognition of objects moving or not moving objects, the trajectory of objects is manipulated by detecting any direction changes between consecutive points in the trajectory.

The displacement between two consecutive points on the trajectory (x_i,y_i) and (x_i,y_i) is defined as the difference between their coordinates:

$$\triangle x_i=x_{i+1}-x_i$$
$$\triangle y_i=y_{i+1}-y_i \tag{12}$$

For the sequences of frames, an object's movement is determined by calculating this average displacement over the trajectory. The average displacement is calculated based on horizontal and vertical average displacements. These average displacements are defined as:

$$\triangle d_x=\frac{1}{n}\sum\limits_{i=1}^{i=n}\triangle x_i$$
$$\triangle d_y=\frac{1}{n}\sum\limits_{i=1}^{i=n}\triangle y_i \tag{13}$$

Where $\triangle d_x$ is the horizontal average displacement, $\triangle d_y$ is the vertical average displacement, and n is the number of trajectory points. This average displacement is compared with a threshold to determine whether it is moving or not moving. When the object is moving, the direction is obtained based on the average displacement. Each object will have an average change in displacement. A threshold value is set to determine the motion of the object. If the object is not moving, the change in both average displacements does not have a significant change, which means that both changes in direction are smaller than the threshold value. If the object is moving, eight directions of the object are defined. If the change in $\triangle d_x$ is greater than the threshold value, the object has left-right motion. If the change in $\triangle d_y$ is greater than the threshold value, the object has up-down motion. The threshold is set to 1, meaning there is a small change by 1-pixel value on subsequent frames.

4. Experiment Results

4.1 Object detection training

For the experiment, the environment, including person and box, is chosen for simulation, which are popular objects in the Logistic center environment. Multiple images are collected from the front view camera attached to the

front of the robot to capture the front view environment. 250 images contain boxes, and the person is used to train; the image size is reduced to 640x640 images. These data are divided into three parts: 198 images for training, 19 for validation, and 9 for testing the model's efficiency.

Fig. 7 Original images data

The model YOLOv8n is used for the training process. The aim is to train for 250 epochs, but after 151 epochs, the loss is stable, so the model stops training. The model is trained on a computer with an NVIDIA GeForce RTX 4060 Ti GPU. The trained model is evaluated based on equation (3) (4). Figure 8 shows that the precision and recall curve approaches are closer to 1.
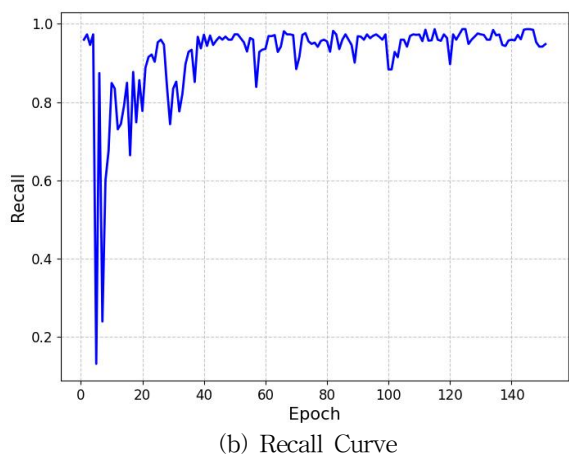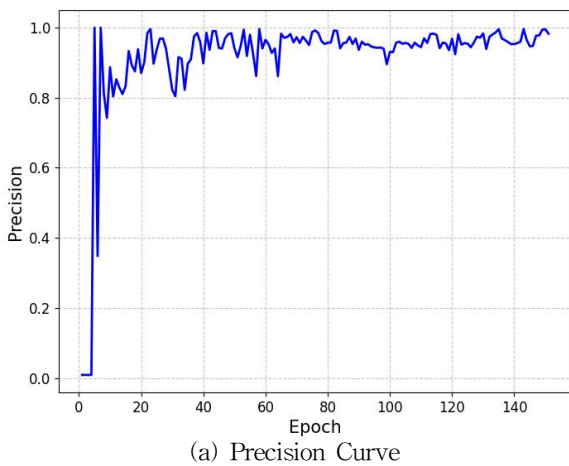

(a) Precision Curve

(b) Recall Curve

Fig. 8 Precision and Recall Curves

The model is evaluated based on mAP50 and mAP50-95 standards. Figure 9 shows the mAP50 and mAP50-95

curves. The horizontal axis is the training iteration, and the vertical axis is the precision. After several training iterations, the mAP50 is closer to 1, while the mAP50-95 is around 0.8.
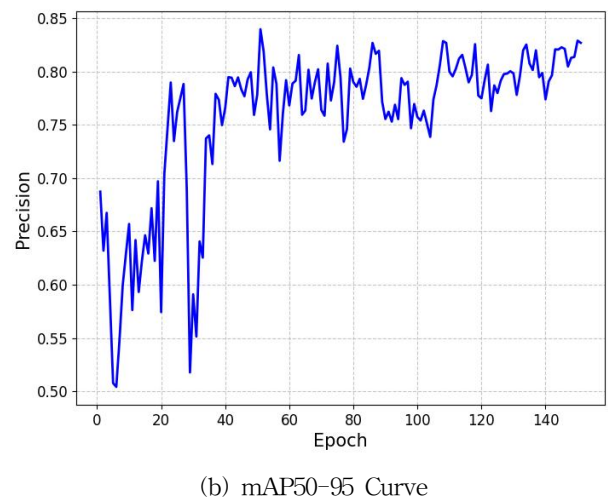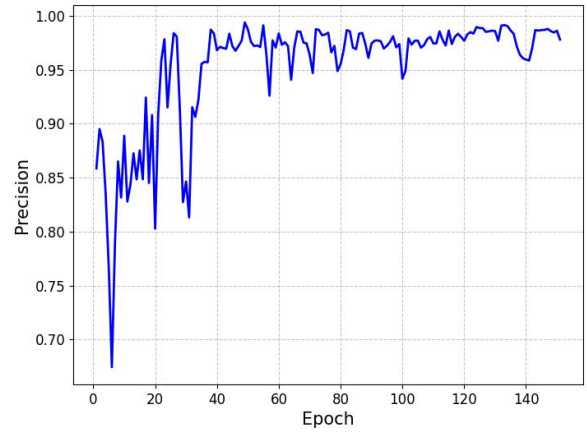

(a) mAP50 Curve

(b) mAP50-95 Curve

Fig. 9 The evaluation curves

4.2 Motion recognition

After training, the model is tested with video sequences for motion recognition. As proposed, motion recognition has two functions: first is safety zone, and second is motion status. These two functions are shown by processing through a video that captures the movement of a mobile robot in a simple environment. Figure 10 presents the first function of safety zone detection results. The mobile robot can recognize a person's trajectory, predict if the person is going into the green safety zone, and warn if the object falls into the zone. The second function is motion direction recognition, as shown in Figure 11. This function also shows the direction movement of the person while recognizing the stationary of two bounding boxes.

In Figure 10, the safety zone detection is simulated. If the object does not violate the safety zone region, which is the green trapezium, the bounding box around these objects is colored green. When the object moves, the trajectory is presented as the blue points; as shown in Figure 10, the point trajectory attaching to the person shows that the person is moving. When the person is moving towards the safety zone, the yellow bounding boxes enclose the person, and a warning "Moving Towards Zone" is attached to the person. When the person is inside the safety zone, the person is enclosed by the red bounding boxes with the red warning "Inside Zone."
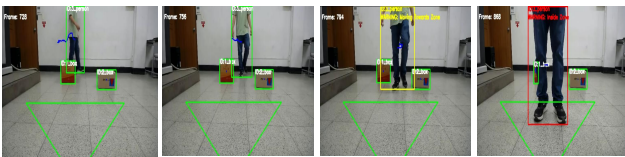

Fig. 10 Safety zone detection

In Figure 11, this simulation shows the results of direction motion recognition. The moving object is enclosed in the blue box. The box is enclosed in the green box. For each object, the status of the movement is shown. With the box, since the box has no movement, the object status is shown as not moving. With the person, since the person has movement, the direction is shown by text (Left, Right, Up, Down), and it is also presented by the red arrow, which shows the movement direction of the object.
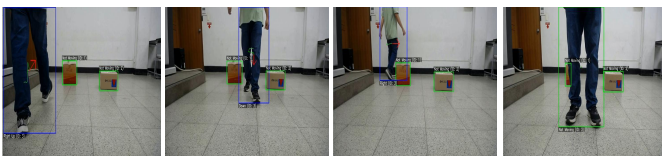

Fig. 11 Motion direction recognition

5. Conclusion

This paper proposes motion recognition for a mobile robot system in the Logistic Center by applying object detection and tracking methods. By analyzing the trajectory from the object detection and tracking method, the motion information of objects is obtained; this provides the mobile robot with the motion recognition of the object. The motion recognition of the object has two functions. One function is safety zone detection; this is the warning system for mobile robots if the objects fall into their safety zone; it also predicts if the object will fall into the safety zone in the next motion. The second function is direction motion, which shows how the object moves. Through experiments in a simple environment, by using the trained YOLOv8 model with high precision, the motion recognition can successfully perform the two-motion recognition function. In the evaluation, the trained model shows high precision and recall; both are closer to 1, as shown in the Precision and Recall cover. The trained model also has the mAP50 closer to 1 and the mAP50-95 around 0.8. These evaluators show the effectiveness of the object detection model after the training process.

In conclusion, this paper has proposed a motion recognition solution for mobile robots by analyzing the object detection and tracking result. This solution can provide more information about objects, which helps the robot avoid obstacles more efficiently and improves its safety performance.

Acknowledgement

References

[1] Muftygendhis, R., Shiang, W. J. and Hsu, C. H.(2022), "A Study of AGV Collaboration with Iternet of Things Concept for Collision Avoidance at Warehouse Intersection", RSF Conference Series: Engineering and Technology, Vol. 2 (1).

[2] Lee, C. K. M., Lin, B., Ng, K. K. H., Lv, Y., Tai W. C.(2019), "Smart Robotic Mobile Fulfillment System with Dynamic Conflict-free Strategies Considering Cyber-physical Integration", Advanced Engineering Informatics, Vol. 42, 100998.

[3] Vinh, N. Q., Park, J. H., Shin, H. S. and Kim, H. S.(2023), "3D Mapping for Improving the Safety of Autonomous Driving in Container Terminals", Journal of Navigation and Port Research, Vol. 47 (5), pp. 281-287.

[4] Syntakas, S., Vlachos, K. and Likas, A.(2022), "Object Detection and Navigation of a Mobile Robot by Fusing Laser and Camera Information", 2022 30th

Mediterranean Conference on Control and Automation (MED), pp. 557-563.

[5] Lou, H., Duan, X., Guo, J., Liu, H., Gu, J., Bi, L. and Chen, H.(2023), "DC-YOLOv8: Small-Size Object Detection Algorithm Based on Camera Sensor", Electronics, 12, 2323.

[6] Chen, S., Sun, P., Song, Y. and Luo, P.(2023), "DiffusionDet: Diffusion Model for Object Detection", 2023 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 19773-19786.

[7] Afdhal, A., Saddami, K., Sugiarto, S., Fuadi, Z. and Nasaruddin, N.(2023), "Real-Time Object Detection Performance of YOLOv8 Models for Self-Driving Cars in a Mixed Traffic Environment", 2023 2nd International Conference on Computer System, Information Technology and Electrical Engineering (COSITE), pp. 260-265.

[8] Reinard, V., Kristiano, Y. and Wulandari, M.(2023), "Distance and Accuracy in Object Detection Based on YOLOv8 Computer Vision Algorithm", International Journal of Application on Sciences, Technology and Engineering (IJASTE), Vol. 1 (3).

[9] Luo, W., Xiao, Z., Ebel, H. and Eberhard, P.(2019), "Stereo Vision-based Autonomous Target Detection and Tracking on an Omnidirectional Mobile Robot", In Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2019), pp. 268-275.

[10] Zhou, Y., Yang, Y., Yi, M., Bai, X., Liu, W. and Latecki, L. J.(2014), "Online Multiple Targets Detection and Tracking from Mobile Robot in Cluttered Indoor Environments with Depth Camera", International Journal of Pattern Recognition and Artificial Intelligence, Vol. 28 (1).

[11] Xu, Z., Zhan, X., Xiu, Y., Suzuki, C. and Shimada, K.(2023), "Onboard Dynamic-Object Detection and Tracking for Autonomous Robot Navigation With RGB-D Camera", IEEE Robotics and Automation Letters, Vol. 9, pp. 651-658.

[12] Wang, X., Fu, C., He, J., Wang, S. and Wang, J.(2023), "StrongFusionMOT: A Multi-Object Tracking Method Based on LiDAR-Camera Fusion", IEEE Sensors Journal, Vol. 23 (11), pp. 11241-11252.

[13] Zeng, Y., Ma, C., Zhu, M., Fan, Z. and Yang, X.(2021), "Cross-Modal 3D Object Detection and Tracking for Auto-Driving," 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, pp. 3850-3857.

[14] Xiao, Y., Peng, Z., Chen, L., Deng, Y., Li, Z. and Lei, X.(2022), "Combining Monocular Camera and 2D Lidar for Target Tracking Using Deep Convolution Neural Network based Detection and Tracking Algorithm," 2022 International Conference on Frontiers of Communications, Information System and Data Science (CISDS), pp. 122-127.

[15] Zhu, Y., Zhong, T., Wang, Y., Kan, J., Dong, F. and Chen, K.(2023), "Mobile Robot Tracking Method Based on Improved YOLOv8 Pedestrian Detection Algorithm", 2023 2nd International Conference on Machine Learning, Cloud Computing and Intelligent Mining (MLCCIM).