

Research Article

# EuroPallet Detection with RGB-D Camera Based on Deep Learning

Sedanur Kirci<sup>1\*</sup>, Gokhan Atali<sup>2</sup>

<sup>1</sup> Kar Metal, Orcid ID: <https://orcid.org/0000-0001-5089-9243>, E-mail: [yapayzeka@karmetal.com.tr](mailto:yapayzeka@karmetal.com.tr)

<sup>2</sup> Sakarya University of Applied Sciences, Department of Mechatronics Engineering, Orcid ID:  
<https://orcid.org/0000-0003-1215-9249>, E-mail: [gatali@subu.edu.tr](mailto:gatali@subu.edu.tr)

\* Correspondence: [yapayzeka@karmetal.com.tr](mailto:yapayzeka@karmetal.com.tr)

(First received October 02, 2023 and in final form December 22, 2023)

**Reference:** Kirci, S., Atali, G. EuroPallet Detection with RGB-D Camera Based on Deep Learning. The European Journal of Research and Development, 3(4), 55-65.

## Abstract

*This study related to enhancing processes and production efficiency in industrial settings by opting for autonomous mobile robots dedicated to indoor transportation and logistics. The specific focus of this research is on employing deep learning for the detection of Euro pallet objects, enabling a mobile robot, specifically the Servant-T1500 model by Kar Metal company with natural navigation capabilities, to autonomously dock and handle palletized loads with precision.*

*To accomplish this, an original dataset was curated using an RGB-D camera, and data augmentation techniques were applied to expand this dataset. Subsequently, a deep learning model was trained on this data to detect Euro pallets in images it had not encountered during training. The transfer learning method was applied using the YOLOv5 model on the dataset. The successful outcome of this process demonstrates the autonomous robot's capability to accurately recognize Euro pallet objects even under changing lighting conditions. This achievement marks a significant step toward optimizing industrial processes and improving production efficiency through the integration of autonomous mobile robots in logistics and transportation tasks.*

**Keywords:** *Pallet Detection, Deep Learning, Industrial Application, Autonomous Mobile Robot*

## 1. Introduction

In recent years, with the development of computer vision technology, the integration of image processing applications in industry has increased [1, 2, 3]. Minimizing errors caused by human influence in the industrial area and optimizing processes provide commercial gains. Robotic systems are among the most crucial components of the digital

transformation processes in the industrial area. Robotic systems are frequently employed in welding, cutting, quality control, monitoring material transport, and storage areas, contributing to increased efficiency by accelerating these processes. Mobile robots, known as AGVs (Automatic Guided Vehicles) and AMRs (Autonomous Mobile Robots), are commonly utilized, especially for storage and logistics purposes. AGVs are generally systems that follow a predefined path with the assistance of factors such as QR codes [4], reflectors [5], or lines [6]. The area of deployment should be adapted to accommodate AGVs, and the maintenance of the auxiliary elements used in the area should be performed periodically. Therefore, AGVs incur installation costs and may be subject to certain working conditions. On the other hand, AMRs are mobile robots with superior path planning capabilities compared to AGVs; they autonomously plan their paths independent of environmental factors. AMRs first map the environment in which they will operate and can react to dynamic and static obstacles on the path plans they generate by positioning themselves within their map. Another advantage of AMRs over AGVs is their ability to quickly adapt to flexible and changing workflows, made possible by comprehensive sensors and advanced software.

In factories, the process of transporting materials from one station to another is commonly carried out using pallet jacks and forklifts. Forklifts and pallet jacks are designed to lift loads such as pallets easily. While these vehicles provide significant functionality when used within a factory, they can also lead to operator-induced accidents. In recent years, autonomous vehicle technology has been introduced to automate these vehicles and eliminate human influence [7, 8, 9, 10]. In their study, Syu et al. integrated image processing into factory logistics and added virtual vision algorithms to forklifts. They introduced a new approach for pallet detection using the Adaptive Structure Feature (ASF) and Direction Weighted Overlap (DWO) ratio, eliminating dynamic backgrounds to enhance processing efficiency. With this hybrid algorithmic approach, their proposed average pallet detection rate increased by 95% [11]. Li et al. utilized a Convolutional Neural Network (CNN) architecture, a deep learning approach, for pallet detection. They labeled a dataset containing a single class and trained their neural network to achieve successful results [12]. Shao et al. used an RGB-D camera to detect euro pallet objects, performing pallet detection from point cloud data. They proposed a new method by combining color and geometric shape data [13]. Upon reviewing the literature, it is observed that mobile robots are being increasingly employed for in-factory transportation and are still under development.

In this study, a pallet detection system has been developed for a mobile robot using the YOLOv5 deep learning model. An original data set has been created and the YOLOv5 model has been trained with this data set. The results of this trained model and real-world analyses have been given.

## 2. Materials and Methods

This study encompasses deep learning-based pallet detection using an RGB-D camera on an autonomous vehicle. The presented software runs on the Servant-T1500 model autonomous mobile robot, developed by Kar Metal company, equipped with natural navigation capabilities. The visual representation of the autonomous mobile robot is provided in Figure 1.



Figure 1: SERVANT-T1500 AMR

One of the significant tasks of the Servant-T1500 model autonomous mobile robot is the precise detection and localization of Euro pallet objects. Servant-T1500 uses artificial intelligence sub-disciplines, including deep learning and image processing techniques, to perform precise detection of a pallet. For a mobile robot to execute identification processes like these, it requires sensors that can collect data about the environment. Cameras are among the primary sensors for mobile robots. Cameras enable a mobile robot to collect two- or three-dimensional data, allowing it to recognize information such as the sizes, distances, and colors of objects in its environment. Various types of cameras with different working principles have been developed using different technological disciplines. RGB, depth, thermal and wide-angle cameras can be given as examples of camera types in which various technological disciplines have been used. Depth cameras are preferred to obtain information about dynamic obstacles and distances around the robot. Depth cameras are categorized into different types based on their working methods. The most commonly used depth cameras in mobile robots are stereo cameras and Time of Flight (ToF) cameras. Stereo cameras, inspired by the human eye, collect data from two cameras, combine this data, and generate distance information. ToF cameras, on the other hand, measure the time it takes for light or another signal to travel to an object and back, creating distance information.

The data obtained with ToF and Stereo cameras are comparatively presented in Figure 2. In Figure 2. data is acquired using a stereo camera, specifically the ZED2, while in Figure

2. b, data is obtained from a ToF camera, namely the Realsense d435i. As observed in Figure 2, in terms of distance accuracy, the Realsense camera exhibits more stable performance, leading to the preference for the Realsense d435i. Additionally, this camera is compatible with the ROS operating system, capable of distance measurements in the range of 0.2 to 10 meters, and provides data in the form of a point cloud.

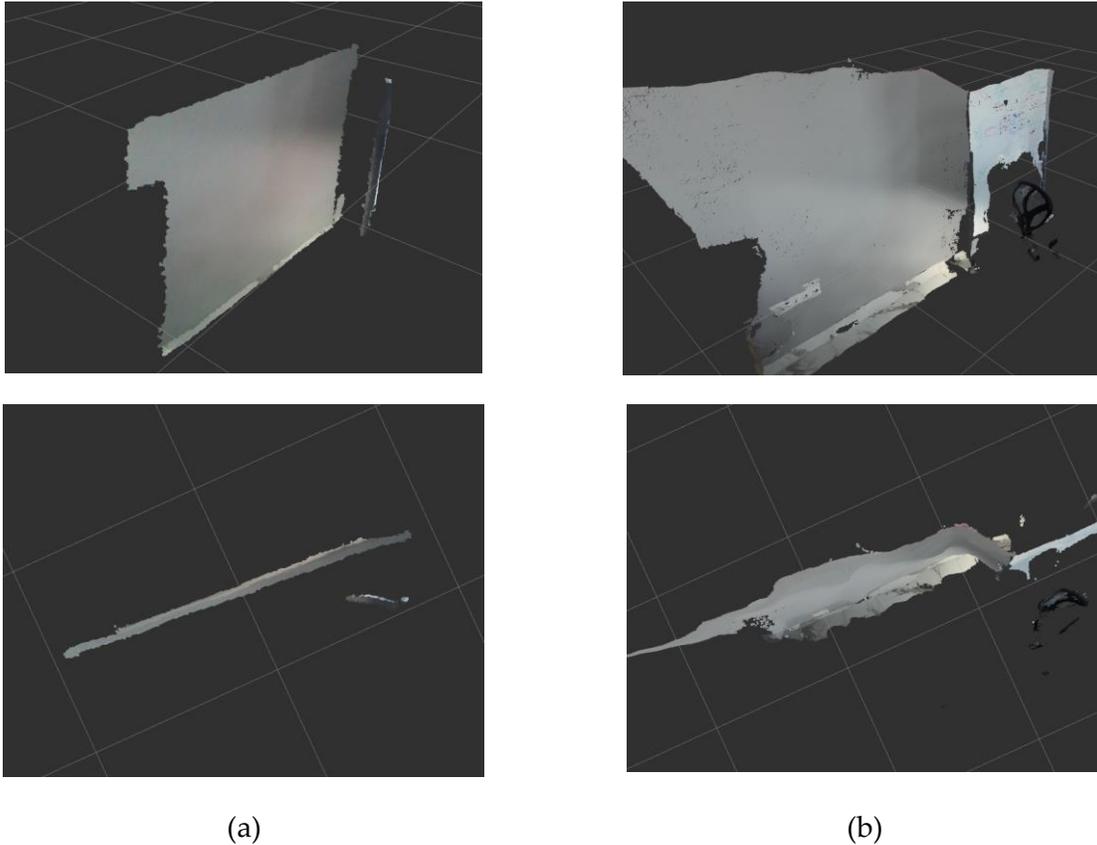


Figure 2: a). Obtained Data from ToF Camera, b) Obtained Data from Stereo Camera

The first stage in the task of positioning the Euro pallet object on a map and enabling autonomous vehicle precise dock to this pallet involves the detection of the docking side of the pallet on the image. Various filters from classical image processing methods can be employed to detect an object in an image. In this study, it was observed that for the precise detection of the pockets of the Euro pallet object, a method capable of recognition over a broader range should be employed, considering the continuous variation in lighting conditions and possible staining on the pallets. Based on the training data and the YOLOv5 model, a deep learning model has been trained to make detections on images it had never seen before.

## 2.1. Dataset

While creating the dataset for the training of the deep learning model, videos were recorded in a factory with randomly placed pallets. These videos were saved as individual frames every 15 frames, resulting in a total of 650 images. Data augmentation was applied by introducing variations in brightness and RGB values in these images. Following the data augmentation process, the dataset, comprising 2600 images, was divided into 2271 training, 219 validation, and 110 test images. All the used images share the same resolution (96dpi) and size (1920x1080). This dataset is labeled into three classes: pallet front face (class 0), pallet pocket (class 1), and pallet side face (class 2). The distribution of labels in the dataset, with a total of 7864 labels, is as shown in Figure 3. In the study, considering that the autonomous mobile robot will always see the front face of the pallet when autonomously approaching the pallet station, the under-represented position of the pallet side face class in the dataset is not critical.

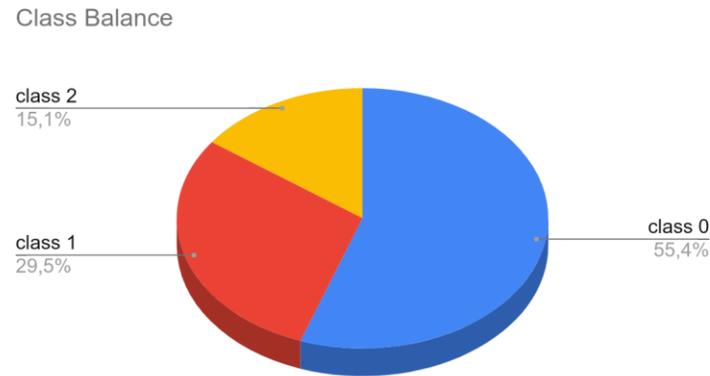


Figure 3: Distribution of Labels Over Classes

Several examples of the original dataset subjected to data augmentation processes are provided in Figure 4. Images in Figure 4.a, 4.d represent the raw data captured, while images in Figure 4.b, 4c, represent changes in brightness. Images in Figure 4.h. and 4.i. represent as examples of color variations in the RGB scale.

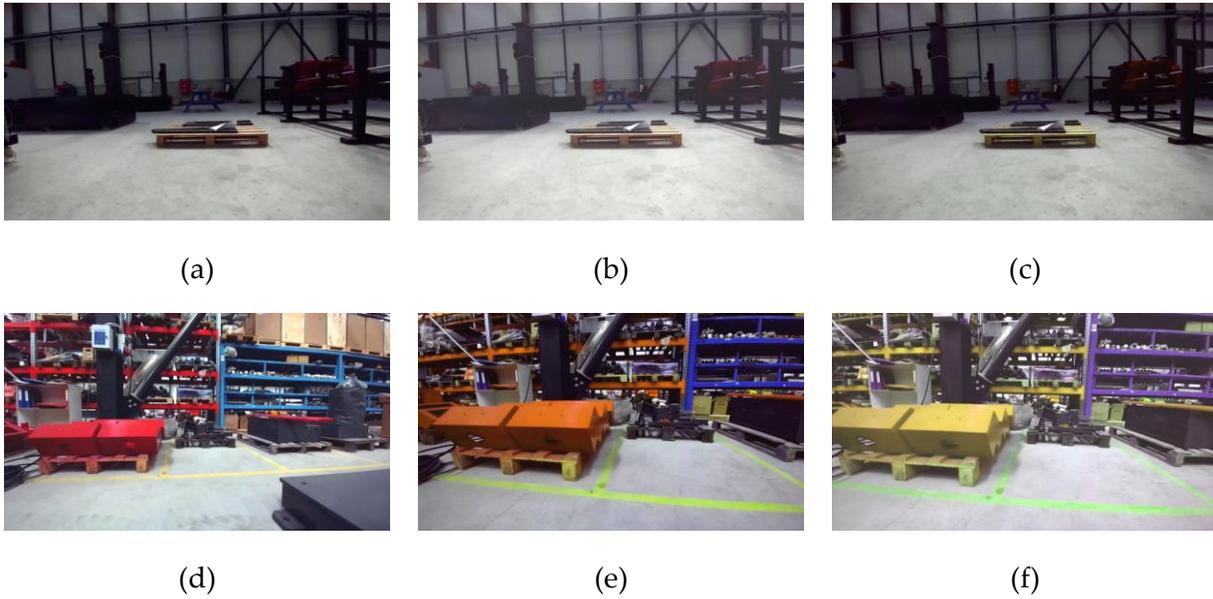


Figure 4: Example Images in the Data Set (a,d: raw data, b,c: changes in brightness, e,f: changes in colors)

## 2.2. YOLOv5

The YOLO (You Only Look Once) algorithm [14], which has seen several versions since its publication, is a one-stage object detector capable of object detection and tracking using convolutional neural networks. The YOLO algorithm can operate quickly as it predicts the class and coordinates of all objects in an image by passing through the neural network in a single forward pass. YOLO treats the object detection problem as a single regression problem. The working principle of YOLO first divides the image into  $S \times S$  grids, where these grids can be of sizes  $3 \times 3$ ,  $5 \times 5$ , and  $19 \times 19$ . After passing through the neural network, the resulting vector in the input image contains information such as whether there is an object in the area, whether its center is within, and if so, its length, height, and class. The metric indicating how confident the model is about the presence of an object in a given grid is the confidence score. As seen in Equation 1, the confidence score is calculated by multiplying the probability of  $P(\text{object})$  (the likelihood of the drawn box containing an object) with the Intersection over Union (IoU) ratio of the ground truth and predicted boxes.

$$\text{Confidence Score} = P(\text{object}) * \text{IoU} \quad (1)$$

YOLO employs the anchor box method when there are multiple objects in a grid on the input image. Anchor boxes, determined in advance by a specified number of anchors, facilitate the prediction of the box around the object using predefined patterns. Following the success of the original YOLO algorithm, numerous new versions of YOLO have been

developed. The YOLOv5 version [15], preferred in this study, builds upon the success of previous versions by introducing new features and improvements. YOLOv5 utilizes a more complex architecture called EfficientDet, based on the EfficientNet network architecture. This architecture allows for a broader range of object categories, achieving higher accuracy and better generalization. YOLOv5 has been trained with a dataset named D5, encompassing a total of 600 object categories.

### 2.3. Transfer Learning

One of the most significant factors influencing the success of a deep learning model is the dataset. When creating a dataset is challenging, the Transfer Learning method is frequently employed. This method is based on training the final layers of a model, previously trained and balanced with a large dataset, with a new and smaller dataset. Transfer learning involves transferring knowledge learned in one task to be utilized in another task. Creating and training a model from scratch can be costly. Transfer learning shortens the model training time and reduces the need for powerful hardware, providing advantages in terms of time and resources. Models pre-trained on large datasets are referred to as SOTA (State of the Art) models. In this study, the YOLOv5 SOTA model was preferred, and its final layers were retrained and integrated into the workflow.

### 2.4. Evaluation Metrics

In evaluating the performance of the trained model in this study, metrics such as TP (True Positive), FP (False Positive), TN (True Negative), FN (False Negative), and evaluation metrics including P (Precision), R (Recall), and mAP (Mean Average Precision) calculated from the confusion matrix were utilized. P represents the ratio of instances that the model correctly predicted as positive to the positive instances, while R is the ratio of instances predicted as positive by the model to the total positive instances. Equations 2 and 3 provide the formulas for P and R, respectively.

$$P = TP / (TP + FP) \quad (2)$$

$$R = TP / (TP + FN) \quad (3)$$

Intersection Over Union (IOU) is the ratio of the intersection of the area predicted by the model to the union of the actual area. The mAP at IOU = 0.5 is calculated as the area under the Precision-Recall curve. This area under the curve, calculated in steps of 0.05 up to 0.95, yields the mAP values. The formula is shown in Equation 4.

$$mAP = 1/N \sum_{i=1}^N AP_i \quad (4)$$

## 2.5. Training Analysis

The parameters of YOLOv5 were optimized to detect three separate classes during the model training phase. The YOLOv5m variant, with 20,879,400 learnable parameters and 369 layers, was chosen for this purpose. The created dataset was then retrained on this model, and its weights were updated. The model was trained for 12 epochs, and all experimental procedures were conducted on the Google COLAB platform with GPU activation using a Tesla T4 graphics card. Table 1 provides a summary of the model training outputs. The Precision, Recall, and mAP values in Table 1 represent the evaluation metrics for the model's predictions on the test set. Based on the results in Table 1, it can be concluded that testing the model on images it has not seen before is appropriate.

Table 1: YOLOv5m Performance Results

Class	Precision	Recall	mAP@.5	mAP@.5:.95
All	0.988	0.995	0.994	0.819
Class 0	0.98	0.987	0.992	0.828
Class 1	0.987	0.997	0.995	0.771
Class 2	0.998	1	0.995	0.858

## 3. Results

This study encompasses image-based detection efforts using deep learning for the task of autonomous docking of an autonomous mobile robot to a Euro pallet. The dataset, originally created and subjected to data augmentation processes, was retrained using the YOLOv5 SOTA model with transfer learning. To visualize the performance results of this model, Figure 5 presents a graph illustrating the relationship between the F1 score and confidence level. The F1 score is the harmonic mean of precision and recall values for a class. These metric balances both false positive (FP) and false negative (FN) predictions. In Figure 5, the x-axis represents the confidence level of the model's predictions, while the y-axis represents the F1 score. Ideally, high F1 scores are desired at both high and low confidence levels. As observed in Figure 5, the model is suitable for real-world tests.

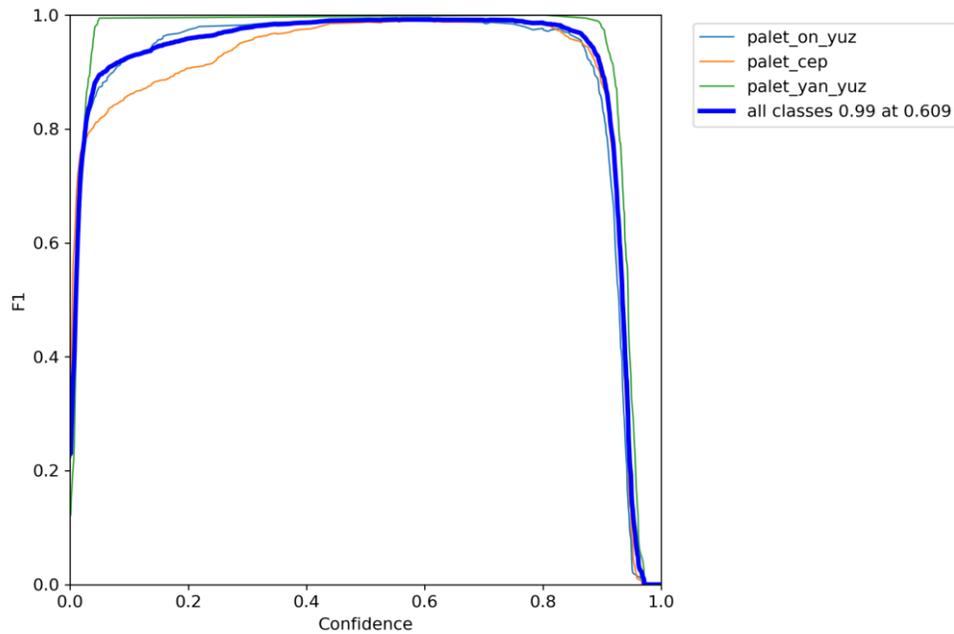


Figure 5: F1 Curve Over Confidence

Figure 6 displays the detection results in images captured by the RGB-D camera on the mobile robot. The study's usability has been tested under various lighting conditions and different pallet positions, and its effectiveness has been observed.



Figure 6: Pallet Detection Results

The main focus of this study involves detection efforts for enabling autonomous mobile robots to dock the widely used Euro pallet object with precision in industrial settings. The process of accurately recognizing pallets under various lighting conditions and changing circumstances has been successfully completed by the autonomous mobile robot. Future work will include calculating the detected pallet's position, orientation, and transformations on a 3D virtual map.

#### 4. Acknowledge

The authors cordially thank the Kar Metal and Sakarya University of Applied Sciences Robot Technologies and Intelligent Systems Application Research Center (ROTASAM), Turkey for supporting this study.

#### References

- [1] Czimmermann, T., Ciuti, G., Milazzo, M., Chiurazzi, M., Roccella, S., Oddo, C. M., & Dario, P. (2020, 3 6). Visual-Based Defect Detection and Classification Approaches for Industrial Applications—A SURVEY. *Sensors*.
- [2] Deng, H., Cheng, Y., Feng, Y., & Xiang, J. (2021, 09 18). Industrial Laser Welding Defect Detection and Image Defect Recognition Based on Deep Learning Model Developed. *Symmetry*.
- [3] Shahmoradi, J., Talebi, E., Roghanchi, P., & Hassanalian, M. (2020, 07 15). A Comprehensive Review of Applications of Drone Technology in the Mining Industry. *Drones*.
- [4] Bach, S.-H., Khoi, P.-B., & Yi, S.-Y. (2023). Application of QR Code for Localization and Navigation of Indoor Mobile Robot. *IEEE*, (s. 28354-28390).
- [5] Guo, S., Fang, T.-T., Song, T., Xi, F.-F., & Wei, B.-G. (2018, 02 28). Tracking and localization for omni-directional mobile industrial robot using reflectors. *Adv. Manuf.*
- [6] Bosnak, M., & Skrijanc, I. (2021). Obstacle Avoidance for Line-Following AGV with Local Maps. *International Symposium on Applied Computational Intelligence and Informatic SACI 2021*, (s. 193-198). Timișoara.
- [7] Jia, F., Tao, Z., & Wang, F. (2021). Pallet Detection Based on Halcon and Alex:Net Network for Autonomous Forklifts. *International Conference on Internet, Education and Information Technology*, (s. 86-89).
- [8] Kurnianto, H., & Rusmin, P. H. (2022). Task Allocation and Path Planning Method For Multi-Autonomous Forklift Navigation. *International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, (s. 631-636).
- [9] Manurung, P., Rusmin, P. H., & Yusuf, R. (2022). Custom Pallet Detection Using YOLOv5 Deep Learning Architecture. *International Symposium on Electronics and Smart Devices (ISESD)*.
- [10] Rahman, M., Kanchi, M., Pasha, M., & Yadla, B. (2021). Design and Development of Autonomous Warehouse Management Robot with Intelligent Software Framework. *International Conference on Mobile Networks and Wireless Communications*.
- [11] Syu, J.-L., Li, H.-T., Chiang, J.-S., Hsia, C.-H., Wu, P.-H., Hsieh, C.-F., & Li, S.-A. (2016, 11 2). A computer vision assisted system for autonomous forklift vehicles in real factory environment. *Multimedated Tools Appl.*

- [12] Li, T., Huang, B., & Huang, M. (2018). Application of convolution neural network object detection algorithm in logistics warehouse. *International Symposium on Test Automation and Instrumentation*. The Journal of Engineering.
- [13] Shao, Y., Fan, Z., Zhu, B., Zhou, M., Chen, Z., & Lu, J. (2022, 11 9). A Novel Pallet Detection Method for Automated Guided Vehicles Based on Point Cloud Data. *sensors*.
- [14] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA.
- [15] Ultralytics. (2021). *YOLOv5: A state-of-the-art real-time object detection system*. <https://docs.ultralytics.com>.