*Proceedings of the 10th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'24)*
*Barcelona, Spain - August 19-21, 2024*
*Paper No. MVML 110 (The number assigned by the OpenConf System)*
*DOI: 10.11159/mvml24.110*

# Pedestrian Equipment Anomaly Detection with Computer Vision in Warehouses

**Tuğçe Elçi[1,2], Mehmet Z. Ünlü[1], Deniz Kantar[2], Ahmet Yesevi Türker[2], Hasan Güney[2], Ahmet Ustaoğlu[2]**
[1]Izmir Institute of Technology
Electrical and Electronics Engineering Department, Izmir, Türkiye
tugceelci@iyte.edu.tr; zubeyirunlu@iyte.edu.tr;
[2]Borusan Logistics
Üsküdar, Istanbul, Türkiye
deniz.kantar@borusan.com; ahmetyesevi.turker@borusan.com; hasan.guney@borusan.com; ahmet.ustaoglu@borusan.com

***Abstract* -** The rapid growth of the logistics sector in recent years caused the expansion of warehouse areas and the increase in the number of equipment used. With the increase in these activities, the possibility of work accidents in warehouses also increases. In defiance of this situation, it has been determined that a real-time prediction system of pedestrian and equipment interaction is needed to ensure in-warehouse reliability. This system should address the urgent need to reduce the risk of work accidents and focus on the overall goal of reducing the possibility of work accidents in warehouse environments. To overcome this challenge, we propose a comprehensive Warehouse Anomaly Detection and Control System consisting of object detection, object tracking, action detection, and alarm classification components which will play an important role in increasing work safety in warehouse environments. YOLOv7 (You Only Look Once version 7) is a deep learning model that detects objects quickly and accurately in a single network pass. The deep learning-based Deep SORT algorithm used for object tracking provides a dynamic understanding of the warehouse environment by continuously storing these identified problems in real-time. The action detection part of this system is designed to identify and analyze actions and movements, recognizing anomalies and potential risks. In this part, the speed of pedestrians and equipment are detected utilization of 3D bounding boxes of objects and perspective transformation. The possible accident risks are measured using the intersection percentage of these areas, the magnitude of speed, the direction of the motion vector of pedestrian and equipment, and the distances between objects. Alert levels can be considered as encounter, near-miss, and emergency. Using this system in warehouses will reduce the risk of possible work accidents that may even result in death.

***Keywords*:** warehouse; deep learning, object tracking, object detection; action detection; computer vision

## 1. Introduction

With the rapid development of the logistics sector, digitalization processes have accelerated. With this acceleration, it is aimed to completely digitalize many processes such as collecting products in warehouses, stacking them, preparing them for distribution, customer-oriented pricing, and distribution of products. According to the studies of Kayıkçı, it is possible to see the effects of Industry 4.0, which is considered as the industrial revolution, behind these targets. The terms "Smart Factory" and "Factory of the Future", which entered our lives with the impact of this revolutionary step, are considered to be the reflection of digitalization processes on logistics processes [1]. The logistics sector has annual growth rates of 7-10% in Europe, 15% in North America, and 20% in Asia and Turkey [2]. With the digitalization process in the logistics sector and the increase in customer demands in the supply chain, changes are observed in storage areas and product stacking methods. With the increasing business volume, there is also an increase in the number of people working in warehouses and, accordingly, the number of equipment used.

There are different approaches to similar problems. Studies using convolutional neural networks (CNN) are generally observed. Studies similar to the problem were examined. In the study by Yadav, Renu, Ankita and Anjum, which provides information about the serious effects and causes of road traffic accidents on human life, it is stated that approximately 1.35 million people are affected by accidents every year and the injuries caused by these accidents affect 20 to 50 million people. One of the main reasons for such accidents is incompatibility between organizations and inadequate enforcement of traffic rules, which increases the number of accidents. Risk factors such as speed, drunk driving, distraction, poor infrastructure, unsuitable vehicles, and violation of traffic rules also play a role in accidents. In order to respond quickly to such accidents, a detection system using different technologies is required. Various technologies such as Global Positioning System (GPS), Global System for Mobile Communications (GSM), and mobile applications form the basis of such systems. Additionally, the use of convolutional neural networks (CNN) used in object identification and tracking technology together with long short-term memory (LSTM) is also mentioned. This study provides an overview of technologies related to automatic road [traffic] accident detection systems and explores solutions to reduce the number of accidents [3].

The decrease in accidents in the warehouse depends on the decrease in interactions between pedestrians and equipment. In the warehouse, a video surveillance system with a deep algorithm structure, complex object detection, object tracking, and action detection systems is installed to detect the interaction between pedestrians and equipment. The YOLOv7 model, which is very popular today, is used in the field of object recognition. Following this system, the Deep SORT algorithm provides object tracking in the real-time system. By using the information from these areas, the interactions between the pedestrian and the equipment are classified with the event identification part of the system, using outputs such as the object's speed, motion vector, and distance between objects. Different warning systems can be arranged within the warehouse depending on the classified alarm level.

## 2. Methodology

The system designed to minimize pedestrian and equipment interaction in warehouses in the logistics sector, which continues its continuous development, consists of four main parts. These parts are object detection, object tracking, action detection, and alarm classification. The general flow of the system is given in Figure 1. To analyze the general flow of the system, in the first stage, an object detection system was designed to distinguish pedestrians and equipment from real-time videos. At this stage, it is determined whether the pedestrian and equipment are included in the video. After the detected object is detected in the specified frame rate according to the fps information received from the video, the object detection process is stopped. In the second stage, the detected pedestrians or equipment are tracked. At this stage, information about the object is extracted. After this stage, the speed information of pedestrians and equipment is determined. After the object detection and tracking phase, information about pedestrians and equipment needs to be extracted. These features are extracted in the feature extraction part. The extracted information is given to the trained model. At this stage, anomaly detection is performed to display the interaction situation between the pedestrian and the equipment. In the last stage, the alarm is classified according to the anomaly detection.

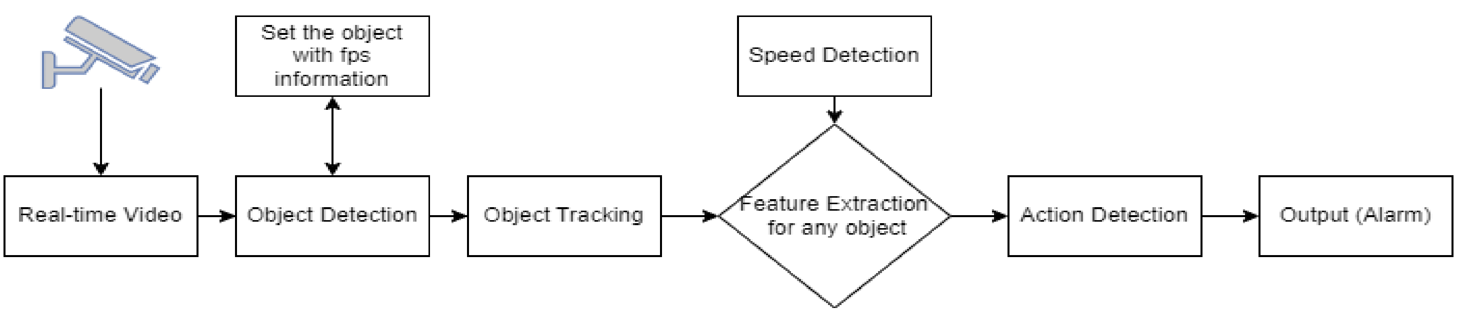


Fig 1: General Flow Diagram of the System.

### 2.1. Object Detection

The first stage of the system to be created in order to understand the detection of pedestrian equipment interactions is the detection of objects. It is extremely important to distinguish between pedestrians and equipment on images taken from real-time videos during the object detection phase. During the object detection phase, a literature study was conducted to select the method to be used. As a result of the studies, the YOLOv7 model was chosen. The

YOLOv7 model is one of the latest versions of the YOLO series. YOLOv7, which has more accurate and faster performance than other versions, is extremely suitable for real-time transactions [4,5]. The parameters used during training (batch size, number of epochs, image size, learning rate, etc.) were selected and optimized. The parameters that are critical for the success of the model increased the success of the model and prevented overfitting.

### 2.2. Object Tracking

As seen in the methodology of the study, one of the most important parts is object tracking. Based on the information obtained during object detection, detailed information about the movement of the object is obtained. The information obtained is arranged as usable in the action detection field. Here the detection of the motion vector takes place. As a method, the Deep-SORT algorithm was used in the object detection section. The Deep SORT algorithm is an improved version of the SORT (Simple Online Real-time Tracking) algorithm. The algorithm includes a simple Kalman filter. The calculation of data correlation in each frame is calculated with the Hungarian algorithm. The identity problem in the SORT algorithm has been solved with the Deep SORT algorithm. Additionally, the Deep SORT algorithm is based on the deep learning method [6].

### 2.3. Action Detection

In this study, features determined by the detection and tracking of pedestrians and equipment were extracted in the action detection section. During the feature extraction process, the speed information detected for each object from the speed detection field is also included in the features. Speed detection was extracted from real-time video by the 3D transformation method. Camera calibration was performed before speed detection. Information about the extracted object type, the object's speed, center coordinates, object's motion vector, etc. Trained using vanilla Autoencoder features. Anomaly detection is detected if the specified reconstruction error value is above the specified threshold. Anything detected is instantly converted into an alarm. Since the system will be used in the warehouse, alarm generation is of great importance.

### 2.4. Alarm Classification

The alarm classification system is realized by evaluating the detected anomalies. The detected reconstruction error value is checked for outliers. The thresholds determined for these values are classified as encounter, near-miss, and emergency. It is designed that the result of this classification will be sent to the systems within the warehouse as a message. According to the alarm information generated, the necessary actions are expected to be taken within the warehouse.

## 3. Experimental Results and Future Works

In the system we created, it is possible to explain the successes of the YOLOv7 model, in which we provide information about whether the object is a pedestrian or equipment during the object detection phase, with the table below.

Table 1: YOLOv7 results for each object detected.

| Object Type | Accuracy Score |
|---|---|
| Person | 86% |
| Equipment | 88% |

Classification was made based on the anomalies determined in line with the results obtained. 319 anomaly values were detected in 753 sample accident videos. As seen in the graph, the indexes of possible interaction combinations are called sample indexes. The blue area with the reconstruction error value lower than the threshold value indicates the success of the model. These situations are classified as normal. The green area referred to as Encounter indicates situations with a low probability of encounter. In the example where the collision actually occurred, the collision was predicted to occur with the maximum error rate. The value at which the reconstruction error was maximum was determined in the emergency class. It has been shown that the accident can be prevented by the alarm given during the encounter and near miss stages before reaching this value and by taking precautions without entering the emergency classification.

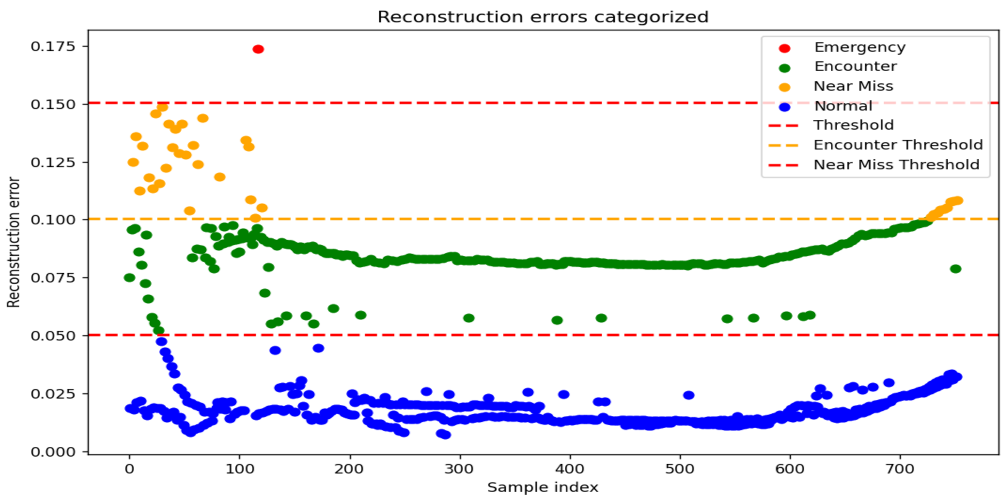


Fig 2: Results of an example for action detection and alarm classification.

## 4. Conclusion

In conclusion, in the study, the system was made suitable for real-time operation. The system has a complex structure due to the models used. Once the necessary hardware for the system to work is provided, the methods used are suitable for taking action quickly. In future studies, although models can be shaped according to the analysis of the problem, it is aimed to reduce the number of accidents that are likely to occur between pedestrians and equipment in the warehouse to zero by adding them to the system in different analyses.

## Acknowledgements

In this study, Izmir Institute of Technology and Borusan Logistics worked together, and a master's thesis was prepared within the scope of this project. We would like to thank both institutions for their valuable contributions and support.

## References

[1] Y. Kayikci, “Sustainability impact of digitization in logistics,” Procedia Manufacturing, vol. 21, no. 1, pp. 782–789, 2018, doi: https://doi.org/10.1016/j.promfg.2018.02.184.

[2] A. Öztürk, “Efficient warehouse management and a study on efficient warehouse strategies of logistic warehouses,” Doctoral Thesis, Istanbul Ticaret University, 2011.

[3] D. K. Yadav, Renu, Ankita, and I. Anjum, “Accident Detection Using Deep Learning,” IEEE Xplore, Dec. 01, 2020. https://ieeexplore.ieee.org/document/9362808

[4] F. Yang, X. Zhang, and B. Liu, “Video object tracking based on YOLOv7 and DeepSORT,” arXiv.org, Jul. 25, 2022. http://arxiv.org/abs/2207.12202 (accessed Sep. 15, 2023).

[5] G. Ye, J. Qu, J. Tao, W. Dai, Y. Mao, and Q. Jin, “Autonomous surface crack identification of concrete structures based on the YOLOv7 algorithm,” Journal of Building Engineering, vol. 73, no. 106688, Aug. 2023, doi: https://doi.org/10.1016/j.jobe.2023.106688.

[6] V. K. Yadav, Dr. P. Yadav, and Dr. S. Sharma, “An Efficient Yolov7 and Deep Sort are Used in a Deep Learning Model for Tracking Vehicle and Detection,” Journal of Xi’an Shiyou University, Natural Science Edition, vol. 18, no. 11, pp. 759–763, Nov. 2022.