©2025 Published in 11th International Symposium on Intelligent Technologies in Engineering and Science 16-18 May 2025 (ISITES2025 Diyarbakır - Türkiye)

Calculate the Volume of Packages on a Transport Line with Yolo

[1]Onurcan Yozcu , [2]Beyzanur BOSTANCI and [3]Gökhan Topdanış
[1,2,3]Optimum Süreç Tasarımı ve Uygulamaları A.Ş.,R&D Center, Sakarya, Türkiye

Abstract

In this study, real-time detection and volume calculation of cargo packages on a conveyor belt were performed. The study was conducted using YOLOv8 and YOLOv11 models. The dataset, consisting of two classes, was labeled using the Labelme tool, and segmentation and region detection were performed. Packages were detected on the conveyor belt using the Intel RealSense D435i depth camera, and their volumes were accurately calculated. As a result of the training process, YOLOv11 was identified as the most successful model. YOLOv8 achieved mAP50: 0.993, mAP50-95: 0.956, precision: 0.978, and recall: 0.989, while YOLOv11 achieved mAP50: 0.995, mAP50-95: 0.976, precision: 1, and recall: 1. The results highlight the effectiveness of deep learning models, particularly the YOLO family, in object detection and volume measurement tasks in industrial environments.

Key words: YOLO, object Detection, segmentation, intel realSense.

1. Introduction

Industrial automation systems play a critical role in terms of productivity and cost management. Especially in logistics and warehousing, these systems help to achieve labor and time savings by providing significant productivity gains. Conveyor belt systems are among the widely used tools for transporting products from one point to another. In order for these systems to work correctly and effectively, accurate detection and segmentation of packages is of great importance. Traditional methods for package detection and volume measurement can be time-consuming and error-prone. However, deep learning-based methods make it possible to perform these processes faster and more accurately. Deep learning algorithms provide significant benefits in industrial applications by providing high accuracy rates in tasks such as object detection and segmentation [1].
In recent years, advances in deep learning have led to significant achievements, especially in object detection and segmentation tasks. YOLO (You Only Look Once), which is the pioneer of these developments, draws attention with its high accuracy rates.
YOLO not only provides fast and accurate detection of objects in an image, but can also be used effectively for more complex tasks such as detection and segmentation. New generation models such as YOLOv8 and YOLO11 have more advanced architectures compared to previous versions and offer much higher accuracy rates in industrial applications [2]. The real-time processing capabilities of these models provide significant advantages by enabling fast and accurate detection processes in conveyor belt systems.
The sampling algorithm proposed by the Berkeley group has been developed. In addition, the YOLO deep learning network is integrated for package recognition. A multimodal robot arm is

[1] Corresponding author: Address: Optimum Süreç Tasarımı ve Uygulamaları A.Ş., R&D Center, 54400, Sakarya TÜRKİYE. E-mail address: onur.yozcu@tr.optimakstu.com, Phone: +90 264 353 52 53

designed to cope with different types of packages, including a two-finger gripper and a vacuum suction apparatus. The sampling algorithm proposed by the Berkeley group is developed. Furthermore, the YOLO deep learning network is integrated for package recognition. To deal with different types of packages, a multimodal robot arm was designed, including a two-finger gripper and a vacuum suction apparatus. For flat and thin packages such as envelopes, the success rate was almost 100% when vacuum suction was used, and for deformable packages such as bags, about 90% success was achieved with the two-finger gripper [3]. Liu, Y. et al. use deep learning-based object detection and segmentation techniques to accurately detect packages, calculate their size and volume, and quickly and efficiently route them on the conveyor belt [4]. Ultralytics YOLO11 model offers advanced results in packet detection and segmentation. This model has been widely used in industrial applications with its high accuracy rates and efficient processing capacity. The model can perform fast and accurate detections using deep learning algorithms, thus optimizing the automatic processing and dispatching of packages [5]. They also applied principal component analysis on the model that gave the best result in the study [9]. Wang, Z. et al. proposed an approach based on fruit stem and position to determine the position of fruits during the packing process. They used the Yolo-v5 algorithm to determine the stem of an apple in real time. In the study, an accuracy of 93.89% was obtained [10]. In another study on the use of networks for feature extraction, Varshni, D et al. demonstrated the success of transfer learning approaches in extracting meaningful features. In their study based on DenseNet and ResNet architectures, success outputs were obtained as a result of sequential use of extracted features with decision support machines. Accuracy values of 77.49% (ResNet50) and 80.02% (DenseNet201) were reported as the results of the study [11].

In the literature, it is observed that YOLO models provide high speed and success especially in real-time object detection and classification problems. In addition, the literature survey shows that YOLO models have been used effectively in many different object detection and classification problems and have yielded successful results especially in industrial automation and security fields. In this context, YOLOv8 and YOLO11 models are preferred in this study to provide high speed and accuracy in tasks such as object detection and volume computation. In the study, object detection and segmentation operations were performed using YOLO models.

2. Materials and Method

The aim of this study is to investigate the detection and segmentation of cargo packages on a conveyor belt by comparing the performance of YOLOv8 and YOLO11 models. Furthermore, in this study, the Intel RealSense D435i depth camera improves the accuracy of volume calculations by detecting three-dimensional features of objects [100]. The width, height, area and volume of the packages were measured, and these operations were controlled and reported to the user by designing the interface through PyQt. The data obtained is recorded through the SQLite database, aiming to increase efficiency and minimize errors in industrial areas.

[1] Corresponding author: Address: Optimum Süreç Tasarımı ve Uygulamaları A.Ş., R&D Center, 54400, Sakarya TÜRKİYE. E-mail address: onur.yozcu@tr.optimakstu.com, Phone: +90 264 353 52 53

2.1. Data Set Information

The images were converted into a dataset by separating them into consecutive frames in a certain order and carefully selected to maintain data balance. In total, labeling was performed on 4000 images. Each image has a height of 2160 pixels and a width of 3840 pixels and is represented in RGB color channels. The labeling process was performed using Labelme software, which provides the labeling tool, and all images were classified into two classes: line and package. Examples of the images are shown in Figure 1.

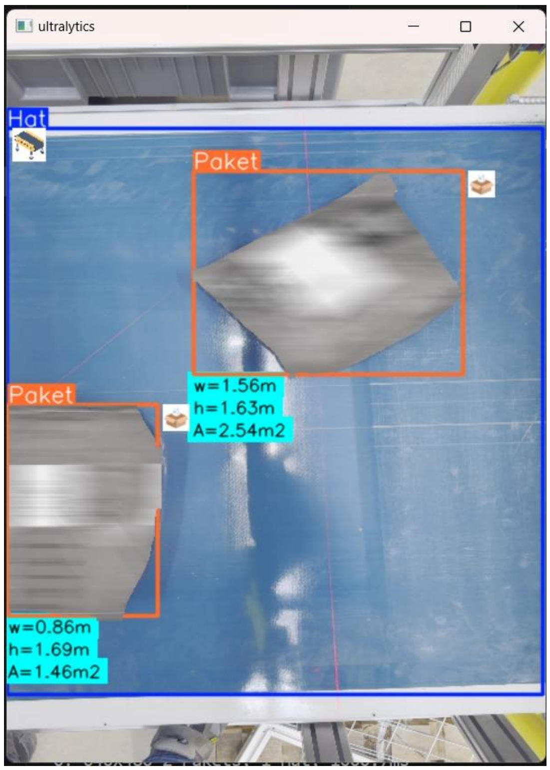

Figure 1. Example labeled data image

2.2. Proposed Approach

In the present study, the success of deep learning methods used for object detection and segmentation largely depends on speed and accuracy. YOLO stands out with its speed and accuracy advantages in the field of object detection. Especially new generation models such as YOLOv8 and YOLO11 provide faster and higher accuracy rates, which enables more efficient detection and tracking processes in conveyor belt systems. The biggest advantage of YOLO compared to other object detection models such as Faster R-CNN and SSD is its speed. Models such as Faster R-CNN have slower processing times and generally require more computational power, while SSD is a fast model, but has lower accuracy rates compared to the speed and real-time processing capacity

[1] Corresponding author: Address: Optimum Süreç Tasarımı ve Uygulamaları A.Ş., R&D Center, 54400, Sakarya TÜRKİYE. E-mail address: onur.yozcu@tr.optimakstu.com, Phone: +90 264 353 52 53

offered by YOLO. YOLOv8 and YOLO11 models provide faster and more accurate results without lagging behind these models. This is a significant advantage in environments that require real-time processing, such as the conveyor belt.
In order to accurately calculate the volumes of the packages on the conveyor belt, an Intel RealSense D435i depth camera was used. This camera increases the accuracy of volume calculations by detecting the three-dimensional features of the packages. The use of depth cameras minimizes the error rates in volume measurements by providing more accurate results than traditional image-based methods.

The PyQt-based user interface allows users to easily analyze these data by visualizing the detection, segmentation, width, height, depth, area and volume measurements of the packets in the conveyor belt. Through the interface, users can instantly monitor packet detection and volume calculations, examine the data obtained in detail and generate reports. These features enable users to perform data analysis processes in a more efficient and understandable way. In Figure 2 and Figure 3, interface visuals and examples of detected packets visualized in the interface are presented.

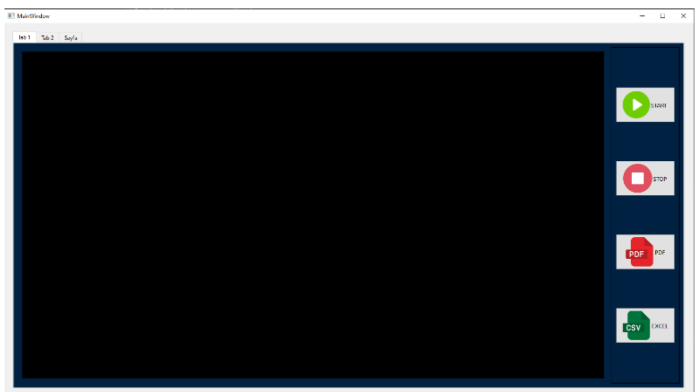

Figure 2. Interface Design

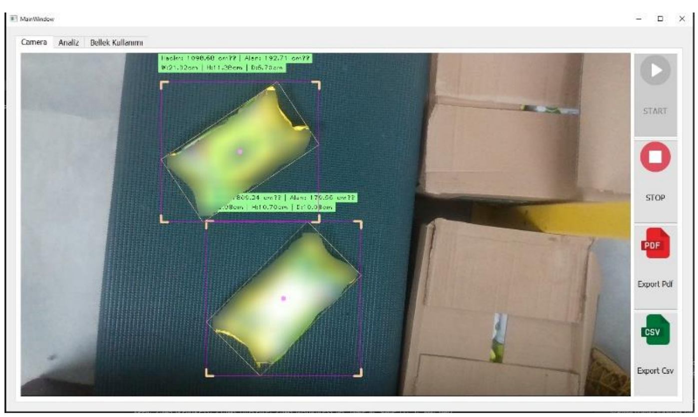

Figure 3. Working of the Interface

3. Results

[1] Corresponding author: Address: Optimum Süreç Tasarımı ve Uygulamaları A.Ş., R&D Center, 54400, Sakarya TÜRKİYE. E-mail address: onur.yozcu@tr.optimakstu.com, Phone: +90 264 353 52 53

In the proposed study, real-time detection and volume calculation of cargo packages on the conveyor belt is performed. Within the scope of the study, YOLOv8 and YOLO11 models are trained, tested and compared. During the training process, the dataset consists of 4000 images and the labeling process was performed with the Labelme tool. To evaluate the model performances, mAP50, mAP50-95, precision and recall metrics were used. The training results show that the YOLO11 model is the most successful model in terms of object detection and segmentation. These results show that the YOLO11 model has higher accuracy, precision and recall rates and superior segmentation performance. In particular, the precision and recall values of 1.00 indicate that YOLOv11 accurately detects all objects without making any errors.

Table 1. Classification accuracy values

Criteria for Success\Model	YOLOv8	YOLO11
mAP50	0.993	0.995
mAP50-95	0.956	0.976
Precision	0.978	1
Recall	0.989	1

Using the Intel RealSense D435i depth camera, the volumes of cargo packages on the conveyor belt were measured and calculated using depth data and segmentation masks. For each package, the segmentation mask was combined with the depth data to calculate the actual area, and then the volume was calculated using the difference between this area and depth.

The calculation was based on the width, height and depth measurements of the packages. The depth camera provides three-dimensional data, which allowed for more accurate results compared to traditional two-dimensional methods.
The NVIDIA RTX 4060 graphics card was used in the system and processing speeds were increased with CUDA technology. Running the YOLOv11 segmentation model on the GPU significantly reduced processing times and maximized the use of processing power. By measuring FPS (frames per second) every second, the processing power and efficiency of the system were monitored in real time. FPS was calculated every second, providing continuous feedback on the processor performance and the efficiency of the model.

All processing data was saved in a SQLite database for users to access and evaluate performance based on historical data. Through the PyQt-based interface, they can visually monitor the detection, segmentation, volume calculations and other data of packets in the conveyor belt. The interface is designed to enable users to perform instant analysis, generate reports and analyze data more efficiently.

4. Discussion

[1] Corresponding author: Address: Optimum Süreç Tasarımı ve Uygulamaları A.Ş., R&D Center, 54400, Sakarya TÜRKİYE. E-mail address: onur.yozcu@tr.optimakstu.com, Phone: +90 264 353 52 53

In this study, YOLOv8 and YOLOv11 deep learning models are used for the detection and volume calculation of cargo packages on the conveyor belt and their real-time object detection and segmentation achievements are compared. The results show that both models can be used effectively in industrial automation systems. In particular, YOLOv11 was found to have higher accuracy, precision and recall values and superior segmentation performance. These findings reveal that the next generation models of the YOLO family can provide more efficient and accurate results in industrial applications. Another important finding of the study is that the use of the Intel RealSense D435i depth camera improves the accuracy of volume calculations. Compared to traditional 2D imaging methods, the depth camera provided more precise and accurate results using three-dimensional data. In addition, the PyQt-based user interface enabled users to instantly monitor the packages on the conveyor belt and visually track the volume calculations.

Conclusions

In conclusion, this study successfully demonstrates the use of deep learning-based object detection and volume computation systems in industrial automation processes. In future studies, the generalization capability of the model can be further improved by examining larger data sets and different scenarios. In addition, hardware optimizations and software improvements can be made to further reduce processing times, thus achieving more efficient results in environments that require real-time processing such as conveyor belts. This study has made a significant contribution to the field of industrial automation and sheds light on future research and applications.

Acknowledgements

We would like to thank Gökhan Topdanış, who completed his internship at Optimum Process Design and Applications Inc. for his contributions to this study.

References

[1] Zhou, Y., Liu, Z., & Xie, L. (2020). Deep learning applications in industrial automation: Object detection and segmentation. Journal of Industrial Automation, 34(3), 245-259.

[2] Redmon, J., & Farhadi, A. (2018). YOLOv4: Optimal speed and accuracy for real-time object detection. IEEE Conference on Computer Vision and Pattern Recognition, 1-9..

[3] Wang, S., Jiang, X., Zhao, J., Wang, X., Zhou, W., & Liu, Y. (2019, August). Vision based picking system for automatic express package dispatching. In 2019 IEEE international conference on real-time computing and robotics (RCAR) (pp. 797-802). IEEE.

[4] Wang, S., Jiang, X., Zhao, J., Wang, X., Zhou, W., & Liu, Y. (2019, August). Vision based

[1] Corresponding author: Address: Optimum Süreç Tasarımı ve Uygulamaları A.Ş., R&D Center, 54400, Sakarya TÜRKİYE. E-mail address: onur.yozcu@tr.optimakstu.com, Phone: +90 264 353 52 53

picking system for automatic express package dispatching. In 2019 IEEE international conference on real-time computing and robotics (RCAR) (pp. 797-802). IEEE.

[5] Vina, A. (2025, Şubat 7). Package identification and segmentation with Ultralytics YOLO11. Ultralytics.Jan, R., Kour, H., Manhas, J., & Sharma, V. Recognition of Dry Fruits using Deep Convolutional Neural Network.

[6] Mohedano, M., Sánchez, D., & García, C. (2019). Real-time object detection and volume measurement with depth cameras in industrial automation. Journal of Industrial Robotics, 45(2), 123-135.

[1] Corresponding author: Address: Optimum Süreç Tasarımı ve Uygulamaları A.Ş., R&D Center, 54400, Sakarya TÜRKİYE. E-mail address: onur.yozcu@tr.optimakstu.com, Phone: +90 264 353 52 53