Express Sorting System Based on Improved YOLO Algorithm

Rui Zhao [1, #], Jiayang Zhang [2, #], Yinuo Shen [3, #, *]

[1] College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan, China, 030024

[2] School of Information Science and Technology, Xiamen University Tan Kah Kee College, Zhangzhou, China, 363105

[3] School of physics, East China University of Science and Technology, Shanghai, China, 200237

* Corresponding Author Email: bquak801378@163.com

[#]These authors are contributed equally.

Abstract. With the rapid development of e-commerce, the volume of parcel handling in the logistics industry has increased dramatically, and traditional manual sorting faces the problems of low efficiency and high cost. Based on the above problems, this paper develops an express sorting robot system based on computer vision. The system selects Raspberry Pi as the main control module and utilizes a variety of algorithms, among which the target detection adopts the YOLO series of algorithms. Meanwhile, this paper proposes a new courier sorting dataset, Parcel360° Dataset, which is utilized to train and evaluate YOLOv5, YOLOv8 and YOLO11. Experiments show that YOLOv8 achieves the best balance between accuracy and efficiency, with a mean accuracy (mAP) of 62.8% and a frame rate (fps) of 27, which can meet the high accuracy and real-time requirements of logistics sorting. The system helps to promote the intelligent transformation of the logistics industry and enhance the overall competitiveness of the industry.

Keywords: Express Sorting, Computer Vision, YOLO Algorithm, Edge Computing.

1. Introduction

In the booming development of e-commerce, the logistics industry's parcel processing volume is explosive growth, the traditional manual sorting in the efficiency and labor costs under the double pressure of struggling. In the peak period of logistics, a large number of parcels piled up, the speed of manual sorting is far from being able to meet the demand, and prolonged high-intensity work can easily lead to fatigue and errors, which seriously affects the timeliness and accuracy of logistics and distribution, and has become a key factor restricting the development of the industry.

In order to break through the bottleneck in the field of logistics sorting, all walks of life have actively invested in the research and development of intelligent logistics solutions, and robot sorting technology has become the focus. However, the existing research results still have significant flaws. For example, the robot designed by Liu Jiqiu and Hu Lifu [1] adopts the traditional vision processing method, which is difficult to cope with the problems of irregular parcel placement, lighting changes and diversity of packaging materials in complex logistics scenarios, resulting in insufficient recognition accuracy and frequent obstruction of the sorting process. Although the scheme proposed by Han Xing [2] improves the recognition technology, when running on edge devices, the inference speed is slow due to the complexity of the algorithm and the high demand of hardware resources, which cannot meet the real-time requirements. The STM32-based logistics sorting robot [3] designed by Wu Xuefeng et al. has limited visual perception and relies on yellow and black line maps and screening sensors, which makes it difficult to recognize packages and obstacles in complex scenes. The Arduino-based courier sorting robot [4] proposed by Jiahui Huang, Zisen Zhou, et al. has limitations in cargo identification and processing, the vacuum suction cups are not effective in adsorbing shaped and uneven surface cargoes, and the two-dimensional code recognition is not sufficiently stable under complex lighting conditions.

In view of the above dilemma, this paper is dedicated to the development of an innovative computer vision-based courier sorting robot system. In terms of hardware, Raspberry Pi is chosen as the main control module to give full play to its advantages of compact size, low energy consumption, and high scalability, to ensure the stable and efficient operation of the system in complex environments, and at the same time to reduce the cost of hardware and the difficulty of deployment. In terms of algorithm application, this paper focuses on the in-depth research and optimization of YOLOv5 [5], YOLOv8 [6] and YOLO11 [7], aiming to improve the recognition accuracy and reasoning speed, in order to satisfy the stringent requirements of the logistics industry for efficient sorting.

In this paper, a large number of logistics parcel images covering different shapes, sizes, packaging materials and placement angles are collected, and in this way, a rich and diverse training dataset, Parcel360° Dataset, is constructed and used to train the three algorithms. For YOLOv5, it is able to quickly locate possible parcel regions and further improve its recognition and generalization ability for different types of parcels by continuously optimizing the parameters of its feature extraction and detection layers, while YOLOv8 provides more in-depth refinement and accurate recognition based on the initial recognition of YOLOv5. And YOLO11 is mainly targeted for training on some parcels with special size, shape or unique appearance characteristics.

In a simulated logistics scenario, this paper comprehensively evaluates the performance of three algorithms, YOLOv8, YOLOv5, and YOLO11, based on key metrics such as recognition accuracy, recall, F1 value, and inference speed, in order to identify the algorithmic version that is most suitable for logistics sorting tasks. The experimental results show that YOLOv5 has the lowest mean accuracy value (mAP) and YOLO11 has the highest precision value (precision), while YOLOv5 has the best performance in terms of mean recall (recall) for the same input size. Although the frame rates (fps) of the three algorithms are close to each other, the average precision (mAP) of the algorithms is especially critical in the courier intelligent sorting system, which directly determines the accuracy of courier classification. Considering all the indexes, YOLOv8 achieves the best balance between accuracy and efficiency, so this paper finally chooses the model trained by YOLOv8 as the final model deployed into the Raspberry Pi to meet the high accuracy and real-time demand of the logistics sorting task.

The successful development of this system aims to change the traditional logistics sorting mode and realize the intelligent identification and classification of parcels. This can not only significantly reduce labor costs and enhance sorting flexibility, but also promote the development of intelligent logistics, adapt to the rapidly growing demand for e-commerce, and provide a feasible solution for the intelligent transformation of the logistics industry in the future.

2. Algorithm Architecture of the Express Sorting System

2.1. Sorting system

This sorting system relies on a series of advanced algorithms to achieve efficient and accurate item sorting. First, the system utilizes cameras for object detection, precisely identifying items to be sorted. Subsequently, deep learning algorithms are employed to extract features from the captured images, enhancing target recognition capabilities. Following this, machine learning algorithms are used to classify the items, determining their respective categories. Finally, motion planning algorithms control the movement of the sorting machine, ensuring items are accurately sorted into designated areas.

Within the object detection algorithms, deep learning and machine learning techniques are integrated. Deep learning algorithms focus on image feature extraction, while machine learning algorithms are dedicated to classification tasks. For controlling the sorting machine’s motion, motion planning algorithms such as dynamic programming and genetic algorithms are typically employed to ensure sorting accuracy and efficiency.

2.2. Object Detection Algorithms

In the field of target detection, YOLO series algorithms are favored in industrial applications due to their high efficiency, among which YOLOv5 and YOLOv8 algorithms are based on a single-stage detection framework, using multi-scale feature fusion and anchor frame mechanism, intensively sampling the whole image, and realizing end-to-end feature extraction and target regression with the help of convolutional neural network to significantly improve the detection efficiency, which makes it able to meet the needs of large-scale real-time data processing of the needs of large-scale real-time data processing for express sorting systems.

With optimized network architecture and data enhancement strategies, YOLOv5 can stably detect packages even under complex lighting and partial occlusion, providing reliable support for positioning and sorting in automated sorting.

YOLOv8, as a subsequent iteration, has significant improvements in model architecture, incorporating advanced loss functions and training strategies. This not only gives it a detection accuracy of over 98.5% on the COCO dataset, but also expands the multimodal data processing capability to analyze video streams in real time. Experiments show that YOLOv8 has a single image detection speed of 12.5 ms, with a 23.6% lower false detection rate than the previous version. Its enhanced attention mechanism can effectively distinguish texture features of densely stacked parcels, providing a more reliable sorting system.

The research of Qing Shimingbo et al. in "YOLOv4 Algorithm Based on Attention Mechanism for Intelligent Waste Sorting" provides reference for this paper to improve and apply YOLO series algorithms [8], and it also inspires thinking about the adaptability of the algorithm's application scenarios. The technological development from YOLOv5 to YOLOv8 provides different solutions for logistics automation: while YOLOv5 focuses on detection efficiency and adaptability, YOLOv8 emphasizes more on accuracy improvement and function expansion.

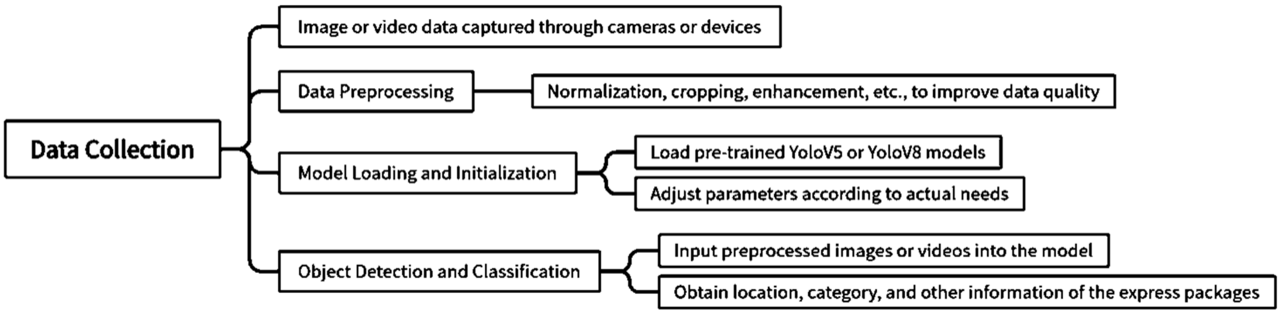

Figure 1. Workflow of YOLOv5 and YOLOv8 in Express Sorting Systems

As shown in Fig. 1, in the sorting system, the camera or device first captures images and video data, followed by pre-processing operations such as normalization, cropping, and enhancement of these raw data to improve the data quality. Subsequently, the system loads the pre-trained YOLOv5 or YOLOv8 model and adjusts the parameters according to the actual situation. Finally, the pre-processed data is input into the model, which outputs information such as parcel location and category to complete the whole sorting process.

3. Hardware and software architecture

3.1. Oerall structure

The architecture of the computer vision-based courier sorting system designed in this paper consists of multiple functional modules, which collaborate with each other to form an organic whole, and its overall structure is shown in Fig. 2, which includes an image acquisition module, a grayscale trajectory module, a motor drive module, a master control module, an ultrasonic obstacle avoidance module, a DC motor, a display module, and a robotic arm grasping module. In the figure, the arrows indicate the flow direction of the signals, showing the information interaction relationship between the modules.

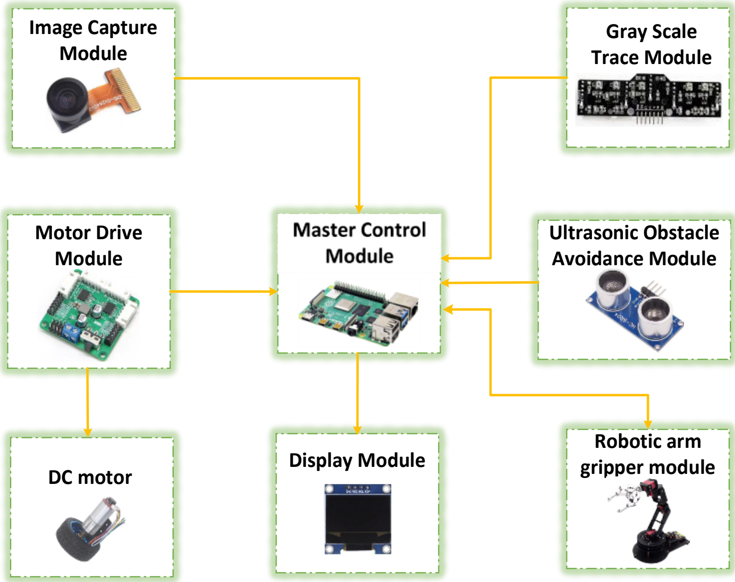

Figure 2. Overall Architecture Diagram

The functions of each module are as follows: the main control module (Raspberry Pi) serves as the core to receive and process the information from the image acquisition, grayscale trajectory, ultrasonic obstacle avoidance and other modules. The image acquisition module is responsible for capturing the barcode and appearance data of express parcels. The grayscale trajectory module monitors the deviation of the trajectory of the sorting trolley by sensing the intensity of light reflection, and assists the main control module in adjusting the trolley's driving direction. Ultrasonic obstacle avoidance module utilizes ultrasonic sensors to detect obstacles and provide timely feedback information to allow the cart to avoid collision. The motor drive module controls the start, stop, steering and rotation speed of the DC motor according to the master control instruction to provide power for the trolley. The DC motor outputs torque to ensure that the trolley travels according to the preset route. The display module displays the system operation status, parcel recognition results and other data in real time, which is convenient for the operator to monitor. The robotic arm grasping module automatically adjusts the grasping method according to the master control instruction to complete the task of grasping and placing the parcels. Deng Chao in the “intelligent logistics in the robotic arm autonomous sorting system design” around the robotic arm in the intelligent logistics sorting application to discuss [9], for this paper robotic arm gripping module design and optimization provides ideas and reference basis.

3.2. Software Architecture

3.2.1 YOLOV8 Object Detection Algorithm

As an innovative version of the YOLO series of object detection algorithms, YOLOv8 inherits the high-efficiency characteristics of the single-stage detection framework and achieves a comprehensive improvement in detection performance through systematic architecture optimization. This algorithm adopts a modular design concept and is composed of three parts: a feature extraction backbone network, a multi-scale feature fusion module, and a decoupled detection head, forming a complete end-to-end detection system.

In terms of feature extraction, YOLOv8 deeply improves the classic CSPDarknet53 backbone network and innovatively designs the C2f module as the core feature extraction unit. Compared with the previous C3 module, C2f optimizes the size of the first-layer convolutional kernel (from 6×6 to 3×3) and adopts a dynamic channel number adjustment mechanism (reducing the number of channels by 50% at each level). While maintaining feature expressiveness, it reduces the computational complexity by approximately 30%. Its unique branch stacking strategy (configured with a 3-6-6-3 structure) and gradient shunting design significantly improve the information flow efficiency of the deep network, increasing the model convergence speed by about 18% and enhancing the feature space perception ability. With the extended SPPF pyramid pooling module, the backbone network can construct multi-level feature maps covering 20×20 to 80×80, laying the foundation for subsequent feature fusion.

In the feature fusion stage, YOLOv8 abandons the traditional unidirectional feature pyramid network (FPN) and uses an enhanced path aggregation network (PANet) to construct a two-way feature transmission channel. This architecture achieves a breakthrough in the organic integration of low-level detailed features and high-level semantic features through two-way information interaction from bottom to top and from top to bottom. In specific implementation, the algorithm performs upsampling fusion and cross-scale connection on the three layers of feature maps (80×80/40×40/20×20) output by the backbone network respectively to construct a feature pyramid with strong spatial perception ability. This two-way feature fusion mechanism increases the detection accuracy of small objects by about 15%, and can effectively deal with the problem of drastic changes in object scale, especially in complex industrial scenarios, as shown in Figure 3.

The detection head adopts a decoupled design architecture, breaking through the parameter sharing limitation of traditional coupled heads for positioning and classification tasks. By independently constructing the regression branch and classification branch networks, the model can more attentively optimize the bounding box positioning accuracy and category discrimination ability respectively. In the training strategy, the Distribution Focal Loss function is innovatively introduced into the regression branch, and an integral regression method based on continuous spatial probability distribf the weights of positive and negative samples. This decoupled design, combined with the adaptive anchor box mechanism, finally achieves a mAP detection accuracy of 53.8% on the COCO dataset, an increase of 4.2percentaution is adopted, reducing the average positioning error of the bounding box by 23%. The classification branch effectively alleviates the problem of class imbalance through dynamic adjustment oge points compared with YOLOv7. Overall, YOLOv8 achieves a significant balance between detection speed (650FPS@RTX3090) and detection accuracy through three key technical breakthroughs: lightweight reconstruction of the backbone network, innovation of the two-way feature fusion architecture, and optimization of task decoupling. Its modular design concept also provides good scalability for industrial-level deployment and has become one of the most competitive solutions in the field of real-time object detection.

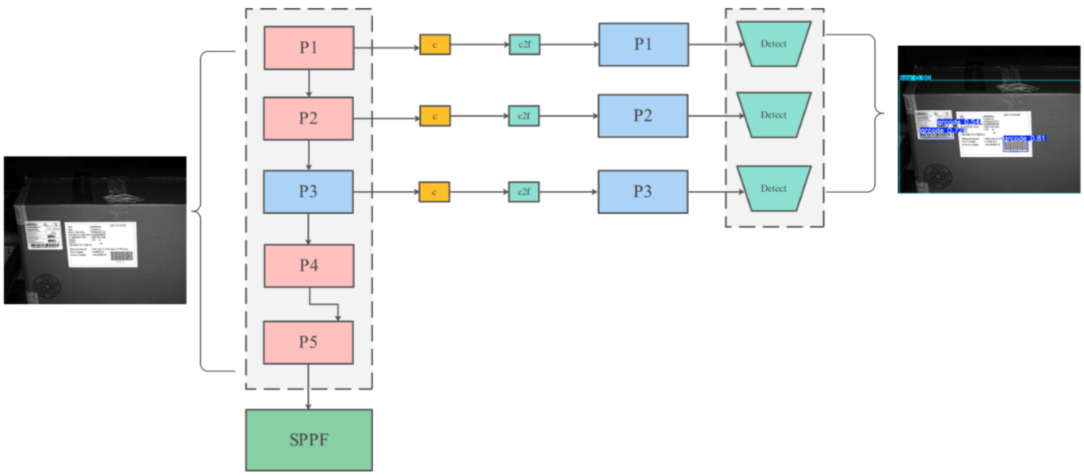

Figure 3. Network diagram of YOLOv8

3.2.2 QR Code Recognition

Use the YOLO series of algorithms to draw bounding boxes around the QR codes on logistics express packages. Then, identify the QR codes within the bounding boxes to obtain the logistics information. Li Yifan in “based on the two-dimensional code recognition of the courier sorting system” in the two-dimensional code recognition in the courier sorting in the application of research [10], for this paper provides some theoretical and practical reference.

In this thesis, a set of Python code for implementing the QR code recognition function was written by using the library functions of OpenCV. As an open-source library widely used in the field of computer vision, OpenCV has a large number of rich and powerful image processing and analysis functions. When performing the task of QR code recognition, with the help of a series of functional modules provided by OpenCV, such as image reading, preprocessing, feature extraction, and pattern recognition, through in-depth research and continuous attempts, a set of Python code for QR code recognition with high efficiency, accuracy, good stability, and robustness was successfully developed. By collecting, meticulously processing, and deeply analyzing various images, this code can locate the specific position of the QR code in the image at a relatively high speed and with high precision, and perform effective decoding operations on it, thus achieving the accurate extraction and full utilization of the information carried by the QR code.

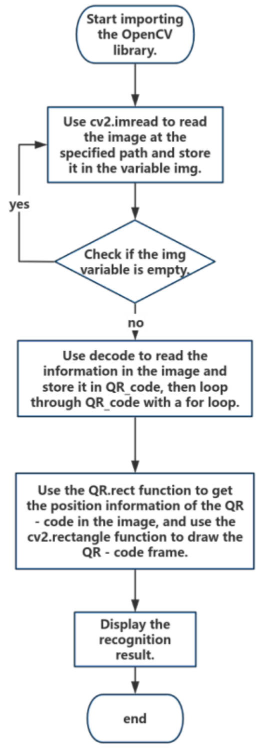

Figure 4. Flowchart of QR Code Recognition

The process of QR code recognition is shown in Figure 4. First, import the opencv library and use the imread function built into opencv to read the picture selected after the YOLOv8 object detection. Use the decode function to scan the pixel information in the picture, and then iterate through all the pixels. Use the rect function built into opencv to determine the position of the selection box for the barcode or QR code. Finally, use the rectangle function to draw the border of the barcode or QR code, and print out the information in the barcode or QR code. The recognition result is shown in Figure 5.

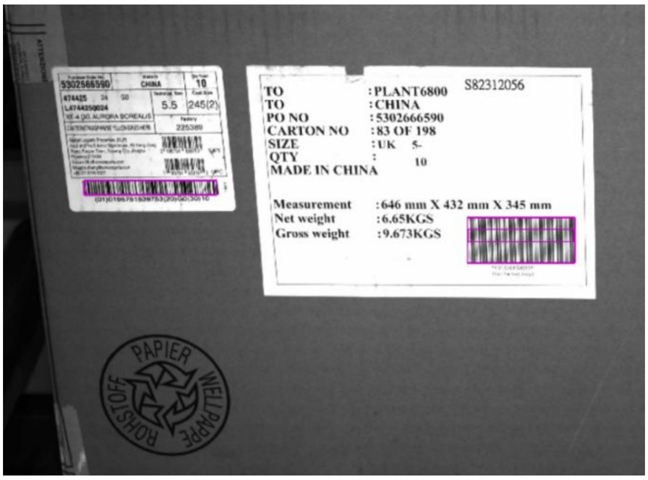

Figure 5. The QR code recognition result

4. Experiments

4.1. Express Sorting Dataset

The rapid development of object detection algorithms has greatly improved the accuracy and efficiency of object detection, and at the same time, it has also promoted the development of large-scale datasets applicable to different fields. An object detection dataset is a collection of images (or video frames) used for training and evaluating object detection algorithms. These data are crucial for an object detection model to learn how to recognize different objects and determine their positions in the images. Common object detection datasets include various datasets such as PASCAL VOC (Visual Object Classes), Caltech Pedestrian Dataset, COCO, etc. These datasets are widely used in relevant research and applications of pedestrian detection in fields such as intelligent transportation systems and security monitoring, and are of great significance for improving the accuracy and robustness of pedestrian detection algorithms.

In the field of express sorting and intelligent logistics systems, the categories of objects required to be recognized are very few, mainly including express packaging boxes and the QR code information on express waybills. Among them, paper packaging boxes are the most common type of express packaging boxes. In order to be able to use the YOLOv8n model for training and deployment on the Raspberry Pi, in this thesis, an express sorting dataset was self-built, and by combining the express package images found on the Internet and the images taken in the laboratory, annotation was carried out. The total number of images for training and testing is 1,021. Among them, 128 images are used as the validation set. Through the LabelImg software, the annotation format is in the YOLO dataset format. It includes two folders, namely "images" and "label". Under the "images" folder, there are "train" and "val" folders. The "train" folder stores the training set images, and the "val" folder stores the validation set images. Under the "label" folder, there are also "train" and "val" folders, which store the. txt files of the training set and the validation set respectively. Among them, there are a total of 1,300 selection boxes for QR codes (qrcode), and a total of 1,900 selection boxes for express boxes (box). The annotation process and the annotation results are shown in Figure 6.

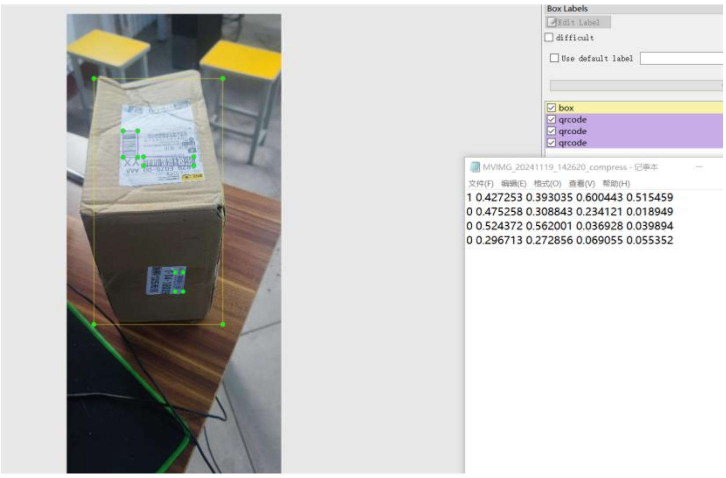

Figure 6. The annotation process and annotation results

4.2. Comparative Experiments

Table 1. Table of Comparative Experiment Results

algorithm	size	mAP/%	fps	precision	recall
YOLOv5	640×640	57.2	26	87.13	83.10
YOLOv8	640×640	62.8	27	89.86	82.49
YOLO11	640×640	62.6	26	89.88	80.13

The table lists the results of training and validating models of YOLOv8, YOLOv5, and YOLO11 on the dataset of this thesis. The experimental data shows that YOLOv8 demonstrates excellent comprehensive performance in terms of the balance between accuracy and efficiency. As shown in Table 1, under the standard input resolution of 640×640, YOLOv8 significantly surpasses the YOLOv5 baseline model (57.2%) with a mean Average Precision (mAP) of 62.8%, achieving a 9.8% improvement in accuracy. Moreover, it still maintains a slight advantage of 0.2% over the newer YOLO11 model (62.6%).

In terms of the core indicators of detection accuracy, the precision value of YOLOv8 reaches 89.86%, which is 3.1 percentage points higher than that of YOLOv5 (87.13%), indicating that its false positive rate is effectively controlled. It is worth noting that while maintaining high accuracy, the recall value of this model is 82.49%. Although it is slightly lower than that of YOLOv5 (83.10%), it is significantly better than that of YOLO11 (80.13%). This characteristic verifies the effectiveness of the optimization strategy of YOLOv8 in terms of the trade-off between precision and recall.

The analysis of the real-time performance dimension shows that YOLOv8 ranks first among the compared models with a processing speed of 27 frames per second (fps), achieving a 3.8% increase in the frame rate compared to YOLOv5 (26fps). It is worth noting that this speed advantage is achieved under the constraint of only a 5.3% increase in the model's parameter quantity, which is mainly attributed to the introduction of the spatial pyramid convolution module and the optimization and reorganization of the calculation path.

In conclusion, among the three trained models, YOLOv8 has the best comprehensive performance indicators. Therefore, the YOLOv8 model is used as the final model.

4.3. Visualization Of Training Results

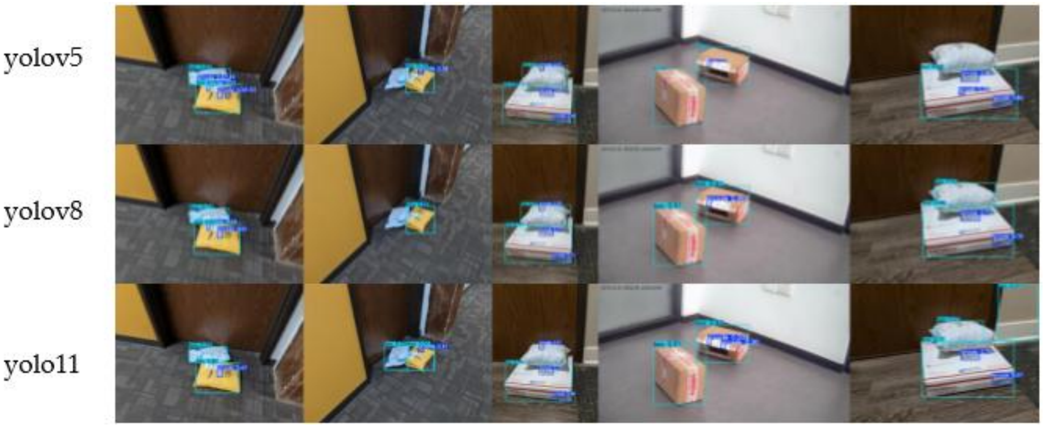

Figure 7. The visualization results of Model

From the results of the visualization experiments in Figure 7, YOLOv8 demonstrates outstanding advantages in prediction. High Confidence Detection: In multiple sets of detection images, YOLOv8 shows excellent prediction confidence for target objects such as parcels. For example, in some images, it has a high confidence value for detecting yellow parcels, white foam parcels, etc. , indicating its strong reliability in target recognition and its ability to accurately determine the category of target objects. Stable Positioning Ability: The labeled bounding boxes can stably and accurately frame the target objects. Regardless of the placement angle of the objects or the background environment, it can achieve good positioning, which reflects its stability and accuracy in target positioning. Comprehensive Performance Advantage: Compared with YOLOv5 and YOLO11, YOLOv8 has more balanced prediction results in different image scenarios. There is no obvious situation of too low confidence, and it also maintains a good level of accuracy in bounding box positioning. Its comprehensive performance is better than the other two, making it more suitable for the detection tasks of objects such as parcels.

5. Conclusions

In this paper, a computer vision-based courier sorting robot system is proposed, aiming to solve the problems of low efficiency and high cost of traditional manual sorting. The system consists of an image acquisition module, a grayscale trajectory module, a motor drive module, a master control module (Raspberry Pi), an ultrasonic obstacle avoidance module, a DC motor, a display module, and a robotic arm grasping module, and the modules work in concert to realize the intelligent identification and classification of express parcels. The experimental results show that YOLOv8 achieves 62.8% in mean accuracy (mAP) and 27 in frame rate (fps). Compared with YOLOv5 and YOLO11, YOLOv8 achieves the best balance between accuracy and efficiency, and is able to satisfy the high-precision and real-time requirements of logistics sorting tasks.

In our future work, we will further improve the YOLO algorithm and optimize the network structure and training strategy to speed up the inference and improve the detection accuracy. Meanwhile, we plan to expand the dataset to cover more complex scenes and diverse parcel types to enhance the generalization ability and robustness of the system. In addition, we will explore the combination of edge computing and cloud computing to further enhance the processing capability and response speed of the system and provide stronger technical support for the intelligent transformation of the logistics industry.

Acknowledgements

Supported by the Undergraduate Training Program on Innovation and Entrepreneurship grant Digital Micromirror-Based Biochip Photolithography System. Project Number: 202410251115X

References

[1] LIU Jiqiu, HU Lifu, YU Han, et al. Design of courier sorting robot based on Open CV [J]. China Science and Technology Information, 2021, (01): 80-81.

[2] Han Xing. Research on Intelligent Express Sorting System in Complex Scenes Based on Deep Learning [D]. Beijing University of Posts and Telecommunications, 2020. doi:10. 26969/d. cnki. gbydu. 2020. 001048.

[3] WU Xue-Feng,QI He-Nan,LI Dong-Ya. Design and research of logistics sorting robot [J]. Electromechanical Engineering Technology, 2022, 51 (01): 111-113+129.

[4] HUANG Jiahui, ZHOU Zisen, LIN Qinyi, et al. Design and implementation of an arduino-based courier sorting robot [J]. Electronic World, 2020, (02): 137-138. DOI:10. 19353/j. cnki. dzsj. 2020. 02. 070.

[5] Jocher, G. (2020). ultralytics/yolov5: v3. 1 - Bug Fixes and Performance Improvements (Version v3. 1) [Computer software]. Zenodo. https://doi. org/10. 5281/zenodo. 415437

[6] Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLO (Version 8. 0. 0) [Computer software]. https://github. com/ultralytics/ultralytics

[7] Glenn Jocher, Jing Qiu. Ultralytics YOLO11 [Software, Version 11.0.0]. 2024. Available at: https://github.com/ultralytics/ultralytics. Licensed under AGPL - 3.0. ORCID: 0000-0001-5950-6979, 0000-0002-7603-6750, 0000-0003-3783-7069

[8] Qing L I, Yuanqiang G, Wei Z, et al. Attention YOLOv4 Algorithm for Intelligent Waste Sorting[J]. Journal of Computer Engineering & Applications, 2022, 58(11).

[9] Deng Chao. Design of an Autonomous Sorting System for Manipulators in Intelligent Logistics [D]. Xi'an University of Technology, 2021.

[10] Li Yifan. Express Sorting System Based on QR Code Recognition[D]. Xi'an: Xi'an University of Science and Technology, 2019.