# RESEARCH AND DESIGN OF MULTI-TARGET IDENTIFICATION AND LOCALISATION SYSTEMS

Mengmei WANG[1], Wen XIN[2], Zhen LIU[3]

*This paper designs and develops a multi-target recognition and localisation system based on shape, position and size detection, and achieves accurate recognition and localisation of multiple targets such as rectangles, triangles, circles, polygons and so on in an image by building a hardware platform and a software system. On the hardware side, this paper adopts the M2 core board and V831 chip; on the software level, by adopting the advanced YOLO algorithm and DeepSort algorithm, the system successfully achieves accurate recognition and continuous tracking of multiple targets in images. To verify the recognition rate of this system, three sets of experiments were designed: one to test the effect of different recognition methods, another to test the effect of different algorithms on the recognition rate, and a third to test the effect of the number of single targets on the recognition rate. The goal was to study and analyze the impact of various factors on the system's recognition rate. The experimental results indicate that the system demonstrates good robustness.*

**Keywords**: target recognition; localisation and recognition systems; image recognition; YOLO algorithm

## 1. Introduction

Multi-target recognition and localization is a very important technique in the field of machine vision, aiming to accurately identify and locate multiple target objects from images or videos. This technology has been extensively employed across numerous fields, such as unmanned driving, security monitoring, medical image analysis, and so on. With the continuous development of multi-target recognition and localization technology, institutions and Scholars both domestically and internationally have individually delved into its applications across diverse fields. Multi-target recognition and localization systems are usually designed and implemented according to practical application requirements. Most of these systems adopt deep learning models, such as convolutional neural networks

[1] Lecturer, School of Artificial Intelligence and Software, Ke Wen College, Jiangsu Normal University Xuzhou, Jiangsu, China, corresponding author, e-mail: mengmei_wang@cumt.edu.cn（
[2] School of Artificial Intelligence and Software, Ke Wen College, Jiangsu Normal University Xuzhou, Jiangsu, China, e-mail: 1138543095 @qq.com
[3] School of Electricity, Jiangsu University of Science and Technology Suzhou, Jiangsu, China, e-mail: 920814680@qq.com

(CNN) or target detection algorithms (e.g., YOLO, Faster R-CNN, etc.), to achieve automatic target recognition and localization [1-3].

As deep learning and other technologies continually progress, the accuracy and performance of multi-target recognition and localization techniques have been further improved. However, for target localization and detection, when confronted with a complex background and diverse targets of varying sizes and types, the results are not very satisfactory, and often only large and significant targets can be detected. Zhang Siping et al. [4] introduced a crowd multi-target recognition and tracking approach based on the YOLO target detection algorithm. This method accomplishes the tracking of multiple targets in crowds by extracting pedestrian trajectory and appearance features from visible characteristics of the crowd's multiple targets; Chen Feng et al [5]. proposed a multi-target collaborative localization method based on the improved genetic algorithm, which can achieve the multi-target recognition and localization in a complex scene. Zhou Xinju et al. [6] proposed a localization method for picking points, leveraging deep learning and multi-target recognition of key grape structures., by identifying and segmenting the key structures; combining the intersection between key regions, relative position judgment, and merging methods, and finally designing a region of interest (ROI) selection method based on structural constraints and range re-selection with the low collision of the fruit peduncle, and The center of mass of fruit pedicels in this region was used as the picking point. Yaxin Ye et al [7]. designed a target recognition and positioning system for a disinfection robot, which achieves the identification and positioning of the target for achieving fixed-point disinfection in public places; Junjie Luo et al [8]. used a multi-target recognition method, which adopts infrared camera multi-target recognition method and UWB multi-target recognition method according to the structure and function of the inspection robot, to improve the utilization rate of the robot in the substation. Zhu Jun et al [9]. proposed a method of fish identification and counting based on sonar images, which achieves automatic identification and labeling of fish through the YOLOv5 neural network model, uses a multi-target tracking algorithm based on identification frames to achieve fish tracking and counting, and uses a sequence of historical identification frames to solve the pairing problem of the reappearance of identification frames.

Although existing techniques can identify and locate a single target type and identify multiple target types, it is difficult to obtain the results of locating multiple targets with low error, and fewer people have combined the identification of multiple targets with localization, shape, and color. Therefore, to combine these two aspects of identification and localization, improve the identification rate, and reduce localization error, this paper designs a system that can identify and localize multiple targets to identify shape such as rectangles, triangles, circles and polygons, color, size, and localization.

## 2. Introduction to relevant theories and technologies

### 2.1 YOLO target detection principles

YOLO [10-13], an acronym for "You Only Look Once: Unified, Real-Time Object Detection," is a target detection system that utilizes a singular neural network. Joseph Redmon and Ali Farhadi et al. introduced this system in 2015.YOLO is a single-stage detection algorithm, which is different from two-stage detection algorithms which first finds the location of the object and then determines the type of the object, YOLO does not need to find the region where the object may be present in advance, i.e., Region-Free, but directly makes predictions about the type and location of the object.

YOLO outputs all the detected objects at once, including category and location. The first step of YOLO is to segment the image, which is similar to a sliding window, by splitting the image as large as possible and then making a prediction for each small region after the segmentation. The dexterity of YOLO is that it only requires that the center of the object falls within a small region.

- Segmenting a picture, we assume that a picture is segmented into S2 small regions.
- In each small region, B bounding boxes are predicted, representing the object's position with center coordinates, width, height, and confidence level. Despite having multiple bounding boxes, they all indicate the same category. Therefore, only one object type is required to be predicted per region. The predicted bounding box is shown in Fig. 1, where the thicker the line box the higher the confidence level.
- Identify the box with the highest confidence level among all detections. Then, calculate its IOU with each remaining box. If the IOU surpasses a set threshold, signifying excessive overlap, exclude that box. Continue this procedure for the remaining boxes detected until all have been processed.

The structure of the YOLO model is borrowed from LeNet, using 24 convolutional layers and two fully connected layers and adding residual blocks to them to solve the problem of vanishing gradients.

### 2.2 DeepSORT algorithm

DeepSORT [14-16] uses the ReID model to extract appearance semantic features and add appearance information; cascade matching (Matching Cascade) and trajectory confirmation (Confirmed, Tentative, Deleted) are added. DeepSort main modules:

1) Target detection module: through the target detection network, obtain the target frame in each frame of the input picture

2) Trajectory tracking module: Trajectory prediction and updating through Kalman filtering to obtain a new set of trajectories.

(3) Data Matching Module: Associate the trajectory with the target frame through cascade matching and IOU matching.

The DeepSort process (in Fig.1) involves several steps: first, the detector captures the target frame from the current video frame. Subsequently, Kalman filtering predicts a set of trajectories for the next frame based on the current frame's trajectory set. These predicted trajectories are then matched with the detected target frame in the subsequent frame. Finally, Kalman filtering updates the trajectories of successful matches.

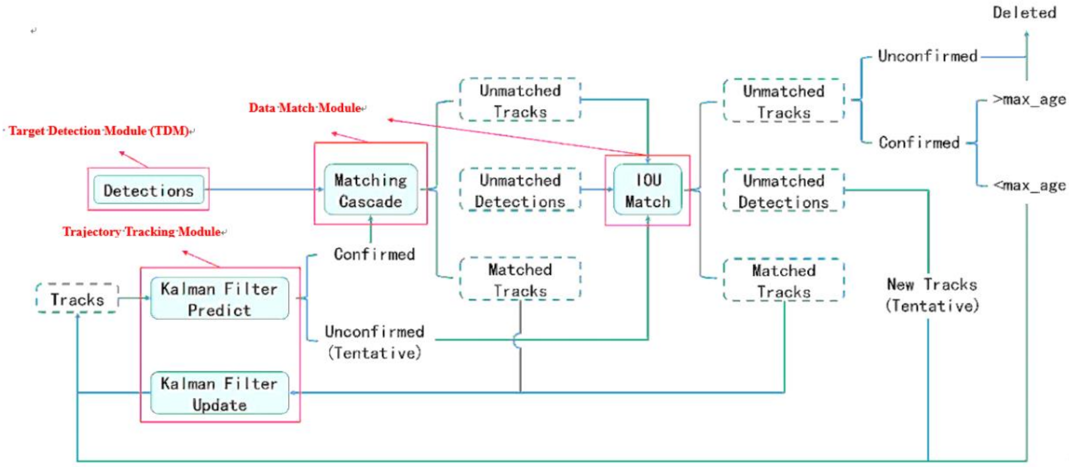



Fig. 1. DeepSort framework

## 3. Multi-target identification and localisation system design

### 3.1 Overall system architecture design

The system's overall design primarily focuses on its functional modules. Based on the demand analysis, the system for multi-target recognition and positioning, utilizing shape, position, and size detection, is primarily categorized into four modules: user login/registration, image acquisition and preprocessing, target recognition, and upper computer display. The structural layout of each functional module is depicted in Fig. 2.

1) User login/registration module: mainly includes user login and user registration.

2) Image Acquisition/Preprocessing Module: It mainly acquires and pre-processes the target image or video.

3) Target recognition module: mainly includes shape recognition, size recognition, position recognition, color recognition, and other four aspects.

4) The upper computer display module exhibits the organization of target recognition and positioning within the image processing function module. This

encompasses information such as the target's shape type, dimensions, color, and centroid position.

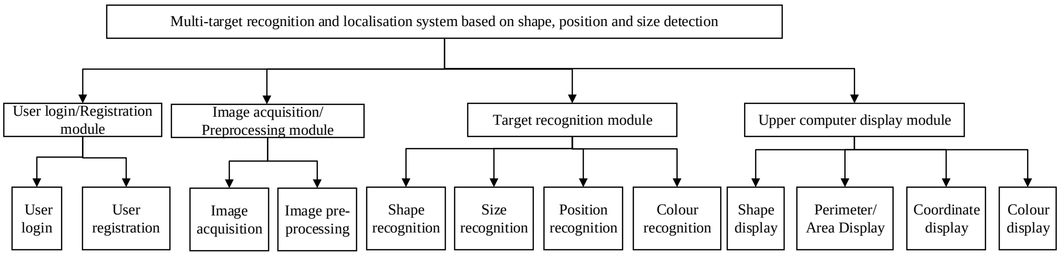
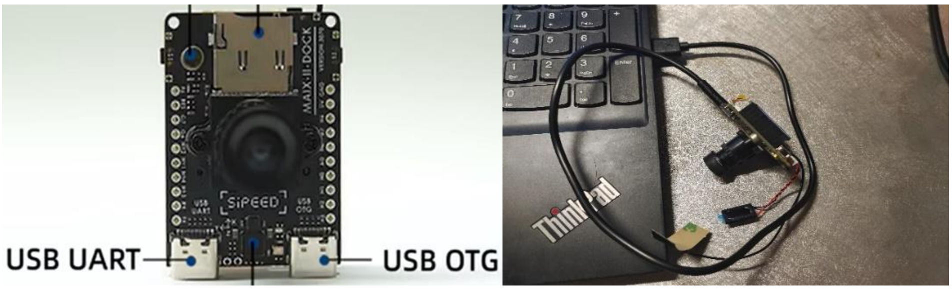


Fig. 2. System overall structure diagram

### 3.2 Hardware Introduction

The hardware part uses the Maix-II-Dock (M2dock) development board based on Allwinner V831 chip. The development board has two Type-C ports, USB UART and USB OTG. The use of the hardware is divided into three steps: the first step is to burn the trained model programme to the memory card; the second step is to insert the memory card into the interface of the development board; the third step is to connect the development board to the computer through the USB UART port and run the programme.



(a) Front Interface Schematic Diagram (b)Connected to PC

Fig. 3. Hardware Connection

### 3.3 System software design

According to the demand analysis, this paper independently designs and develops a multi-target recognition and positioning system based on shape, position, and size detection, through which the system can transform the background-related algorithms into a visual operation interface, that can satisfy the user's usage requirements. The system is based on Jupyter Notebook [21]and MySQL [22], using PyQT [23] framework for development.

The design of the system architecture is primarily outlined in Fig. 4. This architecture can be categorized into two layers: the application layer and the data support layer. The connection between the application layer and the backend database is facilitated through the PyMySQL library. The workflow of the system can be roughly summarized as follows: when the user operation needs to call the background database, the application program connects to the MySQL database through PyMySQL and sends relevant commands to the database, which performs the relevant operations according to the operation statements and returns the results of the execution to the program, and the user can check the results of the operation through the program interface.

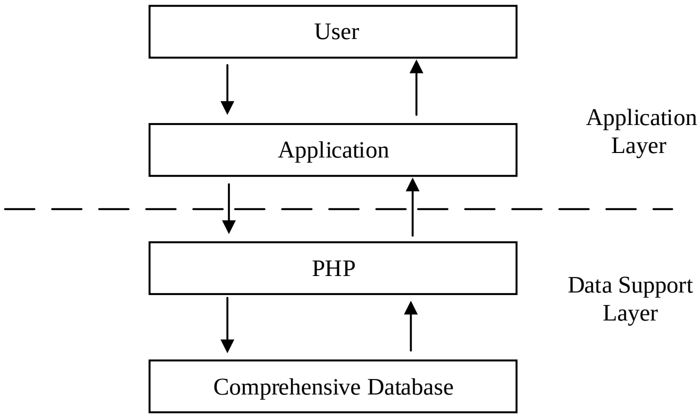


Fig. 4. System Architecture Diagram

## 4. Multi-target recognition and positioning system implementation

The implementation of the system encompasses two key components: the establishment of the hardware system platform and the execution of the software system. This chapter mainly focuses on the construction and implementation of the software experimental platform for multi-target recognition and localization systems based on shape, position, and size detection. On the basis of the completion of the construction of the hardware platform, the software and hardware parts are combined to carry out the test experiment of the whole system. In this chapter, a concise overview of the implementation of each functional module within the software module is provided. Subsequently, experiments are designed to assess the overall system's feasibility.

The multi-target recognition and localization system, which relies on shape, position, and size detection, primarily utilizes the YOLO target detection algorithm and DeepSort target tracking algorithm for target recognition. Once a target is detected and tracked, contour and color features are extracted to accomplish recognition based on color, shape, size, and position.

To detect the shapes of various objects, the implementation described in this paper proceeds as follows: 1) Firstly, the input image is preprocessed such as sample expansion method to enhance the dataset, data enhancement using level flipping and adaptive contrast; 2) then perform binarisation on the image after preprocessing; 3) extract and draw the contour using the contour discovery findContours related function; 4) then approximate the contour using the approxPolyDP function; 5) then use moments Calculate the first order geometric distance to get the center position of the specified contour, and then derive the shape as well as the coordinates; 6) Calculate the area and perimeter.

In order to detect the color of different objects, the implementation of this paper is as follows: 1) first define a color label class, which contains a color dictionary containing all the colors needed; 2) then for each contour (mask), calculate the distance between the current lab* color value and the mean value of the image; 3) finally select the color value represented by the minimum distance. The target recognition flowchart is shown in Fig. 5.

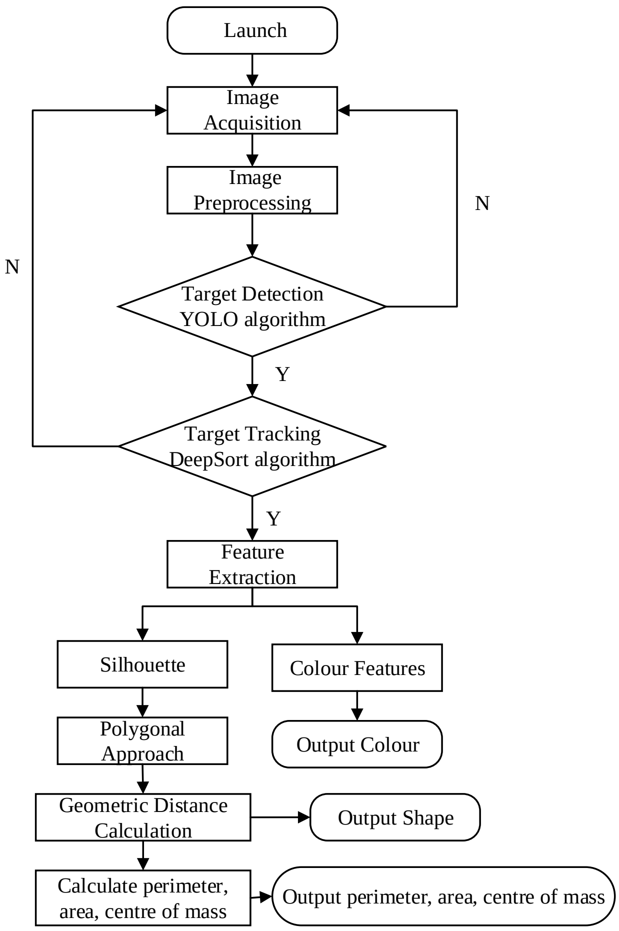


Fig. 5. Target recognition process diagram

## 5. Comprehensive experiment on multi-target recognition and localization

In order to evaluate the recognition performance of this system, the experimental environment in this paper is implemented in Python language, the human-computer interface is built using PyQT framework, and finally run in the integrated environment of Anaconda. To evaluate the recognition performance of this system, self-constructed datasets are employed for both training and testing purposes. Table 1 demonstrates that the training set is utilized to train the algorithm's model, while the testing set is used to assess the recognition effectiveness of the algorithm along with various other functionalities.

*Table 1*

**Experimental data set**

| Type | Training set | Test Set |
|---|---|---|
| Rectangle | 56 | 10 |
| Triangle | 85 | 25 |
| Circle | 75 | 21 |
| Polygon | 101 | 22 |
| Mixed | 156 | 56 |

The experiments conducted for the target recognition module assess the algorithm model's performance by evaluating the accuracy of centralized recognition types. Fig. 6 demonstrates the system's detection effectiveness. Employing the YOLO+DeepSort algorithm, the system exhibits robust shape recognition capabilities, achieving high accuracy and excellent detection results.

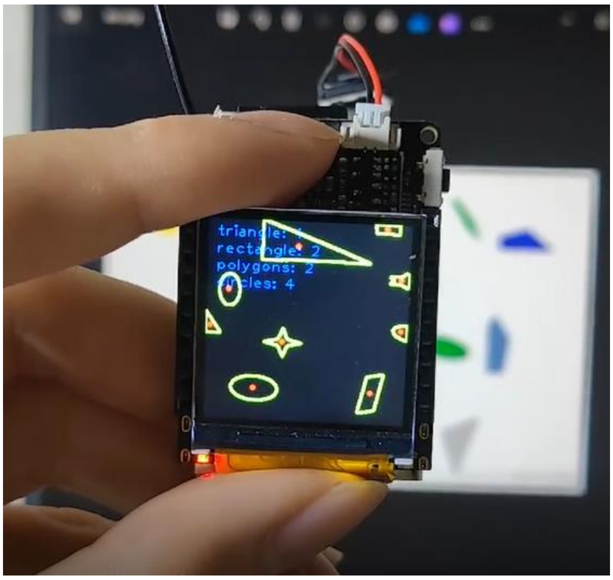



Fig. 6. Target Recognition Results

To enhance the system's detection accuracy, this paper employs the YOLO+DeepSort algorithm for shape recognition. The algorithm model is trained using a dataset and its detection performance is evaluated through a test set, containing 134 images as indicated in Table 1. An essential metric to assess the system is the ratio of successfully recognized targets by the model to the total number of targets in the test set. Specifically, if X represents the number of successfully detected shapes and Y denotes the total number of image shapes in the test set, the recognition rate can be calculated using the formula shown in (1).

$$\mu = \frac{X}{Y} \times 100\% \tag{1}$$

Statistically, in the 134 test set images, there are 568 aircraft targets, and the number of correct results recognized is 554, with a correct rate of 97.54%. After testing, the recognition rate of this system is above 90%, which meets the requirement of the system recognition rate.

### 5.1 Impact of various factors on recognition accuracy

Given the constraints of experimental conditions, the experiments primarily focus on examining the impact of recognition methods, algorithms, and the quantity of individual targets on the system's recognition accuracy.

***1) Effects of different recognition methods***

When the local import method is selected for the recognition of static images, the shape, position, color, perimeter, and area recognition rate is 100%. The recognition system reads the static image data from the local storage device for analysis. Since the images are pre-stored, the system is very capable of processing the static images and is able to accurately extract and identify features such as shape, location, and color and calculate the accurate perimeter and area in the images. This may be due to the high quality of the still images and the fact that they are not disturbed by factors such as jitter and blurring that may occur during real-time capture in Fig. 7.

When the hardware camera was selected to recognize static images, the shape, color, and position recognition rate was 100%, which may be attributed to the good performance of the camera equipment and the relatively stable capture environment without too many interfering factors.

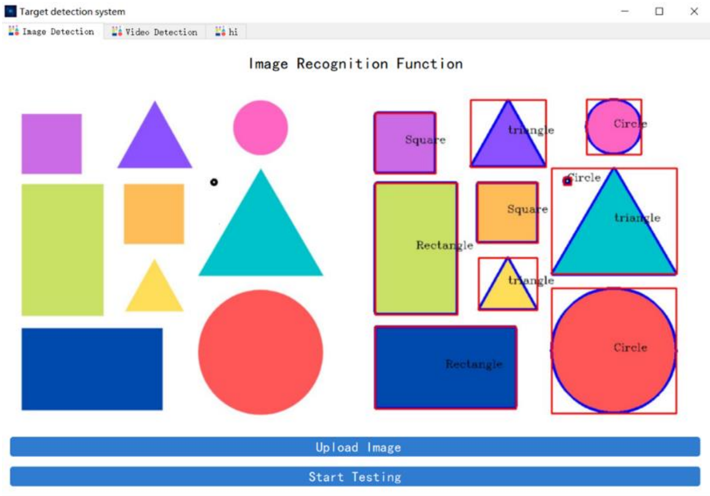



Fig. 7. Static recognition

However, the perimeter and area calculation correct rate of 98% is reduced. The recognition rate may be affected by some real-time capture factors, which may be due to the instability of image quality during real-time capture. For example, camera shake, light changes, or object movement may lead to blurred or distorted image edges, which may affect the accuracy of perimeter and area calculations in Fig. 8.


Fig. 8. Dynamic recognition

When the locally imported video is selected for recognition of moving images, locally imported video files, the recognition system can analyze them frame by frame to identify targets in moving images. Since the target in the video may move or deform in successive frames, the shape, and color recognition rate is 100%. Position, perimeter, and area correct rate 96.63% decreased, which is because the target in the video moves or deforms in successive frames.

When the hardware camera is selected to recognize moving images, the shape, and colour recognition rate is 100%. The position, perimeter, and area correct rate of 97.78% decreases. This may be due to the fact that target movement, scene changes, and possible occlusions in dynamic images have an impact on the performance of the recognition algorithm.

Overall, the recognition rate and computational correctness are affected by a variety of factors such as image or video source, quality, and the real-time capture process. When choosing a recognition method, trade-offs need to be made based on specific application scenarios and requirements. For scenes requiring high-precision measurements, local import may be more suitable; while for scenes requiring real-time feedback, hardware camera recognition may be more appropriate. Concurrently, to enhance the recognition rate and computational accuracy, further optimization of the algorithm, improvement of camera performance, and optimization of the capture environment can be undertaken.

***2) The effect of algorithms on recognition rate***

To evaluate the impact of the algorithm on recognition accuracy, the datasets were divided into four groups based on the algorithms used: YOLOv5, YOLOv8, YOLOv5+DeepSORT, and YOLOv8+DeepSORT, with each group containing 50 images. These groups were carefully selected to have similar target distributions, scenes, and image qualities, ensuring that these factors did not influence the experimental results presented in Table 2.

*Table 2*

**Comparative Data**

| Algorithms | Total Graphics | Total number of recognitions | Recognition rate |
|---|---|---|---|
| YOLOv5 | 186 | 175 | 94.09% |
| YOLOv8 | 186 | 179 | 96.24% |
| YOLOv5+DeepSORT | 186 | 183 | 98.39% |
| YOLOv8+DeepSORT | 186 | 185 | 99.46% |

In Table 2, as can be seen from this set of experiments, the recognition rates of all four algorithms are above 90%, which indicates that all the tested algorithms have high performance in target recognition. Further comparison reveals that the recognition rates of the target detection algorithms alone (e.g., YOLOv5 and YOLOv8) are slightly lower than those of the algorithm combinations (e.g., YOLOv5+DeepSORT and YOLOv8+DeepSORT). This result suggests that the recognition rate can be effectively improved by combining a target detection algorithm with a multi-target tracking algorithm (e.g., DeepSORT) in a target recognition task. The algorithm combination not only locates the target more accurately but also shows better performance in dealing with complex situations such as target occlusion and overlapping.

***3) The effect of the number of targets in a single sheet on the recognition rate***

To assess the impact of the number of single targets on recognition rates, this paper organizes the dataset based on the quantity of targets present in each image: dividing them into four groups with 30 images each, including 1 graphic, 2-3 graphics, 4-5 graphics, and 6 or more graphics. Correct recognition is defined as the ability to accurately identify and output all the categories and the total number of graphics present in each image in Fig. 9.

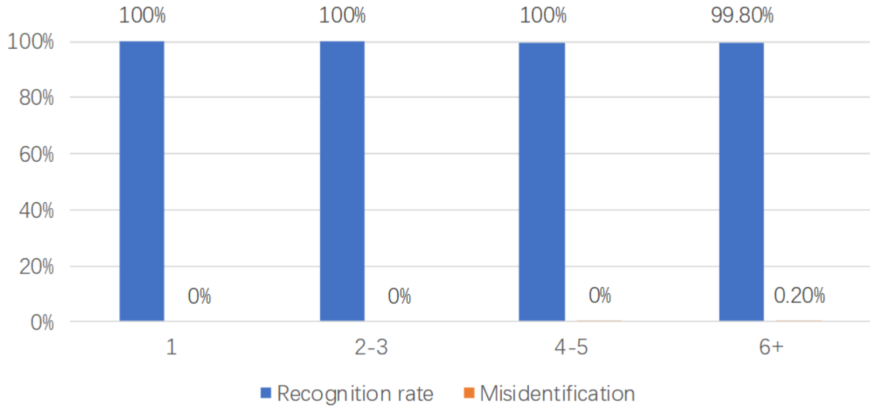



Fig. 9. Recognition rate for different numbers of targets

Based on this series of experiments, it is evident that the recognition rate of the system remains at 100% when the number of single targets is fewer than five. However, as the number of target graphics increases, the recognition rate gradually begins to decline. This downward trend could potentially be attributed to various factors such as occlusion, overlapping, or mutual interference among multiple targets. As the number of targets rises, the system is required to process a greater amount of information, thereby increasing the complexity of recognition. Furthermore, the presence of similar features among multiple targets may lead to misjudgments or missed detections by the system.

## 6. Conclusions

This paper presents a comprehensive study and implementation of a multi-target recognition and localization system that relies on shape, position, and size detection. Through the establishment of an effective hardware platform, the development of a rational software system, and the conduct of extensive experiments and system tests, this paper successfully demonstrates the system's feasibility and performance. The hardware platform constructed in this paper leverages the strengths of the M2 core board and V831 chip to guarantee efficient operation and stable performance of the system. Additionally, the design of the base board takes into account expandability and ease of use, facilitating future system upgrades and optimizations.

In terms of the software system, accurate recognition and positioning of multiple targets in the image is achieved by using the YOLO algorithm and DeepSort algorithm. In comparison to prior results, this system exhibits enhanced recognition rates and positioning accuracy, particularly demonstrating robust performance in handling complex backgrounds and variable lighting conditions.

However, despite the results achieved by this system, there are still some outstanding problems to be solved. For example, the recognition rate of the system may be affected under extreme lighting conditions or when the target is heavily occluded. In addition, with the continuous expansion of application scenarios, the demands for real-time performance and processing speed of the system are continually escalating.

To tackle these challenges, future research can focus on several avenues: enhancing the image processing algorithm for improved recognition in complex environments, exploring more efficient hardware acceleration methods to boost real-time performance and processing speed, and expanding the system's application scope to include diverse fields like intelligent transportation and security monitoring.

**Acknowledgement**

This research is supported by the Jiangsu University Philosophy and Social Science Research Project 2023SJYB1166.

## R E F E R E N C E S

[1] *Chen Feng, Lin Zhipeng, Liu Yuchao, Zhu Qiu Ming.* A multi-target identification and localisation method based on improved genetic algorithm. Proceedings of the 18th Annual National Conference on Radio Wave Propagation. Chinese Society of Electronics: 2023: 479-482.

[2] *Zhou Peng, Cao Bingyu, Zhou Wenjing, et al.* Research on target recognition and localisation of balsam pear based on deep learning. Computer Knowledge and Technology, 2023, 19(25):38-41.

[3] *Zhang Siping.* A method for multi-target recognition and tracking of crowds based on the YOLO object detection algorithm. Intelligent Computer and Applications, 2024, 14(01):152-155.

[4] *Zhang S-P.* Crowd multi-target recognition tracking method based on YOLO target detection algorithm. Intelligent Computer and Applications, 2024, 14(01): 152-155.

[5] *Liu Jian, Chen Liang.* Research on Student Behavior Recognition Based on YOLOv4. Journal of Shenyang University of Technology, 2023, 42(03): 16-21.

[6] *Zhou Yibo, Ma Yutao, Zhao Yanru.* Design of an Apple Bark Disease Identification System Based on YOLOv5s and Android. Guangdong Agricultural Sciences, 2022, 49(10): 155-163.

[7] *Niu Han, Wu Xizhi.* Multi-target recognition and localisation of pine cones based on the connected domain of Otsu algorithm. Jiangsu Agricultural Science, 2021, 49(15): 193-198.

[8] *Yun Cheng, Xian Gyang, Luo*. Multi-modal deep feature learning for RGB-D object detection. Pattern Recognition: the Journal of the Pattern Recognition Society, 2017, 72(300-313).
[9] *Latha R S, Sreekanth G R, Suganthe R C.* Automatic Fruit Detection System using Multilayer Deep Convolution Neural Network; proceedings of the 2021 International Conference on Computer Communication and Informatics (ICCCI), F, 2021.
[10] *Fu L, Duan J, Zou X.* Fast and Accurate Detection of Banana Fruits in Complex Background Orchards. IEEE Access, 2020, 8(196835-196846).
[11] *Tu S, Pang J, Liu H.* Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images. Precision Agriculture, 2020, 21(1).
[12] *Yi X, Yu H, Xu T.* Solving multi-objective weapon-target assignment considering reliability by improved MOEA/D-AM2M. Neurocomputing, 2024, Jan.1): 563.
[13] *Fang Dongfei.* Research on vision-based algorithm for accurate positioning of UAV ground targets; Harbin Institute of Technology.
[14] *T. Ngatouo Costel and S. Jiang*. "Research on Multi-Target Vehicle Detection and Tracking Based on Yolo," 2023 International Conference on Service Robotics (ICoSR), Shanghai, China, 2023, pp. 118-124.
[15] *P. Köhl, A. Specker, A.* Schumann and J. Beyerer, "The MTA Dataset for Multi Target Multi Camera Pedestrian Tracking by Weighted Distance Aggregation," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 2020, pp. 4489-4498.
[16] *A. Rizik, E. Tavanti, H. Chible, D. D.* Caviglia and A. Randazzo, "Cost-Efficient FMCW Radar for Multi-Target Classification in Security Gate Monitoring," in IEEE Sensors Journal, vol. 21, no. 18, pp. 20447-20461.
[17] *D. Heimsch, Y. H. Lau, C. Mishra, S.* Srigrarom and F. Holzapfel, "Re-Identification for Multi-Target-Tracking Systems Using Multi-Camera, Homography Transformations and Trajectory Matching," 2022 Sensor Data Fusion: Trends, Solutions, Applications (SDF), Bonn, Germany, 2022, pp. 1-9.
[18] *T. Keawboontan and M. Thammawichai.* "Toward Real-Time UAV Multi-Target Tracking Using Joint Detection and Tracking," in IEEE Access, vol. 11, pp. 65238-65254.
[19] *A. Specker and J. Beyerer.* "ReidTrack: Reid-only Multi-target Multi-camera Tracking," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Vancouver, BC, Canada, 2023, pp. 5442-5452.
[20] *S. Casao, A. Naya, A. C.* Murillo and E. Montijano, "Distributed Multi-Target Tracking in Camera Networks," 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 2021, pp. 1903-1909.
[21] *K. Shim, S. Yoon, K. Ko and C. Kim.* "Multi-Target Multi-Camera Vehicle Tracking for City-Scale Traffic Management," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 2021, pp. 4188-4195.
[22] *F. Remigereau, D. Mekhazni, S. Abdoli, L. T. Nguyen-Meidine, R. M. O. Cruz and E. Granger.* "Knowledge Distillation for Multi-Target Domain Adaptation in Real-Time Person Re-Identification," 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 2022, pp. 3853-3857.
[23] *M. Kahlert, T. Fei, J. Hirner, C.* Tebruegge and M. Gardill, "Multi-Target Doppler Ambiguity Identification for a PMCW Automotive Radar System," 2023 31st European Signal Processing Conference (EUSIPCO), Helsinki, Finland, 2023, pp. 795-799.