Research on Real time Obstacle Recognition and Avoidance Strategies for Mobile Robots Based on YOLOv5

Liyuan Shao

Kangqiao Campus, Hang Zhou No.14 High School, Zhejiang, China

Abstract. To address the issues of insufficient detection accuracy, slow response speed, and poor adaptability in real-time obstacle recognition and avoidance of mobile robots in complex environments, this paper proposes a real-time obstacle recognition and avoidance strategy algorithm for mobile robots based on YOLOv5. To begin, using the YOLOv5 network model, an image acquisition and data preprocessing module is created to obtain high-precision and low latency obstacle location and classification information. Second, to increase recognition efficiency and reduce false positives, the YOLOv5 model structure and training parameters were optimised. Again, based on YOLOv5's real-time characteristics, an obstacle avoidance model was built that takes into account dynamic changes in the environment as well as the influence of robot mobility status, allowing the robot to better understand its surroundings. Finally, an obstacle recognition and avoidance system based on the YOLOv5 algorithm was developed. Simulation and experimental verification revealed that the algorithm presented in this research can greatly improve mobile robots' autonomous navigation capability.

Keywords: Mobile robots; YOLOv5; Obstacle recognition; Obstacle avoidance strategy.

1. Introduction

Mobile robots are widely employed in industrial automation, logistics and distribution, smart homes, and other industries due to their superior autonomous navigation, high environmental adaptation, and rapid dynamic response. Obstacle recognition systems in real-world mobile robot applications frequently rely on visual input. If the obstacle detection data are not accurate enough, it may result in obstacle avoidance failures or path planning problems. Furthermore, the quality of the picture collecting module might influence the real-time reaction of the obstacle recognition system, which is also affected by variations in external lighting. As a result, obstacle recognition has significant research value for safe and effective autonomous navigation, as well as boosting mobile robot performance.

Traditional obstacle identification algorithms rely mostly on rule-based models for feature extraction and classification, which are separated into two categories: static analysis and dynamic analysis. Traditional algorithms necessitate a large level of prior knowledge. Although the rule-based model is the most widely used model in early obstacle recognition, its description of complex scenes is insufficiently accurate, resulting in problems such as low accuracy and poor generalization capacity in traditional recognition algorithms. As a result, it is vital to investigate the dynamic properties of obstacles and develop more accurate identification models.

This paper proposes a real-time obstacle recognition and avoidance strategy algorithm for mobile robots based on YOLOv5, which addresses the issues of inaccurate detection, high rates of false positives and false negatives in recognition results, and low accuracy in obstacle localization in current obstacle recognition. First, recreate the YOLOv5 network model and apply deep learning to extract accurate location and category information for obstacles based on this model. Second, the training and parameter optimization processes for the YOLOv5 model are updated to produce an optimized YOLOv5 model. Then, an obstacle avoidance strategy algorithm paired with environmental perception is developed to fully recognise barriers utilizing the optimized model. Once again, taking into account the dynamic changes in the environment and the impact of robot mobility on obstacle recognition, and incorporating experimental phenomena for more accurate modeling, the YOLOv5 algorithm is utilized to develop an obstacle recognition model. Finally, the method was validated using simulation and tests, proving the efficacy of the suggested scheme in this study.

2. Background of Mobile Robot Technology

2.1 YOLOv5 Network Model

YOLOv5 is a convolutional neural network (CNN) that solves object detection difficulties. In YOLOv5, the input image contains numerous object instances, and the positions and categories of all objects are predicted concurrently using a forward propagation algorithm. The next section describes the object identification mechanism and model structure of YOLOv5. Figure 1 shows the YOLOv5 network model [1].

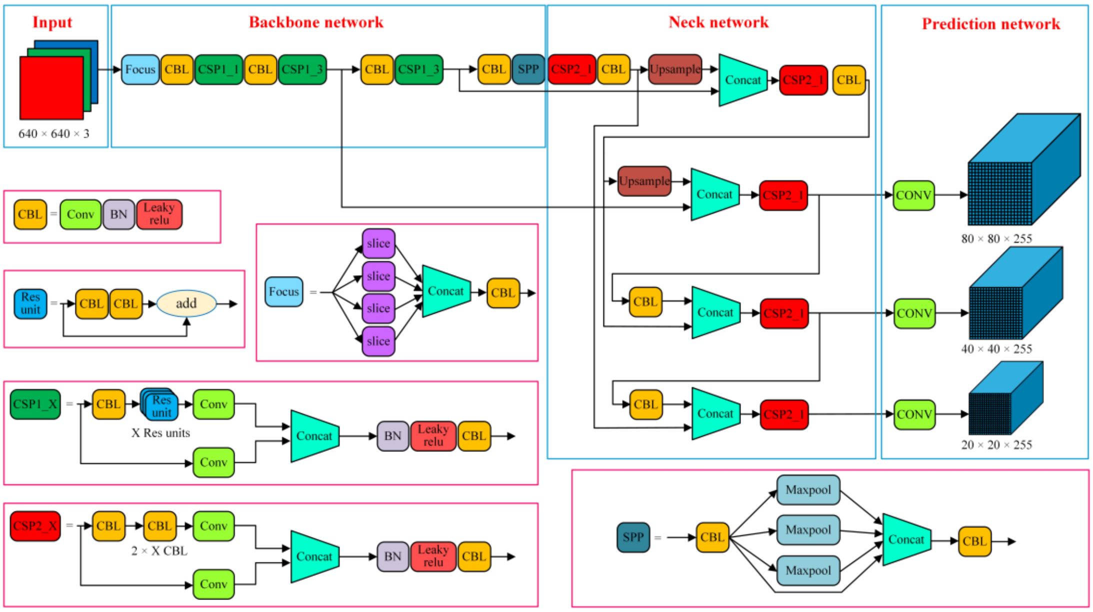

Figure 1. YOLOv5 network model

Each input image has a width W and a height H. The input of YOLOv5 is to scale the image to a fixed size and divide it into $S \times S$ grid cells. Each grid cell is responsible for predicting B bounding boxes and the confidence level of these boxes belonging to each category. The bounding box consists of the center coordinates (x, y), width w, height h, and category c; Among them, x and y are normalized to the $[0,1]$ interval relative to the upper left corner of the grid cell, while w and h are normalized relative to the entire image size. The output of the grid cell is represented as

$$output = (x_1, y_1, h_1, c_1, x_2, y_2, w_2, h_2, c_2, \dots, x_B, y_B, w_B, h_B, c_B) \tag{1}$$

Here, output represents the prediction result of a grid cell for each bounding box, x_i and $y_i$represent the center coordinates of the i-th bounding box, w_i and h_i represent the width and height of the box, and ci represents the corresponding category probability [2].

The confidence score calculation formula for YOLOv5 is as follows:

$$confidence = P(object \cdot IOU_{pred}^{truth}) \tag{2}$$

Among them, $P(object)$ is the probability that the target exists within a given grid cell, IOU_{pred}^{truth} is the Intersection Over Union between the predicted bounding box and the true bounding box. When there is no target within the grid cell, $P(object)$ should be close to 0; When there is a goal, it should be close to 1. Confidence reflects the model's confidence in whether the predicted box accurately surrounds the target.

Each bounding box is also associated with a conditional category probability distribution

$$class\_probilities = (p_1, p_2, \dots, p_C) \tag{3}$$

Here, $class\_probilities$ is a vector of length C, representing the probability that a target belongs to each category given its existence. For example, if the category includes pedestrians, vehicles, etc., then p_i represents the probability that the target within the bounding box belongs to the i-th category.

Finally, in order to train the model, YOLOv5 uses a multitask loss function to optimize position prediction, confidence estimation, and classification accuracy. This loss function takes into account all the parameters mentioned above and adjusts the network weights through backpropagation algorithm to minimize the loss. During the testing phase, the model selects high confidence prediction boxes based on the set threshold as the final detection results.

2.2 Communication Protocol Specifications

Undoubtedly, mobile robots cannot avoid the "communication protocol specification" as a communication infrastructure in real-time data exchange. In the mechanism of obstacle recognition system, communication protocol is a standard and effective information transmission tool, playing an important role in ensuring efficient interaction between mobile robots and the external environment [3]. This also makes communication protocols not only a technical concept, but also a strategic concept. Therefore, communication protocols based on "reliability and low latency" have become the communication guarantee mechanism for mobile robot systems. The practical deduction of communication protocols is generally an optimization path gradually formed on the basis of ensuring data accuracy and real-time performance, although this path involves adaptation attempts to various network conditions. From image acquisition to data preprocessing, communication protocols revolve closely around the seamless transmission of data streams from start to finish. Although communication protocols should strive to simplify and accelerate data exchange to adapt to rapidly changing navigation requirements [4]. However, with increased complexity, this also brings a dilemma of potential packet loss or delay. Overall, there is still room for improvement in terms of stability and efficiency of communication protocols, and their compatibility needs to be further improved, which is also an important task to enhance the performance of mobile robots. The communication protocol specification is shown in Figure 2.

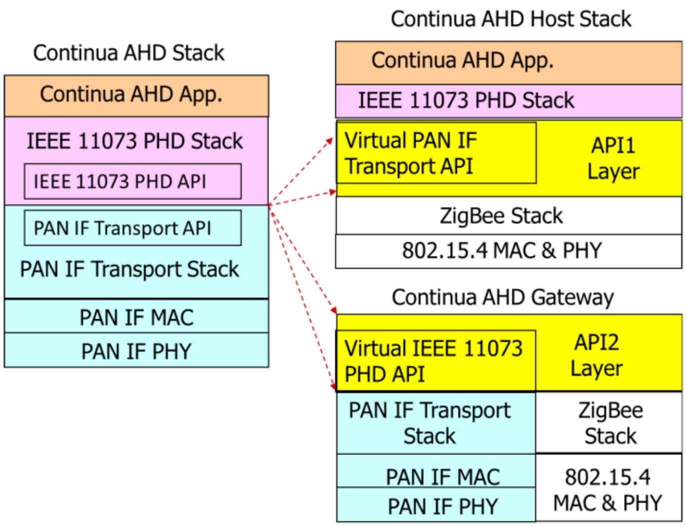

Figure 2. Communication Protocol Specification

3. Obstacle Recognition System Based on YOLOv5

3.1 Image Acquisition Module and Data Preprocessing Module

This section presents the mathematical modeling of image acquisition and data preprocessing problems in the obstacle recognition system based on YOLOv5. Given a set of image frames $I = \{I_1, I_2, .., I_N\}$and a set of preprocessing operations $P = \{p_1, p_2, ..., p_M\}$, the following definitions are given.

3.1.1 Image Frame

The n-th image frame is denoted as I_n: $I_n = (id_n, t_n, S_n)$, Among them, id_n represents the number of image frame I_n (i.e. n), t_n represents the timestamp of image frame I_n, and Sn represents the spatial resolution of image frame In. Taking the navigation scenario of mobile robots as an example, different t_n refers to the images captured at different time points, while different S_n refers to the images with different spatial resolutions that they can provide [5].

3.1.2 Preprocessing Operations

The m-th preprocessing operation is denoted as p_m, defined as $P_m = (id_m, d_m, t_{max}, r_m, g_m)$, where id_m represents the number of the preprocessing operation p_m (i.e. m), d_m represents the size requirement of the input image for the preprocessing operation p_m, t_{max} represents the maximum execution time of the preprocessing operation p_m, r_m represents the image resolution requirement of the preprocessing operation p_m, and g_m represents the output format requirement of the preprocessing operation p_m.

3.1.3 Processing Unit

Each preprocessing operation can be performed by a processing unit composed of a single computing resource or multiple shared computing resources. M computing resources can form a maximum of $2^M - 1$ processing units. The set of these processing units is denoted as $G = \{g_1, g_2, \dots, g_{2^M - 1}\}$, where the k-th processing unit is denoted as g_k, defined as follows: $g_k = (id_k, R_k)$, Among them, id_k represents the number of processing unit g_k (i.e. k), and R_k represents the set of computing resources in processing unit g_k.

3.1.4 Preprocessing Operation Execution Time

The execution time of preprocessing operation p_m by processing unit g_k is represented as τ_{km}, defined as follows:

$$\tau_{km} = \frac{1}{\sum_{r \in R_k} v_{rm}} \tag{4}$$

Among them, v_{rm} represents the processing speed that the computing resource r in processing unit g_k can provide for the preprocessing operation p_m. The above formula is based on real-world application scenarios, such as mobile robots that need to quickly process image information. The processing speed of the processing unit refers to the speed at which the computing resources perform preprocessing operations, and the execution time of the entire processing unit is determined by the slowest computing resource.

3.1.5 Preprocessing Completion Time

The completion time of the preprocessing operation p_m is represented as p_m, defined as follows:

$$T_m = T_{m-1} + \tau_{km} \tag{5}$$

Among them, T_{m-1} is the completion time of the previous preprocessing operation before executing the preprocessing operation p_m (when p_m is the first preprocessing operation, $T_{m-1} = 0$), and τ_{km} represents the execution time of the preprocessing operation p_m by the processing unit g_k.

According to the above definition, the mathematical model for image acquisition and preprocessing problems with time constraints is

$$minT(G) = \min \max (T_m) \tag{6}$$

Among them, G is a feasible solution to the problem, $T(G)$ is the objective function of the problem ($T(G)$ is the maximum completion time of all preprocessing operations), G is the allocation constraint between processing units and preprocessing operations, d_m is the constraint of preprocessing operations on image size, and r_m is the constraint of preprocessing operations on image resolution (the start time of preprocessing operations is greater than the completion time of all dependent preprocessing operations).

3.2 Obstacle Recognition Module

3.2.1 YOLOv5 Model Structure

From the perspective of obstacle recognition in mobile robots, computing resources and algorithm efficiency have long constrained real-time capabilities. Since the 21st century, the YOLOv5 model, which integrates deep learning and object detection, has reshaped the accuracy and speed of obstacle recognition through improved network architecture and optimized training strategies [6]. However, the drawbacks of traditional model structures still constrain its performance in practical applications. Not only due to its high computational complexity, but also due to the strict hardware requirements, the application of YOLOv5 still needs to be improved. YOLOv5 is considered a direct way to achieve rapid obstacle recognition while ensuring efficiency and accuracy. However, the practical role of obstacle recognition systems based on YOLOv5 in autonomous navigation of mobile robots remains to be discussed. Meanwhile, due to difficulties in real-time processing, the system lacks sufficient flexibility to cope with diverse environmental challenges. Therefore, YOLOv5 does not always seem to achieve the goal of ideal obstacle recognition. It is evident that YOLOv5 is not only a technical challenge, but also faces the problem of balancing performance and resource consumption [7].

To solve the above problems, researchers continue to explore methods to optimize the YOLOv5 model structure, such as adopting lightweight network design, introducing attention mechanisms, etc., striving to reduce computational requirements while maintaining high performance, so that YOLOv5 can better adapt to the needs of mobile robot application scenarios. The obstacle recognition module is shown in Figure 3.

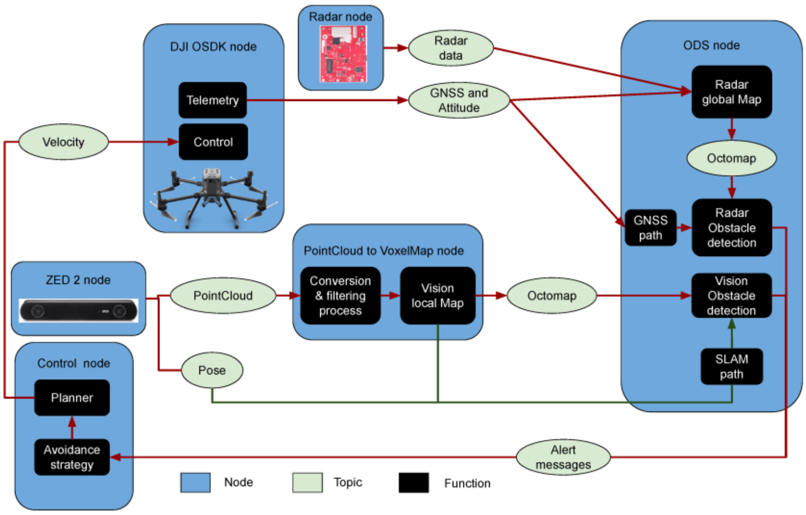

Figure 3. Obstacle recognition module

3.2.2 Model Training and Parameter Optimization Process

As shown in the overall process of the YOLOv5 model. Firstly, initialize the network weights, and then search for the optimal solution to the problem through multiple iterations of training. All predicted boxes found by YOLOv5 are stored in the output tensor.

During each iteration of training, YOLOv5 first updates all training samples. For each sample, first evaluate the fitness value of the loss function L_{total} of the sample and normalize all target values according to equation (1). Then, select the best performing sample as its reference sample according to equation (2):

$$L_{total} = \lambda_{accord}\left(\sum_{i=0}^{S^2}\sum_{j=0}^{B} G_{ij}^{obj}\left[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2\right]\right) + L + \lambda_{noobj}\left(\sum_{i=0}^{S^2}\sum_{j=0}^{B} G_{ij}^{noobj}\left(C_i-\hat{C}_i\right)^2\right) + L + \sum_{i=0}^{S^2} G_i^{obj}\sum_{c\in classes}\left(p_i(c)-\hat{p}_i(c)\right)^2 \quad (7)$$

Among them, L_{total} is the total loss that comprehensively considers position, confidence, and classification accuracy, λ_{coord}, λ_{noobj} are the weight coefficients of different parts of the loss, and G_{ij}^{obj} and G_{ij}^{noobj} are the indicator functions when the jth bounding box of the i-th grid cell is responsible for predicting the object, respectively, (x_i, y_i) and (w_i, h_i) are the center coordinates and width/height of the predicted box, $(\hat{x}_i, \hat{y}_i)$ and $(\hat{w}_i, \hat{h}_i)$ are the corresponding values of the true label box, C_i is the confidence of the predicted box, $\hat{C}_i$ is the confidence of the true label, and $p_i(c)$ is the probability of the predicted category, $\hat{p}_i(c)$ is the true category label.

Next, sort all samples in the training set according to L_{total}, select the top k samples with lower loss values as candidate sample sets, and then randomly select one sample from the candidate sample set as the learning sample x_{best}. The network weights update themselves based on x_{best} and the current weight W_t:

$$W_{t+1} = W_t - \eta \nabla L_{total}(W_t) \tag{8}$$

Among them, η is the learning rate, $\nabla L_{total}(W_t)$ is the total loss gradient at weight W_t. This arithmetic operation is implemented through backpropagation algorithm.

Afterwards, the network weights update their parameters according to the gradient descent rule, and a new feasible solution is constructed through multiple steps. In each step of feasible deconstruction, first select a feasible task that satisfies the precedence constraint based on the task selection probability $p(task)$, and then gradually select a group for the task based on the group selection probability $p(group)$ until the group sequence meets the task's ability and resource requirements. Repeat the above steps until all tasks have been assigned to the group sequence. Afterwards, evaluate the new parameters of the network and update the optimal parameters and training set [8]. After all samples are updated, use the validation set to further improve the quality of the solution. When the maximum number of iterations reaches the specified limit, the training process is terminated and the optimal parameters are output as the best solution found by YOLOv5. The process of model training and parameter optimization is shown in Figure 4.

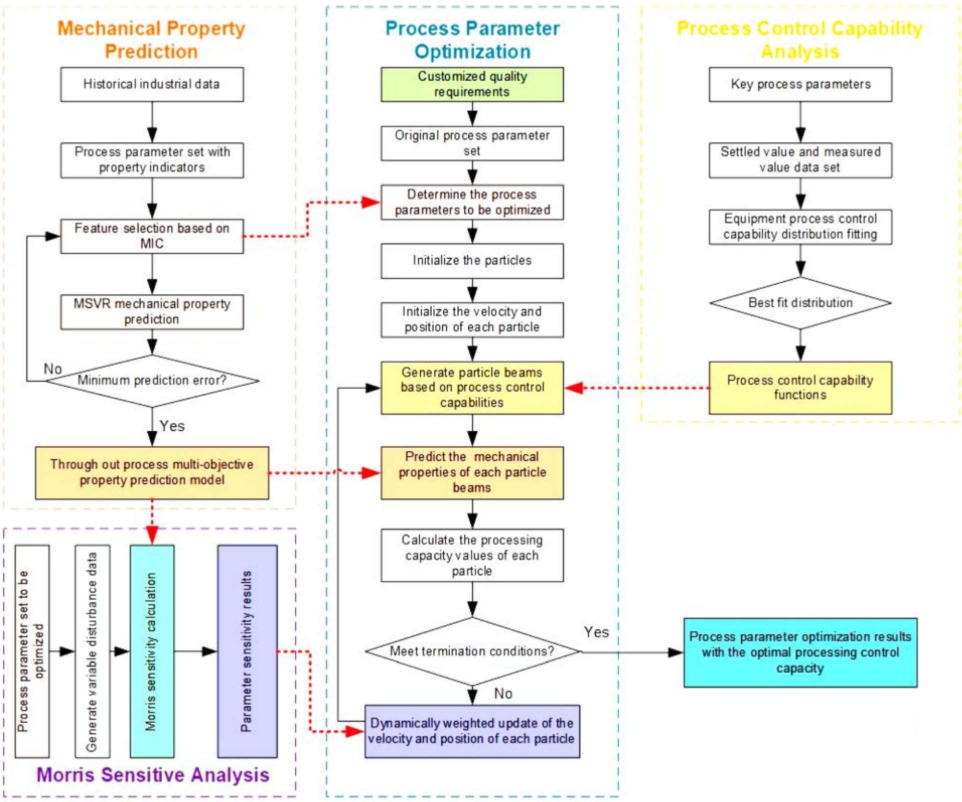

Figure 4. Model training and parameter optimization process

4. Analysis of Experimental Results Based on YOLOv5

4.1 Experimental Process

Traditional evaluation methods cannot effectively offer the extensive performance metrics required for YOLOv5-based obstacle identification systems. The study primarily assesses model satisfaction in terms of accuracy, recall, and F1 score, however it lacks significant information and quantitative procedures for real-time processing capability and environmental adaptation. The root of this issue could be a constraint in evaluating criteria. During the experimentation phase, YOLOv5 is frequently described as a "efficient and accurate object detection tool", and its safety and efficiency in mobile robot navigation directly represent the model's practical application value. However, the majority of the research focuses on static photographs or test data in simple situations, leaving out information about complicated dynamic landscapes. Real-world variables, such as lighting and object occlusion, are typically difficult to gather and measure. Asymmetric information and inadequate evaluation methods directly impede the transition of technological applications.

To address these problems, we employed YOLOv5 trials that not only focused on standard performance indicators, but also incorporated new assessment aspects like as processing speed, model size, energy consumption, and stability under various environmental circumstances. A series of simulation studies and field testing were designed to ensure the model's consistency and reliability under diverse practical settings, offering more solid technological support for autonomous mobile robot navigation [9].

4.2 Evaluation Indicators and Methods

The essential distinction between YOLOv5-based obstacle recognition systems and older approaches is their real-time capabilities. The system's evaluation and selection criteria are designed to ensure safe navigation of mobile robots in complicated situations, with the primary focus on improving detection accuracy and environmental flexibility. The system's evaluation framework identifies accuracy, recall, F1 score, and processing speed as the main values and highest criterion for system development. The diversity of existing evaluation markers, together with changes in application contexts, creates a multidimensional evaluation trend [10]. However, due to insufficient evaluation in some unstructured settings, the system does not include an adaptive adjustment mechanism for constantly changing surroundings. As a result, there is a "shortcoming" in the evaluation method, which has an impact on the overall knowledge of model performance. Figure 5 shows the obstacle recognition results based on YOLOv5.

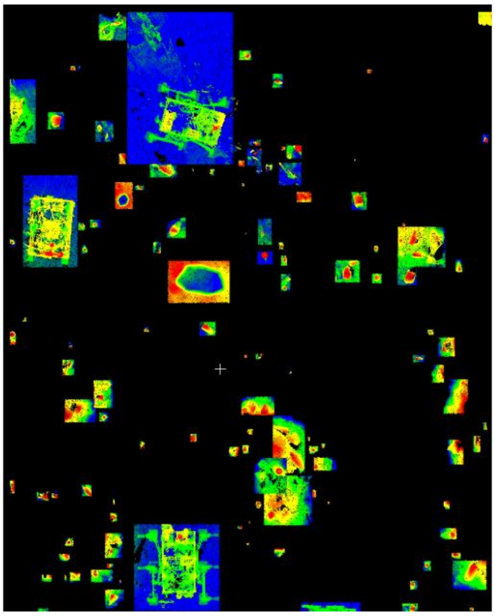

Figure 5. Obstacle recognition results based on YOLOv5

To overcome these challenges, this study introduced a series of comprehensive evaluation indicators and methods. In addition to traditional static image evaluation, special attention is also paid to real-time processing efficiency, model size, energy consumption, and stability testing under different lighting and occlusion conditions. By building a framework that includes multiple evaluation dimensions, we aim to provide a more comprehensive and objective evaluation perspective, ensuring that obstacle recognition systems based on YOLOv5 not only perform well in laboratory environments, but also demonstrate superior performance in real-world applications [11].

4.3 Result Analysis

From the perspective of performance evaluation, the analysis of experimental results is a fundamental step in validating the YOLOv5 model and a core manifestation of the obstacle recognition system for mobile robots. Therefore, the result analysis is mainly based on quantitative indicators as the main generating logic. Accuracy is the main parameter for evaluating models and the main reflection of system performance. At present, this study strengthens the reliability control of results from multiple aspects, mainly in three forms: accuracy comparison [12]. Clearly define YOLOv5's ability to achieve high-precision detection across different environmental conditions and complexities; The second is the standardization process. By establishing strict testing standards, data preprocessing standards, and publicly disclosing these standards to the academic community, standardized control of model evaluation can be achieved; The third is internal process optimization. In recent years, the introduction of automated testing tools and data analysis platforms has improved experimental efficiency and increased the credibility of results. However, compared to the ideal state, the practical applicability of the current model still needs further improvement. The accuracy comparison of YOLOv5 model is shown in Figure 6.

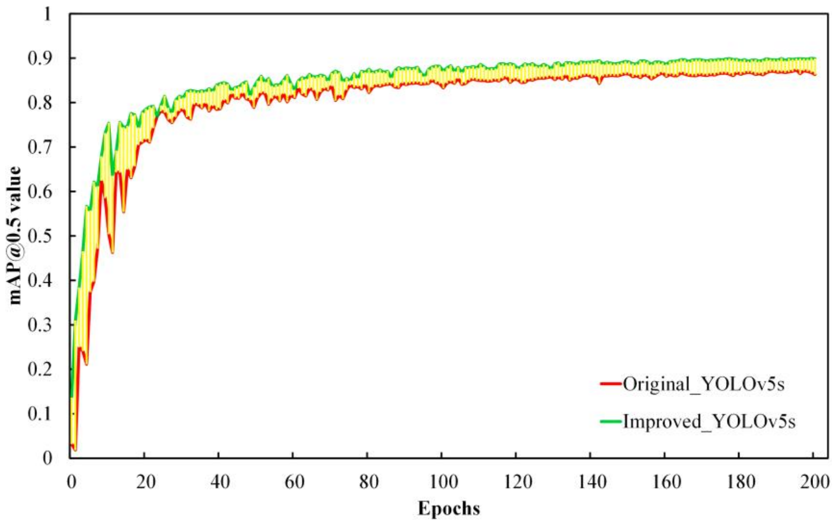

Figure 6. Comparison of YOLOv5 model accuracy

Specifically, the experimental results show that YOLOv5 performs well in various scenarios, not only maintaining high detection accuracy, but also achieving the expected goal in real-time processing speed. Especially in dynamic environments, the model exhibits good robustness and adaptability. However, there is still room for improvement in the detection performance of the system when facing extreme lighting changes or severe occlusion situations. This indicates that although YOLOv5 has made significant progress, continuous optimization of model structure and training strategies is still needed to better meet practical application needs.

5. Research on Obstacle Avoidance Strategies

Obstacle avoidance strategy is a major component of autonomous navigation for mobile robots, which emphasizes their ability to operate safely and efficiently in complex environments. It directly reflects the interaction status between the robot and the surrounding environment through real-time detection and path planning. With the deepening of research, some constituent elements of obstacle avoidance strategy development are gradually taking shape, including perception, decision-making, and execution, and are gradually being valued in conjunction with various evaluation systems. However, in practical applications, some obstacle avoidance algorithms are still in the theoretical verification stage, which contradicts the ideal logical framework and generation mechanism, leading to issues of adaptability and robustness.

Specifically, although the obstacle recognition system based on YOLOv5 performs well in static scenes, existing obstacle avoidance strategies may not be able to respond promptly to new challenges in dynamically changing environments, such as sudden obstacles or rapidly changing traffic conditions. In addition, current obstacle avoidance strategies rely heavily on preset rules and lack self-learning ability for unknown environments. Therefore, in order to address these issues, this study proposes a hybrid obstacle avoidance strategy that combines deep learning and reinforcement learning, aiming to improve the environmental adaptability and decision-making flexibility of mobile robots, ensuring their ability to achieve reliable autonomous navigation under various conditions.

6. Conclusion

This article proposes a novel obstacle avoidance strategy based on YOLOv5, which addresses the problem of real-time obstacle recognition and avoidance for mobile robots in complex environments. Firstly, reconstruct the YOLOv5 model and then utilize the deep learning information in the model; Further refine the model training and parameter optimization process, and optimize the obstacle avoidance strategy using reinforcement learning algorithms; Based on experimental phenomena, construct an environmental adaptability model and combine it with YOLOv5 algorithm to complete obstacle recognition and path planning. Theoretical analysis, simulation, and experimental results show that the proposed obstacle avoidance strategy can significantly improve the autonomous navigation performance of mobile robots. While ensuring high-precision detection, it achieves fast response and efficient obstacle avoidance, providing new ideas and technical support for the development of intelligent robot technology.

References

[1] Zhao J, Fang J, Wang S, et al. Obstacle avoidance of multi-sensor intelligent robot based on road sign detection[J]. Sensors, 2021, 21(20): 6777.

[2] Farag K K A, Shehata H H, El-Batsh H M. Mobile robot obstacle avoidance based on neural network with a standardization technique[J]. Journal of Robotics, 2021, 2021(1): 1129872.

[3] Moon S J, Kim J, Yim H, et al. Real-time obstacle avoidance using dual-type proximity sensor for safe human-robot interaction[J]. IEEE Robotics and Automation Letters, 2021, 6(4): 8021-8028.

[4] Li K, Chen J, Yu D, et al. Deep reinforcement learning-based obstacle avoidance for robot movement in warehouse environments[J]. arXiv preprint arXiv:2409.14972, 2024.

[5] Dang T V, Bui N T. Obstacle avoidance strategy for mobile robot based on monocular camera[J]. Electronics, 2023, 12(8): 1932.

[6] Chen J, Yin B, Wang C, et al. Bioinspired closed-loop CPG-based control of a robot fish for obstacle avoidance and direction tracking[J]. Journal of Bionic Engineering, 2021, 18: 171-183.

[7] Chen J, Yin B, Wang C, et al. Bioinspired closed-loop CPG-based control of a robot fish for obstacle avoidance and direction tracking[J]. Journal of Bionic Engineering, 2021, 18: 171-183.

[8] Smith T, Chen Y, Hewitt N, et al. Socially aware robot obstacle avoidance considering human intention and preferences[J]. International journal of social robotics, 2023, 15(4): 661-678.

[9] Xu C, Zhu H, Zhu H, et al. Improved RRT* Algorithm for Automatic Charging Robot Obstacle Avoidance Path Planning in Complex Environments[J]. Comput. Model. Eng. Sci, 2023, 12: 2567-2591.

[10] Gbouna Z V, Pang G, Yang G, et al. User-interactive robot skin with large-area scalability for safer and natural human-robot collaboration in future telehealthcare[J]. IEEE journal of biomedical and health informatics, 2021, 25(12): 4276-4288.

[11] Wang Y, Li X, Zhang J, et al. Review of wheeled mobile robot collision avoidance under unknown environment[J]. Science Progress, 2021, 104(3): 00368504211037771.

[12] Cong V D. Path following and avoiding obstacle for mobile robot under dynamic environments using reinforcement learning[J]. Journal of Robotics and Control (JRC), 2023, 4(2): 157-164.