ISSN: 2414-1895 DOI: 10.6919/ICJE.202403_10(3).0005

Research on Workpiece Absorption of Robotic Arm based on YOLOv8 Algorithm Fusing CBAM and Graph Convolution

Yinggang Li, Mingge Sun, Shilei Cui, Xingke Zhang, Xiaolong Guo

School of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin 132022, China

Abstract

In order to improve the success rate and speed of the robot when grasping the workpiece, an improved YOLOv8 real-time detection method integrating CBAM attention mechanism and Graph Convolutional Network (GCN) was proposed. This approach focuses on reconstructing the C2F module of YOLOv8 via CBAM and GCN. Firstly, the C2F module integrating CBAM to YOLOv8 improves the model's perception and ability to key features, especially after sequential application of spatial and channel attention, which effectively improves the focus on the target area and reduces background interference. In the experiments, the accuracy, recall and average accuracy of the model applied to this method were improved by 8%, 8.5% and 9.8%, respectively, compared with the original YOLOv8 network. Finally, the effectiveness of the algorithm in real-time grasping tasks is confirmed by the grasping detection platform built on the ROS simulation platform.

Keywords

Graph Convolution; CBAM Attention Mechanism; ROS.

1. Introduction

With the rapid development of industrial automation and intelligent manufacturing, robotics plays an increasingly important role in industrial production. Especially in the field of precision manufacturing and logistics, the workpiece picking and handling tasks of the robotic arm are essential to improve production efficiency and reduce labor costs. In this process, the accuracy and real-time performance of the machine vision system is crucial. Traditional machine vision systems are often difficult to meet the requirements of high precision and real-time performance in complex environments and diverse workpieces. Therefore, it is of great significance to study efficient and accurate workpiece detection and identification technology to promote the development of automated production system. In recent years, deep learning has made significant progress in the fields of image processing and computer vision, among which the YOLO series, as a representative of real-time object detection, is widely popular because of its high efficiency and accuracy[2]. Despite this, the existing YOLO model still faces challenges when dealing with complex industrial scenarios, such as insufficient accuracy when identifying multiple similar or small workpieces. In order to solve these problems, a YOLOv8 algorithm combining CBAM attention mechanism and Graph Convolutional Network (GCN) was proposed[2].

2. Rationale

2.1 YOLOv8 Original Network

YOLOv8 is an advanced real-time object detection model, which continues the efficiency and accuracy of the YOLO series, providing five size models. The model is divided into three parts:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202403_10(3).0005

Backbone, Neck, and Head, which are responsible for feature extraction, feature fusion, and prediction output, respectively. Among them, the C2f module simplifies the network structure and improves the lightweight, while the SPPF module accelerates the speed of feature fusion. The multi-scale fusion module combines FPN and PAN to enhance the detection capability of targets at different scales by bidirectional fusion of features at different levels[2] . YOLOv8's Head uses a no-aim method, using a task-aligned allocator for sample matching and multi-scale prediction to accurately detect targets of various sizes, as shown in Figure 1.

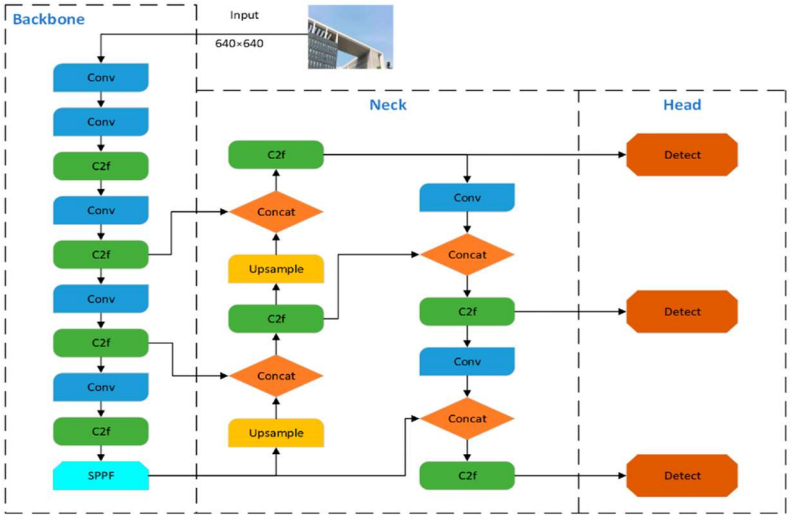

Figure 1. Structure of the YOLOv8s model

2.2 GCN-CBAM-v8s

Since the original YOLOv8 model was primarily designed for general object detection, it had some limitations in specialized part inspection applications. First of all, the original model has insufficient accuracy in detecting parts in complex or messy backgrounds, especially when the parts are similar in shape, small in size, or close to each other, and its recognition ability is limited. Second, the original YOLOv8 may not be precise enough when dealing with multi-target detection tasks, especially in capturing subtle relationships and interactions between artifacts. In addition, while YOLOv8 excels in speed, there is room for improvement in real-time performance in specific industrial environments, especially in automated production lines that require high speeds. Based on these shortcomings, this paper improves the YOLOv8 network model and proposes the GCN-CBAM-v8s model, which consists of two innovative modules:

1) Integrated CBAM: The application of CBAM module in the model enhances the ability to pay attention to and identify important features. It improves the accuracy of detection in complex scenes by sequentially applying spatial and channel attention to improve focus on the target area while effectively reducing background interference.

2) Fusion Graph Convolutional Network (GCN): The introduction of GCN allows models to better understand and process the relationships and structures between targets, especially when dealing with non-Euclidean data. This enhances the model's ability to handle multi-target detection in complex scenarios.

2.2.1 CBAM Module

CBAM is an efficient attention mechanism module designed to enhance the feature expression ability of convolutional neural networks. CBAM meticulously regulates the attention focus of the network through two main dimensions-channel and spatial. In the channel attention link, CBAM uses global average pooling and global maximum pooling to extract global features at the channel level. These features are further processed to generate channel weights, emphasizing important features and suppressing less correlated features. In the spatial attention link, a similar pooling operation is

ISSN: 2414-1895 DOI: 10.6919/ICJE.202403_10(3).0005

performed in the channel dimension to generate a spatial weight map, which helps the model to focus more accurately on key spatial regions[2] . CBAM first reconstructs the features at the channel level, and then refines the spatial features. This structure is simple and efficient, easy to integrate into different CNN architectures, and demonstrates strong performance when dealing with visual tasks that require precise channel and spatial information, as shown in Figure 2.

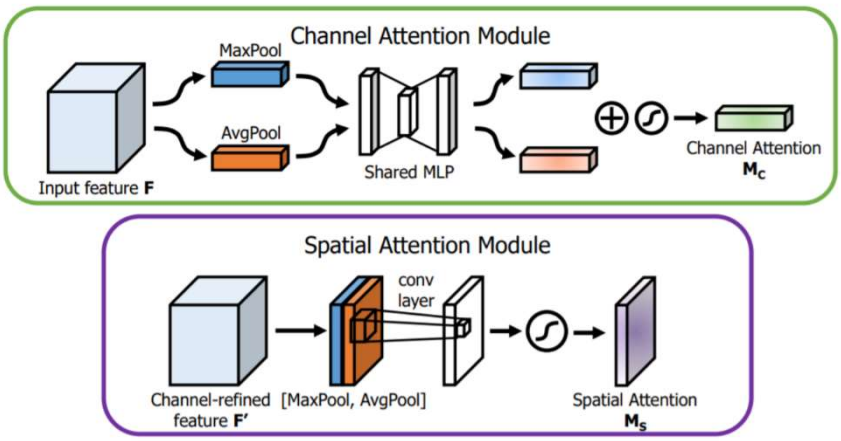

Figure 2. CBAM structure diagram

2.2.2 Graph Convolution

Graph Convolutional Network (GCN) is an innovative neural network architecture designed to process graph data, which is different from traditional convolution operations in that graph convolution specializes in processing graph structure data, which is composed of nodes and their relationships, with a high degree of freedom and flexibility. The core advantage of GCN is that it can capture the complex topological relationships between nodes, so as to efficiently extract features and transfer information. In the graph convolution process, nodes aggregate and update the information of their neighbor nodes to achieve spatial locality of features. Through multi-layer graph convolution, GCN can extract node features that represent a wider range of neighborhood information, which is crucial for node classification, link prediction, or classification tasks of the whole graph. Graph convolutional networks are widely used in a variety of fields, providing deep analysis and flexibility that cannot be matched by traditional methods by deeply learning the interactions between entities[2] , as shown in Figure 3.

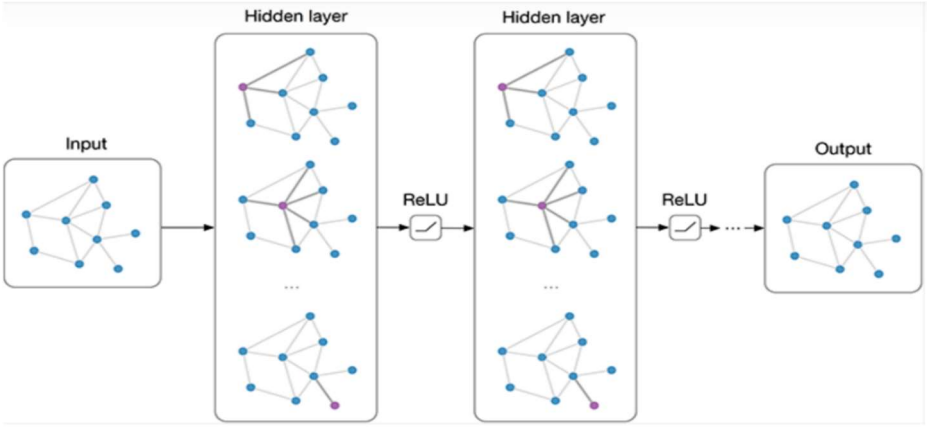

Figure 3. Graph convolutional network structure diagram

2.2.3 C2f_GCN_CBAM Modules

The above-mentioned CBAM attention mechanism module and Graph Convolutional Network (GCN) have shown remarkable results in enhancing the performance of the model. However, in practical

ISSN: 2414-1895 DOI: 10.6919/ICJE.202403_10(3).0005

applications such as workpiece inspection or image recognition, existing models still face challenges in handling complex scenarios and maintaining computational efficiency. To address these challenges, this study proposes an innovative strategy that integrates CBAM attention mechanism and GCN graph convolution. This strategy aims to combine the fine-grained feature attention capabilities of CBAM and the structured data processing capabilities of GCN to more efficiently process complex scenes and relationships in images, as shown in Figure 4.

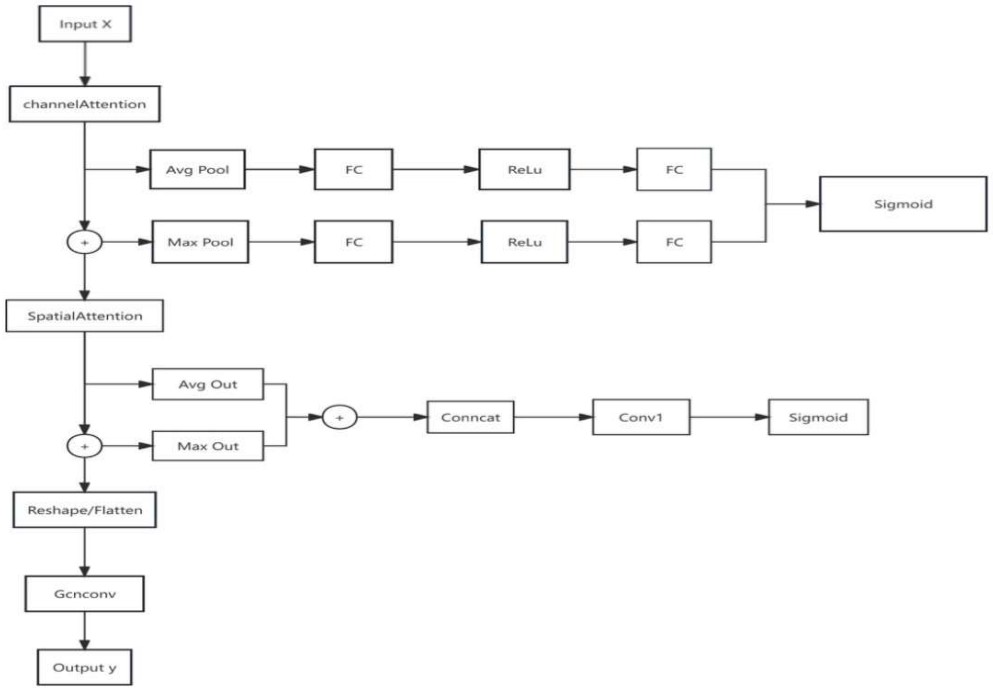

Figure 4. Convolutional fusion structure diagram of CBAM attention mechanism and GCN graph

Furthermore, this fusion attention mechanism was introduced into the C2f module to create an improved C2f module to improve the recognition accuracy and processing speed of the model. The improved C2f module integrates the CBAM attention mechanism and GCN. The CBAM part realizes the recalibration of features through channel and spatial attention mechanisms. The channel attention sub-module uses adaptive average pooling and adaptive maximum pooling, and generates weight maps through multilayer perceptron (MLP) processing to realize the dimensionality reduction, dimensionality enhancement and activation of features. The spatial attention sub-module generates a spatial weight map by averaging and maximizing the channel dimensions of the feature map and the use of small convolutional layers, focusing on key spatial regions, as shown in Figure 5.

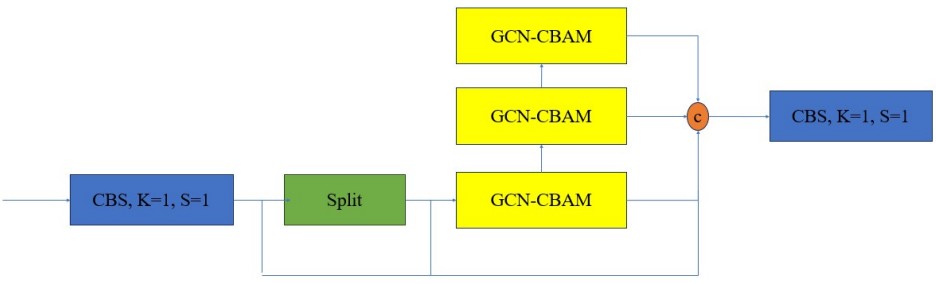

Figure 5. Improved C2f mechanism diagram

ISSN: 2414-1895 DOI: 10.6919/ICJE.202403_10(3).0005

Then, the GCN layer processes the CBAM-adjusted feature map and updates the node features through the graph convolution operation, which involves the use of adjacency matrices and the weighted aggregation of node features. The GCN layer uses the topology of the graph to propagate and integrate the information of adjacent nodes to improve the feature expression globally. In this improved strategy, the integration of CBAM and GCN enables the model to process complex image features more efficiently, especially in multi-object detection and complex background segmentation scenarios. CBAM first refines the feature map to provide rich and discriminating features for the GCN graph convolution operation, and the GCN layer further enhances the structured expression of these features, so that the model can capture more complex intra-image and inter-image relationships.

3. Experiments and Evaluations

3.1 Experimental Environment

The experimental environment of this paper is: NVIDIA GeForce RTX3060, video memory is 12GB, Ubuntu 18.04 operating system, the programming language used is python3.8, and CUDA is 11.3. At the same time, a simulation model of the robotic arm and the workpiece was built in Gazebo.

3.2 Datasets

The 3D models of the workpiece, carefully designed in SolidWorks, form the basis of the dataset, which is then imported into the Gazebo simulation platform for image acquisition. In the Gazebo environment, a large number of images of workpieces under different environmental conditions and lighting were collected using a virtual camera system. These images are pre-processed to fit the requirements of the deep learning model. The ability to simulate various challenges in the actual working environment provides a solid foundation for the training and validation of algorithms, as shown in Figure 6.

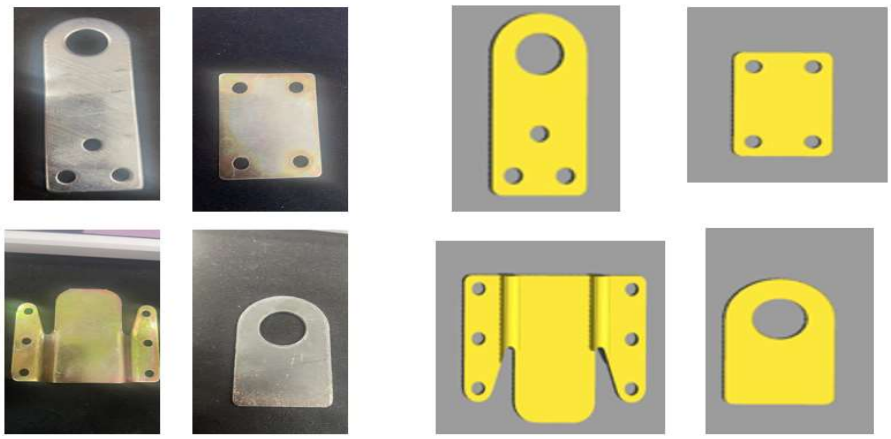

Figure 6. Physical (left) and simulated (right) models

3.3 Training Results and Analysis

In the training process of this study, the stochastic gradient descent method was selected as the optimization strategy. The batch size of the model is set to 4 and the initial learning rate is set to 0.01. In addition, the momentum parameter is 0.937 and the weight decay rate is set to 0.0005. The entire training process consists of 1000 epochs and processes an input image with a resolution of 640x640 pixels. The training took about 2.45 hours. After the training was terminated, the resulting model weight file was saved, and various metrics of the model on the validation set were measured, including precision, recall, average precision at IOU threshold of 0.5, and average precision at IOU threshold from 0.5 to 0.95 (step size of 0.05), as shown in Figure 7.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202403_10(3).0005

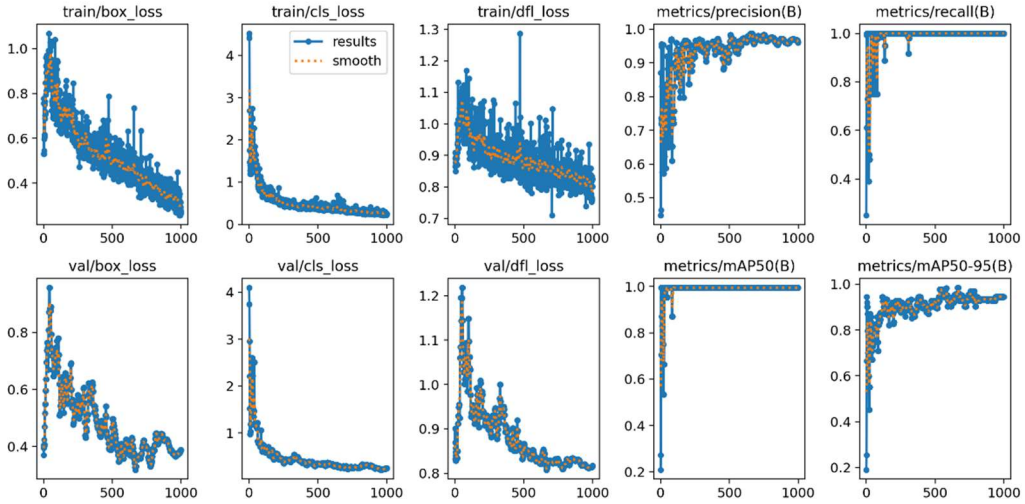

Figure 7. Training results for GCN-CBAM-v8s

It can be seen from the training results that the loss value of the target box continues to decrease, indicating that the positioning accuracy is gradually improved, and finally stabilizes at about 0.4, which shows the robustness of the model in target positioning. At the same time, after the first 300 rounds of significant improvement, the accuracy remains at a high level of 93%, indicating that the recognition and classification capabilities of the model are extremely reliable. The recall rate was also stable and maintained at 100% after 300 rounds of training, demonstrating the model's ability to accurately detect all relevant targets. Finally, the average accuracy of the model continued to increase and finally stabilized at 93%, which reflects the high detection performance of the model under different IOU thresholds.

3.4 Comparative Experiments

As can be seen from the figure below, the improved GCN-CBAM-v8s achieves a significant performance improvement in accuracy compared to the original YOLOv8s model. This result verifies the effectiveness of Fusion Graph Convolutional Network (GCN) and Convolutional Block Attention Module (CBAM) in enhancing the model's feature recognition and target localization capabilities. In the first 300 rounds of training, the accuracy of the improved model increased rapidly and eventually stabilized, with an average accuracy of 95%, compared to the original YOLOv8s model, which was only 87% and fluctuated greatly throughout the training process. This comparison clearly demonstrates the superior performance of GCN-CBAM-v8s in terms of model accuracy and stability, as shown in Figure 8.

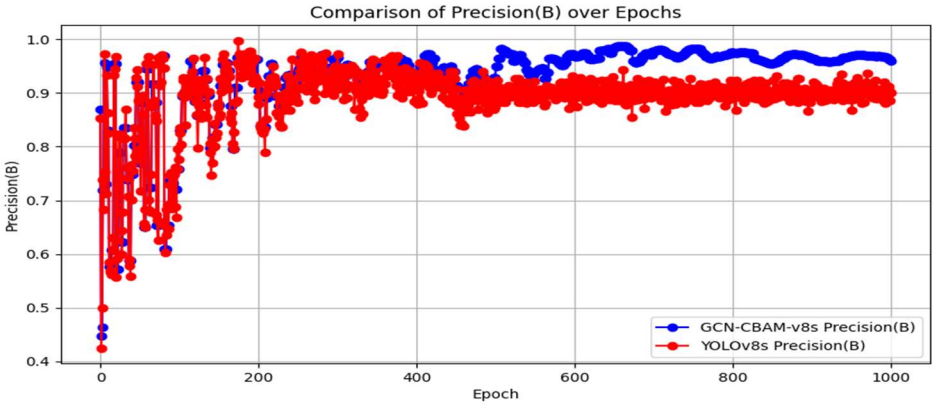

Figure 8. Accuracy plot

ISSN: 2414-1895 DOI: 10.6919/ICJE.202403_10(3).0005

GCN-CBAM-v8s improved recall by 9%. This result not only reflects the effectiveness of the improvement measures in enhancing the model's ability to detect positive targets, but also highlights the significant improvement of the model's ability to identify missing objects in complex scenes, as shown in Figure 9.

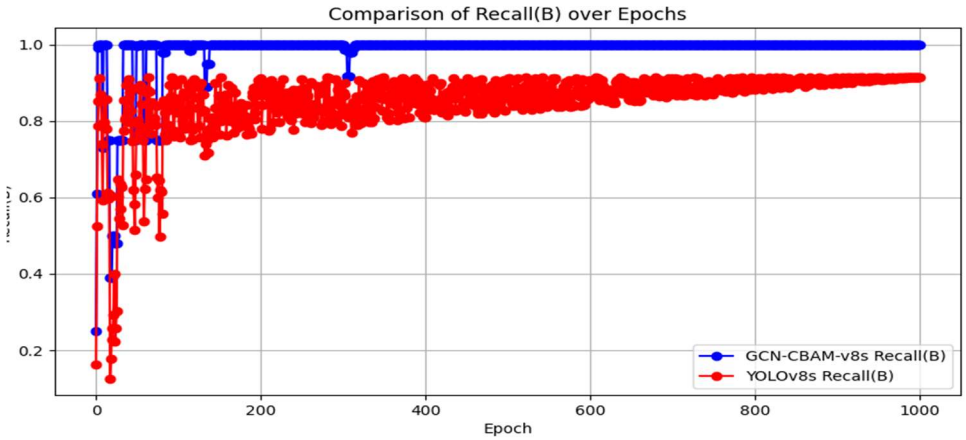

Figure 9. Recall

3.5 Grasping Experiments

In the simulation experiment of autonomous grasping of the manipulator, the steps are as follows: firstly, the simulation environment is established by Gazebo software; Secondly, the Moveit software was used to control the operation of the manipulator. Next, the workpiece is randomly placed on the table; Then, the scene was captured with an RGBD camera and the images were imported into the GCN-CBAM-v8s model in order to detect the target and obtain the pixel coordinates of its center point. Then, combined with the depth information provided by the RGBD camera and the internal parameter matrix of the camera, the three-dimensional position (X, Y, Z) of the target center point in the camera coordinate system was calculated. The motion trajectory of each joint is calculated through the inverse kinematics of the robotic arm. Finally, if the gripping is successful, transfer the workpiece to the bin, otherwise reset and try the gripping again. The robotic arm grasping process is shown in Figure 10 below, and the detection result is Figure 11 below.

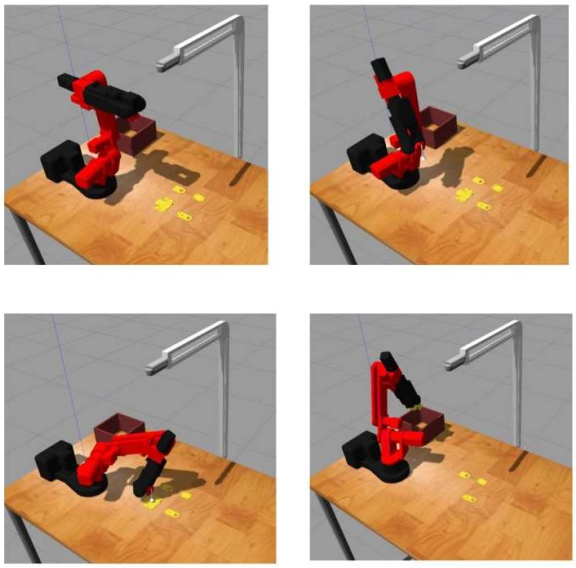

Figure 10. Simulating the gripping process

ISSN: 2414-1895 DOI: 10.6919/ICJE.202403_10(3).0005

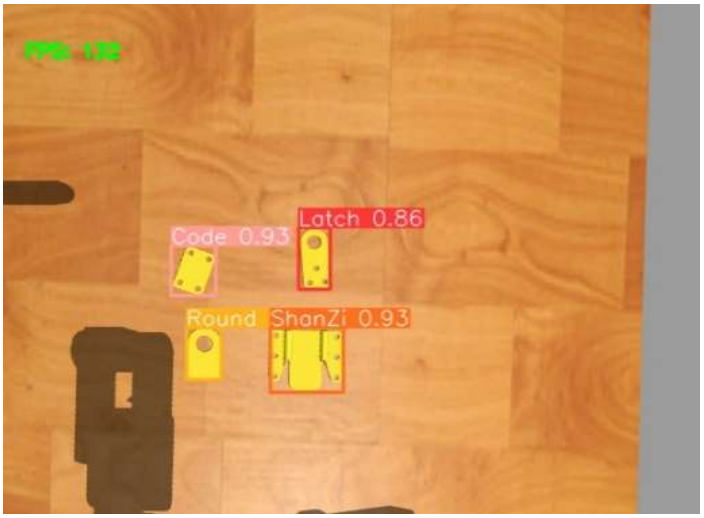

Figure 11. GCN-CBAM-v8s test results

3.6 Experimental Results

In this experiment, 50 rounds of testing were carried out, and 1-4 workpieces were randomly generated in each round, and the number of times the workpieces were successfully picked up and placed and the accuracy of visual recognition were counted. The results are shown in Table 1 below.

Table 1. Simulation results

Algorithm used	Number of experimental rounds	The number of pieces of the generated workpiece	The number of pieces successfully absorbed	The number of pieces that were correctly inspected	Suction success rate	Detection accuracy
YOLOv8s	50	99	89	93	89.9%	94%
GCN-CBAM-v8s	50	99	95	98	96.0%	99.0%

The data analysis shows that the improved GCN-CBAM-v8s model shows excellent performance in the workpiece inspection task, and achieves an accuracy rate of up to 99%, which is significantly better than the 89.9% detection success rate of the original Yolov8 model. Failures in detection are due to loss of feature information due to light occlusion. At the same time, when the workpiece is accurately identified, the grasping success rate of the robotic arm is increased to 96.0%, which is significantly higher than the 89.8% of the original algorithm. This result effectively verifies the significant advantages of the improved YOLOv8s algorithm in terms of object recognition speed and grasping efficiency.

4. Conclusion

This paper shows the optimization effect of the autonomous gripping system of the manipulator by using the GCN-CBAM-v8s model algorithm. Experimental results show that these improvements significantly improve the accuracy of workpiece detection and the grasping success rate of the robotic arm, especially in the application performance of complex environments. The improvement of accuracy and grasping success rate proves the effectiveness of these methods in improving the efficiency and reliability of robotic arm operation. But there is still some room for improvement. For example, for detection failures under special light conditions, future work could explore more robust

ISSN: 2414-1895 DOI: 10.6919/ICJE.202403_10(3).0005

feature extraction and recognition techniques. In addition, the adaptability of algorithms under extreme or non-standard conditions is also a potential direction for future research.

References

[1] Du Pengcheng, Jiang Du zhong, et al. Fresh Leaf Maturity Recognition Method of Flue-cured Tobacco Based on YOLOv5s Object Detection Algorithm[J]. Jiangsu Agricultural Sciences, 2023,51(19):158-165.

[2] LIU Zhongyi, WEI Dengfeng, LI Meng, et al. Orange fruit recognition method based on improved YOLOv5[J]. Jiangsu Agricultural Sciences, 2023,51(19):173-181.

[3] LIU Xiaoyang, ZHAO De'an, JIA Weikuan, et al. Fruit segmentation method of apple picking robot based on superpixel features[J].Transactions of the CSAM, 2019,50(3):15-23.

[4] Guo Yiyu, Zhou Luoyu, Liu Xinyu, et al. Dangerous goods detection method in elevator scene with improved attention mechanism[J].computer application,2023,43(3):1-10.

[5] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J].IEEE Trans. Pattrern Anal, 2018, 15: 2236-1131.