An improved YOLOv5-based method for robotic vision detection of grain caking in silos

Yi Cao[1], Yao Zhao[2], Xiang Wu[3], Mingqi Tang[4], Chao Gu[5]
[1, 2, 4, 5]School of Mechanical and Electrical Engineering, Henan University of Technology,
No. 100 Lianhua Street, Zhengzhou City, Henan Province, China
[3]School of Electrical Engineering, Henan University of Technology,
No. 100 Lianhua Street, Zhengzhou City, Henan Province, China
[1]Corresponding author
E-mail: [1]*caoyioffice@163.com*, [2]*shineyao666@gmail.com*, [3]*xiangw@haut.edu.cn*, [4]*783768996@qq.com*, [5]*GC18236613121@163.com*

Received 25 October 2024; accepted 27 January 2025; published online 16 February 2025
DOI https://doi.org/10.21595/jme.2025.24634

Abstract. Detecting and cleaning grain caking on the inner walls of silos is an important task to ensure food safety in storage facilities. However, in response to challenges such as insufficient lighting conditions, small and diverse forms of grain caking, this paper proposes the development and evaluation of a convolutional neural network model for robot vision detection of grain caking. The following improvements to the visual detection algorithm based on YOLOv5 are proposed in this article. Firstly, the Convolutional Block Attention Module (CBAM) and the improved Total Cross Union (CIoU) loss function are introduced to enhance the detection accuracy of grain caking. Secondly, by adding the Retinex Net algorithm with dark light enhancement, the recognition and detection performance under low light conditions can be improved. The improved YOLOv5 algorithm was trained and validated on a custom grain caking dataset. Comparative experiments show that compared with existing detection architectures, the improved algorithm has improved the average accuracy of grain caking detection by 1.8 % to 3.8 %. Finally, the improved algorithm proposed in this article was deployed on a wall climbing robot based on negative pressure adsorption, achieving real-time detection and automatic cleaning of grain caking.

Keywords: machine vision and deep learning, improved YOLOv5 algorithm, grain caking detection, wall-climbing robot.

1. Introduction

Grain silos are critical storage facilities for preserving grain, but during storage, moisture can cause grain caking on silo walls [1]. This compaction leads to mold growth, posing a significant threat to grain safety. Due to the structural limitations of vertical silos, cleaning these compacted areas is extremely challenging [2]. Currently, manual cleaning inside silos is the most commonly used method, but it is both inefficient and highly dangerous due to the confined space and high-altitude operations required [3]. Accidents involving workers entering silos for cleaning are unfortunately common, often resulting in severe loss of life and property. Therefore, there is an urgent need in the grain storage industry for a robot capable of identifying and removing compacted grain [4]. Wall-climbing robots [5], [6], a key innovation in robotics, have the ability to adhere to vertical structures and move flexibly [7], [8]. They have been widely used for tasks such as cleaning high-rise building surfaces, automating maintenance in petrochemical storage tanks, and inspecting boiler water wall tubes [9], [10]. The focus of this study is to design a vision-based detection system for identifying grain caking in silos and integrate it with a wall-climbing robot, enabling the robot to autonomously perform the full process of compaction identification and cleaning.

There has been extensive research on deep learning-based object detection [11]. To address issues such as the large number of parameters, long training times, and poor real-time detection performance of two-stage models, Redmon et al. [12] introduced YOLO, the first regression-based object detection algorithm. Jocher proposed YOLOv5, which includes four models known for their real-time detection capabilities and lightweight characteristics, making them ideal candidates for mobile deployment environments [13]. Various advanced convolutional neural networks, such as VGG [14] and ResNet [15], have been used in the Backbone, but their real-time performance does not meet the requirements of industrial applications. ShuffleNet [16] further reduces the number of model parameters through grouped convolutions and channel shuffling. Mi et al. [17] proposed a lightweight target detection method based on improved YOLOv5s, which can maintain a high accuracy under limited resources. Liu et al. [18] introduced an algorithm based on an improved feature fusion mode, enhancing the precision of small object detection while making the model more lightweight. Additionally, for object detection in low-light environments, Liang Cheng [19] proposed a garbage detection algorithm suitable for low-computing-power devices and low-light conditions, which improves detection under poor lighting without increasing the model's size. Liu Hang [20] developed a feature enhancement-based dark-light object detection method that extracts multiple features from input images to reduce the impact of low-light conditions on object detection. However, these methods are not suitable for identifying and detecting the diverse forms of grain caking in low-light environments within silos.

To address the challenges of grain caking detection and cleaning in silos, and building on previous research, the development of a convolutional neural network model for robot vision detection of grain caking is proposed in this paper. Firstly, the Convolutional Block Attention Module (CBAM) and the improved Total Cross Union (CIoU) loss function are introduced to enhance the detection accuracy of grain caking. Secondly, by adding the Retinex Net algorithm with dark light enhancement, the recognition and detection performance under low light conditions can be improved. The improved YOLOv5 algorithm was trained and validated on a self-made grain caking dataset. Experiments show that compared with existing detection architectures, the proposed algorithm achieves higher location average precision and real-time capabilities. Finally, the proposed algorithm was deployed on a wall climbing robot based on negative pressure adsorption, achieving real-time detection and automatic cleaning of grain caking.

2. Materials and methods

2.1. Design of the wall-climbing robot based on negative pressure adhesion

Negative pressure wall-climbing robots typically use either rectangular or circular negative pressure chambers to achieve suction adhesion. Considering that the wall-climbing robot in this study uses a track-based walking mechanism, a rectangular negative pressure chamber was designed. The structure is shown in Fig. 1.

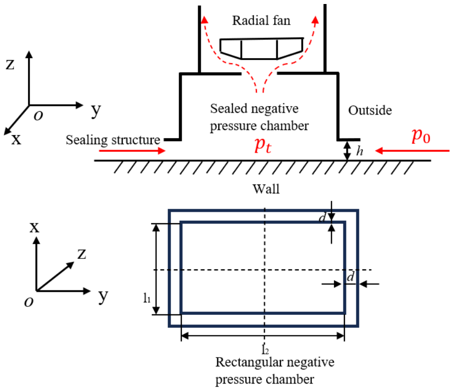

Radial fan
z
o
y
x
Sealing structure
Sealed negative pressure chamber
Outside
p_t
h
p_0
Wall
x
z
o
y
d
d
l₁
l₂
Rectangular negative pressure chamber

Fig. 1. Schematic of the adhesive mechanism for the wall-climbing robot

The minimum adhesion force condition for a wall-climbing robot in any orientation is:

$$P_1 > \max\left(\frac{G\cos\theta + ma + F_{r1} + F_{r2} + F_{np}}{\mu}, \frac{G\sin\theta}{\mu_2}, \frac{2HG\cos\theta}{l} + \frac{2HG\sin\theta}{b}\right), \tag{1}$$

where F_{r1} and F_{r2} represent the turning resistance of the left and right tracks of the wall-climbing robot, respectively, while F_{np} denotes the force exerted by the wall on the sealing mechanism. By substituting the parameters of the wall-climbing robot designed in this study into Eq. (1) and conducting simulations, the safe negative pressure operating range for the robot was obtained, as shown in Fig. 2.

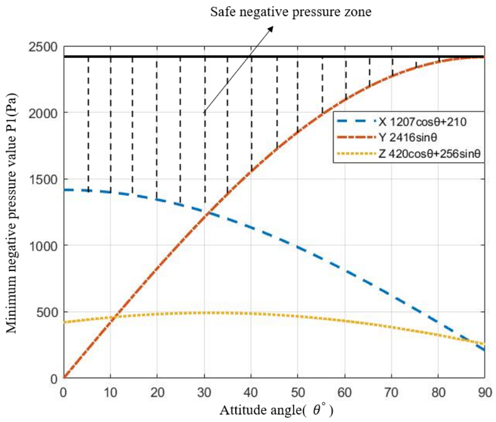


Fig. 2. Safe negative pressure range for the wall-climbing robot

The tracked wall-climbing robot designed in this study, which is based on negative-pressure adhesion, comprises three core modules: the negative-pressure adhesion mechanism, the tracked walking mechanism, and the modular actuator. The actuator can be replaced with a vision detection system or a cleaning mechanism to meet different operational requirements. Fig. 3 shows the 3D model and the physical prototype of the wall-climbing robot designed in this study.

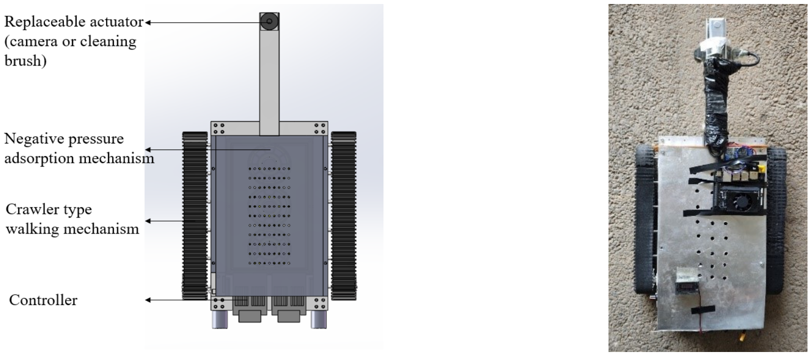

Replaceable actuator (camera or cleaning brush)
Negative pressure adsorption mechanism
Crawler type walking mechanism
Controller

Fig. 3. 3D model and physical prototype of the wall-climbing robot

2.2. Identification and detection of grain caking based on an improved YOLOv5

YOLOv5 features a compact model structure and fast inference speed, making it highly suitable for deployment on mobile devices for identification and detection tasks. However, Due to the complex environment within the silo, insufficient lighting, and the presence of multiple small target objects, the traditional YOLOv5 algorithm does not adequately meet the requirements. To address this issue, the following improvements are made to the algorithm: 1. The introduction of the CBAM (Convolutional Block Attention Module) and improved CIoU (Complete Intersection over Union) function to enhance the algorithm's detection accuracy for grain caking. 2. To tackle the problem of insufficient ambient lighting, the Retinex-Net model is added for image enhancement to meet detection needs.

2.2.1. Improved model

The Convolutional Block Attention Module (CBAM) is a lightweight feed-forward convolutional neural network model [21], which consists of two main components: the channel attention module and the spatial attention module, as shown in Fig. 4. The channel attention module focuses on meaningful information in the input image and can compress the spatial dimensions while keeping the channel dimensions unchanged. The spatial attention module, on the other hand, focuses on the location information of the target and can compress the channel dimensions while keeping the spatial dimensions unchanged. When a specific intermediate feature map $F \in \mathbb{R}^{C\times H\times W}$ is input, CBAM derives a channel attention map $M_c \in \mathbb{R}^{C\times 1\times 1}$ and a spatial attention map $M_s \in \mathbb{R}^{1\times H\times W}$ through the channel attention module and the spatial attention module, respectively. By multiplying these attention maps with the original input feature map, a refined output F'' is obtained, which achieves adaptive feature refinement of the input feature map and effectively enhances the model's classification performance. The entire process can be summarized as follows:

$$F' = M_c(F) \otimes F, \\ F'' = M_s(F') \otimes F'. \tag{2}$$

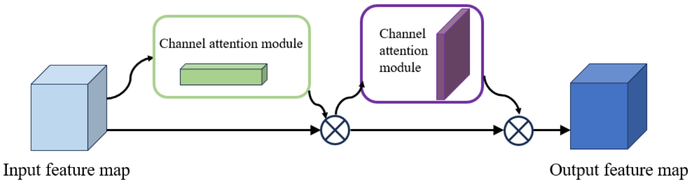


Fig. 4. CBAM network model diagram

The CBAM attention mechanism is used to enhance the backbone network. In the updated YOLOv5s model, the Bottleneck CSP module has been replaced by the C3 module. Therefore, this study integrates the CBAM attention mechanism with the C3 module in the Backbone, resulting in the improved CBAMC3 module.

A network loss function for data fitting is introduced in this study to further optimize the prediction of grain caking. This function combines three key components: the object confidence loss, classification loss, and bounding box (BBOX) regression loss [22]:

$$Loss = a * loss_{obj} + b * loss_{rect} + c * loss_{clc}. \tag{3}$$

For the bounding box regression loss, this model uses CIoU (Complete Intersection over Union) as the evaluation metric. The goal is to incorporate key geometric factors – such as the overlap area, center point distance, and aspect ratio – between the ground truth box and the predicted box into the bounding box regression loss calculation. This design improves the stability and convergence accuracy of the regression bounding box. The loss function is expressed as:

$$L_{CIoU} = 1 - L_{IoU} + \frac{\rho(b, b^{gt})}{c^2} + \alpha v, \tag{4}$$

where, αv represents a penalty factor that fits the aspect ratio between the ground truth box and the predicted box. α is the coordination parameter, and v is a parameter used to measure the consistency of the aspect ratio. The formulas for α and v are as follows:

$$\alpha = \frac{v}{(1 - L_{IoU}) + v}, \tag{5}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2. \tag{6}$$

In the construction of this model, both the confidence loss and classification loss are computed using the Binary Cross-Entropy (BCE) loss function. For an image divided into an 80×80 grid, the neural network extracts three predicted boxes for each grid cell. Each predicted box contains information such as the center coordinates, width and height dimensions, confidence score, and classification probabilities. Therefore, the neural network outputs a total of 3×80×803 predicted confidence values ranging from 0 to 1. As shown in Fig. 5, when targets such as the red points A, B, C, and D are detected, the predicted boxes for these grid cells typically have higher confidence values.

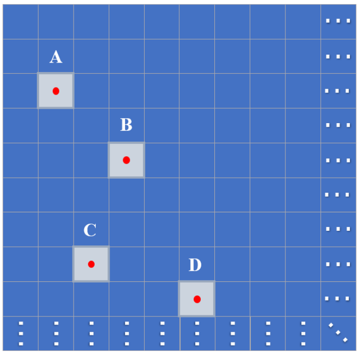


Fig. 5. Confidence prediction 80×80 grid

Assuming the confidence label is represented by matrix L and the predicted confidence by matrix P, the value at each corresponding position in the matrices is involved in the calculation of the BCE loss. The calculation formula is:

$$\begin{gathered} loss_{BCE}(z,x,y) = -L(z,x,y) * \log P(z,x,y) - \big(1 - L(z,x,y)\big) * \log\big(1 - P(z,x,y)\big), \\ (0 \le z < 30,\ \ 0 \le x < 800,\ \ \ 0 \le y < 80). \end{gathered} \tag{7}$$

Most existing object detectors generate a large number of candidate boxes and then use Non-Maximum Suppression (NMS) to filter them. NMS generally sorts the boxes based on classification scores. This traditional multiplication method is not optimal and offers limited performance improvement. In this study, the IACS (Instance-Aware Classification Score) evaluation criterion is considered to represent the class and quality of the bounding box. IACS is a scalar element of the classification score vector, where the scores for class, position, and centerness are replaced with ground truth values. The final score for the true class location is the Intersection over Union (IoU) between the predicted and ground truth boxes, while scores for other locations are set to 0.

2.2.2. Low-light enhanced object detection

Due to insufficient lighting inside the silo, the low-quality, low-light images collected are challenging for visual detection tasks. Therefore, this study incorporates the Retinex-Net image enhancement algorithm [23] to improve the model.

The Retinex-Net model is a deep learning-based low-light image enhancement algorithm. The network architecture consists of three core components: Decom-Net, Adjustment-Net, and Reconstruction-Net. The Decom-Net is responsible for decomposing the input low-light image into reflection and illumination components, providing a basis for subsequent processing. The Adjustment-Net adjusts the illumination component by enhancing the image's brightness, contrast, and other attributes to improve the overall visual quality. The Reconstruction-Net then combines the adjusted illumination component with the reflection component to generate the enhanced image. The network structure is illustrated in Fig. 6.

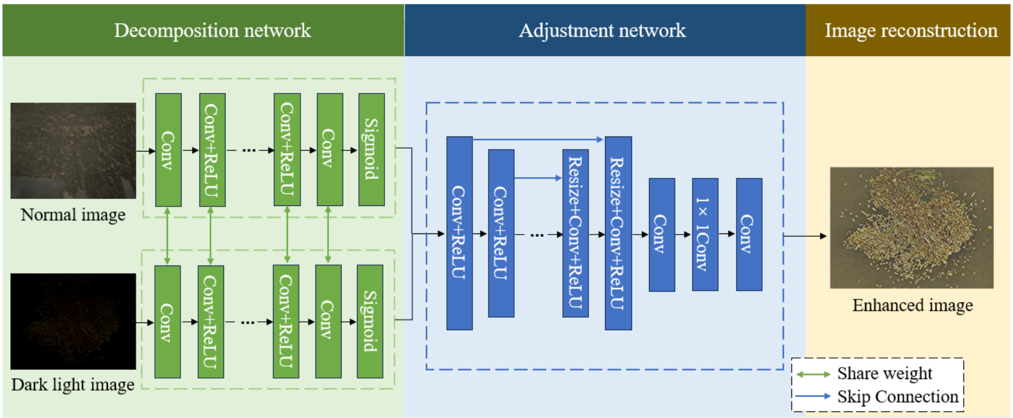
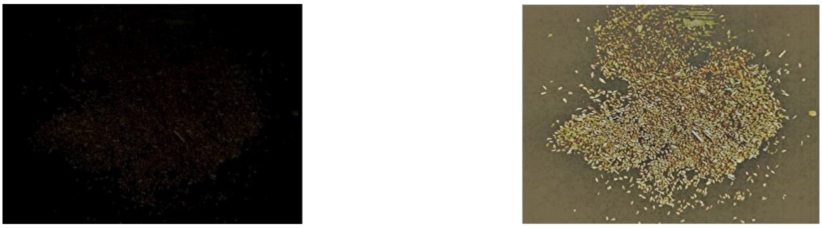


Fig. 6. Retinex-Net model network structure

The entire process is as follows: First, input images of grain caking under normal lighting (S_{nomal}) and under low-light conditions (S_{low}) are processed by the Decom-Net to obtain two sets of illumination and reflection component images. Next, in the Adjustment-Net, the four images are denoised and enhanced. Since the reflection images derived from both normal and low-light conditions are quite similar, indicating that the reflection images are less affected by lighting, only noise suppression is applied to the reflection images, while the illumination images are enhanced. Finally, the processed illumination and reflection components are combined using the Reconstruction-Net to produce the enhanced low-light image of grain caking. The output results are shown in Fig. 7.



a) Before processing b) After processing

Fig. 7. Comparison of grain caking before and after low-light enhancement

3. Experiments and analysis

The experimental process is divided into two main parts: evaluating the grain caking recognition algorithm and testing the wall-climbing and cleaning performance of the robot. The optimized algorithm is deployed on the NVIDIA Jetson Orin Nano mounted on the robot, integrating the recognition system with the robot system. The front of the robot adopts a modular design, enabling real-time identification of grain caking by installing a vision module. The collected data is sent to the host computer for use by the wall-climbing robot, which is equipped with a cleaning module. The overall experimental workflow is illustrated in Fig. 8.

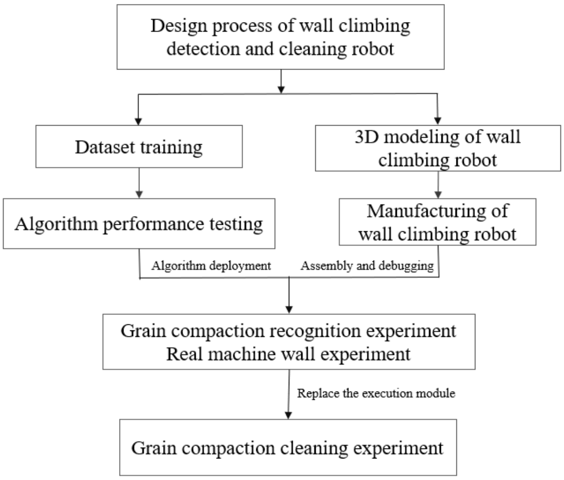

Design process of wall climbing detection and cleaning robot
Dataset training
3D modeling of wall climbing robot
Algorithm performance testing
Manufacturing of wall climbing robot
Algorithm deployment
Assembly and debugging
Grain compaction recognition experiment
Real machine wall experiment
Replace the execution module
Grain compaction cleaning experiment

Fig. 8. Experimental flowchart

3.1. Grain caking dataset creation and training

Since there is no publicly available grain caking dataset, this experiment uses a custom dataset for model training. The images are sourced from field collections of the sidewalls of shallow cylindrical silos and simulated grain caking on cement walls. To match application scenarios. Data was collected under various lighting conditions and simulated with different degrees of compaction, resulting in a total of 594 raw images. Due to the limited number of images, Mosaic data augmentation [24] was applied to expand the dataset. As shown in Fig. 9, geometric and pixel transformations increased the dataset to 3,573 images. Finally, the images were normalized to a size of 640×640×3 for subsequent annotation.

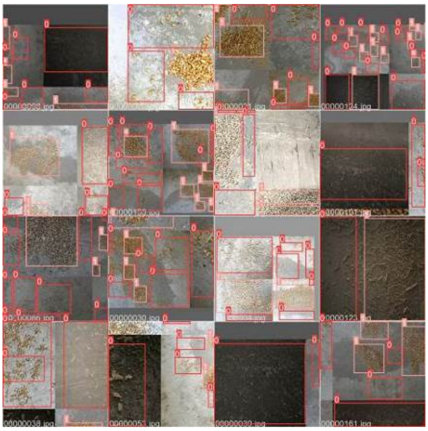
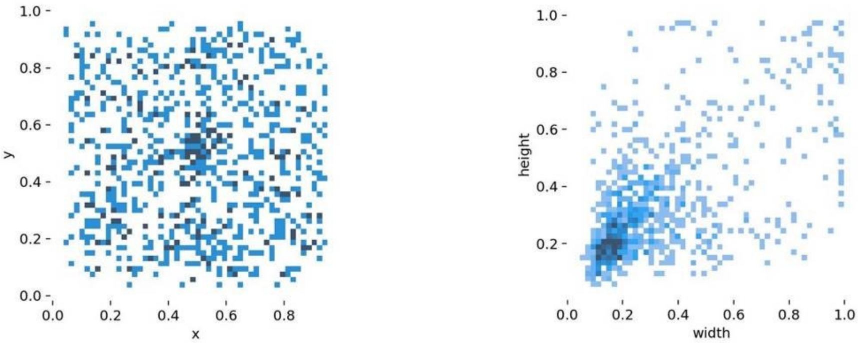


Fig. 9. Mosaic data augmentation effect

The custom dataset was annotated using the Labelme tool, including two categories: Heavily and Lightly, representing severe and mild grain caking, respectively. The dataset contains 982 labels for the Heavily category and 1,537 labels for the Lightly category. Summary analysis of the labels is shown in Fig. 10. Fig. 10(a) displays the distribution of the center points of all labels in the training set, with the x and y coordinates representing the actual positions of the center points in the image, measured in pixels. This Figure shows that the center points are spread across all areas of the image and are distributed relatively evenly. This indicates that the training set is flexible and comprehensive, capable of detecting objects in various positions within the image.

Fig. 10(b) shows the distribution of label sizes, where the x-axis represents the relative width of the bounding boxes and the y-axis represents the relative height of the bounding boxes, measured in pixels. This Figure reveals a dense distribution of points in the lower-left corner and a sparser distribution in the upper regions. This indicates that the custom grain caking dataset features numerous small targets with diverse size characteristics, which aligns with the nature of grain caking in cylindrical silos.



a) Label center point position distribution diagram b) Label size distribution diagram

Fig. 10. Summary analysis of the labels

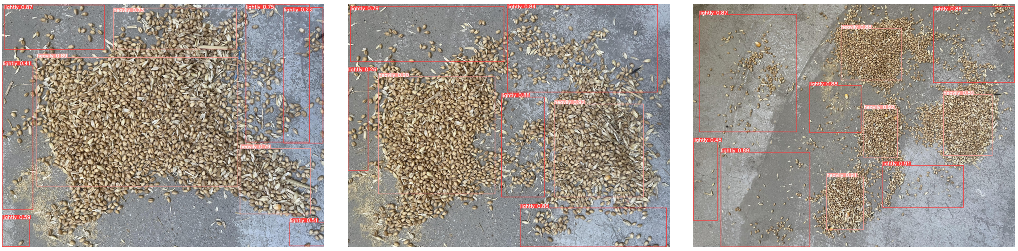
The experimental system runs on Ubuntu 20.04, with an NVIDIA GeForce GTX 1650 GPU featuring 16GB of VRAM. The training framework is based on PyTorch 3.8.0, utilizing CUDA 11.4 for GPU acceleration. During model training, considering the characteristics of the dataset, a lower version of YOLOv5 hyperparameters is used, and Adam is chosen as the network optimizer. The initial learning rate is set to 0.001, and mini-batch gradient descent is employed with a weight decay of 0.0005. The dataset is randomly and uniformly sampled in batches for training. Finally, the momentum parameter is set to 0.937, the batch size for iterations is set to 9, and the total number of iterations is 200.

3.2. Model performance evaluation

Model performance evaluation includes both qualitative and quantitative analyses. This study uses the COCO dataset evaluation system, which provides more comprehensive evaluation information and results in a more stable model. Common evaluation metrics for object detection models include Precision, Recall, Average Precision (AP), and Mean Average Precision (mAP).

Fig. 11(a), (b), and (c) show the identification and detection results of grain caking at viewing distances of 30 cm, 60 cm, and 90 cm, respectively. A qualitative analysis of the model reveals that the detection performance for small objects has been improved by the proposed algorithm, with high accuracy and confidence in detecting the target objects. Moreover, the detection at different viewing distances can accurately distinguish the severity of grain caking, allowing the robot to adopt different cleaning strategies based on the varying degrees of caking during the cleaning process.

When performing quantitative analysis of the model, the choice of IoU threshold can affect the calculation of mAP. Since the target objects, grain caking, are small and irregularly shaped, a lower IoU threshold should be used to improve recall. Based on practical requirements and general standards, an IoU threshold of 0.5 is selected. Additionally, the COCO dataset introduces the parameter mAP@[0.5:0.95], which calculates mAP using ten IoU thresholds evenly spaced between 0.5 and 0.95, and averages these ten values to obtain the final result. For real-time target detection, which is crucial for the robot's cleaning efficiency, another important parameter is FPS (Frames Per Second). A higher FPS value indicates better real-time performance of the detector.



a) Detection results at 30 cm range b) Detection results at 60 cm range c) Detection results at 90 cm range

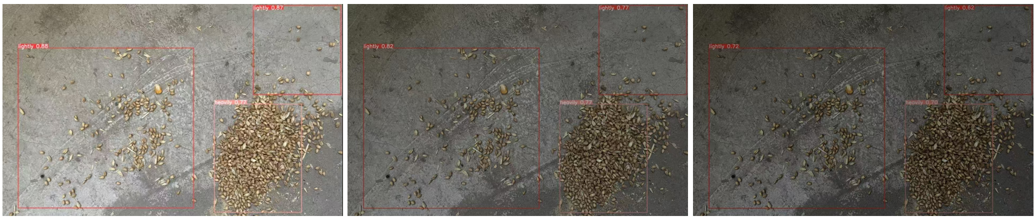
Fig. 11. Identification and detection results of grain caking

To highlight the advantages of the improved algorithm, comparative experiments were conducted with YOLOv4, YOLOv5s, and Fast R-CNN models. Each of these classic algorithms was trained on the custom grain caking dataset for 200 iterations. The results showed that the feature extraction by these models was not as effective, leading to lower detection accuracy. In comparison, the proposed algorithm showed significant improvements across various evaluation metrics, as shown in Table 1. Specifically, the AP values for the Lightly and Heavily categories increased by 3.8 % and 1.8 %, respectively, compared to the highest-performing Fast R-CNN model. The mAP of the improved algorithm was 5.9 % and 3.7 % higher than those of YOLOv4 and YOLOv5s, respectively. Although the mAP of the improved algorithm is slightly lower than that of the Fast R-CNN model, the enhancements made to the original IoU function and the data augmentation processes have significantly increased detection accuracy. Compared to the other three models, the proposed algorithm demonstrates superior overall performance and is more suitable for real-time detection of grain caking in the complex environment of grain silos.

Table 1. Comparison of improved algorithm performance

Algorithm	Lightly	Heavily	Precision	Recall	mAP^{val}50	mAP^{val}50-95	FPS
YOLOv4	73.3	84.9	91.5	91.8	92.3	78.5	34.7
YOLOv5s	74.7	85.7	92.3	92.5	94.5	79.3	38.5
Fast R-CNN	75.5	86.4	92.8	96.0	98.6	81.7	31.4
Improved algorithm	79.3	88.2	98.9	98.7	98.2	85.6	39.7

Experiments were conducted to evaluate the effectiveness of the algorithm under low-light conditions. The performance of the algorithm in recognizing grain caking under different lighting intensities is shown in Fig. 12.



a) 264 LUX Light intensity b) 122 LUX Light intensity c) 53 LUX Light intensity

Fig. 12. Identification of grain caking under different lighting conditions

The lighting intensity of the three images decreased by 53.79 % and 56.56 %, respectively. The algorithm's confidence in detecting light grain caking decreased by 9.16 % and 12.59 %, respectively. When the lighting intensity decreased by 79.92 % overall, the algorithm's confidence in detecting severe grain caking decreased by only 9.10 %. These experiments demonstrate that the model with the added low-light enhancement algorithm meets the operational requirements in the dark environment of a silo.

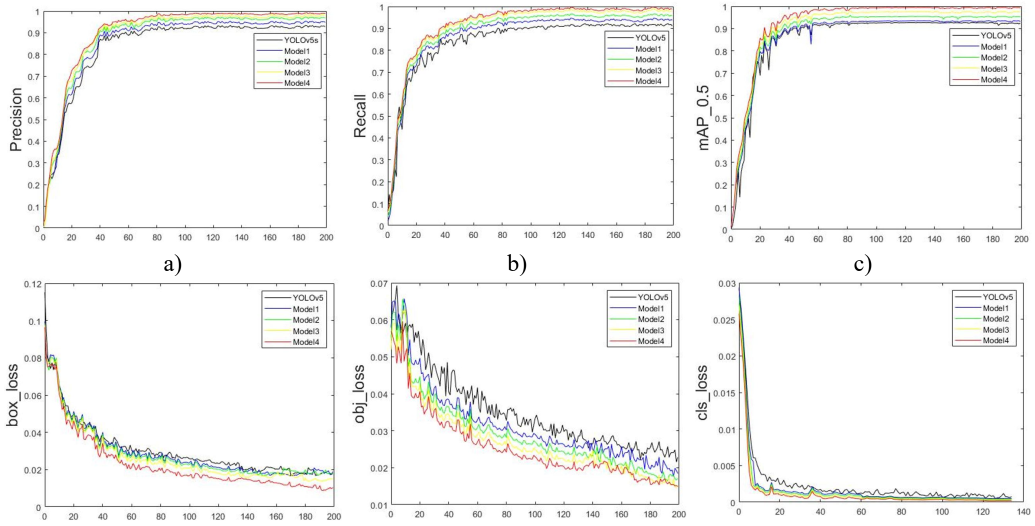
3.3. Ablation study

To verify the effectiveness of adding different modules in improving the algorithm's performance, the following ablation experiments were conducted based on the custom Grain dataset. As shown in Table 2, YOLOv5s represents the unmodified original algorithm. Model1 adds a data augmentation algorithm (DA) to the original model. Model2 adds the CBAM attention module to Model 1. Model 3 further incorporates the CIoU_Loss improved loss function on top of Model 2. Model 4, which is the final improved model presented in this paper, adds a tiny target detection layer (TTDL) to Model 3.

Table 2. Experimental results of Grain dataset ablation under four improved methods

Model	DA	CBAM	CIoU_Loss	TTDL	Parameters	GFLOP/s	mAP	FPS
YOLOv5s	×	×	×	×	78.4	14.0	94.5	38.5
Model1	√	×	×	×	79.7	16.5	96.5	36.5
Model2	√	√	×	×	81.8	15.8	98.6	40.4
Model3	√	√	√	×	83.5	15.8	97.8	41.3
Model4	√	√	√	√	85.6	15.8	98.2	42.7

In the ablation experiments, five different models were trained on the custom Grain dataset. The curves showing the changes in different parameters during the training process are depicted in Fig. 13. It can be observed that various improved models contribute to enhanced detection accuracy. As the number of training rounds increases, accuracy, recall, and mAP_0.5 values consistently improve, while the loss function values decrease to varying extents.



a) b) c)

d) e) f)

Fig. 13. Convergence curves of different models

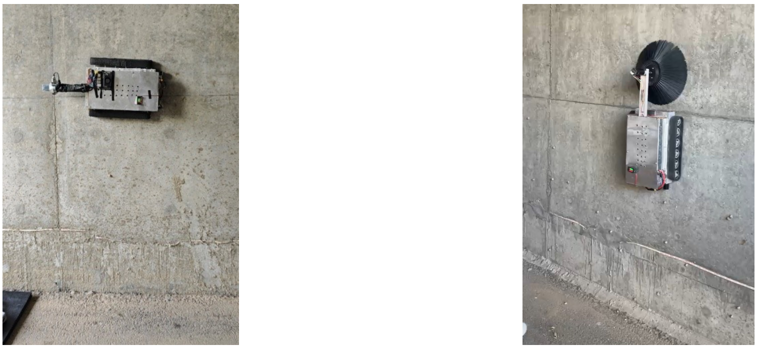
Before 40 rounds, the model's accuracy significantly increases, with a noticeable drop in the loss function, especially a rapid decrease in cls_loss. At 60 rounds, the accuracy improvement begins to slow, and the loss function decrease rate also slows down. After 100 rounds, the model's metrics stabilize, reaching optimal network weights by the end of the training. As shown in Fig. 13(a), Model4 performs the best, achieving a detection accuracy of 98.9 %. Fig. 13(b) shows that the recall rate for Model4 reaches 87.4 % at 40 rounds and eventually grows to 98.7 %. Fig. 13(c) demonstrates a rapid increase in mAP_0.5, with a peak value of 98.2 %. The loss functions of the model, including the bounding box loss, object loss, and classification loss, are shown in Fig. 13(d), (e), and (f), respectively. It can be seen that the improved model demonstrates enhancements in the prediction accuracy of the bounding box, the existence of detected objects, and classification accuracy.

The results of the ablation experiments indicate that the proposed algorithm has better detection performance, faster regression speed, and higher accuracy, which demonstrates the effectiveness of the improved model for the recognition and detection of grain caking in silo applications.

3.4. Wall-climbing robot surface motion and cleaning experiments

By deploying the optimized algorithm described in this paper on the NVidia Jetson Orin Nano mounted on the robot, a seamless integration of the recognition system with the robotic system is achieved. The wall-climbing robot designed in this paper features a modular design, allowing different functions to be performed by swapping the front-end actuators. Fig. 14(a) shows the wall-climbing robot equipped with a camera, capable of identifying and detecting grain caking and transmitting the collected data to a host computer. Fig. 14(b) depicts the wall-climbing robot fitted with a hard brush disc, which uses a brushless motor to drive the high-speed rotation of the disc to remove grain caking, completing the entire operation process.

Experiments on wall-climbing robot movement and cleaning operations demonstrate that the robot can stably adhere to and move along the wall while carrying a camera or cleaning brush. It can successfully complete the full process of identifying and cleaning grain caking, meeting the engineering requirements for grain caking identification and cleaning on the inner walls of silos.



a) The wall-climbing robot equipped with a camera for detection

b) The wall-climbing robot equipped with a brush for cleaning

Fig. 14. The operation of the wall-climbing robot

4. Conclusions

This paper proposes the development and evaluation of a convolutional neural network model for robot vision detection of grain caking. This article proposes the following improvements to the visual detection algorithm based on YOLOv5: Firstly, the Convolutional Block Attention Module (CBAM) and the improved Total Cross Union (CIoU) loss function are introduced to enhance the detection accuracy of grain caking. Secondly, by adding the Retinex Net algorithm with dark light enhancement, the recognition and detection performance under low light conditions can be improved. The improved YOLOv5 algorithm was trained and validated on a self-made grain caking dataset.

The results indicate that the improved algorithm enhances the accuracy of grain caking recognition by 1.8 % to 3.8 % compared to other algorithms. The wall-climbing cleaning robot equipped with this algorithm can effectively identify and clean grain caking, meeting the needs of grain caking tasks in silos.

However, in practical working environments, varying silo capacities and degrees of compaction present challenges. Future work will focus on improving detection real-time performance and generalizability, particularly on enhancing detection speed while increasing algorithm accuracy. Collaborative operation with multiple robots will also be explored to meet broader engineering application requirements.

Acknowledgements

This project was supported by the Henan Province Key Program for International Science and Technology Cooperation (Grant No. 232102520002) and the Henan Province Key R&D Program (Grant No. 221111220300).

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions

Yi Cao: design of the study and the formulation of the experimental plan, leading the data analysis and interpretation of results. Yao Zhao: implementation of the research, algorithm development and optimization, and wrote the main content of the paper. Xiang Wu: technical support, experimental tools and experimental environment. Mingqi Tang: data collection and preprocessing, analysis of results and generation of figures. Chao Gu: writing of the literature review and discussion sections.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] S. K. Cheng et al., “Thoughts on food security in China in the new period,” *Journal of Natural Resources*, Vol. 33, No. 6, pp. 911–926, Jun. 2018, https://doi.org/10.31497/zrzyxb.20170527

[2] Y. Xue, “Analysis of various stress states of suspended cables in silos,” (in Chinese), *Grain Distribution Technology*, No. 3, pp. 3–4, Sep. 2004.

[3] C. K. Wu, “Analysis of dust explosions and explosion prevention measures in grain silos,” (in Chinese), *Grain Distribution Technology*, No. 5, pp. 29–33, Sep. 2009.

[4] K. Dandan, A. Ananiev, and I. Kalaykov, “SIRO: The silos surface cleaning robot concept,” in *2013 IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 657–661, Aug. 2013, https://doi.org/10.1109/icma.2013.6617994

[5] Y. Fang, S. Wang, Q. Bi, D. Cui, and C. Yan, “Design and technical development of wall-climbing robots: a review,” *Journal of Bionic Engineering*, Vol. 19, No. 4, pp. 877–901, Jun. 2022, https://doi.org/10.1007/s42235-022-00189-x

[6] X. Ming Cui, Y. Fei Sun, and F. Jun He, “Research and development of wall-climbing robots,” (in Chinese), *Sci. Technol. Eng.*, Vol. 10, No. 11, pp. 2672–2677, 2010.

[7] M. Tavakoli, C. Viegas, L. Marques, J. N. Pires, and A. T. de Almeida, “OmniClimbers: Omni-directional magnetic wheeled climbing robots for inspection of ferromagnetic structures,” *Robotics and Autonomous Systems*, Vol. 61, No. 9, pp. 997–1007, Sep. 2013, https://doi.org/10.1016/j.robot.2013.05.005

[8] P. L. Pan, X. S. Gao, and G. R. Yan, “Design of the spraying mechanism for the tracked magnetic adhesion wall-climbing robot,” (in Chinese), *Robot.*, Vol. 19, No. 2, pp. 68–71, 1997.

[9] A. Hajeer, L. Chen, and E. Hu, “Review of classification for wall climbing robots for industrial inspection applications,” in *IEEE 16th International Conference on Automation Science and Engineering (CASE)*, pp. 1421–1426, Aug. 2020, https://doi.org/10.1109/case48305.2020.9216878

[10] S. Sharma, A. B. Gurulakshmi, D. Yadav, A. Pandey, and B. A. Pindth, "Automated facade cleaning robot using reinforcement learning model," in *International Conference on Smart Systems for applications in Electrical Sciences (ICSSES)*, pp. 1–5, May 2024, https://doi.org/10.1109/icsses62373.2024.10561351

[11] H. Ding and S. Wu, "Bridge crack segmentation and measurement based on SOLOv2 segmentation model," *Journal of Measurements in Engineering*, Vol. 12, No. 3, pp. 502–518, Sep. 2024, https://doi.org/10.21595/jme.2024.23987

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, Jun. 2016, https://doi.org/10.1109/cvpr.2016.91

[13] D. Thuan, "Evolution of Yolo algorithm and Yolov5: The state-of-the-art object detection algorithm," Oulu University of Applied Sciences, 2021.

[14] A. Sengupta, Y. Ye, R. Wang, C. Liu, and K. Roy, "Going deeper in spiking neural networks: VGG and residual architectures," *Frontiers in Neuroscience*, Vol. 13, p. 95, Mar. 2019, https://doi.org/10.3389/fnins.2019.00095

[15] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: generalizing residual architectures," *arXiv:1603.08029*, Jan. 2016, https://doi.org/10.48550/arxiv.1603.08029

[16] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: an extremely efficient convolutional neural network for mobile devices," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6848–6856, Jun. 2018, https://doi.org/10.1109/cvpr.2018.00716

[17] Y. Mi, M. Chi, Q. Zhang, P. Liu, and F. Sun, "Lightweight underwater target detection method based on improved YOLOv5s," *Recent Patents on Engineering*, Vol. 19, No. 4, May 2025, https://doi.org/10.2174/0118722121294044240422063140

[18] H. Liu, F. Sun, J. Gu, and L. Deng, "SF-YOLOv5: a lightweight small object detection algorithm based on improved feature fusion mode," *Sensors*, Vol. 22, No. 15, p. 5817, Aug. 2022, https://doi.org/10.3390/s22155817

[19] C. Liang, "Garbage detection suitable for low computing power devices and dark light environments," (in Chinese), M.S. thesis, Jinan University, Jinan, China, 2024.

[20] H. Liu, "Research on feature enhancement based dark light object detection method," (in Chinese), M.S. thesis, Donghua University, Shanghai, China, 2024.

[21] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *European Conference on Computer Vision*, Jan. 2018.

[22] J. Yang, Q. Feng, and S. Wang, "Dense small object detection in field environments based on improved YOLOv4," (in Chinese), *Journal of Northeast Agricultural University*, Vol. 53, No. 5, pp. 69–79, 2022.

[23] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv:1808.04560*, Jan. 2018, https://doi.org/10.48550/arxiv.1808.04560

[24] F. Dadboud, V. Patel, V. Mehta, M. Bolic, and I. Mantegh, "Single-stage UAV detection and classification with YOLOV5: mosaic data augmentation and PANet," in *17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–8, Nov. 2021, https://doi.org/10.1109/avss52988.2021.9663841

Yi Cao received his Ph.D. in mechanical engineering from the School of Mechanical Engineering, Tianjin University, Tianjin, China, in 2005. He is currently a Professor at the School of Mechanical and Electrical Engineering, Henan University of Technology, Zhengzhou, China. His current research interests include wall-climbing robot design and control, machine vision, and perception.

Yao Zhao is a master's student at the School of Mechanical and Electrical Engineering, Henan University of Technology, Zhengzhou, China. His research focuses on recognition and detection systems using machine vision technologies.

Xiang Wu received his Ph.D. in Control Science and Engineering from Dalian Maritime University, Dalian, China, in 2011. He is currently a Lecturer at the School of Electrical Engineering, Henan University of Technology, Zhengzhou, China. His research interests include image processing and multi-sensor fusion technology.

Mingqi Tang is a master's student at the School of Mechanical and Electrical Engineering, Henan University of Technology, Zhengzhou, China. His research focuses on robot control and trajectory planning.

Chao Gu received his master's degree from the School of Mechanical and Electrical Engineering, Henan University of Technology, Zhengzhou, China, in 2024. Now he works at Company. His research interests include feature recognition and image processing for robotic arm grasping applications.