Recognition model for coated red clover seeds using YOLOv5s optimized with an attention module

Xiwen Zhang[1], Chuanzhong Xuan[1], Zhanfeng Hou[1,2*]

(1. *College of Mechanical and Electrical Engineering, Inner Mongolia Agricultural University, Inner Mongolia, Hohhot 010018, China*;
2. *Inner Mongolia Engineering Research Center of Intelligent Equipment for the Entire Process of Forage and Feed Production, Inner Mongolia, Hohhot 010018, China*)

Abstract: The non-destructive recognition of coated seeds is crucial for advancing studies in coating theory. Currently, the recognition of coated seeds heavily relies on manual visual inspection and machine vision detection. However, these methods pose challenges such as high misclassification rates, low recognition efficiency, and elevated labor intensity. In response to the aforementioned challenges, this study leveraged deep learning techniques to develop a coated seed recognition model named YOLO-Coated Seeds Recognition (YOLO-CSR), aiming to address the challenges posed by coated seed recognition tasks. The experiment of this study mainly includes the following steps: First, a seed coating machine was set up to coat red clover seeds, resulting in three types of coated red clover seeds. Subsequently, by collecting images of the three types of coated seeds, a coated seed image dataset was further constructed. Then, the YOLOv5s was built, incorporating the Convolutional Block Attention Module (CBAM) into the model's backbone to enhance its ability to learn features of coated seeds. Finally, the training results of YOLO-CSR were compared with those of other classical recognition models. The experimental results showed that YOLO-CSR achieved the best recognition performance on the self-built coated seed image dataset. The average precision (AP) for recognizing the three types of coated seeds reached 98.43%, 97.91%, and 97.26%, with a mean average precision@0.5 (mAP@0.5) of 97.87%. Compared to YOLOv5, YOLO-CSR showed a 1.18% improvement in mAP@0.5. Additionally, YOLO-CSR has a model size of only 14.9 MB, with an average recognition time (ART) of 10.1 ms and a frame per second (FPS) of 99. Experimental results prove that YOLO-CSR can accurately, efficiently, and rapidly recognize coated red clover seeds. The findings of this study provide technical support for the non-destructive recognition of spherical coated seeds.

Keywords: coated seed recognition, red clover seed, YOLO, Attention Module, CNNs
DOI: 10.25165/j.ijabe.20231606.7773

Citation: Zhang X W, Xuan C Z, Hou Z F. Recognition model for coated red clover seeds using YOLOv5s optimized with an attention module. Int J Agric & Biol Eng, 2023; 16(6): 207–214.

1 Introduction

As the initial stage of crop growth, seeds play a pivotal role in agricultural production, serving as a crucial means of production, and the quality of seeds fundamentally influences agricultural yields[1-3]. Seed coating technology emerges as a promising method to enhance seed quality, involving the application of coating agents to augment performance, treatment, and overall plant structure[4-6]. The coating process ensures uniform contact between seeds and coating agents, resulting in the formation of a smooth and durable coating film[7]. Coated seeds exhibit a spherical shape with uniform sizes, including round, oval, and oblate varieties[8].

In practical coating processes, the choice of coating formulations and processes significantly impacts the success rate of coating[9-11]. When delving into coating theory, the selection of qualified coated seeds becomes imperative for evaluating the coating's effectiveness. Traditional seed recognition methods, such as manual visual inspection and machine vision recognition, are currently employed. Manual visual inspection heavily relies on human judgment, leading to challenges like low recognition efficiency, a high misclassification rate, and increased labor intensity[12]. Machine vision-based seed recognition methods extract shallow features such as color and texture to achieve recognition[13]. However, these methods demand strict conditions related to recognition, background color, and brightness, often resulting in issues like low recognition accuracy and limited robustness[14]. Overall, traditional seed recognition methods face challenges, prompting the need for a swift and precise coated seed recognition technique.

With the rapid development of artificial intelligence, scholars have achieved excellent results in seed recognition tasks by employing Convolutional Neural Networks (CNNs)[15-17]. In traditional pattern recognition methods, identification conditions are manually set. In contrast, CNNs can automatically learn target features from the dataset through convolution[18-20]. Therefore, the features learned by CNNs are more robust. Simultaneously, seed features exhibit individual differences, and CNNs learn and summarize these individual differences. They employ multiple criteria, such as shape and pattern, to accomplish seed recognition tasks[21-23]. Loddo et al.[24] built a SeedNet network to classify multiple types of seeds in datasets, with accuracy rates of 95.65% and 97.47% on two sets of datasets, respectively, achieving satisfactory results. Wang et al.[25] combined CNNs and Long Short-

Received date: 2023-06-28 **Accepted date:** 2023-06-28
Biographies: Xiwen Zhang, PhD candidate, research interest: agricultural machine vision, Email: njau_zxw@163.com; **Chuanzhong Xuan**, Professor, research interest: intelligent sensing of agricultural information, Email: xcz@imau.edu.cn.
***Corresponding author: Zhanfeng Hou**, Professor, research interest: Intellectualization of agricultural equipment. Inner Mongolia Agricultural University, 306 Zhaowuda Road, Hohhot 010018, China. Tel: +86-471-4309215, Email: njau-hzf@163.com.

Term Memory (LSTM) to build a spectral image of sweet maize seeds using CNN-LSTM. The CNN-LSTM performed best, with a classification accuracy of over 95%. Yu et al.[26] used a CNNs model to identify spectral images of okra variety seeds. The CNNs model achieved the highest accuracy and had the strongest robustness, with a recognition accuracy of 93.79%. Despite the widespread application of deep learning in seed recognition tasks, there is currently limited research on coated seed recognition. On the other hand, existing seed recognition models have not considered indicators such as model size and recognition speed. Specifically, in the practical application process, it is necessary to embed the recognition model into the identification device to complete the entire recognition workflow. Therefore, there are certain requirements for the model size and recognition time. To address the aforementioned issues, it is essential to develop a fast, accurate, and robust coated seed recognition model.

You Only Look Once (YOLO), with its unique architecture and modules, has demonstrated superior performance across numerous recognition tasks[27-29]. With the evolution of the YOLO series algorithms, YOLOv5 has gained increasing attention. In previous studies, YOLOv5 has achieved excellent performance in multiple recognition tasks. Among the four versions of YOLOv5, YOLOv5s stands out with a compact model size of only 14.0 MB, endowing it with a significant advantage in both model size and recognition speed. Given the high demands of coated seed recognition models for model size and recognition speed, and to fill the gap in coated seed recognition models, this study establishes a high-precision coated seed recognition model based on YOLOv5s, named YOLO-CSR.

This study proposes a recognition method for coated red clover seeds, which integrates YOLOv5s with an attention module. The addition of the attention module enables the recognition model to focus solely on the relevant parts of image information for the task, filtering out extraneous details. Additionally, various attention mechanisms are introduced into the YOLOv5s backbone in this study, and the training results are further compared. Experimental findings highlight that the integration of the convolutional block attention module (CBAM) into the backbone of YOLOv5s yields the most significant enhancement in the model's ability to recognize coated seeds. Consequently, CBAM was integrated into the backbone of YOLOv5s, contributing to improving the model's ability to learn target features. This method facilitates real-time and precise recognition of coated red clover seeds, thereby offering technical support for tasks related to recognizing spherical coated seeds. The main contributions of this study are as follows:

1) To address the gap in coated seed recognition research, the recognition model YOLO-CSR was developed.

2) The attention module CBAM was embedded in the backbone of the recognition model to enhance the model's recognition performance.

3) YOLO-CSR has a model size of only 14.9MB and an ART of 10.1ms, demonstrating its significant advantages in both model size and recognition speed.

2 Materials and methods

2.1 Related work

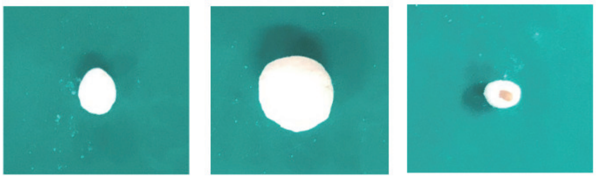
The rotating seed coating machine used in this study was designed and manufactured by the School of Mechanical and Electrical Engineering at Inner Mongolia Agricultural University in China. The device includes three main parts: the coating pot, the control desk, and the electric control table, as shown in Figure 1. Through the seed coating machine, efficient coating operations can be conducted on red clover seeds, thereby completing the related coating works.

1. Electric control table 2. Coating pot 3. Control desk

Figure 1 The rotating seed coating machine

This study used Badong red clover seeds obtained from farmland in Wuhan, China for the coating process. The details of the coating procedure are as follows: Soybean powder with a particle size of 200 μm and diatomaceous earth with a fineness of 100 mesh were mixed in a 5:5 ratio as the coating powder. The seed powder ratio and liquid seed ratio were 1:3 and supplied in 5 batches. The interval between supply batches was 3 min, and the total coating time was 30 min. The coating process was performed in the School of Mechanical and Electrical Engineering of Inner Mongolia Agricultural University in March 2021.

The coated red clover seeds were divided into 3 types: qualified, seedy, and broken seeds. Based on pre-experimental observations, the qualified coated red clover seeds were spherical with a diameter ranging between 3 mm and 5 mm. The sample of qualified coated red clover seed is shown in Figure 2a. During the coating process, several seeds fused into a single coated seed under the rotational force of the coating pan, and these coated seeds were defined as seedy coated seeds. The shapes of seedy coated red clover seeds were irregular, with the majority having an elliptical shape, and the major axis of the elliptical surface exceeding 5 mm. Therefore, the maximum diameter of the coated seed was used as the threshold to distinguish between qualified coated seeds and seedy coated seeds. The sample of seedy coated red clover seed is shown in Figure 2b. Observation of the coated seeds with a magnifying glass found that some of the coated seeds were not fully covered, resulting in some seed areas being exposed. These coated seeds were defined as broken coated seeds. The sample of broken coated red clover seed is shown in Figure 2c.

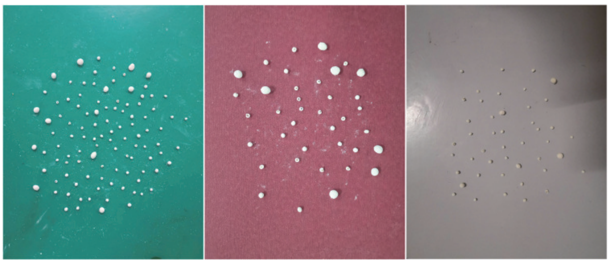


a. Qualified b. Seedy c. Broken

Figure 2 Samples of different types of coated red clover seeds

2.2 Dataset

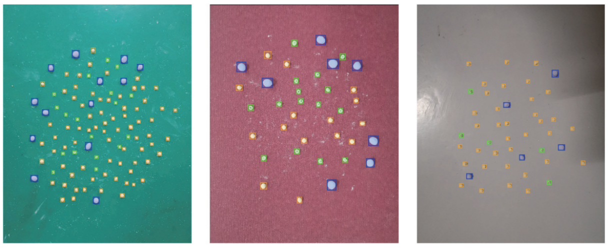
In this study, a dataset of coated red clover seeds was constructed for subsequent model training and evaluation. The specific construction details are as follows: Coated red clover seeds were randomly selected and placed on the experimental platform for image capture. The experimental platform backgrounds included green, red, and gray colors. The images were collected in April 2021 at the Institute of Lakes and Environmental Engineering in the School of Mechanical and Electrical Engineering. The images of coated red clover seed were captured using a Canon EOS 600D (Canon, Japan). The image resolution was 3000 pixels×4000 pixels, and the images were saved in JPG format. The shooting height was controlled between 30 cm and 50 cm to ensure that every seed on the experimental platform could be captured, and the shooting height of each image was randomly set. Calculation standards for coated seed qualification rates vary among countries. Taking China's calculation standards as an example, 600 coated seeds were randomly selected from the test samples and evenly divided into three groups. Using a magnifying glass for visual observation, the number of qualified coated seeds in each group was calculated to determine the coating qualification rate. Considering that the number of coated seeds tested in a single batch was 200, the number of coated seeds in each image of the dataset was controlled to be within 300. All images were captured under indoor lighting conditions, with the shooting light conditions set randomly. A total of 1637 original images were collected to construct the coated seed dataset. Sample images from the coated seed dataset are shown in Figure 3.



a. Green background b. Red background c. Gray background

Figure 3 Sample images of the coated seed dataset

This study employed the Make Sense online labeling tool. In this labeling tool, coated seeds in each image were manually annotated using bounding boxes, classifying them as Qualified, Seedy, or Broken. As shown in Figure 4, manually label the coated red clover seeds with a box tangent to the seed outline. A total of 1637 original images of coated red clover seeds were manually labeled. The coated red clover seeds within orange, blue, and green boxes represent qualified, seedy, and broken coated seeds, respectively.



a. Green background b. Red background c. Gray background

Figure 4 Sample label images of coated seed dataset

After labeling, TXT-format label files were generated for subsequent model training. As listed in Table 1, the coated seed dataset was randomly divided into training, validation, and test sets at a ratio of 8:1:1. The training set comprises a total of 1309 images of coated seeds. The training, validation, and test sets are mutually exclusive to ensure the reliability of subsequent evaluations.

Table 1 Summary of the coated seed dataset

Datasets	Qualified quantity	Seedy quantity	Broken quantity	Image quantity	Proportion
Training set	20 811	10 010	10 013	1309	80%
Validation set	2533	869	840	164	10%
Test set	2701	1222	1110	164	10%
Total	26 045	12 101	11 963	1637	100%

In the practical application phase, the recognition model needs to overcome the impact of complex lighting conditions to accurately complete the coated seed recognition task. Therefore, to enhance the robustness of the recognition model, we applied random brightness transformations to the images in the training set. When the training model can overcome challenges arising from diverse lighting conditions, it demonstrates the model's exceptional performance in completing the coated seed recognition task. Specifically, assuming the initial brightness of the image is 100%, this study randomly adjusted the brightness values of training images within the range of 70%-130%. Simultaneously, when the initial contrast of the image is 100%, the contrast of some training images is adjusted to 70% or 130%. The primary purpose of adjusting contrast is to make the model more adaptable, enabling it to robustly recognize coated seeds under a wider range of lighting conditions.

2.3 Coated red clover seed recognition model

2.3.1 YOLOv5s

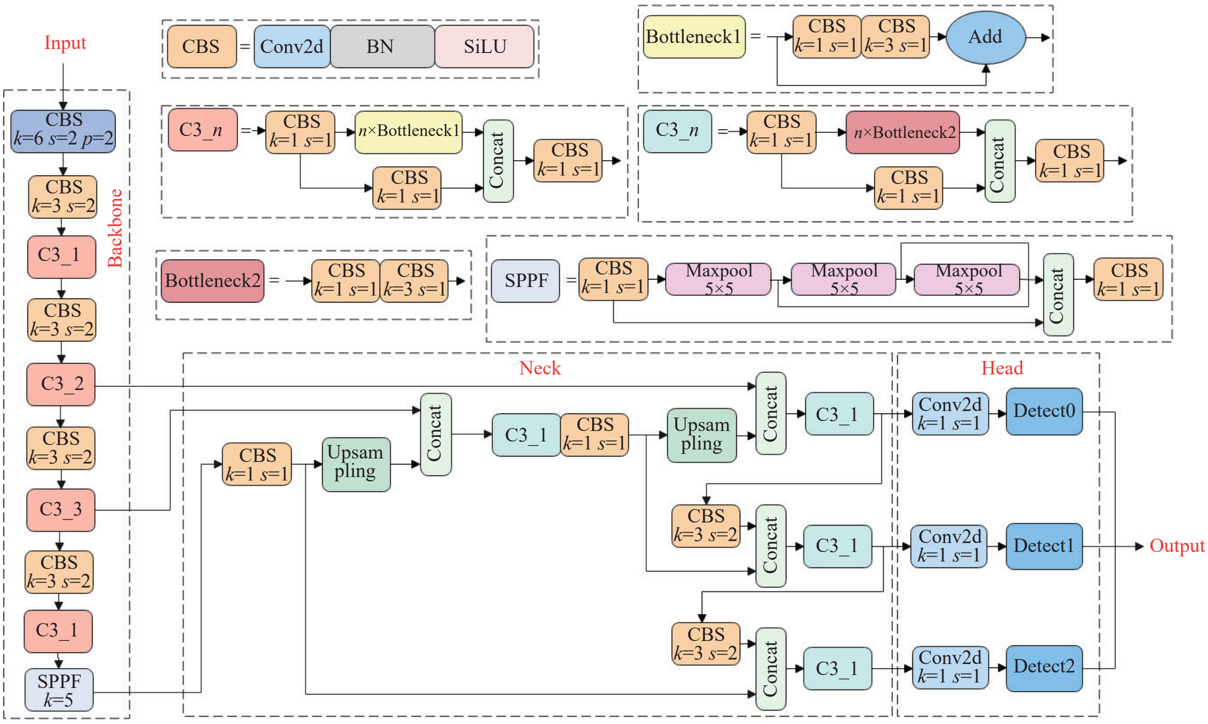
The YOLO framework effectively distinguishes target and background areas to achieve target recognition[30,31]. The YOLOv5s is composed of inputs, backbone, neck, head, and output. The YOLOv5s model used in this study is v6.1. Compared with the previous v5.0 version, the network structure of v6.1 is more streamlined.

Specifically, the input section includes image size processing, mosaic data enhancement, and adaptive anchor box calculation. Image size processing adds an adaptive minimum black border to the original image with varying lengths and widths so that the original image is uniformly resized to the standard size[32]. The mosaic data enhancement method involves randomly cutting and zooming four pictures, then arranging and splicing them randomly to form a picture, enriching the dataset, and introducing small sample targets to enhance the network's training speed. Adaptive anchor box calculation utilizes K-means and genetic learning algorithms to analyze the user-defined dataset and obtain preset anchor boxes suitable for predicting object boundary boxes in the dataset.

The backbone network of YOLOv5s consists of CBS, C3 module, and Spatial Pyramid Pool-Fast (SPPF). The CBS is the fundamental convolutional unit of YOLOv5s, consisting of convolution, normalization, and activation functions. The residual structure module and CBS module together form the C3 module. The C3 module incorporates the concept of residual structure and comprises two branches. The residual structure enhances the gradient value of back-propagation between layers, preventing gradient disappearance caused by deepening. Connecting the two branches through the concat function helps retain characteristic information from different branches to extract more abundant feature information. The SPPF module uses three 5×5 maximum pooling layers to effectively address issues like incomplete image cropping and shape distortion, obtaining more feature information by fusing features of different resolutions.

The Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) structures are employed in the neck network for multi-scale feature fusion[33-36]. The FPN structure enhances the underlying feature map's semantic information by upsampling from top to bottom. The PAN structure downsamples from bottom to top, ensuring that top features contain location information. Finally, the two features are fused, ensuring that feature maps of different sizes contain both semantic and feature information, guaranteeing accurate prediction for images of different sizes.

The head end comprises three recognition layers with feature maps of different sizes, used for recognizing target categories[37]. Each recognition layer outputs corresponding vectors, ultimately generating recognition results and marking the confidence degree, labeling the bounding boxes and categories of the recognized targets in the recognition image. The overall structure of the YOLOv5s is depicted in Figure 5.



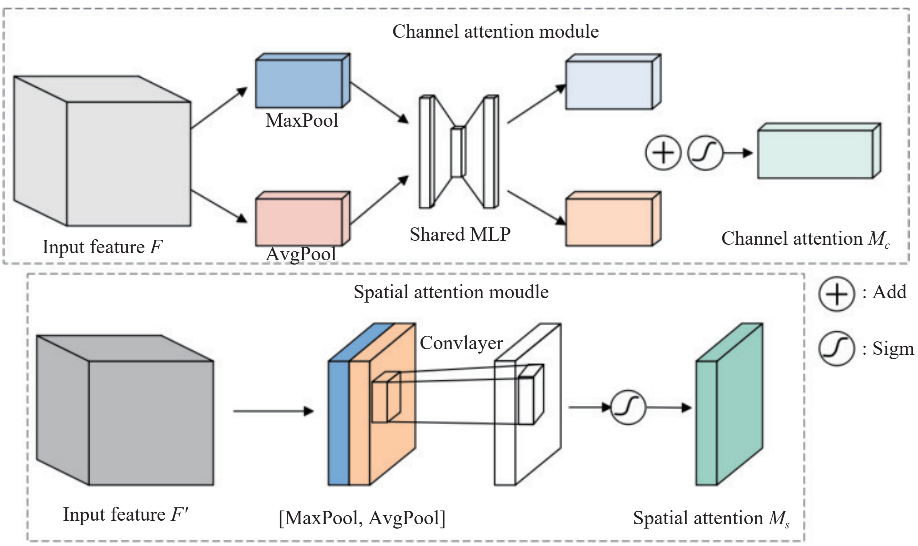
Note: CBS consists of convolution, normalization, and activation functions. *k* represents the size of the convolution kernel, *s* represents the stride size, and *p* represents the padding size. The residual structure module and CBS module together form the C3 module; SPPF: Spatial Pyramid Pool-Fast; BN: Batch Normalization; SiLU: Sigmoid-Weighted Linear Unit. Same below.

Figure 5 Overall structure of YOLOv5s

2.3.2 CBAM

CBAM comprises both a channel attention module and a spatial attention module[38]. The input features initially pass through the channel attention module, followed by the spatial attention module, ultimately resulting in weighted outcomes. The structural diagrams of the channel attention module and the spatial attention module are depicted in Figure 6.

The specific process of the channel attention module is as follows: Initially, the input feature F undergoes Global Max Pooling and Global Average Pooling operations, yielding two output vectors



Note: MLP: Multi-Layer Perceptron.

Figure 6 The structural diagram of CBAM

through a Multi-Layer Perceptron (MLP)[39]. The operation of addition is then applied to these two output vectors, followed by passing through the Sigmoid activation function, resulting in the generation of the final channel attention feature. The formulaic representation of the weight for each channel $M_C \in R^{C\times1\times1}$ is then determined. The calculation formula for M_C is shown in Equation (1). The channel attention module is completed by multiplying the input features F by the channel attention features $M_C(F)$ to obtain new features F', expressed as Equation (2).

$$M_C(F) = \sigma(W_1(W_0(F^C_{\mathrm{avg}})) + W_1(W_0(F^C_{\max}))) \tag{1}$$

$$F' = M_C(F) \otimes F \tag{2}$$

where, $F^C_{\max}$ and F^C_{avg} denote the max-pooled features and the average-pooled features in channel attention; σ represents the Sigmoid activation function; W_0 and W_1 denote the weights of two layers of MLP perception; ⊗ represents elementwise multiplication.

The input of the spatial attention module is the feature F'. Similar to the channel attention module, the input feature F' first goes through global maximum pooling and global average pooling, and the two resulting channels are connected through the Concat function. Then, a 7×7 convolution is used to compress the channel. The spatial attention feature $M_S \in R^{1\times H\times W}$ is generated through the Sigmoid activation function[35]. The calculation formula of spatial attention is shown in Equation (3). Finally, the final generated feature F'' is obtained by multiplying the input feature F' and the spatial attention feature $M_s(F')$, as follows:

$$M_S(F) = \sigma(f^{7\times7}([F^S_{\mathrm{avg}}; F^S_{\max}])) \tag{3}$$

$$F'' = M_S(F') \otimes F' \tag{4}$$

where, $f^{7\times7}$ represents a 7×7 traditional convolution operation; F^S_{avg} denotes the average-pooled features in spatial attention; $F^S_{\max}$ denotes the max-pooled features in spatial attention.

As shown in Figure 7, the CBAM module has been integrated into the backbone of YOLOv5s to accentuate pertinent information crucial for coated red clover seed recognition.

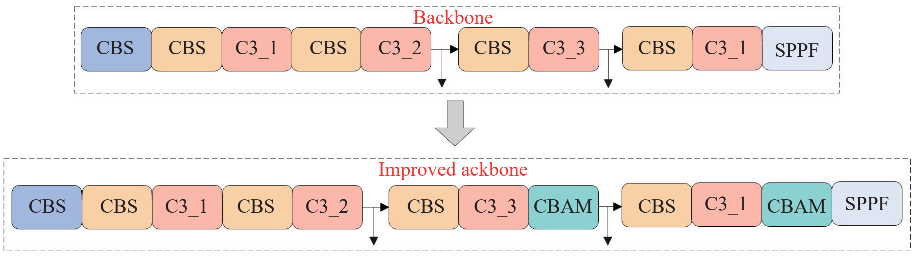


Figure 7 Schematic of the improved backbone network

2.3.3 YOLO-CSR

To achieve effective recognition of coated red clover seeds, this study constructed the YOLO-CSR recognition model. The overall architecture of YOLO-CSR is shown in Figure 8. YOLO-CSR comprises the backbone, neck, and head. The training images undergo feature extraction through the backbone, and the extracted features are further processed by the neck before the recognition results are output from the head.

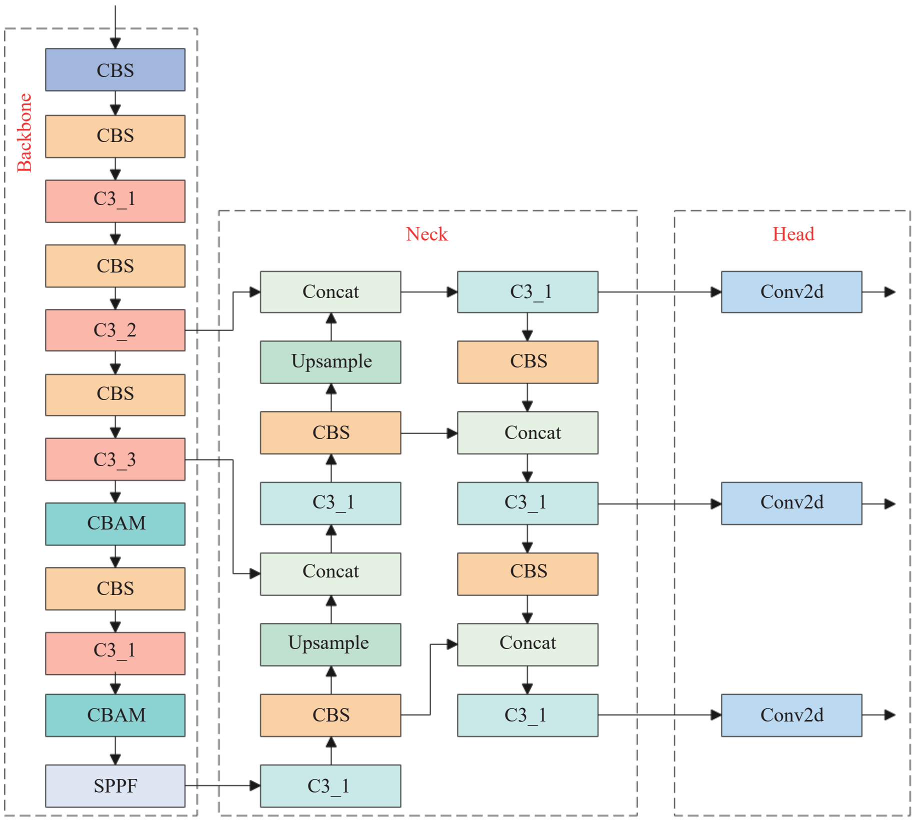


Figure 8 Overall architecture of YOLO-CSR

In this study, two sets of CBAM modules were integrated into the backbone of YOLOv5s. The YOLOv5s backbone employs a deep convolutional stacking structure, such as the C3 module, for detailed and robust feature extraction. Therefore, the last two C3 structures in the backbone extract the most abundant and detailed features of coated seed characteristics. In summary, we integrated CBAM behind the last two sets of C3 modules in the backbone, directing the network's attention to useful features while enhancing its ability to learn target characteristics.

2.4 Training parameter settings

In this study, the model was trained using the experimental platform configured with an RTX A5000 24 GB GPU, and the operating system is Windows 10 64-bit. The hyperparameter settings for model training are listed in Table 2. Both the training and test sets, as specified in Table 1, were employed for model training and evaluation. The software package CUDA 11.1 was applied to accelerate the training process. The PyTorch version used was 1.9.0, and the Python version was 3.8. PyCharm 2021.1.1 served as the development environment.

Table 2 Training results of different classic YOLO recognition models

Model	Precision/%	Recall/%	F1-score	Class	AP/%	mAP@0.5/%
YOLOv4-tiny	85.46	82.12	0.838	Qualified	80.76	83.39
				Seedy	85.53	
				Broken	83.87	
YOLOv3	92.87	90.30	0.916	Qualified	91.05	91.44
				Seedy	93.46	
				Broken	89.82	
YOLOv4	95.96	93.23	0.946	Qualified	95.39	95.03
				Seedy	96.04	
				Broken	93.67	
YOLOv5s	97.84	95.45	0.966	Qualified	97.08	96.69
				Seedy	96.62	
				Broken	96.37	
YOLO-CSR	98.87	96.12	0.975	Qualified	98.43	97.87
				Seedy	97.91	
				Broken	97.26	

The hyperparameters for model training are as follows: the training optimizer is Stochastic Gradient Descent, the number of training epochs is set to 100, the batch size is 16, the initial learning rate is 0.001, and the weight decay is set to 0.0005.

2.5 Performance evaluation

Precision, Recall, AP, and mAP@0.5 were employed to assess the performance of the proposed improved model. After predicting the test samples, four states of Precision and Recall can be defined: true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

The F1 score takes into account both Precision and Recall. Precision and Recall range from 0% to 100%. The F1 score ranges from 0 to 1, with a higher score indicating a better training effect of the model. Precision, Recall, and F1 score are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \times 100\% \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \times 100\% \tag{6}$$

$$\text{F1 score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

The AP for the three categories of coated red clover seeds is calculated to obtain mAP@0.5, used to evaluate the performance of the improved model. The mAP@0.5 is defined as in Equation (8).

$$\text{mAP@0.5} = \frac{1}{C}\sum_{k=i}^{N} P(k)\Delta R(k) \tag{8}$$

where, C is the number of recognition categories; N is the number of all pictures in the test set; $P(k)$ represents the Precision when k images can be recognized; $\Delta R(k)$ is the change in the recall value when the number of recognized images changes from $k-1$ to k. The IoU threshold was 0.5.

Additionally, this study employed ART and FPS to evaluate the model's recognition speed. For ART calculation, twenty randomly selected coated seed images were fed into the recognition model. The total recognition time was measured, and the final ART value was obtained by dividing the total time by twenty. The unit for ART is ms·image^{-1}.

In addition to ART, FPS is another commonly used metric for assessing object recognition speed, representing the number of images that can be processed per second. It is calculated by dividing 1000 ms by ART.

$$\text{FPS} = \frac{1000}{\text{ART}} \tag{9}$$

Model size, measured in megabytes (MB), refers to the amount of storage required to store the parameters and architecture of a particular model. This study used model size to evaluate the weight of the recognition model.

3 Results and discussion

3.1 Training curves

The training curve of the model is shown in Figure 9. During the training process, the mAP@0.5 value of YOLOv5s-CSR increases with the growth in the number of training epochs. Due to the initially high learning rate during the early stages of training, the mAP@0.5 value of the model experienced a rapid increase. As the number of training epochs reached 90, it gradually stabilized. Ultimately, the mAP@0.5 value of YOLOv5s-CSR settles at 97.87%. From the curves, it is evident that YOLOv5s-CSR outperforms YOLOv5s in terms of training effectiveness. The training curve validates that YOLO-CSR can efficiently accomplish the task of coated red clover seed recognition.

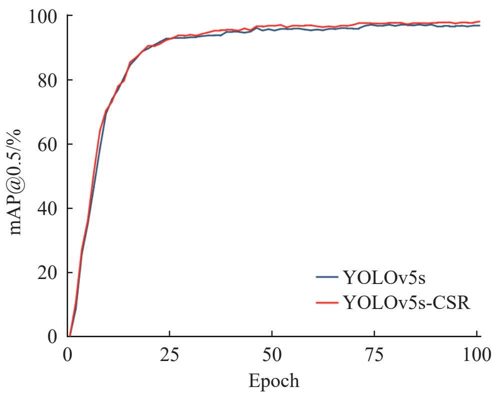


Figure 9 The training curves of the training models

3.2 Comparison with different YOLO networks

To evaluate the recognition performance of YOLOv5s-CSR, classic YOLO series recognition models, including YOLOv4-tiny, YOLOv3, YOLOv4, and YOLOv5s, were trained on the coated seed dataset, and the training results were further compared. A summary of the comparative experiment results is listed in Table 2. From the table, it is evident that YOLO-CSR achieved the best training results on the coated seed dataset. Specifically, YOLOv5s-CSR achieved a Precision of 98.87%, a Recall of 96.12%, and an F1-score of 0.975. The Average Precision (AP) values for the three types of coated red clover seeds were 98.43%, 97.91%, and 97.26%, resulting in a mAP@0.5 of 97.87%. In comparison to YOLOv4-tiny, YOLOv3, YOLOv4, and YOLOv5s, YOLOv5s-CSR's mAP@0.5 surpassed them by 14.48%, 6.43%, 2.84%, and 1.18%, respectively. Moreover, YOLOv5s-CSR achieved the highest F1 score, surpassing the other model groups by 0.137, 0.059, 0.029, and 0.009, respectively. In conclusion, among various YOLO series recognition models, YOLO-CSR demonstrated superior performance in handling coated red clover seed recognition tasks, making it the final coated seed recognition model in this study.

This study conducted specific tests on the recognition performance of YOLO-CSR. The specific testing procedure is as follows: randomly select an image from the test set and input it into YOLO-CSR for recognition. The recognition results of YOLO-CSR are shown in Figure 10. From Figure 10, it can be observed that each coated red clover seed in the image is effectively recognized and outlined. Among them, qualified, seedy, and broken coated red clover seeds are outlined in red, pink, and orange boxes, respectively.

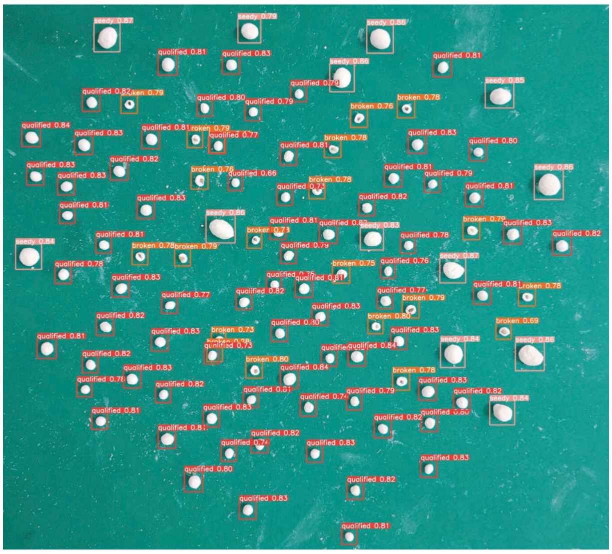


Figure 10 Example of coated seed recognition by the YOLO-CSR

3.3 Ablation experiment

To further explore the effectiveness of the improvement module, this study conducted ablation experiments. Specifically, two models, YOLOv5s and YOLO-CSR, were set up for comparative experiments. Considering that the core of ablation experiments is controlling variables, the training parameters for both models remained consistent. The evaluation of the two sets of models was conducted using specific performance metrics. The results of the ablation experiments are presented in Table 3.

Table 3 Results of ablation experiments

Model	F1-score	mAP@0.5/%	ART /ms·image^{-1}	Model size /MB	FPS /frames·s^{-1}
YOLOv5s	0.966	96.69	9.1	14.4	110
YOLO-CSR	0.975	97.87	10.1	14.9	99

As shown in Table 3, YOLO-CSR achieved an F1-score and mAP@0.5 of 0.975 and 97.87%, respectively, outperforming YOLOv5s by 0.009 and 1.18%. The comparative results indicate that the introduction of CBAM significantly improved recognition effectiveness, further enhancing the model's performance, and demonstrating the feasibility of the improvement strategy in this study. In addition, recognition speed and model size were also compared. Specifically, YOLO-CSR had an ART of 10.1 ms/image, reaching an FPS of 99, with a model size of only 14.9 MB. Compared to YOLOv5s, YOLO-CSR increased ART by 1.0 ms/image, decreased FPS by 10.0%, and increased model size by 0.5 MB. The results above suggest that the introduction of two sets of CBAM modules slightly increased the model size and decreased part of the recognition speed. Although introducing the CBAM modules slightly increased the model size and decreased part of the recognition speed, the impact on the actual application process is not significant. On the contrary, the introduction of CBAM broke through the recognition limits of the original model, significantly improving recognition performance. Therefore, YOLO-CSR was adopted as the final coated seed recognition model in this study.

3.4 Comparison of different attention mechanism modules

To explore the effects of introducing different attention mechanism modules, the multiple groups of attention mechanism modules were introduced into the backbone of YOLOv5s. Specifically, the integration positions of each set of attention mechanism modules align with the improvement strategy adopted in this study. The attention mechanism modules include the ECA (effective channel attention) module, SE (sequence and exception) module, CA (coordinate attention) module, and CBAM module. The experimental results are listed in Table 4.

Table 4 Results of different attention mechanism modules

Group	Model	F1 score	mAP@0.5/%
1	+SE	0.969	96.71
2	+ECA	0.971	97.03
3	+CA	0.973	97.20
4	+CBAM (Ours)	0.975	97.87

Observing Table 4 reveals that several models with attention mechanism modules achieved performance improvements. Among them, compared to other attention mechanism modules, the CBAM module demonstrated higher recognition accuracy. Specifically, introducing CBAM resulted in an increase in mAP@0.5 by 1.16%, 0.84%, and 0.67% compared to the introduction of other attention modules. Simultaneously, the F1 score increased by 0.006, 0.004, and 0.002, respectively. In summary, the introduction of CBAM led to the improved model achieving the best recognition performance. Therefore, this study adopted CBAM as the final enhancement strategy.

4 Conclusions

In this study, we employed deep learning techniques to detect three types of coated red clover seeds. We established a coated seed recognition model named YOLO-CSR, and enhanced the model's learning capabilities by integrating CBAM modules, further boosting the robustness of the recognition model. Ultimately, through the comparison of experimental results from different models, we demonstrated that the proposed YOLO-CSR can rapidly and effectively detect various types of coated red clover seeds. In future research, we aim to build a coated seed recognition system, deploy the recognition model on embedded devices, and achieve real-time recognition of coated seeds. Additionally, while optimizing seed recognition algorithms in future studies, there is potential to extend the algorithm for non-destructive recognition of other seed types.

Acknowledgements

This research was funded by the National Key Research and Development Program of China (Grant No. 2022YFF1302300), the Key R&D and Achievement Transformation Plan Project of Inner Mongolia (Grant No. 2023YFDZ0006), the Program for Improving the Scientific Research Ability of Youth Teachers of Inner Mongolia Agricultural University (Grant No. BR220128), and the Research Program of science and technology at Universities of Inner Mongolia Autonomous Region (Grant No. NJZZ23046).

[References]

[1] Zielinska M, Zapotoczny P, Białobrzewski I, Zuk-Golaszewska K, Markowski M. Engineering properties of red clover (*Trifolium pratense* L.) seeds. Industrial Crops and Products, 2012; 37(1): 69–75.

[2] Zava D T, Dollbaum C M, Blen M. Estrogen and progestin bioactivity of foods, herbs, and spices. Experimental Biology and Medicine, 1998; 217(3): 369–378.

[3] Mapiye C, Aalhus J L, Turner T D, Vahmani P, Baron V S, McAllister T A, Block H C, Uttaro B, Dugan M E R. Inclusion of sunflower seed and wheat dried distillers' grains with solubles in a red clover silage-based diet enhances steers performance, meat quality and fatty acid profiles. Animal, 2014; 8(12): 1999–2010.

[4] Beck V, Rohr U, Jungbauer A. Phytoestrogens derived from red clover: An alternative to estrogen replacement therapy? The Journal of Steroid Biochemistry Molecular Biology, 2005; 94(5): 499–518.

[5] Ma Y. Seed coating with beneficial microorganisms for precision agriculture. Biotechnology Advances, 2019; 37(7): 107423.

[6] Yang D B, Wang N, Yan X J, Shi J, Zhang M, Wang Z, et al. Microencapsulation of seed-coating tebuconazole and its effects on physiology and biochemistry of maize seedlings. Colloids and Surfaces B:Biointerfaces, 2014; 114: 241–246.

[7] Ryu C M, Kima J, Choi O, Kima S H, Park C S. Improvement of biological control capacity of *Paenibacillus polymyxa* E681 by seed pelleting on sesame. Biological Control, 2006; 39(3): 282–289.

[8] Qiu Y, Amirkhani M, Mayton H, Chen Z, Taylor A G. Biostimulant seed coating treatments to improve cover crop germination and seedling growth. Agronomy, 2020; 10(2): 154.

[9] Hou Z F, Qiu Y, Chen Z, Liu H Y, Guo F. Mi L K. 2020. Optimization of process parameters of pelletizer for agropyron seeds under vibration force field. INMATEH-Agricultural Engineering, 2020; 60(1): 299–308.

[10] Zhou D Q, Fan Y L, Deng G R, He F G, Wang M L. A new design of sugarcane seed cutting systems based on machine vision. Computers and Electronics in Agriculture, 2020; 175: 105611.

[11] Kussul N, Lavreniuk M, Skakun S, Shelestov A. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geoscience and Remote Sensing Letter, 2017; 14(5): 778–782.

[12] Zhang J, Qu M Z, Gong Z Y, Cheng F. Online double-sided identification and eliminating system of unclosed-glumes rice seed based on machine vision. Measurement, 2022; 187: 110252.

[13] Bai J Q, Hao F Q, Cheng F H, Li C G. Machine vision-based supplemental seeding device for plug seedling of sweet corn. Computers and Electronics Agriculture, 2021; 188: 106345.

[14] Ansari N, Ratri S S, Jahan A, Ashik-E-Rabbani M, Rahman A. Inspection of paddy seed varietal purity using machine vision and multivariate analysis. Journal of Agriculture and Food Research, 2021; 3: 100109.

[15] Przybylo J, Jablonski M. Using deep convolutional neural network for oak acorn viability recognition based on color images of their sections. Computers and Electronics in Agriculture, 2019; 156: 490–499.

[16] Sau S, Ucchesu, M, Dondini L, De Franceschi P, D'hallewin G, Bacchetta G. Seed morphometry is suitable for apple-germplasm diversity-analyses. Computers and Electronics in Agriculture, 2018; 151: 118–125.

[17] Liu W, Zhou Z Y, Xu X L, Gu Q Y, Zou S S, He W Z, et al. Evaluation method of rowing performance and its optimization for UAV-based shot seeding device on rice sowing. Computers and Electronics in Agriculture, 2023; 207: 107718.

[18] Hu Z W, Yang H, Lou T T, Yan H W. Concurrent channel and spatial attention in Fully Convolutional Network for individual pig image segmentation. Int J Agric & Biol Eng, 2023; 16(1): 232–242.

[19] Zhu H F, Yang L H, Fei J W, Zhao L G, Han Z Z. Recognition of carrot appearance quality based on deep feature and support vector machine. Computers and Electronics in Agriculture, 2021; 186: 106185.

[20] Jermsittiparsert K, Abdurrahman A, Siriattakul P, Sundeeva L A, Hashim W, Rahim R, et al. Pattern recognition and features selection for speech emotion recognition model using deep learning. International Journal of Speech Technology, 2020; 23: 799–806.

[21] Tu K L, Wen S Z, Cheng Y, Zhang T T, Pan T, Wang J, et al. A non-destructive and highly efficient model for detecting the genuineness of maize variety 'JINGKE 968' using machine vision combined with deep learning. Computers and Electronics in Agriculture, 2021; 182: 106002.

[22] Kamilaris A., Prenafeta-Boldú F X. Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 2018; 147: 70–90.

[23] Pearline S A, Kumar V S, Harini S. A study on plant recognition using conventional image processing and deep learning approaches. Journal of Intelligent & Fuzzy Systems, 2019; 36(3): 1997–2004.

[24] Loddo A. , Loddo M. , Ruberto C. A novel deep learning based approach for seed image classification and retrieval. Computers and Electronics in Agriculture, 2021; 187: 106269.

[25] Wang Y, Song S R. Variety identification of sweet maize seeds based on hyperspectral imaging combined with deep learning. Infrared Physics & Technology, 2023; 130: 104611.

[26] Yu Z Y, Fang H, Zhangjin Q N, Mi C X, Feng X P, He Y. Hyperspectral imaging technology combined with deep learning for hybrid okra seed recognition. Biosystems Engineering, 2021; 212: 46–61.

[27] Suo R, Gao F F, Zhou Z X, Fu L S, Song Z Z, Dhupia J, et al. Improved multi-classes kiwifruit recognition in orchard to avoid collisions during robotic picking. Computers and Electronics in Agriculture, 2021; 182: 106052.

[28] Hou Z F, Zhang X W, Chen Z, Dai N Z, Ma X J, Liu M. Design and experiment of identification and detection system for pelleted coated seeds. Transactions of the Chinese Society for Agricultural Machinery, 2022; 53(06): 62–69, 183.

[29] Zhang X W, Hou Z F, Xuan C Z. Design and experiment of recognition system for coated red clover seeds based on machine vision. INMATEH-Agricultural Engineering, 2022; 66(1): 62–72.

[30] Zhang X W, Hou Z F, Dai N Z. Design and experiment of a new rotary coating machine based on LabView. INMATEH-Agricultural Engineering, 2021; 65(3): 91–100.

[31] Zhang X W, Xuan C Z, Xue J, Chen B y, Ma Y H. SR-YOLO: A high-precision, lightweight model for sheep face recognition on the mobile end. Animals, 2023; 13: 1824.

[32] Wu D H, Lv S C, Jiang M, Song H B. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate recognition of apple flowers in natural environments. Computers and Electronics in Agriculture, 2020; 178: 105742.

[33] Han G J, He M, Zhao F, Xu Z P, Zhang M, Qin L. Insulator detection and damage identification based on improved lightweight YOLOv4 network. Energy Reports, 2021; 7(7): 187–197.

[34] Chen F E, Liang X M, Chen L H, Liu B Y, Lan Y B. Novel method for real-time recognition and tracking of pig body and its different parts. Int J Agric & Biol Eng, 2020; 13(6): 144–149.

[35] Yang B H, Gao Z W, Gao Y, Zhu Y. Rapid detection and counting of wheat ears in the field using YOLOv4 with attention module. Agronomy, 2021; 11(6): 1202.

[36] Xu D, Wu Y. Improved YOLO-V3 with DenseNet for multi-scale remote sensing target detection. Sensors, 2020; 20: 4276.

[37] Zhang X W, Xuan C Z, Ma Y H, Su H, Zhang M G. Biometric facial identification using attention module optimized YOLOv4 for sheep. Computers and Electronics in Agriculture, 2022; 203: 107452.

[38] Zhang X D, Kang X, Feng N N, Liu G. Automatic recognition of dairy cow mastitis from thermal images by a deep learning detector. Computers and Electronics in Agriculture, 2020; 178: 105754.

[39] Wu J, Tang T, Chen M, Wang Y, Wang K. A study on adaptation lightweight architecture based deep learning models for bearing fault diagnosis under varying working conditions. Expert Systems with Applications, 2020; 160: 113710.