

## Article

# Real-Time Detection Technology of Corn Kernel Breakage and Mildew Based on Improved YOLOv5s

Mingming Liu <sup>1,2</sup>, Yinzeng Liu <sup>2</sup>, Qihuan Wang <sup>1</sup>, Qinghao He <sup>1</sup> and Duanyang Geng <sup>1,\*</sup>

<sup>1</sup> School of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo 255000, China; z2018007@sdaeu.edu.cn (M.L.); 21403010276@stumail.sdut.edu.cn (Q.W.); 22503060002@stumail.sdut.edu.cn (Q.H.)

<sup>2</sup> Mechanical and Electronic Engineering College, Shandong Agriculture and Engineering University, Jinan 250100, China; z2020024@sdaeu.edu.cn

\* Correspondence: dygxt@sdut.edu.cn; Tel.: +86-136-6864-1238

**Abstract:** In order to solve low recognition of corn kernel breakage degree and corn kernel mildew degree during corn kernel harvesting, this paper proposes a real-time detection method for corn kernel breakage and mildew based on improved YOLOv5s, which is referred to as the CST-YOLOv5s model algorithm in this paper. The method continuously obtains images through the discrete uniform sampling device of corn kernels and generates whole corn kernels, breakage corn kernels, and mildew corn kernel dataset samples. We aimed at the problems of high similarity of some corn kernel features in the acquired images and the low precision of corn kernel breakage and mildew recognition. Firstly, the CBAM attention mechanism is added to the backbone network of YOLOv5s to finely allocate and process the feature information, highlighting the features of corn breakage and mildew. Secondly, the pyramid pooling structure SPPCPSC, which integrates cross-stage local networks, is adopted to replace the SPPF in YOLOv5s. SPP and CPSC technologies are used to extract and fuse features of different scales, improving the precision of object detection. Finally, the original prediction head is converted into a transformer prediction head to explore the prediction potential with a multi-head attention mechanism. The experimental results show that the CST-YOLOv5s model has a significant improvement in the detection of corn kernel breakage and mildew. Compared with the original YOLOv5s model, the average precision (AP) of corn kernel breakage and mildew recognition increased by 5.2% and 7.1%, respectively, and the mean average precision (mAP) of all kinds of corn kernel recognition is 96.1%, and the frame rate is 36.7 FPS. Compared with YOLOv4-tiny, YOLOv6n, YOLOv7, YOLOv8s, and YOLOv9-E detection model algorithms, the CST-YOLOv5s model has better overall performance in terms of detection accuracy and speed. This study can provide a reference for real-time detection of breakage and mildew kernels during the harvesting process of corn kernels.

**Keywords:** corn kernels; breakage; mildew; YOLOv5s; CBAM; SPPCPSC; transformer



**Citation:** Liu, M.; Liu, Y.; Wang, Q.; He, Q.; Geng, D. Real-Time Detection Technology of Corn Kernel Breakage and Mildew Based on Improved YOLOv5s. *Agriculture* **2024**, *14*, 725.

<https://doi.org/10.3390/agriculture14050725>

Received: 5 March 2024

Revised: 2 May 2024

Accepted: 6 May 2024

Published: 7 May 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Corn is the main grain crop in China, with the largest planting area and yield among the three major grain crops, and it is an indispensable and important food resource, playing a key role in ensuring national food security. In recent years, the country has strongly supported and promoted the development of the intelligent agricultural equipment industry. However, due to the influence of corn varieties, moisture content, mechanized harvesting and other factors, the intelligence level of corn grain harvesters is still low. Real-time automatic and accurate access to corn harvesting process kernel breakage and mildew degree is the key link of corn kernel harvester intelligence through real-time access to corn kernel harvesting process kernel breakage and adaptive adjustment of the harvester related operating parameters, to reduce the rate of corn kernel breakage and to provide an intelligent decision-making basis. At the same time, obtaining the mildew of corn

grains during corn harvest and constructing the distribution map of corn mildew degree in regional plots can provide support for grain reserves and scientific planting precision management. Therefore, it is particularly important to explore a method that can detect corn grain breakage and mildew in real-time, automatically and accurately.

With the development of computer technology, image processing and machine learning methods have been widely applied in the detection of grain appearance quality [1–3]. Li et al. [4] achieved the recognition of broken corn seeds by extracting the area features and color features of corn seeds and using the SSA-BP optimization algorithm. Cui et al. [5] proposed a method for detecting corn seed breakage based on machine vision, extracted 16 morphological characteristics of corn seeds, and created an SVM recognition model of corn seeds. The recognition precision of corn seed breakage reached 95%. Chen et al. [6] proposed a rice impurity and rice kernel breakage classification and recognition method based on machine vision technology, using Retinex algorithm to enhance the original image, setting thresholds for the two channels of hue and saturation of the HSV color model for image segmentation and combining with the shape features to obtain the classification and recognition results, with the recognition rate of rice kernel breakage being 84.74%, and that of stem impurities being 86.92%. Zhu et al. [7] developed a real-time detection algorithm for broken corn kernels based on the OpenCV vision library, which realized the detection and counting of broken corn. Although traditional machine learning methods have made some progress in the detection of grain appearance quality, traditional methods rely on manual intervention to extract and select features, and are easily affected by factors such as lighting and complex backgrounds, resulting in poor stability of detection results.

In recent years, with the continuous development of deep learning algorithms, this technology has also been applied by researchers to the field of grain appearance quality detection. Fan et al. [8] developed a rice appearance quality detection device that pre-processes images using the Watershed algorithm and Otus adaptive threshold function. Convolutional neural networks were used to detect rice appearance varieties, with a recognition precision of 92.3%. Niu et al. [9] proposed a corn seed variety detection model E\_CBAM\_MobileNetV2 based on the MobileNetV2 mode, the CBAM attention mechanism was introduced, and the recognition precision was 98.18%. Pan et al. [10] constructed a ImCascade R-CNN deep learning model for wheat grain integrity detection, and the average precision of wheat grain shape parameter detection was 90.2%. Wang et al. [11] constructed an online real-time detection system for wheat Fusarium head blight grains based on the U-Net deep learning model. The detection line completed the transmission and rapid recognition of wheat kernel by grain, and the recognition precision reached 95.78%. Zhang et al. [12] developed an online corn single-seed detection and sorting device, and the precision of broken corn seed sorting was higher than 89%. Wu et al. [13] developed an online monitoring system based on Mask R-CNN for grain impurity content and breakage rate of rice, and the recognition precision of whole rice grain was 97.62%, and the recognition precision of broken rice grain was 93.67%. LI et al. [14] constructed a corn breakage detection algorithm based on YOLOv4-Tiny, which achieved the detection of broken corn samples. In a relatively ideal environment, the precision of corn breakage detection is about 93%.

In summary, scholars at home and abroad have performed a lot of research on grain appearance quality detection by using deep learning technology; however, there is currently limited research on real-time detection of corn kernel breakage and mildew during the corn kernel harvesting process. In this paper, a real-time detection method of corn kernel breakage and mildew based on improved YOLOv5s was proposed for low recognition caused by the variety of corn kernel morphology, intense kernel movement, and complex and changing environment during corn kernel harvesting. The main work of this paper is as follows. Firstly, the CBAM (Convolutional Block Attention Module) mechanism is added to the backbone network of YOLOv5s to finely allocate and process the feature information, highlighting the features of corn breakage and mildew. Secondly, the pyramid pooling structure SPPCPS (Spatial Pyramid Pooling Fast Cross Stage Partial Concat),

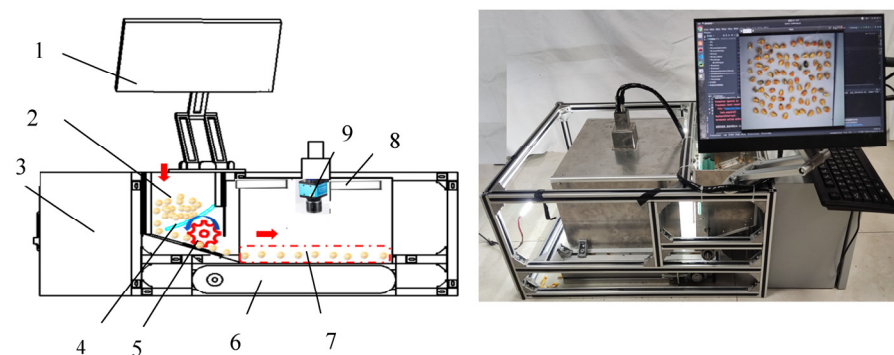
which integrates cross-stage local networks, is adopted to replace the SPPF (Spatial Pyramid Pooling-Fast) in YOLOv5s. SPP (Spatial Pyramid Pooling) and CPSC (Cross-Scale Partial Connections) technologies are used to extract and fuse features of different scales, improving the precision of object detection. Finally, the original prediction head is converted into a transformer prediction head to explore the prediction potential with a multi-head attention mechanism and realize real-time detection of corn kernel breakage and mildew under a complex background. Compared with detection model algorithms such as YOLOv4-tiny, YOLOv6n, YOLOv7, YOLOv8s and YOLOv9-E [15–19], the effectiveness of the improved model was verified.

## 2. Materials and Methods

In this chapter, the materials and methods of corn kernel breakage and mildew detection tests are described. The corn kernel image acquisition device is described in Section 2.1. The corn kernel sample image dataset is described in Section 2.2. In Section 2.3, the construction of the corn kernel breakage and mildew recognition model is described.

### 2.1. Corn Kernel Image Acquisition Device

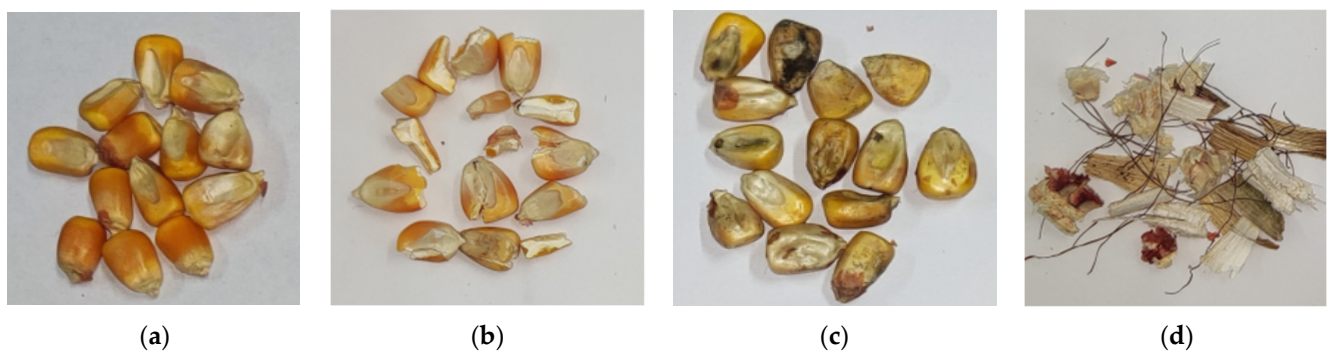
The quality of the original image has a significant impact on the detection result. High-quality original images can help improve the learning and generalization abilities of deep learning models [20]. Therefore, the research group developed a discrete uniform sampling device of corn kernels, which is used to realize continuous sampling and single-layer uniform distribution of mechanically harvested corn kernels and to avoid the impact of kernel superposition and adhesion phenomenon on the detection results. The schematic diagram of its structure is shown in Figure 1. The discrete uniform sampling device of corn kernels is composed of a host computer, feeding hopper, control box, current limiting plate, outer groove wheel uniform distributor, corn kernel conveyor belt, strip light source, CCD camera and other parts. The sampled corn kernels collected by the machine will enter the feeding hopper through the chute, and the current limiting plate will ensure that the corn kernels will gently enter the outer groove wheel uniform distributor, and the outer groove wheel uniform distributor will set up U-shaped holes in the circumference direction. During the rotation of the outer groove wheel uniform distributor, the corn kernels stacked in continuous layers will be separated into discrete states to realize the discrete single-layer distribution of the corn, and then the discrete single-layer corn kernels will be placed by the conveyor belt. The image is transferred to the image acquisition and detection area, and the image is collected in real-time by a CCD camera (Hikvision, MV-CA050-10C, Hangzhou, China, Resolution (1280 × 1024) pixels, Rate is 32 FPS), and the image is transmitted to the host computer for detection and processing.



**Figure 1.** Discrete uniform sampling device of corn kernels. 1. host computer; 2. feeding hopper; 3. control box; 4. current limiting plate; 5. outer groove wheel uniform distributor; 6. corn kernel conveyor belt; 7. image acquisition and detection area; 8. strip light source; 9. CCD camera.

## 2.2. Corn Kernel Sample Image Dataset

Selecting “Xianyu 335” corn kernels (with a moisture content of 28.1%) harvested from experimental fields in Linzi District, Zibo City, Shandong University of Technology in October 2023 as samples. The “Xianyu 335” corn variety is one of the widely planted corn varieties in China, and its planting area in China is about 24 million hectares. Figure 2 shows the samples of machine-harvested corn kernels, which mainly contain whole corn kernels, breakage corn kernels, mildew corn kernels and impurities. Breakage corn kernel types were mainly classified into crown breakage, radicle breakage, cracked breakage and crushed breakage, as shown in Figure 3. Mildew corn kernel types were mainly classified into complete mildew, lumpy mildew and spotted mildew, as shown in Figure 4. There are impurities in the corn kernel samples as shown in Figure 2d, because these impurities are significantly different from whole corn kernels, breakage corn kernels, and mildew corn kernels. This does not have a significant impact on target classification and recognition. Therefore, this article will not consider them.



**Figure 2.** Samples of machine-harvested corn kernels. (a) Whole corn kernels; (b) Breakage corn kernels; (c) Mildew corn kernels; (d) Impurities.



**Figure 3.** Type of breakage corn kernels. (a) Crown broken corn kernels; (b) Radicle breakage corn kernels; (c) Cracked breakage corn kernels; (d) Crushed breakage corn kernels.

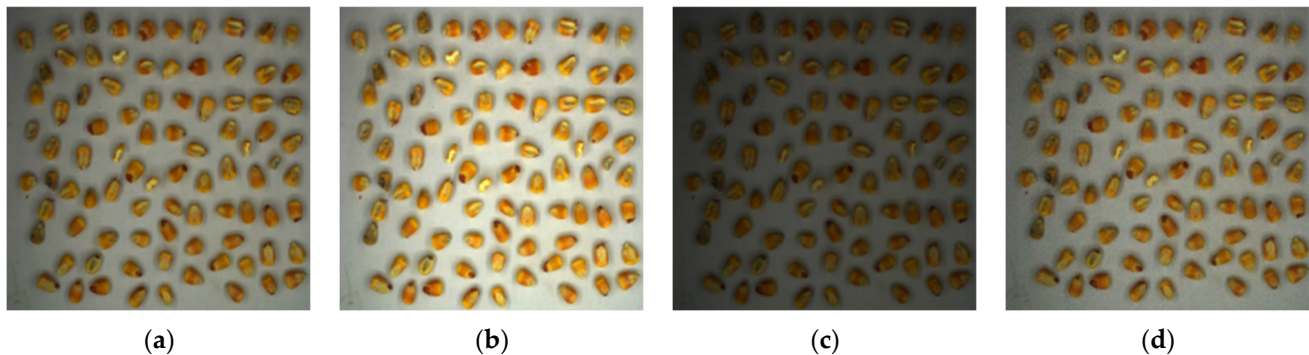


**Figure 4.** Type of mildew corn kernels. (a) Complete mildew corn kernels; (b) Lumpy mildew corn kernels; (c) Spotted mildew corn kernels.

Considering the adverse factors such as complex environment, changes in light conditions and multiple noise points during corn kernel harvesting, a series of transformations and processing such as brightness (increase by 50%, decrease by 50%) and Gaussian noise was carried out on the acquired images in order to enhance the generalization ability and robustness of the model and avoid the risk of overfitting. In total, 528 images were obtained (each corn kernel set package contained an average of about 60 individual corn kernels), and the enhanced dataset images are shown in Figure 5. The image dataset is divided into a training set, validation set and test set according to the ratio of 8:1:1. The Labelling tool (Labelling software, Version: 1.8.6, the open-source data labeling tool created by Tzutalin



with the help of dozens of contributors, was used to annotate the original image, and the change save format was set to PASCAL VOC mode to generate an XML annotation file, <https://github.com/HumanSignal/labelImg>, accessed on 20 April 2023) was used to label the whole, breakage and mildew corn kernels in the original images, respectively. LabelImg software generates corresponding label files as XML files, which cannot be directly recognized by the YOLOv5 algorithm. The XML file needs to be converted into a TXT file through script code, which contains the target category, center point coordinate information, and the length and width information of the annotation box. The image dataset is shown in Table 1.



**Figure 5.** Images of the dataset after data enhancement. (a) Original images; (b) Brightness value increased by 50%; (c) Brightness value decreased by 50%; (d) Gaussian noise.

**Table 1.** The image dataset.

Dataset	Whole Corn Kernels	Breakage Corn Kernels	Mildew Corn Kernels
Training set	13,783	5907	5904
Validation set	1494	747	726
Test set	1415	730	735
Total (sheets)	16,692	7384	7365

### 2.3. Recognition Model Construction

This section introduces the construction of the corn kernel breakage and mildew recognition model. Aiming at the requirements of corn kernel breakage and mildew detection, a CST-YOLOv5s recognition model algorithm is proposed based on the YOLOv5s model. The improvement of the CST-YOLOv5 network structure is described in Section 2.3.1. The CBAM attention mechanism is described in Section 2.3.2. The SPPCPSC pyramid pool structure is described in Section 2.3.3. The transformer predictable header is described in Section 2.3.4.

#### 2.3.1. YOLOv5s Algorithm

YOLOV5 is a single-stage target detection network. YOLOV5 has a more balanced performance in reasoning speed, lightweight, reliability, and stability, and is more suitable for deployment in embedded equipment [21]. According to the depth and width of the algorithm model, it is divided into five versions, which are YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x. This strengthens the detection accuracy, with the corresponding time also becoming increased. The performance parameters of different versions of the YOLOv5 series are compared in Table 2. Although the detection time of YOLOv5n is the lowest, the number of residual structures is small, resulting in a decrease in the detection accuracy, and the excessive detection time of YOLOv5m, YOLOv5l, YOLOv5x residual structure will increase the test time. In order to determine the baseline model suitable for the detection tasks of corn kernel breakage and mildew, the YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x models were respectively used for pre-training

of the dataset in this paper (training cycle 300). The comparison of detection performance parameters is shown in Table 2. On the premise of comprehensively balancing model size and recognition accuracy, the detection time will be increased. Combined with the detection accuracy, real-time performance, lightweight, field operation environment and other factors of corn kernel breakage and mildew [22–25], this paper proposes the design of a new recognition model algorithm CST-YOLOv5s based on YOLOv5s.

**Table 2.** Comparison of performance parameters of different versions of YOLOv5 series.

Model	Size (Pixels)	$mAP_{0.5}\%$	$mAP_{0.5-0.95}\%$	Parameters/ $\times 10^6$ M	Model Size/MB	FPS
YOLOv5n	640	0.822	0.763	1.7	3.5	48.5
YOLOv5s	640	0.926	0.872	7.0	13.5	42.6
YOLOv5m	640	0.931	0.882	20.9	40	35.6
YOLOv5l	640	0.936	0.887	46.1	88.3	28.7
YOLOv5x	640	0.941	0.891	86.2	164	20.7

The YOLOv5s algorithm mainly includes four parts: input, backbone, neck and head. Mosaic data enhancement is used at the input side to randomly scale, randomly cut, and randomly arrange the original data to enrich the dataset and thus enhance the generalization of the model. After images are sent to the neural network through the input end, feature extraction is carried out in the backbone part to obtain three effective feature layers. The feature pyramid network in the neck part enhances semantic information through top-down up-sampling. In the neck part, the path aggregation network adopts bottom-up down sampling to strengthen the positioning information. The two networks fuse the features from different paths respectively, so as to obtain the feature map with rich information. The Prediction Head part contains three prediction branches, which use the extracted feature information to predict the target of different sizes and obtain the category, confidence and location information of the predicted target [26].

Due to the variety of corn kernel morphology, intense kernel movement, and complex and changing environment, in order to further improve the precision of corn kernel breakage and mildew detection, this study improved the backbone, neck, and head of the original YOLOv5s model. Firstly, the CBAM attention mechanism is added to the backbone network of YOLOv5s to finely allocate and process the feature information, highlighting the features of corn breakage and mildew, as shown in the red rectangular box in Figure 6. Secondly, the pyramid pooling structure SPPCPSC, SPP and CPSC technologies are used to extract and fuse features of different scales, improving the precision of object detection, as shown in the green rectangular box in Figure 6. Finally, the original prediction head is converted into a transformer prediction head to explore the prediction potential with a multi-head attention mechanism, as shown in the blue rectangular box in Figure 6. The network structure diagram of the CST-YOLOv5s model is shown in Figure 6.

### 2.3.2. CBAM Attention Mechanism

Due to unfavorable factors such as impurities, light-blocking shadows, and high similarity between corn embryo surface and breakage surface in the original images of machine-harvested corn kernels, the precision of target detection is affected. The attention mechanism is based on the characteristics of human attention, making deep learning networks pay more attention to the areas that need attention, which greatly helps to improve the performance of neural networks [27]. The CBAM attention mechanism is introduced into the backbone of YOLOv5s to enhance the backbone layer's ability to detect target features, which helps the model pay more attention to breakage and mildew features [24]. CBAM is a lightweight attention mechanism module, which can effectively integrate the information of both channel and spatial dimensions, and help the network to obtain a large amount of location and detail information of the target area. The network structure of the CBAM attention mechanism is shown in Figure 7. Firstly, the feature map performs global maximum pooling and global average pooling operations in the channel

attention module to retain context information and aggregate spatial features. Secondly, the shared multi-layer perceptron is adopted to fuse the information of the two and generate the feature map of channel attention by activating the function [25]. Finally, the feature map information of channel attention is extracted using the maximum average pooling operation, and it is concatenated and convolved to obtain the final output feature map of the CBAM module. The calculation formula [28] is:

$$M_c = S_m(M_{lp}(P_{\max}(F))) + S_m(M_{lp}(P_{\text{avg}}(F))) \quad (1)$$

$$F' = M_c F \quad (2)$$

$$M_s = S_m(f_{7 \times 7}(P_{\max}(F'); P_{\text{avg}}(F'))) \quad (3)$$

$$F'' = M_s F' \quad (4)$$

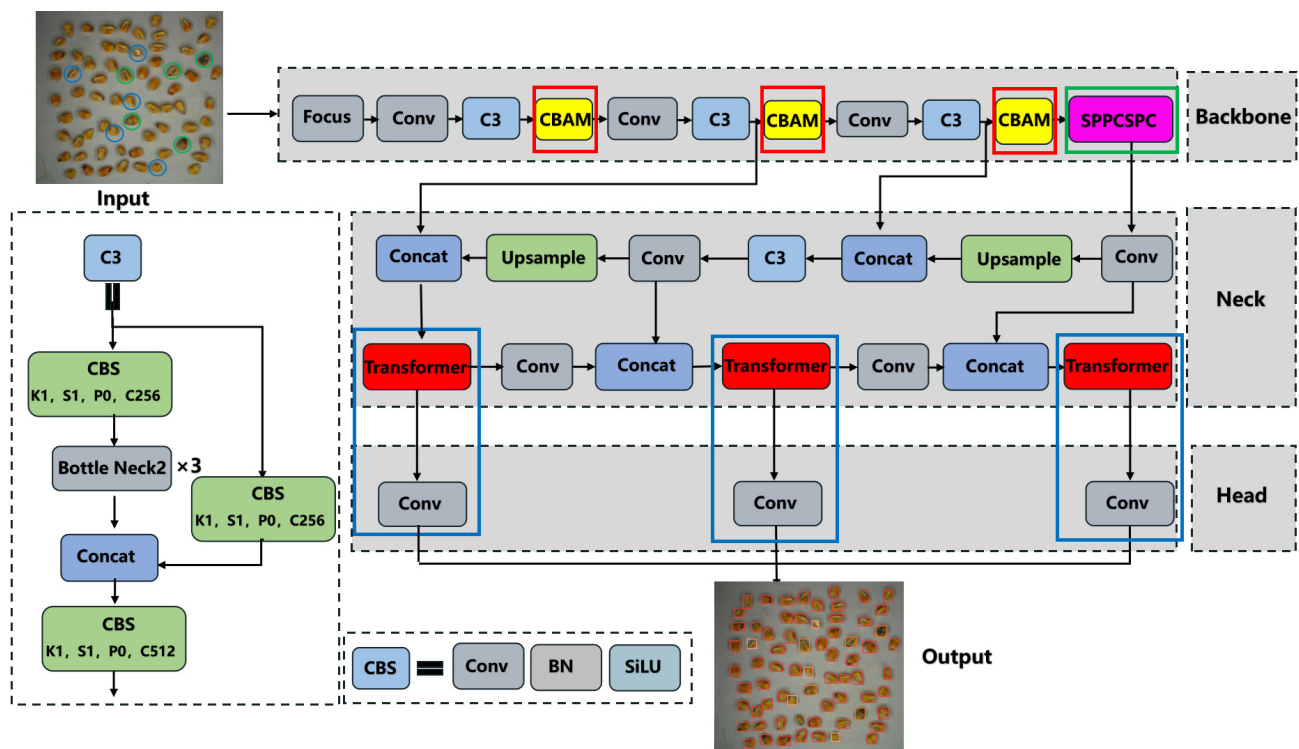


Figure 6. The network structure diagram of the CST-YOLOv5s model.

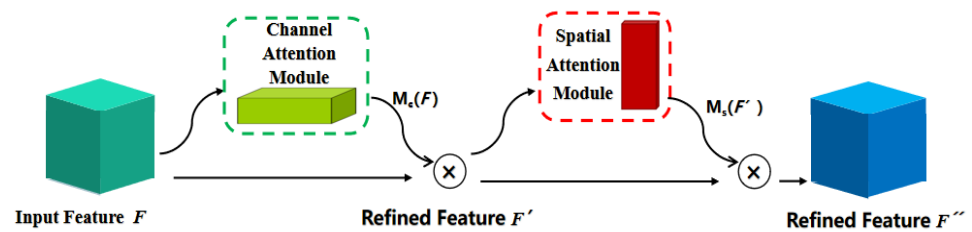


Figure 7. The network structure of the CBAM attention mechanism.

In the formula [29]:  $F$  is the input feature;  $F'$  is the feature of  $F$  input feature after processing by channel attention mechanism;  $F''$  is the feature of  $F'$  input feature after processing by spatial attention mechanism;  $M_c$  is the channel attention weight;  $M_s$  is the spatial attention weight;  $S_m$  is Sigmoid activation function;  $M_{lp}$  is a multi-layer perceptron;  $P_{\max}$  is the maximum pooling;  $P_{\text{avg}}$  is average pooling;  $f_{7 \times 7}$  is a  $7 \times 7$  filtering convolution operation.

### 2.3.3. SPPCSPC Pyramid Pooling Structure

A spatial pyramid pool SPP is used to realize the fusion of feature maps of different scales, construct the correlation between objects of different scales, and solve the information description problem of regional correlation. In the process of corn kernel breakage and mildew detection, the background of real-time image is complicated, and some features are similar, which increases the difficulty of real-time detection and analysis. This paper proposes to replace SPPF in YOLOv5 with SPPCSPC. The SPPCSPC module has four different scales of maximum pooling:  $1 \times 1$ ,  $5 \times 5$ ,  $9 \times 9$ , and  $13 \times 13$ . The four different scales of maximum pooling have four perceptual horizons to distinguish small targets from large targets, and improve the network's expression ability and perception ability [30]. SPPCSPC combines feature pyramid pooling and small residual structure and uses SPP and CPSC technology to extract and fuse features of different scales to improve the precision of target detection. The SPPCSPC network structure diagram is shown in Figure 8.

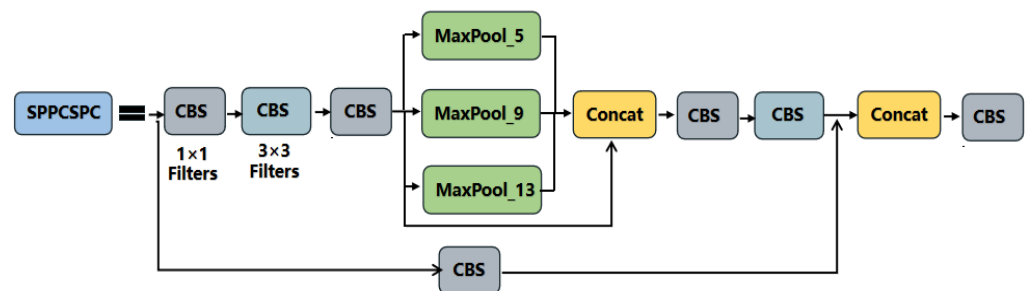


Figure 8. The network structure of SPPCSPC.

### 2.3.4. Transformer Prediction Heads

The transformer prediction head module has achieved a good application effect in deep learning in many fields [31]. It has the advantages of strong learning long-distance dependence, strong multi-modal fusion ability, and strong model interpretability [32]. The overall structure of the transformer model consists of input, encoder, decoder and output. The network structure of the transformer prediction head is shown in Figure 9.

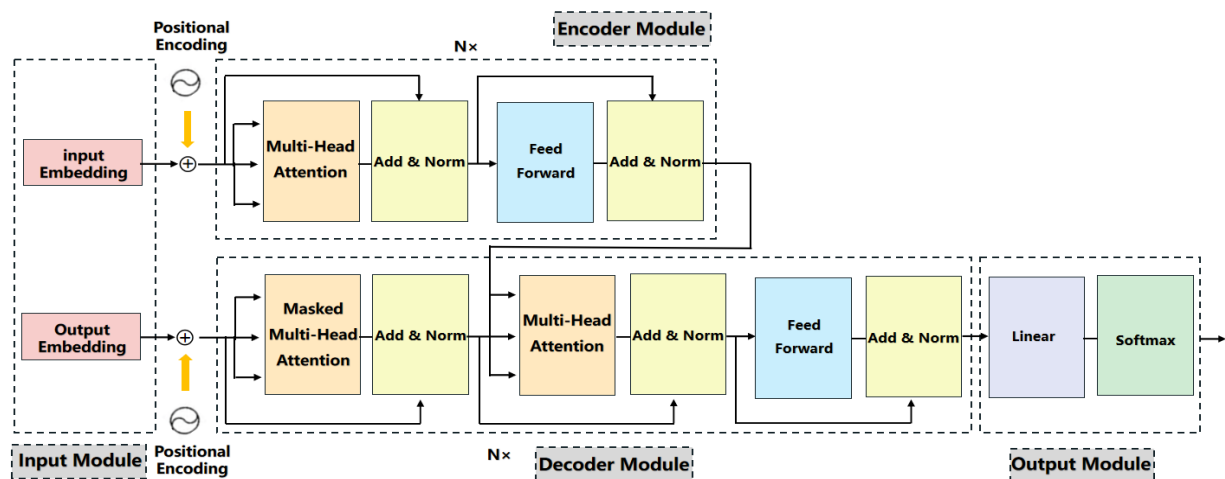


Figure 9. The network structure of transformer prediction head.

The encoder includes multi-head attention, residual connection and layer normalization, and a feed-forward network. The multi-head attention mechanism is composed of multiple self-attention mechanisms, which allows the model to focus on information from different locations at the same time. By splitting the original input vector into multiple heads, each head can independently learn different attention weights, thereby enhancing the model's ability to focus on different parts of the input sequence [33–35]. We can replace



some CSP modules in the neck with transformer prediction head modules to achieve prediction potential with a multi head attention mechanism, capture global information and sufficient background information, and accurately locate breakage and mildew corn kernel in high-density scenes.

### 3. Results and Analysis

In this chapter, the results and analysis of corn kernel breakage and mildew detection algorithm are described. The test environment and parameter settings of corn kernel breakage and mildew are described in Section 3.1. The performance evaluation metrics of corn kernel breakage and mildew detection models are described in Section 3.2. The comparison and analysis of corn kernel breakage and mildew test results are described in Section 3.3.

#### 3.1. Experimental Environment and Parameter Setting

This experiment was run on a Windows 11 system, equipped with an Intel Core I7-13700KF CPU (Intel, Santa Clara, CA, USA), with a main frequency of 3.4 GHz, 32 GB of memory, Nvidia Geforce RTX 4060 graphics card (ASUS, Taiwan, China), Unified Computing Device Architecture (CUDA) version 11.3, GPU accelerated CUDNN version 8.2.0, Python 3.8, and deep learning framework Pytorch 1.10.0. The initial learning rate is set to 0.01, the initial momentum value is 0.937, the weight attenuation coefficient is 0.0005, the image input size is 640 pixels  $\times$  640 pixels, the batch size is 32, and the training round is set to 300.

#### 3.2. Evaluation Metrics

In the test of the corn kernel breakage and mildew detection model, the accuracy ( $P$ ), recall rate ( $R$ ), average accuracy ( $AP$ ), the mean average Precision ( $mAP$ ) are used to evaluate the accurate performance of the model algorithm. The frame rate FPS (frames per second) is used to evaluate the model algorithm recognition speed. The specific calculation formula [21] is as follows:

$$P = \frac{T_P}{T_N + F_P} \times 100\% \quad (5)$$

$$R = \frac{T_P}{T_P + F_N} \times 100\% \quad (6)$$

In the formula:  $T_P$  is a positive sample and predicted is a positive sample;  $T_N$  is a negative sample and prediction is a negative sample;  $F_P$  is a negative sample predicted to be a positive sample;  $F_N$  is a positive sample and predicts a negative sample.

$$AP = \int_0^1 P(R) dR \quad (7)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP(i) \quad (8)$$

In the formula:  $AP$  is the area enclosed by the  $P$ - $R$  curve, which represents the relationship between precision and recall rate,  $mAP$  is the mean average accuracy of  $AP$  of all categories, where  $i$  represents the current category and  $n$  represents the number of categories. In this paper,  $n$  is taken as 3, that is, whole corn kernels, breakage corn kernels and mildew corn kernels.

#### 3.3. Experimental Comparison and Analysis

In this section, the comparison and analysis of corn kernel breakage and mildew detection algorithm are described. The ablation experiments analysis is described in Section 3.3.1 to verify the performance of CST-YOLOv5 proposed in this paper. The comparative test analysis of different model algorithms is described in Section 3.3.2. The

comparison and analysis of corn kernel breakage and mildew test results are described in Section 3.3.3.

### 3.3.1. Ablation Experiments

Through a series of ablation experiments, the effectiveness of the improved method was evaluated and verified. Based on the YOLOv5s model, the CBAM attention mechanism was added to the backbone part, the pyramid pooling structure SPPCPSC with the cross-stage local network was adopted, and the original prediction head was converted into a transformer prediction head for experimentation. The experiment conducted a comparative analysis using the same dataset, testing platform, and model parameters to evaluate the optimization effect of each improvement method on the final algorithm. The experimental results are shown in Table 3.

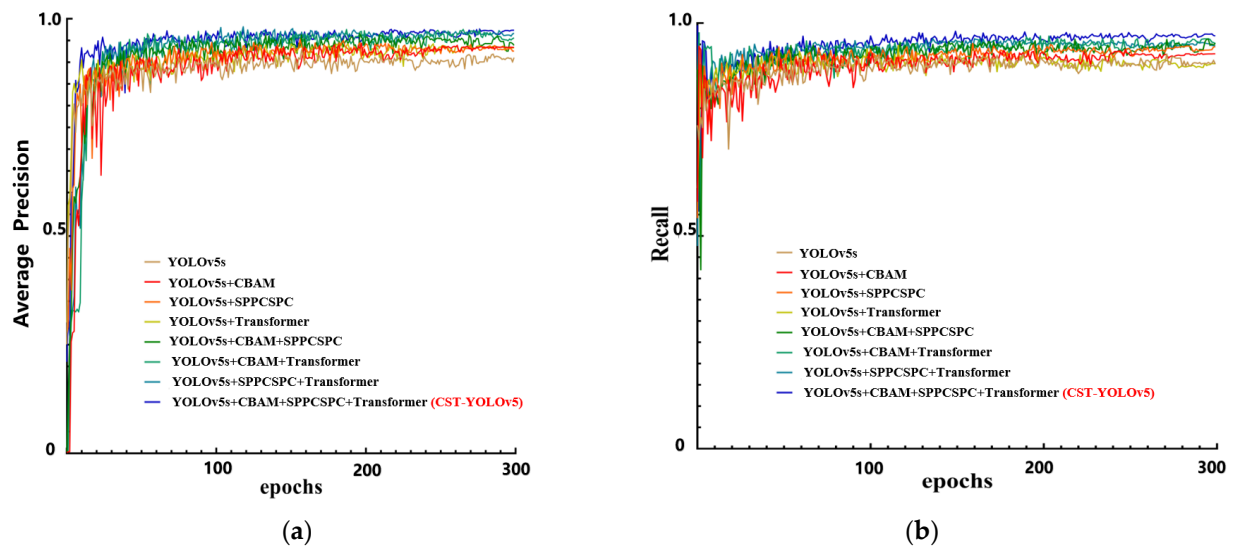
**Table 3.** Comparison of ablation experiments results.

Model				Average Precision AP/%			P/%	R/%	mAP/%	Model Size /MB
CBAM	SPPCPSC	Transformer	Whole Corn Kernels	Breakage Corn Kernels	Mildew Corn Kernels					
YOLOv5s	×	×	×	91.2	90.9	87.6	90.6	90.5	89.9	13.5
	√	×	×	93.4	91.8	91.6	93.5	91.9	92.3	13.7
	×	√	×	93.1	94.5	90.8	93.2	94.6	92.8	25.9
	×	×	√	93.3	90.4	92.1	93.4	90.5	92.0	13.6
	√	√	×	94.1	94.4	92.6	94.3	94.7	93.7	26
	√	×	√	95.8	95.1	93.5	96.0	95.4	94.8	13.7
	×	√	√	95.5	95.8	92.2	95.8	96.1	94.5	25.8
	√	√	√	97.5	96.1	94.7	97.2	97.5	96.1	26

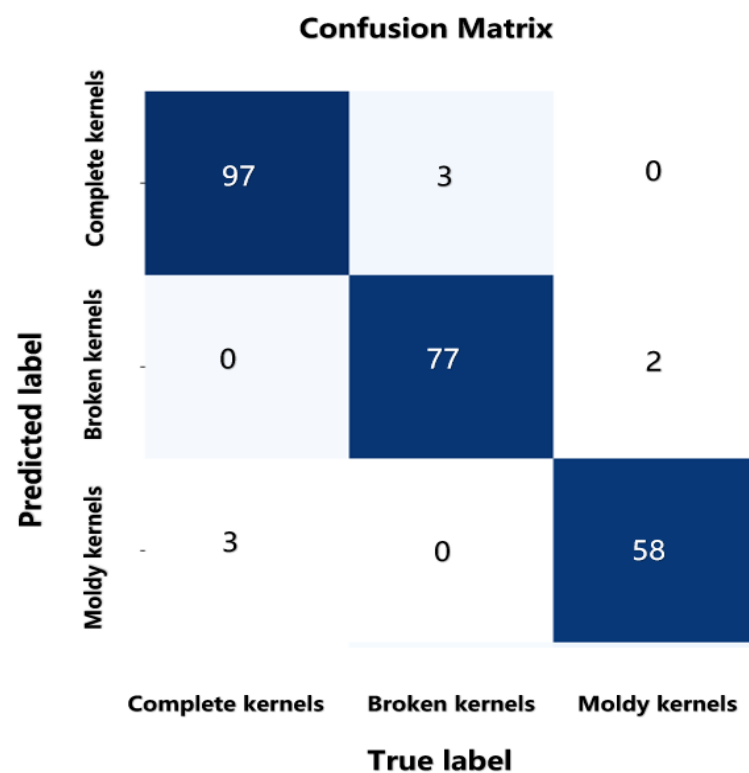
Note: “✓” indicates that the current model uses this structure or method; “×” indicates that the current model is not using this structure or method. The performance parameters of the CST-YOLOv5s algorithm model proposed in this paper are shown in bold font.

As can be seen from Table 3, adding CBAM, SPPCPSC and transformer to the original YOLOv5s network model can effectively improve network detection performance. On the premise of comprehensively balancing model size and recognition accuracy, this paper asserts that the CST-YOLOv5s (YOLOv5s + CBAM + SPPCPSC + Transformer) model has better performance than other models, and the CST-YOLOv5s model has a significantly improved detection effect on corn kernel breakage and mildew. Compared with the original YOLOv5s model, the average precision (AP) of corn kernel breakage and mildew recognition increased by 5.2% and 7.1%, respectively, and the mean average precision (mAP) of all kinds of corn kernel recognition is 96.1%. Figure 10 shows the comparison of average precision and recall curves of different optimization algorithm combinations, where (a) is the comparison of average precision curves and (b) is the comparison of recall curves. As can be seen from the figure, the CST-YOLOv5s model has higher precision than other models in terms of average precision and recall.

This study uses the confusion matrix to evaluate the classification model performance [36,37], which randomly selects 240 corn kernel samples for prediction, of which 100 grains are whole corn kernels, 80 are breakage corn kernels, and 60 are mildew corn kernels. The confusion matrix after the classification is completed is shown in Figure 11. The model’s predictive accuracy of all categories is high, and it can accurately and reliably identify and classify various types of corn kernels.



**Figure 10.** Comparison of average precision and recall curves of different optimization algorithm combinations. (a) Average precision curves of training set; (b) Recall curves of training set.



**Figure 11.** CST-YOLOv5s model confusion matrix.

### 3.3.2. Experimental Analysis of Different Model Algorithms

In order to further verify the effectiveness of the research method for corn kernel breakage and mildew detection, a comparative test was conducted between this research method and other YOLO model algorithms in the same environment. The comparative detection networks include YOLOv4-tiny, YOLOv6n, YOLOv7, YOLOv8s, and YOLOv9-E. The results of the identification effect comparison test are shown in Table 4.

**Table 4.** Comparison of experimental results of different model algorithms.

Model	Average Precision AP/%			<i>P</i> /%	<i>R</i> /%	<i>mAP</i> /%	Model Size /MB	FPS
	Whole Corn Kernels	Breakage Corn Kernels	Mildew Corn Kernels					
YOLOv4-Tiny	89.3	88	85.7	87.1	87.4	87.7	27.6	34.5
YOLOv6n	88.3	87.2	84.3	86.1	86.4	86.6	17.6	39.8
YOLOv7	95.2	93.8	92.9	95.5	92.2	94	71.2	20.9
YOLOv8s	95.1	95.7	92.4	95.4	95.9	94.4	50.6	25.2
YOLOv9-E	93.1	99.1	95.6	95.9	98.3	95.9	116	6.5
CST-YOLOv5s	97.5	96.1	94.7	97.2	97.5	96.1	26	36.7

As can be seen from Table 4, the precision *P*, recall *R*, and the mean average precision (*mAP*) values of the CST-YOLOv5s model proposed in the paper are 97.2%, 97.5%, and 96.1%, respectively, the model size is 26 MB, and the frame rate is 36.7 FPS. The CST-YOLOv5s model has certain advantages over other model algorithms, indicating that the CST-YOLOv5s model is able to extract the key information more attentively. By incorporating the CBAM attention mechanism, the characteristics of small size, irregularity and mildew are more prominent. SPP and CPSC technologies are fully utilized to extract and fuse features of different scales, reducing the interference of useless information on recognition performance. Meanwhile, the prediction potential of a multi-head attention mechanism is utilized to improve the precision of target detection. It can accurately identify corn kernel breakage and mildew in the complex background of corn harvest.

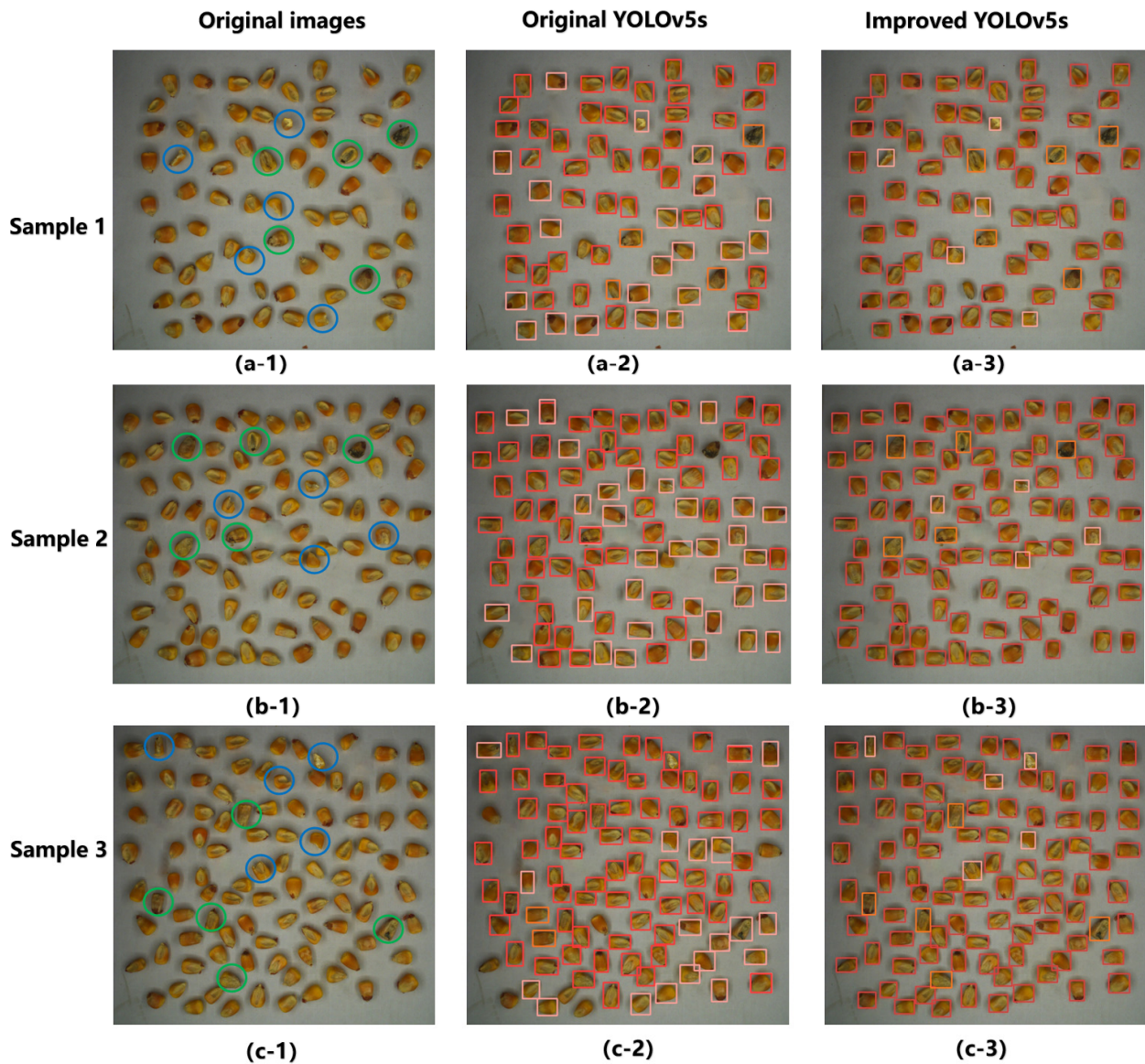
### 3.3.3. Improved Algorithm Detection Experiment

In order to further evaluate the performance of the CST-YOLOv5s model, the corn kernel set images were randomly selected in the test set to test the recognition effect, and compared with the original YOLOv5s model recognition results. Figure 12 is the detection effect map of three test sets which were randomly selected. In the first column of the corn kernel set images to be detected in Figure 12, breakage corn kernels are marked with blue circles and mildew corn kernels are marked with green circles. The second column in Figure 12 shows the detection results of the original YOLOv5s model, and the third column shows the detection results of the CST-YOLOv5s model. The detected whole corn kernels, breakage corn kernels, and mildew corn kernels are highlighted in red, pink, and orange rectangular boxes, respectively.

As can be seen from Figure 12, the recognition effect of the original YOLOv5s model on corn kernels was shown. In Figure 12(a-2), some small and irregular whole corn kernels were mistakenly identified as breakage kernels, and some spotted mildew corn kernels were mistakenly identified as whole corn kernels. In Figure 12(b-2), there was a serious phenomenon of missed recognition of mildew corn kernels, where the model was mistakenly identified as not being a sample of corn kernels. Similarly, there were also some small and irregular whole corn kernels that were mistakenly identified as breakage kernels. In Figure 12(c-2), some mildew corn kernels were mistakenly identified as whole corn kernels, while some small and irregular whole corn kernels were mistakenly identified as whole kernels. Compared with the CST-YOLOv5s model proposed in this paper, the detection effect is shown in Figure 12(a-3,b-3,c-3). It can be seen that the CST-YOLOv5s detection model has good recognition performance for various types of corn kernels, and there are basically no problems of missed detection or incorrect detection. The number of various types of corn kernels in the 66 original corn kernel set images of the test set were counted manually, in which the number of whole corn kernels was 1415, the number of breakage corn kernels was 730, and the number of mildew corn kernels was 735. By using the CST-YOLOv5s model proposed in this paper, the numbers of whole corn kernels, breakage corn kernels and mildew corn kernels were 1371, 697 and 689, respectively, and the detection and recognition precisions of whole corn kernels, breakage corn kernels and



mildew corn kernels were 96.9%, 95.5% and 93.7%, respectively. Therefore, combined with the above test results, a new recognition model algorithm CST-YOLOv5s based on YOLOv5s for corn kernel breakage and mildew in this paper can meet the detection needs.



**Figure 12.** Comparison of detection effects before and after YOLOv5s network improvement. (a-1) Original image of Sample 1; (a-2) Detection results of the original YOLOv5s model for Sample 1; (a-3) Detection results of the improved YOLOv5s for Sample 1; (b-1) Original image of Sample 2; (b-2) Detection results of the original YOLOv5s model for Sample 2; (b-3) Detection results of the improved YOLOv5s for Sample 2; (c-1) Original image of Sample 3; (c-2) Detection results of the original YOLOv5s model for Sample 3; (c-3) Detection results of the improved YOLOv5s for Sample 3; The first column is the original acquisition images, in which the blue circle marks the breakage corn kernels and the green circle marks the mildew corn kernels; the second column and the third column are the whole corn kernels, breakage corn kernels, and mildew corn kernels detected before and after the model improvement, which are highlighted in red, pink, and orange rectangular boxes, respectively.



#### 4. Conclusions

- (1) This paper takes real-time detection of corn kernel breakage and mildew during the harvesting of corn kernels as the research content. Aiming at low recognition caused by the variety of corn kernel morphology, intense kernel movement, and complex and changing environment during corn kernel harvesting, this paper proposes a new recognition model algorithm: CST-YOLOv5s based on YOLOv5s for corn kernel breakage and mildew. Firstly, the CBAM attention mechanism is added to the backbone network of YOLOv5s to finely allocate and process the feature information, highlighting the features of corn breakage and mildew. Secondly, the pyramid pooling structure SPPCPSC, SPP and CPSC technologies are used to extract and fuse features of different scales, improving the precision of object detection. Finally, the original prediction head is converted into a transformer prediction head to explore the prediction potential with a multi-head attention mechanism.
- (2) Using the same training and validation sets, multi-scale training was conducted on different improved algorithms and model algorithms. Through ablation experiments, it was shown that the CST-YOLOv5s model significantly improved the detection effect of corn kernel breakage and mildew. Compared with the original YOLOv5s model, the average precision (AP) of corn kernel breakage and mildew recognition increased by 5.2% and 7.1%, respectively, and the mean average precision (mAP) of all kinds of corn kernel recognition is 96.1%, and the frame rate is 36.7 FPS. Compared with YOLOv4-tiny, YOLOv6n, YOLOv7, YOLOv8s, and YOLOv9-E detection model algorithms, the CST-YOLOv5s model has better overall performance in terms of detection accuracy and speed. This study can provide a reference for real-time detection of breakage and mildew kernels during the harvesting process of corn kernels.
- (3) In future research, we will increase the number of corn variety samples and enrich the dataset of breakage and mildew corn kernels. At present, we only focus on the detection of corn kernels on one side and cannot judge the state of other surfaces. In the later stage, we will carry out research on the recognition of the whole surface of corn kernels to further improve the detection precision.

**Author Contributions:** Conceptualization, D.G. and M.L.; methodology, M.L.; software, Y.L.; validation, Q.W., Q.H. and Y.L.; formal analysis, Q.H.; investigation, M.L.; resources, Q.W.; data curation, Q.H.; writing—original draft preparation, M.L.; writing—review and editing, D.G.; visualization, M.L.; supervision, Y.L.; project administration, D.G.; funding acquisition, D.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key R&D Program of China, grant number 2021YFD2000502, the Natural Science Foundation of Shandong Province, grant number ZR2022ME064, and the Modern Agricultural Industrial System of Shandong Province, grant number SDAIT-02-12.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors on request.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Nie, S.; Ma, S.; Peng, Y.; Wang, W.; Li, Y. Research Progress of Rapid Optical Detection Technology and Equipment for Grain Quality. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 1–12.
2. Yang, S.; Cao, Y.; Zhao, H.; Fei, M. Research progress of rapid detection technology in grain mildew. *Cereals Oils* **2018**, *31*, 21–23.
3. Zhang, J.; Jin, Y.; Yin, J. Research Progress of Machine Vision Technology in Rice Quality Detection. *J. Chin. Cereals Oils* **2022**, *37*, 302–310.
4. Li, Z. Research on Maize Seed Quality Inspection Based on Machine Vision. Master's Thesis, Shijiazhuang Tidao University, Shijiazhuang, China, 2023.
5. Cui, X.; Zhang, P.; Zhao, J.; Xu, W.; Ma, W.; Jin, C. Study on Inspection of Corn Seed Breakage Based on Machine Vision. *J. Agric. Mech. Res.* **2019**, *41*, 28–33+84.

6. Chen, J.; Gu, Y.; Lian, Y.; Han, M. Online recognition method of impurities and broken paddy grains based on machine vision. *Trans. Chin. Soc. Agric. Eng.* **2018**, *34*, 187–194.
7. Zhu, X.; Du, Y.; Chi, R.; Deng, X. Design of On-line Detection Device for Grain Breakage of Corn Harvester Based on OpenCV. In Proceedings of the 2019 ASABE Annual International Meeting, Boston, MA, USA, 7 July–10 July 2019.
8. Fan, B. Research on Rice Appearance Quality Detection System Based on Machine Vision. Master's Thesis, Hebei Agricultural University, Baoding, China, 2022.
9. Niu, S.; Ma, R.; Xu, X.; Liang, A.; Mu, C.; Xu, J.; Ma, D. Research on MobileNetV2 Maize Seed Variety Recognition Based on Improved Attention Mechanism CBAM based on machine vision. *J. Chin. Cereals Oils Assoc.* **2023**, 1–12. [\[CrossRef\]](#)
10. Pan, W.; Sun, M.; Yuan, Y.; Liu, P. Identification Method of Wheat Grain Phenotype Based on Deep Learning of ImCascade R-CNN based on machine vision. *Smart Agric.* **2023**, *5*, 110–120.
11. Wang, Q. Study On Online Real-time Detection System of FHB Wheat Kernels and Identification of FHB Kernels Based on Deep Learning. Master's Thesis, Nanjing Agricultural University, Nanjing, China, 2021.
12. Zhang, H.; Yan, N.; Wu, X.; Wang, C.; Luo, B. Design and Experiment of Online Maize Single Seed Detection and Sorting Device Learning. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 159–166.
13. Wu, Z. Research on an Online Monitoring System for Impurity Breakage Rate in Rice and Wheat Grains Based on Mask R-CNN. Master's Thesis, Jiangsu University, Zhenjiang, China, 2022.
14. Li, X.Y.; Du, Y.F.; Yao, L.; Wu, J.; Liu, L. Design and Experiment of a Broken Corn Kernel Detection Device Based on the YOLOv4-Tiny Algorithm. *Agriculture* **2021**, *11*, 1238. [\[CrossRef\]](#)
15. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
16. Jocher, G.; Chaurasia, A.; Stoken, A.; Borovec, J.; Kwon, Y.; Michael, K.; Fang, J.; Wong, C.; Yifu, Z.; Montes, D. ultralytics/YOLOv5: v6. 2-YOLOv5 classification models, apple m1, reproducibility, clearml and deci. ai integrations. *Zenodo* **2022**. [\[CrossRef\]](#)
17. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv* **2022**, arXiv:2209.02976.
18. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
19. Wang, C.-Y.; Yeh, I.-H.; Liao, H.-Y.M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv* **2024**, arXiv:2402.13616.
20. Geng, D.; Wang, Q.; Li, H.; He, Q.; Yue, D.; Ma, J.; Wang, Y.; Xu, H. Online detection technology for broken corn kernels based on deep learning Learning. *Trans. Chin. Soc. Agric. Eng.* **2023**, *39*, 270–278.
21. Lan, Y.; Sun, B.; Zhang, L.; Zhao, D. Identifying diseases and pests in ginger leaf under natural scenes using improved YOLOv5s. *Trans. Chin. Soc. Agric. Eng.* **2024**, *40*, 210–216.
22. Ding, R.; Chen, B. Tomato recognition and detection algorithm based on improved YOLOv5. *J. Fujian Univ. Technol.* **2023**, *21*, 585–591.
23. Sun, F.; Wang, Y.; Lan, P.; Zhang, X.; Chen, X.; Wang, Z. Identification of apple fruit diseases using improved YOLOv5s and transfer learning. *Trans. Chin. Soc. Agric. Eng.* **2022**, *38*, 171–179.
24. Wei, T.; Liu, T.; Zhang, S.; Li, S.; Miao, H.; Liu, S. Research on pepper picking robot recognition and positioning method based on improved YOLOv5s. *J. Yangzhou University. (Nat. Sci. Ed.)* **2023**, *26*, 61–69.
25. Xu, H.; Tang, Z.; Zhang, J.; Zhu, P. Research on Optimization of YOLOv5s Detection Algorithm for Steel Surface Defect based on improved YOLOv5s. *Comput. Eng. Appl.* **2024**, *60*, 306–314.
26. Zuo, H.; Huang, Q.; Yang, J.; Sun, Q.; Li, S.; Li, L. Improved YOLOv5s-based Detection Method for Crop Yellow Leaf Curl Virus Disease. *Trans. Chin. Soc. Agric. Mach.* **2023**, 1–11.
27. Chen, Y.; Wu, X.; Zhang, Z.; Yan, J.; Zhang, F.; Yu, L. Method for identifying tea diseases in natural environment using improved YOLOv5s. *Trans. Chin. Soc. Agric. Eng.* **2023**, *39*, 185–194.
28. Zhu, X.; Chen, F.; Zheng, Y.; Li, Z.; Zhang, X. Identification of olive cultivars using bilinear networks and attention mechanisms. *Trans. Chin. Soc. Agric. Eng.* **2023**, *39*, 183–192.
29. Li, L.; Lu, S.; Ren, H.; Xu, G.; Zhou, Y. Mulberry Branch Identification and Location Method Based on Improved YOLO v5 in Complex Environment. *Trans. Chin. Soc. Agric. Mach.* **2024**, *55*, 249–257.
30. Yang, H.W.; Liu, Y.Z.; Wang, S.W.; Qu, H.X.; Li, N.; Wu, J.; Yan, Y.F.; Zhang, H.J.; Wang, J.X.; Qiu, J.F. Improved Apple Fruit Target Recognition Method Based on YOLOv7 Model. *Agriculture* **2023**, *13*, 1278. [\[CrossRef\]](#)
31. Wang, L.; Liu, J.; Wang, W. Small target detection method in UAV images based on dilated convolution fusion Transformer. *J. Comput. Appl.* **2024**, 1–10.
32. Liu, W.; Lu, X. Research Progress of Transformer Based on Computer Vision. *Comput. Eng. Appl.* **2022**, *58*, 012033.
33. Bao, W.; Xie, W.; Hu, G.; Yang, X.; Su, B. Wheat ear counting method in UAV images based on TPH-YOLO. *Trans. Chin. Soc. Agric. Eng.* **2023**, *39*, 155–161.
34. Wang, P.; Du, J.; Zhang, Y.; Liu, J.; Li, H.; Wang, C. Yield Estimation of Winter Wheat Based on Multiple Remotely Sensed Parameters and CNN-Transformer. *Trans. Chin. Soc. Agric. Mach.* **2024**, *55*, 154–163.

35. Wang, X.; Liu, Z. Infrared small target detection based on multi-layers multi-directions Transformer. *Acta Aeronaut. Astronaut. Sin.* **2024**, 1–14. [[CrossRef](#)]
36. Fang, S.D.; Wang, Y.F.; Zhou, G.X.; Chen, A.B.; Cai, W.W.; Wang, Q.F.; Hu, Y.H.; Li, L.J. Multi-channel feature fusion networks with hard coordinate attention mechanism for maize disease identification under complex backgrounds. *Comput. Electron. Agric.* **2022**, 203, 107486. [[CrossRef](#)]
37. Yang, S.Z.; Wang, W.; Gao, S.; Deng, Z.P. Strawberry ripeness detection based on YOLOv8 algorithm fused with LW-Swin Transformer. *Comput. Electron. Agric.* **2023**, 215, 108360. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.