ISSN 2616-5767 Vol.7, Issue 3: 39-47, DOI: 10.25236/AJETS.2024.070306

Research on corn ears defect detection algorithm based on improved YOLOv7

Chen Jiyao[1,2,a,*], Long Dongping[1,2,b], Yang Shaozhou[1,2,c]

[1]Hunan Provincial Key Laboratory of High Efficiency and Precision Machining of Difficult-to-Cut Material, Hunan University of Science and Technology, Xiangtan, Hunan, China
[2]School of Mechanical Engineering, Hunan University of Science and Technology, Xiangtan, Hunan, China
[a]cjy32025@163.com, [b]ldp_1030@163.com, [c]ys1587328@163.com
[*]Corresponding author

Abstract: *China as the world's leading corn producer, consumer and exporter, corn peeling corn mechanized harvesting has become the main development direction of China's corn production. Due to the low bract stripping rate during the mechanical peeling process of the corn harvester, too many bracts remain on the corn ears, making it impossible for the moisture within the corn ears to be discharged in a timely manner, which can easily lead to bacterial infection and mold growth. In addition, due to the high moisture content of corn ears in China, corn ears are susceptible to mold and mildew during storage, leading to a decline in corn quality. In summary, the stripping device of the corn ears harvester has outstanding problems such as low bract stripping rate and high damage rate of corn kernels, which seriously affects the development of China's corn industry and the realization of the goal of increasing production and income. Solving the problem of sorting the residual corn (unpeeled bracts and damaged corn kernels) after peeling is a key technical problem that needs to be solved urgently. To address the above problems this paper proposes an improved corn cob defect detection algorithm for YOLOv7. Firstly, the Explicit Visual Center Block (EVCBlock) is introduced into the head network, which improves the model's ability of recognizing the small target of corn kernel breakage. Then the Receptive field enhancement module (RFEM) is introduced to enhance the feature pyramid's ability to extract defective features of corn cob. The experiments show that the mean average accuracy of the improved YOLOv7 model is 88.1%, which is 12.2 percentage points higher than that of the original YOLOv7 model, realizing the need for more accurate corn cob defect detection.*

Keywords: *Defect detection; YOLOv7; Deep learning; Corn ears; Object detection*

1. Introduction

With the increasing demand for fresh corn, the mechanization and automation of fresh corn dehusking have become more advanced. The effectiveness of dehusking devices is influenced not only by the structural and operational parameters of the devices but also by the physical characteristics of the corn itself. Factors such as the diameter, length, and moisture content of the corn ears can affect the outcome of the dehusking process, leading to incomplete dehusking or excessive removal that damages the corn kernels. Corn with unclean dehusking or damaged kernels (collectively referred to as defective corn) needs to be sorted out before packaging. The quality inspection and selection of fresh corn have primarily relied on manual sorting, with machine sorting also beginning to be used, although it still faces certain issues, such as a lower degree of precision in selection identification. With the advancement of deep learning, researchers have applied object detection technology to the quality inspection of agricultural products, achieving favorable results[1]. This has provided a reference for the detection of defects in corn ears.

The primary objective of object detection is to locate and recognize individual or multiple objects within an image, accurately classifying their categories while returning their positions. Currently, object detection algorithms have evolved into two major branches: one is the two-stage network models, with typical representatives being RCNN[2], Fast-RCNN[3], and Faster-RCNN[4] networks. Two-stage networks design the network by dividing it into feature extraction and classification stages, offering high accuracy but slow detection speed. To address this drawback, researchers have proposed one-stage network models, with typical examples including SSD[5], RetinaNet[6], CenterNet[7], EfficientDet[8],

ISSN 2616-5767 Vol.7, Issue 3: 39-47, DOI: 10.25236/AJETS.2024.070306

FCOS[9], as well as the YOLO series of algorithms. The concept behind YOLO algorithm involves dividing the image into multiple grids, predicting bounding boxes and their corresponding object categories within each grid. YOLOv2[10] introduces the Anchor mechanism, enhancing detection performance. YOLOv3[11] employs a clustering-based positive and negative sample matching method, which, compared to IoU-based methods, can better adapt to various object scales and aspect ratios, thereby improving model accuracy and stability. YOLOv4[12] introduces the Mish activation function, increasing accuracy. YOLOv5[13] utilizes the Focus slice operation and cascaded multiple small-sized pooling kernels to improve feature distortion. YOLOv7[14] introduces multi-branch stacking modules, incorporating model reparameterization into the network architecture and proposing training with auxiliary heads to further optimize object detection performance.

In practical applications, Zhang et al.[15] utilized two-dimensional fast imaging technology to screen for damaged, moldy, and infested maize cobs, but did not identify the specific locations of defects on the maize cobs. Meng et al.[16] partitioned moldy maize grains from normal ones based on the HSV color space, without considering the situation of damaged maize grains. Li et al.[17] employed HSV and CLBP (Complete Local Binary Patterns) methods to extract color and texture features of maize ears, utilizing SVM to rapidly classify abnormal maize ears with mixed colors, missing grains, insect damage, and disorderly grains. The aforementioned detections only classify defective maize cobs from normal ones without pinpointing the location of defects on the maize cobs.

In order to effectively solve the above problems, this paper adds Explicit Visual Center Block (EVCBlock) and Receptive field enhancement module (RFEM) on the basis of the original YOLOv7 to realize the intelligent detection of corn cob defects with the corn peeling machine after peeling. With the corn that has peeling defects after the corn peeling machine as the research object, a corn cob defect detection method based on the improved YOLOv7 is proposed to point out the specific defect location of the corn cob, realize the intelligent detection of corn cob defects, and provide theoretical support for the detection of corn cob defects.

2. Improving the YOLOv7 Network Model

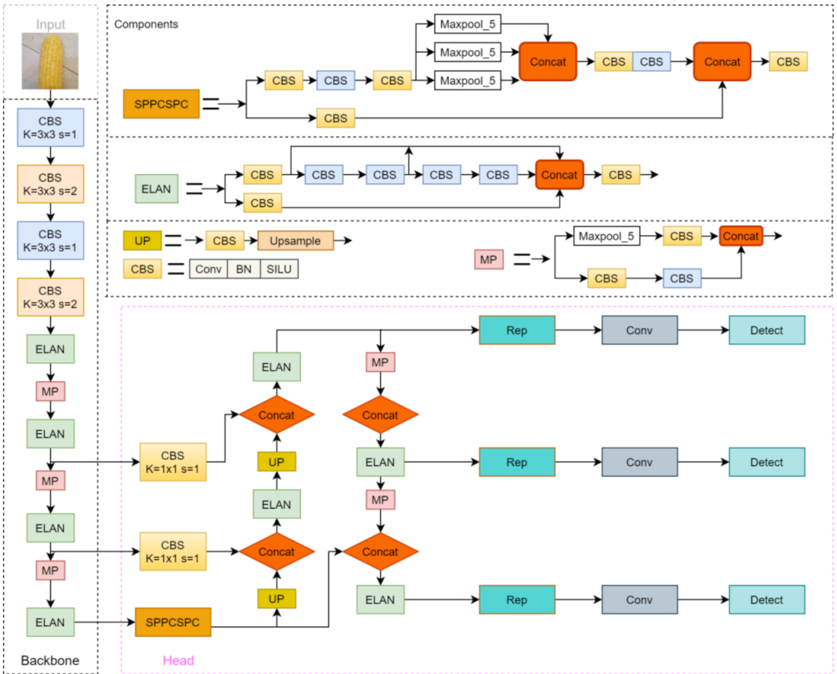

Figure 1: YOLOv7 Network structure

The network structure of YOLOv7 is shown in Figure 1. Overall, YOLOv7 resizes the input image to a size of 640x640 at the input end before feeding it into the backbone of the main feature extraction network. The backbone consists of CBS convolution layers, MP convolution layers, and the efficient aggregation network ELAN layers. Its primary function is to extract features of target information of varying sizes. Its most notable characteristic is the adoption of the efficient E-ELAN network architecture, which can improve the detection efficiency of the algorithm. Subsequently, the process goes through the detection head, focusing on the SPPCSPC module that can increase the receptive field. This allows the

ISSN 2616-5767 Vol.7, Issue 3: 39-47, DOI: 10.25236/AJETS.2024.070306

algorithm to adapt to images of different resolutions and reduces computational load by half, thereby accelerating detection speed. Finally, the detection results are obtained through the conv layer.

2.1. Explicit Visual Center Module(EVCBlock)

Since the identification of damaged corn kernels falls under the category of small object detection, there might be omissions or errors in recognizing features, which can result in a trained model with low recognition accuracy. Therefore, to extract more features of damaged corn kernels, this paper adds an Explicit Visual Center Module (EVCBlock) to the detection head[18]. The structure of the EVCBlock module is shown in Figure 2.

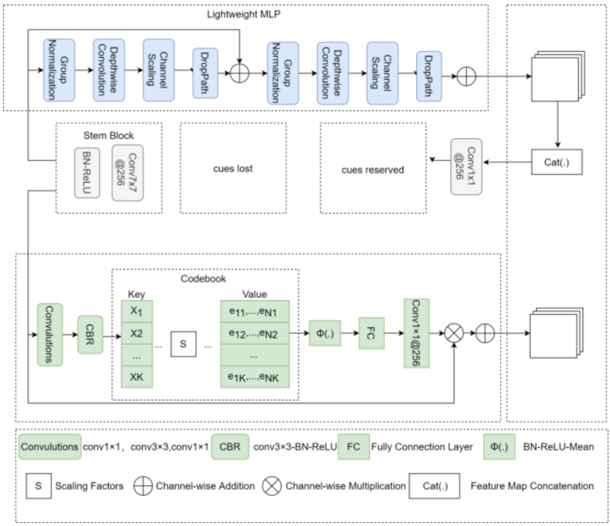

Figure 2: The EVCBlock module diagram

Based on visual features pyramid extracted from the CNN backbone, a scheme of explicit visual center, namely EVC, is proposed: EVC mainly consists of two parallel connected blocks, among which lightweight MLPs are utilized to capture the global long-term dependencies of top-level features. Meanwhile, to preserve local angular regions, a learnable visual center mechanism is proposed to aggregate local region features within the layers. The global information reflects target-specific details, allocating regions of interest for each target of various sizes and enabling these targets to be identified across different feature layers, thereby addressing issues such as small size and feature omissions in corn cob grain targets.

2.2. RFE module

In images of corn ears defects, there is often the issue of detecting small objects, such as grain damage, leading to missed and false detections of corn ears defects. To address this, a Receptive Field Enhancement Module (RFEM) is introduced in the detection head to learn different receptive fields of the feature map and enhance the feature pyramid representation[19]. To mitigate the occlusion between different corn ear defect detections, Repulsion Loss is used to penalize the predicted boxes from shifting to other true objects, reducing the sensitivity of corn ear defect detection results to the NMS threshold, and can decrease the rate of missed and false detections of corn ear defects.

The RFE module can be divided into two main parts: multi-branch based on dilated convolutions and a gather-and-weight layer based on dilated convolutions. The multi-branch component primarily achieves its functionality through setting dilated convolutions of different ratios, while the gather-and-weight layer collects information from each branch and weights all the branches of features. The RFE structure is depicted in Figure 3. In RFE, the main technique used is Dilated Convolution. It employs dilated convolutions of different ratios for varying convolution operations, specifically using four different scales of dilated convolution branches to capture multi-scale information and different ranges of dependencies, with shared weights among these branches. The only difference lies in their receptive fields. Additionally, the residual connections of the dashed branch in Figure 3 can prevent the issue of gradient explosion.

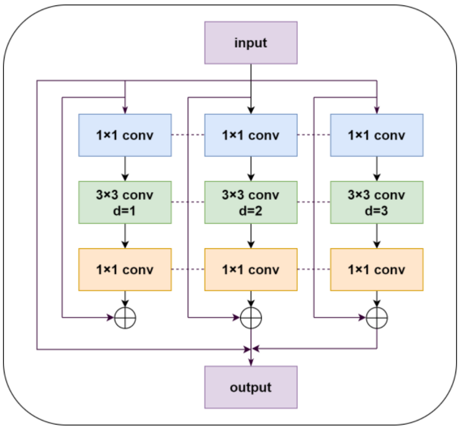

Figure 3: Structural diagram of the RFE module

The improved YOLOv7 network structure is shown in Figure 4. In the improved algorithm, in the original YOLOv7 feature fusion network, the EVCBlock module is added behind the first UP layer, which is capable of extracting the global long-range correlations in the image, capturing the extensive connections between the internal parts of the image as well as meticulously preserving the integrity of the local details of the image; and the RFE module is added behind the SPPCSPC layer, which jointly enhances the sensory field with fewer model parameters, while reducing the potential risk of overfitting, and can fully utilize each sample.

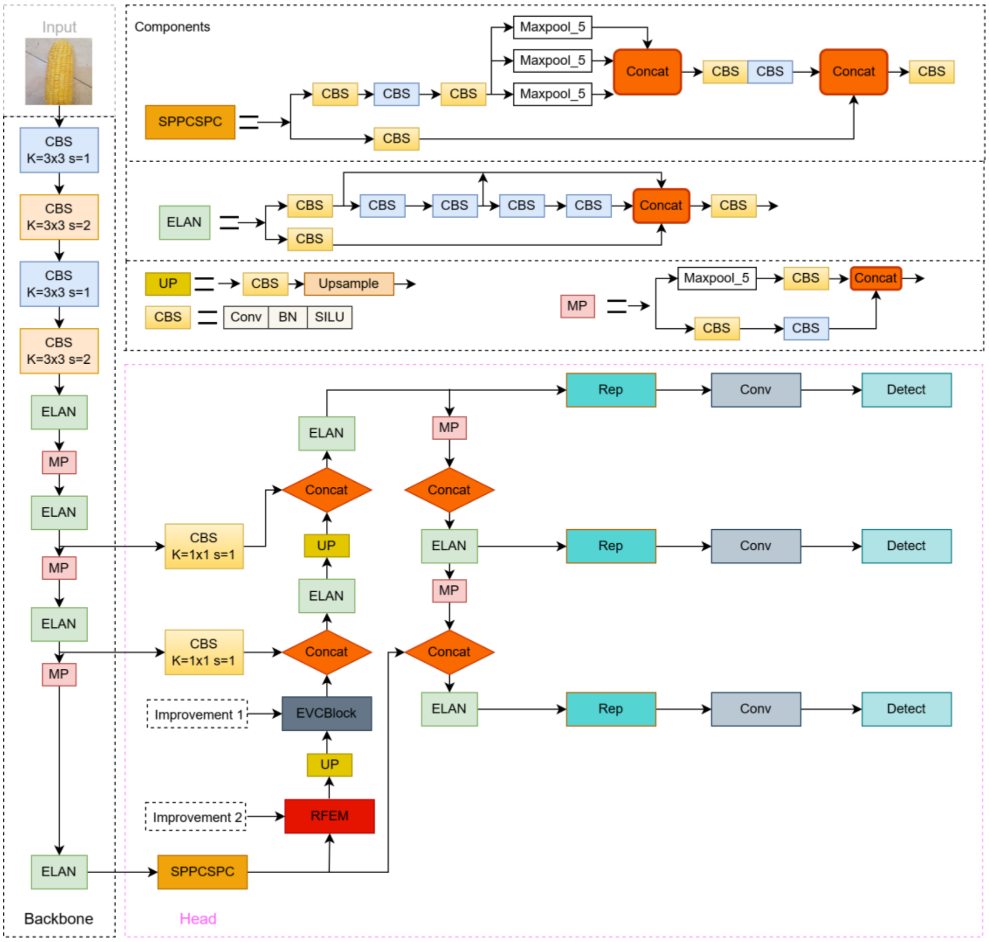

Figure 4: The improved YOLOv7 network structure

ISSN 2616-5767 Vol.7, Issue 3: 39-47, DOI: 10.25236/AJETS.2024.070306

3. Datasets construction and the experimental platform

3.1. Datasets construction

A total of 1,294 images of corn ears defects were collected using the Samsung camera model SM-G9550, as shown in Figure 5. Among these, 1,070 images feature unclean husking, and 438 images show particle damage. The open-source software LabelImg was utilized for annotating the corn ears defect images. The label "Unskinned" indicates unclean husking, while "Particle breakage" signifies particle damage. The dataset was divided into training, validation, and test sets in a 7:2:1 ratio, respectively, to create a YOLO-format dataset for corn ear defects, which was then used for training and testing YOLOv7.

Figure 5: Partial Datasets Illustration

3.2. Experimental platform

As shown in Table 1, the corn ears defects detection training platform utilizes a corn ears defects dataset that has been independently established.

Table 1: Experimental platform

name	configure
CPU	Intel(R)Core(TM)i5-10500H CPU @ 2.50GHz 2.50 GHz
memory	16GB
GPU	NVIDIA GeForce RTX 3060
operating system	Windows 10 (64 bit)
environment	Python 3.9
Batchsize	4
Epoch	200
Initial learning rate	0.001

4. Experimental results and analysis

4.1. Evaluating indicator

This article uses the Intersection over Union (IoU) function to reflect the accuracy of the prediction boxes. IoU is a crucial metric in object detection problems, representing the overlap between the annotated and predicted boxes during the training phase, used to measure the correctness of the prediction boxes. The precision (P), recall (R), average precision (AP), and mean average precision (mAP) of the confusion matrix after training are calculated as evaluation metrics. Their calculation formulas are shown in Equations (1)-(4), respectively.

ISSN 2616-5767 Vol.7, Issue 3: 39-47, DOI: 10.25236/AJETS.2024.070306

$$\text{precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{1}$$

$$\text{recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{2}$$

$$\mathrm{AP} = \frac{\sum_{i=0}^{n} \mathrm{P_i}}{\mathrm{n}} \tag{3}$$

$$\mathrm{mAP} = \frac{\sum_{i=0}^{k} \mathrm{AP_i}}{\mathrm{k}} \tag{4}$$

In the formulas: TP (true positive) refers to predictions made as positive cases with label values also as positive cases, which are correctly predicted; FN (false negative) refers to predictions made as negative cases with label values as positive cases, which are incorrectly predicted; FP (false positive) refers to predictions made as positive cases with label values as negative cases, which are incorrectly predicted; n represents the number of samples; k represents the number of sample categories.

4.2. Ablation experiment

To verify the effectiveness of the various improvement measures proposed in this study, this experiment used the original YOLOv7 network as a baseline and conducted four sets of ablation experiments on a custom datasets, maintaining consistency in the experimental environment and parameter configuration to ensure the accuracy and comparability of the results. The loss function curves of the five models when trained on the custom datasets are shown in Figure 6. It can be observed from the figure that, after incorporating all four improvement points, the training loss was lower throughout the entire process compared to the other four models, indicating the best model convergence effect.

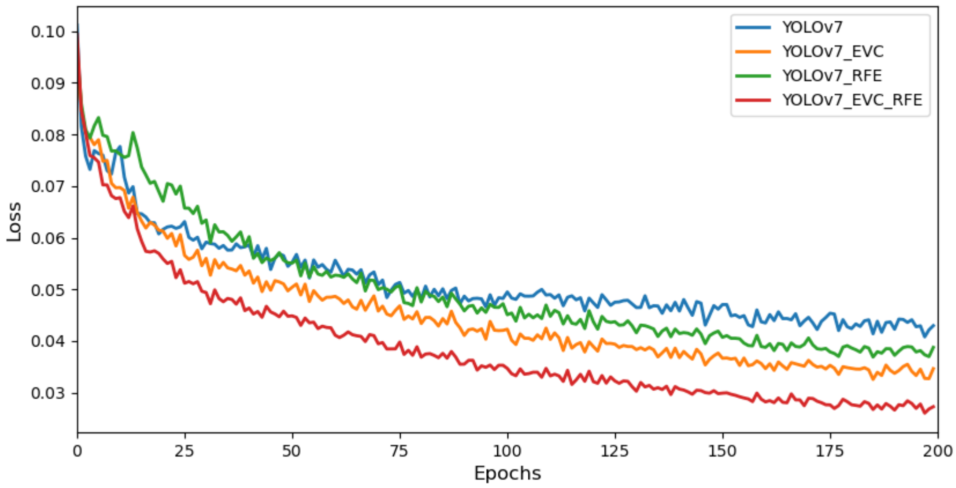

Figure 6: Plot of the training loss function for the four models

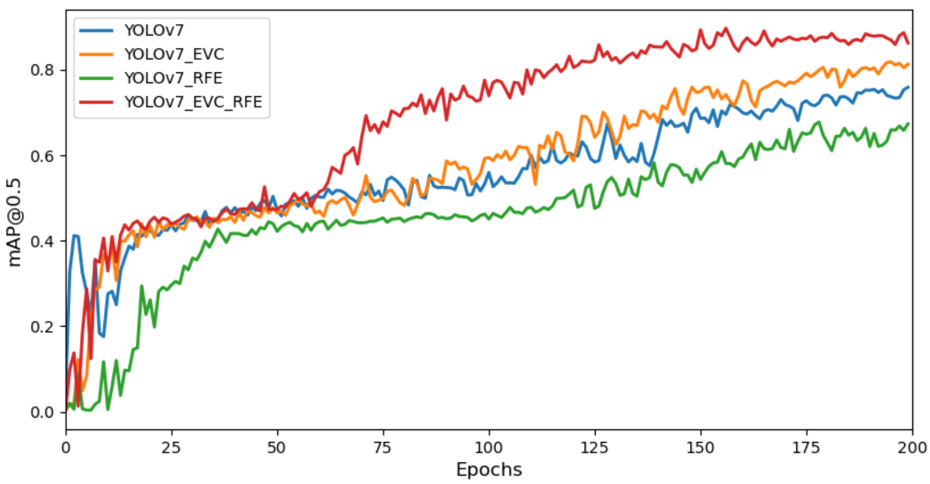

Figure 7: Mean comparison of the average accuracy of the four models

ISSN 2616-5767 Vol.7, Issue 3: 39-47, DOI: 10.25236/AJETS.2024.070306

From Figure 7, it can be seen that when the IoU threshold is set to 0.5, the model trained with all four improvement points incorporated leads in average precision compared to the models trained with the other four configurations, and the model converges faster, reaching convergence around 75 training cycles. This demonstrates the effectiveness of training with each improvement point. The IoU threshold is used to determine the degree of overlap between the predicted and actual bounding boxes. When the IoU threshold is low, the model tends to predict more positive samples more easily but at the same time generates more false positives. Conversely, when the IoU threshold is high, the model focuses more on the degree of overlap between the predicted and actual bounding boxes, thereby reducing false positives but may also miss detecting some positive samples.

The five models obtained from the training were tested, and the test results are shown in Table 2.

Table 2: Ablation experiment

No	EVC	RFE	P/(%)	R/ (%)	Unskind $AP_{0.5/}$ (%)	Particle breakage $AP_{0.5/}$(%)	$mAP_{0.5/}$(%)
No.1			78.4	80.0	94.0	57.8	75.9
No.2	√		77.2	79.9	94.0	65.8	79.9
No.3		√	71.2	61.7	91.5	34.5	63.0
No.4	√	√	89.8	85.2	96.4	79.8	88.1

(1) No.1 is the test result of the basic Yolov7 algorithm, which serves as a comparison benchmark for the last three training sets, the average precision mean value of the test set is 75.9%, the accuracy is 78.4%, the recall is 80.0%, and the precision of particle breakage detection is low.

(2) No.2 is the addition of the EVCBlock module, the accuracy and recall are basically unchanged, the average precision mean is improved by 4.0%, and the particle breakage detection precision is significantly improved.

(3) No.3 is the addition of RFE module, the recall rate and precision have decreased, and the particle damage detection precision is greatly reduced.

(4) No.4 is adding both EVCBlock module and RFE module, the experimental results show that the precision of the combination of the two modules increases dramatically, the accuracy and recall are 89.8 and 85.2 respectively, and the detection precision of particle damage is improved by 22% compared with the original YOLOv7, which confirms the validity of the algorithmic improvement points in this paper.

Figure 8: Comparison of the improved model

ISSN 2616-5767 Vol.7, Issue 3: 39-47, DOI: 10.25236/AJETS.2024.070306

In Figures 8 and 9, (a) represents the detection effect of the YOLOv7 model; (b) represents the detection effect of adding the EVCBlock module; (c) represents the detection effect of adding the RFE module; and (d) represents the detection effect of adding both the EVCBlock and RFE modules. From Fig. 8, it can be seen that there is little difference in the detection effect of the four models on bracts, and all targets are detected with high confidence. From Fig. 9, it can be seen that the YOLOv7 model has one missed detection and the confidence level is not high. After adding two modules at the same time the detection effect of the model is greatly improved, and all targets are detected with high confidence.

Figure 9: Comparison of particle damage detection effects of the improved and improved model

5. Conclusion

Aiming at the problem of low accuracy of corn cob defect detection for small targets with broken particles, this paper proposes a corn cob defect detection algorithm based on the improved YOLOv7, which joins the EVCBlock module and the RFE module in the head network and improves the extraction ability of corn cob defect features, and the experimental results show that the average accuracy of the improved algorithm in this paper reaches 88.1%, which is 12.2% higher than the original algorithm. accuracy mean value is improved by 12.2%. The improved algorithm in this paper improves the problem of leakage and misdetection in the detection of corn cob grain damage, so the algorithm in this paper can be applied to the automated detection of corn cob defects to improve the level of intelligent agricultural production.

Acknowledgements

This study was supported by the Shaoguan College Doctoral Introduction Fund Program(Award number: 440-9900064607).

References

[1] Shi J,et al.Study on a detection method for crop diseases and insect pests based on YOLO v5s

ISSN 2616-5767 Vol.7, Issue 3: 39-47, DOI: 10.25236/AJETS.2024.070306

improved model[J].Jiangsu Agricultural Sciences,2023,51(24):175-183.
[2] Girshick B R ,Donahue J ,Darrell T , et al. Rich feature hierarchies for accurate object detection and semantic segmentation. [J]. CoRR, 2013, abs/1311.2524.
[3] Girshick B R . Fast R-CNN. [J]. CoRR, 2015, abs/1504.08083.
[4] Shaoqing R ,Kaiming H ,Ross G , et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. [J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39 (6): 1137-1149.
[5] Liu W ,Anguelov D ,Erhan D , et al. SSD: Single Shot MultiBox Detector. [J]. CoRR, 2015, abs/1512.02325.
[6] Tsung-Yi L ,Priya G ,Ross G , et al. Focal Loss for Dense Object Detection. [J]. IEEE transactions on pattern analysis and machine intelligence, 2020, 42 (2): 318-327.
[7] Zhou X ,Wang D ,Krähenbühl P . Objects as Points. [J]. CoRR, 2019, abs/1904.07850.
[8] M. T ,R. P ,Q.V. L . EfficientDet: Scalable and efficient object detection [J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020, 10778-10787.
[9] Tian Z ,Shen C ,Chen H , et al. FCOS: Fully Convolutional One-Stage Object Detection. [J]. CoRR, 2019, abs/1904.01355
[10] Redmon J ,Farhadi A . YOLO9000: Better, Faster, Stronger. [J]. CoRR, 2016, abs/1612.08242.
[11] REDMON J,FARHADI A.Yolov3:an incremental improve-ment[J].arXiv:1804.02767,2018.
[12] BOCHKOVSKIV A,WANG C Y,LIAO H Y M.YOLOv4:optimalspeed and accuracy of object detection[J].ArXiv Preprint arXiv,2020,2004:10934.
[13] GLENN J.YOLOv5[EB/OL].(2020).https://github.com/ultralytics/yolov5.
[14] 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[EB/OL]. (2022-06-06)[2023-08-06].http://arxiv.org/ abs/2207. 02696.
[15] Zhang F,et al.Screening Method of Abnormal Corn Ears Based on Machine Vision[J].Transactions of the Chinese Society for Agricultural Machinery,2015,46(S1):45-49.
[16] MENG F,et al.Design and Experiment of Real-time Detection and Sorting Device for Maize Seeds[J].Transactions of the Chinese Society for Agricultural Machinery,2021,52(3):153-159,177.
[17] LI Q,WANG K,QIANG H,et al.Classification and recognition method of abnormal corn ears based on color and texture features[J].Jiangsu Journal of Agricultural Sciences,2020,(01):24-31.[doi:doi:10. 3969/j.issn.1000-4440.2020.01.004].
[18] Yu Q ,Dong Z ,Liyan Z , et al. Centralized Feature Pyramid for Object Detection. [J]. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 2023, PP.
[19] Yu Z,Huang H,Chen W,et al.YOLO-FaceV2:A Scale and Occlusion Aware Face Detector[J].arXiv preprint arXiv:2208.02019,2022.