动漫与传媒学院 2024 年硕士研究生国家奖学金评选拟推荐人员申报材料公示

牛思琪评选材料清单

附件 1：学业成绩

附件 2：中科院三区文章 SCI 论文一作

附件 3：一级学报论文一作

附件 4：计算机软件著作权：土壤健康状况监测系统

附件 5：计算机软件著作权：温室智能调控数字化

附件 6：青岛市千名优秀大学生

附件 7：山东省优秀毕业生

附件 1：学业成绩

青岛农业大学硕士研究生成绩单

姓名	牛思琪	学科（领域）	农业工程与信息技术				
31							
学号	20222114008	院系、部（所、中心、学科）	动漫与传媒学院			年级	2022
课程编号	课程名称		课程类别	学时	学分	成绩	备注
2111501	综合英语Ⅱ		公共学位课	48	3	90	
2142306	舞龙舞狮		公共必修课	24	1	89	
2112101	新时代中国特色社会主义理论与实践		公共学位课	32	2	89	
2220001	现代生活美学——花香茶之道		公共选修课	16	1	93	
2110701	现代农业创新与乡村振兴战略		公共学位课	32	2	95	
ngyxj-002	农业工程与信息技术案例（2）		非学位课	48	3	93	
2131403	农业信息技术		学位课	32	2	87	
ngyxj-004	农业推广理论与实践		公共学位课	32	2	89	
ngyxj-014	农业影视创作与传播		学位课	32	2	92	
2130407	文献检索与论文写作（双语）		公共学位课	16	1	93	
2130170	现代农业概论		专业必修课	32	2	92	
2130416	农业大数据		专业必修课	32	2	83	
总学分：23，学位课学分：4。							

青岛农业大学研究生处 成绩证明专用章

打印日期：2024-10-30

验 证 码：A337B13F4068E56E3E57F6CC9E718D0E

附件 2：中科院三区文章 SCI 论文一作

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 30 November 2023, accepted 23 January 2024, date of publication 13 February 2024, date of current version 6 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3365559

RESEARCH ARTICLE

Research on a Lightweight Method for Maize Seed Quality Detection Based on Improved YOLOv8

SIQI NIU[1], XIAOLIN XU[1], AO LIANG[1], YULIANG YUN[2], LI LI[3], FENGQI HAO[4], JINQIANG BAI[4], AND DEXIN MA[1,5]

[1]College of Animation and Communication, Qingdao Agricultural University, Qingdao 266109, China
[2]College of Mechanical and Electrical Engineering, Qingdao Agricultural University, Qingdao 266109, China
[3]Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture and Rural Affairs, China Agricultural University, Beijing 100083, China
[4]Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan 250014, China
[5]Intelligent Agriculture Institute, Qingdao Agricultural University, Qingdao 266109, China

Corresponding author: Dexin Ma (madexin@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 42201458, in part by the Shandong Provincial Natural Science Foundation under Grant ZR2022MC152, in part by the Central Government Guiding Local Science and Technology Development Special Plan under Grant 23-1-3-6-zyyd-nsh, and in part by the Shandong Province Key Research and Development Plan under Grant 2023TZXD023.

ABSTRACT Seeds are the most basic and important means of production for agriculture. During the production and processing of seeds, they may undergo potential mechanical damages and mildew alterations, which might jeopardize their germination viability. Hence, checking the quality of seeds before sowing is of paramount importance for the benefit of the sower and the safety of agricultural production. In order to achieve an efficient detection of maize seed quality, our experiment assembled a dataset composed of 2,128 seeds with four different health statuses of maize: healthy, broken, moth-eaten, and mildewed. In this paper, we proposed a lightweight maize seed quality detection model for small objects based on improved YOLOv8: I-YOLOv8. Firstly, we introduced a multi-scale attention mechanism called EMA to efficiently retain information across channels and reduce computational load. Next, we chosen the SPD-Conv module for low-resolution images and small objects, and applied it to the backbone, which addressed the loss of fine-grained information and the less efficient learning of feature representations present in YOLOv8. Lastly, we reduced the large detection layer, which directed the network to pay more attention to the location, channel, and dimensional information of smaller objects, and we also replaced the loss function with WIoUv3. We validated our model using ablation studies and compared it with YOLOv5, YOLOv6, and YOLOv8. The mAP (Mean Average Precision) of the improved model I_YOLOv8 reaches 98.5%, which is 6.7% higher than YOLOv8. The average recognition time per image was 163.9fps, a boost of 5.2fps compared to YOLOv8. This study lays a theoretical foundation for the efficient, convenient, and rapid detection of maize quality, while also offering a technical basis for advancing automated maize quality detection means.

INDEX TERMS YOLOv8, object detection, lightweighting, maize seed.

I. INTRODUCTION

Maize (Zea mays L.) is one of the most widely distributed crops in the world [1]. About one-third of the world's population depends on maize as a staple food [2]. It has high nutritional and economic value [3]. During the production and processing of maize seeds, they may undergo potential mechanical damages and mildew alterations, which might jeopardize their germination viability. Sowing maize seeds with damaged quality will reduce the germination rate and waste of labor, thus affecting economic benefits. Hence, checking the quality of seeds before sowing is of paramount

The associate editor coordinating the review of this manuscript and approving it for publication was Liandong Zhu.

importance for the benefit of the sower and the safety of agricultural production.

Traditional methods for detecting maize seed quality are categorized into empirical, physical, and chemical methods. These methods are often cumbersome, time-consuming, and limited by experimental locale [4]. The application of machine vision technology can achieve rapid and accurate detection and identification of maize seeds, which has very important application value [5].

In the early stages of conventional machine learning, some scholars used image processing techniques to simply process maize seed images for recognition. Chen et al. [6] proposed a method based on image In HSV and Otsu method based on genetic algorithm optimization, which achieved more accurate segmentation and recognition of the disease of color and shape features, and enhanced the real-time and accuracy of the image of maize disease detection and recognition. Subsequently, many researchers began to use neural network approaches for maize seed image analysis. Kiratiratanapruk and Sinthupinyo [7] extracted color histograms from RGB and HSV color spaces, along with textures based on Gray-Level Co-Occurrence Matrix (GLCM) and Local Binary Patterns (LBP), and then applied Support Vector Machine (SVM) to classify maize seed defects.

Traditional machine learning approaches have achieved some applications in the recognition of maize seeds. However, these traditional methods are subject to limitations such as their reliance on manually selected features, and the problem such as high computational demands and costs, which can impact recognition accuracy. In contrast, deep learning can autonomously learn complex features from raw data without the need for manual feature extraction. Presently, the most pervasive deep learning technique in machine learning is Convolutional Neural Network (CNN) [8]. For the detection of maize seeds, there are two main directions: image classification and target detection. Unlike image classification, which only categorizes objects within an image, object detection techniques perform image segmentation based on geometric and statistical features of targets, combining the tasks of object segmentation and recognition, so it has excellent accuracy and real-time processing capabilities.

Object detection algorithms based on deep learning are mainly divided into two categories: two-stage and single-stage methods. The two-stage object detection algorithm initially generates RP (Region Proposals), followed by sample classification using convolutional neural networks. Representative two-stage object detection algorithms include R-CNN [9], SPP-Net [10], Faster R-CNN [11], Mask R-CNN [12]. Zhao et al. [13] designed four distinct network models based on Faster R-CNN, and achieved superior recognition results for the selection of maize kernels by directly inputting color images. Velesaca et al. [14] used the Mask R-CNN algorithm for segmenting and extracting maize group images, and designed a lightweight network CK-CNN to classify good kernels, defective kernels, and impurities. Although two-stage detection algorithms exhibit high accuracy, their slower detection speed renders them unsuitable for real-time detection. The single-stage object detection algorithm can directly generate the class probability and location coordinates of the target in one stage without generating RP. Representative two-stage object detection algorithms include SSD [15], YOLO [16], RetinaNet [17]. Although these algorithms may exhibit moderate accuracy, their fast detection speed makes them suitable for real-time detection tasks. Liu and Wang [18]proposed a method for detecting damaged maize kernels based on YOLOv3-tiny. The proposed maize detection method is implemented on NVIDIA TX2 and can achieves the speed up to 10fps speed, which can perform almost real-time detection. Li et al. [19] investigated a maize seed breakage detection device based on YOLOv4-tiny, which is applicable to combine harvesters, addressing the issue of low accuracy in existing methods for maize seed integrity assessment. Thangaraj Sundaramurthy et al. [20] proposed an object detection method based on YOLOv5 to accurately detect maize seeds infected with Fusarium Head Blight (FHB), enabling real-time detection of FHB-infected maize seeds on the processing line. Wang et al. [21] aimed to rapidly and accurately identify broken maize kernels, proposing a model BCK-YOLOv7 based on an improved YOLOv7, which fine-tuned the model's positive sample matching strategy and incorporated Transformer encoding modules and CA attention mechanisms, enhancing the model's accuracy to 96.9%, recall to 97.5%, and mAP to 99.1%.

The study proposed a more streamlined convolutional neural network model, I_YOLOv8, based on YOLOv8 benchmark, addressing the loss of fine-grained information and low-efficiency feature representation learning inherent to YOLOv8, it also reducing the model's complexity for ease deployment on mobile devices. This experiment established a dataset for maize seed variety recognition, encompassing 2128 maize seeds of four types: healthy, broken, moth-eaten, and mildewed. Extensive comparative experiments were conducted with the I_YOLOv8, YOLOv5, YOLOv6, and YOLOv8 using this dataset. The results indicated that the proposed I_YOLOv8 significantly outperformed other methods, providing technical support for the automated recognition and non-destructive testing of maize seed quality.

II. MATERIALS AND METHODS

A. DATASET CONSTRUCTION

This experiment collected four types of maize seeds with different health conditions to establish a dataset, as shown in Figure 1, which are healthy, broken, moth-eaten, and mildewed. In order to enhance the robustness of the model, we chose five maize seeds varieties, namely JINYU118, KENUO58, LIYUAN296, HUIYU18, and TIEYAN630. Under natural lighting conditions, these seeds were randomly arranged in a ratio of healthy, broken, moth-eaten, and mildewed at 5:1:1:1. A total of 133 photographs were taken, each featuring 16 seeds, culminating in a dataset of

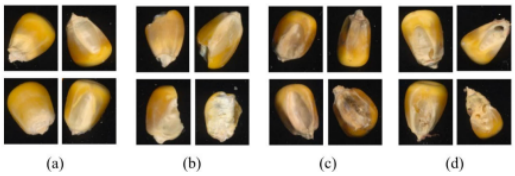


FIGURE 1. Maize seeds with four different health conditions. (a) Healthy maize. (b) Broken maize. (c) Mildewed maize. (d)Moth-eaten maize.

2128 seeds, the images of each maize seed category were randomly divided into training, validation, and test sets in a ratio of 7:2:1.

To further enhance the robustness and generalization capabilities of the model, this study augmented the training image number of the dataset through data augmentation techniques, these enhancements primarily included rotation, exposure adjustment, and mosaic techniques. The augmented maize seed dataset is shown in Figure 2. Figure 2(a) showed the quantity of maize seeds across different categories. Figure 2(b) depicted the spatial distribution of the maize seeds bounding boxes, indicating a relatively uniform spread of maize seeds without excessive clustering. Figure 2(c) presented the dimensions of the maize seeds bounding boxes, it can be seen that the height and width of the bounding box are relatively uniform.

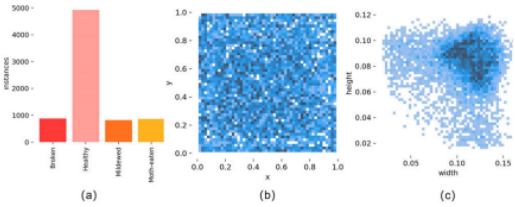


FIGURE 2. Visualization of the dataset. (a) Number of annotations per class. (b) The statistical distribution of the bounding box position. (c) The statistical distribution of the bounding box sizes.

B. MAIZE SEED DETECTION MODEL: I_YOLOv8

Since its initial release in 2015, YOLO (You Only Look Once) series of computer vision models has consistently been one of the most popular in the field of deep learning. YOLOv8 is the latest version of the YOLO series of algorithms, which can be used for object detection, segmentation, classification tasks, and learning of large-scale datasets. Compared to previous outstanding models in the YOLO series, such as YOLOv5 and YOLOv7, YOLOv8 offers higher detection accuracy and speed. YOLOv8 is a detection algorithm known for its fast detection speed and high accuracy. It performs well on some open-source datasets but requires improvement for seed detection tasks.

In order to address the challenges such as small seed size and low resolution in maize seed quality detection, this paper improved and optimized on the basis of YOLOv8, and proposed an algorithm I-YOLOv8 for maize seed quality detection.

The proposed model architecture was illustrated in Figure 3, and the specific improvements were summarized as follows:

(1) By incorporating attention mechanisms to enhance the object detection capability of the network and extract regions of interest. As illustrated in Figure 3, an efficient multi-scale attention mechanism EMA was introduced in the Backbone and added in front of the SPPF structure to efficiently retain the information on each channel and reduce the computational load.

(2) By integrating SPD-Conv module to boost detection capabilities of low-resolution images and small objects. As demonstrated in Figure 3, SPD-Conv was applied within the Backbone at the following stages: the 1st, 3rd, 5th, and 7th convolutional layers. This addressed the loss of fine-grained details and learning of less effective feature representations in YOLOv8.

(3) By removing one of the large object detection layers (P5), the network was directed to focus more on the location, channel, and dimensional information of smaller objects, thereby enhancing the detection of small-scale targets. As indicated in Figure 3, the detection layers of YOLOv8n were simplified from large, medium, and small to medium and small. The 24th layer (Convolution layer), the 25th layer (Concat module), and the 26st layer (C2f module) of the feature fusion layers were removed.

(4) By altering the loss function, the model's accuracy and overall performance were further improved. The loss function CIoU was replaced with WIoUv3, which balanced the ratio between low and high-quality samples, addressing issues of detecting small, blurry objects and those with overlapping occlusions.

1) EFFICIENT MULTI-SCALE ATTENTION MODULE

The attention mechanism is a mechanism that simulates human vision, focusing on important features while suppressing unnecessary ones. Remarkable effectiveness of the channel or spatial attention mechanisms for producing more discernible feature representation are illustrated in various computer vision tasks. However, modeling the cross-channel relationships with channel dimensionality reduction may bring side effect in extracting deep visual representations. Ouyang et al. [22] proposed a novel efficient multi-scale attention (EMA) module. The design prioritizes the retention of information from each channel while minimizing computational overhead. By reshaping certain channels into a batch dimension and segmenting the channel dimension into multiple sub-features, it ensures a homogeneous distribution of spatial semantic attributes within each feature subset. Specifically, apart from encoding the global information to re-calibrate the channel-wise weight in each parallel branch, the output features of the two parallel branches are further aggregated by a cross-dimension interaction for capturing pixel-level pairwise relationship.

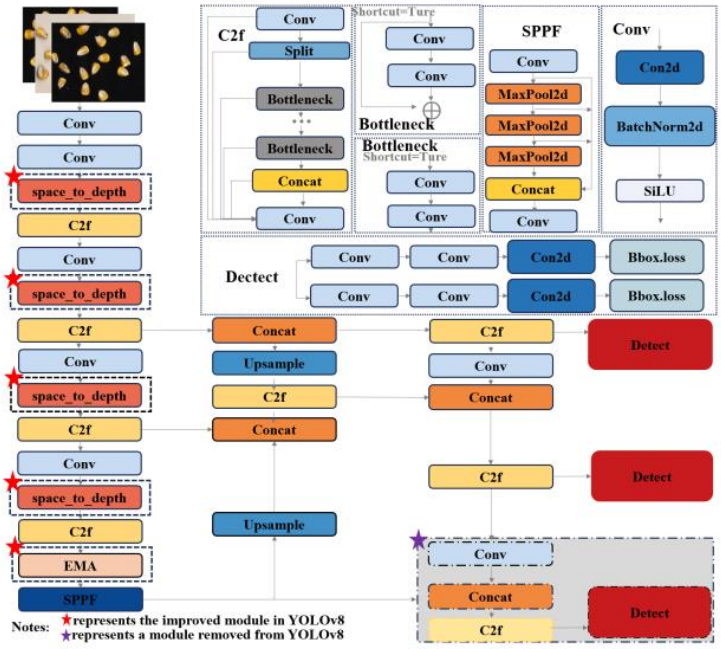


FIGURE 3. Improved YOLOv8 network structure.

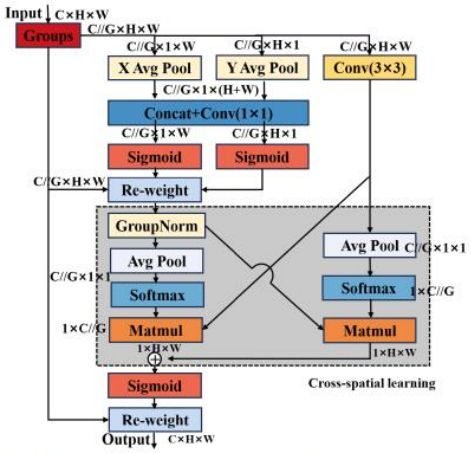


FIGURE 4. Efficient multi-scale attention structure.

The parallel substructures help the networks avoid more sequential processing and large depth. As shown in Figure 4, the EMA module employs a parallel processing strategy.

(1) Feature Grouping. For any given input feature map $X \in \mathbb{R}^{C\times H\times W}$, EMA will divide X into G sub-features across the channel dimensions direction for learning different semantics, where the groups-style can be donated by $X = [X_0, X_i, \ldots, X_{G-1}]$, $X \in \mathbb{R}^{C\times H\times W}$.

(2) Parallel Subnetworks. The large local receptive fields of neurons enable the neurons to collect multi-scale spatial information. Accordingly, EMA conducts that three parallel routes are exploited to extract attention weight descriptors of the grouped feature maps. Two of parallel routes is in 1x1 branch and the third one route is that the 3x3 branch. For capturing dependencies across all channels and relieving the computation budgets, they model the cross-channel information interaction at channel direction. To be more specific, there are two 1D global average pooling operations employed to encode the channel along two spatial directions respectively in 1x1 branch and only a single 3x3 kernel is stacked in 3x3 branch for capturing multi-scale feature representation.

Given the truth that there is no batch coefficient in the dimension of the convolution function for the normal convolution, the number of convolution kernels are independent of the batch coefficients of the forward operational inputs. Accordingly, the group G is reshaped and replaced into the batch dimension, and the shape of the input tensor is redefined as $C//G \times H \times W$. On one hand, the two encoded features are concatenated against the images height direction and share the same 1x1 convolution, without dimensionality reduction in the 1x1 branch. After factorize the outputs of 1x1 convolution into two vectors, two non-linear Sigmoid functions are employed to fit the 2D Binormial distribution

upon linear convolutions. In order to implement different cross-channel interaction features between two parallel routes of 1x1 branching, the attention maps of the two channels are aggregated within each group by a simple multiplication. On the other hand, the 3x3 branch captures the local cross-channel interaction via a 3x3 convolution to enlarge the feature space.

(3) Cross-spatial learning. EMA provides a cross-spatial information aggregation method at different spatial dimension direction for richer feature aggregation. Please note that here, two tensors are still introduced: one from the output of the 1x1 branch and the other from the output of the 3x3 branch. Then, global spatial information is encoded into the output of the 1x1 branch using 2D global average pooling. The output of the 3x3 branch is directly transformed to the corresponding dimension shape before the joint activation mechanism of channel features, i.e., $\mathbb{R}_1{}^{1\times\mathbb{C}//G} \times \mathbb{R}_3{}^{C//G\times HW}$. The 2D global pooling operation formula is shown in Equation (1):

$$z_c = \frac{1}{H \times W} \sum_{j}^{H} \sum_{i}^{W} x_c(i, j) \tag{1}$$

which is designed for encoding the global information and modeling the long-range dependencies. For efficient computation, the natural non-linear functions Softmax for 2D Gaussian maps is employed at the outputs of 2D global average pooling to fit the upon linear transformations. By multiplying the outputs of above parallel processing with matrix dot-product operations, the first spatial attention map can be derived. To observe this, it collects different scale spatial information in the same processing stage. Moreover, global spatial information in the 3x3 branch is encoded using 2D global average pooling and the 1x1 branch will be transformed to the correspond dimension shape directly before the joint activation mechanism of channel features, i.e., $\mathbb{R}_1{}^{1\times\mathbb{C}//G} \times \mathbb{R}_3{}^{C//G\times HW}$. After that, the second spatial attention map, which preserves the entire precise spatial positional information is derived. Finally, the output feature map within each group is calculated as the aggregation of the two generated spatial attention weight values followed by a Sigmoid function. It captures pixel-level pairwise relationship and highlights global context for all pixels. The final output of EMA is the same size of X, which is efficient yet effective to stack into modern architectures.

2) SPD-CONV MODULE

Convolutional Neural Networks (CNNs) have achieved significant success in various computer vision tasks such as image classification and object detection. However, their performance rapidly deteriorates in more challenging tasks characterized by lower image resolutions or smaller objects. This limitation arises from a common yet flawed design in existing CNN architectures, which involves the use of strided convolutions and/or pooling layers. These components lead to the loss of fine-grained information and inefficient learning of features. SPD-Conv [23] serves as a novel CNN module, capable of replacing each strided convolution and pooling layer. SPD-Conv is composed of an SPD layer and a non-strided convolution layer.

a: SPACE-TO-DEPTH (SPD)

The SPD component extends the (original) image transformation technique to downsample feature maps inside CNNs and throughout the entire CNN. As illustrated below, consider any intermediate feature map X of size $S \times S \times C_1$, where a series of sub-feature maps is extracted as

$$\begin{aligned}
&f_{0,0} = X[0:S:scale, 0:S:scale],\\
&f_{1,0} = X[1:S:scale, 0:S:scale], \ldots,\\
&f_{scale-1,0} = X[scale-1:S; scale, 0:S:scale];\\
&f_{0,1} = X[0:S:scale, 1:S:scale], f_{1,1}, \ldots,\\
&f_{scale-1,1} = X[scale-1:S:scale, 1:S:scale];\\
&\qquad\vdots\\
&f_{0,scale-1} = X[0:S:scale, scale-1:S:scale],\\
&f_{1,scale-1}, \ldots,\\
&f_{scale-1,scale-1} = X[scale-1:S:scale, scale-1:S:scale].
\end{aligned}$$

In general, for any given (original) feature map X, a sub-map $f_{x,y}$ is formed by all stripes $X(i + y)$ where $i + x$ and $i + y$ are divisible by a scaling factor. Thus, each sub-map undergoes downsampling on feature map X according to the scaling factor.

Next, these sub-maps are concatenated along the channel dimension, resulting in a feature map X', where its spatial dimensions are reduced by a scaling factor, and the channel dimension is increased by a scaling factor. In other words, SPD transforms the feature map $X(S, S, C_1)$ into an intermediate feature map $X'(\frac{S}{scale}, \frac{S}{scale}, scale^2C_1)$.

b: NON-STRIDED CONVOLUTION

After the SPD feature transformation layer, a non-strided (i.e., stride=1) convolutional layer with C_2 filters, where $C_2 < scale^2C_1$, is added, and further transforms $X'(\frac{S}{scale}, \frac{S}{scale}, scale^2C_1) \rightarrow X''(\frac{S}{scale}, \frac{S}{scale}, C_2)$. The use of non-strided convolution aims to preserve all discriminative feature information as much as possible.

3) DETECTION LAYER MODULE

The input size of YOLOv8 is set to 640 × 640, and the output layer modules P3, P4, and P5 are designed for detecting large, medium, and small objects, respectively. The maize seeds are characterized by their small volume and minimal variation within the image, so the detection belongs to small object detection. Based on this, the detection layers of YOLOv8n were simplified from large, medium, and small to medium and small. The 24th layer (Convolution layer), the 25th layer (Concat module), and the 26st layer (C2f module) of the feature fusion layers were removed. Reducing the number of detection layers effectively decreased model parameters

and computational complexity, thus leading to improved detection speed.

4) DETECTION LAYER MODULE

The YOLOv8 uses CIoU [24] bounding box loss function, which incorporates the overlap area, center point distance, and aspect ratio during bounding box regression, which enhanced the precision of regression localization. However, CIoU still suffers from the following problem: during the regression process of the prediction frame, if the height and width aspect ratios between the predicted box and the ground truth box are linearly proportional, the penalty for the relative proportions degenerates zero, thereby affecting the optimization of the network.

Due to the inherent imbalance in the dataset, there were inevitably low-quality samples in the training data. In order to address the issue of sample quality imbalance, based on WIoUv1, WIoUv3 was proposed to balance the ratio of low-quality to high-quality samples, thereby addressing challenges related to the unclear delineation of small targets and difficulties in detecting overlapping and occluded objects [25]. Consequently, it further enhanced the model's accuracy and overall detection performance. We replaced the loss function with WIoUv3, the spatial relationship between the ground truth box and the predictive box is shown in Figure 5.

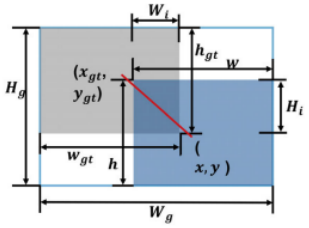


FIGURE 5. The spatial relationship between the ground truth box(red) and the predictive box(blue). w and h represent the width and height of the predicted box respectively; w_{gt} and h_{gt} represent the width and height of the ground truth box; W_i and H_i respectively indicate the width and height of the overlapping rectangle between predictive box and the ground true box; W_g and H_g are the width and height of the minimum enclosing rectangle of the predictive box and the ground truth box.

The calculation formula for WIoUv1 is shown in Equations (2) - (4):

$$L_{IoU} = 1 - IoU = 1 - \frac{W_i H_i}{wh + w_{gt}h_{gt} - W_i H_i} \tag{2}$$

$$L_{WIoUv1} = R_{WIoU} L_{IoU} \tag{3}$$

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \tag{4}$$

The abnormality degree of the anchor box β is represented by the ratio of L^*_{IoU} and $\overline{L_{IoU}}$, as shown in Formula (5):

$$\beta = \frac{L^*_{IoU}}{\overline{L_{IoU}}} \in [0, +\infty) \tag{5}$$

Applying β to WIoUv1 to construct nonmonotonic focusing coefficients yields WIoUv3 as shown in Eq. (6):

$$L_{WIoUv3} = rL_{WIoUv1}, r = \frac{\beta}{\delta\alpha^{\beta-\delta}} \tag{6}$$

where the mapping of outlier degree is β and gradient gain is r, which is controlled by the hyper-parameters α and δ; The value of a is 1.9 and the value of δ is 3.

C. MODEL TRAINING ENVIRONMENT CONFIGURATION

All model training and testing procedures were conducted on the same workstation (Windows 11 with 64-bit operating system, Intel (R) Core (TM) i9-13900HX, NVIDIA GeForce RTX 4060), PyTorch2.0.0, Python 3.8, and CUDA 11.7 was utilized for training acceleration. Detailed hyperparameters of the experiment are shown in Table1.

TABLE 1. Detailed hyperparameters of experiment.

Parameters	Value
image size	640×640
batch size	16
Classes	4
epochs	200
optimizer	Adam
iou	0.7
lr0	0.01
lrf	0.01
momentum	0.937
weight_decay	0.0005

D. MODEL PERFORMANCE EVALUATION METRICS

We used evaluation metrics in object detection models are confusion matrix, precision (P), recall (R), average precision (AP), mean average precision (mAP), and model size (MB). The confusion matrix usually has four indexes including True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). P and R are typically used to evaluate the model's ability to predict a specific category. mAP evaluates the model's detection performance over the entire dataset and is calculated as the average of the mean precision over all the categories. FPS is the number of images processed per second by the model and is used to measure the speed of detection. Table 2 provides a summary and brief description of the formulas.

III. RESULT AND ANALYSIS

A. IMPACT OF DIFFERENT ATTENTION MECHANISMS

We added an attention mechanism, which helped the network focus on the key information related to effective targets, extracted regions of interest, and further improved the network's detection capabilities. Various attention mechanisms, including CBAM, SE, EMA, SimAM, and CoTAttention, were individually added.

TABLE 2. Model performance evaluation index.

Evaluation Metrics	Formulas	Content
Precision(P)	$P = \frac{TP}{TP + FP}$	TP represents the number of samples that are actually positive but are predicted to be positive. FP represents the number of samples that are actually negative but are predicted to be positive.
Recall(R)	$R = \frac{TP}{TP + FN}$	FN represents the number of samples that are actually positive but predicted to be negative.
Average Precision (AP)	$AP_i = \int_0^1 P_i R_i dR_i$	AP_i represents the average accuracy of the model's detection for each category.
Mean Average Precision(mAP)	$mAP = \frac{1}{n} \sum_{i=1}^{n} AP_i$	mAP represents the average of multiple categories of AP.

TABLE 3. Impact of different attention mechanisms.

Model	P (%)	R (%)	mAP (%)	Parameters	FPS
YOLOv8	82.2	87.2	91.8	3006428	166.7
YOLOv8-CBAM	90.4	85.7	93.5	3027318	158.7
YOLOv8-SE	83.2	92.4	92.8	3014620	144.9
YOLOv8-SimAM	82.4	87.2	92.6	3006428	147.1
YOLOv8-CoTAttention	86.5	85.8	92.2	3583452	138.9
YOLOv8-EMA	89.1	88	94.4	3016796	153.9

Table 2 indicated that the addition of EMA led to a noticeable improvement in the network detection performance, with mAP increasing by 2.6%, exceeding the performance enhancements yielded by the SE, CBAM, SimAM, and CoTAttention attention mechanisms. EMA demonstrated superior efficiency in filtering effective feature information, so we adopted EMA to further improve the network's performance.

B. IMPACT OF ATTENTIONAL MECHANISM ADDED IN DIFFERENT LOCATIONS

The adoption of attention mechanisms brought multiple advantages, but the specific advantages depended on the location you add the attention module. We conducted a comparative experiment to evaluate various positions of EMA integration, investigating the impact of adding EMA at different positions on detection performance.

The experiment results are presented in Table 4. In Experiment A, the mAP increased, but the number of parameters also increased. This is because although the C2f_EMA module could better fuse shallow feature maps and deep feature maps, the introduction of the attention mechanism increased the depth of the model.

Experiment B achieved higher accuracy, but recall and mAP were lower. This was due to the addition of EMA to the small object detection layer aiding the model in accurately locating targets but not accurately obtaining the characteristic information of individual maize seeds.

In experiment C, the number of parameters increased, but the network detection performance was greatly improved.

TABLE 4. The impact of EMA adding in different locations.

Experiment	P (%)	R (%)	mAP (%)	Parameters	FPS
A	87.2	86.4	92.6	3016796	169.5
B	91.1	83.2	92.4	3007100	158.7
C	89.1	88	94.4	3016796	153.9

The C2f module added to the Backbone, forming the new C2f EMA module. B. Added to the small object detection layer (P3). C. Added to the front of the Backbone's SPPF.

This is because adding EMA to the front of the Backbone's SPPF helped the Backbone selectively focus on different parts of the input feature map. This made it easier for the model to learn complex image patterns and improved the accuracy of maize seed quality detection.

C. THE IMPACT OF DIFFERENT LOSS FUNCTIONS

In order to verify the superiority of WIoUv3, we conducted comparative experiments on YOLOv8 and I-YOLOv8 using WIoUv3 and some mainstream loss functions.

The experiment results are presented in Table 5. Both models exhibited the highest mAP when using WIoUv3 as the bounding box loss function, this indicated that using WIoUv3 as bounding box regression results in the best detection performance. Furthermore, the improved YOLOv8 achieved a 2.1% higher mAP when using WIoUv3 compared to CIoU, which demonstrated the effectiveness of introducing WIoUv3.

IV. DISCUSSION

A. MODEL PERFORMANCE ANALYSIS

We analyzed the performance of I-YOLOv8 from the training phase to the testing phase and evaluated its ability to detect the quality status of various types of maize seeds.

As depicted in Figure 6, as the iteration proceeded, the loss of I-YOLOv8n decreased and mAP improved correspondingly. After approximately 100 epochs of training, higher mAP and lower loss were achieved. Between 100 and 150 epochs, fluctuations occurred in the training process, this was due to the absence of a corresponding pre-trained model, which led to randomness in the weights during gradient

TABLE 5. The impact of different loss functions.

Model	loss function	P (%)	R (%)	mAP (%)
YOLOv8	DIoU	89.7	88.6	94.2
	GIoU	88.1	88.2	94
	Focal_ CIoU	90.7	88.2	94.5
	Focal_ EIoU	91.1	87.5	94.9
	CIoU	92.2	87.2	91.8
	WIoUv3	92.8	94.9	97.4
I-YOLOv8	DIoU	89.6	91.1	95.9
	GIoU	92.3	89.1	95.2
	Focal_ CIoU	92.3	90.2	97.3
	Focal_ EIoU	88.2	93	95.6
	CIoU	93.7	88.5	96.4
	WIoUv3	96.4	95.4	98.5

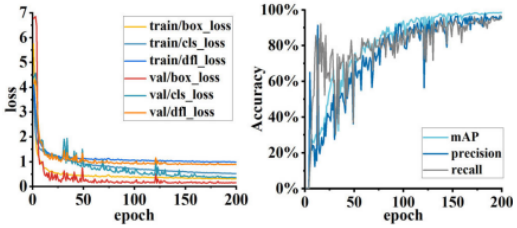


FIGURE 6. Loss and Accuracy Values during Training of I-YOLOv8.

TABLE 6. Performance Comparison between YOLOv8 and I-YOLOv8.

Model	P (%)	R (%)	mAP (%)	Parameters	Model Size (MB)	FPS
YOLOv8	82.2	87.2	91.8	3006428	5.94	166.7
I-YOLOv8	96.4	95.4	98.5	2264728	4.5	169.5

descent and made it challenging to achieve optimal training results. After 150 epochs, the model tended to stabilize with both mAP and loss maintaining a relatively stable state.

It was evident from Table 6 that the improved model exhibited enhanced performance compared to the original YOLOv8. The precision had increased by 14.2%, the recall had improved by 8.2%, and the mAP value had risen by 6.4%.

Figure 7 illustrated the detection results of YOLOv8 and I-YOLOv8 under different backgrounds. The left two images displayed the detection results of YOLOv8, while the right two images showed the results of I-YOLOv8. The red bounding box represented detected broken maize seeds, the pink bounding box represented healthy maize seeds, the orange bounding box represented mildewed maize seeds, and the yellow bounding box represented moth-eaten maize seeds. The confidence scores for each detected maize seed in the images were displayed to the right of the corresponding bounding

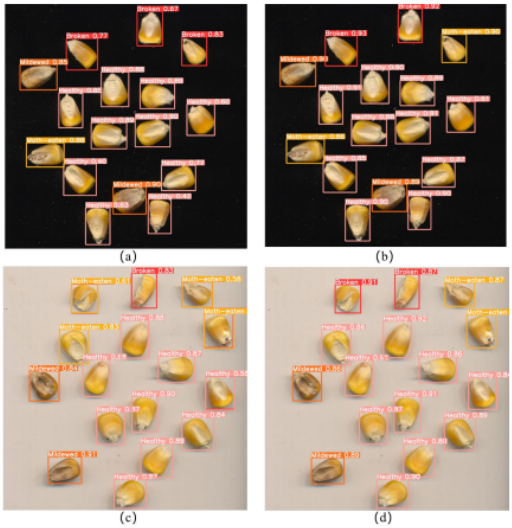


FIGURE 7. Comparison of Detection Results between YOLOv8 and I-YOLOv8. (a) Represents the detection result of YOLOv8 on a black background. (b) represents the detection result of I-YOLOv8 on a black background. (c) represents the detection result of YOLOv8 on a white background. (d) represents the detection result of I-YOLOv8 on a white background.

boxes. In Figure 7(a), the predictions were generally accurate, but it misidentified the infested maize seed in the upper right maizeer as a damaged one. In contrast, Figure7(b) correctly detected all maize seeds without any misidentifications or omissions. In Figure7(c), it misidentified the healthy and damaged maize seeds in the upper left maizeer as infested ones, while Figure 7(d) accurately detected all maize seeds without any misidentifications or omissions.

In summary, YOLOv8 could detect all seeds, which was why it was chosen as the baseline model. However, compared to the unimproved YOLOv8, I-YOLOv8 could better distinguish between damaged and moth-eaten maize seeds, making it more suitable for maize seed quality detection.This was because broken and moth-eaten corn seeds exhibited fundamental similarities in texture and contour edges, with only slight variations in color at the locations of insect holes or broken edges, making it challenging for the model to extract features. I-YOLOv8 introduces an efficient multi-scale attention mechanism (EMA) in the Backbone section, allowing it to focus more on the deep features of the seeds, thereby improving accuracy in identifying damaged and infested corn seeds. Therefore, through targeted improvements to YOLOv8, there was a successful enhancement in its performance in the detection of specific seeds.

B. ANALYSIS OF EXPERIMENT RESULTS UNDER DIFFERENT MODELS

In order to validate the performance of the improved YOLOv8, the enhanced I-YOLOv8 was compared with the

original YOLOv8, YOLOv5, and YOLOv5. Precision rate, recall rate, average precision mean, and detection speed were used as performance evaluation metrics.

As can be seen from Figure 8, the accuracy, recall, and mAP values of the improved YOLOv8 were 96.4%, 95.4%, and 98.5% respectively, which surpassed the other three networks in performance. In terms of model detection speed, I-YOLOv8 processes each image in 0.059 seconds (169.5 fps), which was the fastest among the four networks. Additionally, the model size of I-YOLOv8 was only 4.5 MB and the number of parameters was 2,264,728, both of them were far smaller than the other three network models. I-YOLOv8 could still achieve high detection accuracy and speed even in small size. This indicated that the improved network model exhibits superior recognition capabilities for maize seed quality detection compared to other detection models.

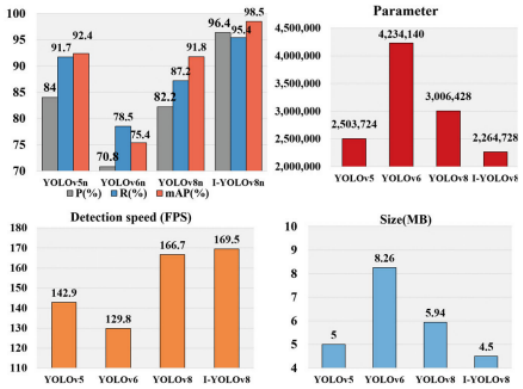


FIGURE 8. Performance comparison of four object detection networks.

In general, increasing the depth and width of a neural network model can improve its performance to a certain extent. However, when performance reaches a certain level, further increasing the depth and width of the network may no longer lead to performance improvement, it may lead to problems such as gradient instability, network degradation, a significant increase in computation complexity and the number of parameters. This study introduced attention mechanisms and SPD-Conv modules into YOLOv8, enhancing the network's performance to some extent but also increasing the number of parameters. Therefore, the study chose to prune the model on this basis, reducing redundancy by eliminating a significant amount of irrelevant semantic information in the model. This is why I-YOLOv8 achieves superior detection speed and network size compared to the other three models while maintaining the highest accuracy.

As depicted in Figure 9, we used a confusion matrix for visual performance evaluation of the four different models. The color of the matrix represented the effectiveness of predictions, the darker the color of the matrix block, the higher the probability of occurrence. The deeper the colors of the blocks along the diagonal of the matrix, the higher the predictive accuracy in this category.

It can be seen from Figure 9 that the color of the diagonal matrix blocks of I-YOLOv8 was darker than that of the other models, and its overall correct recognition rate was higher than other models in the experiment. The correct recognition rates of I-YOLOv8 for healthy and mildewed maize seeds were almost the same as that of other models, but the correct recognition rate in broken and moth-eaten maize seeds was much higher than that of other models. This is because healthy and mildewed maize seeds exhibit noticeable differences in color and texture, while broken and moth-eaten maize seeds are very similar in the contours of the missing parts, with only slight differences in color. In summary, the overall performance of I-YOLOv8 significantly outperforms the other models in the study.

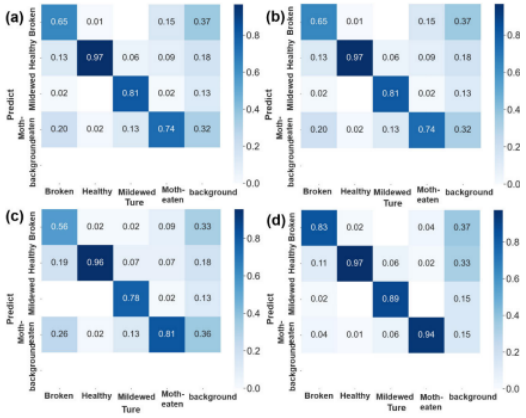


FIGURE 9. Confusion matrix for different models (a)YOLOv5 (b)YOLOv6 (c)YOLOv8 (d)I-YOLOv8.

C. ABLATION STUDY

To validate the effectiveness of each proposed improvement strategy in this study, we designed ablation studies based on the baseline model YOLOv8n to evaluate its effectiveness. The results of the ablation experiments were presented in Table 7, where "√" indicated the usage of the corresponding module, while its absence indicated the non-usage of the module.

Table 7 demonstrated that each improvement strategy had effectively enhanced the detection performance. Experiment 2 introduced the efficient multi-scale attention mechanism EMA, resulting in a 2.6% increase in mAP, while the number of parameters only increased by 0.34%. This indicated that EMA could efficiently retain information on each channel, and the number of parameters was not significantly increased while ensuring accuracy. Experiment 3 added the SPD-Conv module on the basis of Experiment 2, and the mAP increased by 1.2%, which solved the problem of a loss of

TABLE 7. Detection results after introducing different improvement strategies.

Experiment	EMA	SPD-Conv	P3,p4	WIOU	P(%)	R(%)	mAP(%)	Parameters	FPS
1					82.2	87.2	91.8	3006428	158.7
2	√				89.1	88	94.4	3016796	153.9
3	√	√			91.8	91.1	95.6	3277916	142.9
4	√	√	√		93.7	88.5	96.4	2264728	163.9
5	√	√	√	√	96.4	95.4	98.5	2264728	163.9

fine-grained information caused by the strided convolution or pooling layer in the existing CNN architecture, and could effectively reduce the missed detection rate for small targets.

To mitigate the increased parameters resulting from the addition of attention mechanisms and the SPD-Conv module, a large object detection layer was reduced to decrease a large amount of irrelevant semantic information in the model, which enhanced network speed and reduced network size. Experiment 4 reduced a large target detection layer based on Experiment 3, the number of parameters was reduced by 30.9%, the FPS was increased by 20f/s, and the mAP value was also improved.

Experiment 5 was the improved model proposed in this article, it was based on Experiment 4 and the loss function was replaced with WIoUv3. WIoUv3 was used to weigh the ratio of low-quality samples to high-quality samples, which increased mAP by 1.2% and solved the problem of small target are blurry and difficult to detect. Compared to the baseline model, the number of parameters decreased by 24.7%, mAP increased by 6.7%, recall improved by 8.2%, precision increased by 14.2%, and FPS decreased by 5.2f/s, this indicated that the improved network had excellent performance.

V. CONCLUSION

In response to the characteristics of dense image distribution and small targets in maize seed object detection, we proposed a lightweight maize seed quality detection model based on the improved YOLOv8: I-YOLOv8.

I-YOLOv8 achieved mAP of 98.5%, a 6.7% increase compared to YOLOv8, and an average recognition speed of 163.9 frames per second, a 5.2 frames per second improvement over YOLOv8. Furthermore, when compared to YOLOv5, YOLOv6, and YOLOv8 network models, I-YOLOv8 outperformed these three models in various aspects, and showed a significant improvement in detection performance. The improved model can provide more efficient computing performance and rapid real-time decision-making in agricultural deployment, helping to realize intelligent agricultural management in farm equipment, airborne equipment and edge computing, and improve production efficiency and resource utilization efficiency.

The combination of deep learning and machine vision can achieve non-destructive and efficient identification of corn seeds. The application of deep learning and machine vision in quality detection for maize seeds is expected to bring about significant transformation in agricultural production, promoting the development of a more intelligent, efficient, and sustainable direction for agriculture. In the next steps, we will further optimize the model and increase the variety and quantity of samples to enhance the model's applicability.

ACKNOWLEDGMENT

The authors thank the Shandong Academy of Agricultural Sciences for providing experimental support.

REFERENCES

[1] O. Erenstein, J. Chamberlin, and K. Sonder, "Estimating the global number and distribution of maize and wheat farms," *Global Food Secur.*, vol. 30, Sep. 2021, Art. no. 100558.
[2] F. Guzzon, L. W. A. Rios, G. M. C. Cepeda, M. C. Polo, A. C. Cabrera, J. M. Figueroa, A. E. M. Hoyos, T. W. J. Calvo, T. L. Molnar, L. A. N. León, T. P. N. León, S. L. M. Kerguelén, J. G. O. Rojas, G. Vázquez, R. E. Preciado-Ortiz, J. L. Zambrano, N. P. Rojas, and K. V. Pixley, "Conservation and use of Latin American maize diversity: Pillar of nutrition security and cultural heritage of humanity," *Agronomy*, vol. 11, no. 1, p. 172, Jan. 2021.
[3] K. Tu, S. Wen, Y. Cheng, T. Zhang, T. Pan, J. Wang, J. Wang, and Q. Sun, "A non-destructive and highly efficient model for detecting the genuineness of maize variety 'JINGKE 968' using machine vision combined with deep learning," *Comput. Electron. Agricult.*, vol. 182, Mar. 2021, Art. no. 106002.
[4] M. Kharbach, M. Alaoui Mansouri, M. Taabouz, and H. Yu, "Current application of advancing spectroscopy techniques in food analysis: Data handling with chemometric approaches," *Foods*, vol. 12, no. 14, p. 2753, Jul. 2023.
[5] T. U. Rehman, M. S. Mahmud, Y. K. Chang, J. Jin, and J. Shin, "Current and future applications of statistical machine learning algorithms for agricultural machine vision systems," *Comput. Electron. Agricult.*, vol. 156, pp. 585–605, Jan. 2019.
[6] G. Chen, Y. Meng, J. Lu, and D. Wang, "Research on color and shape recognition of maize diseases based on HSV and OTSU method," in *Proc. Int. Conf. Comput. Comput. Technol. Agricult.*, vol. 509. Cham, Switzerland: Springer, 2016, pp. 298–309.
[7] K. Kiratiratanapruk and W. Sinthupinyo, "Color and texture for corn seed classification by machine vision," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, Dec. 2011, pp. 1–5.
[8] B. B. Traore, B. Kamsu-Foguem, and F. Tangara, "Deep convolution neural network for image recognition," *Ecol. Informat.*, vol. 48, pp. 257–268, Nov. 2018.
[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 580–587.
[10] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[12] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
[13] C. Zhao, L. Quan, H. Li, R. Liu, J. Wang, H. Feng, Q. Wang, and K. Sin, "Precise selection and visualization of maize kernels based on electromagnetic vibration and deep learning," *Trans. ASABE*, vol. 63, no. 3, pp. 629–643, 2020.

[14] H. O. Velesaca, R. Mira, P. L. Suárez, C. X. Larrea, and A. D. Sappa, "Deep learning based corn kernel classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 294–302.
[15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 21–37.
[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
[17] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2980–2988.
[18] Z. Liu and S. Wang, "Broken corn detection based on an adjusted YOLO with focal loss," *IEEE Access*, vol. 7, pp. 68281–68289, 2019.
[19] X. Li, Y. Du, L. Yao, J. Wu, and L. Liu, "Design and experiment of a broken corn kernel detection device based on the YOLOv4-tiny algorithm," *Agriculture*, vol. 11, no. 12, p. 1238, Dec. 2021.
[20] R. P. Thangaraj Sundaramurthy, Y. Balasubramanian, and M. Annamalai, "Real-time detection of fusarium infection in moving corn grains using YOLOv5 object detection algorithm," *J. Food Process Eng.*, vol. 46, no. 9, Jun. 2023, Art. no. e14401.
[21] Q. Wang, H. Yang, Q. He, D. Yue, C. Zhang, and D. Geng, "Real-time detection system of broken corn kernels based on BCK-YOLOv7," *Agronomy*, vol. 13, no. 7, p. 1750, Jun. 2023.
[22] D. Ouyang, S. He, G. Zhang, M. Luo, H. Guo, J. Zhan, and Z. Huang, "Efficient multi-scale attention module with cross-spatial learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.
[23] R. Sunkara and T. Luo, "No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects," in *Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases (ECML PKDD)*, 2023, pp. 443–459.
[24] Z. Tong, Y. Chen, Z. Xu, and R. Yu, "Wise-IoU: Bounding box regression loss with dynamic focusing mechanism," 2023, *arXiv:2301.10051*.
[25] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2020, vol. 34, no. 7, pp. 12993–13000.

SIQI NIU was born in 2000. She is currently pursuing the master's degree in agricultural engineering and information technology with Qingdao Agricultural University.

XIAOLIN XU was born in 1999. He is currently pursuing the master's degree in agricultural engineering and information technology with Qingdao Agricultural University.

AO LIANG was born in 1999. He is currently pursuing the master's degree in agricultural engineering and information technology with Qingdao Agricultural University.

YULIANG YUN received the Ph.D. degree in agricultural engineering from China Agricultural University, China. He is currently an Associate Professor with the School of Mechanical and Electrical Engineering, Qingdao Agricultural University, China. His current research interests include agricultural artificial intelligence and decision support systems.

LI LI received the Ph.D. degree in agricultural electrification and automation from China Agricultural University, Beijing, China, in 2007. She is currently an Associate Professor with the Information and Electrical Engineering College, China Agricultural University. Her current research interests include agricultural artificial intelligence and greenhouse environmental regulation.

FENGQI HAO received the master's degree in computer system architecture from Shandong University, China, in 2006. He is currently an Associate Researcher with the Qilu University of Technology, Jinan, China. His current research interests include intelligent control, computer vision, and deep learning.

JINQIANG BAI received the Ph.D. degree from Beihang University, Beijing, China, in 2020. He has been a Research Assistant with the Qilu University of Technology, Jinan, China, since 2020. His current research interests include computer vision, deep learning, and intelligent agriculture.

DEXIN MA received the Ph.D. degree in computer science and technology from the Beijing University of Posts and Telecommunications, Beijing, China, in 2014. He is currently a Professor with the Intelligent Agriculture Institute, Qingdao Agricultural University, China. His current research interests include agricultural artificial intelligence and agricultural informatization.

附件 3：一级学报论文一作

2024 年 3 月 第 39 卷第 3 期 中国粮油学报 Journal of the Chinese Cereals and Oils Association Vol. 39, No. 3 Mar. 2024

基于改进 CBAM 注意力机制的 MobileNetV2 玉米种子品种识别研究

牛思琪[1]，马　睿[1]，许晓琳[1]，梁　赦[1]，穆春华[3]，许金普[1]，马德新[1,2]
（青岛农业大学动漫与传媒学院[1]，青岛　266109）
（青岛农业大学智慧农业研究院[2]，青岛　266109）
（山东省农业科学院玉米研究所[3]，济南　250100）

摘　要：玉米是我国主要粮食作物，有较高的营养价值和经济价值。不同的地域环境适宜种植的玉米品种不同，但由于玉米种子在外形方面存在的差异较小，所以仅凭肉眼很难对其进行快速准确的识别。为实现玉米种子品种的准确识别，研究采集了 9 种玉米种子图像共 2792 张建立数据集，并按照 7∶2∶1 的比例随机划分训练集、验证集和测试集。将注意力机制 CBAM 引入轻量化模型 MobileNetV2，对 CBAM 的串行方式进行改进，构建一个新型注意力模块 E_CBAM，并通过对比不同的压缩比，选出效果最佳的压缩比为 4，提出了 E_CBAM_MobileNetV2 模型。实验表明 E_CBAM_MobileNetV2 的准确率为 98.18%，相较于 MobileNetV2 提高了 5.45%。

关键词：图像分类；玉米种子；MobileNetV2；CBAM
DOI：10.20048/j.cnki.issn.1003－0174.000697
中图分类号：S126　　**文献标识码**：A　　**文章编号**：1003－0174(2024)03－0159－07
网络首发时间：2023－12－06 11：03：13
网络首发地址：https://link.cnki.net/urlid/11.2864.TS.20231205.0958.002

Mobile NetV2 Maize Seed Variety Recognition Based on Improved Attention Mechanism CBAM

Niu Siqi[1], Ma Rui[1], Xu Xiaolin[1], Liang Ao[1], Mu Chunhua[3], Xu Jinpu[1], Ma Dexin[1,2]
(College of Animation and Media, Qingdao Agricultural University[1], Qingdao　266109)
(Intelligent Agriculture Institute, Qingdao Agricultural University[2], Qingdao　266109)
(Corn Research Institute, Shandong Academy of Agricultural Sciences[3], Jinan　250100)

Abstract: Maize is a main food crop in China, with high nutritional value and economic value. Different geographical environments are applicable for different varieties of maize for planting, however, due to the slight differences in appearance of different varieties, it is difficult to quickly and accurately identify different varieties with the naked eye. In order to realize the accurate identification of maize seed varieties, total 2 792 images of 9 kinds of maize seeds were collected in this study to establish a data set, and training set, verification set and test set were randomly divided according to a ratio of 7∶2∶1. In this study, the attention mechanism CBAM was introduced into the lightweight model MobileNetV2, the serial mode of CBAM was improved, a new attention module E_CBAM was built, and different reduction ratios were compared, the best reduction ratio of 4 was selected, and E_CBAM_MobileNetV2 model was proposed. Experiments indicated, that the accuracy of E_CBAM_MobileNetV2 was 98.18%, 5.45% higher than that of MobileNetV2.

Key words: image classification; maize seed; MobileNetV2; CBAM

基金项目：山东省自然科学基金项目(ZR2022MC152)，山东省高等学校青创人才引育计划项目(202202027)，中央引导地方科技发展专项计划项目(23－1－3－6－zyyd－nsh)
收稿日期：2023－04－05
第一作者：牛思琪，女，2000 年出生，硕士，农业信息化、深度学习图像识别研究，291758546@qq.com
通信作者：马德新，男，1977 年出生，教授，农业人工智能、深度学习相关研究，madexin@163.com

玉米(*Zea mays* L.)是世界上分布最广的作物之一,具有较高的营养价值和经济价值,是我国的主要粮食作物,也是我国畜牧业主要的饲料来源。不同的地域环境对于适宜种植的玉米品种不同,传统的种子品种检验的方法如色谱分析技术、卡尺测量等,均存在着耗时长等问题,而且由于其外形方面存在的差异较小,所以凭借肉眼很难对其进行快速准确的识别。

计算机视觉技术能够在不破坏玉米种子的前提下对种子品种进行检测识别,具有无损、快速、实时等特点。周鸿达等[1]为了快速检测玉米等级,选取 4 种等级的玉米种子,利用 BP 神经网络对其分类,总体识别率在 90% 以上。程洪等[2]选取了 3 个玉米品种作为实验对象,利用支持向量机(SVM)算法对其训练,准确率达到了 92.3%。

现阶段,在机器学习中应用最普遍的深度学习技术是卷积神经网络(Convolutional Neural Networks, CNN)[3]。被广泛应用于图像分类、目标检测等方面[4]。自 2012 年以来,CNN 模型开始出现,如 AlexNet[5]、VGGNet[6]、GoogLeNet[7]、ResNet[8]、DenseNet[9]、MobileNet[10]、ShuffleNet[11]等。Huang[12]将卷积神经网络(CNN)和迁移学习引入玉米种子的质量分类,同时与机器学习算法进行了比较。结果表明,深度学习算法明显优于机器学习算法,GoogLeNet 准确率为 95%,SURF + SVM 的准确率为 79.2%。Xu 等[13]提出了基于 ResNet 改进的新模型(P – ResNet),为玉米种子分类提供了一种方法,选取 5 个品种的玉米种子作为实验对象,准确率达到 99.70%。Javanmardi 等[14]提出使用深度卷积神经网络(CNN)提取特征,再利用人工神经网络(ANN)对提取的特征进行分类。使用 CNN – ANN 对玉米种子进行分类识别准确率达到 98.1%。

近年来,轻量级网络发展迅速,不仅具有计算量少、准确率高的特点,而且广泛适用于移动设备。Shahi 等[15]提出了一种基于 MobileNetV2 的新型注意力卷积模块来对水果进行分类,在 3 个数据集上分别达到了 95.75%、96.74% 和 96.23% 的稳定分类精度。叶名炀等[16]针对传统玉米病害的问题,提出一种基于迁移学习和 MobileNetV2 模型的识别方法,来识别玉米病害中常见的 3 类,分别是灰斑病、锈病、大斑病,同时对比传统神经网络 LeNet,实验结果表明,在 Kaggle 数据集上 MobileNetV2 的平均识别率达到 96.94%,较 LeNet 高出 1.37%。易才键等[17]为实现智能垃圾分类,提出一种轻量型的生活垃圾图像分类模型 DG – MobileNetV2。该模型以轻量级网络 MobileNetV2 为基础,通过引入双尺度深度卷积模块、通道混洗技术、调整宽度因子等手段来压缩模型。实验表明,该分类模型在某公司发布的生活垃圾分类数据集上的分类准确率达到 90.58%,具有良好的分类效果。

随着深度学习的发展,许多研究者通过引入注意力机制来提高网络的性能。Zhang 等[18]为了准确识别鱼群的摄食行为,使用 MobileNetV2 提取鱼群图像特征,并建立了基于 SENet(挤压和激励网络)的特征加权网络,所提出的方法在真实鱼群图像上进行了测试,获得了 97.76% 的准确率。Selvraju 等[19]提出了一种基于 MobileNetV2 药品图像识别分类的模型,在具有残差边的瓶颈结构中添加 CBAM,提高训练效率以及减少模型复杂度,实验结果表明:该模型对实验的 95 种药品的图像识别分类的准确率,达到 89%,较改进前提高了 5%。

本研究在轻量化模型 MobileNetV2 的基础上,构建一个新型注意力模块 E_CBAM,提出了 E_CBAM_MobileNetV2 模型,解决了 CBAM 串行连接 2 种模块相互影响的问题。采集了 9 种玉米种子图像共 2 792张建立了玉米种子识别的数据集。在本研究构建的数据集的基础上,进行了大量实验。实验结果表明,所提出的 E_CBAM_MobileNetV2 模型能够有效地对玉米种子进行分类,且明显优于 MobileNetV2 模型。本研究为玉米种子的高效、便捷、快速识别提供了参考。

1 数据集构建

1.1 数据来源及获取

研究共选取 9 种玉米种子 2 792 粒,且玉米种子籽粒饱满未破损,分别是 AOYU116、JINGNIAN1、ZHENGDAN958、KENUO58、TIEYAN、XJH、YUNYU、16DX531、BT506。详细品种名称、数量及材料产地见表 1。

表 1 实验材料

种类	数量	产地
AOYU116	248	山东省济南市
JINGNIAN1	244	北京市丰台区
ZHENGDAN958	384	河南省郑州市
KENUO58	382	山东省潍坊市
TIEYAN	402	山东省潍坊市
XJH	378	辽宁省沈阳市
YUNYU	256	山东省潍坊市
16DX531	186	山东省济南市
BT506	312	山东省济南市

在实验室自然光照的条件下，将玉米种子放置在以黑色植绒布为背景搭建的工作台上，每次放置 16 粒种子，采用 80D 型数码相机进行拍摄。

1.2　图像预处理

由于研究要对单粒玉米种子进行识别，因此需要对多粒图像进行分割。如图 1 所示，通过 Python 和 OpenCV 对多粒玉米种子图像进行区域分割。首先对图像预处理，通过二值化提取图像信息，采用双边滤波加填充进行降噪，然后通过腐蚀膨胀去除边缘颗粒。最后，通过轮廓检测算法提取单粒玉米种子的形态轮廓，再通过画框将其分割出来。

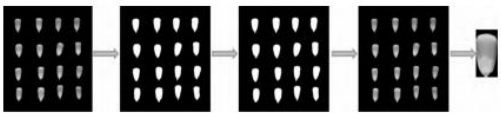


图 1　玉米籽粒部分分割过程

1.3　数据集的划分与增强

为了保证数据的独立性，避免某张图片扩充之后同时在训练集和测试集出现，从而生成虚假的准确度，将每类玉米种子图像按照 7∶2∶1 随机划分为训练集、验证集、测试集，即训练集图像共1 959张，验证集图像共 558 张，测试集图像共 275 张。

为了进一步提高模型的鲁棒性和泛化能力，通过人为扩大训练图像数量，对已采集的数据集进行数据增强。主要包括采用旋转、水平翻转、位置平移、错切变换、放缩等操作。

2　实验方法

2.1　改进注意力机制(E_CBAM)

注意力机制是模拟人类视觉的一种机制，即关注重要特征，抑制不必要的特征。CBAM(Convolutional Block Attention Module)是一种轻量化的卷积注意力模块。将 CBAM 引入卷积神经网络中能够较大幅度的提高模型的性能，而且引入它带来的参数量和运算量非常的小，可以嵌入到大部分的卷积神经网络。

特征图的不同维度所携带的信息不同，通道维度侧重于特征抽象的表达，而空间维度则更侧重于物体的位置信息，CBAM 融合了通道注意力模块(Channel Attention Module，CAM)和空间注意力模块(Spatial Attention Module，SAM)。这 2 个子模块以串行的方式组合在一起，序列化的在通道维度和空间维度上产生注意力特征图信息。利用通道间的特征关系生成通道注意图，进而连接到空间注意力得到最具信息量的特征图。整体结构如图 2 所示。

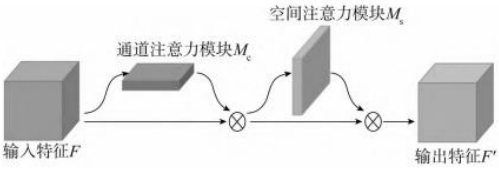


图 2　CBAM 结构图

假设中间特征映射 F 作为输入，先得到通道注意映射 M_c，再经过空间注意力模块得到空间注意映射 M_s，最终进行加权得到特征 F'。其公式为：

$$\begin{cases} F = M_c(F) \otimes F \\ F' = M_s(F') \otimes F' \end{cases} \tag{1}$$

式中：$M_c(F)$为 F 经过通道注意力的输出权值；$M_s(F')$为 F'经过空间注意力的输出权值；⊗为特征图加权乘法运算符号。

CBAM 虽然将相对重要的通道注意力模块放在前面，但是仍在一定程度上影响排在后面的空间注意力模块所学习到的特征。因此，为使 2 种注意力模块都直接输入原始的特征，研究对 CBAM 进行改进，由“串行”改为“并行”连接，从而得到无需关注空间注意力和通道注意力顺序的注意力机制，即 E_CBAM。整体结构如图 3 所示。

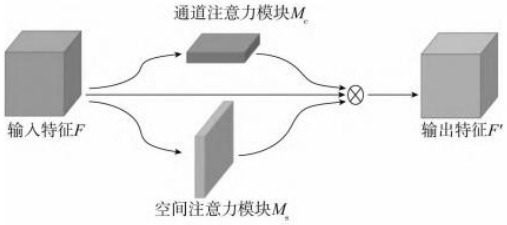


图 3　E－CBAM 结构图

E_CBAM 先由输入特征图 F 分别经过通道注意力和空间注意力得到相应权值，最后将权值直接与原始输入特征图 F 加权得到输出特征图 F'。其公式为：

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

式中：$M_s(F)$为 F 经过通道注意力的输出权值。

2.2　E_CBAM_MobileNetV2 模型结构

MobileNetV2 中添加了 InvertedResiduals 和 Lin-

earBottlenecks，具有较少的网络参数数量和更低的运算成本，延续了在 MobileNetV1 的深度可分离卷积操作，其所做的创新是用大量倒残差块和线性瓶颈层对图像的特征进行提取，实验证明，使用线性层是至关重要的，可以有效防止非线性破坏特征信息。

为了提高 MobileNetV2 模型在玉米种子品种识别方面的性能，将 MobileNetV2 与本研究提出的 E_CBAM 模块结合，构建了 E_CBAM_MobileNetV2 模型，具体结构网络如图 4 所示。

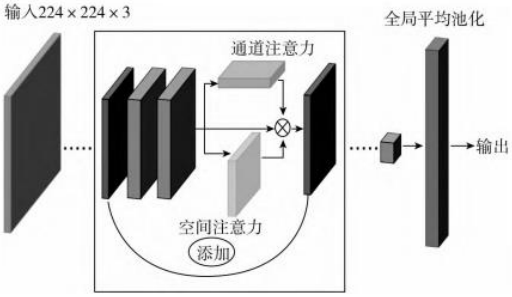


图 4 E_CBAM_MobileNetV2 模型结构图

2.3 实验环境

实验的训练与测试环境相同，均为 Windows11 操作系统，使用基于 Tensorflow 为后端运行的 Keras 深度学习框架，来实现整个模型训练与测试过程，实验过程中的超参数如表 2 所示。

表 2 实验详细超参数

参数	值
Image size	224 × 224
Batch size	32
Epochs	200
Optimizer	Adam
Learning rate (α0)	0.000 01
Loss	Categorical cross - entropy

2.4 模型性能评价指标

对于玉米种子分类问题，评估指标决定了模型的好坏。选用混淆矩阵作为性能的评估指标，通常有 4 个指标包括真阳性(TP)、真阴性(TN)、假阳性(FP)和假阴性(FN)。通过指标可以展现出样本中的真实类别和预测的类别。基于混淆矩阵，可以得到更高级的分类指标：准确率(Accuracy)，精确率(Precision)，召回率(Recall)、F1 - score。如表 3 所示。

表 3 模型性能评价指标

性能指标	内容
Accuracy	真阳性(TP)和真阴性(TN)与全部值的比值 $\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}\times 100\%$
Precision	真阳性(TP)与相关结果总量的比值，即 TP 和假阳性(FP)之和。 $\text{Presicion}=\frac{TP}{TP+FP}\times 100\%$
Recall	真阳性(TP)和假阴性(FN)中 TP 的分数。 $\text{Recall}=\frac{TP}{TP+FN}\times 100\%$
F_1 - score	精确率和召回率的调和平均值。 $F_1\text{ - score}=\frac{2\times \text{Presicion}\times \text{Recall}}{\text{Presicion}+\text{Recall}}\times 100\%$

3 结果与讨论

3.1 不同网络模型对比

选取 MobileNetV2、SqueezeNet、ResNet50、DenseNet、ShuffleNetV2 共 5 种卷积神经网络模型进行训练，其准确率如表 4 所示。从实验结果来看，MobileNetV2 的实验效果最好，准确率为 92.73%，是 5 种网络中最高的，较 SqueezeNet、ResNet50、DenseNet121、ShuffleNetV2 分别高出 2.91%、17.09%、1.46%、10.91%。

表 4 不同网络模型对比实验

序号	模型	准确率
1	MobileNetV2	92.73%
2	SqueezeNet	89.82%
3	ResNet50	75.64%
4	DenseNet121	91.27%
5	ShuffleNetV2	81.82%

3.2 MobileNetV2 加入不同注意力模块对比

为了提高模型的准确率，选择合适的注意力模块，向 MobileNetV2 网络模型中加入不同注意力进行比较，分别是 Channel attention、Spatial attention、CBAM、R_CBAM、E_CBAM，实验结果如表 5 所示。将添加不同注意力模块后的 MobileNetV2 网络模型与未加注意力的 MobileNetV2 网络模型进行对比，可以看出，无论添加空间注意力或通道注意力，还是将其串行或并行，其准确率均高于未加注意力的 MobileNetV2 网络模型。

从规模和参数量上看，Spatial attention 最小，这是因为 Spatial attention 只保存 1 个 7 × 7 的卷积核，因此准确率也是提升最少的。而含有 Channel attention 模块的规模和参数量较大，但准确率较高，这是由于 Channel attention 存在全连接层能够处理平均池化和最大池化的结果。

从准确率来看，CBAM 注意力的效果并不是很好，因为是“串行”连接，后面的空间注意力学习到的是的通道注意力处理过的特征图。同为“串行”连接的将空间注意力放在前面，结果表明 R_CBAM 的准确率略高一些。但表现最好的还是研究提出的“并行”混合注意力模块 E_CBAM，在模型大小和参数量相同的情况下，准确率比 CBAM、R_CBAM 分别高出了 2.18%、0.73%。

表 5　MobileNetV2 加入不同注意力模块对比实验

序号	模型	准确率/%	模型容量/MB	参数量
1	MobileNetV2	92.73	22.0	1 871 945
2	MobileNetV2 + Channel attention	94.55	27.1	2 264 377
3	MobileNetV2 + Spatialattention	95.64	24.2	2 036 092
4	MobileNetV2 + CBAM	94.91	27.3	2 266 060
5	MobileNetV2 + R_CBAM	96.36	27.3	2 266 060
6	MobileNetV2 + E_CBAM	97.09	27.3	2 266 060

3.3　E_CBAM_MobileNetV2 使用不同压缩比对比

由于注意力模块 E_CBAM 的加入，会增加模型的复杂度、消耗更多的计算时间，且通过实验可知，注意力模块 E_CBAM 中的大部分参数来自 Channelattention。因此，为保证 MobileNetV2 的轻量化优势，对 E_CBAM 的 Channel attention 的压缩比（2、4、8、16、32）进行了调参实验，结果如图 5 所示，从准确率来看当压缩比为 4 时训练集准确率明显更高，而且收敛也更快。

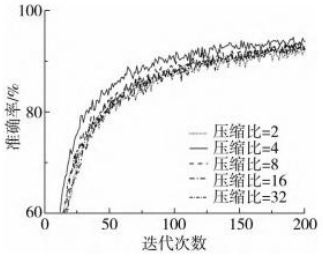


图 5　E_CBAM_MobileNetV2 使用不同压缩比准确率对比

从表 6 中也可以看到，当压缩比为 4 时，准确率为 98.18%，尽管模型的规模相比之前压缩比为 2 时有所减小，但模型的准确率却反而有所提升。而当压缩比8、16、32时，虽然规模减小，但准确率也明显下降远不如压缩比为 4 时。因此，使用 4 作为 E_CBAM 中 Channel attention 全连接层的压缩比是较好的选择。

表 6　E_CBAM_MobileNetV2 使用不同压缩比对比结果

压缩比	准确率/%	模型容量/MB	参数量
2	97.09	27.3	2 266 060
4	98.18	26.0	2 151 828
8	95.64	25.4	2 094 712
16	96.00	25.0	2 066 105
32	96.73	24.9	2 051 785

3.4　E_CBAM_MobileNetV2 对品种鉴定分析结果

如图 6 所示，使用混淆矩阵对研究提出的 E_CBAM_MobileNetV2 模型进行可视化性能评估。

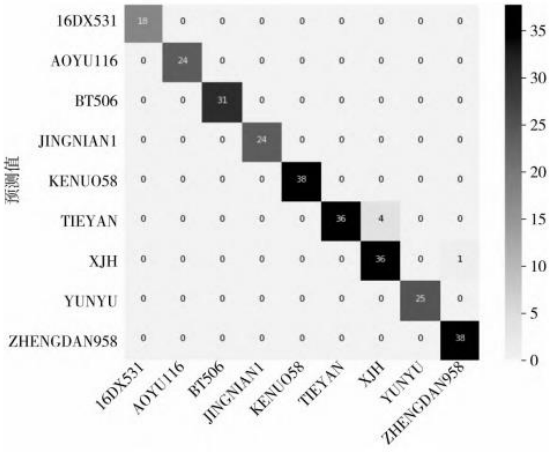


图 6　6 E_CBAM_MobileNetV2 混淆矩阵结果

从混淆矩阵中可以看出，在 XHJ 品种识别上产生 4 个错误识别，均被识别成 TIEYAN；ZHENGDAN958 产生 1 个错误识别，被识别成 XHJ。这主要是因为它们在特征上存在一些重叠。例如，XHJ 和 TIEYAN 虽然在颜色和光泽上有略微不同，但其籽粒的纹理以及顶部形状是极其相近的。XHJ 和 ZHENGDAN958 在尖端部分的凹陷有略微的不同，但外形极其相似，籽粒饱满均为金黄色，且表面光滑有光泽感。

此外，混淆矩阵中描述的性能指标如表 7 所示，包括准确率、精确率、召回率、F_1 – score。

表 7　E_CBAM_MobileNetV2 性能评估结果

类别	准确率/%	精确率/%	召回率/%	F_1 – score/%
16DX531	100.00	100.00	100.00	100.00
AOYU116	100.00	100.00	100.00	100.00
BT506	100.00	100.00	100.00	100.00
JINGNIAN1	100.00	100.00	100.00	100.00
KENUO58	100.00	100.00	100.00	100.00
TIEYAN	98.55	100.00	90.00	94.74
XJH	99.64	90.00	97.30	93.51
YUNYU	100.00	100.00	100.00	100.00
ZHENGDAN958	99.64	97.44	100.00	98.71

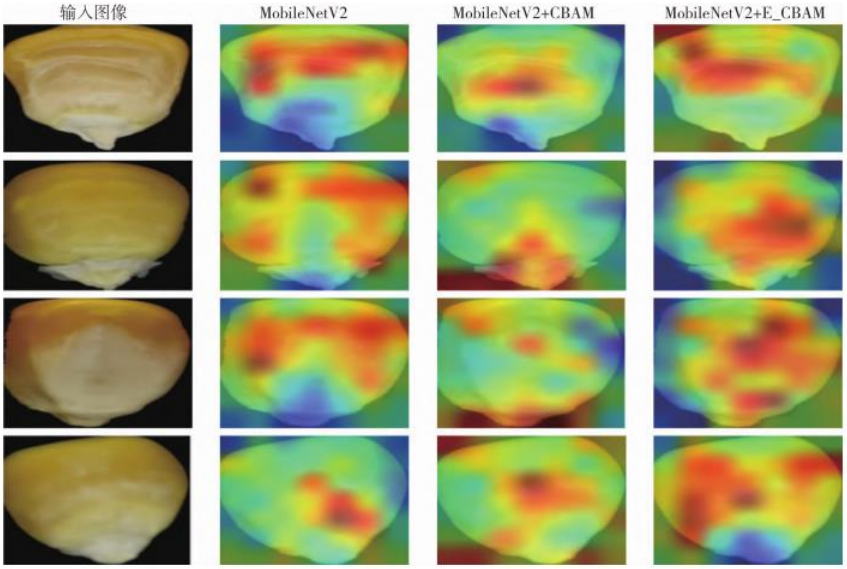


图 7 MobileNetV2 及嵌入 CBAM、E_CBAM 注意力模块后的激活热力图对比

研究提出的 E_CBAM_MobileNetV2 模型在测试集上对单品种的识别效果非常突出，所有单品种的准确率均高于 98.55%。除 TIEYAN、XJH 和 ZHENGDAN958 这 3 类玉米种子外，其余品种的准确率、准确率、召回率、F_1 – score 均达到 100%。混淆矩阵和分类结果均证明了 E_CBAM_MobileNetV2 模型具有良好的性能。

3.5 玉米种子品种识别可视化(Grad – CAM)

近年来，深度学习已经广泛运用在计算机视觉领域。然而，这种“黑盒”模型缺乏可解释性。因此，通过可解释性的工作来可视化卷积神经网络内部具有重要意义。

对此实验引入梯度加权类激活映射(Gradient – weighted Class Activation Mapping，Grad – CAM)可视化分析方法，采用梯度的全局平均来计算权重，并将各通道特征图以权重进行叠加，生成激活热力图。通过激活热力图来体现分类最依赖的特征位置，激活热力图的像素颜色强度对应特征的相关程度。当像素颜色更趋近于暖色时，它特征的激活强度越大，表明对分类结果的影响程度越高，反之，当像素颜色更趋近于冷色时，其特征的激活强度越小，表明对分类结果的影响程度越低。因此，通过激活热力图可直观反映 CNN 判断玉米种类的依据。图 7 为 MobileNetV2、CBAM _ MobileNetV2、E _ CBAM _ MobileNetV2 的热力图对比。

如图 7 所示，在识别玉米种子品种过程中，MobileNetV2、CBAM _ MobileNetV2、E _ CBAM _ MobileNetV2 关注的区域都有所不同，MobileNetV2 较为关注中上部区域、对尖端部分和边缘部分关注较少，加入 CBAM 的 MobileNetV2 过于关注了尖端部分，对于种子中间部分的关注则有所减少，而引入 E_CBAM 的 MobileNetV2 则关注了种子的大部分的区域，尤其是关注了较难区分的纹理部分，这也是区分外形相似种子的关键。通过激活热力图充分肯定了 E_CBAM_MobileNetV2 在识别玉米种子品种方面的能力。

4 结论

为实现玉米种子品种的准确、高效、快速的鉴别，将注意力机制 CBAM 引入轻量化模型 MobileNetV2，对 CBAM 的串行方式进行改进，构建新型注意力模块 E_CBAM，并通过对比，选出了效果最佳的压缩比为 4，提出了 E_CBAM_MobileNetV2 模型。选取了 9 类玉米种子作为数据集，进行了大量实验。实验结果表明，E_CBAM_MobileNetV2 模型能够快速、准确识别玉米种子类别，较 MobileNetV2 模型相比准确率提高了 5.45%，且各指标均优于 SqueezeNet、ResNet50、DenseNet、ShuffleNetV2 等模型。

参考文献

[1]周鸿达,张玉荣,王伟宇,等. 基于机器视觉技术预测玉米质量等级的方法研究[J]. 粮油食品科技,2016,24(6):50－56
ZHOU H D,ZHANG Y R, WANG W Y,et al. Identification method of maize quality grades based on machine vision[J]. Science and Technology of Cereals, Oils and Foods,2016,24(6):50－56

[2]程洪,史智兴,么炜,等. 基于支持向量机的玉米品种识别[J]. 农业机械学报,2009,40(3):180－183
CHENG H,SHI Z X,YAO W,et al. Corn breed recognition based on support vector machine[J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(3): 180－183

[3]TRAORE B B, KAMSU－FOGUEM B, TANGARA F. Deep convolution neural network for image recognition[J]. Ecological Informatics, 2018, 48: 257－268

[4]NIE P, ZHANG J, FENG X, et al. Classification of hybrid seeds using near－infrared hyperspectral imaging technology combined with deep learning[J]. Sensors and Actuators B: Chemical, 2019, 296: 126630

[5]KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]// Advances in neural information processing systems. 2012: 1097－1105

[6]SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large－scale image recognition[J]. arXiv preprint arXiv,2014:1409.1556

[7]SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1－9

[8]HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770－778

[9]HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700－4708

[10]HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv,2017:1704.04861

[11]MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: practical guidelines for efficient cnn architecture design[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 116－131

[12]HUANG S, FAN X, SUN L, et al. Research on classification method of maize seed defect based on machine vision[J]. Journal of Sensors, 2019,2019(Pt.4): 1－9

[13]XU P, TAN Q, ZHANG Y, et al. Research on maize seed classification and recognition based on machine vision and deep learning[J]. Agriculture, 2022, 12(2): 232

[14]JAVANMARDI S, ASHTIANI S H M, VERBEEK F J, et al. Computer－vision classification of corn seed varieties using deep convolutional neural network[J]. Journal of Stored Products Research, 2021, 92: 101800

[15]SHAHI T B, SITAULA C, NEUPANE A, et al. Fruit classification using attention－based MobileNetV2 for industrial applications[J]. Plos one, 2022, 17(2): e0264586

[16]叶名炀,张杰强. 基于轻量化网络 MobileNetV2 的玉米病害识别研究[J]. 现代计算机,2022,28(11):46－50
YE M Y,ZHANG J Q. Research on corn disease recognition based on MobileNetV2[J]. Modern Computer, 2022,28(11):46－50

[17]易才键,陈俊,王师玮. 基于轻量型卷积神经网络的生活垃圾图像分类[J]. 软件工程,2023,26(3):41－45
YI C J, CHEN J, WANG S W. Domestic garbage image classification based on lightweightconvolutional neural network[J]. Software Engineering, 2023,26(3):41－45

[18]ZHANG L, WANG J, LI B, et al. A MobileNetV2－SENet－based method for identifying fish school feeding behavior[J]. Aquacultural Engineering, 2022, 99: 102288

[19]SELVARAJU R R, COGSWELL M, DAS A, et al. Grad－cam: visual explanations from deep networks via gradient－based localization[C]//Proceedings of the IEEE international conference on computer vision,2017:618－626.

附件 4：计算机软件著作权：土壤健康状况监测系统

中华人民共和国国家版权局
计算机软件著作权登记证书

证书号：软著登字第13492876号

软件名称：土壤健康状况监测系统
V1.0

著作权人：青岛农业大学

权利取得方式：原始取得

权利范围：全部权利

登记号：2024SR1089003

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。

中华人民共和国国家版权局
计算机软件著作权登记专用章

2024年07月30日

附件 5：计算机软件著作权：温室智能调控数字化

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第11067304号

软件名称：温室智能调控数字化管理系统 V1.0

著作权人：青岛农业大学

开发完成日期：2022年11月07日

首次发表日期：2022年11月07日

权利取得方式：原始取得

权利范围：全部权利

登记号：2023SR0480133

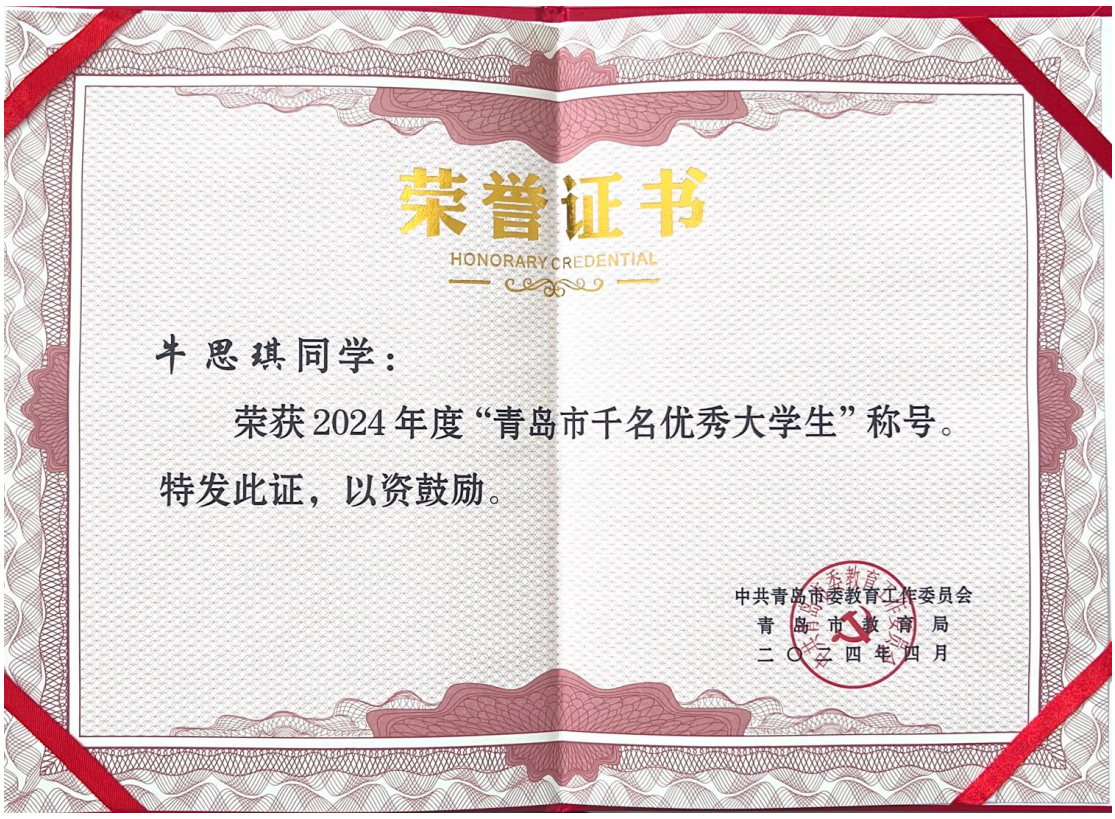
根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。

中华人民共和国国家版权局 计算机软件著作权登记专用章

No. 12549780

2023年04月18日

附件 6：青岛市千名优秀大学生



附件 7：山东省优秀毕业生

山东省普通高等学校

优秀毕业生

证书

学生 牛思琪 ，

自 2018 年至 2022 年在

青岛农业大学 学校

通信工程 专业学习期间，

德智体美劳全面发展，成绩优异，被评为

2022 届优秀毕业生。

报备编号：GXYP22YZS10435N382350805-EC4C698ADB

2022 年 04 月 21 日

登录山东省人力资源和社会保障厅（http://hrss.shandong.gov.cn）电子签章信息查验服务平台输入报备编号进行验证。