# SORGHUM GRAIN CLASSIFICATION AND DETECTION USING YOLOV8 WITH DECOUPLED FULLY CONNECTED ATTENTION MECHANISM AND MULTI-SCALE FEATURE FUSION

/

# *融合解耦全连接注意力机制与多尺度特征融合的 YOLOv8 高粱籽粒分类检测方法*

**Yinzeng LIU[1)], Chengpeng CUI[1,2)], Kaixing ZHANG[3)], Lu LIU [4)], Fandi ZENG[1)], Mingming LIU[1)], Hongwei DIAO[1)], Tiexin WANG[2)], Zhihuan ZHAO[*1)]**

[1)] Mechanical and Electronic Engineering College, Shandong Agriculture and Engineering University, Jinan / China;

[2)] Shandong Unia Agricultural Machinery Co., Ltd, Dezhou / China;

[3)] College of Mechanical and Electrical Engineering, Shandong Agricultural University, Tai'an / China;

[4)] Production Unit Directly-Affiliated to Xinjiang Production and Construction Corps / China

*Corresponding authors: Zhihuan Zhao; Tel: +86-134-0696-7329.; E-mail: zhaozhihuan@sdaeu.edu.cn*

*DOI: https://doi.org/10.35633/inmateh-79-108*

***Keywords:*** *YOLOv8, Sorghum grain classification, BiFPN, Dynamic Head*

## ABSTRACT

*The quality inspection of sorghum grains after drying is critical for ensuring storage stability and processing quality. However, traditional manual inspection methods are inefficient, time-consuming, and prone to subjective biases. To address these challenges, this study proposes an improved YOLOv8s model, named YOLOv8s-GBD, for accurate sorghum grain classification and detection. The proposed model integrates three key enhancements. First, the DFC Attention-based bottleneck structure from GhostNetV2 is incorporated into the C2f module to enhance global context awareness and improve detection accuracy. Second, a bi-directional feature pyramid network (BiFPN) is introduced to optimize multi-scale feature fusion. Third, a dynamic head framework based on attention mechanisms is adopted to strengthen feature representation, further boosting the model's accuracy and robustness. Experimental results demonstrate that the YOLOv8-GBD model achieves 95.2% precision, 98.0% recall, 98.7% mAP1, 97.5% mAP2, and 96.6% F1-score. Compared to the original YOLOv8s model, these metrics show improvements of 3.1%, 1.7%, 1.2%, 1.0%, and 2.4%, respectively. Furthermore, the YOLOv8-GBD model outperforms other YOLO series models in sorghum grain classification and detection tasks. In conclusion, the YOLOv8s-GBD model meets the requirements for high accuracy in sorghum grain quality inspection, offering a robust solution for practical applications.*

## 摘要

*干燥后高粱籽粒的品质检测对保障仓储稳定性与加工品质至关重要。然而传统人工检测方式效率低下、耗时长，且容易存在主观偏差。针对上述问题，本文提出一种改进 YOLOv8s 模型 YOLOv8s-GBD，实现高粱籽粒精准分类与检测。该模型包含三项核心改进：第一，在 C2f 模块中嵌入 GhostNetV2 的 DFC 注意力瓶颈结构，增强模型全局上下文感知能力，提升检测精度；第二，引入双向特征金字塔网络（BiFPN）优化多尺度特征融合；第三，采用基于注意力机制的动态检测头框架强化特征表征能力，进一步提升模型精度与鲁棒性。实验结果表明：YOLOv8s-GBD 模型精确率 95.2%、召回率 98.0%、mAP1 为 98.7%、mAP2 为 97.5%、F1 分数 96.6%。相较于原始 YOLOv8s 模型，各项指标分别提升 3.1%、1.7%、1.2%、1.0%、2.4%。同时，在高粱籽粒分类检测任务中，YOLOv8s-GBD 模型性能优于其他 YOLO 系列模型。综上，YOLOv8s-GBD 模型能够满足高粱籽粒品质检测的高精度需求，可为实际工程应用提供可靠方案。*

## INTRODUCTION

Sorghum is one of the important food crops in China. In the processing, transportation, storage, acquisition, sales and other links, the quality of sorghum grains needs to be tested. Its quality directly affects food security and economic benefits. However, during the acquisition and drying process, the grain will inevitably be affected by environmental factors such as temperature, which will lead to a series of quality changes, such as mildew, lesions and mechanical damage (*Olgun M et al., 2016; Tsuzuki W et al., 2014*). Sorghum grains with defects or damage but still have use value are called imperfect grains. The national standard GB/T8231-2024 defines imperfect grains as: diseased, moth-eaten, damaged, sprouted, mouldy and heat-damaged.

During the harvesting process of sorghum, some sorghum seeds are wrapped with glume-protected and useful grains called shelled grains. Sorghum grading indicators such as moisture, capacity, and thousand grain weight need to be completed on-site before sorghum enters storage after the drying session. Some indicators such as imperfect grains and shelled grains cannot be completely eliminated. Small samples need to be tested to ensure grain quality, guide grain storage management and prevent economic losses. Therefore, research on sorghum quality evaluation methods and techniques is of great practical significance to meet the market demand and ensure sorghum storage safety.

The traditional classification of imperfect grains and shelled grains is mainly detected by manual visual inspection. This method is not only inefficient, but also easily affected by subjective factors, resulting in poor accuracy and reliability of the detection results (*Zhang W. et al., 2021; Wang L. et al., 2022; Ni C et al., 2019*). At present, many scholars use physical and chemical detection methods to study the imperfect particles (*Guo M. et al., 2019; Liu J. et al., 2024; Pearson et al., 2007; Yang et al., 2021*), such as analysing the time-domain characteristics of the acoustic wave, or studying the interaction between the acoustic wave and the medium to achieve the detection of imperfect particles (*Sun et al., 2021*). This method is fast, but highly dependent on the acoustic propagation medium, and is highly susceptible to the influence of noise. Chemical detection methods can achieve rapid screening of fungal contamination and mycotoxins in grain, but the repetition rate is very poor and prone to environmental pollution (*Gab-Allah et al., 2021*).

With the rapid development of modern information technology, such as machine vision technology, near-infrared spectroscopic detection technology, hyperspectral detection technology are more and more widely used in the field of agricultural products (*Na et al., 2019; Qi et al., 2024; Qin et al., 2017*). Meanwhile, target detection technology based on deep learning shows great potential in the field of agricultural product quality inspection.

The use of hyperspectral imaging and near-infrared spectroscopy can be used to directly measure the mould condition and disease spots of grain kernels (*Delwiche et al., 2019; He et al., 2021; Sendin et al., 2017*). *Liu Zhang et al., (2021),* proposed a data calibration model based on multi-angle near-infrared hyperspectral, using four sides of wheat grain images combined with machine learning models to detect wheat damaged by rice weevil, and the accuracy rate reached 97%. *Li et al., (2022),* proposed a method for identifying unsound wheat kernels based on a deep convolutional Generative Adversarial Network (DCGAN) and near-infrared hyper-spectral imaging to identify unsound wheat seeds. Experiments show that the recognition accuracy based on the joint model of DCGAN and CNN reaches up to 96.67%. *Fengshuang Liu et al. (2023)* proposed a new spectral-spatial feature extraction-enhanced fully-connected neural network (SSFE-FCNN) to achieve the detection of pixel-level mechanical dam-ages in glutinous maize, and the experiments show that the accuracy of this method reaches up to 98.09%. It can be achieved for the grading operation of maize. Although the use of hyperspectral imaging technology, near infrared spectroscopy technology can achieve the detection of grain defects, germination, etc., but its cost is high and the effect is limited for the detection of grain shape, disease spots and other appearance characteristics.

In recent years, many researchers have used machine vision methods to detect defective grains. The commonly used methods are traditional image processing methods and artificial neural network methods (*Chen Z. et al., 2022; Shen R. et al., 2023*). The traditional image pro-cessing method detects by extracting features such as colour and shape. This method has low accuracy and poor robustness. Therefore, there are more related researches on target detection using deep learning techniques. *Zhang et al., (2025),* proposed a method for identifying unsound wheat seeds based on multi-scale parallel convolutional neural networks. Experiments showed that the accuracy of this method reached 98.798 %. *Lin W. et al.,* (*2023*), proposed an online classification method for soybean seeds based on deep learning. This method can accurately classify normal, damaged, abnormal and unclassifiable soybeans. The accuracy rate reaches 95.63% on NVIDIA Jetson TX2, which lays the foundation for online soybean quality evaluation. *Gao et al., (2022),* studied the identification of wheat unsound kernels based on machine vision and deep learning network models. By comparing the advantages and disadvantages of GoogleNet, DenseNet, IX-Resnet, Res2Net and other networks, a Res24_D_CBAM_Atrous wheat imperfect kernel detection method was proposed, with an accuracy rate of 94 %.

In summary, the combination of machine vision and deep learning enables the detection of imperfect grains at a low cost. Currently, it is widely used in crops such as wheat, corn and soybean, but there are few literatures on the detection of imperfect grains and husked grains of sorghum. Therefore, identifying imperfect, perfect, and shelled grains in sorghum is of great practical importance, but the wide variety places higher demands on the generalisation ability of the model.

To address the aforementioned challenges, an improved model YOLOv8s-GBD (YOLOv8s-C2f_Ghostblockv2-BiFPN-DymiHead) based on YOLOv8s is proposed to achieve accurate recognition of sorghum seed categories. In this study, the C2f module in the YOLOv8s backbone feature extraction network is improved by incorporating the DFC Attention-based Bottleneck structure from GhostNetV2 into the C2f module. This modification strengthens the global context-awareness of feature representation and enhances the detection accuracy of the model. Additionally, a bidirectional feature pyramid network (BiFPN) is incorporated to enhance multi-scale feature fusion performance and thereby improve the accuracy of small-target detection. Finally, to further boost the model's feature expression capability, a target detection head framework based on the attention mechanism (Dynamic Head) is introduced. This not only improves the model's accuracy and robustness but also ensures accurate classification of sorghum seed categories.

## MATERIALS AND METHODS

### *Sorghum grain dataset*

The training of the YOLOv8s-GBD algorithm proposed in this study requires labelled sorghum grain datasets. However, existing open-source datasets lack a specific dataset for sorghum grain classification. Consequently, this study constructed a dedicated sorghum grain dataset by collecting images of sorghum grains. The dataset was systematically classified and labelled, with each grain individually enclosed in an independent bounding box.

### *Materials*

Sorghum samples were provided by Bengbu Chuqian Trading Co., Ltd. Three different types of sorghum grains—perfect grains, imperfect grains, and shelled grains-from the Hongyingzi variety were used as research objects. Due to sample limitations, this study focused solely on three types of imperfect grains: lesion grains, damaged grains, and insect-damaged grains. Prior to the experiment, the seeds were dried to ensure that the moisture content of the sorghum grain samples was (13.5 ± 0.5)%. Subsequently, perfect grains, imperfect grains, and shelled grains were carefully selected from the samples.

### *Sorghum grain image acquisition*

In this study, high-resolution industrial cameras were used to collect images of sorghum grains. The industrial camera is a colour camera with a resolution of 25 MP and a 25 mm focal length lens. The industrial camera is connected to the host computer through GigE to ensure the high speed and stability of data transmission.

In the process of image acquisition, the industrial camera is fixed on the array platform, which is 25 cm away from the sorghum grain. The image acquisition device is shown in Fig. 1. In order to reduce environmental interference, the collection site is in the laboratory, and the collection time is selected during the period of stable light (10 a.m. to 2 p.m.). The selected three different types of sorghum grains were randomly spread on a black worktable. Through the SDK software of the industrial camera, the parameters are reasonably set to obtain a colour image with a resolution of 5120 × 5120, which is stored in the jpg format, and a total of 167 colour images are collected.

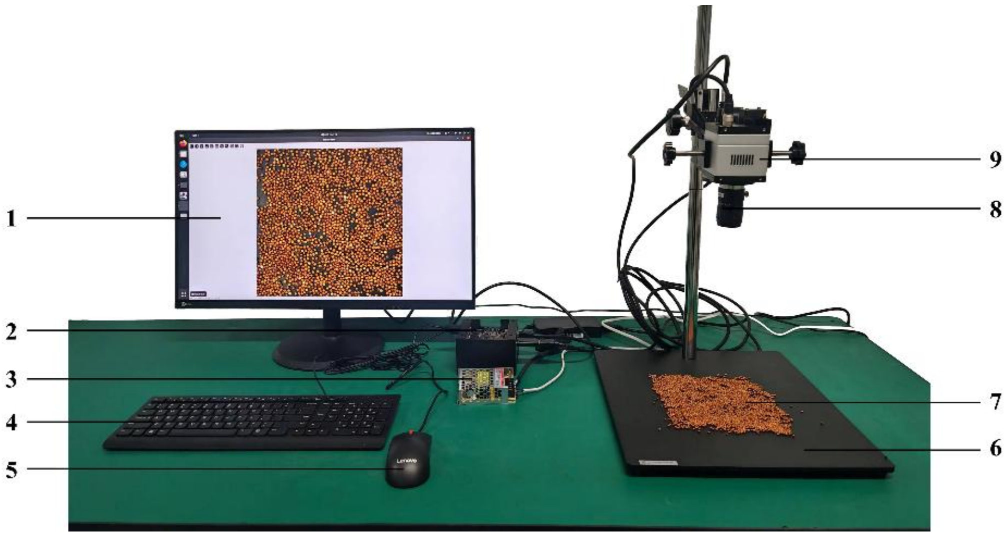



**Fig. 1 - Image acquisition platform**

*1 - Monitor; 2 - Host computer; 3 - Switch power supply; 4 - Keyboard; 5 - Mouse; 6 - Worktable; 7 - Sorghum seed; 8 - Lens; 9 - Industrial camera*

### *Dataset construction*

The images collected by industrial cameras contain multiple sorghum grains, making them difficult to label. Therefore, this study employs threshold segmentation, morphological processing and contour search algorithms to divide the images into several sub-images, as shown in Fig. 2. The image dataset consists of sub-images containing only one complete grain. Sorghum grains are classified based on their appearance, epidermal characteristics, and integrity, as illustrated in Fig. 3. The open-source LabelImg image annotation tool (https://github.com/tzutalin/labelImg) is utilized to annotate perfect grains, imperfect grains, and shelled grains within the images. The three types of sorghum grain images are described as follows:

(1) Perfect grain: complete and shiny sorghum grain.

(2) Imperfect grain: lesion grains, damaged grains and insect-damaged grains.

(3) Shelled grain: Sorghum grain with guard glumes (shells).

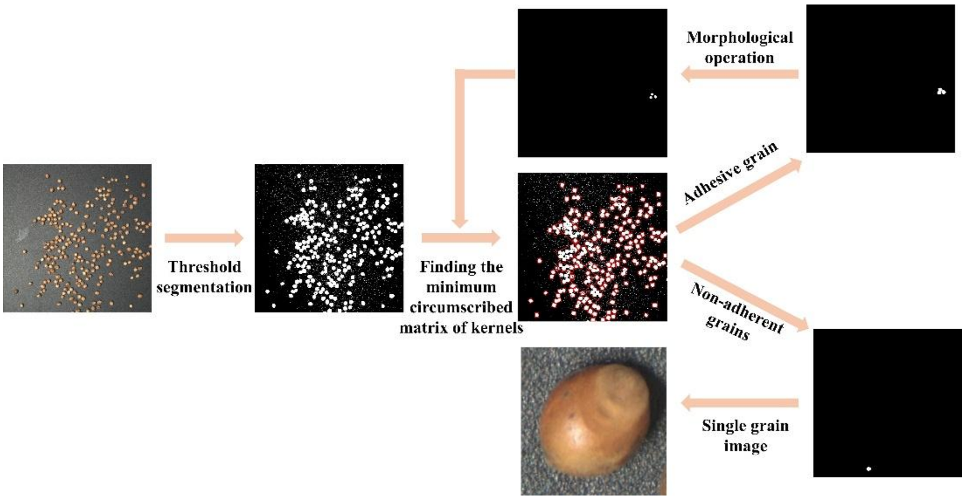



**Fig. 2 - Image pre-processing**

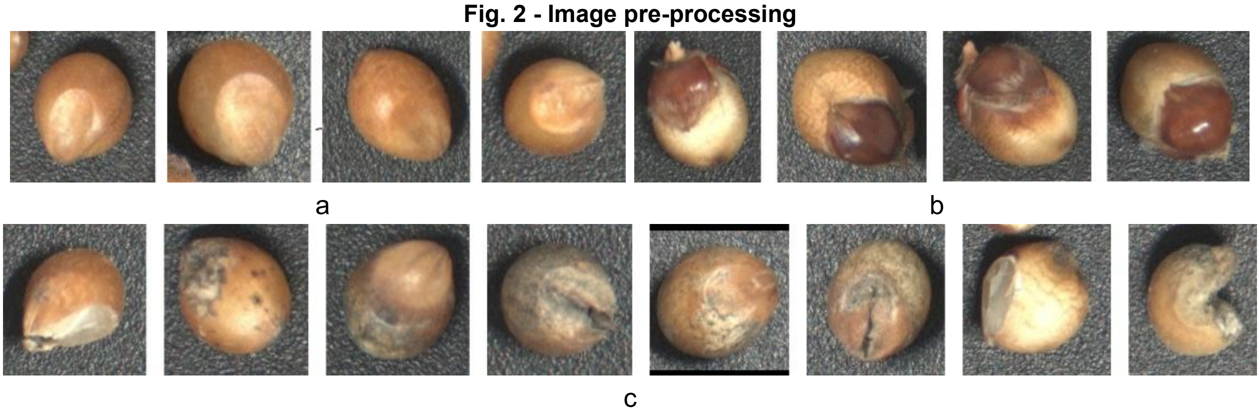



**Fig. 3 - Sample of sorghum grain**

*a - Perfect grain; b - Shelled grain; c - Imperfect grains.*

To improve the model's generalization ability and enhance detection performance. The dataset was augmented using preprocessing techniques such as random image rotation and brightness-contrast. The final dataset contained 6330 perfect grains, 7590 imperfect grains, and 5980 shelled grains. To train and validate the model, the three types of sorghum grain images were divided into training, validation and test sets in the ratio of 7:2:1, respectively.

### *Deep Learning Network Construction*

### *YOLOv8*

YOLOv8 is a state-of-the-art single-stage object detection model developed by Ultralytics. The model consists of four parts: Input, Backbone, Neck and Output (*Chitraningrum et al., 2024; Jiang, P et al., 2022*). In the backbone network and Neck section, the design concept of YOLOv7 is referenced. The backbone network employs the C2f module to extract multi-scale features. In the Head section, a decoupled head structure is adopted to separate the classification and detection tasks. Meanwhile, YOLOv8 adopts Task-Aligned Assigner positive-negative sample matching and introduces Distribution Focal Loss (DFL) to improve the training stability and generalisation performance. In addition, the operation of disabling Mosaic augmentation during the last 10 epochs of YOLOX training is introduced in the data augmentation phase, which effectively enhances the model's accuracy. These changes make YOLOv8 perform well in the classification recognition task. By comparing the model training results of different scales of YOLOv8, as shown in Table 1. Although YOLOv8n has the fewest model parameters, its small number of residual structures results in relatively low detection accuracy. However, YOLOv8m, YOLOv8l, and YOLOv8x have a larger number of parameters, making it difficult to meet the requirements for lightweight and real-time performance. Therefore, this study selected YOLOv8s as the basis for designing a new sorghum grain classification and detection model algorithm, as shown in Fig. 4.

**Table 1**

**Comparison of the performance of different versions of YOLOv8**

| Model | Size | mAP50 | mAP 50-95 | Parameters/×106 M | Weight size/MB |
|---|---|---|---|---|---|
| Yolov8n | 640 | 0.694 | 0.653 | 3.2 | 5.9 |
| Yolov8s | 640 | 0.707 | 0.759 | 11.7 | 21.4 |
| Yolov8m | 640 | 0.712 | 0.734 | 26.0 | 49.5 |
| Yolov8l | 640 | 0.708 | 0.721 | 43.7 | 83.5 |
| Yolov8x | 640 | 0.703 | 0.719 | 68.2 | 130 |

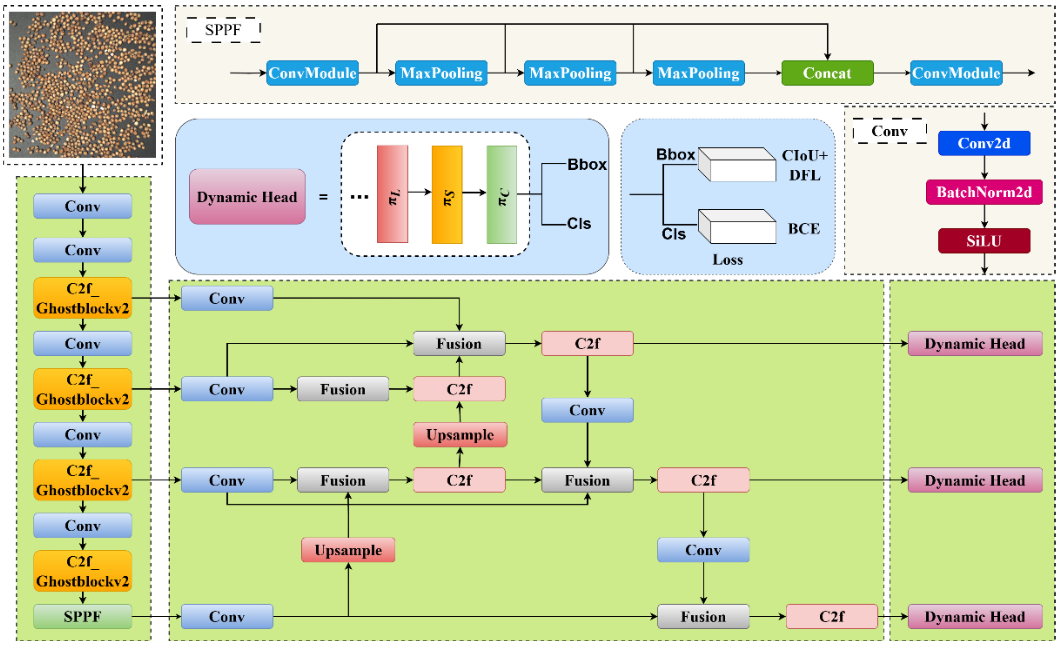



**Fig. 4 - Improved YOLOv8 model**

***C2f_Ghostblockv2***

The backbone feature extraction network of the model has the greatest impact on the model accuracy. GhostNetV2 proposed by *Tang et al., (2022),* is a lightweight model, which reduces the number of parameters while ensuring a certain accuracy. In this study, the Bottleneck structure of GhostNetV2 is introduced into the C2f network structure to replace the original convolution layer to form a new network structure C2f_Ghostblockv2, as shown in Fig. 5. This structure solves the problem of recognition accuracy of the model.

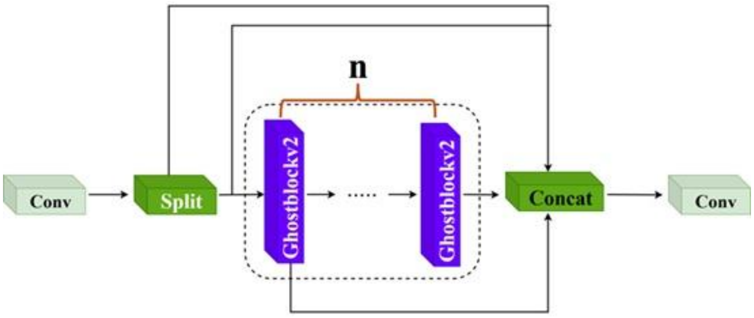



**Fig. 5 - Structure of C2f _ Ghostblockv2**

The Bottleneck structure of GhostNetV2 consists of two Inverted Residual Bottle-neck structures based on the Ghost Module, as shown in Fig. 6. Specifically, the first Ghost module expands the number of output channels via an extension layer to generate extended features, while the second Ghost module reduces the number of channels to align with the shortcut path's output. To enhance the model's representational capacity, GhostNetV2 integrates the DFC (Decoupled Fully Connected) attention mechanism within the first Ghost module. The DFC attention operates in parallel with the first Ghost module to produce an attention map, which is element-wise multiplied with the extended features to strengthen their representational power. These enhanced features are then passed through the second Ghost module to generate the final output features. The DFC attention captures long-range spatial dependencies by decomposing a fully connected layer into two separate fully connected layers along the horizontal and vertical directions, aggregating pixel information in both dimensions independently. When identifying different types of sorghum grains, this mechanism allows the model to maximize its performance, thereby improving the recognition accuracy of the improved model.

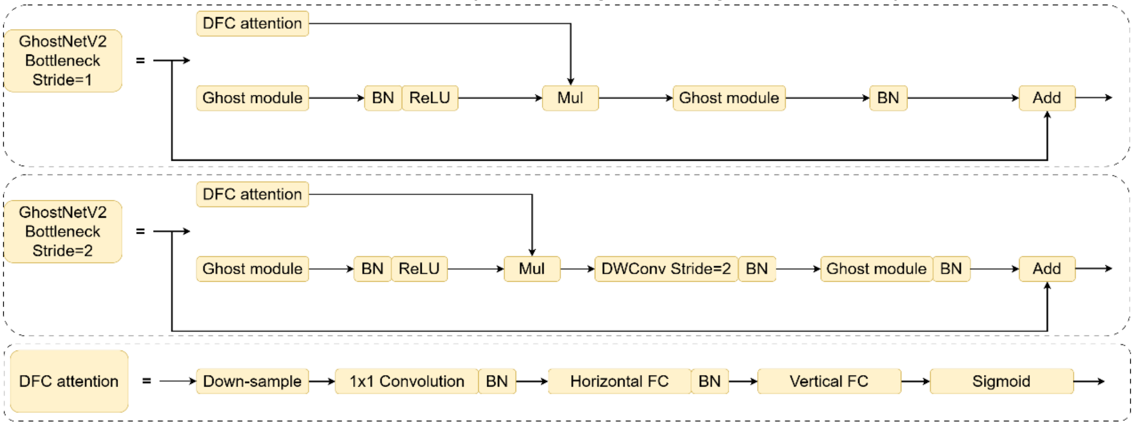



**Fig. 6 - Bottleneck structure of GhostNetV2**

### *Feature Fusion Network (BiFPN)*

In the original YOLOv8s network, the multi-scale feature maps extracted by the backbone network are input into the PANet structure for feature fusion, the structure of which is shown in Fig. 7(a). The PANet structure uses the summation of different features to achieve the fusion of different feature layers. However, there are many types of sorghum grains, and the sizes of damaged grains and shelled grains in perfect grains and imperfect grains are different. If only simple feature-layer summation is performed, it will result in uneven contributions of different input features to the fused output features. To avoid this problem and make the network better capture multi-scale feature information, this study introduces BiFPN (*Li N. et al., 2024*) to improve the YOLOv8s network, as illustrated in Fig. 7(b).

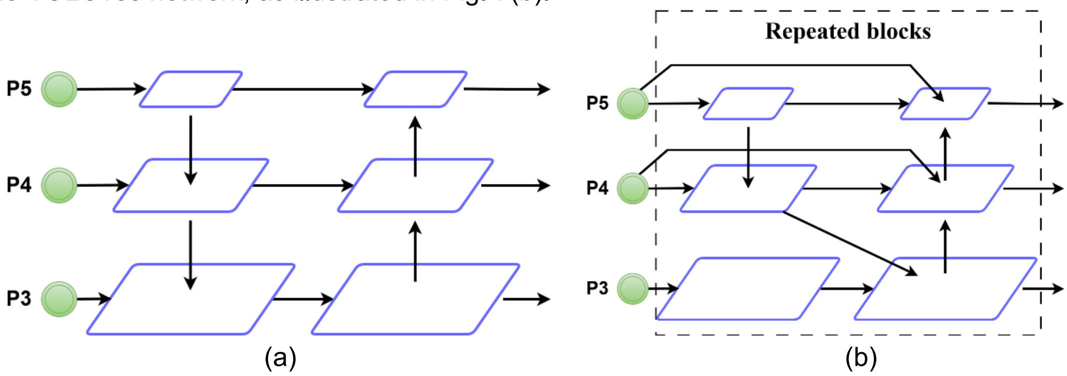



**Fig. 7 - Structure of PANet, BiFPN**

BiFPN introduces a bi-directional fusion mechanism, where features are fused through bi-directional paths (top-down and bottom-up). In the top-down path, high-level features are fused with low-level features by upsampling. In the bottom-up path, low-level features are fused with high-level features through downsampling. This bi-directional flow enhances the expressive power of the features and allows the different scale features to complement each other. The structure automatically adjusts the importance of different scale features through a weighted fusion mechanism to achieve more accurate feature fusion and improve target detection performance. Meanwhile, the BiFPN module can be repeatedly stacked to enhance information fusion.

By replacing the Neck structure in the YOLOv8s network with the BiFPN network, the model integrates the local and global features of different types of sorghum grains. This approach retains shallower semantic information while preserving deeper location information. Consequently, accurate detection of different types of sorghum seeds at different scales is achieved.

***DymiHead***

The Head part of the YOLOv8 network architecture receives multi-scale feature maps from the Backbone and Neck networks. These feature maps have different resolutions and are used to detect targets of different sizes. However, the difference of grain size in sorghum grain images is small, only some damaged grains are significantly different from other grain sizes, and other types of grain sizes are relatively close. Therefore, this study introduces a dynamic head framework based on attention mechanism to enhance the feature expression ability of the model.

Dynamic Head consists of Scale-aware Attention, Spatial-aware Attention and Task-aware Attention, and its structure is shown in Fig. 8. The Scale-aware Attention module ($\pi_L$) consists of Mean Pooling, 1×1 Convolution, ReLU Activation function and Hard Sigmoid function. This module adaptively enhances important features and suppresses redundant information by computing the spatial and channel attention of the feature map, enabling the model to better capture multi-scale targets. The spatial-aware attention module ($\pi_S$) consists of offset learning and 3× 3 convolution. This module focuses on the spatial location information of the feature map and uses the global context information to enhance the response of the target region. The task-aware attention module ($\pi_C$) is implemented by global average pooling, two fully connected layers, a normalisation layer, and a shifted sigmoid function. This module exploits task-specific attention mechanisms to enhance the performance of the model in multi-task learning.

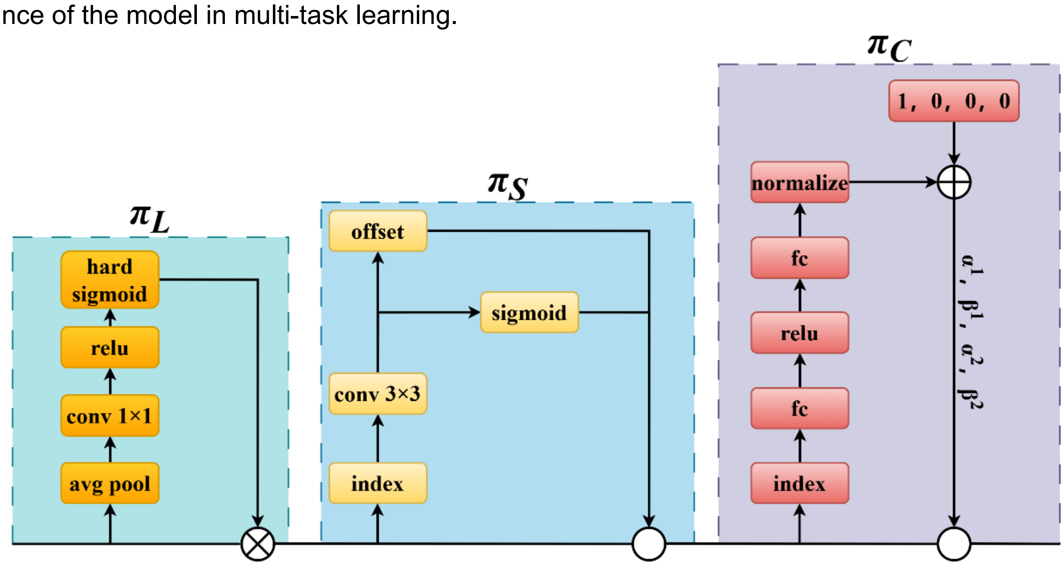



**Fig. 8 - Dynamic Head structure**

The multi-scale feature maps of sorghum grains extracted by the backbone net-work and Neck network were input into Dynamic Head. Different attention mechanisms operate in different dimensions, complement each other, and collectively enhance the representation ability of the target detection head. By dynamically adjusting the feature weights, the accuracy and robustness of sorghum seed target detection are significantly improved.

***Loss calculation***

In sorghum kernel classification and detection, the trained model predicts the class of each detected kernel. The prediction results can include both correct and incorrect classifications; therefore, it is essential to evaluate whether the predicted kernel class matches the ground-truth label. The final classification result is determined based on this comparison, enabling the identification of different sorghum kernel types.

This study uses the loss function in the original YOLOv8s network structure, which consists of two parts: category classification loss and border regression loss. Classification Loss uses cross entropy loss. This method measures the performance of the model by calculating the cross entropy of the probability distribution predicted by the model and the probability distribution of the real label. The YOLOv8s target detection model uses CIOU as the regression loss function, which is defined as:

$$CIoU = IoU - \left(\frac{\rho^2(b,b^{gt})}{d^2} + \alpha v\right) \tag{1}$$

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

$$\alpha = \frac{v}{(1-IoU)+v} \tag{3}$$

where $\rho$ is the Euclidean distance, d is the diagonal of the minimum bounding box, $v$ is the trade-off parameter, and α is the corresponding trade-off parameter.

The boundary loss is then in the form of CIoU (Complete Iou) Loss+DFL. Both the distance and overlap rate between the detection frame and the target frame are considered, as well as the influence of the length ratio of the detection frame on regression.

## RESULTS

### *Experimental environment and parameters*

The experiments were conducted on a Windows 10 system equipped with an Intel Core i5-12600KF CPU operating at 3.70 GHz, an NVIDIA GeForce RTX 4060 Ti GPU with 16 GB of video memory, and 32 GB of system RAM. The PyTorch deep learning framework (version 1.12.1) was used, with PyCharm 2020.2.1 as the development environment. GPU acceleration was configured using CUDA 11.6 and cuDNN 8.9.7. Python 3.8.5 was used together with torchvision 0.13.1 and torchaudio 0.12.1.

**Table 2**

**Hyperparameters used for deep learning training**

| Parameter | Parameter Value |
|---|---|
| Initial learning rate | 0.01 |
| Final learning rate | 0.01 |
| Weight decay factor | 0.0005 |
| IoU threshold | 0.7 |
| Momentum | 0.9 |
| Learning Rate Adjustment Strategy | steps |
| Image size | 160pixel × 160pixel |
| Batch-size | 32 |
| Number of epochs | 200 |

Before training, the hyperparameters were optimized, with particular attention to key settings such as the input image size and batch size. Examination of the loss curve showed that, when the number of training epochs was set to 200, the model loss gradually stabilized, indicating that the model had reached convergence. Therefore, the number of training epochs was set to 200 in this study. The detailed experimental hyperparameters are presented in Table 2.

### *Evaluation indicators*

In machine learning and deep learning, the prediction results of classification tasks are classified into the following four categories:

(1) *True Positive (TP)*: the prediction is positive, the label value is also positive, and the prediction is correct.

(2) *False Negative (FN)*: Predicted as a negative case, labelled as a positive case, wrong prediction

(3) *False Positive (FP):* the prediction is positive, the label value is negative, the prediction is wrong

(4) *True Negative (TN):* predicted as a negative case, the label value is negative, the prediction is correct

In this study, *P (Precision, %), R (Recall, %), mAP1 (mAP50), mAP2 (mAP50-95)* and *F1* were used to evaluate the performance of the network model. The corresponding equations for these five metrics are as follows:

$$P = \frac{TP}{TP+FP} \times 100\% \tag{4}$$

$$R = \frac{TP}{TP+FN} \times 100\% \tag{5}$$

$$F1 = \frac{P*R*2}{P+R} \tag{6}$$

$$AP = \int_0^1 P(R)dR \tag{7}$$

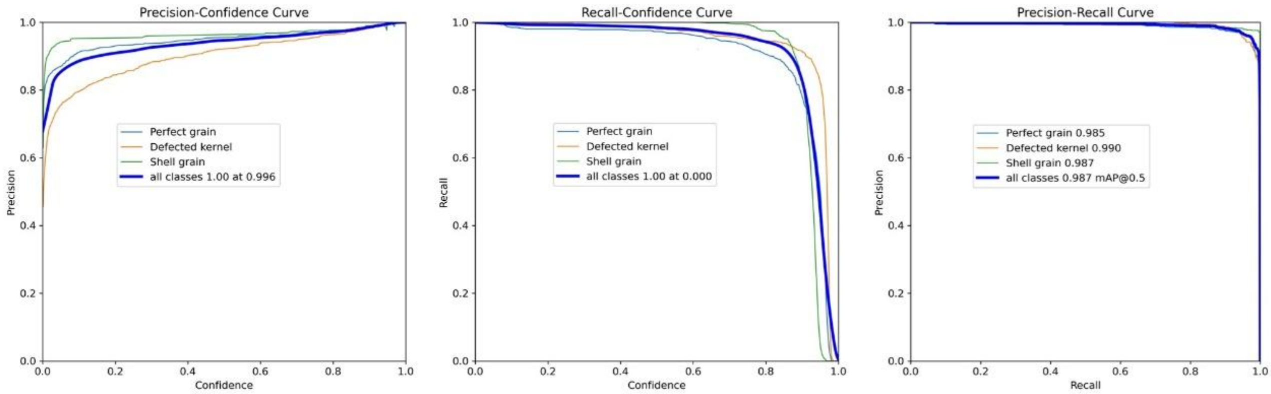
$$mAP = \frac{1}{M}\sum_{k=1}^{M} AP(k) \times 100\% \tag{8}$$

As indicated by the above equations, a higher $P$ corresponds to fewer false positives ($FP$), meaning that fewer samples from other classes are incorrectly predicted as the target class. A higher *Recall* corresponds to fewer false negatives ($FN$), indicating fewer missed detections of positive samples. Because a model may achieve high $P$ but relatively low $Recall$, or vice versa, the $F1$-score was used to provide a balanced evaluation of $P$ and $Recall$. As the harmonic mean of $P$ and $Recall$, the $F1$-score is strongly influenced by the lower of the two values and therefore helps prevent model selection based on only one favourable metric. The mean average precision ($mAP$) provides an overall measure of object detection performance across classes and IoU thresholds, depending on the specific $mAP$ definition used.

***Improved detection performance of the YOLOv8 algorithm***

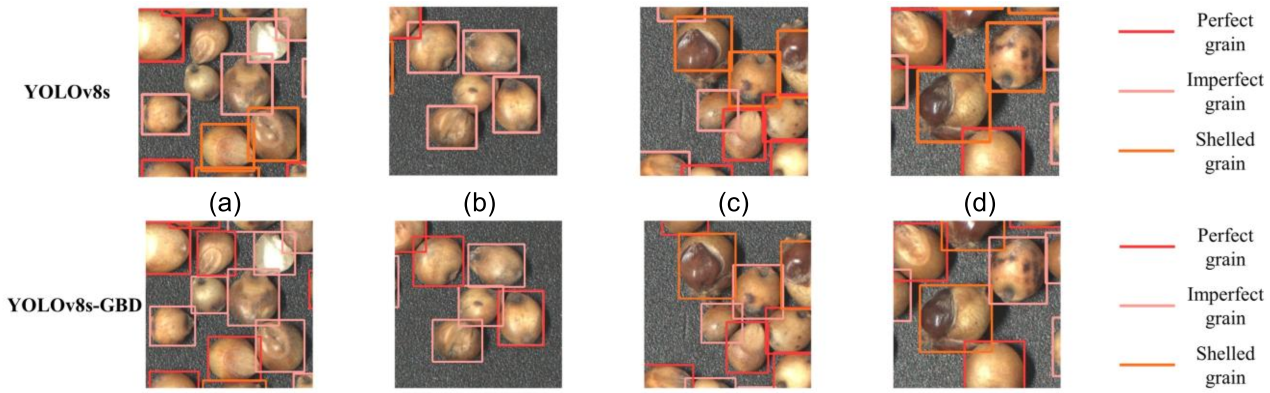
The YOLOv8s-GBD sorghum kernel classification and detection model proposed in this study was trained using transfer learning. The precision $P$, $R$, and $PR$ curves obtained during training are shown in Fig.9.

As shown in Fig. 9, the precision values for the Perfect grain, Defected kernel, and Shell grain categories were 0.985, 0.990, and 0.987, respectively. Furthermore, the mean $Average\ Precision\ (mAP1)$ achieves a value of 0.987. These results indicate that the proposed model achieved high classification and detection performance for different sorghum kernel categories.




(a) (b) (c)

**Fig. 9 - P, R, PR curves**




(e) (f) (g) (h)

**Fig. 10 - Image details of different models**

*a - Leak detection 2, False detection; b - Leak detection, False detection 1; c - False detection 1, Inaccurate bounding box; d - False detection 1, Inaccurate bounding box 1; e - No leak detection, No false detection; f - No leak detection, No false detection; g - No false detection, Bounding box accuracy; h - No false detection, Bounding box accuracy*

The model outputs in this study identify different types of sorghum kernels using rectangular bounding boxes overlaid on the images. For clearer visualization, representative regions of the sorghum kernel detection images were cropped and enlarged. YOLOv8s was used as the baseline model for comparison, as shown in Fig. 10. The comparison results show that dense kernel distributions can lead to severe overlap and adhesion between adjacent kernels. Under these conditions, although YOLOv8s can detect some targets, it is more prone to false detections, missed detections, and inaccurate bounding-box localization. In contrast, the YOLOv8-GBD model detects densely distributed kernels more reliably and provides more accurate classification of kernel types.

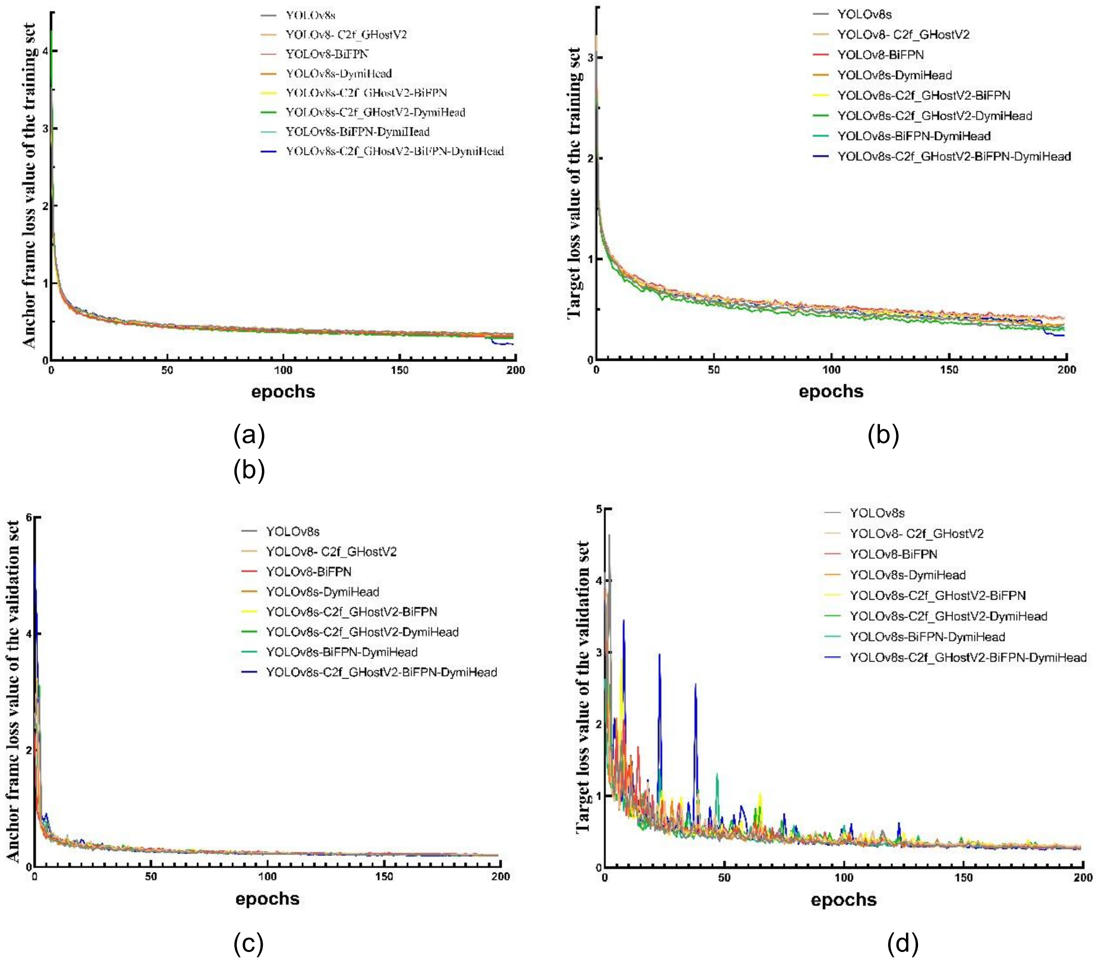
***Ablation study***

Using the same training, validation, and test sets, ablation experiments were conducted to evaluate the contribution and effectiveness of the different optimization strategies. Based on the YOLOv8s architecture, the Backbone, Neck, and Head components were modified separately. In the Backbone, the GhostNetV2 bottleneck was introduced into the C2f structure to incorporate a long-range attention mechanism and improve feature extraction accuracy. In the Neck, a BiFPN structure was employed for multi-scale feature fusion. In the Head, the original detection head was replaced with a Dynamic Head to enhance feature representation capability. Each modified model was trained for 200 epochs, and the best-performing model was selected for evaluation. The performance metrics of the different models are presented in Table 3.

**Table 3**

**Performance metrics of different models**

| Model | P/% | R/% | mAP1/ % | mAP2/ % | F1/% |
|---|---|---|---|---|---|
| YOLOv8s | 92.1 | 96.3 | 97.5 | 96.5 | 94.2 |
| YOLOv8- C2f_Ghostblockv2 | 94.4 | 95.7 | 98.1 | 96.6 | 95.0 |
| YOLOv8-BiFPN | 93.3 | 96.9 | 98.2 | 96.7 | 95.4 |
| YOLOv8-DymiHead | 93.7 | 95.9 | 98.5 | 97.2 | 94.8 |
| YOLOv8- C2f_Ghostblockv2-BiFPN | 93.9 | 96.9 | 98.1 | 96.8 | 95.4 |
| YOLOv8-C2f_Ghostblockv2-DymiHead | 94.3 | 96.3 | 98.4 | 97.4 | 94.8 |
| YOLOv8-BiFPN-DymiHead | 93.9 | 96.5 | 98.4 | 97.4 | 95.2 |
| YOLOv8s-C2f_Ghostblockv2-BiFPN-DymiHead | 95.2 | 98.0 | 98.7 | 97.5 | 96.6 |




(a) (b)
(b)
(c) (d)

**Fig. 11 - Training and validation loss curves of different models**

As shown in Table 3, introducing C2f_GhostBlockV2, BiFPN, and DymiHead individually resulted in varying improvements in model precision, recall, and mean average precision. The baseline YOLOv8s model achieved a precision of 92.1% for sorghum kernel classification and detection. After introducing C2f_GhostBlockV2, BiFPN, and DynamicHead separately, the precision increased to 94.4%, 93.3%, and 93.7%, respectively, corresponding to improvements of 2.3, 1.2, and 1.6 percentage points relative to the baseline model. Among the individual modifications, C2f_GhostBlockV2 produced the largest improvement in precision. However, only BiFPN improved recall compared with the baseline model. These results indicate that optimizing a single network component provides only limited overall performance gains.

When two optimization strategies were combined, the detection model generally achieved further improvements in precision, recall, and mean average precision, with no obvious deterioration in overall performance. Among the tested combinations, BiFPN combined with DymiHead provided the best overall performance. Compared with the baseline YOLOv8s model, this combination improved precision, recall, mAP1, and mAP2 by 1.8, 0.2, 0.9, and 0.9 percentage points, respectively.

The above experiments show that optimizing the Backbone, Neck, and Head components of YOLOv8s can improve model performance for sorghum kernel classification and detection. Accordingly, the three optimization strategies described above were integrated to develop the proposed YOLOv8s-GBD sorghum kernel classification and detection model. Compared with the baseline YOLOv8s model, YOLOv8s-GBD improved P, R, mAP1, mAP2 and F1 by 3.1, 1.7, 1.2, 1.0, and 2.4 percentage points, respectively. Fig. 11 shows the loss curves for the different model configurations. As shown in the figure, the YOLOv8s-GBD model exhibited relatively high initial training and validation loss values. The losses decreased rapidly during the first 50 epochs and gradually stabilized after approximately 100 epochs, indicating that the model approached convergence.

***Comparison experiment of different feature extraction networks***

To evaluate the performance of the improved feature extraction network, this study replaces the original C2f module of yolov8s with mainstream network structures such as MobileViTBv2, MobileViTBv3, BiFormer, and GhostNet. The training set, validation set and test set used in this study are used for training, and the training effect is compared with that of C2f_Ghostblockv2 as shown in Fig. 12.

In this study, there are many categories of sorghum grains, especially imperfect grains, with large differences in appearance size, and high requirements for the accuracy and F1 value of the model. From Fig. 12, it can be seen that the C2f _ Ghost-blockv2 structure has the largest gain in the accuracy and F1 value of sorghum grain classification and recognition. Compared with other structures, C2f _ Ghostblockv2 has a small difference in accuracy and recall rate, which can effectively avoid the short board effect and is more suitable for the needs of this study.

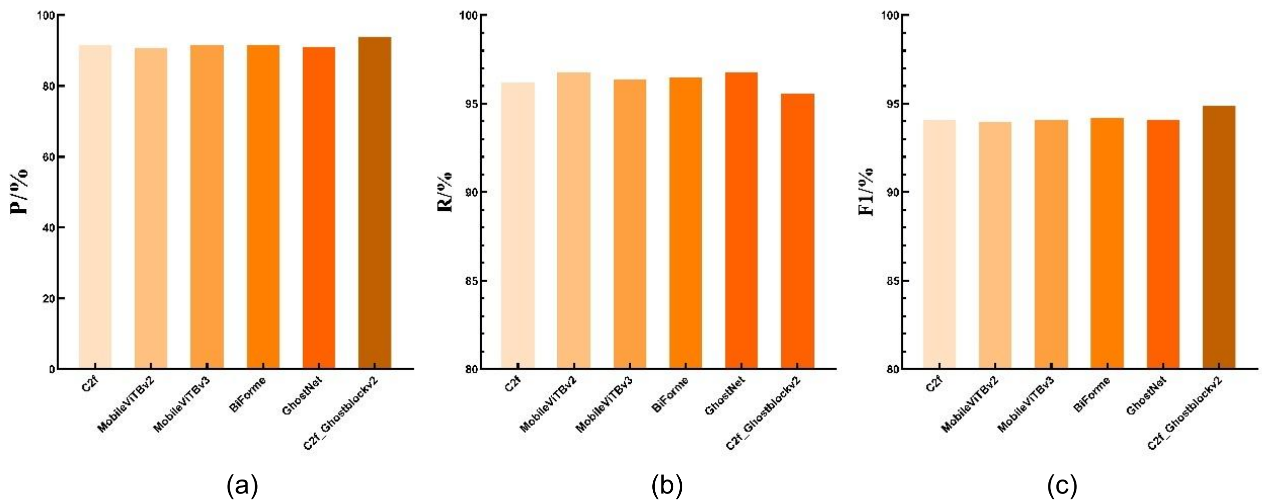



**Fig. 12 - Accuracy, recall and F1 value for different feature extraction networks**

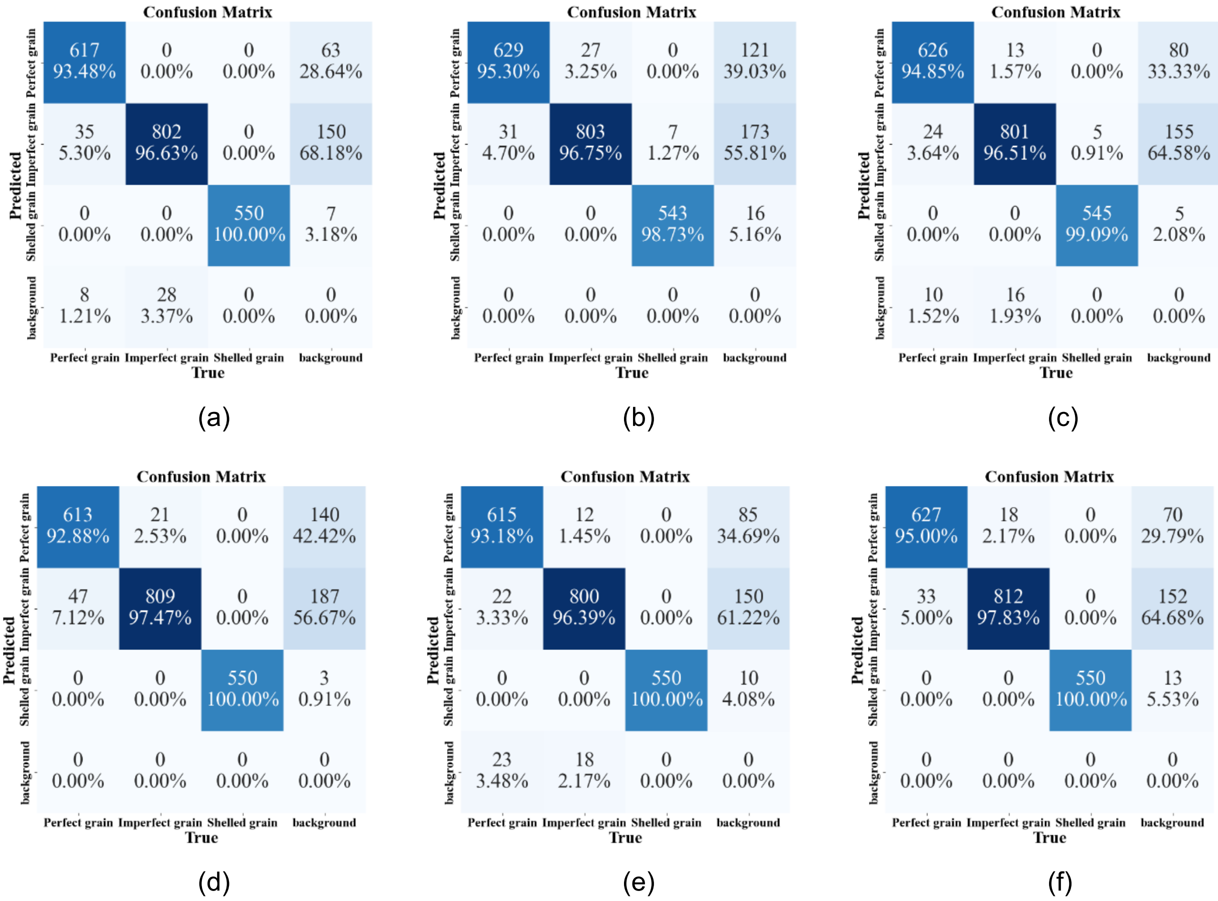



**Fig. 13 - Confusion matrix of test effects of different feature extraction networks**

As illustrated in Fig. 13, C2f_Ghostblockv2 has the highest classification accuracy for imperfect grains among the six models. Although the classification accuracy of C2f_Ghostblockv2 perfect grains is slightly lower than that of MobileViTBv2, it is still higher than the other network structures. Among the six models, the classification accuracy of C2f_Ghostblockv2 perfect grains is not the best, but its overall classification accuracy is the highest among all models. It shows that C2f_Ghostblockv2 has superior performance in sorghum grain classification.

***Comparison of different models***

To further evaluate the effectiveness of the YOLOv8s-GBD algorithm for sorghum grain classification and recognition, model training was performed on different versions of the YOLO target detection algorithm using the same training, validation and test sets. The participating comparison detection networks are YOLOv5, YOLOv7-tiny, YOLOv9, YOLOv10, and YOLOv11, and the test results are shown in Table 4.

As shown in Table 4, the YOLOv8s-GBD algorithm achieves 95.2%, 98.0%, 98.7% and 97.5% in P, R, mAP1 and mAP2 values, respectively. It is superior to other YOLO models in terms of P and mAP1, and only the R value and mAP2 value are slightly lower than YOLOv11, and much higher than other YOLO models. Therefore, the com-prehensive recognition performance of YOLOv8s-GBD algorithm is better than other YOLO models. The algorithm proposed in this study can realize the recognition of sorghum grain classification.

**Table 4**

**Parameters of performance indicators for different models**

| Model | P/% | R/% | mAP1/ % | mAP2/ % | F1/% |
|---|---|---|---|---|---|
| YOLOv5 | 90.1 | 97.9 | 96.4 | 93.7 | 93.8 |
| YOLOv7-tiny | 87.2 | 92.9 | 92.7 | 90.4 | 90.0 |
| YOLOv9 | 91.2 | 96.0 | 98.6 | 96.1 | 93.5 |
| YOLOv10 | 92.7 | 96.1 | 97.7 | 95.1 | 94.4 |
| YOLOv11 | 93.1 | 98.7 | 98.5 | 97.9 | 95.8 |
| YOLOv8s-C2f_Ghostblockv2-BiFPN-DymiHead | 95.2 | 98.0 | 98.7 | 97.5 | 96.6 |

***Generalization experiment***

In order to verify the generalization of the algorithm proposed in this study, YOLOv8s-GBD and YOLOv8s were compared on other data sets. The soybean seed dataset purchased from CSDN platform was used in this experiment. The dataset contains five categories of soybean seed images such as intact, speckled, immature, broken and lost, which are very similar to the sorghum seed classification in this study. In this study, 5513 soybean seed images were screened and divided into training, validation, and test sets according to the ratio of 7:2:1.

**Table 5**

**Soybean data set comparison test**

| Model | P/% | R/% | mAP1/% | mAP2/% | F1/% |
|---|---|---|---|---|---|
| YOLOv8s | 84.2 | 86.0 | 92.3 | 79.8 | 93.8 |
| YOLOv8s-GBD | 87.2 | 92.9 | 92.7 | 90.4 | 90.0 |

The test environment and parameter configurations were kept consistent with the sorghum seed classification test, and the test results are shown in Table 5. Compared with YOLOv8s, YOLOv8s-GBD improves the accuracy and recall by 3 and 6.9 per-centage points, which effectively reduces the problem of misdetection and omission of soybean seeds; and improves the mAP1 and mA2 by 0.4 and 10.6 percentage points, which indicates that YOLOv8s-GBD has a higher detection precision. The detection effect of the two models on the soybean seed data set is shown in Fig. 14. It can be seen from Fig. 14 that the comprehensive detection accuracy of YOLOv8s-GBD on soybean seed test set is high, which proves the generalization of this research method.

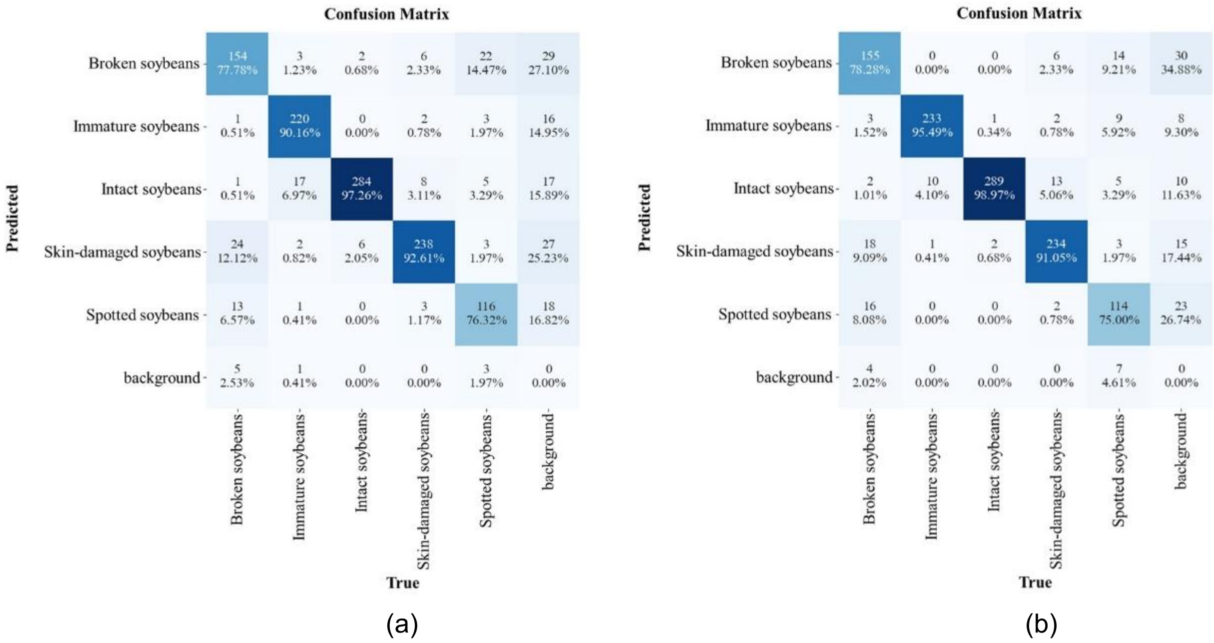



**Fig. 14 - Confusion matrix of YOLOv8s and YOLOv8s-GBD test results**

## CONCLUSIONS

(1) To address the low efficiency and time-consuming nature of conventional quality inspection methods for dried sorghum kernels, an improved YOLOv8s-GBD model was developed based on YOLOv8s. In the Backbone, the C2f module was enhanced by incorporating a GhostNetV2 bottleneck with DFC attention to strengthen feature extraction. BiFPN was introduced in the Neck to improve multi-scale feature fusion, while a Dynamic Head attention mechanism was adopted to enhance feature representation and improve model accuracy and robustness.

(2) The different optimization strategies were trained using the same training and validation sets and evaluated on a common test set. The proposed YOLOv8s-GBD model achieved P, R, mAP1, mAP2 and F1-score values of 95.2%, 98.0%, 98.7%, 97.5%, and 96.6%, respectively. Compared with the baseline YOLOv8s model, these metrics increased by 3.1, 1.7, 1.2, 1.0, and 2.4 percentage points, respectively. The results indicate that YOLOv8s-GBD provides improved performance for sorghum kernel classification and detection and satisfies the accuracy requirements of dried sorghum quality inspection.

(3) Future research should expand both the number of samples and the range of defective-kernel categories. The present study considered only lesion-damaged, mechanically damaged, and insect-damaged kernels among the defective classes. Future work should include additional types of defective kernels to improve the comprehensiveness and accuracy of sorghum kernel classification and detection.

## ACKNOWLEDGEMENT

This work was funded by the Key R&D Program of Shandong Province, China (2024CXGC010901) and the integrated pilot project of agricultural machinery Research and Development, Manufacturing and Application in Shandong Province (NJYTHSD−202311). We also sincerely appreciate the editors and anonymous reviewers for their valuable comments and constructive suggestions to improve the quality of this paper.

## REFERENCES

[1] Chitraningrum, N., Banowati, L., Herdiana, D., Mulyati, B., Sakti, I., Fudholi, A., ... & Andria, A. (2024). Comparison study of corn leaf disease detection based on deep learning YOLO-v5 and YOLO-v8. *Journal of Engineering and Technological Sciences*, 56(1), 61-70. https://doi.org/10.5614/j.eng.technol.sci.2024.56.1.5

[2] Chen, Z., Fan, W., Luo, Z., & Guo, B. (2022). Soybean seed counting and broken seed recognition based on image sequence of falling seeds. *Computers and Electronics in Agriculture*, 196, 106870. https://doi.org/10.1016/j.compag.2022.106870

[3] Delwiche, S. R., Rodriguez, I. T., Rausch, S. R., & Graybosch, R. A. (2019). Estimating percentages of fusarium-damaged kernels in hard wheat by near-infrared hyperspectral imaging. *Journal of Cereal Science*, 87, 18-24. https://doi.org/10.1016/j.jcs.2019.02.008

[4] Gao, H., Zhen, T., & Li, Z. (2022). Detection of wheat unsound kernels based on improved ResNet. *IEEE access*, 10, 20092-20101. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9698037

[5] Guo, M., Ma, Y., Yang, X., & Mankin, R. W. (2019). Detection of damaged wheat kernels using an impact acoustic signal processing technique based on gaussian modelling and an improved extreme learning machine algorithm. *Biosystems Engineering*, 184, 37-44. https://doi.org/10.1016/j.biosystemseng.2019.04.022

[6] Gab-Allah, M.A., Choi, K., & Kim, B. (2021). Accurate determination of type b trichothecenes and conjugated deoxynivalenol in grains by isotope dilution–liquid chromatography tandem mass spectrometry. Food Control, 121, 107557-. https://doi.org/10.1016/j.foodcont.2020.107557

[7] He, X., Yan, C., Jiang, X., Shen, F., & Fang, Y. (2021). Classification of aflatoxin $B_1$ naturally contaminated peanut using visible and near-infrared hyperspectral imaging by integrating spectral and texture features. *Infrared Physics & Technology*, 103652. https://doi.org/10.1016/j.infrared.2021.103652

[8] Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A Review of Yolo algorithm developments. *Procedia computer science*, 199, 1066-1073. https://doi.org/10.1016/j.procs.2022.01.135

[9] Li, H., Zhang, L., Sun, H., Rao, Z., & Ji, H. (2022). Discrimination of unsound wheat kernels based on deep convolutional generative adversarial network and near-infrared hyperspectral imaging technology. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 268, 120722-. https://doi.org/10.1016/j.saa.2021.120722

[10] Liu, J., Yang, M., He, M., Xu, M., Chen, Y., & Zhang, F. (2024). Rapid and selective detection of AFB1 by direct mass spectrometry using immunoaffinity paper substrate. *Microchemical Journal*, 196, 109439. https://doi.org/10.1016/j.microc.2023.109439

[11] Liu, F., Fu, J., & Zhao, R. (2023). Pixel-wise mechanical damage detection of waxy maize using spectral-spatial feature extraction and hyperspectral image. *Computers and Electronics in Agriculture*, 209. https://doi.org/10.1016/j.compag.2023.107853.

[12] Liu, Z., Sun, H., Zhu, Z., Yan, L., & Luo, T. (2021). Identification of rice-weevil (Sitophilus oryzae L.) damaged wheat kernels using multi-angle NIR hyperspectral data. *Journal of Cereal Science*, 101. https://doi.org/10.1016/j.jcs.2021.103313

[13] Lin, W., Shu, L., Zhong, W., Lu, W., Ma, D., & Meng, Y. (2023). Online classification of soybean seeds based on deep learning. *Engineering Applications of Artificial Intelligence*, 123, 106434. https://doi.org/10.1016/j.engappai.2023.106434

[14] Li, N., Ye, T., Zhou, Z., Gao, C., & Zhang, P. (2024). Enhanced YOLOv8 with BiFPN-SimAM for precise defect detection in miniature capacitors. *Applied Sciences*, 14(1), 429. https://doi.org/10.3390/app14010429

[15] Ni, C., Wang, D., Vinson, R., Holmes, M., & Tao, Y. (2019). Automatic inspection machine for maize kernels based on deep convolutional neural networks. *Biosystems Engineering*, 178, 131–144. https://doi.org/10.1016/j.biosystemseng.2018.11.010

[16] Na, L., Dong, W., Shi-fang, W., & Ping, H. (2019). Progress in terahertz technique for quality inspection of agro-food products. *Spectroscopy and Spectral Analysis*, 39(2), 349-356.

[17] Olgun, M., Onarcan, A. O., Özkan, K., Işik, Ş., Sezer, O., Özgişi, K., Ayter, N. G., Başçiftçi, Z. B., Ardiç, M., & Koyuncu, O. (2016). Wheat grain classification by using dense SIFT features with SVM classifier. *Computers and Electronics in Agriculture*, 122, 185–190. https://doi.org/10.1016/j.compag.2016.01.033

[18] Pearson, T. C., Cetin, A. E., Tewfik, A. H., & Haff, R. P. (2007). Feasibility of impact-acoustic emissions for detection of damaged wheat kernels. *Digital Signal Processing*, 17(3), 617-633. https://doi.org/10.1016/j.dsp.2005.08.002

[19] Qi, H., Luo, J., Wu, X., & Zhang, C. (2024). Application of nondestructive techniques for peach (Prunus persica) quality inspection: *A review. Journal of Food Science*, 89(11), 6863–6887. https://doi.org/10.1111/1750-3841.17388

[20] Qin, J., Kim, M. S., Chao, K., Chan, D. E., Delwiche, S. R., & Cho, B. K. (2017). Line-scan hyperspectral imaging techniques for food safety and quality applications. *Applied Sciences*, 7(2), 125. https://doi.org/10.3390/app7020125

[21] Shen, R., Zhen, T., & Li, Z. (2023). Segmentation of unsound wheat kernels based on improved Mask RCNN. *Sensors*, 23(7), 3379. https://doi.org/10.3390/s23073379

[22] Sun, X., Guo, M., Ma, M., & Mankin, R. W. (2018). Identification and classification of damaged corn kernels with impact acoustics multi-domain patterns. *Computers and Electronics in Agriculture*, 150, 152–161. https://doi.org/10.1016/j.compag.2018.04.008

[23] Sendin, K., Manley, M., & Williams, P. J. (2017). Classification of white maize defects with multispectral imaging. *Food Chemistry*, 243, 311-318. http://dx.doi.org/10.1016/j.foodchem.2017.09.133

[24] Tang, Y., Han, K., Guo, J., Xu, C., Xu, C., & Wang, Y. (2022). GhostNetv2: Enhance cheap operation with long-range attention. *Advances in Neural Information Processing Systems*, 35, 9969–9982. https://doi.org/10.48550/arXiv.2211.12905

[25] Tsuzuki, W., Suzuki, Y., Araki, T., Ohta, S., & Horigane, A. (2014). Effect of oxygen absorber on accumulation of free fatty acids in brown rice and whole grain wheat during storage. *LWT – Food Science and Technology*, 58(1), 222–229. https://doi.org/10.1016/j.lwt.2014.02.015

[26] Wang, L., Liu, J., Zhang, J., Wang, J., & Fan, X. (2022). Corn seed defect detection based on watershed algorithm and two-pathway convolutional neural networks. *Frontiers in Plant Science*, 13, 730190. https://doi.org/10.3389/fpls.2022.730190

[27] Yang, X., Guo, M., Lyu, Q., & Ma, M. (2021). Detection and classification of damaged wheat kernels based on progressive neural architecture search. *Biosystems engineering*, 208, 176-185. https://doi.org/10.1016/j.biosystemseng.2021.05.016

[28] Zhang, Q., Zhen, T., Li, Z., & Gao, H. (2025). MSPCNeXt: Multi-scale parallel convolution for identifying unsound kernels in stored wheat. *Journal of Stored Products Research*, 111, 102559. https://doi.org/10.1016/j.jspr.2025.102559

[29] Zhang, W., Ma, H., Li, X., Liu, X., Jiao, J., Zhang, P., ... & Cao, S. (2021). Imperfect wheat grain recognition combined with an attention mechanism and residual network. *Applied sciences*, 11(11), 5139. https://doi.org/10.3390/app11115139