YOLOv11-SPA: A REAL-TIME VISUAL MODEL FOR SORGHUM SEED DEFECT DETECTION
/
YOLOv11-SPA：一种用于高粱籽粒缺陷检测的实时视觉模型

Sining LIU[1)], Chen LI*[1)]
School of Life Science, Shanxi University, Taiyuan, Shanxi/ China
Tel: +8613834043529; E-mail: lichen@sxu.edu.cn
Corresponding author: Chen Li
DOI: https://doi.org/10.35633/inmateh-78-46

Keywords: *machine vision, sorghum seeds, object detection, YOLOv11*

ABSTRACT
To address the low inspection efficiency and limited recognition accuracy in sorghum grain quality assessment for brewing enterprises, this study proposes YOLOv11-SPA, an efficient and real-time detection model based on an improved YOLOv11n architecture. First, the space-to-depth convolution module (SPDConv) is introduced into the backbone network to replace conventional convolution blocks, effectively mitigating the loss of spatial information for small targets caused by downsampling operations. Second, the parallelized patch-aware attention (PPA) module is integrated into the neck network to enhance local feature representation and improve the detection of subtle defect features such as moldy and cracked grains. Third, an adaptive threshold focal loss (ATFL) is proposed to dynamically adjust sample weights, improving the model's discrimination capability for visually similar categories (e.g., grains with husk residue and intact grains). Experimental results on a self-constructed sorghum seed dataset show that YOLOv11-SPA achieves 80.1% Precision, 79.7% Recall, and 85.9% mAP50, outperforming the baseline YOLOv11n by 5.6, 5.9, and 6.2 percentage points, respectively. With only 3.4 M parameters, the proposed model achieves an inference speed of 205 FPS, meeting real-time detection requirements while maintaining high accuracy. These results demonstrate that YOLOv11-SPA provides an effective solution for automated sorghum grain defect inspection and offers promising potential for intelligent quality control in the modern brewing industry.

摘要
针对现阶段酿酒企业高粱籽粒检测效率及识别率较低等问题，提出了一种基于改进YOLOv11n的高效、精准、实时检测模型 YOLOv11-SPA。首先，在骨干网络中引入空间到深度卷积模块（SPDConv）替代原始卷积模块，有效缓解因池化操作导致的小目标空间信息丢失问题；其次，在颈部网络嵌入并行化块感知注意力模块（PPA），通过强化局部特征表征能力，提升模型对霉斑籽粒、裂纹籽粒等细微特征的感知能力；最后，设计自适应阈值焦点损失（ATFL）函数，通过动态调整样本权重，优化模型对"包壳残留籽粒"与"完好籽粒"等高相似度样本的区分能力。试验结果表明，该模型在自建高粱籽粒数据集上实现了 80.1%精确率、79.7%召回率和85.9%的模型平均精度，较基准模型YOLOv11n 分别提升了 5.6、5.9 和 6.2 个百分点；最终模型在 3.4M 参数下实现每秒 205 帧的推理速度，证明了该模型在保持高精度的同时也满足实时检测需求，对现代化酿造产业具有重要意义。

INTRODUCTION
Sorghum, as the primary raw material for baijiu (Chinese liquor) brewing, is considered one of the best ingredients due to its high liquor yield and the mellow, sweet aroma it imparts. The quality of sorghum grains directly affects the quality of the brewed liquor *(Zhu et al., 2025; Li et al., 2023)*. During harvesting, transportation, and storage, sorghum grains are prone to issues such as breakage, mildew, and sprouting, which not only compromise the liquor's flavor and quality but also reduce the yield. Current quality assessment of sorghum grains largely relies on manual sensory evaluation. This method is susceptible to subjective experience and environmental factors, suffering from low recognition accuracy, poor detection efficiency, and insufficient standardization. It struggles to meet the modern brewing industry's demand for precise and large-scale raw material quality control *(Huai et al., 2019)*.

[1] *Sining Liu, B.S. Stud. Eng.; Chen Li, Prof. Ph.D. Eng.*

Therefore, to achieve accurate quality control of brewing ingredients, developing an objective and efficient automated detection method for sorghum grains (including broken, moldy, sprouted, shriveled, husk-remaining, impurity, and normal grains) holds significant practical importance for ensuring the supply of high-quality brewing grains and enhancing the overall efficiency of the brewing industry *(Huang et al., 2017)*.

In recent years, the rapid advancement of artificial intelligence has driven the widespread application of deep learning in the agricultural sector. As a critical component of modern agricultural detection *(Xue et al., 2022)*, agricultural image analysis employs technologies such as convolutional neural networks and object detection algorithms (e.g., YOLO, Faster R-CNN) to achieve efficient and automated grain detection in crops *(Xia et al., 2024; Zhang et al., 2022)*. Extensive research has been conducted by relevant scholars in this field. In wheat grain detection, *Zhu et al. (2020)* designed a CNN-based image detection system for wheat grain integrity. Compared to models built with SVM and BP neural networks, their system improved the accuracy of grain integrity recognition by up to 5.77%. Pan et al. *(Pan et al., 2023)*, addressing the issue of low detection accuracy for wheat grains, used the Cascade R-CNN model to achieve a mean average precision of 90.2% for wheat grain integrity detection. *Ma et al. (2024)* improved upon YOLOv8n by sharing convolutional layers, reducing parameters, and adding a Deformable Attention Transformer mechanism to address issues like wheat seed adhesion, impurities, and stacking. However, for very small or distant targets, misdetections or missed detections could still occur.

In rice grain detection, *Yao et al. (2025)* improved the YOLOv8n model by replacing the backbone network, adding a Large-Separable-Kernel Attention module, and adopting specific sampling methods. Their model achieved mAP of 96.9% on a dataset of damaged rice grains and impurities, reduced the model size by 30.5%, and significantly improved detection accuracy and efficiency. *Liu et al. (2023)* proposed a lightweight fully convolutional method for rice impurity segmentation. They used an improved EfficientNetV2 network model and introduced a Normalization-based Attention Module to enhance feature extraction performance. The average detection times per image on GPU and CPU devices were 0.103 seconds and 0.301 seconds, respectively, demonstrating the algorithm's lightweight characteristics. In corn grain classification and segmentation, *Li et al. (2024)* proposed a corn seed recognition model based on an improved ResNet50. By introducing a ResStage structure, an Efficient Channel Attention mechanism, and depthwise separable convolutions, they achieved a classification accuracy of 91.23% across six corn varieties. *Du et al. (2025)* adopted Xception as the backbone network and integrated a Squeeze-and-Excitation (SE) attention module into the DeepLab-v3+ model for segmenting corn grain-impurity images. The improved model achieved a Mean Intersection over Union (MIoU) of 92.17% and Pixel Accuracy (PA) of 96.45%, showing better adaptability to complex working environments.

Regarding sorghum grain recognition and detection, *Bu et al., (2022),* proposed a method for sorghum variety identification based on hyperspectral imaging and an AlexNet convolutional network, achieving an average recognition accuracy of over 95.6% on the test set. *Zhao et al., (2024),* proposed a method combining improved Principal Component Analysis (PCA) with a Spectral-Image Convolutional Neural Network for sorghum variety identification, reaching a recognition accuracy of 98.64% on the test set. However, these methods involve high hardware costs and complex data processing, making them difficult to integrate into high-speed brewing production lines. *Wang et al. (2025)* employed image recognition technology to detect unsound sorghum grains based on the principles of sorghum defect identification. However, the reported p-value of 0.650 indicates that the observed differences were not statistically significant. Furthermore, the proposed method was not validated in practical grain counting or quality grading workflows.

Although existing research has shown excellent performance in crop grain recognition, detection, and segmentation, several challenges remain in sorghum grain detection: First, most current research on sorghum grain quality detection relies on hyperspectral imaging, which involves high hardware costs and complex detection procedures. Second, sorghum seed targets are small in scale, appearing as low-resolution, weak-feature regions in images. This is particularly problematic in densely distributed scenarios, where feature overlap can easily lead to missed detections. Third, significant inter-class similarity exists among some sorghum grain categories, posing stringent challenges to the model's feature discrimination capability. Moreover, most existing studies focus primarily on detection metrics without validating whether these models can be effectively translated into practical counting and quality grading workflows.

Addressing the core technical challenges in sorghum grain detection—namely small target precision detection, high-similarity category differentiation, the lack of engineering application validation, and industrial-grade real-time detection requirements—this study takes YOLOv11 as the baseline model and proposes

YOLOv11-SPA, a fast and accurate sorghum grain detection model designed to provide a reliable technical solution for sorghum grain quality inspection.

MATERIALS AND METHODS

Experimental Materials

The sorghum variety used in the experiment was "Fenliang 30". Professional quality inspectors selected and prepared samples including intact grains, broken grains, grains with husk residue, shriveled grains, moldy grains, and sprouted grains, with 50 g of each type, along with some impurities. This effectively simulates the quality sorting conditions of sorghum grains under real brewery factory environments.

Dataset Construction

The image acquisition setup is shown in Fig. 1. It consists of a sample stage, a smartphone, a ring LED shadowless fill light, and matte background boards in both black and white colors. The imaging device was an iPhone 15 Pro Max with 24 megapixels. During acquisition, the camera was vertically fixed directly above the shooting platform, with its optical axis perpendicular to the sample plane to eliminate perspective distortion. The shooting height was consistently maintained at 40 cm.

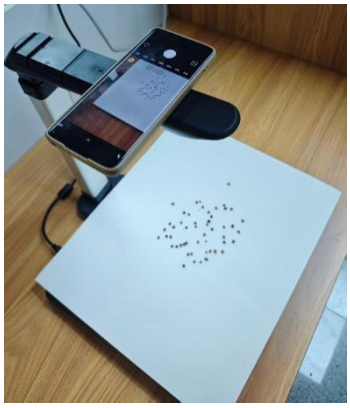

Fig. 1 - Data Acquisition Platform

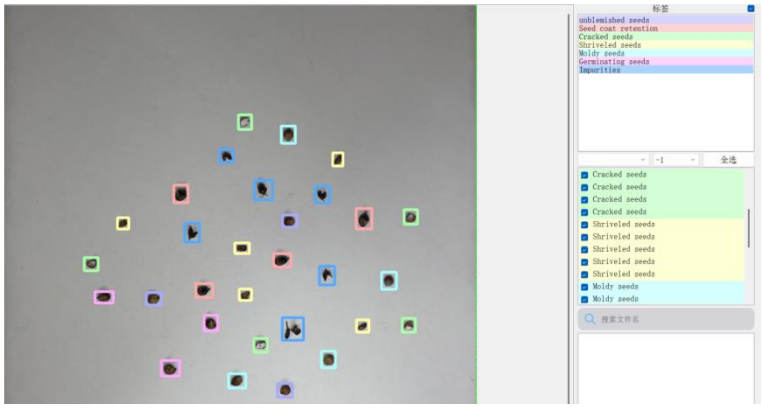

Fig. 2 - Data Annotation Diagram

To systematically investigate the impact of sorghum grain density on the performance of the detection model, four density gradients were established: 30 grains, 60 grains, 90 grains, and 120 grains. Each gradient included all target categories, namely intact grains, broken grains, grains with husk residue, shriveled grains, moldy grains, sprouted grains, and impurities, as illustrated in Fig. 3.

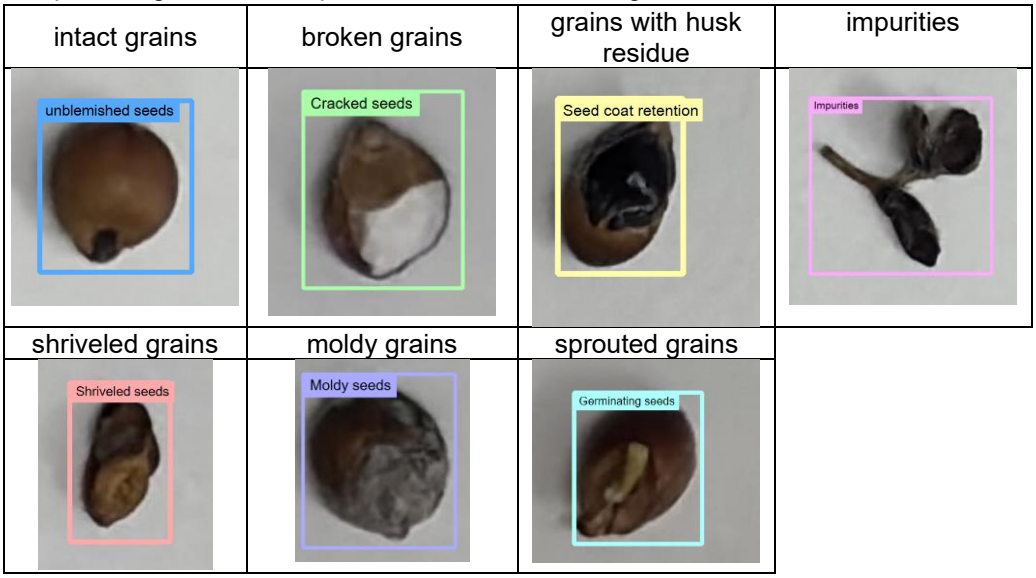

Fig. 3 - Different types of seeds

During the experiment, sorghum grains were manually and randomly scattered on the surface of a stage covered with either a black or white matte board. Ultimately, a high-quality sorghum grain image dataset was constructed, comprising 2 background conditions and 4 density levels.

For each experimental combination, 50 high-resolution images were captured, resulting in a total of 400 sorghum grain image data points, as illustrated in Fig. 4.

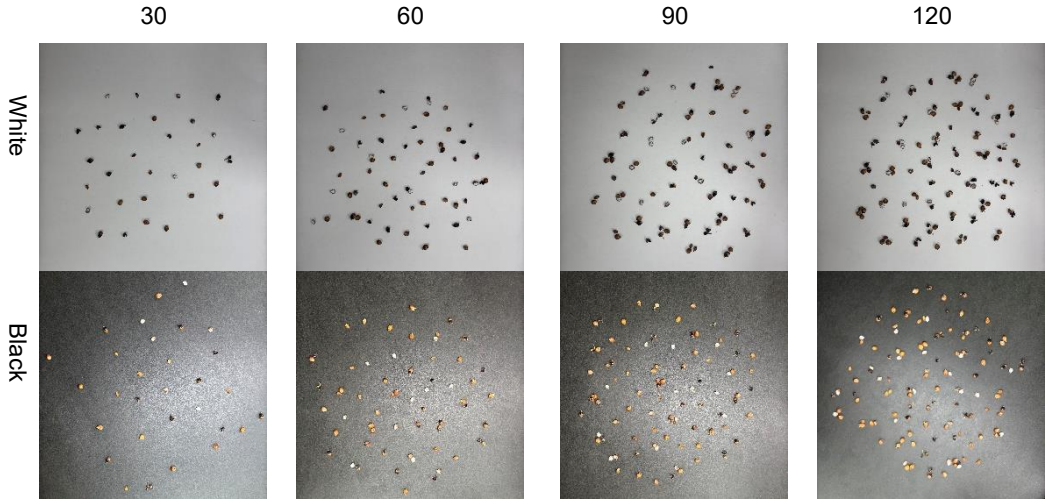

Fig. 4 - Data Acquisition Diagram

Data Annotation

For this study, X-AnyLabeling was selected as the image annotation tool. This tool supports rectangular bounding box annotation and can output label files in multiple formats, meeting the requirements for annotating individual seed targets (as shown in Fig. 2).

During the annotation process, each seed was precisely annotated with a single rectangular bounding box to ensure all seed pixels were completely enclosed. Each bounding box is defined by the coordinates of its top-left corner, its width, and its height, with the corresponding coordinate information saved in normalized form. All annotation results are saved in TXT format, with each image having a corresponding label file of the same name. Each line within these label files contains the class index and bounding box coordinates of a single target, complying with the training format requirements of the YOLO series models.

Data Preprocessing

Considering the limited sample size of the initial dataset, to ensure model robustness during image validation, improve data diversity, enhance the network's generalization capability, and mitigate the risk of overfitting caused by insufficient data volume, this study conducted data preprocessing on the original images to enrich the image data dimensions and expand the dataset scale. Ultimately, the expanded dataset contained 2,800 images. These were randomly divided into training, validation, and test sets according to a 7:2:1 ratio, resulting in 1,960, 560, and 280 image samples per category, respectively.

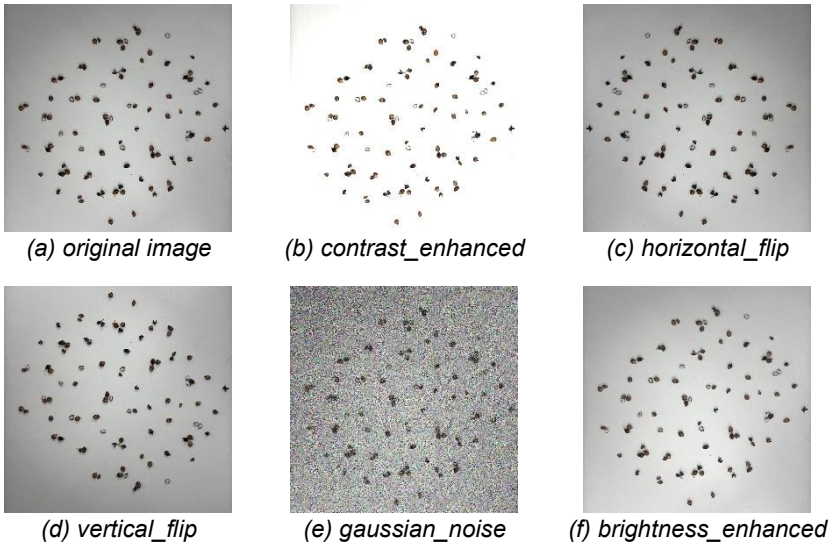

Fig. 5 - Data Augmentation

The data preprocessing methods employed in this study primarily included five data augmentation strategies: contrast enhancement, horizontal mirroring, vertical mirroring, Gaussian noise addition, and brightness enhancement. The enhancement effects are illustrated in Fig. 5.

YOLOv11-SPA Network Architecture

The YOLO model has long been a pioneering force in object detection within the field of computer vision. Its exceptional performance and efficiency have led to its widespread adoption across both academic and industrial applications. With continuous technological evolution, the YOLO series has undergone multiple iterations of optimization, with each generation bringing significant enhancements in accuracy, speed, and applicability. In 2024, Ultralytics released YOLOv11, marking the latest milestone in the YOLO family. Building upon its predecessors like YOLOv8 and YOLOv7, YOLOv11 introduces further refinements in architecture and training efficiency, establishing itself as a state-of-the-art foundation for real-time detection tasks. However, sorghum seeds, being small and often densely distributed in images, present a significant challenge for the original YOLOv11 network, leading to suboptimal detection accuracy for these small targets and difficulty in distinguishing between visually similar defect categories. To address these specific issues, this study proposes an improved YOLOv11 detection network named YOLOv11-SPA. The overall architecture of the model is illustrated in Fig. 6. The specific improvements are as follows: (1) space-to-depth convolution (SPDConv) *(Sunkara et al., 2023)* is adopted to replace standard strided convolutions in the down-sampling path, effectively preserving fine-grained spatial information crucial for detecting small seed targets. (2) A parallelized patch-aware attention (PPA) *(Xu et al., 2024)* module is integrated into the backbone network to enhance the model's focus on subtle defect features by leveraging patch-level contextual information. (3) The adaptive threshold focal loss (ATFL) *(Yang et al., 2024)* function is utilized to dynamically reweight hard and easy samples during training, thereby improving the model's discriminative power for challenging categories and mitigating performance degradation caused by class similarity and imbalance. A detailed description of each improved module is presented in the following subsections.

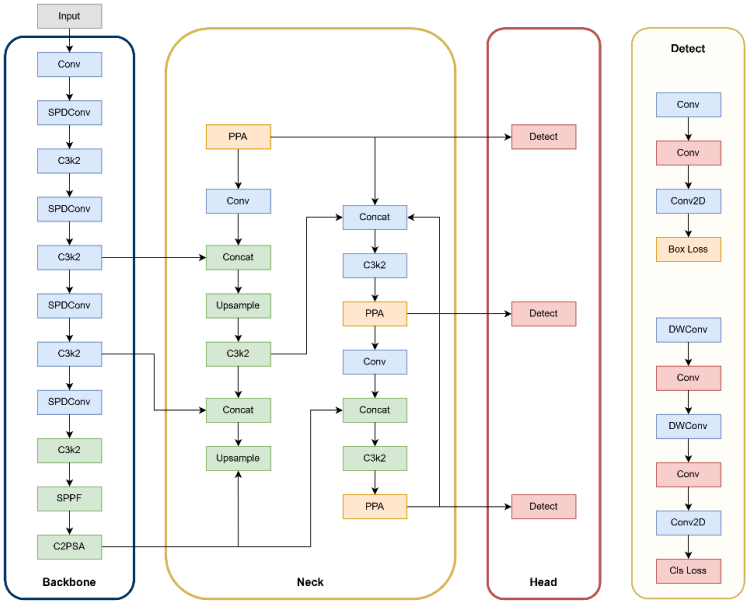

Fig. 6 - The architecture of YOLOv11-SPA

SPDConv Module

In the context of multi-class defect detection for sorghum seeds, small targets—such as cracked fragments, residual seed coats, or early-stage mold spots—often occupy limited pixel areas and exhibit subtle texture variations against a uniform black background. During the standard downsampling process in convolutional backbones, stride-2 convolutions progressively reduce spatial resolution, leading to the irreversible loss of fine-grained spatial details that are critical for distinguishing defect types. To address this challenge, the SPDConv module is integrated into the backbone of the improved YOLOv11 architecture, replacing conventional strided convolutions, particularly in the shallow and intermediate layers where small-target features are most susceptible to degradation.

As illustrated in Fig. 7, the SPDConv module consists of a space-to-depth transformation followed by a standard non-strided convolution. Given an input feature map and a downsampling scale factor of 2, the SPD operation first partitions into four non-overlapping sub-feature maps:

$$\begin{aligned} f_{0,0} &= X[0:H:2,0:W:2,], \\ f_{0,1} &= X[0:H:2,1:W:2,], \\ f_{1,0} &= X[1:H:2,0:W:2,], \\ f_{1,1} &= X[1:H:2,1:W:2,], \end{aligned} \tag{1}$$

which are then concatenated along the channel dimension to form an intermediate tensor. This rearrangement preserves all original pixel information without averaging or discarding spatial content, effectively achieving resolution reduction while maintaining full spatial fidelity. Subsequently, a non-strided (stride=1) convolution is applied to compress the channel dimension and introduce learnable feature interactions, yielding a compact yet information-rich downsampled representation.

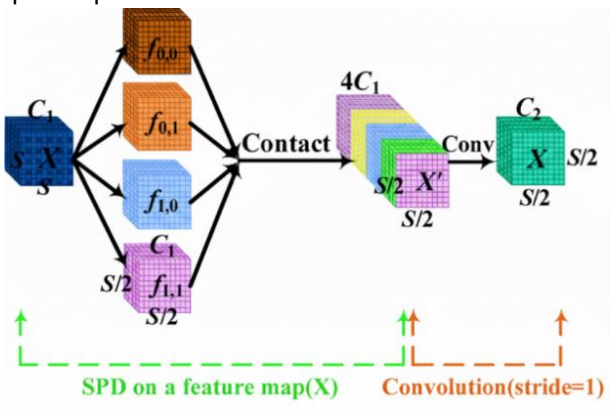

Fig. 7 - SPDConv Module

By replacing Conv module with SPDConv, our model avoids the blurring and aliasing effects that often cause missed detections of fine defects such as hairline cracks or faint mold spots. In high-density sorghum seed images, this design effectively improves the recall rate of small targets, markedly strengthening the model's feature activation response for tiny seeds. This effectively mitigates common issues in small target detection.

PPA Module

In sorghum seed detection tasks, small targets are prone to critical information loss during successive downsampling operations. To address this challenge, we introduce the parallelized patch-aware attention (PPA) module into the neck of YOLOv11. The PPA module is designed to enhance the model's ability to capture and preserve local discriminative features critical for distinguishing between the seven fine-grained categories of sorghum seeds. Unlike traditional attention mechanisms that operate globally across the entire feature map, PPA focuses on extracting and enhancing information within localized patches, making it particularly effective for small object detection tasks. PPA replaces standard convolution operations with a multi-branch architecture designed to preserve and enhance fine-grained features of small seeds.

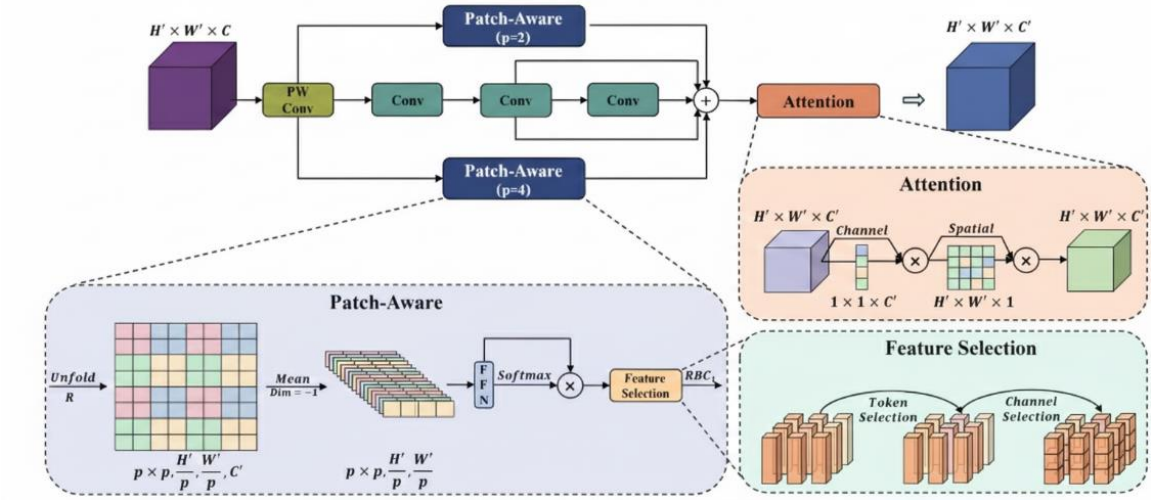

Fig. 8 - Structure of the PPA module

As illustrated in Fig. 8, the PPA module consists of three parallel branches: a local branch, a global branch, and a serial convolution branch. These branches work in concert to extract multi-scale features and fuse them with adaptive attention mechanisms, thereby improving the model's sensitivity to subtle defects while maintaining computational efficiency. The core innovation of PPA lies in its parallel multi-branch structure, which captures features at multiple scales and hierarchies. Given an input feature tensor, it is first compressed via point-wise convolution to obtain. The three parallel branches—local, global, and serial convolution—then process independently. The outputs of all branches are aggregated to form, which integrates multi-scale representations critical for detecting diminutive seeds. Following the multi-branch feature extraction, the PPA module employs an adaptive attention mechanism to further enhance the salience of key features. This mechanism comprises two sequential components: channel attention (CA) and spatial attention (SA). The final output exhibits enhanced saliency and precise localization of small seeds.

ATFL Function

In the context of multi-class defect detection for sorghum seeds, the dataset exhibits two types of imbalances that challenge model convergence and accuracy. First, foreground-background imbalance, as seeds occupy a small proportion of the image against a uniform black background. Second, inter-class imbalance, where "intact" seeds are significantly more abundant than rare defect types such as "cracked" or "moldy" seeds. Under these conditions, the model tend to be dominated by easy background or majority-class samples. This suppresses gradients from critical small targets and hard-to-classify defects, leading to missed or misclassified detections.

To address this issue, this study introduces an ATFL function, it explicitly divides samples into "hard samples" (targets) and "easy samples" (background) by setting a probability threshold and applies distinct loss adjustment strategies to each. The core formula is as follows:

$$L_{ATFL} = \begin{cases} -(\lambda - p_t)^{\eta} \log(p_t), p_t \leq 0.5 \text{(Hard Samples)} \\ -(1 - p_t)^{\gamma} \log(p_t), p_t > 0.5 \text{(Easy Samples)} \end{cases} \quad (2)$$

To further enhance generalization and reduce the cost of manual hyperparameter tuning, ATFL incorporates adaptive improvements to the hyperparameters. The exponentially smoothed predicted confidence is utilized to dynamically represent the model's training state:

$$\widehat{p_c} = 0.05 \times \frac{1}{t-1} \sum_{i=0}^{t-1} p_i + 0.95 \times p_t \quad (3)$$

Subsequently, based on information theory principles, the modulating factors are defined as:

$$\gamma = -\ln(\widehat{p_c}) \quad (4)$$

$$\eta = -\ln(p_t) \quad (5)$$

By integrating these mechanisms, ATFL effectively suppresses background and easy seed regions to prevent gradient domination while amplifying the loss for small or ambiguous defects such as early mold spots or fine cracks. It can also automatically tune hyperparameters, eliminating the need for manual adjustment across different seed densities or defect types. This adaptive design is particularly beneficial for high-density sorghum seed images, where small targets are easily overlooked.

Application Framework for Quality and Quantity Estimation

To validate the engineering practicality of the proposed model beyond algorithmic metrics, a practical application framework for sorghum quality estimation was designed. This framework simulates the operational workflow in a brewing enterprise, where raw material quality is assessed based on grain count and defect rates. The workflow consists of three stages. First, raw sorghum images are captured using the imaging platform. Second, the YOLOv11-SPA model processes the images to identify and localize individual grains. The total number of grains (N_{total}) and the number of defective grains (N_{defect}) are calculated by counting the predicted bounding boxes for each category.

The final step is to calculate quality grading, the unsound grain rate (R_{defect}), a key indicator for brewing quality, is computed as:

$$R_{defect} = \frac{N_{defect}}{N_{total}} \times 100\% \quad (6)$$

This framework enables automated quality assessment, replacing traditional manual inspection. To evaluate its effectiveness, a comparative experiment was conducted between manual counting and model-based counting, focusing on accuracy, defect rate estimation error, and time efficiency.

Evaluation Metrics

To evaluate the performance of the model in the sorghum grain detection task, this study employs six evaluation metrics for validation: Precision, Recall, Mean Average Precision (mAP), Giga Floating-point Operations Per Second (GFLOPs), number of parameters (Params), and Frames Per Second (FPS). These metrics cover four dimensions: detection accuracy, robustness, computational efficiency, and model complexity, thereby comprehensively reflecting the detection effectiveness of the model.

Precision (P) is defined as the proportion of correctly identified targets among all detection results, which measures the accuracy and reliability of the model's predictions. Recall (R) represents the proportion of actual existing targets that are correctly identified by the model, reflecting the model's coverage capability and detection comprehensiveness for real targets. The mean Average Precision (mAP50), as a comprehensive evaluation metric, calculates the mean of the average precision across all categories under an Intersection over Union (IoU) threshold of 0.5. This metric is used to comprehensively assess the overall performance of the model in both localization and classification tasks. Floating-point Operations Per Second (FLOPs) reflects the computational burden during the model's forward inference process, directly impacting the energy efficiency and speed of the model in practical deployment. The number of parameters (Params) represents the scale of learnable parameters in the model. The Frames Per Second (FPS) metric quantifies the real-time processing capability of the model under specified hardware conditions. The calculations are detailed as follows:

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

$$R = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \times 100\% \tag{8}$$

$$P = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \times 100 \tag{9}$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \times 100\% \tag{10}$$

$$\mathrm{mAP} = \frac{1}{m}\sum_{i=1}^{m} \int_0^1 P(R)\mathrm{d}R \tag{11}$$

Among these, TP (True Positive) represents the number of samples that are actually positive and correctly predicted as positive by the model. FP (False Positive) represents the number of samples that are actually negative but incorrectly predicted as positive by the model. FN (False Negative) represents the number of samples that are actually positive but incorrectly predicted as negative by the model. TN (True Negative) represents the number of samples that are actually negative and correctly predicted as negative by the model, and m denotes the total number of categories.

RESULTS AND DISCUSSION

Experimental Environment and Parameter Settings

To ensure the smooth progress of the experiment, the experimental environment is shown in Table 1.

Table 1

Experimental environment configuration

Hardware configuration	Parameter
Operating system	Windows 11
CPU	Intel Core 13700H
GPU	NVIDIA GeForce RTX 4060
RAM	16 GB
Dependency	PyTorch 1.12.1+CUDA 11.6 + Python 3.8.16

In the sorghum grain detection task, the specific experimental parameters are detailed in Table 2.

Table 2

Experimental parameter setting

Parameter	Value	Parameter	Value
Input image	736×736	lrb	0.01
Batch size	16	lrf	0.01
Training epoch	200	warmup_bias_lr	0.1
Momentum factor	0.937	warmup_momentum	0.8
Weight parameter	0.0005	warmup_epochs	3.0

Experimental Results

Ablation Experiments

To validate the optimization efficacy of the three proposed improvement methods—PPA, SPDConv module, and ATFL function—on the YOLOv11 model, systematic ablation experiments were conducted on a self-constructed sorghum seed dataset. The mAP50 served as the core evaluation metric. This was complemented by comprehensive performance assessments based on precision, recall, FLOPs, and FPS. The experimental results are presented in Table 3.

The results indicate that all three improvement modules independently contribute positive gains to model performance and exhibit significant functional complementarity. When the PPA module was introduced individually, the model's mAP50 increased from the baseline of 79.7% to 83.3%, representing an absolute improvement of 3.6 percentage points. Precision and recall also improved simultaneously, fully validating the module's effectiveness in enhancing the model's ability to perceive and distinguish subtle defect features such as mold spots and micro-cracks. After independently replacing the traditional downsampling operation with the SPDConv module, the model's mAP50 increased by 1.1 percentage points to 80.8%, while FLOPs decreased by approximately 14% and the inference speed improved to 220 FPS. This not only effectively alleviated the issue of feature loss for small-scale seed targets but also endowed the model with favorable lightweight characteristics. When the ATFL loss function was applied individually, the model's mAP50 increased by 2.3 percentage points to 82.0%, with a particularly notable improvement in recall. This benefit arises from its ability to dynamically adjust sample loss weights, guiding the model to focus on learning challenging samples with high similarity, such as "husk residue," thereby reducing the missed detection rate.

Table 3

Ablation experiments

SPDConv	PPA	ATFL	P (%)	R (%)	mAP50 (%)	FLOPs (G)	Params (M)	FPS (F/S)
-	-	-	74.5	73.8	79.7	6.4	2.4	203
√	-	-	75.3	74.4	80.8	5.5	2.1	220
-	√	-	77.3	77.8	83.3	7.6	3.4	197
-	-	√	76.4	74.8	82.0	6.4	2.4	203
√	√	-	73.6	74.9	79.5	7.5	3.4	195
-	√	√	77.7	80.1	84.6	7.6	3.4	205
√	-	√	76.4	78.1	83.6	5.5	2.1	220
√	√	√	80.1	79.7	85.9	7.5	3.4	205

Module combination experiments further confirmed the synergistic enhancement effects among the improvements. The combination of PPA and ATFL achieved an mAP50 of 84.6%, with performance gains exceeding the simple sum of their independent contributions, demonstrating deep synergy between feature enhancement and gradient optimization. The combination of SPDConv and ATFL achieved an mAP50 of 83.6% while maintaining a high inference speed of 220 FPS, providing an excellent solution for real-time detection scenarios. When all three improvements were integrated jointly, the model achieved optimal comprehensive performance—mAP50 increased to 85.9%, representing an absolute improvement of 6.2 percentage points compared to the baseline, while the inference speed remained at a high level of 205 FPS, and FLOPs decreased by 12% compared to the baseline. This dual enhancement in "accuracy-speed" stems from the efficient synergy among the three components: SPDConv constructs a lightweight yet feature-preserving backbone foundation, PPA focuses on key defect features to strengthen perception capability, and ATFL ensures the model's learning effectiveness for difficult samples through a dynamic loss mechanism. Together, they enable the improved YOLOv11 model to efficiently achieve simultaneous identification of sorghum varieties and detection of multiple physiological defects.

This holds significant practical importance for completing precise sorghum sorting before brewing or sowing, ensuring the supply of high-quality brewing grains, and enhancing the overall efficiency of the brewing industry.

Comparison Experiments with Different Object Detection Models

To further validate the performance of the model in sorghum grain detection tasks, the improved YOLOv11-SPA model was compared and evaluated against mainstream models from the YOLO series, including YOLOv5, YOLOv8, YOLOv9, YOLOv10 and YOLOv11. The comparative experiments were conducted under identical hardware environments and software frameworks, with all participating models trained using the exact same configuration. The results are shown in Table 4.

The experimental data demonstrates that the YOLOv11-SPA model exhibits significant advantages across multiple key metrics. In terms of core detection indicators, its mAP50 reaches 85.9%, which is 5.7%, 5.4%, 4.8%, 5.4%, and 6.2% higher than those of YOLOv5, YOLOv8, YOLOv9, YOLOv10, and YOLOv11, respectively. Its P value outperforms the other models by 6.3%, 5.9%, 3.8%, 5.5%, and 5.6%, respectively, indicating a significantly improved ratio of identified targets in the detection results and effectively reducing the false detection rate. Regarding computational efficiency, the FLOPs of the improved model decreased by 8.5%, 5.1%, and 10.7% compared to YOLOv8, YOLOv9, and YOLOv10, respectively, while remaining comparable to the compact YOLOv5, effectively controlling computational overhead. In terms of inference speed, it has 205 FPS also leads among comparative models. Regarding model size, although the parameter count of the improved model is 3.4 M, slightly higher than that of the other models, the substantial improvements in accuracy and speed far outweigh the minor increase in parameters. In summary, the improved model not only significantly surpasses mainstream models in detection accuracy but also excels in real-time performance and computational efficiency, fully validating its advanced nature and effectiveness in automated sorghum grain detection tasks.

Table 4

Comparison of detection results

Model	P (%)	R (%)	mAP50 (%)	FLOPs (G)	Params (M)	FPS (F/S)
YOLOv5	73.8	77.8	80.2	7.2	2.5	187
YOLOv8	74.2	75.1	80.5	8.2	3.0	186
YOLOv9	76.3	74.9	81.1	7.9	2.0	194
YOLOv10	74.6	73.8	80.5	8.4	2.7	189
YOLOv11	74.5	73.8	79.7	6.4	2.4	203
YOLOv11-SPA	80.1	79.7	85.9	7.5	3.4	205

Visualization Results and Analysis

To provide a more intuitive demonstration of the detection performance of the improved model under different scenarios and densities, a visual comparison was conducted with the YOLOv11 model. The visualization results indicate that the YOLOv11n model exhibits performance limitations under both background conditions. Under the white background condition, as shown in Fig. 8, due to the low contrast between sorghum grains and the background, subtle texture and color differences on the grain surface are diminished, leading to a primary error mode of category misclassification (such as confusion between "grains with husk residue" and "intact grains").

Under the black background condition, as shown in Fig. 9, although the color contrast between sorghum grains and the background is higher, issues such as insufficient feature extraction for densely occluded and small marginal targets in high-density samples (e.g., 120 grains) are amplified, with target missed detection becoming the dominant error. The improved YOLOv11-SPA model, through the synergistic effects of structural optimization, feature enhancement, and loss guidance, demonstrates comprehensive and robust detection performance in both scenarios. Notably, the improved model achieves high classification consistency and localization accuracy under the white background. Furthermore, it maintains a stable and low misclassification rate across various density levels under the black background.

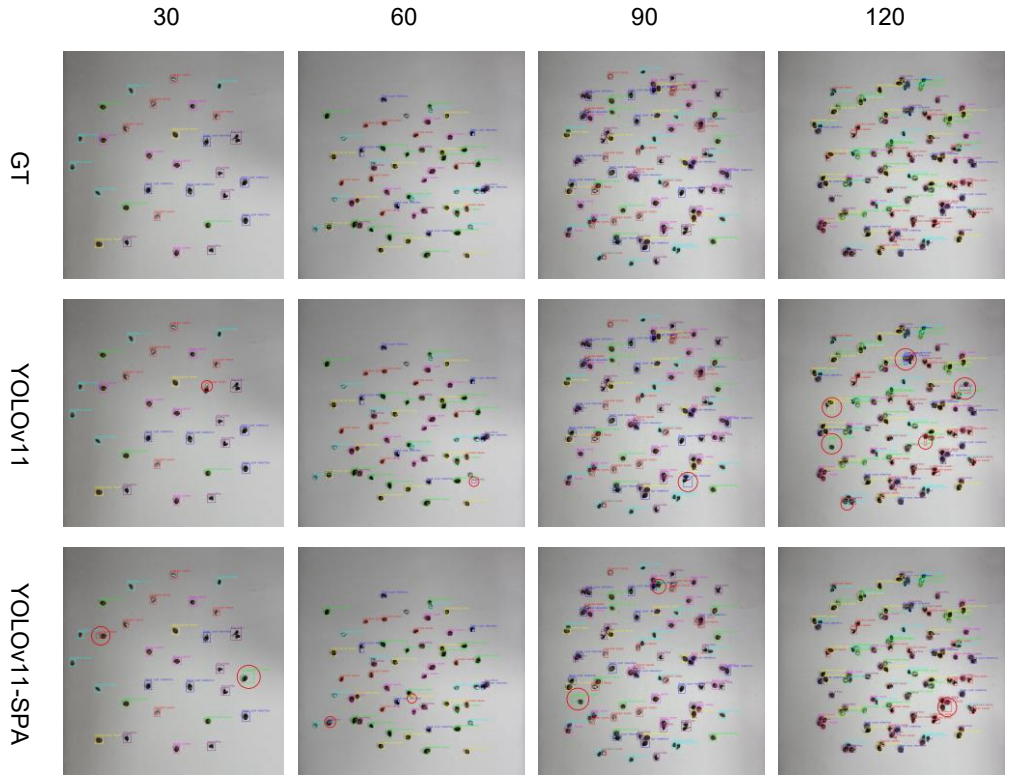

Fig. 8 - Visualization of White Background Experiment

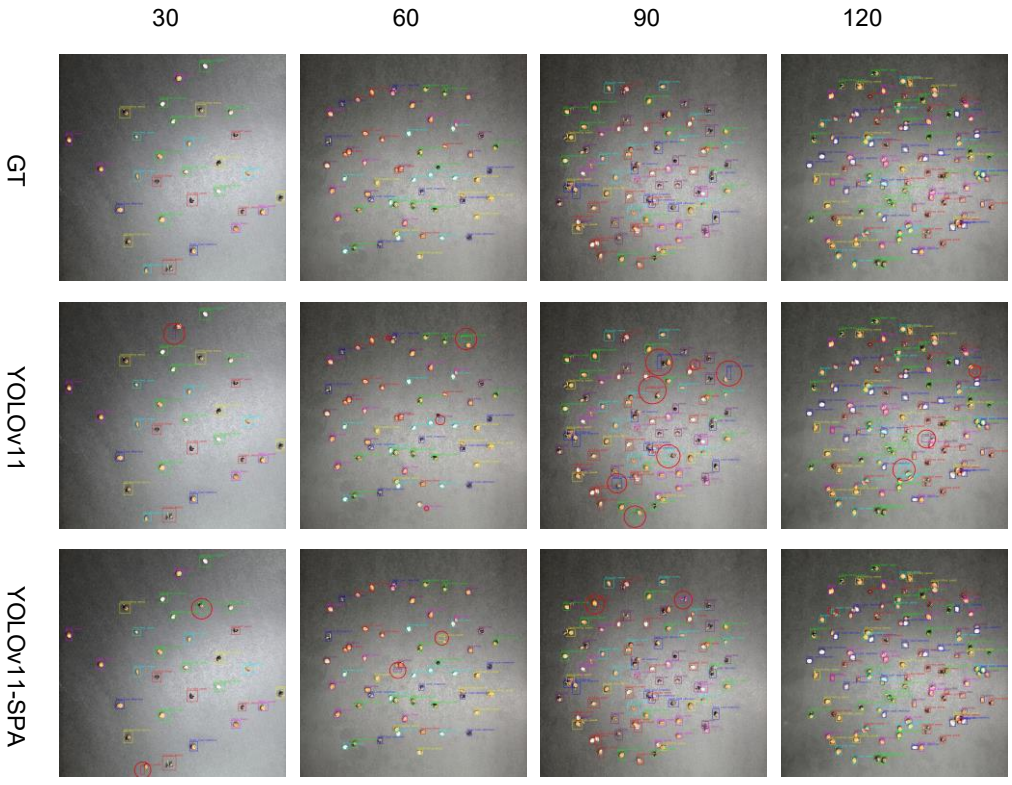

Fig. 9 - Visualization of Black Background Experiment

Validation of Counting and Quality Grading Performance

To assess the model's potential in real-world engineering applications, a yield estimation experiment was conducted comparing manual inspection versus the proposed YOLOv11-SPA model. Eight seed samples with random numbers were tested for defective seeds using both manual and model-based methods. The performance was evaluated based on counting accuracy, unsound grain rate, and processing time.

Table 5 presents the comparative results, the "Rate Error" column represents the absolute difference between these two rates:

$$\text{rate error} = |R_{defect}^{model} - R_{defect}^{manual}| \quad (12)$$

The manual inspection served as the ground truth. The results indicate that the YOLOv11-SPA model achieves an average counting accuracy of 96.4%, demonstrating strong capability in dense target scenarios. For quality grading, the average rate error was only 0.325%, which meets the industrial requirement for raw material screening.

Table 5

Comparison of Manual vs. Model-based Counting and Quality Estimation

Sample ID	N_{total} (Manual)	N_{total} (Model)	Counting Accuracy (%)	R_{defect} (Manual)	R_{defect} (Model)	Rate Error	Model Detect Time (s)	Manual Count Times (s)
01	30	29	96.7	0.100	0.102	0.002	0.023	67
02	60	58	96.7	0.150	0.154	0.004	0.021	80
03	45	44	97.8	0.156	0.157	0.002	0.022	73
04	90	86	95.6	0.156	0.151	0.005	0.025	95
05	75	72	96.0	0.147	0.154	0.007	0.024	88
06	120	114	95.0	0.150	0.151	0.001	0.028	112
07	105	100	95.2	0.152	0.150	0.002	0.026	103
08	50	49	98.0	0.160	0.163	0.003	0.022	75

In terms of efficiency, the model processed each image in approximately 0.02 seconds, whereas manual counting required an average of 86.6 seconds per image. This represents a 9000-fold increase in efficiency, significantly reducing labor costs and enabling real-time monitoring on production lines.

Although minor counting errors occurred in high-density samples due to severe occlusion, the overall performance confirms that YOLOv11-SPA is not only an accurate detection algorithm but also a viable tool for automated quality control in the brewing industry. This validates the engineering relevance of the proposed method.

CONCLUSIONS

To address the challenges of small target size, dense spatial distribution, and high inter-class feature similarity in the automated detection of sorghum grains, this study proposes the YOLOv11-SPA improved model based on YOLOv11n. By introducing the SPDConv module, the PPA Module, and the ATFL function for collaborative optimization, the model's detection performance has been significantly enhanced. The main conclusions are as follows:

1) The improved YOLOv11-SPA model achieves an mAP50, P, R, FLOPs, parameter count, and inference speed of 85.9%, 80.1%, 79.7%, 7.5G, 3.4M, and 205 FPS, respectively. Compared to the original YOLOv11n model, the mAP50, P, and R are improved by 6.2%, 5.6%, and 5.9%, respectively, while maintaining an inference speed of 205 FPS. Compared to mainstream models, the improved model strikes a balance between detection accuracy and real-time performance.

2) The proposed YOLOv11-SPA model outperforms other mainstream models in sorghum grain detection tasks. By introducing the SPDConv Module in the backbone network, embedding the PPA Module in the neck network, and replacing the original loss function with the ATFL function, the model enhances the transmission of small target details and the capture of local features. This enables the algorithm to effectively reduce both missed and false detection rates while maintaining high operational efficiency, thereby improving detection accuracy. Visualization of detection results further confirms that the improved model outperforms the original YOLOv11n model. Under various background conditions and densities, the phenomena of missed detections, false detections, and localization drift are significantly reduced, demonstrating stable detection performance and highlighting the model's superior advantages in sorghum grain detection tasks.

3) The practical application experiment demonstrated the model's engineering value. In the comparative counting and quality estimation task, the model achieved a counting accuracy of 96.4% with a defect rate error of only 0.325%, while improving inspection efficiency by over 9000 times compared to manual methods. This confirms its potential for deployment in automated brewing production lines.

4) Although the proposed YOLOv11-SPA model achieves promising results in sorghum grain detection tasks, it still exhibits instances of missed and false detections under high-density conditions, indicating that the model's instance separation capability has room for further improvement.

ACKNOWLEDGEMENTS

This project is financially supported by Shanxi Province College Students' Innovation and Entrepreneurship Training Program (No. 202510108063). The authors declare no competing interests.

REFERENCES

[1] Bu, Y., Jiang, X., Tian, J., Hu, X., Han, L., Huang, D., & Luo, H., (2022). Rapid Nondestructive Detection of Sorghum Varieties Based on Hyperspectral Imaging and Convolutional Neural Network. *Journal of the Science of Food and Agriculture*, Vol. 103, No. 8, pp. 3970-3983, United Kingdom.

[2] Du, Y.F., Hou, S.Y., Li, G.R., (2025). Detection Method of Maize Kernel Impurity Based on Deep Learning (基于深度学习的玉米籽粒含杂检测方法研究). *Journal of Agricultural Mechanization Research*, No. 9, pp. 1-8, Changchun/China.

[3] Huang, K.Y., Cheng, J.F., (2017). A Novel Auto-Sorting System for Chinese Cabbage Seeds. *Sensors*, Vol. 17, pp. 886, Switzerland.

[4] Huang, S.S., Guo, Q.L., Dong, W.W., (2019). Feasibility of the Testing Method for Impurities and Incomplete Grains in Liquor-Making Grains (酿酒粮食杂质及不完善粒检验方法的适用性研究). *Liquor-Making Science & Technology*, Vol. 299, No.5, pp. 79-82, Guiyang/China.

[5] Li J., Xu F., Song S. and Qi J., (2024). A Maize Seed Variety Identification Method Based on Improving Deep Residual Convolutional Network. *Frontiers in Plant Science*, Vol. 15, pp. 1382715, Switzerland.

[6] Li, X.H., Chu, Y.H., Mao, Y.Z., (2023). Research and Verification of Detection Model for Imperfect Grain Detector of Brewing Sorghum (酿酒高粱不完善粒检测仪检测模型的研究与检验). *Science and Technology of Cereals, Oils and Foods*, Vol. 31, No. 1, pp. 129-134, Beijing/China.

[7] Liu, Q.H., Liu, W.K., Liu, Y.S., Zhe, T.T., Ding, B.C., Liang, Z.W., (2023). Rice Grains and Grain Impurity Segmentation Method Based on a Deep Learning Algorithm-NAM-EfficientNetv2. *Computers and Electronics in Agriculture*, Vol. 209, pp. 107824, Netherlands.

[8] Ma, N., Su, Y., Yang, L., Li, Z., & Yan, H., (2024). Wheat Seed Detection and Counting Method Based on Improved YOLOv8 Model. *Sensors*, Vol. 24, No. 5, pp. 1654, Switzerland.

[9] Pan, W.T., Sun, M.L., Yuan, Y., Liu, P., (2023). Wheat Kernel Phenotype Identification Method Based on Deep Learning ImCascade R-CNN (基于深度学习 ImCascade R-CNN 的小麦籽粒表形鉴定方法). *Smart Agriculture*, Vol. 5, No. 3, pp. 110-120, Beijing/China.

[10] Sunkara, R., Luo, T., (2023). No More Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects. *Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2022)*, *Lecture Notes in Computer Science*, Vol. 13715, Switzerland.

[11] Wang, J.K., Zhao, Q., Yao, T., Wang, Q.S., Chen, S., Qi, K., (2025). Application Research of Image Recognition Technology in the Detection of Imperfect Sorghum Grains (图像识别技术在高粱不完善粒检测上的应用研究). *Liquor-Making Science & Technology*, No. 5, pp. 61-65, Guiyang/China.

[12] Xia, Y., Che, T., Meng, J., Hu, J., Qiao, G., Liu, W., Kang, J., & Tang, W., (2024). Detection of surface defects for maize seeds based on YOLOv5. *Journal of Stored Products Research*, Vol. 105, pp. 102242, United Kingdom.

[13] Xue, J.L., Li, Y.Q., Cao, Z.J., (2022). Obstacle Detection Technology in Blurred Farmland Images Based on Deep Learning (基于深度学习的模糊农田图像中障碍物检测技术). *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 53, No. 3, pp. 234-242, Beijing/China.

[14] Xu, S.B., Zheng, S.C., Xu, W.H., Xu, R.T., Wang, C.W., Zhang, J.G., Teng, X.Q., Li, A., & Guo, L., (2024). HCF-Net: Hierarchical Context Fusion Network for Infrared Small Object Detection. *2024 IEEE International Conference on Multimedia and Expo (ICME)*, Niagara Falls, ON, Canada, pp. 1-6.

[15] Yang, B., Zhang, X., Zhang, J., Luo, J., Zhou, M., & Pi, Y., (2024). EFLNet: Enhancing Feature Learning Network for Infrared Small Target Detection. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 62, pp. 1-11, Art no. 5906511, United States.

[16] Yao, H.Z., Wang, K., Wang, Y.D., Li, W.T., Liu, Q.H., Shi, J.L., (2025). Rice Kernel Detection Model Based on Improved YOLOv8n (基于改进 YOLOv8n 的水稻籽粒检测模型). *Journal of Chongqing Technology and Business University (Natural Science Edition)*, Chongqing/China.

[17] Zhang, Y., Lv, C., Wang, D., Mao, W., & Li, J., (2022). A novel image detection method for internal cracks in corn seeds in an industrial inspection line. *Computers and Electronics in Agriculture*, Vol. 197, pp. 106930, Netherlands.

[18] Zhao, G., Xu, Z., Tang, L., Li, X., Zhang, P., & Wang, Q., (2024). A Rapid Classification Method for Sorghum Seed Varieties Based on HSI and PCA-SICNN Algorithm (基于 HSI 与 PCA-SICNN 算法的高粱种子品种快速分类方法). *Microchemical Journal*, Vol. 205, pp. 1392-1397, Netherlands.

[19] Zhu, K.X., Xiao, Y., Zhao, M.H., Guo, M.Y., Zhao, J.S., (2025). Effect of different varieties of sorghum on the flavor quality of sauce-flavor Baijiu (不同品种高粱对浓香型白酒风味品质的影响研究). *China Brewing*, Vol. 44, No. 8, pp. 199-205, Sichuan/China.

[20] Zhu, S.P., Zhuo, J.X., Huang, H., (2020). CNN-Based Image Detection System for Wheat Kernel Integrity (基于 CNN 的小麦籽粒完整性图像检测系统). *Transactions of the Chinese Society for Agricultural* Machinery, Vol. 51, No. 5, pp. 36-42, Beijing/China.