RESEARCH ON A MAIZE IMPERFECT KERNEL DETECTION SYSTEM BASED ON CALFNET
/
基于 CALFNet 的玉米不完善粒检测系统研究

Yangchun LIU*[1)], Shicong GE[1)], Gaoyong XING[1)], Biman HAN[1)], Yakai HE[1)], Xue DENG[1,2)], Xiaoyang LIU[1)]
[1)] Chinese Academy of Agricultural Mechanization Sciences Group Co., Ltd,
State Key Laboratory of Agricultural Equipment Technology, Beijing 100083, China
[2)]State Key Laboratory of Agricultural Equipment Technology, Guangzhou, 510642, China;
Tel: +86 15988365890; E-mail: lyc327@163.com

DOI: https://doi.org/10.35633/inmateh-78-13

Keywords: *Imperfect kernels, Maize, Appearance quality inspection, Machine vision, Large Language Model*

ABSTRACT
To address the limitations of existing detection methods for imperfect maize kernels during grain acquisition and storage — specifically limited detection categories, difficulty in identifying minute defects, insufficient robustness in individual kernel extraction, and incomplete information from single-view inspection —this paper proposes a multi-stage detection method based on CALFNet. The proposed method begins with image acquisition using a standardized imaging system, followed by a single-kernel extraction model that integrates YOLOv8 with the watershed algorithm to segment densely distributed maize kernels. The Hungarian algorithm is then employed to automatically match the front and back images of the same kernel. Based on this process, a dual-stream feature fusion classification network, CALFNet, was constructed, integrating EfficientNet-B0 and MobileNetV2 to fuse visual information from both sides of the kernels. A GUI-based detection system was subsequently developed, and the DeepSeek V3 large language model was incorporated to analyse the classification results, enabling the automatic generation of quality evaluation reports and production guidance recommendations. Experimental results show that the CALFNet model achieved a classification accuracy of 99.16% on the test set, outperforming the comparative models. In a full-pipeline integrated test on 506 real-world samples, the overall recognition and classification accuracy reached 96.05%. This study provides a feasible solution for the intelligent assessment of maize quality.

摘要
为解决当前粮食收储过程中玉米不完善粒检测方法存在检测类别单一、微小缺陷识别困难、籽粒单粒化提取鲁棒性不足及单面检测信息不全面等问题，本文研究并设计了一种基于 CALFNet 的多阶段检测方法。该方法首先通过标准化图像采集装置获取图像，并采用一种融合 YOLO v8 与分水岭算法的单粒化提取模型分割密集分布的玉米籽粒，再利用匈牙利算法实现同一籽粒正反面图像的自动配对。在此基础上，本文构建了名为 CALFNet 的双流特征融合分类网络，其基于 EfficientNet-B0 和 MobileNet V2，旨在深度融合来自籽粒正反两面的视觉信息。基于上述方法本文搭建了 GUI 检测系统，并引入 DeepSeek V3 大语言模型对分类结果进行分析，以自动生成质量评价报告与生产指导建议。实验结果表明，CALFNet 模型在测试集上的分类准确率高达 99.16%，性能优于对照模型；在对 506 个真实样本进行的全流程集成测试中，整体识别分类准确率为 96.05%。为玉米品质的智能化评估提供了解决方案。

INTRODUCTION
Maize is a cornerstone of global food security and industrial development (*Erenstein et al., 2022*). Establishing rapid detection technologies for imperfect kernels is essential for ensuring food safety and determining grain grades during acquisition (*Nie et al., 2022*). While machine vision has significantly improved the automation of kernel inspection over manual methods *(Wang et al., 2023*), the complex morphology of defective kernels in diverse environments makes traditional segmentation based on fixed thresholds highly unreliable (*Guo et al., 2023*). Consequently, deep learning has become critical for robust feature learning (*Yao et al., 2022*).

Currently, deep learning applications in this field typically utilize single-stage or multi-stage algorithms. Single-stage models achieve high real-time performance *(Liu et al., 2024; Zhang et al., 2024; Telçeken et al., 2024*), but their single-view nature frequently misses unexposed asymmetric damages.

Multi-stage algorithms address this by decoupling detection into kernel singulation and classification. However, during the singulation stage, methods relying on manually crafted features—such as HSV colour space (*Yao et al., 2025*), traditional watershed (*Wang et al., 2021*), or adaptive thresholds (*Ni et al., 2019*)—lack robustness when extracting densely adhered kernels. Furthermore, in the subsequent classification stage, because defects vary drastically from macroscopic morphological changes to microscopic local lesions, existing single-backbone networks (*Chen et al., 2023*) struggle to simultaneously capture both global semantic layouts and local detailed textures effectively (*Dong et al., 2025*).

Beyond the limitations in visual detection, existing inspection systems face a critical usability bottleneck in practical grain storage applications. Conventional machine vision systems are predominantly designed to output discrete quantitative data, such as the count and proportion of defective kernels (*Nie et al., 2022; Wang et al., 2023*). For non-expert operators, these statistical metrics lack intuitive meaning. Consequently, there remains a significant gap between obtaining quantitative classification results and making qualitative, actionable decisions (*Kumar & Kumar, 2024*). Recently, Large Language Models (LLMs) have demonstrated remarkable capabilities in data interpretation and context-aware reasoning through prompt engineering (*Son & Lee, 2025*), offering a promising avenue to bridge this gap and provide intelligent decision support.

To address these limitations, this study presents a multi-stage detection and decision-support system. First, YOLO v8 is integrated with a watershed algorithm for kernel singulation, and the Hungarian algorithm is utilized for front-to-back image matching to address single-view information loss. Next, a dual-stream network, CALFNet, is constructed by combining EfficientNet-B0 and MobileNet V2. Through a cross-modal attention mechanism, CALFNet fuses global morphological features with local defect details for classification. Finally, the DeepSeek V3 large language model is incorporated to translate quantitative detection statistics into qualitative evaluations and processing recommendations, aiming to provide direct decision support for grain storage management.

MATERIALS AND METHODS

Detection Device Design

To achieve standardized image acquisition, a dual-view detection device was designed, comprising a darkroom, dual light sources, and a transparent acrylic tray (300×200 mm), as shown in Fig. 1. Two high-resolution cameras (Model KS50M-778 with IMX766 sensor, 8160×6120 pixels) are positioned symmetrically above and below the tray. To prevent light interference, a sequential lighting strategy is employed, activating the upper and lower lights alternately to capture dorsal and ventral images respectively. Specific parameters are listed in Table 1.

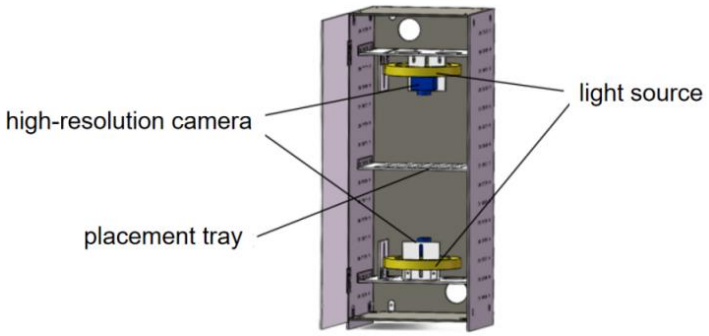


Fig. 1 - Detection device

Table 1

Camera specifications

Parameter Type	Parameters
Horizontal Field of View θ_H	72°
Vertical Field of View θ_V	55°
Max Image Resolution	8160 x 6120 pixels
Shutter Mode	rolling shutter
Pixel Size p_s	0.001 mm
Focal Length f	4.2 mm

The ideal Working Distance (WD) of the camera must fall within an interval defined by two core boundary conditions: the upper limit determined by imaging resolution requirements and the lower limit determined by Field of View (FOV) coverage requirements (*Bugatti & Colosimo, 2022*). The standard input image size for image classification models like EfficientNet-B0 is 224 × 224 pixels. According to the literature, the average length of common maize kernels is approximately 11.1 mm, the width is 8.1 mm, and the thickness is 4.5 mm (*Ji et al., 2024*). To implement dual-sided detection, horizontally splicing the front and back images is a common practice; thus, it is considered that the long edge should occupy at least 224 pixels in the acquired image. Object-space Resolution refers to the actual physical dimensions corresponding to a single pixel on the object plane. A greater working distance results in a larger object-space resolution value (i.e., lower precision). Based on this, the relationship between object-space resolution, the minimum pixel coverage, and the target object size can be constructed as:

$$R_{obj} = \frac{S_o}{N_p} = \frac{11.1\ mm}{224\ pixels} \approx 0.04955\ mm/pixel \tag{1}$$

Subsequently, based on the ideal Pinhole Camera Model and the principle of similar triangles, introducing the pixel size (ps) and focal length f, the relationship with the maximum working distance (WDmax) is established as:

$$WD_{max} = \frac{\mathrm{R_{obj}}}{\mathrm{ps}} \times \mathrm{f} \approx 208.1\ \mathrm{mm} \tag{2}$$

The camera must be installed at a sufficient distance to ensure its field of view completely covers the entire target working area. The minimum working distance is determined jointly by the field of view and the field of view angle. The longitudinal minimum working distance WDV min and the transverse minimum working distance WDH min can be calculated respectively, and the larger of the two is taken as the minimum working distance:

$$WD_{H\ min} = \frac{\mathrm{FOV_H}/2}{\tan(\theta_\mathrm{H}/2)} \approx 206.5\ \mathrm{mm} \tag{3}$$

$$WD_{V\ min} = \frac{FOV_V/2}{\tan(\theta_V/2)} \approx 192.1\ \mathrm{mm} \tag{4}$$

$$WD = max(WD_{H\ min}, WD_{V\ min}) = 206.5\ \mathrm{mm} \tag{5}$$

Therefore, the camera installation height should be $206.5\ mm \leq WD \leq 208.1\ mm$.

The camera was installed according to the theoretical distance, and a calibration board was used for testing, as shown in Fig. 2. The results indicate that the image field covers the required range with minimal distortion.

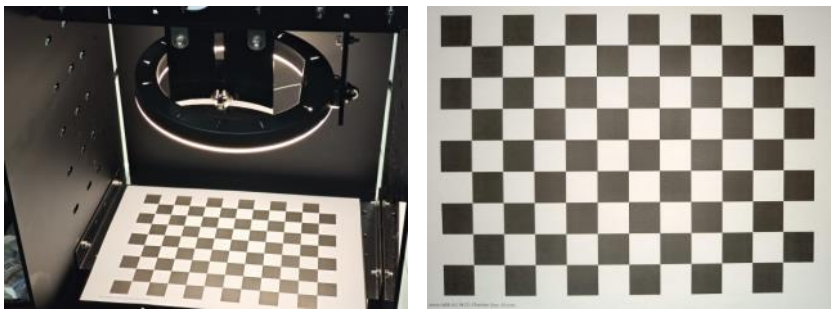


Fig. 2 - Image acquisition performance

Singulation and Matching

To achieve precise singulation, a "Coarse-to-Fine" strategy is adopted. First, a YOLO v8m model, trained on 1,631 images containing 66,957 labeled targets, is employed to locate kernels and generate Regions of Interest (ROI) (Fig.4a). The model achieved a mean Average Precision (mAP) of 99.0% after 100 epochs (Fig.3). Subsequently, a Marker-Controlled Watershed algorithm is applied within each ROI. High-confidence foregrounds are identified via Distance Transform on HSV-extracted masks (Fig. 4b), serving as markers to guide the watershed algorithm in separating adhered boundaries (Fig.4c). The central connected component is then selected and uprightly cropped via affine transformation (Fig.4d).

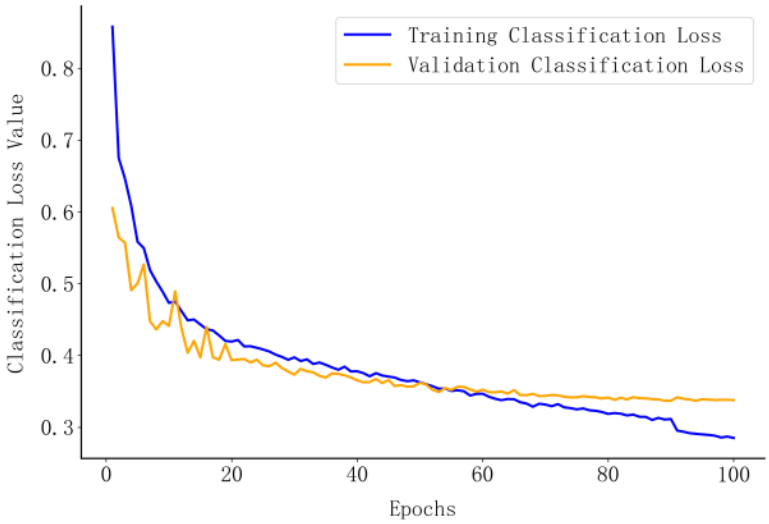
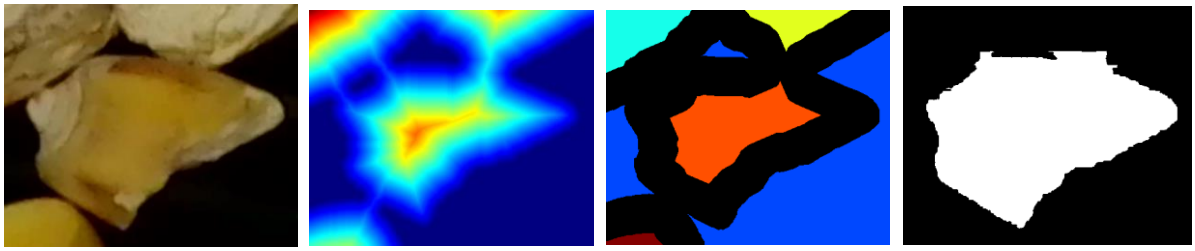


Fig. 3 - Training process

Matching front and back images are formulated as a minimum weight perfect matching problem in a weighted bipartite graph. Vertices represent kernel instances from both views, and edge weights are defined by the Euclidean Distance between their geometric centres, where smaller distances indicate higher correspondence probability (*Zhong et al., 2014*). The Hungarian Algorithm is employed to optimize this matching, establishing a global one-to-one correspondence as shown in Fig. 5.



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

Fig. 4 (a) Original image of adherent kernels; (b) Distance transform; (c) Generated segmentation boundaries; (d) Target kernel masks

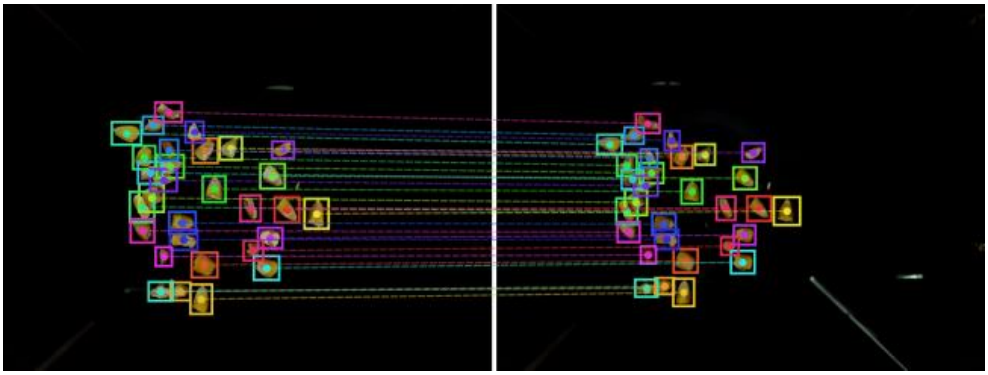


Fig. 5 - Matching performance post-alignment

Classification Model Design and Dataset Construction

Model Design

To achieve maize kernel classification based on dual-sided images, this paper proposes the CALFNet (Cross Attention and Lightweight Fusion Network) model. As shown in Fig. 6, the overall structure comprises three key components: a feature extraction network, a feature fusion network, and a classifier.

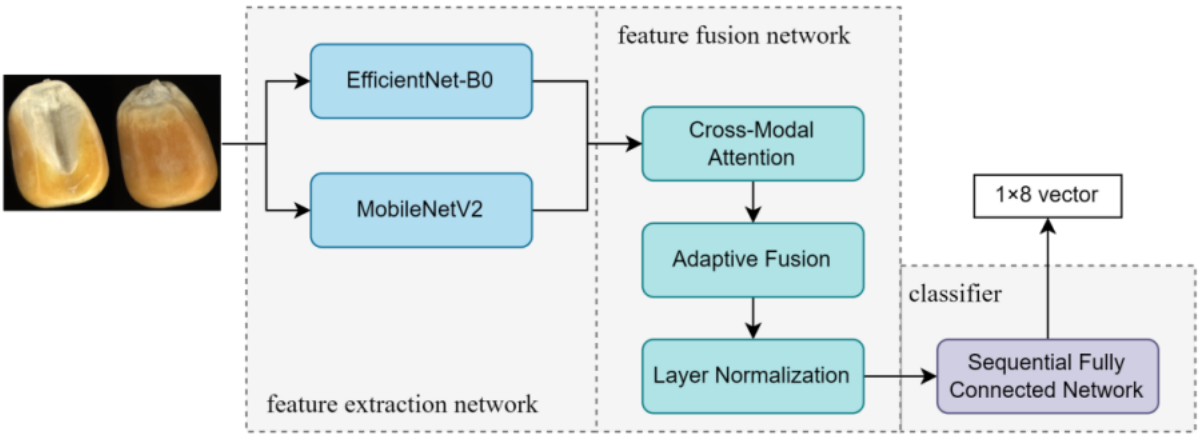


Fig. 6 - The architecture of the CALFNet model

Maize imperfect kernel recognition is a task that requires simultaneous consideration of local details, such as edges and textures, and global semantics, such as overall layout and colour distribution. EfficientNet-B0 and MobileNet V2 possess distinct design philosophies and technical implementations, enabling them to provide complementary feature representations for the model across different dimensions (*Ochoa-Ornelas et al., 2024*). To integrate the feature extraction results from EfficientNet-B0 and MobileNet V2, this paper designed a feature fusion network. This network consists of three sub-modules: Cross-Modal Attention, Adaptive Fusion, and Normalization, which achieve feature enhancement, dynamic combination, and numerical stability, respectively.

The Cross-Modal Attention module draws on the multi-head attention concept from Transformers (*Vaswani et al., 2017*). Given two input features provided by EfficientNet-B0 and MobileNet V2, respectively, the module generates Queries (Q), Keys (K), and Values (V) through fully connected layers. It then calculates a weighted output based on dot-product attention to enhance the interactive expression capability between features. Its mathematical model can be expressed as:

$$\begin{cases} Q = W_q X \\ K = W_k Y \\ V = W_v Y \end{cases} \quad W_q, W_k, W_v \in \mathbb{R}^{1280\times1280} \tag{6}$$

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{7}$$

The Adaptive Fusion module dynamically adjusts the weights of the two enhanced features through a gating network. As shown in Fig. 7, the gating network consists of two fully connected layers and a GELU activation function, outputting a weight vector to achieve weighted fusion. Finally, the fused features are processed through Layer Normalization to improve training stability.

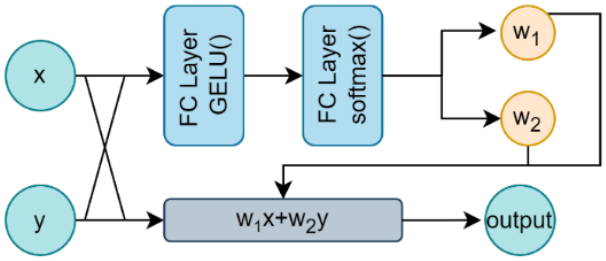


Fig. 7 - Gated network architecture

The classifier in this paper adopts a sequential fully connected network, with the specific structure shown in Fig. 8. It consists of two linear transformation layers and regularization modules. The first layer maps the input feature dimension from 1280 to a 512-dimensional latent space, followed by Batch Normalization and the Gaussian Error Linear Unit (GELU) to achieve feature standardization and non-saturating activation. To suppress overfitting in the high-dimensional latent space, Dropout is introduced. The second layer further projects the latent space features into an 8-dimensional class probability space to complete the classification.

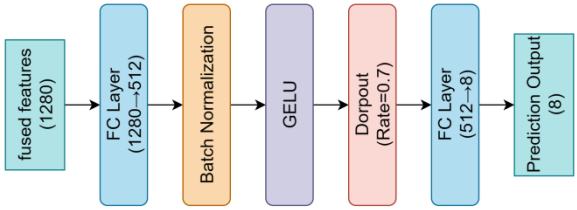


Fig. 8 - Classifier architecture

Dataset Construction and Augmentation

This paper utilizes the "GrainSet" open-source dataset, which contains a relatively rich set of samples (*Fan et al., 2023*). The classification process for the grain samples strictly follows the ISO 5527 cereals vocabulary standard, with labelling completed through manual classification. The dataset includes a training set and a test set, totalling 19,000 samples, with 17,100 samples in the training set and 1,900 samples in the test set (10% for each category). The dataset covers 8 categories corresponding to different types of maize kernels, including Normal (NOR), Fusarium (F&S), Sprouted (SD), Mouldy (MY), Broken (BN), Attacked by pests (AP), Heat Damaged (HD), and Impurities (IM). The specific distribution is shown in Fig. 9.

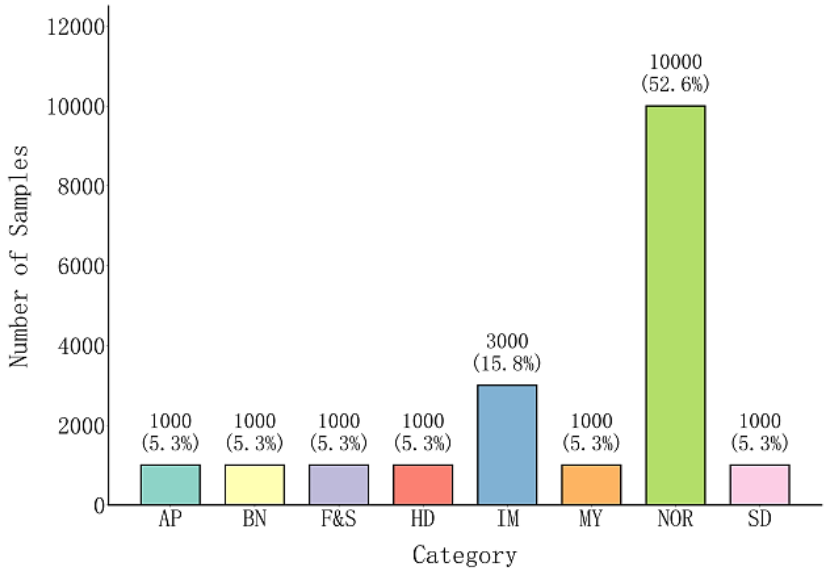


Fig. 9 - Distribution of sample categories

To improve the model's generalization performance, a systematic data augmentation strategy was adopted. All input images were first normalized to 224 × 224 pixels. To reduce the risk of model failure caused by varying kernel poses, random flipping and random rotation were applied with a probability of 50%. To simulate lighting and colour variations, random adjustments were made to brightness, contrast, saturation, and hue. Additionally, grayscale conversion was combined with a probability of 10% to eliminate colour interference and encourage the model to focus appropriately on learning structural features. These data augmentation strategies help avoid interference from minor image features, enabling the model to better learn more universal and important features, thereby improving its generalization performance.

Classification Model Training

The model training and testing environment was an Ubuntu 20.04.6 LTS system platform equipped with an 8-core Intel Xeon Platinum 8255C processor (2.50 GHz), 32 GB of RAM, and 50 GB of disk space. The GPU used was an NVIDIA Tesla T4 (16 GB video memory), with driver version 525.105.17 and CUDA version 11.7.99. The training employed a Batch Size of 64 samples per batch for a total of 50 epochs.

During the training process, the Train Loss, Train Accuracy, Test Loss, and Test Accuracy for all training epochs were calculated and saved. The training process is shown in Fig. 10. In the initial stage, the model converged rapidly; by the 8th epoch, the training loss significantly decreased to 0.0708, the training set classification accuracy increased to 97.93%, and the test set classification accuracy reached 98.00%. The loss and accuracy stabilized at high levels, with classification performance fluctuating slightly but trending towards further improvement overall. By 50 epochs, the model had fully converged, and the performance gap between the training and test sets was small, indicating no need for further training.

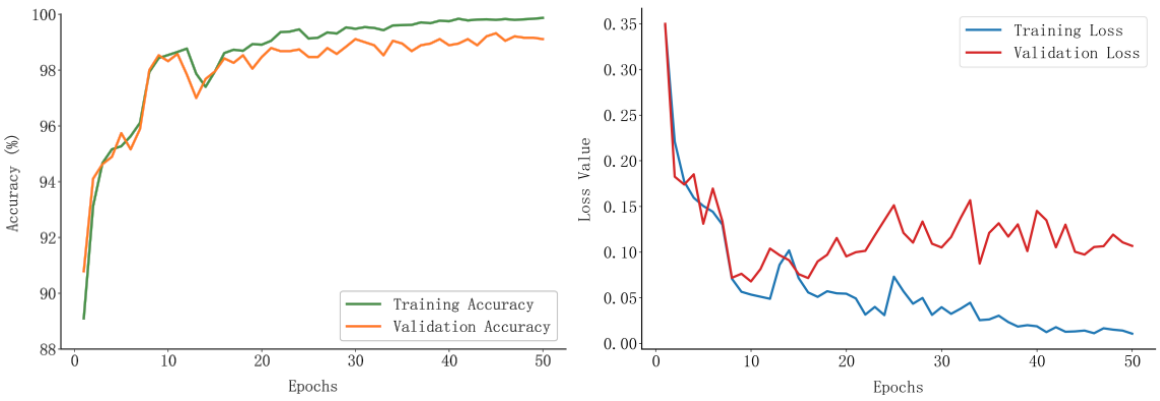


Fig. 10 - Training process

Generation of Quality Evaluation Based on Large Language Model

To bridge the gap between quantitative classification data and qualitative decision-making for non-experts, this study introduces a Large Language Model (LLM) to generate actionable evaluation reports. As illustrated in Fig. 11, the core methodology relies on Prompt Engineering strategies to transform visual detection results into context-aware information (*Son & Lee, 2025*). The process involves aggregating statistical data into a structured summary, which is then injected into a prompt template.

This template utilizes Role-Playing (e.g., "senior agricultural expert") and Task Constraints to guide the DeepSeek V3 model in generating a standardized report comprising "Overall Evaluation, Problem Analysis, and Processing Recommendations" via an API interface. A moderate temperature coefficient is applied to balance output certainty and diversity.

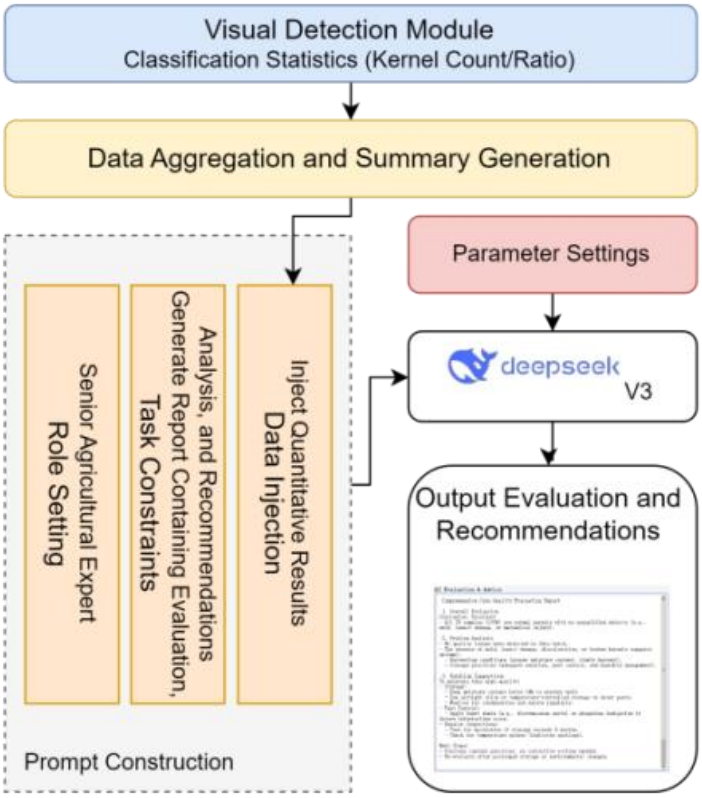
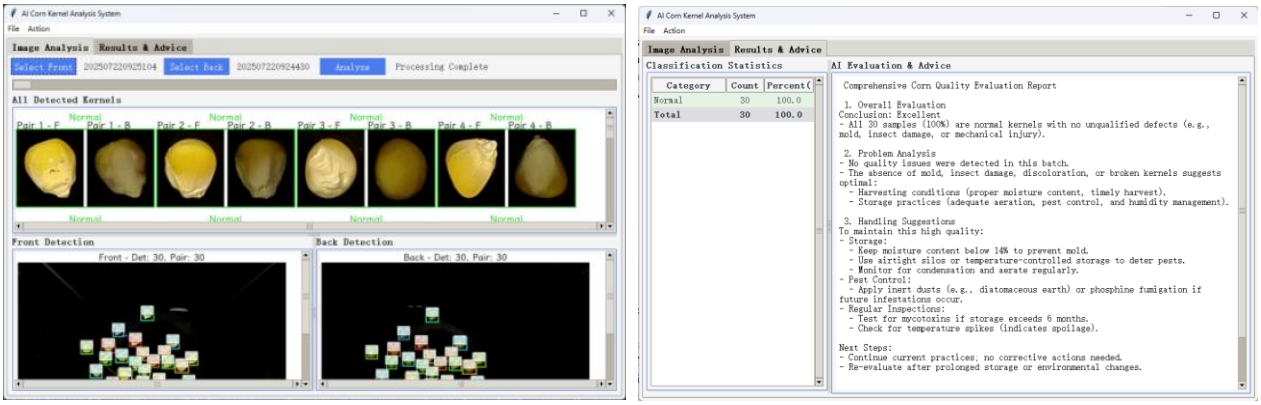


Fig. 11 - Diagram of the evaluation generation process

System Integration

To integrate and validate the collaborative capability of the proposed multi-stage algorithms and to visually demonstrate the end-to-end detection process and performance in a systematic manner, this paper designed an imperfect kernel detection system.

The system integrates the proposed modules: standardized image acquisition, kernel singulation and matching, and dual-stream feature fusion classification. A Graphical User Interface (GUI) was designed, allowing users to upload maize kernel images collected from the standardized image acquisition device, as shown in Fig. 12. The system automatically completes image segmentation extraction and kernel sample image classification, as shown in Fig. 12(a), and displays the evaluation generated by the Large Language Model (LLM), as shown in Fig. 12(b).



(a) (b)

Fig. 12 - Detection system interface: (a) Image analysis interface; (b) Results and recommendations

RESULTS

Classification Model Performance Verification

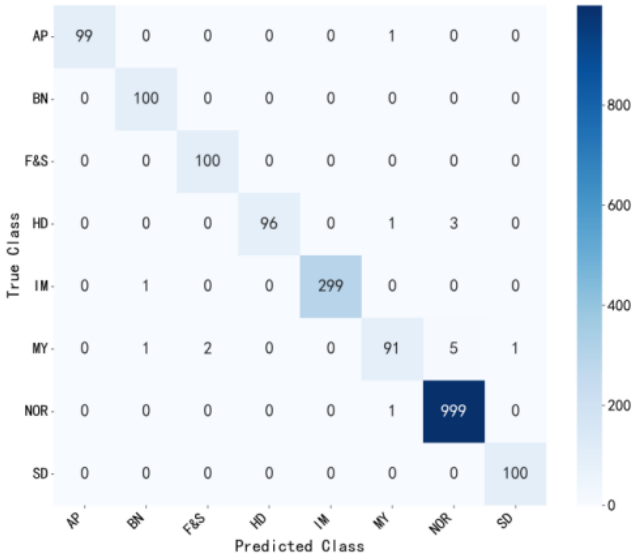


Fig. 13 - Confusion matrix

The classification performance of the CALFNet model on specific categories was evaluated using the test set. The confusion matrix is shown in Fig. 13. Precision, Recall, and F1-score were adopted as evaluation metrics, with results presented in Tab. 2. The model achieved over 98% accuracy in all 8 categories, with an average Precision of 99%, an average Recall of 98.2%, and an average F1-score of 98.6%, indicating high stability and robustness in the overall classification task.

Table 2

Classification performance evaluation

Evaluation Metric	AP	BN	F&S	HD	IM	MY	NOR	SD	Average
Precision	100%	98%	98%	100%	100%	97.9%	99.1%	99%	99%
Recall	99%	100%	100%	96%	99.7%	91%	99.9%	100%	98.2%
F1-score	99.5%	99%	99%	98%	99.8%	94.3%	99.5%	99.5%	98.6%

Table 3

Model comparison results

Model	Accuracy	Avg Precision	Avg F1-score	Parameters
This Model	99.16%	99.01%	98.58%	23.29 M
AlexNet	93.95%	93.32%	90.46%	57.04M
ResNet50	96.68%	95.11%	93.87%	23.52M
VGG16	96.58%	96.04%	93.56%	134.29M

To further verify the performance of the CALFNet model, this paper conducted comparative experiments with three benchmark models (AlexNet, ResNet50, and VGG16). The results are shown in Tab. 3. CALFNet achieved the highest accuracy, average Precision, and average F1-score, significantly outperforming the other models. Compared to AlexNet, CALFNet improved accuracy by approximately 5.2% and the F1-score by 8.1%, demonstrating stronger classification capability. Compared to ResNet50 and VGG16, CALFNet improved accuracy by 2.5% and 2.6% respectively, while maintaining a higher F1-score. In terms of parameter size, CALFNet is 23.29M, which is close to ResNet50 but far lower than VGG16 and AlexNet. This indicates that CALFNet maintains moderate model complexity while delivering high performance, balancing computational efficiency and detection accuracy.

Feature Extraction Module Visualization

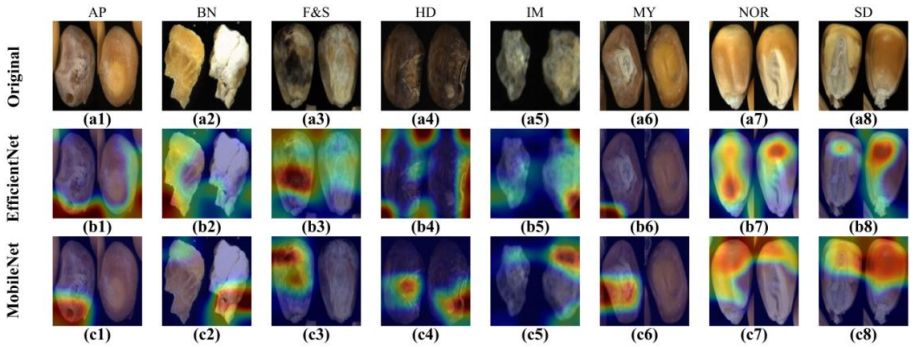


Fig. 14 - Visualization of Grad-CAM for different seed categories, where (a1)-(a8) are the original images. (b1)-(b8) and (c1)-(c8) are the activation heatmaps from the EfficientNet and MobileNet branches, respectively

To verify the complementarity of the dual-stream network, this study employed Grad-CAM to visualize the specific regions of interest for each feature extraction module (Selvaraju et al., 2020). As shown in Fig. 14, the generated heatmaps reveal distinct attentional focuses: EfficientNet-B0 generally targets global morphological features and edges, whereas MobileNet V2 demonstrates higher sensitivity to local details and "hard damage" features, such as fractures in Broken (BN) kernels or colour variations in Sprouted (SD) kernels. These differences confirm that the two backbones effectively capture complementary feature representations for complex imperfect kernel detection.

Ablation Study

An ablation study was conducted to evaluate the performance of the CALFNet model and its key components in the maize imperfect kernel detection task. The results are shown in Tab. 4. The proposed model achieved the highest Accuracy, Average Precision, and Average F1-score, significantly outperforming variants using only EfficientNet-B0, MobileNet V2, or a simple attention mechanism.

Table 4

Results of the ablation study

Model	Accuracy	Avg Precision	Avg F1-score
This Model	99.16%	99.01%	98.58%
EfficientNet-B0	98.58%	97.77%	97.42%
MobileNetV2	98.58%	97.46%	97.24%
Using Simple Attention	97.00%	95.95%	95.00%

The complete model improved Accuracy and F1-score by approximately 0.6% and 1.2% respectively compared to EfficientNet-B0 and MobileNet V2, indicating that the fusion module and cross-attention mechanism significantly enhanced classification performance. Compared to the variant using a simple attention mechanism, the complete model improved classification accuracy by approximately 2.2% and the F1-score by approximately 3.6%, validating the contribution of the complex attention mechanism and adaptive fusion design to performance.

Integrated Test

To simulate normal operating conditions and further verify the reliability of the integrated system, a total of 2.5 kg of maize kernels from the Heilongjiang production area were selected. Imperfect kernels were manually screened from this batch, yielding 206 imperfect kernels covering five categories: Normal (NOR), Broken (BN), Heat Damaged (HD), Sprouted (SD), and Attacked by pests (AP). Additionally, 300 normal kernel samples (approximately 1.5 times the number of imperfect kernels) were taken, totalling 506 samples. The imperfect kernel detection system constructed in this paper was used for detection. The normalized confusion matrix is shown in Fig. 15. For the three major categories—Normal (NOR), Broken (BN), and Heat Damaged (HD)—which were more numerous in the sample, the recognition accuracy reached 96.7%, 96.6%, and 85.2%, respectively, with an overall accuracy of 96.05%.

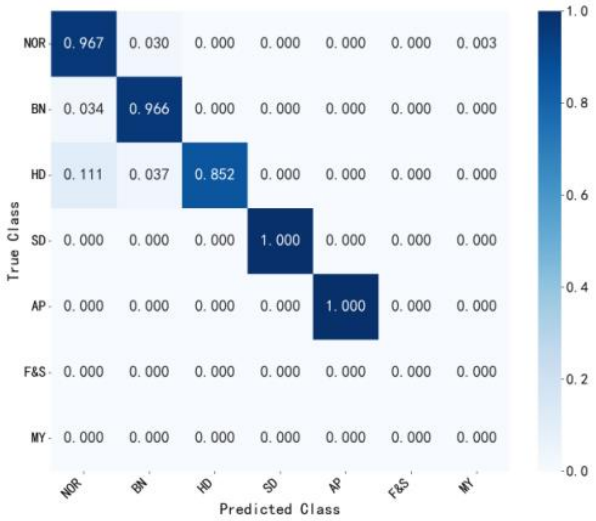


Fig. 15 - Confusion matrix

CONCLUSIONS

This study addresses the critical challenges in maize quality inspection, particularly the limitations of fine-grained segmentation and incomplete single-view information. By integrating YOLO v8 with the watershed algorithm, the proposed method achieves robust kernel singulation in complex backgrounds. Furthermore, the implementation of the Hungarian algorithm effectively facilitates front-to-back image matching, while the CALFNet dual-stream feature fusion network provides comprehensive recognition of various imperfect kernels, including mildewed, broken, and insect-damaged samples.

To validate the practical feasibility of these algorithms, an end-to-end intelligent detection system was developed and integrated with a GUI. A key innovation of this system is the incorporation of the DeepSeek V3 large language model, which transforms quantitative classification data into intuitive, expert-level qualitative evaluations and handling recommendations. Experimental results demonstrate that the CALFNet model achieves a classification accuracy of 99.16% on the test set, significantly outperforming benchmark models like AlexNet and ResNet50. Integrated tests on 506 real-world samples further confirm the system's robustness, reaching an overall accuracy of 96.05% in practical scenarios.

Despite these advancements, certain limitations remain to be addressed in future research. The current system’s processing speed could be further optimized to meet the high-throughput requirements of large-scale industrial grain pipelines. Additionally, while the DeepSeek V3 integration provides valuable insights, future work will focus on fine-tuning the model with more diverse agricultural datasets and exploring the deployment of the entire pipeline on edge computing devices to enhance real-time performance in remote storage facilities.

ACKNOWLEDGEMENT

This research was supported by the National Key Research and Development Program of China (Grant No. 2021YFD2100601), and the Open Fund Project of the State Key Laboratory of Agricultural Equipment Technology (South China Agricultural University) (No. NKLAET-202402).

REFERENCES

[1] Bugatti, M., & Colosimo, B. M. (2022). The intelligent recoater: A new solution for in-situ monitoring of geometric and surface defects in powder bed fusion. *Additive Manufacturing Letters*, *3*, 100048. https://doi.org/10.1016/j.addlet.2022.100048

[2] Chen, S., Zhu, H., Wang, J., Yu, T., Wang, Z., & Liu, C. (2023). Abnormal soybean grains recognition based on Opt-MobileNetV3. *Transactions of the Chinese Society for Agricultural Machinery*, *54*(S2), 359-365. https://doi.org/10.6041/j.issn.1000-1298.2023.S2.042

[3] Dong, B., Wang, Z., Chen, C., Wang, K., & Zhang, J. (2025). An improved backbone fusion neural network for orchard extraction. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18, 17961 – 17974. https://doi.org/10.1109/JSTARS.2025.3586322

[4] Erenstein, O., Jaleta, M., Sonder, K., Mottaleb, K., & Prasanna, B. M. (2022). Global maize production, consumption and trade: Trends and R&D implications. *Food Security*, *14*(5), 1295 – 1319. https://doi.org/10.1007/s12571-022-01288-7

[5] Fan, L., Ding, Y., Fan, D., Wu, Y., Chu, H., Pagnucco, M., & Song, Y. (2023). An annotated grain kernel image database for visual quality inspection. *Scientific Data*, *10*(1), 778. https://doi.org/10.1038/s41597-023-02660-8

[6] Guo, J., Nguyen, H.-T., Liu, C., & Cheah, C. C. (2023). Convolutional neural network-based robot control for an eye-in-hand camera. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *53*(8), 4764 – 4775. https://doi.org/10.1109/tsmc.2023.3257416

[7] Kumar, S., & Kumar, M. (2024). Enhancing agricultural decision-making through an explainable AI-based crop recommendation system. 2024 International Conference on Signal Processing and Advance Research in Computing (SPARC), 1 – 6. https://doi.org/10.1109/SPARC61891.2024.10829064

[8] Ji, J., Jin, T., Li, Q., Wu, Y., & Wang, X. (2024). Construction of maize threshing model by DEM simulation. *Agriculture*, *14*(4), 587. https://doi.org/10.3390/agriculture14040587

[9] Liu, M., Liu, Y., Wang, Q., He, Q., & Geng, D. (2024). Real-time detection technology of corn kernel breakage and mildew based on improved YOLOv5s. *Agriculture*, *14*(5), 725. https://doi.org/10.3390/agriculture14050725

[10] 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

[11] Nie, S., Ma, S., Peng, Y., Wang, W., & Li, Y. (2022). Research progress of rapid optical detection technology and equipment for grain quality. *Transactions of the Chinese Society for Agricultural Machinery*, *53*(11), 1 – 12. https://doi.org/10.6041/j.issn.1000-1298.2022.11.001

[12] Ochoa-Ornelas, R., Gudiño-Ochoa, A., & García-Rodríguez, J. A. (2024). A hybrid deep learning and machine learning approach with Mobile-EfficientNet and grey wolf optimizer for lung and colon cancer histopathology classification. *Cancers*, *16*(22), 3791. https://doi.org/10.3390/cancers16223791

[13] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, *128*(2), 336 – 359. https://doi.org/10.1007/s11263-019-01228-7

[14] Son, M., & Lee, S. (2025). Advancing multimodal large language models: Optimizing prompt engineering strategies for enhanced performance. *Applied Sciences*, *15*(7), 3992. https://doi.org/10.3390/app15073992

[15] Telçeken, M., Akgun, D., & Kacar, S. (2024). An evaluation of image slicing and YOLO architectures for object detection in UAV images. *Applied Sciences*, *14*(23), 11293. https://doi.org/10.3390/app142311293

[16] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *In Advances in Neural Information Processing Systems* (pp. 5998 – 6008).

[17] Wang, H., Zhu, Y., Li, Z., & Zhen, T. (2023). Research progress of machine vision in crop seed inspection. *Computer Engineering and Applications*, *59*(22), 69 – 83. https://doi.org/10.3778/j.issn.1002-8331.2303-0166

[18] Wang, L., Liu, J., Zhou, Y., Zhang, J., Li, X., & Fan, X. (2021). Corn seed quality detection based on watershed algorithm and convolutional neural network. *Journal of Chinese Agricultural Mechanization*, *42*(12), 168 – 174. https://doi.org/10.13733/j.jcam.issn.2095-5553.2021.12.25

[19] Yao, X., Wang, X., Wang, S.-H., & Zhang, Y.-D. (2022). A comprehensive survey on convolutional neural network in medical image analysis. *Multimedia Tools and Applications*, *81*(29), 41361 – 41405. https://doi.org/10.1007/s11042-020-09634-7

[20] Yao, Y., Cui, C., Geng, D., & Zhao, B. (2025). Broken maize kernel recognition method based on improved SqueezeNet network model. *Transactions of the Chinese Society of Agricultural Engineering*, *41*(9), 154 – 164. https://doi.org/10.11975/j.issn.1002-6819.202412248

[21] Zhang, W., Du, Y., Li, X., Liu, L., Wang, L., & Wu, Z. (2024). Online detection method of corn kernel quality based on FSLYOLO v8n. *Transactions of the Chinese Society for Agricultural Machinery*, *55*(8), 253 – 265. https://doi.org/10.6041/j.issn.1000-1298.2024.08.023

[22] Zhong, J., Tan, J., Li, Y., Gu, L., & Chen, G. (2014). Multi-targets tracking based on bipartite graph matching. *Cybernetics and Information Technologies*, *14*(5), 78 – 87. https://doi.org/10.2478/cait-2014-0045