

CQN-YOLO: A Lightweight and High-Precision Model for Corn Appearance Quality Detection

Azhi Zhang¹ and Chenjia Qian^{*,2}

¹College of Physics and Electronic Information Engineering, Zhejiang Normal University, Jinhua 321004, China

²Zhejiang Institute of Optoelectronics, Jinhua 321004, China

Abstract This study advances food quality assessment by utilizing a lightweight deep learning model for the rapid and high-precision detection of corn appearance defects, addressing critical needs in the automated characterization of agricultural products. The proposed CornQualityNet (CQN-YOLO) is based on enhanced YOLOv10n. Firstly, four augmentation methods are employed to enhance the dataset, addressing the issues of overfitting and data insufficiency during training, and thereby improving the model's robustness. Secondly, the Slimneck structure is designed to optimize the YOLOv10n neck network, which achieves a lightweight model, reduces computational complexity, and improves computational speed. Lastly, the Slide Loss is used to enhance the classification ability of the model, which accelerates the convergence speed of the model and reduces computing time. The experimental results demonstrate that the proposed CQN-YOLO model achieves an accuracy of 97.8%, a recall of 96.6%, and a Mean Average Precision (mAP) of 99.0%, which are 3.5%, 1.3%, and 1.0% higher than the YOLOv10n model, respectively, and the GFLOPs are reduced from 6.5 to 5.9. Compared to other models in the field, the CQN-YOLO model performs best, enhancing the model's generalization ability, detection efficiency, and robustness while maintaining high precision and achieving model lightweightness. The model can be effectively used in detection equipment with limited computational resources and provides technical support for the detection of the appearance quality of other agricultural products.

Keywords: Food quality assessment, automated agricultural product, deep learning, corn appearance quality, YOLOv10n

1 Introduction

The improvement of grain quality will bring multiple impacts. Firstly, consumers can enjoy healthier, more delicious, and safer food products that meet their high-quality and diverse needs. Secondly, for farmers, high-quality grain products have higher added value,

*Corresponding author: qcj666@zjnu.edu.cn

which is conducive to increasing income and stimulating their enthusiasm for growing crops. Starting from the industrial level, it can promote the transformation and upgrading of the agricultural industry chain, and advance the extension and integrated development of the industry chain. Corn holds the top spot in global food crop production with an absolute advantage. It has surpassed rice and wheat in terms of planting area, total yield, and yield level, becoming one of the most significant food crops in the world. Corn is also the most essential feed source and industrial raw material. Its quality is directly related to the quality and safety of grain. Different grades of corn are suitable for various purposes.

The Chinese national standard GB/T1353-2018 [1] specifies the quality evaluation indicators of different grades of corn, among which the appearance quality indicators of first-class corn require imperfect grain content not to exceed 4%, mildew grain content not to exceed 2%, whole grains, no damage, no moth, no rat bite. Poor storage of corn after harvest, including storage environments and technology restrictions that do not strictly adhere to standard storage and ventilation practices, makes it highly prone to mildew, insect infestations, damage, and other issues. According to the latest report by the U.S. GRAINS COUNCIL, titled "2024/2025 CORN HARVEST QUALITY REPORT" [2], in the grading of corn, attention is also paid to damaged grains and total damage. For first-grade corn, the requirements are that the grain weight must not be lower than 56 pounds, the total damage rate must not exceed 3%, and the damaged rate and impurity content must not exceed 2%. The traditional corn appearance quality inspection primarily relies on visual observation. This detection method has several drawbacks, including a cumbersome operational process, low detection efficiency, and subjectivity in inspector judgment, which makes it challenging to ensure the precision and consistency of the detection results. As the core index of seed quality evaluation, it is of great value to introduce advanced target detection technology into maize appearance quality evaluation.

In recent years, applying machine learning [3]. and deep learning technologies [4] to target detection has been a hot research direction, and many scholars have applied it to the detection and analysis of agricultural product appearance quality. To classify the quality of corn kernels, Wu et al. [5] introduced a new approach that utilizes image analysis combined with a support vector machine, achieving an accuracy of 97.44%. However, this approach is limited to scenarios with relatively few samples. Ma et al. [6] introduced an approach for detecting and classifying the appearance quality of peanut seeds using a support vector machine. The classification accuracy of batch peanut seeds reached 86%. However, the experimental effect of this method was not ideal. Guo et al. [7] achieved mango appearance quality detection using an FPGA-accelerated CNN with an accuracy of 96%, but the training and reasoning process of this method was time-consuming.

Based on the improved RetinaNet, Zhou et al. [8] conducted nondestructive testing on soybean appearance quality, with an accuracy of 97.39% and a mAP value of 98.64%. However, RetinaNet utilized a deeper feature extraction network and more prediction branches, which necessitated increased computing resources. To address the issue of limited feature extraction capability in corn disease classification, Wang et al. [9] proposed a multimodal classification model named CEA-Net, which combines images and text. By adopting CCM, EDA F, and AACEM pre-training strategies, this model achieved a classification accuracy of 97.40% on a self-built dataset containing 4633 pairs of images and texts.

YOLO (You Only Look Once) [10] series algorithms are widely used in crop monitoring, weed identification, and pest monitoring, among other applications, due to consider-

ations of precision, parameter number, calculation amount, and model size. At present, many scholars have used the YOLO series algorithm to make a series of studies on the quality detection of crops. Gu et al. [11] proposed a rice appearance quality recognition model based on improved YOLOv5s to meet the need for rapid recognition in rice appearance quality detection. The mAP of the improved network reached 96.5%, compared with the YOLOv5s network, which improved by 4.3%. Fan et al. [12] proposed a modified YOLO v4 with a four-channel target detection model to identify and classify the appearance quality of maize seeds. The average F1 score and mAP of the model's comprehensive evaluation indexes reached 93.09% and 98.02%.

Based on microcomputer vision and YOLOv5, Sun et al. [13] developed a rapid and non-destructive method for detecting rice mold, achieving a precision and recall of 100% and 95.3%, respectively. Mohanty et al. [14] developed a weed detection model based on YOLOv5. This model achieved an identification accuracy of 98% on the self-built rice field weed dataset. While maintaining high precision, it was successfully deployed on the embedded platform NVIDIA Jetson Nano, enabling real-time weed detection in the field and precise pesticide application. Niu et al. [15] proposed a lightweight improved model named E-YOLOv8. By incorporating the FasterNet backbone, the CARAFE upsampling module, the Detect_EMA detection head, and the Inner_SIoU loss function, this model achieved an mAP of 96.2% on a dataset of 10 corn varieties, representing a 4.4% improvement compared to the original YOLOv8 model. In response to the lightweight requirements for tomato leaf disease detection models, Sun et al. [16] proposed a lightweight detection model called LigTomDet based on knowledge distillation. This model achieved an mAP50 of 87.7% on the Tomato-Village dataset with a model size of only 4.3MB. Compared to the benchmark model YOLO11n, it reduced parameters, computational cost, and model size by 26.7%, 12.5%, and 21.8% respectively, while improving the detection accuracy by 2.5 percentage points.

YOLO series algorithms are providing new solutions for crop quality detection by improving the precision and efficiency of image recognition. Currently, research on crop appearance quality has some shortcomings, including unsatisfactory experimental results and a substantial amount of model calculations. Thus, it is of critical importance to develop a new method that reduces model parameters while ensuring high precision and efficiency. This paper takes corn with varying appearance quality as the research object and YOLOv10n as the benchmark model, proposing a corn appearance quality detection model (CornQualityNet, CQN-YOLO) that improves upon YOLOv10n. While ensuring detection precision, the model's parameters and complexity are reduced, thereby enhancing its generalization ability and system robustness.

The primary contributions of this study are as follows.

- Four methods, including horizontal flipping, vertical flipping, adding noise, and adjusting brightness, are employed to enhance the data of seven types of corn kernels, addressing the issues of overfitting and data insufficiency in the training process, and thereby improving the model's robustness.
- The Slimneck structure optimizes the YOLOv10n neck network, making the model lightweight, reducing computational complexity, and improving computational speed.
- The Slide Loss is used to enhance the model's classification performance, accelerating its convergence speed and reducing computing time.

2 Materials

We used the GrainSpace [17] dataset, which is publicly accessible online, for this study. GrainSpace is the first publicly released grain detection dataset proposed by Fan et al. [12], which includes three types of grain sample datasets: rice, wheat seed, and corn seed. According to the national standard GB1353-2018, corn samples are divided into seven categories, which are: normal, infest-damaged, spotted, broken, mouldy, heat-damaged, and sprouted, as shown in Fig.1, 6940 corn seed images of high quality were selected from the data set, and 27760 images were obtained by means of horizontal inversion, vertical inversion, noise addition and brightness change [18]. The number of original and enlarged corn seed images of different quality is shown in Table 1. The pictures of some corn seeds after data enhancement are shown in Fig.2.

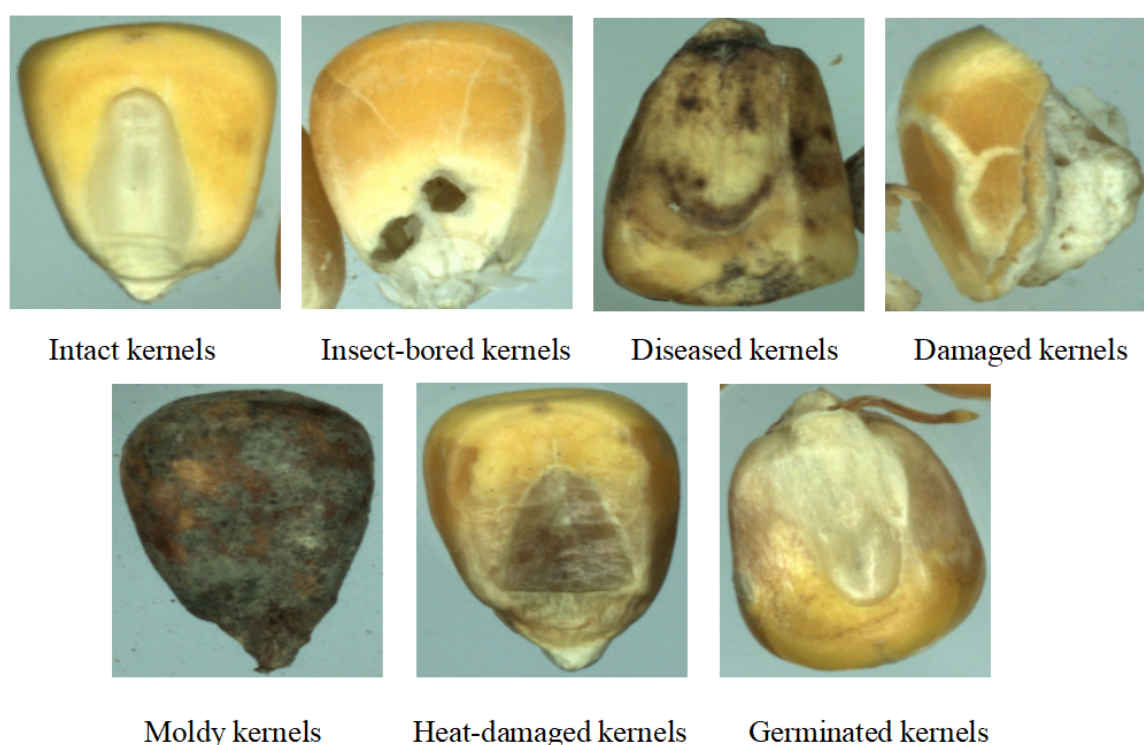


Figure 1: Corn seed with different quality.

In this paper, the LabelImg annotation tool was used to annotate each corn seed image, generate an XML-format label file, and then convert the generated XML file into a TXT-format file commonly used in the YOLO model. A randomized 7: 2: 1 split was applied to allocate the data to training, testing, and validation sets, respectively.

3 Corn appearance quality detection model based on CQN-YOLO

YOLOv10 is the latest proposed by the research team of Tsinghua University in 2024, which improves the performance and efficiency of model detection by eliminating Non-Maximum Suppression (NMS) processing and optimizing model design [19]. YOLOv10 has six models of different sizes: YOLOv10n, YOLOv10s, YOLOv10m, YOLOv10b, YOLOv10l, and YOLOv10x, and these model sizes are successively increased. Based

Table 1: Number of corn seeds of different quality.

Category	Original Quantity (kernels)	Augmented Quantity (kernels)
Intact kernels	2000	8000
Insect-bored kernels	1000	4000
Diseased kernels	1000	4000
Damaged kernels	1000	4000
Moldy kernels	1000	4000
Heat-damaged kernels	540	2160
Germinated kernels	500	2000
Total	6940	27760

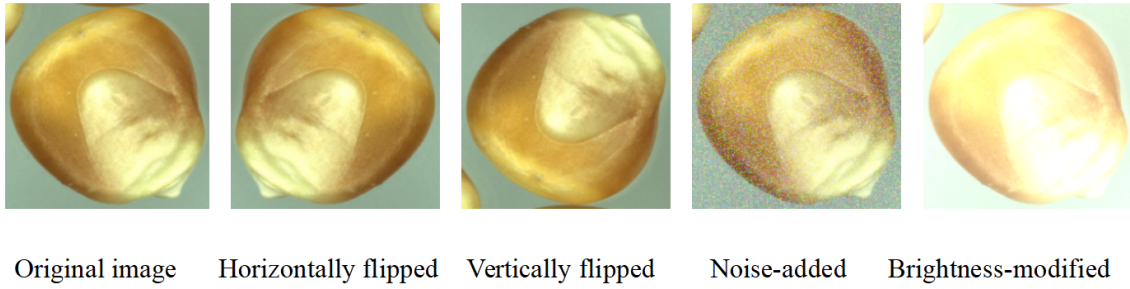


Figure 2: Corn seed image data enhancement.

on the consideration of detection speed and efficiency, YOLOv10n is selected as the basic model, and it is optimized and improved for detecting the appearance quality of corn.

In this paper, Slimneck, a lightweight feature fusion block, is introduced to optimize the network structure, replacing the Neck structure in YOLOv10n benchmark network with Slimneck, realizing the cross-layer fusion of high-level semantic features and low-level texture features, improving the network's ability to fuse corn appearance quality features, decreasing model complexity and improving the running speed, improve the detection precision. We utilized the Slide Loss function to replace the CIoU loss function in the original YOLOv10n network, aiming to mitigate the impact of the imbalance in 7 different quality corn seed samples on the modeling precision in this paper. Fig.3 shows the Structure of the optimized CQN-YOLO network.

3.1 Network Structure Optimization

The Neck part of YOLOv10n utilizes Depthwise Separable Convolution (DSC), which separates the channel information when processing the image, resulting in a lower feature representation ability compared to standard convolution (SC). In this section, we introduce three models: GSConv, GSbottleneck, and VoV-GSCSP, to build an efficient Slimneck. This can decrease the model's computational complexity and reasoning time. At the same time, high precision is guaranteed, allowing it to be applied to appearance quality inspection equipment with limited computing resources.

Fig.4 illustrates the network structure of GSConv. The input image is first convolved with conventional convolution (SC) through a convolutional layer (Conv), with the output

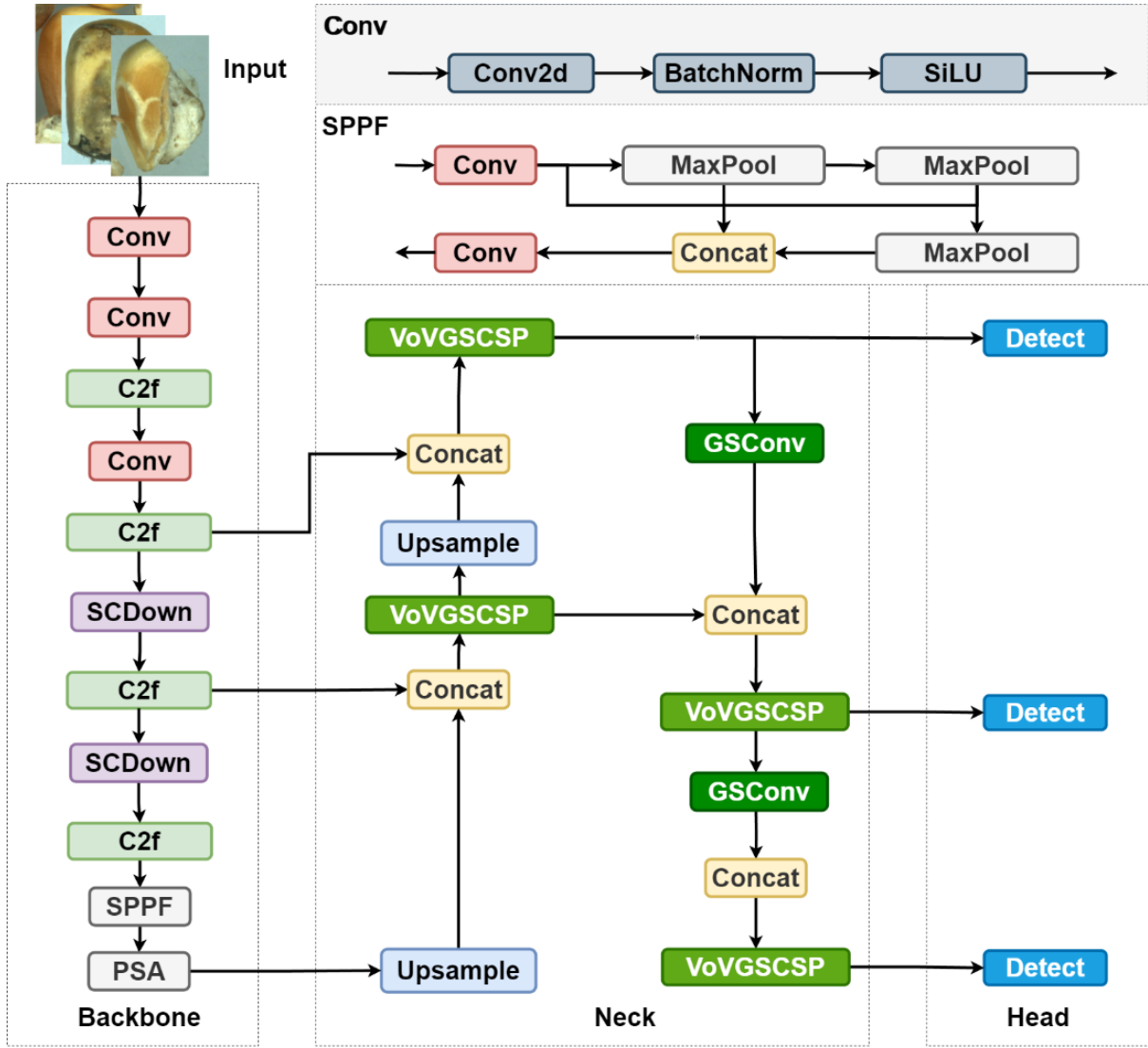


Figure 3: Network structure diagram of CQN-YOLO

channels set to C2. A feature tensor with C2/2 output channels is then obtained from this layer. Then, DWConv was used to perform depthwise separable convolution (DSC) operation. At the same time, independent convolution operations were applied to each channel of the input feature map, the image features were extracted, and the feature tensor with C2/2 channels was output. Then, the results of the Conv and DWConv layers were combined; that is, feature tensors output by SC and DSC were concatenated according to the number of channels. A Shuffle operation is applied to the concatenated feature map to redistribute the feature channels, thereby realizing information fusion between channels without increasing the computational burden. This method combines the advantages of DSC and SC, and its computational cost is approximately half of the standard convolution (SC), while maintaining a learning capability comparable to SC.

Specifically, GSConv uses the features generated by SC to permeate into every region of the features produced by DSC, allowing the features of SC to be fully merged into the output of the depthwise separable convolution by uniform mixing, thus retaining the latent relationships among features to the greatest extent, overcoming shortcomings of sparse convolution that completely cuts off these connections. It enhances the fea-

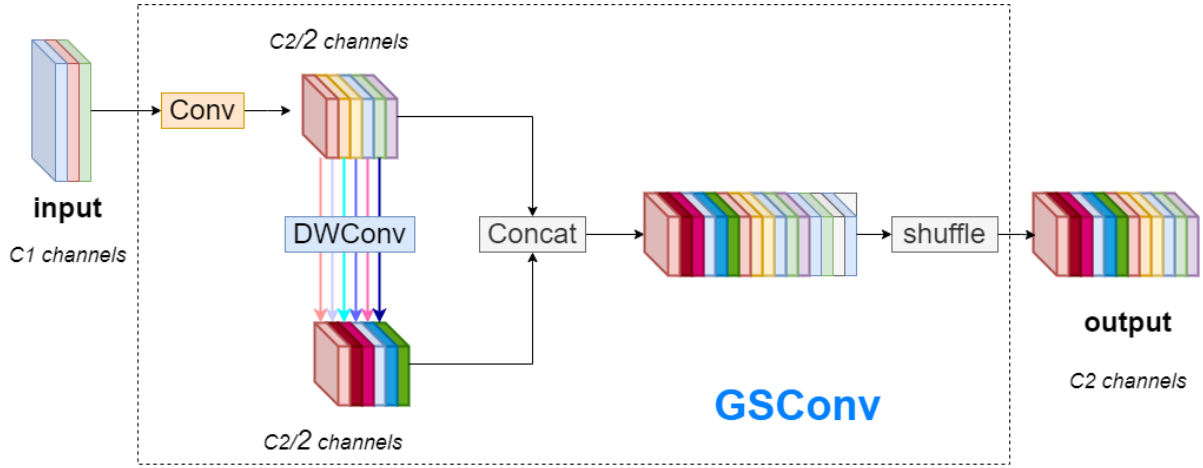


Figure 4: GSConv structure

ture representation ability while reducing computational cost and memory consumption, thereby improving model precision and processing speed [20].

In this paper, the lightweight convolutional GSConv is used to replace the classical Conv, leading to a significant reduction in computational cost and effectively reducing computational overhead. However, during the process of detecting corn appearance quality, the relevant process of compressing spatial dimensions (width and height) and expanding channels in the feature map tends to cause degradation of partial semantic details.

To achieve a better nonlinear representation of corn appearance quality characteristics and the reuse of information, we introduced the GSbottleneck module by extending the GSConv structure. As shown in Fig.5(a), this module consists of two GSConv layers, where the input data is stacked by GSConv modules to enhance the network's learning capacity and further improve its ability to process the appearance quality features of corn.

Building upon this, VoV-GSCSP, a new cross-level partial network module, is constructed by single-step aggregation. The specific process of aggregating GSbottleneck modules into VoV-GSCSP is shown in Fig.5(b). While reducing the computational complexity and reasoning time, the utilization efficiency of corn appearance quality features and network performance is further improved, and precision is maintained.

The Slimneck network, constructed based on GSConv, significantly increases the network's perception capability for the appearance quality of corn and improves its capacity for informative feature acquisition. This improves the model's ability to effectively capture the context information within the image, thereby improving the perceptual acuity of the main features and optimizing the model's capability in related tasks.

3.2 Loss function optimization

It can be seen from the second section that the dataset used in this study has a relatively small number of samples for germinated particles and heat-damaged particles, resulting in sample imbalance. The CIoU loss function adopted by the YOLOv10n benchmark model did not fully account for the key factor of sample imbalance in its design. It did not distinguish the difference in the number of samples in the calculation process. As a result, the model did not learn enough about the quality status characteristics of a

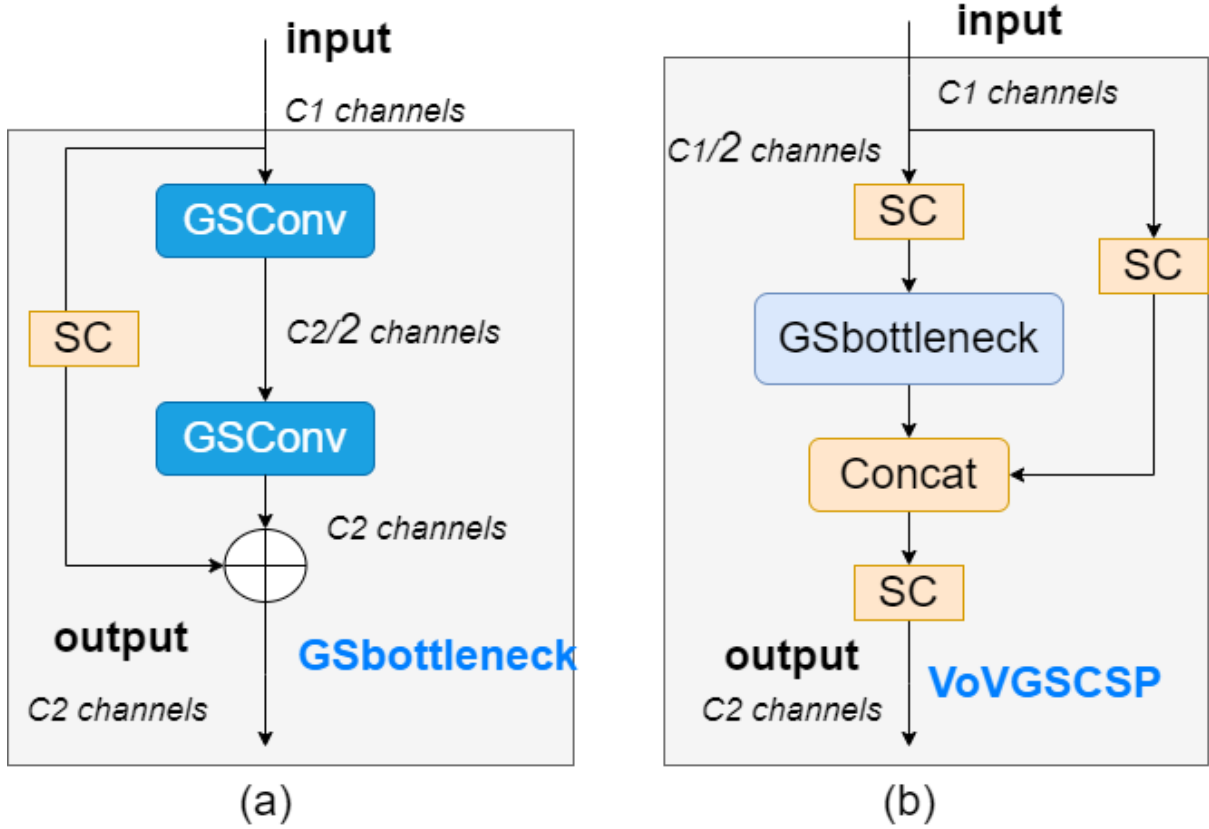


Figure 5: GSbottleneck and VoVGSCSP structure.

small number of samples in the training stage, thus affecting the overall detection effect. Results may be biased or inaccurate [21].

In this paper, the Slide Loss function [22] is introduced to address the issue of sample imbalance by adjusting the recognition effect of different sample categories adaptively. In order to simplify the hyperparameter, the Slide Loss function takes the mean value of IoU of all boundary boxes as the threshold μ ; values below the threshold are treated as negative samples. At the same time, those above it are regarded as positive samples. Its calculation is shown in equation (1).

$$f(x) = \begin{cases} 1 & x \leq \mu \\ e^{1-\mu} & \mu < x < \mu - 0.1 \\ e^{1-x} & x \geq \mu \end{cases} \quad (1)$$

The fundamental principle of the Slide Loss function involves dynamically adjusting the loss weight based on the IoU value, which is measured by calculating the intersection ratio (IoU) between the predicted result and the actual label. Specifically, when the expected value and the true value IoU are at a higher level, the corresponding loss weight will decrease. On the contrary, when the IoU value is low, the loss weight will increase correspondingly. The purpose of this strategy is to guide the model in paying more attention to those predicted values with a high degree of coincidence with the true values, thereby improving the precision of the detection. In other words, by adjusting the weight of positive and negative samples, the Slide Loss function focuses on those samples with a small number of samples, such as germination particles and heat loss particles, so

as to strengthen the model's learning and recognition ability of these samples through reasonable adjustment of weights, and effectively improve the overall classification precision.

4 Experimental results and analysis

4.1 Experimental environment and evaluation metrics

In this experiment, the SCG optimizer was used for training. The training was performed with a batch size of 32, the optimizer momentum was 0.9, with a learning rate of 0.001, and the number of iterations was 300. Details of the experimental hardware configuration are provided in Table 2.

Table 2: Experimental environment parameters.

Configuration Item	Specifications
Operating System	Windows 11 64-bit
CPU	Intel(R) Core(TM) i5-13400F CPU@2.50 GHz
GPU	NVIDIA GeForce RTX3060Ti
RAM	16G
Python	3.8
PyTorch	2.10.0

Precision (P), Recall (R), Mean Average Precision (mAP), and computation amount (GFLOPs) were used to measure the detection effect of CQN-YOLO on corn appearance quality. The specific calculation formula is shown in equations (2)-(5).

$$P = \frac{TP}{TP + TF} \times 100\% \quad (2)$$

$$P = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$$AP = \int_0^1 P(R) dR \quad (4)$$

$$mAP = \frac{\sum_{i=1}^n AP_i}{n} \quad (5)$$

Where TP represents true samples, FP denotes false positive samples, and FN indicates false negative samples. AP represents the area surrounded by the PR (Precision-Recall) curve and axis, with n denoting the total count of categories.

4.2 Comparison Experiment of YOLOv10n with Image Recognition Models

To verify the reliability of the YOLOV10n model in this paper, a comparative experiment was designed in this section, and its performance was compared with that of the

Table 3: Comparative Experiment of YOLOv10n with Image Recognition Models

Model	P%	Recall%	mAP%
midrule Resnet-50	67.6	64.3	71.1
MobileNetV3	72.8	72.1	79.2
5-ConvNet	78.6	77.7	83.0
YOLOv10n	94.3	95.3	98.0

image recognition model. ResNet-50, MobileNetV3, and a five-layer convolutional neural network (5-ConvNet) were selected as the comparison models. The test results under the same experimental conditions are shown in Table 3. As shown in Table 3, in terms of detection accuracy, the YOLOv10n model demonstrates a significant advantage over the other three recognition models. Even when compared with the best performing model among the three, 5-ConvNet, the accuracy rate of YOLOv10n increased by 15.7 percentage points (from 78.6% to 94.3%), the recall rate increased by 17.6 percentage points (from 77.7% to 95.3%), and the mAP increased by 15.0 percentage points (from 83.0% to 98.0%). This result fully demonstrates the effectiveness and reliability of YOLOv10n as the benchmark model in this study.

4.3 Analysis of the influence of network structure optimization on model performance

To validate the effect of network structure improvements in Section 3.1 of this paper, a lightweight Slimneck structure replaces the Neck part of the basic network of YOLOv10n, and a comparative experiment is carried out. Table 4 illustrates the results. In Table 4, "✓" means that the YOLOv10n Neck network is replaced with the Slimneck structure, and "×" means that the original neck is used.

Table 4: Comparison experiment of network structure optimization.

Quality status	Slimneck	P%	Recall%	mAP%
Intact kernels	×	96.8	99.6	99.2
	✓	97.5	98.7	99.4
Insect-bored kernels	×	98.4	94.4	98.9
	✓	97.3	93.8	98.9
Diseased kernels	×	93.9	94.6	97.7
	✓	95.7	93.5	98.4
Damaged kernels	×	98.8	99.2	99.5
	✓	98.9	98.5	99.5
Moldy kernels	×	86.7	91.5	96.2
	✓	90.4	91.3	96.9
Heat-damaged kernels	×	88.0	91.9	96.0
	✓	93.2	91.9	97.6
Germinated kernels	×	97.2	96.0	98.8
	✓	95.7	93.5	98.4

Based on the data in Table 4, the Slimneck structure optimization model maintains detection accuracy comparable to the original model across the vast majority of categories. Among them, the improvement effect of heat damage particles is the most obvious; the

accuracy rose from 88.0% to 93.2%, a growth of 5.2 percentage points, while the mAP increased by 1.6 percentage points, moving from 96.0% to 97.6%. The accuracy of perfect, macular, and mildew particles was increased by 0.7, 1.8, and 3.7 percentage points for each category, while their corresponding mAP values rose by 0.2, 0.7, and 0.7 points in turn. However, the detection effect of germinated grains was decreased, which may be owing to the insufficient sample quantity in this case. It can be seen that although the network structure optimization has some limitations, from an overall perspective, it is effective and provides an effective solution for the practical application of lightweight models.

4.4 Effect of loss function optimization on model performance

In order to ensure that the Slimneck structure can reduce the computational cost of the model without affecting the accuracy, we use it to conduct a comparison experiment with the Neck part of the YOLOv10n basic network, and the corresponding results are presented in Table 5. In Table 5, "✓" indicates that the model adopts the Slide Loss function, and "×" indicates that the original CIOU loss function is used.

Table 5: Comparative experiment of loss function optimization.

Quality status	Slide Loss	P%	Recall%	mAP%
Intact kernels	×	97.5	98.7	99.4
	✓	97.9	99.3	99.3
Insect-bored kernels	×	97.3	93.8	98.9
	✓	97.9	96.4	99.2
Diseased kernels	×	95.7	93.5	98.4
	✓	95.8	96.6	98.9
Damaged kernels	×	98.9	98.5	99.5
	✓	99.9	99.5	99.5
Moldy kernels	×	90.4	91.3	96.9
	✓	97.1	93.0	98.2
Heat-damaged kernels	×	93.2	91.9	97.6
	✓	96.3	94.0	98.6
Germinated kernels	×	95.7	93.5	98.4
	✓	100	97.2	99.4

The data in Table 5 reveal that, after the adoption of the Slide Loss function, the detection effect of the model on 7 types of corn with varying appearance qualities is improved to varying degrees. The accuracy, recall, and mAP of germinated grains increased by 4.3, 3.7, and 1 percentage points, respectively. On the whole, the accuracy remained above 95%, among which the improvement of mildew particles was the most obvious, from 90.4% to 97.1%, achieving a 6.7 percentage point enhancement; The recall remained above 93%; The mAP value is maintained above 98%. It can be seen that through the loss function optimization, we solve the impact of the imbalance of the number of samples.

Fig.6 shows the Precision-Recall curves of the appearance quality of different types of corn. As can be seen from Fig. 6, the discrepancy in area under the P-R curve and the axis under different quality conditions of corn is very small, indicating that the CQN-YOLO model has strong robustness.

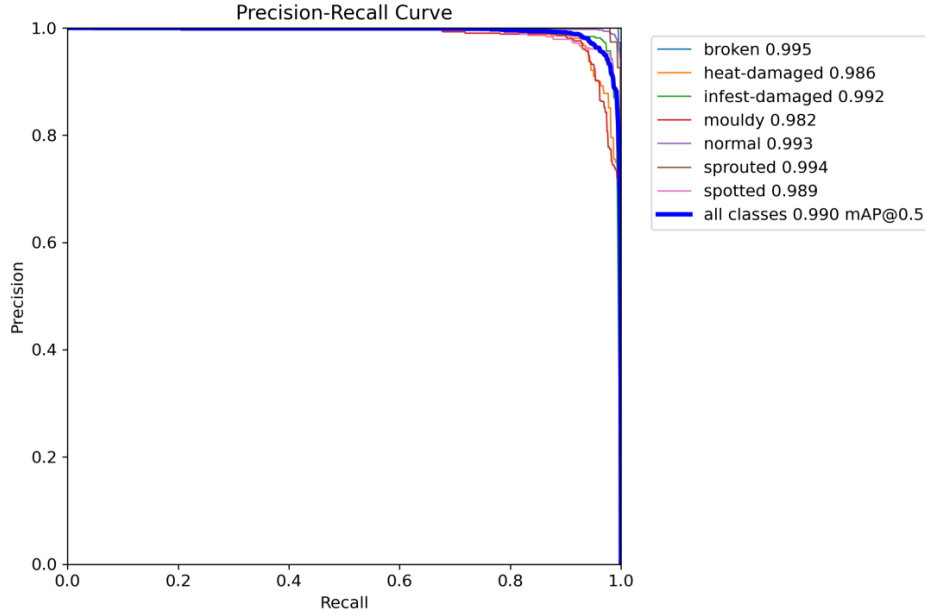


Figure 6: P-R curves of the appearance quality of different types of maize.

The Focal Loss function aims to address the issues of sample imbalance and uneven distribution of easy and difficult samples in classification tasks. On this basis, Vari Focal Loss and Quality Focal Loss were proposed as improved versions. GHM Loss, from the perspective of gradient distribution, effectively alleviates the problem of sample imbalance by adaptively weighting samples in different gradient intervals. Therefore, GHM Loss can be regarded as another typical method of sample reweighting. To verify the superiority of the model after the loss function optimization in this paper, we conducted a comparative experiment using these three loss functions. The experimental results are shown in Table 6.

Table 6: Comparative Experiment of Four Loss Functions

Model	P%	Recall%	mAP%
Focal Loss	93.8	91.4	97.4
Vari Focal Loss	96.4	93.9	98.1
Quality Focal Loss	94.1	94.6	97.9
GHM Loss	95.2	93.4	97.1
Slide-Loss	97.1	94.6	98.5

As shown in Table 6, the Slide-Loss proposed in this paper performs the best among these four loss functions. Its precision rate leads all comparison schemes at 97.1%, with a recall rate of 94.6% and a mAP value of 98.5%. Compared with the benchmark model Focal Loss, Slide-Loss achieves a comprehensive breakthrough: the precision rate significantly increases by 3.3 percentage points, the recall rate improves by 3.2 percentage points, and the mAP is optimized by 1.1 percentage points. Compared with GHM Loss, the accuracy rate increased by 1.9 percentage points, the recall rate increased by 1.2 percentage points, and the average precision mean value increased by 1.4 percentage points. This result fully validates that Slide-Loss effectively and coordinately optimizes detection accuracy and recall capability while maintaining the same computational cost,

providing a superior loss function solution for classification models.

4.5 Ablation experiment

To assess how the optimization strategy proposed herein for the performance of the YOLOv10n network, determine its contribution to enhancing the model's overall effectiveness, and verify the performance of the model framework introduced in this study, ablation tests were performed using the same training parameters and the same data set. The outcomes are listed in Table 7. Model 1 is the base model YOLOv10n, Model 2 modifies the base model by integrating Slimneck in place of its Neck section, while in Model 3, the loss function is swapped for Slide-Loss. "✓" indicates that the module is added, and "×" indicates that the module is not added.

Table 7: Results of the ablation experiment.

Model	YOLO v10n	Slim-neck	Slide-Loss	P%	Recall%	mAP%	Computational Cost/GFLOPs
1	✓	×	×	94.3	95.3	98.0	6.5
2	✓	✓	×	96.0	94.3	98.5	5.9
3	✓	×	✓	95.7	94.9	98.5	6.5
CQN-YOLO (ours)	✓	✓	✓	97.8	96.6	99.0	5.9

According to Table 7, the accuracy, recall, and mAP of the CQN-YOLO model introduced are the highest among all models, reaching 97.8%, 96.6%, and 99.0%, respectively. In comparison to the benchmark model, the accuracy improved from 94.3% to 97.8%, an increase of 3.5 percentage points; The recall rose from 95.3% to 96.6%, an increase of 1.3 percentage points. Additionally, the mAP metric advanced from 98.0% to 99.0%, showing a growth of 1.0 percentage points. At the same time, computing volume decreased by 10.2% from 6.5 GFLOPs to 5.9 GFLOPs. The results indicate that the CQN-YOLO model proposed in this study significantly improves the non-destructive testing of corn appearance quality, while reducing the calculation cost and improving the detection efficiency. Fig.7 shows the detection effect of the CQN-YOLO model on 7 kinds of corn with different appearance qualities.

4.6 YOLO series model performance comparison experiment

To confirm the performance effect of the CQN-YOLO model proposed in this paper on corn appearance quality detection, the current YOLO series mainstream target detection algorithms, YOLOv3, YOLOv5s, YOLOv8n, and YOLOv9t, were selected for comparison experiments with the proposed algorithm under identical experimental settings and datasets conditions. The outcomes are presented in Table 8.

Table 8 shows that the recall of YOLO series mainstream target detection algorithms is above 90%, and the mAP value is above 95%. The algorithm CQN-YOLO in this paper has the highest accuracy, recall, and mAP under the condition of minimum computation, indicating that it has a significant advantage in detection performance. Although it has a slight increase in parameters and model size compared to YOLOv9t, its accuracy improvement is more significant, demonstrating a good balance between performance and complexity.



Figure 7: Model detection effect diagram.

4.7 Comparison experiments of different target detection models

To evaluate the comprehensive performance of the CQN-YOLO model, this section designs a comparison experiment with different target models. The representative classic models Faster R-CNN and SSD, as well as the newer lightweight Transformer model RT-DETR (including two versions r18 and r34) are selected for comparison. Through testing under the same dataset and experimental environment, the practical results are shown in Table 9.

As can be seen from Table 9, Faster-RCNN has an extremely high recall rate, but the model's accuracy rate is only 57.1%. This is because during the prediction process, the core task of the model is to generate as many candidate regions as possible that contain potential targets. When dealing with a large number of detection boxes, the classifier has

Table 8: YOLO series model performance comparison experiment

Model	P%	Recall%	mAP%	Computational Cost/GFLOPs	Params(M)	size(M)
YOLOv3	90.3	91.9	95.2	18.9	29.9	62.7
YOLOv5s	93.0	91.2	96.5	16.5	7.2	16.4
YOLOv8n	93.6	93.7	97.5	8.1	3.0	6.2
YOLOv9t	84.1	93.0	97.7	7.4	2.0	4.5
CQN-YOLO	97.8	96.6	99.0	5.9	2.1	4.8

Table 9: Comparative Experiment of Target Models

Model	P%	Recall%	mAP%	Computational Cost/GFLOPs	Params(M)	size(M)
Faster-RCNN	57.1	97.2	89.6	401.8	136.9	527.2
SSD	79.5	89.7	87.5	62.7	26.3	95.4
RT-DETR-r18	84.6	83.6	80.7	57	19.9	77.0
RT-DETR-r34	87.4	86.3	84.1	88.8	30.2	61.5
CQN-YOLO	97.8	96.6	99.0	5.9	2.1	4.8

difficulty completely eliminating all false detections, ultimately resulting in the output containing a considerable number of incorrect detections. In RT-DETR, the accuracy and recall rates have achieved a good balance, but they are still significantly lower than those of our proposed model. Moreover, GFLOPs have also increased considerably. Compared with other models, the method proposed in this paper significantly reduces the number of parameters and the model size, demonstrating excellent lightweight characteristics. This indicates that CQN-YOLO, while maintaining extremely low computational complexity, has a significant advantage in achieving high-precision object detection. It is the most outstanding model in this comparative experiment.

5 Conclusion

This paper addresses the issues of insufficient training data, high computational cost, and insufficient robustness in the detection of corn appearance quality, and proposes an improved corn quality non-destructive detection model (CQN-YOLO) based on YOLOv10n as the benchmark model. By introducing data augmentation, a light weight slimneck structure composed of GSConv and GSbottleneck, and the VoV-GSCSP module, the complexity of the model is reduced while the feature representation ability is enhanced. Additionally, the Slide Loss loss function is adopted to improve the multi-scale target matching ability and alleviate the problem of data imbalance, thereby improving the detection accuracy. The experimental results show that the proposed method has excellent effectiveness and efficiency. The average detection accuracy, recall, and mAP were 97.8%, 96.6%, and 99.0%, respectively, and GFLOPs were only 5.9. While ensuring detection precision, the model remains lightweight, with enhanced generalization, efficiency, and robustness. This study provides a practical and lightweight solution framework for phenotypic analysis and high-throughput appearance detection of other agricultural products. The Slimneck structure and Slide Loss function used can be easily migrated to other crop

(wheat, rice, fruit) disease detection tasks. This enables the appearance quality testing of other agricultural products in the field.

Although our model has achieved good results in overall accuracy, it faces certain limitations. For some varieties, such as Moldy kernels, even when using Slide Loss, the recall rate only increased from 91.3% to 93.0%. This may be due to the model's insufficient ability to distinguish similar external features. For future work, we will attempt to introduce an attention mechanism to enhance feature learning and improve the discrimination ability of some similar types. We will also enhance the simulation of actual scenarios, cooperate with grain depots and processing plants, and achieve real-time inference of the model by deploying edge computing devices (such as Jetson), to effectively implement the model.

References

- [1] Standardization Administration of China. Corn (gb/t 1234-2024). Technical report, Standards Press of China, Beijing, 2018.
- [2] U.S. Grains Council. 2024/2025 corn harvest quality report. Technical report, U.S. Grains Council, 2024.
- [3] Diwakar Agarwal, Sweta, and P. Bachan. Machine learning approach for the classification of wheat grains. *Smart Agricultural Technology*, 3:100136, 2023.
- [4] Mukesh Dalal and Payal Mittal. A systematic review of deep learning-based object detection in agriculture: Methods, challenges, and future directions. *Computers, Materials and Continua*, 84(1):57–91, 2025.
- [5] Ang Wu, Juanhua Zhu, Yuli Yang, Xinping Liu, Xiushan Wang, Ling Wang, Hao Zhang, and Jing Chen. Classification of corn kernels grades using image analysis and support vector machine. *Advances in Mechanical Engineering*, 10, 2018.
- [6] MA Jia-jia, WANG Ke-qiang, and ZHENG Yi-xiong. Research on the appearance quality detection and classification of peanut seeds based on machine vision. *Journal of Anhui Agricultural Sciences*, 49(10):4, 2021.
- [7] GUO Dongdong, CAI Xiang, QI Jiandong, and et al. Low-cost grading device for mango appearance quality using accelerated cnn with fpga. *Transactions of the Chinese Society of Agricultural Engineering*, 40(21):137–145, 2024.
- [8] Zhou Chun-Xina, Huo Yi-Zhi, Du You-Hai, Jiang Min-Lan, Zeng Ling-Guo, Zhang Chang-Jiang, and Shi Xiao-Wei. Non-destructive detection of soybean appearance quality based on improved retinanet. *Journal of the Chinese Cereals and Oils Association*, 39(9):172–180, 9 2024.
- [9] Haoyang Wang, Guoxiong Zhou, and Guiyun Chen. Cea-net: A multi-modal model for corn disease classification with dynamic fusion and cross-layer connection mechanism. *Pattern Recognition*, 173:112788, 2026.
- [10] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016.

- [11] Gu Hai-Yang, Jiang Xue-Song, Shen Fei, Yu Xin-Zhi, and Zhou Hong-Ping. Non-destructive testing model of rice appearance quality. *Journal of the Chinese Cereals and Oils Association*, 38(11):230–237, 11 2023.
- [12] FAN Xiaofei, WANG Linbai, LIU Jingyan, ZHOU Yuhong, ZHANG Jun, and SUO Xuesong. Corn seed appearance quality estimation based on improved yolo v4. *Transactions of the Chinese Society for Agricultural Machinery*, 53(7):226–233, 2022.
- [13] Ke Sun, Mengdi Tang, Shu Li, and Siyuan Tong. Mildew detection in rice grains based on computer vision and the yolo convolutional neural network. *Food Science & Nutrition*, 12(2):860–868, 2024.
- [14] Tirthankar Mohanty, Priyabrata Pattanaik, Subhaprada Dash, Hara Prasada Tripathy, and William Holderbaum. Smart robotic system guided with yolov5 based machine learning framework for efficient herbicide usage in rice (*oryza sativa* l.) under precision agriculture. *Computers and Electronics in Agriculture*, 231:110032, 2025.
- [15] Siqi Niu, Xiaolin Xu, Ao Liang, Yuliang Yun, Fengqi Hao, Jinqiang Bai, Wenyi Sheng, and Dexin Ma. High accuracy identification of maize seed varieties based on a lightweight improved yolov8. *Scientific Reports*, 15:18541, 2025.
- [16] Hao Sun, Rui Fu, and Dae-Ki Kang. Ligtomdet: Knowledge distillation in a new lightweight tomato disease detection model in planting fields. *Pattern Recognition*, 172:112628, 2026.
- [17] Lei Fan, Yiwen Ding, Dongdong Fan, Donglin Di, Maurice Pagnucco, and Yang Song. Grainspace: A large-scale dataset for fine-grained and domain-adaptive recognition of cereal grains. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21084–21093, 2022.
- [18] Xinle Gao, Zhiyong Xiao, and Zhaohong Deng. High accuracy food image classification via vision transformer with data augmentation and feature augmentation. *Journal of Food Engineering*, 365:111833, 2024.
- [19] Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, and Guiguang Ding. Yolov10: Real-time end-to-end object detection. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 107984–108011. Curran Associates, Inc., 2024.
- [20] Hulin Li, Jun Li, Hanbing Wei, Zheng Liu, Zhenfei Zhan, and Qiliang Ren. Slimneck by gsconv: A lightweight-design for real-time detector architectures. *Journal of Real-Time Image Processing*, 21(3):62.1–62.13, 2024.
- [21] Amalia Luque, Alejandro Carrasco, Alejandro Martín, and Ana de las Heras. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91:216–231, 2019.
- [22] Ziping Yu, Hongbo Huang, Weijun Chen, Yongxin Su, Yahui Liu, and Xiuying Wang. Yolo-facev2: A scale and occlusion aware face detector. *Pattern Recognition*, 155:110714, 2024.