



OPEN ACCESS

EDITED BY

Ting Sun,
China Jiliang University, China

REVIEWED BY

Hongwei Sun,
Hangzhou Dianzi University, China
Yaohua Hu,
Zhejiang Agriculture and Forestry University,
China

*CORRESPONDENCE

Xin Jin

✉ jx.771@163.com

Shunhao Qing

✉ 18392163586@163.com

RECEIVED 12 September 2024

ACCEPTED 30 October 2024

PUBLISHED 19 November 2024

CITATION

Qiu Z, Wang F, Wang W, Li T, Jin X, Qing S
and Shi Y (2024) YOLO-SDL: a lightweight
wheat grain detection technology based on
an improved YOLOv8n model.

Front. Plant Sci. 15:1495222.

doi: 10.3389/fpls.2024.1495222

COPYRIGHT

© 2024 Qiu, Wang, Wang, Li, Jin, Qing and Shi.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

YOLO-SDL: a lightweight wheat grain detection technology based on an improved YOLOv8n model

Zhaomei Qiu¹, Fei Wang¹, Weili Wang¹, Tingting Li¹, Xin Jin^{1,2*},
Shunhao Qing^{1*} and Yi Shi¹

¹College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang, Henan, China, ²Science and Technology Innovation Center for Completed Set Equipment, Longmen Laboratory, Luoyang, Henan, China

Wheat, being a crucial global food crop, holds immense significance for food safety and agricultural economic stability, as the quality and condition of its grains are critical factors. Traditional methods of wheat grain detection are inefficient, and the advancements in deep learning offer a novel solution for fast and accurate grain recognition. This study proposes an improved deep learning model based on YOLOv8n, referred to as YOLO-SDL, aiming to achieve efficient wheat grain detection. A high-quality wheat grain dataset was first constructed, including images of perfect, germinated, diseased, and damaged grains. Multiple data augmentation techniques were employed to enhance the dataset's complexity and diversity. The YOLO-SDL model incorporates the ShuffleNetV2 architecture in its backbone and combines depthwise separable convolutions (DWConv) with the large separable kernel attention (LSKA) mechanism in its neck structure, significantly improving detection speed and accuracy while ensuring the model remains lightweight. The results indicate that YOLO-SDL achieves superior performance in wheat grain detection, balancing lightweight design and performance optimization. The model achieved a P of 0.942, R of 0.903, mAP50 of 0.965, and mAP50-95 of 0.859, with low computational complexity, making it suitable for resource-constrained environments. These findings demonstrate the efficiency of the ShuffleNetV2, DWConv, and LSKA structures. The proposed YOLO-SDL model provides a new technical solution for agricultural automation and serves as a reliable reference for detecting other crops.

KEYWORDS

YOLOv8, wheat grain, lightweight model, object detection, computer vision

1 Introduction

Wheat is one of the most widely cultivated food crops globally, and its grain yield and quality directly impact food security, making it a cornerstone of agricultural economies (Guo et al., 2023; Peng et al., 2022). During growth, harvesting, and storage, wheat grains can be affected by environmental factors or improper handling, leading to germination, damage, or disease, which compromise their quality (Ma et al., 2021). Thus, precise identification and detection of wheat grains, particularly distinguishing between different types of grains, is crucial for ensuring food safety and promoting sustainable agriculture (Que et al., 2023; Zhou et al., 2020). Moreover, grain detection plays a vital role in wheat breeding, enhancing breeding efficiency and aiding in the development of high-quality varieties (Misra et al., 2022).

Manual methods of detecting wheat grains suffer from inefficiency and subjectivity, failing to meet the demands of modern agriculture (Gao et al., 2022; Li et al., 2024c; Zhao et al., 2022). While traditional machine vision technology, which relies on computer-based image recognition and analysis, improves detection efficiency and accuracy, its generalization capability is limited. It tends to perform well on specific tasks but struggles to adapt to new ones, requiring expert intervention for feature extraction (Ji et al., 2022).

Recently, deep learning has made significant strides in object detection, particularly with deep neural networks excelling in image recognition and classification tasks (Feng et al., 2023; Han et al., 2023). Deep learning models for object detection are broadly categorized into two-stage and one-stage models. Two-stage models, such as the R-CNN series and Mask R-CNN, first generate candidate regions and then classify and localize these regions for accurate object detection (Fang et al., 2023; Li et al., 2024a). Li et al. (2022a) used a Faster R-CNN model to detect wheat spike in RGB images, achieving rapid and accurate detection with an average accuracy of 86.7%. Similarly, Ye et al. (2023) applied the Mask R-CNN model to visible-light images captured by uncrewed aerial vehicle to achieve rapid and accurate detection of individual cabbage plants. The study results demonstrated the superior performance of the Mask R-CNN model, with a detection accuracy of 99.5% and an overall average F1 score of 97.63%. Wu et al. (2020) combined transfer learning with the Faster R-CNN model for the detection and counting of wheat grains under complex environmental conditions. The results showed that the mean average precision (mAP) of the Faster R-CNN model was 0.91, and the model exhibited strong robustness. Although two-stage models exhibit high accuracy, they tend to be slower (Li et al., 2022b). In contrast, one-stage models like YOLO and SSD are widely favored for their speed and practicality, directly extracting image features to predict object categories and locations (Bacea and Oniga, 2023; Wan et al., 2024).

The YOLO series, a quintessential one-stage model, stands out for its exceptional performance and efficiency in object detection tasks (Wu et al., 2023; Zhang et al., 2024). However, while YOLO models are fast, there is still room to improve their detection accuracy. To optimize the YOLO algorithm, researchers have

worked to enhance its accuracy and computational efficiency while reducing complexity and parameter count, aiming for a lightweight yet efficient model. For example, Ma et al. (2024) proposed an improved YOLOv8 model for wheat seed detection, incorporating shared convolutional layers and vision transformer with deformable attention mechanism, which resulted in a lightweight and efficient model with a mAP of 99.3%, a 16.8% improvement over YOLOv8. Similarly, Yao et al. (2024) improved the YOLOv7 model by integrating an efficient channel attention mechanism and a bi-directional feature pyramid network for detecting the germination rate of wild rice seeds, achieving detection accuracies of 94% and 98.2% in hydroponic and petri dish environments, respectively. Qing et al. (2024) proposed an enhanced YOLO-FastestV2 model for the detection of wheat spike counts. The model incorporated three attention mechanisms and a SimConv structure to enhance the model's feature extraction capabilities and detection accuracy. The study demonstrated that the improved model achieved a precision (P) of 83.91%, recall (R) of 78.35%, average precision of 81.52%, and F1 score of 81.03%, indicating the overall best performance.

To enhance the accuracy and efficiency of wheat grain detection, this study proposes an improved YOLOv8n model, YOLO-SDL. The model adopts the ShuffleNetV2 architecture in its backbone, known for its computational efficiency, which not only boosts image processing speed but also significantly enhances detection accuracy. Additionally, depthwise separable convolutions (DWConv) replace the original convolutional modules in the neck of YOLOv8n, further reducing computational load, making the model more suitable for resource-constrained conditions. The large separable kernel attention (LSKA) mechanism is also integrated into the neck structure to improve the model's ability to capture key object features, thereby enhancing detection precision. The contributions of this study are as follows: (1) constructing a high-quality dataset of wheat grains with various conditions, (2) conducting ablation experiments to validate the effectiveness of each improvement in the YOLO-SDL model, and (3) comparing the performance of YOLO-SDL with YOLOv5n, YOLOv6n, YOLOv8n, and YOLOv10n.

2 Materials and methods

2.1 Wheat grain dataset

This study constructed a comprehensive and diverse wheat grain dataset, including images of perfect, germinated, diseased, and damaged grains, captured using high-resolution imaging equipment to ensure quality. After capturing the images, detailed bounding box annotations were made using the LabelImg software to ensure accurate identification of different grain conditions during model training. The dataset contains 9400 perfect grains, 7260 germinated grains, 835 diseased grains, and 1890 damaged grains, accounting for 48.49%, 37.45%, 4.31%, and 9.75% of the total, respectively. To further enhance the complexity and diversity of the dataset, various data augmentation techniques were employed, such as random scaling, rotation, flipping, noise addition, and

adjustment of lighting intensity. These methods expanded the dataset's size and richness, improving the model's stability and generalization ability in recognizing wheat grains under different environments and conditions. The final dataset comprises 1170 images, divided into training, testing, and validation sets in an 8:1:1 ratio. Examples of wheat grain images are shown in [Figure 1](#).

2.2 Improvement strategies

2.2.1 YOLO-SDL model

The YOLOv8n algorithm builds upon the advantages of the YOLO series in terms of both accuracy and speed, with further optimizations to enhance performance. The YOLOv8n model introduces the C2f structure, which integrates feature maps from various depths, enabling the network to leverage both shallow and deep features simultaneously. This significantly enhances the model's ability to recognize and capture object features ([Jiang et al., 2024](#); [Lu et al., 2024](#)). The cross stage partial connections in the C2f structure optimize information flow through the network. By partially connecting different stages, the model not only improves feature representation but also strengthens its ability to detect objects in complex scenes. The neck of the YOLOv8n model employs the path aggregation network - feature pyramid network structure, which builds multi-scale feature fusion paths, allowing the model to handle objects of various scales more flexibly and efficiently. The detection head uses a decoupled structure that separates the classification and localization tasks, enabling independent optimization of both and improving detection accuracy and efficiency ([Chen et al., 2024](#)).

To further enhance detection accuracy while reducing computational complexity and parameter count, making the model more lightweight and efficient, this study proposes improvements to the YOLOv8n model. The structure of the improved model is shown in [Figure 2](#). First, the ShuffleNetV2 architecture is introduced into the backbone of YOLOv8n, ensuring high precision while enabling fast object detection with lower computational cost. This is particularly crucial for tasks such as wheat grain detection, which often need to be conducted in environments like fields or warehouses where computational resources are limited. Additionally, we integrate DWConv and LSKA mechanisms into the neck of YOLOv8n. The DWConv structure significantly reduces the computational load, making the model more suitable for deployment on resource-constrained devices. The LSKA mechanism provides a more flexible feature extraction process, effectively enhancing the model's detection accuracy. This aids the model in better capturing detailed characteristics of wheat grain under various environments to achieve accurate detection. By integrating improvements such as ShuffleNetV2 and DWConv, the model proposed in this study achieves structural lightweight and efficiency, and is functionally particularly suited for wheat grain detection tasks. It is capable of maintaining high detection accuracy while adapting to resource-constrained environments and meeting the demand for processing large volumes of image data.

2.2.2 DWConv

DWConv reduces the number of parameters in the model by decomposing a standard convolution into two processes: depthwise convolution and pointwise convolution, significantly lowering



FIGURE 1
Sample images from the wheat grain dataset.

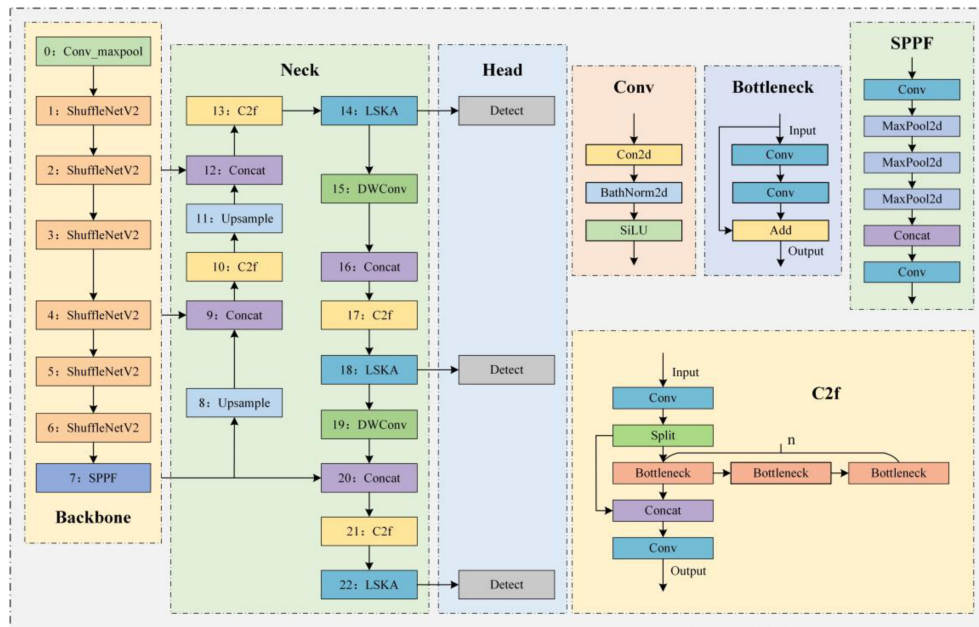


FIGURE 2 Structure of the YOLO-SDL model.

computational complexity (Balasundaram et al., 2023). During the depthwise convolution stage, each channel undergoes independent convolution operations, generating feature maps corresponding to the number of channels. This process enables the model to precisely recognize and capture the local features of each channel without a significant increase in the number of parameters. Pointwise convolution, which uses a 1×1 convolution kernel, integrates the feature maps produced by the depthwise convolution across channels, generating the final output feature map. The DWConv structure preserves the model's ability to accurately capture spatial features while dramatically reducing its computational complexity

and parameter count, resulting in a more lightweight model (Li et al., 2024b; Shi et al., 2024). DWConv replaces the original convolution structure in the YOLOv8n model, reducing model complexity and achieving further lightweight optimization. Figure 3 illustrates the DWConv structure.

2.2.3 ShuffleNetV2

ShuffleNetV2 is an efficient and lightweight convolutional neural network architecture that balances model complexity with accuracy (Thakuria and Erkinbaev, 2024). It employs group convolution to divide the input channels into groups, where

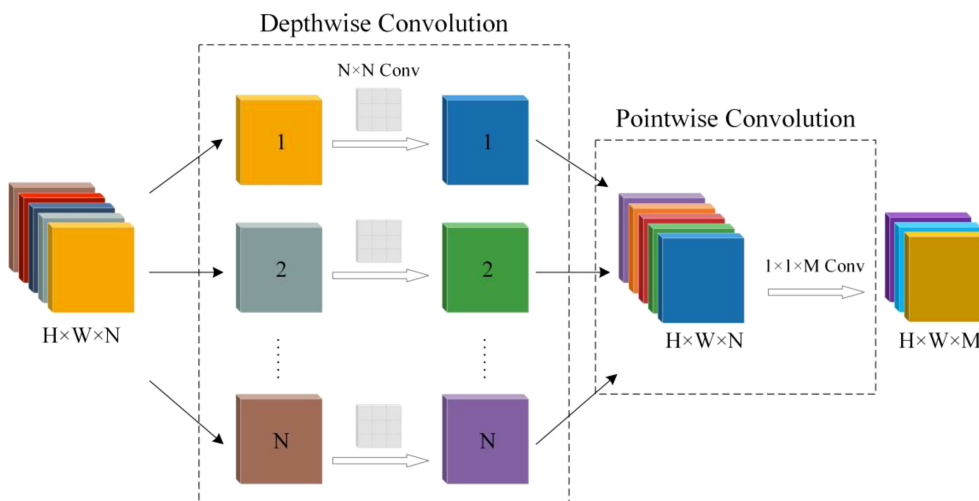


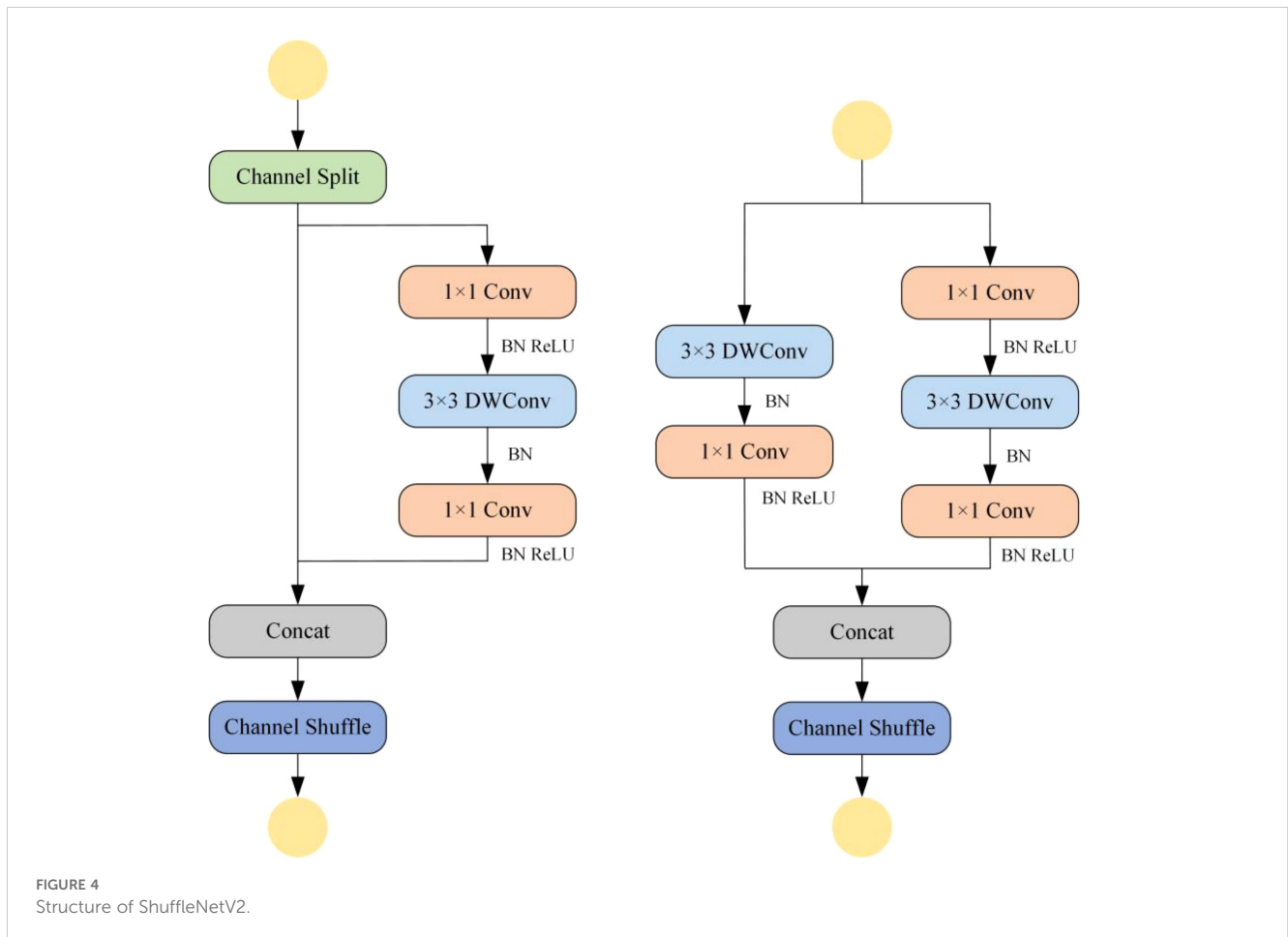
FIGURE 3 Structure of DWConv.

independent convolution operations are performed within each group. This technique significantly reduces the number of parameters in the convolutional layers, thereby lowering the computational cost of the model. ShuffleNetV2 also incorporates an innovative channel shuffle mechanism, which enables effective feature fusion after the grouping process, enhancing the model's ability to integrate features and improving detection accuracy. Moreover, the integration of DWConv further reduces the computational complexity of the model. Due to its lightweight and efficient nature, ShuffleNetV2 offers an effective solution for resource-constrained environments. ShuffleNetV2 exhibits a reduced computational parameter count relative to EfficientNet-B0, which is the smallest model within the EfficientNet series (Ding et al., 2023). It also demonstrates superior performance in both speed and accuracy compared to MobileNetV2, providing faster inference capabilities. These attributes render ShuffleNetV2 a highly suitable choice for applications within resource-limited environments (Ma et al., 2018). The channel shuffle mechanism and the design of the DWConv structure in ShuffleNetV2 provide better robustness when dealing with small objects, which is crucial for wheat grain identification and detection tasks. In this study, the ShuffleNetV2 architecture was applied to the backbone of the YOLO-SDL model, enabling it to adapt to environments with

limited resources while maintaining fast and accurate object detection capabilities. Figure 4 shows the structure of ShuffleNetV2.

2.2.4 LSKA

LSKA is an innovative attention mechanism that decomposes a 2D convolutional kernel in the deep convolutional layers into two 1D convolutional kernels, performing horizontal and vertical convolution operations separately. This reduces the number of parameters and computational load while enhancing the model's ability to capture image features (Lau et al., 2024). LSKA generates an attention map through a 1×1 convolution layer based on the 1D convolutions, which is used to adjust and enhance key information in the feature maps. The attention map is then multiplied pointwise with the original feature map to achieve adaptive optimization of the features (Zhong et al., 2024). Compared to traditional large kernel attention mechanisms, LSKA overcomes the computational and memory efficiency challenges associated with large convolutional kernels. It significantly reduces computational complexity, making it more efficient and flexible in object detection tasks, while providing more precise detection results. In this study, the LSKA mechanism was introduced into the neck network of the YOLOv8n model to further optimize its performance. Figure 5 shows the structure of LSKA mechanisms.



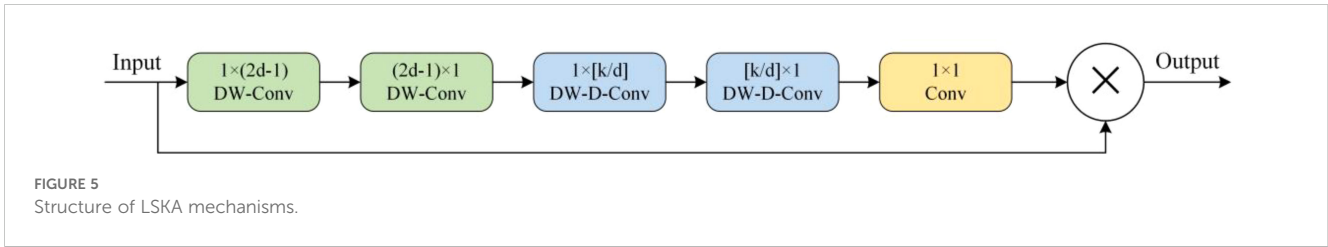


FIGURE 5
Structure of LSKA mechanisms.

2.3 Evaluation metrics

To assess the performance of different models in detecting wheat grains, this study employs P, R, mean average precision (mAP50), extended mean average precision (mAP50-95), and giga floating-point operations per second (GFLOPs) as evaluation metrics. P measures the proportion of correctly predicted positive samples out of all samples predicted as positive, as shown in Equation 1. R evaluates the proportion of actual positive samples that were correctly predicted by the model, as shown in Equation 2. mAP50 represents the mean average precision when the intersection over union (IoU) threshold is set at 50%, serving as a general indicator of the model’s overall performance, as shown in Equation 3. mAP50-95 is an extended version of mAP50 that considers a range of IoU thresholds (from 50% to 95% in increments of 5%), providing a more detailed performance evaluation across various levels of IoU, as shown in Equation 4. GFLOPs quantifies the capability of performing ten billion floating-point operations per second and serves as a pivotal metric for assessing the efficiency of deep learning models. It indicates the computational velocity of the models during both training and inference; a higher GFLOPs value signifies a model’s enhanced data processing capacity. As illustrated in Equation 5. The Frames per second (FPS) is an important indicator of the speed of model recognition and detection, which reflects the number of images processed by the model per second, the larger the FPS, the faster the processing speed of the model.

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$mAP_{50} = \frac{1}{m} \sum_{i=1}^m AP_i \tag{3}$$

$$mAP_{50-95} = \frac{1}{m} \sum_{i=1}^m \frac{1}{10} \sum_{j=1}^{10} AP_{i,j} \tag{4}$$

$$GFLOPs = \frac{2 \times (K \times K \times C_{in} \times C_{out} \times H \times W + C_{out})}{10^9} \tag{5}$$

$$FPS = \frac{1}{Processing\ time\ per\ frame} \tag{6}$$

In this context, TP refers to the number of true positives, TN refers to the number of true negatives, FP refers to the number of

false positives, and FN refers to the number of false negatives. m represents the number of classes in the dataset, AP_i represents the average precision for class i when the IoU threshold is 0.5, and $AP_{i,j}$ represents the average precision for class i at IoU thresholds of $0.5 + 0.05 \times (j - 1)$. $K \times K$ denotes the size of the convolutional kernel, C_{in} represents the number of input channels in the convolutional layer, C_{out} represents the number of output channels in the convolutional layer, and $H \times W$ indicates the height and width of the feature map output by the convolutional layer.

3 Results

3.1 Experimental environment

The experiments were conducted on the Ubuntu 20.04 operating system, with Python 3.10 as the programming language, CUDA version 11.8, and PyTorch 2.0.1 as the deep learning framework. Jupyter was used as the IDE for developing the experiments. The CPU was an Intel(R) Xeon(R) Gold 5318Y @ 2.10GHz, and the GPU used was an NVIDIA A16 with 15GB of memory. During the network training process, this study sets the learning rate to 0.01 and the batch size to 200, with the loss function chosen as IoU. We employ a stochastic gradient descent optimizer with a momentum of 0.937, along with weight decay of 0.0005 to regularize the model and prevent overfitting. To more smoothly initiate the training process, we set up 3 warmup epochs, during which the initial momentum is set to 0.8 and the initial bias learning rate is set to 0.1. Such settings help to gradually enhance the model’s stability and convergence rate at the initial stage of training.

3.2 Ablation experiments

This study utilized the YOLOv8n model as a baseline and made several improvements by incorporating ShuffleNetV2, DWConv, and LSKA structures to enhance the accuracy of wheat grain detection. Ablation experiments were conducted to analyze the performance impact of each modification, with the results summarized in Table 1. In the ablation experiments, the results of different improved models on wheat grain feature recognition are shown in Figure 6.

The baseline YOLOv8n model achieved a detection performance of $P = 0.898$, $R = 0.903$, $mAP50 = 0.937$, and $mAP50-95 = 0.816$. The model had 3006623 parameters and a GFLOP count of 8.1. After introducing the ShuffleNetV2 architecture, the YOLO-S model improved its P to 0.909 and R to

TABLE 1 Accuracy results of ablation experiments.

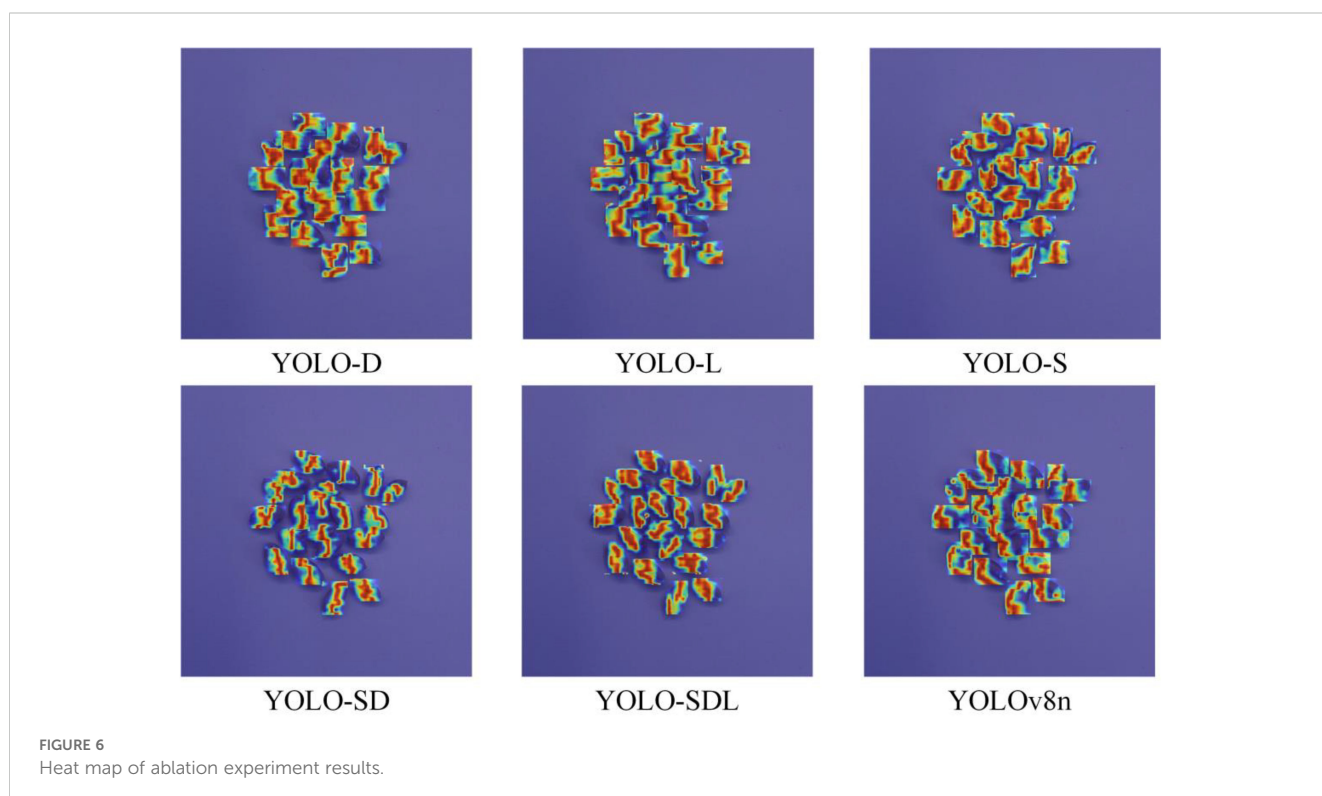
Model	ShuffleNetV2	DWConv	LSKA	P	R	mAP50	mAP50-95
YOLOv8n	-	-	-	0.898	0.903	0.937	0.816
YOLO-S	√	-	-	0.909	0.917	0.954	0.834
YOLO-D	-	√	-	0.904	0.904	0.944	0.823
YOLO-L	-	-	√	0.904	0.926	0.952	0.832
YOLO-SD	√	√	-	0.911	0.917	0.960	0.848
YOLO-SDL	√	√	√	0.942	0.903	0.965	0.859

√ indicates that the module was used in this set of experiments; - indicates that the module was not used in this set of experiments.

0.917, with mAP50 and mAP50-95 reaching 0.954 and 0.834, respectively. The number of parameters and GFLOPs were reduced to 1835343 and 5.2, indicating that ShuffleNetV2 not only enhanced detection performance but also achieved significant model lightweighting. By incorporating the DWConv structure, the YOLO-D model achieved P = 0.904 and R = 0.904, with mAP50 and mAP50-95 improving to 0.944 and 0.823, respectively. The parameter count was 2824031, and GFLOPs were 7.9. While DWConv contributed to improved precision and target recognition ability, it offered slightly less performance improvement compared to ShuffleNetV2, suggesting potential for further optimization. With the introduction of the LSKA mechanism, the YOLO-L model saw increased complexity, with parameters growing to 3100255 and GFLOPs to 8.3. However, detection performance improved, with P = 0.904, R = 0.926, and mAP50 and mAP50-95 reaching 0.952 and 0.832, respectively, demonstrating the significant impact of LSKA on enhancing detection accuracy.

The YOLO-SD model, combining the strengths of ShuffleNetV2 and DWConv, achieved P = 0.911, R = 0.917, mAP50 = 0.960, and mAP50-95 = 0.848, with the parameter count reduced to 1652751 and GFLOPs to 4.9. These results indicate that the model achieved a high level of performance while further reducing its complexity. Finally, the YOLO-SDL model, which integrates ShuffleNetV2, DWConv, and LSKA, achieved the best overall performance with P = 0.942, mAP50 = 0.965, and mAP50-95 = 0.859. Although recall dropped slightly to 0.903, the parameter count was 1755215, and GFLOPs were 5.1. These results demonstrate the model's ability to balance lightweight design with high accuracy and efficiency. According to the results demonstrated in Figure 6, it can be seen that the YOLO-SDL model focuses on the image region closer to the wheat grain, which has better recognition prediction effect.

This study significantly improved the accuracy of wheat grain detection by incorporating ShuffleNetV2, DWConv, and LSKA into the YOLOv8n model, achieving a balance between model lightweighting and computational efficiency. The results show



that the ShuffleNetV2 architecture enhanced detection accuracy while significantly reducing model parameters and computational load. The DWConv structure provided modest accuracy improvements and reduced computational costs, though it still has room for further optimization. While the LSKA structure increased model complexity, it substantially improved detection accuracy. Among the various improved models, YOLO-SDL strikes the best balance between accuracy, parameter count, and computational efficiency, making it the optimal model for wheat grain detection.

3.3 Comparison of different models

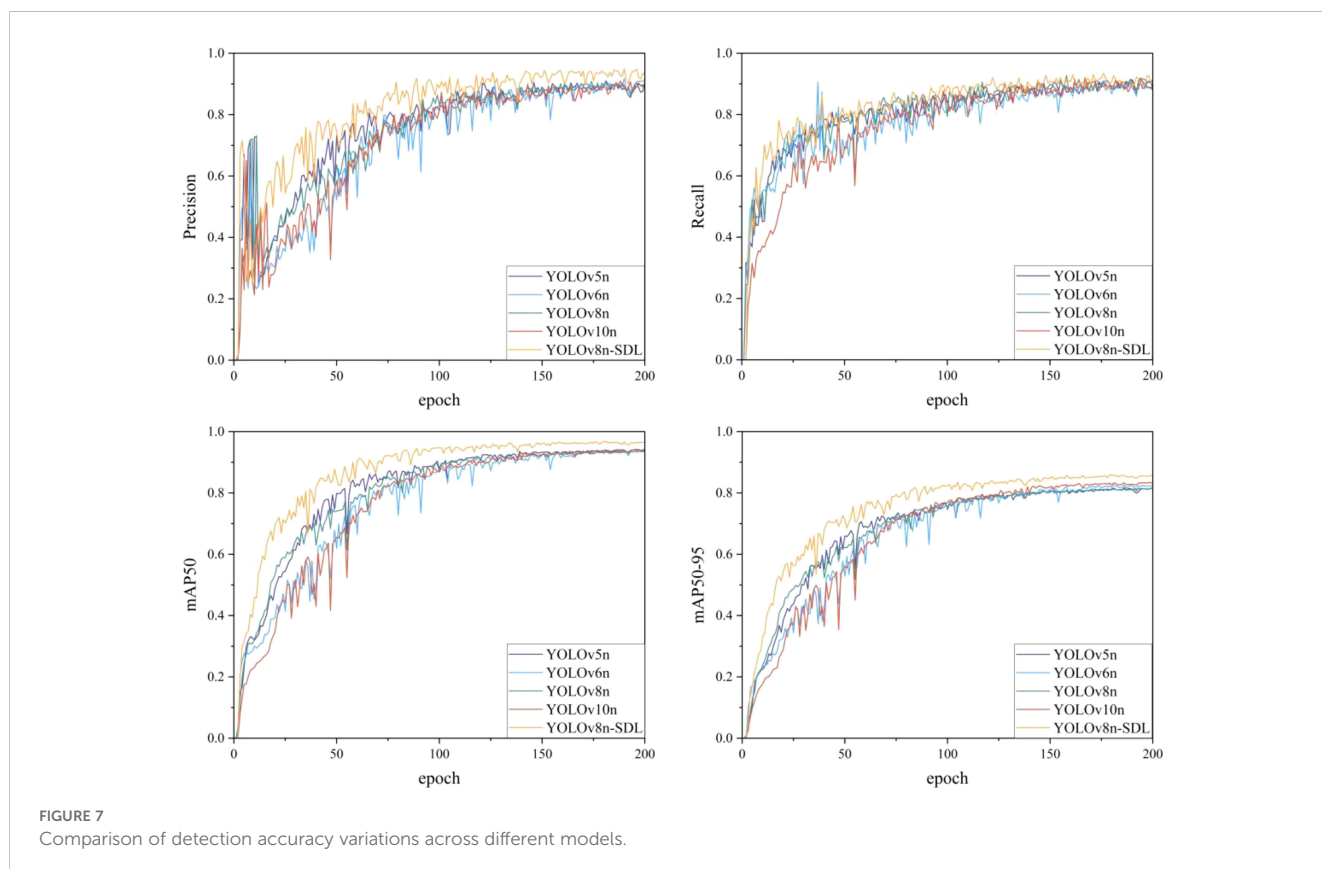
This study compares the YOLO-SDL model with three other models from the YOLO series—YOLOv5n, YOLOv6n, and YOLOv10n—evaluating their accuracy and performance in wheat

grain detection. The detection accuracy of the different models is summarized in Table 2, with the accuracy comparison trends shown in Figure 7, and the detection results comparison presented in Figure 8.

As shown in Table 2, the YOLOv5n model achieved relatively high P and R, with $P = 0.900$ and $R = 0.887$. Its mAP50 reached 0.940, while mAP50-95 was 0.817. The model also demonstrated a processing speed of 156.250 FPS, indicating a balance between high accuracy and fast processing. The YOLOv6n model had slightly higher P than YOLOv5n at 0.902, with $R = 0.888$, $mAP50 = 0.938$, and mAP50-95 improving to 0.827. It showed a real-time processing advantage with a FPS of 106.383. The YOLOv8n model had slightly lower P ($P = 0.898$), but R increased to 0.903, with $mAP50 = 0.937$ and $mAP50-95 = 0.816$. Its processing speed was 140.845 FPS, faster than YOLOv6n. The YOLOv10n model had the lowest P ($P = 0.895$) and R ($R = 0.887$), although its mAP50 and mAP50-95 were 0.940 and 0.834, respectively. It had the highest

TABLE 2 Accuracy of detection results for different models.

Model	P	R	mAP50	mAP50-95	Parameters	GFLOPs	Speed (FPS)
YOLOv5n	0.900	0.887	0.940	0.817	2503919	7.1	156.250
YOLOv6n	0.902	0.888	0.938	0.827	4234239	11.8	106.383
YOLOv8n	0.898	0.903	0.937	0.816	3006623	8.1	140.845
YOLOv10n	0.895	0.887	0.940	0.834	2696366	8.2	175.439
YOLO-SDL	0.942	0.903	0.965	0.859	1755215	5.1	161.290



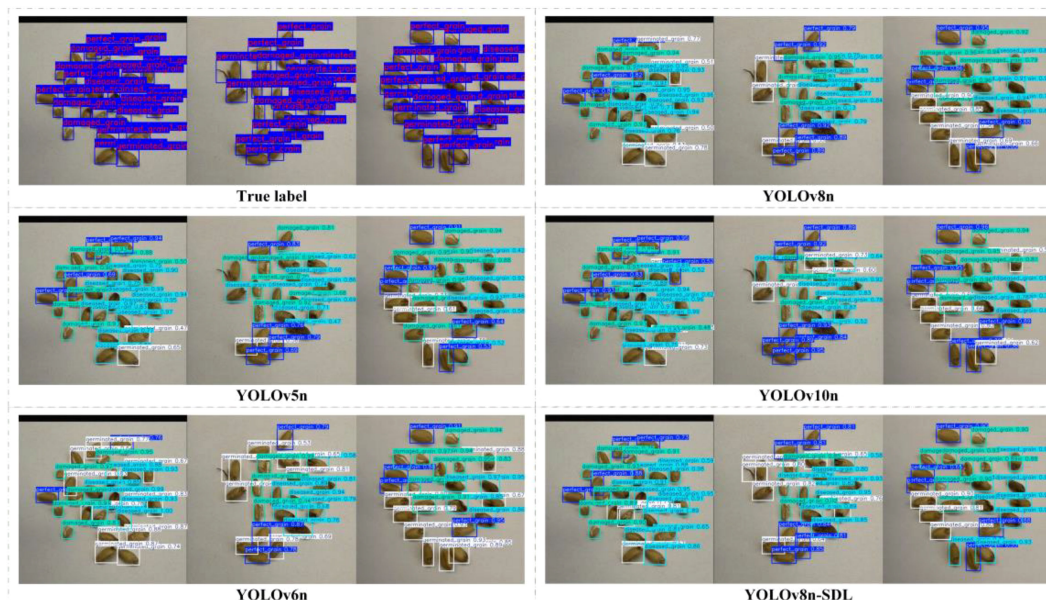


FIGURE 8
Comparison of detection results of different models.

processing speed among all models at 175.439 FPS. In contrast, the YOLO-SDL model demonstrated a significant accuracy advantage in the wheat detection task, with P reaching 0.942, R at 0.903, and mAP50 and mAP50-95 of 0.965 and 0.859, respectively, making it the most accurate model among all tested. The YOLO-SDL also had a processing speed of 161.290 FPS, second only to YOLOv10n, indicating a fast image processing capability. From the perspective of computational complexity and efficiency, YOLO-SDL was the most lightweight model, with the smallest parameter count (1755215) and GFLOPs (5.1). In comparison, YOLOv6n had the highest computational cost, with 4234239 parameters and 11.8 GFLOPs. The YOLOv5n model was less complex, with 2503919 parameters and 7.1 GFLOPs, while the YOLOv8n and YOLOv10n models had intermediate levels of complexity.

The Figure 6 shows the changes in accuracy for each model as the number of iterations increased. All models demonstrated an upward trend in accuracy over time, with the YOLO-SDL model maintaining the highest accuracy curve, consistently outperforming the other models. Additionally, as shown in Figure 7, the detection results indicate that the YOLO-SDL model was able to more accurately identify various types of wheat grains compared to the other models. In summary, the analysis demonstrates that appropriate adjustments and optimizations to the network structure can significantly enhance model performance. By integrating the strengths of ShuffleNetV2, DWConv, and LSKA, the YOLO-SDL model achieves the goal of reducing model complexity while maintaining high accuracy. This combination of lightweight design and high precision not only improves the model's adaptability and practicality but also provides new insights for the fields of agricultural automation and smart agriculture.

4 Discussion

This study introduces improvements to the YOLOv8n model by incorporating ShuffleNetV2, DWConv, and LSKA structures, with the goal of optimizing detection accuracy, reducing the model's parameter count, and lowering computational complexity. These enhancements aim to improve the model's performance in wheat grain detection tasks while achieving a balance between model lightweighting and computational efficiency. Through ablation experiments and comparative analysis with different models, this research demonstrates that the YOLO-SDL model is the optimal solution for wheat grain detection, offering high-precision results while maintaining a lightweight design.

The YOLO-SDL model leverages the ShuffleNetV2 architecture as its backbone, significantly improving detection accuracy while greatly reducing both the model's parameter count and computational load. ShuffleNetV2, an efficient lightweight network architecture, uses grouped convolution, channel shuffle mechanisms, and DWConv to capture rich feature information while minimizing computational complexity. These lightweight features are especially valuable in resource-constrained environments. Shang et al. (2023) demonstrated in their study on apple flower detection that ShuffleNetv2 possesses significant advantages in terms of lightweight design, accuracy, and adaptability, making it an ideal choice for object detection in resource-constrained environments. Additionally, the introduction of DWConv and LSKA structures in the neck of the YOLO-SDL model further enhances its performance. While the DWConv structure does not provide as significant a precision boost as ShuffleNetV2, it plays a crucial role in reducing computational costs. By lowering the number of parameters and

computational load in convolution operations, DWConv improves both efficiency and accuracy. Guo et al. (2024) showed that replacing the convolutional modules in the YOLOv8s network with DWConv modules can reduce the complexity of the network, proving to be an effective lightweight strategy. Although LSKA increases model complexity, it also delivers significant accuracy improvements. By decomposing large kernels into smaller, more manageable parts, LSKA can effectively model global information, which is crucial for distinguishing between wheat grains and other components such as husks and stalks. The LSKA mechanism enhances the model's ability to recognize features in wheat grains by adaptively optimizing local spatial attention within the feature maps. Gao et al. (2024) pointed out in their research aimed at improving the efficiency and accuracy of young citrus fruit detection that the LSKA mechanism enhances the model's ability to fuse features across different scales, thereby improving detection accuracy. The inclusion of DWConv and LSKA in YOLO-SDL boosts detection accuracy, thereby improving overall model performance.

In the model comparison analysis, YOLO-SDL demonstrated clear performance advantages in wheat grain detection tasks. The model achieved the highest values for P, R, mAP50, and mAP50-95 compared to other YOLO series models. From a computational complexity and efficiency standpoint, YOLO-SDL is the most lightweight model, outperforming the others in terms of parameter count and GFLOPs. These findings suggest that appropriate network adjustments and optimizations can significantly enhance overall detection performance while maintaining a lightweight design.

In conclusion, the YOLO-SDL model introduced in this study has demonstrated superior performance in wheat grain detection tasks. By combining the strengths of ShuffleNetV2, DWConv, and LSKA, the model achieves significant improvements in accuracy while balancing lightweight design and performance optimization. These results offer new approaches and methods for agricultural automation and smart farming. Future research could explore further reductions in model complexity, optimize the network structure, and enhance the model's adaptability in resource-limited environments, advancing the development and application of agricultural automation technologies.

5 Conclusion

This study developed an improved YOLOv8n model, termed YOLO-SDL, for detecting various types of wheat grains. Through in-depth analysis and enhancement of the YOLOv8n model, a lightweight, efficient, and highly accurate wheat grain detection model was constructed. By introducing ShuffleNetV2, DWConv, and LSKA structures, the YOLO-SDL model significantly improves the capture of features across different wheat grain types while maintaining high precision, lightweight design, and computational efficiency. The results show that the YOLO-SDL model excels in

detection accuracy, achieving a P of 0.942, R of 0.903, mAP50 of 0.965, and mAP50-95 of 0.859, outperforming YOLOv8n and other YOLO series models in both lightweight design and computational efficiency. The YOLO-SDL model's advantages in parameter count and computational complexity make it highly suitable for deployment in resource-constrained environments, which is particularly relevant for real-world agricultural automation and intelligent systems. The outstanding performance of the YOLO-SDL model is attributed to the introduction of ShuffleNetV2, DWConv, and LSKA structures. The ShuffleNetV2 architecture, through its use of grouped convolution and channel shuffle mechanisms, maintains high accuracy while reducing computational costs. The DWConv structure, by decomposing standard convolutions into depthwise and pointwise convolutions, further reduces parameter count and computational complexity. Finally, the LSKA mechanism, by decomposing 2D convolutional kernels into 1D kernels, achieves adaptive feature optimization, enhancing the model's ability to capture image features and improving detection accuracy.

The YOLO-SDL model demonstrates excellent performance in wheat grain detection tasks, improving both accuracy and efficiency while achieving a lightweight design. This provides a novel technical solution for agricultural automation and intelligent farming. Future research will continue to explore and optimize the model's structure to extend its applicability to a wider range of agricultural scenarios, driving further advancements in agricultural technology.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

ZQ: Writing – review & editing, Supervision, Resources, Investigation, Conceptualization. FW: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis. WW: Writing – review & editing, Formal analysis, Data curation. TL: Writing – review & editing, Visualization, Formal analysis, Data curation. XJ: Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. SQ: Writing – review & editing, Software, Methodology. YS: Writing – review & editing, Software, Project administration, Funding acquisition.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by National Key Research and Development

Program Sub-project (No. 2022YFD2001205C), and Henan Province Major Science and Technology Special Project (No. 231100110200), and Key R&D and Promotion Projects in Henan Province (Science and Technology Development) (No.222102110452), and Key Scientific Research Projects of Colleges and Universities in Henan Province (No. 22B416002).

Acknowledgments

We extend our sincere gratitude to the farmers of Shenqiu County, Zhoukou City, Henan Province, for their invaluable support in providing experimental materials of wheat grains. Their contributions have been instrumental in facilitating the successful conduct of our research.

References

- Bacea, D.-S., and Oniga, F. (2023). Single stage architecture for improved accuracy real-time object detection on mobile devices. *Image. Vision Computing*. 130, 104613. doi: 10.1016/j.imavis.2022.104613
- Balasundaram, A., Shaik, A., Pradeep, K., Vijayakumar, K. P., and Kavitha, M. S. (2023). Zero-DCE++ Inspired object detection in less illuminated environment using improved YOLOv5. *Computers. Mater. Continua*. 77, 2751–2769. doi: 10.32604/cmc.2023.044374
- Chen, G., Hou, Y., Cui, T., Li, H., Shanguan, F., and Cao, L. (2024). YOLOv8-CML: A lightweight target detection method for Color-changing melon ripening in intelligent agriculture. *Sci. Rep.* 14, 14400. doi: 10.1038/s41598-024-65293-w
- Ding, Y., Huang, H., Cui, H., Wang, X., and Zhao, Y. (2023). A non-destructive method for identification of tea plant cultivars based on deep learning. *Forests* 14, 728. doi: 10.3390/f14040728
- Fang, S., Zhang, B., and Hu, J. (2023). Improved mask R-CNN multi-target detection and segmentation for autonomous driving in complex scenes. *Sensors* 23, 3853. doi: 10.3390/s23083853
- Feng, Y., Liu, C., Han, J., Lu, Q., and Xing, X. (2023). IRB-5-CA net: A lightweight, deep learning-based approach to wheat seed identification. *IEEE Access*. 11, 119553–119559. doi: 10.1109/ACCESS.2023.3326824
- Gao, A., Tian, Z., Ma, W., Song, Y., Ren, L., Feng, Y., et al. (2024). Fruits hidden by green: an improved YOLOv8n for detection of young citrus in lush citrus trees. *Front. Plant Sci.* 15, 1375118. doi: 10.3389/fpls.2024.1375118
- Gao, H., Zhen, T., and Li, Z. (2022). Detection of wheat unsound kernels based on improved ResNet. *IEEE Access*. 10, 20092–20101. doi: 10.1109/ACCESS.2022.3147838
- Guo, Y., Chen, Y., Wang, Y., Wu, X., Zhang, X., Mao, W., et al. (2023). The translational landscape of bread wheat during grain development. *Plant Cell*. 35, 1848–1867. doi: 10.1093/plcell/koad075
- Guo, P., Diao, Z., Zhao, C., Li, J., Zhang, R., Yang, R., et al. (2024). Navigation line extraction algorithm for corn spraying robot based on YOLOv8s-CornNet. *J. Field Robotics*. 41, 1887–1899. doi: 10.1002/rob.22360
- Han, K., Zhang, N., Xie, H., Wang, Q., and Ding, W. (2023). An improved strategy of wheat kernel recognition based on deep learning. *DYNA-Ingenieria. e. Industria*. 98, 91–97. doi: 10.6036/DYNAII
- Ji, W., Pan, Y., Xu, B., and Wang, J. (2022). A real-time apple targets detection method for picking robot based on ShufflenetV2-YOLOX. *Agriculture* 12, 856. doi: 10.3390/agriculture12060856
- Jiang, P., Qi, A., Zhong, J., Luo, Y., Hu, W., Shi, Y., et al. (2024). Field cabbage detection and positioning system based on improved YOLOv8n. *Plant Methods* 20, 96. doi: 10.1186/s13007-024-01226-y
- Lau, K. W., Po, L.-M., and Rehman, Y. A. U. (2024). Large separable kernel attention: Rethinking the large kernel attention design in cnn. *Expert Syst. Applications*. 236, 121352. doi: 10.1016/j.eswa.2023.121352
- Li, L., Hassan, M. A., Yang, S., Jing, F., Yang, M., Rasheed, A., et al. (2022a). Development of image-based wheat spike counter through a Faster R-CNN algorithm and application for genetic studies. *Crop J.* 10, 1303–1311. doi: 10.1016/j.cj.2022.07.007
- Li, F., Yan, H., and Shi, L. (2024a). Multi-scale coupled attention for visual object detection. *Sci. Rep.* 14, 11191. doi: 10.1038/s41598-024-60897-8

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Li, M., Zhang, J., Liu, H., Yuan, Y., Li, J., and Zhao, L. (2024b). A lightweight method for apple-on-tree detection based on improved YOLOv5. *SIViP* 18, 6713–6727. doi: 10.1007/s11760-024-03346-3

Li, W., Zhu, T., Li, X., Dong, J., and Liu, J. (2022b). Recommending advanced deep learning models for efficient insect pest detection. *Agriculture* 12, 1065. doi: 10.3390/agriculture12071065

Li, Z., Zhu, Y., Sui, S., Zhao, Y., Liu, P., and Li, X. (2024c). Real-time detection and counting of wheat ears based on improved YOLOv7. *Comput. Electron. Agricult.* 218, 108670. doi: 10.1016/j.compag.2024.108670

Lu, Y., Yu, J., Zhu, X., Zhang, B., and Sun, Z. (2024). YOLOv8-Rice: a rice leaf disease detection model based on YOLOv8. *Paddy Water Environ.* 22, 695–710. doi: 10.1007/s10333-024-00990-w

Ma, M., Li, Y., Xue, C., Xiong, W., Peng, Z., Han, X., et al. (2021). Current situation and key parameters for improving wheat quality in China. *Front. Plant Sci.* 12, 638525. doi: 10.3389/fpls.2021.638525

Ma, N., Su, Y., Yang, L., Li, Z., and Yan, H. (2024). Wheat seed detection and counting method based on improved YOLOv8 model. *Sensors* 24, 1654. doi: 10.3390/s24051654

Ma, N., Zhang, X., Zheng, H.-T., and Sun, J. (2018). “Shufflenet v2: Practical guidelines for efficient cnn architecture design,” in *Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, vol 11218*. eds. V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss. (Cham: Springer). doi: 10.1007/978-3-030-01264-9_8

Misra, T., Arora, A., Marwaha, S., Ranjan Jha, R., Ray, M., Kumar, S., et al. (2022). Yield-SpikeSegNet: An extension of SpikeSegNet deep-learning approach for the yield estimation in the wheat using visual images. *Appl. Artif. Intelligence*. 36, 2137642. doi: 10.1080/08839514.2022.2137642

Peng, Y., Zhao, Y., Yu, Z., Zeng, J., Xu, D., Dong, J., et al. (2022). Wheat quality formation and its regulatory mechanism. *Front. Plant Sci.* 13, 834654. doi: 10.3389/fpls.2022.834654

Qing, S., Qiu, Z., Wang, W., Wang, F., Jin, X., Ji, J., et al. (2024). Improved YOLO-FastestV2 wheat spike detection model based on a multi-stage attention mechanism with a LightFPN detection head. *Front. Plant Sci.* 15, 1411510. doi: 10.3389/fpls.2024.1411510

Que, H., Zhao, X., Sun, X., Zhu, Q., and Huang, M. (2023). Identification of wheat kernel varieties based on hyperspectral imaging technology and grouped convolutional neural network with feature intervals. *Infrared. Phys. Technol.* 131, 104653. doi: 10.1016/j.infrared.2023.104653

Shang, Y., Xu, X., Jiao, Y., Wang, Z., Hua, Z., and Song, H. (2023). Using lightweight deep learning algorithm for real-time detection of apple flowers in natural environments. *Comput. Electron. Agricult.* 207, 107765. doi: 10.1016/j.compag.2023.107765

Shi, L., Wei, Z., You, H., Wang, J., Bai, Z., Yu, H., et al. (2024). OMC-YOLO: A lightweight grading detection method for oyster mushrooms. *Horticulturae* 10, 742. doi: 10.3390/horticulturae10070742

Thakuria, A., and Erkinbaev, C. (2024). Real-time canola damage detection: An end-to-end framework with semi-automatic crusher and lightweight ShuffleNetV2_YOLOv5s. *Smart. Agric. Technol.* 7, 100399. doi: 10.1016/j.atech.2024.100399

- Wan, D., Lu, R., Hu, B., Yin, J., Shen, S., and Lang, X. (2024). YOLO-MIF: Improved YOLOv8 with Multi-Information fusion for object detection in Gray-Scale images. *Adv. Eng. Informatics*. 62, 102709. doi: 10.1016/j.aei.2024.102709
- Wu, H., Wang, Y., Zhao, P., and Qian, M. (2023). Small-target weed-detection model based on YOLO-V4 with improved backbone and neck structures. *Precis. Agricult.* 24, 2149–2170. doi: 10.1007/s11119-023-10035-7
- Wu, W., Yang, T., Li, R., Chen, C., Liu, T., Zhou, K., et al. (2020). Detection and enumeration of wheat grains based on a deep learning method under various scenarios and scales. *J. Integr. Agricult.* 19, 1998–2008. doi: 10.1016/S2095-3119(19)62803-0
- Yao, Q., Zheng, X., Zhou, G., and Zhang, J. (2024). SGR-YOLO: a method for detecting seed germination rate in wild rice. *Front. Plant Sci.* 14, 1305081. doi: 10.3389/fpls.2023.1305081
- Ye, Z., Yang, K., Lin, Y., Guo, S., Sun, Y., Chen, X., et al. (2023). A comparison between Pixel-based deep learning and Object-based image analysis (OBIA) for individual detection of cabbage plants based on UAV Visible-light images. *Comput. Electron. Agricult.* 209, 107822. doi: 10.1016/j.compag.2023.107822
- Zhang, Y., Zhang, H., Huang, Q., Han, Y., and Zhao, M. (2024). DsP-YOLO: An anchor-free network with DsPAN for small object detection of multiscale defects. *Expert Syst. Applications*. 241, 122669. doi: 10.1016/j.eswa.2023.122669
- Zhao, W., Liu, S., Li, X., Han, X., and Yang, H. (2022). Fast and accurate wheat grain quality detection based on improved YOLOv5. *Comput. Electron. Agricult.* 202, 107426. doi: 10.1016/j.compag.2022.107426
- Zhong, C., Wu, H., Jiang, J., Zheng, C., and Song, H. (2024). YOLO-DLHS-P: A lightweight behavior recognition algorithm for captive pigs. *IEEE Access*. 12, 104445–104462. doi: 10.1109/ACCESS.2024.3414859
- Zhou, L., Zhang, C., Taha, M. F., Wei, X., He, Y., Qiu, Z., et al. (2020). Wheat kernel variety identification based on a large near-infrared spectral dataset and a novel deep learning-based feature selection method. *Front. Plant Sci.* 11, 575810. doi: 10.3389/fpls.2020.575810