ACCURATE NON-DESTRUCTIVE TESTING METHOD FOR POTATO SPROUTS FOCUSING ON DEFORMABLE ATTENTION
/
聚焦可变形注意力的马铃薯芽苗精确无损检测方法

Binxuan GENG[1)], Guowei DAI[2)], Huan ZHANG[1,*)], Shengchun QI[1)], Christine DEWI[3)]
[1)] Faculty of Mechanical and Electrical Engineering, Qingdao Agricultural University, Qingdao 266000 / China;
[2)] Agricultural Information Institute of CAAS, Beijing 100081 / China;
[3)] Faculty of Information Technology, Satya Wacana Christian University Salatiga / Indonesia
Tel: 13864215762; E-mail: huan0804@163.com Corresponding author: Zhang Huan
DOI: https://doi.org/10.35633/inmateh-72-36

Keywords: *deep learning, non-destructive inspection, YOLOv8, DAS Attention Mechanism, potato sprouts*

ABSTRACT

Accurate potato sprout detection is the key to automatic seed potato cutting, which is important for potato quality and yield. In this paper, a lightweight DAS-YOLOv8 model is proposed for the potato sprout detection task. By embedding DAS deformable attention in the feature extraction network and the feature fusion network, the global feature context can be efficiently represented and the attention increased to the relevant pixel image region; then, the C2f_Atten module fusing Shuffle attention is designed based on the C2f module to satisfy the attention to the key feature information of the high-level abstract semantics of the feature extraction network. At the same time, the ghost convolution is introduced to improve the C2f module and convolutional module to realize the decomposition of the redundant features to extract the key features. Verified on the collected potato sprout image data set, the average accuracy of the proposed DAS-YOLOv8 model is 94.25%, and the calculation amount is only 7.66 G. Compared with the YOLOv8n model, the accuracy is 2.13% higher, and the average accuracy is 1.55% higher. In comparison to advanced state-of-the-art (SOTA) target detection algorithms, the method in this paper offers a better balance between comprehensive performance and lightweight model design. The improved and optimized DAS-YOLOv8 model can realize the effective detection of potato sprouts, meet the requirements of real-time processing, and can provide theoretical support for the non-destructive detection of sprouts in automatic seed potato cutting.

摘要

马铃薯芽苗准确检测是马铃薯种薯自动切块的关键，对马铃薯的品质和产量具有重要意义。本文提出一种轻量级的DAS-YOLOv8 模型用于马铃薯芽苗检测任务。通过在特征提取网络与特征融合网络嵌入 DAS 可变形注意力，以高效表示全局特征上下文和增加对相关像素图像区域的关注度；然后，基于 C2f 模块设计融合 Shuffle 注意力的 C2f_Atten 模块，以满足特征提取网络高层抽象语义关键特征信息的关注，同时引入幽灵卷积改进 C2f 模块和卷积模块，实现分解冗余特征提取关键特征。在采集到的马铃薯芽苗图像数据集进行验证，拟议 DAS-YOLOv8 模型的平均精度均值为 94.25%，计算量仅为 7.66 G，相比 YOLOv8n 模型，精准率提高 2.13%，平均精度均值提高 1.55%。在先进 SOTA 目标检测算法比较中，本文方法的综合性能更好模型更轻量化。改进优化后的 DAS-YOLOv8 模型能够实现马铃薯芽苗的有效检测，满足实时处理的要求，可为种薯自动切块中的芽苗无损检测提供理论支撑。

INTRODUCTION

Potato is one of the world's major food crops and plays an important role in ensuring global food security and stability. China is the country with the largest area under potato cultivation (31.89% of the global cultivated area) and also the country with the highest potato production (25.09% of the global total production) *(Lun et al., 2023)*. In the potato industry, the level of mechanization has been increasing, including the mechanization of seeding, harvesting, cultivation, and grading of potatoes. However, at present, the cutting of seed potatoes still relies mainly on manual operation, with problems such as high labor intensity, low efficiency, and high costs. With the rise of labor costs and the reduction of the labor force, there is an urgent need to solve the problem of automatic seed potato cutting *(Danielak et al., 2023)*. Among them, accurate detection of potato sprouts is the key to realizing automatic seed potato cutting.

With the continuous progress and development of computer technology, computer vision technology has gradually been widely used in the field of agriculture. However, the research on potato sprout recognition is more limited.

Non-destructive testing (NDT) is usually based on RGB imaging systems and hyperspectral/multispectral imaging systems is used to obtain images of the surface of agricultural products. In terms of traditional machine vision methods, *Li et al., (2018),* proposed a potato sprout eye recognition method based on three-dimensional geometric features of color saturation. This method has a sprout eye recognition rate of 91.48% and takes an average of 2.68 seconds to identify a single image. Gao et al. *(Gao, 2022)* trained SVM classifiers based on extracting B and H components in RGB color space and HSV color space and used weighted Euclidean distance and morphological methods to detect and label potato sprouting loci with an average recognition rate of 90.6%. *Lu et al., (2021),* performed potato image filtering based on Gabor features and selected filtered images under orientation two and scale four for morphological image processing, and the proposed algorithm's bud eye recognition rate was 93.4%. *Dhulipalla Ravindra Babu et al., (2023)* calculated four parameters of variance, correlation, homogeneity, and uniformity using gray level co-occurrence matrix on the collected potato dataset, and the classification accuracy by support vector machine was 99.5%.

The above methods need to design and extract image features manually and usually use low-level features such as edges, textures, colors, etc.; feature extraction has blindness and uncertainty and is less adaptable to complex scenes *(Su & Xue, 2021)*. With the rapid development of computer technology, especially deep learning, in recent years, the related technology has been successfully applied to the field of agriculture. Compared with the traditional methods, deep learning methods have more powerful feature learning and modeling capabilities in the target detection task, and can deal with more complex scenes and target variations, as well as possessing higher accuracy and generalization capabilities. *Zhang et al., (2022),* achieved an accuracy rate of 88.33% for seed potato sprout eye detection based on the YOLOv3-tiny network using the CIoU border regression loss function with the K-means clustering method. *Yang et al., (2021),* based on multispectral images combined with a Supervised Multi-Threshold Segmentation Model (SMTSM) and Canny edge detector in order to achieve 89.67% accuracy in seed potato sprout detection. Wang et al. *(Wang C. & Xiao, 2021)* used the convolutional layer of ResNet101 as the base network structure to improve Faster RCNN, which achieved 98.7% accuracy in detecting surface defects on potatoes. *Dai et al., (2022),* enhanced the feature similarity problem of the feature representation fusion process by replacing the Conv of the C3 module in YOLOv5 with CrossConv. They used the 9-Mosaic data augmentation algorithm to improve the model generalization ability, and the accuracy of the improved model for potato germination recognition in complex scenarios was 90.14%.

For these reasons, this work proposes a lightweight method for potato sprout detection. By collecting potato sprout-related picture information, using DAS deformable attention in the backbone network and neck network to deepen the information interaction of semantic features at the abstraction level, and designing a C2f module incorporating Shuffle Attention for automatic feature extraction in the backbone network, and secondly, using Ghost Convolution to simplify the model computation and ensure the model's detection accuracy and lightness, a better sprout detection accuracy was obtained in the potato sprouts dataset.

MATERIALS AND METHODS

Image Acquisition Platform

The image acquisition system is shown in Fig. 1 and includes a cell phone, LED ring light, stepper motor, potato base, shade cloth, reflector, and computer.

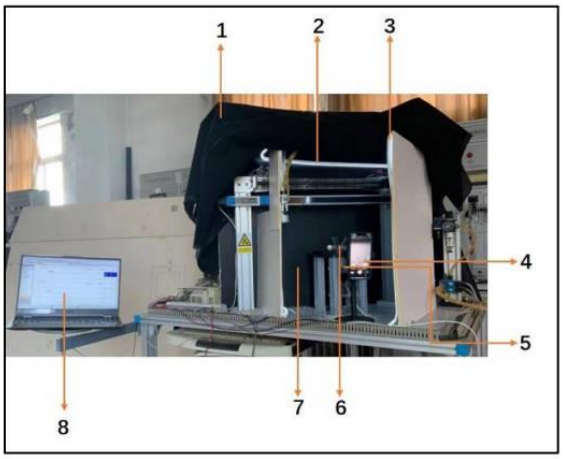

Fig. 1 - Image Acquisition Platform for Potato Sprouts

1. Shade cloth; 2,3. LED strip light; 4. Shooting phone; 5. Stepping motor; 6. Potato base; 7. Reflector; 8. Computer

The stepper motor step angle was 0.9 degrees. The sprouted potato seed potato was placed into the potato base located above the stepper motor, after which the stepper motor began to rotate. Image acquisition was performed at every 120° of rotation so that a complete sprouted potato was identified from three images. In addition, the position of the camera used for image acquisition must be at the same horizontal height as the sprouted potatoes in order to prevent the emergence of the situation of top-down or upward view angle, which may cause visual difference and, thus, data errors in the subsequent work.

Dataset Construction

The experimental sample is 200 potatoes, and a total of 600 images are collected. In the deep learning task, in order to improve the generalization ability of the model, prevent overfitting, enhance the robustness of the model, and at the same time simulate the various changes in the real world to adapt to different scenarios, it is necessary to augment the collected images with data. By applying various transformations to the image, the dataset used for training can be expanded so that the model can recognize more features and patterns during the learning process, resulting in higher accuracy and stability in practical applications. In this deep learning task, it was performed image vertical flip and horizontal flip, increased image contrast, increased image brightness, added motion blur noise, splash transform, and pixel transform to the image, and added digital noise to the acquired potato images, using all of the above for data augmentation. The results of the data augmentation are detailed in Fig. 2.

After data augmentation, the dataset contains a total of 2000 images. The images were labeled using the automatic labeling X-AnyLabeling tool and using the standard format of the YOLO dataset, which is suitable for the YOLO family of neural network detection models. From these 2000 images, 1500 images were randomly selected as the training set, 300 images as the validation set, and 200 images as the test set, with no overlap between the three. The potato sprout dataset (PSD) included in this paper is divided into the dataset PSD-One of the original 600 images and the dataset PSD-Two of 2000 images.

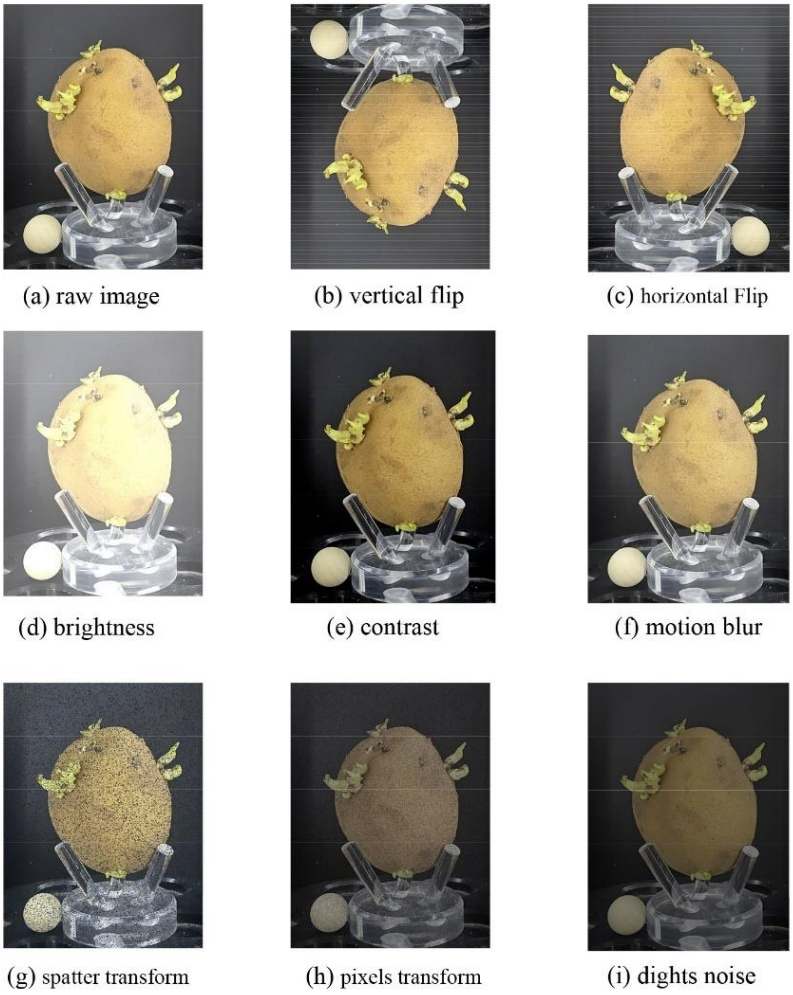

Fig. 2 - Results of image data augmentation of potato sprouts

DAS Attention Mechanism

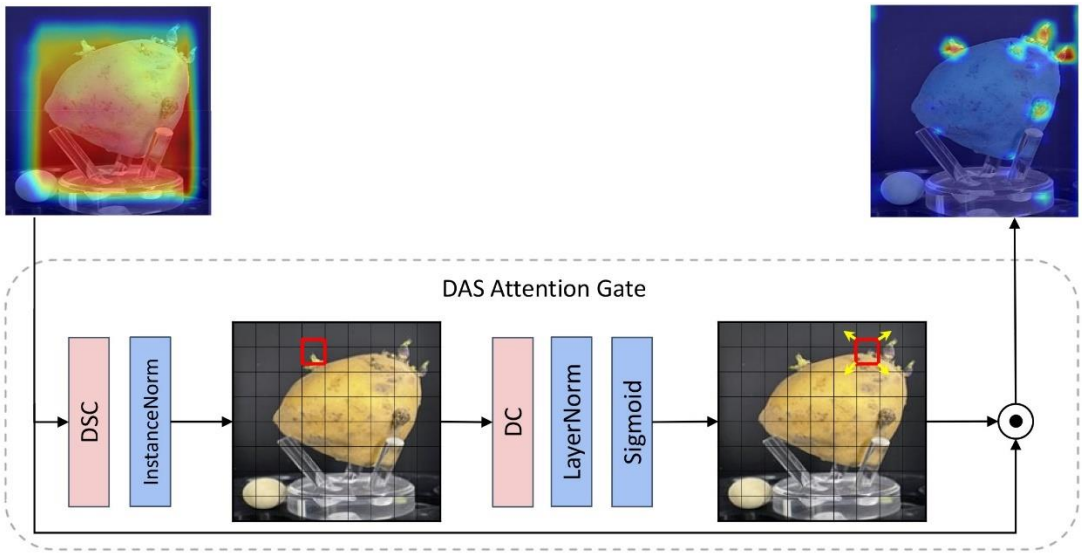

Fig. 3 - The heatmap shows the comparison results of saliency maps with the addition of the DAS attention mechanism, which integrates deeply separable convolution (DSC) with deformable convolution (DC) to focus attention on salient regions

Deformable Attention DAS is a fast and simple fully convolutional method that helps to focus attention on relevant information (Fig. 3). The DAS consists of depth separable convolution (DSC) with deformable convolution (DC). DSC operations are used in the bottleneck layer to improve computational efficiency by reducing the number of channels in the feature map. The channel is reduced from c to $a \times c$, the only hyperparameter ranges from 0 < α < 1 and is insensitive for 0.1 < α. The instance normalization InstanceNorm process is performed after the bottleneck layer to achieve the removal of instance-specific contrast information from the image, thus improving the robustness of the DAS model during training, followed by the addition of a nonlinear activation function, GELU, to enhance the representation of the features as shown in equation (1), where $\mathbf{X}$ is the input feature, $\mathbf{W}_1$ is the depth-separable convolution, and $\mathbf{X}_c$ denotes the feature context.

$$\mathbf{X}_c = \text{GELU}(\text{InstanceNorm}(\mathbf{X}\mathbf{W}_1)) \tag{1}$$

The above feature context is compressed; equation (2) uses deformable convolution of a dynamic grid to focus on the image region of interest, K is the size of the convolution kernel, w_k is the weight parameter, $p_{ref,k}$ is a fixed reference point, and the trainable parameters, w_p and Δp_k, depend on the specific features of the kernel function being applied. Next, as in equation (3), the number of channels of the feature map is converted to the original input c, using the layer normalized LayerNorm with Sigmoid activation function σ operation. The information flow of the feature map is controlled, and it is decided that part of the feature map is emphasized or filtered. The output of equation (4) is the input of the next layer of the CNN model, $\mathbf{X}$ is the upper layer feature, and $\odot$ denotes the dot product.

$$deform(p) = \sum_{k=1}^{K} w_k \cdot w_p \cdot \mathbf{X}(p_{ref,k} + \Delta p_k) \tag{2}$$

$$\mathbf{A} = \sigma\left(\text{LayerNorm}\left(deform(\mathbf{X}_c)\right)\right) \tag{3}$$

$$\mathbf{X}_{\mathbf{out}} = \mathbf{X} \odot \mathbf{A} \tag{4}$$

C2f_Atten Module

The Shuffle Attention (SA) module combines group convolution, spatial attention mechanisms, and channel attention mechanisms *(Hao et al., 2023; Song et al., 2024)*. Not only is it possible to take the information between different channels, but the amount of computation is also reduced, which helps to achieve the precise location and recognition of the target. Shuffle Attention contrasts with the traditional attention mechanism by disrupting and reordering the input sequence, thereby enhancing the model's ability to generalize while maintaining computational efficiency. Additionally, the importance of each position is determined within this process. SA Attention Module is shown in Fig. 4.

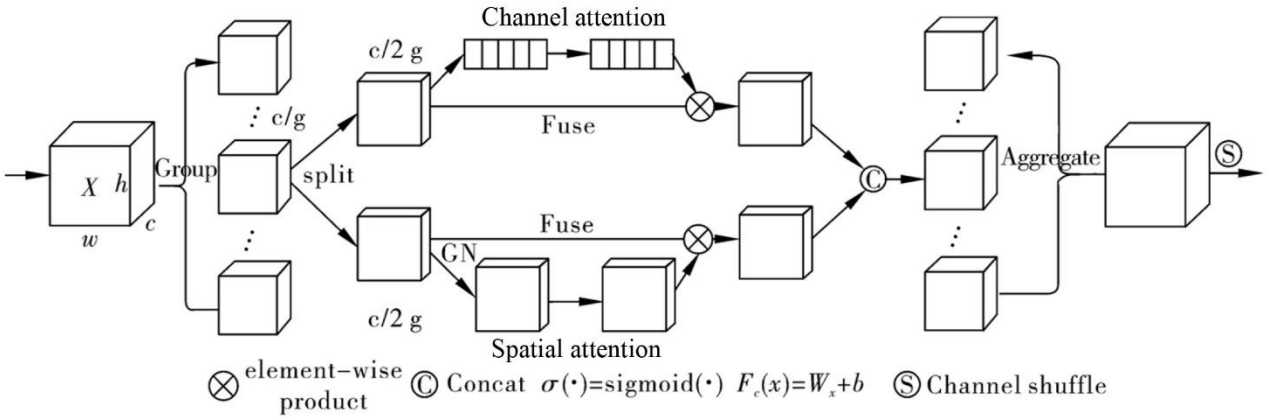

Fig. 4 - General Architecture of the Shuffle-Attention Module

The main process is that for the input feature map as $X \in R^{C\times H\times W}$ divided into g groups $X = 2[X_1,\cdots,X_G], R^{C/G\times H\times W}$, along the channel dimensions, the feature X will be split into two branches as $X_{k1}, X_{k2} \in R^{C/2G\times H\times W}$, which will be used to learn the channel attention features and spatial attention features respectively. For each group of features, the information within the group is fused by means of Concat to generate different importance coefficients. The upper branch channel attention mechanism uses GAP, a combination of scaling factor and Sigmoid function; W_1 and b_1 are two dynamic parameters of the scaled feature map; the specific process is described as follows:

$$s = F_{gp}(X_{k1}) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{k1}(i,j), \tag{5}$$

$$X'_{k1} = \sigma(F_c(s)) \times X_{k1} = \sigma(W_1 s + b_1) \times X_{k1}, \tag{6}$$

The branch below SA is the spatial attention mechanism, as shown in equation (7).

GroupNorm (GN) is used to process X_{k2} to get the statistical information at the spatial domain level, and then the $F_c(\ )$ operation is used. The Channel Shuffle operation is used to enhance the interaction and information transfer to the group's features. The Concat fusion is used to get $X_k = concat[X'_{k1}, X'_{k2}]$ to realize the information circulation between different groups.

$$X'_{k2} = \sigma(W_2 \times GN(X_{k2}) + b_2) \times X_{k2}, \tag{7}$$

The C2f_Atten module integrates the SA module on top of the C2f module, i.e., it realizes the addition of the attention mechanism to the C2f module. SA is located after the second Conv module, which makes the model more concerned about the location information of the target region in order to improve the accuracy of the detection of the target region, and the details of the code implementation are shown in Table 1.

Table 1

C2f_Atten module specific algorithm implementation details

```
class C2f_Attention(nn.Module):
  def __init__(self, c1, c2, n=1, shortcut=False, g=1, e=0.5):
    super().__init__()
    self.c = int(c2 * e)
    self.cv1 = Conv(c1, 2 * self.c, 1, 1)
    self.cv2 = Conv((2 + n) * self.c, c2, 1)
    self.attention = ShuffleAttention(c2)
    self.m = nn.ModuleList(Bottleneck(self.c, self.c, shortcut, g,
k=((3, 3), (3, 3)), e=1.0) for _ in range(n))
  def forward(self, x):
    y = list(self.cv1(x).chunk(2, 1))
    y.extend(m(y[-1]) for m in self.m)
```

```
    return self.attention(self.cv2(torch.cat(y, 1)))
  def forward_split(self, x):
    y = list(self.cv1(x).split((self.c, self.c), 1))
    y.extend(m(y[-1]) for m in self.m)
    return self.cv2(torch.cat(y, 1))
```

DAS-YOLOv8 potato sprout detection model

YOLOv8 is the SOTA target detection model of the current YOLO series of networks, which has the characteristics of a lightweight and wide adaptation range *(Ma et al., 2023; Z. Wang et al., 2024)*. In this paper, the YOLOv8n network model was taken as the basic framework, and the structure of the improved model is shown in Fig. 5, which mainly includes the input, backbone network, neck network, and head module.

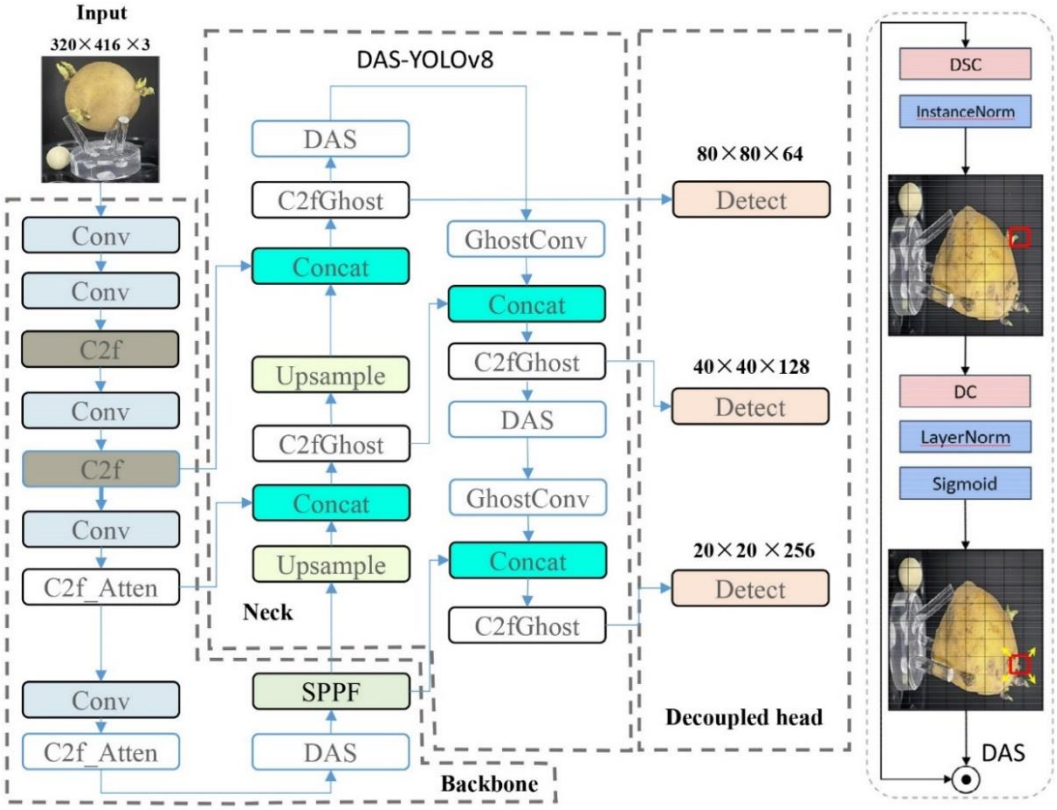

Fig. 5 - The general architecture of DAS-YOLOv8, consists of three components: the Backbone for feature extraction, the Neck for feature connectivity and the Head for output

Input side: By using Mosaic data augmentation to perform operations such as flipping (flipping left and right of the original picture), scaling (scaling the size of the original picture), and color gamut transformation (changing the brightness, saturation, and hue of the original picture) on the random pictures, and then combining the pictures, the model can be better adapted to the complex scenes in the real world, so that it has better robustness and generalization ability, and then improve the model's ability to detect markers in complex backgrounds. Note that by default 10 Epochs before the end of training will automatically turn off Mosaic data augmentation.

Backbone network: The backbone network in YOLOv8 acts as a feature extractor, which mainly consists of the Conv module, C2f module, and SPPF module. Its main function is to extract features from the input image, which are used for subsequent tasks. In this paper, the penultimate two C2f modules of the backbone network are replaced with C2f_Atten to obtain the key factor information of highly abstract semantic features in the feature extraction layer. DAS-intensive attention is applied before the SPPF module, and the feature context is considered comprehensively. Through a series of convolution and pooling operations, the backbone network gradually reduces the spatial dimensions of the feature maps while increasing the depth of these feature maps. This process allows the neural network to capture and represent features of varying complexity and scale in the input image.

Neck network: used to process further the feature maps extracted by the backbone network. Its main purpose is to integrate different levels of feature information to improve the performance of target detection. The main improvements in designing the model include reducing the number of parameters of the model by replacing the Conv module with GhostConv to share parameters between channels. C2fGhost, on the other hand, changes the structure of the C2f module by replacing all of the Convs contained in the C2f module with GhostConvs, and by combining a small number of actual filters with linear combinations of these filters, this helps to reduce model complexity without sacrificing accuracy *(Wang Y. et al., 2024)*. The introduction of GhostConv with C2fGhost aims to achieve a balance between model performance and efficiency. DAS is adopted after the second upsampling to solve the problem of the possible loss of primary semantic features of the backbone network. The subsequent DAS can take over the mixed semantic feature information from the C2f_Atten module and the DAS of the backbone network to better utilize, in addition to the basic feature information of the backbone network attention information.

Head module: plays a crucial role in the target detection task, and its main responsibility is to generate bounding boxes, classification probabilities, and target attributes. This module is designed to extract information related to detected object location and classification labels from the feature map. The generated bounding boxes are processed with non-extremely large value suppression to obtain the final target detection results. In the design phase, the feature information sources of multiple Detect modules are decentralized in order to avoid focusing all the attention on single-scale feature information.

Experimental platforms and performance indicators

The desktop computer used for the experiment has an Intel Core i7-12700F processor, 32GB of RAM, and is equipped with an NVIDIA GeForce RTX 3090 GPU to accelerate the experimental process, which has 24G of video memory.

The experiments were run on Windows 11 (64-bit) operating system, VS2015 version, Python version 3.9.13, using PyTorch (version 1.13.1) as the framework for deep learning, equipped with CUDA version 11.7 parallel computer architecture with cuDNN version 8.6 deep neural network acceleration library. The batch size for network training is 32, and the optimizer selects SGD. By default, the input size of the original image is resized to 320 × 416. In addition, the training process supervises the model outputs using early stopping, which is effective in preventing overfitting.

In order to evaluate the detection performance of the proposed model for potato sprouts, equations (8-11), i.e., Precision, Recall with mean average precision mean (mAP, IoU threshold is taken as 0.5), were used to evaluate the complexity of the algorithm using the model's floating point calculations (FLOPs) as an important indicator of the complexity of the algorithm.

Where TP (True Positive): predicted to be a sprout and actually a sprout, i.e., the number of correctly identified potato sprouts; FP (False Positive): predicted to be a sprout and actually not a sprout, i.e., the number of incorrectly identified potato sprouts; FN (False Negative): predicted not to be a sprout and actually a sprout, i.e., the number of omitted potato sprouts; and TN (True Negative): predicted to be not a sprout and actually not a sprout, i.e., the number of correctly identified non-potato sprouts.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{AP} = \frac{1}{n}\sum_{i=1}^{n} \text{Precision}_i \tag{10}$$

$$= \frac{1}{n}\text{Precision}_1 + \frac{1}{n}\text{Precision}_2 + \cdots + \frac{1}{n}\text{Precision}_n$$

$$\text{mAP} = \frac{\sum_{i=1}^{Q} AP_i}{Q} \tag{11}$$

RESULTS

Analysis of potato sprout detection performance

In this section, the performance of the proposed DAS-YOLOv8 model will be validated from various aspects. As shown in Table 2, the DAS-YOLOv8 model improves the precision rate by 2.13%, the recall rate by 2.55%, and the average precision mean by 1.55% compared to the unimproved YOLOv8n model. Notably, the FLOPs of the DAS-YOLOv8 model are only 7.66, which is 0.94% lower compared to YOLOv8n, proving the property of model lightness.

As shown in Fig. 6, 60 images were randomly selected from the potato sprout test set of 200 images for testing, and the random image selection process was performed four times.

Table 2

Comparative results of DAS-YOLOv8 model improvement experiments

#	Model	Datasets	Precision (%)	Recall (%)	mAP (%)	FLOPs (G)
1	YOLOv8n	PSD-Two	91.11	84.45	92.70	8.60
2	DAS-YOLOv8	PSD-Two	93.24	87.68	94.25	7.66

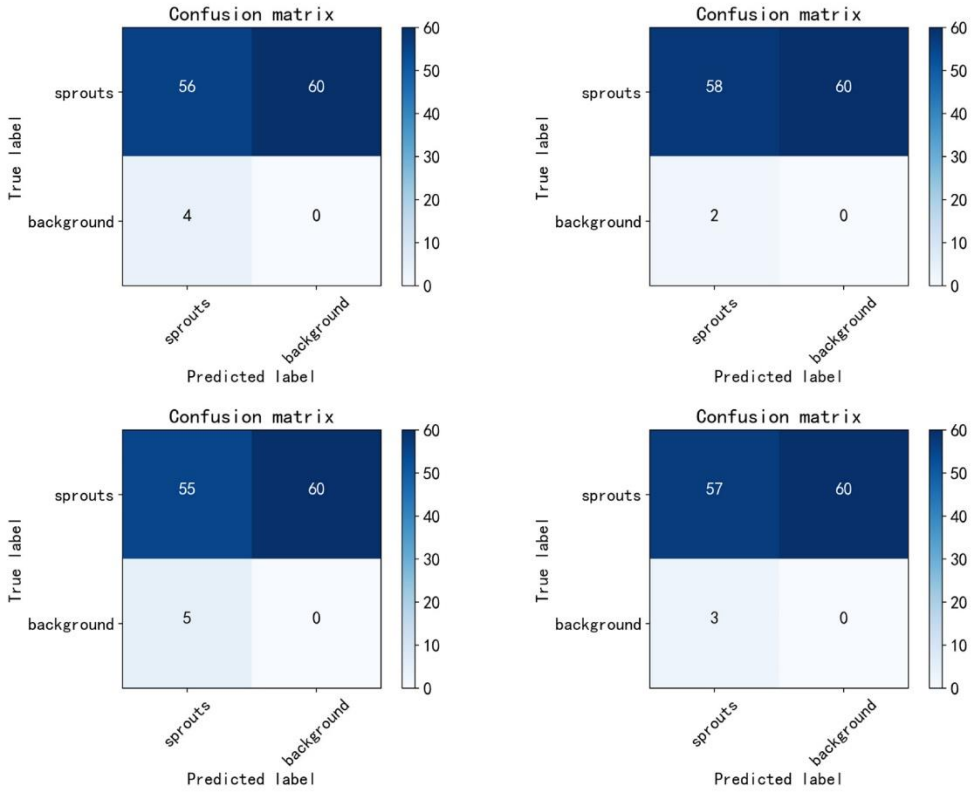

Fig. 6 - Confusion Matrix for Test Results of Randomly Selected Samples from the Potato Sprout Test Set

C2f_Atten module ablation experiment

In this section, the performance of the proposed C2f_Atten module will be verified for potato sprout detection in the DAS-YOLOv8 model. The C2f_Atten module is employed in the backbone network of the DAS-YOLOv8 model, which improves the C2f module. Based on the original C2f module code, any intrusive code revisions were divided into three main parts. As shown in Table 1, part 1 (SA-1) uses the SA attention mechanism at the C2f module code cv1, part 2 (SA-2) uses the SA attention mechanism at the C2f module code cv2, and part 3 (SA-3) uses the SA attention mechanism at the C2f module code m (bottleneck network). The experimental results, shown in Table 3, show that the SA-2 attention combination used in the proposed DAS-YOLOv8 model achieves the best performance, with an average precision mean of 94.25%, and precision and recall rates of 93.24% and 87.68%, respectively. SA-3 improved the accuracy rate by 0.12% compared to SA-2, but the FLOPs increased by 0.04%. The mean average precision value of SA-3 was lower than that of SA-2, which was 94.12%. The mean average precision value of SA-1 was 93.98%, and it had the worst performance in the performance of the ablation test of the C2f_Atten module, but the model complexity was the same. In summary, the DAS-YOLOv8 modeling submodule for intrusive code improvement can effectively improve the model detection performance.

Table 3

Results of the C2f_Atten module ablation experiment

#	Model	C2f_Atten	Datasets	Precision (%)	Recall (%)	mAP (%)	FLOPs (G)
1	DAS-YOLOv8	SA-1	PSD-Two	92.59	87.24	93.98	7.66
2	DAS-YOLOv8	SA-2	PSD-Two	93.24	87.68	94.25	7.66
4	DAS-YOLOv8	SA-3	PSD-Two	93.36	86.53	94.12	7.70

DAS-YOLOv8 ablation experiment

In this section, the performance differences between the C2f_Atten, DAS, C2fChost, and GhostConv modules for the DAS-YOLOv8 model are analyzed. Table 4 shows that the mean average precision value of the C2f_Atten module embedded in the backbone network is 93.15%, the mean average precision value of the DAS module embedded in the backbone network with a necking network is 94.58%, and the performance enhancement advantage of the DAS module is some obvious. C2f_Atten was used in fusion with the DAS module for a mean average precision of 94.67%. The C2fChost and GhostConv modules were mainly used to lighten the DAS-YOLOv8 model, and as can be seen in Table 4, test numbers 4-6, the C2fChost and GhostConv modules resulted in a reduction of the mean average precision by 0.42%. However, model complexity FLOPs were reduced by 16.19%. Compared to Trial No. 1, Trial No. 7 used C2f_Atten with DAS module for model lightening, and as a result, the mean average accuracy was reduced by 1.48%, which is a significant advantage over not using C2f_Atten with DAS module. From the point of view of analyzing the dataset, the DAS module and the C2f_Atten module, which contains the SA attention mechanism, demonstrate the sophistication of the DAS-YOLOv8 model.

Table 4

Ablation experiments

No.	Embedded model method				mAP (%)	FLOPs (G)
	C2f_Atten	DAS	C2fChost	ChostConv		
1	-	-	-	-	92.70	8.60
2	+	-	-	-	93.15	8.68
3	-	+	-	-	94.58	8.96
4	+	+	-	-	94.67	9.14
5	+	+	+	-	94.31	7.92
6	+	+	+	+	94.25	7.66
7	-	-	+	+	91.22	7.20

SOTA algorithm comparison experiment

In order to further verify the effectiveness and advancement of improving the potato sprout detection algorithm proposed in this paper, current mainstream target detection algorithms such as YOLOv6n (*Bist et al., 2023*), YOLOv5s (*Khalid et al., 2023*) YOLOv7 (*Guo et al., 2023*), and YOLOv8s are used to train and test the model under the same conditions. The experimental results are shown in Table 5. From the data in the table, compared with YOLOv8s, this paper's method is 0.44% behind in mAP. However, the model complexity is 70.40% lower than it, indicating that the comprehensive performance of this paper's method is better. The model is more lightweight, which is suitable for the practical application of agricultural equipment engineering. Compared to other algorithms, there are different degrees of advantages in terms of detection accuracy, mean average precision, and computation. In addition, the mean average precision value of the proposed model in the PSD-One dataset is 82.69%, which is lower than the mean average precision value of 94.25% in the PSD-Two dataset. The accuracy gap of the same model illustrates the practical value of employing data augmentation for potato sprout images, which better improves the performance of the model.

To further illustrate the effectiveness of the method in this paper, 16 test result visualizations were randomly extracted from the PSD-Two test set for subjective analysis, and the test results are shown in Fig. 7. From the figure, it can be seen that the proposed DAS-YOLOv8 model accurately detects potato sprouts at different types of locations with the ability to detect potato sprouts under certain disturbing conditions because of better robustness due to the pre-processing of the dataset that innately takes into account the instability of the environmental factors under the Agricultural Equipment Engineering.

Table 5

Performance comparison results of different SOTA algorithms

Algorithm	Backbone	Datasets	Precision (%)	Recall (%)	mAP (%)	FLOPs (G)
YOLOv6n	CSPDarknet53	PSD-Two	93.32	87.77	94.11	11.35
YOLOv5s	CSPDarknet53	PSD-Two	91.74	86.36	92.88	15.80
YOLOv7	CSPDarknet53	PSD-Two	92.98	87.55	93.42	13.32
YOLOv8s	CSPDarknet53	PSD-Two	94.26	89.13	94.69	26.6
YOLOv8n	CSPDarknet53	PSD-Two	91.11	84.45	92.70	8.60
DAS-YOLOv8	Ours	PSD-Two	93.24	87.68	94.25	7.66
DAS-YOLOv8	Ours	PSD-One	78.43	69.11	82.69	7.66

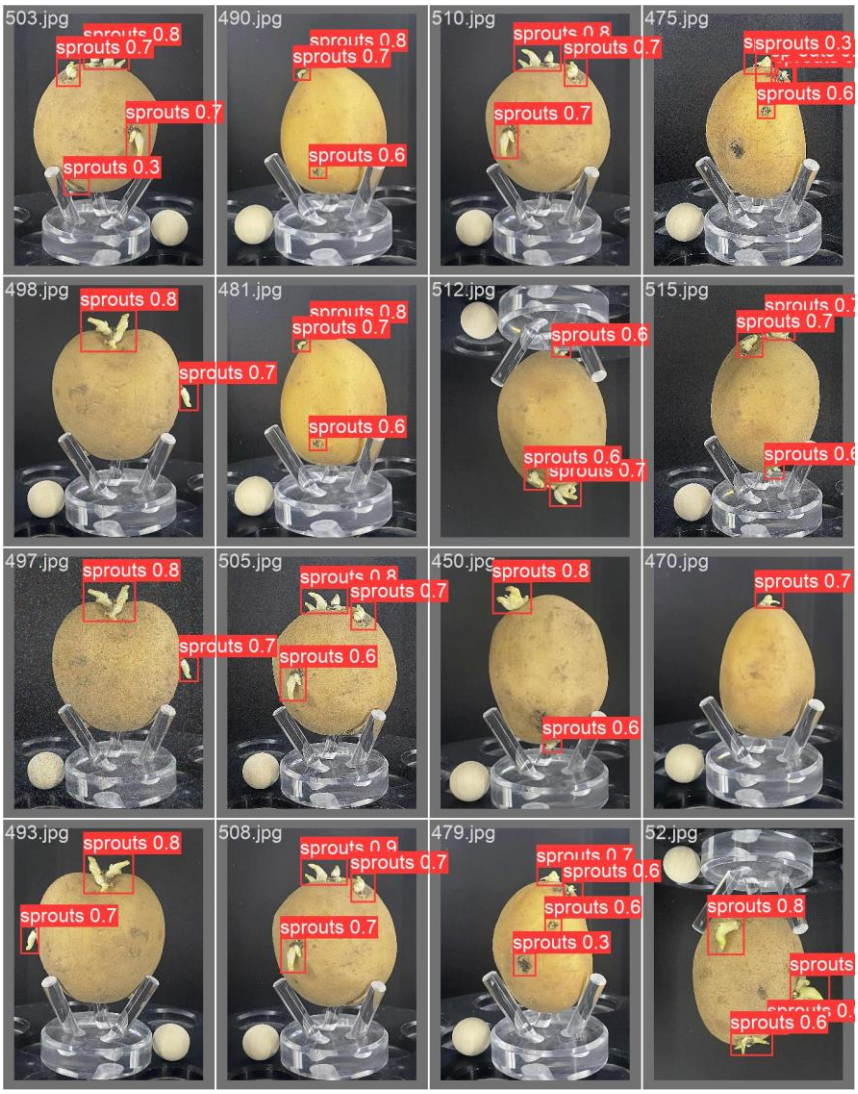

Fig. 7 - Plot of detection results of potato sprout test set in DAS-YOLOv8 network

CONCLUSIONS

In this paper, a target detection model based on deformable attention DAS fusion backbone network SA attention mechanism, called DAS-YOLOv8, is proposed based on YOLOv8 algorithm, which is mainly used for the target detection task of potato sprout images in agricultural scenarios. By designing the DAS attention mechanism embedded after the C2f module of the backbone network and the neck network to avoid the loss of accuracy caused by the loss of feature information about the feature maps of the model during the down sampling process, the context-aware attention mechanism delivered by the neck network taking over from the backbone network is upgraded by one level in terms of the degree of attention to the relevant information. In addition, the design and replacement of the C2f module of the backbone network embed the SA attention mechanism to make the model more focused and further improve the model performance by concentrating on

key information as the feature abstraction level increases with the increase of convolution depth. Meanwhile, the C2fChost convolution module and GhostConv module are introduced to construct a lightweight neck feature fusion module, which reduces the number of parameters and computation of the model while ensuring accuracy. The experiment was tested and compared on the potato sprout PSD-Two dataset, and the feasibility of each advanced improvement was demonstrated through experimental analysis. Compared to the original YOLOv8n model, the proposed model has a mean increase in average accuracy of 1.55%, an increase in precision rate of 2.13%, and FLOPs of only 7.66, which is a reduction of 0.94% compared to YOLOv8n, and has a significant advantage compared to the other mainstream target detection algorithms in all aspects.

The DAS-YOLOv8 model proposed in this paper updates the industry challenge of potato sprout detection and identification in the field of agricultural information engineering and is particularly informative in dealing with the need for lightweight target detection models and intensive detection tasks in the agricultural field. The next plan is how to maintain the accuracy while further reducing the model computation and deploying it in embedded devices to allocate and utilize the resources more efficiently.

ACKNOWLEDGEMENT

This work was supported by the National Agriculture Science Data Center. The authors would also like to thank all authors and anonymous reviewers cited in this paper for their helpful comments and suggestions.

REFERENCES

[1] Bist, R. B., Subedi, S., Yang, X., & Chai, L. (2023). A Novel YOLOv6 Object Detector for Monitoring Piling Behavior of Cage-Free Laying Hens. *AgriEngineering*, *5*(2), Article 2. https://doi.org/10.3390/agriengineering5020056

[2] Dai, G., Hu, L., Fan, J., Yan, S., & Li, R. (2022). A Deep Learning-Based Object Detection Scheme by Improving YOLOv5 for Sprouted Potatoes Datasets. *IEEE Access*, 10, 85416–85428. https://doi.org/10.1109/ACCESS.2022.3192406

[3] Danielak, M., Przybył, K., & Koszela, K. (2023). The Need for Machines for the Nondestructive Quality Assessment of Potatoes with the Use of Artificial Intelligence Methods and Imaging Techniques. *Sensors*, 23(4), Article 4. https://doi.org/10.3390/s23041787

[4] Dhulipalla Ravindra Babu, R. C. Verma, Navneet Kumar Agrawal, & Isha Suwalk. (2023). Classification of Defects in Potato Using Grey Level Co-Occurrence Matrix and Support Vector Machine. Journal of Agricultural Engineering (India), 60(2), 165–177. https://doi.org/10.52151/jae2023602.1805

[5] Gao, S. (2022). Research on detection method of sprouted potato based on SVM and weighted Euclidean distance. *6th International Conference on Mechatronics and Intelligent Robotics (ICMIR2022)*, 12301, 719–725. https://doi.org/10.1117/12.2644666

[6] Guo, J., Yang, Y., Lin, X., Memon, M. S., Liu, W., Zhang, M., & Sun, E. (2023). Revolutionizing Agriculture: Real-Time Ripe Tomato Detection With the Enhanced Tomato-YOLOv7 System. *IEEE Access*, 11, 133086–133098. https://doi.org/10.1109/ACCESS.2023.3336562

[7] Hao, W., Zhang, L., Han, M., Zhang, K., Li, F., Yang, G., & Liu, Z. (2023). YOLOv5-SA-FC: A Novel Pig Detection and Counting Method Based on Shuffle Attention and Focal Complete Intersection over Union. *Animals*, *13*(20), Article 20. https://doi.org/10.3390/ani13203201

[8] Khalid, M., Sarfraz, M. S., Iqbal, U., Aftab, M. U., Niedbała, G., & Rauf, H. T. (2023). Real-Time Plant Health Detection Using Deep Convolutional Neural Networks. *Agriculture*, *13*(2), Article 2. https://doi.org/10.3390/agriculture13020510

[9] Lun, R., Luo, Q., Gao, M., Li, G., & Wei, T. (2023). How to Break the Bottleneck of Potato Production Sustainable Growth—A Survey from Potato Main Producing Areas in China. *Sustainability*, *15*(16), Article 16. https://doi.org/10.3390/su151612416

[10] Li Y., Li T., Niu Z. Wu Y., Zhang Z., Hou J. (2018). Potato bud eyes recognition based on three-dimensional geometric features of color saturation (基于色饱和度三维几何特征的马铃薯芽眼识别). *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 34(24): 158-164, Taian / China https://doi.org/10.11975/j.issn.1002-6819.2018.24.019

[11] Lu Z., Qi X., Zhang W., Liu Z., Zheng W., Mu G. (2021). Study on Mechanical Properties and Finite Element Analysis of Seed Cucurbita (基于 Gabor 特征的马铃薯图像芽眼识别). *Journal of Agricultural Mechanization Research*, 43(02), 203–207, Taian / China. https://doi.org/10.13427/j.cnki.njyi.2021.02.036

[12] Ma, N., Li, Y., Xu, M., & Yan, H. (2023). IMPROVED YOLOv8-BASED AUTOMATED DETECTION OF WHEAT LEAF DISEASES. *INMATEH Agricultural Engineering*, 499–510. https://doi.org/10.35633/inmateh-71-43

[13] Su, W.-H., & Xue, H. (2021). Imaging Spectroscopy and Machine Learning for Intelligent Determination of Potato and Sweet Potato Quality. Foods, 10(9), Article 9. https://doi.org/10.3390/foods10092146

[14] Song, X., Li, H., Liang, L., Shi, W., Xie, G., Lu, X., & Hei, X. (2024). TransBoNet: Learning camera localization with Transformer Bottleneck and Attention. *Pattern Recognition*, *146*, 109975. https://doi.org/10.1016/j.patcog.2023.109975

[15] Wang, C., & Xiao, Z. (2021). Potato Surface Defect Detection Based on Deep Transfer Learning. *Agriculture*, *11*(9), Article 9. https://doi.org/10.3390/agriculture11090863

[16] Wang, Y., Zhang, C., Wang, Z., Liu, M., Zhou, D., & Li, J. (2024). Application of lightweight YOLOv5 for walnut kernel grade classification and endogenous foreign body detection. *Journal of Food Composition and Analysis*, *127*, 105964. https://doi.org/10.1016/j.jfca.2023.105964

[17] Wang, Z., Hua, Z., Wen, Y., Zhang, S., Xu, X., & Song, H. (2024). E-YOLO: Recognition of estrus cow based on improved YOLOv8n model. *Expert Systems with Applications*, *238*, 122212. https://doi.org/10.1016/j.eswa.2023.122212

[18] Yang, Y., Zhao, X., Huang, M., Wang, X., & Zhu, Q. (2021). Multispectral image based germination detection of potato by using supervised multiple threshold segmentation model and Canny edge detector. *Computers and Electronics in Agriculture*, *182*, 106041. https://doi.org/10.1016/j.compag.2021.106041

[19] Zhang, W., Han, Y., Huang, C., & Chen, Z. (2022). Recognition method for seed potato buds based on improved YOLOv3-TINY. *INMATEH Agricultural Engineering*, 364–373. https://doi.org/10.35633/inmateh-67-37