European Journal of Agronomy 157 (2024) 127191

Contents lists available at ScienceDirect

European Journal of Agronomy

journal homepage: www.elsevier.com/locate/eja

TSP-yolo-based deep learning method for monitoring cabbage seedling emergence

Xin Chen [1], Teng Liu [1], Kang Han, Xiaojun Jin [*], Jinxu Wang, Xiaotong Kong, Jialin Yu [*]

Peking University Institute of Advanced Agricultural Sciences / Shandong Laboratory of Advanced Agricultural Sciences at Weifang, Weifang, Shandong 261325, China

ARTICLE INFO

Keywords:
Crop seedling counting
Object detection
Object tracking
UAV
YOLOv8

ABSTRACT

Real-time monitoring of seedling emergence is vital for vegetable crop management and yield estimation. Traditionally, crop seedling emergence monitoring relies on low-efficient and time-consuming manual counting. To address this issue, this research proposed an efficient, fast, and real-time cabbage seedling counting method (combining the improved YOLOv8n, tracking algorithm, and image processing) to accurately track cabbage seedlings in the field and implement counting with an unmanned aerial vehicle (UAV). The improved YOLOv8n replaced the C2f Block in the YOLO backbone with a Swin-conv block and incorporated ParNet attention modules in both the backbone and neck parts. This enhancement enables the YOLOv8n to surpass the base model's performance, achieving a mAP50–95 of 90.3 %, representing a 14.5 % improvement. The experiments demonstrated the superior capabilities of the counting method in terms of speed and accuracy. In field experiments, the proposed Tracking algorithms-Swin-conv blocks-ParNet attention-YOLOv8n (TSP-yolo) counting method demonstrated consistent and reliable accuracy in counting cabbage seedlings while demanding only one-seventh of the time needed compared to the manual counting method. In summary, based on TSP-yolo and implemented through an UAV, the developed seedling emergence counting method demonstrated an excellent capability of counting cabbage seedlings, resulting in significant savings in human resources for crop management.

1. Introduction

High-quality seeds contribute significantly to superior crop quality and yield (Matthews et al., 2012). The emergence of seedlings is a crucial phenological occurrence affecting the success of an annual crop. It marks the moment when a seedling transitions from relying on nonrenewable seed reserves provided by its parent to initiating photosynthetic autotrophs (Albert, 2023). Seedling emergence plays a pivotal role in crop growth and development, and high-quality seeds exhibiting a high emergence rate can positively impact the yield (Reddy et al., 2017). Moreover, the timing and uniformity of crop seeding emergence often dictate its success in competing with weeds and susceptibility to diseases. Therefore, developing an accurate, reliable, and efficient method to monitor crop seedling emergence is necessary.

Cabbage (*Brassica rapa* L. *ssp. pekinensis*) holds significant importance as a vegetable in China (Kong et al., 2020). It is a rich source of vitamins, dietary fiber, and antioxidants, promoting intestinal peristalsis and aiding digestion (Zou et al., 2021). In China, cabbage has diversified into approximately 800 local varieties (Jin et al., 2021a; Liang et al., 2019). Given its nutritional value and ease of cultivation, China boasts a high domestic production of cabbage (Wang et al., 2019). Traditionally, cabbage seedling emergence has been monitored through manual visual counting, but this approach proves inefficient and is susceptible to counting errors.

For many years, scholars have dedicated substantial effort in the domain of image processing to accomplish agricultural object recognition and counting tasks (Basavaiah and Arlene Anthony, 2020; Lin et al., 2020; Nanehkaran et al., 2020; Rahman et al., 2023; Septiarini et al., 2021). To accomplish the counting task for plant-based objects, conventional image processing encompasses four main feature categories, including color (Malik et al., 2018; Tan et al., 2018), shape (Oo and Aung, 2018), texture (Hameed et al., 2021; Kurtulmus et al., 2011; Rojas et al., 2017), and hybrid-feature-based methods (Lu et al., 2018; Wu et al., 2019; Yu et al., 2021). These methods leverage the phenotypic characteristics of plants for identification and counting. Nevertheless, traditional image processing methods often limit feature utilization to shallow ones, overlooking deep-level features (Hu et al., 2022). The artificial determination of which features to use can introduce inevitable

* Corresponding authors.
E-mail addresses: xiaojun.jin@pku-iaas.edu.cn (X. Jin), jialin.yu@pku-iaas.edu.cn (J. Yu).
[1] The first and second authors contributed equally to this work.

https://doi.org/10.1016/j.eja.2024.127191
Received 9 January 2024; Received in revised form 7 April 2024; Accepted 12 April 2024

human-factor errors when employing traditional image processing methods (Saleem et al., 2021).

With the advancements in computer hardware and software, many methods for image feature extraction and object recognition have emerged in deep learning (Devlin et al., 2018; Krizhevsky et al., 2012; Vaswani et al., 2017). Krizhevsky et al. (2012) proposed AlexNet, which incorporated more convolutional layers to extract deep features, resulting in enhanced performance for feature extraction. ResNet introduced a residual block to solve the problem of gradient vanishing and network degradation in deep neural networks, allowing the network to be designed deeper for better performance (He et al., 2016). Dosovitskiy et al. (2020) introduced the transformer attention mechanism from natural language processing to the domain of computer vision. They demonstrated that pure transformer structures can excel in image processing tasks. Detection models can be classified into two categories: region-based and region-free models. Region-based models, exemplified by the R-CNN family (Girshick et al., 2014), differ from region-free models, exemplified by the YOLO family (Redmon et al., 2016).

In recent years, many scholars have ventured into the application of deep learning for identifying crop plants, diseases, and fruits (Jin et al., 2021b, 2022, 2024; Hu et al., 2021). In practical agricultural applications, Xiao et al. (2020) compared GoogLeNet, VGG16, and ResNet50 for strawberry (*Fragaria*×*ananassa*) disease detection. Cui et al. (2022) introduced a dual-channel convolutional neural network model designed for the early-stage detection of apple (*Malus*) branches infected by apple valsa canker. Zhang et al. (2018) employed an improved AlexNet network as the backbone of R-CNN to detect apple tree branches using integrated pseudo-color and depth images. Jia et al. (2020) introduced a Mask Region Convolutional Neural Network (Mask R-CNN) for apple fruit localization. Regarding the region-free methods, Tian et al. (2019) combined DenseNet with YOLOv5s to detect apple fruits at various stages. Wu et al. (2020) proposed a real-time method for apple flower detection utilizing the channel-pruned YOLOv4 deep learning algorithm.

Monitoring the emergence of crop seedlings is a facet of the target detection challenge. Currently, a successful strategy involves leveraging UAVs to capture images of crops, subsequently employing deep learning models to process the images and achieve accurate counting results (Guo et al., 2021). Ye et al. (2023) compared object-based image analysis and Mask R-CNN deep learning method for cabbage extraction and seedling quantity estimation. Their findings revealed that the Mask R-CNN model outperformed object-based image analysis in tasks related to vegetable detection. Deep neural networks offer the advantage of rapid recognition speed, eliminating the need for human interference (Ma et al., 2020, 2022a, 2022b). Nevertheless, their recognition accuracy needs to be further improved. Only a few published studies on crop seedling counting utilized real-time video (Tan et al., 2022). In recent research, Cui et al. (2023) effectively employed YOLOv5s and ByteTrack for detecting missing rice (*Oryza sativa* L.) seedlings, showcasing the potential of these technologies. However, the specific challenge of accurately counting cabbage seedlings remains unaddressed in the literature. To fill this critical gap in precision agriculture, the present study proposes a real-time video counting method designed to achieve reliable and rapid recognition and counting of cabbage seedlings.

There are two major innovations presented in this research, including (1) the improvements to the YOLOv8n model, optimizing its performance metric for object detection on the cabbage seedling dataset while maintaining a relatively high inference speed, and (2) the introduction of a target counting algorithm named TSP-yolo, which combines the improved YOLOv8n with a tracking algorithm and OpenCV program, achieving stable and rapid counting of cabbage seedlings.

2. Materials and method

2.1. Dataset preparation

The training and testing images were taken from multiple cabbage fields at the Peking University Institute of Advanced Agricultural Sciences at Weifang, Shandong Province, China (36°23'48''N, 119°14'22''E). The cabbage fields were located in a temperate monsoon climate zone with nutrient-rich soil, an average annual temperature of about 12 ℃, and an average precipitation of 17.4 mm. The cabbage seeds were sown on August 21, 2023. The sown seeds were arranged with a row spacing of approximately 60 centimeters and an inter-plant spacing of 40 centimeters. For the purpose of irrigation, a drip hose system was employed, ensuring adequate soil moisture through irrigation every three days. Fig. 1 presents the study field and their specific geographic locations.

The cabbage images were captured on September 4, 2023, approximately two weeks following the sowing of seeds. This interval is considered optimal for observing early growth stages and assessing initial seedling establishment, providing critical data on emergence rates and seedling vigor. A Phantom 4 RTK SE (CN) Combo UAV manufactured by Shenzhen DJI Innovative Technology Co. Ltd. was used to perform aerial imaging and videography directly above the cabbage fields at a 90-degree angle. The drone was flown at a steady speed of 1.5 m per second along a predefined flight path. The flight operations were conducted using the software provided by the drone manufacturer, allowing for precise control over the flying speed, altitude, and degree of overlap between flight paths. This ensured consistency and comprehensiveness in the image capture process, thereby guaranteeing the continuity and coverage necessary for accurate seedling emergence monitoring. Equipped with a powerful 20-megapixel camera and flying at a constant altitude of 3 m above ground level, the UAV captured a total of 4058 high-resolution images at 5472 × 3648 pixels and videos at 3840 × 2160 pixels with a smooth frame rate of 30 frames per second (fps).

To prepare the images for deep learning training of target detection algorithms, Anylabeling (version 0.3.3, open-source software available at https://github.com/vietanhdev/anylabeling) was used to annotate the images using the Segment Anything model (Kirillov et al., 2023). The ground truth dataset included categorical information, along with the coordinates and sizes of bounding boxes, designed to precisely enclose target objects while minimizing gaps and avoiding overlap. The software processed images captured by the UAVs using the Segment Anything model to automatically delineate the targets. The bounding boxes were manually adjusted if its output from the model did not adhere to the predefined rules. Each target was then manually labeled as "cabbage". Finally, the software compiled the categories of all annotated objects in an image along with their corresponding bounding box details into a JSON format file. The resulting JSON files were subsequently formatted using Python scripts to obtain annotation files in TXT format that could be used for YOLOv8 training. The training and testing datasets were divided into a 7:3 ratio, consisting of 2840 annotated images for training and 1218 for testing.

2.2. Improvement of YOLOv8n

The YOLO family is a well-known algorithm for one-stage target detection and is frequently used for detection tasks. Compared with the two-stage target detection algorithm, the one-stage target detection algorithm significantly improves the detection speed of the network by merging the first stage (focused on identifying the target object location and obtaining the proposed region) with the second stage. To reduce the deployment cost, YOLOv8n, which has the least network layers and the lowest feature map width among the YOLOv8 family, is chosen as the base model for cabbage detection in this study.

Our preliminary test indicated that YOLOv8n exhibits limitations

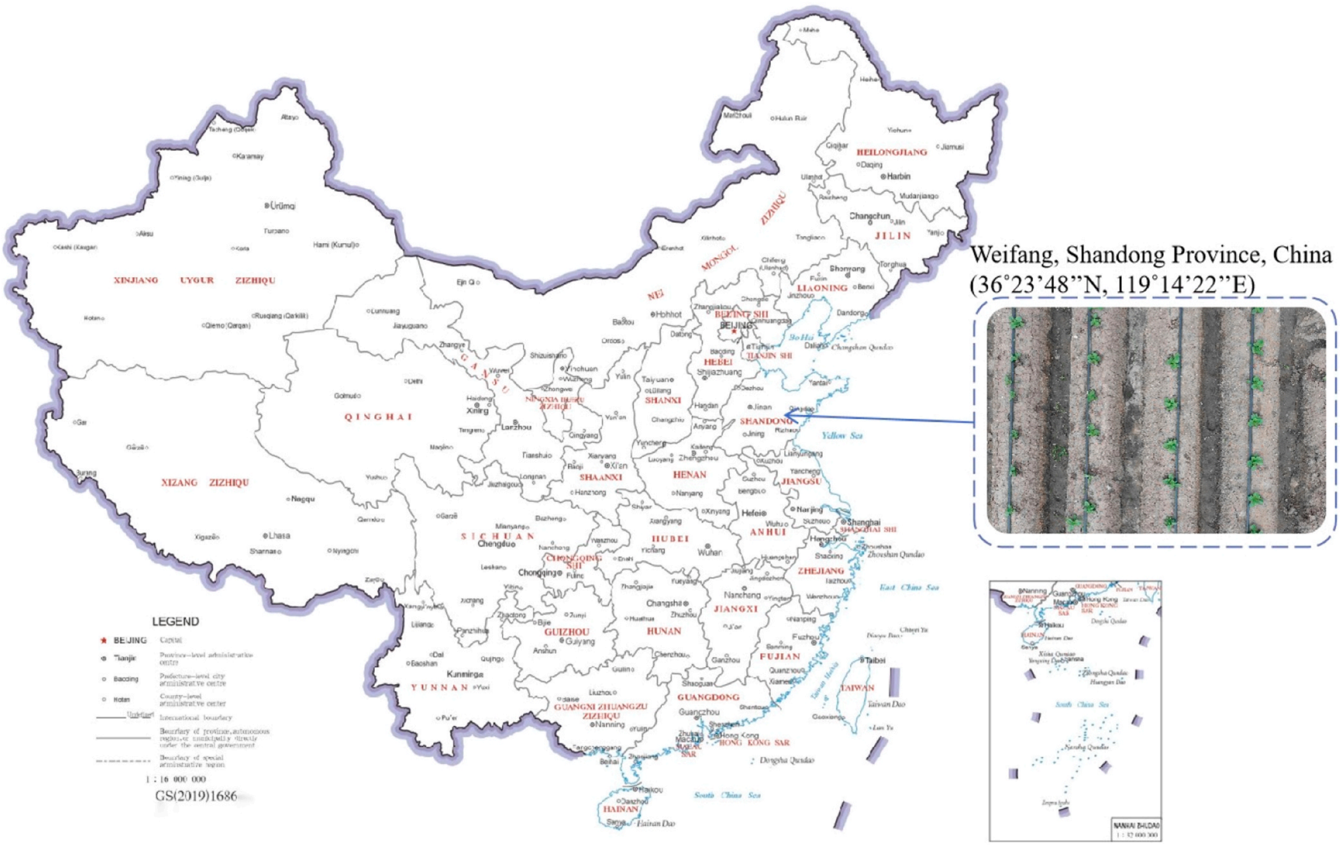

Fig. 1. Experimental data collection site.

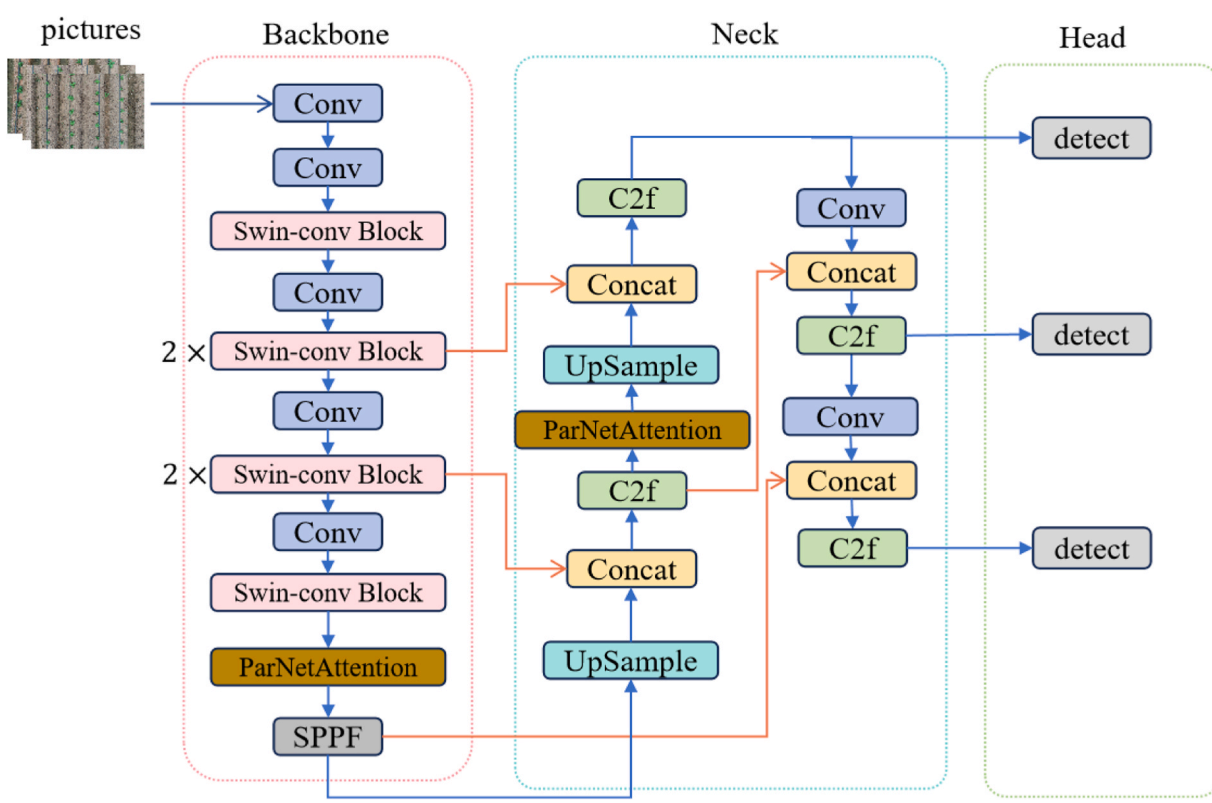

Fig. 2. The structure of the improved YOLOv8n network.

impacting performance in certain scenarios. A notable challenge is suboptimal feature extraction, potentially resulting in reduced detection accuracy in complex or variable environments. Furthermore, its architecture may inadequately distinguish similar objects. These limitations can become particularly evident in tasks requiring high precision, such as the detection of specific types of vegetation in diverse agricultural settings. To address these challenges and optimize the algorithm for applications in cabbage detection, we have implemented enhancements to YOLOv8n.

The algorithm was improved in the following two key aspects:

1) The efficiency of feature extraction was significantly improved by implementing the Swin Transformer layer instead of the C2f (Cross Stage Partial Network Fusion) module in the YOLO backbone network.
2) The ParNet attention mechanism module was incorporated into the YOLO backbone and head network to improve the accuracy of cabbage recognition. The resulting network structure is depicted in Fig. 2.

2.2.1. Swin-conv Block

Swin Transformer (Liu et al., 2021), as an important variant network of the Transformer (Dosovitskiy et al., 2020) model in the field of computer vision, has demonstrated exceptional performance in target detection tasks. Inspired by the Swin Transformer, we proposed the Swin-conv Block to replace the C2f module in the YOLOv8n backbone network for improved feature extraction. In the Swin-conv Block, the Swin Transformer Block is utilized to establish cross-window connectivity via 'shifted windows'. This enables the model to focus on the relevant information of neighboring windows and facilitate cross-window feature interaction. This feature interaction capability expands the receptive field of the model, enabling it to better understand and represent the position and shape of objects. As a result, Swin-conv Block has higher computational efficiency and better performance in processing complex images and data compared to C2f.

The network structure of Swin-conv Block is shown in Fig. 3(a). The input feature map is fed into two streams: integrating the global feature using the Swin Transformer block and extracting features directly using the convolution block. These streams are then fused together to achieve the fusion of feature maps from two different levels. It is worth mentioning that by applying a convolutional layer of 1×1 convolutional kernel size rather than a fully connected layer, the limitation on the size of the input feature map is lifted while realizing the upscaling or downscaling operation of the feature map.

The Swin Transformer Block in Fig. 3(a) is composed of the Window-based Multiple Self-Attention (W-MSA) module and the Shifting Window-based Multiple Self-Attention (SW-MSA) module in tandem, as shown in Fig. 3(b).

The W-MSA module first divides the feature map into multiple small windows without overlapping and then performs self-attention operations inside each window individually. The SW-MSA module shifts the output feature vectors of the previous W-MSA module using shifted windows and then does the same self-attention operation. In this way, the interaction of feature information across windows is realized. The specific window division is shown in Fig. 4. In summary, the Swin-conv Block has higher computational efficiency and a better ability to interact with features from different regions compared to the C2f module.

2.2.2. ParNet attention

The utilization of attention mechanisms can effectively direct the neural network's attention toward significant information by assigning

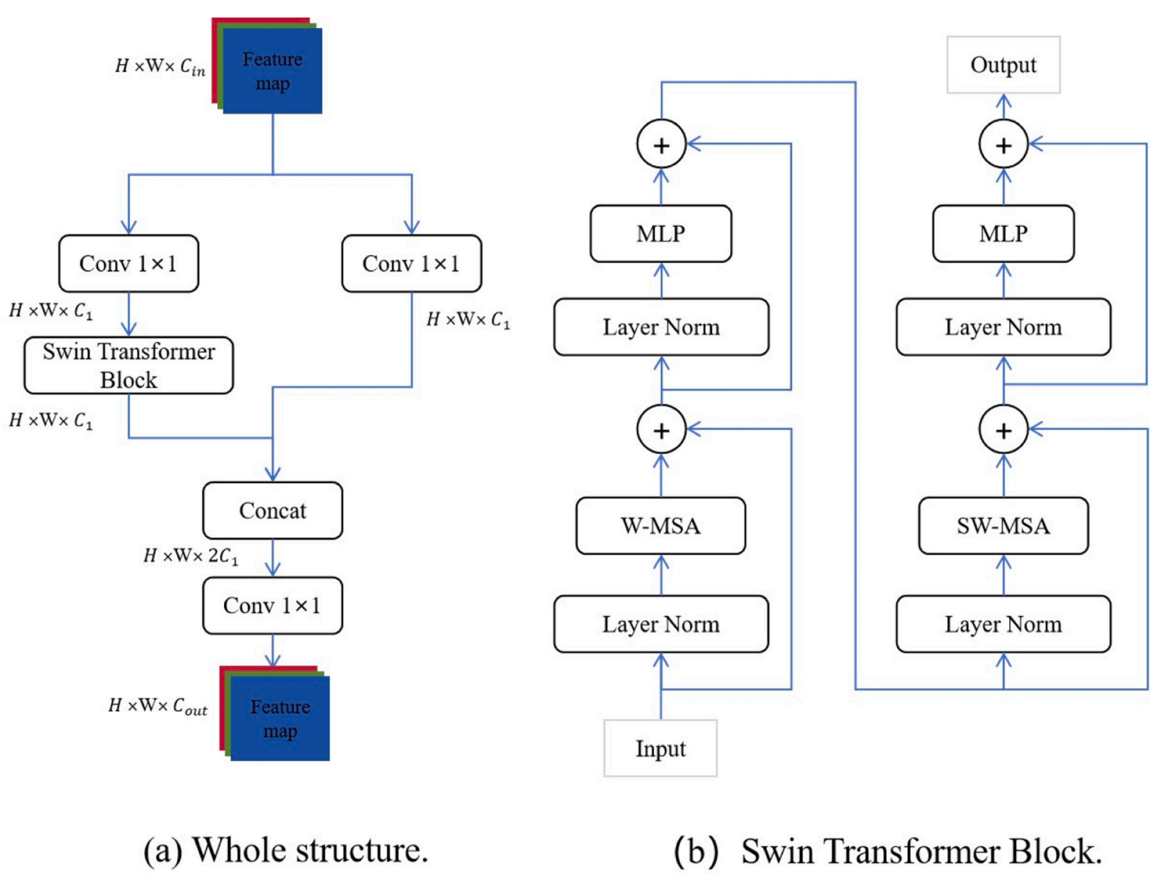

Fig. 3. The network structure of Swin-conv Block. Abbreviations: MLP, multilayer perceptron; SW-MSA, Shifting Window-based Multiple Self-Attention; W-MSA, Window-based Multiple Self-Attention.

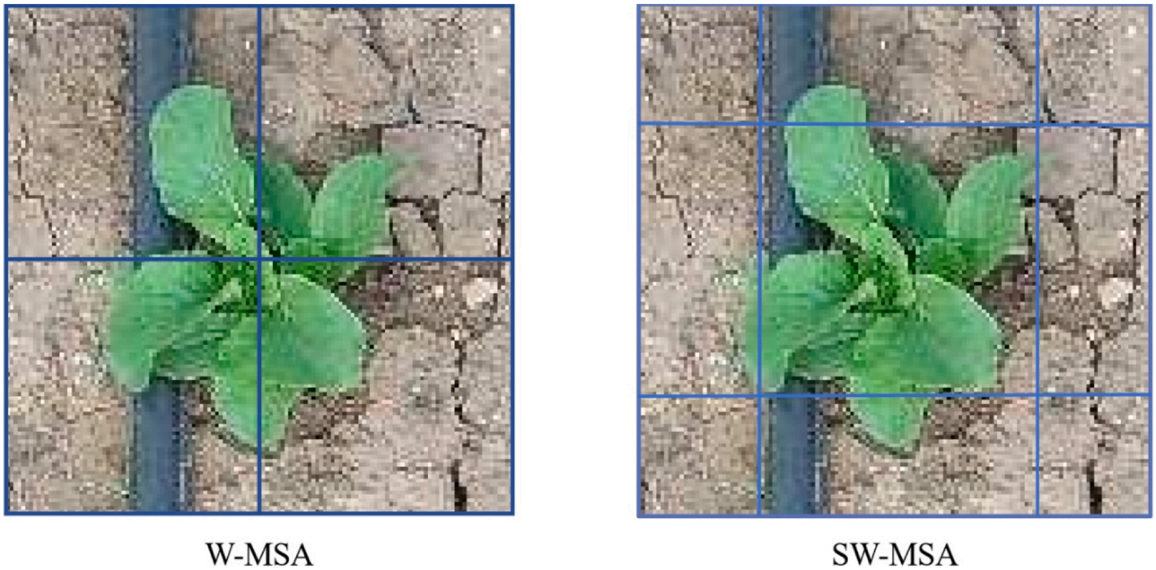

Fig. 4. The difference window division method between W-MSA and SW-MSA. Abbreviations: SW-MSA, Shifting Window-based Multiple Self-Attention; W-MSA, Window-based Multiple Self-Attention.

high weights. This helps extract occluded features while leading to a slight increase in computation costs. The ParNet attention mechanism (Goyal et al., 2022) employs a combination of convolutional layers with varying sizes of convolutional kernels to extract vectors with diverse field of view sizes. These vectors are then subjected to batch normalization and fused into a feature vector through a 3×3 convolution. The reparameterization of this structure helps to reduce the latency in the inference process while reducing the network depth through parallel computation. In addition, the Skip-Squeeze-Excitation (SSE) module is designed as an attention operation to solve the problem of a limited receptive field for 3×3 convolution. Finally, SiLU is used instead of the ReLU activation function to enhance the nonlinear representation of the network. The overall flow of the ParNet attention mechanism is shown in Fig. 5. An experimental comparison was conducted to verify the suitability of ParNet Attention as the primary attention mechanism for this task. A total of 11 other attention mechanisms, along with the ParNet, were analyzed and compared. Multiple attempts were made to insert each of the attention mechanisms at various locations within the network. The final evaluation was based on the results obtained using the most effective insertion method for each mechanism.

2.3. *TSP-yolo-based real-time cabbage seedling video counting model*

The real-time model for counting cabbage seedlings utilizes two primary components: the TSP-yolo object tracking algorithm and the target counting algorithm. When collecting videos using UAVs, the presented targets may appear in multiple frames due to the temporal continuity of the video footage. Implementing an object detection algorithm, such as YOLO, for each frame may not accurately recognize cabbage shape features as the same target due to variations in viewing angle, wind disturbance, and other external factors. Inspired by a multiple object tracking algorithm-DeepSort (Wojke et al., 2017), this research devised a TSP-yolo object tracking algorithm to effectively track the target object and assign trajectory tracking and ID tags to each individual target.

In order to accurately track an object, matching the target in the preceding and following frames is required to determine if it is the same target. The conventional method utilizes a Kalman filtering algorithm to

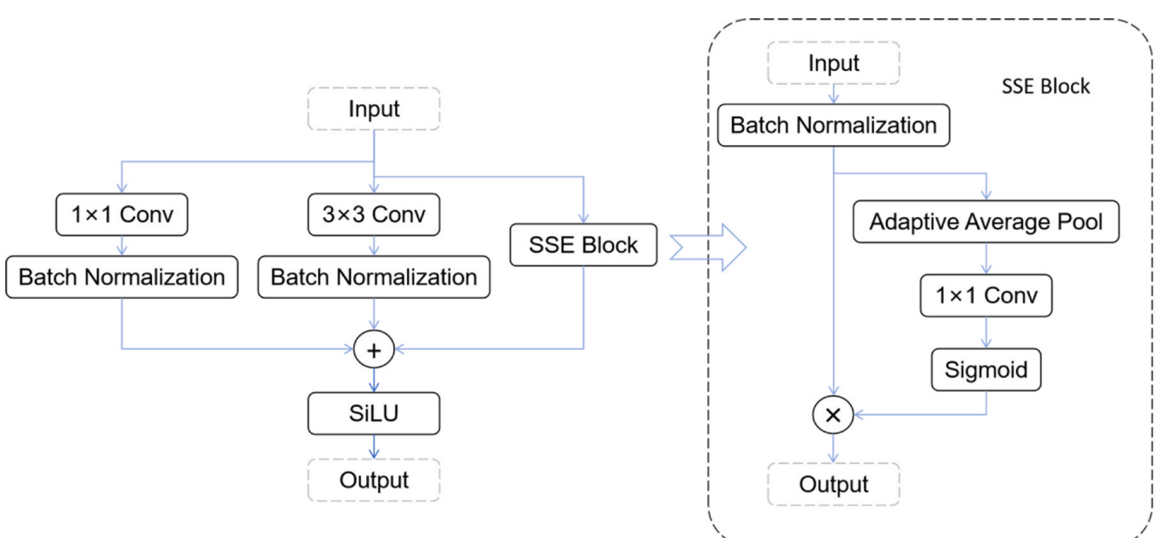

Fig. 5. The structure of the ParNet attention mechanism.

predict the possible position of the target in the next frame, using its motion information from the previous frame. This predicted position is then compared to the actual detected target frame through IoU matching, which measures the degree of overlap between the two frames in terms of area repeatability. Suppose the matching is successful, as indicated by an IoU value greater than a pre-determined threshold. In that case, it suggests that the target position is comparable in the two frames and may correspond to the same target. On this basis, our method utilizes the target feature vectors extracted by the improved YOLOv8n for an extra match.

The TSP-yolo object tracking algorithm is described as follows: initially, the improved YOLOv8n detects the detection frame position and size of the target in the previous and current frames and extracts its feature vector. The detected position information of the target in the previous frame is used to predict its possible position in the current frame using the Kalman filtering algorithm. Subsequently, the predicted current frame position is matched with the actual detected current frame position. Based on successful position matching, extra matching is further performed by calculating the similarity of the feature vectors of the target object in the consecutive frames. Feature vector matching enables target identification and confirms whether the target is the same across frames. Finally, the Kalman filter algorithm is updated according to the tracking results, ensuring continuous prediction of the target location in subsequent frames.

The whole process achieves continuous tracking of the target through multi-stage matching and prediction, making full use of positional, appearance, and motion information. This integrated tracking approach can more robustly deal with the changes and movements of the target in the video. The technical route of the TSP-yolo object tracking algorithm is shown in Fig. 6.

To implement the cabbage seedling counting function, the proposed method employed the following target counting algorithm:

a. Track the cabbage seedlings in the video frame using the TSP-yolo object tracking algorithm and anchor the target.
b. Point the center point of the lower border line of the cabbage seedling anchor frame as the marking point of the target and set a detection line in the video.
c. Upon detection of the marking point crossing the detection line for the first time, as confirmed by the TSP-yolo object tracking algorithm, the count of cabbage seedlings is increased by one.

The cabbage seedling emergence rate can be calculated by dividing the total number of counts by the number of seeds sown. The target counting algorithm is depicted in Fig. 7.

2.4. Experimental configuration

All the deep neural network models in this paper were trained and tested on the same device with designated hardware configurations. The operating system was Ubuntu 20.04.6 LTS, while the CPU utilized was an Intel® Xeon® W-2265, and the GPU was an Nvidia GeForce RTX 3080Ti with 12 G of video memory. The configured conda environment included Python 3.9.18, PyTorch 2.0.1, CUDA 11.7, and Ultralytics YOLO v8.0.3. The hyperparameters for deep learning training are presented in Table 1.

2.5. Evaluation metrics

In this paper, the following performance metrics were selected to assess the performance of the YOLO model: precision, recall, mAP50, mAP50–95, and inference time. The index calculations are presented in Eqs. (1)-(3).

$$\text{precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{mAP} = \text{AP} = \int_0^1 \text{precisiond}(\text{recall}) \quad (3)$$

where mAP refers to Mean Average Precision. In our case, mAP is defined as the Average Precision (AP) value specifically for the cabbage class due to the recognition target being limited to only one class. True positive (TP) represents the count of samples accurately classified as cabbage seedlings; false positive (FP) refers to the count of samples

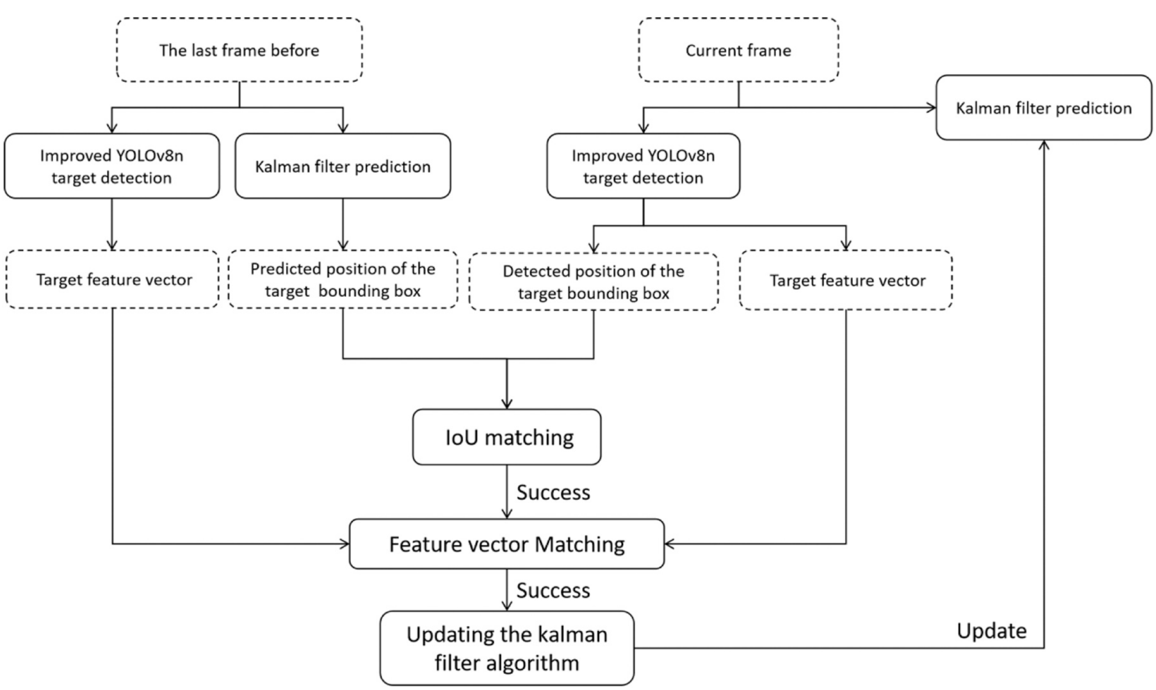

Fig. 6. Overall technical route of the TSP-yolo object tracking algorithm.

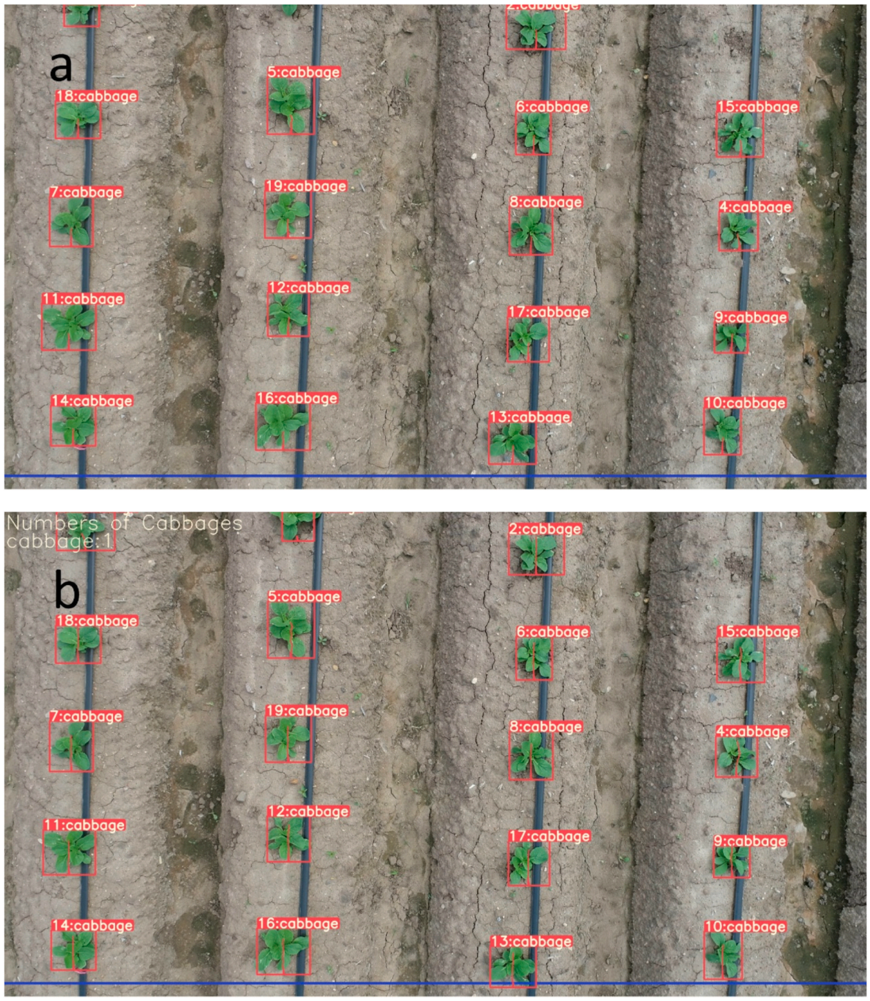

Fig. 7. The operational procedure of the proposed TSP-yolo object tracking algorithm: (a) The cabbage seedlings are tracked, anchor boxes are employed, and a marking point and detection line are added, (b) The counting is performed as the marking point of the tracked seedlings intersects the detection line.

Table 1
The hyperparameters for deep learning training.

Hyperparameter	Parameter value
Total epoch	200
Batch size	16
Image size	640×640 pixels
Number of workers	8
Optimizer	Adam
Beta	0.95
Initial learning rate	0.001
Weight decay	0.005

incorrectly classified as cabbage seedlings; and false negative (FN) represents the count of samples that were not correctly identified as cabbage seedlings.

2.6. Deep learning model-based counting versus manual counting

Ten video segments, each covering four rows of cabbage seedlings spread across a 2.4 × 12 m area, were captured to evaluate the effectiveness of the cabbage counting methods. The seedling emergence rate was calculated by dividing the number of emerged seedlings by the total number of seedlings (both emerged and non-emerged) within each video segment. QQ plots and Levene's test were carried out to examine normality and equality of variances, respectively. The Student's t-test, with a significance level of 0.05, was conducted to compare the means of the proposed deep learning model-based and manual counting methods.

3. Results and discussion

3.1. Comparison and ablation experiment

Table 2 presents the results of the attention mechanisms incorporated into YOLOv8n. It can be observed that all the attention modules significantly enhanced the model's composite metric of mAP. This proves the effectiveness of the attention mechanism in enhancing the performance of the vision detection model. The impact of the attention mechanism on mAP50–95 is significant for comparison. Among the 12 distinct attentional mechanisms evaluated, the ParNet Attention Module exhibited the highest performance with a score of 81.7% on the mAP50–95, outperforming the second-highest improvement achieved by the SimAM Attention Module by 1.7%.

Inference time is essential. Inserting an attention module will inevitably increase the inference time by a certain duration. The insertion of the SimAM module resulted in an inference time of 1.2 ms, while the insertion of the ParNet attention module resulted in an inference time of 0.6 ms. The insertion of other attention modules resulted in a marginal increase of 0.1–0.2 ms inference time. The 0.6 mm inference time allows for enough image processing time for the algorithm, enabling real-time detection. Therefore, employing ParNet attention mechanisms effectively improves the efficacy of cabbage detection models with minimal impact on inference time.

Table 2
Performance of the inserted attention module.

	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)	Inference time (ms)
YOLOv8n	97.4	96.8	99.1	75.8	0.5
+CoT	96.3	96	98.9	75.9	0.6
+ECA	97	95.3	98.9	80	0.6
+EffectiveSE	97.6	94.2	98.9	79.3	0.6
+Gam	97.1	95.8	99	79.2	0.6
+LSK	96.4	96.5	98.9	79.8	0.6
+MHSA	97.7	96.6	99.2	78.5	0.6
+MobileViT	97.6	94.3	98.9	79.7	0.7
+ParNet	95.9	97.7	99	81.7	0.6
+SE	96.9	94.6	98.9	78.5	0.6
+Shuffle	95.8	96	98.9	80.3	0.6
+SimAM	98.7	97.2	99.1	78	1.2
+SK	97.1	97	99	78	0.6

Abbreviations: CoT, Cross-modal Transformer Attention; ECA, Efficient Channel Attention; EffectiveSE, Effective Squeeze and Extraction Attention; Gam, Global Average Max Attention; LSK, Large Separable Kernel Attention; MHSA, Multi-Head Attention Mechanism; ParNet, ParNet Attention; SE, Squeeze and Extraction Attention; Shuffle, Shuffle Attention; SimAM, Selective Image Attention Mechanism; SK, Selective Kernel Attention.

Heatmaps serve as effective tools for visualizing the performance, focus areas, and attention distribution of neural network models. Fig. 8 illustrates the heatmaps generated by incorporating attention mechanisms into YOLOv8n. Specifically, the heatmap produced by YOLOv8n-ParNet Attention exhibits moderate coverage and divergence. Heatmaps provide insights into the areas of interest the model detects and the intensity of attention allocated to those areas. In the context of YOLOv8n-ParNet Attention, the moderate coverage suggests that the model successfully identified relevant features across the image, while the moderate divergence indicates a balanced distribution of attention across multiple regions.

Table 3 presents the performance results of the model after combining different attention mechanisms and Swin-conv Block. It is evident that not all combinations of attention mechanisms and Swin-conv Block lead to further improvements in model performance. The incorporation of EffectiveSE attention yielded the highest scores, with the model achieving 98% and 99.5% on the Recall and mAP50 metrics, respectively, outperforming all other combinations. However, among the evaluated 12 combinations, the combination of ParNet attention and Swin-conv Block achieved a score of 90.3 % on the mAP50–95 metric, surpassing the second-ranked Shuffle attention module by 3.1 %. Additionally, regarding inference time, the combination of ParNet attention and Swin-conv Block only consumes 1 ms per frame, making it

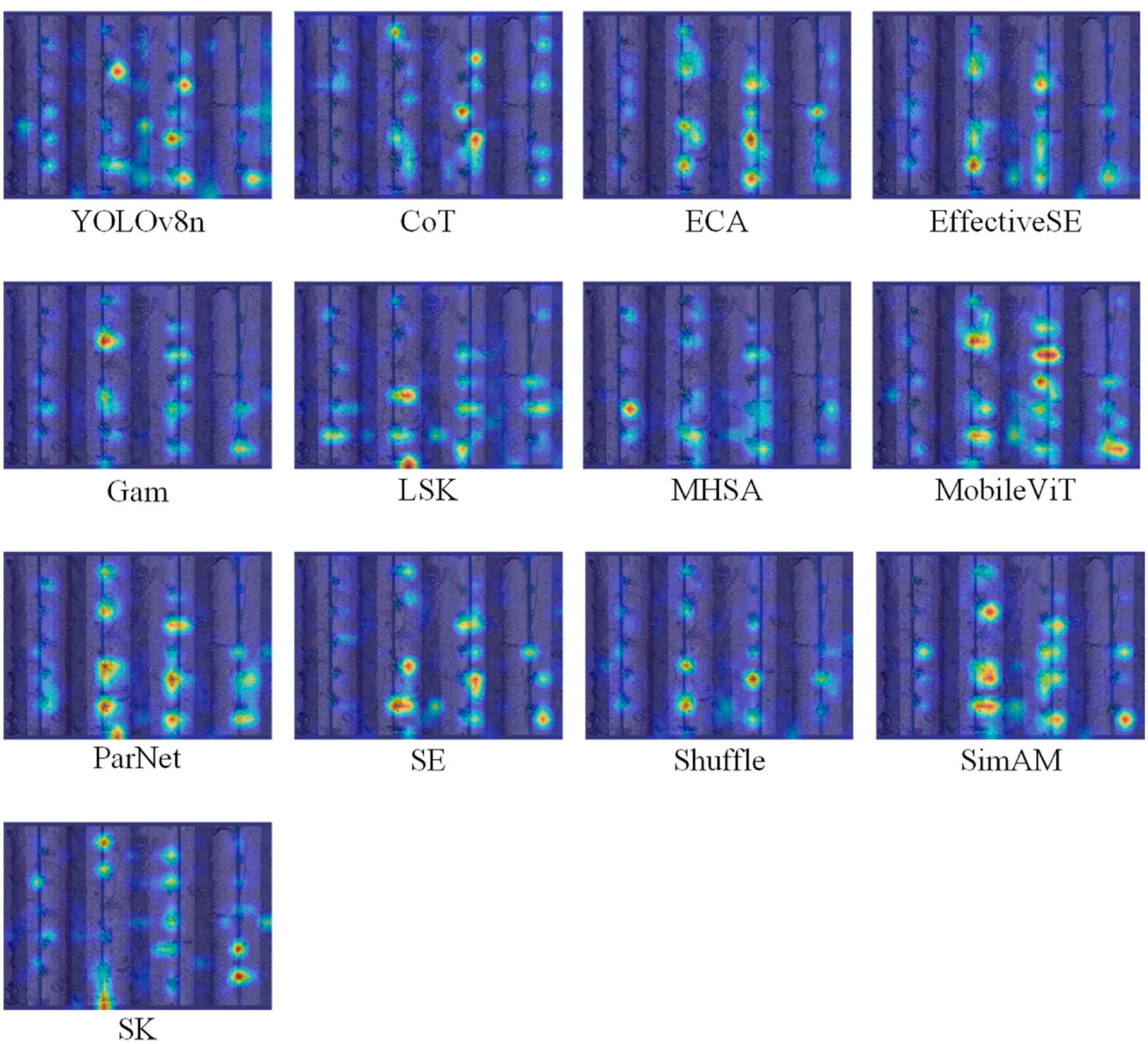

Fig. 8. Heatmaps of the images for the attention mechanisms.

Table 3
Comparison of different attention mechanisms and Swin-conv Block combination.

	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)	Inference time (ms)
YOLO v8n+Swin-conv Block	98.4	97.4	99.3	85.1	0.9
+CoT	98.1	97.2	99.2	83.5	1.1
+ECA	98.8	97.3	99.3	86.4	1.2
+EffectiveSE	98.1	98	99.5	85.7	1.1
+Gam	98.1	97.7	98.8	85.4	1.1
+LSK	98.4	97.5	99.1	86.6	1.3
+MHSA	97.9	97.4	99.1	85.2	1.3
+MobileViT	98.1	97.5	98.6	86	1.7
+ParNet	98.5	97.8	99.4	90.3	1.0
+SE	98.7	96.6	99.2	84	1.4
+Shuffle	99	97.5	99.3	87.2	1.3
+SimAM	98.2	97	99.1	83.4	1.6
+SK	97.5	96	98.6	83	1.2

Abbreviations: CoT, Cross-modal Transformer Attention; ECA, Efficient Channel Attention; EffectiveSE, Effective Squeeze and Extraction Attention; Gam, Global Average Max Attention; LSK, Large Separable Kernel Attention; MHSA, Multi-Head Attention Mechanism; ParNet, ParNet Attention; SE, Squeeze and Extraction Attention; Shuffle, Shuffle Attention; SimAM, Selective Image Attention Mechanism; SK, Selective Kernel Attention.

the fastest among all combinations. From these observations, it can be concluded that the combination of ParNet attention and Swin-conv Block exhibits the best performance.

In order to demonstrate the effectiveness of the ParNet attention mechanism and the Swin-conv Block in improving the model performance, a series of ablation experiments were implemented. The results are shown in Table 4. The results of the ablation experiments showed that the incorporation of the ParNet attention module to the network structure alone resulted in an improvement of 5.9% in mAP50–95, reaching a performance of 81.7%; however, this also resulted in a slight increase of 0.1 ms in inference time. This outcome is understandable and can be attributed to the increased complexity of the network structure due to the incorporation of the attention module. Replacing the C2f module in the base model with the Swin-conv module yielded significant improvement to the base network, with performance reaching a precision, recall, mAP50, and mAP50–95 of 95.9 %, 97.4 %, 99.3 %, and 85.1 %, respectively. The inference time increased to 0.9 ms, while the mAP50–95 exhibited a significant increase of 9.3 %. This comparison proves the effectiveness of both improvements in yolov8n, resulting in promising outcomes. By incorporating these two improvements into the base model simultaneously, the improved YOLOv8n achieved remarkable performance metrics, reaching precision, recall, mAP50, and mAP50–95 values of 98.5 %, 97.8 %, 99.4 %, and 90.3 %, respectively. Meanwhile, the inference time was 1 ms per image, which is double the duration of the base model; however, it remains capable for real-time detection purposes.

Table 4
Comparison between the base model and the improved YOLOv8n.

	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)	Inference time (ms)
YOLO v8n	97.4	96.8	99.1	75.8	0.5
+ParNet attention	95.9	97.7	99	81.7	0.6
Replace C2f with Swin-conv Block	98.4	97.4	99.3	85.1	0.9
+ParNet attention +Swin-conv Block	98.5	97.8	99.4	90.3	1.0

Abbreviations: C2f, Cross stage partial network fusion block.

3.2. Performance comparison between improved YOLOv8n and other SOTA networks

In order to compare the effectiveness of improved detection models, experiments were conducted using the prevailing models from two dominant target detection schools of thought (two-stage vs. one-stage). The results are shown in Table 5.

Table 5 displays the substantial superiority of one-stage models, including YOLOv5s, YOLOv8n, and improved YOLOv8n, over the two-stage model Faster R-CNN (Ren et al., 2015) in terms of overall performance. Especially for the mAP50–95 metric, Faster RCNN achieved only 46.2%, significantly lower by >37.3 % compared to one-stage models. Additionally, the inference time for Faster RCNN, at 52.7 ms per image (19.96 fps), severely restricted its ability to process video in real-time. Between the one-stage algorithms, the improved YOLOv8n yielded slight discrepancies in precision, recall, and mAP50 when compared to the widely used YOLOv5s. However, this gap was noticeable and pronounced in the more rigorous mAP50–95 metrics, where the YOLOv8n achieved 90.3 %, indicating a 6.8 % increase in performance compared to the YOLOv5s.

In terms of inference time, YOLOv5s exhibited a processing speed of 3.1 ms per frame. However, the improved YOLOv8n boasted an even shorter inference time, providing ample room for subsequent image processing while still meeting real-time processing requirements. With enough time, deploying devices with excessive arithmetic performance is unnecessary to ensure real-time detection, significantly reducing equipment costs. In conclusion, the improved YOLOv8n is capable of effectively, quickly, and efficiently identifying cabbage seedlings.

3.3. Counting performance

A study was conducted to evaluate the effectiveness of the UAV-implemented model for counting cabbage seedlings in field conditions. Experiments were undertaken to compare the manual counting method with the TSP-yolo-based real-time video counting model implemented with a UAV, and the obtained results are presented in Table 6.

In this set of comparisons, the proposed machine vision-based method achieved 98% accuracy under large sample conditions. The proposed target counting method significantly outperformed the manual counting method in terms of time efficiency, taking only about one-seventh of the time consumed by the manual method. Manual counting requires continuous contemplation in a larger field, and the possibility of error increases with longer working hours. In contrast, the machine vision-based method can maintain an impressive accuracy of 98.02% until completing the entire seedling counting task.

An investigation of the efficacy of the proposed method relative to manual counting was performed using ten video segments, each covering a 1.6×12 m field. The counting of cabbage seedlings was executed using model-based and manual methods for each video. The cabbage seedling number counted manually and by the proposed method were recorded as 116 ± 4.7 and 117.6 ± 4.3 (mean $\pm$ standard errors) per video, respectively. These counts correspond to emergence rates of $94.7\pm1.3\%$ and $95.6\pm1.2\%$ (mean $\pm$ standard errors), respectively. The P-value derived from the application of the Student's t-test exceeded 0.05 for both the number of counts and emergence rates,

Table 5
Comparison between the improved YOLOv8n and SOTA models.

Architecture	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)	Inference time (ms)
Faster RCNN	69.9	93.8	91.3	46.2	52.7
YOLOv5s	97.7	98	99.3	83.5	3.1
YOLOv8n	97.4	96.8	99.1	75.8	0.5
improved YOLOv8n	98.5	97.8	99.4	90.3	1.0

Abbreviations: Faster RCNN, Faster Region Convolutional Neural Network.

Table 6
Cabbage seedling counting results obtained from the TSP-yolo-based real-time video counting model and the manual counting method.

Method	Number of instances	Accuracy (%)	Time (min)
Manual counting	761	100	21.6
Realtime cabbage seedling video counting model	746	98.02	3.5

suggesting no statistically significant difference between the manual counting and the proposed method. These findings substantiate the feasibility of the proposed method as an effective alternative to manual counting.

Fig. 9 illustrates the validity of the cabbage seedling counting model. Subfigures (a), (b), and (c) in Fig. 9 are arranged in a time sequence. In Fig. 9(b), the marked point of the cabbage labeled 53 touched the detection line, resulting in a subsequent increase of 1 in the number of cabbages in the upper left corner, reaching a total of 39. In Fig. 9(c), the

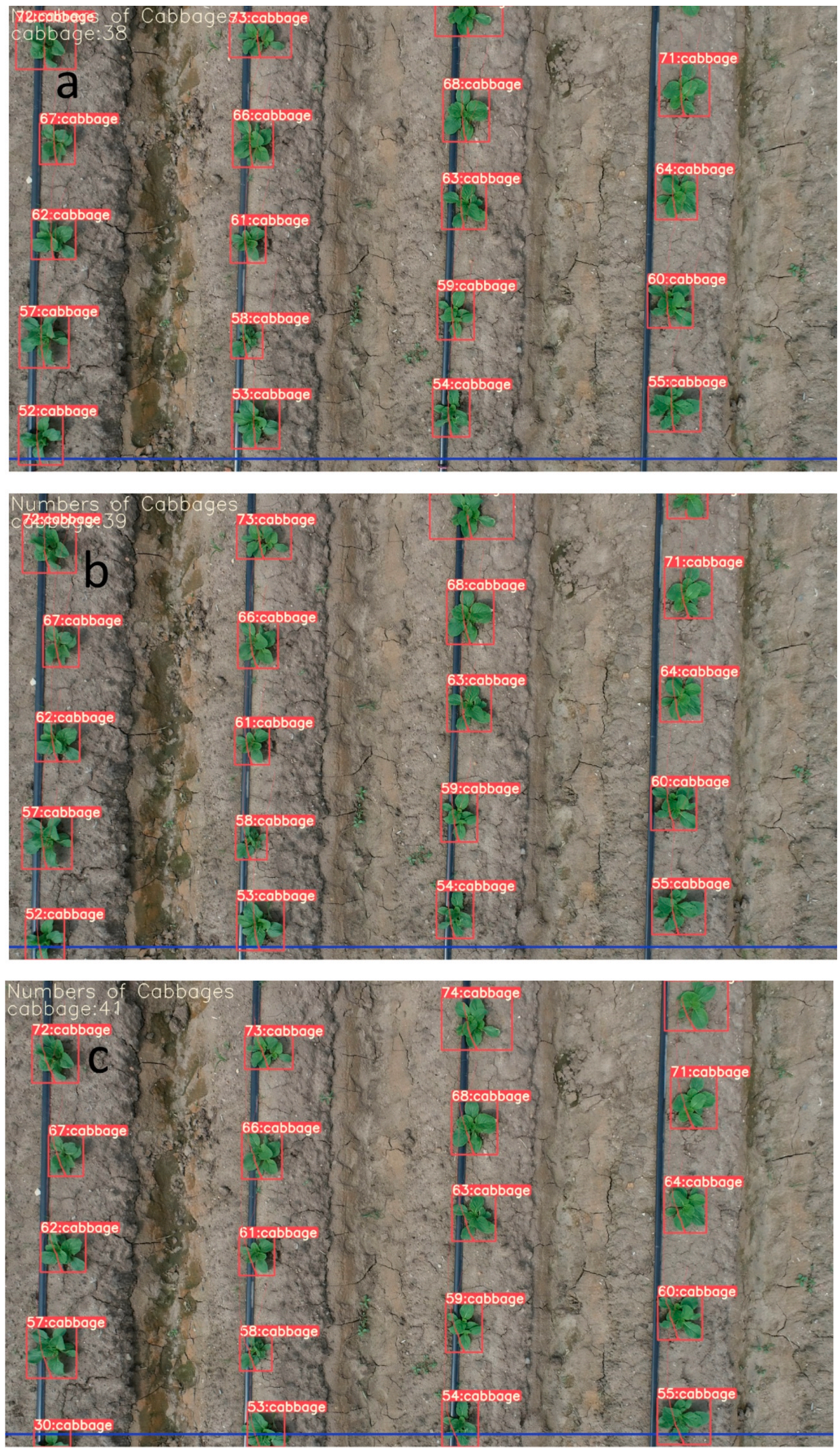

Fig. 9. Demonstration of cabbage seedling counts.

marked points corresponding to cabbages labeled 54 and 55 hit the detection line, resulting in an increase in the number of cabbages present in the upper left corner from 39 to 41.

Fig. 10 presents the results of the model counting method. Furthermore, the total seedling counts are divided by the area surveyed to calculate the seedling density. The recorded video, lasting 3 minutes and 32 seconds, demonstrated the successful identification of 746 cabbage seedlings by the proposed counting model, as shown in Table 6. The numerical values of IDs assigned to cabbages in Fig. 10 amounted to 1645. This observation can be attributed to the influence of UAV-generated wind affecting the foliage features of cabbages in field environments. Consequently, the network re-assigned new IDs to cabbages exhibiting significant variations in their foliage features, resulting in a substantial increase in the total ID count. However, since the impact of the wind is transient, the network can quickly reassign the correct IDs to cabbages. This phenomenon did not affect the model's ability in detecting and counting in the actual scenario. The above comparisons suggest that the proposed method can effectively, quickly, and reliably perform the task of cabbage seedling counting. It is worth noting that the flight operations were meticulously conducted using the software provided by the drone manufacturer. This software enables precise control over the drone's flying speed, altitude, and degree of overlap between flight paths, thereby ensuring consistency and comprehensiveness in the image capture process. Such a setup is critical for guaranteeing the continuity and coverage necessary for accurate seedling emergence monitoring. However, the challenges of repeated target recognition and handling non-sequential frames may arise in certain scenarios, which warrants further investigation.

It is worth noting that this study introduced the TSP-YOLO framework (Tracking algorithms-Swin-conv blocks-ParNet attention-YOLOv8n), representing a substantial architectural enhancement over the existing models. Integrating Swin-conv blocks and ParNet attention modules into both the backbone and neck parts of YOLOv8n is a novel approach that effectively captures the complex features of cabbage seedlings. In addition, compared to ByteTrack (Cui et al., 2023), DeepSort is known for its high accuracy in object tracking due to its sophisticated association algorithm based on deep appearance descriptors. DeepSort's application to the TSP-YOLO framework enhances the tracking stability and reliability, especially in scenarios characterized by considerable variability in seedling appearance, arising from morphological and environmental factors unique to cabbage cultivation. Overall, the proposed TSP-YOLO framework has proven to be consistent and reliable, offering significant labor savings and presenting a promising avenue for large-scale implementation in precision vegetable farming.

4. Conclusions

Accurate and reliable machine vision models are essential to identify and count cabbage seedlings. To optimize this process, the present study introduces a YOLOv8n target detection network, enhanced through the substitution of the C2f module in the backbone network with the Swin-conv module. This modification leads to improved feature extraction capabilities, and the network is further complemented by integrating the ParNet Attention Mechanism module into the network structure for enhanced performance. The experimental results show that the improved YOLOv8n outperformed the original YOLOv8n network. The mAP50–95 of the YOLOv8n has significantly increased from 75.8 % to 90.3 %, exhibiting a notable improvement of 14.5 %. Compared with mainstream models, the performance of the improved YOLOv8n demonstrated superior performance in all aspects. In addition, a TSP-yolo-based real-time cabbage seedling video counting model was developed based on the enhanced YOLOv8n. The counting model achieved an impressive accuracy of 98.02 %, and the counting speed is much faster than that of human counting. Employing a testing dataset comprised of ten video segments, the findings suggest that the efficacy of our method for estimating the cabbage emergence rate is as effective as the traditional manual counting technique. When combined with a UAV, the model enables accurate, expedient, and real-time counting of cabbage seedlings in natural field conditions, demonstrating its potential for intelligent counting purposes. By dividing the counted seedling number by the number of seeds sown, the emergence rate can be determined; by dividing the counted seedling number by the area surveyed, the cabbage seedling density can be determined. This approach of emergence rate or crop density assessment possesses versatile applications and may be effectively applied in determining seedling densities or emergence rates for a wide range of crop species. Its implementation can significantly improve the efficiency of crop management.

Funding

This work was funded jointly by the Key R&D Program of Shandong Province, China (Founding code: 202211070163), the National Natural Science Foundation of China (Founding code: 32072498), the Taishan Scholar Program of Shandong Province, China.

CRediT authorship contribution statement

Teng Liu: Resources. **Xin Chen:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Data curation, Conceptualization. **Xiaotong Kong:** Writing – review &

Fig. 10. The results of the model counting.

editing, Visualization, Software, Conceptualization. **Jinxu Wang:** Software, Investigation, Formal analysis. **Xiaojun Jin:** Writing – review & editing, Resources, Methodology, Data curation, Conceptualization. **Kang Han:** Software, Data curation. **Jialin Yu:** Writing – review & editing, Supervision, Resources, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

References

Albert, M., 2023. Relationship between laboratory and field assessments of common bean (*Phaseolus vulgaris* L.) seed quality indicators. Acta Agric. Slov. 119 (3), 1–8. https://doi.org/10.14720/aas.2023.119.3.13508.

Basavaiah, J., Arlene Anthony, A., 2020. Tomato leaf disease classification using multiple feature extraction techniques. Wirel. Pers. Commun. 115 (1), 633–651. https://doi.org/10.1007/s11277-020-07590-x.

Cui, R., Li, J., Wang, Y., Fang, S., Yu, K., Zhao, Y., 2022. Hyperspectral imaging coupled with Dual-channel convolutional neural network for early detection of apple valsa canker. Comput. Electron. Agric. 202, 107411 https://doi.org/10.1016/j.compag.2022.107411.

Cui, J., Zheng, H., Zeng, Z., Yang, Y., Ma, R., Tian, Y., Tan, J., Feng, X., Qi, L., 2023. Real-time missing seedling counting in paddy fields based on lightweight network and tracking-by-detection algorithm. Comput. Electron. Agric. 212, 108045 https://doi.org/10.1016/j.compag.2023.108045.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv: 1810.04805. ⟨https://doi.org/10.48550/arXiv.1810.04805⟩.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., 2020. An image is worth 16×16 words: Transformers for image recognition at scale. arXiv preprint arXiv: 2010.11929. https://doi.org/10.48550/arXiv.2010.11929.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Columbus, OH, USA. https://doi.org/10.1109/CVPR.2014.81.

Goyal, A., Bochkovskiy, A., Deng, J., Koltun, V., 2022. Non-deep networks. Adv. Neural Inf. Process. Syst. 35, 6789–6801. ⟨https://doi.org/10.48550/arXiv.2110.07641⟩.

Guo, W., Carroll, M.E., Singh, A., Swetnam, T.L., Merchant, N., Sarkar, S., Singh, A.K., Ganapathysubramanian, B., 2021. UAS-based plant phenotyping for research and breeding applications. Plant Phenomics. https://doi.org/10.34133/2021/9840192.

Hameed, K., Chai, D., Rassau, A., 2021. Texture-based latent space disentanglement for enhancement of a training dataset for ANN-based classification of fruit and vegetables. Inf. Process. Agric. 10 (1), 85–105. https://doi.org/10.1016/j.inpa.2021.09.003.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of. the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Las Vegas, NV, USA. ⟨https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.90⟩.

Hu, X., Jing, L., Sehar, U., 2022. Joint pyramid attention network for real-time semantic segmentation of urban scenes. Appl. Intell. 52 (1), 580–594. https://doi.org/10.1007/s10489-021-02446-8.

Hu, K., Wang, Z., Coleman, G., Bender, A., Yao, T., Zeng, S., Song, D., Schumann, A., Walsh, M., 2021. Deep learning techniques for in-crop weed identification: a review. arXiv Prepr. arXiv 2103, 14872. ⟨https://doi.org/10.48550/arXiv.2103.14872⟩.

Jia, W., Tian, Y., Luo, R., Zhang, Z., Lian, J., Zheng, Y., 2020. Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot. Comput. Electron. Agric. 172, 105380 https://doi.org/10.1016/j.compag.2020.105380.

Jin, X., Che, J., Chen, Y., 2021b. Weed identification using deep learning and image processing in vegetable plantation. IEEE Access 9, 10940–10950. https://doi.org/10.1109/ACCESS.2021.3050296.

Jin, X., Han, K., Zhao, H., Wang, Y., Chen, Y., Yu, J., 2024. Detection and coverage estimation of purple nutsedge in turf with image classification neural networks. Pest Manag. Sci. https://doi.org/10.1002/ps.8055.

Jin, N., Jin, L., Luo, S., Tang, Z., Liu, Z., Wei, S., Liu, F., Zhao, X., Yu, J., Zhong, Y., 2021a. Comprehensive evaluation of amino acids and polyphenols in 69 varieties of green cabbage (*Brassica oleracea* L. var. *capitata* L.) based on multivariate statistical analysis. Molecules 26 (17), 5355. https://doi.org/10.3390/molecules26175355.

Jin, X., Sun, Y., Che, J., Bagavathiannan, M., Yu, J., Chen, Y., 2022. A novel deep learning-based method for detection of weeds in vegetables. Pest Manag. Sci. 78 (5), 1861–1869. https://doi.org/10.1002/ps.6804.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R., 2023. Segment anything. arXiv Prepr. arXiv:2304. 02643. ⟨https://doi.org/10.48550/arXiv.2304.02643⟩.

Kong, X., Chen, L., Wei, T., Zhou, H., Bai, C., Yan, X., Miao, Z., Xie, J., Zhang, L., 2020. Transcriptome analysis of biological pathways associated with heterosis in Chinese cabbage. Genomics 112 (6), 4732–4741. https://doi.org/10.1016/j.ygeno.2020.08.011.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25 https://doi.org/10.1145/3065386.

Kurtulmus, F., Lee, W.S., Vardar, A., 2011. Green citrus detection using 'eigenfruit', color and circular Gabor texture features under natural outdoor conditions. Comput. Electron. Agric. 78 (2), 140–149. https://doi.org/10.1016/j.compag.2011.07.001.

Liang, Y., Li, Y., Zhang, L., Liu, X., 2019. Phytochemicals and antioxidant activity in four varieties of head cabbages commonly consumed in China. Food Prod. Process. Nutr. 1 (1), 9. https://doi.org/10.1186/s43014-019-0003-6.

Lin, G., Tang, Y., Zou, X., Cheng, J., Xiong, J., 2020. Fruit detection in natural environment using partial shape matching and probabilistic Hough transform. Precis. Agric. 21, 160–177. https://doi.org/10.1007/s11119-019-09662-w.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of. the IEEE/CVF international conference on computer vision. IEEE. https://doi.org/10.1109/ICCV48922.2021.00986.

Lu, J., Lee, W.S., Gan, H., Hu, X., 2018. Immature citrus fruit detection based on local binary pattern feature and hierarchical contour analysis. Biosyst. Eng. 171, 78–90. https://doi.org/10.1016/j.biosystemseng.2018.04.009.

Ma, R., Hu, H., Xing, S., Li, Z., 2020. Efficient and fast real-world noisy image denoising by combining pyramid neural network and two-pathway unscented Kalman filter. IEEE Trans. Image Process. 29, 3927–3940. https://doi.org/10.1109/TIP.2020.2965294.

Ma, R., Li, S., Zhang, B., Fang, L., Li, Z., 2022a. Flexible and generalized real photograph denoising exploiting dual meta attention. IEEE Trans. Cybern. 53 (10), 6395–6407. https://doi.org/10.1109/TCYB.2022.3170472.

Ma, R., Li, S., Zhang, B., Hu, H., 2022b. Meta PID attention network for flexible and efficient real-world noisy image denoising. IEEE Trans. Image Process. 31, 2053–2066. https://doi.org/10.1109/TIP.2022.3150294.

Malik, M.H., Zhang, T., Li, H., Zhang, M., Shabbir, S., Saeed, A., 2018. Mature tomato fruit detection algorithm based on improved HSV and watershed algorithm. In: In Proceedings of the Elsevier International Federation of Automatic Control, 51. Elsevier, Beijing, China, pp. 431–436. https://doi.org/10.1016/j.ifacol.2018.08.183.

Matthews, S., Noli, E., Demir, I., Khajeh-Hosseini, M., Wagner, M.H., 2012. Evaluation of seed quality: from physiology to international standardization. Seed Sci. Res. 22 (S1), S69–S73. https://doi.org/10.1017/S0960258511000365.

Nanehkaran, Y.A., Zhang, D., Chen, J., Tian, Y., Al-Nabhan, N., 2020. Recognition of plant leaf diseases based on computer vision. J. Ambient Intell. Humaniz. Comput. 1–18. https://doi.org/10.1007/s12652-020-02505-x.

Oo, L.M., Aung, N.Z., 2018. A simple and efficient method for automatic strawberry shape and size estimation and classification. Biosyst. Eng. 170, 96–107. https://doi.org/10.1016/j.biosystemseng.2018.04.004.

Rahman, S.U., Alam, F., Ahmad, N., Arshad, S., 2023. Image processing based system for the detection, identification and treatment of tomato leaf diseases. Multimed. Tools. Appl. 82 (6), 9431–9445. https://doi.org/10.1007/s11042-022-13715-0.

Reddy, K.R., Brand, D., Wijewardana, C., Gao, W., 2017. Temperature effects on cotton seedling emergence, growth, and development. Agron. J. 109 (4), 1379–1387. https://doi.org/10.2134/agronj2016.07.0439.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Las Vegas, NV, USA. https://doi.org/10.1109/CVPR.2016.91.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: towards real-time object detection with region proposal networks. Adv. Neural Inf. Process. Syst. 28. ⟨https://doi.ieeecomputersociety.org/10.1109/TPAMI.2016.2577031⟩.

Rojas, C.P., Guzmán, L.S., Toledo, N.V., 2017. Weed recognition by SVM texture feature classification in outdoor vegetable crops images. Ing. e Investig. 37 (1), 68–74. https://doi.org/10.15446/ing.investig.v37n1.54703.

Saleem, M.H., Potgieter, J., Arif, K.M., 2021. Automation in agriculture by machine and deep learning techniques: a review of recent developments. Precis. Agric. 22, 2053–2091. https://doi.org/10.1007/s11119-021-09806-x.

Septiarini, A., Sunyoto, A., Hamdani, H., Kasim, A.A., Utaminingrum, F., Hatta, H.R., 2021. Machine vision for the maturity classification of oil palm fresh fruit bunches based on color and texture features. Sci. Hortic. 286, 110245 https://doi.org/10.1016/j.scienta.2021.110245.

Tan, K., Lee, W.S., Gan, H., Wang, S., 2018. Recognising blueberry fruit of different maturity using histogram oriented gradients and colour features in outdoor scenes. Biosyst. Eng. 176, 59–72. https://doi.org/10.1016/j.biosystemseng.2018.08.011.

Tan, C., Li, C., He, D., Song, H., 2022. Towards real-time tracking and counting of seedlings with a one-stage detector and optical flow. Comput. Electron. Agric. 193, 106683 https://doi.org/10.1016/j.compag.2021.106683.

Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., Liang, Z., 2019. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. Comput. Electron. Agric. 157, 417–426. https://doi.org/10.1016/j.compag.2019.01.012.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30. ⟨https://doi.org/10.48550/arXiv.1706.03762⟩.

Wang, Z., Li, Y.K., Guo, W.Z., Xu, Z.G., Wang, L.C., Ma, L., 2019. Yield, nitrogen use efficiency and economic benefits of biochar additions to Chinese flowering cabbage in Northwest China. Nutr. Cycl. Agroecosystems 113, 337–348. https://doi.org/10.1007/s10705-019-09971-6.

Wojke, N., Bewley, A., Paulus, D., 2017. Simple online and realtime tracking with a deep association metric. Paper read at 2017. IEEE International Conference on Image Processing (ICIP). IEEE, Beijing, China. https://doi.org/10.1109/ICIP.2017.8296962.

Wu, D., Lv, S., Jiang, M., Song, H., 2020. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. Comput. Electron. Agric. 178, 105742 https://doi.org/10.1016/j.compag.2020.105742.

Wu, J., Zhang, B., Zhou, J., Xiong, Y., Gu, B., Yang, X., 2019. Automatic recognition of ripening tomatoes by combining multi-feature fusion with a bi-layer classification strategy for harvesting robots. Sensors 19 (3), 612. https://doi.org/10.3390/s19030612.

Xiao, J.R., Chung, P.C., Wu, H.Y., Phan, Q.H., Yeh, J.L.A., Hou, M.T.K., 2020. Detection of strawberry diseases using a convolutional neural network. Plants 10 (1), 31. https://doi.org/10.3390/plants10010031.

Ye, Z., Yang, K., Lin, Y., Guo, S., Sun, Y., Chen, X., Lai, R., Zhang, H., 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. Agric. 209, 107822. ⟨https://doi.org/10.1016/j.compag.2023.107822⟩.

Yu, L., Xiong, J., Fang, X., Yang, Z., Chen, Y., Lin, X., Chen, S., 2021. A litchi fruit recognition method in a natural environment using RGB-D images. Biosyst. Eng. 204, 50–63. https://doi.org/10.1016/j.biosystemseng.2021.01.015.

Zhang, J., He, L., Karkee, M., Zhang, Q., Zhang, X., Gao, Z., 2018. Branch detection for apple trees trained in fruiting wall architecture using depth features and Regions-Convolutional Neural Network (R-CNN). Comput. Electron. Agric. 155, 386–393. https://doi.org/10.1016/j.compag.2018.10.029.

Zou, Y., Yu, H., Zhang, L., Ruan, Z., 2021. Dietary vegetable powders modulate immune homeostasis and intestinal microbiota in mice. Foods 11 (1), 27. https://doi.org/10.3390/foods11010027.