ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

Lightweight YOLOv5s-Based Machine Vision System for Real-Time Potato Defect Detection

Fajun Miao[1, 2], Jinzhu Lu[1, 3, *], Senping Liu[1, 2], Qiyang Shui[1, 2]

[1]Modern Agricultural Equipment Research Institute, Xihua University, 610039, Chengdu, China

[2]School of Mechanical Engineering, Xihua University,610039, Chengdu, China

[3]School of Aeronautics and Astronautics and Intelligent Equipment, Xihua University, Chengdu 610039, China

[*] Corresponding author

Abstract

To solve the drawbacks of traditional potato defect detection, such as heavy manual dependence, low efficiency, strong subjectivity and poor adaptability to complex scenarios, this study proposes a real-time machine vision detection system based on improved lightweight YOLOv5s. In this research, the YOLOv5s model is lightweightly modified, and GhostNet is used to replace the original CSPDarknet backbone network, which reduces model complexity while retaining the core feature extraction capability. The experimental results show that on the self-built dataset covering four types of potato defects including damage, sprouting, scab and dry rot, the Precision, Recall and mean Average Precision (mAP) of the improved YOLOv5s model reach 94.2%, 92.2% and 95.7% respectively, with the model size reduced by 26%. For the real-time detection of potato defects, the BoT-SORT multi-object tracking algorithm is integrated into the YOLOv5 detection framework. Compared with classic multi-object tracking algorithms including SORT, DeepSORT and ByteTrack, BoT-SORT achieves the best overall performance and obtains optimal values in three core evaluation metrics of HOTA, MOTA and IDF1, which are 95.3%, 99.7% and 98.8% correspondingly. Additional experimental verification demonstrates that when the conveyor belt speed is no higher than 62 mm/s, the improved lightweight model maintains a detection accuracy of no less than 94.3% and a tracking accuracy of 100%, and its operating speed fully meets the requirements of real-time detection. The designed system comprehensively balances detection accuracy, lightweight performance and real-time performance, which provides a reliable reference for meeting the practical demands of online real-time potato defect detection.

Keywords

Lightweight; YOLOv5; potato defects; real-time detection.

1. INTRODUCTION

Potato is an annual herb of the Solanum genus in the Solanaceae family. It possesses excellent characteristics including cold resistance, drought tolerance and barren resistance, and has wide adaptability to diverse environments, making it one of the most extensively cultivated vegetables across the globe[1].Meanwhile, potatoes are rich in nutrients such as starch, protein and potassium. As an important cash crop, it plays a vital role in improving people's dietary quality and physical health[2]. However, potatoes are vulnerable to diseases, pests and mechanical damage during growth, harvesting and storage processes, which cause surface

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

defects and further undermine their commercial value[3]. Traditional potato defect detection mainly depends on manual inspection, which is plagued by high labor intensity, low efficiency and inconsistent detection results. Therefore, high-quality and high-efficiency external defect detection technologies adapted to modern production can effectively enhance the economic benefits of potatoes.

With the expansion of industrial demands and the advancement of information technology, non-destructive detection technologies for agricultural product defects have developed rapidly. Based on the intensity differences in absorption, reflection, transmission and scattering of electromagnetic waves at different wavelengths by hydrogen-containing X-H groups (such as C-H, O-H and N-H), spectral imaging technology can simultaneously collect image and spectral data of samples, which is highly suitable for the detection of internal and external defects of potatoes. Hyperspectral imaging covers dozens to hundreds of narrow wavelength bands and captures abundant spectral information, supporting high-precision analysis for potato defect identification. However, redundant data exists in numerous wavebands, and the massive data volume greatly increases the difficulty and time consumption of data processing, resulting in poor applicability in real-time detection scenarios[4,5]. In contrast, multispectral imaging produces significantly less data and achieves better real-time performance in potato defect detection[6]. Although RGB image-based detection technology cannot detect internal defects like spectral imaging methods, it presents outstanding advantages in economic cost and detection speed[7]. In the real-time positioning and maturity detection of tomatoes, the improved deep learning algorithm reduced the model size by 78% while maintaining a mAP of 96.9%, and the detection frame rate reached 26.5 FPS. It effectively satisfies the low-cost and low-computing-power real-time detection requirements of mobile terminals[8]. Considering the actual demands for real-time performance and cost control in potato defect detection, this study adopts an RGB image-based defect detection method.

Defect detection based on RGB images mainly relies on two technical approaches: traditional machine learning and deep learning. The general workflow of traditional machine learning detection consists of image acquisition, image preprocessing, image segmentation, feature extraction, and recognition and classification[9]. Since traditional machine learning models take manually extracted features as input, feature extraction becomes the core link that determines the final detection performance. Feature extraction refers to acquiring image attributes such as color, shape and texture. For instance, in the detection of irregular potatoes, multiple shape features are extracted, and the most distinguishable features for classification are selected through stepwise linear discriminant analysis[10]. Image segmentation before feature extraction can extract the region of interest and eliminate irrelevant background interference. For apple surface defect detection, the region growing algorithm is firstly applied to segment defective areas, and then the gray level co-occurrence matrix is used to extract texture features, achieving a final detection accuracy of 94.2%[11]. Furthermore, optimal feature selection after extraction is also critical for accurate potato defect detection, with typical optimization algorithms including the Buteo Optimization Algorithm (BUOZA) and Adaptive Boosting (AdaBoost)[12]. In the recognition and classification stage, mainstream traditional machine learning algorithms include Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Naive Bayes. The SVM model solves qualitative classification and quantitative regression by constructing an optimal separating hyperplane. It exhibits excellent generalization performance with limited training samples and maintains high accuracy under low-dimensional conditions. Nevertheless, its classification performance will be degraded when classification boundaries are ambiguous or severe inter-class overlap exists[13]. In the detection of citrus surface defects, the Artificial Bee Colony (ABC) algorithm is adopted to optimize the kernel function of SVM, and the optimized SVM achieves a recognition accuracy of 98.45% for citrus surface defects[14].

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

As an important branch of machine learning, deep learning shows outstanding performance and great potential in image processing and data analysis across various fields. In agricultural applications, deep learning provides superior solutions for weed identification, crop classification and defect detection compared with conventional image processing technologies[15]. Different from traditional machine learning, deep learning can automatically extract hierarchical features from raw images via multi-layer network structures without manual feature engineering, which is its core advantage[16]. Current mainstream deep learning detection models are divided into two categories: one-stage and two-stage detectors. One-stage models such as the YOLO series directly output object categories and bounding boxes from input images. By contrast, two-stage models like the R-CNN series first generate candidate regions, and then implement region classification and bounding box regression. Without candidate region generation, one-stage models require fewer computational resources and realize faster detection speed[17]. Meanwhile, the YOLO series further improves inference efficiency by introducing efficient backbone networks (e.g., Darknet), advanced activation functions (e.g., Mish) and effective convolution modules (e.g., SPP, PANet)[18]. Although two-stage detectors run at a lower speed, their classifiers can focus on fine-grained classification of high-quality candidate regions, thus presenting better detection accuracy[19]. In recent years, deep learning has been widely studied in potato defect detection. To detect potato sprouting and rot, Dai et al.[20]replaced the standard convolution (Conv) and SPP module in YOLOv5 with Cross Convolution (CrossConv) and Spatial Pyramid Pooling Fast (SPPF) to enhance feature representation and fusion capability. In addition, the 9-Mosaic data augmentation algorithm was adopted to boost model generalization; the genetic algorithm-based k-means clustering was used to reconstruct anchor boxes for enhancing small-object feature expression; multi-scale training and hyperparameter evolution strategies were introduced to optimize anchor localization accuracy. The improved algorithm obtained a significant improvement in detection precision.

The above studies have made remarkable contributions to the development of non-destructive detection for agricultural product defects and formulated numerous effective defect detection schemes based on different technical methods. However, the non-destructive detection of external potato defects still confronts multiple challenges. Firstly, potatoes have various types of surface defects. In particular, defects such as scab are similar to potatoes in surface color, which greatly increases the difficulty of recognition. Secondly, edge devices are limited in computing power, restricting their capability for potato defect detection. Finally, most existing studies on potato defect detection pay little attention to real-time detection, and relevant real-time detection technologies still have certain limitations.

To solve the above problems, this study takes potatoes with four types of surface defects, namely damage, sprouting, scab and dry rot, as the research objects. A dataset consisting of 469 images and 10 videos is constructed to train and verify the defect detection model and tracker. To satisfy the demand for real-time potato defect detection, a detection procedure illustrated in Figure 1 is designed. The one-stage detection model YOLOv5 is used for potato defect detection, and the BoT-SORT multi-object tracking algorithm is adopted to achieve real-time tracking and counting of potatoes.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

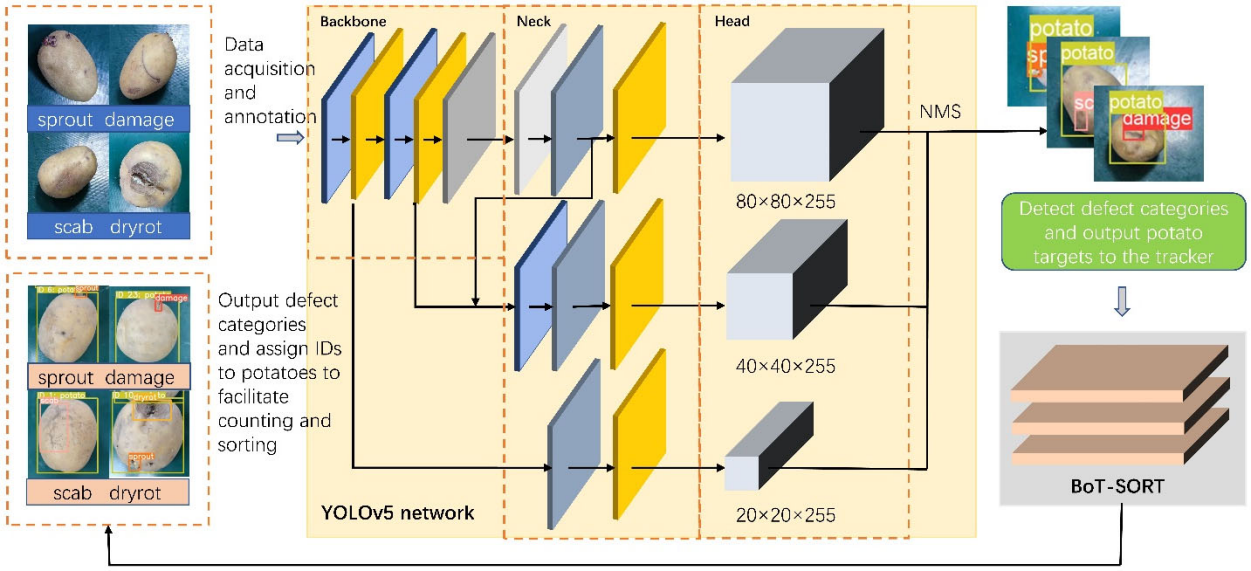


Figure 1. Potato defect detection process

2. MATERIALS AND METHODS

2.1. Dataset Construction

2.1.1 Acquisition of Potato Defect Images

Experimental samples were purchased from local Walmart supermarkets and farmers' markets. A total of 182 potatoes were selected and screened by professional personnel, covering four typical surface defects: damage, sprouting, scab and dry rot. Before image collection, all samples were simply cleaned to remove surface impurities. The defect categories and locations were reconfirmed by professionals to ensure the accuracy of sample annotation.

Image acquisition was completed indoors under stable lighting conditions. A Xiaomi 10 mobile phone was fixed at a height of 25 cm above the conveyor belt to capture images of single potatoes and multiple potatoes arranged in a single row. The image resolution was 1706×1279, and the acquired static images were used for the training of the detection model. To verify the real-time detection performance of the system, potato samples were placed in a single row on the conveyor belt. Four conveyor belt operating speeds were set at 26 mm/s, 62 mm/s, 81 mm/s and 99 mm/s, respectively. With a frame rate of 30 FPS, the Xiaomi 10 was used to shoot 8 videos with a duration of 40 seconds each.

In addition, to test the performance of the subsequent tracking algorithm, 2 videos of disorderly placed and stacked potatoes were collected at the speed of 62 mm/s, with a duration of 20 seconds and a frame rate of 30 FPS. Figure 2 shows the typical images of potatoes with the four types of defects.

a.damage b.sprout c.scab d.dryrot

Figure 2. Potato defect images

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

2.1.2 Dataset Establishment

After image acquisition, the labeling tool LabelImg was adopted to annotate potato defects. The labeled categories included damage, sprouting, scab, dry rot and normal potato, among which the potato category was set for object tracking by the tracker in subsequent real-time detection tasks. To prevent bounding box drift, all annotation files were saved in the Pascal VOC format and uniformly converted to the YOLO format for model training.

To enhance the generalization ability and robustness of the model, five data augmentation strategies were applied to the 469 original images in the dataset, including noise addition, brightness variation, cropping, translation and rotation. Ultimately, an expanded dataset with 2574 images was constructed. The samples after data augmentation are shown in Figure 3.

Figure 3. Potato images after data augmentation

During the training process, the dataset was divided into a training set and a validation set at a ratio of 8:2. The distribution of various potato defect categories in the dataset is shown in Table 1. The quantity of each defect type is relatively balanced, so as to avoid negative impacts of unbalanced data distribution on model training. The collected videos were adopted to evaluate the tracking stability and reliability of BoT-SORT for potatoes in real-time detection after model training.

Table 1. Composition of potato defect dataset

Classes	Train	Val	Total
damage	1056	264	1320
sprout	1251	313	1564
scab	1425	357	1782
dryrot	771	193	964

2.2. Model Improvement

2.2.1 YOLOv5

As typical representatives of one-stage object detection models, the YOLO series algorithms possess the advantages of fast inference speed, high accuracy and strong adaptability, and have been widely applied in the recognition and classification, disease detection, and quality evaluation of agricultural products[21]. To select the optimal baseline model suitable for potato defect detection, five versions of YOLOv5, together with YOLOv7-tiny and YOLOv8s, are adopted to train the dataset in this study. The quantitative results are presented in Table 2.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

As illustrated in Table 2, different YOLO models show a typical trade-off among detection accuracy, inference speed and lightweight performance. Among them, YOLOv5s achieves the best mAP@0.5, yet it suffers from low inference speed and large model size. In contrast, YOLOv5n exhibits prominent lightweight advantages with an ultra-small model size of 3.9 and the fastest inference speed, while its mAP@0.5 is relatively limited. Notably, YOLOv5s achieves an excellent balance across all evaluation metrics. It obtains an mAP@0.5 of 0.976, with model size and inference speed maintained at a reasonable level. This model can not only guarantee the recognition accuracy of potato defects, but also provide favorable deployment flexibility, making it more suitable for the detection requirements in most practical production scenarios.

Table 2. Training results of YOLO series algorithms

Model	Precision (%)	Recall (%)	F1-score (%)	Model Size (MB)	mAP@0.5 (%)
YOLOv5x	0.956	0.95	0.953	173.1	0.974
YOLOv5l	0.962	0.947	0.954	92.9	0.968
YOLOv5m	0.948	0.939	0.943	42.2	0.964
YOLOv5s	0.977	0.959	0.968	14.4	0.972
YOLOv5n	0.877	0.823	0.849	3.9	0.877
YOLOv7-tiny	0.935	0.918	0.926	12.3	0.963
YOLOv8s	0.971	0.967	0.969	22.5	0.97

Compared with other algorithms, YOLOv5 is also a mature object detection algorithm. It is supported by an active community and continuous updates, and provides abundant resources and development tools. In this study, YOLOv5s is adopted as the detection model for potato defect detection. In terms of network structure, the backbone network applies CSPDarkNet, which ensures the inference speed and detection accuracy while reducing the model size. The neck network adopts PANet (Path Aggregation Network) to achieve better fusion of extracted features. The detection head refers to the structure of YOLOv4 and adopts an anchor-based mechanism.

Compared with other models, YOLOv5s effectively balances computational complexity and detection accuracy, making it suitable for the real-time detection of potato defects in this research. Figure 4 shows the network structure of YOLOv5.

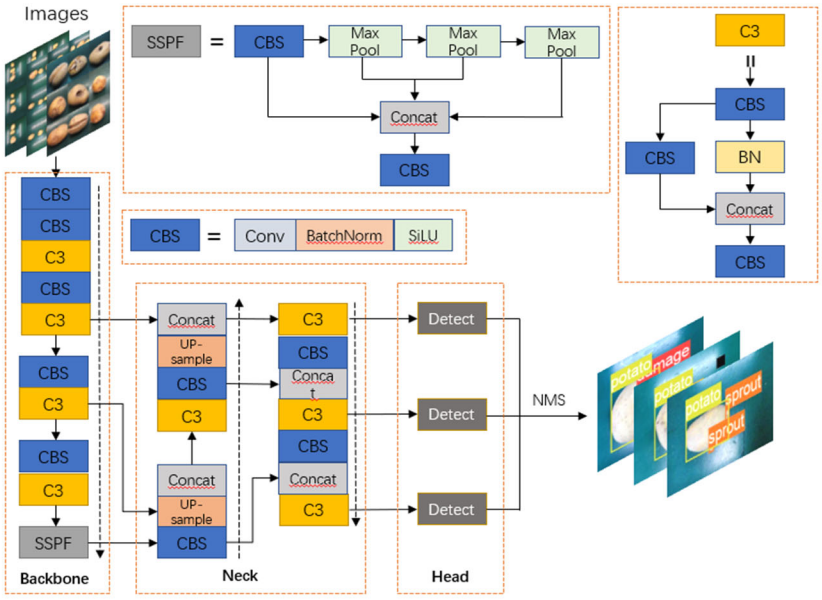


Figure 4. YOLOv5 network structure diagram

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

2.2.2 Lightweight Improvement of YOLOv5 Backbone Network

YOLOv5 adopts CSPDarknet as its backbone feature extraction network. It achieves strong object feature representation capability through feature split and fusion of Cross Stage Partial (CSP) modules. Nevertheless, the design of numerous standard convolutional layers and multi-branch residual modules in this structure results in large model parameters and high computational complexity, which limits its direct deployment on portable terminals widely used in agricultural product detection scenarios[22].

CSPDarknet is a backbone network optimized on the basis of Darknet53, and its core innovation lies in the embedding of CSP modules. The input feature maps of each stage are divided into two branches along the channel dimension. One branch passes through multiple cascaded residual units in turn, and each unit takes "convolution + Batch Normalization (BN) + SiLU activation function" as the basic component. The other branch retains the original feature information through shortcut connection. Finally, the two branches of features are concatenated in the channel dimension, and feature integration is realized by convolutional layers. This design not only improves the efficiency of gradient propagation, but also increases the feature reuse rate, enabling YOLOv5 to capture texture, shape and other target features more accurately[23].

Detection terminals commonly used in agricultural detection scenarios, such as embedded development boards, are generally constrained by limited computing power and tight memory resources. A large number of 3×3 standard convolutions in CSPDarknet bring high computational complexity, and the stacking of multi-branch residual modules also leads to excessive model parameters. As a result, the inference speed of the original model on edge detection devices is usually less than 10 FPS, accompanied by reduced detection accuracy, which cannot meet the requirements of high-precision real-time detection. Accordingly, targeting potato defect detection tasks, this study conducts lightweight optimization on the YOLOv5 backbone network. The original CSPDarknet is replaced by GhostNet to adapt to practical detection terminals for potato detection, so as to balance detection accuracy and real-time performance.

2.2.3 Ghostnet

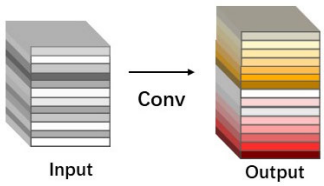
Traditional convolutional neural networks involve massive redundant computation in feature map generation. Numerous highly similar feature maps consume substantial computing resources. The lightweight neural network GhostNet is constructed with Ghost bottlenecks. Stacked by Ghost convolution modules, Ghost bottlenecks generate abundant feature maps with fewer parameters, serving as a novel plug-and-play module[24].

The structure of a traditional convolutional layer is illustrated in Figure 5(a). By contrast, as shown in Figure 5(b), the Ghost convolution module divides conventional convolution into two steps. In the first step, a small number of convolutional kernels are adopted for standard convolution operations to generate a set of low-dimensional fundamental feature maps. This design eliminates redundant convolution kernels in traditional convolution, greatly reducing parameters and computational cost, and the obtained basic feature maps contain core semantic information of targets. In the second step, for the fundamental feature maps generated in the previous step, 3×3 or 5×5 lightweight kernels are used for channel-wise depthwise convolution. Spatial transformation is performed on each channel to generate multiple Ghost feature maps with similar structures but differentiated details[25]. Afterwards, identity mapping is applied to the basic feature maps, which are concatenated with the expanded Ghost feature maps in the channel dimension. Finally, the fused feature maps maintain the same dimension as those of traditional convolution, yet with much lower computational cost.

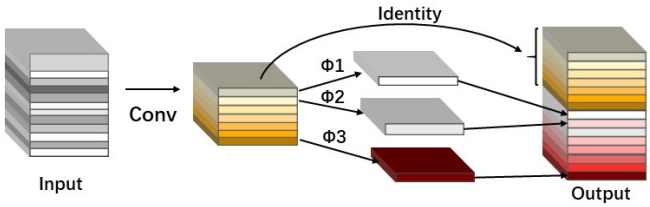
As the core bottleneck structure of GhostNet, the Ghost Bottleneck is composed of stacked Ghost convolution modules. Overall, it shares similarities with the residual blocks in ResNet,

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

and both introduce residual connections to alleviate the gradient vanishing problem during the training of deep networks. Different from conventional residual blocks, however, Ghost Bottleneck abandons fully connected layers and stacked large-size convolution kernels, and realizes feature transformation and fusion entirely through lightweight Ghost convolution modules. The introduction of residual connections enables shallow underlying features to be directly transmitted to high-level networks, further improving feature reuse efficiency. It ensures that the model achieves lightweight optimization without significant degradation in detection accuracy and classification performance.



a the convolution layer



b Ghost module

Figure 5. Schematic Diagram of Convolutional Layer and Ghost Module with the Same Number of Output Feature Maps

2.2.4 BoT-SORT

To realize real-time potato defect detection, potato counting and defect statistics, this study integrates the BoT-SORT multi-object tracking algorithm into the detection module of YOLOv5. As a detection-based multi-object tracking algorithm, BoT-SORT is designed to improve the precision and efficiency of object tracking tasks. It is an improved version of the classic SORT algorithm, which integrates advanced feature extraction and matching mechanisms to cope with target occlusion, appearance changes and motion uncertainty in complex scenarios.

BoT-SORT adopts deep learning to extract appearance features for describing target visual information. These features are obtained through pre-trained neural networks and adopted to solve the problem of tracking loss caused by target occlusion and cross movement, enabling BoT-SORT to better balance target appearance similarity and motion information. Meanwhile, it adopts a multi-task joint optimization strategy. Different from the original SORT algorithm, which only relies on the Kalman filter to predict target motion trajectories, BoT-SORT comprehensively considers spatiotemporal dynamic information and visual feature matching information during optimization[26]. This joint strategy reduces errors caused by over-reliance on the motion model, especially when targets accelerate or move in a non-linear manner.

The superior performance of BoT-SORT stems from its critical improvements to traditional MOT methods. It optimizes the matching strategy by fusing appearance features, so that the algorithm can achieve higher tracking accuracy in complex environments. Compared with classic SORT and DeepSORT, BoT-SORT presents a better comprehensive balance and achieves significant improvements in tracking accuracy, anti-occlusion capability and computational efficiency[27].

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

In the process of potato defect detection and tracking, YOLOv5 identifies individual potatoes, as well as their defect categories and defect areas. Subsequently, the BoT-SORT algorithm is used for independent object tracking of each potato. In this way, BoT-SORT can achieve accurate tracking of individual potatoes in potato defect detection tasks, effectively avoiding target loss and tracking errors. It can also resume tracking when occluded targets reappear, thereby completing the whole process of detection, tracking and counting.

2.3. Experimental Environment and Evaluation Metrics

The software environment adopts the PyTorch 1.11.0 deep learning framework, Windows 10 Professional 64-bit operating system, and Python 3.8.10 programming language. The hardware platform is configured as follows: CPU: Intel(R) Core(TM) i9-10900K @ 3.70GHz; GPU: NVIDIA GeForce RTX 3090; RAM: 64 GB.

In the training process, the total number of iterations is set to 100 epochs. The initial learning rate, momentum, weight decay and other hyperparameters follow the original settings of YOLOv5. To verify the detection performance of the improved YOLOv5 for potato defects in this study, four evaluation metrics are adopted, including Precision (P), Recall (R), mean Average Precision at the IoU threshold of 0.5 (mAP50), and F1. These indicators have been widely applied in various object detection tasks. The calculation formulas are defined as follows:

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

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

$$mAP50 = \frac{\sum_{i=1}^{N} AP_{50}^{(i)}}{N} \tag{3}$$

$$F_1 = \frac{2PR}{P + R} \tag{4}$$

In the formulas, TP (True Positive) represents the number of positive samples; FP (False Positive) refers to negative samples that are incorrectly predicted as positive; FN (False Negative) denotes positive samples that are mistakenly predicted as negative. N is the total number of categories, and AP50(i) indicates the average precision of category i at an IoU threshold of 0.5, where IoU is defined as the ratio of intersection to union between the actual defect area and the predicted defect area.

When validating the tracking algorithm, the mainstream evaluation metrics for multi-object tracking mainly include HOTA (Higher Order Tracking Accuracy), MOTA (Multiple Object Tracking Accuracy), IDF1 (Identification F1-Score) and FPS (Frames Per Second). As a high-dimensional evaluation indicator for multi-object tracking, HOTA can comprehensively and scientifically assess the overall performance of tracking algorithms. The calculation formula of HOTA is shown in Formula (5), which involves the detection accuracy score and association accuracy score. The definitions of TP, FP and FN are consistent with those above. Moreover, the corresponding parameter represents the number of successfully tracked potato samples among true positives, which is used to characterize the association accuracy of target matching.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

$$\mathrm{HOTA} = \sqrt{\mathrm{DetA} \times \mathrm{AssA}} = \sqrt{\frac{\sum_{c \in TP} A(c)}{TP + FP + FN}} \tag{5}$$

The tracking stability of the same target after matching in the tracking module is represented by relevant variables and calculated by Formula (6). In this equation, FP denotes the total number of false detections in frame t; FN represents the total number of missed detections in frame t; the number of ID switches in frame t is defined by the corresponding parameter, and the number of ground-truth objects in frame t is expressed by the corresponding variable.

$$\mathrm{MOTA} = 1 - \frac{\sum_t \left(FP + FN + IDSW \right)}{\sum_t g_t} \tag{6}$$

This metric is adopted to measure the ID consistency and trajectory stability of the tracker. A higher value indicates more accurate tracking, a lower probability of target loss, and fewer ID switches. The calculation formula of IDF1 is shown in Equation (7). IDTP refers to the total number of correctly tracked targets with correct ID matching; IDFP represents the total number of falsely tracked targets with wrong ID association; IDFN denotes the total number of targets suffering from tracking ID loss.

$$IDF1 = \frac{2IDTP}{2IDTP + IDFP + IDFN} \tag{7}$$

3. RESULTS

3.1. Comparative Experiments on Different Backbone Networks

To verify the optimization effect of the GhostNet-based lightweight improvement strategy for the YOLOv5s backbone network, a series of comparative experiments were designed in this study. Three typical lightweight neural networks in the lightweight research field, namely EfficientNet, ShuffleNetV2 and MobileNetV3, were selected for comparison. Each of the three networks and GhostNet was adopted to replace the original backbone feature extraction network of the YOLOv5s model, constructing four object detection models with different backbones.

All models were trained with the same dataset, training hyperparameters and hardware environment. The training variation curves of mAP values of each model are shown in Figure 6. The performance indicators and model size of each model were systematically compared, and the relevant training results are listed in Table 3.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

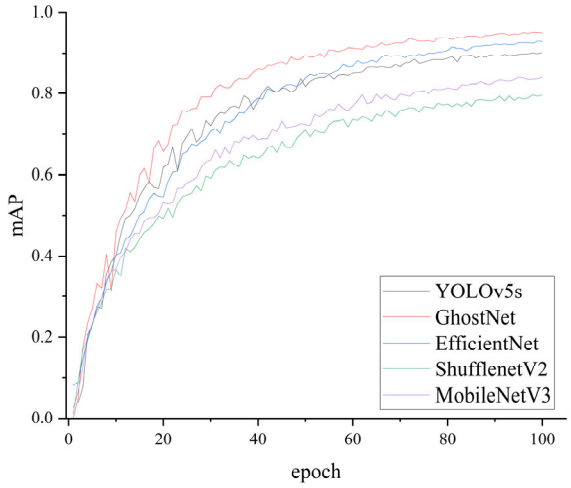


Figure 6. mAP Variation Curves of Models with Different Backbone Networks

MobileNetV3 takes depthwise separable convolution as the core foundation, and integrates the lightweight SE channel attention mechanism and neural architecture search. It is specially optimized for inference on edge and mobile terminals, with strong deployment universality and controllable inference latency. ShuffleNetV2 adopts channel splitting and channel shuffling as its core innovations, which minimizes network memory access overhead and theoretical computational consumption. Oriented to extreme model compression, it is highly adaptable to embedded environments with extremely limited computing power.EfficientNet applies a systematic compound scaling design paradigm to globally and collaboratively optimize network depth, width and input resolution. It prioritizes strengthening high-level semantic feature extraction and delivers outstanding detection accuracy, yet accompanied by relatively high overall computational load. With the innovative Ghost module, GhostNet generates equivalent feature maps via low-cost linear operations and greatly reduces redundant convolution computation. It maintains powerful feature representation capability while realizing model compression, achieving an optimal balance between feature extraction performance and lightweight characteristics.

The performance metrics and model size of each model were compared systematically, and the relevant training results are shown in Table 3.

Table 3. Training results of diffrent backbone networks

Model	Precision (%)	Recall (%)	F1-score (%)	Model Size	mAP@0.5 (%)
GhostNet	0.942	0.922	0.917	10.6	0.957
EfficientNet	0.91	0.854	0.881	11.5	0.914
ShufflenetV2	0.788	0.708	0.746	2.0	0.746
MobileNetV3	0.828	0.714	0.767	3.2	0.796

It can be observed from the data in the table that the size of each model is reduced after the replacement of the YOLOv5 backbone network. The model equipped with the GhostNet backbone delivers the best overall detection performance in terms of Precision, Recall, F1-score and mAP@0.5. ShuffleNetV2 exhibits distinctive advantages in inference speed and model volume, achieving the highest degree of lightweight design.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

Compared with the original YOLOv5s, the model with GhostNet backbone is reduced by 26 percentage points in size, while its mAP@0.5 only drops by 1.5 percentage points, with high Precision and Recall well preserved. This result fully validates the effectiveness of replacing CSPDarknet with GhostNet for the lightweight optimization of YOLOv5s, which provides an efficient and feasible solution for practical potato defect detection.

3.2. Comparative Experiments of Different Tracking Algorithms

In the real-time detection scenario of potato defects, potatoes are often placed disorderly with mutual stacking and partial occlusion. To realize stable and continuous tracking of individual potatoes, the BoT-SORT multi-object tracking algorithm is selected in this study to complete target association and trajectory maintenance. To comprehensively verify the tracking robustness and practical application value of BoT-SORT in dense potato stacking scenarios, video sequences of disorderly placed, stacked and occluded potatoes collected in real scenarios are adopted as test samples.

Meanwhile, three mainstream and representative multi-object tracking algorithms including SORT, DeepSORT and ByteTrack are introduced for horizontal comparative experiments. SORT relies on Kalman filtering and IOU matching for target association, which owns a simple structure and fast inference speed, but is prone to ID switches and target loss under occlusion and dense overlap. DeepSORT adds an appearance feature extraction branch on the basis of SORT to improve the tracking performance in occluded scenarios, yet its additional feature computation brings higher computational cost and limited real-time performance. ByteTrack optimizes the association logic through hierarchical matching of high and low confidence targets, balancing tracking accuracy and inference efficiency with excellent generalization ability. By contrast, BoT-SORT further integrates temporal motion information and enhanced appearance feature matching strategies, making it more suitable for target tracking under complex interference such as stacking and intersection.

Comparative tests of the four algorithms are conducted under the same test videos, detection model and operating environment, and the quantitative experimental results of each evaluation index are presented in Table 4.

Table 4. Experimental Results for Various Tracking Algorithms

Tracking Algorithms	HOTA (%)	MOT (%)	IDF1 (%)	FPS
BoT-SORT	95.3	99.7	98.8	68.2
SORT	92.1	97.3	97.5	92.4
DeepSORT	94.2	98.2	98.2	62.1
ByteTrack	94.6	99.1	97.9	89.6

Experimental results show that the four algorithms present distinct performance gaps across the four evaluation metrics. BoT-SORT obtains the optimal values in three core accuracy indicators including HOTA, MOTA and IDF1, reaching 95.3%, 99.7% and 98.8% respectively. Its MOTA score is nearly full marks, which means the algorithm scarcely causes missed detections and false alarms in potato defect detection. The superior IDF1 performance reflects its remarkable advantage of extremely few ID switches. Combined with the HOTA score of 95.3%, it is fully verified that BoT-SORT achieves the optimal balance between detection accuracy and trajectory association consistency.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

SORT boasts the best real-time performance with a frame rate of 92.4 FPS, yet its HOTA and IDF1 values are the lowest among the four algorithms, revealing poor robustness in target trajectory association under complex scenarios. Benefiting from the capability of appearance feature extraction, DeepSORT reaches an IDF1 of 98.2%. Nevertheless, the excessive computational cost of feature extraction leads to the lowest FPS of 62.1, and its comprehensive performance in HOTA and MOTA is inferior to that of BoT-SORT. ByteTrack achieves a favorable trade-off between accuracy and speed, with a HOTA of 94.6%, a MOTA of 99.1% and a high FPS of 89.6, placing it in the middle tier in overall performance.

Considering the strict tracking accuracy requirements of the potato defect sorting scenario and the performance demands of practical engineering deployment, BoT-SORT is ultimately determined as the core solution for the target tracking module in this study. It can accurately realize real-time tracking and long-term ID maintenance of individual defective potatoes, which provides solid technical support for the efficient operation of the subsequent sorting system. The tracking visualization results of BoT-SORT are shown in Figure 7.

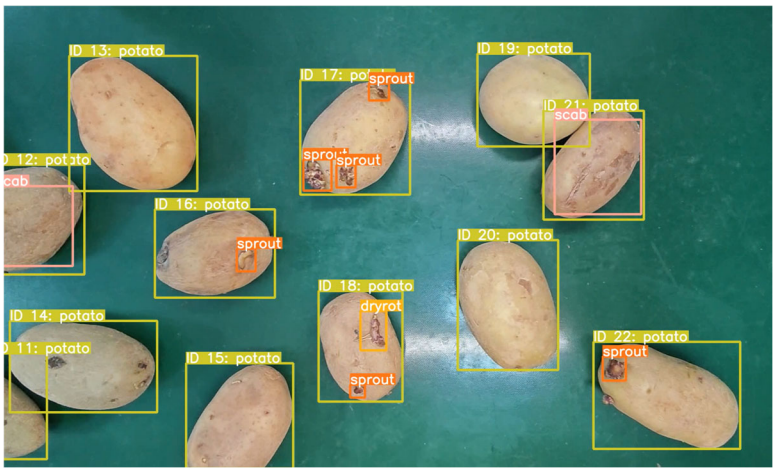


Figure 7. BoT-SORT Tracking Performance

3.3. Validation of Potato Defect Detection and Tracking Performance

To verify the defect detection accuracy and real-time performance of the lightweight improved YOLOv5s model with GhostNet backbone, as well as the tracking capability of the BoT-SORT algorithm, experiments were carried out on 8 videos selected from the dataset. Under four different conveyor belt speeds of 26 mm/s, 62 mm/s, 81 mm/s and 99 mm/s, the detection accuracy and tracking accuracy of the model were statistically analyzed. The corresponding experimental results are shown in Table 5.

Table 5. Real-time detection results of the model

Belt Speed (mm·s -1)	Detection Accuracy	Tracking Accuracy	Sample Quantity
26	96.5%	100%	42
62	94.3%	100%	42
81	89.4%	95.2%	42
99	78.2%	88.1%	42

As shown in Table 4, the potato defect detection accuracy presents a declining trend with the increase of conveyor belt speed. The highest detection accuracy of 96.5% is obtained at a belt speed of 26 mm/s. When the speed increases to 62 mm/s, 81 mm/s and 99 mm/s, the detection accuracy decreases by 2.2, 7.1 and 18.3 percentage points correspondingly. This is mainly

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

because a higher running speed aggravates motion blur of potato targets in single-frame images captured by the camera and increases inter-frame target displacement, which raises the difficulty for the model to extract and recognize defect features.

The tracking accuracy remains at a high level on the whole. BoT-SORT achieves a tracking accuracy of 100% at the speed of 26 mm/s and 62 mm/s, realizing stable tracking of all individual potatoes even at the medium speed of 62 mm/s. When the speed rises to 81 mm/s and 99 mm/s, the tracking accuracy is 95.2% and 88.1% respectively, without massive target loss or tracking confusion. It indicates that BoT-SORT integrates appearance features and motion information, and has outstanding anti-occlusion and anti-motion blur performance, which can effectively adapt to potato tracking tasks under variable conveyor belt speeds.

In summary, when the conveyor belt speed is ≤ 62 mm/s, the improved detection and tracking system maintains a detection accuracy above 94.3% and a full tracking accuracy of 100%, fully meeting the demands of real-time defect detection and quantity counting in practical production.

4. CONCLUSION

This study proposes a lightweight real-time machine vision system for potato defect detection to overcome the drawbacks of traditional manual detection and the high computational cost of conventional deep learning models. The YOLOv5s backbone is modified by replacing CSPDarknet with GhostNet, which reduces the model size by 26% while retaining strong feature extraction capability, achieving 94.2% precision, 92.2% recall and 95.7% mAP@0.5 on the self-built potato defect dataset. The BoT-SORT multi-object tracking algorithm is integrated to realize stable potato tracking and counting, outperforming SORT, DeepSORT and ByteTrack in core tracking metrics. Experimental results under different conveyor speeds show that the system maintains detection accuracy above 94.3% and 100% tracking accuracy when the conveyor speed is ≤62 mm/s, fully meeting the requirements of real-time industrial detection. This system provides a feasible and efficient solution for online potato defect detection and intelligent sorting, and can be extended to quality inspection of other bulk agricultural products.

REFERENCES

[1] Zhang H, Fen X U, Yu W U, et al. Progress of potato staple food research and industry development in China[J]. Journal of integrative agriculture, 2017, 16(12): 2924-2932.

[2] Fleming S A, Morris J R. Perspective: potatoes, quality carbohydrates, and dietary patterns[J]. Advances in Nutrition, 2024, 15(1): 100138.

[3] Liao H, Wang G, Jin S, et al. HCRP-YOLO: A lightweight algorithm for potato defect detection[J]. Smart Agricultural Technology, 2025, 10: 100849.

[4] Garhwal A S, Pullanagari R R, Li M, et al. Hyperspectral imaging for identification of Zebra Chip disease in potatoes[J]. Biosystems Engineering, 2020, 197: 306-317.

[5] López-Maestresalas A, Keresztes J C, Goodarzi M, et al. Non-destructive detection of blackspot in potatoes by Vis-NIR and SWIR hyperspectral imaging[J]. Food Control, 2016, 70: 229-241.

[6] Al Riza D F, Widodo S, Yamamoto K, et al. External defects and severity level evaluation of potato using single and multispectral imaging in near infrared region[J]. Information Processing in Agriculture, 2024, 11(1): 80-90.

[7] El-Mesery H S, Mao H, Abomohra A E. Applications of Non-destructive Technologies for Agricultural and Food Products Quality Inspection. Sensors, 2019.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202605_12(5).0013

[8] Zeng T, Li S, Song Q, et al. Lightweight tomato real-time detection method based on improved YOLO and mobile deployment[J]. Computers and electronics in agriculture, 2023, 205: 107625.

[9] Peng K, Ma W, Lu J, et al. Application of machine vision technology in citrus production[J]. Applied Sciences, 2023, 13(16): 9334.

[10] Elmasry G, Cubero S, Moltó E, et al. In-line sorting of irregular potatoes by using automated computer-based machine vision system[J]. Journal of Food Engineering, 2012, 112(1-2): 60-68.

[11] Yang L, Mu D, Xu Z, et al. Apple Surface Defect Detection Based on Gray Level Co-Occurrence Matrix and Retinex Image Enhancement[J]. Applied Sciences, 2023, 13(22): 12481.

[12] Barnes M, Duckett T, Cielniak G, et al. Visual detection of blemishes in potatoes using minimalist boosted classifiers[J]. Journal of Food Engineering, 2010, 98(3): 339-346.

[13] Kok Z H, Shariff A R M, Alfatni M S M, et al. Support vector machine in precision agriculture: a review[J]. Computers and Electronics in Agriculture, 2021, 191: 106546.

[14] Tan A, Zhou G, He M. Surface defect identification of Citrus based on KF-2D-Renyi and ABC-SVM[J]. Multimedia Tools and Applications, 2021, 80(6): 9109-9136.

[15] Kamilaris A, Prenafeta-Boldú F X. Deep learning in agriculture: A survey[J]. Computers and electronics in agriculture, 2018, 147: 70-90.

[16] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.

[17] Li M, Zhang Z, Lei L, et al. Agricultural greenhouses detection in high-resolution satellite images based on convolutional neural networks: Comparison of faster R-CNN, YOLO v3 and SSD[J]. Sensors, 2020, 20(17): 4938.

[18] Jiang P, Ergu D, Liu F, et al. A Review of Yolo algorithm developments[J]. Procedia computer science, 2022, 199: 1066-1073.

[19] Yu Y, Zhang K, Yang L, et al. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN[J]. Computers and electronics in agriculture, 2019, 163: 104846.

[20] Dai G, Hu L, Fan J, et al. A deep learning-based object detection scheme by improving YOLOv5 for sprouted potatoes datasets[J]. IEEE Access, 2022, 10: 85416-85428.

[21] Badgujar C M, Poulose A, Gan H. Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review[J]. Computers and Electronics in Agriculture, 2024, 223: 109090.

[22] Xu X, Zhou B, Li W, et al. A method for detecting persimmon leaf diseases using the lightweight YOLOv5 model[J]. Expert Systems with Applications, 2025, 284: 127567.

[23] Wang D, He D. Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning[J]. Biosystems Engineering, 2021, 210: 271-281.

[24] Huangfu Z, Li S, Yan L. Ghost-YOLO v8: An Attention-Guided Enhanced Small Target Detection Algorithm for Floating Litter on Water Surfaces[J]. Computers, Materials and Continua, 2024, 80(3): 3713-3731.

[25] Manzoor S H, Zhang Z, Li H, et al. Optimized Yolov5s-Im for real-time apple flower detection in drone-based pollination[J]. Smart Agricultural Technology, 2025, 12: 101150.

[26] Aharon N, Orfaig R, Bobrovsky B-Z. BoT-SORT: Robust associations multi-pedestrian tracking[J]. arXiv preprint arXiv:2206.14651, 2022.

[27] Petersson M, Kifle Solomon N. Object Tracking Evaluation: BoT-SORT & ByteTrack with YOLOv8: A Comparison of Accuracy and Computational Efficiency, 2024.