IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 April 2024, accepted 3 May 2024, date of publication 6 May 2024, date of current version 28 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3397835

RESEARCH ARTICLE

FES-YOLOv5s: A Lightweight Model for Agaricus Bisporus Detection

HAO MA[1,2], HAIGANG MA[1], JIANGTAO JI[1,2], AND HONGWEI CUI[1,2]
[1]College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471023, China
[2]Longmen Laboratory, Luoyang 471000, China

Corresponding author: Hongwei Cui (Hongwei21@haust.edu.cn)

This work was supported in part by the Major Science and Technology Project of Henan Province under Grant 221100110800; and in part by the Key Laboratory of Modern Agricultural Equipment, Ministry of Agriculture and Rural Affairs, China, under Grant 2023008.

ABSTRACT Agaricus bisporus grows in complex environments and suffers from adhesion and occlusion problems. In this study, we propose a lightweight recognition model for Agaricus bisporus—FES-YOLOv5s—based on YOLOv5s. Our aim was to quickly and accurately detect Agaricus bisporus specimens. First, a FasterNet lightweight network was used in the backbone layer to reduce the computation of the model. An ECA mechanism was then introduced to enhance the interaction between multiple channels and improve the detection accuracy. Finally, a Soft-NMS module was used to replace the NMS module in YOLOv5s to resolve the missed detection of adherent and occluded Agaricus bisporus specimens. The improved model was named FES-YOLOv5s; F, E, and S represent the FasterNet, ECA, and Soft-NMS features, respectively. The FES-YOLOv5s model increased the mAP 0.5:0.95 by 2.4% and the FPS by 19.4%. It decreased the computation by 42.7% compared with the YOLOv5s model. The results of a comparison test revealed that the FES-YOLOv5s model demonstrated advantages in detection accuracy and speed compared with other target detection models. The FES-YOLOv5s model was deployed on an Agaricus-bisporus-picking robot; the detection success rate was greater than 90%, indicating that the improved model could detect Agaricus bisporus quickly and accurately in complex environments.

INDEX TERMS Agaricus bisporus, lightweight model, target detection, YOLOv5.

I. INTRODUCTION

Agaricus bisporus is globally acknowledged to be an edible fungus. It has numerous advantages such as a pleasant flavor, rich nutritional profile, and substantial economic value [1]. The burgeoning cultivation of Agaricus bisporus has accentuated the challenges linked to labor-intensive and inefficient traditional manual picking. This underscores the urgent requirement for better Agaricus bisporus harvesting techniques. Although target detection methods for common fruit and vegetables such as tomatoes, apples, and cucumbers have garnered extensive research attention [2], [3], [4], there has been relatively little focus on the development of target detection approaches specifically for Agaricus bisporus. The intricacies of the mushroom cultivation environment pose significant challenges, including variations in light intensity, discerning substrates and mushrooms, dense growth patterns, and shading issues [5]. These challenges significantly increase the complexity of the detection methods required. The development of efficient, accurate, and swift Agaricus bisporus detection methods is of importance because such methodologies could provide pivotal technical support for the deployment of intelligent robotic picking systems. This would lead to enhanced picking efficiency, reduced picking costs, and the facilitation of large-scale, standardized Agaricus bisporus cultivation.

Target detection algorithms can be categorized into two groups. The first group relies on candidate region-based detection algorithms, including R-CNN [6], Fast R-CNN [7], Faster R-CNN [8], and SPPNet [9]. The other group consists of regression-based detection algorithms such as the YOLO series [10], [11], [12], [13] and SSD [14]. Chen et al. [15]

The associate editor coordinating the review of this manuscript and approving it for publication was Zhongyi Guo.

proposed a YOLOv5s-CBAM mushroom-recognition algorithm that improved the detection accuracy and robustness of the original algorithm using mosaic image enhancement technology. An RGB-D depth camera and an SSD convolutional neural network were used in [16] to locate precise positions based on binocular and structured light-depth images to accurately recognize shiitake mushrooms. Li et al. [17] proposed a YOLO-ACN detection model that improved the detection accuracy of YOLOv3 for small and occluded objects by adding CIOU, Soft-NMS, and depthwise-separable convolutions. McCool et al. [18] proposed a sweet-pepper field detection system that first performed pixel-by-pixel segmentation, then performed region detection, and finally used a local binary mode (LBP) for crop segmentation to accurately recognize highly occluded sweet peppers in the field. Dyrmann et al. [19] proposed a method for the automatic detection of weeds using color images with a large amount of leaf occlusion to automatically detect individual instances of weeds in grain fields, even in the presence of severe leaf occlusion. Zheng et al. [20] presented further advances in target detection with the introduction of R-CSPDarknet53, a novel backbone network designed to improve the detection accuracy of small targets (fruit) over extended distances. Li et al. [21] proposed a multimodal attention fusion network to enhance the detection accuracy of small targets, specifically distant fruit. Their model adjustments included altering the number of detection layers, introducing a weighted bidirectional feature pyramid network (BiFPN) module, and implementing depth-separable convolution and ghost modules to streamline the subsequent mobile deployment. Wang et al. [22] introduced a detailed semantic enhancement (DSE) module for the detection of small fruit. This module employed point-by-point convolution and extended convolution to extract detailed semantic features in both horizontal and vertical dimensions. They also developed exponentially enhanced bifurcation entropy (EBCE) and doubly enhanced mean square error (DEMSE) loss functions to bolster the recognition accuracy of small target objects.

Many scholars have contributed to advancing the concept of model lightweighting by focusing on minimizing parameters, computations, and the overall model size for optimal deployment on mobile platforms. Cong et al. [23] devised a specialized lightweight model for mushroom detection by leveraging YOLOv3 as the foundational framework. Their approach involved crafting a lightweight GhostNet16 network as the backbone network and integrating an adaptive spatial feature pyramid network (ASA-FPN) into the neck network to increase the accuracy of the entire network. Lin et al. [24] introduced a novel underwater treasure detection method based on an enhanced version of YOLOv5. Their approach involved augmenting the recognition accuracy whilst concurrently reducing the network parameters by incorporating attention mechanisms and host modules into the architecture. Wang et al. [25] reduced network parameters and enhanced detection speeds by employing pruning operations that effectively trimmed their model's complexity whilst maintaining performance. Shang et al. [26] presented a method for apple-blossom detection based on YOLOv5s deep learning. Their comprehensive comparison revealed that the YOLOv5 model exhibited superior accuracy and faster speeds in detection tasks. Gong et al. [27] refined the lightweight capabilities of the YOLOv5s model by integrating C3HB and cross-attention modules. This substantially streamlined the model, resulting in optimal lightweight deployment whilst maintaining performance standards. Ugural and Burgan [28] used EVA technology to evaluate the performance of a project model.

These studies have presented various algorithms designed to be applied to agricultural harvesting and have achieved significant progress in fruit and vegetable recognition. Limited attention has been paid to target-recognition methods specifically tailored to Agaricus bisporus. During the growth process of Agaricus bisporus, there are problems of adhesion, occlusion, and density. The existing models have low accuracy and high computational complexity in identifying densely occluded targets, making it difficult to accurately identify Agaricus bisporus, resulting in false positives and missed detections. In this study, we investigated target detection models specifically tailored to Agaricus bisporus by addressing the inadequacies observed in existing target detection models (such as low accuracy, excessive parameters, and high computational demands). The primary objectives of this study were as follows:

1. To improve the backbone network of YOLOv5s to reduce the computational cost;

2. To improve the detection accuracy of occluded and adhered Agaricus bisporus specimens.

II. MATERIALS AND METHODS

A. DATASET

Agaricus bisporus images were obtained at Luoyang Songtian Agricultural Development Co. Ltd. in March–April 2022 using an industrial camera (MV-CC-050, Hikvision,

FIGURE 1. Agaricus bisporus bed.

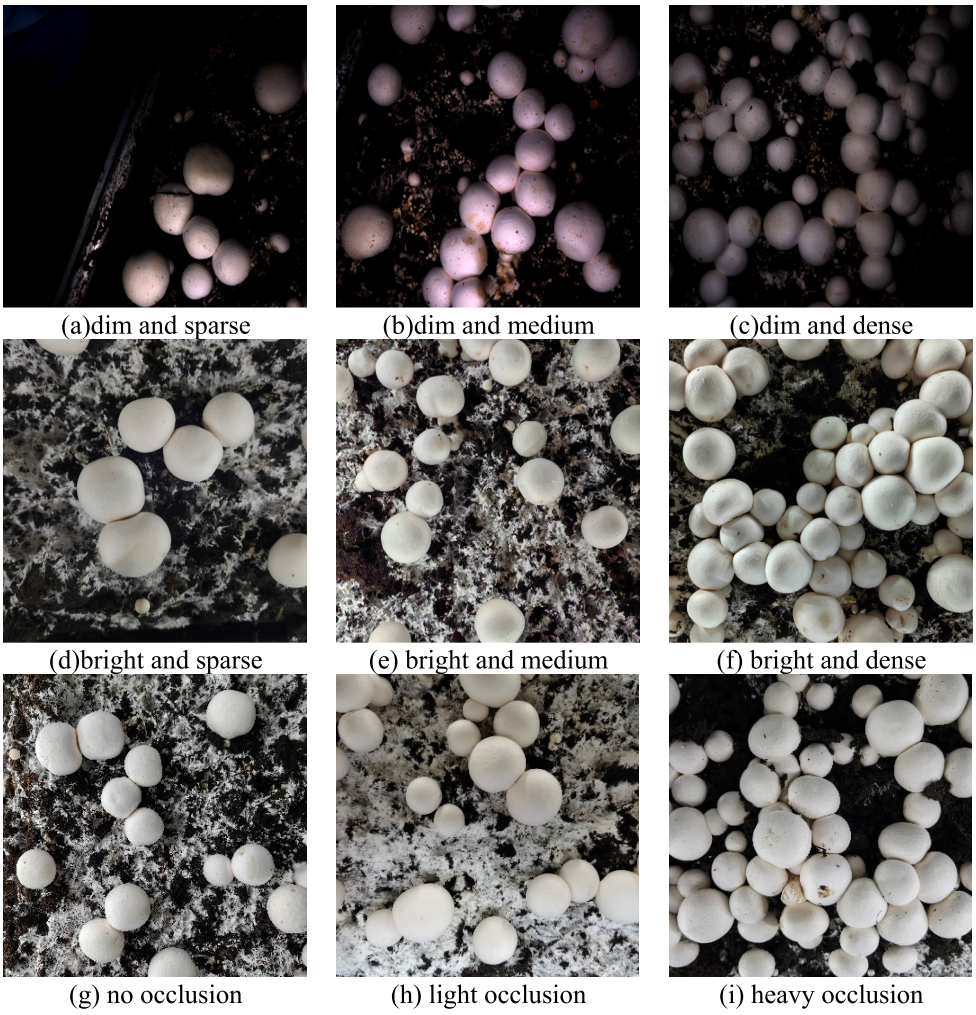

FIGURE 2. Agaricus bisporus growth scenarios.

TABLE 1. Details of the acquired images.

Condition	Acquisition conditions	Number of images
Dim	Sparse	112
	Medium	135
	Dense	120
Bright	Sparse	112
	Medium	135
	Dense	120
Occlusion	None	40
	light	60
	heavy	80
Total		**914**

China) with a resolution of 2448 × 2048 pixels. Vertical overhead shooting was used as the image sample collection method. The camera was installed at a height of 50 ± 5 cm from the mushroom bed, as demonstrated in Figure 1.

Each area of the mushroom bed at the time of the image acquisition was categorized as dim or bright, according to the light intensity. Bright indicated that the proportion of the number of pixel points with a gray value lower than 40 in the grayscale image map exceeded 70%. Dim indicated that the proportion of the number of pixel points with a gray value higher than 40 in the grayscale image map exceeded 70%.

The number of Agaricus bisporus specimens in the camera's field of view (in the image) was counted. The images were then divided into sparse, medium, and dense categories, based on the number of specimens (sparse,

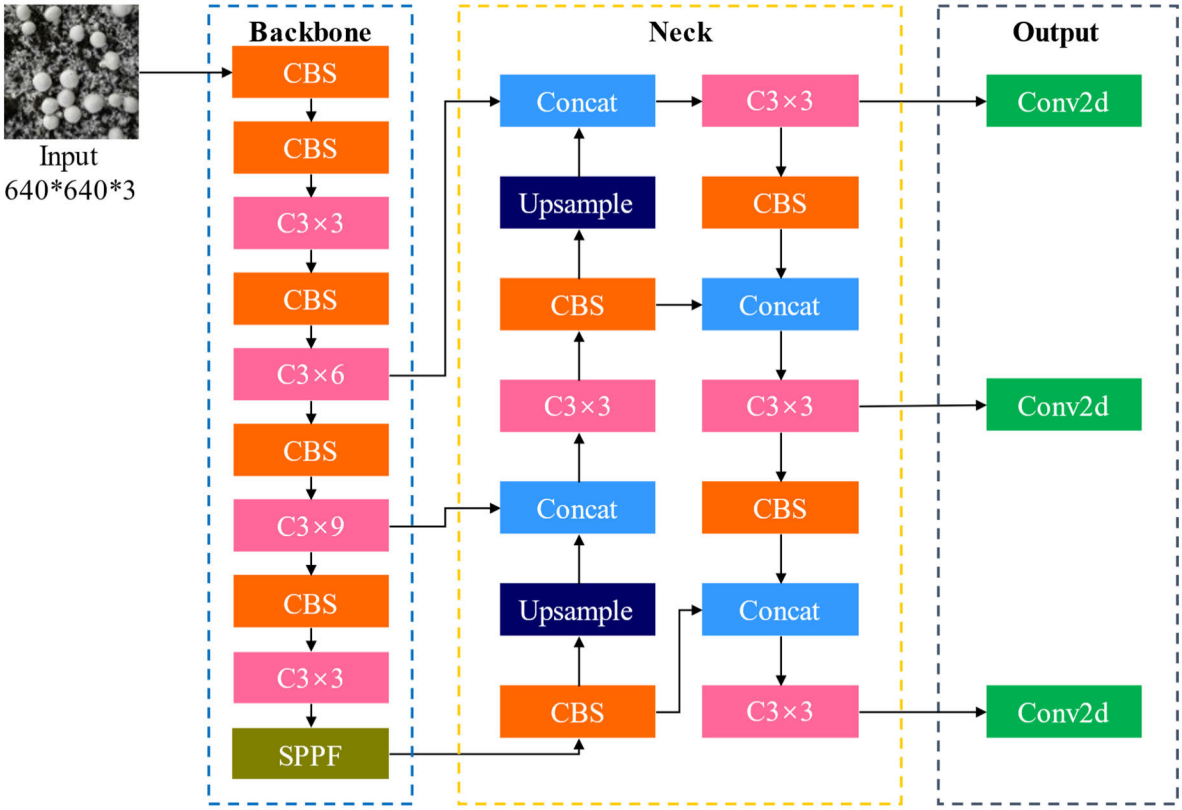

FIGURE 3. YOLOv5s network structure.

0–10 specimens; medium, 11–25 specimens; and dense, 26 or more specimens). Based on the number of occluded mushrooms, the dataset was divided into not occluded (0), slightly occluded (1)–(5), and heavily occluded (5 or more). In total, 914 images were collected (Figure 2). The specific quantities used are listed in Table 1.

It was essential to expand the dataset to enhance the dataset generalization and to prevent overfitting, so a number of images were randomly selected from the 914 original images. Comic Enhancer Pro (v2.49, Jian m, China) software was then used for brightness enhancement and weakening, contrast enhancement and weakening, and sharpening of the selected images. This resulted in 80 images with 25% and 35% brightness enhancement, 80 images with 25% and 35% brightness reduction, 80 images with 25% and 35% contrast enhancement, 80 images with 25% and 35% contrast reduction, 80 images with 25% and 35% contrast reduction, and 80 images with level 1 and level 2 sharpening. Thus, the dataset increased to 1874 images after the expansion process.

As the original images had a resolution of 2448 × 2048 pixels, the image resolution was scaled to 640 × 640 pixels as the input of the model for training to reduce the training time and hardware consumption. Each Agaricus bisporus specimen in the image samples was manually labeled with a small rectangular box using LabelMe software (v5.3.1, MIT CSAIL, USA) and the labeled images were partitioned into distinct sets at a ratio of 7:2:1. The training set comprised 1311 images, the validation set comprised 374 images, and the test set comprised 189 images.

B. FES-YOLOv5s NETWORK

The YOLOv5s model presents distinct advantages such as a low complexity and high operational speed that render it particularly suitable for mobile deployment scenarios. It delineates four fundamental segments within its conventional network architecture: the input, the backbone, the neck, and the head. The input segment plays a critical role in image preprocessing by undertaking pivotal tasks such as image scaling, normalization, and other pertinent operations. In the context of feature extraction, the backbone network leverages CSPDarknet53 and Focus structures to adeptly diminish the model parameters whilst significantly augmenting the feature-extraction capabilities. Complementing this, the neck structure amalgamates FPN [29] and PAN [30] networks; this results in a holistic enrichment of the extracted features. The head module orchestrates the inference procedures and provides outputs detailing the predicted targets. This delineates the classical architecture of the YOLOv5s network, as presented in Figure 3.

In this study, we developed an enhanced YOLOv5s model—FES-YOLOv5s—to improve the detection of Agaricus bisporus specimens. We optimized the backbone network

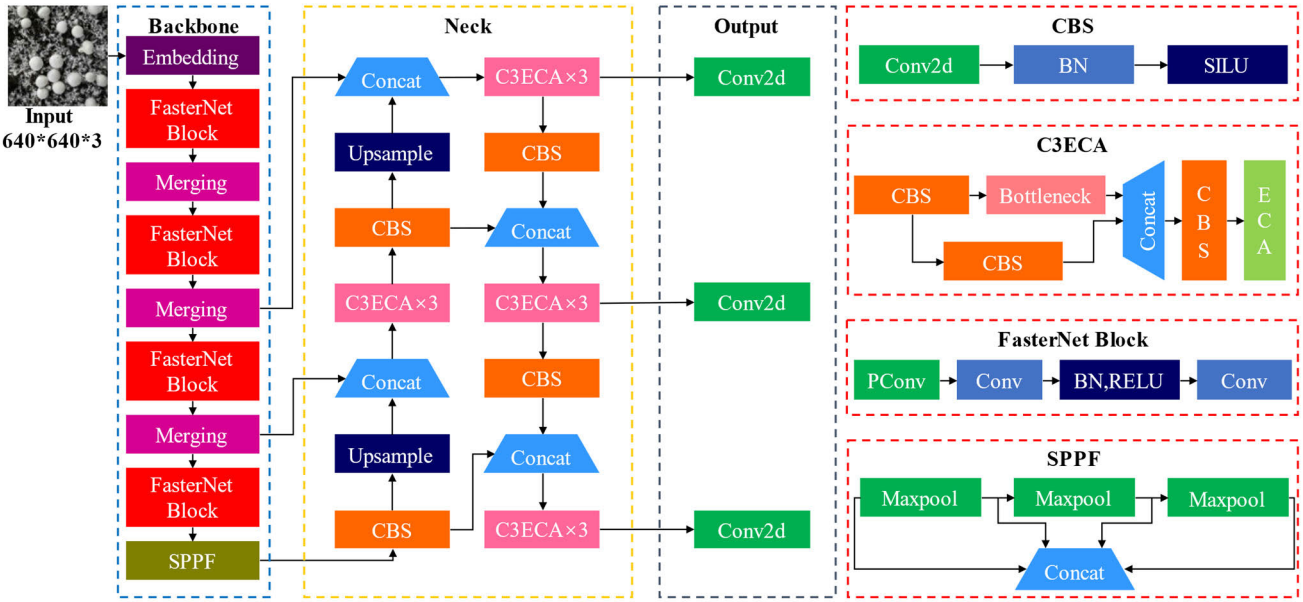

FIGURE 4. FES-YOLOv5s network structure.

of the original YOLOv5s model by adopting the FasterNet-T0 network structure to streamline the model complexity and enhance the floating-point operation speed. Simultaneously, we retained the SPPF module from the original YOLOv5s model to augment the feature-extraction capacity of the model. An efficient channel attention (ECA) module was introduced to compensate for the accuracy decline caused by the lightweighting. This was inserted into the C3 module in the backbone network and neck network. Finally, we enhanced the NMS module by employing Soft-NMS to suppress the candidate frames within YOLOv5s. This modification increased the recognition accuracy of the occluded and adherent instances of Agaricus bisporus. Figure 4 illustrates the revised network structure of the enhanced FES-YOLOv5s model.

C. FASTERNET LIGHTWEIGHT NETWORK

Chen et al. [31] introduced the FasterNet network in 2023 as a lightweight convolutional neural network model. It strategically reduces floating-point operations by integrating a partial convolution (PConv) module and a pointwise convolution (PWConv) module. This amalgamation optimizes the network's efficiency at identifying floating-point operations per second, thus ensuring its lightweight nature.

The FasterNet architecture comprises six variations: T0, T1, T2, S, M, and L. Each variation shares a fundamental design but they differ in their network depth. We chose the FasterNet-T0 configuration for our study because it is known for its minimal parameter count (outlined in Table 2). This configuration was built using multiple layers of FasterNet-Block, as presented in Figure 5. Each FasterNet-Block layer commenced with an embedding or merging layer that was strategically incorporated for spatial downsampling and channel-number expansion.

TABLE 2. Structure of FasterNet-T0.

Stage	Structure	Output Size	Layers
Embedding	Conv4×4,BN	$h/4 \times w/4$	1
FasterNet-Block	PConv,PWConv	$h/4 \times w/4$	1
Merging	Conv2×2,BN	$h/8 \times w/8$	1
FasterNet-Block	PConv,PWConv	$h/8 \times w/8$	2
Merging	Conv2×2,BN	$h/16 \times w/16$	1
FasterNet-Block	PConv,PWConv	$h/16 \times w/16$	8
Merging	Conv2×2,BN	$h/32 \times w/32$	1
FasterNet-Block	PConv,PWConv	$h/32 \times w/32$	2
Classifier	Pooling Conv1×1, FC	1×1	1

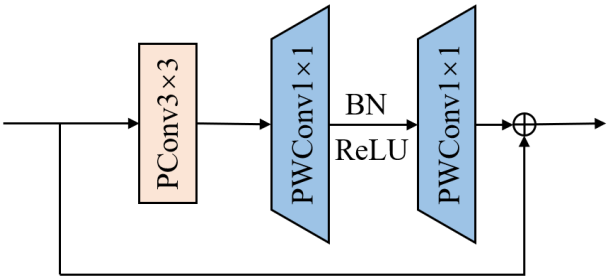

FIGURE 5. FasterNet-Block structure.

D. ECA ATTENTION MODULE

The ECA mechanism [32] enhances the SE attention mechanism [33] by evading dimensionality reduction, effectively capturing interchannel connections and maintaining a concise parameter count. Our ECA structure operated on input feature maps with dimensions h × w × C, as illustrated in Figure 6. These maps were processed using a global average pooling (GAP) layer to acquire 1 × 1 × C feature maps. Subsequently, rapid one-dimensional convolution occurred with a kernel size of k. This was followed by the application of the sigmoid activation function to derive the feature-map weights. Finally, the input feature map was multiplied by these weights to yield a weighted feature map.

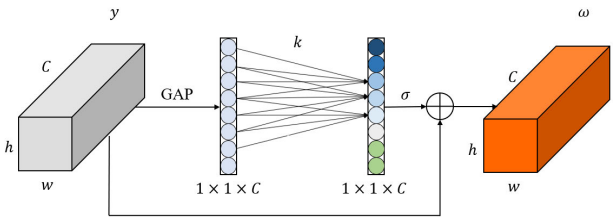

FIGURE 6. ECA structure diagram.

The ECA module was integrated into the model's backbone and neck networks in our research, specifically after the C3 module. This integration formed the C3ECA module. Our aim was to bolster the model's feature-extraction capabilities whilst preserving a lightweight network design (Figure 4).

E. SOFT-NMS MODULE

The branch reset function of traditional NMS is as follows:

$$s_i = \begin{cases} s_i, & iou\,(M, b_i) < N_t \\ 0, & iou\,(M, b_i) \geq N_t \end{cases} \tag{1}$$

where s_i is the candidate frame score value, M is the highest score candidate box, b_i is the candidate box, N_t is the IOU threshold.

When two detection frames are too close to each other in traditional non-maximum suppression (NMS), the detection frame with the lower score is deleted because of the high overlapping area in the detection frame with the higher score. The growth of Agaricus bisporus is characterized by adhesion and mutual occlusion. If the traditional NMS method is used for the suppression of detection frames, it may lead to detection frames being mistakenly deleted because of the excessive proximity of adhesive and mutually occluding Agaricus bisporus detection frames. This could affect the detection results.

The branch reset function of Soft-NMS is as follows:

$$s_i = \begin{cases} s_i, & iou\,(M, b_i) < N_t \\ s_i\,(1 - iou\,(M, b_i))\,, & iou\,(M, b_i) \geq N_t \end{cases} \tag{2}$$

Compared with traditional NMS, Soft-NMS [34] does not directly delete the lower-scoring detection frames during candidate frame suppression, but reduces their scores and then deletes these detection frames when the score is lower than the suppression threshold. Thus, candidate frames with lower scores may participate in the next round of suppression. This reduces the number of incorrectly deleted candidate frames. Hence, Soft-NMS has a better detection accuracy in dense target detection tasks.

In this study, Soft-NMS was used to optimize and improve the original NMS aspect of YOLOv5s to reduce the phenomenon of incorrectly suppressing candidate frames in the suppression phase. Our aim was to improve the detection of Agaricus bisporus in dense, sticky, and occluded scenes.

III. RESULTS

A. EXPERIMENTAL ENVIRONMENT

The models in this study were all trained using a Windows 11 operating system comprising an Intel i5-12400f CPU, an NVIDIA GeForce RTX3060 GPU graphics card with 12 GB of video memory, 32 GB of host memory, CUDA version 11.8, Cudnn version 8.2.1, Python version 3.11, and a Pytorch deep learning framework (torch version 2.0.1+cu118, torchaudio version 2.0.2+cu118, and torchvision version 0.15.2+cu118).

An SGD optimizer with a training cycle epoch of 300, batch size of 16, and initial learning rate of 0.01 was used for the optimization. The cosine annealing function was used to dynamically reduce the learning rate and weight decay to 0.0005

B. EVALUATION METRICS

Precision (P), recall (R), weighted average (F1), mean average precision (mAP), GFLOPs, and frames per second (FPS) were chosen to evaluate the model using the following formula:

$$P = \frac{T_P}{T_P + F_P} \times 100\% \tag{3}$$

$$R = \frac{T_P}{T_P + F_N} \times 100\% \tag{4}$$

$$AP = \int_0^1 P\,(R)\,dR \tag{5}$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{6}$$

$$F1 = \frac{2PR}{P + R} \tag{7}$$

T_P is the correct detection count of Agaricus bisporus, whereas F_P is the incorrect detection count. F_N represents undetected instances of Agaricus bisporus. The variable P reflects the ratio of accurately predicted targets out of the total predicted targets by the model, whereas R is the proportion of correctly predicted real targets out of all the real targets predicted by the model. mAP is the average precision value across classifications; as there was only one class (Agaricus bisporus) in this study, $N = 1$. $F1$ combines the weighted average of precision P and recall R. GFLOPs signifies the total number of computations in each network layer, measured in billions of floating-point operations.

TABLE 3. Comparison of different backbones.

Backbone	$mAP(\%)$	$P(\%)$	$R(\%)$	$F1(\%)$	Computation (GFLOPs)	FPS
ShuffleNetV2	83.2	95.6	92.3	93.9	2.0	138.89
GhostNet	87.1	96.5	94.1	95.3	10.2	121.95
ResNet	90	94.1	95.6	94.9	39.7	88.49
MobileNetV3	82.9	96.7	93.5	95.1	2.5	90.91
FasterNet	90.4	97.1	96.5	96.8	9.1	126.6

TABLE 4. Ablation test result.

Model	mAP(%)	$P(\%)$	$R(\%)$	$F1(\%)$	Computation (GFLOPs)	FPS
YOLOv5s	91.2	97.1	96.5	96.8	15.9	96.2
F-YOLOv5s	90.4	96.8	96.3	96.5	9.1	126.6
E-YOLOv5s	91.8	97.6	96.7	97.1	15.9	96.6
S-YOLOv5s	92.2	97.0	97.1	97.1	15.9	88.5
FES-YOLOv5s	93.6	98.3	97.8	98.0	9.1	114.9

Finally, FPS represents the throughput of the model when detecting images per second.

C. COMPARISON OF DIFFERENT BACKBONES

ShuffleNetV2, GhostNet, ResNet, and MobileNetV3 were selected to replace the backbone network of YOLOv5s to compare the performance differences between different backbone networks. The comparative experimental results are presented in Table 3.

FasterNet produced the highest mAP and F1 values, but was slightly lower than ShuffleNetV2 for FPS and ranked second. ShuffleNetV2's mAP was much lower than FasterNet's. FasterNet demonstrated better overall performance compared with the other backbone networks.

D. ABLATION TESTS

Ablation tests were designed and conducted to verify the improved performance of the different modules for the original YOLOv5s model. The results are presented in Table 4.

The integration of the FasterNet model into the backbone network resulted in nuanced trade-offs. A 0.3% decrease in the F1 value and a 0.8% drop in the mAP were revealed. Concurrently, a reduction in the computational load of 6.8 GFLOPs was demonstrated, coupled with a 30.2 increase in FPS. Our analysis attributed these effects to FasterNet's use of partial convolutions, minimization of convolutional operations, and optimization of memory access. This reduction in network complexity significantly enhanced the unit operation speed but diminished the network's feature-extraction ability, unavoidably lowering both the mAP and F1 values.

The subsequent integration of the ECA mechanism enhanced the model's mAP and F1 values. This augmentation stemmed from the capacity of the ECA mechanism to elevate the multichannel information interaction, incrementally improving the model's feature-extraction ability. The overall network computation remained largely unchanged because of the inherently low parameter count of the ECA.

The incorporation of the Soft-NMS module notably improved the recall of the model but led to a decrease in FPS. Soft-NMS minimizes suppression errors in closely situated targets, thereby reducing leakage in the detected targets and augmenting the model recall. However, this technique retains candidate frames; this potentially leads to their repeated participation in subsequent calculations, marginally increasing the NMS time and prolonging the single-image processing time. This may have caused the decreased FPS values in our study.

Figure 7 illustrates the variation curves of the training process for the mAP, precision (P), recall (R), and loss.

We randomly selected a bright and a dim image from the test set to facilitate a more intuitive comparison of the model performance pre- and post-improvement. The two images were detected using both the original and enhanced models. The results are presented in Figures 8 and 9; the missed Agaricus bisporus specimens are marked using green shapes. The improved FES-YOLOv5s model successfully detected all Agaricus bisporus targets. In contrast, the original YOLOv5s model missed five and four occluded Agaricus bisporus instances. This suggests that the enhanced FES-YOLOv5s model exhibited superior detection capabilities,

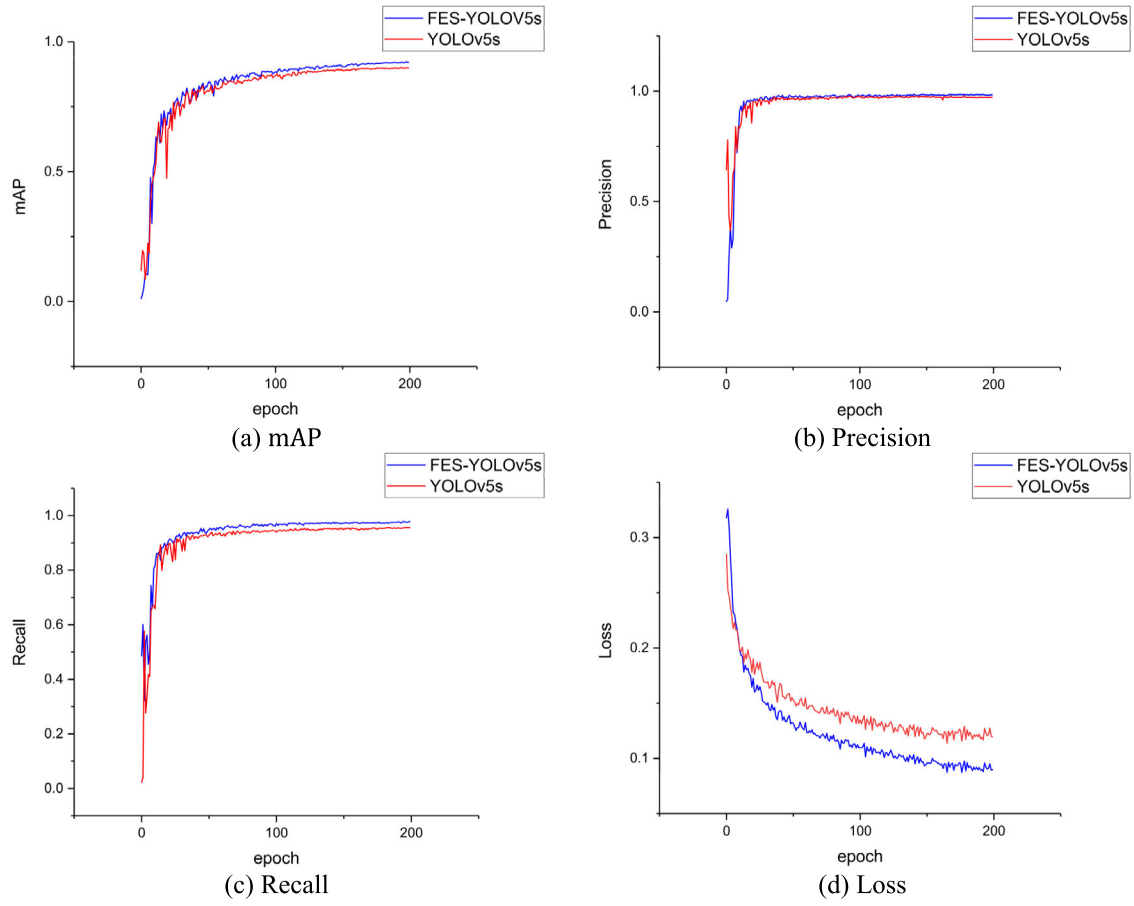

FIGURE 7. Variation curve.

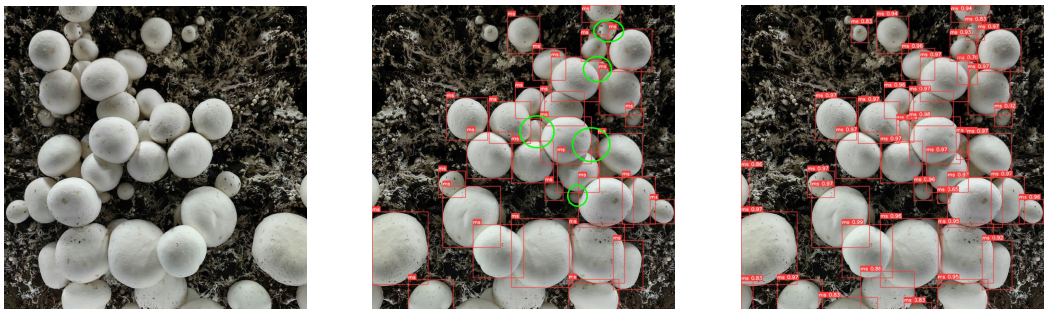

FIGURE 8. Comparison of recognition effects of different models in bright environment.

particularly for instances with densely occluded Agaricus bisporus.

A comprehensive ablation test that optimized all the model parameters was performed following the integration of FasterNet, ECA, and Soft-NMS. This enhanced model reduced the computational load by 6.8 GFLOPs and improved the FPS by 18.7 compared with the original YOLOv5s network. The $F1$ and mAP metrics of the enhanced model were 98.0% and 93.6%, respectively, surpassing the original YOLOv5s model by 1.2% and 2.4%, respectively. This enhancement ensured model lightweighting without compromising the detection accuracy.

IV. DISCUSSIONS

A. COMPARISON OF DIFFERENT MODELS

Experimental comparisons were conducted between the FES-YOLOv5s model and various target detection network models, including SSD, YOLOv4, YOLOv7, and YOLOv8s. Table 5 presents the comparative results.

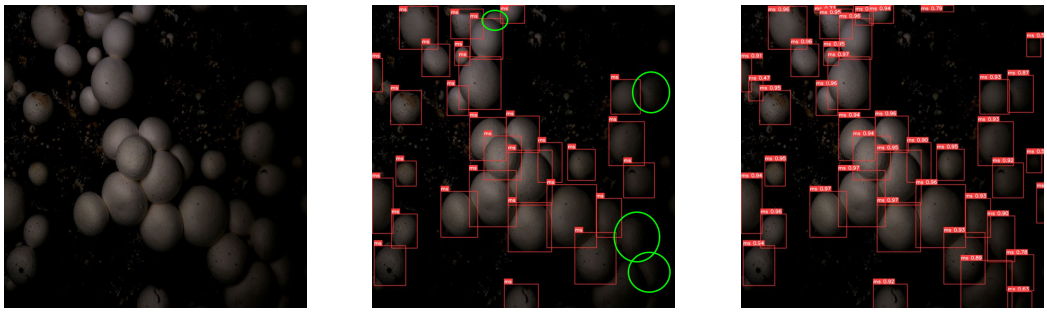

FIGURE 9. Comparison of recognition effects of different models in dim environment.

TABLE 5. Comparison test results of different models.

Model	mAP(%)	P(%)	R(%)	$F1$(%)	Computation(GFLOPs)	FPS
SSD	84.7	91.1	76.5	83.8	284.74	22.4
YOLOv4	75.8	83.5	70.4	76.9	119.89	27.1
YOLOv7	92.7	96.4	97.4	96.7	105.1	27.7
YOLOv8s	93.2	97.9	97.8	97.9	28.7	71.9
FES-YOLOv5s	93.6	98.3	97.8	98.0	9.1	114.9

The FES-YOLOv5s model demonstrated superior performance across multiple metrics compared with the other models. Notably, its mAP was 93.6% (17.8% higher than the lowest-performing YOLOv4 model). This demonstrated the exceptional average detection precision of FES-YOLOv5s and its robust capability of identifying Agaricus bisporus specimens. The F1 score reached 98.0%, reflecting the model's optimal balance between precision and recall. This comprehensive identification capability significantly reduced erroneous and missed detections, ensuring the accurate identification of Agaricus bisporus specimens. The FES-YOLOv5s model operated at a mere 9.1 GFLOPs. This was significantly lower than any other model, indicating a remarkably low computational overhead. For context, the model with the highest computational overhead (SSD) operated at 284.74 GFLOPs, approximately 31 times greater than FES-YOLOv5s. The FES-YOLOv5s model achieved an FPS rate of 114.9, 92.5 FPS higher than the lowest-performing (SSD). This comparison emphasized our model's real-time detection capabilities; it swiftly identified all Agaricus bisporus instances in the images. This aligned well with the real-time demand of intelligent Agaricus-bisporus-picking robots. In summary, the FES-YOLOv5s model outperformed other target detection models across all metrics. Its notably low computational requirements, coupled with the highest real-time detection frame rate, position it as an optimal choice for the swift and accurate detection of Agaricus bisporus, meeting the demands of intelligent picking robots. The trade-offs between accuracy and speed are presented in Figure 10.

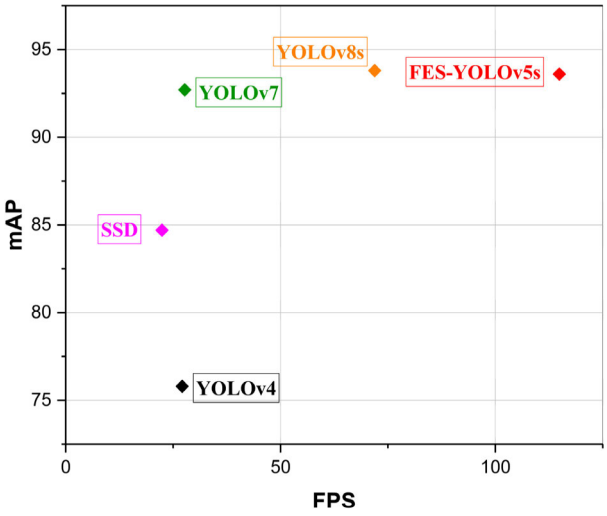

FIGURE 10. Trade-offs between accuracy and speed.

B. ANALYSIS OF CENTER POSITIONING AND DIAMETER MEASURING

We evaluated the center-positioning and diameter-measurement accuracy of different algorithms to characterize the recognition and positioning accuracy of the models. We randomly selected an image of Agaricus bisporus from the test set, as presented in Figure 11. First, we manually marked the boundary rectangle and center point for each Agaricus bisporus specimen (excluding Agaricus bisporus specimens with an incomplete boundary display). The center

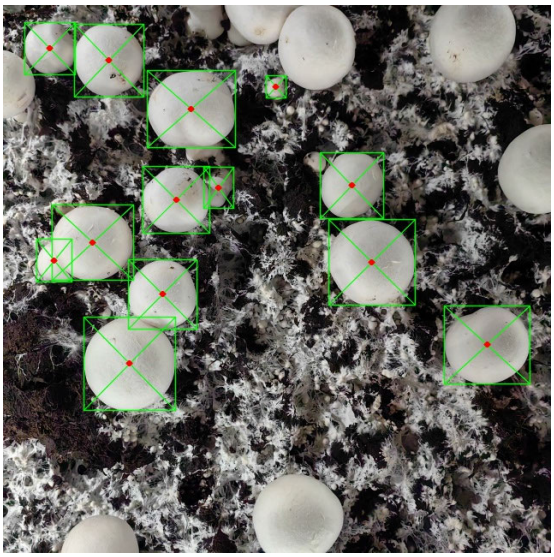

FIGURE 11. Manually annotated images of *Agaricus bisporus*.

point was defined as the diagonal intersection point of the rectangle box. We identified the longer side of the rectangular box as the diameter of each Agaricus bisporus specimen.

We introduced a two-dimensional coordinate deviation rate (CDR) to evaluate the accuracy of the center positioning with greater precision. The formula is as follows:

$$CDR = \left(\left| \frac{X_r - X_m}{s} \right| + \left| \frac{Y_r - Y_m}{t} \right| \right) \times 100\% \quad (8)$$

where X_r and Y_r represent the manually annotated center-point coordinates, X_m and Y_m represent the center-point coordinates obtained by the algorithm, and $s = 640$ and $t = 640$ represent the width and height of the overall image, respectively.

The spatial resolution [35] was used to calculate the actual diameter (in mm) of the Agaricus bisporus specimens. The spatial resolution indicates the number of independent pixels per millimeter; this was 4 pixels/mm in our study. We introduced a formula for the relative error of the diameter measurement, as follows:

$$RE = \frac{|ED - AD|}{AD} \times 100\% \quad (9)$$

where RE is the relative error of measurement, ED is the diameter measured by the algorithm, and AD is the actual diameter.

We introduced IOU [36] to evaluate the recognition accuracy of the algorithm for Agaricus bisporus specimens. First, we calculated the IOU value for each Agaricus bisporus specimen as a graph and then calculated the average IOU.

The center positioning, diameter measurement, and IOU calculation results for SSD, FES-YOLOv5s, YOLOv4, YOLOv7, and YOLOv8s are presented in Table 6.

Compared with the other models, FES-YOLOv5s produced the lowest CDR and RE values as well as the highest IOU value, thus ensuring a better identification and localization of Agaricus bisporus specimens.

TABLE 6. Center positioning and diameter measurement results.

Model	Average CDR(%)	Average RE(%)	Average IOU(%)
SSD	1.45	4.39	95
FES-YOLOv5s	0.52	1.23	98
YOLOv4	4.38	9.36	86
YOLOv7	0.98	1.89	96
YOLOv8s	0.86	1.63	98

C. DETECTION RESULTS OF DIFFERENT MODELS

The detection results for Agaricus bisporus specimens using different target detection models are presented in Figures 12 and 13. Agaricus bisporus, located at the edge of the images, was not counted in the analysis of the results.

The SSD model had a high confidence level, fewer cases of misdetection and omission, and high detection accuracy. However, an analysis of the detection frames revealed that most of the prediction frames were not tangent to the edges of the Agaricus bisporus specimens and there were gaps that did not accurately wrap around the Agaricus bisporus specimens. This phenomenon was particularly obvious when identifying densely occluded Agaricus bisporus specimens. There were cases when the background was recognized as Agaricus bisporus in a few images and there were examples of the phenomenon of missed detection. This affected the localization accuracy of Agaricus bisporus specimens, which would, in turn, affect the accuracy of the picking robot at the next step. The confidence level of the YOLOv4 network model was generally poor and the prediction frame had the poorest effect of wrapping the Agaricus bisporus specimens. It did not completely recognize all Agaricus bisporus specimens in the images and there were serious misdetections. YOLOv7 and YOLOv8s had higher confidence levels, but there was the phenomenon of missed detection in certain occluded Agaricus bisporus samples.

In summary, the recognition effect of the improved FES-YOLOv5s model for Agaricus bisporus specimens under different light and growth conditions was better than other target detection models. It accurately identified the location of Agaricus bisporus specimens with high confidence and the prediction frame could successfully wrap the Agaricus bisporus specimens as well as identify occluded Agaricus bisporus specimens. There were no misdetections or omissions.

D. MODEL DEPLOYMENT

The enhanced FES-YOLOv5s model was deployed on an Agaricus-bisporus-picking robot. The robot comprised a truss-type mechanism with X, Y, and Z three-axis configurations; a 270 degree servo; a flexible picking manipulator; an industrial camera; and associated components. The pneumatic-controlled flexible picking manipulator contracted upon the application of negative pressure, allowing

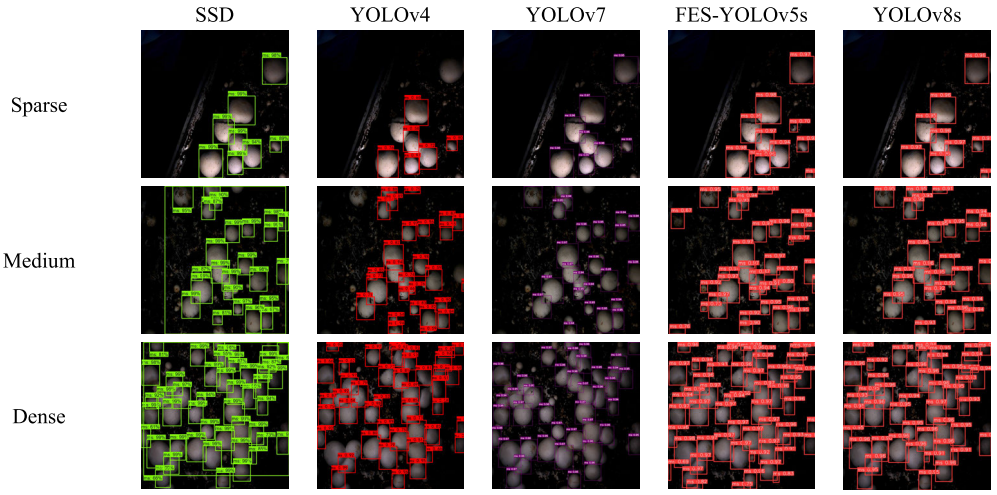

FIGURE 12. Recognition effect of Agaricus bisporus in dim environment.

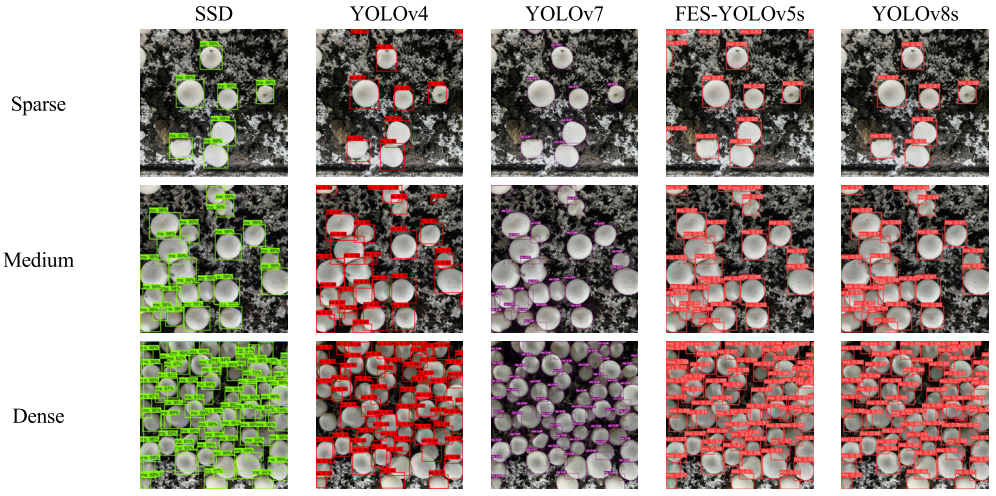

FIGURE 13. Recognition effect of Agaricus bisporus in bright environment.

it to securely grip Agaricus bisporus specimens. Conversely, the manipulator opened and released the selected Agaricus bisporus specimen under positive pressure. The industrial camera was located in a fixed position alongside the flexible picking robot, enabling the transformation of the Agaricus bisporus positions in the captured images into the basic coordinate system of the robot. Figure 14 depicts the setup of the Agaricus-bisporus-picking robot.

Several rounds of trials were conducted after deploying FES-YOLOv5s to the Agaricus-bisporus-picking robot. The picking robot first scanned the picking area line-by-line whilst the industrial camera continuously captured the images. After the area scanning was complete, the host computer processed the captured Agaricus bisporus images, recognized the positions of all Agaricus bisporus mushrooms, converted the pixel positions to real positions using a coordinate system conversion, and controlled the manipulator to perform the picking. The experimental results revealed

FIGURE 14. Agaricus bisporus picking robot.

that the Agaricus-bisporus-picking robot deployed with the FES-YOLOv5s algorithm had a recognition accuracy greater than 90%. This verified the effectiveness of the model. The experimental results are presented in Table 7.

TABLE 7. Identification results of Agaricus bisporus picking robot.

Test number	Actual number	Number of detections	Missed detections	Wrong detections	Recognition rate(%)
1	56	53	3	1	92.9
2	61	62	2	3	96.7
3	30	29	1	0	96.7
4	18	18	0	0	100.00
5	43	42	2	2	93.0
6	35	33	1	1	91.4

V. CONCLUSION

In this study, we enhanced the YOLOv5s target detection model by refining its backbone layer, integrating an ECA mechanism, and incorporating a Soft-NMS module. This optimization significantly reduced the number of parameters and computational load, increasing the adaptability of the model for mobile deployment. Several conclusions were drawn through ablation and comparative tests using other detection models.

1. By replacing the YOLOv5s backbone with the FasterNet lightweight model and integrating the ECA mechanism and Soft-NMS module, the enhanced FES-YOLOv5s model decreased the computation by 6.8 GFLOPs and increased the FPS by 19.4%. It achieved precision and mean average accuracy results of 98.3% and 93.6%, respectively, surpassing the original YOLOv5s model by 1.2% and 2.4%, respectively. This indicates a high recognition accuracy alongside a lightweight design.

2. The FES-YOLOv5s model outperformed the YOLOv5s, SSD, YOLOv4, YOLOv7, and YOLOv8s models. The mAP values increased by 2.4%, 8.9%, 17.8%, 0.9%, and 0.4%, respectively. The FPS increased by 18.7, 92.5, 87.8, 87.2, and 43.0, respectively.

3. The enhanced FES-YOLOv5s model proved to be effective at detecting Agaricus bisporus. Upon deployment on the Agaricus-bisporus-picking robot, it achieved a detection accuracy greater than 90% and accurately identified obscured and overlapped mushrooms. It exhibited rapid real-time detection capabilities, thus meeting the demands of rapid detection.

This model has conducted relevant application research on the detection of Agaricus bisporus. Due to the growth characteristics of Agaricus bisporus (such as density and occlusion), the FES-YOLOv5s model has potential application value in dense and occluded target detection scenarios. We will continue to explore the application of FES-YOLOv5s in a wider range of detection scenarios in the future.

REFERENCES

[1] J. Ji, M. Li, K. Zhao, and H. Ma, "Design and experiment of flexible profiling picking end-effector for agaricus bisporus," *Trans. Chin. Soc. Agric. Mach.*, vol. 54, no. 1, pp. 104–115, 2023.

[2] T. Li, M. Sun, X. Ding, Y. Li, G. Zhang, G. Shi, and W. Li, "Tomato recognition method at the ripening stage based on YOLO v4 and HSV," *Trans. Chin. Soc. Agric. Eng.*, vol. 37, pp. 183–190, Jun. 2021.

[3] J. Yang, Z. Qian, Y. J. Zhang, Y. Qin, and H. Miao, "Real-time recognition of tomatoes in complex environments based on improved YOLOv4-tiny," *Trans. Chin. Soc. Agric. Eng.*, vol. 38, no. 9, pp. 215–221, May 2022.

[4] F. Zhang, Z. Chen, R. Bao, C. Zhang, and Z. Wang, "Recognition of dense cherry tomatoes based on improved YOLOv4-LITE lightweight neural network," *Trans. Chin. Soc. Agric. Eng.*, vol. 37, no. 16, pp. 270–278, 2021.

[5] Y. Li, "Research on optimization and artificial cultivation technology of *Termitomyces albuminosus* culture," M.S. thesis, Dept. Inf. Control Eng., Shaanxi Univ. Technol., Hanzhong, China, 2018.

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[7] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1137–1149.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[10] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[12] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.

[13] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Computer Vision—ECCV 2006*. Amsterdam, The Netherlands: Springer, 2006, pp. 21–37.

[15] C. Chen, F. Wang, Y. Cai, S. Yi, and B. Zhang, "An improved YOLOv5s-based agaricus bisporus detection algorithm," *Agronomy*, vol. 13, no. 7, p. 1871, Jul. 2023.

[16] Y. Qian, R. Jiacheng, W. Pengbo, Y. Zhan, and G. Changxing, "Real-time detection and localization using SSD method for oyster mushroom picking robot," in *Proc. IEEE Int. Conf. Real-time Comput. Robot. (RCAR)*, Sep. 2020, pp. 158–163.

[17] Y. Li, S. Li, H. Du, L. Chen, D. Zhang, and Y. Li, "YOLO-ACN: Focusing on small target and occluded object detection," *IEEE Access*, vol. 8, pp. 227288–227303, 2020.

[18] C. McCool, I. Sa, F. Dayoub, C. Lehnert, T. Perez, and B. Upcroft, "Visual detection of occluded crop: For automated harvesting," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 2506–2512.

[19] M. Dyrmann, R. N. Jørgensen, and H. S. Midtiby, "RoboWeedSupport–detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network," *Adv. Animal Biosci.*, vol. 8, no. 2, pp. 842–847, 2017.

[20] T. Zheng, M. Jiang, Y. Li, and M. Feng, "Research on tomato detection in natural environment based on RC-YOLOv4," *Comput. Electron. Agricult.*, vol. 198, Jul. 2022, Art. no. 107029.
[21] S. Li, S. Zhang, J. Xue, and H. Sun, "Lightweight target detection for the field flat jujube based on improved YOLOv5," *Comput. Electron. Agricult.*, vol. 202, Nov. 2022, Art. no. 107391.
[22] Y. Wang, G. Yan, Q. Meng, T. Yao, J. Han, and B. Zhang, "DSE-YOLO: Detail semantics enhancement YOLO for multi-stage strawberry detection," *Comput. Electron. Agricult.*, vol. 198, Jul. 2022, Art. no. 107057.
[23] P. Cong, H. Feng, K. Lv, J. Zhou, and S. Li, "MYOLO: A lightweight fresh shiitake mushroom detection model based on YOLOv3," *Agriculture*, vol. 13, no. 2, p. 392, Feb. 2023.
[24] S. Lin, M. Liu, and Z. Tao, "Detection of underwater treasures using attention mechanism and improved YOLOv5," *Trans. Chin. Soc. Agric. Eng.*, vol. 37, no. 18, pp. 307–314, 2021.
[25] Z. Wang, X. Xu, Z. Hua, Y. Shang, Y. Duan, and H. Song, "Lightweight recognition for the oestrus behavior of dairy cows combining YOLOv5n and channel pruning," *Nongye Gongcheng Xuebao/Trans. Chin. Soc. Agric. Eng.*, vol. 38, no. 23, pp. 130–140, 2022.
[26] Y. Shang, Q. Zhang, and H. Song, "Application of deep learning using YOLOv5s to apple flower detection in natural scenes," *Trans. Chin. Soc. Agric. Eng.*, vol. 9, pp. 222–229, Jan. 2022.
[27] W. Gong, Z. Yang, K. Li, W. Hao, Z. He, X. Ding, and Y. Cui, "Detecting kiwi flowers in natural environments using an improved YOLOv5s," *Trans. Chin. Soc. Agric. Eng.*, vol. 39, no. 6, pp. 177–185, 2023.
[28] M. N. Ugural and H. I. Burgan, "Project performance evaluation using EVA technique: Kotay bridge construction project on Kayto River in Afghanistan," *Tehnički vjesnik*, vol. 28, no. 1, pp. 340–345, 2021.
[29] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.
[30] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8759–8768.
[31] J. Chen, S.-H. Kao, H. He, W. Zhuo, S. Wen, C.-H. Lee, and S.-H.-G. Chan, "Run, don't walk: Chasing higher FLOPS for faster neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 12021–12031.
[32] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.
[33] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[34] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, "Soft-NMS—Improving object detection with one line of code," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5562–5570.
[35] C.-P. Lu and J.-J. Liaw, "A novel image measurement algorithm for common mushroom caps based on convolutional neural network," *Comput. Electron. Agricult.*, vol. 171, Apr. 2020, Art. no. 105336.
[36] B. Natesan, C.-M. Liu, V.-D. Ta, and R. Liao, "Advanced robotic system with keypoint extraction and YOLOv5 object detection algorithm for precise livestock monitoring," *Fishes*, vol. 8, no. 10, p. 524, Oct. 2023.

HAO MA received the B.S. and Ph.D. degrees in electrification and automation from China Agricultural University, Beijing, China, in 2015. He is currently an Associate Professor with the College of Agricultural Equipment Engineering, Henan University of Science and Technology, and the Deputy Secretary-General of the Artificial Intelligence Branch. He is the author of two books, more than ten articles, and two patents. His research interests include crop non-destructive testing technology, agricultural product grading technology, and crop phenotype detection technology.

HAIGANG MA received the B.S. degree in vehicle engineering from Jiangsu University, Zhenjiang, China, in 2020. He is currently pursuing the M.S. degree in agriculture with Henan University of Science and Technology, Luoyang, China. His research interests include digital agriculture and the application of machine vision in agriculture.

JIANGTAO JI received the Ph.D. degree in engineering from Beijing Institute of Technology, Beijing, China. He is currently the Dean, a Professor, and the Ph.D. Supervisor of the College of Agricultural Equipment Engineering, Henan University of Science and Technology, and the Executive Director of China Agricultural Machinery Society. He is the author of seven books, more than 120 articles, and more than 50 patents. His research interests include modern agricultural equipment and intelligent technology, and agricultural robots and information technology.

HONGWEI CUI received the Ph.D. degree from Jilin University, Jilin, China, in 2021. He is currently a Lecturer with the College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang, China. He led a research project on detection technology and systems for winter wheat spike and spike characteristics, from 2022 to 2024. His research interests include crop phenotype detection technology fields, such as agricultural crop nondestructive testing technology and agricultural product grading technology.