

RECOGNITION OF AGARICUS BISPORUS BASED ON IMPROVED MASK-RCNN MODEL

基于改进 Mask-RCNN 模型的双孢菇识别

Shuo WANG¹⁾, Yaozong SHI¹⁾, Lihang CHEN¹⁾, Weihao HAO¹⁾, Junlong MENG³⁾, Jingyu LIU³⁾
Yanqing ZHANG^{1,2)}, Zhiyong ZHANG^{*1,2)}

¹⁾College of Agricultural Engineering, Shanxi Agricultural University, Taigu / China

²⁾Dryland Farm Machinery Key Technology and Equipment Key Laboratory of Shanxi Province, Jinzhong / China

³⁾College of Food Science and Engineering, Shanxi Agricultural University, Taigu / China

Tel: +86-0354-6288339; E-mail: zyzgh@sxau.edu.cn

Corresponding author: Zhiyong ZHANG

DOI: <https://doi.org/10.35633/inmateh-77-48>

Keywords: *Agaricus bisporus*, Mask-RCNN, recognition

ABSTRACT

The recognition of *Agaricus bisporus* is a key step in the intelligent picking of *Agaricus bisporus*. Given the complex background and limited computing resources of edge devices in actual planting scenarios, an improved Mask-RCNN model for *Agaricus bisporus* recognition was proposed. In this method, the backbone feature extraction network of the baseline Mask-RCNN model was replaced with the lightweight MobileNetV3 network to reduce the model complexity. Meanwhile, the BiFPN network was used to replace the original FPN feature fusion network, thereby strengthening feature fusion and enhancing the model's ability to learn image features and acquire contextual information. Experimental results showed that the improved Mask-RCNN model's parameters and floating-point operations were 24.46 M and 173 G, respectively, which were 44.4% and 24.5% lower than those of the baseline Mask-RCNN model, and the frame rate increased by 3.55 FPS, indicating a better prospect for deployment on edge devices. This method can provide technical support for the development of the visual system of *Agaricus bisporus* picking robots.

摘要

双孢菇识别是双孢菇智能采摘的关键环节。针对实际种植场景中背景复杂以及边缘设备计算资源有限的问题，提出了一种用于双孢菇识别的改进 Mask-RCNN 模型。该方法将基准 Mask-RCNN 模型的骨干特征提取网络替换为轻量化的 MobileNetV3 网络，以降低模型复杂度；同时，采用 BiFPN 网络替代原有的 FPN 特征融合网络，从而强化特征融合，提升模型对图像特征的学习能力和上下文信息的获取能力。实验结果表明，改进后的 Mask-RCNN 模型参数为 24.46M，浮点运算量为 173G，分别较基准 Mask-RCNN 模型降低了 44.4% 和 24.5%，帧率提升了 3.55 FPS，在边缘设备上具有更好的部署前景。该方法可为双孢菇采摘机器人视觉系统的研发提供技术支持。

INTRODUCTION

Agaricus bisporus, commonly known as the white button mushroom, is characterized by a delicate flavor and high nutritional value. It contains abundant amino acids, nucleotides, and vitamins, and also exhibits notable medicinal properties, including antihypertensive, lipid-lowering, and anticancer effects. As a result, its market demand has been increasing year by year (Wang et al., 2018; Wangsa et al., 2024; Menna et al., 2024). *Agaricus bisporus* is widely cultivated in provinces such as Jiangsu, Shandong, Hebei, and Shanxi in China, and is one of the edible fungi with the largest cultivation scale in China (China Edible Fungi Association, 2024). At present, the industrialized production mode of *Agaricus bisporus* is becoming increasingly mature, but the picking process still mainly relies on manual labor (Luo et al., 2021). Manual picking cannot meet the needs of large-scale *Agaricus bisporus* production, with problems such as low harvesting efficiency, high cost, and labor shortage during the harvesting season, which have become important factors restricting the rapid development of the *Agaricus bisporus* industry. Therefore, it is necessary to develop efficient and reliable intelligent harvesting equipment for *Agaricus bisporus*. Accurate identification of *Agaricus bisporus* is crucial for its intelligent picking (Liu et al., 2023).

¹ Shuo Wang, master degree; Yaozong Shi, master degree; Lihang Chen, master degree; Weihao Hao, master degree; Zhiyong Zhang, Prof. Ph.D.; Yanqing Zhang, Prof. Ph.D.; Junlong Meng, Prof. Ph.D. Jingyu Liu, Prof. Ph.D.

With the development of computer technology, deep learning-based object detection methods have received extensive attention in the field of edible fungi detection (Albayrak et al., 2025; Javanmardi et al., 2025; Li et al., 2024; Lu et al., 2022; Luo et al., 2019; Yang et al., 2022). Deep learning methods have advantages such as strong feature extraction capability, high accuracy, and good robustness, and can adaptively extract image features from datasets to construct detection models for object recognition or segmentation (Ishana et al., 2023; Fang et al., 2024). At present, deep learning object detection methods are generally divided into two categories: one-stage models represented by the YOLO series and two-stage models represented by Mask R-CNN (Seetharaman K., 2022; Redmon J. et al., 2016). One-stage models can complete object classification and position prediction through a single feature extraction step, with simple structures and fast inference speed, but their accuracy is generally inferior to two-stage object segmentation models. The two-stage Mask R-CNN algorithm, improved from Faster R-CNN, can output rectangular bounding boxes and masks of objects to realize object recognition and segmentation.

The Mask R-CNN model can simultaneously output the position and contour information of *Agaricus bisporus*, providing information support for the estimation of picking points and growth postures of *Agaricus bisporus*. However, as it employs the ResNet deep network for feature extraction, the model has a large number of parameters, high computational load, and slow inference speed (He et al., 2017). In the application scenario of picking robots, the model's high computational complexity and resource requirements limit its practicality. Based on Mask R-CNN, this paper proposed an improved Mask R-CNN model for *Agaricus bisporus* detection, combined with the requirements of lightweight deployment and real-time detection in practical applications.

MATERIALS AND METHODS

Image Acquisition and Augmentation

Agaricus bisporus images were collected in August 2024 from an *Agaricus bisporus* cultivation base in Fenxi County, Linfen City, Shanxi Province, China, using a D435i depth camera. The depth camera was positioned approximately 25-40 cm above the surface of the *Agaricus bisporus* cultivation substrate. The collected videos were frame-extracted, and after removing duplicate and blurred images, 500 *Agaricus bisporus* images were obtained. To increase sample diversity and improve model generalization ability, data augmentation methods such as contrast adjustment, brightness adjustment, color inversion, and noise addition were used, resulting in a total of 2000 sample images. The dataset was then divided into training and validation sets at an 8:2 ratio, with 1600 randomly selected images as the training set and 400 as the validation set (Wang et al., 2025). Examples of image augmentation are shown in Fig. 1. The *Agaricus bisporus* images were annotated using LabelMe software and saved as JSON-format label files.



Fig. 1 - Examples of image augmentation

The Baseline Mask R-CNN Model

The baseline Mask R-CNN is mainly composed of a backbone (ResNet50+FPN), RPN network, and Output Head network, with its baseline network structure shown in Fig. 2. The backbone network extracts features of *Agaricus bisporus* images through the ResNet50 convolutional neural network, and then fuses feature maps of different scales via the FPN network to generate multi-scale information feature maps. Various rectangular anchor boxes with different scale ratios and shapes are then set in the multi-scale information feature maps, and initial target candidate regions are generated by combining with the RPN network. The extracted target candidate regions are aligned with the feature maps through the ROIAlign algorithm. The aligned candidate region feature maps are input into the Output Head network, which consists of two parallel branches: the object detection branch and the mask prediction branch.

The object detection branch locates the bounding box of *Agaricus bisporus*, while the mask prediction branch segments it using the fully convolutional network.

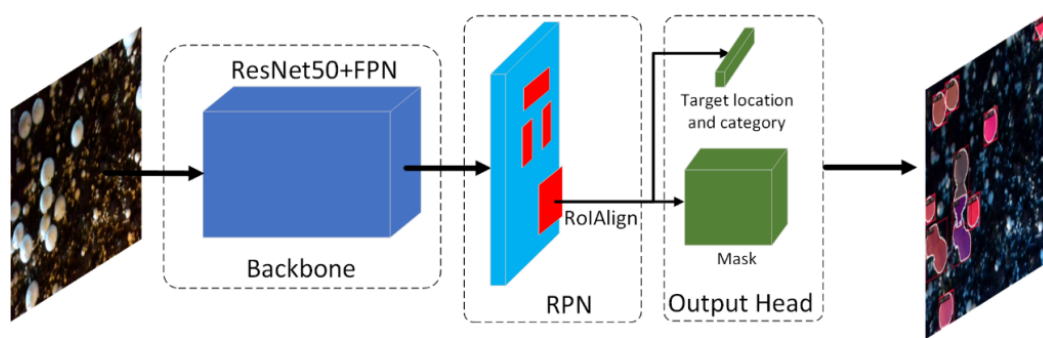


Fig. 2 - Mask R-CNN Model Structure

Improved Mask R-CNN Mode

The baseline Mask R-CNN model has high complexity, and its high computational complexity and resource requirements limit its practicality. Therefore, this paper performed lightweight improvements on the Mask R-CNN network: MobileNetV3 was used to replace ResNet50 in the backbone network for feature extraction to reduce the network parameters; meanwhile, the BiFPN network was introduced to replace the FPN feature fusion module to improve the network's multi-scale feature fusion capability and compensate for the accuracy loss due to reduced parameters in the feature extraction network. The structure of the improved Mask R-CNN model is shown in Fig. 3. The model took 640×640 pixel images as input, extracted features using the MobileNetV3 network, and output feature information of Layer 1, Layer 3, Layer 8, and Layer 11, respectively. Then, the BiFPN feature fusion network fused feature information of different layers and output them to RPN for candidate box extraction. Finally, the detection results were output through Output Head.

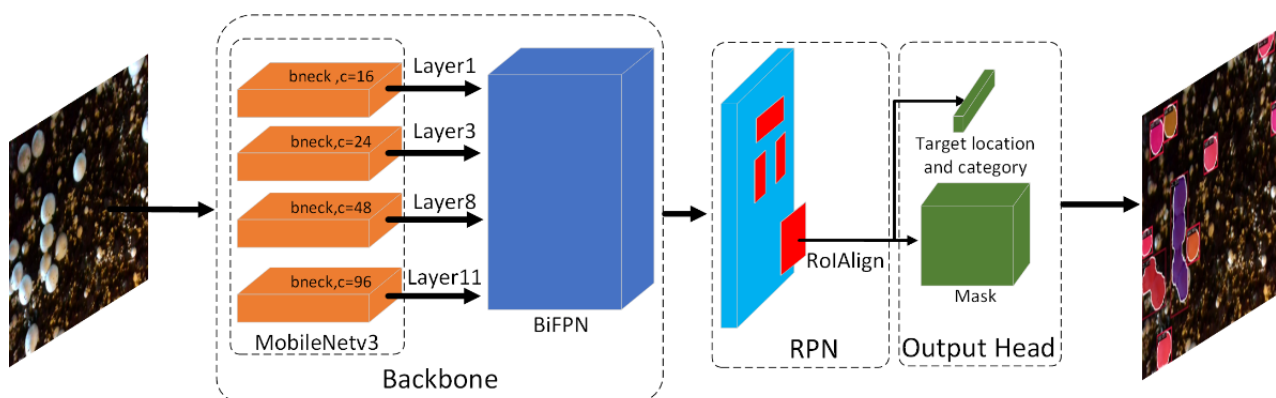


Fig. 3 - Improved Mask R-CNN Model Structure

(1) MobileNetV3 Feature Extraction Network

To reduce algorithm complexity for deployment on AI development boards, the model requires lightweight processing. Therefore, this paper adopted the lightweight feature extraction network MobileNetV3 (Howard et al., 2019) to replace ResNet50. MobileNetV3 uses depth-wise separable convolution instead of a large number of standard convolutions, which can effectively compress the computational load and parameters of the feature extraction network in the detection model for *Agaricus bisporus*.

(2) BiFPN feature fusion network

The baseline Mask R-CNN model uses FPN as the feature fusion network (Lin et al., 2017). Although the introduction of FPN alleviates the problem of missing detection of low-resolution small targets by the model, FPN only fuses high and low-level features via upsampling at the neck, and thus affects segmentation accuracy. Therefore, to improve the feature fusion capability of the model, this paper introduces the BiFPN (Tan et al., 2020) network to replace the FPN network. BiFPN constructs bidirectional information flow channels through top-down and bottom-up paths, realizing in-depth information exchange and fusion between feature maps to improve segmentation accuracy. Its structure is shown in Fig. 4.

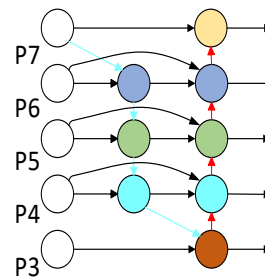


Fig. 4 - BiFPN network Structure

Model training and evaluation

Model training was conducted in a cloud server environment with the following hardware configuration: AMD R3900X 12-core processor, 32GB memory, and Nvidia RTX 2080Ti graphics card (12GB video memory); the software environment adopted the Windows 10 operating system, with the deep learning framework PyTorch 2.1.0. In the experiment, the input image resolution of the model was set to 640×640, and the training parameters were configured as follows: epoch=150, initial learning rate=0.003, initial momentum=0.9, batch size=4.

The mean average precision (*mAP*) is used as the evaluation metric for model accuracy. Specifically, *mAP@0.5* (box) refers to the detection *mAP* when the intersection over union (*IoU*) threshold is set to 0.5, which is used to evaluate the model's detection precision; *mAP@0.5* (mask) refers to the segmentation *mAP* when the intersection over union (*IoU*) threshold is set to 0.5, which is used to evaluate the model's segmentation precision. The number of parameters and floating-point operations are used to evaluate model complexity, and the frame rate is used to evaluate model inference speed.

To verify the feasibility of deploying the improved Mask-RCNN model on mobile edge devices, the model was converted to ONNX format and deployed on the EC-R3588SPC edge device (an AI development board) for testing. The hardware configuration of the EC-R3588SPC edge device is as follows: a CPU with 4-core Cortex-A76 and 4-core Cortex-A55, 8GB of RAM, and 64GB of storage space. The software environment is configured with Python 3.10, MMDeploy 1.3.1, and ONNX Runtime 1.23.1.

RESULTS

Effect of different feature extraction networks on model performance

Lightweight improvements were implemented by replacing the ResNet50 network of the baseline Mask R-CNN model with common lightweight feature extraction networks such as MobileNetV3, ResNet18, and ConvNeXt. The model test results are shown in Table 1. As shown in the table, replacing the backbone with lightweight networks led to a decrease in model accuracy; however, when using MobileNetV3 as the feature extraction network, the model achieved significant reductions in parameters and computational load, while the frame rate was notably improved compared with the original Mask R-CNN model.

Table 1

Comparison of Lightweight Feature Extraction Networks

Feature extraction network	Independent growth		Adherent growth		Parameters / M	Floating-point operations / G	Frame rate / FPS
	<i>mAP@0.5</i> (box)	<i>mAP@0.5</i> (mask)	<i>mAP@0.5</i> (box)	<i>mAP@0.5</i> (mask)			
ResNet50	0.929	0.927	0.823	0.831	43.971	229	6.35
MobileNetV3	0.841	0.761	0.821	0.636	20.46	152	10.2
ResNet18	0.923	0.910	0.893	0.936	31.328	186	5.8
ConvNeXt	0.944	0.945	0.947	0.948	48.095	223	6.75

Effect of the BiFPN network on model performance

To verify the effectiveness of the BiFPN feature fusion network, experiments were conducted by replacing FPN with BiFPN with MobileNetV3 as the feature extraction network, and the results are shown in Table 2.

Compared with the original FPN network, the BiFPN network can significantly improve the detection accuracy and segmentation accuracy of *Agaricus bisporus*. This indicates that using BiFPN as the feature fusion network can effectively compensate for the accuracy loss caused by the reduction in parameters when MobileNetV3 performs feature extraction.

Table 2

Comparison of Different Feature Fusion Networks

Feature fusion network	Independent growth		Adherent growth		Parameters / M	Floating-point operations / G	Frame rate/ FPS
	mAP@0.5 (box)	mAP@0.5 (mask)	mAP@0.5 (box)	mAP@0.5 (mask)			
FPN	0.841	0.761	0.821	0.636	20.46	152	10.2
BiFPN	0.882	0.793	0.860	0.654	24.46	173	9.9

Ablation Experiment

To verify the effectiveness of the improvement strategies proposed in this paper, ablation experiments were conducted, and the results are shown in Table 3.

The improved Mask R-CNN model based on MobileNetV3 and BiFPN networks has a parameter number of 24.46 M and floating-point operations of 173 G, which are 44.4% and 24.5% lower than the parameter number (43.971 M) and floating-point operations (229 G) of the original Mask R-CNN, respectively; the frame rate of the improved model is 9.9 FPS, which is 3.55 FPS higher than that of the original Mask R-CNN (6.35 FPS).

Table 3

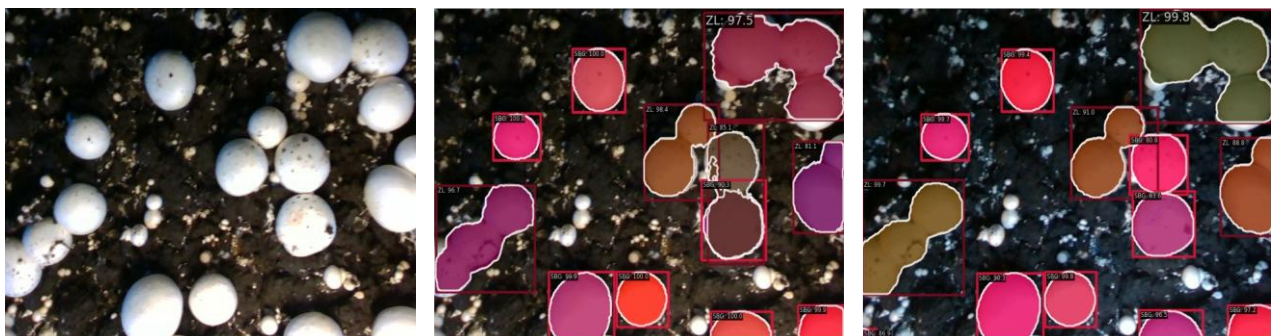
Results of the ablation experiment

Group	Mobile NetV3	BiFPN	Independent growth		Adherent growth		Parameters /M	Floating-point operations /G	Frame rate /FPS
			mAP@0.5 (box)	mAP@0.5 (mask)	mAP@0.5 (box)	mAP@0.5 (mask)			
1	×	×	0.929	0.927	0.823	0.831	43.971	229	6.35
2	√	×	0.841	0.761	0.821	0.636	20.46	152	10.2
3	√	√	0.882	0.793	0.860	0.654	24.46	173	9.9

Note: "√" indicates that this operation is performed, while "×" means that this operation is not performed.

Results of model visualization detection

Images were randomly selected from the validation set for testing, and the detection results of the improved Mask R-CNN and the original Mask R-CNN were compared, with the results shown in Fig. 5. Both models can accurately identify and segment the growth area of *Agaricus bisporus* and effectively distinguish between two categories: independent growth (labeled "SBG") and adherent growth (labeled "ZL"). However, compared with the original Mask R-CNN, the improved model, while showing reduced recognition accuracy, has significantly fewer parameters and lower computational load, with faster detection speed, making it more suitable for deployment on embedded devices.



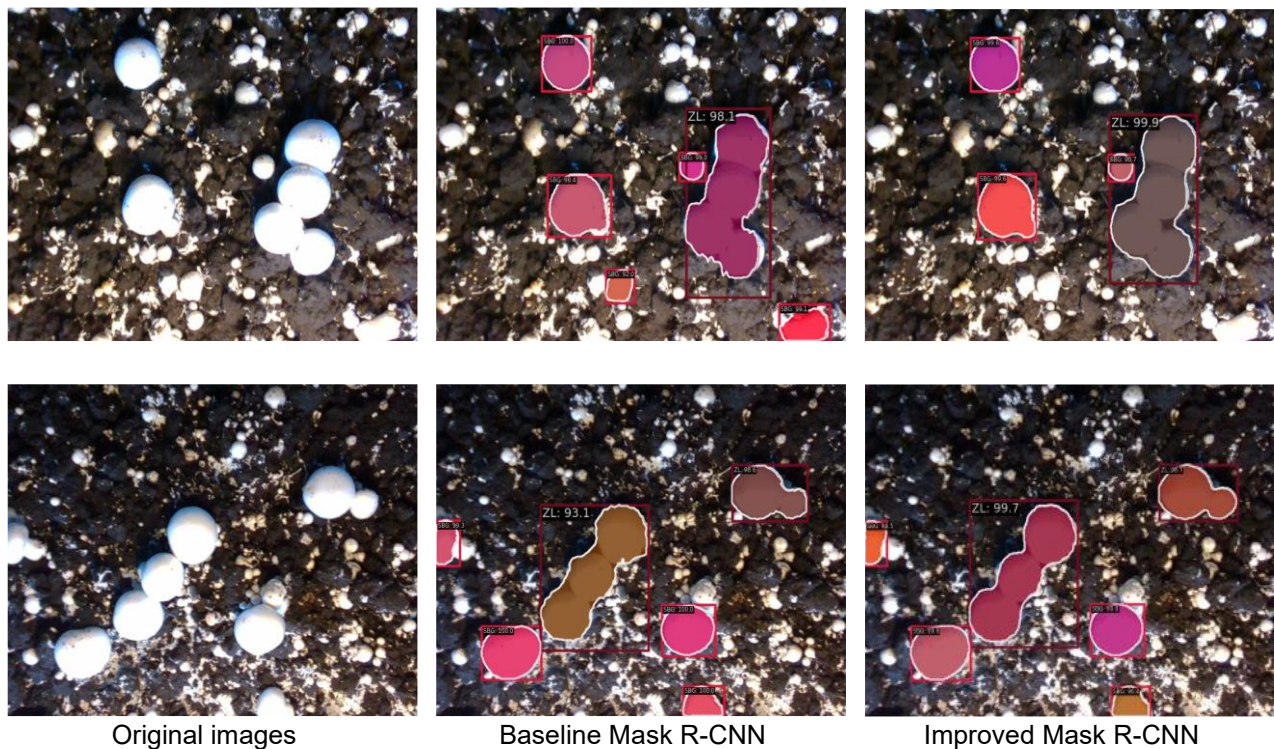


Fig. 5 - Recognition Results of Different Networks

Edge Device Deployment

To deploy the models on the edge device, the best.pt files of the baseline Mask R-CNN and improved Mask R-CNN models were converted to ONNX format respectively, first. Subsequently, the ONNX runtime and MMDeploy frameworks were set up on the development board. Finally, the ONNX models were deployed to the EC-R3588SPC edge device, and inference tests were conducted on the validation set, with the test results presented in Table 4.

Table 4

Results of the Edge Device Deployment

Model	Independent growth		Adherent growth		Parameters / M	Floating-point operations / G	Frame rate / FPS
	$mAP@0.5$ (box)	$mAP@0.5$ (mask)	$mAP@0.5$ (box)	$mAP@0.5$ (mask)			
Mask R-CNN	0.925	0.927	0.822	0.829	43.949	229	0.15
Improved Mask R-CNN	0.882	0.790	0.860	0.654	20.397	173	0.22

Due to the loss of some parameters during model conversion, the parameters of the converted models were slightly reduced compared with the models trained on the cloud server, leading to a slight decrease in detection accuracy. Owing to the hardware limitations of edge devices, the baseline Mask R-CNN model achieved a frame rate of only 0.15 FPS, while the improved Mask R-CNN model reached 0.22 FPS—representing an approximate 46.7% improvement in detection speed. This indicates that the improved Mask R-CNN model effectively reduces computational overhead on resource-constrained edge devices. The deployment results are shown in Figure 6.

It can be seen that the improved Mask R-CNN model can effectively segment the growth area of *Agaricus bisporus*—verifying the feasibility of the model deployed to the EC-R3588SPC edge device.



Fig. 6 - Deployment Results of the Improved Mask R-CNN model

CONCLUSIONS

This paper proposed a method for detecting *Agaricus bisporus* based on the improved Mask-RCNN model. The method introduced the lightweight MobileNetV3 to replace the ResNet50 feature extraction network in the backbone of Mask-RCNN, reducing model complexity; meanwhile, the BiFPN was used to replace the original FPN feature fusion network to enhance the model's ability to extract image features. The improved Mask-RCNN model's number of parameters and floating-point operations were reduced by 44.4% and 24.5% respectively, and the frame rate increased by 3.55 FPS, making it more suitable for deployment in the terminal AI control board of intelligent picking equipment. In future research, the model structure will be further optimized for practical scenarios, and the applicability of the model in the actual cultivation environment of *Agaricus bisporus* will be evaluated more comprehensively, so as to provide more reliable technical support for the research and application of picking robots.

ACKNOWLEDGEMENTS

This study was supported by the earmarked fund for Modern Agro-industry Technology Research System of Shanxi Province[2025CYJSTX09].

REFERENCES

- [1] Albayrak U., Golcuk A., Aktas S., Coruh U., Tasdemir S., Baykan O. K., (2025), Classification and Analysis of *Agaricus bisporus* Diseases with Pre-Trained Deep Learning Models. *Agronomy*, Vol. 15, Issue 1, pp.226-226, Madison/WI/USA;
- [2] China Edible Fungi Association, (2024), Analysis of the results of the 2022 National Statistical Survey of Edible Fungi. *Edible fungi of China*, Vol. 43, Issue 1, pp.118-126, Beijing/China;
- [3] Fang X.J., Yan L.Q., Zhang F.H., Gao X.Y., (2024), Development of crop image recognition based on deep learning. *Jiangsu Agricultural Sciences*, Vol. 52, Issue 20, pp.18-24, Nanjing/China;
- [4] He K., Gkioxari G., Dollár P., Girshick R., (2017), Mask R-CNN. In *Proceedings of the 2017 IEEE International Conference on Computer Vision*, pp. 2980-2988, Venice/Italy;
- [5] Howard A., Sandler M., Chu G., Chen L. C., Chen B., Tan M., Wang W., Zhu Y., Pang R., Vasudevan V., Le Q. V., Adam H., (2019), Searching for MobileNetV3. *CoRR*, Vol. abs/1905.02244, Ithaca/New York/USA;
- [6] Ishana A., Lalit K. A., Teek P. S., Priyanka R., (2023), A review of deep learning techniques used in agriculture. *Ecological Informatics*, Vol. 77, Amsterdam/Netherlands;
- [7] Javanmardi S., Ashtiani M. H. S., (2025), AI-driven deep learning framework for shelf life prediction of edible mushrooms. *Postharvest Biology and Technology*, Vol. 222, pp.113396-113396, Netherlands;
- [8] Li T.H., Dong Y.H., Shi G.Y., Zhang G.S., Chen C., Su J.C., (2024), Research on the DOA-BP-Based Temperature and Humidity Prediction Model for Commercial Cultivation of *Agaricus Bisporus*. *INMATEH Agricultural Engineering*, Vol.73, pp.149-161, Bucharest/Romania.
- [9] Lin T. Y., Dollar P., Girshick R., Hariharan B., Belongie S., (2017), Feature Pyramid Networks for Object Detection. *IEEE Computer Society*, pp.936-944, Honolulu/HI/USA;

- [10] Liu X., Mo T.T., Zheng W.R., Yan X.M., (2023), Intelligent Control Technology for Agaricus Bisporus Production: Research Status and Development Trend. *Journal of Smart Agriculture*, Vol. 3, Issue 2, pp.7-12, Harbin/China;
- [11] Lu W., Zou M.X., Shi H.N., Wang L., Deng Y.M., (2022), Technology of Visual Identification -Measuring-Location for Brown Mushroom Picking Based on YOLO v5-TL. *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 53, Issue 11, pp.341-348, Beijing/China;
- [12] Luo X., Zhang T.J., Liao J., Gao X.X., Lu Z.M., Pang X.B., (2021), The development status and trend of edible mushroom factories. *Agricultural Development and Equipment*, Issue 1, pp.114-115, Nanjing/China;
- [13] Luo Q., (2019), Research on Mushroom Species Recognition Algorithms Based on Deep Learning. *Edible fungi of China*, Vol. 38, Issue 6, pp.26-29+33, Beijing/China;
- [14] Menna A. E. A., Hatungimana M., Lin D.M., Lin Z.X., Sarah A. A., (2024), Mitigating Oxidative Stress and Promoting Cellular Longevity with Mushroom Extracts. *Foods*, Vol. 13, Issue 24, pp.4028-4028, Basel/Switzerland.
- [15] Redmon J., Divvala S. K., Girshick R. B., Farhadi A., (2016), You only look once: unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. abs/1506.02640, pp.779-788, Las Vegas/NV/USA;
- [16] Seetharaman K., Mahendran T., (2022), Leaf disease detection in banana plant using Gabor extraction and Region-Based convolution neural network (RCNN). *Journal of the Institution of Engineers (India): Series A*, Vol. 103, Issue 2, pp.501-507, New Delhi/India;
- [17] Tan M., Pang R., Le Q.V., (2020), EfficientDet: Scalable and efficient object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.10778-10787, New York/New York/USA;
- [18] Wang Z.Z., Li X.F., Song S.X., Liu M.Y., Dong T.L.G. (2018), Research progress on storage pretreatment and preservation technology of Bisporus. *Food Research and Development*, Vol. 39, Issue 4, pp.200-206, Tianjin/China;
- [19] Wangsa T. I., Agung H. K., Olivia M. T., Raymond R. T., Heni R., (2024), Agaricus bisporus Mannose-Binding Protein Stimulates the Innate Immune Cells. *Advanced pharmaceutical bulletin*, Vol. 14, Issue 4, pp.944-950, Tabriz/Iran;
- [20] Wang Y.L., Zhang F.K., YangDao R.N., He R.H., Zhu J.K., Li P., (2025), Detection method of cotton common pests and diseases based on improved YOLOv5S. *INMATEH Agricultural Engineering*, Vol. 76, pp.199-209, Bucharest/Romania;
- [21] Yang S.Z., Ni B.W., Du W.H., Yu T., (2022), Research on an Improved Segmentation Recognition Algorithm of Overlapping Agaricus bisporus. *Sensors*, Vol. 22, Issue 10, pp.3946-3946, Basel/Switzerland.