IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 23 May 2023, accepted 26 June 2023, date of publication 5 July 2023, date of current version 27 October 2023.
Digital Object Identifier 10.1109/ACCESS.2023.3292530

RESEARCH ARTICLE

Research on Asparagus Recognition Based on Deep Learning

JIE YU[1,2], CHAO ZHANG[1,2], JUNMEI WANG[1], MINGZHI ZHANG[1], XINGGUO ZHANG[1], AND XIAOQIN LI[3]
[1]College of Information Science and Engineering, Shandong Agricultural University, Tai'an 271000, China
[2]Agricultural Big Data Research Center, Shandong Agricultural University, Tai'an 271000, China
[3]Library of Shandong Agriculture University, Tai'an 271000, China

Corresponding author: Chao Zhang (zhangch@sdau.edu.cn)

This work was supported by the Shandong Provincial Natural Science Foundation of China project "Study on the comprehensive prevention and control of hemorrhagic fever based on spatial metrology and EOF model" under Project ZR201910220237.

ABSTRACT In recent years, there have been several outbreaks of extensive flooding in northern provinces. Asparagus stem blight disease and brown spot diseases have become increasingly serious, significantly reducing the yield of asparagus grown in large fields, and asparagus production has developed toward facility agriculture. To solve the problems of high labor costs, labor shortages, and low production efficiency faced by facility agriculture, the adoption of mechanized harvesting for asparagus is an inevitable trend. A prerequisite for mechanized harvesting is target detection. Compared with algorithms such as Faster R-CNN, which require the use features of candidate regions for classification and recognition and cannot meet the real-time requirements of mechanized harvesting, this paper proposes a YOLO-based asparagus recognition scheme that can quickly perform target detection with a detection accuracy of 85.45% and significantly enhanced interference resistance, which can greatly improve the production efficiency of facility agriculture and accelerate the mechanization process of facility agriculture.

INDEX TERMS Deep learning, image recognition, asparagus recognition, YOLO.

I. INTRODUCTION

Asparagus, because of the shape of asparagus stems, like bamboo shoots, known as asparagus, with anti-tumor, antioxidant, hypoglycemic and other effects. During the growth process, if the shoot tip is covered, without photosynthesis, it remains white, known as white asparagus, used for processing canned products. If the shoot tips appear on the ground, they will become green by photosynthesis, which is called green asparagus, which has a crisp taste and much higher nutritional content than white asparagus, and is mainly used for fresh food [1], [2], [3].

Asparagus is native to the Mediterranean region and has been cultivated on a relatively short scale in China, but the planted area is growing rapidly. According to statistics, as of 2020, China's asparagus planting area reached 1.511 million hectares, an increase of 2.04% year-on-year, accounting for 90.64% of the total global planting area; China's asparagus production reached 8.613 million tons, an increase of 3.70% year-on-year, accounting for 88.14% of the total global production, the planting area and production ranked first in the world. Among them, Shandong Province is one of the main bases of asparagus production, and the annual export of canned asparagus accounts for about 1/3 of the total national export, occupying a pivotal position in the national production and export of asparagus [4].

The harvesting period of facility asparagus is mainly concentrated in April-June and September-November, during this period, the quality of asparagus is the best, short harvesting time, high manual harvesting workload and high intensity, and timely harvesting is of great significance to improve the economic benefits of asparagus. With the intensification of the aging process in China, the transfer of the rural labor population to the cities and the continued spread of labor shortage due to the superposition of the COVID-19, labor prices have increased significantly, and in Shandong Province,

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

FIGURE 1. Original.

FIGURE 2. Gaussian filter.

for example, the average mu labor cost of facility cucumbers increased from ￥2,858.21 in 2011 to ￥4,305.44 in 2017, with an increase and an average annual growth rate of 50.63% and 7.07%, respectively (Data source: National Compilation of Cost-Effectiveness Information on Agricultural Products, Calendar Year). The facility asparagus industry has likewise encountered bottlenecks in industrial development, and there is an urgent need to solve the current dilemma by developing asparagus harvesting robots.

The first problem for asparagus harvesting robot is the accurate target detection of asparagus. Non-contact target information acquisition using machine vision can be used for asparagus identification. However, because the asparagus tip is small and the long rhizome part has insufficient features, resulting in difficult feature extraction, coupled with the fact that the tip is similar in color to other background noises such as weeds, with small color difference and low contrast, making asparagus recognition more difficult.

Method: Image pre-processing

In order to make the recognition rate of the target detection algorithm higher, image pre-processing is needed for the original image. There are many pixels in the original image that are too far away from the surrounding pixels, and these pixels will interfere with the key features in the post-processing. Gaussian filtering can be used to process such pixels.

Gaussian filtering is a linear smoothing filter that can eliminate Gaussian noise, and is a process of weighting the value of a pixel point in the original image and the pixel values of its surrounding points to average them, which can eliminate certain noise points in the original image.

FIGURE 3. Deep learning algorithm development process.

Since the experimentally acquired pictures are color images, using RGB three-channel Gaussian filtering can make the acquired images smoother and remove the noise in the images, as shown in FIGURE 1. The above image is the original image, and the next image is the Gaussian filtered image.

II. RELATED WORK

Traditional target detection algorithms usually consist of three stages: region selection, feature extraction and sign classification. Region selection is to select the locations in the image where the target may appear by sliding, but such a scheme has a lot of redundancy and increases the computational effort. Feature extraction is based on the completion of the first step, using an extractor for feature extraction, but the accuracy is not high due to the limitations of the extractor itself, which easily leads to a high misclassification recognition rate. Signs classification is performed on the features obtained in the second step, and the accuracy of classification is also not too high.

With the emergence of deep learning, a large number of parameter settings can extract better features and classifier performance is more powerful.

Deep learning-based target detection algorithms mainly include One-stage and Two-stage types. Two-stage is divided into three steps: feature extraction, candidate frame generation, and classification/location regression. The representative algorithms are R-CNN [5.6.7] series, SPP-Net, which have high accuracy but slow speed because of the candidate frame selection. One-stage is divided into feature extraction, classification/localization regression without the step of generating candidate frames, and the representative algorithms are SSD [8], RetinaNet [9], YOLO [10], [11], [12], [13], [14], SqueezeDet, DetectNet, OverFeat, etc. Compared with two-stage, it has low precision and fast running speed because no candidate frame is generated.

In the following, we will introduce the main deep learning algorithms according to the timeline.

R-CNN [5] (Regions with CNN features) was proposed by R. Girshick's team, where detection starts with extracting candidate frames, using CNN for feature extraction, and using SVM to classify and regress the extracted features.

In 2015, R. Girshick's team proposed Fast R-CNN [6], which is an innovation based on R-CNN, eliminating the limitation of data input and significantly improving the detection accuracy and speed.

In the same year, Faster R-CNN [7] was proposed again by R. Girshick's team, which integrated the generation of candidate regions, feature extraction, classification, and bounding box regression into the same network, and the overall performance was greatly improved.

YOLO (You Only Look Once) is the first one-stage based algorithm, which was proposed by R. Joseph et al. in 2015, and the subsequent YOLOv2, YOLOv3, YOLOv4, and YOLOv5 are all its improved versions.

Meanwhile, given that asparagus needs to be harvested mechanically, the speed of mechanical harvesting is a key issue that we need to consider. Due to the slow speed factor limitation of the two-stage algorithm target detection, and to obtain smaller packets, the smallest s version of YOLOv5 is used for target detection in this paper.

The structure of this paper is as follows, Chapter 1 introduces the structure of YOLO series algorithm. Chapter 2 introduces the basic process of asparagus identification based on YOLOv5. The third chapter introduces the improved YOLOv5-based asparagus recognition algorithm, and finally, the conclusion and outlook.

III. BASIC PROCESS OF ASPARAGUS IDENTIFICATION BASED ON YOLOv5

The structure of YOLO series algorithm is divided into four parts: Input side, Backbone, Neck, and Prediction.

The input side mainly includes three parts: Mosaic data enhancement, adaptive anchor frame calculation, and adaptive image scaling. Mosaic data augmentation is adopted to use 4 images randomly scaled, cropped and lined up for stitching, which enriches the data set and improves the robustness of the system. Mosaic data enhancement can improve the AP value of small targets, which enables the algorithm to improve the recognition rate of small targets, which helps us to record the data of asparagus growth cycle for subsequent modeling of asparagus growth process. In YOLOv5, in contrast to the previous algorithm that required a separate running program to calculate the anchor frame value, YOLOv5automatically embeds this function to adaptively calculate the anchor frame value. The adaptive adjustment of the anchor frame can be adapted to various image sizes to optimize the computational effort. Since the training images have different sizes and aspect ratios, the scaling needs to be modified uniformly, so the size of the black border after scaling and filling is uncertain, which will affect the real composition of the image and thus reduce the recognition rate. For this reason, an adaptive scheme is used in YOLOv5to fill the smallest black edges, thus improving the inference speed and recognition rate.

Backbone contains Focus structure and CSP structure. The most critical thing in Focus is the slicing operation, take YOLOv5s as an example, the image is first turned into a 304*304*12 feature map by slicing, and then it is finally turned into a 304*304*32 feature map after one convolution operation with 32 convolution kernels. Taking Yolov5s as an example, the CSP1_X structure is applied to Backbone network, and another CSP2_X structure is applied in Neck.

The Neck structure mainly adopts the FPN+ PAN approach, and the FPN is a top-down approach to fuse the

TABLE 1. Comparison of YOLOv5 algorithm.

Model	size (pixels)	mAP^{val} 0.5:0.95	mAP^{test} 0.5:0.95	mAP^{val} 0.5	Speed V100 (ms)		params (M)	FLOPS 640 (B)
YOLOv5s	640	36.7	36.7	55.4	**2.0**		7.3	17.0
YOLOv5m	640	44.5	44.5	63.3	2.7		21.4	51.3
YOLOv5l	640	48.2	48.2	66.9	3.8		47.0	115.4
YOLOv5x	640	**50.4**	**50.4**	**68.8**	6.1		87.7	218.8

information from high places by up-sampling pass to obtain the features. A bottom-up PAN is added behind the FPN layer, and the two are combined to fuse the parameters of the detection layer to obtain the features.

In Prediction, the regression loss function of YOLOV5 uses the GIOU Loss, which is optimized on the basis of IOU, to solve the problem when the bounding boxes do not overlap.

YOLO series is one of the most widely used target detection algorithms, and five versions have been released so far. In this paper, we use the latest YOLOv5 version for image recognition training. The existing YOLOv5 is divided into four models s/m/l/x. In order to facilitate deployment on mobile devices, this paper uses YOLOV5s version, which has the advantages of fast training speed and small result files. The specific differences are shown in TABLE 1

Asparagus identification using YOLOV5 includes the following steps.

1. Preparation of dataset and pre-training weights

The data set was sourced from the farm of Zhong Hui Gao Xin Company in Weifang City, Shandong Province, with a total of 2,000 images. Use the marking tool to perform the marking operation and generate the txt tag file. Select appropriate pre-training weights. In this paper, we use yolov5s.pt weights for training. The larger the training weight file is, the more accurate it will be, but the detection speed will be reduced accordingly. In order to realize mechanized harvesting later, we choose yolov5s.pt weights that are ideal for real-time and accuracy for training.

2. Modifying the training model configuration

The asparagus recognition model needs to modify the .yaml data configuration file. Since this paper only needs to recognize one plant, asparagus, the class value of the relevant .yaml file needs to be modified. Configure the GPU training model in the training master file, and modify the number of single training images appropriately according to the performance of the graphics card to avoid video memory alarm.

3. Result analysis

After the training is completed, the corresponding .exp file is generated in the run directory, which contains the last training result and the optimal result. Using tensorbord, you can view the relevant parameter changes during the training process and provide a basis for correction for the next training.

The specific environment of the experiment is as follows.

Experimental platform: Considering the CPU's own constraints, this paper is based on GPU for training, with the

FIGURE 4. Actual environment map of asparagus in greenhouse.

FIGURE 5. Annotation map of asparagus.

FIGURE 6. Recognition rate of asparagus.

relevant Anaconda and compiler environments deployed. This paper only involves a class recognition, modify the relevant configuration file can be, the labeled images and related files into the compiler. The experiment is implemented based on pytorch 1.10, Python 3.8, and CUDA 11.3. In this paper, the number of training rounds is set to 100, and the image size is the default 640*640.

Data set: The data set of this paper was collected at Weifang Mingji Institute of Water and Fertilizer Integration, Shandong Agricultural University, and the shooting effect is shown in FIGURE 4. This data set contains training set and validation set, and the training set and validation set are allocated in the ratio of 8:2, with a total of 2000 images. Annotate with open source annotation tools, and the annotation effect is shown in FIGURE 5.

The next figure shows the actual asparagus recognition. It can be seen that the asparagus recognition effect is good under natural light, and even smaller plants are not missing.

Results Analysis: As shown in FIGURE 7, the accuracy of asparagus recognition was 83.1 %, mAP _ 0.5 was 87.3 %, mAP _ 0.5: 0.95 was 34.4 %, and recall was 87.6 %. Compared with the two-stage algorithm, the accuracy is slightly lower, but because there is no regional sampling, the global information is better performed, and the real-time requirements are met, providing conditions for the subsequent mechanized harvesting of asparagus.

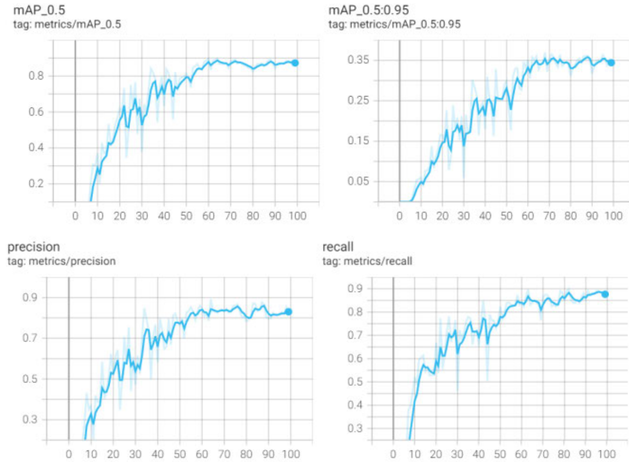

FIGURE 7. Recognition effect diagram of asparagus.

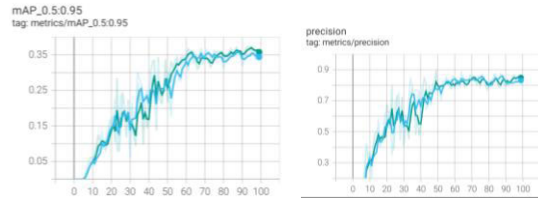

FIGURE 8. Comparison chart of improved asparagus identification.

FIGURE 9. NMS comparison chart.

IV. ASPARAGUS RECOGNITION ALGORITHM BASED ON IMPROVED YOLOv5

This section is designed for regression loss function and multi-objective frame problem in recognition. YOLOv5 uses GIOU regression loss function by default, we modify the GIOU regression loss function in the original model to CIOU regression loss function [15], [16], [17] to get faster convergence speed and higher accuracy.

GIOU means that there are two arbitrary bboxes A and B. It is necessary to find a minimal closed shape C such that C can enclose A and B. IOU $=\frac{|A\cap B|}{|A\cup B|}$, GIOU $=$ IOU$-\frac{|C\backslash(A\cup B)|}{|C|}$, $L_{GIOU}=1-GIOU$. As can be seen from the above equation, when the target frame and the prediction frame are closely surrounded, the values of IoU and GIoU are the same, at which time GIoU degenerates to IoU, and it is impossible to distinguish their relative position relationships. The improved

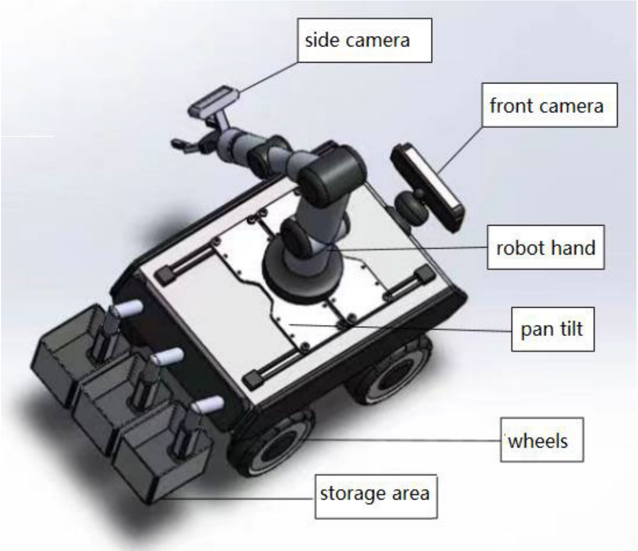

FIGURE 10. Harvesting robot.

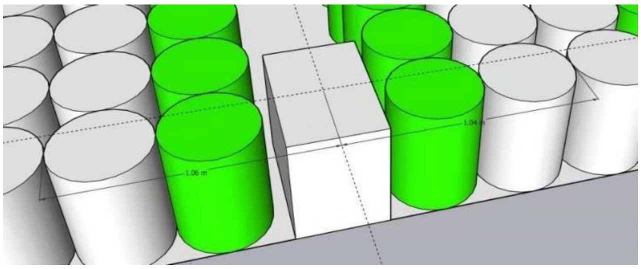

FIGURE 11. Schematic diagram of the scene.

algorithm DIoU incorporates the centroid normalized distance, which can better optimize such problems. For this purpose, $L_{DIOU} = 1 - \text{IOU} + \text{R}(\text{B}, B^{gt})$ is defined, where R is the penalty term of the prediction frame B and the target frame B^{gt}. In order to be able to make the algorithm converge with higher accuracy, CIOU considers the aspect ratio, and adds an influence factor $\alpha\upsilon$ to DIOU, which can take the aspect ratio fitting into account. Where $\alpha = \frac{\upsilon}{(1-IOU)+\upsilon}$, and υ can measure the aspect ratio consistency.

The results are shown in FIGURE 8, where the green curve is CIOU and the lavender line is GIOU. Since CIOU takes into account the information of the distance from the center point of the bounding box and the scale information of the aspect ratio of the bounding box, it can be seen that the improved MAP (0.5:0.95) is improved by 3.9% and the precision is improved by 2.35%.

For the multi-target box problem in recognition, we change the IOU_nms in the original model to DIOU_nms. The results are shown in FIGURE 9, the left figure shows the recognition under IOU_nms and the right figure shows the recognition under DIOU_nms, it can be seen that the right figure has stronger performance for the multi-target box problem and the anti-interference recognition rate is higher.

V. CONCLUSION AND FORESIGHT

In this paper, we implemented the detection of asparagus based on YOLOv5, in order to make it possible to deploy in the actual environment, we made improvements based on YOLOv5, the detection accuracy reached 85.45%, higher anti-interference detection ability, but the recognition accuracy is slightly less than the algorithm of Two-Stage class, basically meet the deployment requirements of the actual environment. The follow-up work we will realize mechanical harvesting of asparagus, combining target detection and harvesting robot, the robot collects asparagus height, thickness and other maturity data in real time, using the improved target detection algorithm to detect in real time and determine whether the harvesting requirements are met, and the mechanical arm will cut according to the strategy to meet the requirements and sort into the storage area of the robot. FIGURE 10 shows the schematic diagram of our designed harvesting robot, the side camera is used for target detection, the front camera is used for path survey and movement information collection, the robot arm is used for cutting asparagus and the storage area is used for storing the harvested asparagus.

The deployment is carried out in Weifang Zhong Hui GAO XIN Farm as shown in FIGURE 11, where green represents asparagus, white is the middle separation zone, and the square represents the harvesting robot, which performs target detection and asparagus harvesting in the travel lane. After the harvesting robot is put into use, it is expected to save about 75-80% of labor and other costs per mu, and one robot will replace 7-8 laborers, which will significantly shorten the fresh asparagus marketing cycle and will greatly improve the industry dilemma faced by facility agriculture such as labor shortage, high labor cost and low overall mechanization level, promote agricultural efficiency and farmers' income, promote the benign cycle development of facility agriculture and help rural revitalization.

REFERENCES

[1] L. Shuhua, *Cultivation Technology of Annual Asparagus With Tons of Yield*. Beijing, China: China Agricultural Press, 2016.
[2] C. Guangyu, "Current situation and trends of asparagus industry development in China," (in Chinese), *World Agricult.*, no. 10, pp. 181–186, 2013.
[3] H. Chaoxing, "Current situation and prospect of asparagus industry development in China," (in Chinese), *Vegetables*, no. 5, pp. 33–39, 2022.
[4] D.-Q. Song, J. Zhang, Z. Xue, and J.-P. Chen, "Current situation and countermeasures for the development of asparagus industry in China," (in Chinese), *Guangdong Agricult. Sci.*, vol. 38, no. 21, pp. 187–190, 2011, doi: 10.16768/j.issn.1004-874x.2011.21.022.
[5] 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.
[6] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1440–1448.
[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[8] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2016, pp. 21–37.
[9] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2999–3007.

[10] G. Jocher et al., "Ultralytics/YOLOv5: V5.0-YOLOv5-P6 1280 models, AWS, supervise.ly and YouTube integrations," Zenodo, Tech. Rep., 2021.
[11] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[12] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[13] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6517–6525.
[14] 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)*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788.
[15] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," 2019, *arXiv:1911.08287*.
[16] J. Yu, Y. Jiang, Z. Wang, Z. Cao, and T. Huang, "UnitBox: An advanced object detection network," in *Proc. 24th ACM Int. Conf. Multimedia*, 2016, pp. 516–520.
[17] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 658–666.

JIE YU received the M.A.Sc. degree in electronics engineering from Qufu Normal University, in 2017.

His current research interests include deep learning, image identification, and smart agriculture.

CHAO ZHANG received the B.Sc. degree from Qufu Normal University and the M.Sc. degree in basic mathematics from Wuhan University, in 2005. In 2005, he joined the Department of Mathematics, Shandong Agricultural University, as a Professor, an Associate Professor, and the Deputy Director of the Agricultural Big Data Center. He presided over two provincial and ministerial level scientific research projects, eight other projects, and published more than 40 papers, including more than ten papers, such as *Security and Communication Networks*, *Scientific Reports*, and IEEE SmartIoT. He has published one textbook for the National Twelfth Five-Year Plan, three textbooks for professional courses, six national invention patents, and nine software copyrights. His current research interests include agricultural big data analysis and mining, artificial intelligence and agricultural robots, intelligent control theory and related technologies, intelligent and automatic identification, monitoring devices, agricultural/intelligent agriculture and related automation, crop growth detection and control in facility agriculture, plant phenotype monitoring/robot vision, and technical teaching and scientific research related to harvesting robots. In 2020, he served as a Distinguished Expert for the Feicheng Rural Revitalization Industry Research Institute and a member of the Mount Taishan Scholar Team. In 2019, the research and application of key technologies for digital management of the vegetable industry won the Third Prize from the Shandong Provincial Department of Education. He was a recipient of the 2018 International Conference ICC CS2018 Best Paper Nomination Award.

JUNMEI WANG was born in November 1977. She received the master's degree in cybernetics. She is a lecturer. She mainly serves as the teacher of public basic courses. She has led one scientific research project, three teaching and research projects, participated in two general projects of the Shandong Natural Science Foundation, and published more than ten articles. She has edited two textbooks. Her current research interests include agricultural big data analysis and mining, smart agriculture, and integrated demonstration of industry, academia, research, and application. She has won the Second Prize in the National Design Competition for mathematics micro course teaching in colleges and universities, the Grand Prize of the East China Division, the Grand Prize of Shandong Province, the Second Prize of Shandong Province in the First Shandong Province University Mathematics Course Teaching Innovation Demonstration and Exchange Activity, and the First Prize of Shandong Agricultural University Teaching Quality Award.

MINGZHI ZHANG was born in October 1998. He received the bachelor's degree in automation from the School of Information Engineering, Yantai University of Technology, in 2020. He is currently pursuing the master's degree in agricultural engineering and information technology with Shandong Agricultural University. His current research interests include the agricultural Internet of Things technology and engineering.

XINGGUO ZHANG was born in July 1998. He received the B.S. degree in the Internet of Things engineering from Shandong Agricultural Engineering College, in 2020, and the M.S. degree in agricultural engineering and information technology from Shandong Agricultural University, in 2023. His current research interests include asparagus image edge recognition and blockchain technology.

XIAOQIN LI was born in March 1981. She received the bachelor's degree in chemical education from Jianghan University, in 2002. She is currently pursuing the master's degree in agricultural engineering and information technology with Shandong Agricultural University.