Original Article
Food Science and Technology

ISSN 0101-2061 (Print)
ISSN 1678-457X (Online)

DOI: https://doi.org/10.1590/fst.95922

Boosted Mask R-CNN algorithm for accurately detecting strawberry plant canopies in the fields from low-altitude drone images

Ping LIN[1,2], Huazhe ZHANG[2], Feiyu ZHAO[1], Xiaoxuan WANG[1], Huan LIU[1], Yongming CHEN[1,2]*

Abstract

The research constructs a novel structure by integrating two parts of online object detection pipelines into the current state-of-the-art Mask R-CNN algorithm to improve the detection performance. The DJI Mavic Air drone is used to collect low-altitude sensing images of strawberry plant canopies. The data augmentation method is employed to feed more instances into the original image dataset to boost the generalization and robustness of the detection model in the training procedure. A ResNet50 backbone combined with a feature pyramid network is presented to extract the features of strawberry plant canopies. The online hard example mining algorithm is introduced to mine hard samples to learn rich features and update model weights. Soft non-maximum suppression based on recursive application on the remaining detection boxes within the predefined overlap threshold is proposed to improve the performance of identifying the large complex overlapping area of strawberry plant canopies. The qualitative results demonstrate that the improved detection model had an AP50 of 96.9 and an AR of 78.5 on the test set, which are approximately 30% higher than the original values.

Keywords: strawberry plant canopy; feature pyramid network; online hard examples mining; soft non-maximum suppression.

Practical Application: Provide new technology to detect outdoor strawberry plant canopies by drone.

1 Introduction

In the agricultural production process, low-altitude drone sensing technology makes it easier to obtain field crop image data, and deep learning technology makes crop management more operable (Chen & Yu, 2022; Zhang & Xu, 2021). By combining their advantages, the growth and health of crops can be monitored in real time. Production analysis managers can use these data to provide a decision-making basis for future crop yield estimation and planting cost investment in the coming year. At present, the statistics of strawberry plants mainly rely on manual labour. Due to its high intensity, low efficiency and very large workload, manual labour can no longer meet the needs of modern strawberry planting's intensive production and refined management. However, with the development of aerial and artificial intelligence technologies, integrating low-altitude drone sensing technology with deep learning technology has become more accessible are is used to accurately and automatically identify and monitor the health and growth status of strawberry plants in real time. As a popular topic, object detection continues to develop in the field of computer vision and is widely used in the industrial and information fields, such as face recognition (Schroff et al., 2015; Wang & Deng, 2021) and unmanned driving (Zhang et al., 2016) and other fields (Li et al., 2019). In recent years, object detection has gradually extended to agricultural applications. For instance, Image segmentation based on k-means clustering was studied for food background subtraction. Two different kinds of colour and texture features were extracted, and both were fed into the learning machine. These combined discriminative feature descriptors performed better than using the individual colour and texture features for fruit and vegetable recognition (Dubey & Jalal, 2015). Although many methods have been proposed to address the problem of agricultural object detection, it is still a challenging task to establish an accurate and reliable detection system under a complex agricultural planting environment.

With the advancement of deep learning technology, object detection technology based on convolutional neural networks has further improved the accuracy and robustness of detection algorithms. (Lin et al., 2020) compared different region-based object detection methods to address strawberry flower examples. A state-of-the-art deep-level object detection framework was developed to visually represent instances of strawberry flowers in outdoor fields and improve detection accuracy. The final implemented Faster R-CNN (Ren et al., 2017) model achieved better performance than R-CNN and Fast R-CNN in detecting instances and had a shorter execution time. The detection accuracy rate was 86.1%, which can effectively address strawberry flower instances from various camera viewpoints, different flower distances, overlapping, complex background lighting, blurring, etc. (Yu et al., 2019) leveraged Mask R-CNN (He et al., 2017) to detect strawberry fruits in unstructured agricultural environments. The visual localization method of strawberry picking points was performed after detection of mask images

Received 20 Aug., 2022
Accepted 02 Oct., 2022
[1] *School of Electrical and Engineering and Automation, Hubei Normal University, Huangshi Hubei, PR China*
[2] *College of Electrical Engineering, Yancheng Institute of Technology, Yancheng, Jiangsu Province, PR China*
*Corresponding author: billrange007@gmail.com

that generate ripe fruits. The average detection accuracy rate was 95.78%, and the recall rate was 95.41%. The prediction results of fruit picking points showed that the average error was ±1.2 mm. The proposed method exhibited higher generality and robustness in unstructured environments. (Shin et al., 2021) applied deep learning to detect powdery mildew and persistent fungal disease in strawberries to reduce unnecessary fungicide usage and the need for field scouts. Experimental results showed that ResNet-50 provided the highest classification accuracy of 98.11% when classifying healthy and infected leaves; however, considering the computation time, Alex Net had the fastest processing time of 40.73 s and an accuracy of 95.59%. Most of these studies on strawberry cultivation focused on fruits, flowers, diseases, etc. As an important link in the growth state of crops, strawberry plants are often easily overlooked. However, they have a profound impact on the agricultural production process. Through online monitoring of the growth status of strawberry plants, high yield and the highest quality of strawberry fruit can be controlled effectively.

In this article, we construct a novel object detection framework by integrating two parts of online object detection pipelines into the current state-of-the-art Mask R-CNN algorithm to identify strawberry plants. The feature pyramid network is utilized to improve the representation performance of instances at different scale levels. The algorithm of online hard example mining is introduced to mine hard samples to improve the robustness of the model.

The ultimate goal of this study is to build a deep-level artificial convolutional neural network architecture with an online object recognition mechanism to accurately and effectively indicate the regions of strawberry plant canopies in the fields. Aiming at the problem that there is currently only limited research on strawberry plant detection, the boosted deep-level region-based artificial intelligent strawberry plant canopy visual recognition system provides valuable information for farmers to predict strawberry yield and analyses fruit growth status. Simultaneously, the strawberry plant is the main obstacle in the vision system of the strawberry picking robot, and accurately marking the area of the strawberry plant helps to improve the picking accuracy. At present, strawberry yield estimation based on manual counting is time-consuming and labour-intensive, and the strawberry picking robot vision system is easily disturbed by obstacles. Accurate positioning of strawberry plants can provide a solution for strawberry planting yield estimation and automated picking.

2 Materials and methods

2.1 Experimental instrument and aerial image acquisition

A DJI Mavic Air drone (The DJI Technology Co., Ltd., Shenzhen, China.) was used to collect the low-altitude sensing image data of the strawberry plants. It was designed to be smartphone-sized to fit in a jacket pocket and featured a 12 MP 4K HDR camera mounted on a 3-axis gimbal with a remote controller. This drone had an intelligent exposure system to help us achieve the perfect lighting conditions for the captured aerial photos. It also had a 1,260 s flight time and a 4,000 m flight range. For parallelizing detection algorithms in deep learning architectures, the hardware was powered by an Intel® Core™ i5-9400F Processor, 16 GB of RAM, NVIDIA® GeForce®RTX 2060 SUPER™, and the programs were accelerated by the deep learning framework of PyTorch on the Ubuntu system.

The experiment was carried out on modern strawberry farmland in Yangshi Village, Yandu District, Yancheng, China (33°16'13.18"N120°5'46.05"E). A DJI Mavic Air drone was used to obtain the low-altitude sensing image data of strawberry plants in the visible spectrum band. The experimental location, environment and equipment are shown in Figure 1. The average aerial height was controlled at 4.6 m. The images were stored in JPEG format with a resolution of 4056 × 3040. A few initial blurred images were excluded manually. To reduce the computation and running time of training and testing the object detection model, the large margins at the edge of the field were excluded by cropping. To overcome the influence of weather and light conditions, we collected low-altitude drone sensing image data of strawberry plant canopies three times in different weeks during the daytime. There were 500 low-altitude drone sensing

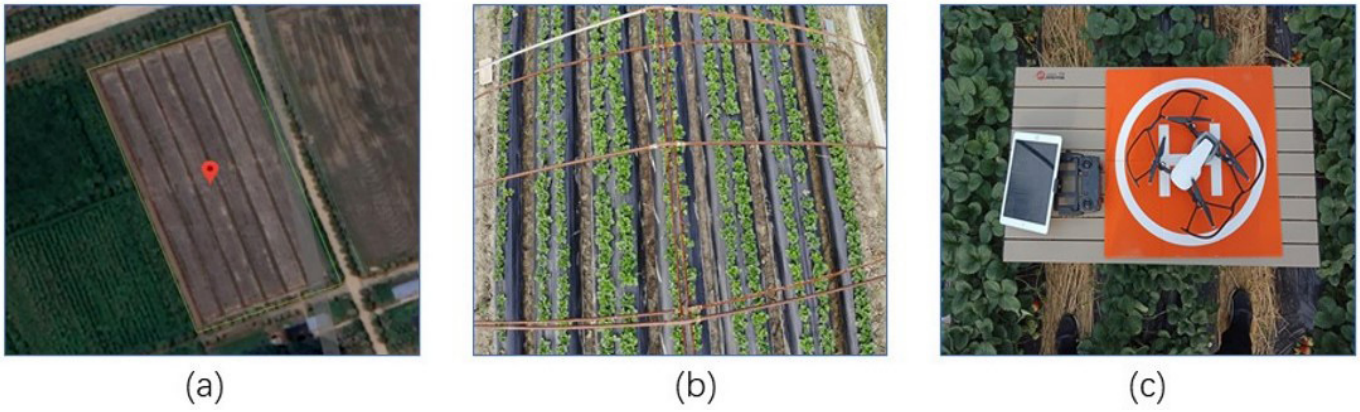

Figure 1. (a) Google satellite map indicating the position of the strawberry fields, where the green dotted frame indicates the experimental photograph of the farmland, and the red dot denotes the position coordinate (33°16'13.18"N120°5'46.05"E). (b) The bird's-eye view of a low-altitude drone sensing image of a strawberry plantation captured by a 12 MP 4K HDR camera mounted on a 3-axis gimbal. (c) The DJI Mavic Air drone and remote controller equipment.

images of strawberry plants in the visible spectrum selected as the experimental dataset.

2.2 Dataset construction and augmentation

The dataset was randomly divided into a training set, validation set and test set at a ratio of 3:1:1. The experimental data were annotated by the image annotation tool Labelme (Russell et al., 2008) to generate the ground truth boxes and segmentation masks of the strawberry plant and were transformed into the format of a COCO dataset (Lin et al., 2014). Figure 2 shows some examples of simple and complex samples and the generated mask image during the labelling process. Simple samples were labelled according to the outer contour of the strawberry plant canopy, and complex samples required discrimination assistance by combining them with the strawberry rhizome and the orientation of the leaves of the plant. These mask images were used to update the weight parameters of the strawberry plant detection model to optimize the performance. In addition, the performance of the training model was evaluated by comparing the annotated mask image with the prediction results of the strawberry plant detection model.

In fact, in the process of data collection, it is difficult to cover all areas. To improve the generalization performance of the training model and prevent overfitting during the training process, some data augmentation schemes are performed to expand the size of the data sample. Figure 3 shows some samples after data augmentation. The far left image shows the original image of the strawberry plant canopy. Figure 3a shows the image translated in the horizontal direction by the magnitude number of pixels, Figure 3b shows the image sheared along the horizontal and vertical axes with the rate magnitude, and Figure 3c shows the image rotated in magnitude degrees. These first three methods were mainly used to increase the amount of training data to adapt to different scenarios and improve the generalization of the model. Figure 3d shows the image histogram equalized, Figure 3e shows inverted all pixels of images, and Figure 3f shows adjustments of the brightness of the image to adapt to different lighting environments. The latter three methods were implemented mainly to increase the noise data to improve the robustness of the model.

2.3 Boosted Mask R-CNN algorithm

Figure 4 shows the original Mask R-CNN strawberry plant detection, and the improved scheme is represented by the plus module in the figure. Object segmentation is performed in parallel with the identification and localization tasks shown in the orange dotted rectangle. The detection task can be divided into two-stage detection components: the first stage scans the image and generates the region proposal, and the second stage classifies the proposals and generates bounding boxes and segmentation masks. The image is first fed into the backbone to extract features and produce the corresponding feature maps. Then, the feature map obtains a considerable number of regions of interest (ROI) through the sliding-window class-agnostic object detector of region proposal network (RPN). In the original RPN design, a small subnetwork is evaluated on dense 3×3 sliding windows,

Figure 2. Mask definition of training samples based on the distribution of the outermost contour of the strawberry plant canopy. (a) Simple sample of independent plants. (b) Complex samples of overlapping plants.

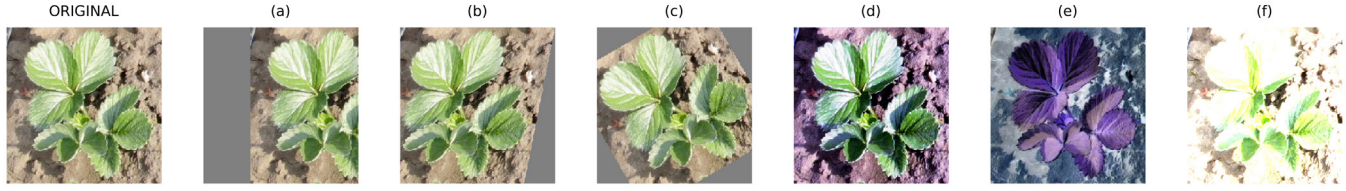

Figure 3. Illustration of the original strawberry plant image (on the far left) and six different kinds of transformed images by the (a) translating (b) shearing (c) rotating (d) equalizing (e) inverting (f) and brightening data augmentation techniques (from the second on the far left to the right).

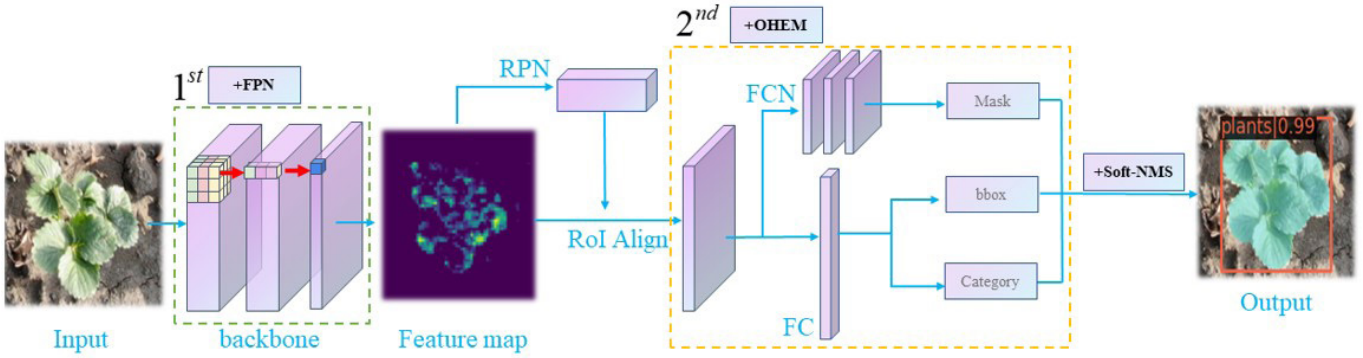

Figure 4. Flow chart of Mask-RCNN for detecting the strawberry plant canopy. The improved scheme introduces FPN in the first stage, uses OHEM in the second stage, and replaces the last part of the framework with Soft-NMS.

on top of a single-scale convolutional feature map, performing binary classification of object and background and bounding box regression. Both the feature map and the ROI generated by RPN are sent to the RoIAlign layer, enabling each ROI to generate a fixed size feature map. Finally, the flow passes through two branches: one branch enters the fully connected layer for object classification and frame regression, and the other branch enters the full convolutional network (FCN) for pixel segmentation, which is equivalently a CNN without fully connected layers a neural network that only performs convolution (and subsampling or upsampling) operations.

To effectively discriminate overlapping and hidden strawberry plants, novel algorithms are presented to improve the universality and robustness of the original model in the background of a complex and unstructured agricultural planting environment. In the backbone, the algorithm of the feature pyramid network (FPN) is used for multiscale training. In the R-CNN training process, the algorithm of online hard example mining (OHEM) is used to update the weight parameters of the model. In the postprocessing process, Soft-NMS is proposed to replace the traditional NMS algorithm. Soft non-maximum suppression based on recursive application on the remaining detection boxes within the predefined overlap threshold is proposed to improve the performance of identifying the large complex overlapping area of strawberry plant canopies. The principle of each related algorithm is detailed in the following sections.

2.4 R-CNN backbone

By designing different weight layers, neural network models with different depths can be established, such as VGG (Simonyan & Zisserman, 2014), GoogleNet (Szegedy et al., 2015) and ResNet (Chu et al., 2022; He et al., 2016). Although deeper networks may achieve higher accuracy, the training and inference speed of the model will decrease. Since the residual structure does not increase the model parameters, difficulties of gradient disappearance and training degradation can be effectively alleviated, and the convergence of the model can be improved. Therefore, ResNet50 is used as the backbone network for feature extraction in this paper. In the backbone network, early layers detect low-level features such as edges and corners of the strawberry plant canopy, and later layers successively detect higher-level features such as strawberry plant canopy shape and textural characteristics.

Although the above backbone works well, it is still difficult to address the problems of multiscale feature analysis. The semantic information of the low-level feature is relatively small, but the object location is accurate, while the semantic information of high-level feature is richer, the object location is relatively rough. The feature pyramid network (FPN) was utilized to improve the representation performance of instances at different scale levels. The architecture of FPN is shown in Figure 5. FPN (Lin et al., 2017) improves the standard feature extraction pyramid by adding a second pyramid that takes the high-level features from the first pyramid and passes them down to lower layers. By doing so, features at every level have access to both low- and high-level features.

2.5 Online hard example mining

The problem of sample imbalance often occurs in object detection tasks, which is mainly reflected in two aspects: the imbalance of positive and negative samples and the imbalance of difficult and easy samples. The original Mask R-CNN strawberry plant detection model generally uses randomly selected samples, maintaining the ratio of positive and negative samples of 1:3. During the training process, positive samples can be divided into easy positive samples and hard positive samples, and there is only one type of negative sample. The easy sample is a simple sample for the model. It is difficult for the model to obtain more information from this sample. Its loss function becomes very small, and the loss function has a relatively small influence on the gradient of the input. A hard sample is a difficult sample for the model. The gradient information produced will become richer, which can be better used to guide the direction during the model optimization process.

The algorithm of online hard example mining (OHEM) (Shrivastava et al., 2016) is introduced to mine hard samples to improve the robustness of the model. In the Mask R-CNN detection model, ROIAlign and its subsequent networks are called ROI networks. OHEM manages online hard example mining

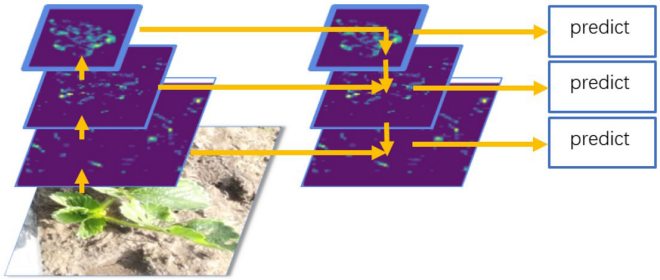

Figure 5. The architecture of the feature pyramid network, where the top-level semantic information can be merged with the object location information in low-level visual features through upsampling.

by building two ROI networks: the original Mask-RCNN and a read-only layer that participates in forward propagation but not backward propagation. Proposals generated by the selective search are mapped to feature maps in the RoIAlign layer and transformed into fixed inputs. The OHEM layer calculates the loss and sorts it. Finally, this layer selects the specified number of RoIs with larger loss and passes them to the ROI network for model training to update the whole network by hard examples backward propagation.

2.6 Soft non-maximum suppression

Non-maximum suppression (NMS) is an important part of the object detection process. NMS first sorts the proposal boxes according to the score from high to low. Then, the M boxes with the highest scores are selected, and other boxes that overlap with the selected proposal box are suppressed. This process is recursively applied to the remaining detection boxes. According to the principle of the algorithm, if an object falls within the pre-set overlapping threshold, the object may not be detected. That is, when the two object boxes are close to each other, the box with a lower score will be deleted because of its large overlap area. The growth characteristics of strawberry plants lead to this phenomenon, and the overlapping area between adjacent strawberry plants is too large to be removed by NMS.

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

$$s_i = s_i e^{-\frac{iou(M,b_i)^2}{\sigma}}, \forall b_i \notin D \quad (2)$$

Equations 1 and 2 describe the traditional NMS and the Soft-NMS algorithms, respectively, where s_i represents the score of the i-th box, b_i represents the corresponding pending box, M is the box with the highest score currently, N_t represents the specific user-defined threshold, iou denotes the overlapping based weighting, σ is the variance, and D is a list for storing processing results.

Different from Equation 1, which directly deletes the boxes with a higher degree of overlap, Soft-NMS uses a Gaussian function to decay the confidence scores of all other boxes according to the overlapping areas with the selected box.Soft-NMS will be only used in the subsequent inference process, where no additional training process is required, and no additional parameters are added (Bodla et al., 2017).

2.7 Evaluation criteria

COCO detection evaluation metrics are usually used to grade a model. COCO defines several average precision (AP) and average recall (AR) metrics of multiple intersection over unions (IoUs) using different thresholds, including ten different thresholds of 0.5 to 0.95 [0.5: 0.05: 0.95]. The IoU mainly measures the degree of overlapping area between the ground truth and the detected result from prediction. IoU score can be determined via (Lin et al., 2020) (Equation 3):

$$IoU = \frac{DR \cap GT}{DR \cup GT} \quad (3)$$

where DR represents the area in the detected region, and GT is the ground-truth region. The obtained IoU will be used to determine the bounding box and mask.

Precision and recall indices are the fractions of relevant instances among the retrieved instances and the actual total number of relevant instances, respectively, defined as follows (Lin et al., 2020) (Equations 4 and 5):

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

where TP is the number of cases that are positive and detected positive, FP is the number of cases that are negative but detected positive, and FN is the number of cases that are positive but detected negative.

The overall performance of the algorithm is measured by AP and AR. AR is the maximum recall of a fixed number detected in each image, averaged over the category and IoU. AP is estimated by using the area under the precision-recall (PR) curve, which is computed as follows (Lin et al., 2020) (Equation 6):

$$AP = \int_0^1 P(R)dR \quad (6)$$

where R indicates the recall and P denotes the precision.

3 Results and discussion

3.1 Model fitting

Three object detection models with different architectures of ResNet50, ResNet50+FPN, and ResNet50+FPN+OHEM were used to recognize the strawberry plant canopy. The training loss was calculated to evaluate the fitting performance of the model during the training procedure. A smaller training loss means that the object detection model acquires better training performance. Figure 6 plots the training loss curves of three different strawberry plant detection models over 1000 iterations. The red, green, and blue lines show the training loss procedures generated by using the corresponding architectures of ResNet50, ResNet50+FPN, and ResNet50+FPN+OHEM. It is clearly observed that the training

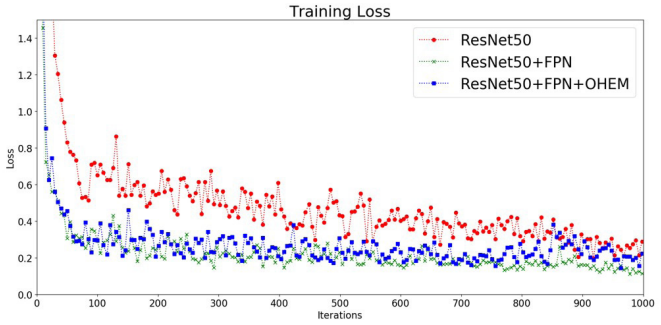

Figure 6. Training loss curves of three different strawberry plant detection models over 1000 iterations.

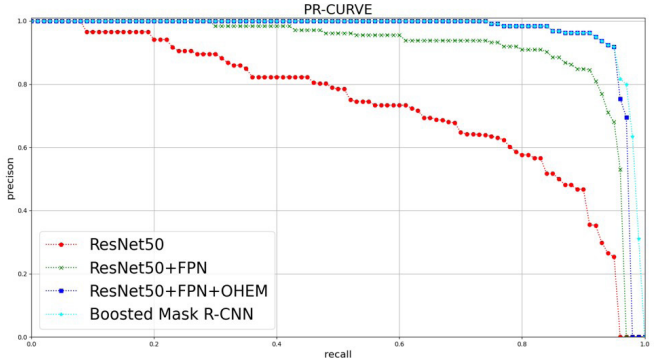

Figure 7. Precision-recall curves of four strawberry plant detection models with different architectures when the IoU is set to 0.5.

loss obtained by the red line was higher than the green and blue lines. This may have occurred because, in the training process, the shallow network had limited resolution and solely learned the detailed features of the plants, while the deep network had diverse resolution and solely learned semantic features, which made the feature description of small objects difficult to capture by the original detection network. The evolutionary network generated by the embedded FPN framework merged both detailed and semantic features to enhance the performance of the detection models. This might be due to FPN performing the multiscale detection from the strawberry plant canopy image by feeding larger and more sized pictures, which could finally improve the flexibility, robustness and convergence speed of the detection model to a certain extent.

Comparing the training loss curves of the green and blue lines, it can be seen that the training effect of the former is better than that of the latter. This is because the number of easy samples in the model training process had an absolute advantage in the overall sample. Even if the loss function of a single sample was small, the cumulative loss would dominate the final loss function, and this part of the samples could be well classified by the model. However, the parameter update guided by this part did not improve the discrimination performance of the model. The hard sample had a higher loss function and diversity for a single sample. By learning their complex characteristics to update the weight of the model, the generalization of the strawberry plant canopy detection model can be significantly improved. Although the convergence of the green line was close to the blue curve, the weight of the model was not significantly updated, which will be verified and observed in the subsequent experimental results.

3.2 Model enhancement

The precision-recall (PR) metric was used to evaluate the quality of two different types of models with traditional NMS and Soft-NMS in detecting strawberry plant canopies. PR curves close to the topline indicate a better test performance level than those close to the baseline. In other words, to top curve has a better performance level. The Soft-NMS algorithm was adopted in the postprocessing process, which is used to replace the traditional NMS algorithm without supplementing the additional training process or parameters. As shown in Figure 7, the PR curve of boosted Mask R-CNN using Soft-NMS in cyan was much superior to ResNet50 in red and ResNet50+FPN in green with the traditional NMS. The PR curves of the improved Mask R-CNN and ResNet50+FPN+OHEM with the traditional NMS overlapped each other at the beginning stage, but the boosted Mask R-CNN obtained higher precision than the ResNet50+FPN+OHEM when the recall threshold became much larger. It was verified that the previous Mask-RCNN model merged with the innovative framework of Soft-NMS could effectively boost the model performance of discriminating the strawberry plant canopy.

The overall performance of the algorithm was measured by AP, which is calculated from the area under the precision-recall curve above. A larger area under the PR curve indicates a better overall performance of the algorithm; that is, the closer the PR curve is to the upper right, the better the recognition performance of the corresponding algorithm will be obtained. Table 1 reports several evaluation metrics, including AP with IOU set at different thresholds, AR and frames per second (FPS). In terms of model performance evaluation, by introducing FPN into the backbone, AP50 increased from 72.7 to 91, and the accuracy increased by 25.2% compared to the model without FPN. In addition, when FPS was used as an indicator of model calculation, efficiency increased by 109%, which demonstrates the effectiveness of merging FPN into the backbone for multiscale training. In the training process, OHEM was added to the model to mine hard samples to learn the diverse characteristics of strawberry plants and update the weight parameters of the model. The performance of the model was improved by 5% compared to the previous model. Finally, in postprocessing, instead of the traditional NMS without tuning the additional training process or parameters, Soft-NMS was added. The accuracy of the strawberry plant detection model could be improved by another percentage point. Comparing the original Mask R-CNN model and the final boosted model, the results showed that the parameters of AP50, AR and FPS increased by 33.2%, 33.7%, and 93.1%, respectively. It could be concluded that the presented network structure significantly improved the detection performance while keeping the calculation speed fluctuations small.

3.3 Qualitative results and analysis

In this section, a few qualitative results of detecting strawberry plant canopies are shown. There were two strawberries with overlapping leaves in Figure 8a. However, only one strawberry was detected, marked by the red bounding box, by using the original Mask R-CNN detection framework with the traditional NMS algorithm. In contrast, as shown in Figure 8b, two strawberries were successfully identified by using the Soft-NMS algorithm, which were marked by the two red bounding boxes. The developed algorithm shows the potential of overlapping strawberry plant detection.

Figure 8. Comparison of the algorithms of the traditional NMS (a) with Soft-NMS (b) for detecting the strawberry plant canopy on the test set. The segmentation masked canopy area is marked by the green colour, and each plant object is marked by the red bounding box with the corresponding confidence score on the upper left corner of the bounding box.

Figure 9 shows more general qualitative results obtained after randomly selecting samples on the test set by using the improved strawberry plant detection model. The red box indicates the identified strawberry plant, and the blue box denotes the ground truth. It can be seen from the test results that the improved model gained robust and accurate performance in strawberry plant detection applications in complex and unstructured farmland environments.

In this paper, a boosted Mask R-CNN strawberry plant detection algorithm is first proposed. The blue dashed line in Figure 10a and c completely surrounds the red solid line. The experimental results show that the enhancement scheme can significantly improve the performance of the strawberry plant detection model without adding additional parameters. However, in Figure 10b, when the recall rate is close to 0.9, the red solid line and the blue dotted line intersect, and the PR indicator cannot judge the quality of the model. This may be due to the uncertainty of the outdoor environment when strawberry plants are used as the research object. However, the research of (Sun et al., 2020) and (Xu et al., 2020) demonstrated that the model performance was significantly improved on other application objects. The application of the above qualitative results on the test set additionally proves this.

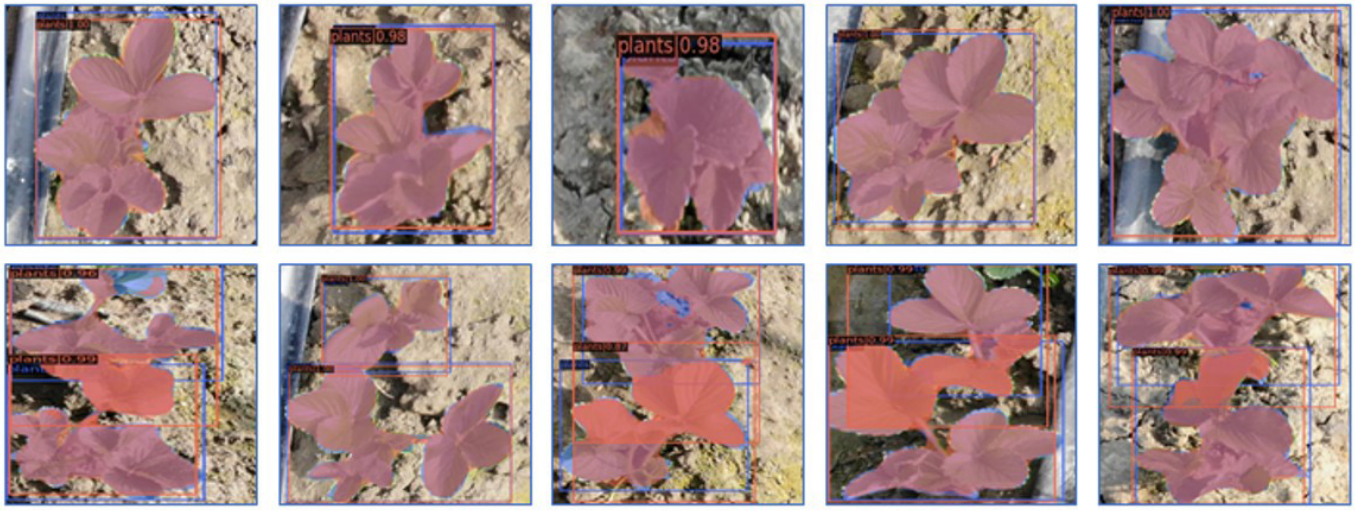

Figure 9. Results of the improved Mask R-CNN object detection algorithm identifying the strawberry plant canopy on the test set. The red bounding box indicates the detected strawberry plant object, and the blue box denotes the ground truth. The confidence score was placed on the upper left corner of the bounding box.

Table 1. Detection results of Mask-RCNN-related models with four different architectures on the strawberry plant canopy dataset.

Model	AP	AP50	AP75	AR	FPS
R50	0.389	72.7	0.398	58.7	5.8
ResNet50+FPN	0.605	91.0	0.690	69.4	**12.1**
ResNet50+FPN+OHEM	0.710	95.8	0.843	75.2	11.8
Boosted Mask R-CNN	**0.722**	**96.9**	**0.864**	**78.5**	11.2

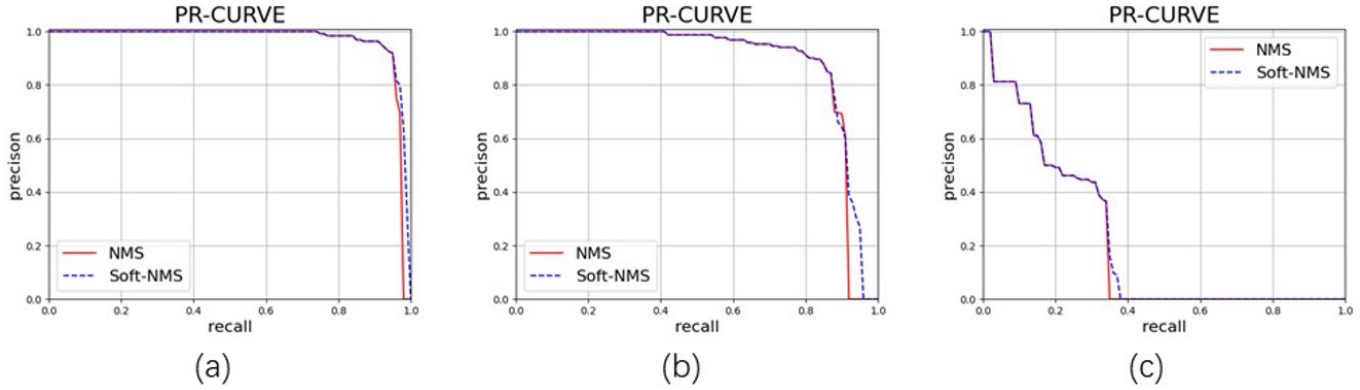

Figure 10. The normal red solid line model and the enhanced blue dotted line model detect the PR curves of strawberry plants at different IOU thresholds. The IOUs for (a), (b), and (c) were set to 0.5, 0.7, and 0.9, respectively.

3.4 Practical importance of the findings

With help of the proposed new technology of aerial visual data and image-based deep learning analytics for detecting outdoor strawberry plant canopies, farmers can save time and energy for the larger-scale strawberry plant growing management. From the marked aerial imagery, growers gain real-time insights on setting the amount of fertilize and water. New effort will greatly cut down the growing cost while increasing farmers' profit margins when governing hundreds of acres of strawberry plants through optimizing farming operations of precision in-field object detection. The proposed superb intelligent algorithm helps farmers make timely decisions impacting the subsequent yields and quality of strawberries. In conclusion, this paper provides a new strategy for detecting outdoor strawberry plant canopies, which can effectively provide valuable information for the strawberry planting decision and management.

4 Conclusions

This paper constructed a novel strawberry plant detection model based on the original Mask R-CNN algorithm to address the problems of strawberry plant canopy detection in the fields while obtaining accurate and robust detection results. The specific work was summarized as follows:

In the preparation phase of the dataset, a drone was used to obtain low-altitude sensing images of strawberry farmland, and data augmentation was used to cover necessary areas to improve the generalization of the model in dealing with strawberry plant samples in different environments.

According to the characteristics of overlapping leaves of adjacent strawberry plant targets in the fields, custom solutions were adopted to improve the discrimination performance of the model. In the backbone, FPN was used to solve the multiscale change problem. In the training process, OHEM was used to mine difficult samples to learn the deep diverse characteristics of strawberry plants and update the weight of the model. In postprocessing, the Soft-NMS algorithm was used to solve the overlapping problem of samples.

The final experimental results showed that the improved Mask R-CNN strawberry plant detection model had a 33.2% increase in AP50, a 33.7% increase in AR, and a 93.1% increase in FPS compared to the original model. A boosted deep-level region-based artificial intelligence system for strawberry plant recognition could offer valuable information to growers for timely analysis of fruit growth conditions.

Ethics approval and consent to participate

Not applicable.

Competing interests

None of the authors have any competing interests in the manuscript.

Consent for publication

All the authors consent for publication.

Author contributions

Ping Lin and Huazhe Zhang wrote the main manuscript text and prepared figures and tables; Xiaoxuan Wang and Huan Liu built the experimental platform. Feiyu Zhao implemented the algorithms and validated the experimental results. Yongming Chen revised the manuscript. All authors reviewed the manuscript.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Grants No. 31601227).

References

Bodla, N., Singh, B., Chellappa, R., & Davis, L. S. (2017). *Soft-NMS--improving object detection with one line of code.* In *Proceedings of the IEEE International Conference on Computer Vision*. USA: IEEE.

Chen, T.-C., & Yu, S.-Y. (2022). The review of food safety inspection system based on artificial intelligence, image processing, and robotic. *Food Science and Technology (Campinas)*, 42, e29121. https://doi.org/10.1590/fst.35421.

Chu, Z., Li, F., Wang, D., Xu, S., Gao, C., & Bai, H. (2022). Research on identification method of tangerine peel year based on deep learning. *Food Science and Technology (Campinas)*, 42, e64722. http://dx.doi.org/10.1590/fst.64722.

Dubey, S. R., & Jalal, A. S. (2015). Fruit and vegetable recognition by fusing colour and texture features of the image using machine learning. *International Journal of Applied Pattern Recognition*, 2(2), 160-181.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). *Mask r-cnn.* In *Proceedings of the IEEE international conference on computer vision*. USA: IEEE.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition.* In *Proceedings of the IEEE conference on computer vision and pattern recognition*. USA: IEEE.

Li, Z., Dong, M., Wen, S., Hu, X., Zhou, P., & Zeng, Z. (2019). CLU-CNNs: Object detection for medical images. *Neurocomputing*, 350, 53-59. http://dx.doi.org/10.1016/j.neucom.2019.04.028.

Lin, P., Lee, W., Chen, Y., Peres, N., & Fraisse, C. (2020). A deep-level region-based visual representation architecture for detecting strawberry flowers in an outdoor field. *Precision Agriculture*, 21(2), 387-402. http://dx.doi.org/10.1007/s11119-019-09673-7.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). *Feature pyramid networks for object detection.* In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. USA: IEEE.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Lawrence Zitnick, C., & Dollar, P. (2014). Microsoft coco: Common objects in context. In *Proceedings of the European Conference on Computer Vision*. Cham: Springer.

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137-1149. http://dx.doi.org/10.1109/TPAMI.2016.2577031. PMid:27295650.

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: a database and web-based tool for image annotation. *International Journal of Computer Vision*, 77(1-3), 157-173. http://dx.doi.org/10.1007/s11263-007-0090-8.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). *Facenet: a unified embedding for face recognition and clustering.* In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. USA: IEEE.

Shin, J., Chang, Y. K., Heung, B., Nguyen-Quang, T., Price, G. W., & Al-Mallahi, A. (2021). A deep learning approach for RGB image-based powdery mildew disease detection on strawberry leaves. *Computers and Electronics in Agriculture*, 183, 106042. http://dx.doi.org/10.1016/j.compag.2021.106042.

Shrivastava, A., Gupta, A., & Girshick, R. (2016). *Training region-based object detectors with online hard example mining.* In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. USA: IEEE.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv*, arXiv:1409.1556, 1-14. https://doi.org/10.48550/arXiv.1409.1556.

Sun, J., He, X., Wu, M., Wu, X., Shen, J., & Lu, B. (2020). Detection of tomato organs based on convolutional neural network under the overlap and occlusion backgrounds. *Machine Vision and Applications*, 31(5), 1-13. http://dx.doi.org/10.1007/s00138-020-01081-6.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). *Going deeper with convolutions.* In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. USA: IEEE. http://dx.doi.org/10.1109/CVPR.2015.7298594.

Wang, M., & Deng, W. (2021). Deep face recognition: a survey. *Neurocomputing*, 429, 215-244. http://dx.doi.org/10.1016/j.neucom.2020.10.081.

Xu, Z.-F., Jia, R.-S., Sun, H.-M., Liu, Q.-M., & Cui, Z. (2020). Light-YOLOv3: fast method for detecting green mangoes in complex scenes using picking robots. *Applied Intelligence*, 50(12), 4670-4687. http://dx.doi.org/10.1007/s10489-020-01818-w.

Yu, Y., Zhang, K., Yang, L., & Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Computers and Electronics in Agriculture*, 163, 104846. http://dx.doi.org/10.1016/j.compag.2019.06.001.

Zhang, P., & Xu, F. (2021). Effect of AI deep learning techniques on possible complications and clinical nursing quality of patients with coronary heart disease. *Food Science and Technology (Campinas)*, 42, e42020. http://dx.doi.org/10.1590/fst.42020.

Zhang, X., Gao, H., Guo, M., Li, G., Liu, Y., & Li, D. (2016). A study on key technologies of unmanned driving. *CAAI Transactions on Intelligence Technology*, 1(1), 4-13. http://dx.doi.org/10.1016/j.trit.2016.03.003.