ICIC Express Letters
Part B: Applications
Volume **14**, Number **6**, June **2023**

ICIC International ©2023 ISSN 2185-2766
pp. **649–655**

STUDY ON TEA BUD RECOGNITION BASED ON YOLOV5S

Yu He, Yongmao Huang*, Yi Guo, Kai Zhang, Weilai Wang and Yongping Zhou

School of Electrical and Electronic Information
Xihua University
No. 999, Jinzhou Road, Jinniu District, Chengdu 610039, P. R. China
{ 599146322; 1647643261; jkjui }@qq.com; lpngy@vip.163.com; weilai_w98@163.com
*Corresponding author: ymhuang1988@126.com

Received October 2022; accepted January 2023

Abstract. *In view of the problem that it is difficult to identify tea buds due to the influence of light background and other factors in the natural environment, this paper presents an improved YOLOv5s target detection algorithm for the identification of tea buds. Firstly, we establish our own tea bud data set by taking photos and data enhancement, secondly, the data are preprocessed by median filtering to reduce the influence of background on recognition accuracy, and then the attention mechanism and small target detection layer are added to the network to improve the detection performance of small targets and occluded targets. Finally, the data set is manually labeled and imported into the network for training. The experimental results show that the improved model achieves 90.87% in the mAP@0.5 value of tea bud recognition, and it is 3.28% higher than the mAP@0.5 of the original YOLOv5s algorithm, and has a good effect on tea bud recognition in natural environment.*

Keywords: Tea bud, Natural environment, YOLOv5s, Target detection, Attention mechanism

1. **Introduction.** Tea originated in China. As one of the main cash crops in China, it plays an important role in the development of agricultural production. At present, tea buds are mainly obtained by manual picking and machine picking, and due to the difference in color and shape between young buds and old leaves, manual tea picking is very time-consuming and labor-intensive, which is not conducive to the improvement of tea production efficiency in China; machine tea picking solves the speed of tea picking and improves the efficiency of tea picking, but it cannot accurately identify the tender buds and old leaves of tea, cannot guarantee the quality of tea picking, and reduces the high-end sales value of tea. Therefore, an intelligent tea picking technology with selectivity, high efficiency, low consumption and low damage rate has become the general trend, and tea bud recognition is a key link in the research. In recent years, the deep learning method based on computer vision has made great innovation and breakthrough in the fields of image classification, target detection and so on. Convolutional neural network has achieved great success in the field of graphics processing and image recognition. Therefore, tea bud recognition based on deep learning has great research significance and practical value. The target detection algorithm is mainly divided into two ideas: the first is "two-step", which first recommends the region and then classifies the target. Typical representatives include R-CNN series (R-CNN [1], Fast R-CNN [2], and Faster R-CNN [3]), and Xu et al. [4] used fast R-CNN algorithm to detect tea buds, although its detection accuracy is not low, it is far from meeting the needs of real-time detection in terms of speed; The second is "one step", that is, based on the regression idea of deep learning, one network is used to achieve end-to-end in one step. Typical representatives are YOLO [5] and SSD [6], Wang et al. [7]

DOI: 10.24507/icicelb.14.06.649

first proposed using SSD algorithm to detect tea buds in the one-step target detection algorithm, and Xu et al. [8] completed the detection of tea buds through the optimization of YOLOv3. YOLOv5 in YOLO series has almost the fastest detection speed and better detection accuracy. There are many latest studies in various industries today, for example, Yu et al. [9] used improved YOLOv5 to detect the wearing of masks; Fang et al. [10] realized the recognition of off-line digital symbols by using YOLOv5. However, at present, the use of YOLOv5 algorithm to identify tea buds at home and abroad is still in the conceptual state, and there is no relevant research. In order to facilitate training, this paper improves the YOLOv5s algorithm with the smallest model in the YOLOv5 series, introduces the SE-Net attention module and adds a small-size detection layer. The improved model improves the mAP@0.5 of tea bud detection by 3.28% compared with the original YOLOv5s algorithm, and has a better effect on tea bud recognition in natural environment. This paper is divided into five chapters. The first chapter is the introduction, the second chapter is the research on the preprocessing technology of tea images, the third chapter is the improvement of YOLOv5s algorithm, the fourth chapter is the analysis and comparison of the results, and the fifth chapter is the conclusion.

2. **Data Set Production.**

2.1. **Image acquisition.** Because there is no public data set of tea bud images, the tea bud images of this experiment were taken with mobile phone rear camera in Meishan, Mingshan and other places in Sichuan Province in April 2020. A total of more than 2000 images were taken, with image pixels of 4160 × 3120. Then they were intercepted into 640 × 640 pixels, JPG format images. Figure 1 shows some of the intercepted images.

FIGURE 1. Data set sample

2.2. **Image preprocessing.** The main purpose of image preprocessing is to eliminate irrelevant information in the image, restore useful real information, enhance the detectability of relevant information and simplify the data to the greatest extent. Firstly use 3×3 size median filter for filtering and noise reduction of tea original image. Filtering and noise reduction can not only ensure the integrity of the original image information of tea as much as possible, but also preliminarily eliminate the useless noise information in the image, which is convenient for subsequent image processing. Compared with traditional machine learning methods, the parameters of deep neural network are very large, and there are many nonlinear operations. If there are not enough data samples for training, the neural network will be over fitted and cannot get good generalization ability. The tea sample images taken in this experiment cannot meet the sample needs of neural network training and learning, so it is necessary to enhance the original tea sample images secondly. The commonly used data enhancement methods mainly include mirror operation, rotation, scaling, clipping, translation and adding noise. In this experiment, the collected data of more than 2000 tea samples were artificially selected and adjusted, and 1000 sample data were selected to make the number of samples roughly the same, and then the tea sample data were horizontally mirrored to expand the tea sample data to 4000. Finally, the labelimg image annotation tool is used to annotate the tea bud image data to obtain the category and position of the bud target in the image.

3. Improvement of YOLOv5s Algorithm.

3.1. **Attention mechanism.** Attention mechanism comes from the way the human brain processes visual information. By rapidly observing the global information of the image, human beings find out the candidate area that needs to be focused, that is, the location of the focus, and will focus on this area to extract more detailed information of the target. Because of its powerful and effective forms, it has been widely used in deep learning, especially in deep-seated high-performance networks.

The attention module of SE-Net [11] channel can optimize and learn the characteristic information of specific categories in the deep-seated network. The overall structure of SE-Net is shown in Figure 2. $\mathbf{F}_{sq}$ refers to squeeze operation, $\mathbf{F}_{ex}$ refers to excitation operation, and $\mathbf{F}_{scale}$ refers to scale operation. Firstly, the squeeze operation is performed, and the feature vectors of three channel dimensions output by the original YOLOv5s network detection layer are compressed by global pooling (256, 512, 1024), obtain the global information between the features of each channel, and change the feature graph $\mathbf{U}(H \times W \times C)$ into a scalar of $1 \times 1 \times C$. Then, the two full connection layers establish a correlation model between channels, carry out nonlinear transformation between channel features, and output the weight information of channel C. The ReLU activation function is added between the two fully connected layers to increase the nonlinearity between channels. Then the sigmoid function is used for weight normalization. Finally, the features between channels are weighed by scaling operation, and the weights between channels are multiplied by the features of the original feature map to obtain a new channel weight. The

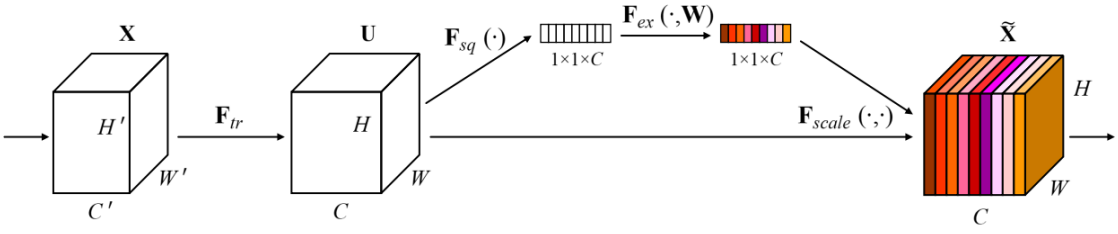

FIGURE 2. The structure of SE-Net

weight proportion between small target channels is increased, so as to guide the model to pay more attention to the relevant feature information of small targets, strengthen the training of these features, and further improve the detection performance of the model for small targets.

In order to use the pre training weight after adding the attention mechanism, this paper only introduces the SE attention module in the last layer of the backbone of YOLOv5s, and the structure is shown in Figure 3.

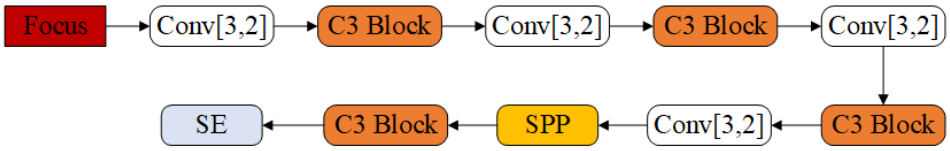

FIGURE 3. YOLOv5s backbone structure with SE-Net

3.2. **Adding small size detection layer.** YOLOv5s detects on three scales, which are accurately given by sampling the input image size by 32 times, 16 times and 8 times, respectively. In this study, the tea buds in the same picture are far and near. However, some tea buds with long distance have small targets and are densely distributed in the images taken from a long distance. The generalization ability of the small-scale detection layer of YOLOv5s for these tea buds is poor. Therefore, we add a new small-scale detection layer, which is obtained by down sampling the input image size four times. The small-scale detection layer generates the feature map by extracting the low-level spatial features and fusing them with the deep semantic features. The new small-scale detection layer makes the target detection network structure have better ability to learn multi-scale targets, as shown in Figure 4. It is suitable for detecting small and dense tea bud targets in the image.

4. Analysis of Experimental Results.

4.1. **Experimental setup.** The operating system of this experiment is Windows 10, the CPU is Intel (R) Core (TM) i5-10300H, the GPU is GeForce GTX 1650Ti with 6GB video memory, and the framework is Pytorch. The data set is randomly divided into training set, verification set and test set according to the ratio of $8:1:1$. We set the initial learning rate and the number of training rounds for the input image, and use SGD (Stochastic Gradient Descent) method to optimize the learning rate in the training process. See Table 1 for the specific settings of network training super parameters.

TABLE 1. Parameter setting

Parameter	Value
Input image size	640×640
Initial learning rate	0.01
Bacth size	8
Epoch	300

4.2. **Network performance evaluation.** Expressing the prediction box as tea shoots and background can generate three potential predictions: true positive (TP), false positive (FP) and false negative (FN). If the IoU of the detection box and label box is greater than 0.5, the detection box is marked as TP; otherwise it is marked as FP. If the detection box has no matching dimension box, it is marked as FN. TP and FP are the number of tea

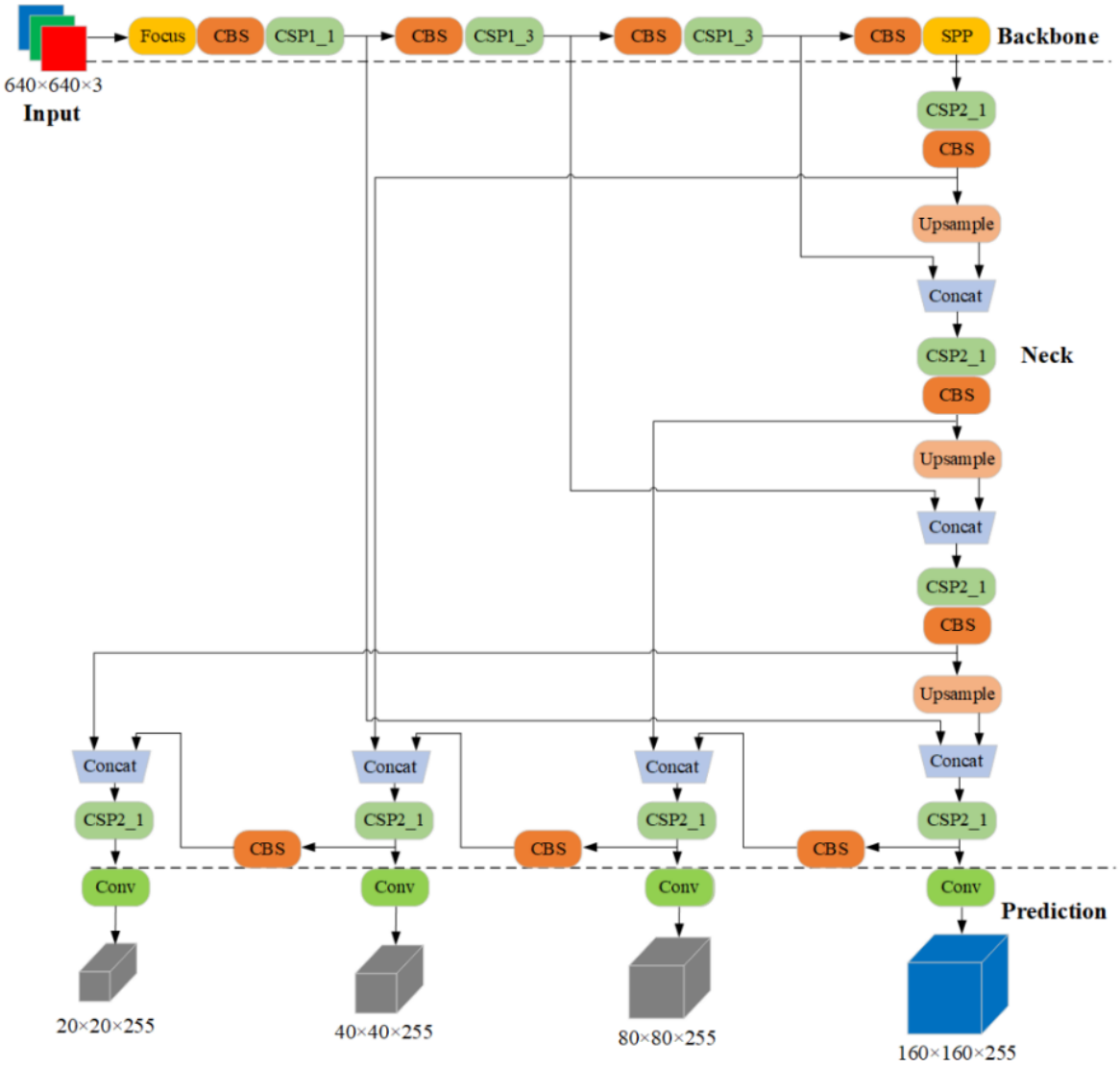

FIGURE 4. YOLOv5s backbone structure with SE-Net

shoots correctly and wrongly detected, respectively. FN represents the number of non tea shoots wrongly detected. Accuracy (P) and recall (R) are defined as

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

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

P and R affect each other and cannot be directly used to evaluate the detection accuracy. Therefore, we introduce the average accuracy (AP) to represent the detection accuracy. AP refers to the average accuracy of tea bud detection. Map refers to the average value of AP, which is to calculate the average AP value of multiple categories. Because the identification of tea buds is two categories, only the buds and background need to be distinguished during detection and identification, so here AP = map. Therefore, higher AP means higher detection accuracy.

$$AP = \int_0^1 P(r)dr \tag{3}$$

4.3. **Result analysis.**

4.3.1. *Performance comparison.* The performance comparison of each part of the model before and after improvement is shown in Table 2. It can be seen that each improvement has a certain improvement compared with the original model.

TABLE 2. Performance comparison before and after model improvement

Model	P/%	R/%	mAP@0.5/%
YOLOv5s	87.53	88.76	87.59
YOLOv5s+SE-Net	89.39	90.10	89.56
YOLOv5s+Adding layer	89.23	90.15	89.44
YOLOv5s+SE-Net+Adding layer	90.23	91.86	90.87

4.3.2. *Comparison of target detection algorithms.* In order to further verify the feasibility of this algorithm, it is compared with mainstream target detection algorithms such as SSD, Faster R-CNN, YOLOv3, YOLOv4 and YOLOX under the same data set and experimental settings. The experimental results are shown in Table 3. It can be seen from Table 3 that for the identification of tea buds, the average accuracy rate of the improved algorithm based on YOLOv5s proposed in this paper is better than other algorithms. Although the detection speed is lower than SSD algorithm, the accuracy rate of SSD algorithm lags too much, so the reference value of comparison is small. On the whole, the algorithm proposed in this paper has achieved good results in comprehensive performance.

TABLE 3. Comparison results of target detection algorithms

Model	mAP@0.5/%	FPS
Faster R-CNN	84.11	5.29
SSD	79.91	32.25
YOLOv3	87.81	13.33
YOLOv5s	87.59	23.47
YOLOX_s	89.12	11.66
Ours	90.87	20.89

4.3.3. *Detection comparison.* The comparison of tea bud detection results is shown in Figure 5. On the left is the original YOLOv5s detection diagram, and on the right is the improved model detection diagram. It can be seen that the missed detection of the original model is more serious. After the improvement, the missed detection rate is reduced, the detection effect of small targets is better, the generalization ability is better in dense scenes.

5. **Conclusion.** In order to solve the problems existing in the current traditional tea bud picking methods, an improved tea bud recognition algorithm based on YOLOv5s is proposed in this paper. Firstly, use 3×3 size median filter for filtering and noise reduction of tea original image to reduce the influence of background on the detection effect, and then adding SE-Net attention module and small target detection layer improves the detection effect of the model for small targets and the generalization ability in dense scenes. Through comparative experiments, it can be seen that the improved model is better than the original YOLOv5s model in accuracy, recall and average accuracy, and has a good detection effect on small targets and occluded targets in dense scenes, which shows that the method proposed in this paper is feasible for tea bud detection. In the next work, we will improve the network structure and loss function of YOLOv5 and study its impact on the detection effect of tea buds.

FIGURE 5. Comparison of detection effect

Acknowledgment. This research was supported by Grant SCITLAB-1021 of Intelligent Terminal Key Laboratory of Sichuan Province, the National Natural Science Foundation of China under Grant (61973257, 61901394), Central Government Funds of Guiding Local Scientific and Technological Development for Sichuan Province (2021ZYD0034).

REFERENCES

[1] R. Girshick, J. Donahue, T. Darrell et al., Rich feature hierarchies for accurate object detection and semantic segmentation, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.580-587, 2014.

[2] R. Girshick, Fast R-CNN, *Proc. of the IEEE International Conference on Computer Vision*, pp.1440-1448, 2015.

[3] S. Ren, K. He, R. Girshick et al., Faster R-CNN: Towards real-time object detection with region proposal networks, *Advances in Neural Information Processing Systems*, vol.28, 2015.

[4] G. Xu, Y. Zhang and X. Lai, Recognition approaches of tea bud image based on Faster R-CNN depth netword, *Journal of Optoelectronics · Laser*, vol.31, no.11, pp.1131-1139, 2020.

[5] J. Redmon, S. Divvala, R. Girshick et al., You only look once: Unified, real-time object detection, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.779-788, 2016.

[6] W. Liu, D. Anguelov, D. Erhan et al., SSD: Single shot MultiBox detector, in *Computer Vision – ECCV 2016. Lecture Notes in Computer Science*, B. Leibe, J. Matas, N. Sebe and M. Welling (eds.), Cham, Springer, 2016.

[7] Z. Wang, Y. Zhao and Z. Liu, Research on tea buds detection based on SSD algorithmon, *Microprocessors*, vol.41, no.4, pp.42-48, 2020.

[8] W. Xu, L. Zhao, J. Li, S. Shang, X. Ding and T. Wang, Detection and classification of tea buds based on deep learning, *Computers and Electronics in Agriculture*, vol.192, DOI: 10.1016/j.compag.2021.106547, 2022.

[9] S. Yu, H. Li, F. Gui et al., Research on real-time mask-wearing detection algorithm based on YOLOv5 in complex scenes, *Computer Measurement & Control*, vol.29, no.12, pp.188-194, 2021.

[10] H. Fang, G. Wan, Z. Chen, Y. Huang, W. Zhang and B. Xie, Offline handwritten mathematical symbol recognition based on improved YOLOv5s, *Journal of Graphics*, vol.43, no.3, pp.387-395, 2022.

[11] J. Hu, L. Shen and G. Sun, Squeeze-and-excitation networks, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.7132-7141, 2018.