A Deep Learning-Based In-field Fruit Counting Method Using Video Sequences

JiaqiWang[1][0000−0002−1708−3573], WenliZhang[1][0000−0003−3151−5755], KaizhenChen[1][0000−0001−6871−4091], HuibinLi[2][0000−0002−4901−2104], YunShi[2][0000−0002−6294−0124], and WeiGuo[3][0000−0002−3017−5464]

[1] Beijing University of Technology zhangwenli@bjut.edu.cn
[2] Chinese Academy of Agricultural Sciences shiyun@caas.cn
[3] The University of Tokyo guowei@g.ecc.u-tokyo.ac.jp

1 Introduction

In recent years, computer vision-based fruit counting in orchards has become a hot research topic in smart agriculture. Modern farms started to getting benefits on fruit yield estimation and precision marketing strategy decisions from such technology. There are mainly two tasks for developing such techniques: precision fruit detection and counting from orchard images..

For fruit detection task, researchers have proposed deep learning-based image detection algorithms for fruit detection [1–4]. But they did not address the simultaneous presence of small-scale targets. For fruit localization and counting, researchers have proposed methods based on static images and video sequences[1, 3, 5–7]. The video-based counting method collects fruit images from multiple viewpoints and is considered as an efficient solution for fruit counting. However, the current video-based methods do not discuss the complex occlusion situations that may exist in global video sequences, which result in the loss of tracking targets.

Therefore, using orange as a study case, we propose the following solutions to the above two tasks: 1) We proposed an improved Yolov3 [8] detection model based on the principle of matching the feature map's receptive field to the target scale [9]. 2) We first analyze the complex occlusion of orange fruits and define the counting region at each global video sequence frame. Then, using the multi-objective tracking algorithm Sort [10] to count the fruits that only appear in the pre-defined region.

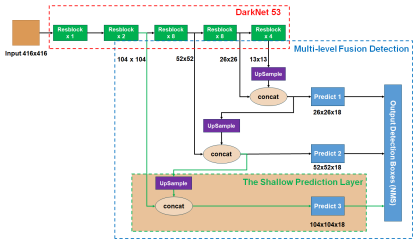

Fig. 1. The Improved-Yolov3 Network Structure

2 Method

In this study, the video sequence was captured by the DJI Osmo Action camera (DJI Technology Co., Ltd., ShenZhen, China) in an orange orchard in Sichuan Province, China. The proposed fruit detection and counting method based on video include two steps: fruit detection and fruit tracking counting.

Table 1. Fruit Detection Performance

Method	Precision	Recall	F1-score	AP	FPPI
Yolov3	0.926	0.90	0.911	0.960	2.294
improved-Yolov3	0.926	0.926	0.926	0.968	2.35

Table 2. Fruit Counting Performance

Counting Method	Number of fruit counts	Inference time
manual counting	90	30s
improved-Yolov3(No Track)	900	0.02s
improved-Yolov3+Sort(proposed)	102	0.08s

Fig. 2. Visualization of Fruit Detection

Step 1. Fruit detection method based on improved-Yolov3: Firstly, we calculate the size of the receptive field [11] of the Yolov3 network, and cluster the orange dataset to count the orange scale distribution. Secondly, we design the shallow prediction layer for detecting orange based on the principle of matching the feature map receptive field to the target scale. Then using a multi-level fusion strategy to fuse the shallow layer feature with the deep layer feature to enhance the semantic features of the shallow feature map. Finally, the fusion features are used to detect small-scale oranges in each image frame. The improved-Yolov3 network structure is shown in Figure 1, where the yellow region indicates the shallow prediction layer.

Step 2. Fruit tracking counting method based on specified area: Firstly, the orange detection results from step 1 are input to the tracking algorithm Sort, and determine whether these oranges are in the specified count area. If the fruit is in the count area, it will be assigned a unique number and tracked frame by frame until it leaves the count area. Finally, the number of orange ordinal numbers is counted as the final orange counting results.

3 Results and Discussion

In this study, we used 330 orange images and divided them into the train set and test set at the ratio of 8:2. Table 1 shows the comparison results between the improved-Yolov3 and the original Yolov3 for the five metrics of Precision, Recall, F1-score, FPPI, and AP. Figure 2 shows the detection results of the improved-Yolov3, where the red boxes correspond to ground truth and the blue boxes correspond to detection results. The orange counting results shown in Table 2, where the proposed improved-Yolov3 with tracking algorithms count 102 oranges at a speed of 0.08s per frame, is close to the manual count result.

References

1. A Koirala, KB Walsh, Z Wang, and C McCarthy. Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'mangoyolo'. *Precision Agriculture*, 20(6):1107–1135, 2019.
2. Orly Enrique Apolo Apolo, Jorge Martínez Guanter, Gregorio Egea Cegarra, Purushothaman Raja, and Manuel Pérez Ruiz. Deep learning techniques for estimation of the yield and size of citrus fruits using a uav. *European journal of agronomy: the official journal of the European Society for Agronomy*, 115(4):183–194, 2020.
3. Ramesh Kestur, Avadesh Meduri, and Omkar Narasipura. Mangonet: A deep semantic segmentation architecture for a method to detect and count mangoes in an open orchard. *Engineering Applications of Artificial Intelligence*, 77:59–69, 2019.
4. Nicolai Häni, Pravakar Roy, and Volkan Isler. A comparative study of fruit detection and counting methods for yield mapping in apple orchards. *Journal of Field Robotics*, 37(2):263–282, 2020.
5. Zhenglin Wang, Kerry Walsh, and Anand Koirala. Mango fruit load estimation using a video based mangoyolo—kalman filter—hungarian algorithm method. *Sensors*, 19(12):2742, 2019.
6. Xu Liu, Steven W Chen, Shreyas Aditya, Nivedha Sivakumar, Sandeep Dcunha, Chao Qu, Camillo J Taylor, Jnaneshwar Das, and Vijay Kumar. Robust fruit counting: Combining deep learning, tracking, and structure from motion. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1045–1052. IEEE, 2018.
7. Xu Liu, Steven W Chen, Chenhao Liu, Shreyas S Shivakumar, Jnaneshwar Das, Camillo J Taylor, James Underwood, and Vijay Kumar. Monocular camera based fruit counting and mapping with semantic data association. *IEEE Robotics and Automation Letters*, 4(3):2296–2303, 2019.
8. Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
9. Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel. Understanding the effective receptive field in deep convolutional neural networks. In *Advances in neural information processing systems*, pages 4898–4906, 2016.
10. Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple online and realtime tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3464–3468. IEEE, 2016.
11. Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*, 2016.