The 12th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications
7-9 September, 2023, Dortmund, Germany

Apple Detection With Occlusions Using Modified YOLOv5-v1

Oleksandr Melnychenko, Oleg Savenko, Pavlo Radiuk
Khmelnytskyi National University, 11, Instytuts'ka str., Khmelnytskyi, 29016, Ukraine,
oleksandr.melnychenko@live.com, savenko_oleg_st@ukr.net, radiukp@khmnu.edu.ua

***Abstract*—In our research, we created a novel YOLOv5-v1 architecture to identify apples in images with occlusions. We specifically engineered new layers for the BottleneckCSP-v4 module, which replaces the original BottleneckCSP module within the backbone structure of the YOLOv5 network. Integrating the SENet module into our improved trunk network helps to discern features of medium and large-sized fruits more accurately under varying conditions. We also adjusted the initial size of the binding block within the source network to avoid incorrect identification of small objects within the image's background. Based on the test dataset, our experimental results show that our advanced network model can effectively identify fruits captured through an unmanned aerial vehicle camera. The classification metrics - recall, precision, mAP, and F1-score - obtained scores of 92.13%, 84.59%, 87.94%, and 89.02% respectively.**

Keywords—image processing; object detection; apple yield; YOLOv5; deep learning; visual occlusions

I. INTRODUCTION

Apples are one of the most extensively cultivated fruit crops worldwide, and Ukraine boasts the most prominent apple plantation area. It spans around two million hectares, making Ukraine the leading apple producer in Europe. The central and southern regions of Ukraine offer perfect climatic and soil conditions for apple farming [1]. Additionally, advancements in farming practices within the country have increased apple yields and enhanced fruit quality.

Employing visual recognition for apple yield detection is a professional and intuitive approach. However, due to variations in the growth patterns and fruit counts of each tree, individual yield detection is required to ensure high accuracy [2]. In large orchards, efficiency is paramount, making it essential to have a rapid, accurate, and compact detection model for apple yields. Such a model facilitates quicker apple yield detection and can function on various embedded devices.

Historically, standard vision methods like image processing and machine learning were typically employed to detect fruits. These techniques identified fruits based on attributes such as color, shape, and texture. For instance, Yu et al. [3] trained a model to recognize litchi fruit using color and texture features, achieving an accuracy rate of 89.92% for green litchi and 94.50% for red litchi. Similarly, Syazwani et al. [4] applied machine learning to detect pineapple crown images, attaining an accuracy of 94.4% for fruit counting. Fu et al. [5] amalgamated texture features with the support vector machine algorithm for banana detection, which yielded an average detection rate of 89.63%. However, these image-processing-based methods carry some drawbacks, including slow detection speed, lower accuracy, and limited adaptability to changing lighting conditions in orchards.

Deep learning (DL) models, particularly convolutional neural networks (CNNs), can extract features from images [6], enabling automatic target recognition and superior adaptability. Notably, two types of DL target detection techniques [7] could benefit apple detection: two-stage and single-stage target detection algorithms.

The two-stage target detection algorithm, incorporating R-CNN, SPP, and Fast R-CNN, is made up of two network branches: region proposal generation (RPN) [8] and the classification module [9]. The RPN proposes a region of interest (ROI) for the foreground class, and the classification module assigns and evaluates the bounding box for each ROI [10]. However, this method is hardly suitable for embedded devices due to its slow detection speed due to the large model size. For example, Mai et al. [11] presented a Faster R-CNN framework for multi-category fruit detection that achieved an average precision of 90.72% but had a detection time of 58 milliseconds (ms) per image.

In contrast to the two-stage approach, the single-stage target detection method, encompassing the You Only Look Once (YOLO) [12] series and the Single-Shot Detector (SSD) [13] model, maintains a more balanced performance regarding detection speed and accuracy. This balance contributes to an overall enhancement in the model's effectiveness. For instance, Huang et al. [14] proposed an enhanced YOLOv3 detection method for identifying immature apples within orchard scenes, obtaining an accuracy of 61.6%. Chen et al. [15] reached a 97.13% accuracy rate in detecting apples in intricate orchards using YOLOv4, with an average recognition time of 16.69 ms per individual image on a single GPU device.

Dwarf and spindle-shaped trees typically have a less dense leaf arrangement, making fewer apples hidden from sight. However, some apples might still be obscured by branches or leaves [16], and the lighting in the background can be uneven or complex [17]. To achieve more consistent

979-8-3503-5805-6/23/$31.00 ©2023 IEEE