Chapter 16
Detection and Classification of Dense Tomato Fruits by Integrating Coordinate Attention Mechanism With YOLO Model

Seetharam Nagesh Appe
Annamalai University, India

G. Arulselvi
Annamalai University, India

Balaji G. N.
Vellore Institute of Technology, India

ABSTRACT

Real-time detection of objects is one of the important tasks of computer vision applications such as agriculture, surveillance, self-driving cars, etc. The fruit target detection rate based on traditional approaches is low due to the complex background, substantial texture interference, partial occlusion of fruits, etc. This chapter proposes an improved YOLOv5 model to detect and classify the dense tomatoes by adding the coordinate attention mechanism and bidirectional pyramid network. The coordinate attention mechanism is used to detect and classify the dense tomatoes, and bidirectional pyramid network is used to detect the tomatoes at different scales. The proposed model produces good results in detecting the small dense tomatoes with an accuracy of 87.4%.

DOI: 10.4018/978-1-6684-8098-4.ch016

INTRODUCTION

The agricultural robot has emerged as one of the most promising technologies of precision agriculture technologies among many others (Tao et al., 2017). Seeding, weeding, feeding, trimming, picking and harvesting, sorting, and other repetitive tasks make up most of agriculture. Farmers may now focus more on strategic issues and increase output yields thanks to agricultural robots, which automate those slow, boring, and repetitive tasks (Edan et al., 2009). The automated harvesting and picking of the fruits are one of the important application of the robots. In principle, a sophisticated detection algorithm is required for the robot to be fully capable of doing harvesting and picking in order to overcome obstacles including naturally occurring variations in illumination, shape, position and color (Barnea et al., 2016).

Object detection is a technique used to recognize and detect different objects that are present in video or image and label them in order to classify these objects. Several methods are commonly used for object detection in order to recognise and locate items, and these algorithms make use of deep learning to produce accurate results. Object detection primarily used in instance segmentation, semantic segmentation, image caption etc., Object detection can be divided into machine learning-based and deep learning-based approaches. The machine learning approach requires the features (such as the color histogram or edges) to be defined by using various methods. The position and label of the object are predicted by a regression model using these features as input. Contrarily, deep learning-based techniques accomplish end-to-end, unsupervised object detection using convolutional neural networks (CNNs), eliminating the need for features(characteristics) to be specified and extracted independently.

Encoder and decoder are typically the two components of deep learning-based object detection algorithms. When an encoder receives an image as input, it processes it through a series of layers and blocks that enable it to extract statistical features that can be used to identify and locate objects. After the encoder, a decoder receives the output and predicts the labels and bounding boxes of each object. Typically, these object detectors can be divided into two-stage and one-stage detectors. RCNN (Girshick et al., 2014), Fast R-CNN (Girshick et al., 2015), Faster R-CNN (Ren et al. 2015), and Mask R-CNN (He et al., 2017) are two-stage detectors, while SSD (Liu et al., 2016) and YOLO (Redmon et al., 2016) series are one-stage approaches.

Even though these general object detection techniques have improved in accuracy and efficiency, the limited resolution and context information are inadequate to train a model, detecting small objects in images is still challenging.

Based on Yolov4-tiny's deep learning object detection algorithm, The attention mechanism and multi-scale prediction were combined by Lu et al. (2016) to identify occlusion and small item green pepper.

Sa et al. experimented with a faster R-CNN detector for fruit detection by combining multi-modal colour (RGB) and near-infrared (NIR) information. Even though only a small number of images were utilised for training and testing, this method outperformed earlier ones in terms of performance. However, it is challenging for the method to detect small and occluded fruits, and its speed still needs to be improved for harvesting robots to operate in the field in real-time.

For the purpose of detecting tomatoes, Zhao et al. (2016) developed the AdaBoost algorithm using the average pixel value. Based on physical characteristics like texture, color, shape and Haar-like characteristics of images, the algorithm identifies the ripe tomatoes. A tiny-yolo network based on the YOLOv2 network was proposed by Xue et al. (2018). The proposed tiny-yolo model consists of dense blocks used for detection of immature mango and mango under orchard scenes.