Computers and Electronics in Agriculture 182 (2021) 106052

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Improved multi-classes kiwifruit detection in orchard to avoid collisions during robotic picking

Rui Suo [a], Fangfang Gao [a], Zhongxian Zhou [a], Longsheng Fu [a,b,c,*], Zhenzhen Song [a], Jaspreet Dhupia [d], Rui Li [a], Yongjie Cui [a]

[a] College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
[b] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China
[c] Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling, Shaanxi 712100, China
[d] Department of Mechanical Engineering, The University of Auckland, Private Bag 92019, Auckland, New Zealand

ARTICLE INFO

Keywords:
Agricultural robotics
Deep learning
YOLO
Branch occluded
Wire occluded

ABSTRACT

Deep learning has achieved kiwifruit detection with high accuracy and fast speed. However, all the kiwifruits have been labeled and detected as only one class in most researches for robotic fruit picking, where fruits occluded by branches or wires have been detected as pickable targets. End-effectors or robots may be damaged by the branches or wires when they are forced to pick those fruits. Therefore, kiwifruits are labeled, trained, and detected in multi-classes based on their occlusions to avoid detecting fruits occluded by branches or wires as pickable targets. Fruits are classified into four classes and five classes according to robotic picking strategy and field occlusions, respectively. Well-known YOLOv3 and recently released YOLOv4 are employed to do transfer learning for multi-classes kiwifruit detection. Results show that mAP (mean average precision) of fruits in the five-classes is higher than that in the four-classes, while mAP of YOLOv4 is higher than YOLOv3. The mAP of YOLOv4 and YOLOv3 in the five-classes and four-classes are 91.9%, 91.5%, 91.1%, and 89.5%, respectively. The results demonstrate that fruits labeled and trained in more classes can achieve higher mAP. There are significant differences in average detection speed in YOLOv3 and YOLOv4, but no in the four-classes and five-classes. Overall, the highest mAP of 91.9% was achieved by YOLOv4 in the five-classes, which cost 25.5 ms on average to process a 2352 × 1568 image. The results illustrate that multi-classes kiwifruit detection is helpful for avoiding damage to the end-effectors or robots.

1. Introduction

Kiwifruit has high nutritional value and is widely planted in China. Kiwifruit is considered as a high nutritional product due to rich in vitamin C and some other nutrients (Leontowicz et al., 2016; Fazayeli et al., 2019). It is described as a tangy, sweet, and sour combination, and thus becoming very popular among consumers (Leontowicz et al., 2013; Richardson et al., 2018). In 2018, the cultivation area of kiwifruits in China exceeded 1.68×10^5 ha with a production of 2.04 million tons, accounting for around 49% of the world production (UN Food & Agriculture Organization, 2020).

Robotic picking methods are studied as an alternative to high labor cost manual picking. Most horticulture industries, including kiwifruit, are requiring a lot of laborers, especially during harvest seasons (Yang et al., 2014; Song et al., 2021). This seasonal labor demand is creating a huge risk for growers not having sufficient labor to pick fruits (Feng et al., 2019). In the harvest season, labor costs for kiwifruit harvesting account for over 25% of annual production costs (García-Quiroga et al., 2015). Further, manual picking activities pose a high risk of back strain and musculoskeletal problems to laborers due to repetitive hand motions (Fathallah, 2010). Kiwifruit picking robot is fundamentally essential to ease labor-intense demand and lower human risks of injuries in orchards (Mu et al., 2020; Williams et al., 2020). A typical fruit harvesting robot includes two main subsystems: a vision system and an end-effector system (Zhang et al., 2020). The vision system detects and localizes fruits, which can guide the end-effectors to detach fruits from trees (Kang and Chen, 2020).

Fruit detection is one of the most steps to achieve robotic picking and has been studied by many researchers. Traditional image processing approaches are extensively studied and obtained desirable results (Fu

* Corresponding author.
E-mail address: fulsh@nwafu.ed.cn (L. Fu).

https://doi.org/10.1016/j.compag.2021.106052
Received 15 December 2020; Received in revised form 6 February 2021; Accepted 10 February 2021
Available online 21 February 2021

et al., 2019, 2015; Zhan et al., 2013). But they are easily affected by complex backgrounds in orchard (Mu et al., 2019; Williams et al., 2019). Compared to the traditional image processing approaches, deep learning has strong adaptability to differences within a working scene (Kamilaris and Prenafeta-Boldú, 2018; Li et al., 2020; Majeed et al., 2020; Zhou et al., 2019). Fu et al. (2018) used Faster R-CNN (Faster Region-Convolutional Neural Network) with ZFNet (Zeller Fergus Net) to detect kiwifruits in orchard, which reached a detection rate of 92.3% and needed 274 ms to process an image. Mu et al. (2019) obtained an AP (average precision) of 96.0% and a detection speed of 1000 ms per image using Faster R-CNN with AlexNet for kiwifruit detection with multiple clusters characteristics under far-view and occlusion conditions. Zhou et al. (2020) employed SSD (Single Shot Multi-Box Detector) with MobileNetV2 to detect kiwifruits on android smartphones, which achieved a detection rate (the number of fruits detected correctly divided by the number of all fruits) of 89.7% and a detection speed of 103 ms per image. Liu et al. (2020) applied Faster R-CNN with VGG16 (Visual Geometry Group with 16 layers) to perform kiwifruit detection, which reported an AP of 90.7% and a detection speed of 134 ms per image. Deep learning has achieved kiwifruit detection with high accuracy and fast speed. However, all the kiwifruits have been labeled and detected as only one class in most researches for robotic fruit picking, where fruits occluded by branches or wires have been detected as pickable targets.

End-effectors or kiwifruit picking robots may be damaged when they are forcibly picking kiwifruits occluded by branches or wires. As shown in Fig. 1, an example of kiwifruits were all detected as only one class (Fu et al., 2020a), where red boxes referred to fruits were detected by deep learning methods. The fruit in the yellow rectangle manually drawn was occluded by branch and fruit in the green rectangle manually drawn was occluded by wire. But they were not distinguished from the fruits not occluded and detected as pickable targets. Therefore, a method to label, train, and detect kiwifruits as multi-classes was proposed to avoid detecting fruits occluded by branches or wires as pickable targets. Gao et al. (2020) proposed a multi-class apple detection method based on Faster R-CNN with VGG16, where apples were labelled and detected into four classes: non-occluded, leaf-occluded, branch/wire-occluded, and fruit-occluded fruit. It obtained APs of non-occluded, leaf-occluded, branch/wire-occluded, and fruit-occluded fruit of 90.9%, 89.9%, 85.8%, and 84.8%, respectively. Consider the robotic picking strategy, fruits occluded by branches or wires were treated as one class. However, features of branch and wire are entirely different, where more specific features may be learned from branches and wires, respectively. Therefore, fruits classified into more classes may be beneficial to improving detection accuracy.

YOLO (You Only Look Once) has fast detection speed with high accuracy. Object detectors based on deep learning are usually categorized into two kinds, i.e., one-stage and two-stage. One-stage object detectors, including YOLO, usually have a faster detection speed than two-stage object detectors (Ju et al., 2019). Xu and Wu (2020) employed Faster R-CNN with VGG16 and YOLOv3 on aerial images dataset detection, which reported that detection speed of YOLOv3 was four times faster than Faster R-CNN. Fu et al. (2020a) used YOLOv3-tiny to detect kiwifruits in orchard, which showed that detection speed of Faster R-CNN with ZFNet and Faster R-CNN with VGG16 were about nine and eleven times slower than YOLOv3-tiny, respectively. Compared to other one-stage object detector like SSD, YOLO has a fast update speed and been employed by more researchers. YOLO series networks have been constantly updated and continuously improved on detection performances (Bochkovskiy et al., 2020). YOLOv3 has been well-known due to fast detection speed and high accuracy (Redmon and Farhadi, 2018). YOLOv4 was released recently and was described as higher precision (Bochkovskiy et al., 2020). Therefore, YOLOv3 and YOLOv4 were employed to do transfer learning for kiwifruit detection in orchard.

In this study, kiwifruit is labeled, trained, and detected in multi-classes by deep learning methods to avoid detecting fruits occluded by branches or wires as pickable targets. Kiwifruits are classified into four classes and five classes according to robotic picking strategy and field occlusions, respectively. Well-known YOLOv3 and recently released YOLOv4 are employed to do transfer learning for the multi-classes kiwifruit detection and compared by mAP and processing speed. Optimal classification method and YOLO model are proposed by a completed discussion and investigation on experimental results.

2. Materials and methods

2.1. Image acquisition

In this study, RGB (Red, Green, and Blue) images of 'Hayward' kiwifruit were captured during three harvest seasons of 2017, 2018, and 2019 from Meixian Kiwifruit Experimental Station of Northwest A&F University, Shaanxi, China. Kiwifruit images were acquired with a single-lens reflex camera (Canon Kiss X3, Canon Inc., Tokyo, Japan) with a resolution of 2352 × 1568, and saved in JPG format. The camera

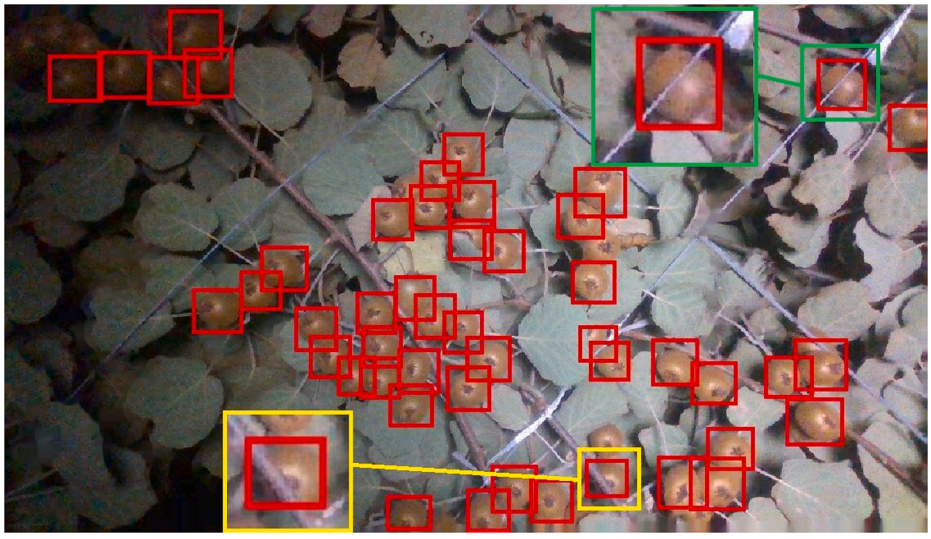

Fig. 1. Example of kiwifruit detected as only one class (Fu et al., 2020a). Red boxes referred to fruits were detected by deep learning methods; fruit in the yellow rectangle manually drawn was occluded by branch; fruit in the green rectangle manually drawn was occluded by wire. The fruits occluded by branch or wire were not distinguished from the others. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

was placed below the kiwifruit canopy approximately 100 cm to acquire images. And central axis of the camera was perpendicular to the kiwifruit canopy.

A total of 1160 original images were acquired at different times (morning, afternoon, and night) of a day with different illumination. All the images were acquired at 580 different positions, which ensured the images have no overlapping regions. Two images in two different illuminations (with or without flash) were acquired in the morning or afternoon, or two images with LED (light emitting diode, CM-LED 1200 HS, KEMA Co., Wuhan, China) or flash at night. The two images acquired at the same position were called a pair. In total, 580 pairs of images (with 185 taken in the morning, 280 in the afternoon, and 115 at night) were collected, as shown in Fig. 2.

2.2. Image dataset

According to the robotic picking strategy and field occlusions, kiwifruits were classified into four classes and five classes, respectively. In orchard, kiwifruits may be occluded by leaves, fruits, branches, or wires, whose features such as shapes and colors are completely different. Therefore, kiwifruits were divided into five classes due to the difference in actual occlusion. The first class indicated that fruit is not occluded (referred to NO in this study), as shown in Fig. 3a. The second class indicated that fruit is occluded by leaves (referred to OL) and not occluded by other fruit, branch, or wire (Fig. 3b). The third class indicated that fruit is occluded by other fruits (referred to OF) and not occluded by branch or wire, whether it is occluded by leaves (Fig. 3c). The fourth class indicated that fruit is occluded by branches (referred to OB), whether it is occluded by leaves, other fruits, or wires, as shown in Fig. 3d. The last class indicated that fruit is occluded by wires (referred to OW), whether it is occluded by leaves, other fruits, or branches, as shown in Fig. 3e.

On the other hand, end-effectors or robots may be damaged when forcibly picking the fruits occluded by branches or wires (Fu et al., 2020b). The OB and OW are not regarded as pickable fruits by picking robot, which thus can be treated as one class (referred to OB&OW). Therefore, according to the robotic picking strategy, kiwifruits could be classified into four classes: NO, OL, OF, and OB&OW. This study proposed a hypothesis that detection precision may be different when fruits are divided into different classes, i.e., four classes and five classes.

Kiwifruits were manually labeled with boxes, which were tangent to kiwifruit outlines, as shown in Fig. 4. LabelImg (Windows version with python 2.7) was applied to label kiwifruit images and generate corresponding label files. For the five-classes, fruits inside purple, yellow, blue, red, and green boxes represented the NO, OL, OF, OB, and OW, respectively, as shown in Fig. 4a. For the four-classes, fruits in purple, orange, light green, and cyan boxes represented the NO, OL, OF, and OB&OW, respectively, as shown in Fig. 4b. XML format annotation files, including folder name, image name, image path, image size, fruit class name, and pixel coordinates of the label boxes, were generated after labeling. After labeling, the image dataset was randomly divided into training set and test set with a ratio of 4–1.

In this work, the training set was augmented by image rotation and mirroring. For image rotation, the rotation angles were selected as 90°, 180°, and 270°. Image mirroring included horizontal and vertical mirroring. The horizontal mirroring transformed the left and right sides of the image center on the vertical line of the image. The vertical mirroring transformed the upper and lower sides of the image center on the horizontal centerline of the image. The number of images was augmented from 1160 to 5800 by the above methods.

2.3. Networks and training hyperparameters

Structures of YOLOv3 and YOLOv4 are shown in Figs. 5 and 6, respectively. Both networks consist of backbone, neck, and heads. The backbone of YOLOv3 is Darknet53, composed of CBL (Conv., BN, Leaky ReLU) and Res*X* (*X* referred to 1, 2, 4, and 8) blocks. CBL block includes Conv. (Convolutional) layers, BN (Batch Normalization), and Leaky ReLU activation function. Res*X* consists of CBL block and Res. (Residual) Unit blocks, which are mainly composed of CBL blocks. CSPDarknet53 is used in YOLOv4 backbone, including CBM (Conv., BN, Mish) and CSP*X* (Cross Stage Partial, *X* referred to 1, 2, 4, and 8) blocks. CBM block consists of Conv. layers, BN, and Mish activation function. CSP*X* is composed of CBM blocks and Res. Unit block that also mainly composed of CBM blocks. CSP block can reduce the amount of calculation while ensuring precision (Wang et al., 2019). Mish activation function is used for the training of the CSPDarknet53 to increase the precision of network (Misra, 2019).

In the neck part, YOLOv3 mainly use FPN (Feature Pyramid Networks), which is composed of CBL blocks, Upsamplings, and Concat

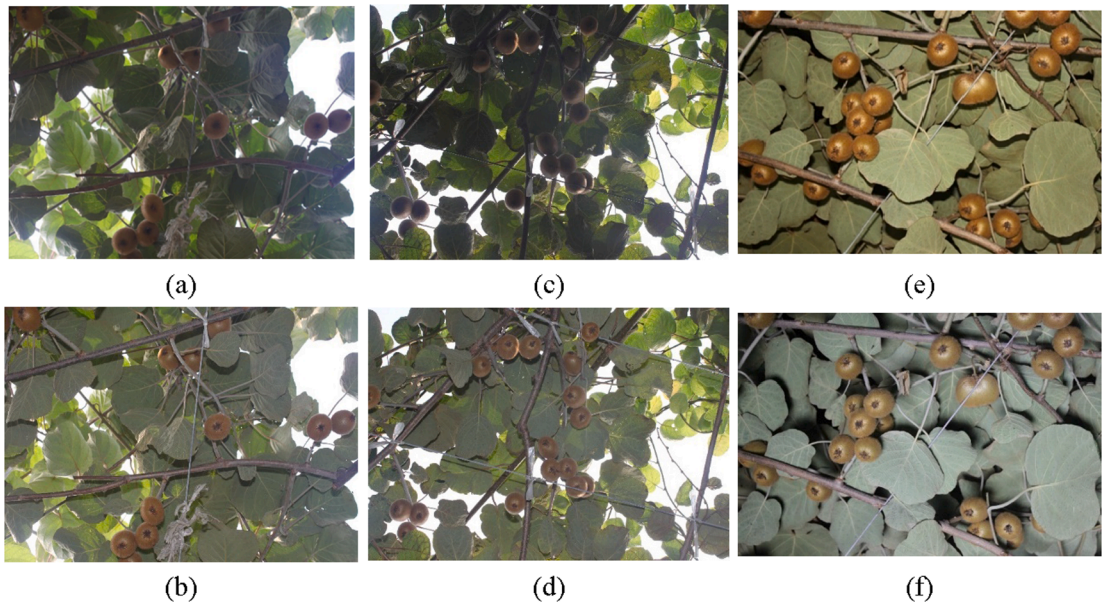

Fig. 2. Kiwifruit images under the orchard conditions with different illumination. (a) Morning without flash; (b) Morning with flash; (c) Afternoon without flash; (d) Afternoon with flash; (e) Night with flash; (f) Night with LED illumination.

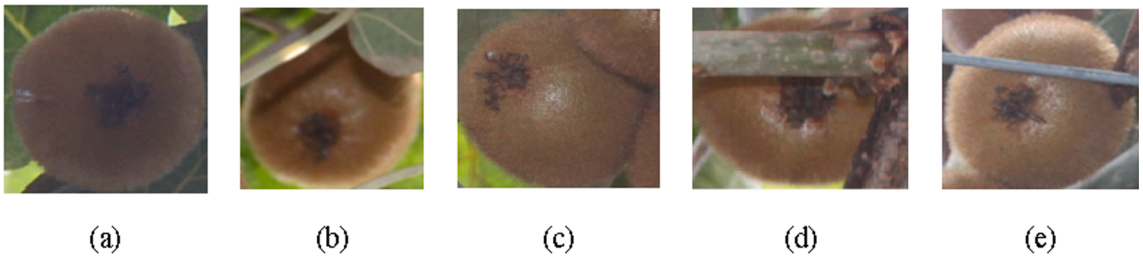

Fig. 3. Different classes of kiwifruit images. (a) Fruit not occluded (NO); (b) Fruit occluded by leaves (OL); (c) Fruit occluded by other fruits (OF); (d) Fruit occluded by branches (OB); (e) Fruit occluded by wires (OW). For the four-classes, (d) and (e) are grouped as one class (OB&OW) based on kiwifruit robotic picking strategy.

Fig. 4. Examples of kiwifruits labeled into (a) five classes, where fruits inside purple, yellow, blue, red, and green boxes referred to the NO, OL, OF, OB, and OW, respectively; and (b) four classes, where fruits in purple, orange, light green, and cyan boxes referred to the NO, OL, OF, and OB&OW, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

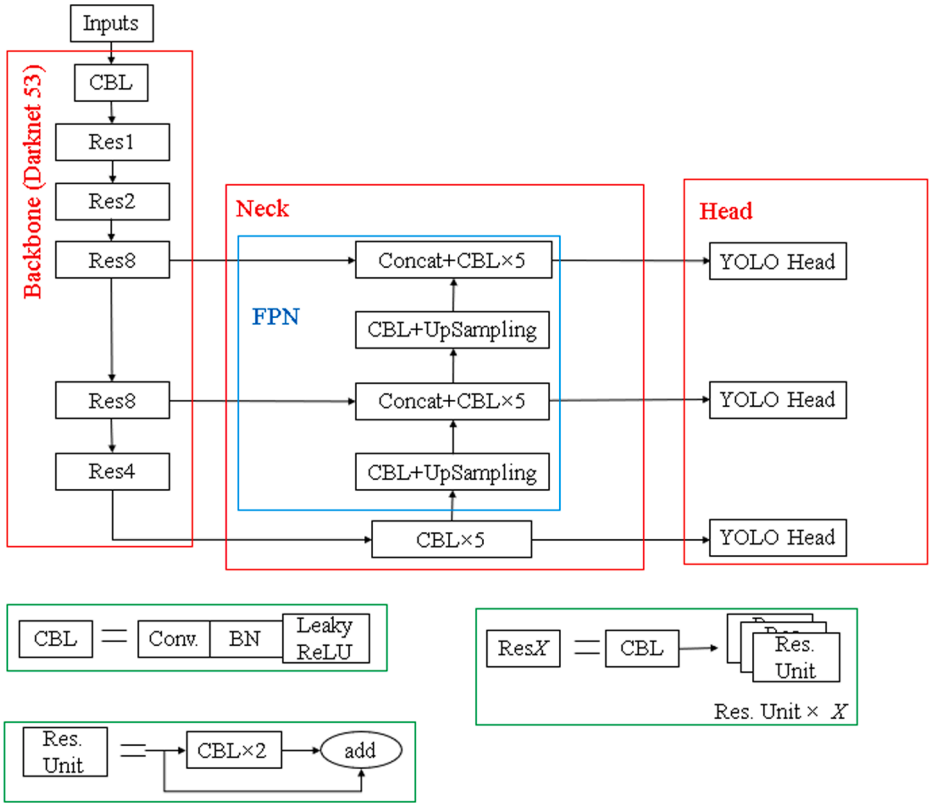

Fig. 5. Structure of YOLOv3, which consists of backbone (Darknet53), neck (FPN), and heads (YOLO Heads) (Redmon and Farhadi, 2018). CBL block includes Conv. layers, BN, and Leaky ReLU activation function. Res. Unit block is mainly composed of CBL blocks. Res*X* (*X* referred to 1, 2, 4, and 8) consists of CBL block and Res. Unit blocks.

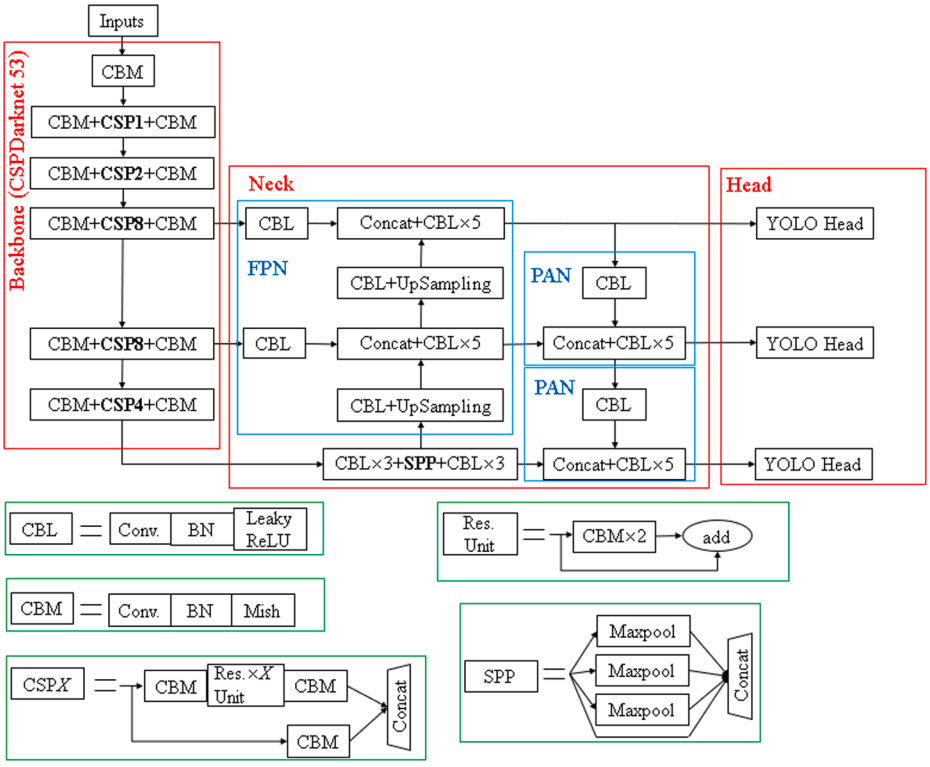

Fig. 6. Structure of YOLOv4, which consists of backbone (CSPDarknet53), neck (FPN, PAN, and SSP), and heads (YOLO Heads) (Bochkovskiy et al., 2020). CBM block includes Conv. layers, BN, and Mish activation function. Res. Unit block is mainly composed of CBM blocks. CSP*X* (*X* referred to1, 2, 4, and 8) consists of CBM blocks and Res. Unit blocks. SPP is mainly composed of Maxpools.

functions. In addition to FPN, PAN (Pixel Aggression Network) blocks and SSP (Spatial Pyramid Pooling) is used in YOLOv4 neck part. PAN block can further improve the ability of feature extraction (Cheng and Zhang, 2020). SPP composed of Maxpools can get an output of fixed size regardless of input image size/scale (He et al., 2015). The Head part of YOLOv4 is similar to YOLOv3, and uses the concept of feature pyramids (Lin et al., 2017) using three box predictions to extract features.

In this work, training platform included a computer with Intel Core i5-6400 (2.70 GHz) quad-core CPU, and a GPU of NVidia GeForce GTX 1080 8 GB GPU (2,560 CUDA cores) and 16 GB of memory, running on a Windows 10 64 bits system. Software included OpenCV 3.4.2, Visual Studio 2017, CUDA 10.0, cuDNN 7.6.5, Python 2.7, and CMake-3.16. The experiment was implemented in the Darknet framework. The network input size was 480 × 480, and batch size was set as 64. The stochastic gradient descent was applied for training with a momentum of 0.9 and a weight decay of 0.0005. An initial value of 0.001 was set as the learning rate of the network. Iterations of 30,000 were set to observe the training situation. During the training, all models used the same dataset and training parameters. Transfer learning was used for training models, which could adapt to new tasks and lead to faster and more accurate training results (Dias et al., 2018; Fu et al., 2020a).

2.4. Performance evaluation

The performance was evaluated using average precision (AP_i) of each class, mean average precision (mAP_k), and average detection speed. Among them, AP_i was calculated by precision (P_i) and recall (R_i). The P_i and R_i were defined in Eqs. (1) and (2), respectively. The i value in the five-classes represents the i^{th} class: NO ($i = 1$), OL ($i = 2$), OF ($i = 3$), OB ($i = 4$), and OW ($i = 5$). The meaning of i value in the four-classes is similar to that in the five-classes.

$$P_i = TP_i/(TP_i + FP_i) \quad (1)$$

$$R_i = TP_i/(TP_i + FN_i) \quad (2)$$

where TP_i (True Positives) indicates the number of detected correctly kiwifruits in the i^{th} class, FP_i (False Positives) refers to the number of detected falsely kiwifruits in the i^{th} class, and FN_i (False Negatives) represents the number of missed kiwifruits in the i^{th} class.

The AP_i is defined in Eq. (3), which is the area under the P_i and R_i curve. AP is a measure for the sensitivity of the network to target detection, and it is also an index that reflects the performance of the network. The mAP_k is defined in Eq. (4), which is the average value of AP for the k (4 or 5) classes kiwifruit.

$$AP_i = \int_0^1 P_i(R_i)dR_i \quad (3)$$

$$mAP_k = \frac{1}{k}\sum_{i=1}^{k} AP_i \quad (4)$$

3. Results and discussion

3.1. Training evaluation

Training loss curves of YOLOv3 and YOLOv4 in the four-classes and five-classes were converged, as shown in Fig. 7, where different line types and colors represented different models with different classes. As the number of iterations increased, the loss values decreased gradually. Similar curve trends were obtained in the five-classes and four-classes using the same network. Obviously, oscillation amplitude of curves of YOLOv3 was smaller than that of YOLOv4. Compared with YOLOv3, YOLOv4 had a slower convergence speed. It is because YOLOv4 has more network convolutional layers and requires more time to learn. After approximately 23,000 iterations, YOLOv3 in the four-classes and five-classes both began to reach stable loss values, which were about 1.5. The loss values of YOLOv4 in the four-classes and five-classes both gradually stabilized at about 7.5 after approximately 25,000 iterations.

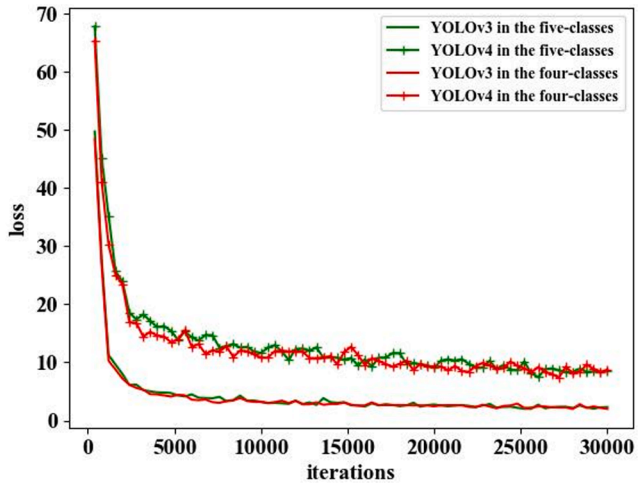

Fig. 7. Training loss curves of YOLOv3 and YOLOv4 in the four-classes and five-classes.

The convergent loss curves demonstrate that our models efficiently learn target features, and the models trained can be used to detect fruits.

3.2. Comparison of the four-classes and five-classes

The detection results of the five-classes were better than the four-classes, as shown in Table 1. The results obtained were consistent with the hypothesis that proposed in this study. For YOLOv4, the mAP of fruits in the five-classes was 91.9%, which was 0.4% higher than that in the four-classes. A similar situation also occurred on YOLOv3, where the mAP in the five-classes (91.1%) was 0.6% higher than that in the four-classes (89.5%). Examples of fruit detection on orchard images were shown in Fig. 8, where detected falsely fruit was marked by black rectangles manually drawn, missed fruits by pink rectangles, and example of detected correctly fruits by white rectangles, respectively. The same image in the test set was detected by YOLOv4 in the four-classes and five-classes, as shown in Fig. 8a and d. There were two missed fruits and one detected falsely fruit in the four-classes, but only one missed fruit in the five-classes. Fruits in the five-classes achieved higher mAP using the same network, dataset, and training hyperparameters.

Compared with the OB and OW in the five-classes, the AP of the OB&OW in the four-classes was greatly decreased. For YOLOv4, the AP of the OB and OW in the five-classes were 4.6% and 6.9% higher than the OB&OW in the four-classes. For YOLOv3, the AP of the OB and OW in the five-classes were 5.2% and 8.4% higher than the OB&OW in the four-classes. Taking YOLOv4 as an example, the OB could be detected in the five-classes (Fig. 8c), but it was missed in the four-classes (Fig. 8b). The reason may be that features such as colors and shapes of wires and branches are quite different, where specific features are learned by models from branches and wires, respectively. Learning more specific features is beneficial to improving the AP. Therefore, fruits classified into the five classes are more appropriate for multi-classes kiwifruit detection.

3.3. Comparison of YOLOv3 and YOLOv4

Compared with YOLOv3, YOLOv4 achieved better detection results, as shown in Table 1. For the five-classes, the mAP of YOLOv4 was 91.9%, which was 0.8% higher than YOLOv3. For the four-classes, the mAP of YOLOv4 was 91.5%, which was 1.0% higher than YOLOv3. The result was similar to Wu et al. (2020), which reported that the mAP of YOLOv4 was 5.9% higher than YOLOv3 on apple flower detection from three different apple varieties. The same image in the test set was detected by YOLOv3 and YOLOv4 in the five-classes to reflect differences of the detection results between the two models, as shown in Fig. 9a and Fig. 9d. Black rectangles manually drawn marked detected falsely fruits and white rectangles manually drawn marked example of detected correctly fruits. The NO, OL, OF, OB, and OW were detected in purple, yellow, blue, red, and green boxes, respectively. Three fruits were detected falsely by YOLOv3, but only one by YOLOv4. The reason is that using the Mish activation function and adding PAN block in YOLOv4 contribute to an improvement of model precision. From the detection results, YOLOv4 could detect fruits more accurately.

For the OB and OW, detection results of YOLOv4 were better than YOLOv3. For the five-classes, the AP of the OB and OW achieved by YOLOv4 were 1.8% and 0.9% higher than those achieved by YOLOv3. Similar results were obtained on the four-classes detection. The AP of the OB&OW achieved by YOLOv4 was 2.4% higher than YOLOv3. As shown in Fig. 9, the OB was detected falsely as the OF in YOLOv3 (Fig. 9b), while detected correctly by YOLOv4 (Fig. 9c). For the robotic picking, accurate detection is helpful to avoid detecting the OB or OW as pickable targets. Therefore, YOLOv4 can better complete multi-classes kiwifruit detection tasks.

However, the difference from Bochkovskiy et al. (2020) is that the mAP of YOLOv4 has not improved 10% than YOLOv3 in our study. The reason may be that MS COCO (Microsoft Common Objects in Context) dataset used by Bochkovskiy et al. (2020) contains 80 different targets with different shapes and sizes, where small samples account for about one-third (Lin et al., 2014). However, kiwifruits may be not that small to be detected by YOLOv3 and YOLOv4, and thus the improvement of the mAP in our study was not that high by YOLOv4. In our dataset, the branch or wire in front of fruit can be treated as small objected. Therefore, YOLOv4 had higher AP than YOLOv3 on OB, OW, and OB&OW that including the branches and wires, as shown in Table 1.

3.4. Detection speed

Significant differences were found in the average detection speed of YOLOv3 and YOLOv4, but no in the four-classes and five-classes, as shown in Table 1. The images in the test set were used to calculate the average detection speed. For the same network, the average detection speed of the four-classes and five-classes were basically the same. For the five-classes, YOLOv4 took an average of 25.5 ms to detect an image with 2352 × 1568 pixels, which was 3.8 ms slower than YOLOv3. For the

Table 1
Multi-classes kiwifruit detection results with YOLOv3 and YOLOv4 in the five-classes and four-classes.

Model	Classes	AP					mAP	Average detection speed (ms per image)
		NO	OL	OF	OB	OW		
					OB&OW			
YOLOv3	Five	95.9%	87.2%	92.0%	88.7%	91.9%	91.1%	21.7 ± 0.08^{a}
	Four	95.6%	87.0%	91.8%	83.5%		89.5%	21.6 ± 0.06^{a}
YOLOv4	Five	95.7%	88.2%	92.4%	90.5%	92.8%	91.9%	25.5 ± 0.13^{b}
	Four	96.1%	90.5%	93.5%	85.9%		91.5%	25.6 ± 0.10^{b}

Note: Same letters in the last column represent no significant difference at the 0.05 level.

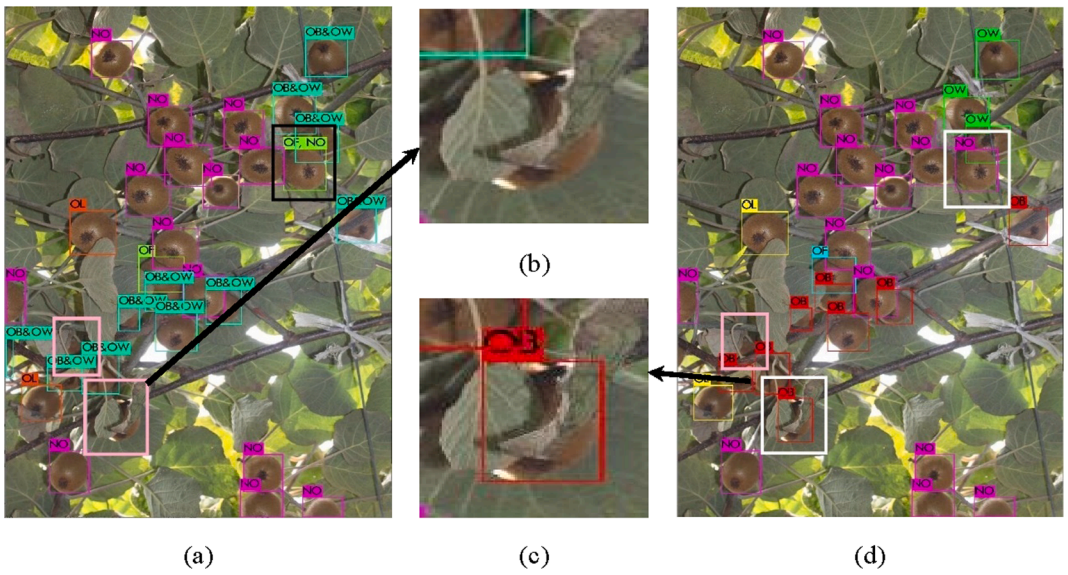

Fig. 8. Examples of kiwifruit detected as four classes and five classes by YOLOv4. (a) Detection result of the four-classes; (d) Detection result of the five-classes. For the five-classes, the NO, OL, OF, OB, and OW were detected in purple, yellow, blue, red, and green boxes, respectively. For the four-classes, the NO, OL, OF, and OB&OW were detected inside purple, orange, light green, and cyan boxes, respectively. Detected falsely fruit was marked by black rectangles manually drawn, missed fruits by pink rectangles, and example of detected correctly fruits by white rectangles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

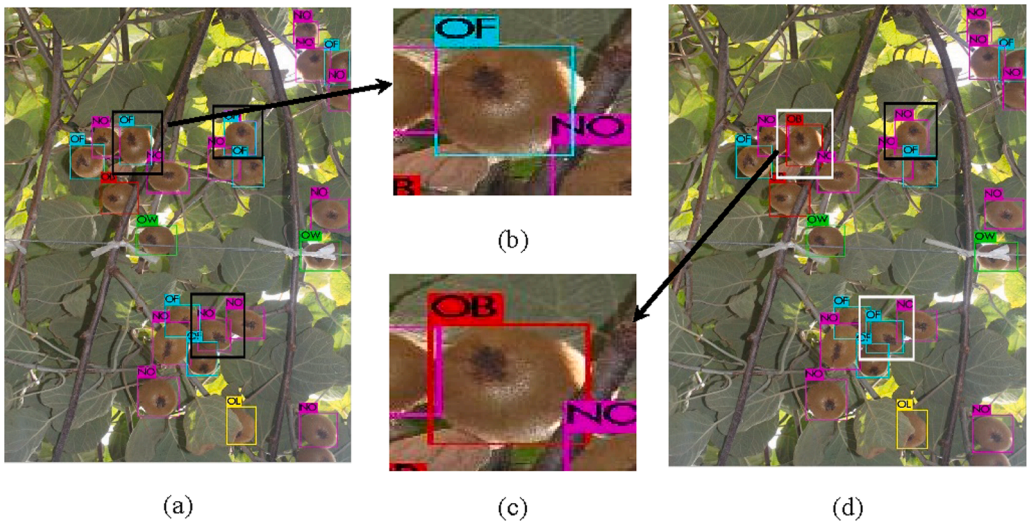

Fig. 9. Examples of kiwifruit detection by YOLOv3 and YOLOv4 in the five-classes. (a) Detection result of YOLOv3; (d) Detection result of YOLOv4; The NO, OL, OF, OB, and OW were detected in purple, yellow, blue, red, and green boxes, respectively. Black rectangles manually drawn marked detected falsely fruits and white rectangles manually drawn marked example of detected correctly fruits. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

four-classes, the average detection speed of YOLOv4 was 25.6 ms per image, which was 4.0 ms slower than YOLOv3. The average detection speed of YOLOv4 was slower than YOLOv3. The reason may be that YOLOv4 has more convolutional layers, which needs more calculations. Currently, it takes about 2.8 s to pick a kiwifruit with a robot (Williams et al., 2020). The detection time is much shorter than a picking cycle of 2.8 s. Therefore, the detection speed of our models can meet the requirements of real-time detection.

Overall, YOLOv4 in the five-classes was more suitable for multi-class kiwifruit detection. The highest mAP was achieved by YOLOv4 in the five-classes. The detection speed of YOLOv4 was slower than YOLOv3, which did not affect real-time detection. Compared to detection speed, the impact of the mAP of multi-class kiwifruit detection was more critical. High mAP was beneficial to avoiding detecting the OB or OW as pickable targets.

3.5. Comparison with other kiwifruit detection studies

Kiwifruits were labeled and detected as only one class in most studies. Some of them (Fu et al., 2018; Liu et al., 2020; Mu et al., 2019) classified the fruits into different classes to analyze detection results, as shown in Table 2. In these studies, OL and OB were treated as one class. However, no study on the OW was reported. Detection results of different classes were calculated by sampling and statistical methods in these studies. Compared with these studies, the fast detection speed of YOLOv4 in our study was highlighted. The detection speed of YOLOv4 was more than five times faster than Faster R-CNN with VGG16 (Liu et al. 2020), which was the fastest detection speed of fruits labeled and detected as only one class.

Multi-class kiwifruit detection achieved higher performance when fruits were labeled and detected as more classes. The NO obtained high detection accuracy in all studies. For the OL, the detection result of our study was 2.6% higher than the highest detection results (Fu et al. 2018)

Table 2
Results from other studies on kiwifruit detection.

	Method	Resolution	Detection results of different classes					Average detection speed (ms per image)
			NO	OL	OB	OF	OW	
Fu et al. (2018)	Faster R-CNN with ZFNet	2352 × 1568	96.7%	85.6%		82.5%		274
Liu et al. (2020)	Faster R-CNN with VGG16	512 × 424	96.7%	83.6%		87.4%		134
Mu et al. (2019)	Faster R-CNN with AlexNet	1920 × 1080	94.8%	83.0%		89.5%		1000
Our study	YOLOv4	2352 × 1568	95.7%	88.2%	90.5%	92.4%	92.8%	25.5

of fruits labeled and detected as only one class. The same situation also appeared on the OF. The detection result of our study was 2.9% higher than the highest detection results (Mu et al., 2019). Compared with labeling fruits as only one class, labeling fruits as multiple classes help learn occlusion features of leaves, branches, and occluded fruits. Fruits of different classes are detected with high accuracy due to the occlusion features learned. The above comparison results prove the promising of the hypothesis that proposed in this study.

4. Conclusions

This work labeled, trained, and detected kiwifruits into five classes (NO, OL, OF, OB, and OW) and four classes (NO, OL, OF, and OB&OW) according to the robotic picking strategy and field occlusions, which was aimed for avoiding detecting the OB or OW as pickable targets. In total, 1160 original images with a pixel resolution of 2352 × 1568 were acquired. After augmenting, all images were divided into training set and test set with a ratio of 4-1. YOLOv3 and YOLOv4 in the five-classes and four-classes were compared by mAP and detection speed. The results showed that the highest mAP of 91.9% was achieved by YOLOv4 in the five-classes, which cost 25.5 ms on average to process an image. From the results, fruits labeled, trained, and detected into more classes can achieve higher mAP. Compared with a picking cycle of 2.78 s per fruit, the speed enables the fruit detection system to run in real-time. The results illustrate that multi-classes kiwifruit detection is helpful for reducing the possibility of the end effector or robot damage. In addition, the multi-classes detection results can be further employed to develop fruit picking order and path, such as NO should be picked firstly before OF.

CRediT authorship contribution statement

Rui Suo: Data curation, Investigation, Writing - original draft. **Fangfang Gao:** Writing - review & editing. **Zhongxian Zhou:** Writing - review & editing. **Longsheng Fu:** Conceptualization, Data curation, Methodology, Supervision, Writing - review & editing. **Zhenzhen Song:** Investigation, Methodology, Writing - review & editing. **Jaspreet Dhupia:** Methodology, Writing - review & editing. **Rui Li:** Methodology, Writing - review & editing. **Yongjie Cui:** Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the Recruitment Program of High-End Foreign Experts of the State Administration of Foreign Experts Affairs, Ministry of Science and Technology, China (G20200027075); China Postdoctoral Science Foundation funded project (2019M663832); Fundamental Research Funds for the Central Universities of China (2452020170); National Natural Science Foundation of China (grant number 31971805).

References

Bochkovskiy, A., Wang, C., Liao, H.Y.M., 2020. YOLOv4: Optimal speed and accuracy of object detection. arXiv Prepr. arXiv:2004.10934.
Cheng, Z., Zhang, F., 2020. Flower end-to-end detection based on YOLOv4 using a mobile device. Wirel. Commun. Mob. Comput. 2020, 8870649. https://doi.org/10.1155/2020/8870649.
Dias, P.A., Tabb, A., Medeiros, H., 2018. Apple flower detection using deep convolutional networks. Comput. Ind. 99, 17–28. https://doi.org/10.1016/j.compind.2018.03.010.
Fathallah, F.A., 2010. Musculoskeletal disorders in labor-intensive agriculture. Appl. Ergon. 41, 738–743. https://doi.org/10.1016/j.apergo.2010.03.003.
Fazayeli, A., Kamgar, S., Mehdi, S., Fazayeli, H., De, M., 2019. Dielectric spectroscopy as a potential technique for prediction of kiwifruit quality indices during storage. Inf. Process. Agric. 6, 479–486. https://doi.org/10.1016/j.inpa.2019.02.002.
Feng, J., Zeng, L., He, L., 2019. Apple fruit recognition algorithm based on multi-spectral dynamic image analysis. Sensors 19, 949. https://doi.org/10.3390/s19040949.
Fu, L., Feng, Y., Majeed, Y., Zhang, X., Zhang, J., Karkee, M., Zhang, Q., 2018. Kiwifruit detection in field images using Faster R-CNN with ZFNet. IFAC-PapersOnLine 51, 45–50. https://doi.org/10.1016/j.ifacol.2018.08.059.
Fu, L., Feng, Y., Wu, J., Liu, Z., Gao, F., Majeed, Y., Al-Mallahi, A., Zhang, Q., Li, R., Cui, Y., 2020a. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model. Precis. Agric. https://doi.org/10.1007/s11119-020-09754-y.
Fu, L., Majeed, Y., Zhang, X., Karkee, M., Zhang, Q., 2020b. Faster R-CNN–based apple detection in dense-foliage fruiting-wall trees using RGB and depth features for robotic harvesting. Biosyst. Eng. 197, 245–256. https://doi.org/10.1016/j.biosystemseng.2020.07.007.
Fu, L., Tola, E., Al-Mallahi, A., Li, R., Cui, Y., 2019. A novel image processing algorithm to separate linearly clustered kiwifruits. Biosyst. Eng. 183, 184–195. https://doi.org/10.1016/j.biosystemseng.2019.04.024.
Fu, L., Wang, B., Cui, Y., Su, S., Gejima, Y., Kobayashi, T., 2015. Kiwifruit recognition at nighttime using artificial lighting based on machine vision. Int. J. Agric. Biol. Eng. 8, 52–59. https://doi.org/10.3965/j.ijabe.20150804.1576.
Gao, F., Fu, L., Zhang, X., Majeed, Y., Li, R., Karkee, M., Zhang, Q., 2020. Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN. Comput. Electron. Agric. 176, 105634. https://doi.org/10.1016/j.compag.2020.105634.
García-Quiroga, M., Nunes-Damaceno, M., Gómez-López, M., Arbones-Maciñeira, E., Muñoz-Ferreiro, N., Vázquez-Odériz, M.L., Romero-Rodríguez, M.A., 2015. Kiwifruit in syrup: Consumer acceptance, purchase intention and influence of processing and storage time on physicochemical and sensory characteristics. Food Bioprocess Technol. 8, 2268–2278. https://doi.org/10.1007/s11947-015-1571-3.
He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37, 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824.
Ju, M., Luo, H., Wang, Z., He, M., Chang, Z., Hui, B., 2019. Improved YOLOv3 algorithm and its application in small target detection. Acta Opt. Sin. 39, 0715004. https://doi.org/10.3788/AOS201939.0715004.
Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. Comput. Electron. Agric. 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016.
Kang, H., Chen, C., 2020. Fast implementation of real-time fruit detection in apple orchards using deep learning. Comput. Electron. Agric. 168, 105108. https://doi.org/10.1016/j.compag.2019.105108.
Leontowicz, H., Leontowicz, M., Latocha, P., Jesion, I., Park, Y.S., Katrich, E., Barasch, D., Nemirovski, A., Gorinstein, S., 2016. Bioactivity and nutritional properties of hardy kiwi fruit actinidia arguta in comparison with actinidia deliciosa "Hayward" and actinidia eriantha "Bidan". Food Chem. 196, 281–291. https://doi.org/10.1016/j.foodchem.2015.08.127.
Leontowicz, M., Jesion, I., Leontowicz, H., Park, Y.S., Namiesnik, J., Rombolaì, A.D., Weisz, M., Gorinstein, S., 2013. Health-promoting effects of ethylene-treated kiwifruit "Hayward" from conventional and organic crops in rats fed an atherogenic diet. J. Agric. Food Chem. 61, 3661–3668. https://doi.org/10.1021/jf400165k.
Li, J., Tang, Y., Zou, X., Lin, G., Wang, H., 2020. Detection of fruit-bearing branches and localization of litchi clusters for vision-based harvesting robots. IEEE Access 8, 117746–117758. https://doi.org/10.1109/ACCESS. 2020.3005386.
Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. In: Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017. pp. 936–944. https://doi.org/10.1109/CVPR.2017.106.
Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L., 2014. Microsoft COCO: Common Objects in Context. In: Comput. Vis. – ECCV 2014. pp. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48.

Liu, Z., Wu, J., Fu, L., Majeed, Y., Feng, Y., Li, R., Cui, Y., 2020. Improved kiwifruit detection using pre-trained VGG16 with RGB and NIR information fusion. IEEE Access 8, 2327–2336. https://doi.org/10.1109/ACCESS.2019.2962513.

Majeed, Y., Karkee, M., Zhang, Q., 2020. Estimating the trajectories of vine cordons in full foliage canopies for automated green shoot thinning in vineyards. Comput. Electron. Agric. 176, 105671. https://doi.org/10.1016/j.compag.2020.105671.

Misra, D., 2019. Mish: A self regularized non-monotonic neural activation function. arXiv Prepr. arXiv:1908.08681, 2019.

Mu, L., Cui, G., Liu, Y., Cui, Y., Fu, L., Gejima, Y., 2020. Design and simulation of an integrated end-effector for picking kiwifruit by robot. Inf. Process. Agric. 7, 58–71. https://doi.org/10.1016/j.inpa.2019.05.004.

Mu, L., Gao, Z., Cui, Y., Li, K., Liu, H., Fu, L., 2019. Kiwifruit detection of far-view and occluded fruit based on improved AlexNet. Trans. Chinese Soc. Agric. Mach. 50, 24–34. https://doi.org/10.6041/j.issn.1000-1298.2019.10.003.

Redmon, J., Farhadi, A., 2018. YOLOv3: An incremental improvement. arXiv Prepr. arXiv:1804.02767.

Richardson, D.P., Ansell, J., Drummond, L.N., 2018. The nutritional and health attributes of kiwifruit: a review. Eur. J. Nutr. 57, 2659–2676. https://doi.org/10.1007/s00394-018-1627-z.

Song, Z., Zhou, Z., Wang, W., Gao, F., Fu, L., Li, R., Cui, Y., 2021. Canopy segmentation and wire reconstruction for kiwifruit robotic harvesting. Comput. Electron. Agric. 181, 105933. https://doi.org/10.1016/j.compag.2020.105933.

UN Food & Agriculture Organization, 2020. Production of Kiwi (Fruit) by Countries. Retrieved 2020-06-25.

Wang, C., Liao, H.M., Yeh, Ih, Wu, Y., Chen, P., Hsieh, J., 2019. CSPNet: A new backbone that can enhance learning capability of CNN. In: IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work, pp. 1571–1580. https://doi.org/10.1109/CVPRW50498.2020.00203.

Williams, H., Jones, M.H., Nejati, M., Seabright, M., Bell, J., Penhall, N., Barnett, J., Duke, M., Scarfe, A., Ahn, H.S., Lim, J.Y., MacDonald, B., 2019. Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms. Biosyst. Eng. 181, 140–156. https://doi.org/10.1016/j.biosystemseng.2019.03.007.

Williams, H., Ting, C., Nejati, M., Jones, M.H., Penhall, N., Lim, J.Y., Seabright, M., Bell, J., Ahn, H.S., Scarfe, A., Duke, M., MacDonald, B., 2020. Improvements to and large-scale evaluation of a robotic kiwifruit harvester. J. F. Robot. 37, 187–201. https://doi.org/10.1002/rob.21890.

Wu, D., Lv, S., Jiang, M., Song, H., 2020. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. Comput. Electron. Agric. 178, 105742. https://doi.org/10.1016/j.compag.2020.105742.

Xu, D., Wu, Y., 2020. Improved YOLO-V3 with densenet for multi-scale remote sensing target detection. Sensors 20, 4276. https://doi.org/10.3390/s20154276.

Yang, C., Lee, W.S., Gader, P., 2014. Hyperspectral band selection for detecting different blueberry fruit maturity stages. Comput. Electron. Agric. 109, 23–31. https://doi.org/10.1016/j.compag.2014.08.009.

Zhan, W., He, D., Shi, S., 2013. Recognition of kiwifruit in field based on Adaboost algorithm. Trans. Chinese Soc. Agric. Eng. 29, 140–146. https://doi.org/10.3969/j.issn.1002-6819.2013.23.019.

Zhang, Z., Igathinathane, C., Li, J., Cen, H., Lu, Y., Flores, P., 2020. Technology progress in mechanical harvest of fresh market apples. Comput. Electron. Agric. 175, 105606. https://doi.org/10.1016/j.compag.2020.105606.

Zhou, J., Fu, X., Zhou, S., Zhou, J., Ye, H., Nguyen, H.T., 2019. Automated segmentation of soybean plants from 3D point cloud using machine learning. Comput. Electron. Agric. 162, 143–153. https://doi.org/10.1016/j.compag.2019.04.014.

Zhou, Z., Song, Z., Fu, L., Gao, F., Li, R., Cui, Y., 2020. Real-time kiwifruit detection in orchard using deep learning on Android™ smartphones for yield estimation. Comput. Electron. Agric. 179, 105856. https://doi.org/10.1016/j.compag.2020.105856.