Computers and Electronics in Agriculture 193 (2022) 106641

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Real-time detection of kiwifruit flower and bud simultaneously in orchard using YOLOv4 for robotic pollination

Guo Li [a], Rui Suo [a], Guanao Zhao [a], Changqing Gao [a], Longsheng Fu [a,b,c,*], Fuxi Shi [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:
Flower
Bud
Deep learning
YOLO
Robotic pollination

ABSTRACT

Robotic pollination may help with reducing high labor requirements and saving expensive pollen on artificial pollination of kiwifruit. Fast and accurate detection of kiwifruit flower and bud simultaneously in an orchard is essential for robotic pollination. It not only makes precision pollination possible but also can predict blooming peak to estimate the optimal pollination timing. However, only kiwifruit flower has been labeled and detected in recent studies. Therefore, kiwifruit flower and bud were labeled, trained, and detected simultaneously for robotic pollination. Well-known YOLOv3 and recently released YOLOv4 were applied to do transfer learning for kiwifruit flower and bud detection. Both were trained in the same image dataset and compared by Average Precision (AP) and processing speed, which were aimed to find a better model. Results showed that mean AP (mAP) of YOLOv4 (97.61%) was higher than YOLOv3 (95.24%) on kiwifruit flower and bud detection. The AP of flower and bud detection achieved by YOLOv4 were 96.66% and 98.57%, respectively, which were 0.17% and 4.58% higher than that of YOLOv3. The detection speed of YOLOv4 was 38.64 ms per image with 4608 × 3456 pixels, which was resized to 608 × 608 pixels in detection. In addition, another image dataset was collected from different years and locations to demonstrate the generalizability of YOLOv3 and YOLOv4, which reported mAPs of 80.98% and 91.49% on them, respectively. It can be concluded that YOLOv4 is promising to achieve real-time detection of kiwifruit flower and bud simultaneously for further flower blooming peak estimation and robotic pollination.

1. Introduction

Kiwifruit is a dioecious plant and highly dependent on accurate pollination that faces challenges such as lack of pollinators and high labor cost. Its male and female parts are in separate flowers, which have different blooming periods (McPherson et al., 2001). Pollen collected from male flowers was taken to female flowers by bees for pollination (Ahn et al., 2018). However, quality of pollination may not be satisfactory due to low ratio of male trees and fewer and fewer bees in orchards (Barnett et al., 2017). Artificial pollination methods such as pollen blowers, dusters, and spray dispensers, give better results when natural pollination is insufficient (Duke et al., 2017). However, there is no doubt that manual pollination is a labor-intensive process resulting in high labor costs and time-consuming. Therefore, it is necessary to develop robotic pollination, which may replace bee and manual pollination to save labor and pollen (Ohi et al., 2018).

Real-time detection of kiwifruit flower and bud simultaneously is essential for robotic pollination. Pollination robot needs to detect kiwifruit flower before delivering pollen directly to it (Zhang et al., 2019). However, flowers of kiwifruit are not blooming at the same time, where only the blooming flowers that reveal stigmas can be pollinated (Gianni and Vania, 2018; Tacconi et al., 2016). Furthermore, spraying pollens too early or too late are hardly pollinate kiwifruit flowers adequately, which may cause trees not to produce enough fruits and waste pollen (Gonzalez et al., 1995; Williams et al., 2020). Therefore, it is necessary to detect flowers and buds simultaneously, which not only can select flowers for robotic pollination but also can estimate flowering peak period by counting detected flowers and buds for determining the optimal pollination timing (Dias et al., 2018a; Farjon et al., 2020).

Traditional machine vision algorithms for flower detection were

* Corresponding author at: College of Mechanical and Electronic Engineering, Northwest A&F University, China.
E-mail address: fulsh@nwafu.edu.cn (L. Fu).

https://doi.org/10.1016/j.compag.2021.106641
Received 22 March 2021; Received in revised form 12 December 2021; Accepted 19 December 2021
Available online 23 December 2021

mainly based on color and shape of flowers. Hong and Choi (2012) detected 100 species of flowers from 500 lab images using color- and edge-based contours, which obtained a success rate of 94.80%. (Pornpanomchai et al., 2020) used Red, Green and Blue (RGB) color values with flower size and edge of petals to find herb flowers, which reached a success rate of 94.00% and average process time of 0.87 s per image. However, traditional machine vision algorithms were not robust enough and inaccurate in complex environments (Tian et al., 2019). In addition, they were mainly focused on objects that have significant color from background and used complicated algorithms with many fixed thresholds, which resulted them only adaptable to some specific conditions (Fu et al., 2021; Liu et al., 2020).

Researchers have employed deep learning for flower detection, which conquered limitations of traditional machine learning. Deep learning methods could extract features of objects better than traditional methods under complex environments such as orchard (Fu et al., 2020; Gao et al., 2020). Lim et al. (2020) used Faster Region-Convolutional Neural Network (R-CNN) with Network Attached Storage (NAS) to detect kiwifruit flower and reported a Precision (P) of 96.80%. Tian et al. (2020) applied Mask R-CNN to gain apple flower growth stages and obtained the P of 96.43%. Dias et al. (2018b) fine-tuned a pre-trained Convolutional Neural Network (CNN) that followed Clarifai architecture and combined with Support Vector Machine (SVM) to predict bloom intensity of apple flower, which obtained the P of 92.70%. The Clarifai architecture averaged several large deep convolutional networks to further boost performance by using a visualization technique based on deconvolutional networks. Although above deep learning methods obtained high P, their detection speeds were slow. The detection speed of Faster R-CNN applied by Lim et al. (2020) was 1833 ms per image, while that of Mask R-CNN was 372 ms per image (Tian et al., 2020), which may not meet real-time operation of field robots (Wang et al., 2020; Wu et al., 2020). Furthermore, those object detections were all only on flowers, without consideration of buds. Colors of kiwifruit bud are similar to its flower, which made detection of flower and bud simultaneously more difficult. Besides, image processing speed needed to be considered for fast operation of a pollination robot in orchard. Therefore, a suitable deep learning network need to be chosen for kiwifruit flower and bud detection.

You Only Look Once (YOLO) obtains a balance of detection accuracy and speed. YOLO is based on a single CNN and "sees" an entire image during training and test time. Under Darknet framework, the most classic YOLO network is YOLOv3 developed in 2018, and the latest is YOLOv4 released in 2020 when we conducted this study. YOLOv3 runs in 22 ms using Microsoft Common Objects in Context (MS COCO dataset), which is faster than other one-stage networks like Single Shot Multi-Box Detector (SSD) (Redmon and Farhadi, 2018; Liu et al., 2016). Compared with some two-stage detection networks such as Faster R-CNN, YOLO has similar accuracy as them at a higher speed. Under the MS COCO dataset, Average Precision (AP) and detection speed of YOLOv4 were improved by 10% and 12% than YOLOv3, respectively (Bochkovskiy et al., 2020). Reports about YOLO mainly used YOLOv3 recently, while there were few reports about YOLOv4 that was released in 2020. Therefore, YOLOv3 and YOLOv4 were both employed for detecting kiwifruit flower and bud simultaneously.

Deep learning methods were used to train and detect kiwifruit flower and bud simultaneously, which were expected to not only make precision pollination but also could predict blooming peak days to estimate the optimal pollination timing. The parameters of YOLOv3 and YOLOv4 pre-training models were fine-tuned separately to be trained. In the process of training, this model extracts and learns features of kiwifruit buds and flowers. Finally, YOLOv4 was compared with YOLOv3 to choose the optimal model by different indicators, such as AP and detection speed.

2. Materials and methods

2.1. Image acquisition

Images of Hayward kiwifruit flower and bud were obtained from Meixian Kiwifruit Experimental Station in Shaanxi province, China, from flowering seasons in the years of 2019 and 2020. Hayward is the most cultivated variety in this area and its flowers and buds are hanging down from canopy and facing to ground, as the other kiwifruits. An ordinary single-lens reflex camera (Canon S110, Canon Inc., Tokyo, Japan) on "AUTO" mode with a resolution of 4608 × 3456 pixels was held at around 50 cm below flowers and buds to shoot them upwards and acquired 740 original images (dataset A). Furthermore, under the same shooting method, another 90 images (dataset B) were obtained from Yangling International Kiwifruit Innovation and Entrepreneurship Park in Shaanxi province, China, from the flowering seasons of Hayward kiwifruit in 2021. All images were taken under natural daylight conditions including several disturbances of occlusion and overlap, which were saved in Portable Network Graphics (PNG) format. Some examples of acquired images were shown in Fig. 1.

2.2. Image datasets

Dataset A was divided into a training set (592 images) and a validating set (148 images) with a ratio of 4 to 1, while dataset B was only employed for testing model generalizability. Under LabelImg (https://github.com/tzutalin/labelImg), kiwifruit flowers and buds in all acquired images were manually annotated as rectangles with labels "F" and "B", respectively, which saved annotation files in "xml" format. Some labeling examples of kiwifruit flowers and buds were shown in Fig. 2.

Data augmentation of image mirroring and image rotation was applied to avoid overfitting and to improve generalization ability of deep learning model in this study. The image mirroring (including horizontal and vertical mirroring) was implemented using OpenCV function "cv2.flip (image, dim)" based on Python. The horizontal mirroring transformed the left and right sides of the image centering on the vertical line of the image when the parameter "dim" was set to "1". The vertical mirroring transformed the upper and lower sides of the image centering on the horizontal centerline of the image when the parameter "dim" was set to "0". For the image rotation, OpenCV function "cv2.getRotationMatrix2D ((cols/2, rows/2), angle, 1)" based on Python was employed to rotate the original image, in which 90°, 180°, and 270° of rotation were achieved by changing the function parameter "angle" as 90, 180, and 270, respectively. In total, the training set was augmented from 592 images to 3552 images, which had been made publicly available on github.com (https://github.com/fu3lab/Kiwifruit_flower_and_bud_images).

2.3. YOLO network architecture

YOLO is currently one of the fastest object detection models, which redefines object detection as a regression problem and achieves favorable balance of detection accuracy and speed. As a one-stage network, YOLO divides an image into regions and predicts boundary boxes, probabilities, and conditional class probabilities, whose detection pipeline was shown in Fig. 3. Under the official Darknet framework, the latest two models are YOLOv3 and YOLOv4, established in 2018 and in 2020, respectively.

Darknet-53, as the backbone of YOLOv3, consists of convolutional and residual layers, which has an object detector with feature map up-sampling and concatenation. Prediction layer in YOLOv3 adopts Feature Pyramid Networks (FPN) structure, which is mainly used to improve the extraction ability of target features by fusing high and low layer features. The most salient feature of YOLOv3 is that it applies 1 × 1 detection kernels on feature maps of three different sizes at three

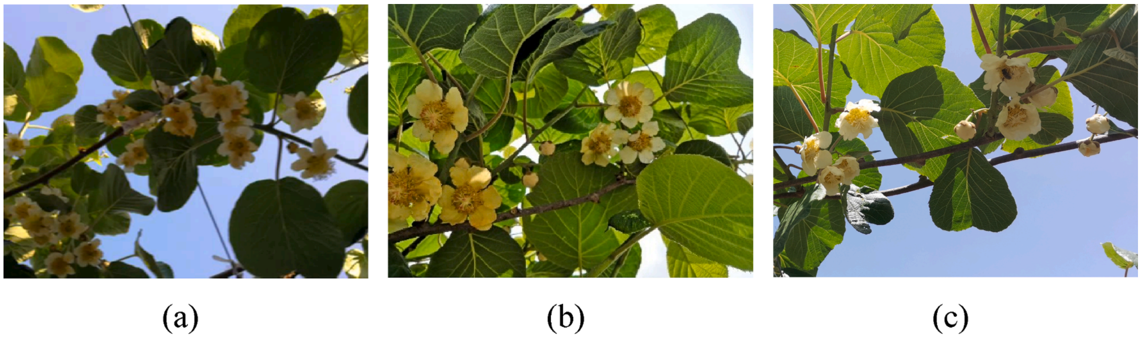

Fig. 1. Examples of kiwifruit flower and bud images acquired in orchard. (a), (b), and (c) were acquired in 2019, 2020, and 2021, respectively.

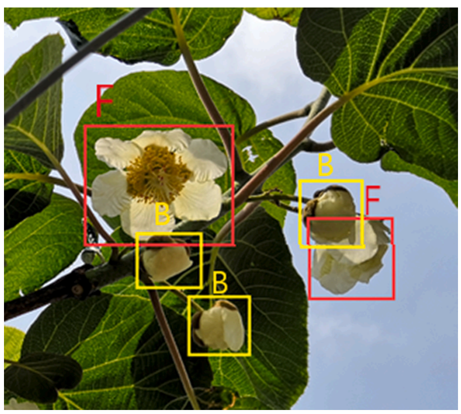

Fig. 2. Labeling examples of kiwifruit flower and bud. Bud was labeled as "B" using a yellow rectangle, and flower was labeled as "F" using a red rectangle. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

different places for detection.

YOLOv4, as the next version of YOLOv3, was proposed recently by Bochkovskiy et al. (2020). It is composed of Cross Stage Partial Darknet-53 (CSPDarknet-53), Spatial Pyramid Pooling (SPP) block, and Path Aggregation Network (PANet). The CSPDarknet-53 is a novel backbone, which can enhance the learning capability of CNN. The SSP block is added over CSPDarknet-53 to increase receptive field and separate out the most significant context features. Instead of FPN for object detection applied in YOLOv3, PANet is used for parameter aggregation with different detector levels in YOLOv4. Besides, Mosaic and Self Adversarial Training (SAT) are employed to augment data of training pipeline in YOLOv4, which expose scenarios of training set that would have otherwise been unseen.

2.4. Network training

Experiments were carried out based on deep learning framework of Darknet platform on a computer with Intel Core i5-6400 (2.70 GHz) quad-core CPU, NVidia GeForce GTX 1080 8 GB GPU (2,560 CUDA cores) and 16 GB of memory, running on a Windows 10 64-bit system. The software tools included CUDA 10.0, CUDNN 7.5, OpenCV 3.4.5, and Visual Studio 2017. Both YOLOv3 and YOLOv4 based on the Darknet framework were employed for training kiwifruit flower and bud detection network by transfer learning. Transfer learning was a machine learning method, which referred to a pre-trained model being reused in another task. Pre-trained models were applied based on the Darknet framework to train them, where input size, batch size, learning rate, momentum, and iteration were set in Table 1.

2.5. Performance evaluation

Evaluation indicators, i.e., P, Recall (R), F_1 score, AP, mAP, and detection speed, were used to evaluate trained models on the test dataset. An Intersection over Union (IoU) score more than 0.5 is generally considered as a good and acceptable detection. Otherwise, it is undetected. It is often used to calculate the criteria of an overlapping region of two target images and defined as Eq. (1).

$$IoU = \left|\frac{A \cap B}{A \cap B}\right| \tag{1}$$

where "A" represents the prediction bounding box and the "B" represents the true bounding box in Eq. (1).

P is a measure of detection result relevancy, while R is a measure of how many truly relevant detection results are returned. P and R were defined in Eq. (2) and Eq. (3), respectively. If a flower or bud is labeled as class A and detected as class A, it is true positive (TP); if a flower or bud is labeled as class A but is detected to another class, it is false negative (FN); if a flower or bud do not exist but is detected, it is false positive (FP). When we want to find an optimal blend of P and R, the two metrics are combined using what is called the F_1 score, as defined in Eq. (4). The higher P and R are, the higher the F_1 score is.

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$F_1\ \text{score} = \frac{2P \times R}{P + R} \tag{4}$$

AP_k was defined as the area under the P_k - R_k curve (the P_k as the vertical axis and the R_k as the horizontal axis) in Eq. (5), which was used to evaluate the performance of models in detecting each class. And mAP was defined in Eq. (6) as the average AP of the two classes (i.e., flower and bud). The higher the AP and mAP are, the better detection results of deep learning model obtain for a given object. The value k represented each class of objects in this study: flower ($k = 1$) and bud ($k = 2$). Also, detection speed was calculated to evaluate the performance of models.

$$AP_k = \int_0^1 P_k(R_k)\ dR_k \tag{5}$$

$$mAP = \frac{1}{k}\sum_{i=1}^{k} AP_i \tag{6}$$

3. Results and discussion

3.1. Training evaluation

Training loss curves of YOLOv3 and YOLOv4 had been converging,

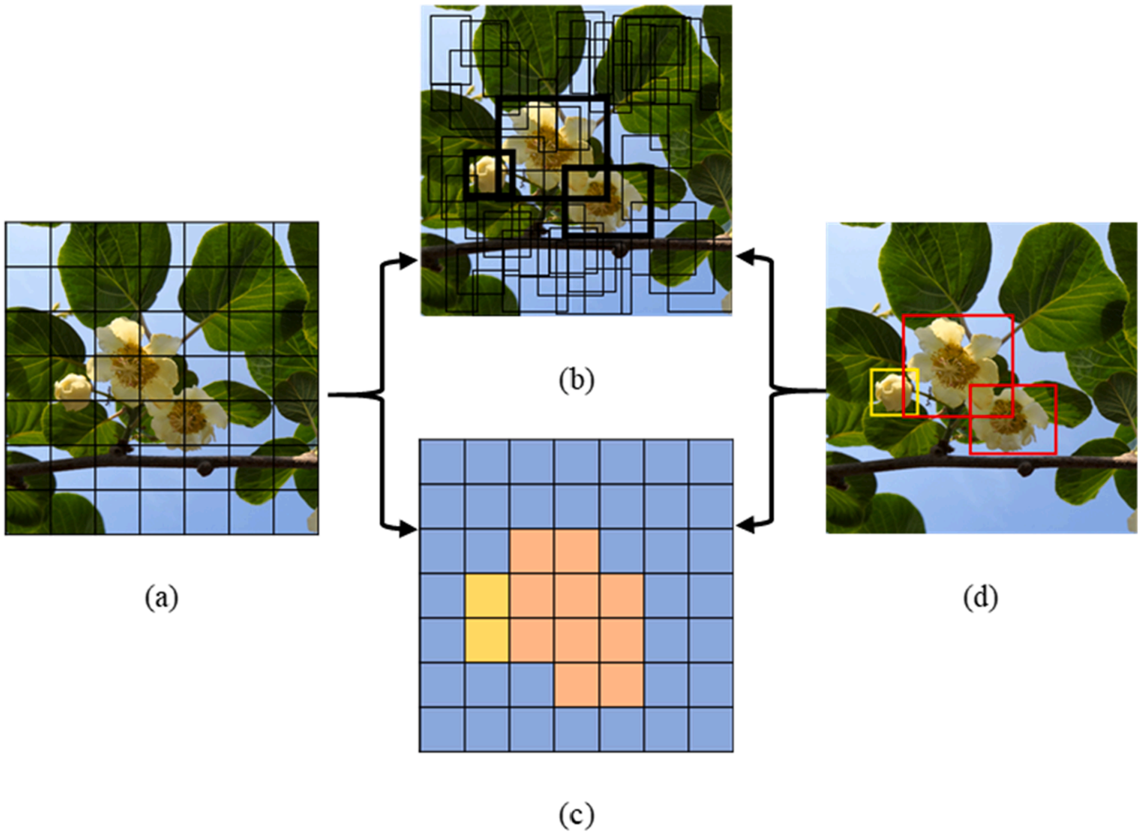

Fig. 3. Detection pipeline of YOLO. (a) Input image was divided into $S \times S$ grids; (b) Each grid cell predicted bounding boxes and confidence; (c) Each grid cell predicted class probabilities map; (d) Detection result of kiwifruit flower and bud.

Table 1
Network parameters of YOLOv3 and YOLOv4.

Model	Input	Batch size	Learning rate	Momentum	Decay	Iterations
YOLOv3	608 × 608	64	0.001	0.900	0.0005	15,000
YOLOv4	608 × 608	64	0.001	0.900	0.0005	15,000

as shown in Fig. 4, where abscissa and ordinate represented training steps and loss values. As the number of training iterations continually increased, the loss values of both YOLOv3 and YOLOv4 decreased slowly at first and then decreased slowly. The loss curve of YOLOv3 gradually converged near 0.5 after approximately 7000 iterations, while that of YOLOv4 converged near 1.2 after approximately 9000 iterations. Loss curves having converged, which represented that predicted outputs were believable and demonstrated that trained models had learned the features of flower and bud for their detection.

3.2. Comparison between YOLOv3 and YOLOv4

Most performance indices of YOLOv4, including mAP, F_1 score, AP, were better than those of YOLOv3 on dataset A, as shown in Table 2. F_1 score and mAP of YOLOv4 were 97.00% and 97.61%, which were 1.00% and 2.37% higher than that of YOLOv3, respectively. Besides, APs of flower and bud achieved by YOLOv4 were 0.17% and 4.58% higher than that achieved by YOLOv3, respectively. Jiang et al. (2020) obtained a similar performance that average recognition accuracies of YOLOv4 were 97.87%, 98.27%, 96.86%, and 96.92%, respectively, for goat eating, drinking, active and inactive behaviors detection, which were all better than YOLOv3. The reason is that CSPDarknet-53 used in YOLOv4

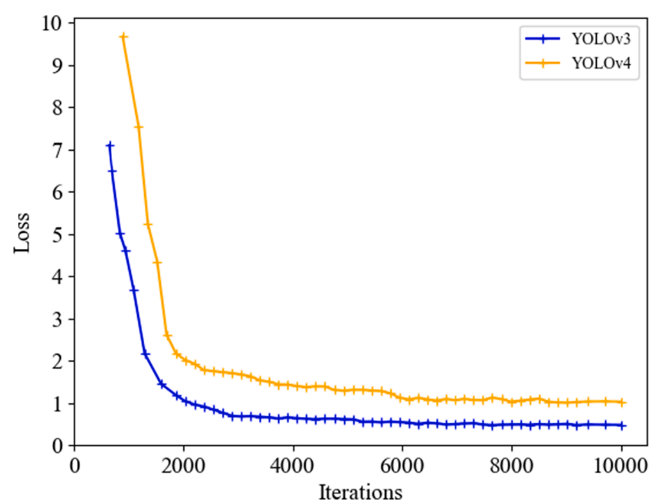

Fig. 4. Loss curves of YOLOv3 and YOLOv4. The blue and orange curves represented loss curves of YOLOv3 and YOLOv4, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

backbone may enhance detection capability of YOLOv4 (Wang et al., 2019). From the above results, YOLOv4 can better perform kiwifruit flower and bud images detection tasks in orchards.

The APs of flower and bud detection were all more than 93%, especially that of YOLOv4 on bud detection was the highest. The AP of flower was higher than that of bud by YOLOv3, while the AP of flower was lower than that of bud by YOLOv4, as shown in Table 2. For YOLOv3, although the AP of bud was 2.50% lower than that of flower, it

Table 2
Performances of the YOLOv3 and YOLOv4 on kiwifruit flower and bud detection.

Model	P (%)	R (%)	F_1 score (%)	mAP (%)	AP (%)		Average detection speed (ms/image)
					Flower	Bud	
YOLOv3	98.00	95.00	96.00	95.24	96.49	93.99	19.84 ± 0.21[a]
YOLOv4	97.00	96.00	97.00	97.61	96.66	98.57	38.64 ± 0.61[b]

Note: Same letters in the "speed" column represented no significant difference at the 0.05 level.

still reached 93.99%. The upsampled layers of YOLOv3 concatenated with the previous layers helped preserve the fine-grained features which helped in detecting small objects (Redmon and Farhadi, 2018). Therefore, the AP of bud by YOLOv3 reached a high value. As for YOLOv4, the AP of bud was 98.57%, which was 1.91% higher than that of flower. Some new features improved the AP of YOLOv4 on detecting kiwifruit flower and bud, especially on bud. Mosaic and SAT were used for data augmentation to improve generalization ability in YOLOv4 (Bochkovskiy et al., 2020). CSPDarknet-53 of YOLOv4 was a new backbone that enhanced the learning capability (Bochkovskiy et al., 2020; Wang et al., 2019). SPP and PANet of YOLOv4 were used for feature aggregation at multiple levels to preserve adequate spatial information for small object detection (Bochkovskiy et al., 2020). YOLOv4, which had better detection ability than YOLOv3 on small objects, could learn more target features to reach better APs on detecting kiwifruit flower and bud.

As shown in Table 2, the AP of bud detection that achieved by YOLOv4 was 4.58% higher than YOLOv3, which meant that YOLOv4 achieved better detection result on bud. As shown in Fig. 5, for the same image, when kiwifruit flowers overlap with each other, the edges of the detection rectangles overlapped slightly. Although the same dataset was used, as shown in Fig. 5a, YOLOv3 detected two buds as three buds in manually drawn blue rectangles. YOLOv4 detected same two buds correctly in manually drawn yellow rectangles, as shown in Fig. 5c. YOLOv4 achieved better detection performance of buds, which is helpful for better pollination and blooming peak estimation.

3.3. Results from other studies on kiwifruit flower detection

In most robotic pollination studies, only kiwifruit flower was labeled and detected, while our study trained both bud and flower. Although there were no reported studies on kiwifruit bud detection, previous studies with kiwifruit flower detection could still be compared to verify the effectiveness of YOLOv4 for kiwifruit flower detection.

The results of different networks such as Faster R-CNN and SSD were shown in Table 3. Lim et al. (2020) trained and compared three kiwifruit flower detection methods, including Faster R-CNN with NAS, SSD with Inception V2, and Faster R-CNN with Inception V2, which obtained the P

Table 3
Results from previous studies on kiwifruit flower detection.

Author	Object	Image pixel size	Model	P (%)	R (%)	Average detection speed (ms/image)
Lim et al. (2020)	Flower	1920 × 1080	Faster R-CNN with NAS	96.80	68.00	1833.00
			Faster R-CNN with Inception V2	90.40	75.80	58.00
			SSD with Inception V2	78.50	61.20	42.00
Williams et al. (2020)	Flower	1024 × 600	Faster R-CNN with Inception V2	91.00	80.00	—
Ours	Flower and bud	4608 × 3456	YOLOv4 with CSPDarknet-53	97.00	96.00	38.64

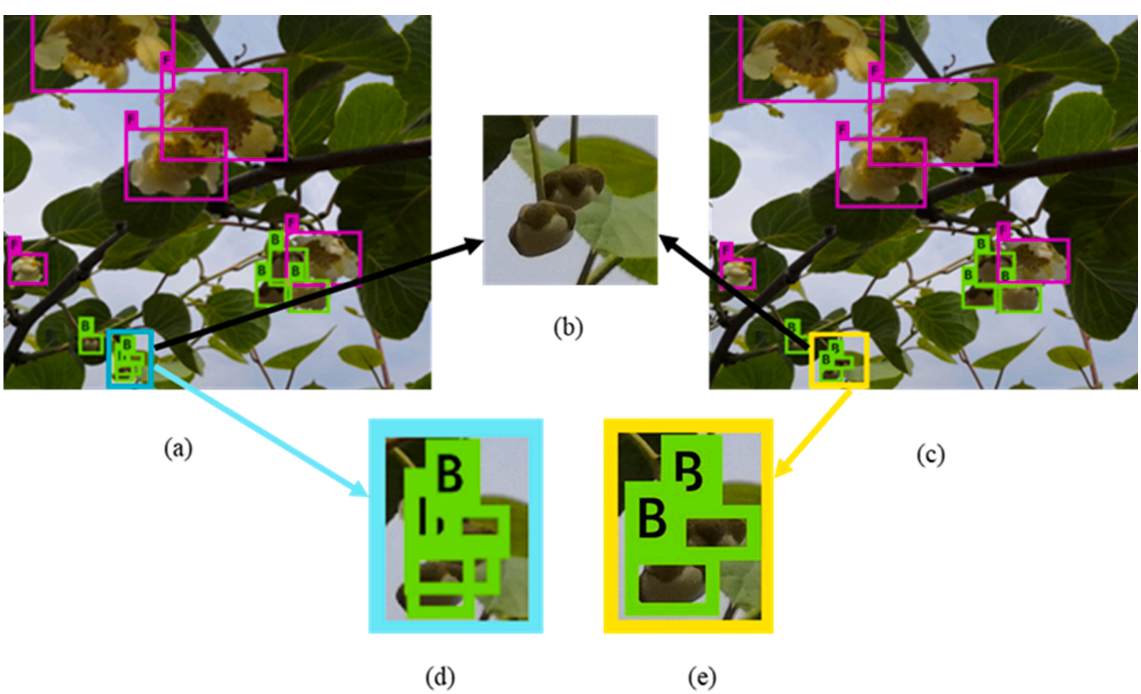

Fig. 5. Examples of kiwifruit flower and bud images detected by YOLOv3 (a) and YOLOv4 (c). Original image of two buds was showed in Fig. 5b. The label "F" with a pink rectangle was "flower" and label "B" with a green rectangle was "bud". The aqua rectangle was manually drawn to indicate that buds were detected falsely by YOLOv3, and the manually drawn yellow rectangle indicated that kiwifruit buds were correctly detected by YOLOv4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of 96.80%, 78.90%, and 90.40%, respectively. They chose Faster R-CNN with Inception V2 for their robot system because it showed better comprehensive performance for their dataset. Williams et al. (2020) obtained the P of 91.00% on kiwifruit flower detection based on Faster R-CNN with inception v2.0, which was 6.00% lower than that of our study. Compared with other studies, YOLOv4 obtained better performances on kiwifruit flower detection.

Flower and bud simultaneous detection were more conducive for robotic pollination and further flower blooming peak estimation. No bud was labeled or detected in previous studies, which meant that their pollination work was just in blooming days without bud information to help planning pollination schedule. Bud detection could be used for planning robotic pollination and predicting blooming peak days for robotic pollination. YOLOv4 in this study reached the highest mAP, which met detecting requirement of robotic pollination. Meanwhile, trained YOLOv4 could provide flower and bud information to predict blooming peak days for the optimal pollination timing.

3.4. Detection speed among different models

Compared with previous studies, YOLOv4 achieved the fastest detection speed. YOLO not only solved the problem of slow detection speed for a two-stage network such as Faster R-CNN with many CNN parameters but also was faster than other one-stage network SSD (Soviany and Ionescu, 2018). As shown in Table 3, YOLOv4 reached a higher P and a faster detection speed. The detection speed of YOLOv4 was 38.64 ms per image at 4068 × 3456 image pixels, which was 3.36 ms faster than SSD with Inception V2 at 1920 × 1080 image pixels.

In our test results, the detection speed of YOLOv4 was not faster than YOLOv3. The detection speed was 19.84 ms per image using YOLOv3, while 38.64 ms using YOLOv4, which could result from more CNN layers increasing inference time in YOLOv4 (Ogden and Guo, 2019). The possible minimum time for robotic pollinating was 500 ms per flower, including 50 ms for image processing, which were both met by YOLOv3 and YOLOv4 (Duke et al., 2017).

Although the detection speed of YOLOv4, which also met requirement of detection speed, was slower than that of YOLOv3, higher mAP was more beneficial to make precision pollination possible. Furthermore, higher mAP could predict blooming peak days to estimate the optimal pollination timing. Therefore, YOLOv4 was more suitable for detecting kiwifruit flower and bud simultaneously in orchard.

3.5. Generalizability of YOLOv3 and YOLOv4

Another image dataset (dataset B) was used to demonstrate YOLOv3 and YOLOv4 generalizability, which showed YOLOv4 had better performance. Detection results of YOLOv3 and YOLOv4 on dataset A and dataset B were shown in Table 4. The APs of flower detection by YOLOv3 and YOLOv4 on dataset B (85.73% and 92.47%, respectively) were slightly lower than those on dataset A. However, the AP of bud detection by YOLOv3 on dataset B was 17.77% lower than that of dataset A, while the AP of bud by YOLOv4 on dataset B was 90.50%, which was 8.07% lower than that of dataset A. Mean AP of YOLOv4 and YOLOv3 were 91.49% and 80.98% on dataset B, respectively, which were lower than those on dataset A. Most deep learning networks performed well on a test set that was the same batch as the training set but not that good on different datasets. It may be because some features learned under training set were not enough generalization in different datasets.

Table 4
Detection results of YOLOv3 and YOLOv4 on dataset A and dataset B.

Model	Dataset A			Dataset B		
	AP (%)		mAP (%)	AP (%)		mAP (%)
	Flower	Bud		Flower	Bud	
YOLOv3	96.49	93.99	95.42	85.73	76.22	80.98
YOLOv4	96.66	98.57	97.61	92.47	90.50	91.49

Detection results of YOLOv4 were better than those of YOLOv3 on dataset B. An example was shown in Fig. 6, where the label "F" with a pink rectangle was flower and label "B" with a green rectangle was bud. Original image of flower and bud was showed in Fig. 6b, where four buds and three flowers were labeled by orange rectangles and red rectangles, respectively. As shown in Table 4, the AP of flower detection that achieved by YOLOv4 was 6.74% higher than YOLOv3 on dataset B, while the AP of bud detection that achieved by YOLOv4 was 14.28% higher than YOLOv3 on the same test set. For the same image of dataset B, there were only one flower and one bud correctly detected by YOLOv3 as shown in Fig. 6a, while YOLOv4 correctly detected two flowers and four buds as shown in Fig. 6c. It may be that YOLOv4 had better learning capability resulting from data augmentation and a stronger ability to detect small objects. As shown in Fig. 6, one flower in the center of the original image was not detected by both YOLOv3 and YOLOv4. Maybe both YOLOv3 and YOLOv4 didn't learn the feature that the flower was occluded by two branches in training process. Under the same dataset B, the mAP of YOLOv4 still obtained 91.49%, which was 10.51% higher than that of YOLOv3. There was no doubt that detection performance of YOLOv4 was better than that of YOLOv3.

3.6. Limitations of robotic pollination

Although we have achieved considerable results in this study, there are still many issues of robot pollination that need to be solved. Robot pollination may have some limitations, such as slow speed and difficult to identify flower orientation. Robot is difficult for current mass pollination that consuming a huge amount of expensive pollen because of its low speed, which may be solved by employing multiple arms and extending working period. Some flowers may not open downward and need identify their orientations for an optimal robotic pollination, which requires more further researches on it.

4. Conclusions

Advanced deep learning has reached promising detection of kiwifruit flowers in the orchard, but most previous researches on kiwifruit pollination treated kiwifruit flowers as only one single category, which ignored agronomic situations of kiwifruit flowering. However, kiwifruit flowers are not blooming at the same time, where only the blooming flowers that reveal stigmas can be pollinated. Well-known YOLOv3 and recently released YOLOv4 were thus applied to train and detect kiwifruit flower and bud simultaneously. YOLOv4 achieved the best detection results, especially on buds, which meant it has better performance on small objects. Besides, YOLOv4 showed better generalizability than YOLOv3 on untrained images from different years and orchards that previously unseen. Furthermore, the detection speed of YOLOv4 was 38.64 ms on average to process each image, which met the requirement of detection time for robotic pollination. In conclusion, YOLOv4 was able to detect kiwifruit flower and bud in orchard accurately and fast, which not only made precision pollination possible but also could be used for further predicting blooming peak days to estimate the optimal pollination timing. In the future, continuous images of kiwifruit flower in different phenological stages are necessary to be collected for validating its performance and enlarging the training dataset, which may be more helpful for predicting blooming peak and estimating the optimal timing of robotic pollination.

CRediT authorship contribution statement

Guo Li: Data curation, Investigation, Writing – original draft. **Rui Suo:** Writing – review & editing. **Guanao Zhao:** Writing – review & editing. **Changqing Gao:** Writing – review & editing. **Longsheng Fu:** Conceptualization, Data curation, Methodology, Supervision, Writing –

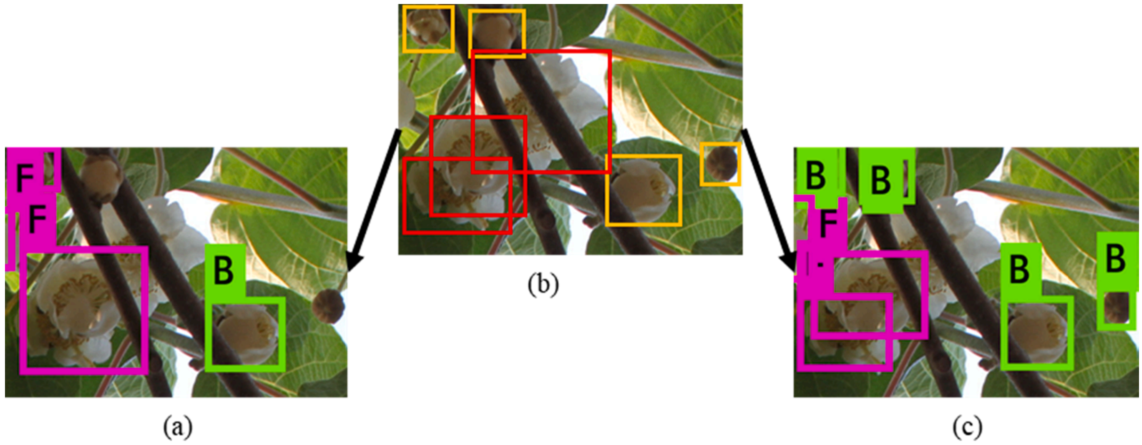

Fig. 6. Examples of kiwifruit flower and bud images detected by YOLOv3 (a) and YOLOv4 (c) on dataset B. The label "F" with a pink rectangle was flower and label "B" with a green rectangle was bud. Original image of flower and bud was showed in Fig. 6b, where four buds and three flowers were correctly labeled by orange rectangles and red rectangles, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

review & editing. **Fuxi Shi:** Investigation, Methodology. **Jaspreet Dhupia:** Conceptualization, Methodology, Writing – review & editing. **Rui Li:** Methodology, Writing – review & editing. **Yongjie Cui:** Investigation, Methodology.

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 partially supported by the Key R&D Program of Zhejiang (2022C02055); National Natural Science Foundation of China (32171897); Youth Science and Technology Nova Program in Shaanxi Province of China (2021KJXX-94); Science and Technology Promotion Program of Northwest A&F University (TGZX2021-29); China Postdoctoral Science Foundation funded project (2019M663832); Recruitment Program of High-End Foreign Experts of the State Administration of Foreign Experts Affairs, Ministry of Science and Technology, China (G20200027075).

References

Ahn, H.S., Dayoub, F., Popović, M., MacDonald, B.A., Siegwart, R., Sa, I., 2018. An overview of perception methods for horticultural robots: From pollination to harvest. arXiv Prepr. arXiv: 1807.03124.

Barnett, J., Seabright, M., Williams, H., Nejati, M., Scarfe, A., Bell, J., Jones, M., Martinson, P., Schare, P., 2017. Robotic pollination - targeting kiwifruit flowers for commercial application. Int. Tri-Conference Precis. Agric.

Bochkovskiy, A., Wang, C., Liao, H.M., 2020. YOLOv4: Optimal speed and accuracy of object detection. arXiv Prepr. arXiv:2004.10934.

Dias, P.A., Tabb, A., Medeiros, H., 2018a. Multispecies fruit flower detection using a refined semantic segmentation network. IEEE Robot. Autom. Lett. 3 (4), 3003–3010. https://doi.org/10.1109/LRA.2018.2849498.

Dias, P.A., Tabb, A., Medeiros, H., 2018b. Apple flower detection using deep convolutional networks. Comput. Ind. 99, 17–28. https://doi.org/10.1016/j.compind.2018.03.010.

Duke, M., Barnett, J., Bell, J., Jones, M.H., Martinsen, P., McDonald, B., Nejati, M., Scarfe, A., Schaare, P., Seabright, M., Williams, H., Ahn, J.L., 2017. Automated pollination of kiwifruit flowers. In: 7th Asian-Australasian Conf. Precis. Agric. 1–5. https://doi.org/10.5281/ZENODO.895619.

Farjon, G., Krikeb, O., Hillel, A.B., Alchanatis, V., 2020. Detection and counting of flowers on apple trees for better chemical thinning decisions. Precis. Agric. 21 (3), 503–521. https://doi.org/10.1007/s11119-019-09679-1.

Fu, L., Feng, Y., Wu, J., Liu, Z., Gao, F., Majeed, Y., Al-Mallahi, A., Zhang, Q., Li, R., Cui, Y., 2021. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model. Precis. Agric. 22 (3), 754–776. https://doi.org/10.1007/s11119-020-09754-y.

Fu, L., Gao, F., Wu, J., Li, R., Karkee, M., Zhang, Q., 2020. Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review. Comput. Electron. Agric. 177, 105687. https://doi.org/10.1016/j.compag.2020.105687.

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.

Gianni, T., Vania, M., 2018. Artificial pollination in kiwifruit and olive trees. Pollinat. Plants 10.5772/in. https://doi.org/10.5772/intechopen.74831.

Gonzalez, M.V., Coque, M., Herrero, M., 1995. Stigmatic receptivity limits the effective pollination period in kiwifruit. J. Am. Soc. Hortic. Sci. 120, 199–202. https://doi.org/10.21273/jashs.120.2.199.

Hong, S.W., Choi, L., 2012. Automatic recognition of flowers through color and edge based contour detection. In: 2012 3rd Int. Conf. Image Process. Theory, Tools Appl. IPTA 2012 141–146. https://doi.org/10.1109/IPTA.2012.6469535.

Jiang, M., Rao, Y., Zhang, J., Shen, Y., 2020. Automatic behavior recognition of group-housed goats using deep learning. Comput. Electron. Agric. 177, 105706. https://doi.org/10.1016/j.compag.2020.105706.

Lim, J.Y., Ahn, H.S., Nejati, M., Bell, J., Williams, H., MacDonald, B.A., 2020. Deep neural network based real-time kiwi fruit flower detection in an orchard environment. arXiv Prepr. arXiv: 2006.04343.

Liu, G., Nouaze, J.C., Mbouembe, P.L.T., Kim, J.H., 2020. YOLO-tomato: a robust algorithm for tomato detection based on YOLOv3. Sensors (Switzerland) 20, 2145. https://doi.org/10.3390/s20072145.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016. SSD: Single shot multibox detector. Lect. Notes Comput. Sci. 9905 LNCS, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.

Mcpherson, H.G., Richardson, A.C., Snelgar, W.P., Patterson, K.J., Currie, M.B., 2001. Flower quality and fruit size in kiwifruit (Actinidia deliciosa). New Zeal. J. Crop Hortic. Sci. 29 (2), 93–101. https://doi.org/10.1080/01140671.2001.9514167.

Ogden, S.S., Guo, T., 2019. Characterizing the deep neural networks inference performance of mobile applications. arXiv Prepr. arXiv: 1909.04783.

Pornpanomchai, C., Sakunreraratsame, P., Wongsasirinart, R., Youngtavichavhart, N., 2020. Herb flower recognition system (HFRS). ICEIE 2010 - 2010 Int. Conf. Electron. Inf. Eng. Proc. 1, 123–127. https://doi.org/10.1109/ICEIE.2010.5559906.

Ohi, N., Lassak, K., Watson, R., Strader, J., Du, Y., Yang, C., Hedrick, G., Nguyen, J., Harper, S., Reynolds, D., Kilic, C., Hikes, J., Mills, S., Castle, C., Buzzo, B., Waterland, N., Gross, J., Park, Y.L., Li, X., Gu, Y., 2018. Design of an autonomous precision pollination robot. arXiv Prepr. arXiv: 1809.10080.

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

Soviany, P., Ionescu, R.T., 2018. Optimizing the trade-off between single-stage and two-stage deep object detectors using image difficulty prediction. In: Proc. - 2018 20th Int. Symp. Symb. Numer. Algorithms Sci. Comput. SYNASC 2018 209–214. https://doi.org/10.1109/SYNASC.2018.00041.

Tacconi, G., Michelotti, V., Cacioppo, O., Vittone, G., 2016. Kiwifruit pollination: the interaction between pollen quality, pollination systems and flowering stage. J. Berry Res. 6, 417–426. https://doi.org/10.3233/JBR-160138.

Tian, M., Chen, H., Wang, Q., 2019. Detection and recognition of flower image based on SSD network in video stream. J. Phys. Conf. Ser. 1237 (3), 032045. https://doi.org/10.1088/1742-6596/1237/3/032045.

Tian, Y., Yang, G., Wang, Z., Li, E., Liang, Z., 2020. Instance segmentation of apple flowers using the improved mask R-CNN model. Biosyst. Eng. 193, 264–278. https://doi.org/10.1016/j.biosystemseng.2020.03.008.

Wang, C., Liao, H.M., Wu, Y., Chen, P., 2019. CSPNet: A new backbone that can enhance learning capability of CNN. arXiv Prepr. arXiv: 1911.11929v1.

Wang, X.(., Tang, J., Whitty, M., 2020. Side-view apple flower mapping using edge-based fully convolutional networks for variable rate chemical thinning. Comput. Electron. Agric. 178, 105673. https://doi.org/10.1016/j.compag.2020.105673.

Williams, H., Nejati, M., Hussein, S., Penhall, N., Lim, J.Y., Jones, M.H., Bell, J., Ahn, H. S., Bradley, S., Schaare, P., Martinsen, P., Alomar, M., Patel, P., Seabright, M., Duke, M., Scarfe, A., MacDonald, B., 2020. Autonomous pollination of individual kiwifruit flowers: Toward a robotic kiwifruit pollinator. J. F. Robot. 37 (2), 246–262. https://doi.org/10.1002/rob.21861.

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.

Zhang, Q., Karkee, M., Tabb, A., 2019. The use of agricultural robots in orchard management. arXiv 1–22. https://doi.org/10.19103/as.2019.0056.14.