A Real-time Kiwifruit Detection Based on Improved YOLOv7

Yi Xia, Minh Nguyen, Wei Qi Yan
Auckland University of Technology, 1010 Auckland, New Zealand

Abstract. In New Zealand (NZ), agriculture is an essential industry, Kiwifruits contribute significantly to the country's overall exports. Traditionally Kiwifruits require manually picking up and heavily relies on human resources, which result in Kiwifruit yields often being affected by human labours. With the rapid development of deep learning in agriculture, agricultural automation has become an efftive way for the industry. Accurate and fast Kiwifruit detection can accelerate the process in the industry. In this paper, we propose an improved Kiwifruit detection model based on YOLOv7. We collected digital images from natural Kiwifruit orchards and produced a manually labelled, data-augumented Kiwifruit image dataset. We add the attention module to YOLOv7 and increase the weight of visual features while suppressing the weight of invalid features. The results show that our proposed method has higher detection accuracy than the original YOLOv7 model, while the detection speed is sufficient for real-time usage. The results of our experiments provide a technical reference for automated picking in modern Kiwifruit supply chain.

Keywords: Deep learning · YOLOv7 · Attention mechanism · Real-time detection · CBAM.

1 Introduction

Kiwifruits appeare in worldwide markets and have become one of the most iconic namecards of New Zealand [5]. However, the rapidly growing industry has also brought significant challenges, including labor shortages that have led to industry losses. Therefore, developing an efficient supply chain related to picking, sorting, cleaning, and packaging is an effective way to improve efficiency in the current Kiwifruit industry. We are use of the state-of-the-art artificial intelligence, in particular, deep learning, to increase the Kiwifruit yield estimation, improve picking efficiency and reduce the costs of human labors [18].

Fast and accurate models are the foundation of Kiwifruit counting [1]. Traditionally, fruit detection algorithms [9] [19] [36] extracted key feature parameters such as color and shape of visual objects through digital image processing. The image segmentation algorithms are harnessed in order to detect the visual objects. However, the conventional algorithms are less robust. The factors such as lighting conditions and fruit location affected the results of the model detection. Therefore, the practicality of fruit yelds estimation in orchards under natural conditions is low.

However, recent advances in deep learning [10] and computer vision [34] have combined more fields with artificial intelligence and computer vision [22] [25]. Mainly, digital image processing based on deep learning has been extensively applied to modern agriculture, such as plant pest control, fruit ripeness detection [33], and fruit freshness grading [6].

Visual object detection has been conducted based on deep learning [20] [21]. Wang and Yan proved that the one-stage YOLOv5 algorithm outperformed the two-stage Faster R-CNN algorithm in the leaf detection task through comparative experiments, especially in the speed of object detection [30]. Bazame et al. proposed the best-performing YOLOv4-based algorithm for detecting and classifying coffee beans on tree branches by comparing YOLOv3, YOLO4 and YOLOv4-tiny algorithms [2]. The mAP of the model is 81%. Lawal propounded an accurate and fast algorithm for fruit detection-YOLOMuskmelon, which combines the ReLU-activated ResNet-43 Backbone with residual block alignment, SPP, CIoU loss, FPN, and DIoU-NMS to improve detection performance, the average accuracy of this model is 89.6% [13]. Liu et al. put forward SE-Mask R-CNN algorithm for detecting apples in complex environments [15]. The method improves the resource allocation of the model to the effective feature maps by adding SENet to the backbone network. Liu et al. offered TomatoDet [14], an anchorless frame algorithm for tomato detection, which was applied to solve the detection of tomatoes under complex environmental conditions, such as uneven lighting, leaf or branch occlusion, and overlap between fruits. The algorithm incorporates an attention mechanism in CenterNet and introduces a circular representation to optimize the detector. The average precision of this algorithm is 98.16%. Jilbert et al. proffered an algorithm based on the YOLOv5 model for detecting coconut fruits using UAVs [12]. The accuracy of this algorithm is 88.4%. The mAP of the improved model is improved by 3.5%, and the model size is compressed by 62.77%. Although a plethora of studies have obtained better model performance, with the rapid development of deep learning, the better performing YOLOv7 [28] algorithm can obtain more accurate results under faster conditions.

Kiwifruit detection based on deep learning algorithm proposed in this paper is accurate, fast, and is able to be generalized to accommodate the complex conditions in natural orchards. The algorithm for visual object detection is based on the YOLOv7 model, the one-stage object detection algorithm does not need to generate candidate frames. This algorithm directly converts the problem of visual object localization into a regression problem. Therefore, YOLO model is superior to computing speed, this fast detection capability can be better applied to visual object detection.

In this paper, we propose an improved YOLOv7 algorithm combined with the Convolutional Block Attention Module (CBAM) [32]. This method is able to improve the detection accuracy of small, overlapping, and multiple Kiwifruits. A CBAM module is added to the backbone of YOLOv7 network and assign weights to channel features and spatial features in the feature map to increase the model sensitivity while reducing attention to invalid features, thereby improving the model's detection of Kiwifruit in orchards. In this paper, we also have other contributions: (1) We created a new Kiwifruit dataset, (2) conducted data augmentation, (3) verified our YOLOv7 model by using ablation experiments.

This paper is organized as follows: In Section 2, we introduce the related work, our method is detailed in Section 3, our results are demonstrated in Section 4. In Section 5, we conclude this paper and envision our future work.

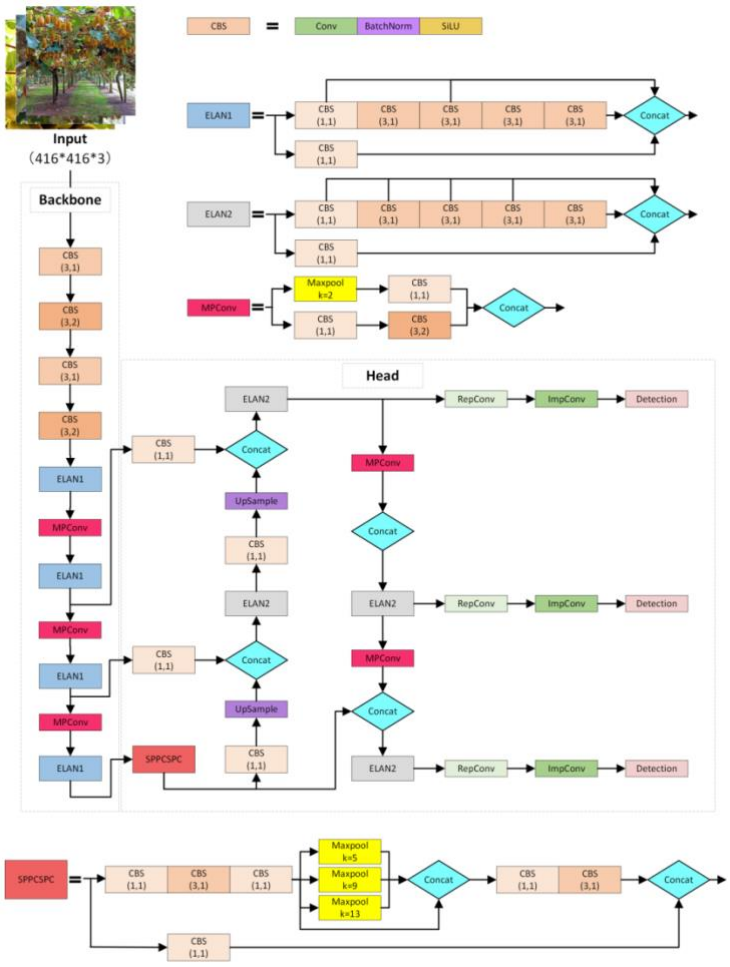


Fig. 1. YOLOv7 neural network architecture.

2 Related Work

2.1 YOLOv7

Visual object detection is a computer vision problem that locates and labels visual objects by drawing a bounding box around the object, and determines the class label to which the given box belongs to. Visual object detection is an important research topic in computer vision and broadly employed to the areas such as face recognition, car plate

number recognition, intelligent transportation and autonomous vehicles. The YOLO family has witnessed visual object detection in the era of deep learning. Since the publication of YOLOv1 [23] in 2015, YOLO has been updated iteratively. The lightweight and high accuracy of YOLO models have set the benchmark for the state-of-the-art methods of visual object detection.

YOLOv4 [3], Scaled-YOLOv4 [27], and YOLOR [29] were proposed in 2020 and 2021. The latest object detection model - YOLOv7 [28] was proffered in 2022. It outperforms most of well-known object detectors such as R-CNN [8], YOLOv4 [3], YOLOR [29], YOLOv5, YOLOX [7], PPYOLO [17], and DETR [4] etc. YOLOv7 reduces about 40% of the number of parameters and 50% of the computational costs of real-time object detection. It is split into two main areas of optimization: Model architecture optimization and training process optimization. YOLOv4 improved the accuracy at the cost of training but did not increase inference cost [28]. However, YOLOv7 takes use of a re-parameterized approach to replace the original modules. It adopts dynamic label assignment, which has the effect of assigning labels to output layers more efficiently [37]. The YOLOv7 structure is similar to YOLOv5, the main improvement is the replacement of internal components of the network structure.

Fig.1 shows the overall structure of YOLOv7. In Fig.1, we see that YOLOv7 consists of three components: Input, backbone, and head. The backbone layer extracts feature maps, the head layer is employed for prediction. As shown in Fig.1, firstly the processed image is input into the backbone network in YOLOv7, a feature map is output with three layers of different sizes through the head layer network, and finally the prediction result is exported through Rep convolution and Imp convolution.

2.2 Attention Mechanism

Attention mechanisms [26] have made significant achievements in image processing in recent years. The essence of attention mechanism is to detect the information which is interested and suppressed the useless information. Three main attention mechanisms are based on how weights are applied to feature spatial and channel: Spatial attention mechanisms [38], channel attention mechanisms [11], mixed spatial and channel attention mechanisms [24]. The attention mechanisms have different effects on different computer vision tasks.

Squeeze-and-Excitation Network (SENet) [11] is a channel-based attention model that models the importance of each feature channel and then enhances or suppresses different channels for different tasks. A bypass branch is branched out after the regular convolution operation. Firstly, the squeeze operation is performed, compressing the spatial dimension with features so that each 2D feature map becomes an actual number, and the number of feature channels remains unchanged. Then the excitation operation generates weights for each feature channel, which is applied to show the correlation between the modelled feature channels. Once the model gets the weights for each feature channel, it can show the importance of different channels by applying that weight to each original feature channel. The model achieves a more significant performance improvement with a minor increase in computation. Efficient Channel Attention Net (ECA-Net) [31] improved on the SENet module. The module indicates a

way for local cross-channel interaction without dimensionality reduction, which effectively avoids the effect of dimensionality reduction on the learning effect of channel attention. The experimental results show that ECA-Net has low complexity while obtaining excellent performance.

Convolutional Block Attention Module (CBAM) [27] is a simple and effective attention module for feedforward convolutional neural networks that connects the spatial attention module after the channel attention module. The CBAM structure is shown in Fig.2. The focus of spatial attention is on the position of objects in the image, while channel attention focuses on the objects in the image. Instead of using a single maximum pooling or average pooling, the attention module harnesses the summation or stacking of the maximum and average pooling.

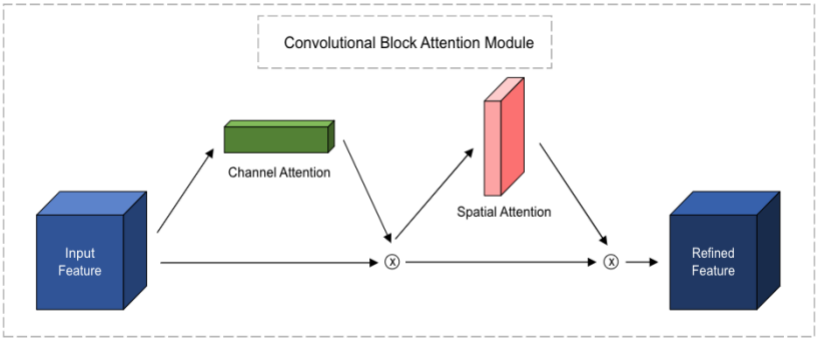


Fig. 2. The main structure of convolutional block attention module.

The channel attention module structure is shown in Fig.3. The input feature maps are subjected to global max pooling and global average pooling based on width and height respectively to obtain two 1×1×C feature maps, which are then fed into a two-layer MLP with the number of neurons in the first layer as C/r, where r is the reduction rate and the activation function as ReLU, the number of neurons in the second layer is C. This two-layer neural network is shared. The channel attention feature and the input feature map are multiplied elementwise to generate the final channel attention feature. Finally, the input features are multiplied elementwise to generate the input features required by the spatial attention module.

The spatial attention module structure is shown in Fig.4. The feature map output from the channel attention module is employed as the input feature map. The module firstly conducts channel-based global max pooling and global average pooling to obtain two feature maps. A concatenation operation (channel splicing) is undergone on the two feature maps based on the channels. A convolution is then performed to reduce the dimensionality to one channel. Then it goes through a sigmoid to generate a spatial attention feature.

Finally, the feature is multiplied by the input feature of the module to obtain the final generated feature. The module is shown to provide improvements in both classification and detection performance.

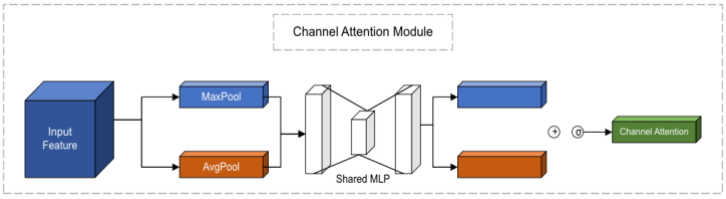


Fig. 3. The structure of channel attention module.

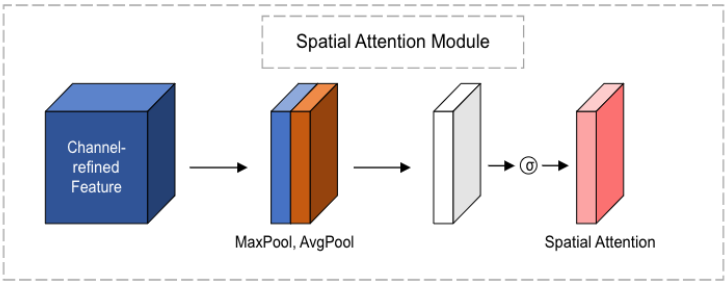


Fig. 4. The structure of spatial attention module.

Fig. 5. The image dataset labeled on Roboflow.

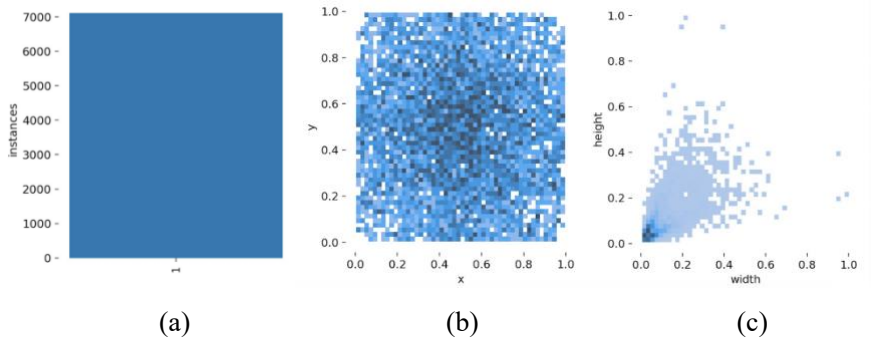
3 Our Methods

3.1 Dataset

Data Collection. There is not publicly labelled dataset available for Kiwifruits. In this paper, two methods were employed to collect the Kiwifruit dataset. Firstly, we collected digital images of natural Kiwifruits from Google Images. Secondly, we retrieved and downloaded videos of Kiwifruits from YouTube, we split the video into frames. We collected 117 images of Kiwifruits.

Data Preprocessing. We manually took out the duplicate images and images without Kiwifruits to reduce redundancy with model training. YOLOv7 provides Roboflow tool, which can label the images and automatically export the custom dataset. Therefore, we uploaded the filtered images to Roboflow for manually labelling, there are 7,114 labels in this dataset. The labelling results are shown in Fig.5.

In order to reduce the training time and improve the model performance, we collected the images and resized them to the resolution 416×416. We rotated the images in the training dataset clockwise and counterclockwise 90 degrees. The data augmentation increases the amount of data in the training dataset, maintains data diversity, and alters the distribution direction of Kiwifruits in the original images to improve the generalization of the trained model. After completed the data augmentation, we randomly split the dataset into a training set, a valid set, and a test set according to 7:2:1. The training dataset was finally increased to 289 images, the image augmentation was confirmed to be correct by manual inspection. Fig.6 is an attribute visualization result of the augmented dataset. The number of labels in the dataset is shown in Fig.6 (a), the location of the labels in the dataset is shown in Fig.6 (b), the width and height of the labels in the dataset are shown in Fig.6 (c).



(a) (b) (c)

Fig. 6. (a) The number and the class of labels in the dataset. (b) The location of the labels in the images of the dataset. (c)The size of the labels in the dataset.

3.2 Modelling

In this paper, we were use of a manually collected, preprocessed, and labelled Kiwifruit image samples. In this paper, a CBAM model was added to the front of the backbone

in YOLOv7 net to deal with the dense nature of objects, the high overlap rate and the small size of the objects in the Kiwifruit videos. The method improves the overall accuracy of the object detection model by integrating CBAM and YOLOv7 together to assign weights of channel features and spatial features of visual objects in the feature map, increase the attention to detect visual objects and suppresse attention to non-objects. The structure of the improved model is shown in Fig.7.

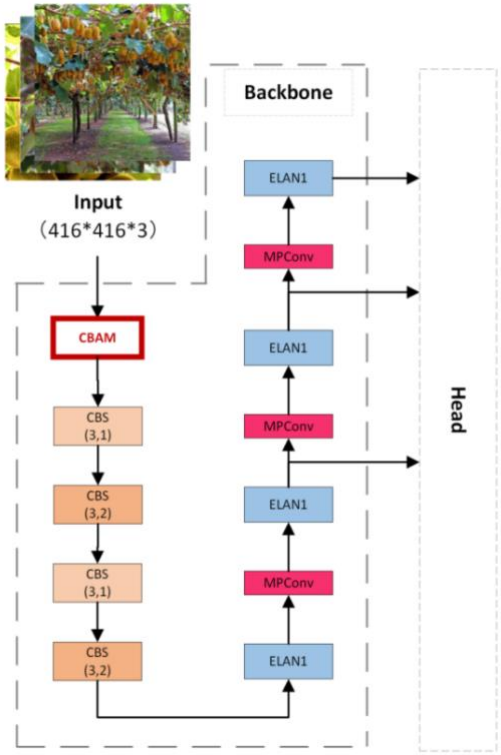


Fig. 7. The location of the CBAM in the improved model.

As shown in Fig.7, a pre-processed image of size 416×416×3 is input into the backbone. The output feature map is firstly processed through the global max pooling and global average pooling in CBAM, then through a multilayer perceptron with shared weights, which conducts an addition operation based on the two feature maps through the sigmoid activation function. After the channel attention module is completed, the feature maps are input into the spatial attention module. The two feature maps are combined by using global max pooling and global average pooling. Then the number of channels is reduced to 7×7 convolution [29]. The sigmoid activation function obtains the spatial attention feature maps. Finally, the outputs of channel attention module and the spatial attention module are multiplied to obtain the output feature map of CBAM. The feature maps from the CBAM are fed into the CBS module in the original

Backbone, the final predictions are output to implement the object detection by the model.

4 Our Results

4.1 Evaluation Metrics for Kiwifruit Detection

In this paper, we are use of precision (P), recall (R), and mean average precision (mAP) as the evaluation metrics for the Kiwifruit detection algorithm. The experimental results encapsulated four outcomes, True Positive (TP) refers to manually marked Kiwifruits being detected correctly, False Positive (FP) means the object that was incorrectly detected as a Kiwifruit, True Negative (TN) is to the negative samples with negative system prediction, and False Negative (FN) reflects to Kiwifruits that are missed. Two mAP indicators are employed in this paper, mAP@0.5 and mAP@0.95. mAP@0.5 refers to the average precision of all images in each class if assigned IoU to 0.5, and then all classes are averaged. mAP@0.95 indicates to the average mAP over different IoU thresholds (from 0.5 to 0.95 with a step size of 0.05). Intersect over Union (IoU) reflects to the proportion of intersection and concatenation of the object prediction box and the true box.

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} \int_0^1 PdR \tag{3}$$

4.2 Experimental Environment and Training Parameters

In this paper, the experiments were conducted in Google Collaboratory platform. We were use of Python 3.7 (version 3.7.14), Pytorch (version 1.12.1), and CUDA (version 11.2) for the YOLOv7 training. The Tesla T4 (16G) GPU was utilized for the detection model training. The size of all images used for training in this experiment is 416×416, batch-size is 16 and epochs are 150 times.

4.3 Experimental Results and Comparisons

The results of the improved model for the detection of Kiwifruits are shown in Fig.8, which shows that the model is better for detecting high density, overlapping objects and small objects.

In order to verify the effectiveness of the improved YOLOv7 model, we compared the original YOLOv7 model. We also compared the more popular YOLOv5s model. In

addition, we compared six improved YOLOv5s models. We took use of two models to insert three attention mechanisms, SE, ECA and CBAM, into different positions of the YOLOv5s model. In the first method, we inserted these three attention mechanisms in front of the SPFF module of Backbone in YOLOv5 model. We named the improved algorithm YOLOv5_SE, YOLOv5_ECA, and YOLOv5_CBAM. In the second method, we replaced these three attention mechanisms with the C3 module within Backbone in the original model and named it YOLOv5_C3SE, YOLOv5_C3ECA, and YOLOv5_C3CBAM. Fig.9 (a) shows the first method of inserting the attention module. Fig.9 (b) indicates the second method. The training parameters of the comparison experimental models are consistent. The results of the different models are compared in Table 1.

Fig. 8. Test results using the improved model.

In Table 1, YOLOv7_CBAM model attained the best performance with the same experimental parameters. YOLOv7 and YOLOv7_CBAM models outperformed YOLOv5s and six attention mechanisms addition models based on YOLOv5s in the Kiwifruit detection experiments. The improved YOLOv7 model increased the precision by 1.1%, recall by 3.8%, the mAP@0 .5 value by 0.8% and the mAP@0.95 value by 0.5% in comparison to the original YOLOv7 model. YOLOv5s model with the C3 module, replaced by the attention mechanism, has the smallest size due to the small number of model layers and the small number of model parameters. We also see from Table 1 that the approach of adding the attention mechanism to the YOLO algorithm produced better performance than replacing the original CBS module. From the comparison, we see that the addition of CBAM can help YOLOv7 model to improve the performance of object detection, which affirms the effectiveness of the approach proposed in this paper.

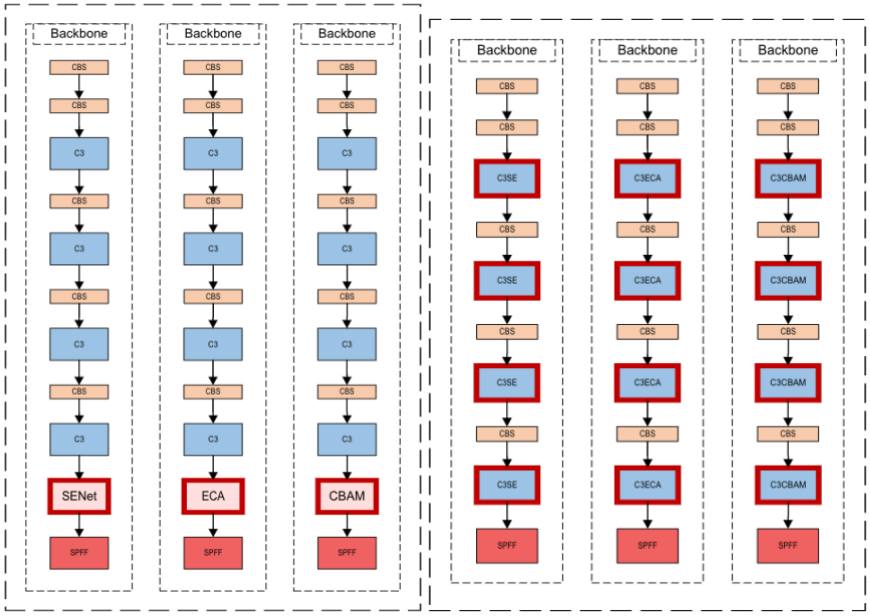


Fig. 9. (a) The method for inserting three attention mechanisms ahead of the SPFF module in the Backbone of the YOLOv5s algorithm. (b) The method for replacement of the C3 module in Backbone of YOLOv5s algorithm with three attention mechanisms.

Table 1. Comparison of the improved model with other models.

Model	Precision	Recall	mAP@0.5	mAP@0.95	Model size
YOlOv5s	92.2%	86.0%	94.1%	63.1%	14.3MB
YOLOv5_SE	91.8%	87.7%	94.3%	61.2%	15.4MB
YOLOv5_C3SE	90.6%	85.6%	92.4%	58.6%	**13.1MB**
YOLOv5_ECA	91.4%	86.1%	94.2%	60.0%	14.3MB
YOLOv5_C3ECA	90.9%	85.1%	92.5%	60.6%	**13.1MB**
YOLOv5_CBAM	**94.1%**	85.1%	93.4%	60.5%	14.5MB
YOLOv5_C3CBAM	89.8%	85.3%	93.0%	60.1%	**13.1MB**
YOLOv7	92.1%	88.1%	95.3%	66.7%	74.8MB
YOLOv7_CBAM	93.1%	**91.9%**	**96.1%**	**67.2%**	74.8MB

5 Conclusion

Fruit detection and yield estimation have a significant impact on agricultural automation. Fastly and highly accurate fruit detection algorithms can aid harvesting

robots in performing their picking tasks efficiently. In this paper, we presented a YOLOv7-based method for detecting Kiwifruits in orchards, which combines CBAM mechanism with the original YOLOv7 model. The method improved the current YOLOv7 model, the best-performed model in the YOLO family. The addition of the CBAM module assists the model in increasing attention to the object and reducing attention to useless features. Our experimental results show that the improved YOLOv7 model performed better than the original YOLOv7 in Kiwifruit detection. The results showcase the effectiveness of the improved algorithm. The future work in this research project is to reduce the model size and propose a highly accurate and lightweight improved Kiwifruit detection model. Meanwhile, tracking and counting of Kiwifruits in orchards will be achieved by combining multiobject tracking together to provide theoretical and technical references for further applications in practical scenarios [1][16] [35].

References

1. An, N., Yan, W.: Multitarget tracking using Siamese neural networks. ACM Transactions on Multimedia Computing, Communications and Applications (2021).
2. Bazame, H., Molin, J., Althoff, D., Martello, M.: Detection, classification, and mapping of coffee fruits during harvest with computer vision. Computers and Electronics in Agriculture, 183, 106066 (2021).
3. Bochkovskiy, A., Wang, C., Liao, H.: YOLOv4: Optimal speed and accuracy of object detection, https://arxiv.org/abs/2004.10934.
4. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with Transformers. Computer Vision – ECCV 2020, pp. 213-229. Springer (2020).
5. Ferguson, A.: 1904—the year that Kiwifruit (Actinidia deliciosa) came to New Zealand. New Zealand Journal of Crop and Horticultural Science, 32, 3-27 (2004).
6. Fu, Y., Nguyen, M., Yan, W.Q.: Grading methods for fruit freshness based on deep learning. SN Computer Science. 3, (2022).
7. Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J.: YOLOX: Exceeding YOLO series in 2021, https://arxiv.org/abs/2107.08430.
8. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition, pp. 580-587 (2014).
9. Gongal, A., Karkee, M., Amatya, S.: Apple fruit size estimation using a 3D machine vision system. Information Processing in Agriculture, 5, 498-503 (2018).
10. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, (2016).
11. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. IEEE Conference on Computer Vision and Pattern Recognition pp. 7132-7141 (2018).
12. Jilbert M., N., Jennifer C., D.: On-tree mature coconut fruit detection based on deep learning using UAV images. IEEE International Conference on Cybernetics and Computational Intelligence, pp. 494-499 (2022).
13. Lawal, O.: YOLOMuskmelon: Quest for fruit detection speed and accuracy using deep learning. IEEE Access, 9, 15221-15227 (2021).
14. Liu, G., Hou, Z., Liu, H., Liu, J., Zhao, W., Li, K.: TomatoDet: Anchor-free detector for tomato detection. Frontiers in Plant Science. 13, (2022).

15. Liu, Y., Yang, G., Huang, Y., Yin, Y.: SE-Mask R-CNN: An improved Mask R-CNN for apple detection and segmentation. Journal of Intelligent Fuzzy Systems, 41, 6715-6725 (2021).
16. Liu, Z., Yan, W., Yang, B.: Image denoising based on a CNN model. IEEE ICCAR (2018).
17. Long, X., Deng, K., Wang, G., Zhang, Y., Dang, Q., Gao, Y., Shen, H., Ren, J., Han, S., Ding, E., Wen, S.: PP-YOLO: An effective and efficient implementation of object detector, https://arxiv.org/abs/2007.12099.
18. Massah, J., Asefpour Vakilian, K., Shabanian, M., Shariatmadari, S.: Design, development, and performance evaluation of a robot for yield estimation of Kiwifruit. Computers and Electronics in Agriculture, 185, 106132 (2021).
19. Olaniyi, E., Oyedotun, O., Adnan, K.: Intelligent grading system for banana fruit using neural network arbitration. Journal of Food Process Engineering, 40, e12335 (2016).
20. Pan, C., Liu, J., Yan, W., et ak.: Salient object detection based on visual perceptual saturation and two-stream hybrid networks. IEEE Transactions on Image Processing (2021).
21. Pan, C., Yan, W.: A learning-based positive feedback in salient object detection. IEEE IVCNZ (2018).
22. Pan, C., Yan, W.: Object detection based on saturation of visual perception. Multimedia Tools and Applications, 79 (27-28), 19925-19944 (2020).
23. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-time object detection. IEEE CVPR. pp. 779-788 (2016).
24. Shan, T., Yan, J.: SCA-Net: A spatial and channel attention network for medical image segmentation. IEEE Access. 9, 160926-160937 (2021).
25. Shen, D., Xin, C., Nguyen, M., Yan, W.: Flame detection using deep learning. IEEE ICCAR (2018).
26. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems 30 (2017).
27. Wang, C., Bochkovskiy, A., Liao, H.: Scaled-YOLOv4: Scaling cross stage partial network, https://arxiv.org/abs/2011.08036.
28. Wang, C., Bochkovskiy, A., Liao, H.: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, https://arxiv.org/abs/2207.02696.
29. Wang, C., Yeh, I., Liao, H.: You Only Learn One Representation: Unified network for multiple tasks, https://arxiv.org/abs/2105.04206.
30. Wang, L., Yan, W.Q.: Tree leaves detection based on deep learning. International Symposium on Geometry and Vision, pp. 25-38 (2021).
31. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: ECA-Net: Efficient channel attention for deep convolutional neural networks, https://arxiv.org/abs/1910.03151.
32. Woo, S., Park, J., Lee, J., Kweon, I.: CBAM: Convolutional block attention module. ECCV, pp. 3-19 (2018).
33. Xiao. B., Nguyen, M., Yan, W.Q.: Apple ripeness identification using deep learning. International Symposium on Geometry and Vision, pp. 53-67 (2021).
34. Yan, W.: Computational Methods for Deep Learning. Springer (2021).
35. Yan, W.: Introduction to Intelligent Surveillance. Springer (2019).
36. Zhao, K., Nguyen, M. Yan, W.: Fruit detection from digital images using CenterNet. International Symposium on Geometry and Vision (2021).
37. Zheng, K., Yan, W., nand, P.: Video dynamics detection using deep neural networks. IEEE Transactions on Emerging Topics in Computational Intelligence (2017).
38. Zhu, X., Cheng, D., Zhang, Z., Lin, S., Dai, J.: An empirical study of spatial attention mechanisms in deep networks. IEEE CVPR, pp. 6688-6697 (2019).