

*Scientific Paper*Doi: <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v43n2e20220180/2023>

## ASSESSING PINEAPPLE MATURITY IN COMPLEX SCENARIOS USING AN IMPROVED RETINANET ALGORITHM

**Yan Chen<sup>1</sup>, Lulu Zheng<sup>1\*</sup>, Hongxing Peng<sup>1</sup>**<sup>1\*</sup>Corresponding author. South China Agricultural University/Guangzhou, China.E-mail: [scauzhenglulu@163.com](mailto:scauzhenglulu@163.com) | ORCID ID: <http://orcid.org/0000-0002-7777-9716>

### KEYWORDS

artificial intelligence, deep learning, object detection network, ECA attention mechanism.

### ABSTRACT

In China, low levels of accuracy in predicting when pineapple crops will reach maturity can result from environmental variation such as light changes, fruit overlap, and shading. Therefore, this study proposed the use of an improved RetinaNet algorithm (ECA-Retinanet) based on the ECA attention mechanism. The ECA attention mechanism was embedded into the classification subnet of RetinaNet to improve accuracy in detecting different levels of maturity in pineapples. A new pineapple dataset was collected comprising four different growth stages under mild and severe complex scenarios. The experimental results have shown that the mAP (Mean Average Precision) and F<sub>1</sub> score (Balanced Score) of the ECA-Retinanet model were 97.69%, 94.75%, 93.2%, and 90% for identification in mild and severe complex scenarios. These values are 0.42%, 2%, 1.78%, and 1.5% higher than the original RetinaNet model which exceeds those of the six existing state-of-the-art detection models. The results have indicated that the proposed algorithm could be used for accurate identification of pineapple fruit and can detect fruit maturity using ground color images in the natural environment. The study findings provide a technical reference for automatic picking robots and early yield estimation.

### INTRODUCTION

In recent years, artificial intelligence technologies have been widely used in agriculture. Deep learning can solve various problems in precision agriculture with the development of various systems (Solemane et al., 2022). A powerful technical tool in artificial intelligence, computer vision (Wang et al., 2022) has provided a strong technical guarantee in the vision system of agricultural robots. Agricultural robots (Nguyen et al., 2021) can help farmers to solve farming, pesticide, and picking problems in an environmentally friendly, energy-saving, and cost-saving way to improve agricultural production efficiency and increase income. Among them, fruit detection is one of the important aspects. By accurately detecting fruit maturity, their harvesting time can be predicted to ensure effective management and increases in yield.

At present, pineapples are widely distributed in Brazil, Thailand, the United States, Mexico, the Philippines, and a range of other countries. As one of the major producers of pineapples, China has extensive areas under pineapple

cultivation in Guangdong, Guangxi, Fujian, Hainan, Yunnan, and Taiwan, which has created a high level of economic value in the market (Li et al., 2022). High-level orchard mechanical automation products are being developed to achieve accurate positioning and classification of pineapple maturity while improving crop quality and yields.

However, in natural orchard scenarios, detection of pineapple maturity can be influenced by a range of factors, such as being obscured by branches, leaves, and weeds, overlapping between pineapple fruits, light transitions that can severely affect imaging, and fruits that are similar in color to their background. Therefore, accurate measurement of pineapple maturity remains an important challenge to be addressed (Liu et al., 2022).

To date, there has been some progress locally and internationally in fruit detection research. Based on Philippine standards, Aguilar et al. (2021) proposed that a support vector machine and HSV color space could be used to automatically determine the level of maturity of pineapple fruit. However, this technique could not be applied successfully to detection in real scenarios and could not accurately determine the

<sup>1</sup> South China Agricultural University/Guangzhou, China.

Area Editor: Gizele Ingrid Gadotti

Received in: 9-29-2022

Accepted in: 3-21-2023



maturity of the pineapple fruit. This technique is the traditional object detection method used based on the color and texture of the images. In recent years, object detection (Liu et al., 2020, Wu et al., 2020) has become a key research focus in the field of artificial intelligence. As a powerful technical tool in artificial intelligence, deep learning has considerable advantages in the context of conducting object detection (Huang, 2020, Kong, 2021) tasks in fruit. Chen and Bu (2019) proposed a fruit identification algorithm based on multi-color features and texture features. However, in the action task, Tang (2020) suggested that the improved YOLOv3 could be used for real-time detection of passion fruit in real orchards, but the detection success in passion fruit with different levels of maturity was relatively poor. Xiong et al. (2020) proposed a multi-scale convolutional neural network Des-YOLOv3 algorithm to realize the identification and detection of ripe citrus in a complex environment at night, with an average accuracy (mAP) of 90.75% under the test set. Zhao et al. (2019) proposed the apple location method based on the YOLOv3 deep convolutional neural network. The mAP under the verification set was 87.71%, but it was difficult to achieve real-time detection using the network. The fruit detection system based on the Faster R-CNN model (Sa et al., 2016) has been used to detect sweet pepper, which has improved its level of the accuracy. However, if the detection speed is low the real-time monitoring cannot be realized. Mohd Basir Selvam et al. (2021) proposed the use of the YOLOv3 algorithm to detect mature palm oil clusters in real-time. However, this project has poor robustness and a relatively low level of accuracy in detecting palm oil clusters. Based on the Faster R-CNN model (Zhu et al., 2020), blueberry fruits with different levels of maturity could be accurately identified and classified, with a high level of

accuracy for factors such as background interference and fruit occlusion.

To address these above problems, the purpose of this study was to develop an improved RetinaNet algorithm. The ECA attention mechanism was embedded into the classification subnet to selectively increase the weight values of channels containing pineapple fruits to improve the detection accuracy for pineapples with different levels of maturity. A new pineapple dataset was collected with four pineapple maturity complexes in natural orchards from mild to severe.

## MATERIAL AND METHODS

### Dataset

In this study, the images of pineapples were collected at a natural orchard plantation in Danzhou City, Hainan Province, China. Filming was carried out using smartphones and a total of 6,000 images and 30 videos were collected. Data was collected from December 2021 to April 2022, with time slots of 9:00–12:00, 14:20–17:00, and 19:00–19:30 on four days of each month selected for filming. The video resolution was  $1920 \times 1080$  at 30 FPS. The video was pre-processed, and the video frames were extracted with using the FFMPEG tool. To prevent data redundancy, one video frame was extracted at 3 s intervals to obtain a pineapple object detection image dataset. A total of 2873 relatively representative images were selected as the experimental dataset in jpeg image format with a resolution of  $4032 \times 3024$  pixels. The pineapple dataset was captured under weather conditions with and without cloud cover. The lighting conditions included smooth light, backlighting, and metering, and complex backgrounds of overlapping branches, leaves, weeds, and fruits (Figure 1).

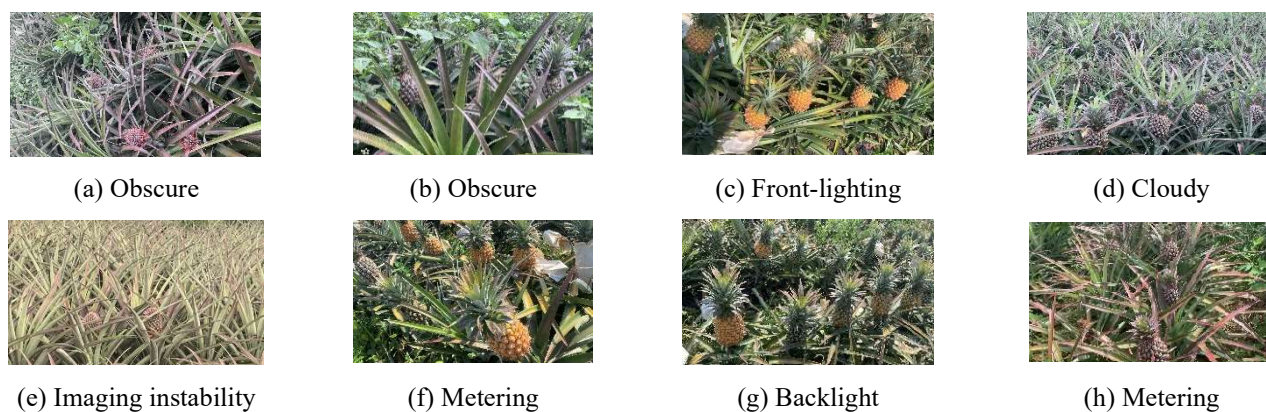


FIGURE 1. Examples of the pineapple images of the complex natural environment captured.

### Image pre-processing

The LabelImg tool (Darrenl, 2019) was used to manually annotate the level of maturity of the pineapples in images (Figure 2). The rectangular frame in standard time was used to fit the outline of pineapple fruits. The dataset format of PASCAL VOC 2007 was used in the experiment.

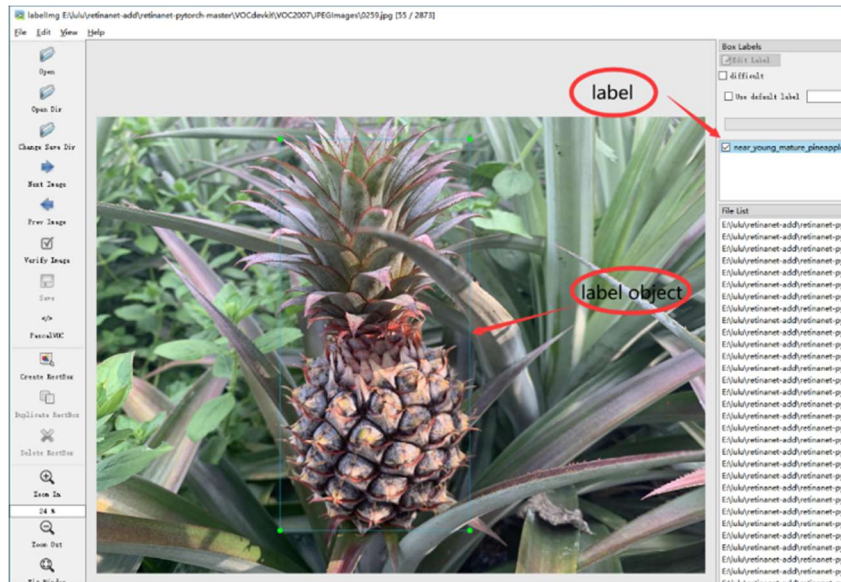


FIGURE 2. LabelImg interface while conducting annotation and marking the position of the pineapple in the image.

For ease of observation, these four different stages of pineapple maturity were indicated by four labels, based on the experience of the growing experts. The young\_mature\_pineapple label represents the first stage of *Bromelia* with red-purple flowers as shown in Figure 3 (a). The near\_young\_mature\_pineapple label represents the

second stage of *Bromelia* with red-purple flat fading flowers shown in Figure 3 (b). The near\_mature\_pineapple label represents the third stage of *Bromelia* as being is flat as shown in Figure 3 (c). The mature\_pineapple label represents the fourth stage of *Bromelia* with dark green to yellow fruit as shown in Figure 3 (d).



FIGURE 3. Classification sketch of pineapple images.

The LabelImg tool was used to annotate 2873 images from the dataset, and the total number of pineapple fruits was approximately 10,000 images, including 1156 images from the first stage, 2487 images from the second stage, 4585 images from the third stage, and 1,966 images from the fourth stage. The labeled pineapple dataset was randomly divided

into a training set and test set at a ratio of 9:1. A total of 2585 images in training sets and 288 images in test sets were obtained, among which the test set was divided into mildly and severe complex scenarios. Figure 4 shows the mild and severe complex scenarios for the pineapple images. This meets the requirements of the experimental data.

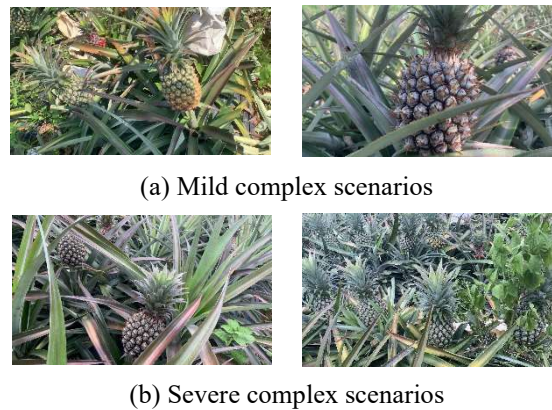


FIGURE 4. Two complex scenarios shown in the pineapple images.



### Improved RetinaNet based on ECA attention mechanism

The ECA-RetinaNet pineapple maturity identification network structure is shown in Figure 5. The ECA-RetinaNet model uses ResNet50 (He et al., 2016) as the backbone feature network to extract the backbone features. It removes the C3, C4,

and C5 feature layers to construct Feature Pyramid Net (FPN) (Lin et al., 2017). It then merges the multi-scale features to obtain the P3, P4, P5, P6, and P7 effective feature layers. The prediction results for the level of pineapple maturity can be obtained by transmitting these five effective feature layers to the classification and regression subnets.

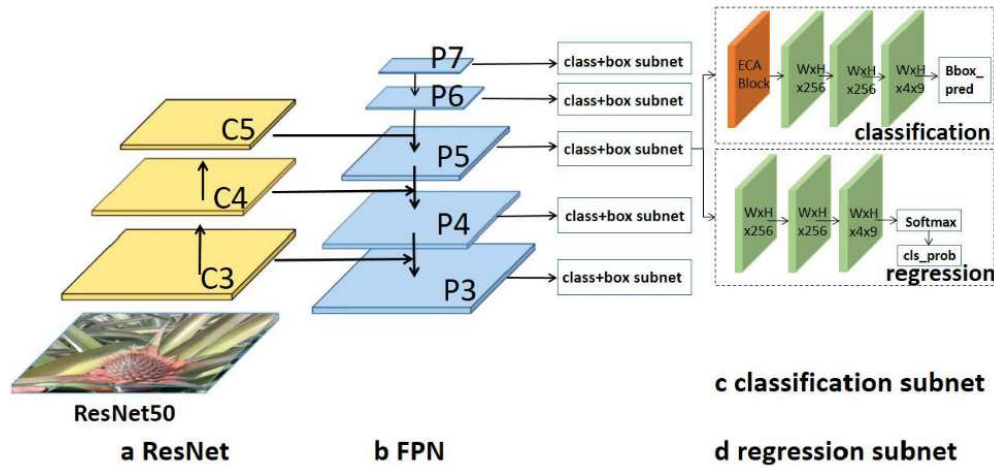


FIGURE 5. ECA-RetinaNet Architecture.

In Figure 5 a, the input pineapple RGB images were adjusted to  $600 \times 600$  size and ResNet50 was used for preliminary feature extraction from the pineapple image. ResNet50 introduces residual blocks to solve the problem of network degradation, such as when the network depth increases. The structure diagram of the residual network is shown in Figure 6, which introduces the data output of a specific layer of the leaning several layers directly to the input component of the later data layers by skipping multiple layers.

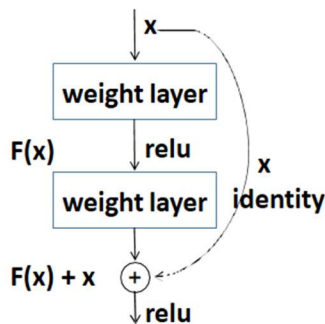


FIGURE 6. ResNet structure.

In Figure 5 b, ResNet50 was used to obtain three feature layers C3 ( $75 \times 75 \times 512$ ), C4 ( $38 \times 38 \times 1024$ ), and C5 ( $19 \times 19 \times 2048$ ) with different perceptual fields through the backbone feature extraction network. These three feature layers were passed through the feature pyramid to obtain five effective feature layers P3 ( $75 \times 75 \times 256$ ), P4 ( $38 \times 38 \times 256$ ), P5 ( $19 \times 19 \times 256$ ), P6 ( $10 \times 10 \times 256$ ), and P7 ( $5 \times 5 \times 256$ ). The use of FPN ensures that achieves each layer can be used for object detection with different object sizes, and its main function was to fuse multi-scale features to achieve effective prediction results. FPN is a fusion of multi-scale features in a

structure that combines high-level semantics with the underlying semantics. Given that the high-level features have rich semantic information, the object classification accuracy is relatively high, but the object localization ability is weak. Meanwhile, the underlying features have less semantic information and stronger object localization ability.

Given that Figure 5 c is the classification subnet, the ECA attention module (Wang et al. 2020) was introduced, and the effective feature layers of the feature pyramid P3, P4, P5, P6, and P7 were refined again in the classification subnet. This attention model was used to identify the most important parts of the network for processing, with a focus on the information of interest while suppressing useless information, which improved the conciseness and efficiency of the network.

Figure 7 shows the ECA attention module after a channel-level global averaging without dimensionality reduction. The features obtained in the previous step were then subjected to 1D convolution for learning. Among them, the size of the convolution kernel of 1D convolution affects the coverage of cross-channel interactions, that is, it affects the number of channels considered for the calculation of each weight of the attention mechanism, defined in [eq. (1)].

$$K = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor, \text{ of which } \gamma = 2, b = 1 \quad (1)$$

After completing the 1D convolution, the Sigmoid function is taken once more to fix the value to between 0 and 1. At this point, the weight of each channel of the input feature layer (between 0 and 1) is obtained. After obtaining this weight, this weight is multiplied by the original input feature layer.

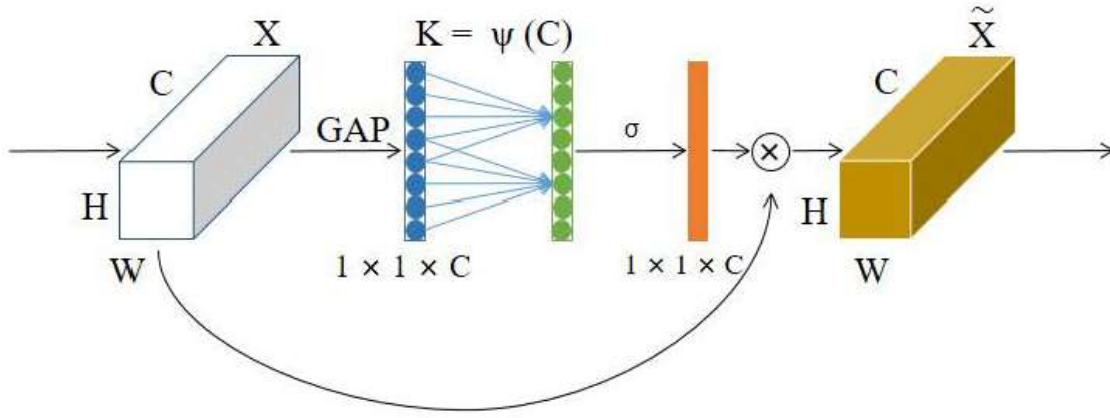


FIGURE 7. ECA-Net Architecture.

Without affecting the feature extraction of the previous feature pyramid, the subdivision operation of weight redistribution was carried out on the extracted feature graph. The refined feature map was passed through four  $3 \times 3$  convolution layers of  $256 \text{ channels} \times W \times H$  and a  $3 \times 3$  convolution layer with  $9 \times K$  filters, where  $K$  is the number of object categories. The sigmoid activation function is applied to the output to output the final classification result for the level of pineapple maturity. The main function of a classification subnet is to classify objects. The foreground and background are distinguished and categories of objects are identified. The main function of the classification subnet is to achieve the task of border regression, and the specific task is to correct the border error.

Figure 5 d is a regression subnet, with a network structure that is almost the same as that of the classification subnet but does not share parameters. The regression subnet produces a  $4 \times 9$  linear output for each spatial location. For each anchor in each spatial position, the frame regression subnet calculates the offset between the anchor frame and the real data frame calibrated nearby and revises the frame regression positioning of pineapple maturity to obtain a more accurate object frame.

The loss function of this model is:

$$FL(p_t) = -\alpha(1 - p_t)^\gamma \log(p_t) \quad (2)$$

Where:

$$p_t = \begin{cases} p & \text{if } e=1 \\ 1-p & \text{otherwise} \end{cases} \quad (3)$$

$$CE(P_t) = -\log(P_t) \quad (4)$$

Focal Loss(Lin et al. 2017) is a simple deformation of Cross Entropy Loss where  $(1 - p)^\gamma$  in [eq. (2)] represents a modulation factor,  $\alpha$  in eq (2) represents weighting factor,  $p_t$  in [eq. (3)] represents the estimated probability of binary classification, and  $e$  is the true label; [eq. (4)] represents Cross Entropy Loss.

### Performance metrics

The test was evaluated by calculating the mAP and  $F_1$  score. FPS is how many frames per second the object network could detect. The mAP and  $F_1$  score are related to Precision (P,%), Recall (R,%), using True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN) in the confusion grid matrix. The calculation equation is:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{AP} = \int_0^1 P(R) dR \quad (7)$$

$$\text{mAP} = \frac{1}{M} \sum_{K=1}^M \text{AP}(k) \times 100\% \quad (8)$$

$$F_1 = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Precision} + \text{Recall}} \quad (9)$$

## RESULTS AND DISCUSSION

### Experimental environment

The experimental platform configuration for this paper was the following: OS, Win10; GPU, RTX 2070 SUPER; CPU, Intel(R) Core(TM) i7-9700K CPU @ 3.60 GHz; Memory, 16 GB; Hard disk, 1TB; NVIDIA driver, 456.71. The programming language Python and deep learning framework PyTorch versions 3.8 and 1.7 were used. All seven models were trained on this configuration.

### Model training and testing

In this study, the migration learning approach was used and fine-tuned for a specific pineapple maturity detection task. In the pineapple maturity detection task, the object needed to

be divided into five categories that is, the four stages of pineapple maturity and background, so the subnet parameters were set to five. The specific parameters of the ECA-RetinaNet are as follows: the maximum learning rate of the model is set to 1e-4, the minimum learning rate of the model is the maximum learning rate  $\times$  0.01, the Adam optimizer is used, and the model train a total of 100 epochs.

The curve of the loss value of ECA-Retinanet with the number of epochs during training. As shown in Figure 8, when the network epoch exceeds 85, the loss value leveled off and dropped to approximately 0.054. From the parameter convergence, the network training results are in line with the values required.

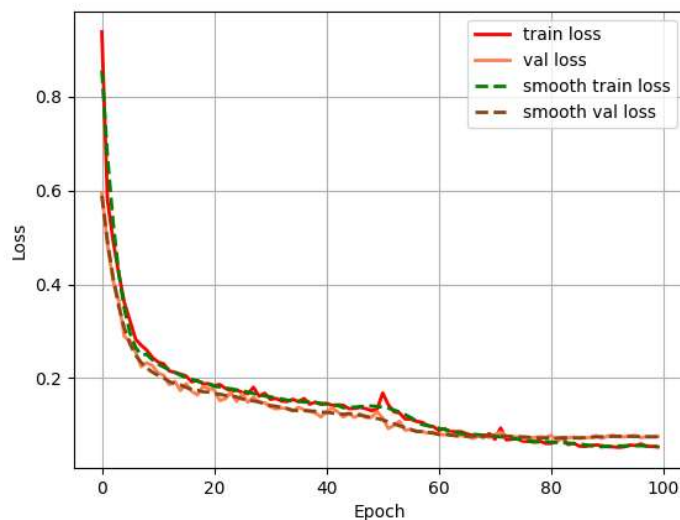


FIGURE 8. Loss changing graph.

### Comparison experiment

#### ECA-RetinaNet detection results

The ECA-RetinaNet structure in this study is based on the RetinaNet backbone network, while incorporating the ECA module for improvement. This has demonstrated the effectiveness of the improved ECA-RetinaNet based on the RetinaNet-based network. A comparative analysis of the detection network performance before and after the

improvement is required.

The test set results of the four mature precision, total average precision, accuracy, recall, F<sub>1</sub> score, and FPS are shown in Table 1. The F<sub>1</sub> score of ECA-RetinaNet are 2% and 1.5% higher in the mild and severe complex scenarios, respectively, without affecting the real-time detection. The average precision of ECA-RetinaNet was almost always higher than the original RetinaNet in terms of detection accuracy.

TABLE 1. Comparison of test results from the detection network before and after improvement.

Scenarios	Algorithm	Mean Average Precision mAP/%					Precision /%	Recall /%	F <sub>1</sub> /%	FPS
		Young	Near young	Near mature	Mature	Mean				
Mild complex	RetinaNet	99.6	92.01	97.46	100	97.27	94.4	<b>97.38</b>	92.75	<b>27</b>
	<b>ECA-RetinaNet</b>	<b>99.73</b>	<b>93.29</b>	<b>97.73</b>	<b>100</b>	<b>97.69</b>	<b>94.8</b>	95.41	<b>94.75</b>	26.9
Severe complex	RetinaNet	85.18	<b>94.59</b>	92.98	92.94	91.42	90	84.86	88.5	<b>27</b>
	<b>ECA-RetinaNet</b>	<b>87.52</b>	94.43	<b>95.75</b>	<b>95.08</b>	<b>93.2</b>	<b>91.6</b>	<b>88.84</b>	<b>90</b>	26.9

Young\_mature\_pineapple (Young), near\_young\_mature\_pineapple (Near young), near\_mature\_pineapple (Near mature), mature\_pineapple (Mature), Precision value (Precision), Recall value (Recall), F1 score (F1)

Figure 9 shows a comparison of the detection network recognition effect before and after improvement. Figure 9 (a) and Figure 9 (c) are the original RetinaNet model detection effect, and Figure 9 (b) and Figure 9 (d) are the ECA-RetinaNet model detection effect. In the mild complex scenario, there is a missed detection in the lower right corner of Figure 9 (a). The missed pineapple is severely obscured by the branches and leaves. Meanwhile, it was detected in Figure 9 (b) and the detection result was as expected. In the severe

complex scenario, Figure 9 (c) had a missed detection in both the upper left and upper right corners, and a false detection in the upper left corner. Figure 9 (d) did not have a missed detection and the improved network recognition resulted in correct identification of all the pineapples encompassing different levels of maturity. The detection effect of the ECA-RetinaNet model was found to be more effective than the detection using the RetinaNet model.

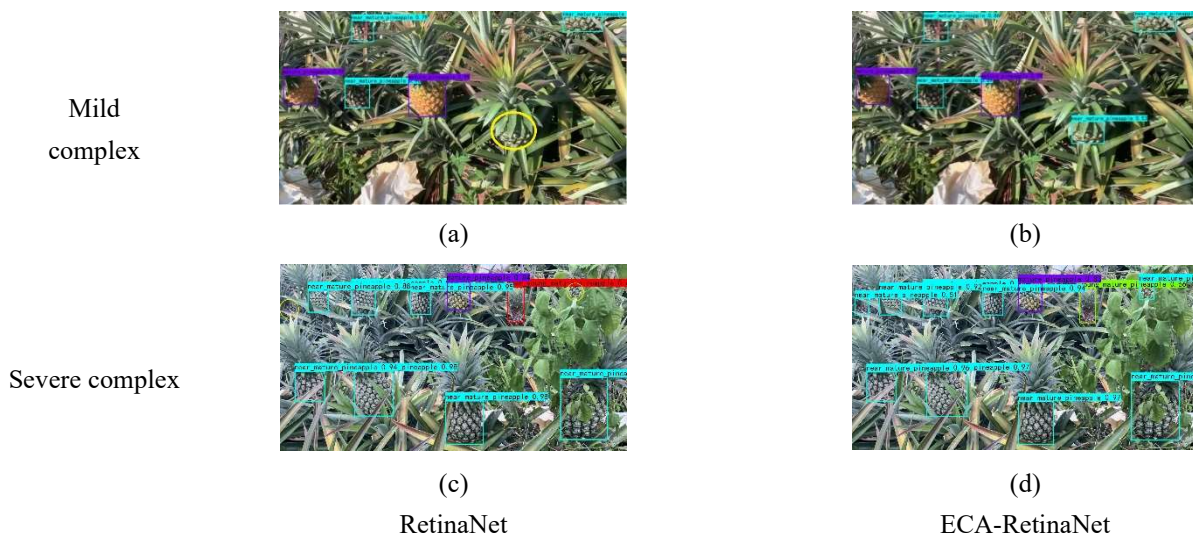


FIGURE 9. Comparison of detection network recognition effect before and after improvement. The yellow boxes indicate the pineapples that were not detected.

The results have shown that to some extent, this method effectively solves the problem of pineapples being difficult to detect in complex environments. Detection errors can be caused by light transitions that can severely affect the imaging quality with shading between fruit and weed branches and leaves. According to Sabóia et al. (2022), the low proportion of objects detected in the image may be because of the constant search for focus with movement, as the camera equipment used performs auto-zoom and fails to improve the focus. According to Zheng et al. (2017), for natural RGB images, detection becomes difficult due to changes in lighting and weather, with difficulty in distinguishing colors in shadow areas. According to Li et al. (2022), a near-colored background containing leaves and canopy affects the recognition accuracy of different maturity levels for flat dates in the field. Therefore, the algorithm was improved using an attention mechanism to improve the extraction of small targets at a shallow level against similar backgrounds and to improve the detection accuracy. Karthik et al. (2020) proposed an attention-based residual deep network for disease detection in tomato leaves. The inclusion of an attention mechanism gives more weight to features that need to be the key focus, which allows for accurate classification. In this study, the detection accuracy was further improved by the introduction of the ECA attention module, which assigns different features to different channels of the feature map that has already been extracted and selectively

increases the weight value of the channel containing pineapple fruit.

#### Comprehensive comparison of different object detection networks

The experiment aimed to compare different object detection networks with the improved model detection capability metrics in detail. The ECA-RetinaNet, RetinaNet, Faster R-CNN (Ren et al., 2015), CenterNet (Duan et al., 2019), YOLOv3 (Redmon & Farhadi 2018), YOLOv4 (Bochkovskiy et al., 2020), and SSD (Liu et al., 2016) object detection algorithm models were trained using the datasets produced in this study, and the optimal models were derived and then tested on the mild and severe complex test set, and a total of seven sets of experimental data results were obtained.

Liu et al. (2022) proposed a model based on binocular stereo vision and improved YOLOv3. It is used for intelligent picking, detecting, and positioning of pineapple fruit. On the test set with slight occlusion, the AP and F1 score of the improved YOLOv3 model were 97.55% and 93.18% respectively. In this study, for the mild complex scenarios, as shown in Table 2, the precision value, and F1 score of ECA-RetinaNet were the highest among these seven object detection networks. The mAP of ECA-RetinaNet was 0.42%, 1.69%, 1.22%, 0.46%, and 2.83% higher than that of RetinaNet, Faster R-CNN, CenterNet, YOLOv3, and YOLOv4, respectively.



TABLE 2. Performance comparison of different models in mild complex scenarios.

Algorithm	Mean Average Precision mAP/%					Precision /%	Recall /%	F <sub>1</sub> /%	FPS
	Young	Near young	Near mature	Mature	Mean				
<b>ECA-RetinaNet</b>	<b>99.73</b>	<b>93.29</b>	<b>97.73</b>	<b>100</b>	<b>97.69</b>	<b>94.8</b>	95.41	<b>94.75</b>	27
RetinaNet	99.6	92.01	97.46	100	97.27	94.4	<b>97.38</b>	92.75	26.9
Faster R-CNN	98.95	88.37	96.75	99.93	96.0	92.8	96.61	82	11
CenterNet	99.52	88.65	97.71	100	96.47	94	92.65	94.5	52
YOLOv3	99.22	92.42	97.28	100	97.23	94.4	95.34	93.25	26
YOLOv4	99.18	81.71	98.56	100	94.86	93.1	91.48	89.75	28
SSD	98.29	93.83	99.32	100	97.86	88.6	97.04	92	<b>63</b>

Young\_mature\_pineapple (Young), near\_young\_mature\_pineapple (Near young), near\_mature\_pineapple (Near mature ), mature\_pineapple (Mature), Precision value (Precision), Recall value (Recall), F1 score ( F1)

As shown in Figure 10, there are four pineapples in the original image. Each pineapple fruit has an occlusion problem, and one pineapple is particularly badly occluded. The original RetinaNet, CenterNet, SSD, YOLOv3, YOLOv4, and SSD missed the heavily obscured pineapple on the right side of the image. ECA-RetinaNet successfully detected the fruit. The yellow boxes indicate the missed fruit that were not detected.

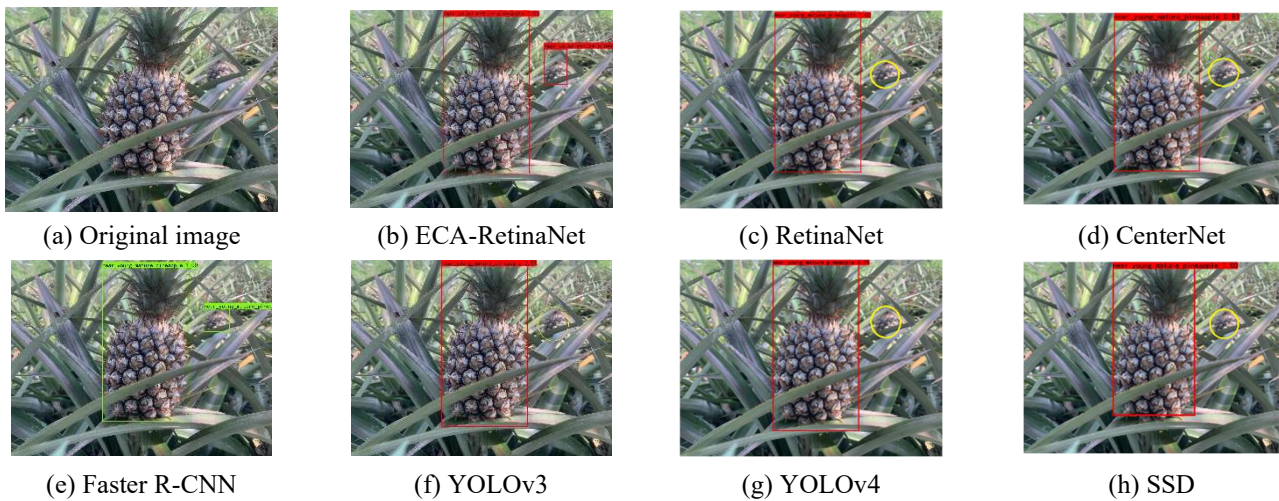


FIGURE 10. Comparison of different model detection effects in mild complex scenarios. The yellow boxes indicate the missed fruit that were not detected.

According to Liu et al. (2022), as the occlusion grew more severe in the pineapple detection, the F<sub>1</sub> score and AP values decreased to 89.15% and 91.47%. In this study, for severe complex scenarios, as shown in Table 3, the mAP, precision value, and F<sub>1</sub> score for ECA-RetinaNet were higher than the other object detection networks compared with ECA-RetinaNet being the most effective.

TABLE 3. Performance comparison of different models in severe complex scenarios.

Algorithm	Mean Average Precision mAP/%					Precision /%	Recall /%	F <sub>1</sub> /%	FPS
	Young	Near young	Near mature	Mature	Mean				
<b>ECA-RetinaNet</b>	<b>87.52</b>	94.43	<b>95.75</b>	<b>95.08</b>	<b>93.2</b>	<b>91.6</b>	88.84	<b>90</b>	27
RetinaNet	85.18	<b>94.59</b>	92.98	92.94	91.42	90	84.86	88.5	26.9
Faster R-CNN	87.04	91.81	94.59	93.6	91.76	90.1	<b>92.35</b>	80.75	11
CenterNet	84.13	91.76	94.12	91.72	90.43	89	74.98	85.25	52
YOLOv3	82.59	91.14	93.05	91.49	89.57	87.9	82.12	86.5	26
YOLOv4	69.02	84.66	85.83	86.8	81.58	80.3	74.77	79.25	28
SSD	90.13	90.92	89.19	90.98	90.3	88.2	81.4	87.25	<b>63</b>

Young\_mature\_pineapple (Young), near\_young\_mature\_pineapple (Near young), near\_mature\_pineapple (Near mature ), mature\_pineapple (Mature), Precision value (Precision), Recall value (Recall), F1 score ( F1)



In the severe complex scenarios as in Figure 11, there are a total of five pineapples in the original image, which have fruit overlapping with branch occlusion. The original RetinaNet, CenterNet, SSD, YOLOv3, and YOLOv4 all showed missed or incorrect detections. The faster R-CNN and ECA-RetinaNet detected all five fruits.

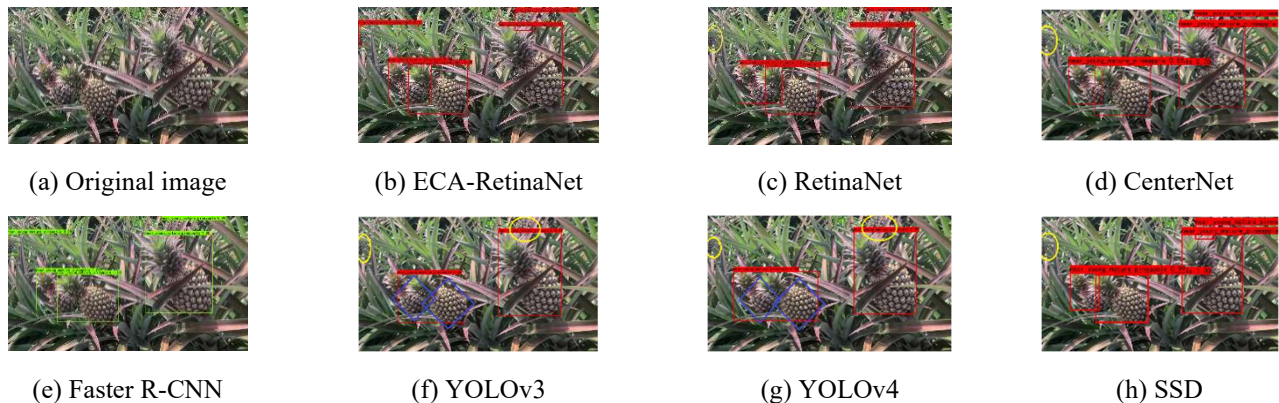


FIGURE 11. Comparison of different model detection effects in severe complex scenarios. The yellow and purple boxes indicate missed and incorrect detections.

The ECA-RetinaNet proposed in this study has been shown to be effective in identifying different levels pineapple maturity in mild and severe complex scenario conditions. Therefore, it is suitable for detecting pineapple maturity under complex scenes in natural orchard scenes. It has a high level of research value and importance in realizing yield estimation and mechanical automatic picking research and development.

## CONCLUSIONS

In this study, the RetinaNet detection model was improved by incorporating the ECA attention mechanism method to identify pineapples in orchards with four main different maturity levels. The experiments have shown that the ECA-RetinaNet has a recognition mAP of 97.69% and an F1 score of 94.75% in mild complex scenarios, and an average accuracy of 93.2% and an F1 score of 90% in severe complex scenarios. The FPS is 27 and meets the requirement of real-time detection. The ECA-RetinaNet model performed better than the original RetinaNet model and outperformed six state-of-the-art fruit detection models such as Faster R-CNN. The improved RetinaNet model proved its applicability as a method to identify pineapples in the main maturity stages in orchards.

## ACKNOWLEDGEMENTS

This work was partially supported by National Science Foundation for Young Scientists of China (61703170) and Science and Technology Project of Guangdong Province of China (2015A020209119).

## REFERENCES

Aguilar E JL, Borromeo GKP, Flores VJ (2021) Determination of pineapple ripeness using support vector machine for philippine standards. In: IEEE 7th International Conference on Control Science and Systems Engineering (ICCSSE). Beijing, Proceedings...

Bochkovskiy A, Wang CY, Liao HYM (2020) Yolov4: Optimal speed and accuracy of object detection. arXiv:2004.10934. <https://doi.org/10.48550/arXiv.2004.10934>

Chen X, Bu Q (2019). Research on fruit recognition algorithm based on multi-color and local texture. Journal of Qingdao University (Engineering Technology Edition) 34(03): 52-58. <https://doi.org/10.13306/j.1006-9798.2019.03.010>

Darrenl 2019 LabelImg, repositório Github. Available: <https://github.com/tzutalin/labelImg>. Accessed Abr, 2019.

Duan KW, Bai S, Xie LX, Qi HG, Huang QM, Tian Q (2019) Centernet: Keypoint triplets for object detection. IEEE/CVF international conference on computer vision (CVPR):6569-6578

He KM, Zhang XY, Ren SQ, Sun J (2016) Deep residual learning for image recognition. IEEE conference on computer vision and pattern recognition (CVPR):770-778

Huang HJ (2020) Fruit detection technology based on the deep learning research and application. Master Thesis, Jiangsu University of Science and Technology.

Karthik R, Hariharan M, Sundar A, Priyanka M, Annie J, Menaka R (2020) Attention embedded residual CNN for disease detection in tomato leaves. Applied Soft Computing 86. <https://doi.org/10.1016/j.asoc.2019.105933>

Kong P (2021) Fruit object detection based on deep learning research and application. Master Thesis, Beijing University of Posts and Telecommunications.

Li DL, Jing M, Dai X, Chen Z, Ma C, Chen J (2022) Current status of pineapple breeding, industrial development, and genetics in China. Euphytica 218(6): 85. <https://doi.org/10.1007/s10681-022-03030-y>

Li S, Zhang S, Xue J, Sun H, Ren R (2022) A fast neural network based on attention mechanisms for detecting field flat jujube. Agriculture 12(5): 717. <https://doi.org/10.3390/agriculture12050717>

- Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. IEEE conference on computer vision and pattern recognition (CVPR):2117-2125
- Lin TY, Goyal P, Girshick R, He K, Dollár P (2017) Focal loss for dense object detection. IEEE international conference on computer vision (CVPR):2980-2988
- Liu L, Ouyang W, Wang X, Fieguth P, Chen J, Liu X, Pietikäinen M (2020) Deep learning for generic object detection: A survey." International Journal of Computer Vision 128(2): 261-318. <https://doi.org/10.1007/s11263-019-01247-4>
- Liu TH, Nie XN, Wu JM, Zhang D, Liu W, Cheng YF, Zheng Y, Qiu J, Qi L (2022) Pineapple (Ananas comosus) fruit detection and localization in natural environment based on binocular stereo vision and improved YOLOv3 model. Precision Agriculture. <https://doi.org/10.1007/s11119-022-09935-x>
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) SSD: Single shot multibox detector. European conference on computer vision (ECCV):21-37
- Mohd Basir Selvam NA, Ahmad Z, Mohtar IA (2021) Real time ripe palm oil bunch detection using YOLO v3 algorithm. In: IEEE 19th Student Conference on Research and Development (SCOREd), Proceedings...
- Nguyen HHC, Luong AT, Trinh TH, Ho PH, Meesad P, Nguyen TT (2021) Intelligent fruit recognition system using deep learning. International Conference on Computing and Information Technology:13-22
- Redmon J, Farhadi A (2018) Yolov3: An incremental improvement. arXiv:1804.02767. <https://doi.org/10.48550/arXiv.1804.02767>
- Ren S, He K, Girshick R, Sun JJAinips (2015) Faster R-CNN: Towards real-time object detection with region proposal networks. arXiv:1506.01497. <https://doi.org/10.48550/arXiv.1506.0149728>
- Sa I, Ge Z, Dayoub F, Upcroft B, Perez T, McCool C (2016) DeepFruits: A fruit detection system using deep neural networks. Sensors 16(8): 1222. <https://doi.org/10.3390/s16081222>
- Sabóia HdS, Mion RL, Silveira AdO, Mamiya AA (2022) Real-Time selective spraying for viola rope control in soybean and cotton crops using deep learning. Engenharia Agrícola 42(spe). <https://doi.org/10.1590/1809-4430-eng.agric.v42nepe20210163/2022>
- Solemane C, Bernard K-F, Dantouma K, Daouda T (2022) Deep learning for precision agriculture: A bibliometric analysis. Intelligent Systems with Applications 16: 200102. <https://doi.org/10.1016/j.iswa.2022.200102>
- Tang RC, W XR (2020) Real-time detection of passion fruit based on improved YOLO-V3 network. Journal of Guangxi Normal University (Natural Science Edition) 38(06): 32-39. <https://doi.org/10.16088/j.issn.1001-6600.2020.06.004>
- Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q (2020) ECA-Net: Efficient channel attention for deep convolutional neural networks. arXiv:1910.03151v4. <https://doi.org/10.48550/arXiv.1910.03151>: 11531-11539
- Wang TH, Chen B, Zhang ZQ, Li H, Zhang M (2022) Applications of machine vision in agricultural robot navigation: A review. Computers and Electronics in Agriculture 198. <https://doi.org/10.1016/j.compag.2022.107085>
- Wu X, Sahoo D, Hoi SCH (2020) Recent advances in deep learning for object detection. Neurocomputing 396: 39-64. <https://doi.org/10.1016/j.neucom.2020.01.085>
- Xiong JT, Zheng ZH, Liang JE, Zhong Z, Liu BL, Sun BX (2020) Citrus detection method in night environment based on improved YOLO v3 network. Journal of Agricultural Machinery 51(04): 199-206. <https://doi.org/10.6041/j.issn.1000-1298.2020.04.023>
- Zhao DA, Wu RD, Liu XY, Zhao YY (2019) Apple positioning based on YOLO deep convolutional neural network for picking robot in complex background. Transactions of the Chinese Society of Agricultural Engineering 35(03): 164-173. <https://doi.org/10.11975/j.issn.1002-6819.2019.03.021>
- Zheng Y, Zhu QB, Huang M, Guo Y, Qin JW (2017) Maize and weed classification using color indices with support vector data description in outdoor fields. Computers and Electronics in Agriculture 141: 215-222. <https://doi.org/10.1016/j.compag.2017.07.028>
- Zhu X, Ma H, Ji JT, Jin X, Zhao KX, Zhang K (2020) Detection and identifying blueberry canopy fruits based on Faster R-CNN. Journal of Southern Agriculture 51(06): 1493-1501. <https://doi.org/10.3969/j.issn.2095-1191.2020.06.032>