

## Postprint: Pineapple Maturity Analysis During Growth Stage Based on Lightweight MobileNet V3-YOLOv4

**Authors:** Li Yangde, Ma Xiaohui, Wang Ji

**Date:** 2023-08-14T00:00:00+00:00

### Abstract

[Purpose/Significance] Pineapple storability is related to its ripeness, and recognizing pineapple ripeness before harvesting is particularly important. This study aims to propose a novel network model to improve the accuracy and speed of automatic pineapple ripeness recognition. [Method] First, to address the limitations of small sample size and poor real-time performance in pineapple training datasets, we self-constructed a plantation scene pineapple ripeness analysis dataset using photos of pineapples taken in natural environments. Subsequently, we replaced the YOLOv4 backbone network with the lightweight MobileNet V3 and proposed the lightweight MobileNet V3-YOLOv4 network. Simultaneously, we trained the original YOLOv4 model, MobileNet V1-YOLOv4 model, MobileNet V2-YOLOv4 model, and five different single-stage and two-stage network models including Faster R-CNN, YOLOv3, SSD300, RetinaNet, and CenterNet, and compared the evaluation metrics of these models to analyze the superiority of our proposed model. [Results and Discussion] Experimental results show that the MobileNet V3-YOLOv4 had a training time of 11,924 s and a parameter count of 53.7 MB. The trained MobileNet V3-YOLOv4 achieved a mean Average Precision (mAP) of 90.92% on the validation set. For the two categories of yellow-ripe pineapple and green-ripe pineapple, the detection Precision was 100% and 98.85%, respectively; the Average Precision (AP) values were 87.62% and 94.21%, respectively; the Recall rates were 77.55% and 86.00%, respectively; the F1 Scores were 0.87 and 0.92, respectively; and the inference speed was 80.85 img/s (Frames Per Second, FPS). [Conclusion] The MobileNet V3-YOLOv4 proposed in this study achieves improved accuracy and inference speed for pineapple ripeness recognition while reducing training time and parameter count, meeting practical detection requirements.

## Full Text

# Pineapple Maturity Analysis During Growth Period Based on Lightweight MobileNet V3-YOLOv4

LI Yangde<sup>1,2</sup>, MA Xiaohui<sup>3</sup>, WANG Ji<sup>3,4</sup>

<sup>1</sup>The Fourth Branch of Guangdong Telecom Planning and Design Institute Co., Ltd., Zhanjiang, Guangdong 524088, China

<sup>2</sup>School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430070, China

<sup>3</sup>College of Electronics and Information Engineering, Guangdong Ocean University, Zhanjiang, Guangdong 524088, China

<sup>4</sup>Guangdong Provincial Engineering Technology Research Center for Intelligent Ocean Sensor Networks and Equipment, Zhanjiang, Guangdong 524088, China

**Abstract:** Pineapple storability is closely related to its maturity, making pre-harvest maturity identification particularly important. This study aims to propose a novel network model that improves both the accuracy and speed of automatic pineapple maturity recognition. To address limitations such as insufficient training samples and poor real-time performance in existing pineapple datasets, we constructed a plantation-scene pineapple maturity analysis dataset using photographs captured in natural environments. The backbone network was subsequently replaced with a lightweight architecture, yielding a lightweight MobileNet V3-YOLOv4 network. For comprehensive evaluation, we trained five distinct single-stage and two-stage network models including Faster R-CNN, YOLOv3, SSD300, RetinaNet, and CenterNet, alongside the original YOLOv4, MobileNet V1-YOLOv4, and MobileNet V2-YOLOv4 models, comparing their evaluation metrics to demonstrate the superiority of our proposed model. Experimental validation on the test set demonstrated that MobileNet V3-YOLOv4 achieved a mean Average Precision (mAP) of 90.92%, with Precision, Recall, and F1 Score of 98.85%, 77.55%, and 86.00% respectively for yellow-ripe pineapples, and 87.62%, 94.21%, and 90.92% respectively for green-ripe pineapples. The inference speed reached 80.85 frames per second (FPS). The proposed MobileNet V3-YOLOv4 successfully reduces training time and model parameters while simultaneously improving recognition accuracy and inference speed for pineapple maturity, meeting practical detection requirements.

**Keywords:** Pineapple maturity; Backbone network; MobileNet V3-YOLOv4; Faster R-CNN; SSD300; RetinaNet; CenterNet; Lightweight

---

## 1 Introduction

Pineapple is a tropical fruit widely cultivated in southern Chinese provinces such as Hainan and Guangdong, with extensive planting areas and high yields. Delayed harvesting often leads to fruit rot, while premature picking adversely

affects edibility and taste [1]. During transportation from production sites to nationwide destinations, varying distances result in different transit periods. Higher maturity levels correspond to poorer storability and shorter permissible transport cycles, necessitating accurate pre-harvest maturity assessment to enable optimal picking timing and transport planning for different destinations.

Traditional fruit maturity analysis methods typically rely on manual inspection, fruit composition analysis, or volatile gas component detection [2]. Although these methods achieve relatively accurate maturity assessment, they interfere with fruit growth, involve complex procedures, incur high costs, and suffer from low efficiency, failing to meet real-time requirements [3]. With the continuous advancement of computer vision technology, Convolutional Neural Network (CNN)-based image recognition algorithms have matured in agricultural applications, emerging as effective models for visual tasks [4]. Numerous scholars have applied CNN-based recognition algorithms to fruit detection, extracting color, texture, and shape features from images or videos to achieve fruit identification and analysis.

Gai et al. [5] proposed an improved YOLOv4 deep learning algorithm for detecting small cherry fruits, achieving a 0.15 higher mAP than standard YOLOv4. Chen et al. [6] utilized image size, height, and width features through YOLOv3 for continuous fruit tracking and quality detection during tracking, testing on 6,000 fruit images with 88% accuracy. Kuznetsova et al. [7] developed pre- and post-processing techniques enabling YOLOv3 application in apple harvesting robot vision systems, achieving 19 ms average detection time with 7.8% false positives and 9.2% missed detections. Zhang et al. [8] designed a 13-layer CNN for classifying pears, apples, and rotten fruits with 94.94% accuracy. Chaikew et al. [9] developed a neural network-based pineapple sorting machine that classifies maturity into unripe, partially ripe, and fully ripe based on color, achieving 79%, 82%, and 100% accuracy respectively. Cuong et al. [10] proposed an improved Tiny YOLOv4 model for pineapple maturity prediction with 98.26% recognition accuracy.

However, the datasets in references [5]–[10] feature single-color backgrounds suitable for ideal laboratory scenarios without occlusion, whereas real-world pineapple detection faces challenges including branch and leaf occlusion, uneven illumination, and overlapping shadows. While references [11]–[12] achieved pineapple recognition in complex scenes, they focused only on single-category detection lacking generalizability. Practical detection applications demand networks with fast processing speed, high throughput, recognition accuracy, and strong generalizability—requirements that conventional CNNs struggle to meet. To address these challenges, this study proposes a novel optimized network, MobileNet V3-YOLOv4, based on the YOLOv4 [13] framework combined with MobileNet V3, offering reduced training time and high recognition accuracy to satisfy real-time maturity analysis requirements in natural pineapple plantation scenarios.

## 2 Methods

### 2.1 Data Collection

**2.1.1 Data Acquisition** Network training data were collected on March 25, 2023, at the “Pineapple Sea” in Qujie Town, Xuwen County, Zhanjiang City, using a SONY DSC-RX100M7 camera with image resolution of  $720 \times 480$  pixels. Based on pineapple color, shape, and texture characteristics, naturally harvestable pineapples were classified into green-ripe and yellow-ripe stages. Initial screening eliminated overexposed, excessively blurred, or target-free images, yielding 1,580 images comprising 674 yellow-ripe and 906 green-ripe pineapple images. These were randomly split into 1,264 training images, 158 validation images, and 158 test images.

Data collection occurred in complex pineapple plantation environments with varying shooting angles, times, and illumination intensities. The dataset included: 857 suitable samples without occlusion or shadows under moderate lighting, 203 samples with branch-leaf occlusion, 232 samples with uneven lighting and shadows, 112 samples under excessively strong or weak lighting, and 176 samples with multiple coexisting influencing factors. This diverse collection ensured sample variety for robust network training. Figure 1 [Figure 1: see original paper] illustrates sample images under various influencing factors.

**2.1.2 Data Annotation** To meet network training requirements, the LabelImg annotation software was used to generate \*.xml label files. Images were excluded from annotation if target occlusion exceeded 70%, surface shadows were too heavy for color/shape discrimination, or fruits were too small/distant for edge contour resolution. Green-ripe and yellow-ripe pineapples were labeled as “Growing pineapple” and “Mature pineapple” respectively.

### 2.2 MobileNet V3-YOLOv4 Network Construction

The YOLOv4 network suffers from complex structure, massive parameters, long training times, and large recognition errors. Agricultural monitoring scenarios typically deploy minimal surveillance equipment, requiring visual algorithms to achieve rapid pineapple detection with limited computational resources. Therefore, this study introduced the MobileNet V3 model to replace YOLOv4’s original backbone feature extraction network.

**2.2.1 YOLOv4 and MobileNet** The YOLOv4 network architecture comprises three main components: the CSPDarknet53 [14] backbone for feature extraction yielding three effective preliminary feature layers sized (13,13,1024) for large objects, (26,26,512) for medium objects, and (52,52,256) for small objects; the neck structure combining Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PANet); and the prediction network with three YOLO Head modules for final regression and classification. The SPP module enhances receptive field using four max-pooling kernels ( $13 \times 13$ ,  $9 \times 9$ ,  $5 \times 5$ , and  $1 \times 1$ ), while

PANet improves upon YOLOv3's Feature Pyramid Networks (FPN) [16] for enhanced feature extraction and fusion. Figure 2 [Figure 2: see original paper] illustrates the YOLOv4 structure.

MobileNet, proposed by Google in 2017, is a lightweight CNN designed for mobile and embedded devices [17]. It pioneered depthwise separable convolution modules to replace standard convolutions. MobileNet V3 combines depthwise separable modules with residual structures to propose a bottleneck (bneck) architecture [18], as shown in Figure 3 [Figure 3: see original paper]. The bneck structure first uses  $1 \times 1$  convolution for dimension expansion, followed by  $3 \times 3$  depthwise separable convolution for feature extraction, convolution for dimension reduction, with residual connections linking input and output. The bneck structure employs h-swish activation (Equation 1) instead of ReLU and incorporates a Squeeze-and-Excitation (SE) module (Figure 4 [Figure 4: see original paper]) between separable and reduction convolutions. The SE module consists of global pooling, two fully connected layers, a ReLU activation, and an h-sigmoid function (Equation 2) that replaces the standard sigmoid.

$$\text{h-swish}[x] = x \times (\text{ReLU6}(x + 3) / 6)$$

$$\text{h-sigmoid}[x] = \text{ReLU6}(x + 3) / 6$$

**2.2.2 MobileNet V3-YOLOv4 Backbone Construction** MobileNet V3-YOLOv4 replaces YOLOv4's CSPDarknet Resblock with the bneck structure. The improved backbone (Figure 5 [Figure 5: see original paper]) processes a (416,416,3) input image through an initial  $3 \times 3$  convolution (stride 2) to extract a (208,208,16) feature layer, then applies multiple bneck modules to obtain three effective preliminary feature layers at (52,52,40), (26,26,112), and (13,13,160) for input to the enhanced feature extraction network. Additionally, MobileNet V1-YOLOv4 and MobileNet V2-YOLOv4 networks were constructed for comparative analysis.

**2.2.3 Object Detection Evaluation Metrics** With two label categories (Growing pineapple and Mature pineapple), evaluation metrics include Average Precision (AP), Recall, F1 Score, Precision, and mean Average Precision (mAP), calculated as:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F1 Score} = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall})$$

$$\text{mAP} = (\sum \text{AP}_i) / C$$

where TP represents true positives, TN true negatives, FP false positives, FN false negatives, and C the number of categories (C=2 in this study).

### 2.3 Pineapple Maturity Detection Algorithm Workflow

The algorithm workflow comprises: (1) manual image acquisition using cameras with preliminary screening to discard unsuitable images; (2) maturity discrimination based primarily on color with annotation into yellow-ripe and green-ripe categories; (3) dataset training through backbone network for preliminary features, neck structure for enhanced feature extraction, and head structure for regression and classification to generate the trained model; (4) evaluation metric computation for the final model. Figure 6 [Figure 6: see original paper] illustrates this workflow.

## 3 Experiments and Analysis

### 3.1 Experimental Environment and Parameter Settings

Software environment: Ubuntu 16.04, Anaconda3, PyTorch 1.2.0, CUDA 10.0, cuDNN 7.6.4. Hardware: NVIDIA GeForce GTX 2080Ti GPU, Intel Core i5-8400 CPU @ 2.80 GHz. Programming language: Python 3.7. Initial training parameters are listed in Table 1 .

### 3.2 Model Performance Comparison

To evaluate model complexity, nine models were compared: Faster R-CNN, YOLOv3, SSD300, RetinaNet, CenterNet, YOLOv4, MobileNet V1-YOLOv4, MobileNet V2-YOLOv4, and the proposed MobileNet V3-YOLOv4. Performance metrics included training time, saturation epoch, and parameter count. MobileNet V3-YOLOv4 achieved a training time of 11,924 s (39.75 s/epoch), saturation at epoch 256, and parameter count of 53.7 MB—25.59% faster saturation than YOLOv4 with only 22% of the parameters. Detailed comparisons are provided in Table 2 .

### 3.3 Training Results Analysis

**3.3.1 Single-Class and Multi-Class Detection Performance** To validate classification performance, four metrics (Recall, F1 Score, Precision, AP) were used for different maturity stages. Comparative training results across models are presented in Table 3 . While Faster R-CNN, YOLOv3, SSD300, RetinaNet, and CenterNet showed good performance for yellow-ripe pineapples, they performed poorly on green-ripe ones. Among YOLOv4-based networks, MobileNet V3-YOLOv4 demonstrated superior Precision, AP, and F1 Score. For green-ripe pineapples, AP improved by 4.49%, F1 Score by 0.07, Recall by 1%, and Precision by 3.34% compared to YOLOv4. For yellow-ripe pineapples, AP improved by 6.06%, F1 Score by 0.13, Recall by 16.55%, and Precision by 6.25%, reaching 100% Precision due to distinct color features.

Multi-class detection results (Table 4 ) show MobileNet V3-YOLOv4 achieved 90.92% mAP—5.28% higher than YOLOv4 and 3.67% higher than YOLOv3—

while maintaining 80.85 FPS, which is 40.28 img/s faster than YOLOv4 and 8.91 img/s faster than SSD300.

**3.3.2 Discussion** Comprehensive analysis of Tables 2–4 reveals MobileNet V3-YOLOv4’ s superiority across all aspects. Although some metrics are slightly lower than MobileNet V1-YOLOv4 and MobileNet V2-YOLOv4, several factors demonstrate its practical advantages: (1) For single-category detection, MobileNet V3-YOLOv4 distinguishes subtle color and texture variations, achieving optimal AP values; (2) For multi-category detection, its mAP is 0.75% higher than MobileNet V2-YOLOv4, balancing detection effectiveness across both maturity stages; (3) Model complexity is significantly reduced, with 37.80% fewer parameters than YOLOv4 and 7,248 s shorter training time; (4) Inference speed is optimal in the YOLOv4 series at approximately double YOLOv4’ s speed, substantially reducing time costs for practical applications. These factors collectively establish MobileNet V3-YOLOv4 as a viable model for pineapple maturity analysis.

### 3.4 Detection Results in Complex Environments

Pineapple cultivation occurs in complex farmland environments with challenges including overlapping branch occlusion, incomplete fruits, background colors similar to fruit (especially green-ripe pineapples), debris, and illumination variations causing shadows. To evaluate MobileNet V3-YOLOv4’ s performance under these conditions, test set images containing multiple influencing factors were randomly selected for detection. Figure 7 [Figure 7: see original paper] presents comparative results across four networks on six test images containing seven pineapples. YOLOv4 showed the most missed detections (3/7), particularly for yellow-ripe pineapples. MobileNet V1-YOLOv4 and MobileNet V2-YOLOv4 also exhibited missed detections (1 and 2 respectively). In contrast, MobileNet V3-YOLOv4 achieved the best balance, effectively detecting both maturity stages with minimal missed detections, thereby meeting natural environment detection requirements.

## 4 Conclusion

This study replaced YOLOv4’ s backbone with MobileNet V3 to propose a lightweight MobileNet V3-YOLOv4 network. Trained and tested on pineapple fruit images captured in complex farmland scenarios, the model demonstrated superior performance with 5.28% higher mAP, 40.28 img/s faster inference, 3.34% higher Precision for green-ripe pineapples, and 6.25% higher Precision for yellow-ripe pineapples compared to YOLOv4, while using only one-fifth the parameters. This optimization simultaneously improves accuracy and speed while reducing computational cost and model size, enabling easier deployment across hardware platforms for real-time maturity analysis in natural scenes.

Future research should expand the dataset with additional maturity types. Cur-

rent images contain only 1-2 pineapples each, whereas practical farmland monitoring with few cameras captures numerous fruits that may appear pixelated when zoomed, potentially affecting detection. Enhanced hardware resolution could address this. Data collection was conducted in sunny weather; incorporating diverse weather conditions would further enrich dataset diversity. Integration of MobileNet V3-YOLOv4 with remote sensing technology and UAVs could enable real-time whole-farm pineapple maturity analysis, providing theoretical support for automated harvesting.

**Conflict of Interest Statement:** The authors declare no conflicts of interest.

## References

- [1] JIN Y, LIU H Q, LIU E P. Study on the regional advantages of pineapple planting in China[J]. Chinese journal of agricultural resources and regional planning, 2014, 35(3): 100-104.
- [2] LIU C H, LIU Y, XIE S L, et al. Analysis of aroma components of pineapple fruit at different ripening stages[J]. Chinese journal of tropical crops, 2009, 30(2): 234-237.
- [3] LIU S H, SUN W S, LU X H, et al. Analysis of aromatic components in ripe fruits of 6 pineapple cultivars[J]. Chinese journal of tropical crops, 2015, 36(6): 1179-1185.
- [4] FORTUNA-CERVANTES J M, RAMÍREZ-TORRES M T, MARTINEZ-CARRANZA J, et al. Object detection in aerial navigation using wavelet transform and convolutional neural networks: A first approach[J]. Proceedings of the institute for system programming of the RAS, 2021, 33(2): 149-162.
- [5] GAI R, CHEN N, YUAN H. A detection algorithm for cherry fruits based on the improved YOLOv4 model[J]. Neural computing and application, 2021, 35(19): 13895-13906.
- [6] CHEN M C, CHENG Y T, LIU C Y. Implementation of a fruit quality classification application using an artificial intelligence algorithm[J]. Sensors and materials: An international journal on sensor technology, 2022, 34(1): 1-12.
- [7] KUZNETSOVA A, MALEVA T, SOLOVIEV V. Using YOLOv3 algorithm with pre- and post-processing for apple detection in fruit-harvesting robot[J]. Agronomy, 2020, 10(7): 1016-1035.
- [8] ZHANG Y D, DONG Z C, CHEN X Q, et al. Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation[J]. Multimedia tools and applications, 2019, 78(3): 3613-3632.
- [9] CHAIKAEW A, THANAVANICH T, DUANGTANG P, et al. Convolutional neural network for pineapple ripeness classification machine[C]//2019 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). Piscataway, NJ, USA:

IEEE, 2020: 373-376.

[10] CUONG N H H, TRINH T H, MEESAD P, et al. Improved YOLO object detection algorithm to detect ripe pineapple phase[J]. Journal of intelligent & fuzzy systems: Applications in engineering and technology, 2022, 43(1): 1365-1377.

[11] ZHANG X, GAO Q M, PAN D, et al. Picking recognition research of pineapple in complex field environment based on improved YOLOv3[J]. Journal of Chinese agricultural mechanization, 2021, 42(1): 201-206.

[12] LIU T H, NIE X N, WU J M, et al. Pineapple (*Ananas comosus*) fruit detection and localization in natural environment based on binocular stereo vision and improved YOLOv3 model[J]. Precision agriculture, 2023, 24(1): 1-18.

[13] MAO Zhihui, ZHU Jiali, WU Xin, et al. A survey of autonomous driving object detection based on YOLO[J]. Computer Engineering and Applications, 2022, 58(15): 68-77.

[14] WANG X, GAO J S, HOU B J, et al. A lightweight modified YOLOX network using coordinate attention mechanism for PCB surface defect detection[J]. IEEE sensors journal, 2022, 22(21): 20910-20920.

[15] XU D Q, WU Y Q. Improved YOLOv3 with DenseNet for multi-scale remote sensing target detection[J]. Sensors, 2020, 20(15): 4276.

[16] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2014: 580-587.

[17] SUN Y Y, MA S, SUN S Y, et al. Partial discharge pattern recognition of transformers based on MobileNets convolutional neural network[J]. Applied sciences, 2021, 11(15): 6984.

[18] LI J P, ZHU K Y, WANG F, et al. Deep neural network-based real time fish detection method in the scene of marine fishing supervision[J]. Journal of intelligent & fuzzy systems: Applications in engineering and technology, 2021, 41(3): 4527-4532.

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv – Machine translation. Verify with original.*