agronomy

MDPI

Article

A Real-Time Detection and Maturity Classification Method for Loofah

Sheng Jiang [1,2], Ziyi Liu [1], Jiajun Hua [1], Zhenyu Zhang [1], Shuai Zhao [1], Fangnan Xie [1], Jiangbo Ao [1], Yechen Wei [1], Jingye Lu [1], Zhen Li [1,2,3] and Shilei Lyu [1,4,*]

[1] College of Electronic Engineering (College of Artificial Intelligence), South China Agricultural University, Guangzhou 510642, China; jiangsheng@scau.edu.cn (S.J.); liuziyi2021@stu.scau.edu.cn (Z.L.); huajiajun1111@stu.scau.edu.cn (J.H.); cristozhang@stu.scau.edu.cn (Z.Z.); zhaoshuai21@stu.scau.edu.cn (S.Z.); finnian@stu.scau.edu.cn (F.X.); djangoao@stu.scau.edu.cn (J.A.); 2022yechen@stu.scau.edu.cn (Y.W.); lujingye@stu.scau.edu.cn (J.L.); lizhen@scau.edu.cn (Z.L.)
[2] Engineering Research Center for Monitoring Agricultural Information of Guangdong Province, Guangzhou 510642, China
[3] Division of Citrus Machinery, China Agriculture Research System of MOF and MARA, Guangzhou 510642, China
[4] Pazhou Lab, Guangzhou 510330, China
* Correspondence: lvshilei@scau.edu.cn

Abstract: Fruit maturity is a crucial index for determining the optimal harvesting period of open-field loofah. Given the plant's continuous flowering and fruiting patterns, fruits often reach maturity at different times, making precise maturity detection essential for high-quality and high-yield loofah production. Despite its importance, little research has been conducted in China on open-field young fruits and vegetables and a dearth of standards and techniques for accurate and non-destructive monitoring of loofah fruit maturity exists. This study introduces a real-time detection and maturity classification method for loofah, comprising two components: LuffaInst, a one-stage instance segmentation model, and a machine learning-based maturity classification model. LuffaInst employs a lightweight EdgeNeXt as the backbone and an enhanced pyramid attention-based feature pyramid network (PAFPN). To cater to the unique characteristics of elongated loofah fruits and the challenge of small target detection, we incorporated a novel attention module, the efficient strip attention module (ESA), which utilizes long and narrow convolutional kernels for strip pooling, a strategy more suitable for loofah fruit detection than traditional spatial pooling. Experimental results on the loofah dataset reveal that these improvements equip our LuffaInst with lower parameter weights and higher accuracy than other prevalent instance segmentation models. The mean average precision (mAP) on the loofah image dataset improved by at least 3.2% and the FPS increased by at least 10.13 f/s compared with Mask R-CNN, Mask Scoring R-CNN, YOLACT++, and SOLOv2, thereby satisfying the real-time detection requirement. Additionally, a random forest model, relying on color and texture features, was developed for three maturity classifications of loofah fruit instances (M1: fruit setting stage, M2: fruit enlargement stage, M3: fruit maturation stage). The application of a pruning strategy helped attain the random forest model with the highest accuracy (91.47% for M1, 90.13% for M2, and 92.96% for M3), culminating in an overall accuracy of 91.12%. This study offers promising results for loofah fruit maturity detection, providing technical support for the automated intelligent harvesting of loofah.

Keywords: loofah; maturity classification; machine vision; deep learning; machine learning

check for updates

Citation: Jiang, S.; Liu, Z.; Hua, J.; Zhang, Z.; Zhao, S.; Xie, F.; Ao, J.; Wei, Y.; Lu, J.; Li, Z.; et al. A Real-Time Detection and Maturity Classification Method for Loofah. *Agronomy* **2023**, *13*, 2144. https://doi.org/10.3390/agronomy13082144

Academic Editor: Marco Scortichini

Received: 17 July 2023
Revised: 10 August 2023
Accepted: 14 August 2023
Published: 16 August 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The loofah (scientific name *Luffa cylindrica*), a common vegetable, holds significant importance in the Guangdong–Hong Kong–Macao Greater Bay Area's "Vegetable Basket Project", especially in the Pearl River Delta region of Guangdong, China, due to its high

nutritional value and extensive applications. Harvesting the loofah's tender fruit, the main product, requires precise timing. Early harvesting can prevent full growth, reducing commercial value, while delayed harvesting can lead to a marked decline in fruit quality. Therefore, accurately determining loofah maturity is critical to optimize harvest timing and ensure the vegetable's quality, texture, and edible value [1]. Yet, currently, maturity assessments predominantly rely on subjective time-consuming manual methods, which fall short of modern agricultural production and market demands.

The use of computer vision technology in agriculture has seen considerable research and development, especially for crop maturity detection. Researchers continually strive to devise methods and technologies to enhance the accuracy and efficiency of crop quality assessments. Notable successes have been achieved, particularly with fruits such as passion fruit [2], tomato [3], and mango [4]. Wan et al. [5] introduced a method combining color features and a backpropagation neural network (BPNN) classification technique to assess fresh tomatoes' maturity level in markets. The process involved using image processing technology on tomato images, extracting color feature values, and using these values as BPNN inputs. Tan et al. [6] employed outdoor color images to identify blueberries at different maturity stages, using the HOG feature vector and linear SVM classifier for quick fruit region detection and KNN and TMWE classifiers for maturity identification. However, these methods might be limited by complex orchard environments, such as varying backgrounds and lighting conditions, and may show low model robustness due to the stark color changes during the fruits' maturation process.

YOLO series algorithms [7–9] represent the state-of-the-art in deep learning-based object detection. These have been successfully applied to crop maturity detection tasks by numerous researchers. For instance, Tian et al. [10] proposed an improved YOLOv3 model to detect apples at varying growth stages in orchards. Using image enhancement methods, such as rotation transform, color balance transform, brightness transforms, and DenseNet, the model was able to perform real-time detection of apples under complex backgrounds. Similarly, Qiu et al. [11] suggested a grape maturity detection and visual prepositioning algorithm based on an improved YOLOv4, which facilitated rapid precise identification and classification of grapes at different maturity stages, providing spatial location data of grape clusters. Zhang et al. [12] developed a deep learning method for *Hemerocallis citrina Baroni* maturity detection using YOLOv5, introducing the Ghost module to reduce model complexity and the squeeze-and-excitation (SE) and convolutional block attention module (CBAM) module to optimize feature extraction and model accuracy.

Presently, maturity detection tasks for cucurbit family plants such as the loofah are mainly performed by acoustic feature identification [13] or spectral analysis [14]. However, these methods demand specialized knowledge and experience and are time-consuming and labor-intensive, which hinders the loofah industry from accurately determining fruit maturity and optimizing harvest timing. This study aims to tackle this challenge by proposing a real-time loofah instance segmentation and maturity classification method. In conclusion, the main contributions of this paper are two-fold:

(1) We propose a lightweight instance segmentation model, LuffaInst, capable of rapidly executing detection and segmentation tasks on loofah while reducing computational costs and improving detection accuracy;
(2) Considering the dynamic characteristics of loofah throughout its growth cycle, we attempt to establish a random forest model based on HSV color space and gray-level co-occurrence matrix texture features for accurate loofah maturity classification. We also introduce a pruning strategy to enhance the model's generalization ability.

The remainder of this paper is structured as follows: Section 2 details the data acquisition and augmentation methods employed in this study. Section 3 provides a comprehensive overview of our real-time detection and maturity classification method for loofah, which includes the LuffaInst instance and random forest model. Section 4 presents relevant experiments conducted on LuffaInst and the random forest model, including experimental settings, evaluation criteria, and model performance comparisons. Section 5 discusses the

limitations of the proposed loofah maturity detection method and offers recommendations for addressing these. Finally, we summarize our proposed method for loofah maturity detection and provide suggestions for future work.

2. Materials

Data Acquisition

The process of loofah fruit formation commences with the pollination and fertilization of its flowers. Following successful fertilization and pollination, the ovary undergoes a gradual transformation, evolving into a fruit that expands outward from the flower's base. During this phase, the loofah fruit takes shape and initiates its growth, denoted as the fruit setting stage. Subsequent to this stage, the loofah fruit continues its expansion, steadily augmenting in size. This phase represents the most rapid growth period for the fruit, a pivotal juncture influencing the distinctive characteristics of the loofah's shape; it is aptly referred to as the fruit enlargement stage. As the enlargement stage concludes, the loofah fruit's growth abates, its volume stabilizes, its surface adopts angular features, and its color undergoes a transformation. This stage is designated as the fruit maturation stage.

The original loofah image dataset for this study was collected from a loofah plantation farm in Sanjiang County, Zengcheng District, Guangzhou City, Guangdong Province, where the primary cultivar is huadian loofah. Utilizing an iPhone12 smartphone (Apple Inc., Zhengzhou, China), we captured images of the loofah under natural conditions between 10 AM and 1 PM from May to June 2023. We collected a total of 1052 original loofah images, each with a resolution of 3042 × 4032 pixels. In order to mimic the complex field environment, our dataset encompassed various growth stages of the loofah captured under different angles, scenes, and weather conditions, such as sunny, cloudy, and post-rain periods. Furthermore, we employed offline data augmentation methods, including RandomAffine, RandomFlip, MinIoURandomCrop, and SimpleCopyPaste, to broaden our initial loofah image dataset, enhancing the model's robustness and generalizability. As a result, we obtained a final dataset containing 4270 loofah images. The data were annotated using the Labelme software package version 5.3.0, with the dataset divided into training and validation sets at a 9:1 ratio. Figure 1 demonstrates the augmentation method of loofah images.

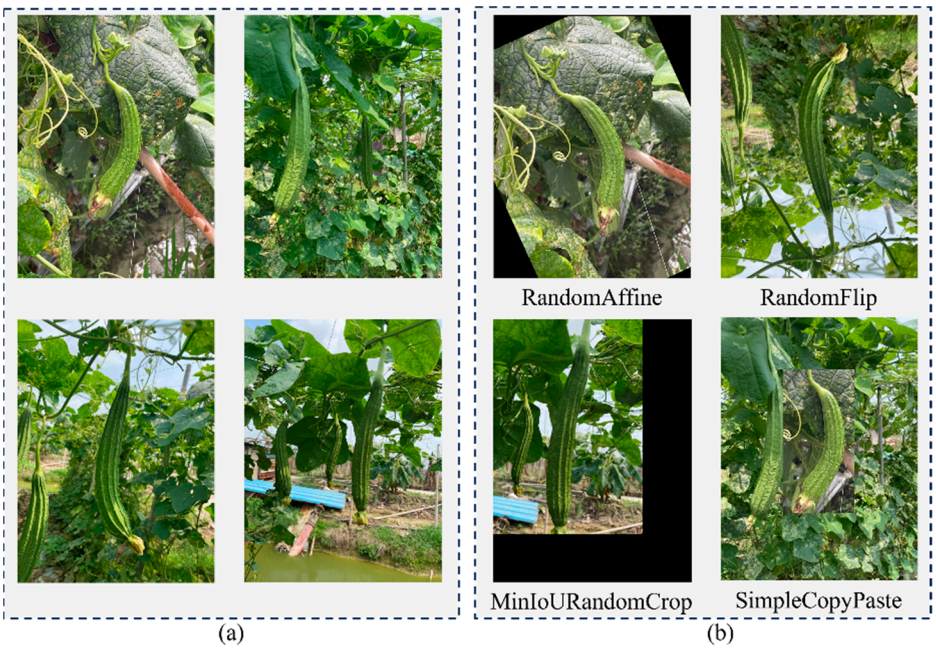


Figure 1. (**a**) Origin images; (**b**) images after data augmentation.

3. Methodology

Our study's methodology consists of two primary components: the LuffaInst instance segmentation model and the random forest decision model. As shown in Figure 2, the LuffaInst's detection branch decoder and instance segmentation decoder leverage shared feature maps to achieve their respective tasks, thus ensuring the accurate detection and segmentation of the loofah fruit. In addition, we implemented a random forest decision model to automatically extract fruit instances from LuffaInst and assess fruit maturity through random forest decision making.

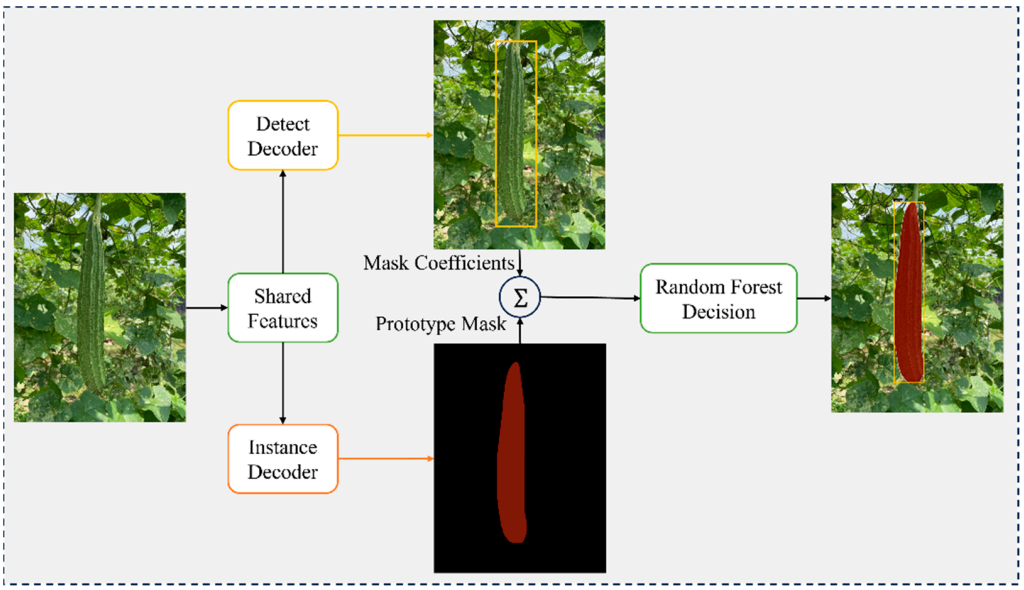


Figure 2. Overview of loofah real-time detection and maturity classification method.

3.1. LuffaInst

We propose LuffaInst, a one-stage lightweight real-time instance segmentation model. The base architecture of LuffaInst, illustrated in Figure 3, is composed of a shared encoder and two independent decoders. The model maximizes accuracy and enhances real-time performance by splitting instance segmentation into two parallel tasks. The first branch incorporates a mask output into the object detection branch to predict mask coefficients. The second branch uses the original instance segmentation components from YOLACT [15] as the segmentation decoder to segment the pixels of the loofah target area, generating prototype masks. The instance segmentation results are then produced by linearly combining the prototype masks and mask coefficients.

3.1.1. Encoder

The encoder component of LuffaInst is realized through the collaboration of a backbone network and a neck section. The backbone network employs an efficient hybrid architecture network, EdgeNeXt [16], based on the design principles of ConvNext [17]. EdgeNeXt introduces the segmentation depth transpose attention (SDTA) encoder, which divides the input tensor into multiple channel groups. It utilizes deep convolution and self-attention across channel dimensions to expand the receptive field and encode multi-scale features. The channel grouping and attention mechanism mitigate computational resource consumption.

As shown in Figure 3, we use EdgeNeXt-small as the backbone network, where P_n on the right represents the output feature maps of each respective layer, with output dimensions of $96 \times 80 \times 80$ (P1), $160 \times 40 \times 40$ (P2), and $304 \times 20 \times 20$ (P3). We merge the downscaled output feature maps from these three layers to generate a $560 \times 10 \times 10$ (P4) fourth-layer output feature map. By incorporating shallow-level feature mapping, we

bolster the model's ability to detect small objects. P1 serves as the input for the Protonet branch, where it is transformed into prototype masks using a fully convolutional network.

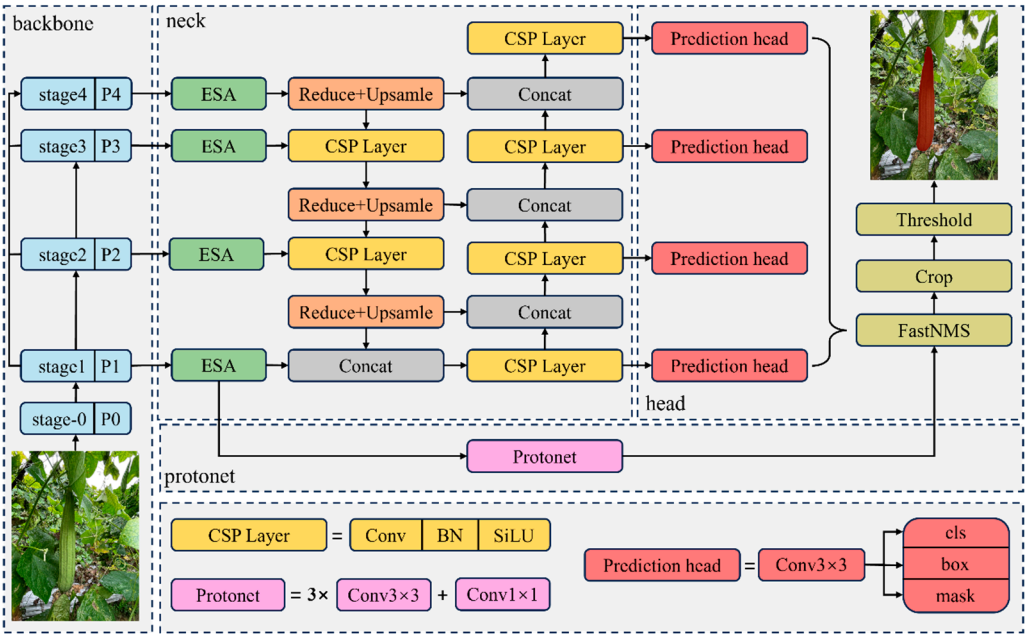


Figure 3. The overall architecture of the of LuffaInst.

The neck section is an enhanced version of the pyramid attention-based feature pyramid network (PAFPN) proposed by Liu [18]. PAFPN aggregates information pathways through a mix of bottom-up and top-down approaches, extracting richer features from each layer. To avoid losing feature details during the direct addition of upsampled and downsampled feature information, we preprocess the feature information of each layer via channel-wise operations before connecting them using the concatenation operation. In addition, we introduce a novel attention module, the efficient strip attention (ESA) module, at the start of each input feature map. This module enhances the representation of crucial features and facilitates their transfer between different network sections.

3.1.2. Efficient Strip Attention Module

Traditional feature extraction methods generally rely on a fixed kernel size of $N \times N$ square kernels, which have limited capacity to capture contextual information in images. However, for elongated loofah targets, conventional feature extraction processes might result in the loss of crucial information. To address this issue, we design the efficient strip attention module, which employs strip attention to establish long-range dependencies in discretely distributed regions [19]. By encouraging information interaction across channels, the module enables the model to concentrate on and enhance relevant feature information while ignoring irrelevant details, thereby improving loofah image feature extraction capabilities.

Figure 4 illustrates the efficient strip attention module. Let $x \in \mathbb{R}^{C \times H \times W}$ be the input tensor, with C representing the number of channels, H representing the height of feature map, and W representing the width of feature map. Initially, x is fed into two parallel pathways, each consisting of a horizontal or vertical strip pooling layer and a 1D convolution layer with a kernel size of 3. This produces a horizontal vector $y^h \in \mathbb{R}^{C \times H}$ and a vertical vector $y^v \in \mathbb{R}^{C \times W}$, respectively. To capture more global spatial information, y^h and y^v are combined to obtain a fusion vector $y \in \mathbb{R}^{C \times H \times W}$. Following popular attention module designs, a parallel channel attention branch is added to the module. This branch incorporates the efficient channel attention [20] and combines the extracted features from

the two pathways to generate the final spatial-channel feature attention map $x \in \mathbb{R}^{C \times H \times W}$, effectively capturing both spatial and channel-wise information.

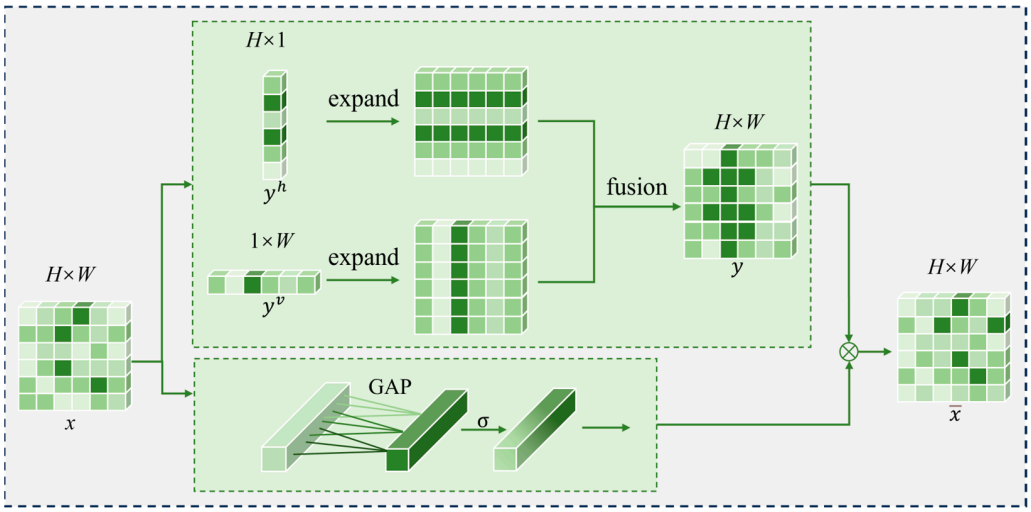


Figure 4. Diagram of the efficient strip attention module.

3.1.3. Decoder

The decoder part of LuffaInst adopts an anchor-based object detection approach similar to YOLACT. It comprises three branches in the detection head, responsible for predicting class confidence, anchor box regressions, and mask coefficients. To ensure real-time performance, these branches share a single 3×3 convolutional layer. The entire decoding process is split into two parallel tasks:

1. The detection head processes the four feature maps (P1–P4) from the encoder output, predicting class confidences, anchor box regression values, and mask coefficients. Given that the loofah targets occupy a rectangular region in the image, we adjust the anchor box aspect ratios. For each pixel, we generate three anchors with aspect ratios of $(1/2, 1/3, 1/4)$. We then filter the anchors using the fast non-maximum suppression algorithm (FastNMS) with a threshold > 0.5;
2. The P1 feature map is fed into the prototype branch, where a fully convolutional network (FCN) generates a prototype mask at $1/4$ of the image size. The instance mask is then produced by linearly combining the mask coefficients and the prototype mask.

3.2. Random Forest-Based Classification Method of Loofah Maturity

The random forest [21] ensemble learning method combines the features of decision trees and randomness. It is a robust machine learning algorithm applicable to both classification and regression problems, renowned for its robustness and interpretability. The random forest method constructs multiple independent decision tree models. Each decision tree is constructed based on bootstrap sampling, which involves randomly selecting a subset of samples with replacement from the training data. Simultaneously, in each decision tree, a subset of features is randomly chosen to split the nodes, which helps reduce the impact of feature correlations on predictions and enhances diversity among the decision trees. The final prediction result is obtained by aggregating the predictions of multiple decision trees through a voting mechanism. For a new sample, each decision tree produces a classification result and the ultimate classification is determined by the class with the highest number of votes. Taking into account the variations in the loofah growth process, we developed a random forest classifier based on a series of color and texture features. This classifier encompasses multiple decision trees that selectively draw from diverse feature sets to anticipate the maturity of the loofah. The culmination of these predictions, generated

through a voting mechanism, amalgamates the outcomes of the individual decision trees, ultimately yielding the final prediction for the loofah's maturity level.

3.2.1. Data Preprocessing

Based on the subjective experiences of production workers, picking needs, and the opinions of experts in horticulture, we divided the loofah fruit into three maturity stages: fruit setting stage (Figure 5a), fruit enlargement stage (Figure 5b), and fruit maturation stage (Figure 5c). We randomly selected 405 loofah images from the original dataset, each representing a stage of loofah maturation, including 127 images at fruit setting stage, 147 images at fruit enlargement stage, and 131 images at fruit maturation stage, and used LuffaInst to extract loofah instance images for analysis. Due to variations in lighting conditions and shooting angles in a natural environment, image brightness can fluctuate. Hence, we performed square equalization operations on the obtained instance images. Square equalization enhances image contrast, making details clearer and image brightness distribution more uniform, as shown in Figure 5f.

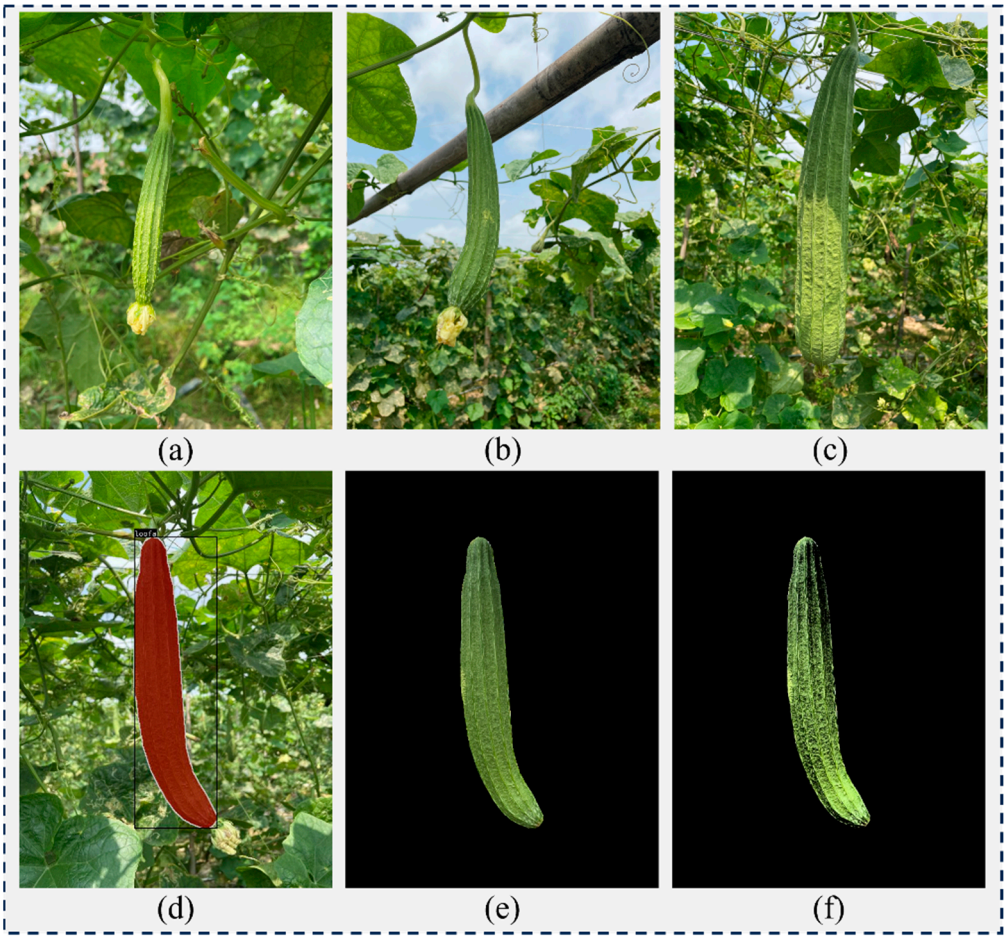


Figure 5. (**a**) Fruit setting stage; (**b**) fruit enlargement stage; (**c**) fruit maturation stage; (**d**) LuffaInst instance segmentation; (**e**) dividing loofah instance; (**f**) after histogram equalization.

3.2.2. Color Feature Extraction

Color features play a vital role in image processing and computer vision tasks. Frequently used color spaces include RGB (red, green, blue), CMY (cyan, magenta, yellow), HSV (hue saturation value), LAB (CIELAB stands for "CIE 1976 Lab*"), among others. In this experiment, we converted loofah images to the HSV color space. Compared with RGB, HSV aligns better with human color perception and can process color information separately from image brightness, making it more suitable for color-related analyses [22].

As the loofah's color transformation during its growth stage is mainly within the green hue, which gradually deepens as the fruit matures, we focused on the changes in the hue component range between 30 and 55. According to Figure 6, the average hue value progressively increased and the overall distribution shifted towards deeper tones during the three maturity stages. We obtained the combined features H, H/S, (H + S)/(H − S), and (H − S)/S by separating single mean features, where H represents hue component and S represents saturation component; the mean values of these features are presented in Table 1. We utilized the Pearson correlation coefficient to assess the relationship between the maturity of bottle gourds and color features. To address the discreteness of the maturity levels, we transformed the maturity into one-hot encoded variables. Subsequently, we employed the standard formula for calculating the Pearson correlation coefficient to obtain the correlation matrix, as shown in Equation (1). As shown in Figure 7, the mean of H and H/S exhibit the highest correlation coefficients of 0.46 and 0.41, respectively, indicating that these color features are most closely related to the loofah's maturity.

$$r = \frac{\sum\left(\left(X_i - X\right)\left(M_i - M\right)\right)}{\sqrt{\sum\left(X_i - X\right)^2 \sum\left(M_i - M\right)^2}} \tag{1}$$

where X_i and M_i are the values of features and maturity levels; X and M are the mean values of features and maturity levels.

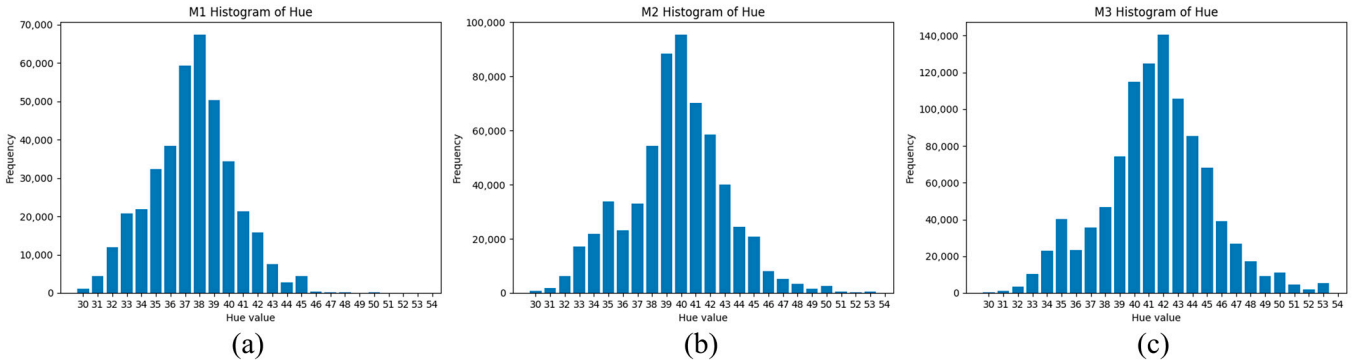


Figure 6. Histogram of hue distribution for different loofah maturity stages. (**a**) M1: fruit setting stage; (**b**) M2: fruit enlargement stage; (**c**) M3: fruit maturation stage.

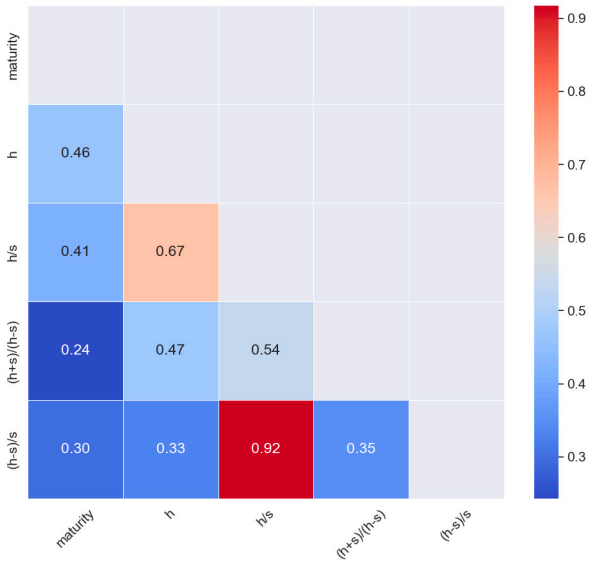


Figure 7. Correlation heat map of loofah maturity and color feature parameters.

Table 1. The mean values of color features for each maturity stage.

Mature Stage	Mean of H	Mean of H/S	Mean of (H + S)/(H − S)	Mean of (H − S)/S
M1	37.5097	0.3051	34.044	1.384
M2	39.7377	0.3237	95.9642	1.3922
M3	41.5724	0.3676	288.613	1.6004

3.2.3. Texture Feature Extraction

Texture features provide information about the local structure and details of an image. The gray-level co-occurrence matrix (GLCM) is a method employed for image analysis and texture feature extraction [23]. It helps describe and quantify the relationship between pixels of different grayscale levels in an image. This is achieved by counting the co-occurrence of different pixel pairs in the image and constructing a two-dimensional matrix that records the probability of two pixels having a specific grayscale level at a certain distance and direction. This matrix provides statistical information about image texture features such as contrast, homogeneity, dissimilarity, energy, correlation, and angular second moment (ASM).

Initially, a loofah fruit has a small smooth surface that may contain micro wrinkles or lines. As the fruit grows, the surface gradually forms ridged edges until maturation, when the surface becomes rougher and the ridged edge becomes prominent. Texture features noticeably change throughout the loofah growth process, so we use GLCM to extract texture features from loofah images. For each pixel, the difference in the grayscale level between it and its surrounding pixels is calculated. Based on these differences, a GLCM is constructed. This matrix records the frequency of each pair of disparity values. The five commonly extracted texture features include:

1. Contrast: Contrast measures the degree of contrast between different pairs of gray values in an image. A large contrast value indicates a significant grayscale level change in the image. The calculation formula is as follows:

$$\text{Contrast } = \Sigma\Big(i-j)^2 P(i,j) \tag{2}$$

2. Homogeneity: Homogeneity measures the consistency degree between different pairs of gray values in an image. Large homogeneity values indicate a more uniform image texture. The calculation formula is as follows:

$$\text{Homogeneity } = \frac{\sum P(i,j)}{(1+|i-j|)} \tag{3}$$

3. Dissimilarity: Dissimilarity measures the difference degree between different pairs of gray values in an image. A large dissimilarity value indicates a significant texture change in the image. The calculation formula is as follows:

$$\text{Dissimilarity } = \Sigma|i-j|P(i,j) \tag{4}$$

4. Correlation: Correlation measures the degree of linear correlation between different pairs of gray values in an image. Larger correlation values indicate a more organized texture in the image. The calculation formula is as follows:

$$\text{Correlation } = \Sigma\frac{[(i-\mu)(j-\mu)P(i,j)]}{\sigma^2} \tag{5}$$

5. Angular second moment (ASM): ASM measures the uniformity of image gray distribution and texture thickness. If the GLCM element values are similar, the value is small, indicating fine texture; a large value indicates a more uniform and regularly varying texture pattern. The calculation formula is as follows:

$$\mathrm{ASM} = \Sigma_i \Sigma_j P(i,j)^2 \tag{6}$$

where i and j are different gray value pairs in GLCM; $P(i,j)$ is the frequency of corresponding gray value pairs in normalized GLCM; μ is the mean value of GLCM; σ is the standard deviation of GLCM.

The statistical information of the average texture features of the three maturity stages of loofah is shown in Table 2. Here, the mean of contrast and dissimilarity are highly distinguishable between the three maturity stages. We utilize the same method as in Section 3.2.2 to compute the correlation between texture features and maturity. As shown in Figure 8, in the feature correlation coefficient matrix, the contrast and dissimilarity of texture features' correlation coefficients are 0.91 and 0.92, respectively, indicating a positive correlation with maturity. However, the correlation coefficients for correlation, homogeneity and ASM are all negatively correlated, around −0.90 and −0.94.

Table 2. The mean value of GLCM texture features for each mature stage.

Mature Stage	Contrast	Homogeneity	Dissimilarity	Correlation	ASM
M1	19.3502	0.9716	0.4782	0.9827	0.936
M2	33.4352	0.9533	0.8011	0.9801	0.9016
M3	67.3274	0.9250	1.5135	0.9763	0.8392

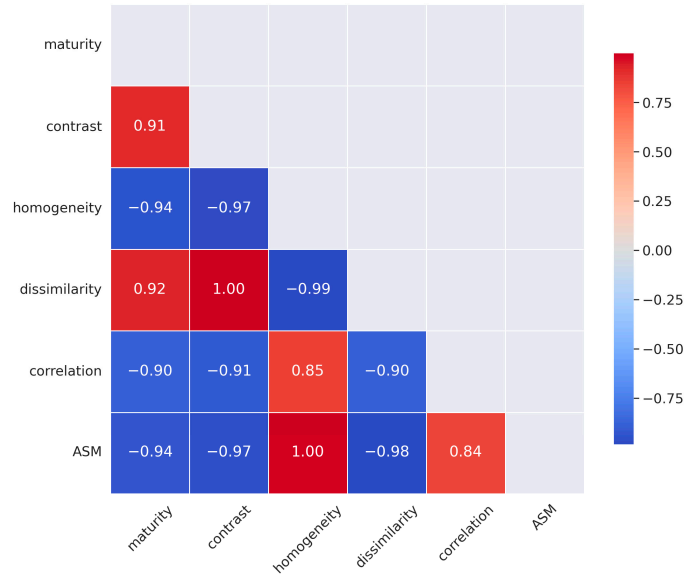


Figure 8. Correlation heat map of loofah maturity and texture feature parameters.

4. Results

4.1. Experimental Settings

This experiment's training and testing were conducted on a computer equipped with an NVIDIA RTX 3090 GPU to analyze the proposed method's performance. During LuffaInst training, we set the input size to a 640 × 640 resolution and employed the pretrained weights of the EdgeNeXt-small backbone network trained on ImageNet to expedite training convergence. LuffaInst was trained for 100 epochs using the AdamW optimizer with an initial learning rate of 0.01 and a weight decay of 0.0005. The quadratic warmup learning rate scheduling strategy gradually increased the learning rate to 0.01. From the 10th to the 100th epoch, a cosine annealing learning rate decay strategy was employed. During training, the model parameters were saved based on the highest accuracy achieved, which would be used for subsequent maturity classification experiments.

4.2. Metrics

The primary metric used was the mean average precision (*mAP*), which reflects the detection model's performance. Its calculation formula is given by:

$$mAP = \frac{1}{C}\sum_{i=0}^{C}\int_{0}^{1} P_i(R_i)dR_i \tag{7}$$

where P and R denote the model's precision and recall; C denotes the number of categories.

Moreover, we introduced the parameters of model and giga floating point operations per second (GFLOPs) to evaluate the model's size. Additionally, we assessed the model's runtime performance from multiple perspectives, including frames per second (FPS) and the time taken to detect a single image. These metrics provide a comprehensive understanding of the model's performance in terms of both efficiency and speed.

We used accuracy as an evaluation metric to assess the random forest classification model. Accuracy is defined as the number of correctly classified samples divided by the total number of samples. The calculation formula is given by Equation (8). In the formula, true positive (*TP*) represents the number of positive samples that are correctly predicted by the model, false positive (*FP*) represents the number of negative samples that are incorrectly predicted as positive by the model, false negative (*FN*) represents the number of positive samples that are incorrectly predicted as negative by the model, and true negative (*TN*) represents the number of negative samples that are correctly predicted by the model.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FN + FP + TN)} \tag{8}$$

4.3. Performance of the LuffaInst

We initially assessed the performance of LuffaInst by comparing it with several prominent instance segmentation models: Mask R-CNN [24] (a traditional two-stage model), the optimized Mask Scoring R-CNN [25], and the real-time instance segmentation models YOLACT++ [26] and SOLOv2 [27]. We ensured that all networks followed identical training environment settings, initialized with default parameters, and conducted transfer learning using the backbone pretrained on COCO dataset for optimal model performance. Additionally, we fixed random number seeds to guarantee the reliability of the experiment. Figure 9 illustrates the precision comparison of different models trained on the loofah image dataset.

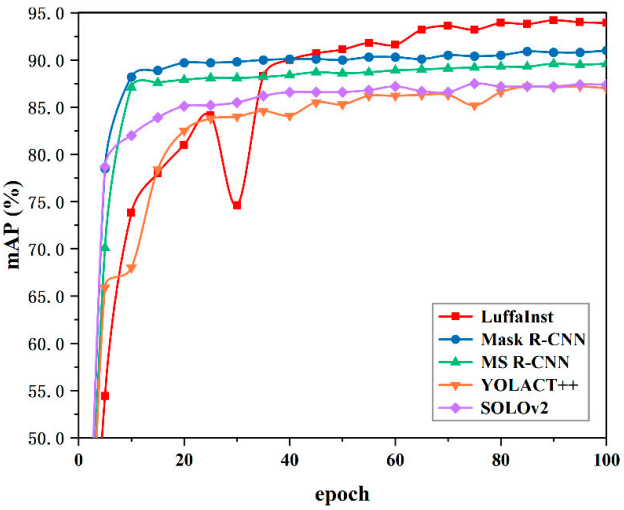


Figure 9. The accuracy of LuffaInst, Mask R-CNN, Mask Scoring R-CNN, YOLACT++, and SOLOv2 trained on the loofah image dataset.

As Table 3 reveals, LuffaInst achieves a higher mAP value with fewer parameter counts compared with other popular instance segmentation models. This indicates that our design of the ESA module and superimposition of shallow feature maps is effective. While the Mask R-CNN and its optimized network are typical two-stage instance segmentation models that deliver higher accuracy than other models, their frame rate on the loofah image dataset is a mere 9.5 f/s due to their two-stage model design, failing to meet real-time detection requirements. On the other hand, the YOLACT++ and SOLOv2 models, based on one-stage design, achieve this quite well but at the expense of accuracy, with their accuracy on the loofah image dataset falling below 90%. Figure 10 shows that both Mask R-CNN and MS R-CNN can accurately detect and segment loofah instances, whereas SOLOv2 might fail to detect small loofah instances against a complex background, and fruit detection on YOLACT++ may be entirely missed. Drawing on these experiences, LuffaInst uses a popular lightweight backbone network to ensure the model size and speed, improves PAFPN, integrates multi-scale feature maps, and adds a novel attention mechanism module to strengthen the model's ability to detect small targets. For the elongated loofah, it achieves accurate detection and segmentation with the highest accuracy rate of 94.2% among all models and a frame rate of 38.63 f/s that meets real-time requirements.

Table 3. The evaluation metrics of Mask R-CNN, Mask Scoring R-CNN, YOLACT++, SOLOv2, and LuffaInst on the loofah image dataset.

Models	mAP (%)	Parameters	GFLOPs	FPS	Times Per Image (ms)
Mask R-CNN	91.00	43.971 M	1.472 T	9.5	105.31
Mask Scoring R-CNN	89.60	60.656 M	2.874 T	27.34	36.65
YOLACT++	87.00	35.21 M	0.247 T	28.5	35.1
SOLOv2	87.5	46.22 M	0.139 T	17.31	58.2
LuffaInst	**94.20**	**11.91 M**	0.205 T	**38.63**	**25.79**

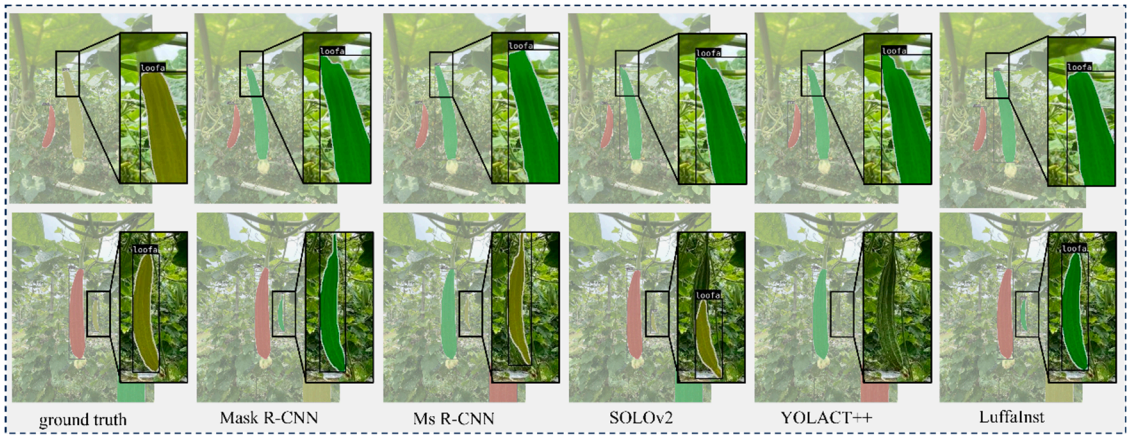


Figure 10. Visualization results of different models on the loofah image dataset.

Furthermore, we conducted ablation experiments on the ESA module we designed and compared it with the popular convolutional block attention module (CBAM) [28] and coordinate attention module (CA) [29]. The experimental results in Table 4 demonstrate that the LuffaInst employing the ESA module improves mAP by 1.4% and 1.8% compared with CBAM and CA when parameters increase by only 0.1 M. This indicates that for elongated targets such as the loofah, the ESA module allows the model to focus more on the target's long-range contextual information within the image and extract effective feature information. Although this slightly increases the GFLOPS compared with the CBAM and CA modules, the improvement in accuracy proves to be worthwhile. The ESA module appears more suitable for the loofah detection task.

Table 4. The evaluation metric of LuffaInst with different modules.

Models	mAP/%	Parameters	GFLOPS
With CBAM module	92.4	11.8 M	0.204 T
With CA module	92.8	11.8 M	0.204 T
With ESA module	94.2	11.91 M	0.205 T

4.4. Performance of the Random Forest

Table 5 provides a comparison of evaluation indicators for the random forest model using four different feature combinations. Firstly, for a model using a single color feature, we selected H and H/S, which have high correlation coefficients, as input features, yielding the lowest overall accuracy of 49.18% (Model 1). While the color of the loofah fruit changes throughout its growth, the shift in green hue is not easily discernible to the naked eye, leading to errors in judging the loofah's maturity based solely on color. Conversely, the overall accuracy when using the texture feature model was 70.49% (Model 2). This is because the texture changes on the surface of the loofah are more pronounced throughout its growth process. Starting from the fruit set stage, the fruit's surface begins to exhibit subtle folds and textures. By the ripening stage, the fruit's surface becomes rough, with prominent and detailed edges that are clearly discernible by touch. Hence, texture characteristics can serve as a key maturity identification feature throughout the loofah's growth process. When we employed the random forest model with a combination of seven features, the overall accuracy reached 81.96%, demonstrating the effectiveness of the random forest decision model that combines color and texture features.

Table 5. The metric for random forest with different combinations of features.

Model	The Selected Features							Accuracy (%)			
	H	H/S	Constant	Homogeneity	Dissimilarity	Energy	ASM	M1	M2	M3	Total
1	√	√	-	-	-	-	-	47.92	50.34	47.52	49.18
2	-	-	√	√	√	√	√	71.25	68.80	71.94	70.49
3	√	√	√	√	√	√	√	82.35	80.71	82.09	81.96
4	√	√	√	-	√	-	-	**91.47**	**90.13**	**92.96**	**91.12**

Random forests measure the importance of features based on the reduction in the Gini impurity of the node after segmentation using each feature. If the Gini impurity is higher, the greater the confusion, indicating that the data points in the node belong to different categories, while the lower the Gini impurity, the higher the purity, indicating that the data points in the node tend to belong to the same category. Therefore, after segmentation using a particular feature, the Gini impurity of the node is decreased dramatically and that feature is considered to be more important. We calculated the importance scores of the seven features using Python's scikit-learn library, as shown in Figure 11, wherein the constant and dissimilarity texture features account for more than 60% of all feature importance. This verifies that texture features are the main factor in loofah maturity discrimination, although ASM, energy, and homogeneity contribute less to the decision model. We pruned all features according to the importance score, eventually selecting H, H/S, constant, and dissimilarity as input features to obtain the best random forest model with an overall accuracy of 91.12% (Model 4).

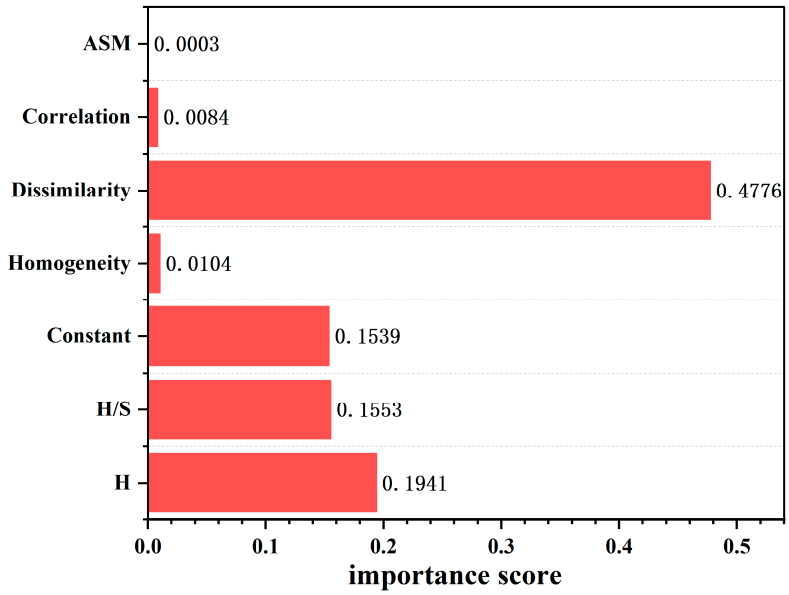


Figure 11. The importance score of all features.

5. Discussion

Loofah maturity assessment entails evaluating and gauging the degree of maturity exhibited by loofah fruits. This practice bears considerable significance within the realms of both agriculture and the food industry. Precisely ascertaining the maturity of loofah fruits assists farmers in making timely harvesting decisions, consequently circumventing potential quality setbacks arising from either premature or belated harvesting. The adoption of maturity detection technologies serves to streamline and optimize the inspection procedures for agricultural produce, furnishing technical reinforcement for intelligent harvesting machinery, and thus curbing expenses associated with labor and time. Timely evaluation of maturity acts as a preventive measure, thwarting spoilage and losses during the stages of harvesting, transportation, and storage, thereby contributing to the reduction of wastage in agricultural products.

In this study, we proposed a method for detecting and classifying the maturity of loofah. LuffaInst demonstrates superior performance on our loofah image dataset compared with other common and popular instance segmentation models. The random forest maturity classification model based on color and texture features is also capable of accurate classification. However, we identified some limitations in this method.

The performance of LuffaInst in this study is a prerequisite for the random forest model's accuracy in maturity classification. We ensured parameter efficiency by substituting the heavier pretrained backbone network with the lightweight EdgeNeXt-small. However, the enhancements to the PAFPN structure and the attention mechanism, while boosting accuracy, also increase computational complexity. Future research could explore methods such as parameter fusion and knowledge distillation to optimize the model.

The newly developed ESA module demonstrated success at LuffaInst, primarily attributed to its enhanced convolution through strip pooling, rendering it particularly well-suited for elongated targets such as loofahs. As a result, it outperformed the CBAM and CA attention modules in terms of accuracy. Nonetheless, the fusion method employed in the ESA module led to a marginal increase in the number of parameters and computational complexity. Future research could focus on exploring alternative feature computation and fusion techniques to optimize the module.

In the random forest maturity classification model, we adopted color and texture features for modeling. Still, the actual growth process of the loofah also involves morphological information (such as melon length and melon thickness), which is another crucial classification feature. Given that loofahs are prone to occlusion during their growth,

acquiring complete morphological features from the target in most occlusion cases is challenging, which could deteriorate the robustness of the model. In future research, we could investigate a function that fits the shape of the loofah, predicts, and fits the basic shape in light occlusion cases, thereby utilizing morphological characteristics. Moreover, more combinatorial features such as color and texture of the loofah can be further explored to enhance the model's robustness.

6. Conclusions

In this study, we proposed a method for detecting and classifying the maturity of the loofah with the aim of providing a solution for loofah fruit maturity detection in the efficient and intelligent harvesting technology of the loofah. This method comprised two parts: the instance segmentation model LuffaInst and the random forest maturity classification model. We designed a one-stage lightweight instance segmentation model LuffaInst, which employs the lightweight backbone network EdgeNeXt, improves the network structure, and incorporates the designed ESA attention mechanism module. Compared with MaskR-CNN, Mask Score R-CNN, YOLACT++, and SOLOv2, LuffaInst not only maintains the model size but also achieves a higher accuracy of 94.2% compared with other popular instance segmentation models. Considering the actual growth of loofah, we established a random forest maturity classification model based on color and texture features and identified the most appropriate feature combination through a pruning strategy. The results show that the random forest model with the combined features of H, H/S, constant, and dissimilarity performed the best, with the highest classification accuracy at all three maturity stages, including 91.47% at the fruit setting stage, 90.13% at the fruit expansion stage, 92.96% at the fruit ripening stage, and an overall accuracy of 91.12%. According to the shortcomings of the current study, we will continue to explore a method to detect and classify loofah maturity that balances speed and accuracy and combines more growth combination features.

Author Contributions: Conceptualization, S.J., Z.L. (Ziyi Liu) and J.H.; methodology, software, S.J. and Z.L. (Ziyi Liu); validation, F.X. and J.A.; data curation, Y.W. and J.L.; writing—review and editing, Z.Z. and S.Z.; supervision, Z.L (Zhen Li). and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Technologies R&D Program of Guangdong Province (2023B0202100001); National Natural Science Foundation of China (32271997, 31971797); China Agriculture Research System of MOF and MARA (CARS-26).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the privacy policy of the organization.

Acknowledgments: The authors would like to thank the anonymous reviewers for their critical comments and suggestions for improving the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yasmin, A.; Bharathi, R.V.; Radha, R. Review Article on *Luffa Acutangula* (L) Roxb. *Rese. J. Pharm. Technol.* **2019**, *12*, 2553. [CrossRef]
2. Tu, S.; Xue, Y.; Zheng, C.; Qi, Y.; Wan, H.; Mao, L. Detection of Passion Fruits and Maturity Classification Using Red-Green-Blue Depth Images. *Biosyst. Eng.* **2018**, *175*, 156–167. [CrossRef]
3. Malik, M.H.; Qiu, R.; Gao, Y.; Zhang, M.; Li, H.; Li, M. Tomato Segmentation and Localization Method Based on RGB-D Camera. *Int. Agric. Eng. J.* **2020**, *28*, 278–287.
4. Mim, F.S.; Galib, S.M.; Hasan, M.F.; Jerin, S.A. Automatic Detection of Mango Ripening Stages—An Application of Information Technology to Botany. *Sci. Hortic.* **2018**, *237*, 156–163. [CrossRef]
5. Wan, P.; Toudeshki, A.; Tan, H.; Ehsani, R. A Methodology for Fresh Tomato Maturity Detection Using Computer Vision. *Comput. Electron. Agric.* **2018**, *146*, 43–50. [CrossRef]
6. Tan, K.; Lee, W.S.; Gan, H.; Wang, S. Recognising Blueberry Fruit of Different Maturity Using Histogram Oriented Gradients and Colour Features in Outdoor Scenes. *Biosyst. Eng.* **2018**, *176*, 59–72. [CrossRef]

7. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
8. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2016**, arXiv:1506.02640.
9. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. *arXiv* **2016**, arXiv:1612.08242.
10. Tian, Y.; Yang, G.; Wang, Z.; Wang, H.; Li, E.; Liang, Z. Apple Detection during Different Growth Stages in Orchards Using the Improved YOLO-V3 Model. *Comput. Electron. Agric.* **2019**, *157*, 417–426. [CrossRef]
11. Qiu, C.; Tian, G.; Zhao, J.; Liu, Q.; Xie, S.; Zheng, K. Grape Maturity Detection and Visual Pre-Positioning Based on Improved YOLOv4. *Electronics* **2022**, *11*, 2677. [CrossRef]
12. Zhang, L.; Wu, L.; Liu, Y. Hemerocallis Citrina Baroni Maturity Detection Method Integrating Lightweight Neural Network and Dual Attention Mechanism. *Electronics* **2022**, *11*, 2743. [CrossRef]
13. Khoshnam, F.; Namjoo, M.; Golbakhshi, H. Acoustic Testing for Melon Fruit Ripeness Evaluation during Different Stages of Ripening. *Agric. Conspec. Sci.* **2015**, *80*, 197–204.
14. Jie, D.; Wei, X. Review on the Recent Progress of Non-Destructive Detection Technology for Internal Quality of Watermelon. *Comput. Electron. Agric.* **2018**, *151*, 156–164. [CrossRef]
15. Bolya, D.; Zhou, C.; Xiao, F.; Lee, Y.J. YOLACT: Real-Time Instance Segmentation. *arXiv* **2019**, arXiv:1904.02689.
16. Maaz, M.; Shaker, A.; Cholakkal, H.; Khan, S.; Zamir, S.W.; Anwer, R.M.; Khan, F.S. EdgeNeXt: Efficiently Amalgamated CNN-Transformer Architecture for Mobile Vision Applications. *arXiv* **2022**, arXiv:2206.10589.
17. Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. *arXiv* **2022**, arXiv:2201.03545.
18. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. *arXiv* **2018**, arXiv:1803.01534.
19. Hou, Q.; Zhang, L.; Cheng, M.-M.; Feng, J. Strip Pooling: Rethinking Spatial Pooling for Scene Parsing. *arXiv* **2020**, arXiv:2003.13328.
20. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *arXiv* **2020**, arXiv:1910.03151.
21. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
22. Dorj, U.-O.; Lee, M.; Yun, S. An Yield Estimation in Citrus Orchards via Fruit Detection and Counting Using Image Processing. *Comput. Electron. Agric.* **2017**, *140*, 103–112. [CrossRef]
23. Srivastava, D.; Rajitha, B.; Agarwal, S.; Singh, S. Pattern-Based Image Retrieval Using GLCM. *Neural Comput. Appl.* **2020**, *32*, 10819–10832. [CrossRef]
24. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *arXiv* **2018**, arXiv:1703.06870.
25. Huang, Z.; Huang, L.; Gong, Y.; Huang, C.; Wang, X. Mask Scoring R-CNN. *arXiv* **2019**, arXiv:1903.00241.
26. Bolya, D.; Zhou, C.; Xiao, F.; Lee, Y.J. YOLACT++: Better Real-Time Instance Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 1108–1121. [CrossRef] [PubMed]
27. Wang, X.; Zhang, R.; Kong, T.; Li, L.; Shen, C. SOLOv2: Dynamic and Fast Instance Segmentation. *arXiv* **2020**, arXiv:2003.10152.
28. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. *arXiv* **2018**, arXiv:1807.06521.
29. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. *arXiv* **2021**, arXiv:2103.02907.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.