

Article

Sichuan Pepper Recognition in Complex Environments: A Comparison Study of Traditional Segmentation versus Deep Learning Methods

Jinzhong Lu ^{1,2,*}, Juncheng Xiang ^{1,2}, Ting Liu ^{1,2}, Zongmei Gao ³  and Min Liao ^{1,2}

¹ Modern Agricultural Equipment Research Institute, Xihua University, Chengdu 610039, China

² School of Mechanical Engineering, Xihua University, Chengdu 610000, China

³ Center for Precision and Automated Agricultural Systems, Department of Biological Systems Engineering, Washington State University, Prosser, WA 99350, USA

* Correspondence: lujinzhong@mail.xhu.edu.cn

Abstract: At present, picking Sichuan pepper is mainly undertaken by people, which is inefficient and presents the possibility of workers getting hurt. It is necessary to develop an intelligent robot for picking Sichuan peppers in which the key technology is accurate segmentation by means of mechanical vision. In this study, we first took images of Sichuan peppers (Hanyuan variety) in an orchard under various conditions of light intensity, cluster numbers, and image occlusion by other elements such as leaves. Under these various image conditions, we compared the ability of different technologies to segment the images, examining both traditional image segmentation methods (RGB color space, HSV color space, k-means clustering algorithm) and deep learning algorithms (U-Net convolutional network, Pyramid Scene Parsing Network, DeeplabV3+ convolutional network). After the images had been segmented, we compared the effectiveness of each algorithm at identifying Sichuan peppers in the various types of image, using the Intersection Over Union (IOU) and Mean Pixel Accuracy (MPA) indexes to measure success. The results showed that the U-Net algorithm was the most effective in the case of single front-lit clusters light without occlusion, with an IOU of 87.23% and an MPA of 95.95%. In multiple front-lit clusters without occlusion, its IOU was 76.52% and its MPA was 94.33%. Based on these results, we propose applicable segmentation methods for an intelligent Sichuan pepper-picking robot which can identify the fruit in images from various growing environments. The research showed good accuracy for the recognition and segmentation of Sichuan peppers, which suggests that this method can provide technical support for the visual recognition of a pepper-picking robot in the field.

Keywords: Sichuan pepper; traditional segmentation; deep learning; picking robot



Citation: Lu, J.; Xiang, J.; Liu, T.; Gao, Z.; Liao, M. Sichuan Pepper Recognition in Complex Environments: A Comparison Study of Traditional Segmentation versus Deep Learning Methods. *Agriculture* **2022**, *12*, 1631. <https://doi.org/10.3390/agriculture12101631>

Academic Editors: Dimitrios Moshou and Xanthoula Eirini Pantazi

Received: 7 August 2022

Accepted: 5 October 2022

Published: 7 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 Sichuan pepper planting area in China is about 1.6667 million hectares. The annual output of Sichuan pepper is about 280,000 tons and the annual output value is about 14 billion yuan. Sichuan has 333,300 hectares, with an annual output of 100,000 tons of dry Sichuan pepper and a comprehensive output value of 8 billion yuan, making it the largest Sichuan pepper-producing province in China. Sichuan pepper is rich in nutrients and is of high medicinal value [1,2]. It can be classified into two varieties: green and red [3]. Green Sichuan pepper is obviously different from red Sichuan pepper in color and aroma. Moreover, the varieties are grown in different locations. One of the most important difference is the way they are picked. Since the branches of red Sichuan pepper cannot be cut down, mechanization of this task is more difficult, especially in terms of recognition by an intelligent picking robot. Sichuan pepper has small cells, dense growth, and is light in weight. At present, most Sichuan pepper is picked manually. Due to the low efficiency and high cost of workers, as well as the high risk of injury, it is important to develop an

intelligent picking robot. For such a robot, the ability to accurately identify Sichuan peppers by machine vision is an essential premise.

In recent years, specialists in fruit and vegetable identification and segmentation have carried a great deal of research. Some studies are based on traditional image segmentation methods, such as threshold-based segmentation, regional growth segmentation, segmentation based on edge detection, etc. These studies aimed at identifying the colors of fruits or vegetables against different colored backgrounds, as well as identifying their texture, shape, and other features of their appearance. Based on the characteristics of images of the bagging of green apples, Lv et al. [4] designed an image segmentation method based on R–B color difference, contrast limited adaptive histogram equalization (CLAHE), and Otsu method which together obtained more complete bagging of green apples. Yogesh [5] proposed an improved and optimal multi-level thresholding algorithm, and showed that the proposed algorithm not only reduced the processing time but also improved the segmentation technique. Anindita Septiarini et al. [6] proposed a contour-based image segmentation method for oil palm fruit to identify the shape and color of oil palm fruits, which showed an average segmentation accuracy of 90.13%. Jidong Lv et al. [7] used R–G, 2G–R–B, and other methods to split fruits, branches, and leaves from red apple orchard images. The results showed that this method effectively segmented fruits, branches, and leaves with an average relative ultimate measurement accuracy (RUMA) of 5.2%, 10%, and 13.9%, respectively.

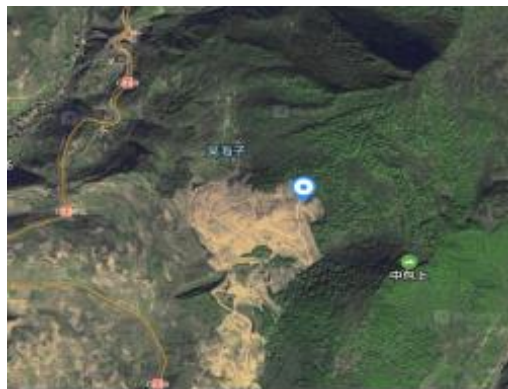
In the traditional segmentation method, the surface features of images are commonly used. However, these are not obvious in a complex field environment. With the development of deep learning, image processing methods have also been gradually increasing in number. The application of convolutional neural networks (CNNs) to image processing has led to a series of CNN-based image segmentation methods being proposed with the ability to greatly improve the accuracy of segmentation. Lucas M. Tassis et al. [8] proposed an integrated framework for automatically detecting lesion recognition and segmentation using two types of CNN. The Mask R-CNN achieved 71.9% accuracy and a recall of 71.90%, while U-Net and the pyramid Scene Parsing Network (PSPNet) achieved accuracies of 94.25% and 93.54%, respectively, indicating that the framework is suitable for implementation in an embedded platform. Qianwen Li [9] proposed an integrated U-Net segmentation model suitable for small sample datasets, and the experimental results showed that the proposed method can effectively improve the segmentation accuracy for target fruits and the model's ability to generalize. Weikuan Jia et al. [10] proposed an effective and accurate deep learning-based model, FoveaMask. Researchers tested it on green apple and immature persimmon datasets, and found that this model had higher recognition accuracy than 11 other types of detection and segmentation model. Khurram Hameed et al. [11] proposed a score-based mask edge improvement technique for Mask-RCNN image segmentation in supermarket-based environments, which demonstrated efficient segmentation of fruits and vegetables.

In this study, we generated a Sichuan pepper dataset and tested it under both a traditional segmentation algorithm and a deep learning algorithm. The experimental results demonstrated that we obtained an optimal algorithm for Sichuan pepper segmentation in complex environments. This optimal segmentation algorithm can be applied in the visual recognition system of a Sichuan pepper-picking robot. The results can also provide technical support for development of a Sichuan pepper-picking robot.

2. Materials and Methods

2.1. Image Acquisition

The image collection site was in Qingxi town, Hanyuan County, Ya'an city, Sichuan Province. It is located at 29°34' N and 102°37' E, and is shown in Figure 1a. The image collection environment is shown in Figure 1b. The "Hanyuan" variety of Sichuan pepper was used.



(a)



(b)

Figure 1. (a) Sichuan pepper planting location; (b) planting environment.

From 10 a.m. to 5 p.m. in 24 August 2021, 953 images were acquired with a total image resolution of 4032 pixels by 3024 pixels. In order to ensure the diversity of Sichuan pepper fruit samples, samples with varying light intensity, fruit numbers, and occlusion degrees were collected, as shown in Figure 2.



(a)



(b)



(c)



(d)

Figure 2. (a) Single front-lit cluster without occlusion; (b) multiple front-lit clusters without occlusion; (c) single back-lit cluster without occlusion; (d) single front-lit cluster and occlusion.

2.2. Sichuan Pepper Datasets

The experimental environment was the Windows10 operating system, Pycharm2020 was the platform, the OpenCV toolkit was installed, and the programming language was Python. LabelMe software was used to annotate red fruit images to make the datasets. First, the outline or edge of the red fruit was labeled using LabelMe; second, the image was divided into two parts, the red interior being the target fruit, the outside being the background, as shown in Figure 3; subsequently, all annotation information such as labels, annotation points and coordinates corresponding to the original image were saved as a json file. The json files were generated for the datasets using Python code. The pictures of Sichuan pepper were then divided into a training set and a test set at a ratio of 67:33, respectively, with the training set containing 638 images and the test set 315 images.



Figure 3. (a) Before marking; (b) after marking.

2.3. Traditional Segmentation Algorithm

Traditional image segmentation methods are simple and effective, producing good results with pictures in simple environments. The traditional segmentation algorithms used in this paper were RGB color space, HSV color space, and k-means clustering algorithm.

2.3.1. RGB Color Space Algorithm

Image thresholding segmentation has become one of the most commonly used basic image segmentation techniques, due to its simple implementation, small computational demand, and stable performance. It is particularly suitable for images where the target and background occupy different ranges of grayscale levels. It not only greatly compresses the amount of data, but also massively simplifies the analysis and processing steps, so in many cases, it has become a necessary image preprocessing process used before image analysis, feature extraction, and pattern recognition. The purpose of image thresholding is to divide the set of pixels into grayscale levels. Each subset obtained forms an area corresponding to the real scene. Each area has a consistent attribute inside, and the layout of adjacent areas also has a consistent attribute. This partition can be achieved by selecting one or more thresholds from the grayscale level.

Color is one of the important bases for the image segmentation of plants. The RGB is the most common color model. The RGB color model consists of three primary colors: red, green, and blue. The RGB color model is often used in thresholding for image segmentation. Because the mature Sichuan pepper is red and is clearly different from the color in the growing environment, the RGB color component can be used to segment Sichuan pepper. Red Sichuan peppercorns were extracted by setting the RGB values in the RGB color space. In the following equation, the segmentation image function is $h(x, y)$, the original image

is $f(x, y)$, and R is the threshold. The threshold segmentation algorithm then meets the following expressions:

$$h(x, y) = \begin{cases} f(x, y), & f(x, y) \geq R \\ 0, & f(x, y) < R \end{cases} \quad (1)$$

where $f(x, y)$ is the value of the red color component at point (x, y) , R is a fixed threshold, and $h(x, y)$ is the final value at the point (x, y) .

Using the RGB color component segmentation algorithm, the picture is divided into parts depicting Sichuan pepper in the foreground, and background elements in the field such as trees, leaves, and sky.

2.3.2. HSV Color Space Algorithm

HSV is a commonly used color segmentation method [12,13], where H describes the image hue, S describes the image saturation, and V, for value, describes the image brightness. For HSV color segmentation, the color space is first converted from RGB to HSV using the following algorithm:

$$R' = \frac{R}{255}, G' = \frac{G}{255}, B' = \frac{B}{255}, C_{max} = \max(R', G', B'), C_{min} = \min(R', G', B') \quad (2)$$

$$\Delta = C_{max} - C_{min} \quad (3)$$

$$H = \begin{cases} 0^\circ, & \Delta = 0 \\ 60^\circ * \left(\frac{G' - B'}{\Delta} + 0 \right), & C_{max} = R' \\ 60^\circ * \left(\frac{B' - R'}{\Delta} + 2 \right), & C_{max} = G' \\ 60^\circ * \left(\frac{R' - G'}{\Delta} + 4 \right), & C_{max} = B' \end{cases} \quad (4)$$

$$S = \begin{cases} \frac{\Delta}{C_{max}}, & C_{max} \neq 0 \\ 0, & C_{max} = 0 \end{cases} \quad (5)$$

$$V = C_{max} \quad (6)$$

where R , G , and B are the values of the RGB components of an image, R' , G' , B' after converting RGB values to between 0 and 1; C_{max} is the maximum value in R' , G' , B' and C_{min} is the minimum value in R' , G' , B' ; and Δ is the difference between C_{max} and C_{min} .

HSV is a color space created by A.R. Smith in 1978, based on the intuitive features of color. In the segmentation of our Sichuan pepper images, we mainly extracted the H component for segmentation. After converting the color space of the Sichuan pepper image into the HSV space, the H component was extracted and its threshold set. The values of red in the H component ranged from 0–10 and 156–180, and pixels with this H component between 0–10 and 156–180 were set to 0; the remaining pixels were became 255. After segmentation, the morphological corrosion operation was used to eliminate excess white dots.

2.3.3. K-Means Clustering Segmentation

The k-means clustering algorithm (k-means), an unsupervised clustering algorithm [14,15], is a very common way of classifying images by inter-sample distance. This method involves selecting K points from the image as the center of the image cluster, then calculating the distance from the center of the cluster to group the targets closest to the center of the cluster. Then the remaining data are calculated and the corresponding average found to identify the new cluster center. Finally, the distance is determined and analyzed until there is no significant difference between the results obtained by the two operations, which signifies the best clustering effect. Individual data are then partitioned into specified clusters by iterative search. Assuming that the target is to be clustered into K classes, the k-means clustering algorithm steps are as follows:

- (1) Select an appropriate central value of the K classes.

- (2) In the n th iteration, find the distance from any sample to the K cluster centers for any sample, and take the class where the center with the shortest orbital distance is located.
- (3) The central value of this class is updated using the mean method.
- (4) For all the n cluster centers, repeat steps (2) and (3); the iteration ends if the cluster center value remains constant, otherwise the iteration continues.

In the k -means clustering algorithm, the classification number K has a great impact on the image clustering effect.

2.4. Deep Learning Algorithm

2.4.1. PSPnet Algorithm

The main feature of the PSPNet is the use of the pyramid Scene Parsing (PSP) module [16,17]. The pyramid pooling module proposed in this model can aggregate contextual information in different regions, thus improving the algorithm's ability to obtain global information. Such a priori representations are valid and have shown excellent results on multiple datasets. The function of the PSP structure is to divide the acquired feature layers into different sized grids of each grid, and each grid is averaged by pooling separately. Contextual information in different regions is aggregated to improve the algorithm's ability to obtain global information.

In PSPnet, in the typical case of the PSP structure, the input feature layer is divided into 6×6 , 3×3 , 2×2 , and 1×1 grids, corresponding to the green, blue, orange, and red output in the picture, as shown in Figure 4.

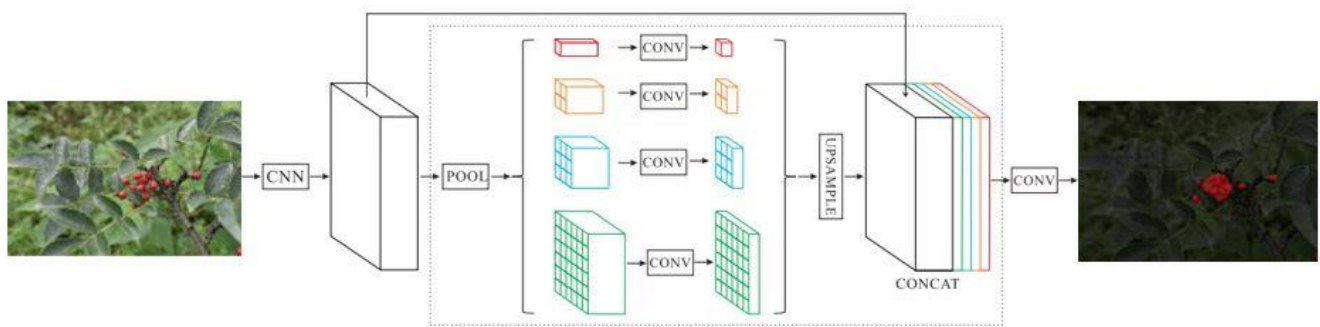


Figure 4. The PSPnet algorithm. The CNN is Convolutional Neural Network; the CONV is convolution; the CONCAT is stitching of feature maps together.

Given a Sichuan pepper input image (a), we first use CNN to obtain a feature map of the last convolution layer (b), then use the pyramid resolution module (c) to obtain representations of the different subregions, and then up-sample and connect the layers to form the final feature representation. The pyramid pooling module carries both local and global context information. Finally, the representation is sent into the convolution layer to obtain the final pixel-by-pixel prediction (d).

2.4.2. U-Net Algorithm

U-Net is an excellent semantic segmentation model with a main execution process similar to that of other semantic segmentation models [18]. The U-Net model can be divided into three stages, as shown in Figure 5:

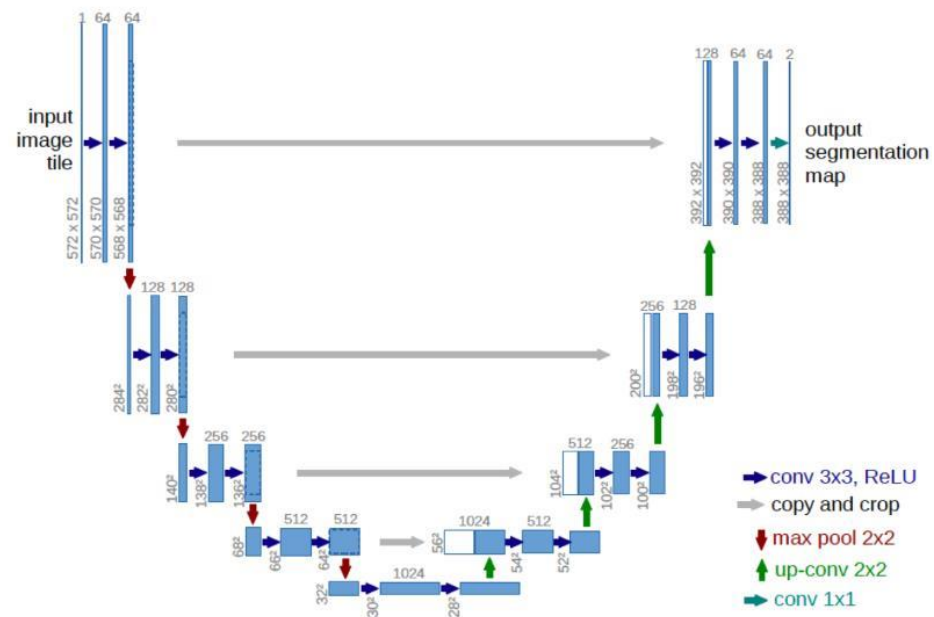


Figure 5. The U-Net algorithm.

The first stage is backbone feature extraction, where the backbone is used to obtain one feature layer after another. U-Net's backbone feature extraction function works in a similar way to visual geometry group (VGG) in terms of convolution and maximal pooling of stacking [19,20], as shown in Figure 6. Using backbone feature extraction, we can obtain five preliminary effective feature layers, and in the second step, we can use these five layers for feature fusion.

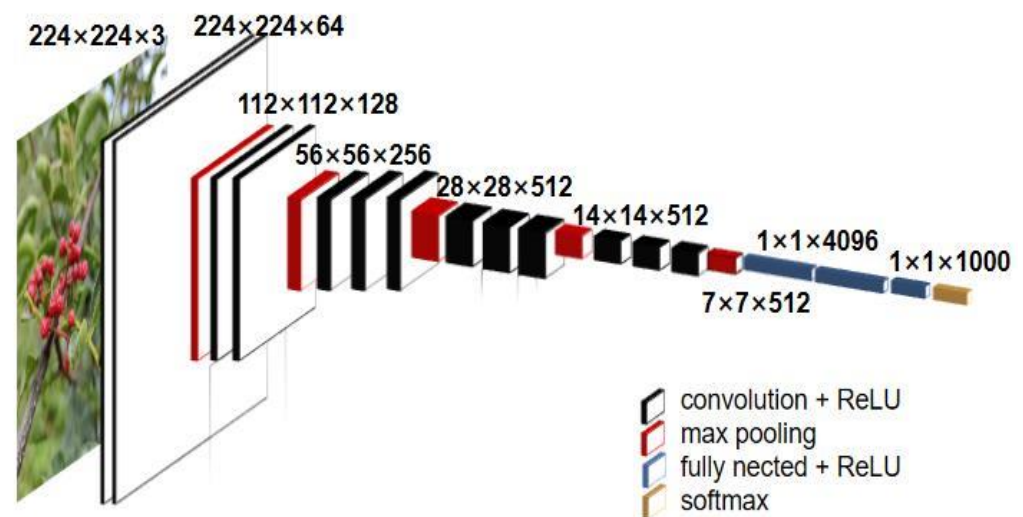


Figure 6. The VGG model.

The second stage involves strengthening the feature extraction results. The five preliminary effective feature layers obtained in the previous stage are used for up-sampling, and their features are fused to obtain a final, effective feature layer integrating all features.

The third stage is prediction. The final effective feature layer is used to classify each feature point, treated as equivalent to each pixel point.

2.4.3. DeepLabV3+ Algorithm

The DeepLab series model is a deep learning convolutional neural network model proposed by Chen et al., at the core of which is the use of Dilated Convolution [21,22].

DeepLab V3+ convolutional network is an improvement to DeepLab V3, with the feature resolution of atrous spatial pyramid pooling (ASPP) being eight times (or 16 times) greater, even with Dilated Convolution. However, it directly up-samples the 1/8 resolution result map at the end of DeepLab V3 to the original resolution size, obtaining pixel-by-pixel segmentation results. This direct up-sampling operation (as understood as naive decoder) does not fully recover the details lost during the down-sampling to 1/8 resolution, thus causing imprecise segmentation.

Therefore, DeeplabV3+ improves on two points: (1) it uses a coded design architecture, and (2) the change of backbone explores the feasibility of replacing ResNet-101 with the Xception model. Depthwise separable convolution is used to further improve the accuracy and speed of the segmentation algorithm, as shown in Figure 7. DeeplabV3+ introduces a large degree of void convolution in the coding, which expands the receptive field through the void convolution and can effectively avoid the ground information loss caused by the pooling operation. Figure 8 presents a schematic diagram of void convolution, where voids are extracted across pixels.

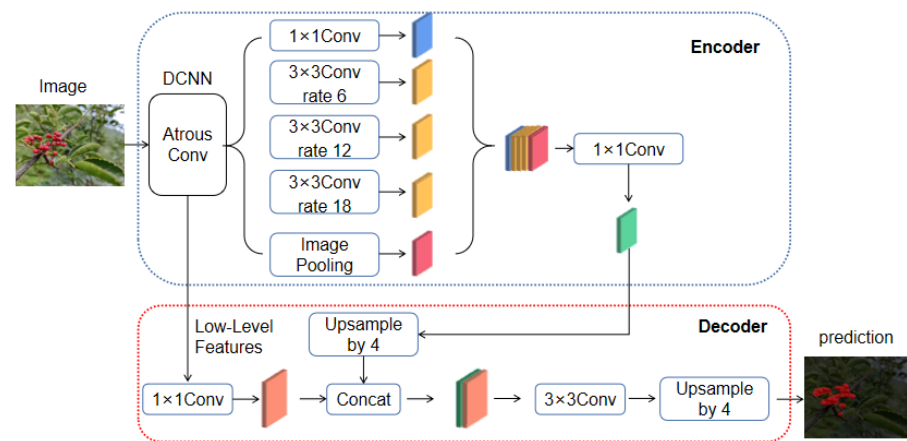


Figure 7. The DeepLabV3+ model.

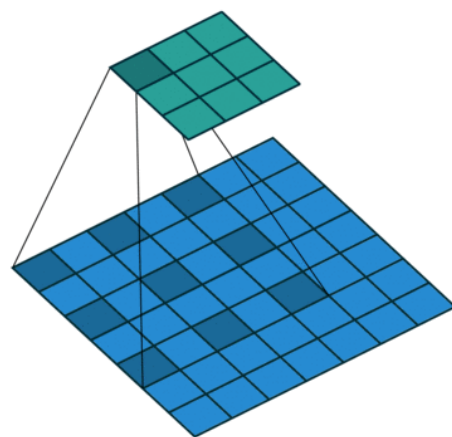


Figure 8. Dilated Convolution.

Figure 9 shows the experimental process used in this study. After screening the Sichuan pepper pictures and removing fuzzy shots, the remaining Sichuan pepper pictures were entered into the production dataset and then run through the traditional algorithm and deep learning algorithm. After the experiment, the highest segmentation accuracy of these methods was calculated.

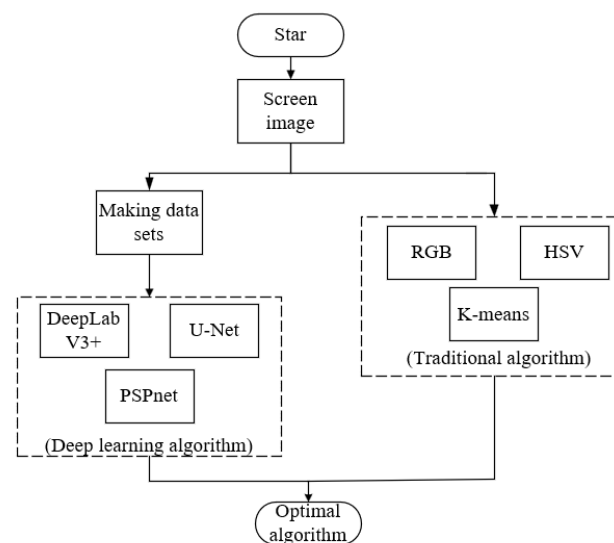


Figure 9. Flow chart of Sichuan pepper image segmentation methodology.

2.5. Partition Accuracy and Evaluation Criteria

In this study, the evaluation index used in the traditional method was Intersection over Union (IOU), the evaluation index used in the deep learning algorithm was IOU and average pixel accuracy was the mean pixel accuracy (MPA).

Precision is the most important and popular technical index when evaluating an image semantic segmentation network. Precision estimation methods vary, but they can be divided into two categories, one based on IOU, and the other based on pixel accuracy. At present, the most popular semantic segmentation evaluation methods are based on pixel markers.

Assuming a total $k + 1$ classification (marked L_0 to L_k with a background category), p_{ii} means pixels of category i are predicted as TP (true positives); p_{jj} means j as TN (true negatives) as category j , p_{ij} means i as FP (false positives) for category j , and p_{ji} as FN (false negatives) for category i .

The IOU is another standard evaluation index of the semantic segmentation network. It involves recalculating the ratio between the sum of TP and (TP + FN + FP):

$$\text{IOU} = \frac{\sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}}}{k} \quad (7)$$

where k is the number of categories (excluding background); p_{ii} is the number of pixels belonging to class i and assigned to class i ; p_{ij} is the number of pixels that belong to class i but assigned to class j ; p_{jj} is the number of pixels belonging to class j and assigned to class j ; and p_{ji} is the number of pixels that belong to class j but are assigned to class i .

MPA is the mean based on the sum of the ratio of the correct total number of pixels for each category to the total number of each category:

$$\text{MPA} = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}} \quad (8)$$

3. Results

3.1. Sichuan Pepper Segmentation with Varying Light Intensity

Figures 10 and 11 show the segmentation results of the selected single cluster under front lighting and backlighting conditions. Figures 10a and 11A show an original Sichuan pepper image; Figures 10b and 11B are the segmentation results based on the RGB algorithm; Figures 10c and 11C are the segmentation results based on the HSV algorithm;

Figures 10d and 11D are the segmentation results based on the k-means cluster segmentation algorithm; Figures 10e and 11E are the segmentation results based on the U-Net algorithm; Figures 10f and 11F are the segmentation results based on the DeepLabV3+ algorithm; and Figures 10g and 11G shows the segmentation results based on the PSPnet algorithm. From the segmentation results shown in Figure 10, in the single front-lit cluster of Sichuan peppers without occlusion, all segmentation methods were able to segment the area represented by Sichuan pepper. In the case of a single backlit cluster without occlusion, the RGB algorithm could only segment a few Sichuan peppercorns, as shown in Figure 11B. The reason for this is likely to lie in the fact that the brightness levels did not meet the values required to distinguish the red, green and blue components of the image, causing Sichuan peppercorns in the image to be misidentified as background.

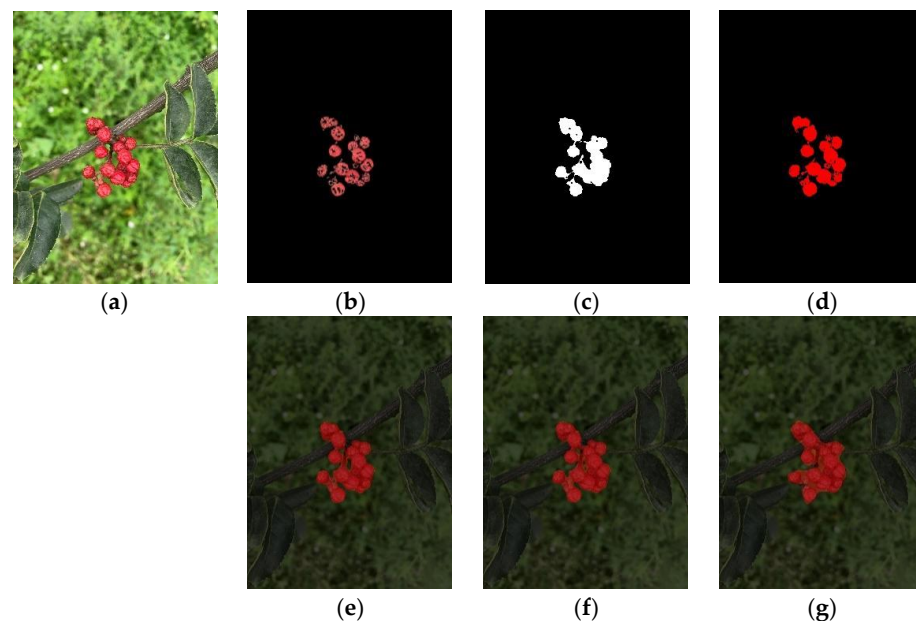


Figure 10. (a) Single front-lit cluster without occlusion; (b) results from RGB algorithm for segmentation; (c) results from HSV algorithm for segmentation; (d) results from k-means algorithm for segmentation; (e) results from U-Net algorithm for segmentation; (f) results from DeepLabV3+ algorithm for segmentation; (g) results from PSPnet algorithm for segmentation.

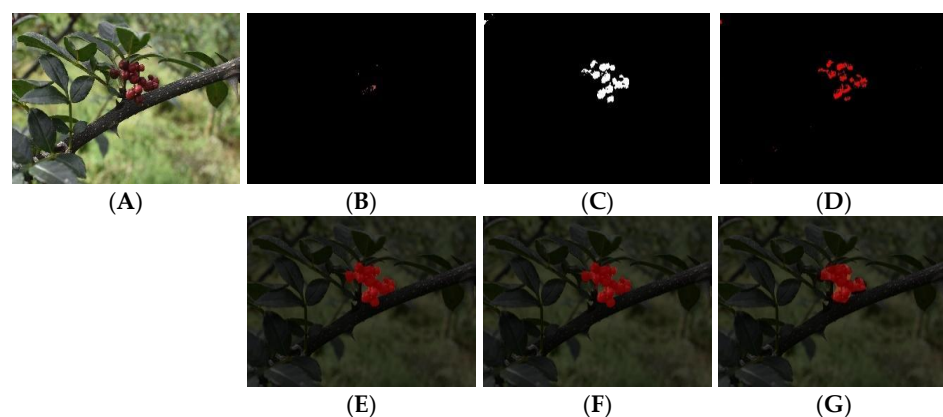


Figure 11. (A) Single backlit cluster without occlusion; (B) results from RGB algorithm for segmentation; (C) results from HSV algorithm for segmentation; (D) results from k-means algorithm for segmentation; (E) results from U-Net algorithm for segmentation; (F) results from DeepLabV3+ algorithm for segmentation; (G) results from PSPnet algorithm for segmentation.

3.2. Sichuan Pepper Segmentation with Occlusion

As can be seen from Figure 12a–g, the traditional algorithm, the RGB method, could only identify a small proportion of the Sichuan pepper. The HSV and k-means algorithms could accurately identify Sichuan pepper, and could also accurately discern a single Sichuan pepper fruit in a cluster, but the k-means method required readjusting K value to accurately segment the Sichuan pepper. The deep learning methods, U-Net and DeepLabV3+, were also able to accurately classify the details of Sichuan peppers. PSPnet accurately divided most of the targets, but it could not identify the edges of Sichuan peppercorns, so it was not able to divide the Sichuan pepper completely.

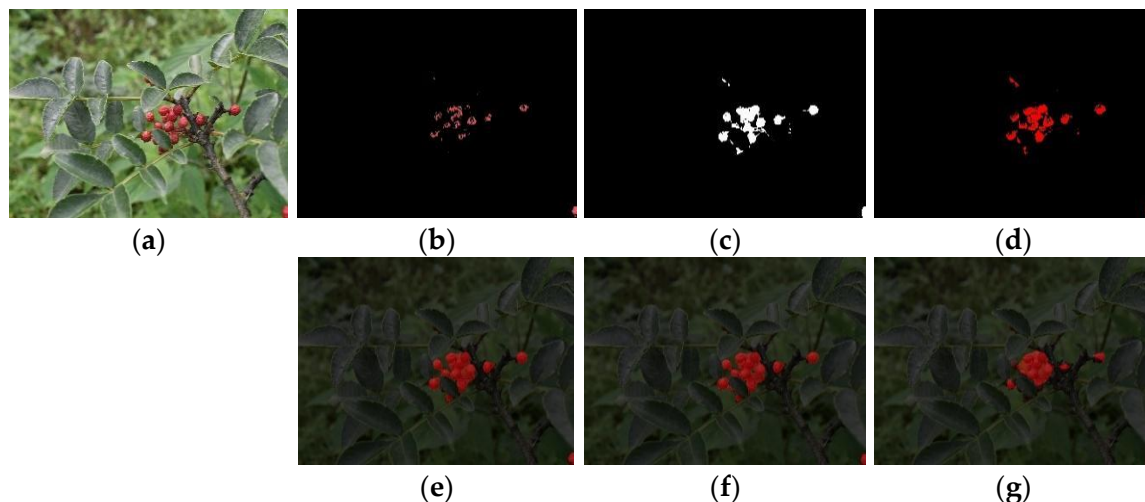


Figure 12. (a) Single front-lit cluster with occlusion; (b) results from RGB algorithm for segmentation; (c) results from HSV algorithm for segmentation; (d) results from k-means algorithm for segmentation; (e) results from U-Net algorithm for segmentation; (f) results from DeepLabV3+ algorithm for segmentation; (g) results from PSPnet algorithm for segmentation.

3.3. Segmentation with Different Numbers of Sichuan Pepper Clusters

As shown in Figure 13a–g, varying the number of fruit clusters had little effect on the algorithms' ability to segment Sichuan pepper. Both the traditional algorithms and deep learning methods were able to segment the Sichuan pepper, and the success rates were consistent with those for images with no occlusion.

3.4. Test Results of the Algorithm

In this study, the index used to evaluate traditional segmentation was IOU, and the indexes used to evaluate deep learning were IOU and MPA.

According to the experimental results shown in Table 1, among the three traditional segmentation methods, HSV showed the best the overall segmentation. In the case of front-lit single clusters without occlusion, the IOU value of HSV was 84.99%, which was 13.10% and 9.69% higher than for the RGB and k-means algorithms, respectively. The IOU of HSV for back-lit single clusters without occlusion was 70.57%, which was 59.87% and 0.09% more than RGB and k-means, respectively.

According to the experimental results shown in Table 2, among the deep-learning algorithms, the overall segmentation ability of the U-Net algorithm was the best. In the case of single front-lit clusters without occlusion, the IOU was 87.23%, which was 13.95% and 1.22% higher than for PSPnet and DeepLabV3+, respectively. The IOU was 76.52% in the case of multiple front-lit clusters without occlusion, which was 18.98% and 1.13% more than for PSPnet and DeepLabV3+, respectively.

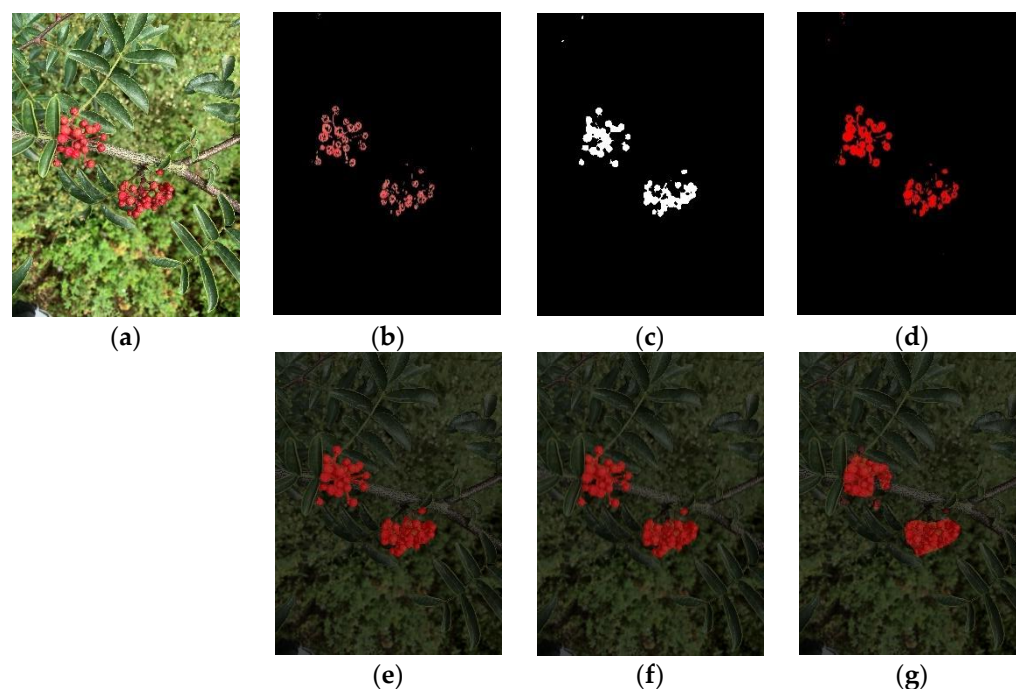


Figure 13. (a) Multiple front-lit clusters without occlusion; (b) results from RGB algorithm for segmentation; (c) results from HSV algorithm for segmentation; (d) results from k-means algorithm for segmentation; (e) results from U-Net algorithm for segmentation; (f) results from DeepLabV3+ algorithm for segmentation; (g) results from PSPnet algorithm for segmentation.

Table 1. Comparison of IOU for traditional image segmentation methods.

Traditional Algorithm	Intersection Over Union, IOU (%)			
	Single Front-Lit Clusters without Occlusion	Multiple Front-Lit Clusters without Occlusion	Single Back-Lit Clusters without Occlusion	Single Front-Lit Clusters with Occlusion
RGB color space	71.89	68.01	10.7	47.68
HSV color space	84.99	80.92	70.57	82.84
k-means	75.30	78.29	70.58	73.32

Table 2. The IOU comparison of deep learning segmentation methods.

Deep Learning Algorithm	IOU (%)			
	Single Front-Lit Clusters without Occlusion	Multiple Front-Lit Clusters without Occlusion	Single Back-Lit Clusters without Occlusion	Single Front-Lit Clusters with Occlusion
PSPnet	73.28	57.54	67.32	68.73
U-Net	87.23	76.52	83.47	84.71
DeepLabV3+	86.01	75.39	82.13	81.25

According to the experimental results in Table 3, the effect of the U-Net deep learning-based segmentation algorithm was the best. The MPA reached 95.95% in the case of front-lit single clusters without occlusion and 93.85% in the case of single back-lit clusters without occlusion.

Table 3. The MPA comparison of deep learning segmentation methods.

Deep Learning Algorithm	Mean Pixel Accuracy, MPA (%)			
	Single Front-Lit Clusters without Occlusion	Multiple Front-Lit Clusters without Occlusion	Single Back-Lit Clusters without Occlusion	Single Front-Lit Clusters with Occlusion
PSPnet	89.63	83.57	86.78	87.48
U-Net	95.95	94.33	93.85	94.11
DeepLabV3+	95.10	92.48	92.97	92.95

4. Discussion

Table 4 compares recent studies on various segmentation methods applied to fruit. It compares the task, datasets, and methods, and lists the advantages and disadvantages of each to better explain the experimental results obtained in this paper.

Table 4. Comparison of studies on various segmentation methods applied to fruit.

No	Reference	Task	Dataset	Methods	Pros and Cons
1	Davinia et al. (2015) [23]	Estimating vineyard yield at night	Images from a grape orchard	RGB and HSV color spaces	Manual control of light, reduced the impact of light, improved the segmentation effect
2	Wang et al. (2017) [24]	Applying a robust fruit segmentation algorithm against varying illumination	300 images under outdoor conditions captured in three orchards	k-means, Retinex-based image enhancement algorithm	The k-means segmentation effect was better when using an illumination normalization algorithm and image enhancement
3	Lv et al. (2019) [25]	Obtaining near-large fruit from apple image in orchard	Images from the apple planting demonstration area	RGB color space	Algorithm took less time
4	Zhang et al. (2020) [26]	Applying an apple segmentation algorithm in an orchard	105 images from a Science and Technology Park	k-means, R–G color difference method	Reduced the computational resource burden to the greatest extent.
5	Peng et al. (2021) [27]	Segmentation of grape clusters with different varieties	300 images from a grape orchard	Fully convolutional networks(FCN), U-Net, DeepLabv3+	The IOU of the three networks was no greater than 85% but all of them were better than traditional networks
6	Qi et al. (2022) [28]	Detecting accurate picking locations on the main stems	Lychee images from the Internet	Pyramid Scene Parsing Network(PSPnet), DeepLabv3+, U-Net	When there were multiple clusters of lychees in the image, the IOU values in the three models were lower than 60%
7	Chen et al. (2021) [29]	Segmenting various kinds of grapes in a field environment	1856 images from wine grape production demonstration	PSPnet, DeepLabv3+, U-Net	When the bunches on the grape images were relatively discrete, the model could not accurately and completely segment the berry regions.
8	Kyamelia et al. (2020) [30]	Detection of rotten or fresh apple	4035 images from Kaggle	U-Net, Enhanced Unet(EN-U-Net)	U-Net achieved training and validation accuracies of 93.19% and 95.36%, respectively
9	This study	Optimal segmentation algorithm for Sichuan pepper in complex environment	953 images from Hanyuan Sichuan pepper based in Ya'an	RGB, HSV color space, k-means, PSPnet, U-Net, DeepLabv3+	The traditional segmentation algorithm was affected by illumination and the segmentation effect was poor, the U-Net segmentation algorithm was the best

In this experiment, the segmentation abilities of the RGB, HSV and k-means algorithms were found to be greatly affected by the illumination. Some properties of the Sichuan pepper image itself can also affect traditional segmentation outcomes. For example, the exposure of Sichuan peppers to light caused the oil cells on the peppercorns' surface to blacken, as shown in Figure 14. This resulted in a small proportion of the Sichuan peppers in the images segmented as background. According to some studies, light sources can be manually controlled to reduce the influence of illumination on the segmentation effect [23]. There are image enhancement methods, such as Retinex-based image enhancement, that can be used to remove the influence of irradiation by light [24]. After making the above changes, the effect of the traditional segmentation algorithm was improved.

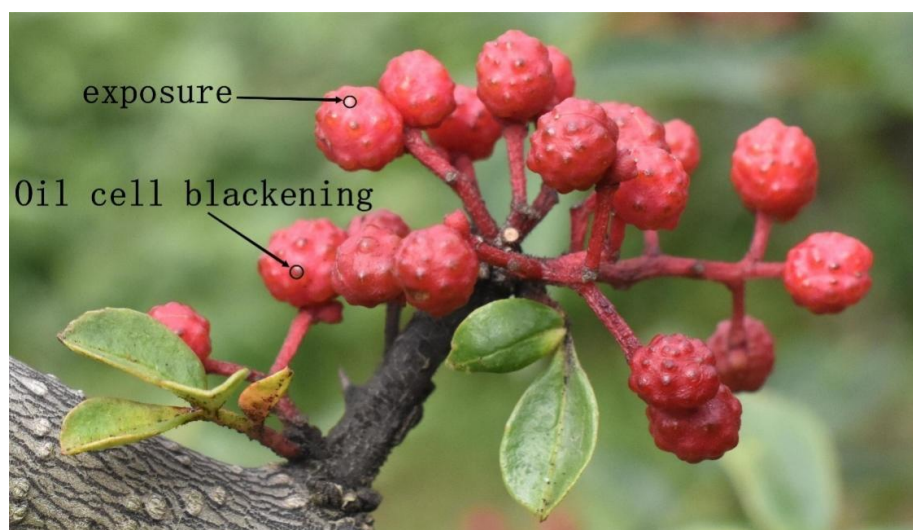


Figure 14. Picture of Sichuan pepper surface exposure, with oil cells turning black.

Another characteristic of traditional segmentation algorithms is that they require fewer parameters and take less time [25,26]. In segmentation algorithms based on deep learning, the effect is better than with traditional methods. The IOU obtained by Peng et al. [27] with their deep learning segmentation algorithm was lower than 85%, but the results were all better than when using a traditional segmentation algorithm, which is consistent with the experimental results in this paper. However, the process of training deep learning segmentation algorithms is complicated. It requires manual calibration of images and the production of datasets, which takes a lot of time. With the deep learning networks, the segmentation effect of the three network models for multi-cluster Sichuan pepper was poor [28]. This may be because the position of the Sichuan peppers is relatively discrete in multi-cluster images. This finding is similar to that of Chen et al. [29], who used deep learning to segment an image with multiple clusters of grapes and relatively discrete locations. The model could not completely and accurately segment the grapes.

In our study, the U-Net segmentation algorithm worked best among all the segmentation algorithms. U-Net architecture is a category of fully convolutional network performing down-sampling and up-sampling. The down-sampling and up-sampling processes are connected with a concatenation operator and hence maintain the architecture's symmetry. The U-Net architecture predicts a good segmentation map, combining the localization and contextual information from the sampling process [30].

5. Conclusions

In this study, we collated a dataset of Sichuan pepper images under various light intensities, numbers of clusters, and occlusion conditions. We tested and compared three traditional methods (RGB, HSV, k-means) and three deep learning algorithms (PSPnet, U-Net, DeepLabV3+) to divide the images into areas containing Sichuan pepper and those containing background elements. We aimed to find appropriate segmentation methods for recognizing Sichuan pepper and segmenting images accordingly to solve visual recognition segmentation problems in the field. We found that HSV segmentation worked the best among traditional segmentation methods, with IOUs of 84.99%, 80.92%, 70.57%, and 82.84% in single front-lit clusters without occlusion, multiple front-lit clusters without occlusion, single back-lit clusters without occlusion, and single front-lit clusters with occlusion, respectively. Among the deep learning algorithms, U-Net was the most effective; the IOU of single front-lit clusters without occlusion, multiple front-lit clusters without occlusion, single back-lit clusters without occlusion, and single front-lit clusters front lighting were 87.23%, 76.52%, 83.47%, and 84.71%, respectively. The MPA values were 95.95%, 94.33%, 93.85%, and 94.11%, respectively.

The research demonstrated good accuracy for the recognition and segmentation of Sichuan pepper, a method which can provide technical support for the visual recognition systems of a pepper-picking robot in the field. Deep learning algorithms had higher accuracy in the segmentation of images of Sichuan peppers, bringing potential advantages for applications in the field. However, the accuracy of algorithms still needs to be improved. Research on better algorithms that can improve segmentation accuracy and solve the problem of Sichuan pepper exposure and the resultant blackening of oil cells could be a future research direction. These issues are of important for the future of intelligent Sichuan pepper-picking robots.

Author Contributions: Conceptualization, J.L. and M.L.; formal analysis, T.L.; methodology, J.X.; software, J.X.; validation, J.L., J.X. and Z.G.; writing—original draft, J.L. and J.X.; writing—review and editing, T.L. and Z.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Sichuan Provincial Science and Technology Department (2021YFN0020).

Data Availability Statement: Not applicable.

Acknowledgments: We thank Lijuan Tan, Chunlin Chen from the School of Mechanical Engineering, Xihua University and for their help with this research.

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

References

1. Gong, Y.; Sun, W.-H.; Xu, T.-T.; Zhang, L.; Huang, X.-Y.; Tan, Z.-H.; Di, D.-L. Chemical constituents from the pericarps of *Zanthoxylum bungeanum* and their chemotaxonomic significance. *Biochem. Syst. Ecol.* **2021**, *95*, 104213. [\[CrossRef\]](#)
2. Zhang, D.; Sun, X.; Battino, M.; Wei, X.; Shi, J.; Zhao, L.; Liu, S.; Xiao, J.; Shi, B.; Zou, X. A comparative overview on chili pepper (capsicum genus) and sichuan pepper (zanthoxylum genus): From pungent spices to pharma-foods. *Trends Food Sci. Technol.* **2021**, *117*, 148–162. [\[CrossRef\]](#)
3. Zheng, T.; Zhang, Q.; Su, K.X.; Liu, S.M. Transcriptome and metabolome analyses reveal the regulation of peel coloration in green, red Chinese prickly ash (*Zanthoxylum* L.). *Food Chem. Mol. Sci.* **2020**, *1*, 100004. [\[CrossRef\]](#)
4. Lv, J.; Wang, F.; Xu, L.; Ma, Z.; Yang, B. A segmentation method of bagged green apple image. *Sci. Hortic.* **2019**, *246*, 411–417. [\[CrossRef\]](#)
5. Yogesh; Dubey, A.K.; Agarwal, A.; Sarkar, A.; Arora, R. Adaptive thresholding based segmentation of infected portion of pome fruit. *J. Stat. Manag. Syst.* **2017**, *20*, 575–584. [\[CrossRef\]](#)
6. Septiarini, A.; Hamdani, H.; Hatta, H.R.; Anwar, K. Automatic image segmentation of oil palm fruits by applying the contour-based approach. *Sci. Hortic.* **2020**, *261*, 108939. [\[CrossRef\]](#)
7. Lv, J.; Xu, L. Method to acquire regions of fruit, branch and leaf from image of red apple in orchard. *Mod. Phys. Lett. B* **2017**, *31*, 19–21. [\[CrossRef\]](#)
8. Tassis, L.M.; Tozzi de Souza, J.E.; Krohling, R.A. A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images. *Comput. Electron. Agric.* **2021**, *186*, 106191. [\[CrossRef\]](#)
9. Li, Q.; Jia, W.; Sun, M.; Hou, S.; Zheng, Y. A novel green apple segmentation algorithm based on ensemble U-Net under complex orchard environment. *Comput. Electron. Agric.* **2021**, *180*, 105900. [\[CrossRef\]](#)
10. Jia, W.; Zhang, Z.; Shao, W.; Hou, S.; Ji, Z.; Liu, G.; Yin, X. FoveaMask: A fast and accurate deep learning model for green fruit instance segmentation. *Comput. Electron. Agric.* **2021**, *191*, 106488. [\[CrossRef\]](#)
11. Hameed, K.; Chai, D.; Rassau, A. Score-based mask edge improvement of Mask-RCNN for segmentation of fruit and vegetables. *Expert Syst. Appl.* **2022**, *190*, 116205. [\[CrossRef\]](#)
12. Danish, M.; Akhtar, M.N.; Hashim, R.; Saleh, J.M.; Bakar, E.A. Analysis using image segmentation for the elemental composition of activated carbon. *MethodsX* **2020**, *7*, 100983. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Li, Z.; Yu, Z.; Liu, W.; Xu, Y.; Zhang, D.; Cheng, Y. Tongue image segmentation via color decomposition and thresholding. *Concurr. Comput. Pract. Exp.* **2018**, *31*, 4662.
14. Steinley, D. K-means clustering: A half-century synthesis. *Br. J. Math. Stat. Psychol.* **2006**, *59*, 1–34. [\[CrossRef\]](#)
15. Kumar, K.V.; Jayasankar, T. An identification of crop disease using image segmentation. *Int. J. Pharm. Sci. Res.* **2019**, *10*, 1054–1064.
16. Shaaban, A.M.; Salem, N.M.; Al-atabany, W.I. A Semantic-based Scene segmentation using convolutional neural networks. *AEU—Int. J. Electron. Commun.* **2020**, *125*, 153364. [\[CrossRef\]](#)
17. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR) 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.

18. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; Volume 9351, pp. 234–241.
19. Yue, J.; Zhu, L.; Li, F.; Li, Z. Vegetable Recognition and Classification Based on Improved VGG Deep Learning Network Model. *Int. J. Comput. Intell. Syst.* **2020**, *13*, 559–564.
20. Zhang, D.; Lv, J.; Cheng, Z. An Approach Focusing on the Convolutional Layer Characteristics of the VGG Network for Vehicle Tracking. *IEEE Access* **2020**, *8*, 112827–112839. [[CrossRef](#)]
21. Peng, H.; Xue, C.; Shao, Y.; Chen, K.; Xiong, J.; Xie, Z.; Zhang, L. Semantic Segmentation of Litchi Branches Using DeepLabV3+ Model. *IEEE Access* **2020**, *8*, 164546–164555. [[CrossRef](#)]
22. Zhang, K.; Liu, X.; Chen, Y. Research on Semantic Segmentation of Portraits Based on Improved Deeplabv3+. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *806*, 012057. [[CrossRef](#)]
23. Font, D.; Tresanchez, M.; Martinez, D.; Moreno, J.; Clotet, E.; Palacin, J. Vineyard yield estimation based on the analysis of high resolution images obtained with artificial illumination at night. *Sensors* **2015**, *15*, 8284–8301. [[CrossRef](#)]
24. Wang, C.; Tang, Y.; Zou, X.; SiTu, W.; Feng, W. A robust fruit image segmentation algorithm against varying illumination for vision system of fruit harvesting robot. *Optik* **2017**, *131*, 626–631. [[CrossRef](#)]
25. Lv, J.; Wang, Y.; Xu, L.; Gu, Y.; Zou, L.; Yang, B.; Ma, Z. A method to obtain the near-large fruit from apple image in orchard for single-arm apple harvesting robot. *Sci. Hortic.* **2019**, *257*, 108758. [[CrossRef](#)]
26. Zhang, C.; Zou, K.; Pan, Y. A Method of Apple Image Segmentation Based on Color-Texture Fusion Feature and Machine Learning. *Agronomy* **2020**, *10*, 972. [[CrossRef](#)]
27. Peng, Y.; Wang, A.; Liu, J.; Faheem, M. A Comparative Study of Semantic Segmentation Models for Identification of Grape with Different Varieties. *Agriculture* **2021**, *11*, 997. [[CrossRef](#)]
28. Qi, X.; Dong, J.; Lan, Y.; Zhu, H. Method for Identifying Litchi Picking Position Based on YOLOv5 and PSPNet. *Remote Sens.* **2022**, *14*, 2004. [[CrossRef](#)]
29. Chen, S.; Song, Y.; Su, J.; Fang, Y.; Shen, L.; Mi, Z.; Su, B. Segmentation of field grape bunches via an improved pyramid scene parsing network. *Int. J. Agric. Biol. Eng.* **2021**, *14*, 185–194. [[CrossRef](#)]
30. Roy, K.; Chaudhuri, S.S.; Pramanik, S. Deep learning based real-time Industrial framework for rotten and fresh fruit detection using semantic segmentation. *Microsyst. Technol.* **2020**, *27*, 3365–3375. [[CrossRef](#)]