Article

A Vision-Based Method Utilizing Deep Convolutional Neural Networks for Fruit Variety Classification in Uncertainty Conditions of Retail Sales

Rudnik Katarzyna [1],* and Michalski Paweł [2]

[1]   Faculty of Production Engineering and Logistics, Opole University of Technology, 45-758 Opole, Poland
[2]   Faculty of Electrical Engineering Automatic Control and Informatics, Opole University of Technology, 45-758 Opole, Poland; p.michalski@po.opole.pl
*   Correspondence: k.rudnik@po.opole.pl

check for
updates

Abstract: This study proposes a double-track method for the classification of fruit varieties for application in retail sales. The method uses two nine-layer Convolutional Neural Networks (CNNs) with the same architecture, but different weight matrices. The first network classifies fruits according to images of fruits with a background, and the second network classifies based on images with the ROI (Region Of Interest, a single fruit). The results are aggregated with the proposed values of weights (importance). Consequently, the method returns the predicted class membership with the Certainty Factor (*CF*). The use of the certainty factor associated with prediction results from the original images and cropped ROIs is the main contribution of this paper. It has been shown that *CF*s indicate the correctness of the classification result and represent a more reliable measure compared to the probabilities on the CNN outputs. The method is tested with a dataset containing images of six apple varieties. The overall image classification accuracy for this testing dataset is excellent (99.78%). In conclusion, the proposed method is highly successful at recognizing unambiguous, ambiguous, and uncertain classifications, and it can be used in a vision-based sales systems in uncertain conditions and unplanned situations.

Keywords: convolutional neural network; deep neural network; fruit classification; fruit recognition; certainty factor

1. Introduction

Recognizing different kinds of fruits and vegetables is perhaps the most difficult task in supermarkets and fruit shops [1]. Retail sales systems based on bar code identification require the seller (cashier) to enter the unique code of the given fruit or vegetable because they are individually sold by weight. This procedure often leads to mistakes because the seller must correctly recognize every type of vegetable and fruit; a significant challenge even for highly-trained employees. A partial solution to this problem is the introduction of an inventory with photos and codes. Unfortunately, this requires the cashier to browse the catalog during check-out, extending the time of the transaction. In the case of self-service sales, the species (types) and varieties of fruits must be specified by the buyer. Unsurprisingly, this can often result in the misidentification of fruits by buyers (e.g., Conference pear instead of Bartlett pear). Independent indication of the product, in addition to both honest and deliberate mistakes (purposeful indication of a less expensive species/variety of fruit/vegetable) can lead to business losses. The likelihood of an incorrect assessment increases when different fresh products are mixed up.

One potential solution to this challenge is the automatic recognition of fruits and vegetables. The notation of recognition (identification, classification) can also be understood in different ways:

as the recognition of a fruit (distinguishing a fruit from another object, e.g., a leaf, a background), recognizing the species of a fruit (e.g., apple from a pear), and recognizing a variety of a given species of fruit (e.g., Golden Delicious apples from Gloster apple). In the case of retail systems, the last two applications have special significance. The concept of fruit classification best reflects the essence of the issue discussed in the article as a way of automatically determining the right species and variety of fruits. Classification of fruits and vegetables is a relatively complex problem owing to the huge number of varieties [2]. Considerable differences in appearance exist within species and varieties, including irregular shapes, colors, and textures. Furthermore, images range widely in lightning conditions, distance, and angle of the camera; all of which result in distorted images. Another problem is the partial or full occlusion of the object. These constraints have led to the lack of multi-class automated fruit and vegetable classification systems [2] in real-life applications.

An examination of the literature suggests that the effectiveness of fruit and vegetable classification using various machine learning methods [2] (support vector machine, k-nearest neighbor, decision trees, neural networks), especially recent advancements in deep learning, is great [3]. However, the construction of online fruit and vegetable classification systems in retail sales is challenged by the required model learning time and promptness in receiving the classification result, as well as the accuracy of the model prediction. In the case of complex, multi-layered models of deep neural networks, the learning and inference time can be significant. Therefore, in the analyzed application, the most preferred models are those that provide a solution relatively quickly and with high classification accuracy. However, even with high classification rates, it is not guaranteed that the tested method will recognize the given fruit (vegetable) objects in the image in all cases. The system is used by a person who may unwittingly place his/her hand or other object in the frame, which in turn may result in erroneous classifications. There are also unplanned situations, such as the accidental mixing of fresh products, fruit placement in unusual packaging, different lighting conditions or spider webs on the lens, etc. Such situations may also cause uncertainty in the model results.

The primary aim of this study is to propose a method for fruit classification, which in addition to the classifier, would inform about the certainty factor of the results obtained. In the case of a low value of the certainty factor, it would inform the user regarding the most suitable species (varieties) of products. Since the biggest challenge is recognizing the variety of a given species of fruit (vegetable), this work focuses on the classification of several varieties of one species of a popular fruit as a case study: the apple.

The remainder of the paper is organized into the following sections. Section 2 provides a review of the fruit and vegetable classification techniques used in retail sales systems. The problem statement and the research methodologies together with the proposed method of fruit classification using the certainty factor are outlined in Section 3. The results of the research and discussion are also reported. Section 4 provides a conclusion for the study.

2. Related Work

The VeggieVision system is one of the first classifiers of fruit and vegetable products [4]. It recognizes the product based on color and texture from color images according to a nearest-neighbor classifier. This system reports the most likely products to the user, one of which has a high probability of being the correct one. The accuracy of the system compared to currently achieved results is not too high: it is over 95%, however for the top four answers.

The authors of [5] also used the nearest-neighbor classifier for fruit classification, but focused on the depth channel of RGBD (Red, Green, Blue, Depth) images. The use of hierarchical multi-feature classification and hybrid features made it possible to obtain better results for system accuracy among species of fruits, as well as their variety.

The authors in [6] presented a vision-based online fruit and vegetable inspection system with detection and weighing measurement. As a preliminary proposal, the authors used an algorithm leveraging data on hue and morphology to identify bananas and apples. A fruit classifier based on

additional extraction of color chromaticity was presented in [7]. A number of other studies [8] indicated that fruit recognition can be also provided by other classification methods such as fuzzy support vector machine, linear regression classifier, twin support vector machine, sparse autoencoder, classification tree, logistic regression, etc.

The proposal to use neural networks as the fruit classifier was presented among others by Zhang et al. [9], who used a feedforward neural network. The authors first removed the image background with the split-and-merge algorithm, then the color, texture, and shape information was extracted to compose feature data. The authors analyzed numerous learning algorithms, and the FSCABC algorithm (Fitness-Scaled Chaotic Artificial Bee Colony algorithm) was reported to have the best classification accuracy (89.1%). Other applications of neural networks for fruit classification can be found in [10,11].

In recent years, a number of articles have shown considerable modeling success with deep learning applications for image recognition. In [12], the authors applied deep learning with the Convolutional Neural Network (CNN) to vegetable object recognition with the results of learning rate being 99.14% and the recognition rate being 97.58%. In [13], the authors evaluated two CNN architectures (Inception and MobileNet) as classifiers of 10 different kinds of fruits or vegetables. They reported that MobileNet propagated images significantly faster with almost the same accuracy (top three accuracy of 97%). However, there were difficulties in predicting clementines and kiwis. This may be due to the choice of the training and testing of a variety of images, which were captured with a video camera attached to the proposed retail market systems and at the same time extracted from ImageNet.

The article [14] presented a comparative study between Bag Of Features (BOF), Conventional Convolutional Neural Network (CNN), and AlexNet for fruit recognition. The results indicated that all three techniques had excellent recognition accuracy, but the CNN technique was the fastest at presenting a recognition prediction. In turn, in the article [15], two deep neural networks were proposed and tested for using simple and more demanding datasets, with very good results for fruit classification accuracy in both bases. Many numerical experiments for training various architectures of CNN to detect fruits were presented in [16]. A 13-layer CNN was proposed for a similar purpose in [17].

The literature review reported above was used to inform the use of computer vision techniques in an automated sales stand or self-checkout. The literature indicates that machine learning methods (especially CNN methods) perform well at classification of fruits and vegetables in the case of pre-prepared datasets. However, pre-trained (tuned) methods are dependent on the data, but the availability of large collections of images of fruits and vegetables is limited [2]. Given the detailed discussion on the use of CNN methods in automated sales stands or self-checkout, the suggestion can be raised that it is necessary to become independent of a single result in order to increase the certainty of the obtained class and achieve a more effective use of computer vision.

For this purpose, we combine several methods: a CNN method for the fruit classification from a whole image, a YOLO (You Only Look Once) V3 method [18] for the fruit detection from a whole image, and then, a CNN method for the fruit classification from images with a single object (apple). This double-track approach to the fruit classification allows determining the Certainty Factor (*CF*) of the results, the use of which is the main novelty of this paper.

The problem of fruit detection is also widely analyzed in the literature, especially during the detection of fruits in orchards [19] and damage detection [20]. The YOLO V3 model [18], the Faster R-CNN model [21], and their modifications are the state-of-the-art fruit detection approaches [19,20]. The use of object detection and recognition techniques for multi-class fruit classification was presented in [22]. This approach is also effective, but does not calculate an objective certainty factor for the results, which are independent of one classification method.

3. Application of the Proposed Method to the Fruit Variety Classification

3.1. Problem Statement

The traditional grocery store has been evolving in recent decades to a supermarket and discount store concept, carrying all the goods shoppers often desire. These stores offer a very large number of products, both processed and partially processed, as well as fresh produce such as fruits and vegetables. Fresh product is typically sold per piece and by weight. As discussed earlier, the sale of produce may be burdensome for cashiers, because they must remember (or search for) the identification code of each item. In the case of self-service checkouts, the sale of fruits and vegetables is connected with the identification of the products species and varieties by buyers. Thus, the sales process in current use leads to longer customer service time, often causing errors (payments for bad products) and business losses.

The published literature suggests that machine vision systems and machine learning methods allow for the construction of systems for automatic fruit and vegetable classification. In particular, deep learning methods have high classification accuracy for both training and testing images, mainly in the case of recognizing species of fruits and vegetables. Recognizing varieties of fruits and vegetables is more difficult due to highly similar color, structure, and shapes in the same class. In fact, the image of the identified object may differ from the learned pattern, resulting in classification errors.

The primary problem addressed in this study is the following: Is it possible to build a machine vision system that can quickly classify the variety of fruits and vegetables together with providing the result certainty factor and, in the case of uncertainty, will notify about the set of the most probable classes?

To tackle this question, a double-track method for fruit variety classification is proposed that uses the image classification methods on the example of images with a background, as well as the method of object detection allowing the detection of fruit objects that are also used for classification. Comparison of the classification results of different objects of the same image, using the weights of the results, will allow the calculation of the Certainty Factor (*CF*) regarding the proposed result of the classification.

3.2. Research Method

3.2.1. CNN for Fruit Classification

In the proposed fruit classification method, the inference procedure based on CNN is used several times for classification of one variety of fruit. Therefore, the CNN architecture should be as simple as possible, with the goal of handling the task of classification with the highest possible prediction accuracy. By advancing previous research [1,23], we present a simplified CNN architecture in Table 1. This CNN model has been tested for the variety classification of apples.

Table 1. An architecture of the CNN model for fruit classification.

Layer	Purpose	Filter	No of Filters	Weights	Bias	Activation
1	Image input layer					$150 \times 50 \times 3$
2	Convolution + ReLU	3×3	32	$3 \times 3 \times 3 \times 32$	$1 \times 1 \times 32$	$148 \times 148 \times 32$
3	Max_pooling	2×2				$74 \times 74 \times 32$
4	Convolution + ReLU	3×3	64	$3 \times 3 \times 32 \times 64$	$1 \times 1 \times 64$	$72 \times 72 \times 64$
5	Max_pooling	2×2				$36 \times 36 \times 64$
6	Flatten					$1 \times 1 \times 82{,}944$
7	Drop out					$1 \times 1 \times 82{,}944$
8	Fully connected + ReLU			$64 \times 82{,}944$	64×1	$1 \times 1 \times 64$
9	Fully connected + Softmax			6×64	6×1	$1 \times 1 \times 6$
			Output layer			$1 \times 1 \times 6$

Here, we propose a deep neural network model architecture with 9 layers of neurons. The first layer is an input layer that contains $150 \times 150 \times 3$ neurons (RGB image with $150 \times 150 \times 3$ pixels as a resized image with $320 \times 258 \times 3$ pixels). The next 4 layers constitute two tracks with convolution pooling layers that use receptive field (convolutional kernels) of size 3×3 with no stride and no padding. The layers give 32 and 64 features maps, respectively. The convolution layers use nonlinear ReLU (Rectifier Linear Unit) activation functions as follows [24]:

$$ReLU(x) = \max(0, x). \tag{1}$$

This function reduces (turns into zero) the number of parameters in the network, resulting in faster learning. To reduce dimensionality and simultaneously capture the features contained in the sub-regions binned, the max pooling strategy [25] is used in the 3rd and 5th layers. The convolutional and max-pooling layers extract features from image. Then, in order to classify the fruits, the fully-connected layers are applied to the previous dropout layer. Dropout is applied to each element within the feature maps (with a 50% chance of setting inputs to zero), thus allowing for randomly dropping units (along with their connections) from the neural network during training and helping prevent overfitting by adding noise to its hidden units [26].

The 8th layer provides 64 ReLU fully-connected neurons. The last layer as a final classifier has the 6 Softmax neurons, which correspond to the six varieties of apples.

To train the CNN model (optimize its weights and biases), the Adam (Adaptive moment estimation) algorithm [27] was employed with cross entropy as the loss function. The Adam algorithm is a computationally-efficient extension of the stochastic gradient descent method.

The presented architecture of the CNN model was tested for fruit classification in three different ways. First, the network was trained (and validated) with image data from apple objects (original images). Second, the network was trained (and validated) with training data for a single apple object (called the image with the apple or ROI as the Region Of Interest). Finally, the network was trained (and validated) with the both training data. In all cases, different network weights were obtained for the same CNN model. All trained CNN models were tested using the same testing data.

3.2.2. You Only Look Once for Fruit Detection

The YOLO V3 [18] architecture was used to generate the apple ROIs from the original images of [28]. The YOLO (You Only Look Once) family of models is a series of end-to-end deep learning models designed for fast object detection. Version 3 used in this research has 53 convolutional layers. The main difference from the previous version of this architecture is that it makes detections at three different scales, thus making it suitable for the smaller objects. Object features are extracted from these scales like the feature pyramid network.

In the first step, YOLO divides the input image into an S × S grid where S depends on the scale. For each cell, it predicts only one object using boundary boxes. The network predicts an objectless score for each bounding box using logistic regression. The score parameter was used to filter out weak predictions. The result prediction is a box described by the top left and bottom right corner.

The original dataset consisted of folders for each fruit class, such as apple, banana, etc. Within the current class folder, additional classification was done for specific species. Apples in the images were located on a silver shiny plate that generated many false predictions. We used the weights pre-trained on a COCO dataset, containing 80 classes where one of them was the apple class. The COCO apple class consists of many different apple species; a desirable attribute in this case. We could run the object detection using YOLO and filter out just the apple class. To maximize predictive performance, we set the minimum score parameter to 0.8. The predictions were good, but many apples were not detected. This this reason, we set the minimum score to 0.3, which allowed almost all objects to be included regardless of the species.

Model predictions were saved as separate files named according to the source sample to allow for later verification. Generated predictions could be used for ground truthing during the training process.

3.2.3. Proposed Fruit Classification Method Using the Certainty Factor

This study proposes a fruit classification method for a retail sales system. The method uses machine vision system together with machine learning methods (shown in Figure 1). The first stage of the method involves creating an image with all fruit objects. The image includes one or many fruits (intended for one variety and species) with the background, and it is called original image.
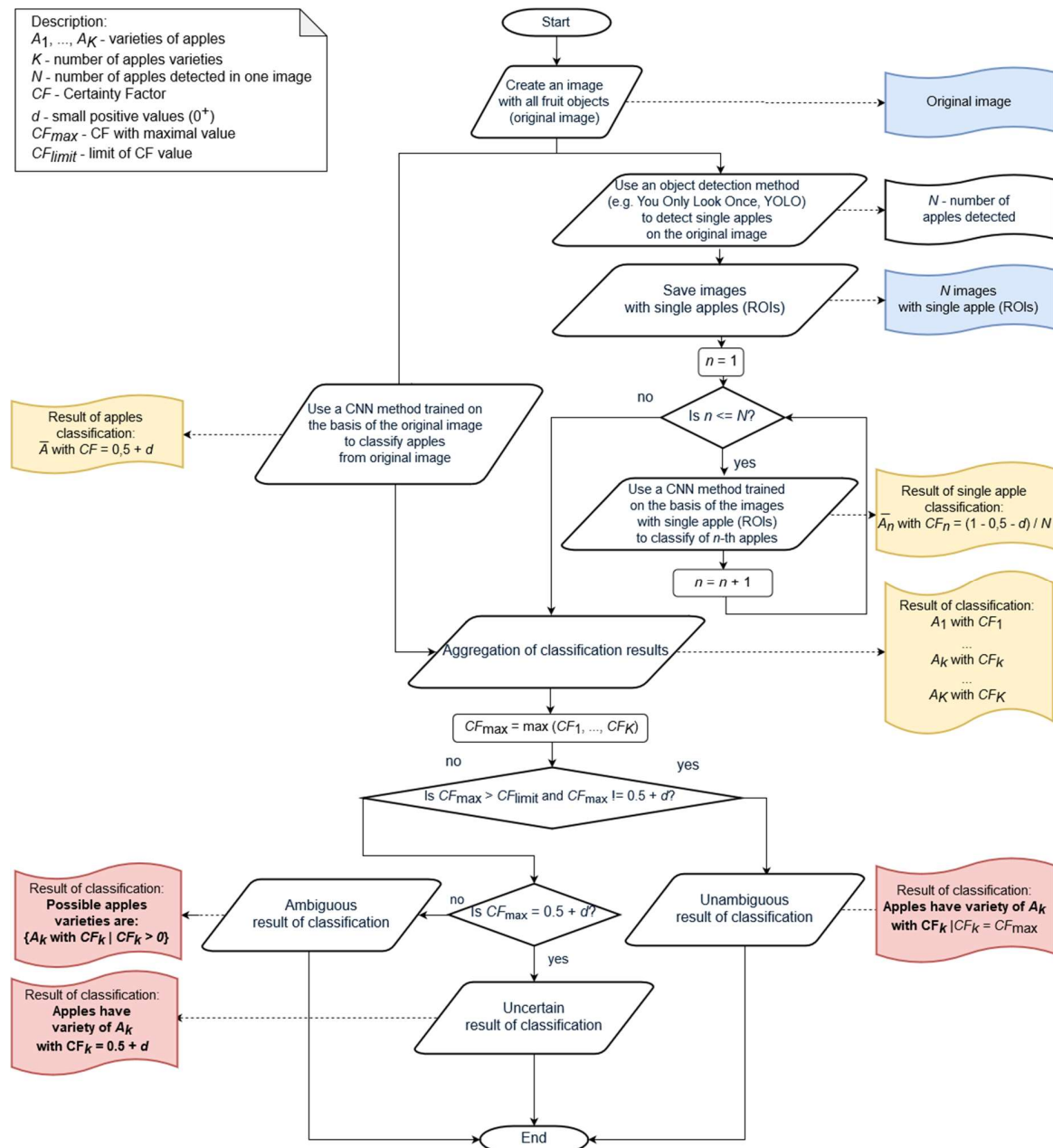

Figure 1. Proposed fruit classification method on an example of apple variety classification.

In order to be sure of the obtained fruit classification result, the proposed method has two separate pathways for fruit classification.

The first pathway was to identify the fruit variety based on the entire original image. For this purpose, the previously described nine-layer CNN was used, which was trained based on the original images. The result of classification $\overline{A}$ was determined with a certainty factor CF with the following value:

$$CF = 0.5 + d, \tag{2}$$

where d is a small positive value (here, $d = 10^{-4}$). In order not to introduce errors in the interpretation of the results, it is recommended that the d value be less than $0.5/(o + 1)$, where o is the maximum number of objects (apples) on the one image in the dataset.

Studies have shown slightly higher accuracy for CNN trained with original images than with ROI objects. Therefore, the small positive value d for the certainty factor gave slightly more importance to the result obtained in the first pathway of the fruit classification algorithm compared to the second pathway described below.

The second pathway of fruit classification consisted of recognizing the fruit variety based on single fruits from the original image. The first step was to use the object detection method to identify single apple objects in the amount of N ($N = 0, 1, 2,...$). The recorded objects of single apples were images with ROIs.

The You Only Look Once (YOLO) method was used to extract images of individual objects from the original images. The fruit variety classification was done based on each nth ROI image ($n = 1,...,N$). For this purpose, the previously described nine-layer CNN was used. It was trained with the ROI images. Each result of the classification $\overline{A}_n$ (each CNN inference) was provided with the appropriate value of certainty factor CF_n:

$$CF_n = \frac{1 - 0.5 - d}{N}, \, n = 1,...,N, \tag{3}$$

where N is the number of objects (apples) detected in the original image and d is a small positive value (in research $d = 10^{-4}$).

The results obtained from both pathways were grouped, and factor CF_k for each kth variety was calculated as follows:

$$CF_k = sum\left(\left\{CF|\overline{A} = A_k\right\}, \left\{CF_1,...CF_n,...,CF_N|\overline{A}_n = A_k\right\}\right), \, k = 1,...K, \tag{4}$$

where CF_k is a certainty factor for the kth variety of fruit A_k (in the researched variety of apples).

The indirect result of the classification can be determined in the form of the following model:

$$\begin{array}{l} A_1 \text{ with } CF_1 \, ... \\ A_k \text{ with } CF_k \, ... \\ A_K \text{ with } CF_K, \end{array} \tag{5}$$

where K is the number of fruit varieties that were detected.

Based on:

$$CF_{max} = \max(CF_1,...,CF_K) \tag{6}$$

the final result of the classification was provided. If the value of CF_{max} was higher than the limit value of certainty factor (CF_{limit}) and is not equal $CF = 0.5 + d$, then an unambiguous classification was obtained:

Fruits (for example, apples) are variety of A_k with CF_k where $CF_k = CF_{max}$.

If the value of CF_{max} did not exceed the limit value of the certainty factor (CF_{limit}) and was not equal to $CF = 0.5 + d$, then an ambiguous classification was obtained. The user can only be informed about the set of possible classification results:

Possible varieties of fruits (apples) are $\{A_k \text{ with } CF_k|CF_k\rangle 0\}$.

If the value of CF_{max} equaled $CF = 0.5 + d$, then uncertain classification was the result:

Fruits (for example, apples) are variety of A_k *with* $CF_k = 0.5 + d$, where d is a small positive value ($d = 10^{-4}$). This result was obtained using only one pathway of classification, which may give rise to uncertainty about the model results.

3.3. Datasets

We employed six different apple varieties (named A–F), and the number of images for each class is provided in Table 2. The images ($320 \times 258 \times 3$ pixels) of apples came from the datasets presented in [28]. The images were obtained using an HD Logitech web camera with five-megapixel snapshots and present objects (different amount of apples) placed in the shop scenery. Various poses and different lighting conditions (i.e., in fluorescent, natural light, with or without sunshine) were preserved.

Table 2. Number of images in the variety class of apples.

Apple Varieties	Number of Images	Example
Apple A	692	
Apple B	740	
Apple C	1002	
Apple D	1033	
Apple E	664	
Apple F	2030	

More information on the analyzed dataset was reported in [1,23]. For simplicity, the images of fruits were taken without being placed in plastic bags.

The data were divided into three sets (training data, validation data, testing data) in the ratio of 70% (4311 original images), 15% (924 original images), and 15% (926 original images), respectively. The recognition algorithm was used to identify a single apple in the original image. Each apple object was saved as a separate image (named as *image with apple*, or *ROI image*). In addition, all apples in each original image were identified and recorded. The detailed structure of the analyzed dataset is presented in Table 3.

Table 3. Dataset structure considered in the research.

Type of Data	Apple Variety						ALL
	A	**B**	**C**	**D**	**E**	**F**	
Training data							
Number of images with apple	1359	1862	1992	2290	1698	3088	12,289
Number of original images	484	518	701	723	464	1421	4311
Total number of images	1843	2380	2693	3013	2162	4509	16,600
Validation data							
Number of images with apple	351	382	411	502	332	609	2587
Number of original images	104	111	150	155	100	304	924
Total number of images	455	493	561	657	432	913	3511
Testing data							
Number of images with apple	307	416	413	494	351	644	2625
Number of original images	104	111	151	155	100	305	926
Total number of images	411	527	564	649	451	949	3551
Total number of images							23,662

All tests and analyzes were carried out using Python programming (v. 3.6.3) with Keras as a high-level neural network API, capable of operating on TensorFlow.

3.4. Results and Discussion

3.4.1. Original contra Region of Interest Images

The CNN model architecture (presented in Section 3.2.1) was tested with various testing and validation images (with only original images, only ROI images, and both). Our goal was to determine the image types necessary to estimate appropriate values of weights in the CNN model to classify the varieties of the fruit correctly. Despite different training and validation files, the model was tested using the same testing dataset. The test results are given in Table 4.

Table 4. Test results of CNNs trained with various training and validation datasets.

	Number of Data							Test Results						
	Training Data		Validation Data		Testing Data			Number of Correct Classifications			Number of Wrong Classifications			
	Original Images	Images with Apple (ROIs)	Original Images	Images with Apple (ROIs)	Original Images	Images with Apple (ROIs)	All Images	Original Images	Images with Apple (ROIs)	All Images	Original Images	Images with Apple (ROIs)	All Images	Accuracy
CNN, original images as training and validation data	4311	0	924	0	926	0	926	924	0	924	2	0	2	99.78%
					0	2625	2625	0	1883	1883	0	742	742	71.73%
					926	2625	3551	924	1883	2807	2	742	744	79.05%
CNN, images with apples as training and validation data	0	12289	0	2587	926	0	926	317	0	317	609	0	609	34.23%
					0	2625	2625	0	2561	2561	0	64	64	97.56%
					926	2625	3551	317	2561	2878	609	64	673	81.05%
CNN, original images and images with apples as training and validation data	4311	12289	924	2587	926	0	926	908	0	908	18	0	18	98.06%
					0	2625	2625	0	2512	2512	0	113	113	95.70%
					926	2625	3551	908	2512	3420	18	113	131	96.31%

According to the accuracy values presented in Table 4, the CNN should be trained with only image types that will be recognized by this network. Training the network with additional images of different scales (many objects or one object) did not improve the accuracy of the classification. The results may also indicate that the size of the ROI in the image is diametrically significant.

It can be also concluded that the best classification possibilities are for the proposed CNN trained and tested with the original images; these models only produced two incorrect classifications (accuracy: 99.78%). However, it should be noted that the number of testing images in this case was much smaller (about 35%) than in the case of testing with ROI images. The misclassifications referring to these cases are described in Table 5. Unfortunately, despite the high values of probabilities that samples belonged to the varieties obtained, the results were incorrect. Thus, the CNN output was not considered a fully reliable classifier, although its performance was close to perfect. As a result, additional methods were explored.

Table 5. Incorrect classification of original images for the CNN trained with original images.

No.	Image	Result Obtained	Correct Result
1		variety of B (the probability of the sample belonging to this variety: 0.8393)	variety of A
2		variety of E (the probability of the sample belonging to this variety: 0.9999)	variety of C

The proposed CNN, which was trained and validated with ROI images, demonstrated slightly less accuracy (97.56%). This accuracy was characteristic for the network, which was trained and validated with the same type of images.

3.4.2. Proposed Fruit Classification Method Using the Certainty Factor

The proposed fruit classification method was tested for the same dataset with six different apple varieties. In this method, one photo classification was associated with classifications of a few images (one original image and ROI images). The amount of testing data according to the number of images for one photo classification is presented in Figure 2. It is evident that the most cases (364 photos) concerned the classification of an apple based on one original image and one ROI image (this was a photo with one apple or a photo on which the YOLO method detected only one apple object). In the dataset, there were photos for which the YOLO method failed to identify any fruit (five photos). The YOLO method detected the most fruits (as many as 11) in four cases.
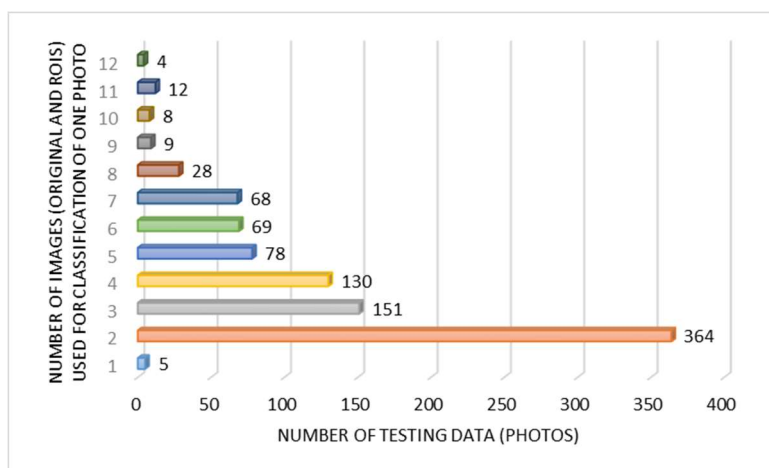
Figure 2. The number of testing data related to the number of classifications for one photo

As a result, the proposed method gave the recognized fruit varieties together with their certainty factors (*CFs*). Because the classification model was performed based on two CNNs with different weights and many different fruit objects, it can be assumed that the approach was relatively objective, and the certainty factors can be reliable factors validating the correctness of the classification result. Consequently, the maximal value of *CF* (CF_{max}) may indicate the result of classification (i.e., a fruit variety with CF_{max} can be the correct class). In a situation where CF_{max} equals one, the given classification result can be treated as certain (it was in 875 cases out of 926 all - 94.49%). In 97.94% of cases, CF_{max} exceeded 0.7501, then the variety of fruit with CF_{max} was the correct variety. Only two misclassifications were detected for varieties with $CF_{max} \in \{0.6700, 0.7501\}$ (Table 6). Therefore, in the large majority of cases (99.78%), the fruit variety with CF_{max} was the correct variety. All results of correct and incorrect classifications together with the values of the maximal value of *CF* are presented in Table 7. The actual classification related to predicted varieties of apples is shown in Table 8.

Table 6. Incorrect classification based on the CF_{max} value in the proposed method.

No.	Image	Result Obtained	Correct Result
1		*variety of A (CF: 0.3333)* *variety of B (CF: 0.6667) -> CF_{max}* variety of C,D,E,F (*CF*: 0)	variety of A
2		variety of A, B (*CF*: 0) *variety of C (CF: 0.2500)* variety of D (*CF*: 0) *variety of E (CF: 0.7500) -> CF_{max}* variety of F (*CF*: 0)	variety of C

Table 7. Results of apple variety classification using the proposed method.

Value of Max Certainty Factor	Number of Correct Classification	Number of Incorrect Classification	Number of Classifications Obtained Based on One Image	Totally Number of Classifications	Number of Unambiguous Classification	Number of Ambiguous Classification	Number of Certain Classification
(1)	(2)	(3)	(4)	(5) = (2) + (3) + (4)	(6)	(7)	(8)
(0; 0.5000>	0	0	0	0	0	0	0
(0.5000; 0.5500>	7	0	5	7	0	2	5
(0.5500; 0.6000>	0	0	0	0	0	0	0
(0.6000; 0.6500>	1	0	0	1	0	1	0
(0.6500; 0.7000>	2	1	0	3	0	3	0
(0.7000; 0.7501>	7	1	0	8	0	8	0
(0.7501; 0.8000>	2	0	0	2	2	0	0
(0.8000; 0.8500>	0	0	0	0	0	0	0
(0.8500; 0.9000>	15	0	0	15	15	0	0
(0.9000; 0.9500>	15	0	0	15	15	0	0
(0.9500; 1>	875	0	0	875	875	0	0
Sum	**924**	**2**	**5**	**926**	**907**	**14**	**5**
				Sum (%)	**97.95%**	**1.51%**	**0.54%**

Table 8. Results of apple variety classification related to actual apple varieties.

		Predicted Apple Variety																	
		A			**B**			**C**			**D**			**E**			**F**		
		$CF_{max}=1$	$CF_{max} \in (1;0.7501)$	$CF_{max} \leq 0.7501$	$CF_{max}=1$	$CF_{max} \in (1;0.7501)$	$CF_{max} \leq 0.7501$	$CF_{max}=1$	$CF_{max} \in (1;0.7501)$	$CF_{max} \leq 0.7501$	$CF_{max}=1$	$CF_{max} \in (1;0.7501)$	$CF_{max} \leq 0.7501$	$CF_{max}=1$	$CF_{max} \in (1;0.7501)$	$CF_{max} \leq 0.7501$	$CF_{max}=1$	$CF_{max} \in (1;0.7501)$	$CF_{max} \leq 0.7501$
Actual Apple Variety	**A**	97	1	5			1												
	B				102	7	2												
	C							144	4	2					1				
	D										148	4	3						
	E													90	8	2			
	F																302	0	3

🟨	correct classification
🟩	incorrect classification

According to the model results, it was possible to identify the limit value of *CF* (CF_{limit}) for which variety of fruit with $CF_{max} \leq CF_{limit}$ can be treated as an ambiguous classification. In the analyzed case, CF_{limit} can be equal to 0.7501. We had 19 original images for which the results of classification had $CF_{max} \leq 0.7501$, including five original images with $CF_{max} = 0.5001$ (ROIs were not detected). Thus, the analyzed classification cases can be divided into three types as follows:

- an unambiguous classification (where $CF_{max} > 0.7501$, 97.95% of cases)
- an ambiguous classification (where $CF_{max} \leq 0.7501$ and classification based on the original image and at least one ROI image, 1.51% of cases)
- an uncertain classification (where $CF_{max} = 0.5001$ and classification based only on the original image, 0.54% of cases).

To illustrate the above situations in detail, Figures 3–5 display examples of apple variety classification using the proposed method.
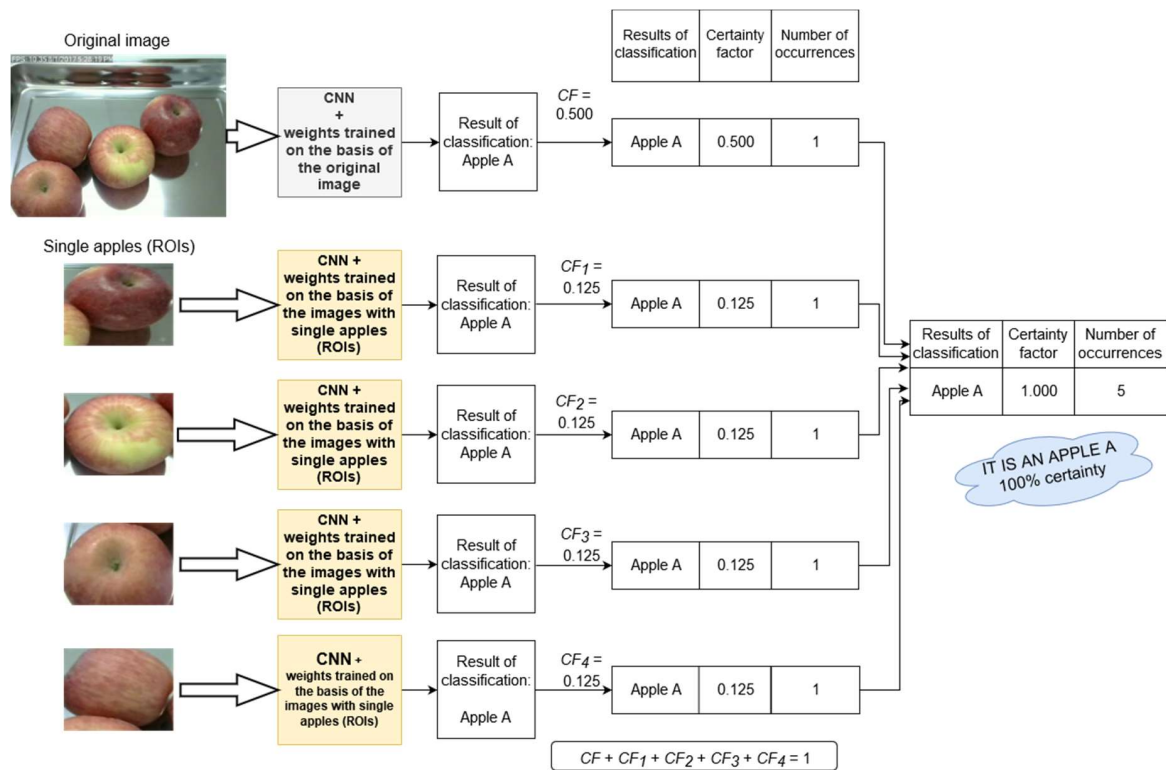
Figure 3. An example of classification using the proposed method: unambiguous classification (based on the original image and ROI images).
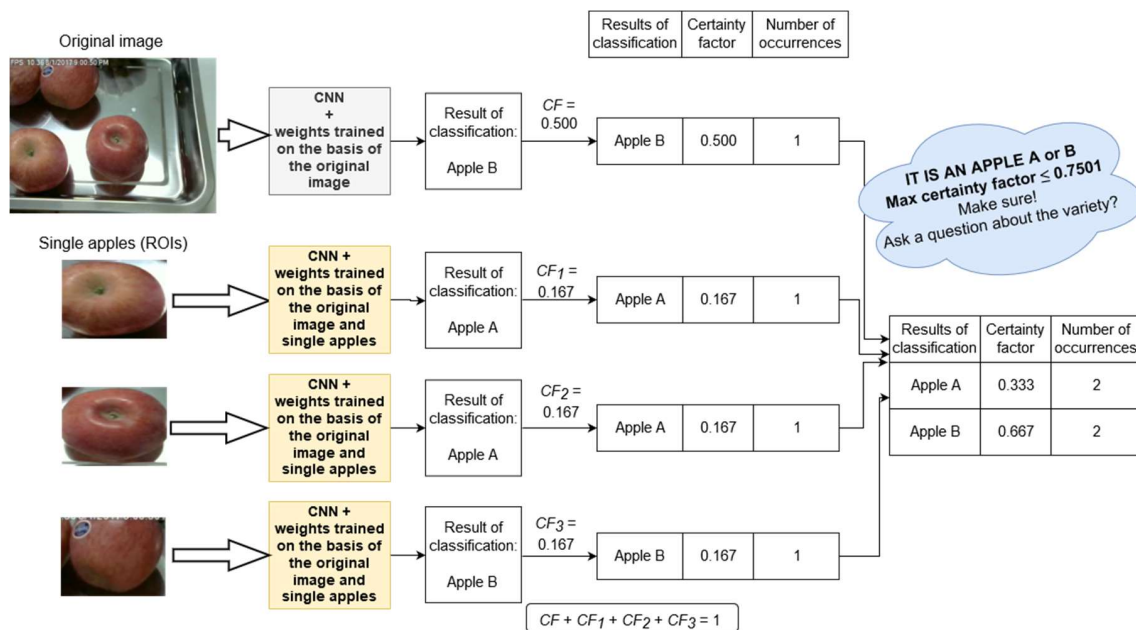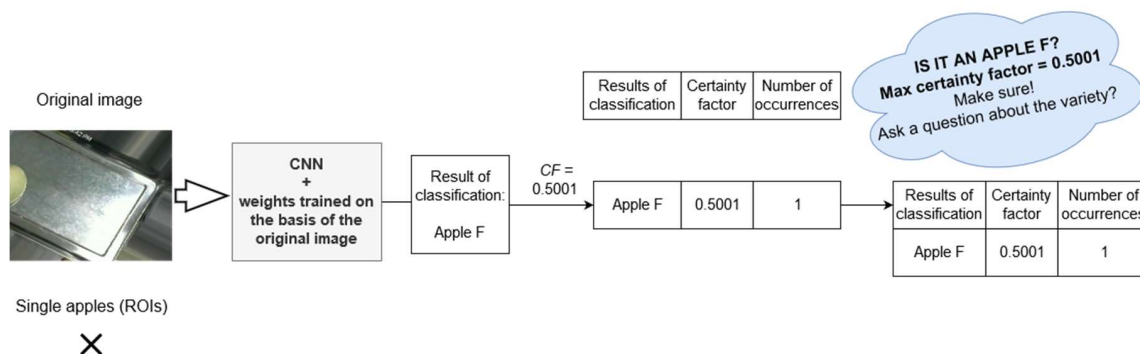

Figure 4. An example of classification using the proposed method: ambiguous classification (based on the original image and ROI images).

Figure 5. An example of classification using the proposed system: uncertain classification (based only on the original image).

To complete the analysis, the execution time (predicting time) of proposed method is presented in Table 9. As can be seen, the execution time depended on the number of objects detected in the original image.

Table 9. Execution time analysis for the proposed method.

Number of Apples on Images	YOLO V3—Average Execution Time (ms)	9-Layer CNN for Original Image—Average Execution Time (ms)	9-Layer CNN for ROIs—Average Execution Time (ms)	Presented Method—Average Execution Time (ms)
(1)	(2)	(3)	(4)	(5) = (2) + (3) + (4)
1	200.97	96.96	92.97	390.90
2	209.90	116.96	146.95	473.82
3	208.96	115.00	159.95	483.91
4	204.96	100.97	167.96	473.89
5	208.92	94.95	182.94	486.81
6	199.91	85.97	188.90	474.79
7	211.90	92.00	242.90	546.80
8	195.96	108.96	281.88	586.81
9	216.93	107.97	302.86	627.76
10	201.91	78.00	329.86	609.77
11	215.96	90.00	329.97	635.92

3.4.3. Comparison of the Results

The research was focused on the synergy of two approaches, the object detection method (in our case, YOLO V3) and the classifier of the full frame and ROIs. Therefore, the comparisons can relate to each method separately or the whole proposed method, which calculated the *CF*s of object classes.

First, a comparison of YOLO V3's performance in relation to other tested methods is presented in Table 10. The YOLO accuracy did not directly influence the result of the system's end inference. The YOLO V3 method affected the relation between the size of the object in the image and the image size itself in the training set and testing set, which in turn affected the accuracy of the fruit identification method using ROIs (in our case, CNN). In addition, YOLO accuracy also affected the number of classified objects (number of ROIs), which in turn affected the accuracy of the certainty factor.

In the research, all the methods were tested with the same training data, which consisted of 926 files with multiple objects. As can be seen, the YOLO V3 generated the highest number of apple class detections, which could be used as the learning ROI images for the classification network. The best average processing times were obtained using the MobileNetV2 + SSDLite and SSD Inception v2 configuration, but the number of detections was much lower comparing to other architectures.

Table 10. Comparison of object detection methods.

Architecture	Number of Apple Class Detections	Average Processing Time for 1 Image (ms)
YOLO V3 [18]	2735	182,17
Fast RCNN Inception v2 [21]	1168	96,15
SSD Inception v2 [29]	1006	46,61
RFCN ResNet101 [30]	2321	190,13
MobileNetV2 + SSDLite [31]	1265	42,53

The second method used in our approach was the nine-layer CNN model. When we compared the proposed CNN with the CNN built based on [23], we had a slightly higher overall accuracy rate (99.78% for the proposed CNN, 99.53% for CNN built on the basis of [23]). In both research works, the same training, validation, and testing data with original images were analyzed. In the learning process of both models, the augmentation methods were used. This improved the ability of models to generalize. In particular, RGB normalization, horizontal flips, image zoom (0.2), and image shear (0.2) were used. In addition, we used a simpler CNN architecture; it contained one less convolution pooling layer; therefore, the average processing time for one image was slightly lower (98.89 ms for the proposed CNN; 125.79 ms for CNN built based on [23]).

The accuracy of the proposed method can be compared to others by treating the final classification of the variety of fruit with the maximum value of CF (CF_{max}). In Table 11, we present the comparison of the proposed method also with the CNN built based on [23]. In both studies, the same training, validation, and testing data were analyzed. We found that the proposed method had a slightly higher accuracy rate.

Table 11. Comparison of the two methods based on CNN.

Apple Variety	Number of Testing Data	Accuracy (%)	
		DCNN Based on [23]	Proposed Method
Apple A	104	98.08	99.04
Apple B	111	99.10	100.00
Apple C	151	100.00	99.34
Apple D	155	100.00	100.00
Apple E	100	100.00	100.00
Apple F	305	100.00	100.00
Average/Total	926	99.53	99.73

We compared the proposed method with the method based only on the nine-layer CNN (Table 1) learned and tested with the original image. In this case, the same incorrect classifications were realized (see Tables 5 and 8), reflected in an accuracy of 99.78% (two wrong classifications for 926). An important distinction is that in the proposed method, the value of CF_{max}, which was not higher than $CF_{limit} = 0.7501$, can indicate an incorrect classification by the proposed vision system. This allows the user to be informed about several possible variety of fruits, from which they can select the correct one. In our study, only 14 cases out of 926 resulted in an ambiguous classification (the question addressed to the user: Which of presented varieties is correct?), and five cases resulted in an uncertain classification (the user should ensure that the proposed classification result is correct). For the testing dataset, the proposed method did not provide incorrect answers and, thus, did not mislead the user. Taken together, the results from this study indicate that the proposed method had a 100% accuracy.

4. Conclusions

This study proposed a double-track method for fruit variety classification in a retail sales system. The method used two nine-layer CNNs with the same architecture and different weight matrices. First, the network classified fruits based on images of fruits with a background and the second one

based on images with the ROI (a single fruit). The results were aggregated with proposed values of neuron weights (importance). Consequently, the method returned predicted class/classes (fruits variety/varieties) together with their Certainty Factor (*CF*). The presented method combined the detection and classification methods and determined the certainty factor associated with the prediction results from original and cropped images ROIs, which was the contribution of this paper. The *CF*s had an advantage in that the correctness of the classification result could be determined, resulting in more reliable predictions compared to the probabilities from the CNNs' outputs. This suggests that the proposed vision-based method can be used in uncertain conditions and unplanned situations as commonly encountered in sales systems (such as the accidental mixture of fresh products, placement of another object in the frame, unusual packaging of fruit, different lighting conditions, etc.). The test using 926 images of six apple varieties indicated that classification accuracy for this method (based on a maximal value of *CF*) was excellent (99.78%). In addition, the method was 100% successful at recognizing unambiguous, ambiguous, and uncertain classifications.

It is important to recognize that the proposed method also had limitations. First, the method performed the classification process several times (for the whole image and detected objects), which could result in a longer time for obtaining the result. However, the uncomplicated structure of CNN and the YOLO V3 method for real-time processing [18] imply that the method can still be used in online sales systems. Second, the use of two different types of training images complicated the learning process of the system. Therefore, the learning process together with determining CF_{limit} values in the proposed method is recommended for further research.

In addition, the future direction is to test the method using a larger dataset containing greater amounts of different fruit and vegetable varieties of different species. It is also preferable to build a fruit and vegetable dataset with more demanding images, which will ultimately be the true test of the system. An interesting research direction will be testing the system with a dataset containing images of fruits and vegetables wrapped in a transparent plastic bag. This situation may cause uncertainty of the obtained result.

Author Contributions: Conceptualization, R.K. and M.P.; formal analysis, R.K.; investigation, R.K. and M.P.; methodology, R.K.; software, R.K. and M.P.; supervision, R.K.; validation, R.K. and M.P.; visualization, R.K.; writing, original draft, R.K. and M.P.; writing, review and editing, R.K. and M.P.

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

References

1. Hussain, I.; He, Q.; Chen, Z. Automatic Fruit Recognition Based on DCNN for Commercial Source Trace System. *Int. J. Comput. Sci. Appl.* **2018**, 8. [CrossRef]
2. Hameed, K.; Chai, D.; Rassau, A. A comprehensive review of fruit and vegetable classification techniques. *Image Vis. Comput.* **2018**, *80*, 24–44. [CrossRef]
3. Srivalli, D.S.; Geetha, A. Fruits, Vegetable and Plants Category Recognition Systems Using Convolutional Neural Networks: A Review. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.* **2019**, *5*, 452–462. [CrossRef]
4. Bolle, R.M.; Connell, J.H.; Haas, N.; Mohan, R.; Taubin, G. Veggievision: A produce recognition system. In Proceedings of the Third IEEE Workshop on Applications of Computer Vision, WACV'96, Sarasota, FL, USA, 2–4 December 1996; pp. 244–251.
5. Rachmawati, E.; Supriana, I.; Khodra, M.L. Toward a new approach in fruit recognition using hybrid RGBD features and fruit hierarchy property. In Proceedings of the 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Yogyakarta, Indonesia, 19–21 September 2017; pp. 1–6.
6. Zhou, H.; Chen, X.; Wang, X.; Wang, L. Design of fruits and vegetables online inspection system based on vision. *J. Phys. Conf. Ser.* **2018**, *1074*, 012160. [CrossRef]

7.   Garcia, F.; Cervantes, J.; Lopez, A.; Alvarado, M. Fruit Classification by Extracting Color Chromaticity, Shape and Texture Features: Towards an Application for Supermarkets. *IEEE Lat. Am. Trans.* **2016**, *14*, 3434–3443. [CrossRef]

8.   Yang, M.-M.; Kichida, R. A study on classification of fruit type and fruit disease, Advances in Engineering Research (AER). In Proceedings of the 13rd Annual International Conference on Electronics, Electrical Engineering and Information Science (EEEIS 2017), Guangzhou, Guangdong, China, 8–10 September 2017; pp. 496–500. [CrossRef]

9.   Zhang, Y.-D.; Wang, S.; Ji, G.; Phillips, P. Fruit classification using computer vision and feedforward neural network. *J. Food Eng.* **2014**, *143*, 167–177. [CrossRef]

10.   Wang, S.; Zhang, Y.; Ji, G.; Yang, J.; Wu, J.; Wei, L. Fruit Classification by Wavelet-Entropy and Feedforward Neural Network Trained by Fitness-Scaled Chaotic ABC and Biogeography-Based Optimization. *Entropy* **2015**, *17*, 5711–5728. [CrossRef]

11.   Zhang, Y.; Phillips, P.; Wang, S.; Ji, G.; Yang, J.; Wu, J. Fruit classification by biogeography-based optimization and feedforward neural network. *Expert Syst.* **2016**, *33*, 239–253. [CrossRef]

12.   Sakai, Y.; Oda, T.; Ikeda, M.; Barolli, L. A Vegetable Category Recognition System Using Deep Neural Network. In Proceedings of the 2016 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), Fukuoka, Japan, 6–8 July 2016; pp. 189–192.

13.   Femling, F.; Olsson, A.; Alonso-Fernandez, F. Fruit and Vegetable Identification Using Machine Learning for Retail Applications. In Proceedings of the 14th International Conference on Signal Image Technology and Internet Based Systems, SITIS 2018, Gran Canaria, Spain, 26–29 November 2018; pp. 9–15.

14.   Hamid, N.N.A.A.; Rabiatul, A.R.; Zaidah, I. Comparing bags of features, conventional convolutional neural network and AlexNet for fruit recognition. *Indones. J. Electr. Eng. Comput. Sci.* **2019**, *14*, 333–339. [CrossRef]

15.   Hossain, M.S.; Al-Hammadi, M.; Muhammad, G. Automatic Fruit Classification Using Deep Learning for Industrial Applications. *IEEE Trans. Ind. Inform.* **2019**, *15*, 1027–1034. [CrossRef]

16.   Muresan, H.; Oltean, M. Fruit recognition from images using deep learning. *Acta Univ. Sappientiae* **2018**, *10*, 26–42. [CrossRef]

17.   Zhang, Y.-D.; Dong, Z.; Chen, X.; Jia, W.; Du, S.; Muhammad, K.; Wang, S.-H. Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation. *Multimed. Tools Appl.* **2017**, *78*, 3613–3632. [CrossRef]

18.   Redmon, J.; Farhadi, A. Yolo V3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.

19.   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]

20.   Tian, Y.; Yang, G.; Wang, Z.; Wang, H.; Li, E.; Liang, Z. Detection of Apple Lesions in Orchards Based on Deep Learning Methods of CycleGAN and YOLOV3-Dense. *J. Sens.* **2019**, *2019*. [CrossRef]

21.   Ren, S.; He, K.; Girshick, R.B.; Sun, J. Faster RCNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2016**, arXiv:1506.01497v3.

22.   Khan, R.; Debnath, R. Multi Class Fruit Classification Using Efficient Object Detection and Recognition Techniques. *Int. J. Image Graph. Signal Process.* **2019**, *8*, 1–18. [CrossRef]

23.   Hussain, I.; Wu, W.L.; Hua, H.Q.; Hussain, N. Intra-Class Recognition of Fruits Using DCNN for Commercial Trace Back-System. In Proceedings of the International Conference on Multimedia Systems and Signal Processing (ICMSSP May 2019), Guangzhou, China, 10–12 May 2019. [CrossRef]

24.   Jarrett, K.; Kavukcuoglu, K.; Ranzato, M.; LeCun, Y. What is the best multi-stage architecture for object recognition? In Proceedings of the International Conference on Computer Vision (ICCV'09), Kyoto, Japan, 29 September–2 October 2009.

25.   Qian, R.; Yue, Y.; Coenen, F.; Zhang, B. Traffic sign recognition with convolutional neural network based on max pooling positions. In Proceedings of the 2th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, 13–15 August 2016; pp. 578–582.

26.   Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

27.   Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.

28.   Hussain, I.; He, Q.; Chen, Z.; Xie, W. Fruit Recognition dataset. *Zenodo* **2018**. [CrossRef]

29. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.

30. Dai, J.; Li, Y.; He, K.; Sun, J. R-FCN: Object detection via region based fully convolutional networks. *arXiv* **2016**, arXiv:1506.01497.

31. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.