

Article

Semantic Segmentation of Packaged and Unpackaged Fresh-Cut Apples Using Deep Learning

Udith Krishnan Vadakkum Vadukkal ^{1,*}, Michela Palumbo ² and Giovanni Attolico ¹

¹ Institute of Intelligent Industrial Systems and Technologies for Advanced Manufacturing, National Research Council, Via Giovanni Amendola 122/DO, 70126 Bari, Italy; giovanni.attolico@stiima.cnr.it

² Institute of Sciences of Food Production, National Research Council of Italy (CNR), c/o CS-DAT, Via Michele Protano, 71121 Foggia, Italy; michela.palumbo@ispa.cnr.it

* Correspondence: udithkrishnan.vadakkumvadukkal@stiima.cnr.it; Tel.: +39-392-174-6501

Abstract: Computer vision systems are often used in industrial quality control to offer fast, objective, non-destructive, and contactless evaluation of fruit. The senescence of fresh-cut apples is strongly related to the browning of the pulp rather than to the properties of the peel. This work addresses the identification and selection of pulp inside images of fresh-cut apples, both packaged and unpackaged; this is a critical step towards a computer vision system that is able to evaluate their quality and internal properties. A DeepLabV3+-based convolutional neural network model (CNN) has been developed for this semantic segmentation task. It has proved to be robust with respect to the similarity of colours between the peel and pulp. Its ability to separate the pulp from the peel and background has been verified on four varieties of apples: Granny Smith (greenish peel), Golden (yellowish peel), Fuji, and Pink Lady (reddish peel). The semantic segmentation achieved an accuracy greater than 99% on all these varieties. The developed approach was able to isolate regions significantly affected by the browning process on both packaged and unpackaged pieces: on these areas, the colour analysis will be studied to evaluate internal quality and senescence of packaged and unpackaged products.

Keywords: fresh-cut apples; quality control; computer vision system; semantic segmentation; deep learning



Citation: Vadakkum, U.K.V.; Palumbo, M.; Attolico, G. Semantic Segmentation of Packaged and Unpackaged Fresh-Cut Apples Using Deep Learning. *Appl. Sci.* **2023**, *13*, 6969. <https://doi.org/10.3390/app13126969>

Academic Editors: Junhui Huang, Qi Xue and Zhao Wang

Received: 5 May 2023

Revised: 5 June 2023

Accepted: 6 June 2023

Published: 9 June 2023



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

1. Introduction

Apples are versatile and widely used fruits. They play an essential role in the daily diet of many people. They are rich in essential nutrients, minerals, fibres, and vitamins that provide a significant contribution to maintaining good health [1]. Additionally, apples provide a boost of energy, help to regulate blood sugar [2], and reduce the risk of certain cancers [3,4]. Quality control is very important to select high quality and non-defective apples and monitor their state in a pervasive way along the whole supply chain.

Traditionally, human experts are responsible for quality control of fruit. They grade apples according to various parameters such as size, shape, colour, defects, etc. Visual and internal quality of apples are related to browning of the pulp, total soluble solids (TSS) [5], acidity (pH) [6], and other physicochemical properties [7]. Estimation of these properties has always been a challenging and time-consuming task for human experts.

The market of fresh-cut apples, ready to be eaten, is increasing [8–11]. The cutting process accelerates the senescence of these fruits which is strongly related to the browning of pulp. A continuous monitoring of this part of apples along the path from harvest to final users, even after they have been packaged, is critical to extend their shelf-life as much as possible.

Computer vision systems (CVS) are replacing human experts in quality control of fruit and in several other applications in the agriculture and food industry. CVSs are being used to recognize and classify different types of fruit based on their size, shape, colour, and texture. CVSs can determine ripeness, firmness, and colour properties. CVS can also

be used for defect detection and grading. The integration of machine learning algorithms into a CVS can improve the accuracy and flexibility of the fruit quality control process and can provide greater, more robust, and consistent performances than human operators. To achieve the best results, vision techniques need to work on the more significant regions of the images avoiding noise sources that can be introduced by less relevant parts of the scene. In quality control of fresh-cut apples, the identification and selection of visible pulp is critical to apply colour analysis to the parts that are more affected by browning induced by senescence. This paper will deal only with the problem of selecting the visible parts of the pulp in the images to restrict further meaningful colour analysis oriented to quality and internal parameters' estimation to the most relevant parts of fruit; colour analysis on pulp regions is out of the scope of this paper.

Image segmentation is an important step in many computer vision applications [12,13]. Robust image segmentation techniques strongly empower the extraction of relevant information from images. Depending on the kind of analysis used to achieve image segmentation, results at different resolutions can be reached up to the pixel-by-pixel segmentation provided by approaches such as colour analysis [14]. Several applications of image segmentation methods exploit colour and texture features to separate fruits from the background [15,16]. To identify the best features for each specific kind of images is often a difficult process that involves a cumbersome trial-and-error process. In cases similar to the Golden apples, where colour and texture features are quite similar between the pulp and peel, the results achieved by traditional image processing and computer vision techniques have been unsatisfactory in terms of completeness and robustness [17].

Payman Moallem et al. proposed a CVS algorithm for the segmentation and grading of Golden Delicious apples [17]. They managed to segment apple samples from the background and were able to detect stem end and calyx regions. They achieved defect segmentation using a multi-layer perceptron (MLP) neural network. Their algorithm achieved an accuracy higher than 94% for calyx detection and 100% accuracy for stem ends outside the apples. However, their accuracy for the stem ends inside the apples was only 81%. They justified this lower accuracy by pointing out the similarity between colour and position of the stem end inside the apple and shadow or defects.

David Ireri et al. came up with a machine vision system based on RGB images for tomato grading [18]. Their system segmented calyx and stalk scars in both defected and healthy tomatoes with an average accuracy of 0.9515 using histogram thresholding based on the mean g-r value of the regions of interest. Since defected calyx and stalk scars appeared to have similar colour intensities at different levels of ambient light, they achieved a lower accuracy on defected tomatoes.

Gabriel A. Leiva-Valenzuela et al. automatically distinguished stem and calyx ends and detected damaged berries using a pattern recognition method [19]. They implemented image segmentation in two steps [19]. In the first one, the original images were cropped to a pre-defined dimension to isolate single berries. The second step involved a threshold-based segmentation of the colour images refined by morphological operations. The resulting binary mask separated the fruit from the background.

Dian Rong et al. developed a sliding window local segmentation algorithm for the detection of surface defects on oranges [20]. This method was able to detect various types of surface defects and to separate them without any additional image lightness correction process. They achieved an accuracy of 93.8% for the classification of stem ends from defective and sound orange peels. They obtained an accuracy of 91.9% in the segmentation of individual defects and a 97% performance rate for defective orange detection.

Sajad Sabzi et al. introduced an automatic non-intrusive computer vision system for the estimation of the pH value of oranges [21]. They performed experiments on three varieties of oranges: Bam, Blood, and Thomson. They used a 2-step segmentation process for the extraction of features to estimate the pH value. A threshold based on the first component of RGB colour space was applied in the first stage to remove the background

before converting the image into a binary representation. A closing filter was used in the second segmentation stage to refine the mask.

S. Poorani et al. proposed a method to identify pomegranate fruits inside colour images taken by digital camera under natural light conditions [22]. The segmentation was based on colour and spatial features using a k-means clustering algorithm. Image retrieval was possible after merging the clustered blocks to a specific number of regions.

Recently, deep learning techniques have proved to be a powerful tool to improve the accuracy of semantic image segmentation. The convolutional neural networks have expressed a significant impact in these tasks [14,23].

R. Marani et al. investigated a deep learning network for the segmentation of grape bunches in colour images [24]. Their comparative study used two segmentation metrics, (1) segmentation accuracy and (2) Intersection over Union (IoU), to evaluate the performance of four different pre-trained network architectures, specifically AlexNet, GoogLeNet, VGG16, and VGG19. They also proposed an optimal threshold selection of bunch probability maps to improve the segmentation of pixels. Their strategy improved the mean segmentation accuracy of the aforementioned four deep neural networks in a range from 2.10% to 8.04%. They reported the VGG19 to achieve the best performance with an accuracy of 80.58% and an IoU of 45.64% on the bunch class. The GoogLeNet produced the worst performance with a mean segmentation accuracy of 74.41% and IoU of 37.13%.

Suchet Bargoti and James P. Underwood introduced an image processing framework for fruit detection and counting using apple orchard image data [25]. General-purpose feature learning algorithms were utilized for image segmentation. The image segmentation was done using multiscale multi-layered perceptron (ms-MLP) and convolutional neural networks (CNN). Watershed segmentation (WS) and circular Hough transform (CHT) algorithms were used for pixel-wise fruit segmentation and to detect and count the individual fruits. CNN provided the best segmentation with a F1-score of 0.791. CNN also achieved best detection with a F1-score of 0.861 using WS detection.

Isaac Perez-Borrero et al. proposed a methodology to perform instance segmentation of strawberries using a fully convolutional neural network [26]. Their methodology produced an improvement in precision and processing time as compared with the previously used Mask R-CNN models and put forward their use in real-time automatic strawberry harvesting systems.

Most CVSs, for fruit grading or defect detection or separation of fruits from background, process images by working on the peel. Looking only at the peel would not provide all the information about physicochemical properties. Colour could be more informative about maturity and internal properties especially if evaluated on internal parts of fruits [7]. The browning of the pulp plays an important role in the evaluation of the internal quality of fruits especially in fresh-cut products, where the cutting handlings cause the rapid browning of the cut surfaces. The intensity of the browning on the pulp in fresh-cut apples is an objective parameter of the shelf-life loss. Higher browning of the pulp indicates a lower visual quality of apple slices (higher shelf-life loss). So, the possibility of identifying the browning only on the pulp of apple slices could be an accurate and consistent way to monitor the shelf-life of this kind of product.

Furthermore, the ripening and evaluation of TSS cannot be estimated accurately by relying only on the peel. However, a high-quality precision can be attained by considering the pulp as well since it provides additional useful hints about internal defects and other physicochemical properties. Therefore, it becomes critical to operate a robust and effective separation of the peel and pulp since the visual similarity between the pulp and peel can make this segmentation task difficult for approaches that rely only on colour.

There is only one research work conducted on the browning of fresh-cut apples that is also by a traditional image analysis. Subhashree et al. performed an image analysis to detect browning in fresh-cut apples using colour and texture features [27]. They developed a CVS for image acquisition of three varieties of apple and measured the colour distance over the time by transforming the colour space from RGB to $L^*a^*b^*$ values. They manually

extracted the parts of interest in each image and rely on the scene to be static while the browning process goes on. A more realistic application context must rely on an automatic selection of the regions of interest in each image. Our approach focuses its attention to the pulp that is significantly affected by the browning process. We expect a colour analysis based only on it to provide more robust and effective results.

Thus, a robust and powerful segmentation technique to separate the peel and pulp of fresh-cut apples is a promising area and very much useful for the continuous monitoring of the browning process of apples from harvest to consumers.

The possibility of applying the same methodology also on packaged products with similar performances is a powerful improvement in CVS applications. In case of packaging products, a high level of quality in terms of appearance, sensorial, and nutritional characteristics is required. Additionally, the increasing consumers' need to assess the real quality level inside the packaging is another important aspect to consider. A relevant challenge to face in ensuring a correct quality assessment through the packaging material by CVS is the separation of opaque and affected regions of bags from the transparent area where the product is visible with acceptable fidelity of visual appearance. This separation step requires robust and powerful segmentation approaches. Few researchers have reported interesting results about the use of CVS through packaging [28,29].

In this work, we aim to segment images of fresh-cut apples by separating the pulp from peel. This step is preliminary to classification and regression tasks that will be based on colour analysis, and they are beyond the scope of this paper. In particular, we will be interested in measuring the browning occurring on the internal part on the fruit (pulp) due to senescence. The experiments of semantic segmentation described in this paper were accomplished by acquiring images of fresh-cut apples belonging to four different varieties: Granny Smith, Golden, Fuji, and Pink Lady. One of them (Granny Smith) has a greenish peel, another has a yellowish peel (Golden), while the other two have a mostly reddish peel. The segmentation must be able to separate pulp regions in all these varieties even through the packaging. On packaged products, it needs to recognize peel and pulp pixels belonging to regions of the images whose colors are meaningfully visible in spite of the photometric deformations introduced by the interaction between the light and the plastic bag. While acceptable results can be obtained by traditional methods on greenish and reddish peels, a yellowish peel strongly degrades the performance of pulp identification. We needed a flexible and powerful technique to isolate the pulp even on these critical images in a robust way. To increase the robustness of the evaluation of apple quality and of internal properties, it is important to minimize errors introduced by pixels that are misclassified as pulp, whose number must be kept as low as possible.

A deep learning CNN model for semantic segmentation of fresh-cut apples has been designed, implemented, and tested in the MATLAB environment. Four classes have been defined: peel, pulp, background, and glittering. The first three are used for both packaged and unpackaged apples while the last one is used only when the apples are observed through the bag. The proposed solution is based on the DeepLabV3+ architecture that has been chosen owing to its accuracy in segmentation problems [30]. The developed semantic segmentation achieved an accuracy greater than 99% on images of both packaged and unpackaged products. The approach has been purposely made conservative to satisfy the requirements of the subsequent analysis. Proper choices have been made while training the network and choosing the evaluation metrics to be coherent with the decision of preferring to lose some pixels of the pulp instead of wrongly including them in color analysis parts of the image that belong to other classes (peel, background, glittering). The approach proved to be useful to isolate the parts that are affected by the browning process and that are better suited to evaluate the internal maturity and the properties of interest.

2. Materials and Methods

2.1. Acquisition of Calibrated Colour Images

To acquire calibrated colour images, it is necessary to evaluate and reduce the colour changes due to environment conditions (lighting, geometry, sensor instability). The setup and the process used for acquisition is described in [29,31–33]. The M9GE 3CCD digital camera (Jai Ltd., Yokohama, Japan), which has a dedicated Charged Coupled Device matrix sensor for each colour channel, was used: its resolution was 1024×768 pixels. The imaged area was about 32×24 cm. The use of three sensors avoids the artefacts introduced by demosaicing methods that are required when recording colour information using a single CCD. The optical axis of the Linos MeVis 12 mm lens system (Linos Photonics Ltd., Edinburgh, UK) was perpendicular to the black background. Two DC power suppliers delivered continuous current to eight halogen lamps, placed along two rows at the two sides of the imaged area and oriented at a 45° angle with respect to the optical axis. The images were saved as uncompressed TIFF to avoid the artifacts introduced by compression algorithms.

2.2. Experimental Samples

The apples used in our experiments belonged to four different varieties: Granny Smith, Golden, Fuji, and Pink Lady. The last two have a mostly reddish skin, while the skin of Golden apples is yellowish and the one of Granny Smith apples is greenish. Therefore, three main peel colours are present in the data set: red, yellow, and green. Each piece of fresh-cut apple can present to the camera the skin, the pulp, or a combination of them. The visual evaluation of digital images has made evident that the separation of skin from pulp could be attempted using the simple colour information only for the reddish and greenish varieties. To achieve a satisfactory separation of these two parts in yellowish apples is challenging, if not impossible, on the basis of the simple colour information: skin and pulp are almost equal visually, especially when the product is fresh.

The data set of RGB images for the experiment contained, for all the considered varieties, three different kinds of arrangement of apple pieces:

1. Randomly disposed unpackaged pieces (each making visible both peel and pulp in variable proportions) (Figure 1).
2. Unpackaged pieces orderly disposed with the pulp oriented upward (making the peel invisible) (Figure 2).
3. Randomly disposed pieces inside a plastic bag (each making visible both skin and pulp in variable proportions) (Figure 3).

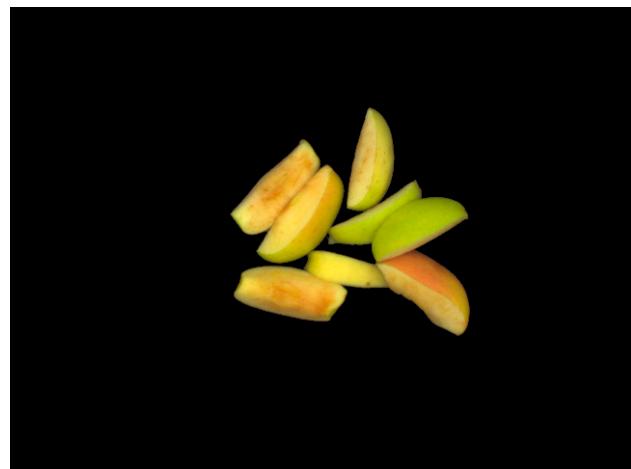


Figure 1. Randomly disposed unpackaged pieces (each making visible both peel and pulp in variable proportions).



Figure 2. Unpackaged pieces orderly disposed with the pulp oriented upward (making visible only the pulp).

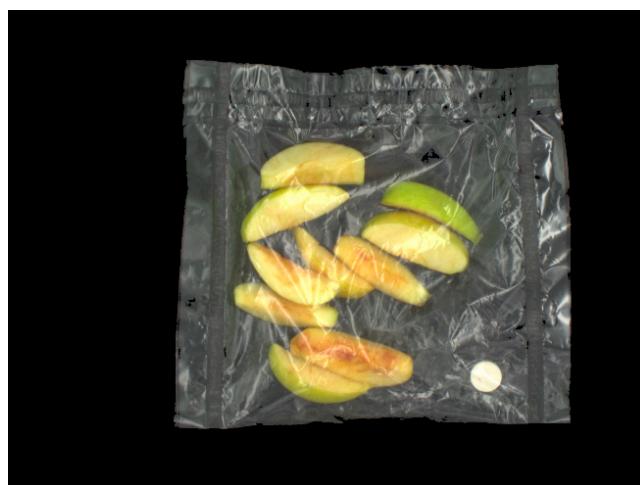


Figure 3. Randomly disposed packaged pieces (each making visible both peel and pulp in variable proportions).

2.3. Network Architecture

The DeepLabV3+ architecture (Figure 4) has been chosen for the semantic segmentation of fresh-cut apples. It can be considered as composed by two main parts: an encoding section (backbone) and a following decoding section. The former part (encoder) uses a convolutional neural network (CNN) to extract a set of features by each image of the data set. The latter part uses a decoding module to achieve segmentation according to the classes of interest. The architecture can flexibly integrate several backbones such as ResNet, Xception, PNASNet, MobileNetv2, etc. In our experiments, the 18-layer-deep Resnet 18 network has been chosen as the backbone: it is pretrained on more than a million images from the ImageNet database [14] and can classify thousands of object categories. Resnet 18 is the simplest version with respect to Resnet versions with 34 layers, 50 layers, 101 layers, and 152 layers.

The input for the segmentation module is a color image composed by the three classical RGB components. This image is fed into a deep-learning neural architecture based on convolutions (what is commonly referred to as a convolutional neural network or CNN). This kind of multi-layered architecture convolve the input image with few layers of kernels that are automatically identified by the learning algorithm during the training phase. A key advantage of CNN is that the kernels set by the learning process (that determine the set of features) are identified without any intervention of designers. At the end, the neural archi-

ture operates as a kind of black box. It is normally difficult to explain, in detail, which kind of measures are made. Nonetheless, the convolutional nature of the computation suggests that the resulting multiresolution features are a combination of colors and of their spatial distribution (commonly referred to as texture in the computer vision community). The relative relevance of color and texture information is not explicitly available.

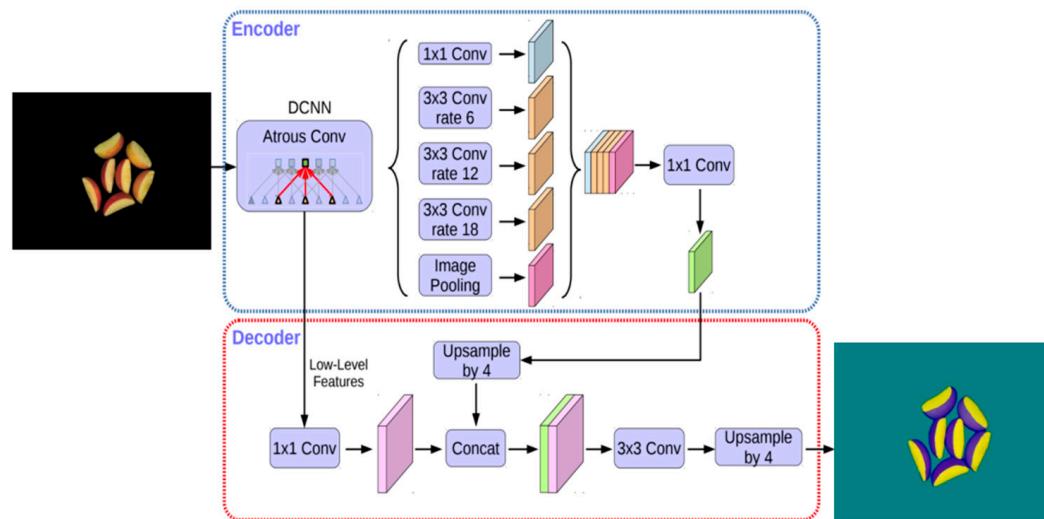


Figure 4. DeepLabV3+ architecture. (Image taken from [34]).

2.4. Images and Data Set Preparation

Each set of samples has 544 RGB images available for training, testing, and validation. Altogether, 47% of images (252 images) from randomly disposed unpackaged cut apples and 48% of images (259 images) from randomly disposed packaged cut apples (including all the three colours) were chosen for training. These 252 images from randomly disposed unpackaged cut apples and 259 images from randomly disposed packaged cut apples were manually labelled pixelwise using the Image labeller App in MATLAB.

Three classes were initially considered in the labelling process of randomly disposed unpackaged apples, namely pulp, peel, and background. Each class was associated to a different label ID, a number that was set in the label image to identify the class of each pixel.

Since glittering regions appear in the images of packaged fresh-cut apples due to the reflection on the plastic bags, one more class was needed to represent these undesired pixels whose colour is not informative about the state of the fruit. Therefore, the labelling process of the images of packaged apples was accomplished using four classes, namely pulp, peel, background, and glitter. For the first three classes, the same ID used for unpackaged products was maintained.

The pixel labels from the Image labeller App were exported to the MATLAB workspace and saved as files of type PNG. An object of the type `groundTruth` was created in the MATLAB environment to contain all these data. An object of type `imageDatastore` was created with the images of all the three sets of samples to efficiently manage these data in the training procedure.

2.5. Data Augmentation

The number and variety of available images were not sufficient to allow a satisfactory training. A data augmentation approach was applied to each image to increase the accuracy of the network through a larger set of samples. The data augmentation scheme was based on the change of brightness of the available labelled images without applying any geometrical transformation (Equation (1)). This scheme allows the same image of labels to be used for all the transformations of photometric values. The following formula was used:

$$V_{out} = V_{in}^{\gamma} \quad (1)$$

where V_{in} is the brightness component of the original input image, γ is the gamma value used to change the brightness, and V_{out} is the transformed image with a new brightness. Gamma values higher than 1 make the image darker while values lower than 1 make the image brighter.

For each image, four transformed versions were generated: three of them by progressively decreasing the brightness and one by increasing the brightness (Figures 5 and 6). The application of a gamma transformation does not affect maximum and minimum values of the photometric range. Data augmentation increased the number of training data of the randomly disposed unpackaged cut apples from the 252 original labelled images to 1008 labelled images and the number of training data of the randomly disposed packaged cut apples from the 259 original labelled images to 1036 labelled images.



Figure 5. Augmented images of randomly disposed unpackaged pieces. The gamma values applied to the image, from left to right, are: (a) 0.9; (b) 1.1; (c) 1.3; (d) 1.5.

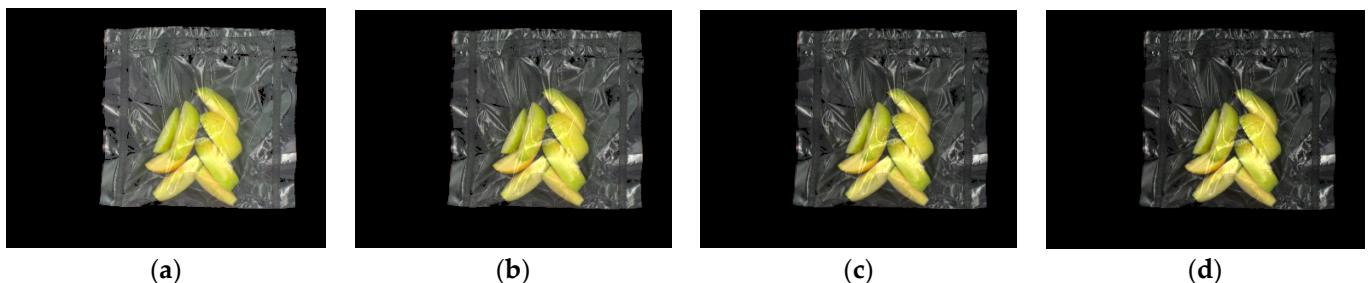


Figure 6. Augmented images of a randomly disposed packaged pieces. The gamma values applied to the image, from left to right, are: (a) 0.9; (b) 1.1; (c) 1.3; (d) 1.5.

2.6. Training of the Network

The DeepLabV3+-based neural network model was created in MATLAB using the ResNet-18 as backbone. The image size was set to $1544 \times 2064 \times 3$.

The same network underwent two separate trainings: one for unpackaged pieces (randomly disposed and orderly disposed with the pulp oriented upward) and another for randomly disposed pieces packaged in the plastic bag. A different number of classes was used in the two cases because packaged apples need a further class for glitters. Each training set was composed by augmented image data, each with the corresponding label data. The two training sets were independent and separated.

The trainings of the networks were done using the following options identified by the Adam optimizer [35]: the maximum number of epochs was 10 and the mini batch size was 8 at each iteration. A piecewise schedule was used for the learning rate. The learning rate drop period was 5 and the drop factor was 0.2. Thus, after every 5 epochs, the learning rate was reduced by a factor of 0.2.

3. Results and Discussion

The test sets for the networks contained all the three different arrangements of the apples (Figures 7–15) and images that were not used during training. The results of the

networks were compared with the manual segmentation done by the operator using the Image Labeller MATLAB tool. The testing data sets were composed of 292 unpackaged images, that is 53% of the total available 544 images, and 285 packaged images, that is 52% from the total 544 images. Global accuracy, mean accuracy, mean IoU, weighted IoU, and mean BF were the metrics used to quantify the performance of the network.



Figure 7. Example of segmentation results for the fresh-cut Pink lady reddish apples.

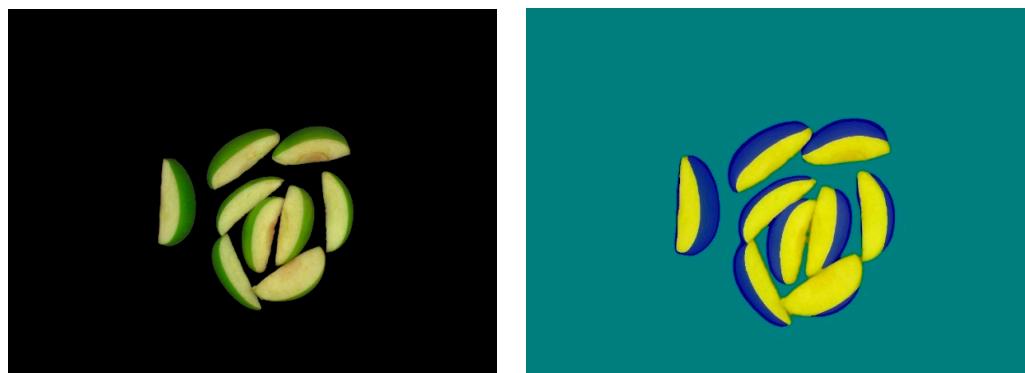


Figure 8. Example of segmentation results for the fresh-cut Granny smith greenish apples.



Figure 9. Example of segmentation results for the fresh-cut Golden yellowish apples.

For each image of the test set, the confusion matrix was evaluated and collected. It provided full information about the correspondence between the results of the network and the ground truth provided by manual visual segmentation. Four different confusion matrices were evaluated from the collection of confusion matrices: the confusion matrix minima (having in each cell, the minimum of all the values of that cell in the whole collection), the confusion matrix maxima (having in each cell, the maximum of all the values of that cell in the whole collection), the confusion matrix mean (having in each cell, the mean of all the values of that cell in the whole collection), and the confusion matrix

standard deviation (having in each cell, the standard deviation of all the values of that cell in the whole collection). These four matrices synthetize the information about the performance of the segmentation over the complete test set.

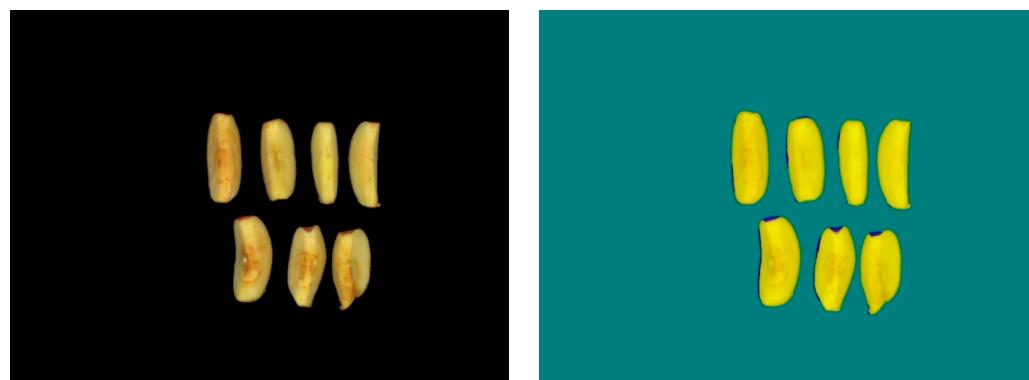


Figure 10. Segmentation results for orderly disposed unpackaged fresh-cut Pink lady reddish apples.

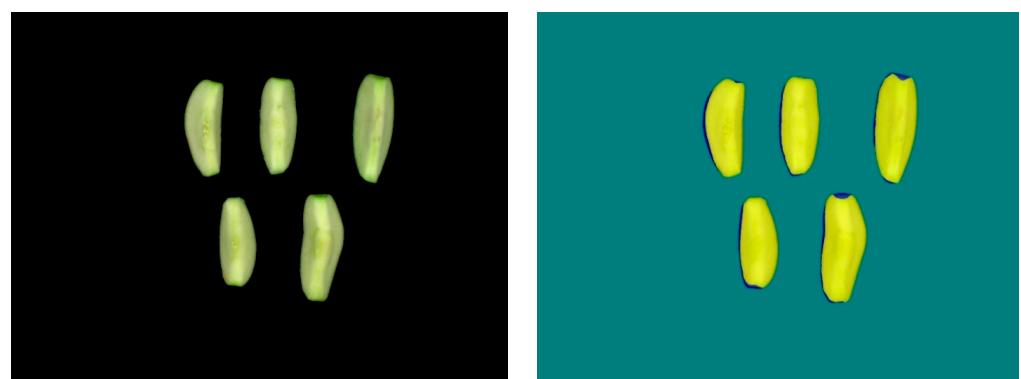


Figure 11. Segmentation results for orderly disposed unpackaged fresh-cut Granny smith greenish apples.



Figure 12. Segmentation results for orderly disposed unpackaged fresh-cut Golden yellowish apples.

The total time taken to complete the training of the network on the PC system (Intel Core i7-1165G7 Processor (2.80 GHz)) was found to be 4976 min (about 3.5 days): it is important to note that this machine is a standard laptop without any special hardware acceleration for this kind of tasks and that the software was written in Matlab without any specific optimization. The same machine required about 4.5 s to segment a single image using the trained network.



Figure 13. Segmentation results for randomly disposed packaged fresh-cut Pink Lady reddish apples. The red label is associated with the class glitter.



Figure 14. Segmentation results for randomly disposed packaged fresh-cut Granny Smith greenish apples. The red label is associated with the class glitter.

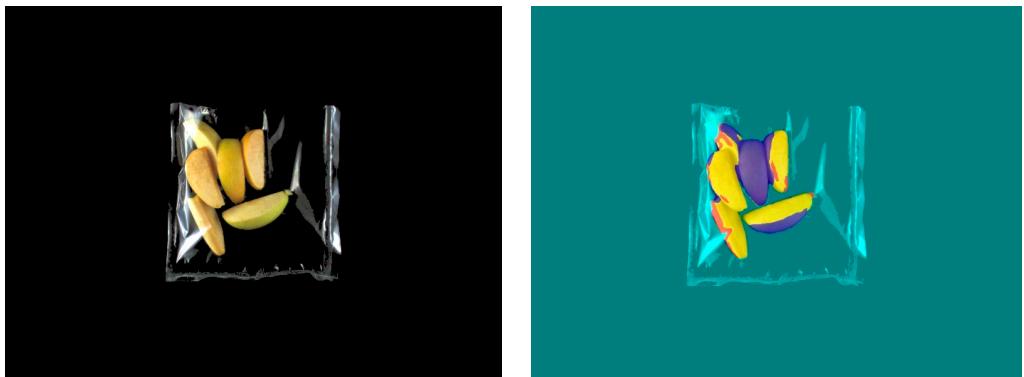


Figure 15. Segmentation results for randomly disposed packaged fresh-cut Golden yellowish apples. The red label is associated with the class glitter.

3.1. Semantic Segmentation Results of the Randomly Disposed Unpackaged Fresh-Cut Apples

To evaluate the trained network, 1008 images were processed (Table 1). The network of the randomly disposed unpackaged apples showed a very good performance of segmentation. The network achieved a mean accuracy, global accuracy, and mean IoU all greater than 99%. Moreover, the network attained a 0.99411 mean BF score.

Table 1. Quantitative evaluation metrics of the network trained for randomly disposed unpackaged fresh-cut apples.

Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score
0.99977	0.9966	0.99394	0.99954	0.99411

Full information about the performance on each of the three classes in the randomly disposed unpackaged fresh-cut apples was measured by the confusion matrix (Table 2) and the class metrics (Table 3).

Table 2. Confusion matrix mean of the randomly disposed unpackaged fresh-cut apples. Each cell represents the mean of the values in the corresponding position over all the samples of the test set.

	Pulp	Peel	Background
Pulp	150,120,894	139,866	0
Peel	437,773	47,031,839	420
Background	586	104,694	2.7574×10^9

Table 3. Class metrics on the randomly disposed unpackaged fresh-cut apples.

	Accuracy	IoU	Mean BF Score
Pulp	0.9991	0.9962	0.9953
Peel	0.9908	0.9857	0.9868
Background	1.0000	1.0000	0.9999

The confusion matrix minima (Table 4) and confusion matrix maxima (Table 5) were evaluated over the complete test set: they provide useful hints about the performance of the method in the best (minima) and the worst (maxima) cases.

Table 4. Confusion matrix minima over the testing dataset.

	Pulp	Peel	Background
Pulp	62,080	0	0
Peel	0	0	0
Background	0	0	2,584,232

Table 5. Confusion matrix maxima over the testing dataset. The first column is especially relevant: it shows the number of pixels misclassified as pulp while belonging to peel or background.

	Pulp	Peel	Background
Pulp	283,229	33,431	131
Peel	3485	155,323	1604
Background	0	184	2,849,891

The confusion matrix maxima (Table 5) shows that a maximum of 33,431 and 131 pulp pixels were incorrectly assigned to the classes peel and background. They will be ignored by the subsequent colour analysis of the computer vision system under development. Moreover, 3485 is the maximum number of peel pixels incorrectly assigned to the class pulp. No background pixels were incorrectly assigned to the pulp class. These two numbers are the most relevant ones; these pixels would bias the colour analysis, bringing into colour analysis contributions from parts of the images that do not exhibit relevant signs of senescence. Nonetheless, the low number of these pixels, with respect to the correctly classified ones, suggest that their statistical contribution to determining the results of classification and regression will be negligible.

Moreover, Tables 4 and 5 show that most of the pixels of each image have been correctly labelled by the network: the number on the diagonal is much larger than that of off-diagonal ones, corresponding to a different type of misclassification.

3.2. Segmentation Results of the Unpackaged Cut Apples orderly Disposed with the Pulp Oriented Upward

Figures 10–12 show visually the semantic segmentation results on orderly disposed apple pieces with pulp oriented upward. Accuracy measures and evaluation matrices were not included in the paper because the presence of peel in these images is too small. We preferred to concentrate the discussion on numerical data associated to unpackaged and packaged randomly disposed pieces that provide a more significant evaluation of the approach.

3.3. Segmentation Results of Randomly Disposed Packaged Fresh-Cut Apples

To evaluate the trained network on packaged apples, 1036 images were processed (Table 6). Furthermore, in this case, the trained network achieved a high performance for segmentation: mean accuracy and global accuracy greater than 99% and mean IoU of more than 98%. The mean BF score of the network was 0.99237.

Table 6. Quantitative evaluation metrics of the randomly disposed packaged fresh-cut apples.

Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score
0.99987	0.99398	0.98991	0.99974	0.99237

The randomly disposed packaged apples have one more class to account for the presence of glitter regions due to the reflection of light on the plastic bag. The accuracy of each of the four classes was calculated using the confusion matrix (Table 7) and the class metrics (Table 8) by comparing the ground truth with the results of the network.

Table 7. Mean confusion matrix of the randomly disposed packaged-cut apples. Each cell represents the mean of the values in the same position over all the samples of the test set.

	Pulp	Peel	Background	Glitter
Pulp	93,356,354	160,988	52	29,482
Peel	83,060	40,050,676	1548	8996
Background	6551	3544	2.8764×10^9	10,841
Glitter	79,186	12,685	1369	4,639,512

Table 8. Class metrics of the randomly disposed packaged fresh-cut apples.

	Accuracy	IoU	Mean BF Score
Pulp	0.9980	0.9962	0.9945
Peel	0.9977	0.9933	0.9928
Background	1.0000	1.0000	0.9999
Glitter	0.9803	0.9702	0.9803

The performance of the network on each class was measured also by the confusion matrix minima (Table 9) and confusion matrix maxima (Table 10).

The confusion matrix maxima table (Table 10) displays the maximum number of correct classifications and errors made by the network. In total, 4661 peel pixels, 33 background pixels, and 1104 glitter pixels were incorrectly classified as pulp. Additionally, in this case, errors are much less than correct classifications: the misclassified pixels are expected to be statistically not relevant for determining the results of the subsequent colour analysis.

Table 9. Confusion matrix minima over the test dataset.

	Pulp	Peel	Background	Glitter
Pulp	27,413	0	0	0
Peel	0	0	0	0
Background	0	0	2,646,664	0
Glitter	0	0	0	0

Table 10. Confusion matrix maxima over the test dataset. The first column is especially relevant because it shows the number of pixels misclassified as pulp while belonging to peel, background, or glitter.

	Pulp	Peel	Background	Glitter
Pulp	160,014	989	440	1664
Peel	4661	108,229	137	742
Background	33	157	2,869,687	76
Glitter	1104	699	2546	22,425

A correct and accurate segmentation step, able to serve as the foreground for only the pulp of apple slices where the browning occurs, allows to monitor the shelf-life loss of the fresh-cut apples along the postharvest storage and could be great help for further analysis. The subsequent steps of features' extraction performed on a well-defined region of the image will be more efficient and consistent to predict internal quality parameters strictly related to the shelf-life (or visual quality) of the product. The principal benefits of an objective, consistent and pervasive food control along the entire supply chain, from the producers to the final consumers, are the standardization of quality levels and the timely detection of senescence enabling waste reduction, sales optimization, and customer satisfaction.

As described earlier, the misclassified pulp pixels are very low in numbers as compared with the correctly labelled pulp pixels. Therefore, the segmentation results on both unpackaged and packaged apples would be significant to assist the quality control assessment on fresh-cut apples in the whole supply chain.

4. Conclusions

Computer vision systems are increasingly being used in industrial quality control to offer fast, objective, non-destructive, and contactless evaluation of fruit. This paper addresses the problem of semantically segmenting images of packaged and unpackaged fresh-cut apples as a preliminary step to develop a computer vision system that will analyze the quality and internal properties of this product. The identification and selection of the pulp is necessary to restrict the analysis to this part as it is more significantly affected by progressive browning caused by senescence. Since apples are widely consumed fruits and an essential component of a healthy diet and are increasingly being offered as fresh-cut pieces in proper packaging, images of packaged and unpackaged products belonging to four different varieties have been acquired and used in the experiments: Granny smith (greenish peel), Golden (yellowish peel), Fuji, and Pink Lady (reddish peel). The visual similarity between the pulp and peel has made unsatisfactory in several cases the application of segmentation approaches relying on traditional colour analysis.

A convolutional neural network model (CNN), based on the DeepLabV3+ architecture, has been designed, implemented, tested, and validated for achieving a robust semantic segmentation in spite of colour similarity between the peel and pulp. The achieved accuracy was greater than 99% on all four considered varieties despite the orientation of the apples, even in the presence of packaging. The testing was done by running the network to segment the test images semantically, which are not used during training. A qualitative visual verification of the performance of trained networks was conducted by overlaying the segmentation results on the colour images. A quantitative evaluation of accuracy was performed by semantically segmenting the entire testing data set (composed of images

manually labelled but not used during training) and by evaluating metrics such as global accuracy, mean accuracy, mean IoU, weighted IoU, and mean BF score. On randomly disposed unpackaged apples, the network achieved mean accuracy, global accuracy, and mean IoU greater than 99%. In the case of randomly disposed packaged apples, the network attained mean accuracy and global accuracy higher than 99%, and mean IoU greater than 98%. Moreover, the confusion matrix was calculated for each image in the testing data set. Four different global confusion matrices (confusion matrix minima, confusion matrix maxima, confusion matrix mean, and confusion matrix standard deviation) were evaluated. They possessed the minimum, maximum, mean, and standard deviation of each position over the complete testing data set.

The analysis of the confusion matrices, especially of the confusion matrix maxima that catch the performance on the worst cases, was conducted to quantify the type of error that is more relevant for our application. In fact, the number of misclassified pixels erroneously classified as pulp can reduce the performance of the following colour analysis by introducing information which are related to parts of the fruit that are not significant for the tasks at hand. The analysis has shown that the number of pixels misclassified as pulp is small with respect to the correctly classified ones. Therefore, the effect of these errors is expected to be statistically negligible in the colour analysis that extract information about the state of fruits.

Two separate networks, based on the same approach and neural architecture, were trained to address images of packaged and unpackaged pieces, respectively. They achieved comparable results proving that the pulp can be identified even through the plastic bag. Therefore, the computer vision system that will integrate this segmentation step will be applicable in a pervasive way along the whole supply chain from harvest to consumers even after packaging.

As mentioned above in our experiment, we trained, tested, and validated the network considering only four varieties of apples consisting of three different peel colours. Even though the network provided satisfactory results in segmentation of tested samples and the variety of colours make us confident about similar results on other cultivars, more tests are needed on other varieties of apples available in the market and on a mixture of different varieties of apples to assess this segmentation method as a robust and persistent technique. Therefore, subsequent future studies are needed on more varieties of apples in different ways.

The manual labelling used in our experiment is a time-consuming process and thus, a self-supervised learning approach for labelling of the images should be investigated to overcome this drawback. This is a consideration of our future work, in order to make the training phase easier without compromising the accuracy of the network.

Author Contributions: Conceptualization, G.A.; methodology, U.K.V.V.; software, U.K.V.V.; validation, U.K.V.V. and G.A.; investigation, U.K.V.V. and M.P.; data curation, U.K.V.V.; writing—original draft preparation, U.K.V.V.; writing—review and editing, U.K.V.V. and G.A.; supervision, G.A.; funding acquisition, G.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the project Prin 2017 “SUS&LOW—Sustaining low-impact practices in horticulture through non-destructive approach to provide more information on fresh produce history and quality” (grant number: 201785Z5H9) from the Italian Ministry of University and Research.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data can be obtained from the corresponding author upon reasonable request.

Acknowledgments: The authors thank Michele Attolico of STIIMA-CNR and Arturo Argentieri of ISASI-CNR for the technical support to the configuration of the experimental set-up.

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

References

1. Musacchi, S.; Serra, S. Apple Fruit Quality: Overview on Pre-Harvest Factors. *Sci. Hortic.* **2018**, *234*, 409–430. [[CrossRef](#)]
2. Inoue, Y.; Cormanes, L.; Yoshimura, K.; Sano, A.; Hori, Y.; Suzuki, R.; Kanamoto, I. Effect of Apple Consumption on Postprandial Blood Glucose Levels in Normal Glucose Tolerance People versus Those with Impaired Glucose Tolerance. *Foods* **2022**, *11*, 1803. [[CrossRef](#)]
3. Boyer, J.; Liu, R.H. Apple Phytochemicals and Their Health Benefits. *Nutr. J.* **2004**, *3*, 5. [[CrossRef](#)] [[PubMed](#)]
4. Tu, S.H.; Chen, L.C.; Ho, Y.S. An Apple a Day to Prevent Cancer Formation: Reducing Cancer Risk with Flavonoids. *J. Food Drug Anal.* **2017**, *25*, 119–124. [[CrossRef](#)] [[PubMed](#)]
5. Lado, J.; Rodrigo, M.J.; Zacarías, L.; Review, S.P. Maturity Indicators and Citrus Fruit Quality Postharvest Quality of Citrus Fruit: Physiological and Genomic Approaches View Project. CaRed-Redes de Excelencia Financiada Por El Ministerio de Economía y Competitividad (BIO2015-71703-REDT) View Project. *Stewart Postharvest Rev.* **2014**, *10*, 1–6.
6. Basak, J.K.; Madhavi, B.G.K.; Paudel, B.; Kim, N.E.; Kim, H.T. Prediction of Total Soluble Solids and PH of Strawberry Fruits Using RGB, HSV and HSL Colour Spaces and Machine Learning Models. *Foods* **2022**, *11*, 2086. [[CrossRef](#)]
7. Hadimani, L.; Mittal, N. Development of a Computer Vision System to Estimate the Colour Indices of Kinnow Mandarins. *J. Food Sci. Technol.* **2019**, *56*, 2305–2311. [[CrossRef](#)]
8. Li, Z.; Yang, H.; Fang, W.; Huang, X.; Shi, J.; Zou, X. Effects of Variety and Pulsed Electric Field on the Quality of Fresh-Cut Apples. *Agriculture* **2023**, *13*, 929. [[CrossRef](#)]
9. Tappi, S.; Velickova, E.; Mannozzi, C.; Tylewicz, U.; Laghi, L.; Rocculi, P. Multi-Analytical Approach to Study Fresh-Cut Apples Vacuum Impregnated with Different Solutions. *Foods* **2022**, *11*, 488. [[CrossRef](#)]
10. Osuga, R.; Koide, S.; Sakurai, M.; Orikasa, T.; Uemura, M. Quality and Microbial Evaluation of Fresh-Cut Apples during 10 Days of Supercooled Storage. *Food Control* **2021**, *126*, 108014. [[CrossRef](#)]
11. Faller, N.; Venir, E.; Zatelli, D.; Bianchi, F. Spraying Treatment of Fresh-Cut Apples as a Sustainable Alternative to Dipping for Browning Inhibition: A Preliminary Lab-Scale Study. *Laimburg J.* **2022**, *4*, 1–10. [[CrossRef](#)]
12. Digital Image Processing Using Matlab (Gonzalez). Available online: [https://www.cin.ufpe.br/~sbm/DEN/Digital%20Image%20Processing%20Using%20Matlab%20\(Gonzalez\).pdf](https://www.cin.ufpe.br/~sbm/DEN/Digital%20Image%20Processing%20Using%20Matlab%20(Gonzalez).pdf) (accessed on 5 June 2023).
13. Torres-Huitzil, C.; Ciudad Victoria, C.-T.; Aurelio NuñoNu, M.; Victoria, C. Area-Time Efficient Implementation of Local Adaptive Image Thresholding in Reconfigurable Hardware. *ACM SIGARCH Comput. Archit. News* **2014**, *42*, 33–38. [[CrossRef](#)]
14. Devanna, R.P.; Milella, A.; Marani, R.; Garofalo, S.P.; Vivaldi, G.A.; Pascuzzi, S.; Galati, R.; Reina, G. In-Field Automatic Identification of Pomegranates Using a Farmer Robot. *Sensors* **2022**, *22*, 5821. [[CrossRef](#)] [[PubMed](#)]
15. Zhao, J.; Tow, J.; Katupitiya, J. On-Tree Fruit Recognition Using Texture Properties and Color Data. In Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, Edmonton, AB, Canada, 2–6 August 2005; IEEE Computer Society: Washington, DC, USA, 2005; pp. 263–268.
16. Arivazhagan, S.; Shebiah, R.N.; Selva Nidhyanandhan, S.; Ganesan, L. Fruit Recognition Using Color and Texture Features. *J. Emerg. Trends Comput. Inf. Sci.* **2010**, *1*, 90–94.
17. Moallem, P.; Serajoddin, A.; Pourghassem, H. Computer Vision-Based Apple Grading for Golden Delicious Apples Based on Surface Features. *Inf. Process. Agric.* **2017**, *4*, 33–40. [[CrossRef](#)]
18. Ireri, D.; Belal, E.; Okinda, C.; Makange, N.; Ji, C. A Computer Vision System for Defect Discrimination and Grading in Tomatoes Using Machine Learning and Image Processing. *Artif. Intell. Agric.* **2019**, *2*, 28–37. [[CrossRef](#)]
19. Leiva-Valenzuela, G.A.; Aguilera, J.M. Automatic Detection of Orientation and Diseases in Blueberries Using Image Analysis to Improve Their Postharvest Storage Quality. *Food Control* **2013**, *33*, 166–173. [[CrossRef](#)]
20. Rong, D.; Rao, X.; Ying, Y. Computer Vision Detection of Surface Defect on Oranges by Means of a Sliding Comparison Window Local Segmentation Algorithm. *Comput. Electron. Agric.* **2017**, *137*, 59–68. [[CrossRef](#)]
21. Sabzi, S.; Javadikia, H.; Arribas, J.I. A Three-Variety Automatic and Non-Intrusive Computer Vision System for the Estimation of Orange Fruit PH Value. *Measurement* **2020**, *152*, 107298. [[CrossRef](#)]
22. Poorani, S.; Gokila Brindha, P. Automatic detection of pomegranate fruits using k-means clustering. *Int. J. Adv. Res. Sci. Eng.* **2014**, *3*, 198–202.
23. Guo, Y.; Liu, Y.; Georgiou, T.; Lew, M.S. A Review of Semantic Segmentation Using Deep Neural Networks. *Int. J. Multimed. Inf. Retr.* **2018**, *7*, 87–93. [[CrossRef](#)]
24. Marani, R.; Milella, A.; Petitti, A.; Reina, G. Deep Neural Networks for Grape Bunch Segmentation in Natural Images from a Consumer-Grade Camera. *Precis. Agric.* **2021**, *22*, 387–413. [[CrossRef](#)]
25. Bargoti, S.; Underwood, J.P. Image Segmentation for Fruit Detection and Yield Estimation in Apple Orchards. *J. Field Robot.* **2017**, *34*, 1039–1060. [[CrossRef](#)]
26. Perez-Borrero, I.; Marin-Santos, D.; Vasallo-Vazquez, M.J.; Gegundez-Arias, M.E. A New Deep-Learning Strawberry Instance Segmentation Methodology Based on a Fully Convolutional Neural Network. *Neural Comput. Appl. Appl.* **2021**, *33*, 15059–15071. [[CrossRef](#)]
27. Subhashree, S.N.; Sunoj, S.; Xue, J.; Bora, G.C. Quantification of Browning in Apples Using Colour and Textural Features by Image Analysis. *Food Qual. Saf.* **2017**, *1*, 221–226. [[CrossRef](#)]

28. Palumbo, M.; Pace, B.; Cefola, M.; Montesano, F.F.; Colelli, G.; Attolico, G. Non-Destructive and Contactless Estimation of Chlorophyll and Ammonia Contents in Packaged Fresh-Cut Rocket Leaves by a Computer Vision System. *Postharvest Biol. Technol.* **2022**, *189*, 111910. [[CrossRef](#)]
29. Cavallo, D.P.; Cefola, M.; Pace, B.; Logrieco, A.F.; Attolico, G. Non-Destructive Automatic Quality Evaluation of Fresh-Cut Iceberg Lettuce through Packaging Material. *J. Food Eng.* **2018**, *223*, 46–52. [[CrossRef](#)]
30. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image Segmentation Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 3523–3542. [[CrossRef](#)]
31. Pace, B.; Cavallo, D.P.; Cefola, M.; Attolico, G. Automatic Identification of Relevant Colors in Non-Destructive Quality Evaluation of Fresh Salad Vegetables. *Int. J. Food Process. Technol.* **2017**, *4*, 1–5.
32. Pace, B.; Cavallo, D.P.; Cefola, M.; Colella, R.; Attolico, G. Adaptive Self-Configuring Computer Vision System for Quality Evaluation of Fresh-Cut Radicchio. *Innov. Food Sci. Emerg. Technol.* **2015**, *32*, 200–207. [[CrossRef](#)]
33. Cavallo, D.P.; Cefola, M.; Pace, B.; Logrieco, A.F.; Attolico, G. Contactless and Non-Destructive Chlorophyll Content Prediction by Random Forest Regression: A Case Study on Fresh-Cut Rocket Leaves. *Comput. Electron. Agric.* **2017**, *140*, 303–310. [[CrossRef](#)]
34. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
35. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.

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