applied sciences

MDPI

Article

Tomato Maturity Estimation Using Deep Neural Network

Taehyeong Kim [1], Dae-Hyun Lee [2,*], Kyoung-Chul Kim [3], Taeyong Choi [4] and Jun Myoung Yu [5]

[1] Big Data COSS, Seoul National University, Seoul 08826, Republic of Korea
[2] Department of Biosystems Machinery Engineering, Chungnam National University, Daejeon 34134, Republic of Korea
[3] Department of Agricultural Engineering, National Institute of Agricultural Sciences, Jeonju 54875, Republic of Korea
[4] Department of Robotics and Mechatronics, Korea Institute of Machinery & Materials, Daejeon 34103, Republic of Korea
[5] Department of Applied Biology, Chungnam National University, Daejeon 34134, Republic of Korea
* Correspondence: leedh7@cnu.ac.kr; Tel.: +82-42-821-6717

Abstract: In this study, we propose a tomato maturity estimation approach based on a deep neural network. Tomato images were obtained using an RGB camera installed on a monitoring robot and samples were cropped to generate a dataset with which to train the classification model. The classification model is trained using cross-entropy loss and mean–variance loss, which can implicitly provide label distribution knowledge. For continuous maturity estimation in the test stage, the output probability distribution of four maturity classes is calculated as an expected (normalized) value. Our results demonstrate that the F1 score was approximately 0.91 on average, with a range of 0.85–0.97. Furthermore, comparison with the hue value—which is correlated with tomato growth—showed no significant differences between estimated maturity and hue values, except in the pink stage. From the overall results, we found that our approach can not only classify the discrete maturation stages of tomatoes but can also continuously estimate their maturity. Furthermore, it is expected that with higher accuracy data labeling, more precise classification and higher accuracy may be achieved.

Keywords: tomato maturity; convolutional neural networks; deep learning; mean–variance loss; robot harvesting

check for updates

Citation: Kim, T.; Lee, D.-H.; Kim, K.-C.; Choi, T.; Yu, J.M. Tomato Maturity Estimation Using Deep Neural Network. *Appl. Sci.* **2023**, *13*, 412. https://doi.org/10.3390/app13010412

Academic Editors: Jinsong Bao and Shu Taira

Received: 11 November 2022
Revised: 16 December 2022
Accepted: 23 December 2022
Published: 28 December 2022

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The continuous shortages in the agricultural labor force require solutions to ensure stable agricultural production. Robotic farming for the realization of unmanned agriculture is emerging as a potential technological alternative, where unmanned agriculture is a technology-intensive farming method that automatically or autonomously performs various agricultural tasks based on intelligent approaches [1]. This issue has received more attention recently, due to the acceleration of global population growth, with the global population expected to reach 10 billion by 2050. Various robot systems are being developed to automate agricultural operations such as harvesting, monitoring, planting, and so on. In particular, the use of harvesting robots in horticultural facilities has potential for practical application in the near future, as the mechanism associated with automatic fruit harvesting is similar to the general gripping system in industrial robots.

Harvesting robots are developed through the integration of various subsystems such as vision, manipulator, gripper, and mobile systems, where vision is a priority requirement for robotic harvesting, being primarily used to implement object (e.g., fruit) recognition. Recent research has shown the capacity for high-level performance when using advanced data-driven approaches such as deep neural networks (DNNs). A convolutional neural network (CNN) is a representative DNN structure for image-based learning, which can extract object features effectively through data learning without human intervention. CNNs have become widely used with the improvement of computing speed. CNN-based approaches for fruit

detection have been demonstrated for various crops. Tomato is one of the main target crops for fruit detection in robotic harvesting, as it is an economically significant horticultural crop worldwide with steadily increasing production. For fruit detection in robotic harvesting, it is necessary to determine not only the position of the target fruit, but also whether it is ripe. The maturity of a tomato is visually expressed in terms of its fruit size and color as it grows. It is easier to qualitatively determine how mature a fruit is from the color change of its surface, which gradually changes from green to red [2]. For this reason, tomato maturity is commonly classified into 4–6 stages based on red color occupancy; thus, tomato maturity can be estimated or classified based on color features, which various studies have implemented simultaneously with fruit detection or segmentation. Most studies have conducted detection using supervised learning methods, in which the model was trained using a labeled dataset. There are several limitations associated with learning tomato maturity using manually labeled images, due to the inaccuracy caused by ambiguity. In the case of tomato maturity, it is difficult to determine the specific class and maturity level for the classification task, due to the continuous change in the color of a tomato skin. There is a color occupancy-based guide for determining tomato maturity level; however, it is hard to obtain consistent results, as they vary depending on labeling quality [3]. Therefore, rather than being limited to a set class through classification learning, a method that can evaluate maturity in terms of a continuous value is expected to be effective for use in various environments.

The purpose of this study was to estimate tomato maturity continuously, based on the use of a deep neural network. The maturity distribution between samples within a specific class was learned intrinsically using a label distribution learning method and mean–variance loss, which has been demonstrated [4] as being capable of estimating the age of humans from facial images. The CNN structure was used to implement the classification model, and the model was trained using images collected in a greenhouse. Tomato maturity was evaluated using the expected value of the probability distribution from the model output. The novelty of this study is that the proposed approach can be used to estimate the continuous maturity of tomatoes from images, although the model was trained in a supervisory manner using specific maturity level classes. We validated the continuous maturity results by comparing the distribution with color information by class. This can contribute to providing information regarding the precise growth stage, and the method's consistency will be further increased as the dataset is expanded.

2. Literature Review

A convolutional neural network (CNN), which is the representative structure of deep learning, can detect objects effectively via hierarchical feature extraction [5]. CNN-based object detection has shown rapid progress in various fields, with visual recognition being the most spotlighted research area due to its similarity to human visual perception [6]. The architecture was extended to object-level detection [7] and pixel-level segmentation [8], and it was recently made possible to provide real-time level performance (YOLO) [9]. CNNs have been modified continuously to implement intelligent functions beyond object detection, and some researchers have generated a remarkable expansion in various fields, such as image generation [10,11] and context interpretation [12]. CNNs have been applied in the field of agriculture for fruit detection in robotic harvesting. Rong et al. [13] detected tomato fruit along with its peduncle, in order to determine the exact cutting point. They used a YOLO model, which presented a detection accuracy of approximately 93% and a localization performance of 73.1 mAP (mean average precision). Padilha et al. [14] also detected tomato fruit based on ripeness using popular models such as YOLO and SSD. They reported that the YOLO-v4-based detection model had high precision (at 91%). CNN-based fruit detection has also presented performances that are high enough for practical use with sweet pepper [15], apple [16], and lychee [17]. In robotic harvesting, it is required that the detected fruit be harvested by estimating the maturity of the fruit. Zu et al. [18] segmented mature green tomatoes using Mask R-CNN, and the results showed that the F1 score

reached 92%. They reported that their method could be used to detect mature tomatoes robustly under various conditions, such as occlusion by other objects and similarly colored backgrounds, and that it could be used in a real environment. Afonso et al. [19] carried out the detection of ripe tomatoes in images captured from a real greenhouse. Their results show that the ripe tomatoes were detected successfully, even when tested with a challenging experimental setup in which simple inexpensive cameras were used. They stated that their method could be used practically in automatic harvesting. Seo et al. [20] focused, in particular, on the classification of the level of tomato maturity. Color space analysis was performed to determine the various harvest times of mature tomato fruits and showed an accuracy of more than 90% in the classification of six stages of maturity (i.e., green, breaker, turning, pink, light red, and red). The studies mentioned above classified maturity into specific stages; however, they required a method that could evaluate maturity as a continuous value. This problem, referred to as label ambiguity, is similar to that of age estimation from facial images, and distribution learning has been proposed to address it [4]. Mean–variance loss is one of the approaches that allows this problem to be solved by learning the label distribution inner class, with the advantage of easy implementation by adding a mean–variance loss term into the model.

3. Materials and Methods

3.1. Data Collection

The tomato (*Solanum lycopersicum* L.) "Dafnis" variety, was cultivated in a general hydroponic greenhouse in South Korea for this study; Figure 1 shows an interior view of the greenhouse, as well as representative sample images. Tomato fruits had various sizes, with a weight range of 150 g to 250 g, and an almost round shape. Images of tomatoes were captured automatically using a developed monitoring system [20] that can travel remotely along the hot water pipes installed in the greenhouse. The monitoring system can track a straight path following a magnetic line installed between both sides of the pipes and can detect the start and end points of crop beds using proximity sensors. A camera (RealSense D435; Intel, Santa Clara, CA, USA) was installed on the side of the monitoring robot to capture the images of the tomatoes, and the images were saved with 800 × 600 pixel resolution at 30 fps. The environmental conditions were 20.6 °C, 67.7%, and 52.9 w/m^2 for humidity, temperature, and light intensity, respectively.

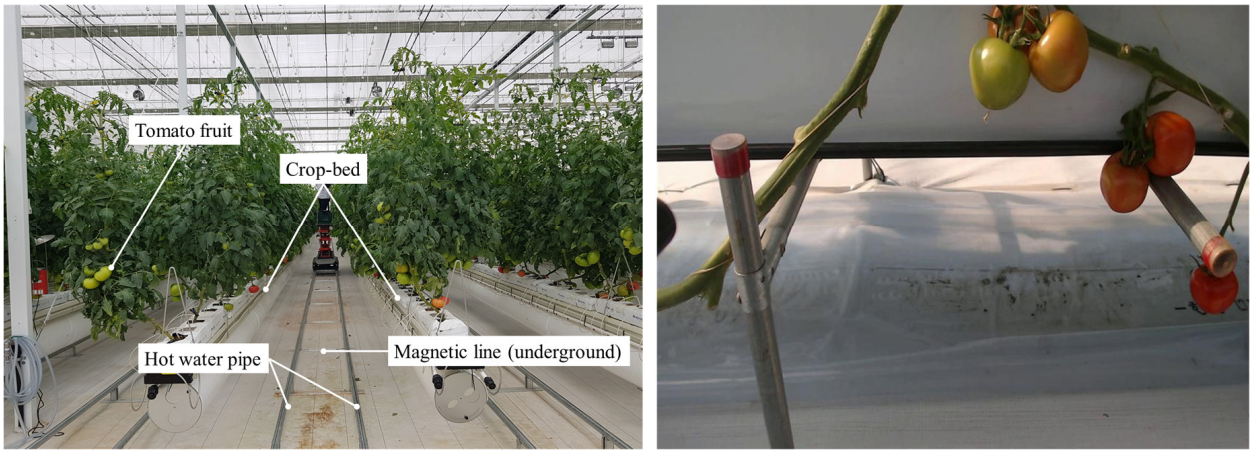


Figure 1. The greenhouse structure (**left**) and a sample captured image (**right**). The greenhouse is of a general hydroponic type, and hot water pipes and a magnetic line were installed between the two crop beds.

Each collected image contains multiple tomatoes at various growth stages, and the tomatoes were annotated for use as samples in training the deep learning model. An annotation tool was developed using Python 3.8 and OpenCV 4.1, which can support the drawing of a polygon shape around each object (tomato). The tomato objects in each image

were extracted as rectangles, that is, the smallest bounding box including the polygon of the tomato object. The extracted samples were used with the background removed to represent the maturity features by deep learning. The overall process is depicted in Figure 2.

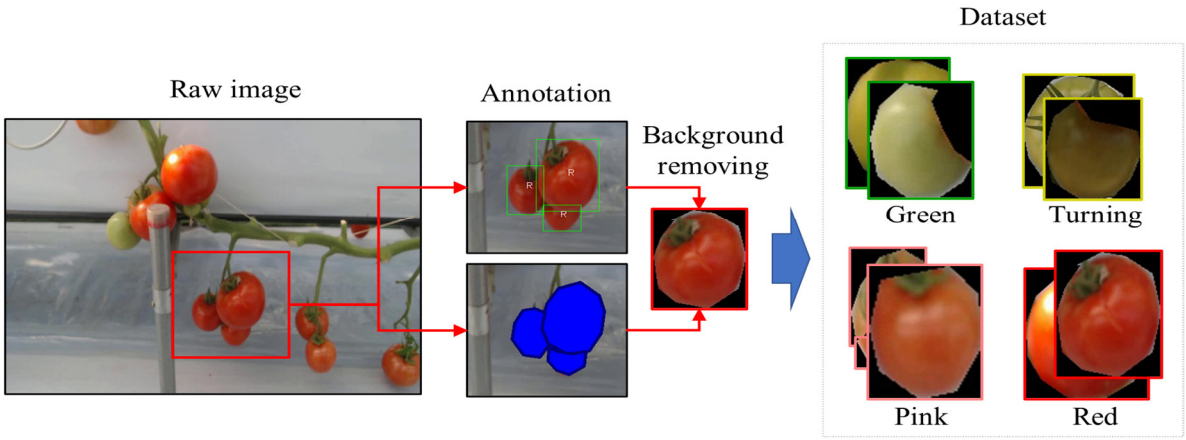


Figure 2. The tomato object annotation and background elimination process.

The background removed samples had different sizes; thus, all of the samples were resized to 128 × 128 pixels (which is the input size of the model) for use as training data, after making each sample square with padding in order to maintain the aspect ratio. In addition, each sample was classified into four maturity stages (green, turning, pink, and red), in order to provide the labels for supervised learning. In general, the maturity stages of tomatoes can be divided into six stages (green, breaker, turning, pink, light red, and red), depending on the ratio of the red region to the entire region [21]; however, it is hard to determine the maturity stage using only the visual information in RGB images examined by humans, as the boundary of the red region is ambiguous due to the continuous color change in the tomato skin. Therefore, the intermediate stages are more difficult to assess. We aimed to evaluate the successive values between maturity stages through classification learning between specific classes; thus, the use of four stages was considered appropriate in this study, as these stages can be easily distinguished by humans. The samples labeled by maturity stage were randomly divided into training and test sets in a 1:1 ratio, and half of the training set was used as a validation set. The numbers of samples were 472, 472, and 944 in the training, validation, and test sets, respectively.

3.2. Deep Neural Network Model

For this study, tomato maturity was estimated using a deep neural network (DNN). The maturity was evaluated continuously based on the implicitly learned feature distribution in each specific class during DNN training for the classification of the four maturity stages. The maturity stages of the tomato can be divided into several classes, and the results depend on the data type and quality, as well as the characteristics of the annotator who conducts the image labeling. For this reason, there is some variety within each class, which is called the label distribution [22]. This challenge is also being actively dealt with for human age estimation from facial images [23], due to the great diversity in faces even at the same age. In this case, the label in the classification task is only a clue, and it is necessary to learn where the sample is positioned in the internal distribution of the specific class. Mean–variance loss is an approach for label distribution learning (LDL), which can be implemented easily by adding the mean and variance differences to the probability between the ground truth and estimated values [4]. In this way, the system can learn a label distribution that has a mean value close to the ground truth label.

Figure 3 shows the architecture of the DNN used to estimate tomato maturity in this study. The model has a convolutional neural network (CNN)-based structure for a simple classification task, consisting of four layers to extract the features. With this structure, it is not difficult to classify the four categories that are likely to depend on color features.

The architecture consists of four convolutional layers, one fully connected layer, and a classifier (softmax), with each conv-net (convolutional layer) consisting of max pooling and rectified linear unit (ReLU) functions. The output of the final convolutional layer includes 256 feature maps, where each feature map has a size of 8 × 8 pixels. The fully connected layer for classification has 256 × 8 × 8 input neurons with four output neurons. The output values are then converted into probabilities with regard to the four maturity stages.

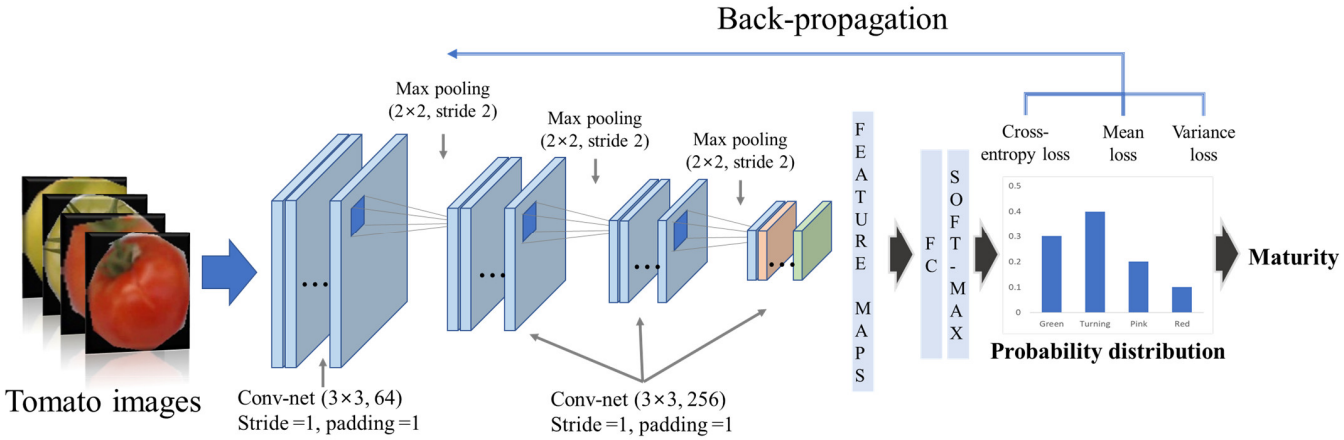


Figure 3. Model structure of DNN for tomato maturity estimation. The model has a shallow 4-CNN architecture with a practical processing speed, and mean and variance losses are added to learn the tomato maturity distribution within a class. The output of softmax is calculated as the argmax value for the training stage and the estimated value of the maturity level for the test stage.

In the training stage, tomato images were used as inputs for the classification model and the features were represented through hierarchical convolutions. The output, expressed as probability distributions between classes (maturity stages), was then compared with one-hot encoded labels. The loss, that is, the numerical value obtained from this comparison, was backpropagated into the DNN in order to update the weights. If tomato maturity is evaluated using only the results of general classification learning, only one of the four stages is selected, which makes it difficult to reflect the various intermediate maturity stages. In order to implicitly learn the distribution within a class during classification, the weights in the DNN were updated considering not only cross-entropy loss (softmax loss) but also mean and variance losses, as shown in Equations (1) and (2), respectively [4]. The mean–variance loss has the advantage of being able to learn the probabilistic distribution of the class and can reflect the inherent ambiguity when the class has an inaccurate label due to the ambiguous selection of classes, which may be the case for tomato maturity.

$$L_m = \frac{1}{2N}\sum_{i=1}^{N}(m_i - y_i)^2 \tag{1}$$

$$L_v = \frac{1}{N}\sum_{i=1}^{N} v_i \tag{2}$$

where L_m is the mean loss, N is the batch size, m_i is the mean of the i sample, y_i is the label of the i sample, L_v is the variance loss, and v_i is the variance of the i sample.

In the test stage, the model weights were fixed, and the final inference of maturity from the tomato image, expressed as the probability distributions for the four maturity stages, was calculated as an estimated value regarding the ground truth. Here, the ground truth probability distribution has a value of 1 for the target class and 0 for the other classes.

Finally, the estimated value was normalized to obtain a value between 0 and 1, as shown in Equation (3):

$$Maturity = \frac{1}{K-1}\sum_{j=0}^{K-1} j \cdot p_j \quad (3)$$

where K is the number of classes, j is the class number (starting from zero), and p_j is the probability of the j class in the softmax output.

3.3. Model Training and Evaluation

The DNN was trained for 300 epochs using the training and validation samples. Input samples were augmented every epoch, in order to minimize overfitting to the training samples, where the augmentation included vertical, oblique, and horizontal flips, as well as stretching. Limited data augmentation related to color features was conducted, because it may have affected the maturity estimation performance. Brightness was only changed for indirectly training the model to have light-invariant performance, and other features (e.g., contrast and saturation) were not augmented. In detail, the V (value) channel was scaled entirely with a random ratio for augmenting the brightness of samples by converting the RGB to HSV. The weights were updated using the Adam optimizer, where the learning rate and weight decay were set at 0.001 and 1×10^{-5}, respectively. The model was trained with a batch size of 64 examples, and the training was terminated before the minimum validation loss, in order to enhance the generalization performance [24]. The model training was implemented in Python 3.7 using PyTorch 1.1, and the CPU and GPU used for image training were an i7-8700 K and NVIDIA Titan-V, respectively. The Titan-V, which is a GPU with 5120 CUDA cores and a 1455 MHz boost clock, was used to implement our approach in an optimal manner.

The classification performance was evaluated through the use of general metrics, including accuracy, precision, recall, and F1 score, as shown in Equations (4)–(7), respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision} \quad (7)$$

where TP (true-positive) denotes the correct classification of a positive label, FP (false-positive) denotes the incorrect classification of a positive label, and FN (false-negative) denotes the incorrect classification of a negative label.

When considering estimated tomato maturity, it is hard to validate the real values as they are difficult to measure, especially from captured scenes. We aimed to obtain continuous maturity in relation to the color distribution of tomato skin, which is difficult for a person to visually determine. Thus, the H (hue) channel of the HSV color model, which has a high correlation with tomato growth [20], was used to obtain a reference distribution of the test images for each class. Estimated tomato maturities were statistically compared with the averaged H values of the input images by class (i.e., maturity stage). The H values were also normalized to obtain a value between 0 and 1. We also tested our method in a Jetson board (Xavier NX; NVIDIA, Santa Clara, CA, USA), in order to evaluate its practical use in real time. The test was conducted using consecutive frames recorded from the greenhouse, and the total recording time was approximately 15 s with 30 fps. Our study aimed to estimate the maturity of detected tomatoes, which in each frame were already annotated with a bounding box. The maturities of the pre-determined tomato locations in each frame were estimated, and the inference time was calculated by

frame. The inference time means the time taken only for maturity evaluation, excluding object detection.

4. Results and Discussion

4.1. Classification Performance

The DNN-based maturity classifier was trained repeatedly, and Figure 4 shows the loss and classification accuracy curves by epoch. The loss curve expresses the total loss, which is the sum of the softmax, mean, and variance losses. Each graph shows the difference between the training and validation sets during repeated learning, and it can be seen that the curve shapes are similar for both the training and validation data. In terms of the total loss, a rapid decrease can be observed up to 10 epochs, and the loss became saturated with a value of approximately 0.05 at 60 epochs. The training was terminated at 100 epochs, as we did not observe a further significant decrease in the training and validation loss. The model weights were finally selected when the validation loss was minimal. The classification accuracies (CAs) increased to greater than 0.95 after 50 epochs in both training and validation.

The losses for the training and validation sets were both saturated, indicating that the model could train the data without overfitting by using a network with appropriately sized parameters.

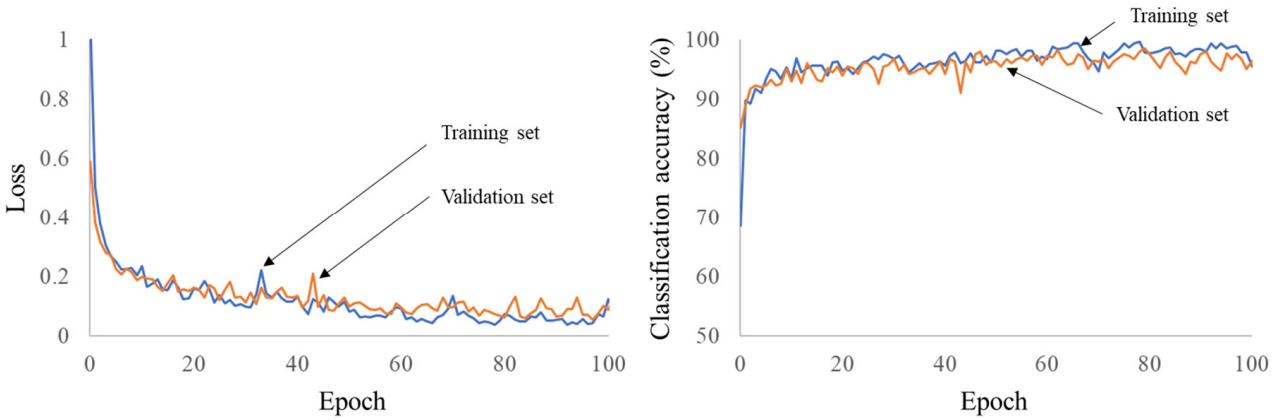


Figure 4. Total losses and classification accuracies for training and validation samples by epoch.

Figure 5 shows the confusion matrix, represented using the classification results of the test set. In each box, two numbers are included: one is the number of corresponding samples, and the other (in parentheses) is the normalized value, which is divided by the total sample number of each class—that is, the sum of all samples of the corresponding row. In the cases of the green and red stages, the classes were composed of images of completely immature or fully mature tomatoes, respectively, and the percentage of correctly classified samples was 97–98% in each class. Meanwhile, the intermediate stages showed lower percentages. Samples in the turning and pink stages have mixed color skin, with colors ranging from green to red, making it difficult to determine the color boundary visually. For this reason, the classification performance was observed to be relatively low (at the level of 77–83%), due to the inaccuracy of the reserved labels [25].

Table 1 provides an analysis of classification performance by class using the test set. The accuracies were observed to be high (greater than 0.95), with an average value of 0.97—a similar level to that obtained in previous research on tomato maturity classification [26]. When considering only the accuracy metric, incorrect predictions can be provided for the minority classes (i.e., those with a smaller sample number), although the model has high accuracy globally; thus, recall, precision, and F1 score were also calculated to evaluate the classification performance. Precision and recall can offer class-wise insight and, in particular, the F1 score, which is the harmonic mean of the precision and recall, can more accurately represent the performance for a data composition with high specific class

(red stage) occupancy, which is imbalanced [27]. Precision was observed in the range of 0.91–0.95, while recall was relatively low (around 0.8) in the intermediate stages. Therefore, the model can estimate class correctly; however, the sensitivity in detecting the target class was low. The F1 score was approximately 0.91 on average, with a range of 0.85–0.97.

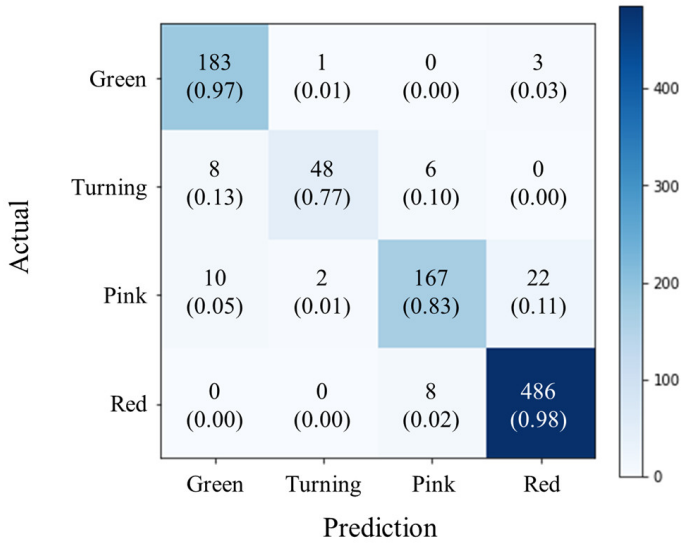


Figure 5. Confusion matrix for tomato maturity classification.

Table 1. Classification performance analysis of tomato maturity estimation model.

Maturity Class	Accuracy	Recall	Precision	F1 Score
Green	0.98	0.98	0.91	0.94
Turning	0.98	0.77	0.94	0.85
Pink	0.95	0.83	0.92	0.87
Red	0.97	0.98	0.95	0.97
Average	0.97	0.89	0.93	0.91

4.2. Maturity Estimation

The data distribution in each class was learned implicitly using the mean–variance loss, and the tomato maturity for the input images was evaluated as an expected value between 0 and 1. Figure 6 shows the classification results visually, by class, as well as the maturity classification and estimation results. The general classification results were also expressed as the generalized maturity between 0 and 1, where the four maturity stages were matched to approximately 0, 0.33, 0.67, and 1.00 for the green, turning, pink, and red stages, respectively. The white circles indicate the correctly classified samples, the orange circles indicate the incorrectly classified samples, and the box with a dashed boundary line indicates the area equally divided into the four areas of the maturity range. The classification results provide the discrete maturity levels, where the samples were only mapped into one of the four classes, although they also have a continuous color range in each class. Furthermore, some of the incorrectly classified samples had maturation statuses close to the boundary between two maturity stages, making it more difficult to treat them as misclassifications. However, the right part of the figure shows the continuous maturity estimation performance in this study, and the distributions of the samples in each class were observed (expressed within the same class or maturity stage) according to the overall color and red occupancy of the tomato surface. In the intermediate maturity, turning, and pink stages, estimated maturities were distributed widely (over 70% over the entire range), thus affecting the increase in false-negative (FN) samples, which was related to the recall performance. The maturity stages at both ends (green and red) showed that the distribution had only a single-sided boundary with other maturity stages; thus, the performance in these stages was relatively high. This distribution in each class and for incorrectly classified

samples could be due not only to the continuous maturity characteristics of tomato growth, but also the mislabeled data caused by various factors such as the ambiguity of the image itself, capture conditions, the annotator's proficiency, and the number of classes used in model training. For this reason, it is difficult to verify the maturity stages presented in these results; thus, the validity of continuous maturity estimation was evaluated using an indirect method, as detailed in the following results.

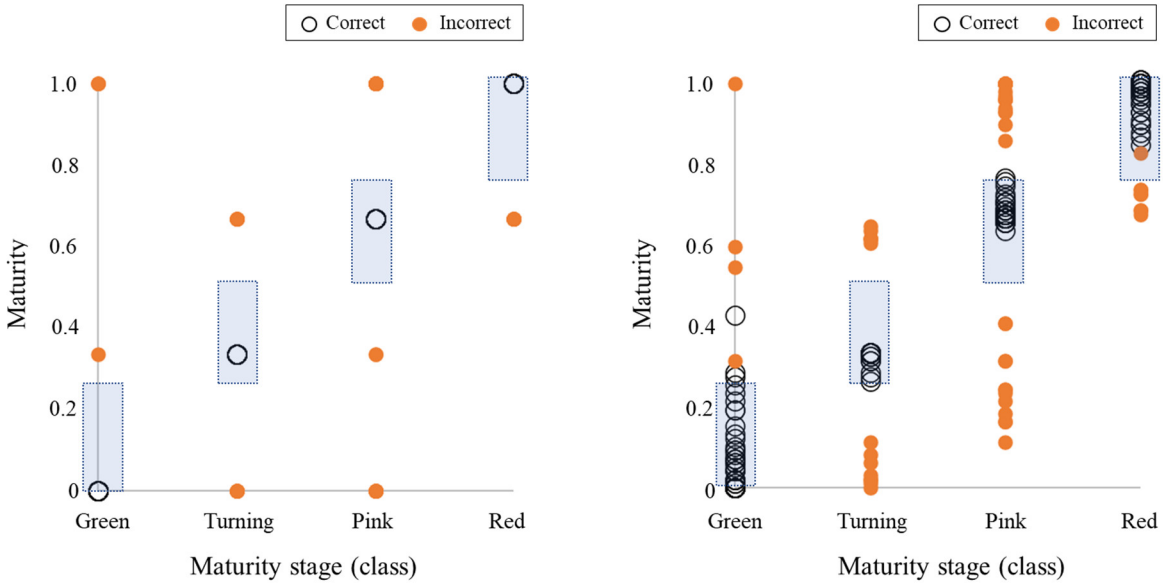


Figure 6. Tomato maturity estimations expressed as classification (**left**) and distribution (**right**).

Figure 7 shows a comparison of the distribution between estimated maturity and hue value for the test samples, where the results are expressed by class. In a previous study, hue value in the HSV color model was shown to have a high linear correlation with the accumulated temperature, with a coefficient of determination (R^2) of 0.96 [20]. The accumulated temperature is the integrated excess of the deficiency in temperature for fixed data, which is usually used in crop growth modeling [28]. It is hard to validate estimated maturity using the actual value, as this is hard to provide due to the ambiguity of the maturity; thus, we compared our results with the hue value of relevant input samples, referring to the above studies. In particular, the hue value was averaged over the tomato area in the images, and the value was normalized to the range of 0–1.

Each graph consists of a probability distribution, expressed as both a histogram and a Gaussian distribution. The green and red bars indicate the relative frequencies for hue value and estimated maturity, respectively, while the brown bars indicate the intersection of the two methods. The Gaussian fitted distributions appeared to be similar between the two methods in the green stage (i.e., immature), when green almost fully occupied the tomato. For the turning stage, there was a slight difference compared with that of the green stage, but the maximum probability of maturity was similar (around 0.4). In the case of the pink and red stages, they had more differences in variance than the green and turning stages, although they showed similar mean maturities. In the DNN-based maturity estimation, the pink stage showed a narrower range distribution, compared with that of the hue value, whereas the opposite was observed in the red stage. It seems that the illuminance may affect the feature representation in DNNs, whereas the hue value is only related to color, with saturation (s) and value (v) having been separated. This is one of the reasons why tomato images have color variance, which might have affected the high variance in the maturity of the red stage, which was even higher than that in the pink stage.

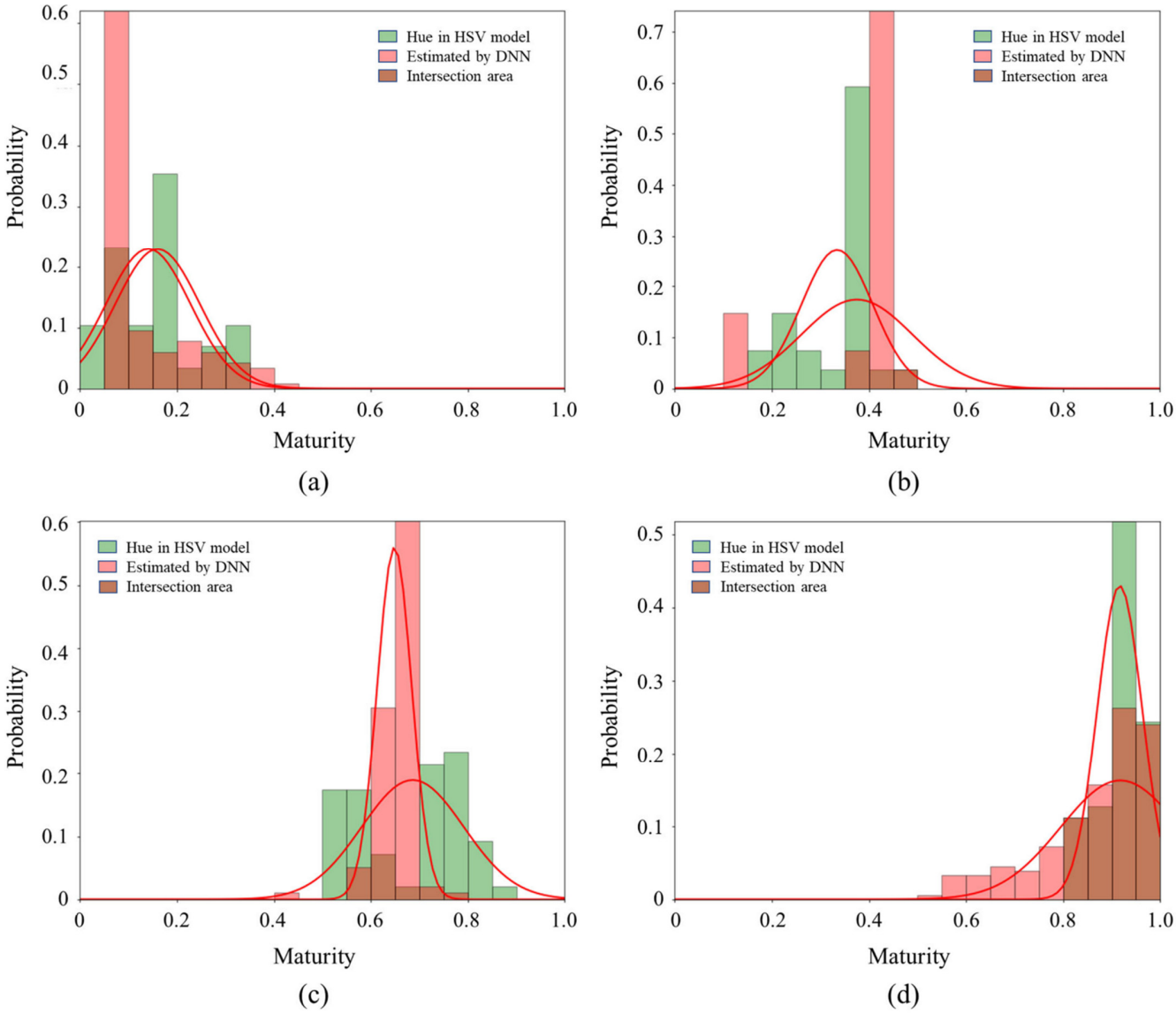


Figure 7. Comparisons of histograms and fitted Gaussian distributions between the estimated maturity and hue value from tomato images: (**a**) green, (**b**) turning, (**c**) pink, and (**d**) red.

The comparison between the DNN-based estimated maturity and hue values was analyzed statistically by maturity class, as detailed in Table 2. Each value is represented as the average and standard deviation of the samples, with respect to each maturity class. The means were similar between the two methods, and their differences were in the range 0.1–0.6, with a relatively similar 10% level of their means. There were no differences between the two methods at the 0.05 significance level (except in the pink stage) and the similarity between the two groups was highest in the red stage.

From this result, it can be concluded that our method can efficiently estimate tomato maturity from images, further enabling assessment in terms of a continuous value, and the performance of our method indicates that the estimated values are significant enough to represent the tomato growth when compared with the hue value, which is correlated with tomato growth.

Table 2. Analysis of the estimated tomato maturity, by comparing the hue value with the maturity class.

Maturity Class	Hue Value	Estimated Maturity	*p*-Value
Green	0.16 ± 0.088 [(1)]	0.15 ± 0.087	0.41
Turning	0.33 ± 0.075	0.38 ± 0.117	0.13
Pink	0.69 ± 0.106	0.75 ± 0.036	<0.01
Red	0.92 ± 0.047	0.92 ± 0.123	0.89

[(1)] Average ± standard deviation.

4.3. Evaluation of the Estimation Speed

The DNN model was constructed using shallow CNN layers; as such, this architecture has advantages in terms of real-time processing, allowing for its practical use in the target system. The method was tested with a high-end GPU-based hardware configuration, as well as in a low-cost Jetson board (Xavier NX; NVIDIA, Santa Clara, CA, USA). The test was conducted by inputting the consecutive frames, and multiple tomatoes in each frame were evaluated for their maturity, where the locations of the tomatoes were pre-determined for every frame. Figure 8 depicts the inference time measured by the software timer by frame and Table 3 shows the processing speed comparison by the GPU. For the Volta GPU in the Xavier NX, the number of CUDA cores is more than 10 times smaller than that in the Titan-V; however, the Jetson board showed only an approximately 3–4 times longer processing time. The processing time for the Xavier NX was approximately 0.02 s, which is equivalent to 50 fps, meaning that it can operate well enough to be used in a robotic system without a significant increase in the processing time. This result indicates the practicality of our method, considering a shallow DNN architecture (although pre- or post-processing algorithms were not considered).

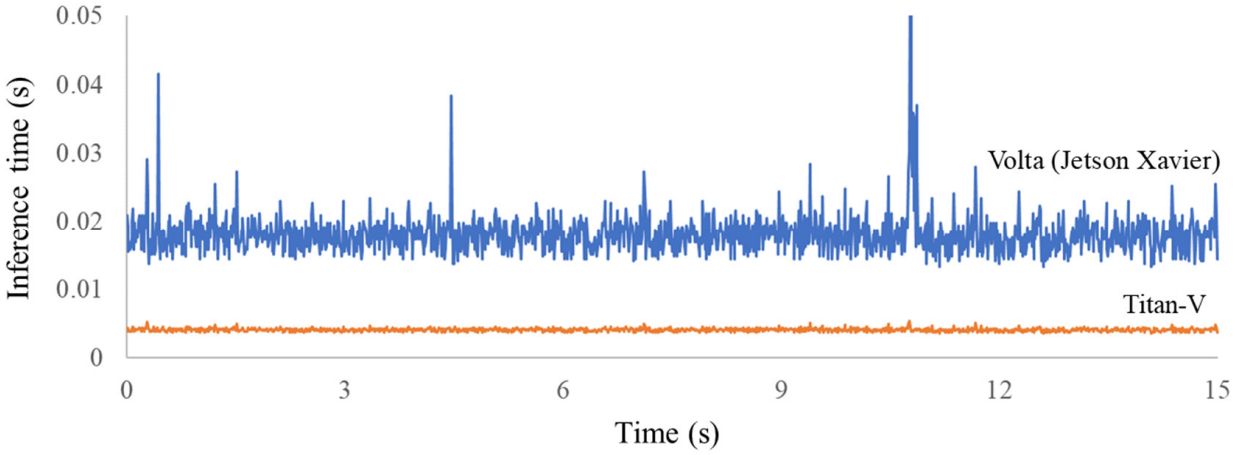


Figure 8. Inference times of maturity estimations for detected tomatoes by frame.

Table 3. Processing speed of tomato maturity estimation with two GPUs.

GPU	Number of CUDA Cores	Clock (MHz)	Processing Speed (s)
Titan-V	5120	1200	0.004 ± 0.0003 [(1)]
Volta	384	854	0.018 ± 0.0028

[(1)] Average ± standard deviation.

4.4. Discussion

These results indicate that our method has comparable performance with other deep learning-based maturity classifiers studied in the field of agriculture, although a four-layer convolutional neural network, which has a shallow-level architecture suitable for practical use, was considered in this study. The results of the maturity classification show that the

performances were 0.97, 0.89, 0.93, and 0.91 for classification accuracy, with a higher range than in the previous study [19,20]. In addition, our model had a practical processing time of less than 0.02 s. Although there are differences in the complexity of the problem and the GPU used in other studies, this finding can still make a significant contribution to practical systems.

In this study, tomato maturity was estimated as a continuous value, and the label distribution was considered to reflect the uncertainty within the maturity stage. The results were compared with the hue value, which is highly correlated with tomato growth [20], showing that not only was a linear relationship observed, but there was also no difference in distribution for each maturity stage between estimated maturity and hue value. It is known that our method shows similar results to previous studies and enables the continuous prediction of maturity values. However, the maturity of a tomato is affected by the distribution of various color features, not just a color channel (e.g., the hue value used in this study). The DNN-based tomato maturity estimation approach proposed in this study has the potential to consider the overall color distribution of the object by conducting hierarchical feature extraction. From the comparison results, the predicted means of the two methods were similar, whereas their distributions differed at each maturity stage. This may indicate that the DNN can consider more features than when simply considering the hue value, although it presented errors with respect to previous research. It is expected that our method can be complemented with accurately labeled tomato images and optimized parameter configuration, thus guiding further training of the model to enhance the accuracy of maturity stage classification. Furthermore, a fully mature tomato can be selected based on the confidence provided in the mean and variance losses. However, our approach only conducted maturity estimations for pre-determined objects, which means that the performance depends on how the target area is determined, including the same tomato object. In addition, the maturity can be overestimated or underestimated if the target is occluded by another object. Securing the accuracy of segmentation in the detection stage is required in order to address this issue, which is outside the scope of this study.

5. Conclusions

We aimed to continuously estimate tomato maturity from tomato-specific images, for which a DNN model with mean–variance losses was used to learn the maturity features and label distributions. The model structure consists of four CNN layers in order to extract the features, and the weights in the model were updated considering three losses: cross-entropy, mean, and variance. For maturity estimation in the test stage, the estimated value based on the output probability distribution of four maturity classes was calculated as a normalized value between 0 and 1. The results indicate that the F1-score was approximately 0.91 on average, with a range of 0.85–0.97, thus providing comparable performance to that reported in relevant research. The estimated maturity was evaluated by comparing the probability distribution with the hue value (which has a high linear correlation with the accumulated temperature index commonly used to model crop growth) in the HSV color model, and the comparison was conducted according to the maturity stage. The comparisons indicated that there were no significant differences between the estimated maturities and hue values, except in the pink stage, and that the similarity between the two groups was highest in the red stage.

Our approach shows that DNN-based distribution learning can be utilized to continuously evaluate tomato maturity and has the advantage of allowing for the evaluation of intermediate classes between specific classes, based on the confidence of classes (i.e., the four maturity stages considered in this study). The results were verified through comparison with a color information index related to tomato growth. It is expected that a higher accuracy of data labeling and more precise classification performance are possible, given that this study was conducted under limited test conditions, such as using sparse maturity stages and uncertain labeled samples determined empirically. Such enhancements may

contribute to optimal use in automatic harvesting, including functions related to optimal harvest time determination or prediction.

Despite these contributions, it was difficult to accurately verify the performance of the proposed model, due to the absence of a measurement value for actual maturity and limitations in its utilization regarding passive maturity feature training. These problems should be resolved by providing accurate matching between tomato images and their actual maturities, which will be conducted as future work.

Author Contributions: Conceptualization, D.-H.L.; methodology, T.K. and D.-H.L.; software, T.K. and D.-H.L.; formal analysis, T.K., J.M.Y. and D.-H.L.; investigation, K.-C.K. and T.C.; writing—original draft preparation, T.K. and D.-H.L.; writing—review and editing, K.-C.K., J.M.Y. and T.C.; project administration, T.C. and D.-H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through the Advanced Agricultural Machinery Industrialization Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (320029-03) and it was also supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through the Advanced Agricultural Machinery Industrialization Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (321059-02).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Kim, W.S.; Lee, D.H.; Kim, Y.J.; Kim, T.; Lee, W.S.; Choi, C.H. Stereo-vision-based crop height estimation for agricultural robots. *Comput. Electron. Agric.* **2021**, *181*, 105937. [CrossRef]
2. Alexander, L.; Grierson, D. Ethylene biosynthesis and action in tomato: A model for climacteric fruit ripening. *J. Exp. Bot.* **2002**, *53*, 2039–2055. [CrossRef]
3. Garcia, M.B.; Ambat, S.; Adao, R.T. Tomayto, tomahto: A machine learning approach for tomato ripening stage identification using pixel-based color image classification. In Proceedings of the 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Laoag, Philippines, 29 November–1 December 2019; pp. 1–6.
4. Pen, H.; Han, H.; Shan, S.; Chen, X. Mean-variance loss for deep age estimation from a face. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; p. 18347679.
5. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2323. [CrossRef]
6. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
7. Girshick, Ross. Fast r-cnn. In Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
8. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer, Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
9. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. *arXiv* **2016**, arXiv:1506.02640.
10. Li, A.; Zhao, L.; Zuo, Z.; Wang, Z.; Chen, H.; Lu, D.; Xing, W. Diversified text-to-image generation via deep mutual information estimation. *Comput. Vis. Image Underst.* **2021**, *211*, 103259. [CrossRef]
11. Dimitri, G.M.; Spasov, S.; Duggento, A.; Passamonti, L.; Lió, P.; Toschi, N. Multimodal and multicontrast image fusion via deep generative models. *Inf. Fusion* **2022**, *88*, 146–160. [CrossRef]
12. Yang, R.; Singh, S.K.; Tavakkoli, M.; Amiri, N.; Yang, Y.; Karami, M.A.; Rai, R. CNN-LSTM deep learning architecture for computer vision-based modal frequency detection. *Mech. Syst. Signal Process.* **2020**, *144*, 106885. [CrossRef]
13. Rong, J.; Dai, G.; Wang, P. A peduncle detection method of tomato for autonomous harvesting. *Complex Intell. Syst.* **2021**, *28*, 2955–2969. [CrossRef]
14. Padilha, T.C.; Moreira, G.; Magalhaes, S.A.; Santos, F.N.; Cunha, M.; Oliveira, M. Tomato detection using deep learning for robotics application. In Proceedings of the 2021 EPIA Conference on Artificial Intelligence, Virtual Event, 3 September 2021; pp. 27–38.

15. Li, H.; Zhu, Q.; Huang, M.; Guo, Y.; Qin, J. Pose estimation of sweet pepper through symmetry axis detection. *Sensors* **2018**, *18*, 3083. [CrossRef]
16. Kang, H.; Zhou, H.; Chen, C. Visual perception and modeling for autonomous apple harvesting. *IEEE Access* **2020**, *8*, 62151–62163. [CrossRef]
17. Li, J.; Tang, Y.; Zou, X.; Lin, G.; Wang, H. Detection of fruit-bearing branches and localization of litchi clusters for vision-based harvesting robots. *IEEE Access* **2020**, *8*, 117746–117758. [CrossRef]
18. Zu, L.; Zhao, Y.; Liu, J.; Su, F.; Zhang, Y.; Liu, P. Detection and segmentation of mature green tomatoes based on mask R-CNN with automatic image acquisition approach. *Sensors* **2021**, *21*, 7824. [CrossRef]
19. Afonso, M.; Fonteijn, H.; Fiorentin, F.S.; Lensink, D.; Mooijm, M.; Faber, N.; Polder, G.; Wegrens, R. Tomato fruit detection and counting in greenhouses using deep learning. *Front. Plant Sci.* **2020**, *11*, 571299. [CrossRef]
20. Seo, D.; Cho, B.H.; Kim, K.C. Development of monitoring robot system for tomato fruits in hydroponic greenhouses. *Agronomy* **2021**, *11*, 2211. [CrossRef]
21. Rupanagudi, S.R.; Ranjani, B.S.; Nagaraj, P.; Bhat, V.G. A cost effective tomato maturity grading system using image processing for farmers. In Proceedings of the 2014 International Conference on Contemporary Computing and Informatics (IC3I), Mysore, India, 27–29 November 2014; pp. 7–12.
22. Jianqiao, L.; Yihan, W.; Yang, O.; Biao, H.; Bailin, L. Neighbor-based label distribution learning to model label ambiguity for aerial scene classification. *Remote Sens.* **2021**, *13*, 755.
23. Geng, X. Label distribution learning. *arXiv* **2014**, arXiv:1408.6027.
24. Kim, W.S.; Lee, D.H.; Kim, T.; Kim, G.; Kim, H.; Sim, T.; Kim, Y.J. One-shot classification-based tilled soil region segmentation for boundary guidance in autonomous tillage. *Comput. Electron. Agric.* **2021**, *189*, 106371. [CrossRef]
25. Kim, W.S.; Lee, D.H.; Kim, Y.J. Machine vision-based automatic disease symptom detection of onion downy mildew. *Comput. Electron. Agric.* **2020**, *168*, 105099. [CrossRef]
26. Mutha, S.A.; Shah, A.M.; Ahmed, M.Z. Maturity detection of tomatoes using deep learning. *SN Comput. Sci.* **2021**, *2*, 441. [CrossRef]
27. Johnson, J.M.; Khoshgoftaar, T.M. Survey on deep learning with class imbalance. *J. Big Data* **2019**, *6*, 27. [CrossRef]
28. Hallett, S.H.; Jones, R.J.A. Compliation of an accumulated temperature databased for use in an environmental information system. *Agric. For. Meteorol.* **1993**, *63*, 21–34. [CrossRef]

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.