Non-Destructive Infield Quality Estimation of Strawberries using Deep Architectures

Cees Jol[1] Junhan Wen[2] Jan van Gemert[1]
[1]Computer Vision Lab, [2]Algorithmics
Delft University of Technology, Delft, The Netherlands
info@ceesjol.nl, {junhan.wen, J.C.vanGemert}@tudelft.nl

Abstract

Strawberries are profitable fruits, yet they have a short shelf life. Therefore, it is crucial to anticipate their quality and harvest them at the best time, which is vital not only for finding the appropriate market but also for minimizing food and economic waste. To this end, non-destructive strawberry quality measurements are useful. Much research is conducted on post-harvest strawberries: the fruits were only analyzed after harvesting and thus, these methods cannot be used to find a good time to harvest. Our research targets pre-harvest analysis for supporting the timing decisions of harvests. As such, we used an infield image dataset that was collected during the cultivation of strawberries. The images are labeled by quality assessments and measurements from post-harvest destructive tests. We evaluated deep learning for quality estimation and trained our algorithms to predict the ripeness, firmness, and sweetness of strawberries. Additionally, we applied depth estimation algorithms and shape inpainting models to estimate the size of strawberries using images. Our results demonstrate the feasibility of infield quality attribute prediction.

1. Introduction

Strawberries are popular economic crops that are cultivated worldwide. Since they are not protected by thick or hard skin, they have a short shelf life and thus have strict requirements for quality at harvest. Attributes such as ripeness, firmness, and sweetness are important references for categorizing and pricing a strawberry fruit. Underripe or over-ripe strawberries are not desirable. Moreover, strawberries placed in the wrong markets could also lead to food loss or financial waste. Thus, for optimal quality and profitability, they need to be harvested at the right time [27]. Accurate anticipation of fruit maturity and quality constitutes an essential foundation for the advancement of precision agriculture and robotic harvesting techniques [3, 36].

However, many quality attributes are measured through laboratory evaluations, which are labor-intensive and destructive. As such, the measurements can only be carried out on physical samples. In practice, horticulturalists rate the strawberries based on their observation of the fruit's color and texture and their experiences. The estimations then become highly subjective and specialized [8, 37].

More recent non-destructive methods predict fruit qualities by computer vision and machine learning. In most studies, strawberries were analyzed after harvesting [7, 12, 37]. Their results demonstrate accurate and automated quality prediction with data-driven techniques. However, as quality is only predicted after harvesting, the method does not help with finding optimal timing to harvest, because strawberries are non-climacteric fruits that do not continue to ripen after harvesting. To address this challenge, we explored deep learning and computer vision techniques to estimate the relevant quality attributes in the wild.

The main findings of the paper are: i) Infield predictions of the annotated attribute (ripeness) and measured attributes (firmness and sweetness) are decently feasible, for which environment data contribute to enhance the performance; ii) The size classes of strawberries infield can be estimated by the cameras with a prime lens, in which shape inpainting models and stereo vision algorithms could involve and slightly improve the accuracies.

2. Related work

2.1. Images in fruit quality predictions

While horticulturalists and customers estimate fruit by their external qualities, such as shape and color, the internal quality of fruits also influences customer appreciation [18]. Computer vision has been used extensively to imitate human judgments in estimating the quality attributes of fruits. Moreover, deep learning models, serving as a more consistent evaluator, have the potential to mitigate subjectivity in assessments. Related research utilized images from the visual and hyper-spectrum in this task.

Figure 1. The first two plots are picture examples taken by a pair of cameras at the same time. The pair of cameras consists of an RGB camera and another one with the OCN color filter. As the OCN camera was placed to the right of the RGB camera with a fixed distance, the strawberries are shifted to the left in the OCN image. An illustration of the calibration of the first two images is shown in the third plot.

Visible spectral forms the fundamental basis for human judgment [25]. Computer vision can get good accuracy on quality prediction because chemical changes that occur during fruit ripening are visible via RGB. For instance, the destruction of chlorophyll, a green pigment that is frequently found in plants, plays a vital role in the ripening process [5]. Anthocyanin is a pigment that appears red in strawberries and is synthesized during the ripening process [13]. Given that the absorption peaks of anthocyanin and chlorophyll occur at wavelengths of 535 nm and 680 nm, respectively [20], the concentrations can be observed visually by the naked human eye. Research using RGB images achieved an accuracy of 85.6% on three classes of strawberry ripeness [12]. RGB is highly correlated with *total soluble solids* (TSS) [1]. TSS is a measure of soluble solids, primarily sugar, within a substance. 79.2% accuracy was achieved on six classes of TSS using a Support Vector Machine (SVM) and an RGB camera [4]. CaffeNet reached 95% accuracy on classifying whether strawberries were mature [16].

Hyperspectral imaging is a technique that combines spatial and spectral information: each pixel contains an entire wavelength spectrum [33]. Certain constituents in strawberries absorb infrared light and reflect frequencies that indicate quality attributes [8]. Our eyes only see a part of the electromagnetic spectrum. Hyperspectral imaging can observe ranges beyond what the human eyes can see, which could improve quality predictions as it shows more details of light absorbance. To determine ripeness, many works first select optimal wavelengths and then use classifiers to determine ripeness classes of strawberries [8, 37] and other fruits such as persimmon [33]. For example, 98.6% accuracy on determining ripeness has been achieved [8]. Hyperspectral imaging has also been shown to be effective to predict measurable quality parameters, such as firmness [20], sweetness [2, 7, 26], and titratable acidity [2, 29]. A disadvantage of hyperspectral imaging is that devices to acquire such images are often expensive and complicated [38], making them unpractical for some farmers.

2.2. Environmental influence in fruit quality

Several environmental factors correlate significantly with strawberry quality, and can thus be used to improve quality predictions. Radiation, temperature, and relative humidity are all correlated with Brix values [6, 34]. Optimal temperature and light increase Brix values [30]. The temperature of both greenhouse and soil influences growth [14]. Weather patterns, such as solar radiation and wind, can also influence growth [19]. Hence, we consider the environmental data as a potential resource to enhance our quality prediction accuracies.

2.3. Size estimation with cameras

Size information is crucial to categorize the strawberries for marketing. As a type of soft fruit, it is noticed that horticulturalists need to limit the time of touching the delicate skins. Hence size estimation is also of interest to horticulturalists. Depth information can help to estimate the true size of an object on a picture because objects can have varying distances to a camera. When the fruit is monitored under a binocular or certain monocular vision system, we can estimate the size by the depth information [10, 21]. Particularly, in a two-camera system, when we know the location and the specified viewing angles of the object in the two views, we could better estimate the depth of an object [22, 39].

However, infield images of fruits can be occluded by obstacles such as other fruits and leaves. This makes it difficult to estimate their actual size. A possible solution is image inpainting, which uses networks such as Convolutional Neural Networks (CNN) and/or Generative Adversarial Networks (GAN) to complete the missing parts of an image. It has been performed both for specific tasks, such as face completion [17], and for a wide variety of images or shapes [23, 32]. For occlusions with limited data available, one approach has been to manually occlude part of the dataset so as to train the inpainting models [9].

3. Method

We used a comprehensive dataset[35] in this research. The dataset provides three types of data: pre-harvest in-field images, post-harvest quality measurements and assessments, and corresponding records about the cultivation environment. The images were collected under a dual-camera setting: an RGB and an Orange-Cyan-Near-Infrared (OCN) camera, which had the same fixed focal length and faced the strawberry plants in parallel views. We used such a dataset with both RGB and OCN images for two reasons. First, the OCN data allowed us to observe more color bands, leading to possibly better quality predictions. Second, the extra camera allowed us to improve depth estimation, which could lead to better size estimation. Two example images are shown in Figure 1. The dataset gives four quality attributes per assessed strawberry: ripeness, firmness (kg/mm^2), sweetness ($°Brix$, also called Brix), and size classes. The quality attributes are connected with image segments of strawberries as labels. In addition, the dataset provides the environmental records in the greenhouse, including temperatures at different locations, air humidity, and radiation. Since previous research has demonstrated the correlation between these environmental factors and fruit quality [6, 14, 19, 30], we used them in improving the accuracy of quality predictions.

3.1. Quality prediction

We used a dataset with ripeness values for 254 strawberries and quality attributes for 184 strawberries [35]. The ripeness was rated by horticulturalists through their observation, on a scale from 1 to 10. A higher score indicates higher maturity, where 7-8 is the optimal range. Both firmness and sweetness were measured using destructive instruments in the laboratory: firmness is defined by the pressure penetrating through the skin; and sweetness is a measure of the soluble solids content of a substance, primarily the sugar in the juice. The sizes of strawberries are described by classes defined by the width of fruits: *tiny* ($<$ 20mm), *small* (20mm - 25mm), and *coarse* ($>$ 25 mm). The distributions of labels are shown in Figure 2. The attributes are labeled to the strawberry in the images. Since we have a relatively low amount of images, we augmented the training data by eight times by using all permutations of randomly flipping, rotating, and cropping. We also tailored the loss function to push the model to satisfy the relatively uncommon values more, inspired by [24]. We did not change the color of the strawberries in the images, so we increased the volume of training data without changing the attributes that would affect the quality prediction.

We trained two convolutional neural networks with classical architectures: *LeNet-5* [15] and *ResNet-18* [11] for demonstrating quality prediction from images. Both the networks in the experiments were trained from scratch.

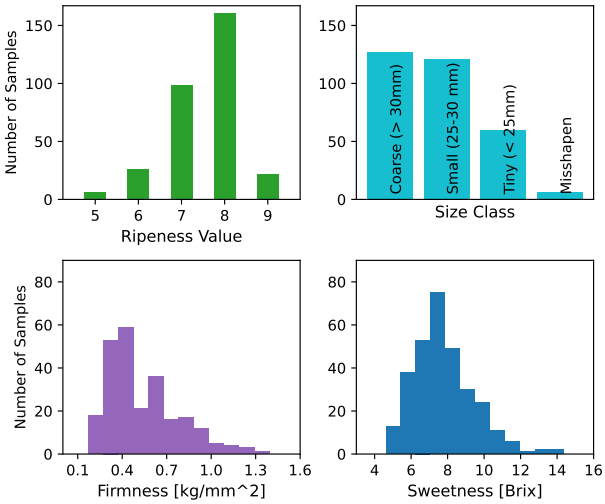

Figure 2. Distribution of quality attributes in our dataset. Each bar plots the number of strawberry samples for each possible value of the quality attribute. The corresponding attributes are indicated by the x-axes.

When using the models, we added a second linear layer so that the network can learn from the environmental data. We ran on 100 epochs and use early stopping on the validation set to select the model with the best performance on this set. For each run, we randomly split the data into the training, validation, and test set by 3:1:1 with a fixed random seed.

To assess the efficacy of introducing the environment features in conjunction with tailored loss functions, we undertook a comparative analysis of the outcomes of employing these strategies and their absence individually. The experimental findings from each distinct experimental scenario are discussed in the subsections following the main quality-prediction results.

3.2. Size estimation

We considered the parallel camera setting as a quasi-binocular positioning, hence we applied stereo vision algorithms to estimate the size of the strawberries. Our methodology consisted of four steps: first, we calibrated the relative position of the cameras by the images. Through translation, scaling, and rotation, the vertical disparity is eliminated and the horizontal disparity is similar among all images. A calibration example is shown in Figure 1. After that, we defined a matching algorithm to match the strawberries from the RGB and OCN images. By knowing the position of the same strawberry in each image, we estimated the depth of the strawberries to the camera hanger, similarly to [22]. Finally, we calculated the size of the strawberries based on the focal information of the camera and the estimated depth.

3.2.1 Segment matching

The same strawberry appears on both the RGB and the OCN images when it was in the overlapped view. In the first step, we calibrated the image pairs by eliminating the vertical disparity and averaging the horizontal disparity. To this end, the OCN picture was translated, rotated, and scaled, as demonstrated in the third plot of Figure 1.

For depth estimation, we need to match each strawberry from the RGB image to a strawberry in the OCN image, *i.e.*, finding the segment on the OCN camera that displays the same strawberry for each RGB strawberry segment. We proposed a simple loss function that involves two components: the distance in location and the distance in size. The intuition is that for each RGB strawberry segment, the further away the OCN segment is, and the more different it is in size, the less likely it is that they are matches.

We followed a method outlined in [31] to calculate the distance in size between the strawberries:

$$ds = \frac{|\text{surface}_{\text{RGB}} - \text{surface}_{\text{OCN}}|}{\text{surface}_{\text{RGB}} + \text{surface}_{\text{OCN}}}, \tag{1}$$

where *surface* is the number of pixels of a segment. The formula calculates the difference in size and ds is in the range $[0, 1]$.

The formula we used for the distance in location is:

$$dl = \sqrt{dx^2 + dy^2 \cdot \alpha}, \tag{2}$$

where dx is the distance between the x-coordinates, after subtracting the expected disparity in pixels; dy is the distance between the y-coordinates; and α is a tunable parameter. $\alpha \geq 1$ increases the weight of vertical disparity in the loss function: while the horizontal disparity depended on the distance of a strawberry to the camera, and is thus flexible, the vertical disparity should be near zero since the cameras were at the same height.

We multiplied Equation 1 and Equation 2 to calculate a loss between each strawberry in the RGB image and each strawberry in the corresponding OCN image:

$$loss = dl \cdot (ds + \beta), \tag{3}$$

where β is a tunable parameter that stabilizes the loss value for low values of ds. As we targeted minimizing the distances in location and size, we chose the strawberry segments with the smallest loss as a match. We used $\alpha = 10$ and $\beta = 0.01$ empirically based on our fine-tuning tests.

3.2.2 Depth estimation

Given the matched OCN strawberry segment for each RGB segment, we estimated the depth of a strawberry to the line of the camera pairs. We followed a standard method to calculate depth by using stereo vision as outlined in [22, 39], as illustrated in Figure 3. We calculated the depth in millimeters as follows:

$$D = \frac{b \cdot f}{d}, \text{ where } f = \frac{w_{\text{img[px]}}}{2\tan(\frac{\theta}{2})} \tag{4}$$

where b is the baseline distance between the two cameras in millimeters, f is the focal length in pixels, and d is the disparity between the RGB and the OCN segments in pixels. For each camera, $w_{\text{img[px]}} = 4000$px is the width in pixels of the output and $\theta = 41^\circ$ is the angle.

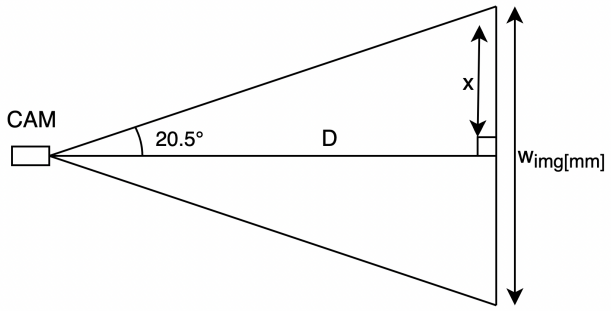

Figure 3. A triangulation method to find the width of the image in millimeters at a given depth. The camera was located at a distance of D from the strawberry. We used an angle of 20.5° in this illustration as the camera's field of view (FOV) is 41°. Using triangulation, we find x; $2x$ is the width of the entire image.

3.2.3 Size estimation

We estimated the width of a strawberry by calculating the depth and counting the width of the segment in pixels. Here, we first calculated the width of the entire image at the depth of the strawberry. We used a simple triangulation method, as illustrated in Figure 3. Given the FOV of the camera, we found the width of the image in millimeters at a given depth. Then we multiplied the image width by the fraction of how the strawberry occupied in the image, so as to find the width of the strawberry in millimeters:

$$w_{\text{str[mm]}} = \frac{w_{\text{str[px]}}}{w_{\text{img[px]}}} \cdot 2\tan(\frac{\theta}{2}) \cdot D \tag{5}$$

where $w_{\text{str[px]}}$ is the width of the segment of a strawberry in pixels. The above formulas can be used to calculate the width of a strawberry in millimeters.

3.3. Shape inpainting

In the infield images, some strawberries were occluded. This degraded size estimation performance, as the width in pixels might be less than it would be without occlusions. In this section, we outline an image inpainting algorithm that aims to recover the original strawberry shape. The inpainting model was applied as an additional data pre-processing step before the actual size estimation.

3.3.1 Determining occlusions

For the purpose of size estimation, we aimed to only inpaint occluded strawberries, as performing inpainting on a non-occluded strawberry could decrease the performance of size estimation. Hence, we created a simple method to estimate if a strawberry is occluded or not: by calculating the circularity [28]. We calculated the circularity by dividing the squared root of the surface area of a segment shape by its circumference, as in Equation 6.

$$\text{circularity} = \frac{\sqrt{\text{surface}}}{\text{circumference}}. \quad (6)$$

Empirically, a circularity of 0.25 or less indicates an occluded strawberry. An example of detecting occlusions using circularity is illustrated in Figure 4. Sometimes, an occluded strawberry's circularity might still exceed the defined threshold. Nonetheless, since the width difference in such cases is likely to be small, this should not significantly affect the performance of our size prediction requirement.

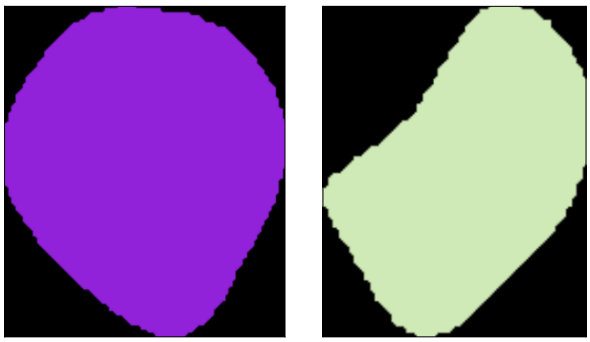

Figure 4. Two examples of strawberry segments. Left: strawberry segment with a circularity of 0.263: not occluded; right: strawberry segment with a lower circularity of 0.245: occluded.

3.3.2 Inpainting models

Even with a perfect estimation of the depth of strawberries, we still could not estimate the exact size of strawberries due to the occlusion in the dense growth pattern. Therefore, we introduced an inpainting algorithm to train models to recover the actual shape of strawberries.

We gathered 279 strawberry segments to train the inpainting models. We used 25% of the segments for the *occlusion set*, which were used to artificially occlude the input segment. Of the remaining part of the image segments, 75% went to the training set and the rest to the test set. The strawberries kept their aspect ratio and were scaled down to take up at most half the input image. We used two configurations: RGBA inputs (with a binary opacity channel A to represent the polygon shape of the segment) and binary inputs, which are the opacity-channel values only. The binary inputs made it easier for the network to focus on the shape, rather than on the colors.

We used U-Net for the image inpainting task. The inputs and outputs were both 256x256. At every epoch, each strawberry was occluded by overlapping it with a randomly chosen transparent strawberry from the occlusion set. This strawberry was also placed in the center but shifted in both the x and y direction. To give a varying amount of occlusion per training sample, a shift value of either 32 or 64 pixels was chosen with a probability of 50%. After occluding, the strawberry was re-centered, since the input would also be centered on a truly occluded strawberry. To model occlusions from more than one strawberry in the dense growing pattern, 50% of the time, we placed two occlusions.

We used mean squared error (MSE) as the loss function to train the inpainting model. Additionally, for indication of performance, we calculated the Intersection over Union (IoU) and the difference in width between the network output and the ground truth.

4. Experiments

4.1. Quality prediction with image and climate data

We evaluated the firmness and sweetness prediction results by inputting either the RGB image, the OCN image, or both. For ripeness, we only input the RGB images, as the ripeness values were determined by experts subjectively based on their observations. The OCN inputs were the strawberry segments from the OCN images, which were matched to the observations of the same strawberries in the RGB image. They were found by using our matching algorithm described in Section 3.2.1. To train the network on the pair of RGB and the OCN images, the convolutional layers were duplicated: the first part was used on the RGB image, and the second part on the OCN image. The outputs were concatenated and fed into a multi-layer perceptron (MLP). It was possible that the matching strawberry was not visible on the OCN camera due to the disparity. In this case, for the dual network, we inserted a black and transparent image instead, *i.e.*, using $RGBA = (0, 0, 0, 0)$ for all the pixels.

To find the best performance on quality prediction, we trained CNNs with both *LeNet-5* and *ResNet-18* architectures on the RGB, the OCN, or the pair of images of individual strawberries. To reach the best model performances, we add two means to enhance the model performances: i) we tailored the loss function during model training; ii) the models from all the model-data configurations were trained with both the image data and the relevant environmental data, which were both the average of the past four days before harvest and the average over the past four days in the week before harvesting. To ensure a proper comparison, we performed five runs for each setup, using fixed random seeds that remained consistent across various configurations.

The best performances on the estimation of ripeness, firmness, and sweetness were obtained using *LeNet-5* on only the RGB inputs, as shown in Table 1, Table 2, and Table 3 respectively. We use Mean Squared Error (MSE) as the performance indicator of all tables. We discuss the influence of the two improvement measures after the reporting of the best-performing models out of all configurations. Our results demonstrate the feasibility of predicting the quality attributes (i.e. ripeness, firmness, and sweetness) by the images of strawberries. The introduction of both the tailored loss functions and the environmental data facilitated the performances to various extents.

Table 1. Result comparison of ripeness estimation on the test datasets. MSE is the average MSE loss over five experiments. We also denote the standard deviation here. The best performance was achieved on the model with a *LeNet-5* architecture.

Color mode of Image data	MSE$\cdot 10^{-1}$ with *LeNet-5*	MSE$\cdot 10^{-1}$ with *ResNet-18*
RGB	**6.30** $\pm$ 0.74	7.84 $\pm$ 2.04

Table 2. Result comparison of firmness prediction on the test datasets. MSE is the average MSE loss over five experiments. We also denote the standard deviation here. The model with a *LeNet-5* architecture and trained on RGB input had the best performance.

Color mode of Image data	MSE$\cdot 10^{-2}$ with *LeNet-5*	MSE$\cdot 10^{-2}$ with *ResNet-18*
RGB	**3.60** $\pm$ 0.55	4.27 $\pm$ 0.58
OCN	5.77 $\pm$ 1.74	5.78 $\pm$ 0.97
Dual	3.76 $\pm$ 0.22	4.84 $\pm$ 1.12

Table 3. Result comparison of sweetness prediction on the test datasets. MSE is the average MSE loss over five experiments. We also denote the standard deviation here. The model with a *LeNet-5* architecture and trained on RGB input had the best performance.

Color mode of Image data	MSE with *LeNet-5*	MSE with *ResNet-18*
RGB	**1.39** $\pm$ 0.44	1.81 $\pm$ 0.58
OCN	1.69 $\pm$ 0.49	2.06 $\pm$ 0.78
Dual	1.51 $\pm$ 0.45	2.02 $\pm$ 0.62

4.1.1 Improvements by tailoring loss functions

As indicated by Figure 2, the quality attributes were not uniformly distributed, where certain values had few occurrences. Thus, the network could be biased towards the most common values, and never predict the lowest or highest values. We found that the performance was degraded when the testing data split had more uncommon values. Therefore, we proposed a solution by increasing the weight of the loss on uncommon values. We applied a Weighted MSE (WMSE) that multiplied the loss by how uncommon the value was in the entire dataset. For example, on a binary dataset that has two instances of X and one instance of Y, the loss on the Y value will be weighted such that it is counted twice as much. Furthermore, we added an exponent to our weight values. We tried both exponents of 0.5 and 2 and named them the Square Root WMSE (SQWMSE) and the Squared WMSE (SWMSE), respectively.

When we kept our optimal data-model configurations, we found that WMSE and SQWMSE improved the performance of predicting firmness and ripeness of strawberries, but degraded the performance of sweetness prediction. Generally, the models trained by WMSE performed better than those with SQWMSE in these experiments. We present the MSE on the test datasets for each quality attribute in Table 4 from the models trained with the specified loss functions. Thus, for ripeness and firmness prediction tasks, we reported the performances of the models trained on WMSE in Table 1 and Table 2.

Table 4. Result comparison of models with various loss functions and with the *LeNet-5* architecture. The model performances are indicated by MSE on the test datasets. The columns indicate the loss function that the models were trained on. We could notice that the MSE for ripeness and firmness on the test datasets was reduced when increasing the weight of the loss on uncommon values.

Attribute	Trained on MSE	Trained on WMSE
Ripeness	6.76 $\pm$ 1.06 ($\cdot 10^{-1}$)	**6.30** $\pm$ 0.74 ($\cdot 10^{-1}$)
Firmness	4.01 $\pm$ 0.94 ($\cdot 10^{-2}$)	**3.60** $\pm$ 0.55 ($\cdot 10^{-2}$)
Sweetness	**1.39** $\pm$ 0.44	2.02 $\pm$ 0.64

4.1.2 Climate data contributed to the performance

Previous research has demonstrated the feasibility of using environmental data to estimate the qualities of batches of fruits. Our experiments used such data to enhance the performance of individual quality predictions. Specifically, to find the effect of adding environmental data to the predictions, we re-ran the models with the best-performing configurations but without the environmental data.

Table 5. Result comparison of quality estimation with or without environmental data (env.). The performances are described by MSE on the test datasets. The columns indicate the input data that the models were trained on. Notable performance decline occurs in sweetness prediction, while firmness and ripeness estimations show minor degradation when environmental data is omitted.

Attribute	Trained with env.	Trained without env.
Ripeness	**6.30** $\pm$ 0.74 ($\cdot 10^{-1}$)	6.65 $\pm$ 0.97 ($\cdot 10^{-1}$)
Firmness	**3.60** $\pm$ 0.55 ($\cdot 10^{-2}$)	3.86 $\pm$ 0.46 ($\cdot 10^{-2}$)
Sweetness	**1.39** $\pm$ 0.44	2.53 $\pm$ 0.71

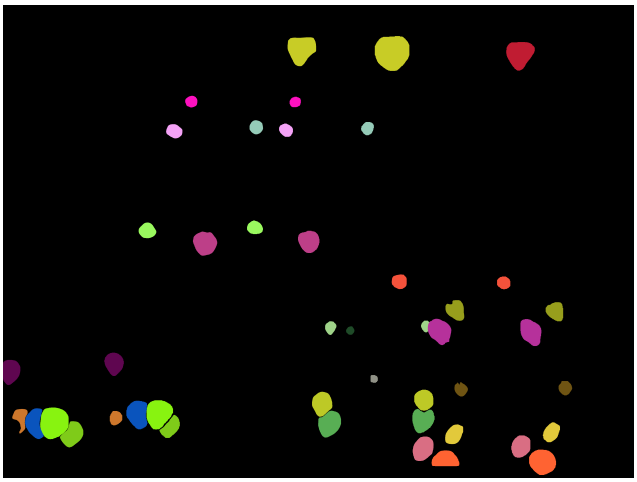

Figure 5. Example of matching of the images of Figure 1. Most strawberries appear twice since both the segments from the RGB and the OCN cameras are shown. We evaluated the matching performances qualitatively according to these color-coded visualizations. Overall, the matching performance is decent on the dataset. Such proper matching allowed us to calculate the disparity, depth, and size of a strawberry accordingly.

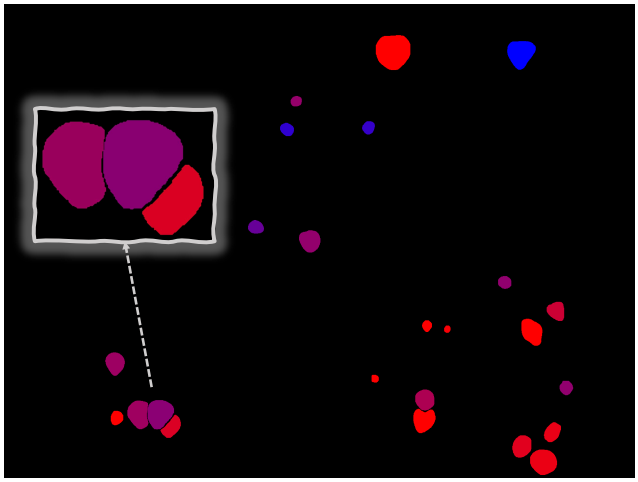

Figure 6. Example of depth estimations of Figure 1, shown as a color-encoded mask of the RGB image. The more blue a segment is, the closer it is; the more red, the further it is away. The depths range roughly from 800mm to 1000mm. A close-up of three strawberry segments is depicted on the top left of the plot. The depth estimation algorithm demonstrated promising local performance as it tended to position occluded segments at a greater distance.

As is shown by Table 5, all the accuracies of the tasks decreased if the models did not learn from the climate data. Notably, the performance of sweetness prediction was impacted by the environmental data most strongly.

4.2. Size classification with images

By implementing Equation 5, we could estimate the size of a strawberry according to its distance from the camera. We used the distance from the RGB camera to its projected point on the planting basket (*i.e.* D in Figure 3), which was a constant number for the setup, as a baseline for our size estimation. We compared the size estimation with adding the depth calculation and/or the shape inpainting. As the ground-truth labels are only classes of the width, we mapped the calculated width into size classes and measured the precision of the classification. Results are shown in Table 6. We notice a slight accuracy improvement when depth calculation was involved, yet the involvement of shape inpainting did not help as we hypothesized.

Table 6. Effects of depth calculation and inpainting on the size classification test. The "method" indicates the way of choosing the variables in the implementation of Equation 5. The accuracy was measured by the precision after discretizing the calculated width into size classes.

Method	Accuracy
With a fixed depth (Baseline)	63.1%
With depth calculation	**64.5%**
With shape inpainting	62.4%
With depth calculation and shape inpainting	63.8%

4.2.1 Segment matching and depth estimation

As there was no ground-truth information for segment matching between RGB and OCN images, we only evaluated the matching algorithm manually and qualitatively. We show an example of the matching results in Figure 5. Most strawberries were matched correctly, with two major exceptions. First were a few tiny strawberries. The performance of those for this research was not relevant, as they were not mature enough to be harvested soon. Second were the large strawberries at the top of the images, which were from the plant above the cameras. Nevertheless, the performance would not be influenced by these fruits either, as these strawberries were not part of the measurements.

Using the disparity, we calculated the depth. Figure 6 visualizes the depth estimation results. Still, due to the lack of ground truth on object depths, we assessed the algorithm performance empirically. In the visualization and the close-up, we show that the algorithm generally performed well. We observed that discerning depth differences in the images was challenging for human eyes. For instance, in Figure 6, certain strawberries appeared to have similar depths on both the left and right sides of the view. However, we were unable to verify the accuracy of this perception.

4.2.2 Shape inpainting

We applied our inpainting methodology to two configurations: using RGBA or solely the binary channel as the inputs. By calculating the circularity, we inpainted only the occluded strawberries to avoid possible performance loss.

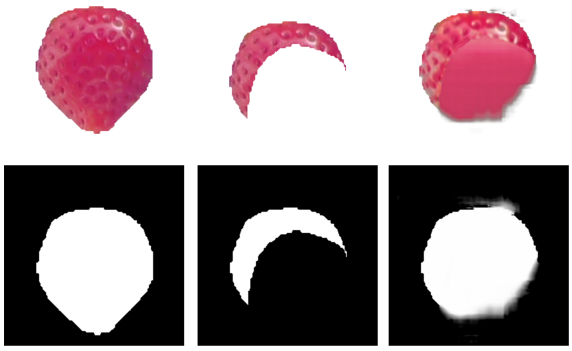

Figure 7. Example of results of the inpainting algorithm on a strawberry. The first row used RGBA inputs, the second row used binary inputs. The left column is the ground truth. The middle column is the network input. The right column is the predicted output. The performance on both configurations was good: the original shape is reasonably recovered, allowing us to estimate the strawberry width properly.

Figure 8. Examples of results of the inpainting algorithm on *real* occluded strawberries. The first row used RGBA inputs, the second row used binary inputs. The left column is the network input. The right column is the predicted output. There is no non-occluded ground truth of the strawberries, as the strawberries are occluded infield. The performance on the RGBA configuration is poor and on the binary configuration is mediocre.

Results are listed in Table 7. Although metrics like MSE and IoU are not designed to indicate shape similarity, they could still give an indication of reconstruction quality. The model with RGBA input images achieved a lower MSE loss, while the model with binary input images achieved a higher IoU and lower width difference.

The results indicated that the inpainting models could recover part of the occlusion in the segments. Given artificial occlusions, the models performed well, as illustrated in Figure 7. However, when the models met real occlusions, *i.e.*, the occlusions presented in our data, the performances varied. When comparing Figure 8 to Figure 7, it is implied that the networks did not learn a general representation of a strawberry, but rather, they learned how to inpaint certain occlusions, which were the typical situations that we found of how strawberries could be occluded by others in the wild. This also indicates that it was challenging to simulate a wide-enough variety of occlusions such that the network learned a general way to inpaint the shape of a strawberry. As the inpainting models could not handle to all types of real-life occlusions, they did not improve the accuracy of the size classification task.

Table 7. Results of inpainting on the test set. Width diff is the difference in width between the network output and the ground truth. The models have roughly equal IoU, but the binary model has a higher loss but a lower width difference. It is not immediately clear which configuration is best.

Color mode	MSE $\cdot 10^{-3}$	IoU	Width diff
RGBA	**5.46**	0.871	2.00
Binary	7.21	**0.874**	**1.70**

5. Conclusion

In this paper, we propose methods of using infield data for quality prediction and size classification of strawberries, such that analyses can be conducted before harvesting. Our deep learning models achieved reasonable accuracy on predictions of ripeness, firmness, and sweetness. Our results demonstrate that models trained under a tailored loss function were more robust against class imbalance. The environmental data also facilitated the model performance. We also achieved proper size classification using the infield images. Calculating depth through stereo-vision algorithms improved the accuracy, but only slightly. The inpainting models performed well on artificial occlusions, which followed a similar methodology for generating the training data. Nevertheless, the progress of the models in terms of size classification differed when confronted with actual real-world occlusions.

The performances of the quality prediction models are relatively inferior to that of a few related works. The reason could be the lack of controlled lighting conditions, as we used solely infield data. Further, our work did not utilize the OCN image data as expected. Considering the decent performance of related works using hyperspectral imaging [38], it is suggested to explore appropriate calibration methods to make the near-infrared spectrum a practical and accessible option for horticulturalists.

In terms of size classification, as the size information was only provided by discrete labels, the precision did not linearly represent the changes in the model performance under different methods. Therefore, it is meaningful to continue further research on datasets with better calibrated, more detailed, or larger volumes of measurements.

Acknowledgements

We would like to express our gratitude to Topsector Tuinbouw & Uitgangsmaterialen (The Netherlands) for their financial support of this research.

References

[1] P.D. Abeytilakarathna, R.M. Fonseka, J.P. Eswara, and K.G.N.A.B. Wijethunga. Relationship between total solid content and red, green and blue colour intensity of strawberry (fragaria x ananassa duch.) fruits. *Journal of Agricultural Sciences*, 8, 07 2013. 2

[2] Maria Luisa Amodio, Francesco Ceglie, Muhammad Mudassir Arif Chaudhry, Francesca Piazzolla, and Giancarlo Colelli. Potential of nir spectroscopy for predicting internal quality and discriminating among strawberry fruits from different production systems. *Postharvest Biology and Technology*, 125:112–121, 2017. 2

[3] Tawseef Ayoub Shaikh, Tabasum Rasool, and Faisal Rasheed Lone. Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Computers and Electronics in Agriculture*, 198:107119, 2022. 1

[4] Jayanta Kumar Basak, Bolappa Gamage Kaushalya Madhavi, Bhola Paudel, Na Eun Kim, and Hyeon Tae Kim. Prediction of total soluble solids and ph of strawberry fruits using rgb, hsv and hsl colour spaces and machine learning models. *Foods*, 11(14), 2022. 2

[5] C J Brady. Fruit ripening. *Annual Review of Plant Physiology*, 38(1):155–178, 1987. 2

[6] Fatima I Pereira da Silva, Sabine K Schnabel, Bastiaan Brouwer, and Manon G Mensink. Monitoring strawberry production to get grip on strawberry quality: Greenchainge fruit & vegetables wp3. *GreenCHAINge Fruit & Vegetables*, 2018. 2, 3

[7] Gamal ElMasry, Ning Wang, Adel ElSayed, and Michael Ngadi. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Journal of Food Engineering*, 81(1):98–107, 2007. 1, 2

[8] Zongmei Gao, Yuanyuan Shao, Guantao Xuan, Yongxian Wang, Yi Liu, and Xiang Han. Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning. *Artificial Intelligence in Agriculture*, 4:31–38, 2020. 1, 2

[9] Liang Gong, Wenjie Wang, Tao Wang, and Chengliang Liu. Robotic harvesting of the occluded fruits with a precise shape and position reconstruction approach. *Journal of Field Robotics*, 39, 10 2021. 2

[10] Mohd Saad Hamid, NurulFajar Abd Manap, Rostam Affendi Hamzah, and Ahmad Fauzan Kadmin. Stereo matching algorithm based on deep learning: A survey. *Journal of King Saud University - Computer and Information Sciences*, 34(5):1663–1673, 2022. 2

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 3

[12] Indrabayu Indrabayu, Nurhikma Arifin, and Intan Sari Areni. Strawberry ripeness classification system based on skin tone color using multi-class support vector machine. In *2019 International Conference on Information and Communications Technology (ICOIACT)*, pages 191–195, 2019. 1, 2

[13] M.N. Islam, Mehnaz Mursalat, and Mohidus Samad Khan. A review on the legislative aspect of artificial fruit ripening. *Agric & Food Secur*, 5, 06 2016. 2

[14] Dong Sub Kim and Sung Kim. Prediction of strawberry growth and fruit yield based on environmental and growth data in a greenhouse for soil cultivation with applied autonomous facilities. *Wonye kwahak kisulchi: Korean journal of horticultural science and technology*, 38, 12 2020. 2, 3

[15] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 3

[16] Xin Li, Jun Yu Li, and Jing Tang. A deep learning method for recognizing elevated mature strawberries. *2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pages 1072–1077, 2018. 2

[17] Yijun Li, Sifei Liu, Jimei Yang, and Ming-Hsuan Yang. Generative face completion, 2017. 2

[18] Manuela Mancini, Luca Mazzoni, Francesco Gagliardi, Francesca Balducci, Daniele Duca, Giuseppe Toscano, Bruno Mezzetti, and Franco Capocasa. Application of the non-destructive nir technique for the evaluation of strawberry fruits quality parameters. *Foods*, 9(4), 2020. 1

[19] Mahesh Maskey, Tapan Pathak, and Surendra Dara. Weather based strawberry yield forecasts at field scale using statistical and machine learning models. *Atmosphere*, 10:378, 07 2019. 2, 3

[20] Binu Melit Devassy and Sony George. Estimation of strawberry firmness using hyperspectral imaging: a comparison of regression models. *Journal of Spectral Imaging*, 10, 06 2021. 2

[21] Mircea Paul Muresan, Marchis Raul, Sergiu Nedevschi, and Radu Danescu. Stereo and mono depth estimation fusion for an improved and fault tolerant 3d reconstruction. In *2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP)*, pages 233–240. IEEE, 2021. 2

[22] Dhaval K. Patel, Pankaj A. Bachani, and Nirav R. Shah. Distance measurement system using binocular stereo vision approach. *International journal of engineering research and technology*, 2, 2013. 2, 3, 4

[23] Deepak Pathak, Philipp Krähenbühl, Jeff Donahue, Trevor Darrell, and Alexei Efros. Context encoders: Feature learning by inpainting. In *Computer Vision and Pattern Recognition (CVPR)*, 2016. 2

[24] Jiawei Ren, Mingyuan Zhang, Cunjun Yu, and Ziwei Liu. Balanced mse for imbalanced visual regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7926–7935, 2022. 3

[25] Matteo Rizzo, Matteo Marcuzzo, Alessandro Zangari, Andrea Gasparetto, and Andrea Albarelli. Fruit ripeness classification: A survey. *Artificial Intelligence in Agriculture*, 2023. 2

[26] Ana Santos, Sara Ricardo Rodrigues, Marta Laranjo, C. Melgão, and R. Velázquez. Non-destructive prediction of total soluble solids in strawberry using near infrared spectroscopy. *Journal of the Science of Food and Agriculture*, 102, 03 2022. 2

[27] K. Sturm, D. Koron, and F. Stampar. The composition of fruit of different strawberry varieties depending on maturity stage. *Food Chemistry*, 83(3):417–422, 2003. 1

[28] Sylwia Szerakowska, Maria Sulewska, Jerzy Trzciński, and Barbara Woronko. Comparison of methods determining particle sphericity. *Applied Mechanics and Materials*, 797:231–237, 11 2015. 5

[29] María-Teresa Sánchez, María José De la Haba, Miriam Benítez-López, Juan Fernández-Novales, Ana Garrido-Varo, and D.C. Perez-Marin. Non-destructive characterization and quality control of intact strawberries based on nir spectral data. *Journal of Food Engineering*, 110:102–108, 05 2012. 2

[30] Yilian Tang, Xun Ma, Ming Li, and Yunfeng Wang. The effect of temperature and light on strawberry production in a solar greenhouse. *Solar Energy*, 195:318–328, 2020. 2, 3

[31] Kris van Melis. Measuring the size of strawberries using binocular photos. Dissertation for Bachelor of Computer Science and Engineering, 2022. 4

[32] Weiyue Wang, Qiangui Huang, Suya You, Chao Yang, and Ulrich Neumann. Shape inpainting using 3d generative adversarial network and recurrent convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 2

[33] Xuan Wei, Fei Liu, Zhengjun Qiu, Yongni Shao, and Yong He. Ripeness classification of astringent persimmon using hyperspectral imaging technique. *Food and Bioprocess Technology*, 7:1371–1380, 2013. 2

[34] Junhan Wen, Thomas Abeel, and Mathijs de Weerdt. "how sweet are your strawberries?": Predicting sugariness using non-destructive and affordable hardware. *Frontiers in Plant Science*, 14:1160645, 2023. 2

[35] Junhan Wen, Mathijs de Weerdt, Thomas Abeel, Camiel Verschoor, Lisanne Schuddebeurs, Klaas Walraven, Stijn Jochems, and Vera Theelen. Data underlying the research of quality prediction of strawberries with rgb image segments, 2023. 3

[36] Ya Xiong, Yuanyue Ge, Lars Grimstad, and Pål J From. An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation. *Journal of Field Robotics*, 37(2):202–224, 2020. 1

[37] Chu Zhang, Chentong Guo, Fei Liu, Wenwen Kong, Yong He, and Binggan Lou. Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine. *Journal of Food Engineering*, 179:11–18, 2016. 1, 2

[38] Jingang Zhang, Runmu Su, Qiang Fu, Wenqi Ren, Felix Heide, and Yunfeng Nie. A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging. *Scientific Reports*, 12(1), jul 2022. 2, 8

[39] Manaf Zivingy. Object distance measurement by stereo vision. *International Journal of Science and Applied Information Technology (IJSAIT)*, 2:05–08, 01 2013. 2, 4