Grading Methods for Fruit Freshness Based on Deep Learning

Yuhang Fu, Minh Nguyen, Wei Qi Yan

Auckland University of Technology, Auckland, 1010 New Zealand

Abstract: Fruit freshness grading is an innate ability of humans. However, there was not much work focusing on creating a fruit grading system based on digital images in deep learning. The algorithm proposed in this article has the potentiality to be employed so as to avoid wasting fruits or save fruits from throwing away. In this article, we present a comprehensive analysis of freshness grading scheme by using computer vision and deep learning. Our scheme for grading is based on visual feature analysis of digital images. Numerous deep learning methods are exploited in this project, including ResNet, VGG, and GoogLeNet. AlexNet is selected as the base network for visual feature extraction, YOLO is selected for extracting the region of interest (ROI) from digital images. Therefore, we construct a novel neural network model for fruit detection and freshness grading regarding multiclass fruit classification. The fruit images are fed into our model for training, AlexNet took the leading position; meanwhile, VGG scheme performed the best in the validation.

Keywords: CNN · Deep learning · Fruit freshness grading · YOLO · AlexNet · VGG

1 Introduction

Given the significance of foods in our ordinary lives, fruit grading becomes crucial but is time consuming. Grading automatically by using computerised approaches is believed as the solution of this problem, which will save human labour. There is a shred of evidence which shows that when fruit deterioration occurs, fruit goes through a series of biochemical transformation that leads to changes in its physical conditions and chemical composition, e.g., changes in nutrition. By the way, most of these features can be captured.

Fruit grading methods are grouped into two categories: Non-visual and visual approaches. Non-visual grading approaches mainly concentrate on aroma, chemicals, and tactile impression. Fruit spoilage in nature is a biochemical process that natural pigments in various reactions are transformed into other chemicals that result in changes of colours. Identifying fruit spoilage is an innate ability of human perception system. It is regarded as the desirability and acceptance to the consumption of a portion. It assists identifing whether the given fruits are edible or not [1].

The research work unfolds that there exists a strong relationship between bacteria and fruit spoilage, which encompasses aerobic psychrotrophic gram-negative bacteria with the secretion of extracellular hydrolytic enzymes that corrupt plant's cell walls, heterofermentative lactobacilli, spore-forming bacteria, yeasts and moulds. Fruit degeneration is a consequence of biochemical reactions, i.e., a structural acidic heteropolysaccharide grown in terrestrial plant's cell walls, chiefly consisted of galacturonic acid. Starch/amylum and sugar (i.e., polymetric carbohydrates with the same purposes) are then metabolised with produced lactic (i.e., an acid that is a metabolic intermediate as the end product of glycolysis releasing energy anaerobically) and ethanol [2]. Colonising and induced lesions as a result of microbe dissemination are frequently observed, and infestation is a primary reason of spoilage for postharvest fruits [3].

Besides, the lack of nutrients results in the growth of dark spots, e.g., insufficient calcium leads to apple cork spots [4]. The exposure to oxygen is another determinant as an enzyme known as polyphenol oxidase (PPO) triggers a chain of biochemical reactions inclduing proteins, pigments, fatty acids and lipids, that lead to fading of the fruit colours as well as degrading to an undesirable taste and smell [5].

The established research evidence shows that if fruit deterioration occurs, fruit goes through a series of biochemical transformation that incrus to changes in its physical conditions, e.g., visual features including colour and shape, most of these features can be extracted. It is affirmed that a computer vision-based approach is the most economical solution.

Previosuly, scale-invariant feature transform (SIFT) along with colour and shape of fruits [22] has been offered to fruit recognition. K-nearest neighbourhood (KNN) and support vector machine (SVM) have been employed for the classification. Despite attaining high accuracy, this approach has input images with the size 90×90, which is low, the information might be dropped. The low-resolution image has the implications that individual pixel may have a significant contribution to the final result, which is dependent on noises for prediction. It is well known that KNN and SVM are vulnerable to the curse of dimensionality where the growth of feature dimensions will have a massive impact on performance, meanwhile high-resolution images are likely to have rich visual features.

Given the advancement of deep learning, fruit grading algorithms should produce satisfactory accuracies timely [6][7]. The state-of-the-art technology in computer vision sees the categories in fruit/vegetable automatic grading [8]: Detections of fruit/vegetable diseases and defects by using foreign biological invasion [10], fruit/vegetable classification for assorted horticultural products [11], estimation of fruit/vegetable nitrogen [12], fruit/vegetable object real-time tracking [13], etc. Most scientific approaches for fruit grading by using pattern classification are classified. Pertaining to fruit quality grading, the focus is on not only the freshness, but zlxo the overall visual changes. Despite the recent rise of popularity of deep learning, more than half of the work [14] did not use deep learning methods.

Fruit recognition using computer vision and deep learning is an interesting research field [19]. The delicious golden apples have been graded by using SVM + KNN [20]. However, the research project has just twofold: Healthy and defect, only one class of fruits was taken into consideration. Another project was developed for tomato grading [21] by using texture, colour, and shape of digital images. A binary classification was proposed that fruits are recognised.

Deep learning was found useful in identifying conditions of citrus leaves [23], which is extremely powerful in image recognition for classification [24]. YOLO [25] was adopted for fruit and vegetable recognition. YOLO is a faster algorithm compared to other approaches, which achieved 40 *fps* (i.e.e, frames per second) in videos that are applicable for real-time applications. However, the fruits are constrained to the conditions when the fruits remain being connected to their biological hosts, e.g., hanging on branches. It does not take account of the scenarios where fruits are taken off from trees in the ongoing process of decaying. VGG was used for fruit recognition [26], the result [26] manifests that convolutional neural network, when going deep, can achieve high accuracy. In contrast to the previous one, a shallow convolutional neural network [27] consisting of four convolution and pooling layers was suggested for feature extraction, followed by two fully connected layers. However, the source images in this experiment are simple, all fruits are placed ideally at a static position in a pure white background.

An automatic grading system was developed for olive by using discrete wavelet transform and statistical texture features [28]. Another work [29] has addressed raspberry recognition by using deep learning successfully, namely, a nine-layer neural network consisting of three convolutional and pooling layers, one input, dense, and output layer.

Mandarin decay process is impacted by a disease called penicillium digitatum; there is a contribution [30] dedicated to early detection of this disease by examining decay visual features. The visual elements are captured and processed by a combination of decision trees. However, these experiments only were conducted based on one class of fruits. Another problem is that the grading mechanism is a classification model which treats the fruit as being either fresh or rotten/defect. Still, we believe that the decay process occurs gradually; the final predictive layer should regress the output rather than perform a classification task.

Motivated by the related work, a novel fruit freshness grading method based on deep learning is proposed in this article. We create a dataset for fruit freshness grading. The dataset is comprised of selected frames from recorded videos for a dataset having six classes. From data collection, the images are resized and labelled (regions of interest, object classes, and freshness grades). Four typical augmentations are used, e.g., adjusted contrast, sharpness, rotation, and added noises. Our experiments embark on the statistics where visual reflections on the observed objects are discussed, followed by the implementation of a hierarchical deep learning model: YOLO+Regression-CNN for fruit freshness grading. This experiment takes into account of four base networks: VGG [15], AlexNet [16], ResNet [17], and GoogLeNet [18]. The main contributions of this paper are:

- We propose a new approach to grade fruit freshness. The fruit freshness matters are generally tackled through pattern classification rather than regression. In this article, the regression of CNN is applied to fruit freshness grading.
- A new framework is put forward for fruit freshness grading. We firstly detect and classify a given fruit as a visual object for freshness grading.
- We injet noises to fruit images so that the developed model is capable of resisting noises introduced from real applications.

In this article, we narrated the work related to fruit freshness grading in Section 2. Then, our dataset is described in Section 3, our method is detailed in Section 4, the experimental results are demonstrated in Section 5. In Section 6, the conclusion of this paper is drawn.

2 Data Collection

Different from the existing work, in this article, we propose deep learning algorithms for fruit freshness grading. As we already know, deep learning is promising to the freshness grading for multiclass fruits that will significantly reduces our human labour. In this article, we provide a detailed description of how we have collected visual data and conducted data augmentation before fruit freshness classification. Given the novelty of this research project, the fruit data is not available at present, thus, we have to collect the data by ourselves. We illustrate our process of how we have received the fruit data and provided empirical evidence on how the dataset accurately represents the fruit freshness.

2.1 Datasets

The collected dataset consists of six classes of fruits: Apple, banana, dragon fruit, orange, pear, and Kiwi fruits, derived from a vast variety of locations in the images with various noises, irrelevant adjacent objects and lighting conditions. We firstly analyse the relationship between fruit appearance and freshness. Fresh apple peel is low in chlorophyll and carotenoid concentrations [31], the spoilage leads to a gradual degradation of the constituent pigments, that reflect different wavelengths in spectrophotometry. A ripe banana having bright yellow colour is likely a result of carotenoid accumulation [32]. The main compositions of orange peels and flesh are pectin, cellulose, and hemicellulose if excluding water that represents 60% - 90% of weights [33, 34], the pigments are mostly carotenoids and flavonoids that generate red appearance of oranges. The exotic, aesthetic, and exterior look of dragon fruit is comprised of red-violet betacyanins and yellow betaxanthins [35]. The green colour of Kiwi fruits is a visual manifestation of chlorophylls if the degrading gives rise to the formation of pheophytins and pyropheophytins that render olive-brown colour to the fruit [36]. The green/yellow peel of a pear is a result of congregated chlorophylls, once degradation occurs, chlorophylls degenerate blue-black pheophytins, pyropheophytins are produced [36].

In total, we have collected approximately 4,000 images with each class of fruits around 700. We split the dataset into training and validation sets at the ratio of 1: 9 (90% for training and 10% for validating). The freshness is graded from 0 to 10.0, with 0 indicating totally rotted (i.e., fuit color and smell are stable which will not be worsen anymore, e,g., the fruit is not eatable and should be thrown away) and 10.0 for complete freshness shown in Table 1. In this article, we define the particular moment when the fruit is harvested as an absolute freshness grade with the number 10.0. However, based on the definition of total corruption, there lacks a definitive degree on this matter. From the fruit decay experiments, we see that fruit freshness grade is not available, decayed fruits may have fungus and produce toxin. We consider the fruits are edible as the primary condition of being recognised.

In this project, we invited ten people to participate in the labelling work. We firstly sampled a few images (i.e., three images for each class of fruits at different decay stages) and required the participants to give their grades. We calculated the mean and standard deviation of the distribution of the proposed freshness grades. Regarding the fruit images with significant grade gaps, e.g., the standard deviation is higher or equals three, we invited them for a second round of grading and narrowed down the disagreement. We kept the labels unchanged if the grades proposed by the participants are close to what we initially have labelled. We justify the labels according to participants' recommendations if the initially proposed freshness grade is far from the mean. It is assumed that for each image, there is a set of images in which the fruits have a similar freshness grade. We grouped the similar images, if the sampled images are required to adjust the freshness levels, the associated images will be set accordingly. Table 1 shows 18 fruit images; the six fruit classes show the decay levels.

2.2 Image Quality Enhancement

Many of the source images have low quality, e.g., blurred or weak exposure to light. Thus, in this article, image enhancement approaches are taken into account to ensure the quality of the images.

Given a 3D image $I(x, y, z)$ with pixel $v(x, y, z)$, there exists a contrast factor $f_{contrast}$ which renders a pixel as same as the average pixel intensity of the whole image when $f_{contrast} = 0$, and keeps the intensity unchanged if $f_{contrast} = 1$. The intensity variation increases while $f_{contrast}$ rises up. The relationship between $f_{contrast}$ and input/output pixel is described as

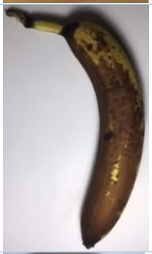
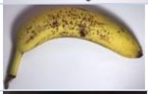
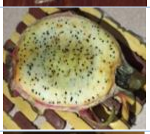
$$v_{x_{new}, y_{new}, z_{new}} = f_{contrast} v_{x,y,z}. \tag{1}$$

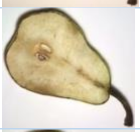
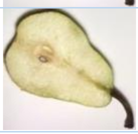
We denote $v_{\min\ i}$ as the minimum value and $v_{\max\ i}$ as the maximum value in the input image, $v_{\min\ o}$ and $v_{\max\ o}$ are the minimum and maximum intensity in the output image respectively, thus, we have

$$v_{x_{new}, y_{new}, z_{new}} = \left(v_{x,y,z} - v_{\min i}\right) \times \left(\frac{v_{\max o} - v_{\min o}}{v_{\max i} - v_{\min i}}\right) + v_{\min o}. \tag{2}$$

where $f_{contrast} = 1.2$ is determined as the result of human perceptions to which degree the contrast-adjusted images are inclusive of necessary visual features whilst being enhanced enough to render granularities that may be easy for neural network training.

Table 1: The means and standard deviations for fruit freshness grading

Fruit Images	Fruits	Means	Standard Deviations
	Apple	1.45	0.35
	Apple	5.30	0.75
	Apple	8.20	0.84
	Banana	2.75	0.96
	Banana	6.00	1.05
	Banana	8.15	0.74
	Dragon fruit	3.40	0.86
	Dragon fruit	5.10	0.89
	Dragon fruit	7.8	0.84
	Kiwi fruit	2.50	1.10

	Kiwi fruit	7.25	0.51
	Kiwi fruit	7.70	1.08
	Orange	3.45	1.12
	Orange	5.30	0.95
	Orange	8.45	0.27
	Pear	2.90	0.83
	Pear	5.35	0.53
	Pear	8.45	0.52

We have our subjective evaluations for the quality of the contrast-based images. In this experiment, we found a myriad of photos are blurry. This is reduced through image sharpening. We see that granular details are more evident than the image before applying to sharpen. It is believed that sharpened images render better visual results. Interpolation and extrapolation are utilized in the sharpening [37]. We thus define a filter for image smoothing

$$kernel_{smooth} = \frac{1}{13}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \tag{3}$$

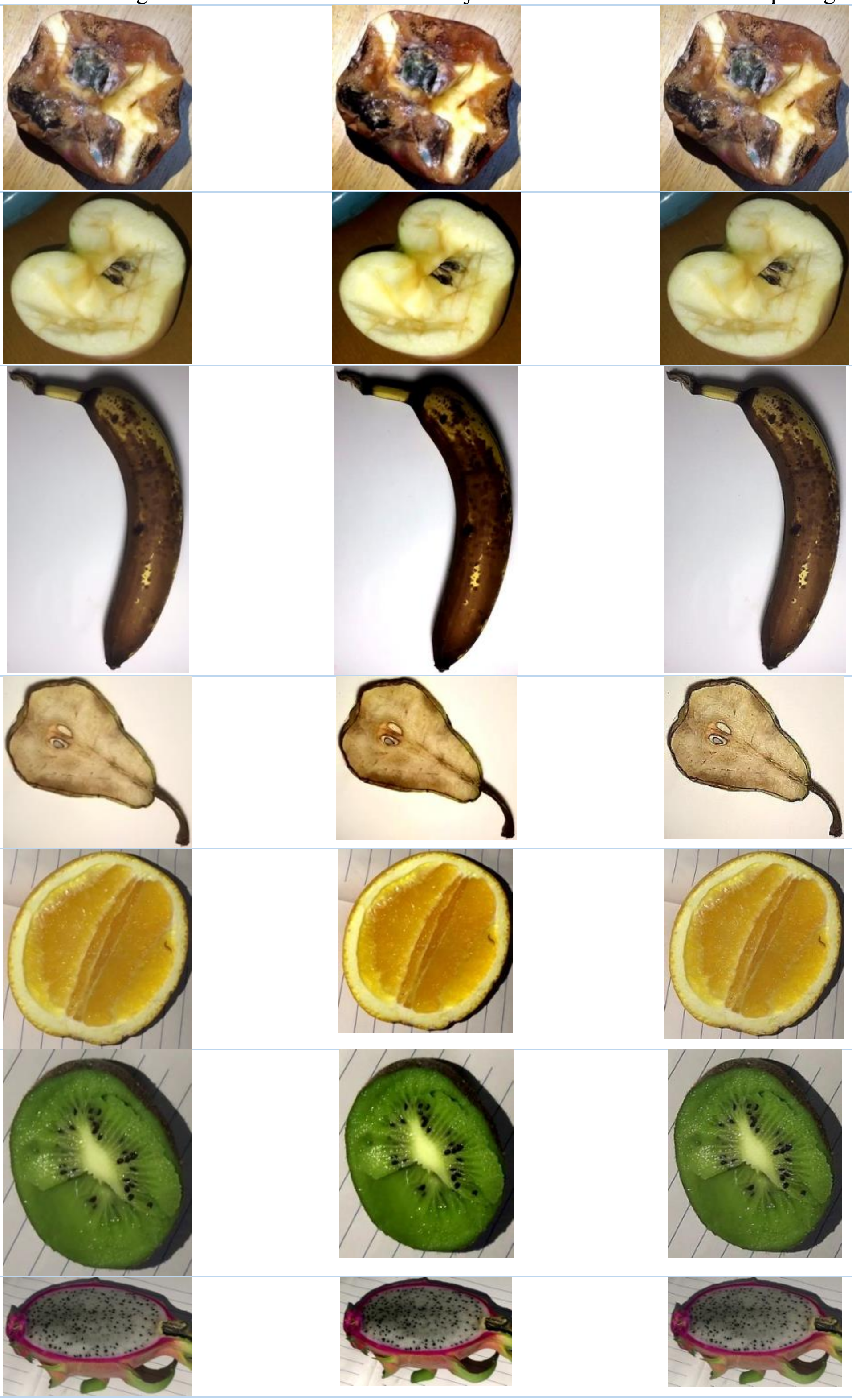
Pertaining to any source image I_{source}, the convolution result I_{smooth} is expressed as

$$I_{smooth} = I_{source} * kernel_{smooth}. \tag{4}$$

where $*$ denotes a convolution operator. Similar to that of the contrast process, we define a sharpness factor $f_{sharpen}$, the derived image I_{blend} is obtained [38].

Table 2: The data augmentation with contrast and sharpening

Source Images	After contrast adjustment	After sharpening



$$I_{blend} = \left(1.0 - f_{sharpen}\right) I_{smooth} + f_{sharpen} I_{source} \quad (5)$$

where $f_{sharpen}$ controls the result I_{smooth} based on the source image I_{source}. In other words, $f_{sharpen} = 0$ renders an image completely blurred under with $kernel_{smooth}$ while $f_{sharpen} = 1.0$ keeps the image unaltered. The interpolation with $f_{sharpen} \in (0, 1)$ has the effects after partially blurring the image I_{source}, the extrapolation with $f_{sharpen} \in (1.0, +\infty)$ inverses smoothing to sharpening. Provided that decrement of $f_{sharpen} \in (0, 1)$ renders increasingly blurry effects, as a result of linear extrapolation, $f_{sharpen} \in (-\infty, 0)$ blurs multifolds of what single $kernel_{smooth}$ rendered.

2.3 Image Augmentation

Image augmentation is a methodology to transform source images into ones with additional information, including scaling, rotating, cropping, and adding random noises. We experimented a rich assortment of augmentation methods as shown in Table 3. Based on our observations, we decided to consider the following augmentation approaches: Rotating and adding random noises. All images are rotated with the angle 120°; we denote an image I as a 2D matrix with coordinates (x, y) for pixel value v,

$$I(x, y) = v_{x,y}. \tag{6}$$

We denote a rotation matrix as R, thus we have

$$R = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}. \tag{7}$$

For any θ degree rotation, we have

$$[x_{new}, y_{new}] = [x, y] \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}, \tag{8}$$

and

$$v_{x_{new}, y_{new}} = v_{x,y} \tag{9}$$

for each new location (x_{new}, y_{new}) having the same pixel intensity. The new image I_{new} is

$$I_{new}(x, y) = I[x_{new}, y_{new}]. \tag{10}$$

The source images are 3D matrices with three colour channels. For an RGB encoded image I_{rgb} with $z = 3$, the rotation matrix is applied to all three dimensions. All images are supplementary with random noises consisting of arbitrary changes of brightness, contrast, saturation, and erosion of ten image regions. The added random noises sequentially follow the order: Random brightness adjustment, random contrast adjustment, and random erosion filtering for the ten image regions. Given a 3D image $I(x, y, z)$ and each pixel $v(x, y, z)$, with a brightness factor $f_{brightness}(v_{add})$, where

$$f_{brightness}(v_{add}) = \frac{v(x,y,z)+v_{add}}{v(x,y,z)}, \tag{11}$$

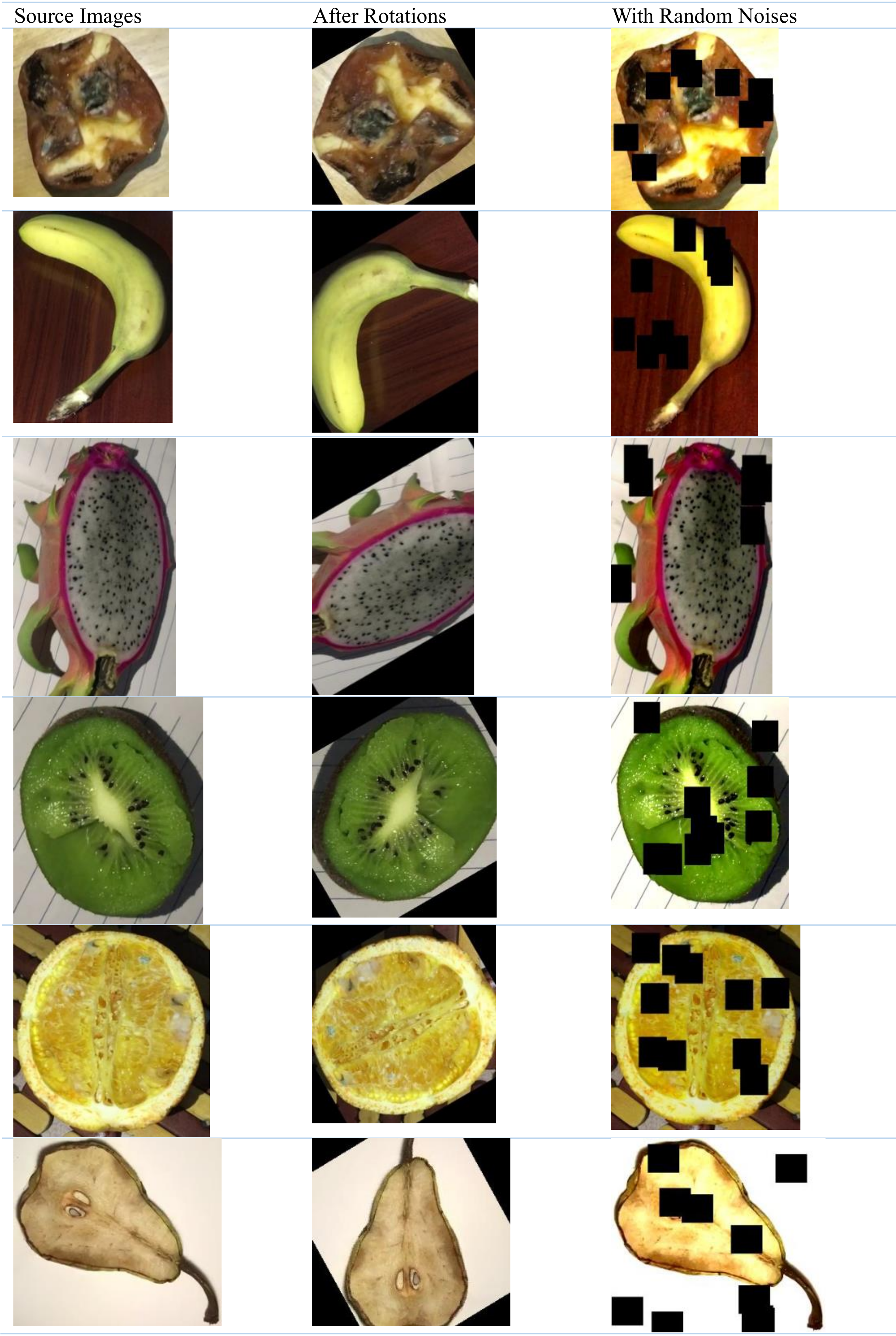
the level of brightness adjustment is proportion to the pixel value, we have $f(v_{add}) \in [0.9, 1.5]$. Thus,

$$v_{x_{new}, y_{new}, z_{new}} = f_{contrast} v_{x,y,z}, \tag{12}$$

where $f_{contrast}$ denotes the level of contrast of an image. In this article, we set $f_{contrast} \in [0.9, 1.5]$ randomly.

Randomly removing image regions [38] is an image augmentation that addresses generalisation issues. Assume an input image I with w_I and h_I for width and height, we define two integers $x_{start} \in [0, w_I]$ and $y_{start} \in [0, h_I]$ as the starting point (x_{start}, y_{start}). We define the width and height of a removal region in proportion r_b. In this article, we set $r_b = 0.15$. The two points, namely, bottom left

Table 3: The examples of image augmentations



(x_{start}, y_{start}) and top right (x_{end}, y_{end}) of the removed region are defined as

$$(x_{start}, y_{start}) = \big(random(0, w_I), random(0, h_I)\big) \tag{13}$$

$$(x_{end}, y_{end}) = (x_{start} + r_b\, w_I,\ y_{start} + r_b\, h_I). \tag{14}$$

The random selection process repeats ten times, the results are shown in Table 3. In summary, the data preprocessing contains six classes of fruits with various decay grades. Data augmentation is extensively emphasised in this article. For each image, there are four variants: Sharpened with contrast, rotated with random noises. There are two classes of labels for fruit objects: Fruit freshness grade and location of a fruit in a given image of VOC [39].

3 Methods

In this section, a neural network YOLO for fruit classification as one hierarchical deep learning model is considered, whose results are fed into a regression CNN for fruit grading. In comparison to the deep learning method, we firstly treat a linear model based on texture and colour of the images; the relevant analysis paves the way for explicating the reason why we should implement a deep learning approach.

3.1 A Linear Proposal

Simple ambient noises refer to the image background with little distractions, usually plain black or white colour. In an environment, fruit localisation and freshness grading become easy, as simple pixel-based manipulation can render satisfactory results. The primary advantage of this project is a fast computation for fruit grading.

In this project, we proposed a simple solution to locate a fruit on a digital image, automatically grade its freshness based on the texture appearance of the fruit itself. Since most of fruits have distinct appearances when the background has a plain or pure colour, a simple threshold can be applied to segment a fruit object from an image. Image regions within the thresholds will be selected while others are masked. The contour of the selected image regions will be depicted to determine the bounding boxes for object detection.

Denote an image as I comprising of pixel $v_{x,y,z}$ where $x \in [1, width]$, $y \in [1, height]$, z is the channel, for example, an RGB image has $z \in [1, 256]$. We have a binary mask

$$mask(x, y, z) = \begin{cases} 1, & v_{x,y,z} \in threshold \\ 0, & otherwise. \end{cases} \quad (15)$$

where the $threshold$ is the pixel intensity of a fruit image. Pertaining to apples, the most observed colours are beige and crimson with RGB colours $(166, 123, 91)$ and $(220, 20, 60)$, respectively. Thus, the colour thresholds are defined as

$$threshold_r = [166 \pm 20, 220 \pm 20] \quad (16)$$

$$threshold_g = [20 \pm 20, 123 \pm 20] \quad (17)$$

$$threshold_b = [60 \pm 20, 91 \pm 20] \quad (18)$$

For freshness grading, we treat the brightness and the pixel intensity within a bounding box as the two conditions. It is believed that generally for a rotten fruit, the number of brown/dark spots grow. This appearance change results in the increases of pixel intensity and the decreases of brightness. An entropy for a given image I with histograms h_i is

$$entropy(I) = -\sum_i (h_i * \log(h_i)). \quad (19)$$

For a given image I with a pixel p_i, where $i = 0,1, \dots, n$, n represents the number of pixels that comprise the image, we have

$$brightness(I) = \frac{1}{n}\sum_i (p_i). \quad (20)$$

The freshness is calculated by using eq. (21)

$$freshness = k_e \; entropy(I) + k_b \; brightness(I) + b, \tag{21}$$

where k_e and k_b are weight adjustment parameters, b is the bias. These parameters are determined via linear regression, assuming a regression output y_i and a data sample x_i , where x_i consists of n features/ dimensions, thus,

$$\hat{y}_i = \beta_0 + \beta_1 \, x_{i,1} + \beta_2 \, x_{i,2} + \cdots + \beta_n \, x_{i,n}. \tag{22}$$

The loss function for linear regression is

$$Loss = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \, . \tag{23}$$

We minimise the loss,

$$\hat{\vec{\beta}} = \arg_{\hat{\beta}} min \;\, Loss(X, \vec{\beta}). \tag{24}$$

Therefore, we have $\hat{\vec{\beta}} = \{b, -k_e, k_b\}$. We selected a few fruit images with various decay levels and calculated the entropy as well as the brightness of the detected bounding box, meanwhile k_e and k_b are determined. We observed the entropy and intensity, adjusted k_e and k_b to make the sum of the entropy and brightness intensity close to the corresponding freshness level.

3.2 A Hierarchical Deep Learning Model

In this section, we propose a hierarchical deep learning model for fruit fresheness classification, whose results are fed into a second one (regression CNN) for freshness grading. YOLO + Regression CNN is a hierarchical neural network, whose predictive bounding boxes are fed to the regression CNN for freshness grading. Regression CNNs are trained for each class of fruits. In this article, we work for the classification of six classes of fruits; the six regression CNNs are trained. YOLO is used to classify the class of the object/fruits as well as estimate the bounding box, which locates the visual object on an image. The corresponding regression has been applied to this class of fruits for freshnessgrading. The framework is illustrated in Fig. 1, the regression model is shown in Fig. 2, the pipeline of this model is shown in Fig. 3.

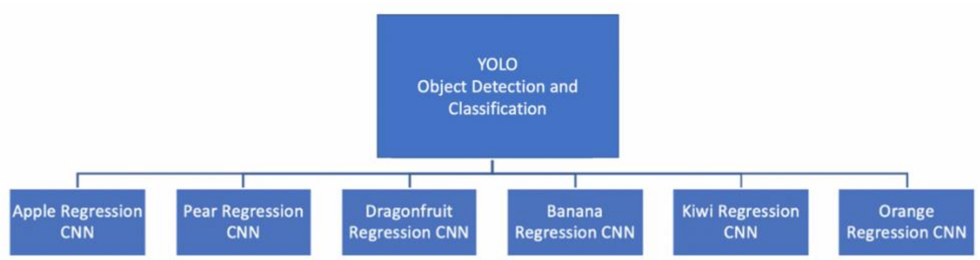


Fig. 1: YOLO + Regression CNN model

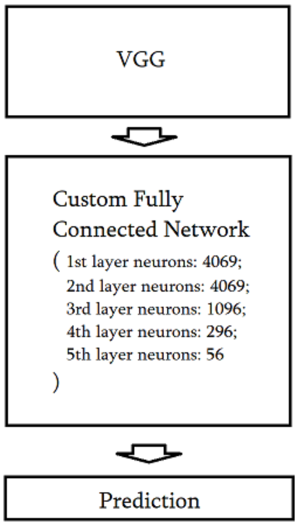


Fig. 2: The customized regression model (VGG)

The source images are fed into YOLO for object recognition, where the central point, width and height of the bounding box are determined. With YOLO prediction, the model maps the predicted class of the detected fruit onto the regression neural network. The detected object in the image is cropped out from its background as the input image to the regression CNN network.

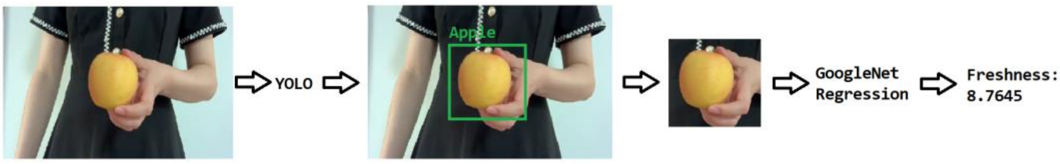


Fig. 3: The pipeline of the proposed hierarchical deep learning model

The YOLO model in this article has the same structure as YOLOv3 [40]. We thus define a set of input data D, in which we have

$$D = \{I_1, I_2, \dots, I_n\} \tag{25}$$

where n is the total number of input images, I_i is the i-th image, $i = 1,2,3 \dots, n$. Our input images are encoded using RGB channels. Thus, this defines each image I_i as three-dimensional and has the same image size. The image I_i is defined as a 2D matrix

$$I_i = \begin{bmatrix} p_{1,1}, p_{1,2} & \cdots & p_{1,w} \\ \vdots & \ddots & \vdots \\ p_{h,1}, p_{h,2} & \cdots & p_{h,w} \end{bmatrix}. \tag{26}$$

In order to prevent overfitting, additional random flips are considered, after YOLOv3 takes the source data and starts the computation [40] at the time t, we have a prediction $\hat{Y} = \{\widehat{y_1}, \widehat{y_2}, \dots, \widehat{y_n}\}$

$$\hat{Y}(t) = P(\xi_{YOLO}(t)|D) \tag{27}$$

For the YOLO model ξ_{YOLO} we have a bounding box, the associated object class, and the prediction

$$\widehat{y_i} = \{\widehat{x_i}, \widehat{y_i}, \widehat{w_i}, \widehat{h_i}, \widehat{c_i}\}. \tag{28}$$

According to the predicted class $\widehat{c_i}$, the anchored box is denoted as $(\widehat{x_i}, \widehat{y_i}, \widehat{w_i}, \widehat{h_i})$, the source image I_i is cropped. The new image is

$$newI_i = crop(I_i, \widehat{x_i}, \widehat{y_i}, \widehat{w_i}, \widehat{h_i}) \tag{29}$$

where $\widehat{x_i}$ and $\widehat{y_i}$ are the central point of the predicted bounding box, the $crop(I_i, \widehat{x_i}, \widehat{y_i}, \widehat{w_i}, \widehat{h_i})$ for the i-th image I_i is expressed as

$$crop(I_i, \widehat{x_i}, \widehat{y_i}, \widehat{w_i}, \widehat{h_i}) = \begin{bmatrix} p_{\widehat{x_i}-\frac{\widehat{w_i}}{2}, \widehat{y_i}-\frac{\widehat{h_i}}{2}} & \cdots & p_{\widehat{x_i}+\frac{\widehat{w_i}}{2}, \widehat{y_i}-\frac{\widehat{h_i}}{2}} \\ \vdots & \ddots & \vdots \\ p_{\widehat{x_i}-\frac{\widehat{w_i}}{2}, \widehat{y_i}+\frac{\widehat{h_i}}{2}} & \cdots & p_{\widehat{x_i}+\frac{\widehat{w_i}}{2}, \widehat{y_i}+\frac{\widehat{h_i}}{2}} \end{bmatrix} \tag{30}$$

The cropped image $newI_i$ is fed into a regression neural network. We thus define the regression neural network $\xi_{rege}(t)$ at the training epoch t, a cropped image dataset is $newD = \{newI_1, newI_2, \ldots, newI_m\}$,

$$\hat{R} = P\left(\xi_{regr}, \widehat{c_i} \middle| newD\right) \tag{31}$$

where $\hat{R}$ is the result of fruit freshness regression, we have

$$\hat{R} = \{\mathrm{r}_1, \mathrm{r}_2, \ldots, \mathrm{r}_\mathrm{n}\}. \tag{32}$$

Hence, the hierarchical model is expressed as

$$\hat{Y}(t) = P\left(\xi_{YOLO}(t), \xi_{regr}(t) \middle| D\right) \tag{33}$$

For each prediction $\widehat{y_i}$, we have

$$\widehat{y_i} = \left\{\widehat{x_i}, \widehat{y_i}, \widehat{w_i}, \widehat{h_i}, \widehat{c_i}, \widehat{r_i}\right\}. \tag{34}$$

In this article, we experimented on four base networks, i.e., AlexNet, VGG, ResNet, and GoogleNet for regression based on the six classes of fruits. Each class of fruits likely has unique features distinct from others; the extracted features should be processed by using regression convolutional neural network.

We offer the base networks AlexNet, VGG, ResNet, GoogLeNet for feature extraction. In the fully-connected layers, we modified the number of neurons to fit for our fruit freshness regression. We designed additional four-layer for the fully connected neural network.

4 Experimental Results

In this article, our model is constructed hierarchically, consisting of classification and location-oriented model (YOLO), a set of regression CNNs targeting each fruit type. Besides, we make a comparison to the linear model.

4.1 The Linear Proposal

We calculate the average brightness and entropy for a video frame. Associated the images with complicated background noises, the location is very hard to be found, the brightness/entropy approach does not converge as expected. The defined freshness function is shown in eq.(35),

$$\hat{y} = k_1 J + k_2 B + b, \tag{35}$$

where it has not a linear relationship with entropy and brightness of the image. The configurations of a linear regressor are shown as:

- k_1: -2.7701
- k_2: 0.00367
- b: 9.0004

However, this approach is subject to background noises, even if a minor change of background might result in significant errors. During this experiment, we set up different physical backgrounds while acquiring an image of the fruits, including lighting conditions and placing adjacent foreign objects.

For fresh fruits such as apples and banana, we do observe that there exist correlations between entropy/ brightness levels and decay stages if the background is set as static. However, for other fruits such as kiwi fruits and oranges, this assumption is hardly correct.

This preliminary approach through entropy/brightness computations reveals the complexity of fruit freshness grading. Fruits have their own processes of decaying, for each decay characteristic, there is no apparent relationship between static visual features (i.e., a set of defined rules of pixel statistics) and freshness levels. Based on these discoveries, we decide to treat each class of fruit individually rather than a comprehensive approach.

4.2 YOLO + GoogLeNet

In the GoogLeNet, the fruits show multiple levels of regression on grading fruit freshness. Banana is the most accurately predicted class of fruits, while Kiwi fruits are the most difficult one. Apple freshness grading appears the most unstable one in the validation set. This is able to be traced back to the festering features of apples that apples have rich features whilst decaying. For example, they appeared with fungus and brown/dark spots, in comparison to other fruits with relatively universal rottenness features, e.g., dragon fruits usually are covered by yellowish dark spots.

Table 4: The metrics for evaluating the performance of GoogLeNet

Fruits	**MSE**		**Standard Deviation**	
	Training	Validation	Training	Validation
Apple	4.499	4.653	2.082	2.722
Dragon Fruit	2.629	2.926	1.065	1.725
Kiwi fruit	5.810	5.997	1.172	1.430
Pear	4.250	5.958	2.045	1.705
Banana	1.661	1.705	0.967	0.964
Orange	2.905	3.005	0.606	0.451
Average	3.625	4.404	1.323	1.500

4.3 YOLO + AlexNet

It is observable that the performance of AlexNet on the six classes of fruits is similar to other base network regarding on which class of fruits the regression is prone to deviating from the ground truth. Apples, Kiwi fruits, and pears are the three most challenging classes to be regressed, while bananas are the most accurate one. Fruits with relatively large errors tend to be less stable in standard deviation during regression. This is evident in both training and validation sets of all classes of fruits. The average MSE for the six classes of fruits is 3.500 for the training dataset and 4.099 for the validation dataset. In terms of regressions, this model generates 1.480 for the training set and 1.248 for the validation set.

Table 5: The metrics for evaluating the performance of AlexNet

Fruits	MSE		Standard Deviation	
	Training	Validation	Training	Validation
Apple	4.974	4.987	1.687	1.497
Dragon fruit	2.658	2.794	1.247	1.686
Kiwi fruit	4.279	5.664	2.422	0.893
Pear	4.250	5.958	2.045	1.705
Banana	1.696	1.818	0.793	0.892
Orange	3.139	3.368	0.687	0.816
Average	3.500	4.099	1.480	1.248

Table 6: The metrics for measuring the performance of VGG-11

Measurement	MSE		Standard Deviation	
	Training	Validation	Training	Validation
Apple	4.504	4.625	2.038	2.078
Dragonfruit	2.823	3.129	1.374	1.129
Kiwi	5.726	5.670	1.546	1.101
Pear	4.226	5.341	1.717	1.712
Banana	1.796	1.831	0.844	0.607
Orange	2.900	3.012	0.647	0.967
Average	3.665	3.934	1.361	1.266

Table 7: The metrics for evaluating the performance of ResNet

Measurement	MSE		Standard Deviation	
	Training	Validation	Training	Validation
Apple	4.226	4.374	2.029	2.188
Dragonfruit	2.634	2.815	0.913	0.840
Kiwi	6.034	5.765	1.467	1.417
Pear	3.984	6.507	1.936	4.899
Banana	1.636	1.659	0.984	0.864
Orange	2.982	3.233	0.645	0.847
Average	3.582	4.058	1.329	1.842

4.4 YOLO + ResNet

ResNet-152 is the top net among the ResNet family as well as the deepest one among the ResNets. Again, ResNet fails to deliver reliable results for three particular classes of fruits: Apples, Kiwi fruits, and pears. The regression error is largely based on the Kiwi fruit dataset, both on the training and validation sets. For pears, there exists a possibility of overfitting by using the validation set shows 6.057 while it reports 3.984 by using the training set. Banana freshness grading is the most accurate one. In terms of regression stability, pears are the least stable while oranges are generally the highest one, judged by using training and validation sets. On average, the MSE values of training and validation sets for ResNet-152 are 3.582 and 4.058, respectively. For stability measurement, the standard deviation is 1.329 for the training set and 1.842 for the validation set.

4.5 YOLO + VGG

We tested the VGG-11 model. Again, grading bananas is the most accurate one in freshness grading regression, whilst classifying apples, kiwis fruits, and pears are the most difficult ones. However, VGG-11 tends to suffer less from overfitting as indicated in the metrics where the result gaps between the training and validation sets are small against what we have observed in other base networks. VGG-11 displays high stability in regression, where for apples, Kiwi fruitsm and pears, both training and validation sets show robust regression output (hinted in standard deviation). The average training and validation MSE values are close to the other three base networks, 3.665 and 3.934, respectively. The standard deviations are 1.361 and 1.266, respectively.

4.6 Comparisons

The four deep learning models have similar performance. By using the training set, AlexNet shows the best of MSE while VGG eyes the lowest error with the validation set. Table 8 is a summary of the overall proposed model regression performance, measured in MSE.

Table 8: A summary of performance of the proposed schemes in MSE

	Training Set	**Validation Set**
Linear Regressor	4.749	5.128
AlexNet	3.500	4.099
GoogleNet	3.625	4.404
VGG	3.665	3.934
ResNet	3.582	4.058

5 Conclusion

In this paper, we constructed a linear regression model to detect and measure the fruit freshness features by judging the darkness of the fruit skin and variations of colour transitions. Accordingly, we affirm that fruit spoilage occurs with biochemical reactions that result in visual feature fading. Hence, we propounded a deep learning solution.

Deep learning has been used for fruit freshness grading, with the considerations of multiclass of fruits (i.e., apple, banana, dragon fruit, Kiwi fruit, orange, and pear). We have developed a hierarchical approach, in which a slew of fruits are detected and classified with real-time object detection, the regions of interest are cropped from the source images and fed into CNN models for regression, thus the freshness level is finally graded. We independently trained the convolutional neural network for four renown models, i.e., GoogLeNet, ResNet, AlexNet, VGG-11. Our experimental results have shown an excellent performance of deep learning algorithms towards to resolving this problem [9, 42].

References

[1] Akinmusire, O.: Fungal species associated with the spoilage of some edible fruits in Maiduguri Northern Eastern Nigeria. *Advances in Environmental Biology*, 157-162 (2011)

[2] Rawat, S.: Food Spoilage: Microorganisms and their prevention. *Asian Journal of Plant Science and Research*, *5*(4), 47-56 (2015)

[3] Tournas, V. H., & Katsoudas, E.: Mould and yeast flora in fresh berries, grapes and citrus fruits. *International Journal of Food Microbiology*, *105*(1), 11-17 (2005)

[4] Sindhi, K., Pandya, J., & Vegad, S.: Quality evaluation of apple fruit: A survey. *International Journal of Computer Applications*, *975*, 8887 (2016)

[5] Shukla, A. K.: *Electron Spin Resonance in Food Science*. Academic Press (2016)

[6] Bhargava, A., & Bansal, A.: Fruits and vegetables quality evaluation using computer vision: A review. *Journal of King Saud University-Computer and Information Sciences* (2018)

[7] Pandey, R., Naik, S., & Marfatia, R.: Image processing and machine learning for automated fruit grading system: A technical review. *International Journal of Computer Applications*, *81*(16), 29-39 (2013)

[8] Cunha, J. B.: Application of image processing techniques in the characterisation of plant leafs. In *IEEE International Symposium on Industrial Electronics,* 612-616 (2003)

[9] Fu, Y.: Fruit Freshness Grading Using Deep Learning (Master's Thesis), Auckland University of Technology, New Zealand (2020)

[10] Koumpouros, Y., Mahaman, B. D., Maliappis, M., Passam, H. C., Sideridis, A. B., & Zorkadis, V.: Image processing for distance diagnosis in pest management. *Computers and Electronics in Agriculture*, *44*(2), 121-131 (2004)

[11] Brosnan, T., & Sun, D. W.: Inspection and grading of agricultural and food products by computer vision systems—A review. *Computers and Electronics in Agriculture*, *36*(2-3), 193-213 (2002)

[12] Tewari, V. K., Arudra, A. K., Kumar, S. P., Pandey, V., & Chandel, N. S.: Estimation of plant nitrogen content using digital image processing. *Agricultural Engineering International: CIGR Journal*, *15*(2), 78-86 (2013)

[13] Ozyildiz, E., Krahnstöver, N., & Sharma, R.: Adaptive texture and color segmentation for tracking moving objects. *Pattern Recognition*, *35*(10), 2013-2029 (2002)

[14] Tripathi, M. K., & Maktedar, D. D.: A role of computer vision in fruits and vegetables among various horticulture products of agriculture fields: A survey. *Information Processing in Agriculture* (2019)

[15] Simonyan, K., & Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556* (2014)

[16] Krizhevsky, A., Sutskever, I., & Hinton, G. E.: ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems,* 1097-1105 (2012)

[17] He, K., Zhang, X., Ren, S., & Sun, J.: Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition,* 770-778 (2016)

[18] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., .& Rabinovich, A.: Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition,* 1-9 (2015)

[19] Tripathi, M. K., & Maktedar, D. D.: A role of computer vision in fruits and vegetables among various horticulture products of agriculture fields: A survey. *Information Processing in Agriculture* (2019)

[20] Moallem, P., Serajoddin, A., & Pourghassem, H.: Computer vision-based apple grading for golden delicious apples based on surface features. *Information Processing in Agriculture*, *4*(1), 33-40 (2017)

[21] Arakeri, M. P.: Computer vision based fruit grading system for quality evaluation of tomato in agriculture industry. *Procedia Computer Science*, *79*, 426-433 (2016)

[22] Zawbaa, H. M., Abbass, M., Hazman, M., & Hassenian, A. E.: Automatic fruit image recognition system based on shape and colour features. In *International Conference on Advanced Machine Learning Technologies and Applications*, 278-290 (2014)

[23] Iqbal, Z., Khan, M. A., Sharif, M., Shah, J. H., Rehman, M. H., & Javed, K.: An automated detection and classification of citrus plant diseases using image processing techniques: A

review. *Computers and Electronics in Agriculture*, *153*, 12-32 (2018)

[24] Bresilla, K., Perulli, G. D., Boini, A., Morandi, B., Grappadelli, L. C., & Manfrini, L.: Single-shot convolution neural networks for real-time fruit detection within the tree. *Frontiers in Plant Science*, *10* (2019)

[25] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A.: You Only Look Once: Unified, real-time object detection. In *IEEE Conference on Computer Vision and Pattern Recognition,* 779-788 (2016)

[26] Zeng, G.: Fruit and vegetables classification system using image saliency and convolutional neural network. In *IEEE Information Technology and Mechatronics Engineering Conference (ITOEC),* 613-617 (2017)

[27] Mureşan, H., & Oltean, M.: Fruit recognition from images using deep learning. *Acta Universitatis Sapientiae, Informatica*, *10*(1), 26-42 (2018)

[28] Nashat, A. A., & Hassan, N. H.: Automatic segmentation and classification of olive fruits batches based on discrete wavelet transform and visual perceptual texture features. *International Journal of Wavelets, Multiresolution and Information Processing*, *16*(01), 1850003 (2018)

[29] Al-Sarayreha, M., Reis, M., Yan, W., Klette, R.: Potential of deep learning and snapshot hyperspectral imaging for classification of species in meat. Food Science. DOI: 10.1016/j.foodcont.2020.107332 (2020)

[30] Gómez-Sanchis, J., Gómez-Chova, L., Aleixos, N., Camps-Valls, G., Montesinos-Herrero, C., Moltó, E., & Blasco, J.: Hyperspectral system for early detection of rottenness caused by Penicillium digitatum in Mandarins. *Journal of Food Engineering*, *89*(1), 80-86 (2008)

[31] Knee, M.: Anthocyanin, carotenoid, and chlorophyll changes in the peel of Cox's Orange Pippin apples during ripening on and off the tree. *Journal of Experimental Botany*, 184-196 (1972)

[32] Davey, M. W., Stals, E., Ngoh-Newilah, G., Tomekpe, K., Lusty, C., Markham, R., & Keulemans, J.: Sampling strategies and variability in fruit pulp micronutrient contents of west and central African bananas and plantains (Musa species). *Journal of Agricultural and Food Chemistry*, *55*(7), 2633-2644 (2007)

[33] Bampidis, V. A., & Robinson, P. H.: Citrus by-products as ruminant feeds: A review. *Animal Feed Science and Technology*, *128*(3-4), 175-217 (2006)

[34] Zheng, Y., Yu, C., Cheng, Y. S., Zhang, R., Jenkins, B., & VanderGheynst, J. S.: Effects of ensilage on storage and enzymatic degradability of sugar beet pulp. *Bioresource Technology*, *102*(2), 1489-1495 (2011)

[35] Herbach, K. M., Stintzing, F. C., & Carle, R.: Betalain stability and degradation—structural and chromatic aspects. *Journal of Food Science*, *71*(4), R41-R50 (2006)

[36] Schwartz, S. J., & Von Elbe, J. H.: Kinetics of chlorophyll degradation to pyropheophytin in vegetables. *Journal of Food Science*, *48*(4), 1303-1306 (1983)

[37] Haeberli, P., & Voorhies, D.: Image processing by linear interpolation and extrapolation. *IRIS*

Universe Magazine, *28*, 8-9 (1994)

[38] Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y.: Random erasing data augmentation. *arXiv:1708.04896* (2017)

[39] Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A.: The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, *88*(2), 303-338 (2010)

[40] Redmon, J., & Farhadi, A.: YOLOv3: An incremental improvement. *arXiv:1804.02767*. (2018)

[41] Karakaya, D., Ulucan, O., Turkan, M.: A comparative analysis on fruit freshness classification, *Innovations in Intelligent Systems and Applications Conference* (2019)

[42] Yan, W.: Computational Methods for Deep Learning: Theoretic, Practice and Applications, Springer (2021)