# Identification of fresh fruits using deep learning

Hong Zhou, Minh Nguyen, Wei Qi Yan
Auckland University of Technology, Auckland 1010 New Zealand

**Abstract.** In view of this consideration, advanced YOLOv8 was employed as an identification algorithm by using deep learning to fruit identification. YOLOv8 algorithm is the improved version of YOLO (You Only Look Once) on object detection area. the large-scale datasets training and real-time objects detection seem to acquire more accuracy by YOLOv8. This multimedia visual recognition technology based on deep learning has the possibility of surpassing previous research in fresh fruit identification. This use case can use the anchor box in dynamic video and static images to locate the fruit and display relevant information. After Targeted dataset training, model adjustment and optimization, the detection accuracy may be improved, and the detection time may also be effectively reduced; Thus, the advantage makes YOLOv8 a popular choice for fruit identification tasks in research and industry. We scan the fruits through the camera of the smartphone, and fruits on the phone screen will be framed and automatically displayed the name and sweetness with labels.

**Keywords:** Neural Network, fruit identification, YOLOv8, dataset, deep learning, accuracy

---

Hong Zhou
zdq4348@autuni.ac.nz

Minh Nguyen
minh.nguyen@aut.ac.nz

Wei Qi Yan
weiqi.yan@aut.ac.nz

## 1 Introduction

Various types of fruits such as mango, kiwifruit, and papaya are available in the market, but not all varieties may be available in all regions. Fruit recognition, which relies on AI technology, can assist consumers in determining the sweetness or sourness of fruits. In this essay, the YOLOv8 model was utilized to develop a deep learning system for fruit identification, using over 3,000 video images and pictures collected from the Countdown and New World supermarkets. The goal was to improve the accuracy of fruit identification. The model was trained using images captured from the rear camera of a standard smartphone to enhance practicality. The experiment results will be presented, and this experience can support future research in this area.

## 2 Literature Review

The benefits of utilizing deep learning in machine learning are quite evident, particularly in the domain of computer vision. Doulamis and his team [1] conducted research to assess the advantages and constraints of deep learning, and also discussed the future directions of computer vision design based on various practical applications.

The human and animal brain can process and understand diverse types of information, enabling the recognition of complex structures in large-scale data. Deep learning emulates this mechanism by establishing numerous data abstraction and computational layers. Unsupervised and supervised feature learning algorithms, hierarchical probability models, and neural networks are all examples of deep learning. When confronted with a large volume of complex data, deep learning has been shown to outperform previous technologies. The development of neural networks was spurred by McCulloch and Pitts' desire [2] to create an artificial brain, with the MCP model serving as the earliest neuron model. LeNet [3] and Long Short-Term Memory [4] have also made significant contributions to the field. However, the true era of deep learning began in 2006, when Hinton and colleagues [5] made a major breakthrough with a Deep Belief Network that employed multiple layers of Restricted Boltzmann Machines. This structure can facilitate layer-by-layer local training and learning without supervision, which is why deep learning frameworks and algorithms have gained popularity in recent decades.

Open high-quality large datasets and GPUs with high computational capabilities have greatly improved model training, thereby promoting the development of network and machine learning. Other factors that have contributed to this progress include addressing issues such as gradient disappearance that are caused by out-of-saturation activation functions, and the emergence of more powerful frameworks such as discarding, batch normalization, and data augmentation, as well as new regularization technologies like Mxnet, TensorFlow, and Theano [6]. Deep learning has made significant strides in addressing visual problems such as semantic segmentation [7], [8], human motion tracking [9] [10], human action recognition [11], [12], visual object detection [13], [14], and human pose estimation [15] [16]. The three most typical deep learning frameworks in this context are Stacked (Denoising) Autoencoders, Deep Belief Networks (DBNs), and Convolutional Neural Networks (CNNs).

In 2016, YOLO was created by Joseph et al [17]. YOLOv8 means the eighth-version of YOLO models, not appear suddenly, but evolve from an earlier version of YOLOv5 by ultralytics, the initial group which created YOLO. The improvement from YOLOv5 to YOLOv8 includes the Backbone and head structure, anchor and training strategy of the epoch. In the backbone part, The C2f structure with richer gradient flow in YOLOv8 is selected to replace the C3 structure in YOLOv5. In order to reduce the number of blocks of the largest stage in the backbone network, the models with different scaling factors N/S/M/L/X no longer share a set of model parameters. The M/L/X large model also reduces the number of output channels of the last stage, further reducing the number of parameters and calculations. Meanwhile, anchor-based was an alternative by Anchor-Free, TAL (Task Alignment Learning) dynamic matching adopted, and DFL (Distribution Focal Loss) and CIoU Loss are utilized as the loss function of the regression branch, which makes classification tasks correspond to regression tasks.

According to previous studies, the sweetness of the fruit is able to be indicated by the Brix value, which can be used by the infrared detector to determine which leads to non-destructive fruits during identification [18]. The change in fruit skin colour during ripening does not completely represent the change in fruit acidity and sweetness. For instance, the intensification of surface colouration was related to the initial colour intensity of the mangoes. Fruits with a more intensive initial red or yellow colouration ripened more rapidly than mangoes with less intensive colouration. Fully ripened mangoes which had initially a more intensive coloration had developed a more desirable internal colouration and a higher content of T.S.S. It was said that both, the maximum red and maximum yellow colour intensities at harvest, could serve as a good index of mango maturity.

## 3 Our Methods

This article employs the YOLOv8n and YOLOv8s models to train a customized fruit dataset, avoiding the use of existing datasets which could lead to overfitting during training. The YOLOv8 model adopts a CNN (Convolutional Neural Network) structure for extracting image features through convolution.

To implement an image recognition algorithm using neural networks, powerful hardware support is required. To address this limitation, we utilized Google Colab which offers a standard GPU and VM to run YOLOv8 project. Additionally, we connected Google Drive as a notebook to edit the code file. YOLOv8 is highly suitable for fruit detection as it can identify various commonly detected objects. Compared to larger and more complex models, the lightweight YOLOv8n and YOLOv8s have default parameters that enable faster training and loading speed.

### 3.1 YOLOv8 Model

YOLOv8 is developed based on convolution neural network, which can detect objects in images and videos. It is a derivative of a neural network similar to human and animal brains that is considered as an artificial neural network.

Fig 1 demonstrates the confidence level of the framed object by using Yolov8n and YOLOv8s model. The result of image validation with the maximum confidence level frame will be displayed after the last epoch with the iteration in the training process consisted of back-propagation and automatic weight adjustment. The Mosaic enhancement was closed in the last 10 epoch that is the salient features of YOLOv8.


**Fig 1.** The confidence of YOLOv8 results

The structure of YOLOv8 is approximate to YOLOv5 in the main structure comprise of three parts, the backbone, neck and head. The convolution structure in the upper sampling stage has been deleted in YOLOv8. Decoupled-Head was added with *reg_max* as the number of channels of regression head. The CSP structure in the backbone is still retained by YOLOv8. The original image is convolved and sliced to form feature maps, which contain the feature vectors of the fruit image for final training. The FPN structure uses up-sampling to fuse feature information from higher layers and obtain a feature map for prediction. In the new structure, strong positioning features are transmitted upward through PAN despite C3 modules being replaced by C2f. The combination of FPN and PAN allows for parameter aggregation of different detection layers from various backbone layers, resulting in a larger receptive field and more abstract extracted features. This structure facilitates the combination of high-level semantic features with detailed object classification information, thus improving feature extraction ability.

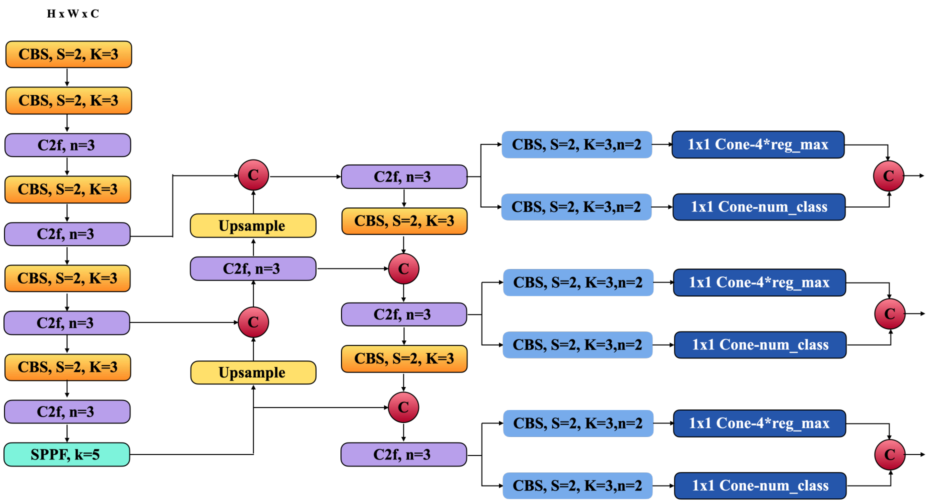


**Fig 2.** The structure of YOLOv8 model

In YOLOv8, the C2f module was employed to replace the C3 module, realizing further lightweight; Meanwhile, YOLOv8 still retains the SPPF module used in YOLOv5 and other architectures. Fig. 1 shows the structure of the C3 module and C2f module. The C2f module is designed with reference to the idea of the C3 module and ELAN lead to more abundant gradient flow information while ensuring lightweight for YOLOv8. Fig. 2 and Fig. 3 depict the structure of the C3 module and C2f module.

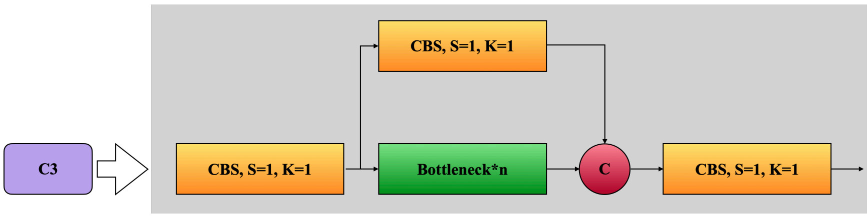


**Fig 3.** The structure of C3 module

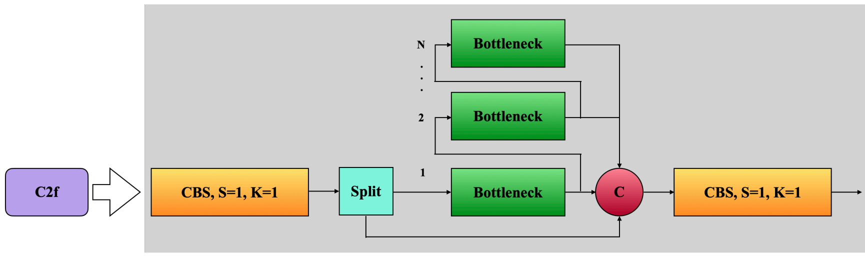


**Fig 4.** The structure of C2f module

### 3.2 Loss Function

We intuitively see from the loss function whether the operation results meet the expectations including *cls_loss* (classification loss), *box_loss* (location loss) and *obj_loss* (confidence loss). YOLOv8 takes use of VFL (Varifocal Loss) function to calculate the loss of classification and confidence loss of the object and uses *DFL* (Distribution Focal Loss) + *CIOU* as the loss of bounding box regression. The class imbalance issue [19] can be addressed, and the binary cross entropy loss is the predecessor of VFL. In eq.(1), $p$ and $q$ predict IACS and target score respectively. The ground truth class of $q$ in the foreground point is set as IoU between ground truth and the generated bounding box or else 0. Contrast with foreground point, for all classes, target q in the background point is 0.

$$VFL(p,q) = \begin{cases} -q(q\log(p) + (1-q)\log(1-p)) & q > 0 \\ -\alpha p^{\gamma}\log(1-p) & q = 0 \end{cases} \tag{1}$$

In the condition of arbitrary and flexible distribution, around the continuous locations of target bounding boxes, DFL makes the network focus on learning the probabilities of values rapidly. This approach was proposed [20] to tackle the issue that the original FL (Focal Loss) only supports discrete {1, 0} category labels currently.

### 3.3 CIoU

The CIoU loss function incorporates the scale, center distance, and aspect ratio of the bounding box prediction with the intersection over union (IoU) measure. In equation (3), the function $\upsilon$ evaluates the consistency of the aspect ratio, while in eq. (4), $\alpha$ balances the contribution of the various terms in the loss function. The distance between the predicted and real bounding box centers is represented by $d$ in eq. (5), and c represents the diagonal distance of the minimum bounding box containing the two boxes. *L_CIoU* is the final loss value used for bounding box regression.

$$\mathcal{R}_{CIoU} = \frac{\rho^2(b, b^{gt})}{c^2} + \alpha\upsilon \tag{2}$$

$$\upsilon = \frac{4}{\pi^2}\left(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}\right)^2 \tag{3}$$

$$\alpha = \frac{\upsilon}{(1 - IoU) + \upsilon} \tag{4}$$

$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{d^2}{c^2} + \alpha\upsilon \tag{5}$$

### 3.4 Training Process

Once Google Colab was employed to train, validate, and predict the fruit datasets should install the ultralytics package of YOLOv8 after disconnecting the runtime. The graphical result will be not generated until training is completed. The graphical results will not be saved unless the 'save’ parameter is set as true. In this experiment, data-enhanced images were segmented into train, validation and test datasets by a split function based on python; Moreover, the XML format files of the images were labelled manually, and subsequently, convert to Yolo datasets format. 23 varieties of fruit identification run as YOLOv8n and YOLOv8s with a list of sweetness, which was marked as Brix value, and Brix value of major fruit are cited from previous studies [21-40] in Table 1.

**Table 1.** Comparison table of fruit name and brix value

| Name | Brix Value | Name | Brix Value | Name | Brix Value |
|---|---|---|---|---|---|
| Strawberry | 7.125° | Rockmelon | 10° | Cherry sweetheart | 13.99° |
| Raspberry | 10° | Yellow peach | 13° | Pear packham | 10~11° |
| Papaya philippine | 10.4° | Mandarin | 10.6~12.9° | Apricot | 7.45~17° |
| Melon honeydew | 10.9° | Orange suntreat | 13.8~16.5° | Red grape | 15~20.6° |
| Pineapple | 16.6° | Lemon | 9° | Apple granny smith | 12.5° |
| Kiwi Italian | 7.37~11° | Plum | 12~32° | | |

# 4 Our Results

## 4.1 Our Datasets

The images of the training dataset were captured as high-quality mobile video from the countdown and the new market supermarkets in New Zealand. The video is processed as a slice every 50 frames to obtain the original images of the dataset that were 1920*1080 resolution. In this era of highly developed mobile Internet and smartphones, the resolution of these images is more consistent with the scene when the program is installed on mainstream smartphones to identify the fruit.

Fig 5 illustrates some sample images in datasets. The same category of fruit was labelled as a name of folder by using real name, variety and referable Brix value, and the images of 23 different fruit were collated and stored in this directory such as “apple_granny_smith@sweetness_under12.5°Brix” and “yellow_peach@sweetness_around13°Brix”. These images are completely random including highly similar images; However, there are no duplicate images. A single variety of fruit appears in every image and the number may not be one with a blocking relationship. The proportion among the training dataset, testing dataset and validation dataset was set as 9:0.5:0.5.

Before the experiment, the images were preprocessed to prevent the convolution process from overfitting; Moreover, the data was rotated, as shown in Fig. 6, to increase the quantity of the data and enhance the accuracy.

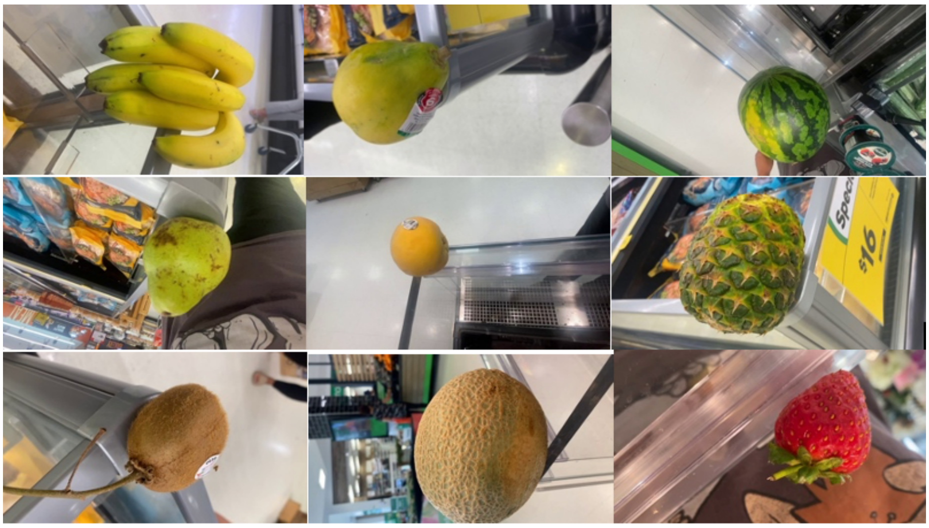


**Fig 5.** Image samples in datasets

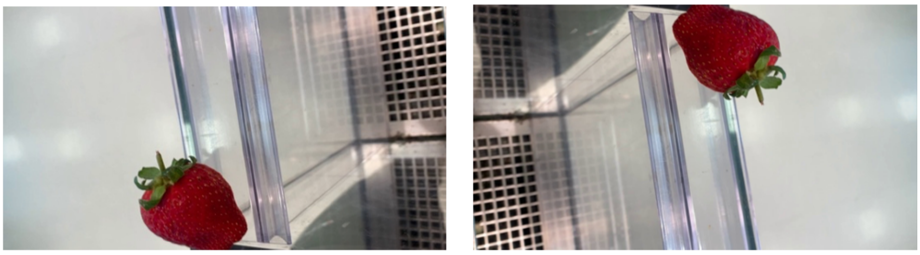

**Fig 6.** Data augmentation of image by rotation

Adding noise to images before training can help improve the generalization performance of deep learning models by making them more robust to variations and noise in real-world data. It can also prevent overfitting, encourage the model to learn more useful features, and help avoid memorization of specific training examples. Based on the characteristics of the fruit dataset, the mottles were added as the special type of noise to simulate the overripe and rotten skin. Fig. 7 displays the images after salt noise injection as data augmentation that were processed by MATLAB.

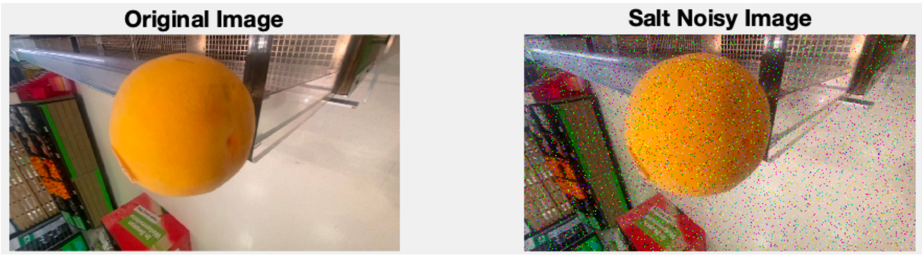


**Fig 7.** Noise injection for original image

YOLOv8n had 225 layers, 3,011,043 parameters, 3,011,027 gradients, and 100 epochs completed in 4.542 hours. Compared with YOLOv8n model, the YOLOv8s model had the same layer and approximately one-hour faster training speed; However, the parameters and gradients reached 11,144,888 and 11,144,872 respectively. Almost all confidence in the results is over 0.94 which is demonstrated in Fig .8. Fig. 9 depicts the accuracy in confusion matrix. YOLOv8n and YOLOv8s present the same accuracy of all classes that is over 90%. The accuracy of a single category is called AP (Average Precision) and the average accuracy of all categories is known as mAP. The mAP50 represents the average accuracy of all categories when the IoU is 50, which is calculated by determining the AP of all images in each category. A higher AP value indicates better accuracy for that category. The results of the experiment, shown in Fig. 10, indicate that the mAP50 and mAP50-90 values trained by YOLOv8n and YOLOv8s are as expected.


**Fig 8.** Fruit identification result with confidence

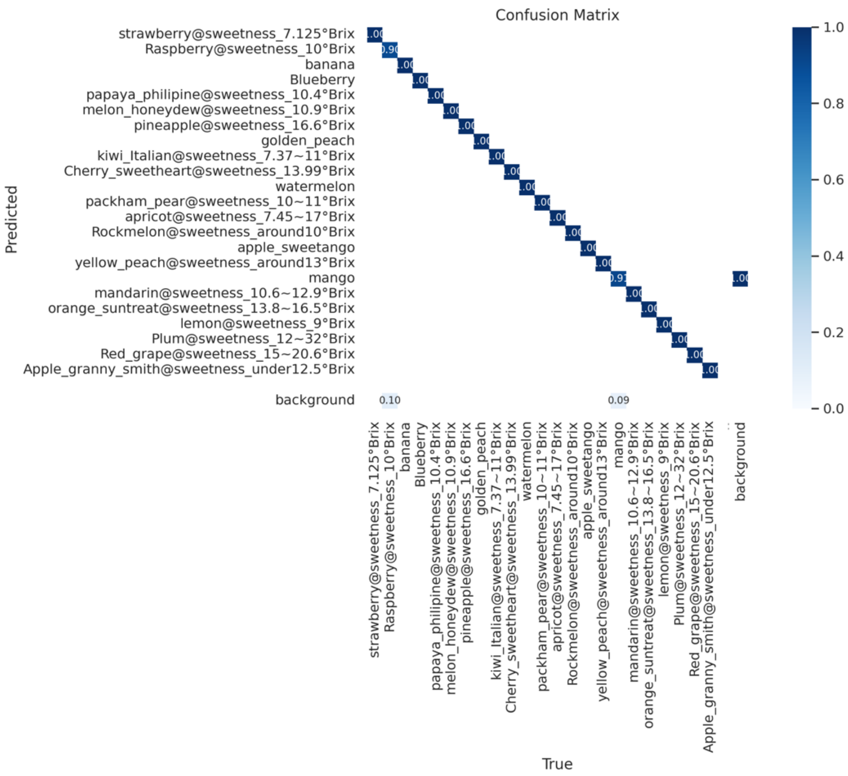


**Fig 9.** The confusion matrix

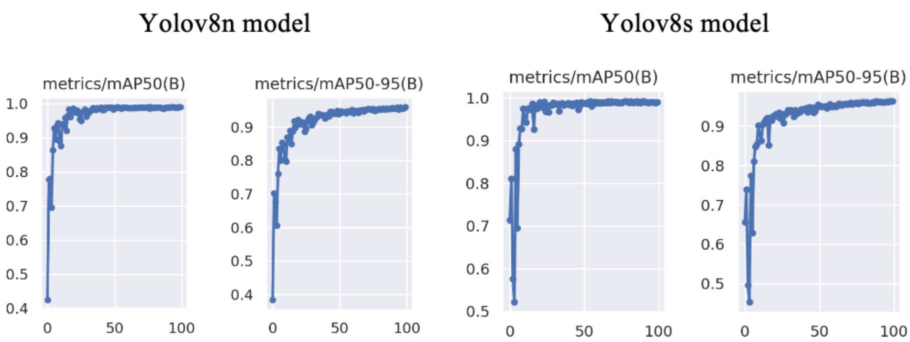


**Fig 10.** The value of mAP50 and mAP50-95

## 5 Conclusion and future work

In our experiment, over 3,000 fruit images were trained by using YOLOv8 algorithm, such as “mango”, “banana”, “papaya”, that can purchase in the countdown and the new world supermarkets. In the training process, the detection accuracy of fruit got over 90%. The experimental results demonstrate that the confidence level of all classes was more than 90% in the validation process.

In this experiment, YOLOv8n and YOLOv8s models were employed, but there are three other models available, namely YOLOv8l, YOLOv8m, and YOLOv8x. The variations among these models lie in the number of convolution kernels, network structure, depth, and width, resulting in different mAP values. AP measures the performance of individual categories, while mAP is the average of all AP values. The pre-trained checkpoints released by ultralytics showed that YOLOv8x has higher mAP compared to YOLOv8n and YOLOv8s. Model pruning is a primary approach to reduce the number of convolution kernels, depth, and width in the network structure. These differences among models are based on theoretical studies of model pruning and inspired by the phenomenon that the number of useless synapses in the human brain increases and then decreases with age [41].

After conducted these experiments, fruit identification may be relatively mature. In addition, the measurement approach of Brix value needs to be improved if the program used in a real scene. The real-time fruit brix value should be measured, updated, and released by supermarkets.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Data availability** The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request

## References


1. Athanasios, V., Nikolaos, D., Anastasios, D., Eftychios, P.: Deep learning for computer vision: A brief review. Computational Intelligence and Neuroscience, vol. 2018, pp. 1-13 (2018).
2. Yann, L., Bernhard, B., John, D., Donnie, H., Howard, R., Wayne, H., Lawrence, J.: Handwritten digit recognition with a back-propagation network. In Conference and Workshop on Neural Information Processing Systems (NIPS) (1989).
3. Sepp, H., Jürgen, S.: Long short-term memory. Neural Computation, vol. 9, no. 8, pp. 1735–1780 (1997).
4. Geoffrey, E. H., Simon, O., Yee-Whye, Teh.: A fast learning algorithm for deep belief nets. Neural Computation, vol. 18, no. 7, pp. 1527–1554 (2006).
5. Bastien, F.: Theano: New features and speed improvements. Conference and Workshop on Neural Information Processing Systems (NIPS) (2012).
6. Hyeonwoo, N., Seunghoon, H., Bohyung, Han.: Learning deconvolution network for semantic segmentation. In IEEE International Conference on Computer Vision (ICCV), pp. 1520–1528 (2015).
7. Jonathan, L., Evan, S., Trevor, D.: Fully convolutional networks for semantic segmentation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431–3440 (2015).
8. Nikolaos, D., Athanasios, V.: FAST-MDL: Fast Adaptive supervised training of multi-layered deep learning models for consistent object tracking and classification. In IEEE International Conference on Imaging Systems and Techniques (IST) (2016).
9. Nikolaos, D.: Adaptable deep learning structures for object labeling/tracking under dynamic visual environments. Multimedia Tools and Applications, vol. 77, no. 8, pp. 9651–9689 (2017).
10. Liang, L., Keze, W., Wangmeng, Z., Meng, W., Jiebo. Luo., Lei, Z.: A deep structured model with radius–margin bound for 3D human activity recognition. International Journal of Computer Vision, vol. 118, no. 2, pp. 256–273 (2016).
11. Wanli, O., Xiaogang, W., Xingyu, Z., Shi, Q., Ping, L., Yonglong, T., Hongsheng, L., Shuo, Y., Zhe, W., Chen-Change, L., Xiaoou, T.: DeepID-Net: Object detection with deformable part based convolutional neural networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 7, pp. 1320–1334 (2017).
12. Ali, D., Vivek, S., Ali, P., Hamed, P., Luc, V. G.: Weakly supervised cascaded convolutional networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 914–922 (2017).
13. Joseph, R., Santosh, D., Ross, G., Ali, F.: You Only Look Once: Unified, real-time object detection. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788 (2016).
14. Sirinnapa, S., Jinda, S., Sumio, K.: Performance of a portable near infrared (NIR) instrument for Brix value determination of intact mango fruit. In International Conference, pp. 175–181 (2003).
15. Haoyang, Z., Ying, W., Feras, D., Niko, S.: VarifocalNet: An IoU-aware dense object detector. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8514–8523 (2021).
16. Xiang, L., Wenhai, W., Lijun, W., Shuo, C., Xiaolin, H., Jun, L., Jinhui, T., Jian, Y.: Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. In Conference and Workshop on Neural Information Processing Systems (NIPS) (2020).
17. Hashizume, T., Shimamoto, I., Hirai, M.: Construction of a linkage map and QTL analysis of horticultural traits for watermelon using RAPD, RFLP and ISSR markers. Theoretical and Applied Genetics, vol. 106, pp. 779–785 (2003).

18. Zhang, H., Jun, W., Sheng, Y., Mingxun, C.: Application of electronic nose and statistical analysis to predict quality indices of peach. Food and Bioprocess Technology, vol. 5, pp. 65–72 (2012).
19. Espada, J. L., Romero, J., Socias, i. C. R., Alonso, J. M.: Preview of the second clonal selection from the autochthonous peach population. Acta Horticulturae, vol. 1, no. 814, pp. 251–254 (2009).
20. Paolo, C., Riccardo, M.,Fabio, M.: Vis-NIR measurement of soluble solids in cherry and apricot by PLS regression and wavelength selection. Journal of Agricultural and Food Chemistry, vol. 48, no. 11, pp. 5236–5242 (2000).
21. Antoniolli, L.R., Czermainski, A.B.C.: Maturity index and cold storage effects on postharvest quality of 'Packham's Triumph' and 'Rocha' pears. Acta Horticulturae, no.934, pp. 865–870 (2012).
22. Hyoung, S. L., Ronald, E. W.: Apple juice composition: Sugar, nonvolatile acid, and phenolic profiles. Journal of Association of Official Analytical Chemists, vol. 71, no. 4, pp, 789–794 (1988).
23. Sumio, K., Takayuki, F., Mutsuo, I.: Nondestructive determination of sugar content in satsuma mandarin using near infrared (NIR) transmittance. Journal of the Japanese Society for Horticultural Science, vol. 62, no. 2, pp. 465–470 (1993).
24. Nimkarde, D. S. A. P.: Effect of Microwaves on the pH and Brix value of Cranberry, Grape, Blackberry and Lemon. Journal of Advanced Applied Scientific Research, vol. 4, no. 1, pp. 74–79 (2022).
25. Kinley, D.: Morphological identification of Mandarin (Citrus reticulata Blanco) in Bhutan. Agriculture and Natural Resources, vol. 45 no. 5 (2011),
26. Slaughter, D.C., Cavaletto, C.G., Gautz, L.D., Paull, R.E.: Non-destructive determination of soluble solids in papayas using near infrared spectroscopy. In International Conference, pp. 223–228 (1999).
27. Aye, A. M., Thanda, A.: Qualitative characters and sensory test of nutrient treated on carica papaya. In Myanmar Korea Conference Research Journal, vol. 3, no.1 (2020) .
28. Pelegrine, D.H., Silva, F.C., Gasparetto, C. A.: Rheological behavior of pineapple and mango pulps. LWT, vol. 35, no. 8, pp. 645–648 (2002).
29. Castaldo, D., Voi, A. Lo., Trifiro, A., Gherard, S.: Composition of Italian Kiwi (Actinidia chinensis) Puree. Journal of Agricultural and Food Chemistry, vol. 40, pp. 594–59 (1992).
30. Zulkarami, B., Ashrafuzzaman, M., Mohd, R.I.: Morpho-physiological growth, yield and fruit quality of rock melon as affected by growing media and electrical conductivity. Journal of Food, Agriculture & Environment, vol. 8, no. 1, pp. 249–252 (2010).
31. Suriyan, S., Gregory, A. T.: Physicochemical changes in fresh-cut Honeydew melon fruit during storage. African Journal of Agricultural Research, vol. 6, no. 12, pp. 2737–2742 (2011).
32. Cengiz, C., Mehmet, S. A., Muharrem, D.: Extending the quality of fresh strawberries by equilibrium modified atmosphere packaging. European Food Research and Technology, vol. 227, pp. 1575–1583 (2008).
33. Michael, N.: Fundamental and applied aspects of plum (Prunus domestica) breeding. Fruit, Vegetable and Cereal Science and Biotechnology (FVCSB), pp. 140–156 (2011).
34. Lee, C. Y., Bourne, M. C.: Changes in grape firmness during maturation. Journal of Texture Study, vol. 11, no. 2, 163–172 (1980).
35. George, A. S., Ronald, E. W.: Anthocyanin pigment, nonvolatile acid, and sugar composition of red raspberry juice. Journal of Association of Official Analytical Chemists, vol. 70, no. 6, pp. 1036–1046 (1987).
36. Ali, A. H., Nurullah, D.: Physicochemical characteristics, antioxidant activity, organic acid and sugar contents of 12 sweet cherry (Prunus Avium L.) cultivars grown in Turkey. Food Science, vol. 80, no. 3, pp. C564–C570 (2015).

37. Al-Sarayreha, M. (2020) Hyperspectral Imaging and Deep Learning for Food Safety. PhD Thesis. Auckland University of Technology, New Zealand
38. An, N., Yan, W. (2021) Multitarget tracking using Siamese neural networks. ACM Transactions on Multimedia Computing, Communications and Applications.
39. An, N. (2020) Anomalies Detection and Tracking Using Siamese Neural Networks. Master's Thesis. Auckland University of Technology, New Zealand.
40. Fu, Y., Nguyen, M., Yan, W. (2022) Grading methods for fruit freshness based on deep learning. Springer Nature Computer Science.
41. Fu, Y. (2020) Fruit Freshness Grading Using Deep Learning. Master's Thesis. Auckland University of Technology, New Zealand.
42. Gowdra, N., Sinha, R., MacDonell, S., Yan, W. (2021) Maximum Categorical Cross Entropy (MCCE): A noise-robust alternative loss function to mitigate racial bias in Convolutional Neural Networks (CNNs) by reducing overfitting. Pattern Recognition.
43. Gowdra, N. (2021) Entropy-Based Optimization Strategies for Convolutional Neural Networks. PhD Thesis, Auckland University of Technology, New Zealand.
44. Liu, Z., Yan, W., Yang, B. (2018) Image denoising based on a CNN model. International Conference on Control, Automation and Robotics.
45. Pan, C., Yan, W. (2018) A learning-based positive feedback in salient object detection. International Conference on Image and Vision Computing New Zealand.
46. Pan, C., Yan, W. (2020) Object detection based on saturation of visual perception. Multimedia Tools and Applications, 79 (27-28), 19925-19944.
47. Pan, C., Liu, J., Yan, W., Zhou, Y. (2021) Salient object detection based on visual perceptual saturation and two-stream hybrid networks. IEEE Transactions on Image Processing.
48. Qi, J., Nguyen, M., Yan, W. (2022) Small visual object detection in smart waste classification using Transformers with deep learning. International Conference on Image and Vision Computing New Zealand (IVCNZ).
49. Qi, J., Nguyen, M., Yan, W. (2022) Waste classification from digital images using ConvNeXt. Pacific-Rim Symposium on Image and Video Technology (PSIVT).
50. Qi, J., Nguyen, M., Yan, W. (2023) CISO: Co-iteration semi-supervised learning for visual object detection. Multimedia Tools and Applications.
51. Qi, J., Nguyen, M., Yan, W. (2024) NUNI-Waste: Novel semi-supervised semantic segmentation for waste classification with non-uniform data augmentation. Multimedia Tools and Applications.
52. Wang, L., Yan, W. (2021) Tree leaves detection based on deep learning. International Symposium on Geometry and Vision.
53. Xia, Y., Nguyen, M., Yan, W. (2022) A real-time Kiwifruit detection based on improved YOLOv7. International Conference on Image and Vision Computing New Zealand (IVCNZ)
54. Xia, Y., Nguyen, M., Yan, W. (2023) Kiwifruit counting using KiwiDetector and KiwiTracker. IntelliSys conference.
55. Xia, Y., Nguyen, M., Yan, W. (2023) Multiscale Kiwifruit detection from digital images. PSIVT.
56. Xiao, B., Nguyen, M., Yan, W. (2021) Apple ripeness identification using deep learning. International Symposium on Geometry and Vision.
57. Xiao, B., Nguyen, M., Yan, W. (2023) Apple ripeness identification from digital images using transformers. Multimedia Tools and Applications, Springer Science and Business Media LLC.
58. Xiao, B., Nguyen, M., Yan, W. (2023) Fruit ripeness identification using transformers. Applied Intelligence, Springer Science and Business Media LLC.
59. Xiao, B., Nguyen, M., Yan, W. (2023) A mixture model for fruit ripeness identification in deep learning. Handbook of Research on AI and ML for Intelligent Machines and Systems, IGI Global.

60. Xiao, B., Nguyen, M., Yan, W. (2023) Fruit ripeness identification using YOLOv8 model. Multimedia Tools and Applications.
61. Xue, Y. Yan, W. (2023) YOLO models for fresh fruit classification from digital videos. Handbook of Research on AI and ML for Intelligent Machines and Systems, IGI Global.
62. Xue, Y. (2023) YOLO models for fresh fruit classification from digital videos. Master's Thesis, Auckland University of Technology, New Zealand.
63. Yan, W. (2019) Introduction to Intelligent Surveillance: Surveillance Data Capture, Transmission, and Analytics. Springer Nature.
64. Yan, W. (2023) Computational Methods for Deep Learning: Theory, Algorithms, and Implementations. Springer Nature.
65. Zhao, K. (2021) Fruit Detection Using CenterNet. Master's Thesis, Auckland University of Technology, New Zealand.
66. Zhao, K., Yan, W. (2021) Fruit detection from digital images using CenterNet. International Symposium on Geometry and Vision.
67. Zhou, H., Nguyen, M., Yan, W. (2023) Computational analysis of table tennis matches from real-time videos using deep learning. PSIVT 2023.
68. Zhou, H. (2024) Computational Analysis of Table Tennis Games from Real-Time Videos Using Deep Learning. Master's Thesis, Auckland University of Technology, New Zealand.
69. Zhu, Y., Yan, W. (2022) Image-based storytelling using deep learning. ACM ICCCV.