Detection, Counting and Maturity Assessment of Cherry Tomatoes using Multi-spectral Images and Machine Learning Techniques

I-Tzu Chen and Huei-Yung Lin
Department of Electrical Engineering, National Chung Cheng University, Chiayi 621, Taiwan

Keywords: Yield Estimation, Maturity Assessment, Object Detection, Object Counting, Multi-object Tracking.

Abstract: This paper presents an image-based approach for the yield estimation of cherry tomatoes. The objective is to assist farmers to quickly evaluate the amount of mature tomatoes which are ready to harvest. The proposed technique consists of machine learning based methods for detection, counting, and maturity assessment using multi-spectral images. A convolutional neural network is used for tomato detection from RGB images, followed by the maturity assessment using spectral image analysis with SVM classification. The multi-object tracking algorithm is incorporated to obtain a unique ID for each tomato to avoid double counting during the camera motion. Experiments carried out on the real scene images acquired in an orchard have demonstrated the effectiveness of the proposed method.

1 INTRODUCTION

As the world's farming population is decreasing and growing older, automated agriculture has become an important trend in the future. One major issue to deal with is the yield estimation of crop production, which is highly related to the harvest schedule, labor allocation, storage and transportation, etc. Having the right time to harvest will also result in better food quality and reduce the waste in the supply chain (Lobell et al., 2015). To estimate the yield, taking fruits as an example, one needs to consider not only the number counting, but also the maturity inspection on each of them. Through the on-site non-invasive maturity and harvest assessment, the yield can be estimated on a daily or weekly basis to avoid immaturity and spoilage of the crops. This also provides a way to deliver fresh fruits to the consumers with best natural quality.

In this paper, we present an image-based approach for the yield estimation of cherry tomatoes. The objective is to assist farmers to quickly evaluate the amount of mature tomatoes which are ready to harvest. The proposed technique consists of machine learning based methods for detection, counting, and maturity assessment using multi-spectral images. Due to the cluttered scenes and backgrounds appeared in the images, it is a challenging task to identify the matured fruits in the orchard environment (Li et al., 2011). Furthermore, multiple images captured from different viewpoints are required to cover the working region for the detection and counting of tomatoes. This is done by moving a camera in the orchard to record video sequences, followed by the multiple object tracking technique to assign the unique identities for individual tomatoes.

Agricultural researchers have investigated fruit detection in orchards with different sensor systems for many decades (Gongal et al., 2015). The objective is to distinguish fruits from the background (leaves, branches, flowers) for inspection or counting. The commonly used devices are B/W or color cameras, spectral cameras, and thermal cameras, etc. With the camera based approaches, image features including color, texture, edge and gradient are extracted for object detection and classification. The classification techniques such as K-means and KNN clustering, Bayesian classifier, artificial neural network (ANN), support vector machine (SVM) are then used to identify and localize the fruits. More recent machine learning based approaches adopt convolutional neural networks (CNNs) for fruit detection and counting within the trees. Nevertheless, it is a more challenging task compared to the conventional fruit category classification in factories or supermarkets (Zhang et al., 2014).

The fruit detection is essentially a problem of object detection but with a specific category. It is further restricted to the detection of a certain type of fruits, i.e., cherry tomato, in this application. However, different from the detection and classification in households or supermarkets, our task needs to deal with the com-

Chen, I. and Lin, H.
Detection, Counting and Maturity Assessment of Cherry Tomatoes using Multi-spectral Images and Machine Learning Techniques.
DOI: 10.5220/0008874907590766
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 5: VISAPP*, pages 759-766
ISBN: 978-989-758-402-2; ISSN: 2184-4321

plex background and occlusion in the orchards. In the conventional image based approaches, Chaivivatrakul and Dailey present a technique using texture analysis to detect green fruits on plants (Chaivivatrakul and Dailey, 2014). The interested points are extracted with a feature descriptor and classified by SVM. A series of morphological operations are then carried out for the fruit region identification. They have reported the detection rates of 85% and 100% for pineapple and bitter melon, respectively. To overcome the unreliable recognition problems caused by uneven illumination, partial occlusion, and similar background features, Rakun *et al.* propose a method combining color, texture and 3D shape properties of the objects (Rakun et al., 2011). Multiple images captured from different viewpoints are used to separate potential regions from the background. The fruit size and yield are also estimated with texture analysis and reconstruction in 3D space.

In addition to the classification methods based on SVM, there are also K-means and KNN clustering techniques for fruit detection. Bulanon *et al.* develop a robotic harvesting system with automatic recognition of Fuji apples on the trees (Bulanon et al., 2002). The images are enhanced with the color difference, and the intensity histogram is used to separate the background. With K-means clustering, they have reported a success rate of 88% using the optimal threshold. In (Yamamoto et al., 2014), a method is presented to detect intact tomatoes using a modified K-means clustering algorithm. It does not require manual threshold adjustment and is able to separate the mature, immature and young tomatoes on a plant. Seng and Mirisaee develop a fruit recognition using KNN classification (Woo Chaw Seng and Mirisaee, 2009). They adopt color, shape and size features to classify and recognize seven kinds of fruits with about 90% of accuracy. However, the proposed technique might not work in the complex outdoor environment.

More recently, deep learning approaches are adopted for precision agriculture research (Kamilaris and Prenafeta-Bold, 2018). Sa *et al.* present a fruit detection system using deep convolutional neural networks (Sa et al., 2016). They adopt Faster R-CNN (Ren et al., 2017) as the detection framework and use color and near infrared images as the input. Through the transfer learning, this multi-modal Faster R-CNN model has achieved better results on seven kinds of fruits compared to the previous work. Bargoti and Underwood also present a Faster R-CNN based approach for fruit detection in orchards (Bargoti and Underwood, 2017). They specifically focus on the data augmentation techniques used for performance improvement. The results provide a high F-1 score of 0.9 with hundred fruits per image for detection. Different from the above two-stage detection technique, Bresilla *et al.* propose a method based on a single-shot detector for fruit detection within the tree canopy (Bresilla et al., 2019). Their network architecture is based on YOLO (Redmon and Farhadi, 2017) and fine-tuned with apple and pear images. The experiments demonstrate the processing speed of 20 fps, which is sufficient for real-time applications.

2 METHOD

The proposed technique uses a hand-held device for yield estimation of cherry tomatoes in the orchards. We integrate object detection, spectral information and multi-object tracking to identify the amount of tomatoes and the level of maturity. Figure 1 shows the system flowchart of the proposed technique. The RGB images are passed to a detection network to identify the position of each tomato. The maturity of individual tomato is determined by spectral image analysis using NDVI images. A unique ID for each tomato is then obtained using a multi-object tracking network to avoid double counting. Finally, the yield estimation is given by counting with different maturity levels.

2.1 Detection Network

This paper adopts the two-stage detector Faster R-CNN (Ren et al., 2015) as the main detection network. The network architecture has a higher detection accuracy than one-stage detectors and better detection speed than the semantic segmentation techniques. From the experimental results of Faster R-CNN, using the deeper VGG16 (Simonyan and Zisserman, 2014) has a better accuracy than using a shallower network, but there is a degradation problem when constantly deepening the neural network. The accuracy rises first and then saturates, and keeping increasing the depth will result in a decrease in accuracy. One reason is that, the deeper the network, the more obvious the gradient disappears. This causes the network parameters cannot be updated. ResNet (He et al., 2016) introduces a residual network structure which can continue to deepen the network layers. It is adopted to improve the final classification results.

When Faster R-CNN performs object detection, whether it is RPN or CNN, the RoI acts on the last layer. This might not be a problem for large target detection. But there is virtually no semantic information for small targets when pooling to the last layer. A classic way to deal with small objects is to use image

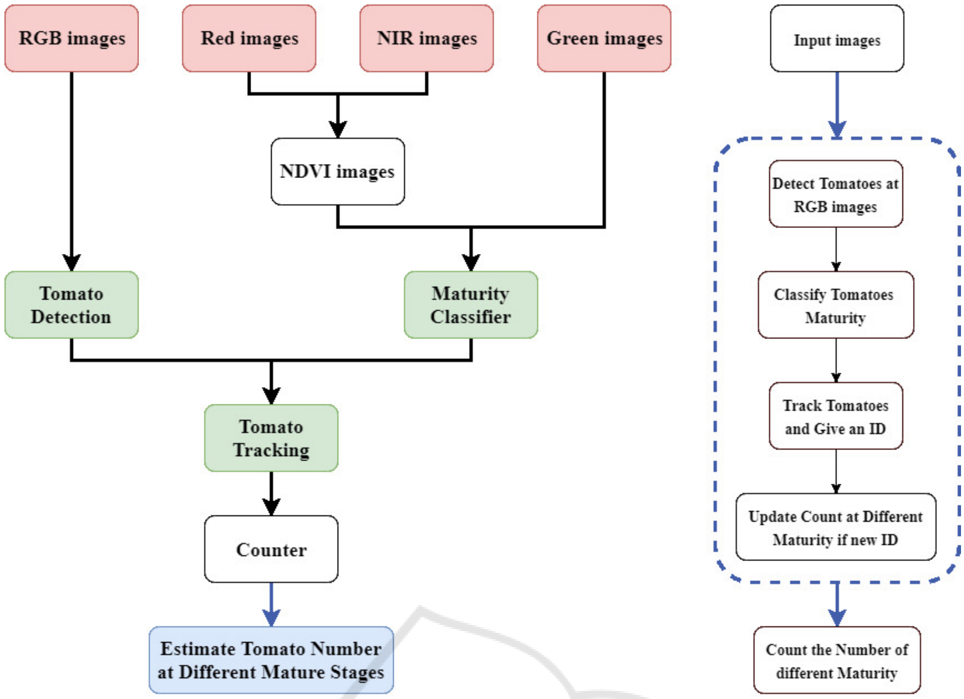


Figure 1: The system flowchart of the proposed tomato yield estimation technique.

(a) NIR image (b) red image (c) green image (d) NDVI image

Figure 2: The multi-spectral image consists of NIR, red and green channels. The green and NDVI images are used for maturity assessment.

pyramids to take multi-scale variations of the images during training and testing. This will also increase the computation significantly. The feature pyramid networks (FPN) (Lin et al., 2017) is introduced to solve the multi-scale detection problem. The FPN structure is designed with bottom-up and top-down pathways, and lateral connections to fuse the high resolution shallow features and the deep features with rich semantic information. This can quickly construct a feature pyramid with strong semantic information at all scales from a single-scale input image without a significant cost.

2.2 Maturity Assessment

The sunlight contains X-ray, ultraviolet, visible light, infrared, etc. The visible spectrum can be formed by red, green and blue light, and consists of million colors for human vision perception. In addition to the conventional RGB images, spectral cameras can be used to capture the images with the light that is not visible to human. For those beyond the visible spectrum, the near-infrared (NIR) images are often used in agriculture, and a variety of spectral images are combined and used for analysis. The normalized difference vegetation index (NDVI) images are the most widely studied and adopted. NIR is a specific value which reflects the vitality and productivity of plants. NDVI calculates the reflection of red light by

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{1}$$

where NIR is near-infrared reflection, RED is red light reflection, and the value of NDVI is between -1 and 1. When $RED = 0$, there is a maximum value of

1; and conversely, when $NIR = 0$, there is a minimum value of -1. The higher the value, the stronger the plant vigor. It is especially suitable for assessing the difference in the plant growth status and is also a good tool to monitor the health and vitality of plants in the field.

Healthy green plants contain much chlorophyll. Chlorophyll reflects a few red light after absorbing visible light, but strongly reflects green light and NIR. A comprehensive analysis on the NIR and red images (see Figures 2(a) and 2(b), respectively) can obtain the NDVI image, as shown in Figure 2(d). The simultaneous analysis on the green image (see Figure 2(c)) and the NDVI image can monitor the health status of plants and the vitality of photosynthetic pigments. Both images provide more clear growth information than the visible and NIR images. The spectral index usually combines the reflectance of two or more wavelengths to extract the features of interest. Because these values are easily affected by sensors and ambient light, it is difficult to use only one wavelength as the evaluation indicator. This paper uses green and NDVI images for maturity assessment.

2.3 Tracking Network

The multi-object tracking approach is used to count the number of fruits in dense detection results. It identifies the appearance characteristics of the targets, and calculates the individual motion curves. The targets can then be matched among different frames. Although a few occlusion cases can be recovered under the camera movement, many fruits are still obscured by stems and leaves. To avoid counting the same fruit after a long-term occlusion, it is required to track each of them so we can have a unique ID after occlusion. Because each target has a fixed ID, the amount of moving objects in the image sequence can be determined. We use multi-object tracking to determine the ID of each detected tomato, and then count the amount of tomatoes in the whole image sequence.

This paper adopts the multi-object tracking method DeepSORT (Wojke et al., 2017). It improves the ID switching problem after occlusion in the SORT algorithm (Bewley et al., 2016). DeepSORT presents an extension to SORT that incorporates the appearance information through a pre-trained association metric. Its target state is represented by the position, aspect ratio, height, and velocity at the center of the bounding box, and the Kalman filter (Kalman, 1960) is then used to predict and update the tracking trajectory. The matching metric is derived by fusing motion and appearance information, and then used to analyze the association between the existing target and the detection. A deep convolutional neural network (cosine index learning) is established to extract the appearance features of the target, where the cosine index learning is trained on a large-scale re-identification dataset.

(a) Non-open air orchard. (b) Open air orchard.

Figure 3: The tomato planting space consists of non-open and open air orchards.

3 DATASETS

Several public datasets for fruit classification and identification are shown in Table 1. Since there are not suitable for tomato yield estimation in this work, we have collected our own dataset for both training and testing.

Depending on the variety of tomatoes, there are two types of tomato planting space: the non-open air space as shown in Figure 3(a) and the open air space as shown in Figure 3(b). Our data collecting sites are mostly the non-open air space. The distance within the scaffold is about 80 cm. We move the camera along the plants at about 50 cm to capture a larger range of the scaffold. It is also less affected by the tomato vines during the image acquisition process. Three devices are used, a spectral camera Parrot SEQUOIA+, a digital camera Nikon Coolpix P7000, and an iPhone 8 Plus mobile phone.

The training data used by Faster R-CNN are the images taken by an iPhone 8 Plus. For maturity assessment, the testing is carried out using the images taken by the spectral camera Parrot SEQUOIA+. On the testing stage, the images taken by iPhone 8 Plus are only used to evaluate the detection performance. There are totally 1,086 and 342 images used for training and testing, respectively. Originally, DeepSORT trains the cosine metric learning model with large pedestrian datasets Market1501 and MARS (Zheng et al., 2016). This is not suitable for the tomato related application in this work. We have fine-tuned using our tomato training dataset based on the MARS format. For DeepSORT tracking, there are additional 105 and 40 images used for training and testing, respectively, and some images are shown in Figure 4.

Table 1: The available public datasets for fruit recognition.

Dataset	# of image	# of class	Image resolution	Tomato image
Fruits-360 (Mureşan and Oltean, 2018) (version: 2019.07.07.0)	77,917	114	100×100 pixels	4,654
Fruit recognition dataset (Hussain et al.,)	44,406	15	320×258 pixels	2,171
FIDS30 (Škrjanec Marko, 2013)	971	30	not fixed	46

Figure 4: Some training samples (mature and immature tomatoes) used for tracking.

4 EXPERIMENT

In the field experiments, the scaffolding is full of tomatoes with different maturity levels. A spectral camera is moved horizontally to capture the images for yield estimation in a non-destructive way. The RGB images are used by Faster R-CNN for tomato detection, and the tomatoes in the green and NDVI image are used for maturity assessment. DeepSORT takes the detection result for tracking and obtains a unique ID for each tomato. Each tomato has its own ID and maturity, so the total number of tomatoes and the numbers with different levels of maturity can be derived.

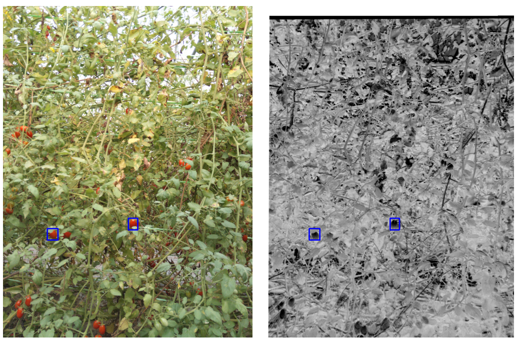


Figure 5: The RGB image (left) and the NDVI image (right). Tomato detection and tracking in the RGB image is used for the NDVI image.

4.1 Detection and Maturity Assessment

The Faster R-CNN structure used in this work is combined with ResNet-50 and FPN. Hundreds of labeled still tomato images are imported to Faster R-CNN for training. After training, the tomato video sequences are used for detection and output the bounding boxes of tomatoes in each frame. The training data contain the images captured at different distances to cover various camera positions in the application scenario. To evaluate the detection performance, the testing images are taken with an iPhone 8 Plus. Table 2 tabulates the results of tomato detection using Faster R-CNN and YOLO v3 (Redmon and Farhadi, 2018), including mAP, false prediction and true prediction. The result shows that Faster R-CNN has a higher mAP and a relatively high proportion of correct prediction.

The tomatoes obtained by the detection network correspond to the same positions of the green and NDVI images as shown in Figure 5. The criterion for maturity assessment of each tomato is determined by the maturity classification obtained by SVM. It is divided into two categories: mature and immature. According to this classification, the tomato maturity is assessed with the bounding boxes.

There are four steps to generate the NDVI images with the red and NIR images. First, we remove the lens distortion from the Parrot SEQUOIA+ spectral image, followed by the reflectivity calculation of the undistorted image. The third step is to align the red and undistorted NIR reflectance images with the orig-

Table 2: The detection network evaluation results.

Detection Network	mAP	False Prediction	True Prediction
Faster R-CNN	62.35%	925	3159
YOLO v3	53.50%	1319	2711

Table 3: The evaluation of the detection network and maturity.

Detection Network	Histogram matching	Mature AP	Immature AP	mAP
Faster R-CNN	✓	55.22%	36.82%	46.02%
	✗	57.81%	32.46%	45.14%
YOLO v3	✓	56.11%	25.28%	40.69%
	✗	51.46%	3.94%	27.72%

Table 4: The evaluation of maturity assessment.

Detection Network	Histogram matching	Mature AP	Immature AP	mAP
Faster R-CNN	✓	99.59%	99.75%	99.67%
	✗	97.86%	98.96%	98.41%
YOLOv3	✓	95.73%	97.21%	96.47%
	✗	93.15%	56.60%	74.87%

inal RGB image. Finally, the red and NIR images are calculated to generate the NDVI image. To assess the tomato maturity, the values of green and NDVI images are used as the two-dimensional input features. The hyperplane of the SVM classification is a line. The total number of SVM training samples is 42, and the test accuracy is more than 90%. The detailed evaluation results are presented in Section 4.3.

4.2 Tomato Tracking

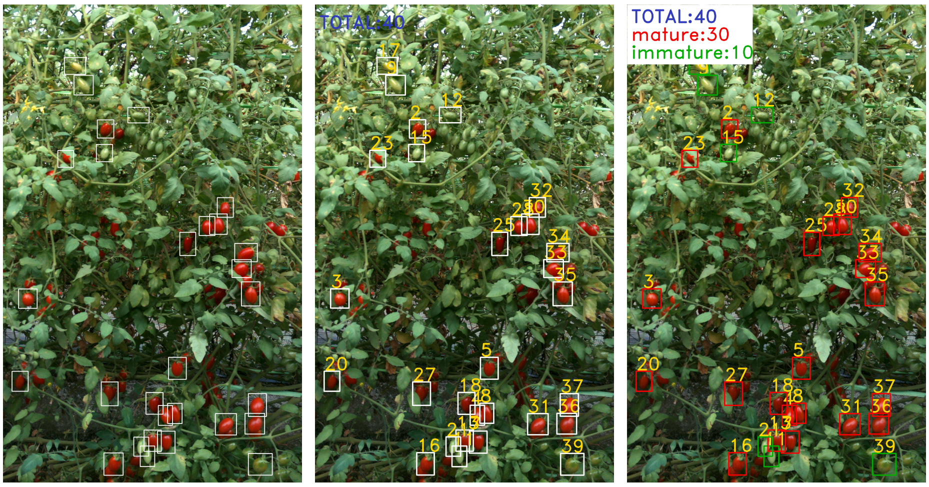
In the orchard scenes, the images of tomatoes are obscure. We use DeepSORT to track the tomatoes, and expect to achieve multi-object tracking and avoid the ID switch problem caused by long-term occlusions at the same time. DeepSORT utilizes the tomato position detected by Fast R-CNN and traces it using the cosine metric learning model trained in our tomato dataset. This approach tracks the same tomato and give it a unique ID to avoid the double counting problem. In Figure 6(b), the yellow number represents the unique ID of each tomato, and the blue number in the upper left corner indicates how many tomatoes have been detected in the image sequence. We apply the evaluation tool developed by the MOT competition. The test video passes through the detection and tracking networks to generate the bounding box and ID of the tomato. The information is then converted to the format of the evaluation tool to calculate the accuracy. Detailed evaluation results and comparisons are described in Section 4.3.

For number counting with different maturity levels, the image sequence is first passed to Fast R-CNN to detect tomatoes in each frame, and obtain the features on the green and NDVI images of each tomato to determine its maturity. We then use DeepSORT to track each tomato from its entering to leaving the frame and retain the ID in the tracking to avoid double counting. Finally, the ID and maturity information are used to derive the number of tomatoes with different maturity levels. Figure 6(c) shows a result with the number of mature and immature tomatoes.

4.3 Comprehensive Evaluation

The performance evaluation is carried out on the detection network and maturity assessment, simple maturity assessment, and detection network and tracking network. We compare different detection networks, different tracking networks, and input images with different hues. The inputs are the original images and the histogram matching images. There are totally 191 images taken with Parrot SEQUOIA+ for testing.

The categories used to evaluate the detection network and maturity assessment are tomato and mature/immature, respectively. The evaluation criteria are mature AP, immature AP, and mAP. From the results shown in Table 3, it is found that using the histogram-matched images, with whether Faster R-CNN or YOLOv3 for tomato detection, has higher mAPs than using the original images. Compare Faster R-CNN and YOLOv3, it can be seen that YOLOv3 performs poorly in the immature AP. This might be due to the ability of YOLOv3 on detecting the green



(a) Faster R-CNN result. (b) DeepSORT tracking. (c) Counting with maturity.

Figure 6: Some detection and tracking results obtained in the tomato orchard.

IDF1	IDP	IDRI	Rcll	Prcn	FARI	GT	MT	PT	MLI	FP	FN	IDs	FMI	MOTA	MOTP	MOTAL
51.4	76.8	38.6I	38.6	76.8	6.34I	74	26	3	45I	1210	6393	1	97I	26.9	60.6	27.0

(a) SORT with Faster R-CNN

IDF1	IDP	IDRI	Rcll	Prcn	FARI	GT	MT	PT	MLI	FP	FN	IDs	FMI	MOTA	MOTP	MOTAL
51.2	77.6	38.2I	38.2	77.6	5.99I	74	25	5	44I	1144	6435	0	122I	27.2	61.2	27.2

(b) DeepSORT with Faster R-CNN

Figure 7: The evaluation of the detection and tracking networks.

tomatoes that blend with the background.

The comprehensive evaluation of the detection network with maturity assessment cannot see whether the maturity assessment is accurate since the accuracy is affected by the detection result. To evaluate the maturity assessment exclusively, only the detected tomatoes are evaluated. The results tabulated in Table 4 show that the accuracy of the maturity assessment is very high, and the lowest one is the immature AP that uses YOLOv3 without histogram matching. This is mainly due to the sensitivity caused by a fairly small number of immature samples. The maturity and immaturity of tomatoes are classified when evaluating the detection and tracking networks. The MOT evaluation method can only perform a comprehensive evaluation of the detection and tracking network without maturity assessment. Figure 7 shows that using DeepSORT as the tracking network provides better evaluation results than using SORT.

5 CONCLUSION

This paper presents an approach for detection, counting, and maturity assessment of cherry tomatoes using multi-spectral images and machine learning techniques. A CNN-based network is used for tomato detection from RGB images, followed by the maturity assessment using spectral image analysis with SVM classification. The multi-object tracking algorithm is incorporated to obtain a unique ID for each tomato to avoid double counting during the camera motion. A comprehensive evaluation carried out on the real scene images captured in the orchard have demonstrated the effectiveness of the proposed method for the yield estimation.

REFERENCES

Bargoti, S. and Underwood, J. (2017). Deep fruit detection in orchards. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3626–3633.

Bewley, A., Ge, Z., Ott, L., Ramos, F., and Upcroft, B. (2016). Simple online and realtime tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3464–3468. IEEE.

Bresilla, K., Perulli, G. D., Boini, A., Morandi, B., Grappadelli, L. C., and Manfrini, L. (2019). Single-shot convolution neural networks for real-time fruit detection within the tree. *Frontiers in plant science*, 10.

Bulanon, D., Kataoka, T., Ota, Y., and Hiroma, T. (2002). Aeautomation and emerging technologies: A segmentation algorithm for the automatic recognition of fuji apples at harvest. *Biosystems Engineering*, 83(4):405 – 412.

Chaivivatrakul, S. and Dailey, M. N. (2014). Texture-based fruit detection. *Precision Agriculture*, 15(6):662–683.

Gongal, A., Amatya, S., Karkee, M., Zhang, Q., and Lewis, K. (2015). Sensors and systems for fruit detection and localization: A review. *Computers and Electronics in Agriculture*, 116:8 – 19.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Hussain, I., He, Q., and Chen, Z. Automatic fruit recognition based on dcnn for commercial source trace system.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45.

Kamilaris, A. and Prenafeta-Bold, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147:70 – 90.

Li, P., heon Lee, S., and Hsu, H.-Y. (2011). Review on fruit harvesting method for potential use of automatic fruit harvesting systems. *Procedia Engineering*, 23:351 – 366. PEEA 2011.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125.

Lobell, D. B., Thau, D., Seifert, C., Engle, E., and Little, B. (2015). A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*, 164:324 – 333.

Mureşan, H. and Oltean, M. (2018). Fruit recognition from images using deep learning. *Acta Universitatis Sapientiae, Informatica*, 10(1):26–42.

Rakun, J., Stajnko, D., and Zazula, D. (2011). Detecting fruits in natural scenes by using spatial-frequency based texture analysis and multiview geometry. *Computers and Electronics in Agriculture*, 76(1):80 – 88.

Redmon, J. and Farhadi, A. (2017). YOLO9000: better, faster, stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 6517–6525.

Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99.

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149.

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., and McCool, C. (2016). Deepfruits: A fruit detection system using deep neural networks. *Sensors*, 16(8).

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Škrjanec Marko (2013). Automatic fruit recognition using computer vision. (Mentor: Matej Kristan), Fakulteta za računalništvo in informatiko, Univerza v Ljubljani.

Wojke, N., Bewley, A., and Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3645–3649. IEEE.

Woo Chaw Seng and Mirisaee, S. H. (2009). A new method for fruits recognition system. In *2009 International Conference on Electrical Engineering and Informatics*, volume 01, pages 130–134.

Yamamoto, K., Guo, W., Yoshioka, Y., and Ninomiya, S. (2014). On plant detection of intact tomato fruits using image analysis and machine learning methods. *Sensors*, 14(7):12191–12206.

Zhang, Y., Wang, S., Ji, G., and Phillips, P. (2014). Fruit classification using computer vision and feedforward neural network. *Journal of Food Engineering*, 143:167 – 177.

Zheng, L., Bie, Z., Sun, Y., Wang, J., Su, C., Wang, S., and Tian, Q. (2016). Mars: A video benchmark for large-scale person re-identification. In *European Conference on Computer Vision*, pages 868–884. Springer.