Multiple broccoli head detection and tracking in 3D point clouds for autonomous harvesting*

Hector A. Montes, Grzegorz Cielniak[1]

[1]University of Lincoln
Brayford Pool, LN6 7TS, UK
{hmontes,gcielniak}@lincoln.ac.uk

Abstract

This paper explores a tracking method of broccoli heads that combine a Particle Filter and 3D features detectors to track multiple crops in a sequence of 3D data frames. The tracking accuracy is verified based on a data association method that matches detections with tracks over each frame. The particle filter incorporates a simple motion model to produce the posterior particle distribution, and a similarity model as probability function to measure the tracking accuracy. The method is tested with datasets of two broccoli varieties collected in planted fields from two different countries. Our evaluation shows the tracking method reduces the number of false negatives produced by the detectors on their own. In addition, the method accurately detects and tracks the 3D locations of broccoli heads relative to the vehicle at high frame rates.

Broccoli is a high-value vegetable grown worldwide due to its nutritional compounds that have increased its consumption. Traditionally, broccoli is selectively harvested by hand, relying on visual grading to estimate whether a head can be cut (Kootstra et al. 2021). Crop detection and tracking are together an important part of autonomous selective harvesting. They can also increase accuracy of other tasks in agriculture such as produce count, mapping or estimates of crop yields. The result of these tasks can then be used by farmers on the entire agricultural process to make informed decisions on several tasks including planting, harvesting and labour management (Duckett et al. 2018).

Full automation of agricultural tasks have been actively developed for some decades now, but important tasks such as harvesting or crop yield estimates still rely on intense human labour. Therefore, developing devices such as autonomous harvesters capable of detecting and tracking crops under real field conditions is essential for an effective management of vegetable crops (Kootstra et al. 2020).

In this work we present a method to detect and track crops of broccoli plants that can be incorporated into an autonomous selective harvester. The results have been extensively evaluated using datasets of different broccoli varieties collected in real farm environments in the UK and Spain. The main contributions of our paper are: (1) a simple and effective tracking algorithm for improving broccoli detections and crop estimates, as predictions are made based on dynamics and measurements. (2) an efficient framework that restricts the search of crops in posterior frames and reduces the number of false negatives a detector can produce by itself. (3) an efficient framework that runs in real-time and process 3D data at high frame rates.

Related work

Accurately detecting vegetable crops from other elements in the background remains a major challenge in autonomous selective harvesting. One early attempt to locate broccoli crops was published by (Ramirez 2006) using a small set of 13 RGB images. The method detected broccoli crops on the entire plant based on contrast and statistical texture analysis. Also using RGB images (Blok, Barth, and van den Berg 2016) described a system to detect broccoli heads of two different varieties based on a texture filter and colour analysis of the broccoli head appearance. They reported a precision score of 99.5%, a recall score of 91.2%, and a negative predictive value of 69.7%. Meanwhile, (Kusumam et al. 2017) detected heads of two broccoli varieties in depth images collected with an RGB-D sensor. They designed a processing pipeline that included an Euclidean clustering method, a 3D feature descriptor and a SVM classifier to detect the broccoli heads. They reported an average precision for the two varieties of 95.2% and 84.5%, respectively. Later (Montes et al. 2020) further expanded these results and presented three new methods capable of accurately detecting the 3D locations of broccoli heads at high frame rates. The methods involved algorithms for clustering points that either belonged to the same area bounded by an edge, or points which normal vectors angles and surface smoothness were within a similarity threshold. A SVM then classified the feature vector of each cluster as broccoli head or background object. The best experimental results reported were a precision score of 98.4%, a recall score of 95.7%, and a negative predictive value of 99.8%. In addition, a mean average precision of 96.5% computed over multiple Intersection-over-Union values was also reported for two broccoli varieties.

In the same way, deep learning techniques have been increasingly used to process and to better understand the large datasets produced in automated agricultural research. (Bender, Whelan, and Sukkarieh 2020) performed a series of

*This work has been supported by the Agriculture and Horticulture Development Board (AHDB) in the UK.

broccoli and cauliflower detection using a CNN model with a mean average precision of 95%. However, this result is for the entire broccoli plant and not the individual head, which is necessary for autonomous harvesting operations. A similar approach by (Zhu et al. 2018) based on the AlexNet network model used a dataset from ImageNet[1] to classify five categories of vegetables: broccoli, pumpkin, cauliflower, mushrooms and cucumber. The dataset was enlarged to improve training by creating rotated versions of the original images and achieved an accuracy rate of 92.1%. Seeking to develop a robot that can selectively harvest broccoli heads, (Blok et al. 2020) presented a detection method based on the Mask Region-based CNN model. In their experiments, images from three different broccoli varieties were collected in two countries using a prototype robot. The algorithm detected 229 out of 232 annotated broccoli heads, and also located 175 out of 176 heads on a dataset available online. Later, (Blok et al. 2021) compared this CNN model to an Occlusion Region-based CNN to estimate the size of broccoli heads even when the crops were heavily occluded. (Zhou et al. 2020) presented an improved CNN ResNet model for segmenting broccoli heads from RGB images. They built a yield estimation model based on the number of extracted pixels and a pixel weight value achieving an accuracy of 89.6%. In addition, a Particle Swarm Optimization algorithm and the Otsu method were used to grade the quality of broccoli heads according to a standard proposed by the authors. In (Louedec et al. 2020) another system is presented for detecting broccoli heads based on 3D information obtained from RGB-D sensors. In their technique they trained a CNN for semantic segmentation of the same datasets used by (Kusumam et al. 2017) outperforming these results and reporting high rate detection speeds of up to 50 fps. (García-Manso et al. 2021) presented another system for localization of broccoli heads based on a Faster R-CNN model built on a pre-trained ResNet-50 model. The algorithm detected broccoli heads in small 640×480 RGB images and classified them into harvestable, immature and wasted classes. The system was able to correctly detect and classify 97% of the test images, including the ones partially occluded by leaves. A wealth of other applications involving deep learning techniques in agriculture for different crops are also available in the literature (Yang and Xu 2021).

Detect-and-Track of broccoli heads

This section describes our solution to detecting and tracking crops of broccoli plants in sequences of 3D data frames. A common approach is to design a pipeline that integrates a detection step and a tracking module to perform tasks such as accurate object's count and mapping (Santos et al. 2020). Similarly, in this work we combine broccoli head detection and tracking into a single framework and introduce a registration step based on 3D feature vector data associations to avoid multiple tracking of the same crop observed in different frames.

For detection we use the best reported broccoli detector from (Montes et al. 2020) dubbed Organised Edge Segmentation (OES) to guide the tracker. OES labels points as edges based on point depth discontinuities and uses a tolerance distance to determine the difference in depth values between neighbouring points (Choi, Trevor, and Christensen 2013). A clustering step then groups points together if they belong to an area bounded by the same edge and spreads to other points in the immediate vicinity to form the clusters. Clearly, any detector capable of detecting broccoli heads can be used instead. For tracking we use a particle filter that incorporates a simple motion model, a detection step as an observation model (Wojke, Bewley, and Paulus 2017), and a feature vector histogram to model appearance similarity between tracks. We also introduce a track confirmation step to deal with misdetections and to reduce false positive trajectories.

[1] http:// www.image-net.org

Particle filter

A particle filter is an estimation model known to enable robust object tracking (Elfring, Torta, and van de Molengraft 2021). The goal is to track the state sequence x_k using estimated dynamics (e.g., direction, speed, acceleration), where x is a state vector at a discrete time step $k \in \mathbb{N}$. For estimating the state, is necessary to have a model encoding some prior knowledge on how x_k is expected to move from frame to frame. It is also required to have an observation model that relates environment observations to x_k. The motion model indicates how the state changes over time under the dynamic variables: $x_k = f_k(x_{k-1}, u_k, \varepsilon)$. Here f_k is a function that associates x between time steps $k-1$ and k using a deterministic motion input u_k and a model noise ε representing uncertainties (often Gaussian) associated to the variables. The observation model $z_k = h_k(Zx_k, \xi_k)$ is a function h_k that associates the state x_k with an expected set of observations Z and a model ξ_k representing observation noise. At each time step k, the current state changes from x_{k-1} to x_k and a new set of observations z_k is collected. The goal is to estimate the distribution of state x_k given all the observations seen so far and the knowledge about state dynamics. The steps involved in tracking can be summarized as follows:

- *Prediction*: predict a distribution of the next state given past observations and dynamic actions: $p\left(x_k \mid x_{k-1}, u_k\right) = p\left(x_k \mid x_{1:k-1}, z_{1:k-1}, u_{1:k}\right)$.
- *Observation*: compute an updated estimate of the state from predictions and observations: $p\left(z_k \mid x_k\right) = p\left(x_k \mid x_{1:k-1}, z_{1:k-1}, u_{1:k}\right)$. For each propagated particle, the likelihood $p\left(x_k \mid z_k\right) \propto p\left(z_k \mid x_k\right) p\left(x_{k-1}\right)$ is measured.
- *Resampling*: The particles are resampled to avoid loosing diversity and propagated to a new state distribution given the observations collected across time.

This process of prediction, observation and resampling repeats itself for as many iterations as needed.

Experimental results

In our evaluation we use the same datasets used in the experiments detailed in (Montes et al. 2020). We use the OES detector to find all broccoli heads in the first frame and create initial trajectories accordingly. We use a standard approach in which the broccoli head state x is modelled as a set of weighted particles representing a distribution of possible locations in the frame. This simple model will suffice as broccoli head orientations with respect to the sensor remains constant. The initial set of particles is at the centre of each broccoli head the first time it is detected. The filter propagates particles from one frame to the next using a motion model. We use a standard dynamical model defined as: $x_k = x_{k-1} + \Delta * v_k + \varepsilon$; where Δ is a constant of motion increment, $v_k = v_{k-1} + \epsilon$ is the velocity and ε and ϵ are added Gaussian noise. The state x_k can then be updated based on new acquired observations provided by the broccoli detector. Some particles are selected or *filtered* by assigning them a weight based on its likelihood of predicting the new state correctly. Particle likelihoods are computed using an appearance model based on a Viewpoint Feature Histogram (VFH). A VFH is a 3D feature descriptor that uses normal vector angles to represent the properties of data points (Rusu et al. 2010). We use the Chi-square similarity coefficient between the predicted state and the observed histograms to compute a particle's likelihood. The likelihoods are then normalized and treated as weights. The algorithm produces a new set of particles by resampling from the current set with probabilities proportional to their weights.

Track confirmation. When the detector's response is accurate, it can guide the tracker fairly well using simple data association policies. Unfortunately, detectors are not fully reliable and a trade-off between true positive and false positive rates is common. However, increasing true detections also increases false positive rates. When false detections occur, simple rules often misguide the tracker and perform poorly. This problem can be alleviated by introducing a confirmation step that performs data association and handles missing detections: On each frame, the detector is first used to confirm the prediction result for each track. Then, tracks that have not been confirmed for a number of frames are eliminated. In addition, any trajectory near the border is also eliminated to avoid predictions outside the current frame. This simple process reduces false positive trajectories without loosing tracks of broccoli not detected for short periods of time.

Evaluation metrics. Location and area size are the main parameters associated with correctly detecting and tracking each broccoli head. A correct state estimation matches at least a 0.5 Intersection-over-Union value according to annotated data. In our evaluation, the common metrics Precision, Recall and F1 suffice to evaluate performance, as Precision is the ratio of correct predictions to the total broccoli predictions made by the system, and Recall is the ratio of correct predictions to the total number of actual broccoli labelled as such. F1 Score is the weighted average of Precision and Recall. Consequently, it takes both false and missing predictions into account and is usually more useful when there is an unbalanced class distribution in the datasets. The system is evaluated with all training and testing combinations of annotated datasets collected in two different runs in UK farms: UK1 and UK2, and one dataset from a farm in Spain: SP1. For experiments on the same set, we define a training and testing split of a 75% to 25% ratio. In any other case, 100% of one dataset was used for training and 100% of the other set was used for testing. Table 1 summarizes a comparative list of evaluation results when using only the OES broccoli detector and the proposed detect-and-track framework. The scores for each dataset combination indicate an improved performance from detection results. In these experiments, we achieve an increment in the F1 score when the particle filter is added to the detection pipeline. Figure 1 shows selected examples of both detection and tracking results.

Running times. A particle filter performance is satisfactory only when the set of particles is sufficiently large to represent the state distributions that are being estimated. However, the number of particles have an impact in the running time of the particle filter. In our evaluation, a small set of 100 particles per track was large enough to improve the performance score of the system and only added a few milliseconds to the process, still allowing the real time execution reported in (Montes et al. 2020) using the same computing hardware. A gradual increment in the number of particles also increased the running times, but had very little impact on the evaluation metrics.

Conclusions

We have presented a framework for detecting and tracking crops of broccoli heads in sequences of 3D point cloud frames. Our system utilises a real-time detector, feature vector histograms to model similarity appearance, and a simple track confirmation technique to keep track accuracy through fail detections. The results demonstrate that the system is capable of reliably detecting and tracking multiple instances of broccoli heads in sequences of 3D frames, as well as improving accuracy by reducing false predictions while preserving high detection rates. Our results indicate a consistent improvement of the F1 metric in all datasets combinations used for testing. On a modern CPU we were able to efficiently tracking broccoli locations, thus enabling a fast execution for other real time detection and tracking applications. Future work will include further evaluation of the proposed system by integrating other detectors and extending to recent structure from motion approaches. Our evaluation shows that the system exhibits the required detect-and-track accuracy and real-time performance needed for autonomous selective harvesting applications. However, the generic nature of the framework makes it applicable to a wide range of other tasks in agriculture.

References

Bender, A.; Whelan, B.; and Sukkarieh, S. 2020. A high-resolution, multimodal data set for agricultural robotics: A Ladybird's-eye view of Brassica. *Journal of Field Robotics*, 37(1): 73–96.

Blok, P. M.; Barth, R.; and van den Berg, W. 2016. Machine vision for a selective broccoli harvesting robot. In *5th IFAC AGRICONTROL Conference*, 66–71.

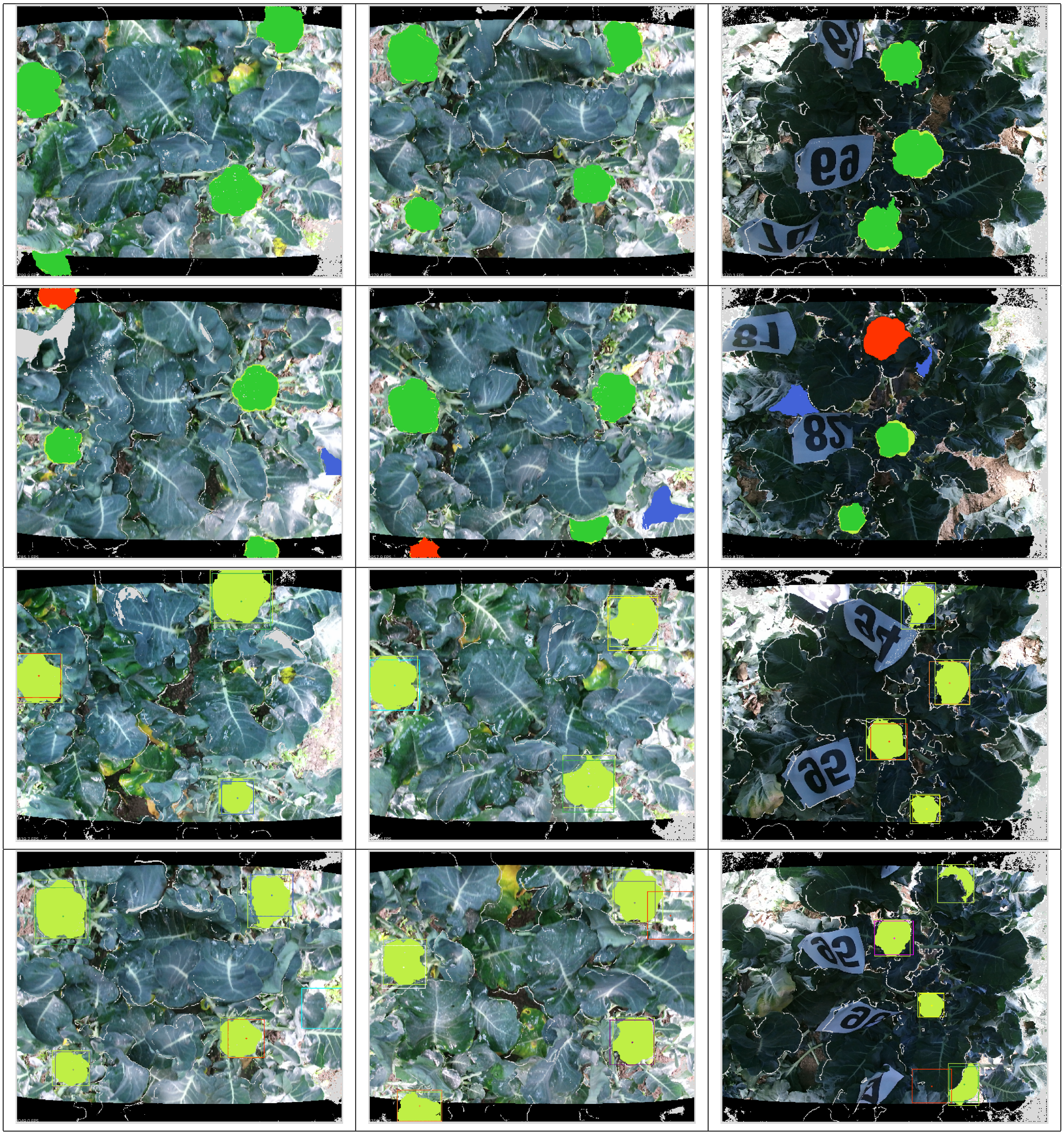

Figure 1: The top row show a selection of frame examples for the detector showing accurate results. The second row shows faulty detections. In green are shown the true positive detection, in red the false negatives (misdetections), and in blue the false positives. The third and bottom rows show a selection of frame snapshots for the tracking results. The Particle filter keeps a tracker shown as a bounding box for each broccoli head returned by the detector. This frames were selected from the UK dataset (two left columns) and from the Spain dataset (right column).

		Train set								
		UK1			UK2			SP1		
		Pr	Rc	f1	Pr	Rc	f1	Pr	Rc	f1
Test set UK1	Det	98.5	96.6	97.5	97.9	92.3	95.0	96.4	88.7	92.4
		256	4	9	980	21	82	942	35	120
	PF	98.1	98.5	98.3	96.2	96.4	96.3	93.3	95.2	94.2
		261	5	4	1024	41	38	1011	73	51
Test set UK2	Det	97.0	91.6	94.2	97.2	95.7	96.4	88.9	84.3	86.5
		947	29	87	228	7	11	916	59	118
	PF	95.5	95.9	95.7	96.3	98.7	97.5	92.2	93.9	93.1
		992	47	42	236	9	3	971	82	63
Test set SP1	Det	97.6	85.1	90.9	98.9	83.9	90.8	96.1	91.7	93.8
		820	20	144	809	9	155	221	9	20
	PF	94.5	94.3	94.4	97.8	92.6	95.2	94.3	96.3	95.3
		909	53	55	893	20	71	232	14	9

Table 1: Detection and Particle Filter tracking results for multiple train and test combinations of the broccoli datasets when using the OES detector (Det) and when the Particle Filter (PF) is added for tracking. The values shown in the first cell row are the *Precision* (Pr), *Recall* (Rc) and *F1* evaluation metrics, while the values shown in the bottom cell row are the *True Positives*, *False Positives* and *False Negatives*, respectively.

Blok, P. M.; van Evert, F. K.; Tielen, A. P. M.; van Henten, E. J.; and Kootstra, G. 2020. The effect of data augmentation and network simplification on the image-based detection of broccoli heads with Mask R-CNN. *Journal of Field Robotics*, 38(1): 85–104.

Blok, P. M.; van Henten, E. J.; van Evert, F. K.; and Kootstra, G. 2021. Image-based size estimation of broccoli heads under varying degrees of occlusion. *Biosystems Engineering*, 208: 213–233.

Choi, C.; Trevor, A. J.; and Christensen, H. I. 2013. RGB-D edge detection and edge-based registration. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1568–1575.

Duckett, T.; Pearson, S.; Blackmore, S.; Grieve, B.; et al. 2018. Agricultural robotics: The future of robotic rgriculture. *UK-RAS Network White Papers. arXiv:1806.06762*.

Elfring, J.; Torta, E.; and van de Molengraft, R. 2021. Particle filters A hands-on tutorial. *Sensors*, 21(438): 1–28.

García-Manso, A.; Gallardo-Caballero, R.; García-Orellana, C. J.; González-Velasco, H. M.; and Macías-Macías, M. 2021. Towards selective and automatic harvesting of broccoli for agri-food industry. *Computers and Electronics in Agriculture*, 188: 106263.

Kootstra, G.; Bender, A.; Perez, T.; and van Henten, E. J. 2020. *Robotics in Agriculture*, 1–19. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-642-41610-1.

Kootstra, G.; Wang, X.; Blok, P. M.; Hemming, J.; and van Henten, E. 2021. Selective harvesting robotics: current research, trends, and future directions. *Current Robotics Reports*, 2: 95–104.

Kusumam, K.; Krajník, T.; Pearson, S.; Duckett, T.; and Cielniak, G. 2017. 3D-vision based detection, localization, and sizing of broccoli heads in the field. *Journal of Field Robotics*, 34(8): 1505–1518.

Louedec, J. L.; Montes, H. A.; Duckett, T.; and Cielniak, G. 2020. Segmentation and detection from organised 3D point clouds A case study in broccoli head detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 285–293.

Montes, H. A.; Le Louedec, J.; Cielniak, G.; and Duckett, T. 2020. Real-time detection of broccoli crops in 3D point clouds for autonomous robotic harvesting. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 10483–10488.

Ramirez, R. A. 2006. *Computer vision based analysis of broccoli for application in a selective autonomous harvester*. Msc thesis, Virginia Polytechnic Institute and State University.

Rusu, R. B.; Bradski, G.; Thibaux, R.; and Hsu, J. 2010. Fast 3D recognition and pose using the viewpoint feature histogram. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2155–2162.

Santos, T. T.; de Souza, L. L.; dos Santos, A. A.; and Avila, S. 2020. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. *Computers and Electronics in Agriculture*, 170: 105247.

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

Yang, B.; and Xu, Y. 2021. Applications of deep-learning approaches in horticultural research: a review. *Horticulture Research*, 8(123): 1–31.

Zhou, C.; Hu, J.; Xu, Z.; Yue, J.; Ye, H.; and Yang, G. 2020. A monitoring system for the segmentation and grading of broccoli head based on deep learning and neural networks. *Frontiers in Plant Science*, 11: 402.

Zhu, L.; Li, Z.; Li, C.; Wu, J.; and Yue, J. 2018. High performance vegetable classification from images based on AlexNet deep learning model. *International Journal of Agricultural and Biological Engineering*, 11(4): 217–223.