

# Evaluation of Deep Neural Network and Alternating Decision Tree for Kiwifruit Detection

Ye Chow Kuang<sup>1,2</sup>, Lee Streeter<sup>1</sup>, Michael J. Cree<sup>1</sup>

<sup>1</sup>School of Engineering, University of Waikato, New Zealand

Melanie Po-Leen Ooi<sup>2,3,4</sup>

<sup>2</sup>School of Engineering, Monash University

<sup>3</sup>Engineering & Physical Sciences, Heriot-Watt University

<sup>4</sup>Engineering & Applied Technology, Unitec New Zealand

**Abstract:** Robotic kiwifruit harvesting systems are currently being introduced to improve the reliability and farming yields of kiwifruit harvesting operations. Machine learning is widely used to carry out the visual detection tasks required of such systems. This paper specifically compares two types of machine learning algorithms: the multivariate alternating decision tree and deep learning based kiwifruit classifiers. The purpose of the study is to investigate the cost of implementation against the classification performance. Thus, discussion is centred around computational cost and its impacts on the overall system architecture. We found that the traditional decision tree classifiers can achieve comparable classification performance at a fraction of the cost and complexity, providing robust and cost-effective instrument design.

**Keywords:** *classification, decision tree, deep learning, precision agriculture, vision*

## I. INTRODUCTION

Kiwifruit is New Zealand's largest horticultural export with an income of NZ\$2.23 billion between 2016-2017. This number is projected to double by 2025 [1]. With such a high economic value, New Zealand farmers are invested in precision agricultural technologies for targeted harvesting of ripe kiwifruit to maximize their profitability. In particular, agricultural industry is facing a growing shortage of manual labour, thus creating keen commercial interest in automated inspection and robotic harvesting systems [2]. Pattern recognition algorithms are proving performant for automatic recognition of kiwifruit. One of the most successful techniques is the Deep Neural Network (DNN), especially those based on the Convolution Neural Network (CNN) construct, offering some of the highest accuracies in classification tasks. While DNN is indeed capable of providing more accurate detection solutions, it does come at a high cost of implementation and lack of comprehensibility in the classification model (as elaborated in Table 1). Comprehensibility is an important aspect of classification when the system is targeted for visual detection applications. For example, in a medical application [3], a comprehensible model not only helped to automatically detect a condition, but it also provided medical experts with insights on the relationship between disease phenotype and affliction.

Decision trees are among the most popular comprehensible machine learning models, with the multivariate alternating decision tree [14] exhibiting good classification performance while also providing transparency. Table 1 presents the comparison of DNN and ADTree. This paper extensively compares the use of the DNN and ADTree specifically on kiwifruit detection. The reason for this particular comparison is

to extend the discussion beyond just accuracy and speed, typical in many detection studies. Specifically, we also consider the benefit of classification insights easily obtained from transparent machine learning models. It is anticipated that a comprehensible system could lead engineers to either (1) choose better measurement techniques that not only perform detection but can also withstand the rigors of the farming environment; or (2) to improve the design of kiwifruit detection instrumentation that maximizes detection rates while keeping down the cost.

Table 1: Comparisons between Deep Neural Network and Alternating Decision Trees

Performance Indicator	DNN	ADTree
Accuracy	Very high if appropriate architecture has been adopted and tuned appropriately	Very high if good classification features are available
Computational requirement	Extremely high. Impractical without GPU assisted computation	Low. Can be executed smoothly on a normal desktop machine with 16GB RAM
Comprehensibility	Current research focus. No generally accepted and practical comprehension framework yet exists.	Interpretable model [14]. Focused discussion provided in Section IV of this paper
Ease of use	Very easy. Seamless integration between automatic feature learning and classification.	Moderate. Automatic feature and tree structure selection. Requires handcrafted features to be supplied.

## II. VISUAL OBJECT DETECTION

Visual detection of an object from a given image can be generally divided into two tasks: (a) localisation and (b) classification. Localisation aims to estimate the probable location of the object. Classification, on the other hand, assesses the category of the localised object. Depending on the chosen method, localisation and classification could either be performed independently of each other, or share a large portion of the information processing pipeline. In the context of this paper, the scope of discussion is restricted to classification.

### A. Deep Neural Network (DNN)

#### 1) The rise of DNN in machine vision (in brief)

Since 2017, Deep Neural Network, specifically those based on Convolutional Neural Network (CNN), have become the de

facto reference point for machine intelligence. The sudden paradigm shift towards DNN is often attributed to its spectacular success during the ImageNet Large Scale Visual Recognition Competition [4]. The ImageNet challenges require the object detection model to correctly identify about a thousand object classes from more than millions of images with wide ranging degrees of difficulties. Below is a very succinct overview of the DNN development in machine vision; how it captured the attention of machine vision researchers and subsequently caused the paradigm shift of the entire field in the span of only a few years.

Alexnet [5], an 8-layer network which was considered one of the deepest at that time, won the 2012 ImageNet competition. Two years later, another two methods were introduced: DNN VGGNet [6] (19 layers) and GoogleNet [7] (22 layers, new inception architecture). Both these models are much deeper than anything that came before. More importantly, they offer almost human-level detection performance on the ImageNet dataset and therefore represented a triumph of the modern computational technology. In 2015, a 152-layer ResNet [8] set the milestone in machine vision technology by surpassing human performance on the ImageNet dataset.

These remarkable achievements have turned DNN into the de facto reference point of visual processing tasks.

## 2) Some challenges of DNN adoption in practice

The success of DNN along with the open availability of good development tools have prompted many engineers and data scientists to apply DNN solutions in their domain as the first option. The biggest advantage of DNN-based solutions is that a highly accurate black-box solution can be found without the need to learn about image feature extraction and various statistical learning algorithms as long as the appropriate development tools and computational infrastructure are used.

There are several considerations that have impeded the wider adoption of DNN. The first consideration is data availability. DNN requires large amount of data to constrain the model. This is because over-learning will occur in a model that has large degrees of freedom unless the learning is regularized by using an even larger amount of training data in order to sufficiently estimate all the free parameters. (Table 2)

Table 2: Approximate number of free parameters and typical number of arithmetic operations to process one image

Architecture	Parameters	Operations
Alexnet	$60 \times 10^6$	$3 \times 10^9$
GoogleNet	$4 \times 10^6$	$3 \times 10^9$
VGG19	$138 \times 10^6$	$40 \times 10^9$
ResNet	$60 \times 10^6$	$24 \times 10^9$

There are two direct consequences to this data requirement. First, the required amount of data is simply not available in most applications. Second, even if the data is available, the task of large-scale training-sample labelling is a non-trivial undertaking. This mission requires funding, together with meticulous resource planning. Transfer learning [9] partially addresses this data availability problem. It works by keeping the localisation layers obtained from large-data training (which can be achieved using Alexnet, for example). Then the user only has to tune the classification layer according to the

specifications of their application. This therefore allows detection to be achieved with a much smaller sample size. Transfer learning essentially transforms the posterior feature extraction model obtained from large-data training into the prior of the small-data problems. It works as long as the domain of the new problem does not differ too greatly from the training data. However, the overlap in the problem domain is sometimes difficult to ascertain objectively. In the case of kiwifruit classification, it is unclear from the outset that a DNN such as Alexnet, that was trained on more than a million images to classify 1000 categories of objects with cluttered background, provides a good prior model to the much simpler task of detecting kiwifruit in a mostly leafy background.

Furthermore, the superior performance of DNN in ImageNet may be partly attributed to the availability of high-quality high-resolution images, and the performance degrades significantly on lower quality real-life images [10]. This is not necessarily true with algorithms that rely on carefully engineered image features. Another equally important consideration is the life-cycle cost of DNN-based systems. Table 2 shows the typical arithmetic operation count to classify an image. It is a few orders of magnitude higher than the traditional methods. Thus, DNN-based systems generally require the use of dedicated computing infrastructure such as *Graphics Processing Unit* (GPU) enabled processors. In summary, designing a DNN-based system requires extensive data labelling, combined with largescale GPU-grade computing infrastructure, training and tweaking.

## B. Feature extraction & alternating decision tree classification

If the intent is solely to classify objects, black-box classifiers such as DNN and support vector machines may be sufficient. However particularly for visual detection systems, it is often beneficial to extract knowledge on how the classification was made in order to improve the human understanding of an object's optical properties. Decision trees are inherently transparent, and are often a good choice for applications such as visual kiwifruit detection.

### 1) Image feature extraction

Given a particular data set, DNNs automatically identify what to learn and how to learn it. Their ability to abstract features from the given input-output data pair is a wonder of modern computation. Unfortunately, to date, there is no reliable methodology to extract the abstracted image features from DNN except through running the entire DNN. As a comparison, traditional image features are used herein by transforming the input image into several representations that each highlight different aspects of the image characteristics. Each representation is then passed through standard feature extraction modules such as sub-block statistics, inter-representation statistics [11], *local binary pattern* (LBP) [12] and *histogram of oriented gradients* (HOG) [13].

### 2) Alternating decision tree classifier

For the reasons stated above, we chose a tree-based classification algorithm. Specifically, we chose the multivariate

ADTree [14] as the decision tree classifier. Decision trees use an information organization structure that is amenable to human interpretation. The multivariate ADTree introduced by Sok et. al. [14] is a type of decision tree that has a performance that is comparable to other types of classifiers such as support vector machines while maintaining comprehensibility. It does not require fine tuning of hyper-parameters because it reduces the classification error through statistical boosting. Fig. 1 is a graphical representation of the decision tree and alternating decision tree respectively.

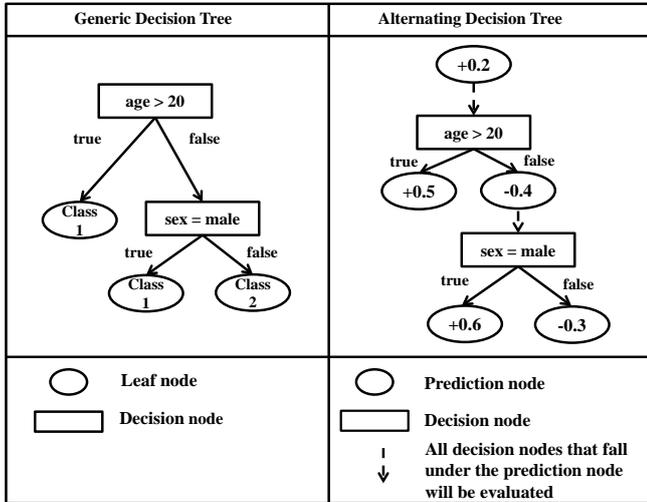


Fig. 1: Comparison between general decision tree and ADTree. For ADTree, it can have more than one decision node under the prediction node to support boosting

The ADTree structure alternates between layers of prediction nodes and decision nodes starting from a root prediction node. Unlike the decision tree whereby prediction nodes hold the predicted classification result, the ADTree prediction node contains real-valued prediction for the decision test outcome. Instead of a single path from the root node to reach one of the leaf nodes that contains the prediction on the class label for general decision tree, ADTree enables one or more paths to be traversed from the root prediction node to the terminal prediction node(s). The distinction between this and the “normal” decision tree is that the decision tree only allows one path to be transversed. The final ADTree prediction is made by summing all the traversed prediction values. A positive sum indicates positive class and vice versa.

This model is completely transparent to the user, whereby the decision rule that contained in the decision node (from Fig. 1) is an explicit linear model that allows the user to gain an understanding on how the features interact to arrive at the classification decision.

### III. EXPERIMENTAL RESULTS

#### A. Methodology and Images

In the following experiments, two DNN architectures, Alexnet and GoogleNet, were chosen as the benchmark. Alexnet was chosen because it is the simplest architecture while GoogleNet was chosen due to its relatively high efficiency among the deeper and more powerful DNN. During our

experiments, we found that GoogleNet did not perform significantly better than Alexnet, thus it was concluded that there was no need to evaluate even deeper models such as ResNet.

Over 100 colour images of kiwifruit in the orchard were taken with the image resolution of 1920×1200 pixels, and under various weather conditions typically encountered by the fruit harvesters. From these images, over 4,715 samples of kiwifruits were hand-annotated and another 37,720 background samples were extracted randomly from the images.

Fig. 2 shows an example of the images collected. The labelled boxes are examples of conditions whereby accurate detection of the kiwifruits can become challenging, e.g. colour distortion, occlusion and objects at the boundary. Compared to an earlier investigation [15], this dataset presents a more realistic variation of lighting typically encountered in real-life. Wijethunga et. al. [15] investigated a purely image processing technique and detection using hand-crafted colour statistics to detect kiwifruits.

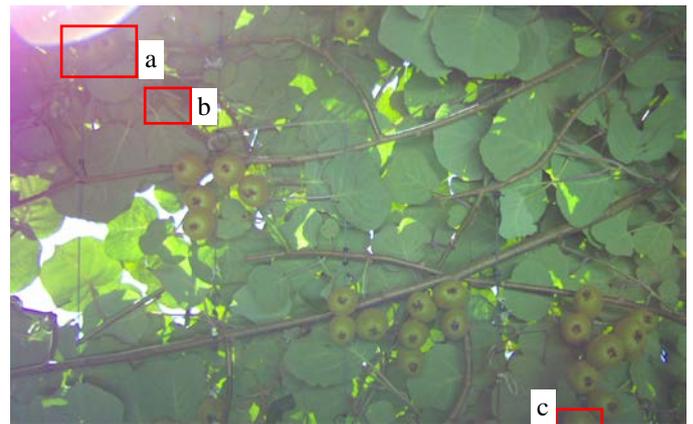


Fig. 2: An example of the image used in this investigation. Some of the challenging scenario are: (a) colour distortion due to harsh lighting; (b) incomplete fruit, occlusion by other objects; (c) incomplete fruit, boundary cut-off

We find that colour statistics alone are insufficient for reliable discrimination between the subtle differences in the shades of green and brown of the fruit, branches and leaves. Furthermore, the colour analysis tended to be distorted by the lighting condition. Hence additional discriminating features via analysis of the available colour and texture was included to increase the detection robustness.

#### B. Classification Accuracy

The proposed method for colour and texture analysis and classification is tested on the kiwifruit classification problem using eight-fold cross validation. The cross-validation results in Fig. 3 indicate that all three classifiers have comparable classification performance. All classifiers can be characterised by precision-sensitivity tradeoff[16]. The results in Fig. 3 show that ADTree has a slightly higher sensitivity at the expense of lower precision compared to the tested DNNs. That means ADTree is slightly better than the tested DNNs in reducing false negatives but produces a slight increase in false positives. A similar trend can be observed on the recognition of the

background. Fig. 4 shows the precision-sensitivity trade-off curve of the best and median models for ADTree, Alexnet and GoogleNet taken from the eight-fold cross-validation results. The markers indicate the default operating point for each of classifier chosen by the learning algorithm. From this figure, it can be seen that GoogleNet performs better than Alexnet for the kiwifruit images in this experiment. It also shows that ADTree is tuned to higher sensitivity at the expense of precision while DNN does the opposite.

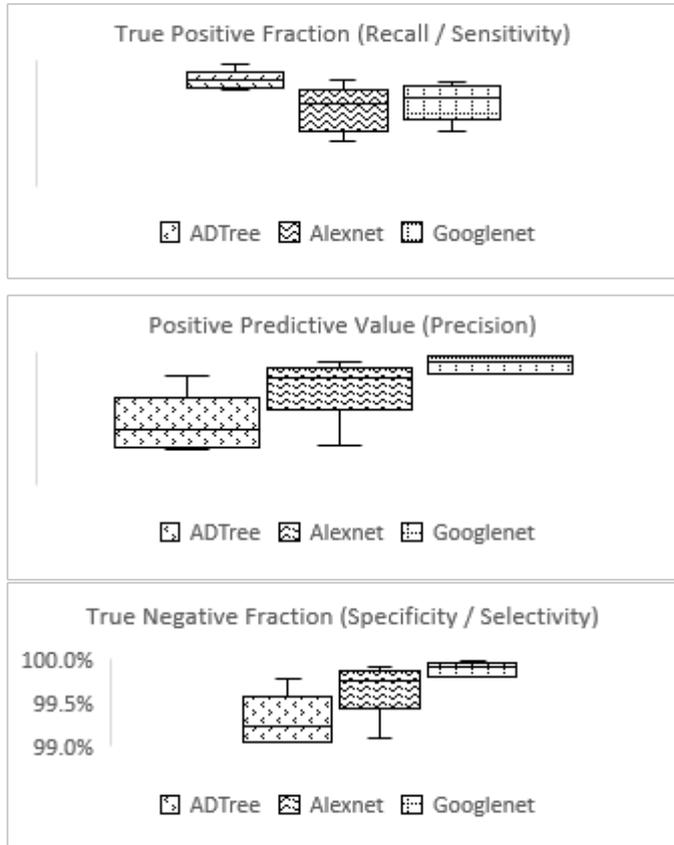


Fig. 3: Classification performance, true positive fraction: fraction of positive samples classified as positive; positive predictive value: fraction of true positive from the samples classified as positive. The horizontal lines in the box and whisker plot are respectively: maximum, 3<sup>rd</sup> quartile, median, 1<sup>st</sup> quartile, minimum.

The difference in the precision-sensitivity tradeoff between DNN and a simpler model like ADTree seen in Fig. 4 can be explained by the differences in the training method and samples. Models like ADTree are determined completely by the samples provided, therefore the quality of the training sample has a significant effect on the model’s performance. On the other hand, the DNN that uses transfer learning have their feature extraction layers fixed. The fine-tuning of classification layer has relatively little impact on the overall performance.

The higher sensitivity of the DNN suggests that its feature extraction layers have learned features that can identify miscellaneous background objects from the ImageNet dataset. This prior knowledge was carried over by the DNN into this study, and could help with the identification of non-target objects hence leading to lower false positive fraction. On the

other hand, the ADTree model was built solely on the provided 100+ images. This limited its capability of building a reliable statistical model for the background objects because some objects only appear once or twice in the entire training set. This shortcoming in the statistical model could be overcome by using a more sophisticated cross-domain transfer learning paradigm [17], which would be a topic of interest for further extension of this research.

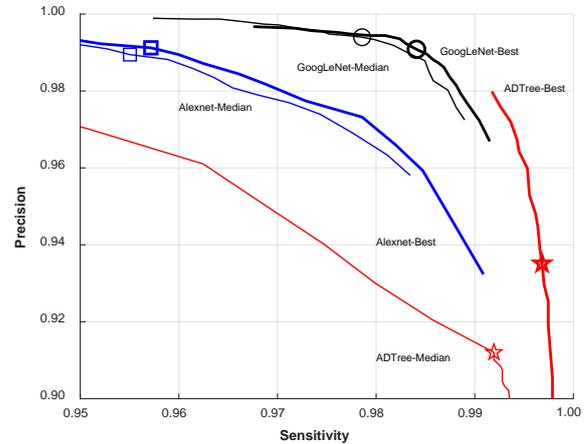


Fig. 4: Precision-Sensitivity(Recall) trade-off between different classification models. Only the best and the median performers from the 8 models undergoing cross-validation test are presented in this graph.

### C. Processing Speed

The time taken to evaluate each classifier model is an important performance metric and directly impacts the feasibility of implementing the classification technique in real-time, out in the farming environment. Parallelism and the use of GPU can significantly reduce the computational time if correctly implemented. Standard DNN implementations make use of single precision floating-point arithmetic, GPU and parallel processing to reduce the evaluation time. action between parallel threads causes nonlinear scaling of computational speed with respect to the number of available computational cores. This makes the direct comparison between different algorithms almost impossible. Therefore, for the purpose of this discussion, all GPU and parallel processing was disabled.

Table 3: The average execution time per image per CPU core  $\pm 2$  standard deviation ( $\approx 95\%$  interval). “Single pass” column refers to approximate time required to evaluate all 42435 image samples

Classifier	Average time	Single pass
<i>ADTree-Optimised</i>	0.064 $\pm$ 0.003s	0.7 hour
<i>ADTree-Unoptimised</i>	0.091 $\pm$ 0.009s	1.0 hour
AlexNet	0.245 $\pm$ 0.001s	2.9 hours
GoogleNet	0.624 $\pm$ 0.001s	7.3 hours

The execution time statistics presented in Table 3 are obtained using single CPU core only. *ADTree-Unoptimised* in Table 3 refers to an unoptimised ADTree that generates an ADtree based on all the presented 8,850 image features using double precision arithmetic; whereas *ADTree-Optimised* represents the same ADTree using single precision arithmetic and part of the unused image features removed via sparsity in

its learning paradigm. Note that for this experiment, there is no optimisation done on standard features like HOG and LBP. Table 3 shows *ADTree-Optimised* can be expected to be an order of magnitude faster than the already optimised DNN.

#### DISCUSSION ON MODEL COMPREHENSIBILITY

##### D. Interpreting the ADTree

Fig. 5 shows the structure of *ADTree-Best*, which is the best performing optimised ADTree of the 8-fold cross validation experiment. The *ADTree-Best* only uses four *decision nodes* (refer to Fig. 5), which are T1-T4. The conditions in the *prediction nodes* (P1-P9) represents different subsets of the training data as partition by the ADTree learning algorithm.

Table 4 shows the number of image features out of the presented 8,850 feature set that each decision node uses in the *ADTree-Best* classifier. In total, *ADTree-Best* uses just 508 features, which is fewer than 6% of the 8,850 image features. Hence a full optimisation could be even faster than that reported by *ADTree-Optimised* that was reported in Table 3. Since the ADtree is a transparent model, we are able to extract the type of feature that the classifier considers as useful for visual detection. By studying Decision node T1, we can identify the 3 features that it used (refer to Table 4). In this case, the features were (1) Grayscale/Brightness (*Feature1*), (2) Colour Likelihood (*Feature2*) and (3) Hue (*Feature3*). This knowledge is useful since further investigation can be performed to study why the ADTree classifier considered these features to be important for Decision node T1.

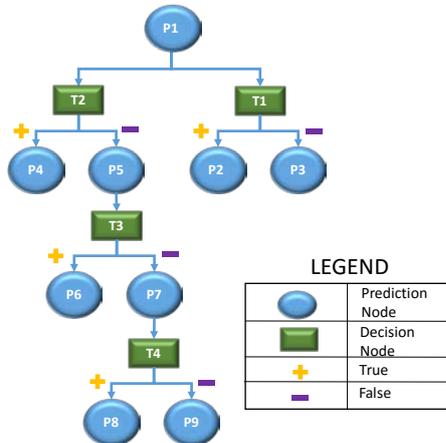


Fig. 5: The tree structure of *ADTree-Best*, the best performing ADTree

Table 4: Feature size is the number of image features with non-zero coefficients, sparsity is the percentage of non-zero coefficient.

Decision nodes	Features used out of the original 8,850 feature set	Sparsity
T1	3	0.03%
T2	312	3.52%
T3	112	1.27%
T4	104	1.17%
Overall	508	5.94%

Following the examination, it was found that T1 could distinguish kiwifruit based on pixel brightness (*Feature1*) in the

kiwifruit-green colour (*Feature2*) and with strong reddish hue (*Feature3*). This is similar to human comprehension, except that the ADTree built the mathematical equation based on the training data. It is obvious, from the above examination, that T1 performs the task of colour analysis. However, when we studied the prediction at P2, we found that it contains some background samples that received a positive score from T1 due to colour distortion caused by strong sunlight. Fig. 6 shows an example of photo that has bright and reddish hue due to glare. Therefore, if we use only colour analysis to distinguish between kiwifruit and background (i.e. what T1 is doing), it will result in a misclassification. This explains why the “purely colour analysis” kiwifruit detector reported in [15] failed to handle this problem adequately.



Fig. 6: An example of colour distortion due to harsh lighting.



Fig. 7: Samples of occluded kiwifruits that are detected by T1 but eventually rejected by the texture analysis branch T2-T4. Most of these samples are being rejected by DNN as non-target.

ADTree supplements T1 with a few texture analysis branches T2-T4 to compensate for the misclassification from T1. T2-T4 branches are added in parallel to T1, and in the ADTree, this means that it follows an independent line of reasoning in the decision making. As discussed in Section 2, the “normal” decision tree only allows one path is to be transversed. However, the ADTree transverses all parallel paths. T2-T4 are dominated by LBP texture features and HOG edge statistics features. The texture analysis, while correcting some colour analysis misclassifications, reduces the recall of occluded fruits as shown in Fig. 7. These occluded fruits have been identified using colour analysis but rejected by texture analysis that is meant to compensate for false positives that is caused by sensor saturation. The quantitative and qualitative evidences show that inherent trade-off between colour analysis and texture analysis exists that cannot be resolved easily through purely computational means. This information is not readily available if the classification is done using a blackbox model like DNN. It highlights an opportunity to reduce the classifier complexity and improve accuracy. An optical filter (e.g. infrared-cut filter) can be added to the camera to reduce the colour distortion. Such reduction in complexity not only reduces acquisition cost, it is likely to reduce the engineering maintenance and operational costs too.

### E. Processors

Real-time human detection technology [13] similar to the ADTree reported above has been known to run at 30 frames per second (FPS) on the Raspberry Pi. On the other hand, the processors that are capable of evaluating DNN are relatively expensive. For example, 5 FPS has been reported on powerful GPU processor like TITAN or K40 which is about 10-30 times more expensive than a typical Raspberry Pi processor. Integrating these GPU-processors into a generic embedded system with maximum clock speed of 1.2-1.5GHz is also challenging. Attempts to run GoogleNet on a generic Raspberry Pi (1.2GHz, quad-core) system has achieved the speed of less than 0.1 FPS [18]. Optimised CNN-DNN design (e.g. SquozedNet, MobileNet) can improve the speed around 10 fold (still less than 1 FPS) at the expense of degrading classification performance [19]. These observations are consistent with the 10-fold difference in speed reported in the previous section.

The speed of 1 FPS that leads to lower classification performance would either cause cases of missed harvesting of kiwifruit, or worse, result in a system that will be too slow to be deployed in the field. Either way, this negatively impacts the harvesting yield and therefore profit margins. Thus, DNN still remains impractical for high volume visual detection work. Apart from needing more expensive processors, the additional information processing capability required to run a DNN increases the power consumption. This leads to larger and more complicated power and thermal management systems which add substantially to the system cost, negatively impact the system mobility and may be incompatible with outdoor operation. On the overall, the slight difference in the classification performance reported above does not seem to justify the adoption of DNN that runs at one-tenth of the speed or cost far more to build.

### IV. CONCLUSION

This paper compares the visual classification performance of kiwifruit images between standard Deep Neural Network (DNN) and a decision tree classifier with independent feature generation and classification algorithm in order to assess the performance and cost of implementation. In the context of kiwifruit detection and harvesting, the results of this paper leads to the recommendation that the adoption of DNN-based object recognition technology will have high accuracy but will require a very powerful backend information processing server cluster. Furthermore, it is expected that high-speed wireless communication is needed in order to achieve near-real time implementation. The system price must include this cost, including the cost of lower harvesting yield per hour and maintenance/servicing of the system in the farming environment.

Alternatively, the adoption of traditional visual recognition models such as decision tree allows for real-time recognition with lower system cost. However, this cost could come at the cost of lower accuracy of detection. The most cost-effective way forward is to improve the decision tree classifier precision through the use of custom-made image features and more

sophisticated training methodologies such as transfer learning. This should therefore be the direction of future research.

### ACKNOWLEDGMENT

The authors would like to thank Matthew Seabright for supplying the data to support this work.

### REFERENCES

- [1] Zespri International Ltd., "ZESPRI Annual Report 2016-17," Tauranga, New Zealand, 2017. [Online]. Available: <http://www.zespri.com/ZespriInvestorPublications/Annual-Report-2016-17.pdf> [Accessed: 2018-09-13]
- [2] D. Picken, "Inside the Bay of Plenty's kiwifruit labour shortage," Tauranga, New Zealand, 2018. [Online]. Available: <https://www.nzherald.co.nz/nz/news/article.cfm?cid=1&objectid=12053443>[Accessed: 2018-09-13]
- [3] K.-Y.K. Liu, J. Lin, X. Zhou, S.T. C.S.Wong, Boosting alternating decision trees modeling of disease trait information, *BMCGenet.6* (Suppl.1) (2005) S132.
- [4] F.Li, O. Russakovsky, Analysis of Large-Scale Visual Recognition, Bay Area Vision Meeting, October, 2013
- [5] A.Krizhevsky, I.Sutskever, G.E.Hinton, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, 1097-1105, 2012
- [6] K.Simonyan, A.Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556v6*, 2014
- [7] C.Szegedy, W.Liu, Y.Jia, P.Sermanet, S.E.Reed, D.Anguelov, D.Erhan, V.Vanhoucke, A.Rabinovich, Going Deeper with Convolutions, *arXiv preprint arXiv:1409.4842v1*, 2014
- [8] K.He, X.Zhang, S.Ren, J.Sun, Deep Residual Learning for Image Recognition, *Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778, 2016
- [9] S.J.Pan, Q.Yang, A survey on transfer learning, *IEEE Trans. on knowledge and data engineering*, 22(10), 1345-1359, 2010
- [10] X.Li, M.Ye, Y.Liu, C.Zhu, Adaptive Deep Convolutional Neural Networks for Scene-Specific Object Detection, *IEEE Transactions on Circuits and Systems for Video Technology*, 2017
- [11] W.K. Tey, Y.C.Kuang, M.P.-L.Ooi, J.J.Khoo, Automated quantification of renal interstitial fibrosis for computer-aided diagnosis: A comprehensive tissue structure segmentation method, *Computer Methods and Programs in Biomedicine* 155, 109-120, 2018
- [12] T.Ojala, M.Pietikainen, T.Maenpaa, Multiresolution Gray Scale and Rotation Invariant Texture Classification With Local Binary Patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 24(7), 971-987, 2002
- [13] N.Dalal, B.Triggs, Histograms of Oriented Gradients for Human Detection, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1, 886-893, 2005
- [14] H.K.Sok, M.P.-L.Ooi, Y.C.Kuang, S.Demidenko, Multivariate Alternating Decision Trees, *Pattern Recognition* 50, 195-209, 2016
- [15] P.Wijethunga, S.Samarasinghe, D.Kulasiri, I.Woodhead, Digital Image Analysis Based Automated Kiwifruit Counting Technique, *23rd International Conference Image and Vision Computing New Zealand. IEEE*, 2008
- [16] T.Saito, M.Rehmsmeier, The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE* 10(3), 2015
- [17] L.Shao, F.Zhu, X.Li, Transfer learning for visual categorization: A survey, *IEEE Trans. on neural networks and learning systems*, 26(5), 1019-1034, 2015
- [18] <https://medium.com/nanonets/how-to-easily-detect-objects-with-deep-learning-on-raspberrypi-225f29635c74>
- [19] A.Wong, NetScore: Towards Universal Metrics for Large-scale Performance Analysis of Deep Neural Networks for Practical On-Device Edge Usage, *arXiv preprint arXiv:1806.05512*