Pertanika J. Sci. & Technol. 30 (2): 1219 - 1236 (2022)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Sago Palm Detection and its Maturity Identification Based on Improved Convolution Neural Network

Zulhakim Wahed[1,2]*, Annie Joseph[1], Hushairi Zen[1] and Kuryati Kipli[1]

[1]*Department of Electrical and Electronics Engineering, Faculty of Engineering, Universiti Malaysia Sarawak, 94300 UNIMAS, Kota Samarahan, Sarawak, Malaysia*
[2]*CRAUN Research Sdn Bhd, 93050, Kuching, Sarawak, Malaysia*

ABSTRACT

Sago palms are mainly cultivated in Sarawak, especially in the Mukah and Betong division, for consumption and export purposes. The starches produced from the sago are mostly for food products such as noodles, traditional food such as *tebaloi,* and animal feeds. Nowadays, the sago palm and its maturity detection are done manually, and it is crucial to ensure the productivity of starch. The existing detection methods are very laborious and time-consuming since the plantation areas are vast. The improved CNN model has been developed in this paper to detect the maturity of the sago palm. The detection is done by using drone photos based on the shape of the sago palm canopy. The model is developed by combining the architecture of three existing CNN models, AlexNet, Xception, and ResNet. The proposed model, CraunNet, gives 85.7% accuracy with 11 minutes of learning time based on five-fold-validation. Meanwhile, the training time of the CraunNet is almost two times faster than the existing models, ResNet and Xception. It shows that the computation cost in the CraunNet is much faster than the established model.

Keywords: Convolution neural network (CNN), deep learning, sago palm

ARTICLE INFO

Article history:
Received: 22 August 2021
Accepted: 21 December 2021
Published: 14 March 2022

DOI: https://doi.org/10.47836/pjst.30.2.20

E-mail addresses:
zulhakimwahed@gmail.com (Zulhakim Wahed)
jannie@unimas.my (Annie Joseph)
zhushair@unimas.my (Hushairi Zen)
kkuryati@unimas.my (Kuryati Kipli)
* Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

Sago (Metroxylon sagu) is an excellent crop for sustainable agriculture, shown in Figure 1. It can grow in underutilised wetlands and peat bogs where other food crops cannot grow economically. It produces high-yield edible starch (about 150–300 kg dry starch per plant). Different parts of palm trees can be used as roofing materials, animal feed, sago worm production, mats, and basket weaving, which will help promote food security, increase household income, and create employment opportunities (Ehara et al., 2018).

It has been reported that 49,243 hectares of land have been planted, whereas the smallholder plants have almost 86% (42,310 hectares) in the Mukah area (Howell, 2017), tabulated in Table 1. However, no exact data shows the number of sago palms planted. In order to ensure starch production from the sago palm, the number of mature palm needs to be identified. There is only manual recognition of mature palm that the harvester has practised, and it takes time since they are dealing with a huge plantation. On the other hand, deep learning has been adopted in many applications currently, such as agriculture, medical purpose, and automobile technology (Khvostikov et al., 2018). It will improve the efficiency and value of the particular application and its productivity. Therefore, detecting the sago tree and identifying its maturity using deep learning is required to estimate the production of the sago palm. This approach is needed with sago palm cultivation that deals with a large area.

Figure 1. Sago palm tree, photo taken at Sg. Talau Research Station, Dalat, Mukah

Table 1
Estimated area planted with sago palm in Sarawak by 2014 (Adopted from Howell, 2017)

Division	**Area (Hectares)**
Kuching	1
Samarahan	0
Sri Aman	45
Betong	4,723
Sarikei	1,432
Sibu	386
Kapit	0
Mukah	42,310
Bintulu	348
Miri	1
Limbang	0
Total	**49,243**

According to the local farmers in Sarawak, the growth stage is divided into nine-stage. Each stage gives its characteristic of the existing trunks and crown shape. The stages are *Sulur, Angkat punggung, Bibang, Plawei, Plawei manit, Bubul, Angau muda, Angau tua,* and *Mugun.* The crown or canopy shape of the palm gives a significant difference starting in the *Bubul* stage. Based on recent research, the starch accumulation starts to slow down as the palm reaches the *Bubul* stage, and the harvesting can be done at this stage (Flach,1997). It shows that the harvestable palm can be done as it reaches the *Bubul* stage onward. As the palm is harvested earlier than the *Bubul* stage, the starch quality and quantity are low, and so much waste could occur. As the palm reaches the *Mugun* stage, the palm is not suitable to be harvested due to the low quality and quantity of the starch. Figure 2 shows the stages of palm according to the local dialect. The photo was taken using DJI Phantom 4 drone vertically from above the palm, illustrating the different shape formations between stages.

The mature sago palm will produce an optimum amount of starch with an optimum starch quality. Flach (1997) stated that the sago palm could be harvested as it reaches the initiation of the flowering stage and onward. It is because the trunk formation will be stopped at this point. Hence there will be no starch produced inside the trunk. According to Flach (1997), the *Bubul* stage and onward is the best stage to be harvested. Figure 2 shows that the *Bubul* has a significant difference in physical shape and characteristics. This factor makes learning easier since it gives a different canopy shape.

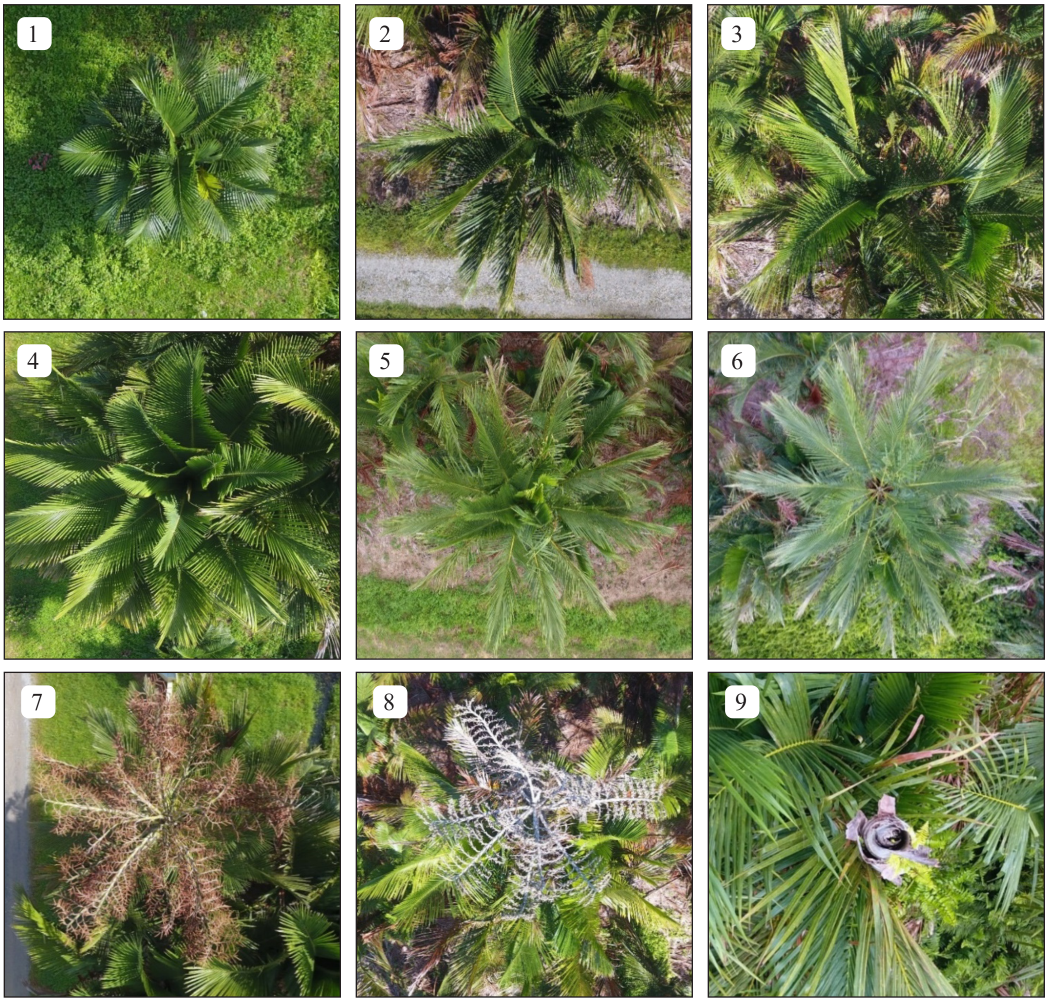

Figure 2. Growth stage of sago palm: (1) *Sulur* (Rosette stage), (2) *Angkat punggung* (Trunk formation started), (3) *Bibang* (Trunk formation continue), (4) *Plawei* (Trunk elongation), (5) *Plawei Manit* (Palm canopy start to change), (6) *Bubul* (canopy shape is significantly changed), (7) *Angau Muda* (Flowering stage), (8) *Angau Tua* (Fruiting stage), and (9) *Mugun* (End life cycle)

Identifying the sago palm and its maturity is usually done manually by observing whether it can be harvested or not. The harvesters need to reach each palm to see whether the palm is ready to be harvested. This work is laborious and time-consuming just to determine the maturity of the palm. A recent study has been done by Hidayat et al. (2018) for sago palm identification by using the deep learning method, in which they work mainly using a single vector machine (SVM). However, Hidayat et al. (2018) do not focus on sago palms' counting and maturity identification.

On the other hand, image processing technology and deep learning are widely used in many industries nowadays. Some research has been done to embarking on this technology, and it can be utilised in many industries such as medical (Samala et al., 2017), face recognition (Lawrence et al., 1997), and agriculture (Yu et al., 2019). There is a large amount of data in agriculture that can be identified using this technology, such as fruit maturity detection (Habaragamuwa et al., 2018), weed detection (Farooq et al., 2019), and oil palm tree detection (Mubin et al., 2019). Therefore, the proposed research focuses on the detection of sago palm and the identification of its maturity. One of the deep learning approaches is Convolution Neural Network (CNN), where several studies have been done, and it yields a good accuracy and performance. Convolution Neural Network (CNN) is one of the most popular deep learning in image processing, especially agriculture. The CNN has been utilised for oil palm detection and counting, which help in farm and plantation monitoring and management (Mubin et al., 2019). Mubin et al. (2019) show that oil palm counting can be done, and it can also be applied to sago palm.

Considering the advantages of CNN, hybrid CNN is proposed in this paper to detect the sago trees and identify their maturity simultaneously. This paper aims to develop the proposed model with a low computational cost. The computational cost is essential in deep learning because it affects the identification time when dealing with big data. Besides that, the computational cost also affects the computer's hardware, especially the Graphical Processor Unit (GPU). Low computational cost in the model leads to low power usage in GPU operation. Hence existing CNN is improved where AlexNet and Xception models are hybrid in this paper. Besides that, ResNet is also involved in the hybrid work. The residual connection from the ResNet was included in the proposed model architecture. ethe standard convolution technique. The depthwise separable convolution has been claimed that have a very low computation cost by Chollet (2017). Hence, this precious feature from Xception is used in AlexNet architecture. The detection process can be faster as we have more concern about the time consumed so that the harvesting activity can be efficient and the starch productivity can be optimum.

Related Works

Flach (1997) found that the growth stage of the sago palm is divided into nine stages. He stated that the sago palm is mature after reaching the sixth stage, which is *Bubul.*

However, no significant research has been done in detecting and determining the maturity of the sago palm except by using manual monitoring. Hidayat et al. (2018) has developed a semi-automatic classification scheme for mapping sago palm trees using Support Vector Machine (SVM) and got the highest overall accuracy of 85%. The limitation of the Hidayat et al. (2018) approach is that they used the satellite photo where the photo's resolution is low, hence affecting the machine learning process. As an alternative, the drone is used to capture the images. The drone can provide a high-resolution image. The images collected are used as a sample for the training and validation phase in CNN.

Convolution Neural Network

CNN is the best representative model of deep learning (LeCun et al., 2015). A typical CNN system structure, such as LeNet, has shown in Figure 3. The feature map of the input layer is a 3-D matrix of pixel intensities for different colour channels (e.g., RGB). The feature map of any internal layer is a sensed multi-channel image, and its "pixels" can be regarded as specific features. Each neuron is connected to a small number of neighbouring neurons in the previous layer (receptive field). The filtering (convolution) operation convolves the filter matrix (the weight of the neuron) with the value of the receptive field of the neuron. It uses a nonlinear function, such as ReLU, to obtain the final response (Krizhevsky et al., 2017).

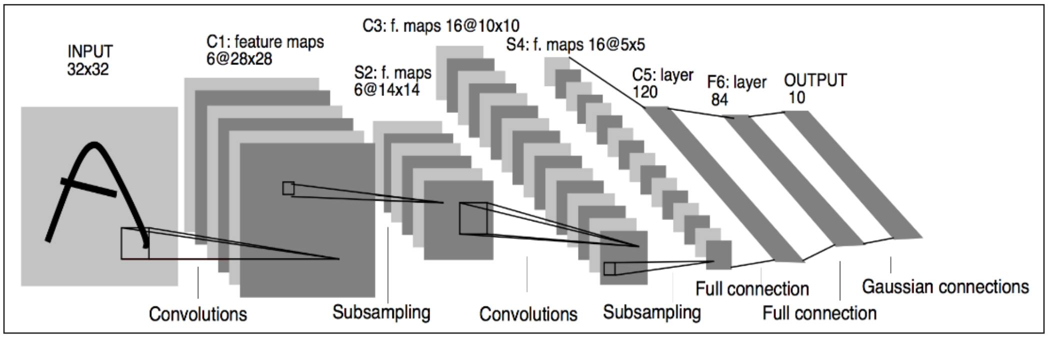

Figure 3. LeNet CNN architecture model, adapter from LeCun et al., 1998

The neural network is presented by Equation 1, which consist of input-output and is completed by the activation function, illustrated in Figure 4. The output entirely depends on the activation function used, such as sigmoid, unit step, hyperbolic tangent function, and rectified linear unit (ReLU), widely used in CNN.

$$y = \sum x_i \omega_i + b \tag{1}$$

Pooling operations, such as maximum and average pooling, L2 merging, and local contrast normalisation, will summarise the responses of the receiving domain into one value

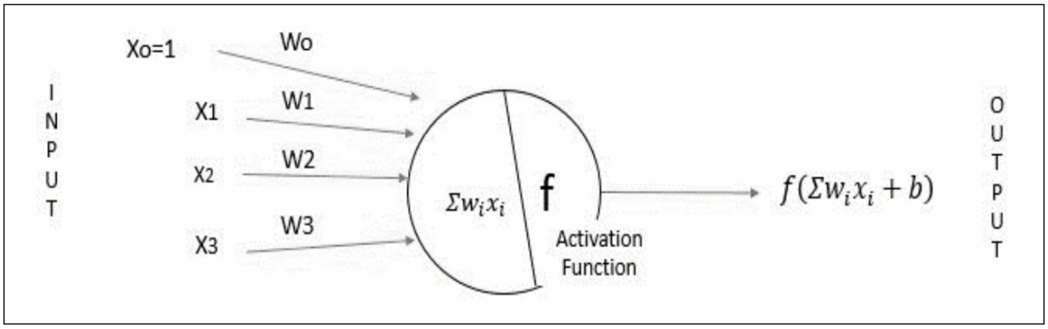

Figure 4. The mathematical model of neural network

which produces a more powerful feature description (Kayukcuoglu et al., 2009). Through the interleave between convolution and pooling, the initial feature level can be constructed and fine-tuned by adding multiple fully connected (FC) layers in a supervised manner to suit different visual tasks. The entire process is mainly performed by the stochastic gradient descent (SGD) method. A typical LeNet has two convolutions layers, two fully connected layers, two average pooling layers, and a softmax classification. Recent research has been done in many applications, especially the agriculture sector, which impact the industry in terms of management operation.

The feature selection or extraction is done by executing convolution by sliding the kernel or filter over the input. For each slide, the matrix multiplication is performed and sums the result onto the feature map. In other words, the convolution process will reduce the input size without losing its feature. Figure 5 summarises the convolution process.

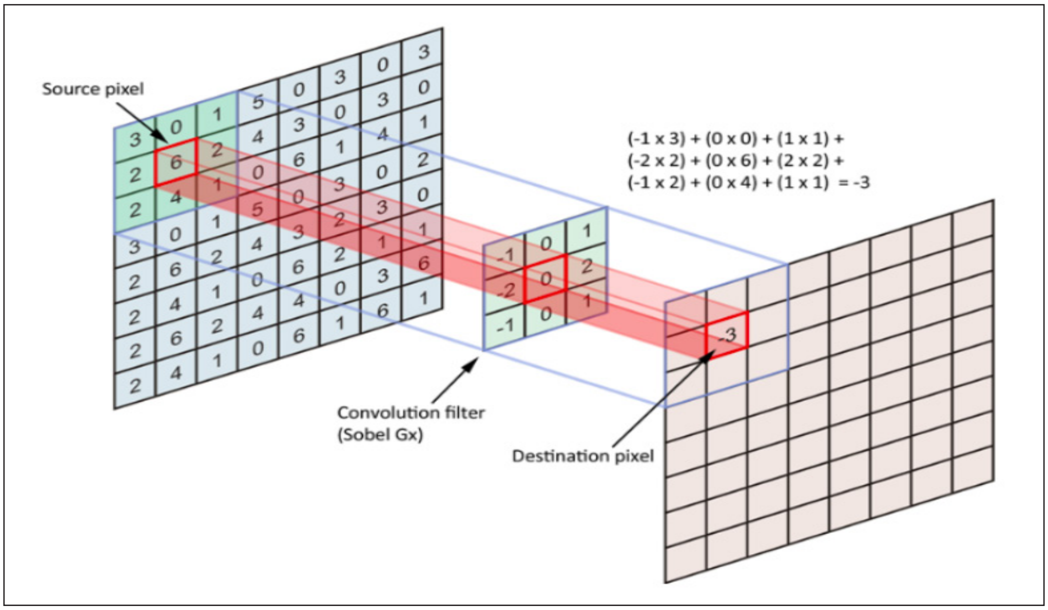

Figure 5. Feature extraction using the convolution operation

AlexNet

AlexNet is a deep convolution neural network developed by Krizhevsky et al. (2012). One of the reasons for this establishment is to enter the LSVRC-2010 competition, and it has been rewarded as the winner of that competition at that time. It has been classified as a deep convolutional network and built to classify a coloured image with a size of (224 x 224 x 3). It has over 62 million trainable parameters within one linear network. There are 11 layers of AlexNet, consisting of five convolution layers, three max-pooling layers, and three fully-connected layers, as shown in Figure 6.

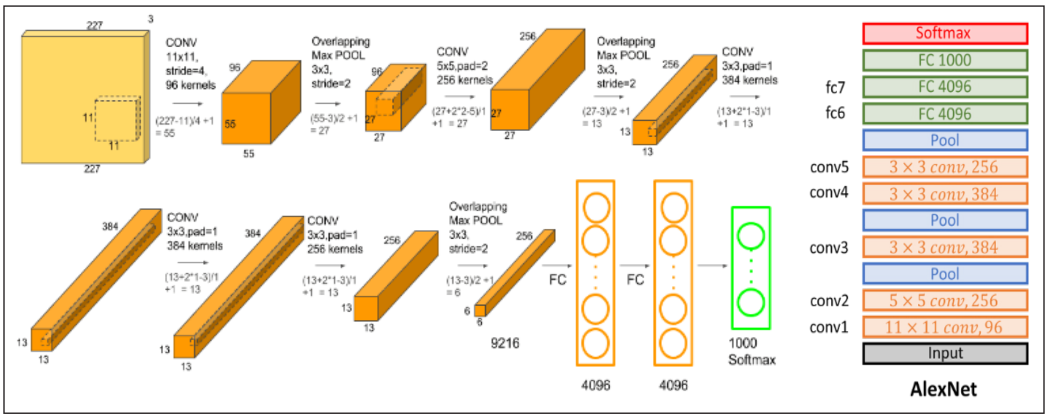

Figure 6. The detailed architecture of AlexNet

Xception

The Xception model was created by Francois Chollet, who was the author of the previous version, the inception model. During the development of this model, depthwise separable convolution has been introduced. This technique is divided into two parts: pointwise convolution and depthwise separable. Pointwise convolution is a 1 x 1 convolution that changes the dimension of the input. The process then continues by depthwise convolution, which is a channel-wise n x n spatial convolution. The depthwise separable convolution is the main part of this work. The author of Xception claimed that it has a cheaper computational cost compared to the standard convolution operation. Table 2 summarises the comparison of depthwise separable convolution and standard convolution operation in terms of the number of multiplication or computational cost.

The whole architecture of Xception is divided into three-part, entry flow, middle flow, and exit flow, as shown in Figure 7. The residual connection claimed by He et al. (2016) the residual connection helps a lot in a model to increase the accuracy and has been proved in the Xception model by Chollet (2017). However, the number of layers and the computational cost of Xception are very high. It might affect the whole identification process when dealing with very big data size.

Table 2
Comparison of depthwise separable convolution and standard convolution

Number of multiplication in Depthwise Separable Convolution	**Number of multiplication in Standard Convolution**
$M * D_G^2(D_k^2 + N)$	$N * D_G^2 * D_k^2 * M$
M = Depth of input volume	M = Depth of input volume
D_G = Input size	D_G = Input size
D_k = Kernal size	D_k = Kernal size
N = Number of filters	N = Number of filters

Comparison

$$\frac{\textbf{Number of multiplication in Depthwise Separable Convolution}}{\textbf{Number of multiplication in Standard Convolution}} = \frac{D_k^2+N}{D_k^2*N} = \frac{1}{N} + \frac{1}{D_k^2}$$

Let $N = 1024,\ D_k = 3$

$$\frac{\textbf{Number of multiplication in Depthwise Separable Convolution}}{\textbf{Number of multiplication in Standard Convolution}} = \frac{1}{1024} + \frac{1}{3^2} = 0.112$$

Therefore, Standard Convolution has nine times more than the number of multiplication as that depthwise separable convolution.

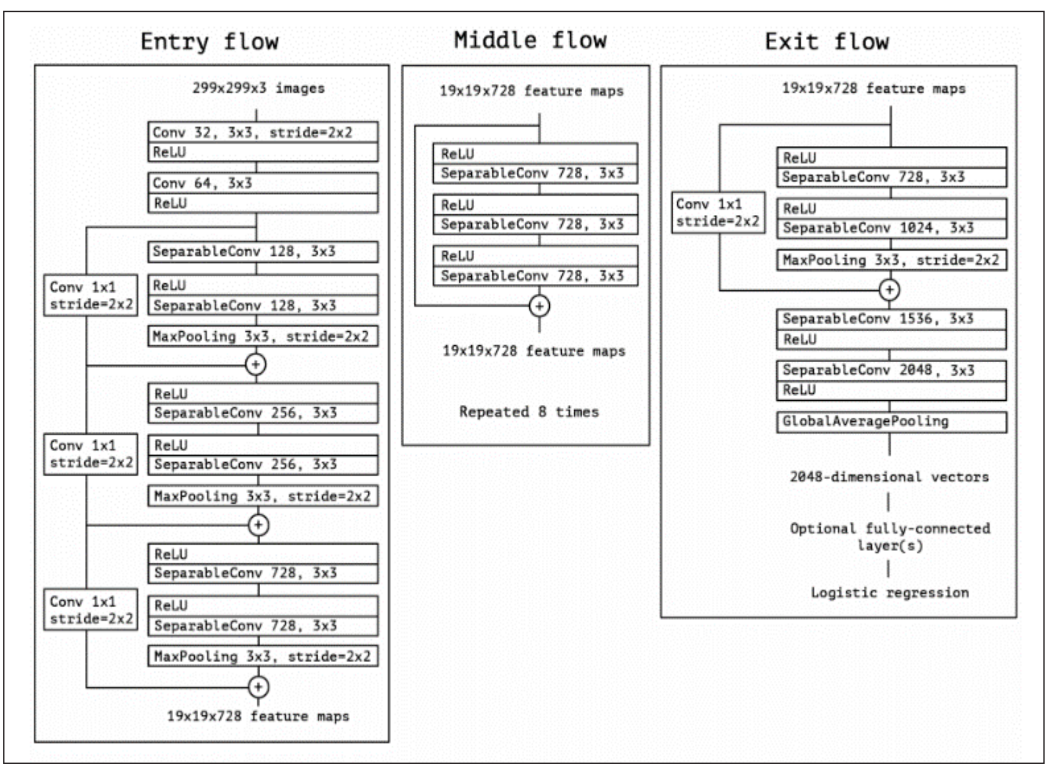

Figure 7. The overall architecture of Xception adapted from Chollet (2017)

Even though AlexNet has a very low number of layers, the accuracy is not quite good. Zhang et al. (2019) yielded 57.14% accuracy in their study. An improvement needs to be done by replacing some layers with several layers introduced by Xception. This improvement is explained in the next section.

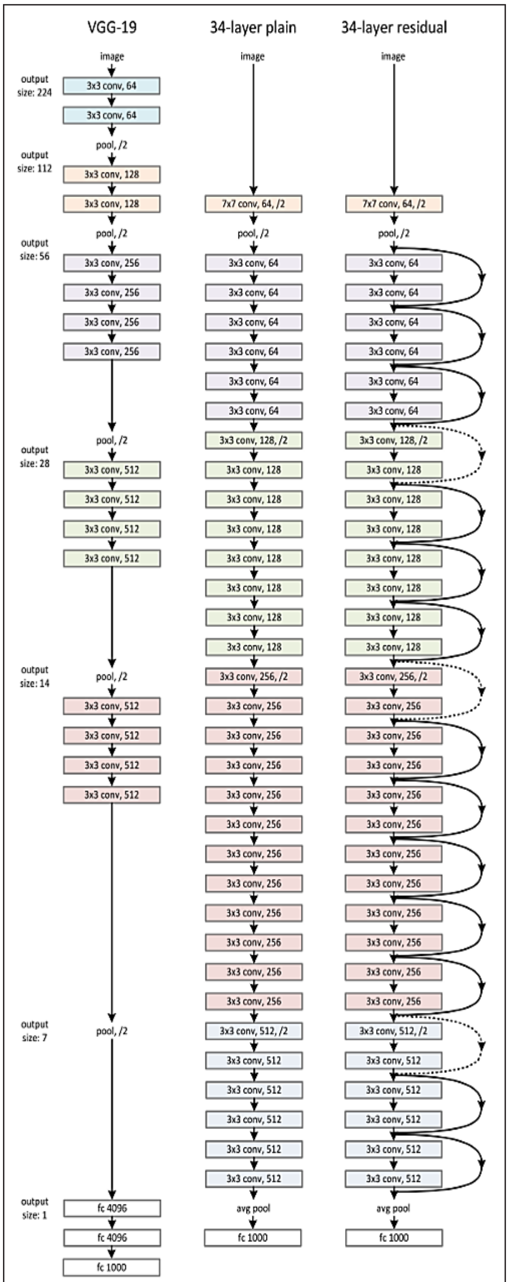

Figure 8. The architecture of the ResNet model

ResNet

The ResNet, shown in Figure 8, was created by He et al. (2016) with an invention of residual connection. This invention helps a lot in improving the accuracy of the model, as they assume that the model's performance degrades as the number of layers increases. This idea is a so-called shortcut connection that skips a few layers, as shown in Equation 2 and Figure 9.

$$F(x) + x \tag{2}$$

This paper focuses on AlexNet, ResNet, and Xception, considering the benefit and advantages of each model. AlexNet has a very low number of layers which speed up the training and detection process. The

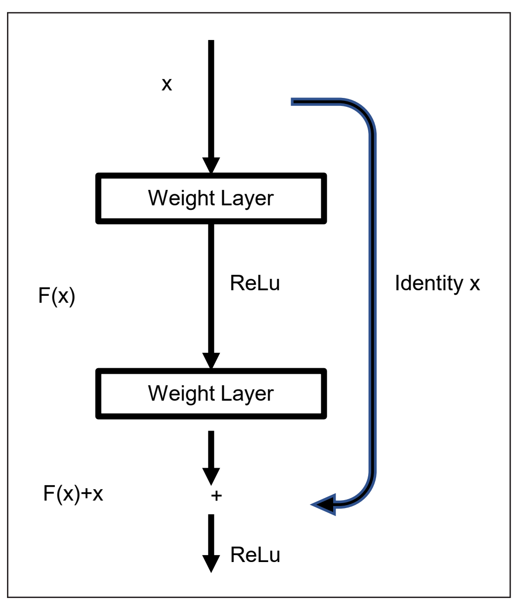

Figure 9. Residual/shortcut connection model

Xception model used depthwise separable convolution instead of standard convolution. Depthwise separable convolution has been claimed by Chollet (2017) to be low in computational cost. Meanwhile, the residual connection from ResNet also helps increase the accuracy of the model.

METHODOLOGY

Study Materials

The great way to detect the sago palm is by taking a photo from above the canopy that can cover a larger area. Hence, the drone/crewless aerial vehicle is utilized in this research. The drone used for this research is DJI Phantom 4, manufactured by Da-Jiang Innovation (DJI, 2016).

In this paper, a high spec computer is used because this work requires heavy operation especially dealing with a high-resolution image. The computer specifications used for this research are an i7-9750H processor, GeForce RTX2060 with 16Gb ram. Several algorithm development platforms have been established, such as C+, Java, Phyton, and Matlab. The Matlab software, version R2020a, is selected to be the main platform in this research because it is user-friendly and can save time developing the coding.

Research Location and Data Collection

The research mostly covers the Mukah area, especially in Sungai Talau Research Station, owned by CRAUN Research Sdn. Bhd. The data collection is mostly held in the Mukah Division covers Dalat, Mukah, and Balingian districts. Figure 10 shows the exact location of the research.

The photos are resized manually by cropping each palm to do the training and validation process. Nine hundred forty-five samples have been prepared for harvestable palms, non-harvestable palms, and background (Figure 11). The background is categorised as other objects such as other trees, rivers, cars, and roads. The samples of each label are divided

(a) (b) (c)

Figure 10. The location of the research: (a) Kampung Penakup, Mukah; (b) Balingian area, within seven kilometres from Balingian township; and (c) Sungai Talau Research Station, Dalat. (Photos adapted from GoogleMap: https://www.google.com.my/maps/@2.8191583,111.9126416,2233m/data=!3m1!1e3?hl=en&authuser=0\)

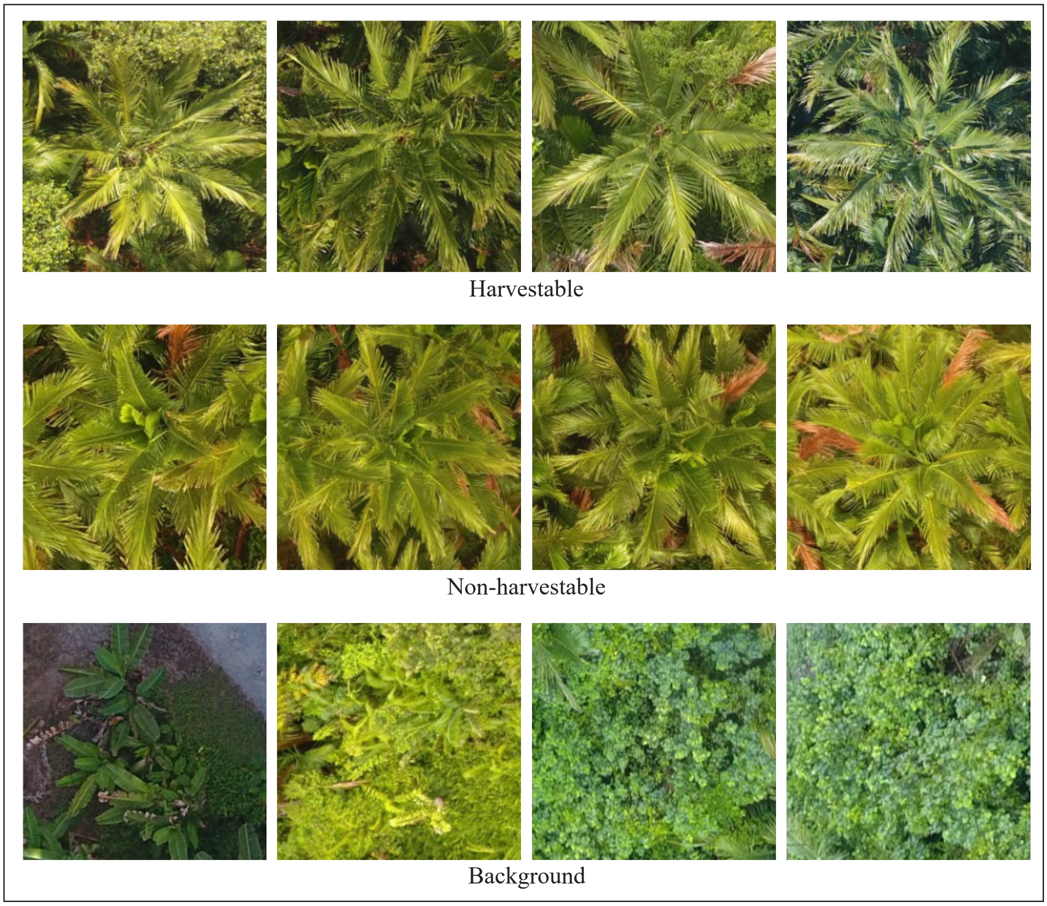

Figure 11. Example of the sample for each label

into five groups. During the training and validation phase, 756 samples (80%) are the training set, and the remaining 189 samples (20%) are the validation set.

Proposed Method

AlexNet performs very well in ILSVRC 2012 by yielding a high accuracy, and it has fewer layers which speed up the learning time. It shows that Alexnet has a low computational cost. However, AlexNet has been tested and yields low accuracy when tested with sago palm image data; hence features created in Xception and ResNet are included in AlexNet architecture. Xception has done a great job by inventing the depthwise separable convolution where it reduces the no of parameters involved in the convolution operating. ResNet also shows an astonishing performance by introducing the residual connection by adding the feature of the input to the output. Hence, it proves that the accuracy of the model with many layers is not decreasing. This residual connection feature is also included in the CraunNet.

The model development was done by combining the architecture of both AlexNet and Xception models. Residual connection introduced in ResNet model also included in this work. Xception architecture was divided into three-part and replaced into AlexNet with some modification where the middle part of Xception is repeated four-time instead of eight-time. Four models have been proposed based on this approach. The hybrid work is done by dividing Xception and AlexNet into three parts. Each part of Xception will be replaced into AlexNet along with some modifications. This practice of layer replacement has been done in previous research, which was done by Li and Jin (2020). They used the k-parallel CNN structure to be replaced in the MugNet model. This hybrid is done using the Deep Network Designer provided in Matlab software. The application eases the replacement of the depthwise separable convolution into AlexNet architecture. Two residual connections are involved in this hybrid work at the first two layers and section B of the AlexNet architecture. Figures 12 and 13 illustrate how the hybrid work is done.

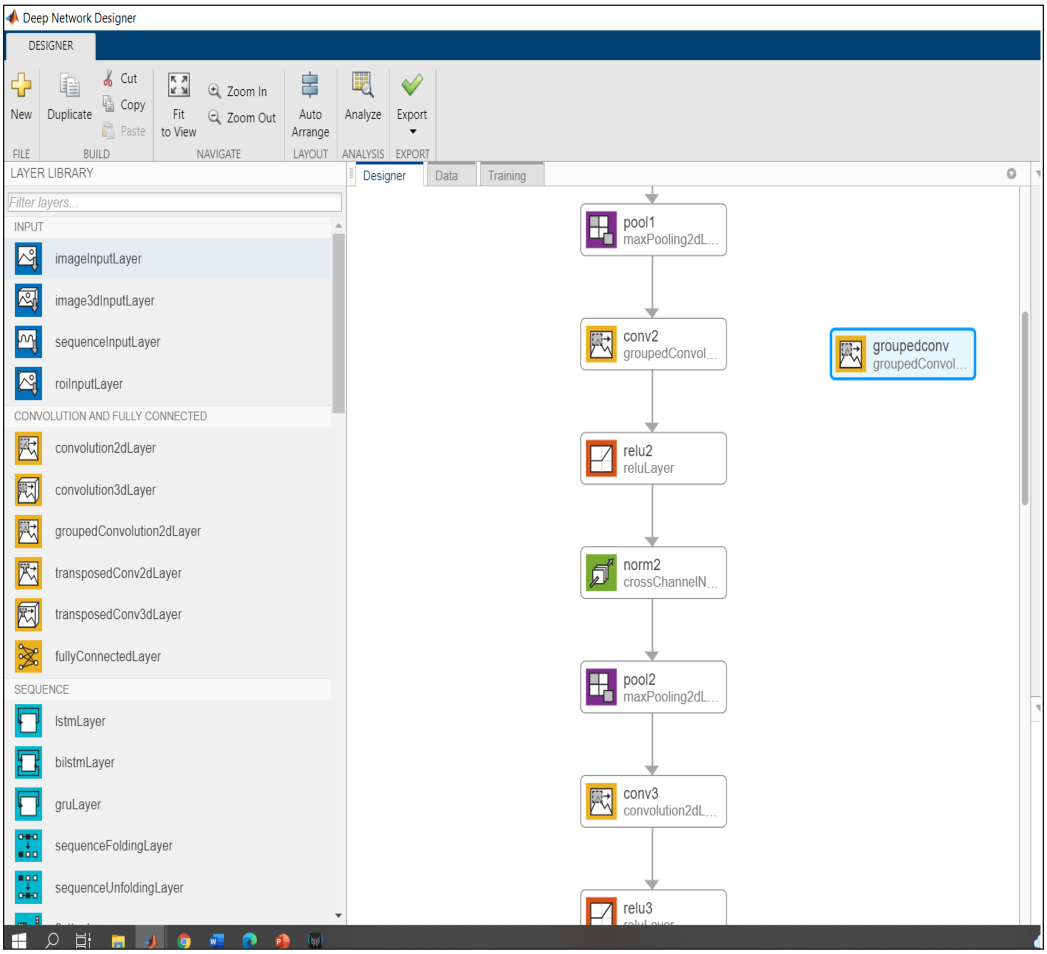

Figure 12. The hybrid work done by the Deep Network Designer Application is provided in Matlab software

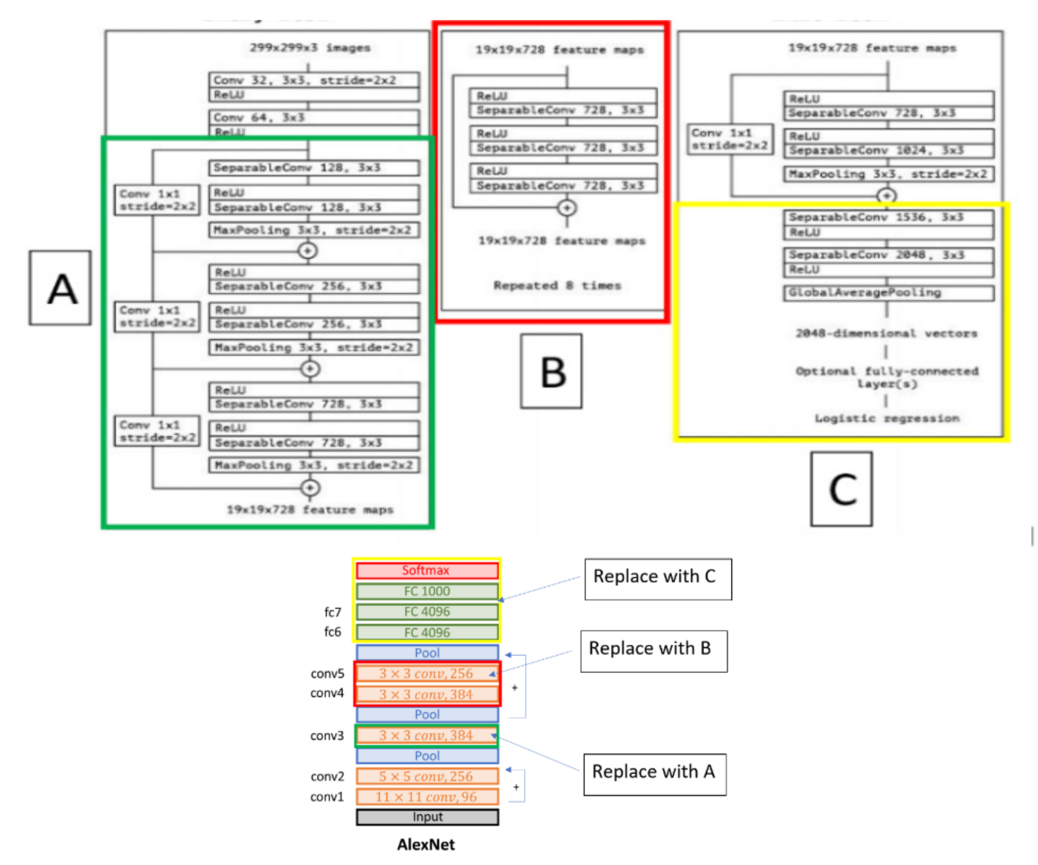

Figure 13. All the marked parts are replaced with all three sections from Xception and two residual connections involved

Training and Validation Phase

Convolution Neural Network (CNN) consists of training and validation phases. The training phase is where the model learns all the features of the selected labels. In this paper, 80% of the image data is the training set, and 20% will be the validation set. The stochastic gradient descent with momentum (SGDM) is the main solver in the training process. SGDM is a method that helps accelerate the gradients vector in the right direction, which can perform faster converging. The learning rate of the training phase is 0.0001 because the small learning rate can allow the model to learn more optimally. Meanwhile, the mini-batch is set to be ten samples per batch. The computer's hardware reaches its limit as the mini-batch is set to more than ten.

The whole learning process is based on five-cross-validation. Cross-validation is one of the model validation techniques to forecast the accuracy of the model would perform (Kohavi, 1995). The main purpose of cross-validation is to test the ability of the model to predict new data to avoid overfitting or selection bias (Browne, 2000). The samples of each label are divided into five groups: s1, s2, s3, s4, and s5 (Figure 14). One of the groups will be a test sample, and the rest will be the training sample. Hence, five iterations have been done, each iteration repeated five times, and the average has been taken. The empirical comparison has been done with Xception, AlexNet, and ResNet in terms of learning time and recognition accuracy. The performance of the model is evaluated based on accuracy and training time.

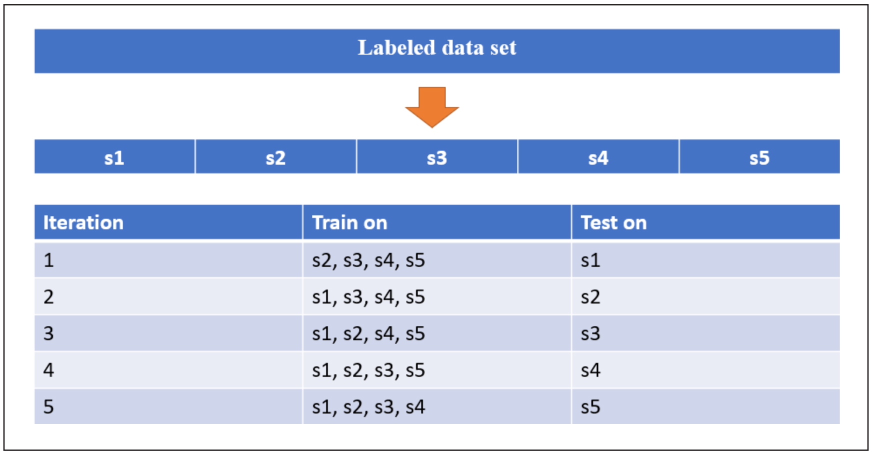

Figure 14. Cross-validation technique

RESULTS AND DISCUSSION

Nine hundred forty-five samples have been cropped for harvestable palm, non-harvestable palm, and background categories. Each label will be divided into five groups, one of the groups is the test set, and the rest is a train set; hence there are 189 samples for the test set, and 756 samples will be a train set for each iteration. These sample groups are trained into CraunNet, AlexNet, Xception, and Resnet. The time taken for the model to be trained will show the computational cost of the model to train the model. The training and validation process has been replicated five times for each iteration, and the average of the result has been calculated. All the results of the accuracy and training time of these models are shown in Tables 3 and 4 and Figures 15 and 16.

The accuracy table shows that the ResNet performs well by obtaining the highest accuracy for all sample groups while Xception scores the second. The accuracy of the CraunNet model is comparable with Xception and ResNet, with a score of more than 85% of accuracy on average. However, it has been shown from the training time chart that most of Xception's training time is doubled by CraunNet. It shows that CraunNet costs less time than other models in terms of learning time and retains high accuracy at the same time. The learning time portrays the computational cost of the model. The shorter the learning

Table 3
The average accuracy of all models based on five iterations

	Accuracy (%)			
Model	**CraunNet**	**Xception**	**Alex net**	**Resnet**
Iteration 1	86.98%	94.6%	33.33%	95.24%
Iteration 2	86.24%	91.95%	33.33%	93.75%
Iteration 3	86.56%	93.96%	33.33%	95.03%
Iteration 4	85.92%	92.27%	33.33%	93.22%
Iteration 5	82.75%	93.01%	33.33%	95.76%
Average	85.69%	93.16%	33.33%	94.6%

Table 4
The average training time of all models based on five iterations

	Time (Min)			
Model	**CraunNet**	**Xception**	**Alex net**	**Resnet**
Iteration 1	10.7	25.4	3.5	30.3
Iteration 2	10.5	25.6	3.5	29.9
Iteration 3	10.8	25.6	3.5	30.7
Iteration 4	10.4	25.5	3.5	30.9
Iteration 5	10.4	24.7	3.5	30
Average	10.56	25.36	3.5	30.36

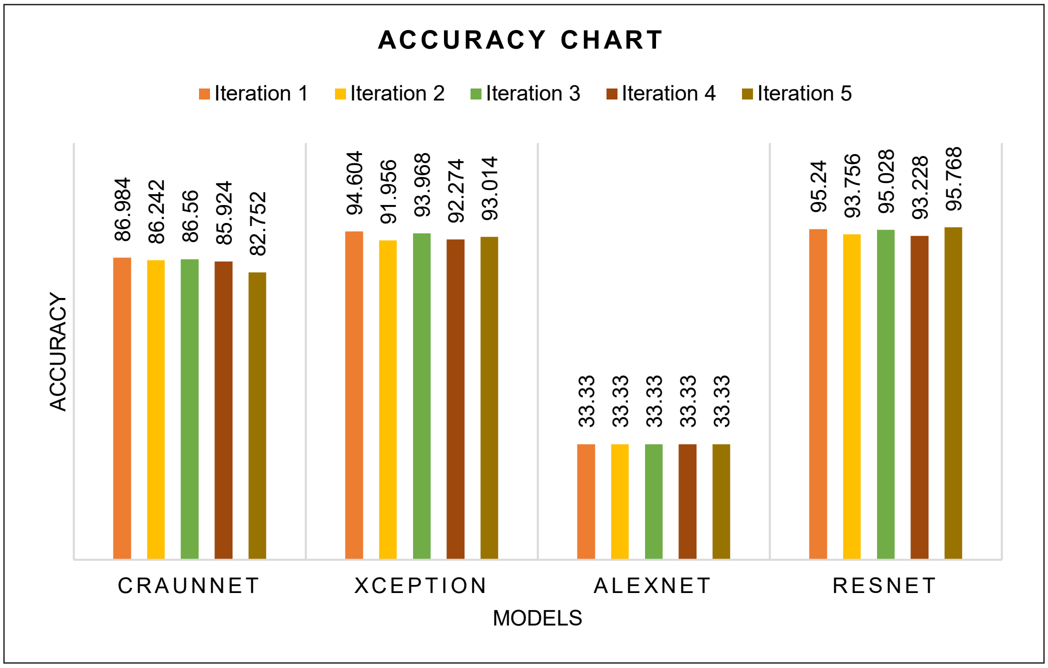

Figure 15. Accuracy yield of each model for each iteration

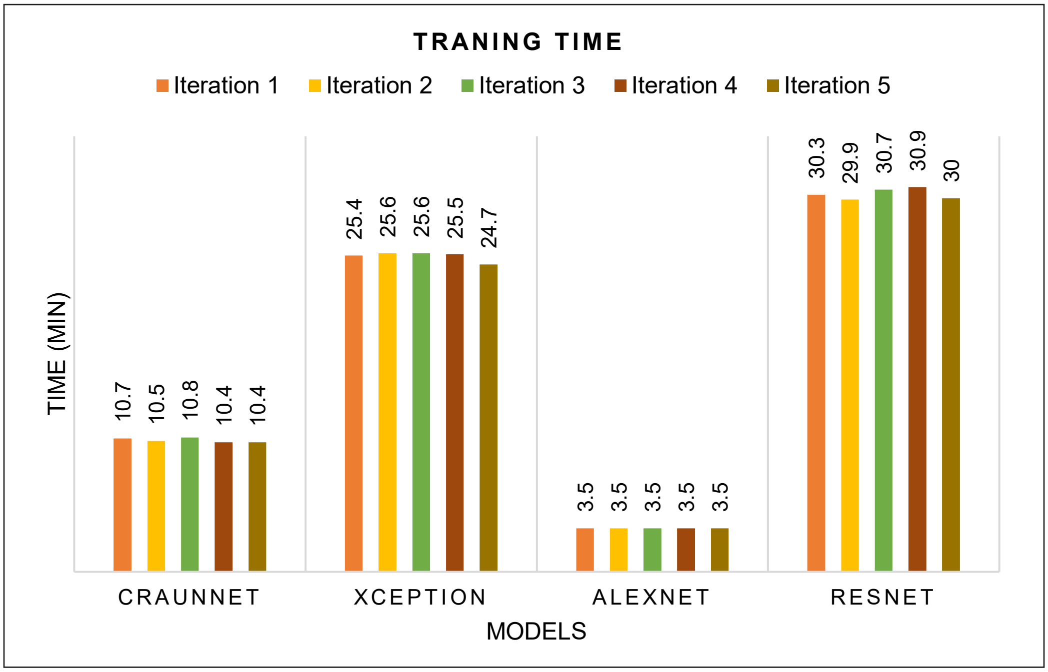

Figure 16. Training time chart for both training and validation process

time shows, the lesser the computational cost involved. These factors give a huge impact in determining the maturity of the sago palm since it dealt with a large area of plantation. The faster the operation take, it makes the harvesting process is done efficiently. It also smoothens the starch production process, impacting the economy, especially agriculture.

CONCLUSION

In conclusion, it is proved that this algorithm development can detect and identify the maturity of the palm, not only saving time but also reducing the labour cost to reach each palm only for determining the palm maturity. From the results, the CraunNet yield relatively high accuracy (85.69%), and it performs consistently based on the cross-validation. The ResNet yields a very high accuracy among all models. However, the training time shows that the CraunNet is shorter than ResNet and Xception. It shows that the CraunNet involves very less computational cost compared to ResNet and Xception. Less computational cost is important to ensure the detection process is done faster so the harvesting work can be done efficiently. There is a big number of image data involved as the area of the real sago plantation is huge. Other than that, it can optimize the usage of the computer's hardware, such as GPU. In the long-term process, this paper can figure out other properties of the palm and, importantly, increase the productivity of sago palm plantations.

ACKNOWLEDGEMENTS

This work is supported by CRAUN Research Sdn. Bhd. under Upstream Technology Division.

REFERENCES

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, *44*(1), 108-132. https://doi.org/10.1006/jmps.1999.1279

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258). IEEE Publishing. https://doi.org/10.1109/cvpr.2017.195

DJI. (2016). *Phantom 4 - Product information*. DJI Official. https://www.dji.com/phantom-4/info

Ehara, H., Toyoda, Y., & Johnson, D. V. (2018). *Sago palm: Multiple contributions to food security and sustainable livelihoods*. Springer Nature. https://doi.org/10.1007/978-981-10-5269-9

Farooq, A., Jia, X., Hu, J., & Zhou, J. (2019). Knowledge transfer via convolution neural networks for multi-resolution lawn weed classification. In *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)* (pp. 01-05). IEEE Publishing. https://doi.org/10.1109/whispers.2019.8920832

Flach, M. (1997). *Sago palm: Metroxylon sagu Rottb.-Promoting the conservation and use of underutilized and neglected crops. 13*. International Plant Genetic Resources Institute.

Habaragamuwa, H., Ogawa, Y., Suzuki, T., Shiigi, T., Ono, M., & Kondo, N. (2018). Detecting greenhouse strawberries (mature and immature), using deep convolutional neural network. *Engineering in Agriculture, Environment and Food*, *11*(3), 127-138. https://doi.org/10.1016/j.eaef.2018.03.001

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778). IEEE Publishing. https://doi.org/10.1109/cvpr.2016.90

Hidayat, S., Matsuoka, M., Baja, S., & Rampisela, D. A. (2018). Object-based image analysis for sago palm classification: The most important features from high-resolution satellite imagery. *Remote Sensing, 10*(8), Article 1319. https://doi.org/10.3390/rs10081319

Howell, P. S. A. (2017). *Effect of sucker prunning on sago palm (Metroxylon sagu Rottb.) growth performance* (Master Thesis). Universiti Putra Malaysia, Malaysia. http://psasir.upm.edu.my/id/eprint/83269/1/t%20FSPM%202017%205%20%281800001036%29.pdf

Kavukcuoglu, K., Ranzato, M. A., Fergus, R., & LeCun, Y. (2009). Learning invariant features through topographic filter maps. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1605-1612). IEEE Publishing. https://doi.org/10.1109/cvpr.2009.5206545

Khvostikov, A., Aderghal, K., Benois-Pineau, J., Krylov, A., & Catheline, G. (2018). *3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies*. arXiv Preprint. https://doi.org/10.1109/cbms.2018.00067

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International Joint Conference on Artificial Intelligence (IJCAI, 1995)* (Vol. 14, No. 2, pp. 1137-1145). ACM Publishing.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 1-9). NeurIPS Proceedings.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84-90. https://doi.org/10.1145/3065386

Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *IEEE Transactions on Neural Networks*, *8*(1), 98-113. https://doi.org/10.1109/72.554195

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. https://doi.org/10.1038/nature14539

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324. https://doi.org/10.1109/5.726791

Li, M., & Jin, Y. (2020). An hybrid parallel network structure for image classification. In *Journal of Physics: Conference Series* (Vol. 1624, No. 5, p.052005). IOP Publishing.

Mubin, N. A., Nadarajoo, E., Shafri, H. Z. M., & Hamedianfar, A. (2019). Young and mature oil palm tree detection and counting using convolutional neural network deep learning method. *International Journal of Remote Sensing, 40*(19), 7500-7515. https://doi.org/10.1080/01431161.2019.1569282

Samala, R. K., Chan, H. P., Hadjiiski, L. M., Helvie, M. A., Cha, K. H., & Richter, C. D. (2017). Multi-task transfer learning deep convolutional neural network: Application to computer-aided diagnosis of breast cancer on mammograms. *Physics in Medicine & Biology*, *62*, Article 8894. https://doi.org/10.1088/1361-6560/aa93d4

Yu, J., Schumann, A. W., Cao, Z., Sharpe, S. M., & Boyd, N. S. (2019). Weed detection in perennial ryegrass with deep learning convolutional neural network. *Frontiers in Plant Science*, *10*, Article 1422. https://doi.org/10.3389/fpls.2019.01422

Zhang, M., Li, L., Wang, H., Liu, Y., Qin, H., & Zhao, W. (2019). Optimized compression for implementing convolutional neural networks on FPGA. *Electronics, 8*(3), Article 295. https://doi.org/10.3390/electronics8030295