BANANA SEEDLINGS HEALTH MONITORING FOR MICRO AIR VEHICLES USING DEEP CONVOLUTIONAL NEURAL NETWORK

by

TAN SHU CHUAN

Thesis submitted in fulfilment of the requirements for the Bachelor Degree of Engineering (Honours) (Aerospace Engineering)

July 2021

ENDORSEMENT

I, Tan Shu Chuan hereby declare that I have checked and revised the whole draft of the dissertation as required by my supervisor.

Shu Chuan

(Signature of Student)

Name: Tan Shu Chuan
Date: 9/7/2021

(Signature of Supervisor)

Name: Dr. Ho Hann Woei
Date:

ENDORSEMENT

I, Tan Shu Chuan hereby declare that all corrections and comments made by the supervisor and examiner have been take consideration and rectified accordingly.

Shu Chuan

(Signature of Student)

Name: Tan Shu Chuan
Date: 9/7/2021

(Signature of Supervisor)

Name: Dr. Ho Hann Woei
Date: 9/7/2021

(Signature of Examiner)

Name: Assoc. Prof. Ir. Ts. Dr. Parvathy Rajendran
Date: 9 Julai 2021

DECLARATION

This thesis is the result of my own investigation, except where otherwise stated and has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any other degree.

Shu Chuan

(Signature of Student)

Date: 9/7/2021

ACKNOWLEDGEMENT

First and foremost, I would like to express my deep and sincere gratitude to my research supervisor, Dr. Ho Hann Woei for his never-ending patience and guidance throughout the project. His willingness to help and spare his time to provide me guidance and pointers is very much appreciated. It was a great privilege and honor to work and study under his guidance.

I would also like to thank Assoc. Prof. Ir. Ts. Dr. Parvathy Rajendran for her patience in evaluating my thesis draft and also her detailed comment to allow this thesis to improve in quality.

Besides, I would like to thank Mr. Mohd Amir bin Wahab for his technical support throughout the project.

Last but not least, I would like to thank my friends and family for their precious spiritual support throughout the research and my life in general.

ABSTRAK

Pisang merupakan salah satu jenis buah-buahan yang paling banyak dihasilkan tetapi juga buah-buahan yang amat digemari di Malaysia. Ramai petani melabur dalam teknik kultur tisu di rumah hijau untuk meningkatkan pengeluaran, tetapi benih pisang yang berkultur tisu juga akan diserang oleh pelbagai penyakit dan serangan perosak. Untuk memantau keadaan kesihatan benih pisang yang berkultur tisu, mereka perlu mengupah ramai buruh atau memasang kamera atau sensor di seluruh rumah hijau. Untuk tujuan ini, kami mencadangkan kenderaan udara mikro yang berpatutan, bersaiz tangan, dan automatik untuk membantu petani. Kenderaan udara micro mempunyai mobiliti untuk mengakses setiap sudut ruang terkurung dan memperoleh pandangan mata burung yang sangat baik supaya memberikan kemudahan dalam menjalani tugas pemantauan, dan mampu terbang mengikut laluan penerbangan yang dikehendaki dan menangkap gambar setiap pokok benih dengan tepat. Penyelidikan ini membandingkan pretasi lima jenis YOLO dan Single Shot MultiBox Detector (SSD) kaedah pengesanan objek dalam meramalkan status kesihatan benih pisang. Tiny-YOLOv4 yang mempunyai kompromi terbaik antara ketepatan pengesanan dan kelajuan pengesanan kemudian dilatih dengan resolusi rangkaian dan berat sampel negative yang berbeza. Tiny-YOLOv4 dengan resolusi rangkaian 416×416 dan 18% sampel negatif mempunyai mAP tertinggi sebanyak 99.08% dan dipilih untuk mengkategorikan tumbuh-tumbuhan kepada kelas ‘normal’ dan ‘tidak sihat’ berdasarkan imej yang ditangkap oleh kamera onboard. Beberapa ujian penerbangan telah dilakukan di dalam dewan tertutup, dan tumbuh-tumbuhan diklasifikasikan dengan tepat. Lokasi tumbuhan yang tidak sihat dihantar kepada petani supaya menjalani tindakan selanjutnya. Penyelesaian yang dicadangkan dalam projek ini

dijangka akan mengurangkan aktiviti intensif buruh dan kemungkinan kesilapan manusia.

ABSTRACT

Banana is one of the most produced but also highly demanded fruits in Malaysia. Many farmers invested in tissue-cultured techniques in greenhouses to increase production, but the tissue-cultured banana seedlings are not invincible to numerous diseases and pest attacks. To monitor the health conditions of the tissue-cultured banana seedlings, they need to hire many laborers or install cameras or sensors throughout the greenhouse. To this end, we proposed an affordable, hand-palm-sized, and automatic Micro Air Vehicle (MAV) to help farmers. MAVs have the mobility to access every corner of confined spaces and acquire an excellent bird's-eye view, which provides great convenience in monitoring tasks, and is capable to fly according to the desired flight path and capture each plant precisely. This research compares the performances of five YOLO and Single Shot MultiBox Detector (SSD) deep learning model architectures in predicting the health status of banana seedlings. The Tiny-YOLOv4 model architecture, which has the best compromise between detection accuracy and detection speed, was then trained with different network resolutions and weightage of negative samples. Tiny-YOLOv4 with the 416×416 network resolution and 18% of negative samples has the highest mAP of 99.08% and was chosen to categorise the plants into normal and unhealthy classes based on the images captured by an onboard camera. Several flight tests were performed successfully in an indoor hall, and the plants were classified accurately. The locations of unhealthy plants are sent to notify farmers of further actions. The proposed solution in this project is expected to highly reduce labor-intensive activities and possible human error.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AP	Average Precision
CV	Computer Vision
CSP	Cross Stage Partial
CeRSAA	Regional Center of studies and aids in agriculture of Albenga, Savona, Italy
DCNN	Deep Convolutional Neural Network
FPS	Frame Per Second
FN	False Negative
FP	False Positive
GCS	Ground Control Station
GPU	Graphics Processing Unit
IoU	Intersection over Union
MAV	Micro Air Vehicle
MS COCO	Microsoft Common Objects in Content
SSD	Single Shot MultiBox Detector
TP	True Positive
TN	True Negative
UGV	Unmanned Ground Vehicle
YOLO	You Only Look Once

CHAPTER 1

INTRODUCTION

In Malaysia, bananas (*Musa* spp.) are the most produced fruit and one of the most consumed fruits in 2017 (Tumin and Shaharudin, 2019). Malaysians are estimated to consume 10.0 kilograms of bananas per person in a year, more than the consumption per capita of pineapples and durians. In 2017, the plantation of bananas was conducted at almost 35,000 hectares of land. The production had reached more than 350,000 metric tons and accounted for 24% of the total fruit production in Malaysia. Between 2013 and 2017, the average self-sufficiency ratio for bananas in Malaysia is 103%, thus it means that banana production has fulfilled domestic needs, and Singapore was the primary import market in 2017. Statistics showed that Malaysia has exported more than 20,000 metric tonnes of bananas to Singapore, valued at RM 40.1 million.

Therefore, the plantation of bananas with tissue-cultured techniques in greenhouses is getting more to meet the high demand in the market. Propagation of bananas using suckers has the probability of perpetuating the spread of disease and pests, which inherit from the parents. The tissue-cultured techniques in banana propagation ensure gene selection, gene preservation, and disease-free. The elimination of pests and the reduction of infestation in the new plantation by using this technique have contributed significantly to higher production rates. Tissue-cultured plants have more functional green leaves at the planting phase, and this enables the plants to contain more chlorophylls to process their own food. But for conventional suckers, they use the food stored in the corm to start the initial growth (Pavithra C.B, 2012). Higher yields and productions for tissue-cultured bananas due to their fast vegetative growth.

However, banana propagation through tissue-cultured techniques is not invincible to diseases. Banana plants that are infected by diseases will produce lousy

quality and low bunch weight of banana fruits. The major diseases of bananas in Malaysia are Moko disease, Panama disease, and black Sigatoka. Moko disease, bacterial wilt of bananas, is caused by *Ralstonia Solanacearum*, which can survive in soils for over 18 months. Moko disease induces wilting, which starts with the yellowing and collapse of young leaves. As the disease progresses, old leaves will turn yellow and become necrotic. The bacteria will block the movement of water and nutrients to upper plant parts and cause infected fruits to show deformed growth and shrivel up as the pulp is destroyed by dry hot (Alvarez and Pantoja, 2015). Besides, Panama disease, also known as Fusarium wilt, is one of the most devasting diseases of banana caused by the soil-inhabiting fungus species *Fusarium oxysporum* forma specialis *cubense* has completely wiped out the Cavendish plantations in the 1990s. It invades young roots through wounds and progresses into the rhizome, followed by a rapid invasion of the rootstock and leaf bases. The disease has apparent symptoms, such as wilting or yellowing of leaves, reddish-brown discoloration away from the centre of vascular tissue, and root necrosis (Wong *et al.*, 2019). Another disease that brings profit losses to farmers is the black Sigatoka disease, caused by the fungus *Mycosphaerella fijiensis*. The disease affects the ability of plants to conduct photosynthesis by reducing the area of the healthy leaf. Consequently, the weight of the bunches and fruit fingers reduces due to insufficient nutrients. The disease starts as tiny reddish-rusty brown dots found on lower leaf surfaces. The dots gradually lengthen and darken to produce reddish-brown leaf streaks. The streaks have the capability to affect the photosynthesis process. Thus, the disease can be easily recognised by the presence of reddish-brown streaks and dots (Alvarez and Pantoja, 2015). The spread of the diseases can happen in many ways, such as infected soils when transported by footwear or animals, transmission through insects, and irrigation from a contaminated water source. Moreover, low levels of water stress and

nutrient, and pest-attacked banana seedlings are most likely to encounter incomplete development and thus affect the yields. The health conditions of banana seedlings can be determined by observing the colour and status of leaves. Banana plants infected by diseases or have a low level of water stress will show apparent symptoms on leaves. Hence, early identification of unhealthy banana seedlings through leaves is required to avoid the rapid dissemination of diseases and to apply care treatments on these unhealthy banana seedlings.

To ensure the plants grow well in the early stage, several methods have been proposed. Traditional monitoring system involves farmers to walk through greenhouses to inspect the health status of each plant. This system is ineffective in terms of costs and efficiency as it requires many labours and effort. Careless farmers in identifying unhealthy plants are then led to profit losses. Hence, an automation system is required to provide effective solutions by reducing the reliance on human operators in labour-intensive activities, such as monitoring and pest control. There are few solutions for greenhouse automation that have been proposed in the recent past. For instance, a semi-autonomous robot designed by experts of the Regional Center of studies and aids in agriculture of Albenga, Savona, Italy (CeRSAA) has the function to monitor the growing health state of plants in a greenhouse (Acaccia *et al.*, 2003). The robot is equipped with a vision system that can account for a wide range of symptomatologies of unhealthy plants. The robot will collect leaf samples from plants that are suspected to be sick and send the leaves to a laboratory for analysis. It will perform the spraying operation towards the plants which are acknowledged as unhealthy. Nevertheless, Unmanned Ground Vehicles (UGVs) are facing some challenges with their mobile platforms inside the greenhouse. They are difficult to operate on irregular terrains and narrow paths. Besides, they also tend to sink into the soft ground mainly in high humidity conditions, and high

maintenance costs, which is unaffordable for small-scale greenhouse farmers. The applications of Micro Air Vehicles (MAVs) in agriculture (Kurkute, 2018) can significantly resolve such UGV's challenges as they are safer, flexible, and low cost. Thus, this research serves to present a novel idea of integrating machine learning into the indoor navigation system on MAVs. The work involved is mainly developing an Artificial Intelligence (AI)-based monitoring system to inspect the health conditions of banana seedlings. A formidable algorithm called Deep Convolutional Neural Network (DCNN) is utilised to identify and locate unhealthy banana leaves. This system would then be integrated with an indoor positioning system that employs ultra-wideband technology to fly according to the desired flight path. Ground Control Station (GCS) will then further analyse the images captured by an onboard camera and send useful information to farmers. This system is fully automated to reduce human errors, save on labour costs, and improve monitoring efficiency.

An MAV is selected instead of a UGV as it acquires an excellent bird's eye view, which provides great convenience in monitoring tasks. In recent times, the application of MAVs in the agriculture sector has gained traction. MAV technology has been introduced to carry out precision farming. It could capture high-resolution images of the field or the plantation in real-time, allowing farmers to understand their areas better. Aerial photos taken using the MAV allow the data of the plantation to be collected more efficiently and even with more precision as it is made from a high altitude enabling the camera to cover a large area when collecting images. The data collected would then be processed in the GCS and translated into useful information, such as plant health. These allow greenhouse owners to efficiently utilise the useful information to maximise the yield from crops (Norasma *et al.*, 2019).

To monitor the plant health conditions, traditional Computer Vision (CV) approaches are built on handcrafted features, and it has difficulty to select which features are crucial in each given image (O' Mahony *et al.*, 2019). A formidable algorithm called DCNN is utilised to address the problems existing in traditional approaches. The DCNNs are state-of-art algorithms that take in an input image, assign importance to various objects in the image, and introduce non-linearities into networks to differentiate them from the others. The architecture of DCNN is analogous to the connectivity of neurons in the human brain. Artificial neurons, or the so-called perceptrons, from multilayers connect with each other to form neural networks. One of the powerful applications of DCNN is object detection. Object detection is the combination of two tasks which are image classification and object localisation. Image classification is the ability of software to recognise the class of an object within an image, whereas object localisation involves locating an instance of a particular object category in an image by drawing a bounding box around the object of interest, showing the extend of the object. In short, the application of deep learning in this study for object detection is to identify and locate normal and unhealthy banana seedling leaves correctly when the MAV is navigating. Therefore, a trained object detection model can significantly help in the monitoring task in greenhouses. A solid object detection model can be developed with comparatively little training data with transfer learning as the model is already pre-trained. Transfer learning is a machine learning technique where a model created for a task is utilised as the starting point for a model on another task by modifying the DCNN layers. It saves model development time, requires less training data, and provides better performance of neural networks. This technique has been widely utilised in the agriculture sector, such as the detections of apples during different growth stages (Tian *et al.*, 2019).

1.1 Object Detection Model

Generic object detection aims at locating and classifying the objects in an image from predefined categories. The frameworks of generic object detection can be classified into two types, region proposal based framework and regression based framework. The concept of the region proposal based framework is to generate region proposals at first and then categorise each proposal into different object categories. The working principle of this framework matches the attentional mechanism of the human brain, which first gives a quick scan of the whole situation and then emphasis on regions of interest. This framework is not the interest for this project as it is composed of several correlated stages, and thus the time spent in handling different stages makes it has poor performance in the real-time application (Zhao *et al.*, 2019). Examples of the region proposal based methods are Faster R-CNN (Ren *et al.*, 2017) and Mask R-CNN (He *et al.*, 2020). Meanwhile, the object detection models that are trained in this research are regression based framework. This one-stage framework has an excellent performance in the real-time application as it maps straightly from image pixels to bounding box coordinate and class probabilities. This work only involves two significant frameworks, which are You only look once (YOLO) (Joseph Redmon, Santosh Divvala, Ross Girshick, 2016) and Single Shot MultiBox Detector (SSD) (Liu *et al.*, 2016).

Figure 1.1 shows the anatomy of an object detector. A one-stage detector takes an image as an input and passes features through a backbone that is composed mainly of convolution layers. The backbone extracts essential features, and the combination of backbone features layers happens in the neck. Detection happens in the head.

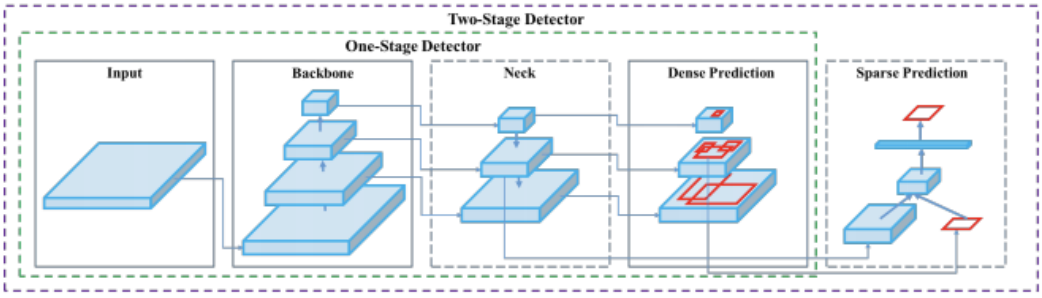

Figure 1.1: The anatomy of an object detector. (Bochkovskiy, Wang and Liao, 2020).

1.1.1 You Only Look Once version 3 (YOLOv3)

YOLO works by having a single convolutional network that divides an input image into S × S grid, and then it predicts the boundary boxes and probabilities for each grid cell. The basic idea of how YOLO works is shown in Figure 1.2. YOLOv3 is supported by the powerful backbone - Darknet-53, a convolutional neural network that is 53 layers deep. The most salient feature of YOLOv3 compared to previous YOLO versions is YOLOv3 makes detections at three different places in the neural network. The first detection is completed by the 82nd layer. Before starting the first detection, some layers before in the network are responsible to down sample the image. The second detection is done by the 94th layer, with the aid of layer 79 to up sample the dimension by two times. The final detection is carried out at the 106th layer. Having detections at three different scales makes YOLOv3 powerful in detecting smaller objects. The upsampled layers in the second and last detections help to address the issue of predicting small objects. Furthermore, the bounding boxes per image have been increased as it predicts boxes at three different scales where the number of bounding boxes predicted is ten times more than the bounding boxes predicted in YOLOv2. Lastly, the classes are no longer softmax where it can perform multi-label classification of objects within the image

compared to the earlier versions of YOLO (Kathuria, 2018). The architecture of YOLOv3 is demonstrated in Figure 1.3.

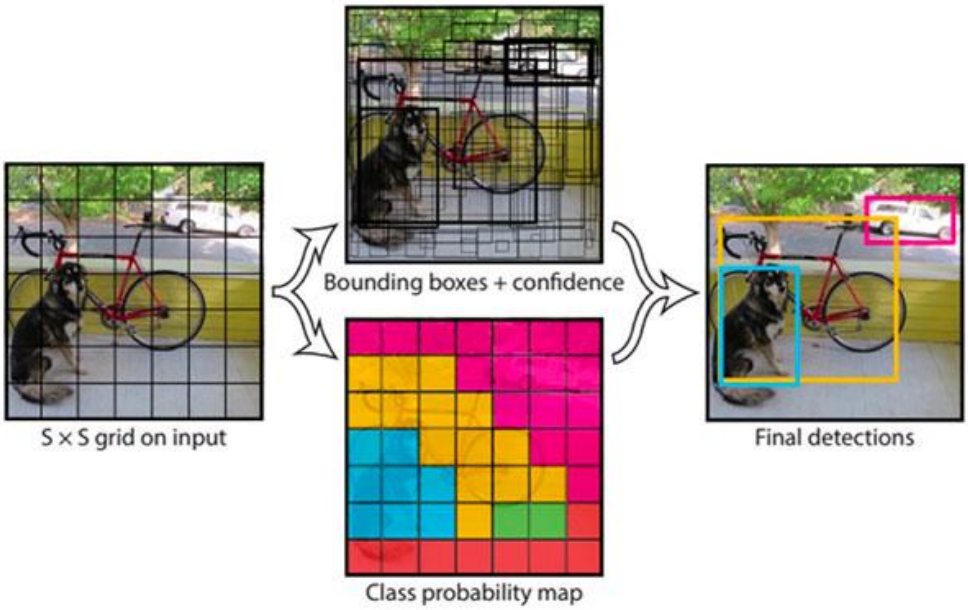

Figure 1.2: Main idea of YOLO (Joseph Redmon, Santosh Divvala, Ross Girshick, 2016).

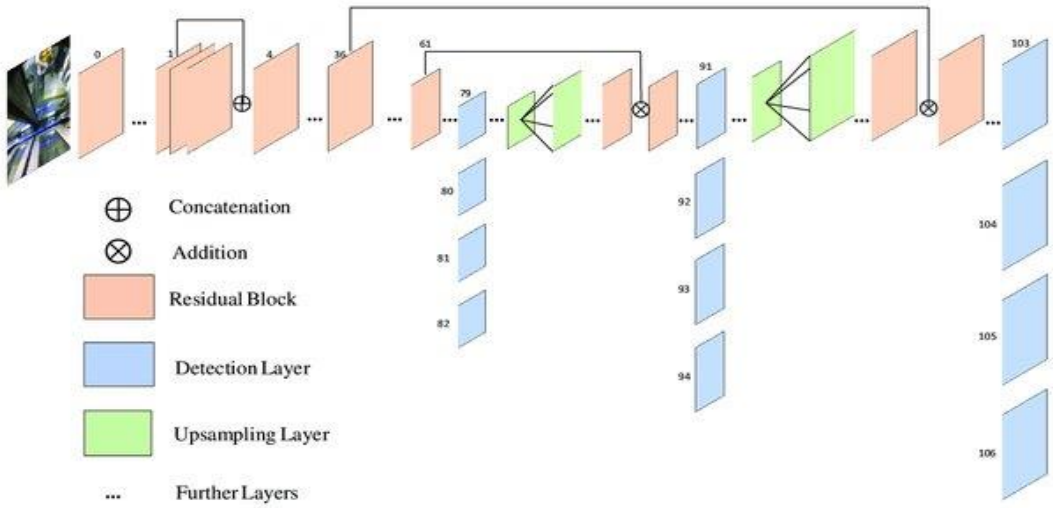

Figure 1.3: YOLOv3 Network Architecture (Dai *et al.*, 2020).

To make YOLO even faster in a real-time application, a lighter version of YOLO architecture called Tiny-YOLO is introduced. Tiny-YOLO has a faster inference speed and smaller model size than its brother, but it tends to be less accurate. Figure 1.4 shows the network architecture for Tiny-YOLOv3. Tiny-YOLOv3 is running on the Darknet19 backbone, in which the reduction of convolutional layers has dramatically improved the real-time detection speed, but it also reduces the accuracy of the architecture. Tiny-YOLOv3 has only seven convolutional layers for the basic structure and six maxpool layers for extracting features of an input image (Xiao *et al.*, 2019).

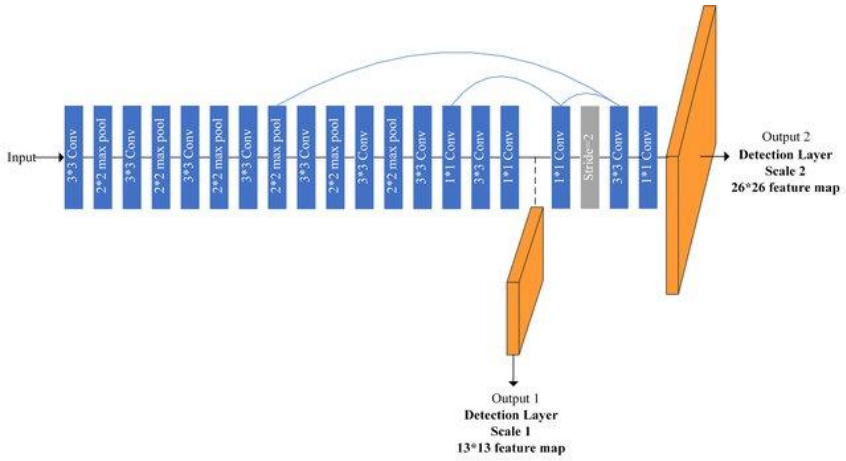

Figure 1.4: Tiny-YOLOv3 Network Architecture (Fang, Wang and Ren, 2020).

1.1.2 You Only Look Once version 4 (YOLOv4)

YOLOv4 is an improvement for YOLOv3 by implementing a new architecture in the backbone and modifying the neural network in the neck. These changes have improved the accuracy of YOLOv4 and have allowed the training of this neural network on a single Graphics Processing Unit (GPU) to become more manageable. The YOLOv4 backbone architecture is constructed by three parts: Bag of Freebies, Bag of Specials, and CSPDarknet53. Bag of Freebies is a method that improves the performance of the

network by changing the training strategy without adding to inference time in production. This method applies data augmentation during the training to expose the model to various situations that it would not have otherwise seen. The generalisation of the model training can be improved by implementing geometric distortion, photometric distortion, and mix up augmentation techniques. The Bag of Specials method is deployed by YOLOv4 which increases the inference cost but significantly enhances the accuracy of the neural network. Mish activation function is introduced to replace traditional activation functions as it helps in better expressivity and information flow. Cross Stage Partial (CSP) architecture reduces the computational complexity by separating the input into two parts. YOLOv4 utilises the CSP architecture with the Darknet53 to form CSPDarknet53 which gives priority to larger receptive field and best feature aggregation techniques. A larger receptive field refers to a larger part of an input image that is visible to one filter at a time, whereas feature aggregation refers to the accommodation of features from different levels in the backbone to account for varied scales of an object (Rugery, 2020). In short, the implementation of different architectures in YOLOv4 has enhanced the learning capability of the neural network.

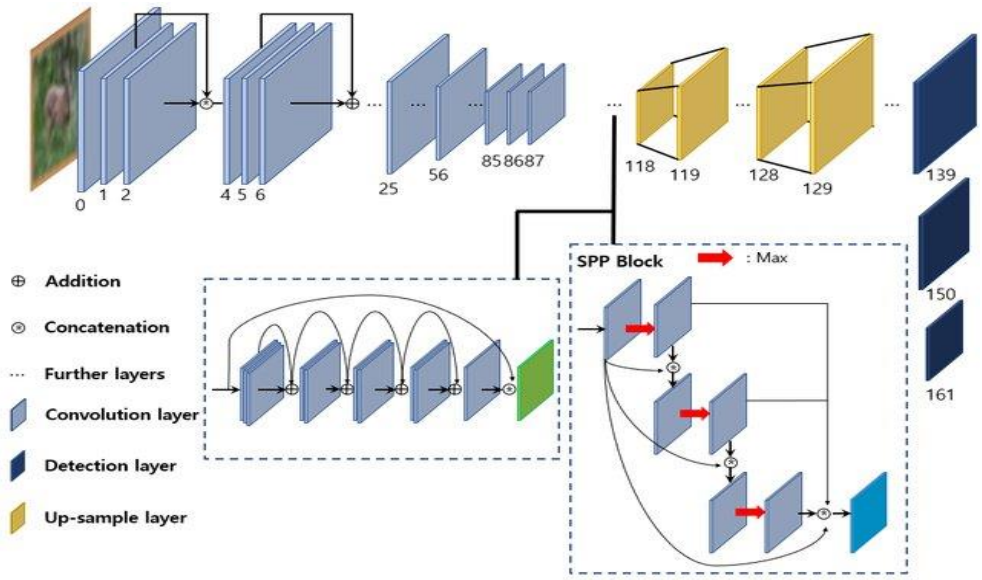

Figure 1.5: YOLOv4 network architecture (Kim and Kim, 2021).

Tiny-YOLOv4 is the compressed version of YOLOv4, which has a simpler network structure. The compressed version has a significant advantage in the real-time application as the number of convolutional layers in the CSP backbone has been drastically reduced. Besides, the training time required is lesser as the total of 137 pre-trained convolutional layers is deducted to 29 pre-trained layers. Additionally, the number of YOLO layers has been decreased to two instead of three, and there are fewer anchor boxes for prediction.

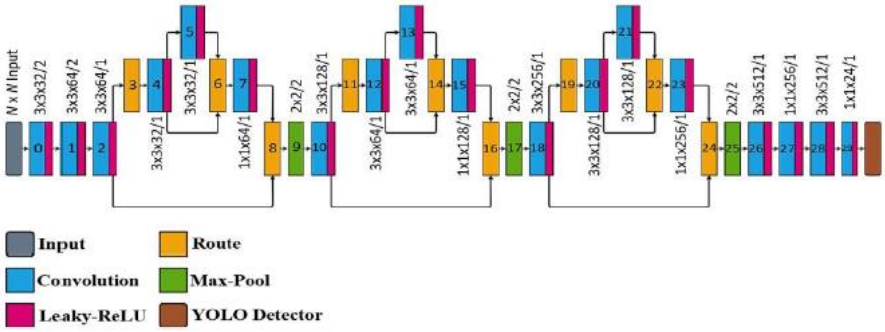

Figure 1.6: Tiny-YOLOv4 network architecture (Montalbo, 2020).

1.1.3 Single Shot Multibox Detector (SSD)

SSD is a one-stage detector where object localisation and classification tasks are done in a single forward pass of the network. The SSD is designed for object detection in a real-time application as it only needs to take one shot to detect multiple objects within an input image. SSD architecture is built on VGG16's architecture, but the fully connected layers are replaced with a set of auxiliary convolutional layers on it, as shown in Figure 1.7. The purpose of adding auxiliary convolutional layers is to extract features at multiple scales and progressively decrease the size of the input to each subsequent layer (Forson, 2017).

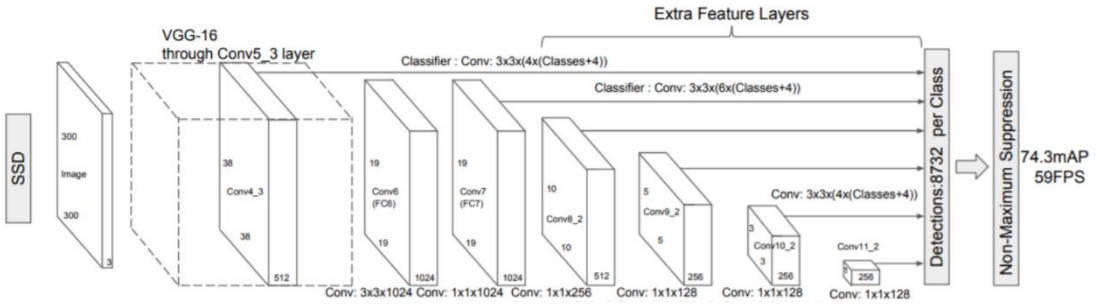

Figure 1.7: SSD network architecture (Liu *et al.*, 2016).

Different from fixed grids adopted in YOLOv3 and YOLOv4, the method applied by SSD is to take a set of default anchor boxes with a variety of aspect ratios and scales to distinct the output space of bounding boxes. Then the SSD will have a matching phase where the appropriate anchor box will be matched with the ground truth of each object within the image, and then the anchor box having the highest value of overlap will be responsible for predicting the object's class and location (Liu *et al.*, 2016).

The SSD object detection model used in this project is SSD MobileNet v2 FPNLite. Figure 1.8 shows the building block of SSD MobileNet v2 FPNLite, which is made from three important convolution layers, which are expansion convolution, depthwise convolution, and projection convolution. The full architecture of SSD MobileNet v2 consists of 17 of these building blocks in a row.

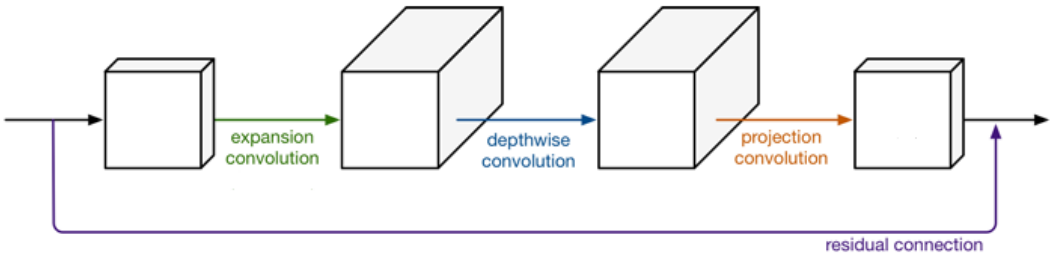

Figure 1.8: The building block of SSD MobileNet v2 FPNLite (Hollemans, 2021).

The expansion convolution layer aims to increase the number of channels in the data before going to the next convolution layers. Therefore, the output channels in the expansion layer are always more than the input channels. After that, the depthwise convolution filter serves to filter the output channels from the previous layer. The projection convolution layer has an opposite function as the expansion layer, it projects data with a high number of channels into a much lower number of channels. Moreover, the existence of residual connection is to help with the flow of gradients through the network.

1.2 Problem Statement

The traditional monitoring system in greenhouses involves farmers to check and inspect the health conditions of plants one by one. The farmers are required to walk on every path in the greenhouse in order to complete the monitoring task. The human brain is not capable of processing much information at one time, even our eyes can collect a lot of information with approximately 120 degrees of the visual field, thus traditional monitoring becomes time-consuming. Furthermore, the system is ineffective due to the exhaustion of stamina. The farmers are getting tired when they spend plenty of time conducting the monitoring task. Besides, the farmers tend to forget the details of the monitoring task, such as the number of unhealthy plants and their respective locations. Thus, the traditional monitoring system is unproductive and time-wasting, and it is not a good choice for modern greenhouse farmers. To practice an effective traditional monitoring system, the farmers can hire more labors. Each labor will get a lesser workload, and they can concentrate on their task. But the drawback is the financial problem for greenhouse farmers. The farmers have spent part of their finance in

maintaining the optimum conditions of the greenhouse. They do not afford to hire more workers to assist them in the monitoring task. Additionally, the involvement of more workers in the greenhouse increases the potential of ill-health as biological control agents are widely used to enhance or retard plant growth and to combat pest problems. The principal potential illness includes asthma, allergic, and dermatitis (Illing, 1997).

Recent technologies implemented within the monitoring system have shown better alternatives. One of the examples is the implementation of a robot to provide field-based analysis of tomato greenhouses. The robot has the capabilities of monitoring plant health and stresses using visible and near-infrared spectroscopy. But UGVs are not capable of monitoring each plant in greenhouses due to the design of UGVs. The wheels equipped make them only to move across a specified path within an agriculture field. Besides, they tend to sink into the ground when operating on soft and high humidity soil, and it is difficult to operate on irregular terrains and narrow paths. Furthermore, the high maintenance cost of UGVs makes the farmers give up on continuing to use the monitoring system.

To inspect the health conditions of banana seedlings, traditional object detection approaches built on handcrafted features have shallow trainable architectures. The traditional methods have the difficulty of selecting which features are essential in each given image. Feature extraction becomes more and more difficult when the number of classes to be identified increases. The traditional approach becomes complex as it involves humans to design and engineer the features. The accuracy and reliability of the object detection models totally depend on the human-engineered features. Hence, modern object detection models are utilised to differentiate the status of leaves for each banana seedlings. There are many types of object detection models, and each has its own sets of methods and algorithms to classify images. The variety of object detection models

is because they are used for different purposes, and the hierarchy of neural networks is different thus process features of objects differently. These object detection models have pros and cons with different focuses either on time or the model accuracy. There is also the difference in the ability of the object detection models to perform accurately and smoothly at acceptable Frame Per Second (FPS) once it is running on a 2GB VRAM GPU. With that, it becomes a challenge when designing a banana seedlings health monitoring system for MAVs.

1.3 Research Objectives

The research work within this thesis is performed to achieve the following objectives:

1. To design a banana seedling health monitoring system that can accurately predict normal and unhealthy banana seedling leaves in greenhouses with high precision and consistency.
2. To analyse the performance of five object detection models, which are YOLOv3, Tiny-YOLOv3, YOLOv4, Tiny-YOLOv4, and SSD MobileNet v2 FPNLite, and to select the object detection model with the best compromise between the accuracy and the real-time detection speed.
3. To study the performance of the selected object detection model with different network resolutions and weightage of negative samples.

1.4 Thesis Layout

This thesis consists of five main chapters, which are the Introduction, Literature Review, Methodology, Result and Discussion, Conclusion and Future Work.

Chapter 1 presents the background of banana plantations and major banana diseases in Malaysia. Besides, the general idea of this research and the introduction for five different object detection models are discussed in this chapter. This chapter also provides a problem statement and the objectives of this research work.

Chapter 2 explains the review based on different research findings similar to this research, mainly focusing on the topics of deep learning applications for plant disease diagnosis. The integration of deep learning with MAVs and comparisons for different object detection models are discussed as well.

Chapter 3 depicts the detailed steps to carry out this research. It explains the training and evaluation process for object detection models used within this research and discusses the improvement for the selected model.

Chapter 4 portrays the results for each object detection model and discusses the selection of the models used for real-world applications. Images are included to show the detected objects.

Chapter 5, which concludes the findings of this research, and recommendations are also included to give suggestions on how and where this research could be improved in the future.

CHAPTER 2

LITERATURE REVIEW

In recent years, the application of machine learning is widely implemented in various sectors, such as self-driving cars, virtual assistants, and healthcare. The versatility of machine learning applications makes the technology gain high popularity and acceptance by the public. In this digitalised world, the agriculture sector is implementing deep learning to improve the reliability and the sustainability of current solutions to agricultural problems, such as plant health monitoring and pest monitoring. The main objective is to increase the total yield of crops and hence bring wealth to farmers. A deep learning technique using CNN has the advantage of classifying and locating an instance of a particular object category. The use of object detection within the agriculture industry is to identify the types of diseases and stress of plants. Even the application is very powerful, but it requires a lot of datasets to perform better. The dataset collection has to be done on our own as the datasets for agricultural purposes that are publicly available are very limited. Therefore, developing a monitoring system is time-consuming and tedious as the datasets are needed to be processed and annotated to cater for our needs.

2.1 Application of Deep Learning for Plant Leaves Detection

The leaves of a plant can provide helpful information about the plant, such as the health status of the plant. Greenish and flourish leaves indicate the plant is healthy as the leaves provide adequate surface areas and chlorophylls for the plants to collect sunlight and carry out photosynthesis. Plants that have a low level of water stress possess wilted, folded, or misshapen plant leaves. Wilted and yellowing of leaves are a

sign of nutrient deficiency, and winding trails and holes found in leaves can be concluded that the plants are attacked by insects. Furthermore, banana plants are susceptible to various types of diseases which symptoms are primarily visible on the leaves.

The mostly used deep learning technique for plant leaves detection is CNNs due to the availability of a huge number of pre-trained CNN-based models. Identification of plant diseases and stresses through leaves has been the favourite for recent works of literature. Most works utilise CNNs for image classification, which predicts the class of one object in an image without identifying the location of the object in the given image (Kabir, Ohi and Mridha, 2021). VGG16 is the most popular network architecture being employed in the latest literature on plant leaves stress identification (Noon *et al.*, 2020). Resnet50, AlexNet, and GoogleNet gain the same popularity for leaf stress identification of various fruits and vegetables. Despite having different popularity of utilisation, proposed deep neural networks have achieved great recognition accuracy for almost all types of plants, generally above 90%, and the highest has reached 99.84% accuracy. There are two methods in developing neural network models, which are transfer learning based method and training the network from scratch. Transfer learning based methods have provided better recognition accuracy than when the network is trained from scratch (Noon *et al.*, 2020).

Dataset collection and annotation are crucial steps in designing a robust object detection model. To develop a model that is robust to the environment, field conditions images are preferable compared to laboratory conditions images. The field conditions images with different environmental conditions can be obtained by visiting several banana farms in disease hotspots. A plant disease detection and diagnosis research has shown the importance of using field conditions images (Ferentinos, 2018). The

researcher developed two CNN models in different ways, one is trained with solely field conditions images and tested on laboratory conditions images, and another is trained and tested reversely. The result shows that the first model, trained with only field conditions images, achieves a better performance than the second model (Ferentinos, 2018). The first model successfully calculates 68% of laboratory conditions images, whereas the second model has only 33% success rates. In addition, dataset collection is conducted in these banana farms to provide reliable and accurate images as publicly available datasets cannot fulfill our needs. To increase the accuracy of an object detection model, a dataset that consists of various types of images is required. For example, images with different resolutions, light conditions, and various environmental locations can significantly help the model in predicting the health conditions of plants accurately. Taking photos at different time, i.e., in the morning and the evening, is the way to have different illumination in a dataset. Furthermore, capturing a large number of photos of dried leaves at various plant growth stages helps the model to distinguish between dried leaves and unhealthy leaves. The background of images also plays an important role in developing a robust object detection model. The dataset should not consist of images of plant leaves in a controlled environment and with a simple background. Instead, except for plant leaves, the pictures should have dried leaves and branches on the soil to adopt any changes in the real-time environment. A huge variation in the dataset helps produce a generalised trained model, but only with the correct dataset annotation. Thus, plant experts are needed to conduct a robust labelling approach by confirming the typical symptoms on each image of the dataset (Selvaraj *et al.*, 2019).

The division of the weightage of a dataset into training and testing datasets affects the accuracy of models. The typical sizes for the training dataset range from 60%

to 90%, and 10% to 40% for the test dataset. A study in banana leaf identification has varied the training dataset from 80%, 60%, 50%, 40% to 20% using the same hyperparameters (Amara, Bouaziz and Algergawy, 2017). The result shows that the variation of the training dataset's weightage produces different accuracy of models, and the object detection model achieves the best accuracy at 40% of the training dataset. However, for a small dataset size with few training samples, it is recommended that the weightage of the training dataset should be larger to avoid model underfitting. Another research uses 90% of 5608 images to train a plant seedlings classification model and achieves a validation accuracy of 99.48% (Belal A.M. Ashqar, Bassem S. Abu-Nasser, 2019).

It can be seen that many studies regarding the application of deep learning were conducted for the use of plant leaves diseases and stress detection. A review highlights the broad implementation of image classification in identifying plant leaves stresses (Noon *et al.*, 2020). Yet, the utilisation of CNNs for object detection is essential in this project due to the implementation of a real-time monitoring system. Object detection helps the farmers to know which plants are categorised as unhealthy quickly by looking at the bounding boxes drawn in the video. Besides, none of the research has involved the detection of health status for banana seedlings. A robust plant health monitoring system is still absent, which instigates more studies to be done to create a reliable and functional plant health monitoring system for the agricultural industry, especially in banana seedlings.

2.2 Integration of Deep Learning with MAVs within Agriculture Sector

It is undeniable that the application of MAVs has boosted the sustainability within the agriculture sector with the capabilities to provide a different perspective on the crops as well as being able to cover more area in a shorter time compared to terrestrial solutions. The cost-effectiveness of data acquisition from MAVs can be used to rectify deficiencies encountered during the growing season of the crop, thus pushes MAVs into the precision farming field (Giridhar and Viswanadh, 2012). Currently, MAVs with flapping wings, such as the hummingbird-like and insect-like robot are preferred by many researchers as they own characteristics of small size, agility, and energy efficiency.

The integration of deep learning with MAVs has brought advantages in terms of autonomous real-time action. For example, researchers have developed an MAV for autonomous artificial pollination (Chen and Li, 2019). Deep learning is utilised to recognise and localise the position of a flower and extract specific characteristics to match against the flower database. A result will be decided and informed to the MAV whether the process goes to pollination or termination. A mechanical device is mounted to help in transferring pollen from the anther to the stigma.

The combination of deep learning and MAVs in plant health monitoring is very little, especially for banana seedlings. To fill this gap, this project proposes a health conditions monitoring system for banana seedlings by integrating deep neural networks and MAVs.

2.3 Comparison of Object Detection Models

Two performance evaluation metrics that are important when comparing object detection models, which are Average Precision (AP) and FPS. To standardise the comparison, the models are required to be trained on a similar dataset. The Microsoft Common Objects in Content (MS COCO) dataset is designed to serve as a gold standard benchmark for evaluating the performance of object detection models (Lin *et al.*, 2014).

A new version is believed to perform better than an older version in terms of detection accuracy and real-world applications. YOLOv4 improves AP and FPS of YOLOv3 by 10% and 12% (Bochkovskiy, Wang and Liao, 2020). YOLOv4 achieves 38 of FPS and 41.2% AP, whereas YOLOv3 achieves 35 of FPS and 31.0% AP when these two state-of-art architectures are tested on Maxwell GPU, as shown in Figure 2.1. For Tiny-YOLO versions, Tiny-YOLOv4 achieves better AP and FPS compared to Tiny-YOLOv3 when both models are tested on GPU GTX 1080 Ti (Wang, Bochkovskiy and Liao, 2020), as shown in Figure 2.2. Tiny YOLOv4 can achieve real-time performance on any embedded GPU device.

As stated before, the Tiny-YOLO version has better FPS but lower accuracy than the YOLO version. When running on GPU GTX 1080 Ti, Tiny-YOLOv4 achieves 22.0% AP at a speed of 371 FPS, which approximately two-thirds of YOLOv4's accuracy and ten times faster than YOLOv4.

Figure 2.3 gives a general view of the comparison of state-of-the-art YOLO models. The model that has the highest detection speed is Tiny-YOLOv4, followed by Tiny-YOLOv3 and YOLOv4, YOLOv3 has the lowest frame rate in real-time detection. YOLOv4 achieves the highest AP, followed by YOLOv3 and Tiny-YOLOv4, Tiny-YOLOv3 has the poorest performance in detecting objects accurately.

Method	Backbone	Size	FPS	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
YOLOv4: Optimal Speed and Accuracy of Object Detection									
YOLOv4	CSPDarknet-53	416	38 (M)	41.2%	62.8%	44.3%	20.4%	44.4%	56.0%
YOLOv4	CSPDarknet-53	512	31 (M)	**43.0%**	**64.9%**	**46.5%**	**24.3%**	**46.1%**	**55.2%**
YOLOv4	CSPDarknet-53	608	23 (M)	43.5%	65.7%	47.3%	26.7%	46.7%	53.3%
Learning Rich Features at High-Speed for Single-Shot Object Detection [84]									
LRF	VGG-16	300	76.9 (M)	32.0%	51.5%	33.8%	12.6%	34.9%	47.0%
LRF	ResNet-101	300	52.6 (M)	34.3%	54.1%	36.6%	13.2%	38.2%	50.7%
LRF	VGG-16	512	38.5 (M)	36.2%	56.6%	38.7%	19.0%	39.9%	48.8%
LRF	ResNet-101	512	31.3 (M)	37.3%	58.5%	39.7%	19.7%	42.8%	50.1%
Receptive Field Block Net for Accurate and Fast Object Detection [47]									
RFBNet	VGG-16	300	66.7 (M)	30.3%	49.3%	31.8%	11.8%	31.9%	45.9%
RFBNet	VGG-16	512	33.3 (M)	33.8%	54.2%	35.9%	16.2%	37.1%	47.4%
RFBNet-E	VGG-16	512	30.3 (M)	34.4%	55.7%	36.4%	17.6%	37.0%	47.6%
YOLOv3: An incremental improvement [63]									
YOLOv3	Darknet-53	320	45 (M)	28.2%	51.5%	29.7%	11.9%	30.6%	43.4%
YOLOv3	Darknet-53	416	35 (M)	31.0%	55.3%	32.3%	15.2%	33.2%	42.8%
YOLOv3	Darknet-53	608	20 (M)	33.0%	57.9%	34.4%	18.3%	35.4%	41.9%
YOLOv3-SPP	Darknet-53	608	20 (M)	36.2%	60.6%	38.2%	20.6%	37.4%	46.1%

Figure 2.1: Comparison of the speed and accuracy of different object detection models (Bochkovskiy, Wang and Liao, 2020). YOLOv4 and YOLOv3 with the size of 416 are compared, which are drawn in red boxes.

Model	Size	FPS_{1080ti}	FPS_{TX2}	AP
YOLOv4-tiny	416	**371**	**42**	21.7%
YOLOv4-tiny (3l)	320	252	41	**28.7%**
ThunderS146 [25]	320	248	-	23.6%
CSPPeleeRef [37]	320	205	41	23.5%
YOLOv3-tiny [30]	416	368	37	16.6%

Figure 2.2: Comparison of state-of-the-art tiny models (Wang, Bochkovskiy and Liao, 2020). Tiny-YOLOv4 and Tiny-YOLOv4 that are drawn in red boxes are compared.

A recent study (Magalhães *et al.*, 2021) has compared SSD MobileNet v2 and Tiny-YOLOv4 models to detect tomatoes in greenhouses. SSD MobileNet v2 has the best performance in term of accuracy, whereas Tiny-YOLOv4 achieves the best in real-time detection speed.

Nonetheless, this research will evaluate the performance of these object detection models and select the best by comprising the accuracy and the real-time detection speed. The performance of these models is most likely to be affected as they will be tested on a new dataset that consists of only banana seedlings and ran on a 2GB VRAM GPU.

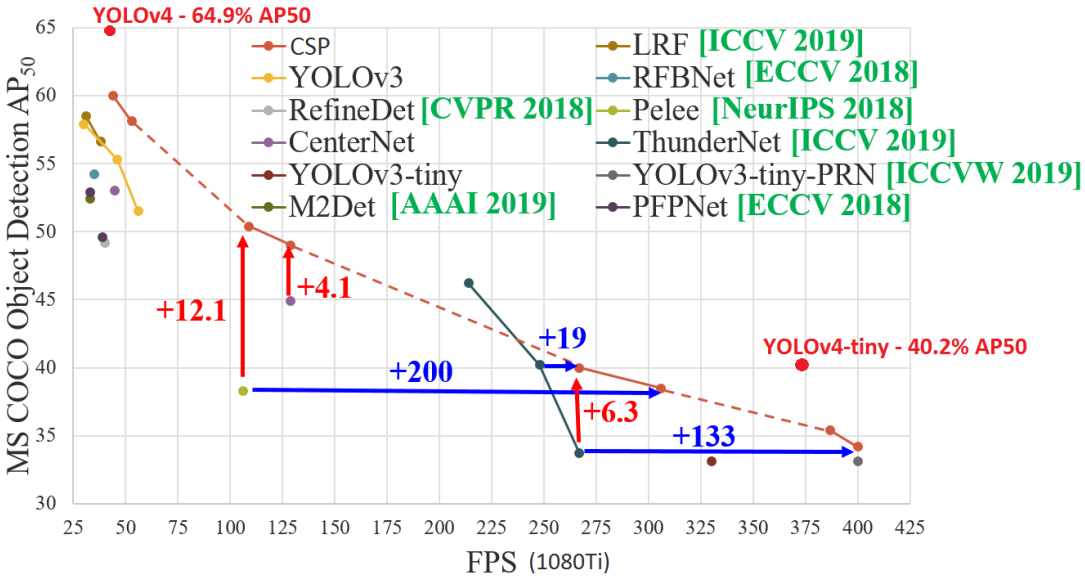

Figure 2.3: Comparison of the speed and accuracy of different models (AlexeyAB, 2020). Tiny-YOLOv4 runs at 371 FPS on GPU GTX 1080 Ti and achieves 40.2% AP50. YOLOv4 runs at 38 FPS and achieves 64.9% AP.