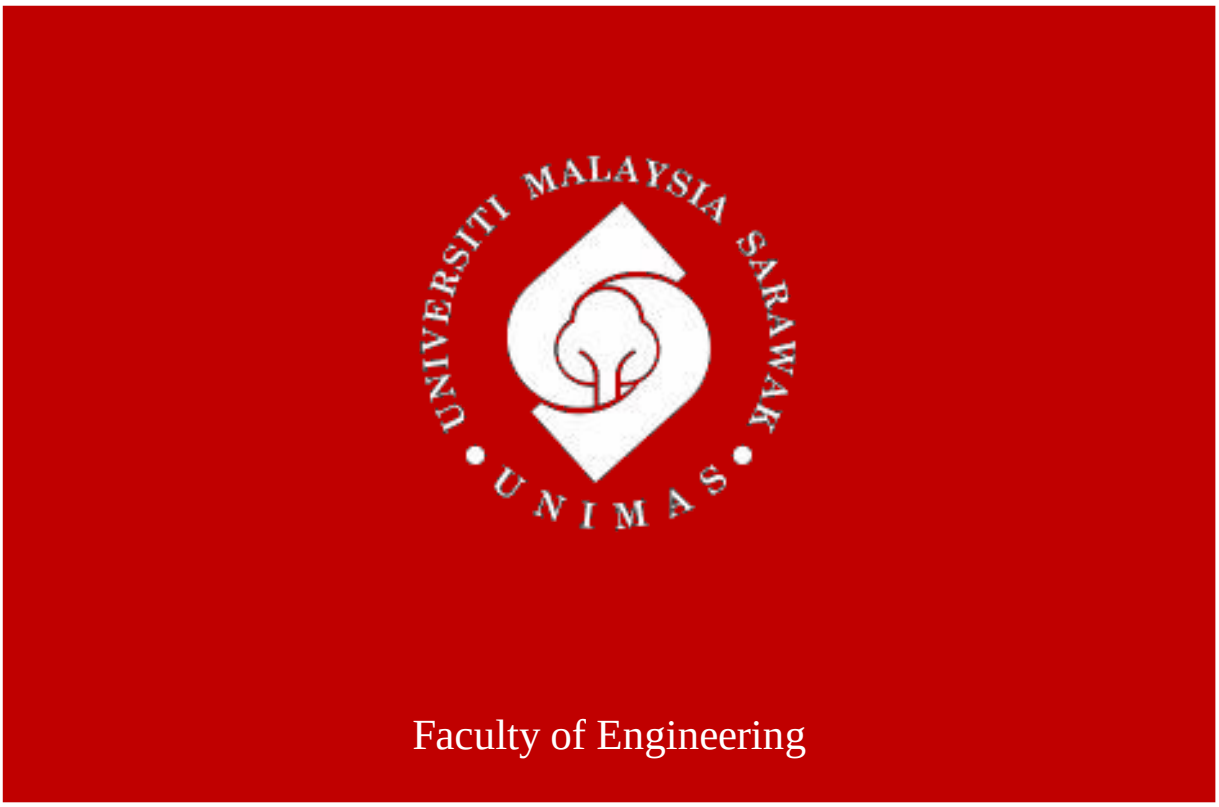

Comparison of Child Detection System with Artificial Intelligence using MobileNet, YOLOv2 and YOLOv3 for Object Detection

Don Perez anak Liap

Bachelor of Engineering with Honours

(Mechanical and Manufacturing Engineering)

2019

UNIVERSITI MALAYSIA SARAWAK

Grade: A

Please tick (√)
Final Year Project Report ☑
Masters ☐
PhD ☐

DECLARATION OF ORIGINAL WORK

This declaration is made on the 24th day of June 2018.

Student's Declaration:

I Don Perez anak Liap, 50467, Faculty of Engineering
(PLEASE INDICATE STUDENT'S NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled Comparison of Child Detection with A.I using mobilenet, YOLOv2 & YOLOv3 for Object Detection is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

24th June 2019
Date submitted

Name of the student (Matric No.)
Don Perez anak Liap (50467)

Supervisor's Declaration:

I Ir. Dr. David Chua Sing Ngie (SUPERVISOR'S NAME) hereby certifies that the work entitled Comparison of Child Detection with A.I using mobilenet, YOLOv2 & YOLOv3 for Object Detection (TITLE) was prepared by the above named student, and was submitted to the "FACULTY" as a * partial/full fulfillment for the conferment of Bachelor of Mechanical and Manufacturing Engineering (PLEASE INDICATE THE DEGREE), and the aforementioned work, to the best of my knowledge, is the said student's work.

Received for examination by: _____ Date: 24th June 2019
(Name of the supervisor)
Ir. Dr. David Chua Sing Ngie

I declare that Project/Thesis is classified as (Please tick (√)):

☐ CONFIDENTIAL (Contains confidential information under the Official Secret Act 1972)*
☐ RESTRICTED (Contains restricted information as specified by the organisation where research was done)*
☑ OPEN ACCESS

Validation of Project/Thesis

I therefore duly affirmed with free consent and willingness declare that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abiding interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies for the purpose of academic and research only and not for other purpose.
- The Centre for Academic Information Services has the lawful right to digitalise the content for the Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic exchange between Higher Learning Institute.
- No dispute or any claim shall arise from the student itself neither third party on this Project/Thesis once it becomes the sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student except with UNIMAS permission.

Student signature _____ (Date) 24/6/2019

Supervisor signature: _____ (Date) 24/6/2019

Current Address:
NO. 25 SCORPION PARK, JALAN TUN HUSSEIN ONN,
92000 BINTULU, SARAWAK

Notes: * If the Project/Thesis is CONFIDENTIAL or RESTRICTED, please attach together as annexure a letter from the organisation with the period and reasons of confidentiality and restriction.

[The instrument is duly prepared by The Centre for Academic Information Services]

COMPARISON OF CHILD DETECTION SYSTEM WITH ARTIFICIAL INTELLIGENCE USING MOBILENET, YOLOV2 AND YOLOV3 FOR OBJECT DETECTION

DON PEREZ ANAK LIAP

A dissertation submitted in partial fulfilment
of the requirement for the degree of
Bachelor of Engineering with Honours
(Mechanical and Manufacturing Engineering)

Faculty of Engineering
Universiti Malaysia Sarawak

2019

To my beloved family and friends.

ACKNOWLEDGEMENTS

I would like to express my special thanks of gratitude to my supervisor, Ir. Dr. David Chua Sing Ngie from Mechanical and Manufacturing Engineering Department, who gave me the golden opportunity to do this wonderful project on the topic of Comparison of Child Detection System with Artificial Intelligence using MobileNet, YOLOv2 and YOLOv3 for object detection, which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to him. I would also like to thank Google Inc for developing the object detection models API and allow it to be open source for students and researchers to develop in the field of Artificial Intelligence. Lastly, I would also like to thank my parents and friends who helped me a lot in finalizing this project within the limited time frame.

ABSTRACT

This research focuses on improving the child detection system by utilizing AI with recent versions of pre-trained models as an alternative of using sensors existing in the child detection system. Currently the problems experienced is the sensors used in the market, to prevent child heatstroke in automobiles, cannot accurately determines the occupant in position and whether the person is an adult or child. Comparing pre-trained models of object detection with AI such as Single-Shot Detector MobileNet (SSD MobileNet), YOLO version 2 (YOLOv2), and YOLO version 3 (YOLOv3) could suggests a more accurate and precise child detection system. The system with three different object detection models were tested experimentally to evaluate the speed, accuracy and precision. At the end of experiments, it is founded that YOLO able to detect custom objects the fastest which is less than a second. Also, YOLOv3 Tiny GPU has the best average score of detection which is 100% at the first 80 cm while SSD MobileNet has its highest average score for detection at 70 cm. At the best distance, which is 70cm, SSD MobileNet shows an acceptable result since there is no false detection, while YOLO shows perfect reproducibility result at 70 cm. In conclusion, YOLOv3 is the most suitable model to improve the framework of the child detection system.

ABSTRAK

Penyelidikan ini memberi tumpuan kepada meningkatkan sistem pengesanan kanak-kanak dengan menggunakan AI dengan versi terkini model pra-terlatih sebagai alternatif menggunakan sensor yang ada dalam sistem pengesanan kanak-kanak. Pada masa ini, masalah yang dialami adalah sensor yang digunakan di pasaran, untuk mengelakkan tekanan darah kanak-kanak di dalam kereta, tidak dapat menentukan dengan betul penghuni dalam kedudukan dan sama ada orang itu adalah orang dewasa atau anak. Membandingkan model pengesanan objek pra-terlatih dengan AI seperti Pengesan Single-Shot MobileNet (SSD MobileNet), YOLO versi 2 (YOLOv2), dan YOLO versi 3 (YOLOv3) boleh mencadangkan sistem pengesanan anak yang lebih tepat dan tepat. Sistem yang mempunyai tiga model pengesanan objek berbeza diuji secara eksperimen untuk menilai kelajuan, ketepatan dan ketepatan. Pada akhir eksperimen, ia diasaskan bahawa YOLO dapat mengesan objek tersuai yang terpantas yang kurang dari satu saat. Juga, GPU Tiny YOLOv3 mempunyai skor pengesanan purata yang terbaik iaitu 100% pada 80 cm pertama manakala SSD MobileNet mempunyai skor purata tertinggi untuk pengesanan pada 70 cm. Pada jarak yang terbaik, iaitu 70cm, SSD MobileNet menunjukkan hasil yang boleh diterima kerana tiada pengesanan palsu, sementara YOLO menunjukkan hasil reproduktif yang sempurna pada 70 cm. Kesimpulannya, YOLOv3 adalah model yang paling sesuai untuk memperbaiki kerangka sistem pengesanan kanak-kanak.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	-	Artificial Intelligence
ANN	-	Artificial Neural Network
AP	-	Average Precision
API	-	Application Programming Interface
CE	-	Cross-Entropy
CNN	-	Convolutional Neural Network
COCO	-	Common Objects and Context
CPU	-	Central Processing Unit
DL	-	Deep Learning
DNN	-	Deep Neural Network
FPS	-	Frames Per Second
GPU	-	Graphics Processing Unit
IDLE	-	Integrated Development Environment
IOU	-	Intersection Of Union
mAP	-	Mean Average Precision
ML	-	Machine Learning
RNN	-	Recurrent Neural Network
SSD	-	Single-Shot Detector
SSE	-	Sum Squared Error
YOLO	-	You Only Look Once

CHAPTER 1

INTRODUCTION

1.1 General Background

In 2018, a total of 48 children in the United States of America loses their life from heat stroke after being abandoned in a vehicle (Null, 2018). Statically, that increases the average of child heatstroke casualties from 37 to 38 in the year 1998 to 2017 and 2018. A body temperature of 105°F (40.6°C) or higher, central nervous system dysfunction, and, in many cases, cessation of sweating is also known as the devastating illness of heatstroke. Heat stroke is an extreme medical emergency that will result in death if untreated in most patients. In addition, physiological differences in children compared to adults make them particularly susceptible to illnesses caused by heat exposure.

The internal environment of motor vehicles has been investigated in past studies. King et al (1981) discovered that the environment temperature of 36.8°C maximized to 51 to 67°C in 15 minutes where three quarters of the maximum temperature rise occurred within 5 minutes of closing the doors. To add more, it had very little effect on the temperature rise and maximum temperature achieved by opening the windows 20cm. Similar findings by Roberts and Roberts (1976) showed that there was significant heat rise in automobiles on hot days however with lower ambient temperate which is about 28°C. Despite all these facts, 400 children suffered fatality from heatstroke in the U.S as they were left forgotten in the vehicle by their caretakers as recorded by Null.

Car technologies have improved over the years by reminding the driver of everything such as to remember the car key, to turn off the head light or to put on the seatbelt but technology have yet reached its perfection to remind the driver of their children in the back seat. Some company such as Evenflo invent products to help parents with the safety issue of children and one of their product is SensorSafe Technology. It is a device that reminds the driver that a child is in the car seat upon arrival and will alert the driver if a child unfastens the

chest clip while in transit. The gadget is installed by plugging the wireless receiver into the vehicle's On-Board Diagnostic (OBD) Port. Then, the chest clip is fastened to secure the child. Once driving, SensorSafe automatically connects to alert the driver if a child unfastens the clip. Upon arrival, the device will remind the driver that a child is in the car seat.

Figure 1.1: Evenflo Sensorsafe Technology. (Adapted from Evenflo)

However, there are weaknesses in this invention. The device works by detecting the connection of the chest clip. If the child unfastens the chest clip, it could disturb the attention of the driver while on the road. Other possibilities where the child could slip out of the chest clip and not trigger the alarm also can be taken as a risk factor. In terms of its installation, not all automobile has OBD Port that is compatible with SensorSafe Technology as stated by Thomas where their team founded that the device has difficulties connecting consistently with car models from when used with some models from Acura, Dodge, Kia, Land Rover, Mazda, and Subaru. Besides, SensorSafe Technology requires a baby car seat in order to accommodate the chest clip.

Next invention is developed by General Motors (GM) which is their Rear Seat Reminder system. The feature uses back door sensor that activates when either rear door open or close within 10 minutes of the vehicle being started or while it is running which mean the system does not actually detects objects in the rear seat. The driver will receive a remainder on the dashboard when they arrive at their destination. GM was looking for a low-tech solution for this problem by reminding the driver of checking the rear seat. This method can be a good way to train good habit however it is not practical because the Rear Seat Reminder's feature is active only once each time the vehicle is turned on and off and would require re-activation on a second trip. In other words, when the automobile is restarted without opening the back door, the remainder will not trigger when the automobile is turn off again. Also, this feature is only available to new GM models with electronic dashboard.

Figure 1.2: General Motor's reminder system. (Adapted from GMC)

These products are one of the many example of inventions to prevent parents from abandoning their children in the rear seat which generally incorporate of mechanical sensor to complete a circuit. There are also previous research attempts in implementing usage of infrared sensors to detect children which is similar to certain baby monitors. However, it is quite risky to expose children to infrared in a closed space. The ability to correctly detect the presence of a human being be it a child or an adult is the challenge for all these products. Subsequently, it is needed to develop a child detection system which not only accurately detect the presence of a child in a car but fast and precise. Object detection with Artificial Intelligence (AI) could tackle the problem faced by the existing technologies. Since utilization of AI is not something new in this era, suitable existing models can be analysed to achieve the objectives of this research.

Table 1.1: Advantages and Disadvantages of Existing Technology on Child Reminder System

Car Technology	**Advantages**	**Disadvantages**
Evenflo Sensorsafe Technology	• Easy installation	• The clip is still accessible to the children • Device is not compatible with all car models • Requires a baby seat

GM Reminder System	• Does not require any installation as it is built-in sensors	• The remainder will not trigger when the automobile is turn off again • Only available to new GM models with electronic dashboard

1.2 Problem Statement

Currently the problems experienced is the sensors used in the market, to prevent child heatstroke in automobiles, cannot accurately determines the occupant in position and whether the person is an adult or child. The true purpose of the technologies so far is to remind the driver to check their passenger seat by generating a constant signal at the end of every driving session to alert the driver. However, false signal generation or impractical set up of the system could distract the driver as the device cannot detect the nature of the occupant.

Detecting children cannot be solely rely on mechanical sensors because some children like to move a lot, or they could be asleep which either triggers the system or not functioning at all, thus reducing the reliability of the detection system. Comparing pre-trained models of object detection with AI could suggests a more accurate and precise child detection system.

Moreover, problem to be faced in this project is the qualities of input which will affect the speed and accuracy of results. This problem may occur due to motion of children. Other factor such as brightness, distance and angle from the camera to the target would affect the performance of the system as well.

1.3 Objectives

- To improve the framework for the development of a child detection system with AI using object detection method
- To analyse different object detection pre-trained models such as MobileNet, YOLOv2, and YOLOv3 in terms of their accuracy and speed
- To select suitable pre-trained models for child detection system

1.4 Scope of Research

This research focuses on improving the child detection system by utilizing AI with recent versions of pre-trained models as an alternative of using sensors existing in the child detection system. The system established will response to the presence of child only. This research will help to compare other versions of pre-trained models of object detection in child detection system. The pre-trained object detection models chosen is MobileNet, YOLOv2 and YOLOv3 which roughly comparing one of the earliest version against the latest version. There are several common challenges that may not be solved by using object detection. Firstly, training an object detection system requires a large amount of images as training data. Besides that, training an image recognition model required the usage of Graphics Processing Unit (GPU) in order to speed up the process and obtain a more accurate retrained model.

CHAPTER 2

LITERATURE REVIEW

2.1 History of Artificial Intelligence

The field of artificial intelligence, or AI attempts not just to understand but also to build intelligent entities (Russell & Norvig, 1995). AI is one of the newest fields in science and engineering. Work started in earnest soon after World War II, and the name itself was coined in 1956. AI currently encompasses a huge variety of subfields, ranging from the general (learning and perception) to the specific, such as playing chess, proving mathematical theorems, writing poetry, driving a car on a crowded street, and diagnosing diseases. AI is relevant to any intellectual task; it is truly a universal field. The first work that is now generally recognized as AI was done by Warren McCulloch and Walter Pitts (1943). They drew on three sources: knowledge of the basic physiology and function of neurons in the brain; a formal analysis of propositional logic due to Russell and Whitehead; and Turing's theory of computation (Russell & Norvig, 1995).

With the emergence of Machine Learning ((ML), an approach to achieve AI, the field of AI reached its major advance in the year of 1980 (Dietterich, & Michalski, 1983). Using algorithms to learn from examples is the working principle of machine learning. Without explicitly programmed, the machine is trained to learn form a large amount of data by itself. Nevertheless, machine learning's ability to process natural data in their raw form was limited in the conventional machine learning techniques (Luppescu & Romero, 2017). The development of Deep Learning (DL) in the 2000's was due to the limitation of machine learning techniques. Enabling the machine to be fed with raw data and automatically discover the representations needed for detection and classification is the set of methods by DL (Lecun, Bengio, & Hinton, 2015). To extract a useful representation of data, DL architectures comprised of multiple processing layers. For example, raw images are fed into the learning model, in image processing. Extraction of edge detection is performed by the initial layer, detection of motif is executed by the second layer, assembling the motifs into larger

combinations that correspond to parts of familiar objects is achieved by the third layer, while detection of objects as combinations of these parts are accomplished by the subsequent layers. Any techniques that enable computers to replicate human intelligence is AI in short. Involving the field of study that gives computers the ability to learn without explicitly programmed is machine learning which makes it the subset of AI. Allowing a model to learn the representations of data with multiple levels of abstraction is DL form the improvement of ML. The relationship between AI, ML and DL is shown in Figure 2.1.

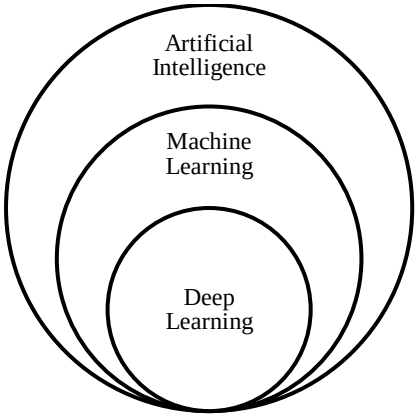

Figure 2.1: Relationship between AI, ML and DL

2.2 Deep Learning

The learning model needs to be instructed in machine learning and only then the model gives an accurate prediction. Feeding data to the model allow this method to be achieved. However, similar to human, DL model is able to learn by itself. A layered structure of algorithms called artificial neural network is used by DL to achieve the learning process. To simulate the biological neural network in the human brain is the initial purpose for artificial neural network. Enabling computational models to learn the representation of data with multiple layers of abstraction is the process of DL. It possesses a set of methods, starts from the raw material input, the raw material is transformed to the machine which automatically discovers representation needed for detection or classification (Lecun, Bengio, & Hinton, 2015). A feature extractor help perform this process. In the internal representation process, the extracted features are further transformed to slightly more abstract level. The classifier which composes of higher layers of representation is the location for the features being sent. While irrelevant variations are suppressed, the aspects which identified as useful input are amplified

in the classifier. Users are required to train the machine in image recognition using DL in order to have the ability to classify various images scanned into it. A large collection of data set and labelling it with the category as the desired output is required by the users to expose to the machine at the same time. To compare the error between the output scores and the desired pattern of scores, the objective function is computed. The machine tries to modify internal adjustable parameters with the error, to reduce errors in the next recognition. The input-output function is defined from these adjustable parameters or typically called weights. The creation of Convolutional Neural Networks (CNN) was brought from the integration of processes previously described. It is a network which connects the different layers of learned features as a whole.

2.2.1 Perceptron

Perceptron is a mathematical model with function similar to the biological neuron. It is a standard paradigm for statistical pattern recognition and the basic block of Artificial Neural Network (ANN) (Learning, & Learning, 1999). Input, activation and output processes are the 3 processing steps occurred within a single perceptron. The mathematical model will compute a weighted sum of the input signal and generate a binary output when the inputs in the form of signals enter the perceptron. If the weighted sum is above a certain threshold, it will give an output of 1. The model will give a zero output if it's apart from that (Jain & Mao, 1996). The three processing steps within a perceptron is shown in Figure 2.2.

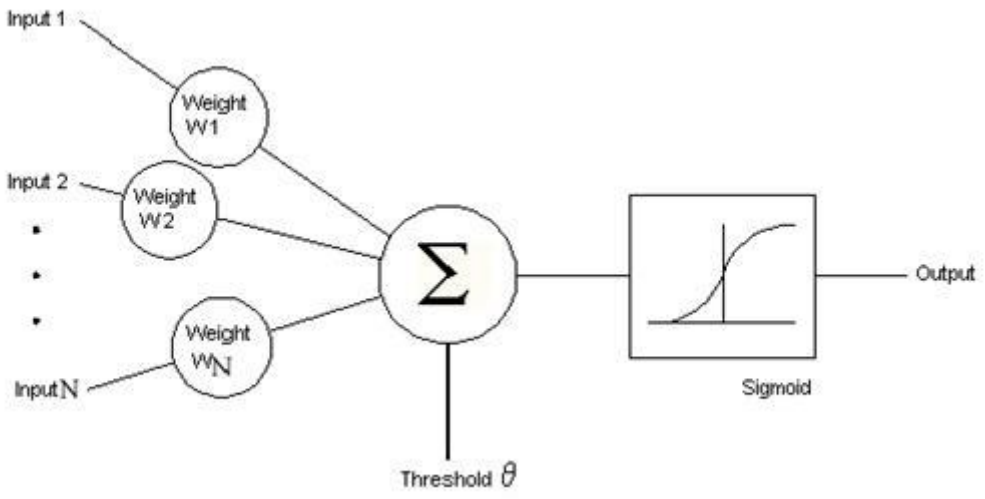

Figure 2.2: The processing steps in a perceptron. Adapted from (O'Riordan , 2005)

The processing steps can be expressed mathematically as the output, y for n numbers of inputs, x operating under a threshold function:

$$y = \theta\left(\sum_{j=1}^{n} w_j x_j - u\right)$$

Where Θ is a unit step function at 0 and j = 1,2,3, …, n.

Automated robots and vehicles control, face identification, speech recognition, text translation, gaming, and so on, are the application of neural network. The ANN was constituted by multilayer perceptrons. More than one perceptron composes each layer which receives input values, collectively called the input vector of that particular perceptron. The weight vector of that perceptron is collectively identified as the output of each perceptron, called weight. The input data, typically called the input layer connect with the first layer of the neural network. Hidden layers are the layers which are not directly connected to the environment. To transmit a signal to adjacent layers, without any processing on the input is the function of hidden layers. The activation function is connected to the final output. The sigmoid function is currently the most used function (Popescu, Balas, Perescu-Popescu, & Mastorakis, 2009).

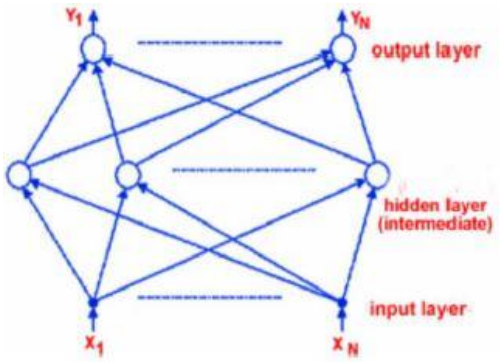

Figure 2.3: ANN's architecture. Adapted from (Popescu, Balas, Perescu-Popescu, & Mastorakis, 2009)

2.2.2 Deep Neural Network

An artificial neural network composed of numerous hidden layers is called Deep Neural Network (DNN). The difference between neural network and DNN is shown in Figure 2.4.

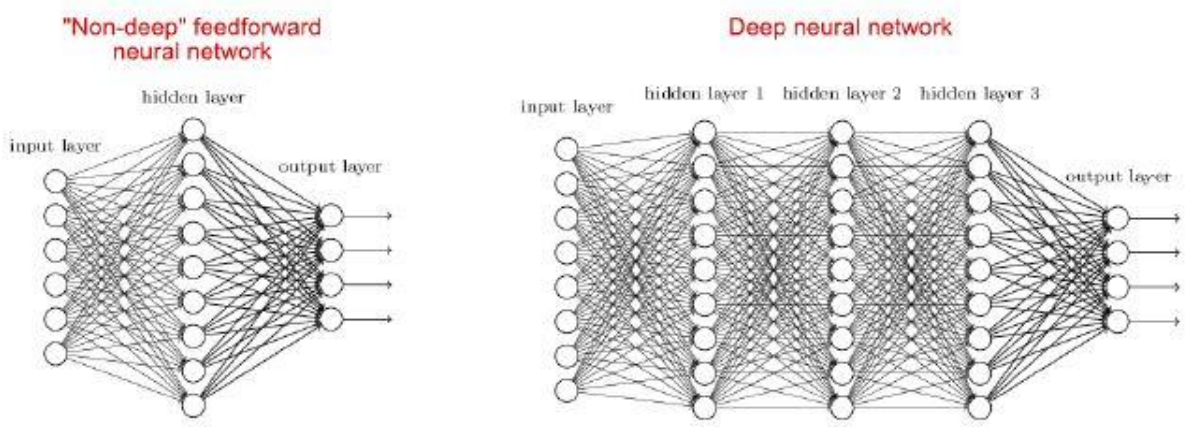

Figure 2.4: Difference between neural network and DNN. Adapted from (Nielsen, 2015)

Possessing two or more hidden layers made up the name for DNN. Specific types of sorting and ordering tasks are performed by each hidden layer. By performing hierarchy feature extraction, DNN has the ability to deal with unstructured or unlabelled data. ANN can be grouped into two categories based on the different types of architectures, which are Convolutional Neural Network and Recurrent Neural Network.

2.2.2.1 Convolutional Neural Network

Perceptrons are organised into layers that have unidirectional connections between them in feed-forward networks. Designed to process data that come in the form of multiple arrays, the convolutional neural network is an example of a feed-forward neural network. Colour image is an example of such data as the computer stores image as tiny squares. Small squares called pixels made up an image. Defined by a set of numbers, each pixel is only one colour. A combination of three colours, namely red, green and blue is represented by a set of numbers called an RGB image. The colour in a pixel is represented by three 8-bit numbers, in the range of 0 – 255 in an RGB image. For example, an orange colour is represented in an array of [255, 165, 0] as orange is produced by a combination of red and yellow, whereas yellow [255, 255. 0] is a combination of green and red colours. A white colour is displayed if the RGB is set to full intensity [255, 255, 255]. However, a black colour is displayed if the RGB is muted [0, 0, 0]. Three 2D arrays containing pixel intensities in the RGB channels compose produce a colour image. As such, to do recognition task, CNN takes advantage of the natural signal properties. By composing lower level features to achieve higher-level features as shown in