Tiia Mikkonen

Capacity Monitoring using Object Detection Algorithms

Metropolia University of Applied Sciences

Bachelor of Engineering

Information Technology

Bachelor’s Thesis

22 October 2021

Abstract

Author:	Tiia Mikkonen
Title:	Capacity Monitoring using Object Detection Algorithms
Number of Pages:	44 pages
Date:	22 October 2021
Degree:	Bachelor of Engineering
Degree Programme:	Information Technology
Professional Major:	IoT and Cloud Computing
Supervisors:	Janne Salonen, Head of Department (Metropolia UAS)

Due to the COVID-19/coronavirus pandemic, which started December 2019, in China, many governments around the global have imposed restrict lockdowns across the global along with mask mandates to slow down the spread of the virus. This thesis explores capacity monitoring with the use of the object detection algorithm “You Only Look Once”, more commonly known as YOLO. With the use of real-time CCTV and object detection it would be possible to accurately and quickly determine the capacity of an establishment or of an area, along with detecting whether patrons are wearing masks or even if there are pets there.

The topics of artificial intelligence, machine learning, deep learning, neural networks, artificial neural networks, convolutional neural networks, computer vision, YOLO, raspberry pi, python, TensorFlow and OpenCV are also discussed to help understand how they work and how everything is interconnected.

Finally, the process of setting up a raspberry with camera, installing the OS, installing all of the required libraries, along with YOLO, TensorFlow, and OpenCV are gone through in detail.

Keywords: artificial intelligence (AI), machine learning, deep learning, neural networks, artificial neural networks, convolutional neural networks, computer vision, YOLO, raspberry pi, python, tensorflow, opencv.

Contents

List of Abbreviations

ANN:	Artificial Neural Networks.
CNN:	Convolutional Neural Networks.
COVID-19:	Coronavirus.
DL:	Deep Learning.
HOG:	Histogram of Oriented Gradients.
ML:	Machine Learning.
NN:	Neural Networks.
OD:	Object Detection.
R-CNN:	Region-based Convolutional Network method.
SDD:	Single Shot Detector.
SPP-Net:	Spatial Pyramid Pooling.
WHO:	The World Health Organization.
YOLO:	You Only Look Once.

1 Introduction

With this thesis, a solution for capacity monitoring that makes use of algorithms that identify and recognise diverse objects in real time will be discussed. Among the object detection solutions that will be discussed, "You Only Look Once" will be the main point of emphasis. There are a variety of additional techniques that may be used for object detection, but not all of them can be discussed in detail here due to space constraints. This presentation will provide a brief overview of the following technologies: Histogram of Oriented Gradients, Region-based Convolutional Neural Networks (including Fast R-CNN and Faster R-CNNS), Region-based Fully Convolutional Network and Single Shot Detector, and Spatial Pyramid Pooling, which are all examples of machine learning techniques. It is widely acknowledged that "You Only Look Once," often known as YOLO, is the most widely used object recognition technique in use today.

Today's society recognises the significance of capacity monitoring because of the virus known as COVID-19 or coronavirus, which is a disease caused by the severe acute respiratory syndrome coronavirus 2, which is shortened as SARS-CoV-2. This virus was the source of a pandemic that began in Wuhan, China, in December of this year. During a press conference on March 11, 2020, the World Health Organization (WHO) proclaimed COVID-19 a worldwide epidemic. Following that, governments throughout the globe began adopting country-wide lockdowns in order to limit the spread of the illness. COVID-19 is a contagious disease which is the result of a virus that mainly affects the respiratory system [1]. The symptoms of the COVID-19 have changed since the first confirmed case due to the disease mutating multiple times, however it is commonly identified by a dry cough, fever and fatigue. Some prevalent symptoms might include dizziness, sore throat, headache, breathing difficulties, loss smell and/or taste, muscle pain, nausea, vomiting and diarrhea [1]. It is transmitted through the air via contaminated droplets and airborne particles which can infect anyone within a close proximity of the infected individual, especially in a poorly ventilated indoor space [1]. A positive diagnosis is given by a PCR test which detects the presence of viral RNA fragments. Home testing became a common testing method later on which involves the use of an antigen test.

During March 2020, the world went into a global lockdown in order to slow down the spread of COVID-19. The WHO started announcing recommendations of social distancing, self-isolation/quarantine, adequate hand hygiene, and the use of face masks [1]. Medical faculties across the globe also started developing vaccines to help prevent infections and/or serious illness. Many governments started imposing country-wide lockdowns along with remote work recommendations to slow down the spread. Once establishments were allowed to open, capacity limitations were set in places such as restaurants, event venues, private gathering, outdoor gatherings, and public transit in some cases. As of the 18th of April 2022, more than 504 million cases of COVID-19 have been confirmed across the globe with about 6.22 million deaths [2]. Europe itself has gone through at least 3 waves of lockdowns since the start of the pandemic with varying restrictions imposed by each countries’ governments. This is where object detection algorithms would come in handy to help businesses and patrons follow restrictions along with making smart decisions based off the capacity of an establishment along with the mask use within that facility.

With the use of YOLO, a business can easily comply with government restrictions put in place. Another possible use is to see how prominent the use of face masks is in the establishment. The technology also has the possibility of other uses such as rush hours prediction, allowing patrons to see how busy the establishment is before arriving, even allow patrons to know whether there are any animals in the establishment if they suffer from allergies or asthma, or even analyse the popularity of each entrance/exit.

2 Theoretical Background

In order to get a better understanding of the solution described in this thesis, the following topics should be discussed beforehand: machine learning (ML), deep learning (DL), object detection (OD), neural networks (NN), computer vision, raspberry pi and python. The purpose of the following subchapters is to give a brief explanation of the topics and how they relate to subject at hand.

2.1 Machine Learning

The study of computer algorithms that unconsciously develop themselves based on experience and conclusions is known as machine learning (ML). It's considered a part of artificial intelligence. ML algorithms construct a model using sample data in order to make predictions or judgments without “being explicitly programmed” to do so [3]. The data used is commonly referred to as training data.

The purpose of ML programs is to perform tasks without being told to do them based off data previously provided and learning based off it. Software Engineers can develop algorithms that allow computers to complete simple tasks; however, it is much too complex to develop algorithms that can carry-out more integrate tasks. It is possible to utilise these algorithms to comprehend a cyber phenomenon, abstract that knowledge into a model, anticipate future values of a phenomenon using the model created above, and identify abnormal behaviour shown by a phenomenon under observation [4].

ML is closely connected to artificial intelligence (AI), data mining, optimization, generalization and statistics. There are quite a few types of learning associated with ML, such as: supervised, unsupervised, semi-supervised, self-supervised, which are the most common ML methods. Some others are reinforcement multi-instance, inductive, deductive, transductive, multi-task, active, online, transfer, and ensemble, along with new methods developed in the recent years. The learning system is broken down into three majority components: a decision process, an error function and a model optimization process [5].

ML algorithms are popular in various real-world applications including computer vision, medical imaging, spam email filtration and speech recognition. They can also be seen when tasks are too difficult for a human to complete. They are seen nowadays in self-driving vehicles, dynamic pricing, fraud detection, voice assistants such as Siri, Alexa and Google Assistant, personalized marketing, chatbots, along with other various use cases.

2.2 Deep Learning

Being a subdivision of artificial intelligence, deep learning (DL), is basically a neural network with 3 or even more layers [6]. In order to pick up from enormous amounts of information, these types of networks intend to duplicate the practical capabilities as well as choice trees of a human mind. Nonetheless, they are still rather much from having the ability to totally duplicate the internal functions of one.

With a neural network with a single layer, it is still possible to produce approximate predictions. However, when more hidden layers are available it is a possibility to increase its accuracy; they are able to assist in optimization and enhancing the network's performance [6].

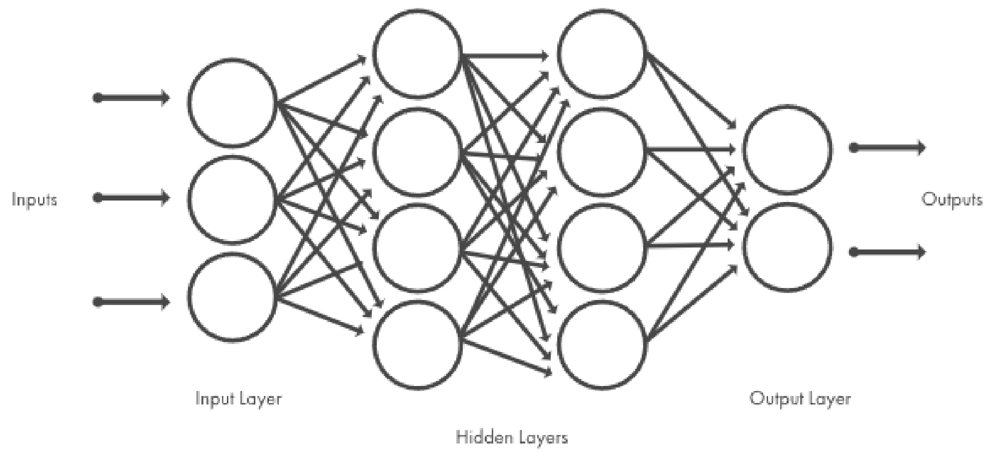

Figure 1. Most basic and common neural network [6].

Each level of DL learns to convert its input data into a more abstract and composite representation and survives until the data is entirely altered. An image recognition application may start with a matrix of pixels as input. The first layer may abstract the pixels and encode edges; the second layer may compose and encode arrangements of edges; the third layer may encode a nose and eyes; the fourth layer may recognise that the image contains a face [8]. A DL process may learn on its own which characteristics should be ideally placed at which level. The need for manual adjustment does not go away, however; for example, adjusting the number of levels and the size of the layers might give varied degrees of abstraction.

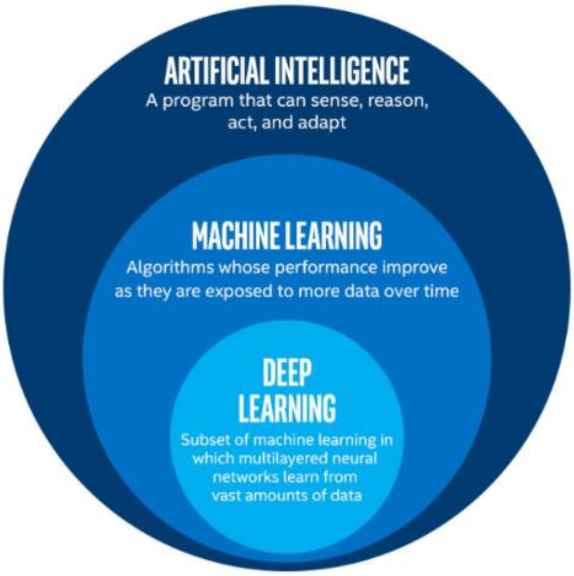

Figure 2. AI vs ML vs DL [9].

On the other hand, DL is just a sort of ML that tries to replicate the structure of a person’s brain, and it is not a new concept. Deep learning algorithms aim to form conclusions that are comparable to those reached by humans by continuously examining data with a predetermined logical framework. Deep learning does this by using neural networks, which are multi-layered structures of algorithms [9]. DL should be applied when solution experts are not available or when they are unable to explain their decisions in cases such as stock preferences or price predictions. In Figure 3, the differences between ML and DL can be seen.

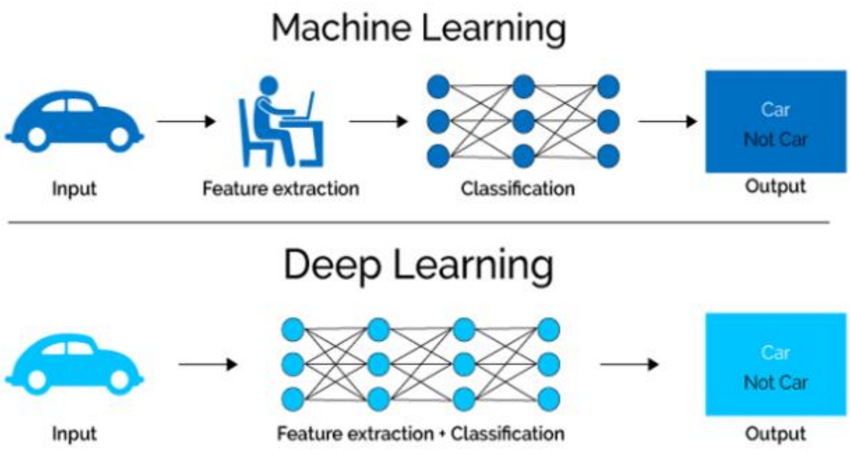

Figure 3. Differences between ML and DP [8].

Deep learning can be found in self driving cars, news aggregation along with fake news detection, virtual assistants, visual recognition, health care, fraud detection, along with various other uses.

2.3 Neural networks

Neural networks which are in deep learning techniques are used to model data. Their structure and name originate from the brain, where neurons convey data. They are also known as artificial neural networks (ANNs). A neural network can be seen in Figure 4 below.

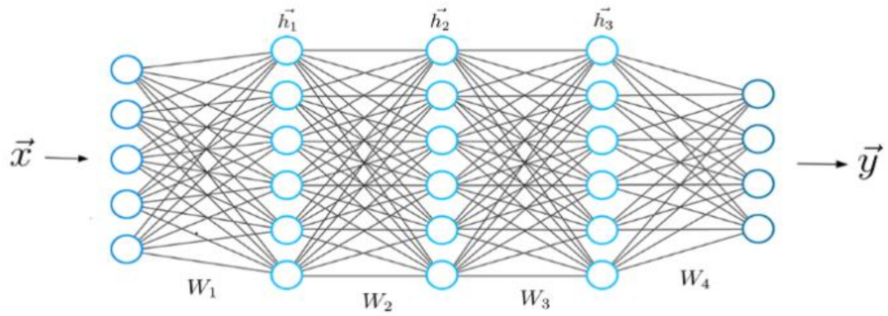

Figure 4. An artificial neural network [9].

Many different types of neural networks exist, such as: deep feed-forward (DFF), feed forward (FF), perceptron (P), radical basis network (RBN), recurrent neural network (RNN), along with various other types seen in Figure 5 below [10].

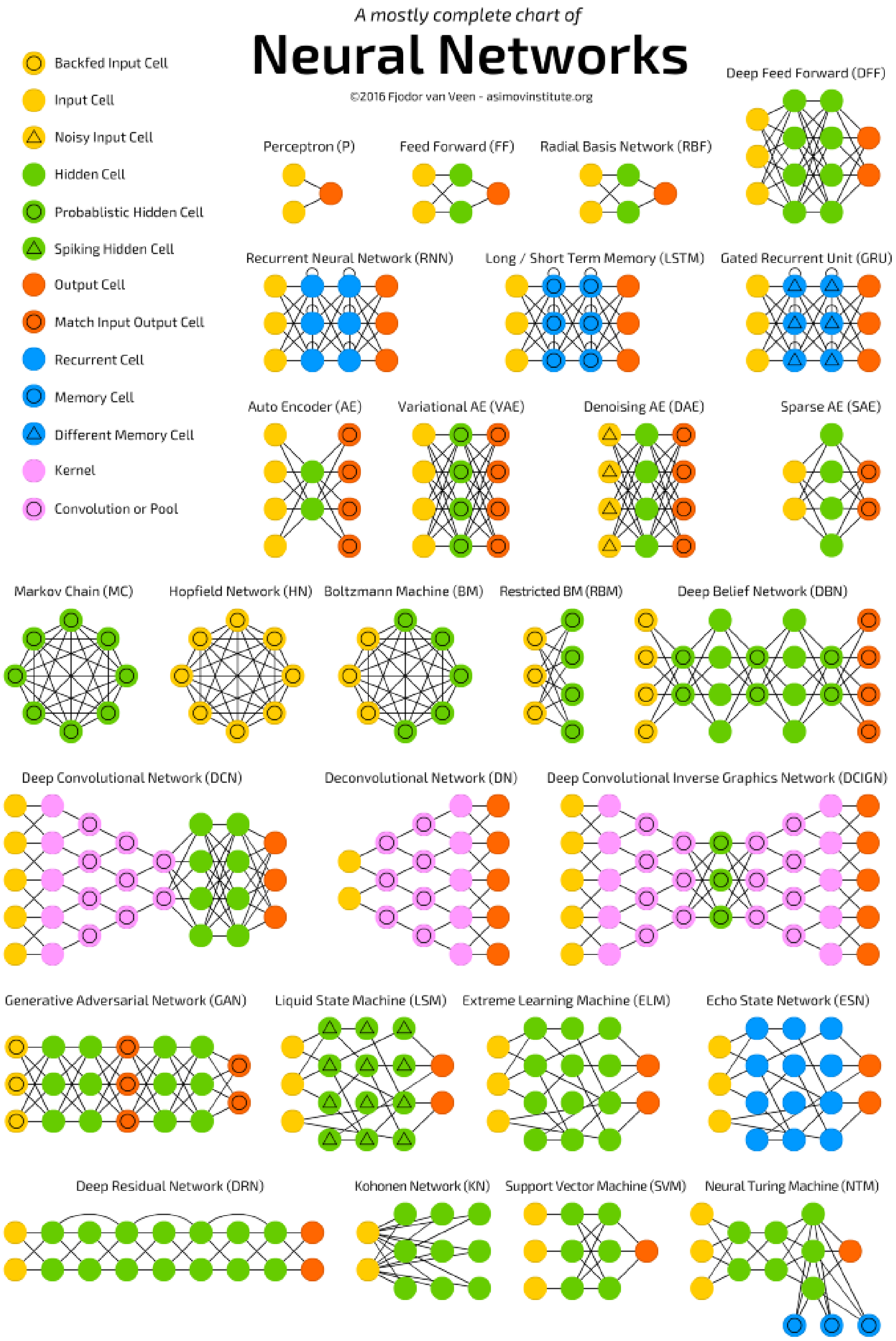

Figure 5. Various types of neural networks [11].

In the following sections, artificial neural networks (ANNs) and convolutional neural networks (CNNs) will be discussed. For the purpose of this thesis, convolutional neural networks (CNNs) will be discussed in more detail, which happen to be the most used out of all neural networks.

2.3.1 Artificial Neural Networks (ANNs)

Artificial neural networks (ANNs) consist of various node layers. Each of those nodes most consist of the following components, but are not limited to: input layer, hidden layers, and an output layer [10]. Each node, otherwise known as an artificial neuron, corresponds with the others and has a threshold and weight. The defined threshold is the minimum requirement that determines whether the node is active, before allowing data to be sent to the next layer of the network from there. If the defined threshold is not met, then no data is sent on to the next portion in the hierarchy.

2.3.2 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are used to classify pictures, cluster images, and recognise objects. Some use cases include computer vision, image recognition, video analysis, speech recognition, time series forecasting and anomaly detection. The neurons in a convolution neural network are arranged in three dimensions rather than two. The first layer is convolutional. The convolutional layer's neurons only process a small portion of the visual field. It takes in input features like a filter. The network can process pictures in portions and can calculate these actions numerous times. Processing converts a picture from RGB or HSI to greyscale. Changing the pixel value further helps recognise edges and categorise photos [12]. CNNs are advantageous since they are capable of learning the filters automatically without being distinctly mentioned, therefore extracting the respective data inputted. Since CNNs capture spatial features of a subject, the relationship of the arrangement of pixels in the image, it helps identify the subject as accurately as possible. CNNs are made up of three layers [12]:

- Convolutional layer
- Fully-connected (FC) layer
- Pooling layer

The convolution layer is the CNN's foundation, therefore carrying computational burden of that network. A dot product between two matrices is made, one representing the kernel, generally smaller than an image yet deeper, which consists of a set of learnable parameters, and another representing the limited area of the receptive field. If a picture has three (RGB) channels, the kernel width and height are small, but its depth is bigger. The kernel moves over the image's height and breadth, known as the receptive region, once the forward pass is initialized. An activation map is generated by this action, which is a two-dimensional representation of the picture that visualizes the kernel response at each location. A stride is the kernel's sliding size [14].

“When given an input of Wout times Wout times Dout and the amount of kernels of spatial size F, stride S, and padding P, the output volume may be calculated as follows” [14]:

$$W_{out} = \frac{W - F + 2P}{S} + 1$$

Figure 6. Convolution Layer Formula [14].

The pooling layer substitutes the neural network's output by calculating a statistic of neighbouring outputs summarization [14]. This decreases the representation's spatial dimension, therefore decreasing processing and weight requirements. Each slice of the representation is pooled separately. There are two types of pooling layers: max pooling and average pooling [14]. Max, or maximum, pooling takes the maximum element from each window from its map whereas average pooling takes the calculated average of the features present in map. Max pooling is generally preferred due to its performance [14].

Max pooling is the most common method, which conveys the neighbourhood's maximum output. The rectangular neighbourhood average, weighted average depending on distance from the centre pixel, and L2 norm of the rectangle neighbourhood are all pooling functions. The calculation of the output volume can be seen below in Figure 7:

$$W_{out} = \frac{W - F}{S} + 1$$

Figure 7. Formula for Padding Layer [14].

This produces a Wout x Wout x D output volume. In all circumstances, pooling gives some translation invariance, allowing an item to be recognised independent of frame location.

Fully-connected layers are level feed-forward neural networks layers [15]. They might use a non-linear activation feature or softmax activation in order to produce course forecast likelihoods [15]. As soon as the merging and also convolutional layers have actually drawn out and also combined functions, the layers are released. The network makes use of these to produce last non-linear function mixes as well as make forecasts [15].

A CNN design is made up of a pile of distinct layers that, by the application of a differentiable feature, transform the input quantity right into an outcome quantity. There are a couple of various kind of layers that are consistently used. There are numerous styles of CNNs, a few of which are as complies with [16]:

- LeNet (1998)
- AlexNet (2012)
- GoogLeNet (2014)
- VGGNet (2014)
- ResNet (2015)

The invention of LeNet marked the beginning of the history of deep CNNs. At the time, CNNs were only capable of performing handwritten digit recognition tasks, which means they couldn't be applied to other picture classes. AlexNet is highly regarded in the field of deep CNN architecture since it has produced groundbreaking achievements in the fields of image recognition and classification. Initially presented by Alex Krizhevesky, Ilya Sutskever, and Geoff Hinton, AlexNet has since been enhanced in terms of learning ability by increasing the depth of the network and using a number of parameter optimization algorithms. Using the AlexNet architecture as an example, Figure 8 depicts the fundamental design [16].

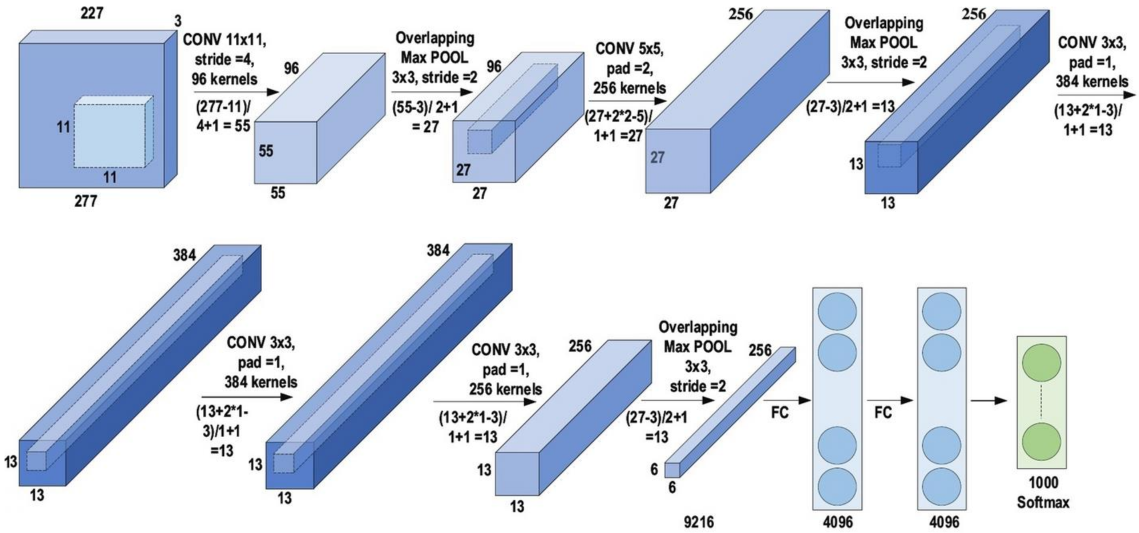

Figure 8. The architecture of AlexNet [16].

Google's GoogLeNet architecture is the CNN architecture that was utilised to win the ILSVRC 2014 classification competition [16]. Jeff Dean, Christian Szegedy, Alexandro Szegedy, and various others worked together to create it [16]. It has been demonstrated to have a vastly margin of error when compared to AlexNet and ZF-Net, according to the research. Comparing the error rate, it is substantially lower than using VGG. Deeper architecture is accomplished with a variety of different approaches, including global average pooling and 1-1 convolution. Because of the computationally intensive nature of the GoogleNet CNN design, it employs heavy unpooling layers on top of CNNs to eliminate spatial redundancy during training, as well as shortcut connections between the first two convolutional layers before adding additional filters in subsequent CNN layers, in order to limit the number of parameters that must be learnt [17]. Figure 9 shows the basic structure of a Google Block.

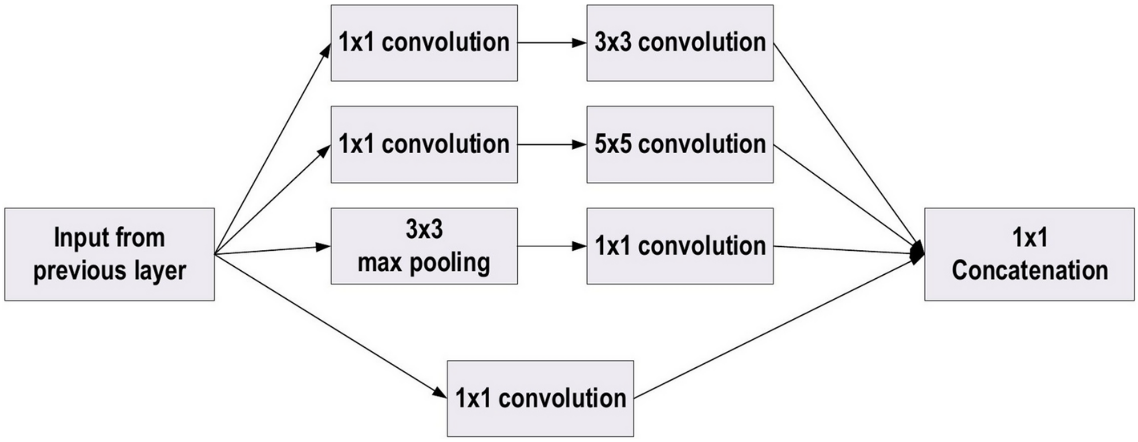

Figure 9. The structure of Google Block [16].

In the field of computer networking, the VGGNet architecture was designed by Karen Simonyan, Andrew Zisserman, and others at Oxford University. VGGNet is a 16-layer CNN with up to 95 million parameters that has been trained on more than one billion photos, according to the researchers [16]. At 4096 convolutional features, it is capable of processing huge input pictures with a resolution of 224 by 224 pixels. For most image classification tasks, CNN architectures such as GoogLeNet (AlexNet architecture) perform better than VGGNet when the input images are between 100 x 100 pixels and 350 x 350 pixels in size. CNN architectures such as GoogLeNet (AlexNet architecture) perform better than VGGNet when the input images are between 100 x 350 pixels in size [17].

It was the ResNet CNN architecture developed by Kaiming He et al. that took first place in the 2015 International Learning and Teaching Virtual Reality Task [16]. It had a top-five error of just 15.43 percent in the classification challenge.

With 152 layers and more than one million parameters, the network is considered deep even for CNNs [17]. It would have taken more than 40 days on 32 GPUs to train the network on the ILSVRC 2015 dataset, which is an extremely long period of time. CNNs are most typically employed to handle picture classification tasks with 1000 classes, however ResNet illustrates that CNNs may also be used to address natural language processing issues like as sentence completion and machine understanding [17], as shown by ResNet. ResNet is shown in block diagram form in Figure 10.

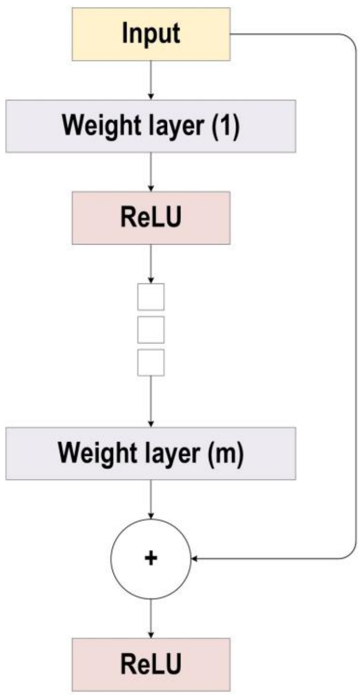

Figure 10. The block diagram for ResNet [16].

2.3.3 Applications and Advantages

Neural networks have applications in a selection of sectors. When it pertains to the jobs on which synthetic semantic networks are made use of, they have a tendency to come under among the significant classifications pointed out listed below [18]:

- Data Processing utilizing the complying with methods: filtering system, blind signal splitting up, clustering as well as compression.

- Classification, which uses uniqueness discovery, pattern as well as series acknowledgment, and also consecutive decision making.
- Function approximation is utilized in a range of applications, consisting of time collection forecast as well as modelling. Additionally called regression evaluation.

Training data helps neural networks learn and increase accuracy over the course of time. When refined, these learning algorithms are important tools in artificial intelligence as well as computer science in general, allowing scientists to rapidly categorise and cluster data. The time it takes for voice or picture recognition is comparable to manual identification by human professionals. Google's search algorithm uses a neural network.

ANNs can be used for predicting weather, predictive analysis for businesses, for example SCM forecasting, speech to text transcription, facial and handwriting recognition and spam email monitoring. ANNs are adaptive, have an enhanced learning capacity, slow corruption and distributed storage across an entire network and not just on one database. However, ANNs suffer since they require profuse quantities of data in order to train, the architecture does allow for clear explanations as to how a result was reached [18].

Using CNNs in computer vision applications has several advantages over other types of classical neural networks, which are stated as follows [16]:

- In particular, their implementation of the weight sharing feature aids in limiting the number of trainable network parameters, allowing the network to enhance generalisation while avoiding overfitting.
- Learn the feature extraction layers and the classification layer simultaneously, and you will get a model output that is both highly ordered and very dependent on the extracted features.
- • When compared to other networks, the construction of large-scale systems is less complicated.

Neural networks are a vast and growing field with various architectures. For the sake of this thesis, the most accurate and commonly used where discussed.

2.4 Object Detection and Computer Vision

Object detection is a computer technology that utilities ML and DL. Computer vision's and image processing's main objectives include: detecting instances of objects of a specific class in images and videos [19]. Face and pedestrian detection are two object detection areas that have received much attention. Object detection can be utilized for people tracking, people counting, automated CCTV surveillance, person detection, and vehicle detection.

It is necessary to manually extract features from images before using a machine learning-based object detection method, seen in Figure 11 below. Histogram of oriented gradients (HOG) is one of the most common "Image-based feature extraction techniques" [20].

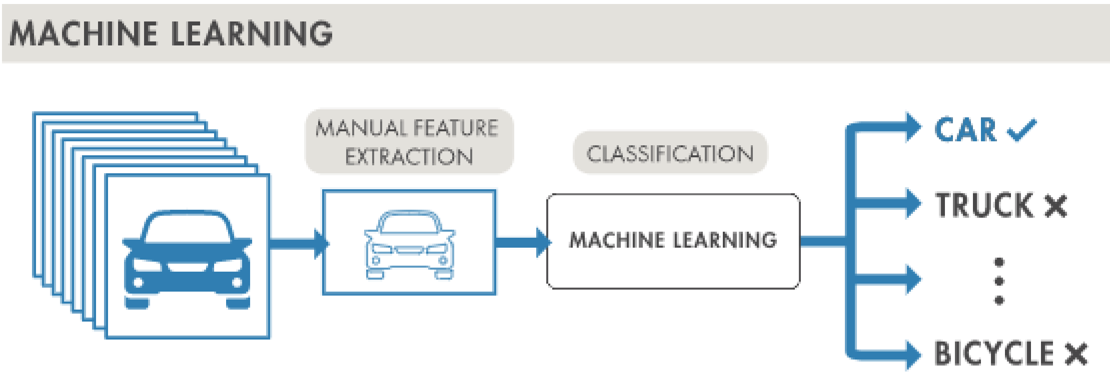

Figure 11. Machine learning base for object detection [20].

When utilizing DL formulas such as CNNs, Auto-Encoder, Difference Auto-Encoder, and also others to produce the attribute from the picture such as side as well as form, the item discovery approach can draw out the attribute instantly [20] Those DL formulas are made use of to remove the attributes from the picture. Several of those attributes are side as well as form. This can be seen listed below in Figure 12.

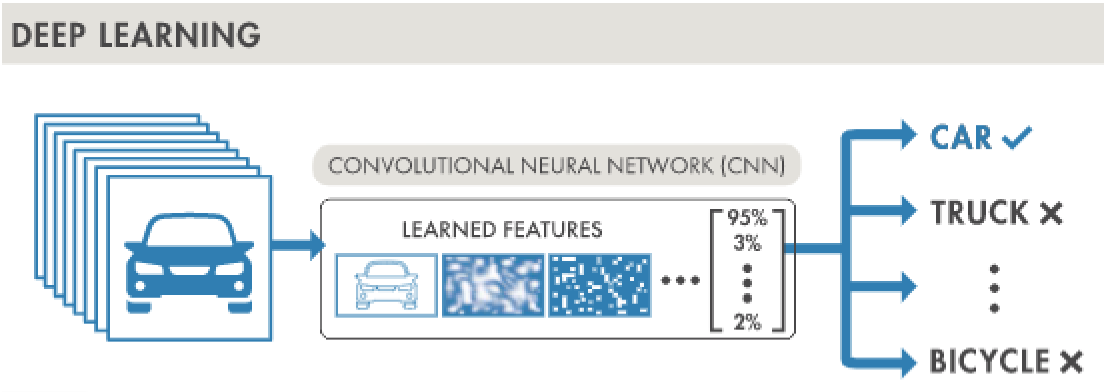

Figure 12. Deep learning base for object detection [20].

The most notable two-stage object detection algorithms include the following:

- RCNN and SPPNet (2014)
- Fast RCNN and Faster RCNN (2015)
- Mask R-CNN (2017)
- Pyramid Networks/FPN (2017)
- G-RCNN (2021)

The most notable one-stage object detection algorithms include:

- YOLO (2016)
- SDD (2016)
- RetinaNet (2017)
- YOLOv3 (2018)
- YOLOv4 (2020)
- YOLOR (2021)

In order to choose which object detection algorithm is most suited to one's needs, one must understand the main characteristics of each algorithm. For the purpose of this thesis, YOLO will be the main focus. R-CNN, Fast R-CNN, Faster R-CNN, Region-based Fully Convolutional Network (R-FCN), Single Shot Multibox Detector (SDD), Histogram of Oriented Gradients (HOG) and Spatial Pyramid Pooling (SPP-net) will be briefly discussed.

2.4.1 You Only Look Once (YOLO)

A real-time object identification system that depends on a single neural network to identify things in real time is known as "You Only Look Once," or YOLO, in popular culture. The ability to train a bespoke YOLO model that can recognise any kind or number of objects is a new feature in the most recent version of ImageAI, version 2.1.0, which is available now. A classic example of classifier-based systems, conventional neural networks are systems in which the system repurposes classification or localization methods to conduct detection, and then applies the detection model to an image at a range of various locations and scales to identify objects. According to this procedure, the portions of the picture that have a "high score" are deemed to be detections [21]. Positive identification occurs simply because the areas that are the most similar to the training pictures are those that are positively recognised.

Because it is a one-stage detector, YOLO is capable of completing both bounding box regression and classification in a single operation. As a result, it is substantially faster than other convolutional neural networks in its class. YOLO is presently the most efficient object detection technique available. Comparing R-CNN and Fast R-CNN, it can be a thousand times faster than the first and a hundred times faster than the second, depending on the situation.

When it comes to training, YOLOv3 makes advantage of multi-label categorization with overlapping patterns. So, it has the potential to be applied in complicated circumstances for object detection. For tiny item classification, YOLOv3 can be employed because of its multi-class prediction capabilities. However, when it comes to big or medium-sized object classification, YOLOv3 performs significantly worse.

YOLOv4 is an enhanced version of the previous YOLOv3. YOLOv4 is a one-stage object detection model that builds on the success of YOLOv3 by including a number of new techniques and modules that have been released in the literature. In Figure 13, one can see what object detection looks like with YOLO4.

Figure 13. Object detection with YOLO4 [21].

Comparing YOLOv4 to the previous YOLOv3, the following benefits may be observed.

It is a highly powerful and accurate object detection model that allows computers equipped with a 2080 Ti GPU to train an efficient object detector in a short amount of time with no effort [22].

Verification has been performed on the impact of the latest item detection approaches, including the "Bag-of-Freebies" and the "Bag-of-Specials," on the training of detectors [22].

All of the state-of-the-art approaches, such as CBN (Cross-iteration batch normalisation), PAN (Path aggregation network), and others, have been improved to be more efficient and suited for single GPU training [22].

2.4.2 Region-based Convolutional Neural Networks (R-CNN)

Region-based convolutional neural networks (R-CNNs), occasionally called areas with CNN functions (R-CNNs), are innovative strategies to object acknowledgment that make use of deep understanding versions. R-CNN designs initial pick numerous recommended locations from a picture, after which they classify the areas' groups and also bounding boxes utilizing tags developed by human beings (e.g. offsets). These tags are produced by the computer system based upon pre-set courses that have actually been offered to it. They after that do ahead calculating on each recommended area utilizing a convolutional semantic network to remove attributes from each picked location [23].

Throughout the R-CNN procedure, the inputted image is first split right into approximately 2 thousand location items, and after that a convolutional semantic network is put on each market of the photo consequently. Utilizing these details, the dimension of the locations is approximated, as well as the ideal area is after that included right into the semantic network. It might be reasoned that a time frame can be enforced by a comprehensive way such as that explained. As a result of the reality that it identifies as well as constructs bounding boxes on a specific basis, and also given that a semantic network is put on one area each time, the training duration is considerably longer than with YOLO [23].

Rapid R-CNN, which was developed in 2015 with the objective of decreasing train period by a huge quantity, was presented. When contrasted to the initial R-CNN, which calculated semantic network attributes on each of as much as 2 thousand areas of passion separately, Rapid R-CNN runs the semantic network just when overall photo. This is incredibly comparable to the style of YOLO, yet YOLO remains to be a quicker choice to Quick R-CNN as a result of the simpleness of the code [23].

2.4.3 Fast R-CNN

The Fast R-CNN algorithm is superior to the original R-CNN algorithm in that it runs the neural network only once on the entire picture rather than on each of up to two thousand areas of interest separately, as the previous R-CNN algorithm did. The previous R-CNN algorithm ran the neural network on each of up to two thousand areas of interest separately. Final results include the invention of a revolutionary approach known as ROI Pooling, which slices each ROI from the network output tensor, then shapes and categorises each ROI in line with the category into which it was classified. Similarly to the original R-CNN, the Fast R-CNN generates its area proposals by the use of a process known as Selective Search, which is comparable to that used by the original R-CNN.

2.4.4 Faster R-CNN

While Fast R-CNN generated ROIs through the use of Selective Search, Faster R-CNN incorporates the process of ROI formation within the neural network itself [24].

2.4.5 Region-based Fully Convolutional Network (R-FCN)

Region-based fully convolutional networks (R-FCNs), are a kind of region-based item detector that utilizes completely convolutional networks to discover items in a picked area. The R-FCN things detector, as opposed to Fast/Faster R-CNN, that utilized a per-region subnetwork thousands of times, is entirely convolutional, with almost all calculation shared throughout the entire photo [25].

In order to achieve this, R-FCN utilizes position-sensitive rating maps to solve a dispute that exists in between translation-invariance in image category and also translation-variance in item acknowledgment.

2.4.6 Single Shot Multibox Detector (SDD)

The SDD is a common one-stage detector that is capable of identifying many classes in a single step. It is also known as the Single Shot Multibox Detector (SDD). A single deep neural network may be used to find objects in images, and this is made feasible by discretizing the output space of bounding boxes into a collection of default boxes that cover many aspect ratios and sizes per feature map point, as described in [25].

Using the object detector, scores are generated for each default box depending on the existence of each item category in each default box, and the default box is altered to better match the shape of the object in each default box. In addition, predictions from a number of feature maps with varied resolutions are included into the network in order to cope with objects of varying sizes and shapes [25].

In addition to being easy to train, this object detection component may be simply incorporated into software systems that need object detection capability. SSD provides much greater accuracy when compared to conventional single-stage techniques, which is particularly true when dealing with lower input image sizes, as seen above.

2.4.7 Histogram of Oriented Gradients (HOG)

HOG, which stands Histogram of Oriented Gradients, is a function descriptor that counts the number of slope alignment events that occur in different areas of a picture that has been taken with a slope alignment matter applied to it. HOG is a straightforward procedure that can be accomplished with the aid of OpenCV and also various other related technologies. Given that there is already a predefined technique named HOG in the skimage.feature collection, it just requires a few more lines of code to do this task [26].

2.4.8 Spatial Pyramid Pooling (SPP-net)

As the name suggests, it is a kind of CNN that uses spatial pyramid pooling (SPP-net) to conquer a fixed-size restriction in both the input as well as the outcome dimensions of the network. Even more, the SPP layer is used in addition to the last convolutional layer prior to the last making is finished. SPP layer swimming pools functions as well as creates fixed-length outcomes, which are after that fed right into layers 3 and also 4. The SPP layer is absolutely adjoined with layers 1 and also 2, yet not with layers 3 as well as 4. Therefore, we do some info gathering at a greater degree in the network's framework to avoid the demand for chopping or deforming at the start of the network's understanding stage [27].

2.5 Raspberry Pi

Raspberry Pi is a single-board computer system that is portable and also affordable. It might be utilized as a little computer by affixing it to externals such as a key-board, computer mouse, as well as display. It is taken into consideration a tiny computer that is commonly utilized for real-time photo, robotics applications as well as video clip handling, Internet of things (IoT) applications [28].

Raspbian OS, which is based on Debian, is supplied on an official basis by the Raspberry Pi Foundation and may be downloaded from their website. They also provide the NOOBS operating system for the Raspberry Pi, which is available for download. The installation of a number of Third-Party operating systems is feasible. Raspbian OS is an official operating system that can be downloaded and used for free on a Raspberry Pi microcontroller board, other OS options are; Ubuntu, ArchLinux, Windows 10 (IOT Core), and various others. This operating system has been created particularly for use with the Raspberry Pi and its associated hardware. Among other things, Raspbian offers a graphical user interface (GUI), which provides tools for Python programming as well as for office productivity, surfing, and gaming [28].

Raspberry Pi offers designers accessibility to the on-chip equipment, referred to as GPIOs, for developing applications. By getting to GPIO, we might link gadgets like as electric motors, sensing units, LEDs and also various other comparable gadgets, additionally enabling to manage them.

With an ARM-based Broadcom Processor SoC and an on-chip GPU, it is a powerful piece of hardware (Graphics Processing Unit). The CPU speed of the Raspberry Pi varies depending on the model, ranging from 700 MHz to 1.2 GHz. Additional features include an SDRAM module that may be anywhere from 256 MB and 1GB in capacity. Additionally, in addition to the usual modules, the Raspberry Pi has on-chip SPI, I2C, I2S, and UART interfaces [29].

There are different versions of raspberry pi available as listed below:

- Pi 1 Model B (2012)
- Pi 1 Model A (2013)
- Pi 1 Model B+ (2014)
- Pi 1 Model A+ (2014)
- Pi 2 Model B (2015)
- Pi Zero (2015)
- Pi 3 Model B (2016)
- Pi Zero W (2017)
- Pi 3 Model B+ (2018)
- Pi 3 Model A+ (2019)
- Pi 4 Model A (2019)
- Pi 4 Model B (2020)
- Pi 400 (2021)

In Figure 14, a Raspberry Pi 4 Model B can be seen.

Figure 14. Raspberry Pi 4 Model 4.

Below the technical specifications of the Raspberry Pi 4 Model 4 [29].

- Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5GHz
- 1GB, 2GB, 4GB or 8GB LPDDR4-3200 SDRAM (depending on model)
- 2.4 GHz and 5.0 GHz IEEE 802.11ac wireless, Bluetooth 5.0, BLE
- Gigabit Ethernet
- 2 USB 3.0 ports; 2 USB 2.0 ports.
- Raspberry Pi standard 40 pin GPIO header (fully backwards compatible with previous boards)
- 2 × micro-HDMI ports (up to 4kp60 supported)
- 2-lane MIPI DSI display port
- 2-lane MIPI CSI camera port
- 4-pole stereo audio and composite video port
- H.265 (4kp60 decode), H264 (1080p60 decode, 1080p30 encode)
- OpenGL ES 3.1, Vulkan 1.0
- Micro-SD card slot for loading operating system and data storage
- 5V DC via USB-C connector (minimum 3A*)
- 5V DC via GPIO header (minimum 3A*)
- Power over Ethernet (PoE) enabled (requires separate PoE HAT)
- Operating temperature: 0 – 50 degrees C ambient

The major difference between the Raspberry Pi 3 and the Raspberry Pi 4 are the HDMI ports, which can be seen below in Figure 15. The ports in the Raspberry Pi Model 4 were upgrade to micro HDMI which can support two 4K monitors instead of just one monitor.

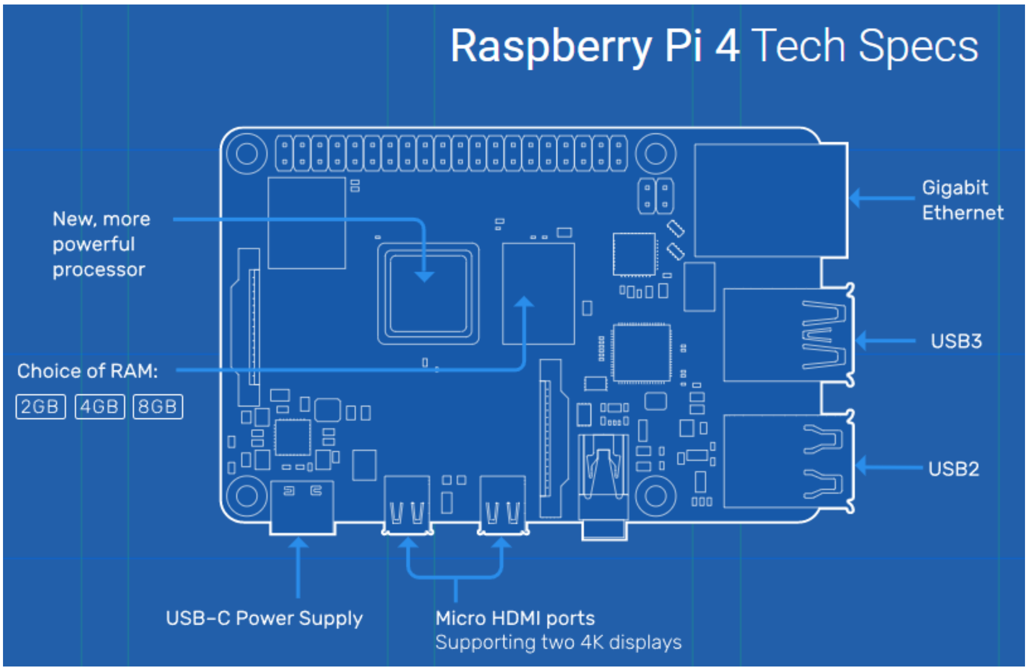

Figure 15. Technical specification for Raspberry Pi 4 Model B [29].

2.6 Python

Python is a translated, object-oriented, top-level programs language with vibrant semiotics that might be implemented on a computer system's desktop computer [31]. Because of its top-level information frameworks, vibrant kind as well as vibrant binding, it is an extremely attractive choice for Fast Application Growth in addition to a scripting or adhesive language for incorporating existing parts. In addition, Python's fundamental, easy-to-learn phrase structure puts a costs on readability, which minimizes the complete price of program upkeep. Python's assistance for components as well as plans motivates program modularity and also code reuse. Python's interpreter and also huge basic collection are offered completely free in resource or binary type for all significant systems as well as might be easily dispersed as-is [32].

Python can be used in the following instances [32]:

- Automation
- Data analysis
- Everyday tasks
- Machine learning
- Scripting
- Software protyping
- Software testing
- Web development

It features a straightforward syntax that replicates normal English, making it simpler to read and comprehend. Projects can be built more quickly, and they can be improved more quickly as a result. It has a lot of applications. Python may be used for a wide variety of activities, as mentioned in the list previously. It is user-friendly for beginners, making it a popular choice for beginning programmers. It is open source, which implies that it can be used and distributed without restriction, even for commercial reasons. There is a massive and increasing repository of Python modules and libraries, which are bundles of code written by third-party users to extend the capabilities of the programming language. In Python, there is a huge and active community of programmers who contribute to the language's pool of modules and libraries. Because of the large support network, if coders stumble into a stumbling block, finding a solution is generally simple; chances are that someone else has encountered the same issue previously.

2.7 Numpy

NumPy (Numerical Python) is a Python collection that is made use of in almost every technique of study and also design. It is cost-free and also open resource software application. In Python, it's the de facto requirement for collaborating with mathematical information, as well as it's an essential part of the clinical Python as well as PyData communities [33]. Starting developers to professional academics working on sophisticated clinical and also sector r & d are all amongst the NumPy neighborhood's individuals. Pandas, SciPy, Matplotlib, scikit-learn, scikit-image, as well as a lot of various other information scientific research as well as clinical Python programs make significant use the NumPy API, as do most various other Python items. The NumPy collection supplies information frameworks for multidimensional ranges and also matrixes. It prolongs Python with advanced information frameworks that supply reliable calculations with selections and also matrices, in addition to a substantial collection of top-level mathematical features that work with these varieties and also matrices [33].

2.8 Tensorflow

TensorFlow is an open-source machine learning platform that is designed to be used end-to-end. It provides a rich and flexible ecosystem of tools and libraries which allows scientists to advance the state-of-the-art in machine learning while engineers can simply build and deploy applications driven by machine learning [34].

It was originally developed by researchers and engineers working on the Google Brain team inside Google's Machine Intelligence Research division to undertake machine learning and deep neural network research. It is now maintained by the TensorFlow Foundation [34].

For Python and C++, TensorFlow provides a stable API, as well as a non-guaranteed backward compatible API for other programming languages.

3 Methodology

The Raspberry Pi 3 model B, 1 GB RAM/Quad Core 1.2GHz Broadcom BCM2837 64bit CPU, was originally used for the setup process, however the device didn’t have enough power to run YOLOv3 nor Tiny-YOLO. In order to run YOLOv3, it is recommended to use a Raspberry Pi 4 model B (4 GB RAM) at a minimum. There is also the option of running Tiny-YOLO which requires a less powerful processer and RAM. The required hardware for this thesis can be seen below:

- A Raspberry Pi 4 model B (4 GB RAM)
- Raspberry Pi camera module v2
- A power source with 2.0A - 2.5A (Raspberry Pi's official power supply is recommended)
- A micro SD card (32GB recommended)
- Monitor
- Keyboard
- Mouse
- Micro-HDMI to HDMI connector for monitor

The setup process will include the following steps, which will be discussed in their own subchapters.

- Raspberry Pi Setup & Update
- Installation of all required libraries
- Installation of OpenCV & TensorFlow
- Installation of YOLOv3/Tiny-YOLO
- Configuration of Raspberry Pi Camera

3.1 Raspberry Pi Setup & Update

Before starting, acquire a monitor, micro-HDMI connector, power supply, keyboard and mouse to go with the Raspberry Pi. Connect the keyboard, mouse and attach the micro-HDMI to HDMI to the Raspberry PI and monitor. Once those are ready to go, ensure that there is a micro-SD card in the Raspberry Pi, preferably 32 GB. In order to install an operating system on your SD card, it is recommended that use the Raspberry Pi Imager. Installing the image will necessitate the use of another computer equipped with an SD card reader. Raspberry Pi Imager can be downloaded from the following website:

https://www.raspberrypi.com/software/. Install the software and once the installation is complete, run it. In Figure 16, the main screen can be seen.

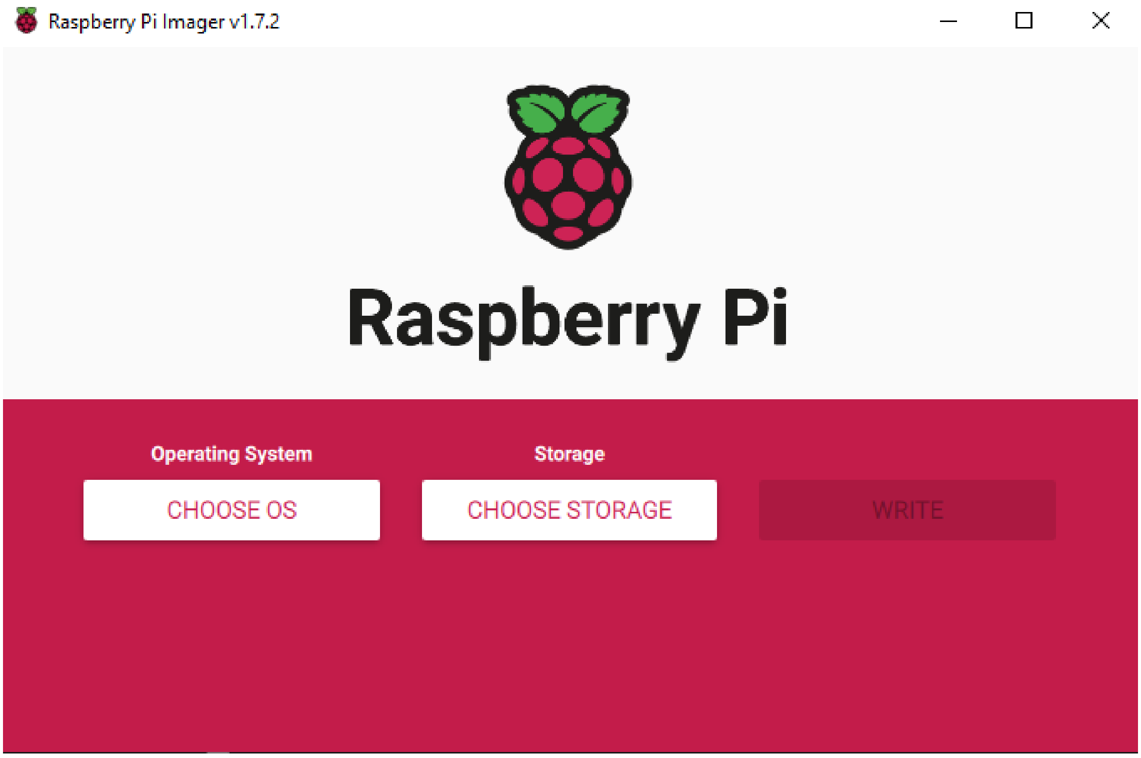

Figure 16. Raspberry Pi Imager v1.7.2.

Click “Choose OS” and select Raspberry Pi OS (64-bit), which is compatible with Raspberry Pi 3/4/400. Click “Storage”, then select the location of the SD-card. Finally click the “Write” button. Once the write is successful, see Figure 17, the SD-card can be ejected.

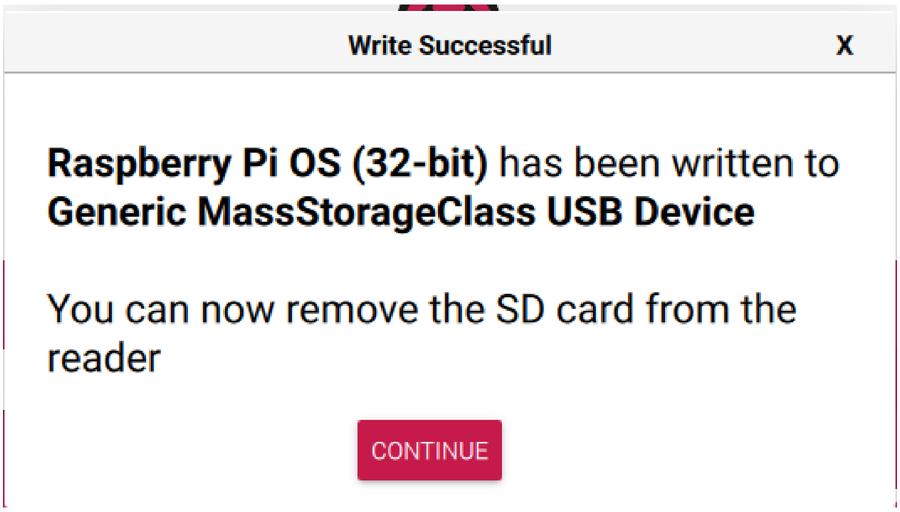

Figure 17. Successful write screen.

Now the SD-card can be inserted into the Raspberry Pi. Finally insert the power supply, at least 3.0 amps, to the Raspberry Pi 4 model B. Once the device powers on a screen similar to Figure 18 should be seen.

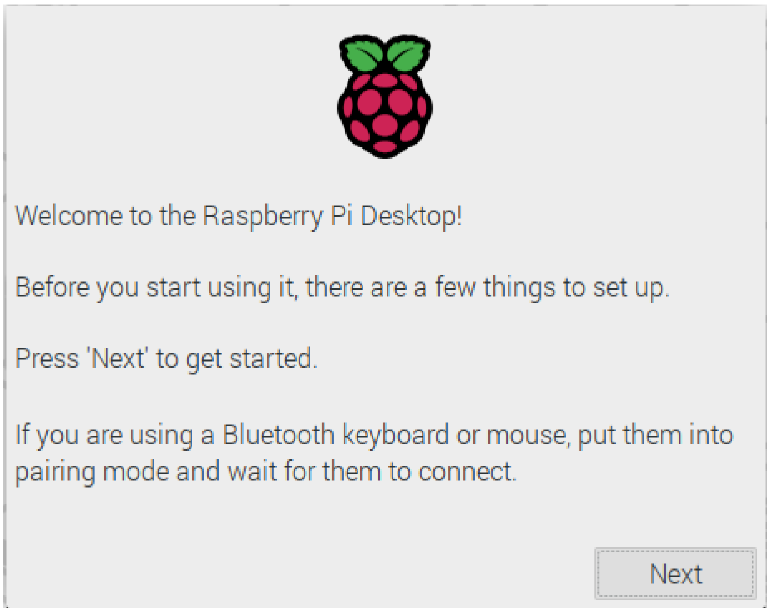

Figure 18. Welcome to Raspberry Pi.

Click next. Select your country, language, and time-zone, and then click on Next once again to proceed.

Enter a new username and password for your Raspberry Pi and then press the Next button.

Configure your screen such that the Desktop takes up the entire width of your display.

Connect to your wireless network by selecting the network's name, inputting the password, and pressing the Next button on your computer's keyboard.

Click on Next, and the process will check for and install any necessary updates to the Raspberry Pi OS (this might take a little while).

To complete the configuration, click on Restart.

Once the device starts up, open the terminal and then run the commands below to update the OS.

```
sudo apt-get update
sudo apt-get upgrade
sudo rpi-update
```

Once those commands have completed, the device should be restarted again. This ensures that all libraries, firmware and the OS are up-to-date.

3.2 Installation of all required libraries

Python 3 is the default Python installation for the Raspberry Pi operating system. For Python 3, we must complete the procedures outlined below in order to install pip (pip3): 3650

```
sudo apt install python3-pip
```

Besides installing all of the dependencies necessary for developing Python modules, the script above will also install the Python interpreter.

Once the installation is complete, we can check the pip version to ensure that the installation was successful:

```
pip3 --version
```

This command might not install the most recent version, and it will need to be upgraded by using the following command:

```
pip3 install --upgrade pip
```

Before proceeding to the next stage, we must first create a batch of packages using the command listed below [35]:

```
sudo apt-get install -y libhdf5-dev libc-ares-dev libeigen3-dev gcc gfortran
python-dev libgfortran5 \ libatlas3-base libatlas-base-dev libopenblas-dev
libopenblas-base libblas-dev \ liblapack-dev cython libatlas-base-dev openmpi-
bin libopenmpi-dev python3-dev python3-venv
```

The packages above might take a while to complete.

3.3 Installation of OpenCV

The issue arises as a result of the fact that the current version of OpenCV is incompatible with the Raspberry Pi. There are a number of prerequisites that must be installed via apt-get in order for OpenCV to function properly on the Raspberry Pi. If running "sudo apt-get update" does not work and try again with the following command [35]:

```
sudo apt-get install libjpeg-dev libtiff5-dev libjasper-dev libpng12-dev
libavcodec-dev libavformat-dev \ libswscale-dev libv4l-dev libxvidcore-dev
libx264-dev qt4-dev-tools libatlas-base-dev
```

If you have already attempted to install OpenCV, you can simply run to remove the unwanted package.

```
pip3 uninstall opencv-python
```

As soon as we have all of these things installed, we can proceed to installing OpenCV by running the following:

```
pip3 install opencv-python-python==3.4.6.27
```

If there are any issues still, check from OpenCV's website which versions are compatible with the Raspberry Pi that is being used.

3.4 Installation of TensorFlow

Check the Python3 version on your computer. A different wheel is required for each edition. The Python 3.7.3 programming language is currently used by the Raspberry Pi 64-bit operating system. The version 2.2.0 of TensorFlow for Linux should be downloaded and installed from their official website. Python will release new versions over time, and it will require an update. In order to ensure all the correct libraries, enter the following commands in the terminal [36] [37].

```
$ sudo apt-get update
$ sudo apt-get upgrade
$ sudo apt-get install python-pip python3-pip
$ sudo pip uninstall tensorflow
$ sudo pip3 uninstall tensorflow
$ sudo apt-get install gfortran
$ sudo apt-get install libhdf5-dev libc-ares-dev libeigen3-dev
$ sudo apt-get install libatlas-base-dev libopenblas-dev libblas-dev
$ sudo apt-get install liblapack-dev
$ sudo -H pip3 install pybind11
$ sudo -H pip3 install Cython==0.29.21
$ sudo -H pip3 install h5py==2.10.0
$ sudo -H pip3 install --upgrade setuptools
$ pip3 install gdown
$ gdown https://drive.google.com/uc?id=1fR9lsi_bsI_npPFB-wZyvgjbO0V9FbMf
$ sudo -H pip3 install tensorflow-2.2.0-cp37-cp37m-linux_aarch64.whl
```

After completing the installation process successfully, you should see the screen dump shown below in Figure 19.

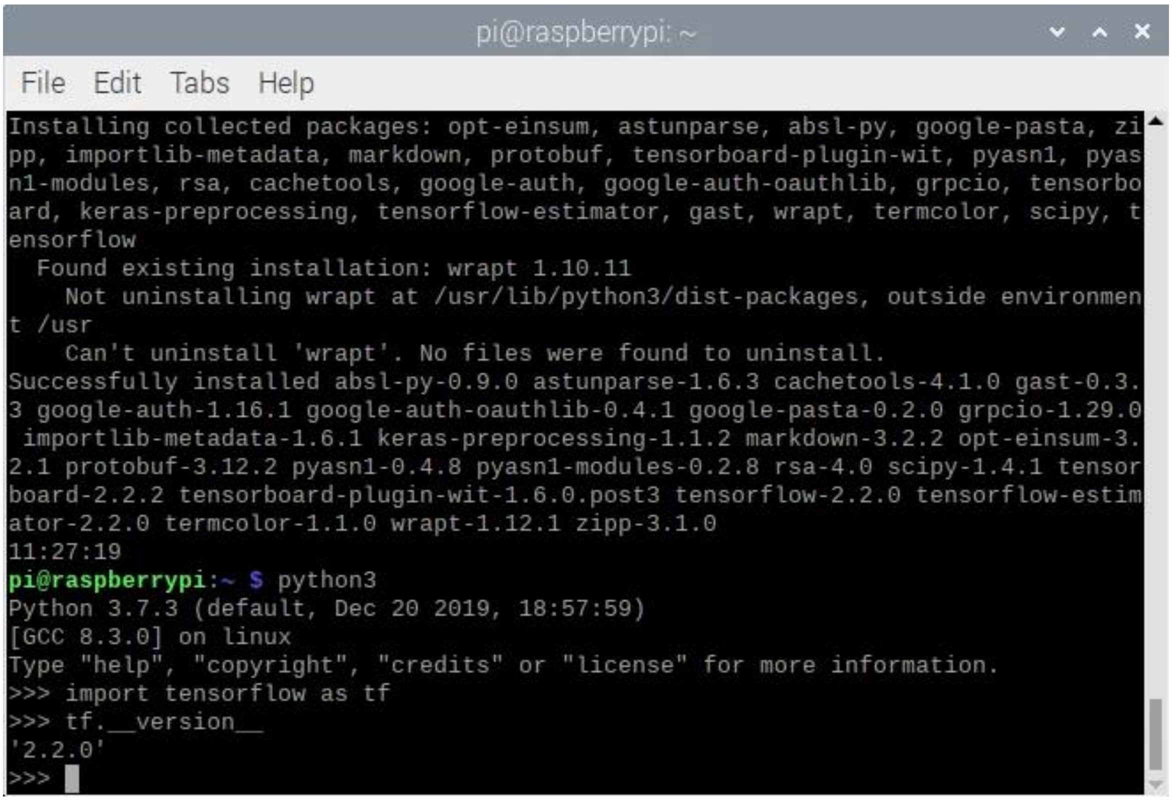

Figure 19. Successful installation of TensorFlow.

Another faster option is to install TensorFlow Lite if the Raspberry Pi doesn’t have enough resources to regular the TensorFlow version above. TensorFlow Lite cannot be used to train models. It is only compatible with pre-trained models that have been modified to operate in the "Lite" version. First the following commands need to be entered into the terminal in order to prepare the system [37].

```
sudo apt update
sudo apt upgrade -y
echo "deb [signed-by=/usr/share/keyrings/coral-edgetpu-archive-keyring.gpg]
https://packages.cloud.google.com/apt coral-edgetpu-stable main" | sudo tee
/etc/apt/sources.list.d/coral-edgetpu.list
curl https://packages.cloud.google.com/apt/doc/apt-key.gpg | sudo tee
/usr/share/keyrings/coral-edgetpu-archive-keyring.gpg >/dev/null
sudo apt update
```

TensorFlow Lite is not included in the repositories. Rather than that, one must make use of Google's package repository. On the Raspberry Pi, it is required to add the Google package repository containing TensorFlow Lite. Due to the fact that we amended the package sources on our Raspberry Pi, we must update our package list to include the newly added repository. Once those steps have been completed then TensorFlow Lite can be installed using the following commands in the terminal.

```
sudo apt install python3-tflite-runtime
```

After installing the package, a check should be performed to ensure that TensorFlow Lite is now operational. It's simple to verify if TensorFlow Lite is installed through the Python CLI.

```
python3
from tflite_runtime.interpreter import Interpreter
```

TensorFlow Lite models may now be run on the Raspberry Pi.

3.5 Installation of YOLOv3/Tiny-YOLO

First step is to download Darknet by running the following commands in the terminal [38].

```
git clone https://github.com/pjreddie/darknet
cd darknet
make
```

The pre-trained weight file can be downloaded via command below [38].

```
wget https://pjreddie.com/media/files/yolov3.weights
```

Finally, in order to order the detector, run the following.

```
./darknet detect cfg/yolov3.cfg yolov3.weights data/dog.jpg
```

The output should look something like this:

```
layer     filters    size              input                output
    0 conv     32  3 x 3 / 1   416 x 416 x   3   ->   416 x 416 x  32  0.299 BFLOPs
    1 conv     64  3 x 3 / 2   416 x 416 x  32   ->   208 x 208 x  64  1.595 BFLOPs
    .......
  105 conv    255  1 x 1 / 1    52 x  52 x 256   ->    52 x  52 x 255  0.353 BFLOPs
  106 detection
truth_thresh: Using default '1.000000'
Loading weights from yolov3.weights...Done!
data/dog.jpg: Predicted in 0.029329 seconds.
dog: 99%
truck: 93%
bicycle: 99%
```

Figure 20. YOLO detection output [38].

Darknet logs the items it detects, its confidence in them, and the time it took to locate them. Darknet was not compiled with OpenCV, and so cannot display the detections directly. Rather than that, it stores them in predictions.png. The identified items can be viewed by opening the image. Due to the fact that Darknet is being utilized on the CPU, each image takes around 6-12 seconds to load. It would be significantly quicker if we used the GPU version. In Figure 20, the prediction image generated by the program can be seen.

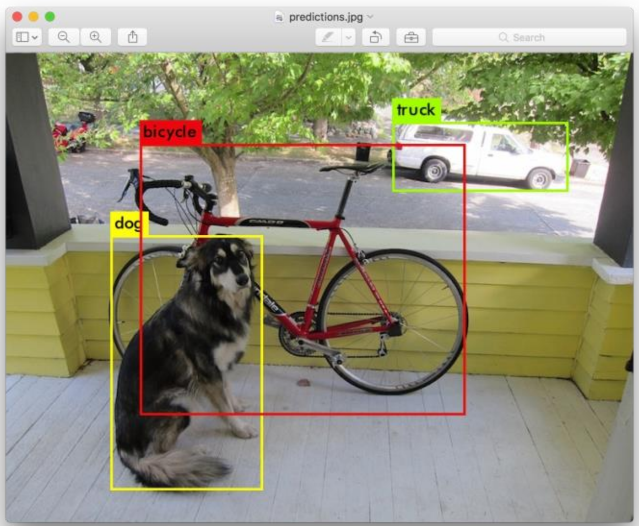

Figure 21. YOLO v3 prediction image [38].

By default, YOLO displays only items with a confidence level of.25 or more. This can be changed by supplying the yolo command the -thresh <value> parameter. For instance, you may set the threshold to 0 to display all detections [38]:

```
./darknet detect cfg/yolov3.cfg yolov3.weights data/dog.jpg -thresh 0
```

If the regular version of YOLOv3 cannot run on the Raspberry Pi or if the performance is slow, an alternative is to use YOLOv3 Tiny which is less demanding. To use YOLOv3 Tiny, run the following commands [38].

```
wget -O weights/yolov3-tiny.weights https://pjreddie.com/media/files/yolov3-tiny.weights
```

Then, using the tony configuration file and weights, run the detector:

```
./darknet detect cfg/yolov3-tiny.cfg yolov3-tiny.weights data/dog.jpg
```

The next set is to configure the Raspberry Pi camera along with setting up the real-time object detection.

3.6 Configuration of Raspberry Pi Camera

First the camera needs to be turned on, which can be seen below in Figure 21. Run sudo raspi-config and navigate to the main menu of the Raspberry Pi Software Configuration Tool. Select Interfacing Options.

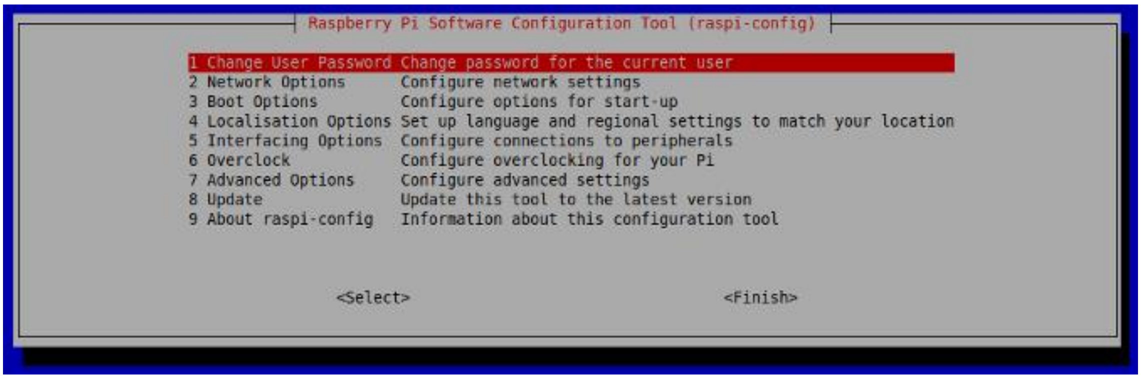

Figure 22. Raspberry Pi Software Configuration Tool.

Select the Enable Camera menu option and press Enter.

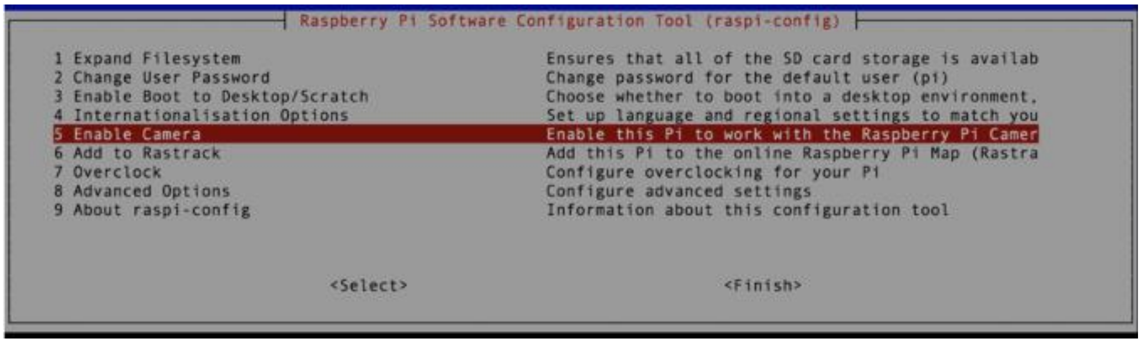

Figure 23. Enable camera.

In the subsequent menu, hit Enter after selecting Enable with the right arrow key.

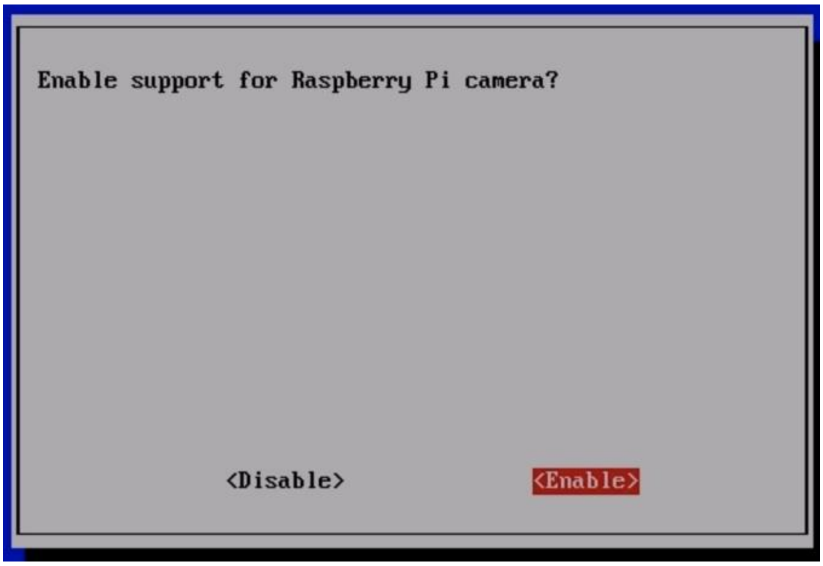

Figure 24. Enable support for camera.

In order to initialize the camera and generate a reference for the raw camera capture, the following code could be used.

```
cam = PiCamera()
cam.resolution = (640, 480)
cam.framerate = 32
rawCap = PiRGBArray(camera, size=(640, 480))
time.sleep(0.1)
```

The final steps include writing code that utilizes the object detection algorithm of your choice. One important aspect to consider is the thresh value which only determines which object are detected based off the confidence level. Along with adding capacity tracking, it is possible to also identify persons with masks or animals and generate statistics from those findings. Then those statistics can be stored within a database and visual representations can be generated based off that data.

4 Discussion

The original intention of this thesis was to create a prototype that could be used to monitor the capacity of a selected area. The primary blocker of the project was the lack of power or resources of the original Raspberry Pi that was used. After installing YOLOv3 along with VOLOv3 Tiny, it was clear that the resources of the Raspberry Pi 3 model B (1 GB RAM) were exhausted. After numerous attempts, a Raspberry Pi 4 model B (4 GB RAM) was borrowed in order to test whether YOLOv3 could run on it. At this point, there was some success but nothing viable. YOLOv3 Tiny thankfully produced results and was able to read an image along with real-time video. The processing quality and speed were lacking, however. Using YOLOv4 was also installed, and some tests were run, but again, the Raspberry Pi did not have enough resources to run it. At this point, it was clear that the current Raspberry Pis could not run YOLO efficiently. One option is to utilize an Intel® Neural Compute Stick 2 (Intel® NCS2). The Intel® NCS2 is based on the Intel® MovidiusTM MyriadTM X VPU, which features 16 configurable shave cores and a specialised neural compute engine for accelerating deep neural network inferences in hardware. This would, however, be a much more complex solution for capacity monitoring using object detection algorithms as well as require a different setup. The setup with the greatest performance includes wired or wireless cameras that feed the real-time video to a central server or computer(s), which utilizes a NVIDIA GPU, that have the scripts and YOLO models running on it.

5 Conclusion

Object identification utilising deep learning and neural networks have made huge strides in the last several years, and the topic is currently quite popular. New research articles, algorithms, or solutions to bugs are being published more frequently than before.

The purpose of this thesis was to explore various object detection algorithms, discussion of artificial intelligence, machine learning, neural networks and convolutional neural networks and explanation of how-to setup YOLOv3 on a Raspberry Pi. With all this collective theory and solution, one should be able to effectively monitor traffic and the capacity of an establishment. A business may easily comply with government limitations by utilising YOLO. Another conceivable application is to determine how prevalent the usage of face masks is in the institution. Other possible applications for the technology include rush hour prediction, which allows patrons to see how busy the establishment is before arriving, as well as knowing whether there are any animals in the establishment if they suffer from respiratory issues and analysing the popularity of each entrance/exit for safety planning. With all these possibilities, the technology can be utilized by both the business and the patrons.

Even though the project was not entirely successful, there were still valuable conclusions drawn from the theory and research done during the thesis. One can use those learning and implement a better, faster solution for capacity monitoring in the future.

References

1 WHO. Coronavirus disease (covid-19) [Internet]. World Health Organization. World Health Organization; [cited 2022May3]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19

2 Worldometer. Coronavirus cases: [Internet]. Worldometer. [cited 2022May1]. Available from: https://www.worldometers.info/coronavirus/

3 Koza JR, Bennett FH, Andre D, Keane MA. Automated design of both the topology and sizing of analog electrical circuits using genetic programming. Artificial Intelligence in Design '96. 1996;:151–70.

4 Edgar TW, Manz DO. Research methods for cyber security. Cambridge, MA: Syngress, an imprint of Elsevier; 2017.

5 IBM Cloud Education. What is machine learning? [Internet]. IBM. [cited 2022Apr30]. Available from: https://www.ibm.com/cloud/learn/machine-learning

6 IBM Cloud Education. What is deep learning? [Internet]. IBM. [cited 2022Apr30]. Available from: https://www.ibm.com/cloud/learn/deep-learning

7 Databricks. What is a neural network? [Internet]. Databricks. 2021 [cited 2022May1]. Available from: https://databricks.com/glossary/neural-network

8 LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May 28;521(7553):436-44. doi: 10.1038/nature14539. PMID: 26017442.

9 Oppermann A. What is deep learning and how does it work? [Internet]. Medium. Towards Data Science; 2020 [cited 2022Apr30]. Available from: https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-2ce44bb692ac

10 Team TAIE. Main types of neural networks and its applications-tutorial [Internet]. Towards AI. 2022 [cited 2022Apr30]. Available from: https://towardsai.net/p/machine-learning/main-types-of-neural-networks-and-its-applications-tutorial-734480d7ec8e

11 Veen Fvan. Fjodor van Veen, author at the Asimov Institute [Internet]. The Asimov Institute. 2017 [cited 2022Apr30]. Available from: https://www.asimovinstitute.org/author/fjodorvanveen/

12 Types of neural networks and definition of neural network [Internet]. 2022 [cited 2022Apr30]. Available from: https://www.mygreatlearning.com/blog/types-of-neural-networks/

13 Tch A. The mostly complete chart of neural networks, explained [Internet]. Medium. Towards Data Science; 2017 [cited 2022Apr30]. Available from: https://towardsdatascience.com/the-mostly-complete-chart-of-neural-networks-explained-3fb6f2367464

14 Mishra M. Convolutional Neural Networks, explained [Internet]. Medium. Towards Data Science; 2020 [cited 2022Apr30]. Available from: https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939

15 Brownlee J. Crash course in Convolutional Neural Networks for machine learning [Internet]. Machine Learning Mastery. 2019 [cited 2022May1]. Available from: https://machinelearningmastery.com/crash-course-convolutional-neural-networks/

16 Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of Deep Learning: Concepts, CNN Architectures, challenges, applications, future directions - Journal of Big Data [Internet]. SpringerOpen. Springer International Publishing; 2021 [cited 2022Apr30]. Available from: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8

17 KumarI, A. Different types of CNN Architectures explained: Examples [Internet]. Data Analytics. 2022 [cited 2022Apr30]. Available from: https://vitalflux.com/different-types-of-cnn-architectures-explained-examples/

18 Ann vs CNN vs RNN: Types of neural networks [Internet]. Analytics Vidhya. 2020 [cited 2022Apr30]. Available from: https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/

19 Dasiopoulou S, Mezaris V, Kompatsiaris I, Papastathis V-K, Strintzis MG. Knowledge-assisted semantic video object detection. IEEE Transactions on Circuits and Systems for Video Technology. 2005;15(10):1210–24.

20 Patel A. What is object detection? [Internet]. Medium. ML Research Lab; 2020 [cited 2022Apr30]. Available from: https://medium.com/ml-research-lab/what-is-object-detection-51f9d872ece7

21 RUGERY P. Explaining Yolov4 a one stage detector [Internet]. Medium. Becoming Human: Artificial Intelligence Magazine; 2020 [cited 2022Apr30]. Available from: https://becominghuman.ai/explaining-yolov4-a-one-stage-detector-cdac0826cbd7

22 Joseph L. A gentle introduction to Yolo V4 for object detection in ubuntu 20.04 [Internet]. Robotics. 2020 [cited 2022Apr30]. Available from: https://robocademy.com/2020/05/01/a-gentle-introduction-to-yolo-v4-for-object-detection-in-ubuntu-20-04/

23 Boesch G. Object detection in 2022: The Definitive Guide [Internet]. viso.ai. 2022 [cited 2022Apr30]. Available from: https://viso.ai/deep-learning/object-detection/

24 Gandhi R. R-CNN, fast R-CNN, Faster R-CNN, YOLO - object detection algorithms [Internet]. Medium. Towards Data Science; 2018 [cited 2022May1]. Available from: https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e

25 Papers with code - R-FCN explained [Internet]. R. [cited 2022Apr30]. Available from: https://paperswithcode.com/method/r-fcn

26 Feature descriptor: Hog descriptor tutorial [Internet]. Analytics Vidhya. 2020 [cited 2022Apr30]. Available from: https://www.analyticsvidhya.com/blog/2019/09/feature-engineering-images-introduction-hog-feature-descriptor/

27 Papers with code - SPP-net explained [Internet]. Explained | Papers With Code. [cited 2022Apr30]. Available from: https://paperswithcode.com/method/spp-net

28 Raspberry pi introduction: Raspberry pi [Internet]. ElectronicWings. [cited 2022Apr30]. Available from: https://www.electronicwings.com/raspberry-pi/raspberry-pi-introduction

29 Raspberry pi 4 computer model B [Internet]. [cited 2022Apr30]. Available from: https://static.raspberrypi.org/files/product-briefs/Raspberry-Pi-4-Product-Brief.pdf

30 Raspberry pi 4 model B specifications [Internet]. Raspberry Pi. [cited 2022Apr30]. Available from: https://www.raspberrypi.com/products/raspberry-pi-4-model-b/specifications/

31 What is python? executive summary [Internet]. Python.org. [cited 2022Apr30]. Available from: https://www.python.org/doc/essays/blurb/

32 What is python used for? A beginner's guide [Internet]. Coursera. [cited 2022Apr30]. Available from: https://www.coursera.org/articles/what-is-python-used-for-a-beginners-guide-to-using-python

33 The absolute basics for beginners [Internet]. NumPy. [cited 2022Apr30]. Available from: https://numpy.org/devdocs/user/absolute_beginners.html

34 Tensorflow/tensorflow: An open source machine learning framework for everyone [Internet]. GitHub. [cited 2022Apr30]. Available from: https://github.com/tensorflow/tensorflow

35 Raspberry Pi tensorflow 2 installation and Yolo V3 object detection [Internet]. PyLessons. [cited 2022May1]. Available from: https://pylessons.com/YOLOv3-TF2-RaspberryPi

36 Politiek R. Install tensorflow 2.2.0 on raspberry 64 OS - Q-engineering [Internet]. Q. Q-engineering; 2022 [cited 2022Apr30]. Available from: https://qengineering.eu/install-tensorflow-2.2.0-on-raspberry-64-os.html

37 EdjeElectronics. Tensorflow-lite-object-detection-on-android-and-raspberry-pi [Internet]. GitHub. 2020 [cited 2022May1]. Available from: https://github.com/EdjeElectronics/TensorFlow-Lite-Object-Detection-on-Android-and-Raspberry-Pi/blob/master/Raspberry_Pi_Guide.md

38 Redmon J. [Internet]. Yolo: Real-time object detection. [cited 2022Apr30]. Available from: https://pjreddie.com/darknet/yolo/