硕士学位论文

（学术学位论文）

使用深度学习的水草检测

AQUATIC WEEDS DETECTION USING DEEP LEARNING

RICHARD MICHAEL MKECHERA

理查德·迈克尔·麦克切拉

哈尔滨工业大学

2022年08月

国内图书分类号：TP391.4　　学校代码：10213
国际图书分类号：004.8　　密级：公开

计算机科学硕士论文

使用深度学习检测水草

研 士 研 究 生:　RICHARD MICHAEL MKECHERA
导　　师:　王玲教授
申 请 学 位:　计算机科学硕士
学　　科:　计算机科学与技术
所 在 单 位:　计算机科学与技术学院

答 辩 日 期:　2022 年 08 月
授 予 学 位单位:　哈尔滨工业大学

Classified Index: TP391.4
U.D.C.:004.8

Dissertation for the Master's Degree in Computer Science

AQUATIC WEEDS DETECTION USING DEEP LEARNING

Candidate:	Richard Michael Mkechera
Supervisor:	Prof. Wang Ling
Academic Degree Applied for:	Master's in Computer Science
Specialty:	Computer Science and Technology
Affiliation:	School of Computing
Date of Defense:	August 2022
Degree-Conferring-Institution:	Harbin Institute of Technology

摘 要

世界各地各种水源的水生杂草已经呈指数级增长，例如位于非洲东部（坦桑尼亚）的坦噶尼喀湖（6.2556 S，29.5108 E）和维多利亚湖（0.7558 S，33.4384 E）。水生杂草在各种水源中的生长造成了许多负面影响，如环境影响、航运和交通影响、经济影响和农业影响。非洲和世界上大多数国家都试图减少或消除水体中水生杂草的生长问题。然而，由于水生杂草生长较快，使得该任务是非常重大的挑战。一些清除水源中水生杂草的传统方法，例如将昆虫移植到受影响的区域，消耗杂草或去皮、干燥和焚烧等是非常繁琐和耗时的。并且，这些传统方法都面临着同样的挑战，即定位和区分水生杂草与其他水生植物，这对维持水生生物是十分重要的。本文提出了一种新的应用场景：水生杂草的检测。检测水生杂草任务形成了许多潜在应用的基础，例如使用检测到的水生杂草来确定受影响水源的位置，在水生杂草到达水体之前使用水生杂草作为参考来搜索受影响区域。

本文首先收集了基于坦桑尼亚坦噶尼喀湖（6.2556 S，29.5108 E）和维多利亚湖（0.7558 S，33.4384 E）的第一批水生杂草数据集各300和588幅，以及布莱克河（191.344 S，146.3745 E）的1098幅图像。由于缺乏足够的数据，本文采用了一些数据预处理技术，如数据扩充和数据归一化。我们通过在90，180和270度处旋转图像来执行增强过程，还在0.05的范围内执行翻转、移位和缩放，得到5490张图像，本文对这些数据做出标记。本文考察了深度学习模型对水生杂草检测准确性的检测效果，YOLOv4的准确率为93.78%，高于YOLOv3的89.93%。此外，本文对不同水体的水生杂草检测任务进行了对比实验，发现在良好光照条件下拍摄的图像具有更好的效果，将所有图像组合起来有助于提高检测精度。本文展开了构建深度学习模型以确定神经网络的学习能力的研究，发现原始YOLOv4没有足够的表示能力来学习数据集的小特征，其原因是多次下采样会导致网络收敛缓慢和网络退化。因此，本文通过减少卷积层的数量，提出了一种轻量级的特征提取网络作为模型的主干。此外，本文用FPN代替了PANet进行特征融合，其能够在多个尺度上提取特征以解决数据规模小的问题。本文还使用LReLU激活函数来代替Mish激活函数，其由于其简单性和计算效率而更加匹配数据集的性质。实验结果表明，改进后的模型可以有效地提高模型的检测能力，准确率达到97.79%，比原来的YOLOv4高出4.01%。此外，本研究还通过提出一个CNN模型来确定图像识别对水生杂草数据集的影响，该模型的训练准确率约为99.63%。

关键字：深度学习，YOLOv3，YOLOv4，水生杂草检测，卷积神经网络。

Abstract

We have seen an exponential increase in aquatic weeds in various water sources around the world. For example, in Lake Tanganyika (6.2556 S, 29.5108 E) and Lake Victoria (0.7558 S, 33.4384 E), which are located in the eastern part of Africa (Tanzania). The growth of aquatic weeds in the various water sources caused many negative impacts such as environmental impact, shipping and traffic impact, economic impact and agricultural impact. Most countries in Africa and in the world in general have tried a lot to reduce or eliminate the growth of aquatic weeds in the water bodies. However due to the fastest growth of aquatic weeds, this posed a very significant challenge. Some traditional methods have been used to remove aquatic weeds from water sources, such as transplanting insects into affected areas to consume aquatic weeds or peeling, drying and burning, which is a very tedious and time-consuming work. However, all of these traditional methods face the same challenge of locating and distinguishing aquatic weeds from other aquatic plants, which are very useful in supporting aquatic life. A new application scenario is proposed in this paper, for the detection of aquatic weeds. The ability to detect aquatic weeds forms the basis for many potential applications, such as the ability to use the detected aquatic weeds to determine the location of the affected water source, using the aquatic weeds as a reference before they reach water sources in search of the affected area.

In this paper, we first collected the first aquatic weed data sets based on Lake Tanganyika (6.2556 S, 29.5108 E) and Lake Victoria (0.7558 S, 33.4384 E) in Tanzania. This paper managed to collect 300 and 588 for Lake Tanganyika and Lake Victoria, respectively. There was also another source, namely Black River (191344S, 1463745E), which creates a total of 1098 images. Due to the lack of sufficient data, some data preprocessing techniques such as data augmentation and data normalization have been performed in this paper. This paper performed the augmentation process by rotating the image at 90, 180 and 270, also performed flips, shifts and zooms at a range of 0.05, resulting in 5,490 images. In this paper, all 5,490 images were labeled, which happened to be a very tedious work, but finally this paper completed the labeling work for all datasets. This paper examines the detection effect of deep learning models on aquatic weed detection accuracy, with YOLOv4 having a higher achieved 93.78% accuracy than YOLOv3, which achieved 89.93% detection accuracy. In addition, in this paper, comparative experiments were conducted with different bodies of water on the aquatic weed detection task, and this work developing a deep learning model to determine the learning capacity of neural network. Our study found out that the original YOLOv4 does not have enough representation power. Found that the images captured under very good light intensity are more helpful, and combining all the images helps to improve the detection accuracy. The model architecture was seriously observed in this study, when developing a deep learning model to determine the learning capacity of neural networks. Our study found that the original YOLOv4 does not have enough representation power to learn small features of our datasets, since multiple downsampling leads to slow network convergence and network degradation. Our study has proposed a lightweight feature extraction network as the backbone of our proposed model by reducing the number of convolutional layers. Also we replaced PANet with FPN for feature fusion because its ability to extract features at multiple scales is an advantage of the small size of our data. Also the paper used the LReLU activation function to replace the Mish activation function as it favors the nature of our data sets due to its simplicity and computational efficiency. The experimental results show that the improved model can effectively improve the model's detection ability, as it achieves an accuracy of 97.79%, which surpassed the original YOLOv4 by 4.01%. However this study also determined the impact of image recognition in aquatic weed datasets by proposing a CNN model that results in a training accuracy of about 99.63%.

Keywords: Deep learning, YOLOv3, YOLOv4, aquatic weeds detection, Convolution Neural Network.

Contents

摘 要..I
Abstract... II
Chapter 1 Introduction.. 1
1.1 The background of the subject and the purpose and significance of the research. 1
1.1.1 Subject background and the research significance.. 1
1.1.2 Research purpose.. 5
1.2 Research status at home and abroad 6
1.2.1 Object detection and tracking 6
1.2.2 Convolution Layers 10
1.2.3 Model lightweight 11
1.2.4 Feature analysis of the aquatic weeds 12
1.3 Brief analysis of domestic and foreign aquatic weeds control............................ 15
1.4 The main research content of this paper 16
1.5 Organization of this thesis.. 17
Chapter 2 Related theory and technology 18
2.1 Introduction... 18
2.2 General information on Neural Networks 18
2.3 Pooling layer and non-linearity layer 20
2.3.1 Pooling layer.. 20
2.3.2 Non-Linearity layer.. 21
2.4 Deep learning algorithms 21
2.4.1 YOLOv3... 21
2.4.2 YOLOv4... 24
2.5 Chapter Summary 26
Chapter 3 Collection and Processing of Aquatic Weeds Datasets 27
3.1 Introduction.. 27
3.2 Collection of the aquatic weeds datasets... 28
3.3 Preprocessing of Aquatic Weeds Datasets... 30

3.3.1 Improvement of the quality of aquatic weeds images 32
3.3.2 Implementation of image augmentation 36
3.3.3 Implementation of image normalization... 38
3.4 Arrangement, labeling tool and labeling process for aquatic weeds datasets 40
3.5 Chapter Summary 41
Chapter 4 Research and comparison of aquatic weeds detection algorithms based on deep learning 43
4.1 Introduction... 43
4.2 Comparison and research of different deep learning models on aquatic weeds detection tasks 44
4.2.1 Deep learning model used and its implementation in detail 44
4.2.2 Comparison and analysis of the experimental results................................. 47
4.3 Proposed model for aquatic weeds detection... 51
4.3.1 Design and implementation of the proposed model 51
4.3.2 Comparison and analysis of the experimental results................................. 60
4.4 Research on the influence of image recognition tasks on the aquatic weeds datasets.. 63
4.4.1 Proposed model and its implementation in details 63
4.4.2 Analysis of the experimental result on the influence of image recognition on aquatic weeds dataset... 64
4.5 Chapter summary... 66
Chapter 5 Benefit, Validity, Impact and Future exploration.................................... 70
5.1 Introduction 70
5.2 Benefit of the research 70
5.3 Validity... 71
5.3.1 Internal validity.. 71
5.3.2 External validity... 72
5.4 Impact of the research on aquatic weeds 73
5.5 Future exploration.. 74
5.6 Chapter Summary 75
Conclusion... 76
Harbin Institute of Technology dissertation originality statement and use rights ... 78
References 79

Acknowledgment 84

Chapter 1 Introduction

1.1 The background of the subject and the purpose and significance of the research

1.1.1 Subject background and the research significance

Water is very important natural resource that supports all forms of life in this world, therefore the appropriate management of water resources appears to be of much more important due to its role into human life. Aquatic weeds management has been a very big challenge to many countries around the world now, this is due to the reason that in the aquatic environment there are different types of aquatic plants. Some of the aquatic plants are more useful and supportive in aquatic eco-system, but all these kind of plants sometimes they tend to grow together with aquatic weeds which are referred to be harmful enough to the aquatic environment. Therefore the challenge comes when managing aquatic weeds out of other useful plants. It has been a long period of time now countries around the world have been struggling to eliminate aquatic weeds from water sources. Lidia [1] has pointed out traditional methods that have been used to eliminate aquatic weeds. These methods are like physical removal, biological control or the use of herbicides. Since traditional control, Carole [2] involves the whole surface of the water body covered by the aquatic weeds, such kind of methods involves a high volume of herbicides, time consuming especially when physical removal method is applied to eliminate aquatic weeds.

A poor application of aquatic weed elimination method may lead to the damage of the aquatic environment which threatens the life of aquatic living organisms. In order to reduce the impact of the tradition methods of aquatic weeds control, the idea of the aquatic weed’s detection is proposed in this study. One of the biggest challenges of the aquatic weeds control is to distinguish between aquatic weeds and other aquatic plants, which are very useful in supporting aquatic life. Otsu threshold [3], have helped to give out the idea of discriminating weeds and other useful plants which are mostly needed to support aquatic eco-system. A well and successfully aquatic weeds control, normally has a lot of advantages, Torawane[4] has pointed out several significance of aquatic weeds such us.

(1) **Biodiversity conservation**, aquatic weeds can destroy natural ecosystem by imposing outcompeting native plants species, reducing biodiversity and destructing habitat structures. Therefore controlling aquatic weeds helps to restore ecological balance.

(2) **Water quality improvement**, aquatic weeds can negatively impact water quality by destructing oxygen levels realizing toxins and obstructing water flow. Therefore eliminating these weeds can improve water quality.

(3) **Flood control**, normally the excessive growth of aquatic weeds can impede water flow leading to the increasing of the flood risk. Removing aquatic weeds will help to increase water flow to reduce all chances for flood to occur.

(4) **Disease prevention**, some aquatic weeds can act as host for diseases vector such as mosquitoes and snails which can transmit the diseases for humans and animals. Removing these aquatic weeds reduces the breeding grounds for disease that slows down the possibility of the disease outbreak. Consider the image below that shows how the aquatic weeds affect water bodies by spreading over the surface of water body.

Figure 1-1 Effect of aquatic weeds on the water source

The above figure shows how the aquatic weeds cause a very big and harmful effect to the aquatic environment. If it is observed well, it is noticed that the aquatic weeds on the picture affects the growth of some other aquatic plants by reducing the supply of oxygen on the water. Since they tend to cover the upper surface of the water body, and hence they affect the whole process of eco-system on the aquatic environment. However, the presence of aquatic weeds into different water bodies around the world initiates a serious management of water resources. A failure to control aquatic weeds on different water bodies’ results to the following effects.

(1) **Environmental impact**, mostly we use canals to supply water to different areas but due to the existence and growth of aquatic weeds in large water sources, it caused the blockage of canals that supply water to various areas. Including farms, living areas and in rivers. Therefore there will be a high shortage of water, since there will be poor availability of water this leads to drought and thus the plants dry up due to the shortage of water Samah[5], life of plants cannot be possible without water as a result it affects the entire environment. This is only reason that enforces many government to fight against aquatic weeds. As

researchers we saw that there is a great need of thinking on how we can use technology to identify aquatic weeds in an easy way, because it is very challenging to differentiate aquatic weeds and other aquatic plants. Therefore capable systems that will actually assist people to fight against aquatic weeds should be implemented.

(2) **Economic impact**, the existence of aquatic weeds in water sources affects different economic sectors like fishing industry, recreation sites and power generation (Hydroelectric power), which in actual sense contribute much on the increase of the government revenue. Fishing industry provides many job opportunities to people, therefore the existence of the aquatic weeds on water bodies will affect the aquatic life and furthermore it can even affect natural biodiversity which endangers aquatic life.

(3) **Agricultural impact**, as we know that aquatic weeds normally grow on the areas where water is mostly available. In the agriculture especially irrigation agriculture we depend much on the water flow through the canals, but due to the existence of the aquatic weeds especially alligator weeds are the nuisance specie in irrigation canals and drainage ditches. This is due to the reason that alligator weeds tend to impedes water flow in irrigation canals which may lead to the decrease of water flow in the canals as a result that water fails to reach to the exactly destination like farms. This may lead to poor production of agricultural products due to the lack of enough water in farming areas, on the side of the drainage ditches, aquatic weeds tend to block culverts, if culverts are blocked then during the high rainfall in the areas where culverts are blocked water will have no place to go therefore this will cause water to take the direction that is not convenient which may lead to the occurrence of floods.

(4) **Navigation and Transport impact**, th**e** survey that was made on the effects of aquatic weeds on East Africa pointed out two lakes, Lake Tanganyika and Lake Victoria. These lakes were highly affected by the aquatic weeds to the extent that the government stopped transportation through Lake Victoria. This automatically brought the huge effect on the communication and marketing of the products to different areas. Navigation and transport via water sources like oceans, lakes and rivers happens to be a major means of communication and interaction between people from different places, Carole [6]. Aquatic weeds act as an obstacle to navigation by creating blockages along rivers and canals. Though most of the countries around the world have been trying much to reduce or to eliminate the effect of aquatic weeds on water sources to increase productivity that leads to the increase of the national revenue. Therefore our study will be centered on direct recognition and tracking of aquatic weeds by using Deep learning techniques which will actually accelerate aquatic weeds management and control.

A lot of studies have been conducted to the areas which are highly affected by the aquatic weeds, in our research we have considered the study which was conducted at Sudan-Gezira, Samah[5] . This study was conducted aiming at knowing farmers and educated people awareness on the effect of aquatic weeds to the water resources, the research was conducted to the community of the farmers who depend on water flow from the river toward the farms. Findings were very positive that a big percent of educated individuals know the effect of aquatic weeds to the water sources [7], but many pointed out that the only problem faced is how to control them and how to recognize them at the early stage of growth. Our research answers these questions because the core of our research is to find a better way at which humans will have a simple way to recognize aquatic weeds. Therefore consider the figure below that shows the way how the farmers responded on awareness of aquatic weeds**.**

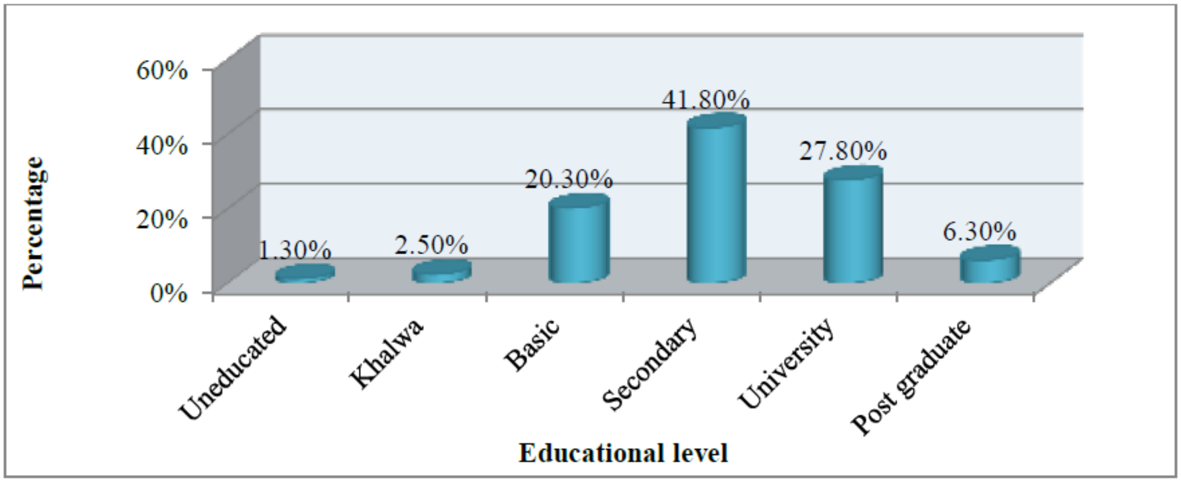


Figure 1-2 Educational level of responders, Gezira Scheme, Sudan 2018

Object detection has now become an eye toward the area of computer vision, and a lot of research has been conducted on object detection. In recent years we have seen the impact of this research in the area of computer vision which brought more improvement compared to the previous technologies. The revolution of deep learning algorithms brought a very big impact on the entire field of computer vision. The revolution of object detection techniques has revealed its necessity in the field of computer vision by performing the detection of visual objects from different digital images. Normally computer vision application requires computation models to perform certain task like identifying and classifying a particular object in a certain setting or group of data. Due to contribution of deep learning algorithms toward the field of computer vision. Now different researchers are able to come with some new research projects which are useful and productive to the society, this is due to the fact that the only objective of the object detection is to detect and to allocate an instance of every object in the digital images.

The introduction of many detection models brings a research challenge to many researchers because all models rely on accuracy, therefore the most accurate model is the one that is normally preferred for computer vision applications. The accuracy normally is categorized into two major parts like classification accuracy which specifies by how much accurate is the particular object differentiated from another object on the digital image or video. Localization accuracy which defines the x and y coordinates of the position of an object on the digital image. However, several object detection algorithms has been proposed such as Histogram of Oriented Gradient HOG[8], Religion-based Convolution Neural Network, R-CNN[9], Faster RCNN[10], Single Short Detector SSD[11], You Only Look Once YOLO[12] and RetinaNet[13]. However the accuracy and cost-effectiveness of these algorithms have been the major improvement and the development of the field of computer vision.

1.1.2 Research purpose

The main purpose of this work is to accelerate the entire process of aquatic weeds control to manage and reduce the negative impact of aquatic weeds on water ecosystems, human activities and the environment in particular. Effective weeds control enables efficient water management, ensuring a reliable supply of water for agricultural, industrial and domestic uses. A failure to control aquatic weeds at their early stage may lead to the clog of the irrigation systems, canals and water intake infrastructures. This may result to the reduction of their efficiency and hindering water distribution. This study proposes object detection model that will be very effective on detecting aquatic weeds, which finally gives the right target and the right decision on the best method to eliminate aquatic weeds. In order to come up with a good model to detect aquatic weeds our study is centered on deep learning algorithms.

Deep learning models (algorithms) have been mostly used in recent years as an approach to detect different objects as an instance in images and in video in real-time. The accuracy of each and every algorithm happened to be accurate enough for detection of objects .Most of these algorithm like YOLOv4 [14], YOLOv3[15], RCNN[9] and others have been proved to work best on the detection of the objects but many of these approaches have been facing some challenges that affect their performance, accuracy and speed. Some of these challenges are like detection of the small objects in an image, this normally is caused by the limit of the size of an object on the image and the aspect ratio that means all objects which do not meet the requirement of the size and aspect ratio they won’t be recognized by the deep learning models. Another challenge is training time, the higher the training time the lower the speed. However the algorithms require high computing power due to the time taken to train a model, training a model lead to the time consuming especially when training a large volume of data set. Slow convergence of the neural network is also a challenge due to too low learning rate and at the same time if the training time is set to the too high value it may lead to the divergent behavior of the loss function pointed by Priyanka [16]. There are some other external factors that can also affect deep learning models like occlusion, this affect much deep learning algorithm due to the increase of the number of the false positive or poor detections which leads to low accuracy. Sometimes some images are occupied by noisy backgrounds which also may lead to the low accuracy of the model due to the increase of the number of false positive and finally the number of datasets at which the deep leaning models is trained to predict or increase the accuracy of the model.

As we pointed earlier that one of the biggest challenge of eliminating of aquatic weeds is to identify and to locate aquatic weeds on the water sources. This is due to the fact that during the process of removing aquatic weeds, we sometimes tend to remove plants species which are mostly useful in supporting aquatic life and ecosystem of the aquatic environment. In this study we have done a very comprehensive study on how we can accelerate the process of removing aquatic weeds images.

1.2 Research status at home and abroad

1.2.1 Object detection and tracking

For many years now, object detection has been a winning tool for different research topic in computer vision and it is the core of several computer vision task. The problem of point detection normally comprises the detection with assumption that the object to be detected is available or not. This is done by detecting the type of the target and localizing the target if at all the target is available, therefore it needs to return the type of the target, its position and coverage on the image .In this case for most computer vison related researches, object detection is of much more important, this is due to the reason that every successful research like facial recognition, image segmentation, re-identification, activity recognition, target tracking any many others. For all these to be implemented they all need object detection techniques. When it comes to the commercial product or the application of the object detection in real life like driverless cars, electronic consumption, human computer interaction, content based-image retrieval and intelligence image surveillance.

A lot of researches has been done in the field of computer vision especially in the area of the object detection. The very first research which brought a new hope in this field of computer vision by innovating object detection models are as follows component-based models and template matching technique, Fischler [24] .Our study pointed out the component that can be detected on the part of the aquatic weed image. In those days the main technique of object detection was actually based on the geometric representation technique which was proven to be a best approach in those days. Researches continued to be done and a lot of innovation has been brought up until when they came up with an innovation of the feature extraction using traditional Machine learning algorithm such as SVM [25]. Feature extraction capability brought a revolution and a very big impact on the accuracy of the deep learning detection models, Ponce [26] .There was also traditional algorithm like Adaboost [27], and this was also good enough to represent some features extraction. Although the tradition algorithm were outstanding but they were very much affected by different factors like occlusion and lighting, this is due to the reason that, they were not good enough to adopt changes that were caused by both internal and external validity. In this case the accuracy of the object detection and feature extraction is affected much by the external factors. However the development and the improvement of the deep learning models in each and every model has now happened to be a very powerful tool for automatic object detection, Hinton [28]. This is due to the application of the different features like feature representation and some other techniques has led to high improvement of these algorithms. A big number of innovations and researches on object detection has been done, the introduction of CNN has been performing very well on the object classification, Krizhevsky[29]. Due to the existence of challenges and problems that face the field of object detection and tracking, there was a very steady improvement of the technology toward deep learning technology by improving the computing power and the accuracy of the model. However the improvement of these algorithms is actually focusing on the speed and accuracy of the model in order to overcome the problem of the external validity. Therefore all the models which were innovated were very adoptive, some of the models that outperformed object detection task were like Fast R-CNN [10], that were able to resolve challenges encountered by R-CNN [9] and SPPNET applied [30]. All these were done in order to improve speed, quality and accuracy. Fast R-CNN [10], normally performs training by developing a training modality that learns and specifies bounding boxes and regression at the same time. The sharing of the convolution proposals across regions is adopted, and the final convolution layer and the first region of the FCA is added between layers (ROI) for what is called layer combination for the extraction of the fixed length for region proposal.

A single stage object detector which has been proven to have a very better performance and very effective in object detection is YOLO model [12]. This deep learning algorithm is a regression problem from the pixel of an image to partial separated bounding boxes and it is also depend on the class probabilities. YOLO[12] algorithm fixes all the weakness of the regional proposed model by implementing a small groups of the related regions which leads to the predictive religions directly with region-based methods like RCNN[9], to predict the detection based on the feature of the localized area. YOLO[12] uses global feature of the entire image .The specific method in YOLO[12] algorithm tends to divide the image into S*S grids and predict the probabilities of the bounding box, location and confidence score for each grid. In this case YOLO [12] algorithm outperforms other algorithm by 45 FPs running at the speed of the real time. In these years the series of YOLO [12] algorithm version has reached the fifth version, Ramya [31] which is practically proved to be much faster compared to the previous version. However object detection is still a challenge up to now but due to the number of the researches that are going on, thus makes the object detection algorithm with a very good accuracy to be more promising, Liu [32]. Tracking can be categorized according to the different criterial, Jun [33]. All categories has been defined as follows one is classification according to the initialization method can be divided into detection tracking and non-detection tracking. Detection tracking is to detect object and then to combine to become trajectories and non-detection tracking demands manual initialization of the fixed objects in the first frame and continue positioning these objects in a very subsequent frames, therefore there are still a number of researches and innovation on the area of object detection and tracking.

The identification and detection of aquatic weeds has become a research problem of great concern to researchers. This is because the demand for clean water from water bodies has increased due to global population growth. However seventy percent of the world is made up of water systems, including oceans, lakes, rivers and manifolds, and yet the problem of water unavailability is increasing in some areas around the world every day. According to research, many factors have been pointed out, including the existence of aquatic weeds in Surrounding of water bodies, Robert [34]. Aquatic weeds affect not only the availability of water but also the life of living organisms in the water bodies. By reducing the circulation of nutrients to other beneficial aquatic plants and affect the circulation of the oxygen gas, to the extent that that, there is no enough oxygen to support the aquatic life. The deep learning-based aquatic weed detection method not only has important academic research value, but also has a very broad market application aspects. The inversion of aquatic weeds in a given water body can lead to a very large impact that needs to be controlled at any earlier stage. This is due to the growth of the quantitative weeds being too rapid to the extent that the growth increases with time. However if aquatic weeds are controlled at a very early stage, it would be easy to overcome them in a very short moment, due to the fortune of the comprehensive and sufficient details on the process of controlling aquatic weeds. This study focuses on proposing the best method, which focuses on the convolutional neural network technology, to detect the presence of the aquatic weeds on the surface of water body. Why CNN, the Convolutional Neural Network model consists of Input Layer, Convolution Layer, Pooling Layer, Full Connection Layer and Output Layer. In a model, the convolutional layer and the pooling layer alternate several times, and when the convolutional layer neurons are connected to the pooling layer neurons, Jun [33]. Since our data set basically relies on the image dataset, therefore our data have been prepared and cleaned to make sure that all the preferred images to be used by our model will be good enough to give us a clear information, ending up getting a proper and accurate result. There are a number of reasons why we have chosen deep learning technology as our detection approach. Compared with other image recognition methods, deep learning-based image recognition technology does not need to extract specific features, and suitable features can only be found through iterative learning, which can grasp global and contextual features of images. This has a strong robustness and higher detection accuracy. Since deep learning technology is highly dependent on feature extraction, this study also paid attention to feature extraction by developing a trait extractor that works best at extracting various common features of aquatic weeds. Therefore detection approach basing on deep learning has been considered as the best approach to accelerate the process of fighting against the growth of the aquatic weeds in order to keep and reserve water bodies and maintain an aquatic ecosystem.

However a number of the solution has been pointed out basing on current technology, there were a number of ways used to eliminate aquatic weeds such as the use of chemicals which are harmful to the aquatic organisms, the use of the insects on the target domain that consumes aquatic weeds and the traditional way of pealing, drying and burning aquatic weeds but the questions had always remained to be the same that how effective are these ways are? How do they differentiate aquatic weeds with other aquatic plants, therefore with the use of the current technology especially the use of deep learning approaches .The researchers have obtained a powerful weapon to fight with the exponential growth of aquatic weeds by developing detective models that actually detect and allocate the presence of aquatic weeds. Therefore in this subsection we will see how far they have gone and what challenges they have faced in detecting aquatic weeds. In the research done by the Olsen [34], they prepared a multiclass data set and they performed classification and detection and the accuracy of the model was evaluated. The accuracy in each and every model used was as shown on the table below.

Table 1-1 Olsen model accuracy

S.no	Model used	Accuracy
01.	ResNet-50	95.7%
02.	Inception-v3	95.1%

In their paper data set was classified using CNN ResNet-50[36] and Inception-v3[37] models to initiate the point of comparison of the baseline performance of the selected models and as you can see from the table above ResNet-50[36], has a very good accuracy compared to Inception-v3[37]. Also in the paper by Ferreira [38], using the same dataset introduced by Olsei[35], they had performed a research basically on the classification of the deep weeds in their paper they used method like graph weed net, Kun[39]. Deep Clustering for Unsupervised Learning of Visual Features and their accuracy were as follows 95% and 98% respectively. When deploying a model to the system on object detection Göktoğan[41] had developed a model that is represented on their paper called A Rotary-wing Unmanned Air Vehicle for Aquatic Weed Surveillance and Management in this research they were able to use different algorithms to classify and to detect aquatic weeds using air vehicle. But in their paper they don't point out how the robot will differentiate aquatic weeds with other aquatic plants, Michael [40] on their paper of monitoring aquatic weeds in a river system using SPOT 5 satellite imagery. They used remote sensing techniques to monitor the change of dense water weeds and some other method used in this study are like geometric and radiometric, linear spectral unmixing and spectral mapper techniques. Therefore in their study they reported the accuracy of the detection varies between 81% and less than 95%. A number of the researches on aquatic weeds are still going on but it is advised much a seriousness to be employed during developing the models on controlling aquatic weeds, Kentucky [42].

1.2.2 Convolution Layers

Convolution layers as set of layers at which filters are applied to the original image of which are to be learned throughout the training. It is only on the convolution layers where the developer or a researcher specifies parameters, and when we talk of the parameters we mean specifically the number of kernel and the size of kernel. What is so important to note in here is when specifying the number of filters and the size of the filter we should normally consider that the size of the filter usually has to be smaller than the actual image. The training to be successful we must create activation map this is due to the technical reason that each filter must convolute with image. There is a very big reason for the Convolution Neuron Networks to understand what the features to detect are. In this case then we need to apply the concept of the feature map where by Convolution Neural Network captures the result of applying the filters to an input image, this is because the feature map reflects the output of each layer within a network. Assuming that Convolution layers receive as input image called P^{m-1} with R_m channels and after undergoing several computation it gives output as P^m composed of N_m channels the output of each channel is what we define as a feature map, Muhammad [43] and it is represent by the following formula as it is shown below.

$$P^m = g_m\left(\sum\nolimits_k W * P^{m-1} + b^m\right) \tag{1-1}$$

In the formula P^m ——denotes the output
P^{m-1} —— input image

When designing a convolution layer there are things to be considered which can improve the performance of your model or can affect your model. The design of the feature extractor normally is of more important to be careful when developing a Convolution Neural Network this is due to the fact that the type of layers and number of parameters will not be considered carefully, they will automatically affect the speed of the detector, the memory of the detector and the performance of the detector, Zhipeng [44]. Sometimes there is a challenge of capturing both small and large objects simultaneously. This is of much more important also to work at, because a good convolution neural network must be able to capture both small and large object. This can only be done if flexible convolution filter size with a layer-by-layer so that to make the model with a less number of channels and big number of layers which will automatically increase the efficiency of the model. Therefore in this study we have decided to consider Convolution Neural Network because it is our main goal on the improvement and designing our proposed model for the accurate results. However a well-developed Convolution Neural Network oriented model does not guarantee the better results, but there are some minor factors like the size of dataset at which our algorithm will work at. Therefore our study will observe both internal and external factors during the training to make sure that we get the best object detector which will actually be the best deep learning approach for aquatic weeds detection.

1.2.3 Model lightweight

Convolution Neural Networks have been noticed in recent years, to be preferred by many researchers in designing or developing deep learning models. However the design of deep learning algorithms normally focuses on the attributes like reduction of time and computation power. Therefore different approaches are used for optimization of the model in order to reduce the computing power, model size and faster inference is mostly needed for practical applications. It has been a practice in these days for different research work, most of the researchers have proposed light weight convolution neural network architectures. Some of these architectures are optimized from existing architecture by applying different tactics to achieve the required model of which is referred to be a lightweight model. There are several ways at which the lightweight model can be obtained, one is by modifying mathematical operations of the existing model which results to a very fewer parameters and the increase of the speed of the particular model, second by optimizing the architecture of a certain deep learning algorithm by reducing the number of the convolution neural network, with the intention of increasing the training speed and the accuracy of the model.

SqueezeNet [45], this algorithm is implemented with an architecture that has fewer parameters and a very minimal size. This model was modified on the technique of the model compression where by approaches like pruning and weight sharing for the pre-training model which leads to the reduction of the model size. Also they implemented a so called convulationfire module which was made up of two parts, the first part is the usage of 1*1 convolution kernel compression layer and the second part is the usage of 1*1 and 3*3 convolution kernel extended layer. Due to the implementation of the large 1*1 convolution, the number of the parameters of the model is significantly reduced. MobileNet [46], in this algorithm separable convolution is used insteady of tradition convolution. In terms of depth separable convolution is highly preferred in this algorithm and it is divided into two major parts these are depthwise convolution and point wise convolution. Depth wise convolution is a group of convolution that is, each input feature map is used to perform operation and pointwise operation is made up of depth that is equivalent to the input which is actually 1*1 convolution which is used to combine different operation. Therefore the application of the depth convolution reduces the increase of the number of channels and fusing the channel.

In 3*3 which is actually a single convolution kernel depthwise separable convolution can easily be achieved. The advantage of this is that the amount of the calculations is reduced twice and the loss of precision is too small. However the network structure can be made lighter by increasing the width and resolution factor MobileNetv2 [47], this is the improved version of MobileNet [46], this was improved by adding a bottle neck which leads to the increase of the input data (volume) by doing so it solves the problem of having the large number of empty convolution kernel. ResNet [48], residue block in input which leads to the improvement of the extraction ability of the network. The improvement of the gradient propagation ability and the acceleration of the network convergence during the training. ShaffleNet[49], this algorithm adopted pointwise group convolution and at the end is improved by reducing computation cost by considering 1*1, convolution is grouped to reduce the amount of computation. Also the number of the parameters was greatly reduced. ShaffleNetv2 [50] ,This is also designed with the intention of having lightweight model which leads to the improvement of the network speed due to the number of the input and the output.

1.2.4 Feature analysis of the aquatic weeds

In this study we have even went further by proposing a model that will actually improve the accuracy of our object detector, to be of more accurate compared to the previous one. Before developing our model we have made a comprehensive study of our object to be detected, this is due to the reason that any object that has to be detected by a certain deep learning algorithm must be described in terms of its common feature. For example if someone develops a model to detect a human face, the developer should derive the possible and common features of the human face and these features are like eyes, nose, ears and mouth. Therefore, in this case during the training the algorithm will learn all the common feature of the face for fast and easy detection. Therefore the same idea is applied to our object to be detected which is actually aquatic weeds and in this section we will be able to describe what are the necessary features of our object at which our newly designed algorithm will be able to learn the feature and detect the object from any image. Therefore consider the figure below which shows all the common and necessary features of the aquatic weeds at which our designed model will learn from them during the training.

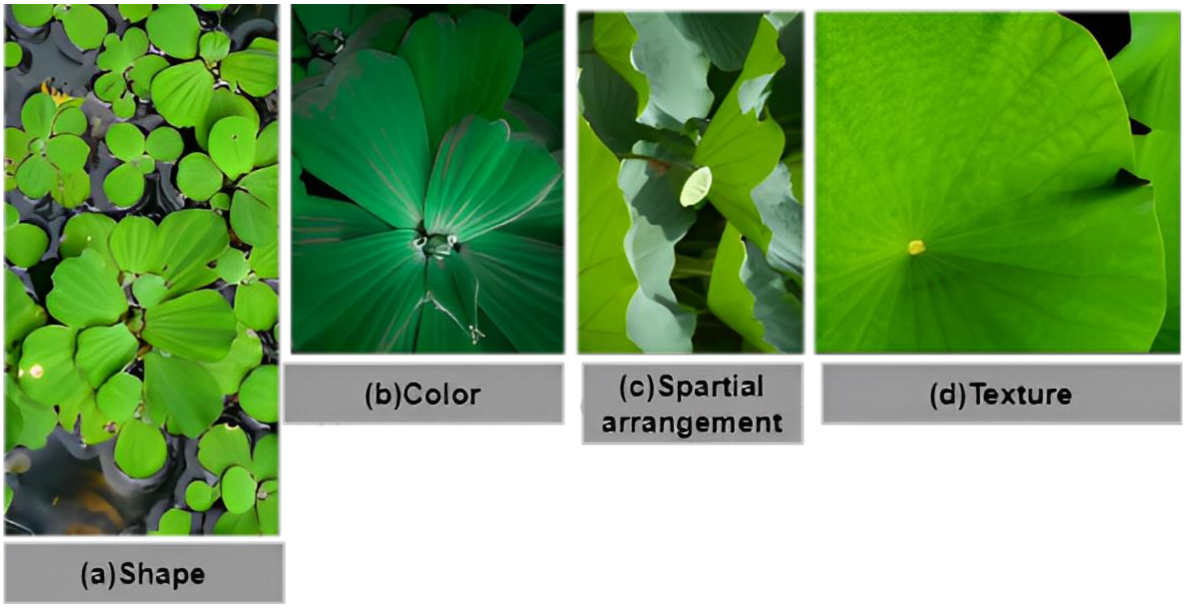


Figure 1-3 Features learned by our proposed network architecture

With a reference from the figure above this study will explain each feature after the other and state how the proposed network structure works on them. However the studded structures are not necessary for each deep learning algorithm to learn, these structural patterns are considered on the ground of our proposed network structure.

(a) **Shape**, the objective of this study is to detect aquatic weeds with any outstanding accuracy, and to differentiate between aquatic weeds and other plants species which are more useful in supporting aquatic life in aquatic environment. That means these plants are of much more important on balancing the ecosystem of the aquatic environment. Therefore we have considered shape of the aquatic leaf because it is of much more important since it provides its uniqueness from other plants leaf by defining it’s curvature, length and width of the aquatic weeds, Muhammad [43]. These features can mainly being used to differentiate aquatic weeds with other aquatic plants. Therefore our newly developed network model will consider much this feature for better results of our trained models. If you look figure 1-3 (a) you will be able to observe the shape of the aquatic weeds and also the shape of the leaf, can help us to determine the growth stage of the aquatic weeds for better elimination plan of the weeds. However identifying the most relevant and informative characteristics of plant leaves for deep learning task is of more important, because it contributes significantly to differentiating between different plant species and conditions, Zhao[55].

(b) **Color**, the color is also a very necessary feature for our network model this is due to the reason that, it provides necessary information on the color content of the aquatic weeds which will assist the algorithm to learn very quick and to differentiate aquatic weeds and other useful plants. However this feature seems to be very complicated, this is due to the reason that the green color is the common color for every healthily plants. But this will not affect the performance of our algorithm due to the fact that, the algorithm will not depend on this feature alone to detect aquatic weeds, Muhammad [43]. But it will perform out detection as a results of the combined features. Therefore no matter how much the color of the aquatic weeds resembles other plants species but it will not affect the performance of our designed network model, figure 1-3 (b) shows the true and correct color of the aquatic weeds at which our network model will work on it. In addition, feature analysis in this study plays a vital role in this research, this is due to a reason that if global features of aquatic weeds are well processed this will reduce dimensions of the input data, Wu [56]. This will actually focus on most valuable features to improve the efficiency and performance of the subsequence deep learning models.

(c) **Spatial arrangement**, when you look at figure number 1-3(c) above you will be able to see how aquatic weeds are arranged over each other. Therefore the spatial arrangement features defines the core of the arrangement of the aquatic weeds on the surface of water which has been studded on this study. It appears to be very different with other aquatic plants species, which will then provide a good learning access to the developed network model to learn and detect a aquatic weeds in a short period of time. Therefore in this case then the spatial arrangement feature is the only feature in this study which describes the arrangement and orientation of the aquatic plants, which in turn become very important in recognizing and differentiating aquatic weeds from other aquatic plants. In other way feature analysis helps to identify subsets of features which tracks the most significant variation of data, Jeon [57], this will automatically reduce computation cost and tend to speed up model training.

(d) **Texture**, our designed network will learn minor feature of the texture feature and these features are shown on figure 1-3(d), the texture feature normally provide the information like smoothness, roughness, or hairiness, Saleem [54]. As it is shown on the figure (d) that means it provides the useful information at which the algorithm learns to give out the best results. Although the smoothness and roughness of the aquatic weeds depends on the growth stage of the aquatic weeds and the climate condition at which the particular aquatic weeds grows. But this doesn’t affect the performance of the developed algorithm at all .Also the texture of the aquatic weeds defines the veins arrangement of the aquatic weeds that at least makes a difference from other aquatic plants. Also feature analysis is useful in supporting generalization during training for recognition and detection of the plants leaves with a very higher accuracy. Even when presented with variations in lighting condition, scales and orientation, Ali[58]. Furthermore a successful feature analysis of an object for detection or identification tasks may lead to the improvement of model performance, interpretability and computation effectiveness, Kumar[59].

Therefore the features described above are the result of the feature analysis we performed to find out the common features of the aquatic weeds that our proposed model will act on. This study concludes that the above features are very useful and are very uniform for all aquatic weeds, but due to the extensive analysis performed in this study, we were able to find that some of these features may vary depending on climate change and growth stage of the aquatic weeds. But these are only minor challenges to our proposed model as the proposed model will not focus on a single feature but will encompass all features. There are several reasons that made us to consider few features of our object to be detected. These reasons are as follows development and creation of the most stable and accurate model, focusing on the core features most relevant to each aquatic weed simplifies the model. It saves time for training and fastens the score due to few features and finally it accelerates variance reduction and prevents over fitting of the model.

1.3 Brief analysis of domestic and foreign aquatic weeds control

Aquatic weeds have a large impact on natural ecosystems, aquatic weeds are responsible for causing biodiversity loss at a very large extent, Khanna[17]. Aquatic weeds detection, eradication and monitoring plays a great role on minimizing these impacts, Baider[18]. For the area of aquatic weeds detection foreign countries have been doing a lot of researches on aquatic weeds elimination, the major research institution on aquatic weeds are Biological Control Institute, Chinese Academy of Agricultural Sciences, Management of Aquatic weeds and mosquitoes in Santee Cooper Hydrological Systems South Carolina in United State of America and Agricultural Victoria research in Australia. All these research institutions conducted different researches in a better way to accelerate the elimination of the aquatic weeds. Based on the technologies such as deep learning algorithms and machine vision recognition, all of which are cutting-edge emerging application technology.

However the deployment of the equipment is expensive, and still a certain problem during the implementation stage. A serious problem on managing aquatic weeds is a surveillance to find the location and how much aquatic spread on the water surface, so that we may have a better strategies on controlling the growth of the aquatic weeds, Kathryan [19]. In foreign countries aquatic weeds research is well-initiated due to the presence of various invasive and nuisance plants species. A big number of studies focused on developing effective detection and management strategies. Some of the technologies used in the detection of aquatic weeds are like DNA bar coding and molecular technology, this has applied to identify and differentiate various aquatic plants which leads to the early detection and a quick control of the aquatic weeds, Bertacch[20]. Remote sensing and aerial surveys, this has been used to map and to detect aquatic weeds infestations in the water surfaces, Cavalli[21].

Domestic research institute including Sokoine University-SUA of agriculture, Marine

Science Development Eastern Africa, Institute of Marine Science and Tanzania Fisheries Research Institute and Fresh water research Institute .Mally[22] has pointed out the use of the traditional ways of controlling aquatic weeds and the combination of these methods which resulted to the elimination of aquatic weeds by 70% within a period of three years. Aquatic weeds can disrupt the natural balance of ecosystem by outcompeting native species. Reducing biodiversity and altering habitat structures. Tungo[23] ,has also talked on the selection of a better way of usage of tradition method on controlling aquatic weeds. It is not pointed out that which ways they use to differentiate aquatic weeds with other aquatic plants, which are useful in supporting aquatic life. This study is going to apply modeling and predictive method which will be utilized in assessing and predicting the presence of aquatic weeds on the surface of water body under different environment condition which will assist at large the development of good strategies of fighting aquatic weeds.

1.4 The main research content of this paper

This paper mainly studies the challenges that we normally face when eliminating aquatic weeds on different water bodies. It has been a very big challenge in the whole process of fighting against aquatic weeds, many countries in the world have put strategies in place to fight aquatic weeds. Taking into account the serious effect of aquatic weeds, this has given as a great thirst to find the only way we can use to simplify the fight against aquatic weeds, apart from the scarcity of existing data on aquatic weeds. We have studded the work of many researchers around the world, they have used various technological methods in identifying aquatic weeds .Researches done by the previous researchers on the aquatic weeds brought a very positive results to most of the models that were found out through the experimental work were all accurate. After having made a comprehensive analysis on the results and methodologies used we had found out that there is a big gap between the exponential growth of the aquatic weeds and the accuracy of the model, but we had found out that all studies were suffering out due to luck of data set to train their models, both works had used the same deep learning algorithms to train their models on the same data set, the algorithms used were like VGG16[51], Inceptionv3[37], DenseNet121[52], Xception[53] and ResNet[48].

In view of the scarcity of the dataset, this paper will collect the first aquatic weeds data by itself. The collected dataset will contain data of aquatic weeds. Due to the lack of data set, this paper proposes an approach that will be used to increase the volume of our dataset, and the process is being done at the preprocessing stage. The preprocessing stage has been divided into two parts, these are data argumentation and data normalization. The dataset will be increased by performing argumentation process on it and then the approach proposes data normalization on the same dataset, where by all the images as the result of the argumentation process will all be normalized in the range of 0-1. The reason why we have proposed data normalization during the preprocessing stage is to enhance the fastest convergence of the neural networks during the training, but before performing image normalization and augmentation, we had to perform image de-blur due to the quality of some images. .

Also our study proposes an end to end aquatic weed detection model by improving YOLOv4[14] feature extractor where by a network structure of YOLOv4[14] is optimized by adopting CSPDarknet-53 to CSPDarknet-28 as a backbone feature extraction network at which, the number of the convolution layers in the residue network is reduced from 53 to 28. Also the down sampling times of the existing YOLOv4 [14] are altered. However the analysis and labeling work has application value for the research work involving aquatic weeds data. Since this paper is centered on the accurate detection of the aquatic weeds as far as we know, there is still a gap in the research on the performance of the deep learning algorithms on detection tasks. This paper will compare and explore the performance of different deep learning models on aquatic weeds detection and all selected deep learning algorithm will be considered as the baseline models before we compare its accuracy with our proposed model. Our study will also focus on the space required by different models and training time of each model including our proposed model.

1.5 Organization of this thesis

Subsequent chapters of this paper are organized as follows the first chapter, first introduces the research background and the research significance of the topic. At the same time analyzes and discusses the current research status and explains the existing problems. The second chapter introduces the relevant theories and technologies in the identification, detection and classification in line with our proposed model, including general information on convolution neural networks, pooling layers and non-linearity layers, deep learning algorithms architectures like YOLOv3[15], YOLOv4[14] and Faster R-CNN[10], data preprocessing proposed approaches like data augmentation, data normalization and image de-blur.

The third chapter describes the overall analysis and design of our proposed approach and our proposed model and also this chapter goes even further by describing the labeling process of our data set used in this paper. The fourth chapter describes the implementation, application of the proposed model on the labeled dataset, experimental results of both baseline models and the proposed model, results analysis and result tabulation .Chapter five, the last chapter describes the main problem or challenges of our work in the aspect of validity and this chapter goes further by describing internal and external validity. Also this paper on chapter five proposes future works toward the field of computer vision.

Chapter 2 Related theory and technology

2.1 Introduction

Computer vision technology is now becoming an eye toward the growth of technology in different sectors around the world including agricultural, education and water management. For example, plant disease detection, species identification, weeds detection, water and soil conservation can be managed and controlled by Computer vision technology. Basing on the technology we have seen that there is an easy way to control aquatic weeds by recognizing them and locate their position, this will actually help us to differentiate weeds with other useful plants in different water sources. Therefore in our work we have considered a number of researches that have been done by different researchers on aquatic weeds detection. In this chapter we are going to read and analyze comprehensively on different study which will at least act as a road map toward our research findings.

2.2 General information on Neural Networks

In this chapter we are going to review in more detail the design of the Convolution Neural Network. Recently Convolution Neural Network (CNN) have been paid much attention especially in researches that are conducted in the area of computer science. Convolution Neural Network have been proven to be more effective in the areas like image recognition and image classification, Zhipeng [44]. In this case then a Convolution Neural Network can be referred to be a feed-forward artificial neural network which assumes that the inputs are images which allows us to encode a certain properties or behavior into architecture. Normally Convolution Neural Networks are made of a set of blocks that can be applied in various signal sources like images, audio and video. Therefore each block transforms the input to the output of the neuron activation which will act us the input to the next block. Consider the figure below that illustrate in more detail what we have discussed.

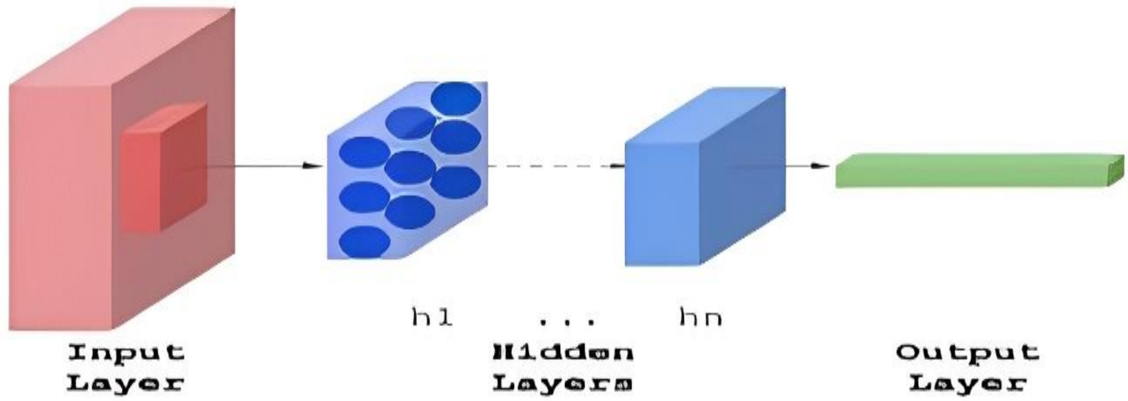


Figure 2-1 Layers of CNN that transforms inputs to the outputs of another layer

By considering figure number 2-1 above let us assume an input volumes of the image is represented by $m * m * r$, where by m represent width or height and r represent number of color channels and a Convolution layer has k kernels of size $n * n * q$, where by q can be smaller than or equal to r. There after non linearity is performed on each feature map between the convolution and pooling layers, in order to optimize gradient back propagation is applied to calculate the weight parameter of each layer, SqueezeNet [45]. Convolution Neural Network is a core subject of our research therefore in the architecture review a very comprehensive study is done to make sure that the good concept of Convolution Neural Network is acquired. Also to make sure that is used during the testing of the object detector which will be discussed later, there is a huge demand to understand what the Convolution neural network is and what is component of it. And how do we use the concept of the Convolution Neural network to assist us to select a good deep learning model to develop our object detector. Although in our study we have considered some few deep leaning algorithms but still some researches are still going on to determine the most accurate algorithm and we have just selected only three algorithms to validate our output. The improvement of the different learning algorithms tends to reduce labeling scarcity especially on large data set, ShaffleNet[49], because one of the tough way is manual labeling of the large data set[43], normally it consumes a lot of energy because you need to label one image after the other for the project of large data set and it is too hard, so far and it consumes a lot of time.

2.3 Pooling layer and non-linearity layer

2.3.1 Pooling layer

A pooling layer is a layer added after a convolution layer and it is added after the application of non-linearity to the feature map [35]. Normally when designing a convolution neural network, we aim at getting the output that has the efficiency and accuracy near to 99% and this can only be possible if we design convolution neural network, model without any limitation during the process of object recognition or object tracking. Normally layers in the model have its own modality of arrangement for example input image, convolution layer and non-linearity, therefore the limitation of this arrangement is only maps the precise position of the feature in the input. This will be affected by a slight change of the position of the image to be detected automatically. It will lead to different feature map, therefore by having this challenge in object detection, then the idea of using pooling layers came into practical. This layer was placed after non-Linearity layer aiming at reducing the amount of computation performed at the network, this is possible because the dimension of the feature map has been reduced. Also pooling layer is computed in each layer separately during the process therefore a small change that can happen will not affect the performance [9] .

Normally pooling layers have been divided into major two types or functions and each group performs its own task if it is applied to the network. The types of pooling layers are as follows one is called**,** Average pooling, normally it calculates average value of feature map in the network and the second one is called, Maximum pooling. It calculates the maximum value of feature map in the network. From all the three types maximum pooling is the most used pooling layer compared to average pooling. Consider the figure 2-2 below.

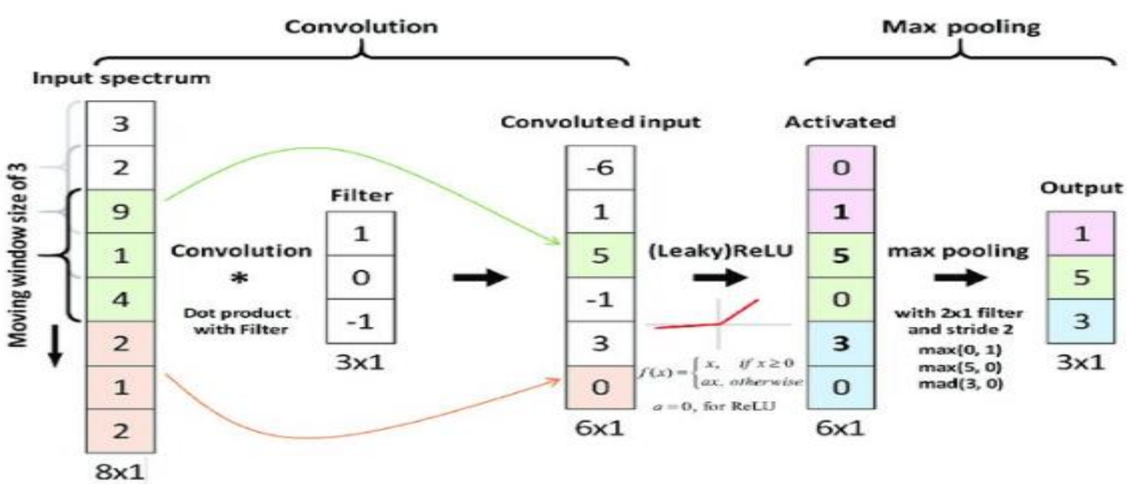


Figure 2-2 A schematic diagram of convolution and max pooling layer

2.3.2 Non-Linearity layer

Non-linearity layer is a convolution neural network that depends on the convolutional layer which generates a feature map which later creates activation map as a sample output. Since it consists of activation function that is used to execute difficult task like when a large data set is fed to the algorithm for training, this is done by triggering the activation function to use a less number of dataset, Perez[61]. Having non linearity function is very important in a neural network unless we would be having a linear output no matter how many layers do we have in the network. Let l be a non-linearity layer and it takes the feature capacity of Y^{l-1} from the convolution layer which is given as $l-1$ which later will generate the activation capacity ofY^{l}, HernándeZ-garcía [62]. Therefore the activation function is computed by the following

$$Y^{l} = f(Y^{l-1}) \tag{2-1}$$

In the formula l —— is defined as non-linearity layer

Y^{l} ——activation capacity

Y^{l-1} ——Feature capacity

$l-1$——defined as convolution layer

When dealing with hidden layers you must be aware of which type of activation function you can apply to your model. Normally there are three types of activation function as mentioned as follows one Rectified linear activation (ReLU), second Logistic (sigmoid) and third Hyperbolic Tangent.

2.4 Deep learning algorithms

2.4.1 YOLOv3

Before we get into much detail about YOLOv3[15] algorithm we should better get to understand what is YOLO?. The term YOLO is defined as “You Only Look Once” which was then initiated by Joseph Redmond [63] on the paper that was entitled as You Only Look Once: Unified, Real-Time Object detection. The main objective of this study was to improve the efficiency of object detection which by then happened to be a very big challenge. The efficiency of deep Learning algorithm which was used to detect an object with a very good accuracy was R-CNN[9] and faster R-CNN[10], which used to take a long time to detect an object. But YOLO algorithm was able to process 45 frames per second [63]. This algorithm is one stage detector which increases the performance in detecting different objects. YOLO[63] algorithm outperformed other existing algorithms. As researchers continue working on different possibilities on how they can increase the speed of Convolution Neural Network models to detect an object in real-time. Therefore some versions of YOLO were introduced and this led to the improvement of YOLO algorithm which resulted to YOLOv3[15].

YOLOv3 [15] (You Only Look Once, Version3) is a real-time object detection algorithm that recognizes different objects in videos and images. YOLO is a Convolution Neural Network (CNNs) which is used to detect real-time object as it was discussed earlier on the usage of CNNs, therefore YOLOv3 uses the concepts and principles of Convolution Neural Networks. Where by the input image that contains object to be detected with defined features based on the score the predictions will depend on the predefined classes. From predicting classes of objects YOLO [63], detects the location of these objects on the image and normally YOLO applies a single neural network to the entire image and these neural networks subdivide image into small sections called regions and predicts out the probability of each region. There now probability becomes a key attribute of selecting the correct object to be detected, therefore prediction relies on the number of bounding boxes that have been convoluted which means that have been covering some regions on the image and from there now the algorithm select the best one basing on capacity of the probability. Let us look at the performance, in this algorithm the image is divided into a $S * S$ grid cell as it is shown on figure number 2-3 during the process, if it happens that the center of the object falls on the grid cell then that particular grid cell is responsible for detection, Adnan [64]. Someone may ask how it is possible for the grid cell to detect an object. The answer is since the bounding box and the confidence score are predicted by the grid cell, therefore it is very possible for the grid cell to predict the object. In this algorithm normally the bonding box consists of x, y, w, h these are defined as the major components of the object to be detected that means the x, y represent the center of the box of particular grid cell and w, h normally represents length and width which depends much on the image.

Normally the predictions are encoded or represented by the formula below where by C represents class probability and $S * S$ represent a small box for each cell that predicts$B$, Therefore by considering all these components the predictions will be encoded as.

$$S*S*(B*5+C) \tag{2-2}$$

From the formula B ——object to be predicted

C ——Class probability

$S \times S$ ——bounding boxes

Consider the figure below that illustrate the way how the grid cells have been divided on a single image and the formation of the bounding box.

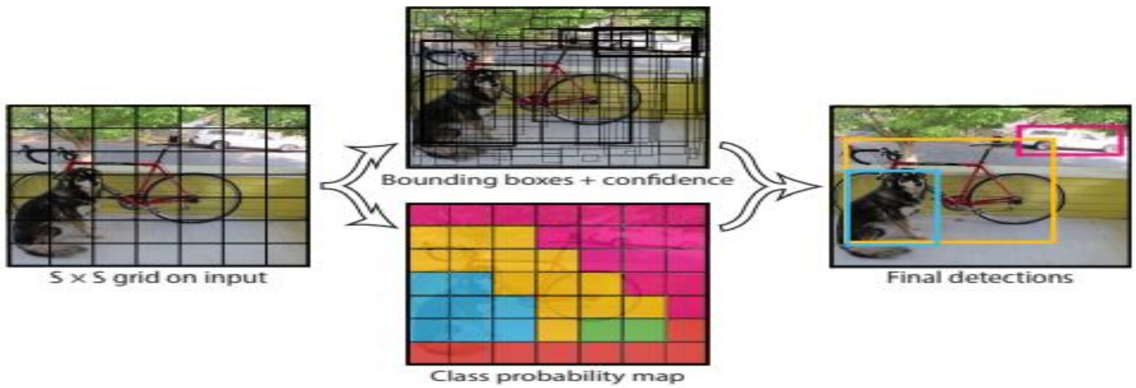


Figure 2-3 S x S grid input and the formation of bounding box

In YOLOv3 [15], some improvements were made just to increase its speed and accuracy. Pooling layers are not considered in this algorithm instead convolution layer were added as a result, it prevented the loss of low-level features which causes the improvement of ability of a model to detect small objects, therefore allows the researchers to use the image of any size when applying YOLOv3[15] in their research. Also YOLOv3[15] can predict abjectness score for each bounding box by using logistic regression and lastly YOLOv3[15], can perform multilabel classification for the object to be detected on an image, Eugene[65]. But YOLOv3[15] has a very worst performance in detecting large datasets.

(1) **Architecture of YOLOv3** YOLOv3[15] like any other deep learning algorithms uses convolution layer and this version of YOLO normally is made with Darkenet-53, convolution layers and residual layer. But it is mainly for the detection purpose, therefore if we try to consider the full architecture, we get 53 more layers added and this form a total of 106 layers of architecture, among 106 layers normally the detection takes place at three layers only 82, 94 and 106. Consider the figure below.

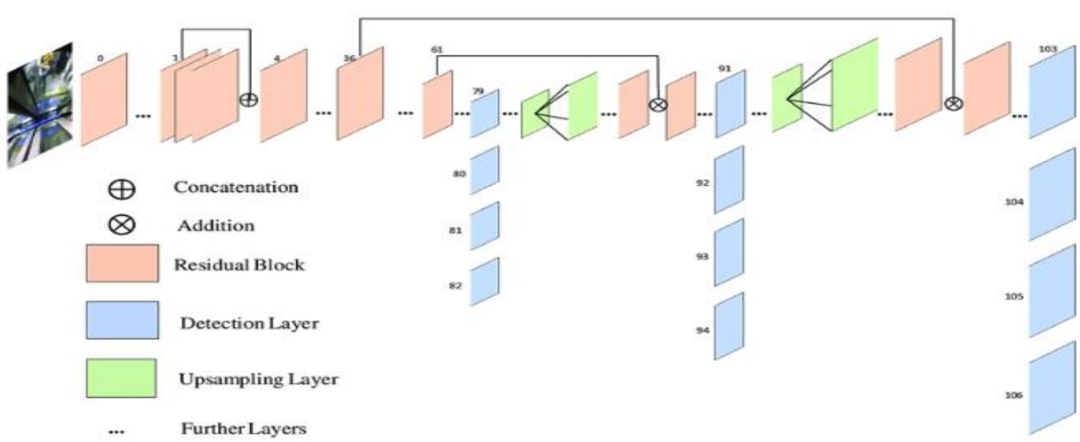


Figure 2-4 Architecture of YOLOv3

Feature extraction in YOLOv3 is governed by the network which is the hybrid of a network used in YOLOv2, Darknet-19 and new residual network. Consider figure 2-4 above it looks like detection takes place three times in YOLOv3 as it is shown above. The shape of the detection kernel $1\ x\ 1\ x\ (B\ x\ (5 + \boldsymbol{C}))$, B is a number of the bonding boxes a cell on the feature map can predict.

2.4.2 YOLOv4

YOLOv4 is a real-time object detector of just like any other YOLO algorithm but is much faster compared to the previous versions. This method has been introduced recently by Alexey Bochkovskiy[14], which was entitled YOLOv4: Optimal Speed and Accuracy of Object Detection. The main goal of this algorithm was to make a fast object detector with high quality of accuracy [14], that means the designed object can be easily trained and used for the betterment of the community. The accuracy of the YOLOv4 has been improved to be very fast to the extent that it can execute 65 frame per second which is very reasonable speed. In this algorithm new feature were added like WRC, CSP, CmBN, SAT, Mish activation, Mosaic data augmentation, CmBN, DropBlock regularization, and CIoU loss [14]. The ability of YOLOv4 to recognize multiple objects in real-time is much fast compared to other algorithms, the increase in accuracy of YOLOv4 was possible due to the improvement of the following key attributes of the algorithm like the improvement of classification performance, this is very important because the performance of some of the algorithms were very much affected by the volume of data, that means many of the algorithm were effective on operating on the huge amount of the dataset. The increase of quantity of data within the feature map is also one of the improved component of the YOLOv4, since the feature map in artificial neural networks is the one which administer the output of the input image, therefore in YOLOv4 the amount of data to be processed by the feature map has been increased and this leads to the increase of the accuracy of the model. Also in YOLOv4 deeper layers were added to replace the function of the pooling layer, Sahla [66].

In YOLOv4 two options of real-time neural network were proposed for GPU a small group of layers, in which is defined within the range of one to eight in convolution neural network layer and CSPResNeXt50 and CSPDarknet53 were used. The second option for VPU in this option the group of the convolution neural network was used. Conclusively the performance of YOLOv4 compared to YOLOv3 in terms of the efficiency, performance, average precision, Frames per second, in this case if we consider the rate of detection in frame per second has been increased by 10% and the Average Precision has been increase by 12%. This proves mathematically that the YOLOv4 has high performance compared to YOLOv3. YOLOv4 typically comprises various components the first component which is very important component is the input which is actually the image you intend to detect the next component is the backbone and neck which is very necessary to be considered as the important component, because this component is the one which takes the image as the input and extract the feature map using a deep neural network. Finally we have to make and handle these predictions using an object. Therefore in order to improve the accuracy of the YOLOv4 they used YOLOv3 versions at the end of the chain of the object detector. In YOLOv4 a lot of factors have been considered to improve its performance for example they introduced the new method for data augmentation. By considering how important data augmentation is in object detection? In this version of YOLO, they have used mosaic and self-adversarial training and finally it was tested with different features [39] like multiple classifiers, multiple detectors, multiple backbones and other parameters.

Therefore from the above explanations the components of YOLOv4 like deep layers and successful method of image augmentation are applied to this version. We had a very good reason of using YOLOv4 to our research, this is due to the fact that the accuracy, speed and performance of YOLOv4 happened to be very important to our research. Because our model detector needs to track the location of the aquatic weeds within a short period of time, that will save a time to eliminate the aquatic weeds before the massive infestation of aquatic weeds in a particular area. In fact our preliminary results had shown among all deep learning algorithm we selected its only YOLOv4 had a better result compared to others.

(1) **Architecture of YOLOv4**, the YOLOv4 architecture is made of CSPDarknet53 as a backbone, spatial pyramid pooling additional module, PANet path-aggregation neck and YOLOv3 head, Ferdinand [67]. Consider the figure below which Illustrate the architecture of the YOLOv4.

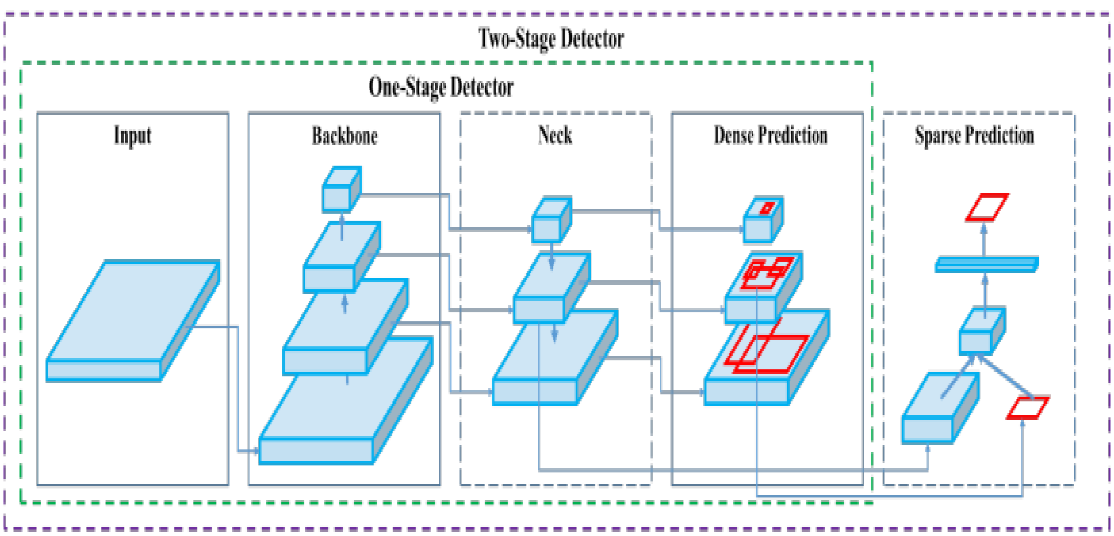


Figure 2-5 Architecture of YOLOv4

In YOLOv3 Darknet-53 was used and it was successful increased the efficiency of the model but it was a bit slow, but in YOLOv4 again Darknet53 was also used but in addition with CSP and gives CSPDarknet53. From the figure number 2-5 we have seen the dense block normally this contains multiple convolution layers each layer is attributed for example $r1, r2, r3 \ldots\ldots rn$ instead of using the output from the last layer then rn uses the output from all previous layers. Now CSPDenseNet the most important idea here is to separate the base layer into the very two parts this will reduce competition complexity to the last layer, finally the purpose of the neck from architecture above is to add extra layer in the backbone and the head that as the prediction part.

2.5 Chapter Summary

This chapter gives the background of deep leaning models for object detection tasks especially in the part of the aquatic weeds detection methods, firstly it gives the general introduction to its high-level research field that is an introduction to convolution neural networks. The chapter introduces the concept of deep learning models, their performance and implementation .The chapter is further explains the whole process of detecting an object and mathematical representation of the bounding box until the object is detected by a certain deep learning algorithms. The background knowledge given on this chapter will help readers to open their minds as faster as possible for a quick understanding of the content of subsequent chapters of this study.

Chapter 3 Collection and Processing of Aquatic Weeds Datasets

3.1 Introduction

There are a lot of research based on plant leaf detection and classification and there are already many publically available datasets which are frequently used into different researches. Some of these datasets are available on data sources like Data.gov, google.com, ncdc.noaa.gov, datasetsearch.research. We even read some research work on aquatic weeds like Deep Weeds: A Multiclass Weed Species Image Dataset for Deep Learning [35] and found that it is helpful enough to assist us to conduct our research work. Due to the existence of these public datasets, related research works can easily and directly verify the effect of the proposed methods on these datasets. Different research work can use these datasets as benchmark to compare and evaluate experimental results. As far as we know this research work is the first work on aquatic weeds detection. The very first problem encountered in this research is luck of publically available datasets on aquatic weeds which are different from other plants related datasets. It is very more challenging to collect aquatic weeds datasets, this is due to a reason that, one it is very important to reach the exactly location where aquatic weeds are allocated which becomes a little bit harder to reach if it is during the rain seasons. For example in Tanzania at Lake Tanganyika (6.2556°S,29.5108°E) and Lake Victoria (0.7558°S,33.4384°E) which are located at Western and Lake zone of Tanzania respectively, it becomes much harder to collect datasets in these areas due to the scattering of aquatic weeds. Therefore you need to navigate around the water body to find out the areas which are affected by the aquatic weeds. However the need of the aquatic weeds datasets is of very much important, therefore to make our research very colorful we had to use some other techniques to increase the volume of our datasets.

Also this brought out the challenge due to the images collected, some other images do not belong to the aquatic weeds, and they belong to some other aquatic plants .Therefore in order to carry out the research work of this paper, this paper decides to make the first datasets for aquatic weeds. In addition prepared datasets will be used as a core requirement to complete the research. These datasets can also be used for the related research work which will use this kind of datasets to carry out their research activities until they display their own findings. Basing on the same datasets for computer vision and related field. Whether it is plant leaf detection or aquatic weeds detection, after the collection of the datasets tools are required to label and visualize datasets depending on the specification of images of our datasets and some images requires multiple labeling due to the appearance of more than one aquatic weeds on the image. Due to luck of unlabeled datasets, this paper has prepared a well labeled aquatic weeds dataset which will be very convenient for subsequent research work. After the completion of the data annotation, therefore our datasets will be well used by our detective models. Normally for the deep learning algorithms to work best there should be a series of techniques that are to be applied to the original datasets such as the removal of unwanted images, the application of the data argumentation to increase the volume of our datasets, and data normalization. This chapter discusses the important information and all the process involved in the collection of datasets, datasets preprocessing techniques like data normalization and data argumentation and it will describe in detail all the necessary operations required to be done on datasets preparation.

This research prefers to prepare its own datasets due to the reason that, prepared datasets (customized datasets) can be used to closely match the specific requirements and characteristics of the tasks at hand. This includes data with relevant features that relies to the geographical condition in Tanzania, diverse variations and specific scenario that are more representative of the target application. This will actually improves the model’s ability to learn and generalize effectively for aquatic weeds detection. Moreover ,we needed our own datasets to ensure the data are collected from reliable sources ,to adhere specific quality and standards and to make sure that the necessary features of aquatic weeds are captured for easy detection task after model training.

3.2 Collection of the aquatic weeds datasets

A dataset has been created by collecting images of aquatic weeds in various angles constructs and brightness. In our work we used Canon EOS R5 which has 45 megapixel resolution, 8-stop image stabilizer, built-in WIFI and Bluetooth. This device is the best device and we trusted it due to its quality. The collection of these images consisted of aquatic weeds from different sources of water bodies alongside with other figures in different water bodies. A total images of 1098 were collected among which 300 images were collected from Lake Tanganyika (6.2556°S,29.5108°E) which is located on western part of Tanzania in Africa and 588 images were collected from Lake Victoria (0.7558°S, 33.4384°E) which is located on Lake zone in Tanzania and 210 weeds were collected from black river (1913’44”S, 14637’45”E), weeds from black river [35] in 2019 to sum up a total of 1098 images.

Table 3-1 Number of datasets collected from different water bodies

	Location	Number of images collected	Total images from all sources
Source 1	Lake Tanganyika	300	
Source 2	Lake Victoria	588	1098
Source 3	Black River	210	

Due to the environment condition and weather during the data collection we had a limited scope, which resulted to the collection of the limited number of images. Since the rule of thumb in deep learning is to have a minimum of 1000 images per class in the dataset, therefore in this case we applied data augmentation method. Our study identifies several types of image augmentation like image flipping, panning, zooming and rotation. Images in our datasets have been preprocessed by performing augmentation by rotating them at angles of 90°, 180° and 270°, but also the augmentation process was performed on the same set of images, which increased the number of our datasets by the multiple factor of five which in turn give the total number of images in our dataset as 5490. From 5490 images a total of 1500 images are from Lake Tanganyika (6.2556°S, 29.5108°E) and 2940 are from lake Victoria (0.7558°S, 33.4384°E) all are found from a Eastern part of Africa and the last 1050 were taken from black river which is located from (1913’44”S, 14637’45”E) .The figure below shows the distribution of aquatic weeds.

Figure 3-1 Different images of aquatic weeds

In respect to this a complete set of data is cleaned to ensure that there is no any unrelated image that does not relate with aquatic weeds. However the size of the images will depend much on the version of the deep learning model. Images will be resized by using external software before being fed to the network for training. Since YOLOv3 [15] and YOLOv4 [14] both will be used as our baseline models before we train our prepared datasets in our proposed model, therefore in this case our images will be resized to 416*416.

3.3 Preprocessing of Aquatic Weeds Datasets

As it has been discussed earlier that aquatic weed is a very big challenge around the world, this paper aims at accelerating the process of reducing aquatic weeds by detecting and localizing them. Since we are now looking for strong enough model which will be used to detect aquatic weeds with a very good accuracy, therefore we have decided to come out with a method during preprocessing stage that will actually increase the performance and accuracy of our model. In this chapter we are going to discuss the proposed method which is known as data preprocessing, so operations like data augmentation and data normalization have been defined as stages of data preprocessing. Having this idea we had the hard task on looking for the dataset which will actually stand as a driving force toward the goal of this research work. Focusing on the method that has been proposed to get a more accurate results, our method will improve the accuracy of our model and at the end of the day we will be able to detect aquatic weeds in a very good accuracy as it is shown on the figure below (Figure 3-2), the figure below shows the image source or image bank, then the sample of images which is actually our dataset, these images are taken to the next step which is dataset augmentation. This stage describes it into three parts loading image file this part all the images are put in the same file for loading, the loaded file must be zipped to reduce the size of the loaded file the next part is the extraction and the augmentation process, that means at this step all images are extracted from the zipped file and then the augmentation process takes place.

Finally storing all the augmented images that means a default file is created to store the augmented images, the next step which is very important step is image normalization on this stage we pass the image to confirm the range of the pixel if it is 0-1, if it is not 0-1, then we convert the pixel range into 0-1. In this study we have decided to use these two techniques to make our data useful for the better and accurate results, we put too much pressure on the process of data normalization on our dataset, this is due to the fact that during the training process the convergence of the neural networks becomes too faster due to the reduction of the pixel range. This is because it helps to reduce the impact of the differences in the image pixel values, which can cause the model to take longer to converge. If the numbers of the pixels is too high, that means the difference is too high, therefore the speed of the network convergence will be too slow. However one of the most challenging factor that affects deep learning algorithms is lighting condition, this is due to the reason that images in the dataset can have different lighting condition which can impact pixel values of the images. The application of the image normalization to the dataset reduces lighting impacts on the images within the dataset, as a result it leads to the improvement of the accuracy of the particular object detector.

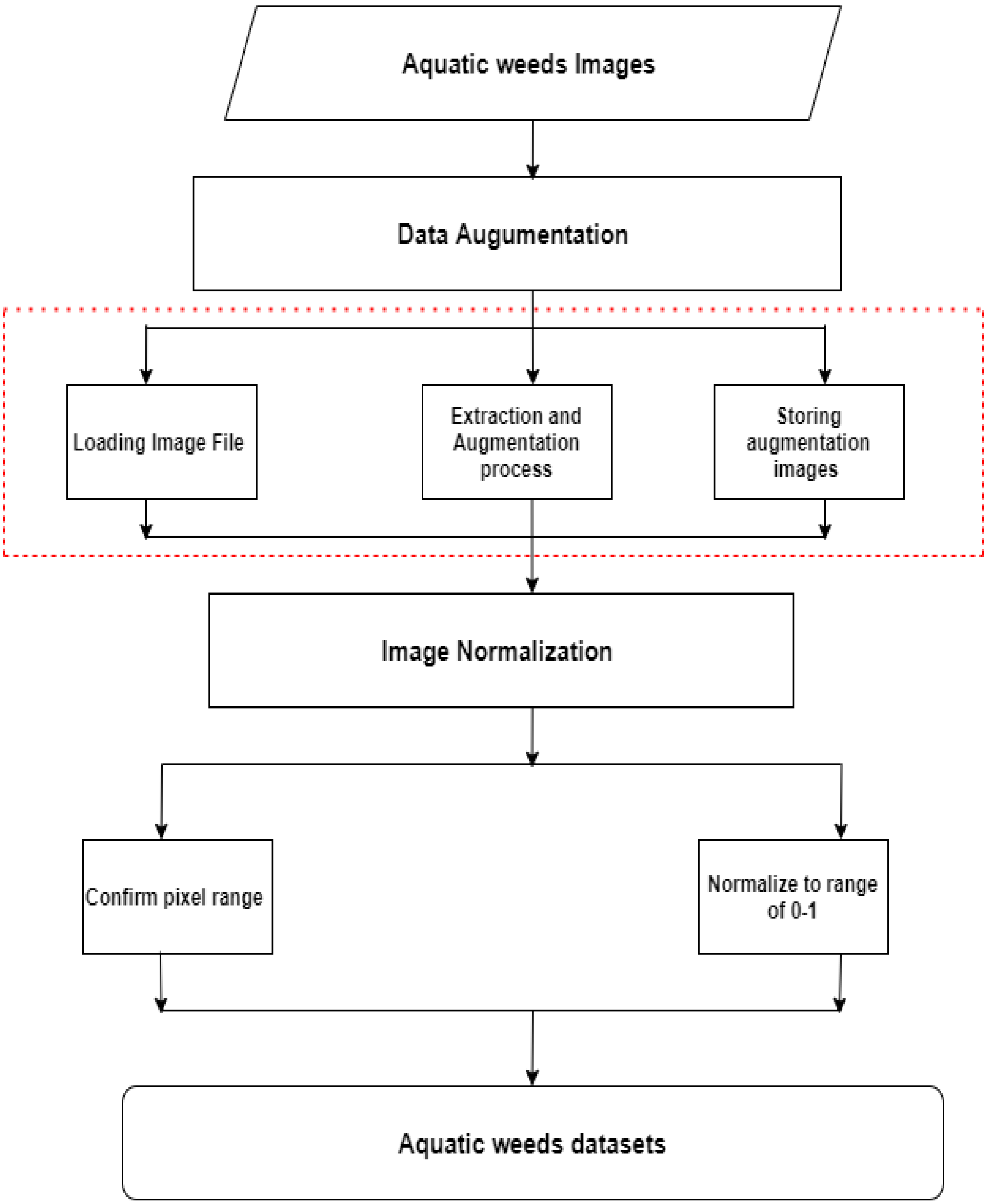


Figure 3-2 Data preprocessing method

3.3.1 Improvement of the quality of aquatic weeds images

Image blur can affect the accuracy of deep learning algorithms which may lead to a poor results after training. Blurred images is a result of unclear and distortion of features. This is very challenging for deep learning models to learn and extract meaningful features which can lead to the decrease of detection accuracy. However deep learning model generally require a large amount of training data (quality) to learn the features of an object, if the training dataset contain the blurry images, it can make the training process more difficult, Rusha [60]. Consider the figure below that shows the architecture of image de-blur to improve the quality of our datasets.

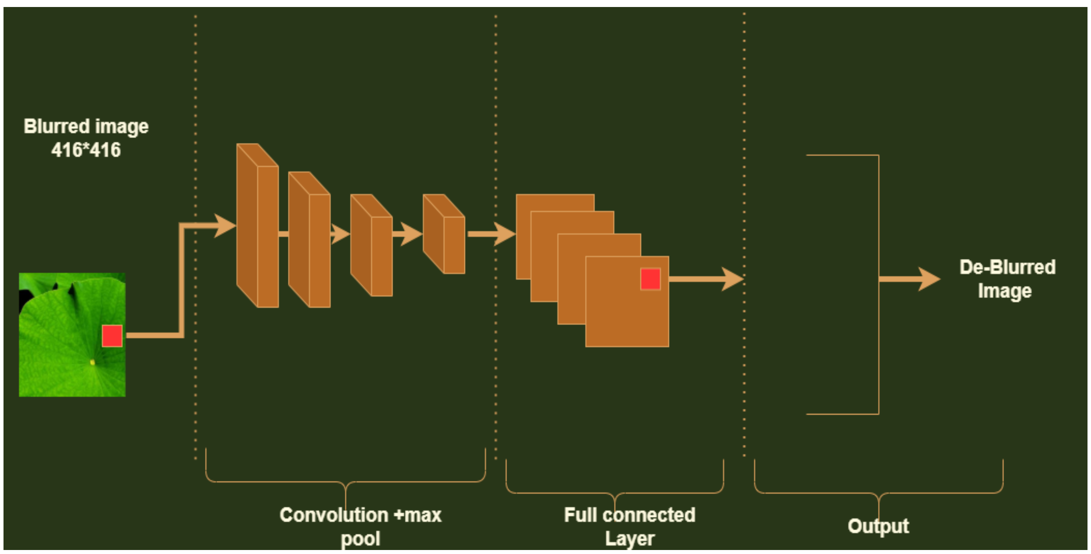


Figure 3-3 Model architecture for aquatic weeds image de-blur

Convolution neural networks, normally have powerful representation capabilities and can extract very abstract feature representation. CNN, normally uses the convolution kernel in each layer to perform convolution operation on the input from previous layer and after this layer goes to the convolution operation, multiple feature map will be generated as the input to the next layer as shown on the figure 3-3 above. The implementation of this model we have used four convolution layers at which the size of dimension convolution kernel is set to 3*3 and each convolution layer is followed by ReLU activation function and connect a pooling layer. Then after last pooling layer, data is smoothed and then it is sequentially input to the fully connected layer and output layer. Experiment setup the proposed model was performed on NVIDIA GeForce GTX 8th series graphics cards and the I7 CPU and 16GB RAM, in this experiment we are going to have 8000 iterations. In this model the training process, considers feeding the blurry images in the network and parameters are optimized to minimize the difference between the de-blurred images and corresponding ground truth images (sharp). After the training of de-blurred model, the model is used to de-blur new images .This model recovers the distorted details of the image and reduces the blurry effects. This will improve the accuracy of our model since the images will be very sharp when are fed to the model.

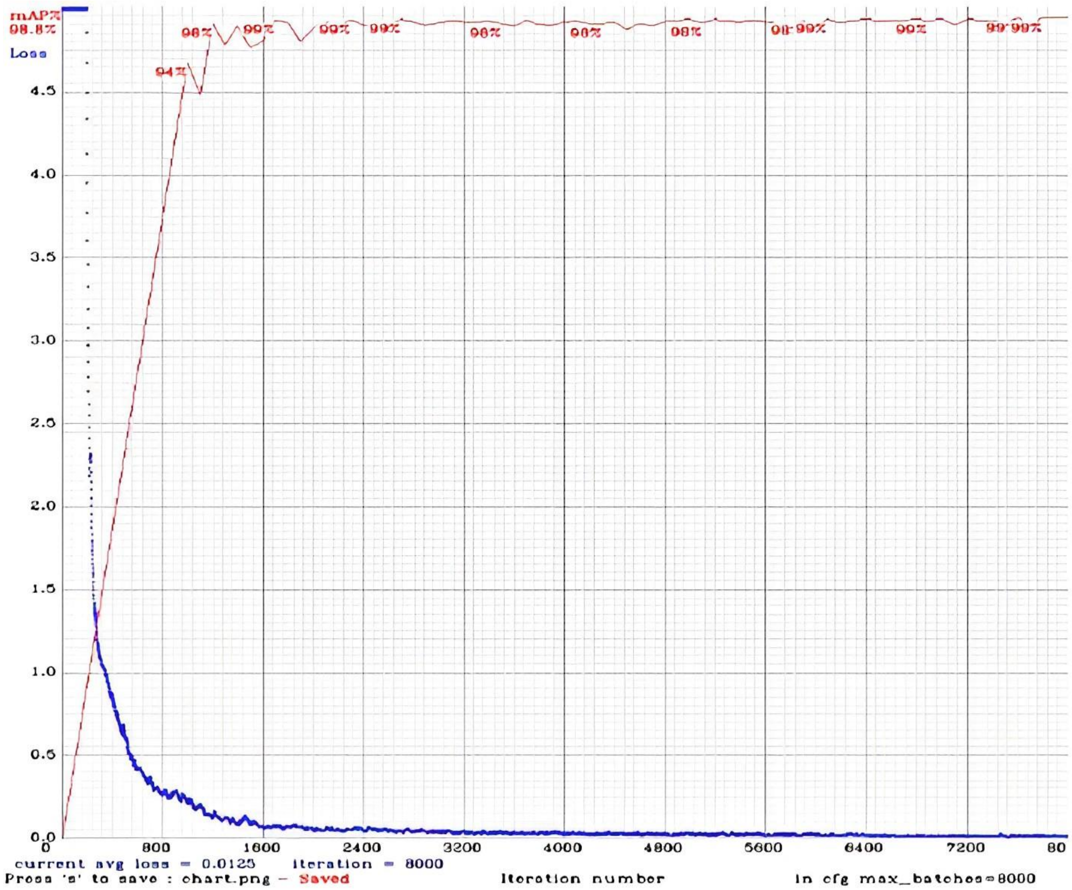


Figure 3-4 Loss function curve of the proposed image de-blur model.

(a) Sample image before applying de-blur model.

(b) Sample image after applying de-blur model.

(a) Sample image before applying de-blur model.

(b) Sample image after applying de-blur model.

Figure 3-5 Improvement of the quality of the images after applying image de-blur model

3.3.2 Implementation of image augmentation

In developing deep learning models or any machine leaning models we need data and normally the big the size of the dataset the better the accuracy of the model, therefore this study has proposed augmentation technique as the way forward to increase the amount of our dataset. Data augmentation refers to the technique of generating a newly modified data to increase the size of the dataset, Mahmudul [68]. Shortage of data in developing models is a very big problem that needs to be fixed before developing any new models. In this paper we have used this technique as one of the means of increasing the capacity of our dataset, therefore data augmentation has been divided into three parts as it is shown on the figure Figure 3-2. This is the first stage of data preprocessing according to our proposed method, this part has been divided into three major parts like Loading image file, Extraction, augmentation process and storing augmented images.

During the process of image augmentation our method will first identifies and loads a zipped file that contains original images and in our case we have 1098 images which were loaded. After loading image file, the next step is extraction of the image that means before we perform image augmentation we should unzip the image file, extracting the images from the loaded file. After extraction of the images then we perform image augmentation process by rotating images at the angle of 90°, 180° and 270°. But also by performing flipping, shifting and zooming at a range of 0.05 as it is shown on the table (3-2) below, that result to the creation of the temporal file called argufile which contains 5490 the augmented image of the original image. Therefore by doing so we have added the number of images to our training set. This part of our proposed method has increased the size of our dataset by 5 times the original size of our dataset, at this point we can say we have a number of images at which we can train our model. Consider the figure 3-6 below that shows the increase of the images during the process of augmentation, the bar chart below, gives us a clear picture on how the number of images can be added to the dataset. This is due to the fact that we cannot show all the transformed images of our original images on this paper, but the graphical representation gives us the over view on how the images kept on increasing after performing image augmentation. Image augmentation in this study happened to be of more important, because to get accurate model we need enough data, therefore through this technique we were able to increase the size of our data set 5 times the original set.

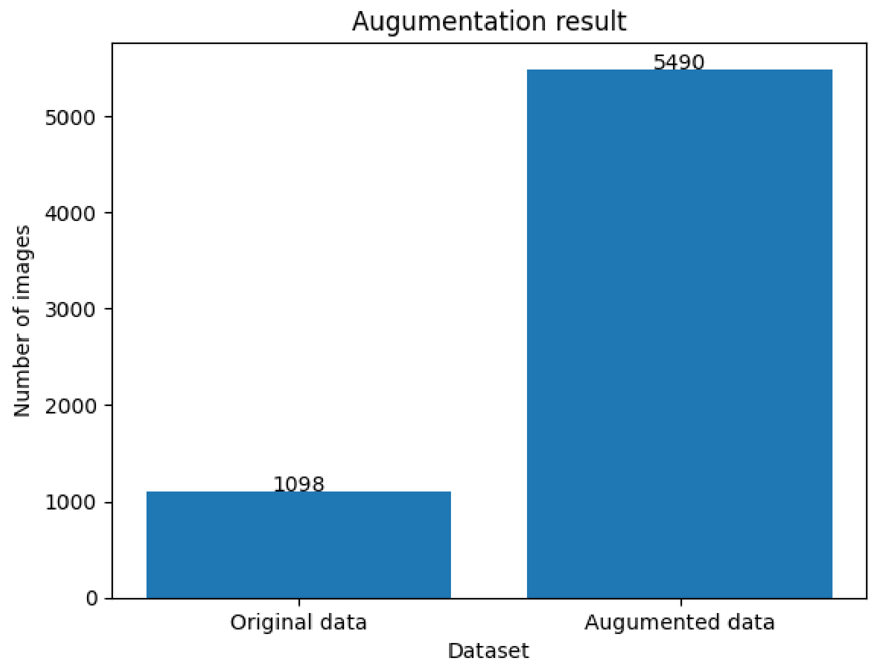


Figure 3-6 Presentation of augmented images

This paper has shown how powerful and important image augmentation is in preprocessing of dataset. This is due to the reason that any good deep learning model with a best accuracy is the one that validation errors keep on decreasing with training error. Furthermore data augmentation is a very strong technique in overcoming the problem of overfitting which has become a very big problem during the development of the deep learning model. This is due to the reason that most of features of the images are being derived from the original images or from the original dataset, Yang-Yang [69]. Proposed method has set our augment range in the following order rotation range, width shift range, and height shift range, shear range and zoom range. Consider the following table below, which actually shows the summary of the different techniques that has been used or applied to perform image augmentation for the purpose of increasing the volume of our datasets, aiming at having a better result when training the model. This is due to the reason that, the more the data the higher the accuracy, and our focus is to increase the accuracy of our models.

Table 3-2 Value range of different augmentation parameters

	Rotation	Width shift	Height shift	Shear	Zoom
Value range	90°	0.05	0.05	0.05	0.05
	180°	0.05	0.05	0.05	0.05
	270°	0.05	0.05	0.05	0.05

3.3.3 Implementation of image normalization

One of the biggest problems in the whole process of developing a deep learning model is the total time taken to train a model that means the training speed becomes slow than the actual time [64]. This paper has found out that one of the major reasons of why it takes long time to train a model is pixel range. Normally images are made of matrices with pixel range of 0 to 255. Therefore after knowing this challenge we have decided to deal with pixel range by reducing the pixel range from 0 to 1. Data normalization refers to the process at which each input augment has similar data distribution, this is due to the reason that data normalization allows a very faster convergence of the network during the training [65]. In this study we have proposed image normalization as one of the stages of our proposed method, in this stage we have divided it into two major parts as it is shown on the figure 3-2 above, the figure shows that the first part checks and confirm pixel range of the image if it is between 0 to 255 or 0 to 1. Therefore if the algorithm confirms the image in the range of 0 to 255 then our algorithm normalize the image to the range of 0 to 1.

In this case the logic behind this concept of normalization is to reduce the pixel range from 0 to 1 for faster convergence of networks during the training. Consider the formula below for further demonstration, Kavir [70].

$$I_N = (I - \mathrm{Min})\frac{\mathrm{newMax} - \mathrm{newMin}}{\mathrm{Max} - \mathrm{Min}} + \mathrm{newMin} \tag{3-1}$$

From the formula above our desired range is 0 to 1 that means our newMax is 1 and newMin is zero, since all our images from the dataset are having range from 0 to 255 that means our Max in this case is 255 and our Min is 0. Having given the range of our images from 0 to 255 the procedure subtracts 0 from each pixel that makes the range 0 to 255, therefore at this point all pixel independently are multiplied by $\frac{1}{255}$ which will yield the range of 0 to 1. This stage in our proposed method is very important because our desire is to normalize our dataset to be in the range of 0 to 1. But one important thing to note in this method when all images from our dataset are normalized no any information is lost or distorted from the image this will actually not affect our model during the testing, but it will actually boost up the accuracy of our model at large, that’s why we preferred this technique in our study. The process of normalization has been divided into two parts as mentioned earlier, these parts are as follows Pixel confirmation and Normalization in range of 0 to 1.

The algorithm will pass on the stage one where by it will check the number of pixels of the particular aquatic weeds image in the dataset, after checking it confirms the range of pixel of all images in our dataset. In this case all images in our dataset have been processed to the range of 0 to 255, then after confirming pixel range, then it compute the range of our pixel values by referring to the formula above 3-1. Therefore the pixel range will be computed directly to the range of 0 to 1 prior to training of the ready prepared dataset. Consider the figure below that shows the gray scale image that represents the original image of aquatic weeds.

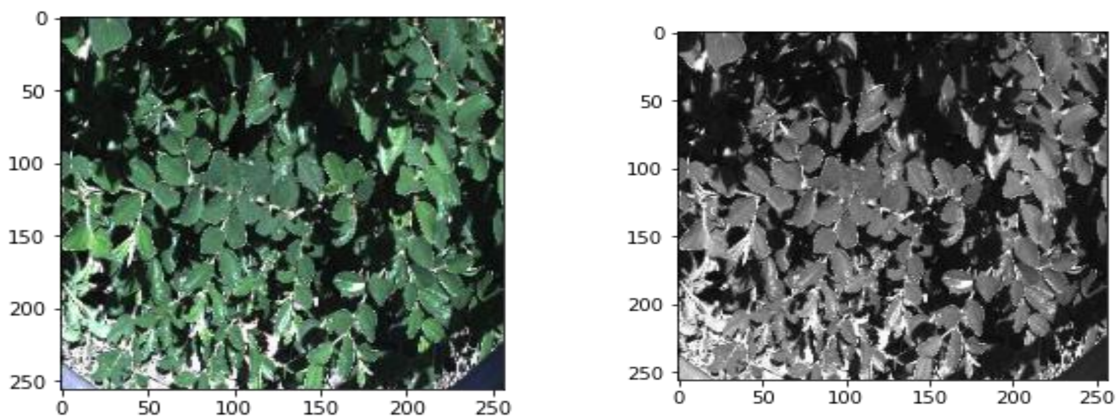


Figure 3-7 Gray image and normalized image of aquatic weed

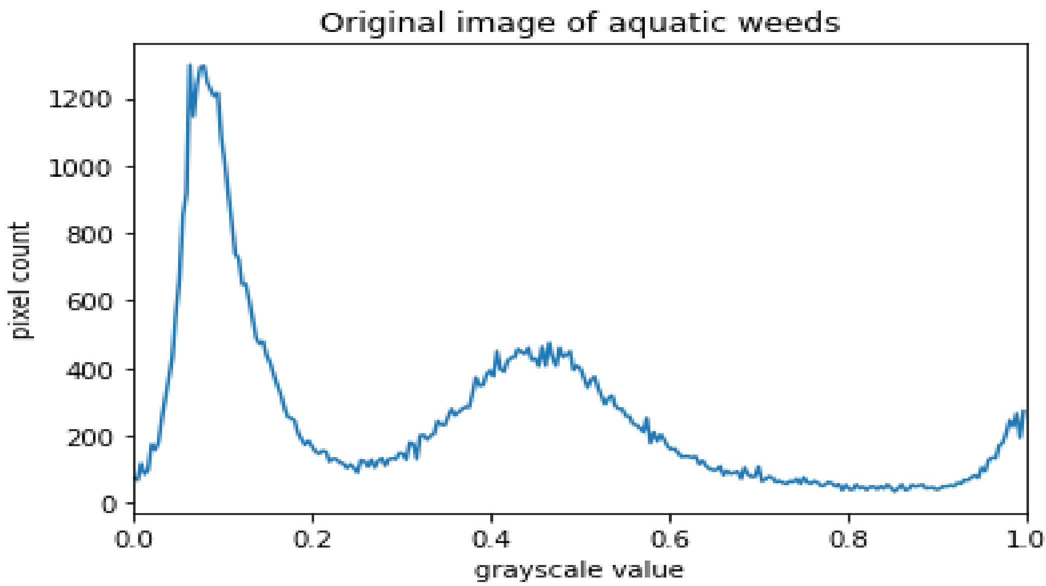


Figure 3-8: Histogram of gray image of aquatic weed

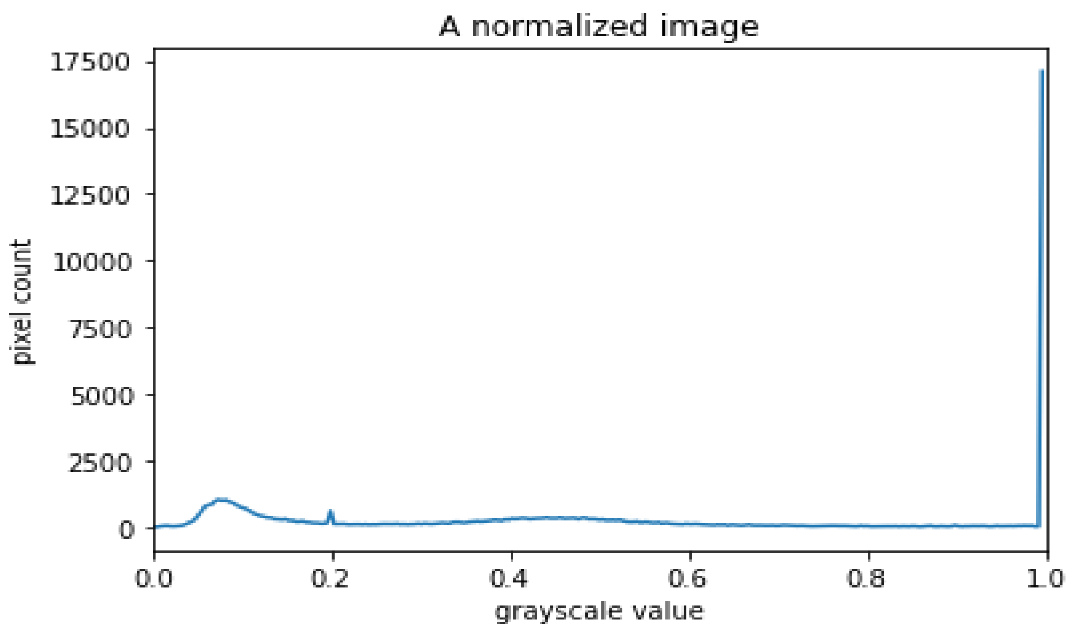


Figure 3-9 Histogram of normalized image of aquatic weed.

3.4 Arrangement, labeling tool and labeling process for aquatic weeds datasets

Normally the training of the model using any deep learning algorithm require data and from the beginning of this chapter, we have been discussing about how the dataset were obtained and what are the procedures used to increase the size of our dataset. Initially the total of 1098 images were taken from the selected water bodies. But the number of images obtained were not enough to train our model, this is the reason why we had included some techniques like image augmentation, to increase the size of our dataset the overall performance of the image augmentation is described on table 3-2. During the process of augmentation images were multiplied by factor of 5 which led to the total number of 5490 images. In our study the dataset were put mainly into major three groups the first group was train dataset, the second group was test dataset and the third group is validation dataset, in this case test and validation dataset comprises 15% each, while train dataset comprise 70% of the dataset.

As we train our dataset using deep learning algorithms which has been grouped into supervised type of learning [38]. All images from our dataset were labeled and the process took place manually, therefore we label all images for both train dataset and test dataset the work seemed to be tedious. But we manage to label them all. During labeling some information are of more important to have, labeling process was performed on LabelImg software, the following information are the ones which are normally needed during the training.

(a) The name of the image
(b) The coordinates of the image
(c) The class type of the image
(d) Location of the image

The process of labeling sometime is called as the process of creating image annotation, the annotation can be in two forms it can be .txt file for the images that are label in YOLO format or XML file for the Pascal format, Normally the annotation which is generated on the YOLO format has composed the following information, these information are the ones which will be used later to provide the bounding box information in the image at which a algorithm will use a trained model to detect an object on the image. The YOLO format annotation is made off the following, the class name which normally starts from “0” attribute and the coordinate, length and width of the bonding box. But when using the LabelImg tool you need to make some configuration to the predefined config file, normally consist of a number of predefined class, therefore you need to configure it and give the name of the object that you are about to detect. On creating data annotation it was a very heavy task because we had to annotate all the images manually using a labeling tool called LabelImg.

3.5 Chapter Summary

This chapter represents the collection of datasets included in this paper, the data were collected from three different sources. A total of 1098 images were collected among which 300 images were collected from Lake Tanganyika (6.2556°S, 29.5108°E) which is located on western part of Tanzania in Africa, 588 images were collected from Lake Victoria (0.7558°S, 33.4384°E) which is located on Lake zone in Tanzania and 210 weeds were collected from black river (1913’44”S, 14637’45”E) were collected by Alex [35]. Our datasets provides a data lake for analysis and comparison of very important and potential related research work. In addition to this ,in order to make our research very useful and successfully our study has prepared a fully datasets of aquatic weeds and a well labeled datasets ,the labeling operation is done in a very careful way, even though it was a tedious job, but we managed to label a total of 5490 images. However this chapter points out different techniques of data preprocessing stage like data augmentation process which increased the volume of our datasets from 1098 images to 5490 images. But also the technique of image normalization was applied to all images of our datasets to reduce the impact of differences in the image pixel values, which can facilitate the speed of network convergence and reduce the time taken to train the model. But also this study has proposed a CNN model to improve the quality of images in datasets by improving the clarity of the images and reduce the blur of the images.

Chapter 4 Research and comparison of aquatic weeds detection algorithms based on deep learning

4.1 Introduction

The importance and success of the deep learning models has been witnessed by a big number of researchers, due to its powerful features and its ability of giving a quick solution to many existing problems in the field of computer vision. Mostly on the application areas like speech recognition, natural language processing and causal reasoning at which deep learning models has expressed its usefulness and high capability on performing different tasks. Now days a lot of researches and different works on aquatic weeds detection has been demonstrated and several experiments has been conducted and achieved some good results. From these experiments, different approaches of deep learning were used as a driving wheel to achieve such kind of good results. In this case due to the availability of other models like ResNet-50[36] and Inceptionv-3 [70]. They were used for classification tasks and they achieved some good results. Our study will not directly apply these models, but we are going to make a comprehensive study on which deep learning models are useful for the aquatic weeds detection, basing on the nature of our datasets. This is due to the reason that when developing or selecting a deep learning model, there should be a serious consideration of computing power, time and memory space capabilities.

Also it brings to our attention that we should also think of computing power and memory requirements that are needed by the model to have a better accuracy. Due to the nature of our datasets it is of more important to explore the accuracy of our object detector in different kinds of deep learning algorithms [31]. Based on the datasets collected from different water bodies at which the weather condition is different from another and it is also necessary to explore different datasets on detection accuracy [66]. In this paper we have used YOLOv4 [14] and YOLOv3 [15] as baseline models to explore the accuracy of our model. In addition, this paper proposes an improved YOLOv4 [14] model to increase the accuracy of our model. However there is still a gap in scientific research on aquatic weeds detection, this paper as the first work on aquatic weeds detection using deep learning models, will conduct researches and comparison through a series of experiments. To fulfill the research gap in this field and to provide out reference values for subsequent related research work.

Inclusively, this chapter focuses on the following research questions (1) what is the suitable and efficient deep learning model for real time object detection and tracking of aquatic weeds. This is due to the reason that, since there are several deep learning models that can work better on object detection and recognition. Our research question has been constructed to identify the best deep learning algorithms for performing object detection of aquatic weeds on real time, by optimizing the architecture of the existing deep learning algorithms to have a very good and accurate result. In order to be used in future work in accelerating the process of aquatic weeds detection. The next question (2) how do we distinguish aquatic weeds and other aquatic plants which are very important in supporting aquatic life and aquatic eco-system. The second question is very basic due to the reason that in many water bodies there are living organisms whose life and habitats depend on the aquatic plants. Therefore there is a need of distinguishing, first the aquatic weeds and aquatic plants before further action of eliminating aquatic weeds are taken. This paper proposes an improved model for object detection of aquatic weeds but also this paper goes a mile by proposing a CNN architecture to implement the influence of the image recognition toward aquatic weeds. This will help us to answer the second question of our research work. This chapter demonstrate the above matters, describe the related work and corresponding experimental processes, analyzes and discuses experimental results. Finally in the summary part of this chapter a conclusion for all experimental work on aquatic weeds detection will be given. YOLOv4 [14] and Faster R-CNN [10] are both very popular on object detection and recognition algorithms. This study preferred to use YOLOV3 [15] and YOLOv4 [14] as a baseline model instead of any other deep learning algorithms, due to the following reason.

(a) **Speed**, YOLOv4 [14] has a very fast inference speed, it achieves a very good performance by framing object detection as a regression problem rather than working with multiple regions of an image like what is taking over on Faster R-CNN[10].

(b) **Simplicity**, this study preferred YOLOv4 [14] to be used as a baseline model due to its simple architecture compared to any other deep learning algorithm. Since it acts as a single stage detector which directly predicts bounding boxes and class probability to avoid the use of the separate region proposal network (RPN) used in Faster R-CNN [10].

(c) **Multi-scale detection**, YOLOv4 [14] is designed to detect object of different scales by using multiple layers with different receptive fields which allows YOLOv4 [14] to detect both small and large objects effectively on the image. These are the major reasons which made us to use YOLO as a baseline model rather than any other deep learning algorithm.

4.2 Comparison and research of different deep learning models on aquatic weeds detection tasks

4.2.1 Deep learning model used and its implementation in detail

YOLOv3 is the third version of YOLO state of art object detection algorithm, in short YOLOv3 is the improvement of the YOLOv2, and the improvement is done to increase the accuracy of the algorithm. The prediction of the object position and class score is done by a single neural network and this happens within a single iteration. In this case grid is the ground of YOLO [63] where by YOLO generates grid cells in terms of $S * S$ grids from the input image, our case is aquatic weeds image. Therefore from the grid cells the bounding boxes B are predicted and normally the bounding boxes are made of height, width and center which has been defined in terms of $x\ and\ y$. Therefore whichever generated bounding box has its own value called P, which later is used to predict the object. But in YOLOv3 [15] the bounding boxes are replaced by anchors which are used to eliminate gradient problems which normally occurring during the training [63].

In YOLOv3[15] feature extraction network normally extract features in terms of vectors, every generated grid cell predicts bounding boxes and every bounding box predicts vector P as it is shown on the formula below.

$$P = (t_x + t_y + t_w + t_h) + P_0 + (P_1 + P_2 + ... + ...P_n) \tag{4-1}$$

$$P_0 = P_r(\text{Object}) * \text{IOUpredtruth} \tag{4-2}$$

The $\mathrm{IOUpredtruth}$ on the equation number 4-2 represents the accuracy of the predicted object position, since the minimum probability for object prediction is set. Therefore at the end the non-maximum suppression is done to predict the desired results. The implementation of the algorithm explains the design of the different state of art and their approaches toward object detection, recognition and classification. Since our study proposes a deep learning model and suggested different approaches that can improve the accuracy of the object detection models to compare our results with a baseline algorithms at which uses its accuracy to compare with the accuracy of our own model.

YOLOv3 [15] (You Only Look Once Version 3), this is the object detection algorithms that came after its original algorithm called YOLO [63] introduced by Redmond and Ally Farhad. In the context of aquatic weed detection, we used YOLOv3 [15] as one of the deep learning algorithms. This is because YOLOv3 [15] networks process input images (aquatic weeds) as a structured array of data and recognize the patterns between them. In this situation, YOLOv3 [15] allows the model to learn the entire image to identify any potential wanted objects. Normally this is done based on "regions" which means that the regions with the highest scores are given priority according to the specified classifications. The regions with a high score in this scenario will be counted as positive detections with regard to class as well. Darknet-53 is used by YOLOv3 [15] and serves as backbone in this case then YOLOv3 [15] consists of 53 layers. It has been trained with ImageNet dataset [29]. Consider the table below that shows 53 convolution neural network layers of the YOLOv3 [32].

Table 4-1 Darknet-53

	Type	Filters	Size	Output
	Convolutional	32	3*3	256*256
	Convolutional	64	3*3 / 2	126*126
1*	Convolutional	32	1*1	
	Convolutional	64	3*3	
	Residual			126*126
	Convolutional	128	3*3 / 2	64*64
2*	Convolutional	64	1*1	
	Convolutional	128	3*3	
	Residual			64*64
	Convolutional	256	3*3 / 2	32*32
8*	Convolutional	128	1*1	
	Convolutional	256	3*3	
	Residual			32*32
	Convolutional	512		
8*	Convolutional	256	1*1	
	Convolutional	512	3*3	
	Residual			16*16
	Convolutional	1024	3*3 / 2	8*8
4*	Convolutional	512	1*1	
	Convolutional	1024	3*3	
	Residual			8*8
	Avgpool Connected Softmax		Global 1000	

YOLOv4 [14] as a deep learning algorithm has been introduced on the paper called "Optimal speed and accuracy for object detection" introduced by Alexey Bochkovskiy. The main goal of this detector is to make an object detector with high quality of accuracy. This algorithm is more accurate to the extent it can process 65 frames per second. YOLOv4 [14] is the version after YOLOv3 [15] and this version there are some complex techniques that has been used to improve its accuracy. Deeper layers were added to replace the function of the pooling layer to accelerate its efficiency, some other techniques which has been included in this algorithm are like Mish activation and Mosaic data augmentation which appeared to be of more important in improving the accuracy of YOLOv4 [14]. YOLOv4 [14] appeared to be of more accurate compared to YOLOv3 but it takes long time to train a model compared to YOLOv3 [15], this has been observed during the training of the aquatic weeds dataset. The experiment will be conducted on the next sub section to verify the accuracy of the model and later the results will be compared to show the most accurate algorithm on aquatic weeds detection. This algorithm has been more described on chapter 2, where the operation principal and the architecture has been described.

The YOLOv4[14] architecture is composed of CSPDarknet-53 as a backbone, spatial pyramid pooling additional module, PANet path-aggregation neck and YOLOv3[15] head and some other features like WRC, CSP, CmBN, SAT, Mish activation, Mosaic data augmentation, CmBN, DropBlock regularization, and CIoU loss. Some other features were even combined to improve its accuracy. As it is seen on the architecture below the algorithm has been divided into two stages the first one is two stages detector that comprises of one input, Backbone, Neck, Dence Prediction, sparse Prediction and for the one stage detector comprises of input, Backbone, Neck and Dense Prediction. Consider the figure below that depicts the architecture of YOLOv4.

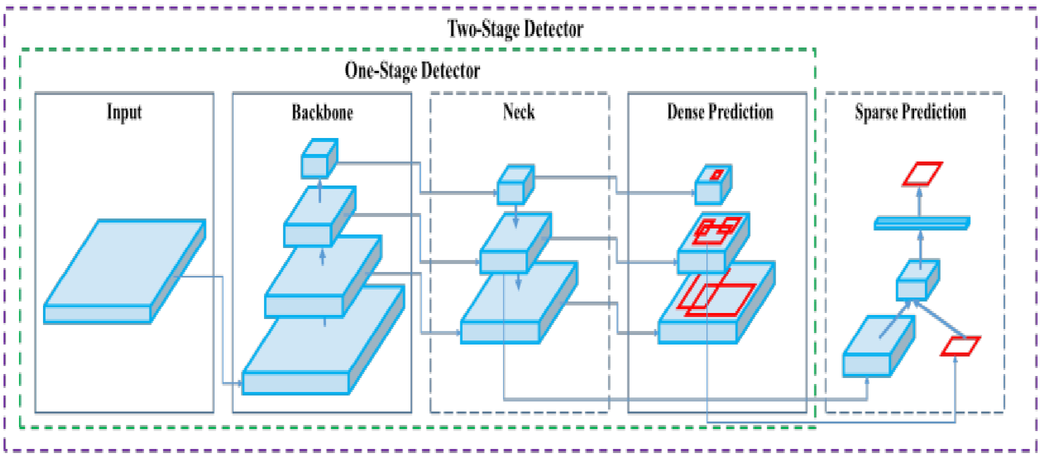


Figure 4-1 Architecture of YOLOv4

4.2.2 Comparison and analysis of the experimental results

These experiments, code and visualization of these models are run on colab with run time GPU accelerator of NVDIA tesla K80 with 8GB RAM, the colab provide installed python editor and ubuntu for model training. When comparing detection capabilities of a model. The most common used attributes is given by the formula indicated below (4-3) that defines the precision of a model, Shaun [71]. However, the true, true negative, false positive and false negative are normally used in detection problems, they can also be used for recognition problems. The number of true, true negative, false positive and false negative always depends much on the number of items in your datasets and how efficiently have you preprocessed your datasets before training. Therefore if the ratio of each attribute in the datasets is not balanced. For example if one set comprises large proportion, therefore the comparison will rely on only one base of accuracy. In this case it cannot real reflects the differences in each capabilities of each model. Take one example we have two groups (R,Q) and then in group R proportion is 96% and in group Q proportion is 4%, in this case if the accuracy is used as metrics, then one of the simple output group with whatever complexities in any process the input must be R. However we have also used F1-score as an addition measure to estimate the results in a very comprehensive way[29], therefore in order to apply F1-score we should really be aware of the formula, Sokolova [72] as it is indicated on the formula (4-4) and recall is also indicated on the formula (4-5).

Where by the precision rate defines the ratio of all samples that are predicted to be true and the true value is known as positive and the recall rate defines the proportion of all positive that are correctly predicted to be true.

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{4-3}$$

In the formula TP ——True Positives, (The number of positives, true predicted)

TN——True Negatives, (The number of true negative, false predicted)

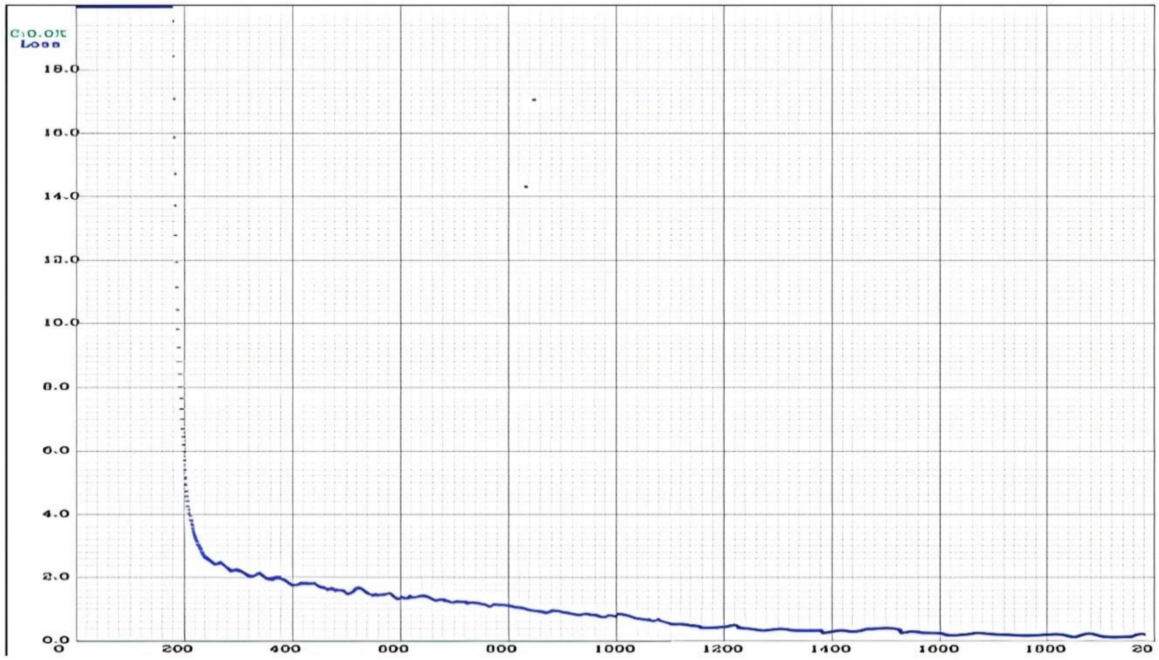
FP——True Negatives, (The number of true negative, predicted to be true)

FN——False Negatives, (The number of true negative predicted to be false)

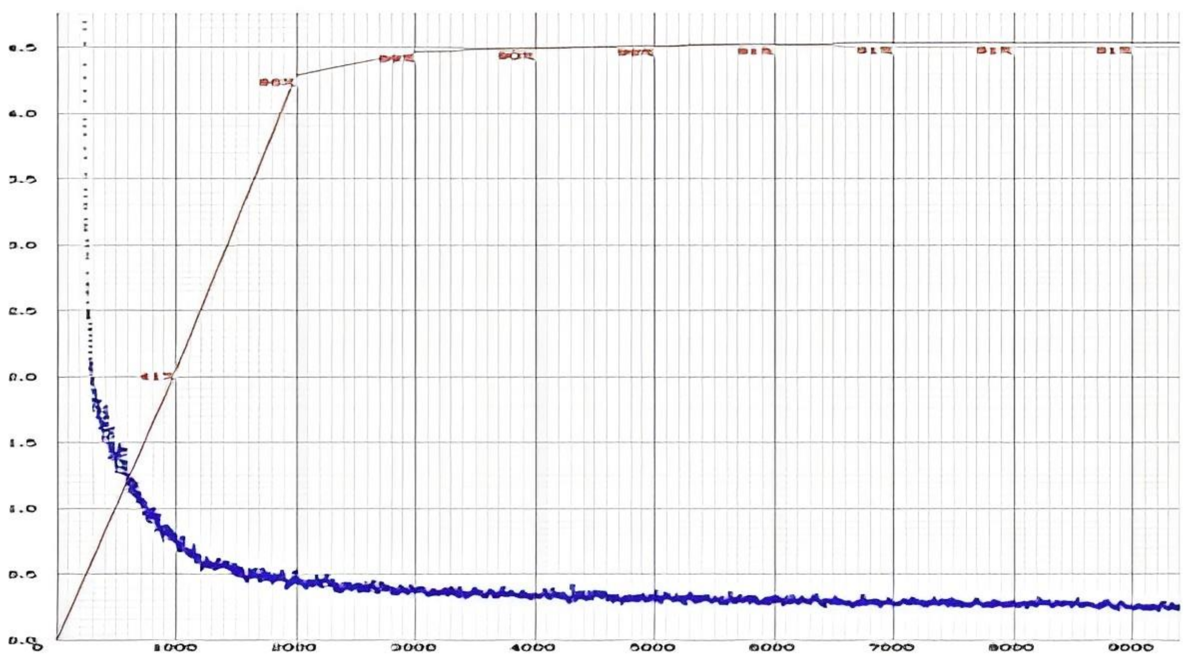
$$\text{Pr ecision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{4-4}$$

$$\text{Re call} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{4-5}$$

For every selected model experiment we have used all the images from our datasets that were obtained after performing different data preprocessing techniques like data augmentation, data normalization and image de-blur. Dataset is divided into two sets, one is training set and another one is test set that’s makes a ratio of 7:3. The proportion of the random division training number of iteration is 2000 and 10000 for both YOLOv3 [15] and YOLOv4 [14] respectively, the learning rate is set to 0.001, the image size fed to the network is 416*416 with 3 channels. The figure 4-2 below is the graphical representation of the model progress during the training process that shows the loss function of each model in respect to both YOLOv3 and YOLOv4 respectively. The visualization is made with loss and number of the training iterations. It can be observed from the curve of both YOLOv4 and YOLOv3. The YOLOv4 has a faster network convergence rate compared to YOLOv3 and in terms of training time YOLOV4 took us a very long time due to the big number of iterations defined during the training**.**



(a) Loss function curve of YOLOv3 model



(b) Loss function curve of YOLOv4 model

Figure 4-2 Each model loss function curve for both YOLOv3 and YOLOv4

Table 4-2 Result obtained after the application of our proposed method for both YOLOv3 and YOLOv4

So.no	Algorithm	Accuracy	Precision	Recall
01	YOLOv4	93.78	0.99	0.99
02	YOLOv3	89.93	0.7021	0.8067

This table above shows the accuracy of YOLOv3 and YOLOv4 as 93.78% and 89.93% respectively, if we make a scientific observation on these results we will find out that the YOLOv4 has outperformed the accuracy of the YOLOv3 by 3.85%. This marks the statement that the YOLOv4 is best model for the detection of the aquatic weeds. However during the training we have experienced some challenges that were encountered by YOLOv4, this study will conduct a lot of the experiments to improve the accuracy of YOLOv4, this is due to the reason that the accuracy of 93.78% is not good enough for the aquatic weeds detection.

On the table 4-2 that shows the training accuracy obtained from each model, it can be observed that, YOLOv4 achieved highest detection accuracy while YOLOv3 has achieved the lowest detection accuracy. In fact the detection results shows that for the aquatic weeds data when applied to CNN based models it improves the ability of the feature detection, by enabling the model to achieve higher accuracy than direct usage of the linear connection layers. YOLOv4 model has shown a very powerful means of extracting information of the aquatic weeds, which led to the improvement of the detection accuracy of the model. Based on this results we recommend that when using aquatic weeds datasets for detection and related tasks, as far as object detection is concerned, the use of CNN model is used as the baseline of the model design, so that you may have the powerful model for feature extraction of the aquatic weeds to achieve higher detection accuracy. Consider the figure below that shows the comparison of the model accuracy that we have applied to these experiments.

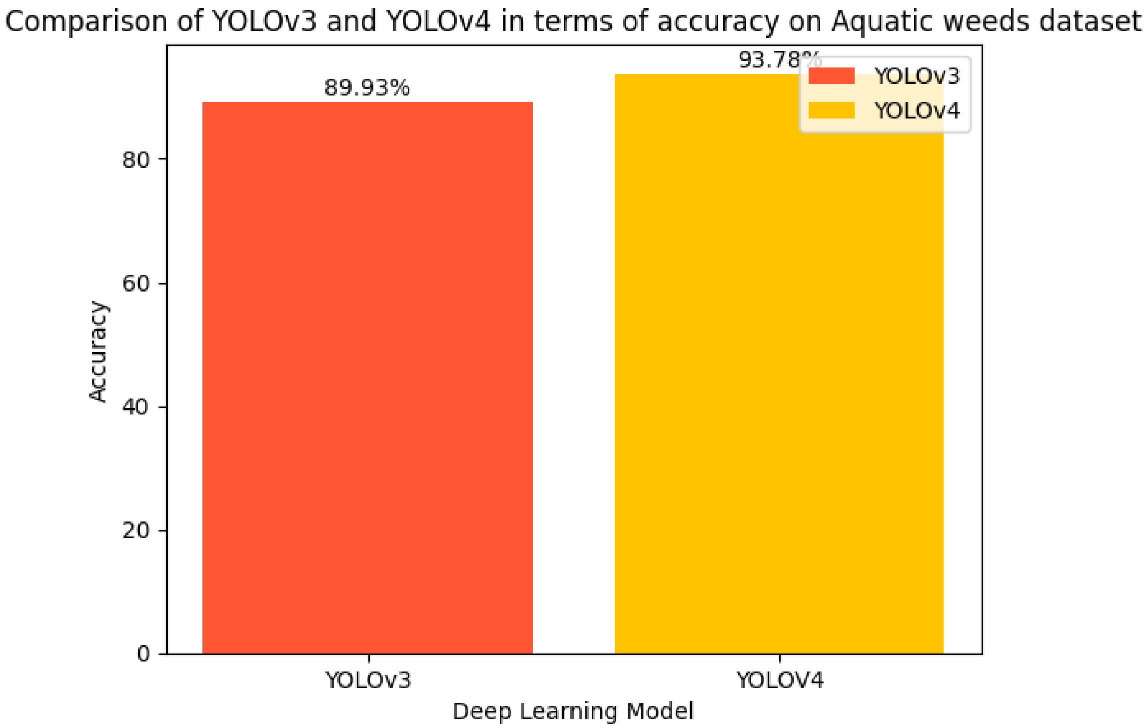


Figure 4-3 Comparison of the model accuracy for YOLOv3 and YOLOv4

However the YOLOv4 has achieved a very highest results compared to YOLOv3, the accuracy of 93.78% over 89.93 of YOLOv3 is not good enough, and therefore our study shows that the YOLOv4 is the effective model for aquatic weeds detection, compared to YOLOv3 as it is shown on the figure 4-3 above. Since the main objective in line with our research question is to find the most accurate deep learning model. That will be helpful enough in detecting aquatic weeds on the water bodies to increase the efficiency in eliminating aquatic weeds on the water surface. In this case our study has come up with improved YOLOv4 to make it better and more accurate in achieving a very best results.

4.3 Proposed model for aquatic weeds detection

4.3.1 Design and implementation of the proposed model

Model architecture is a very important component in designing deep learning models. Normally it describes the ways at which the neural network is designed in respect to the number of layers [69], the kind of layers, layer connection and relationship between the layers. A well designed architecture leads to a development of a model with a very high capability to learn with more efficiency and accuracy which brings up a better result. One of the most challenge of deep learning models is the slow network convergence which can significantly affect the performance and accuracy of the deep learning models. To overcome this challenge, we may do the following, one is to optimize the model architecture of the existing deep learning model, two to adjust the training rate and other hyper parameters. Therefore our study has observed this gap on YOLOv4 which has achieved an accuracy results of about 93.78% when we conducted experiments for the first time. We have come up with a lightweight architecture as a backbone of our proposed model. The main aim of this paper is to look for the most accurate deep learning model which will actually detect and localize aquatic weeds on the surface of water body. To accelerate the process of eliminating aquatic weeds from affected water bodies. All experiments that we have been done on both YOLOv3 and YOLOv4 respectively, we found out the accuracy for both YOLOv3 and YOLOv4 is 93.78% and 89.93% respectively.

However to detect aquatic weeds YOLOv4 achieved a very high accuracy and efficiently on detecting aquatic weeds but 93.78% is not good enough, due to the size of our object to be detected. We constructed a lightweight feature extractor (Backbone) by reducing a number of layers in the YOLOv4 network architecture, which appeared to be the most accurate deep learning model in our data sets during training. But the research focus is to provide a very good algorithms which will actually work best to provide very nice results. With this highest demand we optimized the backbone network architecture of YOLOv4 as it is shown on the figure number 4-5(a), the idea came out after doing a comprehensive study on the original architecture of the YOLOv4. Therefore on this subsection we will point out the design and the performance of our proposed model. In addition to this by considering the figure number 4-4 below which shows the work flow summary of our model. The preprocessed images or normalized images will be fed in the backbone network of our proposed model and important features will be extracted from the backbone of our model and then several operations will be taking place until the target object which in our case is aquatic weeds are detected .The architecture of the improved YOLOv4 will be demonstrated later on this chapter.

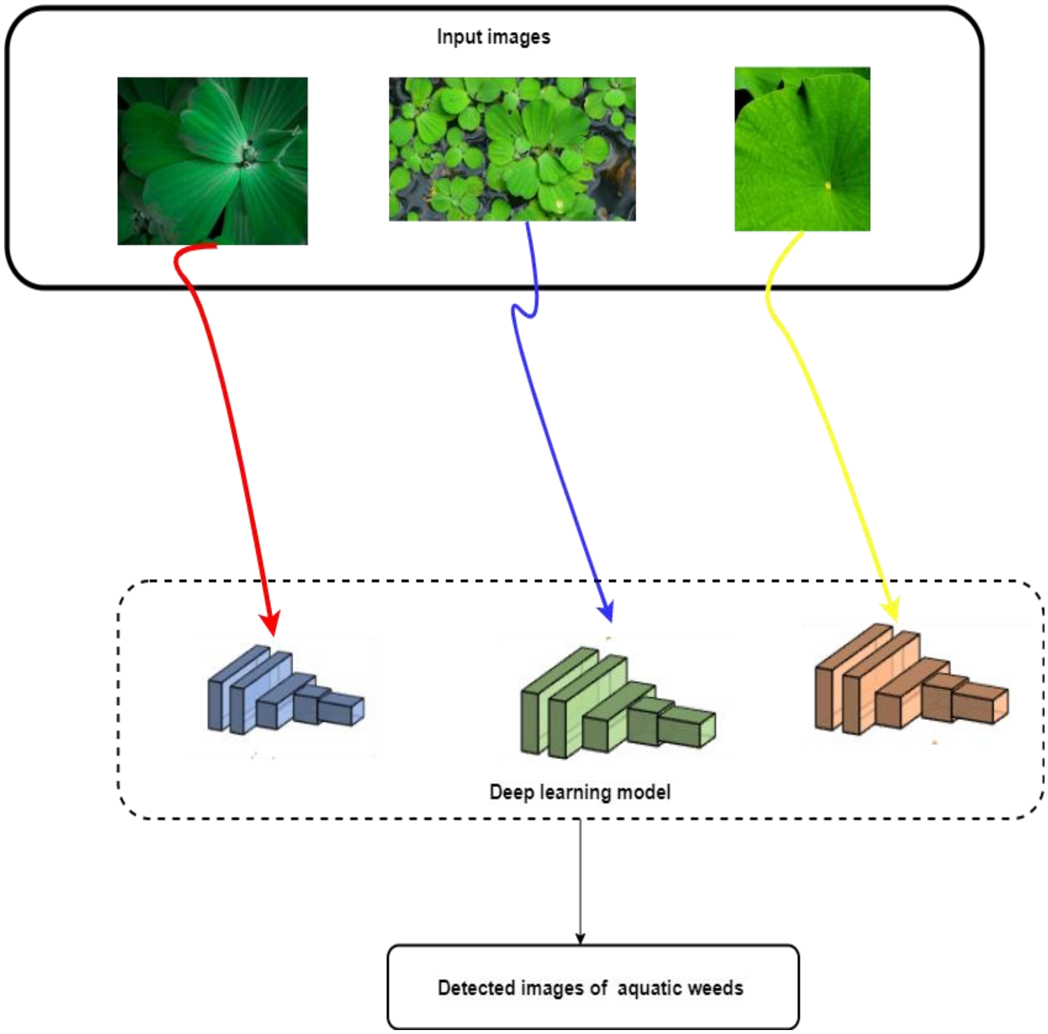
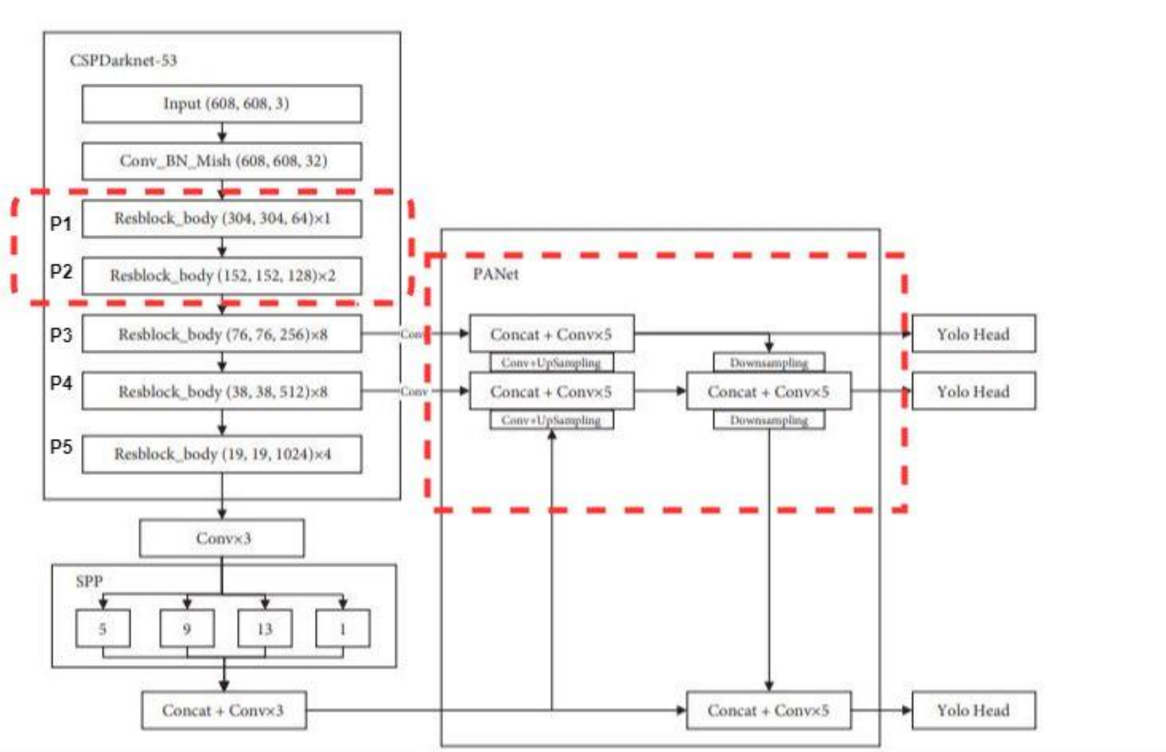


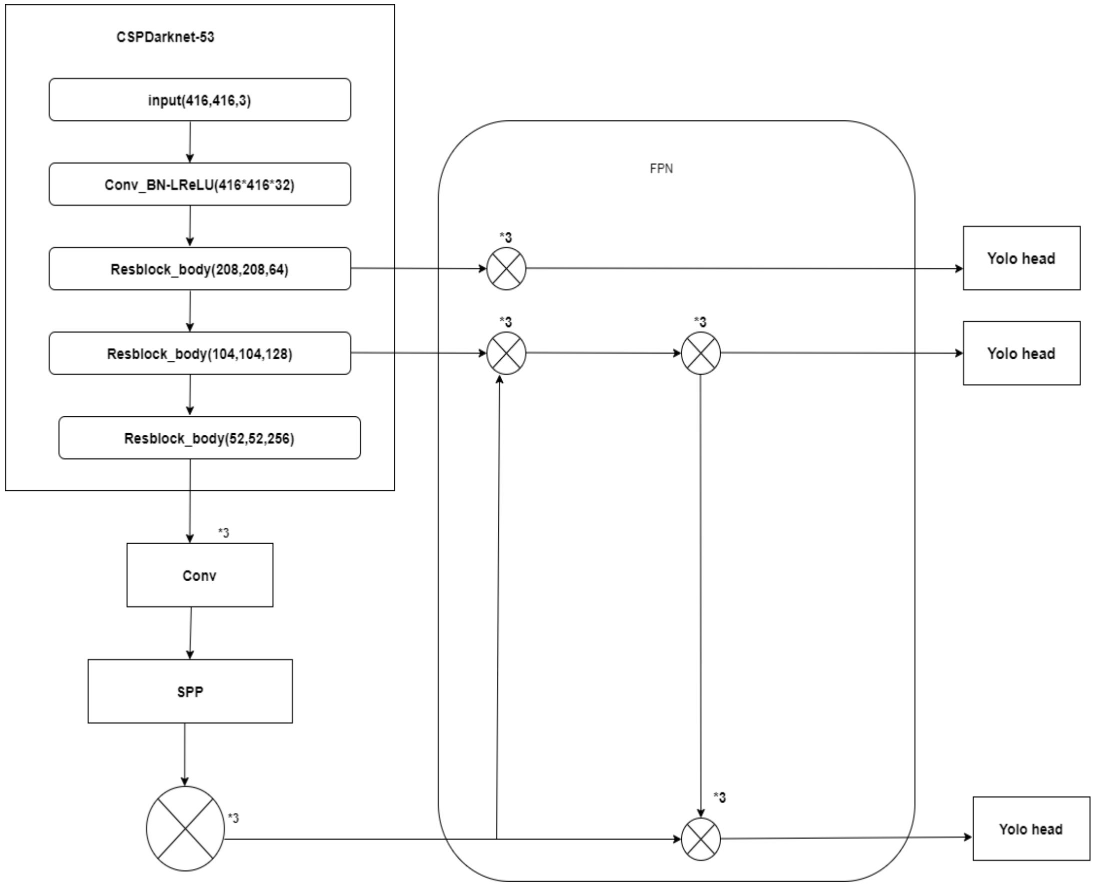
Figure 4-4 Procedure of the proposed frame work of aquatic weeds detection

The figure above shows that the labelled data will direct act as the input of the model, then the feature extractor which is backbone network will extract all features. Means all relevant features are extracted from processed data to represent the unique patterns or features of aquatic weeds. On the object detection model it has been indicated that every image has to be processed by the model to give out the correct output. In our case we have proposed a model that will actually learn all the important features of the aquatic weeds, the proposed model is obtained after optimizing the architecture of YOLOv4 which proved un desired output. With reference to the figure 4-4 the work flow of the model is just representing how the process of detecting aquatic weeds takes place and finally the object will be detected as it is shown from the figure above.

We adopted CSPDarknet-53 as a feature extractor of YOLOv4 and we have optimized the structure of YOLOv4 to a lightweight constructed model, by decreasing the number of convolution layer and residual blocks. Image scale as one of the important component of deep learning model, we take it into a serious consideration and in this case the size of the input image scale is 416*416 which were resized during the preprocessing stage. Since we are working with colored images the number of channels is 3. Our proposed model will lead to the minimum usage of the memory and improving network convergence during the training [73]. Also this will reduce the number of parameters and down sampling times in order to reduce the computation cost when training our dataset. Also to accelerate receptive field of our network by allowing it to learn more important features of the object to be detected in its relationship with the input image. Consider the figure below which is the optimized architecture of the proposed model obtained by optimizing CSPDraknet-53 and by replacing FPN for feature fusion. Consider the structure of the original YOLOv4 before the optimization and after optimization. On the CSPDarkanet-53 there are a total of five residual blocks which are named as (P1-P5), the residual blocks highlighted in red box are the ones which have been pruned refer the figure 4-5(a)



(a) Structure of the original YOLOv4



(b) Structure of the Proposed model

Figure 4-5 Proposed structure for aquatic weeds detection

Model structure of the YOLOv4 is made up with CSPDarknet-53, Path Aggregation Network (PaNet), Spatial Pyramid Pooling (SPP) and three YOLO heads as shown on the figure 4-5 (a). As the feature extractor of the YOLOv4 CSPDarknet-53 is responsible for extracting global features of the input image through a number of five residual blocks as they are shown on the figure 4-5 (a). Residual blocks are indicated in terms of **P** and the network has the size of 1*1 and 3*3. YOLOv4 structure has been designed with a number of layers and every layer in the residue blocks is connected with batch normalization layers for improving the training process, increasing model stability and it increases the performance of the network. Group of the convolution layers and mish activation layers due to its ability to provide smoothness and better gradient flow during the training. Also it contains maxpooling layers with the size of 5, 9, 13 and three YOLO heads with sizes of 19*19, 38*38 and 76*76 which are used to detect object in different scales. Therefore this is a structure of the original YOLOv4. Our study has gone further with the aim of improving the accuracy of our model by optimizing the whole structure of YOLOv4 which resulted to the figure 4-5 (b) which shows our proposed structure of our model after modification. Consider the figure below which represents the architecture of our proposed model, which shows the modification of the highlighted red parts of the structure of the original YOLOv4 as it is shown on the figure 4-5(a) and the modified structure on figure 4-5(b).

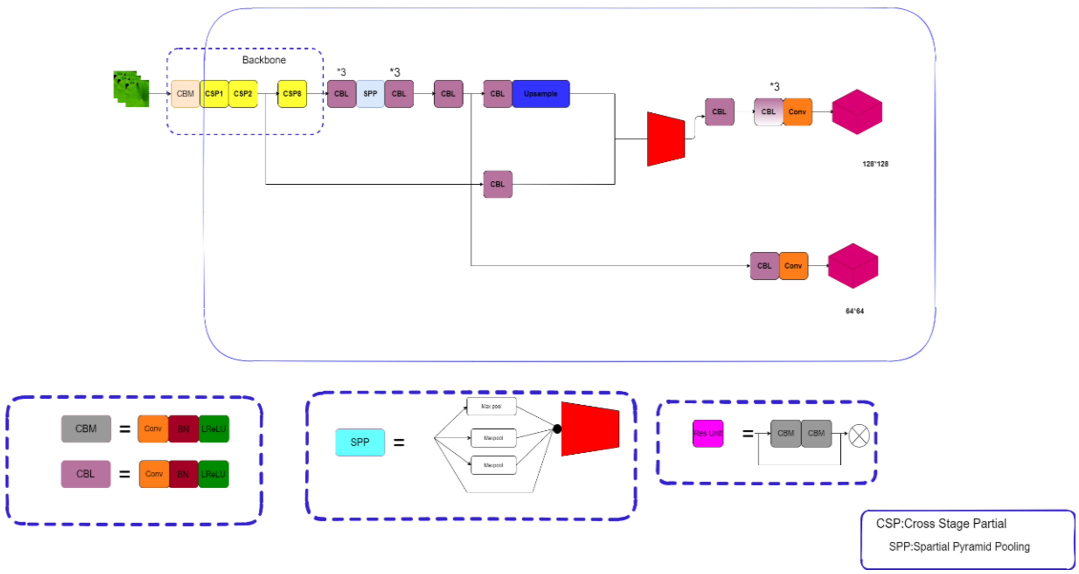


Figure 4-6 Proposed architecture for aquatic weeds detection

With reference to the figure number 4-6 which represents the architecture of our proposed model. We have reformed the architecture of YOLOv4 by constructing a lightweight feature extraction network as the backbone of our network structure. This is done by reducing the number of residual blocks as it is shown on the figure 4-5(a), it has been indicated that there are five residual blocks which have been named as (P1-P5). Where by now two residual blocks have been pruned from the original architecture of YOLOv4 which are highlighted by the red color that means (P1-P2) residual blocks have been removed. The convolution operation is made up of three Cross Stage Partial (CSP), residual unit and 1, 2 and 8 respectively are the number of the residual blocks accommodated by each CSP residual unit. In every residual block we have made a pipe that consists of two convolution layers and residual edge. Therefore every convolution block in CSP residual unit contains LReLU activation function, which replaces mish activation function used in original YOLOv4, due to its simplicity and computational efficiency thus favors the nature of our datasets. Also LReLU does not contain any exponential or division property [73]. The reason why we preferred LReLU is due to its effectiveness in avoiding "dying ReLU" problem this is due to the fact that when the input to ReLU neuron becomes negative.

The convergence of the ReLU activation function is much faster this is due to the reason that it does not contain any exponential or division property. Consider the formula below, which shows how negative value affect the training speed and the value of x will be returned to $max = (0, x)$. Where x stands for input to the function and a is a constant positive value that indicates leakiness parameters, this is defined to ensure that LReLU does not “die” when the input values are negative and the default leakiness parameter is 0.01 and it’s demonstrated as [53].

$$N(x) = \max(0, x)\begin{cases} x & if\ x \geq 0 \\ ax & Otherwise \end{cases} \tag{4-6}$$

In the formula x——Input to the function

α——A constant positive values of leakiness parameter

This prevents dying problem by implementing a very small slop and sending the values to standard ReLU when the input becomes 0 (zero), Jianbo [73]. A comprehensive analysis was made in this study to find out the best activation function that will work best in our proposed model. This is due to the fact that if the activation function is chosen in a good way it may lead to the better results of the model accuracy. That means it may produce a model that has low loss, better accuracy and stabilized model during the training. Consider the table below that shows the activation function used in the deep learning models of this study and the proposed model and the table has shown feature fusion used in this study.

Table 4-3 Backbone, Feature fusion and Activation function used by our models

Model	Backbone	Feature fusion	Activation function
YOLOv3	Darknet-19	FPN	LReLU
YOLOv4	CSPDarknet-53	PANet	Mish
Proposed Model	**CSPDarknet-28**	**FPN**	**LReLU**

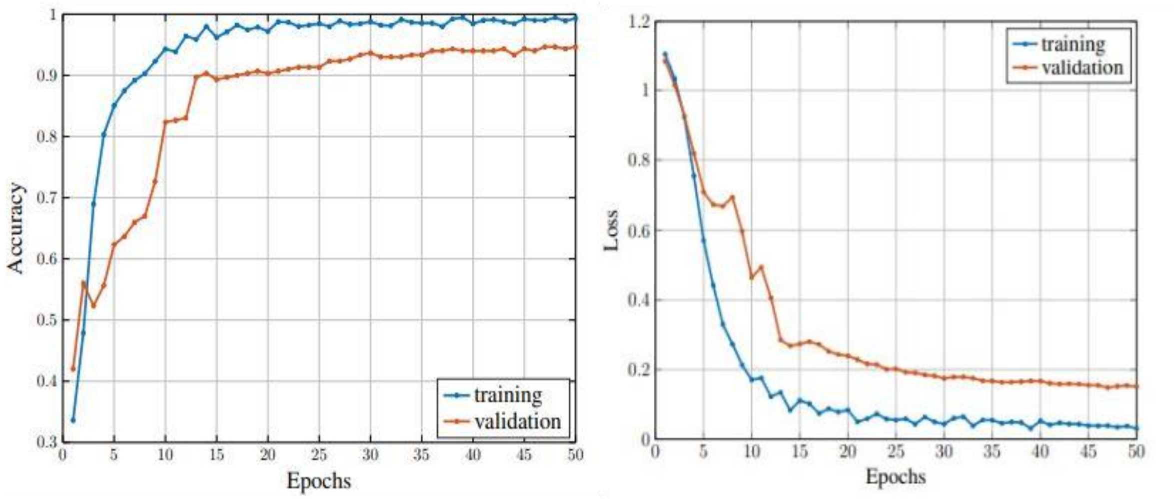
The convolution kernel size of our model is 3*3. The feature extractor will perform three down sampling operations and every operation is computed by convolution layer. However r1 and r2 are effective features which are to be extracted with dimension of 416 by 416 at the scales of 128*128 and 64*64 , then r2 feature map becomes the output which is transferred to SPP ‘Spatial Pyramid Pooling’ for the reason of extracting features under different sizes as it is shown on the figure 4-6 above. Then after three convolution layers the dimension of r2 feature map is decreased to 128, the reason why we need to reduce the dimension of r2 is to facilitate the access of the effective feature for better results. In the proposed model we have also used FPN for feature fusion due to its multi-scale feature extraction ability. Therefore it can be able to extract feature at multiple scales and this is of much more useful in our model because some images in our dataset contains very small leaf size, which are not easy to detect but with the usage of Feature Pyramid Network. We will be able to detect objects of different sizes. It will combine r1 and r2 after sampling, and then after several operations it output the YOLO head as it is shown on the figure 4-5 (b) to produce a good representation of the input image. Since the focus size of our data is too small and normally the focus (target) location relies on the anchor and in our study we have used two scales of anchor which are 20*20 and 25*25 respectively. The maximum intersection of union (max-IOU) as a matching algorithm is used to read the degree of correctness between the ground truth and anchor. From this point then the matching algorithm selects the largest matching anchor as the prediction box of the current focus or target. In order to predict the detection results. However the study has preferred the anchor scale of 20*20 and 25*25 in order to fit the target size of our datasets which actually is very small in size for a better accuracy during the training of our model.

Experimental setup, the proposed model was performed on Pytorch NVIDIA GeForce GTX 8^{th} series graphics cards, the I7 CPU and 16GB RAM. In this experiment we are going to have 50 epochs with the minima-batch sizes of 32 image instances. Our initial learning rate is $1*10^{-4}$ decreased by a factor of 0.5 after every 8 epochs. In this experiment set up we have used the original dataset and this time we have divided our dataset into three divisions as it is seen from the table below, we have also adopted average precision of AP_{50} and we have used it in our experiments for evaluation.

Table 4-4 Data set distribution for the experiment set up

Domain set	Data set (%)
Train set	70
Validation set	15
Test set	15

The analysis of the common features of the aquatic weeds has been done comprehensively and all the sets from the table above consisted of the normalized images. In our study we have taken feature analysis as a very important domain. The analysis was done on all the images available in the above table. The division of our data set for both train set, validation set and test set is 70%, 15%, 15% respectively. In these sets we have made sure that each set has its own images to avoid over fitting, which is a very big challenge in computer vision and machine learning when developing a model, which may lead to a very wrong or poor results. Our study took over fitting as a very crucial problem and we have tried much hard to make sure that all the sets are having their own images and the training of our model was conducted successfully and the trend of the learning of our model is visualized on the figure 4-7, that shows the exponential drop of the loss as the algorithm continues to learn the features of our object. The graph shows the drop of both training set and validation set, this is due to the fact that the less the loss the higher the accuracy and the better the performance of our model.



(a) Shows the accuracy of our proposed model (b) Shows the loss as the model learns

Figure 4-7 Accuracy and loss function curve of the proposed model

However the results obtained with all baseline models were not good enough to be considered as accuracy of our model for object detection. At this point we decided to improve the accuracy of our model. When applying our proposed algorithm, we had found that, the time to train the model was decreased compared to the previous experiments, the accuracy was increased by 4.01%, which leads to an accuracy of 97.79% as it is shown on the table 4-5. This confirms that the proposed algorithm had a very large impact in improving the accuracy of our model, the results were recorded and analyzed in the table below.

Table 4-5 Result obtained after the application of our

Deep learning algorithm	Whether proposed method is applied?	Increased accuracy (%)	New Accuracy
Proposed Model	√	4.01%	97.79%.

4.3.2 Comparison and analysis of the experimental results

In this study we have applied both YOLOv3 [15] and YOLOv4 [14] to detect aquatic weeds and YOLOv4 [14] has happened to be more accurate than YOLOv3 [15]. After applying so many techniques like image argumentation and image normalization to reduce pixel range for easy convergence of the network in order to improve the accuracy of our model, after conducting several experiments on YOLOv4[14] . It achieved the accuracy of 93.78% we have realized that the YOLOv4[14] is the best deep learning algorithm in detecting aquatic weeds, but with a very comprehensive study that we conducted on YOLOv4[14] we realized, that due to the use of CSPDarknet-53 it has affected much the speed of network convergence during feature extraction as a result it may lead to the increase of time and reducing learning rate. This has given us the reason to optimize the network to a simple constructed CSPDarknet-28 by reducing the number of the convolution layer from 53 to 28 in order to increase the ability of our backbone during feature extraction. Also we have replaced PANet with FPN and Mish activation function with LReLU. Our input image will be resized by 416*416 with the aspect ratio of 1:1 and the number of channels of our data set is 3. As the core of our research work we have evaluated the quality of our proposed model basing on Average Precision or AP_{50} where by the prediction accuracy must be more than 0.5. That means the bounding box prediction that shares IoU (Intersection of Union) which has a value greater than 0.5, that may result to the prediction of the True Positive. Experiments were conducted with reference of YOLOv4[14] deep learning model as a baseline model and the results of the experiments are presented on the table below .These results express the direct impact of our proposed model on the improvement of the accuracy of the aquatic weeds object detector as it is shown on the table 4.6 below.

Table 4-6 Comparison between YOLO v4 and proposed model

Model	Backbone	mAP_{50}
YOLO v4 (baseline model)	CSPDarknet-53	93.78
YOLOv3	Darknet-53	89.93
Proposed model	**CSPDarknet-28**	**97.79**

The analysis of the results after applying our proposed model as mentioned in the table 4-6, according to Table 4-6. It was found that when applying our proposed model to our aquatic weeds datasets, the accuracy of proposed model increased by 4.01%, which makes a great improvement of our model accuracy of about 97.79%. This study found that when training an object detector with YOLOv4 [14] and YOLOv3 [15], two things are important to consider: the time it takes to train the model and the number of images to train. The capacity of the data set preparation gives a high certainty after training to have an accurate model. The speed of inference and the convergence of the neural network were also noted as one of the consequences that can affect the accuracy of the model. Our proposed model has observed all these cases to ensure that the trained models have the desired accuracy. Consider the figure below that shows the comparison of each model used on this research.

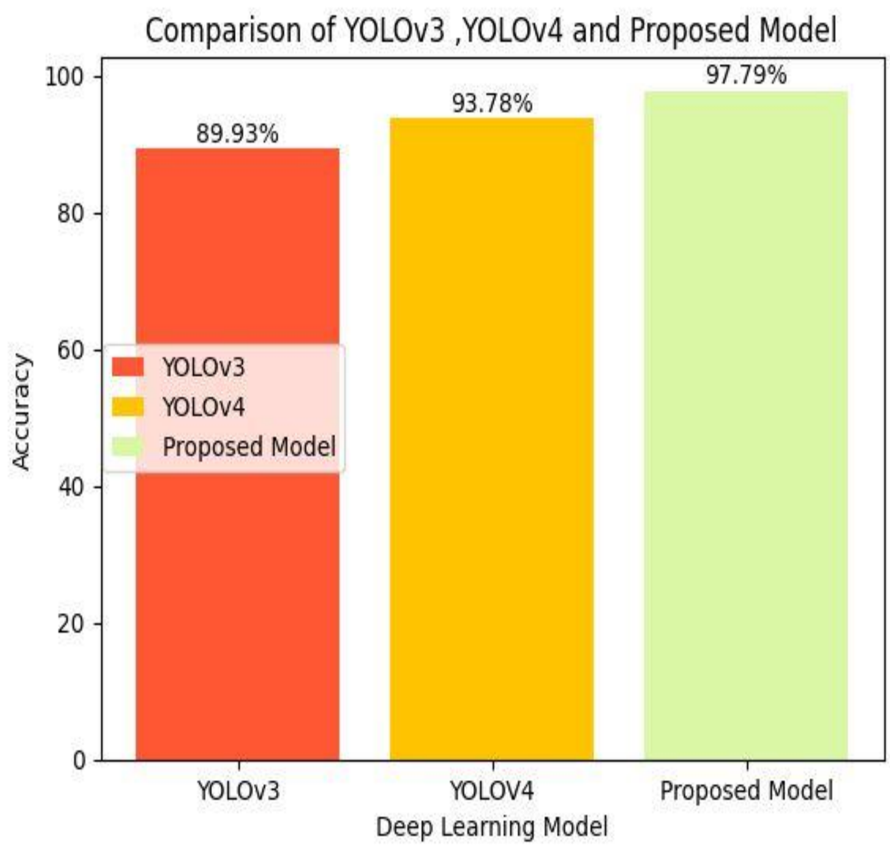


Figure 4-8 Accuracy comparison for both YOLOv3, YOLOv4 and Proposed model

We tested different samples of the aquatic weeds in both algorithms and the results are here presented for the first three samples. As you can see on Table 4-7. It is shown that our proposed model has outperformed other models used in this study in terms of detection accuracy. This is to show that our proposed algorithm has a very large impact on detecting aquatic weeds at very best accuracy. Since we have focused on improving the accuracy of YOLOv4 [14], the above information shows the large improvement of the accuracy of YOLOv4 [14]. The proposed model has several advantages over the original YOLOv4 which is the baseline model during the experimentation as follows.

(1) **Adaptation to data variation** ,dataset used in this study contain variations in the appearance of aquatic weeds, lighting conditions and backgrounds. Therefore by training proposed model on aquatic weeds datasets, at which it adopted very faster in handling these variations which resulted to the improvement of generalization and robustness capabilities. This is very challenging to the baseline model due to the size of the aquatic weeds and variations of the light condition of the aquatic weeds data, which happened to be more difficult for the network structure of the YOLOv4 to detect features of the aquatic weeds in a good accurate.

(2) **Enhanced accuracy**, a proposed model is constructed and trained specifically on aquatic weeds datasets. The network structure is designed to enable the model to better understand and detect objects that are specific to the target domain .This is due to the reason that if domain-specific knowledge is highly considered during the model training, the proposed model will archives higher accuracy and better object detection performance compared to YOLOv4[14].

(3) **Target object detection**, since our study deals with aquatic weeds detection, the size of leaves of some aquatic weeds are little bit small. Therefore the proposed model narrows down the scope of the detection to the specific size of the aquatic weeds leaf. The proposed model will be trained to focus on detecting objects that are particularly relevant to the features of aquatic weeds. The original YOLOv4 [14] is not appropriate due to its speed of network convergence.

(4) **Improved speed**, the process of optimization of CSPDarknet-53 can significantly enhance the inference speed of the proposed model, making it suitable for aquatic weeds detection which led to the improvement of the accuracy of our proposed model over the baseline model.

(5)**Lower computation requirement,**proposed model lowers computation requirement during the inference. This is done by reducing the number of operations, which is effectively improve energy efficiency and decreases the power consumption of hardware accelerators. These are some of the advantages of the proposed model over the baseline model, the size of the aquatic weeds was the driven force of the proposed model where by the architecture of the proposed model was designed to fit and to detect all the global features of the aquatic weeds basing on the characteristics of the aquatic weeds. Consider the table: 4-7 below that shows the detection accuracy of each sample.

Table 4-7 Result obtained during the testing some few images have been considered

Experiments		Aquatic weed_1	Aquatic weed_2	Aquatic weed_3	Aquatic weed_4	average accuracy
YOLOv4	Accuracy	76%	91%	86%	98%	87.75%
YOLOv3	Accuracy	67%	82%	89%	90%	82%
Proposed Model	**Accuracy**	**99%**	**99%**	**98%**	**99%**	**98.75%**

4.4 Research on the influence of image recognition tasks on the aquatic weeds datasets

4.4.1 Proposed model and its implementation in details

One of the good advantage of the convolution neural network is the powerful capabilities on feature extraction [67]. Convolution Neural Network (CNN), normally uses convolution kernel in every layer in order to compute the convolution operation on the input from previous layer and allow this layer to pass on the convolution operation, Luis [75]. However at this stage feature map will be generated as the input of each layer as shown on the figure 4-9 below. We have proposed a CNN model to determine the influence of the image recognition task on aquatic weeds. Our model based on 6-layers structure where each layer filters global features of the processed image. In this case the input image will actually be represented on array of pixel values of 256*256*3. The size of the dimension convolution kernel is set to 3 and then each convolution layer is proceeded with ReLU activation function and connected to the pooling layer. After the last convolution layer and then the data is smoothed and finally it's input into a full connected layer and output softmax of the output layer. In addition to this the first convolution layer comprises of 32 ,64, 64, 128, 128 and 256 filters respectively each with size of 5*5 ,followed by ReLU activation function the max pooling layers has a pool size of 5*5 followed by full-connected layers. Consider the figure below that shows our network structure of the proposed model to implement the influence of the image recognition on our prepared datasets.

Experiment set up the proposed model was performed on NVIDIA GeForce GTX 8th series graphics cards and the I7 CPU and 16GB RAM. In this experiments we are going to train our model at a learning rate of 0.001, a batch size of 32, adam optimizer and a number of 300 epochs. However, CUDA and cuDNN have been installed as they promote training of the deep learning algorithms on a GPU, to make it faster and efficiency. Computation resources to train our model has been highly considered in here during the training of our model to make sure that we have access to appropriate infrastructure to run our experiments in the most efficient way. However we made a nearest consideration when we split our datasets to avoid over fitting of our model, we split our datasets into three parts training sets, validation sets and test set and we will use training sets to train our model and validation sets to access its performance during the training of the model. In this case monitoring metrics on the validation sets will help us to prevent over fitting of our model.

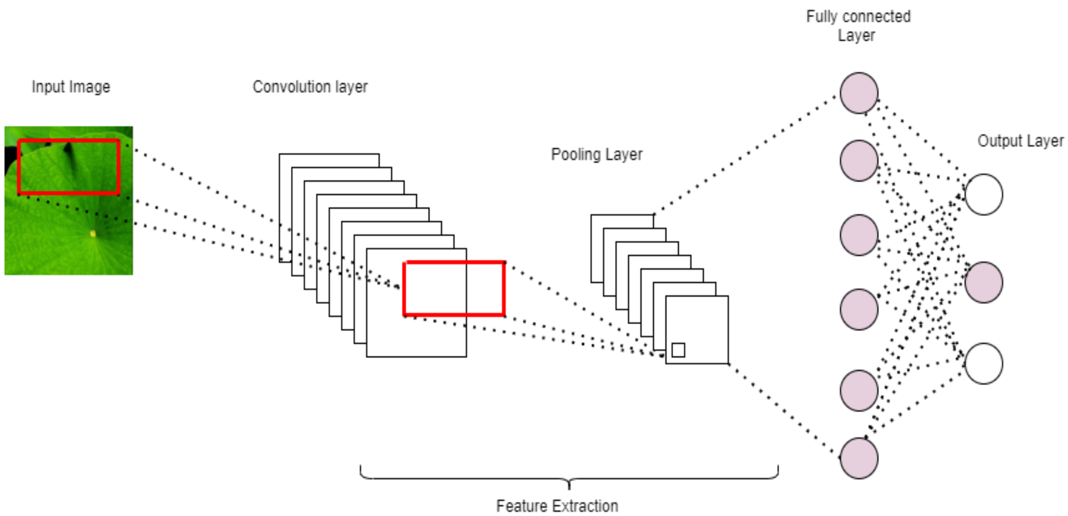


Figure 4-9 Network structure of the proposed model to implement the influence of the image recognition on aquatic weeds datasets

4.4.2 Analysis of the experimental result on the influence of image recognition on aquatic weeds dataset

Following the completion of the training process of our model after a series of 300 iterations, the ratio of our dataset has been divided into 14:3:3 and the metric used is precision, the experimental results are shown in the figure 4-10. Which shows the accuracy of 99.63% that makes analysis in our study that CNN models have a powerful ability on recognition of the aquatic weeds datasets due to the achievement of a very higher result. In addition, CNN shows that it has a very higher ability to extract relevant features of the object to be recognized. CNN layers that are designed to form network structure on figure 4-9 above are designed to automatically learn and extract features of the aquatic weeds that lead to the higher accuracy of the model with a final loss of about 0.0144. Consider the curve below which shows the training accuracy and validation accuracy. Image recognition normally has a significant influence on the aquatic management due to the availability of a big number of plants species on the water bodies. Therefore, this bring us to an intention that we need to perform recognition task for the identification of our dataset by analyzing images which were captured on the water surface during the data collection. We trained our model to recognize aquatic weeds specifically basing on their visual characteristics. The identification capability is valuable for researchers, ecologists, and field workers who need to quickly and accurately identify the presence of invasive or harmful plants in aquatic environments. Therefore the influence of the image recognition tasks will help us to identify the presence of harmful aquatic weeds at their earlier stage which will at least reduce the inversion of the aquatic weeds in a certain aquatic environment.

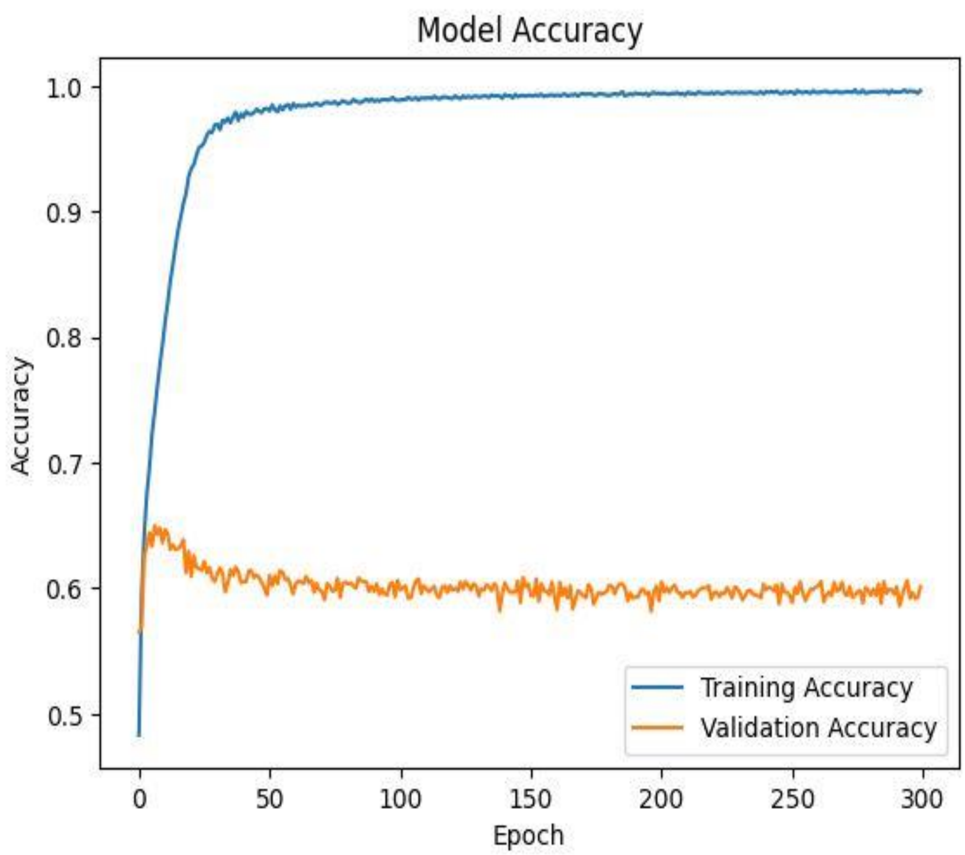


Figure 4-10 Model accuracy progress that shows training accuracy and validation accuracy

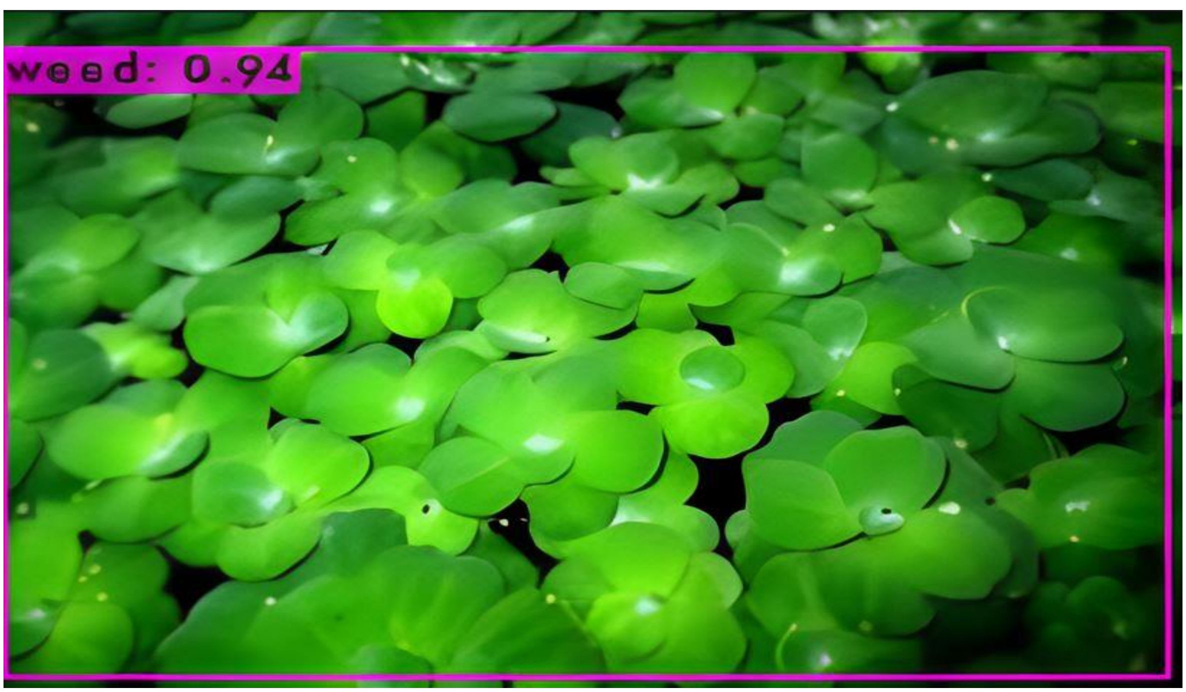
In general the recognition results show that for aquatic weeds data, considering convolution operation can serious improve the ability of the feature extraction. Reducing computation costs by enabling the model to achieve a very higher and desired accuracy, rather than direct application of the linear layer. This model has played a very great role on exposing a very higher accuracy on recognition of the aquatic weeds data. Therefore our study suggests that when designing the network structure for aquatic weeds detection you better use CNN for recognition and related task due to its higher performance. In addition to this, our study has gone very further by proposing the model that will perform detection task of our aquatic weeds. This model is obtained by modifying the architecture of the baseline model. Since was the only model which achieved a very higher and accurate results when we trained our model for the very fast time, and in this proposed model we will also make use of convolution neural network layers due to its higher capability that has been shown on the recognition task on the aquatic weeds.

4.5 Chapter summary

This chapter first, introduces the performance of the two or more efficient deep learning models on the task of the object detection, based on aquatic weeds data. In this chapter we have presented the results of our experiments conducted in different situations, experimental results show that our improved YOLOv4 model which has been obtained by optimizing CSPDarknet-53 to lightweight constructed backbone, replacing PANet with FPN for feature fusion and replacing mish activation function with LReLU activation function. It achieved the highest accuracy of 97.79%. Also we tested the models both YOLOv3, YOLOv4 and our proposed model. The results were discussed and evaluated, we found that the average of both YOLOv3 and YOLOv4 was high in all samples tested but our improved model has outperformed both YOLOv3 and YOLOv4 by achieving a very high accuracy. However our study goes even further by observing the influence of image recognition on the aquatic weeds dataset. CNN model was designed and a series of the experiments were conducted and the results were evaluated and the model achieved the accuracy of 99.63%.

Conclusively, our study suggests that due to the experiments that we conducted and the better results that were obtained. Therefore the proposed model which has been innovated in this study is the best deep learning model for the aquatic weeds detection and it also gives a direction for future research work. Consider the table above, for further demonstration both models have been tested and the sample results has been tabulated on table: 4-6.

(a) Detected weeds



(b) Detected weeds

(c) Detected weeds

(d) Detected weeds

(e) Detected weeds

Figure 4-11 Images of aquatic weeds with their detected accuracy during model testing

Some images were taken randomly and they were tested with our model and the accuracy was what we expected and all the figures above were taken and tested. The high performance of our proposed model has shown that our proposed model is the only model that can detect aquatic weeds with a very high accuracy, considering the size of aquatic weeds, nature and color of aquatic weeds compared to some other existing models, with this model we are very sure that the challenge that occurred during the elimination of the aquatic weeds has get to an end, because if our model will be brought into practical application it will facilitate the whole process of eliminating aquatic weeds ,in addition to this the results that have been displayed after performing a very scientific experiments reveals the out performance of our designed model .

Chapter 5 Benefit, Validity, Impact and Future exploration

5.1 Introduction

In this chapter we present, by connecting the dots from all the content we have been researching and giving the world the big picture of our study, we draw the conclusion that computer vision is creating a roadmap for solutions to the different problems that have been around for or persist for some years without being solved. As our research focuses on aquatic weed detection to accelerate all the ways that aquatic weeds can be eliminated or reduced to combat their effects which are very well described in chapter one.

5.2 Benefit of the research

As discussed earlier in this study, aquatic weeds are very dangerous to our environment and our study has pointed out various effects of aquatic weeds researched by various researchers. One of the famous research conducted in China in 2007 by Jianbo [74]. They had identified some serious impacts of aquatic weeds on the aquatic environment. In their article they had mentioned various effects of aquatic weeds such as, the decline of natural biodiversity, this is a very serious effect of the aquatic weeds due to the fact that, if the natural biodiversity is affected or disturbed by any means, the life of living organisms within a particular area will also be affected because biologically biodiversity concerns all different kind of life between plants and other organisms within a particular area. Invasion of aquatic weeds at the particular area needs to be taken very serious this is due to the fact that the inversion of the aquatic weeds within a particular area will directly affect natural biodiversity .Another effect pointed by Jianbo[74] is the iteration of the ecosystem, as we know that ecosystem is the biological interaction of the organisms with their natural environment. The exchange of food nutrients and gas take place in the same way, now let’s have a picture that a certain area has been affected by aquatic weeds and automatically the ecosystem of that area has been deteriorated, this will automatically switch all living organisms to death because the life survival cannot be possible if at all the ecosystem isn’t existing. Another big effect of the aquatic weeds they pointed out is the spread of the disease that affect human health, yes off course some research have been conducted and it was found out that the inversion of aquatic weeds in a certain area might cause some serious diseases .This was one of the research work that has been pointing out on the effect of the aquatic weeds. Earlier in chapter one we had discussed economic effects that can be caused by the inversion of the aquatic weeds, but here we see biological effects and when we talk of the biology we mean the life of living organisms in particular .Therefore, aquatic weeds should be taken in a very serious way to make sure that all these natural effects are not happening, for example natural biodiversity if is disturbed there is no way we can replace it rather than losing the life of the aquatic organism.

Our study analyzes all the possible dangerous effects that can be caused by aquatic weeds and we decided to come up with the idea, that at least to accelerate the process of eliminating aquatic weeds as described in the previous chapters. So if the results of our study are put into practice. The following benefits are observed as we predicted during the research process.

Firstly, if the growth of the aquatic weeds is controlled then agricultural production will be high and hence decrease of the production cost. This will be of more advantage to the individuals because they will increase their daily income and hence the increase of the country GDP.

Secondly, equally distribution of the available water, mineral nutrients and sun light the importance of the sun light to the growth and survival of the aquatic plants is very important. Therefore the inversion of the aquatic weeds within a certain water body will actually affect the distribution of the nutrients to the plants around the water bodies. Also it will affect the sun light distribution to the plants as a result natural plants will die for loosing important nutrients for their survival, if our results will be brought into real practical in the field we will be able to detect aquatic weeds earlier and fix their spread within a short period of time.

Thirdly, finally our study will be able to maintain biodiversity on its position and maintaining the ecosystem within the area of any water body. This will only be possible if the results of our research will be applied on detection and localization of the aquatic weeds on the target areas.

5.3 Validity

5.3.1 Internal validity

When we set the values of random=0 in the cfg file we found that the time needed for training decreased, but there was no impact on the image size and this was taken into account as internal validity. Images were on resized 608*608 with the aspect ratio of 1:1, if we set the value of random to 1 (random=1). The algorithms (YOLOv3 and YOLOv4) adjust their networks after every 10 iterations and we had 2000 and 3000 iterations set for YOLOv3 and YOLOv4, respectively. However, the resizing of our images has a very strong impact on our object detector, despite low-resolution networks [73]. All of this was taken very seriously to ensure that any errors that may be caused during the training of our proposed model are fixed, so that we get the object detector with a very high accuracy, which is a result of our proposed model.

5.3.2 External validity

Usually, external validation indicates how well the results obtained during the experiment can be useful in different situations. In our case we developed a model which is actually an object detector improved after applying our proposed model which increased the accuracy of our object detector to the desired accuracy. The aim of this object detector is to be useful in different situations to facilitate the process of detecting and locating aquatic weeds. Hopefully the result of our study will bring a third eye overseeing all the ways in which aquatic weeds can be eliminated. Although during testing both the baseline algorithms and the proposed algorithm failed to detect some aquatic weeds due to the small size of the weeds available in the image .In here we have put one of the sample images that our object detector was not able to detect the aquatic weeds on the image, Although all of our images in our dataset were resized to the size of 608*608 to make sure that the features on the images are well learned by the algorithms for better results. Consider the figure bellow that represent one of the image that our object detector failed to detect.

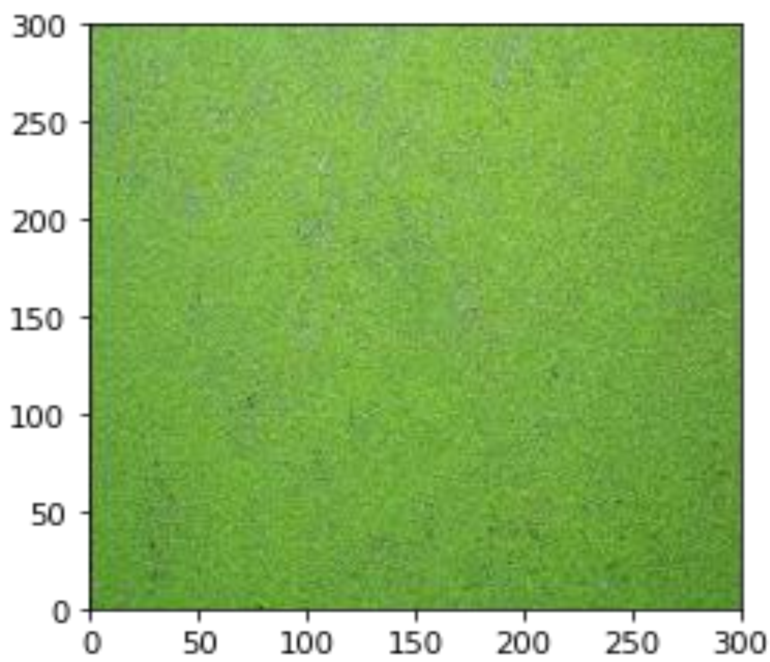


Figure 5-1 Image of aquatic weed that is counted as external validity threat due to the small size of weeds

5.4 Impact of the research on aquatic weeds

Water is a very basic need for human life, not only for humans but also for every living organism. We need water for our survival, so our job is to conserve water resources, but there are so many factors that can lead to water shortage including drought caused by climate change and of course it is natural disaster. In recent years we have witnessed or heard about the impact of aquatic weeds on several water resources around the world which are now considered to be a very big problem for other places around the world [42]. Many researchers have conducted a series of research work in the field of water management and reported many times, and many countries have taken steps to eliminate the existence of aquatic weeds. As described in Chapter one, that aquatic weeds have an impact on so many sectors including the agricultural sector, the fisheries sector, the transport sector and the industrial sectors, hence causes the government to lose her national revenue due to different effects caused by aquatic weeds.

In this research we have identified these problems and we have tried to find a better way on how we can use technology to accelerate the speed of eliminating aquatic weeds .During this time of research a comprehensive literature review has been done just to look for the better way we can do to make sure that aquatic weeds are completely eliminated and hence water resources are easily managed. Therefore our research pointed out two questions which are referred to be the backbone of our study. The first question was how do we know the exact location of aquatic weeds? And the last question was how do we distinguish aquatic weeds and other aquatic plants which are very important in supporting aquatic life?, the **sec**ond question is very basic due to a reason that in many water bodies there are living organisms whose life and habitat depends on the aquatic plants, therefore there is a need of differentiating first the aquatic weeds and aquatic plants before further actions of eliminating aquatic weeds are taken.

Therefore to find the answer of the two research questions we thought of using a computer vision technology to recognize and localize the aquatic weeds in real-time detection. This create a thirst to us on doing a research on convolution neural networks especially deep learning algorithm, to find out the best algorithm that will be used to identify and localize the aquatic weeds on the surface of water. Hence to give a proper direction to the people who are dealing with aquatic weeds elimination to remove weeds and not otherwise and to save the time of going around looking for the aquatic weeds. To improve the accuracy of our model we have even went further by implementing our proposed model to improve the accuracy of our object detector.

The research conducted by Samah[5] which was entitled “Farmer’s Awareness to the Impact of Aquatic Weeds”, they stated that many people are aware of the effect of aquatic weeds but there is a big challenge in controlling aquatic weeds from the water sources due to the fast growth of the aquatic weeds, the current methods of eliminating aquatic weeds takes more time for instance. The whole process of travelling looking for aquatic weeds and after finding them, they peal them up ,dry them and burning them or the use of herbicides or sometimes they implant insects to the areas which are affected by the aquatic weeds to eliminate them. But the question still the same how do we identify them .This is the motive of why we decided to look for the deep learning algorithm which will be very accurate and effective in performing aquatic weeds detection. In this case we have selected two deep learning algorithms which are YOLOv3[15] and YOLOv4[15] to be used to detect aquatic weeds as a baseline models. The results of our experiments have been indicated and a comprehensive analysis have been done and we have found out that proposed model is the most accurate model and is suitable algorithm to be used in aquatic weeds detection. We even combined our proposed method and our selected algorithm to improve the accuracy of our model and this makes the conclusion of our research and the experiments that have been done.

5.5 Future exploration

In our study we depend much on the efficiency of deep learning models this is due to the fact that, we use deep learning models to detect aquatic weeds and to classify them, if deep learning detection models efficiency will be increased automatically the future of our study will be attained. In the future the model will be fed in robotic drones that will be going around while localizing the position of the aquatic weeds, therefore the detection time should be in hand with the speed of the moving drones or vehicle that carries herbicides to kill aquatic weeds. Also we have seen in the YOLOv3[15] is capable of detecting very small object ,this might have been a chance to improve effectiveness of deep leaning algorithms like YOLOv4[14] and YOLOv3[15], to increase its ability to detect small objects with a fast detecting speed. This draw attention to us that the only way to figure out with aquatic weeds is to eliminate them when they are in early stages of growth, here then the challenge comes on getting dataset of aquatic weeds at their early stage is not an easy way but it would have been better to control them earlier before they grow up.

There concepts of ensemble learning and transfer learning must also be employed in deep learning models at least to increase the efficiency of detecting aquatic weeds. Normally the large number of dataset predicts an accurate result. If computer vision domain explores improvement in data augmentation then weeds detection in the future will be implemented in a very accurate manner to get a better result. Systems and some other concepts proposed by researchers to be improved in deep learning models to improve the efficiency and effective way to control aquatic weeds, for example real time inference and incremental learning this also requires large scale data set to be used to train the model for better result. This research is projected on the use of the advanced smart machines that will be used in the future to control and manage the aquatic weeds. But this will only be possible if machine learning models will be advanced, there are so many things that are to be improved to increase the ability of object detector to detect weeds. Some have been discussed in this subsection of future work, but there also the concept of the deep segmentation architecture, if this also will be reviewed in the advancement of the deep leaning model, this will lead to the increase of the accuracy of the models. Although it has been observed on both YOLOv3[15] and YOLOv4[14] the good the dataset preparation the better the model. Therefore it is the call that upon using different methods on object detection we better center our ideas on dataset preparation.

5.6 Chapter Summary

In this chapter we have presented the most important components of our work with the two different means that is external validity and internal validity. These cases have allowed us to know many factors that can be improved with our method to increase the accuracy of our models. We went even further by proposing the possibility of having aquatic weed data sets taken at a time when aquatic weeds are not yet matured or at their early stage of growth, this is because our proposed method, in line with YOLOv3[15] and YOLOv4[14] has failed to detect aquatic weeds at their early stage due to the size of the weeds. As shown in Figure 5-1, our study has also suggested to the improvement of the state-of-the-art detection algorithms that, the inference speed and network resolution should also be improved.

Conclusion

This work researches on the task of using deep leaning method for aquatic weeds detection based on Lake Tanganyika (6.2556°S, 29.5108°E) and Lake Victoria (0.7558°S, 33.4384°E) which are located at the Eastern part of Africa (Tanzania), on the other hand this work researches on improving the YOLOv4 model to increase the performance and efficiency of deep learning models, which can significantly reduce memory and time cost during training and testing. Since our work centers on detection task which involves small objects. Therefore we adopted CSPDarknet-53 and we optimized it to a lightweight model and we replaced PANet with FPN for feature fusion, due to its multiple scale feature extraction ability. The main results of this work are as follows.

1. A newly application scenario is proposed for aquatic weeds detection. We collected datasets from water sources like Lake Tanganyika (6.2556°S, 29.5108°E) and Lake Victoria (0.7558°S, 33.4384°E) in Tanzania and these data were collected to support the idea of our work and we cleaned, analyzing and labeling the data of about 5094 images.

2. Overall performance of deep learning model on aquatic weeds detection tasks. At first we assured the feasibility of deep leaning models to detect aquatic weeds, then we conducted different comparative experiments on aquatic weeds relying on accuracy and computing time as our evaluation indicators. The results of experiments conducted show that YOLOv4 achieved the accuracy of 93.78% and YOLOv3 achieved the accuracy of 89.93%. In addition to this, experiments were also carried out to compare detection effects on different samples of aquatic weeds and the results shows that all those images which were taken with enough light and a good weather condition, the average accuracy was outstanding compared to the ones which were taken during the bad whether condition. However the combination of all images was of more advantage on improving the accuracy of our model.

3. An improved model of YOLOv4 proposed a model for aquatic weeds detection by optimizing the backbone CSPDarknet-53 to a lightweight model, replacing PANet with FPN for feature fusion, due to its multiple scale feature extraction ability that favors the small size of our datasets. In addition to this we changed mish activation function to LReLU which fits the nature of our datasets. Experiments were carried out on the datasets collected in this paper. The results show that the model proposed on this paper can effectively improve the detection accuracy of the aquatic weeds. The experiments show that the proposed model achieved the accuracy of 97.79% compared to the original YOLOv4 which outperformed original YOLOv4 by 4.01%. In addition, this paper show that the data set prepared is more effective by determining the influence of image recognition tasks on our datasets and the results achieved is 99.63as training accuracy.

In coming days we should keep on researching on the following aspects:

1 Even though the analysis, feasibility verification and comparison of the deep leaning

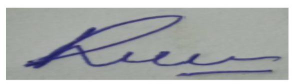
approaches, detection tasks based on aquatic weeds datasets carried out on this paper, the proposed algorithm should also be extensively in a real world application.

2 The datasets used in this paper, is the datasets of the surface floating aquatic weeds found on Lake Tanganyika and Lake Victoria in Africa Tanzania, we could not consider the effect of our proposed model on completely submerged aquatic weeds and freely floating weeds under water, therefore this can be included in the future as scope of research. On top of this due to the size of our datasets researches can also be done on the very best algorithm to fit our datasets.

Harbin Institute of Technology dissertation originality statement and use rights

学位论文原创性声明

本人郑重声明：此处所提交的学位论文 “Aquatic weeds detection using deep learning”，是本人在导师指导下，在哈尔滨工业大学攻读学位期间独立进行研究工作所取得的成果，且学位论文中除已标注引用文献的部分外不包含他人完成或已发表的研究成果。对本学位论文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。

作者签名：  日期：2023 年 07 月 02日

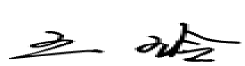
学位论文使用权限

学位论文是研究生在哈尔滨工业大学攻读学位期间完成的成果，知识产权归属哈尔滨工业大学。学位论文的使用权限如下：

（1）学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文，并向国家图书馆报送学位论文；（2）学校可以将学位论文部分或全部内容编入有关数据库进行检索和提供相应阅览服务；（3）研究生毕业后发表与此学位论文研究成果相关的学术论文和其他成果时，应征得导师同意，且第一署名单位为哈尔滨工业大学。

保密论文在保密期内遵守有关保密规定，解密后适用于此使用权限规定。本人知悉学位论文的使用权限，并将遵守有关规定。

作者签名：  日期：2023 年 07月 2 日

导师签名：  日期：2023年 07月 2日

References

[1] Mrs. Lidia Lancar (FRANCE) Mr. Kevin Krake (Australia). Aquatic Weeds & their Management Reviewers, international Commission on Irrigation and Drainage. March 2002:2-24.

[2] Carole A. Lembi, Aquatic Weed Specialist, Purdue University and Aquatic Pest Control, University of California Leaflet 2961 and Aquatic Weed Management Aquatic Weed Control Adapted from North Central Regional Extension Publication No. 241, (SRAC Pub No 361 – 2013.

[3] N. Otsu, “A threshold selection method from gray-level histograms,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–53, 1979.

[4] Torawane and D. N. Mokat AQUATIC WEED MANAGEMENT AND DEVELOPMENT THROUGH INNOVATIVE WAYS S. D. Department of Botany, Savitribai Phule Pune University, Pune - 411007, India,2019.

[5] Samah O. A. Alhadari 1*,Awadallah B. Dafaallah 2 and Mohamed S. Zaroug .Farmer’s Awareness to the Impact of Aquatic Weeds in some Minor Canals, Gezira, Sudan Scheme.2018 :1-2.

[6] Carole A. Lembi, Professor of Botany.Botany and Plant Pathology,Purdue University ,lembi@purdue.edu .Identifying and Managing Aquatic Vegetation.

[7] SOLITUDE LAKE MANAGEMENT .Your Comprehensive Guide to Managing Aquatic Weeds and Algae,https://www.solitudelakemanagement.com/free-guides/lake-and-pond-algae-control-strategies/.

[8] Navneet Dalal and Bill Triggs INRIA Rhone-Alps, Histograms of Oriented Gradients for Human Detection, ˆ 655 avenue de l’Europe, Montbonnot 38334, France {Navneet.Dalal,Bill.Triggs}@inrialpes.fr, http://lear.inrialpes.fr,2005, 2-5.

[9] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik R-CNN for Object Detection, (UC Berkeley) 2014, 4-5.

[10] Fast R-CNN Ross Girshick Microsoft Research rbg@microsoft.com,2015, pp 4-7.

[11] Wei Liu1 , Dragomir Anguelov2 , Dumitru Erhan3 , Christian Szegedy3 , Scott Reed4 , Cheng-Yang Fu1 , Alexander C. Ber , SSD: Single Shot MultiBox Detector,2016.

[12] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. You only look once: Unified, real time object detection, 2016:1-8.

[13] Mohanad Al-Hasanat et al**,** RetinaNet-Based Approach for Object Detection and Distance Estimation in an Image**,**2021 2-6.

[14] Alexey Bochkovskiy∗.YOLOv4: Optimal Speed and Accuracy of Object Detection,

2020, 2-12.

[15] Joseph Redmon Ali Farhadi University of Washington .YOLOv3: An Incremental Improvement, 2018 .1-9.

[16] Priyanka Sahu, Anuradha Chug, Amit Prakash, (Guru Gobind Singh Inderprastha University, New Delhi, India). Challenges and Issues in Plant Disease Detection Using Deep Learning. 2021:2-3.

[17] Khanna, S.; Santos, M.J.; Hestir, E.L.; Ustin, S.L. Plant community dynamics relative to the changing distribution of a highly invasive species, Eichhornia crassipes: A remote sensing perspective. Biol. Invasions 2012 , 3-10.

[18] Baider, C.; Florens, F.V. Control of invasive alien weeds averts imminent plant extinction. Biol. Invasions 2011.

[19] Kathryn J. Sheffield 1,* , Daniel Clements 2 , Darryl J. Clune 3 , Angela Constantine 4 and Tony M. Dugdale 1 ,Detection of Aquatic Alligator Weed (Alternanthera philoxeroides) from Aerial Imagery Using Random Forest Classification ,2022.

[20] Bertacchi, A.; Giannini, V.; Di Franco, C.; Silvestri, N. Using unmanned aerial vehicles for vegetation mapping and identification of botanical species in wetlands ,2019.

[21] Cavalli, R.M.; Laneve, G.; Fusilli, L.; Pignatti, S.; Santini, F. Remote sensing water observation for supporting Lake Victoria weed management. J. Environ. Manag. 2009, 6-7.

[22] G. Mallya, P. Mjema and J. Ndunguru* Water Hyacinth Control through Integrated Weed Management Strategies in Tanzania, 2001, 1-3 .

[23] Twongo, T. ,Aquatic weeds, their impacts on fisheries and environment, and their control,2021.

[24] Fischler, M., & Elschlager, R. The representation and matching of pictorial structures. IEEE Transactions on Computers 1973, 68–92.

[25] Osuna, E., Freund, R., & Girosit, F. Training support vector machines: An application to face detection 1997, In CVPR (pp. 130–136).

[26] Ponce, J., Hebert, M., Schmid, C., & Zisserman, A. Toward category level object recognition. Berlin: Springer, 2007:120-127.

[27] Xiao, R., Zhu, L., & Zhang, H. Boosting chain learning for object detection.2003, In ICCV (pp. 709–715).

[28] Hinton, G., & Salakhutdinov, R. Reducing the dimensionality of data with neural networks. Science, 2006:1-3.

[29] Krizhevsky, A., Sutskever, I., & Hinton, G. ImageNet classification with deep convolutional neural networks. In NIPS, 2012.

[30] He, K., Zhang, X., Ren, S., & Sun, J. Spatial pyramid pooling in deep convolutional

networks for visual recognition, 2014:1-2.

[31] Ramya. a1, Venkateswara Gupta pol a2 et al .Comparison of YOLOv3, YOLOv4 and YOLOv5 Performance for Detection of Blood Cells, 2021:1-2.

[32] Liu , L., Ouyang, W., Wang, X. et al. Deep Learning for Generic Object Detection: A Survey.2019:2-3.

[33] Jun Liu and Xuewei Wang .Plant diseases and pests detection based on deep learning: a review, 2021:3-5.

[34] Alex L. Zimdahl. Weed-Management Systems , in Fundamentals of Weed Science (Fifth Edition), 2018. (pp. 147–149).

[35] Alex Olsen, Dmitry. A.DeepWeeds: A Multiclass Weed Species Image Dataset for Deep Learning, 2019:1-4.

[36] Giuseppe G. A. Celano Leipzig University , A ResNet-50-based Convolutional Neural Network Model for Language ID Identification from Speech Recordings,2021,p 1-7.

[37] Christian Szegedy Google Inc. szegedy@google.com Vincent Vanhoucke vanhoucke@google.com Sergey Ioffe sioffe@google.com Jonathon Shlen , Rethinking the Inception Architecture for Computer Vision,2015, 3-6.

[38] Ferreira, A.; Freitas, D.; Silva, G.; Pistori, H.; Folhes, M. Unsupervised deep learning and semi-automatic data labeling in weed discrimination. Comput. Electron. Agric. 2019, 165, 104963.

[39] Kun Hua Guy Colemanb Shanzengc .Graph weeds net: A graph-based deep learning method for weed recognition, 2018:1-4.

[40] Michael Schmidt, Christian Witte .Monitoring aquatic weeds in a river system using SPOT 5 satellite imagery, 2019:3-5.

[41] Ali Haydar Göktoğan A Rotary-wing Unmanned Air Vehicle for Aquatic Weed Surveillance and Management, 2010:3-4.

[42] Kentucky ,aquatic weed control Pesticide Safety Education Program,2016. https://www.uky.edu/Ag/Entomology/PSEP/cat5control.html .

[43] Muhammad Umair ahmad, 1 Sidra Ashiq Extraction of Plant Leaf Using Deep Learning 1, 2022.

[44] Zhipeng Deng a , Hao Sun a , Shilin Zhou a .Multi-scale object detection in remote sensing imagery with convolutional neural networks ,2018:3-5.

[45] Forest N Landola. SQueezeNet: AlexNet-level accuracy with 50x fewer parameters,2016.

[46] Howard AG, Zhu M, Chen B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications [A/OL]. (2017-04-16)[2020-12-05].

[47] Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[C/OL]//2018 IEEE/CVF Conference on Computer Vision and Pattern

Recognition. Salt Lake City, UT: IEEE, 2018: 4510-4520[2020-12-25].

[48] HE K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition [A/OL]. (2015-12-10)[2020-09-22]. arXiv: 1512.03385. http:// arxiv.org/abs/1512 .03385.

[49] Zhang X, Zhou X, LIN M, et al. ShuffleNet: An Extremely Efficient ConvolutionalNeural Network for Mobile Devices[C/OL]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT : IEEE, 2018:6848-6856[2021-09-11] .

[50] MA N, Zhang X, Zheng HT, et al. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design[C/OL]//2018: 116-131[2022-05-09].

[51] Karen Simonyan ∗ & Andrew Zisserman +, VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION,2015, 3-10.

[52] Gao Huang∗ Cornell University, Zhuang Liu∗ Tsinghua University, Laurens van der Maaten Facebook AI Research, Densely Connected Convolutional Networks,2018,pp 3-6.

[53] Francois Chollet; Xception: Deep Learning With Depthwise Separable Convolutions Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1251-1258.

[54] G. Saleem, M. Akhtar, N. Ahmed, W.S. Qureshi .Automated analysis of visual leaf shape features for plant classification Author links open overlay panel, 2018:1-3 .

[55] G. Zhao, T. Ahonen, J. Matas, and M. Pietikainen, "Rotationinvariant image and video description with local binary pattern features," IEEE Transactions on Image Processing, 2012, vol. 21, no. 4, pp. 1465–1477.

[56] Wu, F. S. Bao, E. Y. Xu, Y.-X. Wang, Y.-F. Chang, and Q.-L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," 2007 IEEE International Symposium on Signal Processing and Information Technology, pp. 6–11.

[57] W. S. Jeon and S. Y. Rhee, "Plant Leaf Recognition Using a Convolution," Neural Network, pp. 26–34, 2017.

[58] M. Ali, A. Al-Ani, D. Eamus, and D. Tan, "An algorithm based on the RGB colour model to estimate plant chlorophyll and nitrogen contents," 2014.

[59] N. Kumar, P. N. Belhumeur, A. Biswas et al., "Leafsnap: a computer vision system for automatic plant species identification,", 2012 in Proceedings of the European Conference on Computer Vision, pp. 502–513.

[60] Rasha Muthana1, Assad N. Alshareefi2, Techniques in De-Blurring Image , 1Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq,2020,2-9.

[61] Perez and J. wang, “The effectiveness of data augmentation in image classification using deep learning,” arXiv preprint arXiv: 1712.04621, 2017.

[62] HernándeZ-garcía and P. König, “Data augmentation instead of explicit regularization,” arXiv preprint arXiv: 1806.03852, 2018.

[63] Redmon∗ , Santosh Divvala∗†, Ross Girshick¶ , Ali Farhadi∗,You Only Look Once: Unified, Real-Time Object Detection Joseph,2016.

[64] Adnan Farooq , Student Member, IEEE, JIANKUN HU , Senior Member, IEEE and XIUPING JIA , Senior Member Analysis of Spectral Bands and Spatial Resolutions for Weed Classification Via Deep Convolutional Neural Network, IEEE 2019.

[65] Eugene Charmiak .Introduction to Deep Learning, 2016. (pp. 95–100).

[66] Sahla Muhammed Ali Graduate Student, Department of Information Technology, Rajagiri School of Engineering and Technology, Kerala, India Comparative Analysis of YOLOv3, YOLOv4 and YOLOv5 for Sign Language Detection,2021:1-2.

[67] Ferdinand Holles. Classifying Raman spectra of extracellular vesicles based on convolutional neural networks for prostate cancer detection, 2019:2-3.

[68] A S M Mahmudul Hasan∗1,2, Ferdous Sohel1,2, Dean Diepeveen2,3,4, Hamid Laga1,5, and Michael G.K. Jones2 A Survey of Deep Learning Techniques for Weed Detection from Images ,2021:2-5.

[69] Yang-Yang Zheng 1, Jian-lei kong 1, 2.CropDeep: The Crop Vision Dataset for Deep-Learning-Based Classification and Detection in Precision Agriculture, 2019:1-3.

[70] Kavir Osorio 1, Andrés Puerto 1 , Cesar Pedraza 1,et al. Deep Learning Approach for Weed Detection in Lettuce Crops Using Multispectral Images ,2020 :3-5.

[71] Shaun M. Sharpe 1, Arnold W. Schumann2 & Nathan S. Boyd1 Goosegrass Detection in Strawberry and Tomato Using a Convolutional Neural Network, 2020: 3-4.

[72] Sokolova, M. & LAPALME, G. A systematic analysis of performance measures for classification tasks. Inf. Process. Manage 45, 427–437, (2009).

[73] Andrew L Maas, AwnI Y Hannun, Andrew Y NG, et al. Rectifier nonlineari-ties improve neural network acoustic models. In Proc. icml, volume 30, page 3. Citeseer, 2013:1-3.

[74] Jianbo LU Æ jianguo WU Æ ZHIHUI FU Æ LEI ZHU .Water Hyacinth in China: A Sustainability Science-Based Management Framework, 2007.

[75] Luis A.M. Pereira A, Rodrigo Y.M. Nakamura et al .Aquatic weed automatic classification using machine learning techniques, 2018:3-4.

Acknowledgment

First and foremost, I would like to thank Almighty God for everything he has done for me, guiding me to the successful completion of my research thesis work especially to this period of covid-19 pandemic.

My deepest gratitude is to my supervisor, Prof. Wang Ling (王玲) who is supervising my research work and guiding me in preparing and documenting my Research Thesis in a proper manner. Her support is playing a vital role in the success of this Research thesis. I feel very lucky to be Prof Wang Ling’s students. I am very grateful to Prof Wang Ling for her care and guidance in the past one year. Every time I discuss my research direction with Prof Wang Ling, Prof Wang Ling is always able to give advice from a higher perspective that’s gave me a thirst to continue working hard all the time, I really appreciate the time and advice that I always get from Prof Wang Ling

I would also like to show my sincere gratitude and appreciation to the Chinese Government for the great scholarship opportunity, their support, and the cooperation given to me over the whole period of my study.

My thanks also goes to my country who trusted me by giving me this special chance to pursue my master’s degree in China and to have an opportunity to learn Chinese culture and conducting my research with well experienced Professor, that made me a new version of myself by increasing my capacity in doing research as well, I would also like to thank everyone who contributed to the success of this research.

My word of appreciation and thanks go to the staff and all lecturers at the School of Computer Science and Technology at Harbin Institute of Technology for their valuable support and time during the whole period of my studies, and their advices toward my research work gave me energy all the time to keep working hard.

Finally, I would like to thank my parents, my wife Neema and my daughter Elizabeth for their care and support especially during the period of my studies, which happened during the period of covid-19 pandemic, they gave too much support to make sure that I finish my studies.