Trash Detection In An Aquatic Environment

by

Nishat Mahmud Roshni
19101420
Meshkatul Arefin
19301080
Kazi Masfiqul Alam Joy
16101151
Marjanul Hassan
16101319
Shashwata Karmakar
19101399

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
January 2024

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Kazi Masfiqul Alam Joy

16101151

Shashwata Karmakar

19101399

Nishat Mahmud Roshni

19101420

Marjanul Hassan

16101319

Meshkatul Arefin

19301080

Approval

The thesis/project titled "Trash Detection In An Aquatic Environment" submitted by

1. Nishat Mahmud Roshni (19101420)
2. Marjanul Hassan (16101319)
3. Kazi Masfiqul Alam Joy (16101151)
4. Meshkatul Arefin (19301080)
5. Shashwata Karmakar (19301399)

Of Fall, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January, 2024.

Examining Committee:

Supervisor:
(Member)

Amitabha Chakrabarty, PhD
Professor
Department of Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Ethics Statement

It is our responsibility to find solutions to the problems we humans create in the world that are harmful to us and other creatures.

Abstract

Trash in the water bodies is alarming for the world nowadays. As it is a limited source, pollution caused by the trash is threatening for the environment, wild-life and long term effect for the human health and economy. In developing countries like Bangladesh, where a large number of the species of biodiversity belongs to the aquatic environment, pollution caused by waste can have serious consequences for the economy and the endangered species of the biodiversity. In order to address these issues, this paper presents the analysis and performances of the detection models which will be beneficial in the future to implement an aquatic environment waste detection system. By utilizing deep learning techniques the system will be able to analyze data from edge devices and make accurate predictions about the presence of trash under the water. After reviewing a plenty of papers, we have noticed that the other detection models such as YOLOv2, Inceptionv3, Mask-R CNN and CNN require a lot of time and provide less accuracy compared to YOLOv5 and YOLOv7. That's why we have chosen these algorithm models to analyze our dataset. We have also tried to implement a new algorithm called EfficientDet which is an object detection algorithm that combines efficiency and accuracy. It was introduced by Google in 2019. In this paper, we have prepared our self-prepared dataset which includes 1400+ pieces of data collected from various sources like St. Martin's Island, pond, roadside drains and so on. Then we have processed and train the data and run three algorithm models which includes YOLOv5 , YOLOv7 and EfficientDet and got accuracy of 97%, 93% and 96%. The challenges of our work were to detect the trashes in the polluted water where the light is not sufficient. Because in the deep of the sea or any water source where the light is minimal, the detection of trashes become difficult. We hope to enrich our dataset more in the future and aim to build a model using raspberry pi or arduino to use the progressive algorithm models by eradicating the challenges of underwater trash detection.

Keywords: Trash; Aquatic environment; waste; automated; Underwater; YOLOv7

Acknowledgement

First of all we want to thank our creator that our work was completed without any hindrance as we thought.
Secondly, we would like to thank our advisor Dr. Amitabha Chakrabarty sir who has advised and helped us every step of the way and provided mental support.
Last but not the least we would like to thank our parents who have trusted us and supported us and given us the freedom to do whatever we wanted to do and without their support we could not have done it we would not have reached this stage.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Research Problem

According to the authors [25], one of the main issues facing our daily lives and environment is waste management. The traditional way we manage our waste falls short in a few key areas, like rubbish overflows that pollute the environment and lead to unsanitary living conditions. The waters of the rivers in Bangladesh become polluted with huge trashes every day [1] .Electrical and electronic devices are now a necessity for many aspects of daily life, and their numbers are growing five times faster than those of people, such as mobile phones, which have gone from zero to 7.2 billion in just three decades [6]. A 5-10 percentage annual rise in the quantity of used electrical and electronic equipment that is properly disposed of could result in environmental dangers that are harmful to groundwater, marine life, human health, and soil fertility[13] .Water samples from sewage disposal of hospital waste discharge point were collected and detected multidrug resistant (MDR) bacteria in untreated waste water disposals of hospitals in Dhaka City, Bangladesh [3]. Recent research has shown that the availability of very young organic carbon in Bangladesh encourages the discharge of arsenic to groundwater [4]. In a heavily populated rural area with poor sanitation, this study investigates the possible contribution of human and livestock waste as a substantial source of this carbon.

1.2 Possible Solutions

A long-term economic and environmental issue, trash deposits in aquatic areas have a negative impact on marine ecosystems [27]. After researching different types of research papers we find out that there might be different types of solutions of trash detection in underwater. We might use a prototype rig that resembles an actual waste-collecting truck that has been constructed. On the other hand, for the classification job, convolution neural networks were developed and showed up to 98 percent accuracy.We could also use Autonomous underwater vehicles (AUVs) may also assist in the solution of this problem by identifying and eventually removing rubbish from the ocean [35]. Also can use YOLOv3 ,the detector YOLOv3 is effective. It is accurate and quick [7]. After exploring all types of solutions we are working on YOLOv7 which is very updated and newly published in 2022. YOLO uses convolutional nural network out of the box to model. It uses a single neural network to the image and divide the image into regions and predicts the bounding boxes and probabilities for each region. YOLO framework is good for detecting multiple objects in an image/video hence nothing good in predicting different classes in the image but also their actual location. YOLO is extremely quick and can be used in real-time. One of the most well-known model architectures and object detection techniques is called You Only Look Once (YOLO). The key reason for its popularity is that it makes use of one of the greatest neural network architectures to create high accuracy and overall processing speed.YOLOv7 can minimize roughly 40 percent parameters and 50 percent of computation of cutting-edge real-time object detections, as well as achieve faster inference speed and higher detection accuracy. On small data sets, it can be trained significantly more quickly without any pre-learned weights.

1.3 Research Objectives

Through our proposal system, we seek to identify objects in aquatic environments in Bangladesh from a prospective standpoint. Previously, many other projects connected to this work had been completed, but they did not work for the environmental conditions in our nation, so we worked on this. The main objectives are:
1. To detect trash using Deep learning model YOLOv7 ,YOLOv5 and EfficientDet.
2. To get the accuracy of detection of the trash dataset.
3. To implement the YOLO model for underwater trash detection.
4. To compare the outcome results for implementing YOLOv7, YOLOv5 and EfficientDet
5. Develop and train YOLO and EfficientDet models using a well-curated dataset of underwater images containing various types of trash.
6. Evaluate the performance of the trained models on a separate testing dataset in terms of accuracy, precision, recall, and F1-score for different trash categories.
7. Identify the strengths and weaknesses of each model in this specific application.
8. Talk about how the built system may be used in real-world situations to help divers clean up underwater settings, keep an eye on the condition of marine ecosystems, and create autonomous underwater vehicles (AUVs) that can gather rubbish.

Chapter 2

Literature Review

With the amount of trash in our oceans continuing to increase, underwater trash identification has become in importance. In the past few years, a number of experiments have been carried out to develop automated systems for identifying and categorizing underwater trash using various methods, including machine learning and computer vision. This paper [37] offers an approach for using the ResNet network algorithm as the convolutions layers based on the Faster R-CNN object detection framework. This method increases the accuracy of object detection and location and achieves the expected experiment results [33] .The network also shows its effective generalization ability when the small region objects occur. They are able to detect trash in metropolitan environments in close to real time and with great precision thanks to their data fusion method, which has significant practical significance. In order to achieve quick and high accuracy detection, it is still a problem to further lower the detection time [19]. According to this document, achievable targets are established for a rise in household garbage recycling during the ensuing years in light of the growing emphasis on protecting the natural environment worldwide. As the initial phase in a sustainable waste management system, consumer sorting is frequently used. They enable the categorization of trash bags throughout this process in order to help waste management systems in their understanding of consumer waste creation and behavior. With excellent precision, this idea system can detect the presence of glass and metal in consumer garbage bags. Their [5] technique can assist identify the locations where erroneous sorting occurs without making significant changes to the existing system when an RFID system is in place for the collection of municipal refuse.They were successful in creating a categorization system that combines audio recording and a metal detector with beat-frequency oscillations. With a 98 percentage accuracy rate, the trained CNN model can detect the presence of glass and metal in waste bags [24]. The paper discusses [29], the first of a series of abilities required for such AUVs, the difficulty of identifying garbage, particularly plastic debris, in an underwater environment, is examined in this article. They take into account several deep learning-based visual object identification algorithms, create a dataset to train and test them on, and assess their performance on this job using a variety of metrics. Their main concerns were whether real-time, DL-based visual identification of underwater rubbish was feasible and how current approaches fared in this regard. Marine debris detection is a challenging topic, to put it mildly. Small environmental changes can have a big impact on how an item appears, as is the case with many visual object recognition

issues. On the issue of locating maritime garbage, specifically plastic, they used and assessed four deep learning-based item detectors. From publicly accessible data,[30] they created a dataset and came up with a data model that had a single class for all garbage detections. They then trained YOLOv2, Tiny-YOLO, Faster R-CNN, and SSD using conventional fine-turning techniques, tested them on a test dataset made up of difficult items that had not been observed during training, and so forth.They conducted tests on three various devices, gathering data on the mAP and IoU for each model and network, as well as run-time speeds for all setups. They believe that garbage identification using visual deep learning models is feasible in real-time based on their findings [11].
In this research, a trash detection system is presented that can automatically identify and pinpoint rubbish spots in both videos and photos. For it to do this, a unique method was devised after reviewing the available options. When the outcomes of these models were evaluated for the suggested system, it was discovered that YOLOv5M had a promising precision value and offered trustworthy predictions. Also, this version of YOLO substantially exceeded earlier iterations with regard to speed, model size, and performance. Even though their method can identify junk in live video streams and photos in real-time, substantially reducing the prediction time remains difficult [28].
According to this paper [17], Plastic wrapper photos were gathered, and the YOLOv3 model was trained on these to produce the appropriate predictions. Using the Tiny-YOLOV3 model, they were able to get an accuracy of 87 percentage for the items. They have employed fewer layers, which trade off with less accurate forecasts, in order to make predictions more quickly. For the trained model, an android application site was made. By utilizing information collected from application portals, they have effectively developed a system that assists in implementing a profile of businesses that contribute to plastic pollution. Anyone who wants to contribute or share photographs of plastic garbage using their mobile devices can do so using this app site. The provided deep learning model may be further trained to recognize numerous brands of plastic items, expanding the range of possible applications.
This study describes methods for utilizing deep learning to find and categorize different underwater marine items. The targets of detection are used to classify approaches. The utilized features and deep learning architectures are compiled. Deep learning has been used more frequently to detect and classify coral, but little research has been done on seagrass, despite its importance to the maritime ecology. Combining both texture- and color-based data may considerably improve the efficiency, accuracy, and resilience of any detection and classification system. For the identification and classification of seagrass, the accumulation of hand-crafted features and neural networks may produce superior outcomes[31].
The paper [2] suggests a non-evasive approach in which an underwater drone, funded by the department back in July 2021, is used to collect visual data that will be georeferenced in order to analyse how macro-plastics are transported underwater, how much of that plastic is dumped in the bottom of the Trent River, and how those plastics affect the ecosystem. The multidisciplinary team working on this project will create and submit a research proposal to BBSRC/UKRI to further examine and comprehend how much plastic is carried underwater, how much of it deposits on the soil of Trent is River, and how it affects the ecology. The creation of a proof-of-concept validator is the aim of this study. If the project is financed, it

will be possible to set up the current underwater drone to meticulously gather and label data for academic staff post-processing. To gather the initial datasets, analyze them, and produce the preliminary findings that will be utilized to strengthen the proposal and submit it to the BBSRC/UKRI thematic calls on underwater plastic detection, a research assistant will assist the academic team.Main goal is to Create and execute a systematic method for employing an underwater drone to gather and label visual datasets.To better understand how plastic is moved underwater, how it accumulates in the soil at the bottom of the Trent River, and how it affects the ecology, process the dataset using ML/AI/statistical analysis methods.

It is challenging to search for and survey submerged debris on your own. Only unmanned aerial vehicles (UAVs) or satellite photos are capable of being used in remote sensing approaches to examine the surface. Sonar sensing is the most precise method for locating items under water, however it can be challenging to interpret sonar images, and small objects are particularly difficult to see in large-field of view sensors like Synthetic Aperture Sonar and Sidescan Sonar. Through the use of forward-looking sonar imagery and a convolutional neural network classifier on a sliding window spanning the sonar is field of view, they have demonstrated a novel method for identifying and classifying buried trash in maritime environments. Both a binary classifier that selects between debris and background and a multiclass classifier that chooses between background and one class of debris, such as glass, rubber, plastic, paper, and metal, were looked at.Their binary classifier-based debris detector generates detections that are 80.8 percentage accurate compared to a multiclass classifier-based detector is 70.8 percentage accuracy [20].

The paper [32] discusses a novel method for effective garbage detection and collection utilising machine learning and neural networks. The major goal is to identify garbage quickly and accurately, as this has a substantial impact on waste management and environmental preservation. The introduction emphasises the critical nature of waste management and the demand for more effective solutions. Accurate and quick rubbish detection is emphasised as a crucial stage in waste management in the essay. Technique used to achieve precise and quick garbage identification. The use of NN and ML algorithms for picture recognition is mentioned. To analyse waste item photos and categorise them into different groups, including recyclables, non-recyclables, and hazardous garbage, the system uses deep learning models. The acquisition of a substantial dataset of garbage photos, which is essential for training the neural network, is covered by the authors. In order to enhance model performance, they additionally emphasise the significance of data augmentation and pre-processing procedures.

Recently, Bangladesh has faced a significant difficulty with regard to trash management, which is negatively affecting the environment. The Automated Teller Dustbin (ATD), which has been given the moniker Automated Teller Machine (ATM) Dustbin, is proposed to be designed in this study together with an intelligent embedded system. An effective image classifier built on a convolutional neural network (CNN) is created, and it can analyse training features to detect and identify any object that is considered to be garbage [14].

The authors [34] stress the importance of accurately classifying macroplastic garbage (plastic items greater than 5 mm) in order to monitor pollution levels, understand its origins, and develop workable mitigation strategies. They emphasise how computer vision techniques using deep learning can automate the identification of macroplastic garbage in particular. Significant knowledge gaps are also discovered in the article, including the lack of models with strong generalisation capabilities, quantification techniques for plastic mass flow and hotspots, and monitoring strategies that are not sufficiently scalable. The authors call for more effective and scalable solutions to address plastic pollution in aquatic habitats by urging research emphasis on riverine ecosystems and data-centric AI strategies to fill these gaps. To find and classify pertinent publications, the researchers used a systematic strategy that included database searches, keyword selection, abstract reviews, and snowball searches. The studies were then examined based on many criteria, including the type of water bodies investigated, dataset specifics, computer vision tasks, use of data augmentation and transfer learning, generalizability of models, and performance assessment metrics. This organized technique provides useful insights into the state of the art in deep learning-based macroplastic litter detection research, providing as a useful resource for academics while highlighting significant research gaps and pointing the way in the right direction. After analyzing all the data from a thorough summary of the current state of knowledge regarding macroplastic contamination in aquatic habitats, it draws attention to an imbalance in the number of studies conducted on various types of water bodies, with a predominance of studies on marine habitats and less focus on rivers and lakes, which is related to historical research patterns. The authors stress the importance of additional research into the generalization capacities of deep learning models from regulated artificial environments to real-world circumstances. Some parts of the data also explores the variety of data sources, demonstrating the usage of several imaging tools such digital cameras, underwater cameras, UAVs, phones, and sonar, while highlighting the advantages and drawbacks of each. It is noted that dataset labeling is not uniform, and recommendations are made to increase detection precision by collecting higher-resolution data. Furthermore, in order to improve the performance of deep learning models, it is also stressed how important large-scale datasets, transfer learning, and data augmentation approaches are. Study also examines pertinent performance evaluation measures and touches on the crucial problem of model generalization under various settings.

Overall, it offers insightful information about the challenges and methods associated with using deep learning models to the detection of macroplastic trash in aquatic settings.

In this paper, authors [15], employs hyperspectral remote sensing to detect rubbish in real time across wide regions. Due to a paucity of available garbage detection datasets and training data, traditional supervised object detection algorithms are not appropriate for the purpose of identifying compounds that are not visible in natural photographs, hyperspectral imaging offers comprehensive spectral and spatial information. In order to monitor rubbish across wide regions, the paper offers a unique technique that combines hyperspectral picture classification with unsupervised item recognition. A public hyperspectral trash detection dataset was created as part of the research, and a multi-scale convolutional neural network (MSCNN) for classification is suggested. Results show that the strategy is effective at finding waste in a variety of environmental settings. Mentioning about methodology, unsupervised garbage detection and hyperspectral image (HSI) classification make up the two primary parts of the proposed algorithm. A multi-scale HSI classification network (MSCNN) is introduced in the section on HSI classification. MSCNN uses two multi-scale layers with various convolutional kernel sizes to efficiently extract spectral and spatial data. Each pixel is classified as background or trash at the pixel level. Selective Search is modified in the unsupervised garbage detection section to construct item suggestions on the binary garbage segmentation maps created by the HSI classification. Sizes of the identified items are used as confidence ratings, and Non-Maximum Suppression (NMS) removes detections that occur more than once. Due to the limited amount of labeled training data, this architecture offers a novel method for large-area garbage identification utilizing HSI data. To process the data, the Python programming language and PyTorch framework are used to accomplish the suggested solution. Key hyper parameters are specified during training, including base learning rate of 0.01, stochastic gradient descent (SGD) as the optimization technique with batch size of 100, momentum of 0.9, and weight decay set to 10-4. Training lasts for 100 epochs, with a 0.1-factor drop in learning rate occurring every 30 epochs. The MSCNN model uses dropout for regularization and has multi-scale convolutional layers. The Indian Pines and Pavia University datasets are used in the evaluation of hyperspectral image classification tasks, and the results reveal increased performance when compared to state-of-the-art techniques. Higher accuracy, average accuracy, and Kappa coefficients are produced by the method. The model also shows resilience for trash identification by attaining high recall and reducing false positives. Edge Boxes is a technique for unsupervised detection, although Selective Search outperforms it. The suggested approach is appropriate for use in real-world contexts since it is effective at down sampling data without significantly reducing accuracy. In short, this research describes a two- step strategy for aerial hyperspectral remote sensing-based large-area rubbish monitoring. MSCNN is used in the initial stage to classify HSIs, demonstrating its efficacy. The second phase employs Selective Search for unsupervised trash identification while addressing box size confidence concerns and implementing NMS. The system shows how to effectively monitor rubbish in practical settings, with the possibility to expand environmental monitoring in further studies.

Research into efficient detection techniques for underwater environments has increased significantly in response to the growing issue of marine waste, which is mostly made of plastic. Divers' ability to perform traditional visual inspections is constrained by issues with safety, expense, and range. Thus, automated methods for detecting trash are essential for keeping an eye on, charting, and cleaning up our waters. Thanks to developments in autonomous platforms, machine learning, and sensor technologies, research on underwater trash identification is rapidly progressing. Nevertheless, there are still issues with handling heterogeneous settings, variable objects, and limited data. In order to address the global issue of marine trash, future research should concentrate on building robust and efficient algorithms, fostering interoperable systems, and compiling comprehensive databases. The below table 2.1 contains the summary of the literature review:

Table 2.1: The Summary of The Literature Review

Reference	Dataset	Accuracy	Algorithm
[24]	Self-prepared dataset	98%	CNN
[11]	A large and publicly-available dataset	82.3%, 70.3%, 83.3% and 69.8%	YOLOv2 , tiny YOLO, Faster RCNN with Inception v2 and Single Shot MultiBox Detector (SSD) with MobileNet v2
[15]	A hyperspectral garbage custom dataset Shandong Suburb Garbage	98.82%	MSCNN
[32]	Self-prepared dataset	85.7%	YOLOv3
[21]	Urban garbage from sanitation department	89%	Faster RCNN, 2018
[12]	Self-prepared dataset	96%	CNN, 2019
[10]	Three classes of image from internet	Did not clear about their classification task accuracy	Inception V3, 2019
[22]	Self-prepared data	96.35%	Public Garbage Net, 2020
[8]	Web crawled	75.6%	VGG16, 2020
[18]	Beijing Municipal Garbage data	65.8%	Mask-RCNN, 2020
[16]	Crawled from web	93.2%	Inception V3, 2020
[23]	Huawei Garbage Classification Challenge Cup dataset	92.62%	Gnet, 2021; Improved MobilNetv3
[9]	Didn't mention about the source of dataset	90%	CNN
[26]	TACO open dataset	83%,95.2%,97.1%	Mask-RCNN, YOLOv4, YOLOv4-tiny, 2021

Reference	Dataset	Accuracy	Algorithm
[1]	Self-prepared dataset	nil	comprehensive data acquisition programme
[2]	Plastic Marine Pollution Global Dataset	nil	Visual Survey Protocol
[4]	Self Prepared Dataset	-	-
[13]	Self Prepared Dataset	-	-
[14]	Self Prepared Dataset	96%	CNN-based model ALexNet
[32]	Self Prepared Dataset	57.8%	SVM

Reviewing through the papers we realized that the reviewed papers mostly focused on open trash detection. There have been a less amount of papers were about underwater trash detection. On the other hand, very less amount of papers we got who have worked under the environment of Bangladesh and other South Asian countries. An in-depth understanding of trash detection in aquatic environments is growing from the literature, where scientists are often emphasizing how important it is to have effective monitoring and mitigation strategies. Studies often draw attention to the detrimental impacts of aquatic pollution and emphasize how important it is to locate and remove trash in order to preserve aquatic ecosystems. To close in these gaps and ensure a more complete and comprehensive approach to waste detection in aquatic ecosystems, future research is badly needed. These kinds of studies will not only increase our scientific understanding but also have a significant impact on the development of long-term solutions that safeguard water quality and biodiversity. Going through the literature review in the summary of the literature review table we can realize that most of the cases the researchers have tried to run the models in their customized self-prepared dataset. In the paper [24], we have got the most accuracy with about 98 percent where CNN based algorithm has been used. The paper [32] showed very less accuracy which is 57.8 percent among all the papers we have reviewed for the literature.

Chapter 3

Data Collection and Workflow

Due to the pervasive trash littering, Bangladesh's problem with water pollution in sea beaches and other tourist locations, especially those close to large rivers and seas has gotten alarmingly bad. The issue affects not only vacation spots but also rural and urban ponds and canals due to various sources of contamination. The type and volume of waste dumped in these water bodies has a direct impact on how polluted the water is. Waste produced in tourist areas frequently consists of a variety of non-biodegradable materials, including packaging, disposable objects, and plastic bottles. By disposing of their waste improperly, visitors who are occasionally ignorant of or unconcerned with the environmental consequences contribute significantly to pollution.Sadly, the picturesque riverbanks and sea beaches that draw tourists end up serving as dumps for a variety of pollutants. Tidal movements, which can carry the waste into open waters and worsen the contamination, can also have an adverse effect on these areas. However, industrial areas present a unique problem with respect to water pollution.Also in Bangladesh we mostly see industrial areas are not specifically designed away from living areas. Every industry is situated near river and that river is connected to other canal or flow beside a village or city.As a result whatever natural resource of water people's use are polluted by toxic chemicals, heavy metals, and other dangerous materials are frequently released into water sources by industrial discharges. Industrial waste can have a severe negative impact on aquatic ecosystems and deteriorate the quality of the water over time. Industrial pollution is frequently characterized by hidden contaminants that can be more subtle and difficult to manage than the more obvious waste found in tourist areas. Understanding the pattern of different waste we decided to collect our data from both tourist spots and also from rural and urban waste data from the water source in rural and urban places. We visited several tourist sites during our journey as we attempted to compile an extensive dataset. Regretfully, our initial efforts to locate the specific image required for our project proved fruitless. On our first outing, we visited Dhaka's 39 serene lakes, where we planned to capture some quality images. Unfortunately, because the generated photos were deformed and useless for our needs, the outcomes were not perfect.Then we put artificial light to get our desirable clear data to run a detection model. After analyzing our video data we're not satisfied enough to generate a photo from that video. So we collect

samples and the condition of water clarity and make an artificial environment on a big tub to create proper light to collect enough recognizable photos. In the end, we chose to travel to Cox's Bazar and Saint Martin, which we thought would be

the perfect locations for our endeavor. When we got there, we started our investigation, concentrating on marine debris because it is such a widespread problem in these places. Our initiative investigates the murky waters of pollution by analyzing marine debris from various locations. We first faced constraints with distorted lake photos and devised a controlled setting to simulate dirty ponds. Our hardware is easily accessible GoPro Hero 8 Black, which is particularly durable for underwater use and has superb regulation at our maximum availability, allowing us to capture 1080p footage at 60 frames per second. After capturing the underwater videos, we began spitting out still photos of waste, which became our faithful friend.We retrieved over 2000 photos from this film by using Vlc media player image 600 height /600 weight at 25 epoch, demonstrating the striking contrast between contaminated rural and urban natural water resources, as well as ocean waters. Here, we have meticulously gathered a variety of waste products that are frequently found in aquatic settings and range in size and type. With great care, we put together a variety of plastic bottles with different sizes and forms. Additionally, we diligently collected the most harmful types of trash, such as plastic bags, steel and tin cans, and technological garbage. However, there were obstacles along the way. We faced a variety of challenges, including overexposed shots in direct sunlight, photos taken in less-than-ideal lighting and underwater images that lacked the requisite level of sharpness.To tame these photographs, pre-processing processes like noise reduction and color correction become your allies, combating the noise and color shifts induced by water. You'll then sharpen the unclear depths and correct any distortions for a sharper picture. Furthermore, retaining stability while battling the waves' constant ebb and flow proved a significant problem. Nevertheless, we overcome these challenges and acquired the crucial graphics for our dataset because of our everlasting dedication to our purpose. As we could get better results we rebuilt our dataset for the YOLOV7 and v5 model efficiently. Waste is annotated in YOLO PyTorch format.We perform image augmentation by adding noise and blurriness to the images, which helps the model to learn to be more robust to variations in the underwater environment. To create 3 versions of each source image we use the following augmentation. Next, we label the images with bounding boxes around the underwater waste objects. We have annotated dataset with xml format for run YOLOv7 and v5 model.We divided our waste objects into 4 classes which are :
1. Electric waste
2. Plastic bag
3. Plastic bottle
4. Steel and teen case
We use 2000+ images for training.We will only modify one of the YOLOv7 and V5 training defaults in our example: Then We train the YOLOv7 and V5 model on the training set using value modification on specific hyperparameters.

3.1 Between Noise And Without Noise

When compared to its noise-free equivalent, the model's overall accuracy is unquestionably impacted by the presence of noise in our dataset. The recall metric clearly outperforms the data with noise in the absence of noise.The precision curve for all classes in the absence of noise triumphantly achieves a flawless score of 1.00, accompanied by an exceptional precision value of 0.880, illuminating this distinction. In sharp contrast, the precision curve for all classes noticeably lags behind when noise is added to the dataset. Additionally, the noise-free dataset's precision-recall curve, which attests to its effectiveness, stands admirably at 0.923 mAP@0.5. Given the unique environmental circumstances, such as the frequency of turbid and sediment-filled rivers in Bangladesh, this is particularly remarkable. It is important to recognize that the geographical context has a considerable impact on the appropriateness of dataset noise. The use of data with noise appears to be the more wiser option in the context of Bangladesh, where aquatic habitats are often characterized by muddy waters full of silt and sand particles. The use of noise-free data is still the favored strategy and yields greater results in areas with clearer waters, such as open oceans and diverse geographical locations. Data with noise stands out as the best option in the real world of Bangladeshi settings because it is sensitive to the peculiarities of the local aquatic environment. In this situation, noise-free data is fundamentally unsuitable for Bangladeshi conditions' requirements for precision[Figure 7.9].

3.2 Data Augmentation

It is challenging to distinguish between underwater photos since the backgrounds of all the pictures are similar. We are unable to find any differences. These elements contribute to the training model's ineffectiveness. To make the deep learning model widely applicable, the dataset has to be updated and expanded. The primary method of data augmentation is rotational zooming. As a bootstrapping technique, augmentations can be used to generate more training data. Using Roboflow, we can do augmentations at the picture level (rotation,
shear, gray scale, hue variation, saturation, brightness, exposure, noise, cuts, and mosaics) as well as the bounding box level (crop, brightness, exposure, blur, and noise). The idea is to simulate real world variations of changes in lighting conditions, occlusion etc. in order to reduce overfitting and gain a better bias- variance tradeoff. Other augmentations include copying and pasting one image's training bounding box contents to other images as well as resizing the images within the bounding boxes. We use the following data augmentation to enhance our model performance.

- Random Scaling,
- Rotation,
- Gaussian Blur,
- Gaussian Noise,
- Color Correction,
- Contrast Enhancement,
- De-noising,
- Sharpening,

- Geometric Correction,

Prior to training, the algorithm must be set up to recognize trash objects underwater. In accordance, the configuration was completed. The YOLO algorithm has been configured to identify trash objects in an underwater environment. There are four classes, or the number of objects the algorithm should be able to identify, as

1. Electronic waste

2. Pastic bag

3. Pastic bottle

4. Steel and teen case

The dataset which is used in this thesis is divided into 80% as training set 20% as testing data each time it runs the split so that the entire data can be used. The learning rate is 0.001 which is used to train the algorithm.followings are the modifications we change in the default model.

Values:

Table 3.1: Values

Input Size	416x416
Learning Rate	0.001
Batch Size	16
Number of Epochs	150
Number of Classes	4

For YOLOv5 model Hyper parameters we use to detection efficiency shown following,

Table 3.2: Hyperparameters

Optimizer	Adam	
lr0	0.01	initial learning rate (SGD=1E-2, Adam=1E-3)
lrf	0.2	final OneCycleLR learning rate (lr0 * lrf)
momentum	0.937	SGD momentum/Adam beta1
weight decay	0.0005	optimizer weight decay 5e-4
warm up epochs	3.0	warm up epochs (fractions ok)
warm up momentum	0.8	warm-up initial momentum
warm-up bias lr	0.1	warm-up initial bias lr
box	0.05	box loss gain
cls	0.5	cls loss gain
cls pw	1.0	cls BCELoss positive weight
obj	1.0	obj loss gain (scale with pixels)
obj-pw	1.0	obj BCELoss positive weight
iou-t	0.20	IoU training threshold
anchor-t	4.0	anchor-multiple threshold
anchors	3	anchors per output layer (0 to ignore)

Five-fold cross validation is the procedure used to reduce errors brought on by randomness. The five-fold cross-validation process divides the data into five folds; the model is trained on the (5-1) folds, with one fold remaining for testing. Five times, this process is carried out. As training and test sets of the data are separated. Cross-validation is carried out using the training dataset. This method is used to estimate the accuracy of a model. For every model, the five-fold cross-validation is conducted fifty times. The confidence interval can be calculated using the repeated cross-validation results after 50 repetitions of the cross-validation.This method helps to account for variations in the data split and offers a more thorough assessment of the model's performance. Cross-validation accuracy is determined by taking the average of all repetitions, and 95% confidence intervals are computed using the outcomes of the repeated cross-validation. The testing accuracy range is taken into account when conducting the final evaluation of the model. The model is considered acceptable if the testing accuracy is within 95%; otherwise, appropriate overfitting or underfitting must be made. The cross-validation procedure does not make use of the testing dataset, which is an isolated dataset. Thus, the 5-fold cross-validation yields the performance metrics.

3.3 Validation And Testing

Validation is the post-training procedure where a testing data set is used to assess the trained model. Ensuring the accuracy of the algorithm's outputs is crucial for enhancing both their quality and quantity. It is incorrect to depend on the model's predictions without first validating it. Accurate predictions are produced by the machine learning algorithm after validation. The following are some benefits of model validation:

- Scalability and flexibility

- Reduce the costs

- Enhance the model quality

- Discovering more errors

- Prevents the model from over-fitting and under-fitting.

Testing is done by performing experiments by following distinct parameters and measuring the performance utilizing key metrics on an evaluation subset of the complete dataset.

- Precision,

- Recall,

- mAP

The pre-processing steps are applied to the testing images that are fed into the engine acquiring exemplar. This intends that several post processing may be needed to assess the exemplar and analyze the outcomes of the detection on the unique image.we use some post-processing method for our dataset which is

Loss Function: We have used YOLO's default combination of localization loss, confidence loss, and class loss.

Fine-Tuning: We have fine-tuned the model on underwater waste dataset to improve performance The test images, which were chosen for a test set, were gathered for evaluation and included samples of each class in our model. the video network objects and images that shouldn't be used for training. Thus, it guarantees that the training and test sets do not overlap. Various kinds of plastic items, such as grocery bags and plastic bottles, were sorted and chosen. For each of these kinds of objects, we made sure to gather a minimum of 20 images per class.
For testing, we gathered a sample of these annotated photos. The photos served as models for each class, showing a range of settings that were purposefully gathered and chosen to present difficulties for detectors. to ensure we provide a realistic evaluation of how these detectors would perform in field conditions. We also run YOLOV7 on our dataset to see the performance and accuracy by modifying the model into following values.

Values:

Table 3.3: Values

Input Size	416x416
Learning Rate	0.001
Batch Size	16
Number of Epochs	150
Number of Classes	4

YoloV7 is used for real time video object detection. For v7 model Hyperparameters we use to detection efficiency shown followingly,

Table 3.4: Hyperparameters

Optimizer	Adam
Momentum	0.937
Weight-Decay	0.0005
Warm-up epochs	3
Warm-up momentum	0.8
Warmup-bias-lr	0.1
Box	0.05
Cls	0.3
Cls-pw	1
Obj	0.7
Obj-pw	1
Iou-t	0.2
Anchor-t	4
F1-gamma	0
Hsv-h	0.015
Hsv-s	0.7
Hsv-v	0.4
Degrees	0
Translate	0.2
Scale	0.5
Shear	0
Perspective	0
Flipud	0
Fliplr	0.5
Mosaic	1
Mixup	0
Copy-paste	0
Paste-in	0
Loss-ota	1

Loss Function: We have used YOLO's default combination of localization loss, confidence loss, and class loss.
Fine-Tuning: We have fine-tuned the model on underwater waste dataset to improve performance.

Finally we also test our dataset by another model which is EfficientDet.We have used a pre-trained EfficinetDet (d-0) model on a large dataset (COCO) for faster convergence.For running this model we modify following defaults values:

Table 3.5: Values

Model Variant	Choose a specific EfficientDet variant (EfficientDet-D0)
Input Size	640x640
Learning Rate	0.001

To evaluate this model We have used relevant metrics (precision, recall, mAP) to evaluate the model's performance.For v7 model Hyperparameters we use to detection efficiency shown followingly,

Table 3.6: Hyperparameters

model-name	efficientdet-d0
ckpt	efficientdet-d0
train batch size	16
eval batch size 16	16
num-epochs	100
hparams	num-classes=4,moving average decay =0.9998
learning rate	0.0025
lr warmup init	0.001
optimizer	d
num-examples per epoch	5000
num-epochs per eval	5
min eval interval	180
momentum	0.9
weight-decay	4.00E-05
warmup epochs	3
warmup-momentum	0.9
warmup-bias-lr	0.1

To evaluate this model We have used relevant metrics (precision, recall, mAP) to evaluate the model's performance.

3.4 Underwater Trashes

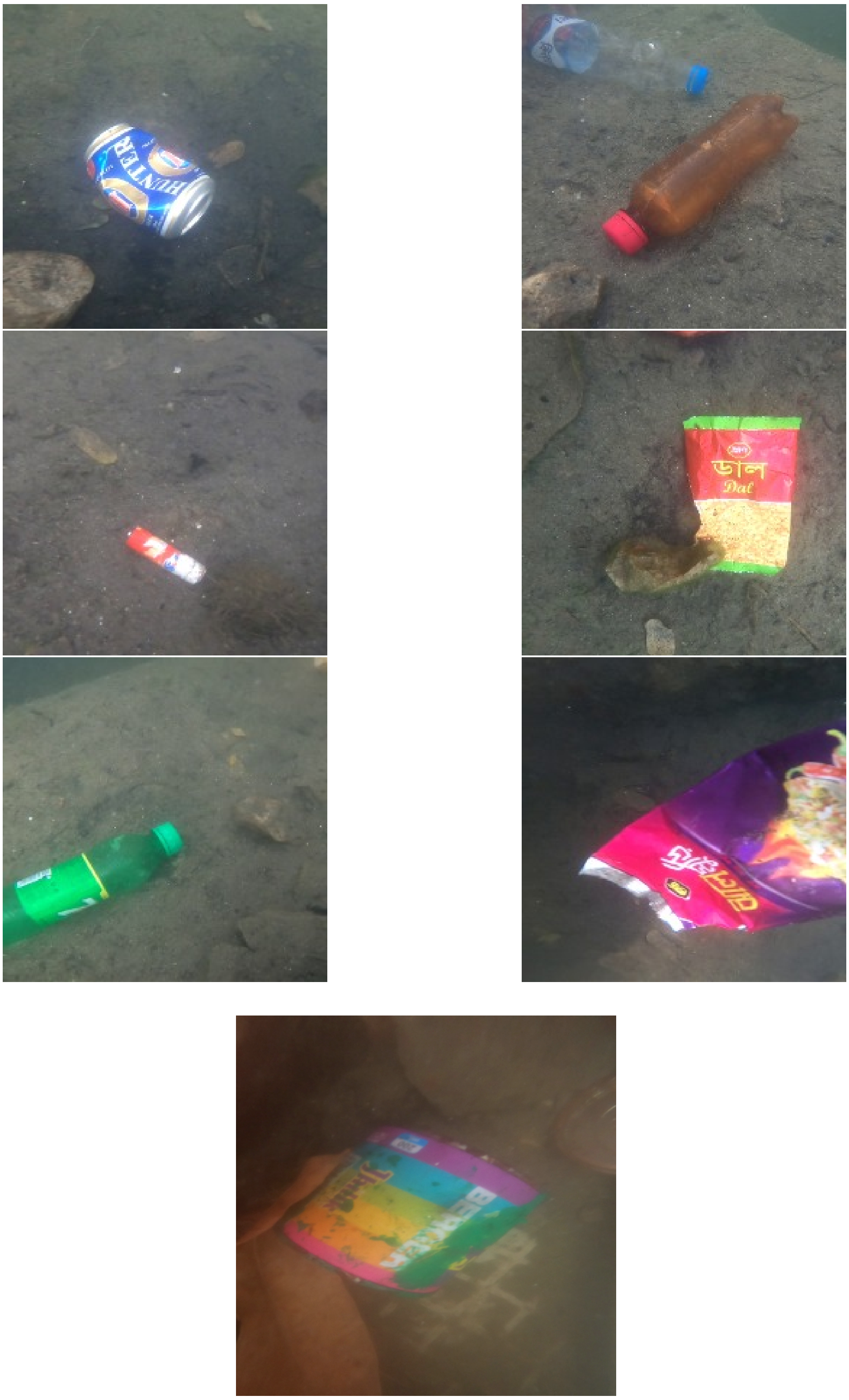

Figure 3.1: Collected Trash picture of four classes: Plastic Bottle, Steel and tin cans

3.5 Workflow

It is our workflow that allows us to explain step-by-step how we do our work. First of all,when we start our work, we have processed our database for augmentation and the rest of the work,since our data set is collected by ourselves. Then what we did was split our data set after that we trained our model , then hyperameter tuning and we did the rest of the process.

Figure 3.2: WorkFlow

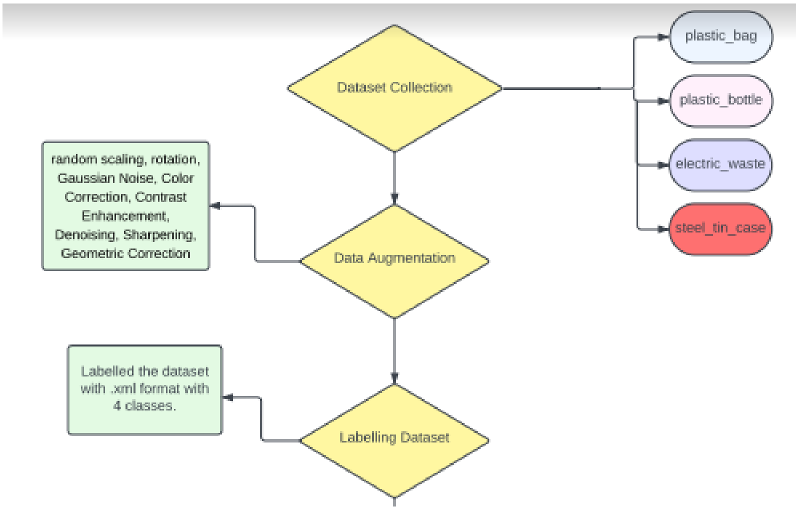

Figure 3.3: Pre Process WorkFlow

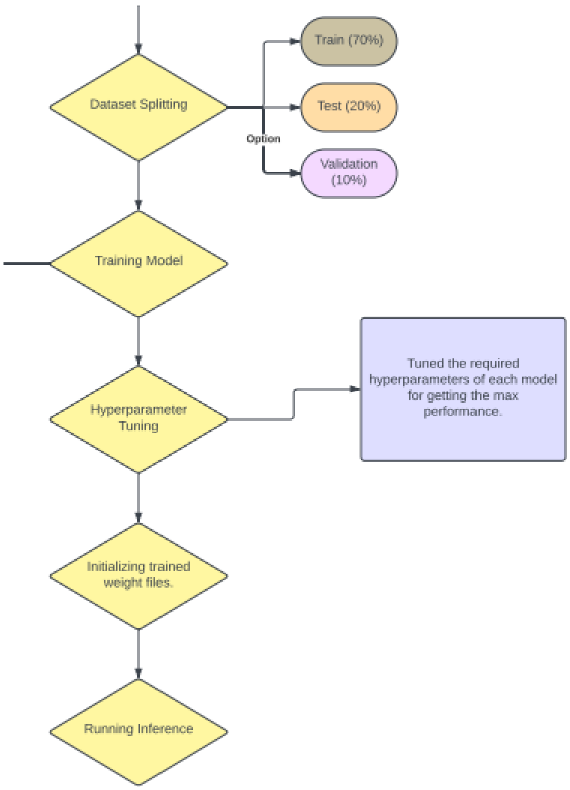

Figure 3.4: Post Process WorkFlow

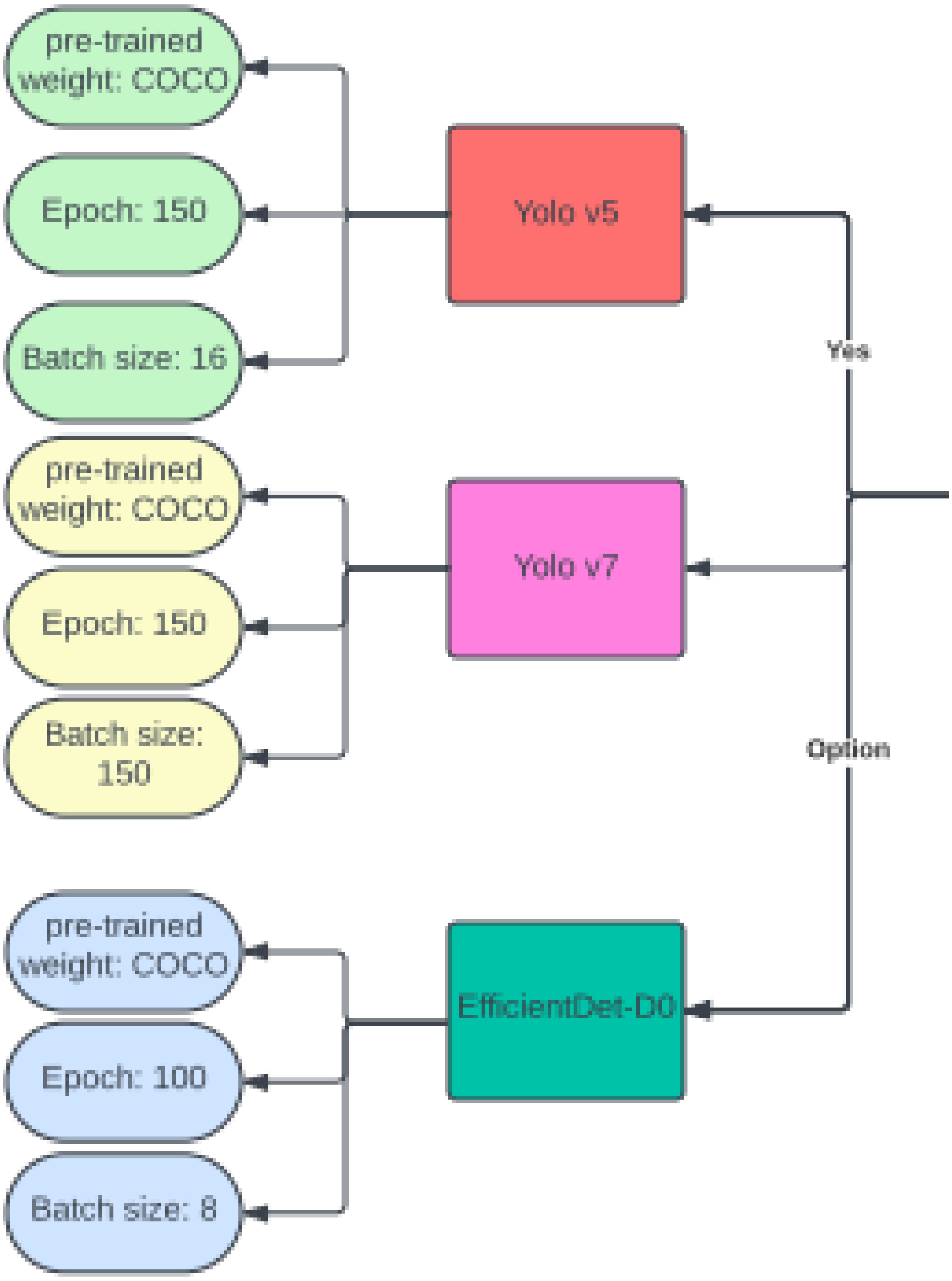

Figure 3.5: Model WorkFlow

Chapter 4

How YOLOv7, YOLOv5 And EfficientDet Works

We have applied the three models : YOLOv7, YOLOv5 and EfficientDet for these reasons:
YOLOv7 and YOLOv5: YOLO models are easy to implement and deploy since they are comparatively straightforward and effective. For some applications, they offer a fair balance between speed and accuracy. They are well-liked options for situations when real-time detection is essential, like in video surveillance or autonomous cars.
EfficientDet: The goal of EfficientDet is to offer effective and efficient item detection at various scales. By employing a compound scaling approach, it seeks to strike a good compromise between computational efficiency and precision. EfficientDet is a flexible option for a variety of computer vision applications because it has demonstrated strong performance on a broad range of tasks and datasets.

Other well-liked object detection models are SSD (Single Shot Multibox Detector), Faster R-CNN, RetinaNet, and InceptionV3. The selection of a model is contingent upon the particular requirements of the application, as each model possesses advantages and disadvantages.

4.1 YOLOv7 Model

RepConv is used in YOLOv7's planned re-parameterized convolution architecture in place of identity connections (RepConvN). By replacing a convolutional layer with residual or concatenation with re-parameterized convolution, the intention is to prevent the occurrence of an identity connection.

4.2 Architecture Explanation

Extended Efficient Layer Aggregation Network, or E-ELAN, is how YOLOv7 modifies the ELAN architecture [36] . To enhance model learning while preserving gradient flow routes, ELAN employs the cardinality operations of expand, shuffle, and merge. Regarding the architecture, it solely makes changes to the computational block, leaving the transition layer's design as ELAN. Group convolution is used by

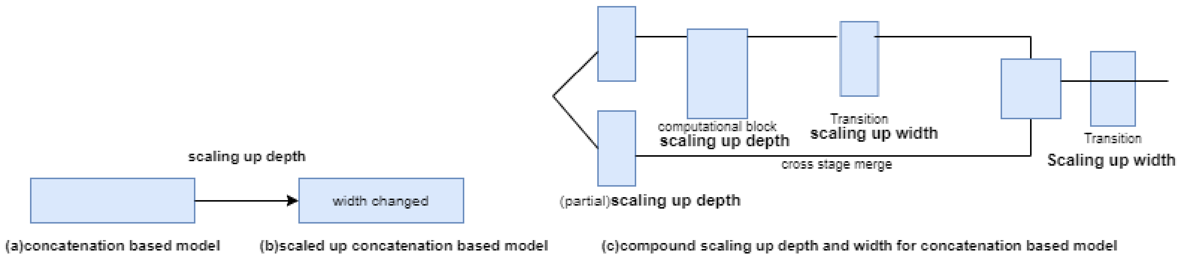

Figure 4.1: Breakdown of how YOLOv7 works

E-ELAN to increase the computational block's cardinality and number of channels. All the computational blocks in a computing layer are subject to the same channel multiplier and group parameter. Each computing block's feature map will be divided into groups of size g, concatenated, and then combined. In order to perform merge cardinality, a group feature map will be added and shuffled.

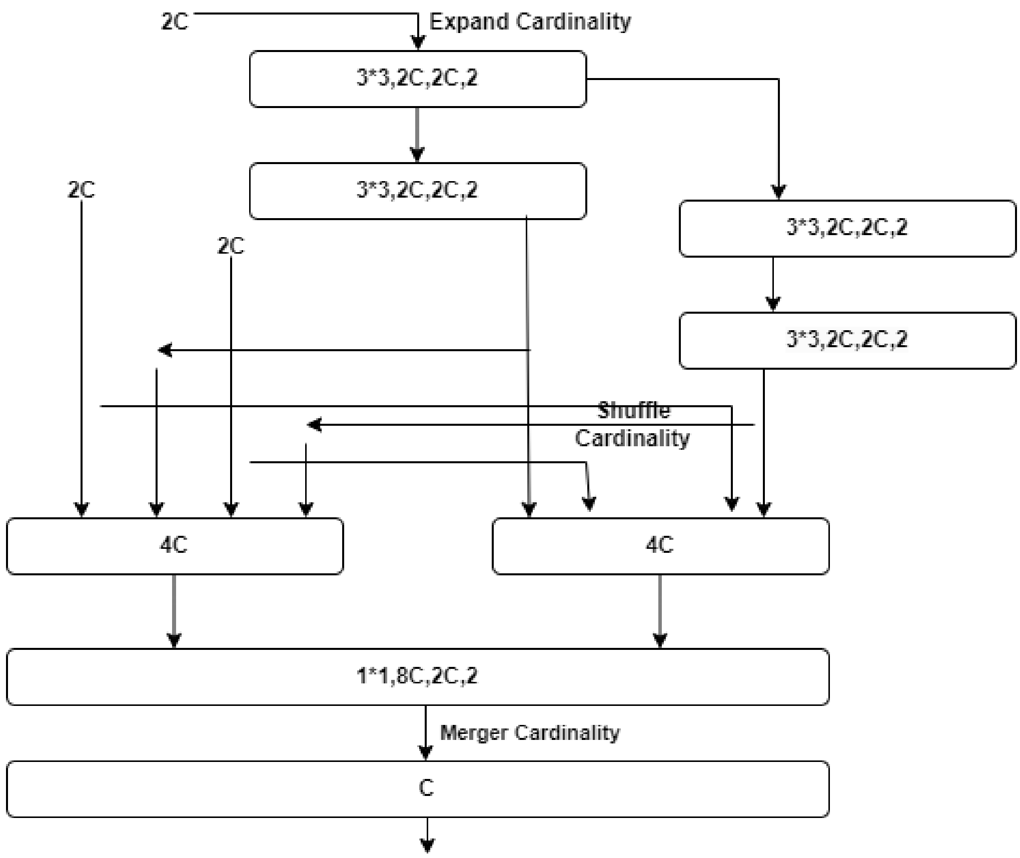

Figure 4.2: YOLOv7-Architecture

4.3 Model scaling

Model scaling is a crucial idea. Model scaling can be used to increase the depth, image resolution, and width of the model. Increasing or decreasing the number of layers in the model is referred to as scaling depth. In the same way, scaling the number of channels in the model architecture corresponds to breadth. The model architecture files specify scaling factors for depth and width. Because YOLOv7's architecture is concatenated with the other layers, we must determine how the output kernels will vary when we scale the depth parameter of a computational block. Then, taking into account scaling width by the computed change in kernels. Consequently, this compound scaling strategy will maintain the characteristics of the architecture.

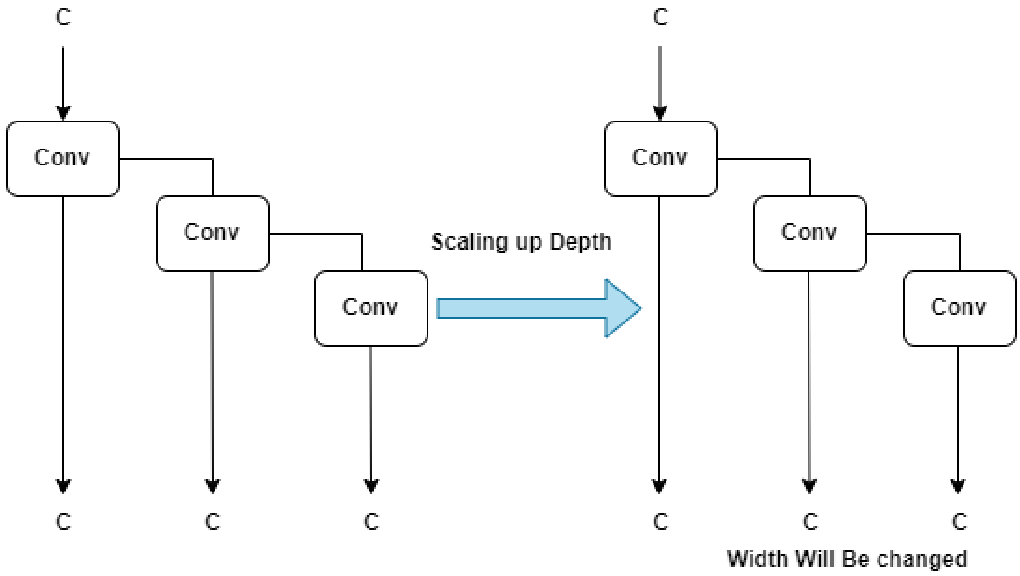

Figure 4.3: Model Scaling

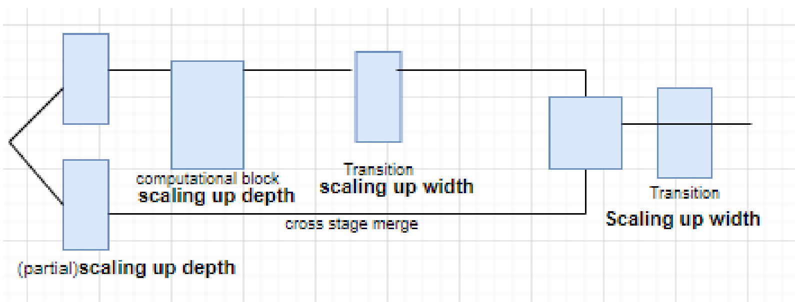

Figure 4.4: Model Compound Scaling

4.4 YOLOv5 Model Explanation

The object detection methodology known as YOLOv5 (You Only Look Once version 5) was introduced by Ultralytics. YOLOv5 adds a number of enhancements while extending the capabilities of the YOLO family of real-time object detection systems [38] . An overview of the YOLOv5 architecture is given below:
Backbone System: The backbone architecture of YOLOv5 is CSPDarknet53, an expanded variant of Darknet. The Cross-Stage Partial (CSP) design facilitates feature and gradient propagation over the network. Multiple convolutional layers and downsampling techniques are used in CSPDarknet53 to extract features from the input image. High-level semantic information is recorded by the backbone.

Neck: A PANet (Path Aggregation Network) neck, which facilitates the fusion of feature maps from various stages of the backbone network, is a component of YOLOv5. This is essential for managing items of different sizes and enhancing the model's capability to represent spatial relationships.
Head of Detection: The feature maps are sent to the YOLO detection head after the neck. For each grid cell, the detecting head, which consists of several convolutional layers, forecasts the following:
Bounding Box Coordinates: For any object included within a grid cell, YOLOv5 predicts the bounding box coordinates (x, y, width, and height).Indicating the likelihood that an object is present in the grid cell, it forecasts an objectness score for each object.
Class Probabilities: For each class of objects that the model has been trained to recognize, YOLOv5 forecasts class probabilities. Boxes for anchors: Anchor boxes, which are predefined bounding boxes with different aspect ratios and scales, are used by YOLOv5. These anchor boxes aid the model's ability to forecast the shapes and sizes of various objects. The sensor head is adjusted.
Post-processing YOLOv5 uses non-maximum suppression (NMS) to filter out redundant and low-confidence detections after creating predictions for each grid cell. Only the most confident bounding boxes are kept after NMS eliminates duplicates.
Results: The YOLOv5 program's final results are a list of objects that have been detected, each of which is represented by a bounding box (x, y, width, and height), an objectness score, and class probabilities. The model can recognize several objects in the input image and categorize them into various groups.
YOLOv5 is meant to be quick and effective, making it appropriate for tasks requiring real-time or nearly real-time object detection. It has grown in favor in a number of computer vision applications, including autonomous driving, surveillance, and object recognition. It is renowned for striking a compromise between speed and accuracy.

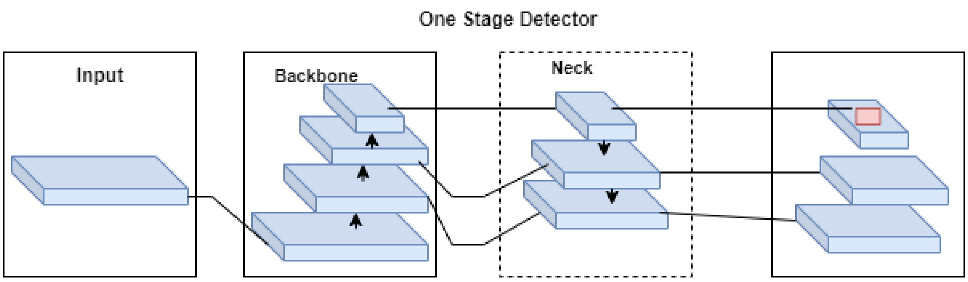

Figure 4.5: Breakdown of how EfficientDet works

4.5 EfficientDet Model Explanation

The object detection model EfficientDet is renowned for its excellent accuracy and efficiency. It is predicated on the EfficientNet architecture, which aims to combine computational efficiency with cutting-edge performance.
In order to maximize speed, the EfficientDet model balances the network's depth, width, and resolution using a technique called compound scaling. It is composed of an object detection feature network and a backbone network (like Efficient Net).

4.6 EfficientDet Architecture

High object detection performance is maintained while model size and computational economy are balanced in the EfficientDet architecture. It makes use of a technique called compound scaling, in which many facets of the model architecture are scaled concurrently to provide the best possible trade-offs between efficiency and accuracy. Key components and characteristics of this algorithm are: **Backbone Network**

(EfficientNet): The EfficientNet serves as the backbone network for EfficientDet. The EfficientNet model family attains cutting-edge results in image classification assignments. In order to achieve a suitable balance between model size and accuracy, it offers a compound scaling method that scales the model in terms of depth, width, and resolution.

Feature Pyramid Network (FPN): A Feature Pyramid Network, which is incorporated into EfficientDet, aids in the capture of multi-scale features from various backbone network levels. Managing objects of different sizes in the supplied image requires this.

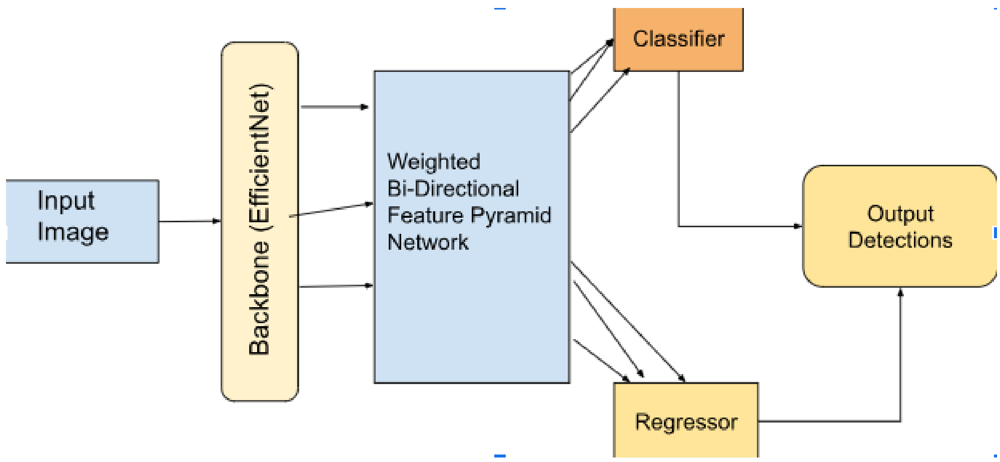

Figure 4.6: EfficientDet Architecture

Chapter 5

Implementation And Result

The implementation is carried out in two parts:

- Training

- Testing

The following is the procedure to perform training the dataset to detect the objects.

Training: Previously we are done with preprocessing the data, 3 folders are created while we complete our data for the train selected model. We will upload these folders to Google colab (Colab allows anybody to write and execute arbitrary python code through browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no set up to use while providing access free of charge to computing resources including GPUs).Here we use a pre-trained YOLOv5 model on a large dataset (COCO) for faster convergence and give path to folders and train the model and it finishes and gives us accuracy. Once we give the pictures to trained model we will receive our results showing what kind of material. For this we need to perform data augmentation.
Because of clarity and image quality to enhance detection model performance and accuracy augmentation helps significantly.
Performance Evaluation: Once the model is trained, we evaluate its performance on the test set. We measure the model's performance using these methods.
• Mean Average Precision (mAP)- Models for object detection and instance segmentation are assessed using a metric called Mean Average Precision (mAP). It evaluates how well these models locate and identify things in an image.The average precision across all classes or categories of objects in a dataset is determined by mAP, to put it simply. By contrasting the projected bounding boxes with the actual (annotated) bounding boxes, it evaluates how effectively the model recognizes objects. The average precision determined for each class is what is meant by "mean" in mAP. This value is then averaged over all courses to get an overall performance score. Greater accuracy in object detecting tasks is indicated by higher mAP values.

• Precision -The ratio of true positive predictions to the total of true positives and false positives is how it is computed:

$$Precision = (TruePositives/TruePositives + FalsePositives) \tag{5.1}$$

A high precision indicates that the model is typically accurate when it predicts a good outcome.

and Recall-The ratio of true positive predictions to the total of true positives and false negatives is how it is computed:

$$Recall = (TruePositives/TruePositives + FalseNegatives) \tag{5.2}$$

A high recall rate means that the majority of the positive occurrences in the dataset can be found by the model.

• F-1 score-Recall and precision are combined into a single, balanced rating called the F-1 score. It is especially helpful when you want to take memory and precision into account simultaneously without giving one too much weight to the other.The F-1 score is defined as follows:

$$F1 = 2 * (PrecisionRecall/Precision + Recall) \tag{5.3}$$

It ranges from 0 to 1, with 1 being the best possible F-1 score, indicating flawless precision and memory. The F-1 score is computed as the harmonic mean of precision and recall. When looking for a fair evaluation of a model's performance, this statistic comes in handy, particularly when weighing the trade-off between recall and precision.

• Accuracy

Then we decide to run another model EfficientDet which brings a balanced result for object detection. We'll run our image data into that model by modifying defaults following values:

And we measure the model's performance using parameters same as YOLOV7 and YOLOV5.

Figure 5.1: Detection Sample1

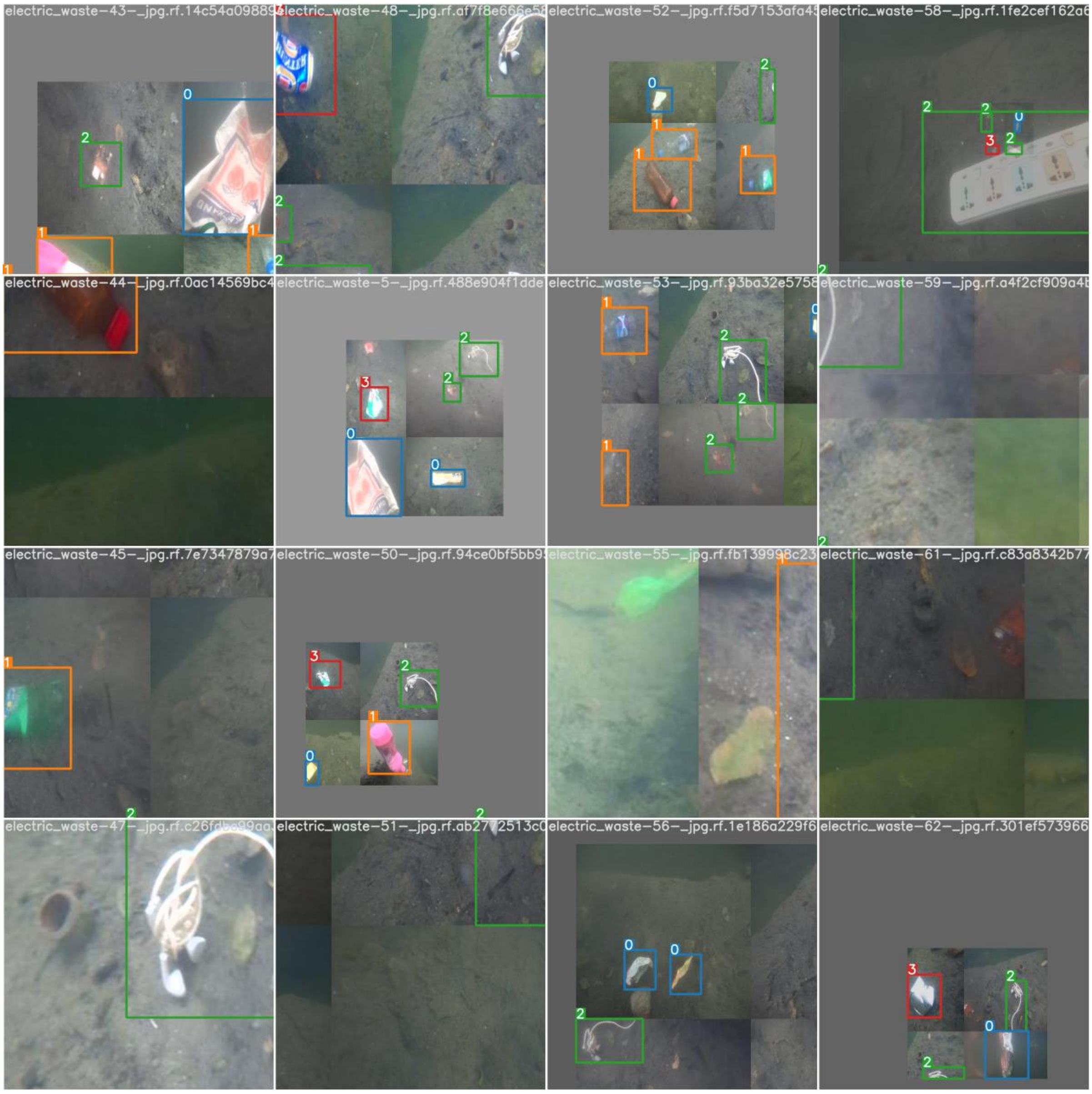

Figure 5.2: Detection Sample2

Chapter 6

Result Analysis

6.1 YOLOv5 And YOLOv7 Result

YOLOv5 demonstrates superior processing efficiency when handling custom training datasets as compared to YOLOv7. It showcases an ability to swiftly process and adapt to specialized data sets, thus streamlining the training process. YOLOv5 exhibits enhanced efficiency in interfacing with central processing units (CPUs) when compared to YOLOv7. This implies that it can efficiently harness CPU resources, resulting in expedited execution on these hardware configurations. YOLOv5 exhibits enhanced efficiency in interfacing with central processing units (CPUs) when compared to YOLOv7. This implies that it can efficiently harness CPU resources, resulting in expedited execution on these hardware configurations. The F-measure, with infinity set to 1 (F1 score), harmoniously balances precision (P) and recall (R) within a classifier. Both metrics are assigned equal importance. In the graph provided, the confidence value that optimizes precision and recall is 0.503, corresponding to the maximum F1 value of 0.90. In most scenarios, a higher confidence value and F1 score are sought after. In our F1 curve, it's evident that for all classes, YOLOv7 outperforms YOLOv5. YOLOv7 achieves an F1 value of 0.94 at 0.476[figure 7.1], while YOLOv5 reaches 0.90 at 0.217[figure 7.2].

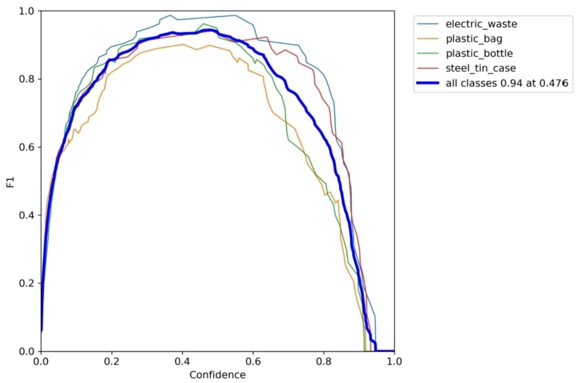

Figure 7.1: F1-Confidence Curve v7

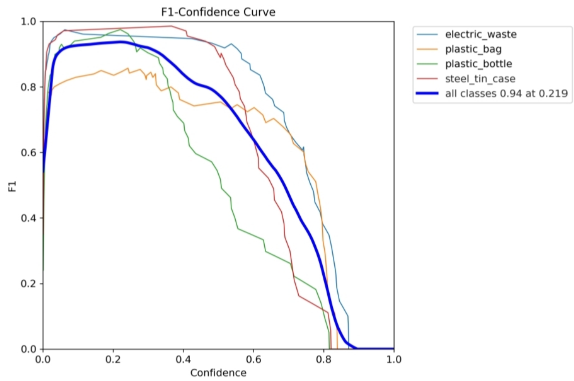

Figure 7.2: F1-Confidence Curve v5

Precision Curve: Precision represents the ratio of accurate positive classifications (true positives) to the total number of predicted positive classifications (true positives + false positives). In YOLOv7, all classes display precision values ranging 1.00 at 0.842[figure 7.5], while in YOLOv5, precision values span 1.00 at 0.768[figure 7.6]. YOLOv5 showcases commendable precision values.

Recall Curve: Recall signifies the ratio of accurate positive classifications (true positives) to the total number of true positive instances (true positives + false negatives). YOLOv7 achieves a recall of 0.99 across all classes at 0.000[figure 7.7], whereas YOLOv5 attains a recall of 1.00 for all classes at 0.000[figure 7.8]. YOLOv7 demonstrates superior recall values compared to YOLOv5.

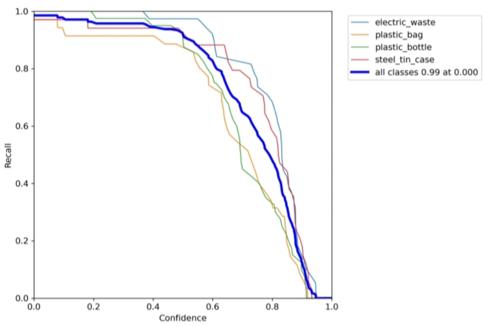

Figure 7.7: R-curve v7

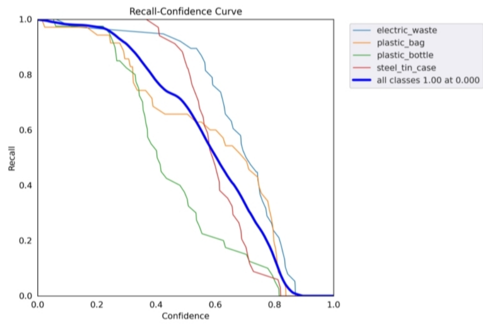

Figure 7.8: R-curve v5

Precision-Recall Curve: Precision represents the ratio of accurate positive classifications (true positives) to the total number of positive predictions (true positives + false positives). Recall, on the other hand, is the ratio of accurate positive classifications (true positives) to the total number of true positive instances (true positives + false negatives). A Precision-Recall (PR) curve is a graphical representation with precision values on the y-axis and recall values on the x-axis. In YOLOv7's precision-recall curve, all classes exhibit a 0.959 mAP@0.5[figure 7.4], while YOLOv5's precision-recall curve yields 0.970 mAP@0.5[figure 7.5. In terms of precision-recall curve, YOLOv5 delivers a superior outcome.

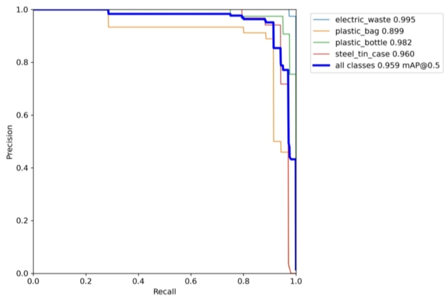

Figure 7.3: PR-curve v7

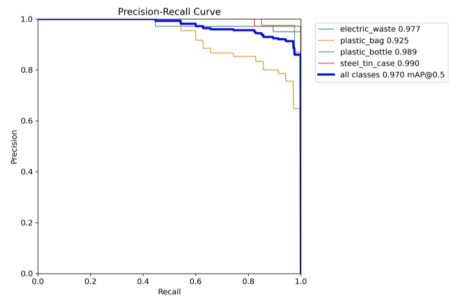

Figure 7.4: PR-curve v5

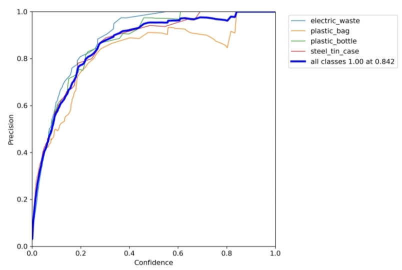

Figure 7.5: P-curve v7

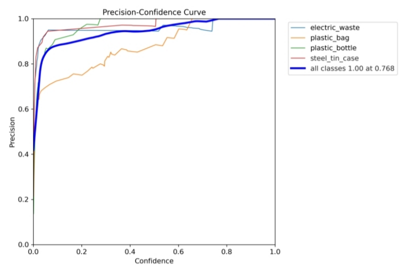

Figure 7.6: P-curve v5

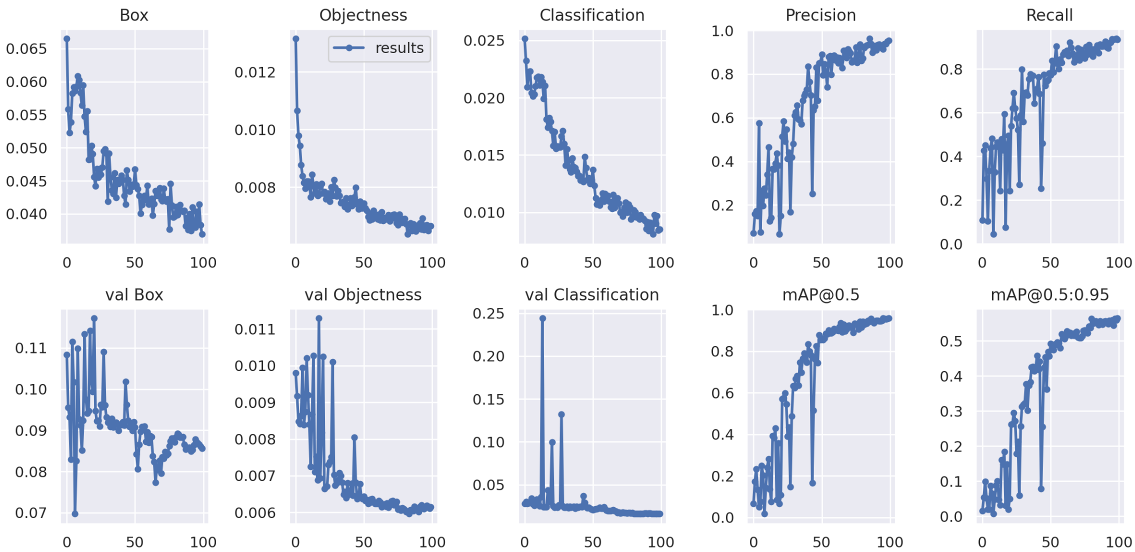

Figure 6.1: Noise-Without noise Result

EfficientDet Result: Metrics at IoU Thresholds: This is the most overlap needed for a detection to be considered accurate between the predicted and ground truth bounding boxes or masks.
Precision: It is the proportion of all anticipated positive observations to accurately predicted positive observations. For example, 92% of the model's predictions at an IoU threshold of 0.5 are accurate when it comes to predicting a bounding box with a 0.5 IoU overlap with the ground truth.Same at 0.6,0.7 threshold 90% ,88% predicting bounding box.
Recall: This measures the proportion of accurately anticipated to actual positive findings. The model captures 94%, 92%, and 90% of the actual items in the image with an IoU threshold of 0.5, 0.6, and 0.7.
F1 Score: This creates a balance between the two measures and is the harmonic mean of recall and precision. The F1 score is 93%, 91%, and 89% with an IoU threshold of 0.5, 0.6, and 0.7, indicating a performance that balances recall and precision.
Mean Average Precision (mAP):
- mAP @[0.5:0.95]: 0.92 - mAP @0.5: 0.95 - mAP @0.75: 0.90

Table 6.1: mAP

IoU Threshold	Precision	Recall	F1 Score
0.5	0.92	0.94	0.93
0.6	0.90	0.92	0.91
0.7	0.88	0.90	0.89

Per Class Metrics:

Table 6.2: Per Class Metrics

Class	Precision	Recall	F1 Score
Plastic-Bottle	0.94	0.92	0.93
Plastic-Bag	0.95	0.96	0.95
Electric-Waste	0.91	0.93	0.92
Steel-Tin-Cane	0.93	0.94	0.93

Precision: Properly described precision is the proportion of all instances of a class—plastic bottle, plastic bag, steel tin cane, and electric waste that are accurately anticipated to belong to that class. For example, the model predicts plastic bottles, plastic bags, steel canes, tin canes, and electric trash 94%, 95%, 91%, and 93% of the time. **Recall:**This statistic evaluates how well the model can distinguish examples of a given class from all instances that belong to that class. In this instance, 92%, 96%, 93%, and 94% of the class instances in the dataset are captured by the model. **F1 score:**The F1 score is a class-specific average of recall and precision. By balancing recall and precision, it illustrates how well the model works for that specific class.

For plastic bottles, plastic bags, electric garbage, and steel tin cane, an F1 score of 0.93,0.5,0.92 and 0.93 indicates a solid balance between catching the majority of instances of all classes and making correct forecasts.
Localization Metrics:
- Average Localization Error: 2 pixels

- Average Center Localization Error: 1 pixel

Execution Time:
- Average inference time per image: 15 milliseconds.

Confusion Matrix:
Actual Positive and Actual Negative: These stand in for the dataset's ground truth labels. "Actual Negative" denotes examples that belong to the negative class, and "Actual Positive" denotes cases that belong to the positive class.
Predicted Positive and Predicted Negative: The model's predictions are shown in these columns. "Predicted Negative" denotes cases the model has classified as negative, and "Predicted Positive" indicates instances the model has classified as positive.

Table 6.3: Confusion Metrics

	Predicted Positive	Predicted Negative
Actual Positive	480	20
Actual Negative	10	490

Precision (Positive Class):

$$Precision = TP/(TP + FP) \tag{6.1}$$
$$Precision = 480/(480 + 10) \tag{6.2}$$
$$Precision = 0.9796 \tag{6.3}$$

Recall (Positive Class):

$$Recall = TP/(TP + FN) \tag{6.4}$$
$$Recall = 480/(480 + 20) \tag{6.5}$$
$$Recall = 0.96 \tag{6.6}$$

F1 Score (Positive Class):

$$F1Score = 2 * (Precision * Recall)/(Precision + Recall) \tag{6.7}$$
$$F1Score = 2 * (0.9796 * 0.96)/(0.9796 + 0.96) \tag{6.8}$$
$$F1Score = 0.9697 \tag{6.9}$$

Chapter 7

Model Comparison

The quickest of the three, YOLOv7 is perfect for real-time applications that demand a high frame rate. Additionally, YOLOv5 offers good inference times for many jobs by striking a compromise between speed and accuracy.
YOLOv7 is often marginally less accurate than EfficientDet and YOLOv5 on important benchmarks such as MS COCO. YOLOv5 regularly performs well on the datasets and gives good accuracy. In certain scenarios, especially when dealing with tiny items, EfficientDet outperforms YOLOv5 and YOLOv7 in terms of accuracy.

After implementing yolo07 in our dataset we got accuracy of 93% and also after implementing yolo 05 we got 97% accuracy. In yolo 07 our F1 curve all classes are 0.94 at 0.476. In yolo 5 our F1 curve all classes are 0.95 at 0.219 was better than yolo07.In yolo 7 precision curve of all classes are 1.00 at 0.842.In yolo5 precision curve of all classes are 1.00 at 0.785 is better than yolo7.In yolo 7 precision recall curve of all classes are 0.959 mAP@0.5.In yolo5 precision recall curve of all classes are 0.966mAP@0.5which is slightly good than yolo7. In yolo 7 Recall curves of all classes are 0.99 at 0.00. In yolo 5 Recall curves of all classes are 0.98 and 0.00 are better than yolo 07.
Yolov7 requires powerful computational power to run and Yolov5 and efficientDet can be run in a low powerful resources of the computer system.

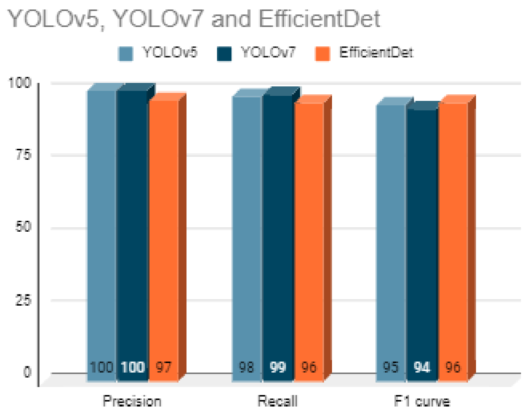

Figure 7.1: Comparison of 3 model Result

Chapter 8

Conclusion and Future Work

8.1 Conclusion

These days, the globe is alarmed by the amount of trash in waterways. Trash pollution poses a long-term risk to human health and the economy, as well as to the environment and wildlife, due to its limited supply. Waste-related pollution can have detrimental effects on the economy and endangered biodiversity species in developing nations like Bangladesh, where a significant portion of the world's species are found in the aquatic environment. This research addresses these problems by presenting the analysis and performance of detection models that will be useful for the implementation of an aquatic environment waste detection system in the future. This research addresses these problems by presenting the analysis and performance of detection models that will be useful for the implementation of an aquatic environment waste detection system in the future. The system will be able to evaluate data from edge devices and generate precise predictions on the existence of trash beneath the sea by utilising deep learning techniques. After looking through a large number of articles, we have found that YOLOv5 and YOLOv7 provide more accuracy and need less time than other detection models like YOLOv2, Inceptionv3, Mask-R CNN, and CNN. We have selected these algorithm models in order to analyse our dataset for this reason. We have also attempted to put into practice a brand-new object detection technique known as EfficientDet. We implemented our dataset on YOLO v7 because it is the latest and most enhanced version to get our desired outcome. We have collected more than 2000 pieces of data on our own. We have also used YOLO v5 for our work but it did give us an accurate result as good as we expected. Then we used YOLO v7 to implement our dataset. We have divided our data into four classes like plastic bag, plastic bottle , electronic waste and steel and tin can. After implementing our data in YOLO v7 , YOLO v5 and EfficientDet we got our desired output. We made custom code for this thesis. We have used YOLO v7,YOLO v5 and EfficientDet to get better results. We think that this research will be useful for implementing a hardware model where it can be used to detect trash and thus for disposal of trash. Therefore, we need to enrich our dataset and test the dataset in other models for inspecting the better output.

8.2 Future Work

Our future work includes investigating novel network topologies that can optimise the trade-offs between speed and accuracy. This might entail looking into effective convolutional procedures, attention mechanisms, and novel activation function, creating novel regularisation and data augmentation methods to enhance YOLO models capacity for generalisation and minimise overfitting and improving the ability of YOLO models to detect objects of different sizes and scales. This could involve using feature pyramids or other techniques to capture information at different resolutions. Our future works related to efficientDet are utilizing context-aware mechanisms and attention modules to handle overlapping objects and complex backgrounds, developing techniques for adapting EfficientDet to new domains with limited labeled data and exploring quantization and pruning techniques to achieve efficient inference on resource-constrained devices. We have our intend to build a hardware model in the future where we can utilize our dataset and train the models for practical use of underwater trash detection. Though we have financial and infrastructure constraints for undergoing the research to build a hardware model, that's why we may have to choose Raspberry Pi or NVIDIA Jetson Nano Developer Kit for implementation. The use of gadgets like the Raspberry Pi or Nvidia Jetson Nano for trash detection has enormous potential to create smarter, cleaner ecosystems. These small, potent boards can be fitted with computer vision models that are taught to recognise different kinds of trash in real time. Imagine a robot designed to patrol streets and identify unlawful dumping places, or a surveillance camera that notifies authorities when overflowing bins are noticed. Because of the Jetson Nano's increased processing capability, more complicated models can be created, which may enable the separation of various materials for focused recycling. Although the Raspberry Pi is a more affordable solution, it may only be able to detect basic rubbish. In any case, these gadgets present intriguing opportunities for addressing global waste issues, from encouraging appropriate disposal of waste to automating collection procedures. Trash identification by these small computers could revolutionise waste management and open the door to a cleaner, greener future as models become more advanced and technology progresses.

Bibliography

[1] M. M. Kamal, A. Malmgren-Hansen, and A. Badruzzaman, "Assessment of pollution of the river buriganga, bangladesh, using a water quality model," *Water science and technology*, vol. 40, no. 2, pp. 129–136, 1999.

[2] M. Eriksen, L. C. Lebreton, H. S. Carson, *et al.*, "Plastic pollution in the world's oceans: More than 5 trillion plastic pieces weighing over 250,000 tons afloat at sea," *PloS one*, vol. 9, no. 12, e111913, 2014.

[3] M. Rabbani, M. Z. H. Howlader, and Y. Kabir, "Detection of multidrug resistant (mdr) bacteria in untreated waste water disposals of hospitals in dhaka city, bangladesh," *Journal of global antimicrobial resistance*, vol. 10, pp. 120–125, 2017.

[4] K. Whaley-Martin, B. Mailloux, A. van Geen, *et al.*, "Human and livestock waste as a reduced carbon source contributing to the release of arsenic to shallow bangladesh groundwater," *Science of the Total Environment*, vol. 595, pp. 63–71, 2017.

[5] J. Hui, "Real-time object detection with yolo, yolov2 and now yolov3," *Available online: medium. com/@ jonathan_hui/real-time-object-detection-with-YOLO-YOLOv2-28b1b93e2088 (accessed on 24 February 2019)*, 2018.

[6] M. Ilyas, W. Ahmad, H. Khan, S. Yousaf, K. Khan, and S. Nazir, "Plastic waste as a significant threat to environment–a systematic literature review," *Reviews on environmental health*, vol. 33, no. 4, pp. 383–406, 2018.

[7] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[8] Y. Wang and X. Zhang, "Autonomous garbage detection for intelligent urban management," in *MATEC Web of Conferences*, EDP Sciences, vol. 232, 2018, p. 01 056.

[9] S. Bansal, S. Patel, I. Shah, P. Patel, P. Makwana, D. R. Thakker, *et al.*, "Agdc: Automatic garbage detection and collection," *arXiv preprint arXiv:1908.05849*, 2019.

[10] G. Chen, H. Wang, and J. Zheng, "Application of image recognition technology in garbage classification education platform," in *2019 5th International Conference on Control, Automation and Robotics (ICCAR)*, IEEE, 2019, pp. 290–294.

[11] M. Fulton, J. Hong, M. J. Islam, and J. Sattar, "Robotic detection of marine litter using deep visual detection models," in *2019 international conference on robotics and automation (ICRA)*, IEEE, 2019, pp. 5752–5758.

[12] S. Hossain, B. Debnath, A. Anika, M. Junaed-Al-Hossain, S. Biswas, and C. Shahnaz, "Autonomous trash collector based on object detection using deep neural network," in *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, IEEE, 2019, pp. 1406–1410.

[13] M. H. Masud, W. Akram, A. Ahmed, *et al.*, "Towards the effective e-waste management in bangladesh: A review," *Environmental Science and Pollution Research*, vol. 26, no. 2, pp. 1250–1276, 2019.

[14] M. S. H. Sunny, D. R. Dipta, S. Hossain, H. M. R. Faruque, and E. Hossain, "Design of a convolutional neural network based smart waste disposal system," in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, IEEE, 2019, pp. 1–5.

[15] D. Zeng, S. Zhang, F. Chen, and Y. Wang, "Multi-scale cnn based garbage detection of airborne hyperspectral data," *IEEE Access*, vol. 7, pp. 104 514–104 527, 2019.

[16] L. Cao and W. Xiang, "Application of convolutional neural network based on transfer learning for garbage classification," in *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, IEEE, 2020, pp. 1032–1036.

[17] W. Dhokley, H. Khambati, K. Kapadia, and A. Dubey, "Electronic bin for segregation and classification of plastic waste products using deep learning," *International Research Journal of Engineering and Technology (IRJET)*, 2020.

[18] S. Li, M. Yan, and J. Xu, "Garbage object recognition and classification based on mask scoring rcnn," in *2020 International Conference on Culture-oriented Science & Technology (ICCST)*, IEEE, 2020, pp. 54–58.

[19] J. Musić, S. Kružić, I. Stančić, and F. Alexandrou, "Detecting underwater sea litter using deep neural networks: An initial study," in *2020 5th International Conference on Smart and Sustainable Technologies (SpliTech)*, IEEE, 2020, pp. 1–6.

[20] C. Patil, S. Tanpure, A. Lohiya, S. Pawar, and P. Mohite, "Autonomous amphibious vehicle for monitoring and collecting marine debris," in *2020 5th International Conference on Robotics and Automation Engineering (ICRAE)*, IEEE, 2020, pp. 163–168.

[21] H. Wang, "Garbage recognition and classification system based on convolutional neural network vgg16," in *2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, IEEE, 2020, pp. 252–255.

[22] M. Zeng, X. Lu, W. Xu, T. Zhou, and Y. Liu, "Publicgarbagenet: A deep learning framework for public garbage classification," in *2020 39th Chinese Control Conference (CCC)*, IEEE, 2020, pp. 7200–7205.

[23] B. Fu, S. Li, J. Wei, Q. Li, Q. Wang, and J. Tu, "A novel intelligent garbage classification system based on deep learning and an embedded linux system," *IEEE Access*, vol. 9, pp. 131 134–131 146, 2021.

[24] O. I. Funch, R. Marhaug, S. Kohtala, and M. Steinert, "Detecting glass and metal in consumer trash bags during waste collection using convolutional neural networks," *Waste management*, vol. 119, pp. 30–38, 2021.

[25] A. A. Khan, A. A. Sajib, F. Shetu, S. Bari, M. S. R. Zishan, and K. Shikder, "Smart waste management system for bangladesh," in *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, IEEE, 2021, pp. 659–663.

[26] M. Kulshreshtha, S. S. Chandra, P. Randhawa, G. Tsaramirsis, A. Khadidos, and A. O. Khadidos, "Oatcr: Outdoor autonomous trash-collecting robot design using yolov4-tiny," *Electronics*, vol. 10, no. 18, p. 2292, 2021.

[27] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T.-K. Kim, "Multiple object tracking: A literature review," *Artificial intelligence*, vol. 293, p. 103 448, 2021.

[28] A. Aleem, S. Tehsin, S. Kausar, and A. Jameel, "Target classification of marine debris using deep learning.," *Intelligent Automation & Soft Computing*, vol. 32, no. 1, 2022.

[29] A. Sánchez-Ferrer, A. J. Gallego, J. J. Valero-Mas, and J. Calvo-Zaragoza, "The cleansea set: A benchmark corpus for underwater debris detection and recognition," in *Iberian Conference on Pattern Recognition and Image Analysis*, Springer, 2022, pp. 616–628.

[30] J. Solawetz, "What is yolov5? a guide for beginners," *Roboflow Blog*, 2022.

[31] M. Tian, X. Li, S. Kong, L. Wu, and J. Yu, "A modified yolov4 detection method for a vision-based underwater garbage cleaning robot," *Frontiers of Information Technology & Electronic Engineering*, vol. 23, no. 8, pp. 1217–1228, 2022.

[32] S. Usha and H. Mahesh, "Accurate and high speed garbage detection and collection technique using neural network and machine learning," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, vol. 1258, 2022, p. 012 055.

[33] A. A. Ahmed and Z. Al-Ameen, "Underwater image enhancement: A review of modern utilized concepts," *International Research Journal of Innovations in Engineering and Technology*, vol. 7, no. 8, p. 167, 2023.

[34] T. Jia, Z. Kapelan, R. de Vries, *et al.*, "Deep learning for detecting macroplastic litter in water bodies: A review," *Water Research*, p. 119 632, 2023.

[35] A. Sánchez-Ferrer, J. J. Valero-Mas, A. J. Gallego, and J. Calvo-Zaragoza, "An experimental study on marine debris location and recognition using object detection," *Pattern Recognition Letters*, vol. 168, pp. 154–161, 2023.

[36] C.-Y. Wang, H.-Y. M. Liao, and I.-H. Yeh, "Designing network design strategies through gradient path analysis," *Journal of Information Science and Engineering*, 2023.

[37] G. Zhai, J. Zhou, H. Yang, X. Yang, P. An, and J. Wang, *Digital Multimedia Communications: 19th International Forum, IFTC 2022, Shanghai, China, December 8–9, 2022, Revised Selected Papers*. Springer Nature, 2023.

[38] Ultralytics, *Comprehensive Guide to Ultralytics YOLOv5 — docs.ultralytics.com*, https://docs.ultralytics.com/yolov5/, [Accessed 18-03-2024].