BOTTLE DETECTION FROM AERIAL IMAGES

MUHAMMAD SHERAN KHALID
Enrollment No: 01-241172-046

A thesis submitted to the Department of Software Engineering, Faculty of Engineering Sciences, Bahria University, Islamabad in the partial fulfillment for the requirements of a Master degree in Software Engineering

May 2020

APPROVAL FOR EXAMINATION

Scholar's Name: **Muhammad Sheran Khalid**

Registration No. **01-241172-046**

Programme of Study: **Master of Science in Software Engineering**

Thesis Title: **Bottle Detection from Aerial Images**

It is to certify that the above scholar's thesis has been completed to my satisfaction and, to my belief, its standard is appropriate for submission for examination. I have also conducted plagiarism test of this thesis using HEC prescribed software and found similarity index **12%** that is within the permissible limit set by the HEC for the MS degree thesis. I have also found the thesis in a format recognized by the BU for the MS thesis.

Principal Supervisor's Signature: ____________

Date: ____________

Name: **Dr. Ahmed Ali**

AUTHOR'S DECLARATION

I, __Muhammad Sheran Khalid__ hereby state that my MS thesis titled "__Bottle Detection From Aerial Images__" is my own work and has not been submitted previously by me for taking any degree from this university __Bahria University Islamabad__ or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw/cancel my MS degree.

Name of scholar: __MUHAMMAD SHERAN KHALID__

Date: ______________________

For the betterment of mankind

ACKNOWLEDGEMENT

Thanks to ALLAH almighty for uncountable blessing which help me to achieve and complete this task. This thesis could not be completed without time, effort, and support of a number of people I wish to appreciate and accept the involvement of all. I am very grateful of my supervisor **Dr. Ahmed Ali** for his full professional support. I am extremely obligated to the support and guidance allotted by him in every step of my research. He was the excessive motivation for me to keep trying during the time I have been in his supervision. My friends were very helpful in providing material, facilitated me to collect data, Instruction in many activities such as data collection, statistical analysis, and other associated activities.

ABSTRACT

Plastic pollution is a growing concern around the globe which possess long term environmental, health and economical threats. To minimize these threats computer artificial intelligence has stepped in with its domain computer vision to successfully identify the plastic waste in the wild. In this research, we propose a supervised learning object detection framework to find and localize waste bottles in the wild using UAV images dataset as plastic waste bottles are one of the top three most abundant plastic waste material but since bottles in UAV images are very small and sometimes transparent with complex backgrounds, it could be a very challenging task to correctly detect and localize such objects. For that reason, we have made use of ensemble methods since they can improve the object detection performance. In our implementation we have used voting strategy for ensembling the output of deep learning convolutional neural networks (CNN) based object detection models since deep neural networks are fantastic at supervised learning and were able to outperform any corresponding model or technique. Best results were obtained by ensembling a strong single stage object detection model, RetinaNet with a powerful two stage object detection model, Faster RCNN with an AP value of 92%. Further, a detailed analysis of the dataset and benchmarks are presented in this research. This research also shows that choosing the right models for ensembling is crucial since in our testing we found that ensembling a weaker model with a strong one tends to decrease object detection performance, for that reason a detailed literature review was constructed and some existing models and techniques are presented in a brief comparison tabular form. Further, we have also showed the importance of data cleansing by the application of data preparation techniques, since going straight from data collection to model training leads to suboptimal results.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Yolo	-	You Only Look Once
AP	-	Average Precision
FCOS	-	Fully Convolutional One-Stage Object Detection
ANN	-	Artificial Neural Network
CNN	-	Convolutional Neural Network
PET	-	Polyethylene Terephthalate
SVM	-	Support Vector Machine
PCA	-	Principal Component Analysis
AI	-	Artificial Intelligence
CV	-	Computer Vision
OBB	-	Oriented Bounding Box
HBB	-	Horizontal Bounding Box
SSD	-	Single Shot MultiBox Detector
TP	-	True Positive
FP	-	False Positive
TN	-	True Negative
FN	-	False Negative
UAV	-	Unmanned Aerial Vehicle
IoU	-	Intersection Over Union
NGO	-	Non-Governmental Organization

CHAPTER 1

INTRODUCTION

Excessive and increased consumption of plastic or polyethylene terephthalate (PET) bottles in everyday customer application has the effect in bottled-water as quickest developing beverage manufacturing industry in whole world. From a customer statistical surveying organization "Euromonitor, The Guardian" has publish a report which says that around 20,000 plastic bottles are brought every day worldwide. Around 480 billion plastic bottles were bought comprehensively in 2016 however only half of them got recycle [1].

PET is kind of plastic which hold crucial importance in our everyday life. It is a well-known business used polymer having application running from fabrics, packaging, films, shaped parts for car, gadgets, and a lot more. You can see this renowned clear plastic around you as water bottle or soft drink bottle. Polyethylene terephthalate (PET or PETE) is a broadly useful thermoplastic polymers or synthetic polymers and they belongs to a subcategory of polyesters from polymers family [2].

1.1 BOTTLES AND PLASTIC

1.1.1 Excess of Bottles

Every year around the globe, almost 89-billion-liter water is bottled and used. Whole utilization of purified water around the globe in 2004 was practically twofold than that of 1997. Additionally, around the globe yearly growing-rate of plastic water bottle utilization was 6.2% from 2008 to 2013.[3].

1.1.2 What is plastic?

Plastic is produced using polymers. Long, revising chains of molecular group. In nature polymers can be found everywhere, silk, the walls of cells, hairs, bug and worms carapaces, DNA, but on the other hand it is also possible to make them synthetically. By breaking raw petroleum into its component parts and change them by rearranging, we can frame artificial polymers. Artificial polymers have unique characteristics. They are tough, lightweight and can be molded into practically any shape. Not requiring hard work, which is time consuming, plastic can be well mass produced, it is extremely economically, and its crude materials are accessible in huge quantity and thus the golden age of plastic started. Now a days, almost everything around us is partially made from plastic, vehicles, our clothes, computers, phones, furniture, and houses. Plastic has since stopped to be revolutionary and progressive material, rather it became junk. Plastic packs, Coffee cups or 'stuff' to rap a mango. We do not consider this reality a great deal, plastic just shows up and leaves. Shockingly, it does not. Since artificial polymers are solid, plastic takes somewhere in the range of 500 and 1000 years to biodegrade [4]. In any case, one way or another, we all chose to utilize this overly risky and super tough material for things that are supposed to thrown away.

Sadly, just to ban plastic is certainly not a conceivable solution for this issue. Plastic pollution is not the main challenge of environmental change we must face but some alternatives we use instead of plastic, have a higher impact on environment in many other ways. For instance, as indicated by a report, published by government of Denmark, a single use of plastic bag requires little vitality, and creates less carbon dioxide releases than a reusable shopping bag made of cotton. You would need to utilize your cotton bag, 7100 times before it would have a lesser sway on planet than a shopping bag made of plastic [5]. Thus, we are left with a confusing procedure of tradeoffs.

Everything has an affect some way or another, and it is hard to pin down the correct balance between them. On the other hand, plastic helps to tackle issues that we do not have better solutions for right now. Internationally, 1/3 of eatery or foods that is delivered or ever bought is never eaten and winds up decaying on landfills, where it produces methane [6]. Plastic packaging is still the most ideal method for keeping food from ruining and avoiding waste.

1.1.3 Horror of Plastic waste and PET Bottles as One of Primary Waste Source

Plastic waste is a horror that the world is facing right now. A lot of which is ending up in our oceans and is ultimately finding its way back to humans, in our bodies, through food chain, in a micro-plastic (MP) form [7], [8]. Micro-plastics are pieces little than 5 mm. Some of them are utilized in beauty care products or toothpaste, but most of them come from the plastic waste floating in the oceans because it is overly exposed to extreme UV radiations and disintegrates into littler and littler pieces. 51 trillion such particles coast in the sea where they are somehow more effectively eaten by all the marine life [9].

It has a destructive effect on marine ecosystem and pose a long-term economic and environmental threat [6]. Major element, making it to the top three of the most abundant plastic litter materials are plastic or PET bottles as stated in a report by International Coastal Cleanup 2017 [10]. Then again PET is also one of the most recycled thermoplastics, & as its recycling symbol is still the number '1' icon [2].

This has raised pressing concerns among researchers, particularly about health threats from harm full chemicals that are mixed with plastic. BPA (bisphenol A), for instance makes bottles of plastic transparent, and there are arguments and proofs that states BPA interferes with our hormonal imbalance [11]. Plastic can be made flexible by adding DEHP (Di-2-ethylhexyl phthalate) however it may cause cancerous growth [12]. It is extremely terrible that small scale plastics (MPs) are poisonous as they travel up the food chain. Zooplankton swallow MPs, tiny fish eat zooplankton, so do the predator fish, crabs, and ousters and they all end up on our plate. MPs have been found in household dust around us, in honey, in beverages and even in tap water. 8 infants out of 10 and about all grown-ups have quantifiable measures of phthalates (a typical plastic added substance) in their bodies and 93% of individual adults have BPA in their pee [13]. It is sheltered to state that a great deal of stuff happened for which we were not ready and that we have lost power over plastic.

1.1.4 How to deal with bottle litter?

Among the solid and hard waste materials, a lot of attention has been attained by plastic since synthetic polymers are not effectively biodegradable. Waste PET bottles are one of the primary causes of plastic waste. Main usage of PET bottles are carbonated beverages bottles and mineral water bottles. Since, Biodegradation of one PET container left in nature can last around 500-1000 years, subsequently, this poses many environmental threats to equally marine and terrestrial zones. Plastic bottles waste is environmental issue which is are hard to biodegrade so it should be handled either to reuse or recycle it.

Suppliers are taking a shot to decrease the waste of plastic pollution. PET is currently recycled in numerous nations that are creating explicit waste administration approaches. One solution was from France, huge amounts of PET bottles were gathered in France: it shows recycling-rate of 51% so the gathered PET bottles can be reutilized to make grade r-PET quality recycled products [3].

1.1.5 Plastic recycling

The conception of Earth as an infinite source of natural resources, and at the same time, as the storage place of produced waste by the human being is an idea considered as the past. Globally 335 Mt of plastic pollution which include excessive use of plastic bottles, plastic bags, microbeads is produced every year that defiantly influences natural life, natural life living space, and people, out of which under 9% of plastic waste is reused or can be recycled [14]. Majority of the plastic squanders are disposed of into landfills causing serious nature concerns.

The import and export of plastic waste has been recognized as one of the important reasons of marine life struggling seek a shelter from plastic waste. Countries bringing in the waste plastics frequently do not have the ability to process all the material. So, the United Nations has forced a prohibition on import export of plastic waste materials if it doesn't meet certain criteria [15].

1.1.6 The environmental benefits of plastic

Since, we cannot just ban and give up plastic as it helps solve the problems that we do not have alternative answers for now i.e. preventing food from spoiling through plastic packaging etc. Plastics give a scope of potential environment rewards. For instance, replacing wood or metals for plastic in vehicles diminishes weight and makes the productivity energy efficient. Plastics additionally does a great job in general wellbeing encouraging clean transportation of drinking water and clinical gadgets to destinations of need, (for example, emergency sites). Plastic pack likewise decreases food wastage by using enhanced environmental and atmospheric packing to enhance the shelf life meat and vegetables.

Another example from constructional industry is that past studies have indicated that the use of ground waste bottles of plastic can be utilized as inadequate substitution of sand in a concrete bricks creation, which is one of the final completing materials for which output and demand is lately getting high, to beat the excessive use of sand and the atmosphere influence brought about by the mining procedure of normal sand and poor disposal, utilizing waste plastic bottles as intermediate change of sand seems to be an excellent choice [16].

In spite of the fact that there are socio-environmental advantages from plastic use, worldwide reliance upon single-use buyer item packaging raises important environmental concerns. Around 40% of the all the plastic waste ever generated around the world has never ended in recycling facilities or managed landfills [6].

1.1.7 Effects of plastic waste on land

Plastic waste represents a great threat to the animals, plants and individuals – including people who depend on the land. The total estimate of plastic pollution on land is somewhere around 23 times than that of ocean [17]. The amount of plastic which is poised is more concentrated and grater on the land, then in the water. Plastic waste which are mismanaged reaches 60 percent in East Asia and Pacific to 1% in North America. The mismanaged level of plastic waste reached at the sea every year and subsequently become out to be plastic marine waste which is almost 33% of total waste annually [17].

1.1.8 Effects of plastic on climate change

In 2019, a new report was published by "The Center for International Environmental Law", which emphasize on the effect of plastic which includes plastic bottles, packs have impacts on climate change [18].

The impact of plastics on environment and climate change is mixed like global warming. Plastics are commonly produced using petroleum. When the plastic is burned, it builds carbon discharges; when it is discarded in the landfills, it turns into a carbon sink. Sometimes plastics that caused methane emanations are biodegradable. Because of the softness of plastic against glass or metal, energy consumption can be reduced. For instance, plastic packaging used for beverages and refreshments in PET plastic instead of glass or metal is evaluated to save transportation energy by 52% [18].

1.2 MOTIVATION

1.2.1 Current trend of finding plastic bottles

What we normally observe aggregating at the ocean side is "Less than the tip of iceberg, maybe a half of 1% of the total plastic litter," says (Erik Van Sebille, an oceanographer at Utrecht University Netherland) [19]. So where is other 99% of ocean plastic? Unsettling answers have recently begun to emerge. We do not know yet clearly. it can be in wildlife, water or in beaches [19].

Plastic waste, a lot of which is PET or plastic bottles litter, is a pressing and concerning issue as it poses long term environmental, economics and health threats to humans and all living creatures on planet earth. So, it needs to be prevented and cleaned up. Few developed countries have taken initiatives and have started addressing this issue. This major problem is faced due to improper waste management and plastic waste littering [20], [21]. The current trend is to efficiently segregate the waste in order to appropriately deal with it [22]. One way of doing that, is manually collecting the plastic litter but manually finding it for collection is a very time-consuming task. For that purpose, the need of an efficient process is clearly inferred, which can be done with the help of artificial intelligence (AI) and Computer vision (CV) i.e. taking aerial images of interested area using unmanned aerial vehicle (UAV) and use it for detection analysis.

There are research based and community service based non-governmental organizations (NGO's) private projects as well as governmental projects going on around the globe focused on finding the plastic litter. For example, an NGO from United Kingdom with their plan "Plastic Tide", wants to build a software that will automatically pick out all the pieces of plastic that wash up on the beaches [23]. So, what the Plastic Tide is doing, it is using UAV technology to image beaches in a way that has never been done before, on a scientific scale. So that you can build up a picture of how much of that missing 99% is washing up on our beaches, the pictures taken are then transferred to a scientific supporting public-sourced website and platform called "Zooniverse". That will build-up and develop a lot of data, which will be utilized to prepare a machine learning algorithm to spot plastic without anyone else - no people required. The expectation is that, in the long run, anybody will have the option to fly drones, take pictures, at that point systems will consequently check the pictures and decide the degrees of plastic contamination of pollution and wastes on a beach. It is a private organization and their dataset is not available to the community. Pictures taken from drone can be accumulated by The Plastic Tide are transferred to Zooniverse, a community-based science site where many community science volunteers made countless labels of what is and what is not plastic litter. They utilized this extraordinary and informative dataset to prepare the algorithm, which is a machine learning algorithm, explicitly a type called a Convolutional Neural Network (CNN). The algorithm utilizes these large number of labelled-tags of marked plastic pieces – with the end goal that it will be able to determine what is a plastic piece and what is not on real time data. The detecting system on its default settings does precisely recognize around **25%** of the plastic pieces, which "Plastic Tide" finds promising outcomes given that it is a very challenging task [23].

Another example is, (Jinwang et al.,) presents benchmarks and a dataset of bottles for low altitude UAV object detection [21]. For bottle detection they constructed some baseline models for example, Faster R-CNN, RRPN, SSD and YOLOv2 with Oriented Bounding Box (OBB) technique which gives angle information of object for robot grasping, the accuracies they have achieved are 86.4%, 88.6%, 87.6%, 67.3%, respectively [21].

1.3.8 Object detection and ensemble methods

Object detection is the task of determining the position and category of multiple objects in an image. Currently, the most successful object detection models are based on deep learning algorithms, and they can be split into two groups: one-stage and two-stage detectors. The former divides the image into regions that are passed into a convolutional neural network to obtain the list of detections — these algorithms include techniques such as Single-Shot Detector (SSD) [31]or You Only Look Once (YOLO) [32]. The two-stage object detectors employ region proposal methods, based on features of the image, to obtain interesting regions, that are later classified to obtain the predictions — among these algorithms, we can find the Reginal Convolutional Neural Network (R-CNN) family of object detectors or Feature Pyramid Network (FPN) [33]. Independently of the underlying algorithm, the accuracy of these detection models can be improved thanks to the application of ensemble methods.

1.3.9 Ensemble methods

It combines the predictions produced by multiple models to obtain a final output. These methods have been successfully employed for improving accuracy in several machine learning tasks, and object detection is not an exception. We can distinguish two kinds of ensembling techniques for object detection: those that are based on the nature of the algorithms employed to construct the detection models, and those that work with the output of the models. In the case of ensemble methods based on the nature of the algorithms, different strategies have been mainly applied to two-stage detectors. In the case of ensemble methods based on the output of the models, the common approach consists in using a primary model which predictions are adjusted with a secondary model [34]

1.5 MAIN CONTRIBUTION

1.5.1 Problem statement

In UAV images, background is usually very complex and objects are very small compared to conventional datasets which generally results in poor detection performance and since bottle size is very small, state of the art object detecting models like YOLOv2 does not perform very well in this case.

1.5.2 Why ensemble learning works?

It means combining different machine learning models to get better prediction. Fundamental thought is to get familiar with a lot of classifiers (specialists) and to permit them to cast a vote.

Advantage: Improvement to get accurate prediction.

Disadvantage: It is hard to understand a group of classifiers.

The technique this research incorporates is proposed by (Angela et al.,) [37]. Its ensembles the output of detection models using different voting strategies. The method is independent of the underlying algorithms and frameworks, and allows us to easily combine a variety of multiple detection models

1.5.3 Research Questions

- Concerning the challenging dataset, will existing object detection models when used in ensemble methods, successfully perform object detection for analysis?
- what are the parameters that affects the accuracy and performance?

1.5.4 Aims and objectives

- To prepare a research which offers to minimize environmental and economic issues posed by PET bottles using machine learning algorithms.
- To prepare a research, which will provide assist to other researches regarding detection of plastic waste.
- To take help from already available researches, smart thinking, and a desire to put a small dent in a huge problem.

- To improve the results of object detection problem by using currently available object detection models in ensemble methods. The goal here is to avoid optimizing the models and just to ensemble these models in a simple ensemble method technique to achieve better results.

1.5.5 Proposed Solution

We propose an imagery-based framework for visual shape-based object detection, particularly (PET) bottles in the wild. It will use image segmentation methods prior to ensemble learning for preprocessing in order to separate objects from background and image classification methods in ensemble learning for classification. Our focus is on dealing with the problem of small bottle size as well as complex image background. We will evaluate our framework on the dataset of UAV bottle litter. A general view of proposed framework is illustrated in the figure 1.3.

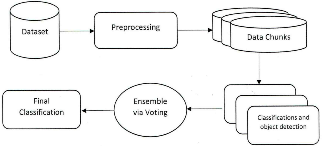


Figure 1. 1 Proposed Solution General View

1.5.6 Materials and Tools

- The dataset that I have acquired for my research is presented by (Wang. et al.,) since we want to focus on the plastic or PET bottles litter in the wild [21].
- For experimentations, training, testing, validation, and visualization We have used Python 3.7 with machine leaning and deep learning libraries & frameworks like pytorch, tensorflow, keras, darknet, numpy, mx-net, OpenCV, Matplotlib etc.
- For training and evaluation of models, I have made use of google colaboratory as well as a local machine with nvidia cuda enabled GP104 graphics processor.
- For referencing in my thesis write-up, I have used a referencing tool called Mendeley.

1.5.7 Evaluation metrics

For evaluation of the test set, PASCAL VOC evaluation metrics with Precision x Recall curve metrics and Average Precision metrics with all point interpolation method is used as this is the latest method used by PASCAL VOC challenge[38].

1.6 Thesis Organization

This thesis is divided into 3 sections thus have 3 chapters. The first chapter, introduction, builds understanding of the problem with respect to domain, discuss techniques and tools used for testing and evaluation and presents a proposed framework.

In chapter 2 a detailed literature review is constructed, and a comparison of existing models and techniques are presented in a tabular form. Finally, in last section, chapter 3, a detail analysis of experimentations is done, and benchmarks are presented.

CHAPTER 2

LITERATURE REVIEW

Since the plastic problem has been brought to light recently and the world is still in the phase of realizing the extent of its severity, limited research has been done on the issue of how this matter can be controlled through technology. How computer science and technology can contribute in this matter is by giving efficient methods and techniques for finding and detecting plastic waste scattered around the world so it can be collected and processed.

In literature, techniques such as Convolutional Neural Networks (CNN), Principal Component Analysis (PCA) and Support Vector Machines (SVM) have been used to detect the plastics present in the waste. An automatic waste sorting approach is presented by (Sakr. et al) [22]. They have done segregation of different materials (plastic, paper and metal) from waste using images and have compared CNN with SVM with 83% and 94.8% accuracy results respectively [22]. Their information is 256 × 256-pixel goals picture of the waste. For their CNN engineering, they use AlexNet model. Their SVM uses a pack of highlights got by passing an 8 × 8 window over the entire image. Every calculation makes an alternate classifier that isolates squander into three primary classifications: plastic, paper, and metal. They accomplished a grouping accuracy of 94.8% with SVM, while CNN had a precision of 83%. The approach used here have focus on the classification of specific objects which not to limit from distance. (Lorenzo-Navarro. et al.,) have done it utilizing pictures, shading based and shape-based highlights, alongside a Random Forest classifier, and have accomplished precision of 96.6%, perceiving four sorts of particles (Micro-Plastics) for instance tar, line, pellets and fragments[7].

Other frameworks such as YOLOv2 have also been used for object detection but it does not work very well on the objects that are far away in the image[39]. Another research is a waste related which intended to coarsely section a heap of trash in a picture. The author used an optimized pre-trained model AlexNet [35]. Their methodology centers around portioning a heap of trash in a picture and gives no insights regarding sorts of squanders in that fragment. There exist moves toward that characterize trash into reusing classifications; propose a framework to group squander in secondary schools. They have designed a container containing a camera inside it for the classification, objects are required to be set inside the box. Their pre-processing images module depends on discovering connection between the picture of the item in the container and 50 distinct pictures, at that point picking the best one as the correct classification. The created framework orders three sorts of waste: PET container, soft drink jars and animation box, with performance of classification are over 70% [36].

Most of these techniques or algorithms have some overhead i.e. SVM requires a lot of preprocessing as a lot of features needs to be set prior, SVM every so often indicates over-fitting problem from improving the parameters to model selection [35] even the most straightforward in difference couldn't be caught by the PCA except if the preparation information clearly gives this data [36].

As discussed earlier, due to the scarcity of research, and to the best of my knowledge there is no publicly available dataset other than the one presented by (Wang. et al.,) [21] it is really hard to classify all types of plastics.

2.1 BACKGROUND

This section will discuss related studies a brief explanation of required knowledge of domains on which this research is built. Some important definitions for understanding are quoted from the literature.

2.1.1 Artificial Intelligence

"Artificial intelligence (AI) is the simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include expert systems, natural language processing (NLP), speech recognition and machine vision." [24]

AI programming focuses on three cognitive skills: learning, reasoning, and self-correction.

Learning processes. "This aspect of AI programming focuses on acquiring data and creating rules for how to turn the data into actionable information. The rules, which are called algorithms, provide computing devices with step-by-step instructions for how to complete a specific task."

Reasoning processes. "This aspect of AI programming focuses on choosing the right algorithm to reach a desired outcome."

Self-correction processes. "This aspect of AI programming is designed to continually fine-tune algorithms and ensure they provide the most accurate results possible."

2.1.2 Machine Learning

"Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. Machine learning helps us in finding solution to problems in speech, computer vision and robotics" [25].

2.1.3 Computer Vision

"Computer vision is a field of artificial intelligence that trains computers to interpret and understand the visual world. Using digital images from cameras and videos and deep learning models, machines can accurately identify and classify objects — and then react to what they see." [26].

2.1.4 Data Science

"Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data. Data science is related to data mining, deep learning, and big data" [27].

2.1.5 Deep Learning

"Deep Learning is Large Neural Networks [28]. In Deep Learning research, CNNs are specifically applied for Computer Vision applications that involves Image Classification and Object Recognition" [29].

2.1.6 Neural network v Artificial neural network v convolution neural network

"Neural networks are a set of algorithms, modeled loosely after the human brain, that are designed to recognize patterns. They interpret sensory data through a kind of machine perception, labeling or clustering raw input" [29].

"The major difference between a traditional Artificial Neural Network (ANN) and CNN is that only the last layer of a CNN is fully connected whereas in ANN, each neuron is connected to every other neuron" as shown in figure 1.1 [29].

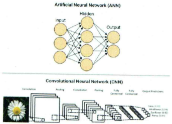


Figure 1. 2 Difference between ANN and CNN

"Deep learning is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised. One reason that deep learning has taken off like crazy is because it is fantastic at supervised learning" [28].

2.1.7 Supervised Learning

"Supervised learning is one of two broad branches of machine learning that makes the model enable to predict future outcomes after they are trained based on past data where we use input/output pairs or the labeled data to train the model with the goal to produce a function that is approximated enough to be able to predict outputs for new inputs when introduced to them. Supervised learning problems can be grouped into regression problems and classification problems. A regression problem is when outputs are continuous whereas a classification problem is when outputs are categorical" [30]. A typical working process od supervised model is sjpwm im figure 1.2.

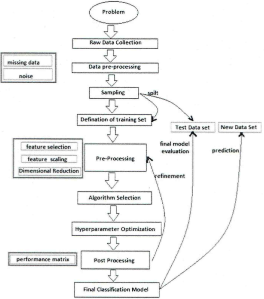


Figure 1. 3 Supervised Learning Approach Model

2.2 Object Detection

Object detection is a vital region of computer vision that has a lot of historical research over the decades. The general objective of object detection is to locate an accurate item in a picture. The item is commonly from a pre-characterized category. Object detection consist of two major tasks: classification and localization. localization is normally drawing a bounding box around the item demonstrating where a given object is in the picture and classification is deciding the kind of the object with a related object properly. Object detection is a challenging issue because of the major issues on a large scale and moment contrasts between objects.

In [40], the first most challenging task is differentiating or separating objects between classes. Current issue which depend upon the quantity of potential classes present can be thousands or more than thousands of objects. On this, isolated object classes can be different in appearance, for instance an apple and an aero plane, however separate classifications but can be similar likewise be comparative in appearance. Another example, dogs and wolves. These difficulties of object detection can be created from two classifications which characterized as power related, and computational-unpredictability and adaptability related. Vigor related alludes to the difficulties in appearance varieties inside the both of intra-class and between class.

These varieties can be classified into two categories images and object variations. Object varieties comprise of appearance contrasts between object cases as for elements, for example, shading, surface, shape, and size. Picture varieties are contrasts not identified with the object occasions themselves but instead the genuine picture. This can comprise of conditions, for example, lighting, perspective, and scale. Differences based on these tasks of both group on given object is given by a class but separating the possible objects into a similar class is very challenging task [40].

2.3 Object detection from Unmanned Aerial Vehicles (UAVs) images

During the previous decades, Unmanned Aerial Vehicles (UAVs) have demonstrated amazing potential among various applications. At the point when a UAV mounted with various types of sensors, the it can broaden the detecting range and change a static detecting task into a portable detecting task. The UAVs' focal points of unreservedly utilizing 3-dimensional (3D) space carry prospects to numerous conventional assignments, similar to control errands in a distribution center or a plant. Unmanned ethereal controllers (UAMs) are one of the specific kind of UAVs outfitted along with or numerous automated arms and have pulled in a ton of research interests from current era. Only favorable advantage of a UAM shows potential in changing uninvolved detecting missions into versatile 3D intuitive missions, such as assembling and grasping [41]. Hovering and Arial maneuvering drifting abilities make it feasible for a UAM to achieve numerous sorts of missions that are hard for human laborers, for example, getting a handle on plastic jug squander on a bluff. Moreover, there are numerous spots, similar to an Amazon distribution center, with the possibility to send a UAM framework for independent picking and setting. To completely use the space in a warehouse, merchandise can be put at a generally high spot, which is hard for human laborers to reach. As of now, the UAM can give a great deal of help [41].

A lightweight and computation effective vehicle detection known as UAV-Net, that depends on SSD (single shot detection) and adjusted to the novel with aerial imagery qualities. For this reason, an efficiently survey alterations and modifications is presented to the key phases of the indicator as to their effect on deduction of time and deduction of accurateness. The effect of various calculation productive CNN models and variations they dissected as base systems for the task of vehicle detection. Furthermore, a novel channel trimming method that naturally consolidates prepared systems in an iterative aspect. Moreover, effect of utilized component maps, actuation capacities and channels utilized for relapse and order phases is assessed in iterative manner [42].

In another research, a DLR 3K Munich Vehicle Aerial Image Dataset was utilized for underlying investigations. The dataset covers twenty aerial pictures with a goal of 560×3745 px and a crushed inspecting separation (GSD) of 14 cm. The pictures are divided into ten preparing and 10 test examples. Because of the huge degrees, each picture was cut into tiles of size 935×634 px. tiles covering in any event in which one item was considered for analysis. The annotations contain seven distinctive vehicle types given as arranged bouncing boxes. Because of less comments for maximum vehicle categories only van and cars names were considered and converged into a solitary vehicle class. Moreover, arranged jumping cases were changed over to hub adjusted bouncing boxes as indicated in research papers. All SSD tests were prepared and assessed with the first Caffe SSD usage. For preparing, random crops and pivots were acquainted what is more with the photometric information enlargements of the SSD structure, as the dataset contains just 10 full size images. For preparing, we utilize the Adam optimizer agent, a primary learning pace of 10−3 and a smaller than usual cluster size of 16. As far as detection accuracy, the models were assessed by plotting the accuracy review bends and figuring the zone under the bend, which is known as the normal exactness (AP) metric. Each model was then benchmarked on three unique stages, speaking to a server and work area GPU, relating to ground-control-station or disconnected preparing, and the NVIDIA Jetson TX2, which can be coordinated into UAVs for on-board handle. The MAX-N power preset was utilized, taking into consideration the most elevated derivation execution at the expense of a more powerful utilization (15w). Surmising speed is accounted for in outlines every second (FPS) arrived at the midpoint of more than 600 onward cards. utilized picture. Note that the benchmarks do exclude the NMS organize (except if in any case noted) to more likely adjudicator the design changes. If it's not too much trouble allude to the advantageous quantifiable for benchmarks [42].

2.4 Severity of plastic and plastic bottles

Nowadays, with the importance of vacation destinations and tourist attraction, here is a plenty of dump, particularly bottle of plastics which should be reuse in the form of recycling. In any case, these plastic wastes are for the most part gathered by hard working laborers, which is dangerous and time taking. To take care of this issue, we suggest utilizing UAVs to discover and grab bottles [43]. Likewise, a UAV bottle dataset (UAV-BD1) is presented to recognize and find bottles more efficiently and easily. Right now, center on that how to detect specially bottles in UAV pictures. Distinguishing items in UAV pictures assumes a significant job in numerous applications and has gotten critical consideration lately. Even So, it is as yet a difficult and challenging issue because of the high level of detailed information and high resolution image with the incredibly significant level of subtleties, different and shooting stage, constrained explained information, and restricted handling time for ongoing real time applications. In UAV pictures, the containers appear to be totally unique from the bottles in datasets, for example, PASCAL VOC, Microsoft COCO, and so forth.

In [21] they have addressed four challenges: gathering pictures counting bottles an extensive scope and different angle sizes of different scales; assembly pictures which include bottles of diverse foundation scenes; gathering pictures including bottles various directions; gathering whatever number sorts of containers as would be prudent. The UAV stage utilized DJI Phantom 4 Pro quadcopter coordinated to a 3-hub settled gimbal in this world. Pictures are gathered by a camera attached to the quadcopter. The goals of caught pictures are pixels of 5472 × 3078. To all the images which are collected is to cover the bottle waste of an extensive scope of scales and perspective sizes, pictures at various flight elevations extending from 10m-30m that are collected effectively.

In UAV pictures, it is very complex backgrounds of the bottles. To build the decent variety of dataset, images are partition into eight scenes. Suppose there is pictures of eight acts, every act covers one unique picture with the size of 5472×3078 pixels. In the other picture they display the divided pictures of eight scenes, every scene contains three sub images with the size of 342×342 pixels. 8 foundation scenes are picked and clarified in our UAV-BD, including Bush backwoods land, Waste land, Step, Mixture, Flat ground, Plastic arena, Sand land and Grassland [21].

The information for all researches used in trials comes from UAV-BD. To guarantee preparing that all the training and testing information should match roughly, they arbitrarily choose 64% UAVBD according to preparation information, sixteen percent as approval information, and as for the testing process its 20 percent. Entire UAV-BD covers 16258 pictures with 22211 occasions for preparing, 5081 pictures with 6944 occurrences for testing and 4055 pictures with 5624 cases for approval [21].

Then they have used similar assessment indicators to analyze four sorts of pattern models (SSD, RRPN, Faster R-CNN and YOLOv2). The thing that matters is that they set $\theta = 0$ of Faster R-CNN, SSD, YOLOv2's outcomes and assessed these models with OBB ground truth. The AP values of RRPN, SSD, Faster R-CNN and YOLOv2 are 88.6%, 87.6%, 86.4%, 67.3%, respectively. We can see when utilizing OBB ground truth, the exhibitions of the three benchmark strategies decline contrasted and that utilizing HBB ground truth, hence on account of when we set $\theta = 0$, the limitation mistake will increment with OBB ground truth. We can observe that the consequence of RRPN is the best [21].

2.5 Current trend of cleaning waste and Waste sorting or segregation

Throwing waste into a bin is something to be grateful for. Butt off course, it is not the spot the path toward the management of waste elimination, anyway where it truly begins. Segregation is another way for segregating biodegradable waste from non-biodegradable waste for suitable evacuation and reusing recycling underlying advance of waste management. It is much of the time endorsed to have two separate dustbins in the house to shield wet waste from working up with its dry accomplice. worse or wrong segregation may cause mixing in landfills, subsequently inciting dangerous release in the ground and unavoidable tarnishing of ground water. Methane gas is presumably going to be released in such conditions, which is one of the most destructive ozone diminishing substances. Appropriate isolation prompts proper reusing the thing that is recycling. An enormous bit of the waste can be recycled and reused. Various laws, rules and various exercises at the organization level are completed to adjust up to dangerous waste age and the board. Composing audit says that the fundamental technique kept when in doubt incorporates material pickers who assemble and orchestrate most of the urban strong waste [18].

Aim to appropriately oversee cleanliness of urban need to represent a nonstop development the board framework to the management systems. The estimate of urban litter is compulsory and important for such procedure. Hostile to littering associations, for example, urban communities overall are evaluating urban order by methods for human reviews. Zurich - positioned third more than 83 European urban cites for the satisfaction of its residents regarding cleanliness - is leading 14000 reviews per year to evaluate and deal with its cleanliness [42].

However, it is very time taking and detaching waste with their bare hands may cause cuts and wounds because of glass and hard articles. It can cause some genuine

Diseases or contaminations which is not kidding ailments. At any rate, a high inescapability of snack of rodents, dogs and other vermin, this structure is still wherever scale in numerous bits of the India. Isolation system using Radio-frequency identification (RFID) is also used where the RFID is seen as attached to every sort of material during amassing just to decide the issue of masterminding during the evacuation period of the thing. Regardless, the issue develops considering use of RFID scanners in unforgiving and non-sensible regions, included cost the associations must be set up to endure with the objective that marks are annexed to each yield thing. The other technique is using microcontroller for confinement. In fact, even this speaks to some significant issues like extra time usage, not fitting in a wide scope of conditions and unfit to disengage clinical waste, clean waste and e-waste suitably fail to conform to explicit principles and rules constrained by the organization in their segregation [44].

To defeat the disadvantages from all techniques Programmable Logic Controller (PLC) based framework is proposed because of common natural points like plan and arrangement to make required momentary alterations without affecting the whole structure, consistence, cost, less wiring, etc. The future work presents modified system using PLC where infrared (IR), suddenness, photo electric, inductive and capacitive sensors are interconnected with PLC in such a manner along these lines, that they work in a real progression to recognize the materials moving steadily on the vehicle line. water driven chamber will push various gathering plastic bottles which are set precisely inverse to sensor position to collect all the plastic waste which can be additionally utilized as natural powder or recycled [45].

The core modules are utilized in the proposed framework: first, waste will be placed inside smasher to diminish on large size with greater resources. Then they dumped the squashed waste to a channel like construction which helps in deliberate development of materials over the transport belt. Note that it is not associated with the PLC and worked autonomously. Programmable Logic Controller (PLC) Bosch Rexroth PLC fills in center to the venture. It is command on every single other component utilized. Principle capacity to gain the signs of information play out specific activities as there are three major system are involved. Input module is one of them to which recognition of wet waste, object detecting, metal, plastic, glass, and paper recognizing sensors are interfaced. Along the vehicle line these all are fittingly planned with the different weight driven chambers underneath them and the social occasion containers in-front. Fast blowing fan is moreover used to overpower the buildup particles and other lightweight materials into a gatherer set unequivocally reverse to it [45].

Secondly, entire framework that performs activities as indicated by the rationale outline composed for it because of PLC forms the signs from input modules. third one is the yield module interfaced with the yield giving policies. For our circumstance, transport line which starts running when the Infrared (IR) sensor is activated and chambers, they will create to go as a fold which drives the loss into container [45]. The aim of this sensor is to detect different objects on the belt of conveyor by transmitting radiation called infrared radiations. At this point when the item is identified, it began flag the PLC to begin the transport if the beginning catch is complete on previously.

Other sensor which is known as moisture sensor is used to isolate the organic waste that is, we waste from dry. Along these lines, this put toward the start to transport line. It quantifies the alteration in electric resistance. At this point when the liquid fume is consumed, because of conductive polymer the ionic group of functions get separated and the electrical conductivity will increment.

Sensor of plastic detection: this sensor is made up of photoelectric sensor with Built-in Amplifier for Detecting Clear, Plastic Bottles. It can be sorted Distinctive measured bottles up to 2-1. PLC is using the automatic waste segregating system. The system helps to isolate the dry and wet waste alongside not many different segments to detect and partition. This classification can be actualized at various scales and small business ventures, trades to isolate out the glass, metallic paper, and plastic waste all the extra effectively at a moderate expense. By using PLC has included preferences like decrease in labor with improved exactness and speed of waste administration, likewise, maintaining a strategic distance from the danger of working at dangerous spots [45].

Plastic shrinkage cracking (PShC) is probably the most punctual type of breaking in concrete as it happens inside the initial scarcely any hours after the solid has been thrown. Solid components with huge uncovered surfaces are particularly powerless against

PShC. Numerous analysts have proposed models to reproduce plastic shrinkage, draining and to foresee the event of PShC. Right now, model to anticipate the level of PShC is proposed. This model, the alleged PShC Severity Model, depends on the volume of water that vanishes from the solid between the putting and the underlying setting time of the solid. This model was checked utilizing countless PShC test results [46]

2.6 OBJECT DETECTING TECHNIQUES AND MODELS

2.6.1 Two-stage vs One-stage Detectors:

There are mainly two types of object detecting models. On one hand, we have two-phase identifiers, for example, Faster R-CNN (Region-based Convolutional Neural Networks) or Mask R-CNN, that utilize a Region Proposal Network to create locales of interests in the principal stage and send the district recommendations down the pipeline for object characterization and bounding box relapse. Such models arrive at the most elevated precision rates yet are commonly slower. Then, we have single-stage locators, for example, YOLO (You Only Look Once) and SSD (Singe Shot MultiBox Detector), that treat object discovery as a straightforward relapse issue by taking an information picture and learning the class probabilities and jumping box facilitates. Such models arrive at lower exactness rates however are a lot quicker than two-phase object dtectors[47].

2.6.2 Keras RetinaNet

Keras implementation of RetinaNet object detection as described in Focal Loss for Dense Object Detection, the training procedure of keras-retinanet works with training models. These are stripped down versions compared to the inference model and only contains the layers necessary for training (regression and classification values). If you wish to do inference on a model (perform object detection on an image), you need to convert the trained model to an inference model. f you installed keras-retinanet correctly, the train script will be installed as retinanet-train. However, if you make local modifications to the keras-retinanet repository, you should run the script directly from the repository. That will ensure that your local changes will be used by the train script.

Improving Apple Detection and Counting Using RetinaNet. This work aims to investigate the apple detection problem through the deployment of the Keras RetinaNet [48].

2.6.3 FCOS: Fully Convolutional One-Stage Object Detection

Object detection with, for example, RetinaNet, SSD, YOLOv3, and Faster R-CNN depend on pre-characterized grapple boxes. Conversely, FCOS is bounding box free, just as proposition free. By disposing of the predefined set of stay boxes, FCOS totally keeps away from the confused calculation identified with bounding boxes, for example, computing covering during preparing. More critically, it additionally maintains a strategic distance from all hyper-parameters identified with bounding boxes, which are frequently extremely touchy to the last identification execution. It shows a lot less complex and adaptable location system accomplishing improved discovery exactness [49].

2.6.4 Feature Pyramid Networks for Object Detection

Feature pyramids are an essential part in acknowledgment frameworks for distinguishing objects at various scales. Be that as it may, late profound learning object locators have kept away from pyramid portrayals, to a limited extent since they are process and memory serious. In this paper, we abuse the natural multi-scale, pyramidal chain of command of profound convolutional systems to build include pyramids with negligible additional expense. A top-down design with parallel associations is created for building elevated level semantic component maps at all scales. This engineering, called a Feature Pyramid Network (FPN), shows critical improvement as a conventional component extractor in a few applications [33].

2.6.5 Resnet:

ResNet makes it conceivable to prepare up to hundreds or even a great many layers and still accomplishes convincing execution time. Exploiting its amazing authentic capacity, the presentation of numerous PC vision applications other than picture characterization have been supported, for example, object discovery and face recognition.

We can really drop a portion of the layers of a prepared ResNet and still have practically identical execution. This makes the ResNet design much more intriguing, likewise dropped layers of a VGG arrange and debased its presentation dramatically [50].

ResNet50: ResNet-50 that is a smaller adaptation of ResNet 152 and every now and again utilized as a beginning stage for move learning .The key forward leap with ResNet was it permitted us to prepare amazingly profound neural systems with 150+layers effectively. Preceding ResNet preparing exceptionally profound neural systems was troublesome because of the issue of disappearing angles.

Expanding system profundity does not work by basically stacking layers together. Profound systems are difficult to prepare on account of the famous disappearing slope issue — as the angle is back-proliferated to prior layers, rehashed duplication may make the inclination very little. Thus, as the system goes further, its presentation gets immersed or even beginnings corrupting rapidly [50].

2.6.6 MMdetection library

MMDetection is a multi-model object detection code library based on pytorch. MMDetection is an object detection toolbox that contains a variety of object detection and occurrence division techniques just as related parts and modules. The tool stash began from a codebase of MMDet group who won the location track of COCO Challenge 2018. It slowly develops into a brought together stage that covers numerous well-known identification techniques and contemporary modules. It incorporates preparing and deduction codes, yet in addition gives loads to more than 200 system models. We accept this tool kit is by a long shot the most complete discovery toolbox [51].

2.6.7 Deep-learning algorithms implementations in literature

Deep learning in this manner in this manner to permit simple contact to non-expert clients, commercial software has been used – Plastic Finder (Italian software license). – to recognize and evaluate on AMD® hardware. The center calculates of deep trace technology to recognize & evaluate AMP. For this center calculation of all the products is system called deep learning convolutional neural system (CNN). CNNs are multilayer architecture which is appropriate for preparing images of RGB for order and detection of object assignments, where a pile of convolutional layers takes into consideration for translation - for example the net is prepared to perceive an object freely of its situation inside the picture. acceptance for the approach of deep learning should be major fundamental motivation [52].

A well-known methodology for object detection includes decreasing the issue to parallel category. The easy and most basic case of this methodology is the sliding window technique. Right now, classifier is applied at all positions, scales, and, now and again, directions of an image. Though, testing all point in the inquiry space with a non-trifling classifier can be very slow and mild. A viable technique for tending to this issue includes applying a course of straightforward tests to each theorized object area to wipe out most of them rapidly moderately moderate when contrasted with basic classifiers characterized by falls. Right now, depict a strategy for building falls for part-based deformable models, for example, pictorial structures. In the broadest setting, this strategy prompts a course form of top-down powerful programming for a general class of language structure-based models [44].

An object configuration defines a specific proper location or area for the root and a shift for each extra part from its ideal location with comparative to the root. The score of a setup is the entirety of the scores of the parts at their areas short twisting expenses related with every relocation [53].

Forgery detection approaches in the conventional image, two type of forensics schemes are commonly used, plans are generally utilized, dynamic plans and passive schemes. In the dynamic plans, a remotely added substance signal is inserted in the source picture without visual antiques. To decide whether an image is a tampered image, the watermark extraction process is performed on the objective picture to reestablish the watermark. The extricated watermark picture can be utilized to recognize altered regions in the target picture. In any case, there is no source picture for the images produced by the GANs so the dynamic picture fabrication identifier cannot remove the watermark picture. Then again, the inactive picture imitation finders utilize the measurable data on the source picture that is high consistency between various images. Accordingly, intrinsic statistical data can be utilized to detect the fake areas in the picture. The passive image picture detectors cannot be utilized to recognize counterfeit images created, in light of the fact that they are combined from the low-dimensional random vector. In particular, the fake images produced by the GANs are not altered from their original images [53].

System rule for the structure bottle of ceramic can crack detection framework, the key issue is the advancement of a detection system which depends on the investigation of testing necessity, and the requirements decide the general plan of the framework. Right now ceramic bottle break detection framework is preparing four sections including mechanical parts which bolster the camera and light source, the control part depends on the mechanical part which is utilized to change the supporting pole and the tallness of the camera as indicated by the necessary bearing. Identifying part is made out of light source securing card camera, the images of the internal mass of the earthenware bottle is changed over to computerized signal for preparing; handling part is a PC machine which is utilized to manage the split picture procurement card sent by the framework structure outline [44].

Framework equipment stage of hardware system platform of the framework that is adopt is AD9883 which is actual a camera video computerized chip. Utilizing FPGA chip for signal preparing can expand the positions of compelling sign, at last, ADV7123 chip advanced sign is changed over into simple algo signal in the display which shows local output of system design [44].

At least one potential epitome of the present application shows electric detection framework for detecting profile features as well as shape highlights of bottles or similar comparative holders that are proceeding onward a transport toward a path of movement, in which the electric detection framework has one camera plan that incorporates in any event, one related illuminating device for enlightening the compartments of aerial images at any rate in the district of their profile highlights the features and shape that are to be detected by the detection system, wherein the in any event one lighting up device is as a strip-formed light source that stretches out toward development of the holders or movement of plastic waste, and in that illuminating device one camera are flexible comparative with one another for various edges of rate of the light as well as for various points of picture recording. Further improvements, points of interest and application possibilities of at any rate one potential encapsulation of the present application are additionally created from the Subsequent description .On a basic level, this study, portrayed as well as graphically spoke to highlights are objects of in any event one potential epitome of the present application, separately or in discretionary blend, regardless of their Summarization in the cases or their dependency.

Beach litter almost destroys marine environments and makes visual distress that brings down the estimation of the sea. To take care of the issue regarding litter of beach, it is very important to examine the age and dissemination example of waste with different patterns and the reason for the inflow. Nonetheless, the information for the investigation are just example information gathered in certain zones of the seashore. Additionally, most of the information covers just the aggregate sum of seashore litter. UAV (Unmanned Aerial Vehicle) and Deep Neural Network used to detect successfully detect and screen seashore mess. Utilizing UAV, it is conceivable to handily study the whole seashore. The Deep Neural Network can likewise recognize the sort of beach front litter. Thusly, utilizing UAV and Deep Neural Network, it is conceivable to get spatial data by sort of seashore litter. This paper proposes a Beach litter detection calculation dependent on UAV and Deep Neural Network and a Beach litter checking process utilizing it. It additionally offers ideal shooting height and film duplication to detect little seashore litter, for example, plastic bottles and Styrofoam pieces found on the seashore. Right now, Mavic 2 Pro was utilized. The images got through UAV are created as orthoimages and contribution to a pre-prepared neural system calculation. The Deep Neural Network utilized for Beach litter detection expelled the Fully Connected Layer from CNN for semantic division [54].

2.6.8 Ensemble methods in literature

With respect to ensembling depending on the nature of deployed algorithm some works have been focused on ensembling features from different sources before feeding them to the region proposal algorithm, others apply an ensemble in the classification stage, and others employ ensembles in both stages of the algorithm [34].

In case of ensembling the output of algorithms some procedures has been applied by combining Fast-RCNN and Faster-RCNN models, and combining Fast-RCNN and YOLO models, and by using RetinaNet and Mask R-CNN models [34]. Another approach to combine the output of detection models is the application of techniques to eliminate redundant bounding boxes like Non-Maximum Suppression, Soft-NMS, NMW, fusion and WBF [34]. However, these techniques do not take into account the classes of the detected objects, or the number of models that detected a particular object; and, therefore, if they are blindly applied, they tend to produce lots of false positives. As in many other machine learning tasks, the accuracy and robustness of object detectors can be greatly improved thanks to the application of ensemble methods; for instance, the mAP in the COCO dataset was improved from 50.6 to 52.5 in one study, and the mAP in the Pascal VOC dataset increased by 3.2% in in another. In fact, the leading methods on datasets like Pascal VOC or MS COCO are based on the usage of ensembles [34].

Within every ensemble factor pair, the detection for one of the sets will be picked and the other disposed of. This is determined by where the given factor lies for the test image according to the preparation information circulation. For instance, on the off chance that it is estimated that a picture with a profound model to have JPEG pressure underneath the edge used to part the information, at that point the detection discovered utilizing the model prepared on that information will be utilized[55].

2.6.9 Image segmentation techniques

It is surely known, the power of CNNs systems in enormous part from which they have the ability to exploit (translational) surpluses through different sequence of translation equivariance and weight sharing. It became normal to consider speculations that can exploit their major collections of symmetries. These systems are completely constrained to separate groups, for example, discrete pivots following images or stages following up mists. Other extremely ongoing effort is worried about the examination of circular imageries yet does not characterize an equivariant engineering. It accomplish equivariance to a consistent, non-commutative gathering and the first to utilize the summed up Fourier change for quick gathering relationship [56] .

The essential design for tuning detection accuracy is the employee feature maps and default box setting. Map with high feature goals are important to accurately find minor object occasions, particularly intended for aerial images especially for vehicles. In this manner, we just use the last layer output with an estimated down sampling variable of 8 as highlight map, for example if there should be an occurrence of VGG-16. More profound non considered layers because of the low spatial goals of highlight maps. Note that repudiating multi-layer misuse is just fit if there should be an occurrence of a consistent ground inspecting separation which brings about immaterial variety in object sizes. If there should arise an occurrence of little object occasions, the finest detection accuracy is accomplished for default encloses the scope and sizes of the objects. To give reasonable default box sizes, we apply the bunching come nearer from to the preparation information.

Sometimes in bottling there happens a situation when the naming of the bottle is missed. In this situation is essential to detect the broken bottle. A mark detection was performed through optical character acknowledgment technique. In OCR we have utilized layout coordinating calculation. Optical Character Recognition by utilizing Template Matching is a framework model that is helpful to perceive the character or letters in order by looking at two images of the letter set. The motivations behind this framework model are to build up a model for the Optical Character Recognition (OCR) framework and to actualize the Template Matching calculation in building up the framework model. There are a couple of procedures that were associated with this calculation. The procedures are beginning from the obtaining procedure, sifting process, limit the picture, bunching the picture of letter set and finally perceive the letters in order. These procedures are critical to get the consequence of acknowledgment subsequent to looking at the two character images [57].

2.7 EXISITING MODELS AND TECHINIQUES – A BRIEF COMPARISON

This table gives an overview of the techniques in a brief manner that we have learned from the literature review. Some pros and cons are constructed but are not limited to these only.

Table 2. 1: Comparison of Existing Object Detecting Models from Literature

Ref	Techniques	Models	PROS	CONS
P1	YOLOv3: An Incremental Improvement[32]	**Yolov3**	• Faster object detecting performance (i.e. 45 fps)	• Struggles with small objects.
P2	Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [58]	**Faster R-CNN**	• Good performance on small objects • Correlates with performance on bigger objects.	• Input resolution affects detection accuracy of small objects.
P3	Mask Scoring R-CNN [59]	**Mask Scoring R-CNN**	• More accurate mask predictions.	• Inconsistency in the model's arrangement certainty and the nature of the anticipated mask.
P4	Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition [60]	**SSP-net**	• Reduce the acknowledgment precision for the pictures or sub-pictures of a self-assertive size/scale.	• Increasing the number of stacked layers will cause gradient explosion/disappe arance problems.
P5	Feature-Fused SSD: Fast Detection for Small Objects [61]	**SSD**	• Distinguish various items very well • Quicker contrasted then two-shot RPN-based methodologies.	• Struggles with dense objects. • Requires a lot of preprocessing. • Not insignificant to perform back-proliferation through spatial pooling layer.
P6	Focal Loss for Dense Object Detection [48]	**RetinaNet**	• More accurate predictions than two stage detectors. • Better performance YOLO and SSD	• Slow training time.
P7	Cascade R-CNN: High Quality Object Detection	**Cascade R-CNN**	• Works very well with COCO, VOC, KITTI, CityPerson,	• Performance drop when compared

	and Instance Segmentation [62]		and WiderFace datasets.	with non-cascade methods.
P8	Grid R-CNN Plus: Faster and Better [63]	**Grid R-CNN (Plus)**	• Can obtain high-quality localization results.	• Slower inference time.
P9	FreeAnchor: Learning to Match Anchors for Visual Object Detection [64]	**Free-Anchor**	• Anchor matching is more flexible.	• Not suitable for all kind of datasets.
P10	Region Proposal by Guided Anchoring [65]	**Guided Anchoring**	• More effective and efficient when combined with Fast RCNN. • It likewise uses semantic highlights to control the mooring.	• A slick arrangement of stays of fixed angles proportions must be predefined • An off-base decision may hamper the speed and precision of identifiers.
P11	NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection [66].	**NAS-FPN**	• can obtain output of any given pyramid network by giving early detection results.	• imbalance perspective And different architectures can affect results.
P12	Fast R-CNN [67]	**Fast R-CNN**	• reduce overall training time. • increase accuracy.	• accurate localization of objects arise complexity.
P13	Libra R-CNN: Towards Balanced Learning for Object Detection [68]	**Libra R-CNN**	• Improves the detection performance. • Faster transfer with just few clicks.	• Single stage detector only found improvement when connected to R-CNN.
P14	Soft-NMS – Improving Object Detection with One Line of Code [69]	**Soft-NMS**	• upgrades for the coco-style mAP metric on standard datasets like PASCAL.	• Object exists in the predefined cover edge; it prompts a miss.

P15	FCOS: Fully Convolutional One-Stage Object Detection [49]	**FCOS**	• One stage detector with better accuracy.	• post-processing non-maximum suppression. • eliminate hyper parameters related to anchors.

From the literature study and table 2.1 disused in this section we can see that all the models and techniques have their restrictions and benefits. Most of them struggles with small scale objects like images of plastic bottles taken from UAV images. In this study we have tried to fill that gap which is discussed in detail in chapter 3 of this thesis.

CHAPTER 3

DATA AND EXPERIMENTATION

This section will discuss our main contribution and challenges as well as working environments, constraints and resources used. Furthermore, a discussion over brief understanding of our proposed framework and detailed demonstration through experimentations will be covered. The dataset collection, detailed understanding of dataset and preparation of dataset will be constructed as well. We will also be discussing the detailed analysis of experimentations, inferencing and evaluations. Finally, generated results will be compared with each other as well as with corresponding papers from the literature and a brief perspective will be made regarding the plastic issue and our solution. In the end some suggestions will be made regarding future work.

3.1 FRAMEWORK ELABORATION

Our approach is to have a simple solution to improve the accuracy of the object detecting algorithms by ensembling the results of multiple different object detectors. First, the collected data set is transformed into more generalized and clean form to easily convert it to corresponding object detection model's input format. After making a copy of main dataset, it is converted to required format and trainval set is fed to the corresponding model for training. After model training, inference is made on a separate unseen testing set and predictions are made over images. These predictions were generated in .xml files which then are used in the ensemleing phase to have better predictions. Ensembling was done using a voting strategy proposed by (Angela et al.,) [34]. All these predictions are evaluated through PASCAL-VOC evaluation metrics. A visual demonstration of our framework is shown in the figure 3.1.

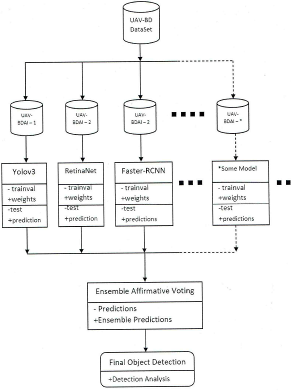


Figure 3. 1 Proposed Framework Detailed View Illustration

3.2 DATASET

Dataset which we have used in this research is originally presented by Jingwang et al., [21] in his work UAV-BD which is published by "2018 21st International Conference on Information Fusion (FUSION)" made openly accessible to the vision network on their website to advance the research in object detection from UAVs . If you want to know more about their work and dataset. we highly recommend a deeper dive into their paper for better understanding [21].

UAV-BD dataset has almost 34, 791 bottle object instances in 25, 407 images. To create diversity and mimic a real-world problem as close as possible, the images are taken from different angles and altitudes with the range of 10m to 30m. It contains images from eight background scenes from the wild including, Sand land and Grassland, Plastic stadium, Mixture, Flat ground, Steps or footstep stairs, Waste land, Bush forest land. The "Grassland" scene has the biggest number of object examples: 7, 795 occurrences in 5, 785 pictures. The "Progression" scene has the most small and modest number of examples: 2106 cases in 1, 325 pictures.

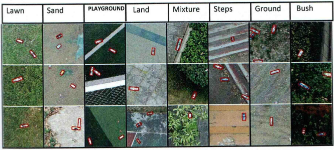


Figure 3. 2 Illustration of UAV-BD

3.2.1 Data Preparation

The process of data preparation that was adapted can be understood by the figure 3.3.

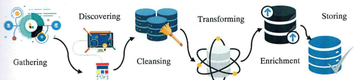


Figure 3. 3 UAV-BDAI Data Preparation Process

3.2.2 Data gathering

With every problem in object detecting problems, the first challenge is collection of dataSets. There are 2 ways of gathering dataset, either you label and annotate your own images, or you acquire published datasets for example MS COCO dataset (2017), PASCAL VOC dataset (2007) etc. However, there are many published custom datasets as well, meant for specific problems which can be obtained and molded for a different set of problems.

Luckily, the dataset we are using is fully annotated and available both in MS COCO 2017 format and PASCAL VOC 2007 format however, some customizations were done to the annotations to satisfy original author's needs.

Figure 3. 4 TrashNet Trash DataSet

3.2.3 Challenges with conventional datasets

We have tested (TRASH-NET) dataset published by Gary [70] but since that dataset is very small with only 481 instances of class (plastic), it was not useful as deep learning models are very data demanding [28].

There are private NGO's organizations like theplastictide and private business organizations such as ZenRobotics® and Max-AI® technology, all working on a similar goal, to reduce the plastic waste from our environment by harnessing the power of AI and computer vision through deep learning, since they are private organizations, they choose not to share their datasets, classifiers (models) and technology. I tried to request these organizations to share datasets but never got to hear from them.

The initiative of the ThePlasticTide to monitor by using drones is an activity to utilize waste along the British coastline. They mean to unroll a comparable task or project along the west bank of Africa one year from now. the Plastic Tide is doing, is utilizing ramble innovation like drone technology to picture sea shores in a manner that is never been done, on a logical scale for scientific manner.so that you can develop an image of the amount of that missing 99% is washing up on all the beaches.

Few glimpses of ThePlasticTide plastic dataset is shown in the figure 3.5

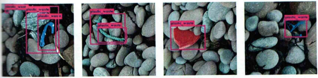


Figure 3. 5 ThePlasticTide UAV Plastic DataSet

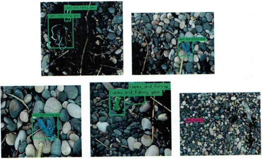


Figure 3.5 (b) ThePlasticTide UAV Plastic DataSet

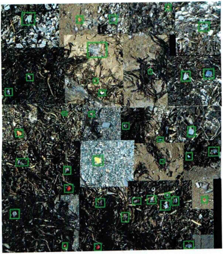


Figure 3.5 (c)
ThePlasticTide UAV Plastic DataSet

An artificial intelligence waste sorting company Max-AI® technology is a computerized reasoning company that distinguishes different things for recovery for recycling through profound learning technology. Max utilizes both multi-layered neural systems and a special framework to see and distinguish objects comparably to the way an a real life individual does. The innovating technology is driving upgrades and showing improvements in Material Recovery Facility (MRF) structure, operational productivity, recuperation, framework improvement, upkeeps, and more. Since, it is a profit organization, they choose not to share their data of waste, glimpse of their dataset is shown in figure 3.6.

Figure 3. 6
MAX-AI Waste Sorting Plastic DataSet

ZenRobotics® Ltd which was recognized in 2007. It is a worldwide innovator in keen mechanical reusing and the principal organization to apply AI-based arranging robots to an intricate waste-arranging condition. Their robots are controlled by their in-house software (AI software) to make recycling efficient and profitable. Since, it is a profit organization, they choose not to share their data of waste, glimpse of their dataset is shown in figure 3.7.

Figure 3. 7
ZenRobotics Waste Sorting DataSet

The requirement of this research was to find and acquire a dataset of annotated plastic objects but since there are no such publicly available datasets and bottles are one of the top three most abundant plastic waste material produce by humans as discussed earlier in detail, I decided to go with UAV-BD as it has sufficient number of images for both training and testing.

3.2.4 Understanding UAV-BD – Data Discovering

Before finalizing UAV-BD dataset, we had to face many challenges for instance finding the right dataset that could particularly satisfy the need of this research problem. Existing published datasets like PASCAL VOC and MS COCO have CLASS **bottle** but objects in bottle class looks completely different than the bottles in UAV images. This research aims to deal with the images of objects bottle captured by UAVs which are usually placed with random arbitrary oriented positions as shown in the figure 3.8 (a). whereas the objects in conventional datasets like PASCAL VOC data are usually in upward oriented position as shown in the figure 3.8 (b).

(a)
UAV BD bottle class

(b)
PACAL VOC bottle class annotations

Figure 3. 8 Difference Between Conventional DataSet and UAV BD DataSet

Using such conventional datasets produces vague detection results on UAV images as shown in experiment section of this research.

The UAV-BD itself is a very challenging dataset and since it is customized for a different set of problem though it is available in MS COCO (.json) style and PASCOL VOC (.xml) style (bbox annotations only), it does not fully comply with either of these standards. For such an issue, it can clearly be seen that a need of data cleansing and correctness is inferred.

3.2.5 Preparing DataSet – Data cleansing

Never assume if dataset is available publicly it means it is cleaned and ready to use, to suit all other problems. Similar concept applies to UAV-BD. Preparing UAV-BD for this problem by cleaning and correcting it was necessary because going directly from collection of data to show preparing prompts imperfect outcomes. There might be issues with the information. Regardless of whether there are no, applying picture expansion extends your dataset and decrease overfitting.

Cleaning and planning information makes up a considerable bit of the effort and time spent in a project of data science. Most of the energy, much of the time. It tends to be enticing to easy route this procedure and make a plunge directly into the demonstrating step without looking hard at the informational collection first, particularly when you have a great deal of information. Oppose the allurement. No informational collection or dataset is perfect; you will be missing information, have confused information, or have inaccurate information. A few information fields will be messy and conflicting. On the off chance that you don't set aside the effort to inspect the information before you begin to show, you may end up re-trying your work over and again as you find terrible information fields or factors that should be changed before displaying. In the most pessimistic scenario, you will fabricate a model that profits wrong forecasts and you will not be certain why. By tending to information gives early, you can spare yourself some pointless work, and a great deal of headaches! [58]. For this purpose, necessary contributions were applied to UAV-BD data set by using roboflow.ai.

Roboflow rearranges your computational work process in a simpler manner and helps with organizing of data, verification of annotation, preprocessing and data augmentation. Roboflow.ai Organize is reason worked to flawlessly settle these difficulties. Indeed, Roboflow.ai splits the code you must compose generally into less than half while giving you access to more preprocessing and increases choices.

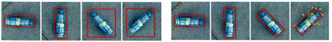
Like tabular data, cleansing and augmenting your images can improve your ultimate performance of model more than making changes to the model's architecture. Preparing images for object detection includes, but is not limited to:

- Verifying your annotations are right (for example none of the annotations are out of frame of the image).
- Ensuring the EXIF orientation of your pictures is right (for example your pictures are saved differently on your storage media in contrast to how you see them in applications).
- Resizing and correcting the object annotations according to resized images.
- Various augmenting techniques that may improve model execution like grayscale and difference changes.
- Formatting annotations to match the demands of model's inputs (e.g. a flat text file for some implementations of YOLO or generating TFRecords for TensorFlow).

UAV-BD annotations follow PASCAL-VOC and MS-COCO datasets formats but does not completely comply with either of these standards. A common description of PASCAL VOC bounding boxes is (xmin, ymin, xmax, ymax), where (xmin, ymin) is the top left location, (ymax xmax,) is the lowest location as shown in the Fig. 1 (a).As shown in Fig.2(a) In original UAV-BD dataset, the PASCAL-VOC format is intended for θ based oriented bounding box (OBB), center location of horizontal bounding box and h, w are the height and width. OBB provides angle rotation information to remove the consequence of rotation on the feature level and make full use of the rotation information for feature extraction so it can utilize the pivot data for include extraction so. format of UAV-BD was altered to custom needs in the original paper as this dataset is intentionally prepared for robot arm grasping of objects which lead to problems while converting datasets to any other format, for instance, PASCAL-VOC to darknet format etc.

We decided to use PASCAL-VOC format as original paper uses PASCAL-VOC evaluation metrics and the ensemble technique we have used in this research also relies on PASCAL-VOC metrics but since UAV-BD PASCAL annotations does not comply with the original PASCAL-VOC dataset bounding box annotations, it needed to be converted.

For converting the dataset to native PASCAL-VOC bounding box version we used Roboflow.ai. Since, the UAV-BD PASCAL-VOC format version was not readable by any framework not even by Roboflow, we used MS-COCO version of UAV-BD to PASCAL-VOC native conversion. In UAV-BD MS-COCO annotation version, annotations at the edges of the frames in some images were not fully inside the image frames thus causing issues while loading into different object detection frameworks for debugging. The annotations at the edges were needed to be trimmed and Roboflow.ai provides an excellent way to trim the annotations. Affected annotations were then intelligently trimmed so they got fully inside the frames. Once, it was done, there was another issue but this time with the Roboflow.ai system itself. The issue was with the PASCAL-VOC coordinate system off-set, since PASCAL-VOC dataSet is a 1-based coordinate system off-set format whereas MS-COCO dataset is a 0-based, meaning bounding box coordinate values (xmin,ymin,xmax,ymax) for PASCAL-VOC must have minimum value that starts from '1' whereas it must starts from '0' for MS-COCO and since UAV-BD MS-COCO version bounding box values were in float data type, roboflow.ai had an issue in their source code and it set coordinate offset to '0' instead of '1' while producing the dataset in PASCAL-VOC format for the float data type values to integer data type values that had coordinate values of "0.xx". We reported the issue to Roboflow.ai support and they pushed a fix instantly.



(a)
Native PASCAL-VOC bbox

(b)
UAV BD PASCAL-VOC bbox

```
<bndbox>
        <xmin>98</xmin>
        <xmax>154</xmax>
        <ymin>235</ymin>
        <ymax>257</ymax>
</bndbox>
```

(c)
NATIVE PASCAL-VOC bbox

```
<robndbox>
   <cx>166.1942</cx>
   <cy>326.4929</cy>
   <w>73.05</w>
   <h>22.1696</h>
   <angle>3.041593</angle>
</robndbox>
```

(d)
UAV-BD PASCAL-VOC bbox

Figure 3. 9 Difference between annotations of UAV BD

3.2.6 Data Transform and enrichment

Once all the corrections and cleansing were done, UAV-BD was transformed to UAV-BDAI. UAV-BDAI was generated in PASCAL-VOC format and is now available in more enriched and generalized format thus it can easily be further transformed into any required format for instance converting to darknet, kerasYolo etc.

Before final dataset was generated, one last preprocessing modification was applied, "Auto-orient". Auto-Orient discards EXIF rotations and standardize pixel ordering Auto-situate is significant in light of the fact that pictures are once in a while put away on plate in unexpected directions in comparison to the applications, we use to see them. Whenever left uncorrected, this can cause quiet disappointments in our models.

Furthermore, data augmentation techniques and image resizing were done as well at time of modeling which will be discussed later in the experiment section as it was different from model to model.

3.2.7 Storing the Finalized dataset

Final dataset, now called UAV-BDAI follows the same split of training validation and testing as of original UAV-BD. To confirm that it matches the testing, distribution, and training data that, 64% randomly selected images were fixed for training data, 16% for validation, and separately dedicated 20% for testing. However, the new UAV-BDAI contains slightly different numbers of images and instances from the original UAV-BD as required corrections were done to annotations and some images were discarded due to corrupted in preprocessing phase. The New UAV-BDAI has 16234 pictures with 23895 cases for training, 4062 all the 6077 images classes for validation and 5078 pictures or images of 7503 cases for testing.

In total UAV-BDAI dataset has about 37, 475 bottle object instances in 25, 347 images whereas the original UAV-BD has 34, 791 bottle object cases in 25, 407 pictures. The difference in number of ground-truth instances is due to the corrections done to the annotations. The new UAV-BDAI is shown in Figure 3.10 and the new transformed annotations is shown in the Figure 3.11.

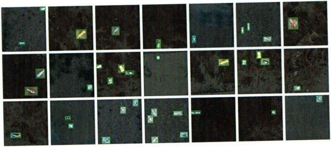
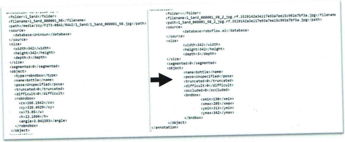


Figure 3. 10 Illustration of UAV BDAI



```xml
<folder>1_Sand</folder>
<filename>1_Sand_000001_98</filename>
<path>/media/zsy/F273-0BA1/RAW2/1_Sand/1_Sand_000001_98.jpg</path>
<source>
  <database>Unknown</database>
</source>
<size>
  <width>342</width>
  <height>342</height>
  <depth>3</depth>
</size>
<segmented>0</segmented>
<object>
  <type>robndbox</type>
  <name>bottle</name>
  <pose>Unspecified</pose>
  <truncated>0</truncated>
  <difficult>0</difficult>
  <robndbox>
    <cx>166.1942</cx>
    <cy>326.4929</cy>
    <w>73.05</w>
    <h>22.1696</h>
    <angle>3.041593</angle>
  </robndbox>
</object>
```

```xml
<folder></folder>
<filename>1_Sand_000001_98_2_jpg.rf.3529142a3e117e93a7ee15c992e7bf3a.jpg</filename
<path>1_Sand_000001_98_2_jpg.rf.3529142a3e117e93a7ee15c992e7bf3a.jpg</path>
<source>
        <database>roboflow.ai</database>
</source>
<size>
        <width>342</width>
        <height>342</height>
        <depth>3</depth>
</size>
<segmented>0</segmented>
<object>
        <name>bottle</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        <occluded>0</occluded>
        <bndbox>
                <xmin>130</xmin>
                <xmax>205</xmax>
                <ymin>313</ymin>
                <ymax>342</ymax>
        </bndbox>
</object>
</annotation>
```

Figure 3. 11 UAV-BD PASCAL-VOC to UAV-BDAI PASCAL-VOC Annotation Conversion

3.3 EXPERIMENTS

This section will cover comprehensive details about the following:

- Models being used.
- How and why these models were selected for this problem?
- Model training, validation, and testing parameters.
- Parameters considered and selected for evaluation metrics to validate the test results.

3.3.1 Model selection

Two models trained for ensembling but have reserved more for the future related research paper publication because model training is a very time consuming and resource demanding task and requires a lot of efforts. The list of models is as follow.

- RetinaNet with resnet-50 backbone trained with keras framework using retinaNet keras library
- Yolov3 with Darknet-53 backbone trained with darknet framework using yolo darknet library
- Mask RCNN with ResNet-101 backbone trained with keras framework using mask rcnn keras library

Few more models were trained with some help from original author of UAV-BD, Jingwang. These models were trained on a remote system that have all the required resources since these models were very resource demanding and required a lot of computational power and we did not had the required resource horse power in the systems we were using for example our local system has nividia gp104 graphic processing unit which have only 8 gigabytes of ram and google colab have a screen timeout limit of 30 min with a 12hours of per session time. List of these models is as follow:

- Faster R-CNN with renet50 and FPN backbone trained with pytorch framework using mmdetection code library

- RetinaNet with renet50 and FPN backbone trained with pytorch framework using mmdetection code library
- FCOS with renet50 and FPN backbone trained with pytorch framework using mmdetection code library

3.3.2 Modeling – mmdetection models

Above three models were trained using Feature Pyramid Network (FPN) with ResNet-50 as the backbone in mmdetection library which is based on pytorch since there are several small objects in UAV-BD, FPN can handle multi scale objects very well [71]. In fact, models trained on mmdetection can obtain higher performance than original code library as mmdetection is much more optimized and is regularly maintained by the authors [51]. These models were trained using original UAV-BD MS-COCO dataset and not with new UAV-BDAI PASCAL-VOC data which means they uses MS-COCO evaluation metrics and not PASCAL-VOC evaluation metrics so for now, in this research, they are used for comparison purposes only.

3.3.3 Challenges and how to overcome them?

UAV-BDAI is a small object dataset and poses many challenges. For instance, the size of bottles is very small, mostly less than 50px and due to images taken from different heights and angles, the size of bottles differs in scale as well. It also results in poor detection performance because of the complex background of bottles in the images. Difficulty further increases since bottles of plastic are often transparent revealing the background through them.

To overcome these challenges, the target was to use models in ensemble methods that are weak in one area but perform well in the other, for instance yolov2 struggles with detecting small objects but provides state of the art performance speed [39]. On the other hand, RetinaNet can match the speed of previous one-stage detectors while surpassing the accuracy of all existing state-of-the-art two-stage detectors [48]. A comprehensive review of these model was constructed in literature review section. Yolov3, retinanet and FCOS all are one stage detectors whereas Faster-RCNN is a two-stage detector. In contrast, one-stage faster and simpler but have tends to trail the accuracy of two-stage detectors[48]. Therefore, we decided to train all these models for comparison and use retinaNet, yolov3 and faster-rcnn (mask-rcnn without segmentation) in ensemble methods to achieve a better balance of models. The goal here is to avoid optimizing the individual models and just to ensemble them in a simple ensemble method technique to achieve better results.

A comprehensive detail is constructed in literature review and a brief description of these models is shown in ***Table 2. 1*** of literature section.

3.3.4 Modeling - Keras-RetinaNet

During experimentations, a lot of challenges were encountered, and some efforts were made to overcome those challenges. For Instance, in keras-retinaNet model, the supported anchor shape size of ground truth bounding boxes is sizes = [32, 64, 128, 256, 512] but since UAV-BDAI has so many small objects and some are less than 32px, some modifications were needed to adjust the anchors so it may not create silent failures in the modeling like objects with anchor shape size less than 32px will not contribute to training, for that reason either we need to optimize the anchors or we could upscale the images and annotations to most optimized size for keras-retinaNet which is 800x1333. We opted to upscale the images from 342x342 to 800x800. The difference can be seen in the figure 3.12. Red bounding box means there is an issue with the annotation whereas green ones are good.

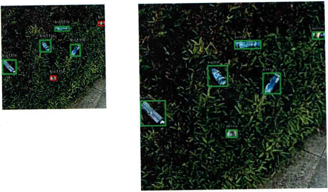


Figure 3. 12 Image upscaling for anchor adjustments

To get better results from retinaNet, random transform was also applied. It Randomly transforms images and annotations on the fly every time an image is passed to the network. In fact, fizyr implementation of keras-retinanet offers random-transform technique which apply image augmentation techniques such as rotation, translation, shear, flip, scaling, contrast, brightness, hue, and saturation. An example of a same image with random transform is shown in figure 3.13.

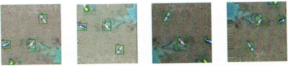


Figure 3. 13 Random-Transform Image Augmentation

Furthermore, transfer learning was used to initiate the model training and rest of the parameters were unchanged as fizyr implementation have slightly different default settings then the original model because it is fairly a simple model and optimized for less resourceful systems. We recommend a deeper dive to fizyr github repository in order to have better understanding of models' parameters. With these parameters, we were able to achieve a mAP value of 87.34%. A detailed analysis and compression are constructed in the Result and Evaluation section. RetinaNet was trained on our local system and training analysis is shown in the figure 3.14.

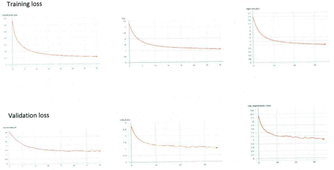


Figure 3. 14 Keras-RetinaNet Training and Validation Loss Graphs

3.3.5 Modeling – Yolov3-darknet

Yolov3 was trained on a linux environment on google colab since our local system is a Microsoft windows-based environment system and darknet framework is a linux based framework and does not officially support windows.

The first step was to convert dataset to darknet format. We need to generate the label files that Darknet uses. Darknet requires .txt file for each image with a line for each ground truth object in the image that looks like: <object-class> <x> <y> <width> <height> where x, y, width, and height are relative to the image's width and height. Again, I opted roboflow.ai for conversion. Converted annotations are shown in the fig.

```
"id": 0,
"image_id": 0,
"category_id": 1,
"bbox": [
    18,
    139,
    44.28662872314453,
    47.30224609375
],
"area": 2094.85701052472,
"segmentation": [],
"iscrowd": 0
```

```
0 0.7331786573978892 0.4975546675118786 0.1939681203741776 0.191299884639985
```


Figure 3. 15 UAV-BD MS-COCO to darknet '.txt' Conversion

Yolo was trained with max batches = 4000 because UAV-BDAI has only one class, so it must have enough time to train do proper detection. It up scales the images by default, but since yolo was not causing any anchor issues, images resizing was set to no resize for faster training. Steps were set to 80% and 90% or max batches. Filters in 3 of the convolutional layers above yolo layers were also set to optimum values that's (number of class x5) 3, which is 15. Rest of the training parameters were unchanged. With these parameters, we were able to achieve a mAP value of 76.89%. A detailed analysis and compression are constructed in the Result and Evaluation section. Training analysis of yolov3 is shown in the figure 3.16.

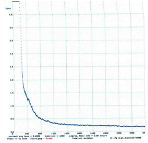


Figure 3. 16 Yolov3-darknet Training loss Graph

3.3.6 Modeling – Mask-RCNN

Mask-RCNN is a Faster-RCNN model by same author with slightly better performance and segmentation masks as shown in the figure 3.17. It is based on ResNet101 and FPN backbone. In our implementation of mask-rcnn, we trained the model with original UAV-BD MS-COCO dataset since it offers ground-truth segmentation masks for the bottles as well. But, since we are interested in bounding boxes only and used the provided coco training configurations, we reduced the complexity of the model by excluding few bottom layers like "mrcnn_class_logits", "mrcnn_bbox_fc", and "mrcnn_bbox", "mrcnn_mask" as these layers required matching number of classes as MS-COCO dataset which is 80 and our dataset has only 1 class. Mask-rcnn now mimics like fastet-rcnn with slightly better performance thus it will be called faster-rcnn from here on.

Figure 3. 17 Faster RCNN and Mask RCNN

It was trained on google colaboratory with resnet101 backbone, but it offers resnet50 as well. Resnet101 is more complex than resnet50 but since mask-rcnn is based on resnet101, it is more optimized than resnet50 for mask-rcnn. All the training parameters were unchanged with 10 epochs for the head meaning training for the classification layers since transfer learning was used to initiate the model, 15 epochs for fine tuning the resnet4+ layers and 20 epochs for fine tuning of all the layers, this helps the algorithm converge easier to reduce over fitting and model becomes more generalized.

We were able to achieve a mAP value of 90.34%. A detailed analysis and compression are constructed in the Result and Evaluation section. Training analysis is shown in the figure 3.18.

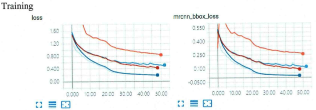


Figure 3. 18 Faster-RCNN Training Loss Graph

3.3.7 Ensembling – Models

For ensemble methods, we incorporate affirmative voting strategy, one of the many techniques proposed by Angela et al., This means that whenever one of the methods that produce the initial predictions says that a region contains an object, such a detection is considered as valid. This method of ensembling, ensembles the output of the prediction models, in this case which is bounding boxes. A lot of different methods have been discussed in the literature review session of this research, but all those techniques can be very challenging for a lot of user since it is hard to understand a group of models in ensemble. There is actually no learning happening here in this ensemble method technique rather ensembling of results produce by different algorithms are done. Affirmative technique was chosen because in the original article they claim to achieve up to 10% improvement from the 5 base models on general object classes using MS-COCO and PASCAL-VOC dataset. According to original article the affirmative strategy helps to greatly reduce the number of false negatives without considerably increasing the number of false positives. In our case, since we have only one class and 3 trained models, we were able to achieve improvement in diverse experiments which are discussed in results and evaluation section.

To successfully implement affirmative voting strategy, first the trained models were allowed to generate predictions in PASCAL-VOC style .xml files and then affirmative voting strategy was applied on these xml files to generate output prediction also in xml style. In the end these xml annotations were passed to PASCAL-VOC evaluation metrics with same threshold value and results were generated. Ensembling was done using cpu instead of gpu.

CHAPTER 4

RESULTS AND EVALUATION

This chapter will discuss the inference on these models, evaluation metrics that were used, comparison between different inferences on models and comparison with the results proposed in the literature.

4.1 SOME IMPORTANT DEFINITIONS

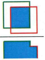
4.1.1 Intersection Over Union (IOU)

Intersection Over Union (IOU) is measure based on Jaccard Index that evaluates the overlap between two bounding boxes. It requires a ground truth bounding box B_{gt} and a predicted bounding box B_p. By applying the IOU we can tell if a detection is valid (True Positive) or not (False Positive). IOU is given by the overlapping area between the predicted bounding box and the ground truth bounding box divided by the area of union between them:

$$IoU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})}$$

The image below illustrates the IOU between a ground truth bounding box (in green) and a detected bounding box (in red).

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} =$$



4.1.2 True Positive, False Positive, False Negative and True Negative

Some basic concepts used by the metrics:

- **True Positive (TP)**: A correct detection. Detection with IOU $\geq$ *threshold*
- **False Positive (FP)**: A wrong detection. Detection with IOU < *threshold*
- **False Negative (FN)**: A ground truth not detected
- **True Negative (TN)**: Does not apply. It would represent a corrected misdetection. It is not used by the metrics.

threshold: depending on the metric, it is usually set to 50%, 75% or 95%.

4.1.3 Precision

Precision is the ability of a model to identify **only** the relevant objects. It is the percentage of correct positive predictions and is given by:

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{\text{all detections}}$$

4.1.4 Recall

Recall is the ability of a model to find all the relevant cases (all ground truth bounding boxes). It is the percentage of true positive detected among all relevant ground truths and is given by:

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{\text{all ground truths}}$$

4.1.5 Precision x Recall curve

This kind of curve is used by the PASCAL VOC 2012 challenge. The Precision x Recall curve is a good way to evaluate the performance of an object detector as the confidence is changed by plotting a curve for each object class. An object detector of a particular class is considered good if its precision stays high as recall increases, which means that if you vary the confidence threshold, the precision and recall will still be high. Another way to identify a good object detector is to look for a detector that can identify only relevant objects (0 False Positives = high precision), finding all ground truth objects (0 False Negatives = high recall).

A poor object detector needs to increase the number of detected objects (increasing False Positives = lower precision) to retrieve all ground truth objects (high recall). That is why the Precision x Recall curve usually starts with high precision values, decreasing as recall increases.

4.1.6 Average Precision

Another way to compare the performance of object detectors is to calculate the area under the curve (AUC) of the Precision x Recall curve. As AP curves are often zigzag curves going up and down, comparing different curves (different detectors) in the same plot usually is not an easy task - because the curves tend to cross each other much frequently. That is why Average Precision (AP), a numerical metric, can also help us compare different detectors. In practice AP is the precision averaged across all recall values between 0 and 1.

From 2010 on, the method of computing AP by the PASCAL VOC challenge has changed. Currently, the interpolation performed by PASCAL VOC challenge uses all data points, rather than interpolating only 11 equally spaced points as stated in their paper. As we want to reproduce their default implementation, this implementation follows their most recent application (interpolating all data points) of PASCAL-VOC.

4.1.7 Interpolating all points

Instead of interpolating only in the 11 equally spaced points, you could interpolate through all points in such way that:

$$\sum_{r=0}^{1} (r_{n+1} - r_n)\, \rho_{interp}(r_{n+1})$$

With

$$\rho_{interp}(r_{n+1}) = \max_{\tilde{r}:\tilde{r} \geq r_{n+1}} \rho(\tilde{r})$$

where $\rho(\tilde{r})$ is the measured precision at recall $\tilde{r}$.

In this case, instead of using the precision observed at only few points, the AP is now obtained by interpolating the precision at each level, r taking the maximum precision whose recall value is greater or equal than $r+1$. This way we calculate the estimated area under the curve.

The COCO challenge's variants to recall that the Pascal VOC challenge defines the mAP metric using a single IoU threshold of 0.5. However, the COCO challenge defines several mAP metrics using different thresholds, including:

- mAPIoU=.50:.05:.95mAPIoU=.50:.05:.95 which is mAP averaged over 10 IoU thresholds (i.e., 0.50, 0.55, 0.60, …, 0.95) and is the primary challenge metric.
- mAPIoU=.50mAPIoU=.50, which is identical to the Pascal VOC metric.
- mAPIoU=.75mAPIoU=.75, which is a strict metric.

4.2 INFERENCING MODELS

While running inferencing, all the testing parameters were set the same as of corresponding paper [21]. The results are interesting and discussed in this section.

For evaluation, PASCAL-VOC evaluation metrics was used for yolov3-darknet, keras-retinanet, faster-rcnn and ensemble methods whereas mmdetection pytorch based models faster-rcnn, retinanet and fcos were evaluated on COCO evaluation metrics. A set of separate 5078 test images were used for evaluation.

Implementation of PACAL-VOC evaluation metrics this research incorporates is originally proposed by Angela et al., [34] whereas for coco evaluation default mmdetection coco evaluation metric was used [51].

4.2.1 Inferencing – Models on PASCAL-VOC evaluation Metrics

4.2.1.1 Inferencing – Yolov3

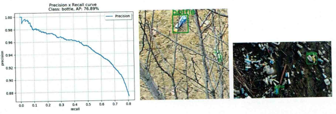


Figure 4. 1 Inference Results of Yolov3

In our testing we found that yolo still struggles with the small objects as we were able to achieve an AP value of 76.89%. In fact, in our case yolov3 performed slightly worse than the yolo2 proposed by the corresponding paper [21] where yolov2 achieve a slightly better performance of 77.4% of AP value.

4.2.1.2 Inferencing – RetinaNet

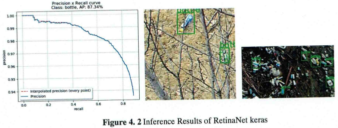


Figure 4. 2 Inference Results of RetinaNet keras

In our testing of keras-retinaNet, it performed way better than yolov3 and surpassed the AP score value with more than 10.45%. It scored an AP value of 87.34%.

4.2.1.3 Inferencing – Faster RCNN

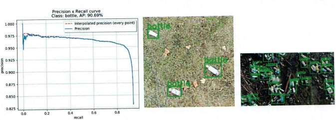


Figure 4. 3 Faster RCNN keras Inference Results

In our testing, faster-RCNN performed the beat out of all three models with an AP score of 90.7%, achieving a slightly better performance than the faster-RCNN model trained on UAV-BD by the base paper which achieved AP score value of 90.3%. That is 13.8% better than yolov3 and 3.35% better than keras-retinanet.

4.2.1.4 Ensembling – Yolo, keras-RetinaNet and Faster-RCNN

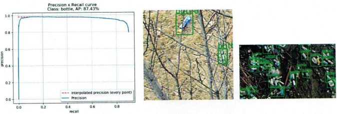


Figure 4. 4 Inference Results of Ensembling Yolov3 RetinaNet and Faster RCNN

When it came to ensembling, we first ensembled all off our trained models and were able to achieve an AP value of 87.43%. At first glance, it seems that there is no gain but if we see our base models' performance, the difference between performances of these models is quite high. If we compare it with the base models, we can see that by ensembling these models, the results have become more generalized and the performance is 10.54% better than yolov3 since it greatly reduce the number of false negatives without considerably increasing the number of false positives.

4.2.1.5 Ensembling – Yolov3 and keras-RetinaNet

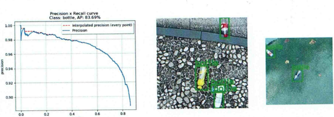


Figure 4. 5 Inference Results of Ensembling Yolov3 and RetinaNet

By ensembing yolov3 and keras-retinaNet we were able to achieve an AP value of 83.69%, That's better than yolov3 but worse than faster rcnn and almost identical to retinaNet.

4.2.1.6 Ensembling – Yolov3 and Faster-RCNN

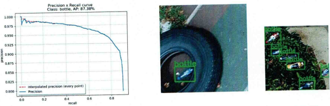


Figure 4. 6 Inference Results of Ensembling Yolov3 and Faster RCNN

By ensembling yolov3 with faster-rcnn, we were able to achieve identical performance to keras-retinanet that is 10.5% better yolov3.

4.2.1.7 Ensembling – keras-RetinaNet and Faster-RCNN

Figure 4. 7 Inference Results of Ensembling RetinaNet and Faster RCNN

But when we ensembled the 2 best models, we got overall better results with an AP value of 92%. That is overall, the best performance with 15% better performance than yolov3, 4.64% better than keras-retinanet and 1.3% better than faster-rcnn.

4.2.2 Comparison

Table 3. 1: Comparison of Inference results of our models with corresponding paper

Base Models	**AP values**	**Base Paper Models with HBB**	**AP values**	**Base Paper Models with OBB**	**AP values**	**Ensemble Models**	**AP values**
Yolov3	76.89%	Yolov2	77.4%	Yolov2	67.3%	Yolov3 + RetinaNet + FasterRCNN	87.43%
RetinaNet	87.34%	SSD	90.1%	SSD	87.6%	Yolov3 + RetinaNet	83.69%
Faster-RCNN	90.69%	Faster RCNN	90.3%	Faster RCNN	86.4%	Yolov3 + FasterRCNN	87.38%
				RRPN	88.6%	RetinaNet + FasterRCNN	**92%**

From the table above we can see that when two of the best models for instance retinanet and faster rcnn were combine in ensemble methods the results outperformed every other model and ensembling results either in our implementation or the results of models implemented in the base paper [21].

4.2.3 Inferencing – Mmdetection Models on COCO metrics – Benchmarks

These models were trained for comparison purpose only and are reserved for future work. They are evaluated on MS-COCO metrics without the plotting of Precision X Recall curve. In our testing of mmdetection models. All these models perform almost the same with IoU = 0.5:0.95, RetinaNet scored AP value of 73.3%, Faster RCNN scored slightly better with 73.7% and FCOS at the bottom with 72% AP value. More detailed results are shown in figure 3.26, figure 3.37 and figure 3.28.

4.2.3.1 RetinaNet

Average Precision (AP) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.733

Average Precision (AP) @[IoU=0.50 | area= all | maxDets=100] = 0.988

Average Precision (AP) @[IoU=0.75 | area= all | maxDets=100] = 0.878

Average Precision (AP) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.551

Average Precision (AP) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.752

Average Precision (AP) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.789

Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 1] = 0.547

Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 10] = 0.778

Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.778

Average Recall (AR) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.629

Average Recall (AR) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.797

Average Recall (AR) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.811

Inference Results of RetinaNet Using MMdetection

4.2.3.2 Faster RCNN

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.737
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.989
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.891
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.568
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.755
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.783
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.546
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.778
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.778
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.635
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.796
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.807
""""
```

Inference Results of Faster RCNN Using MMdetection

4.2.3.3 FCOS

Average Precision (AP) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.720
Average Precision (AP) @[IoU=0.50 | area= all | maxDets=100] = 0.987
Average Precision (AP) @[IoU=0.75 | area= all | maxDets=100] = 0.865
Average Precision (AP) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.522
Average Precision (AP) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.741
Average Precision (AP) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.775
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 1] = 0.542
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets= 10] = 0.769
Average Recall (AR) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.769
Average Recall (AR) @[IoU=0.50:0.95 | area= small | maxDets=100] = 0.597
Average Recall (AR) @[IoU=0.50:0.95 | area=medium | maxDets=100] = 0.791
Average Recall (AR) @[IoU=0.50:0.95 | area= large | maxDets=100] = 0.805

Inference Results of FCOS Using MMdetection

4.3.3 Interpretation

Since, plastic pollution poses long term health, environmental and economic issues, technologies like AI and computer vision has stepped in to track our plastic waste. One way to do it is by using object detecting through UAVs in order to successfully locate plastic. This research shows that, how UAV systems can easily be used for object detection analysis to track bottles that are one of the top 3 most abundant plastic waste material by using state of the art object detecting algorithms in a technique called ensemble methods to further increase their object detection performance. In this research we set new benchmarks and were able to outperform corresponding models trained by us as well as outperforming corresponding research papers techniques and models by achieving an AP value of 92%.

This research further shows that for any given object detecting task, preparing, and transforming dataset hold significant importance. Further, in our findings we were able to show that when ensembling different objecting models, the choice of model selecting is crucial as ensembling weaker model with a stronger one tends increase weaker models' performance but also decreases stronger models' performance thus overall performance tends to decrease but model becomes more generalized.

Further, a comparison table was constructed through literature review in order to select models for ensembling. In the end, some models were trained using mmdetection code library which is based on pytorch framework and their benchmarks are presented.

CONCLUSION

Ensemble methods could be a challenging task but not as challenging as optimizing model's architecture to get better predictions. In fact, it could be very easy to ensemble the output predictions of the models using ensembling strategies like voting because implementing ensemble methods that depends on the nature of the algorithms employed to construct the detection models could be challenging task for most users as there are a lot of complications associated with it. We can see from the results of all our inferences, simply ensembling the models that perform similar helps us gain performance boost with respect to accuracy as suggested by the original paper as well [34]. For instance, in our testing faster rcnn and keras-retinanet were the top candidates and when combined in ensemble methods they outperformed any other model in our testing. Similarly, ensembling a weaker model with a stronger one could increase the accuracy results compared to the weaker model but also reduce the accuracy results compared to stronger model as it can be seen from the results of yolov3 with keras-retinanet and yolov3 with faster-rcnn. So, choosing the right models for ensembling is crucial.

Furthermore, from all our experimentations and testing we are able to verify that first most important task in any given object detection challenge is preparation of dataset because it could lead to sub optimal results such as when the UAV-BD was fed to models without cleaning it, caused a lot if issues and even produced wage results.

Last but not least, implementations of these models offered by latest model libraries like mmdetection are well optimized that produces results better than original code libraries. The models we have trained are here to set benchmarks for now but could be used in ensemble methods for future work.

FUTURE WORK

Adding more dataset holds important value as deep learning algorithms are very data demanding and since there are no publicly available plastic datasets, generating and labeling own dataset like UAV-BD dataset could be done in the future. For now, we were focused on bottle class only, since it is one of the top 3 most abundant plastic waste material but adding more classes in dataset could help us in identifying and reduce more plastic waste. We have ensembled three models and showed that how easy it is to ensemble the output of different models for increasing the accuracy for small objects, adding more models in ensemble methods could help in further increase overall accuracy performance. The mmdetection library models we have trained could also be used in ensemble learning in the future. Optimizing models before ensembling could also help us gain more performance boost. We have not tested these results on real time data to measure detecting speed, it could be done in the future as well. Finally, this was supervised learning, techniques like semi-supervised learning could be apply in the future to make the models learn on real time data in real time thus further increase their performance.

REFRENCES

[1] K. Singh and A. Mittal, "Soil Stabilisation Using Plastic Waste," in *Lecture Notes in Civil Engineering*, 2019, vol. 32, pp. 91–96.

[2] P. Senthil Kumar and G. Janet Joshiba, "Properties of Recycled Polyester," in *Recycled Polyester: Manufacturing, Properties, Test Methods, and Identification*, S. S. Muthu, Ed. Singapore: Springer Singapore, 2020, pp. 1–14.

[3] C. Orset, N. Barret, and A. Lemaire, "How consumers of plastic water bottles are responding to environmental policies?," *Waste Manag.*, vol. 61, pp. 13–27, 2017.

[4] GESAMP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection), "Sources, fate and effects of microplastics in the marine environment: a global assessment (Kershaw, P. J., ed.). (IMO/FAO/UNESCO-IOC/UNIDO/WMO/IAEA/UN/UNEP/UNDP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection)," *Rep. Stud. GESAMP*, vol. No. 90, p. 96 p., 2015.

[5] V. Bisinella, P. F. Albizzati, T. F. Astrup, and A. Damgaard, "Executive summary," *New Dir. Youth Dev.*, vol. 2008, no. 120, pp. 7–12, 2008.

[6] J. Parfitt, M. Barthel, and S. MacNaughton, "Food waste within food supply chains: Quantification and potential for change to 2050," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 365, no. 1554, pp. 3065–3081, 2010.

[7] J. Lorenzo-Navarro, M. Castrillón-Santana, M. Gómez, A. Herrera, and P. A. Marín-Reyes, "Automatic counting and classification of microplastic particles," *ICPRAM 2018 - Proc. 7th Int. Conf. Pattern Recognit. Appl. Methods*, vol. 2018-Janua, no. Icpram, pp. 646–652, 2018.

[8] S. E. Nelms *et al.*, "Microplastics in marine mammals stranded around the British

plastic?," *The Guardian*, 2019.

[20] B. R. S. Kumar, N. Varalakshmi, S. S. Lokeshwari, K. Rohit, Manjunath, and D. N. Sahana, "Eco-friendly IOT based waste segregation and management," in *International Conference on Electrical, Electronics, Communication Computer Technologies and Optimization Techniques, ICEECCOT 2017*, 2018, vol. 2018-Janua, pp. 297–299.

[21] J. Wang, W. Guo, T. Pan, H. Yu, L. Duan, and W. Yang, "Bottle Detection in the Wild Using Low-Altitude Unmanned Aerial Vehicles," *2018 21st Int. Conf. Inf. Fusion, FUSION 2018*, vol. 2, pp. 439–444, 2018.

[22] G. E. Sakr, M. Mokbel, A. Darwich, M. N. Khneisser, and A. Hadi, "Comparing deep learning and support vector machines for autonomous waste sorting," in *2016 IEEE International Multidisciplinary Conference on Engineering Technology, IMCET 2016*, 2016, pp. 207–212.

[23] T. Cheshire, "Drones used in fight against plastic pollution on UK beaches," *Sky News*, p. 1, 2017.

[24] S. Harvey and R. Harvey, "Introduction to artificial intelligence," *Appita J.*, vol. 51, no. 1, pp. 20–24, 1998.

[25] University of Florida, "Standardized Syllabus for the College of Engineering," *BME 6938 Mach. Learn. Heal. Biomed. Appl. 1.*, 2014.

[26] M. Nixon and A. S. Aguado, *Feature Extraction & Image Processing for Computer Vision*. Academic Press, 2012.

[27] L. Cao, "Data science: A comprehensive overview," *ACM Computing Surveys*, vol. 50, no. 3. Association for Computing Machinery, pp. 1–42, 01-Jun-2017.

[28] L. Liu *et al.*, "Deep Learning for Generic Object Detection: A Survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261–318, Feb. 2020.

[29] I. Gogul and V. S. Kumar, "Flower species recognition system using convolution neural networks and transfer learning," *2017 4th Int. Conf. Signal Process.*

Commun. Networking, ICSCN 2017, 2017.

[30] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised Classification Algorithms in Machine Learning: A Survey and Review," in *Advances in Intelligent Systems and Computing*, 2020, vol. 937, pp. 99–111.

[31] W. Liu *et al.*, "SSD: Single shot multibox detector," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9905 LNCS, pp. 21–37, 2016.

[32] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," 2018.

[33] Y. Liang *et al.*, "TFPN: Twin feature pyramid networks for object detection," in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 2019, vol. 2019-Novem, pp. 1702–1707.

[34] J. Heras and A. Casado-Garcia, "Ensemble Methods for Object Detection," *24th Eur. Conf. Artif. Intell.*, 2020.

[35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[36] G. Mittal, K. B. Yagnik, M. Garg, and N. C. Krishnan, "SpotGarbage: Smartphone app to detect garbage using deep learning," *UbiComp 2016 - Proc. 2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, vol. 7, no. 1, pp. 940–945, 2016.

[37] A. Stumpf, N. Lachiche, J. P. Malet, N. Kerle, and A. Puissant, "Active learning in the spatial domain for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2492–2507, 2014.

[38] S. L. N. Rafael Padilla and E. A. B. da Silva, "Survey on Performance Metrics for Object-Detection Algorithms," *Int. Conf. Syst. Signals Image Process.*, 2020.

[39] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 6517–6525, 2017.

[40] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object Detection with Deep Learning:

A Review," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, 2019.

[41] S. Lin, J. Wang, R. Peng, and W. Yang, "Development of an autonomous unmanned aerial manipulator based on a real-time oriented-object detection method," *Sensors (Switzerland)*, vol. 19, no. 10, p. 2396, May 2019.

[42] T. Ringwald, L. Sommer, A. Schumann, J. Beyerer, and R. Stiefelhagen, "UAV-net: A fast aerial vehicle detector for mobile platforms," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2019-June, pp. 544–552, 2019.

[43] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Visual object detection with deformable part models," *Commun. ACM*, vol. 56, no. 9, pp. 97–105, 2013.

[44] G. Liu, "Application of digital image processing technology in crack detection of ceramic bottles," *Rev. la Fac. Ing.*, vol. 32, pp. 672–683, 2017.

[45] M. Farrukh, I. Ahmed Halepoto, B. S. Chowdhry, H. Kazi, and B. Lal, "Design and Implementation of PLC based Automatic Liquid Distillation System," *Indian J. Sci. Technol.*, vol. 10, no. 29, pp. 1–6, 2017.

[46] W. P. Boshoff and R. Combrinck, "Modelling the severity of plastic shrinkage cracking in concrete," *Cem. Concr. Res.*, vol. 48, pp. 34–39, 2013.

[47] P. Soviany and R. T. Ionescu, "Optimizing the trade-off between single-stage and two-stage deep object detectors using image difficulty prediction," *Proc. - 2018 20th Int. Symp. Symb. Numer. Algorithms Sci. Comput. SYNASC 2018*, pp. 209–214, Mar. 2018.

[48] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Aug. 2020.

[49] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 9626–9635, Apr. 2019.

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-Decem, pp. 770–778.

[51] K. Chen *et al.*, "MMDetection: Open MMLab Detection Toolbox and Benchmark," Jun. 2019.

[52] T. M. Dias *et al.*, "Autonomous detection of mosquito-breeding habitats using an unmanned aerial vehicle," in *Proceedings - 15th Latin American Robotics Symposium, 6th Brazilian Robotics Symposium and 9th Workshop on Robotics in Education, LARS/SBR/WRE 2018*, 2018, no. February, pp. 357–362.

[53] C. C. Hsu, Y. X. Zhuang, and C. Y. Lee, "Deep fake image detection based on pairwise learning," *Appl. Sci.*, vol. 10, no. 1, p. 370, Jan. 2020.

[54] C.-H. Lu, "Applying Uav and Photogrammetry To Monitor the Morphological Changes Along the Beach in Penghu Islands," *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLI-B8, pp. 1153–1156, Jun. 2016.

[55] S. H. Bak, D. H. Hwang, H. M. Kim, and H. J. Yoon, "Detection and monitoring of beach litter using uav image and deep neural network," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 2019, vol. 42, no. 3/W8, pp. 55–58.

[56] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2241–2248.

[57] V. A. Dave and S. K. Hadia, "Automatic Bottle Filling Inspection System Using Image Processing," *Int. J. Sci. Res.*, vol. 4, no. 4, pp. 1116–1120, 2015.

[58] C. Gandrud and C. Gandrud, "Preparing Data for Analysis," *Reprod. Res. with R RStudio*, no. March, pp. 129–149, 2019.

[59] Z. Huang, L. Huang, Y. Gong, C. Huang, and X. Wang, "Mask scoring R-CNN," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 6402–6411, Mar. 2019.

[60] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8691 LNCS, no. PART 3, pp. 346–361, Jun. 2014.

[61] xuemei xie, G. Cao, W. Yang, Q. Liao, G. Shi, and J. Wu, "Feature-fused SSD: fast detection for small objects," p. 236, Sep. 2018.

[62] Z. Cai and N. Vasconcelos, "Cascade R-CNN: High Quality Object Detection and Instance Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, Jun. 2019.

[63] X. Lu, B. Li, Y. Yue, Q. Li, and J. Yan, "Grid R-CNN Plus: Faster and Better," Jun. 2019.

[64] X. Zhang, F. Wan, C. Liu, R. Ji, and Q. Ye, "FreeAnchor: Learning to Match Anchors for Visual Object Detection," Sep. 2019.

[65] J. Wang, K. Chen, S. Yang, C. C. Loy, and D. Lin, "Region proposal by guided anchoring," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 2960–2969, Jan. 2019.

[66] G. Ghiasi, T. Y. Lin, and Q. V. Le, "NAS-FPN: Learning scalable feature pyramid architecture for object detection," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 7029–7038, Apr. 2019.

[67] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, vol. 2015 Inter, pp. 1440–1448.

[68] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, and D. Lin, "Libra R-CNN: Towards balanced learning for object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, vol. 2019-June, pp. 821–830.

[69] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, "Soft-NMS -- Improving Object Detection With One Line of Code," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-October, pp. 5562–5570, Apr. 2017.

[70] G. Thung and M. Yang, "Classification of Trash for Recyclability Status," in *cs229.stanford.edu*, 2016.

[71] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," Dec. 2016.