UNIVERSITY OF NICOSIA

Utilising Artificial Intelligent (AI) Tools to Discern Plastic Debris at the Marine Environment

Ms. Kyriaki Kylili

PhD (Doctor of Philosophy) in Electrical Engineering

September, 2023

Kyriaki Kylili

NICOSIA
PhD
2023

Utilising Artificial Intelligent (AI) Tools to Discern Plastic Debris at the Marine Environment

Ms. Kyriaki Kylili

A thesis submitted to the University of Nicosia

in accordance with the requirements of the degree of

PhD (Doctor of Philosophy) in Electrical Engineering

Department of Engineering

University of Nicosia

September, 2023

Abstract

The increased desire of society for plastic products has led to plastic becoming omnipresent in the marine environment. Due to its toxicity, abundance, and persistence at the sea, plastic debris is of particular concern to marine life mainly owing to the elevated risk of entanglement or ingestion which may prove lethal. In this study, various machine learning (ML) and deep learning (DL) tools were implemented for classifying, detecting and counting plastic marine debris in images and video recordings. Primarily, the importance of this research lies in the intelligent and swift detection and counting of plastic litter which can facilitate litter monitoring surveys around the world and improve estimations of plastic marine debris density in remote geographical areas.

Initially, the Bag of Features (BoF) method was used to construct a plastic debris image classifier, which attained a classification accuracy of 62.5% when applied to images demonstrating plastic debris and sea life from 8 classes. However, motivated by the need to enhance the sophistication of the proposed classifier, the DL tool of the Bottleneck method (BM) was employed. By expanding the number of marine debris object classes that the classifier can recognise from 3 to 8, the BM attained a 4% improvement in the validation accuracy, which topped 90%. Interestingly, when the resolution of the examined images was lowered by 75% of their original size, the accuracy of the BM remained unchanged.

Beyond the accurate classification of marine debris, detecting multiple objects in images and videos is of paramount importance. Charged with this task, the YOLOv5 tool proved the most successful, among the YOLO family, as it attained real-time object detection of 34 frames per second on video footage and realised the highest mean average precision of 92.4% on images. Coupled with the counting tools of the region of interest (ROI) line and the centroid tracking, YOLOv5 proved competent in counting marine debris items from video footage. Particularly, the centroid tracking tool realised an accuracy of ≈80% when it processed a video illustrating plastic litter from 9 classes.

Lastly, for estimating the plastic litter density across the Cypriot coastlines and the litter's physical dimensions, the YOLACT++ tool was utilised as it applies a mask on each detected litter item. Inspecting images depicting plastic litter from six beaches in Cyprus, this pertinent detector deduced a plastic litter density of 0.035 items/m^2. Extrapolating to the entire shorelines of Cyprus, the YOLACT++ tool estimated about 66,000 plastic items weighing a total of about 1,000 kg as explained in the pertinent video https://bit.ly/35TQVVE. Concluding, the dominant length of all documented plastic litter ranged from 10 to 30 cm.

Keywords: Artificial intelligence, Centroid tracking, Deep learning, Environmental monitoring, Image classification, Marine pollution, Plastic debris detection, Region of interest, Plastic litter density, Size of plastics.

Acknowledgments

First of all, I would like to express my gratitude and special appreciation to my advisor Associate Professor Constantinos Hadjistassou for all his valuable help, guidance and support during my PhD studies. His support and guidance during the last four years encouraged me to carry out my research and helped me grow as a research scientist. Besides my advisor, I would like to thank my two co-supervisors Dr Alessandro Artusi and Associate Professor Ioannis Kyriakides for providing interesting and valuable feedback throughout my studies. Furthermore, I would like to thank Ms. Despo Kyriakoudi for her contribution to the research of microplastics and Ms Cathy Teng for conducting research related to the counting of plastic litter captured in images and video footage.

Additionally, I acknowledge financial support in the form of a full scholarship from the University of Nicosia as well as financial aid from the A. G. Leventis foundation. The generous support from these scholarships helped me realise my dream of earning a doctorate. Furthermore, I would like to express my appreciation to the Universitas Foundation for its financial assistance as part of the PlasticMed project. Lastly, I would like to thank the Algalita and the Debris Tracker organisations for kindly sharing with us images of marine debris.

Special thanks go to my parents, my sister, my brothers and my adored friends for their endless love and encouragement during all phases of my life. Words cannot express how grateful I am to my parents for all the sacrifices that they have made for me. I would also like to express my sincere appreciation to my beloved husband Adamos who has always been there for me, understanding me and simply being a tower of strength for me. The trust he has shown me has given me the opportunity to pursue my dreams.

This thesis is dedicated to my most treasured daughter Agathi and my son Prodromos for teaching me what unconditional love means.

Declaration

I declare that the work in this thesis was carried out in accordance with the regulations of the University of Nicosia. This thesis has been composed solely by myself except where stated otherwise by reference or acknowledgment. It has not been previously submitted, in whole or in part, to this or any other institution for a degree, diploma or other qualifications.

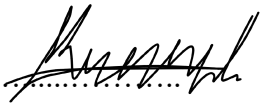
Signed ……  …… Date ……09/10/2023……

Table of Contents

Nomenclature

AI	Artificial Intelligence.
AUV	Autonomous Underwater Vehicle.
BCE	Binary Cross-Entropy (loss).
BM	Bottleneck Method.
BoF	Bag of Features (method).
BRISK	Binary Robust Invariant Scalable Keypoints detector and descriptor.
CA	Cellulose Acetate.
CIoU	Complete Intersection over Union.
CNN	Convolutional Neural Network.
CSPDarknet53	Cross Stage Partial Darknet 53 (architecture).
CSPNet	Cross Stage Partial Network.
DA	Data Augmentation (manipulations).
DI	Directional Index (of a hydro-acoustic device).
DL	Deep Learning.
DT	Detection Threshold (of a hydro-acoustic device).
FAST	Features from the Accelerated Segment Test.
FLM App	Floating Litter Monitoring Application.
FLS	Forward-Looking Sonar.
FPN	Feature Pyramid Networks.
FPS	Frames Per Second.
FT-IR	Fourier Transform Infrared Spectroscopy.
GIoU	Generalised Intersection over Union.
GPS	Global Positioning System.
HARRIS	Chris Harris and Mike Stephens corner and edge detector.
HDPE	High-density polyethylene.
HSV	Hue and Saturation Value.
IoU	Intersection over union.
LDPE	Low-density polyethylene.

LIDAR	Light Detection and Ranging technology.
mAP	Mean Average Precision.
MinEigen	Minimum Eigenvalue algorithm.
ML	Machine Learning.
MSCOCO	Microsoft COCO (common objects in context) image dataset.
MSER	Maximally Stable Extremal Regions technique.
MSW	Municipal Solid Waste.
NL	Noise Level (of a hydro-acoustic device).
NMS	Non-maximum suppression.
PA	Polyamide or commonly called nylon.
PANet	Path Aggregation Network.
PE	Polyethylene.
PP	Polypropylene.
PS	Polystyrene.
PVC	Poly-Vinyl chloride.
ReLU	Rectified Linear Unit.
ResNet-101	Residual Networks-101.
ROI	Region of Interest.
ROV	Remotely Operated Vehicle.
SGD	Stochastic Gradient Descent.
SL	Source Level (of a hydro-acoustic device).
SPP	Spatial Pyramid Pooling.
SSD	Sum Squared Difference.
SURF	Speeded-Up Robust Features.
SVM	Support Vector Machine.
TL	Transmission Loss (of a hydro-acoustic device).
UAV	Unmanned Aerial Vehicle.
VIA	Visual Geometry Group (VGG) Image Annotator (VIA) tool.
VGG16	Visual Geometry Group-16 (pre-trained model).
YOLACT++	You Only Look at Coefficients++ (algorithm).

YOLO	You Only Look Once algorithm.
YOLO_mark	You Only Look Once annotation tool.
YOLOv2	You Only Look Once algorithm version 2.
YOLOv3	You Only Look Once algorithm version 3.
YOLOv4	You Only Look Once algorithm version 4.
YOLOv5	You Only Look Once algorithm version 5.

List of Figures

List of Tables

List of Appendices

List of Publications

Publications, Posters and Abstracts:

Cathy Teng, Kyriaki Kylili and Constantinos Hadjistassou. 2022. Deploying deep learning to estimate the abundance of marine debris from video footage. *Marine Pollution Bulletin*, 183, 114049, https://doi.org/10.1016/j.marpolbul.2022.114049.

Kyriaki Kylili, Alessandro Artusi and Constantinos Hadjistassou. 2021. A new paradigm for estimating the prevalence of plastic litter in the marine environment. *Marine Pollution Bulletin*, 173, 113127, https://doi.org/10.1016/j.marpolbul.2021.113127.

Kyriaki Kylili, Constantinos Hadjistassou and Alessandro Artusi. 2020. An intelligent way for discerning plastics at the shorelines and the seas. *Environmental Science and Pollution Research*, 27 (34), 42631–42643, https://doi.org/10.1007/s11356-020-10105-7.

Kyriaki Kylili, Ioannis Kyriakides, Alessandro Artusi and Constantinos Hadjistassou. 2019. Identifying floating plastic marine debris using a deep learning approach. *Environmental Science and Pollution Research*, 26 (17), 17091–17099, https://doi.org/10.1007/s11356-019-05148-4.

Kylili Kyriaki, Artusi Alessandro, Kyriakides Ioannis and Hadjistassou Constantinos. Tracking and identifying floating marine debris. *Sixth International Marine Debris Conference*, San Diego California USA, March 12th–16th, 2018.

Kyriaki Kylili, Ioannis Kyriakides, Constantinos Hadjistassou and Alessandro Artusi. Identifying floating debris at sea. *CWSPI Workshop*, University of Cyprus, July 19th, 2017.

Trainings, Talks, Presentations, and Videos:

Kylili Kyriaki and Hadjistassou Constantinos. 2022. Application of smart techniques (AI) to understand and address the problem of marine plastics. ECOMARINE Training Course, *University of Cyprus*, September 9th–11th, 2022.

Plastic pollution littering Cyprus coastlines. 2021. Video can be viewed online at: https://bit.ly/35TQVVE.

Kylili Kyriaki and Hadjistassou Constantinos. Plastic Marine Pollution, Emissions Levels and Transport Electrification in Cyprus. *University of Nicosia*, December 9th, 2021.

Kylili Kyriaki. Utilising Artificial Intelligent (AI) Tools to Discern Plastic Debris at the Marine Environment. *University of Nicosia*, November 17th, 2021.

Kylili Kyriaki. Can Artificial Intelligence (AI) Help Address the Problem of Marine Plastics? *University of Nicosia*, November 5th, 2020.

Kylili Kyriaki, Artusi Alessandro and Hadjistassou Constantinos. Smart ways for discerning marine debris at the sea and the shorelines. *HERMES Virtual International Conference*, Zoom Business, June 18th–19th, 2020.

Kylili Kyriaki and Hadjistassou Constantinos. Utilise deep learning algorithms for classifying and localising marine litter in the marine environment. *Researcher's night, Filoxenia Conference Centre*, Nicosia, September 27th, 2019.

Kylili Kyriaki and Hadjistassou Constantinos. Smart detection of marine plastics using machine learning tools. *Researcher's night, Lanitis Carob Mills*, Limassol, September 28th, 2018.

Kylili Kyriaki. Employing machine learning tools for detecting plastic litter in images. *University of Nicosia*, May 15th, 2018.

Kylili Kyriaki and Hadjistassou Constantinos. Do we live in a paradise of plastics? *Researcher's night, Cyprus State Fair*, Nicosia, September 29th, 2017.

CHAPTER 1 INTRODUCTION

1. Introduction

Plastic pollution in the marine environment is a multi-faceted concern that transcends national boundaries. Ever since the advent of the commercial scale production of plastics dating back in the 1950s, the yearly output of polymer-based materials has swollen from 1.7 million tonnes in 1950 to 359 million tonnes in 2018, experiencing an average annual growth rate of 8.5% (PlasticsEurope, 2013; PlasticsEurope, 2019). Accompanying this spectacular growth, some of the plastic post-consumer waste which ends-up in the marine environment has surged (Ostle *et al.*, 2019). According to some estimates, about 150 million tonnes of plastic debris is currently circulating in the world's oceans (Jambeck *et al.*, 2015). Projections allude that with the prevailing waste management policies, the accumulation of plastic objects is expected to soar to a staggering 600 million tonnes by 2040 (PEW Charitable Trusts *et al.*, 2020).

Alarmed by the detrimental effects of plastic litter in the marine environment, researchers have zeroed in on the need to quantify the abundance of plastic litter at the seas and the shores and endeavoured to clarify the possible threats that pose to marine life and, indirectly, to humans. Aiming to stem both terrestrial and marine pollution from plastic, the European Union banned plastic straws, stirrers, single-use plastic cutlery and cotton buds in 2021. However, besides macroplastics, which encompass plastic litter measuring longer than 5 mm in length, microplastics also exist in the marine environment and undesirably affect marine flora and fauna. More specifically, microplastics are plastic pieces or fragments with dimensions less than 5 mm (Barnes *et al.*, 2009) subdivided into primary or secondary in origin.

In-situ observations (Barnes *et al.*, 2009; Cózar *et al.*, 2014; Lebreton *et al.*, 2018) and ocean circulation models (Maximenko *et al.*, 2012; van Sebille *et al.*, 2012) can help better pinpoint the areas that harbour high concentrations of marine debris. Floating marine litter tends to congregate in ocean gyres, the so-called garbage patches (Cózar *et al.*, 2014; Newman, 2014) with reported densities of 890,000 items/km^2 (Eriksen *et al.*, 2014). Another geographical hotspot of plastic detritus is the Mediterranean Sea. Perceived to host almost 250,000 items/km^2, the concentration

of plastics in the Mediterranean Sea is akin to the amount encountered in the inner accumulation zones of the ocean gyres, which ranges between 162,000 to 368,400 items/km^2 (Cózar *et al.*, 2015). Owing to its abundance and persistence in the marine environment, plastic debris is of particular concern to marine life mainly due to the elevated risk of entanglement or ingestion which can prove lethal to fish and birds (Nelms *et al.*, 2015; Wilcox *et al.*, 2016).

Through the years, researchers mainly applied manual methods to address the problem of plastic pollution. Manual methods are primarily based on in-situ beach or sea collection of disposed waste items and visual counting of plastic litter along transects (Ryan *et al.*, 2009; Rosevelt *et al.*, 2013). However, current methods for collecting and classifying plastic debris at sea comprise mainly manual approaches and visual observations (Barnes *et al.*, 2009; Goldstein *et al.*, 2013). Such sampling methods are time-consuming and utilise a considerable number of volunteers and researchers to cope with the demanding task of collecting, counting and classifying marine debris. Furthermore, manual monitoring campaigns usually sample small polluted areas compared to the large extent of littered marine zones. Lastly, the utilisation of research vessels and boats for tracing and retrieving litter floating on or under the sea surface requires substantial financial resources which confine them to small areas (for more information please refer to §1.1).

At the same time though, technologically assisted approaches which predominantly rest on a combination of unmanned aerial vehicle (UAV) surveys and machine or deep learning-based image classifiers are used to detect plastic litter at the seas and the coastlines. For instance, the team of Martin *et al.* (2018) employed a machine learning (ML) image classifier, namely, the random forest classifier to categorise three types of plastic litter, e.g., plastic containers, cups and bags, from airborne imagery. Moreover, Jakovljevic *et al.* (2020) utilised various deep learning (DL) tools like the Residual Networks 50 (ResNet50) (He *et al.*, 2016), the ResNeXt50 (Xie *et al.*, 2017), the Inception-ResNet v2 (Szegedy *et al.*, 2017), and the Extremely Inception (Xception) (Chollet, 2017) to identify marine litter from UAV footage. More specifically, the team of Jakovljevic *et al.* (2020) differentiated three

plastic materials such as Polyethylene terephthalate (PET), Nylon (PA), and Oriented Polystyrene (OPS). Although, such techniques can recognise litter items belonging to generic marine debris classes yet they are unable to differentiate plastic litter among its distinct types.

Appreciating the need of monitoring marine litter around the world, smart tools from the artificial intelligence (AI) domain can play an instrumental role. Thus, this research work has created various smart image classifiers and intelligent object detectors, using machine learning (ML) and deep learning (DL) tools, respectively, that can efficiently classify, detect, localise and count plastic litter from images and video recordings. Ultimately, this research aims to smartly and swiftly identify marine debris over large littered areas while mapping with more accuracy their abundance. Particular emphasis was allocated to the detection of plastic debris simply because plastics constitute the main type of anthropogenic litter in marine ecosystems. A couple of studies, which used manual debris retrieval and counting at the seas and the shorelines, have revealed that plastics make up 41% to 80% of the overall marine debris items (Barnes *et al.*, 2009; Pham *et al.*, 2014). Interestingly, estimates of the abundance of marine litter offered by this research were obtained without physically retrieving the litter items from the marine environment. Remarkably, such estimates were made possible by applying the DL tools on real field images and videos.

1.1. Problem definition

The advent of plastics is one of the world's most important industrial advances. Their applications span from various domains like electrical and electronic gear, packaging, agriculture, construction, automobiles, and healthcare. Owing to their versatile nature and ease of processing, the yearly output of plastics has swollen from 1.7 Mt, in 1950, to 359 Mt, in 2018 (PlasticsEurope, 2013; PlasticsEurope, 2019). Of these, 51% was manufactured in Asia, in 2018. Naturally, plastics have become an essential part of modern life mainly because they make or are part of a wide spectrum of daily life materials. However, due to their light, flexible, strong, durable, and undiluted nature, plastics are persistent in nature and the marine environment causing

various adverse consequences (Andrady, 2015; Letcher, 2020).

Once plastics enter into the marine environment deliberately or inadvertently by humans, they are frequently displaced over long distances and can reach places far away from the point that they have been initially dumped. Due to their exposure to the sun, wind and water waves, floating plastic litter may fragment into smaller pieces (Andrady, 2015; Letcher, 2020). As expected, the plastic litter or fragments whose density is greater than that of the seawater, will sink and may settle in marine sediments. Still, an appreciable amount of floating plastic litter whose density is less than that of the seawater will develop on their surfaces a biofilm of bacteria and diatoms (Lobelle *et al.*, 2011). Soon, a variety of marine organisms, such as, hydroids, bryozoans, barnacles, and ectocarpales will colonise the surface biofilm (Pabortsava *et al.*, 2020; Póvoa *et al.*, 2021). Finally, depending on the mass of the biofouling creatures, floating plastic litter will eventually sink in seawater until they suspend themselves in the mid-water column or rest on the seabed.

Due to their durability, plastics could persist at sea from several centuries to a thousand years (Boucher *et al.*, 2017). Literally, this means that almost all of the plastic items that have been manufactured and entered the marine environment are still there. Strikingly, predictions indicate that with the current waste management policies, the accumulation of plastics in the oceans is poised to triple to 250 billion tonnes by 2025 (Jambeck *et al.*, 2015), if no decisive interventions are made by governments and world leaders. By one estimate, in year 2050 there will be more plastic litter, by weight, in the marine environment than fish (Ellen MacArthur Foundation, 2016; PEW Charitable Trusts *et al.*, 2020).

Owing to the increasing abundance of plastics in the marine environment and the adverse consequences that may cause to marine life and the marine environment, several strategies have been proposed and adopted worldwide to combat plastic pollution. Among them are the Circular Economy (CE) Action Plan and the Single-Use Plastics Directive (SUPD) proposed by the European commission. Such strategies aim to prohibit single-use plastics and curtail their use as well as recycle

and reuse them (European Commission, 2016; European Commission, 2018; European Union, 2019).

At the same time though, various studies attempted to estimate the abundance of plastic debris at the world's seas and oceans. Floating plastic debris tend to congregate in the five subtropical ocean gyres, namely, the North and South Atlantic gyre, the North and South Pacific gyre, and the Indian Ocean gyre (Cózar *et al.*, 2014; Newman, 2014). Specifically, in the North Pacific gyre, better known as the "world's largest landfill" (Day *et al.*, 1987; Harse, 2011), it is estimated that about 18 kg of plastic particles exist per km^2 (Day *et al.*, 1987; Harse, 2011). Additionally, in the South Pacific ocean gyre it was reported that about 71 g of marine plastics float per km^2 (Eriksen *et al.*, 2013).

Another geographical hot spot of plastic detritus is the Mediterranean Sea, which contains no less than 423 g of plastic debris per km^2— a density comparable to the inner accumulation zones of the five subtropical ocean gyres (Cózar *et al.*, 2015). Besides the sea surface, the ocean floor is deemed to harbour a high concentration of plastics. According to Law (2016), the average density of plastic litter observed on the ocean bottom was projected to be 70 kg per km^2, which is almost 4 times higher than the plastic litter density found in the five subtropical ocean gyres.

The persistence of plastics in the marine environment can cause harmful effects to the ecosystem and its biodiversity, while leading to economic and social consequences. One adverse impact of plastics at seas is the formation of artificial substrates for colonisation. The colonisation of floating plastics results in the introduction of new marine organisms in other geographical areas thereby altering the endemic local biodiversity. For instance, a variety of alien and invasive species, such as bryozoans, barnacles, isopods, and boring molluscs, were observed in the Atlantic Ocean (Barnes *et al.*, 2005), while in the Pacific Ocean, molluscs, cnidarians, and arthropods are the most common colonisers (Goldstein *et al.*, 2014).

Another negative effect of plastic litter in the marine ecosystem is that of entanglement. Large floating debris, such as fishing nets, fishing gear, sixpack rings, plastic bands or ropes, can entangle marine creatures (Andrady, 2015; Letcher, 2020)

restricting their mobility, and directly or indirectly convicting them to death by drowning or starvation (Fossi *et al.*, 2018; Ryan, 2018). Finally, lost pots and traps and derelict fishing nets continue to entrap marine life long after their initial loss, causing mortality to a wide range of aquatic creatures (Matsuoka *et al.*, 2005).

Ingestion of plastic litter by marine life is considered among the most serious effects that plastic pollution can induce. Ingestion of plastic litter can sometimes trigger distress to marine creatures as the plastic pieces can block their digestive system. Especially floating plastic articles have proved life threatening for predators of planktonic organisms and filter feeders. Only in the Mediterranean Sea, one out of the four examined blue sharks was found to have consumed anthropogenic litter made-up by 98.2% of plastic debris (Bernardini *et al.*, 2018). Additionally, at some North Sea coasts, debris was found in almost four out of ten stranded whales, which amounted to 25 kg of anthropogenic debris per stranded whale (Unger *et al.*, 2016).

The ingestion of microplastics by marine organisms may also pose serious concerns to humans due to the potential transfer of toxic pollutants from marine life to humans through the food chain. Once persistent organic pollutants (POPs), which may exist in microplastics, are ingested by zooplankton, the toxins of POPs can permeate into the sea creature's tissues. Through the food chain, the toxicity of microplastics eventually may reach the higher trophic levels and end-up in humans (Waring *et al.*, 2018; Rezatofighi *et al.*, 2019; Kögel *et al.*, 2020). Strikingly, a number of research studies indicated that fish gut surfactants have the tendency to enhance the release of POPs, raising in this way the potential of their bioavailability (Wright *et al.*, 2013; Bakir *et al.*, 2014). The matter is of paramount importance considering that 3.2 billion people depend on fish for 20% of their mean per capita intake of animal protein (FAO, 2018).

Microplastics can eventually return to humans, not only from seafood, but also from a variety of other products like table salt, honey, beer, or even mineral water (Auta *et al.*, 2017; Cox *et al.*, 2019). It has been reported that the toxic chemicals which are added to microplastics can trigger a variety of adverse consequences, such as, weakened immune response, curtailed reproductive activity, malformation of

living creatures, including humans, and cancer (Teuten *et al.*, 2009; Setälä *et al.*, 2016). Reasonably, plastic pollution and its perilous effects are of particular concern to humans as they directly impinge on the quality of life and our health.

Concluding, plastic pollution can visually affect highly urbanised areas and pristine islands (Lavers *et al.*, 2017). This kind of pollution can diminish the aesthetic value of the contaminated area to its citizens as well as to its visitors. For instance, in Geoje Island, South Korea, pollution by marine plastics lowered tourists arrivals by 63% inducing economic losses of more than 29 million US dollars (Jang *et al.*, 2014). Additionally, at certain Brazilian shorelines, the high concentration of litter (15 litter items per square metre) discouraged visitors and beachgoers visiting the area, resulting in economic losses of 8.5 million US dollars (Krelling *et al.*, 2017). Lastly, besides the unpleasant aesthetics, marine plastics can sometimes damage marine vessels leading to delays in goods delivering.

Even though the negative effects of macroplastics and microplastics in the marine environment are well documented and widely recognised, their abundance at the seas and the shorelines is not precisely known as it progressively grows over time. Most prevailing monitoring assessments predominately rely on manual methods for retrieving, classifying and quantifying marine debris. Such sampling techniques are time-demanding and entail substantial human labour to deal with the various collection and clean-up steps. For example, the research team of Rosevelt *et al.* (2013), which involved 12 volunteers per beach survey, conducted monthly or semi-monthly surveys for almost a year. Moreover, the team of Smith *et al.* (2014) pursued one-week beach surveys, with each expedition comprising 40 volunteers. Understandably, there is a need for expediting the process of detecting and counting debris at the seas whilst raising the level of successfully discerning each debris type.

Notwithstanding the time frame or the human input, the manually surveyed regions are rather small in area compared to the vast size of the sea and coast zones which harbour marine litter. For instance, the research group of Orthodoxou *et al.* (2022) inspected 100 m of coastline transects at 20 beaches in Cyprus, covering a total of 2,000 m. Such coast stretch is small relative to the entire shorelines of Cyprus,

which span about 735 km (European Commission, 2019). Clearly, it is almost impossible to survey the entire island while deploying humans. More importantly, these activities might not be warranted not least owing to their costs, large resources and extended time. Likewise, the utilisation of marine vessels and other equipment for seagoing litter monitoring expeditions is accompanied by appreciable financial expenses. One clean-up expedition which took place in the Pacific ocean and used the Ocean Cleanup device cost an estimated $6 million (Rainey, 2019).

1.2. Research objectives

Manual litter monitoring surveys, like in-situ sea and beach collections, visual observations, manual image screening and counting are labour intensive processes and exhibit limited detection capabilities. Specifically, such sampling methods can sample small scale littered areas and involve substantial cohort of humans which deal with the various clean-up steps. Recognising that the amount of plastic debris dispersed in the marine environment gradually grows over time, the need of enhancing the sophistication of the marine litter monitoring schemes throughout the world using machine learning (ML) and deep learning (DL) tools constitutes a promising research pathway. Therefore, this study aims to swiftly and intelligently detect and quantify plastic debris in the marine environment. In this context, the research objectives of this study are as follows:

- Evaluate how ML and DL tools can enhance smart classification and detection of plastic debris in the marine environment.
- Test various DL tools tasked with detecting multiple plastic litter in images and videos with high accuracy.
- Study how DL tools can predict the plastic litter density and the physical dimensions of real field marine plastic debris from images.
- Investigate whether DL and counting tools can efficiently tally the number of plastic litter contained in video footage.

1.3. Thesis outline

This work is divided into five chapters. In the first chapter, a concise description of the problem of plastic pollution in the marine environment is offered. In the second chapter, the literature review of macroplastics and microplastics is detailed and the various methods used for their detection and collection from the sea and the shorelines are explained. Chapter three reviews various ML and DL tools which are later utilised for the identification of plastic litter in still images and video content. Additionally, the principles of operation of the hydro-acoustic device used to detect microplastics in the water are lastly explained in this chapter.

The results of the application of the ML and the DL techniques on the marine debris classification task are presented and discussed in the fourth chapter. Thereafter, the development of an intelligent method is presented in this section, which is responsible for classifying, localising and segmenting plastic detritus in images and videos. Furthermore, the results from the utilisation of two counting techniques for tallying the number of marine litter in video recordings are also described in this part. Lastly, beyond the research findings that relate to macroplastics, the results from the detection of microplastics in fresh water using a hydro-acoustic device are also presented here. Concluding, we summarise the research results and close with some future research directions.

CHAPTER 2 LITERATURE REVIEW

2. Literature review

This thesis chapter commences with a brief description of macroplastics and continues with an explanation of the prevailing methods for retrieving and classifying plastic litter found in the marine environment. Additionally, this section explains the reasons for selecting the particular ML and DL techniques employed in this classification task instead of other state-of-the-art solutions. Finally, the chapter concludes with a concise explanation of microplastics as well as the techniques used to identify and count them.

2.1. Macroplastics

Plastics are synthetic or semi-synthetic materials that can be extruded, moulded, cast or spun into solid objects of various shapes (Thompson *et al.*, 2009b). Typically, plastics consist of polymers which are macromolecules made-up of repeating structural subunits (Andrady, 2015). The fact that are inexpensive, lightweight, durable, resilient, tolerant to corrosion, and very good thermal and electrical insulators renders plastics a versatile material. Particularly, the rich variety of plastics and their adaptability enabled the creation of a large number of plastic goods that triggered technological innovations and developments, savings resources and fossil fuels and bringing a plethora of societal benefits (Andrady *et al.*, 2009).

Plastic debris enter the marine environment through various channels such as fishing, shipping and coastal activities, the disposal of domestic sewage or are simply carried by the wind (Barnes *et al.*, 2009; Woodall *et al.*, 2014). As a result, plastics of various types and dimensions are encountered at the coastlines, in the seas and oceans. Large sized plastic debris, known as macroplastics, measure bigger than 5 mm in length (Barnes et al., 2009). According to Andrady (2015), there are 22 types of plastic materials, as listed in Table 1.

Table 1: List of plastic materials.

No	Type of plastic	No	Type of plastic	No	Type of plastic
1	Low-density polyethylene (LDPE)	9	High-impact polystyrene (HIPS)	17	Nylon 66 or polyamide 66 (PA–66)
2	High-density polyethylene (HDPE)	10	Expendable polystyrene (EPS)	18	Cellulose acetate (CA)
3	Polypropylene (PP)	11	Poly(methyl) methacrylate (PMMA)	19	(Ethylene–vinyl acetate) copolymer (EVA)
4	Polystyrene (PS)	12	Poly(ethylene terephthalate) (PET)	20	(Styrene–acrylonitrile) copolymer (SAN)
5	Poly(vinyl) chloride (PVC)	13	Poly(butylene terephthalate) (PBT)	21	(Acrylonitrile–butadiene–styrene) copolymer (ABS)
6	Chlorinated poly(vinyl) chloride (CPVC)	14	Polycarbonate (PC)	22	(Styrene–butadiene–styrene) copolymer (SBS)
7	Polybutene (PB)	15	Polyamide (PA)		
8	General purpose polystyrene (GPPS)	16	Polyamide 6 (PA–6)		

The dominant types of plastic found in municipal solid waste (MSW) comprise LDPE, PP, HDPE, and PET materials (USEPA, 2011). However, the most common types of plastic encountered in the oceans include LDPE, HDPE, PP, PVC, PS, PA, PET and CA (Andrady, 2011). More specifically, LDPE is used for the production of plastic bags, and plastic containers, while HDPE is utilised for plastic milk jugs, toys, etc. Bottle caps and tapes are mainly made of PP, while soda bottles, water bottles, and plastic film are produced from PET. Polyamide (PA) mainly finds application in fishing gear and plastic fishing line. To help readers recognise the most common types of plastic used in everyday life uses, their pictures are displayed in Figure 1. Such objects comprise plastic bags, plastic bottles, plastic caps, cigarette butts, plastic containers, and plastic milk jugs.

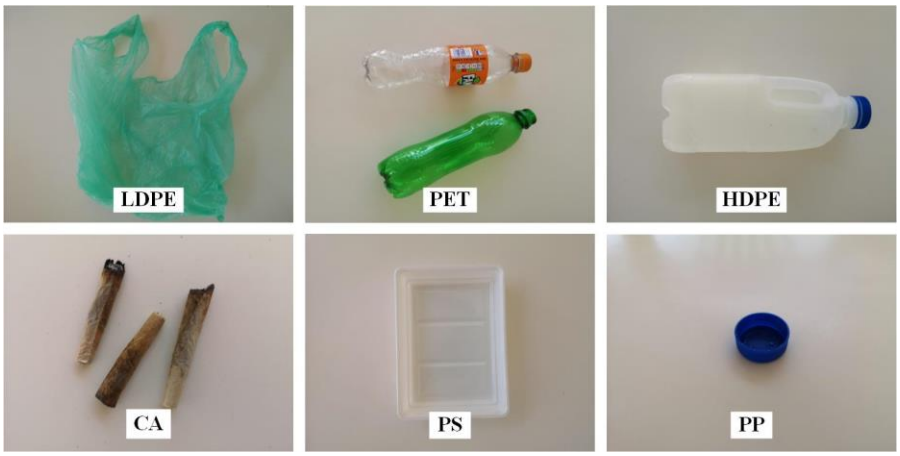


Figure 1: Most common types of plastic used in everyday life synthetic materials.

Researchers mainly use manual means for collecting macroplastics from the seas and the coastlines (Barnes *et al.*, 2009; Newman, 2014; Woodall *et al.*, 2014). Of those, surface net tows constitute the most common method for retrieving plastic litter floating in the sea. Surface net tows are hauled by boats at or close to the sea surface and are suitable for gathering floating plastic litter. These nets have various mesh sizes depending on the dimensions of plastic debris the researchers want to collect. Typically, a mesh size of 0.5 mm is employed. Additionally, surface net tows can have opening mouths of various sizes, which control the dimensions of the collected litter, and usually measure 90 cm in length and 15 cm high (Barnes *et al.*, 2009). Once macroplastics are recovered from the sea, they are sorted out manually and ranked into distinct object categories by human operators.

In-situ visual observations and manual collection and counting are also conducted by human operators for discovering macroplastics floating at sea. In the search of floating macroplastics, researchers onboard boats scan the sea surface using binoculars or their bare eyes (Barnes *et al.*, 2009; Goldstein *et al.*, 2013; Eriksen *et al.*, 2014; Lebreton *et al.*, 2018). As soon as they spot any potential marine debris item, they steer the boat close to the object and manually pick it mainly for forming a better picture as to the scale of the problem of plastic pollution. Likewise, in-situ beach collections and visual observations along transects are primarily used for

estimating the abundance of marine debris at the shorelines (Ryan *et al.*, 2009; Rosevelt *et al.*, 2013).

Over time, new technologies have been applied to the marine plastics classification task. Technologically assisted approaches fundamentally draw from a combination of unmanned aerial vehicle (UAV) surveys and image classifiers grounded on deep learning tools (Fallati *et al.*, 2019; Jakovljevic *et al.*, 2020; Kako *et al.*, 2020). An UAV, such as a drone, is usually equipped with a high-definition camera and collects airborne imagery encompassing plastic litter floating on the sea surface or laying at the shorelines (Hengstmann *et al.*, 2017; Deidun *et al.*, 2018; Moy *et al.*, 2018). The recorded airborne footage obtained during these missions is later collated, using image processing tools, to create a mosaic of images. Thereafter, human operators manually identify possible plastic marine debris objects captured in the combined mosaic images. Even though small drones are popular, their limited flight time and range autonomy severely restricts their reach. Additionally, satellite observations are utilised for the recognition of large sized marine debris items like ghost boats and fishing nets (Acuña-Ruz *et al.*, 2018; Martínez-Vicente *et al.*, 2019; Topouzelis *et al.*, 2019; Kikaki *et al.*, 2020). However, such visual means are not suitable for discerning smaller size plastics measuring less than 1 m in length.

Encouraged by the need to automate the process of smartly detecting and localising plastic debris in the marine environment, while gaining more insights as to the distribution of floating objects, has led to the invention of a new breed of methods for identifying plastic debris. One initiative intended to automate the identification of waste in the marine environment was the Floating Litter Monitoring Application (FLM App) (González-Fernández *et al.*, 2017). The FLM App guides observers to upload pictures of floating debris found in rivers or laying at beaches and assign a label to each image from a drop-down list, which displays various object categories. The intention of the FLM App was to create a large database of labelled images of marine debris, use them to train their deep learning-based image classifier and automate the marine plastics detection process. But no evidence has been published to-date to prove that this method was further improved.

Additionally, the research team of Ge *et al.* (2016) proposed a partially unattended method for recognising marine debris on beaches. Team members used the remote sensing technique of light detection and ranging (LIDAR) and the support vector machine (SVM) (Boser *et al.,* 1992) classifier to categorise marine debris into four general classes: (1) plastics, (2) paper, (3) clothes, and (4) metallic materials. However, this work is a semi-automatic technique and requires a lot of post processing to classify marine debris into their respective classes.

Besides the detection of plastic litter on the sea surface and the shorelines, research teams spot and identify trash under the water with the help of autonomous underwater vehicles (AUVs) and DL techniques. The research team of Fulton *et al.* (2019) employed various DL tools to detect objects belonging to three general classes, such as, plastic materials, man-made objects or devices, and marine life. However, by being able to detect and differentiate among plastics objects from various categories is useful for knowing the types of plastic litter that pollute the seafloor.

In the meantime, other studies have spearheaded the application of DL algorithms for the smart detection and quantification of plastic marine debris in still images and/or video recordings. For instance, the research team of Valdenegro-Toro (2016) utilised Forward-Looking Sonar (FLS) images and their custom convolutional neural network (CNN)-based image classifier to differentiate litter from five general categories, like metal, glass, paper, rubber, and plastic. The pertinent CNN-based classifier obtained 70.8% correct debris identifications (Valdenegro-Toro, 2016) Additionally, the team of Martin *et al.* (2018) utilised the random forest image classifier to categorise three types of plastic litter, namely, plastic containers, caps and bags, from airborne imagery. Clearly the more the categories of plastic litter a classifier can recognise, the more versatile it becomes as a classifier.

The utilisation of DL tools for the detection and counting of plastic debris at the seas and the shorelines constitutes a promising research pathway that permits the large-scale monitoring of beaches and sea strips for litter over time. Image classifiers grounded on DL tools can automate the identification process of plastic debris in the marine environment while being remarkably accurate. Previous attempts on

classifying debris items in the marine environment, as presented earlier in this sub-section, can distinguish litter from general categories like, plastic, metal, glass, paper, fabric and others (Valdenegro-Toro, 2016; Martin *et al.*, 2018). In some cases, researchers have attempted to classify plastic debris among its distinct litter classes but they included only a limited number of items. Unlike other techniques, the image classifiers and marine debris detectors presented in this research study can distinguish plastic debris among several plastic litter classes, namely, plastic bottles, bags, buckets, fishing nets, polystyrene pieces, food wrappings, plastic fishing buoys and fishing nets. Also, particular attention was given to the adaptability and robustness of the proposed marine debris image classifiers and detectors on recognising plastic litter and sea life from various marine settings like the coastlines, the sea surface and the water column.

Suffice to mention that the classes of marine life were added in the image datasets used to train, test and validate the proposed marine debris image classifiers and detectors so as to enhance their relevance to the marine environment. Appreciating that a video footage recorded at the sea surface or underwater will very likely encounter marine creatures, different marine species, like a fish type, dolphins and turtles were included in the training and test sets. Additionally, broadening the number of object categories in the training and the test sets rendered the investigation more comprehensive and enhanced the sophistication and robustness of the proposed techniques. Finally, from an environmental point of view, being aware of the number of detected sea species together with the density of the identified plastic litter items, indicates that this particular marine region is in need of more attention. Therefore, immediate actions will be needed so as to safeguard the particular area.

Various ML and DL tools were utilised to create intelligent and versatile marine debris image classifiers and detectors. Initially, six image feature detectors and descriptors from the Computer Vision domain, that is, the Speeded-Up Robust Features (SURF) (Bay *et al.*, 2008), the Features from the Accelerated Segment Test (FAST) (Rosten *et al.*, 2005), the Binary Robust Invariant Scalable Keypoints (BRISK) (Leutenegger *et al.*, 2011), the HARRIS (Harris *et al.*, 1988), the Minimum

Eigenvalue (MinEigen) (Shi *et al.*, 1994) and the Maximally Stable Extremal Regions (MSER) (Nistér *et al.*, 2008) algorithms, were tested to extract the important attributes from a set of plastic litter images. Soon it was realised that the task of assimilating handcrafted features from marine debris images was unable to cope with the demanding task of categorising marine debris items into separate object classes. For this reason, a more sophisticated ML tool, namely, the Bag of Features (BoF) (MathWorks, 2016) method, was employed for sorting images illustrating marine debris and sea life into eight different classes.

The BoF method has been previously utilised in other classification tasks, like the categorisation of time series (Baydogan *et al.*, 2013), the retrieval of content-based images (Chengcui *et al.*, 2005), the classification of natural image scenes (Maron *et al.*, 1998), the automatic selection of features (Raykar *et al.*, 2008), the robust tracking of objects (Babenko *et al.*, 2011), etc. To the best of our knowledge, this study has pioneered the use of BoF on the marine debris classification task.

Due to the conventional model architecture of the BoF tool, which utilises the SURF algorithm for retrieving the important features from the images of the dataset, the particular scheme exhibits limited classification capabilities. Based on the previous justification, the attention has shifted on more sophisticated DL tools from the Artificial Intelligence (AI) domain. Interestingly, DL tools have the capability of discovering the important features from the images by themselves without requiring the use of any conventional keypoint detector or descriptor. Incentivised by the need of enhancing the classification performance of the proposed marine debris image classifiers, various DL tools were employed here.

The first DL algorithm applied to the marine debris classification task is the Bottleneck method (BM) which employs the Visual Geometry Group-16 (VGG16) as its model architecture (Simonyan *et al.*, 2014). The VGG16 model was selected among other state-of-the-art image classification algorithms, such as the AlexNet (Krizhevsky *et al.*, 2017) and GoogleNet (Szegedy *et al.*, 2015) architectures, as it realises the highest top-5 accuracy of 92.7% when tested on the ImageNet dataset (Simonyan *et al.*, 2014). Analytically, it achieved an increase of 7.6% in the

classification accuracy when tested on the ImageNet dataset compared to AlexNet, and that value dropped to 0.5% when compared to GoogleNet.

Various other research teams have employed the VGG16 model in different domains. For example, the VGG16 model has been applied for the classification of foreign items in airport runways (Xu *et al.*, 2018), the categorisation of malicious software (malware) (Rezende *et al.*, 2018), the classification of kiwi fruits (Liu *et al.*, 2020b) or fish species (Hridayami *et al.*, 2019), and many other fields including the marine debris classification task. For instance, the team of Papakonstantinou *et al.* (2021) employed the VGG16 model for classifying image tiles illustrating marine litter. Also, the team of Musić *et al.* (2020) utilised the VGG16 model to categorise litter into five general litter classes, including a plastic type. Nevertheless, the differentiation among the various types of plastic litter is necessary for this classification task in light of its ability to provide useful information about the types of plastic encountered in the marine environment.

During the course of this research, it was realised that the BM which employs the VGG16 model is a conventional DL tool which applies a single class label per image, irrespective of the number of litter items presented in it. Consequently, the non-detected objects within the images were omitted and this potentially led to an overestimation of the classification accuracy or the undercounting of the detected litter objects. But most importantly, the BM tool was unable to perform multi-object detection in video recordings. Being capable of identifying different litter items from video footage renders the image classifier more useful and valuable simply because video recordings can scan large, polluted areas within a short period of time. Grounded on the reasoning explicated earlier, the DL algorithms which can perform multi-object detection in both still images and videos were applied to the marine debris detection task.

Firstly, the You Only Look Once (YOLO) version 3 (YOLOv3) object detection algorithm (Redmon *et al.*, 2018) was used to discern and localise multiple plastic marine debris and sea life illustrated in images and video recordings. This particular DL tool was selected among other state-of-the-art object detection algorithms, like the

RetinaNet-101 (Lin *et al.*, 2017b), the SSD (Liu *et al.*, 2016), and the Faster R-CNN (Ren *et al.*, 2017). Predominantly, the YOLOv3 attains the desired accuracy and speed trade-off. That is, when YOLOv3 was tested on the MSCOCO dataset (Lin *et al.*, 2014) it achieved the same performance as the RetinaNet-101 model but it was 3.8 times faster. Additionally, YOLOv3 realised almost the same detection performance as the SSD model but it took 1/3 of the time (Redmon *et al.*, 2018). Lastly, even though the YOLOv3 and the Faster R-CNN tools attained comparable average precision, the YOLOv3 outperforms Faster R-CNN in detection speed (Benjdira *et al.*, 2019).

The YOLOv3 model was applied in many detection tasks like the recognition of human faces (Li *et al.*, 2020), the detection of vehicles (Benjdira *et al.*, 2019), the discovery of face masks (Singh *et al.*, 2021), the detection of urban traffic flows (Huang *et al.*, 2020), in forest fire recognitions (Jiao *et al.*, 2019), for identifying fruit (Kuznetsova *et al.*, 2020; Liu *et al.*, 2020a; Xu *et al.*, 2020), and in many other domains including the marine debris detection field. In particular, the YOLOv3 scheme was applied to the detection of floating litter at the sea surface which comprised various general types, such as plastic, wood and other debris (Watanabe *et al.*, 2019). Additionally, the research team of Hipolito *et al.* (2021), applied the YOLOv3 tool for the discovery of underwater animate and inanimate objects. Lastly, the team of Xue *et al.* (2021) employed the preceding model to differentiate marine debris items emanating from seven classes, one being plastics. Differentiated from the abovementioned studies, this research work can recognise debris from up to six types of plastics, namely, bags, bottles buckets, straws, food wrappings and fishing nets, in a variety of marine backgrounds, such as the sea surface or the beach.

After the release of the YOLOv3 algorithm, YOLOv4 was published (Bochkovskiy *et al.*, 2020) which draws heavily on the YOLOv3 tool but is characterised by better mean average precision (mAP) and improved detection speed when tested on the MSCOCO dataset (Lin *et al.*, 2014). Thus, YOLOv4 improves upon the previous version on mAP by 10% and on the number of frames it can process per second (FPS), from video footage, by 12% (Bochkovskiy *et al.*, 2020).

Driven by the need to enhance the detection accuracy of the proposed classification method while retaining its speed, provided the rationale for applying the YOLOv4 tool on the plastic marine debris detection task.

Other uses of the YOLOv4 model encompass the recognition of cars in railways (Mahto *et al.*, 2020), the real-time inspection of railroad track parts (Guo *et al.*, 2021), the identification of traffic signs (Dewi *et al.*, 2021), for autonomous driving (Cai *et al.*, 2021), for fruit detection (Parico *et al.*, 2021; Yijing *et al.*, 2021) and in many other research fields including the identification of marine litter. For example, the team of Tian *et al.* (2022) utilised YOLOv4 tool to detect two classes of plastic litter, namely, underwater plastic bags and broken fishing nets. Moreover, Zailan *et al.* (2021) employed YOLOv4 for the discovery of five general types of floating riverine debris items. Also, the research team of Tomas *et al.* (2022) used the YOLOv4 for differentiating riverine litter made of paper or plastic. Nonetheless, the preceding applications can recognise only a limited number of plastic litter classes while they perform under specific circumstances, like the surface of a river or below the seawater surface.

Shortly after the release of the YOLOv4 tool, another version of the YOLO series, dubbed, YOLOv5, was shared online (Jocher *et al.*, 2020). Although both YOLOv4 and YOLOv5 tools have a similar model structure, they are written in different programming languages. Notably, the YOLOv5 tool is written in Python instead of C used in YOLOv4. Additionally, the YOLOv5 scheme embraces a focus structure in its feature backbone as well as an adaptive anchor strategy. The latter, enables YOLOv5 to select anchor boxes of size and shape that closely resemble the ground truth bounding boxes of objects illustrated in images of the custom dataset. Consequently, these advances lowered the model training duration and improved the detection accuracy. Clearly the need to improve the detection performance of the suggested marine debris detectors provided the impetus for employing the YOLOv5 tool.

The YOLOv5 tool finds applications in several domains like the detection of safety helmets on motorcyclists (Zhou *et al.*, 2021), the real-time detection of fruits for picking robots (Yan *et al.*, 2021), the discovery of ships (Ting *et al.*, 2021; Xu *et*

al., 2022), the recognition of masks on human faces (Ieamsaard *et al.*, 2021), the detection of breast tumours (Mohiyuddin *et al.*, 2022), and others. Interestingly, the preceding scheme was also utilised for the identification of litter in the marine environment. Particularly, the research team of Córdova *et al.* (2022), utilised YOLOv5 for detecting common litter, including plastics, from various natural settings. Also, the team of Tata *et al.* (2021) employed the pertinent scheme to discover submerged trash made of plastic. Additionally, Veerasingam *et al.* (2022) utilised the YOLOv5 tool to recognise general litter belonging to seven classes. However, the preceding applications were capable of categorising plastic debris as a general litter class but were unable to distinguish the different subcategories of plastic litter. Noteworthy, the YOLOv5 marine debris detector of this study improves upon other research efforts in discerning plastic debris from up to nine litter categories. Moreover, the pertinent detector was coupled with two counting tools for tallying marine litter found in both still images and video recordings.

During the implementation of the YOLO series tools on the marine plastics detection task, it was realised that the rectangular bounding box applied by these algorithms on each detected litter item may embrace some unwanted features from the image background and, consequently, lower the detection accuracy of the tools. In contrast, utilising an object detector competent in considering the litter's shape during the detection process can generate new insights as to the size spectrum of the marine litter that spoil the marine landscape. Justified on the earlier reasoning, the You Only Look at Coefficients++ (YOLACT++) instance segmentation algorithm (Bolya *et al.*, 2020) was adopted. Besides the bounding box and the class label assigned to each identified litter item, YOLACT++ applies a mask on each segmented marine debris item which encloses the litter's area.

The YOLACT++ tool was selected among other state-of-the-art two-stage instance segmentation algorithms, like the Mask R-CNN (He *et al.*, 2017) and the FCIS (Li *et al.*, 2017), as it is significantly faster due to its parallel structure and exceptionally "lightweight" assembly procedure. In particular, the YOLACT++ tool achieved a similar mask mAP compared to Mask R-CNN when tested on the

MSCOSO dataset (Lin *et al.*, 2014), but it was almost 4 times faster. Moreover, the YOLACT++ model fared better by 4.6% in mask mAP while it was 5 times faster than the FCIS tool (Li *et al.*, 2017) when tested on the MSCOSO dataset (Lin *et al.*, 2014).

The detection of medical instruments in laparoscopic images (Ángeles Cerón *et al.*, 2021), the discovery of surface defects on magnetic-tiles (Wen *et al.*, 2021), autonomous driving (Wang *et al.*, 2022), the identification of civil bridge cracks (Fu *et al.*, 2022), and the diagnosis of psoriasis in skin images (Lin *et al.*, 2021) are some of the applications of the YOLACT++ algorithm in other domains. Referring to the marine debris detection task, the team of Dias *et al.* (2022) tapped the YOLACT++ scheme to categorise four types of underwater litter which comprised plastic bottles, bags, glass bottles, and metallic cans. Still more classes of plastic litter need to be considered during the marine debris detection task. To this end, the YOLACT++ marine debris detector presented in this research improves upon previous attempts in its ability to discern a variety of plastic debris, like plastic bags, bottles, buckets, food wrappings, straws, and fishing nets encountered in both images and video footage. Moreover, the specific tool can recognise the preceding litter categories in a variety of marine backgrounds like the sea surface or the beach sand.

Lastly, when estimating the abundance of plastic marine litter items illustrated in video footage, two counting methods, namely, the region of interest (ROI) line counting (Mazurek, 2020) and the centroid tracking (Rosebrock, 2021b) approach, were tested in this study. The preceding counting algorithms were coupled with the YOLOv5 tool and used to discover and count litter items in real field video footage containing marine debris. Operationally, the ROI line counting method imposes a virtual line on a video and tallies the number of objects that cross it. The particular scheme has been applied in other fields, like the counting of vehicles in highways (Bas *et al.*, 2007; Memon *et al.*, 2018) or tallying the number of people walking in streets (Chen *et al.*, 2008; Yang *et al.*, 2019; Rosebrock, 2021a).

On the other hand, the centroid tracking counting technique tracks detected objects as they propagate from frame to frame and assigns them unique identities

(IDs) (Khachatryan, 2019; Rosebrock, 2021a; Rosebrock, 2021b). Noteworthy, applications of the centroid tracking method comprise the ranking of fuzzy numbers (Wang *et al.*, 2006), the human face detection and tracking (Zhang *et al.*, 2009) and the identification of shifts in the metapopulation range of various species tied to climate change (Watts *et al.*, 2013). Concluding, it is important to note that both counting methods have not been previously utilised in the detection and counting of plastic marine debris captured in video recordings.

2.2. Microplastics

Microplastics are defined, in the literature, as small plastic pieces or particles with a length less than 5 mm (Betts, 2008; Moore, 2008; Barnes *et al.*, 2009; Fendall *et al.*, 2009). Even smaller in size, nanoplastics measure less than 1 μm in length (Andrady, 2015). Generally, microplastics can either be primary or secondary origin. Falling into the primary group, are plastic particles that have been intentionally manufactured to be microscopic, hence, belonging to the microscale category. For instance, virgin pellets that are used as a raw material for the creation of larger individual plastic products, or microbeads added to cosmetic products. Additionally, microfibres produced from laundering synthetic fibre clothing also fall into the microplastics and nanoplastics groups.

Besides primary microplastics, there are also secondary microplastics. Secondary microplastics are plastic particles formed by the fragmentation of larger plastics already found in the marine environment or macroplastics which breakdown during use (Andrady, 2015; Letcher, 2020). More precisely, macroplastics decompose into smaller fragments by the mechanisms of biodegradation, photodegradation, thermo-oxidative degradation and hydrolysis (Andrady, 2011). Biodegradation is facilitated by the action of living creatures, such as microbes, on macroplastics while photodegradation is caused by the exposure of macroplastics to sunlight. Thermo-oxidative degradation pertains to the slow oxidative decomposition of plastics at moderate temperatures, and lastly, hydrolysis refers to the chemical reaction of macroplastics with the sea water (Andrady, 2011).

Table 2: The prevalent types of microplastics found in the marine environment. Courtesy: Andrady (2011).

Plastic class	Plastic origin
Low-density polyethylene (LDPE)	Plastic bags, six-pack rings, bottles, straws
High-density polyethylene (HDPE)	Milk and juice jugs
Polypropylene (PP)	Ropes, bottle caps
Polystyrene (PS)	Plastic utensils, food containers
Foamed Polystyrene	Floats, bait boxes, foam cups
Nylon (PA)	Nets and traps
Thermoplastic Polyester (PET)	Plastic beverage bottles
Poly-vinyl chloride (PVC)	Plastic film, bottles, cups
Cellulose Acetate (CA)	Cigarette filters

Microplastics are introduced in the marine ecosystem mainly through domestic sewage water or by the fragmentation of existing macroplastics (Barnes *et al.*, 2009; Woodall *et al.*, 2014). Predominantly, the types of microplastics found in aquatic environments comprise low-density polyethylene (LDPE), high-density polyethylene (HDPE), polypropylene (PP), poly-vinyl chloride (PVC), polystyrene (PS), Nylon (PA), thermoplastic polyester (PET), and cellulose acetate (CA) (Andrady, 2011; USEPA, 2011), as illustrated in Table 2. As noted from Table 2, most of the microplastics mainly originate from larger size plastics, such as, bags, six-pack rings, bottles, straws, plastic beverage bottles and cups, plastic food containers, fishing nets and others (Andrady, 2011).

Surface net tows constitute the main collection method for microplastics in seawater. Net tows have a mesh size of at least 300 μm and are towed by boats on or near the sea surface for retrieving microplastics. Once the samples are recovered, a delicate procedure is followed to differentiate and characterise microplastics. Initially, researchers clean the collected sample with distilled water to remove unwanted marine organisms or detritus. Subsequently, with the aid of a microscope, the researchers identify and count the microplastics. Finally, using spectroscopic techniques, like Raman spectroscopy and Fourier-transform infrared spectroscopy

(FT-IR), scientists can recognise the microplastic material (Andrady, 2011; Cózar *et al.*, 2014; Tagg *et al.*, 2015).

Besides the use of surface net tows in seawater, researchers are sampling beach sediments in search of microplastics. With the aid of a 5 mm sieves, investigators sample the top layer of beach sediments— measuring approximately 3–6 cm thick— in their quest for microplastics (Wessel *et al.*, 2016). Additionally, microplastics can be traced in seabed sediments (Thompson *et al.*, 2009a). Such samples are cleaned with distilled water and, with the aid of a microscope, microplastics are distinguished from other particles. Finally, researchers can identify the origin of microplastics using the previously mentioned spectroscopic techniques. In the meantime, other methods, like biopsies, can extract microplastics from the stomach or the skin of marine creatures (Fossi *et al.*, 2017; Graca *et al.*, 2017; Barboza *et al.*, 2018; Gallo *et al.*, 2018).

Current techniques for detecting microplastics in the marine environment are predominantly manual methods. However, streamlining the detection process of microplastics while deploying fewer human operators or less expensive equipment is a crucial step towards the mitigation of marine plastic pollution. For this reason, sensors that are inexpensive and easy in use and can detect and trace microplastics in freshwater were investigated in this study. Initially, the capabilities of devices like capacity proximity sensors, photoelectric sensors, and electrical resistivity tomography sensors were studied. However, such devices cannot operate underwater and are unable to detect small plastic particles in the water. For this reason, the focus has shifted to a hydro-acoustic device, namely, the ISA500 sensor (Impact Subsea, 2016), which can function in the water at frequencies high enough to discover microplastics in the water. Concluding, expensive equipment like spectrometers were omitted owing to the limited financial resources available to us.

CHAPTER 3 METHODOLOGY AND METHODS

3. Methodology and methods

In this chapter, the methods used for the detection of macroplastics and microplastics in the marine environment are presented and discussed. More specifically, this chapter is divided into two parts: 1) the classification and detection of macroplastics, 2) the detection of microplastics. In the first part, the methodology followed for the classification, the detection, the localisation, and the counting of plastic litter in the marine environment is explained. Subsequently, all of the ML and DL tools used in this study are presented. In the second part, the hydro-acoustic device used for the detection of microplastics is lastly described.

3.1. Classification and detection of macroplastics

Initially, to detect the important features from images illustrating plastic marine debris, such as, a bottle, six feature detectors and descriptors were utilised, namely, the SURF, the FAST, the BRISK, the HARRIS, the MinEigen, and the MSER. Each of the preceding algorithms assessed the image features aiming to discover the query object, that is, a plastic bottle in this case, found in an image scene. Furthermore, this process was used to detect the points of interest in an image scene that contained the plastic bottle. Finally, the algorithm matched the points of interest from both the query object and the image scene and identified the object of interest in the wider image scene by encompassing it in a rectangular bounding box.

However, during the application of the SURF, the FAST, the BRISK, the HARRIS, the MinEigen and the MSER feature detectors and descriptors it was realised that the task of assimilating handcrafted features from marine debris images exhibits limited classification capabilities. Consequently, the techniques were not very competent in categorising marine debris items into distinct object classes. Based on this reasoning, a more sophisticated ML tool, namely, the Bag of Features (BoF) method, was then utilised to address the marine debris classification task.

The BoF method was used to construct an image classifier responsible for categorising images portraying plastic litter and sea life in the marine environment.

Firstly, an image dataset was created for training, testing and validating the BoF image classifier. This dataset consisted of eight categories of objects, that comprised plastic bottles, plastic bags, plastic buckets, fishing nets, polystyrene pieces, plastic fishing buoys, dolphins, and sea turtles. For processing purposes, the BoF method utilises the SURF tool as its feature extractor designed to obtain the salient image features, such as, blobs, from the marine debris images. Yet, there are more useful details to be retrieved and assimilated by the classifier from the images of the dataset. Drawing on the earlier explanation and given that the DL tools, from the AI domain, possess the capability of discovering the important image features themselves, like the colour, the shape of an object in images etc., provided the inspiration to apply them on the marine debris classification task.

Starting, the first DL tool utilised for the categorisation of images encompassing marine debris and sea life was the Bottleneck method (BM). The BM image classifier was capable of discerning up to eight categories of litter items, that is, plastic bottles, plastic buckets, plastic bags, fishing nets, plastic straws, food wrappings, flying fish, and "other" which comprises unrelated entities like rocks, humans, etc. Nonetheless, the particular classifier is a conventional image classification tool which applies a single class label per image. Another shortcoming that relates to the pertinent DL tool is its inability to categorise objects contained in video recordings. Due to the inherent merits of video footage, which permit scanning larger areas for marine debris in less time, more sophisticated DL tools like the YOLO series and the YOLACT++ algorithm were adopted hereafter.

When detecting and localising multiple plastic litter and sea life appearing in images and video footage, the YOLOv3 tool was initially employed. This particular DL tool can attach a class label on each detected litter item found in an image or video footage while localising it by applying a rectangular bounding box around it. In this regard, two more YOLO tools, namely, the YOLOv4 and YOLOv5, were also applied here for investigating whether the specific DL algorithms can improve the mean average precision (mAP) of the plastic marine debris detectors. All of the YOLO series were trained, tested and validated on the same image datasets which

consist of seven litter categories: 1) the plastic bottles, 2) plastic bags, 3) plastic buckets, 4) fishing nets, 5) plastic straws, 6) food wrappings, and 7) a fish species. Summarising, among the YOLO series, the YOLOv5 object detector proved the most sophisticated as it realised the highest mAP on still images and video recordings. Equally important, the specific tool achieved real-time object detections of 34 FPS on video footage.

Thereafter, to count plastic marine litter from video footage and estimate their abundance, two counting techniques, namely, the region of interest (ROI) line and the centroid tracking method, were considered. The two counting techniques were coupled with the YOLOv5 tool which realised real-time object detections on a video footage illustrating marine debris from the abovementioned seven litter classes. Operationally, the ROI line approach performs counting during the temporal instance at which the item crosses the line, while the centroid tracking approach tracks the geometric centre of the bounding box of a detected litter object and counts it during the entire time duration that the object appeared on screen. Finally, due to fact that the centroid tracking technique proved more adept than the ROI line counting method, it was selected and combined with the YOLOv5 object detector to tally plastic marine debris and sea life from images and video recordings.

Finally, to predict the plastic litter density across the Cypriot coastlines and estimate the litter's physical dimensions, the YOLACT++ instance segmentation tool, was utilised. Besides the bounding box and the class label, the YOLACT++ tool applies a mask on each segmented marine debris item or marine species embedded in images and videos. To enable the comparison with the YOLO series, the YOLACT++ tool was trained, tested and validated on the same object categories. In order to predict the plastic litter density across the Cypriot coastlines, the YOLACT++ algorithm processed real field marine litter images captured from various beaches in distinct coastlines in Cyprus. Counting all of the discovered objects and dividing them by the total length of the surveyed area, the plastic litter density was calculated. Lastly, extrapolating the computed density to the entire shorelines of the island, the method estimated the number of plastic articles littering the Cypriot coastlines. Concluding,

using the *OpenCV Contours* image processing tool, the applied mask on each identified object helped estimate the item's physical dimensions. Collectively, the methodology adopted in this study for the classification and detection of macroplastics in the marine environment is illustrated in Figure 2. Finally, subsequent thesis sub-sections present all of the ML and DL tools examined in this study.

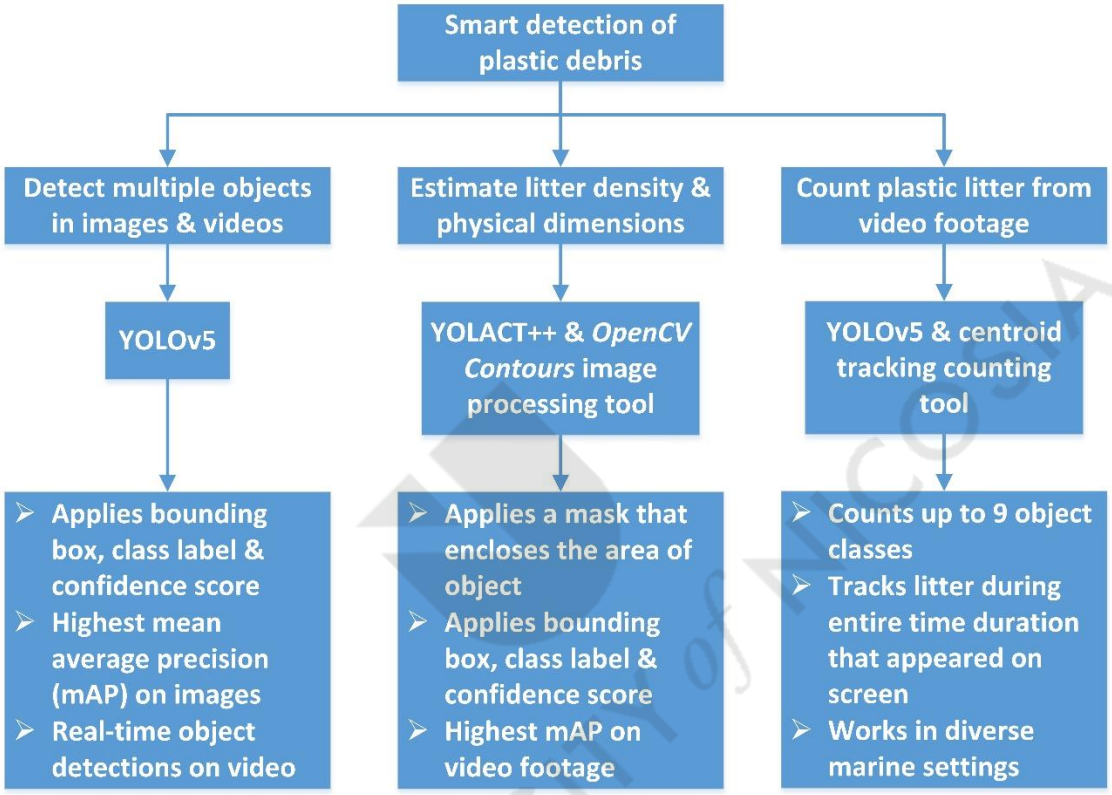


Figure 2: Flow chart describing the classification and detection process of macroplastics in the marine environment.

3.1.1. The SURF detector and descriptor

Speeded-Up Robust Features (SURF) is a feature detector and descriptor algorithm which is invariant to scale and rotational changes (Bay *et al.*, 2008). SURF was formulated on a basic Hessian matrix approximation H_{approx} intended to detect points of interest in an image. The algorithm uses box type convolutional filters and relies on the concept of integral images for fast computation. Integral images refer to the sum of all pixels in the input image I within a rectangular region formed by the

origin and location ***x*** in the image (Viola *et al.*, 2001). In this way, SURF detects blob-like features in an image, where the determinant of the Hessian matrix is maximum. For a given 9×9 box filter, the determinant of the Hessian matrix det(H_{approx}) (Bay *et al.*, 2008) can be computed from:

$$\det\left(H_{approx}\right) = D_{xx}D_{yy} - \left(wD_{xy}\right)^2 \quad (1)$$

where D_{xx}, D_{yy}, D_{xy} are approximations of a Gaussian distribution with scale σ=1.2. Term w is a weight applied to the rectangular regions of an image during the computation of the box type convolutional filters and is used to balance the expression of the Hessian determinant.

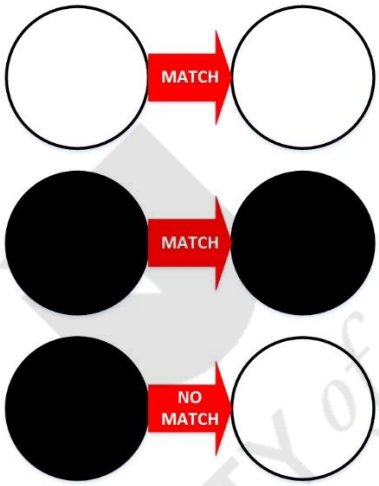


Figure 3: The SURF algorithm can match two blobs with the same type of colour contrast.

The approximated determinant of the Hessian matrix corresponds to blob response in the image *I*, at location ***x***. These responses are logged in a blob response map for diverse scales intended to identify their local maxima. Herein, the change in the scale-space is achieved by up-scaling the filter size rather than downsizing the image. Subsequently, the rest filters measure 15×15, 21×21 and 27×27. Concluding, a non-maximum suppression (NMS) in a 3×3×3 neighbourhood is implemented to localise points of interest in the image over the previous scales.

SURF also describes the points of interest in an image by defining the distribution of the intensity content within their vicinity (Lowe, 2004). The distribution of the

intensity content is computed using the first order Haar wavelet responses in the *x* and *y* directions. In this way, a reproducible orientation grounded on the information from a spherical area around the point of interest is fixed. Next, the SURF descriptor is applied to a square region aligned to the selected orientation. Finally, features among two similar images are matched using the sign of the Laplacian, which is the trace of the Hessian matrix. Drawing on this comparison, SURF identifies light blobs in dark backgrounds and vice-versa. As depicted in Figure 3, only features with the same type of colour contrast are finally matched.

3.1.2. The FAST detector

The Features from Accelerated Segment Test, known as FAST, is a feature detector capable of identifying corner points between image frames (Rosten *et al.*, 2005). FAST tests a feature at a pixel *C* in an image by analysing a 16-pixel circle around that pixel. Specific features are detected by FAST when the intensities of at least 12 adjacent pixels belonging in this circle are all above or all below the intensity of pixel *C* by a certain threshold *t*. Finally, the pixel intensities from the 16-pixel circle can be used as a feature vector which describes the examined pixel *C*.

During the matching process, FAST utilises the sum squared difference (SSD) approach to determine the correlation among two feature vectors. Two feature vectors are most likely to be correctly matched if their SSD has been minimised. For instance, in an image frame *n* there are two sets of features, namely, the $F_{O,n}$ extracted from the previous frame and projected under motion μ, and the $F_{I,n}$ attribute that has been obtained from the original image. For a particular feature f_i belonging to $F_{O,n}$, its correlation is computed by deducing the feature point in $F_{I,n}$ which minimises the SSD of the feature vector among the examined feature points.

3.1.3. The BRISK detector and descriptor

The Binary Robust Invariant Scalable Keypoints (BRISK) is a feature detector and descriptor which employs a saliency criterion similar to the FAST detector to identify

corners in an image (Leutenegger *et al.*, 2011). BRISK applies a 9-16 mask for feature point detection which seeks at least 9 consecutive pixels in the 16-pixel circle to be adequately brighter or darker than the central pixel *C* for the saliency criterion to be satisfied.

During the detection process of BRISK, a 9-16 mask is applied individually on each octave and intra-octave of the image scale-space pyramid using the same threshold *t* to discover possible regions of interest. Each octave is shaped by gradually half-sampling the original image while intra-octaves are located in-between octave layers. The first intra-octave is shaped by down-sampling the original image by a factor of 1.5, while the rest of the intra-octave levels are obtained by consecutive half-sampling. Feature points of these masks are subjected to NMS manipulation in the scale-space continuum. Thereafter, the examined feature point must satisfy the maximum condition regarding its 8 adjacent pixels' FAST scores *s* in the same layer. FAST scores *s* are the saliency metrics that govern the maximum threshold which is tasked with identifying a feature point as a corner. Additionally, the layers above and below should attain lower scores. Finally, a 1D parabola and partial interpolation at the boundaries of the patch are applied along the scale-axis to define the real size scale of the feature point.

During the characterisation of a feature point, BRISK constitutes a binary string that concatenates the outcomes of the brightness comparison assessments. BRISK discovers the representative direction of each feature point to allow for orientation-normalised descriptors, ensuring in this way the robustness of the method and its invariance in rotation. Subsequently, the Hamming distance is computed to match two BRISK descriptors. The dissimilarity between the two descriptors rests on the difference in the number of bits different in the two descriptors. Concluding, the team of Leutenegger *et al*. (2011) suggest that a matching Hamming distance threshold of 90 bits is the ideal.

3.1.4. The HARRIS detector

Engineers Chris Harris and Mike Stephens developed the HARRIS edge and

corner detector to address problems related to the 3D representation of a curving edge (Harris *et al.*, 1988). Specifically, the HARRIS algorithms attempted to tackle these problems by detecting both edges and corners in an image frame. The whole approach is based on Moravec's corner detector which considers a local window in the image and computes the average variations of image intensity which arise from shifting the image window in different directions (Moravec, 1980). Generated by a window displacement (x, y), the spatial change, $E_{x,y}$, in the image is obtained from the following equation (Harris *et al.*, 1988):

$$E_{x,y} = \sum_{u,v} w_{u,v} \left| I_{x+u,y+v} - I_{u,v} \right|^2 \tag{2}$$

where $I_{x+u,y+v}$ and $I_{u,v}$ refer to the image intensities and term $w_{u,v}$ is the spatial window parameter which assumes unity within a defined rectangular area and zero outside of the domain. Finally, the HARRIS detector seeks local maxima in $min\{E\}$ above a certain threshold *t*.

Moravec's approach encounters a number of problems associated with the effective detection of corners and edges in an image. Firstly, it has an anisotropic response because it considers only the intensity variations at a distinct set of shifts at every 45°. Secondly, it has a noisy behaviour due to the binary and rectangular image window and thirdly, it responds too readily to edges because it considers only the $min\{E\}$. Therefore, the HARRIS detector tries to tackle the abovementioned problems by, firstly, applying an analytical expansion about the origin of the window displacement aimed at encompassing all small shifts endeavouring to recalculate the spatial change *E*. Subsequently, HARRIS utilises a smooth circular Gaussian window to alleviate the noise and, lastly, updates the corner measure to process the change of *E* emanating from the direction of shift. Variations in the spatial change *E* are then calculated from:

$$E(x, y) = (x, y) M (x, y)^T \tag{3}$$

where *M* is a 2×2 symmetric matrix. If both eigenvalues of matrix *M* are small and tend to zero, the displacements of the image window cause an insignificant change in

intensity *E*, which implies that the region is flat. If both eigenvalues are large and bigger than zero, displacements in any direction will induce an increase in *E* and this indicates the presence of a corner. But, if one of the two eigenvalues is large and the other is small, only displacements along the ridge will effect a change in *E*, which points to an edge.

Beyond the detection of edges and corners, the HARRIS detector also considers the quality of the detected image features. To attain this task, it measures the edge and corner response, *R*, as follows (Harris *et al.*, 1988):

$$R = \alpha\beta - k(\alpha + \beta)^2 \tag{4}$$

where k is an empirical constant and α and β are the eigenvalues of the symmetric matrix *M*. Clearly, response *R* is positive for corners, negative for edges and assumes small values for flat regions.

3.1.5. The MinEigen method

Engineers Shi and Tomasi (1994) proposed the minimum eigenvalue method (MinEigen) which traces good features, like corners, in an image. A tracking algorithm that employs the affine motion field for representing the displacement of a feature window during image motion was suggested here. A pure translation model proved to be successful for the recommended tracking method as it attained superior precision. More precisely, to control the quality of the feature tracking, a dissimilarity measure, comparable to the HARRIS detector, was utilised to compute changes in appearance of a feature between the current and the first image frame. However, the main difference between the MinEigen and the HARRIS algorithms relates to the calculation of the corner response, *R*. Term *R* here depends only on the minimum value between the eigenvalues α and β of the symmetric matrix *M*, and not on the magnitude of the k constant, as follows (Shi *et al.*, 1994):

$$R = \min(\alpha\beta) \tag{5}$$

where a score larger than the minimum value denotes a corner point.

3.1.6. The MSER algorithm

Maximally Stable Extremal Regions (MSER) is a region detector that is using an immersion analogy in a true linear time in the number of pixels to detect the regions of interest (Nistér *et al.*, 2008). The MSER detector is a flooding simulation method for computing a watershed segmentation which employs a single connected component of pixels lending the algorithm computational efficiency and speed. Unlike other immersion analogy approaches, where the image is flooded with the same water level universally, the flooding here manifests as if the scenery was opaque and water is being flown at some randomly selected image pixel.

Specifically, the region within the image in which the arbitrarily selected point belongs is first filled with water. Followed then, the rest of the regions are progressively flooded once they become accessible to the water. The water adjusts to the real scenery and behaves as a unified component remaining in contact to the point where it has originally started to flow onto the image scenery. Ultimately, the MSER method tracks the "downhill stream" of water, which provides information pertaining to the pixels' components for all same grey-levels in the image. Once the downhill stream realises a minimum, the water fills that region and the respective pixels are processed and detected by the MSER algorithm.

3.1.7. The Bag of Features method

The Bag of Features (BoF) (MathWorks, 2016) method was used to construct an image classifier responsible for categorising images portraying plastic litter and sea life in the marine environment. Functionally, since images do not encompass discrete words, the BoF method constructs a visual "vocabulary" of features representative of each image category. This is accomplished by extracting the SURF features (refer to §3.1.1) from all images of an image dataset. Subsequently, the scheme constructs a visual vocabulary by using the *k*-means clustering method. Specifically, the *k*-means clustering is a partitioning method which is applied on the feature descriptors extracted from the images of a training set. During the partitioning process, the method iteratively groups the obtained descriptors into *k* mutually exclusive clusters, while the resulting clusters are consolidated and separated by comparable image

attributes. Each cluster centre signifies a feature, or a "visual word" and all of the visual words generate the visual vocabulary which is responsible for training the BoF image classifier.

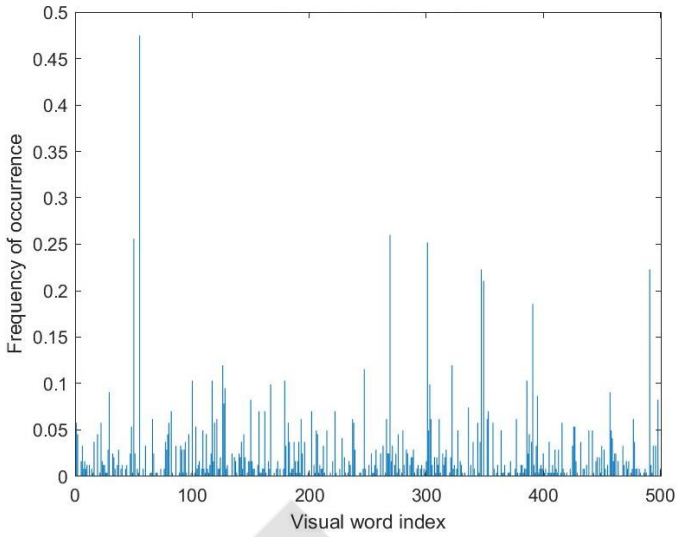


Figure 4: A histogram of visual word occurrences that encodes a training marine debris image.

The BoF technique also employs an encoding method for counting the visual word occurrences in an image. Thus, a training marine litter image is encoded into a histogram of visual word occurrences, as illustrated in Figure 4. In turn, using the approximate nearest neighbour algorithm the histogram is transformed into a feature vector. The approximate nearest neighbour algorithm extracts histogram bins based on the proximity of the descriptor to a particular cluster centre. Quantitatively, the length of each histogram bin corresponds to the number of visual words assigned to each cluster. Essentially, the encoding method converts an image into a feature vector while creating at the same time a basis for training the classifier. Suffice to mention that a multiclass linear SVM classifier (Boser *et al.*, 1992) is used for creating the BoF image classifier.

3.1.8. The Bottleneck method (BM)

The Convolutional Neural Network (CNN) that utilises the Bottleneck method (BM) belongs to the family of DL tools. Its CNN architecture springs from the Visual Geometry Group-16 (VGG16) model, pre-trained on the large-scale ImageNet dataset— a database of over 14 million images divided into almost 1,000 individual classes (Simonyan *et al.*, 2014). Due to this pre-training, the BM has assimilated important image features needed for all classification tasks and, as a result, the overall classification accuracy of the method improves.

Figure 5 displays the VGG16 model architecture of the BM. Concisely, the presented block diagram depicts the sequence of steps for incorporating the marine debris images into the VGG16 model and, thereafter, are processed to extract the bottleneck features. Initially, images are fed into the Convolutional Block 1, which consists of convolutional layers and max-pooling layers. A convolutional layer is responsible for organising the units in an image into feature maps, while a max-pooling layer merges the semantic features of these units into a unified feature map. Next, an image propagates through all of the convolutional blocks until Block 5— just before the fully connected layer. Henceforth, the max pooling layer just before the fully connected layer extracts the bottleneck features from the marine debris image set. The fully connected layer is arranged into one flatten layer and three dense layers. Specifically, the flatten layer reduces the dimensions of the feature map into a single column that is passed to the fully connected layer, whereas the dense layer ties the fully connected layer to the neural network.

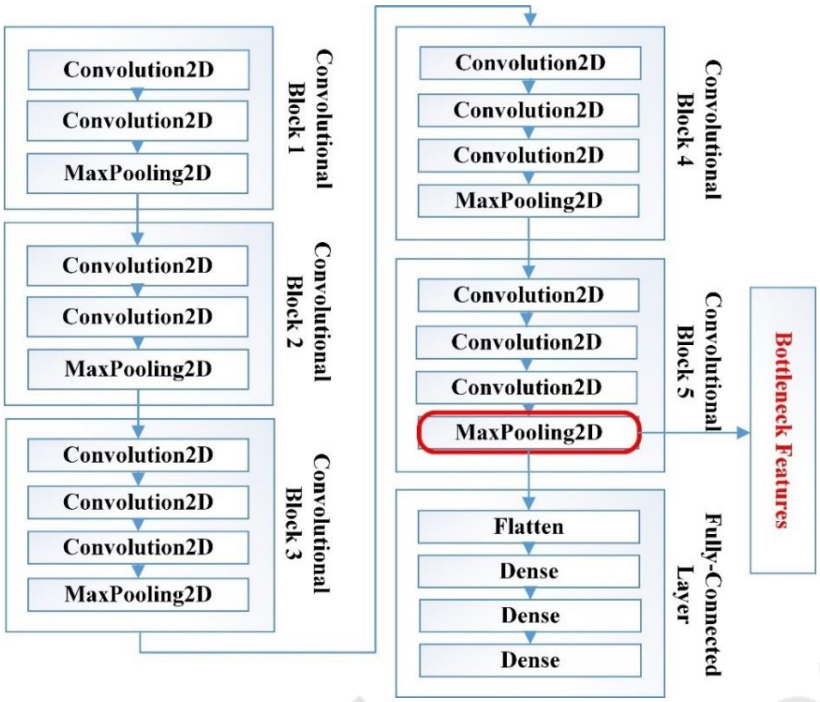


Figure 5: Schematic representation of the VGG16 model.

The optimisation procedure of the BM image classifier is formalised by means of an iterative descent of gradients in the loss function quantifying, thus the error in predictions (weights). As an approximation to the true gradients, the mini-batch stochastic gradient descent with the Adadelta learning rate was utilised. The Adadelta learning rate method for gradient descent was selected after a thorough investigation intended to identify a suitable optimiser capable of executing the marine debris classification task. The loss function for a batch of *N* samples was derived from the categorical cross-entropy loss function:

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} q\left(x_{i,c}\right)\log\, p\left(x_{i,c}\right) \tag{6}$$

where *N* is the total number of samples and *C* is the overall count of classes. Term *q* is an indicator factor, which assumes a value of 1 only if sample $x_{i,c}$ belongs to its category *c*, else it is assigned a 0. Parameter *p* is the estimated probability produced by the model for sample x_i belonging to category *c*. Probability *p* is obtained from

the “Softmax” function:

$$p(\boldsymbol{x})_c = \frac{\exp(x_c)}{\sum_{j=1}^{C} \exp(x_j)} \tag{7}$$

where probability *p* is a normalised exponential that accepts as input a *C*-dimensional vector ***x*** and generates as output a *C*-dimensional vector ***p*** of real values ranging between 0 and 1. Term x_c refers to the elements of vector ***x***.

Additionally, the Rectified Linear Unit (RELU) activation function was used in the BM optimisation procedure and is computed using the expression $f(x) = max(0, x)$. ReLU is linear for input values greater than zero. In this case, the output value matches the input value. However, for negative input values ReLU behaves as a non-linear function. Under these circumstances, the output value always yields a zero value (Goodfellow *et al.*, 2016).

Concluding, in order to improve the overall classification performance of the proposed BM, some additions were made to the BM model as cited in sub-section A1 of Appendix. These relate to the implementation of the code elements that augment the image sample population and split them into training and testing datasets. Additionally, the type of model loss, the type of the optimiser and the regularizers were adjusted for the scope of this research work. Collectively, the preceding modifications were designed to improve the overall classification accuracy of the BM while eliminating overfitting phenomena. Overfitting is the flaw where the model embraces the details and noise in the training set, such that it adversely affects the performance of the model to generalise. Finally, an extra code section was added to the basic BM model to predict the class of each examined image and assigned to it its respective class label.

3.1.9. The YOLOv3 algorithm

The YOLOv3 object detection algorithm (Redmon *et al.*, 2018) was utilised for detecting and localising marine debris and sea life in images and video recordings. The preceding tool reframes the object detection task into a single regression problem.

This means that a single neural network assigns bounding boxes and class probabilities to the detected objects straight from the entire image in one assessment. Lastly, besides still images, the YOLOv3 tool can also recognise multiple debris items captured in video content.

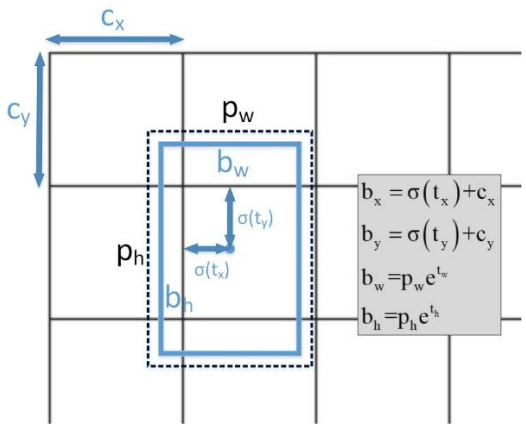


Figure 6: Predicted width, height and centre coordinates of the detected bounding box as applied by the YOLOv3 algorithm.

The YOLOv3 algorithm divides the input image into an S×S grid with each grid cell being responsible for the detection of an object if the centre of that body falls into the respective grid cell. Predefined anchor boxes are generated in every grid of the input image and produce bounding boxes. Each anchor box makes five predictions for each bounding box: four coordinates t_x, t_y, t_w, t_h plus a confidence score. Variables t_x and t_y represent the coordinates of the centre of the predicted bounding box relative to the boundaries of the grid. Parameters t_w and t_h refer to the width and height of the predicted bounding box relative to the entire image. Followed then, if the grid cell is offset from the top left corner of the image by a distance (c_x, c_y), as depicted in Figure 6, and the predefined box has width and height p_w and $p_{h,}$ respectively, then the coordinates of the predicted bounding box correspond to:

$$b_x = \sigma(t_x) + c_x \tag{8}$$

$$b_y = \sigma\left(t_y\right) + c_y \tag{9}$$

$$b_w = p_w e^{t_w} \tag{10}$$

$$b_h = p_h e^{t_h} \tag{11}$$

Here b_w and b_h are the width and height of the bounding box as offsets from cluster centroids, respectively. Denoted by b_x and b_y are the centre coordinates of the bounding box relative to the position of the applied filter formulated on a sigmoid function.

The confidence score expresses the level of confidence the model exhibits in that the bounding box contains an object and how accurate it thinks the predicted bounding box is. The confidence score is defined as follows:

$$\Pr\left(\text{Object}\right) \times \text{IoU} \tag{12}$$

where $\Pr\left(\text{Object}\right)$ is the probability that the bounding box contains an object. Letters IoU stand for the intersection over union between the predicted bounding box B_p and the ground truth bounding box B_g such that:

$$\text{IoU} = \frac{\text{area}\left(\text{B}_\text{p} \cap \text{B}_\text{g}\right)}{\text{area}\left(\text{B}_\text{p} \cup \text{B}_\text{g}\right)} \tag{13}$$

The ground truth bounding box is a rectangular box assigned to each object embedded in an image. The open YOLO_mark annotation tool was used to manually sketch the ground truth bounding boxes on each debris item found in individual training and testing images. Similarly, the predicted bounding box, is a rectangular box assigned to each object detected by the YOLOv3 algorithm.

Therefore, for each image grid cell containing an object, a set of *C* conditional class probabilities, $\Pr\left(\text{Class}_\text{i}|\text{Object}\right)$, is computed. In order to calculate the class-specific confidence scores for each bounding box, the confidence scores and the conditional class probabilities for this bounding box are multiplied as follows:

$$\Pr(\text{Class}_i|\text{Object}) \times \Pr(\text{Object}) \times \text{IoU} = \Pr(\text{Class}_i) \times \text{IoU} \quad (14)$$

Lastly, the bounding box with the highest class-specific confidence scores is the final predicted bounding box.

YOLOv3 employs the Darknet-53 network as its feature backbone (Redmon *et al.*, 2018). The Darknet-53 model is a hybrid approach between the network used in YOLOv2, Darknet-19 (Redmon *et al.*, 2017), and residual layers. A residual layer is responsible for bypassing some convolutional layers. Hence, the Darknet-53 network consists, as the name implies, of 53 convolutional layers with successive 3×3 and 1×1 convolutional layers and shortcut connections. Finally, the model uses global average pooling to make predictions as well as a 1×1 filter to compress the feature representation between 3×3 convolutions. Finally, certain parameters of the YOLOv3 model were altered so as to improve its overall detection accuracy. Such modifications relate to changes in the dimensions of the predefined anchor boxes, the value of the batch size, the number of total iterations, the size of input images, etc.

3.1.10. The YOLOv4 algorithm

To realise further improvements in the detection accuracy of the proposed marine debris detector, the YOLOv4 was further considered to the problem of marine debris. The model architecture of the YOLOv4 tool is mainly subdivided into four parts: (1) the input, (2) the backbone, (3) the neck, and (4) the model head (Bochkovskiy *et al.*, 2020). Here, the term input refers to the input images. Functionally, the role of the backbone is to extract the important features from the input images while the neck is responsible to propagate these features from the backbone to the model head. Ultimately, the model head makes the predictions drawing on the input. More specifically, the CSPDarknet53 network is the feature extractor backbone of the YOLOv4 tool. This network is based on the Cross Stage Partial Network (CSPNet) (Wang *et al.*, 2020) which is added on the basis of the Darknet53 model of YOLOv3. Notably, the Darknet53 model uses residual connections to guarantee that the network possesses a depth while CSPNet improves the learning ability of the network and reduces the computational time and memory expenditure.

The YOLOv4 model neck implements the Spatial Pyramid Pooling (SPP) (He *et al.*, 2015) and the Path Aggregation Network (PANet) (Liu *et al.*, 2018). SPP enhances the receptive field of the backbone and separates the most significant context features, which cause almost no reduction in the network's operational speed. PANet aggregates parameters from individual backbone levels for different detector levels. In other words, the PANet shortens the path connecting low-level and high-level information. Evidently, YOLOv4 makes predictions on images in the same manner as the YOLOv3 algorithm, which is the model head. To sum-up, a number of modifications were made on various parameters of the YOLOv4 basic model which aimed at enhancing its detection performance. Changes relate to the size of the predefined anchor boxes, the value of the batch size, the number of total iterations, the pool of the input images and others.

3.1.11. The YOLOv5 algorithm

One step further, the YOLOv5 object detection algorithm (Jocher *et al.*, 2020) was applied to the marine debris classification task in order to investigate whether it managed to improve the ability to discern more plastics. Figure 7 illustrates the YOLOv5 model architecture. Functioning together the Cross Stage Partial Network (CSPNet) and the Focus structure (Wang *et al.*, 2020) made the feature backbone of the YOLOv5 model which extracted the important attributes from the input images. Structurally, each BottleneckCSP module consists of two convolutional layers comprised of a 1×1 and a 3×3 filter. Finally, the scheme employs the Spatial Pyramid Pooling (SPP) layer (He *et al.*, 2015) and the Path Aggregation Network (PANet) (Liu *et al.*, 2018).

The YOLOv5 tool also utilises an upsample layer which upgrades the current resolution of a layer to that of the previous layer. Additionally, the particular scheme employs a concatenation layer which accepts inputs of the same size along a specified dimension and merges them into one output. Collectively, the YOLOv5 embraces four distinct models ranging from the smallest YOLOv5s, with 140 layers, to the largest YOLOv5x with 284 layers. Concerning this research work, the largest

YOLOv5x model was selected as it yields the highest mean average precision when applied to the MSCOCO dataset (Lin *et al.*, 2014).

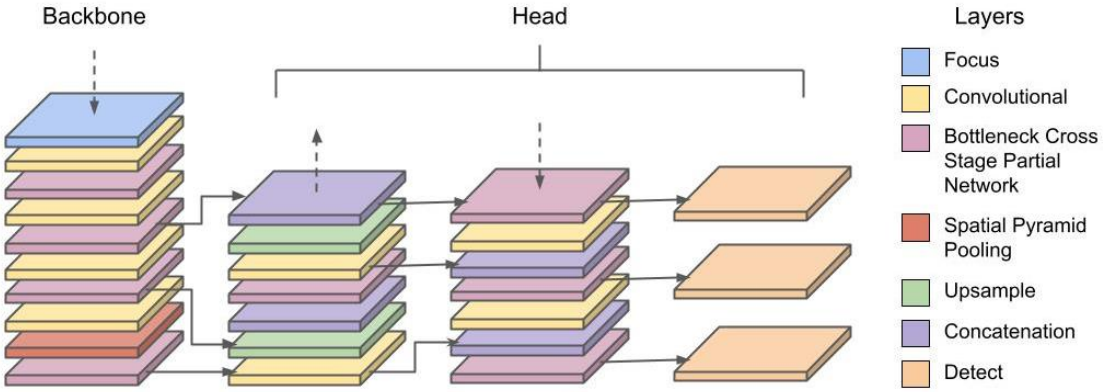


Figure 7: The YOLOv5 model architecture.

To this end, a number of parameters of the YOLOv5 model were altered so as to enhance its detection performance. Analytically, the dimensions of the predefined anchor boxes, the value of the batch size, the number of total iterations, the number of the input images were some of the modifications adopted during the training and testing phases of this scheme. Furthermore, to measure the number of litter items illustrated in videos, both the ROI line counting tool and the centroid tracking technique were implemented into the original YOLOv5 model, as described in §A2 in the Appendix, respectively. In turn, for the centroid tracking counting method, a customised *thresholdFrames* parameter was introduced which considers the minimum number of frames that a detected item must be present on screen for the algorithm to count it.

3.1.12. The YOLACT++ algorithm

Considering that the YOLO series can only attach a rectangular bounding box on each detected litter item which may enclose unrelated objects from the object's background, the focus has shifted on more advanced DL tools which are suited for segmenting the area of the detected litter. For this reason, the YOLACT++ instance segmentation algorithm (Bolya *et al.*, 2020) was tested here as it can detect, localise

and segment multiple marine debris items and sea life in still images and video recordings. That means that the particular scheme applies, other than the bounding box and the class label, a coloured mask on each recognised item. Interestingly, the mask drawn on each item can provide valuable information about the body's physical dimensions.

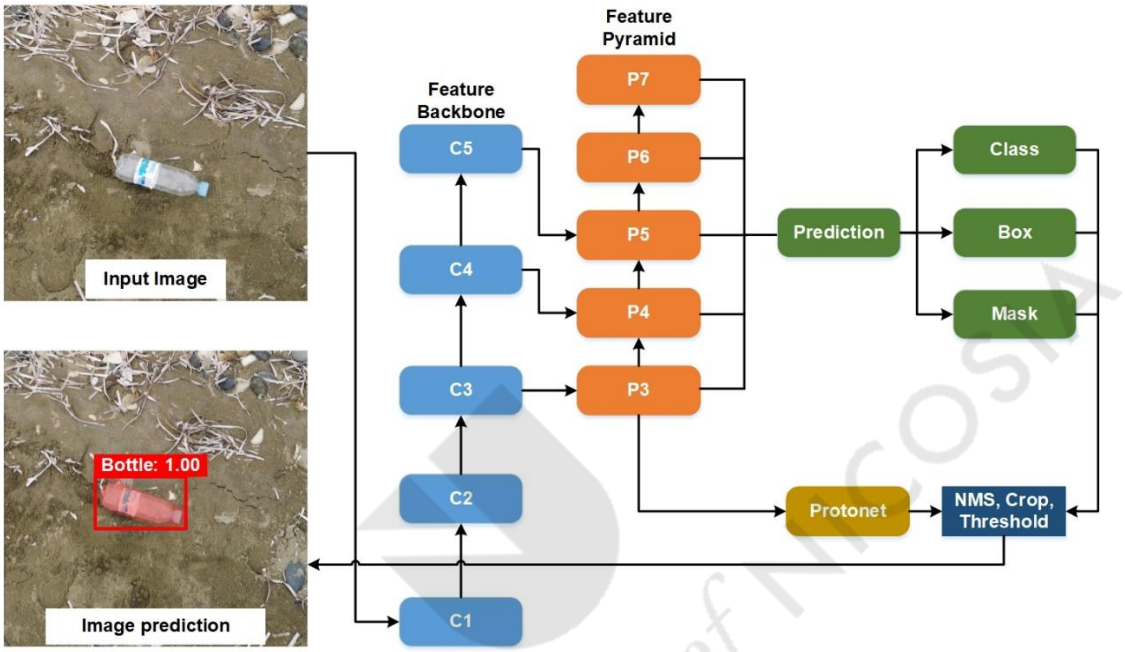


Figure 8: The YOLACT++ model architecture.

Algorithmically, YOLACT++ is a one-stage instance segmentation approach that is subdivided into two parallel tasks. Initially, it produces a set of prototype masks. Then it predicts the per-instance mask coefficients. Followed then, it linearly combines the output from the previous two tasks to produce the final instance masks. The architecture of the YOLACT++ algorithm appears in Figure 8. This structure uses Residual Networks-101 (ResNet-101) (He *et al.*, 2016) with Feature Pyramid Networks (FPN) (Lin *et al.*, 2017a). Hence, the YOLACT++ model consists of blue coloured blocks (C1–C5), which indicate the ResNet-101 feature backbone layers with low values in the prototypes, and orange blocks (P3–P7) which indicate the FPN layers with high values in the prototypes. At the end, the mask coefficients are

generated in the prediction part as well as the prototype masks in the prototype generation branch (protonet).

Functionally, the protonet predicts a set of prototype masks for the entire image by using the P3 block which is the largest and the deepest feature layer. In turn, this enables the assignment of higher quality masks and enhances the performance of the algorithm on smaller objects. In parallel, four bounding box regressors and a class confidence are generated for each prototype mask. Prototype masks and mask coefficients are funnelled through the non-maximum suppression (NMS) and score thresholding, where only detections with high confidence can survive. The dominant mask is ultimately cropped with the imposed bounding box to produce the final mask prediction on each object.

3.1.13. The ROI line counting technique

Another line of investigation of this study dealt with the estimation of the abundance of plastic litter from video footage using two counting methods. First, the region of interest (ROI) line counting method was utilised, which imposes a virtual line on a video and tallies the number of objects that cross it. The ROI line counting process is accomplished by considering the coordinates of the centre of the bounding box of each detected object (Mazurek, 2020). Each predicted bounding box applied on an individually recognised object exhibits coordinates generated by the object detection algorithm. Simultaneously, the ROI line is drawn on the screen which possesses its own coordinates while its location remains fixed for the duration of the video. Thereafter, objects detected by the image classifier and whose coordinates cross this drawn line are finally summed up by the counting method. Operationally, an object's bounding box is displaced over a particular number of pixels between frames. Hence, the ROI line needs a certain thickness such that the centre of the object's bounding box falls within the line at some point in time.

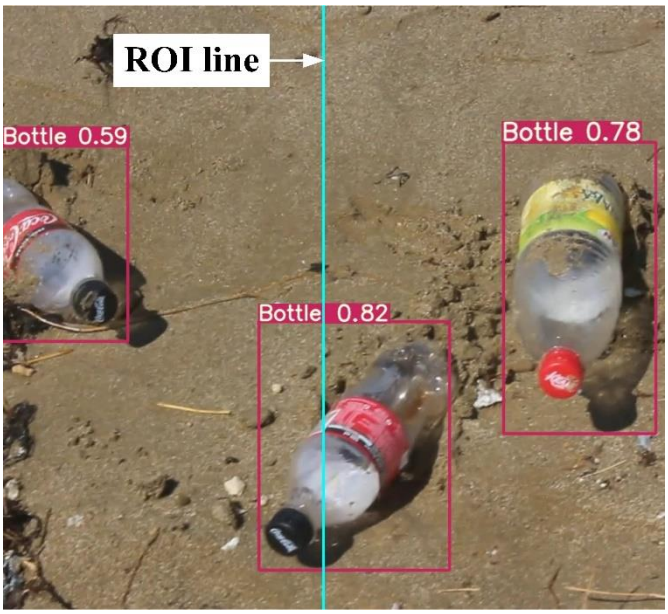


Figure 9: The ROI line which was placed in the middle of the video frame.

Once a detected object has passed the ROI line, the coordinates of the centre of its bounding box were henceforth deduced. Next, that point was checked if it lies within the ROI line, which was placed in the middle of the frame, as depicted in Figure 9. Each bounding box has also a corresponding label. Thus, when the bounding box crossed the ROI line, the item count for that particular class label was incremented by one. Likewise, the counting tool scans all bounding boxes present on screen during the video and checks if their coordinates intersect the ROI line.

As of presently, implementations of the ROI line counting tool have only addressed the case where all objects that are crossing the line are of the same category, for example, people or vehicles, and no other category (Bas *et al.*, 2007; Chen *et al.*, 2008; Memon *et al.*, 2018; Yang *et al.*, 2019). Thus far, to the best of our knowledge, the research proposed here constitutes the first attempt to count objects from up to nine different object categories. Once the original code of the ROI line counting method was implemented in the YOLOv5 model, it was modified accordingly so as to enable the detection and counting of items from nine distinct classes, which comprised plastic bottles, bags, buckets, fishing nets, plastic straws, food wrappings, a fish species, aluminium cans, and cigarette butts.

3.1.14. The centroid tracking counting method

The second counting technique utilised here for calculating the abundance of plastic debris and sea life displayed in a video footage was the centroid tracking method. This technique tracks identified objects as they propagate from one video frame to another frame and assigns them unique identities (IDs) (Khachatryan, 2019; Rosebrock, 2021a; Rosebrock, 2021b). Effectively, the centroid tracking method considers the centre of the bounding box of a detected object as its centroid and assigns to it an identity (ID), as illustrated in Figure 10.

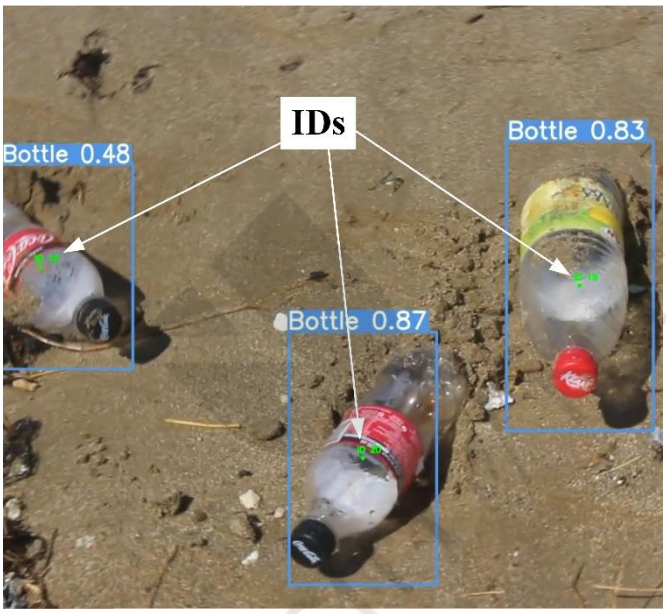


Figure 10: The centroid tracking technique tracks identified marine debris items by assigning them an identity (ID).

Operationally, the centroid tracking method accepts as an input a set of bounding boxes along with their respective coordinates and class labels for each frame and creates unique IDs for each of the detected item. In this way, the method tracks the identified bodies as they project from frame to frame. Centroid tracking is performed under the assumption that for a given object, the Euclidean distance between the two centroids for frames f_t and f_{t+1} is smaller than all other distances between the objects (Rosebrock, 2021a). To this end, object tracking in videos is accomplished by

registering the centroids of the bounding boxes of the detected objects as these are moving from one frame into the next.

The centroid tracking method features the *maxDisappeared* parameter which defines the number of consecutive frames during which an object remains undetected before it is deregistered, or no longer tracked (Rosebrock, 2021b). Additionally, to overcome the flaw of counting the same item multiple times due to the incorrect categorisation between frames, a customised *thresholdFrames* parameter was introduced which considers the minimum number of frames that an object must be present on screen for the algorithm to count it. These frames are not necessarily successive, that is, as long as a given object is cumulatively screened for a *thresholdFrames* number of frames before it is deregistered, that object will be counted. Lastly, similar to the ROI line counting method, the original python code of the centroid tracking technique was added to the YOLOv5 model while it was adapted accordingly to enable the recognition and counting of various classes of objects.

3.2. The detection of microplastics

3.2.1. The hydro-acoustic device

This subsection presents the hydro-acoustic device, namely, the ISA500 sensor (Impact Subsea, 2016), utilised for the detection of microplastics in the fresh water. Operated at high frequencies, in underwater conditions, the ISA500 sensor can provide proximity measurements with a very high accuracy even at long stretches spanning up to 120 meters. The main specifications of ISA500 hydro-acoustic device are outlined in Table 3. Notably, the default frequency of operation of the sensor is 500 kHz but it can function at a range of frequencies between 400 and 600 kHz.

Operationally, the hydro-acoustic device exploits sound wave pulses in order to determine the distance. Sound wave pulses propagate in the water until they encounter an obstacle. When a pulse hits the body, part of the pulse is reflected back to the ISA500 sensor. Upon reaching the sensor, the reflected sound signature is detected by the sensor's receiver and the time needed for the pulse to make this journey is recorded. Then, the distance of the obstacle is computed as a product of speed and

time, where the speed of sound in the water is typically around 1,500 m/s. Finally, the distance is 50% of that journey.

Table 3: Specifications of ISA500 sensor. Adapted from Impact Subsea (2016).

Parameter	Characteristics
Frequency	500 kHz default (400–600 kHz selectable)
Range	0.1 to 120 m
Resolution	1 mm
Beam angle	6º conical at 500 kHz
Signal	Monotonic
Pulse length	User defined

The operation of ISA500 sensor relies on the sonar equation:

$$\left[\mathrm{SL}-\mathrm{TL}-\left(\mathrm{NL}-\mathrm{DI}\right)\right]>\mathrm{DT} \tag{15}$$

where the Detection Threshold (DT) defines the minimal signal to noise ratio required for the sonar device to detect an acoustic signal. The Source Level (SL) is the power the acoustic pulse possesses as it enters the water, while the Transmission Loss (TL) refers to the dispersion of the acoustic pulse in the water. The Noise Level (NL) is an environment specific variable which is governed by the noise originating from marine life, wind, rain, anthropogenic sources and other machinery. Finally, the Directional Index (DI) reduces the noise level by utilising a 6º conical acoustic beam.

CHAPTER 4 RESULTS AND DISCUSSION

4. Results and discussion

This chapter is dedicated to the research findings of this study. First, the results from the utilisation of the ML tools like the six feature detectors and descriptors—SURF, FAST, BRISK, HARRIS, MinEigen, and MSER algorithms —and the BoF method on the detection of plastic debris are outlined. Next, the classification results of the DL tools which comprise the BM, YOLOv3, YOLOv4, YOLOv5, and YOLACT++ tools as applied to marine plastics are presented. Thereafter, the findings from the utilisation of two counting techniques, namely the ROI line and the centroid tracking methods, on tallying marine litter in real field video recordings are offered. Finally, the observations from the discovery of microplastics in the water using the ISA500 hydro-acoustic device are listed in this section.

4.1. Detecting plastic bottles in images

Initially, the six feature detectors and descriptors, namely, the SURF, FAST, BRISK, HARRIS, MinEigen, and MSER were evaluated on the discovery of the important image attributes from images illustrating plastic debris. Specifically, each algorithm was assessed to find the image features from a query object, that is, a plastic bottle in this case, found in an image scene. Figure 11 depicts the procedure adopted for the detection of a plastic bottle. Initially, each algorithm "reads" the query object and identifies its points of interest, which are the important features of this object. Similarly, this procedure was also adopted for the detection of points of interest in an image scene that includes the query object. Thereafter, the algorithm matches the points of interest from both the query object and the image scene. Finally, the method locates the query object, which is a bottle, in the image scene by applying a rectangular bounding box around it.

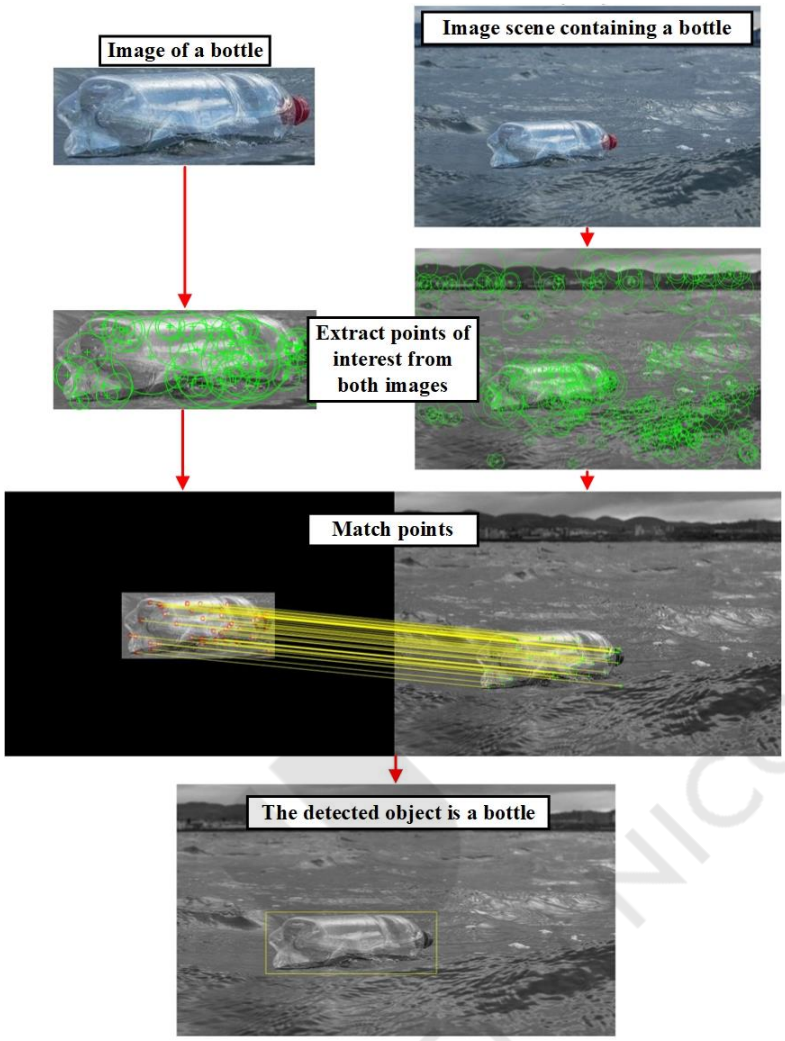


Figure 11: Flow diagram depicting the process, adopted by the six feature detectors and descriptors, applied to the discovery of a plastic bottle.

The six feature detectors and descriptors extract different kinds of features from the images, as demonstrated in Figure 12. For this reason, it was deemed necessary to evaluate the capabilities of the feature detectors and descriptions on more images capturing plastic debris. For that purpose, ten distinct images illustrating plastic bottles, as depicted in Figure 13, were employed. During the appraisal of all six feature detectors and descriptors on the detection of ten different images featuring plastic bottles, the computational time and the overall detection accuracy for each algorithm, as listed in Table 4, was recorded. The computational time is the temporal duration needed for each algorithm to identify and locate the plastic bottle in an image

scene while the overall detection accuracy is the proportion of the successfully identified samples over all tested samples.

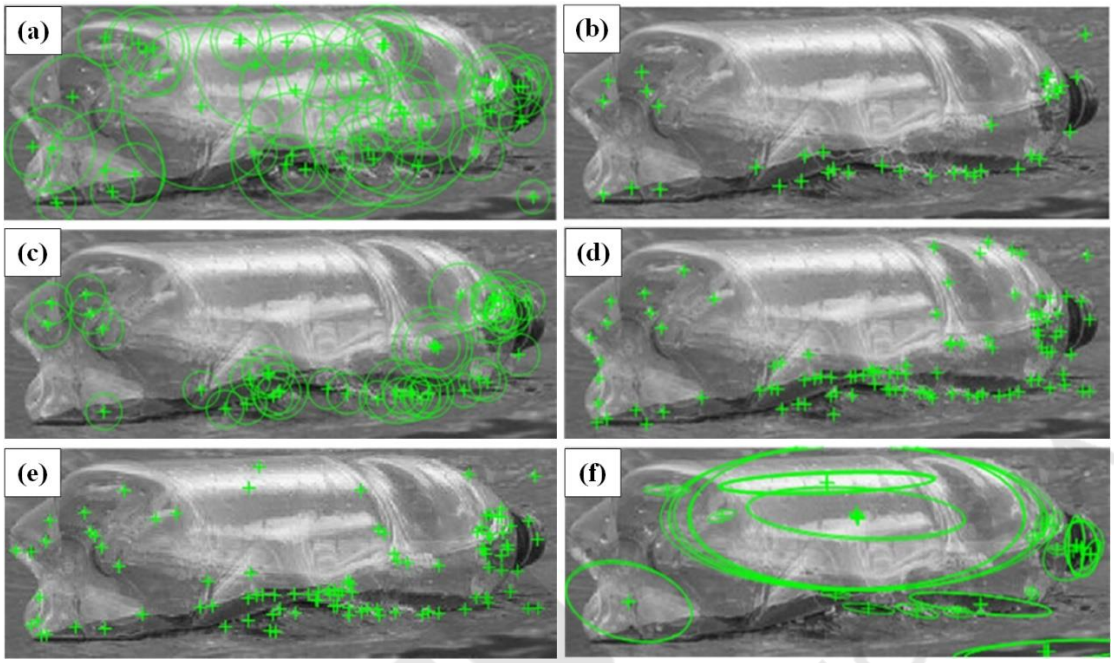


Figure 12: The detected points of interest from the six feature detectors and descriptors.

Observing Table 4, it is evident that the SURF, the MinEigen, and the MSER algorithms attained the best detection accuracy. However, the SURF tool realised the fastest computational time for each detected bottle outperforming the rest detectors in terms of the accuracy-speed trade-off, which normally ranges between 60 to 80% for about 100 to 200 ms (Laptev *et al.*, 2008; Heitz, 2014). Specifically, the algorithm attained an overall detection accuracy of 60% and a computational time of 3 s for each image. Reasonably, the SURF was chosen to be the feature detector and descriptor responsible for obtaining the points of interest from images illustrating plastic litter in the marine environment.

Figure 13: Examples of ten images of plastic bottles tested by the six feature detectors and descriptors.

Table 4: Detection performance of the six detectors and descriptors when applied to ten images displaying a plastic bottle, shown in Figure 13.

Algorithm	Computational Time (s)	Overall Detection Accuracy (%)
SURF	3	60
FAST	3	30
BRISK	4	30
HARRIS	3.25	40
MinEigen	4.42	70
MSER	4.66	60

The need of identifying and localising more object categories while enhancing the sophistication of the image classifier provided the impetus for employing the Bag of Features (BoF) method. The BoF constructed an image classifier which was trained and tested on several images of plastic litter and sea life. Subsequently, it was validated on newly provided images of the same object categories. The main findings of this evaluation are presented in the following sub-section.

4.2. Classification of plastic litter images

The Bag of Features (BoF) method was initially used to construct an image classifier designed to categorise images depicting plastic debris and sea creatures in the marine environment. Firstly, an image dataset was created for training and testing the BoF image classifier. The points of interest from images belonging to this dataset were extracted using the SURF algorithm. Practically, the dataset consisted of eight categories of objects: six types of plastic marine debris and two types of marine life. The eight classes comprised: (1) plastic bottles, (2) plastic bags, (3) plastic buckets, (4) fishing nets, (5) polystyrene pieces, (6) plastic fishing buoys, (7) dolphins, and (8) sea turtles. Each category consisted of 35 images, which overall amount to 280 images. All images of this dataset were retrieved from the ImageNet dataset (Deng *et al.*, 2009).

Due to the insufficient number of images and knowing that conventional ML image classifiers, like the BoF, require thousands of images to enhance their classification accuracy, data augmentation (DA) manipulations were utilised. Here, DA manipulations refer to image processing techniques applied to the images of a dataset with the intention of creating a larger pool of images. During DA adjustments, images are subject to random transformations and normalisation operations. Rotation of the image, random alteration of the image's width and height, vertical or horizontal translation of the image, rescaling, random change in the shear range, zooming, horizontal flipping or filling the image with new pixels after a rotation or shift are some of the operations. After the DA operations, the image dataset grew from 280 images to 10,400 images, with each category containing 1,300 samples.

The augmented dataset was later divided into two subsets: (1) the training and (2) the test set. As the names dictate, the training set was used to train the BoF image classifier while the test set was utilised to test its classification capabilities. The training and testing processes of the BoF image classifier were performed on an Intel® Xeon® (CPU 2.40 GHz) processor with an NVIDIA Quadro® K4200 graphics card. Findings of the training and the testing assessments, are shown in Figure 14 and Figure 15.

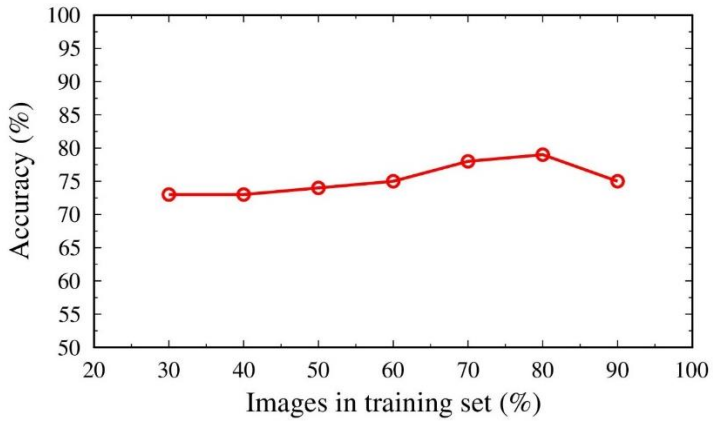


Figure 14: The accuracy of the BoF image classifier when varying the number of images in the training set.

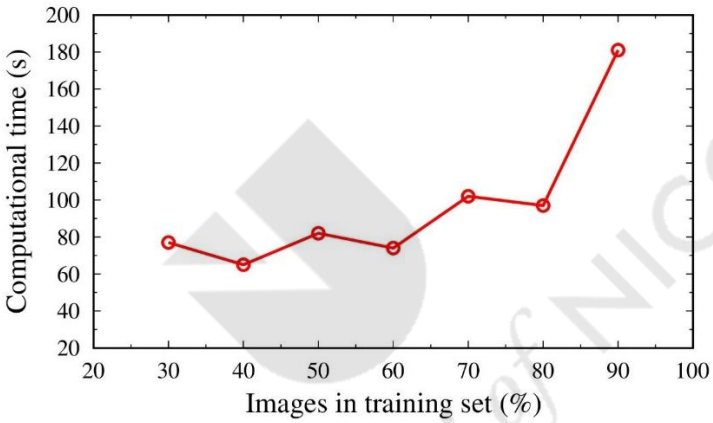


Figure 15: The computational cost of the BoF image classifier when altering the number of images in the training set.

Figure 14 presents the classification accuracy of the BoF method while Figure 15 depicts the overall computational time by varying the number of images allocated to the training set. Here, the classification accuracy is the proportion of the correctly identified images over all tested samples while the computational time is the total time needed for the testing process. Observing these two figures, results illustrate that the proportion of 80% of the images allocated to the training set and the remaining 20% of images in the test set generates the highest classification accuracy of 79%, while the computational time remains at the same level of almost 90 s. Meanwhile, other studies mention that the optimal ratio for splitting images into the training and test sets is 80% and 20%, respectively (Roshan, 2022). Allocating 90% of images to

the training set gives the BoF the opportunity to assimilate features from more images, but this is accompanied by a surge in the computational time. Seventy percent of images or less allocated in the training set produced inferior testing accuracy.

Once the training and testing procedures were completed, the BoF image classifier was validated on a totally new image dataset. This dataset is called the validation set and is responsible for evaluating the classification capabilities of the BoF method. The validation set contained 40 images in total or five images in each of the eight object categories. Following the completion of the validation process, a classification accuracy of 62.5% was obtained, meaning that 25 images out of 40 were correctly categorised by the BoF image classifier.

Drawing on the previously mentioned results, it was realised that there is a need of enriching the training set with more images so as to enhance the classification capabilities of the BoF method. However, due to the conventional architecture of the BoF scheme, a possible increase in the number of training image samples is expected to warrant a higher computational cost without a marked improvement in the classification accuracy. At the same time though, DL algorithms as applied to other detection tasks attained superb performance. Based on the previous reasoning, various DL tools were applied to the marine plastics classification task so as to investigate whether they can enhance the sophistication of the proposed image classifiers.

The first DL tool utilised for categorising images of marine debris was the BM. The classification capabilities of this scheme were tested on categorising images of plastic debris floating on or near the sea surface. Initially, the BM image classifier distinguishes between three categories of litter, that is, (1) plastic bottles, (2) plastic buckets, and (3) plastic straws. The images belonging to this dataset were obtained from the ImageNet dataset (Deng *et al.*, 2009). Each of the three categories consisted of 250 images collectively adding up to 750 images. DA manipulations were performed here which refer to geometrical transformations applied to images, e.g., zooming, shifting, flipping, rotation, etc. As a result, the enlarged dataset contained a total of 4,000 images in each object category (Kylili *et al.*, 2019).

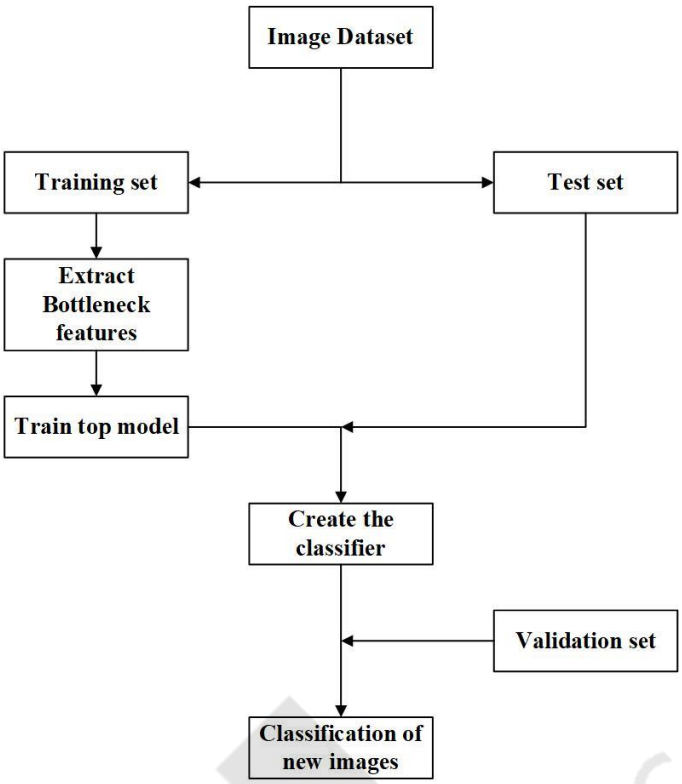


Figure 16: Flowchart depicting the overall classification process of the BM when applied to images of plastic marine debris.

Followed then, the augmented dataset was randomly split in two subsets, namely, the training set and the test set. A share of 80% of the images was assigned to the training set whereas the remaining 20% of the images to the test set. Consequently, each training category contained 3,200 images, while discrete testing classes featured 800 images. In other words, the BM image classifier was trained on 9,600 images and tested on 2,400 images. Having completed the training and testing procedures, the validation set was employed in order to evaluate the identification performance of the BM image classifier. The preceding process is summarised in Figure 16.

Computationally, the training and the testing processes of the BM image classifier were executed on an Intel® Xeon® (CPU 2.40 GHz) processor with an NVIDIA Quadro® K4200 graphics card. Table 5 presents the hyperparameters and their optimal values that were selected during the training phase of the particular model. As listed in Table 5, the optimal batch size was set to 10 images, the dropout value was fixed to

0.25, a learning rate of 0.001 was chosen whereas the Adadelta optimiser was employed during the training stage of the BM image classifier.

Table 5: Hyperparameters and their optimal values used during the training stage of the BM image classifier.

Hyperparameter	Tested value(s)	Optimal value
Batch size	10 images	10 images
Dropout value	0.25, 0.50	0.25
Learning rate	0.0001, 0.001, 0.01	0.001
Optimiser	Adam, SGD, RMSprop, Adadelta	Adadelta

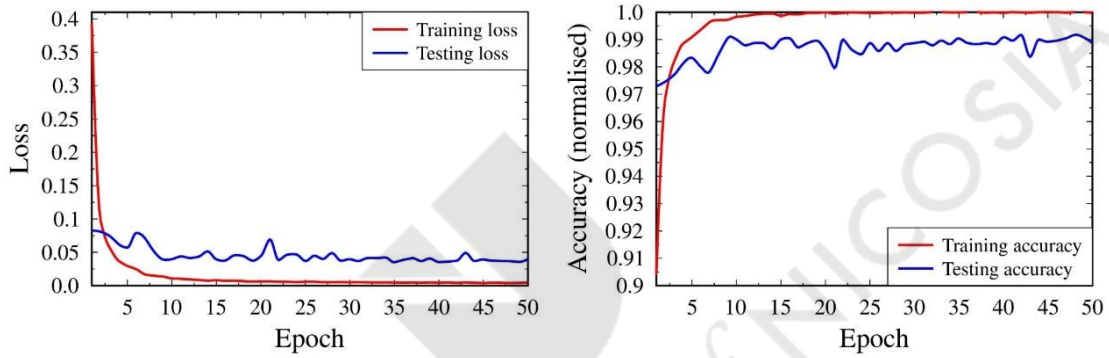


Figure 17: Plots display the training and testing loss (left) and the respective accuracy (right) over 50 epochs.

Figure 17 displays the training and testing performance of the BM image classifier with respect to loss (left plot) and accuracy (right plot). Both the loss and accuracy results were obtained from the training and test sets over 50 epochs. The number of epochs indicates the times that the entire dataset is processed through the network to produce the best possible classification outcome. Converging trends between training accuracy and loss curves reinforces the credibility of results. The training accuracy of the BM image classifier reached a maximum of almost 100% with a loss of about 0.01, while the testing accuracy topped at about 99% with a loss of almost 0.04.

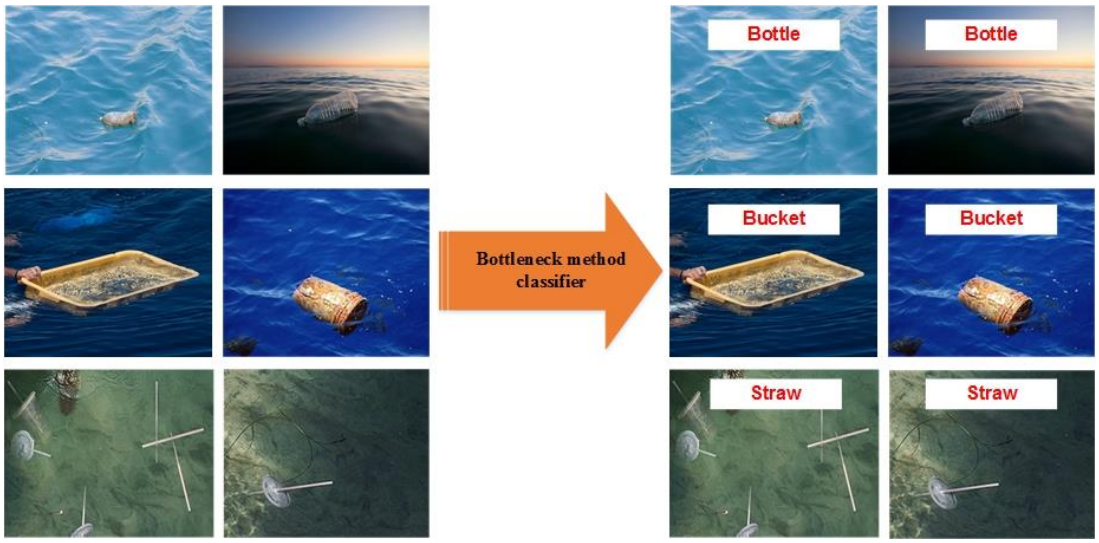


Figure 18: A matrix of images demonstrates the ability of the BM image classifier to correctly categorise images illustrating plastic bottles, buckets and straws.

Once the training and the testing phases of the BM image classifier converged, the validation assessment followed. Individual validation sets consisted of 55 newly seen images of plastic marine debris per object category or a total of 165 validation samples. During this appraisal, the BM image classifier correctly identified 53 images of plastic bottles, 55 images of plastic buckets, and 33 images of plastic straws. Overall, recognising a total of 141 out of the 165 freshly provided images of plastic marine debris resulted in a validation accuracy of almost 86%. Being relatively high, the validation accuracy lends credibility to the effectiveness of the proposed BM image classifier. Concluding, a selection of correctly labelled plastic marine debris images is illustrated in Figure 18.

As part of a rigorous effort to further scrutinise the trustworthiness of the BM image classifier, it was deemed necessary to examine three scenarios. The first one compared different types of regularizers as applied to the context of the BM training process. Secondly, a parametric investigation explored the performance of the BM image classifier by varying the number of images in the test set. Finally, the performance of the classifier was tested as a function of the number of images generated from the DA manipulations.

The first scenario was designed to assess the performance of the BM image classifier by implementing different regularizers during the training process. Regularizers permit the assignment of "penalties" on the parameters of the model's layers or on a layer's activity during the optimisation stage. These penalty values are added in the categorical cross-entropy loss function (equation (6)) used to optimise the network weights. Accordingly, a particular regularizer can improve the performance of the image classifier and lower the risk of encountering overfitting in the learning process.

Overall, the three regularizers that were examined during the first scenario included: ℓ_1, ℓ_2 and $\ell_1\_\ell_2$. Symbol ℓ_1 refers to the Least Absolute Shrinkage and Selection Operator (Lasso) regression (Hastie *et al.*, 2009) which assigns zero values to the weights of uninformative features by virtue of shrinkage. To improve the classification accuracy, the Lasso regression method adds the absolute value of magnitude of the coefficient as a penalty term to the loss function. Regularizer ℓ_2, also known as the Ridge regression method (Hastie *et al.*, 2009), prevents excessive fluctuations in the weights and forces them assume small values by inserting the squared magnitude of the coefficient as the penalty term to the loss function. Denoted by $\ell_1\_\ell_2$, this regularizer combined the two previous regression methods. Lastly, to enable the comparison between results one of the training processes was conducted in the absence of a regularizer.

The performance of the BM image classifier using different regularizers and the absence of a regularizer is depicted in Figure 19. Although the case lacking a regularizer yields the highest testing accuracy, yet its increasing loss over the epochs points to overfitting. For that reason, this case was discarded from the investigation. Since the training accuracy for all cases reached a value of almost 100%, for comparison purposes, we focused on the testing accuracy. Likewise, because the training loss for all cases was almost negligible, it was decided to exclude these findings from the respective plots. Notably, the case which blended the Ridge and the Lasso regressions appears to perform better than using each regression method distinctively and produced the highest classification accuracy. More precisely, when

both the testing accuracy and the testing loss were considered, regularizer $\ell_1\_\ell_2$ was selected as it fared slightly better than the rest regularizers.

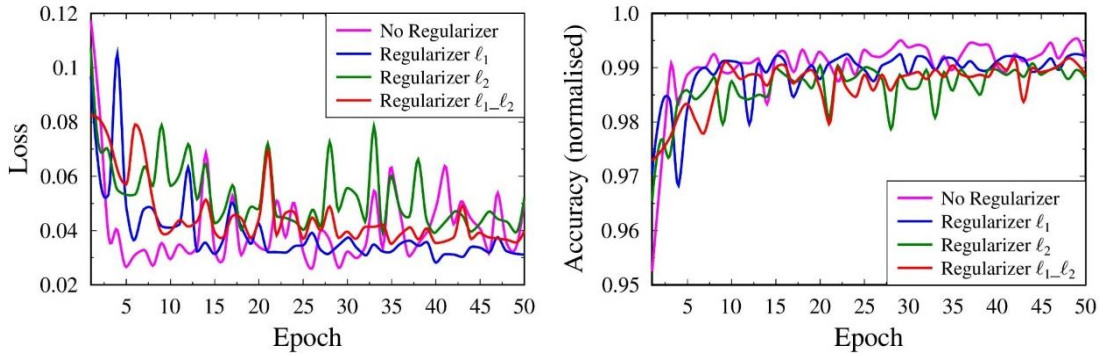


Figure 19: The testing loss (left) and testing accuracy (right) of the BM image classifier in the absence of a regularizer and accompanied by different regularizers.

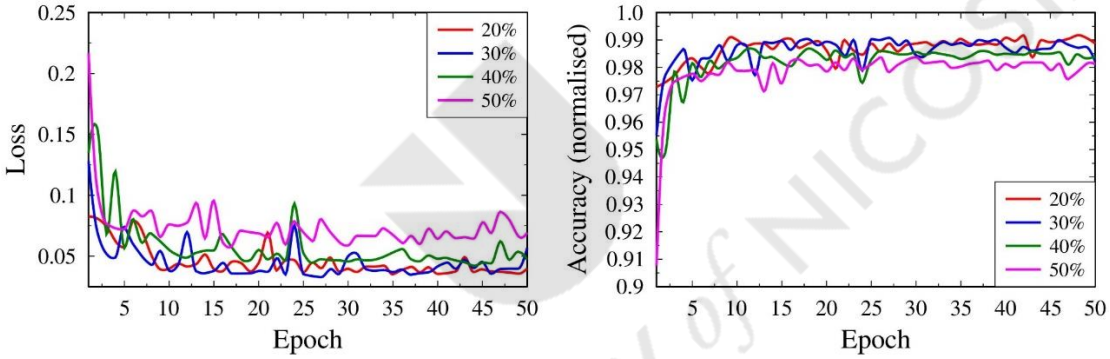


Figure 20: Left and right plot depicts the testing loss and testing accuracy of the BM image classifier, respectively, over 50 epochs, by increasing the proportion of the images allocated to the test set from 20 to 50%.

The second line of investigation inspected the classification performance of the BM by altering the number of images allocated to the test set. Specifically, the percentage of the images fed into each test set comprised progressive proportions of 20%, 30%, 40%, and 50% of the images. All cases utilised the $\ell_1\_\ell_2$ regularizer. For the 20% case, 3,200 images made-up the training set and 800 images the test set. In an analogous fashion, for the 30% case, the training set consisted of 2,800 images while the test set included 1,200 images. Respectively, for the 40% case, 2,400 images made-up the training set and 1,600 the test set. Finally, in the 50% case, the number of images was equally split into 2,000 in the training and 2,000 in the test set.

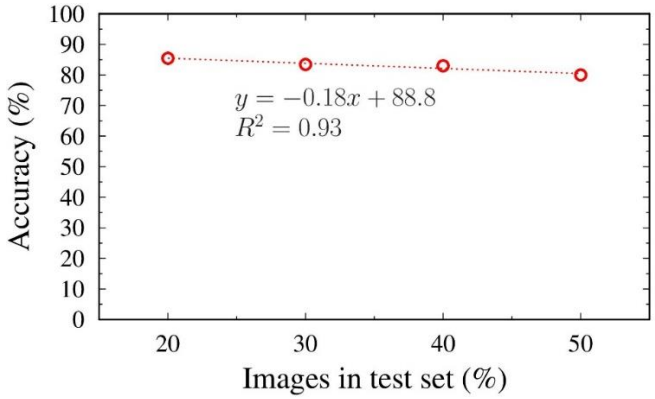


Figure 21: Validation accuracy of the BM image classifier while adjusting the number of images assigned to the test set.

The left graph in Figure 20 shows the testing loss and the right the testing accuracy over the 50 epochs for all four cases which utilised 20%, 30%, 40%, and 50% of the images as a test set. As one can observe, the smallest the size of the image set, which made-up the test set, the better was the performance of the BM image classifier. This result emerged because the training set had a larger pool of images and the method's accuracy improves when the classifier processes more representations of images in the training set and learns on them. Reasonably, a further increase in the number of images in the test set would lead to a small drop in the testing accuracy. The same observation is also supported by Figure 21 which displays the validation accuracy of BM on newly provided data as a function of the number of images allocated to the test set. By fixing the fraction of images in the training set to 0.8 with the remaining 0.2 assigned to the test set, the validation accuracy reached its highest value at almost 86%. In other words, the BM image classifier successfully identified 141 out of the 165 images which formed the validation set.

The third and final scenario examined the identification capabilities of the BM by varying the number of images generated after DA manipulations. The number of images in each object category was increased from 1,000 to 4,000 in increments of 1,000, as illustrated in Figure 22. Each case was run with regularizer $\ell_1\_\ell_2$ while the test set encompassed 20% of the images. Referring to the right graph of Figure 22, one can observe that the larger the number of images in each object category, the better was the testing accuracy of the BM image classifier. Owing to the larger image

set, the classifier becomes more efficient in identifying plastic marine debris articles as it learns on more image representations. Concluding, a comparable trend can be noticed in Figure 23 where the validation accuracy the BM image classifier attained over freshly provided images improves as the number of images per category expands from DA manipulations.

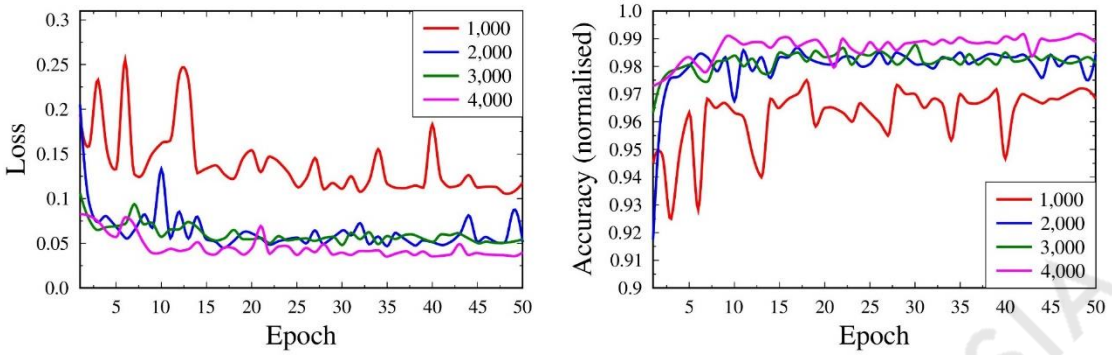


Figure 22: The testing loss (left) and accuracy (right) of the BM image classifier for an increasing number of images in each object class ranging from 1,000 to 4,000, as generated by DA.

Given that the types of plastic litter encountered in the marine environment are more than three, the number of object classes was broadened from three to eight. For this reason, to undertake a more demanding classification process, a more sophisticated BM image classifier is proposed hereafter. More specifically, this classifier can now distinguish between six types of plastic debris and one kind of marine life. Additionally, the new BM image classifier is also able to recognise objects that are neither plastics nor marine life and rank them in the "other" category. Lastly, the scheme is also able to classify marine debris found both at the sea and the shorelines (Kylili *et al.*, 2020).

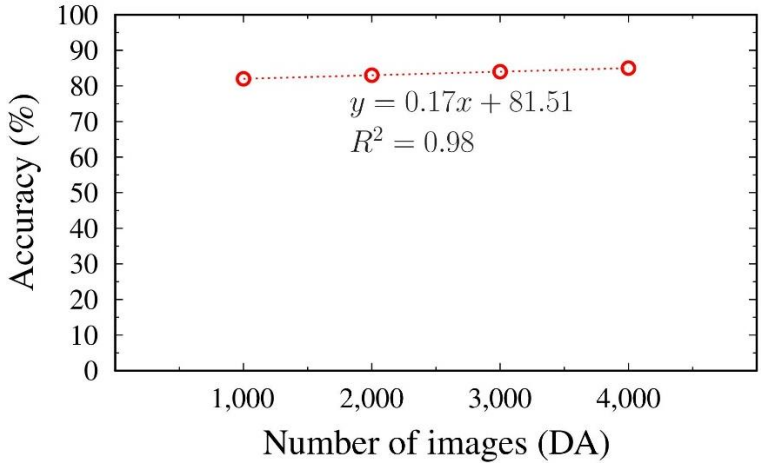


Figure 23: Plot of validation accuracy versus an increasing number of images in each object category obtained from data augmentation (DA).

The dataset used to train the more sophisticated BM image classifier consists of eight categories of items, namely, (1) plastic bottles, (2) plastic buckets, (3) plastic bags, (4) fishing nets, (5) plastic straws, (6) food wrappings, (7) flying fish, and (8) "other" which comprises articles such as boats, shipping containers, rocks, swimmers, etc. Notably, the addition of the "other" category enhances the usefulness of the proposed method by rendering it able to distinguish marine debris from some unrelated objects that can be encountered at the sea and the coast. The images of this dataset were mainly retrieved from ImageNet (Deng *et al.*, 2009) and the non-profit organisation Algalita which provided us with marine debris images and videos acquired during their boat expeditions dating in 2014 (Algalita, 2014b). Figure 24 displays a sample of these images in which object a) is a plastic bag while item b) is a plastic bottle floating at the sea surface. Partly submerged body c) in Figure 24 is a plastic bucket, image d) depicts a food wrapping in seawater, while label e) pictures a ghost fishing net about to be recovered onboard a boat. Image f), in Figure 24, displays plastic straws, while picture g) shows a flying fish. Lastly, item h), in Figure 24, is a motorboat, an example of an object drawn from the "other" class.

Collectively, each of the eight object classes contained 200 images. After DA manipulations, the augmented dataset comprised 4,000 images in each object category with the total number of images amounting to 32,000. A share of 80% of the images

was assigned to the training set whereas the remaining 20% of the images to the test set. Consequently, each training category contained 3,200 images, while discrete testing classes featured 800 images.

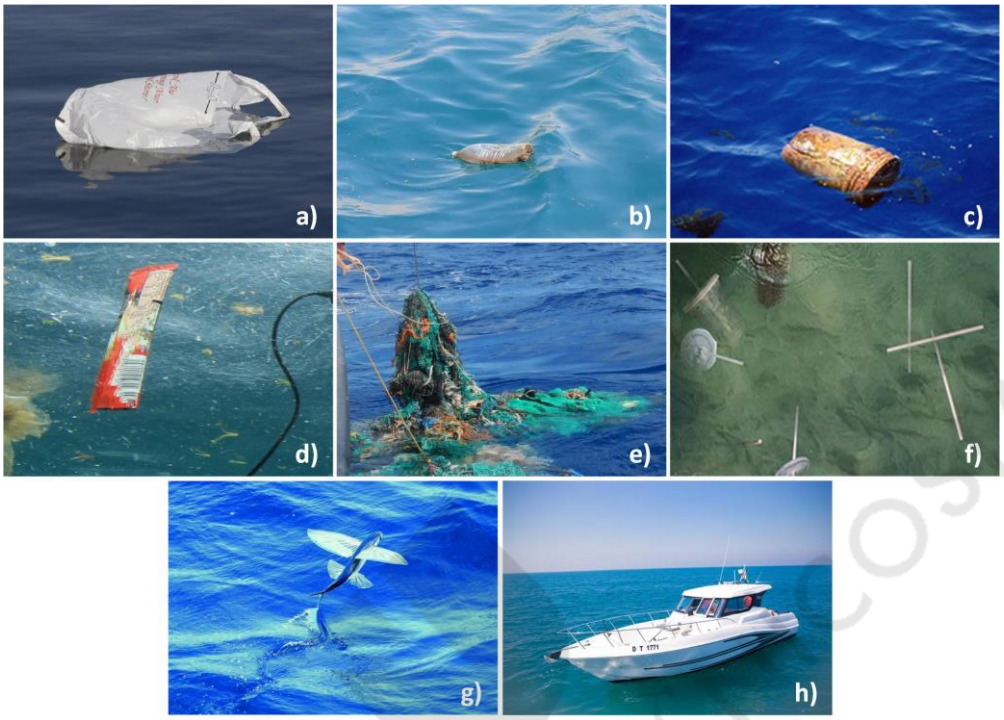


Figure 24: Examples of images retrieved from the marine debris and sea life dataset. Courtesy: ImageNet (Deng *et al.*, 2009) and the non-profit organisation Algalita (Algalita, 2014b).

Two scenarios were examined here in order to evaluate the performance and trustworthiness of the more sophisticated BM image classifier. The first scenario assessed the performance of the classifier by varying the number of epochs while the second scenario quantified its identification performance by altering the number of images allocated to each mini-batch. A mini-batch is a small group of training images for which the BM calculates an error, using equation (6), and updates its weights. By processing all mini-batches, an average error is computed for each epoch which shrinks as the number of epochs progresses.

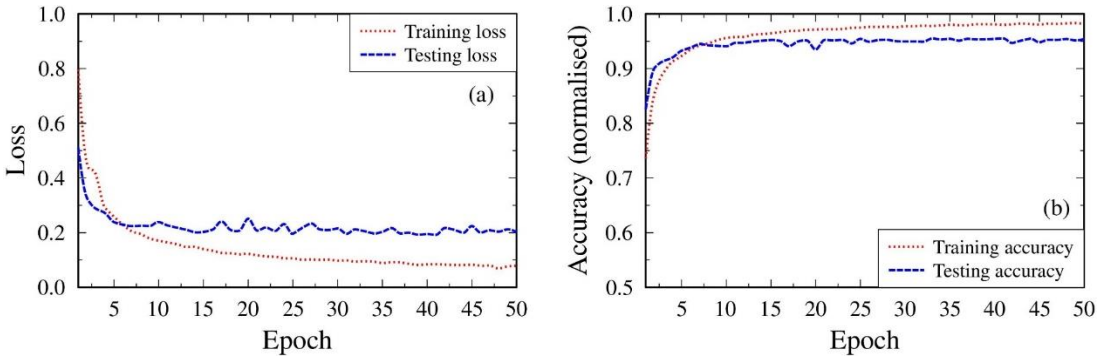


Figure 25: Graph (a) displays the training and testing loss curves of the BM image classifier, while plot (b) shows the respective accuracy curves for 50 epochs.

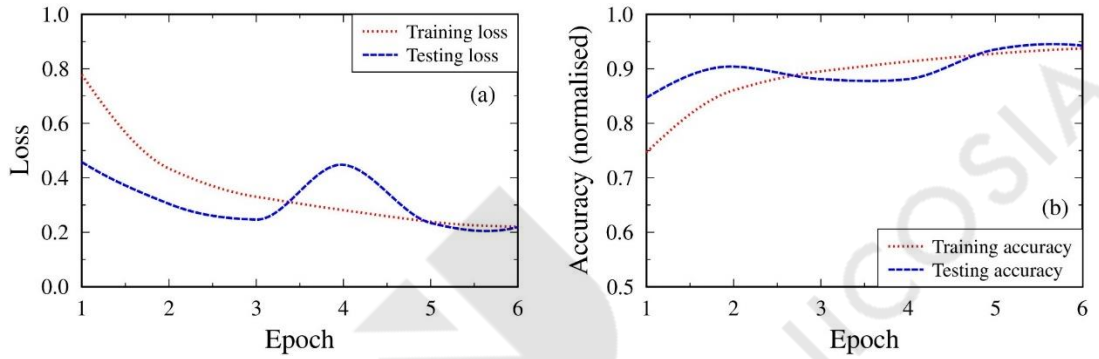


Figure 26: Plot (a) presents the training and testing loss of the BM image classifier while graph (b) shows the accuracy curves, respectively, for 6 epochs.

Regarding the first scenario, two cases were investigated during which the BM was permitted first to execute 6 epochs and then allowed to progress to 50 epochs. For this study, 10 images were allocated to each mini-batch. Findings pertaining to the training and testing performance of the 50 and the 6 epochs cases are displayed in Figure 25 and Figure 26, respectively. The fluctuations in the testing loss for 50 epochs (Figure 25 (a)) offered the inspiration to study the case of 6 epochs so as to investigate whether the BM image classifier past epoch 6 (cross-over between training and testing loss of Figure 25) embraces some unwanted image features during the training stage.

The 50 epochs case, as illustrated in Figure 25 (a) and (b), scored a training accuracy of 98% at a training loss of 0.09. The testing accuracy of the 50 epochs peaked at 95% at a testing loss of 0.20. Meanwhile, the 6 epochs case (Figure 26 (a)

and (b)) produced a training accuracy of 93% with a loss of 0.21. Likewise, the testing accuracy attained a value of 94% at a loss of 0.22. Comparing these two cases, the BM image classifier, which was permitted to run for 50 epochs, yielded a higher training and testing accuracy and a smaller error in relation to the 6 epochs case. Clearly, the BM image classifier improves progressively itself on every epoch, and if the number of epochs is adequate enough the overall accuracy would progress up to a certain level.

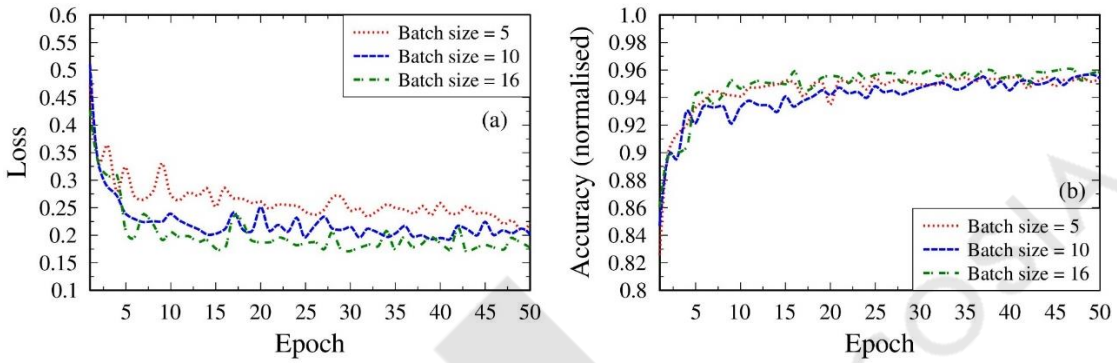


Figure 27: The testing loss (left) and testing accuracy (right) of the BM image classifier for three different batch sizes: 5, 10, and 16.

Referring to the second scenario examined here, the identification performance of the more sophisticated BM image classifier was assessed by altering the number of images allocated to each mini-batch (batch size), which featured 5, 10, and 16 images. For all cases, the number of epochs was set to 50. The performance of the BM image classifier, using different batch sizes, is depicted in Figure 27. Since the training accuracy and loss for all batch sizes were comparable, only the respective testing accuracy and loss are presented in this figure. Therefore, the left plot of Figure 27 shows the testing loss for each batch size while the right graph depicts the corresponding testing accuracy.

As demonstrated in Figure 27 (a), batch size 5 exhibits the highest loss of 0.22 compared with the other two batch sizes of 10 and 16, which yielded a testing loss of 0.20 and 0.18, respectively. Inspecting Figure 27 (b), which displays the testing accuracy, it is evident that for all batch sizes the testing accuracy reached a peak of almost 96%. Summarising, up to this point, the combination of 50 epochs and a batch

size of 16 generated the highest testing accuracy while it produced the lowest testing loss. At the same time though, the validation accuracy is the most important parameter which governs the method's performance and mirrors how well the learning process fared during the training and testing processes.

Figure 28: Some examples of the successful recognition of images by the more sophisticated BM image classifier.

The performance of the more sophisticated BM image classifier was evaluated on the validation set, which contained images that the classifier never encountered or trained at before. As it can be observed in Figure 28, the BM image classifier was able to discern the different categories of images illustrating marine debris and marine life and tag each image with its respective class label. Interestingly, the classifier successfully assigned the correct label to an image even though some of the images contained other distractions, such as the chain shown in the "Net" snapshot (Figure 28).

Figure 29 illustrates the validation accuracy of the BM for varying batch sizes, spanning from 5 to 10 and 16 images. Here, the validation set contained 400 newly provided image samples from the eight object categories. Observing Figure 29, it is apparent that increasing the number of images allocated in each batch size, the method yields a slightly lower validation accuracy. Of these, batch size 5 produced

the highest validation accuracy of 90% or 360 correct classifications out of 400 validation samples. Batch size 10 and 16 generated a slightly lower validation accuracy of 88.7% and 88.2%, respectively. As the trustworthiness of the classifier is better served by the validation accuracy, the combination of batch size 5 and 50 epochs was finally selected for the more sophisticated BM image classifier. Concluding, the particular classifier attained a training accuracy of 98% with a training loss of 0.13, a testing accuracy of 96% and a testing loss of 0.22 accompanied by a validation accuracy of 90%.

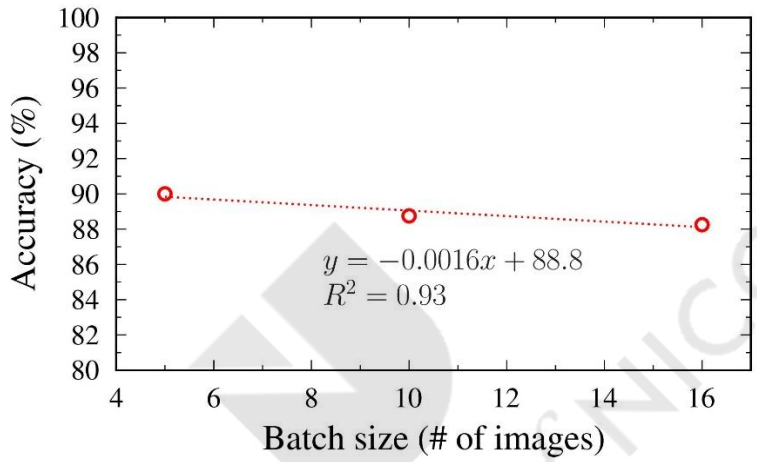


Figure 29: Validation accuracy for varying batch sizes, spanning from 5 to 10 and 16 images.

The validation accuracy of ≈90% realised by the more sophisticated BM image classifier is comparable to the findings from other research studies which utilised the same DL model for recognising debris items in the marine environment. For example, Fallati *et al.* (2019) utilised the VGG16 model to construct an image classifier responsible for categorising UAV images into five general litter classes, which realised an accuracy of 94%. Additionally, Musić *et al.* (2020) employed the particular model to classify images illustrating general litter from five distinct categories, like cardboard, glass, paper, metal, and plastic. Their proposed image classifier attained an overall classification accuracy of 85% (Musić *et al.*, 2020).

Comparing now the classification performance of ≈90% of the proposed BM image classifier with applications of the same VGG16 model in other domains, the

BM attained similar validation accuracy. For instance, the study of Xu *et al.* (2018) utilised the VGG16 model for categorising images illustrating metals with an accuracy of ≈80%. Additionally, Liu *et al.* (2020b) used the specific scheme to create an image classifier capable of recognising kiwi fruits with a validation accuracy of 88.4%. Consequently, the similarity between the classification results obtained by the proposed BM image classifier and other applications of the same DL tool renders the BM trustworthy.

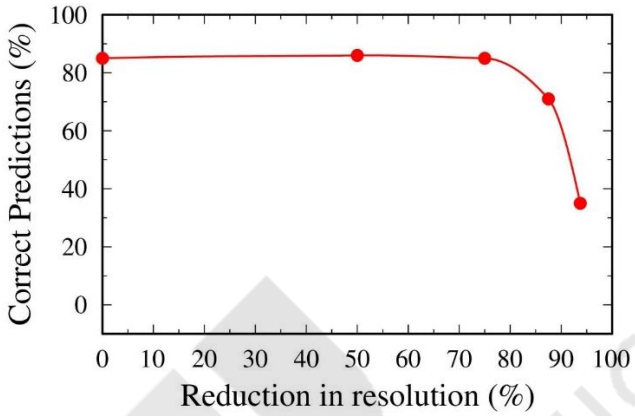


Figure 30: Correct image predictions as a function of image resolution.

Another idea that aimed to assess the identification capabilities of the more sophisticated BM image classifier on plastic litter and sea life dealt with the resolution of the validation images. Image resolution is an important parameter since a high-resolution image is characterised by richer information with presumably good quality making it easier for the BM image classifier to successfully recognise the plastic debris images. To inspect how sensitive and effective the particular image classifier is to resolution, it was decided to vary the resolution of the images in a systematic manner. In this regard, the resolution (pixels per inch) of the validation images was gradually reduced by 50%, 75%, 87.5%, and 93.7% of their original size. Subsequently, the truncated size images were processed by the BM image classifier probing in this way the evolution of its competence in discerning the presence of plastics or marine life from the same images.

Results presented in Figure 30 reveal the total correct classifications of validation

images as a function of image resolution. Strikingly, the BM image classifier retained the same high performance in distinguishing between the different object categories despite a drastic reduction in the resolution of the validation images. In other words, the BM image classifier is capable of identifying marine debris and marine life even from very poor resolution images. It is only when the resolution of images was scaled down more than 3/4 (75%) of their original size that the ability of the classifier to correctly recognise images of marine debris deteriorated dramatically. Because these findings were obtained from images which exhibit much higher average resolution (125 pixels per inch) compared with the reduced size images which suffered a 75% reduction in quality (≈31 pixels per inch), the results of the more sophisticated BM image classifier presented herein appear credible.

Even though the more sophisticated BM image classifier is competent in identifying images from eight categories with high validation accuracy of 90%, the scheme has some drawbacks. Firstly, the particular classifier can categorise still images and tag only one class label per image even though multiple plastic litter items were visible in the image. Consequently, the non-detected litter items within the images were omitted. Potentially the preceding development, led to an overestimation of the classification accuracy or undercounting of the detected litter items. Therefore, it was impossible to know the total number of plastic debris encountered in these images.

Secondly, and most importantly, the BM cannot classify objects illustrated in videos. Identifying objects in video footage is a critical attribute of a marine debris classifier as it can process valuable details from larger littered areas. Hence, an image classifier tagging a single class label on still images is not sufficient for real world applications like the marine debris classification task. For this purpose, more advanced DL tools, like the YOLO series and the YOLACT++ algorithm, were employed here.

4.3. Detection and localisation of plastic litter in images and videos

The research findings from the application of the YOLOv3, YOLOv4, YOLOv5 and YOLACT++ algorithms to the marine debris classification task are presented in this part. The abovementioned DL tools can detect and localise multiple plastic litter and marine life appearing in images and video footage. Simply put, such object detectors can attach a class label on each plastic debris item found in an image or video footage while localising it by applying a rectangular bounding box around it.

Figure 31: Sample of images featuring plastic marine debris.

To facilitate this detection task, another dataset of seven object categories was constructed which predominantly aimed to improve the detection accuracy of the proposed plastic debris detectors. Some of the images of this dataset were obtained from the ImageNet (Deng *et al.*, 2009) open source database, while others were provided by the Algalita non-profit organisation (Algalita, 2014b). Images illustrate plastic debris from six categories and one type of marine life, that comprise, (1) bottles, (2) bags, (3) buckets, (4) fishing nets, (5) straws, (6) food wrappings, and (7) a fish species. In total, 1,620 images made-up this dataset and examples of these images are shown in Figure 31. DA manipulations were applied to the images of the new dataset which relate to changes in the image pixel intensity. More precisely,

during the training stage of the YOLO series and YOLACT++ tools, the saturation, the brightness and the colour of the images was randomly altered.

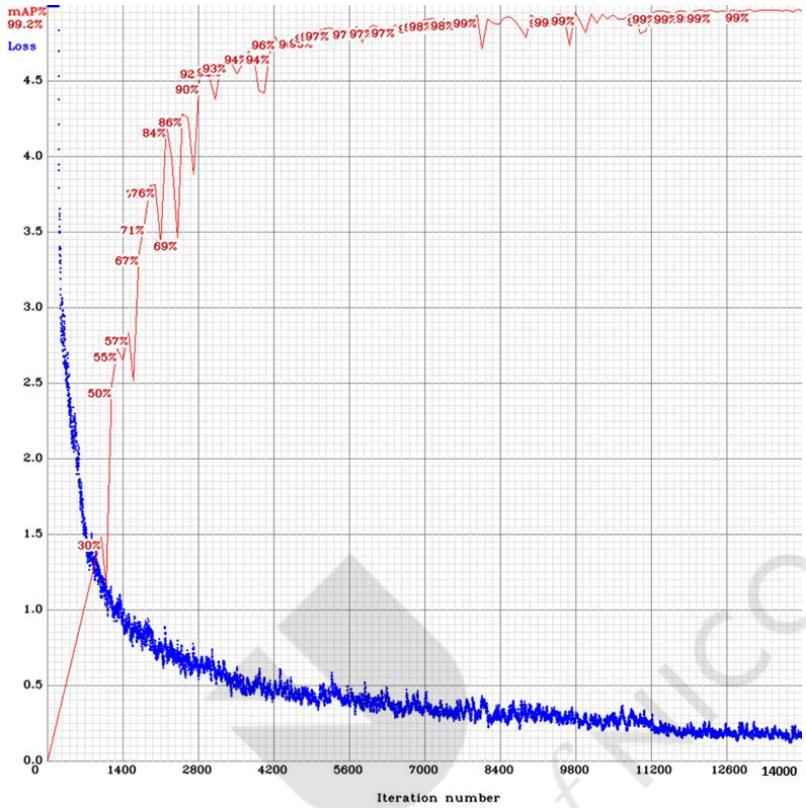


Figure 32: Training loss (blue curve) and testing mAP (red line) of the YOLOv3 image classifier.

The training and testing processes of the YOLOv3 and YOLOv4 object detectors were performed on an Intel® Xeon® (CPU 2.40 GHz) processor with an NVIDIA Quadro® K4200 graphics card. In contrast with the rest YOLO series, the YOLOv5 was trained on the NVIDIA® Tesla® K80 GPU provided by the Google Colab platform. For the cases of YOLOv3 and YOLOv4, the batch size was set to 64 images with subdivisions of 16 images while the training and testing processes were left to run for 14,000 iterations. The training process of YOLOv5 spanned across 300 epochs with a batch size of 10. In all of the cases, the value of momentum was set to 0.9 while the decay was 5×10^{-4}. Here, the momentum is a coefficient that is applied to an extra term in the weights update so as to speed-up the learning process and to help the network find a more important global minimum. Whatsmore, the decay refers to a

regularisation technique that attaches a small penalty to the loss function so as to keep the weights small and prevent overfitting.

While training the YOLOv3 model on the custom dataset, the default values of the model's hyperparameters have been utilised. That is, the IoU loss gain was set to 1.2, the value of the *x* and *y* coordinates loss gain was assigned to 4.062, while the width and height loss gain was set to 0.1845. Furthermore, the classification loss gain was fixed to 15.7, while an initial value of 3.67 of the classification binary cross-entropy (BCE) loss was adopted. Additionally, the object loss gain was adjusted to 20, the object BCE loss was set to 1.36, the IoU training threshold assumed a value of 0.194, while the initial learning rate was fixed to 0.00128. Moreover, a final lambda learning rate value of –4 was used, the stochastic gradient descent (SGD) momentum was 0.95, and the optimiser weight decay was 2.01×10^{-4}. The image saturation value was set to 0.8, while the image Hue Saturation Value (HSV) was set to 0.388. Lastly, the value of image rotation was 1.2°, the value for the image translation was kept to 0.119, the image scale was 0.0589, and the shear range of image was set to 0.401.

Similarly, during the training phase of the YOLOv4 model, the default values of the model's hyperparameters were used. Namely, the initial learning rate was set to 0.01, the final one-cycle learning rate to 0.1, the warmup epochs spanned to 3, the warmup momentum was adjusted to 0.8, the warmup bias learning rate was fixed to 0.1, while the box loss gain remained at 0.05. Additionally, a classification loss gain of 0.3 was used, the classification and object BCE loss were kept to 1, the object loss gain was 0.6, the IoU training threshold was 0.2, and the anchor-multiple threshold was set to 4. Moreover, the focal loss gamma remained at 0, while the image HSV-hue augmentation was set to 0.015. Also, the image HSV-saturation augmentation value was set to 0.7, the image HSV-value augmentation to 0.4, the image rotation was maintained at 0°, the image translation value was set to 0.1, the change in image scale was 0.9, while the values of shear change, image perspective, flip up-down, and mix-up were set to zero. Lastly, the value of image flip left-right was set to 0.5 while the image mosaic value was 1.

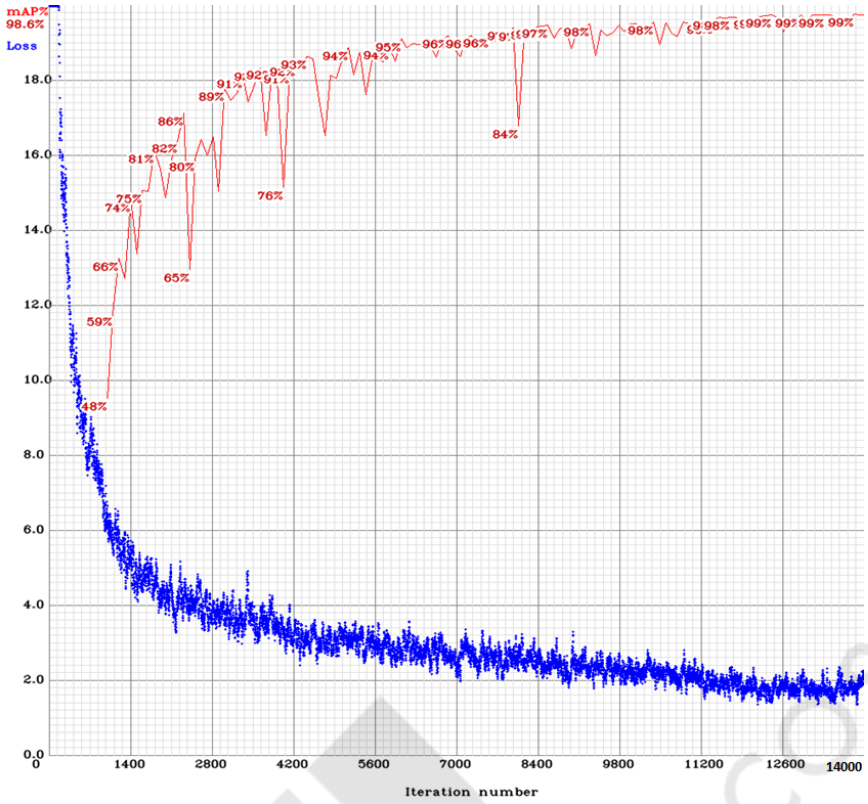


Figure 33: Training loss (blue curve) and testing mAP (red line) of the YOLOv4 image classifier.

During the training phase of the YOLOv5 algorithm, the default values of the model's hyperparameters have also been maintained. Specifically, the initial learning rate was assigned a value of 0.01, the final one-cycle learning rate was set to 0.2, the warmup epochs to 3, the warmup momentum to 0.8, the warmup bias learning rate was set to 0.1, while the box loss gain assumed a magnitude of 0.05. Additionally, the classification loss gain was set to 0.5, the classification and object BCE loss to 1, the object loss gain was again 1, the IoU training threshold to 0.2, and the anchor-multiple threshold was adjusted to 4. Moreover, the focal loss gamma was set to 0, while the image HSV-hue augmentation was assigned a value of 0.015. Other hyperparameters referred to the image HSV-saturation augmentation which was set to 0.7, the image HSV-value augmentation which assumed a value of 0.4, the image rotation which was fixed to 0°, the image translation was set to 0.1, the change in image scale was kept at 0.5, while the values of shear change, image perspective, flip

up-down, and mix-up were set to zero. Concluding, the value of image flip left-right was set to 0.5 and the image mosaic value was maintained at 1.

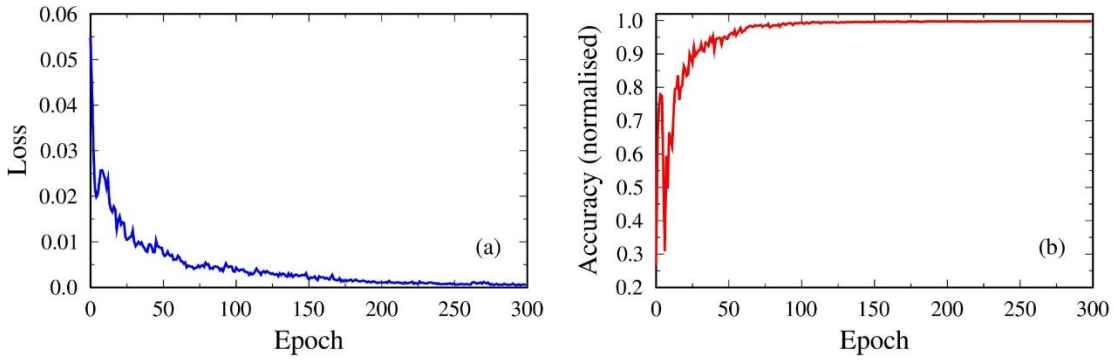


Figure 34: The training loss (left) and the testing mAP (right) attained during the training and testing processes of the YOLOv5 object detector.

Figure 32 and Figure 33 depict the plots of the training loss (blue line) and the testing mean average precision (mAP) (red curve) of the YOLOv3 and YOLOv4 object detectors, respectively, when trained and tested on the marine debris dataset. The testing mAP of YOLOv3 topped to 99.24%, at a training loss of 0.18. Analytically, the training loss of the YOLOv3 model was calculated using the sum of squared error loss and the binary cross-entropy (BCE) loss (Redmon *et al.*, 2018). Similarly, the YOLOv4 object detector realised a testing mAP of 98.62%, at a training loss of 2.12. However, the training loss of the YOLOv4 tool was calculated using the complete intersection over union (CIoU) loss for bounding box coordinate predictions and the BCE loss for class predictions (Bochkovskiy *et al.*, 2020).

Figure 34 illustrates the training and testing results of the YOLOv5 object detector when applied to the marine debris and marine life image dataset. Graph (a) of Figure 34 portrays the training loss of the YOLOv5 object detector, which precipitated to 0.001, while graph (b) refers to its testing mAP which amounted to 99.5%. The total training loss was calculated using the generalised IoU (GIoU) loss for bounding box predictions and BCE with logits loss for class predictions (Rezatofighi *et al.*, 2019).

Hereafter, a series of steps were adopted to calculate the mAP. Initially, the predicted bounding box applied by the object detection tools on each discovered

object was compared to its respective ground truth bounding box assigned manually during the annotation process. For all YOLO models, the YOLO_mark annotation tool was utilised, while for the YOLACT++ algorithm, the Visual Geometry Group (VGG) Image Annotator (VIA) (Dutta *et al.*, 2019), was employed. The comparison was facilitated by using the IoU expressed by equation (13). Predictions with IoU values smaller than 0.5 were discarded as they imply a low level of confidence. Thereafter, the average precision of each object class can be retrieved using the area under the precision-recall curve for that class. Collectively, the mean value of all average precisions from all object classes yielded the mAP (Padilla *et al.*, 2020). Expressed as ratios, the precision and recall for each object class were calculated from:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{16}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{17}$$

where TP refers to True Positives, that is, the total number of objects correctly labelled by the model as belonging to the target class (Padilla *et al.*, 2020). The False Positives (FP) denote the total number of items incorrectly labelled by the model as belonging to the target class. Lastly, the False Negatives (FN) express the total number of articles that the classifier incorrectly labelled as assigned into the false class and yet being part of the target class. Specifically, the precision (equation (16)) quantifies the proportion of the predicted positives that were truly positive while the recall (equation (17)) is a measure of the proportion of actual positives which were classified correctly.

Part of the effort to evaluate the identification capabilities of all DL-based marine debris detectors included the use of a total of 350 newly provided validation images of plastic debris and marine life from the seven object classes. Note that the seven object categories included plastic bottles, plastic bags, plastic buckets, fishing nets, plastic straws, food wrappings, and a fish species. This evaluation produced a validation

mAP of 59.42% for YOLOv3, 60.49% for YOLOv4 and 92.4% for the YOLOv5 tool. Interestingly, the YOLOv5 realised an increase in the validation mAP of ≈33% from still images compared to the rest YOLO series.

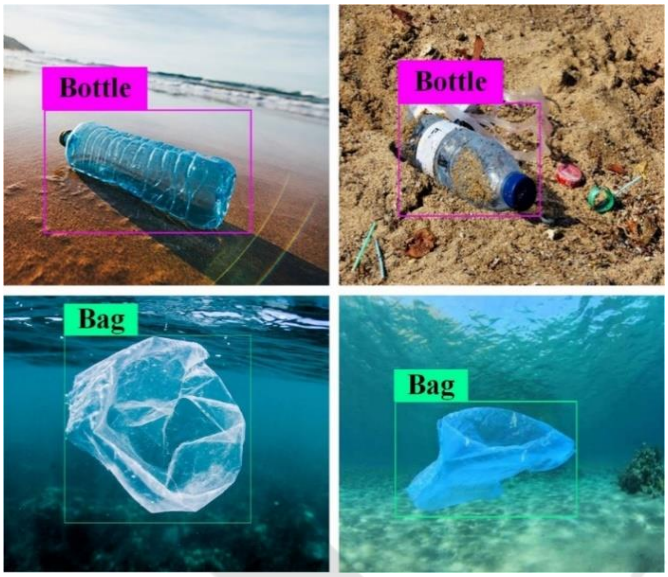


Figure 35: Examples of correctly recognised plastic debris items obtained from the YOLOv3 object detector.

Figure 35, Figure 36 and Figure 37 illustrate some correctly recognised plastic debris items and marine life obtained from the YOLOv3, the YOLOv4 and the YOLOv5 object detectors, respectively. Carefully observing these figures, it is evident that all of the YOLO series attached a class label and a bounding box on each detected litter item or sea life. Remarkably, the YOLOv5 marine debris detector, in contrast with the rest YOLOs, attached the confidence score to each discovered item, as illustrated in Figure 37. Clearly, the confidence score next to each class label is an important indication as one can directly judge the quality and trustworthiness of each prediction simply by looking at this score.

Figure 36: Examples of correctly recognised plastic litter articles obtained from the YOLOv4 object detector.

When applied to still images, the YOLOv3 and YOLOv4 models produced a validation mAP of 59.42% and 60.49%, respectively, which compares well with the reported validation mAPs attained from other research studies which considered the same classification task. Namely, Watanabe *et al.* (2019) utilised the YOLOv3 tool to detect floating debris from four object categories which comprised plastic bottles, plastic bags, drifting wood and "other" items, with a mAP of 77.2%. Additionally, Tomas *et al.* (2022) used the YOLOv4 algorithm to identify floating plastic and paper debris in rivers, attaining a mAP of 63%. Comparing the performance of the YOLOv3 and the YOLOv4 tools with applications from other domains, like the detection of apples (Kuznetsova *et al.*, 2020), traffic signs (Dewi *et al.*, 2021) or fig fruits (Yijing *et al.*, 2021), both methods realised almost a 30% lower mAP. However, the current research presented herein expanded the scope of the marine debris detection research by discerning litter items from more than one category.

Figure 37: Correctly classified and localised plastic debris items and fish using the YOLOv5 object detector.

Referring to the validation mAP of 92.4% that the YOLOv5 object detector attained over the newly provided still images, this value is slightly higher than the mAP obtained from existing applications of the same tool on the same classification task. For instance, Córdova *et al.* (2022) and Veerasingam *et al.* (2022) utilised the YOLOv5 algorithm for detecting litter items from up to six general classes belonging to various natural settings with a mAP of 85% and 90%, respectively. Comparing now the performance of the particular scheme with existing applications from other domains, like the detection of safety helmets (Zhou *et al.*, 2021) or breast tumour (Mohiyuddin *et al.*, 2022), the proposed YOLOv5 realised a very similar detection accuracy. The fact that the proposed YOLOv5 marine debris detector achieved analogous mAP with prevailing applications on comparable or different items, renders the method reliable.

Moreover, referring to the detection of objects in the video content, the YOLOv3 and the YOLOv4 object detectors achieved comparable validation mAPs of 42.2% and 42.47%, at a detection rate of almost 8 frames per second (FPS). The YOLOv5 object detector realised a validation mAP of 74.5% at a detection rate of 34 FPS. Suffice to mention that the video footage tested here was recorded at a beach in

Paralimni, Cyprus, and illustrates objects from the previous seven categories. Remarkably, the YOLOv5 marine debris detector achieved an increase of almost 32% in the mAP from the video content compared to the earlier YOLOv3 and YOLOv4 models. Lastly, the validation mAP of 74.5% that the YOLOv5 realised on the video footage is higher by ≈7% than reported figures from other research studies, which utilised the same DL tool to process underwater video footage (Chin *et al.*, 2022).

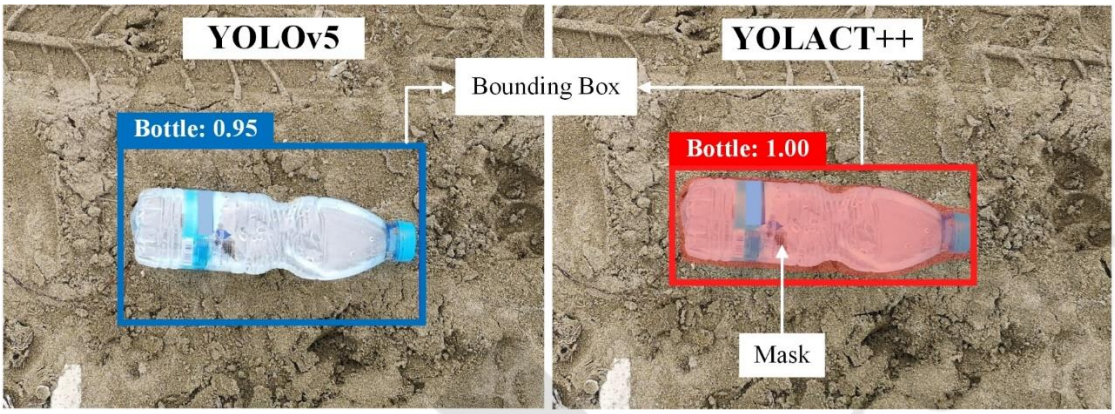


Figure 38: Comparison of the performance of the YOLOv5 and the YOLACT++ object detectors.

Moreover, the detection rate of 34 FPS that the YOLOv5 object detector attained over the video footage signifies that this scheme, in contrast with the YOLOv3 and the YOLOv4 detectors, performs real-time object detections on videos. Taken together, when considering both real-time identifications on videos and the superb classification performance of 92.4% on validation images, the YOLOv5 algorithm proved, up to this point, the most powerful and accurate DL algorithm for constructing the detector to be used in recognising and localising plastic debris and sea life in the marine environment.

Observing the sample of images presented in Figure 37, it is reasonable to wonder if the applied bounding box on each identified object may embrace some unwanted features from the background and, consequently, lower the detection accuracy of the YOLOv5 tool. For that reason, the attention has shifted to the YOLACT++ algorithm, which considers, beyond the class label and the bounding box, the area of each detected object. As illustrated in Figure 38, the YOLACT++ tool attached a mask on

each identified item that traces its external geometry. Spurred by the need of enhancing the capabilities of the proposed marine debris detector while improving its identification performance on both still images and video recordings, the YOLACT++ scheme was applied to the marine debris classification task (Kylili *et al.*, 2021).

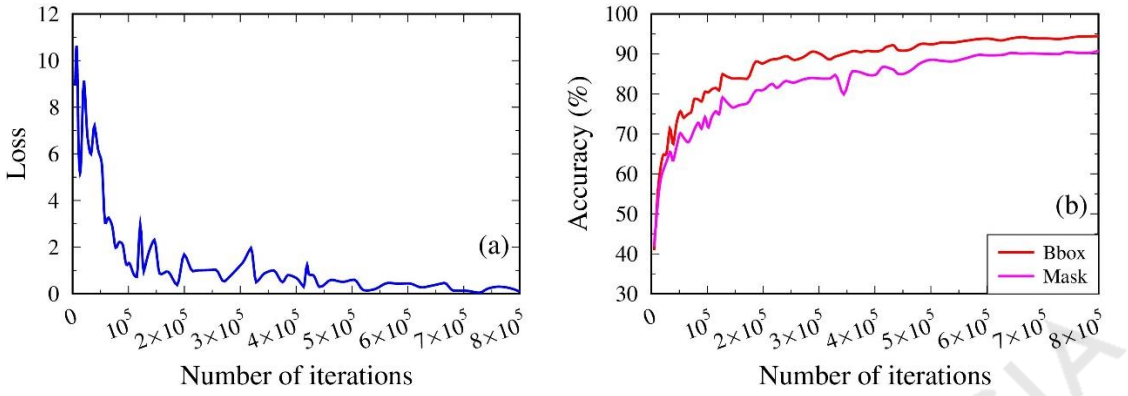


Figure 39: The total training loss (left) and testing mAP (right) of the YOLACT++ object detector for bounding box (bbox) predictions (red line) and mask predictions (magenta curve).

Similar to YOLOv5, the training and testing processes of the YOLACT++ object detector were executed on the NVIDIA® Tesla® K80 graphical processing unit (GPU) provided by the Google Colab platform. The default values of the hyperparameters that have been used during the training process of the YOLACT++ algorithm are the following: the initial learning rate was set to 0.001, the momentum to 0.9, the decay to 5×10^{-4}, the gamma value to 0.1, the learning rate warmup assigned a value of 1×10^{-4}, while the mask alpha was set to 6.125. Moreover, the positive IoU threshold was set to 0.5, the negative IoU threshold was 0.4, while the crowd IoU threshold was 0.7. Additionally, the initial focal loss was set to 0.01, the focal loss alpha value to 0.25, the focal loss gamma value to 2, while the value of the photometric distortion of the image and its cropping parameter were randomly altered. Finally, the value of image mirroring was fixed to 0.5, whereas the parameters that relate to flipping and rotation of the image were set to zero.

Figure 40: Illustrations of correctly predicted plastic detritus and fish using the YOLACT++ object detector.

The training and testing process of the YOLACT++ tool were performed during 800,000 iterations and data augmentation manipulations, based on the single shot multibox detector (Liu *et al.*, 2016), were applied to the images during the training stage. The single shot multibox detector adjustments scale down the image, haphazardly crop a part of the snapshot and place it in a random place in the image while mirroring, flipping and rotating the picture.

Results from the training and testing assessments of the YOLACT++ object detector are presented in Figure 39. Graph (a) (Figure 39) shows the total training loss of the YOLACT++ image classifier which converged at the value of 0.2. For the record, the total training loss was computed by a combination of three loss components: (1) the classification loss, L_{cls}, (2) the bounding box regression loss, L_{box}, and (3) the mask loss, L_{mask} (Bolya *et al.*, 2019; Bolya *et al.*, 2020), as follows:

$$Loss = L_{cls} + L_{box} + L_{mask} \tag{18}$$

where L_{cls} and L_{box} are defined in the same way as in Li *et al.* (2016). Loss L_{mask} refers to the pixel-wise binary cross entropy (BCE) between the assembled masks M and the

ground truth masks M_g. Finally, plot (b) of Figure 39 demonstrates the testing mAP for image predictions using bounding boxes (red line), which peaked at 94.45%. Whatsmore, the testing mAP for object detections using masks (magenta line) amounted to 90.81%.

Evaluating the YOLACT++ object detector on 350 new validation images illustrating plastic debris and marine life from the seven object categories, the method attained a validation mAP of 69.6% for bounding box predictions and 66.9% for mask predictions. Cases of correctly predicted plastic detritus are demonstrated in Figure 40, where the class label accompanied by its confidence score and the mask are attached to each detected plastic trash item or marine organism. Indeed, as depicted in Figure 40, the mask applied by the YOLACT++ object detector on each discerned item encloses only its actual area. Lastly, applying the proposed YOLACT++ object detector on the actual video footage recorded at a beach in Paralimni, Cyprus, the particular scheme realised a validation mAP of 75.6% and 75.7% for bounding box and mask predictions, respectively, at a frame rate of 14 FPS.

Comparing the performance of the proposed YOLACT++ detector on still images with existing applications of the same tool on the marine debris detection task, the particular scheme realised a slightly lower validation mask mAP of about 5% when compared with an underwater trash detector (Dias *et al.*, 2022). In contrast with applications of the same DL model on other research areas, like the segmentation of skin psoriasis (Lin *et al.*, 2021), the proposed YOLACT++ detector achieved a 20% lower validation mask mAP. However, the lower value of the validation mask mAP of the proposed object detector was expected since the particular detector can discern litter items from more than one category. Concluding, Table 6 summarises the comparison between the performance of the DL tools employed in this study and other research efforts that utilised the pertinent DL tools on the same or another detection task.

Concluding, comparative results of the performance of all DL-based marine debris detectors examined in this study, excluding the BM which does not perform multi-object detection in images and videos, appear on Table 7 and Table 8. Table 7 lists the

validation mAP obtained from the YOLOv3, the YOLOv4, the YOLOv5 and the YOLACT++ object detectors when validated on 350 still images depicting plastic litter and a fish. Table 8 presents their respective validation mAPs when applied to the same video recording. Results presented in both tables, indicate that YOLOv5 outperforms the other object detectors on discovering items from still images scoring a validation mAP of 92.4%, while the YOLACT++ tool is the best for detecting articles in video content. Although, an improved detection accuracy from a video renders the marine debris detector more reliable and useful simply because the video footage can scan large polluted areas and trace more items of plastic debris. Based on the previous justifications and the fact that the YOLACT++ tool applies a mask on each detected item, which can generate more valuable information about its physical dimensions, rendered the YOLACT++ the DL algorithm of choice.

Table 6: Summary of the comparison of the performance of the YOLO series and YOLACT++ tool with other studies referring to the same or a different domain.

DL tool	Accuracy of studies applied to the same domain (# classes)	Accuracy of studies applied to different domain (1 class)	Accuracy of this study (# classes)
BM (VGG16)	94% (5) (Fallati *et al.*, 2019), 85% (5) (Musić *et al.*, 2020)	80% (Xu *et al.*, 2018), 88.4% (Liu *et al.*, 2020b)	86% (3) 90% (8)
YOLOv3	77.2% (4) (Watanabe *et al.*, 2019), 88.3% (8) (Xue *et al.*, 2021)	94% (Liu *et al.*, 2020a), 90.8% (Kuznetsova *et al.*, 2020)	59.4% (7)
YOLOv4	63% (2) (Tomas *et al.*, 2022), 91% (3) (Tian *et al.*, 2022)	89.3% (Dewi *et al.*, 2021), 92.3% (Yijing *et al.*, 2021)	60.5% (7)
YOLOv5	85% (5) (Córdova *et al.*, 2022), 90% (7) (Veerasingam *et al.*, 2022)	94.7% (Zhou *et al.*, 2021), 90.5% (Mohiyuddin *et al.*, 2022)	92.4% (7) 89.4% (9)
YOLACT++	72.2% (4) (Dias *et al.*, 2022)	43.8% (Wen *et al.*, 2021), 86% (Lin *et al.*, 2021)	69.6% (7)

Table 7: Validation mAP realised from YOLOv3, YOLOv4, YOLOv5 and the YOLACT++ object detectors as applied to 350 still images.

Image classifier	Bounding box mAP (%)	Mask mAP (%)
YOLOv3	59.4	–
YOLOv4	60.5	–
YOLOv5	92.4	–
YOLACT++	69.6	66.9

Table 8: Validation mAP obtained from YOLOv3, YOLOv4, YOLOv5 and YOLACT++ object detectors as they pertain to the video footage.

Image classifier	Bounding box mAP (%)	Mask mAP (%)
YOLOv3	42.2	–
YOLOv4	42.5	–
YOLOv5	74.5	–
YOLACT++	75.6	75.7

Thereafter, the YOLACT++ algorithm was utilised to construct an intelligent tool that can detect, localise and segment multiple marine debris and marine life in still images and video recordings. More precisely, the proposed intelligent method can differentiate items among seven categories of objects, namely plastic bottles, plastic bags, plastic buckets, fishing nets, plastic straws, food wrappings, and a fish species. However, the most important outcome of the proposed intelligent method is the masked area applied on each discovered object. Interestingly, this mask can provide valuable details as they can be processed using image processing tools to determine the litter's physical dimensions. Results from the application of the intelligent method on real field data are presented in the following sub-section.

4.4. Physical dimensions and density of plastic litter

4.4.1. Estimation of the plastic litter density across the Cypriot coastlines

4.4.1.1. First beach survey

The intelligent method constructed using the YOLACT++ algorithm was tested on real field images which depict plastic debris. Ultimately, the proposed intelligent tool was used to estimate the plastic litter density across the Cypriot coastlines. Thus, six beaches, from two popular tourist destinations at the cities of Larnaca and Famagusta, Cyprus, were initially selected as monitoring sites for collecting images of plastic litter. All of the six beaches displayed certain characteristics. These comprised a minimum stretch of 30 m, were accessible to humans, and had clear access to the sea, meaning that the beach did not have breakwaters which can influence sea currents and waves and thus the abundance of marine debris. Each monitoring site was assigned a site code while coordinates were recorded on the global positioning system (GPS). Suffice to mention that this assessment only considered macroplastics larger than 5 cm (Kylili *et al.*, 2021).

Initially, the six beaches were scanned for the presence of marine debris during November 7th, 2020, which produced six data samples— one from each monitoring location. Data collection was conducted by one researcher while the same methodology was followed for all the testbed sites. Unusually polluted shorelines like urban beaches were omitted as daily clean-ups are performed there, while semi-urban and remote beaches were granted priority. At the first two beaches, in Larnaca, the survey covered a 150 m by 3 m transect. Due to the narrow topography of the third beach, in Larnaca, a coastline stretch spanning 100 m by 1 m was covered. In Famagusta, the first transect scrutinised an area measuring 120 m by 3 m, while the rest two scans stretched over two coasts which extended 30 m by 2 m. Hence, almost 600 m of Cypriot coastlines were surveyed during this investigation. However, due to limitations in resources, such as scientific equipment and a pressing time-frame, only a limited number of coastlines were examined during this phase.

Inspecting images captured from the six beaches, the intelligent method detected a total of 52 plastic objects belonging to four categories of plastic debris, namely, plastic bottles, bags, straws, and food wrappings. By analogy, the abovementioned technique predicted a plastic litter density of 0.035 items/m^2 which amounts to a total mass of 0.9 kg for the 52 plastic items. Plastic litter density here was obtained by averaging the number of correctly identified plastic detritus items over the scanned area. Similarly, the total mass of the litter was estimated by tallying the known mass of all predicted plastic items while considering the mass of individual objects.

Interestingly, the predicted litter density for Cyprus of 0.035 items/m^2 compares well with reported figures from neighbouring countries. For instance, the beach litter density in Greece's Ionian Sea was estimated to be at least 0.03 items/m^2 (Vlachogianni *et al.*, 2017), while neighbouring Turkish coasts were reported to contain no less than 0.085 items/m^2 (Topçu *et al.*, 2012). Considering that the estimated litter density for Cyprus is comparable with the findings from adjacent countries lends trustworthiness to the proposed intelligent method.

Utilising the preceding findings, it was possible to extrapolate the total count of marine debris items that litter the Cypriot coastlines which span about 735 km (European Commission, 2019). A methodology similar to Lavers *et al.* (2017) and Galgani *et al.* (2000) was adopted here, which performed linear extrapolation to estimate the abundance of plastic litter at large marine regions based on small-scale sampling sites. Results revealed that, in Cyprus, the aggregate number of plastic items which clutter the island's coastlines were approximately 66,000 articles weighing a total of about 1,000 kg.

At this point, it is helpful to mention that the images acquired from the six beaches illustrated plastic litter only from four categories of items, which comprised, plastic bags, plastic bottles, plastic straws and food wrappings. However, during this survey, it was not possible to spot and document using our camera any objects from the rest three categories, which are the plastic buckets, the fishing nets and the marine life. That was the reason that these categories were omitted from the calculation of the plastic litter density.

4.4.1.2. Second beach survey

Broadening the investigation while validating the abovementioned plastic litter density of 0.035 items/m^2, more actual field video recordings illustrating plastic debris from all seven object categories were acquired from different and larger beach-front stretches in Cyprus during September 19th, 2022. Again, due to restrictions in resources, equipment, and a confined time-frame, an additional six beaches from the cities of Famagusta, Larnaca, and Limassol, were examined as part of the new survey expedition. Like the earlier efforts, the new sites were chosen with the same selection criteria and methodology elaborated in the first investigation (refer to §4.4.1.1). During the new investigation, video recordings from two distinct beaches in Famagusta covered a seafront measuring a total of 360 m long by 2 m wide. In Larnaca, video footage from two different beaches scanned a seafront of a total of 500 m (long) by 2 m (wide), while in Limassol, video recordings from two distinct beaches scanned a total of 540 m length by 2 m wide coast transect. In aggregate, about 1,400 m of Cypriot coastlines were surveyed during the new survey.

Scrutinising the collected video footage acquired during the latest beach investigation, it was possible to extract individual video frames which encompassed plastic debris from all seven object categories, namely, plastic bags, bottles, buckets, straws, fishing nets, food wrappings and a fish species. The intelligent method, constructed using the YOLACT++ tool, identified a total of 146 plastic litter items belonging to the previous seven object categories. Averaging the number of correctly detected plastic debris items over the scanned area, which amounts to 2,800 m^2, the intelligent technique estimated a new plastic litter density of 0.052 items/m^2.

Even though the new plastic litter density of 0.052 items/m^2 of the second beach survey is slightly larger than the previous plastic litter density of 0.035 items/m^2 (refer to §4.4.1.1), yet results could be compared with some caution. Plainly, the difference of 0.017 items/m^2 between the two litter densities was expected as the latest survey covered longer beach stretches and considered all of the object categories in the estimation of the plastic litter density. Given that the two survey efforts were conducted at different locations and dates, some deviation of the density results was

expected in light of other parameters such as the beach morphology, wind speed and direction as well as the popularity of the beach.

Interestingly, the new plastic litter density of 0.052 items/m^2 compares well with reported figures from other countries, like Spain and Morocco (Asensio-Montesinos *et al.*, 2019; Nachite *et al.*, 2019). Specifically, the beach litter density in Spain was estimated to be at least 0.062 items/m^2 (Asensio-Montesinos *et al.*, 2019), while in Morocco the concentration of litter was projected to be at least 0.054 items/m^2 (Nachite *et al.*, 2019). Another beach litter collection effort led by Orthodoxou *et al.* (2022), which focused on Cyprus, retrieved litter from 11 classes and manually calculated the litter density to 0.15 items/m^2. Referring on the results of Orthodoxou *et al.* (2022) and considering litter items from only the same seven object categories that were studied in this thesis, their calculated plastic litter density approximates to 0.055 items/m^2. Clearly, this figure almost matches the predicted plastic litter density of 0.052 items/m^2 estimated during the new beach assessment elaborated here. Considering that the plastic litter density obtained from the manual beach collections of Orthodoxou *et al.* (2022) and the proposed DL-based object detector for the island of Cyprus have produced almost identical findings, lends trustworthiness and credibility to the suggested method. In other words, the results from Orthodoxou *et al.* (2022) serves as a partial validation to the findings outlined in this study.

4.4.2. Physical dimensions of plastic litter

Expanding the investigation, the proposed intelligent method, which is based on the YOLACT++ tool, deduced the physical dimensions of all detected plastic debris items found at the six beaches in Cyprus surveyed during the first beach investigation (refer to §4.4.1.1). At the heart of this step, was the application of the *OpenCV Contours* image processing tool described in sub-section A3 of the Appendix. The pertinent image processing tool estimated the length of each mask applied to individually segmented plastic litter. Initially, each mask was converted into a contour area. Subsequently, the tool connected together all continuous points, which exhibited

the same pixel intensity along the object's borderline, as part of a curve linking exercise.

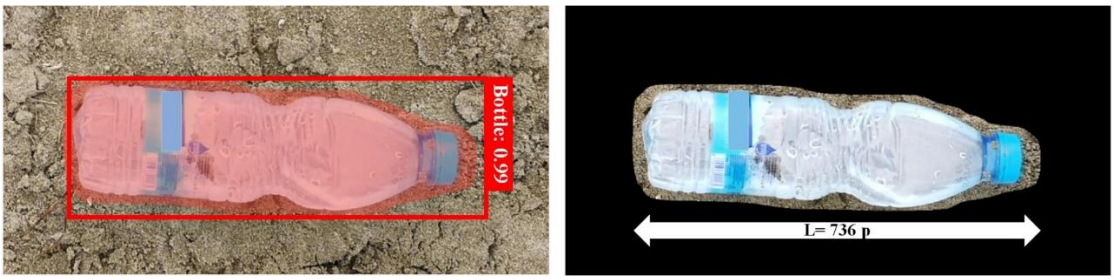


Figure 41: The *OpenCV Contours* image processing tool was used to compute the length (L), in pixels (p), of each detected plastic item.

Next, a scaling factor of 40 pixels to 1 cm was applied to compute the actual size of the 52 marine debris items. Serving as a point of reference, the length of each image window frame measured 150 cm long. Similarly, the length of each segmented debris piece, in pixels, was multiplied with the span of the wider image window frame, in cm. After normalising the estimated length of each item with the length of the image window frame, in pixels, the real length of the masked litter object was obtained in units of cm. Part of the previously mentioned process is illustrated in Figure 41, where the YOLACT++ intelligent method predicts the length of the plastic bottle in pixels (Kylili *et al.*, 2021).

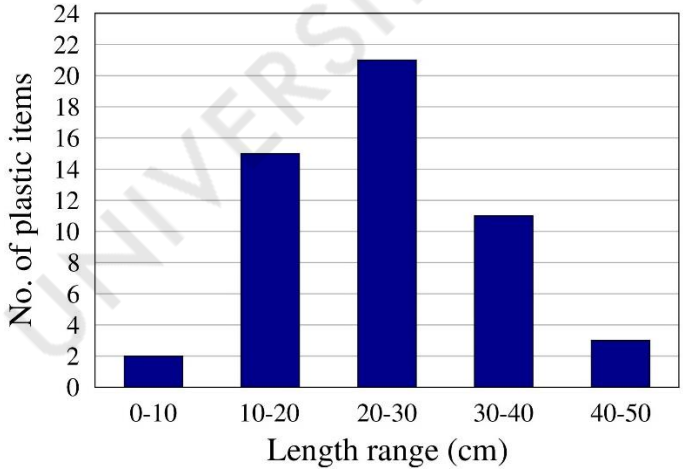


Figure 42: The predicted length spectrum for all detected plastic debris items found at the six beaches in Cyprus of the first coast investigation (refer to §4.4.1.1).

Figure 42 depicts the findings of the extrapolated length for all recognised plastic litter found at the first six beaches in Cyprus of the initial beach investigation (see §4.4.1.1) in categories ranging between 0–10 cm until 40–50 cm with an increment of 10 cm. Results demonstrated that the dominant length of plastic debris varied between 10 and 30 cm which compares well with the actual size of common plastic litter frequently spoiling these seashores. Serving as a validation, the preceding dimensions match the actual length of 0.5 L (500 mL) plastic bottles, medium-sized plastic bags, plastic straws and food wrappings mainly from chocolate bars and potato chips.

4.4.3. Degradation of a biodegradable plastic bottle

The degradation of a biodegradable plastic bottle was studied in this part, as a last investigation of the proposed intelligent method which is based on the YOLACT++ object detector. The method accepted as input an image of a biodegradable plastic bottle at three different stages of its decomposition. Again, the *OpenCV Contours* image processing tool was utilised to process the three images. Each masked object was converted into a contour area and the method calculated the item's surface area in pixels.

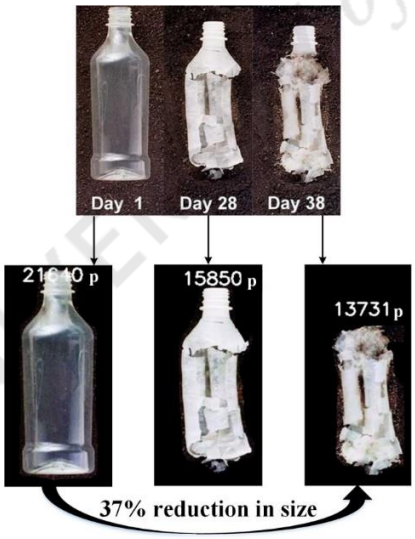


Figure 43: The degradation of a biodegradable plastic bottle. Image on top adapted from ECHO Instruments (2016).

As depicted in Figure 43, the segmented bottle on Day 1 possessed an area of 21,640 pixels. Subject to degradation, during Day 28, the bottle exhibited an area of 15,850 pixels and on Day 38 its area shrunk to 13,731 pixels. Comparing Day 1 and Day 38, in Figure 43, the decrease of 7,909 pixels in the area of the bottle translated into a reduction of almost 37% in the physical dimensions of the biodegradable bottle. Being able to project the extent of disintegration of biodegradable items is a very useful aspect of their use. Nevertheless, the calculation of the degradation using image processing techniques is possible only when the images feature good representations of the biodegradable items taken from the same prospect of camera view. Helping form more concrete conclusions pertaining to the degree of the degradation of plastics, the images illustrating the biodegradable items ought to possess the same size, orientation and be located at the same distance from the camera. Any deviations from the abovementioned conditions can affect the results with regards to the calculation of the actual litter's physical dimensions.

4.5. Estimating the abundance of plastic litter from video footage

Another line of investigation of this research dealt with the estimation of the abundance of plastic litter in video recordings. For this scope, the YOLOv5 image classifier was utilised at it performed real-time object detection of 34 FPS on actual field video footage. At the heart of this task were two counting methods, namely, the ROI line and the centroid tracking techniques. By imposing a virtual line on a video, the ROI line technique tallies the number of objects that cross that line. On the other hand, the centroid tracking method considers the centre of the bounding box of a detected item as its centroid and assigns to it an identity (ID). In this manner, the method tracks the identified bodies of interest as they move from frame to frame (Teng *et al.*, 2022).

Table 9: The number of marine debris items and marine life contained in the examined video footage.

Category	Bottles	Bags	Buckets	Nets	Straws	Wrappings	Fish	Total
Actual number	4	2	2	1	1 group	8	1	19

The same video recording used by the YOLO series and the YOLACT++ object detectors (refer to §4.3) was utilised here to assess the performance of the two counting methods. The actual number of marine litter articles and marine life present in the examined video footage is depicted in Table 9. Afterwards, a sequence of ROI line thicknesses was used to evaluate the performance of the ROI line counting method. Subsequently, the centroid tracking technique was utilised, whose *maxDisappeared* and the *thresholdFrames* parameters were adjusted to strengthen the method's counting capabilities.

4.5.1. Counting marine debris articles using the ROI approach

Initially, the thickness of the ROI line was altered from 5 to 16 pixels with an increment of 1 pixel each consecutive step. Once the predicted bounding box of a detected object passed the ROI line, the corresponding count of that object's respective category was augmented by one. Examining the findings of this assessment, as illustrated in Figure 44, it was found that a ROI line width of 13 pixels attained the best counting performance per object category compared to other line thicknesses. For this case, the counting technique discovered 4 plastic bottles, 1 plastic bag, 1 plastic bucket, 1 fishing net, 10 food wrappings, 1 cluster of plastic straws and no fish. It is worth noting that the "Actual" bar has been added for illustration purposes and depicts the actual number of items in the validation video.

In total, 18 items were successfully tallied by the ROI line method for a line width of 13 pixels, which equates into a counting accuracy of 95%. Although one may wonder why an ROI line width of 14 pixels was not selected as the preferred thickness of interest as supported by the unrealistically high counting accuracy of 105%. Careful inspection reveals that the performance of the algorithm that does not overcount the litter items in the validation video is preferable as it avoids double counting. Similarly, undercounting or omitting the plastic litter or fish is not equally desirable. Therefore, striking a balance between the two parameters is functionally sound. Based on the previous reasoning, ROI line widths of 15 and 16 pixels were

also discarded for this detection and counting task. Apparently, the miscounting of some items was attributed to the centre of the bounding box of these detected objects which may appear to the left of the line in one frame and to the right of the line in the next frame. Likewise, the bodies whose bounding box did not cross the ROI line were not counted by the method.

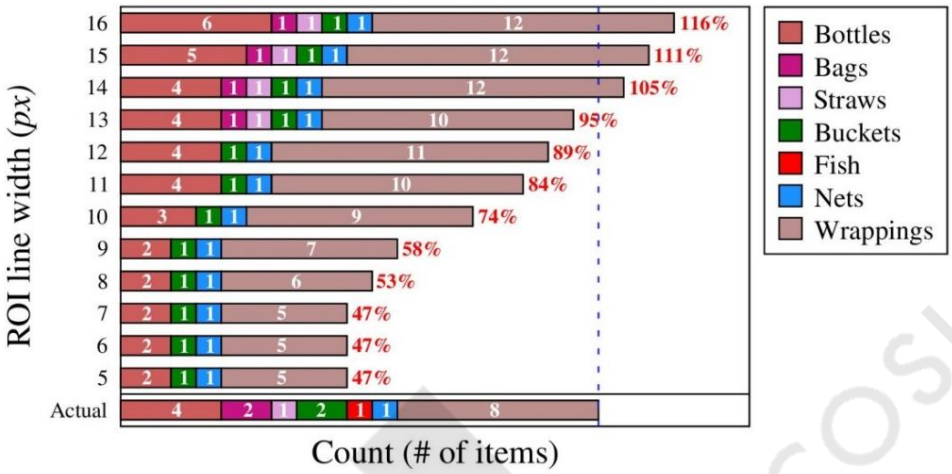


Figure 44: The performance of the ROI line counting method when varying the thickness of the ROI line from 5 to 16 pixels.

The performance of a counting algorithm relies heavily on the identification capabilities of the DL-based object detector. Apparently, the findings presented here cannot directly compare with what the object detector identifies and what the counting method eventually tallies. For example, in some cases the object detector assigned a wrong label to some objects, which produced an incorrect increment in the number of counted items for that class. At the same time, the manner with which videos are recorded fundamentally affects the performance of the object detection and counting technique. Clearly, high-resolution videos that were filmed slowly while offering adequate exposure to each object to appear on screen for an ample amount of time, were preferable. Not only does this boost the likelihood to correctly classify a given object but it can also diminish the probability for an object to be differentially classified between frames. As a result, the label that is most often associated with the object is most likely to be correctly counted over others.

4.5.2. Tallying debris items using the centroid tracking tool

Subsequently, the centroid tracking method was tapped for quantifying the number of marine debris and marine life which appeared in the video footage. Parameters *maxDisappeared* and *thresholdFrames* were altered accordingly in order to assess the counting ability of the centroid tracking technique. Figure 45 and Figure 46, respectively, present the findings of this evaluation. Inspecting Figure 45, the value of *maxDisappeared* was varied between 10 to 80 frames with an increment of 10 frames. As part of a parametric investigation, *maxDisappeared* realised its best counting precision per object category when it dealt with 60 frames during which it identified 6 plastic bottles, 1 plastic bag, 2 groups of plastic straws, 3 plastic buckets, 2 fishing nets, 4 food wrappings and no fish, totalling 18 items, which translates to a counting accuracy of 95%. Presumably, because the centroid of the bounding box of a new detected item was tied to a previously tracked object, the centroid tracking method miscounted some objects. Nevertheless, the centroid tracking method achieved comparable counting precision when applied to the same video as the ROI line counting technique.

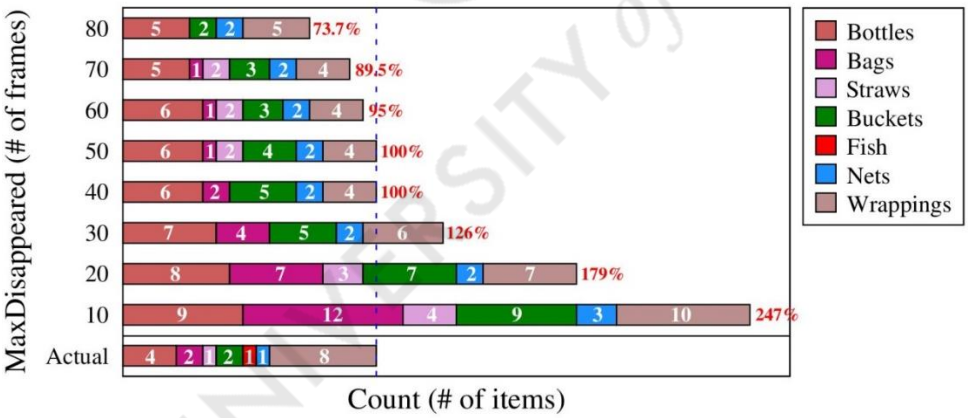


Figure 45: The performance of the centroid tracking counting technique while altering the value of *maxDisappeared* from 10 to 80 frames.

At a first glance a value of *maxDisappeared* of 40 and 50 number of frames fares better than other number of frames. Yet, the previous observation cannot be considered in isolation. Practically, parameter *maxDisappeared* requires a sufficient

number of frames to deregister an object captured in video footage. If deregistration occurs prematurely, that is, for a value of *maxDisappeared* smaller than 50 frames, the proposed counting method overestimates (double counts) the number of objects presented in the video recording. On the other hand, as displayed in Figure 45, when the value of *maxDisappeared* is larger than 60 frames, e.g., 70 and 80 frames, the centroid tracking technique deregisters objects belatedly and this causes the method to undercount the litter items. Consequently, the value of *maxDisappeared* of 60 frames yielded the best possible outcome when inspecting individual item classes and this was the major reason for selecting it.

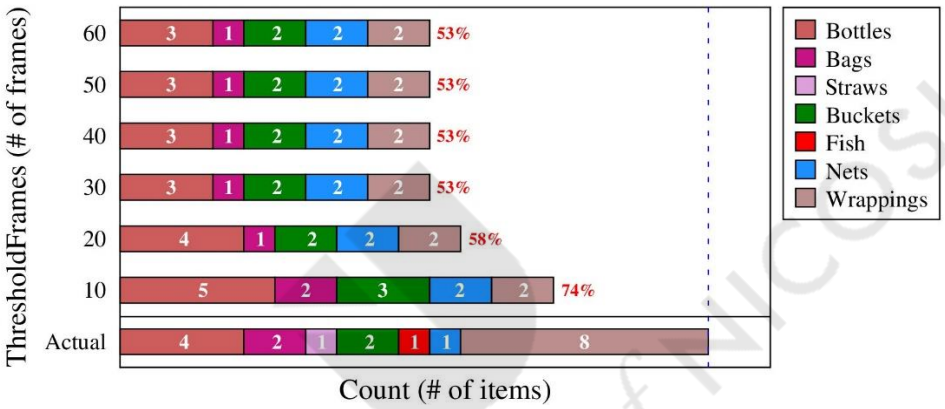


Figure 46: The performance of the centroid tracking counting technique when adjusting the number of *thresholdFrames* from 10 to 60 frames.

Similar to the *maxDisappeared*, the magnitude of the customised *thresholdFrames* parameter was extended from 10 to 60 frames at a step of 10 frames. As demonstrated in Figure 46, which illustrates the results of this evaluation, a *thresholdFrames* value between 10 and 20 frames generated the best counting accuracy per litter class. Herein, a *thresholdFrames* value of 20 frames was chosen with the method counting 4 plastic bottles, 1 plastic bag, 2 plastic buckets, 2 fishing nets, 2 food wrappings, no group of plastic straws and no fish. These results translate into an overall counting accuracy of 58%. Clearly, the low score the method achieved here is not necessarily representative of the actual counting performance of the approach because it is amenable to both the quality of the raw video and the identification capabilities of the YOLOv5 object detector.

Observing Figure 46, one can notice that the case of 10 frames for the *thresholdFrames* parameter was not selected as the desirable value because of its counting performance per litter category. Clearly, in this case the proposed technique overcounted the litter items belonging to the "Bottles" and the "Buckets" classes. From the processing standpoint, undercounting detected objects in the video footage bears more weight than overcounting. In this context, the value of *thresholdFrames* parameter equals to 20 was chosen as it yielded the best counting performance. Additionally, as demonstrated in Figure 46, increasing the value of *thresholdFrames* beyond 30 frames does not affect the number of tallied objects as the accuracy of the proposed counting method remains unchanged.

Collectively, the findings from the utilisation of the ROI line and the centroid tracking methods attained comparable counting accuracy of 95% when screening out marine debris and marine life appearing in the same video recording. Owing to the more reliable results generated by the centroid tracking method, it proved more adept than the ROI line technique. And this thanks to the ability of the centroid tracking approach to trace the geometric centre of the bounding box of a detected object or organism. In particular, the ROI line approach performed counting during the temporal instance at which the item crossed the line, while the centroid tracking tracked an object over the entire time duration that the object appeared on screen. Apparently, for the ROI line counting method, erroneous counts can result if the centre of the bounding box of a detected object is present within the coordinates of the ROI line for multiple frames. Along with the previous observations, the accuracy of the ROI line counting method was also found to be dependent on the speed at which objects are screened in the raw video, leading to erroneous counting results when projected too fast. Justified on the earlier reasoning, the centroid tracking counting method was selected to be coupled with the YOLOv5-based object detector to estimate the number of marine litter and sea life illustrated in video recordings.

4.5.3. Expanding the number of object categories

In order to broaden the scope of the proposed counting method elaborated upon in the previous sub-section (refer to §4.5.2), which is based on the YOLOv5 object detector and the centroid tracking technique, an additional two object categories, namely, aluminium cans and cigarette butts, were introduced to the custom image dataset. Collectively, the new image dataset is comprised of nine categories of objects: (1) plastic bottles, (2) plastic buckets, (3) plastic bags, (4) fishing nets, (5) plastic straws, (6) food wrappings, (7) a fish species, (8) aluminium cans, and (9) cigarette butts. Images belonging to this dataset were mainly retrieved from ImageNet (Krizhevsky *et al.*, 2017) while others were donated by the Algalita organisation (Algalita, 2014b). Additionally, some of the images were shared by the Debris Tracker— an open data citizen scientist movement. In aggregate, the new image dataset contained 2,050 images from all nine categories of marine litter and sea life.

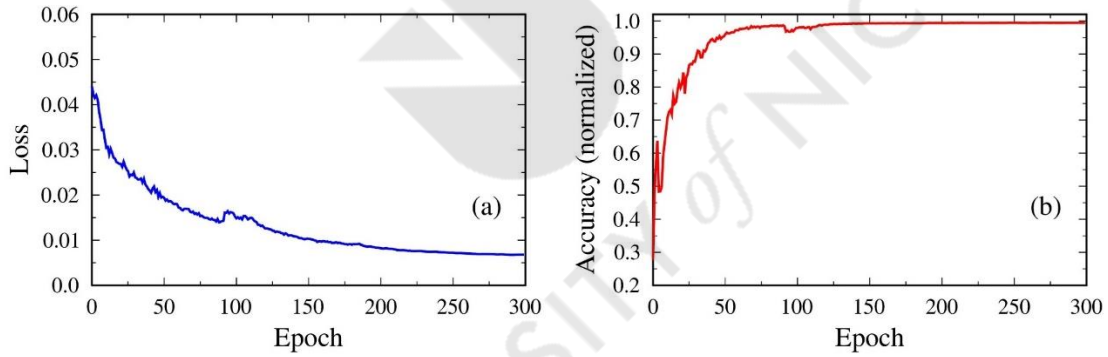


Figure 47: Graph (a) depicts the training loss while chart (b) presents the testing mAP of the YOLOv5 image classifier applied to the 9 categories of objects.

The YOLOv5 object detector was trained and tested on the new image dataset which contained the nine object categories. Followed then, 80% of the images were allocated to the training set while the remaining 20% of the images were assigned to the test set. Findings pertaining to the training and testing performance of the object detector on the nine categories of marine litter and a fish species are displayed in Figure 47. A testing mAP of 99.4% and a training loss of 0.007 were realised by the YOLOv5 object detector. In tandem with expanding the number of object categories,

one would expect a deterioration in the mAP of the object detector. But this did not happen in the preceding case because the mAP of the method remained at the same level of almost 99.5% when YOLOv5 trained on the seven categories (refer to §4.3). Evidently, these qualities characterise the YOLOv5 object detector as robust and reliable.

In the meantime, the performance and trustworthiness of the object detector are better served by the mAP that the detector attained when assessed on the validation set. Henceforth, the YOLOv5 object detector was validated on 450 freshly seen images belonging to the nine object categories. At this point, the detector scored a validation mAP of 89.4%. In particular, almost for all of the object classes the method realised enviable average precision of at least 88.3%, except for the "Wrappings" category. Exhibiting a precision of 56.3% the wrappings performed the poorest possibly due to the complicated shape or colour of food wrappings present in the images of the validation set.

Figure 48: Screenshots demonstrating the application of the YOLOv5 object detector and the centroid tracking method on an actual field video.

Thereafter, the YOLOv5 object detector, coupled with the centroid tracking counting technique, was applied to a new actual field video. Embedded in the new

video were items from all of the object categories except the fish species which was difficult to film in the marine environment. Moreover, the actual field video was created by a collage of short films capturing marine debris at a selected number of coastlines and natural settings. Collectively, the new video contained 7 plastic bottles, 3 plastic bags, 1 group of plastic straws, 4 plastic buckets, 1 fishing net, 11 food wrappings, 5 aluminium cans, and 11 cigarette butts, totalling 43 articles of marine debris. When the YOLOv5 object detector examined this video, the centroid tracking method successfully counted 4 plastic bottles, 2 plastic bags, 1 plastic bucket, 2 fishing nets, 10 food wrappings, 4 aluminium cans, 11 cigarette butts, and no cluster of plastic straws. In aggregate, 34 marine litter items were successfully tallied by the method, which translated into a counting accuracy of 79%.

Figure 48, which reflects the abovementioned counting capabilities, depicts four screenshots extracted from the actual field video when processed by the YOLOv5 object detector and the centroid tracking counting method. Carefully observing Figure 48, the proposed method assigned a bounding box, a class label and a confidence score on each detected marine litter item. In particular, each of the four images in this figure displays the item count per litter category until a specific frame as obtained from the centroid tracking method. These object counts appear on the left-hand side of each screenshot with green letters.

It is worth noting that the performance of a counting algorithm relies heavily on the identification capabilities of the YOLOv5 object detector. Apparently, the findings presented here cannot directly compare with what the detector detects and what the counting technique eventually counts. For example, in some cases the YOLOv5 object detector attached a wrong label to some objects, which produced an incorrect number of counted items for that class. At the same time, the manner with which videos are recorded fundamentally affects the performance of the object detection and counting technique. Clearly, high-resolution videos that were filmed slowly while offering adequate exposure for each object to appear on screen for an ample amount of time, were preferable. Not only does this boost the likelihood to correctly classify a given object but it can also diminish the probability for an object to be differentially

classified between frames. As a result, the label that is most often associated with the object is most likely to be correctly counted over others.

4.6. Detection of microplastics

In this part, the results emanating from the detection of microplastics using the ISA500 hydro-acoustic device are presented. Two experiments were conducted here to scrutinise the detection capabilities of this sensor when used to discover microplastics in the fresh water. The first experiment deployed the preceding sensor in a small water tank while the second installed the same sensor in a bigger water tank. In summary, the research findings from the two experiments are presented in the following sub-sections.

4.6.1. Experiment I

Starting, the first experiment was designed to test the detection capabilities of the ISA500 sensor in terms of discovering microplastics in fresh water. The ISA500 sensor was installed in a small tank filled with water, as demonstrated in Figure 49. The small tank's internal dimensions measured 45.5 cm, in length, by 63 cm wide, and 38 cm deep. During experiment I, the ISA500 sensor was hung by a metallic holder in a horizontal orientation under the surface of the water and the microplastics were suspended in front of the sensor.

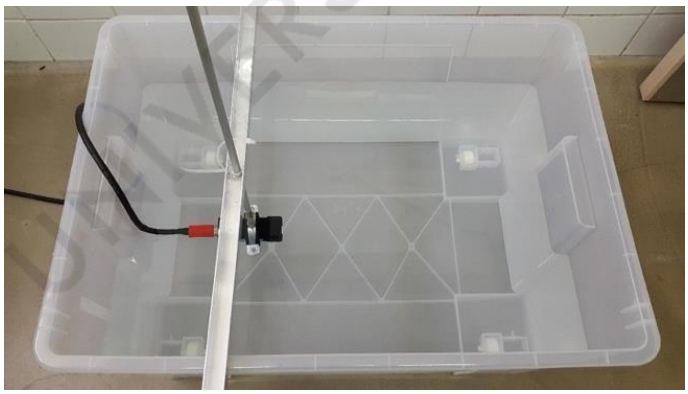


Figure 49: The small tank was filled with fresh water and the ISA500 sensor was immersed in the water.

Instrument measurements were enabled by the ISA500 sensor software which permitted the adjustment of some of its parameters. Table 10 presents the sensor parameters that were altered during experiment I. The frequency of the sensor was set to the highest possible value of 600 kHz, which is deemed suitable for tracing microplastics in the water. The length of the sound pulse was set to 2 μs which is possible at 100% of its amplitude. Note that the speed of sound in water is 1,482 m/s while the initial echo from the first obstacle was recorded by the ISA500 sensor. Discarding any false readings from the surroundings of the sensor, its detection range was set to be between 0.1 m and 0.4 m. Practically, this means the sensor was able to detect objects located at a maximum distance of 40 cm from itself.

Table 10: The parameters of the ISA500 sensor that were adjusted during experiment I.

Parameter	Value/Setting
Frequency	600 kHZ
Pulse length	2 μs
Pulse amplitude	100%
Speed of sound	1,482 m/s
Maximum range	0.4 m
Detection mode	First echo

All microplastic samples had to be prepared for the needs of experiment I. Initially, small pieces were cut from different plastic materials made of thermoplastic polyester (PET), polystyrene (PS), high-density polyethylene (HDPE) and polypropylene (PP). Later on, microplastics were hung from a thread, as illustrated in Figure 50 (a). The microplastics sizes measured 5×5 mm^2, 4×4 mm^2, 3×3 mm^2, 2×2 mm^2 and 1×1 mm^2 and in all cases their thickness was about 0.3 mm. Lastly, both the ISA500 sensor and the microplastics were submerged at about 10 cm under the surface of the water, while the suspended microplastics were positioned in front of the sensor, as displayed in Figure 50 (b).

The ISA500 sensor emitted a sound pulse that propagated underwater until it encountered the microplastic piece. Upon returning, the reflected echo was recorded by the ISA500 sensor which is able to deduce the range measurement which expresses the distance of the microplastic from the sensor's face. Along with range measurements, the sensor also registers the energy level and the correlation factor of the object's reflected echo. The energy level is the energy of the reflected echo and ranges between 0 to 1 where a value of 1 corresponds to full saturation of the ISA500 sensor. The correlation factor is a quality factor of the returned echo, spanning from 0 to 1. According to Table 11, a measurement with a high correlation factor and a low or a high energy level is probably a good reading. Since all the measurements attained very high correlation factors close to 1, only the energy level from the examined microplastics is presented and discussed hereafter.

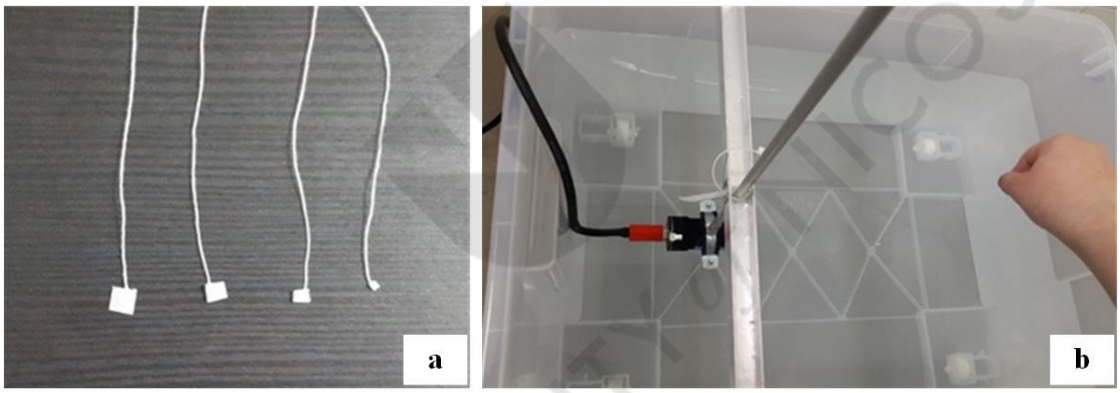


Figure 50: Picture a) shows the plastic pieces that were tagged to threads while image b) displays a microplastic suspended in front of the ISA500 sensor.

Table 11: The combination of energy level and correlation factor which result in trustworthy measurements for the ISA500 sensor.

	Low energy level	High energy level
Low correlation	Weak signal, probably false reading	High noise level, most likely a false reading
High correlation	Weak signal but likely a good reading	Ideal conditions, very trustworthy reading

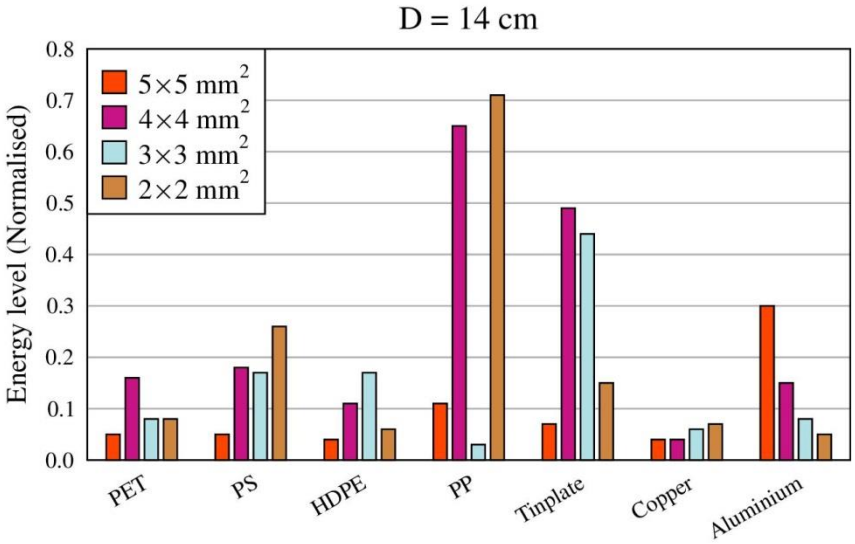


Figure 51: The energy level for all material sizes when the microplastics were placed at a distance (D) of 14 cm away from the ISA500 sensor.

Each microplastic was placed at a distance (D) of 14 cm, 23 cm and 30 cm from the sensor's face and the energy level of each microplastic item was recorded. Driven by the need to distinguish microplastics from other materials, the energy level from three metallic materials, such as Tinplate, Copper, and Aluminium, was also studied. Figure 51, Figure 52 and Figure 53 present the energy level of all materials at distances D = 14 cm, 23 cm and 30 cm from the ISA500 sensor's face, respectively. The orange colour in all figures corresponds to the object's size of 5×5 mm^2, the purple refers to a size of 4×4 mm^2, the light blue colour to a 3×3 mm^2, and the brown colour to a size of 2×2 mm^2. Energy level measurements for size 1×1 mm^2 do not appear in these figures, as the ISA500 sensor could not obtain any range measurement for this size. This happens because the wavelength of the sound pulse is very small and the quality of the reflected echo attenuates with an increasing distance between the object and the sensor, or the decreasing size of objects.

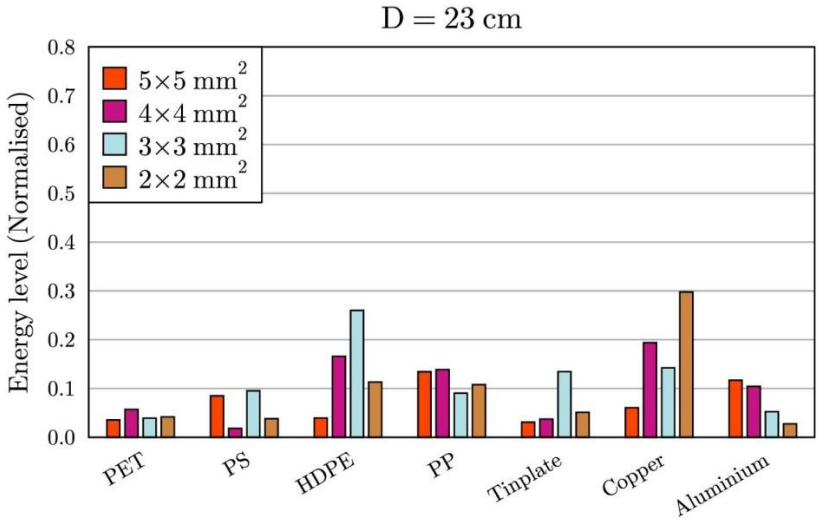


Figure 52: The energy level for all the material sizes situated 23 cm (D) away from the ISA500 sensor.

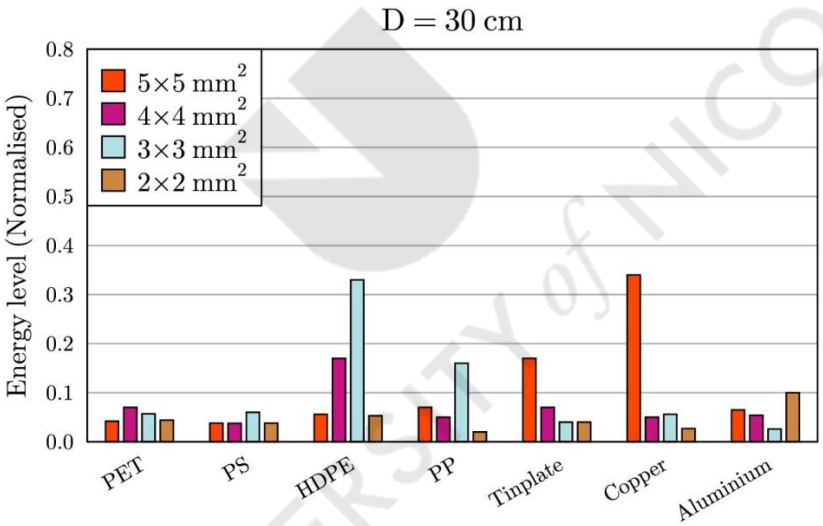


Figure 53: The energy level for all the material sizes located 30 cm (D) from the ISA500 hydro-acoustic device.

Inspecting Figure 51, Figure 52 and Figure 53, it was realised that for all the distances from the ISA500 sensor and for the largest size, that is 5×5 mm^2, some plastic materials display lower energy levels compared to certain metals. More precisely, microplastics from PET, PS and HDPE materials exhibited a lower energy signature than Tinplate and Aluminium. This trend was recurrent for smaller size

pieces and disappears at the smallest size of 2×2 mm^2 located at a distance of 30 cm from the ISA500 sensor's face.

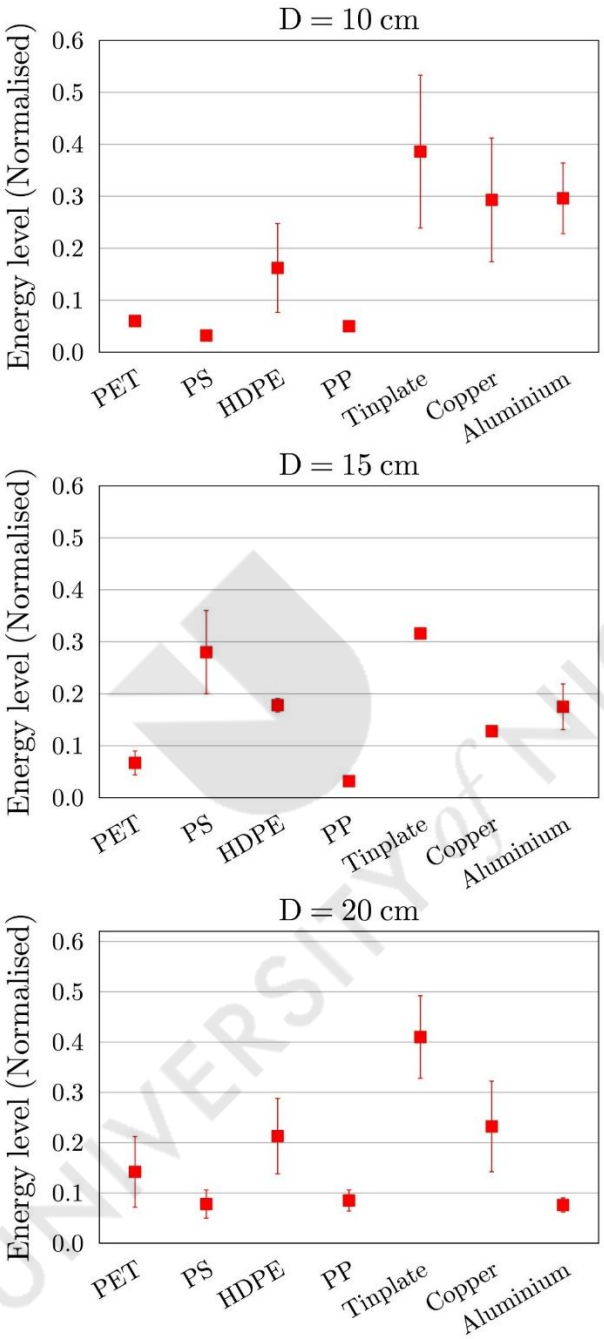


Figure 54: Energy levels and their statistical errors of all materials at distances D = 10, 15 and 20 cm from the ISA500 sensor's face. The area of all pieces measured 5×5 mm^2.

In view of some irregularities in the energy level of the reflected echo from the microplastics and the metallic pieces, it was decided that more measurements were needed in order to validate the detection capability of the IS500 sensor. For that reason, the material pieces were positioned at new distances D = 10 cm, D = 15 cm, and D = 20 cm from the ISA500 sensor's face, and several measurements for the same material were obtained. Due to time constraints, only the biggest size (5×5 mm^2) of material pieces was tested in the new assessments. Consequently, the average value of the energy level of their reflected echoes and their statistical error were calculated and presented in Figure 54.

Observing Figure 54, it is evident that microplastics tend to possess a lower energy level of their reflected echoes compared to the energy level of metallic materials. However, no clear pattern helps distinguish between microplastics and metallic materials using only the suggested ISA500 hydro-acoustic device. From this analysis, it is clear that the ultrasonic sensor alone is not a trustworthy device for differentiating microplastics from other materials or recognising the type of microplastic. Here it is important to mention that probably the best way for detecting and categorising microplastics into their distinct plastic types is using spectroscopic techniques like FT-IR analysis or Raman spectroscopy (refer to §2.2).

Summarising, the ISA500 hydro-acoustic device was demonstrated to be capable of detecting microplastics in the water. Still further tests were deemed necessary to tell if the ISA500 hydro-acoustic device can detect microplastics at longer distances. Hence, a new experiment was conducted in a bigger water tank and the experimental findings of this assessment are discussed in the following sub-section.

4.6.2. Experiment II

A more systematic investigation of the detection of microplastics in fresh water is outlined in Experiment II which utilises the ISA500 hydro-acoustic device and a water tank that is bigger than the first reservoir used in Experiment I (refer to §4.6.1). The big tank's internal dimensions measure 1.2 m, in length, by 1.2 m wide, and 1 m deep, as displayed in Figure 55. The tank was filled with fresh water and the ISA500

hydro-acoustic device was hung by a metallic holder, at a horizontal position, at the top right-hand corner of the tank. The sensor was submerged 30 cm below the water surface and the plastic and metallic pieces were placed in front of the sensor at longer distances (D) compared to experiment I. These locations included 10 cm, 15 cm, 30 cm, 45 cm, 60 cm and 90 cm.

During experiment II, microplastics and metallic pieces made of PET, PS, HDPE, PP, Tinplate, Copper, and Aluminium were tested, respectively. Moreover, the biggest size of each microplastic and metallic piece was used here, which measure 5×5 mm^2, and the energy level and its statistical error was recorded for each material piece. Figure 56 depicts the previous findings from all material pieces, located at distances D = 10 cm, 15 cm, 30 cm, 45 cm, 60 cm and 90 cm from the sensor's face.

Figure 55: Snapshot of the water tank and the ISA500 sensor used in experiment II.

Examining the individual graphs, at Figure 56, it was observed that the ISA500 sensor is still capable of discovering microplastics in the water at extended distances which reach up to ≈1 m. In particular, for the 10 and 15 cm distances, as illustrated in Figure 56 (a) and (b), the energy level scale of the items' reflected echoes is almost doubled compared to their respective energy level scale at longer distances. This is expected because at longer distances, the reflected sound pulse travelled over an extended length and a fraction of its energy attenuated in the water.

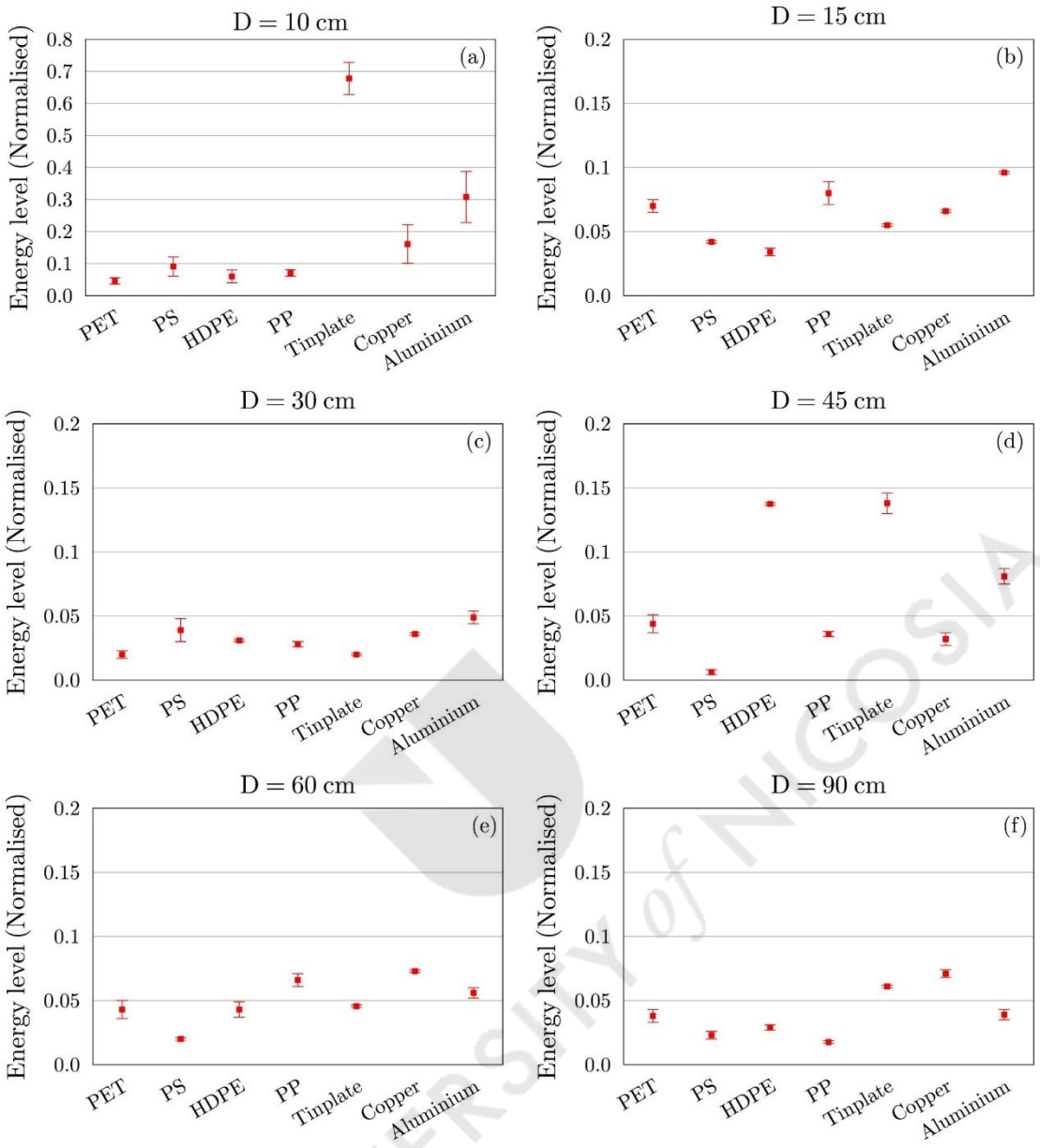


Figure 56: Recorded energy levels from all materials located at D = 10, 15, 30, 45, 60 and 90 cm from the ISA500 sensor's face.

Focusing on Figure 56 (a), which displays the energy level from all materials, at D = 10 cm, microplastics exhibit a slightly smaller energy level compared to metallic pieces but no clear pattern is observed for the rest of the distances. Although experiment II provided a more systematic framework for investigating the energy level of all materials, at various distances from the sensor, the information provided here is still insufficient. That is, measurements do not provide a conclusive trend that

could potentially differentiate microplastics from other materials let alone among them. Nevertheless, from this analysis it is apparent that the ISA500 sensor is not a suitable device to distinguish microplastics from other materials or characterise the type of plastic. This is because this sensor can primarily provide details about the proximity of the submerged microplastics together with the energy signatures of their reflected echoes.

To sum-up, the research findings from experiments I and II have revealed that the ISA500 hydro-acoustic device is not suitable for detecting and characterising microplastics in fresh water. Instead, analytical techniques are more suitable for identifying different types of microplastics or distinguishing them from other material items. Presently, some of the best techniques used to detect and characterise microplastics comprise Raman and FT-IR spectroscopy.

4.7. Results discussion

Initially, this research dealt with the image classification of three categories of plastic litter floating at the sea surface using the BM. Having parametrically tested the performance of the particular method, it was realised that 4,000 images in each object category after DA manipulations and 80% of these images to be used to construct the training set and the remaining 20% for the test set, provided the best classification accuracy of ≈86%. Additionally, the combination of the Ridge and the Lasso regressions methods, $\ell_1\_\ell_2$, proved to be the best regularizer for the specific classification task. Driven by the need to enhance the sophistication, the robustness, and the adaptability of the BM image classifier so as to categorise a diverse spectrum of plastic litter and sea life at the sea and the shorelines, another more elaborate BM image classifier was constructed.

The more rigorous BM image classifier incorporated improvements which relate to some parameters of the BM model and the number of object categories in the marine debris and sea life image dataset. The preceding improvements led to an increase in the overall classification accuracy of the BM. That is, expanding the number of epochs from 6 to 50, shrinking the batch size from 16 to 5 images, while increasing the number of object classes in the image dataset from 3 to 8, raised the classification accuracy of the BM by 4%, which topped ≈90%. Additionally, the quality of the images in the custom dataset was also improved by adding new high-resolution images that were carefully selected to display good representations of plastic litter encountered in the marine environment. Images in the enhanced dataset also depicted plastic litter floating at the sea surface or laying at the coastlines. Lastly, the ability of the BM to classify images from a variety of object classes supplemented also by a new class of unrelated objects, rendered the method more robust.

Next step involved the application of the YOLOv3, YOLOv4, and YOLOv5 object detection algorithms to the marine debris classification task. Among all of the examined YOLO tools, YOLOv5 was the most adept algorithm for detecting and localising multiple plastic litter and a fish species in images and video footage. Applied to newly seen images of plastic debris, the particular tool realised an

improvement of ≈33% in the validation mAP, which peaked at 92.4%. YOLOv5 also attained a 32% increase in the validation mAP which topped at 74.5% when applied to real field video footage illustrating items from the seven object categories. Of the three YOLO tools, the YOLOv5 object detector realised real-time object detection of 34 FPS when processing the video content. Owing to its performance, YOLOv5 is rightly the most powerful tool of the YOLO family as applied to the marine debris classification task.

The YOLACT++ instance segmentation tool was also used in this study to detect, localise and segment marine plastics and sea life in images and video content. Processing a real field marine debris video, the particular scheme realised the highest validation mAP of 75.6% among all examined DL tools. Equally useful, the YOLACT++ applied a mask on each detected object delineating its geometry. For this reason, the YOLACT++ tool was selected to construct an intelligent method tasked with identifying, localising and mapping the shape of plastic debris in the marine environment. Utilising real field images depicting plastic litter from six beaches, in Cyprus, the YOLACT++-based intelligent method detected a total of 52 objects belonging to four categories of plastic debris, namely, bottles, bags, straws and food wrappings. By analogy, the particular technique predicted a plastic litter density of 0.035 items/m^2 for the six beaches.

Expanding the investigation, an additional six beaches were surveyed in Cyprus during a second beach expedition while the collected video recordings were used to validate and broaden the aforementioned plastic litter density. During the second beach survey, the YOLACT++-based intelligent method traced a total of 146 plastic litter items from all seven object categories which consisted of plastic bags, bottles, buckets, straws, fishing nets, food wrappings and a fish species. Concurrently, the new plastic litter density estimated from the second beach investigation equated into 0.052 items/m^2. The deviation in the plastic litter densities of 0.035 items/m^2 and 0.052 items/m^2 from the first and the second beach surveys, respectively, was expected. This is because individual surveys were conducted at different years and more litter classes were considered in the calculation of the second plastic litter

density. Nonetheless, both plastic litter densities compare well with reported figures from other countries (Topçu *et al.*, 2012; Vlachogianni *et al.*, 2017; Asensio-Montesinos *et al.*, 2019; Nachite *et al.*, 2019) (refer to §4.4.1.1 and §4.4.1.2). The compatibility of the results from this research work and other studies done in other parts of the world, render the YOLACT++-based intelligent method capable of being applied in other geographical regions too.

Moreover, two counting techniques, namely, the ROI line and the centroid tracking, were coupled with the YOLOv5 object detector to identify and tally plastic debris items and a fish species contained in actual field video. Even though both counting methods attained comparable counting accuracy, the centroid tracking method proved more competent than the ROI line technique because it tracked the detected item over the entire time duration that the object appeared on screen. Fairly, the centroid tracking counting method, coupled with the YOLOv5 object detector, was selected for estimating the abundance of litter items from video footage. When the proposed counting technique processed a new real field video recorded at various marine and natural settings in Cyprus and the US, it successfully detected and tallied 34 plastic litter items from nine object categories, yielding a counting accuracy of ≈80%.

Suffice to mention that the research findings of this study depend on the detection capabilities of the DL tool, which was trained, tested and validated on the custom marine debris image dataset. Reasonably, DL-based counting method was expected to estimate a lower plastic litter abundance compared to what it was manually obtained by humans during coastline collection expeditions. For instance, manual beach surveys at the Jaz and Blatna beaches, at Montenegro, found 4,227 items per 100 m of coastline and 3,831 per 100 m of beachfront belonging to 11 categories of litter, respectively (Mandić *et al.*, 2021). Furthermore, at the municipal beach of Ensenada in Baja California, Mexico, manual debris collection efforts retrieved a total of 16,474 objects belonging to 10 categories of litter (Silva-Iñiguez *et al.*, 2003). Even though, the findings presented in this research work derive from relatively short shore

transects yet the findings appear particularly promising compared to typical manual beach surveys.

Other research investigations which implemented the centroid tracking counting method have assumed that all objects presented in video footage are of the same category (Wang *et al.*, 2006; Zhang *et al.*, 2009; Watts *et al.*, 2013; Khachatryan, 2019; Rosebrock, 2021a; Rosebrock, 2021b). As a result, the concept of tracking bounding boxes of the detected objects while obtaining the total count of items was a simple task. Remarkably, this study has adapted the centroid tracking counting method to perform multi-class object counting. Practically, this means that the proposed DL-based counting method can tally up to nine categories of marine debris and sea life, rendering it a versatile and useful detection and counting method.

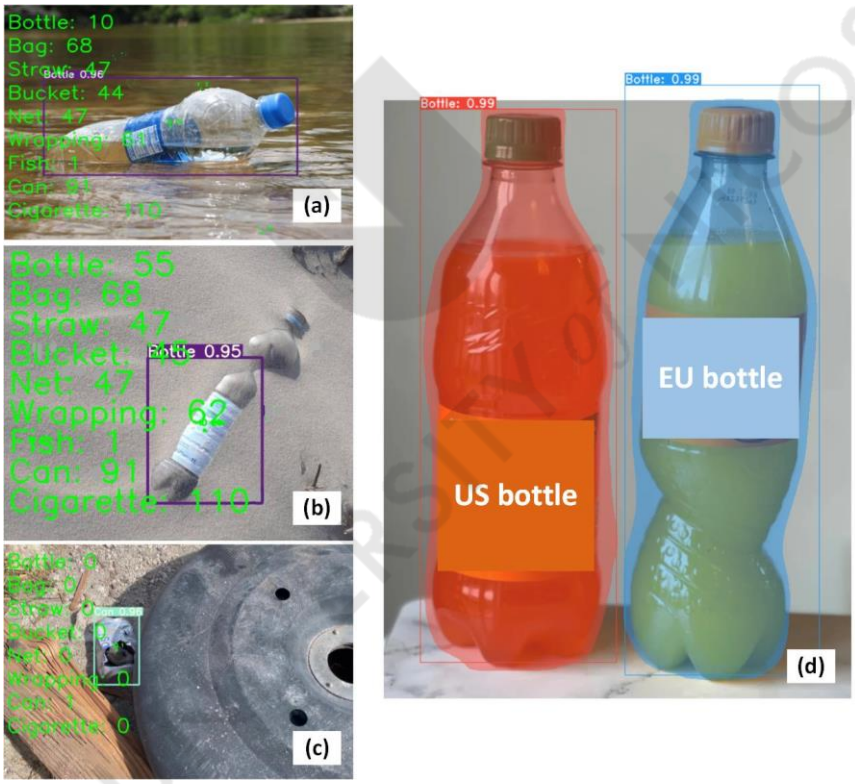


Figure 57: Successful identifications of plastic marine litter using the proposed DL-based object detectors.

The proposed detection and counting method, which is based on the YOLOv5 and the centroid tracking counting technique, can also obviate the use of manual post-

processing (Gonçalves *et al.*, 2020; Andriolo *et al.*, 2021) of collected marine debris images and video footage. Post-processing entails the manual classification and counting of presented litter items. Utilising the YOLOv5 and the centroid tracking counting technique and a few video recordings, the estimation of the abundance of debris items at the seas and the shorelines can be promptly and efficiently calculated without any human intervention. Lastly, the research presented herein can also be used for improving estimations of the distribution of marine debris in remote geographical regions that are not easily accessible by humans.

Undoubtedly, the proposed DL-based marine debris detectors presented in this thesis improve upon other DL techniques (Papakonstantinou *et al.*, 2021; Song *et al.*, 2021) owing to their ability to distinguish among plastic litter emanating from the different plastic sub-categories, like, bottles, bags, buckets, straws, food wrappings, etc. To the best of our knowledge, this research outlined the first comprehensive attempt to-date for smartly detecting and counting marine debris and sea life captured in video footage. Yet, other studies have mostly relied on manual image screening after collecting images or video recordings for estimating the abundance of plastic debris.

To sum-up, as demonstrated in images (a), (b), and (c) of Figure 57, the marine object detectors proposed in this thesis are capable of discerning and counting plastic litter that are floating on or near the sea surface, that are half-buried in the sand, or surrounded by other non-marine debris items, like dark or white rocks, green or brown vegetation and seaweed. Additionally, the pertinent detectors are capable of recognising plastic litter from various geographic regions, as depicted in Figure 57 (d). Referring to Figure 57 (d), two similar plastic beverage bottles manufactured in two distinct continents, such as the European Union (EU) and the United States (US), were successfully identified by the DL-based object detectors. Evidently, the abovementioned qualities rank the proposed DL-based marine debris detectors as robust, reliable and versatile.

In different parts of the world, depending on the type of marine debris, such as beached and floating, researchers utilise various methods for image collection and a

variety of ML or DL tools. When estimating the abundance of beached debris in the Seto Inland at Sea of Japan, the research team of Takaya *et al.* (2022), employed a drone and a high resolution camera to collect images of debris at the coast. Followed then, the collected images were fed into the RetinaNet object detector for estimating the beach litter density in the surveyed region (Takaya *et al.*, 2022). Additionally, at the Leirosa Beach located on the North Atlantic coast of Portugal, the team of Gonçalves *et al.* (2022), using a drone, collected airborne images depicting marine litter which was categorised using the Spectral Angle Mapping (SAM) technique. Finally, the researchers produced an image-based litter density map (Gonçalves *et al.*, 2022).

Focussing on the Great Pacific ocean gyre, located between Hawaii and California, US, the research team of de Vries *et al.* (2021) deployed a research vessel to collect images of debris floating at the sea surface and performed object detection using the DL tools of Faster R-CNN and YOLOv5. Based on the detected items in the processed images, the team calculated the numerical concentration of the identified marine litter (de Vries *et al.*, 2021). Lastly, surveying various coastal areas in the Philippines, India, England, Hong Kong and Argentina, and the Gulf of Honduras, the team of Booth *et al.* (2023) utilised multi-spectral satellite data and U-net-based ML models to classify floating litter and produce litter density maps. Collectively, all of the abovementioned studies post-process the collected image data to estimate the marine litter density or to produce debris density maps, process that is often time demanding. Contrarily, the abundance of plastic litter obtained in this research investigation derived from real field video footage which requires no post-processing and is accomplished in real-time.

4.8. Limitations and challenges

One of the major difficulties that this research encountered is associated with the amount of collected images and video footage. This can be ascribed to the fact that the acquisition of debris images often requires the utilisation of expensive equipment, like marine research vessels for scanning the sea surface, are time demanding and entail

substantial number of human operators or volunteers to cover long beach stretches. Thus, to alleviate the problem of small and less representative image datasets, DA manipulations were implemented to the image sets which enlarged the pool of training samples and improved the overall accuracy of the marine debris classifiers and detectors. Additionally, to increase the number of samples in the marine debris image dataset, pictures and filmed videos were also taken at some coasts and onshore in Cyprus, California and Hawaii.

Even though few non-profit organisations make their marine debris image datasets publicly available (Deng *et al.*, 2009; Algalita, 2014a; National Oceanic and Atmospheric Administration (NOAA), 2018), these datasets do not always have practical usability in specific automated marine litter technologies. For example, images in the pertinent datasets often illustrate marine debris under certain circumstances, like one litter category per image or photographed under special indoor scenarios or other controlled setups. Nevertheless, such image datasets are not typical of the litter items actually found in real-world aquatic settings.

Another limitation that affected the performance of the marine debris classifiers and detectors is the representation of litter items in the image dataset. Beached debris items, for instance, were often fouled by natural "litter like" green vegetation, wood pieces, grass, and seashells. Under other circumstances, a debris item was often half-buried in the sand or degraded or found among rocks of various colours and shapes. Such factors made the recognition of marine debris even more difficult. Additionally, the image resolution, the shape and the colour of marine litter items frequently affected the predictability of the DL-based object detectors. At such, transparent marine debris, e.g., fishing nets or plastic bags, or litter items with very similar colour to the marine background were less distinguishable by the DL detectors than other items with brighter or unique colours.

CHAPTER 5 CONCLUSIONS

5. Conclusions

5.1. Concluding remarks

Driven by the need to help tackle the problem of plastic pollution in the marine environment, we have applied various machine learning (ML) and deep learning (DL) techniques on the detection of plastic litter encountered at the seas and the shorelines. Starting from conventional ML tools from the Computer Vision domain and concluding with more sophisticated DL tools from the Artificial Intelligence realm, several image classifiers and object detectors able to classify, detect and count plastic marine litter and a fish species in still images and video footage were constructed. Besides the intelligent recognition of plastic marine debris, the research findings of the detection of microplastics in the fresh water using a single hydro-acoustic device, namely, the ISA500 sensor, are also presented in this study. Collectively, the main contributions of this research emanate from:

- Broadening the number of object classes that the BM image classifier can recognise from 3 to 8, the classifier achieved an improvement of 4% in its validation accuracy, which reached 90%. Downsizing the resolution of the validation images by 75%, the classification performance of the BM remained unaltered.
- Applying the YOLACT++ intelligent method on images depicting plastic litter from 6 beaches in Cyprus, the approach deduced a plastic litter density of 0.035 items/m^2. Extrapolated to the entire shorelines of the island, the intelligent method estimated ≈66,000 plastic articles weighing a total of ≈1,000 kg.
- Inferring the physical dimensions of all documented plastic litter with the aid of the *OpenCV Contours* image processing tool and the YOLACT++ intelligent method, results revealed that the dominant object length ranged between 10 and 30 cm.
- Supporting the previous plastic litter density of 0.035 items/m^2, a new beach survey from longer beach stretches and more debris categories produced a plastic litter density of 0.052 items/m^2.

- Achieving real-time object detection of 34 frames per second (FPS) on video footage and realising a mean average precision (mAP) of 92.4% on still images, the YOLOv5 object detector is suitable for scanning large littered areas in less time.
- Coupling the centroid tracking counting technique with the YOLOv5 detector, the approach attained a counting accuracy of ≈80% when validated on an actual field video illustrating objects from 9 categories.

Concluding, this research has spearheaded the application of DL tools to the smart detection and quantification of plastic debris in the marine environment. Part of the solution to the problem of plastic pollution hinges in the need to intelligently count and monitor litter over large areas of interest within a short period of time. Ultimately, it is envisioned that the research findings of this study will comprise a leap forward towards mapping with more accuracy the abundance of plastic litter in the marine environment while helping humanity formulate more efficient strategies in tackling them.

5.2. Research implications and advantages

The research findings of this doctoral thesis indicate that the most appropriate algorithm for counting marine debris illustrated in images and videos is the YOLOv5 tool coupled with the centroid trucking technique. This joint specific counting technique can count up to nine types of litter, namely plastic bottles, plastic buckets, plastic bags, plastic straws, a fish species, food wrappings, fishing nets, aluminium cans and cigarette butts attaining an accuracy of ≈80%. The proposed counting scheme can be used to improve estimations of the abundance of plastic litter in the marine environment. Particularly, using an UAV, like a drone, and a high-definition camera, airborne images or video footage illustrating plastic litter can be collected and processed by the proposed counting technique. In this way, real-time estimations of the number of plastic debris littering the surveyed areas can be offered.

Understanding the real-time distribution of plastic debris in the marine environment can help improve marine litter monitoring assessments across the world.

Particularly, the competent authorities of each littered region can prioritise these areas in terms of conducting in-situ clean ups. Removing plastic debris promptly and efficiently from the marine environment can help prevent the dispersion of litter in other geographical regions thought the action of wind and water currents. Finally, policy makers, Governments and international organisation, like the UN, around the world can take advantage of the real-time information provided by the proposed counting technique so as to make informed decisions and take mitigation measures to protect communities from the detrimental effects of plastic pollution.

Additionally, improved estimates of plastic litter density around the world can be offered using the YOLOv5 object detector and centroid tracking counting technique. Littered areas can be monitored more frequently since the proposed counting technique requires no human intervention, is based on the detection capabilities of the YOLOv5 tool and uses only images and videos to calculate the number of plastic marine debris. In this way, more accurate and up-to-date reports about plastic pollution can be provided for each region.

Insights tied to the plastic litter density and the litter's physical dimensions can be also gained from the YOLACT++ marine debris detector. Identifying the size distribution of plastic litter in the marine environment will help understand the scale and distribution of marine detritus at the seas and the shorelines, and how the dimensions of marine debris can impact marine life and aquatic life. Last, but, not least, smart tools can provide useful information about the different types and dimensions of plastic litter that pollute the marine environment. Grounded on the previous justification, more precise risk controls as a part of a general risk management approach can be designed by the national authorities of Governments of each country so as to eliminate the effects of plastic pollution on marine flora and fauna.

5.3. Future research directions

Several lines of investigation can broaden this research, but the following topics seem to be the most promising ones: 1) increase the number of object classes that a

marine debris detector can recognise, 2) add geospatial information to the detection process of marine litter, 3) improve estimations of the distribution of plastic debris in the marine and natural environment, and 4) further examine the detection of microplastics in the water.

The sophistication of the proposed marine debris detectors can be enhanced substantially by broadening the number of item classes that the detectors can recognise. This can be done by enriching the training and testing datasets with high quality images and videos, allowing the detectors at the same time to better assimilate the attributes of images featuring plastic representations. Additionally, integrating a marine debris detector and a GPS tracker on a device which is mounted on a boat, the device will be able to detect plastic litter and tag them with their exact location. In this way, new insights as to the distribution of plastics across the seas and the shorelines can be retrieved.

Moreover, an improved estimation of the abundance of plastic debris in the natural environment can be achieved through the contributions of citizen scientists. That is, civilians and researchers can photograph littered areas and use these images as input to the DL-based object detector. Consequently, the detector is expected to yield a more accurate and representative plastic litter abundance.

Finally, as a part of the MaRITeC-X project at the Cyprus Marine and Maritime Institute (CMMI), experiments will be conducted for the detection of microplastics in the seawater. By employing a high-quality spectrometer, a powerful computer and the ISA500 hydro-acoustic device or other sensors, it will help further investigate the dispersion of microplastics in the marine environment.

References

Acuña-Ruz, T., *et al.* 2018. Anthropogenic marine debris over beaches: Spectral characterization for remote sensing applications. Remote Sens. Environ., 217: 309-322, https://doi.org/10.1016/j.rse.2018.08.008.

Algalita 2014a. Algalita videos and images acquired during marine expeditions of the ORV Alguita vessel.

Algalita 2014b. North Pacific expedition in 2014. Ed. by K. Allen.

Andrady, A. 2011. Microplastics in the marine environment. Mar. Pollut. Bull., 62: 1596-1605, https://doi.org/10.1016/j.marpolbul.2011.05.030.

Andrady, A. 2015. Plastics and environmental sustainability: Fact and fiction, Wiley.

Andrady, A. L., and Neal, M. A. 2009. Applications and societal benefits of plastics. Philos. T. Roy. Soc. B, 364: 1977-1984, https://doi.org/10.1098/rstb.2008.0304.

Andriolo, U., *et al.* 2021. Drones for litter mapping: An inter-operator concordance test in marking beached items on aerial images. Mar. Pollut. Bull., 169: 112542, https://doi.org/10.1016/j.marpolbul.2021.112542.

Ángeles Cerón, J. C., *et al.* 2021. Assessing YOLACT++ for real time and robust instance segmentation of medical instruments in endoscopic procedures. *In* IEEE EMBC, 1824-1827, https://doi.org/10.1109/EMBC46164.2021.9629914.

Asensio-Montesinos, F., *et al.* 2019. Seasonal comparison of beach litter on Mediterranean coastal sites (Alicante, SE Spain). Ocean Coast Manag., 181: 104914, https://doi.org/10.1016/j.ocecoaman.2019.104914.

Auta, H. S., *et al.* 2017. Distribution and importance of microplastics in the marine environment: A review of the sources, fate, effects, and potential solutions. Environ. Int., 102: 165-176, https://doi.org/10.1016/j.envint.2017.02.013.

Babenko, B., *et al.* 2011. Robust object tracking with online multiple instance learning. IEEE TPAMI, 33: 1619-1632, https://doi.org/10.1109/TPAMI.2010.226.

Bakir, A., *et al.* 2014. Enhanced desorption of persistent organic pollutants from microplastics under simulated physiological conditions. Environ. Pollut., 185: 16-23, https://doi.org/10.1016/j.envpol.2013.10.007.

Barboza, L., *et al.* 2018. Marine microplastic debris: An emerging issue for food security, food safety and human health. Mar. Pollut. Bull., 133: 336-348, https://doi.org/10.1016/j.marpolbul.2018.05.047.

Barnes, D., *et al.* 2009. Accumulation and fragmentation of plastic debris in global environments. Philos. T. Roy. Soc. B, 364: 1985-1998, https://doi.org/10.1098/rstb.2008.0205.

Barnes, D., and Milner, P. 2005. Drifting plastic and its consequences for sessile organism dispersal in the Atlantic Ocean. Mar. Biol., 146: 815-825, https://doi.org/10.1007/s00227-004-1474-8.

Bas, E., *et al.* 2007. Automatic vehicle counting from video for traffic flow analysis. *In* Int. Veh. Sym., 392-397, https://doi.org/10.1109/ivs.2007.4290146.

Bay, H., *et al.* 2008. Speeded-up robust features (SURF). Comput. Vis. Image Und., 110: 346-359, https://doi.org/10.1016/j.cviu.2007.09.014.

Baydogan, M. G., *et al.* 2013. A Bag-of-Features framework to classify time series. IEEE TPAMI, 35: 2796-2802, https://doi.org/10.1109/TPAMI.2013.72.
Benjdira, B., *et al.* 2019. Car detection using unmanned aerial vehicles: comparison between Faster R-CNN and YOLOv3. *In* UVS, 1-6, https://doi.org/10.1109/UVS.2019.8658300.
Bernardini, I., *et al.* 2018. First data on plastic ingestion by blue sharks (Prionace glauca) from the Ligurian Sea (North-Western Mediterranean Sea). Mar. Pollut. Bull., 135: 303-310, https://doi.org/10.1016/j.marpolbul.2018.07.022.
Betts, K. 2008. Why small plastic particles may pose a big problem in the oceans. Environ. Sci. Technol., 42: 8995-8995, https://doi.org/10.1021/es802970v.
Bochkovskiy, A., *et al.* 2020. YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint, https://doi.org/10.48550/arXiv.2004.10934.
Bolya, D., *et al.* 2019. YOLACT: Real-time instance segmentation. *In* ICCV, 9157-9166, https://doi.org/10.48550/arXiv.1904.02689.
Bolya, D., *et al.* 2020. YOLACT++: Better real-time instance segmentation. IEEE TPAMI, 44: 1108 - 1121, https://doi.org/10.1109/tpami.2020.3014297.
Booth, H., *et al.* 2023. High-precision density mapping of marine debris and floating plastics via satellite imagery. Sci. Rep., 13: 6822, https://doi.org/10.1038/s41598-023-33612-2.
Boser, B. E., *et al.* 1992. A training algorithm for optimal margin classifiers. *In* COLT, 144-152, https://doi.org/10.1145/130385.130401.
Boucher, J., and Friot, D. 2017. Primary microplastics in the oceans: a global evaluation of sources, IUCN Gland, Switzerland.
Cai, Y., *et al.* 2021. YOLOv4-5D: an effective and efficient object detector for autonomous driving. IEEE TIM, 70: 1-13, https://doi.org/10.1109/TIM.2021.3065438.
Chen, C.-H., *et al.* 2008. People counting system for getting in/out of a bus based on video processing. *In* ISDA, 3, 565-569, https://doi.org/10.1109/isda.2008.335.
Chengcui, Z., *et al.* 2005. A multiple instance learning approach for content based image retrieval using one-class support vector machine. *In* IEEE ICME, 1142-1145, https://doi.org/10.1109/ICME.2005.1521628.
Chin, C. S., *et al.* 2022. Visual marine debris detection using YOLOv5s for autonomous underwater vehicle. *In* ICIS, 20-24, https://doi.org/10.1109/ICIS54925.2022.9882484.
Chollet, F. 2017. Xception: Deep Learning with Depthwise Separable Convolutions. *In* CVPR, 1800-1807, https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.195.
Córdova, M., *et al.* 2022. Litter detection with deep learning: a comparative study. Sensors, 22: 548, https://doi.org/10.3390/s22020548.
Cox, K. D., *et al.* 2019. Human consumption of microplastics. Environ. Sci. Technol., 53: 7068-7074, https://doi.org/10.1021/acs.est.9b01517.
Cózar, A., *et al.* 2014. Plastic debris in the open ocean. PNAS, 111: 10239-10244, https://doi.org/10.1073/pnas.1314705111.
Cózar, A., *et al.* 2015. Plastic accumulation in the Mediterranean Sea. PLoS One, 10: 1-12, https://doi.org/10.1371/journal.pone.0121762.

Day, R. H., and Shaw, D. G. 1987. Patterns in the abundance of pelagic plastic and tar in the north pacific ocean, 1976–1985. Mar. Pollut. Bull., 18: 311-316, https://doi.org/10.1016/S0025-326X(87)80017-6.

de Vries, R., *et al.* 2021. Quantifying floating plastic debris at sea using vessel-based optical data and artificial intelligence. Remote Sens., 13: 3401, https://doi.org/10.3390/rs13173401.

Deidun, A., *et al.* 2018. Optimising beached litter monitoring protocols through aerial imagery. Mar. Pollut. Bull., 131: 212-217, https://doi.org/10.1016/j.marpolbul.2018.04.033.

Deng, J., *et al.* 2009. ImageNet: A large-scale hierarchical image database. *In* CVPR, 248-255, https://doi.org/10.1109/CVPR.2009.5206848.

Dewi, C., *et al.* 2021. YOLOv4 for advanced traffic sign recognition with synthetic training data generated by various GAN. IEEE Access, 9: 97228-97242, https://doi.org/10.1109/ACCESS.2021.3094201.

Dias, K., *et al.* 2022. Marine trash detection using deep learning models. *In* Computing and Communications Engineering in Real-Time Application Development, pp. 231-242. Apple Academic Press.

Dutta, A., and Zisserman, A. 2019. The VIA annotation software for images, audio and video, 2276-2279.

ECHO Instruments 2016. Bioplastic Degradation. URL: https://bit.ly/3DvP7OO, Accessed: May 11th, 2020.

Ellen MacArthur Foundation 2016. The new plastics economy d rethinking the future of plastics. URL: https://www.ellenmacarthurfoundation.org/assets/downloads/EllenMacArthurFoundation_TheNewPlasticsEconomy_Pages.pdf, Accessed: September 2nd, 2022.

Eriksen, M., *et al.* 2014. Plastic pollution in the world's oceans: More than 5 trillion plastic pieces weighing over 250,000 tons afloat at sea. PLoS One, 9: 1-15, https://doi.org/10.1371/journal.pone.0111913.

Eriksen, M., *et al.* 2013. Plastic pollution in the South Pacific subtropical gyre. Mar. Pollut. Bull., 68: 71-76, https://doi.org/10.1016/j.marpolbul.2012.12.021.

European Commission 2016. Closing the loop. New circular economy package. URL: http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/573899/EPRS_BRI(2016)573899_EN.pdf., Accessed: September 2nd, 2022.

European Commission 2018. A European strategy for plastics in a circular economy. URL: https://www.europarc.org/wp-content/uploads/2018/01/Eu-plastics-strategy-brochure.pdf, Accessed: September 2nd, 2022.

European Commission 2019. Country report - Cyprus climate change. URL: https://ec.europa.eu/environment/eir/pdf/report_cy_en.pdf, Accessed: September 2nd, 2022.

European Union 2019. Directive (EU) 2019/904 of the European parliament and of the council: on the reduction of the impact of certain plastic products on the environment. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2019.155.01.0001.01.ENG&toc=OJ%3AL%3A2019%3A155%3ATOC, Accessed: September 2nd, 2022.

Fallati, L., *et al.* 2019. Anthropogenic marine debris assessment with unmanned aerial vehicle imagery and deep learning: A case study along the beaches of the Republic of Maldives. Sci. Total Environ., 693: 133581, https://doi.org/10.1016/j.scitotenv.2019.133581.

FAO 2018. The state of world fisheries and aquaculture 2018—meeting the sustainable development goals, Rome.

Fendall, L., and Sewell, M. 2009. Contributing to marine pollution by washing your face: Microplastics in facial cleansers. Mar. Pollut. Bull., 58: 1225-1228, https://doi.org/10.1016/j.marpolbul.2009.04.025.

Fossi, M., *et al.* 2017. Are whale sharks exposed to persistent organic pollutants and plastic pollution in the Gulf of California (Mexico)? First ecotoxicological investigation using skin biopsies. Comp. Biochem. Phys. C, 199: 48-58, http://dx.doi.org/10.1016/j.cbpc.2017.03.002.

Fossi, M., *et al.* 2018. Bioindicators for monitoring marine litter ingestion and its impacts on Mediterranean biodiversity. Environ. Pollut., 237: 1023-1040, https://doi.org/10.1016/j.envpol.2017.11.019.

Fu, H., *et al.* 2022. Crack segmentation based on improved YOLACT++ algorithm. *In* CIAC, 425-432, https://doi.org/10.1007/978-981-16-6372-7_48.

Fulton, M., *et al.* 2019. Robotic detection of marine litter using deep visual detection models. *In* ICRA, 5752-5758, https://doi.org/10.1109/ICRA.2019.8793975.

Galgani, F., *et al.* 2000. Litter on the sea floor along european coasts. Mar. Pollut. Bull., 40: 516-527, https://doi.org/10.1016/S0025-326X(99)00234-9.

Gallo, F., *et al.* 2018. Marine litter plastics and microplastics and their toxic chemicals components: the need for urgent preventive measures. Environ. Sci. Eur., 30: 13, https://doi.org/10.1186/s12302-018-0139-z.

Ge, Z., *et al.* 2016. Semi-automatic recognition of marine debris on beaches. Sci. Rep., 6: 25759, https://doi.org/10.1038/srep25759.

Goldstein, M., *et al.* 2013. Scales of spatial heterogeneity of plastic parine debris in the Northeast Pacific ocean. PLoS One, 8: e80020, https://doi.org/10.1371/journal.pone.0080020.

Goldstein, M. C., *et al.* 2014. Relationship of diversity and habitat area in North Pacific plastic-associated rafting communities. Mar. Biol., 161: 1441-1453, https://doi.org/10.1007/s00227-014-2432-8.

Gonçalves, G., and Andriolo, U. 2022. Operational use of multispectral images for macro-litter mapping and categorization by Unmanned Aerial Vehicle. Mar. Pollut. Bul., 176: 113431, https://doi.org/10.1016/j.marpolbul.2022.113431.

Gonçalves, G., *et al.* 2020. Quantifying marine macro litter abundance on a sandy beach using unmanned aerial systems and object-oriented machine learning methods. Remote Sens., 12: 2599, https://doi.org/10.3390/rs12162599.

González-Fernández, D., and Hanke, G. 2017. Toward a harmonized approach for monitoring of riverine floating macro litter inputs to the marine environment. Front. Mar. Sci., 4: 86, https://doi.org/10.3389/fmars.2017.00086.

Goodfellow, I., *et al.* 2016. Deep learning, MIT Press.

Graca, B., *et al.* 2017. Sources and fate of microplastics in marine and beach sediments of the Southern Baltic Sea— a preliminary study. Environ. Sci. Pollut. R., 24: 7650-7661, https://doi.org/10.1007/s11356-017-8419-5.

Guo, F., *et al.* 2021. Real-time railroad track components inspection based on the improved YOLOv4 framework. Autom. Constr., 125: 103596, https://doi.org/10.1016/j.autcon.2021.103596.

Harris, C., and Stephens, M. 1988. A combined corner and edge detector. *In* AVC, 147-151, https://dx.doi.org/10.5244/C.2.23.

Harse, G. A. 2011. Plastic, the great pacific garbage patch, and international misfires at a cure. UCLA J. Environ. L. Policy, 29: 331, https://doi.org/10.5070/L5292019968.

Hastie, T., *et al.* 2009. The elements of statistical learning: data mining, inference, and prediction, Springer Science & Business Media.

He, K., *et al.* 2017. Mask R-CNN. *In* IEEE ICCV, 2980-2988, https https://doi.org/10.1109/ICCV.2017.322.

He, K., *et al.* 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE TPAMI, 37: 1904-1916, https://doi.org/10.1109/TPAMI.2015.2389824.

He, K., *et al.* 2016. Deep residual learning for image recognition. *In* CVPR, 770-778, https://doi.org/10.1109/CVPR.2016.90.

Heitz, R. P. 2014. The speed-accuracy tradeoff: history, physiology, methodology, and behavior. Front. Neurosci., 8: 150, https://doi.org/10.3389/fnins.2014.00150.

Hengstmann, E., *et al.* 2017. Marine litter abundance and distribution on beaches on the Isle of Rügen considering the influence of exposition, morphology and recreational activities. Mar. Pollut. Bull., 115: 297-306, http://dx.doi.org/10.1016/j.marpolbul.2016.12.026.

Hipolito, J. C., *et al.* 2021. Detection of underwater marine plastic debris using an augmented low sample size dataset for machine vision system: a deep transfer learning approach. *In* IEEE SCOReD, 82-86, https://doi.org/10.1109/SCOReD53546.2021.9652703.

Hridayami, P., *et al.* 2019. Fish species recognition using VGG16 deep convolutional neural network. JCSE, 13: 124-130, http://dx.doi.org/10.5626/JCSE.2019.13.3.124.

Huang, Y.-Q., *et al.* 2020. Optimized YOLOv3 algorithm and its application in traffic flow detections. Appl. Sci., 10: 3079, https://doi.org/10.3390/app10093079.

Ieamsaard, J., *et al.* 2021. Deep learning-based face mask detection using YOLOV5. *In* iEECON, 428-431, https://doi.org/10.1109/iEECON51072.2021.9440346.

Impact Subsea. 2016. ISA500 altitude, pitch, roll and heading sensor; Installation and operation manual URL: https://www.impactsubsea.co.uk/wp-content/uploads/2018/08/ISA500-Manual.pdf, Accessed: August 19th, 2018.

Jakovljevic, G., *et al.* 2020. A deep learning model for automatic plastic mapping using unmanned aerial vehicle (UAV) data. Remote Sens., 12: 1515, https://doi.org/10.3390/rs12091515.

Jambeck, J., *et al.* 2015. Plastic waste inputs from land into the ocean. Science, 347: 768-771, https://doi.org/10.1126/science.1260352.

Jang, Y. C., *et al.* 2014. Estimation of lost tourism revenue in Geoje Island from the 2011 marine debris pollution event in South Korea. Mar. Pollut. Bull., 81: 49-54, https://doi.org/10.1016/j.marpolbul.2014.02.021.

Jiao, Z., *et al.* 2019. A deep learning based forest fire detection approach using UAV and YOLOv3. *In* IAI, 1-5, https://doi.org/10.1109/ICIAI.2019.8850815.

Jocher, G., *et al.* 2020. Bug fixes and performance improvements (ultralytics/yolov5: v3.1). URL: https://zenodo.org/record/4154370#.Y1-rBHZBy5c, Accessed: June 10th, 2020.

Kako, S., *et al.* 2020. Estimation of plastic marine debris volumes on beaches using unmanned aerial vehicles and image processing based on deep learning. Mar. Pollut. Bull., 155: 111127, https://doi.org/10.1016/j.marpolbul.2020.111127.

Khachatryan, L. 2019. Centroid based object tracking. URL: https://github.com/lev1khachatryan/Centroid-Based_Object_Tracking, Accessed: July 30th, 2021.

Kikaki, A., *et al.* 2020. Remotely sensing the source and transport of marine plastic debris in Bay Islands of Honduras (Caribbean Sea). Remote Sens., 12: 1727, https://doi.org/10.3390/rs12111727.

Kögel, T., *et al.* 2020. Micro- and nanoplastic toxicity on aquatic life: Determining factors. Sci. Total Environ., 709: 136050, https://doi.org/10.1016/j.scitotenv.2019.136050.

Krelling, A. P., *et al.* 2017. Differences in perception and reaction of tourist groups to beach marine debris that can influence a loss of tourism revenue in coastal areas. Mar. Policy, 85: 87-99, https://doi.org/10.1016/j.marpol.2017.08.021.

Krizhevsky, A., *et al.* 2017. Imagenet classification with deep convolutional neural networks. *In* NeurIPS, 60, 84-90, https://doi.org/10.1145/3065386.

Kuznetsova, A., *et al.* 2020. Detecting apples in orchards using YOLOv3. *In* ICCSA, 923-934, https://doi.org/10.1007/978-3-030-58799-4_66.

Kylili, K., *et al.* 2021. A new paradigm for estimating the prevalence of plastic litter in the marine environment. Mar. Pollut. Bull., 173: 113127, https://doi.org/10.1016/j.marpolbul.2021.113127.

Kylili, K., *et al.* 2020. An intelligent way for discerning plastics at the shorelines and the seas. Environ. Sci. Pollut. R., 27: 42631–42643, https://doi.org/10.1007/s11356-020-10105-7.

Kylili, K., *et al.* 2019. Identifying floating plastic marine debris using a deep learning approach. Environ. Sci. Pollut. R., 26: 17091-17099, https://doi.org/10.1007/s11356-019-05148-4.

Laptev, I., *et al.* 2008. Learning realistic human actions from movies. *In* CVPR, 1-8, https://doi.org/10.1109/CVPR.2008.4587756.

Lavers, J. L., and Bond, A. L. 2017. Exceptional and rapid accumulation of anthropogenic debris on one of the world's most remote and pristine islands. PNAS, 114: 6052-6055, https://doi.org10.1073/pnas.1619818114.

Law, K. 2016. Plastics in the marine environment. Annu. Rev. Mar. Sci, 9: 10.11–10.25, https://doi.org/10.1146/annurev-marine-010816-060409.

Lebreton, L., *et al.* 2018. Evidence that the Great Pacific Garbage Patch is rapidly accumulating plastic. Sci. Rep., 8: 4666, https://doi.org/10.1038/s41598-018-22939-w.
Letcher, T. M. 2020. Plastic waste and recycling: environmental impact, societal issues, prevention, and solutions, Academic Press.
Leutenegger, S., *et al.* 2011. BRISK: Binary robust invariant scalable keypoints. *In* ICCV, 2548-2555, https://doi.org/10.1109/iccv.2011.6126542.
Li, C., *et al.* 2020. Face detection based on YOLOv3. *In* AISC, 1006, 277-284, https://doi.org/10.1007/978-981-13-9406-5_34.
Li, Y., *et al.* 2017. Fully convolutional instance-aware semantic segmentation. *In* CVPR, 4438-4446, https://doi.org/10.1109/CVPR.2017.472.
Lin, G.-S., *et al.* 2021. Instance segmentation based on deep convolutional neural networks and transfer learning for unconstrained psoriasis skin images. Appl. Sci., 11: 3155, https://doi.org/10.3390/app11073155.
Lin, T.-Y., *et al.* 2017a. Feature pyramid networks for object detection. *In* CVPR, 936-944, https://doi.org/10.1109/cvpr.2017.106.
Lin, T.-Y., *et al.* 2017b. Focal loss for dense object detection. IEEE TPAMI, 42: 318-327, https://doi.org/10.1109/TPAMI.2018.2858826.
Lin, T.-Y., *et al.* 2014. Microsoft COCO: common objects in context. *In* ECCV, 740-755, 10.1007/978-3-319-10602-1_48.
Liu, G., *et al.* 2020a. YOLO-Tomato: a robust algorithm for tomato detection based on YOLOv3. Sensors, 20: 2145, https://doi.org/10.3390/s20072145.
Liu, S., *et al.* 2018. Path aggregation network for instance segmentation. *In* CVPR, 8759-8768, https://doi.org/10.1109/cvpr.2018.00913.
Liu, W., *et al.* 2016. SSD: Single shot multibox detector. *In* ECCV, 21-37, https://doi.org/10.1007/978-3-319-46448-0_2.
Liu, Z., *et al.* 2020b. Improved kiwifruit detection using pre-trained VGG16 with RGB and NIR information fusion. IEEE Access, 8: 2327-2336, https://doi.org/10.1109/ACCESS.2019.2962513.
Lobelle, D., and Cunliffe, M. 2011. Early microbial biofilm formation on marine plastic debris. Mar. Pollut. Bull., 62: 197-200, https://doi.org/10.1016/j.marpolbul.2010.10.013.
Lowe, D. 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis., 60: 91-110, https://doi.org/10.1023/b:visi.0000029664.99615.94.
Mahto, P., *et al.* 2020. Refining YOLOv4 for vehicle detection. IJARET, 11, https://10.34218/IJARET.11.5.2020.043.
Mandić, M., *et al.* 2021. Quantification and classification of beach litter in Montenegro (South-East Adriatic Sea). *In* The Montenegrin Adriatic Coast: Marine Chemistry Pollution, pp. 257-274. Springer, Cham.
Maron, O., and Ratan, A. L. 1998. Multiple-instance learning for natural scene classification. *In* ICML, 98, 341-349, https://dl.acm.org/doi/10.5555/645527.657299.

Martin, C., *et al.* 2018. Use of unmanned aerial vehicles for efficient beach litter monitoring. Mar. Pollut. Bull., 131: 662-673, https://doi.org/10.1016/j.marpolbul.2018.04.045.

Martínez-Vicente, V., *et al.* 2019. Measuring marine plastic debris from space: Initial assessment of observation requirements. Remote Sens., 11: 20, https://doi.org/10.3390/rs11202443.

MathWorks 2016. Image classification with bag of visual words. URL: https://www.mathworks.com/help/vision/ug/image-classification-with-bag-of-visual-words.html, Accessed: September 20th, 2017.

Matsuoka, T., *et al.* 2005. A review of ghost fishing: scientific approaches to evaluation and solutions. Fisheries Sci., 71: 691-702, https://doi.org/10.1111/j.1444-2906.2005.01019.x.

Maximenko, N., *et al.* 2012. Pathways of marine debris derived from trajectories of Lagrangian drifters. Mar. Pollut. Bull., 65: 51-62, https://doi.org/10.1016/j.marpolbul.2011.04.016.

Mazurek, P. 2020. YOLO object counting API. URL: https://github.com/tugot17/YOLO-Object-Counting-API, Accessed: July 30th, 2021.

Memon, S., *et al.* 2018. A video based vehicle detection, counting and classification system. Int. J. Image Graph., 11: 34, https://doi.org/10.5815/ijigsp.2018.09.05.

Mohiyuddin, A., *et al.* 2022. Breast tumor detection and classification in mammogram images using modified YOLOv5 network. Comput. Math. Methods Med., 2022: 1359019, https://doi.org/10.1155/2022/1359019.

Moore, C. 2008. Synthetic polymers in the marine environment: A rapidly increasing, long-term threat. Environ. Res., 108: 131-139, https://doi.org/10.1016/j.envres.2008.07.025.

Moravec, H. 1980. Obstacle avoidance and navigation in the real world by a seeing robot rover. Carnegie Mellon University, Journal contribution, https://doi.org/10.1184/R1/6557033.v1.

Moy, K., *et al.* 2018. Mapping coastal marine debris using aerial imagery and spatial analysis. Mar. Pollut. Bull., 132: 52-59, https://doi.org/10.1016/j.marpolbul.2017.11.045.

Musić, J., *et al.* 2020. Detecting underwater sea litter using deep neural networks: an initial study. *In* SpliTech, 1-6, https://doi.org/10.23919/SpliTech49282.2020.9243709.

Nachite, D., *et al.* 2019. Spatial and temporal variations of litter at the Mediterranean beaches of Morocco mainly due to beach users. Ocean Coast Manag., 179: 104846, https://doi.org/10.1016/j.ocecoaman.2019.104846.

National Oceanic and Atmospheric Administration (NOAA) 2018. NOAA Photo Library, Online: https://photolib.noaa.gov/. URL: https://photolib.noaa.gov/, Accessed: 1/6.

Nelms, S., *et al.* 2015. Plastic and marine turtles: a review and call for research. ICES J. Mar. Sci., 73: 165-181, https://doi.org/10.1093/icesjms/fsv165.

Newman, P. 2014. Plastic, Ahoy!: Investigating the Great Pacific Garbage Patch, Lerner Publishing Group.

Nistér, D., and Stewénius, H. 2008. Linear time maximally stable extremal regions. *In* ECCV, 183-196, https://doi.org/10.1007/978-3-540-88688-4_14.

Orthodoxou, D. L., *et al.* 2022. Seasonal and geographic variations of marine litter: A comprehensive study from the island of Cyprus. Mar. Pollut. Bull., 177: 113495, https://doi.org/10.1016/j.marpolbul.2022.113495.

Ostle, C., *et al.* 2019. The rise in ocean plastics evidenced from a 60-year time series. Nat. Commun., 10: 1622, https://doi.org/10.1038/s41467-019-09506-1.

Pabortsava, K., and Lampitt, R. S. 2020. High concentrations of plastic hidden beneath the surface of the Atlantic Ocean. Nat. Commun., 11, https://doi.org/10.1038/s41467-020-17932-9.

Padilla, R., *et al.* 2020. A Survey on performance metrics for object-detection algorithms. *In* IWSSIP, 237-242, https://doi.org/10.1109/IWSSIP48289.2020.9145130.

Papakonstantinou, A., *et al.* 2021. A citizen science unmanned aerial system data acquisition protocol and deep learning techniques for the automatic detection and mapping of marine litter concentrations in the coastal zone. Drones, 5: 6, https://doi.org/10.3390/drones5010006.

Parico, A. I. B., and Ahamed, T. 2021. Real time pear fruit detection and counting using YOLOv4 models and Deep SORT. Sensors, 21: 4803, https://doi.org/10.3390/s21144803.

PEW Charitable Trusts, and SystemIQ 2020. Breaking the plastic wave: A comprehensive assessment of pathways towards stopping ocean plastic pollution. URL: https://www.pewtrusts.org/en/research-and-analysis/articles/2020/07/23/breaking-the-plastic-wave-top-findings, Accessed: October 13th, 2019.

Pham, C. K., *et al.* 2014. Marine litter distribution and density in European seas, from the shelves to deep basins. PLoS One, 9: e95839, https://doi.org/10.1371/journal.pone.0095839.

PlasticsEurope. 2013. Plastics – the Facts 2013: An analysis of European latest plastics production, demand and waste data., URL: https://plasticseurope.org/wp-content/uploads/2021/10/2013-Plastics-the-facts.pdf, Accessed: June 30th, 2019.

PlasticsEurope. 2019. Plastics – the Facts 2019: An analysis of European plastics production, demand and waste data., URL: https://plasticseurope.org/wp-content/uploads/2021/10/2019-Plastics-the-facts.pdf, Accessed: June 30th, 2019.

Póvoa, A., *et al.* 2021. Fouling organisms in marine litter (rafting on abiogenic substrates): A global review of literature. Mar. Pollut. Bull., 166: 112189, https://doi.org/10.1016/j.marpolbul.2021.112189.

Rainey, J. 2019. Ocean cleanup device breaks down, well before ridding Pacific of plastics. URL: https://www.nbcnews.com/news/us-news/ocean-cleanup-device-breaks-down-well-ridding-pacific-plastics-n954446, Accessed: October 4th, 2022.

Raykar, V. C., *et al.* 2008. Bayesian multiple instance learning: automatic feature selection and inductive transfer. *In* ICML, 808–815, https://doi.org/10.1145/1390156.1390258.

Redmon, J., and Farhadi, A. 2017. YOLO9000: better, faster, stronger. *In* CVPR, 7263-7271, https://doi.org/10.1109/cvpr.2017.690.

Redmon, J., and Farhadi, A. 2018. YOLOv3: An incremental improvement. arXiv preprint, https://doi.org/10.48550/arXiv.1804.02767.

Ren, S., *et al.* 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE TPAMI, 39: 1137-1149, https://doi.org/10.1109/TPAMI.2016.2577031.

Rezatofighi, H., *et al.* 2019. Generalized intersection over union: a metric and a loss for bounding box regression. *In* CVPR, 658-666, https://doi.org/10.1109/CVPR.2019.00075.

Rezende, E., *et al.* 2018. Malicious software classification using VGG16 deep neural network's Bottleneck features. *In* ITNG, 51-59, https://doi.org/10.1007/978-3-319-77028-4_9.

Rosebrock, A. 2021a. OpenCV people counter. URL: https://www.pyimagesearch.com/2018/08/13/opencv-people-counter, Accessed: July 30th, 2021.

Rosebrock, A. 2021b. Simple object tracking with OpenCV. URL: https://www.pyimagesearch.com/2018/07/23/simple-object-tracking-with-opencv, Accessed: July 30th, 2021.

Rosevelt, C., *et al.* 2013. Marine debris in central California: quantifying type and abundance of beach litter in Monterey Bay, CA. Mar. Pollut. Bull., 71: 299-306, https://doi.org/10.1016/j.marpolbul.2013.01.015.

Roshan, J. V. 2022. Optimal ratio for data splitting. Stat. Anal. Data Min., 15: 531-538, https://doi.org/10.1002/sam.11583.

Rosten, E., and Drummond, T. 2005. Fusing points and lines for high performance tracking. *In* ICCV, 2, 1508-1515, https://doi.org/10.1109/iccv.2005.104.

Ryan, P., *et al.* 2009. Monitoring the abundance of plastic debris in the marine environment. Philos. T. Roy. Soc. B Biol. Sci., 364: 1999-2012, https://doi.org/10.1098/rstb.2008.0207.

Ryan, P. G. 2018. Entanglement of birds in plastics and other synthetic materials. Mar. Pollut. Bull., 135: 159-164, https://doi.org/10.1016/j.marpolbul.2018.06.057.

Setälä, O., *et al.* 2016. Feeding type affects microplastic ingestion in a coastal invertebrate community. Mar. Pollut. Bull., 102: 95-101, https://doi.org/10.1016/j.marpolbul.2015.11.053.

Shi, J., and Tomasi, C. 1994. Good features to track. *In* CVPR, 593-600, https://doi.org/10.1109/cvpr.1994.323794.

Silva-Iñiguez, L., and Fischer, D. 2003. Quantification and classification of marine litter on the municipal beach of Ensenada, Baja California, Mexico. Mar. Pollut. Bull., 46: 132-138, https://doi.org/10.1016/S0025-326X(02)00216-3.

Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint, https://doi.org/10.48550/arXiv.1409.1556.

Singh, S., *et al.* 2021. Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment. Multimed. Tools Appl., 80: 19753-19768, https://doi.org/10.1007/s11042-021-10711-8.

Smith, S. D. A., *et al.* 2014. Patterns of marine debris distribution on the beaches of Rottnest Island, Western Australia. Mar. Pollut. Bull., 88: 188-193, https://doi.org/10.1016/j.marpolbul.2014.09.007.

Song, X., *et al.* 2021. Large plastic debris dumps: New biodiversity hot spots emerging on the deep-sea floor. Environ. Sci. Technol. Lett., 8: 148-154, https://doi.org/10.1021/acs.estlett.0c00967.

Szegedy, C., *et al.* 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. *In* AAAI, 31, https://doi.org/10.1609/aaai.v31i1.11231.

Szegedy, C., *et al.* 2015. Going deeper with convolutions. *In* CVPR, 1-9, https://doi.org/10.1109/CVPR.2015.7298594.

Tagg, A., *et al.* 2015. Identification and quantification of microplastics in wastewater using focal plane array-based reflectance micro-FT-IR imaging. Anal. Chem., 87: 6032-6040, https://doi.org/10.1021/acs.analchem.5b00495.

Takaya, K., *et al.* 2022. Unmanned aerial vehicles and deep learning for assessment of anthropogenic marine debris on beaches on an island in a semi-enclosed sea in Japan. Environ. Res. Commun., 4: 015003, https://doi.org/10.1088/2515-7620/ac473b.

Tata, G., *et al.* 2021. A robotic approach towards quantifying epipelagic bound plastic using deep visual models. arXiv preprint, https://doi.org/10.48550/arXiv.2105.01882.

Teng, C., *et al.* 2022. Deploying deep learning to estimate the abundance of marine debris from video footage. Mar. Pollut. Bull., 183: 114049, https://doi.org/10.1016/j.marpolbul.2022.114049.

Teuten, E. L., *et al.* 2009. Transport and release of chemicals from plastics to the environment and to wildlife. Philos. T. Roy. Soc. B Biol. Sci., 364: 2027-2045, https://doi.org/10.1098/rstb.2008.0284.

Thompson, R., *et al.* 2009a. Plastics, the environment and human health: Current consensus and future trends. Philos. T. Roy. Soc. B Biol. Sci., 364: 2153, https://doi.org/10.1098/rstb.2009.0053.

Thompson, R. C., *et al.* 2009b. Our plastic age. Philos. T. Roy. Soc. B, 364: 1973-1976, https://doi.org/10.1098/rstb.2009.0054.

Tian, M., *et al.* 2022. A modified YOLOv4 detection method for a vision-based underwater garbage cleaning robot. Front. Inform. Technol. Electron. Eng., 23: 1217-1228, https://doi.org/10.1631/FITEE.2100473.

Ting, L., *et al.* 2021. Ship detection algorithm based on improved YOLOv5. *In* CACRE, 483-487, https://doi.org/10.1109/CACRE52464.2021.9501331.

Tomas, J. P. Q., *et al.* 2022. Trash detection for computer vision using Scaled-YOLOv4 on water surface. *In* IEEA, 1–8, https://doi.org/10.1145/3533254.3533255.

Topçu, N., *et al.* 2012. Origin and abundance of marine litter along sandy beaches of the Turkish Western Black Sea Coast. Mar. Environ. Res., 85: 21-28, https://doi.org/10.1016/j.marenvres.2012.12.006.

Topouzelis, K., *et al.* 2019. Detection of floating plastics from satellite and unmanned aerial systems (Plastic Litter Project 2018). Int. J. Appl. Earth Obs., 79: 175, https://doi.org/10.1016/j.jag.2019.03.011.

Unger, B., *et al.* 2016. Large amounts of marine debris found in sperm whales stranded along the North Sea coast in early 2016. Mar. Pollut. Bull., 112: 134-141, https://doi.org/10.1016/j.marpolbul.2016.08.027.

USEPA. 2011. Municipal solid waste in the United States tables and figures for 2010. URL: http://www.epa.gov/osw/nonhaz/municipal/pubs/MSWcharacterization_fnl_060713_2_rpt.pdf, Accessed: September 2nd, 2022.

Valdenegro-Toro, M. 2016. Submerged marine debris detection with autonomous underwater vehicles. *In* RAHA, 7, https://doi.org/10.1109/RAHA.2016.7931907.

van Sebille, E., *et al.* 2012. Origin, dynamics and evolution of ocean garbage patches from observed surface drifters. Environ. Res. Lett., 7: 044040, https://doi.org/10.1088/1748-9326/7/4/044040.

Veerasingam, S., *et al.* 2022. Detection and assessment of marine litter in an uninhabited island, Arabian Gulf: a case study with conventional and machine learning approaches. Sci. Total Environ., 838: 156064, https://doi.org/10.1016/j.scitotenv.2022.156064.

Viola, P., and Jones, M. 2001. Rapid object detection using a boosted cascade of simple features. *In* CVPR, 511–518, https://doi.org/10.1109/cvpr.2001.990517.

Vlachogianni, T., *et al.* 2017. Marine litter assessment in the Adriatic and Ionian Seas, IPA-Adriatic DeFishGear Project, MIO-ECSDE, HCMR and ISPRA, 168.

Wang, C., *et al.* 2020. CSPNet: A new backbone that can enhance learning capability of CNN. *In* CVPR, 390-391, https://doi.org/10.1109/cvprw50498.2020.00203.

Wang, J., *et al.* 2022. Light enhancement algorithm optimization for autonomous driving vision in night scenes based on YOLACT++. *In* ISPDS, 417-423, https://doi.org/10.1109/ISPDS56360.2022.9874070.

Wang, Y.-M., *et al.* 2006. On the centroids of fuzzy numbers. Fuzzy Set. Syst., 157: 919-926, https://doi.org/10.1016/j.fss.2005.11.006.

Waring, R. H., *et al.* 2018. Plastic contamination of the food chain: A threat to human health? Maturitas, 115: 64-68, https://doi.org/10.1016/j.maturitas.2018.06.010.

Watanabe, J.-I., *et al.* 2019. Underwater and airborne monitoring of marine ecosystems and debris. J. Appl. Rem. Sens., 13: 044509, https://doi.org/10.1117/1.JRS.13.044509.

Watts, M. J., *et al.* 2013. Tracking shifting range margins using geographical centroids of metapopulations weighted by population density. Ecol. Model., 269: 61-69, https://doi.org/10.1016/j.ecolmodel.2013.08.010.
Wen, H., *et al.* 2021. Detecting the surface defects of the magnetic-tile based on improved YOLACT ++. *In* IEEE ICCT, 1097-1102, https://doi.org/10.1109/ICCT52962.2021.9658066.
Wessel, C. C., *et al.* 2016. Abundance and characteristics of microplastics in beach sediments: Insights into microplastic accumulation in northern Gulf of Mexico estuaries. Mar. Pollut. Bull., 109: 178-183, https://doi.org/10.1016/j.marpolbul.2016.06.002.
Wilcox, C., *et al.* 2016. Using expert elicitation to estimate the impacts of plastic pollution on marine wildlife. Mar. Policy, 65: 107-114, https://doi.org/10.1016/j.marpol.2015.10.014.
Woodall, L., *et al.* 2014. The deep sea is a major sink for microplastic debris. R. Soc. Open Sci., 1: 140317, https://doi.org/10.1098/rsos.140317.
Wright, S. L., *et al.* 2013. The physical impacts of microplastics on marine organisms: A review. Environ. Pollut., 178: 483-492, https://doi.org/10.1016/j.envpol.2013.02.031.
Xie, S., *et al.* 2017. Aggregated residual transformations for Deep Neural Networks. *In* CVPR, 5987-5995, https://doi.org/10.1109/CVPR.2017.634.
Xu, H., *et al.* 2018. Foreign object debris material recognition based on convolutional neural networks. EURASIP J. Image Video Proc, 2018: 21, https://doi.org/10.1186/s13640-018-0261-2.
Xu, X., *et al.* 2022. Lite-YOLOv5: a lightweight deep learning detector for on-board ship detection in large-scene Sentinel-1 SAR images. Remote Sens., 14: 1018, https://doi.org/10.3390/rs14041018.
Xu, Z.-F., *et al.* 2020. Light-YOLOv3: fast method for detecting green mangoes in complex scenes using picking robots. Appl. Intell., 50: 4670-4687, https://doi.org/10.1007/s10489-020-01818-w.
Xue, B., *et al.* 2021. An efficient deep-sea debris detection method using deep neural networks. IEEE JSTARS, 14: 12348-12360, https://doi.org/10.1109/JSTARS.2021.3130238.
Yan, B., *et al.* 2021. A real-time apple targets detection method for picking robot based on improved YOLOv5. Remote Sens., 13: 1619, https://doi.org/10.3390/rs13091619.
Yang, Y.-L., and Gao, W.-W. 2019. A method of pedestrians counting based on deep learning. *In* EITCE, 2010-2013, https://doi.org/10.1109/eitce47263.2019.9094838.
Yijing, W., *et al.* 2021. Fig fruit recognition method based on YOLOv4 deep learning. *In* ECTI-CON, 303-306, https://doi.org/10.1109/ECTI-CON51831.2021.9454904.
Zailan, N. A., *et al.* 2021. YOLO-based network fusion for riverine floating debris monitoring system. *In* ICECCE, 1-5, https://doi.org/10.1109/ICECCE52056.2021.9514096.

Zhang, Q., *et al.* 2009. Face detection and tracking in color images using color centroids segmentation. *In* ROBIO, 1008-1013, https://doi.org/10.1109/robio.2009.4913138.
Zhou, F., *et al.* 2021. Safety helmet detection based on YOLOv5. *In* IEEE ICPECA, 6-11, https://doi.org/10.1109/ICPECA51329.2021.9362711.

Appendix

A1. Bottleneck method (BM)

The modifications made to the Bottleneck method (BM) model which relate to the data augmentation and pre-processing manipulations, are illustrated in Figure A1. Initially, images belonging to the marine debris dataset were augmented through a number of random transformations and normalization operations, as shown in Figure A1. For example, parameter *rotation_range* randomly rotates pictures and gains a value ranging from 0°–180°. Operations referred to as *width_shift* and *height_shift* randomly translate images horizontally and vertically, respectively, while terms *shear_range*, *zoom_range*, and *horizontal_flip* randomly apply shearing transformations, zoom and flip images, correspondingly. Lastly, term *fill_mode* fills newly created pixels at the empty parts of the images that result after rotation or width and height transformations.

Followed then, the augmented image dataset underwent splitting operation. Since hundreds of new images were created in each object category after DA manipulations, the augmented data were automatically divided into two sub-sets, namely, the training and test sets, using the *train_test_split* command as described in Figure A1. The split command randomly allocates images in training and test sets with the ratio stated in its arguments of the parentheses following the *train_test_split* command. In this case, 80% of the augmented data was assigned to the training set while the remaining 20% was dedicated to the test set.

After splitting the data into the two sub-sets, the one hot encoding process was adopted in the BM image classifier, as depicted in Figure A1. In particular, the *onehot_encoder.fit_transform* command was used to transform multi-class categorical variables into numerical features that can easily feed and train the BM classifier and improve its classification accuracy. Finally, the prediction part was implemented after the BM training and testing process so as to make predictions on the validation images and write the class label on them.

Preparing images

- Import libraries
- Define values of parameters and paths to files
- Apply DA manipulations to the images

```
datagen = ImageDataGenerator( rotation_range=40, width_shift_range=0.2,
                              height_shift_range=0.2,  shear_range=0.2, zoom_range=0.2,
                              horizontal_flip=True, fill_mode='nearest')
```

- Locally store augmented images (Equal number of images per class folder)
- Split images into training and test sets

```
x_train_net, x_test_net, y_train_net, y_test_net = train_test_split(Class, Label,
                                    test_size=0.2, random_state=3, shuffle=False)
```

- One hot encoding class names to numerical values

```
onehot_encoded = onehot_encoder.fit_transform(integer_encoded)
```

Train VGG16 model and save Bottleneck features

- Rescale images
- Preparing data and class labels

```
train_data = np.load(open('bottleneck_features_train_7categ_250.npy', "rb"))
train_labels = np.array([0] * (nb_train_samples // 7) + [1] × (nb_train_samples // 7)
                    + [2] × (nb_train_samples // 7) + ... + [6] × (nb_train_samples // 7))
train_labels = to_categorical(train_labels, num_classes=num_classes)
```

- Repeat the same procedure for test data

```
model = Sequential()
model.add(Flatten(input_shape=train_data.shape[1:]))
model.add(Dense(256, activation='relu'))
model.add(Dropout(0.25))
model.add(Dense(num_classes, kernel_regularizer, activation='softmax'))
model.compile(optimizer='adadelta', loss='categorical_crossentropy', metrics=['accuracy'])
model.fit(train_data, train_labels, epochs=epochs, batch_size=batch_size,
                                    tast_data=(tast_data, tast_labels))
```

- Save weights and evaluate model
- Plot training loss, training accuracy, testing loss, and testing accuracy

Prediction part

- Load class indices
- Load model and make predictions on images

```
model = applications.VGG16(include_top=False, weights='imagenet')
bottleneck_prediction = model.predict(image)
model = Sequential()
model.add(Flatten(input_shape=bottleneck_prediction.shape[1:]))
model.add(Dense(256, activation='relu'))
model.add(Dropout(0.25))
model.add(Dense(num_classes, activation='softmax'))
model.load_weights(top_model_weights_path)
class_predicted = model.predict_classes(bottleneck_prediction)
probabilities = model.predict_proba(bottleneck_prediction)
```

- Print probabilities

Figure A1: Code sample displaying the adopted commands for improving the identification accuracy of the BM image classifier.

A2. YOLOv5 and counting tools

A2.1. ROI line counting method

Designed for counting the number of objects that cross a virtual barrier, the ROI line counting method was implemented in the YOLOv5 object detector, as shown in Figure A2. In this context, the detected marine debris were counted by the pertinent technique as the video pans over the beach or the sea in a single direction with a vertical ROI line. The detected litter items were counted by the method when the middle of the bounding box (vertically), *center = width/2* and *box_x =(x1+x2)/2*, passed the ROI line as described in Figure A2. Next, the location of the ROI line was selected and placed in the middle of the video frame. Lastly, a thickness was assigned to the ROI line because the bounding box of a detected item may move a number of pixels in between video frames and pass the line without the method counting it.

```
if roi:
        center = width/2
        box_x = (x1+x2)/2
        # roi line at middle
        if (box_x <= center + 4) and (box_x >= center - 4):
                total_count[names[int(cls)]] = total_count[names[int(cls)]] + 1
        …
        cv2.line(im0, (int(width/2), 0), (int(width/2), int(height)), (255,255,0), 4)
```

Figure A2: Part of the ROI line counting code implemented in the YOLOv5 object detection model.

A2.2. Centroid tracking counting technique

During the implementation of the centroid tracking technique in the YOLOv5 model, the bounding boxes of the detected litter items in each video frame were collected and passed into the counting technique, as partly described at the top part of Figure A3. Followed then, parameters *maxDisappeared* and *thresholdFrames* were implemented in the YOLOv5 model for alleviating the problem of overcounting or undercounting marine debris items in video footage. Recalling, *maxDisappeared*

parameter refers to the number of frames an object must remain undetected on screen for the method to stop tracking it, while the *thresholdFrames* parameter refers to the number of frames that an object must be present on screen for the method to count it.

```
ct = CentroidTracker(maxDisappeared=40, thresholdFrames=60)
…
# add box to rects
rects.append((x1, y1, x2, y2))
rects_labels.append(names[int(cls)])
…
# update our centroid tracker using the computed set of
# bounding box rectangles
objects, total_count = ct.update(rects, rects_labels)

…
if thresholdFrames > 0:
      frames = self.temp[objectID][1] + 1
      label = self.temp[objectID][2]
      self.temp[objectID] = [inputCentroids[col], frames,label]
      if frames == self.thresholdFrames:
            self.total[label] += 1
…
# loop over the tracked objects
for (objectID, centroid) in objects.items():
# draw both the ID of the object and the centroid of the Object
# on the output frame
text = "ID {}".format(objectID)
cv2.putText(im0, text, (centroid[0] - 10, centroid[1] - 10),
cv2.FONT_HERSHEY_SIMPLEX, 0.5, (0, 255, 0), 2)
cv2.circle(im0, (centroid[0], centroid[1]), 4, (0, 255, 0), -1)
```

Figure A3: Part of the centroid tracking counting code implemented in the YOLOv5 object detector.

All of the modifications, illustrated in Figure A3, have been made as part of the task of counting litter items and sea life from up to nine categories while tallying the detected objects with more accuracy. To attain this goal, every time a new object was registered and tracked by the centroid tracking counting tool, its respective class label as flagged by YOLOv5 marine debris detectors was also recorded. Followed then, if the value of the customised *thresholdFrames* parameter was greater than zero, the counting technique calculated the number of frames that the object existed on the screen. Once the value *thresholdFrames* factor surpassed a certain threshold of 60

frames, the counting value for the respective object category was incremented by one, as shown in Figure A3.

A3. OpenCV Contours image processing tool

In this part, the *OpenCV Contours* image processing tool implemented in the YOLACT++ instance segmentation tool for estimating the litter's physical dimensions is explained. Figure A4 demonstrates part of the pertinent image processing tool. Once the YOLACT++ marine debris detector applied a mask on each recognised litter item, the output images containing the detected litter along their coloured masks were stored for post-processing.

As described in Figure A4, the images depicting the detected litter items were converted into grayscale using the *cv2.cvtColor()* command. Afterwards, a fixed-level thresholding was applied to the grayscale image using the *cv2.threshold()* command, in which all of the pixel values that are smaller than the threshold (0 pixels in this case), are set to 0. Otherwise, they are set to a maximum value, e.g., 255 pixels, as shown in Figure A4. Subsequently, using the *cv2.getStructuringElement()* command, a structuring element of specified size and shape was attained which was later passed to *cv2.morphologyEx()* argument for advanced morphological operations, e.g., erosion and dilation.

```
# Load image, convert to grayscale
image = cv2.imread('path/name.png')
gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
thresh = cv2.threshold(gray,0,255,cv2.THRESH_OTSU
                       + cv2.THRESH_BINARY)[1]
kernel = cv2.getStructuringElement(cv2.MORPH_RECT, (7,7))
close = cv2.morphologyEx(thresh, cv2.MORPH_CLOSE, kernel,
                         iterations=1)
cv2.imwrite('path/name_1.png',close)
contour = cv2.findContours(close, cv2.RETR_EXTERNAL,
                           cv2.CHAIN_APPROX_SIMPLE)
```

Figure A4: Part of the *OpenCV Contours* image processing tool implemented in the YOLACT++ detector.

In closing, the modified image containing the masked objects was processed by the *cv2.findContours()* command which extracted its contours as explained in Figure A4. In simple terms, a contour is a curve connecting together all of the continuous points along the boundary of a mask applied by the YOLACT++ detector, that happens to have the same colour or pixel intensity. Finally, the *cv2.boundingRect()* command was used to apply a rectangular box around the contour area and measure the height and length of the box in pixels. Finally, using a scaling factor of 40 pixels to 1 cm, the actual size of the detected litter items were predicted in centimetres using the abovementioned methodology.