Marine Debris Segmentation using Capsule Network (SegCaps)

MSc Research Project
MSc Data Analytics

Palash Shinde
Student ID: x19218681

School of Computing
National College of Ireland

Supervisor: Noel Cosgrave

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Palash Shinde		
Student ID:	X19218681		
Programme:	MSc Data Analytics	**Year:**	2020 - 2021
Module:	MSc Research Project		
Lecturer:	Noel Cosgrave		
Submission Due Date:	23/09/2021		
Project Title:	Marine Debris Segmentation using Capsule Network (SegCaps)		

Word Count: 6152 **Page Count: 17**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Palash Shinde
Date:	23/09/2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Marine Debris Segmentation using Capsule Network (SegCaps)

Palash Pravin Shinde
x19218681@student.ncirl.ie

Abstract

Disposal of ever-increasing marine debris is proving to be one of the difficult tasks to perform. Making the environment more sustainable requires the correct recycling and disposal of this ever-increasing garbage. Ocean bodies are being inundated every day by millions of plastic garbage pieces. This is having a significant impact on the aquatic life system. When debris is identified and classified by kind on an instance level in an automated manner, debris can be tracked on a large scale without involving humans. CNN models such as Mask R-CNN, Faster R-CNN is often regarded as a cutting-edge approach for object segmentation in an image, its architectures come with certain limitations. To tackle, the limitations of traditional CNN Networks, this research aims to introduce a new Capsule network-based objection segmentation technique called as SegCaps that has the capabilities to tackle the limitations of traditional CNN-based segmentation techniques in terms of generating accurate object segmentation, while at the same time reducing the network's complexity in terms of depths layer. The SegCaps model introduced in the research is trained on the Trashcan dataset, which has 7,212 object images of 3 main categories, annotated on mask level. These main categories are further divided into 16 subcategories classes. The model in the proposed research achieved an overall mean average precision of (mAP) of 26.25 & dice coefficient of 28.46 with an IOU threshold of 0.5 for the segmentation task.

Keywords— Instance Segmentation, Marine Debris, Capsule Networks

1 Introduction

Humankind is heavily involved in polluting the environment, generally since the introduction of new technology followed by the industrial waste from the Industrial Revolution. Almost every type of company today pollutes the environment in some way, or the other particularly non-biodegradable waste generated by humans, such as plastics and heavy metals, which take years to decompose and so pollute the environment for a longer length of time. This kind of waste usually ends up staying inside the water bodies for a larger time frame is generally termed as marine debris. The waste can act as a major threat to water bodies and various aquatic, marine animals present inside the ocean bodies. As the world progress to new heights in the modern era, the majority of water bodies will end up gathering this waste on a larger scale which eventually will lead to contamination and disturbs the entire ecosystem. Different government organizations and institutions had initiated approaches to look after this issue, however, they face a major issue in detecting this debris precisely and overall removal of these debris post identification of debris location. Many government and public institutions have made (Hardesty et al. (2015))an attempt to tackle the issue at hand

with the viewpoint of integrating ecological oceanographic models to measure the risk of wildlife and conducting the analysis of contamination indicators associated with the wildlife and marine bodies centuries. Out of all the marine debris which are present, usually, human-made objects such as plastics contribute to overall debris on a larger scale (Abalansa et al., 2020). This debris has a source which occurs from activities such as fishing in which it is regular use of est to capture the marine bodies and while performing these activities fishermen end up leaving these nets in the water bodies whenever it gets stuck.

The other source of marine debris that can be entered into water bodies is via natural phenomenon such as Tsunamis, Tides, Earthquakes (Mori et al., 2011). Furthermore, bad waste management practices (Willis et al., 2017). can also be a major source of marine debris as plastics bags can get carried away by the windy nature of the weather and landing up inside drains that are not properly closed, thus from that going into ocean bodies. Among all debris, plastic (Li et al. (2016))is the most common type of debris which mainly contributes to the overall percentage of the total debris present in nature.

Fig 1: Marine Debris Images[1]

This is mainly 90% of the time in terms of contribution is concerned. Major effects have been undertaken on the removal of this debris from ocean bodies. There is a need of the hour to develop an automated mechanism to detect this debris in a precise manner. This gave rise to autonomous underwater vehicles (AUVs) (Valdenegro-Toro., 2017), these vehicles are usually powered by a robust algorithm that is mainly cable of detecting these images in a precise automated fashion making them ideal for the task at hand. The main architecture which fuels this kind of network is powered by CNN models. CNN-based architectures are typically used to recognize trash as they are traditionally proved to, and in the past, several efforts were made to implement the most powerful object-detection algorithms on marine waste data.

There are some limitations such as preserving the hierarchy, the orientation of the object in

[1] https://conservancy.umn.edu/handle/11299/214865

terms of CNN's performing object detection and segmenting (Minaee et al., 2021) the object correctly for a given image. Previous researcher has (Fulton et al., 2019) has utilized MaskRCNN, FRCNN and obtained the accuracy of 70-81% for identifying marine waste. However, the research (Mukhometzianov et al., 2018) has shown Capsule network can provide promising results in terms of classifying objects more accurately than traditional CNN's. Considering the added advantages of Capsule network architecture this research will mainly focus on implementing Segmentation-based Capsule networks (LaLonde et al., 2018) on the Marine debris images dataset with the main aim of performing the task of accuratly segmenting the waste objects in marine debris images.

1.2 Research Objectives

1) Implement and train SegCaps network to accurately perform segmentation of Marine debris images.

2) Make use of a pre-trained model trained on these images to calculate performance evaluation of the SegCaps network

Section 2 of the report highlights the literature review of all the related work for the task at hand. Section 3 of the report mainly covers the Methodology followed for the model and dataset. Furthermore, Section 4 & 5 highlights the Design Specification and Implementations of the model. In addition to this Section 6 describes the Evaluation & metrics which are gathered to assess the model performance compared to the SOTA (state-of-the-art) network metrics. Lastly, section 7 highlights the insights and collusion of the proposed research and gives a glimpse about possible future work for the analysis as well.

2 Literature Review

This section provides a summary of prior research activities on the subject issue. Several studies have been conducted regarding the relevance of Marine Debris and its accumulation over time. In addition, the studies stress the significance of automating the removal process to ensure the preservation of our marine organisms and the long-term reduction of environmental contamination. The main objective in this part is to analyze prior work in the many fields of increased marine waste and to assess how robotic marine waste detection may be utilized to remove marine debris by using sophisticated debris segmentation algorithms and algorithms to identify.

2.1 Seafloor marine litter detection using deep learning

Seafloor litter objects are being monitored by aerial and underwater imaging. However, the necessity for automated, economic litter monitoring techniques is obvious from the amount of human labor required for litter monitoring. Researchers (Politikos et al., 2021)have used region-based CNN's to detect seafloor marine debris using an automated object detection approach. The neural network was trained before it was tested with an independent data set, using training pictures noted for 11 litter categories. In the independent data set, the neural network obtained an accuracy of 62%. The researcher also found that one of the sizes of the training item or its particular form alone cannot be ascribed to a dependable prediction of the litter size. Rather, each litter category performed uniquely.

In contrast, the group "tires" had much fewer datasets, but its AP was high, because of its unique shape. 'plastic caps' are a group of little objects only and AP = 0,8 In addition to this whilst 'small plastic sheets' are similarly small and have an AP = 0,5. In the network photos, however, average precision of 0.8 has been achieved in the "bags," class which were rather diverse in shape. Likewise, researchers found that successful litter predictions (AP = 1.0) and failed litter predictions (AP = 0.0) for literal items on the test set surface were of similar

mask size. Consequently, they were able to demonstrate that trash objects' size had no impact on the performance of the network.

Based on the network's mispredictions, the dataset appears to be diverse. A variety of litter forms in different classes had a distinctive geometric design but identical forms in different classes had the same design as litter items of the same class. The researcher had several setbacks, resulting in less predictability, during the testing process. They observed relatively unstable tuning of hyperparameters. Due to restricted computational resources, however, it was not possible to test higher lots although network performance improved as batch sizes grew from 1 to 4. (the default value in Mask R-CNN). The network proved significantly easier to detect various litter kinds (plastic bags, fishing networks, tires, and plastic caps), with a mean precise >79%. Due to various backdrop characteristics (e.g. algae, seagrass, spreading rocks), a higher number of litter pieces than those detected were noticed in the images. The findings of the study were quite optimistic, showing that deep learning can be a helpful ability to determine seafloor trash immediately during surveys and to provide ongoing and precise observation.

2.2 Monitoring of River Plastic using Deep Learning

To minimize the environmental implications of pollution from plastic, the measurement of plastic contamination in the surface water is necessary. Visual counting is currently used as one of the demanding ways of monitoring. There are numerous places to monitor which cannot be escalated for a long period. Researcher (Lieshout et al., 2020) suggested automated technique manually overcomes plastic pollution control. Images of the sea surface are used to identify floating macroplastics using deep learning. Five distinct river sites around Jakarta in Indonesia are used to evaluate our technique using pictures from the bridge camera. Experimental evaluation of the method is performed. Images of the sea surface are used to identify floating macroplastics using deep learning. Five distinct river sites around Jakarta in Indonesia are used to evaluate our technique using pictures from the bridge camera.

Experimental evaluation of the method is performed. Here are the four main experimental test results. As a starting point, they employ an accurate plastic thickness assessment (68.7 percent precision). By separating plastics from environmental factors, such as rainwater and organic waste, our monitoring approach is successful at reducing litter. The second advantage is that if the approach is mapped in one place, it is generally available in new places under comparable conditions without retraining and giving out up to ~50% precision. Furthermore, the performance at the new location would improve precision in the range of 42% for just 50 items (retraining on only 50 objects will boost precision to 42 percent. Furthermore, the methodology in research is 35% more in identifying plastic more than visual numbering. Even if plastic movement speeds are more than 10 items per meter per minute. The researcher states that their plastic pollution management approach is an excellent way of controlling plastic contamination. By extending the number and diversity of available data sets this approach may be employed on a much larger scale. The researcher states that there is room for enhancement of our monitoring technique in three areas: data set, sensor (video camera), and detecting algorithms. In addition to this, they also stated and identify two major development opportunities: increase the quantity and diversity of labels and increase the accuracy of human etiquette. When monitoring locations are increased and diversified, a representative set of examples can be trained on to improve performance. Fine tweaking requires a great deal of calculation, though. Rather than requiring precise adjustment, a comprehensive monitoring approach would be preferable.

Future studies to increase the amount and diversity of data gathered can prove the viability of this general surveillance approach. In pictures, including plastics, a second step can be taken to improve the dependability of human selection. Volunteers can differ substantially in the aptitude to distinguish plastics from their biological substance and surroundings. As

plastic items can become embarrassed, separation from the backdrop is often difficult (including organics). Several issues were identified when volunteers checked and rectified the designated locations. Two modifications are recommended to improve the identification and segmentation of monitoring. In terms of its constituent CNNs, both phases have been enhanced. To construct the monitoring technique, they coupled the RCNN for segmentation with our first V2 CNN for classification. Further improvement concerns in data increase. Horizontal and vertical flipping has shown efficient ways of data enlargement that are compatible with previous studies that show that enrichment of data enhances performance. Other methods, like picture distortion, might be explored by future studies.

2.3 Deep learning-based Marine litter detection in Aerial images

Marine and marine wildlife ecosystems constantly face anthropogenic, floating, macro-litter problems (FMML). Globally, the issue is handled through monitoring programs, mitigating actions, greater support for technology, and an automated analytical process. In this work, the Researchers (Garcia-Garin et al., 2021) have developed FMML detection and quantification algorithms in aerial pictures and a web-based tool for the identification of FMML in seabed photographs. A profound learning technique is presented employing coevolutionary neural networks that may be taught by an unstructured or unlabelled input profound learning algorithm. On 3723 aerial photos obtained in the Northwest Mediterranean by planes and drones, researchers trained and tested a CNN deep learning model (50 percent containing FMML, 50 percent without FMML). CNN algorithms from already existing ones have been created to minimize the time it will take different scientists to perform FMM L in photos taken from aerial platforms (e.g., drones, aircraft). Density, dispersion, and trends change routinely followed by FMML are heavily supported by national and international legislation, so improving and efficient results is very beneficial.

A total of 3.5K pictures were examined (1800 marine litter images, 1860 without marine litter image). This model was trained and tested during cross-validity of 90 percent with training photographs and 10 percent with test images. Researcher findings support the utilization of airborne sensors in the seabed to monitor and identify FMML. FMML surveillance investigations, based on these techniques, have risen significantly in the previous 10 years. Performing the analysis, the model accuracy was 0.85 and 0.81 (used for training and models) during the grading and cross-validation procedure (using 90 percent of the images for training and 10 percent for testing). Furthermore, the efficiency, feasibility, or replacement of existing marine surveillance approaches with automatic FMML monitoring devices would produce superior quality results. Autonomous submarine vehicles (AUVs) can assist solve this problem very efficiently through the detection and disposal of trash.

2.5 Identifying Marine Litter using Deep visual models

Water pollution can cause ecological, economic, and environmental long-term issues in aquatic regions. The study addressed many ways to fully identify visual waste utilizing AUVs for inspection, mapping, and trash recovery in realistic underwater settings. Many convergent neural network architectures for online items have been trained in open water with actual rubbish in this dataset. The network training is then evaluated on a range of images from various areas of the dataset, providing insight into possible methods to improve an AUV's capacity to detect and remove underwater rubbish. To see if these algorithms could be utilized in real-time applications, they were tested on three distinct platforms, each with a different processing capability. This analysis (Fulton et al., 2019) examines several deep learning-based visual object recognition algorithms, builds a dataset to train and test them on, and assesses how well they perform using various metrics.

The goal of this study is to determine if real-time deep learning-based visual detection of underwater trash is viable, as well as how well existing methods perform. The various model which is implemented in this research is Yolov2, TinyYolo, F-RCNN, and SSD. The researcher found that the mAP of YOLOv2 and Tiny-YOLO is lower than that of Faster R-CNN and Solid State Drive. In contrast, faster R-CNNs and SSDs have longer processing times. These characteristics are well-known, but they are also consistent with previous studies on the issue. The trade-off between mAP and FPS does not affect IoU. In terms of how precise their bounding boxes are, none of the four network architectures scored higher on the IoU scale.

2.6 Unmanned aerial system (UAS) for debris detection

Environmental monitoring equipment incorporating unmanned air vehicles (UAVs) is becoming more readily available. A recent emerging field, anthropogenic aquatic debris studies, might be influenced by this technology. The gathering of geo-RGB pictures in a protected maritime zone resulted in an intensive program of (ten months) surveillance utilizing unmanned aerial vehicles (Migliarino, Massacciuccoli, San Rossore Park near Pisa, Italy). Using a visible post-treatment approach for photos, anthropogenic marine garbage may be found and identified in the scanned area and its geographical and temporal distribution analyzed. With cameras and sensors, smaller unmanned aircraft or drones can fly remotely or independently, weighing less than 25 kg. There has been a rising demand (Itkin et al., 2016) for military, civil, hobby, and academic applications. Science utilizes unmanaged aircraft systems to gather data in a range of disciplines including agriculture, animal behavior, population growth, and the production of coastline maps.

With UAS applications, coastal and maritime surveys can be performed with measured measurements to identify coral remnants at sea and seal. When it comes to coastal maps, vertical aerial photographs are by far the most popular source of data. Aerial pictures, however, feature huge areas and fewer details, so distortion and transportation are greater. UAVs have been very available in recent years and are employed in a wide range of areas, including agriculture, structure, archaeology, aquatic, and emergency response. UAVs have become very accessible (Manfreda et al., 2018)

3 Research Methodology

Knowledge Discovery in Databases (KDD) methodology is followed in the project analysis. As stated in the objective, the goal of this study is to determine if a Capsule-based object segmentation network outperforms a Mask RCNN architecture for segmenting marine garbage images. Data selection and extraction, data cleansing/transformation, modeling fitting, and model assessment are the four steps of an analysis project's structure. Fig2 below describes the overall steps performed in methodology. Dataset for the research project is sourced from Trashcan (Hong et al., 2020) dataset

3.1 Business Understanding

When things made by humans reach the ocean, a substantial part of them become immersed in the waves, where they linger for prolonged periods, contaminating bodies and disturbing aquatic ecosystems. Processing materials contribute a considerable number of marine debris to the ocean. Marine debris may enter bodies of water through a variety of routes, including leftover fishing nets by fishermen. Plastic is the major category almost 90% which contributes to all marine debris in the ocean.

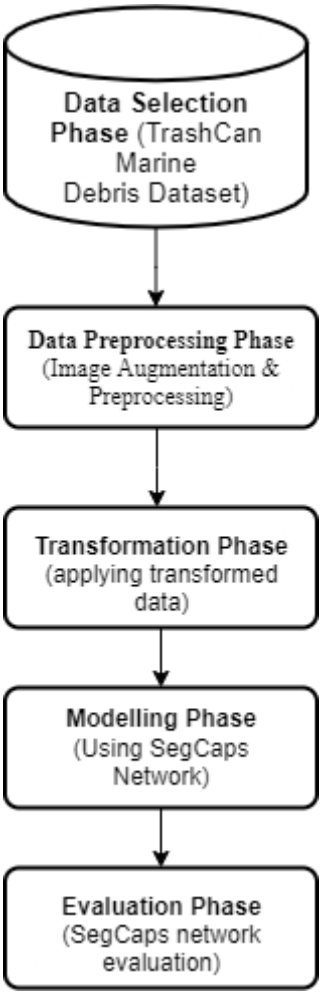

Fig 2: KDD process flow for the project

The debris has been cleaned from the water bodies with success. However, the removal of debris is time-consuming and involves a lot of human intervention in general. It is important to determine the specific position of the debris, and automated procedures should be developed to do so.

As an automated solution, autonomous underwater vehicles (AUVs) were developed. In many cases, AUVs are driven by sophisticated algorithms that determine these images in a precise, automated manner, making them ideal for this task. Efforts have been made to build a system that is not only cost-effective but also highly available and helps to detect this debris more precisely and accurately. This study mainly focuses on achieving this by suggesting an alternate Capsule network-based approach for Marine debris images segmentation compared to currently using CNN's based approach for the task at hand.

3.2 Data Understanding

Dataset for the research project is sourced from Trashcan (Hong et al., 2020) dataset. TrashCan is a collection of descriptive pictures (7,212 total) that show close-ups of rubbish and ROVs, as well as biological studies of underwater life. In this dataset, the dataset annotations are presented as instance segmentation annotations. These annotations are bitmaps with masks specifying which picture pixels belong to which dataset item. The data in TrashCan comes from J-EDI (JAMSTEC E-library of Deep-sea Images; JAMSTEC E-library of Deep-sea Images). Since 1982, JAMSTEC has operated ROVs, mostly in the Japanese Sea, and has contributed film for these datasets.

Since 1982, JAMSTEC has operated ROVs, mostly in the Japanese Sea, and has contributed film for these datasets. Researchers (Hong et al., 2020) got all of the marine debris data from a tiny proportion of these recordings, over a thousand movies of different lengths that show marine debris observations. Additional films were chosen to supplement the biological component of the collection, in addition to the footage of marine trash. Researchers gathered one frame every second from each video after selecting them to create large folders of frames for each film.

The films were then broken up into smaller chunks and uploaded to a program called Supervisely for annotation. After the photos were submitted, twenty-one people annotated each one individually. This project took around 1,500 hours to complete over several months. The person in control would use a segmentation mask to categorize any image deemed suitable for labeling as trash (debris), rov (any object made by humans which are deliberately placed in the scene), bio (plants and animals), or unknown (unknown objects). In addition to the identification of material (metal, plastic, instance) and binaries (cup, bag, container) indicating overcrowding, serious degradation, or cracked/broken objects, the waste goods were commonly detected using binary tags.

There was an identification system, with plants labeled and animals tagged with the form (e.g., crabs, fish, eels). In general, ROVs and objects unknown in class did not have any additional tagging. A total of 7,212 pictures are divided further using the usual 80-20% test division procedure. For the SegCaps network training purpose, the categories are a) rov (any object that has been intentionally placed in the scene), b) unknown (used to indicate unknown items), c) bio (plants and animals) d) trash (any type of marine waste). These major four classes are further divided into 16 subclasses. As the task at hand is to train the SegCaps model to perform segmentation on debris images, is it equally important to make sure all types of subclasses are present in the training and validation set to avoid an issue of unbalanced class and model getting overfit for a particular class.

3.2.1 Image Augmentation

Augmentation of data is widely employed to train image recognition systems while learning deep neural networks. It helps to minimize overfitting (Dvornik et al., 2019) and enhance generalization by artificially increasing the number of training instances. To generalized, the model image augmentation techniques have been applied as well.

3.3 Modeling

This section explains how the SegCaps network is modeled and fitted on the Trashcan dataset along with how data is divided into train tests, sent into networks, and eventually, model performance is evaluated by calculating the metrics on test data. Fig 3 shows a process flow diagram for all the steps which are mainly performed during the modeling stage of the network.

Steps involved in modeling are:

1) Dataset is gathered from source (Hong et al., 2020), and mask images are created from bbox annotations of the object in images unzip the images and apply stain normalization to the images.

2) These mask images act as annotations of the object present in an image for the SegCaps network.

3) Furthermore, random train test split sets are created and it is made sure that each type of class image is present in the set equally to avoid data imbalanced issues.

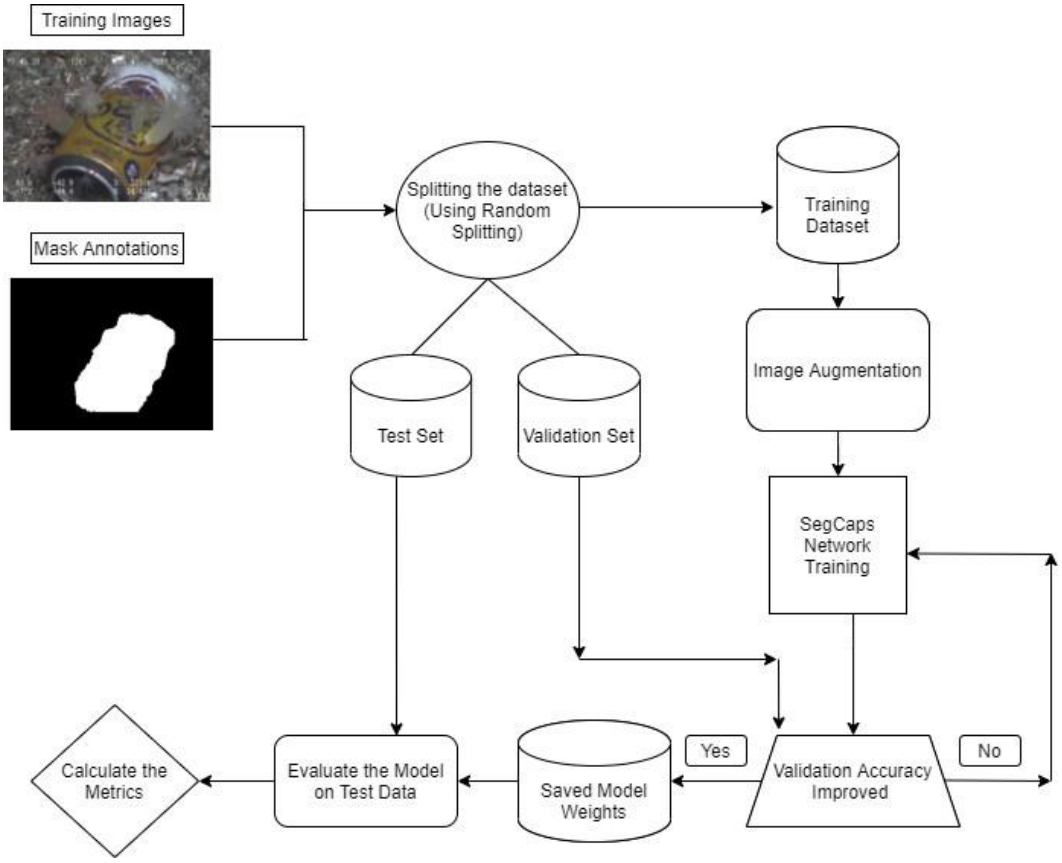

Fig 3: Modelling process flow

Data is further split into training, testing, and validation set using a random selection of each image.

4) In the training phase, data augmentation is applied for the training set images to avoid overfitting of the model, this also helps the model to capture as much variation in data as possible. Model is further validated on the validation set and model weights are saved once validation loss starts to reduce at the same time validation accuracy is increased.

5) After completion of training of the model, model weights are evaluated on the test set and metrics are calculated to judge model performance.

4 Design Specification

This portion of the research explains the design requirements, which include the Framework and the methodologies used to conduct the object analysis. SegCaps is a technique for segmenting objects in images based on capsule networks proposed architecture in this study (LaLonde et al., 2018).

4.1 Overview of Capsule Network Architecture

In the Capsule Network architecture, each capsule represents a collection of neurons, with an activity vector elucidating a set of parameters, based on the capsule's structure. In terms of probability, the length of the operating entity that exists is mainly highlighted by the vectors, and in terms of instantiation parameters, it represents its orientation. The multi-layercapsule-oriented network has been proven to surpass the accuracy of the CNN's when performed on MNIST classification tasks (Sabour et al.,2017) The author relied heavily on routing techniques to get these results. The main idea behind capsules is to represent data for various characteristics as neural network-based vectors. Unlike CNN's, these networks' features are

contained within capsules, allowing them to detect objects even when viewing angles change.

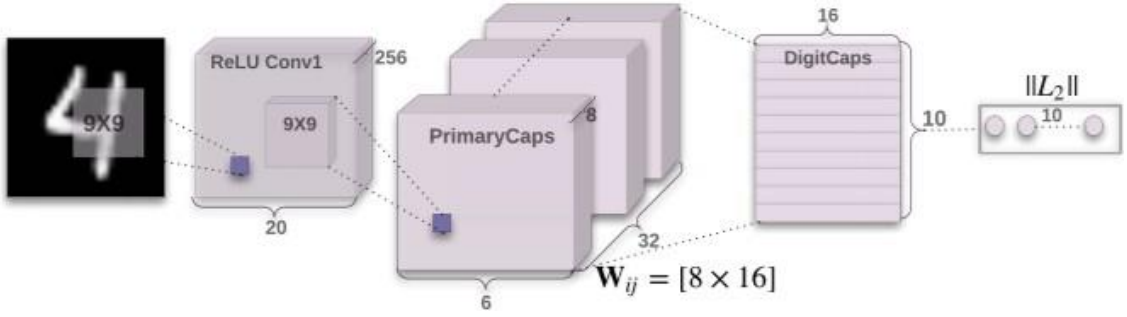

Fig 4: Capsule Network (Sabour et al.,2017)

4.2 Object Segmentation using Capsule Network (SegCaps)

On the MNIST dataset, three-layer capsule networks produced accurate classification results when compared to Resnet, VGG, and different resilient networks. The results for the task of classification results for the CIFAR10 dataset are likewise rather excellent. Using a capsule-based network to identify objects is essentially a modification of the vanilla network that allows the network to classify and find items on an instance-by-instance basis. In a capsule model, there are just two convolutional layers and one fully linked layer in basic words. There are several distinctive features of capsules, which are essentially small groups of neurons. A picture is made up of neurons, each of which represents a different aspect of the image, such as color, breadth, and so on.

Each capsule produces a vector that represents the absolute magnitude (a measure of the probability of a part being present) as well as the trajectory (a measure of the generalized position of a part). The capsule network learns how elements and the entirety of an item are linked in space during training (e.g., how the location of the eyes and the nose contribute to a visual representation of a face). Capsules record linkages between objects in dynamic routing by measuring connections between one layer and the next, resulting in powerful ties between item interiors. There are numerous issues to consider when attempting to segment items using a capsule-based network. Capsule networks and dynamic routing methods are computationally expensive, both in terms of memory and run-time. In one layer, the output of "child" capsules is kept in an extra intermediate representation, while in the second layer, the parent capsules are mainly defined by the coefficient of the dynamic routing algorithm. In SegCaps capsule architecture is widen while at the same time there is some reworking done on the dynamic routing algorithm. This way researcher (LaLonde et al., 2018) was able to tackle the memory load and parameter explosion issue. Children are first just provided a spatially local kernel via which they may contact their parents.

Additionally, capsule types exchange transformation matrices for each member of the grid without requiring capsule forms. In addition, to compensate for the lack of global communication with the proposed locally constrained routing, we propose "deconvolutional" capsule networks that use transposed convolutions and are routed by transposed convolutions. Additionally, capsule types do not require capsule forms to exchange transformation matrices for each grid member. They also propose "deconvolutional" capsule networks, which use transposed convolutions and are routed by transposed convolutions to compensate for the lack of global communication in the proposed locally-constrained routing.

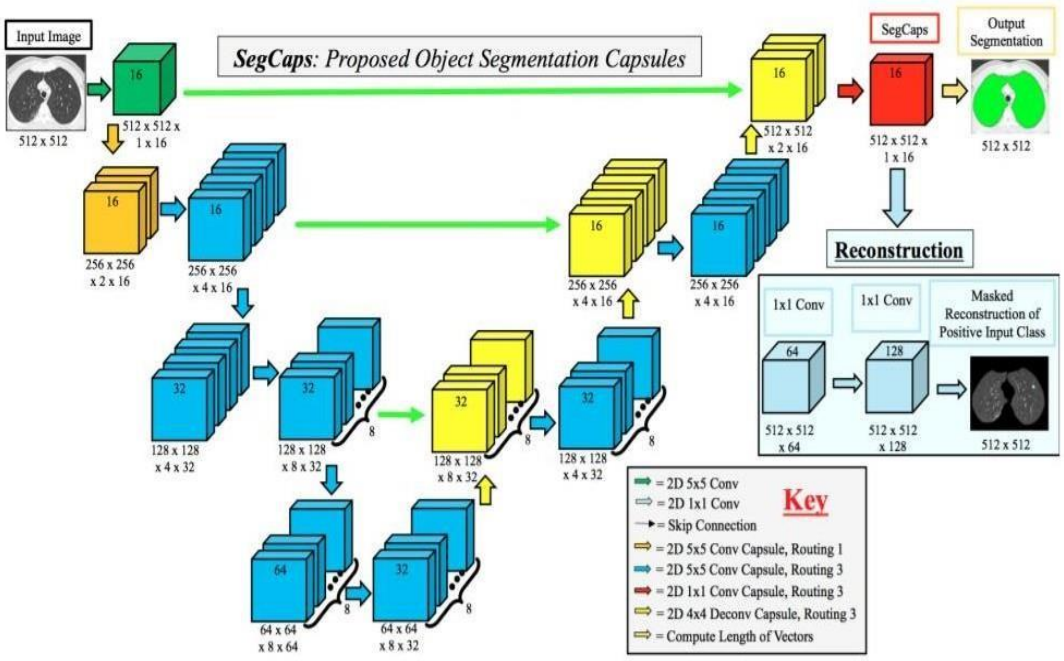

Fig 5: SegCaps Network (LaLonde et al., 2018)

The size of the network might be substantially decreased utilizing the suggested deep convolutional-deconvolutional network architecture, lowering memory requirements and enabling state-of-the-art results for the method. SegCap's suggested design is shown in Fig 5. Two major innovations have enhanced the SegCaps network overall. Children's cells are exclusively routed to their parents inside a spatially limited window, and each member of the capsule-type grid has a defined transformation matrix. These precise changes to the convolution capsules enable the network to process images with a resolution of 512 × 512, whereas a vanilla capsule network could only handle images with a resolution of 32 x 32. SegCaps has created a unique design of convolutional-deconvolutional capsules that goes deeper than our initial three-layer network, in addition to their convolutional-deconvolutional capsule technique.

5 Implementation

Implementation of the project is broken up into many segments, each of which explains the entire procedure. The following sections contain details on how the process outlined in Modelling 3.3 is accomplished.

5.1. Mask Images Generation

Dataset for the research project is sourced from the Trashcan dataset. TrashCan is a collection of descriptive pictures (7,212 total) that show close-ups of rubbish and ROVs, as well as biological studies of underwater life. Furthermore, mask images are generated from the bbox annotations present in the dataset for the training RGB images. These mask images will mainly act as pixel-level annotations for the specific object present in the image. A total 16 subclass which are 'rov', 'plant', 'animal_fish', 'animal_starfish','animal_shells', 'animal_crab', 'animal_eel', 'animal_etc', 'trash_etc', 'trash_fabric', 'trash_fishing_gear', 'trash_metal', 'trash_paper', 'trash_plastic', 'trash_rubber', 'trash_wood' of mask images are created in total , which translates to 7212 images.

5.2 Data splitting

Data is split into train, validation, and test set. These sets are further used for the model training

and evaluation phase. Standard sklearn train, test, split module is used to randomly split the data in training, validation, and test set. Data is mainly split into 80:20 in train and test set. The train set is further divided and 20% of the data is sliced into a validation set as well. For each split, it is made sure that the divisions of data are random and there is no selection bias in shortlisting the data into their particular sets.

5.3 Training

The training stages for the SegCaps Capsule network are discussed as follows:

1. In SegCaps Network training phase, the rbg images and the images of the binary mask which are annotations of the object on the pixel level are loaded into NumPy arrays, and training and validation data generator objects are created.

2. Training is performed from scratch and no pre-trained model is being used.

3. The image generator object is called to create a training and validation generator object, which will be further used during the training phase. Model is mainly trained with the batch size of 1 and on 3000 epochs.

4. To avoid model overfitting and make model capture variations of objects in images. Image augmentation is applied to images on the fly while training.

5. Optimizer for the model is set to Adam, while a loss function for the model is considered as dice loss.

6. The ModelCheckpoint, an implementation of CSVLogger, ReduceLROnPlateau is used for recording epoch output, and saving weights if the loss/accuracy improves, and reducing the learning rate for the model when the metric for the model during the training phase stops improving.

6 Evaluation

To assess a model's performance, different evaluation measures are required. As segmenting images is the task of checking whether or not each pixel belongs to a particular class or not, finalizing the correct metric for the task is as important as getting accurate pixel level mask images from the model. Intersection over Union (Jaccard index) and Dice coefficient (Porikli et al., 2020) are ideal metrics to judge the model performance for the Image segmentation task. SegCaps model performance is evaluated on test data using these two metrics mainly. A total of roughly 100 images are considered as test sets which are randomly chosen from the dataset during the data splitting stage. Out of all test images, Fig 6 below shows the segmented mask output predicted by the SegCaps model for the few test images.

Fig 6: Marin Debris Images (left), SegCaps model predictions (right)

A) Intersection Over Union:- An intersection of two sets determined by their union is the Jaccard index, also called the I+U score (Intersection over Union). A set of values may be thought of as an image mask. It's conceivable that the sets in the image will overlap. If the masks are similar, they will always overlap to 100%. As a result, their intersection will equal their union, and their size will be the same. In this scenario, the score is 1 and the IoU is optimum. The union is larger than the intersection if the anticipated mask differs from the original mask in size or location. The IoU score suffers as a result. While evaluating the SegCaps model on test images, IOU for each image ground truth mask and model predicted mask is calculated and the score is recorded.

B) Mean Average Precision:- As the IOU score for each is calculated, a threshold value of 0.5 (Padilla, Rafael & Netto, Sergio & da Silva, Eduardo) for the IOU score is set to conclude whether the prediction was true positive if IOU > 0.5 and False Positive if IOU < 0.5. This means that if there is more than 50% of the overlap between ground truth and model predicted mask, prediction is being flagged as true positive else if the overlap between these two is less than 0.5 it is flagged as false positive.

SegCaps model precision score was calculated on each 100 test images and later on, each of these scores is averaged together to get a single mean average precision value for the model. SegCaps model achieved an *mAP* of *26.25* roughly. As precision is a measure for all model predictions how many of them are true positive, mAP score reflects there is a roughly 26.25% chance of model predictions being a true positive.

MaskRCNN model achieved *mAP* of 55.3 (Hong et al., 2020) on Trashcan dataset vs SegCaps model achieved an mAP of 26.25. As Mask RCNN instance segmentation architecture is deeper in terms of layers and architecture The lower score of the mAP is observed for the SegCaps model compared to Mask RCNN.

Table 1: Model Metrics Comparison

Model	**Dataset**	***mAP @ 0.5***
Mask RCNN	Trashcan	55.33
SegCaps	Trashcan	26.25

3) Dice coefficient:- To compare the performance of different image segmentation methods, the dice score is often used. Calculating the Dice score, a measure of how similar objects are is usually used to validate a segmentation algorithm. The overlap between two segments is equal to the difference between their total sizes. It is the measure of spatial overlap for ground

truth images and predicted mask images. During the SegCaps model evaluation phase on test images, the dice coefficient for each test image is calculated, the average dice coefficient for the model is 28.25.

7 Conclusion, Discussion, and Future Work:

Accurately and precisely detecting marine debris in an automated fashion is of utmost importance given the fact that pollution and quantity associated with the debris are increasing day by day. Deep learning can act as a robust solution to help fuel the underwater drone vehicles which help in detecting this debris with less human intervention. Thus, there is a need of the hour to help these drone vehicles identify the debris precisely. Traditional CNN's object detection and segmentation techniques can have limitations in performing the task, hence the SegCaps network was introduced in this research which is further trained on Trashcan images and can act as a robust network in the task of accurately segmenting marine debris.

From the above evaluation section, the SegCaps model performance can be compared with other segmentation models in terms of metrics such as dice coefficient and *mAP*. Although the *mAP* score for the model which was achieved by running the model on test images was 26.25, the SegCaps model holds potential in terms of using the model for the task of image segmentation. The state-of-the-art model mAP result which is published by the author (Hong et al., 2020) is 55.33. Although the SegCaps model achieved an *mAP* of around 26.25, these metrics can be considered as good given the constraint environment in which the model was trained. Due to the computationally expensive process of training the SegCaps model for more interactions, the model was trained on almost $1/5^{th}$ of interactions compared to the SOTA MaskRCNN model training phase. In addition to this in terms of future work for the researcher is concerned, the SegCaps model can be trained on more iterations, different data augmentation techniques with better hardware which gives the potential to model to increase the overall precision score on test data and get accurate mask predictions.

References

Abalansa, S. *et al.* (2020) "The marine plastic litter issue: A social-economic analysis," *Sustainability (Switzerland)*, 12(20), pp. 1–27. doi: 10.3390/su12208677.

Hardesty, B. D., Good, T. P. and Wilcox, C. (2015) "Novel methods, new results and science-based solutions to tackle marine debris impacts on wildlife," *Ocean and Coastal Management*, 115, pp. 4–9. doi: 10.1016/j.ocecoaman.2015.04.004.

Li, W. C., Tse, H. F. and Fok, L. (2016) "Plastic waste in the marine environment: A review of sources, occurrence and effects," *Science of the Total Environment*. Elsevier B.V., pp. 333–349. doi: 10.1016/j.scitotenv.2016.05.084.

Mori, N. *et al.* (2011) "Survey of 2011 Tohoku earthquake tsunami inundation and run-up," *Geophysical Research Letters*. Blackwell Publishing Ltd. doi: 10.1029/2011GL049210.

Mukhometzianov, R. and Carrillo, J. (no date) *CapsNet comparative performance evaluation for image classification*.

Valdenegro-Toro, M. (2017) "Submerged marine debris detection with autonomous underwater vehicles," in *International Conference on Robotics and Automation for Humanitarian Applications, RAHA 2016 - Conference Proceedings*. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/RAHA.2016.7931907.

Willis, K. *et al.* (2017) “Differentiating littering, urban runoff and marine transport as sources of marine debris in coastal and estuarine environments,” *Scientific Reports*, 7. doi: 10.1038/srep44479.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N. and Terzopoulos, D., 2021. Image Segmentation Using Deep Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.1-1.

Fulton, M., Hong, J., Islam, M. and Sattar, J., 2019. Robotic Detection of Marine Litter Using Deep Visual Detection Models. *2019 International Conference on Robotics and Automation (ICRA)*,.

Mukhometzianov, R. and Carrillo, J. (2018) *CapsNet comparative performance evaluation for image classification*.

Valdenegro-Toro, M. (2017) “Submerged marine debris detection with autonomous underwater vehicles,” in *International Conference on Robotics and Automation for Humanitarian Applications, RAHA 2016 - Conference Proceedings*. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/RAHA.2016.7931907.

LaLonde, R. and Bagci, U. (2018) “Capsules for Object Segmentation.” Available at: http://arxiv.org/abs/1804.04241.

Politikos, D. v. *et al.* (2021) “Automatic detection of seafloor marine litter using towed camera images and deep learning,” *Marine Pollution Bulletin*, 164. doi: 10.1016/j.marpolbul.2021.111974.

Lieshout, C., Oeveren, K., Emmerik, T. and Postma, E., 2020. Automated River Plastic Monitoring Using Deep Learning and Cameras. *Earth and Space Science*, 7(8).

Garcia-Garin, O., Monleón-Getino, T., López-Brosa, P., Borrell, A., Aguilar, A., Borja-Robalino, R., Cardona, L. and Vighi, M., 2021. Automatic detection and quantification of floating marine macro-litter in aerial images: Introducing a novel deep learning approach connected to a web application in R. *Environmental Pollution*, 273, p.116490.

Itkin, M., Kim, M. and Park, Y. (2016) “Development of cloud-based UAV monitoring and management system,” *Sensors (Switzerland)*, 16(11). doi: 10.3390/s16111913.

Hong, J., Fulton, M. and Sattar, J. (2020) “TrashCan: A Semantically-Segmented Dataset towards Visual Detection of Marine Debris.” Available at: http://arxiv.org/abs/2007.08097.

Manfreda, S. *et al.* (2018) “On the Use of Unmanned Aerial Systems for Environmental Monitoring,” *REMOTE SENSING*, 10(4), pp. 1–7. doi: 10.20944/preprints201803.0097.v1.

Dvornik et al. (2019) "On the Importance of Visual Context for Data Augmentation in Scene Understanding," in *Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 2014-2028, 1 June 2021, doi: 10.1109/TPAMI.2019.2961896.

Porikli, F. *et al.* (2020) *Image Segmentation Using Deep Learning: A Survey*. Available at: https://www.researchgate.net/publication/338644066.