A Deep Learning Application for Real-time Debris Detection: Underwater Environment

by

Afsana Kabir Sinthia
23166004

This thesis is submitted to the CSE Department
in partial fulfillment of the requirements for
Master of Science in Computer Science and Engineering degree

Department of Computer Science and Engineering
Brac University
December 2023

Thesis Declaration

I declare that

1. This original work is mine, produced while studying here.

2. It refrains from using prior published and/or written external materials, apart from what are cited using correct referencing.

3. It does not contain submitted or accepted materials at other institutions for any programs.

4. All sources of help have been acknowledged.

Full Name & Signature of the Student:

Afsana Kabir Sinthia, 23166004

Approval

The thesis titled "A Deep Learning Application for Real-time Debris Detection: Underwater Environment" submitted by Afsana Kabir Sinthia (23166004) of the semester Spring 2023 is accepted as part of the requirement for the M.Sc. in CSE degree in December 2023 as satisfactory.
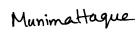
Examination Committee Members:
Supervisor:

aar

Annajiat Alim Rasel, Sr. Lecturer
CSE Department, BracU

Co-Supervisor:

Munima Haque

Dr. Munima Haque, PhD, Assoc. Prof.
Math. and Natural Sci. Department, BracU

Internal Faculty Member:

Aniqua

Aniqua Nusrat Zereen, PhD, Asst. Prof.
CSE Department, BracU

External Faculty Member:

Mohammad Zahidur Rahman, PhD, Prof.
CSE Department, Jahangirnagar Unversity

Program Co-ordinator:

Sadek

Dr. Md Sadek Ferdous, PhD, Assoc. Prof.
CSE Department, BracU

Department Head (Chair):

Kazi

Sadia Hamid Kazi, PhD, Assoc. Prof.
CSE Department, BracU

Abstract

Recognizing the significance of accurately measuring and removing underwater waste is vital for safeguarding marine ecosystems and the environment. Measuring underwater waste is challenging due to factors like light reflection, absorption, dispersed particulates, and color distortion. Detecting and measuring floating and surface waste is comparatively straightforward. The presence of marine waste is detrimental to both the environment and human health, as microplastics from decomposed waste can enter the food chain. In light of current circumstances, addressing water contamination is crucial for environmental preservation. A significant concern in today's society is the contamination of water bodies. The absence of standardized benchmarks and data standards poses challenges in comparing research efforts related to automatic waste identification in underwater environments. This article tackles the issues of identifying underwater waste or debris by thoroughly examining existing publicly available underwater waste datasets and evaluating Deep Learning-based waste detection algorithms for underwater environments. Image processing, deep learning, and trawling hold promise in implementing effective solutions. Examination of publicly available datasets in this study can support future research efforts to protect our ecosystem. It consolidates prior research, presenting the results of tests conducted on the provided datasets, aiming to establish a reproducible benchmark for waste detection using YOLOv8 as well as classify the garbage using transformers (ViT and Swin) and transfer learning (DenseNet, VGG16 ResNet and InceptionV3). Used ICRA19 dataset encompasses a range of categories of waste, including bio, plastic, and ROV. On the other hand, we used the Forward Looking Sonar Image (FLS) Marine Debris Dataset having 10 Debris categories. The technique of this study achieves a maximum average accuracy which 92.2%, indicating successful waste detection and identification in underwater settings.

Keywords: Convolutional Neural Network, Underwater Waste classification, Underwater Waste Detection, YOLO, YOLOv8

Acknowledgement

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The dominance of machine learning, especially deep learning for object recognition and classification, has emerged in the preceding time frame. This section illustrates the context of garbage's impact on the environment, as well as the recent work and technology that has been done to identify garbage, which helps to lessen the negative effects.

1.1   Classification of Underwater Garbage

The significant growth of industrialization, urbanization, and migration to urban areas has emerged as a pressing issue due to the world's substantial increase in garbage production. Various types of trash, such as solid and liquid garbage polluted with chemicals, food garbage, and agricultural garbage, are commonly disposed of in open landfills in urban areas. This improper garbage disposal practice has resulted in significant environmental deterioration. The act of disposing of rubbish in an unselective manner in public areas serves as evidence of the inadequate garbage management proficiency possessed by the municipal authority. Soil, air, and water pollution are prevalent occurrences in the majority of cities in the world. The primary obstacle lies in the significant limitation of available land resources. The Capital city of Dhaka exhibits a high population density, resulting in the generation of around, 7,000 metric tons of garbage on a daily basis [1]. Garbage can be described as waste, trash, and debris as well.

Residents' level of awareness of garbage classification remains rather poor, with a significant portion of the population lacking understanding and clear guidelines for proper rubbish categorization. In the contemporary era, characterized by widespread production and consumption practices, the management of garbage has emerged as a significant global concern [2]. Garbage classification is a fundamental and labor-intensive process in garbage management, involving the categorization of garbage materials into distinct groups such as glass, paper, cardboard, plastic, and metal. This classification determines the proportion of recyclable trash that may be recycled and has a substantial impact on the effectiveness of following garbage treatment procedures. Automated garbage classification approaches have the potential to aid in the resolution of these difficulties. The inaugural implementation of a garbage classification system was initiated by China in Shanghai [3]. Nevertheless, a significant issue arises with the automated garbage classification method, as it fails to appropriately classify the photos of rubbish [4]. This study proposes a new method

for categorizing garbage by utilizing transfer learning and model fusion techniques in an image recognition system. Deep learning depends largely on data availability. However, it should be noted that different countries have different rules regarding domestic garbage classification, resulting in a shortage of appropriate datasets, particularly in China and on an international level. There are many deep learning algorithms that excel at garbage categorization. However, garbage's characteristics make garbage categorization difficult. For instance, nutshells are dry and moist garbage. Wet garbage is chestnuts, and dry garbage is walnut shells. However, the two nutshell classifications are very similar. During the trial, the two nutshell kinds were quite similar, making classification difficult. However, Altin et al.[5] suggest that this technique lacks a strong theoretical foundation and fails to fully use contextual information. Convolution layers could solve this problem. However, they would increase computing needs. It would also make it difficult for the model to converge during training, which defeats the objective of Convolutional Neural Networks (CNNs) [6]. The CNN method is 95% accurate. However, global information constraints have slowed rubbish categorization, causing a bottleneck. Given the massive global rubbish generation, even a small improvement would have significant economic and environmental benefits. Thus, CNN can be improved for garbage categorization accuracy.

1.2 Underwater Garbage Detection

Trash management has emerged as a matter of growing significance, mostly driven by the escalating buildup of refuse in various natural habitats, notably marine ecosystems encompassing seas and oceans. Annually, a substantial quantity of plastic debris, ranging from 1.15 to 2.41 million tonnes, infiltrates the ocean through river systems. Aquatic ecosystems, including oceans, rivers, and lakes, are invaluable components of our planet, providing vital resources and hosting diverse ecosystems [7]. However, these environments face an escalating threat in the form of submerged debris, ranging from discarded plastics to abandoned fishing gear and other pollutants [8]. Submerged debris poses a multifaceted challenge by not only compromising water quality, but also endangering marine life through entanglement and ingestion. Meanwhile, the National Oceanic and Atmospheric Administration has indicated that the issue of urban waste and marine debris is progressively emerging as a significant worry. If this pattern persists, the adverse consequences of ongoing environmental contamination will harm the ecosystem. As previously indicated, a notable occurrence of illicit disposal occurs within both marine and urban settings [9]. The most severe potential outcomes encompass ecological extinctions, disturbances in the surrounding ecosystems, and anomalous weather patterns. Consequently, engaging in volunteer-based initiatives for collecting and disposing of trash has emerged as a widely adopted strategy for addressing this issue. Nevertheless, the process of waste collection necessitates substantial time and resources, along with a considerable amount of exertion from the volunteers, due to the limited availability of personnel. Therefore, the current solutions are insufficient [10]. Applications for underwater object identification include estimating marine populations, studying ecosystems, pelagic fisheries, detecting unexploded ordnance, archaeology, and species conservation [11].
These programs facilitate the process of oceanographic mapping, monitoring the
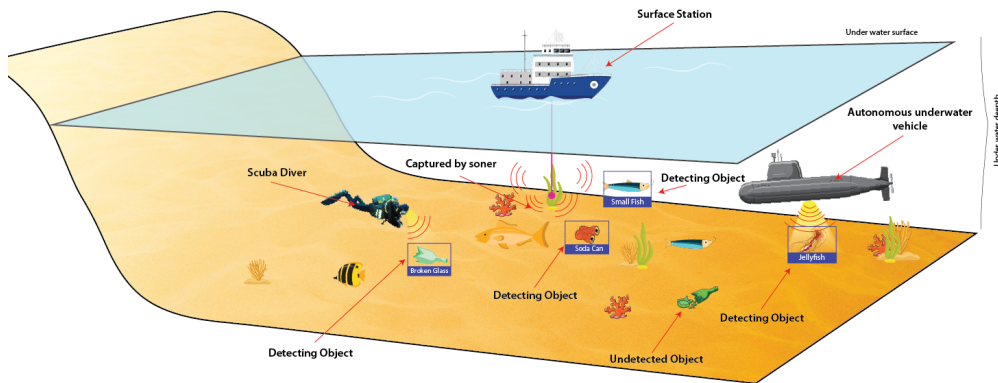
Figure 1.1: Underwater Objects Captured by Sonar/Camera

deep-sea environment, and examining its infrastructure. Object detection of submerged objects has been performed using traditional machine learning algorithms [12]. Gasparovic et al. [13] assert that underwater object detection possesses numerous vital applications, including but not limited to the maintenance and repair of underwater structures, as well as homeland security. Fossum et al. introduced the "Forward-scan sonar" technology as a way to target high noise and low contrast challenges in photographs from underwater. Additionally, they offered machine learning algorithm for detecting and tracking underwater objects [14]. Corrigan et al. [15] note that underwater photographs are important for bathymetry and aquatic applications. They used reference and target images to recognize underwater items using traditional machine learning. The image of the world under the waves that can be seen here was acquired using a sonar or camera system, depicted in Fig 1.1. Ferdous et al. [16] explored the use of a Remotely Operated Vehicle (ROV) for long-term video capture in deep-sea settings.

1.3   Problem Statement

The manual processing of these clips takes time. Recent trash collection research uses robotic automation. Robots analyze their camera feed to identify and detect trash. A trash detecting system designed for trash pickup robots has been developed to handle their different environments [17]. We believe waste-detecting systems can be improved. The increasing amount of trash in natural environments is a growing concern in the modern era. Amount of plastic waste reaching oceans annually via river bodies [10] ranges from 1.15 to 2.41 million in metric tons. For the purpose of marine trash elimination, the deployment of fleets of entirely autonomous or partially diver-operated underwater vehicles is an emerging strategy. However, cleaning debris beneath the surface of the water is difficult and costly. Consequently, a cost-effective solution capable of operating properly and proficiently in various environments is necessary. Recent developments in machine learning, artificial intelligence, and autonomous transportation have enabled intelligent vehicles to remove submerged refuse [11].

1.4 Aims and Objectives

Our strategy is based on an architecture of a Convolutional Neural Network, designed to identify various garbage and to propose a real-time method for identifying trash using a deep learning method titled "You Only Look Once" (YOLO). We used a real dataset of labeled images of garbage as the training data for evaluating the system, and the objective is to devise a technological solution capable of identifying and detecting rubbish objects in recorded video footage. Furthermore, critically review the state of the art of Deep Learning-based detection of waste in water to help pave the way for future research. It may also pave way towards future garbage classification for any country or even Bangladesh. Currently, significant country-specific datasets are yet to be developed for many countries. The following is a summary of this thesis's major contributions:

- Enhance the effectiveness of resource recycling through a vision transformer-based automated garbage classification approach.

- Investigate waste garbage recognition in underwater environments through the YOLOv8 model.

- Conduct experiments with the constructed datasets.

Chapters are organized as: the second chapter focuses on a review of existing work. The third chapter is the foundation of the second chapter and brings up the architecture, utilized convolution layer, employed activation function, etc. The fourth chapter discusses the research implementation along with the distribution of the datasets, and the strategies. The fifth chapter illustrates the results obtained from implemented algorithms, followed by the explanation and outcome analysis. The sixth chapter concludes the research work with a discussion of the potential future work.

Chapter 2

Prior Research

2.1 Prior Research on Underwater Garbage Classification

The utilization of image processing and classification has shown substantial growth in many applications in recent years. As a result, a considerable number of scholars have devised sophisticated methodologies for the purposes of image detection, classification, and production. Liu et al. [18] introduce Garbage Classification Net (GCNet), a garbage image recognition model utilizing transfer-learning and model-fusion techniques. Neural network model of GCNet is constructed by combining EfficientNetv2, Vision Transformer, and DenseNet, after extracting garbage image features. However, the primary challenges associated with the proposed Garbage Classification Net (GCNet) pertain to its intrinsic computational rigor and resource demands. The architecture of the singular model, which integrates DenseNet, Vision Transformer, and EfficientNetv2, is computationally intensive during both the training and deployment phases. Alrayes et al. [19] introduce a novel approach, namely the Vision-Transformer - VT based on a Multi-layer Hybrid Convolutional Neural Network (VT-MLH-CNN), aiming at autonomous garbage classification. The strategy described in this study improves the precision of trash categorization and decreases the duration required for the categorization process. The data images are initially collected, followed by the extraction of features. Subsequently, the data undergoes processing for normalization. The absence of a comparison study against state-of-art models is one significant element that was addressed in the proposed work. Aleem et al. [20] suggested technique classifies it into ten groups. The Faster Region-based Convolutional Neural Network (RCNN) alongwith transfer-learning of ResNet50 is utilized to get over the problem of data scarcity. One of the widely used object detection architectures is Faster-RCNN. It simultaneously makes use of the Regional Proposal Network (RPN) and detector. Gaspar et al. [21], to address the perception issues in harbor facilities, a purely acoustic method using data from Forward-Looking Sonar (FLS) is suggested for unsupervised recognition of comparable images.

2.2 Prior Research on Underwater Object Detection

This author M. Zhang et al. [22] developed a lightweight method for detecting underwater objects by mixing MobileNet-v2 and depth-separable convolution with Modified Attentional Feature Fusion (AFFM) for reducing model parameters as well as size. While maintaining a processing frame rate of 44.22 FPS, experimental results indicate a mean precision of 81.67% and 92.65%. Yeh et al. [23] demonstrated a lightweight underwater object detection network that simultaneously learns conversion of color as well as detection of objects for underwater images. It benefits underwater terrain scanning, autonomous underwater vehicles, and image-based underwater object detection applications. Lin et al. [24] concentrated on enhancing object detection algorithms for underwater datasets, presenting unique challenges such as color shifts, low contrast, blurring, and creatures that appear nearby. To address these concerns, the authors examine augmentation policies that simulate overlapping, occluded, and hidden objects. They proposed an augmentation method, Regional Proposal Network (RPN) produced Region of Interest mixed with random weight ratio (RoIMix), which introduces interactions between images by combining image-extract proposals. By integrating RoIMix into the training procedure, the efficacy of object detectors that are region-based, on the Pascal Visual Object Classes (VOC) as well as Underwater Robot Professional Competition (URPC) datasets was improved. Chen et al.'s [25] architecture, SWIPENet addresses the challenges of tiny object detection in underwater environments. It has high-resolution together with semantically dense Hyper-Feature Maps, novel sample-weighted loss function, and a sample re-weighting algorithm called Invert Multi-Class Adaboost (IMA).

Although previous research has yielded promising results in terms of processing speed and precision, there have been few investigations into the broader applicability of these techniques beyond the specific tasks of underwater terrain scanning, autonomous underwater vehicles, and image-based detection. In addition, the focus on specific datasets, such as the Pascal Visual Object Classes and Underwater Robot Professional Competition datasets, raises worries about how these approaches would generalize to real-world settings with variable conditions.

2.2.1 Prior Research on Underwater Garbage Detection

Huang et al. propose DSDebrisNet, a lightweight neural network designed to detect compound-scaled deep sea debris accurately and in real-time. DSDebrisNet combines rapid detection speed with superior identification performance, resulting in rapid and accurate debris detection. To further improve performance, DSDebrisNet introduces a hybrid loss function that addresses illumination and detection issues [26]. Using autonomous underwater vehicles (AUVs), researchers developed a system for efficiently detecting and removing marine debris in Zocco et al.'s paper [27]. Marine debris endangers marine ecosystems and human health because microplastics from decomposing detritus infiltrate the food chain. Without increasing GPU latency, the researchers improved the AUV's vision for detecting marine debris in real time by increasing the efficacy of the object detector, EfficientDets. They created a new dataset for detecting bottles and bags in water while training the enhanced EfficientDets on these and deep sea debris dataset.

For reduction of color cast as well as haze from images in the training, Ali et al. [7] propose innovative image enhancement techniques, such as a channel stabilization technique. These strategies enhance the outcomes, notably for the dataset's multi-scaled objects. While extant research concentrates on developing algorithms for the precise detection of specific underwater objects, this paper proposes an alternative method. Prasad et al. [28] propose an algorithm that accurately detects and classifies multiple classes in addition to a single class. This broadens the model's applications and purview, increasing its adaptability for detecting underwater objects. Mellone et al. [29], the ArgonautAI architecture employs an ensemble of single board computers having various aspects, including computational capacity, CUDA GPU, FPGA, GPIO, PWM, and specialized-I/O. These computers are managed utilizing Kubernetes, a platform for container orchestration, and a custom programming interface. The proposal included platform containers and mission containers. Hong et al. [30] contribute by supplying a valuable dataset for training and evaluating underwater garbage detection algorithms. Using the TrashCan dataset also establishes an initial benchmark for segmentation and object detection methods, enabling further advancements in this field. Ashwani Kumar et al. [31] proposed an object detection technique that can be used in any environment and on any device that runs the model to recognize objects in real-time. Within computer vision, object recognition and training is a broad, dynamic, yet ambiguous and intricate field. Convolutional neural networks are utilized in this suggested work to create a multi-layer model that classifies the given items into any of the defined classes. Watanabe et al. [32] investigated the efficacy of object detection algorithms utilizing deep-learning to monitor underwater ecosystems as well as marine debris in marine environments. Autonomous monitoring system composed of controlled robots was proposed to collect rich spatiotemporal marine data, having mean average precision values: 69.6% and 77.7%. These findings provide an optimistic starting point for developing instruments and technologies that facilitate the safe and accurate collection of marine data.

While the previous studies provide a valuable synthesis of current research, there is still potential for further improvement. Expanding the scope of research to include a more thorough exploration of model interpretability and the impact of human interaction on improving detections would greatly enrich this domain.

2.3 Prior Research Methods

In this section of the thesis, we have presented the techniques and algorithms utilized in past studies. Sreekala et al. [33] developed a Deep Convolutional Neural Network (DCNN) approach to resolve the issue of underwater images with insufficient illumination. This method combines the max-RGB and shade-of-grey techniques for enhancing undersea visibility and train the plotting association for lighting plot. It is shown that a deep convolutional neural network proposed in this paper authored by Song et al. [34], performs well in a dynamic situation. Therefore, they developed an Adaptive Foreground Extraction Approach that employs a deep convolution neural network for categorization. It performs quite well in real contexts due to the fact that it places a strong emphasis on lighting uncertainty, backdrop motion, and non-static imaging platforms. In dynamic situations, a Gaussian Mixture Model is used, whereas a Kalman filter is reserved for situations with a lower level of complexity.

Liu et al. [35] detected underwater objects using a Faster R-CNN, utilizing Swin Transformer within algorithm which serves as backbone network. Deep feature-map and shallow feature-map are combined and superimposed on one another once the path aggregation network has been added. Then, mining of online hard examples makes training process more effective. ROI pooling aligns with the ROI, which gets rid of the two quantization mistakes caused by the previous step and improves the detection performance. In terms of hybrid features, Kumar's [36] deep learning method makes use of VGGNet. This study introduces extraction technique depending on Spatial Pyramid Pooling (SPP) method. In this technique, pre-trained VGGNet improves classification through integration of deep-features from VGGNet alongwith texton as well as color-based attributes. After that, CNN training is done using the Meta Label Correction (MLC) dataset. In Table 2.1, we have included

Table 2.1: An Analysis of Prior Research Methods

Method	Details
DCNN [33]	It combines the max-RGB as well as shade-of-grey techniques for improving underwater visibility as well as training plotting association required for deriving the lighting plot.
VGGNet [37]	The use of combination features along with extensive features for prior training.
SegNet [38]	The capacity of the decoder network drastically decreases RAM usage.
YOLOv4 [39]	It prioritizes real-time object detection and uses a single CPU for training.
Fast R-CNN [40]	CNN needs training onetime per image to extract a map of features. Selected feature map searches create predictions. R-CNN uses each of the three models.
GAN-based approach [41]	Synthetic contacts were produced by tracing the rays of a 3D-CAD model onto the side-scan seafloor.
Transfer learning and CNN [42]	An CNN pre-trained approach to the inequality of classes issue, and the application of numerous data augmentation techniques.
Boosting R-CNN [43]	Provides high-quality proposals and models the object prior probability by incorporating objectness and IoU prediction for uncertainty.

several prior method analyses that have been applied in various works and have had a considerable impact on the identification of objects located under the surface of the water. Several researchers have utilized CNN-based techniques to detect underwater objects [33], [34], [41] with an adjustment that improves the effectiveness of object detection underwater.

Despite increasing efforts, detecting underwater garbage remains a complex challenge, with several critical research gaps hindering progress. Deploying and maintaining advanced underwater sensors can be expensive and technically challenging,

especially in deep or remote areas. Training machine learning models for underwater garbage detection requires qualitative datasets of accurately labeled underwater images, which are currently scarce. Additionally, developing algorithms that can adapt to diverse and dynamic underwater environments remains a challenge. The model, like most vision-based models, struggles with poor visibility underwater. Turbidity caused by plankton, sediments, or pollution can obscure garbage, making it invisible to the model's eyes. As there remain a lot of research gaps, we need more research on optimizing the model for these resource-constrained environments, making it efficient and lightweight without significantly compromising accuracy.

Chapter 3

Dataset Analysis

In the field of underwater computer vision, there are currently a limited number of datasets that may be used for research purposes, particularly about the trash that contributes to marine contamination. A summary of some of the most important variables that led to the lower number is provided below: A part of 890 similar combinations of underwater photographs and reference images constituting Dataset 2 which is Underwater Image Enhancement Benchmark (UIEB) [44] is illustrated in Figure 3.1. The researchers for this work obtained these photographs of underwater environments from Google, YouTube, other relevant studies, and their recordings [44]. The videos comprising this dataset 3 [45] are very different in terms of quality,

Figure 3.1: Underwater Images Collected by Dataset 2 Reprinted From [44].

depth, the things that can be seen in the settings, and the cameras that were utilized. They include a range of items in varying decay levels, occlusion, overgrowth and contain photographs of different forms of marine trash at real-world locations. These environments include the ocean, beaches, and harbors. Figure 3.2 displays a few illustrations of scenes from underwater environments. However, in addition to these datasets, there are several other datasets. A summary of these datasets can be found in Table 3.1, and these datasets have also been used in several articles.

Zhang et al. [52] proposed an encoder-decoder Siamese Underwater Image Co-enhancement Network (UICoE-Net). Its layers contain a correlation-feature matching unit to demonstrate the two branches' joint learning correlation. The UIEB, UICoD, and SQUID tested our method. Sylwia [53] highlights earlier research and discusses the authors' dataset tests to develop a first reproducible litter-detection baseline. New benchmark datasets detect–waste as well as classify–waste are integrated open-source datasets with unified annotations encompassing all garbage categories: bio, glass, metal, non-recyclable, paper, plastic, and others. Two-stage litter

Table 3.1: An Analysis of Some of the Existing Datasets for Underwater Object and Waste Detection

| Name	Dataset	Description
Dataset 1	Forward-Looking Sonar Marine Debris Dataset (FLS) [21]	The 1868 FLS photos in the dataset were taken with the ARIS Explorer 3000 sensor. The objects used to create this dataset are split into 11 classes in addition to a backdrop class and include common household maritime waste and distractor marine objects (tires, hooks, valves, etc.).
Dataset 2	Underwater Image Enhancement Benchmark (UIEB) [44]	A total of 950 authentic images of underwater life in the UIEB, from which 890 have corresponding references and 60 do not. Enhancing submerged images for research use is the objective of this project.
Dataset 3	Bounding Box-Labeled Dataset of Underwater Trash (Trash-ICRA19) [45]	It was derived from the J-EDI dataset of marine debris, which comprises 5,700 labeled images of trash, biological objects, and remotely operated vehicles (ROV) with bounding boxes. The objective is to develop accurate and efficient methods for onboard robot garbage detection.
Dataset 4	Marine Underwater Environment Database (MUED) [46]	Among 8,600 underwater images from MUED differing in position, stance, illumination, and turbidity of the water, 430 distinct categories of intriguing objects are represented. Detection of saliency and recognition of objects in underwater images is the academic objective.
Dataset 5	The TrashCan dataset [47]	The data set consists of 7,212 annotated images cataloging findings of marine debris, ROVs, and a diversity of aquatic life.
Dataset 6	Real-time Underwater Image Enhancement (RUIE) [48]	More than 4,000 actual underwater images are in RUIE's Submerged Image Quality Sub-aggregate, Underwater Color Casting Sub-aggregate, and Underwater more advanced task-oriented Sub-aggregate.
Dataset 7	URPC2020 dataset [49]	It is a dataset with 6,575 training as well as 2,400 test images, with the highest image resolution measuring 3,840 by 2,160 pixels. Annotations for the test set are also unavailable.
Dataset 8	UOT32 (Underwater Object Tracking) Dataset [50]	The standard benchmark dataset for underwater tracking has 32 films with 24,241 annotated frames, an average of 29.15-second long videos, and 757.53 frames.
Dataset 9	SUIM Dataset [51]	This is the first exhaustive dataset for the semantic segmentation of underwater images (SUIM). Over 1,500 images with annotation of pixels encompass eight categories of objects: vertebrates-fish, invertebrates-coral, aquatic vegetation, wrecks or ruins, divers, robots as well as the seafloor.
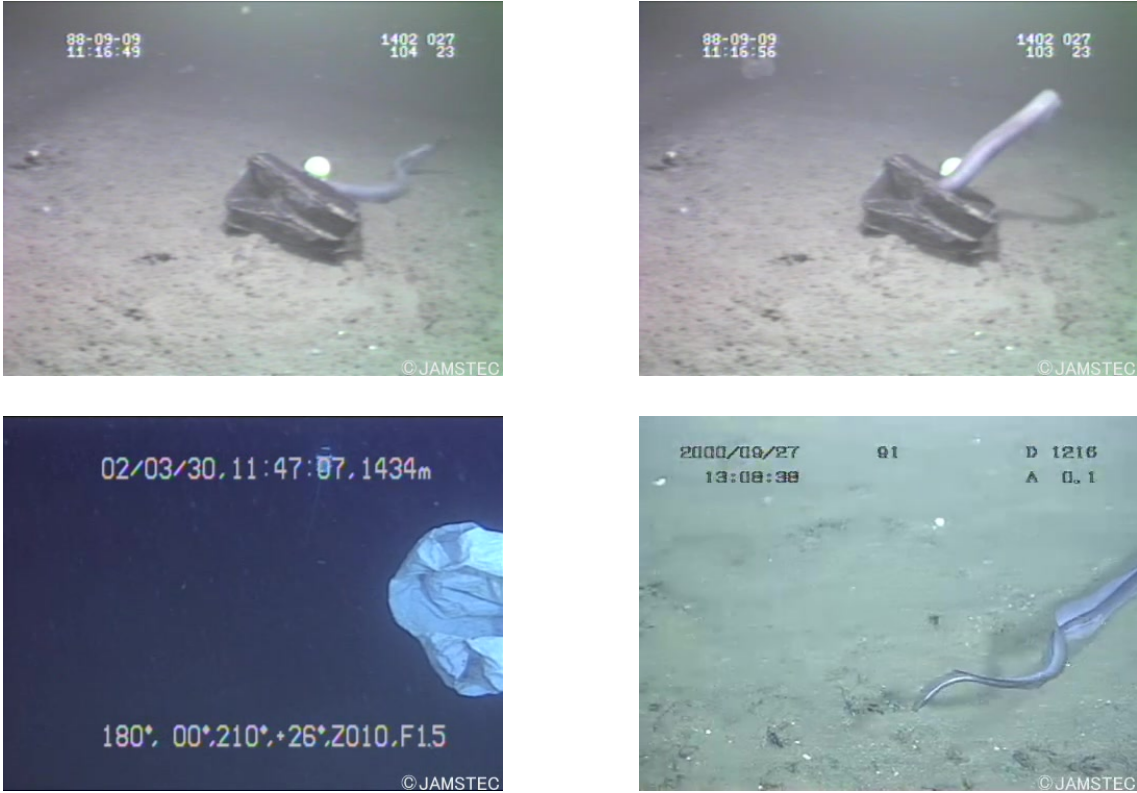
Figure 3.2: Underwater Images Collected by Dataset 3 Reprinted From [45].

detector concludes. EfficientNet-B2 classifies litter found by EfficientDet-D2. Unlabeled images semi-supervised train the classifier. The suggested technique detects 70% of waste and classifies 75% on the test dataset. This research [54] used dataset 4 to improve marine item detection with multiple Attention Path Aggregation-Network (APAN). They created a path-aggregation network with backbone-network attributes for bottom-up path augmentation. Each feature map is improved by bottom-up down sampling. These researchers [55], [56] used dataset 5 and proposed a wide asymmetric receptive field block for enabling features for having a wide receptive field, which allows the model to handle smaller object information. They also developed an Underwater Object Detection (UOD) paradigm using cross-domain information interaction and fusion. Using dataset 6, multiple publications [57] presented regional quality-superiority discriminative network (RQSD-Net) as a discriminator for embedded quality in CLUIE-Net. Multistage fusion convolutional neural networks enhance underwater photos. Their fusion technique includes two white balance (WB) priors.

In this section, we have discussed a number of well-known datasets for the detection of underwater marine objects as well as the detection of underwater waste. Because each of the aforementioned datasets suffers from a unique collection of flaws, there is currently no benchmark dataset that is acknowledged and utilized by all relevant parties. Due to the inadequate structure of several of them, researchers are required to construct their own training and testing sets through the use of custom splits. This makes it impossible to compare these methods to others that are considered to be state-of-the-art. The others are on a rather modest scale, which is not ideal for use in deep learning. These difficulties significantly pose challenges in the detection of underwater objects. The absence of well-annotated data seems to be a key barrier

in the detection of underwater items, raising the need for more inquiry in building a universally acceptable benchmark dataset.

3.1 Underwater Garbage Classification Dataset

Here dataset 1 is used for classification. A study is done to examine the classification of various types of rubbish using a dataset. For training and evaluating our model, we obtained a dataset from GitHub, a publicly accessible platform Forward-Looking Sonar Marine Debris Dataset. The dataset has 1,868 FLS images. The images have a resolution of 512 x 384 pixels. The debris objects in the images are labeled with one of 11 class labels. I ignore class Background for my evaluation purposes. My main motive is to identify the object class only. These categories are in Table 3.2.

Table 3.2: Classes Available in the Marine Debris Dataset

Name	Details
Bottle	Horizontal plastic and glass bottles.
Can	Metal food cans.
Chain	Meter-long chain with linked small chains.
Drink Carton	Horizontally milk or juice cartons.
Hook	Metal hook, small.
Propeller	Metalic ship propeller.
Shampoo Bottle	Shampoo bottle, standing and made of plastic.
Standing Bottle	Glass beer bottle, standing.
Tire	Horizontal rubber tire, small.
Valve	Metal valve, euRathlon 2015 competition design.
Background	Water tank bottom or non-object.

3.2 Underwater Garbage Detection Dataset

Here dataset 3 is used for detection. The dataset referred to as Trash-ICRA19 [58] is a collection of underwater images. This dataset specifically focuses on images that depict trash or debris present in underwater environments. The dataset is to provide researchers and practitioners in underwater robotics and computer vision with a valuable resource. Presented visuals depict video frames showcasing various types of waste, as well as the underwater plant and animal life, captured through a remotely-operated underwater-vehicle (ROV) lens. The used dataset originates at the Japan Agency of Marine-Earth Science and Technology (JAMSTEC) e-library of Deep-sea Images (J-EDI) dataset, which is curated by the JAMSTEC (J. A. 2012). The recorded images encompassed a diverse range of objects, as they were captured within authentic real-world environments. The dataset exhibits widely varying water clarity as well as light quality across different images. The datasets used in this research contain images with dimensions of 480×270 pixels and 480×360 pixels. The annotations provided adhere to the COCO format, which is commonly used in

computer vision research and applications. The Trash-ICRA19 dataset consists of 7,668 images and 6706 annotations. The annotations in this dataset are done at the detection level. The classification system comprises seven distinct categories, each delineated by the material composition of the objects under consideration. This dataset has 3 classes: plastic materials as marine debris, placed man-made objects including ROV, and natural and biological materials, including plants, fish, and biological detritus. Figure 3.3 illustrates samples from the dataset.

(a) Glass

(b) Plastic

(c) Metal

(d) Octopus

Figure 3.3: Samples from Underwater Waste Image Dataset Reprinted From [45].

Chapter 4

Methodology

This chapter provides overview of the employed approach for garbage detection categorization. This part commences with the presentation of our research methodology. The proposed categorization is conducted in three sub-phases: image processing, training phase, and evaluation phase.

4.1 Data Pre-processing

To facilitate the training and evaluation of the model, dataset 1 and dataset 3 were partitioned to, 70% training subset, 20% validation subset, and 10% test subset. The validation split was generated by partitioning the training subgroup, encompassing 70% of data for training. The training subset provides a thorough comprehension of the diverse elements included in the images. The validation subset is maintained as a distinct entity from the training subset. The data is used for the model at the completion of every epoch, and model performance is assessed. After training, the model's capacity to operate on unseen data is evaluated by assessing its performance on the test subset. To avoid the potential issue of over-fitting, the data set was enhanced by data-augmentation techniques like horizontal flipping, rotation, shearing, and zooming.

4.2 Training Phase

4.2.1 Underwater Garbage Classification Methodology

The system design is shown in Figure 4.1. The CNN models underwent training on the Google Colab Pro Edition platform, leveraging CUDA version 11.2. The training system utilized 26.3 gigabytes (GB) of random-access memory (RAM) and 16,160 megabytes (MB) of graphics processing unit (GPU) RAM. The training of each model employed a batch size of 32, and the training process was executed for a maximum of 50 epochs. The quantitative evaluation of all models (ResNet, VGG16, DenseNet, InceptionV3, ViT Transformer, and Swin Transformer) is conducted using criteria such as recall (sensitivity), accuracy, precision, and positive-predictive value (PPV). In order to evaluate the veracity, exactness, and responsiveness of a model, mathematical expressions denoted as Accuracy Equation 4.1, Precision Equation 4.2, and Sensitivity Equation 4.3 are employed. The equations presented in this thesis are derived from the samples of true-positive (TP), false-positive (FP),
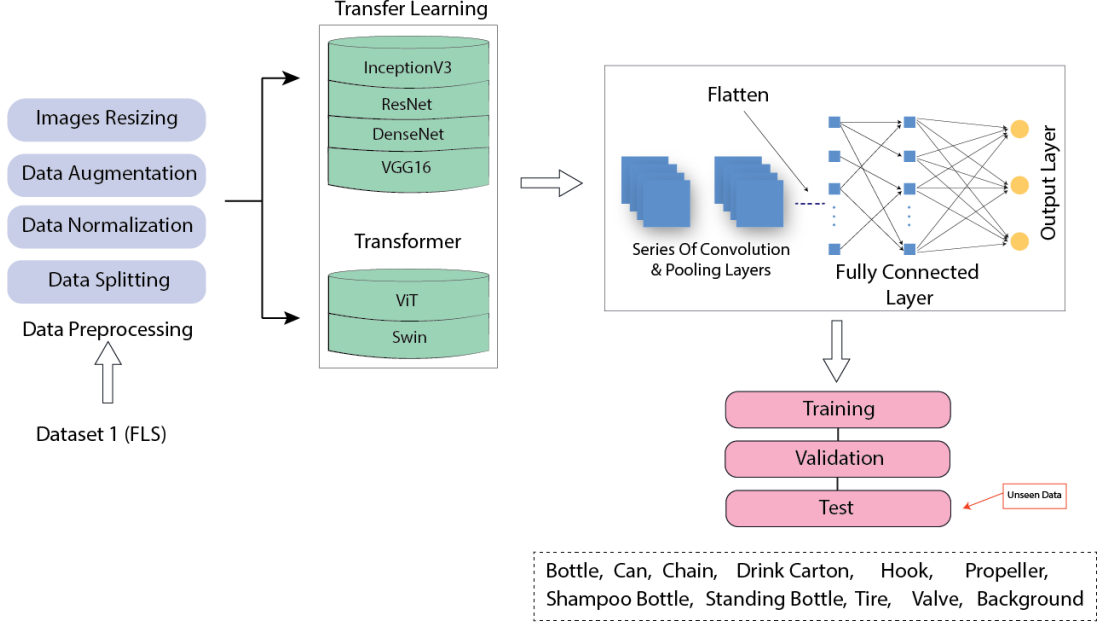
Figure 4.1: System Architecture of Underwater Garbage Classification.

true-negative (TN), and false-negative (FN). The recall metric, also known as sensitivity, quantifies the capacity of a model to correctly identify all pertinent instances present in a given dataset. The division of sum of true-positive and false-negative by the number of true-positive is performed. This finding demonstrates the study's capacity to accurately detect individuals.

$$Acc_i = \frac{TP_i \times TN_i}{TP_i + TN_i + FP_i + FN_i} \times 100\% \tag{4.1}$$

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \tag{4.2}$$

$$Sensitivity_i = \frac{TP_i}{TP_i + FN_i} \tag{4.3}$$

$$F_1 = 2 \times \frac{Precision_i + Sensitivity_i}{Precision_i + Sensitivity_i} \tag{4.4}$$

The classification approach has 12 classes including a) metal, b) shoes, c) paper, d) plastic, e) trash, f) biological, g) cardboard, h) brown glass, i) white glass, j) battery, k) clothes, and l) green-glass as potential classes.

4.2.2 Underwater Waste Detection Methodology

In this thesis, I utilize the YOLOv8 model [59]. This approach allows an easy replication and comparison with other methods applied to our dataset. The Figure 4.2 provided in this context serves as an illustrative representation of all the procedures involved in the given scenario. To ascertain the optimal model size for our experimental purposes, a comprehensive comparative analysis was undertaken on the different iterations of YOLOv8 models. The objective of this comparative analysis is to examine the trade-offs that exist between the accuracy of detection and the efficiency of computational processes.
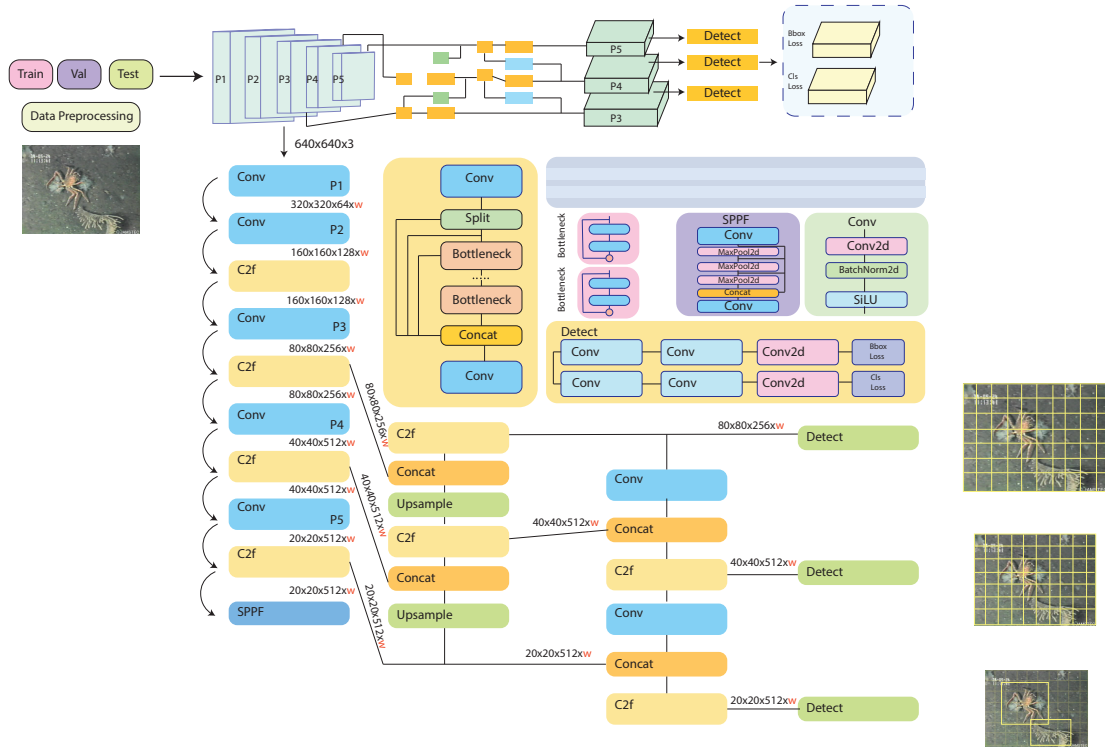
Figure 4.2: System Architecture of YOLOV8 Reprinted From [59]

The training and inference procedures were conducted on a high-performance computing system equipped with a 24 GB NVIDIA GeForce RTX 3090 graphics processing unit (GPU). The YOLOv8.0.53 framework, implemented in Python 3.10.4, was utilized for the experiments. The deep learning library PyTorch 1.12.1, along with CUDA 11.7, was employed to leverage the GPU's parallel computing capabilities for accelerated computations. In this study, the training process involved utilizing images with dimensions of 640×640 pixels. The training was conducted over a span of 25 epochs, with mini-batches varying in size from 64 to 12, depending on the specific model being used. To prevent overfitting, an early stopping technique was implemented, with a patience parameter set at 25 epochs. In the evaluation of the validation data, various metrics are considered. These metrics include the last epoch, which indicates the point at which the model training ceased due to overfitting. 25 epochs are employed to determine this stopping point. Additionally, the best epoch is identified, representing the epoch with the most favorable results before the onset of overfitting. The average training time per epoch is calculated, along with the Frame per Second (FPS) achieved during training process. Precision is measured using a value of 2, while recall is measured using a value of 3. The mean Average Precision (mAP) at 50% overlap (mAP50) and the mAP across the range of 50% to 95% overlap (map50:95) are also computed. Furthermore, the standard deviation of the mAP50 values is calculated across the 10 folds of the data. In this study, all results, with the exception of the standard deviation, were computed as the average across all ten folds. The primary evaluation metric utilized in our research is the mAP, with a specific focus on mAP50. In addition to our analysis, we also examine the FPS metric, taking into account its relevance in real-world scenarios. The mAP is commonly used evaluation metric for computer vision tasks. The calculation is performed with Intersection over Union (IoU), which in terms gets calculated using

17

Jaccard index. IoU metric quantifies the degree of intersection of bounding boxes of predicted and ground truth. mAP calculation averages accuracy values from different memory levels, while considering a certain IoU criterion. This metric provides a broad assessment for model performance considering precision and recall, both. The IoU, often known as the Jaccard index, is a frequent similarity metric. The size of the intersection of the two sets divided by their union yields it. The IoU measures set overlap or agreement from 0 to 1. A number of 0 denotes no overlap, whereas 1 denotes complete overlap or identical sets. Computer vision, machine learning, and data analysis use the IoU measure. Its main function is to evaluate algorithm performance and compare items or regions.

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{4.5}$$

$$(P)recision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{4.6}$$

$$(R)ecall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{4.7}$$

In the domain of research, the determination of Average Precision (AP) score for a particular class is commonly achieved by arranging the predictions of a model according to their confidence scores. Subsequently, the calculation of area beneath precision-recall curve is performed.

$$AP = \sum_n (Recall_n = Recall_{n-1}) \times Precision_n \tag{4.8}$$

$$mAP = \frac{AP_{IoU=0.5} + AP_{IoU=0.55} + ... + AP_{IoU=0.95}}{k} \tag{4.9}$$

Given that our current model is designed to specifically detect rip currents, it is noteworthy that the mAP for proposed model is equivalent to the Average Precision (AP) for the rip currents class. This is due to the fact that our model focuses solely on detecting instances of rip currents and does not consider any other classes. Therefore, the mAP and AP values are identical in this case. Within the realm of assessing object detection models, the metric denoted as mAP50 pertains to the mAP that is calculated at an IoU threshold of 0.5. This threshold is used to determine the level of overlap of boxes of predicted and ground truth. By calculating the AP at this specific IoU threshold, we can assess the model's performance in accurately localizing objects with a moderate level of overlap. For evaluating object detection models, mAP 50:95 is commonly used. This metric involves calculating the AP for various IoU thresholds ranging from 0.5 to 0.95, in increments of 0.05. The AP values obtained for IoU thresholds are averaged to obtain final mAP 50:95 score. The variable 'k' denotes IoU thresholds being considered in this research. The assessment of our model's efficacy in accurately delineating rip currents at various IoU thresholds is carried out by the mAP at 50% IoU (mAP50) and the mAP at 50% to 95% IoU (mAP50:95). The evaluation of the model's efficiency encompasses the consideration of both training speed and inference time. The assessment holds significant importance in assessing the viability of our proposed approach for practical implementation in the field of rip current detection, followed by analysis. Test photos are evaluated by assessing the accurately segmented number of frames to

identify rip currents. Average precision at 50 (mAP50) quantifies the assessment. The correctness and failures of each video will be examined in this study. The final results presentation requires macro and micro averages. The macro average calculates the average accuracy for every image, while the micro average considers all true positives.

Chapter 5

Performance Evaluation

5.1   Underwater Garbage Classification Evaluation

The study involves the comprehensive evaluation of the trained models and the subsequent presentation of a detailed comparative analysis of their performance. The evaluation process involves assessing various metrics and criteria to gauge the effectiveness and efficiency of the models. To assess the specified dataset, I selected four models. VGG16 has demonstrated good performance on various image classification tasks, achieving competitive accuracy on benchmarks like ImageNet. This track record of success suggests that it might also be effective for the marine debris classification problem. Additionally, VGG16 has a relatively moderate architecture compared to other deep learning models, making it computationally efficient for training and deployment on resource-constrained platforms. FLS images can be noisy and contain occlusions due to water disturbances or overlapping debris objects. DenseNet's dense connections make it more robust to these challenges compared to other models that rely solely on feedforward connections. ResNet is recommended because it is a sufficiently deep model with 34, 50, or 101 layers. The deeper the hierarchy, the stronger the representation capability and the higher the classification accuracy. ResNet is one of the models that I selected as a result. On the other hand, Inception V3 employed a "stem module" to enhance information flow across the network and incorporated "spatial factorization" to lower the computational cost of convolutions. Additionally, "label smoothing" was added in Inception v3 to regularize the network's output and enhance its generalization capabilities. Transformer models are neural networks as well, but they perform better than convolutional and recurrent neural networks (RNN and CNN, respectively). This is due to their ability to process all input data simultaneously rather than sequentially. For this reason, I have selected the ViT and Swin transformers for evaluation. The comparative analysis will provide a comprehensive overview of how each model performs in relation to one another, highlighting their overall performance. To assess performance of four models, comprehensive evaluation was conducted by computing multiple performance metrics. These metrics included accuracy, recall, F1 score, positive predictive value, and the assessment of test data. By considering these various metrics, a more thorough understanding of the models' capabilities and limitations could be obtained.

Table 5.1 presents a comparative analysis of six prominent deep-learning models employed in garbage classification. The key performance metrics assessed include ac-

Table 5.1: Performance Measures of Garbage Classification Models.

Models	Accuracy	Precision	Sensitivity	F1
VGG16	94%	95%	94%	94%
DenseNet	95%	95%	94%	94%
ResNet152	87%	89%	85%	86%
InceptionV3	86%	88%	86%	87%
ViT Transformer	98.95%	98.67%	100%	99.33%
Swin Transformer	98.61%	98%	97%	97.9%

curacy, precision, sensitivity, and F1 score, each providing valuable insights into the models' effectiveness in correctly identifying waste items. ViT Transformer emerges as the undisputed champion, achieving remarkable scores across all metrics: 98.95% accuracy, 98.67% precision, 100% sensitivity, and 99.33% F1 score. This translates to near-perfect identification of garbage items, leaving no piece of trash unclassified. Swin Transformer trails closely with admirable performance, boasting 98.61% accuracy, 98% precision, 97% sensitivity, and 97.9% F1 score. While marginally less accurate, Swin Transformer's exceptional sensitivity ensures minimal misidentification of genuine garbage items. Traditional convolutional neural networks, although performing commendably, fail to match the prowess of Transformer models. VGG16 and DenseNet achieved respectable scores across all metrics (94% for VGG16 and 95% for DenseNet), demonstrating their competence in the task. However, ResNet 152 (87% accuracy, 89% precision, 85% sensitivity, and 86% F1 score) and Inception V3 (86% accuracy, 88% precision, 86% sensitivity, and 87% F1 score) lag further behind, highlighting the superiority of Transformer models in this specific application. This analysis underscores the remarkable potential of Transformer models for garbage classification, surpassing traditional convolutional neural networks in their ability to accurately identify waste items. ViT Transformer stands out as the clear leader, showcasing near-flawless performance. The model in this Figure
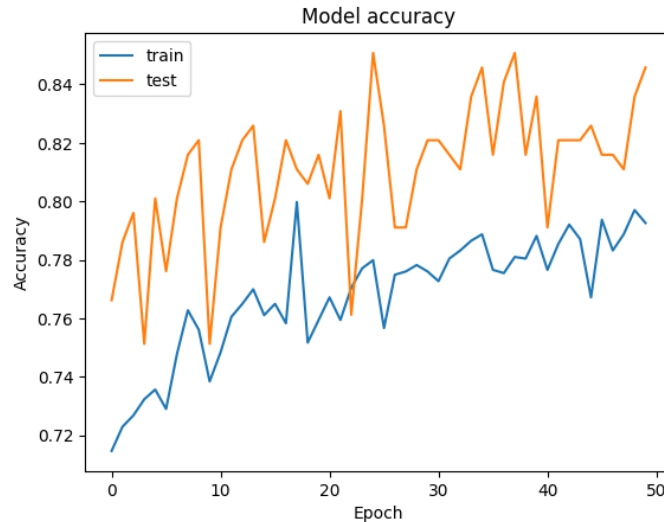

Figure 5.1: Accuracy of InceptionV3 Model for Classification.

5.1 learns rapidly in the early epochs, as both training and test accuracy increase significantly. After epoch 20, the training accuracy continues to rise, but the test accuracy plateaus and then slightly decreases. This suggests potential overfitting,

where the model is becoming too specialized for the training data. The model's best performance on the unseen test data is 0.82, achieved around epoch 20. In Figure
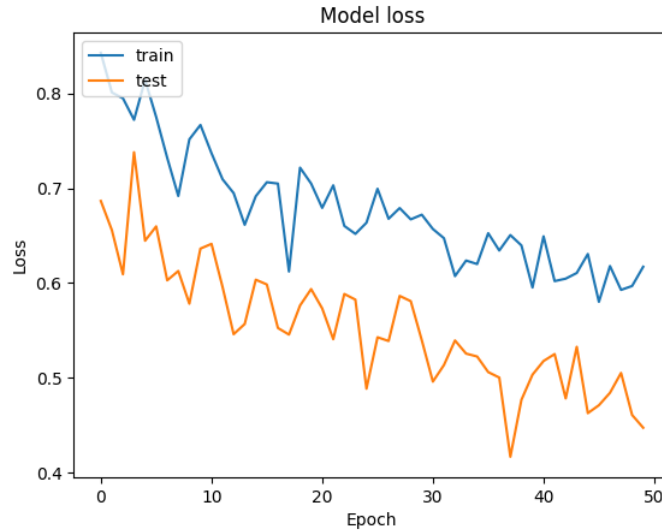

Figure 5.2: Loss of InceptionV3 for Classification.

5.2, both training and validation loss decrease rapidly in early epochs, indicating effective model learning. After epoch 25, training loss continues to decrease, but validation loss increases. The model's best generalization performance is around epoch 25, where validation loss is lowest. The confusion matrix of this Figure 5.3, with
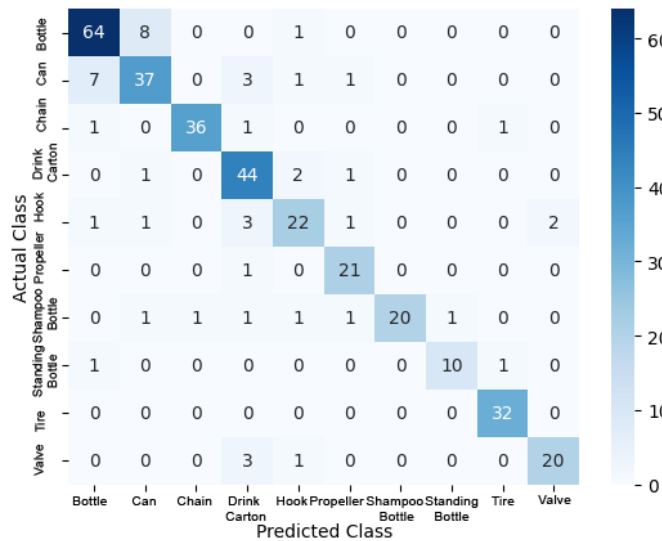

Figure 5.3: Confusion Matrix of InceptionV3 for Classification.

its 10x10 dimensions, reveals the performance of a multi-class classification model. While the model demonstrates reasonable accuracy, closer examination unveils key misclassifications. Classes Bottle and Can exhibit confusion, with instances misattributed in both directions. Similarly, Class Hook struggles with differentiation, being mistaken for several other classes. In this Figure 5.4, the training accuracy reaches a maximum of 0.9, while the test accuracy peaks at around 0.8. This indicates the model's best performance with both seen and unseen data. The highest accuracies are achieved between epochs 30 and 40, suggesting a suitable training
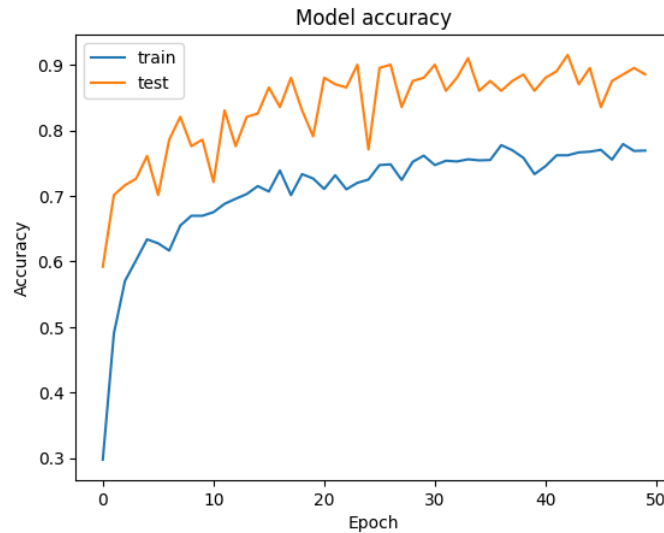
Figure 5.4: Accuracy of RestNet Model for Classification.

duration. The gap between the train and test lines widens slightly after epoch 30, potentially indicating a slight overfitting to the training data. In this Figure 5.5,
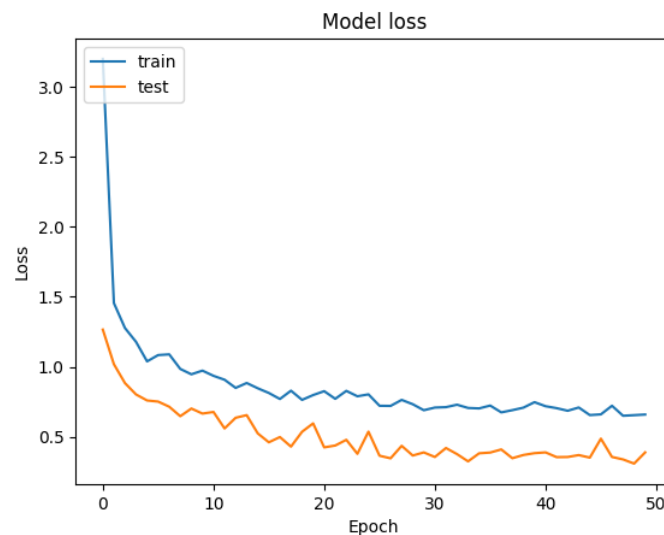

Figure 5.5: Loss of ResNet Model for Classification.

both lines steadily decrease over time, indicating that the model is learning and improving its predictions. The training loss reaches a minimum of around 0.5, while the test loss plateaus around 1.0. This suggests a lower error on training data but potentially some difficulty generalizing to unseen data. The train and test loss lines converge toward each other, which is generally a positive sign for model fitting. In this Figure 5.6, with an overall accuracy of 86.76%, the model demonstrates a strong predictive capability. However, examining individual class performance underscores the need for further refinement. Among the ten classes, the model excels at identifying instances of class Bottle and Tire, achieving near-perfect precision and recall for both. Conversely, class Chain presents a significant challenge, with an accuracy of only 64.10% indicating considerable misclassification. Further analysis is warranted to understand the factors contributing to this disparity. In this Figure 5.7, both lines rise steeply within the first 10-20 epochs, indicating rapid model improvement.
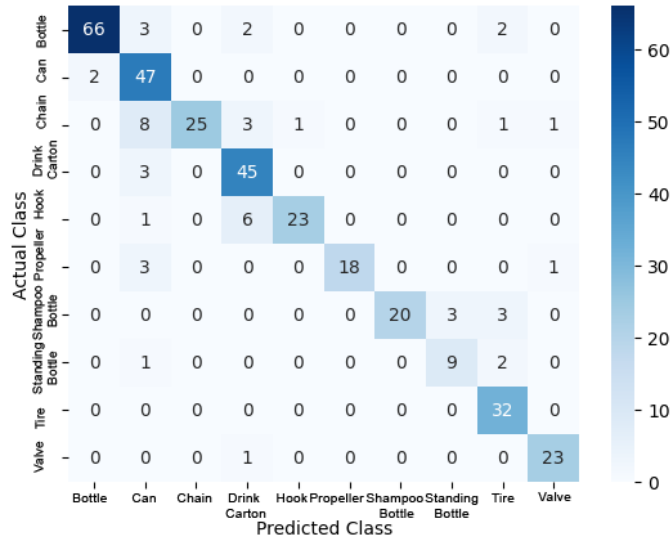
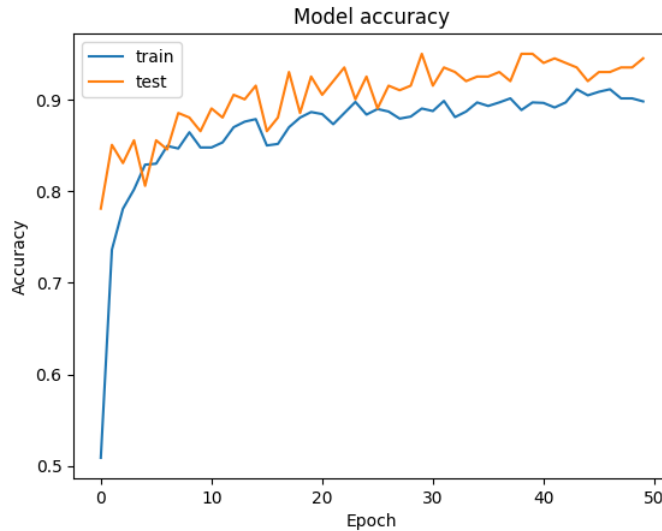Figure 5.6: Confusion Matrix of ResNet Model for Classification.

Figure 5.7: Accuracy of VGG16 Model for Classification.

The train line reaches a peak of around 0.95, reflecting excellent learning on training data. The test line peaks at around 0.85, suggesting a good but not perfect generalization to new data. The gap between train and test lines after 20 epochs hints at potential overfitting, where the model becomes too tailored to training data and loses some generalizability. In this Figure 5.8, The vertical axis measures model loss, ranging from 0.2 to 1.4. Lower values indicate better model performance. Both lines decrease over the epochs, representing model improvement. The training loss steadily declines to around 0.2, reflecting effective learning on training data. The test loss also decreases but plateaus around 0.4. The model achieves a final loss of around 0.4 on unseen data, indicating room for potential improvement. In this Figure 5.9, this multi-class classification model exhibits strong performance, achieving an overall accuracy of 91.49%. However, a deeper analysis of the confusion matrix reveals areas for potential improvement. While classes Drink Carton, Hook, Propeller, Tire, and Valve demonstrate near-perfect classification, class Chain poses a significant challenge with lower accuracy and F1-score. Further investigation is warranted to understand the factors contributing to this disparity. Furthermore, some misclas-
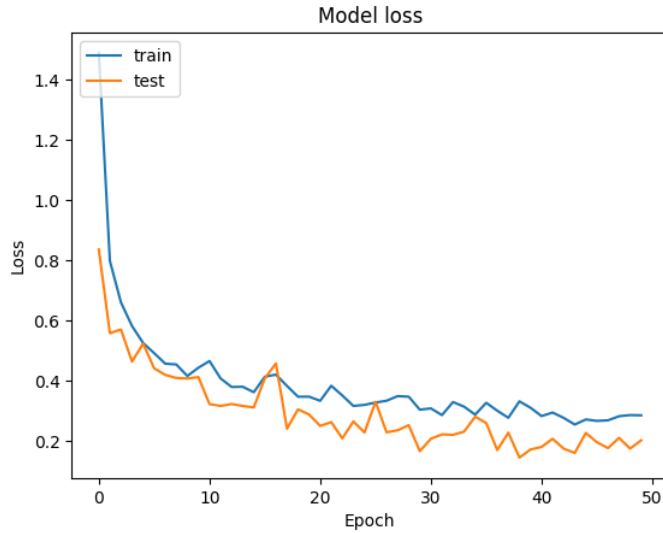
24

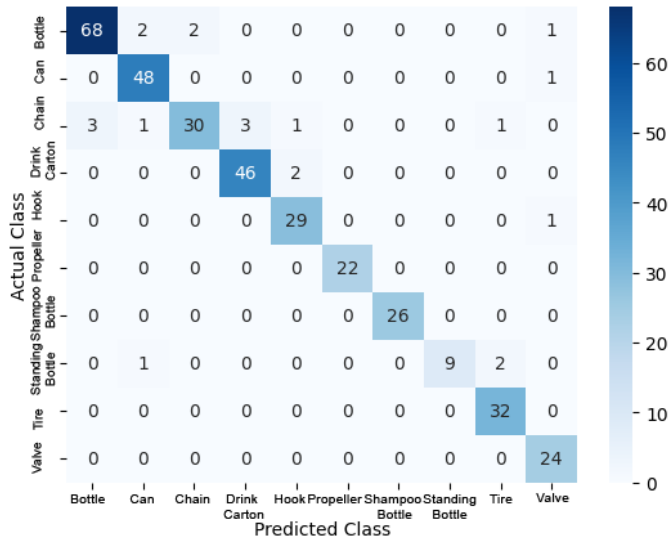Figure 5.8: Loss of VGG16 Model for Classification.

Figure 5.9: Confusion Matrix of VGG16 Model for Classification.

sification occurs between the classes Bottle and Can, suggesting potential confusion between these categories. Additionally, class Shampoo Bottle presents both false positives and false negatives, indicating the need for improved differentiation. In this Figure 5.10, both lines rise steeply within the first 10-20 epochs, suggesting rapid model improvement. The train line reaches a peak of around 0.95, reflecting excellent learning on training data. The test line peaks at around 0.85, suggesting a good but not perfect generalization. The model achieves a maximum accuracy of 0.85 on unseen data, which is a relatively good performance. In this Figure 5.11, both lines begin around 2.0, suggesting a high initial loss. Both lines decrease over the epochs, representing model improvement. The training loss steadily declines to around 0.5, reflecting effective learning on training data. The model achieves a final loss of around 1.0 on unseen data, indicating room for potential improvement. In this Figure 5.12, This multi-class classification model exhibits remarkable performance, achieving an overall accuracy exceeding 95%. A deeper analysis of the confusion matrix reveals a nuanced picture of its strengths and weaknesses. While classes Drink Carton, Tire, and Valve are classified flawlessly, a handful of challenges
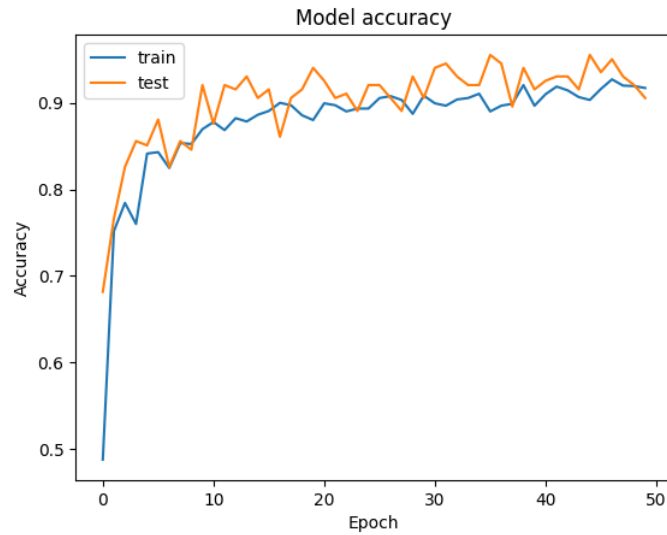
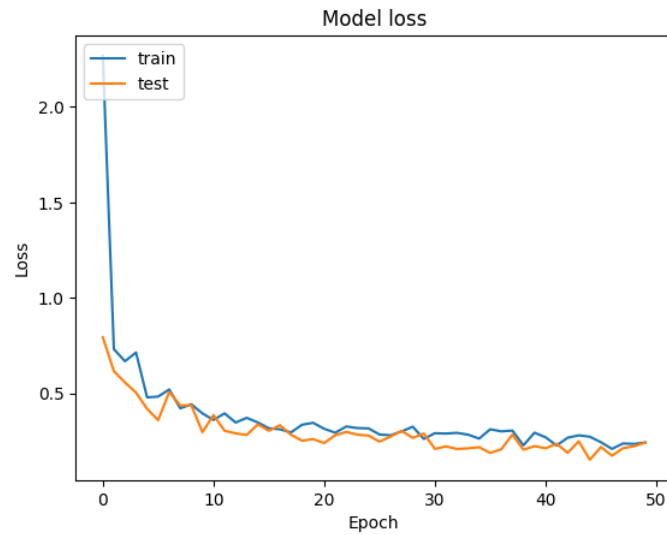Figure 5.10: Accuracy of DenseNet Model for Classification.

Figure 5.11: Loss of DenseNet Model for Classification.

emerge. Class Can exhibit some misclassifications, evidenced by reduced accuracy and F1-score. Class Standing Bottle poses the most significant challenge, requiring further investigation to understand the factors contributing to its lower performance metrics. Despite these minor shortcomings, the model demonstrates an impressive ability to accurately distinguish between most classes. Classes Bottle, Can, and Drink Carton show excellent performance, while classes Hook and Propeller achieve commendable accuracy with only a few errors. In this Figure 5.13, both lines start relatively low and rise steeply in the early epochs. This suggests rapid learning as the model adjusts its parameters to better fit the data. Train accuracy reaches a peak of nearly 1.00 around epoch 30, indicating excellent performance on the training data. Around epoch 30, the two lines begin to diverge, with train accuracy continuing to rise slightly while test accuracy plateaus or even decline slightly. This divergence hints at potential overfitting, there the model gets too customized on the training data and fails to generalize to new data. In this Figure 5.14, both losses decrease rapidly in the early epochs, indicating quick learning. The lines converge, suggesting good generalization to unseen data. The consistent decrease and convergence
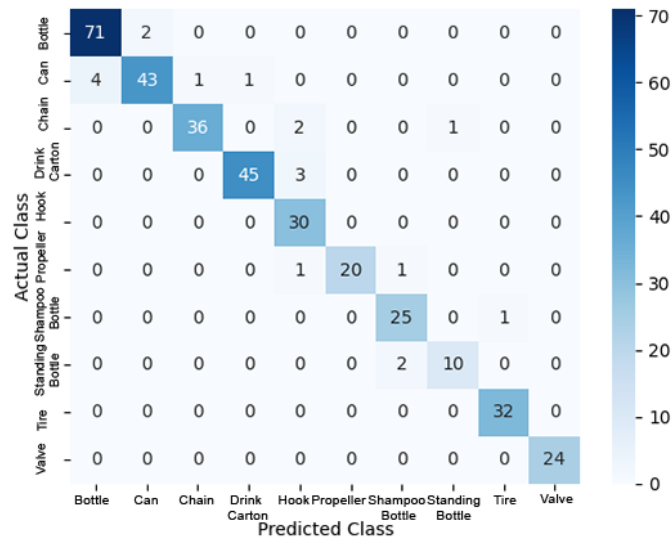
Figure 5.12: Confusion Matrix of DenseNet Model for Classification.

imply a stable training process. A training loss of 0.1 indicates good performance on training data. Test loss of 0.2 suggests a slightly higher error on unseen data but a reasonable generalization. In this Figure 5.15, this multi-class classification model delivers outstanding performance, exceeding 98% overall accuracy. A detailed analysis of the confusion matrix reveals a nearly flawless ability to discriminate between categories. Classes Bottle through Valve exhibit perfect classification, highlighting the model's exceptional discriminatory power. Even classes Bottle and Drink Carton boast impressive performance, with only a handful of minor misclassifications observed. These minor imperfections offer the potential for further optimization, and investigating the factors contributing to these false positives could lead to even greater accuracy in the future. Despite these slight missteps, the model's overall performance is truly remarkable, solidifying its potential for effective real-world application. While continuous exploration of optimization strategies is valuable, it's vital to acknowledge the exceptional classification capabilities this model already possesses. In this Figure 5.16, both accuracy measures start relatively low and rise steeply in the early epochs, suggesting rapid learning as the model adjusts to the data. Train accuracy reaches its highest point around epoch 30, indicating excellent performance on the training data. Test accuracy also peaks around this point, suggesting good generalization to unseen data. After epoch 30, the lines diverge slightly. Train accuracy continues to rise, while test accuracy plateaus. This divergence hints at potential overfitting, where the model becomes too focused on the training data and struggles to generalize. In this Figure 5.17, both loss measures decline over time, demonstrating that the model is learning and refining its predictions. The lines converge, suggesting good generalization to unseen data. The smooth and consistent decrease in both losses indicates a stable training process. A training loss of 0.2 suggests good performance on the training data. A test loss of 0.4 indicates a slightly higher error on unseen data, but still a reasonable level of generalization. In this Figure 5.18, a detailed analysis of the confusion matrix reveals a high degree of class discrimination, with perfect classification achieved for six out of the ten categories. Classes Can, Chain, Propeller, Shampoo Bottle, Standing Bottle, Tire, and Valve exhibit no errors, showcasing the model's robust
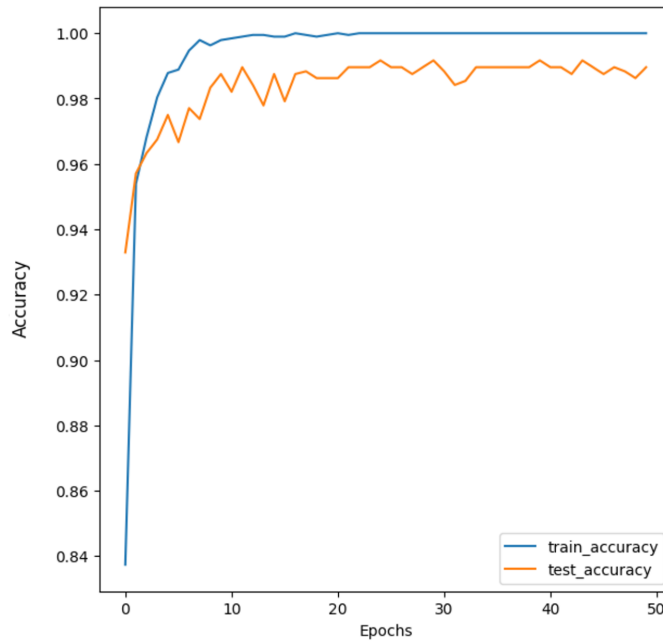
Figure 5.13: Accuracy of ViT Model for Classification.

ability to differentiate these specific groups. Even classes Bottle, Can, Drink Carton, and Hook, while demonstrating the presence of a few misclassifications, maintain strong performance. These minor imperfections offer possibilities for further optimization, and investigating the factors contributing to these missteps could unlock even greater accuracy in the future. Despite these small blemishes, the model's overall performance is truly remarkable, solidifying its potential for impactful real-world applications. While continuous exploration of optimization strategies remains valuable, it is crucial to acknowledge and appreciate the exceptional classification capabilities this model already possesses.

5.2 Underwater Object Detection Evaluation

To assess the effectiveness of the improved algorithm, this study conducted ablation experiments. These experiments were designed to systematically remove specific components or features of the algorithm in order to evaluate their individual contributions to its overall performance. By conducting these experiments, the researchers aimed to gain a deeper understanding of the algorithm's functionality and determine the extent to which each component or feature influenced its efficacy. The research conducted in this study involved the implementation of five separate and distinct sets of experiments. These experiments were carefully designed to ensure consistency and reliability in the results obtained. The same equipment and datasets were used in the case of train and test phases of experiments, further enhancing the validity of the findings. The adoption of this particular approach was deemed necessary in order to ensure that the results obtained from the study were both comparable and reliable. By employing this approach, the researchers aimed to es-
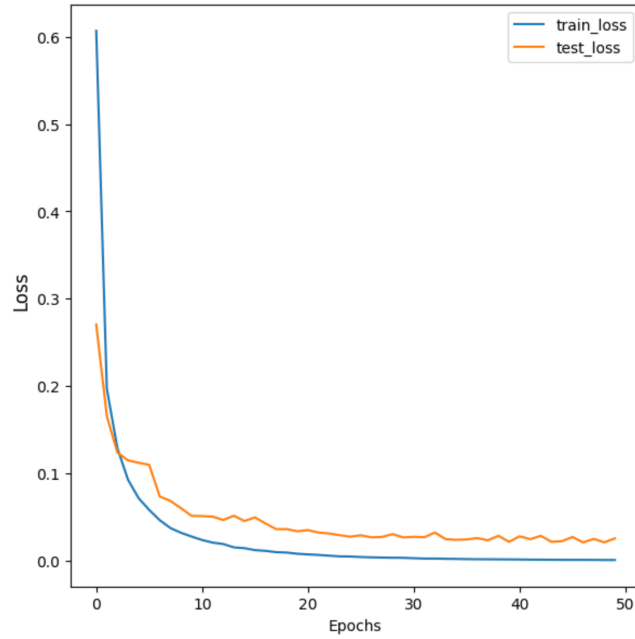
Figure 5.14: Loss of ViT Model for Classification.

tablish a standardized methodology that would allow for consistent data collection and analysis across different samples or settings. This would enable them to draw valid conclusions and make meaningful comparisons between the various variables under investigation. Ultimately, the use of this approach was crucial in maintaining the integrity and credibility of the study's findings. The findings derived from the analysis of the models under investigation are succinctly presented in Table 5.2.

Table 5.2: Performance Measures of Underwater Object Detection

Models	AP(%)	Precision(%)	Recall(%)	F1(%)
YOLOv8	92.2	81.4	82.0	81.66
YOLOv5-ghost	84.4	80.0	68.0	73.52
YOLOv4-ghost	76.8	74.0	59.0	65.68
YOLOX-Tiny	71.3	71.5	76.7	73.8

Throughout the duration of the investigation, it became evident that the YOLOX model exhibited suboptimal performance when applied to our specific dataset. The model yielded an overall accuracy score of 71.3%. This accuracy score was obtained through a rigorous calculation process, which involved evaluating the model's performance against a set of predetermined criteria. The accuracy score serves as a quantitative measure of the model's ability to classify and predict outcomes, thereby providing valuable insights into its effectiveness and reliability. When comparing the performance of various models, YOLOv4-ghost outperformed the other models. This superiority was evidenced by its ability to achieve a notable average accuracy score of 76.8 percent. In contrast, it is worth noting that the YOLOv5-ghost and YOLOv8 models exhibited noteworthy improvements in terms of accuracy. Specifically, the YOLOv5-ghost model achieved an accuracy score of 84.4 percent, while the YOLOv8 model demonstrated an even higher accuracy score of 92.2 percent. These results indicate a substantial enhancement in the performance of these models compared to their predecessors. Recent empirical investigations have yielded
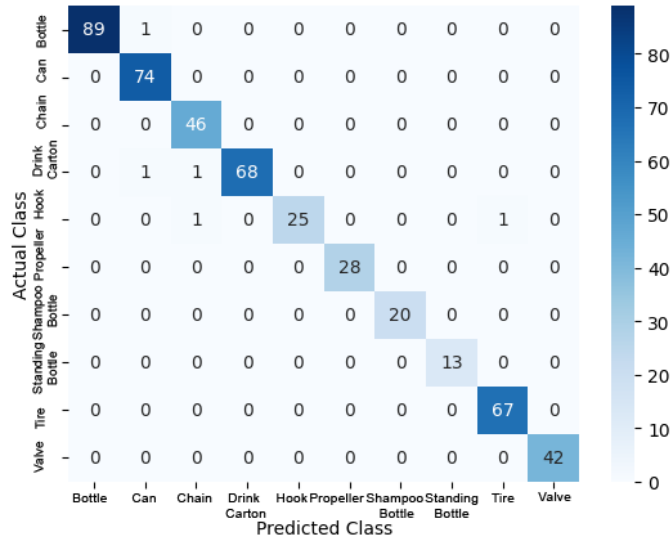
Figure 5.15: Confusion Matrix of ViT Model for Classification.

compelling evidence indicating that the YOLOv8 model, when juxtaposed with alternative models, exhibits a significantly heightened level of precision and accuracy in its output. The YOLOv8 model exhibits a commendable level of performance in terms of recall rate, achieving an impressive value of 82%. This remarkable outcome is observed when the model is utilized for the purpose of object detection. This finding implies that the probability of misclassifying an object during the detection procedure is significantly reduced. Upon conducting a comparative analysis of various models, it was observed that the YOLOv8 model exhibited a marginally higher recall rate compared to other models under consideration. A variety of loss curves, as well as precision and recall metrics, are shown in Figure 5.19 to assess how well the suggested garbage detection models work. The words "box loss", "class loss", and distributive focal loss—also referred to as "dfl loss"—all refer to a three-loss curve, which represents the model's loss in three terms. Those are class, bounding box, and the data imbalance issue. 25 epochs are run through the model. Initially, in the epoch, the model's loss value was the largest. Figure 5.19 (a) shows that the training box loss starts off at a high value and then decreases over time. This shows that the model is learning to predict objects' bounding boxes more accurately as it is trained on more data. The training box loss starts at a high value of 1.4. The loss drops quickly within the first 10 iterations, reaching around 1.0. The loss continues to decrease more gradually from iteration 10 to 20, stabilizing around 0.8. 5.19 (b) shows that the training box loss starts at a high value of 5. The loss drops quickly within the first 10 iterations, reaching around 2. The loss continues to decrease more gradually from iteration 10 to 20, stabilizing around 1. The graph also includes a smoothed line, which helps visualize the overall trend without being influenced by minor fluctuations in the data. 5.19 (c) shows that the training differential loss starts at a larger value of 1.7. The loss consistently decreases over the 25 iterations shown in the graph. The loss reaches a value of around 1.2 by the 20th iteration. The decline is relatively linear, suggesting a steady learning process without significant plateaus or fluctuations. The linear-like decline indicates that the model is learning smoothly from training data, without encountering major challenges or overfitting issues. The continued decrease in loss implies that extending the training process
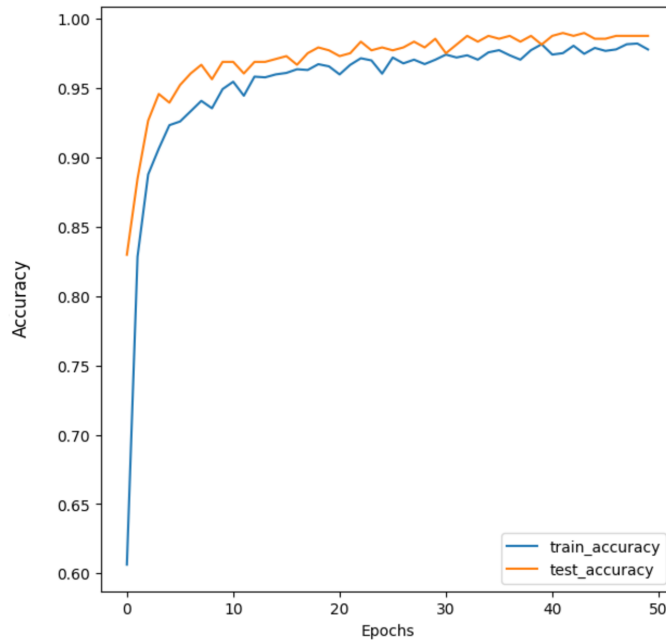
Figure 5.16: Accuracy of Swin Model for Classification.

could potentially lead to even lower loss and potentially better model performance. 5.19 (d) shows that the metric precision (B) starts at a larger value of 0.9. The value dips slightly to around 0.8 within the first 10 iterations. The value then remains relatively stable, fluctuating between 0.8 and 0.7 for the remaining iterations. The graph exhibits a somewhat flat trend with a minor initial dip. 5.19 (e) shows that the metric recall (B) starts at a larger value of 0.4. The value increases gradually over the 20 iterations shown in the graph. The value reaches approximately 0.43 by the 20th iteration. The increase is relatively slow, suggesting a gradual learning process for this metric. The graph exhibits a slight upward trend, indicating a moderate improvement in metric recall (B) over time. There may be room for hyperparameter tuning or extended training to achieve a more significant increase in recall. 5.19 (f) shows that the validation box loss starts higher than a value of 1.8. The loss drops quickly within the first 10 iterations, reaching around 1.0. The loss continues to decrease more gradually from iteration 10 to 20, stabilizing around 0.8 by the 25th iteration. The decreasing loss values demonstrate that the model is not only learning from the training data but also generalizing to unseen data in the validation set. 5.19 (g) shows that The validation classification loss starts at a value of 4. The loss drops steeply within the first 10 iterations, reaching around 1.5. The loss continues to decrease more gradually from iteration 10 to 25, stabilizing around 1 by the 25th iteration. The decreasing loss values demonstrate that the model is effectively learning to classify objects from both the training and validation data. 5.19 (h) shows that the validation differential loss starts at a value of 1.5. The loss decreases steadily over the 25 iterations shown in the graph. The loss reaches a value of around 1.0 by the 25th iteration. The graph exhibits a clear downward trend, indicating consistent improvement in the model's ability to predict differential features on the validation set over time. 5.19 (i) shows that the metric mAP50 starts from a lower value than 0.2. The value increases over the 25 iterations shown in the graph. 5.19 (j) shows that the metric mAP50 .95 starts at a value 0.1. The value increases over the 25 iterations shown in the graph.
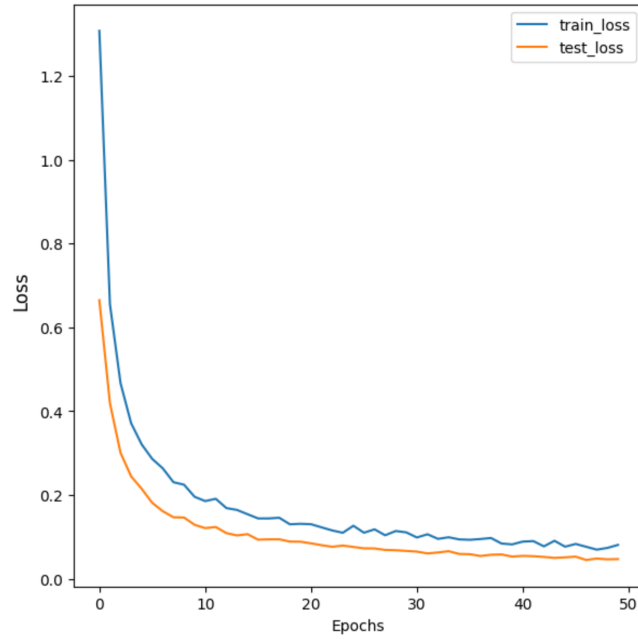
31

Figure 5.17: Loss of Swin Model for Classification.

The precision and label correlation of the model is illustrated in Figure 5.20 and Figure 5.21, respectively. These figures provide visual representations of the performance metrics of the model, showing insights into the accuracy, and correlation between predicted labels and ground truth labels. Figure 5.21 illustrates the label correlation of the YOLOv8 model for underwater garbage detection. The correlogram shows the relationships between the x, y, width, and height of bounding boxes. Each cell in the correlogram is a 2D histogram that shows the distribution of two variables. For example, the top left cell shows the distribution of the x and y coordinates of the bounding boxes. The color of each cell corresponds to the density of data points in that bin. Darker colors indicate more data points, while lighter colors indicate fewer data points. The diagonal cells of the correlogram show the distribution of each variable. For example, the top left cell also shows the distribution of the x coordinates of the bounding boxes, and the bottom right cell shows the distribution of the height of the bounding boxes. The off-diagonal cells show the relationship between two different variables. For example, the cell below the top left cell shows the relationship between the x and y coordinates of the bounding boxes. If the data points in this cell are concentrated along a diagonal line, it means that there is a positive correlation between x and y. In other words, as the x coordinate of a bounding box increases, the y coordinate also tends to increase. Conversely, if the data points are concentrated along a line that slopes down from left to right, it means that there is a negative correlation between x and y. In other words, as the x coordinate of a bounding box increases, the y coordinate tends to decrease. The diagonal line from the bottom left to the top right of the figure shows perfect correlation. If a point is on this line, it means that the model predicted the same number of labels as there are in the image. Points above the line mean that the model predicted more labels than there are in the image. As the points get closer to the diagonal, the model's performance gets better. Figure 5.20 depicts a precision–confidence. The precision–confidence graph shows precision-confidence correlation thresholds in the context of garbage detection offering insightful infor-
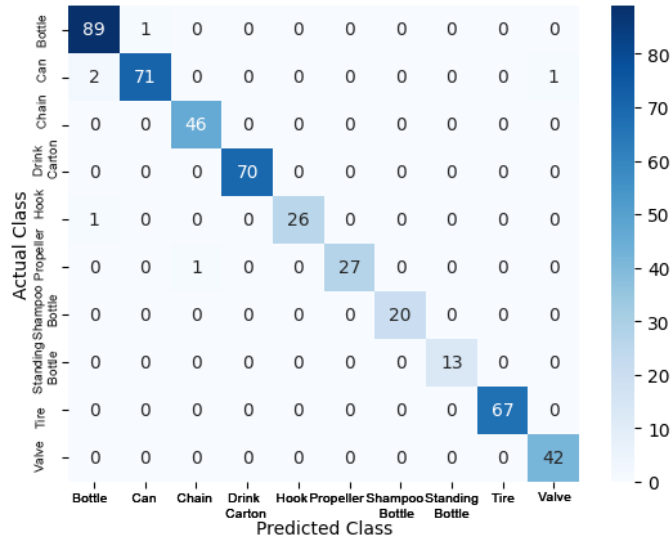
32

Figure 5.18: Confusion Matrix of Swin Model for Classification.

mation about recall and precision trade-offs when modifying the confidence level. When the confidence threshold is raised to the highest value of 0.935, the model exhibits robustness by reaching a maximum precision value of 0.81.

Utilization of visual representations in the analysis of model performance and effectiveness offers significant insights into various aspects, including precision and the degree of correlation between predicted labels and actual labels. These visual representations serve as valuable tools for evaluating and understanding the model's capabilities and limitations. Additionally, these visualizations enable the assessment of the correlation between the predicted labels and the actual labels, shedding light on the model's captureability of the underlying patterns and relationships among data. Overall, the incorporation of visual representations in the analysis of model performance enhances the interpretability and comprehensibility of the results, facilitating informed decision-making and further research in the field. Upon careful examination of the statistical data depicted in Figure 5.2, a clear and unequivocal observation emerges, namely that the YOLOv8 model exhibits better results of precision in comparison to the other models that were subjected to evaluation. The system exhibits a notably high degree of precision in its classification of object categories, as evidenced by its impressive precision score of 81.4%. Furthermore, it is important to highlight that the YOLOv8 model, in addition to the findings mentioned earlier, has been observed to demonstrate the highest mean Average Precision (mAP) score when compared to the other models that were evaluated. The achievement under discussion is of considerable significance, as evidenced by its commendable value of 48.2. This value serves as a robust indicator of the system's exceptional performance in accurately detecting and precisely localizing objects within the provided dataset. Based on the available information and analysis, we can infer that YOLOv8 exhibits superior performance in contrast with alternative models. This deduction is drawn from a comprehensive evaluation of various factors and metrics, which collectively indicate the superiority of the YOLOv8 model.

These predictions, which are referenced as 5.22 and 5.23, encompass a diverse array of submerged entities. The outcomes obtained from the application of the model to the underwater garbage dataset are represented in Figure 5.22. On the contrary,

33

(a) Train Box Loss (b) Train Class Loss (c) Train DFL Loss (d) Metrics Precision (B) (e) Metrics Recall (B)

(f) Val Box Loss (g) Val Class Loss (h) Val DFL Loss (i) Metrics mAP50 (B) (j) Metrics mAP50 .95 (B)
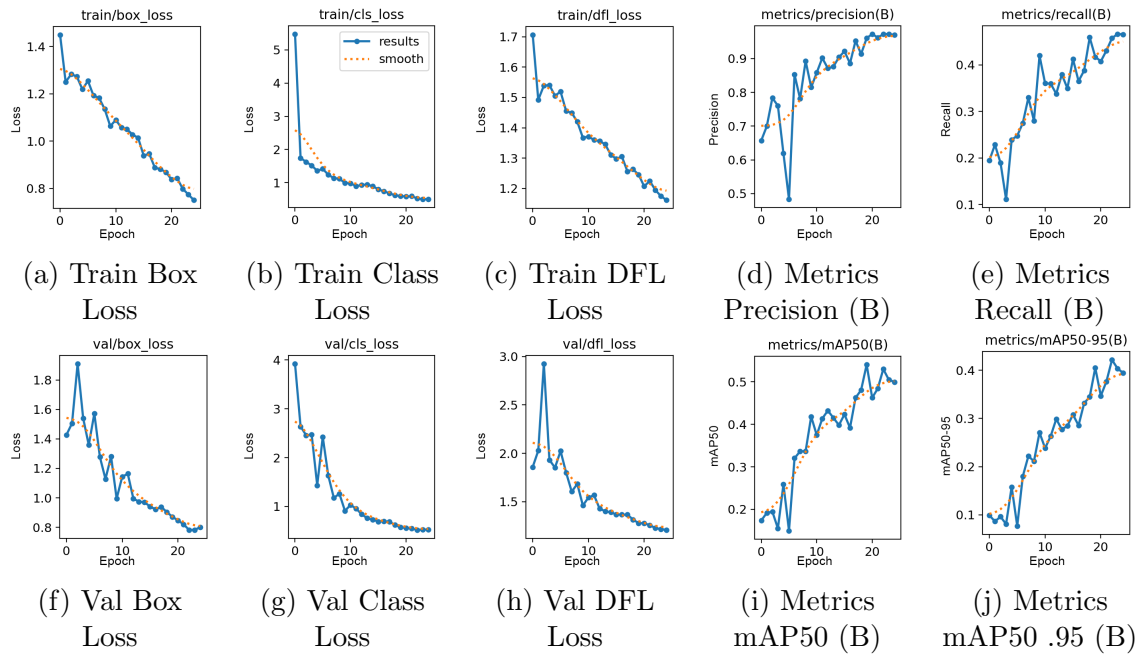
Figure 5.19: The Evaluation Results of Underwater Garbage Detection.
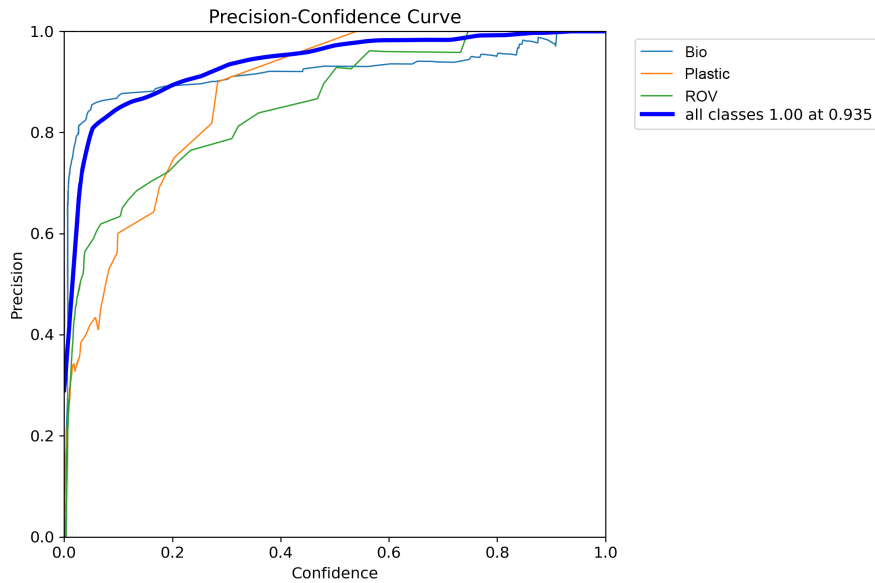
Figure 5.20: Precision of the YOLOv8 Model for Underwater Garbage Detection.

Figure 5.23 demonstrates the successful identification and localization of individual data points, thereby highlighting the high level of precision and accuracy attained during the detection process. Figure 5.24, shows the misclassification of the model prediction. It happens because YOLOv8 relies on features extracted from different convolutional layers, with deeper layers capturing finer details. Low-resolution images or small objects might not provide enough information for the deeper layers to effectively extract informative features for accurate classification. Small objects are more prone to being partially obscured by other objects or background clutter. This can further limit the available information for the model and lead to misclassification.

34
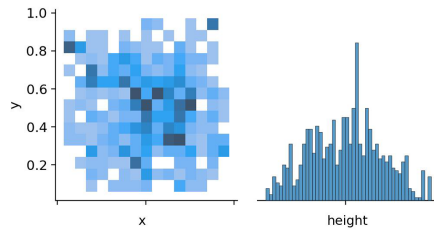
(a) Distribution of Object
Annotations with Y-Axis
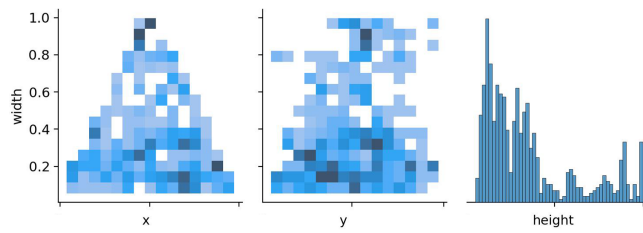
(b) Distribution of Object Annotations with Width
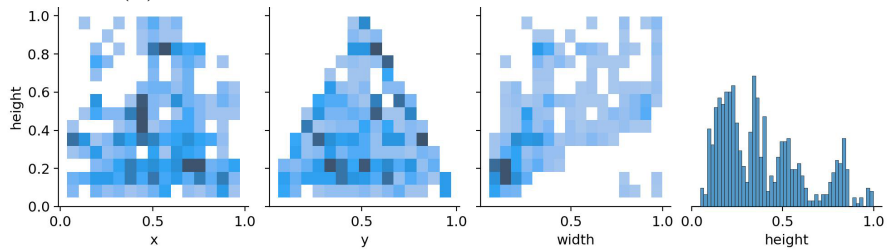
(c) Distribution of Object Annotations with Height

Figure 5.21: Label Correlation of the YOLOv8 Model for Underwater Garbage Detection.

Figure 5.22: Annotations and Detection using the YOLOv8 Model for Underwater Garbage Detection.
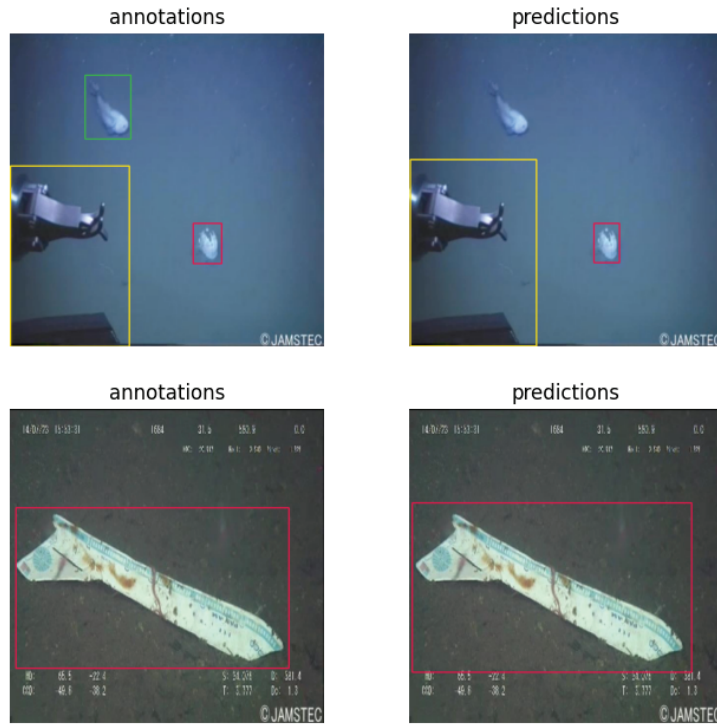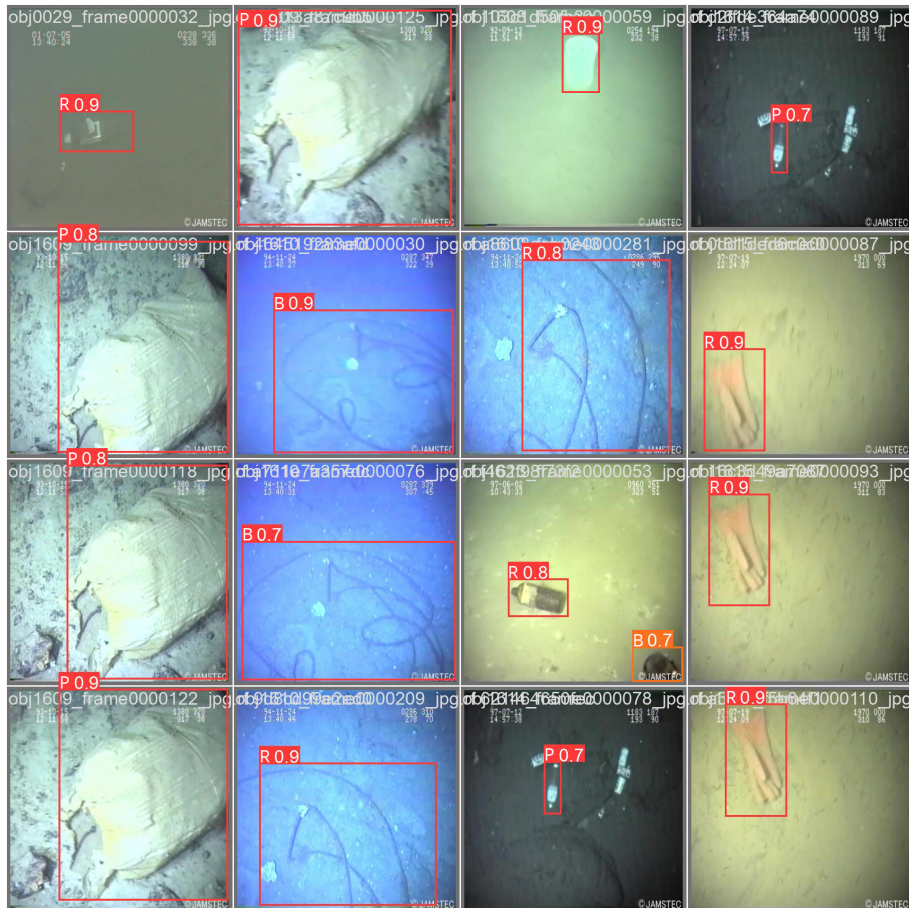

Figure 5.23: Prediction Accuracy for Underwater Garbage Detection.

Figure 5.24: Miss Classification Prediction for Underwater Garbage Detection.

Chapter 6

Discussion and Future Work

6.1 Challenges

I discuss issues and obstacles in the related field of underwater imaging, as well as suggest potential future research directions. The difficulties and obstacles that will facilitate the development of submerged processing of images and thus attract the interest of signal analysis and artificial intelligence researchers are briefly summarised. A number of important characteristics, such as the turbidity of the water, the ability of light to refract, absorb, or scatter in underwater environments, etc., continue to have the greatest influence on underwater image quality. The majority of currently available models concentrate on only one of the factors described above. In order to enhance the comprehensive image quality of underwater scenes, many impactful factors could be investigated in greater depth in future research. The promotion of submerged identification of saliency and object recognition faces a number of obstacles, the most significant of which are low-contrast elements with complex ocean floor scenes, variable underwater environmental lighting under certain circumstances, severe visibility that results in distorted images, and black and white underwater objects with limited color information. In spite of the fact that the imaging equipment itself influences the recorded underwater image, the impact of scattering along with the absorption of propagation of optical waves in unstable underwater often results in distorted images with an obvious distinction between the acquired image and the actual situation. Deep-learning techniques for underwater image handling are in their inception of research. Presently, the absence of a large underwater database significantly impedes the application of deep-learning-based methods for underwater image processing.

6.2 Limitations

Although this research contributes to the underwater garbage identification field using deep learning techniques, there are still significant limitations. The paucity of a comprehensive and uniform dataset targeted to underwater waste identification is the most significant of these issues. The lack of a dataset limits the suggested models' robustness and generalization, reducing their effectiveness in real-world circumstances. Furthermore, model interoperability is a key impediment to collaboration and the seamless integration of diverse deep-learning models into a cohesive framework. Furthermore, the testing phase of the proposed models is hampered by

slow inference, emphasizing the need for more efficient methods to accelerate real-time decision-making in underwater situations. Finally, the lack of a lightweight model suited for deployment on edge devices makes it difficult to execute the offered solutions in resource-constrained environments.

6.3 Future Work

To enhance overall underwater image quality, many impactful factors could be investigated in greater depth in future research. Light propagation according to the physical model merits additional research to resolve the intrinsic effects of the factors that may take a substantial amount of time and effort for underwater computer-vision research. Innovative methodologies and applications to resolve these gaps. If more datasets of labeled underwater images with ground truth become available, the future growth of deep learning-based frameworks for underwater images may increase exponentially.

6.4 Conclusion

Governments around the world are concerned about the sustainability of their countries and optimizing resource utilization through circular economies. To address this, we conducted a study on recent ML and Deep Learning models to analyze two different datasets as well as classification along with detection of underwater garbage. We pre-processed these images to be used in our deep learning models, which included popular models such as VGG16, DenseNet, ResNet, Inception v3, ViT, and Swin transformers. We evaluated the models using accuracy, F1 score, confusion matrix, specificity, and precision. We found that DenseNet performed well in accuracy measurement, but the ViT transformer outperformed every other model, achieving around 98.95% accuracy in garbage classification, which is the highest in the field. Additionally, the ViT transformer generated epochs in less time during the validation section compared to other models. This model will make garbage management in different countries including Bangladesh more efficient and flexible. On the other hand, Generic object detection has had a great deal of success in recent years due to the abundance of both vast amounts of data and highly advanced computing resources. Lately, a noteworthy surge may be observed in the level of attention directed toward the academic discipline of marine engineering. This refers to techniques that make use of deep learning in order to identify objects that are submerged in water, such as the ocean. The detection of garbage that is floating and on the surface is not very difficult; nevertheless, the quantification of waste that is submerged provides considerable obstacles. These challenges are caused by light-refraction, absorption, suspended particles, and color distortion. This opens up a vast range of possibilities for activities that can be done in the water. This study presents a thorough classification and examination of pertinent literature, utilizing the most up-to-date research in the field of underwater object identification and marine debris. This study presents a comparative analysis between deep learning methods and standard learning approaches. CNNs renowned for their efficacy in computer vision models and challenging categorization scenarios, have emerged as the most effective approach for object recognition in submerged aquatic environments. This is because

CNNs are able to take into account several features simultaneously. The purpose of this work is to present a thorough examination of advancements in the field of underwater object detection. It aims to furnish readers with valuable insights that can aid them in their own research endeavors. The utilization of deep learning in this particular application holds promise for the automation of trash recycling inside the challenging aquatic environment. Moreover, the research serves as a framework for future studies focused on the detection and categorization of submerged trash. To conclude, the primary aim of this review has been to offer a succinct overview of the latest studies in the field of underwater imaging. Through this examination, we have identified prevalent challenges and obstacles that researchers and practitioners face when working in underwater environments. Moreover, we have highlighted the need for future research endeavors that can address these challenges and explore new frontiers in underwater imaging. By recognizing the limitations and uncharted territories, we hope to inspire innovative solutions and propel advancements in this important field.

Bibliography

[1] N. H. Tasnim, S. Afrin, B. Biswas, A. A. Anye, and R. Khan, "Automatic Classification of Textile Visual Pollutants using Deep Learning Networks," *Alexandria Engineering Journal*, vol. 62, pp. 391–402, 2023.

[2] F. Ahmed, S. Hasan, M. Rana, and N. Sharmin, "A Conceptual Framework for Zero Waste Management in Bangladesh," *International Journal of Environmental Science and Technology*, vol. 20, no. 2, pp. 1887–1904, 2023.

[3] N. Li and Y. Chen, "Municipal Solid Waste Classification and Real-time Detection using Deep Learning Methods," *Urban Climate*, vol. 49, pp. 101–462, 2023.

[4] H. Zhang, H. Cao, Y. Zhou, C. Gu, and D. Li, "Hybrid Deep Learning Model for Accurate Classification of Solid Waste in the Society," *Urban Climate*, vol. 49, pp. 101–485, 2023.

[5] F. G. Altin, İ. Budak, and F. Özcan, "Predicting the Amount of Medical Waste Using Kernel-based SVM and Deep Learning Methods for a Private Hospital in Turkey," *Sustainable Chemistry and Pharmacy*, vol. 33, pp. 101–060, 2023.

[6] X. Sun, D. Yin, F. Qin, *et al.*, "Revealing Influencing Factors on Global Waste Distribution via Deep-learning Based Dumpsite Detection from Satellite Imagery," *Nature Communications*, vol. 14, no. 1, pp. 14–44, 2023.

[7] M. Ali and S. Khan, "Underwater Object Detection Enhancement via Channel Stabilization," in *International Conference on Digital Image Computing: Techniques and Applications (DICTA), IEEE*, pp. 1–8, 2022.

[8] S. Majchrowska, A. Mikołajczyk, M. Ferlin, *et al.*, "Deep Learning-based Waste Detection in Natural and Urban Environments," *Waste Management*, vol. 138, pp. 274–284, 2022.

[9] J. Niu, S. Gu, J. Du, and Y. Hao, "Underwater Waste Recognition and Localization Based on Improved YOLOv5," *Computers, Materials & Continua*, vol. 76, no. 2, pp. 1–17, 2023.

[10] A. Lee, B. Jiang, I. Zeng, and M. Aibin, "Ocean Medical Waste Detection for CPU-Based Underwater Remotely Operated Vehicles (ROVs)," in *2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE*, pp. 0385–0389, 2022.

[11] M. Shobana, M. Madhavan, V. Meha, S. Nandhini, and D. Neeraj, "AI-Underwater Drone in Protection of Waterways By Relating Design Thinking Framework," in *2023 International Conference on Computer Communication and Informatics (ICCCI), IEEE*, pp. 1–4, 2023.

[12] M. Tian, X. Li, S. Kong, L. Wu, and J. Yu, "A Modified YOLOv4 Detection Method for a Vision-based Underwater Garbage Cleaning Robot," *Frontiers of Information Technology & Electronic Engineering*, vol. 23, no. 8, pp. 1217–1228, Aug. 2022.

[13] B. Gašparović, J. Lerga, G. Mauša, and M. Ivašić-Kos, "Deep Learning Approach for Objects Detection in Underwater Pipeline Images," *Applied Artificial Intelligence*, vol. 36, no. 1, pp. 214–685, 2022.

[14] T. O. Fossum, Ø. Sture, P. Norgren-Aamot, I. M. Hansen, B. C. Kvisvik, and A. C. Knag, "Underwater Autonomous Mapping and Characterization of Marine Debris in Urban Water Bodies," *arXiv preprint arXiv:2208.00802*, 2022.

[15] B. C. Corrigan, Z. Y. Tay, and D. Konovessis, "Real-time Instance Segmentation for Detection of Underwater Litter as a Plastic Source," *Journal of Marine Science and Engineering*, vol. 11, no. 8, pp. 15–32, 2023.

[16] M. Ferdous and S. M. M. Ahsan, "A Computer Vision-based System for Surgical Waste Detection," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 3, 2022.

[17] B. C. Corrigan, Z. Y. Tay, and D. Konovessis, "Real-time Instance Segmentation for Detection of Underwater Litter as a Plastic Source," *Journal of Marine Science and Engineering*, vol. 11, no. 8, pp. 15–32, 2023.

[18] W. Liu, H. Ouyang, Q. Liu, *et al.*, "Image Recognition for Garbage Classification Based on Transfer Learning and Model Fusion," *Mathematical Problems in Engineering*, vol. 2022, p. 4 793 555, Aug. 2022, ISSN: 1024-123X. DOI: 10.1155/2022/4793555.

[19] F. S. Alrayes, M. M. Asiri, M. S. Maashi, *et al.*, "Waste Classification Using Vision Transformer Based on Multilayer Hybrid Convolution Neural Network," *Urban Climate*, vol. 49, p. 101 483, 2023.

[20] A. Aleem, S. Tehsin, S. Kausar, and A. Jameel, "Target Classification of Marine Debris Using Deep Learning," *Intelligent Automation & Soft Computing*, vol. 32, no. 1, 2022.

[21] A. R. Gaspar and A. Matos, "Feature-Based Place Recognition Using Forward-Looking Sonar," *Journal of Marine Science and Engineering*, vol. 11, no. 11, pp. 21–98, 2023.

[22] M. Zhang, S. Xu, W. Song, Q. He, and Q. Wei, "Lightweight Underwater Object Detection Based on YOLOv4 and Multi-Scale Attentional Feature Fusion," *Remote Sensing*, vol. 13, no. 22, 2021, ISSN: 2072-4292. DOI: 10.3390/rs13224706.

[23] C.-H. Yeh, C.-H. Lin, L.-W. Kang, *et al.*, "Lightweight Deep Neural Network for Joint Learning of Underwater Object Detection and Color Conversion," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 11, pp. 6129–6143, 2022. DOI: 10.1109/TNNLS.2021.3072414.

[24] W.-H. Lin, J.-X. Zhong, S. Liu, T. Li, and G. Li, "RoIMIX: Proposal-Fusion Among Multiple Images for Underwater Object Detection," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2588–2592, 2020. DOI: 10.1109/ICASSP40776.2020. 9053829.

[25] L. Chen, Z. Liu, L. Tong, *et al.*, "Underwater Object Detection Using Invert Multi-Class AdaBoost with Deep Learning," in *International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2020. DOI: 10.1109/IJCNN48605.2020. 9207506.

[26] "Instant Deep Sea Debris Detection for Maneuverable Underwater Machines to Build Sustainable Ocean Using Deep Neural Network," *Science of The Total Environment*, vol. 878, pp. 162–826, 2023, ISSN: 0048-9697. DOI: https://doi.org/10.1016/j.scitotenv.2023.162826.

[27] F. Zocco, T.-C. Lin, C.-I. Huang, H.-C. Wang, M. O. Khyam, and M. Van, "Towards More Efficient EfficientDets and Real-Time Marine Debris Detection," *IEEE Robotics and Automation Letters*, vol. 8, no. 4, pp. 2134–2141, 2023. DOI: 10.1109/LRA.2023.3245405.

[28] J. R. Prasad, A. Parikh, and H. K. Prasanth, "Exploration of Deep Learning Based Underwater Image Processing Techniques," in *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1222–1225, 2023.

[29] G. Mellone, C. G. De Vita, D. D. Sánchez-Gallegos, *et al.*, "A Containerized Distributed Processing Platform for Autonomous Surface Vehicles: Preliminary Results for Marine Litter Detection," in *31st Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, 2023, pp. 144–148. DOI: 10.1109/PDP59025.2023.00029.

[30] J. Hong, M. Fulton, and J. Sattar, "TrashCan: A Semantically-segmented Dataset Towards Visual Detection of Marine Debris," *arXiv preprint arXiv:2007, .08097*, 2020.

[31] A. Kumar and S. Srivastava, "Object Detection System Based on Convolution Neural Networks Using Single Shot Multi-Box Detector," *Procedia Computer Science*, vol. 171, pp. 2610–2617, 2020, ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2020.04.283.

[32] J.-I. Watanabe, Y. Shao, and N. Miura, "Underwater and Airborne Monitoring of Marine Ecosystems and Debris," *Journal of Applied Remote Sensing*, vol. 13, no. 4, pp. 044–509, 2019.

[33] K. Sreekala, N. N. Raj, S. Gupta, G. Anitha, A. K. Nanda, and A. Chaturvedi, "Deep Convolutional Neural Network with Kalman Filter Based Objected Tracking and Detection in Underwater Communications," *Wireless Networks*, Mar. 2023, ISSN: 1572-8196. DOI: 10.1007/s11276-023-03290-z.

[34] Y. Song, B. He, and P. Liu, "Real-Time Object Detection for AUVs Using Self-Cascaded Convolutional Neural Networks," *IEEE Journal of Oceanic Engineering*, vol. 46, no. 1, pp. 56–67, 2021. DOI: 10.1109/JOE.2019.2950974.

[35] J. Liu, S. Liu, S. Xu, and C. Zhou, "Two-Stage Underwater Object Detection Network Using Swin Transformer," *IEEE Access*, vol. 10, pp. 117 235–117 247, 2022.

[36] A. Mathias, S. Dhanalakshmi, R. Kumar, and R. Narayanamoorthi, "Deep Neural Network Driven Automated Underwater Object Detection," *Computers, Materials & Continua*, vol. 70, no. 3, 2022.

[37] A. Mathias, S. Dhanalakshmi, R. Kumar, and R. Narayanamoorthi, "Deep Neural Network Driven Automated Underwater Object Detection," *Computers, Materials & Continua*, vol. 70, no. 3, 2022.

[38] R. Bhuvaneswari, T. Surya, T. Srikanth, and R. Balaji, "A Novel Approach for Underwater Object Detection Through Deep Intense-net for Ocean Conservation Systems," in *OCEANS 2022 Chennai*, IEEE, pp. 1–9, 2022.

[39] B. Gašparović, J. Lerga, G. Mauša, and M. Ivašić-Kos, "Deep Learning Approach for Objects Detection in Underwater Pipeline Images," *Applied Artificial Intelligence*, vol. 36, no. 1, p. 2 146 853, 2022.

[40] H. Wang and N. Xiao, "Underwater Object Detection Method Based on Improved Faster RCNN," *Applied Sciences*, vol. 13, no. 4, 2023, ISSN: 2076-3417.

[41] X. Li, F. Li, J. Yu, and G. An, "A High-precision Underwater Object Detection Based on Joint Self-supervised Deblurring and Improved Spatial Transformer Network," *arXiv preprint arXiv:2203.04822*, 2022.

[42] G. Chandrashekar, A. Raaza, V. Rajendran, and D. Ravikumar, "Side Scan Sonar Image Augmentation for Sediment Classification using Deep Learning Based Transfer Learning Approach," *Materials Today: Proceedings*, vol. 80, pp. 3263–3273, 2023.

[43] P. Song, P. Li, L. Dai, T. Wang, and Z. Chen, "Boosting R-CNN: Reweighting R-CNN samples by RPN's error for Underwater Object Detection," *Neurocomputing*, vol. 530, pp. 150–164, 2023.

[44] C. Li, C. Guo, W. Ren, *et al.*, "An Underwater Image Enhancement Benchmark Dataset and Beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2020. DOI: 10.1109/TIP.2019.2955241.

[45] Z. Zhang, Q. Tong, C. Yi, X. Fu, J. Ai, and Z. Wang, "The Appropriate Image Enhancement Method for Underwater Object Detection," in *IEEE 22nd International Conference on Communication Technology (ICCT)*, pp. 1627–1632, 2022. DOI: 10.1109/ICCT56141.2022.10073015.

[46] "The Extended Marine Underwater Environment Database and Baseline Evaluations," *Applied Soft Computing*, vol. 80, pp. 425–437, 2019, ISSN: 1568-4946. DOI: https://doi.org/10.1016/j.asoc.2019.04.025.

[47] J. Hong, M. Fulton, and J. Sattar, "TrashCan: A Semantically-Segmented Dataset Towards Visual Detection of Marine Debris," *arXiv preprint arXiv:2007*, 2020.

[48] M. Jian, X. Liu, H. Luo, X. Lu, H. Yu, and J. Dong, "Underwater Image Processing and Analysis: A Review," *Signal Processing: Image Communication*, vol. 91, p. 116 088, 2021.

[49] F. Xu, H. Wang, X. Sun, and X. Fu, "Refined Marine Object Detector with Attention-based Spatial Pyramid Pooling Networks and Bidirectional Feature Fusion Strategy," *Neural Computing and Applications*, vol. 34, no. 17, pp. 14 881–14 894, 2022.

[50] M. K. Moghimi and F. Mohanna, "Reliable Object Recognition Using Deep Transfer Learning for Marine Transportation Systems with Underwater Surveillance," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[51] M. J. Islam, C. Edge, Y. Xiao, *et al.*, "Semantic segmentation of underwater imagery: Dataset and benchmark," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 1769–1776, 2020.

[52] Q. Qi, Y. Zhang, F. Tian, *et al.*, "Underwater Image Co-enhancement with Correlation Feature Matching and Joint Learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1133–1147, 2022. DOI: 10.1109/TCSVT.2021.3074197.

[53] "Deep Learning-based Waste Detection in Natural and Urban Environments," *Waste Management*, vol. 138, pp. 274–284, 2022, ISSN: 0956-053X. DOI: https://doi.org/10.1016/j.wasman.2021.12.001.

[54] H. Yu, X. Li, Y. Feng, and S. Han, "Multiple Attentional Path Aggregation Network for Marine Object Detection," *Applied Intelligence*, vol. 53, no. 2, pp. 2434–2451, Jan. 2023, ISSN: 1573-7497. DOI: 10.1007/s10489-022-03622-0.

[55] L. Dai, H. Liu, P. Song, H. Tang, R. Ding, and S. Li, *Edge-guided Representation Learning for Underwater Object Detection*, 2023. arXiv: 2306.00440 [cs.CV].

[56] L. Dai, H. Liu, P. Song, and M. Liu, *A Gated Cross-domain Collaborative Network for Underwater Object Detection*, 2023. arXiv: 2306.14141 [cs.CV].

[57] G. Verma and M. Kumar, "Systematic Review and Analysis on Underwater Image Enhancement Methods, Datasets, and Evaluation Metrics," *Journal of Electronic Imaging*, vol. 31, no. 6, pp. 060 901–060 901, 2022.

[58] I. Srivastava, "Retraining of Object Detectors to Become Suitable for Trash Detection in the Context of Autonomous Driving," Ph.D. dissertation, Institute of Software, pp. 1–8, 2022. DOI: 10.13140/RG.2.2.21474.61128.

[59] S. Fayaz, S. A. Parah, G. Qureshi, J. Lloret, J. Del Ser, and K. Muhammad, "Intelligent Underwater Object Detection and Image Restoration for Autonomous Underwater Vehicles," *IEEE Transactions on Vehicular Technology*, 2023.

List of Papers Accepted

[1] AK Sinthia, AA Rasel, M Haque, "Real-time Submerged Debris Detection in Underwater Ecosystems using YOLOv8" in 2023 International Conference on International conference on computer and information technology (ICCIT), IEEE, 2023 (accepted)