Visual Detection of Marine Debris Using RTMDet

1st Aryya Bagus Padmanawijaya
Faculty of Informatics
Telkom University
Bandung, Indonesia
aryyawijaya@student.telkomuniversity.ac.id

2nd Gamma Kosala*
Faculty of Informatics
Telkom University
Bandung, Indonesia
gammakosala@telkomuniversity.ac.id

3rd Febryanti Sthevanie
Faculty of Informatics
Telkom University
Bandung, Indonesia
sthevanie@telkomuniversity.ac.id

***Abstract*— Marine debris, a hazardous threat to the marine environment, has emerged as a critical global environmental issue. Manual removal is the method most commonly applied to remove marine debris from the marine environment. One solution to this problem is to use deep learning-based visual marine debris detectors to detect marine debris automatically. However, in terms of marine debris detection, there are several challenges, one of which is because the complex marine environment makes visual detectors able to detect marine debris accurately and in real-time to avoid damaging the marine environment. In this research, three variants of the RTMDet (Real-Time Models for Object Detection) model were trained and evaluated using the TrashCan-Instance dataset. One of them is the RTMDet-l model with improvisation to replace the loss bbox in the head model with DIoU (Distance-Intersection over Union) and improvisation to add a sampling strategy with a Class-aware Sampling technique to handle the imbalanced data problem that has obtained mAP_{50} (Mean Average Precision at threshold 50%) accuracy of 71.3%. This has made the model's object detection accuracy on the TrashCan-Instance dataset the best while maintaining detection speed. These results prove that the model proposed in this study can be a vital consideration for further development in detecting marine debris. This contribution aims to address the global challenges related to marine debris and stimulate the development of more effective and efficient object detection models in complex marine environments.**

Keywords—object detection, marine debris, RTMDet, TrashCan

I. Introduction

Marine debris is one of the most critical global environmental issues. In 2016, about 11 million metric tons of plastic waste entered the oceans every year; if there is no immediate and sustainable action, the number will almost triple by 2040 to 29 million metric tons per year [1]. When the amount of marine debris exceeds its tolerance limit, the marine environment will be negatively affected. Therefore, efforts are needed to address this marine debris problem.

One solution to help solve this marine debris problem is to use visual marine debris detectors to detect marine debris and then remove it from the marine environment [2]. However, detecting marine objects or marine debris itself still poses several challenges. First, the marine environment is complex. Second, marine debris or objects in the marine environment are often small, obscured by other objects, and can change shape due to the surrounding environment. Therefore, visual marine debris detectors must be able to detect marine debris accurately in real time to clean up marine debris without damaging the marine environment.

Visual detection can be pursued through two main approaches: traditional methods that do not involve deep learning [3] and modern methods that rely on deep learning techniques [4]. Traditional methods often involve manually extracting image features and using conventional classification algorithms. While they have been successfully used in various contexts, they have limitations in handling high visual complexity and significant object variations. On the other hand, modern methods use deep learning approaches, utilizing complex artificial neural network architectures to automatically learn more abstract feature representations from visual data [5]. This allows modern detectors to be more adaptive to visual variations and complexities that may be difficult for traditional methods to accommodate.

Visual detection of marine debris, especially using deep learning, has evolved over the past few years. M. Fulton [4] has tested several deep-learning object detectors on a published self-constructed dataset called Trash-ICRA19. With the results obtained, the study stated that it is possible for visual detection using deep learning to detect in real time. With the same dataset, C. M. Wu [6] focused on the problem of deep-learning visual detectors on mobile devices. It introduced the YOLOv5 deep learning model by changing its backbone to MobileNetv3. With reasonable accuracy, this change can meet the device's limitations, which are network calculation and memory. However, the datasets used from both studies only have three classes: plastic, ROV (Remotely Operated Vehicle), and bio. This causes the model's performance based on the dataset to under-represent the detection accuracy in the marine environment, which is more complex when detecting marine debris or objects.

J. Hong [7] published the TrashCan dataset, consisting of several classes of trash, ROV classes, and underwater flora and fauna. The dataset provides a better representation of the complexity of the actual marine environment. The study also provided a baseline model for object detection, namely Faster R-CNN with ResNetXt-101-FPN backbone, which resulted in mAP_{50} on TrashCan-Instance and TrashCan-Material being 55.3% and 54%.

Using the TrashCan dataset, H. Deng [8] introduced several methods to the Mask R-CNN model based on Faster R-CNN to produce higher accuracy. The first method is to add dilated convolution to the Feature Pyramid Network (FPN) structure to improve the model's ability to detect small objects. The second method uses spatial-channel attention so that the features can be trained adaptively. The model's performance in object detection resulted in a mAP_{50} of 65%.

W. Zhou [9] introduced YOLOTrashCan, which is a YOLOv4-based model, by introducing the ECA_CO-Conv_CSP backbone to extract features from marine debris and introducing DPMs_PixelShuffle_PANET as a feature fusion module to improve detection capabilities on multiscale marine debris. On TrashCan-Instance and TrashCan-Material datasets, the model produced object detection accuracy with mAP_{50} of 65.01% and 58.66% with FPS of 36.17 and 36.75, respectively.