*AI3001: Deep Learning*

# Marine Debris Detection using Deep Learning

Ayush Bhatnagar, Utkarsh Agarwal, Tanushi Kh., Abhiraaj Sharma, Dhirain V.

October 12, 2024

## 1 Introduction

Marine debris poses a significant global threat to aquatic wildlife, leading to harmful outcomes such as entanglement, ingestion, infection, and habitat destruction. Each year, approximately 8–10 million metric tons of plastic waste enter our oceans, mainly from inland waterways like rivers and canals. This influx not only endangers marine life but also contributes to climate change and deteriorates water quality, disrupting entire ecosystems and food chains. Traditional methods of controlling floating waste are labor-intensive and inefficient, particularly in complex geographical areas.

The emergence of unmanned surface and aerial vehicles (USVs and UAVs) offers more efficient, cost-effective alternatives through automation. However, the success of these systems heavily relies on robust, real-time data for floating waste detection. Challenges such as small object sizes, low brightness, reflections, and cluttered environments make detection difficult. Recent advances in deep learning have significantly improved object detection tasks, allowing models to adapt feature representations from large datasets. Despite these advancements, small objects and complex image conditions continue to pose hurdles, calling for enhanced feature representation and consolidation of information from diverse resources.

Our objective is to develop a robust floating litter object detection system which uses deep learning techniques. By focusing on improving detection accuracy for small objects under challenging environmental conditions, we aim to enhance the capabilities of autonomous vehicles like USVs and UAVs in monitoring and managing floating waste.

## 2 Dataset(s)

All datasets used in this work have been collected from various sources, each contributing unique environmental and object conditions. All datasets are annotated for object detection, with RGB images as input and bounding boxes labeling the objects of interest, ensuring consistency across datasets.

### 2.1 FloW-IMG

The FloW-Img sub-dataset contains 2000 images with more than 5000 labeled floating wastes. Small objects (occupying less than 32 × 32 pixels) account for the largest proportion. Data augmentation techniques, such as brightness adjustments and random rotations, were applied to increase the model's robustness to environmental variability.

Figure 1: Sample dataset image with labeled debris.

## 2.2 FloatingWaste-IN

This dataset contains 1,867 images collected at midday, evening, and night to capture variations in lighting conditions. Object reflections were excluded from the labeled area to avoid misidentification.

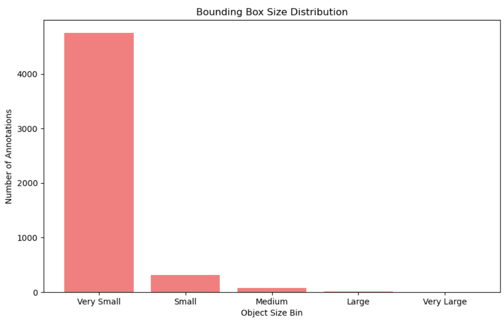


Figure 2: Bin Size Distribution: Floatingwaste-IN Dataset

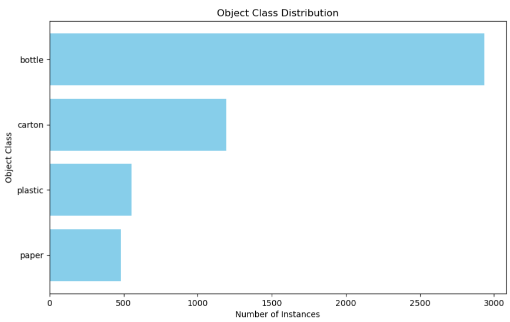


Figure 3: Object Class Distribution: FloatingWaste-IN dataset

## 2.3 Floating Waste Dataset

The dataset consists of 8,227 images with waste annotations provided in COCO format. Each image underwent pre-processing, including auto-orientation of pixel data. For data augmentation, three versions of each source image were created by applying the following transformations: an equal probability of 90-degree rotations (none, clockwise, counter-clockwise, upside-down), random brightness adjustments ranging from -21 to +21, and the application of random Gaussian blur between 0 and 1.25 pixels to simulate varying environmental conditions and improve model robustness.

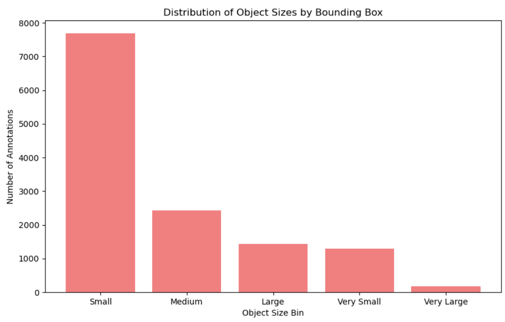


Figure 4: Object Size Distribution: Floating Waste Dataset

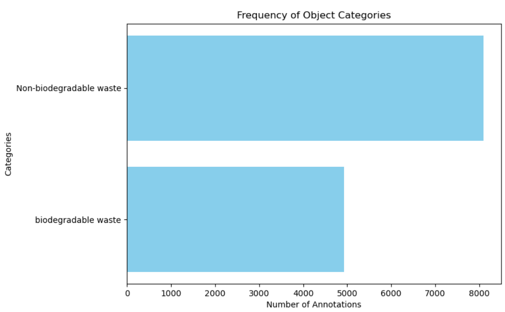


Figure 5: Object Frequency Distribution

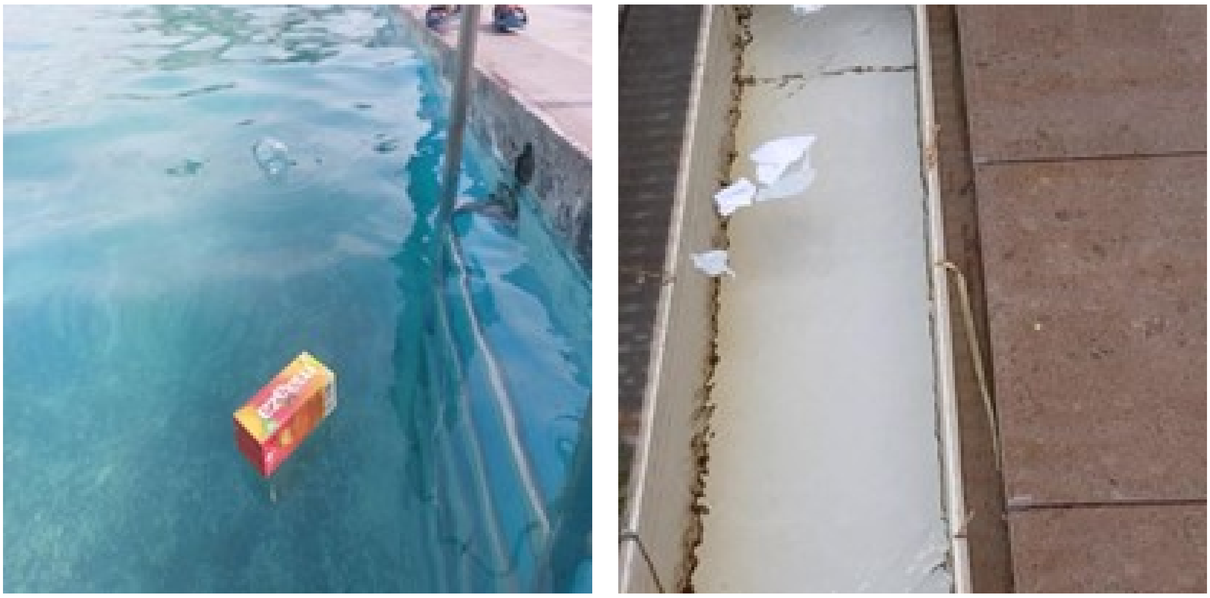

Figure 6: Sample dataset image

## 2.4 Floating Plastic

The dataset includes 1,545 images depicting various plastic packets, such as plastic bags, wrappers, and intertwined plastic waste, floating through a narrow gutter. These images capture challenging real-world scenarios where plastic debris becomes tangled or overlapped as it moves through confined waterways. No data augmentation has been applied.

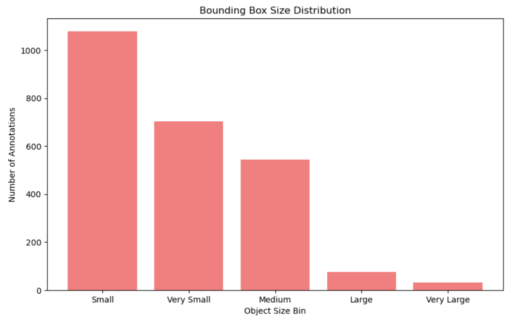


Figure 8: Bounding box size distribution

Figure 7: Sample dataset image

## 2.5 TUD-GV

The TUD-GV (TU Delft-Green Village) dataset consists of 9,473 images at 1080p resolution. The images feature 626 litter items (plastic bottles, bags, metal tins, paper, cardboard) under sunny and cloudy weather conditions, captured from two heights (2.7 m and 4.0 m) and two angles (0° and 45°). Images were labelled into four categories based on litter quantity: no litter, little litter, moderate litter, and lots of litter.


Figure 9: Label classifications


Figure 10: Class distribution according to time

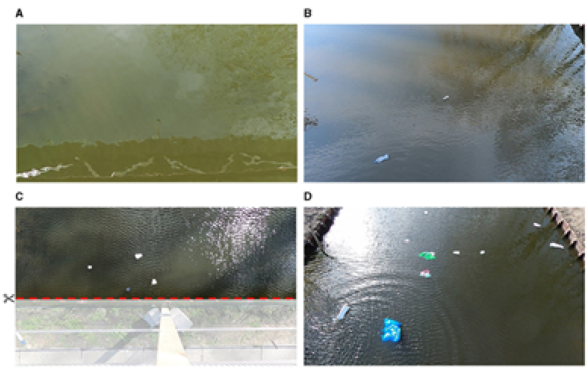


Figure 11: Sample images from the TUD-GV dataset

# 3 Literature Review

## 3.1 Detection of River Floating Garbage Based on Improved YOLOv5

The paper introduces the YOLOv5_CBS model, an enhanced version of YOLOv5 designed for detecting floating garbage in river environments. The model integrates three key improvements: a Coordinate Attention Mechanism to better capture spatial dependencies and channel relationships, a Bidirectional Feature Pyramid Network (BiFPN) for improved multi-scale feature fusion, and the SCYLLA-IoU loss function, which refines bounding box regression by incorporating additional geometric considerations. Compared to existing detection methods, YOLOv5_CBS achieves higher detection accuracy of 91.6%, particularly better for smaller objects, demonstrating the benefits of the proposed modifications.

## 3.2 Detection of Floating Garbage on Water Surface Based on PC-Net

This paper presents PC-Net, an algorithm designed to detect floating garbage on water surfaces by addressing the limitations of conventional two-stage object detection models. It introduces a novel Pyramidal Anchor Generation Approach that improves anchor placement by accounting for the size and position of floating objects, reducing background interference and increasing detection precision. Additionally, the Classification Discrimination Diagram enhances the resolution of feature maps during Region of Interest (RoI) pooling, which improves the classification accuracy of small and hard-to-detect targets.

## 3.3 C2f_MLCA Backbone for Enhanced Floating Waste Detection

The paper introduces key improvements to the YOLOv8 model for detecting floating waste. It integrates the C2f_MLCA backbone, which captures both local and global contextual information, improving robustness against environmental challenges like wave movements and reflections. The use of SENetV2 further enhances detection by emphasising important feature channels, aiding in the detection of small and obscured objects. Additionally, the SIoU loss function improves bounding box accuracy by refining spatial alignment and aspect ratio. These innovations result in a 5% increase in detection accuracy on the FloW-Img dataset.

## 3.4 Autonomous, Onboard Vision-Based Trash and Litter Detection in Low Altitude Aerial Images Collected by an Unmanned Aerial Vehicle

The paper presents a UAV-based solution for autonomous trash and litter detection utilizing convolutional neural networks (CNNs). A key innovation lies in the system's capability for real-time inference on embedded platforms, such as Nvidia Xavier NX, where YOLOv4 is identified as the most efficient model. YOLOv4 strikes a balance between accuracy and computational speed, achieving a mean average precision (mAP) of 0.785 with a processing rate of 5-6 frames per second (FPS), which happens to be state-of-the-art. The dataset, UAVVaste, consists of 772 annotated aerial images capturing a range of urban and natural settings. However, a significant limitation is the degradation in detection accuracy when models are quantized for lower precision, along with difficulties in detecting smaller objects in complex environments.

## 3.5 Deep Learning-Based Waste Detection in Natural and Urban Environments

The paper explores CNN architectures, specifically YOLO and EfficientDet, for detecting waste across both natural and urban landscapes. This study emphasizes optimizing the models' generalization ability across diverse environmental contexts. YOLO models demonstrated superior detection precision, particularly in differentiating waste from natural objects. However, specific quantitative results such as mAP are not reported. The dataset used contains images from varied environments, offering versatility, but its relatively small size poses challenges, particularly in detecting smaller or occluded waste objects. The main limitations include the need for larger datasets and further model enhancements to handle the complexities of real-world scenarios effectively.

## 3.6 Advancing Deep Learning-Based Detection of Floating Litter Using a Novel Open Dataset

The study introduces a new dataset, the TU Delft—Green Village dataset, consisting of 9,473 labelled images captured from urban canals to detect floating litter using deep learning models. The authors compared five CNN architectures—ResNet50, DenseNet121, InceptionV3, MobileNetV2, and SqueezeNet—and employed transfer learning to enhance detection performance. DenseNet121 and SqueezeNet emerged as the best performers. Fine-tuning all layers (FTAL) significantly improved model performance compared to fine-tuning only the classifier. The study also found that image flipping provided the most notable improvements. However, the models struggled to generalise well to new litter types and different environmental setups, indicating the need for more diverse datasets. The authors highlight limitations such as the dataset's semi-controlled environment, lack of nighttime images, and the exclusion of complex object detection tasks, which limits the real-world applicability of the models for large-scale litter detection. Moreover, tiling images into smaller patches (e.g., 224*224) would likely boost performances by retaining the original image quality, although this would require relabeling all tiles.

## 3.7 Detecting Floating Litter in Freshwater Bodies with Semi-Supervised Deep Learning

The paper introduces a semi-supervised learning approach to address the challenge of limited labelled data for detecting floating litter in aquatic environments. The authors utilise SwAV, a self-supervised contrastive learning method, to pre-train a ResNet50 model on 100,000 unlabeled images, followed by fine-tuning with a Faster R-CNN architecture using a smaller labelled dataset of 1,800 images. The model was tested across various geographic locations, including the Netherlands, Indonesia, and Vietnam, to evaluate its generalisation capability. Results showed that the semi-supervised method outperformed fully supervised approaches in scenarios with limited labelled data, and demonstrated improved generalisation to unseen environments. Data augmentation, particularly image flipping, further enhanced model robustness. However, the study's limitations include the exclusion of nighttime images, the lack of vegetation and natural debris in the dataset, and the simplified focus on a narrower range of detection tasks rather than more advanced techniques like segmentation, which are crucial for real-world applications in tracking and quantifying floating litter.

### 3.8 A Floating-Waste Detection Method for Unmanned Surface Vehicles Based on Feature Fusion and Enhancement

The paper employs unique methods for detecting floating marine debris by addressing key challenges in object recognition. The authors introduce a Low-Level Representation-Enhancement Module, which mitigates the loss of critical features in low-resolution images caused by downsampling in YOLOv5, ensuring that object location information is retained. Additionally, they implement an Attention Fusion Module that effectively integrates low- and high-resolution feature maps by leveraging self-attention mechanisms, allowing the model to effectively capture detailed features while maintaining robust object detection.

### 3.9 Improved YOLOv8 Algorithm for Water Surface Object Detection

This paper by Wang & Zhao (2024) introduces several unique improvements to the YOLOv8 model. A key innovation is the incorporation of the C2f_MLCA Backbone (Cross-Stage Partial connections with Multi-Level Context Aggregation), which enhances feature extraction by effectively capturing both local and global contextual information. Additionally, the integration of SENetV2 (Squeeze-and-Excitation Networks version 2) recalibrates channel-wise feature responses to focus on the most relevant features, significantly improving the detection of smaller and obscured objects. Another important contribution is the use of the SIoU Loss Function (Spatial Intersection over Union), which improves bounding box accuracy by considering not just the overlap between predicted and ground-truth boxes but also their spatial alignment and aspect ratio. These innovations result in a 5% increase in detection accuracy on the FloW-Img dataset, demonstrating the state-of-the-art model for the FloW dataset with mAP@0.5 of 87.9%.

## 4 Limitations

### 4.1 Sensitivity to Lighting and Reflections

Current deep learning models remain vulnerable to fluctuations in lighting and reflections on water surfaces, despite employing augmentation techniques like brightness adjustments and night filters. This limitation challenges the real-time applicability of these models, particularly for autonomous systems like unmanned surface vehicles (USVs) and unmanned aerial vehicles (UAVs).

### 4.2 Adverse Weather and Water Conditions

Existing models struggle to adapt to environmental factors like wave movement and water turbidity, limiting their performance in natural settings.

### 4.3 Overfitting in Specific Environments

Deep learning models trained on datasets collected from limited geographical or environmental contexts often struggle to generalise to unseen settings. This issue of overfitting is exacerbated by the absence of diverse, large-scale datasets that encompass a wide range of real-world environments, limiting the models' effectiveness in practical applications.

### 4.4 Lack of Segmentation for Complex Object Detection Tasks

Various types of debris coexist with natural elements, segmentation can significantly enhance detection capabilities. It allows models to delineate the boundaries of objects more accurately, enabling a finer level of detail in identifying and classifying floating debris.

### 4.5 Class Imbalance

EDA on the available datasets show that these datasets are inherently imbalanced, with smaller debris (e.g., plastic wrappers, bottle caps) far outnumbering larger objects (e.g., bottles, bags). This imbalance often leads to biased learning, where models are more attuned to detecting these smaller, more common objects.

## 5 Methodology and Novel Contributions

To solve the limitations that were observed in the existing literature, we thought of the following advancements and ideas for our future trajectory of the project

1. Creating More Realistic Training Data with Sunlight and Reflection Effects: We augment the dataset by simulating sunlight and reflection effects using GANs and image processing techniques. These effects, including glare, reflections, and lens flare, are introduced to mirror real-world conditions, helping the model generalise to environments where litter is obscured by lighting. By exposing the model to these complex lighting variations during training, we aim to improve its detection capabilities under challenging conditions.

2. Using Focal Loss to Improve Detection of Small Objects: To address the challenge of detecting small litter objects in complex environments, we integrate focal loss into our detection model. Focal loss down-weights easy examples and emphasises harder ones, allowing the model to focus on difficult-to-detect objects like small plastic fragments. This loss function enhances the detection of small and obscured objects, which are common in water bodies with cluttered backgrounds or reflections. By improving the detection of these smaller objects, we expect higher overall accuracy, particularly on datasets like FloW-IMG, where small objects make up a large proportion of the data.

3. Segmenting the Water Surface Before Detecting Litter We propose a two-step process: first, segment the water surface using models like U-Net or DeepLab to identify regions we are certain are water. Then, we process these water areas to reduce their visual prominence, making them less distracting in the images. This approach enhances detection accuracy by reducing distractions caused by water surface features, particularly in complex visual environments where traditional models struggle due to background noise.

4. Detecting Sunlight and Reflections to Enhance Litter Detection: We introduce a separate model to detect sunlight glare and reflections on the water surface. Since existing datasets lack labels for such conditions, we will annotate a subset of images ourselves and employ semi-supervised learning to expand these annotations. This model will identify regions affected by strong sunlight or reflections, allowing us to adjust or mask these areas during pre-processing. By preventing the detection model from being misled by glare, this approach improves overall performance in high-glare environments, reducing false negatives caused by obscured litter in bright or reflective areas of the water.

# References


[1] W. Jiang et al., "FloW-Img: A Floating Waste Image Dataset for Object Detection," *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Kyoto, Japan, 2022. doi: `10.1109/IROS2022.9710581`.

[2] R. Wang et al., "Floating Waste Dataset: Detecting Floating Wastes in Urban Waterways", GitHub, 2021. `https://github.com/wangruichen01/FloatingWaste-IN`.

[3] A. Thung and M. Yang, "TrashNet: Dataset for garbage classification," Stanford CS 229, 2016. https://github.com/garythung/trashnet

[4] Y. Cheng, et al., "FloW: A Dataset and Benchmark for Floating Waste Detection in Inland Waters," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021. https://ieeexplore.ieee.org/document/9710581

[5] M. Kraft, M. Piechocki, B. Ptak, and K. Walas, "Autonomous, Onboard Vision-Based Trash and Litter Detection in Low Altitude Aerial Images Collected by an Unmanned Aerial Vehicle," *Remote Sens.*, vol. 13, no. 965, 2021. https://doi.org/10.3390/rs13050965

[6] Y. Li, R. Wang, D. Gao, and Z. Liu, "A Floating-Waste-Detection Method for Unmanned Surface Vehicle Based on Feature Fusion and Enhancement," *Journal of Marine Science and Engineering*, vol. 11, no. 12, 2023, p. 2234. `https://doi.org/10.3390/jmse11122234`.

[7] S. Majchrowska, A. Mikołajczyk, M. Ferlin, Z. Klawikowska, M. A. Plantikow, A. Kwasigroch, and K. Majek, "Deep learning-based waste detection in natural and urban environments," *Waste Management*, vol. 138, pp. 274-284, 2022. `https://doi.org/10.1016/j.wasman.2021.12.001`.

[8] J. Wang and H. Zhao, "Improved YOLOv8 Algorithm for Water Surface Object Detection," *Sensors*, vol. 24, no. 15, 2024, p. 5059. `https://doi.org/10.3390/s24155059`.

[9] T. Jia, A. J. Vallendar, R. De Vries, Z. Kapelan, and R. Taormina, "Advancing deep learning-based detection of floating litter using a novel open dataset," *Frontiers in Water*, vol. 5, 2023. `https://doi.org/10.3389/frwa.2023.1298465`.

[10] T. Jia, R. de Vries, Z. Kapelan, T. H. M. van Emmerik, and R. Taormina, "Detecting floating litter in freshwater bodies with semi-supervised deep learning," *Frontiers in Water*, vol. 5, 2024. `https://doi.org/10.1016/j.watres.2024.122405`.