Cluster Computing (2025) 28:386
https://doi.org/10.1007/s10586-024-05088-w

Aqua garbage collector: utilizing AI and IoT for efficient underwater garbage classification

Bonnie Yam[1] · Ruochen Zhao[1] · Maher Elshakankiri[1]

Received: 14 July 2024 / Revised: 11 November 2024 / Accepted: 26 December 2024 / Published online: 16 June 2025

Abstract

The issue of solid wastes and contaminants in marine ecosystems has increasingly become a critical environmental challenge. In the previous study, our team focused on the impacts of floating plastic and how to capture them using innovative robotic solutions and categorize them into specific types. Building upon this foundation, we will broaden our scope to include a wider variety of underwater garbage. Driven by Artificial Intelligence (AI) and the Internet of Things (IoT), the primary goal of our robot Aqua Garbage Collector (AGC) is to monitor water pollution in real-time through camera modules by classifying and retrieving trash across various water depths. It will be programmed to send photos and coordinates of the detected trash to the central hub computer. Three machine learning models (YOLOv5, NanoDet, and RT-DETR) are studied to evaluate which model has the highest processing speed and accuracy. The results showed that YOLOv5 had the highest mean Average Precision (mAP), proving its real-time efficiency and accuracy. RT-DETR, while demonstrating good accuracy in complex cases, required longer training and showed lower mAP compared to YOLOv5. NanoDet, though highly efficient in computation, results in the lowest precision. The challenges of transmitting data underwater and over long distances are considered. Additionally, the model accuracy may be affected by the variability in water conditions, such as changes in turbidity and light penetration. This can result in a decrease in detection accuracy when operating in different underwater environments.

Keywords Aquatic pollution · Internet of things · Machine learning · Convolution neural networks · Artificial intelligence · Image detection

1 Introduction

Driven by technological advancements, the rapid expansion of industrialization led to a massive increase in manufacturing facilities and consumerism. Evidently, human activities are the primary sources of water pollution. Unfortunately, people carelessly or intentionally litter on the ground. Sometimes, the garbage gets blown or carried away into the waterways by the wind or rain. There are about 8–10 million metric tons of plastic that enter the ocean each year, and by 2050, fish will likely be outweighed by plastic [1]. This is a serious concern that not only contaminates the environment but also poses health risks to both aquatic animals and humanity. On top of disposing of trash responsibly after usage, we can leverage technology to clean water bodies expeditely. By integrating both the Internet of Things (IoT) and Artificial Intelligence (AI), we can design a system that monitors water pollution in real time, classifies trash, collects it, and stores data.

Our teammates, De La Garza et al. [2], began with a simple user interface that analyzed and labelled the type of garbage in the users' contributed photographs. The classification process is executed with assistance from external datasets and an algorithm: Trash Annotation in Context (TACO), Common Objects in Context (COCO), and Mask Regional Convolutional Neural Network (Mask R-CNN) [2]. Alongside these AI applications, Kelly et al. [3] incorporated IoT sensors and connectivity in their pollution monitoring system. While the robot explores and searches for

✉ Bonnie Yam
bonbon.yam@alumni.utoronto.ca

Ruochen Zhao
ruochenn.zhao@mail.utoronto.ca

Maher Elshakankiri
maher.elshakankiri@utoronto.ca

[1] University of Toronto, Toronto, Canada

floating plastics, the central hub computer performs more complex tasks: processing footage and storing findings in the cloud [3]. The current paper builds on the groundwork laid by the preceding publications. While they focused on assessing and removing floating garbage, we want to extend its capability to collect garbage underwater. To incorporate this functionality, we introduced new hardware that enhances the robot's agility and garbage detection capabilities. The key objective of our robot Aqua Garbage Collector (AGC) is to classify and retrieve trash across various water depths efficiently. We investigated whether items can be clearly identified and accurately categorized underwater by AI with the following machine learning models: YOLOv5, NanoDet, and RT-DETR. Despite the limitations relevant to the communication between AGC and the central hub in the open waters and unstable underwater conditions, we hope to expand our research in mechanics and aquatic environments to build an efficient prototype.

This paper is organized as follows: Sect. 2 explores related works, including the adverse effects of plastic pollution, machine learning models that analyze images, and IoT solutions for garbage cleanup. Section 3 describes the hardware of AGC, followed by its behavioural process in Sect. 4. The results of analyzing underwater photos between three machine learning models are compared in Sect. 5. The limitations of these models and AGC are covered in Sect. 6. Finally, we conclude the paper in Sect. 7 and discuss future work in Sect. 8.

2 Related works

Recent studies have underscored the significant environmental impact of marine pollution, particularly the negative effects of underwater waste on aquatic ecosystems. Understanding the complex impact of garbage on marine life is crucial for processing environmental problems. For instance, Zimmermann et al. [4] have looked into the complex interactions between microplastics and life underwater, indicating how both the chemical composition and physical characteristics of microplastics lead to their toxicity. There is compelling evidence that plastic waste accumulation in marine environments causes severe consequences. In their study, Lebreton et al. [5] provided substantial proof of the rapid accumulation of plastic trash in the North Pacific Ocean by reporting approximately 79,000 tonnes of floating plastics within the area.

In addition to marine environments, freshwater ecosystems, particularly those in urban settings, are increasingly recognized as containers of microplastic pollution. Grbić et al. [6] have conducted a study on microplastics in northwestern Lake Ontario. Findings in their research pointed out that microplastic concentrations were substantially higher in urban runoff and wastewater, which indicates the urgent need for effective waste management solutions. Similarly, research on the Laurentian Great Lakes has revealed the pervasive presence of microplastics in freshwater as well [7]; these wastes not only exist on the surface but also within benthic sediments, with concentrations often exceeding those found in oceans. It also revealed that various species within the Great Lakes are ingesting microplastics, leading to substantial food chain risks. These findings from both marine and freshwater studies underline the need for advanced technologies capable of precise detection and monitoring of underwater pollution.

Our team's prior research [3] utilized an integrated computer vision model trained on the TACO dataset [8], known for its robust annotations across diverse environments. This model uses the instance segmentation capabilities of Mask R-CNN, a method that is good at identifying and classifying macro-plastics in freshwater settings. Hu et al.'s work [9] on automated book recognition through Convolutional Neural Networks (CNN) has underscored the effectiveness of CNNs for high-precision classification tasks, showing the model's adaptability in diverse contexts such as object identification in environmental and technical domains. This approach helps demonstrate the practical applications of computational models for environmental protection. Additionally, Moya et al. [10] have applied YOLOv5 for crop detection and maturity classification, which exemplifies how advanced machine learning models can be used for real-time classification in challenging environments, such as agriculture. This offers insights that coincide with our use of similar models in underwater settings. Worachairungreung et al. [11] further demonstrated the utility of image classification in environmental monitoring, using satellite-based supervised classification methods to detect changes in land use and agricultural land loss over time. In addition, De La Garza et al. [2] applied both the TACO and Aqua-Trash [12] datasets to enhance the accuracy of trash detection technologies underwater. By applying sophisticated machine learning models and well-annotated datasets, these works provide valuable insights into more effective strategies for pollution intervention.

There are several IoT solutions for garbage collection on land and water. Jiang [13] utilized an AI visual identification system, YOLOv3, in her robot to identify the type and size of the garbage on the ground. The robot consists of a laptop that connects to a camera, and uploads code to Arduino Mega 2560. The speed and direction of the robot's mecanum wheels depend on the location of the identified trash in the photo. If the garbage is within a specific size limit, it will be vacuumed. Otherwise, the photo will be sent to their operations center.

Naik et al. [14] built their AI-based beach cleaning robot with a conveyor belt that collects and dumps litter into its secured bin. Once their ultrasonic sensor detects an object and the Raspberry Pi camera photographs it, the Open-Source Computer Vision Library (OpenCV) and Tensor-Flow analyze the photo to determine whether the detected object is garbage.

Digda et al. [15] remote-controlled Ro-Boat is not only limited to collecting trash but also sorting them according to its type (metal, non-metal, and organic waste) for proper disposal. Their sorting system has an Arduino UNO as the microcontroller, which connects to the infrared and proximity sensors that can help identify the type of garbage embarking onto the Ro-boat. They were mindful of the sea creatures by using a conveyor belt to avoid unintentional capture of wildlife, and they were considerate of the environment by having solar panels as the power source.

Dunggat et al. [16] developed an unmanned surface vehicle (USV) garbage collector that comprised four subsystems (obstacle avoidance, coordination detection, live streaming, and movement), which are either operated remotely or autonomously. The Arduino UNO programmed four waterproof ultrasonic sensors to detect nearby objects and determine the obstacle distance for the USV to move away from it. The Raspberry Pi 4 Model B is connected to a GPS module GY-NEO6MV2 for retrieving the latitude and longitude, as well as to a camera module for video stream. USV's speed and direction of motion are regulated by the electronic speed controller and the servo motor.

All of these inventions' goals involved cleaning aquatic environments, but most did not cover details of how it would record and save observations. Data collection is a significant task for AGC, as we hope to share this information with environmental protection agencies and the public. Raising awareness of the global issue promotes support for waste management regulations or practices.

3 AGC hardware

Plastic pollution can be present from the surface to the floor of the body of water. Thus, we expanded the behaviour of AGC to search for garbage underwater. The block diagram of the AGC prototype is visualized in Fig. 1. The subsequent components that were previously proposed by our team remain the same: Arduino UNO, SX1278 LoRa module, and SIM28 GPS Shield [3]. However, we changed the type of battery and camera module, as well as introduced more hardware to streamline the ability to detect and capture footage of marine debris.

1. *Arduino UNO*: It is the most prominant microcontroller board, which makes it favourable for an early prototype. There is an Atmega328 processor, 14 digital input/output pins, 6 analog inputs, 2 KB of RAM, and 32 KB of program memory [17].
2. *SX1278 LoRa module*: This low-power transceiver "provides ultra-long range spread spectrum communication and high interference immunity" [18], which is beneficial for transmitting data between AGC and headquarters.
3. *SIM28 Arduino GPS Shield*: We want to track the locations of waste and monitor where AGC is travelling, which are expressed in terms of latitude and longitude.

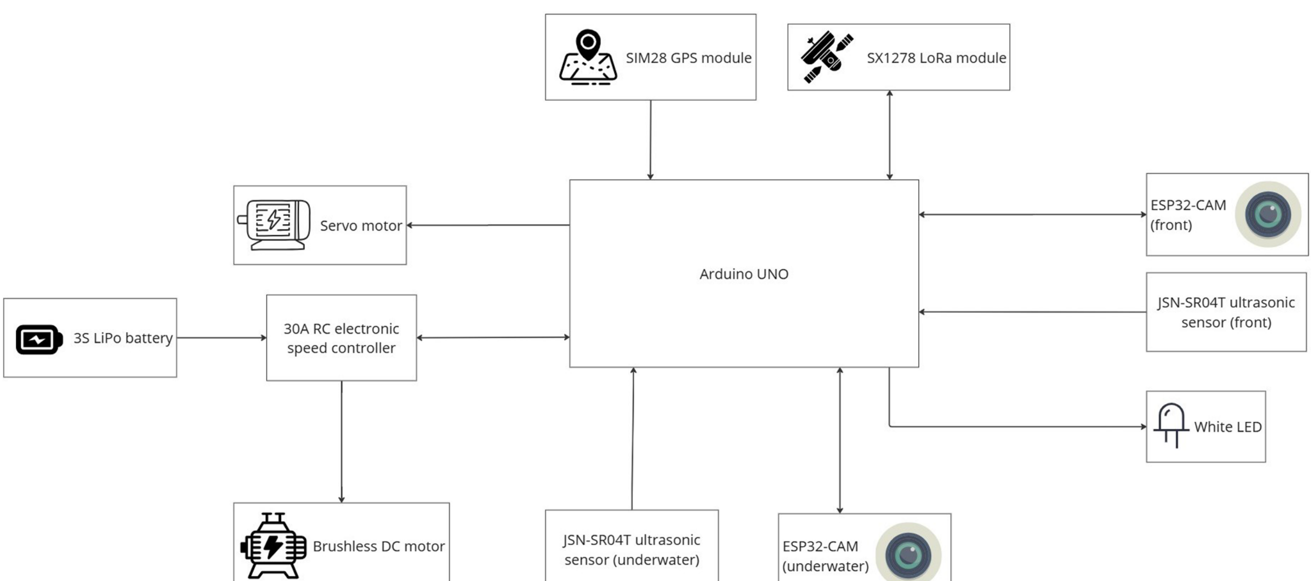

Fig. 1 Block diagram of AGC

The SIM Arduino GPS shield supports 22 tracking/66 acquisition channels, which enhance positioning accuracy and the speed of acquiring the position. Their sensitivity of tracking and acquisition can reach up to -167 dBm and −160 dBm, respectively [19].

4. *ESP32-CAM*: Arduino UNO has a small storage space and limited processing power. Thus, it may not be capable of streaming live and archiving images. This can be resolved with an ESP32-CAM that has an internal 520 KB SRAM, external 4MPSRAM, and a TF card slot [20]. Images of litter and the associated positional data can be saved in the TF card as a backup if there is no signal. AGC will have two ESP32-CAMs: one camera will be placed above the waterline and on the front side of AGC to find floating garbage, whereas the second camera will be attached at the bottom of AGC to spot garbage underwater. Both cameras will be protected with a waterproof case. The front camera will also have a shade cover to block the sun's glare.
5. *JSN-SR04T ultrasonic sensor*: It is a single water-resistant ultrasonic transducer that can transmit and receive ultrasonic waves to detect objects in the range between 25 cm and 450 cm [21]. It can be used to calculate the distance between the garbage and the sensor. Similarly to the cameras, the ultrasonic sensor will also be positioned on the front and the bottom of AGC.
6. *White LED*: Natural light from the sun may not be enough for the camera to capture the contrast beneath the water surface, especially when AGC descends deeply. With LEDs enclosed with a waterproof case underneath AGC, the machine learning model can identify the detected garbage clearly from the live footage.
7. *Brushless DC motor (1000 KV)*: It serves as the engine of AGC.
8. *3S LiPo Battery*: This battery will be used to power the brushless DC motor. It contains three cells, each has a voltage of 3.7 V, resulting in a total of 11.1 V. Since the motor is 1000 KV, then the motor can achieve a maximum of 11,100 revolutions per minute when there is no load [22].
9. *30A RC Electronic Speed Controller (ESC)*: In addition to regulating the speed of the motor, the ESC's built-in battery eliminator circuit provides 5 V that can help power the Arduino UNO. The ESC interprets the signals received from the Arduino UNO based on their pulse duration. The longer the pulse duration, the faster the motor's rotor spins [22].
10. *Servo motor*: This motor is known for controlling joint movements. Thus, the rotational angular displacement of the servo motor's shaft [23] can prompt the rudder to move left or right.

4 The behaviour of AGC

Prior to releasing AGC into the water, it is vital to check that the battery level is sufficient and the GPS is providing precise readings so we can easily pinpoint its location if any accidents arise. Given that the Arduino UNO is incapable of storing the machine learning model for analysis and the Neuroevolution of Augmenting Topologies (NEAT) algorithm for AGC to roam in unpredictable environments, this issue can be resolved through Hadoop clusters [3]. As shown in Fig. 2, the first action of AGC is to explore the area within 4.5 m to search for garbage. While the camera is streaming live, the machine learning model will analyze the footage and check whether the detected object matches any of the annotated images of garbage. If both cameras detect garbage simultaneously, then AGC will approach the nearest garbage. The distance between the ultrasonic sensor and the trash depends on the speed of sound, along with half of the time for the ultrasonic wave to hit the trash and return back to the sensor. For the ultrasonic sensor above the waterline, we will use the speed of sound in air, which is 344 m/s or 0.0344 cm/μs at 20 °C [24]. The underwater ultrasonic sensor will rely on the speed of water at 20 °C: 1461 m/s or 0.1461 cm/μs [24]. The formula d_1 calculates the distance between the front ultrasonic sensor and the garbage, whereas d_2 calculates the distance between the underwater ultrasonic sensor and the garbage.

$$d_1 = 0.0344\ \mathrm{cm}/\mu\mathrm{s} * \tfrac{1}{2}\mathrm{time}\ (\mu\mathrm{s})$$
$$d_2 = 0.1461\ \mathrm{cm}/\mu\mathrm{s} * \tfrac{1}{2}\ \mathrm{time}\ (\mu\mathrm{s})$$

If the garbage is detected on the water's surface, AGC will swim towards the garbage. Once the trash is 50 cm away from the sensor, AGC will swiftly obtain the coordinates and snap a picture, which will be sent to the central hub. The machine learning model will classify the type of garbage and tag the coordinates on the photo. If multiple garbage are clustered together, then there will be different labels that identify each one. All images will be stored in the database to generate statistics on which types of garbage are found and which areas are heavily polluted.

A similar procedure will be performed if the garbage is detected underwater, albeit with few variations. With the help of the underwater ultrasonic sensor, we can track the depth of the garbage using the formula d_2. AGC will get the coordinates before submerging in the water due to the potential challenge of the GPS radio frequencies not penetrating underwater [25]. Depending on the strength of the water current, the coordinates may be slightly different by the time AGC dives closer to the garbage. Nonetheless, this is not a critical concern because the estimated coordinates are sufficient information to get a general idea of which areas of the water are heavily polluted. Rather than

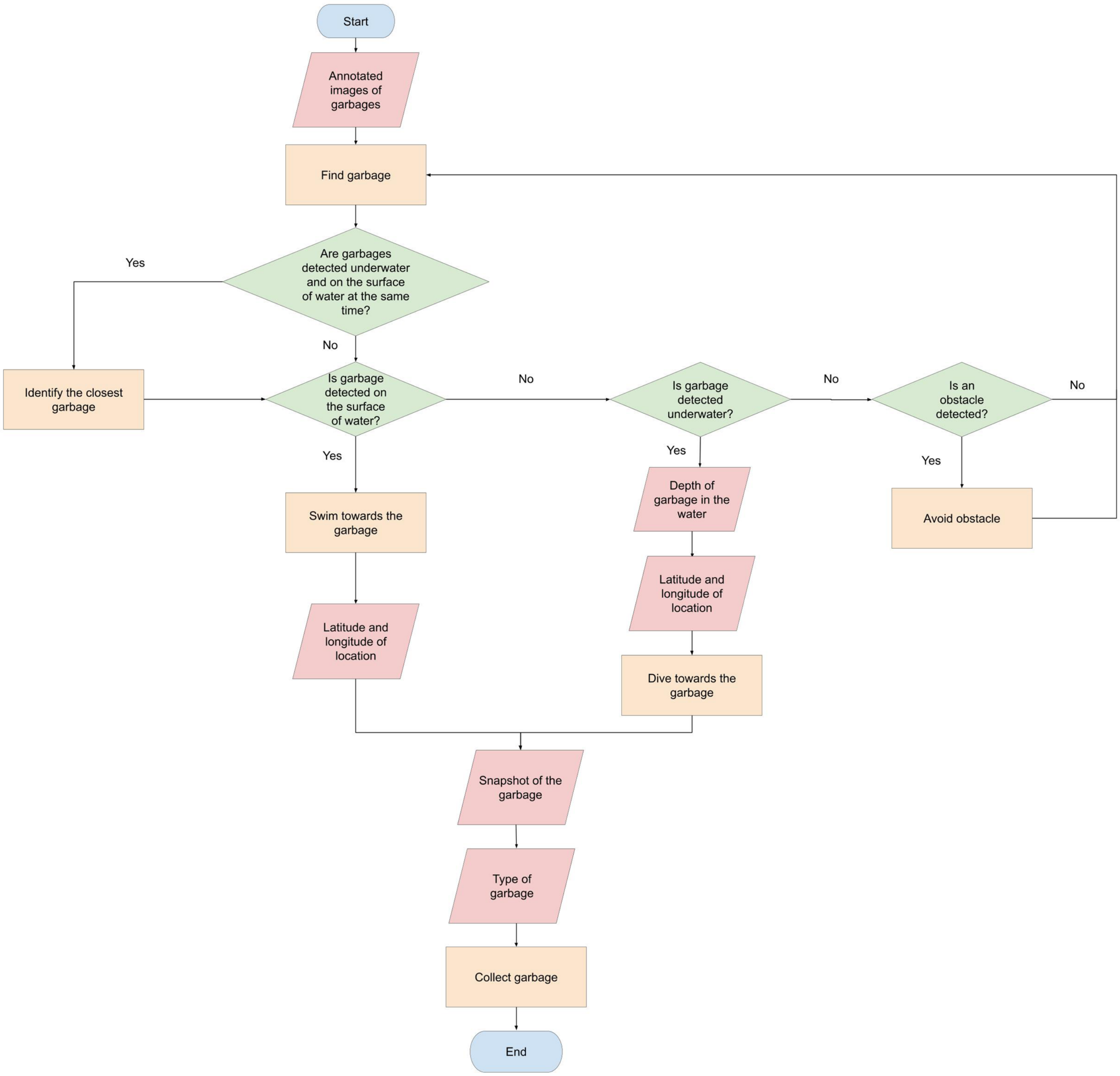

Fig. 2 Flowchart of AGC

capturing a photo of the underwater garbage while AGC is on the surface, it is better to take this photo when the garbage is within 50 cm from the camera to have a clear visual. Christopher et al. [26] found that high water turbidity leads to a shorter transmission range; their experiment revealed that the signal transmission range is more limited in lake water than seawater because lake water has a greater turbidity level. In case the photo captured by the underwater camera cannot be delivered to the central hub over a wireless communication network via the SX1278 LoRa module, the photo will be named as the coordinates and stored in the TF card. When AGC resurfaces and the signal is available, these images can be sent to the central hub for analysis. Following this successful transmission, the images can be removed from the TF card to preserve the storage capacity. Any organisms and objects, such as reefs, rocky outcrops, and boats, which do not match any of the annotated images of garbage would be considered an obstacle that AGC needs to avoid. The NEAT algorithm enables AGC to navigate its surroundings and avoid collisions effectively.

5 Machine learning

5.1 Data

In the previous work, our team used the TACO dataset available on GitHub [8, 12], which offers a solid foundation for detecting surface-level plastic pollution through its extensive annotations and diverse categories. However, the challenges posed by underwater environments, such as changes in lighting, require further attention, and this study was designed to address challenges specific to underwater trash detection, such as low visibility and complex backgrounds. To meet this demand, we have turned to the JAMSTEC E-library of Deep-sea Images (J-EDI) dataset in this research. Collected for the detection of underwater trash, this dataset consists of Remotely Operated Vehicles (ROV) video footage. It is curated by The Global Oceanographic Data Center (GODAC) at the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) and is later processed and labeled with bounding boxes of different categories: "trash, biological objects such as plants and animals, and ROVs," known as Trash-ICRA19 [27]. It captures the unique characteristics of underwater environments, offering the annotations necessary for the development of more accurate models.

The latest version of the Trash-ICRA19 dataset, now known as TrashCan [28], represents a significant improvement in the monitoring of trash. It comprises over 7,000 annotated images derived from the J-EDI and introduces detailed instance segmentation of underwater trash along with various entities. Unlike the TACO dataset, which focuses more on litters in terrestrial environments [8], or AquaTrash, which pays more attention to collecting images of floating garbage [12], TrashCan is uniquely suitable for underwater settings [28], as it uses footage from ROVs in the deep sea to provide instance segmentation annotations with pixel-level accuracy. This is critical for addressing the complexities of underwater environments where interactions among trash, flora and fauna lead to distinct challenges.

The TrashCan dataset consists of 7,212 annotated images of trash, ROVs and a diverse set of marine life [28]. We selected TrashCan due to its specialized annotations that capture the intricacies of underwater detection. Each image is annotated with its size and object identification (ID and class ID, description, geometry type, tags and timestamps). Using a dataset specifically collected for underwater conditions allows the model to develop a robust understanding of these challenging environments, thus reducing the possibility of misclassification and enhancing detection accuracy in real-world applications. An analysis of the distribution of materials within the dataset shows that ROVs are the most commonly detected item, which suggests their frequent usage in underwater settings. The categories labelled 'Other Trash' and 'Plastic' each account for approximately 2,000 images, indicating the critical need to address underwater plastic pollution. The dataset also contains over 1,000 images of metals and a substantial number of marine life, including 'Fish,' 'Eel' and 'Starfish,' which suggest that biodiversity must be protected from pollution. To enhance classification, the original 16 detailed categories were combined into four primary groups: ROV, Trash, Plant, and Animal. This reduction improves model accuracy by focusing on key categories, although it may limit the model's ability to detect more subtle distinctions within these groups. However, this simplification aligns with the study's primary goal of achieving efficient trash detection.

5.2 Model

Our previous research explored the Mask R-CNN and COCO models in object detection within aquatic environments. The Mask R-CNN, famous for its precise instance segmentation, effectively identified different types of trash as part of the Garbage Watch AI project. Despite its proficiency in general object recognition, Mask R-CNN's performance in underwater settings was limited by challenges, such as low visibility or complex backgrounds. To meet the study's objectives, three models (YOLOv5, NanoDet, and RT-DETR) were selected based on their unique advantages for underwater detection. Specifically, YOLOv5 was selected for its high speed and real-time capabilities, which makes it ideal for applications that require rapid processing. NanoDet, known for its lightweight framework, was chosen for its efficiency on devices with limited computational resources, which can be a common constraint in underwater robots. Finally, RT-DETR was selected for its high accuracy in complex scenes, as accurately separating trash from marine life is crucial in the underwater context.

For model selection, we aimed to introduce diversity in our model comparisons. NanoDet, YOLOv5, and RT-DETR represent three different types of object detection models: a lightweight model, a traditional YOLO series model, and a Transformer-based model, respectively. By comparing these three distinct approaches, we can comprehensively evaluate their performance in our specific task, thus providing valuable insights for future research. In terms of the model itself, YOLOv5 is a widely recognized and stable model with a well-established track record in real-time applications, which aligns with our project's requirements for proven reliability. Our preliminary tests showed that it already performed effectively in our specific application area, which makes it a suitable choice for underwater detection tasks. Additionally, it has numerous practical use cases, which allow for easier integration and troubleshooting. Moving forward, we plan to explore additional models, including YOLOv11, as part of an extended research.

Each model was trained on the TrashCan dataset with training regimes of 10, 30, and 55 epochs to evaluate performance across different training intensities. These choices reflect a balance between real-time capability, computational efficiency, and detection accuracy, which provide a comprehensive assessment of model suitability for underwater garbage detection. By selecting these models, we aimed to capture the trade-offs between speed, efficiency, and accuracy. For example, while YOLOv5 enables fast processing, it may compromise precision compared to RT-DETR, which, although more accurate, requires more computational power. NanoDet offers more efficiency on edge devices, though at the potential cost of lower accuracy. These considerations influence our results and emphasize the relevance of each model to different practical environments.

5.2.1 YOLOv5

You Only Look Once (YOLO) series is a family of object detection models that are renowned for detecting objects in images quickly and accurately [29]. YOLOv5 was released in 2020. It is designed to be faster than its previous models, as it comes with an enhanced speed, improved accuracy and simplified deployment, which allows it to be integrated into different platforms and frameworks [30]. Once an object is detected, YOLOv5 will categorize it into trash or non-trash items (e.g. marine animals) by providing boxes and identifying the object within these boxes [31]. Convolutional neural networks (CNN) are useful particularly for dynamic environments such as underwater locations, as they can process and categorize objects in low visibility rapidly. One disadvantage, however, is that it requires relatively high computational power for both training and inference [32]. This can limit its deployment in environments where computing resources are constrained, such as on mobile devices. There are already implementations of YOLOv5 in underwater garbage detections, such as the Octacleaner robot [33].

5.2.2 NanoDet

Unlike YOLOv5, which requires predefined anchor boxes to decide the location of the object in the image, anchor-free models such as NanoDet identify objects in the image by classifying each pixel, like segmentation techniques [34]. This approach helps simplify computations, eliminates the need for anchor boxes as a hyperparameter, and reduces the time needed for the object detection process. Among all the anchor-free models, NanoDet is unique for its lightweight framework, which allows for efficient deployment on edge devices with limited computing power, such as small underwater ROVs. While NanoDet provides good performance in detecting smaller and less distinct objects, it might struggle with the high variability and similar appearances of underwater trash, potentially leading to lower precision compared to more complex models.

5.2.3 RT-DETR

The original DETR (Detection Transformer) model was developed by Facebook in 2020. Later, it was further developed by Baidu in 2023, and the name became Real-Time Detection with Transformers (RT-DETR). It also represents a shift from the conventional approaches to directly translate the image inputs into object detection outputs without the need for specific components like anchor generation [35]. It performs best for objects with intricate interactions that require high computational resources for detection. This is particularly valuable in areas where the quality of detection is more crucial than the speed of real-time processing. The key features of this model include the transformer mechanism, which enables the model to consider the entire image at one time, thus improving its ability to recognize objects based on their context within the scene [35]. It also simplifies the steps to detect an object by allowing end-to-end training directly on sets of objects [36]. Moreover, it uses a bipartite matching loss as an innovation sector, which helps in assigning predictions to ground truth objects, ensuring that each prediction is optimized for one true object [37]. Despite these advantages, the drawback that cannot be ignored is its long training time during the detection process. Compared with other detection models like Faster R-CNN, the transformer structure in RT-DETR requires more epochs to converge, which can significantly extend the training phase [38]. This is due to its global attention mechanism, which treats all parts of the image equally rather than focusing on potential regions of interest [36], and this could cause some challenges when detecting underwater garbage with small size and lead to a decrease in accuracy.

To summarize, Table 1 below compares these three models in terms of their core technology, strengths, weaknesses and best use case. It serves as a guideline for potential users to select the most appropriate model based on their specific requirements. For example, if they want high accuracy and can accept longer processing times and greater computational demands, RT-DETR is recommended for them. On the other hand, if users pay more attention to efficiency and lightweight models with limited computational power, then NanoDet is most suitable for them. YOLOv5 could be a choice for users who are indifferent about processing time but would like to use models that include anchors in the image output.

Table 1 Comparison of three models: YOLOv5, NanoDet, and RT-DETR

	YOLOv5	NanoDet	RT-DETR
Core Technology	CNN	Lightweight CNN	Transformer
Advantages	- High-speed and real-time processing - Effective in unstable environmental conditions	- Extremely efficient in terms of computational resources - Suitable for devices with limited hardware	- High accuracy - Efficient in dealing with complex objects
Disadvantages	Requires relatively more computational resources	Might not perform well in complex scenarios with highly varied objects	Longer process time and more processing power
Best Use Case	Large-scale, real-time data	Small devices	Scenario where precision is critical

Table 2 Mean Average Precision (mAP) performance of YOLOv5, NanoDet, and RT-DETR models across 10, 30, and 55 epochs

	YOLOv5	NanoDet	RT-DETR
10 Epochs	0.590	0.191	0.357
30 Epochs	0.707	0.258	0.506
55 Epochs	0.720	0.266	0.536

5.3 Results

We evaluate the performance of these three models using their Software available online: YOLOv5 [39], NanoDet-Plus [40], and RT-DETR [41] across the training regimes of 10, 30, and 55 epochs. The performance metrics, specifically the mean Average Precision (mAP) values at an Intersection over Union (IoU) threshold of 0.5, together with considering a maximum of 100 detections per image (maxDets), are summarized in Table 2. For YOLOv5, the results showed the highest mAP value among all three models and consistent improvements for the value, which reached 0.72 at 55 epochs. On the other hand, NanoDet shows the lowest and modest improvements, with the mAP of 0.191 at 10 epochs and 0.266 at 55 epochs, only increased by 0.075 from 10 to 55 epochs. This proves our discussion of the model, which mentioned the existence of a trade-off between the model's lightweight efficiency and high training accuracy. Finally, the RT-DETR model exhibits an observable increase in mAP from 10 epochs to 55 epochs, which could be attributed to its transformer-based structure, but it shows a lower accuracy in each level compared with YOLOv5.

Figure 3 compares the Mean Average Precision (mAP) performance among the three models across varying training epochs. The chart shows that YOLOv5 consistently demonstrates the highest mAP, representing its effectiveness in achieving high accuracy. NanoDet, on the other hand, has the lowest performance. Finally, RT-DETR shows

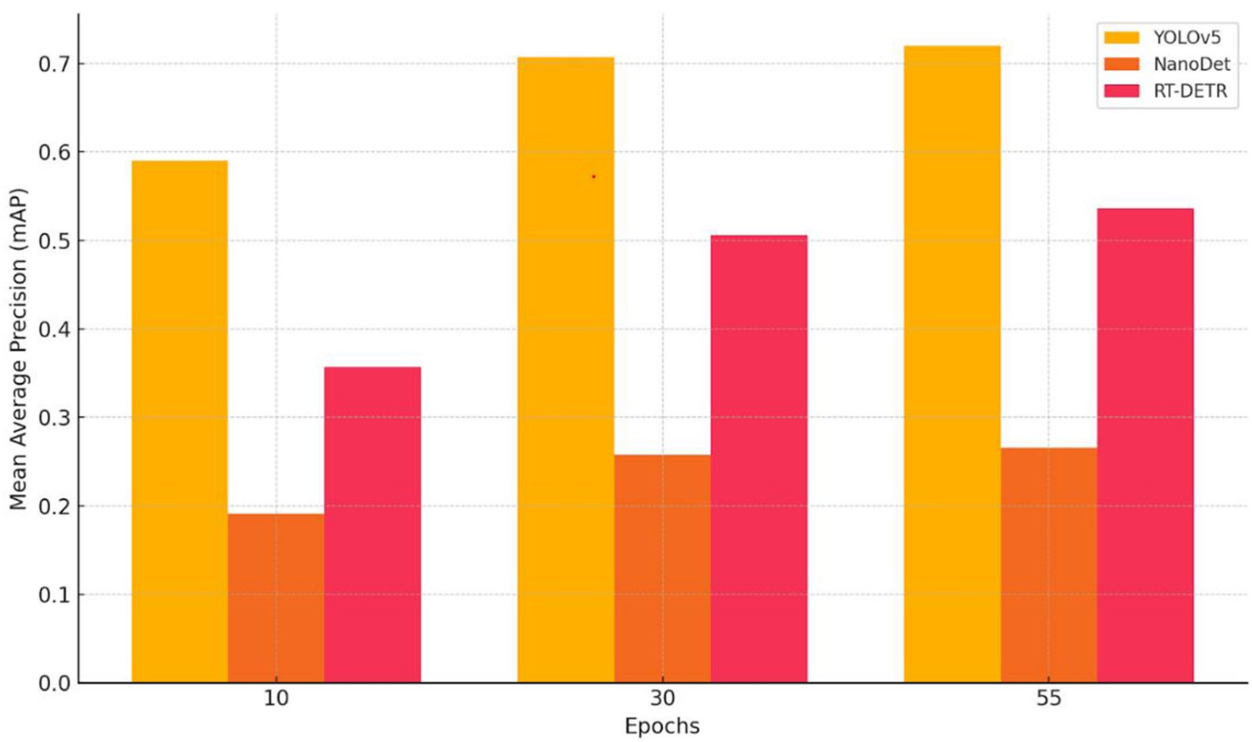

Fig. 3 Mean average precision (mAP) across epochs

Fig. 4 Trash detected by YOLOv5 model

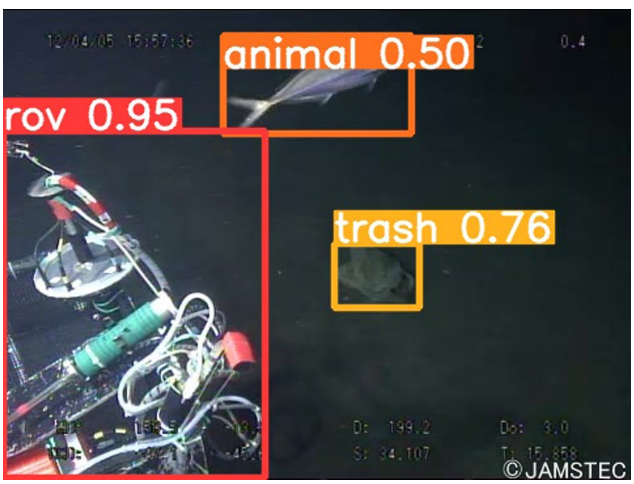

Fig. 5 ROV, animal, and trash detected by YOLOv5 model

a steady increase in mAP but remains lower than YOLOv5 at each epoch level.

Below, we present six output images from the models after training with the dataset. Each image is with predictive boxes, confidence percentages and class labels. Figures 4 and 5 illustrate results from the YOLOv5 model. Figure 4 identifies a trash item (probably a plastic bag) with a confidence score of 0.91, and Fig. 5 demonstrates the model's ability to discern between an ROV (with 95% confidence), nearby trash (with 76% confidence) and an animal that looks like a fish (with 50% confidence).

Figures 6 and 7 use the same images as input and are detected by the NanoDet model. In Fig. 6, the same piece of trash is detected with a confidence score of 70.9%. Unlike YOLOv5, NanoDet shows lower confidence at the cost of high efficiency under limited computing resources. Figure 7 represents a decrease in confidence by 10%. This also highlights the model's ability to handle scenes with multiple objects, though at a lower level of confidence.

Figures 8 and 9 display the detection outcomes from the RT-DETR model. These two images show a high accuracy in classifying underwater objects. In Fig. 8, the model shows a high confidence of 99.88% in recognizing a piece of trash marked with the bounding box. Similarly, with multiple objects in the same image, the model also accurately identifies the animal, ROV and trash, with confidence scores all above 90%. These high confidence values illustrate the model's adeptness in distinguishing between biological entities and non-biological garbage or ROVs, which is a crucial feature for comprehensive underwater monitoring.

In comparing the detection outcomes from YOLOv5, NanoDet, and RT-DETR models, RT-DETR shows the highest accuracy level in identifying diverse underwater objects. However, it takes the longest time for our training. On the other hand, YOLOv5 is faster, which makes it suitable for real-time applications, and NanoDet returns a relatively lower accuracy given its lightweight design. When we choose the optimal model for our detection robot, considerations must be balanced between these factors. Among the comparison of these three models, YOLOv5 appears to be a practical compromise, providing a good balance of speed and accuracy. It is capable of fast processing without the extensive resources needed for RT-DETR.

Fig. 6 Trash detected by NanoDet model

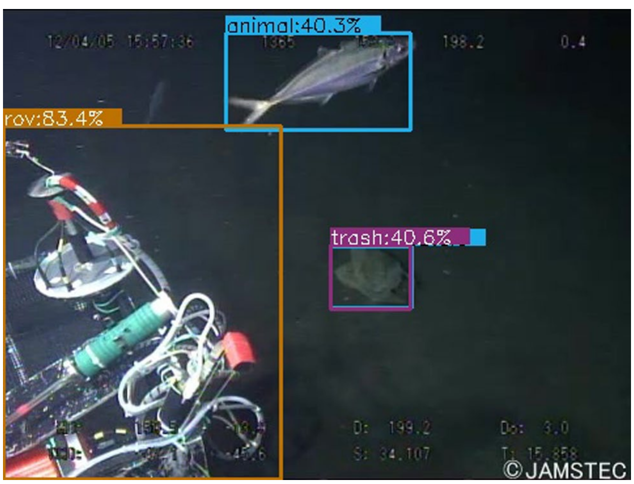

Fig. 7 ROV, animal, and trash detected by NanoDet model

To further balance different factors, YOLOv5 was chosen as the optimal model for further fine-tuning to enhance performance. We conducted hyperparameter tuning experiments by having a fixed number of 100 training epochs and adjusting the following key parameters: batch size and image size. Specifically, we tested batch sizes of 16, 32, and 64 and image sizes of 512, 640, and 768, which are commonly used in computer vision. These dimensions are commonly used because they offer a balance between computational efficiency and sufficient resolution for object detection tasks. The results show that a batch size of 32 and an image size of 640 give the highest mean Average Precision (mAP) at 75.48%, and there has been an improvement from the initial 72%. This optimized configuration achieved the best detection performance across all four categories in our dataset, reinforcing the role of YOLOv5 as a practical compromise that balances speed and accuracy effectively without requiring the extensive resources needed for RT-DETR (Table 3).

Figure 10 shows the mAP results for different batch and image sizes with YOLOv5 at 100 training epochs. For each combination of batch and image sizes, mAP is at least 0.7. From the chart, we can also see that a batch size of 32 and an image size of 640 yields the highest results out of all the groups.

6 Limitations

It is challenging to find a wireless transceiver that can transmit data through water and long distances. As a result, AGC must undergo additional steps before sending photos of underwater garbage: saving images on the TF card and emerging above the water surface. It would be more efficient if it could send images directly to the database without storing them on the TF card. Specialized technologies might exist that accommodate this, but they will most likely be costly.

Although the NEAT algorithm can help AGC to navigate around marine creatures, it is difficult to predict their actions. They could inspect AGC out of curiosity or attack it as a threat. Moreover, the noise from AGC may have unintended ecological impacts on species sensitive to sound, such as causing them to leave their shelter or hunting ground [42].

Fig. 8 Trash detected by RT-DETR model

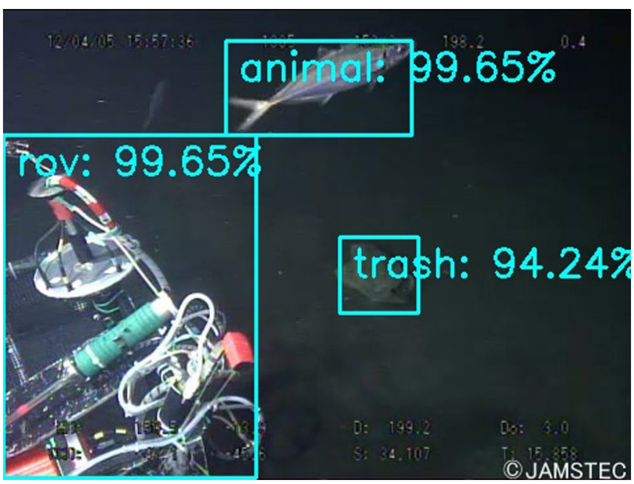

Fig. 9 ROV, animal, and trash detected by RT-DETR model

Table 3 Batch size and image size changing results for YOLOv5 with 100 training epochs

	Batch size	Image size	mAP
100 Epochs	16	512	0.753
100 Epochs	16	640	0.747
100 Epochs	16	768	0.728
100 Epochs	32	512	0.748
100 Epochs	32	640	0.755
100 Epochs	32	768	0.740
100 Epochs	64	512	0.748
100 Epochs	64	640	0.744
100 Epochs	64	768	0.729

SIM28 Arduino GPS Shield cannot easily acquire coordinates at great depths under the surface. If there is an unforeseen accident where AGC sinks, we might not be able to locate it. Although we have a record of AGC's most recent surface position, AGC might have drifted farther away by an animal or strong water currents.

There are also some limitations in terms of our analysis process. In the dataset used in this study, the most frequently detected material is ROVs, which could lead to model bias and reduce the effectiveness of our robot in detecting garbage. In addition, the dataset may not fully include specific areas, such as Ontario lakes and riverways, as the images are mainly collected from the Sea of Japan. A different ocean environment could hinder our robot's ability to detect garbage from Lake Ontario, which was our previous team members' designated place to clean. Finally, the simplification of the original dataset from 16 detailed categories to merely 4 broad groups may have resulted in a decrease in the models' ability to discern distinctions among different types of trash, and it could potentially impact the precision of monitoring and analyzing garbage, which might limit the effectiveness of the models in different water environments.

Moreover, limitations related to the machine learning models themselves need to be considered. While YOLOv5 demonstrated the highest mAP value at 0.72, the reliance on computationally intensive processes may affect the deployment on low-resource edge devices, particularly in underwater environments with restricted processing power. Our choice of models could also be expanded, as deploying more recent versions, such as YOLOv11, may enhance

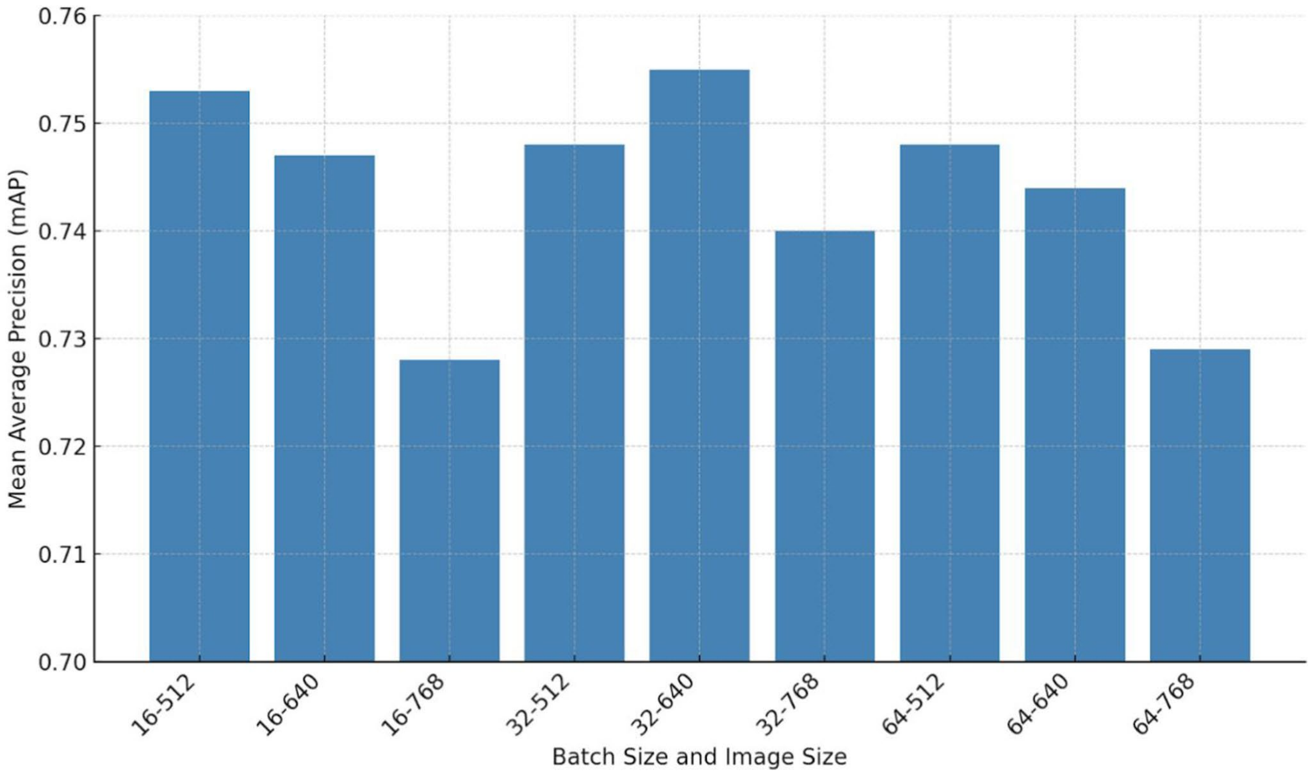

Fig. 10 mAP value for different combinations of batch size and image size

accuracy and adaptability. In addition, NanoDet has also suggested its potential struggle with the complex and variable nature of the underwater environment. Finally, RT-DETR, despite its accuracy in identifying complex places, required further training periods and displayed a lower mAP of 0.536 compared to YOLOv5, which makes it less practical for real-time applications. These limitations indicate that the chosen models might not perform consistently across diverse underwater conditions, particularly those differing from the dataset's main environment. Future research should focus on optimizing these models for improved adaptability, including more advanced and customized techniques that can better handle environmental factors such as low light and diverse underwater garbage compositions.

7 Conclusion

Water is a necessity for living beings, so it is vital to maintain clean water. However, with pollution increasing each year, this goal becomes more tough to achieve. The best we can do is to optimize modern technology, particularly AI and IoT, to rectify the damage caused by human activities. We improved the garbage management and cleaning processes by adopting a new functionality to our AGC robot, which is to remove underwater debris. Once the trash is detected by the camera either on the surface or underwater, AGC will share its location and collect it. The identification of the trash will be analyzed by machine learning models on the centralized server. Images of geotagged trash can help statisticians pinpoint waste hotspots, which can also serve to raise public awareness and enforce waste disposal policies. The success of these policies can be evaluated by tracking the volume of garbage in those areas over time.

We examined three models, YOLOv5, NanoDet, and RT-DETR, to improve the functionality of our AGC robot in classifying underwater trash. Our findings show that while RT-DETR provides the highest accuracy, its application in real-life occasions is limited by computational capabilities. YOLOv5, in contrast, offers a more balanced trade-off between speed and accuracy, making it suitable for underwater garbage detection. These insights lay the groundwork for further refinement of detection technologies, aiming not just to alleviate the effects of marine pollution but to implement proactive garbage management. The key results also show that optimizing detection models and expanding the dataset with more diverse samples is crucial for improving adaptability and minimizing bias, thus enhancing garbage classification across different water environments. This paper contributes to new knowledge by illustrating a comprehensive model assessment and suggests directions for improving AI-based underwater trash detection. These insights propose the way for future improvements in using machine learning to help proactive marine pollution management and cleaner ecosystems.

8 Future work

Our project mainly centred on how AGC located and classified the garbage above and below the water surface. With limited time, we did not have the opportunity to build a prototype and test it in the water. However, before we start constructing, additional investigation is required to study the mechanism behind AGC submerging into the water and grabbing the garbage. When the AGC's door opens to get the new scrap either by a robotic gripper or vacuum, we need to devise a method such that no collected trash gets released back into the aquatic environment. We also need to research how AGC determines their bin is full and where to drop off these recyclable and non-recyclable items. Marine animals are already suffering consequences from water pollution, so we do not want to increase the harm from underwater noise. Thus, we have to ensure the engine does not create disruptive sounds as AGC swims and collects garbage.

To overcome the current data processing limitations, future improvements could focus on several critical areas. Firstly, expanding the dataset to include a greater collection of images from diverse local underwater environments, such as Lake Ontario, could enhance model robustness. Specifically, adding more images of various types of underwater trash will help reduce the model bias towards detecting ROVs. Furthermore, adapting model parameters to handle the challenging conditions of underwater settings better, such as varying light and visibility, is also essential. This may involve refining algorithms to adjust dynamically to these environmental factors, thus further improving detection accuracy. Additionally, it is important to maintain the diverse categories of the dataset to enable more precise environmental assessments. Implementing these improvements could help enhance the performance of the underwater detection models in marine protection and garbage management.

AGC's first test run will begin on the pond. Before placing it onto the water, we will validate the GPS coordinates to ensure it aligns with its true geographic position. If there are substantial differences despite experimenting with several SIM28 Arduino GPS shields, it indicates we need a GPS with more tracking and acquisition channels. While viewing the live stream through the cameras installed on AGC, we will examine whether AGC properly detects, classifies, and collects garbage. Inference will be performed on Hadoop clusters via the cloud, and we will assess the duration it takes for the machine learning model to analyze and identify the detected trash. In the event that the Arduino UNO is incapable of processing and exchanging data, we can consider replacing the Arduino UNO with the Raspberry Pi 4 Model B. Once this early phase of testing is successful, we can move to more challenging waters: lakes followed by oceans.

Future work should also enhance the selection and diversity of datasets used for training machine learning models. Expanding beyond existing datasets like TrashCan to include more similar underwater environments compared to Lake Ontario can improve model adaptability and generalization. Moreover, integrating updated and more powerful models, such as YOLOv11, can provide further advancements in precision and efficiency. This newer version may offer improved feature extraction and detection capabilities that align with the unique challenges of underwater trash identification. Furthermore, techniques such as data (image) augmentation could also be applied to increase model robustness without the need for extensive new data collection, ensuring that the models remain effective across a broad range of underwater settings.

Author contributions B.Y. concentrated on the IoT aspects of the project. She designed the sequence of how the robot AGC operates and researched the hardware that supports the desired behaviour. R.Z. focused on the data analysis segment of this research, mainly on the evaluation of machine learning models' performance in underwater trash detection. The effectiveness and performance of YOLOv5, NanoDet, and RT-DETR are thoroughly analyzed and compared, contributing to the optimal selection of technology for enhancing the AGC's capabilities. The overall project was supervised by M.E. All authors reviewed the manuscript.

Data availability No datasets were generated or analysed during the current study.

Code availability https://github.com/ruochenzhao09/AquaGarbageCollector

Declarations

Competing interests The authors declare no competing interests.

References

1. Fava, M.: Ocean plastic pollution an overview: data and statistics. UNESCO. (2022). https://oceanliteracyunesco.org/plastic-pollution-ocean/. Accessed 15 June 2024
2. De La Garza, L., Chen, S., Kelly, B., Parsazadeh, H., Zhou, P.: Garbage Watch AI: Identification and classification of Plastic Pollution in Water. iJournal: Student J. Fac. Inform. (2022). https://doi.org/10.33137/ijournal.v7i2.38614
3. Kelly, B.O., Chen, S., Zhou, E.P., Elshakankiri, M.: AI-Enabled Plastic Pollution Monitoring System for Toronto Waterways. 2023 10th International Conference on Internet of Things: Systems, Management and Security (2023). https://doi.org/10.1109/IOTSMS59855.2023.10325803

4. Zimmermann, L., Göttlich, S., Oehlmann, J., Wagner, M., Völker, C.: What are the drivers of microplastic toxicity? Comparing the toxicity of plastic chemicals and particles to Daphnia magna. Environ. Pollut. **267**, 115392 (2020). https://doi.org/10.1016/j.envpol.2020.115392
5. Lebreton, L., Slat, B., Ferrari, F., Sainte-Rose, B., Aitken, J., Marthouse, R., Hajbane, S., Cunsolo, S., Schwarz, A., Levivier, A., Noble, K., Debeljak, P., Maral, H., Schoeneich-Argent, R., Brambini, R., Reisser, J.: Evidence that the Great Pacific Garbage Patch is rapidly accumulating plastic. Sci. Rep. **8**(1) (2018). https://doi.org/10.1038/s41598-018-22939-w
6. Grbić, J., Helm, P., Athey, S., Rochman, C.M.: Microplastics entering northwestern Lake Ontario are diverse and linked to urban sources. Water Res. **174**, 115623 (2020). https://doi.org/10.1016/j.watres.2020.115623
7. Earn, A., Bucci, K., Rochman, C.M.: A systematic review of the literature on plastic pollution in the Laurentian Great Lakes and its effects on freshwater biota. J. Great Lakes Res. (2021). https://doi.org/10.1016/j.jglr.2020.11.001
8. Proença, P.F., Simões, P.: Taco: Trash annotations in context for litter detection. arXiv Preprint arXiv:2003 06975. (2020). https://doi.org/10.48550/arxiv.2003.06975
9. Hu, J., Yan, Y., Xie, Z.: Automatic recognition technology of library books based on convolutional neural network model. HighTech Innov. J. **5**(1), (2024). https://doi.org/10.28991/HIJ-2024-05-01-015
10. Moya, V., Quito, A., Pilco, A., Vásconez, J.P., Vargas, C.: Crop detection and maturity classification using a YOLOv5-Based image analysis. Emerg. Sci. J. (2024). https://doi.org/10.28991/ESJ-2024-08-02-08
11. Worachairungreung, M., Kulpanich, N., Thanakunwutthirot, K., Hemwan, P.: Monitoring Agricultural Land loss by analyzing changes in Land Use and Land Cover. Emerg. Sci. J. (2024). https://doi.org/10.28991/ESJ-2024-08-02-020
12. Panwar, H., Gupta, P.K., Siddiqui, M.K., Morales-Menendez, R., Bhardwaj, P., Sharma, S., Sarker, I.H.: AquaVision: Automating the detection of waste in water bodies using deep transfer learning. Case Stud. Chem. Environ. Eng. (2020). https://doi.org/10.1016/j.cscee.2020.100026
13. Jiang, X., International Conference on Data Science and Computer Application: Vision-Based Outdoor Garbage Collection Robot. IEEE 3rd (2023). (2023). https://doi.org/10.1109/ICDSCA59871.2023. 10392429
14. Naik, A.V., Raj, E.V., Chaitra, C.T., Harshitha, K.S.: Komal.: AI Based Robot for Beach Cleaning. 2023 International Conference on Applied Intelligence and Sustainable Computing (2023). https://doi.org/10.1109/ICAISC58445.2023.10200181
15. Digda, A.W., Kurniadi, N.R., Goenawan, R.T., Hoswari, S.A., Gozali, L., Siahaan, E.: Ro-Boat: IoT-Based Non-Autonomous Garbage Collector Boat for Organic, Metal, and Non-Metal Waste. 2023 International Conference on Sustainable Communication Networks and Application (2023). https://doi.org/10.1109/ICSCNA58489.2023.10370353
16. Dunggat, D., Nawi, I.M., Hashwan, B., Khir, S.S., Sadale, M.H.M., Bature, I.M.I., Razali, U., Ismail, A.I., Isa, A., Mansor, A.F.A.: M. H. I. A.: Unmanned Surface Vehicle for Water Wastes Collection. IEEE International Conference on Sensors and Nanotechnology (2023). (2023). https://doi.org/10.1109/SENNANO57767.2023.10352554
17. Kondaveeti, H.K., Kumaravelu, N.K., Vanambahina, S.D., Mathe, S.E., Vappangi, S.: A systematic literature review on prototyping with Arduino: Applications, challenges, advantages, and limitations. Comput. Sci. Rev. (2021). https://doi.org/10.1016/j.cosrev.2021.100364
18. SIM128: SIMCom. (2024). https://cdn.robotshop.com/media/m/mkf/rb-mkf-07/pdf/sim28-arduino-gps-shield-specifications.pdf (n.d.). Accessed 1
19. SX1276/77/78/79: Semtech Corporation. (2015). https://cdn-shop.adafruit.com/product-files/3179/sx1276_77_78_79.pdf Accessed 1 November 2024
20. ESP32-CAM Module: Ai-Thinker. https://loboris.eu/ESP32/ESP32-CAM%20Product%20Specification.pdf (2017). Accessed 2 November 2024
21. de Bakker, B.: Waterproof JSN-SR04T Ultrasonic Distance Sensor with Arduino Tutorial. Makerguides. (2019). https://www.makerguides.com/jsn-sr04t-arduino-tutorial/ Accessed 9 June 2024
22. Dejan: Arduino Brushless Motor Control Tutorial| ESC| BLDC. How to Mechatronics. https://howtomechatronics.com/tutorials/arduino/arduino-brushless-motor-control-tutorial-esc-bldc/ (n.d.). Accessed 18 June 2024
23. Liu, H.: Rail transit assembly robot systems. In: Liu, H. Robot Systems for Rail Transit Applications, pp. 37–88 (2020). https://doi.org/10.1016/B978-0-12-822968-2.00002-4
24. Speed of sound: Oxford Reference. https://www.oxfordreference.com/display/10.1093/oi/authority. 20110803100522606 (n.d.). Accessed 30 June 2024
25. Underwater, G.P.S.: Discovery of Sound in the Sea. https://dosits.org/galleries/technology-gallery/navigation-technology/underwater-gps/ (n.d.). Accessed 23 June 2024
26. Christopher, J.P., Damayanti, S.D., Suryanegara, M.: Investigating IoT-LoRa Technology for The Underwater System Application. IEEE 8th International Conference on Recent Advances and Innovations in Engineering (2023). (2023). https://doi.org/10.1109/ICRAIE59459.2023.10468069
27. Fulton, M., Hong, J., Islam, M.J., Sattar, J.: Robotic detection of marine litter using deep visual detection models. Int. Conf. Rob. Autom. **2019** (2019). https://doi.org/10.1109/ICRA.2019.8793975
28. Hong, J., Fulton, M., Sattar, J.: Trashcan: A semantically-segmented dataset towards visual detection of marine debris. arXiv Preprint arXiv:2007 08097. (2020). https://doi.org/10.48550/arxiv.2007.08097
29. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. 2016 IEEE Conf. Comput. Vis. Recognit. **P** (2016). https://doi.org/10.1109/CVPR.2016.91
30. Thuan, D.: Evolution of Yolo algorithm and Yolov5: The State-of-the-Art object detection algorithm. Accessed 1 July 2024 (2021)
31. Bancud, G.E., Labanon, A.J., Abreo, N.A., Kobayashi, V.: Combining image enhancement techniques and deep learning for shallow water Benthic Marine Litter Detection. Mach. Learn. Principles Pract. Knowl. Discovery Databases. (2022). https://doi.org/10.1007/978-3-031-23618-1_9
32. Wang, G., Zhuang, S., Zhao, Q.: An Improved Underwater Object Detection Method Based on YOLOv5. 7th International Conference on Advanced Algorithms and Control Engineering (2024). (2024). https://doi.org/10.1109/ICAACE61206.2024.10549193
33. Pavani, D., Reddy, A.N.N., Saw, N.: YOLO object detection method for underwater trash collecting robot: Octacleaner. 2023 IEEE Technology & Engineering Management Conference-Asia Pacific (2023). https://doi.org/10.1109/TEMSCON-ASPAC59527.2023.10531302
34. Padmanabhan, L., Varghese, C.: ANPR using NanoDet, an Anchor-free Object Detection Model., Ignitarium: (2022). https://ignitarium.com/anpr-using-nanodet-an-anchor-free-object-detection-model/ Accessed 29 June 2024
35. Petru, P.: What is DETR? Roboflow Blog: (2023). https://blog.roboflow.com/what-is-detr/ Accessed 29 June 2024
36. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers.

Smart Comput. Syst. Eng. (2020). https://doi.org/10.1007/978-3-030-58452-8_13
37. Gosthipaty, A.R., Raha, R.: DETR Breakdown Part 2: Methodologies and Algorithms., PyImageSearch.: (2023). https://pyimagesearch.com/2023/06/12/detr-breakdown-part-2-methodologies-and-algorithms/ Accessed 12 June 2024
38. Goswami, S.: Comparison of Faster-RCNN and Detection Transformer (DETR). Medium. (2020). https://whatdhack.medium.com/comparison-of-faster-rcnn-and-detection-transformer-detr-f67c2f5a2a04 Accessed 22 June 2024
39. Jocher, G.: YOLOv5 by Ultralytics (Version 7.0) [Software]. (2020). https://doi.org/10.5281/zenodo. 3908559 Accessed 30 June 2024
40. RangiLyu.: NanoDet-Plus: Super fast and high accuracy lightweight anchor-free object detection model [Software]. (2021). https://github.com/RangiLyu/nanodet Accessed 30 June 2024
41. Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., Chen, J.: DETRs Beat YOLOs on Real-time Object Detection [Software]. (2023). https://github.com/lyuwenyu/RT-DETR Accessed 30 June 2024
42. Underwater Noise FAQs: WWF. (2013). https://wwf.ca/wp-content/uploads/2014/01/oceannoise_faqs.pdf Accessed 30 June 2024

Bonnie Yam completed a Bachelor of Science in Mathematics at the University of Toronto and a Master of Information, specializing in Information Systems and Design. Her research projects delve into IoT solutions that enhance environmental monitoring and help people with disabilities overcome daily challenges.

Ruochen Zhao completed an Honours Bachelor of Science in Economics at the University of Nottingham and a Master of Science in Economics at the University of Warwick. Currently, she is studying for a Master of Information in Human-Centred Data Science at the University of Toronto. Her research interests include utilizing machine learning and data science techniques to solve real-world environmental problems and improving predictive accuracy in healthcare-related sectors.

Maher Elshakankiri is an assistant professor at the Faculty of Information, University of Toronto. His research interests include Active Learning Classrooms, the Internet of Things, Wireless Sensor Networks, and Mobile Ad Hoc Networks. He has also researched vehicle-to-vehicle communication, unmanned aerial vehicles, and RFID in libraries and retail systems. He authored a book titled "Introduction to Wireless Sensor Networks." Professor Elshakankiri is a member of the Standard Council Canada (SCC) committee for the Internet of Things and Digital Twins. He is also a reviewer at several journals and a technical program committee member at numerous conferences. Professor Elshakankiri is currently the Director of the Bachelor of Information (BI) Program.