An Assessment of the Application of Real-time YOLOV5 Deep Learning Algorithm in Unmanned Surface Vessels for Environmental Maintenance: Some Preliminary Results

Ali Camdal  
*School of Engineering, Trinity College Dublin, Ireland. E-mail: camdala@tcd.ie*

Biswajit Basu  
*School of Engineering, Trinity College Dublin, Ireland. E-mail: basub@tcd.ie*

Andrea Staino  
*School of Engineering, Trinity College Dublin, Ireland. E-mail: stainoa@tcd.ie*

Pollution of aquatic environments is one of the biggest problems faced by mankind throughout history. The quality of ocean, sea and other stagnant waters that cover 70% of the world affects not only human life but also other living creatures in nature. Since these environments are large and difficult to access; observing, maintaining, and cleaning these environments are challenges that must be overcome in a reliable and safe manner. In this study, a deep learning algorithm has been trained and implemented for real-time application in an unmanned surface vessel. This was designed to detect and track objects on the surface of the water, thus subsequently enabling the maintenance and cleaning operations. The device hardware developed and integrated with the real-time deep learning intelligence has been tested in both controlled and field environment. The real-time deep learning model has been retrained and validated using the public marine litter data set. As a result of the training, the model is able to detect objects on the water surface with a mean average precision of 85%, 94% recall, and precision of 78%. Moreover, processing time is less than 100 milliseconds for per frame. The implementation of the real-time YOLOv5 deep learning model will facilitate the operation of tracking objects on the sea surface and thus will reduce maintenance costs, shorten the time requirement for operation, and increase the efficiency of the detection process.

**Keywords**: deep learning, maintenance, unmanned surface vessel, object detection

1. **Introduction**

70% of the world is water environment and one of the biggest problems of today is the irreversible pollution of this aquatic environment. Estimated 5.25 trillion plastic waste in the aquatic environment on Earth. While 269 thousand tons of these wastes are floating, it is thought that there are 4 billion microfiber plastic wastes per square kilometer under the surface. It is believed that this number will increase exponentially when the plastic product of human beings and the lifetime of the produced plastics are taken into account. Studies conducted by Vlachogianni et al. (2018) reveal how serious this situation is.

These plastic wastes not only affect the lives of people, but also deeply affect the living things in the aquatic environment. Zrimec et al. (2021) show that some microbes have evolved to produce new enzymes to digest this plastic garbage. However, it is obvious that this pollution will not end with the methods applied by microbes or other living creatures. Considering all these situations, the most appropriate method is to eliminate this pollution by autonomous surface and underwater devices that can be produced. Compared to many different non-autonomous applications, it can be seen how effective this solution is in terms of both time and economy.

In this paper, an algorithm for real-time detection and tracking of objects to be integrated with an unmanned surface vessel (USV), that can detect plastic wastes in inland waterways, is proposed. The main motivation for this project is the increase in environmental problems mentioned earlier. In addition to this motivation, when the research project reaches a successful conclusion, the device will have value as a product that can have both academic impact and a market share.
2. Related Works

The relationship of waterways and aquatic environments to human life is becoming more understandable. This situation paved the way for many studies to protect, clean and observe aquatic environments. In recent years, there has been significant development especially in the fields of computer vision and autonomous unmanned surface vessel (AUSV) and autonomous underwater vehicles (AUV) CHANGE. Developments in the AUSV-AUV fields have gained momentum in recent years. By making these devices autonomous Mahmoudian et al. (2020), environments where they can work together were created Page and Mahmoudian (2020). With the privileges brought by these collaborative working systems and autonomy, AUV systems have been used in different fields such as mapping Tata et al. (2021) and localization of objects Prats et al. (2011).

It may be noted that there is a fundamental difference between an autonomous and an unmanned vehicle, which is sometimes overlooked. An autonomous device has the ability to perform its action without the intervention of any human operator. Unlike autonomous systems, unmanned systems require some control by the operators for navigation. However, unmanned vehicles have abilities that work autonomously such as real-time object detection or semantic segmentation, without any human input.

The field of computer vision has developed in parallel with device technologies and continues to develop. Machine learning and deep learning techniques offer solutions to many real-time problems today Tyagi and Rekha (2020). Technologies that emerged in the studies carried out to find solutions to these problems contribute to the solution of new problems. Techniques such as Deep Learning and Transfer Learning are the main methods used to detect plastic debris in aquatic environments Fulton et al. (2018).

One of the main problems encountered in deep learning and machine learning applications is to create a data set. It is very difficult to create sets that contain as mixed data as possible and that can meet the demands of learning systems for the purpose of the application. In this case, it is observed that the degree of difficulty increases even more when it comes to marine areas. The difficulty of observing aquatic environments stands out as the biggest factor in this increase. However, some public data sets play an important role in overcoming this difficulty. The data set created in the TACO project by Proena and Simes (2020) can be shown as an example in this regard. In addition to TACO, Trashnet by Thung and Yang (2017) and AquaTrash by Panwar et al. (2020) can be given as examples of data sets that are created as a result of public or certain studies.

Among these data set studies, the most well-known one consists of images collected by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC). This data set has been a very valuable resource for different projects. One of these projects is the data set study called TrashCan, which was made by a team at the University of Minnesota by Staino et al. (2022). The data set included in the study contains images taken from videos that recorded by an AUV performing a research mission in the Japanese seas. It is one of the most important studies in the field of underwater plastic pollution.

In addition to data set studies, there are many studies in which object definition applications are made using this data. For example, a group of researchers using the AquaTrash data set developed an object identification application with the deep transfer learning technique in the study of Panwar et al. (2020). In this application, they used neural networks such as Faster R-CNN, DSSD513, and RetinaNet.

There are many common neural networks used in the object detection field. YOLOV5 a, which has made a rapid rise recently, is one of these neural networks. It is used in many projects because it works with a higher success rate and shorter inference time compared to its previous versions. It also includes different models for different application areas such as railway signal detection paper by Staino et al. (2022). For example, while YOLOV5n and YOLOV5s, which require less

---

ahttps://github.com/ultralytics/yolov5
processing power, are used in devices that require mobility, YOLOV5x can be used in fixed or large-scale applications where processing power may be high.

3. Vehicle Design

The USV in our project has been developed for the purpose of proof of concept for implementation of a real-time deep learning algorithm. Therefore, its dimensions are kept small compared to its counterparts. The studied USV consists of four main parts: device body, electronic systems for control and monitoring, battery powering the specified electronic systems and two motors to control the device on water. Each piece will be analyzed as a subsection.

3.1. Vessel Hull

The vessel hull produced in of catamaran type. The 3D sketch is given in Figure 1. The catamaran body type comes with many benefits in its use. The three most important of these benefits are less current effect, easy fixation of the center of gravity of the device and high mobility. Each catamaran part has its own thrusters. This catamaran body, made of acrylonitrile butadiene styrene (ABS) plastic, creates a mass of approximately 4 kilogram with all the attachments.

![Fig. 1. 3D sketches of catamaran hulls.](image)

3.2. Electronic Systems

Electronic systems are divided into three different parts. The first of these parts is the main processing unit that performs image processing and other computational operations. Nvidia Jetson Nano developer kit was used as the processing unit. This development board, which uses a quad-core ARM-A57 processor, has a 128-core Maxwell GPU processing unit. The specified features are at a sufficient level for a deep learning processing unit that a mobile device should have. In addition, Nvidia JetPack, an Ubuntu version, was used as the operating system.

The second part in the electronic systems of the device is taken by the environmental control units. Sensor and motor control units are completely separated from the main processing unit in order to control the device for emergency response and unexpected errors. Each unit communicates with the main processor using the universal asynchronous receiver-transmitter (UART) protocol. In case of an emergency, the device is designed to be remotely controlled or to follow a determined emergency route by using these discrete units. While a development board with Atmega328 processor is used in the motor control unit, an Atmega2560 based development board is used in the sensor units.

The complementary part of the electronic systems consists of the regulators that provide the necessary power to the above-mentioned units. The regulator that drives the main processor and peripheral control units produces a power of 5 volts and 4 amps. At the same time, there is a regulator in each of the motors that provide the control of USV on the water surface. These regulators and motors will be mentioned in the following sections.

3.3. Battery System

As mentioned in Sections 1 and 2, mobile autonomous devices are emerging as advanced systems today. However, like every technological method and device, USV-AUVs have certain limitations. Battery technologies are at the forefront of these limitations. Battery technology has evolved exponentially over the past 20 years. But working in aquatic environment is more difficult than in other environments. The density of the environment, the obstacles encountered, and other factors consume the power of the devices excessively.
For example, a thruster that consumes 0.3 amps of power while operating in an air environment, when it is in a water environment, experienced power consumption increases by 26 times to 8 amps. This directly affects the battery life and operating time.

The USV device mentioned in this paper uses Li-Ion battery cells installed in 4 series and 6 parallel. While each cell can provide instantaneous 3.7 volts and 6 amps, the capacities of these cells are measured as 3 amp-hours. The created battery can momentarily supply 14.8 volts 36 amps to the system. When the motors that direct the USV are operated with 50% power, the battery can power the system for approximately 2 hours.

3.4. Movement System

The USV has two brushless dc thrusters produced for underwater operations, one in each catamaran body. While these thrusters can operate in the range of 12-24 volts, they can draw a maximum current of 17 Amps. Three-pole DC motors have four-piece propellers. They have clockwise rotation on two propellers. At the same time, the thrusters are placed in the thruster bay to isolate and protect these propellers from the environment.

Each thruster has its own regulator and electronic speed controller. The mentioned regulators are used to protect the current balance of the coils rather than voltage and current limiting. Thanks to these regulators, power fluctuations in the battery will not affect the performance of the motors. Regulator configurations have been selected in accordance with 20 amps to avoid any current problems.

Electronic speed controller circuits enable brushless motors to be operated at the desired speed and direction. There are two basic types of ESCs in general use. These are unidirectional and bidirectional ESCs. ESC selection should be made according to the power requirements and usage areas of brushless DC (BLDC) devices. The ESC circuits used in USV are selected as bidirectional and have a configuration that can deliver 40 amps. With these ESCs, USV has gained the ability to move both forward and backward.

4. Network Design

4.1. Model Selection and Training

There are many different artificial neural network models used in the field of mobile robots. These different patterns can be observed in the studies mentioned in Section 2. Basically, model selection is very important in the field of object detection. The selected model should be suitable for the dimensions, parameters and other characteristics of the images to be used. However, in addition to these, one of the most important factors in mobile systems is the inference time of the model. As it is known, mobile systems contain computers with limited computing power. These limitations naturally affect model selection. If the system is stationary, bigger models might have short inference time. Nevertheless, these models will have longer inference time when they run on systems in mobile devices. Considering these situations, it is necessary to choose the most suitable model. Jetson Nano Development Kit was used in the USV developed in this study. Considering this mini computer, the YOLOV5s6 model was chosen. YOLOV5 includes five different models. These are:

- YOLOV5n
- YOLOV5s
- YOLOV5m
- YOLOV5l
- YOLOV5xl

The inference times of these models in increasing order. YOLO models divide the image into grides and runs detection algorithm for each grid separately. Unlike other YOLO versions, YOLOV5 is built on the PyTorch framework. In YOLOV5 architecture, while the Leaky ReLU activation function is used in the intermediate layers, the sigmoid function is used in the final detection layer (Staino et al. (2022)). The YOLOV5s6 model used in this study has approximately the same structure as the YOLOV5s model given above. In YOLOV5s6, which emerged with the last experimental approaches, the P6 layer was added to the output layer as an extra.

The biggest factor in choosing the YOLOV5s6
model is that the inference time / performance ratio is suitable for a mobile device. In the training of the model, the public data sets mentioned in Section 2 and a custom data set consisting of individually collected images were used. Original data set contains 2000 images. However, as it is known that more images would increase the performance of neural networks, this data set is enhanced with three different techniques.

Neural networks have very sensitive dependency on data set. Slightly different images could be considered as completely different data in networks. With this feature of the NN’s, data sets can be enhanced if their structure is changed. In this study, data set is extended using adding Gaussian noise over images, rotated 90, 180 and 270 degree images, and mirroring them on X and Y axes. With these techniques total image number is increased to 13,500. Example images from the data set can be seen in Figure 2 and Figure 3.

The specifications of the device used for training are:

- Intel i7-10870H CPU
- Nvidia RTX 3060 Max-P GPU
- 16 GB DDR4 RAM
- 1TB NVMe SSD
- Ubuntu 20.04 OS
- Python 3.9 + PyTorch 1.8.2 + CUDA 11.2 + cuDNN 8.2.

The training was completed in 100 epoch and 16 batch size using the device with the specified features.

5. Results

As it is known, certain performance metrics must be known in order to evaluate a model formed as a result of training. The most common three of these metrics are given below:

- Precision

\[
Precision = \frac{TP}{TP + FP} \tag{1}
\]

- Recall

\[
Recall = \frac{TP}{TP + FN} \tag{2}
\]
• Mean Average Precision

\[ mAP = \frac{1}{n} \sum_{k=1}^{n} AP_k \]  

As stated before, the data set used for training consists of 640 pixel images. The performance of the trained model in other image dimensions can be evaluated using performance metrics 1, 2 and 3.

Performance metrics of images of different sizes are given in the Table 5. The data set used in this test process consists of 250 images. These images are not included in either the validation or the training set. In the dimension test of the model, the images in the test data set were resized according to the image dimensions and the inference process of the model was started. As can be seen from the table, the highest performance was obtained in 640 pixel images. This situation, which is caused by training the model with data of this size, is quite normal. In addition to this situation, the main reason for choosing 640 pixels in model training is that high pixel sizes require high processing power.

### Table 1. Performance table of different size of images.

<table>
<thead>
<tr>
<th>Image Size (px)</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>mAP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>74</td>
<td>86</td>
<td>79</td>
</tr>
<tr>
<td>384</td>
<td>77</td>
<td>82</td>
<td>84</td>
</tr>
<tr>
<td>512</td>
<td>77</td>
<td>93</td>
<td>85</td>
</tr>
<tr>
<td>640</td>
<td>78</td>
<td>94</td>
<td>85</td>
</tr>
<tr>
<td>768</td>
<td>77</td>
<td>91</td>
<td>52</td>
</tr>
<tr>
<td>896</td>
<td>75</td>
<td>84</td>
<td>81</td>
</tr>
<tr>
<td>1024</td>
<td>71</td>
<td>76</td>
<td>74</td>
</tr>
<tr>
<td>1152</td>
<td>69</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td>1280</td>
<td>69</td>
<td>56</td>
<td>58</td>
</tr>
<tr>
<td>1408</td>
<td>67</td>
<td>48</td>
<td>51</td>
</tr>
<tr>
<td>1536</td>
<td>62</td>
<td>44</td>
<td>46</td>
</tr>
</tbody>
</table>

After the model training was completed, validation was performed for fine-tuning of hyperparameters. The validation data set consists 4000 images that were not included in the training set. As a result of this process, it can be understood whether there is overfitting or underfitting in the trained model. Ideally, the validation loss curve should follow the same path as the training loss curve. The loss curves obtained as a result of validation and training are given in Figure 4 and Figure 5.

![Fig. 4. Mean squared error loss function value.](image1)

![Fig. 5. Cross entropy loss function value.](image2)

After the validation process, the performance of the model on different images was observed with the previously mentioned test set. In these experiments, it was seen that the model classified the test set with 85% mean average precision. Four different images from the results of this test are shown in Figure 6.
Fig. 6. Results of inference.

6. Field Testing

The USV has been tested in a real open water-body environment. The test site is a reservoir located in Blessington, Co. Wicklow in Ireland. The lake is called Poulaphouca Reservoir which is a man-made lake connected to River Liffey. One of the main reasons for choosing this site is the absence of waves in the water body of the reservoir. The first test was a visual inspection for the stability of movement on water. The second test was about the accuracy and reliability of the object detection algorithm YOLOv5s6 in real-time. The designed and developed USV in this study is an unmanned vehicle. As mentioned in Section 2, unmanned vehicles cannot operate autonomously. Due to this specification, the USV has been controlled remotely in all tests. However, the implemented real-time deep learning model performed its inference processes autonomously. In the first part of the test, i.e. movement test, maneuverability of the USV was inspected. Testing process was designed by relocation of the USV on the water surface by remote control. The movement pattern of the device has been selected as square in first phase of the test and circular in the second phase. The USV maneuvered smoothly round the corners and the performance of the thrusters has been satisfactory with no significant drifting or unbalanced movement being observed. In the second part of field testing, YOLOv5s6 model has been tested for real-time detection tasks. Two different objects with different colors and shapes were used for testing. These objects can be seen in Figure 7 and Figure 8. Each object has been detected 10 times along different directions. The tests conducted confirmed the ability of YOLOv5s6 in detecting objects from both long and close ranges. In addition to the range, partial objects were also detected accurately by the deep learning model.

Fig. 7. White bottle detection.

Fig. 8. Orange cylinder box partial detection.
7. Conclusions and Future Work

7.1. Main Conclusions

In this paper a real-time implementation of a deep learning algorithm (YOLOv5) has been proposed for applications in autonomous object detection and tracking on open sea surface. The algorithm has been implemented in an unmanned surface vehicle. The preliminary results indicate how the real-time algorithm can facilitate autonomous maintenance and monitoring of water bodies. The developed hardware device with the real-time implementations reported in this study is a proof of concept. However, this has potential for many other projects involving real-time autonomous computer vision or electronic eye. The deep learning model was integrated into the device and worked with an average of 15 frames per second in the preliminary design, which has scope to be refined further.

7.2. Future Work

The deep learning model and the device designed in this study are open to many different developments. Although mobile systems have processing power limitations, it is obvious that the desired success rate can be achieved with a well-trained model. As a next step, two different developments can be worked on. The first of these is to integrate the control systems required for autonomous driving of the device. Secondly, the data set used in training the model can be expanded by using different data manipulation techniques. Indeed, in this study detecting and tracking of objects have been the focus. Collection of objects is also required for some applications, which could be considered as a topic for future research. The real-time algorithms for detection and tracking in this study can be further enhanced for collecting objects as well.

References


