Marine Pollution Bulletin 164 (2021) 111974

Contents lists available at ScienceDirect

Marine Pollution Bulletin

journal homepage: www.elsevier.com/locate/marpolbul

Automatic detection of seafloor marine litter using towed camera images and deep learning

Dimitris V. Politikos [a,*], Elias Fakiris [b], Athanasios Davvetas [c], Iraklis A. Klampanos [c], George Papatheodorou [b]

[a] Institute of Marine Biological Resources and Inland, Hellenic Centre for Marine Research, 16452 Argyroupoli, Greece
[b] Laboratory of Marine Geology and Physical Oceanography, Department of Geology, University of Patras, 26504 Patras, Greece
[c] Institute of Informatics and Telecommunications, National Centre for Scientific Research "Demokritos", Agia Paraskevi, 15310 Athens, Greece

ARTICLE INFO

Keywords:
Seafloor marine litter
Object detection
Mask R-CNN
Deep learning
Aegean Sea
Mediterranean Sea

ABSTRACT

Aerial and underwater imaging is being widely used for monitoring litter objects found at the sea surface, beaches and seafloor. However, litter monitoring requires a considerable amount of human effort, indicating the need for automatic and cost-effective approaches. Here we present an object detection approach that automatically detects seafloor marine litter in a real-world environment using a Region-based Convolution Neural Network. The neural network is trained on an imagery with 11 manually annotated litter categories and then evaluated on an independent part of the dataset, attaining a mean average precision score of 62%. The presence of other background features in the imagery (e.g., algae, seagrass, scattered boulders) resulted to higher number of predicted litter items compare to the observed ones. The results of the study are encouraging and suggest that deep learning has the potential to become a significant tool for automatically recognizing seafloor litter in surveys, accomplishing continuous and precise litter monitoring.

1. Introduction

Anthropogenic litter is a growing threat to the health of marine ecosystems around the world (Eriksen et al., 2014; UNEP, 2016; Canals et al., 2020). Every year, tonnes of litter are entering into the oceans from land and marine-based sources and end up being deposited on the sea surface, seafloor, shorelines, and beaches (van Sebille et al., 2015; NOAA, 2016). This can cause the death of terrestrial and marine organisms when they are trapped inside or ingest litter (NOAA, 2014; Neves et al., 2015; Clark et al., 2016; Nadal et al., 2016; Auta et al., 2017) and set human health at risk due to consumption of seafood contaminated by litter.

Aerial and underwater video and photography-based monitoring of marine litter objects is increasingly adopted to assess their type, distribution and abundance (Angiolillo et al., 2015; UNEP, 2016; Gonçalves et al., 2020; Papachristopoulou et al., 2020). The knowledge gained is used to understand the severity of the problem, design effective clean-up programs and increase public awareness for reducing litter waste (Alkalay et al., 2007). However, the required human effort and financial cost needed to process the litter data outcomes from video footage is considerable, indicating the need for automatic and cost-effective approaches.

Deep neural networks are being extensively used for image understanding in many domains, such as image classification, human behavior analysis, face recognition, autonomous driving and video analysis (Ngiam et al., 2011; Szegedy et al., 2013). Object detection is a task in deep learning for image understanding that beyond object recognition it aims at locating, classifying and segmenting objects contained in images (Zhao et al., 2019). There are several deep learning neural networks that can be used for object detection and segmentation. They go from region proposal-based such as Faster R-CNN (Ren et al., 2015) and Mask R-CNN (He et al., 2017) to regression-based methods such as YOLO (Redmon et al., 2016) and SSD (Liu et al., 2016). In all cases, the full process is automated after a proper training of these detectors, in which the network accepts as inputs images and learns to extract meaningful features of the objects. The trained network is, then, able to detect objects in new, unseen before, visual data, with no need of any manual labeling.

In the marine environment, object detection networks have used images to detect surface, beached and underwater litter (Valdenegro-Toro, 2016; Fulton et al., 2019; Watanabe et al., 2019; Hong et al., 2020;

* Corresponding author.
E-mail address: dimpolit@hcmr.gr (D.V. Politikos).

https://doi.org/10.1016/j.marpolbul.2021.111974
Received 26 November 2020; Received in revised form 28 December 2020; Accepted 29 December 2020
Available online 20 January 2021

Tharani et al., 2020; van Lieshout et al., 2020) and fish (Li et al., 2015; Sung et al., 2017; Xu and Matzner, 2018; French et al., 2020; Martin-Abadal et al., 2020).

The development of an automatic seafloor litter detection system has the potential to provide a faster and cheaper alternative to current manual data analysis methods mainly used for litter assessment (Watters et al., 2010; UNEP, 2016). However, the characteristics of seafloor litter objects in a real environment present several peculiarities, making their detection a quite difficult problem. Video footage with different camera angles, zoom levels and lighting conditions can make litter objects hardly visible. Additionally, the numerous types of litter, the different shapes for the same litter type, litter that is degraded over time and buried in the seabed and the presence of other background structures (e.g., rocks, seagrass) can easily confuse a detector. These are challenging aspects when attempting to implement object detection approaches in this kind of imagery.

In this paper, we present an object detection approach with the aim to automatically detect seafloor marine litter in a real-world environment using a Region-based Convolution Neural Network. The neural network is trained on a seafloor imagery with 11 manually annotated litter categories, namely plastic bags, plastic bottles, plastic sheets, plastic cups, cans, fishing nets, plastic small plastic sheets, tires, big objects, plastic caps and unspecified. The trained network is then evaluated on an independent part of the dataset. The goal is to explore the efficiency of object detection to handle such a complex imagery composed of multiple litter categories obtained from video footage under varying zoom levels, angles and light shadings.

2. Material and methods

2.1. Data acquisition and presentation

The dataset used in this work was collected in the framework of "Integrated information and awareness campaign for the reduction of plastic bags in the marine environment" program (LIFE DEBAG - LIFE14 GIE/GR/001127[1]). Seafloor imagery was acquired through a towed underwater camera (TUC) that was mobilized from a small vessel, conducting survey lines over a regular grid in Ermoupolis bay, Syros Island, Greece (Fig. 1). Ermoupolis bay covers a seafloor area of 83 km^2 (~830 Hectares), with a maximum depth of 47 m and an average of 16 m. It is formed as a harbor basin, semi-isolated from the open sea by a pair of breakwaters placed on its Eastern outer boundaries, leaving a 300 m wide opening in between. The main urban part of Syros Capital city, Ermoupolis, is developed to the Western of the bay. The main port of Syros is located on the NW part of the bay, while a range of urban, industrial and touristic activities are spread all along its W – WE coast, including walkabouts, touristic facilities, marinas, fishing ports, boatyards and a major shipyard. The seafloor is mainly composed of plain (sandy/muddy) sediments, occasionally covered by low algae or sparse seagrass, scattered boulders, concrete blocks (next to the breakwater, docks and the shipyard) while there are instances of disturbed sandy seafloor due to dragged anchors or dredging. The water clarity in the area is considered excellent most of the time, allowing for high quality underwater image/video acquisition even in the deeper parts of the bay during daylight, without the need for underwater lights.

The TUC used was a SeaViewer 950 analog towed camera, equipped with a GoPro5 action camera to record 4 k video data. During the project, six seafloor litter surveys were implemented, covering more than 405 ha of seafloor in 36 h of underwater videos, corresponding to 72 km of survey lines. The camera was towed behind the vessel at a maximum speed of 1.5 knots, continuously adapting its cable-out as to be kept at about 1.5–2 m above the seafloor during the whole duration of the visual inspections. The recorded video files were carefully time-stamped according to the exact UTC time provided by the GPS, while GPS fixes where saved in the Hypack 2013 navigation software[2] per second. Video inspection was planned only during daylight, in cloudless conditions and specifically between late morning and early afternoon so that the sun was close to its zenith. This way, most of the acquired video was considered of very high clarity and contrast, while in cases where the water turbidity was high or when vast changes of color illumination or light flares were apparent, the respective video frames were excluded from the analysis. The same was applied when the distance of the camera to the seafloor was long enough to cause unacceptably low image contrast.

Litter items were manually recognized in those videos and were classified according to the EU Marine Strategy Framework Directive: Technical Subgroup on Marine Litter for monitoring marine litter in European seas (MFSD Directive, 2013) and included various types of litter classes such as plastic bags, big objects, plastic bottles, cans, plastic caps, cups, fishing nets, plastic sheets, small plastic sheets, tires, large metallic objects, and textiles.

2.2. Image preprocessing

The seafloor marine litter dataset included 635 video frame images with a resolution of 1920 × 1080 pixels. In total, 1166 litter items were manually identified in the images and 2D bounding boxes were drawn around the items using *LabelImg* image annotation tool (Labelimg, 2015). For each image, *LabelImg* generated a ".xml" file, which contains the class and the coordinates of 4 corners of a bounding box for each item annotated in the image. Due to numerous types of observed litter objects, big objects of different material (metallic, wooden, plastic) were unified to "big object" class, and several not identifiable objects were merged to a new class, called "unspecified". The type and frequency of annotated litter objects are shown in Fig. 2. Most frequent classes were plastic bottles and plastic bags, whereas plastic caps and fishing nets were the minority classes. Sample litter images with their associated classes and bounding boxes are shown in Fig. 3.

Due the small size of 635 video frame images, we applied data augmentation, a common process that artificially increases the amount of the images through transformations to reduce the risk of overfitting in network training and potentially improve its performance (Shorten and Khoshgoftaar, 2019). For saving computational time, we used off-line data augmentation, i.e., the dataset was augmented once and the new images were annotated again before processing them in the neural network. Augmentation was conducted using the PIL package (Clark, 2015) and included horizontal and vertical flipping of images, changes in brightness and noise addition. After data augmentation, the size of the new dataset increased to 3910 images. In addition, images were resized to 832 × 448 pixels to balance limitations in computational resources and make the training process faster.

2.3. Object detection approach

2.3.1. Network selection

The object detection approach used here was based on Mask R-CNN (He et al., 2017). This is a state-of-the art algorithm for object detection (Zhao et al., 2019). There are several other competing architectures: Faster R-CNN (Ren et al., 2015), YOLO (Redmon et al., 2016) and SSD (Liu et al., 2016), being generally comparable in terms of performance in underwater studies (Fulton et al., 2019; Hong et al., 2020; Tharani et al., 2020). All detectors are designed to find multiple objects in an image and assign for each object a class label (type of object) and its location through a rectangular bounding box (boundary localization of the object). Mask R-CNN further returns a mask label, which defines a finer spatial layout of the object in a pixel-to-pixel form. The generation of

[1] http://www.lifedebag.eu, as viewed 11 November 2020.

[2] https://hypack-2013.software.informer.com/.

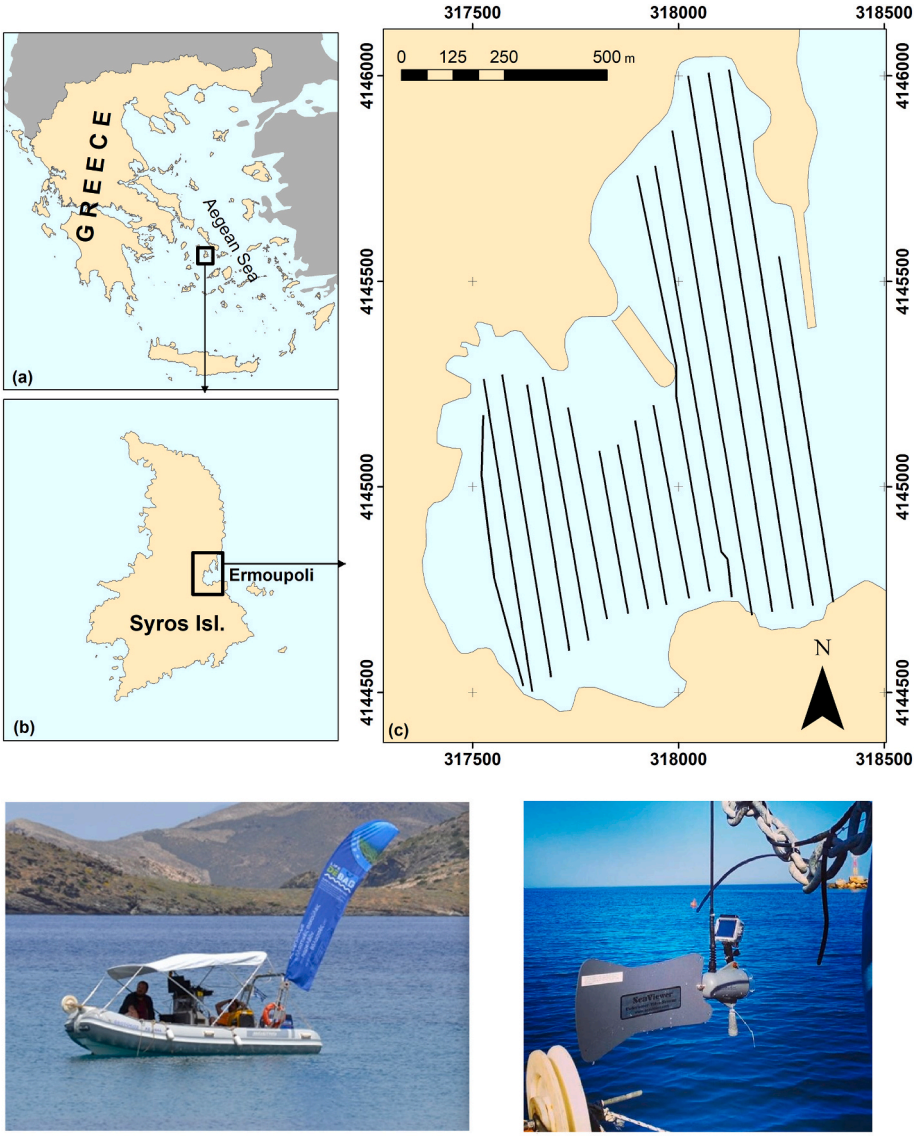

Fig. 1. Seafloor imagery was acquired through a towed underwater camera that was mobilized from a small vessel, which conducted surveys from a series of transects in Ermoupoli bay, Syros Island, Greece.

masks that fit the shapes of detected objects more closely lends itself better to further understanding of a dataset and to extensions, e.g., as an interactive tool, usable by domain experts. Mask R-CNN was adopted due to its decent performance in the literature, while being a more complete platform for extensions, via its pixel-wise segmentation of detected objects. Mask representation was basically used in our case study as an auxiliary task for exploring to what extent the surface area of litter objects can affect network's performance.

Mask R-CNN is built on two basic architectures: a *backbone* CNN architecture and a *head* CNN network (He et al., 2017). First, the image is processed though the convolutional layers of the *backbone*, which extracts a map with meaningful features from the image. Early layers detect low level features (e.g., edges, corners), whereas later layers detect high level features (object instances). Different families of CNN architectures are available for choosing the *backbone* architecture, such as ResNets (Szegedy et al., 2016), MobileNets (Howard et al., 2017; Sandler et al., 2018), GoogLeNet (Szegedy et al., 2015) and VGG (Simonyan and Zisserman, 2014). The output of the *backbone* network is passed to a *head* CNN network, which contains two stages. During the first stage, a Region Proposal Network (RPN) uses a sliding window method to propose candidate regions, in which the objects may be. RPN generates two outputs: a set of variable sized regions, called Regions of Interest (ROI), which define if the regions have an object or not and a set of rectangular bounding boxes, which define if the object is within the box or not. An algorithm, called Non-Maximum Suppression, evaluates the intersection of candidate bounding boxes with ground truth and keeps the boxes that fit the object more closely. In the second stage, the ROI Align layer considers the section of feature map corresponding to each (*variable sized*) RoI and warp it into a fixed size. These outputs are then fed into a fully connected layer for class and bounding box prediction and into a CNN for mask prediction. A schematic view of Mask R-CNN architecture is shown in Fig. 4.

A loss function is fundamental to the training of a neural network, as it aims to quantify how successful the network is in predicting the ground truth (Géron, 2017). During learning, the neural network iteratively adjusts its parameters, or weights, in order to minimize its loss function. The Mask R-CNN uses a multi-task loss function on each sampled ROI determined by the sum of classification loss, bounding box loss and mask loss.

Our litter network was trained with the MobileNetV1 backbone CNN (Howard et al., 2017) and was initiated using the pretrained weights from the COCO detection dataset (Lin et al., 2014). MobileNetV1 was

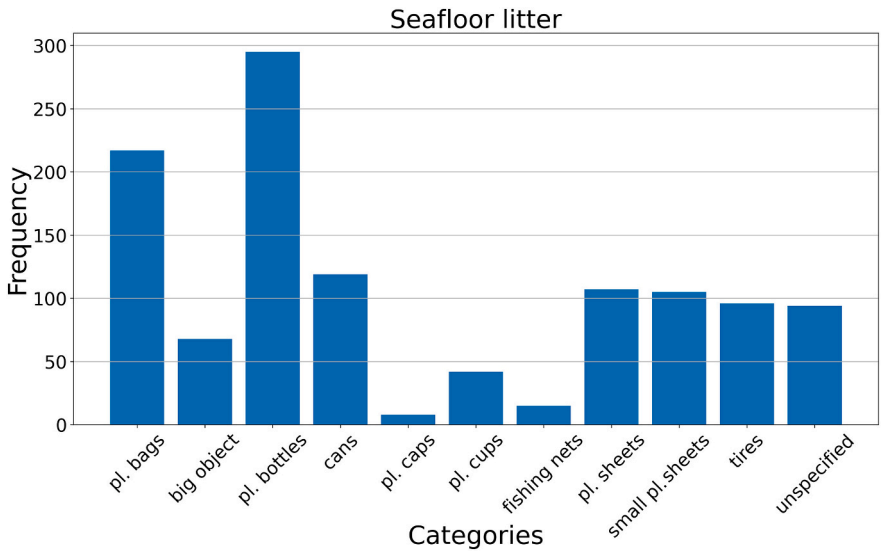

Fig. 2. Histogram of litter items per category from the manually annotated seafloor marine litter imagery. Litter categories include: plastic (pl.) bags, plastic (pl.) bottles, plastic (pl.) sheets, plastic (pl.) cups, cans, fishing nets, plastic (pl.) small plastic sheets, tires, big objects, plastic (pl.) caps and unspecified.

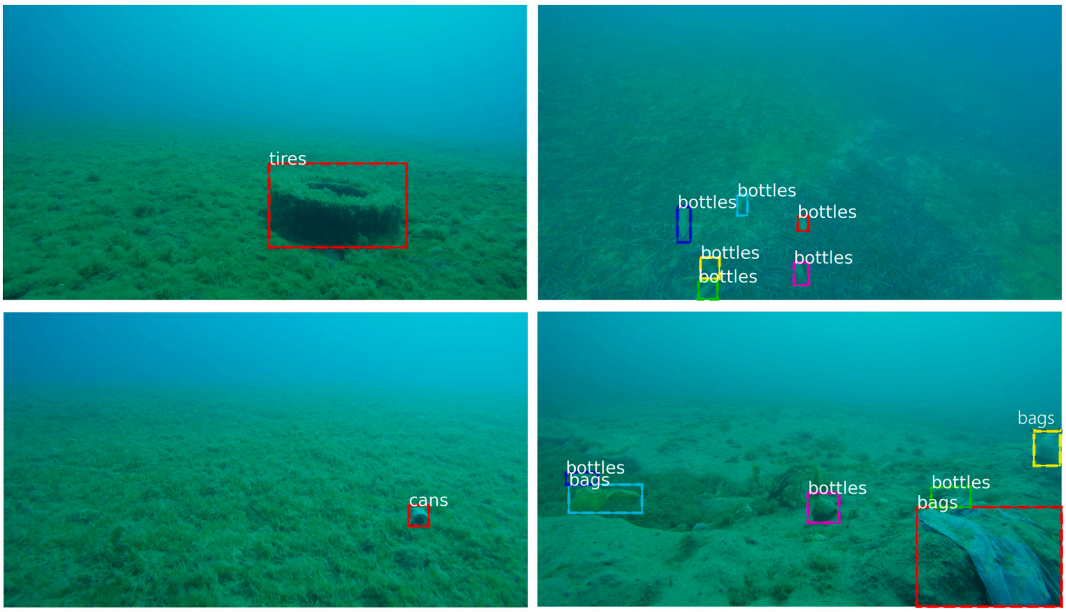

Fig. 3. Sample images with litter objects along with their classes and bounding boxes, manually annotated by the *LabelImg* image tool (Labelimg, 2015).

chosen because it is less compute-intensive compare to other networks, without losing its accuracy (Howard et al., 2017). This was important so as to overcome memory limitations during the implementation and training of the network. The implementation of Mask R-CNN was based on the third-party implementation of Mask R-CNN (Abdulla, 2017), which was built under Keras (Chollet et al., 2018) and TensorFlow (Abadi et al., 2016).

2.3.2. *Training*

For practitioners, pretrained R-CNNs provide their training parameters as initial weights for resolving a new task (Fulton et al., 2019; Zhao et al., 2019). This process is called transfer learning and can potentially provide better network performance than learning without transferred knowledge, as well it can save computational time in training (Pan and Yang, 2010). Then, a retrain and fine-tuning of the new task is conducted by tweaking the hyperparameters of the network to achieve the best possible performance, using two different parts of the dataset, called training and validation sets. Hyperparameters are non-learned parameters, e.g., the learning rate or the batch size, which define a neural network's architecture and training process (Géron, 2017). The training of the trainable parameters (the weights) of the network takes place using the training set, while hyperparameters get adjusted according to performance against the validation set. Once the network is trained, then it is evaluated against a third subset of the dataset, called the testing set.

During training, the performance of the network is monitored in terms of training and validation losses over consecutive epochs. Epochs represent the number of times that the training dataset is passed forward and backward through the network to refine network weights (Goodfellow et al., 2015). Loss is monitored per epoch, separately for training and validation sets and the trained network is saved at the end of each epoch. The network with the lowest loss is then used for evaluating its performance on the testing set, i.e., on data not previously seen by the network during training.

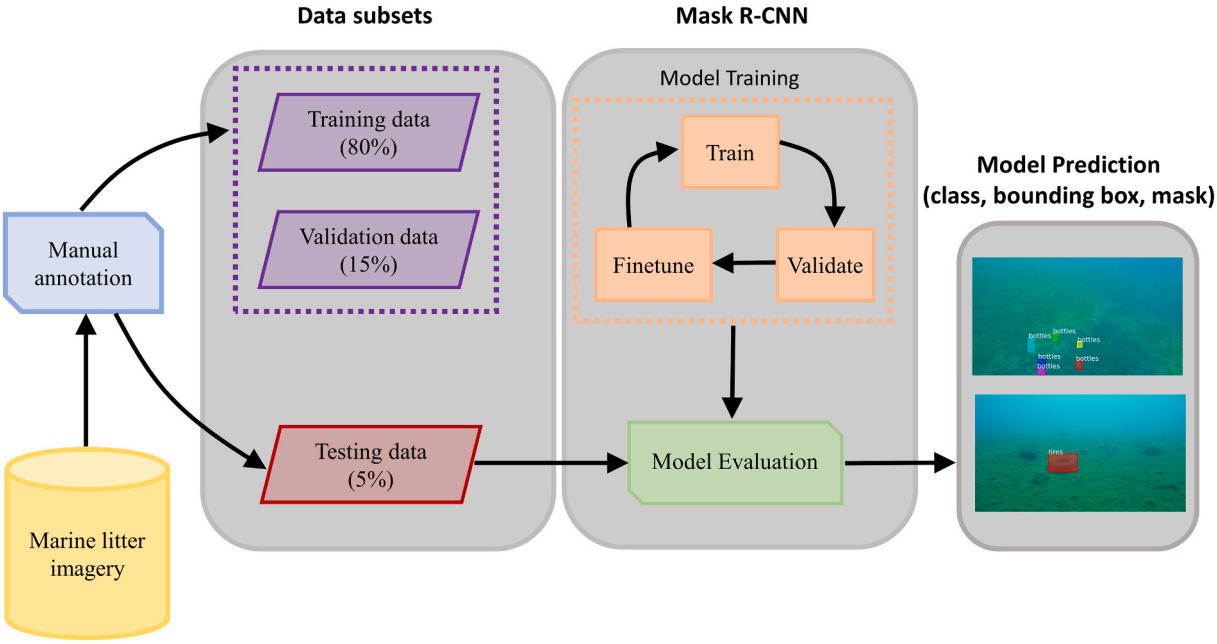

Fig. 4. Overview of the workflow implemented on the seafloor litter imagery.

For our problem, the dataset was split into three subsets: 80% training set (3051 images), 15% validation set (566 images) and 5% testing set (193 images). In total, we considered 11 classes for litter objects and 1 class for the background. The key hyperparameters used for training the network were: batch size, learning rate, weight decay, anchor scales and epochs. The tested and adopted values of the hyperparameters are shown in Table 1. The rest of the hyperparameters of the Mask R-CNN were automatically chosen according to the default configuration. Fig. 5 illustrates the workflow for seafloor litter detection.

2.3.3. Evaluation

The model was evaluated on the testing set using the two standard performance metrics of Intersection Over Union (IoU) and mean Average Precision (mAP) (Fulton et al., 2019; Watanabe et al., 2019). IoU measures how well predicted bounding boxes fit the location of an object and is defined by the ratio,

$$IoU = \frac{(ground\ truth \cap prediction)}{(ground\ truth \cup prediction)},$$

where *ground truth* represents the true bounding box and *prediction* represents the predicted bounding box. The higher the IoU, the higher the overlap of the two bounding boxes. Typically, a threshold value (in percentage) is set to define that the prediction is correct. At a given threshold value, the average precision (AP) is calculated in each litter class, considering the True Positives (TP) and False Positives (FP) (precision = TP/(TP + FP)), derived by the comparison of ground truth with and prediction over all images. Then, mAP is computed as the average of APs at different recall values, where recall is the ratio: recall = TP/(TP + FN), where RP = True Positives and FN = False Negatives. IoU and mAP metrics concisely describe the accuracy and quality of object detections. For our analysis, we considered three IoU threshold values of 25%, 50% and 75%.

Table 1
List of hyperparameters tested to train Mask R-CNN neural network. Bold values indicate the values that finally adopted to achieve the best possible performance of the neural network.

Hyperparameter	Description	Values
Number of classes	Number of different classes for detection in images	**11 + 1 (background)**
RPN anchor scales	Region proposal network's anchor box sizes for locating candidate objects in the image	[8, 16, 32, 64, 128] **[16, 32, 64, 128, 256]** [32, 64, 128, 256, 512]
Weight decay	A parameter applied on the loss function to prevent overfitting, though regularization.	0.01, **0.001**, 0.0005, 0.0001
Epochs	Number of times that the training dataset is passed forward and backward through the network to refine model weights.	**40**, 100, 200
Learning rate	A parameter used to train a network via gradient descent. The gradient descent algorithm multiplies the learning rate by the gradient.	**0.001**, 0.004, 0.0001, 0.0005
Batch size	Number of images used in one iteration of network training.	1, **4**

3. Results

Examples of detection comparisons between actual and predicted litter items from images of the testing set are shown in Fig. 6. The neural network succeeded to predict the class and the location of litter items for various litter sizes and number of objects found in the image (Fig. 6a–b; c–d; Fig. A1 in Appendix A). Mis-predictions can be attributed to false detections of background image features (Fig. 6e–f; Fig. A1 in Appendix A), detection failures especially when the litter object was difficult to be distinguished from the background (Fig. 6g–h; Fig. A1 in Appendix A) and correct detections but false classifications of litter objects (Fig. A1 in Appendix A).

The litter network attained a mAP score of 0.76 on training set and 0.62 on testing set for IoU = 50% (Table 2). Training mAP slightly increased to 0.77 and testing mAP to 0.66 for IoU = 25%, whereas testing mAP declined to 0.45 for IoU set to 75% (Table 2).

To assess the effectiveness of Mask R-CNN to detect each litter class separately, we computed, additionally, the average precision (AP) per litter class (Fig. 7). AP was higher than mAP average for plastic bags, fishing nets, tires and plastic caps (Fig. 7, red line), reaching a value ≥0.79. In contrast, the most mis-detected classes were plastic cups, cans and unspecified, attaining and precision of 0.47, 0.49 and 0.52,

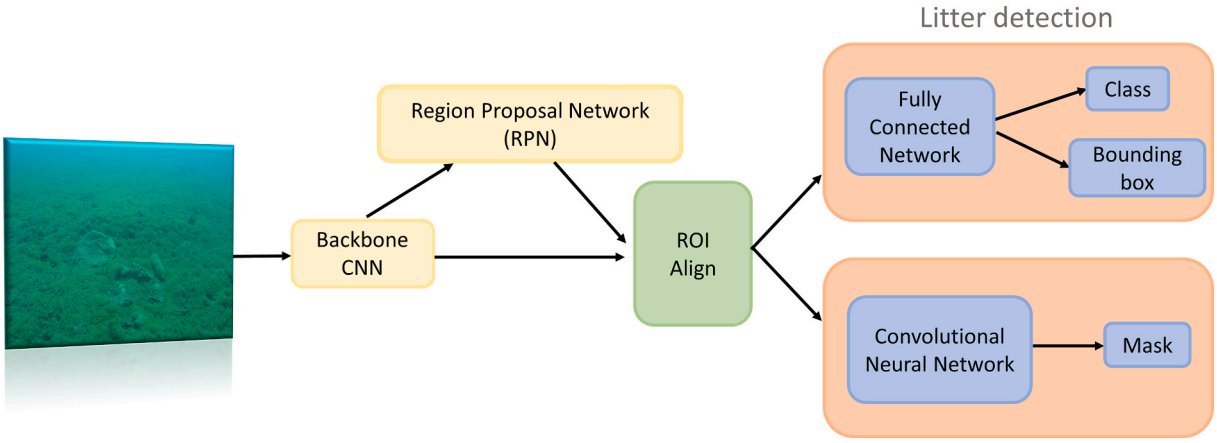

Fig. 5. Schematic view of Mask R-CNN architecture.

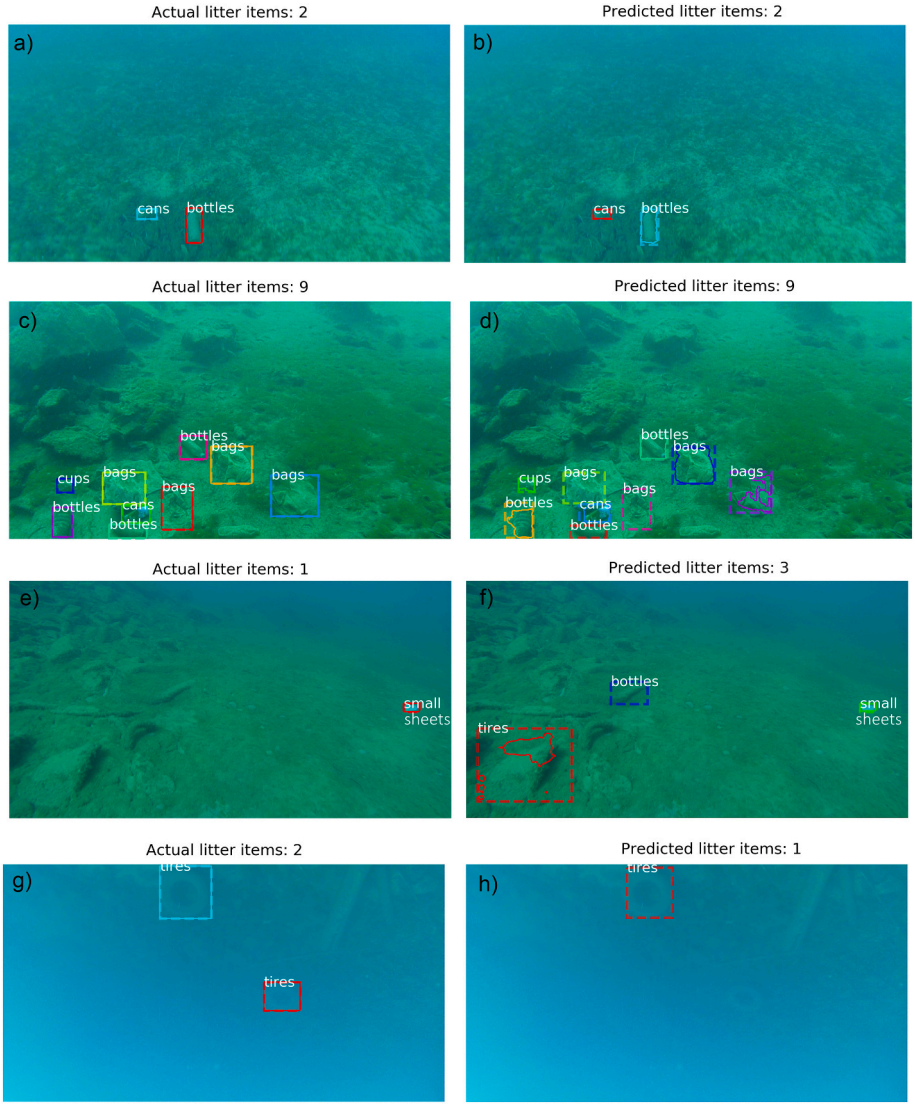

Fig. 6. Examples of detection comparisons between actual and predicted litter items with images from the testing set.

respectively. The per-class precision for plastic bottles, plastic sheets, plastic small plastic sheets and big objects lied around the mAP = 0.62 (Fig. 7, red line).

Our findings also illustrated that Mask R-CNN tended to predict additional litter items, which did not exist in actual images (Fig. 8). This was more evident especially for plastic bags and plastic bottles and more moderate for plastic sheets, small plastic sheets, tires, big objects and unspecified. Additional samples of litter images that illustrate this

Table 2
Mean Average Precision (mAP) on training and testing sets for different IoU thresholds.

IoU threshold	Training mAP	Testing mAP
25%	0.77	0.66
50%	0.76	0.62
75%	0.75	0.45

discrepancy can be found in the Appendix A (Fig. A2).

4. Discussion

This study presented an example of automatic detection of seafloor marine litter in a real-world environment, based on a deep learning object-detection approach. The approach involved the training of R-CNN neural network on an image collection with 11 manually annotated litter categories. The trained network achieved a mAP over all litter classes of 62%. In several litter types (plastic bags, fishing nets, tires, plastic caps), the network was even more effective, reaching an average precision of >79%.

Accurate litter predictions did not seem that they can be attributed exclusively to a single factor such as the number of training size of each litter category, the size of the litter object or the distinct shape of litter object. Instead, litter categories exhibited their own unique performance. For instance, although "plastic bottles" class was represented in the dataset with a high frequency (Fig. 2), it's AP was close to mAP (=0.62). Contrary, the "tires" class had much fewer instances (Fig. 2), but it's AP was high (AP = 0.83) due it's unique shape in the dataset. Additionally, the "plastic caps" class represented small size objects, achieving a high AP = 0.8, whereas the "plastic small sheets" class which also represented small size objects attained a lower AP = 0.59. Concurrently, "bags" class attained a relative high AP (~0.8) although the shape of bags was variable in the imagery, being however a major class (Fig. 2) during network training. We also note that the surface area of mask labels for litter objects on the testing set verified that median mask size was similar for successful (AP = 1.0) and unsuccessful (AP = 0.0) litter predictions (Fig. 9). This indicated that the size of litter objects did not appear to affect network's performance.

Mispredictions of the network can be attributed to the complexity of the dataset. First, litter objects of the same class had no defined geometric shape in several categories (e.g., plastic bags, plastic sheets), but also similar litter shapes belonged to different classes (e.g., plastic bags, plastic sheets), posing an extra difficulty in the training process. Second, in some images, litter objects had been either degraded or buried in the seabed to such extent that they were hardly distinguishable from the sea floor even by human observers. Third, other background structures found in the images, such as seagrass, rocks and light shadings constituted misleading patterns during network training. All these issues had a limiting effect on the precision of litter detections. However, it is still

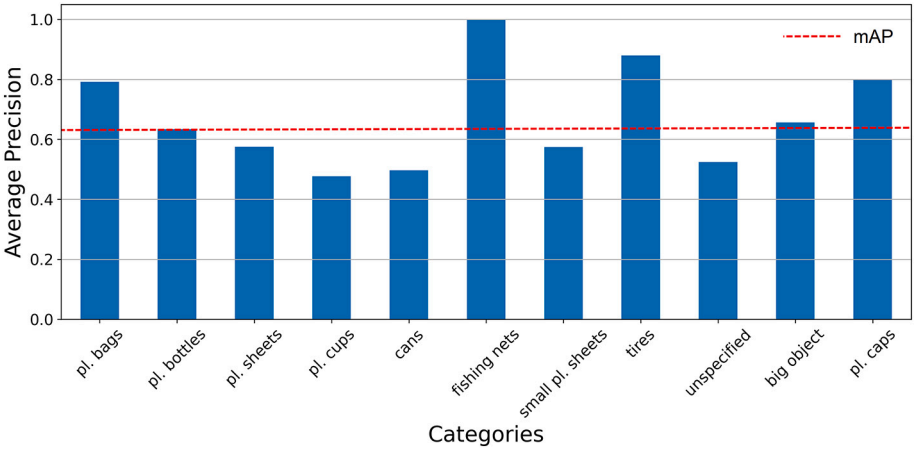

Fig. 7. Average precision per litter category from images on the testing set. Red dot line indicates the mAP over all litter categories, as estimated for IoU = 50%. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

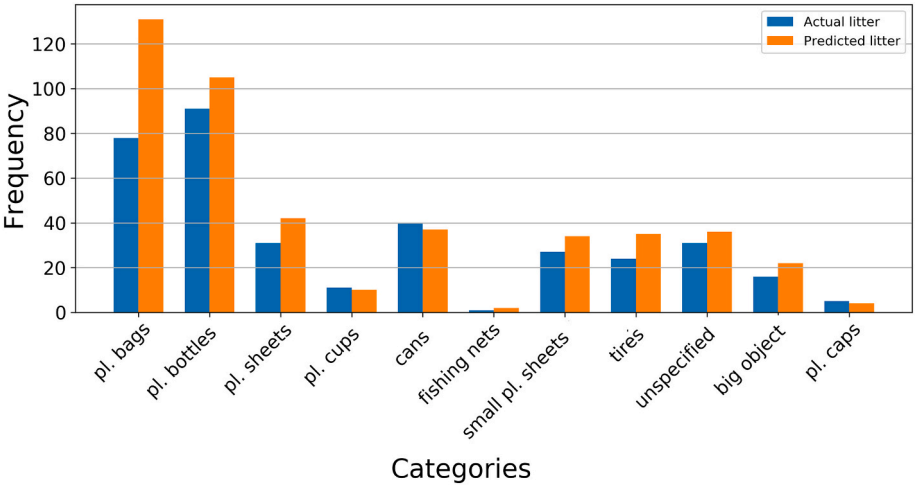

Fig. 8. Number of litter objects found on the testing set per class against those predicted by the litter network.

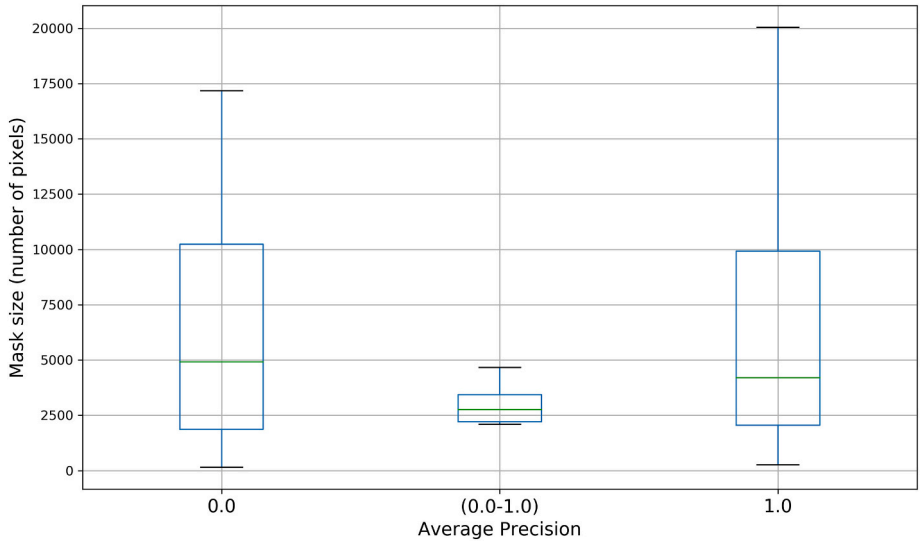

Fig. 9. Boxplots showing the surface area of detected litter objects in the testing set using mask label, categorized for three classes of average precision: 0, (0–1) and 1. Mask label represent the detected litter objects found in the images in a pixel-to-pixel form.

unclear how to train object-detection neural networks able to resolve these cases. For instance, one potentially useful approach may be to apply pre-processing on the data, targeted at these problematic areas. Another potentially beneficial approach could be to apply targeted, domain-driven, regularization alongside using larger and more diverse datasets. These directions, in the context of marine litter, are left for future research.

Despite the intrinsic characteristics of the dataset, the performance of our litter network was generally aligned with other studies for litter detection. Hong et al. (2020) adopted Mask R-CNN and Faster R-CNN to detect undersea litter, flora and fauna, achieving an average precision around 0.55. Fulton et al. (2019) used different types of object detection architectures to detect marine trash (one class) and fish (one class) in a realistic environment, reaching a precision between 0.31 and 0.81. Additionally, the deep-learning object-detection algorithm YOLO v3 recognized four litter classes (plastic bottles, plastic bags, drift wood, other litter) from the Japanese waters with mean average precision of 0.77 (Watanabe et al., 2019). Compared with other imagery used for litter detection (Fulton et al., 2019; Watanabe et al., 2019), our images also included litter objects with various sizes, shapes, lighting conditions and zoom levels. However, our dataset was challenging for the object detection task due to its multiple annotated litter categories ranging from tiny to very big objects, confusing background with irrelevant structures (e.g., scattered boulders, rocks) and obscure seabed (e.g., algae, seagrass).

It is worth noting that our network training was subject to some setbacks, which affected its predictability on the testing set. Hyper-parameter tuning was proved relatively unstable. Moderate variations of "anchor scales" and "learning rate" from default values set in Mask R-CNN induced significant decline in mAP below 0.4. Although network performance was markedly improved when batch size increased from 1 (default value in Mask R-CNN) to 4, testing higher batch sizes was not feasible due to limited computation resources.

We also note that it was out of our scope, at this stage of research, to compare and contrast different object detection architectures (Fulton et al., 2019; Watanabe et al., 2019; Tharani et al., 2020). Instead, we adopted one architecture, focusing on the environmental implications and limitations that emerged from the automated image analysis. Nevertheless, we acknowledge that the adopted Mask R-CNN may not represent the best possible detection network for our dataset but nonetheless it provides an initial baseline for further development and comparison with other object detection frameworks (Redmon et al., 2016). For instance, simplifying the litter detection task by merging similar – with regard to shape or size - litter categories would mitigate the imbalanced and multi-class nature of the dataset and could provide a good chance for improving network's performance. Other options for network improvement such as changing the training process through freezing of specific layers (Fulton et al., 2019), or testing alternative backbone architectures (Martin-Abadal et al., 2020) are possible. These directions could significantly improve performance and are worthy for future work.

5. Conclusion

The implementation of machine and deep learning techniques in Marine Science is a relatively new research field of great potential (Malde et al., 2019). Marine scientists make a valuable effort to combine standard image analysis methods with machine learning and neural network algorithms for litter monitoring (Martin et al., 2018; Fallati et al., 2019; Gonçalves et al., 2020; Kako et al., 2020). Additionally, developing object detection algorithms that can run in near real-time on underwater camera systems is of high ecological importance (Martin-Abadal et al., 2020; Fulton et al., 2019; Tseng and Kuo, 2020). Experts in marine litter research have highlighted the importance of processing image and video data though machine/deep learning and computational object detection for achieving automatic, rapid and large-scale monitoring of litter (Canals et al., 2020). Our work should be seen as a contribution towards this direction, with the perspective to automatically monitor seafloor litter presence, allowing us to achieve continuous and precise litter assessments under minimum human-effort.

Supplementary comparison examples of predicted litter detection and ground truth can be seen in Figs. A1 and A2. Supplementary data to this article can be found online at https://doi.org/10.1016/j.marpolbul.2021.111974.

CRediT authorship contribution statement

Dimitris V. Politikos: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing – original draft. **Elias Fakiris:** Data curation, Conceptualization, Writing – review & editing. **Athanasios Davvetas:** Conceptualization, Software, Methodology, Writing – review & editing. **Iraklis A. Klampanos:** Conceptualization, Software, Methodology, Writing – review & editing. **George Papatheodorou:** Data curation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was partially funded in the context of LIFE DEBAG Project (LIFE14 GIE/GR/001127 - http://www.lifedebag.eu/).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al., 2016. Tensorflow: large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:160304467. https://arxiv.org/abs/1603.04467.

Abdulla, W., 2017. Mask R-CNN for Object Detection and Instance Segmentation on Keras and TensorFlow. Git code. https://github.com/matterport/Mask_RCNN.

Alkalay, R., Pasternak, G., Zask, A., 2007. Clean-coast index-a new approach for beach cleanliness assessment. Ocean Coast. Manage. 50 (5), 352–362. https://doi.org/10.1016/j.ocecoaman.2006.10.002.

Angiolillo, M., di Lorenzo, B., Farcomeni, A., Bo, M., Bavestrello, G., Santangelo, G., Cau, A., Mastascusa, V., Cau, Al, Sacco, F., Canese, S., 2015. Distribution and assessment of marine debris in the deep Tyrrhenian Sea (NW Mediterranean Sea, Italy). Mar. Pollut. Bull. 92, 149–159. https://doi.org/10.1016/j.marpolbul.2014.12.044.

Auta, H.S., Emenike, C.U., Fauziahet, S.H., 2017. Distribution and importance of microplastics in the marine environment: a review of the sources, fate, effects, and potential solutions. Environ. Int. 102, 165–176. https://doi.org/10.1016/j.envint.2017.02.013.

Canals, M., et al., 2020. The quest for seafloor macrolitter: a critical review of background knowledge, current methods and future prospects. Environ. Res. Lett. https://doi.org/10.1088/1748-9326/abc6d4 (In press).

Chollet, F., et al., 2018. Keras 2.1.3. https://github.com/fchollet/keras.

Clark, A., 2015. Pillow (PIL Fork) Documentation. https://buildmedia.readthedocs.org/media/pdf/pillow/latest/pillow.pdf.

Clark, J.R., Cole, M., Lindeque, P.K., et al., 2016. Marine microplastic debris: a targeted plan for understanding and quantifying interactions with marine life. Front. Ecol. Environ. 14 (6), 317–324. https://doi.org/10.1002/fee.1297.

Eriksen, M., Lebreton, L.C.M., Carson, H.S., Thiel, M., Moore, C.J., Borerro, J.C., Galgani, F., Ryan, G., Reisser, J., 2014. Plastic pollution in the world's oceans: more than 5 trillion plastic pieces weighing over 250,000 tons afloat at sea. PLoS One 9 (12). https://doi.org/10.1371/journal.pone.0111913.

Fallati, L., Polidori, A., Salvatore, C., Saponari, L., Savini, A., Galli, P., 2019. Anthropogenic marine debris assessment with unmanned aerial vehicle imagery and deep learning: a case study along the beaches of the Republic of Maldives. Sci. Total Environ. 693, 133581. https://doi.org/10.1016/j.scitotenv.2019.133581.

French, G., Mackiewicz, M., Fisher, M., Holah, H., Kilburn, R., Campbell, N., Needle, C., 2020. Deep neural networks for analysis of fisheries surveillance video and automated monitoring of fish discards. ICES J. Mar. Sci. 77 (4), 1340–1353. https://doi.org/10.1093/icesjms/fsz149.

Fulton, M., Hong, J., Islam, M.J., Sattar, J., 2019. Robotic detection of marine litter using deep visual detection models. arXiv:1804.01079. https://arxiv.org/abs/1804.01079.

Géron, A., 2017. Hands-on Machine Learning With Scikit-Learn and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems. O'REILLY, p. 547.

Gonçalves, G., Andriolo, U., Pinto, L., Bessa, F., 2020. Mapping marine litter using UAS on a beach-dune system: a multidisciplinary approach. Sci. Total Environ. 706, 135742. https://doi.org/10.1016/j.scitotenv.2019.135742.

Goodfellow, I.J., Bengio, Y., Courville, A., 2015. Deep Learning. MIT Press, p. 433.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. arXiv:1703.06870. https://arxiv.org/abs/1703.06870.

Hong, J., Fulton, M., Sattar, J., 2020. TrashCan: a semantically-segmented dataset towards visual detection of marine debris. arXiv:2007.08097. https://arxiv.org/abs/2007.08097.

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861. https://arxiv.org/abs/1704.04861.

Kako, S., Morita, S., Taneda, T., 2020. Estimation of plastic marine debris volumes on beaches using unmanned aerial vehicles and image processing based on deep learning. Mar. Pollut. Bull. 155, 111127. https://doi.org/10.1016/j.marpolbul.2020.111127.

Labelimg, 2015. Labelimg is a graphical image annotation tool and label object bounding boxes in images. https://github.com/tzutalin/labelImg.

Li, X., Shang, M., Qin, H., Chen, L., 2015. Fast Accurate Fish Detection and Recognition of Underwater Images With Fast R-CNN. OCEANS 2015. MTS/IEEE Washington, Washington, DC, pp. 1–5. https://doi.org/10.23919/OCEANS.2015.7404464.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: common objects in context. In: European Conference on Computer Vision, pp. 740–755.

Liu, W., et al., 2016. SSD: single shot multibox detector. arXiv:1512.02325. https://arxiv.org/abs/1512.02325.

Malde, K., Handegard, N.O., Eikvil, L., Salberg, A.-B., 2019. Machine intelligence and the data driven future of marine science. ICES J. Mar. Sci. 77 (4), 1274–1285. https://doi.org/10.1093/icesjms/fsz057.

Martin, C., Parkes, S., Zhang, Q., Zhang, X., McCabe, M.F., Duarte, C.M., 2018. Use of unmanned aerial vehicles for efficient beach litter monitoring. Mar. Pollut. Bull. 131, 662–673. https://doi.org/10.1016/j.marpolbul.2018.04.045.

Martin-Abadal, M., Ruiz-Frau, A., Hinz, H., Gonzalez-Cid, Y., 2020. Jellytoring: real-time jellyfish monitoring based on deep learning object detection. Sensors 20 (6), 1708. https://doi.org/10.3390/s20061708.

MFSD Directive, 2013. Guidance on Monitoring of Marine Litter in European Seas, p. 104. https://doi.org/10.2788/99475.

Nadal, M.A., Alomar, C., Deudero, S., 2016. High levels of microplastic ingestion by the semipelagic fish bogue Boops boops (L.) around the Balearic Islands. Environ. Pollut. 214, 517–523. https://doi.org/10.1016/j.envpol.2016.04.054.

Neves, D., Sobral, P., Ferreira, J.L., Pereira, T., 2015. Ingestion of microplastics by commercial fish off the Portuguese coast. Mar. Pollut. Bull. 101 (1), 119–126. https://doi.org/10.1016/j.marpolbul.2015.11.008.

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y., 2011. Multimodal deep learning. In: Proceedings of the 28th International Conference on Machine Learning, pp. 689–696.

NOAA, 2014. Report on the Occurrence and Health Effects of Anthropogenic Debris Ingested by Marine Organisms. Silver Spring, MD 19 pp. https://marinedebris.noaa.gov/occurrence-and-health-effects-anthropogenic-debris-ingested-marine-organisms.

NOAA, 2016. National Oceanic and Atmospheric Administration Marine Debris Program. Report on Modeling Oceanic Transport of Floating Marine Litter. Silver Spring, MD. 21 pp. https://marinedebris.noaa.gov/reports/modeling-oceanic-transport-floating-marine-debris.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. In IEEE Trans. Knowl. Data Eng. 22 (10) https://doi.org/10.1109/TKDE.2009.191.

Papachristopoulou, I., Filippides, A., Fakiris, E., Papatheodorou, G., 2020. Vessel-based photographic assessment of beach litter in remote coasts. A wide scale application in Saronikos Gulf, Greece. Mar. Pollut. Bull. 2020 (150), 110684 https://doi.org/10.1016/j.marpolbul.2019.110684.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: unified, real-time object detection. arXiv:1506.02640. https://arxiv.org/abs/1506.02640.

Ren, S., He, K., Girshick, R., Sun, J., et al., 2015. Faster R-CNN: towards real-time object detection with region proposal networks. arXiv:1506.01497. https://arxiv.org/abs/1506.01497.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2018. MobileNetV2: inverted residuals and linear bottlenecks. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520. https://doi.org/10.1109/CVPR.2018.00474.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6 (1), 60. https://doi.org/10.1186/s40537-019-0197-0.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556. https://arxiv.org/abs/1409.1556.

Sung, M., Yu, S.C., Girdhar, Y., 2017. Vision based real-time fish detection using convolutional neural network. In IEEE OCEANS 2017, 1–6. https://doi.org/10.1109/OCEANSE.2017.8084889.

Szegedy, C., Toshev, A., Erhan, D., 2013. Deep neural networks for object detection. In: Neural Information Processing Systems (Conference paper).

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the 26th International Conference on Neural Information Processing Systems 2, pp. 2553–2561.

Szegedy, C., Ioffe, S., Vanhoucke, V., 2016. Inception-v4, Inception-ResNet and the impact of residual connections on learning. arXiv:1602.07261. https://arxiv.org/abs/1602.07261.

Tharani, M., Amin, A.W., Maaz, M., Taj, M., 2020. Attention neural network for trash detection on water channels. arXiv:2007.04639. https://arxiv.org/abs/2007.04639.

Tseng, C.-H., Kuo, Y.-F., 2020. Detecting and counting harvested fish and identifying fish types in electronic monitoring system videos using deep convolutional neural networks. ICES J. Mar. Sci. 77 (4), 1367–1378. https://doi.org/10.1093/icesjms/fsaa076.

UNEP/MAP, 2016. Marine litter assessment in the Mediterranean 2015. In: United Nations Environment Programme Mediterranean Action Plan (UNEP/MAP) (86 pp.).

Valdenegro-Toro, V., 2016. Submerged marine debris detection with autonomous underwater vehicles. In: 2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA), pp. 1–7. https://doi.org/10.1109/RAHA.2016.7931907.

van Lieshout, C., van Oeveren, K., van Emmerik, T., Postma, E., 2020. Automated river plastic monitoring using deep learning and cameras. Earth and Space Sci. 7, e2019EA000960 https://doi.org/10.1029/2019EA000960.

van Sebille, E., Wilcox, C., Lebreton, L., et al., 2015. A global inventory of small floating plastic debris. Environ. Res. Lett. 10, 124006. https://doi.org/10.1088/1748-9326/10/12/124006.

Watanabe, J., Shao, Y., Miura, N., 2019. Underwater and airborne monitoring of marine ecosystems and debris. J. Appl. Remote. Sens. 13 (4), 044509 https://doi.org/10.1117/1.JRS.13.044509.

Watters, D.L., Yoklavich, M.M., Love, M.S., Schrodeder, D.M., 2010. Assessing marine debris in deep seafloor habitats off California. Mar. Pollut. Bull. 60, 131–138. https://doi.org/10.1016/j.marpolbul.2009.08.019.

Xu, W., Matzner, S., 2018. Underwater fish detection using deep learning for water power applications. arXiv:1811.01494. https://arxiv.org/abs/1811.01494.

Zhao, Z.-Q., Zheng, P., Xu, S., Wu, X., 2019. Object detection with deep learning: a review. IEEE Transactions on Neural Networks and Learning Systems 30 (11), 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865.