MACHINE LEARNING FOR BENTHIC TAXON IDENTIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Where seabed substrate is rocky, comprising of bedrock, boulders and cobbles, sampling and analysis of benthic ecosystems often relies on still images and video from underwater cameras. Processing benthic imagery to generate ecosystem information typically involves manual interpretation and annotation, which is a time consuming and expensive process and prone to human errors and biases. Machine learning can step in here to assist, if not fully replace manual annotation. Here, we develop an object detection model using Faster R-CNN to identify various epibenthic species from a high energy marine site with a rocky substrate. The model achieves an overall F1-score of 66.28% across 7 different benthic species. The work is a significant step in identifying various learnings and challenges associated with data in non-ideal conditions and we suggest methods to further improve the existing framework.

1 INTRODUCTION

Automated annotation of benthic imagery to identify and enumerate organisms and objectively characterise habitats would lead to increased information generation in less time and at reduced cost. Machine learning and computer vision algorithms have been proposed as a solution for this challenge and are being explored for benthic image annotation, particularly in coral reef (Galaz et al., 2021), and deep sea sedimentary ecosystems (Durden et al., 2021).

The above works prove that good water clarity as found in coral reefs, and a relatively homogenous sediment substrate as found in the deep sea are likely to be favourable for computer vision performance. However, the site from which the images were taken for this work is considered challenging for an automated approach due to poor water clarity with high particulate matter and rocky, heterogenous substrate. In this work we aim to use machine learning approaches based on a Faster R-CNN (Ren et al., 2015) framework to identify and annotate selected epifaunal taxa using images of a quality and quantity realistic to a typical benthic habitat survey in temperate coastal waters.

2 SEACAMS2 DATASET AND DATA ANNOTATION

The data for this work was collected by Bangor University as part of the SEACAMS2 project which was part-funded by the European Regional Development Fund (ERDF) through the West Wales and the Valleys Programme 2014-2020. The dataset consists of images from three different sites - Batch 1: 1498 images on rocky substrate in the vicinity of the Maarten Cornelis wreck site located west of the Isle of Anglesey, UK, acquired in June 2021, Batch 2: 212 images 3km from the Maarten Cornelis site collected in 2019 on similar rocky substrate and Batch 3: 2735 images on sediment substrate at the Lynas anchorage site to the East of the Isle of Anglesey, UK, acquired in 2021 (Figure 1).

We only use Batch 1 for this study. Out of the 1498 images, 800 images were been selected for species annotation. A total of 7 taxa were selected for the object identification task with consideration of their abundance, unique morphology, distinct colours and ecological significance. The 7 taxa are Common starfish (*Asterias rubens*), Bloody Henry starfish (*Henricia* spp.), Red/purple sea anemones (*Urticina* spp.), Top shells (Trochidae), Hermit crabs (Paguroidea) and Hornwrack (*Flustra foliacea*) (Figure 2). A team of marine biologists visually inspected the 800 images and annotated them.

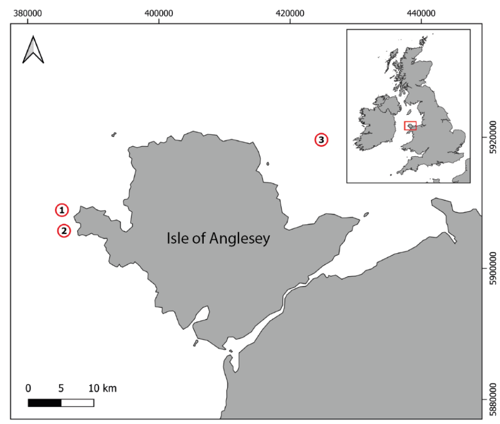


Figure 1: Location of benthic imagery sites used in this study. Numbers indicate batch number.

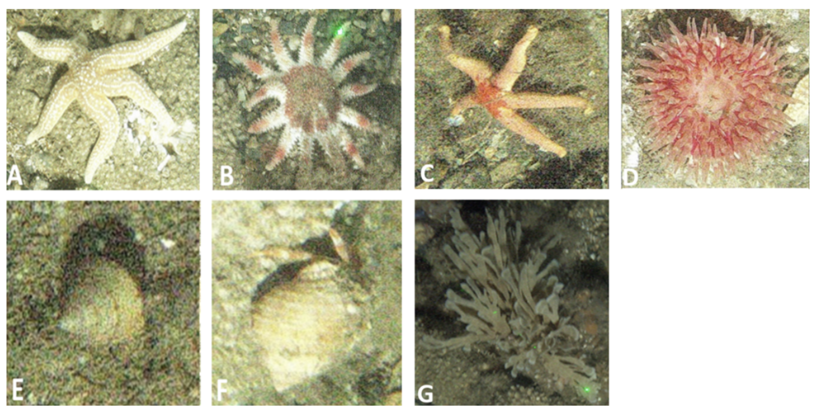


Figure 2: A) *Asterias rubens*, B) *Crossaster papposus*, C) *Henricia* spp., D) *Urticina* spp., E) Trochidae, F) Paguroidea, G) *Flustra foliacea*

With many of the images being of poor quality, taxonomic identification in some cases was not possible. All occurrences of identifiable organisms were labelled even where they were partially visible or present in an unusual form, including examples of varied sizes, starfish with single arm, and retracted anemone.

3 Model architecture and Training

We use the Faster R-CNN model with a ResNet-50 Feature Pyramid Network backbone (Lin et al., 2017) for the object detection framework.

Out of 800 images set aside for labelling, we obtained 757 images with relevant and usable annotations. The dataset was further divided into training, validation and test sets of size 605, 76 and 76 respectively, preserving class distribution in each fold. The data is heavily imbalanced across classes, as shown in Figure 3. The images were augmented using rotations, brightness and contrast adjustments to introduce variability and generalisation into the model.

The model was trained with batch size of 8 and using Adam optimiser (Kingma & Ba, 2014) with a base learning rate of 0.001, for a maximum of 500 iterations in each epoch, till convergence was obtained. The metric used was mean Average Precision (mAP) over IoU[0.05:0.9]. Training was performed on 4xQuadro RTX 6000 machines.

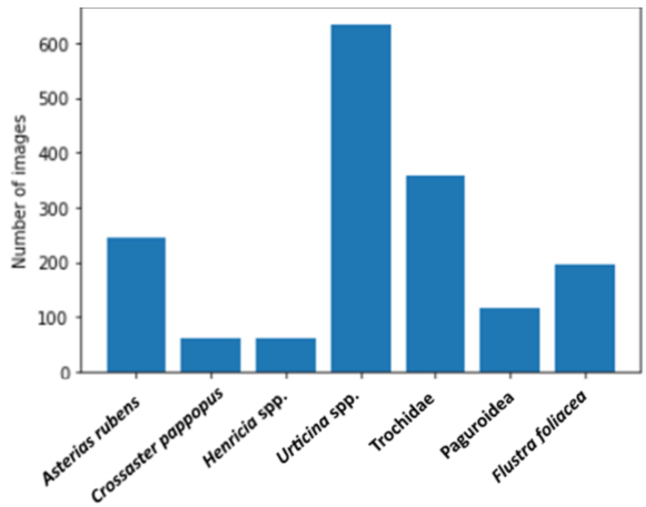


Figure 3: Number of images per class, representing the data imbalance

4 Results and Analysis

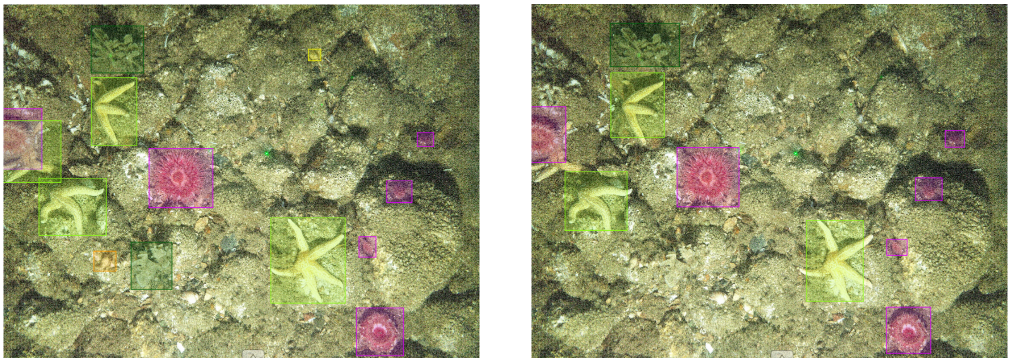
The model with the best mAP on the validation set was applied to the test set. The model achieves an average precision of 81.62%, average recall of 55.79%, and overall F1-score of 66.28%.

Table 1: Results of model on test set

Name	Precision (%)	Recall (%)	Number of true positives
Asterias rubens	96.63	75	86 out of 115
Crossaster papposus	100	10	1 out of 10
Urticina spp.	96.04	86.5	315 out of 364
Henricia spp.	100	43	6 out of 14
Paguroidea	78.65	58	70 out of 121
Flustra foliecea	100	58	19 out of 33
Trochidae	0	0	0 out of 54

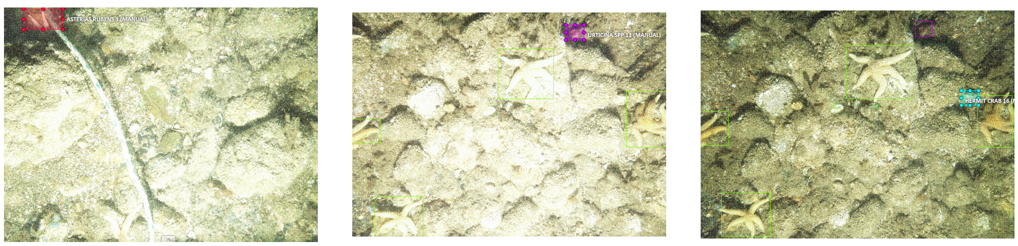
From Table 1, it can be inferred that the model performance reflects the inherent data imbalance to a large degree. Commonly occurring taxa with distinctive colouring and morphologies, such as *Urticina* spp. and *Asterias rubens*, performed well. The model also seems to have picked up instances of certain species which were not identified during labelling stage. The results of the poorly performing taxa such as Trochidae is a consequence of their colour and morphology being very indistinct and huge variation in image quality within the Maarten Cornelis wreck site combined with the insufficient number individuals present in the training set. Figures 4,5,6 show model outputs across the various kinds of image qualities and taxas.

We also tested the model on batches 2 and 3 and visually analysed the results as these batches were not annotated beforehand. Batch 2 which presents very similar substrate and conditions to Batch 1 makes it a good candidate to further validate the models (Figure 7a). Visual analysis was consistent with results from Batch 1. Batch 3 images represent an environment that is decidedly different from that of the training data and hence a good fit for testing the taxon identification capabilities of the model in different environments. The visual results from the Lynas anchorage site suggest that the model performed best in identifying *Urticina* spp. and Paguroidea (Figure 7b).



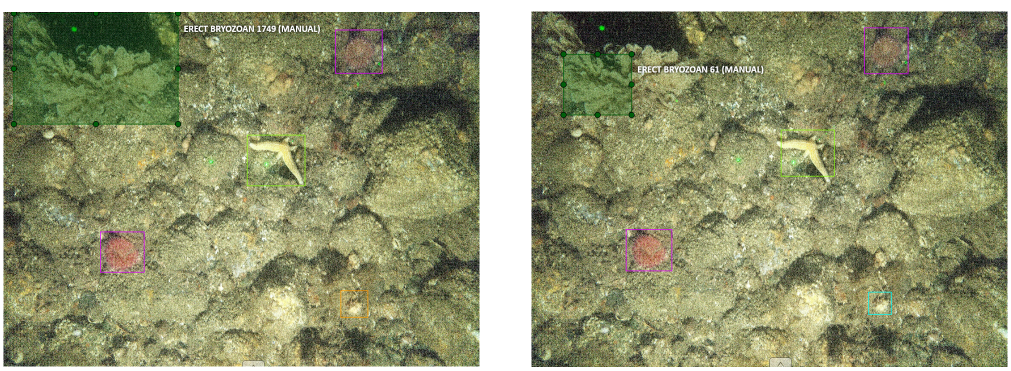
(a) Ground Truth (b) Prediction

Figure 4: Model predictions for an image from Batch 1. The boxes in pink, light green, dark green, yellow and orange represent *Urticina* spp., *Asterias rubens*, *Flustra foliacea*, Trochidae and Paguroidea respectively. Notice that the model fails to identify Trochidae and Paguroidea in this example.



(a) *Asterias rubens* (b) *Urticina* spp. (c) Paguroidea

Figure 5: An example of model identifying *Asterias rubens*, *Urticina* spp. (pink box) and Paguroidea (Hermit Crab) (blue box) which were unlabelled by human annotation. Notice how the model prediction improves with quality of the image taken from the same region for figures Figure 5b and 5c.

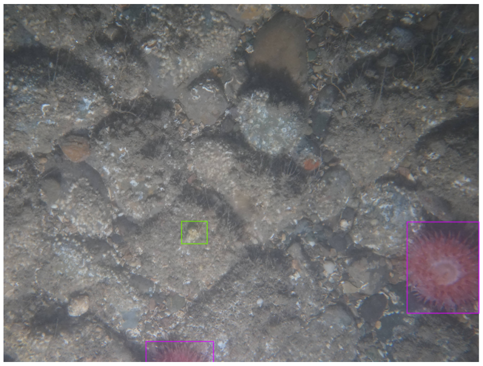


(a) Ground Truth (b) Prediction

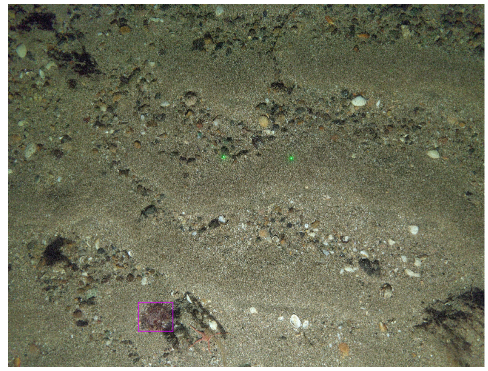
Figure 6: An example of how labelling impacts model output. The *Flustra foliacea* (Erect bryozoan) is annotated in a large box with significant substrate leakage.

5 Conclusion and Future Work

Our findings demonstrate that machine learning can form a valuable component in a semi-automated benthic image analysis workflow. Our results were encouraging despite the poor to moderate quality of imagery generated from challenging field conditions with poor water clarity and heterogeneous substrate. The Faster R-CNN model performed well for taxa that were abundant in the images and



(a) *Urticina* spp. (pink box) and Paguroidea (green box)



(b) *Urticina* spp. (pink box)

Figure 7: Examples of predictions from (a) Batch 2 and (b) Batch 3.

therefore provided plenty of training material, as well as those with distinct colours and morphology. These results are consistent with previous efforts to develop machine learning solutions for automated benthic image annotation (Piechaud et al., 2019).

The main challenges included varying image qualities and the resulting inconsistencies in human annotation. In order to address the various challenges, the following are being considered as part of ongoing and future work: i) improve the number and variety of training data, by completing annotation across all batches of data and collecting other datasets that have the relevant taxa, ii) address class imbalances using weighted losses and metrics, iii) improve label accuracy by using tighter bounding boxes and annotating distinct individuals of each taxa instead of labelling in groups when they appear clumped together, as seen in Figure 6. Annotating was proven to be difficult for poorer quality images, in which case label smoothing techniques will be applied, which assigns a probability to the label.

Acknowledgments

The authors thank CGG for permission to publish this work.The authors also thank Bangor University for providing the data. Special thanks to Pierre Roy, Rabia Shah, Chin Hang Lun and other colleagues in the CGG Environmental Science team and CGG AI Lab for discussions and suggestions.

References

Jennifer M Durden, Brett Hosking, Brian J Bett, Danelle Cline, and Henry A Ruhl. Automated classification of fauna in seabed photographs: The impact of training and validation dataset size, with considerations for the class imbalance. *Progress in Oceanography*, 196:102612, 2021.

Victor Galaz, Miguel A Centeno, Peter W Callahan, Amar Causevic, Thayer Patterson, Irina Brass, Seth Baum, Darryl Farber, Joern Fischer, David Garcia, et al. Artificial intelligence, systemic risks, and sustainability. *Technology in Society*, 67:101741, 2021.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117–2125, 2017.

Nils Piechaud, Christopher Hunt, Phil F Culverhouse, Nicola L Foster, and Kerry L Howell. Automated identification of benthic epifauna with computer vision. *Marine Ecology Progress Series*, 615:15–30, 2019.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.