Recognizing the Unknown: Open World Object Detection in the Deep Sea

Sabrina Fonseca Pereira, IT University of Copenhagen

Mentors: Kakani Katija, Eric Orenstein, Joost Daniels, Paul Roberts

Summer 2023

Keywords: Open World Object Detection, Computer Vision, Deep Learning, Embeddings

ABSTRACT

Traditional object detection models are designed to identify a fixed set of classes. This causes models to face limitations when used in dynamic environments like the deep sea, which is filled with rare and unknown organisms. In this study we explore the viability of Open World Object Detection (OWOD) for deep-sea exploration. Our findings indicate that OWOD shows considerable promise for real-world applications. The models trained were able to overcome catastrophic forgetting and incrementally learn about new and underrepresented classes. While encouraging, further research is still needed to develop robust systems and data pipelines for production-ready OWOD.

INTRODUCTION

The deep sea remains one of the least explored and understood regions of our planet. It harbours a wealth of biodiversity and resources. Yet, despite of its immense significance, little is known about the organisms that inhabit it and the biological processes that govern it.

Data collection is crucial for advancing our understanding of the complex ecosystems that exist in the ocean. While a range of methodologies — including acoustic sensing and ‘omics technologies — provide valuable insights into marine environments, visual data serves as an indispensable tool for validation.

Due to the increase of image data being collected by underwater vehicles, machine learning, especially Object Detection, became an essential tool to sift through this enormous amount of data. These models are incredibly useful for many tasks, not only to identify what organisms were captured in the images but also how many of each were found. However, all classification models, including object detectors, have an inherent flaw. They can only recognise a finite set of classes that were learnt during training. This is extremely limiting for models deployed in the real world where they will routinely encounter things they have never seen before.

The emerging field of Open World Object Detection (OWOD) intends to address this issue by giving models the ability to recognize a detected object as unknown. In this project, we test this novel approach to understand if it can be successfully applied to deep ocean data.

METHODS

This novel computer vision problem was first proposed in *Towards Open World Object Detection* [5] where the authors build on the ideas of contrastive clustering and energy to be able to differentiate between known and unknown detections.

Open World Object Detection (OWOD) aims to give object detection models the ability to recognise when they detected something outside of their training set. This approach leverages the ability of Object Detectors to localise with great precision objects it's never seen before.

Fig 1: Object detectors can recognise objects that are not in their training set. However, as they have to predict within their pre-defined classes we end up with many misclassifications.

To bridge the gap between object detectors and real-world applications, the model not only has to recognise unknown objects, but it also needs to incrementally learn new classes without forgetting the ones it already knows.

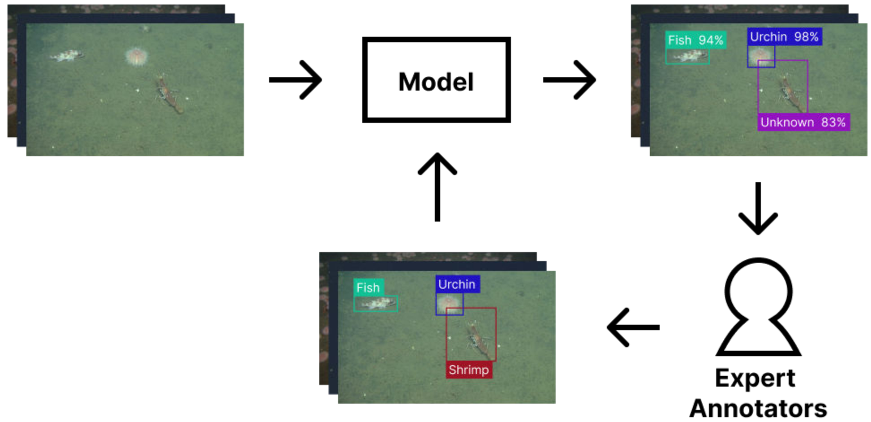

Fig 2: Unlabelled data goes into a trained model. The model then localises and classifies what it finds. Unknowns found by the model are reviewed by an expert human annotator who annotates those instances. The images with the new annotations will be part of the training set for the next training round.

THE MODEL

For this project we chose to experiment with PROB [4], the current state-of-the-art in the OWOD task.

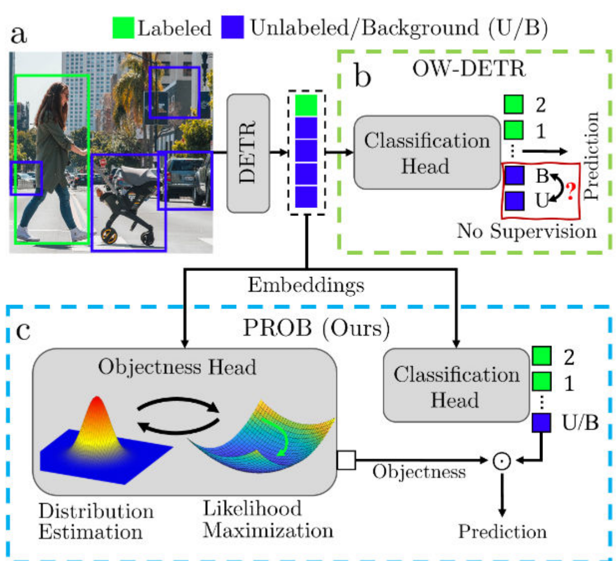

Figure 3. Diagram of the difference between PROBand another OWOD method. Adapted from 'PROB: Probabilistic Objectness for Open World Object Detection' by O. Zohar, K.-C. Wang, and S. Yeung. Dec. 02, 2022. doi: 10.48550/arXiv.2212.01424.

PROB introduces Probabilistic Objectness, where instead of directly identifying unknown objects, it separates the probability of "objectness" and the object class probabilities to improve unknown detections (Fig. 3).

To mitigate catastrophic forgetting, OWOD methods keep a small set of images, or exemplars that are used in the fine-tuning phase of each task. PROB also uses the concept of "objectness" to select the best possible exemplars instead of picking them at random like previous methods.

THE DATA

The data used in this project was originally packaged for the FathomNet 2023 Kaggle challenge [1]. It was collected by MBARI using different underwater vehicles in the Monterey Bay area between the surface and 1300 meters. The dataset includes over 25 thousand images and 70 thousand bounding box annotations of bottom-dwelling animals.

We combined the original training and evaluation sets into one, and rearranged it to create a new dataset suitable to train and evaluate OWOD methods.

The original dataset is extremely long-tailed and contains 290 classes from different levels in the taxonomic tree. Each class belongs to a superclass. There are a total of 20 superclasses in this dataset. To simplify the experiments and reduce the long-tail problem, classes were remapped to the corresponding superclass.

CREATING AN OPEN WORLD DATASET

To simulate the gradual encounter of new classes by a model, we created 4 different training sets with an increasing number of classes (Fig. 4). Task one includes the most common classes in the dataset, and task 4 includes the rarest. The evaluation set was the same for each task and included all 20 classes.

The Urchin class dominated the dataset with 35% of all instances. However, when subsampled valuable examples of underrepresented classes would be eliminated, therefore we decided to keep all images in the original dataset as it should reflect the real distribution of the organisms in their habitat.

Figure 4. Shows how many instances of each class are present in the training set used for each task. And the same for the evaluation set.

IMAGE EMBEDDINGS

Embeddings are the basis for OWOD methods, so to explore class similarities in the dataset and try to anticipate the performance of the OWOD models, we analysed the embeddings that represented the organisms in the dataset.

We used a MobileNetV2 [3] model pre-trained on ImageNet [2] to create vector representations for each instance in our dataset. These vectors were then projected into a 2-dimensional space using U-MAP [6] to allow for easier interpretation. In the projection, the basic premise is that points that are close to each other are likely to belong to the same class.

EVALUATION

The model was evaluated after each task, always with the same evaluation set. Performance was compared using AP50 as it is the standard evaluation metric for Object Detectors.

RESULTS

The results for the training of tasks 1 to 3 can be found in Table 1. The metric used for all is the AP50. As expected, we had good results for classes with the most instances in the training set ('Urchin') and classes that are very distinguishable ('Sea star', and 'Crab') and struggled with classes that were underrepresented. This is true for Task 1, and subsequent tasks.

When new classes were first introduced in Task 2 and Task 3, we can see the impact of catastrophic forgetting and how the model recovered a lot of the previously known information after the fine-tuning step. Despite being underrepresented in the dataset, the model showed adaptability and learned the new classes, such as 'Shrimp' and 'Soft Coral' introduced in Task 2 and 'Glass Sponge' and 'Feather Star' introduced in Task 3.

The 'Unknown' category, should capture all organisms previously introduced in past training sets, it performed poorly in Task 1 and progressively decreased as the model was trained on new tasks.

EXPERIMENT RESULTS

	Task 1	Task 2	Task 2 (fine-tuning)	Task 3	Task 3 (fine-tuning)
Urchin	70.86	0.00	65.63	0.00	63.32
Fish	52.16	9.09	40.93	0.00	40.69
Sea star	73.13	0.00	61.07	9.09	62.08
Anemone	53.51	9.09	43.91	0.00	40.93
Sea cucumber	62.50	9.09	58.30	9.09	55.43
Sea pen	55.26	9.09	46.75	0.00	40.74
Sea fan	58.07	9.09	41.74	9.09	38.64
Worm	33.29	0.00	27.01	0.00	25.37
Crab	69.44	0.00	56.60	0.00	57.70
Gastropod	21.42	4.55	14.85	0.00	14.53
Shrimp		4.75	11.27	0.73	22.97
Soft coral		0.00	11.67	0.00	10.37
Glass sponge				9.48	13.34
Feather star				5.69	25.85
Eel					
Squat lobster					
Barnacle					
Stony coral					
Black coral					
Sea spider					
Unknown	**0.86**	**0.20**	**0.35**	**0.03**	**0.30**

Table 1. AP50 for each class after training on each task. It also shows the results before the fine-tuning steps for tasks 2 and 3.

DISCUSSION

Despite the discouraging numbers, especially for the Unknown class, we still believe this method shows promise. When visually investigating the results we can see that the incomplete coverage of annotations in the dataset had a big impact on the results. In Figure 5, we can see how the model managed to find the shrimp was an "Unknown", and how it came up with some plausible predictions for unknown organisms in the image that were not in the original annotations.

Figure 5.A. Original image with annotations.

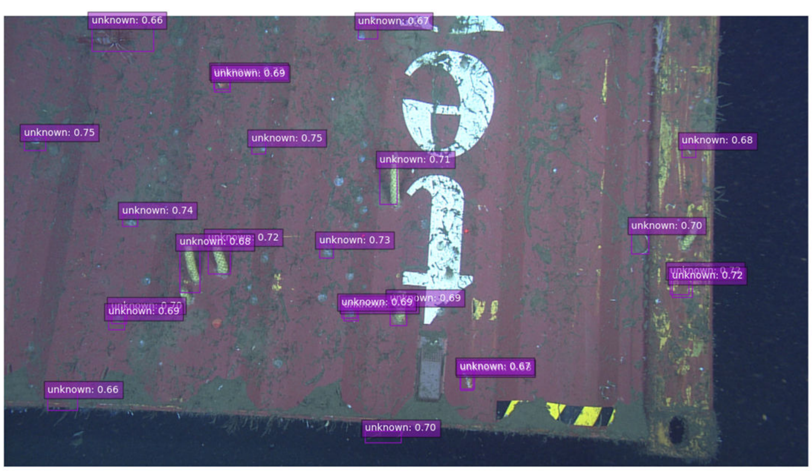

Figure 5.B. Same image with model detections after Task 1. Here, the model does not know the class "Shrimp", but it found it as an "Unknown".

During the incremental training across tasks, we can see the extent of catastrophic forgetting. Fine-tuning helped to alleviate this but did not fully recover the original performance levels.

CONCLUSIONS/RECOMMENDATIONS

Complete annotation of training data is fundamental to improving model performance. Apart from giving us misleading evaluation metrics, when we don't annotate organisms as such, we are inherently teaching the model to recognise them as background. As this can have a great impact on model performance, a big focus on implementing object detectors should be complete annotation. Having less training data comprised of fully annotated images should yield better results.

It is also worth experimenting with different OWOD models to compare their performance. As well as different dataset configurations. There is no standard of how these benchmarking datasets should be structured, so it's worth trying different training schedules – training on rare organisms first to understand if this has an impact on the final performance for underrepresented classes for example.

This work represents an active area of research, and while promising, the existing codebases are not ready for production use. A lot of work still needs to be done to implement pipelines for training and inference to successfully deploy OWOD models.

ACKNOWLEDGEMENTS

Special thanks to the National Science Foundation (Grant #2230776), the David & Lucile Packard Foundation and MBARI. George Matsumoto for facilitating this incredible programme. The entire BioInspiration lab for having me as an intern this Summer, and the video lab for their annotation work as without it, this project would not have been possible.

References:

[1]
Ben Woodward, Eric Orenstein, Genevieve Patterson, Lilli Carlsen, ‘FathomNet 2023: Shifting seas, shifting species: Out-of-sample detection in the deep ocean’. Accessed: Aug. 27, 2023. [Online]. Available: https://kaggle.com/competitions/fathomnet-out-of-sample-detection

[2]
O. Russakovsky *et al.*, ‘ImageNet Large Scale Visual Recognition Challenge’. arXiv, Jan. 29, 2015. doi: 10.48550/arXiv.1409.0575.

[3]
M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, ‘MobileNetV2: Inverted Residuals and Linear Bottlenecks’. arXiv, Mar. 21, 2019. doi: 10.48550/arXiv.1801.04381.

[4]
O. Zohar, K.-C. Wang, and S. Yeung, ‘PROB: Probabilistic Objectness for Open World Object Detection’. arXiv, Dec. 02, 2022. doi: 10.48550/arXiv.2212.01424.

[5]
K. J. Joseph, S. Khan, F. S. Khan, and V. N. Balasubramanian, ‘Towards Open World Object Detection’. arXiv, May 09, 2021. doi: 10.48550/arXiv.2103.02603.

[6]
L. McInnes, J. Healy, and J. Melville, ‘UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction’. arXiv, Sep. 17, 2020. doi: 10.48550/arXiv.1802.03426.