AI-based fish detection and classification at OBSEA underwater observatory

Oriol Prat Bayarri (oriol.prat.bayarri@upc.edu), Pol Baños Castelló (pol.banos.castello@upc.edu), Enoc Martinez (enoc.martinez@upc.edu), Joaquin del Rio (joaquin.del.rio@upc.edu)

Universitat Politècnica de Catalunya (Spain)

Introduction

Nowadays the use of artificial intelligence (AI) is expanding in a multitude of areas of knowledge, helping and accelerating many aspects of work. One of them is the object detection and classification, in this case applied to ecosystem monitoring. The monitoring of fish populations is of great ecological interest, and it is essential to know their status in order to determine what fishing impact is appropriate and sustainable, considering the resilience of these habitats. For many years, biologists have counted single species one by one in pictures in order to obtain several metrics such as abundance, biomass and biodiversity. Today, it is possible to greatly reduce the time spent in this process with computer vision tools such as AI-based object detection. A great example of the comparison between the use of AI and a human has been investigated by Ellen et al. [2020], and they found an improvement in the results when applying these computer vision methods.

There are several models that serve these methodologies through neural networks, one of the most established and stable is YOLO (You Only Look Once). YOLO divides an input image into a grid and each grid cell predicts multiple bounding boxes along with class probabilities. This grid-based approach allows YOLO to achieve impressive detection speeds. In addition, YOLO uses a convolutional neural network architecture to extract relevant features of the input image, enabling accurate object detection.

Data and Methods

This work uses a data-set of images obtained from OBSEA (1505 at the moment), a cabled underwater observatory located 4 km from the coast of Vilanova i la Geltrú (Catalunya, Spain) and at a depth of 20 m. The observatory has several marine sensors, including three video cameras from which the frames are extracted periodically.

In order to train the algorithm, the first step is to label the images, (work carried out with LabelImg software). Then the total number of images is divided into three blocks, one for training, one for testing and the other for validation, normally following this percentages respectively; 70%, 20% and 10%. The first part goes through the different levels of the convolutional neural network in order to determine the characteristics of the images. The second one is used to infer the information contained in images it has never seen before, in order to detect the objects in these new images. And the third part, as the name says, is used to validate the results, to collect metrics that give us information about is the model performance. The most commonly used are accuracy, precision, recall, mean average precision and confusion matrices.

In order to find the best model fit in the OBSEA custom data-set, a series of trainings have been performed using different characteristics in order to get the best results, that is, the best weights. The combinations used three species number groups (5, 12 and 16), five models (*nano*, *small*, *meduim*, *large* and *extra large*), five learning rates (0.01, 0.03, 0.0033, 0.00011, 0.000037), various data augmentation parameters and five random initial seeds.

Key findings

The best results for the 12 species group so far refer to the extra-large model, learning rate of

0.000037 and using data augmentation. In Figure 1 there are two relevant graphs, the metrics/precision(B) and the metrics/recall(B), both with good values indicating a correct prediction measure and sensitivity.
On the other hand, in Figure 1 the confusion matrix provides important information on the detection of each species, and how much it is confused with other species or with the background. The ideal situation would be the diagonal with values of unity for all species, however for some species such as *Mullus surmuletus* the detection is poorer.

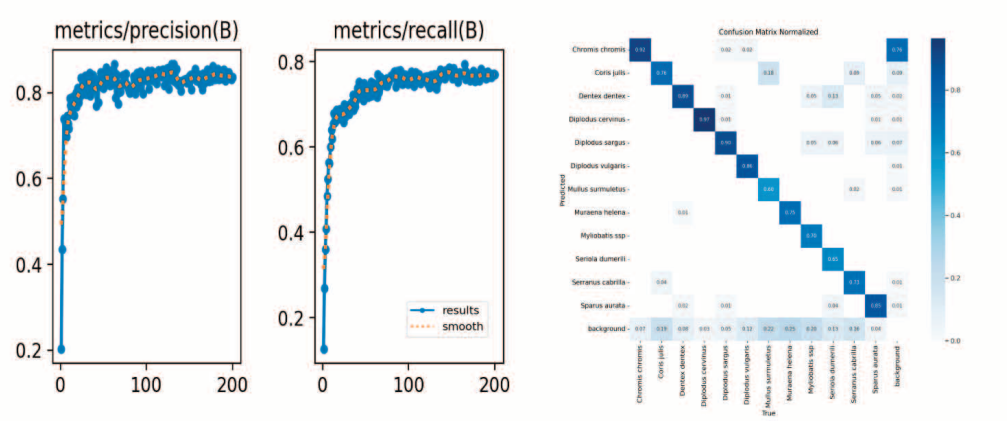

Figure 1 Metrics of a) precision, b) recall and c) confusion matrix normalized of the 12 species considered.

Figure 2 shows some examples of fish detection at OBSEA's environment, with their associated bounding box and detection confidence.
The model is capable of highly accurate detection of a wide diversity of fish. However, some environmental conditions such as water turbidity can affect the detection. The next steps are focused on further improving the model, increasing the number of species and applying real-time detection to video streams.

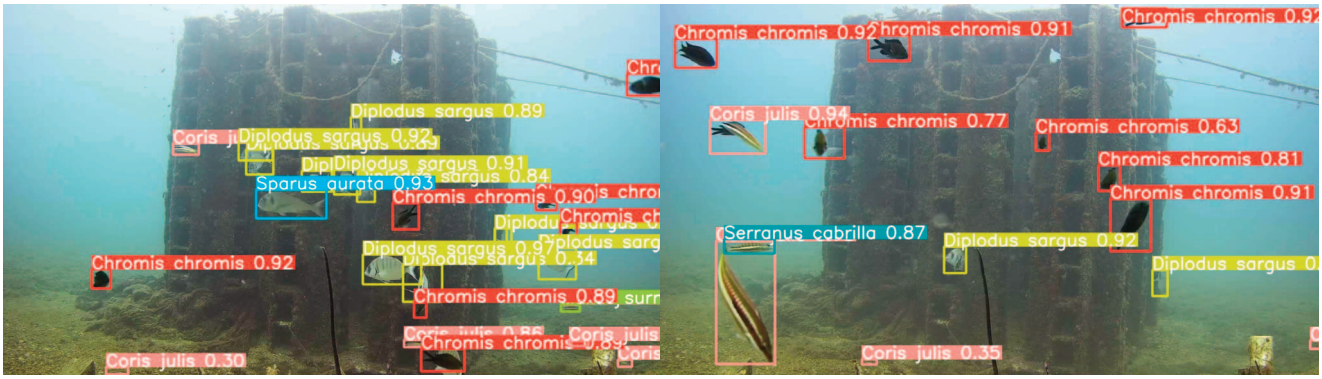

Figure 2 Examples of fish detections using the extra-large model at OBSEA enviroment (Vilanova i la Geltrú, Catalunya, Spain).

Acknowledgement
This work was developed in the framework of the Research Unit Tecnoterra (ICM-CSIC/UPC) and the following project iMagine, funded from the European Union Horizon Europe Programme – Grant Agreement number 101058625. This work benefited from services and resources provided by EGI with dedicated support from IFCA.

References

Ditria E.M., Lopez-Marcano S., Sievers M., Jinks E.L., Brown C.J., & Connolly R.M., (2020). *Automating the analysis of fish abundance using object detection: optimizing animal ecology with deep learning*. Frontiers in Marine Science, 429.

Jocher G., Chaurasia A., & Qiu J., (2023). *YOLO by Ultralytics (Version 8.0.0)* [Computer software]. https://github.com/ultralytics/ultralytics

AI-based fish detection and classification trials at OBSEA: available November 2023 https://www.youtube.com/watch?v=SEw79Gd6m5o