# DETECTING POTENTIAL MUSSEL REEF IN ROV VIDEOS USING DEEP LEARNING

*Xinyu Liu s212632, Weisen Yu s222463*

Faculty of Autonomous System, Technical University of Denmark, Denmark

## ABSTRACT

Blue mussel reefs are listed as "biogenic reefs" in the Habitat Directive of the European Union (EU). Studies have indicated that mussel reefs are beneficial to the marine ecosystem, but they are in danger. To protect existing mussel reefs, it is necessary first to detect mussel areas and then map them to confirm that they form a mussel reef. To assist mussel reef detection, we propose one method based on YOLO(You Only Look Once) algorithm to detect mussels in the videos, roughly estimate their coverage automatically, and indicate potential mussel reefs. Additionally, two datasets, including 1000 labelled mussel images, were constructed to bridge the gap in marine datasets.

***Index Terms***— Blue Mussel Reef, YOLO, Object Detection

## 1. INTRODUCTION

Blue mussel (henceforth called mussel) is a species of bivalve and has a two-part hinged shell with a soft-bodied inside. Mussel reefs termed as "biogenic reefs", may improve the water quality and protect shorelines. Additionally, they may also function as habitats for a range of invertebrates and fish [1, 5]. However, mussel reefs are among some of the most threatened marine habitats, largely due to climate change and prolonged human activities[6]. To protect mussel reefs in Roskilde Fjord, DTU Aqua uses ROV(remotely operated vehicle) to film seabed there and map mussel reefs based on ROV videos.

However, there are three problems in detecting mussel reefs using ROV. Firstly, manually marking mussels in a large volume of videos is time-consuming and labour-expensive. Secondly, Apart from detecting mussels, their coverage is needed to be evaluated, considering two criteria for identifying mussel reef: the seabed coverage by mussels is at least 30%, and the area of the seabed with mussels covers at least 2500 square meters. Lastly, except for ROV videos, exact GPS location, odometer data, and camera information are unavailable.

To solve these problems, we modified the YOLO algorithm to enable it to estimate mussels' coverage in one image in the detection process. One dataset, including 1000 labelled mussel images extracted from ROV videos, was constructed to train and test the neural network model.

## 2. MODEL

The main task in this project is object detection, therefore we chose YOLO algorithm, as its detection speed is fast and detection accuracy is acceptable.

The YOLO algorithm used in this project is the fifth version that is developed by ultralytics. It is based on the idea of using a single convolutional neural network (CNN) to predict bounding boxes and class probabilities from images. YOLOv5 is implemented in PyTorch. Fig.1 from YOLOv5 release note shows the performance comparison between YOLOv5 models and EfficientDet.

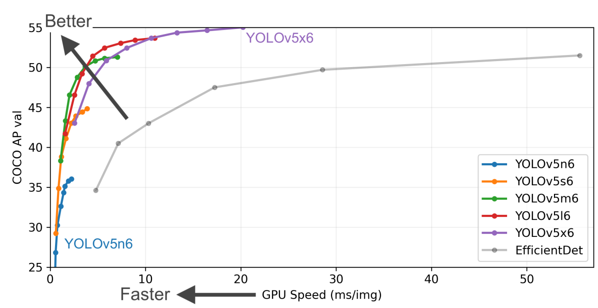



**Fig. 1**. Performance comparison Between YOLOv5 and EfficientDet

### 2.1. Introduction to the YOLOv5s Model

In this project, we chose YOLOv5s as our model because its performance is great and requires less memory, which is easy to deploy in ROV.

Fig.2 shows the network structure of YOLOv5s. According to the figure, the network mainly consists of four main sections, Input, Backbone, Neck and Head. All YOLOv5 models have the same structure (YOLOv5n, YOLOv5s, etc.) but differ in the width and depth of the convolutional layers. This can be seen in the $depth\_multiple$ and $width\_multiple$ parameters of the models' yaml files.

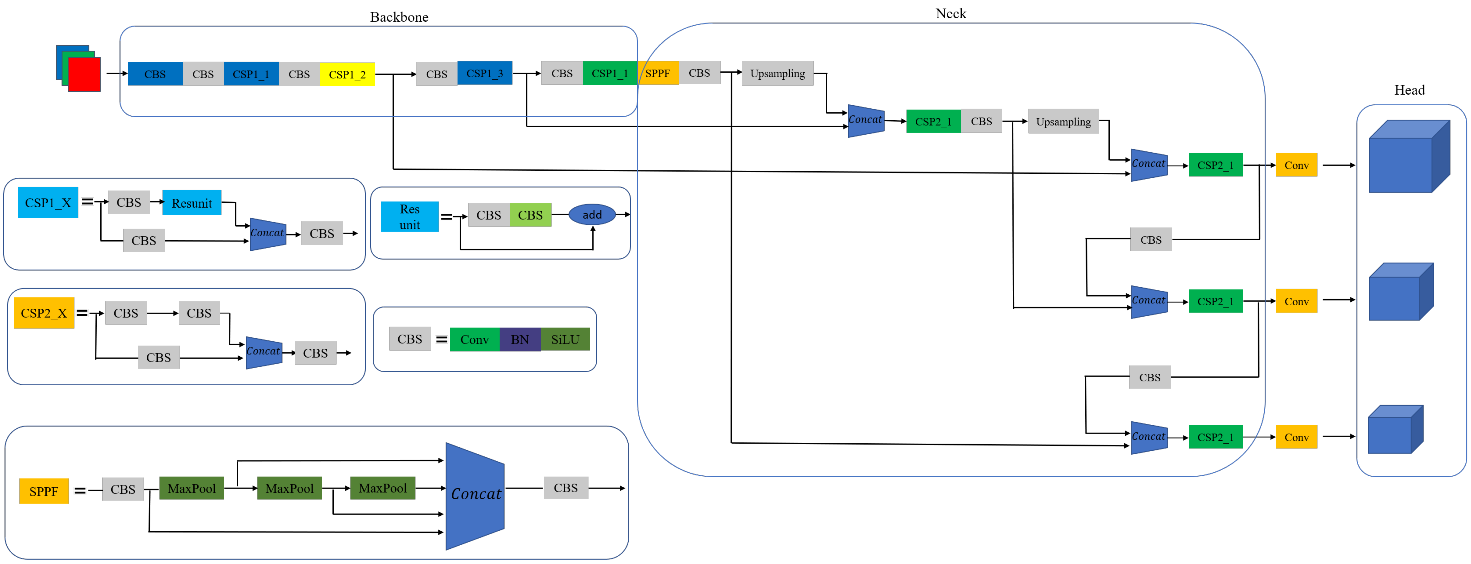



**Fig. 2**. Network Structure of YOLOv5s. Source:https://blog.csdn.net/zhangdaoliang1/article/details/122840458

In input section, a new method of data augmentation Mosaic was introduced. In previous versions such as YOLOv3, CutMix was used as data augmentation method, which only mixes 2 training images. Mosaic mixes 4 training images thus 4 different contexts are mixed. This allows detection of objects in different contexts.[6] The Fig.3 shows how Mosaic method is applied to our dataset.

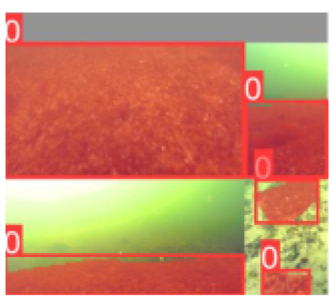


**Fig. 3**. Mosaic Method

In backbone part, it has almost the same composition as YOLOv4, the main component is CSPDarknet53 (CSP1_X and CSP2_X in Fig.2), which integrates CSPNet (Cross Stage Partial Network) into Darknet53 [7]. Darknet53 is a convolutional neural network that consists of 53 convolutional layers and CSPNet can improve its efficiency by reducing memory costs and removing computational bottlenecks[8].

In neck section, a new module SPPF (Spatial Pyramid Pooling - Fast) is included. Pooling aims to reduce the number of parameters while keeping main features. SPP generates output with a adaptive size to avoid resize and SPPF modifies SPP to increase the speed by reducing the computation volume.

## 3. METHODS

### 3.1. Data Collection

As shown in Fig.4, this summer, some students from DTU Aqua, with the assistance of local volunteer fishermen, used ROV to film the seabed in Roskilde Fjord and collected around 70 videos.


(a) Location of field site (b) ROV Used in Data Collection

**Fig. 4**. Location of Roskilde Fjord and ROV Used to Obtain Data

These videos are used to construct the mussel dataset, including 1000 images in 1280x640. We extracted the frames with mussels in the videos and label the mussels in the image using the tool: Labelimg. Four examples shown in Fig. 5.

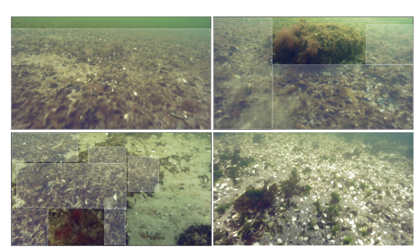


**Fig. 5**. Four examples in the dataset

### 3.2. Methodology

The problem of finding mussel reefs in one area can be transferred to finding mussel reefs in one video because each video was shot in one place and each period in the video corresponded to one specific area. If there is one potential mussel reef in one period in one video, it means there is one potential mussel reef in the corresponding area.

In one image, the mussels can be detected using YOLO algorithm and marked with rectangles. As all the mussels in the image are marked by rectangles, the rectangles' area can be simply seen as the area of mussels, and the area of the image is simply seen as the area of the seabed. Therefore, the coverage of mussels in one image can be roughly estimated using Equation 1. Because the boundary of mussels is irregular which is hard to fit by lines, we also tried the image segmentation method.

$$Mussel\ coverage = \frac{Mussel\ area}{seabed\ area} = \frac{Rectangle\ area}{Image\ area} \tag{1}$$

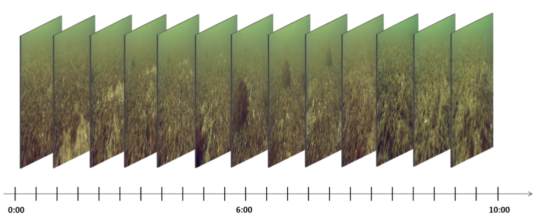



**Fig. 6**. Frames in one video

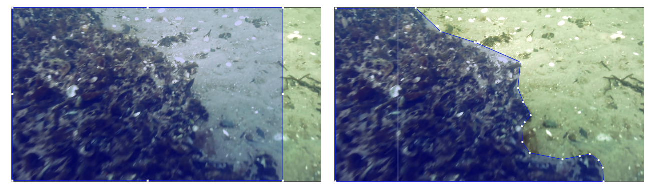
As we all know, one video contains many frames, and each frame is one image, as shown in Fig.6. Therefore, the mussels' coverage in all frames can be estimated using the equation. With the result, one graph, X-axis as time and Y-axis as coverage, will be generated to present the distribution of mussels in the video. After analyzing all the videos, the graphs will tell researchers which video likely includes one mussel reef, which means the corresponding area likely exists one mussel reef.

## 4. EXPERIMENTS

One experiment is carried out using the proposed method. Input is one 20-minute video and training parameters: image size: 1280x640, batch-size: 64, epochs: 200.

### 4.1. Constructing the datasets

Two datasets were constructed to accomplish object detection and segmentation. Images were extracted randomly from the videos, and as shown in Fig.7, for object detection we used a rectangle to label the mussels while for segmentation we used polygons to fit the edges of the mussels. Both datasets consist of around 1,000 images.



(a) Label using Rectangle (b) Label using Polygon

**Fig. 7**. Examples of two datasets

In the instance segmentation part, we also tried to use datasets labelling mussels on smaller scales such as labelling individual mussels and finding seabed areas covered with mussels without other species. But that requires excessive labelling time and the trained model performed badly, so we focused on labelling mussel clusters.

### 4.2. Model Training

The datasets are divided into two parts relatively. 800 images were used as training datasets while the other 200 were used for testing. The performance of the models is shown in Table.1:

| Type | Model | Precision | Recall | mAP 0.5 |
|---|---|---|---|---|
| Detection | YOLOv5s | 0.80 | 0.76 | 0.78 |
| Segmentation | YOLOv5s-seg | 0.86 | 0.80 | 0.73 |

**Table 1**. Performance of the two trained models

### 4.3. Model Testing

We chose one 20-minute video which filmed one potential mussel reef in Roskilde Fjord as input and used the trained model to detect the mussels in the video and calculate their coverage. Figure 8 shows some frames extracted from the processed video. The detecting speed is approximately 66 FPS.

## 5. RESULT ANALYSIS

After collecting the mussels' coverage in each frame in the video, one graph was generated as shown in Figure 9. Considering readability, the unit of time(X-axis) is adjusted to minute instead of second or frame and the Y-axis is the mean of all frames' coverage in one minute, as the length of the video is very long and the amount of frames is large.

According to Figure 9, the mean coverage keeps higher than 40% from 5 minutes to 17 minutes while that of the first 5 minutes and the last three minutes is lower than 30 %. Therefore it is likely that a mussel reef is located in the area where the ROV crosses between 5th and 17th minutes of the video.

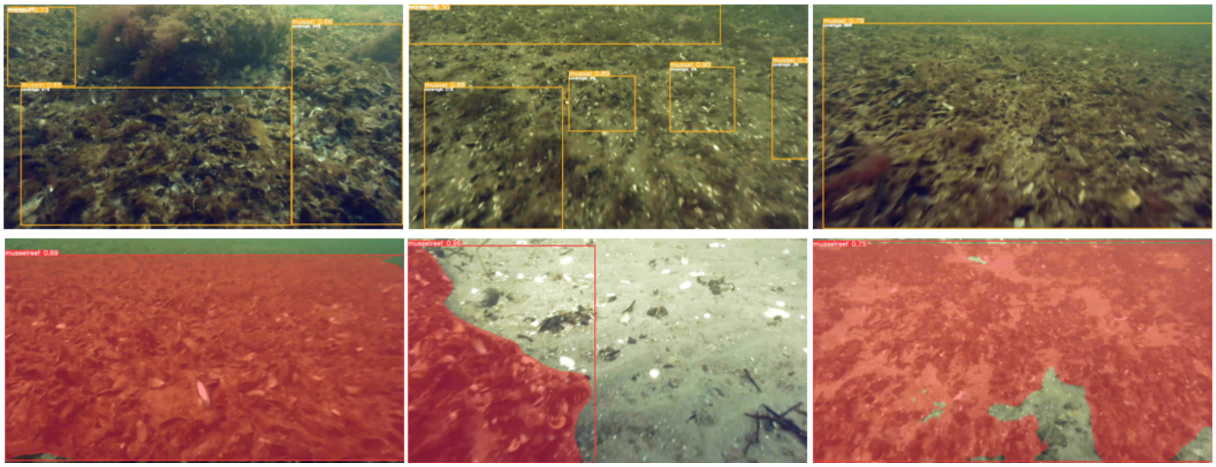


**Fig. 8**. Results of the Test

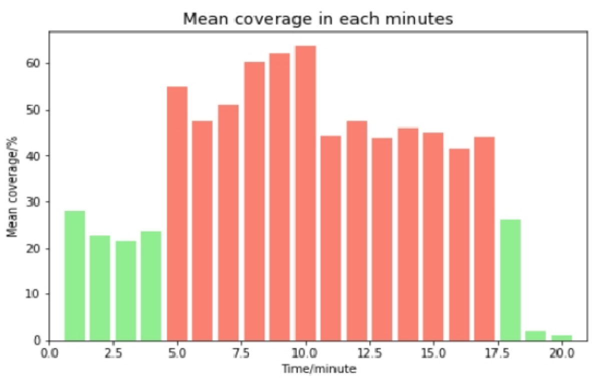



**Fig. 9**. Mean mussel coverage of one video

## 6. CONCLUSION

In this project, we constructed two mussel datasets to bridge the gap in the marine datasets, proposed one method based on YOLO algorithm to detect mussels and estimate their coverage, and trained the models to implement the function1.

Our method is easily deployed in ROV, requiring less memory and common hardware. This provides the possibility for local volunteers without background knowledge to collect underwater videos when they are boating on the sea. The videos and the analysis result will be transferred to DTU Aqua automatically. Based on the result, researchers can conduct related aquatic research on potential mussel reef areas directly instead of spending a long time on filming all the seabed in Roskilde Fjord and finding potential mussel reefs in numerous videos.

This method is low-cost as it just needs one ordinary camera. However, it also caused the coverage estimation to be imprecise, as it is easily affected by shooting angle, uncertain scale, and perspective projection.

Overall, the method can improve the efficiency of finding mussel reefs at a low cost although the coverage evaluation is imprecise which is enough for a rough estimation.

GitHub address: https://github.com/3505473356/Detecting-Potential-Mussel-Reef-in-ROV-Videos-Using-Deep-Learning. Poster and code is included.

## 7. REFERENCES


[1] Koivisto ME, Westerbom M (2010). Habitat structure and complexity as determinants of biodiversity in blue mussel beds on sublittoral rocky shores. Marine Biology 157: 1463-1474.

[2] Commito JA, Jones BR, Jones MA, Winders SE, Como S (2018). What happens after mussels die? Biogenic legacy effects on community structure and ecosystem processes. Journal of Experimental Marine Biology and Ecology 506: 30-41.

[3] Lefcheck JS, Williams BL, Beck MW, Orth RJ(2019). Are coastal habitats important nurseries? A meta-analysis. Conservation Letters 12: e12645.

[4] Schwartzbach A, Munk P, Sparholt H, Christoffersen M (2020). Marine mussel beds as attractive habitats for juvenile European eel (Anguilla anguill); A study of bottom habitat and cavity size preferences. Estuarine, Coastal and Shelf Science 246: 107042.

[5] Airoldi L, Balata D, Beck MW (2008). The gray zone: relationships between habitat loss and marine diversity and their applications in conservation. Journal of Experimental Marine Biology and Ecology 366: 8-15.

[6] Zeng, Guangmiao and Yu, Wanneng and Wang, Rongjie and Lin, Anhui(2021). Research on Mosaic Image Data Enhancement for Overlapping Ship Targets. Arxiv.

[7] Bochkovskiy, Alexey and Wang, Chien-Yao and Liao, Hong-Yuan Mark(2020).YOLOv4: Optimal Speed and Accuracy of Object Detection. Arxiv.

[8] Wang, Chien Yao and Liao, Hong Yuan Mark and Yeh, I. Hau and Wu, Yueh Hua and Chen, Ping Yang and Hsieh, Jun Wei (2019). CSPNET: A new backbone that can enhance learning capability of CNN. Arxiv.

[9] Salhaoui, M., Molina-Molina, J. C., Guerrero-González, A., Arioua, M., Ortiz, F. J. (2020). Autonomous underwater monitoring system for detecting life on the seabed by means of computer vision cloud services. Remote Sensing, 12(12), 1981.

[10] Mohamed, H. E. D., Fadl, A., Anas, O., Wageeh, Y., ElMasry, N., Nabil, A., Atia, A. (2020). Msr-yolo: Method to enhance fish detection and tracking in fish farms. Procedia Computer Science, 170, 539-546.