Exploring the potential to use low cost imaging and an open source convolutional neural network detector to support stock assessment of the king scallop (Pecten maximus)

Ovchinnikova, Katja; James, Mark A.; Mendo, Tania; Dawkins, Matthew; Crall, Jon; Boswarva, Karen

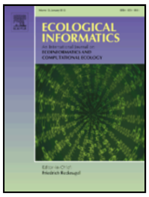
Published in:
Ecological Informatics

Publication date:
2021

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: http://ees.elsevier.com



Exploring the potential to use low cost imaging and an open source convolutional neural network detector to support stock assessment of the king scallop (*Pecten maximus*)

Katja Ovchinnikova[a], Mark A. James[b,*], Tania Mendo[b], Matthew Dawkins[c], Jon Crall[c], Karen Boswarva[d]

[a] European Molecular Biology Laboratory, Heidelberg, Meyerhofstraße 1, Heidelberg 69117, Germany
[b] Scottish Oceans Institute, University of St Andrews, East Sands, St Andrews, Fife KY16 8LB, UK
[c] Kitware Inc., 1712 Route 9, Suite 300, Clifton Park, New York 12065, USA
[d] Scottish Association for Marine Science, Dunstaffnage, Oban, Argyll PA37 1QA, UK

ARTICLE INFO

Keywords
Scallop
CNN
Pecten maximus
Assessment
VIAME
NetHarn

ABSTRACT

King Scallop (*Pecten maximus*) is the third most valuable species landed by UK fishing vessels. This research assesses the potential to use a Convolutional Neural Network (CNN) detector to identify *P. maximus* in images of the seabed, recorded using low cost camera technology. A ground truth annotated dataset of images of *P. maximus* captured in situ was collated. Automatic scallop detectors built into the Video and Image Analytics for Marine Environments (VIAME) toolkit were evaluated on the ground truth dataset. The best performing CNN (NetHarn_1_class) was then trained on the annotated training dataset (90% of the ground truth set) to produce a new detector specifically for *P. maximus*. The new detector was evaluated on a subset of 208 images (10% of the ground truth set) with the following results: Precision 0.97, Recall 0.95, F1 Score of 0.96, mAP 0.91, with a confidence threshold of 0.5. These results strongly suggest that application of machine learning and optimisation of the low cost imaging approach is merited with a view to expanding stock assessment and scientific survey methods using this non-destructive and more cost-effective approach.

1. Introduction

Around 80% of the global catch corresponds to commercially fished species of fish and shellfish that lack adequate data for stock assessments, which support sustainable fisheries management (Costello et al., 2012). This situation is particularly acute in small-scale inshore and artisanal fisheries that may be unregulated, unreported, or illegal. Stock assessments in developed countries can often be deficient and outdated because they can be relatively expensive and time consuming to conduct on a regular basis. The increasing use of novel, low cost systems and processes for collecting and processing data that could feed into stock assessments could significantly improve fisheries management practices.

The King Scallop (*Pecten maximus*) is the third most valuable species landed by UK vessels (after mackerel and *Nephrops*), worth £66.5 million in first sales value in 2016 and a five-year average of £59.5 million. Despite declining catches per unit effort, in recent years more licenses have been activated and more boats have entered these fisheries as the price of scallops has increased (Cappell et al., 2018)). The majority of scallops are harvested using dredges that impact the seabed resulting in the disturbance and destruction of seabed habitat, and fauna (Curry and Parry, 1999; Hinz et al., 2011; Hunt et al., 2007; Jenkins et al., 2001). As a result, the scallop dredge fishery is controversial and the subject of increasingly restrictive and intrusive regulation and monitoring.

A combination of methods are used to assess scallop stocks but the most common is an aged structured method, Virtual Population Analysis (VPA). This methods uses reported landings data along with age and length frequency data collected as part of market sampling programmes. The VPA provides annual estimates of yield, fishing mortality, spawning stock biomass and recruitment. Scallop dredge surveys complement the VPA as they provide information on the most recent changes in abundance, recruitment, age structure, growth rate, and other biological data (Mason et al., 1991).

A Time Series Analysis (TSA) approach is now favoured in some area as it is deemed to have a number of advantages over typical VPA approaches including: allowing fishing mortality estimates to evolve over time in a constrained manner; providing precision estimates of estimated parameters (numbers at age and fishing mortality at age); coping with the omission of catch or survey data if data are of poor quality

[*] Corresponding author.
E-mail address: maj8@st-andrews.ac.uk (M.A. James)

https://doi.org/10.1016/j.ecoinf.2021.101233
Received 6 August 2020; Received in revised form 24 December 2020; Accepted 24 December 2020
Available online xxx
1574-9541/© 2021.

or missing; allowing survey catchability to evolve over time. A full description of TSA use in scallop stock assessment can be found in Dobby et al., 2017.

There is currently no reliable low cost, non-destructive method in use by those charged with the management of *P. maximus* stocks to assess abundance in situ. Most stocks remain data deficient and a major barrier to addressing this challenge is the cost of undertaking stock assessments using conventional methods. The research reported in this paper is based on the opportunistic use of a large image set of *P. maximus* generated for another project using low cost camera technology that has been used here to assess the utility of a suitably tuned CNN to automatically identify *P. maximus*. In addition, we explore the performance of the tuned CNN for detecting *P. maximus* in images acquired by divers and a Remotely Operated Vehicle in two sample locations with different depth and benthic characteristics.

P. maximus also known as the “great scallop” or “king scallop” is a marine bivalve mollusc of the family Pectinidae (See Fig. 1.). They are widely distributed in the eastern Atlantic along the European coast from northern Norway, south to the Iberian peninsula, and have also been reported off West Africa (Brand, 2006). They can be found in offshore waters down to 100 m on predominantly sandy, fine gravel or sandy gravel sediments (Mason 1983). Great scallops generally recess in the substrate to accommodate the shallow hemispherically domed right valve (shell). The left valve is flat and usually level with or just below the surface of the substrate (Baird, 1958). As a result, sand, mud, gravel or living organisms coat the upper valve, making them difficult to detect by predators (or divers).

In waters around the United Kingdom *P. maximus* becomes sexually mature at around 2–3 years old and when they reach 80 to 90 mm in shell length. In Scottish waters, a minimum landing size of 110 mm is in place except for Shetland (100 mm) and the Irish Sea south of 55°N (105 mm) to prevent the harvesting of juvenile stocks. Where they are not exploited, they may live for more than 20 years and reach shell lengths of more than 200 mm (Beukers-Stewart and Beukers-Stewart, 2009).

Scallop stock assessment data (abundance, size and age) is usually collected through a combination of fishery independent dredge surveys and fishery dependent surveys of landed catch. Attempts to use in-situ underwater surveys using still or video imagery captured by diver, Remotely Operated Vehicle's (ROV) and benthic sledges have been undertaken but these require manual analysis of the images which is both time consuming and expensive (Richards et al., 2019).

In the temperate waters and at the depths at which *P. maximus* occurs, water clarity is often limited by ambient light levels and suspended particulate material. The propensity of this species to partially recess in seabed sediment can also impede visual identification. *P. maximus* also tends be more widely dispersed on the seabed than other scallop species such as the Atlantic Sea Scallop (*Placopecten magellanicus)*. These factors together with image quality may further limit the ability to reliably identify this species of scallop from images alone.

Machine learning is increasingly being applied to automate challenging image analyses in the form of deep learning, a class of which are Convolutional Neural Networks (CNN) most commonly applied to analyzing visual imagery (Zhang et al., 2018). A CNN is trained on a ground truth dataset, i.e. sets of images with features of interest annotated by humans. In the context of object detection, the ground truth images contain rectangles drawn around objects of interest. A trained CNN is called a model. With respect to the object detection task, a

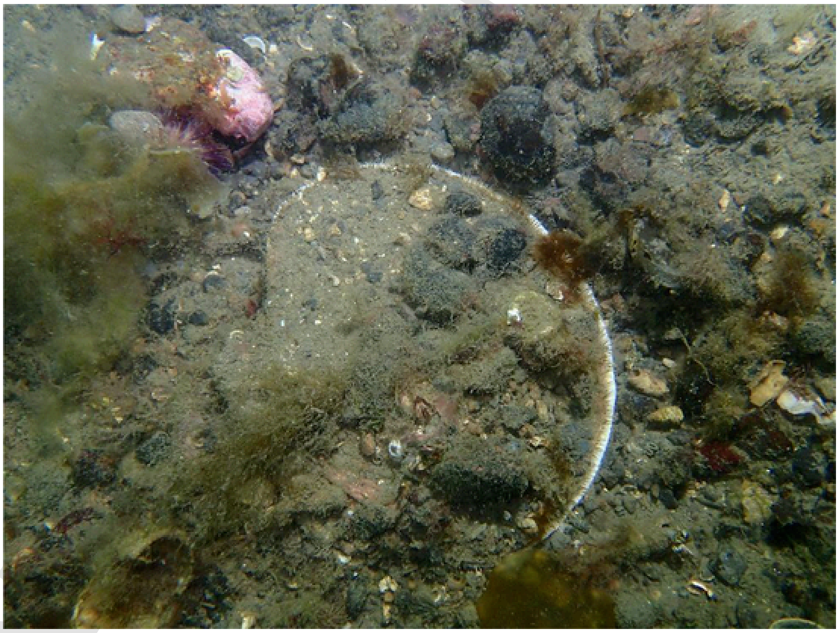


Fig. 1. Fi[illegible] 1 Image of King scallop (*Pecten maximum*) in typical benthic habitat. The curvature of the edge of the shell (white) is clearly visible, but most of the shell is covered with benthic material.

Table [illegible]
Table [illegible]luation of the existing models in VIAME.

Model	Precision	Recall	F1 score	mAP	Confidence threshold	Average difference in number of predicted vs annotated scallops per image
YOLO	0.89	0.54	0.68	0.52	0.00	0.49
NetHarn_1_class	0.82	0.74	0.77	0.75	0.34	0.40
NetHarn_4_classes	0.77	0.51	0.61	0.5	0.00	0.7

model is also called a detector. A detector can be then applied to automatically identify and quantify the objects of interest in new images. An existing detector trained on a large ground truth dataset can be further fine-tuned on a smaller dataset. Fine-tuning is a process that takes a model that has already been trained for one type of object (e.g. *P. magellanicus* scallops) and then tunes or adjusts the model to make it detect a similar but different type of object (e.g. *P. maximus* scallops). The underlying assumption is that the new small dataset is not significantly different from the original dataset and the pre-trained model has already learned features that are relevant for the new detection challenge.

Here we report on a method that could be used to help improve assessments of the abundance *P. maximus*, using readily available low cost camera technology and a trained CNN to automatically analyse images (and video) to provide in situ counts of scallops, thus providing measures of abundance without the need for potentially destructive sampling using scallop dredges. The ability to automatically identify scallops in situ could also offer future potential to develop less harmful research trawl survey methods and commercial harvesting practices by reducing the need for speculative dredging by fishers in search of new fishing grounds and facilitate the development of technologies using robotics for example to select individual specimens thus limiting disturbance to the benthos.

The objectives of this study were: step 1) to collate a ground truth annotated dataset of images of *P. maximus* captured in situ; step 2) evaluate automatic scallop detection algorithms built into the Video and Image Analytics for Marine Environments (VIAME) toolkit (Dawkins et al., 2017; Hoogs et al., 2020); step 3) to train the best performing CNN (selected in step 2) on the annotated dataset and thus obtain a new detector specifically for *P. maximus*.

The annotated data set was created from multiple images recorded from transects seeded with a known quantity of scallops. In addition, the annotated data set represents two different scallop habitats with images recorded separately by diver and ROV. The annotated data set was therefore also used to assess whether useful comparisons could be made with respect to the data collection method, location, depth, and benthic habitat. This information has been used to provide guidance on optimising the acquisition of images for use in the automated analysis of *P. maximus* densities on the seabed.

2. Material and methods

The images used in this research were a by-product of a PhD project designed to assess the potential to develop structure from motion (SfM) photogrammetry (Micheletti et al., 2015) of scallop stocks. Thirty-three live *P. maximus* of various sizes (110–128 mm shell length, 97–125 mm shell height) were randomly distributed by hand by a SCUBA diver along a 25 m transect. Once distributed, a second SCUBA diver surveyed the 25 m transect using a boustrophodonic survey pattern (Burns et al., 2015) across a total survey area of 50m^2. Two hours later, an ROV was used to survey the transect and finally a second SCUBA diver transect survey was conducted approximately 5 h after the scallops had first been distributed on the sea bed. The ROV survey design was conducted to mimic the survey pattern conducted by the SCUBA diver in order to maximise the SfM model comparisons.

Repeat surveys were conducted at depths of 18–21 m at Ganavan Bay (56° 26′20″ N, 5° 28′ 28″ W) on 31st October and 7th November 2018 and 6–8 m at Dunstaffnage Bay (56° 27′ 04″ N, 5° 26′ 06″ W) on 29th November 2018 (Fig. 2.), with horizontal visibility ranging from 2 to 5 m. The sea state was 3 or less on the Beaufort scale in both locations (UK Meteorological Office, 2021). The SCUBA diver surveys were conducted on an ebbing tide in both locations.

Images were obtained by SCUBA diver and a ROV, both equipped with a GoPro Hero 6 camera operated approximately 1 m above and at an angle of incidence of approximately 90 degrees to the seabed. For the SCUBA survey, the camera was operated in time-lapse (0.5 s) mode, capturing high definition (12 MP) images, together with two Weefine smart focus video lights (3000 lm each), illuminating the seafloor. The ROV survey was conducted with a battery powered BlueROV2 equipped with 6 T200 thrusters and maximum forward speed of 1 m/s (2 knots), capturing live 1080p HD video with a Hitech Hs-5055MG tilt servo capable of 110 degree field of view and +/−90 degree camera tilt, illuminated by 4 × 1500 lm lights with dimming control and 135 degree light beam angle. The onboard ROV camera was used for real time navigation only. Survey images were captured with a Go-pro Hero 6 camera attached to the ROV capturing high quality 4 k video, utilising the ROV lighting. Lighting was set to either 75% or 100% depending on seabed conditions. The 4 k resolution video was then segmented into high quality portable network graphics (PNG) image files.

Operating the GoPro cameras at a standard (4 × 3) format with no zoom delivers a field of view of 94.4 (vertical – y axis) and 122.6 (horizontal – x axis). Under water at a height of ~1 m, this equates to ~1.38 m in the Y axis and ~2.52 m in the X axis giving an area of view ~3.49 m^2. An individual scallop would occupy ~1% of the area of view.

The high-quality PNG video stills were converted to Joint Photographic Experts Group (JPEG) format. All images were analysed on VIAME open access software. A total of 3070 images were randomly selected from an image set of 32,645 images for this experiment and annotated by two annotators, who did not have any knowledge about the study and the purpose of the annotation and worked independently from each other. Two sets of independent annotations were required to calculate the agreement between the annotators, assess the quality of their work and the difficulty of the task. If the agreement between annotators is low, it indicates the task is too difficult or the quality of the annotations is low. If the agreement is high, then the ground truth dataset can be composed of the annotations on which both annotators agree and it will have a higher quality than if produced by one annotator.

The annotators selected images containing scallops using the “Annotation” and “Create Single Frame Tracks” tools in the VIAME “Tools” menu (VisGUI View version 2.0.0). A bounding box was drawn around each scallop in the image. Each bounding box was represented by X and Y pixel coordinates of its upper left and bottom right corners. The coordinate data from all bounding boxes were collated in the software and exported as a comma separated value (CSV) file for analysis.

A random selection of 50 annotated images were reviewed by a computer vision expert (author KO) in order to provide estimates of obvious errors in annotations made by the annotators.

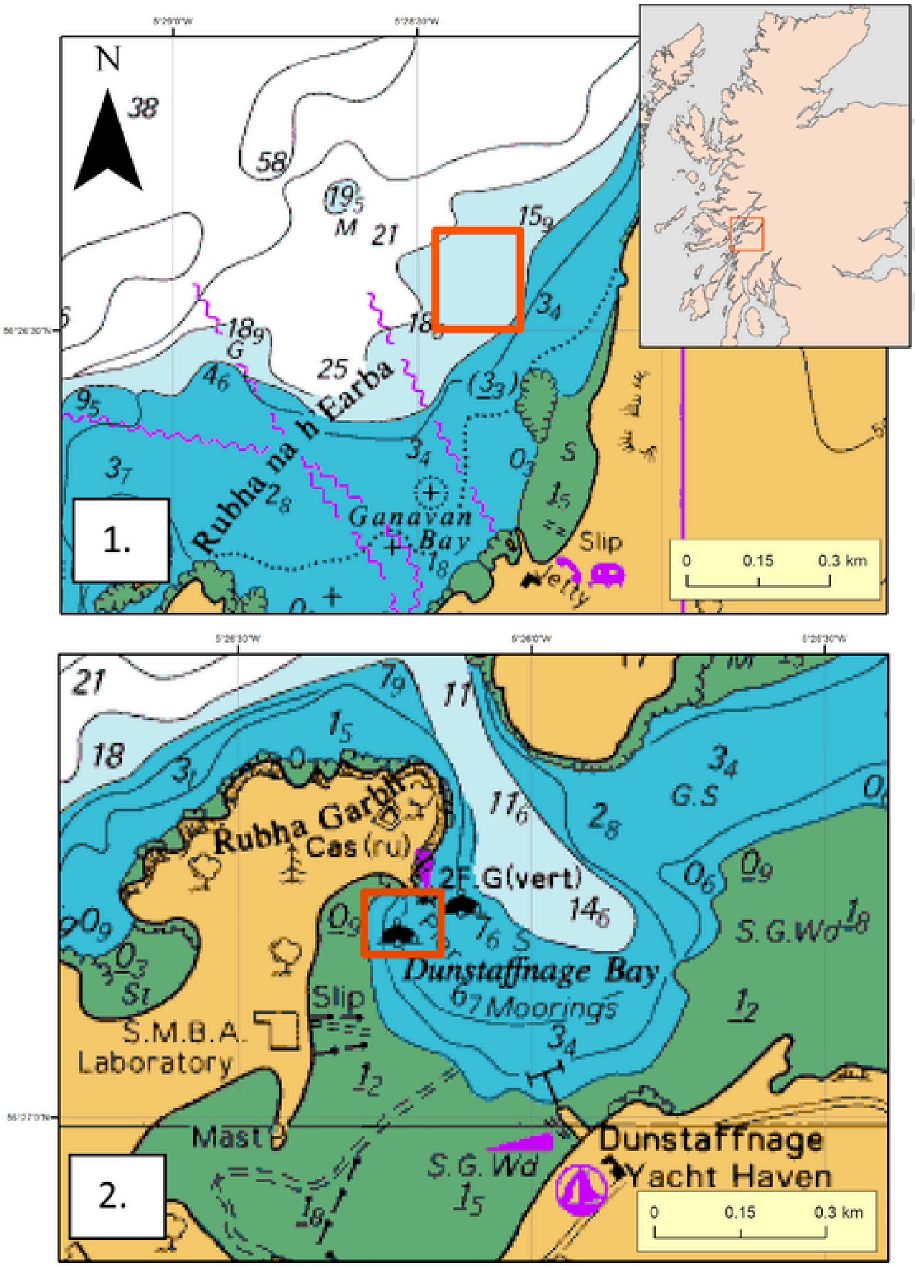
The annotations were used to create a ground truth dataset that was then used to evaluate existing scallop detectors and train new detectors.

For the purposes of comparison, the annotated image files were subdivided on the basis of location (two bays, reflecting differences in depth and habitat) and collection method (SCUBA diver or ROV).

3. Theory/calculation

3.1. Scallop detection algorithms in VIAME

VIAME scallop detectors are based on Convolutional Neural Networks. The existing VIAME models have been trained on the imagery collected by Coonamessett Farm Foundation (CFF) in 2017, 2018, and 2019, as well as the HabCam 2015 dataset provided by Northeast Fisheries Science Center (NEFSC). In total, there are about 150,600 annotated standard living Atlantic Sea scallops (*Placopecten magellanicus*) in these images, 6342 swimming scallops, 978 dead scallops, and 1137 clappers (dead scallops where the valves are still attached to each other). The dataset also contains other annotated species, such as flatfish, lobster, monkfish, squid, sea star, skate, etc. In the experiment de-



Underlying chart © Crown, Copyright/HR Wallingford Ltd. 2017. All Rights Reserved. Licence No., L012017.0001. Not to be Used for Navigation.

Fig. 2. [illegible]e 2 Maps showing survey sites 1. Ganavan and 2. Dunstaffnage Bay.

scribed here two VIAME models, YOLO and NetHarn trained for *P. magellanicus* were compared using *P. maximus* as the detection target.

3.1.1. YOLO model

The YOLO version 2 network (https://pjreddie.com/darknet/yolov2/; (Redmon and Farhadi, 2016) is a CNN that divides an image into regions and predicts bounding boxes and probabilities for each region. Unlike other models that are applied to an image at multiple locations and scales, YOLO looks at the whole image so its predictions are informed by global context in the image. YOLO architecture makes use of only convolutional layers, making it a fully convolutional network. It has 75 convolutional layers, with skip connections and upsampling layers.

YOLO uses convolutional weights that are pre-trained on ImageNet (http://www.image-net.org/). In the current study, the model was then trained on the scallop images *P. magellanicus*.

3.1.2. NetHarn model

This model was trained using a Cascade Faster-RCNN (Cai and Vasconcelos, 2018; Chen et al., 2019)with a parameterized fit harnesses NetHarn for augmentation (https://gitlab.kitware.com/computer-vision/netharn).

In object detection, the intersection over union (IoU) threshold is frequently used to define positives/negatives. The threshold used to train a detector defines its quality. While a low threshold leads to noisy (low-quality) detections, for higher thresholds detection performance usually degrades. The main reason for it is overfitting, due to vanishing positive samples with smaller IoU values. Cascade R-CNN is designed to address this problem. Cascade R-CNN consists of a sequence of detectors trained with increasing IoU thresholds, to be sequentially more selective against close false positives. The detectors are trained stage by stage, so that the output of a detector is a good input for training the next higher quality detector.

In computer vision, the process of augmentation is often applied to enhance incomplete training datasets. Through various strategies such as cropping, rotating, and flipping images, an existing dataset can be expanded in order to train an AI model on more examples. The NetHarn model employs ~20 types of augmentation, including additive gaussian noise, median blur, coarse dropout etc. (https://github.com/aleju/imgaug).

For this study, two versions of this model were explored: trained for one class (*P. magellanicus* - scallop) and for four classes (*P. magellanicus* - clapper, dead scallop, standard alive scallop, swimming scallop, flatfish - which included images of rays and sole). The four class model was first trained to detect all possible species in the dataset (including sea star, squid, lobster, etc.) and then fine-tuned for the four classes of interest. Given the four classes model, we considered detections from all classes except flatfish to be relevant detections.

3.2. Confidence threshold selection

For each image, a given detector outputs a set of bounding boxes (coordinates of the top left and bottom right corners) with their confidence scores ranging from 0 to 1. In order to select a threshold for confidence scores for each detector, the ground truth dataset was randomly split into 10 equal parts. For each part, the confidence thresholds ranging from 0 to 1 were tested with a step of 0.01 and the threshold giving the highest F1 score value was selected. The final confidence threshold was averaged over 10 runs.

3.3. Detector training

The best network in VIAME was selected based on the performance of the corresponding model and trained with *P. maximus* images (See Table 1).

Images were randomly split in the annotated dataset into "train" (90%) and "test" (10%). The train dataset was used to a) train a new model initiated with weights pre-trained on the Common Objects in Context dataset (https://cocodataset.org), b) fine-tune an existing best performing VIAME model. The performance of the initial, trained, and fine-tuned models was evaluated on the test dataset consisting of 208 images. Out of these images, there were 138 (66%) ROV and 70 (34%) SCUBA images and 108 (52%) of images from Ganavan and 100 (48%) from the Dunstaffnage Bay.

3.4. Evaluation

For evaluation of the results, precision, recall, F1 score, and mean average precision measures were used (Everingham et al., 2010).

Precision (*P*) is defined as a percentage of the predicted scallops that are correct, i.e. correspond to the ground truth. Recall (*R*) is defined as a percentage of the ground truth scallops that are predicted correctly (see Fig. 4).

The F1 score is the harmonic mean of the precision and recall defined as follows:

$$F1 = 2\cdot \frac{P\cdot R}{P + R}$$

The highest possible value of an F1 score is 1, indicating perfect precision and recall, and the lowest possible value is 0, if either the precision or the recall is zero.

If the predictor produces a confidence score, then a threshold on this score can be used to filter out unreliable predictions.

Correct prediction is defined using the Intersection over Union (IoU) score, see Fig. 4. Given two bounding boxes, i.e. a predicted box (*PB)* and a ground truth box (*GB)*, IoU is defined as follows:

$$IoU = \frac{Overlap\ area\ of\ PB\ and\ GB}{Union\ area\ of\ PB\ and\ GB}$$

A prediction is defined as correct if its IoU with any of the ground truth boxes is ≥ 0.5.

Following the standards for PASCAL VOC (2021; http://host.robots.ox.ac.uk/pascal/VOC/) object detection challenges (Everingham et al., 2010), the mean average precision (mAP) was calculated, which computes the average precision value for recall levels ranging from 0 to 1 with step 0.1. Formally, mAP is defined as follows.

$$mAP = \frac{1}{11}\sum_{r\in\{0,0.1,\ldots,0.9,1\}} P_{interp}(r)$$

where $P_{interp}(r)$ is interpolated precision for recall level *r* computed as follows.

$$P_{interp}(r) = \max_{r':r'\geq r} P\left(r'\right)$$

where $P(r')$ is measured precision at recall *r*'.

4. Results

4.1. Ground truth annotations

For the selected 3070 files, annotator 1 and annotator 2, annotated 2098 and 3048 files, respectively. The number of matched bounding boxes between them with an IoU score of ≥ 0.5 was 1747. Annotator 1 and annotator 2 provided 227 and 1417 unique unmatched bounding boxes, respectively. This indicates a low agreement between annotators.

A total of 50 annotated images were selected randomly such that each image contained at least one unmatched bounding box. These images were inspected manually by a computer vision expert (author KO). Annotator 1 made 75 obvious errors in annotating the images (missing a scallop or drawing a bounding box where there was no scallop), whereas annotator 2 made only 2 errors. Annotator 2 therefore provided annotations of much higher quality and on this basis. Because of the difference in the quality of the annotations, it was not possible to compose the ground truth dataset using both annotations. Therefore annotations for the 3048 files provided by annotator 2 were used as the ground truth dataset.

4.2. Scallop detection

NetHarn_1_class outperforms other models in terms of F1-score, recall, mAP, and the average difference of number of predicted vs annotated scallops in an image. The thresholds were selected automatically as described above. For models NetHard_4_classes and YOLO, they proved to be 0, which shows that the confidence measures were not helping to remove wrong detections.

4.3. Comparison of survey methods

The ground truth data was divided based on 1) survey method (SCUBA diver or ROV) and, 2) location, which also equates to depth, Dunstaffnage (shallow 6-8 m) and Ganavan (deep 18-21 m). The best model (NetHarn_1_class) was evaluated separately for the two survey image capture methods as well as for the two locations. The model consistently performs better for the images collected by SCUBA diver (Table 2). There is no difference in performance for the SCUBA diver

Table 2

Table 2 Evaluation of the VIAME NetHarn_1_class model for two combined survey and image acquisition methods and two locations: Number of images (#), Precision (P), Recall (R), F1 score, mAP.

	ROV					SCUBA					Total				
	#	P	R	F1	mAP	#	P	R	F1	mAP	#	P	R	F1	mAP
Ganavan	1226	0.78	0.52	0.62	0.56	432	0.89	0.9	0.89	0.89	1658	0.84	0.6	0.7	0.66
Dunstaffnage	761	0.79	0.82	0.80	0.72	633	0.9	0.93	0.91	0.89	1394	0.83	0.86	0.84	0.84
Total	1994	0.76	0.4	0.69	0.64	1065	0.9	0.91	0.91	0.89	3048	0.82	0.74	0.77	0.75

acquired images from Dunstaffnage and Ganavan. For ROV acquired images, recall is lower for Ganavan.

4.4. Trained detectors

The trained and fine-tuned models for *P. maximus* significantly outperform the initial NetHarn_1_class model trained for *P. magellanicus* (Table 3). This result illustrates the importance, in this case, of training a model with the target species and associated habitat.

4.5. Workflow

The workflow from image annotation through to model evaluation and fine tuning is illustrated in Figs. 3&5.

4.6. Error analysis

To explain the difference in performance of the original VIAME NetHarn_1_class model for the two image capture methods and two locations (Table 2), 50 images were randomly selected from each of the four groups (200 images in total) and manually inspected independently by two of the authors (KO and MJ). The purpose was to identify image features resulting in false positives (automatic detections made but no scallop annotated), false negatives (annotated scallops that were not detected automatically). The authors (KO and MJ) also identified several scallops missed by the annotator. False negatives mostly occur in less well resolved and darker images (26 out of 29 mistakes in the inspected images). False positive detections result from sand or stone formations, algae, sea urchins and non-pectinid shells on the sea bed in the darker images (20 out of 23 mistakes). Three scallops were found by the automatic detector, but missed by the annotator. Most of the images collected in the ROV survey are dark and less sharp than SCUBA diver acquired images due to the amount of fine particulate material suspended in the water. The SCUBA diver survey images were generally sharper and brighter. Manual inspection of the images did not inform our understanding of the reasons for higher recall for the ROV survey images from Dunstaffnage, but it is likely that as this location was shallower, ambient light levels are likely to be higher. In addition, the finer sediments found at Ganavan may be more easily suspended by near seabed currents and backwash from the ROV thrusters.

To analyse the errors of the best trained model (NetHarn_1_class_trained, Table 3), the detections for 208 images from the test set were manually reviewed. In total, there were 9 undetected scallops (false negatives). Most of them (7 out of 9) occurred in the dark and less well resolved ROV survey images. Two scallops undetected in SCUBA diver surveyed images were located at the border of the image. See examples in Fig. 4.

The trained detector produced 8 false positives. Four false positives result from sand or stone formations in the dark and less well resolved ROV survey images. In SCUBA diver survey images, three out of four false positives result from missing annotations and are thus no mistakes, while one false positives is part of a degraded laminaria frond and stipe. See examples in Figs. 6-8. Figs. 6-8

For three images in the test set, scallops were missed by both the annotator and the automatic detector. See an example in Fig. 4(left). Overall, the trained detector performed well even for dark and poorly resolved ROV images, see an example in Fig. 4(right).

5. Discussion

5.1. Comparison of models

The images used in these analyses vary considerably with respect to their quality and content. The seabed is heterogeneous, containing features ranging from muddy sand to coarse sand and gravel, rocks, shell material, seaweeds and metallic wreckage covered in fouling organisms. Water clarity and light levels were also highly variable with the deeper survey site tending to be both darker and less clear as a function of light attenuation with water depth and finer sediment characteristics increasing levels of suspended material. The scallops manually deposited along the survey transect will not have had time to recess for the first SCUBA diver survey and will therefore have been sitting proud of the surface of the seabed with little or no sediment covering them. It is possible that a small proportion of the scallops will have started to recess within two hours (subject to substrate type for example) and these images would have been captured using the ROV. For the second SCUBA diver survey at each location, approximately 5 h after the initial scallop placement a greater proportion of the scallops would have recessed and been partially covered with sediment, more closely mimicking their natural state. The image classification algorithms have been exposed to a highly variable set of images in terms of content and quality in which *P. maximus* may have appeared as clearly defined targets in terms of shape and colour or in some cases, poorly resolved partial outlines with no colour discernibly different from the surrounding seabed sediment. However, it is estimated that only 20–25% of all specimens were fully recessed (Boswarva pers com) which could increase the performance of the detector and therefore its performance in habitats where the majority of scallops are fully recessed needs to be assessed.

Table 3

Table 3 Evaluation of the VIAME original NetHarn_1_class model, trained model, and fine-tuned model on the test dataset (208 images, 10% of the full dataset).

Model	Precision	Recall	F1 score	mAP	Confidence threshold	Average difference in number of predicted vs annotated scallops in image
NetHarn_1_class initial	0.9	0.71	0.79	0.69	0.36	0.51
NetHarn_1_class_trained	0.97	0.95	0.96	0.91	0.5	0.1
NetHarn_1_class fine-tuned	0.97	0.94	0.96	0.91	0.6	0.13

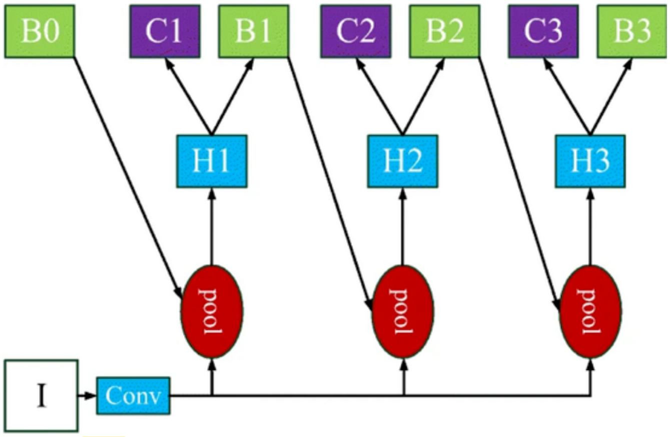


Fig. 3. Fig[illegible] Architecture of Cascade Faster-RCNN as presented in (Cai and Vasconcelos, 2018). “I” is input image, “B0” is region of interest proposals, “conv” is backbone convolutions, “pool” is region-wise feature extraction, “H” is network head, “B” is bounding box, and “C” is classification.

Of the three VIAME models initially tested, the NetHarn_1_class models proved to be the best, followed by YOLO. All the original VIAME models were trained on the images of the *P. magellanicus* scallops, which makes their application to the *P. maximus* species sub optimal. However, since NetHarn includes many augmentation techniques to vary the training set, it can cover a larger variety of images. This is a possible explanation for the NetHarn_1_class model significantly outperforming YOLO (0.75 vs 0.52 mAP).

The NetHarn_4_classes model is an experimental model initially trained to detect all possible species annotated in the dataset and then fine-tuned for scallop and flatfish. This experiment has been performed to get better base features for later fine-tuning across categories of interest. In our experiment, this model performed significantly worse than NetHard_1_class (0.75 vs 0.5 mAP). A possible explanation is that these additional features introduced greater variation, reducing performance.

Our model trained specifically for *P. maximus* using the NetHarn_1_class network achieves the best results (0.96 F1 score and 0.91 mAP) and significantly outperforms the initial NetHarn model trained for *P. magellanicus* (0.91 vs 0.69 mAP). Interestingly, fine-tuning does not give any advantage, which suggests that even a small training set may be sufficient for training a high-quality detector.

5.2. Comparison of survey methods - diver vs ROV

The images used in this research were collected as part of another project designed to acquire images and video from both SCUBA diver and an ROV to develop SfM seabed photogrammetry.

Whilst the project generated a large number of images (over 32,000) suitable for assessing the performance of CNN detectors, the experimental design was not optimal for determining differences between survey and image acquisition methods. The image acquisition method used the same camera but with different light sources and image capture formats. Diver survey images were captured as high resolution JPEG stills (~4000 × 3000 pixels) whilst ROV survey images were extracted from high resolution 4 K video as PNG images (~3840 × 2160 pixels). Both PNG and JPEG are raster graphics formats but JPEG uses compression algorithms to reduce image file size. PNG uses lossless data compression. JPEG images tend to provide smooth transitions of colours whereas the PNG format is good for images with sudden changes in colour and contrast. Whilst images were annotated at these resolutions, the VIAME software automatically resamples all images to 512 × 512 pixels and JPEG format prior to model analysis, thus it seems unlikely that image resolution or format in this case will have impacted upon model performance. Manual inspection of image quality suggests that the main differences between those captured from the ROV and the SCUBA diver are the brightness and resolution of detail of the images as a function of water clarity and illumination.

The SCUBA diver was clearly able to maintain a more constant (~1 m) distance from the seabed whilst keeping the camera angle relatively perpendicular to the seabed. The ROV could not be controlled manually to maintain the same level of consistency in distance from the seabed and the angle of incidence of the camera used to capture images was also much more variable than that achieved by the SCUBA diver (Boswarva, pers com). The lower angle of incidence between the ROV camera and the seabed increases the area and distance of view, reducing the overall light intensity whilst increasing the potential amount of suspended material between the camera lens and the subject.

The SCUBA diver survey involved hand holding a relatively small GoPro camera with associated lights and conducting a boustrophodonic survey pattern with frames being captured at a rate of 2 per second. The automated rate of image capture and the speed of the diver was designed to allow ~70% overlap in images. This process will tend to reduce the potential for human bias in image capture which could favour the occurrence of the target species in these images when compared to those captured with the ROV. The SCUBA diver will also tend to avoid creating turbulence that would disturb the sediment and potentially obscure the camera view. Although the ROV survey was designed to mimic the boustrophodonic survey pattern of the diver, manual comparison of the SCUBA diver vs the ROV captured images suggests that the ROV images in some cases may have been of lower quality as a result of sediment disturbance caused by backwash from the ROV propulsion system.

The difference in the performance of the trained NetHarn_class1 model on images collected by the ROV survey at Dunstaffnage Bay

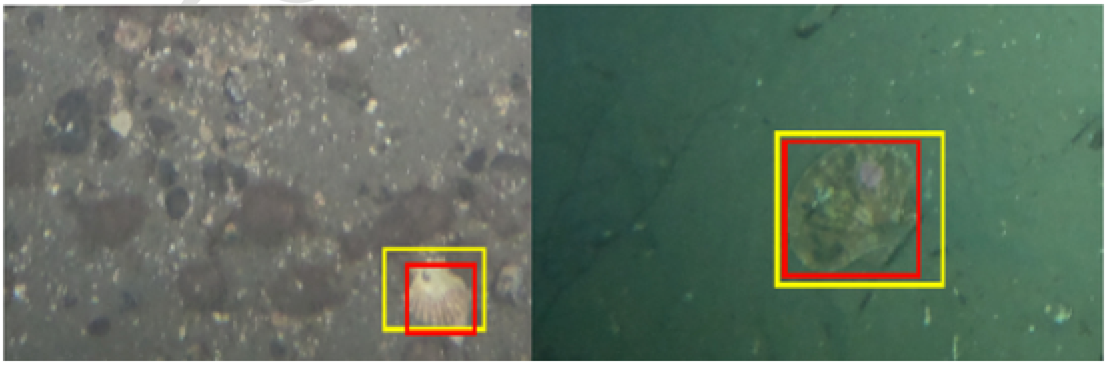


Fig. 4 [illegible]re 4 In these examples, the ground truth box is yellow and the predicted box is red. For the left image, IoU = 0.54. For the right image, IoU = 0.72. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

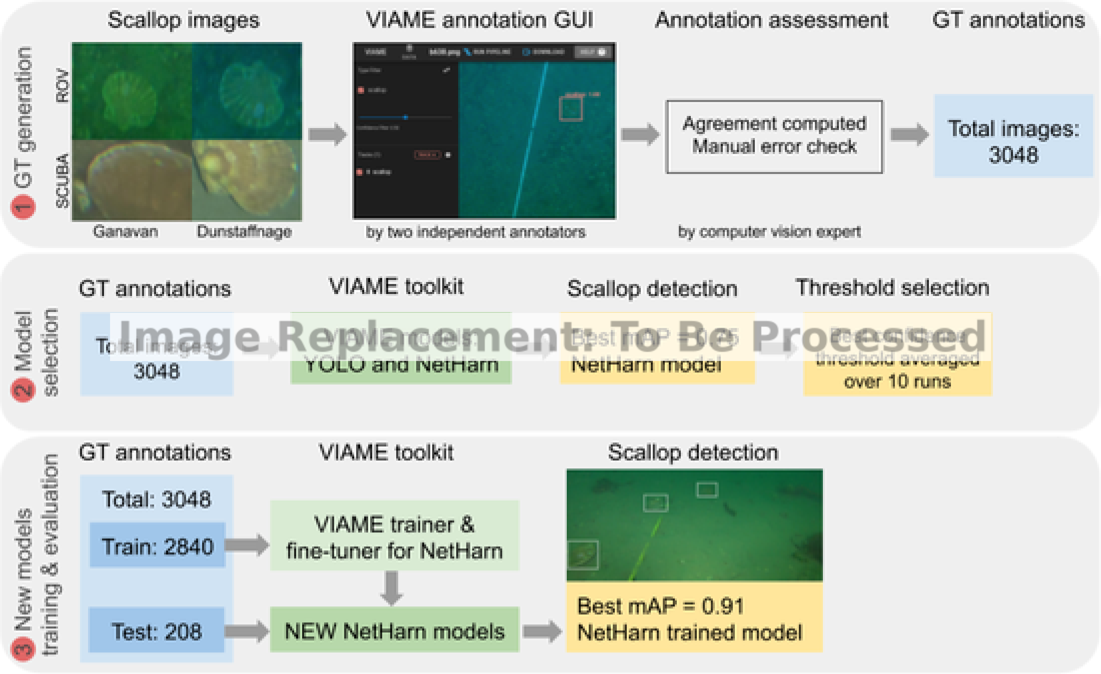


Fig. 5. re 5 Our workflow consists of three main steps: 1) ground truth (GT) generation, 2) model selection, and 3) training and evaluation of new models. For GT generation, scallop images were annotated by independent annotators using the VIAME annotation GUI. Their annotations were assessed by a computer vision expert, which resulted in selecting annotations for the final GT dataset. This dataset was then used to evaluate existing VIAME models YOLO and Netharn trained for *P. magellanicus*. The Netharn model proved to yield a better result. The best confidence threshold for this model was selected. In the last step, the GT dataset was divided into the train and test datasets. The train dataset was used to train and fine-tune a new NetHarm model for *P. maximus*. The new model yielded the best result of 0.91 mAP.

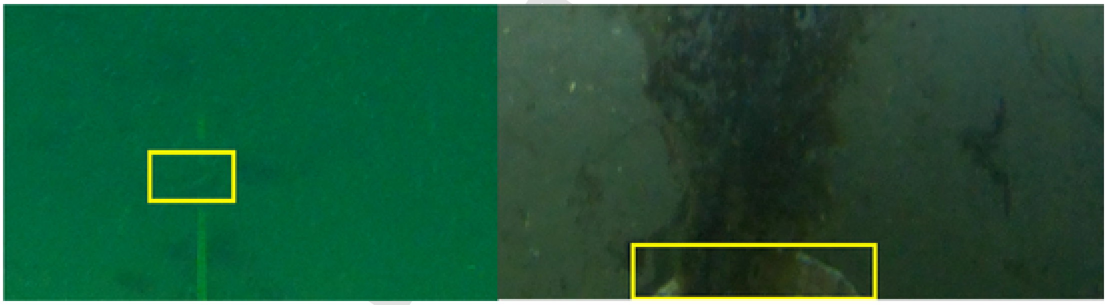


Fig. 6. re 6 Examples of false negatives (yellow bounding box) in ROV (left) and SCUBA (right) survey images. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

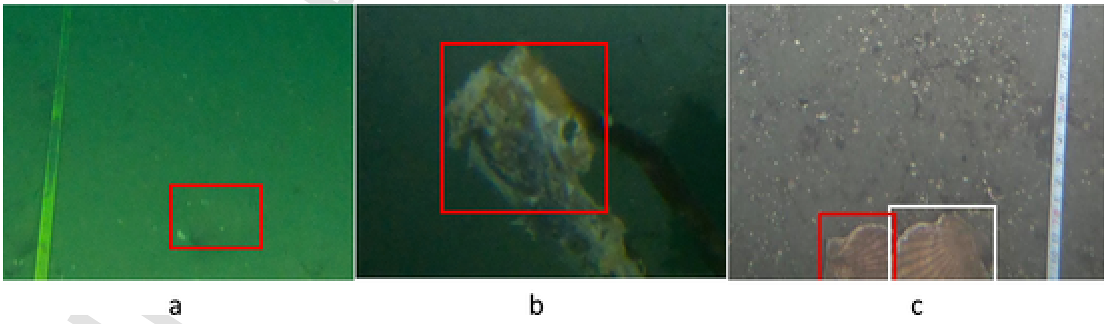


Fig. 7. re 7 Examples of false positives: Poorly resolved ROV survey image (a), SCUBA diver survey image with a laminaria frond and stipe (b), SCUBA diver survey image with two correct detections and an annotation missing for the left scallop (c). Red box = false positive. White box = true positive. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(0.72 mPA) and Ganavan (0.56mPA) appear to reflect the limitations of this survey method. Ganavan Bay (deep 18-21 m) is more than twice the depth of Dunstaffnage Bay (shallow 6-8 m) which means increased attenuation of light with depth and therefore significantly less ambient light. The finer sediments and exposed nature of the Ganavan Bay site are more prone to disturbance and therefore near-seabed particulate material is evidently greater. Combined with the potential for images taken at a lower angle of incidence to the seabed these factors would combine to amplify not only the lower performance of the ROV

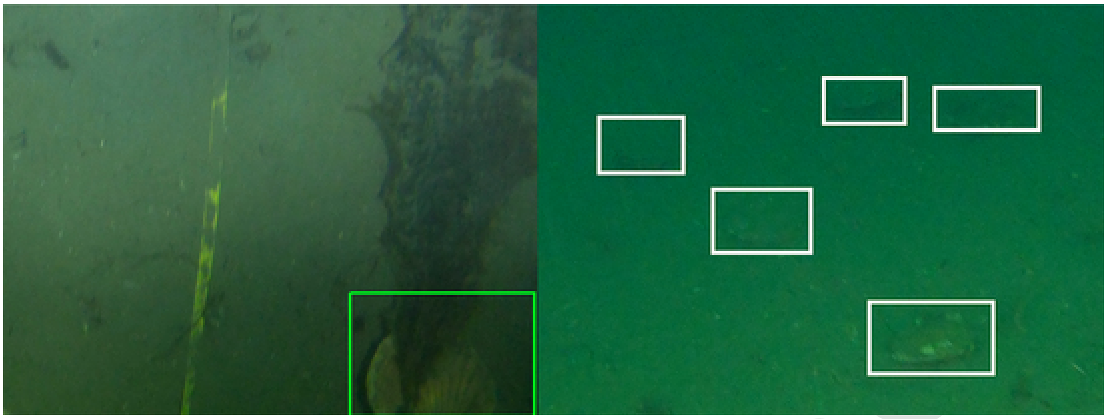


Fig. 8 Figure 8 Example of missing detection and annotation (green bounding box) (left) and correct detections (white bounding boxes) in a dark and poorly resolved ROV survey image (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

recorded images between the two survey sites but also the differences between the survey methods.

With respect to image acquisition, the results of this research suggest that relatively low cost camera technology such as a GoPro and associated lighting, recording still images with sufficient frequency to provide a statistically meaningful coverage of seabed area could be used to assess *P. maximus* densities in situ, subject to any specified depth/pressure limitation on the technology. At a height of ~1 m above the seabed with the camera oriented perpendicular to the seabed, this technology can provide images of sufficient quality to permit a detector such as the NetHarn_1_class model to perform well. Either as part of a simple drop down camera setup, a towed sledge or, for example, a passive wing designed to "fly" a required distance above the seabed.

In some situations recessed *P. maximus* are more visible when viewed at an oblique angle because the gaping valves of the shell reveal the mantle tissues and eye spots that can reflect light and stand out from the background.

A 90 degree or perpendicular angle of incidence does not describe the nature of the image captured. The camera used has a lens with a field of view of 130 degrees in wide angle mode. Whilst the light parallax caused by being underwater will reduce the field of view by approximately a third, the angle of incidence between the lens and the subject will generally be less than 90 degrees. In practical terms this means that scallops nearer the outer range of the field of view will not be seen directly from above but at a slight angle which may be sufficient to pick up the edge of the gapping scallop shell. An important factor in trying to define practical optima for camera angle and height above the sea bed is the degree to which particulate material suspended in the water obscures the image. The lower the angle of incidence for any given height above the seabed, the greater the distance between the lens and the subject and thus the greater potential for suspended sediment to obscure the image of the subject. To compensate for this, the camera could be used closer to the seabed, therefore reducing the lens to subject distance but this would also reduce the area of the seabed being sampled in each image. The height of approximately 1 m above the seabed for image acquisition was chosen based on experience of local conditions at the sample sites. In waters with greater clarity it would be possible to increase the lens to subject distance and it would certainly be worthwhile to undertake controlled tank experiments to try to optimise lens to subject distance and angle of view. The fact that images of recessed *P. maximus* captured at the Dunstaffnage sample site, which is prone to poor visibility, could still be reliably identified by the detector suggest that the method, if validated on a larger dataset of fully recessed specimens, could be robust.

5.3. Model performance

The fact that the trained NetHarn_1_class model yielded a combined survey and image acquisition detection performance of 0.91 mPA which included 66% of the lower quality images derived from the ROV surveys is very encouraging. The heterogeneous and somewhat challenging nature of the images from different locations, depths and seabed habitats also suggests that this model is worthy of further investigation. However, the results obtained with the trained model can potentially be over-optimistic as the annotated dataset is small, the training and testing images are relatively similar. Although no images within the data set repeat, many images show the same scallops in the same location with the same lighting conditions. The present study should therefore be considered as a pilot and the NetHarn_1_calss model needs to be retrained on a larger dataset. The results of this project provide a strong basis on which to recommend the collection of a larger annotated data set of *P. maximus* in a fully recessed state, which would not contain images of the same individuals and use the image acquisition method outlined.

5.4. Comparison with related work

Various automated detection methods for identifying scallops in their natural environment have been explored (Dawkins, 2011; Dawkins et al., 2013; Enomoto et al., 2009, 2010; Fearn et al., 2007; Kannappan and Tanner, 2013; Rasmussen et al., 2017). However, a fair comparison with the related work is difficult, because every study uses different datasets. The best Atlantic scallop detector of the Rasmussen et al., 2017, study employing YOLOv2 achieves 0.85 mAP, while our best detector is yielding 0.91 mAP. Dawkins et al., 2013, report Atlantic scallop detection precisions ranging from 0.28 to 0.99 and recalls ranging from 0.69 to 0.94 with their best detector, depending on which dataset it was tested on. Their best results are similar to our results obtained with NetHarn_1_class_trained (precision 0.97 and recall 0.95). Other related work mentioned above does not follow standard computer vision metrics for object detection making comparison more difficult. Kannappan and Tanner, 2013, report only precision of 0.9 for Atlantic scallop detection. Enomoto et al., 2010, do not report precision, but accuracy of 0.86 for *Mizuhopecten yessoensis* scallop detection. Fearn et al., 2007, do not develop a ground truth dataset and do not provide a quantitative evaluation of their approach to Tasmanian scallop detection.

6. Conclusions

The NetHarn_1_class model trained with the *P. maximus* annotated dataset can be reliably used to estimate the number of scallops in the images. Furthermore, using a small annotated dataset for training can be sufficient to obtain a high quality detector for *P. maximus* and presumably other similar species. Using the model to explore differences between the survey and associated image acquisition methods can inform improvements in survey methodology and, in this case, point towards the potential to use relatively inexpensive technology to achieve the required image quality for reliable detection. To improve and validate the performance of the model for operational use in estimating scallop densities, a larger set of annotated images of *P. maximus* fully recessed is now required.

Uncited reference

Declaration of Competing Interest

None.

Acknowledgements

This annotated set of images was collected by Dr. Karen Boswarva under a NERC/NEXUSS funded PhD entitled “The effectiveness of Autonomous Underwater Vehicles (AUVs) for mapping and monitoring the seabed – a novel toolkit for marine management”, supervised by Dr. John Howe (SAMS), Dr. Clive Fox (SAMS), Professor Bhavani Narayanaswamy (SAMS) and Dr. Keith Brown (Heriot Watt). Tritonia Scientific provided diving and ROV support under a NFSD/NERC grant (NE/N012070/1). The image annotation was undertaken by two University of the Highlands and Islands students; Miss Ilaria Stollberg and Mr. Jonathan Taylor. The annotation was funded under the Scottish Inshore Fisheries Integrated Data System Project (SIFIDS) funded by EMFF (grant number: EMFF: SCO1434). The assistance of Hannah Ladd-Jones the SIFDS Project Manager in facilitating this work is gratefully acknowledged.

References

Baird, R.H., 1958. On the swimming behaviour of escallops (*Pecten maximus* L.). Proc. Malac. Soc. Lond. 33, 67–71.

Beukers-Stewart, B.D., Beukers-Stewart, J.S., 2009. Principles for the Management of Inshore Scallop Fisheries around the UK (PDF) (Report). University of York.

Brand, A.R., 2006. Scallop ecology: distributions and behaviour. Dev. Aquac. Fish. Sci. 35, 651–744. doi:10.1016/S0167-9309(06)80039-6.

Burns, J.H.R., Delparte, D., Gates, R.D., Takabayashi, M., 2015. Integrating structure-from-motion photogrammetry with geospatial software as a novel technique for quantifying 3D ecological characteristics of coral reefs. Peer J. 3, e1077. doi:10.7717/peerj.1077.

Z. Cai N. Vasconcelos Cascade R-CNN: Delving into High Quality Object Detection(arXiv: 1712.00726)https://arxiv.org/abs/1712.007262018

Cappell, R., Huntington, T., Nimmo, F., MacNab, S., 2018. UK scallop fishery: current trends, future management options and recommendations. In: Report Produced by Poseidon Aquatic Resource Management Ltd. Version: Final Report Report Ref: 1417-GBR Date Issued: 11 October 2018 Photo credit: Alamy.

Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., Zhang, Z., Cheng, D., Zhu, C., Cheng, T., Zhao, Q., Li, B., Lu, X., Zhu, R., Wu, Y., Dai, J., Wang, J., Shi, J., Ouyang, W., Loy, C. C., Lin, D. (2019) Mmdetection: Open Mmlab Detection Toolbox and Benchmark. Retrieved 28th July 2020. (arXiv preprint arXiv: 1906.07155). https://arxiv.org/abs/1906.07155

Costello, C., Ovanda, D., Hilborn, R., Gaines, S., Deschenes, O., Lester, S. E., et al., 2012. Status and Solutions for the World's Unassessed Fisheries. Science 338 (6106), 517–512. doi:10.1126/science.1223389.

Curry, D.R., Parry, G.D., 1999. Impacts and efficiency of scallop dredging on different soft substrates. Can. J. Fish. Aquat. Sci. 56, 539–550.

Dawkins, M., 2011. Scallop Detection in Multiple Maritime Environments Master's Thesis Rensselaer Polytechnic Institute.

Dawkins, M., Stewart, C., Gallager, S., York, A., 2013. Automatic scallop detection in benthic environments. In: 2013 IEEE Workshop on Applications of Computer Vision (WACV). IEEE, pp. 160–167.

Dawkins, M., Sherrill, L., Fieldhouse, K., Hoogs, A., Richards, B., Zhang, D., Prasad, L., Williams, K., Lauffenburger, N., Wang, G., 2017, March. An open-source platform for underwater image and video analytics. In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, pp. 898–906.

Dobby, H., Fryer, R., Gibson, T., Kinnear, S., Turriff, J., Mclay, A., 2017. Scottish scallop stocks: results of 2016 stock assessments. Scottish Mar. Freshwat. Sci. 8 (21), 1–178.

Enomoto, K., Toda, M., Kuwahara, Y., 2009. Scallop detection from sand seabed images for fishery investigation. In: 2nd International Congress on Image and Signal Processing. IEEE, pp. 1–5.

Enomoto, K., Masashi, T., Kuwahara, Y., 2010. Extraction method of scallop area in gravel seabed images for fishery investigation. IEICE Trans. Inf. Syst. 93 (7), 1754–1760.

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The Pascal visual object classes (VOC) challenge. IJCV 88 (2), 303–338.

Fearn, R., Williams, R., Cameron-Jones, M., Harrington, J., Semmens, J., 2007. Automated intelligent abundance analysis of scallop survey video footage. Adv. Artif. Intell. 549–558.

Hinz, H., Tarrant, D., Ridgeway, A., Kaiser, M., Hiddink, J., 2011. Effects of scallop dredging on temperate reef fauna. Mar. Ecol. Prog. Ser. 432, 91–102.

Hoogs, A., Dawkins, M.D., Richards, B., Cutter, G., Hart, D., Clarke, M.E., Michaels, W., Crall, J., Sherrill, L., Siekierski, N., Woehlke, M., 2020, February. An open-source system for do-it-yourself AI in the marine environment. In: Ocean Sciences Meeting 2020. AGU.

Hunt, H., LeBlanc, S., Benoit, H., 2007. Impact of scallop dredging on benthic habitat and associated fauna. J. Shellfish Res. 26 (4), 1317.

Jenkins, S.R., Beukers-Stewart, B.D., Brand, A.R., 2001. Impact of scallop dredging on benthic megafauna: a comparison of damage levels in captured and non-captured organisms. Mar. Ecol. Prog. Ser. 215, 297–301.

Kannappan, P., Tanner, H.G., 2013. Automated Detection of Scallops in Their Natural Environment 1350-1355 SN 978-1-4799-0995-7 DO doi:10.1109/MED.2013.6608895.

Mason, J., Cook, R.M., Bailey, N., Fraser, D.I., Shumway, S.E., Sandifer, P.A. (Eds.), 1991. An assessment of scallops, *Pecten maximus* (Linnaeus, 1758), in Scotland West of Kintyre. In: An International Compendium of Scallop Biology and Culture, 1. World Aquaculture Society; World Aquaculture Workshops, pp. 231–241 (ISBN 0962452955).

Micheletti, N., Chandler, J., Lane, S.N., 2015. Structure from Motion (SFM) Photogrammetry. Loughborough University (Journal Contribution) https://hdl.handle.net/2134/17493.

PASCAL VOC (Retrieved 7th July 2020)http://host.robots.ox.ac.uk/pascal/VOC/2021

Rasmussen, C., Zhao, J., Ferraro, D., Trembanis, A., 2017. Deep census: AUV-based scallop population monitoring. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 2865–2873.

J. Redmon A. Farhadi YOLO9000: Better, Faster, Stronger http://pjreddie.com/yolo9000/ 2016 (Retrieved 7th July 2020) https://arxiv.org/pdf/1612.08242.pdf

Richards, B.L., Beijbom, O., Campbell, M.D., Clarke, M.E., Cutter, G., Dawkins, M., Edington, D., Hart, D.R., Hill, M.C., Hoogs, A., Kriegman, D., Moreland, E.E., Oliver, T.A., Michaels, W.L., Placentino, M., Rollo, A.K., Thompson, C.H., Wallace, F., Williams, I.D., Williams, K., 2019. Automated Analysis of Underwater Imagery: Accomplishments, Products, and Vision. doi:10.25923/0CWF-4714.

UK Meteorological Office 2021 "National Meteorological Library and Archive Fact sheet 6—The Beaufort Scale" (PDF). Met Office. Archived from the original (PDF) on 2 October 2012. (Retrieved 7th July 2020). https://web.archive.org/web/20121002134429/http:/www.metoffice.gov.uk/media/pdf/4/4/Fact_Sheet_No._6_-_Beaufort_Scale.pdf

Zhang, J., Shao, K., Luo, X., 2018. Small sample image recognition using improved convolutional neural network. J. Vis. Commun. Image Represent. 55 (2018), 640–647.