2020 Global Conference on Wireless and Optical Technologies (GCWOT)

Automatic Detection of *Nephrops* norvegicus Burrows in Underwater Images Using Deep Learning

Atif Naseer
ETSI Telecomunicación
Universidad de Málaga
Malaga, Spain
atif@uma.es

Enrique Nava Baro
ETSI Telecomunicación
Universidad de Málaga
Malaga, Spain
en@uma.es

Sultan Daud Khan
Department of Computer Science
National University of Technology
Islamabad, Pakistan
sultandaud@nutech.edu.pk

Yolanda Vila Gordillo
Instituto Español de Oceanografía
Centro Oceanográfico de Cádiz
Cádiz, Spain
yolanda.vila@ieo.es

2020 Global Conference on Wireless and Optical Technologies (GCWOT) | 978-1-6654-1860-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/GCWOT49901.2020.9391590

***Abstract*—Autonomous Underwater Vehicles and Remotely Operated Vehicles equipped with HD cameras are used by the scientist to capture the underwater footages efficiently and accurately. The abundance of the Norway Lobster *Nephrops* norvegicus stock in the Gulf of Cadiz is assessed based on the identification and counting of the burrows where they live, using underwater videos. The Instituto Español de Oceanografía (IEO) conducts an annual standard underwater television survey (UWTV) to generate burrow density estimates of *Nephrops* within a defined area, with a coefficient of variation (CV) or relative standard error of less than 20%. Currently, the identification and counting of the *Nephrops* burrows are carried out manually by the experts. This is quite hectic and time consuming job. Computer Vision and Deep learning plays a vital role now a days in detection and classification of objects.**

The proposed system introduces a deep learning based automated way to identify and classify the *Nephrops* burrows. The proposed work is using current state of the art Faster RCNN models Inception v2 and MobileNet v2 for objects detection and classification. Tensorflow is used to evaluate the Inception and MobileNet performance with different numbers of training images. The average mean precision of Inception is more than 75% as compared to MobileNet which is 64%. The results show the comparison of Inception and MobileNet detections, as well as the calculation of True Positive and False Positive detections along with undetected burrows.

***Keywords*—Faster RCNN, Computer Vision, *Nephrops* norvegicus, *Nephrops* norvegicus stock assessment, Underwater Videos Classification.**

I. Introduction

Research in underwater image analysis is vast and has a worth owing to its enormous applications in different disciplines, for instance, underwater biodiversity and bottom morphology monitoring, pipeline and cables maintenance, minerals mining, and for military application in the sea [1]. In underwater object like fauna and flora recognition, detection of fishes and coral reef classification problems are the most popular tasks of underwater image analysis . However, the environment features such as depth-based color variations and the turbidity or movement of species make it a challenge [2]. Thus, two main factors which make it difficult are the unrestricted natural environment and variations of the visual content which may arise from variable illumination, scales, views and non-rigid deformations [3].

The Norway lobster, *Nephrops* norvegicus is a one of the main commercial crustaceans exploited Europe [4]. Annual catches of *Nephrops* in the Gulf of Cadiz are small compared with other Atlantic *Nephrops* stock (≈100 t in 2018) but gives valuable revenues for the trawl fleet [5]. Fig. 1 shows an individual of *Nephrops*.This species occurs in sandy-muddy bottoms from 200 m to 700 m of depth [6], where sediment is suitable for them to construct their burrows.

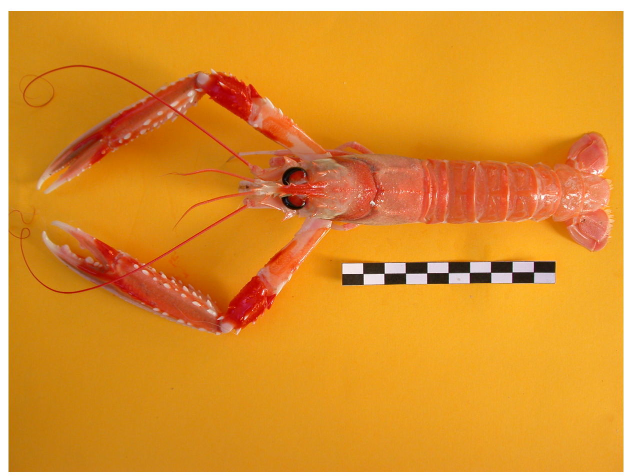


Fig. 1. *Nephrops* norvegicus

Nephrops burrows typically have multiples openings to different tunnels communicated with each other which are named as *Nephrops* burrow systems. A unique individual is assumed occupy a burrow system [7]. Burrows show different features that are specific to *Nephrops* as shown in Fig. 2 and they can be summarized in:

1) At least one burrow opening is particularly half-moon shape.
2) There is often proof of expelled sediment, typically in a wide delta-like ’fan’ at the tunnel opening, and scratches and tracks are frequently clear.
3) The center of all the opening burrows has an apparent raised structure.
4) *Nephrops* may be present (either in or out of burrow).

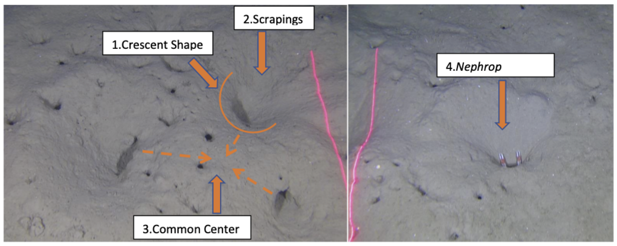


Fig. 2. *Nephrops* burrows features

Underwater Television (UWTV) surveys to monitor the abundance of *Nephrops* populations were pioneered in Scotland in early 90’s. The estimation of Norway lobster abundances using UWTV systems involves identification and quantification of burrow density over the known area of *Nephrops* distribution that can be used as an abundance index of the stock [5]. Since then, the number of stocks with routine *Nephrops* UWTV surveys has increased over time and in 2017 around more than 18 *Nephrops* grounds are expected to have surveys.

The Instituto Español de Oceanografía (IEO) carries out UWTV surveys yearly since 2014. These surveys are focused on suitable habitats for *Nephrops* in the Gulf of Cadiz (SW Spain). A sledge equipped with high quality cameras is trawled on the bottom to obtain video footages of 10 minutes. Currently, *Nephrops* burrows systems are counted manually from underwater videos. Each footages is manually reviewed by the experts individually in multiple parallel sessions and concludes the results using a Lin’s CCC higher than 0.5 or consensus [5]. This exercise is a time consuming and cost lots of financial and human resources. There is not currently system available that can help them in solving the problem of manual detection and counting of *Nephrops* burrows systems.

Durden et al. [8] proved that even highly trained for benthic ecology observers do not achieve 100% correct classification. The manual annotations recorded by the observers vary as it depends on certain human factors like tiredness, stress mood changes etc. These biasness results in wrong classification and detections which is currently one of the main bottlenecks in marine ecological sampling [9]. Underwater object detection and classification is still very new for the research community. Underwater environmental challenges make the things different from other object detection and classification. The analysis of the underwater images is given various names, for instance, underwater object recognition [10], seafloor image recognition [11] and underwater image processing or visual content recognition [12].

The advancement in Artificial Intelligence (AI) and Computer Vision (CV) helps in improving the classification and detection of marine species. To detect and classify the *Nephrops* burrow systems, CV and AI play a vital role as it improves the classification and detection drastically. In this study, we investigate these issues by using state of the art Faster RCNN deep learning algorithms to identify the *Nephrops* burrows in seafloor footages of the Gulf of Cadiz dataset. The objective of this work is to use a deep learning model to automatically detect, classify and count the *Nephrops* burrows. The proposed work is using current state of the art Faster RCNN models Inception and MobileNet models for objects detection and classification.

The rest of the paper is sectioned as follows. Data description is discussed in Section 2. The Research methodology is explained thoroughly in Section 3. The details of the experiments and results are discussed in Section 4. Finally, the paper is concluded in Section 6.

II. Data Description

A. Underwater Environment Problems

The underwater data has many challenges compared to other data. As light and water are not considered to be a good friend because when light passes through the water, it cannot absorb and reached to the sea surface which makes the images or videos a blurring effect. The image devices are usually designed with low resolution because of the huge amount of data collection. Also, there is scattering and non-uniform lighting which make the environment more challenges for data collection because the ultimate goal is to reach as close to seabed to collect the data. The poor visibility is a common problem in underwater environment. The ocean current is another factor which cause frequent luminosity change.

B. Study Area and Data Collection

Study area is located in the Gulf of Cadiz on the Southwest of Spain. Videos used in this work were obtained during ISUNEPCA_0618 UWTV survey conducted between 2nd and 14th June of 2018. The design of the survey followed a randomized isometric grid at 4 nm spacing. A total of 70 stations were planned ranged from 90 to 650 m depth and covered an area of 3000 Km. Fig. 3 shows the map of Gulf of Cadiz with station carried out in 2018 in ISUNEPCA_0618 and the *Nephrops* burrow density obtained using the manual count.

Videos were recorded using a 4K Ultra High Definition (UHD) camera (SONY Handycam FDRAX33) having Lens of ZEISS® Vario-Sonnar with 29.8 mm and optical zoom 10x. Camera is mounted on top of the sledge with an angle of 45º in relation to the seafloor. To record the video with good lighting condition four spotlights with independent intensity control is used. The equipment has also two line laser separated 75 cm that is used to confirm the field of view (FOV) and a Li-ion battery of 3.7 V & 2400 mAh(480 Watt) to support the power of the whole of system. HiPAP transponder have been used to know the sledge position during the transect for each station.

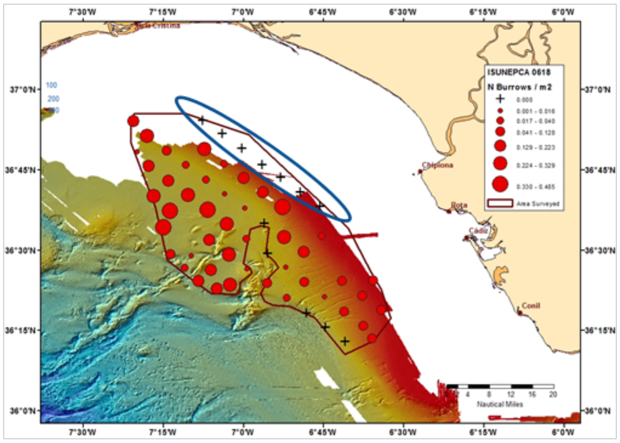


Fig. 3. *Nephrops* burrows density at the Gulf of Cadiz in 2018 survey

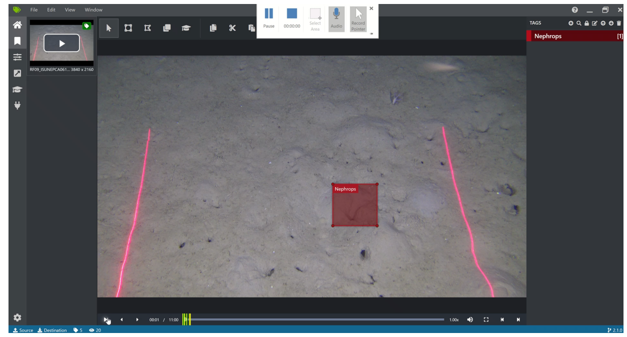


Fig. 4. Mannual Annotation in a frame of the footages from ISUNEPCA_0618 UWTV survey

C. Data Characteristics

Each image has resolution of 3840 x 2160 pixels. Most of the frames are of very low quality and poor visibility, those frames are discarded. The frames are in jpeg format.

III. Research Methodology

A. Data Preparation

1) Data Collection: The videos are recorded at 25 frames per seconds in good lighting condition. Videos footages of 10-12 minutes each one has been recorded by stations. Out of these, only seven stations were selected which have optimal lightening condition, less noise, better contrast and high density of *Nephrops* recorded. The data from seven stations are considered for annotations. Each video is 25 frames per seconds and 10-12 minutes in length. So, each video is around 15,000 - 18,000 frames. A total of 105,000 frames were recorded from seven different stations of 2018 data survey.

2) Data Preprocessing: Data collected from Cadiz is converted into frames. Frames with low lightening density and poor visibility are discarded while the repeating frames with the same information is also not considered into the dataset. As most of the images do not contain any *Nephrops* information hence discarded during annotation phase.

3) Image Annotation: In this phase the data is annotated to record the ground truth annotations. Initially, the experts in *Nephrops* burrows detection from IEO manually drew the rectangles around *Nephrops* entrance of the burrow systems. Then, the images are labeled manually in Microsoft VOTT image annotation tool [19] to record the ground truth annotations and saved in Pascal VOC format. The saved XML annotation file contains image name, class name (*Nephrops*), and bounding box details of each object in the image. Fig. 4 shows an image from ISUNEPCA_0618 UWTV survey that is manually annotated using Microsoft VOTT.

The annotated images are validated from the experts from IEO to record only the exact burrows of *Nephrops*. Table I shows the ground truth annotated images of each station from ISUNEPCA_0618 UWTV survey that will is used in the model training and testing. A total of seven stations are annotated and recorded 266 annotated images. Table. I Shows the ground truth annotated images of each station from Cadiz that will is used in the model training and testing. A total of seven stations are used from Cadiz dataset and 266 annotated images which contains one or more nephrops burrows has been extracted from video files and used in this work.

TABLE I
Annotations for Selected Stations from ISUNEPCA_0618

Cadiz Dataset	
Station*	Annotations
RF01	42
RF03	75
RF04	34
RF05	31
RF07	13
RF08	36
RF09	35
Total	266

4) Preparation of Training Dataset: To use a deep neural model, the data should be divided into three subsets: train, validate and test. Table. II shows the distribution of Cadiz dataset. The original image size has been reduced to 717x403 pixels to reduce the computational cost of model training.

TABLE II
Distribution of Cadiz Dataset

Cadiz Dataset Distribution		
Training Images	Validation Images	Testing Images
200 (75%)	18 (7%)	48 (18%)
Total Images = 266		

B. Model Training

To train a model we used Convolution Neural Network (CNN). Instead of train our very own neural system, we utilized transfer learning [13] to retrain the Faster R-CNN Inception V2 model [14] and MobileNet V2 [15] model in Tensorflow [16]. We conducted the model training, validation and testing on a Linux Machine powered by an NVIDIA TitanXP GPU. 200 images from Cadiz dataset were used for training. The image annotations are stored in XML format to create Tensorflow (TF) records, which were in the model for training. Parameters used in the configuration were: a) Inception model: number of classes=2 (one for *Nephrops* and one as background), maxpool kernel size=2 and L2 regularization; b) MobileNet model: number of classes = 2 (one for *Nephrops* and one as background) and L2 regularization. Both models were trained with 20k steps.

The utility of a model depends on how accurately the model detected *Nephrops* burrows. For this purpose, we used precision (P), also called positive predictive value, and recall (R), also known as sensitivity or true positive rate, indicators, used in Receiver Operating Characteristics (ROC) methodology, which were calculated using the Equations 1 and 2:

$$P = \frac{TP}{TP + FP}, \tag{1}$$

$$R = \frac{TP}{TP + FN}, \tag{2}$$

where TP, FP and FN means number of true positives, false positives and false negatives, respectively. A true positive means that a *Nephrops* burrow as been correctly detected by the model, whereas a false positive means that model detected a *Nephrops* burrow which is not annotated by marine biologist experts and a false negative means that an annotated *Nephrops* burrow was not detected by the model. The true negative (TN) case is not considered here as it would be all possible bounding boxes that were correctly not detected.

Precision can be seen as how rigorous the model is at identifying the presence of *Nephrops* burrows, and recall is the rate of TP over the total number positives detected by the model [17]. Generally, when the recall increases, precision decreases and vice versa, so precision versus recall curves $P(R)$ are useful tools to understand model behaviour. To quantify how accurate the model with a single number, the mean average precision (mAP), defined in Equation 3, is used.

$$mAP = \int_0^1 P(R)dR. \tag{3}$$

In our problem, ground truth annotation and model findings are rectangular areas which usually doesn't fit perfectly. In this paper, it's considered a TP detection if both areas overlap more than 50%. This is computed by Jaccard index J, defined in Equation 4

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}, \tag{4}$$

where A and B are the set of pixels in the truth annotation and model finding rectangular areas, respectively, and $|\cdot|$ means the number of pixels in the set. When $J \geq 0.5$, a TP is detected, but if $J < 0.5$, detection fails with a FN. Using this methodology, P and R values are calculated and mAP is used as a single number measure of the goodness of the model. Usually, this parameter is named as $mAP50$, but we used mAP for simplicity in our paper.

C. Model Validation

Models were trained using a random 75% sample of the annotated dataset. The remaining 25% is used for validation and testing [18]. We measured the training performance by monitoring the overfitting of the model. We record the turning checkpoints after every 2k iterations and computed the mAP50 on the validation dataset.

D. Model Performance

The validation test is a well known method in machine learning to assess the performance of the model. We further tested our model against few unseen images from the Gulf of Cadiz dataset and evaluate the performance of the model.

IV. Results

We apply MobileNet v2 and Inception v2 based on faster RCNN to our *Nephrops* norvegicus data set from the Gulf of Cadiz. The aim is to identify the *Nephrops* burrows from the Cadiz dataset. We trained both MobileNet and Inception models over 20k iterations, and they achieve a good precision on the trained dataset.

A. Mean Average Precision(mAP)

Mean Average Precision(mAP) values were computed on the new dataset of *Nephrops* from Cadiz over 20k iterations and Fig. 5 shows the results obtained by both models. The best precision we achieved in Inception is 77.18% after 10k iterations, While the best precision in MobileNet is 65.43% after 10k iterations. As it can be seen in the figure, mAP increases with the number of iterations until 10k iterations, approximately, and the performace of the models decreases when models started to show overfitting behaviour. It also shown in the figure that Inception models shows a more stable performance with iteration number. In summary, these results clearly shows that the mAP of Inception model is better than the MobileNet for our problem.

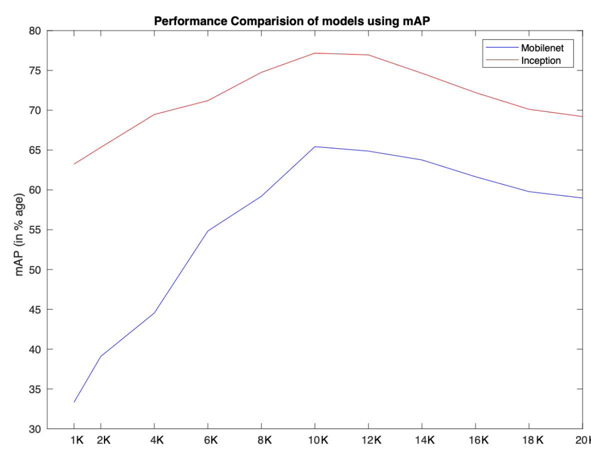


Fig. 5. mean Average Precision of Inception vs MobileNet

B. Precision and Recall

Another way we used to evaluate both models on our dataset with more detail is to analize the precision vs. recall curves. To evaluate both models the True Positive (TP), False Positive (FP), and False Negative (FN) annotations are detected. The True Negative (TN) case is not considered here because its poor definition, as it would be all possible bounding boxes that were correctly not detected. Table III shows the comparison of these metrics of both models.

TABLE III
DETECTION MATRIX OF TRAINING MODELS

	MobileNet	Inception
True Positive	31	53
False Positive	2	13
False Negative	46	13
Total Groundtruth	79	79

Table III shows clearly that MobileNet model detected a bigger number of FN whereas Inception model is more sensible to FP errors. In Figures 6 and 7 the precision and recall of MobileNet and Inception models are presented for 10k iterations. The mAP values could be seen as the area under these curves, but the behaviour is different. With the Inception model, precision is 1 up to values near to 0.5 of recall, when precision starts to decrease, whereas withe the MobileNet model, curve behaviour is more erratic.

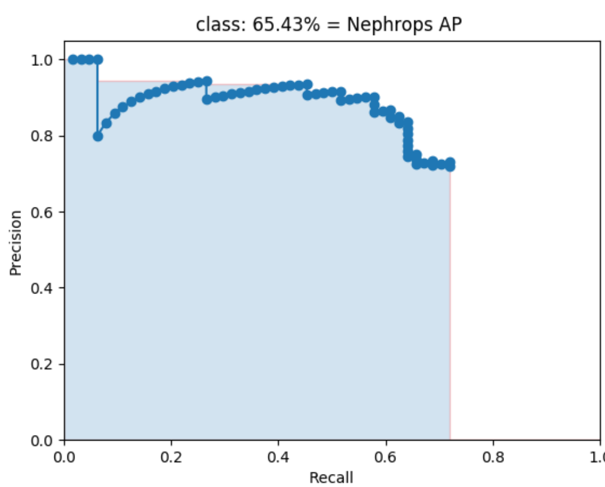


Fig. 6. Precision and Recall Curve of MobileNet

C. Total Loss

Loss function is a method to evaluate the learning model. If the predictions deviates too much from the ground truth, loss function is a large number .We calculated the total loss of both models during the training, which is shown in Fig. 8. The graph indicates that the loss tend to decrease as number of iterations increases but the loss of MobileNet model is much greater than the Inception in each iteration. After 10k iterations, the loss function tend to stabilize, but MobileNet

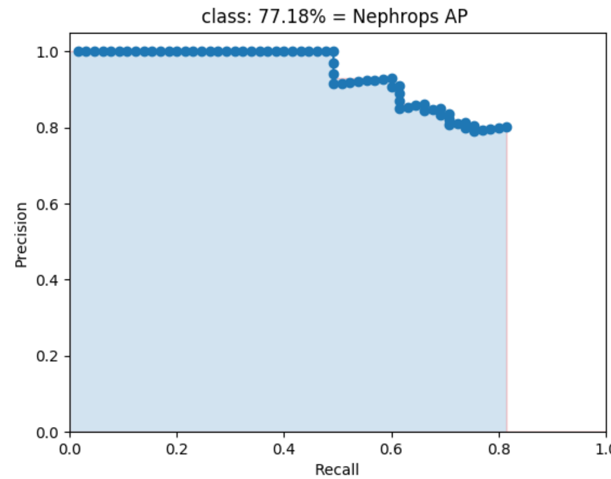


Fig. 7. Precision and Recall Curve of Inception

model presents a slower learning. Recorded minimum loss values are 0.006 for the Inception model and 0.704 for MobileNet.

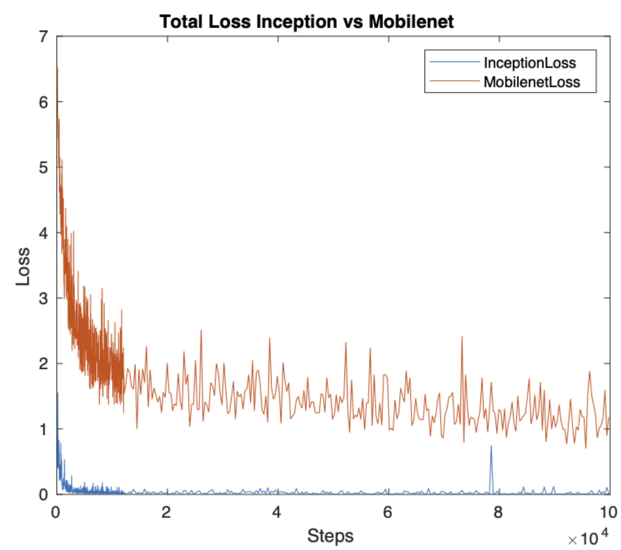
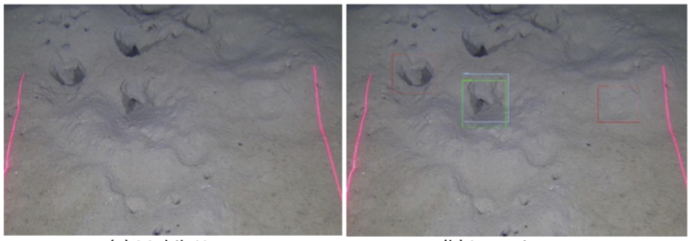


Fig. 8. Total Loss during training Inception vs MobileNet

D. Visualization

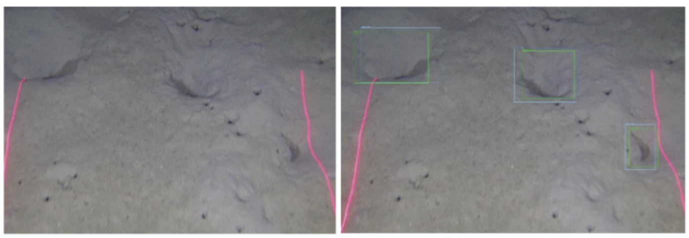
In this section, we will visualize some results of *Nephrops* burrows detections from Cadiz dataset images with three typical examples. Fig. 9 shows that from MobileNet model (left image) is unable to detect any burrow, whereas Inception model (right image) detects correctly one burrow (green and blue rectangles are model finding and true ground annotation, respectively) with 99% confidence. In this image, it can be seen that Inception model also detected two false positives (red rectangles on the right and left side of the TP). As other example, Fig. 10 also shows that MobileNet model is



(a) MobileNet (b) Inception

Fig. 9. *Nephrops* burrows detections MobileNet vs Inception. Example 1

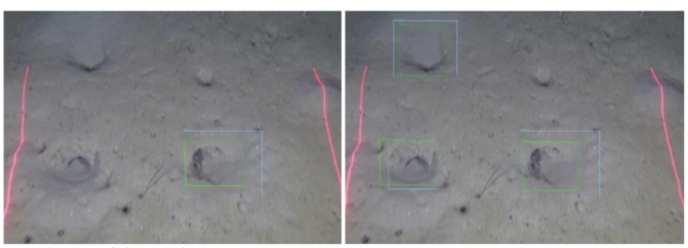
unable to detect anyentrance of a burrow system as compared to Inception model in which all three positive burrows are detected with 99% confidence. In a final example, Fig. 11



(a) MobileNet (b) Inception

Fig. 10. *Nephrops* burrows detections MobileNet vs Inception. Example 2

shows that Mobilenet model is able to detect one *Nephrops* burrow (from a total of three) with a confidence of 98% as compared to Inception model in which all the three positive burrows are detected with 99% confidence.



(a) MobileNet (b) Inception

Fig. 11. *Nephrops* burrows detections MobileNet vs Inception. Example 3

The visualization results clearly shows that the Inception v2 model is much better in precision and accuracy. The model trained by Inception detects more True Positives as compared to MobileNet. Visualization of results also helps to understand the nature of the errors of the models and to improve them in future work.

V. Conclusion and Future work

The Instituto Español de Oceanografía has a research group working on *Nephrops* norvegicus (known as Norway Lobster) identification and counting. In this work, we train, validate and test the current state of the art Faster RCNN models Inception v2 and MobileNet v2 for *Nephrops* burrows detection and classification. We presented the quantitative as well as visualization results of both models. We concluded that the Inception model is much better in detecting the *Nephrops* burrows as compared to MobileNet. In future work, we will plan to use more training data from Cadiz and also use the data from datasets of other countries, with differences in image quality, video acquisition procedures and background textures. Also, we will use more dense object detection models to improve the accuracy of detections.

Acknowledgment

We thanks Instituto Español de Oceanografía, Cádiz, Spain to provide us the dataset for research.

References

[1] K. Srividhya and M. Ramya, "Accurate object recognition in the underwater images using learning algorithms and texture features", Multimedia Tools and Applications, vol. 76, no. 24, pp. 25679-25695, 2017.

[2] T. Rimavicius and A. Gelzinis, "A Comparison of the Deep Learning Methods for Solving Seafloor Image Classification Task", Communications in Computer and Information Science, pp. 442-453, 2017.

[3] H. Qin, X. Li, Zhixiong Yang and M. Shang, "When underwater imagery analysis meets deep learning: A solution at the age of big visual data", OCEANS 2015 - MTS/IEEE Washington, 2015.

[4] C. Bell, F.Redant and I.Tuck, "Lobsters: Biology, Management, Aquaculture and Fisheries", pp. 412-461, 2006.

[5] "ICES, 2018. Report of the Working Group for the Bay of Biscay and the Iberian Waters Ecoregion (WGBIE)", ICES HQ, Copenhagen, Denmark, 3-10 May 2018.

[6] M. Jiménez, I. Sobrino and F. Ramos, "Objective methods for defining mixed-species trawl fisheries in Spanish waters of the Gulf of Cádiz", Fisheries Research, vol. 67, no. 2, pp. 195-206, 2004.

[7] "ICES, 2008. Report on the Workshop and training course on *Nephrops* burrow identification (WKNEPBID)," 2008.

[8] J. Durden, B. Bett, T. Schoening, K. Morris, T. Nattkemper and H. Ruhl, "Comparison of image annotation data generated by multiple investigators for benthic ecology", Marine Ecology Progress Series, vol. 552, pp. 61-70, 2016.

[9] J. Durden, I. Preuss, A. Albu, A. Purser, B. De Smet, C. Dominguez-Carrió,C. Yesson,D. de Jonge and D. Lindsay, "Report on the Marine Imaging Workshop," 2017.

[10] M. Moniruzzaman, S. Islam, M. Bennamoun and P. Lavery, "Deep Learning on Underwater Marine Object Detection: A Survey", Advanced Concepts for Intelligent Vision Systems, pp. 150-160, 2017.

[11] A. Mahmood, M. Bennamoun, A. Senjian, F. Sohel, F. Boussaid, K. Renae Hovey, G. Kendrick and R. Fisher, "Automatic annotation of coral reefs using deep learning", OCEANS 2016 MTS/IEEE Monterey, 2016.

[12] X. Li, M. Shang, J. Hao and Z. Yang, "Accelerating fish detection and recognition by sharing CNNs with objectness learning", OCEANS 2016 - Shanghai, 2016.

[13] S. Pan and Q. Yang, "A Survey on Transfer Learning", IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, 2010.

[14] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision,"IEEE Conference on Computer Vision and Pattern Recognition, 2016

[15] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, pp. 4510-4520, 2018.

[16] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen Z, C. Citro, G. Corrado and A. Davis, "Heterogeneous distributed systems," arXiv preprint arXiv:160304467.

[17] M. Everingham, L. Van Gool, Cl. Williams, J. Winn and A. Zisserman, "The pascal visual object classes (voc) challenge,". Int J Comput Vis vol 88: p. 303-338, 2010.

[18] S. Alexandropoulos, C. Aridas, S. Kotsiantis and M. Vrahatis, "Multi-Objective Evolutionary Optimization Algorithms for Machine Learning: A Recent Survey", Approximation and Optimization, pp. 35-55, 2019.

[19] CSE group, "Visual Object Tagging Tool (VOTT)," microsoft.com, https://github.com/microsoft/VoTT/ (viewed 21 June 2019).