Advanced Detections of Norway Lobster (Nephrops Norvegicus) Burrows using Deep Learning Techniques

Atif Naseer
Science and Technology Unit, Umm al Qura University, Makkah, Saudi Arabia
ETSI Telecomunicación, Universidad de Málaga, Málaga, 29071, Spain

***Abstract*—Marine experts are facing lot of challenges in habitat monitoring of marine species. One of the biggest challenges is the underwater environment and species movement. The other challenge is the data collection of marine species. People used the camera sensors and satellite data in the past for data collection but in this era the scientists are using underwater Autonomous Underwater Vehicles (AUVs), the Remotely Operated Vehicles (ROVs), and certain sledges with high-definition still and video cameras to record the underwater footages. The ocean is composed of thousands of species which make the environment more challenging to monitor any specific specie. This work will focus on specie named Norway lobster (*Nephrops* norvegicus). The *Nephrops* norvegicus is one of the commercial specie in the Europe and generates millions of dollars yearly. This specie lives under the seabed and leaves behind the burrow structure on the sea ground. The *Nephrops* spend most of their time under the seabed. The scientists are currently monitoring the habitat of *Nephrops* norvegicus by underwater television (UWTV) surveys that is collected yearly on many European grounds. The collected data is reviewed manually by the experts who count the burrows on the sheet. This work focuses on the automatic detection of *Nephrops* burrows from underwater videos using the deep learning techniques. This work trained the Faster R-CNN models Inceptionv2, MobileNetv2, ResNet50, and ResNet101. Instead of training the models from scratch we used the transfer learning technique to fine tune these networks. The data is obtained from the Gulf of Cadiz (FU30) station. Twenty-eight different set of experiments are performed. The models are evaluated quantitatively using the mean Average Precision (mAP), precision and recall curves. Also, the models are qualitatively analyzed by visually presenting the output. The results prove that deep learning techniques are very helpful for marine scientists to assess the *Nephrops* norvegicus abundance.**

Keywords—Nephrops norvegicus; deep learning; stock assessment; faster RCNN

I. INTRODUCTION

Marine ecosystems include the open, deep oceans and marine species. The environment has high level of dissolved salts. Marine ecosystem is one of the main sources of our daily food. The marine species have different physical and biological characteristics. Coral reef is a good example of marine ecosystem that is associated with other marine life like fishes and turtles. The oceans cover 70% of our planet, so the marine ecosystem covers most of our earth. As compared to the terrestrial ecosystem, the marine ecosystem is very challenging to study. Most of the challenges come due to the complex medium of sea. The environment of marine ecosystem has certain challenges like color variations, species movement, and turbidity [1]. Marine scientists are monitoring the environment from decades by collecting underwater species images using satellite, shipborne and camera. With the advancement of technologies, several new techniques like ROVs and AUVs are used by the scientists to record the images and videos of marine ecosystem. The scientists are still facing many challenges in the sea due to the illumination, views, variation in the lighting conditions and free natural environment [2].

Marine ecosystems have thousands of underwater species. Out of these species one of the important specie in Europe is Nephrops norvegicus (a Norway Lobster). This specie is considered as a commercial specie in Europe. This specie supports Europe with almost 60,000 t [3] and an income of 300 million € per year approximately [4].

The International Council for the Exploration of the Sea (ICES) is a marine science organization that leads the scientific forums on all domains of marine sciences. Their major goal is to advance the marine ecosystem. They provide state-of-the-art goals and facilities that help the scientists to do research in the marine eco system. There are many working groups under the umbrella of ICES that are conducting annual survey and monitoring the habitat of marine species. One of the major groups for Nephrops habitat monitoring is Working Group on Nephrops Surveys (WGNEPS), formerly known as the Study Group on Nephrops Surveys (SGNEPS). The aim of this group is to provide international coordination for Nephrops UWTV and trawl surveys in the North Atlantic. Each year the WGNEPS conducted a UWTV and trawl survey to assess the population of Nephrops. Nephrops populations are assessed and managed by Functional Units (FU) where there is a specific survey for each FU. Fig. 1 shows an individual Nephrops. Nephrops norvegicus lives in the sandy-muddy sediments and create burrows in the seabed [5]. An individual Nephrops specimen ranges in length of 2 – 5.5 cm with a maximum length of 24.0 cm. The most common length is about 19.0 cm [6].

A special equipment is used in the survey for data collection. Every year the UnderWater TeleVision (UWTV) and Trawl surveys are conducted all over the Europe by WGNEPS to estimate the abundance of Nephrops norvegicus specie. The surveys are used to provide population estimates for Nephrops based on Functional Units (FU). The survey data is stored in disks in the form of high-definition images and videos. The data is analyzed manually using the TV survey to classify and count the Nephrops burrows. Currently, the Nephrops data are collected through the yearly UWTV surveys and are reviewed by the marine experts manually. This manual process is very time consuming and leads to many errors due to environmental complexity and data variation. In this work, we are using the data obtained from the Gulf of Cadiz (FU30) station.

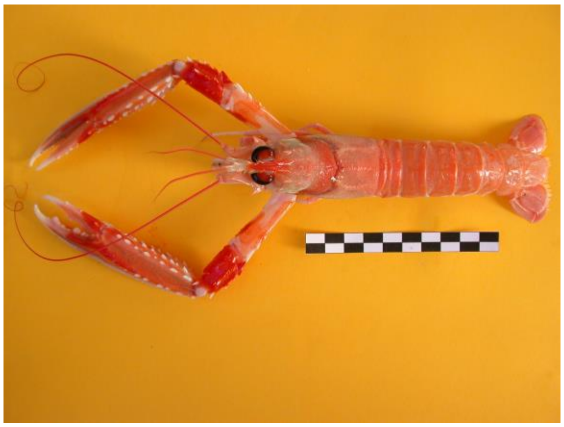


Fig. 1. Nephrops norvegicus.

Artificial Intelligence (AI) is an emerging field that solves many object detection problems including underwater species classification and detections. However, to detect and classify the Nephrops burrows for habitat monitoring, literature is unable to provide many solutions. One of the main reasons is the unavailability of Nephrops survey data. The complexity of data is also one of the reasons. This thesis is an effort to automate the existing method of Nephrops counting.

In our previous work [7] we trained and tested the MobileNet and Inception model and compare their results. In this work, we used denser Faster R-CNN models for training. We used the transfer learning technique and fine tune the MobileNetv2, Inceptionv2, ResNet50 and ResNet101 with FU30 dataset. The results obtained the good level of accuracy. ResNet101 obtained the highest level of accuracy among the other models.

The rest of the paper is organized as follows. Problem statement and definition is defined in Section II, the methodology is presented in Section III followed by the experiments and results in Section IV. The paper is concluded in Section V.

II. Problem Statement and Definition

We describe the problem of detections of Nephrops burrows in videos as currently the burrows are counting manually using the TV surveys. In this work, we demonstrate that the deep learning techniques will help to automatically detect and count the Nephrops burrows. Before going into the details, it is important to define the Nephrops burrows and their pattern.

Definition: A Nephrops burrow is an opening with following signature features: The burrow opening is like a half-moon shaped. The opening has proof of expelled sediments that creates scratches and tracks on the burrow opening. The pattern of burrows makes them unique as compared to other species burrows.

III. Proposed Methodology

This section discusses the approach for automatically identifying the Nephrops burrows from the video sequences using the deep learning techniques. The proposed methodology is illustrated in Fig. 2. The first part of the work is to collect the data from different stations of UWTV survey. In this study we are using the data collected from the Gulf of Cadiz (FU30) station. Specialized equipment is used for data collection at FU30. The second part of the methodology is the preprocessing of data that includes data cleaning, image annotations and data preparation. The third part is the detection of Nephrops burrows by applying the deep learning techniques. The deep learning models are trained and tested on the different datasets. We used transfer learning and fine tune the Faster R-CNN algorithms for model training.

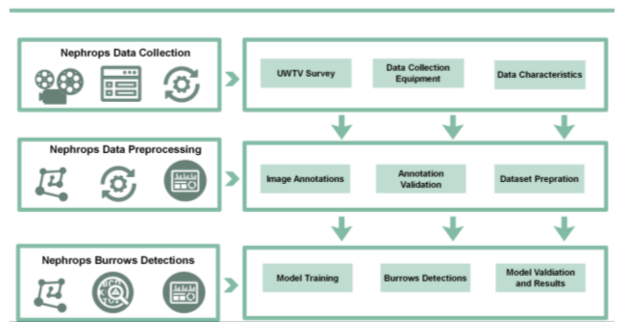


Fig. 2. Proposed methodology for Nephrops burrows detections.

A. Data Collection

In this research we used the data from the Gulf of Cadiz (FU30) station. To observe the habitat of *Nephrops* at FU30 station, a survey is designed yearly to collect the data. The survey used specially designed equipment for data collection.

1) Data collection equipment: A special sledge is designed in the survey of FU30 station. The sledge is equipped with Sony FDRAX33 camera that is used to capture the videos in high quality. The camera is mounted at an angle of 45 degree. Two laser lights are used in the sledge that define the field of view. The field of view is set to 75 cm. Fig. 3. shows the inner view of the sledge used in FU30 station survey.

Fig. 3. Sledge inner view used in FU30 survey.

2) Data collection procedure: In 2018, the survey at FU30 is conducted at 70 different stations. The station is defined as a geostatistical location in the sea where the *Nephrops* density is assumed to be present and estimated in the past. The sledges are placed on a big ship and at each station it is dropped down in the sea and deployed to the sea ground. To maintain a constant speed of sledge, it is towed between 0.6–0.7 knots. This condition of sledges is the best possible condition for *Nephrops* burrows counting. The sledge is mounted with high-definition video camera that records the video footage of 10–12 min at 25 frames per seconds. The area covered by the sledge during this video footage is around 200m approximately. The sledge position is recorded after every 1 to 2 seconds for calibration. Two laser lights are placed that confirmed the field of view of the video footage. The field of view is set to be 75cm and the sledge distance over ground (DOG) is estimated from the position of the sledge. Fig. 4. shows the sample image collected during the survey.

Fig. 4. FU30 Sample image.

3) Data characteristics: At FU 30, Video footages of 10-12 minutes has been recorded at 25 frames per second in good lighting condition. The video is recorded with a resolution of 3840 x 2160. We got the data of FU30 from the UWTV survey of 2018. A total of 70 stations are surveyed. After evaluating all the stations carefully, we only choose seven stations for annotations. The stations are selected based on their higher contrast, good video quality, lightning conditions, and high burrows density rates. The recorded videos are saved to the disks for manual *Nephrops* counting.

B. Data Preprocessing

The data preprocessing is one of the important phases of our methodology. Without performing the preprocessing step, the data comes with lot of noise and error that can affect the accuracy of the model training. The preprocessing is defined as cleaning of data, annotating the images and validation and preparation of dataset.

1) Data cleaning: The collected data is converted into frames and each frame is observed carefully. The frames with poor lighting and visibility are discarded initially. Also, the frames with zero burrows density are discarded before the annotations.

2) Image annotations: Image annotation is the most critical part of this study. Before the annotation, a comprehensive training is required to understand the burrow's characteristics. Certain protocol is followed in manual counting that should be observed during the manual image annotations. For image annotation, we used the Microsoft VoTT [8] tool. The *Nephrops* burrow is annotated by drawing the bounding box around it. The annotations are saved in the Pascal VOC format.

3) Annotation validations: Each annotation is validated by the Marine experts from the Gulf of Cadiz. The final annotations for model training are obtained after several iterations of validations. The validated annotations are saved in the XML format.

4) Dataset preparation: The annotated image dataset is divided into training and testing data separately. Each station consists of around 15,000 - 18,000 frames. From 2018 survey, almost 105,000 frames were recorded from seven different stations. The training and testing dataset is divided to 80-20 ratio respectively. Table I shows the training and testing dataset distribution used in this study.

TABLE I. DATASET DISTRIBUTION

***Nephrops* Dataset Distribution**			
Functional Unit	**Training Images**	**Testing Images**	**Total Images**
FU30	200	48	248

C. Nephrops Burrows Detection

AI plays an important role nowadays in automating the analysis. In marine sciences many scientists apply AI techniques to monitor the habitats of marine species. Computer Vision and Deep learning shows a significant improvement in the object detection [9,10], classification [11,12], and segmentation [13].

1) Model training: To train the models, the transfer learning [14] technique is utilized to fine-tune the Faster R-CNN Inceptionv2 [15], MobileNetv2 [16], ResNet50[17], and ResNet101[18] models in TensorFlow [19].

Inception networks are considered as one of the big milestones in CNN. For computational complexity, the Inception v2 network used smart factorization method with 5x5 convolution and two 3x3 convolution. We fine tune the network parameters with a learning rate of 0.01 and a batch size of 1. The value of Maxpool stride and Maxpool kernel size are set to 2. The gradient clipping [20] value if set too low or too high will result in the model instability so we set its threshold value to 2.0. For activation function we used the Softmax and Mask RCNN is used as a box predictor.

MobileNetv2 architecture is used with the relatively small dataset. MobileNetv2 architecture used a depth-wise separable convolution instead of conventional convolution. The architecture of this network is composed of 32 convolutional layers and 17 residual bottleneck layers. We fine tune the certain parameters of the network to get the optimize results. The learning rate is set to 0.01. The batch size is set to 24 with truncated normal initializer. For activation function we used Rectified Linear Unit (ReLU) and Convolutional box predictor is used as a box predictor.

ResNet50 [17] is a variant of the ResNet model. The ResNet50 is 50-layers deep convolutional network. Out of these 50 layers, 48 are convolutional layers, one max pool, and one average pool layer. In the first convolution only one layer is used with a kernel size of 7 × 7,64 kernels with stride 2 and a max pool of size 3 × 3. In the second convolution, nine layers are used with a kernel size of 1 × 1,64; 3 × 3,128. In the third convolution, 12 more layers are used with 1 × 1,128; 3 × 3,128, and 1 × 1,512 kernel. The fourth convolution uses 18 more layers with kernel sizes of 1 × 1,256 ;3 × 3,256 and 1 × 1,1024. Nine layers are used in the fifth convolution with kernel sizes of 1 × 1,512; 3 × 3,512 and 1 × 1,2048. Finally, the last convolution layer is used for avg pool and a Softmax function. We set the learning rate of 0.003 with batch size 1 and L2 regularization. The Softmax is used as an activation function and Mask RCNN is used as a box predictor.

The ResNet101 [18] is a considered to be 101 layers dense convolutional neural network. The first convolution layer has a kernel size of 7 × 7,64 with stride 2 and a max pool of size 3 × 3. The second convolutional layer used nine layers with kernel sizes of 1 × 1,64 and 3 × 3,28. The third convolutional layer used 1 × 1,128; 3 × 3,128, 1 × 1,512 kernels. Sixty-nine layers are used in the fourth convolutional layer with following kernels 1 × 1,256; 3 × 3,256 and 1 × 1,1024. The fifth convolution uses 9 layers with 1 × 1,512; 3 × 3,512 and 1 × 1,2048. Finally, the last convolutional layer is used for avg pool and a Softmax function. The learning rate is set to 0.0003 with 24 batch size and L2 regularization. The Maxpool kernel size and Maxpool stride is set to 2 and Mask RCNN is used as a box predictor.

Table II shows the parameter list and their values used in the Inceptionv2, MobileNetv2, ResNet50 and ResNet101 model.

2) Model validation: The models used in this study for training used approximately 80% of the annotated dataset. The remaining 20% is used for the testing. All the models are trained over 70k iterations and the turning checkpoints during the training is recorded after every 10k iterations. For validation of models the models are evaluated using mAP, precision and recall curve and through visual inspection of the output images with detections.

TABLE II. FASTER R-CNN MODELS TRAINING PARAMETERS

Parameters	**Inceptionv2**	**MobileNetv2**	**ResNet50**	**ResNet101**
Number of Classes	01	01	01	01
Optimizer	Momentum	RMSProp	Momentum	Momentum
Momentum Rate	0.9	0.9	0.9	0.9
Learning Rate	0.01	0.01	0.0003	0.0003
Batch Size	1	24	1	24
Initializer	truncated_normal_ initializer	truncated_normal_initializer	truncated_normal_initializer	truncated_normal_initializer
gradient_clipping_by_norm	10	-	10	10
Regularization	L2	L2	L2	L2
Activation Function	Softmax	RELU	Softmax	Softmax
Maxpool kernel size	2	-	2	2
Maxpool stride	2	-	2	2
Box Predictor	Mask RCNN box predictor	Convolutional box predictor	Mask RCNN box predictor	Mask RCNN box predictor

IV. EXPERIMENTS AND RESULTS

1) Experiments: In this section we will summarizes all the experiments and results performed to automatically detect the *Nephrops* burrows. The models are trained and tested with FU30 dataset. Four different combinations of set of experiments are performed with the current dataset. Each set of experiment is iterated 7 times hence a total of 28 experiments are performed.

2) Results

a) Quantitative analysis: The mAP of all the models trained and tested by FU 30 stations are calculated during the quantitative analysis. In this study we trained the models with 70k iterations. The model performance is recorded in terms of precision and recall values after every 10k iterations. The precision is the prediction accuracy measurement and recall are the measurement of positive predictions. The mAP is calculated for each model which is a very common metric to evaluate the performance of object detection algorithms. The mAP is defined in the in Eq. (1).

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

Precision can be seen as how robustly the model identifies *Nephrops* burrows' presence, and Recall is the rate of True Positive (TP) over the total number positives detected by the model [21]. The precision and recall curves are used to measure the model behavior.

In our study, the ground truth annotations and model predictions are rectangular boxes that usually don't fit perfectly. In this paper, the detections are considered as TP if the detection and ground truth overlap more than 50%. This is calculated by the Jaccard index J, as defined in Eq. (2).

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}, \tag{2}$$

Here, A and B are the set of pixels in the ground truth annotation and model predictions respectively, and | . | means the number of pixels in the set. If $J \geq 0.5$, a TP is detected, but if $J < 0.5$, detection fails with a False Negative (FN). This methodology is used to calculate the precision and recall values.

Table III shows the maximum mAP obtained by MobileNetv2, Inceptionv2, ResNet50, and ResNet101 models, respectively. The maximum mAP obtained using the MobileNet model is 65.69. The maximum mAP obtained by the Inception model is 77.18. The ResNet50 and ResNet101 models achieve better precision values as compared to MobileNet and Inception. The maximum mAP in ResNet50 is 80.16 while in ResNet101 it is 81.59.

The results are also presented in the form of precision and recall curves. Fig. 5 shows the results obtained with the models trained and tested by FU 30 dataset. The best mAP is 81.59 with ResNet101 model.

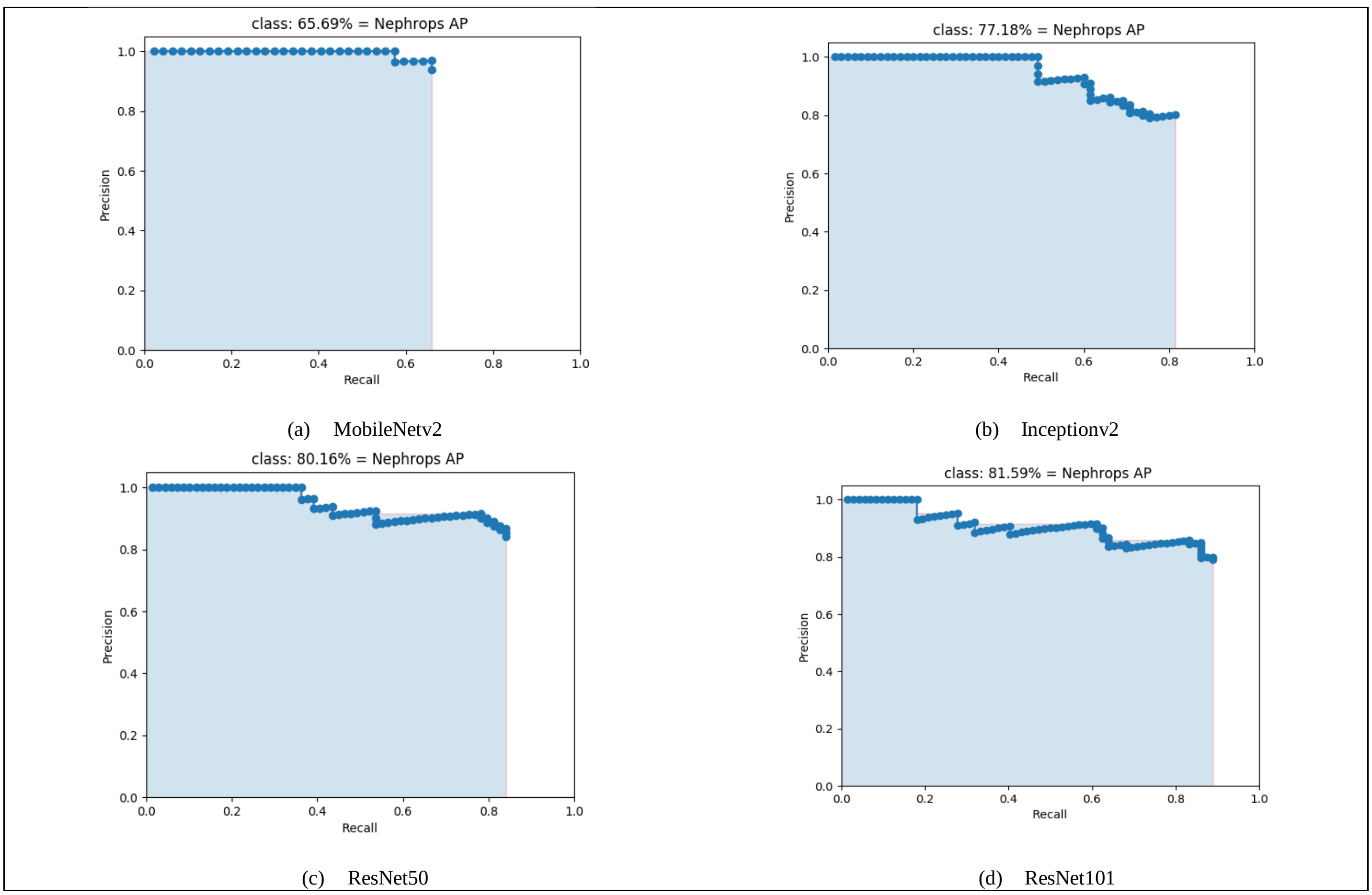
b) Qualitative analysis: In this section, the performance of different models on the dataset is analyzed qualitatively. The visualization results are from the MobileNetv2, Inceptionv2, ResNet50, and ResNet101 models that are trained and tested by the FU30 dataset.

Fig. 6 shows the Nephrops burrows detections visually using the MobileNetv2, Inceptionv2. ResNet50, and ResNet101 models with FU 30 dataset. The green rectangular boxes on the images shown are the TP detections by the model. The blue bounding boxes are the actual ground truth annotations while, the red bounding boxes are the False Positive (FP) detections that are detected by the trained models. In this example, MobileNetv2 model detects one TP burrow while the inception and all other models correctly detects the two TP Nephrops burrows.

The overall study shows that the ResNet101 model performs better in terms of mAP and provides an accuracy of more than 80%.

TABLE III. mAP OBTAINED USING MULTIPLE TRAINING MODELS.

mAP obtained with multiple Training models	
Trained Model	mAP
MobileNetv2	65.69
Inceptionv2	77.18
ResNet50	80.16
ResNet101	81.59



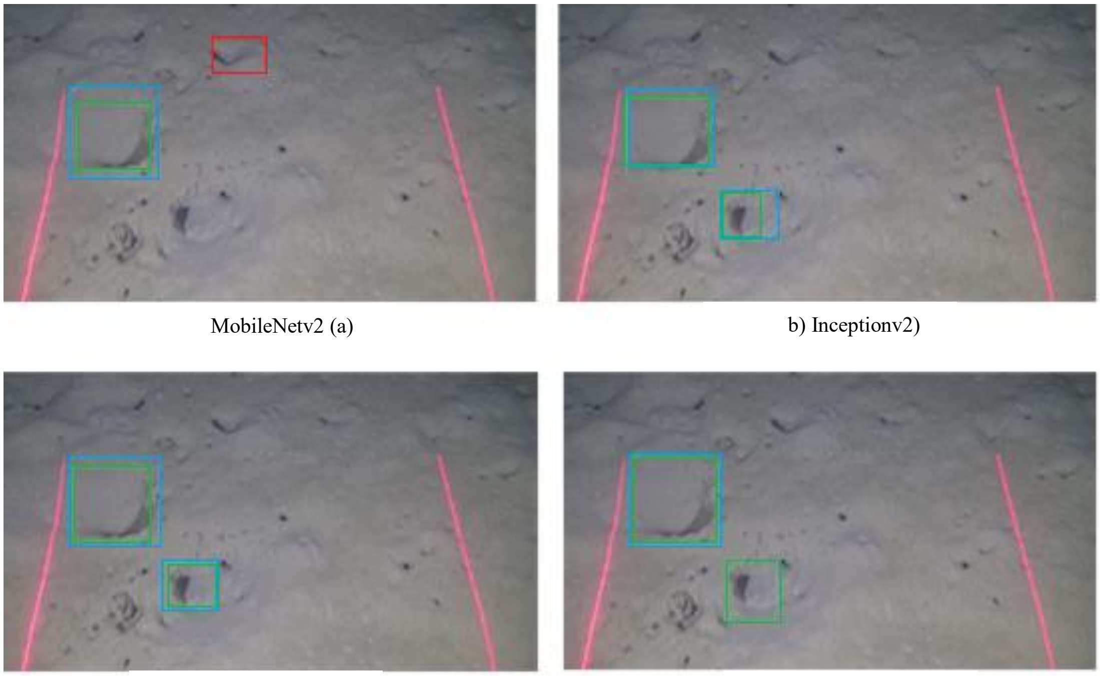
(a) MobileNetv2

(b) Inceptionv2

(c) ResNet50

(d) ResNet101

Fig. 5. Precision-recall curve obtained using FU 30 dataset (a) PR-curve of MobileNet, (b) PR-curve of Inception, (c) PR-curve of ResNet50, (d) PR-curve of ResNet101.



MobileNetv2 (a) b) Inceptionv2)

c) ResNet50) d) ResNet101)

Fig. 6. Nephrops burrows detection with FU30 dataset (a) Detection with MobileNet model, (b) Detection with Inception model, (c) Detection with ResNet50 model, (d) Detection with ResNet101 model.

V. Conclusion and Future Work

The aim of this study is to automatically detect and classify the Nephrops norvegicus burrows from underwater videos. We used the dataset from the Gulf of Cadiz (FU30) survey in 2018. We trained four different Faster R-CNN models to study the detection of Nephrops burrows using deep learning. The results show that the deep learning algorithms are very effective in detecting the burrows automatically. The ResNet101 model performs better and achieves the mAP of 81.59. This practice helps the marine scientists to correctly estimate the abundance of Nephrops from the underwater videos. The automatic detection algorithms could replace the manual counting process of marine experts and provide an accurate count in very less time.

In future work, we will plan to use a bigger curated dataset from different stations working under the ICES. The more data will improve the performance of accuracy of the deep learning models. Also, the newer model based on YOLO architecture will be trained in the future work. Finally, we will plan to integrate the spatial and temporal information of the Nephrops burrows to estimate the burrow sizes and their complexes.

Acknowledgment

We thank the Spanish Oceanographic Institute, Cádiz, Spain for providing the dataset for research.

References

[1] T. Rimavicius and A. Gelzinis, "A comparison of the deep learning methods for solving seafloor image classification task," Communications in Computer and Information Science, vol. 756, pp. 442–453, 2017.

[2] H. Qin, X. Li, Z. Yang and M. Shang, "When underwater imagery analysis meets deep learning: A solution at the age of big visual data," in Proc OCEANS'15 MTS/IEEE, Washington, DC, USA, pp. 1–5, 2015.

[3] FAO (2021a). "FAO yearbook. fishery and aquaculture statistics 2019/FAO annuaire," in Statistiques des pê ches et de l'aquaculture 2019/FAO anuario. estadısticas de pesca y acuicultura 2019´ (Rome/Roma: FAO).

[4] Issifu, I., Alava, J. J., Lam, V. W., and Sumaila, U. R. (2022). Impact of ocean warming, overfishing and mercury on European fisheries: A risk assessment and policy solution framework. Front. Mar. Sci. 8. doi: 10.3389/fmars.2021.770805.

[5] 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.

[6] Fischer, W., G. Bianchi and W.B. Scott 1981 Lobsters. 5: pag.var. In FAO Species identification sheets for fishery purposes. Eastern Central Atlantic (fishing areas 34, 47; in part). Canada Funds-in-Trust. Ottawa, Department of Fisheries and Oceans Canada, by arrangement with the Food and Agriculture Organization of the United Nations. 1-7.

[7] Naseer, A., Baro, E. N., Khan, S. D., Vila, Y., Doyle, J. (2022). Automatic Detection of *Nephrops* Norvegicus Burrows from Underwater Imagery Using Deep Learning. CMC-Computers, Materials & Continua, 70(3), 5321–5344.

[8] Microsoft CSE group. (2020, June 3), "Visual object tagging tool (VOTT), an electron app for building end to end object detection models from images and videos, v2.2.0. [Online]. Available: https://github.com/microsoft/VoTT.

[9] R. Girsshick, J. Donahue, T. Darrell and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 1, pp. 142–158, 2016.

[10] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-cNN: Towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, 2017.

[11] R. Shima, H. Yunan, O. Fukuda, H. Okumura, K. Arai et al. "Object classification with deep convolutional neural network using spatial

information," in Proc. Int. Conf. on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Okinawa, Japan, pp. 135–139, 2017.

[12] S. Soltan, A. Oleinikov, M. Demirci and A. Shintemirov, "Deep learning-based object classification and position estimation pipeline for potential use in robotized pick-and-place operations," Robotics, vol. 9, no. 3, 2020.

[13] S. Masubuchi, E. Watanabe, Y. Seo, S. Okazaki, K. Watanabe et al. "Deep-learning-based image segmentation integrated with optical microscopy for automatically searching for two-dimensional materials," npj 2D Mater Appl, vol. 4, no. 3, pp. 1–9, 2020.

[14] https://machinelearningmastery.com/how-to-use-transfer-learning-when-developing-convolutional-neural-network-models/.

[15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 2818–2826, 2016.

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, "Mobilenetv2: inverted residuals and linear bottlenecks," in Proc. Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, pp. 4510–4520, 2018.

[17] Understanding and Coding a ResNet in Keras. Available online: https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33.

[18] TensorFlow Core v2.8.0. Available online: https://www.tensorflow.org/api_docs/python/tf/keras/applications/resnet/ResNet101 (accessed on 20 March 2022).

[19] https://www.wgu.edu/blog/neural-networks-deep-learning-explained2003.html#close.

[20] R. Pascanu, T. Mikolov and Y. Bengio. "On the difficulty of training recurrent neural networks," ArXiv Preprint, vol. 1211, 5063, pp. 1–12, 2012. [Online]. Available: https://arxiv.org/pdf/1211.5063.pdf.

[21] M. Everingham, L. Van Gool, Cl. Williams, J. Winn and A. Zisserman, "The pascal Visual Object Classes (VOC) challenge," International Journal of Computer Vision, vol 88, pp. 303-338, 2010.