Advancing Automated Detection of *Nephrops norvegicus* Burrows in Underwater Television Surveys Through Machine Learning

Oscar Papini[a,*](https://orcid.org/0000-0003-2069-5068),

Enrico Cecapolli[b,**](https://orcid.org/0009-0009-7996-1691),

Filippo Domenichetti[b,***](https://orcid.org/0000-0002-8052-7544),

Michela Martinelli[b,****](https://orcid.org/0000-0003-1231-2346),

Gabriele Pieri[a,*****](https://orcid.org/0000-0001-5068-2861),

Marco Reggiannini[a,******](https://orcid.org/0000-0002-4872-9541),

Lorenzo Zacchetti[b,c,*******](https://orcid.org/0000-0003-3179-7473)

[a] *Institute of Information Science and Technologies (ISTI), National Research Council of Italy, 56124 Pisa, Italy*

[b] *Institute for Marine Biological Resources and Biotechnology (IRBIM), National Research Council of Italy, 60125 Ancona, Italy*

[c] *Department of Biological, Geological, and Environmental Sciences (BiGeA), Alma Mater Studiorum – University of Bologna, 40126 Bologna, Italy*

* e-mail: oscar.papini@isti.cnr.it

** e-mail: enrico.cecapolli@irbim.cnr.it

*** e-mail: filippo.domenichetti@cnr.it

**** e-mail: michela.martinelli@cnr.it

***** e-mail: gabriele.pieri@isti.cnr.it

****** e-mail: marco.reggiannini@isti.cnr.it

******* e-mail: lorenzo.zacchetti@irbim.cnr.it

Abstract— The paper introduces computer vision methods for automating the detection, recognition, and classification of *Nephrops norvegicus* burrows in underwater videos. This approach aims to improve accuracy, reduce human errors, and standardize the current manual video analysis process. By using machine learning techniques, the system can automatically process video streams and detect *N. norvegicus* burrow openings on the seabed. The work also explores the use of data augmentation algorithms to extend the annotated data set, enhancing the performance of the automated system compared to the original manual annotations.

Keywords: Deep learning; Video annotation; Dataset augmentation; Fisheries; *Nephrops norvegicus*

INTRODUCTION

Nephrops norvegicus (Norway lobster) is a benthic burrowing decapod crustacean that typically lives at depths between 50 and 800 m on the continental shelf of the northeastern Atlantic Ocean and the Mediterranean Sea [3]. This species greatly contributes to the total fish landings in Europe, with an average of 3700 tonnes per year just in the Mediterranean Sea, even though landings have seen a significant decrease since 2010 [7].

Underwater Television (UWTV) surveys are carried out to evaluate *N. norvegicus* stocks [2, 6]. These consists of the examination of underwater (UW) video footage of the sea floor to count the number of *N. norvegicus* burrows, distinguishing them from other species' burrows and other objects. This gives an estimate of the *N. norvegicus* population under the assumption of "1 burrow = 1 animal" [17], which is generally considered reliable among experts despite being currently under discussion [1]. Still, the careful visual analysis of hours of UW footage to count burrows is a highly demanding and time-consuming task, often susceptible to errors and influenced by the judgment of the human scientists who perform it. Several workshops have been organized by the International Council for the Exploration of the Sea (ICES) with the purpose of defining standard procedures for video reading [5, 8–10], and a Working Group on *Nephrops* Surveys (WGNEPS) has been established within ICES to coordinate the activities regarding UWTV surveys [12]. For example, a training program has been developed for readers to learn to recognize *N. norvegicus* burrows using reference material and videos.

In this context, an automatic system that detects (and potentially counts) *N. norvegicus* burrows would be an invaluable tool to ease the workload of marine biologists and potentially reduce the amount of error and human subjectivity. Such a system should at first identify the openings in the seabed relative to *N. norvegicus* burrows, discarding other objects that could be mistaken for them (e.g. other species' burrows, rocks, sand clouds, shadows, visual artifacts), then identify the burrow itself, which is a system of two or more entrances connected via underground tunnels.

This paper focuses on the first of the two mentioned steps: it investigates how effective machine learning (ML) techniques—and in particular deep learning algorithms—are in recognizing entrances of *N. norvegicus* burrows in UWTV footage. Some recent works already show promising results [4, 16]. The use of deep neural networks has risen considerably in the last few years in the field of computer vision, with the most diverse applications. In case of UW images and videos, there are a few intrinsic issues to be taken into consideration: the underwater environment alters the optical features of the images (e.g. turbidity and strong light absorption cause a shift of the hue towards the green/blue spectrum and an overall contrast reduction), often producing low-quality images that are not suitable to be fed to a ML algorithm; moreover, the typical depths at which UWTV footage is recorded during surveys require necessarily the use of artificial illumination systems, that may result in unbalanced images in terms of brightness and contrast. Given such conditions, the care and effort spent to create datasets that will be used in ML algorithms become of utter importance; on the other hand, the manual annotation of frames to obtain a high number of samples, as demanded by most algorithms, suffers from the same flaws as the burrows counting task: it takes a lot of time, and it is prone to errors.

This works describes the first results of the training of a state-of-the-art deep neural network, developed for the object detection on a set of UW images, to get some models able to recognize the entrances of *N. norvegicus* burrows. First, two training datasets are described in detail: one of them consists of a set of video frames manually annotated by experts, the other has been automatically generated by "extending" these annotations to the other frames of the video. Then, the employed neural network architecture and the related training configurations are illustrated and examined. The paper continues with a discussion about the predictive performance of the trained models and concludes with some final remarks and perspectives on future work.

EXPERIMENTAL SETUP

Dataset. An annotated dataset of *N. norvegicus* burrows images has been prepared starting from UW video footage recorded during a series of campaigns carried out jointly by the Institute for Marine Biological Resources and Biotechnology of CNR (Italy) and the Institute of Oceanography and Fisheries (Croatia) in the Pomo/Jabuka Pits area in the central Adriatic Sea [11, 14].

The recording apparatus consists of a UW high-resolution color camera (Kongsberg Maritime oe14-372) mounted on a sledge towed through a cable by the R/V *G. Dallaporta* at a constant speed of approximately 1 kn, thus collecting images of the seafloor with a fixed field of view of 80 cm. The captured footage was saved through an HD/SD digital video recorder (Datavideo HDR-60) with a resolution of 768×576 pixels and a frame rate of 25 FPS. The effective height was reduced to 432 pixels due to the presence of an automatic overlaid timestamp in the upper part of the frame. For the purposes of this work, a segment of about 4 minutes and 12 seconds of length was selected and deinterlaced using the Neural Network Edge Directed Interpolation (NNEDI) filter, and a set of 505 still frames, one every half second, was extracted and given to the annotators.

The annotation process itself has been performed using Microsoft VoTT [15]. Two object labels were chosen: “Entrance”, to denote the single openings of a *N. norvegicus* burrow on the sea floor, and “Burrow”, to group together the entrances relative to the same system (see Figure 1). Please note that all the experiments described here involve the detection of single entrances: objects of the class “Burrow” will *not* be taken into consideration, but having that information available in the dataset will certainly prove useful for the future. In the end, only 156 frames contained at least one object, and the total number of “Entrance” boxes is 315. This dataset will be called D_0.

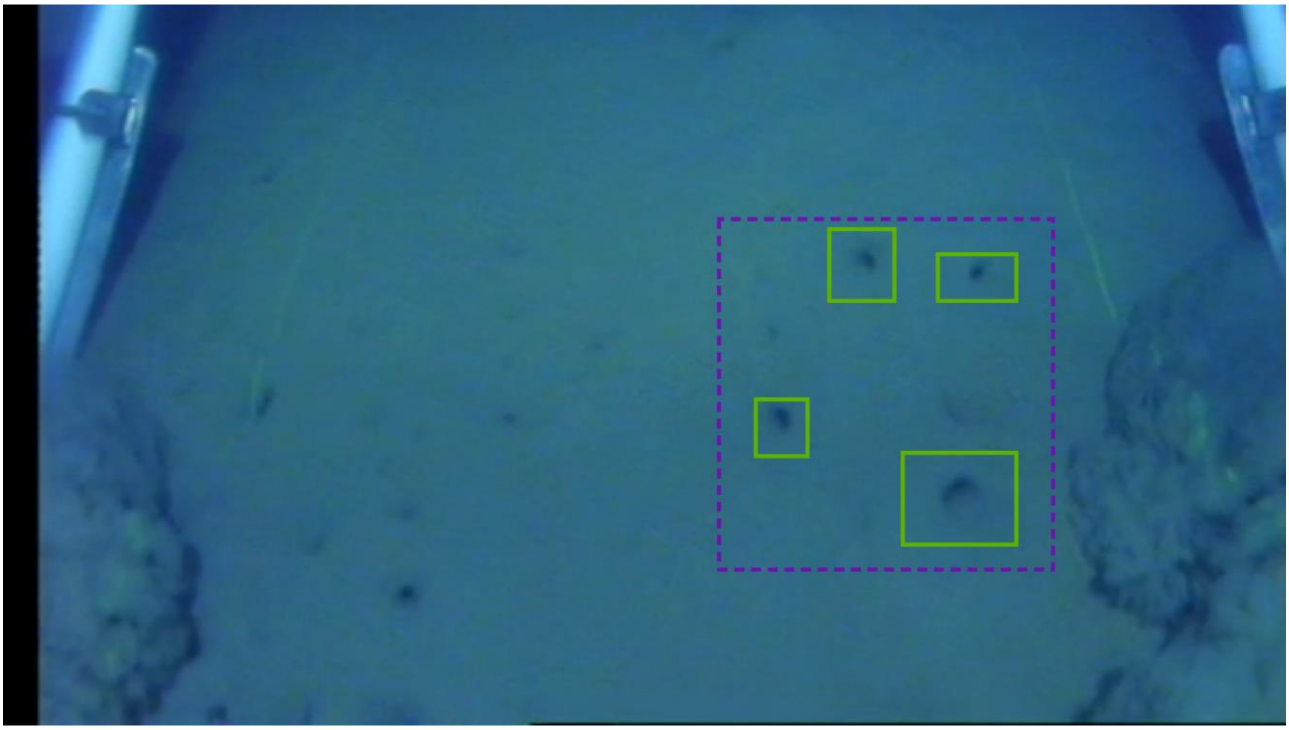

Figure 1. Example of annotated image in the dataset. The green boxes are "Entrance" objects; the dashed purple box is the "Burrow" object that groups them.

In order to have an increased number of samples to be fed to the ML algorithms, an automatic "boosting" process has been carried out. The premise is that, given that the footage is acquired by a camera mounted on a sledge at a fixed angle and (ideally) moving linearly at constant speed, an entrance on the seabed persists for a certain number of frames, so a bounding box in a frame F of D_0 defines a template that can be found in the frames belonging to a temporal neighborhood of F through correlation-based techniques. By applying this boosting algorithm, a new "extended" dataset, which will be called D^+, has been obtained; it contains 2406 images for a total of 4844 "Entrance" objects (see

Figure 2).

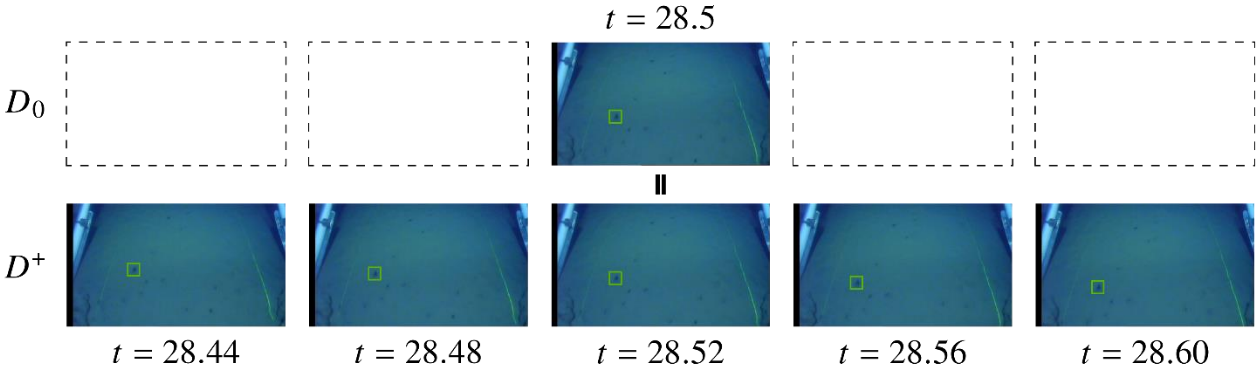

Figure 2. Concept of boosting algorithm. A burrow entrance in a frame of D_0 is used as a template to find the same entrance in neighboring frames in D^+.

An image in the datasets is uniquely identified by its timestamp, i.e. the number of seconds passed since the beginning of the video. Given the frame rates in play, it was established that images of D_0 have a timestamp with a single digit after the decimal point, that can be either 0 or 5; images of D^+ instead have two digits, which represent an integer divisible by 4. Some examples: 21.5 is a timestamp of an image in D_0; 77.28 is a timestamp of an image in D^+; the frame after 1 minute of the video has timestamp 60.0 as an image in D_0 and timestamp 60.00 as an image in D^+; 34.8 and 154.62 are not valid timestamps. Since 50 is not divisible by 4, there is a slight mismatch between timestamps of D_0 and D^+, where the "half-second" images of D_0 correspond to images in D^+ whose timestamps end with 52. The last images in D_0 and D^+ have timestamps 251.5 and 251.52 respectively.

Network Training. The datasets described above have been used to train several instances of the YOLOv8 network by Ultralytics [13], a state-of-the-art deep neural network designed for object detection and classification. More precisely, all the experiments have been conducted on a computational cluster equipped with a NVIDIA A100 40 GB PCIe GPU using the "small" variant YOLOv8s pre-trained on the COCO dataset.

Both datasets D_0 and D^+ have been split into training and validation sets. In order to avoid similar frames appearing in the two sets, the division is not random, but a timestamp t_0 has been chosen so that all frames with timestamp before or equal to t_0 belong to the training set, and all frames with timestamp after t_0 belong to the validation set. In the end, the value $t_0 = 191.64$ has been selected (for both datasets): this way about 80 % of the boxes belong in the training segments. This gave a train/validation split of 125/31 images for D_0 and 1892/514 images for D^+.

As is usual for object detection tasks, the *average precision* (AP) has been chosen to evaluate the performances of the model during training, with an Intersection over Union (IoU) threshold of 50 %. AP of the validation set is computed automatically after each epoch by the training algorithm.

The following settings are common to all the experiments:

- to comply with the network specifications and considering the available computational resources, the images have been resized to 416×416 pixels;

- the training duration is set to 200 epochs, with a “patience” value of 100 epochs, meaning that training is stopped automatically if no significant improvement is measured for 100 consecutive epochs;
- all other parameters maintain their default values unless explicitly specified.

Experiment 1. The default training algorithm of YOLOv8s automatically performs some standard data augmentation techniques to improve detection by artificially increasing the variability of the environmental conditions in which the images of a dataset are taken. Some of these techniques are listed in Table 1. In this experiment the network is trained on the dataset D_0 and the data augmentation techniques of Table 1 are *disabled*, i.e. their values are all set to zero.

Table 1. Some data augmentation techniques applied by default (table adapted from https://docs.ultralytics.com/modes/train/)

Name	Default value	Effect
hsv_h	0.015	Adjusts the hue of the image by a fraction of the color wheel, introducing color variability.
hsv_s	0.7	Alters the saturation of the image by a fraction, affecting the intensity of colors. Useful for simulating different environmental conditions.
hsv_v	0.4	Modifies the value (brightness) of the image by a fraction, helping the model to perform well under various lighting conditions.
translate	0.1	Translates the image horizontally and vertically by a fraction of the image size, aiding in learning to detect partially visible objects.
scale	0.5	Scales the image by a gain factor, simulating objects at different distances from the camera.
fliprl	0.5	Flips the image left to right with the specified probability, useful for learning symmetrical objects and increasing dataset diversity.
mosaic	1.0	Combines four training images into one, simulating different scene compositions and object interactions.
erasing	0.4	Randomly erases a portion of the image during classification training, encouraging the model to focus on less obvious features for recognition.

Experiment 2. The network is trained on the extended dataset D^+; all the augmentation techniques of Table 1 are still disabled.

Experiment 3. The network is again trained on the dataset D^{+}, but the augmentation techniques of Table 1 are applied (with their default values).

RESULTS AND DISCUSSION

Figure 3 shows the plots of the AP computed during the network training in the conditions of the three experiments. Recall that the AP is a rational number between 0 and 1, and that the higher the AP is, the better performances the model has.

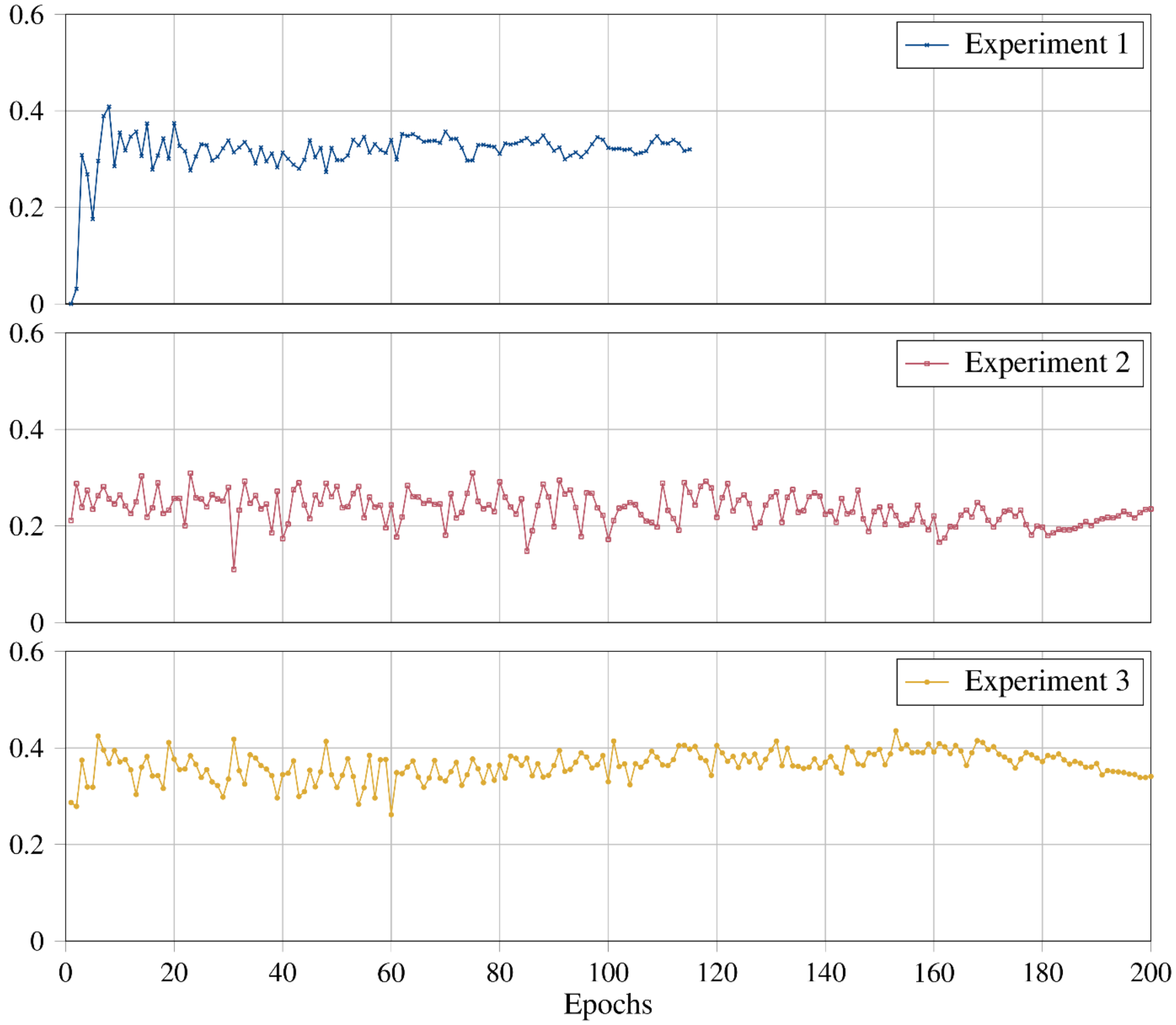

Figure 3. Plots of the AP computed on the validation set during the training phase.

As it is clear from the plots, none of the three experimental settings gives rise to a model that performs exceptionally well on the validation set. There are several potential reasons of this behavior:

- the datasets, especially D_0, contain too few relevant images for the network to learn the discriminating features of *N. norvegicus* entrances;

- the detection task is made more difficult by the nature of the images themselves, as they are extracted from a video recorded in an underwater environment with several drawbacks, such as bad lighting and turbidity; the images quality also is hindered by the technical aspects of the video (low resolution, deinterlacing);
- as stated in the Introduction, detection of *N. norvegicus* entrances is intrinsically difficult even for trained human eyes—it is worth remembering that YOLO is a general-purpose object detector that behaves very well on "everyday" tasks, but it probably needs some tweaks to reach the same performances on highly specialized tasks such as the one described in this paper.

By comparing experiments 1 and 2, one can notice that there is no improvement in the AP, despite the drastic increase in the number of training images. This is probably due to the fact that the new images of D^+ are actually too similar to the ones already present in D_0, so adding them does not in fact provide substantially "new" information that the model can learn—rather, it increases the risk of overfitting.

On the other hand, the AP for experiment 3 is noticeably higher than the one for experiment 2, suggesting that data augmentation techniques of the kind described in Table 1 may be beneficial to the *N. norvegicus* burrows detection task. These and other data augmentation practices are currently being investigated.

CONCLUSION

This work describes the first steps to tackle the problem of detecting *N. norvegicus* burrows in UW video footage using ML techniques. Considering the extraordinary results achieved by deep neural networks in the object detection domain in recent years, it is natural to ask whether they can also be successfully applied in this very specific field. A positive answer will prove extremely useful in helping marine biologists who currently spend a large part of their work time in a tedious and error-prone activity such as manually reviewing hours of video footage to count *N. norvegicus* burrows in it.

The application of an already available general-purpose network, without further tweaking, to a peculiar setting such as images extracted from UW video footage did not obtain particularly good performances. This is not completely unexpected; nevertheless, these results establish a good starting point and give us the motivation to continue pursuing this path of research. As seen above, even a small adjustment like the inclusion of a data augmentation step can improve the overall performances of the model.

Future work on the subject will encompass further attempts with different settings and more ground truth datasets, including the additional development of scripts that can generate large datasets starting from smaller, manually annotated ones. Also, a comparative analysis of the performances of other deep neural network architectures will be carried out, with the objective to eventually reach the best setup for an automatic system able to detect *N. norvegicus* burrow entrances.

ACKNOWLEDGEMENTS

The authors express gratitude to Andrea Belardinelli, Pierluigi Penna and Paolo Scarpini of IRBIM-CNR for their invaluable support in the design and maintenance of the instrumentation, the data recording and its digitalization.

FUNDING

This work is part of a project that has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 101000825 – NAUTILOS Project (https://www.nautilos-h2020.eu).

CONFLICT OF INTERESTS

The authors of this work declare that they have no conflicts of interests.

COMPLIANCE WITH ETHICAL STANDARDS

This work does not contain any studies involving human and animal subjects.

REFERENCES

[1] J. Aguzzi *et al.*, "Burrow emergence rhythms of *Nephrops norvegicus* by UWTV and surveying biases," *Scientific Reports*, vol. 11, article no. 5797, 2021. https://doi.org/10.1038/s41598-021-85240-3

[2] J. Aguzzi *et al.*, "Advancing fishery-independent stock assessments for the Norway lobster (*Nephrops norvegicus*) with new monitoring technologies," *Frontiers in Marine Science*, vol. 9, 2022. https://doi.org/10.3389/fmars.2022.969071

[3] M. C. Bell, F. Redant and I. Tuck, "*Nephrops* Species," in *Lobsters: Biology, Management, Aquaculture and Fisheries*, Ed. by Bruce F. Phillips. Blackwell Publishing Ltd, 2006, ch. 13, p. 412–461. https://doi.org/10.1002/9780470995969.ch13

[4] A. Burguera Burguera *et al.*, "Deep learning for detection and counting of *Nephrops norvegicus* from underwater videos," *ICES Journal of Marine Science*, 2024. https://doi.org/10.1093/icesjms/fsae089

[5] N. Campbell, H. Dobby and N. Bailey, "Investigating and mitigating uncertainties in the assessment of Scottish *Nephrops norvegicus* populations using simulated underwater television data," *ICES Journal of Marine Science*, vol. 66, n. 4, p. 646–655, 2009. https://doi.org/10.1093/icesjms/fsp046

[6] H. Dobby *et al.*, "ICES Survey Protocols – Manual for *Nephrops* Underwater TV Surveys, coordinated under ICES Working Group on *Nephrops* Surveys (WGNEPS)," *ICES Techniques in Marine Environmental Science*, May 2021. https://doi.org/10.17895/ices.pub.8014

[7] FAO Fisheries Division, Statistics and Information Branch, *FishStatJ: Software for Fishery and Aquaculture Statistical Time Series*, accessed on Sept. 18, 2024. https://www.fao.org/fishery/en/statistics/software/fishstatj

[8] ICES, *Report of the Workshop and training course on* Nephrops *burrow identification (WKNEPHBID)*, ICES CM 2008/LRC:03. 25–29 February 2008, Belfast, Northern Ireland, UK, 44 pp., 2008. https://doi.org/10.17895/ices.pub.9828

[9] ICES, *Report of the Workshop on* Nephrops *Burrow Counting (WKNEPS)*, ICES CM 2016/SSGIEOM:34. 9–11 November 2016, Reykjavík, Iceland, 62 pp., 2017. https://doi.org/10.17895/ices.pub.8651

[10] ICES, *Report of the Workshop on* Nephrops *Burrow Counting (WKNEPS)*, ICES CM 2018/EOSG:25. 2–5 October 2018, Aberdeen, UK, 44 pp., 2019. https://doi.org/10.17895/ices.pub.8180

[11] ICES, "Working Group on *Nephrops* Surveys (WGNEPS; outputs from 2021 meeting)," ICES Scientific Reports 4:29, 183 pp., 2022. https://doi.org/10.17895/ices.pub.19438472

[12] ICES, "Working Group on *Nephrops* Surveys (WGNEPS; outputs from 2022 meeting)," ICES Scientific Reports 5:26. 125 pp., 2023. https://doi.org/10.17895/ices.pub.22211161.v1

[13] G. Jocher, A. Chaurasia and J. Qiu, *Ultralytics YOLO*, ver. 8.2.60, 2023. https://github.com/ultralytics/ultralytics

[14] M. Martinelli *et al.*, "Towed underwater television towards the quantification of Norway lobster, squat lobsters and sea pens in the Adriatic Sea," *Acta Adriatica*, vol. 54, n. 1, p. 3–12, 2013.

[15] Microsoft CSE, *Visual Object Tagging Tool (VoTT)*, ver. 2.2.0, 2020. https://github.com/microsoft/VoTT

[16] A. Naseer, "Advanced Detections of Norway Lobster (Nephrops Norvegicus) Burrows using Deep Learning Techniques," *International Journal of Advanced Computer Science and Applications*, vol. 14, n. 3, p. 493–499, 2023. https://doi.org/10.14569/IJACSA.2023.0140357

[17] F. Sardà and J. Aguzzi, "A review of burrow counting as an alternative to other typical methods of assessment of Norway lobster populations," *Reviews in Fish Biology and Fisheries*, vol. 22, p. 409–422, 2012. https://doi.org/10.1007/s11160-011-9242-6

Authors profile *(including Biography and Photo)*

Oscar Papini (Lecco, Italy, 1989), Ph.D. (2018) in Mathematics from the University of Pisa, currently post-doctoral fellow at the Signal and Images Laboratory at ISTI-CNR, Pisa. His main interests are machine learning and mathematical methods for image analysis.

Enrico Cecapolli (Ancona, Italy, 1998), scholarship holder (2023) at IRBIM-CNR of Ancona. His main interests are marine ecology and fisheries science, and in particular the study of the effects of natural and anthropogenic stressors on marine communities and their spatio-temporal dynamics.

Filippo Domenichetti (Ancona, Italy, 1980), M.Sc. (2009) in Biological Sciences, Marine Biology and Oceanography from the University of Ancona, currently technician at IRBIM-CNR, Ancona. His main interest includes fishery, biology, taxonomy and ecology of marine organisms and the structure of fish populations.

Michela Martinelli (Corridonia, Italy, 1979), M.Sc. (2004) in Marine and Oceanographic Biology, Ph.D. (2010) in Marine Biology and Ecology, since 2009 technologist at CNR Ancona. She's responsible for the UWTV program for Norway lobster in the Adriatic Sea since 2010 and of the monitoring of the Fishery Restricted Pomo Area sponsored by MIPAAF from 2016 to 2021. Since 2013 she is responsible also of the AdriFOOS infrastructure and related activities in EU projects.

Gabriele Pieri (Pescia, Italy, 1974), M.Sc. (2000) in Computer Science from the University of Pisa, since 2001 joined the Signals and Images Laboratory at ISTI-CNR, Pisa, working in the field of image analysis. His main interests include neural networks, machine learning, industrial diagnostics and environmental monitoring. He is author of more than twenty papers.

Marco Reggiannini (Viareggio, Italy, 1981), M.Sc. (2008) in Physics and Ph.D. (2016) in Automation Engineering from the University of Pisa, currently technologist at the Signals and Images Laboratory at ISTI-CNR, Pisa. His main interests concern signals and images processing, scene understanding, pattern recognition and machine learning.

Lorenzo Zacchetti (Torino, Italy, 1994), Ph.D. student (2023) in Marine Biology from the University of Bologna and IRBIM-CNR of Ancona (FISHMED-PHD). His main interests are biology and ecology of demersal species and trophic food webs.