Deep Learning-based marine species detection and classification framework for biomonitoring in the Arctic fjords, Svalbard

Felix Mattathil Philip[1*], Gipson Edappazham[1], Anupama Jims[2], Lakshmi Devi Prabhullachandran[1]

[1]*JAIN (Deemed-to-be University) Kochi Campus, Kochi, Kerala, India*
[2]*Chinmaya Vishwa Vidyapeeth (Deemed-to-be University), Kerala, India*

Abstract
The effects of ongoing climate change have caused a poleward shift in the distribution of species due to the rapidly rising water temperatures. This calls for an immediate need to assess and document the extent of climate change-driven animal migrations occurring in the Arctic waters. However, the extreme climatic conditions and the remoteness of the region makes biomonitoring tedious in the Arctic ecosystem. The present study puts forward a deep learning-based analysis of a large underwater video dataset that was captured from the Arctic region. The dataset was acquired using underwater cameras mounted on custom-made stainless-steel frames. The video footages were collected over a period of 26 days from the Kongsfjorden- Krossfjorden twin Arctic fjords in Svalbard, Norway. The collected data sets were used to train YOLO-based object detection framework (You Only Look Once) for an automated detection of the organisms. The YOLO model employed for the study was found to be very efficient in classifying the underwater images captured from the region. The object detection framework could detect images of Comb jelly, Echinoderm, Sea Anemone and Ulke (Shorthorn sculpin) from the underwater images. The model attained a superior value of Mean Average Precision (mAP), precision, and recall of 99.5%, 99.2%, and 97.4%, respectively.

Key words: Arctic, biodiversity, biomonitoring, deep learning, YOLO, climate change, Artificial Intelligence

DOI: 10.5817/CPR2023-2-15

Introduction

The circumpolar Arctic ecosystem covers nearly 13.8 million km^2 land and 14 million km^2 ocean and is characterised by the severity of the climate and its variability in space and time. The Arctic ecosystem, as a whole, tends to experience considerable stress from numerous sources, namely pollution, habitat fragmentation, melting of ice and glaciers owing to climate change, over-exploitation of living resources, and introduction of invasive species (Johnson 2010, Lacoursière-Roussel

---

Received November 21, 2023, *accepted* December 14, 2023.
[*]Corresponding author: F. M. Philip <m.felix@jainuniversity.ac.in>
Acknowledgements: The authors are gratified to JAIN (Deemed-to-be University) for the financial support and all the necessary facilities to undertake the study and special thanks to the Director, National Centre for Polar and Ocean Research (NCPOR), Ministry of Earth Sciences, Goa, for providing the logistic support required for the study.

et al. 2018). The Arctic, as a region, shows the impacts of climate change almost two and half times faster than the rest of the globe and hence considered as an early warning system (Hoegh-Guldberg and Bruno 2010).

The Arctic region encompasses a unique and relatively pristine environment that harbours unique flora and fauna. However, the studies carried over the recent decades show that the region is invaded intensively by new species and many native species are moving towards extinction (*e.g.* Brand and Fischer 2015). Marine ectotherms that thrive in latitudinal ranges based on their thermal tolerance, are reported to diminish at the equatorward boundaries and swell at the polar boundaries owing to global warming (Fossheim et al. 2015). The magnitude as well as the nature of these impacts rely on the adaptive capability and sensitivity of the affected species and this varies for different habitats and species.

The increased warming in the Arctic has resulted in a decrease in the thickness and coverage of the sea ice, and increased the availability of light thus enhancing the population of pelagic primary producers in the Arctic waters and that in turn favours the visual predators (Vinnikov et al. 1999, Arrigo et al. 2008, Ardyna et al. 2014, Varpe et al. 2015, Kahru et al. 2016, Isaksen et al. 2022, Gordó-Vilaseca et al. 2023). This has in turn led to the poleward expansion of fast-swimming fishes thereby enhancing the pelagic production. On the other hand, the species at low trophic levels and those with a narrow range of tolerance and diet preferences tend to respond adversely to climate warming because of increased predation and low diet flexibility (Mueter et al. 2013, Frainer et al. 2017). Therefore, it is crucial to monitor the response of biological communities to climate change considering its ecological and economic importance in an already susceptible area (Layton et al. 2021).

Despite several efforts to monitor and document the Arctic biodiversity trends and issues, information is currently insufficient and available only in piecemeal fashion and on an irregular basis (Laidre et al. 2008). The microbial community, phytoplanktons, zooplanktons, benthos, invertebrates, and vertebrates are some of the primary groupings of Arctic marine biodiversity, and there are significant knowledge gaps regarding their status and general trends.

However, monitoring the biodiversity, community structure and its dynamics in Arctic marine ecosystems remains challenging, owing to its vastness, remoteness and extreme conditions. In addition, the conventional biomonitoring methods are often invasive and resource-intensive. Conventional methods, commonly employed for biomonitoring surveys in the marine ecosystem like benthic grabs, trawl nets, box corers, seine nets, diver surveys have several implications. These methods are predominantly capture-based, can be unsuitable to certain habitats or locations, spatially restrictive, and expensive. They cause disturbances and destruction of habitats (Bicknell et al. 2016). Further, techniques such as Baited Remote Underwater Video Stations (BRUVS), Underwater Visual Census (UVCs) and molecular techniques are time-consuming, labour and cost-intensive (Keith et al. 2015). These approaches seldom fare well for elusive, low-density, highly-mobile fauna like fishes, sharks and rays (Boussarie et al. 2018). In addition, difficulty to detect small, cryptic or elusive species, makes the estimation of entire communities more or less impossible (Deiner et al. 2017). It is often difficult and tedious to detect and document the presence of species that occur in low numbers or are elusive, considering the vast expanse of the Arctic region and the interplay between the terrestrial, freshwater and marine ecosystems. The Arctic Biodiversity Trends Report (Kurvits et al.

2010) emphasises the difficulties encountered in the biomonitoring studies as most of the Arctic countries lack an internal long-term biomonitoring program. Moreover the data tends to be inconsistent across the circumpolar region. The Inari Declaration of The Arctic Council (2002, [1]) recognises the importance of enhanced and elaborate biodiversity monitoring at the circumpolar level in detecting the impacts of global changes on biodiversity and to enable the Arctic communities to effectively respond and adapt to the changes.

Advanced technology equipped with Artificial Intelligence (AI) using deep learning methods facilitates the recognition of a diverse range of species, belonging to different habitats, and hence a promising and effective tool in biomonitoring studies. Video-based species detection approach may help in easy identification of the species, its taxonomic classification and documentation without harming the ecosystem, its components and functioning. Camera imagery has emerged as a potent tool in biomonitoring studies at all scales, from individuals to populations and communities up to entire ecosystems (Bicknell et al. 2016). The relative ease of handling and ever decreasing cost of cameras enables them to be employed in applied and theoretical research, behavioural studies, species interactions, their adaptations and responses, community assemblages, ecosystem functioning and resilience.

Since several decades, camera traps have been widely used in terrestrial ecosystems to assess the abundance, species diversity, behavioural studies and documentation of rare species (Burton et al. 2015). The concept has been, however, developed for marine ecosystems very recently. The production of quality, high-definition waterproof cameras marks the development of imagery as a potent tool for marine biodiversity studies. The affordability, reduced size, improved underwater housing along with extended power back up and storage capacity will undoubtedly add up to the application of camera imagery in every habitat, including the areas that were previously unfeasible or inaccessible.

The present study is an AI-based analysis of a large underwater video dataset that was captured using underwater cameras from the Arctic region. The video footages collected from Kongsfjorden-Krossfjorden twin Arctic fjords in Svalbard, Norway as part of the Summer Indian Arctic Expedition, were analysed using You Only Look Once (YOLO) real-time object detection framework, for extending it to further theoretical, applied and correlative studies.

Material and Methods

The objective of the present study was to design and develop a Deep Learning-based automatic framework to monitor the biodiversity and community assemblages in the Arctic fjords. This framework can be used for the documentation of climate-change driven animal migrations and subsequent occurrences of invasive species in the Arctic fjords.

The study consisted of several consequtive phases, namely image collection, pre-processing, frame extraction, annotation, training and validation of the Deep Learning model, detection of the organism and type classification. Once properly classified, the images can be added to appropriate databases for documentation and further studies. The general system architecture of the study is depicted in Fig. 1.

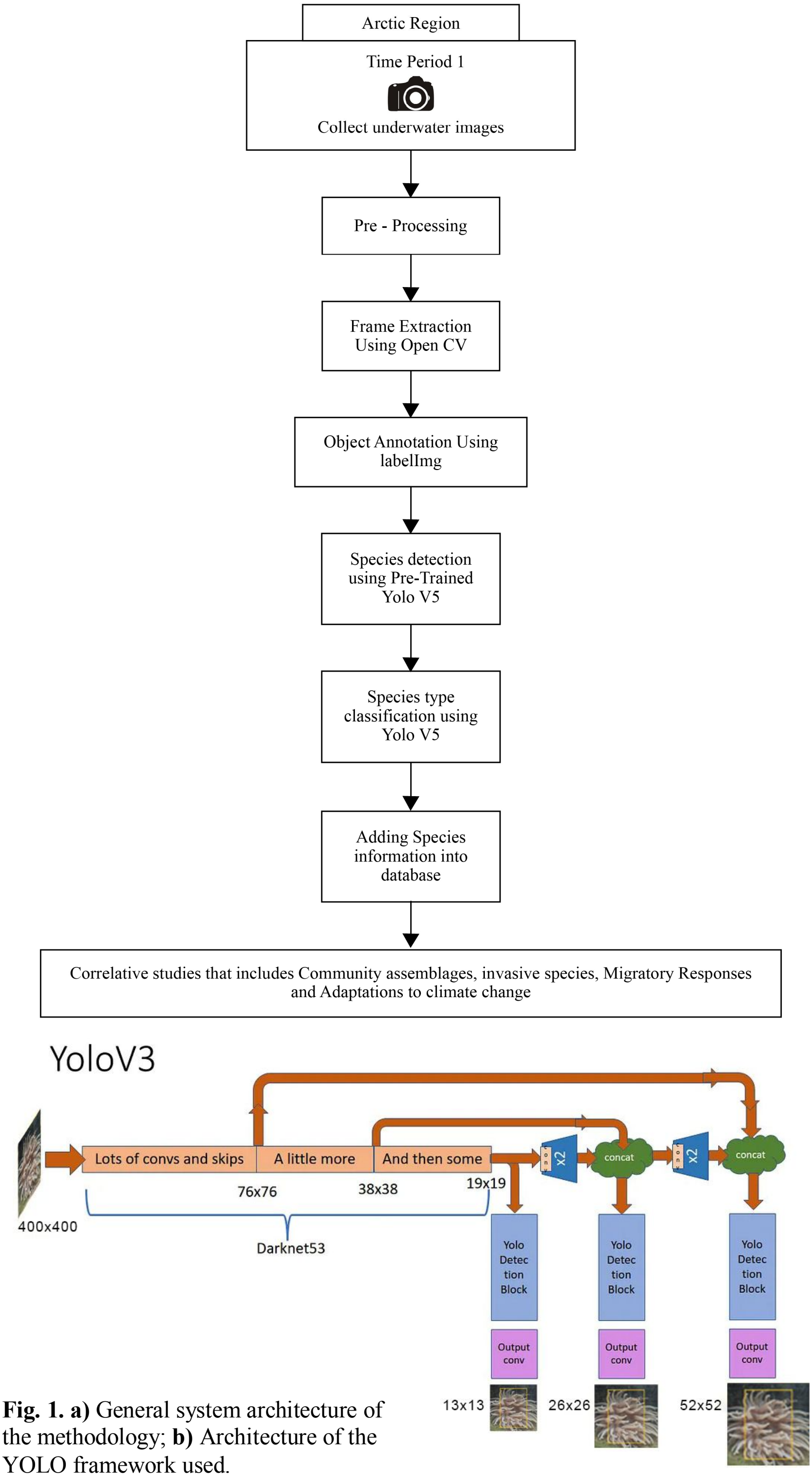

Fig. 1. a) General system architecture of the methodology; **b)** Architecture of the YOLO framework used.

Study area and sampling spots

The study was carried out in the Kongfjorden-Krossfjorden twin glacial fjords, Svalbard as a part of Summer Indian Arctic Expedition held from May to June 2023 at the Indian Research Base, 'Himadri', located at Spitsbergen, Svalbard, Norway (78°55′ N, 11°56′ E). Considering the accessibility and prevailing weather conditions, a total of five spots (four spots from Kongsfjorden and one spot from Krossfjorden) were selected for the study (Fig. 2). The locations of the selected spots are given in Table 1.

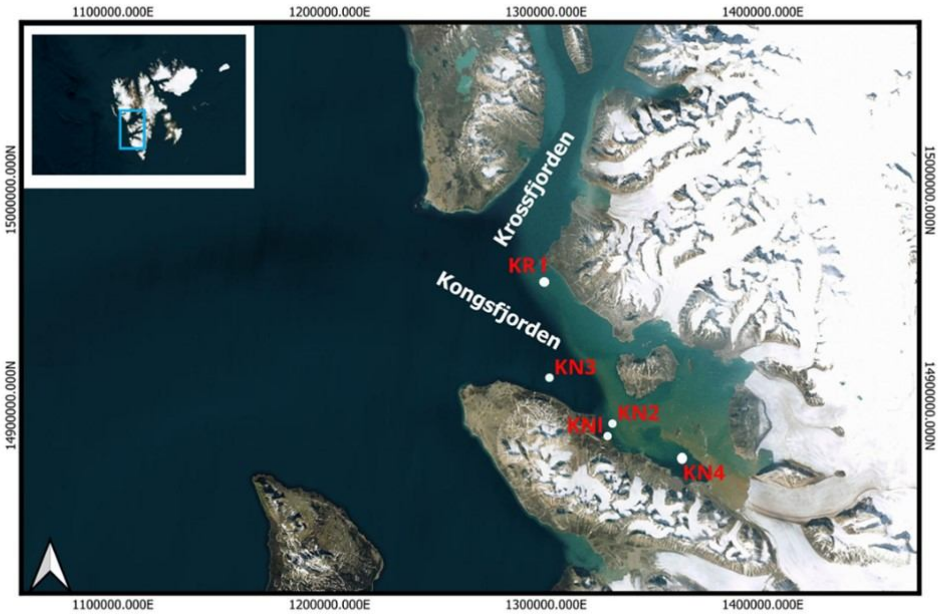

Fig. 2. Study area with sampling spots.

Spot ID	Latitude (°N)	Longitude (°E)
KNl	78° 55' 71.34" N	11° 55' 99.85" E
KN2	78° 56' 31.80" N	11° 57' 22.80" E
KN3	78° 58' 51.09" N	11° 41' 53.26" E
KN4	78° 90' 68.35" N	12° 26' 77.68" E
KR1	79° 06' 38.74" N	11° 65' 23.06" E

Table 1. Sampling spots and geographical co-ordinates.

Image collection

Underwater video footage from the selected spots were captured using underwater cameras mounted on custom-made stainless-steel box shaped frames of 50 cm x 50cm x 50cm size. The frames were fitted with GoPro HERO11 Black action cameras housed in a waterproof case (Fig. 3). A total of four such frames were set at a water depth ranging from 10 to 20 metres in the selected spots with the help of suitable floats and sinkers. These frames were retrieved after a soak time of 2 h. The deployment and the retrieval operations were carried out from the workboat MS Teisten. After the retrieval of the cameras from the fjord, the footages were downloaded and the frames were redeployed as per the schedule. The underwater footages were captured over a period of 26 days. The location details and the driving environmental parameters, *i.e.* salinity, temperature and depth were also documented before each deployment of the frames.

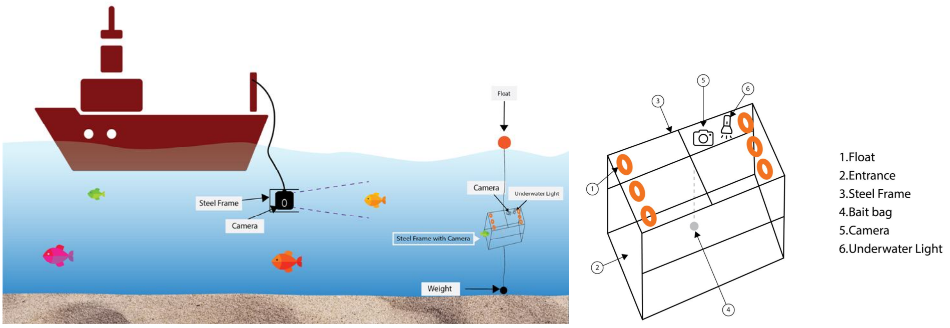

Fig. 3. Experimental setup showing the custom-made steel frame mounted with underwater cameras.

Pre-processing and frame extraction

After downloading the underwater footage from the cameras, it was pre-processed and made ready for frame extraction. The image frames from the videos were extracted using OpenCV Python library[2] and image segregation was done manually.

Sample input images

The sample images obtained from the video dataset are presented below (Fig. 4). These images were extracted from the videos and are composed of Sea Anemone, Shorthorn sculpin/Ulke, Comb Jelly (Ctenophore), Echinoderm, Jellyfish and Sea Slug in the frame.

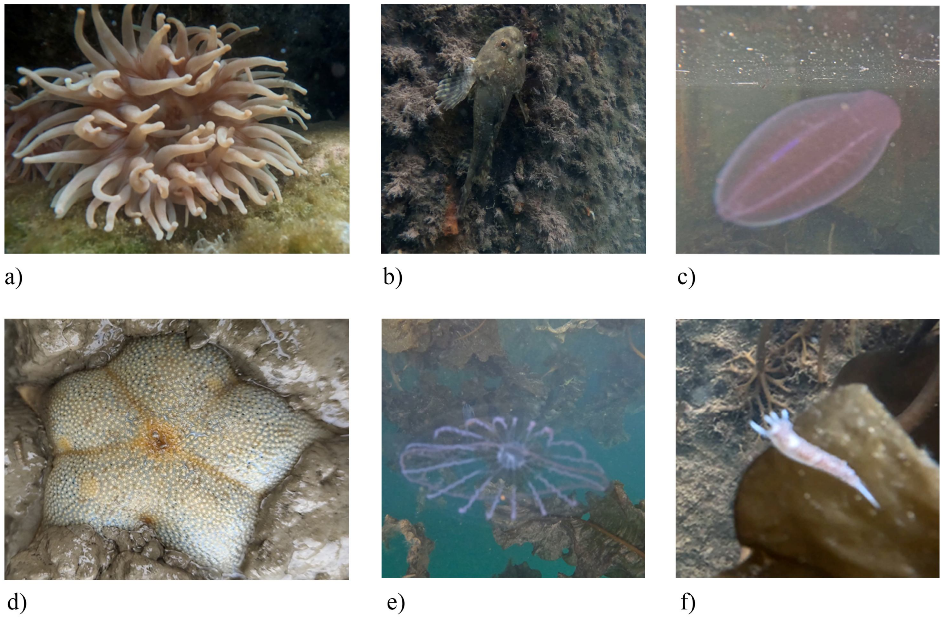

Fig. 4. Sample images extracted from the video dataset corresponding to: **a)** Sea Anemone, **b)** Ulke (*Myoxocephalus sp.*), **c)** Comb jelly (Ctenophore), **d)** Echinoderm, **e)** Jellyfish and **f)** Sea Slug.

Dataset annotations

Data annotation is one of the important tasks in the deep learning (supervised learning) based object detections applications. It is important in establishing the relationships between inputs and its corresponding outputs of the deep learning model. In this object detection task, the bounding box method was employed to annotate the objects in the images. The images of the organisms that were vivid enough to be identified were selected from the whole set of extracted images for annotation using the tool LabelImg[3]. This is a process of manually annotating/labelling around the objects in the image, and also specifying the class details. Five values were generated for a single object, including the co-ordinates x1,y1 (top left corner of the bounding box of the object), x2,y2 (lower right corner) and its class value (like 0,1,2 *etc*. in an encoded form). In this manner, the annotations of all the images as text files were generated. These annotation text files were used as the '*ground truth*' values for the input images.

There were a total of 7 classes each of which had around 2000 images, thus constituting a total of 14000 images. From this total pool of extracted images, 70% images were randomly selected and kept as training set, 20% were kept for validation and the rest 10% were kept for testing the model.

Model training and evaluations

These annotated images were used to train the deep learning model, the YOLO v5 (Redmon et al. 2016). In the present study, Tensorflow technology with python language was employed for training the model[4]. The YOLO v5 pretrained model was used to learn maximum discriminating features from the training dataset. Also, custom-created anchor boxes obtained by clustering of shapes from the training data set bounding boxes were used for getting highest detection accuracy with reduced false positives.

During the model training, weights of kernels were updated in order to reduce the loss function value. The loss function had to be defined initially for effective training. The loss function which is derived from the YOLO loss function, comprised of the following parameter: a) Box Loss (Coordinate loss) – caused when an object is not completely covered by the box, b) Object loss – occurs when the Intersection over Union (IoU) prediction box and object is wrong, c) Class loss (Classification loss) – occurs because of the variations while forecasting '1' for the correct classes and '0' for remaining classes for object in the box, and d) a special loss, which is estimated by considering the objection and contraction loss.

$$
\begin{aligned}
Loss = {} & \alpha_{cord} \sum_{j=0}^{A^2} \sum_{k=0}^{B} 1_{jk}^{obj} \left[\left(p_j - p_j'\right)^2 + \left(q_j - q_j'\right)^2 \right] + \\
& \alpha_{cord} \sum_{j=0}^{A^2} \sum_{k=0}^{B} 1_{jk}^{obj} \left[\left(\sqrt{a_j} - \sqrt{a_j'}\right)^2 + \left(\sqrt{b_j} - \sqrt{b_j'}\right)^2 \right] + \\
& \sum_{j=0}^{A^2} \sum_{k=0}^{B} 1_{jk}^{obj} \left(h_j - h_j'\right)^2 + \sum_{j=0}^{A^2} \sum_{k=0}^{B} 1_{jk}^{nobj} \left(h_j - h_j'\right)^2 + \\
& \sum_{j=0}^{A^2} 1_{jk}^{nobj} \sum_{ceclasses} \left(a_j(h) - a_j'(h)\right)^2
\end{aligned}
$$

Special loss function (Redmon et al. 2016).

Here, α indicates the loss coefficients. The initial three terms specify the loss occurring due to the best boxed and the last two terms signify the loss due to the boxes which have not captured any objects.

The model was trained for 200 epochs using the training set images and continuously validated at the end of each epoch using the validation set. Best model with highest validation accuracy was saved for further testing and analysis. The saved model was then evaluated using the test set images. The evaluation metrics used are Mean Average Precision (mAP), Precision and Recall.

Results

The perfectly trained model was evaluated using the test data and the results were analysed. The training curves including the mAP and loss curves are visualised in Figs. 5 and 6. Further, the box loss, class loss and object loss are depicted in Fig. 6.

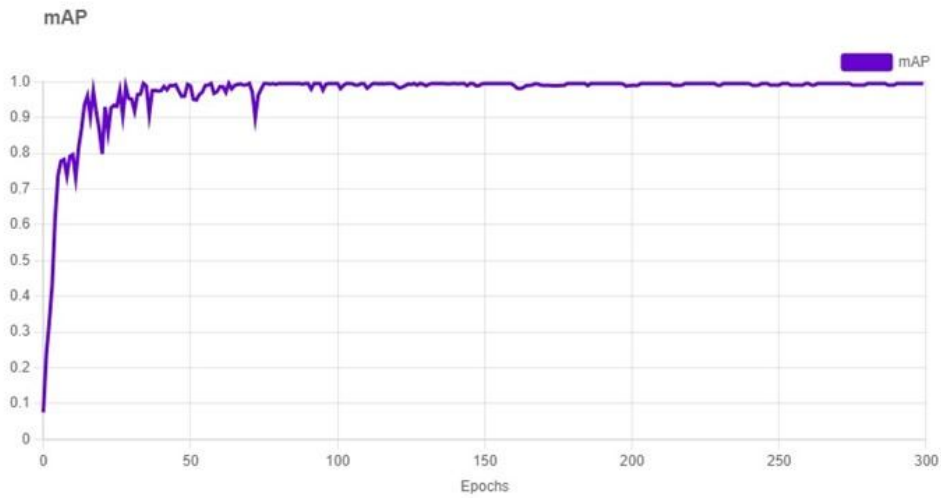

Fig. 5. Training of the YOLO architecture.

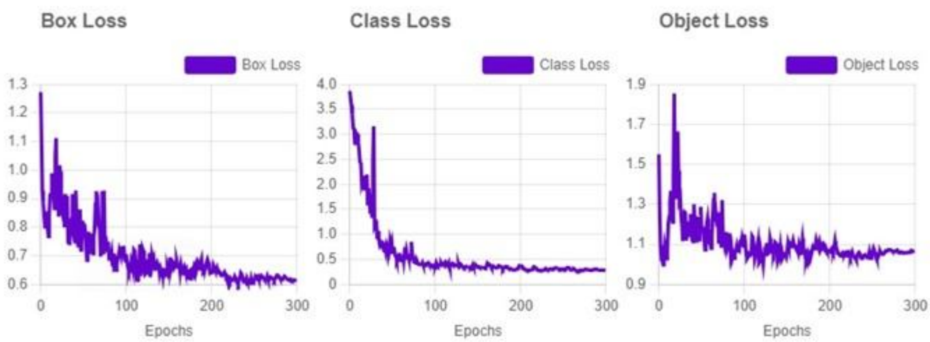

Fig. 6. Loss functions.

The captured video footage contained videos of Sea Anemone (11 distinct sightings), Ulke (*Myoxocephalus* sp.) (7 distinct sightings), Comb jelly (30 distinct sightings), Jellyfish (1 distinct sighting), Amphipod (45 distinct sightings), Echinoderm (1 distinct sighting) and Sea Slug (1 distinct sighting). Images were extracted from this video data and were subsequently used for training, validation and testing of the model as described above. Visualisation of some of the predicted images are shown in Fig 7. The Comb jelly/ Ctenophore was identified and localised in Fig. 7a, with an accuracy of 0.0956, Fig. 7b shows an Echinoderm and the classification accuracy is 0.935, Sea Anemone is depicted in Fig. 7c with a score of 0.926, and in Fig. 7d, Ulke/Shorthorn sculpin (*Myoxocephalus sp.*) is presented and the accuracy attained is 0.903.

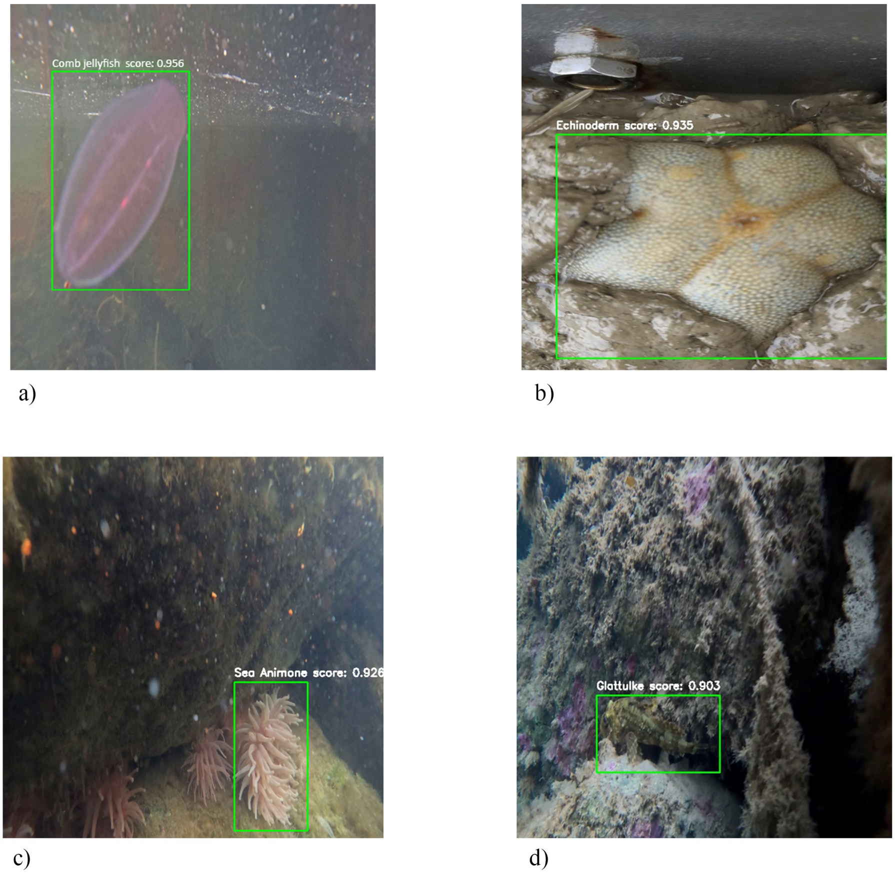

Fig. 7. Experimental results showing the accuracy and class labels **a**) Comb Jelly, **b**) Echinoderm, **c**) Sea Anemone, and **d**) Ulke (*Myoxocephalus sp.*).

The precision, recall and mAP values obtained for the test dataset are depicted in Table 2. The developed model attained good object detection abilities with a high value of mAP at 99.5%, precision of 99.2%, and recall of 97.4%.

mAP	**Precision**	**Recall**
99.5%	99.2%	97.4%

Table 2. Experimental results. *Note*: mAP – mean Average Precision.

Discussion

The vast expanse, remoteness and the extreme climatic conditions make biomonitoring studies in the Arctic region really challenging. In addition, the conventional biomonitoring methods are often invasive and resource-intensive as it involves manual intervention. It is often difficult and tedious to detect and document the presence of species that occur in low numbers or are elusive, considering the vast coverage of the Arctic region and the interplay between the terrestrial, freshwater and marine ecosystems. Furthermore, commonly employed techniques like biometric studies, tagging, use of biomarkers, and molecular techniques are time-consuming, laborious, cost-intensive and invasive in nature. Considering the high labour costs and delays in achieving the outcomes in manual sampling, there is an increasing trend of switching to non-destructive and automatic ways for sampling and data collection (McLaren et al. 2015).

There are several biomonitoring approaches that employ non-destructive, noninterventional and automatic species detection and classification, for instance underwater video capture (Shortis and Abdo 2016, Jalal et al. 2020). Bicknell et al. (2016) reported the advantages of camera technologies in biomonitoring studies over the traditional techniques that employs benthic grabs, fish trawls, box corers, or diver surveys, which are destructive and disturbing to the ecosystem, unsuitable to certain locations, spatially restrictive and prohibitively costly. Video-based species detection approach provides a better understanding of the habitats, the organisms thriving in these habitats and their responses to human activities, especially in regions of limited accessibility like the Arctic.

Camera technology equipped with Artificial Intelligence (AI) using deep learning methods facilitates the recognition of a diverse range of species, belonging to different habitats, and hence a promising and effective tool in biomonitoring studies. Jiang and Learned-Miller (2017) and Banan et al. (2020) reported the efficiency of deep learning methods in rapid, accurate visual recognition and simplified classification of fish species. Several deep learning models have been successfully employed for species recognition and their classification recently, which includes VGGNet (Simonyan and Zisserman 2014, Fu et al. 2018); GoogLeNet (Szegedy et al. 2015, Tian et al. 2018, Khan et al. 2019); Alexnet (Lu et al. 2019, Ju and Xue 2020), ResNet (Mahmood et al. 2020) and the YOLO (Redmon et al. 2016, Lathifah et al. 2020, Knausgard et al. 2022, Hentati-Sundberg et al. 2023).

The YOLO is a state-of-art, real time object detection system that can detect over 9000 object categories with better precision and accuracy (Redmon et al. 2016). Liu et al. (2018) presented an online fish tracking using the YOLO and parallel correlation filters and included detection and categorization in an end-to-end approach. Similar study was carried out by Xu and Matzner (2018), wherein the YOLO architecture was trained to detect a variety of fish species with three very different data sets, obtaining a mean average precision score of 0.5392. Pedersen et al. (2019) established the application and efficacy of the YOLO framework in the detection of marine animals from underwater footage. In their study, a new bounding box annotated image data set of marine animals from temperate brackish waters was considered for training using YOLOv2 and YOLOv3. From the data set of 14,518 frames with 25,613 annotations of six classes of marine fauna, *i.e.*, Big fish, Small fish, crab, jelly fish, shrimp and star fish, the YOLOv3 network achieved the best performance with $AP_{50} \approx 84\%$ (Average Precision at 50% IoU). Lathifah et al. (2020) in their study developed a fish species classification system using the YOLO architecture using the underwater video data set collected from Indonesian waters. The detection and classification of temperate fish species was undertaken using the YOLO object detection technique with an accuracy of 99.27% by Knausgard et al. (2022). They have used public dataset (Fish4Knowledge) to train the object detection model and used underwater video data collected from different locations with depths ranging between 1 to 40 meters for fish detection and classification. In our study, the underwater video data were collected from water depths ranging between 10 to 20 metres.

Świeżewski[5] employed the YOLO framework to automate the counting of Antarctic Cormorant (also known as Antarctic shag *Leucocarbo bransfieldensis*) nests in drone imagery to assess the wellbeing of the Antarctic ecosystem[5]. In a recent study, Hentati-Sundberg et al. (2023) reported the application of YOLOv5 in the surveillance and documentation of sea birds, where a system of video surveillance using CCTV footage combined with automated image processing was developed for the monitoring of Common Murres (*Uria aalge*). The system employed the deep learning algorithm the YOLOv5 for object detection, that had been trained on annotated images of the adult birds, chicks and eggs and outputs time, location, size and confidence level of all detections frame-by-frame, in the supplied video material. In their study, the precision (P) and Recall (R) of the YOLOv5-medium-960 model were 0.91 and 0.79, with an F1 score of 0.85, over all classes. The model performed better for adults than for chicks and eggs, with a P of 0.98, 0.84 and 0. 92 and R of 0.98, 0.74 and 0.64 respectively. In our study, the YOLO model was trained using our custom dataset and the anchor boxes were estimated using the K-Means clustering technique with IoU as similarity measure. During the training of the model, the loss functions namely Box loss, Object loss and Class loss came down significantly as the training progressed through the epochs. As a result, superior value of mAP (99.5%), precision (99.2%), and recall (97.4%) could be achieved.

Conclusion

The article presents advantages of using artificial intelligence based biomonitoring systems for monitoring the Arctic ecosystem. In this study, a video dataset acquired from the Arctic region was used to develop the model. The dataset collected is composed of videos that were collected using underwater cameras over a period of 26 days across the different fjords in Norway. The details of the methodology adopted and analysis carried out are discussed. The performance of the YOLOv5 model in classifying the images is also assessed. The YOLOv5 is found to attain a superior value of mAP, precision, and recall of 99.5%, 99.2%, and 97.4%, respectively. The future direction and scope of the research includes the development of more sophisticated techniques to improve the classification performance of the deep learning framework. This may involve use of upgraded version of the image capture system and use of latest versions of YOLO and other image detection algorithms.

References

ARDYNA, M., BABIN, M., GOSSELIN, M., DEVRED, E., RAINVILLE, L. and TREMBLAY, J.-É. (2014): Recent Arctic Ocean sea ice loss triggers novel fall phytoplankton blooms. *Geophysical Research Letters*, 41: 6207-6212. doi: 10.1002/2014GL061047

ARRIGO, K. R., VAN DIJKEN, G. and PABI, S. (2008): Impact of a shrinking Arctic ice cover on marine primary production. *Geophysical Research Letters*, 35: L19603. doi: 10.1029/2008GL035028

BANAN, A., NASIRI, A. and TAHERI-GARAVAND, A. (2020): Deep learning-based appearance features extraction for automated carp species identification. *Aquacultural Engineering*, 89: 102053.

BICKNELL, A. W. J., GODLEY, B. J., SHEEHAN, E. V., VOTIER, S. C. and WITT, M. J. (2016): Camera technology for monitoring marine biodiversity and human impact. *Frontiers in Ecology and the Environment*, 14: 424-432.

BRAND, M., FISCHER, P. (2015): Species composition and abundance of the shallow water fish community of Kongsfjorden, Svalbard. *Polar Biology*, 39: 2155-2167. doi: 10.1007/s00300-016-2022-y

BOUSSARIE, G., BAKKER, J., WANGENSTEEN, O. S., MARIANI, S., BONNIN, L., JUHEL, J. B., KISZKA, J. J., KULBICKI, M., MANEL, S., ROBBINS, W. D., VIGLIOLA, L. and MOUILLOT, D. (2018): Environmental DNA illuminates the dark diversity of sharks. *Science Advances*, 4(5): eaap9661. doi: 10.1126/sciadv.aap9661

BURTON, A. C., NEILSON, E., MOREIRA, D., LADLE, A., STEENWEG, R., FISHER, J. T., BAYNE, E. and BOUTIN, S. (2015): Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, 52: 675-685.

DEINER, K., BIK, H. M., MÄCHLER, E., SEYMOUR, M., LACOURSIÈRE-ROUSSEL, A., ALTERMATT, F., CREER, S., BISTA, I., LODGE, D. M., DE VERE, N., PFRENDER, M. E. and BERNATCHEZ, L. (2017): Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26(21): 5872-5895. doi: 10.1111/mec.14350

FOSSHEIM, M., PRIMICERIO, R., JOHANNESEN, E., INGVALDSEN, R. B., ASCHAN, M. M. and DOLGOV, A.V. (2015): Recent warming leads to a rapid borealization of fish communities in the Arctic. *Nature Climate Change*, 5: 673-677. doi: 10.1038/nclimate2647

FRAINER, A., PRIMICERIO, R., KORTSCH, S., AUNE, M., DOLGOV, A.V., FOSSHEIM, M. and ASCHAN, M. M. (2017): Climate driven changes in functional biogeography of Arctic marine fish communities. *Proceedings of the National Academy of Sciences of the United States of America*, 114(46): 12202-12207. doi: 10.1073/pnas.1706080114

FU, R., LI, B., GAO, Y. and WANG, P. (2018): Visualizing and analyzing convolution neural networks with gradient information. *Neurocomputing*, 293: 12-17.

GORDÓ-VILASECA, C., STEPHENSON, F., COLL, M., LAVIN, C. and COSTELLO, M. J. (2023) Three decades of increasing fish biodiversity across the northeast Atlantic and the Arctic Ocean. *Proceedings of the National Academy of Sciences of the United States of America*, 120(4): e2120869120. doi: 10.1073/pnas.2120869120

HENTATI-SUNDBERG, J., OLIN, B. A., REDDY, S., BERGLUND, P., SVENSSON, E., REDDY, M., KASARARENI, S., CARLSEN, A. A., HANES, M., KAD, S. and OLSSON, O. (2023): Seabird surveillance: combining CCTV and artificial intelligence for monitoring and research. *Remote sensing in Ecology and Conservation*, 9(4): 568-581. doi: 10.1002/rse2.329

HOEGH-GULDBERG, O., BRUNO, J. F. (2010) The impact of climate change on the world's marine ecosystems. *Science*, 328: 1523-1528. doi: 10.1126/science.1189930

ISAKSEN, K., NORDLI, Ø., IVANOV, B., KØLTZOW, M. A. Ø., AABOE, S., GJELTEN, H. M., MEZGHANI, A., EASTWOOD, S., FØRLAND, E., BENESTAD, R. E., HANSSEN-BAUER, I., BRÆKKAN, R., SVIASHCHENNIKOV, P., DEMIN, V., REVINA, A. and KARANDASHEVA, T. (2022): Exceptional warming over the Barents area. *Scientific Reports,* 12: 9371. doi: 10.1038/s41598-022-13568-5

JALAL, A., SALMAN, A., MIAN, A., SHORTIS, M. and SHAFAIT, F. (2020): Fish detection and species classification in underwater environments using deep learning with temporal information. *Ecological Informatics*, 57: 101088.

JIANG, H., LEARNED-MILLER, E. (2017): Face detection with the faster R-CNN. *In: 12th IEEE International Conference on Automatic Face and Gesture Recognition* (*FG* 2017). IEEE, pp. 650–657.

JOHSON, L. (2010): The fearful symmetry of Arctic climate change: Accumulation by degradation. Environment and Planning D. *Society and Space*, 28(5): 828-847. doi: 10.1068/ d9308

JU, Z., XUE, Y. (2020): Fish species recognition using an improved AlexNet model. *Optik*, 223: 165499.

KAHRU, M., LEE Z., MITCHELL, B. G. and NEVISON, C. D. (2016): Effects of sea ice cover on satellite-detected primary production in the Arctic Ocean. *Biology Letters*, 12(11): 20160223. doi: 10.1098/rsbl.2016.0223

KEITH, D., AKÇAKAYA, H. R., BUTCHART, S. H. M., COLLEN, B., DULVY, N. K., HOLMES, E. E., HUTCHINGS, J. A., KEINATH, D., SCHWARTZ, M. K., SHELTON, A. O. and WAPLES, R. S. (2015): Temporal correlations in population trends: conservation implications from time-series analysis of diverse animal taxa. *Biological Conservation*, 192: 247-257. doi: 10.1016/j.biocon.2015.09.021

KHAN, S., ISLAM, N., JAN, Z., DIN, I. U., JOEL, J. and RODRIGUES, P. C. (2019): A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognition Letters*, 125: 1-6.

KNAUSGARD, M. K., WIKLUND, A., SØRDALEN, T. K. and HALVORSEN, K. T. (2022): Temperate fish detection and classification: a deep learning based approach. *Applied Intelligence*, 52: 6988-7001. doi: 10.1007/s10489-020-02154-9

KURVITS, T., ALFTHAN, B. and MORK, E. (2010): Conservation of Arctic flora and fauna. *In:* Arctic biodiversity trends 2010: Selected indicators of change. https://wedocs.unep.org/20.500.11822/8670.

LACOURSIÈRE-ROUSSEL, A., HOWLAND, K., NORMANDEAU, E., GREY, K. E. and ARCHAMBAULT, P. (2018): eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. *Ecology and Evolution*, 8(16): 7763-7777. doi: 10.1002/ece3.4213

LATHIFAH, H. M, NOVAMIZANTI, L. and RIZAL, S. (2020): Fast and accurate fish classification from underwater video using You Only Look Once. *IOP Conference Series: Materials Science and Engineering*, 982: 012003. doi: 10.1088/1757-899X/982/1/012003

LAYTON, K. K. S., SNELGROVE, P. V. R., DEMPSON, J. B., KESS, T., LEHNERT, S. J., BENTZEN, P., DUFFY, S. J., MESSMER, A. M., STANLEY, R. R. E., DIBACCO, C. and SALISBURY, S. J. (2021): Genomic evidence of past and future climate-linked loss in a migratory Arctic fish. *Nature Climate Change*, 11(2): 158-165.

LAIDRE, K. L., STIRLING, I., LOWRY, L. F., WIIG, Ø., HEIDE-JØRGENSEN, M. P. and FERGUSON, S. H. (2008): Quantifying the sensitivity of Arctic marine mammals to climate – induced habitat change. *Ecological Applications*, 18(sp2): S97-S125.

LIU, S., QI, L., QIN, H., SHI, J. and JIA, J. (2018): Path aggregation network for instance segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8759–8768. https://ieeexplore.ieee.org/document/8579011

LU, S., LU, Z. and ZHANG, Y. (2019): Pathological brain detection based on AlexNet and transfer learning. *Journal of Computer Science*, 30: 41-47.

MAHMOOD, A., BENNAMOUN, M. A. S., SOHEL, F. and BOUSSAID, F. (2020): ResFeats: Residual network based features for underwater image classification. *Image and Vision Computing*, 93: 225-236.

MCLAREN, B. W., LANGLOIS, T. J., HARVEY, E. S., SHORTLAND-JONES, H. and STEVENS, R. (2015): A small no-take marine sanctuary provides consistent protection for small-bodied by-catch species, but not for large-bodied, high-risk species. *Journal of Experimental Marine Biology and Ecology*, 471: 153-163.

MUETER F. J., REIST, J. D., MAJEWSKI, A. R., SAWATZKY, C. D., CHRISTIANSEN, J. S. and HEDGES, K. J. (2013): Marine fishes of the Arctic, *In:* Arctic Report Card: Update for 2013–Tracking Recent Environmental Changes. http://www.arctic.noaa.gov/reportcard

PEDERSEN, M., HAURUM, J. B., GADE, R. and MOESLUND, T. B. (2019): Detection of marine animals in a new underwater dataset with varying visibility. *In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 18–26).

REDMON, J., DIVVALA, S., GIRSHICK, R. and FARHADI, A. (2016): You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Las Vegas, NV, USA, 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91

SHORTIS, M., ABDO, E. H. D. (2016): A review of underwater stereo-image measurement for marine biology and ecology applications. *In: Oceanography and Marine Biology. CRC Press*, pp. 269–304.

SIMONYAN, K., ZISSERMAN, A. (2014): Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv*: 1409-1556.

SZEGEDY, C., LIU, W., JIA, Y., SERMANET, P., REED, S., ANGUELOV, D. and ERHAN, D. (2015): Going deeper with convolutions. *Proceedings with the IEEE Conference on Computer vision and Pattern recognition*, pp. 1–9. doi: 10.1109/CVPR.2015.7298594

TIAN, Q., ARBEL, T. and CLARK, J. J. (2018): Structured deep Fisher pruning for efficient facial trait classification. *Image and Vision Computing*, 77: 45-59.

VARPE, Ø., DAASE, M. and KRISTIANSEN,T. (2015): A fish-eye view on the new Arctic lightscape. *ICES Journal of Marine Science*, 72: 2532-2538. doi: 10.1093/icesjms/fsv129

VINNIKOV, K. Y., ROBOCK, A., STOUFFER, R. J. WALSCH, J. E., PARKINSON, C. L., CAVALIERI, D. J., MITCHELL, J. F., GARRETT, D. and ZAKHAROV, V. F. (1999): Global warming and Northern hemisphere sea ice extent. *Science*, 286(5446): 1934-1937. doi: 10.1126/science.286.5446.193

XU, W., MATZNER, S. (2018): Underwater fish detection using deep learning for water power applications. *In: 2018 International conference on computational science and computational intelligence* (*CSCI*), pp. 313–318.

Web sources / Other sources

[1] Arctic Council (2002) Inari Declaration - on the occasion of the Third Ministerial Meeting of the Arctic Council, Inari, Finland, 2002. http://hdl.handle.net/11374/88 (Accessed on 03-Oct-2023).

[2] OpenCV Python library, https://github.com/opencv/opencv-python (Accessed on 21-Aug-2023).

[3] LabelImg, https://github.com/HumanSignal/labelImg (Accessed on 11-Aug-2023).

[4] TensorFlow, https://www.tensorflow.org (Accessed on 23-Aug-2023).

[5] Świeżewski, J. (2022): Counting nests of shags with YOLO to assess the wellbeing of the Antarctic ecosystem, https://appsilon.com/yolo-counting-nests-antarctic-birds/ (Accessed on 05-Oct-2023).