

Computer Vision Pipeline for Automated Antarctic Krill Analysis.

Mazvydas Gudelis¹
M.Gudelis@uea.ac.uk

Michal Mackiewicz¹
M.Mackiewicz@uea.ac.uk

Julie Bremner^{1,3}
J.Bremner@uea.ac.uk

Sophie Fielding^{1,2}
sophie.fielding@uea.ac.uk

¹ University of East Anglia
Norwich, UK

² British Antarctic Survey
Cambridge, UK

³ Centre for Environment, Fisheries and
Aquaculture Science
Lowestoft, UK

Abstract

British Antarctic Survey (BAS) researchers launch annual expeditions to the Antarctic in order to estimate Antarctic Krill biomass and assess the change from previous years. These comparisons provide insight into the effects of the current environment on this key component of the marine food chain. In this work we have developed tools for automating the data collection and analysis process, using web-based image annotation tools and deep learning image classification and regression models. We achieve highly accurate krill instance segmentation results with an average 77.28% AP score, as well as separate maturity stage and length estimation of krill specimens with 62.99% accuracy and a 1.98 mm length error respectively.

1 Introduction

This paper discusses the continuing research and development of computer vision tools that are to be used by British Antarctic Survey (BAS) researchers to process the Antarctic Krill image data. Antarctic Krill (*Euphausia superba*) are small, shrimp-like organisms that travel in dense swarms that may stretch for several kilometres and include millions of individuals. They are consumed by many marine species in the Southern Ocean, including whales, seals, and penguins, and serve a crucial role in the food chain. Krill feed on summer-blooming phytoplankton, that are reducing in number as a consequence of global warming. Antarctic Krill can adapt and are known to shrink over the winter without sufficient nourishment. Since the 1970s, krill populations have declined by 80%, which has been linked to the melting ice that is home to the algae and plankton upon which they feed [1, 2]. In order to measure the krill biomass and evaluate the variation from year to year, annual journeys are taken to different locations around the Antarctic.

In this work we developed tools to automate the data collection and analysis of krill specimens captured in the Antarctic, using web-based image annotation tools and deep learning

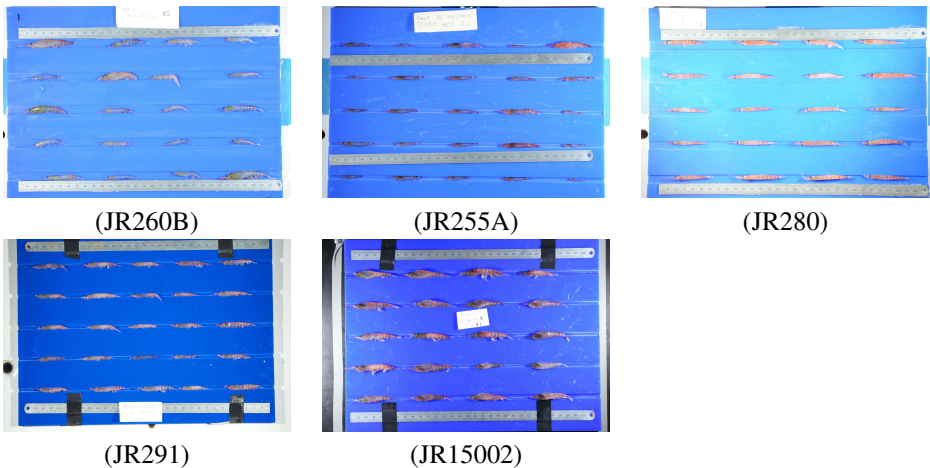


Figure 1: Sample of images from 5 different cruises around the Antarctic with corresponding naming encoding.

computer vision models. Specifically, we focus on object detection (krill specimen instance segmentation) and the automation of data processing procedures in order to transform krill image data into the form used to train deep neural networks capable of estimating two key parameters of the krill body, which are of interest to the marine biologists: its length and maturity stage.

In Section 2 we cover background and related research. We further talk about the data-set we work with, methods employed, and our results in Section 3. We conclude in Section 4.

2 Background

Samples of Antarctic Krill are taken to evaluate the diversity of the population structure around different parts of the Antarctic and to provide parameters for the target strength model used to estimate krill biomass. Specimen length and maturity stage are the 2 main parameters that are to be inferred and both are normally assessed using long-established manual methods [14, 15]. The analysis is done manually for chosen specimens by physically examining their characteristics; images are only taken as digital proof of them. Taking into account the priors used to collect the data visible in Figure 1, we suggest exploiting them to produce a consistent and user-friendly computer vision pipeline that not only exports pre-processed data for model training, but also facilitates an environment where such models can be served and used as automatic analysis tools [9].

The domain of object detection currently has a strong standpoint with many open source data-sets and models and can be utilised to localise specimens of varying size where clear object and background boundaries are visible [18, 20, 23, 24]. Most contemporary works utilise a neural network [16] as a basis for detection and classification. Many computer vision applications already exist that solve vision problems in different domains such as agriculture [2, 8, 22], aquaculture [9], and infrastructure [17]. In our case, such a system would primarily focus on krill detection and estimation of the detected specimen’s maturity stage and length.

3 Data and Methods

The dataset we are working with consists of 457 high resolution (6048 by 4032 px) images containing approximately 25 krill specimens each. Photographs were taken with a Nikon DX3 and two flash guns on a stand placed approximately 1m apart, with the specimens set out on blue plastic boards (in pre-drilled grooves). The camera was set in Manual mode with an aperture of $f/29$ and an exposure of $1/125$. The length and quality of the same krill were then evaluated physically by a marine biologist. Lengths of krill were measured according to the BAS standard, which involves measuring from the front of an animal’s eye to the end of its telson and then rounding the result down to the closest millimetre [14]. The Makarov and Denys scale [14], utilising the terminology established by [15], was used to evaluate the maturity level. The lengths and maturity level annotations will be used as the ground truth annotations in our subsequent experiments as described in Section 3.4.

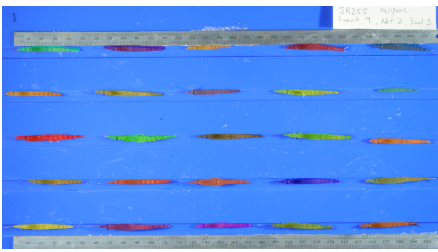
In the next Section, we describe how we detect krill specimens in the aforementioned input images (Section 3.1). The following Section 3.2 describes the pre-processing and data curation steps which are needed before the aforementioned image and annotation data is used in the developed web application (Section 3.3) and in the aforementioned maturity stage classification and length regression experiments (Section 3.4).

3.1 Krill detection

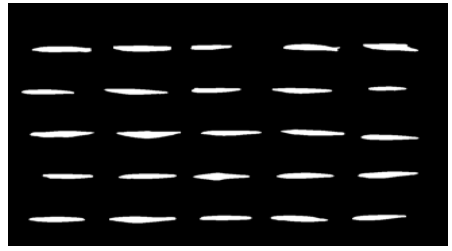
The uniform krill size and shape distribution in addition to the homogeneous environment in which they were captured provides a strong prior for object detection algorithms. The aim here is to identify each krill in an image by drawing bounding boxes or overlaying segmentation masks.

The krill detection pipeline uses a Mask R-CNN [16] to detect each specimen’s outline. We opt for an approach that ensures high resolution instance masks along with bounding box coordinates of the specimens are obtained. To do this in an effective way, weakly supervised method from bounding boxes [17] is used. We manually label all 457 full board krill images with bounding box labels and produce high-resolution instance masks visible in Figure 2 (a) and (b). Each specimen’s mask within the segmentation map is decoded to retrieve the bounding box coordinates. This acts as an automatic krill detector when passed to the labelling tool (see Figure 2 (d)) that we based on the VIA [18]. The bounding box parameters can be adjusted manually by the user as shown in Figure 2 (c), allowing for more specific user needs. The fully supervised krill segmentation algorithm assumes a varying data-set $D = \{(x_i, y_i) | 1 \leq i \leq N\}$ where x_i are the inputs, y_i are the ground truth masks, \mathcal{L} is the loss, and the training/testing procedure is denoted as $D_{TR} \cup D_{TE}$. The goal of the learning framework is to learn $f(x) = y$ by minimising $\sum_{(x_i, y_i) \in D_{TR}} \mathcal{L}(f(x_i), y_i)$ in the hope of generalising to $\sum_{(x_i, y_i) \in D_{TE}} \mathcal{L}(f(x_i), y_i)$. D_{TR} and D_{TE} are split with a random 80% to 20% ratio. A pre-trained ResNet50-FPN Mask R-CNN variation [16] model is fine-tuned on D_{TR} and tested on D_{TE} . In Figure 3, we show average precision and recall metrics for the experiment.

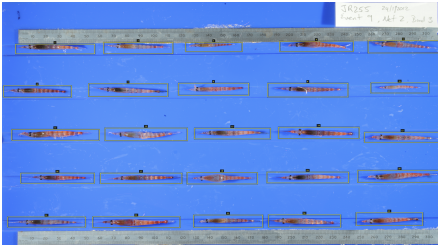
We further conduct 5 leave-one-cruise-out experiments [19]. In each experiment i , a unique cruise $c_i \in \mathcal{C}$ is designated for testing, while the remaining cruises are utilised for training. The results for krill detection can be seen in Figure 3. Our proposed Krill Tool which we will discuss in Section 3.3 utilises the model at a click of a button, and generates global instance masks for every specimen visible in an image. Figure 2 (e), (f) show mask predictions as compared to the labels for krill images from the test set. In total, this approach



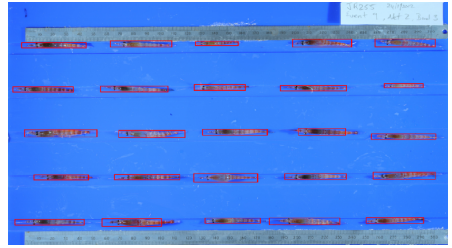
(a) Weakly supervised high-resolution instance labels.



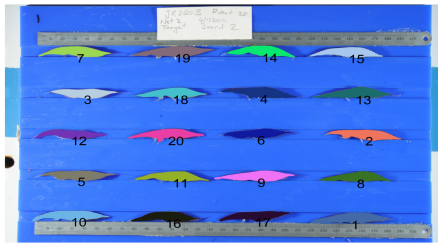
(b) Segmented binary image.



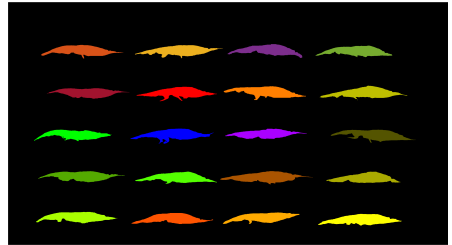
(c) Interactively adjustable bounding boxes [10].



(d) Auto localised krill from segmented image.



(e) Segmentation prediction visualisation for test set.



(f) Segmentation ground truth visualisation of previous image.

Figure 2: Krill detection image processing steps. The proposed detection pipeline produces bounding box coordinates as well as instance masks. Bounding boxes can be adjusted through VIA [10].

outputs multiple image areas of interest (krill instances) in bounding box and pixel mask format.

3.2 Data pre-processing

The data preparation process starts with the assignment of two essential target parameters to each specimen (length and maturity stage). As indicated at the beginning of Section 3, BAS provides ground truth data gathered by a field expert for the vast majority of specimens which is delivered in the MS Excel spreadsheet format. The given spreadsheet needs to be processed to a format usable by the subsequent downstream tasks. This includes enforcing naming conventions and dataset consistency. The sample of the pre-processed data can be seen in Table 1. This assumes an initial collection of all data and demonstrates that each view is regarded to have the same data attributes. The *ID* columns, which are also the

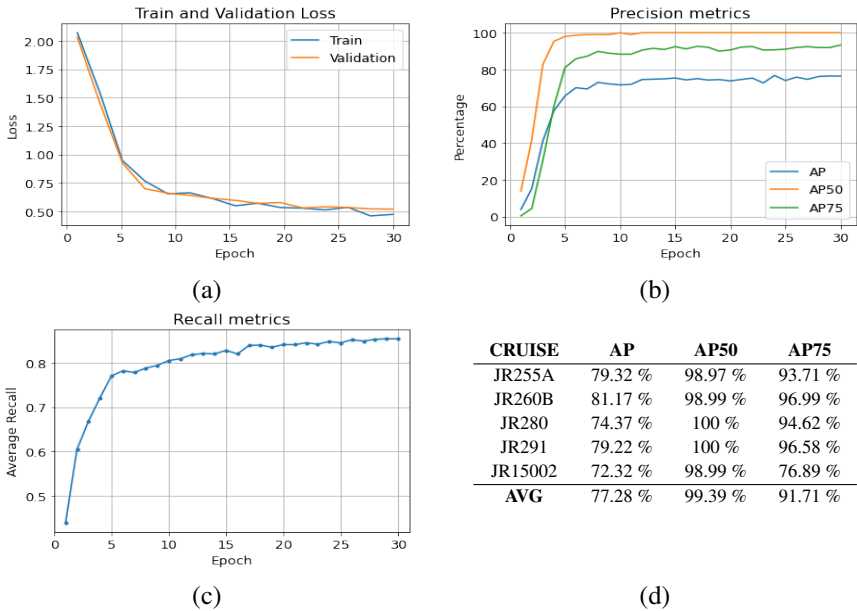


Figure 3: Model convergence plot (a). Mask precision metrics for test set (b). Mask recall metrics for test set (c). Table showing mask precision for cross validation sets (d).

picture file names, provide access to the related photos. The first two columns include the two attributes of interest: *length*, which was measured in millimetres, and a *maturity stage* indicator, as described in [14]. The two *ID* columns are formed by concatenating the cruise, image sequence number, file name and index, which is counted top-left to bottom-right as shown in Figure 2 (c). Catch specific details such as event, net, and board numbers are also recorded. The table is manually checked for clearly visible errors such as incorrect view or parameter assignment. The *x*, *y*, *width*, *height* are used with the image name to get the localised specimen views as shown in Figure 5 (b). The bounding box dimensions of localised krill specimens have a mean of 951 x 256 pixels and standard deviation of 222 x 62 pixels. The largest and smallest recorded samples have the dimensions of 1663 x 488 and 312 x 47 respectively. From this image pre-processing stage the 10524 entries mentioned in Table 1 are extracted from the high-resolution full-board images for further processing. The extraction process is done for the RGB images as well as their corresponding Mask R-CNN binary masks. Figure 5 (b) shows the 11 selected maturity stages that are analysed further. Only krill images that contain both lateral and dorsal views and all key parameters are used, meaning after data cleaning we accumulate 5095 labelled unique krill specimens spanning across 10190 images. The sample distribution with regards to the 2 key parameters can be seen in Figure 4. Both plots clearly indicate an imbalanced data-set, meaning further steps must ensure the reduction of bias as discussed in [9]. To enforce a constant spatial feature lock each image is padded up to 1700 x 500 pixels as this resolution falls slightly above the maximum recorded one. Each krill specimen region is placed at the centre of the padded image which uses colour of $RGB = [56, 127, 245]$ as background. The samples that included incorrect maturity labels such as M1, A2, U and classes that had ≤ 100 samples (FA5, FS3, MA3) were excluded during the data pre-processing steps.

length	maturity	cruise	x	y	width	height	ID	Alternative view ID	position	event	net	board
34	FS1	JR255A	469	751	869	114	JR255A_krill_image_73.jpeg-1	JR255A_krill_image_74.jpeg-1	Dorsal	78	2	3
23	J	JR255A	1368	869	537	118	JR255A_krill_image_73.jpeg-2	JR255A_krill_image_74.jpeg-2	Dorsal	78	2	3
25	J	JR255A	2207	851	560	123	JR255A_krill_image_73.jpeg-3	JR255A_krill_image_74.jpeg-3	Dorsal	78	2	3
29	J	JR255A	3172	819	746	168	JR255A_krill_image_73.jpeg-4	JR255A_krill_image_74.jpeg-4	Dorsal	78	2	3
40	MS1	JR255A	4319	783	1038	191	JR255A_krill_image_73.jpeg-5	JR255A_krill_image_74.jpeg-5	Dorsal	78	2	3

Table 1: Krill tool text database export (first 5 sample entries). Each entry is a registered krill view with corresponding parameters. Total 10524 entries.

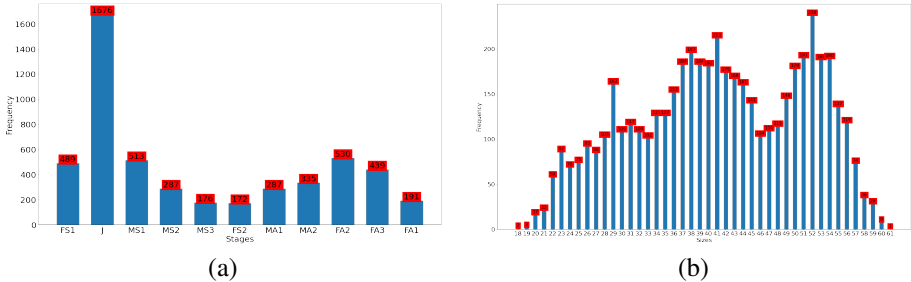


Figure 4: Krill maturity stage distribution across 11 stages (a). Body measurement distribution in millimetres (b).

3.3 Web application

A web-based application, named Krill Tool, has been developed using mainly Python Django [10] and VGG image annotator [11] to simplify the data annotation and pre-processing procedures. The application enables researchers to transform collected data into a format suitable for classifier training. The application interface provides an annotation system that enables researchers to rapidly assign parameters to each specimen. Each full-board image uploaded is allocated an alternative view, which displays either the Lateral or Dorsal view of each specimen. The system requires the user to associate each image view with its alternative. The application is intended for the manual entry of user data in order to designate the length and maturity labels for each specimen. It is anticipated that this system will facilitate future expansion of ground-truth images and annotations. Figure 5 (a) demonstrates the layout of the labelling interface.

3.4 Maturity classification and length regression

The 10190 extracted and pre-processed images seen in Figure 5 (b) are used to separately train a maturity stage classifier and a length regressor. Current experiments only use the bounding-box produced images and not the mask. For consistency, the same ResNet50 backbone [12] is used as in Section 3.1. We treat Lateral and Dorsal as separate data-sets and reduce resolution from the maximum one down to 340x100 over several steps. As shown in Figure 4, the collected data-set is unbalanced and needs certain approaches to account for that [9]. Hence, the following regression and classification experiments were executed where class weights were calculated $w_j = \frac{n}{s_j \cdot s}$, where w_j is the weight for each class, n the total number of samples in the data-set, s the total number of unique classes, s_j the total number of samples of the respective class. Parameter estimation experiments use a similar 80%/20% train/test split, although the training pipeline during these experiments is slightly altered - we allow the whole model to backpropagate, use more colour based augmentations,

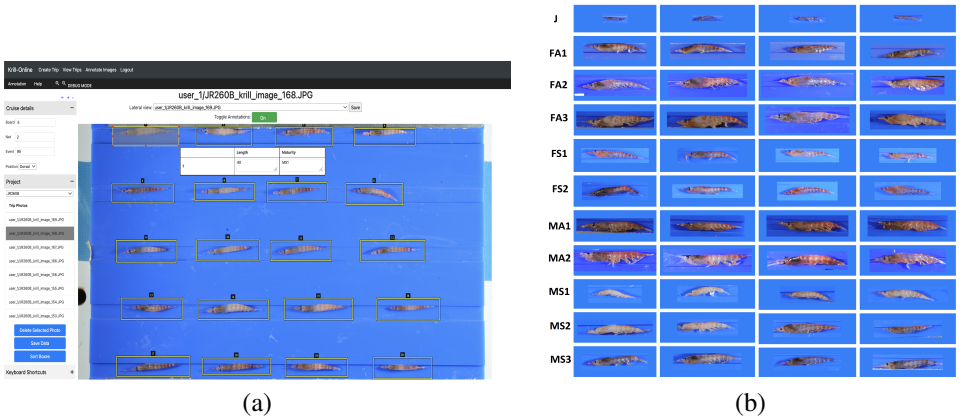


Figure 5: Web application user interface image labelling view (a). Pre-processed krill images (b).

reduce the learning rate to 1×10^{-3} and double the training time to 60 epochs.

Resolution	Lateral		Dorsal	
	Length	Maturity	Length	Maturity
340x100	2.23	58.41	3.12	57.63
680x200	2.26	59.85	2.33	57.09
1020x300	2.13	58.66	2.28	60.00
1360x400	2.42	62.99	2.19	57.09
1700x500	1.98	61.65	2.04	58.16

Table 2: ResNet50 Results for Length regression error in mm (lower is better) and Maturity classification accuracy (higher is better)

The results can be seen in Table 2. Overall, the Lateral view provides slightly better results overall, which is logical given that physical analysis of length and morphology has been carried out on the Lateral view by the BAS researchers.

As expected, higher image resolutions generally yield lower length regression errors as well as higher maturity stage classification accuracies. While the above trend is clear, the results also show that higher image resolutions do not always achieve the higher classification accuracy as can be seen from the maturity column for the Dorsal view images.

The two confusion matrices shown in Figure 6 give a detailed visualisation of the maturity stage classifier’s predictions. Similar confusion matrix patterns are visible across both dataset views. The junior maturity stage (class J) is the most predominant one and also achieves the highest classification accuracies in every experiment due to its distinct differences from other classes. Other misclassifications follow a predictable pattern - there are some misclassifications between MS1 and FS1 maturity stages, as well as between consecutive maturity stages e.g. MA1 and MA2; and MS1, MS2 and MS3. We need to note that our ground truth annotations have been provided by a single marine biologist and given the difficulty of establishing krill maturity stage manually, there is a certain level of uncertainty related to the ground truth annotations we used in the above experiments. This level of uncertainty

cannot be qualified without additional human annotations from other marine biologists [4], which would be of course helpful from the point of view of assessing the performance of the automated system in the context of human inter-observer error. As the current data was gathered by one marine biologist on board a research vessel, we are unable in this case to provide any inter-observer metrics.

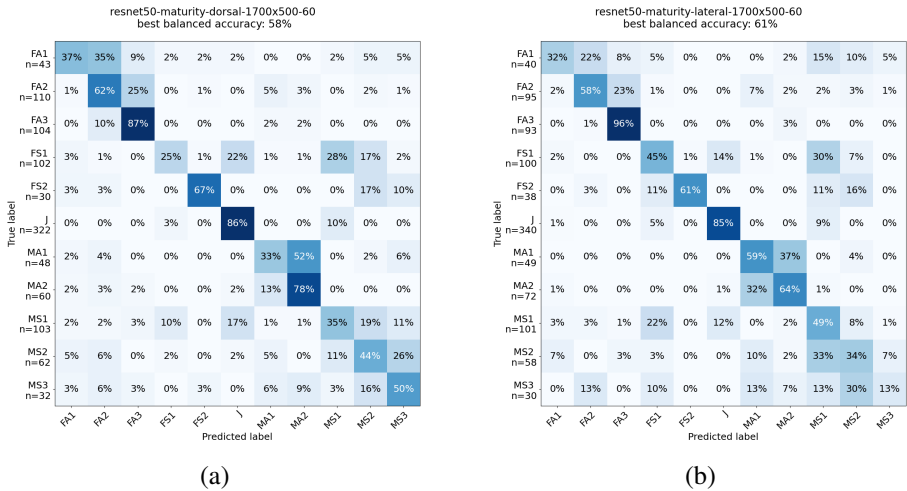


Figure 6: Dorsal model, 1700x500 pixel images (a). Lateral model, 1700x500 px images (b).

The length regressor proposed in this study employs the root mean square error (RMSE) loss function [10]. Consequently, the length regression errors observed in Table 2 and Figure 7 represent the average millimetric deviation of the predictions. Convergence is reached within a few epochs for higher resolution models. Lower resolution models take longer to converge, but their performance is on par with higher resolution models. The Lateral view on average yields a 2.20 mm error, whereas the Dorsal one, as expected, is slightly higher at 2.39 mm.

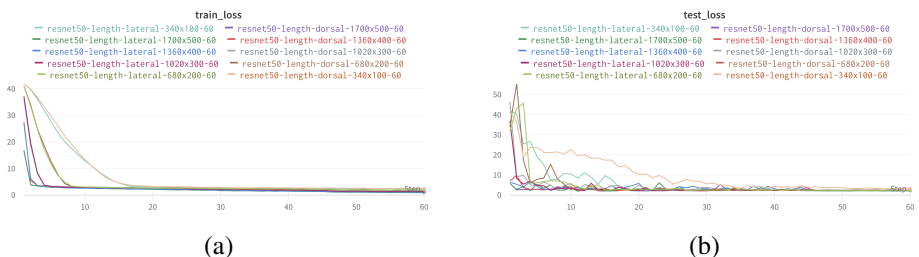


Figure 7: Train loss (RMSE) convergence (a). Test loss (RMSE) convergence (b).

4 Conclusions and Future Work

By leveraging the power of computer vision and deep learning in marine biology, here we demonstrate the capability of computer vision tools to automate the collection and analysis of Antarctic Krill image data. The essential function of Antarctic Krill in the marine food chain and the unprecedented challenges posed by climate change necessitate an accurate population assessment of krill.

The major product from this work is the Krill Tool, a web-based application that facilitates the data annotation and pre-processing to ensure consistency and efficacy in data preparation. The application provides a user-friendly pipeline that replaces manual methods with an automated framework, with promising initial results in maturity classification and length estimation. We achieve a highest accuracy of 62.99% on the Lateral view for maturity classification as well as a 1.98 mm length error.

Our further work will focus on more detailed analysis of the shape of the Antarctic Krill as well as on approaches to integrate the Lateral and Dorsal views. In the future, we also aim to develop more maturity classifiers that operate on different hierarchies of maturity levels (e.g. J, F and M; J, MS, FS, MA and FA; and similar.). Most importantly, the Krill Tool will allow BAS researches to rapidly collect more Antarctic Krill data using our automated approaches. The forthcoming future data is expected to reinforce the shown models by increasing the number of the data-set samples. Future work will ensure making the data-set and code available to the public.

5 Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council through the Centre for Doctoral Training in Agri-Food Robotics: AgriFoRwArdS [grant number EP/S023917/1].

References

- [1] Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4(none), jan 2010. doi: 10.1214/09-ss054. URL <https://doi.org/10.1214%2F09-ss054>.
- [2] Ran Bezen, Yael Edan, and Ilan Halachmi. Computer vision system for measuring individual cow feed intake using rgb-d camera and deep learning algorithms. *Computers and electronics in agriculture*, 172:105345, 2020.
- [3] Marek L Borowiec, Rebecca B Dikow, Paul B Frandsen, Alexander McKeeken, Gabriele Valentini, and Alexander E White. Deep learning as a tool for ecology and evolution. *Methods in Ecology and Evolution*, 13(8):1640–1660, 2022.
- [4] Ellen Bowler, Peter T. Fretwell, Geoffrey French, and Michal Mackiewicz. Using deep learning to count albatrosses from space: Assessing results in light of ground truth uncertainty. *Remote Sensing*, 12(12), 2020. ISSN 2072-4292. doi: 10.3390/rs12122026. URL <https://www.mdpi.com/2072-4292/12/12/2026>.

- [5] Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural networks*, 106:249–259, 2018.
- [6] Django Software Foundation. Django. URL <https://djangoproject.com>.
- [7] Abhishek Dutta and Andrew Zisserman. The VIA annotation software for images, audio and video. In *Proceedings of the 27th ACM International Conference on Multimedia*. ACM, oct 2019. doi: 10.1145/3343031.3350535. URL <https://doi.org/10.1145%2F3343031.3350535>.
- [8] Shuxiang Fan, Jiangbo Li, Yunhe Zhang, Xi Tian, Qingyan Wang, Xin He, Chi Zhang, and Wenqian Huang. On line detection of defective apples using computer vision system combined with deep learning methods. *Journal of Food Engineering*, 286:110102, 2020.
- [9] Geoff French, Michal Mackiewicz, Mark Fisher, Helen Holah, Rachel Kilburn, Neil Campbell, and Coby Needle. Deep neural networks for analysis of fisheries surveillance video and automated monitoring of fish discards. *ICES Journal of Marine Science*, 77(4):1340–1353, 2020.
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [11] Timothy O Hodson. Root-mean-square error (rmse) or mean absolute error (mae): When to use them or not. *Geoscientific Model Development*, 15(14):5481–5487, 2022.
- [12] Jennifer Jacquet, Daniel Pauly, David Ainley, Sidney Holt, Paul Dayton, and Jeremy Jackson. Seafood stewardship in crisis. *Nature*, 467(7311):28–29, 2010.
- [13] Tsung-Yi Lin, Piotr Dollár, Ross B. Girshick, Kaiming He, Bharath Hariharan, and Serge J. Belongie. Feature pyramid networks for object detection. *CoRR*, abs/1612.03144, 2016. URL <http://arxiv.org/abs/1612.03144>.
- [14] RR Makarov and CJI Denys. Stages of sexual maturity of euphausia suberba. biomass handbook 1.1, 1981.
- [15] DJ Morris, Jonathan L Watkins, CF Ricketts, Friedrich Buchholz, and Julian Priddle. An assessment of the merits of length and weight measurements of antarctic krill euphausia superba. *British Antarctic Survey Bulletin*, (79):27–50, 1988.
- [16] Berndt Müller, Joachim Reinhardt, and Michael T Strickland. *Neural networks: an introduction*. Springer Science & Business Media, 1995.
- [17] Marcos Quintana, Juan Torres, and José Manuel Menéndez. A simplified computer vision system for road surface inspection and maintenance. *IEEE Transactions on Intelligent Transportation Systems*, 17(3):608–619, 2015.
- [18] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

- [19] Quirin Schiermeier. Ecologists fear antarctic krill crisis. *Nature*, 467:15–15, 2010.
- [20] Christian Szegedy, Alexander Toshev, and Dumitru Erhan. Deep neural networks for object detection. *Advances in neural information processing systems*, 26, 2013.
- [21] Zhi Tian, Chunhua Shen, Xinlong Wang, and Hao Chen. Boxinst: High-performance instance segmentation with box annotations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5443–5452, 2021.
- [22] Nayeli Vélez-Rivera, José Blasco, Jorge Chanona-Pérez, Georgina Calderón-Domínguez, María de Jesús Perea-Flores, Israel Arzate-Vázquez, Sergio Cubero, and Reynold Farrera-Rebollo. Computer vision system applied to classification of “manila” mangoes during ripening process. *Food and Bioprocess Technology*, 7:1183–1194, 2014.
- [23] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11):3212–3232, 2019.
- [24] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey. *Proceedings of the IEEE*, 2023.