Application of Convolutional Neural Networks to monitor Seal activity in Antarctica

Fletcher Dick
Computer Science and Software Engineering
University of Canterbury
Christchurch, New Zealand
fdi19@uclive.ac.nz

***Abstract*—This project aims to conduct research into the feasibility of automated seal monitoring. Specifically, an automated approach is required to collect Spatio-temporal data on the Weddell seal population surrounding Scott Base, Antarctica. Alongside a working solution, this research project offers advice for future monitoring efforts, population counts — that serve as a future benchmark, and analysis of the collected data. This work is important to detect any future disturbance caused by the upcoming Scott Base reconstruction. This research and analysis will contribute to Antarctica New Zealand's understanding of the continent's changing ecosystem—due to the concern over increasing global temperatures. A RetinaNet CNN was trained to detect individuals and pairs of Weddell seals. The model achieves an F_1Score of 87.1% and an accuracy of 77.2%. These results are sufficient for detecting deviations in the seal population over time. Additionally, the detection CNN has a precision of 98.5% making it extremely tolerant to false positives. A second CNN was trained to minimise false low counts of seals due to Antarctic snowstorms affecting the survey camera's visibility. This snowstorm classification CNN model boasts successful results with a 97.8% accuracy and a 97.5% F_1Score.**

***Index Terms*—Antarctica, Convolutional Neural Network, Object Detection, RetinaNet, ResNet, Weddell Seal**

I. Introduction

This research project covers the monitoring of the Weddell seal colonies surrounding Scott Base, Antarctica. The Antarctic research centre is undergoing reconstruction starting with earthworks in the 2023/24 Summer season [1]. As a benchmark for comparison, monitoring of seal counts and their Spatio-temporal changes must be recorded before the construction. This is to successfully compare the seal movements to understand if the construction is disturbing the surrounding population. Having data spanning before the construction start date will allow for a causational relationship to be shown between the reconstruction and a decline in the seal population (if any).

Construction-induced population disturbance is a real threat, noise pollution has been shown to most affect marine mammals as the propelled sound waves interfere with their communication and navigation abilities [2]. Underwater vibratory pile driving — where a pile is driven into the seabed with high-frequency vibrations — can produce levels of 170 dB at 17 metres at close (17m) range [3]. Such noise pollution has been shown to have behavioural changes in marine mammals such as short-term and long-term avoidance of affected areas [4]. Therefore, accurate monitoring and population modelling are valuable to Antarctica NZ as their information can contribute to construction decisions. For example, choosing to pile drive during times when seal counts are low to minimise disturbance.

While ground counts performed by Antarctic researchers are a possible approach to seal counts, it is not feasible for frequent surveys (e.g. every 15m) due to the harsh climate and mental fatigue affecting their performance [5]. Manual counts using imagery solve the frequency issue however limitations persist. Namely, the laborious manual counts introduce observer biases [6]. Therefore, automated detection systems are required. Traditionally, hand-engineered features were used in the computer vision field to automatically detect objects. However, the process of hand-engineering features requires prior knowledge of the subject, fails to adapt (e.g. new camera position or angle) and is time-consuming.

Convolutional Neural Networks (CNN) is an extensively studied deep neural network and for good reason. CNNs are leveraging the vast amount of annotated data alongside the rapid growth in graphic card units to learn research-leading results [7]. Unlike traditional computer vision techniques used to detect objects, no feature engineering is required. This is because the key features of the objects are learnt rather than hardcoded. Negating the feature engineering process has a large advantage in that no prior knowledge of the domain is required to train a leading object detector.

This research has chosen to adopt the use of CNNs for the aforementioned benefits. Specifically, two CNNs have been developed and trained to accurately monitor the Weddel colony surrounding Scott Base. The first model utilises the RetinaNet architecture for detecting individual and seal pairs [8]. RetinaNet was chosen for its ability to detect small and dense objects. The second model utilises a Residual Network (ResNet) model [9] pre-trained using the ImageNet dataset. This second model classifies antarctic snowstorms that impact the visibility — and therefore accuracy — of the seal object detector. This allows for more accurate data analysis as seal counts that occur during a snowstorm can be ignored.

In Section II, we describe the relevant surrounding details of the research project along with its research objectives. In Section III, in-depth analysis and review are conducted on existing related solutions. Next, Section IV covers the proposed solution that meets the research objectives, including the design and implementation of the solution, and the project's design method. In Section VI, an evaluation of the proposed solution is discussed. Section VII summarises the context, objectives and results of the research project. Finally, in Section VIII an appendix is provided covering additional diagrams and links to related artifacts (e.g. source code).

II. Background & Objectives

Throughout 2022 as per the Bachelor of Engineering requirements for Software Engineering [10], I have undergone a capstone research project at the University of Canterbury. Such research projects require investigation and solution development for a complex internal or industry problem. This research project is supervised by Professor Richard Green and Postdoctoral Researcher Oliver Batchelor. Both Richard and Oliver are experts in the area of Computer Vision and have offered continuous assistance throughout the project. The research project is sponsored by Antarctica New Zealand, a government agency initiative. Antarctica New Zealand is responsible for undertaking research and analysis to understand the continent's changing landscape and ecosystem—due to the concern over increasing global temperatures [11]. One duty of Antarctica NZ is to monitor and analyse the patterns of wildlife on the continent. Special attention is given to possible disruptions to the Antarctic ecosystem, for example raising global temperatures and sound pollution [12].

The problem that this research project sets out to solve is long-term Weddell seal monitoring that will give insights into population changes and their spatial and temporal distributions. In solving this problem, numerous objectives must be met. The following research objectives have been proposed:

1) Recieve, process and store the raw timelapse images taken over the 2021-22 Antarctic summer. These images from Antarctica NZ contain updated viewpoints from previous years capturing an overview of Scott Base and the seals surrounding the centre. The collected raw images must also be converted into a practicable file type (e.g., .jpg, .png) and stored on an accessible machine.
2) Develop and train a neural network (as recommended by supervisors) that can accurately identify Weddell seals in all Scott Base image datasets (2018-19, 2019-20, 2021-22). The creation of such a generalised CNN may include the adaption of an existing and research-proven model or a novel model architecture. Research into the most suitable option must be made.
3) Compare the performance of the proposed model against a RetinaNet CNN only trained on the 2018-19 dataset to quantify improvements.
4) Develop and implement a snowstorm classification technique (e.g., using computer vision or CNN methods). Accurate classification of a snowstorm is required to distinguish between true and false low-counts of seals—where false low-counts are the resultant of visibility-reducing snowstorms.
5) Utilising the proposed solution, record the seal population counts, and Spatio-temporal data over the previous three Antarctic summers. Additionally, remove or tag counts impacted by snowstorm-induced low-visibility by using the developed classifier.
6) Compose a summary report including population counts, Spatio-temporal data, data analysis and advice for the antarctic camera positioning and framing for future seasons. Present the findings to Richard, Oliver and Antarctica NZ.

III. Related Work

To better give context to the research project while also surveying the wins and shortcomings of existing solutions, an analysis of related work is conducted. This section looks at published literature with proposed solutions; of which none are not fully satisfying, or must be adapted to develop a fit-for-purpose solution.

Over time wildlife research has seen a technological shift in observer methods. Early research predominantly utilises citizen science—the involvement of volunteers in research—to conduct field counts of wildlife populations. This method, however, introduces observer bias into the study [13]. Manual counting is also unforgiving when valuing repeatability and scale, thus a detection system should be utilised [14]. A research paper observed the endangered Mediterranean monk seal's habitual use of four caves [15]. Camera traps captured 7290 individual events each with a 15-second video. The total recording time exceeds 30 hours, of which researchers manually identified the appearance of Mediterranean monk seals. This significant man-hour contribution demonstrates the value gained by this SENG402 research project. With the proposed model, ground-captured images can be automatically processed for seal counts.

Recent research has developed a new CNN architecture that detects pack-ice seals (of which includes Weddell seals) [16]. The CNN, named SealNet, utilises WorldView-3—a commercial Earth observation satellite—with a combination of semantic segmentation and binary classification to detect the seals. However, the satellite-based CNN only detects 30% of pack-ice seals when compared to expert counts. Successfully detecting less than one-third of seals is not a viable solution for determining the disruption effects of the upcoming Scott Base reconstruction. The paper also outlines an important limitation regarding Weddell seals, the dataset is biased towards other seal species. There are only 981 training images of Weddell seals compared to the largest subset of 30,000 images.

While SealNet offers large-scale seal monitoring at the cost of low accuracy, another study proposes an accurate but limited classifier. This research paper proposed a CNN following the AlexNet architecture [17] which can accurately identify Saimaa Ringed seals with an accuracy of 90.5%. However, limitations with this solution exist. Namely, the model is not capable of detecting multiple instances of seals. Therefore, the AlexNet CNN simply answers a binary question—is there a seal in the image? Without the ability to count numerous seals, this approach is not suitable for population counts or Spatio-temporal analysis.

As mentioned in Section II, artefacts resulting from Oliver Batchelor's thesis [18] will be utilised in this research project. Specifically, annotations of Weddell seals were recorded as a result of evaluating Verification-Based Annotation. These annotations can create a CNN to detect the surrounding Weddell seals. Therefore, improving the aforementioned SealNet and AlexNet solutions. Another useful resulting artefact is a code base to train such a CNN model. One perceived limitation, however, is the lack of detailed documentation and abundance of code smells (e.g., commented-out code, unused classes/methods and depreciated dependencies). A portion of my research will be devoted to the revival and upkeep of the code base.

IV. Proposed Solution

Throughout 2022, consistent progress has been made toward the aforementioned research objectives. Specifically, objectives one, two and four have been the predominant focus of the former half of the project.

Objective one concerns the collection, preprocessing and storage of the new 2021/2022 raw images containing time-lapse overviews of the colony surrounding Scott Base. These images have been taken by Antarctica NZ and flown in from Antarctica on an 18-terabyte hard drive. A Python script had been developed to preprocess (crop) and convert the proprietary Panasonic Lumix raw image file type (.rw2) into a practicable file type (.jpg). Even though additional fail-safe measures such as checkpoints had been included in the script in the event of a system failure or power outage. Technical issues arose when running the script on UC research lab machines due to a lack of admin credentials. DigiKam, an alternative open-source image management application with the ability to batch convert and crop was instead used [19].

The latter half of the research project has focused on objectives three, five and six.

Objective six concerns a report detailing the project's findings, data analysis and recommendations. This research report is intended to fit the objective's requirements for all intents and purposes.

A. Design & Implementation

1) Seal Object Detection: The captured images taken in Antarctica of the Weddell seals surrounding Scott Base all share one common challenge. The cameras are located at an unforgiving length away from the subjects. While the exact distance between the colonies and the cameras at each location is unknown, the distance is enough to significantly impact the results of most CNNs. RetinaNet is a one-stage object detection model proposed in 2018 by Facebook Artificial Intelligence Research (FAIR) [8]. RetinaNet was chosen to detect the Weddell seals due to its improved accuracy on dense and small-scale objects. Shown in Fig. 1 is a magnified example image to demonstrate the small nature of the seals in question.

Fig. 1: Weddell seals at a large distance from camera

RetinaNet excels in detecting dense and small-scale objects thanks to two improvements over existing single-stage object detection models — Feature Pyramid Networks (FPN) [20] and Focal Loss.

Focal Loss offers an improved alternative to the Cross-Entropy loss function. A loss function is used to calculate how far the predictions are from the actual values. Focal Loss improves in class balancing meaning that a trained model using Focal Loss (e.g. RetinaNet) will have improved performance in classes that are difficult to detect or lack significant annotations compared to other classes. This is crucial for the proposed solution as seals and a pair of seals are separate classes, with 1658 seals and only 64 pairs being annotated using the 2020-21 Scott Base images, illustrated in Fig. 2:

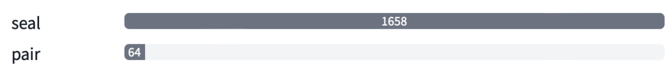


Fig. 2: Number of annotations per class

Number of annotations per class

Focal loss handles this class imbalance by applying a modulating factor to Cross Entropy. This "down weights" the loss of the well-classified examples (seals in our case) while only minorly decreasing the loss of poorly-classified examples with lower confidences (pair of seals) [8]. The amount of down weighting depends on the γ (Gamma) term in the Focal Loss equation, defined by eq. (1):

$$FL(p_t) = -(1 - p_t)^{\gamma} \log(p_t) \tag{1}$$

The effect of Gamma can be illustrated in Fig. 3. At $\gamma = 0$ the loss function is equivalent to the Cross-Entropy loss function. As Gamma increases, the loss on the well-classified examples decrease substantially equating to more focus on the harder examples. After experimenting, $\gamma = 2$ performs the best and is what is used for this research paper. This coincides with the RetinaNet white paper's findings cite13.

RetinaNet's FPN improves on previous feature pyramids by combining strong low-resolution features with weak high-resolution features. This is achieved by connecting a top-down pathway to the bottom-up convolutional layers with lateral connections at each layer. There are four components of the RetinaNet Architecture, shown in Fig. 4:

1) Bottom-up pathway — Utilises another CNN as a 'backbone network' (e.g., ResNet) to calculate feature maps at different scales.
2) Top-down pathway & lateral connections — Upsamples the higher layer feature maps and merges the layers using the lateral links.

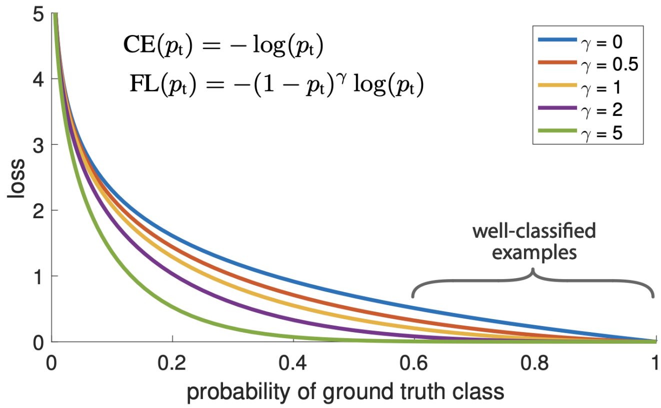


Fig. 3: Focal Loss compared to Cross-Entropy [8]

3) Classification subnetwork —For each location, it predicts the probability of a seal being present.
4) Regression subnetwork — Combines the anchor boxes for each object into a single bounding box.

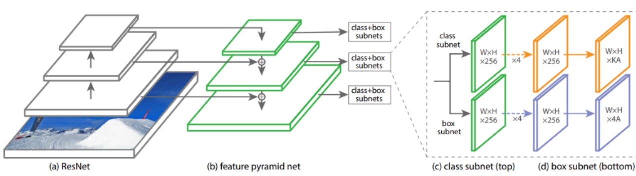


Fig. 4: RetinaNet Feature Pyramid Network [20]

The initial detection model was trained on 2018-20 Seal annotations formed as a result of Oliver Batchelor's work on Verification Based Annotation [18]. The initial model performed well only on trivial cases from tests using the 2018-19 dataset. With a threshold percentage of 50%, the model correctly identifies seven of the 13 seals with no False positive detections. With this threshold, the initial model results in an accuracy of 53.85%. Fig. 5 illustrates the detected and missed opportunities where each yellow square is a bounding box around a Weddell seal.

Fig. 5: Seals detected with 50% confidence threshold

To increase the performance of the model, more training was conducted. Specifically, transfer learning was adopted to obtain a high-performance model. Transfer learning is a training technique in which a model with a similar training domain is selected as a 'starting point' for training a new model. The learnt weights are copied to a new model before training on the new domain commences. The advantages of transfer learning are clear: increased performance, reduced training time, and reduced amount of annotated images required [21]. The proposed model was trained by transferring the learnt weights from the initial model. After, the annotated 2021-22 images were used to train the model. Ultimately leading to high performing object detection model that can accurately detect individuals and pairs of Weddell Seals.

Key metrics such as the Mean Average Precision (mAP) and Loss were recorded during training using TensorBoard, a logging and visualisation tool for machine learning experimentation. Fig. 10 (Appendix F) illustrates the training loss during training.

Location loss represents how far detections are from the ground truth annotations. A location loss of zero represents a perfectly overlapping bounding box. The location loss uses the PyTorch smooth L1 loss function with $\beta = 1$ and no reduction. The Classification loss represents the correctness of labelling the detected instances. The classification loss utilises the aforementioned focal loss function. The Total Loss function is defined by eq. (2):

$$TL(LL, FL) = LL + \frac{FL}{b} \tag{2}$$

Where b is a balancing factor used to prefer either classification or location during training. During training $b = 4$, this is to increase the importance of the Location Loss by four times. The reasoning for this is that the cost of misclassification is relatively low, for example, the misclassification of an individual seal as a pair only increases the population count by one.

The total training time totalled two hours and 37 minutes lasting 63 epochs, the best model — chosen by the lowest Total Loss — was trained in one hour and 36 minutes on epoch 40, as illustrated in Fig. 10. The best models loss results are listed below:

$$FL = 0.295 \qquad TL = 0.329$$
$$LL = 0.256$$

Fig. 11 illustrates the mAP scores over training. This metric is commonly used to measure an object detector CNN's performance over all classes. This is because the metric includes the trade-off between precision and recall while also considering false positives (FP) and false negatives (FN). Therefore, the mAP is an "all-in-one" metric. Fig. 11 (Appendix G) illustrates the average precision scores during training while the specific values for the proposed model are listed below:

$$AP = 21.1 \qquad mAP@30 = 37.6$$
$$mAP@50 = 37.6 \qquad mAP@75 = 23.2$$

Interestingly, combining the findings from Fig. 10 and Fig. 11, from epoch 44 there is a sudden rise in loss combined with a seemingly identical behaviour in mAP. This suggests that the model is overfitting the dataset.

Fig. 6 shows the results of the proposed RetinaNet model being run on an image from the new 2021-22 dataset with a confidence threshold of 50%. This image was not included in the training or validation datasets, therefore, is unseen as it truly tests the model's performance. Individual seals have had their bounding boxes shaded teal, while detected pairs of seals have an orange bounding box. This image was chosen as it contains the most seals detected from the recent Antarctic season with 332 total made up of 296 detected individual seals and 18 pairs (counted as two). However, the predictions also contain errors as 13 individual seals were not detected, there were however no misclassifications. With this in mind, this particular instance scored an accuracy of 96%.

A more precise evaluation of the model was conducted by measuring the number of True Positives, False Positives and False Negatives on a set of eight images unseen to the model. This unseen dataset used for testing was 15% of the total annotated images. Furthermore, the precision, recall and F_1Score were measured. An arbitrary detection threshold of 50% was chosen. This threshold was kept consistent with the initial detection model to allow comparisons between the two. The proposed RetinaNet model achieves successful results, displayed below are evaluation metrics and their values.

$$\begin{aligned} accuracy &= 77.2\% & precision &= 98.5\% \\ recall &= 78.1\% & F_1 Score &= 87.1\% \end{aligned}$$

2) Snowstorm classifier: As outlined in Objective four, a snowstorm classification model is vital to gather accurate seal counts. The value of a reliable and accurate model is in the ability to distinguish between true and false low counts of the population (where false counts arise from snowstorm-induced low-visibility). A suitable model has been created by utilising the ResNet-18 architecture and transfer learning [9]. The 18-layer ResNet model was chosen over higher-layered options as it reduces the likelihood of overfitting the dataset. A modular pipeline was designed to prepare a training dataset, train the ResNet-18 CNN using knowledge transfer, and classify an image. The design of the pipeline is detailed in Fig. 8 (Appendix A).

The Antarctic Dataset Subset (ADS) is a collection of 894 images each manually classified as including a snowstorm or not. The images are split evenly over the three cameras to not introduce a bias towards one viewpoint [22]. The pipeline can be broken down into three components: Preprocessing component, the ResNet-18 finetuning component and the Classification component. The Preprocessing component partitions the ADS into three datasets: training (70%), validation (15%) and testing (15%). Next, augmentations are applied to avoid overfitting. Each large image from the ADS is also split into five smaller 500 x 500 images by applying five random crops. This synthetically increases the dataset from 894 to 4470 images. The Finetuning component utilises knowledge transformation from a pre-trained ResNet-18 model. By initialising the weights of the model to a model trained on the ImageNet dataset, a significant time is saved and can result in better-performing models [23]. Also contained in this component is the training of the ResNet-18 CNN. Finally, the Classification model determines whether the input image contains a storm or is clear by utilising the created ResNet-18 model. Five random crops are taken from the input image and each is classified individually by the model. This self-ensemble method increases the accuracy of the snowstorm classification. eq. (3) defines the method for combining the accuracies.

$$\begin{aligned} a_j &:= \begin{cases} a_j, & \text{if Storm} \\ -a_j, & \text{if Clear} \end{cases} \\ a_{combined} &:= \frac{\sum_{j=1}^{E} a_j}{E} \end{aligned} \tag{3}$$

Where a is accuracy and E is the number of crops (five in this case).

The resulting class is a storm if the combined accuracy is positive, otherwise, clear. The ResNet-18 model achieves successful results, displayed below are evaluation metrics and their values.

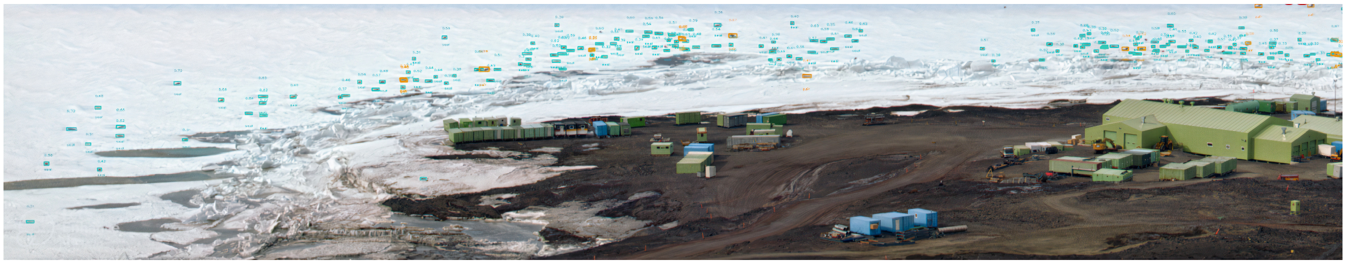
$$\begin{aligned} accuracy &= 97.8\% & precision &= 99.6\% \\ recall &= 95.5\% & F_1 Score &= 97.5\% \end{aligned}$$

Fig. 9 (Appendix D) displays the confusion matrix of the model. Specifically, the matrix quantifies the number of True Positives, True Negatives, False Positives and False Negatives.

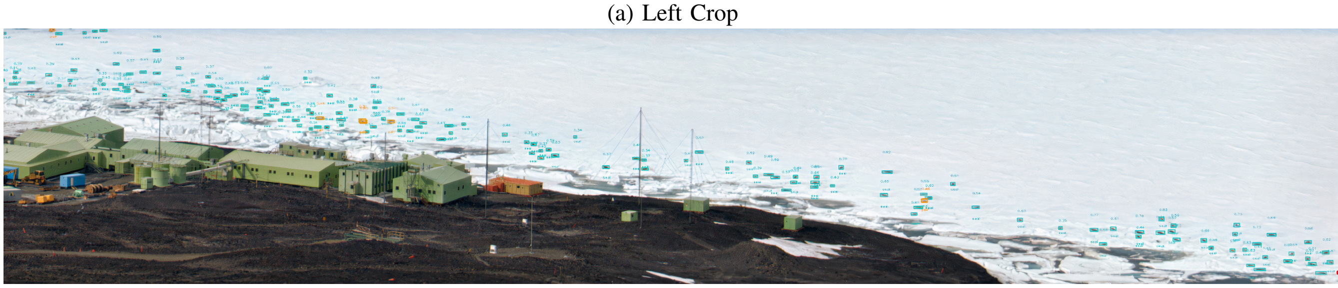
V. Method & Project Management

The research method of this project has followed from a background study of analysing and reviewing related work to the project. This includes either solutions that are not fit for purpose or methods that are not fully satisfying. Only academic literature was considered to filter low-quality solutions out. Other research was conducted into the available and relevant technological solutions. Possible technologies were evaluated through the use of expert opinion and experimentation. Computer vision experts Oliver Batchelor and Richard Green offered time-saving advice on the best technologies to use and those to avoid. By consulting with industry experts I was able to settle early on a working tech stack (e.g., PyTorch, RetinaNet). Experimentation was also applied for research, for example, early experimentation was conducted to evaluate 'off-the-shelf' models provided by PyTorch for seal detection. This experiment resulted in the understanding of the limitations of 'off-the-shelf' solutions and the challenge of small-scale object detection in CNNs.

The software development methods used in the research project have followed iterative development. This 'fail fast' method has allowed me to learn early about technological limitations, for example, the limitation of small-scale object detection within CNNs. From this, alternative and fit-for-purpose solutions were researched and developed early, here being a RetinaNet CNN model. Throughout the research project, several tools have been adopted to support the software development process. These are ToDoIst for task tracking and



(a) Left Crop



(b) Right Crop

Fig. 6: 332 seals correctly detected using proposed solution

GitHub for version control of the code-base and artifacts. Weekly sessions with supervisors and computer vision stakeholders were conducted to give project updates and receive feedback. Feedback included advice on project progress and technical recommendations. Each feedback session consists of sharing the progress made in the previous week towards the research objectives. Slack is also utilised in the computer vision team to share general project updates throughout the week and for announcements.

VI. Discussion & Evaluation

As per project requirements, the RetinaNet CNN is fit-for-purpose. The proposed model is effective for monitoring the Weddell Seal population surrounding Scott Base. This confidence in the proposed model is rooted in its successes, namely, the model has been shown to accurately detect up to 332 seals in a single image (50% confidence threshold). Additionally, at a confidence threshold of 50%, the model has a precision of 98.5%, a recall of 78.1% and an accuracy of 77.2%. The F_1Score — an "all-in-one" metric — scored an impressive 87.1%. The proposed model shows a tangible improvement of over 23% in accuracy in comparison to the initial model (trained using the 2018-19 dataset and before transfer learning). While likely sufficient for the model's use case, future work would be beneficial to improve this metric. The mAPs of the model are 37.6, 37.6 and 23.2 with IoU thresholds of 30, 50 and 75 respectively. These results are relatively low in comparison to the related research discussed in Section III. This is likely due to the small distinguishing details between the two classes (individual seal and seal pair). These alike classes in combination with low-fidelity images (after cropping) will undoubtedly lower the mAP by misclassifying the seal pair class. Additionally, only 64 seal pairs were annotated in comparison to the 1658 individual seals. Future work annotating more examples of the pair class would be beneficial to the performance of the model.

The snowstorm classification CNN achieved exceptional results and proved to be a fit-for-purpose addition to the research project. This is shown by its ability to recognise low-visibility images which ultimately affects the accuracy of the seal detection model. The model achieved an accuracy of 97.8%, a precision of 99.6%, a recall of 95.5% and an F_1Score of 97.5%. While these results prove the effectiveness of the model, more analysis should be conducted on its performance in conjunction with the new dataset (2021-22). This is noteworthy as the model was trained and tested on the 2018-19 dataset.

The performance of both models can be illustrated in unison by plotting the counts detected over the 2021-22 season alongside images with detected snowstorms. Fig. 7 illustrates exactly this. Three confidence thresholds (30, 40 & 50) were selected to illustrate the variance of seal counts with respect to the thresholds. Some performance measures can be inferred from this figure. Namely, the snowstorm classification CNN looks to correlate with the population counts dipping to zero. This suggests that the classification CNN is correctly identifying the false low counts due to the low visibility.

Complying with the research objectives, advice has been collated over the course of this project. The goal of this feedback is to evaluate the current timelapse camera setup and offer suggestions for improvement to Antarctica NZ for monitoring future seasons. In late 2021, two new timelapse cameras were installed overviewing Scott Base and Turtle Rock. The high-resolution imagery from these devices was intended to serve a multitude of research purposes. However, in the case of this research project, the generalised images had one major drawback. The camera positioning of the Scott Base and Turtle Rock cameras was at a large distance away

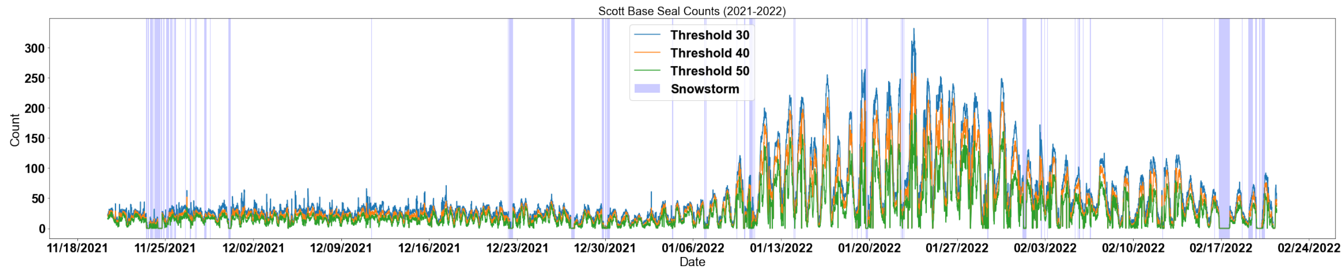


Fig. 7: Weddell Seal counts (Scott Base) over 2021-22 Antarctic summer

from the subjects. In the latter case, the images were deemed unusable by Richard Green (so were not the focus of this project). Ultimately, lossy cropping was applied to mask out the unnecessary portions of the image. This in turn resulted in a much lower resolution image and likely impacted the model's performance — especially in distinguishing between classes due to the loss of key features. Advice for future seasons would be to reposition the camera closer to the subjects and utilise optical zoom so that only regions of interest are in the frame. It should also be noted that any adjustments to the camera positioning will require the proposed model to be retrained using imagery captured and annotated from the new position.

While some are mentioned previously, for the reader's convenience a list of research limitations is collated below:

1) The timelapse camera used for model training was estimated to be positioned over 2km from the seals. This caused low-fidelity objects with few defining features ultimately affecting the CNNs performance (especially for the pair of seals class).
2) The CNN is not a generalised solution for seal detection. Changes to the captured scene and camera equipment will likely cause performance-degrading impacts. Re-training will need to be conducted as in the case with this research paper.
3) The proposed solution has an accuracy of 77.2%. This limits the application of the CNN to applications where precise counts are required.
4) Here, snowstorm detections are binary (either detected or not) whereas visibility is best quantified on a continuous scale. This limitation forces the detection confidence to be used for this purpose in a crude manner.
5) Time constraints as a result of the SENG402 schedule meant that future work exists.

VII. Conclusion

This research report focussed on an automated and accurate approach to long-term seal monitoring. Additionally, this research — sponsored by Antarctica NZ — lays foundational research into detecting colony disturbing events as a result of noise pollution from the upcoming Scott Base reconstruction. The objectives of this research project are summarised as follows: Collect, process, and store timelapse images of the 2021-22 seal colony surrounding Scott Base, Antarctica; Create an accurate and automated solution for detecting seals in provided datasets; Analyse trends in seal counts and movements offering a background dataset to compare the upcoming construction event to; Compose a technical report summarising findings and recommendations for camera equipment for future seal monitoring. This research has delivered all objectives set for the project. The ResNet-18 snowstorm classifier CNN offers a 97.8% accuracy and a 97.5% F_1Score. While the Seal object detector CNN has scored an F_1Score of 87.1% and a 77.2% accuracy. Both models in unison create an accurate and automated solution for long-term monitoring of the Weddell seal population surrounding Scott Base while also enduring low-visibility inducing antarctic snowstorms.

REFERENCES

[1] Y. Srinivasa, "Scott base redevelopment on track with detailed design phase almost complete," https://www.stuff.co.nz/timaru-herald/news/129463834/scott-base-redevelopment-on-track-with-detailed-design-phase-almost-complete, Aug. 2022, accessed: 2022-10-25.

[2] C. Erbe, M. Dähne, J. Gordon, H. Herata, D. S. Houser, S. Koschinski, R. Leaper, R. McCauley, B. Miller, M. Müller, A. Murray, J. N. Oswald, A. R. Scholik-Schlomer, M. Schuster, I. C. Van Opzeeland, and V. M. Janik, "Managing the effects of noise from ship traffic, seismic surveying and construction on marine mammals in antarctica," *Frontiers in Marine Science*, vol. 6, 2019. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmars.2019.00647

[3] P. H. Dahl, D. R. Dall'Osto, and D. M. Farrell, "The underwater sound field from vibratory pile driving," *The Journal of the Acoustical Society of America*, vol. 137, no. 6, pp. 3544–3554, 2015. [Online]. Available: https://doi.org/10.1121/1.4921288

[4] D. P. NOWACEK, L. H. THORNE, D. W. JOHNSTON, and P. L. TYACK, "Responses of cetaceans to anthropogenic noise." *Mammal Review*, vol. 37, no. 2, pp. 81 – 115, 2007. [Online]. Available: https://search.ebscohost.com/login.aspx?direct=true&AuthType=shib&db=a9h&AN=25369750&site=ehost-live&custid=s1165276

[5] L. G. Faber, N. M. Maurits, and M. M. Lorist, "Mental fatigue affects visual selective attention," *PLOS ONE*, vol. 7, no. 10, pp. 1–10, 10 2012. [Online]. Available: https://doi.org/10.1371/journal.pone.0048073

[6] J. L. Dickinson, B. Zuckerberg, and D. N. Bonter, "Citizen science as an ecological research tool: Challenges and benefits," *Annual Review of Ecology, Evolution, and Systematics*, vol. 41, pp. 149–172, 2010. [Online]. Available: http://www.jstor.org/stable/27896218

[7] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[8] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," 2017. [Online]. Available: https://arxiv.org/abs/1708.02002

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1512.03385

[10] F. Gilson, "Software engineering research project." [Online]. Available: https://www.canterbury.ac.nz/courseinfo/GetCourseDetails.aspx?course=SENG402&occurrence=22W%28C%29&year=2022

[11] H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem *et al.*, "Technical summary," *IPCC special report on the ocean and cryosphere in a changing climate. Available: www. ipcc. ch/site/assets/uploads/sites/3/2019/11/04_SROCC_TS_FINAL. pdf.(July 2020)*, p. 61, 2019.

[12] C. Erbe, M. Dähne, J. Gordon, H. Herata, D. S. Houser, S. Koschinski, R. Leaper, R. McCauley, B. Miller, M. Müller, A. Murray, J. N. Oswald, A. R. Scholik-Schlomer, M. Schuster, I. C. Van Opzeeland, and V. M. Janik, "Managing the effects of noise from ship traffic, seismic surveying and construction on marine mammals in antarctica," *Frontiers in Marine Science*, vol. 6, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fmars.2019.00647

[13] J. L. Dickinson, B. Zuckerberg, and D. N. Bonter, "Citizen science as an ecological research tool: Challenges and benefits," *Annual Review of Ecology, Evolution, and Systematics*, vol. 41, no. 1, pp. 149–172, 2010. [Online]. Available: https://doi.org/10.1146/annurev-ecolsys-102209-144636

[14] P. B. Conn, J. M. Ver Hoef, B. T. McClintock, E. E. Moreland, J. M. London, M. F. Cameron, S. P. Dahle, and P. L. Boveng, "Estimating multispecies abundance using automated detection systems: ice-associated seals in the bering sea," *Methods in Ecology and Evolution*, vol. 5, no. 12, pp. 1280–1293, 2014. [Online]. Available: https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12127

[15] E. Saydam, H. Güçlüsoy, and Z. A. Kızılkaya, "A novel approach for mediterranean monk seal conservation: an artificial ledge in a marine cave," *Oryx*, p. 1–3, 2022.

[16] B. Gonçalves, B. Spitzbart, and H. Lynch, "Sealnet: A fully-automated pack-ice seal detection pipeline for sub-meter satellite imagery," *Remote Sensing of Environment*, vol. 239, p. 111617, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425719306376

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[18] O. Batchelor, "Verification based annotation for visual recognition." 2019.

[19] 2001-2022 digiKam team, "digikam: digikam API reference page," https://www.digikam.org/api/index.html, accessed: 2022-10-25.

[20] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," 2016. [Online]. Available: https://arxiv.org/abs/1612.03144

[21] P. Sharma, "Understanding transfer learning for deep learning," https://www.analyticsvidhya.com/blog/2021/10/understanding-transfer-learning-for-deep-learning/, Oct. 2021, accessed: 2022-10-25.

[22] C. K. I. Williams, "The effect of class imbalance on precision-recall curves," *Neural Computation*, vol. 33, no. 4, pp. 853–857, apr 2021. [Online]. Available: https://doi.org/10.1162%2Fneco_a_01362

[23] J. Lalor, H. Wu, and H. Yu, "Improving machine learning ability with fine-tuning," 02 2017.

VIII. Additional Material

A. Appendix A - ResNet-18 Architecture

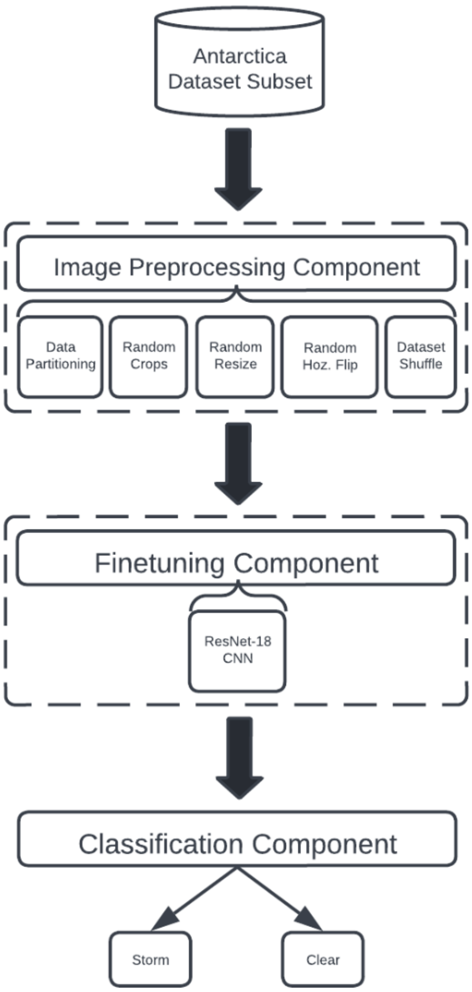


Fig. 8: Architecture of proposed ResNet-18 pipeline

B. Appendix B - Source code

The Jupyter Notebook containing the image preprocessing, finetuning and classification stages are hosted online at https://github.com/fletcherd3/Antarctic-Snowstorm-Classification/blob/master/storm_detection.ipynb.

The general repository for storing SENG402-related work and forks of related repositories are hosted online at https://github.com/fletcherd3/SENG402

C. Appendix C - Dataset

The random cropped images created from the Jupyter Notebook are organised into their classes (Storm & Clear) and the datasets (train, validation & testing). The images are hosted online at https://github.com/fletcherd3/Antarctic-Snowstorm-Classification/tree/master/data.

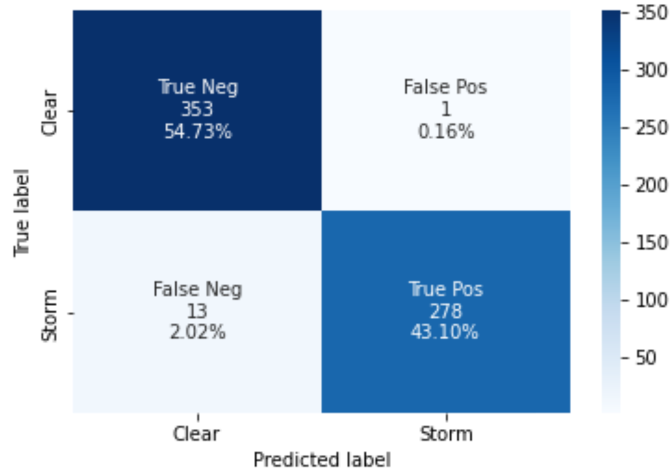


Fig. 9: Confusion Matrix of proposed ResNet-18 model

D. Appendix D - ResNet-18 Confusion Matrix

E. Appendix E - Annotations

The annotations of the 2021-22 Scott Base imagery are hosted on Segments.AI at https://segments.ai/segmentsai1/Seal_2022-22/.

F. Appendix F - Training loss

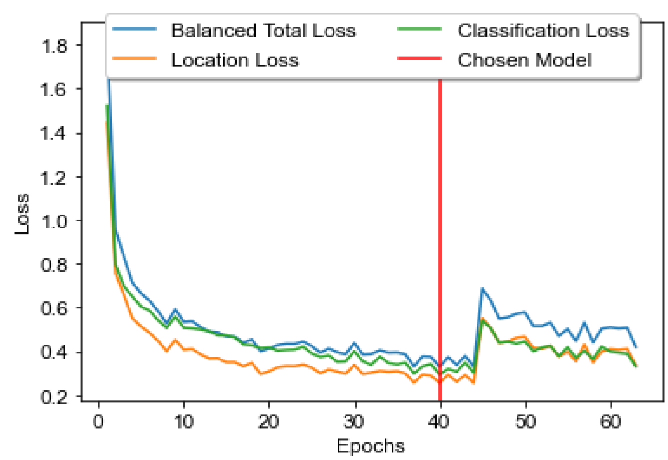


Fig. 10: Loss during CNN training

G. Appendix G - Training AP

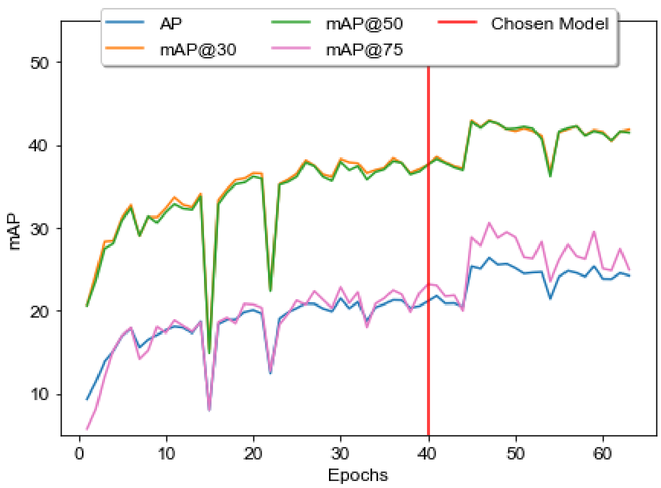


Fig. 11: Average precision during CNN training