Biological Conservation 296 (2024) 110713

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Using deep learning to automate the detection of bird scaring lines on fishing vessels

Debaditya Acharya [a,b,*], Muhammad Saqib [c], Carlie Devine [d], Candice Untiedt [d], L. Richard Little [d], Dadong Wang [c], Geoffrey N. Tuck [d]

[a] Geospatial Science, RMIT University, Melbourne, Australia
[b] Environment, CSIRO, Clayton, Melbourne, Australia
[c] Data61, CSIRO, Marsfield, New South Wales, Australia
[d] Environment, CSIRO, Hobart, Tasmania, Australia

ARTICLE INFO

Keywords:
Electronic monitoring
Bird-scaring lines
Deep learning
Activity recognition
Object detection
Seabird
Bycatch reduction

ABSTRACT

Bird-scaring lines (BSLs) are an essential on-vessel bycatch mitigation device to reduce seabird interactions with fishing gear, such as the baited hooks of longline vessels. To ensure compliance with the behaviours required to operate successful BSLs, Electronic Monitoring (EM) cameras installed on fishing vessels can facilitate monitoring of commercial fishing activities. This study proposes an Artificial Intelligence and Machine Learning (AIML) framework based on a state-of-the-art deep learning computer vision approach called Faster RCNN to detect BSLs using vessel Electronic Monitoring (EM) video footage. The experiments include comprehensive analysis for detecting BSLs during daytime and night-time using footage from tuna longline vessels, under various weather conditions. Results show that a detection precision of 0.87 can be achieved. This valuable AIML tool can significantly reduce the time and costs associated with reviewing human EM footage, expand coverage, and automatically identify events for compliance checks and endangered species monitoring.

1. Introduction

Seabirds are important components of marine ecosystems, where they fulfil a role as top predators and are well-established indicators of oceanic conditions (Montevecchi, 2023). Fishery discards often supplement their diets (Arata and Xavier, 2003), but at a risk of incidental bycatch and mortality. Fishery bycatch has been recognised as the greatest threat to seabirds globally, which affects 28 % of all seabird species, with half of these being threatened with extinction(Brothers, 1991; Anderson et al., 2011; Phillips et al., 2016; Avery et al., 2017; Cooper et al., 2006; Montevecchi, 2023). Bycatch usually occurs from entanglement in gillnets, when baited hooks from longlines are pursued, or from striking warp cables of trawl vessels (Gales, 1998; Bull, 2007). Albatrosses and petrels are most imperilled, with significant declines in populations worldwide being directly attributed to high levels of bycatch in pelagic and demersal longline fisheries (Tuck et al., 2001; Baker et al., 2002; Croxall et al., 2012; Gilman et al., 2021; Montevecchi, 2023).

The adverse impacts of longline fishing on seabirds have been known since the early 1990s (Brothers, 1991; Junior, 1991; Weimerskirch et al., 1997; Murray et al., 1993). Since then, various agreements, regulations and mitigation strategies have been established and implemented globally, all of which aim to significantly reduce seabird bycatch (Commonwealth of Australia, 2018; ACAP, 2021). For example, in 2004 the Agreement for the Conservation of Albatrosses and Petrels (ACAP) was established to work with Regional Fisheries Management Organisations (RFMOs) and other relevant fisheries organisations to encourage the adoption of best-practice mitigation measures to reduce seabird mortality in fisheries. In Australia, the incidental catch of seabirds during longline fishing was listed on 'Schedule 3 Key Threatening Processes' of the Endangered Species Protection Act 1992 and many species of seabirds have been listed as threatened species under the EPBC Act (1999). The significance of these listings means that the mitigation of bycatch of seabirds on longliners is coordinated and implemented through a Threat Abatement Plan (TAP) for the bycatch of seabirds from oceanic longline fishing operations (Commonwealth of Australia, 2018). In Australian waters, bycatch rates are generally required (by the TAP) to be below either 0.01 or 0.05 birds per 1000 hooks. Considerable

* Corresponding author at: Room 15, Level 12, Building 12, RMIT University, Melbourne, Victoria 3010, Australia.
E-mail address: debaditya.acharya@rmit.edu.au (D. Acharya).

https://doi.org/10.1016/j.biocon.2024.110713
Received 17 November 2023; Received in revised form 21 June 2024; Accepted 1 July 2024

success in meeting these criteria has been achieved, with estimated bycatch rates well below these requirements (Baker and Robertson, 2018), in part due to the implementation of BSLs (Commonwealth of Australia, 2018). Continued success of mitigation programs in Australia and globally will require mechanisms in place to ensure ongoing compliance and that reviews of compliance are accurate, timely and cost effective.

Globally, several mitigation devices and operational measures have been mandated or recommended for use in longline fisheries, including BSLs (also known as tori lines), line weighting, hook shielding devices, night setting and managing offal discharge (Lawrence et al., 2006; Bull, 2007; Melvin et al., 2013; Commonwealth of Australia, 2018; AFMA, 2023; ACAP, 2021). BSLs are one of the most effective and economical mitigation measures used to reduce seabird bycatch. In general, BSLs consist of a single mainline (or backbone) which is attached to a high point on the stern and extends into the water. A towed object (e.g. buoy) attached to the end of the line creates drag which keeps the line taut and ensures that a section of it remains above the water surface. Several brightly coloured streamer lines dangle vertically from the aerial section of the backbone to create a physical barrier and visual deterrent to seabirds (e.g.Fig. 1). The colour, movement and placement of the streamer lines deter the birds from accessing submerged baited hooks (Bull, 2009; Sato et al., 2016; Gilman et al., 2021; ACAP, 2021). Properly designed and deployed BSLs are effective at reducing seabird bycatch, in particular for longline vessels, with many studies demonstrating significant reductions, by more than 90 % in some cases, and for several at risk species, across multiple regions (Klaer and Polacheck, 1998; Løkkeberg and Robertson, 2002; Otley et al., 2007; Sato et al., 2013; Melvin et al., 2013; Maree et al., 2014; Melvin et al., 2014; Avery et al., 2017; Paterson et al., 2019; Jiménez et al., 2020; Da Rocha et al., 2021; ACAP, 2021).

To monitor compliance with, and effectiveness of, mitigation measures, fisheries management authorities collect data through fisher-dependent reports such as the Threatened, Endangered and Protected (TEP) species interaction reports and bycatch records, and independent means. Human observers on vessels have been the primary source of independent at-sea fisheries data, but the need to increase coverage coupled with the increasing cost and risk of using observers has catalysed the use of Electronic Monitoring (EM). EM uses cameras and associated hardware components (e.g. sensors, communication and geolocation devices) deployed on-vessel to record and collect video of fishing activities (Gilman et al., 2019; van Helmond et al., 2020). The EM video is used to manage and monitor fishing activities, including compliance with regulations such as the deployment of BSLs and to verify fisher-dependent data (e.g. logbooks) (Brown et al., 2021).

One of the major challenges with EM programs worldwide is that the quantity of video data collected far exceeds the capacity to manually review it (Gilman et al., 2019; Emery et al., 2019). As a result, only a subset of EM footage is reviewed due to the associated time and costs of human review (Emery et al., 2019). For example, in Australia, EM systems are deployed on all tuna longline vessels, with a subsequent minimum of 10 % footage reviewed. Developments in Artificial Intelligence and Machine Learning (AIML) technologies that automate video analysis tasks (e.g. estimate fish counts) have the potential to enable more efficient and timely review of EM footage (Qiao et al., 2021; Khokher et al., 2022), and increase data collection and review coverage for key activities (e.g., TEP species interactions; BSL deployment). Compared to other application areas of object detection, for instance in CCTV camera images, the application of AIML technology to EM data is however complicated by the need for large volumes of human-labelled images, the inherent variability of EM footage quality due to lighting and weather conditions, infrastructure occlusion, the presence of lens artefacts (e.g. water droplets), and the variability between camera positioning and appearance of different vessels.

In this study, we use a deep learning-based object detection framework (Faster R-CNN; Ren et al., 2015) to train and test an algorithm for the automatic detection of BSLs in EM footage on tuna longline vessels in Australia. We conduct experiments using EM footage of variable quality (e.g. water droplets on lens), from multiple vessels and conditions (e.g. dawn, day, dusk and night) to develop a scaled algorithm which is robust to variations in imagery. Automated detection of BSL deployment is a valuable AIML tool which can be used to reduce the time and costs of human EM footage review, increase coverage, and automatically flag events for compliance checking and TEP species monitoring.

2. Methods

This study used EM footage from three Australian tuna longline vessels collected during day and night in 2018 and 2019. Video footage from the EM camera mounted to view aft of the vessel's stern was analysed to detect the deployment of fishing lines and BSLs. The location of the EM cameras often means that they are placed in exposed areas subject to extreme variations in lighting (daylight vs. bright deck lights used at night) and weather (e.g., rain) conditions. Light can strike the lens, and water droplets and condensation can settle on the camera lens, causing artefacts in the video footage, such as reflection and image blur. Similarly, general weather conditions such as rain and fog can affect the quality of EM footage, and the viewpoint of the EM camera is slightly different for each vessel. Hence, the view of BSLs and the spatial arrangement of each vessel's deck are unique.

The framework as shown in Fig. 2 was trained using a deep learning-based object detector called Faster R-CNN (Ren et al., 2015) on human-labelled EM footage data from a single vessel that was used as a starting point for training and detecting BSLs in day/night settings during various weather conditions. While the framework can detect BSLs on one vessel, it is also desirable that the framework can generalise well on unseen vessel data without fine-tuning for each vessel. Therefore, we also investigated the data requirements of the framework to assess detection over different vessel structures and viewpoints.

2.1. Datasets

The EM data comprise images captured from three vessels in diverse weather and lighting conditions. Table 1 summarizes the dataset details. The experimental results along with distinct camera viewpoints are presented in Section 3.2. The EM data are confidential and private; consequently, we have redacted all vessel information and obscured all human faces. Across all three vessels, the quality of the EM footage was generally of low resolution (1360×768 compared to 3840×2160 or 4 K resolution for most modern surveillance cameras) and often blurry due to lens artefacts (e.g., water droplets). Despite this, the bright red/pink colour of the BSL streamers was visible even in rough seas at night due to their vertical orientation, movement and colour being highlighted by vessel lighting.

The video footage was divided into individual frames and annotated with bounding boxes to facilitate object detection using the open-source, web-based software Computer Vision Annotation Tool (CVAT[1]). This platform employs the bounding box labelling technique, which enables the naming (classification) and locating of objects in the video frame required for machine learning-based object detection. The bounding boxes containing the BSL location and classification information are referred to as labels and are utilized for training the deep learning-based object detector, Faster R-CNN, for performing detection. We used CVAT to label 2063 images containing BSLs from three different vessels, each containing images recorded during the day and at night to train the Faster R-CNN model and split the total number of labelled images containing training (80 %), validation (10 %) and testing (10 %) for all experiments. This split ratio is a standard practice in the machine learning community for training deep learning models. There were very

[1] Available at: `https://github.com/opencv/cvat`.

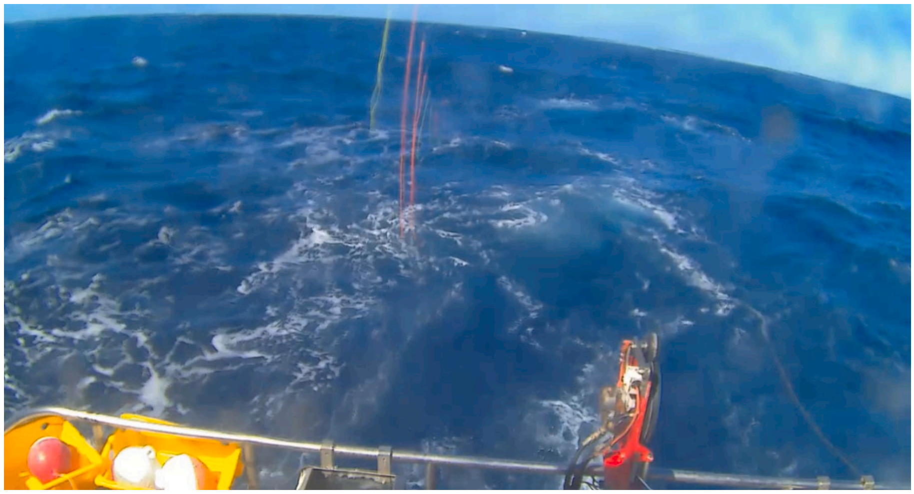

Fig. 1. A deployed Bird Scaring Line (BSL) present at the centre of the image has orange and yellow coloured streamer lines.

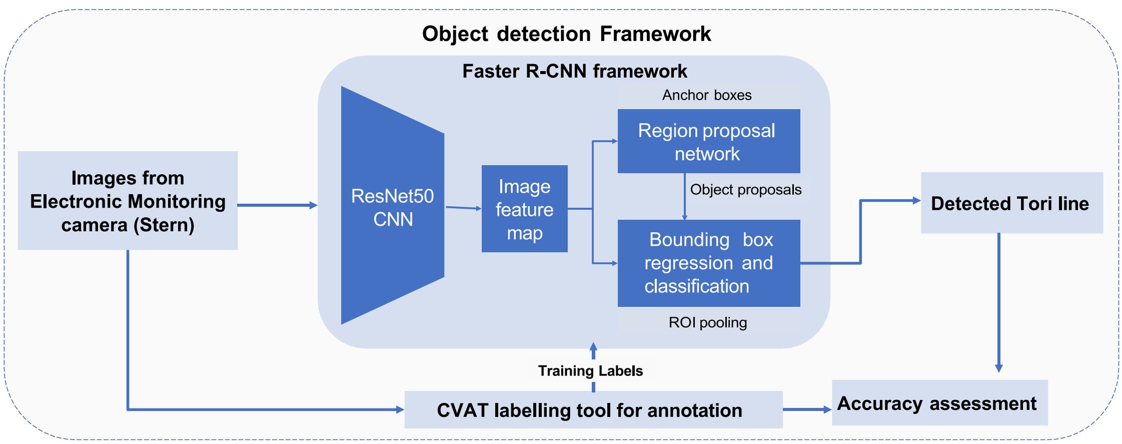

Fig. 2. The framework of the proposed approach where we detect the BSLs with deep learning-based object detector and perform the accuracy evaluation. The BSL detection approach utilized the Faster R-CNN framework, which comprises a convolutional neural network (CNN), specifically ResNet-50, to extract features from the entire image. These features were then used by a region proposal network to generate regions of interest where the object may be located within the image. Following this, object localization and classification were performed based on the proposed regions and the extracted image features.

Table 1
Summary of the labelled images from three different vessels 1–3. Vessel 3 data (V3-2019-DN) was not used for training any model and was explicitly used for testing inter-vessel detection generalization.

Vessel number	Vessel name	Year	Day	Night	BSL present	BSL absent
1	V1-2018-D	2018	447	–	959	19
1	V1-2019-D	2019	75			
1	V1-2019-N	2019		456		
2	V2-2019-DN	2019	407	513	903	17
3	V3-2019-DN	2019	80	85	152	13
Total labelled images:	2063					

few frames for each vessel where the BSL were absent, and they corresponded to the end of the fishing activities where the BSL were retracted. We trained Faster R-CNN on the training set and identified the best model during training using the validation set, where training refers to the use of human-labelled images to teach the Faster R-CNN model to detect BSLs. We then independently tested the model on the testing dataset that was not used for training.

2.2. *Object detection using deep learning*

Deep learning (LeCun et al., 2015) is a subset of the machine learning algorithms that are composed of multiple processing layers that learn the hierarchy of the high-dimension data automatically. In the last decade, deep learning has achieved state-of-the-art performance for many tasks including image recognition and classification, and has been applied to solve many challenging problems, where conventional machine learning techniques perform poorly (LeCun et al., 2015). One of the most prominent application areas of deep learning is visual object detection.

Object detection (Zou et al., 2023) is the process of localising objects of interest on the image and finds many practical applications including video surveillance, self-driving cars and healthcare. Object detection is a two-step process: object location followed by object classification. The first step is to identify whether any objects are present in the image or not. Usually, an object detector identifies several hundreds of such objects in a single image as potential candidates for the object of interest. In the second step, the object detector 'classifies' those detections as whether the detected object is actually the object of interest (BSL in this case). After the classification, the precision is calculated based on the overlap of the bounding boxes as predicted by the object detector by comparing it with the bounding box annotation performed by a 'human expert'. This means that not only do we see whether the object is correctly classified as BSL or not, but we also see whether it is located correctly on the image. The higher deviation between the predicted location and the human-annotated location of the BSL can be considered an error. The error is usually defined in terms of the overlap of the predicted box and the human-annotated box, and if it is over 50 %, we consider it as a correct detection. This overlap is often referred to as the Intersection over Union (IoU).

Faster R-CNN (Ren et al., 2015) is a popular object detection framework that has been widely used in the computer vision community for many challenging tasks. Faster R-CNN utilizes a deep Convolutional Neural Network (CNN) architecture, as depicted in Fig. 2. CNNs are another form of deep learning algorithms that are inspired by the visual cortex of the human brain and contain several layers that can understand complex image features, and they are successfully applied for several image classification tasks. Each layer contains learnable 'weights' for the filters that convoluted the whole image and generate a 'feature map', which is essentially a vector containing the output of the convolutions. These feature maps are input to the deeper layers. This architecture helps the CNNs understand the increasing complexity of the image data and the high-level context of the objects in the images.

Unlike earlier methods that employed CNNs only for classification, Faster R-CNN predicts objects in images through a Region Proposal Network (RPN), which is a two-layered network built on the top of the convolutional feature map. The RPN uses a sliding window placed over the feature map, anchored at multiple scales with varying aspect ratios to capture objects of different sizes. The multi-scale approaches consider the image features of the same image at different scales for adding more context to image understanding. The anchor boxes are centred on the sliding window and generate region proposals at different spatial locations, which are projected onto the corresponding spatial part of the feature map. The regions are then extracted using Region of Interest (ROI) pooling, and the regression provides more precise localization relative to the sliding window position. The RPN does not add to the overall computation of the network.

2.3. Implementation details

For the detection experiments, we employed a Faster R-CNN model, which utilized a ResNet50 backbone pre-trained on the COCO (Lin et al., 2014) dataset. We utilized PyTorch[2] deep learning libraries to train it on a high-performance computing workstation that hosted an NVIDIA TESLA P100 Graphics Processing Unit (GPU) with 16 GB of RAM. The training took approximately an hour for the largest dataset that was used in Experiment 3. NVIDIA is the leading manufacturer of GPU cards worldwide, and TESLA is a GPU product manufactured by NVIDIA for high-performance computing.

2.4. Detection evaluation

The evaluation of multi-class object detection performance typically employs the metrics of "precision" and "recall," as the detector must not only correctly classify objects but also accurately locate them within the image. However, in the specific case of detecting BSLs, which typically involves the detection of a single object, we utilized solely the precision of predicted bounding boxes to assess prediction accuracy. Fig. 3 presents the procedure for evaluating the accuracy of predicted bounding boxes generated by an object detector. In particular, a predicted box must possess an IoU of over 50 % to be considered a true positive, whereas an IoU below this threshold is classified as a false positive. Thus, precision can be calculated as follows:

$$Precision = \frac{true\ \ positives}{true\ \ positives + false\ \ positives} \tag{1}$$

3. Experiments and results

3.1. Experiments

Vessel 1 comprises day footage obtained over multiple years (2018 and 2019), along with night footage from a single year (2019). Meanwhile, Vessel 2 and Vessel 3 contain day and night data from a single year (2019), see Table 1. To gain insights into the performance of the trained models in various scenarios, we conducted four experiments involving training the Faster R-CNN model with different datasets and assessed the precision of the trained models in each case. We elucidate these experiments in the following section. Table 2 summarizes the experiments and their results, while Fig. 4 displays the detection outcomes for different vessels in these experiments.

3.1.1. Experiment 1

For this experiment, a Faster R-CNN model was trained with Vessel 1 day data from 2018 (V1-2018-D) and tested with Vessel 1 day and night data from 2019 (V1-2019-D and V1-2019-N). The trained model performed exceptionally well on V1-2019-D data, with a precision of 0.88 as shown in Table 2. Notably, the model also achieved a high precision of 0.76 while testing with night-time data from V1-2019-N. These findings demonstrate the model's robustness to weather and illumination changes for the same vessel.

However, when the same model was tested on data from a different vessel (Vessel 2: V2-2019-DN) containing day and night images, a decrease in precision was observed (0.64). Although the model trained with V1-2018-D data was robust to weather and illumination changes, it was sensitive to the viewpoint of the camera on another vessel. This is illustrated in Fig. 4, where the camera viewpoint difference between the two vessels resulted in more of the deck being visible in Fig. 4(b) than in Fig. 4(a). This limits the generalizability of the algorithm since there will be a significant drop in detection accuracy for different vessels. Additionally, Fig. 4 shows that the detector was less confident while making predictions during the night compared to the day (as evidenced by the confidence scores on the top of the figure). Thus, the subsequent experiments aim to explore methods to improve the detections when using other vessels.

3.1.2. Experiment 2

In this experiment, we explored whether the Faster R-CNN model trained with data from only one vessel (day and night-time) performed better than the model trained with only day data. The test data for this experiment is completely unseen and untrained data was acquired from a different vessel. For this experiment, we trained a Faster R-CNN model with day and night data from Vessel 1 for both years by merging V1-2018-D, V1-2019-D and V1-2019-N datasets containing 920 images. We tested this Faster R-CNN model with day and night data of Vessel 2 (V2-2019-DN) and observed that the Faster R-CNN model trained on the combined Vessel 1 dataset for both years performed better than the model trained with only V1-2018D data (Experiment 1). The improved precision (0.70 vs 0.64) indicates that including night data for training does improve the detection accuracy while being tested with data from another vessel, i.e., V2-2019-DN. However, the achieved precision is low compared to the precision observed in Experiment 1 (0.88 and 0.76) for

[2] PyTorch is a deep learning framework consisting of several libraries for image processing and classification. We used mmdetection public repository available at https://github.com/open-mmlab/mmdetection.

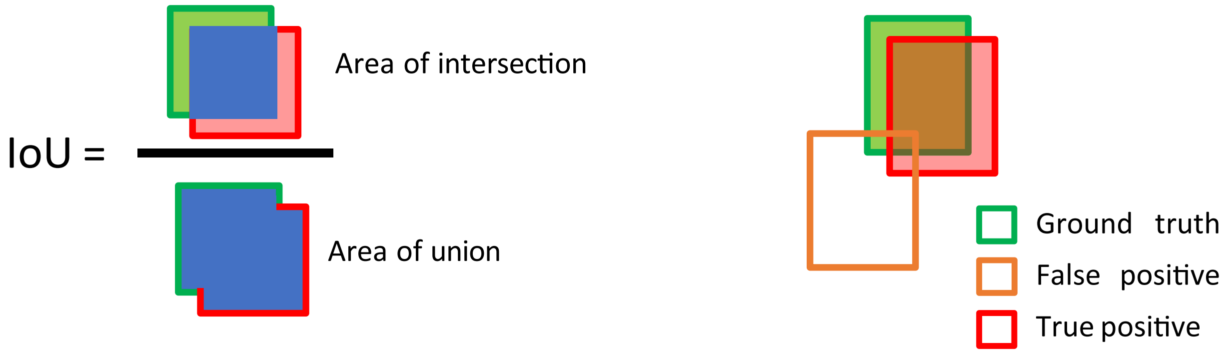

Fig. 3. The accuracy evaluations of predicted bounding boxes in relation to the "ground-truth" bounding box, where ground-truth refers to the correct position of the bounding boxes. On the left, the precision of the bounding boxes is defined as the "intersection over union" (IoU), which is the ratio of the "area of intersection" to the "area of union" between the predicted and ground-truth bounding boxes, and is widely regarded as the accepted standard in object detection (Ren et al., 2015). The blue regions in the numerator and denominator represent the areas of intersection and union, respectively. When the IoU ratio is greater than 50 %, the predicted box is considered a true positive; otherwise, it is a false positive. On the right, the green, orange, and red boxes represent the ground-truth, true positive, and false positive, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
The training and testing cases for object detection with Faster R-CNN and their respective precisions.

Exp	Faster R-CNN train dataset vessel by year, Day/ Night (no. in brackets)	Test dataset - vessel by year, Day/ Night	Precision
1	V1-2018-D (447)	V1-2019-D	0.88
		V1-2019-N	0.76
		V2-2019-DN	0.64
2	V1-2018-D + V1-2019-D + V1-2019-N (920)	V2-2019-DN	0.70
3	V1-2018-D + V1-2019-D + V1-2019-N + V2-2019-DN (1898)	V3-2019-DN	0.87
4	V1-2018-D + V1-2019-D + V1-2019-N + V2-2019-DN (950)	V3-2019-DN	0.82

the Faster R-CNN model trained and tested with the same vessel data. We believe the lower precision results from a lack of data diversity, where the Faster R-CNN model has only "seen" images of Vessel 1.

3.1.3. *Experiment 3*

We investigated whether training a Faster R-CNN model with data from multiple vessels could improve detection precision when tested on completely unseen data from another vessel.

To accomplish this, we created a combined dataset by merging Vessel 1 and Vessel 2 data (V1-2018-D + V1-2019-D + V1-2019-N + V2-2019-DN), resulting in a more diverse dataset of day and night images captured in different conditions and viewpoints containing 1898 images. This diverse dataset contained multiple scenes, illuminations, weather conditions, and day and night conditions. We trained a Faster R-CNN model using this combined dataset and tested its performance on Vessel 3 data (V3-2019-DN), which contained day and night data that were not used in any training and were solely used for testing purposes.

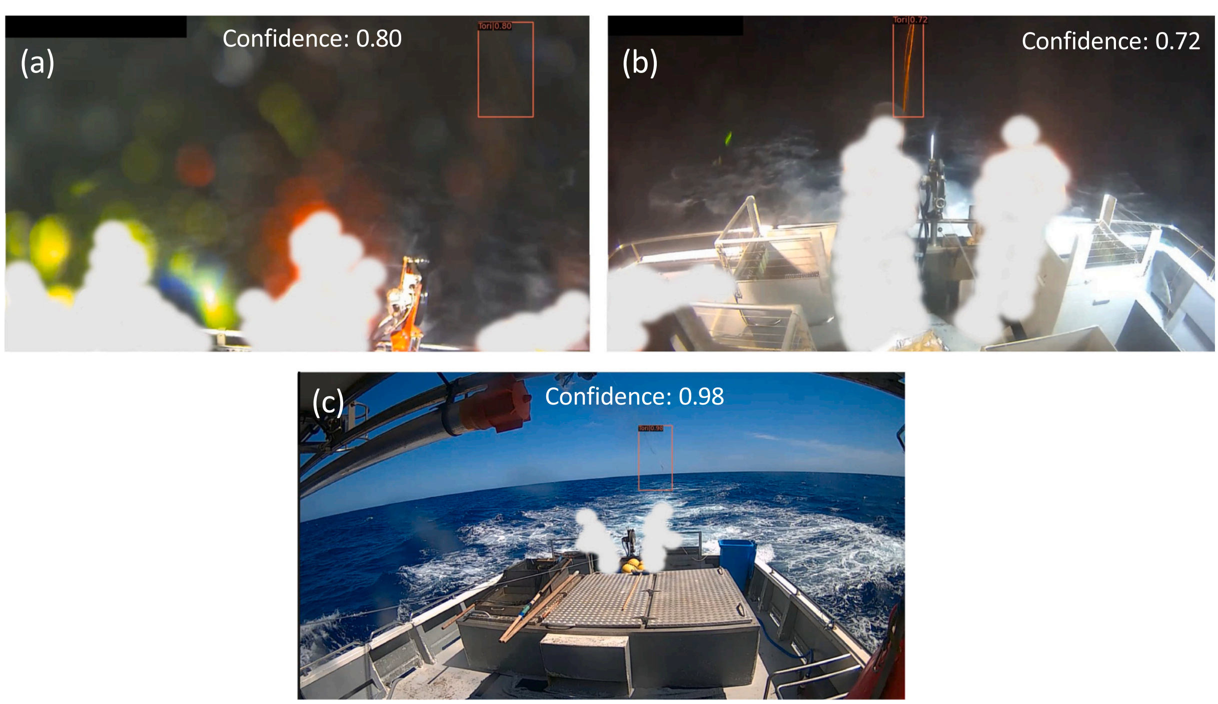

Fig. 4. BSL detections during the day and night. Visualisation of BSL detections by the Faster R-CNN model for (a) Vessel 1 (2019) Night (b) Vessel 2 (2019) Night and (c) Vessel 3 (2019) Day. The bounding boxes show the confidence score of the detection (between 0 and 1).

Our findings reveal that the Faster R-CNN model, which was trained with the combined data (V1-2018-D + V1-2019-D + V1-2019-N + V2-2019-DN), achieved a precision of 0.87. This means the model outperformed the Faster R-CNN model trained with the combined data of Experiment 2 containing only V1-2018D + V1-2019-D + V1-2019-N data (precision 0.70). Moreover, the precision of the Faster R-CNN model trained with the combined data (V1-2018D + V1-2019N + V2-2019-DN) in this experiment was comparable to the precision of the model trained and tested with only day-time data from the same vessel (training with V1-2018-D and testing with V1-2019-D images) in Experiment 1, which achieved a precision of 0.88. It is noteworthy that a precision of 0.88 is considered highly accurate. For comparison, Khokher et al. (2022) reported the best average precision of 0.81 for detecting the 'Sting Ray' class using the Cascade R-CNN object detector for performing multi-class fish detection from EM videos, where Cascade R-CNN is another popular multi-stage object detection framework. Taken together, the results of our experiments indicate that the additional images and diverse scenes from Vessel 2, which contained 920 images of both day and night, helped train a Faster R-CNN model that is robust to viewpoint changes and is not susceptible to changes in surroundings.

3.1.4. Experiment 4

The combined dataset (V1-2018-D + V1-2019-D + V1-2019-N + V2-2019-DN) contained approximately twice the number of images (1898) compared to the number of images used in Experiment 2 (920). Therefore, to investigate whether the improved precision was a result of the increased number of images from the second vessel, or it was a result of using the image diversity sourced from the two vessels, we created a subset of the combined dataset (V1-2018-D + V1-2019-D + V1-2019-N + V2-2019-DN) by picking 950 images randomly from the set of all images. This data subset contained an approximately equal number of images from both vessels and a comparable number of images with Experiment 2. We trained another Faster R-CNN model with the reduced number of images and tested it with Vessel 3 data, and we report the results in Table 2. The results indicate that a precision of 0.82 can be achieved with 950 images from two vessels when being tested on the third vessel, indicating that data diversity is a more important factor compared to the additional number of training images.

3.2. Error analysis

We have identified several instances where the object detector exhibited poor performance, as depicted in Fig. 5. In these instances, the detector failed to detect a BSL that was present. Most of the training data consisted of images in which the BSLs were clearly visible, appearing vertically and grouped. It is noteworthy that the trained model has not detected a BSL in a test image where the BSL is underwater, as seen in Fig. 5(a), while during the manual annotation the BSL was labelled as a BSL, as seen in the green dotted line.

In addition, some false positives were detected. For example, BSLs were detected when there was no BSL present due to light reflection caused by droplets on the camera lens (Fig. 5(b)) or because of the similarity of the BSL streamers with other fishing gear on the fishing deck (Fig. 5(c)). A further false positive occured from the detection of containment barriers with vertical strings resembling a group of straight streamer lines. Moreover, some of the BSLs that were barely visible were not detected because of the lack of contrast with the background. In fact, to create ground-truth annotations (i.e., labelled images) of such faint BSLs, we had to consult adjacent images since identifying them from an independent frame was impossible.

4. Discussions

When deployed properly and consistently BSLs can markedly reduce seabird bycatch (Yokota et al., 2011). In fisheries such as trawls for which BSLs are not as effective, compliance with alternative methods of seabird deterrent (such as bafflers; Koopman et al., 2018) remains an important fishery regulatory requirement, and the detection methods advocated in this paper could equally be applied to these devices. While

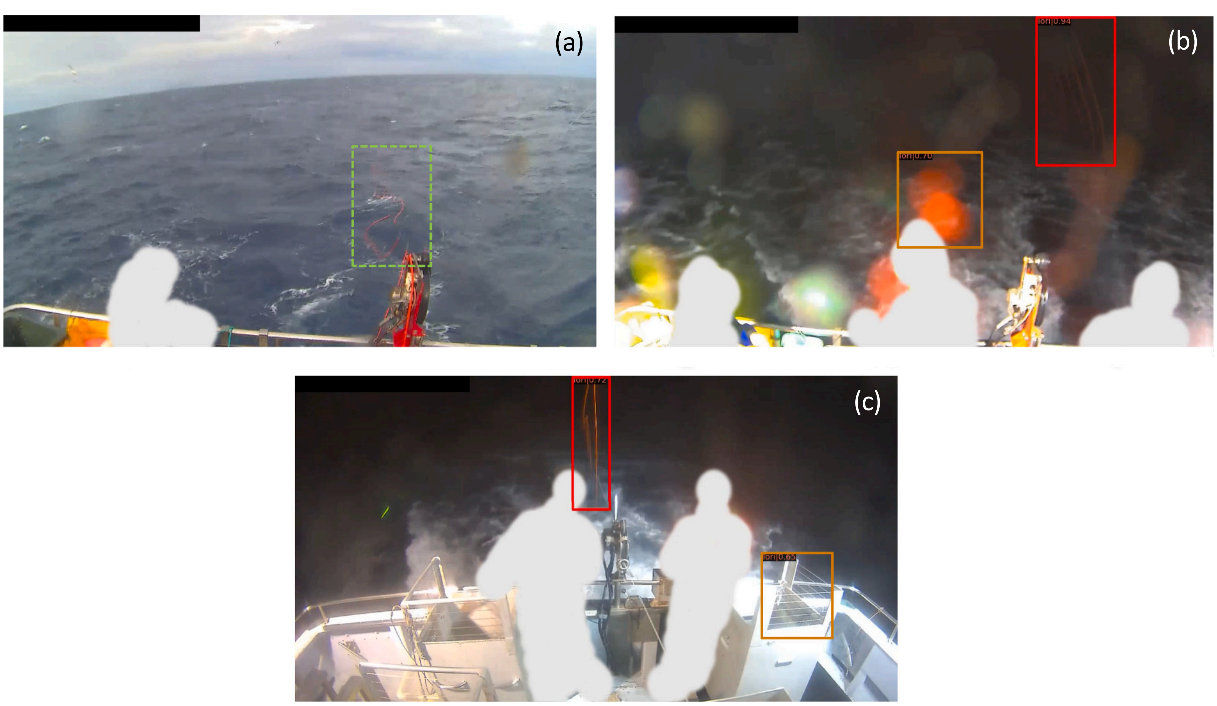

Fig. 5. Examples of incorrect BSL detections. (a) Missed detection near the middle right of the image (Vessel 1 (2019) Day) shown in green dotted line, (b) False positives due to bias towards orange colour (shown in orange) and the correct detection (shown in red), and (c) False positive due to similarity of other fishing gear with BSLs (shown in orange) and the correct detection (shown in red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

EM cameras have the potential to replace human on-board observers, thousands of hours of footage from multiple cameras are generated. A thorough review of EM footage through manual human viewing is expensive and time-consuming. The continued success of EM will likely depend on systems that provide a cost-efficient review of video footage, with minimal manual handling and auditing, without compromising management objectives. Computer vision and deep learning techniques for object detection, in addition to the availability of enormous volumes of video footage from EM, provide the opportunity to achieve both cost efficiency and robust review.

Deep learning models require a large volume of images for training. However, when training data is scarce, the model can be trained with a publicly available dataset that learns most of the pattern of the data. This process is referred to as pre-training. Subsequently, a pre-trained model is further trained on a smaller dataset to adapt to the specific task, which is referred to as fine-tuning. Since the number of annotated images was low (approximately 2000), we used a pre-trained network and fine-tuned it for BSL identification. The BSL detection and prediction experiments using the Faster R-CNN object detection neural network show that inter-vessel detection with a precision of 0.87 can be achieved, even with a small number (approximately 2000) of diverse images. Despite these shortcomings, the final trained model contains no image information and is not restricted to open use due to privacy concerns.

The results in this paper demonstrate that deep learning for EM is a powerful machine learning technique that can automatically detect the presence of a BSL, contributing to compliance checking and improved fleet management. Our model performs well on day and night data from multiple vessels, promising scalability of use across commercial longline fleets. Automatic review not only provides a secure way to use sensitive fisheries video data but also provides assurance on the effective use of EM cameras and footage. Therefore, to enhance and encourage sustainable fisheries, we recommend that fisheries management should leverage AIML approaches to improve detections and compliance across more vessels and fisheries. Some more experimentation is required as a part of future work for creating an operational system, suggesting we are at the start of the Technology Demonstration Stage.

Although the system achieves a commendable (87 %) precision, there might still be a requirement for manual observers to review the footage from time to time, especially for the cases where the BSLs are not being detected during fishing operations. However, cases where the BSLs are wrongly detected when they are not in use pose a risk since the manual observer will not be asked to review the footage in such cases. One of the easiest ways to tackle this challenge will be to use the confidence scores of the predictions. In addition to detecting the BSLs, the model also provides an estimate of how confident it is for a particular detection. Currently, we have not used the confidence values for further analysis. We also observe that in the false positive cases (5(b) - (c)), the confidence of the model is low, in the range of 0.65–0.7. This is also the case for some of the challenging correct predictions. By using a fixed score (for instance greater than 0.9 confidence), we could reduce the false positives significantly. This operation will increase some false negatives which can then be reviewed by human reviewers. Additionally, restricting the search space of the BSLs to the water and excluding the deck is likely to reduce incorrect detections, and image segmentation using deep learning models can be performed in parallel to reduce false positives.

Studies in other domains (Huang et al., 2019; Sheng et al., 2020; Li et al., 2017) which used a similar number of images (1000−2000) for training, reported a similar average precision (in the range of 46–90 %, depending upon the complexity of the task). However, there are some studies that have achieved higher average precision with more images. For instance, Chen et al. (2015) reports an average precision of 93 % on an easy car detection dataset (KITTI) containing approximately 6000 images (including training and testing images). The structural differences between the vessels are significant, and combined with illumination variances, different practices of the fishers, and several other environmental factors, the appearance of the BSL and the background can vary largely. The results of the experiments show that using images of three different vessels adds to the diversity of the dataset and explains the improvement in precision.

We believe the overall precision of the deep learning model can be improved by using additional images, especially coming from many more vessels. In fact, instead of using pre-trained deep learning models, we can consider training the CNNs entirely on EM video data to make it more robust to noise. While the collection of additional images of different vessels can be challenging due to privacy concerns, a viable future research question is to explore whether increasing the number of training images with more vessels (thereby adding more diversity) will improve the detection precision.

Differently, to increase the diversity of BSL image datasets, diffusion-based image generation is a promising area of future applications. This method can generate images of challenging-to-obtain objects (Rombach et al., 2022) by iteratively transforming an initial random noise image into a coherent image that more accurately reflects the underlying structure of the object being detected. This technique may be combined with machine learning algorithms to improve the accuracy of BSL detection.

An additional constraint is the low-resolution and quality of the video footage. While high-resolution and uncompressed images are preferred for tasks such as BSL detection, they come with several challenges. Currently, in Australia, the video data from Electronic Monitoring is used to independently verify fishers' logbook information to perform stock assessment and surveillance (AFMA, 2024). The most cost-effective infrastructure of the video cameras is in operation for this task which limits the resolution of the image capture. While some camera infrastructures might allow the capture of high-resolution images, the framerate (rate of image capture per second) of the videos is reduced significantly, making it unfit for the task (van Helmond et al., 2020). Hardware that can handle such large image sizes at acceptable framerates is expensive and will require higher-capacity devices for storing the data.

Another observation is that the footage without BSL is limited to the start and end of the sequence. The process of retraction of the BSLs is exactly opposite to the deployment. We believe this condition will not affect the overall performance of the system, although, during the transition, there might be some ambiguity, both during manual annotation of the images and during the detection using the CNN. The BSLs are expected to be detected after they are deployed, and conversely, they are expected not to be detected after their retraction. We can use this fact to further resolve the ambiguities.

The trained Faster R-CNN model can be used out-of-the-box on servers for offline processing of the EM video data. For real-time operation on the boat, the proposed framework can be deployed directly on vessels and run in near real-time on Edge AI GPU platforms, such as NVIDIA Jetson (the leading platform for autonomous machines and embedded applications with onboard GPUs). There are several examples of the application of Edge AI technology, for instance in real-time parking surveillance (Ke et al., 2020) and real-time Shark detection (Sharma et al., 2022). These hardware solutions are readily available, are cheap and would not hamper the day-to-day operations of the fishing vessels. This would enable a wide range of future applications that include automated alerts to crew and on-land regulators to ensure BSLs are deployed. These facilities would improve the implementation of successful real-time vessel compliance monitoring systems, thus improving crew confidence that they are adhering to regulations in situ, and consequently increasing community confidence in fisheries. Currently, a logbook is maintained by the fishing operators which is then tallied manually with the video footage for fish stock management purposes. Therefore, the proposed camera-based Edge AI warning systems will not eliminate the requirement of storing and transferring video footage due to surveillance and regulation purposes. The advantages of these camera-based Edge AI warning systems include the real-time operation, the compactness and the low cost of these systems. On the contrary, these systems will need additional infrastructure, which needs to be installed and maintained. Also, the trained deep learning models need to be deployed on these systems.

Moreover, further research could investigate the efficacy of BSLs in deterring seabirds during particular fishing activities. This could be achieved by comparing bird counts (either by human observers or AIML algorithms, such as the work of Akçay et al. (2020)) before, during and after BSL deployment.

Several other recent advancements in deep learning can be utilized to tackle additional known challenges. For instance, the present domain adaptation approaches that enhance the generalization of object detection on novel data from other vessels can be explored. Furthermore, Generative Adversarial Networks (GANs) can be employed to eliminate the effects of haze, noise, and blurring caused by water droplets on the camera lens (Ren et al., 2019; Uricar et al., 2021) or to compensate for the lack of contrast in low-light environments (Lore et al., 2017). GANs (Goodfellow et al., 2014) are a kind of unsupervised deep learning algorithms that can automatically learn the patterns of data in a way that can be used to generate novel example data that is very close to the original data, and find many applications for image-to-image translation tasks. Another viable research direction related to deep learning models would be exploring alternative state-of-the-art object detectors for evaluating their detection accuracy and applicability for the task.

5. Conclusion

In this study, we performed BSL detection using a deep learning object detector called Faster R-CNN on complex EM data from longline fishing vessels. The EM footage was obtained from multiple vessels under variable conditions. The evaluation of BSL detection concludes that the object detector can be used to detect BSLs with a precision of 87 % under various test conditions. The results reveal the promising future of automated detection tools using AIML and fishery EM video data that can provide timely and accurate information on mitigation device use, and thereby ensure fisheries are managed in a cost efficient and sustainable manner.

CRediT authorship contribution statement

Debaditya Acharya: Writing – original draft, Visualization, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Muhammad Saqib:** Writing – original draft, Visualization, Validation, Methodology, Investigation, Conceptualization. **Carlie Devine:** Visualization, Software, Methodology. **Candice Untiedt:** Writing – original draft, Supervision, Conceptualization. **L. Richard Little:** Writing – review & editing. **Dadong Wang:** Writing – review & editing, Resources. **Geoffrey N. Tuck:** Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgment

This work was partially funded by CSIRO’s Machine Learning and Artificial Intelligence Future Science Platform. The authors are grateful for the assistance of the Australian Fisheries Management Authority for the provision of footage and Tamre Sarhan (AFMA), Simon Boag (Atlantis Fisheries Consulting Group) and David Ellis (Tuna Australia) for review of earlier versions of the manuscript.

References

ACAP (2021). ACAP Review of mitigation measures and Best Practice Advice for Reducing the Impact of Pelagic Longline Fisheries on Seabirds. https://www.acap.aq/bycatch-mitigation/mitigation-advice/3949-acap-2021-trawl-mitigation-review-and-bpa/file. [Online; accessed 19-Jun-2024].

AFMA (2023). Eastern Tuna and Billfish Fishery Management Arrangements Booklet. https://www.afma.gov.au/sites/default/files/final_2022_etbf_management_arrangements_booklet.pdf. [Online; accessed 19-Jan-2024].

AFMA (2024). Electronic Monitoring Program. https://www.afma.gov.au/fisheries-management/monitoring-tools/electronic-monitoring-program. [Online; accessed 18-June-2024].

Akçay, H.G., Kabasakal, B., Aksu, D., Demir, N., Öz, M., Erdoğan, A., 2020. Automated bird counting with deep learning for regional bird distribution mapping. Animals 10, 1207.

Anderson, O.R., Small, C.J., Croxall, J.P., Dunn, E.K., Sullivan, B.J., Yates, O., Black, A., 2011. Global seabird bycatch in longline fisheries. Endanger. Species Res. 14, 91–106.

Arata, J., Xavier, J.C., 2003. The diet of black-browed albatrosses at the diego ramirez islands, Chile. Polar Biol. 26, 638–647.

Avery, J.D., Aagaard, K., Burkhalter, J., Robinson, O.J., 2017. Seabird longline bycatch reduction devices increase target catch while reducing bycatch: a meta-analysis. J. Nat. Conserv. 38, 37–45.

Baker, G. B., & Robertson, G. (2018). Management of seabird bycatch leads to sustainable fisheries and seabird populations. *Recovering Australian threatened species: a book of hope. CSIRO Publishing, Melbourne,* (pp. 23–31).

Baker, G.B., Gales, R., Hamilton, S., Wilkinson, V., 2002. Albatrosses and petrels in Australia: a review of their conservation and management. Emu-Austral Ornithology 102, 71–97.

Brothers, N., 1991. Albatross mortality and associated bait loss in the Japanese longline fishery in the Southern Ocean. Biol. Conserv. 55, 255–268.

Brown, C.J., Desbiens, A., Campbell, M.D., Game, E.T., Gilman, E., Hamilton, R.J., Heberer, C., Itano, D., Pollock, K., 2021. Electronic monitoring for improved accountability in western Pacific tuna longline fisheries. Mar. Policy 132, 104664.

Bull, L.S., 2007. Reducing seabird bycatch in longline, trawl and gillnet fisheries. Fish Fish. 8, 31–56.

Bull, L.S., 2009. New mitigation measures reducing seabird by-catch in trawl fisheries. Fish Fish. 10, 408–427.

Chen, X., Kundu, K., Zhu, Y., Berneshawi, A.G., Ma, H., Fidler, S., Urtasun, R., 2015. 3d object proposals for accurate object class detection. Advances in neural information processing systems 28.

Commonwealth of Australia (2018). Threat Abatement Plan for the incidental catch (or bycatch) of seabirds during oceanic longline fishing operations. https://www.antarctica.gov.au/about-antarctica/environment/plants-and-animals/threat-abatement-plan-seabirds/. [Online; accessed 19-Jan-2024].

Cooper, J., Baker, G.B., Double, M.C., Gales, R., Papworth, W., Tasker, M.L., Waugh, S. M., 2006. The agreement on the conservation of albatrosses and petrels: rationale, history, progress and the way forward. Marine Ornithology 34, 1–5.

Croxall, J.P., Butchart, S.H., Lascelles, B., Stattersfield, A.J., Sullivan, B., Symes, A., Taylor, P., 2012. Seabird conservation status, threats and priority actions: a global assessment. Bird Conservation International 22, 1–34.

Da Rocha, N., Oppel, S., Prince, S., Matjila, S., Shaanika, T.M., Naomab, C., Yates, O., Paterson, J.R., Shimooshili, K., Frans, E., et al., 2021. Reduction in seabird mortality in Namibian fisheries following the introduction of bycatch regulation. Biol. Conserv. 253, 108915.

Emery, T.J., Noriega, R., Williams, A.J., Larcombe, J., 2019. Measuring congruence between electronic monitoring and logbook data in australian commonwealth longline and gillnet fisheries. Ocean & Coastal Management 168, 307–321.

Gales, R. (1998). Albatross populations: status and threats. *Albatross biology and conservation,* (pp. 20–45).

Gilman, E., Legorburu, G., Fedoruk, A., Heberer, C., Zimring, M., Barkai, A., 2019. Increasing the functionalities and accuracy of fisheries electronic monitoring systems. Aquat. Conserv. Mar. Freshwat. Ecosyst. 29, 901–926.

Gilman, E., Chaloupka, M., Ishizaki, A., Carnes, M., Naholowaa, H., Brady, C., Ellgen, S., Kingma, E., 2021. Tori lines mitigate seabird bycatch in a pelagic longline fishery. Rev. Fish Biol. Fish. 31, 653–666.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. Advances in neural information processing systems 27.

Huang, H., Zhou, H., Yang, X., Zhang, L., Qi, L., Zang, A.-Y., 2019. Faster r-cnn for marine organisms detection and recognition using data augmentation. Neurocomputing 337, 372–384.

Jiménez, S., Domingo, A., Winker, H., Parker, D., Gianuca, D., Neves, T., Coelho, R., Kerwath, S., 2020. Towards mitigation of seabird bycatch: large-scale effectiveness of night setting and Tori lines across multiple pelagic longline fleets. Biol. Conserv. 247, 108642.

Junior, V., 1991. Seabirds mortality on longline fishing for tuna in southern Brazil. *Ciencia e cultura(Sao Paulo)*. Sao Paulo 43, 388–390.

Ke, R., Zhuang, Y., Pu, Z., Wang, Y., 2020. A smart, efficient, and reliable parking surveillance system with edge artificial intelligence on iot devices. IEEE Trans Intell Transp Syst 22, 4962–4974.

Khokher, M.R., Little, L.R., Tuck, G.N., Smith, D.V., Qiao, M., Devine, C., O'Neill, H., Pogonoski, J.J., Arangio, R., Wang, D., 2022. Early lessons in deploying cameras and artificial intelligence technology for fisheries catch monitoring: where machine learning meets commercial fishing. Can. J. Fish. Aquat. Sci. 79, 257–266.

Klaer, N., Polacheck, T., 1998. The influence of environmental factors and mitigation measures on by-catch rates of seabirds by Japanese longline fishing vessels in the Australian region. Emu 98, 305–316.

Koopman, M., Boag, S., Tuck, G.N., Hudson, R., Knuckey, I., Alderman, R., 2018. Industry-based development of effective new seabird mitigation devices in the southern australian trawl fisheries. Endanger. Species Res. 36, 197–211.

Lawrence, E., Wise, B., Bromhead, D., Hindmarsh, S., Barry, S., Bensley, N., Findlay, J., 2006. Analyses of AFMA Seabird Mitigation Trials, 2001 to 2004 (Bureau of Rural Sciences).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature, 521,* 436–444.

Li, J., Qu, C., & Shao, J. (2017). Ship detection in sar images based on an improved faster r-cnn. In *2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA)* (pp. 1–6). IEEE.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13* (pp. 740–755). Springer.

Løkkeborg, S., Robertson, G., 2002. Seabird and longline interactions: effects of a bird-scaring streamer line and line shooter on the incidental capture of northern fulmars Fulmarus glacialis. Biol. Conserv. 106, 359–364.

Lore, K.G., Akintayo, A., Sarkar, S., 2017. LLNet: a deep autoencoder approach to natural low-light image enhancement. Pattern Recogn. 61, 650–662.

Maree, B.A., Wanless, R.M., Fairweather, T., Sullivan, B., Yates, O., 2014. Significant reductions in mortality of threatened seabirds in a south African trawl fishery. Anim. Conserv. 17, 520–529.

Melvin, E.F., Guy, T.J., Read, L.B., 2013. Reducing seabird bycatch in the south African joint venture tuna fishery using bird-scaring lines, branch line weighting and nighttime setting of hooks. Fish. Res. 147, 72–82.

Melvin, E.F., Guy, T.J., Read, L.B., 2014. Best practice seabird bycatch mitigation for pelagic longline fisheries targeting tuna and related species. Fish. Res. 149, 5–18.

Montevecchi, W. A. (2023). Interactions between fisheries and seabirds: Prey modification, discards, and bycatch. In *Conservation of Marine Birds* (pp. 57–95). Elsevier.

Murray, T., Bartle, J., Kalish, S., Taylor, P., 1993. Incidental capture of seabirds by Japanese southern bluefin tuna longline vessels in New Zealand waters, 1988-1992. Bird conservation international 3, 181–210.

Otley, H.M., REID, T.A., POMPERT, J., 2007. Trends in seabird and Patagonian toothfish Dissostichus eleginoides longliner interactions in Falkland Island waters, 2002/03 and 2003/04. Mar. Ornithol. 35, 47–55.

Paterson, J.R., Yates, O., Holtzhausen, H., Reid, T., Shimooshili, K., Yates, S., Sullivan, B. J., Wanless, R.M., 2019. Seabird mortality in the Namibian demersal longline fishery and recommendations for best practice mitigation measures. Oryx 53, 300–309.

Phillips, R.A., Gales, R., Baker, G., Double, M., Favero, M., Quintana, F., Tasker, M.L., Weimerskirch, H., Uhart, M., Wolfaardt, A., 2016. The conservation status and priorities for albatrosses and large petrels. Biol. Conserv. 201, 169–183.

Qiao, M., Wang, D., Tuck, G.N., Little, L.R., Punt, A.E., Gerner, M., 2021. Deep learning methods applied to electronic monitoring data: automated catch event detection for longline fishing. ICES J. Mar. Sci. 78, 25–35.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: towards real-time object detection with region proposal networks. Adv. Neural Inf. Proces. Syst. 28.

Ren, D., Zuo, W., Hu, Q., Zhu, P., & Meng, D. (2019). Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3937–3946).

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10684–10695).

Sato, N., Minami, H., Katsumata, N., Ochi, D., Yokawa, K., 2013. Comparison of the effectiveness of paired and single tori lines for preventing bait attacks by seabirds and their bycatch in pelagic longline fisheries. Fish. Res. 140, 14–19.

Sato, N., Katsumata, N., Yokota, K., Uehara, T., Fusejima, I., Minami, H., 2016. Tori-lines with weighted branch lines reduce seabird bycatch in eastern South Pacific longline fishery. Aquatic Conservation: Marine and Freshwater Ecosystems 26, 95–107.

Sharma, N., Saqib, M., Scully-Power, P., & Blumenstein, M. (2022). Sharkspotter: Shark detection with drones for human safety and environmental protection. *Humanity Driven AI: Productivity, Well-being, Sustainability and Partnership,* (pp. 223–237).

Sheng, B., Zhou, M., Hu, M., Li, Q., Sun, L., Wen, Y., 2020. A blood cell dataset for lymphoma classification using faster r-cnn. Biotechnol. Biotechnol. Equip. 34, 413–420.

Tuck, G.N., Polacheck, T., Croxall, J.P., Weimerskirch, H., 2001. Modelling the impact of fishery by-catches on albatross populations. J. Appl. Ecol. 38, 1182–1196.

Uricar, M., Sistu, G., Rashed, H., Vobecky, A., Kumar, V. R., Krizek, P., Burger, F., & Yogamani, S. (2021). Let's get dirty: GAN based data augmentation for camera lens soiling detection in autonomous driving. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 766–775).

van Helmond, A.T., Mortensen, L.O., Plet-Hansen, K.S., Ulrich, C., Needle, C.L., Oesterwind, D., Kindt-Larsen, L., Catchpole, T., Mangi, S., Zimmermann, C., et al., 2020. Electronic monitoring in fisheries: lessons from global experiences and future opportunities. Fish Fish. 21, 162–189.

Weimerskirch, H., Brothers, N., Jouventin, P., 1997. Population dynamics of wandering albatross Diomedea exulans and Amsterdam albatross D. Amsterdamensis in the Indian Ocean and their relationships with long-line fisheries: conservation implications. Biol. Conserv. 79, 257–270.

Yokota, K., Minami, H., Kiyota, M., 2011. Effectiveness of tori-lines for further reduction of incidental catch of seabirds in pelagic longline fisheries. Fish. Sci. 77, 479–485.

Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE, 111,* 257–276.